The Social Integration of International Migrants: Evidence from the Networks of Syrians in Germany *

Michael Bailey[†]
Drew Johnston[‡]
Martin Koenen[§]
Theresa Kuchler[¶]
Dominic Russel[||]
Johannes Stroebel[**]

Abstract

We use de-identified friendship data from Facebook to study the social integration of Syrian migrants in Germany. We decompose the significant spatial variation in migrants' integration levels into the rate at which Germans befriend their neighbors in general and the particular rate at which they befriend Syrian migrants versus other Germans. We follow the friending behavior of Germans that move across locations to show that both forces are more affected by local institutions and policies than persistent individual characteristics or preferences of local natives. We explore the characteristics of places with higher integration levels, and show that integration courses causally affect place-specific equilibrium integration levels by shifting the rates of Germans befriending Syrians.

JEL Codes: F22, J15, K37, D85
Keywords: Integration, immigration, social networks, place effects

*This draft: July 15, 2023. This research was facilitated through a research consulting agreement between the academic authors and Meta. Bailey is an employee at Meta. We thank Raj Chetty, Ed Glaeser, Nathan Hendren, Nathan Nunn, Larry Katz, Amanda Pallais, Jesse Shapiro and seminar participants at Harvard, Harvard Kennedy School, NYU, Mannheim, Berlin, the University of Copenhagen, Dusseldorf University, European Union, ifo Institute, Norges Bank, the Stanford Institute for Theoretical Economics, the North American Meeting of the Urban Economics Association, the European Meeting of the Urban Economics Association, the International Conference on Computational Social Science, the NSF Conference on Networks and Economics, the European Networks Conference, the EBRD & Kings College Workshop on the Economics and Politics of Migration, the CESifo Venice Summer Institute, and the HUMANS LACEA Network Seminar for feedback. Stroebel thanks the Carnegie Corporation for financial support through its Andrew Carnegie Fellows program.

[†]Meta Platforms, Inc: mcbailey@meta.com.
[‡]Harvard University: drewjohnston@g.harvard.edu
[§]Harvard University: martin_koenen@g.harvard.edu
[¶]Stern School of Business, New York University, NBER, CEPR, cesifo: tkuchler@stern.nyu.edu
[||]Harvard University: drussel@g.harvard.edu
[**]Stern School of Business, New York University, NBER, CEPR, cesifo: johannes.stroebel@nyu.edu

In 2019, there were 272 million international migrants comprising 3.5% of the world's population (United Nations, 2019). The challenge of fostering communities that harmoniously integrate new arrivals with natives has therefore become of increasing importance to policymakers around the globe (e.g., European Commission, 2020; Bundesregierung, 2021). Yet, because of difficulties with measuring social networks using traditional data sources, researchers have long struggled to understand the determinants of the social integration of migrants in their host communities.

In this paper, we use de-identified data from Facebook, a global online social networking service, to study the factors that shape the social integration of newly arriving migrants. We focus on individuals who recently migrated from Syria to Germany. Following the start of the Syrian Civil War, millions of Syrians fled their home country, with about 800,000 of them settling in Germany since 2014. The social and economic integration of these migrants has been a dominant political issue in Germany in the years since, with policy makers attempting to facilitate this integration through a variety of programs. In 2018 alone, for example, the German government spent €2bn on integration courses that teach migrants the German language and provide information on the country's culture and legal system.

While existing work investigates the *economic* integration of Syrians in Germany—with a particular focus on attempts to bring them into the labor force—data challenges have hindered empirical work to understand the *social* integration of these migrants. How frequently and intensely do Syrian migrants interact with local Germans?[1] How does this differ across demographics and locations? Which Germans form social ties with Syrians? And why are some locations better than others at integrating Syrian migrants? Our unique data and research design allow us to answer these questions.

We begin by identifying Syrian migrants as Facebook users who currently live in Germany, but who specified a hometown or high school in Syria in their Facebook profiles, or who previously had a predicted home region in Syria.[2] This simple methodology generates spatial variation in Syrian migrant population shares across German counties (*Kreise*) that closely resembles German administrative data. We also construct a group of users that we call "German natives" based on self-reported profile information, home region predictions, and German language usage.[3] We use these data to measure Syrian migrants' social integration along three key dimensions: (i) friendships between migrants and German natives; (ii) migrants' German language usage; and (iii) migrants' participation in local social groups.

Syrian migrant users have five local German native friends on average, and 30% of them produce German content such as posts or comments on Facebook. Controlling for Facebook usage patterns, younger and male Syrian migrants have higher levels of social integration than others. We also find large spatial heterogeneity in Syrian migrants' social integration across the 401 German counties: an average Syrian migrant living in a 90th percentile county has more than twice as many native German friends as an average Syrian migrant living in a 10th percentile county. These spatial patterns are highly correlated

[1] While there is no single definition of social integration, the concept is often defined by the frequency of interactions of individuals of different groups (e.g., Phillips et al., 2019). This conceptualization of "social integration" is distinct from that of *assimilation* (Berry, 1997), which is defined in terms of cultural identity, and is not the focus of our work.

[2] Estimated home region is determined by a person's information on Facebook, including the stated city on their Facebook profile, and device and connection information (also see Herdağdelen et al., 2016; Chi et al., 2019).

[3] We describe these criteria in detail in Appendix B. When constructing both the "Syrian migrant" and "German native" samples, we do not make any inference about citizenship status or race and ethnicity. Our intent is instead to create samples of users that appear to have lived in Syria but now live in Germany (Syrian migrant sample), or that have lived in Germany for a substantial amount of time and exclusively or primarily use the German language (German native sample).

across our three measures of social integration. We show that these measures pick up true differences in integration levels rather than sampling variation or differences in Facebook usage; they also align with survey-based measures of integration available at higher levels of geographic aggregation.

We analyze whether the observed spatial differences in social integration are the result of selection (e.g., migrants with higher propensities to integrate select to live in certain regions) or whether they correspond to *causal* effects of location. We first argue that the initial allocation of Syrian refugees to locations—a process that is largely random—suggests that spatial differences in refugee integration are likely the causal effects of place. We confirm this interpretation using a mover research design that follows the (relatively few) Syrian migrants who move across German counties. These movers' social integration patterns quickly adjust from those of their origin towards those of their destination counties. We estimate that the vast majority of the observed regional differences in migrants' social integration are indeed due to causal place-based factors rather than migrant characteristics, consistent with prior work exploiting the random assignment of refugees in other countries (e.g., Auer, Egger and Kunz, 2022; Edin, Fredriksson and Åslund, 2003; Beaman, 2012; Damm, 2014).

We next ask: "What is it about certain places that allows migrants to integrate better? Is this explained by fixed preferences of local natives or the institutional settings in those places? And, if the latter, which institutions matter?" Data challenges and the lack of random assignments of natives to locations have precluded prior attempts to estimate the relative importance of natives' preferences in explaining variation in migrants' integration. Our unique panel data on the characteristics and behaviors of Germans who befriend Syrians allows us to make important progress in answering these questions.

We begin by showing that the level of Syrians' social integration in a location can be decomposed into two forces: (i) the rate at which local Germans befriend their neighbors in general (their *general friendliness*), and (ii) Germans' particular friending behavior towards migrants, given by their relative propensity of befriending local Syrians versus other locals (*relative friending*). Put simply, if Germans in a given location are more likely to befriend *all* of their neighbors, including their German ones, they are also more likely to befriend newly arriving migrants. All else equal, this helps migrants' social integration, even if the level of general friendliness is unlikely to be strongly affected by migrants' behavior or integration policies. In addition, Syrian migrants are more likely to be socially integrated when Germans befriend them at rates more similar to those at which they befriend local Germans. We show that both general friendliness and relative friending vary across locations, with differences in relative friending explaining about two-thirds of the spatial variation in the social integration of Syrian migrants.

We next explore whether spatial differences in relative friending and general friendliness are driven by immutable preferences of the populations of local natives (e.g., if Germans in some regions happen to have a persistently friendlier disposition towards foreigners) or by place-specific factors that can shift the same Germans' friending behavior if they were to move across locations.

Our analysis shows that place-specific factors such as local policies, institutions, or social equilibria play a dominant role in explaining Germans' social behaviors towards migrants, and thus the spatial variation in migrant's integration outcomes. To document this fact, we follow the friending behavior of Germans that move across locations. We find that both general friendliness and relative friending quickly adjust from the levels of the movers' origin locations towards the levels of their destination

locations. Native movers younger than thirty adjust their general friendliness about three-quarters of the way to that of comparable destination natives within a year of moving; their relative friending adjusts fully to that of destination natives. Even for older natives, general friendliness and relative friending adjust more than half the way to those of comparable destination natives within a year of moving. These findings highlight that Syrian migrants' lack of integration in some locations is not primarily the result of immutable preferences or beliefs of the native locals in those locations. Instead, our results show that the probability of the same two individuals—the same German and the same Syrian—becoming friends varies substantially with the institutional frameworks or the prevailing social equilibria across locations.

Why would a German's propensity to befriend a local Syrian migrant differ so much as they move across locations? We next study the determinants of both general friendliness and relative friending—the two drivers of this propensity—by exploring their relationships with county-level characteristics. These correlations can be informative about the mechanisms that drive migrants' integration outcomes even as they capture equilibrium relationships that often complicate assigning a direction of causality. As we show, the correlational analysis can also help identify factors that merit further causal study.

Consistent with ethnographic work that highlights benefits to integration in smaller European towns and cities (Gauci, 2020), we find that migrants' social integration decreases with population density. Our new measures show that two forces drive this result. First, Germans are less likely to befriend any of their (German and non-German) neighbors in cities than in the countryside. Put differently, cities have low general friendliness, a common feature in work that explores the "loneliness of cities" across countries (Hammoud et al., 2021). Second, Germans living in cities are particularly unlikely to befriend migrant neighbors—cities have lower relative friending—consistent with work that highlights that social segregation increases in group size (Chetty et al., 2022*b*).

The 'relative friending' component of integration decreases with a county's population share that was Syrian in 2019, but increases with the share that was Syrian in 2010. This finding speaks to the "ethnic enclaves" literature that finds migrant networks support integration in some settings and hinder it in others (e.g. Lazear, 1999; Edin, Fredriksson and Åslund, 2003; Cutler, Glaeser and Vigdor, 2008; Beaman, 2012; Sale, 2021; Martén, Hainmueller and Hangartner, 2019). Here, earlier migrants may increase relative friending by providing information or connections to support new arrivals' social integration, and their long-term presence could have positively shaped local natives' views towards Syrians. On the other hand, many migrants arriving at the same time may lead to fewer equilibrium migrant-native connections in part because the presence of many others with similar backgrounds facilitates the formation of migrant cliques (Chetty et al., 2022*b*). Those cliques might reduce both the need and desire of migrants to form links with natives as well as the desire of natives to form friendships with migrants.

For policymakers, language and integration courses are one of the few direct tools to foster the social integration of migrants and such courses have been a key component of the German government's integration policy. We next explore the effectiveness of such integration courses, contributing to a literature that has studied various government policies intended to assimilate minority groups or to improve their labor market prospects (e.g., Abdelgadir and Fouka, 2020; Abramitzky, Boustan and Eriksson, 2020; Arendt et al., 2020, 2022; Bandiera et al., 2019; Battisti, Giesing and Laurentsyeva, 2019; Fouka, 2020; Heller and Slungaard Mumma, 2020; Lleras-Muney and Shertzer, 2015; Kanas and Kosyakova, 2022).

We find that in counties with more completed integration courses per Syrian migrant, relative friending is higher, suggesting that these courses might shift equilibrium friending behaviors in a location.

To further explore these patterns, we use an instrumental variables approach to study whether the provision of local integration courses had a *causal* effect on integration outcomes. Our instrument, the local availability of qualified unemployed teachers at the start of the Syrian migration wave, is correlated with the completion of integration courses, even after controlling for the overall unemployment rate. This aligns with anecdotal evidence that the unavailability of qualified teachers substantially limited the government's ability to offer integration courses. We estimate that a 10% increase in 2015-19 integration course completion per Syrian (driven by higher course availability) raised friending integration by 17%. All of this effect comes from raising Germans' equilibrium rate of befriending Syrians closer to their rate of befriending Germans (i.e., by raising relative friending). Germans' general friendliness—the second determinant of migrants' social integration—is unaffected by the completion of integration courses.

In the last section of the paper we return to the determinants of differences in natives' persistent friending behaviors, and study the longer-term effects of exposure to Syrian migrants on subsequent friending patterns. Specifically, we use fluctuations in the presence of Syrian migrants across high school cohorts as a quasi-random source of variation of exposure to such migrants. We find that exposure to Syrian migrants in high school leads to higher probabilities of German natives befriending Syrians even outside the high school setting, consistent with the contact hypothesis, which outlines the circumstances in which social contact between members of different groups can help to reduce prejudice and animosity (Allport, Clark and Pettigrew, 1954; Bursztyn et al., 2021; Boisjoly et al., 2006; Carrell, Hoekstra and West, 2015; Paluck, Green and Green, 2019; Rao, 2019; Corno, La Ferrara and Burns, 2022).

The concept of social integration has long been important in social science research (e.g., Srole, 1956; Coleman, 1988; Putnam, 1995*a*; Alesina, Baqir and Easterly, 1999). Within this literature, our work relates most closely to studies that use surveys or assimilation-related measures to proxy for migrants' social integration. Laurentsyeva and Venturini (2017) provide one overview (see also Niehues, Rother and Siegert, 2021; Schmidt, Jacobsen and Krieger, 2020; Cheung and Phillimore, 2014). In contrast to these studies, we are able to directly measure key elements of migrants' social integration in large-scale administrative data that allows us to explore granular spatial variation in integration outcomes. Our unique panel data on the friending behaviors of Germans in addition to Syrians allows us to obtain a more holistic view on social integration, which, by its nature, depends on the behaviors of both migrants and natives. Our ability to study the friending behavior of natives (not just migrants) enables us to generate novel insights on the determinants of this integration.

Our work also relates to a literature studying the economic integration of refugees in high-income countries. Becker and Ferrara (2019) and Brell, Dustmann and Preston (2020) provide overviews. We complement this literature by showing that migrants' social integration in a region increases with their economic integration and decreases with the overall unemployment rate. We also provide evidence that integration courses had a substantial causal effect on migrants' economic integration.

We also add to a literature that uses experimental and quasi-experimental methods to study the causal effects of local environments on a variety of economic, social, and health outcomes (see Chyn and Katz, 2021, for a review). We believe we are the first to use a mover-based research design to study the

effects of place on migrants' social integration, adding to existing evidence that is observational or relies on quasi-random refugee settlements (e.g. Åslund and Rooth, 2007; Damm, 2014; Braun and Dwenger, 2017; Aksoy, Poutvaara and Schikora, 2020; Jaschke, Sardoschau and Tabellini, 2021; Sale, 2021). We also introduce the use of movers to study the effect of places on *native* rates of befriending migrants, highlighting that place-based effects are not primarily picking up fixed preferences of local natives.

The remainder of this paper is structured as follows. In Section 1 we describe our data, sample, and outcomes of interest, before documenting overall patterns of social integration. Section 2 explores the relationship between individual-level Syrian migrant characteristics and integration outcomes. In Section 3 we generate regional measures of social integration and use movers to study the extent to which they reflect place-based effects. Sections 4 and 5 focus on natives and local institutions, exploring the forces that make migrants more likely to integrate in one place versus another. Section 6 looks at how quasi-random exposure to migrants shapes natives' long-term behavior. We conclude in Section 7.

1 Data and Descriptive Statistics

We work with de-identified data from the online social networking site Facebook. In March 2021, Facebook had over 2.8 billion monthly active users, including 423 million in Europe (Facebook, 2021). Facebook is used widely by Syrian migrants in Germany to share information and communicate with friends and family in Syria and elsewhere (Scheibe, Zimmer and Stock, 2019). Many individuals opened their Facebook accounts prior to arriving in Germany, while others likely created accounts during their migration, as Facebook was frequently cited as a tool used by refugees fleeing to Europe to share information (Dekker et al., 2018; Mall et al., 2015; Ritscher, 2016; Mustafa and Lamb, 2017).

Establishing a "friendship" connection on Facebook requires the consent of both parties, and a person can have at most 5,000 connections. As a result, Facebook connections are usually between individuals who interact in person (Jones et al., 2013). Facebook networks thus resemble real-world social networks more closely than networks on other online platforms where uni-directional links to non-acquaintances such as celebrities are common. As a result, prior studies have used Facebook data to explore the relationship between social connections and a variety of economic and social outcomes such as trade flows, patent citations, travel flows, disease transmission, bank lending, social capital, social program participation, investment decisions, product adoption decisions, housing choices, and beliefs and behaviors surrounding public health (Bailey et al., 2018*a*,*b*, 2019*a*,*b*, 2020*a*,*b*,*c*, 2021; Chetty et al., 2022*a*,*b*; Kuchler, Russel and Stroebel, 2021; Kuchler et al., 2020; Rehbein and Rother, 2020; Wilson, 2019).

1.1 Sample Construction

We construct our primary sample from a sub-population of Facebook users who had active accounts in October 2021, were 18 or older, lived in Germany, and had 25 or more friends. Each user is predicted to have a home region in one of 401 German districts (*Kreis*, *Landkreis*, or *Stadtkreis*), with an average population of just over 200,000 (these are assigned based on user information and activity on Facebook, including their self-reported profile information, and device and connection information). We refer to these geographies as "counties" throughout. From this primary sample, we define two sub-samples.

Syrian Migrant Sample. For many analyses, we focus on users who specify a Syrian hometown or high school in their Facebook profile, or who previously had a predicted home region in Syria. There are about 350,000 such users, which we refer to as "Syrian migrants" (see footnotes 2 and 3 for details).

In Appendix Figures A1 - A3, we compare the demographics and locations of our Syrian migrant sample against information from the corresponding full population constructed using administrative data from the Federal Statistical Office of Germany. Syrian migrant population shares across counties and demographic groups closely correspond to those in the administrative data, highlighting that we observe Syrian and non-Syrian users at roughly the same rates in the Facebook data and consistently so across geography, age, and gender groups.

German Native Sample. We also construct a group of users, which we refer to as "German natives", who meet the criteria described in Appendix B based on self-reported profile information, home region predictions, and German language usage. We identify 18 million such users. The median county has 34,063 German native users; the 10th-90th percentile range is 17,057 to 74,651 German native users. Appendix Figure A4 benchmarks this sample against administrative data. The share of users in the primary Facebook sample that are German natives is somewhat lower than the true population share, a result of our relatively strict assignment criteria. However, the German population shares are consistent across space and gender, with population-weighted correlations between county $\times$ gender shares of German natives in the Facebook sample versus the actual population of 0.94.

1.2 Measures of Migrants' Social Integration

We capture the social integration of Syrian migrants using three primary measures (see Appendix C for detailed definitions):

1. The number of native German friends a Syrian migrant user has in the same or a bordering county;
2. An indicator for whether the Syrian migrant user produces content such as Facebook posts and comments in German; and
3. How many local native Facebook groups (e.g., local sport clubs) a Syrian migrant user joins.

1.3 Sample Summary Statistics

Panel (a) of Table 1 summarizes the Syrian migrant sample. The median Syrian migrant user is 31 years old, with a 10th-90th percentile range of 22 to 48 years. The sample is 32% female, somewhat lower than 40% in the administrative data.[4] The median number of Facebook friends and groups joined is 226 and 56, respectively. The median user in the Syrian migrant sample first used Facebook in Germany 23 quarters ago. About 8% of Syrian migrants list a German college on their profile.

Syrian migrant users have five native local friends on average.[5] This magnitude is consistent with data from the German Socio-Economic Panel (SOEP), a longitudinal survey of German households that

[4] Appendix Figure A2 colors observations in Figure A1 by gender and age, thereby benchmarking our sample against the true population as reported in the administrative data.

[5] Friendship requests between natives and Syrians are initiated at essentially equal rates by each group. On average, Syrians send a friend request in 50.01% of their friendships with native local Germans.

Table 1: Syrian Migrant and German Native Sample Summary Characteristics

Panel (a): Syrian Migrant Sample

	Mean	SD	P10	P25	P50	P75	P90	P99
Age	32.90	10.26	22	25	31	38	48	66
Female (0/100)	32.07	46.68	0	0	0	100	100	100
DE College (0/100)	7.92	27.00	0	0	0	0	0	100
N Friends	347.89	385.84	62	117	226	423	751	2431
N Groups	104.55	137.09	8	22	56	129	256	831
Qs Since 1st on FB in DE	20.30	8.04	7	15	23	25	28	36
N Local Native Friends	5.03	12.24	0	0	1	4	13	87
N Local Syrian Friends	14.99	17.43	1	4	9	20	36	103
Produces DE Content (0/100)	30.40	46.00	0	0	0	100	100	100
N Local Native Groups	0.55	1.41	0	0	0	0	2	9

Panel (b): German Native Sample

	Mean	SD	P10	P25	P50	P75	P90	P99
Age	40.23	13.79	24	29	38	51	60	77
Female (0/100)	51.74	49.97	0	0	100	100	100	100
DE College (0/100)	32.93	47.00	0	0	0	100	100	100
N Friends	253.72	243.28	51	93	181	327	535	1535
N Groups	25.22	34.52	2	6	14	30	59	231
Qs Since 1st on FB in DE	31.87	8.26	18	33	36	36	36	36
N Local Native Friends	122.52	128.88	12	32	79	168	295	687
N Local Syrian Friends	0.09	0.34	0	0	0	0	0	2
Produces DE Content (0/100)	100.00	0.00	100	100	100	100	100	100
N Local Native Groups	3.98	4.92	0	1	2	5	10	26

Note: Table presents summary statistics describing users in our samples. Panel (a) shows users in the Syrian migrant sample. Panel (b) shows users in the German native sample. Each measure is winsorized at the 99% level. Section 1.1 describes the sample construction. Appendix C provides more information on how individual-level outcomes are defined. Appendix Table A1 provides additional summary statistics.

sometimes includes modules that over-sample refugees. In the 2016 wave of the *IAB-BAMF-SOEP Survey of Refugees in Germany*, the average recent Syrian migrant in Germany reported to have 6.2 German acquaintances, a number very close to the number of Facebook friends observed in our data.[6] By contrast, Syrian migrant users have 15 Facebook friendships with other Syrian migrants in the same location. About 30% of Syrian migrant users produce content on Facebook in German. At the median and 90th percentiles, Syrian migrant users are members of zero and two local native groups, respectively.

Appendix Figure A5 presents binned scatter plots showing relationships between our three primary integration outcomes—local native friends, German content production, and local native groups—at the individual level. There are strong positive relationships, both with and without controls for individual-

[6] The exact question asked by the SOEP is: "How many German people have you met since your arrival in Germany with whom you have regular contact?" The average responses reported in the text is based on responses from 1,095 survey respondents. The advantage of our data relative to the SOEP is that our sample size is nearly 320 times as large, allowing us to precisely estimate and explore regional differences in integration outcomes. Our data also allow us to use a mover research design to understand the determinants of this geographic variation.

level demographics and Facebook usage, providing evidence that our measures are capturing related and strongly correlated aspects of social integration (also see Appendix Tables A2 and A3).

Panel (b) of Table 1 summaries characteristics of the German natives sample. The median user is 38 years old, with a 10th-90th percentile range of 24 to 60. The sample is 52% female and 33% of users list a German college on their profile. The median German native has a total of 181 Facebook friends, 79 local native friends, and 0.1 local Syrian migrant friends (users at the 99th percentile have two local Syrian migrant friends), highlighting that most German native users are not Facebook friends with a single Syrian migrant. German natives are members of four local native groups on average.

2 Migrant Characteristics and Integration Outcomes

In this section, we analyze the relationship between individual-level characteristics and integration outcomes of Syrian migrants in Germany. Figure 1 focuses on the number of friendships with German natives in the left column, and the share of Facebook content produced in German in the right column, both measured among the cohort of Syrian migrants with an "observed arrival" in 2015-2016.[7]

Migrants become increasingly socially integrated as they spend more time in Germany. For example, after their first quarter in Germany, Syrian migrant users on average had 1.4 native friends and produced 1.7% of their Facebook content in German; three years later, these numbers were 7.3 friends and 4.2% of content, respectively. The share of content produced in German also increases over the first 18 months after arrival in Germany, but then flattens substantially.

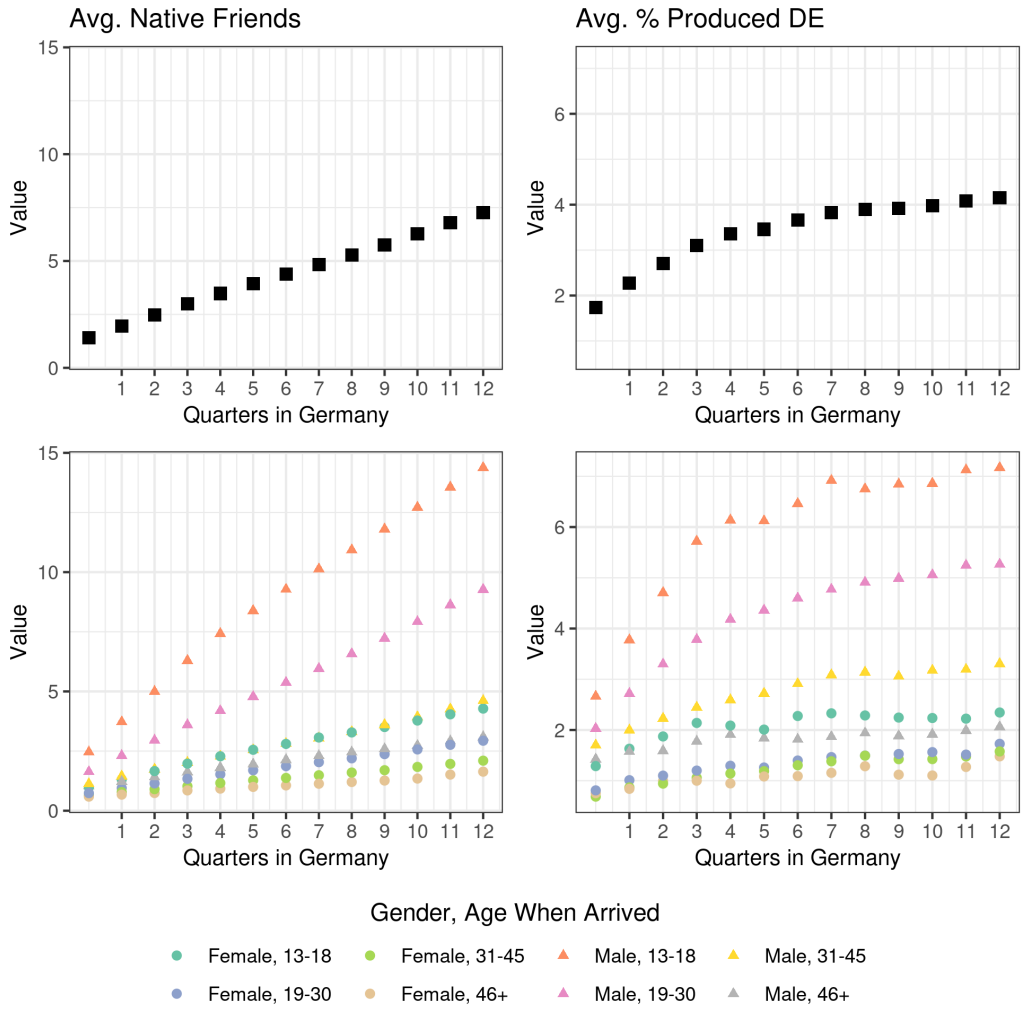
The bottom row of Figure 1 shows considerable heterogeneity in the degree of integration across age and gender groups, with younger and male migrants integrating more quickly than older and female migrants. Three years after arrival, male Syrian migrants who moved to Germany between ages 13-18 had 14.4 native German friends, compared to 4.3 such friends for similarly aged females, and 3.1 such friends for males who arrived after age 45. In Appendix D, we further explore these heterogeneities in integration outcomes across individuals, using a multivariate regression model which allows us to include various controls, including for Facebook usage patterns, as well as state and even family fixed effects. The demographic patterns shown in Figure 1 remain: female and older migrants have fewer local friends than male and younger migrants, respectively.[8] We also show that the demographic differences in integration outcomes across individuals align quantitatively with those in the SOEP survey. It is reassuring that the patterns of social integration we identify in the Facebook data align closely with available survey evidence. The Facebook data, however, is much larger and more detailed, allowing us to explore the spatial variation in integration at very local levels and to better understand the determinants of this variation. We do so in the following sections.

Native characteristics and friendships to migrants. One benefit of our rich data is that we are not only able to observe the social integration of migrants, but also the characteristics of the natives that interact with and befriend migrants. We analyze these in detail in Appendix J and summarize our findings here.

[7] These are Syrian migrant users who first used Facebook *outside* Germany, then began using Facebook inside Germany in 2015 or 2016. Appendix Figure A6 reproduces this plot with additional integration measures.

[8] Appendix Table A4 also presents multivariate regression results for our key language- and group-based measures of social integration, and Appendix Table A5 uses a different variation of our friend-based integration measures. Across all measures, we find highly consistent relationships between demographic characteristics and the social integration of Syrian migrants.

Figure 1: Integration Over Time For 2015-16 Arrival Cohort



Note: Figures show the average values, by quarter, of integration measures for users in the Syrian migrant sample with an observed arrival in 2015 or 2016. The measures are total native friends (left column) and the share of content produced in German (right column). Appendix C provides more details on each measure. The top row shows overall trends. In the bottom row each observation's shape and color represents a gender-by-age group.

Overall, younger and male German natives have more Syrian migrant friends than older and female natives. Because Syrian migrants in Germany are more likely to be young and male than the average German native, one possible explanation for this finding is that homophily plays a strong role in shaping which natives befriend Syrian migrants. Put differently, younger German natives might be more likely to connect with younger Syrian migrants because younger people are more likely to connect in general, rather than because of a particularly friendly behavior toward migrants among younger versus older Germans. Consistent with such an interpretation, we show that it is, in fact, older and female natives that are more likely than others to join pro-immigration groups on Facebook, conditional on the relevant patterns of Facebook usage. In other words, is not necessarily those who are most vocal about support for immigration (measured by supporting pro-immigration groups) that are most likely to befriend Syrian migrants and thereby directly foster the integration.

In Table 2, we explore the extent to which friendship links to Syrians disproportionately come from a small number of Germans that one might call "super integrators." Overall, 71% of all friendships between Germans and Syrians are to Germans with three or fewer Syrian friends and only 0.04% of Germans have more than 10 local Syrian friends. While there are some Germans with more than 50 local Syrian friends—which could include Germans working directly with refugees—they account for only 1.6% of all friendships that Syrians have with Germans.

We conclude that friendships between Syrians and Germans are not overwhelmingly to Germans with a large number of Syrian friends. Instead, most Syrians friendships to Germans are to Germans with few other Syrian friends. The role of possible "super integrators" seems limited.

Table 2: Concentration of Friendships Between Syrian Migrants and German Natives

			Average Age		Share Male		Total Friends	
Number of Migrant Friends	Share of Natives	Share Friendships to Migrants	Native	Migrant Friends	Native	Migrant Friends	Native	Migrant Friends
0	93.96%	0%	43.1	-	0.474	-	262	-
1	4.47%	44.6%	36.4	32.8	0.512	0.865	493	886
2-3	1.19%	26.8%	35.4	31.8	0.524	0.879	644	915
4-5	0.21%	9.0%	35.7	31.6	0.528	0.882	777	927
6-10	0.12%	8.7%	36.8	32.0	0.531	0.872	861	929
11-20	0.03%	5.5%	38.3	32.6	0.548	0.859	965	937
21-50	0.01%	3.7%	39.7	33.3	0.555	0.849	1119	956
51-100	0.002%	1.0%	42.9	33.6	0.601	0.845	1516	994
100+	0.0004%	0.6%	41.2	34.4	0.58	0.854	1981	1087

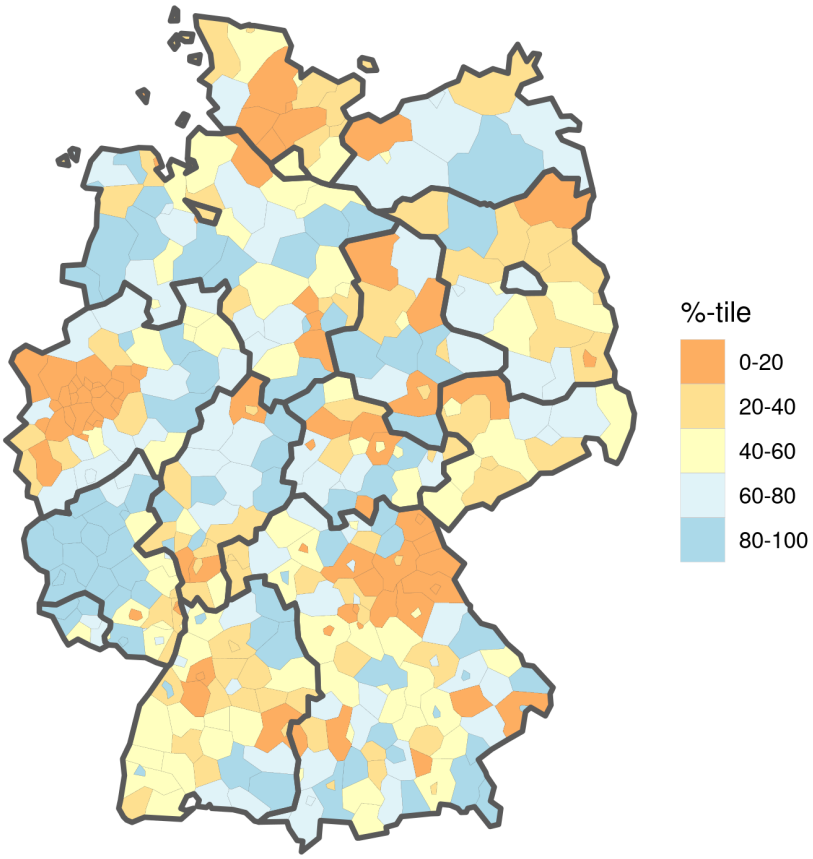
Note: Table shows summary statistics on Germans with various numbers of connections to local Syrian migrants.

3 Regional Variation in Migrants' Social Integration

In this section, we explore how Syrian migrants' social integration outcomes differ across German counties. We first show that there is substantial across-county heterogeneity in integration outcomes that reflects true differences in integration instead of sampling error or differences in patterns of Facebook usage. We then highlight that these spatial differences largely reflect causal place-based effects rather than selection in the type of migrants in each location.

County-Level Estimates. We begin by estimating county-level averages of our measures of Syrian migrants' social integration. Figure 2 maps the resulting county-level measures of friending integration, while Appendix Figures A9 and A10 show analogous maps for our language-based and group-based measures of integration. Syrian migrants in a 90th percentile county make more than twice as many local native friends on average as Syrian migrants in a 10th percentile county (7.9 vs. 3.9). Consistent with anecdotal evidence in Nawras (2017), the social integration of migrants tends to be highest in rural areas: migrants living in counties along the southern border, in Rhineland-Palatinate (along the western border), in Lower Saxony (in the northwest), and in Mecklenburg-Western Pomerania (near the Baltic Sea in the northeast) each have particularly high levels of social integration. By contrast, many mid-sized cities such as Ansbach, Kaiserslautern, and Cottbus rank among the bottom 20% of places in terms

Figure 2: Regional Estimates of Integration – Friending to Native Locals



Note: Figure shows county-level estimates of Syrian migrant integration based on the average number of local native friends among Syrian migrants in each county (residualized on regional patterns of German natives' Facebook usage). Colors correspond to measure ventiles. Darker orange and blue areas indicate the lowest and highest integration counties, respectively.

of the integration of migrants living there. Migrants living in larger cities, including Berlin, Munich, and Cologne, often have intermediate levels of social integration. Interestingly, there do not appear to be systematic differences between East and West Germany, despite their histories as distinct countries.

Panel A of Table 3 shows population-weighted county-level correlations between our various integration measures. The different integration outcomes are positively correlated across counties: those counties where Syrian migrants make more German friends are also the counties where they are more likely to use the German language and more likely to participate in local social groups.

County-Level Estimates: Validation. We next confirm that the differences in integration outcomes shown in Figure 2 reflect true differences in integration.

First, Appendix E shows that the county-level estimates of integration have high reliability, suggesting that the observed differences in integration do not arise from sampling error. For example, Appendix Table A12 shows that if we randomly split the individual-level data into two halves and estimate the county-level average of native friending in each half, the two estimates have a correlation of 0.94.

Second, one might be concerned that differences in our county-level measures of social integration might reflect spatial variation in Facebook usage. While we find no spatial differences in Facebook usage

Table 3: Correlation Between County-Level Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline Integration Measures							
(1) SY Migrants - N Local Native Friends	1.00						
(2) SY Migrants - Produced Content in DE	0.59	1.00					
(3) SY Migrants - N Local Native Groups	0.25	0.49	1.00				
(4) SY Migrants - N Local SY Friends	-0.03	-0.51	-0.41	1.00			
Panel B: Decomposition of Integration Measures							
(5) General Friendliness	0.62	0.29	-0.04	0.11	1.00		
(6) Relative Friending	0.73	0.51	0.40	-0.16	-0.05	1.00	
Panel C: Labor Market Integration Measure							
(7) Share Syrians in Employment or Training	0.45	0.59	0.13	-0.36	0.29	0.33	1.00

Note: Table presents correlations across county-level estimates. Panel A shows the regional averages of Syrian migrants after residualizing on measures of local German natives' intensive and extensive Facebook usage (see Section 3). Panel B shows the regional decomposition measures described in Section 4.1. Row 5 is general friendliness, generated as the regional average of German natives local native friends after residualizing on local patterns of Facebook usage. Row 6 is relative friending, generated as the quotient from dividing the measure in row 1 by the measure in row 5. Panel C shows an external county-level measure of the share of all Syrians that are employed or in training programs according to data from the federal employment agency (see Appendix K). Correlations are weighted by the number of Syrian migrant users in each county. Appendix Table A6 presents analogous signal correlations, which remove noise due to sampling error from the correlations.

among Syrian migrants, there are small spatial differences in Facebook usage patterns of German natives which could influence some measures of Syrian migrants' integration. For example, in a region where fewer natives use Facebook, it might look as if local Syrians were relatively less well integrated according to the"local native friends" measure, even if a key driver might be that we observe a smaller share of actual friendships on Facebook. To account for such concerns, we always residualize the observed average integration outcomes on county-level measures of the intensive and extensive Facebook usage of German natives. However, given the small magnitude of regional differences in natives' Facebook usage patterns, results are essentially the same when using the unresidualized integration measures.[9]

Finally, we validate our regional measures of the social integration of migrants by comparing them to the average number of native acquaintances made by Syrian migrants in Germany as reported in the SOEP (see Section 1). This survey data is only available at less granular geographic levels, so we can only compare the two data sources at the state (and state by age-group) levels. Despite different definitions of friendships and small sample sizes in the SOEP data, the regional measures of social integration are correlated with $\rho \approx 0.5$ across the two data sets, providing further evidence that our Facebook-based measures are picking up true variation in migrants' social integration (see Appendix Figure A8).

3.1 Causal Place-Based Effects

The observed regional variation in integration outcomes of Syrian migrants could be explained by at least two forces. A first possibility is that places have causal effects on integration, either because of

[9]Due to Facebook business restrictions, we are unable to publicly characterize the spatial distribution of natives' Facebook usage patterns. We verify that the high reliability estimates documented above are not driven by usage differences: in Appendix Table A12, we show the split-sample reliability before and after residualizing is similar (0.96 vs 0.94, for friending).

characteristics of the German natives living there, or because of institutional factors associated with the location. A second possibility is that there exist systematic differences in characteristics of Syrian migrants by place that shape their propensity to integrate—for example, if migrants with knowledge of the German language are more likely to live in certain areas. In this section, we show that the observed regional differences largely reflect causal place-based effects on integration.

Königsteiner Schlüssel. First, it is important to note that asylum seekers in Germany are initially dispersed throughout the country in a quasi-random way and according to a formula based on local population and tax revenues (the *Königsteiner Schlüssel*), and that there are restrictions on resettlement. This institutional framework suggests migrants' current locations should largely be independent of any prior propensity to integrate. Indeed, in Appendix Section F we compare the distribution of refugees across places to the official assignment key and find that the two line up very closely, indicating that the assignment key has been followed relatively strictly even during these years of increased migration. Hence, cross-sectional estimates are likely to reflect causal place based effects given the quasi-random assignment via the Königsteiner Schlüssel.

Differences in observable migrant characteristics. We can also directly rule out that *observable* Syrian migrant demographics are driving the regional differences in average integration outcomes. For example, regressing migrant's age, gender, and number of quarters since arriving in Germany on county fixed effects results in R^2s of 0.005, 0.003, and 0.005, respectively, highlighting that these characteristics vary little across counties. This finding is consistent with the fact that regional integration measures with and without individual-level observable controls are highly correlated (see Appendix Figure A7).[10]

Movers Design to Establish Causal Place-Based Effects. Despite the quasi-random assignment via the Königsteiner Schlüssel and no evidence for selection on observables, one might still worry that selection on unobservable characteristics might explain some of the observed regional variation in the integration of Syrian migrants. For example, while some restrictions exist on asylum seekers' movements after settlement, these are less restrictive for individuals who arrived prior to August 2016 or for individuals who have been in Germany for more than three years (see Hilbig and Riaz, 2020).

We next exploit such migrant moves to to separate the role of place-based and non-place-based factors. Specifically, we focus on Syrian migrants who move between non-neighboring German counties, and study changes in the moving migrants' propensity to befriend local natives. This approach builds on recent work using similar designs to study place-based effects in different contexts (e.g., Card, Heining and Kline, 2013; Finkelstein, Gentzkow and Williams, 2016, 2019; Chetty and Hendren, 2018*a*,*b*).[11]

To see the intuition behind this research design, consider a Syrian migrant who moves from Ansbach, where we observe Syrians generally making few native friends, to Saarlouis, where they make more native friends. If the observed differences in the friending behavior of Syrians in Ansbach and Saarlouis were due to (unobservable) characteristics of the Syrians living in those places, we would expect the moving migrant's likelihood of befriending local natives to remain largely unchanged after the

[10] It is also consistent with the fact that adding county fixed effects in column 2 of Table A11 had little effect on the demographic coefficients relative to estimates in column 1.

[11] Note that we will use a panel data, as in Finkelstein, Gentzkow and Williams (2016). This requires weaker assumptions than cross-sectional movers designs such as Chetty and Hendren (2018*a*), Chetty and Hendren (2018*b*), and Finkelstein, Gentzkow and Williams (2019). We provide more detail on the identifying model and assumptions for Figure A15 below.

move. By contrast, if the observed geographic differences in Syrian social integration were primarily due to a causal effect of place, we would expect the moving migrant's likelihood of befriending native locals to increase by the average difference in this likelihoods across the two locations. The *within-migrant* magnitude of the change in the rate of befriending local Germans around a move thus captures the importance of each explanation.

To study migrant movers, we construct a sample of Syrian migrants who were in one county at least four consecutive quarters followed by a different, non-neighboring county for at least six consecutive quarters. We allow a user to be included for multiple moves so long as each move meets these criteria. Our sample includes 33,772 moves and 31,721 unique movers. Appendix Figure A11 shows that the number of moves between counties observed in the Facebook data is highly correlated with the number of moves observed in administrative data.

Figure 3 plots Syrian migrants' probabilities of befriending local natives around a move, where quarter = 0 is the first quarter we observe the migrant in their new location. Counties are grouped into terciles of the integration measure mapped in Figure 2. Panels (a) and (b) focus on users who lived in a bottom and top tercile county in the year prior to moving, respectively. In each panel, the lines correspond to individuals who move to counties in different integration terciles. The vertical axis plots the probability that a migrant makes at least one local German friend in a given quarter, a flow measure of social integration that allows us to study changes in the rate of integration around a move. To avoid picking up possible differences in natives' Facebook usage across locations, we residualize this flow measure of friending on the measures of German natives' Facebook usage in the same location-quarter.

In both panels, the likelihood of migrants making new local German friends is decreasing prior to the move, consistent with individuals investing less effort in making new friends prior to moving. There is little variation in the pre-move rate of making local German friends across the destination tercile, suggesting that individuals moving to a high-integration place behaved similarly prior to the move to individuals moving to a low-integration places.

Following the move, the probability of making local German friends varies systematically by the movers' destination, with higher probabilities for individuals moving to places with higher overall social integration levels. The pattern exists in both panels, which we interpret as evidence for symmetric place-based effects. There is also an additive increase in the rate of making local friends following a move, independent of integration tercile in the origin and destination, consistent with all movers building new local networks in their destinations.[12]

In Appendix G we formally outline and estimate a simple model in which a migrant's rate of befriending local natives is determined by the sum of place-based effects—which we allow to vary across time and with observable migrant characteristics—and other *unobservable* individual-level factors. Since only place-based factors change around a move, this model allows us to estimate the share of regional variation in the social integration of migrants that can be attributed to place-based effects.

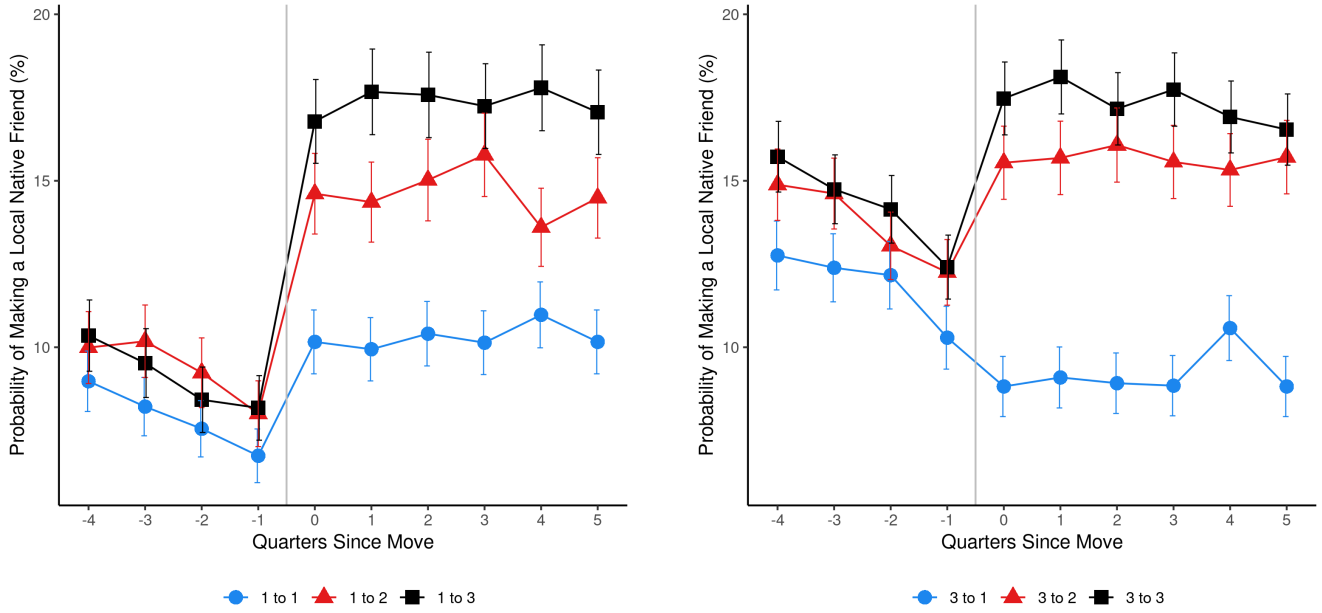
The results confirm that differences in social integration across regions are largely due to causal

[12]In Figure A13, we repeat this analysis, dividing the friendships into two groups based on the identity of the party initiating the friendship request on Facebook. We find that both the probability of incoming and the probability of outgoing friendships move in similar patterns around a move: moving to a high-integration area leads migrants to have more incoming and outgoing friendship requests towards native Germans.

Figure 3: Change in Syrian Migrants' Friending of Local Natives Around a Move

(a) Moving From Bottom Integration Tercile

(b) Moving From Top Integration Tercile



Note: Figures show the quarterly probability that a moving Syrian migrant befriends at least one local German native, relative to the timing of the migrant's move. The population is Syrian migrant users who moved between non-neighboring counties and were in the first and second county for 4+ and 6+ consecutive quarters, respectively. Counties are grouped into terciles (weighted by the number of Syrian migrant users) of the regional friending-based measures of integration in Figure 2. Panels (a) and (b) limit to users who move from a county in the bottom and top tercile of integration, respectively. The different lines show movers to counties in each of the three terciles of social integration. The individual-level outcomes are residualized by the regional measures of Facebook usage described in Section 3. Bars display 95% confidence intervals of the estimates.

place based effects. Specifically, we find that nearly three quarters of the observed regional variation in Syrian migrants' friendship formation with local natives is directly attributable to place-based effects that occur within the first year after their move, rather than individual characteristics. The results are not driven by any particular demographic group and are also fully symmetric, with moves to low-integration places leading to declines in the rates of making native friends of the same magnitude as moves to high-integration places increase that rate.

Causal Place Based Effects – Summary. Both the initial quasi-random assignment via the Königsteiner Schlüssel and our mover design suggest that differences in social integration across localities are largely due to causal place based effect. This is also consistent with prior work that has similarly found substantial evidence for causal place based effects on labor market integration outcomes in other countries.

4 Determinants of Place-Based Integration

Given the large causal effects of locations on migrants' social integration, we now explore why the same migrant is more likely to integrate in one place versus another. One of the unique advantages of our large scale friendship data is our ability to study the role that local natives play in explaining the observed differences in the social integration of migrants.

4.1 Decomposing Migrants' Integration: General Friendliness and Relative Friending

We distinguish two forces that can contribute to regional variation in migrants' social integration.

The first force, which we call *general friendliness*, is the overall rate at which natives in a location befriend others in their community: if local natives in a given location are more likely to befriend any neighbor, they might also be more likely to befriend their Syrian migrant neighbors.

The second force, which we call *relative friending*, is the relative probability of a German native befriending a given local Syrian migrant versus a given local German: the more a local native befriends migrants similar to how they befriend other natives, the easier it is for migrants to integrate socially.

Our unique data allow us to measure these two components separately, and thus improve our understanding of the causal effects of place documented in Section 3. We define a county's general friendliness as German natives' average number of local German friends. Relative friending in a county is defined as migrants' average number of local German friends divided by the county's general friendliness. General friendliness and relative friending thus determine friending integration multiplicatively:

$$\underbrace{NLocalFriends_j^{SY\to DE}}_{Friending\ Integration} = \underbrace{NLocalFriends_j^{DE\to DE}}_{General\ Friendliness} \times \underbrace{\frac{NLocalFriends_j^{SY\to DE}}{NLocalFriends_j^{DE\to DE}}}_{Relative\ Friending}. \tag{1}$$

The variables $NLocalFriends_j^{DE\to DE}$ and $NLocalFriends_j^{SY\to DE}$ correspond to the average number of local native friends among native and Syrian migrant users in county j, respectively, after residualizing on regional patterns of Facebook usage in the native population as before.

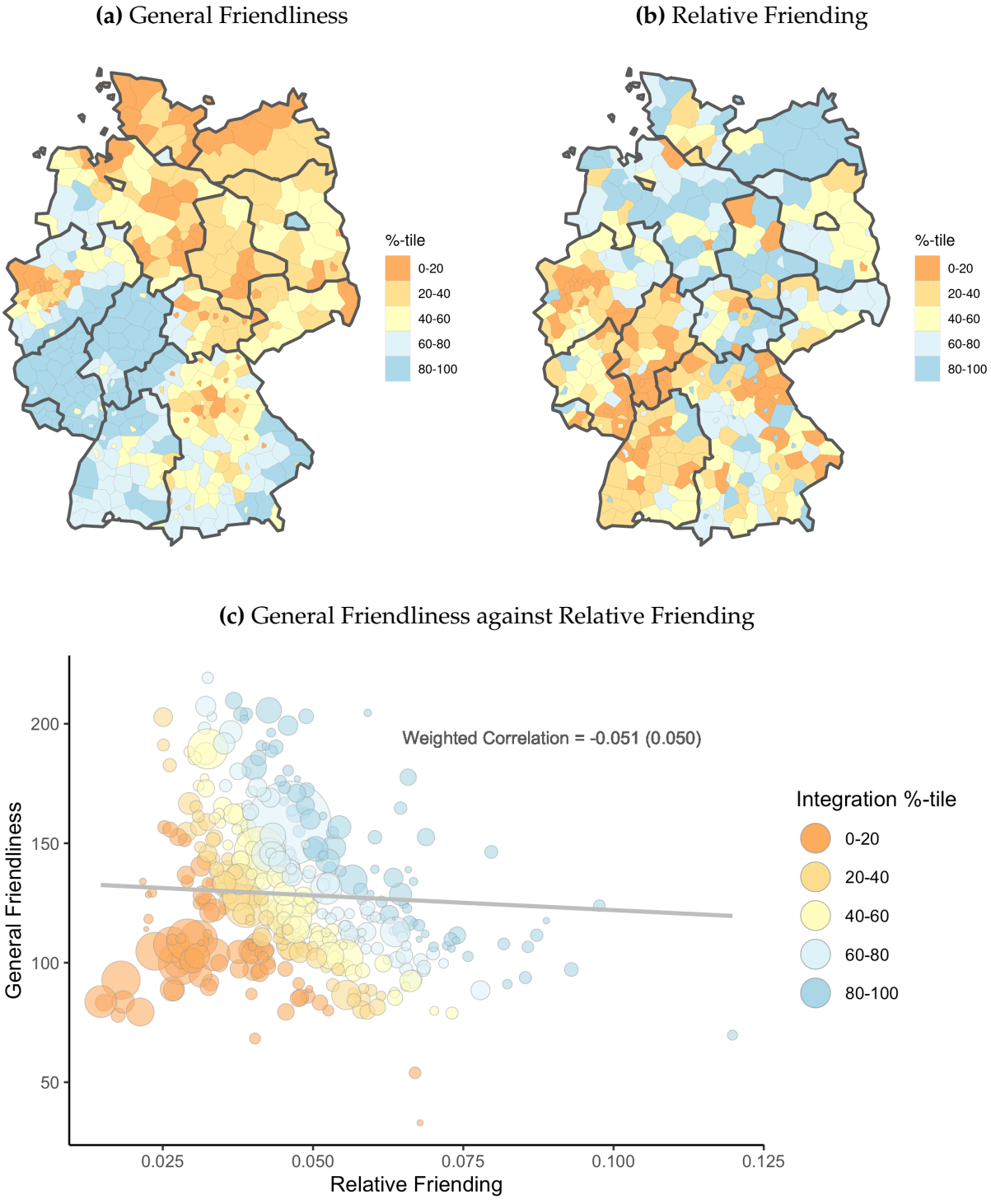
Intuitively, relative friending captures how much harder it is for a Syrian migrant to make a local native friend than it is for a native German to make that friend. To further build intuition for its determinants, it is possible to re-write county-level relative friending as a function of only natives' friending behaviors, using the fact that, within a county, the total number of friendships from local migrants to local Germans must equal the total number of friendships from local Germans to local migrants:

$$Rel.\ Friending = \frac{NLocalFriends_j^{SY\to DE}}{NLocalFriends_j^{DE\to DE}} = \frac{NLocalFriends_j^{DE\to SY}}{NLocalFriends_j^{DE\to DE}} \times \frac{NGer_j}{NSyr_j} = \frac{\frac{NLocalFriends_j^{DE\to SY}}{NLocalFriends_j^{DE\to DE}}}{\frac{NSyr_j}{NGer_j}}. \tag{2}$$

Here, $NGer_j$ and $NSyr_j$ are the numbers of German native and Syrian migrant Facebook users local to county j, respectively. $NLocalFriends_j^{DE\to SY}$ is the average number of local Syrian friends of German natives in county j. Relative friending will thus be equal to one if German natives befriend local Syrian migrants and other local German natives in proportion to their population shares.

Panels (a) and (b) of Figure 4 map general friendliness and relative friending by county, while Panel (c) shows their across-county correlation, with different colors representing different integration levels. General friendliness is higher in Western states and lower in Northern Germany, while relative friending is generally higher in Northern Germany. The industrial areas in the Ruhr area of North Rhine-Westphalia—including the cities of Duisburg, Oberhausen, Bottrop, and Gelsenkirchen—as well as parts of upper Franconia in northern Bavaria have low general friendliness and low relative friend-

Figure 4: Regional Estimates of General Friendliness and Relative Friending



Note: Panel (a) shows county-level estimates of general friendliness, the average number of local native friends among natives in each county (residualized on Facebook usage). Panel (b) shows county-level estimates of relative friending, given by the ratio of the overall friending integration measures and general friendliness (see equation 1, also residualized on Facebook usage). Colors correspond to measure ventiles. Darker orange areas indicate the lower values of general friendliness and relative friending, and darker blue areas indicate higher values. Panel (c) shows a county-level scatter plot of relative friending against general friendliness. The size of bubbles corresponds to the number of Syrian migrants in the county. Darker orange observations have the lowest friending integration (mapped in Figure 2) and darker blue have the highest.

ing; migrants have the lowest integration levels in these places. Overall, general friendliness and relative friending are weakly negatively correlated across counties, with a weighted correlation of -0.05.

To quantify the relative importance of general friendliness and relative friending in explaining county-level differences in integration outcomes, in columns 1 and 2 of Table 4 we separately regress the log of overall friending integration on the log of each component. The R^2 estimates of 0.41 and 0.66 for general friendliness and relative friending, respectively, suggest that relative friending explains 50% more of the geographic variation in Syrian migrants' integration than general friendliness does (also see Appendix I for related analyses).

Table 4: County-Level Relationship Between Integration Measures

	Friending Integration		Language		Employment / Training	
General Friendliness	1.098***			0.183***		0.558***
	(0.13)			(0.07)		(0.08)
Relative Friending		1.056***		0.255***		0.459***
		(0.07)		(0.03)		(0.06)
Friending Integration			0.228***		0.494***	
			(0.04)		(0.05)	
N	401	401	401	401	385	385
R-Squared	0.408	0.664	0.367	0.374	0.353	0.356

Note: Table shows results from multivariate regressions exploring the county-level relationship of integration measures with general friendliness and relative friending. In every specification, the outcomes and all controls are measured in logs. The outcomes are friending integration (columns 1 and 2), the share of Syrian migrants on Facebook who produce German content (columns 3 and 4), and the share of Syrians employed or in training programs (columns 5 and 6) according to data from the federal employment agency (see Appendix K). Regressions are weighted by the number of Syrian migrants in the Facebook data. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

For some policy questions, it is not necessarily central to determine whether good integration outcomes in a given place are driven by high general friendliness or high relative friending. For instance, a policymaker interested in simply assessing the potential of different regions to socially integrate migrants—perhaps because they are interested in determining where to settle new refugees—may be indifferent to which of the components drive this integration. Indeed, columns 3 to 6 of Table 4 show that both components of social integration have strong and similarly-sized positive associations with language- and economic-based measures of integration that policymakers may care about.

However, the distinction between general friendliness and relative friending can be important in other settings. Consider a policymaker seeking to improve a location's integration outcomes. While targeted policies might reduce the gap between natives' rate of befriending migrants versus other locals (i.e., relative friending), increasing the overall friending rate (i.e., general friendliness) is likely more challenging, requiring different policies. In addition, since general friendliness and relative friending shape integration multiplicatively, interventions that raise relative friending will increase integration most where general friendliness is high. Observing each component separately therefore allows policymakers to most effectively target interventions, maximizing the overall social integration of migrants.

4.2 To What Extent Are Native Behaviors Place-Based?

We next ask what role persistent native characteristics (e.g., attitudes toward neighbors or migrants) versus place-based effects (e.g., the structure of local institutions or social equilibria) play in shaping general friendliness and relative friending. Unlike migrants, natives are not randomly assigned to places. We thus use a movers design that explores changes in *natives'* friending patterns as they move between places with different relative friending and general friendliness. When place-based effects dominate fixed individual effects in determining local friending patterns, the native movers' friending behaviors should adjust substantially towards those of natives in the place they move to.

We focus on users who moved between two non-neighboring counties and who lived in the origin and destination counties for at least four consecutive quarters. We consider moves since Q1 2017, when the substantial number of Syrians in Germany allows us to obtain precise measures of relative friending.

Our analysis compares the rate at which movers make friends in the year before and after their move to the differences in the average friending rates of otherwise similar non-movers in each location (see Appendix G for details including a formal discussion of the underlying identifying assumptions). Specifically, the outcome variable $y^{\Delta}_{i,t}$ is the change in *yearly general friendliness* or *yearly relative friending* around a move. Yearly general friendliness is the number of local native friends a user makes in a given year. Yearly relative friending is the ratio of local Syrian migrant friends to local native friends made by a German native in a given year, compared to the relative population shares of Syrian migrants and natives in that location (i.e., an annualized version of the "ratio of ratios" introduced in equation 2). $x^{\Delta}_{i,t}$ is the difference in the corresponding averages among among native stayers in the same place at the same time and in the same gender $\times$ age group as the mover. Appendix Table A9 summarizes the sample of native movers and matched non-movers. We then estimate:

$$y^{\Delta}_{i,t} = \alpha_0 + \alpha_1 x^{\Delta}_{i,t} + \xi_t + \epsilon_{i,t}, \tag{3}$$

where ξ_T are quarter-of-move fixed effects. The slope α_1 provides an answer to the following question: "within a year of being assigned to a new place, to what extent does a moving native's friending behavior adjust to that of observably similar destination non-movers?"[13] An α_1 close to 1 suggests native movers' behavior completely adjusts, whereas an α_1 close to 0 suggests it does not adjust at all.

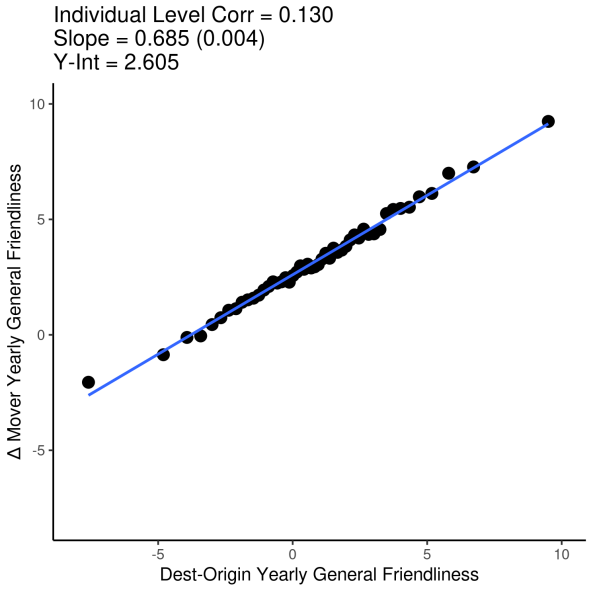
Figure 5 shows conditional binned scatter plots of $y^{\Delta}_{i,t}$ against $x^{\Delta}_{i,t}$, with slopes corresponding to α_1 (Appendix Table A10 provides the underlying regressions, as well as some robustness specifications).[14]

[13] This interpretation is intentionally narrower than that in Section 3.1, where we interpreted α_1 as the share of across-region variation in integration that is explained by place-based effects. In particular, whereas regional differences in the observables for which we allow flexibility (gender, age, and arrival cohort) were essentially non-existent for Syrian migrants, regional differences in native demographics do have the potential to shape overall variation in our measures. For example, since older people are less likely to befriend Syrian migrants, regions with older populations on average may have lower levels of integration. Since we match movers to stayers with similar observables, our estimates will not capture variation in friending patterns across space that is due to the age of the native population. (Though we will show in Section 5 that relative to other factors, the quantitative importance of these county-level differences in natives' gender and age is relatively small.)

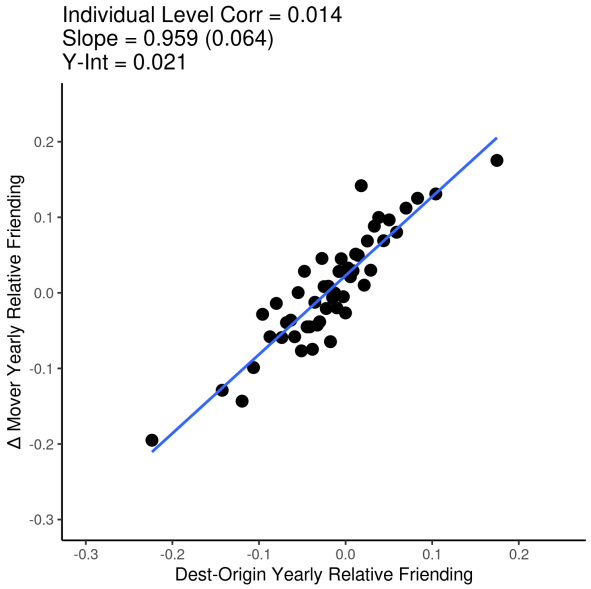
[14] One challenge with our estimation is that we only observe a sample estimate of each mover's $x^{\Delta}_{i,t}$, denoted by $\hat{x}^{\Delta}_{i,t}$. Measurement error in the true differences in friending probabilities of non-movers across locations would thus lead to attenuation bias in α_1. To account for this sampling error, when estimating equation 3, we randomly split the individual-level data of the friending behavior of non-movers used to construct $\hat{x}^{\Delta}_{i,t}$ into two sub-samples and instrument for the value constructed in one sub-sample with the value constructed in the other sub-sample (see Appendix E for details).

Figure 5: Δ Native Mover Behaviors vs. Matched Non-Movers

(a) General Friendliness



(b) Relative Friending



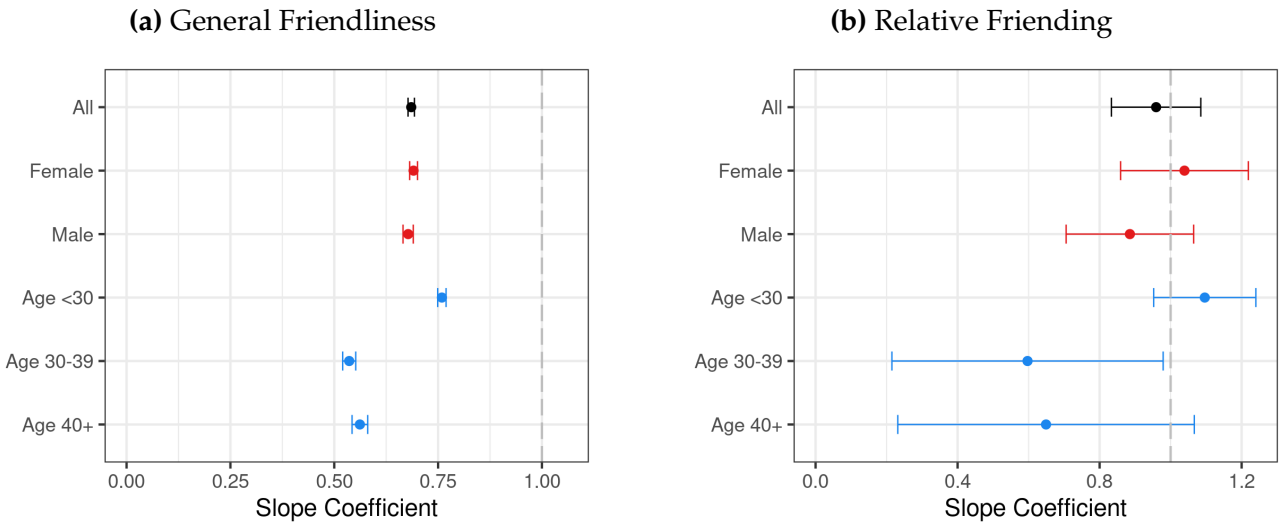
Note: Figures show binned scatter plots describing the change in the friending behavior of German natives before and after a move within Germany. The population is German native users who moved between non-neighboring counties and were in the first and second county for 4+ consecutive quarters each. In both panels the y-axis displays $y^{\Delta}_{i,t}$, an individual level change in movers' behavior the year before vs. after the move, and the x-axis displays $\hat{x}^{\Delta}_{i,t}$, the difference in average outcomes for comparable non-movers at the same time. In panel (a), the outcome is the change in the number of local German native friends made (*yearly general friendliness*) between the years. In panel (b), the outcome is the change in the ratio of the number of local Syrian migrant vs. local native friends, divided by the ratio of the number of local Syrian migrants vs. natives in the Facebook data (*yearly relative friending*) between the years. Panel (b) excludes users who make no local native friends in either the year before or after the move. In both panels we match each mover to a set of non-movers who match on gender and age buckets (18-29, 30-44, 45+). We include observations for which there are at least 1,000 non-movers in both the origin and destination match group. Both panels include quarter-of-move fixed effects. We correct for sampling error in the x-axis measures by randomly splitting the individual-level non-mover data into two halves and instrumenting for one set of averages with the other. See Appendix E for more information this procedure. Standard errors are shown in parentheses.

Panels (a) and (b) show plots for general friendliness and relative friending, respectively. In both panels, the relationship is linear and symmetric around zero, providing evidence of additive place-based effects. In Panel (a), the slope estimate suggests that, within a year of moving to a new place, a native will adjust their general friending 69% of the way to the level of comparable destination natives. In Panel (b), our estimates suggest that movers' relative friending will adjust nearly fully to that of their destination, though the estimates are somewhat less precise, since few natives make any local Syrian migrant friends. Both panels thus provide evidence that institutional factors and local policies play an important role in shaping various components of natives' friending behaviors. The fact that relative friending adjusts almost fully suggests that time-invariant individual-level characteristics such as fixed attitudes towards migrants play only a small role in explaining this outcome on average.

A number of works studying place-based effects in the U.S. find that new places exert stronger effects on younger individuals (Kling, Liebman and Katz, 2007; Chetty, Hendren and Katz, 2016; Chyn, 2018). Motivated by this, we next test whether place-based effects shape general friendliness and relative

friending differentially by age and gender. We do so by running versions of regression equation 3 over samples of users with different ages and genders. Figure 6 presents the corresponding estimates of α_1 (equivalent to the slopes in Figure 5). Native movers under the age of 30 adjust their general friendliness and relative friending around 76% and 110% of the way to the level of comparable destination users, respectively, within a year. By contrast, native movers 40 or older adjust by only 56% and 65%. Put differently, younger natives' overall friending, and friending to Syrian migrants in particular, is more strongly shaped by the characteristics of place. One possible reason for the stronger adjustment by younger movers is that places have cumulative effects on individuals, which become more ingrained over time (a force that would lead our large estimates of place-based effects to *understate* the full role of places on individuals' behaviors). In Section 6, we explore the potential role of such lasting effects by analyzing whether contact between migrants and natives in one setting has lasting effects on natives' friending behavior in other settings.

Figure 6: Δ Native Mover Behaviors vs. Matched Non-Movers - Slope By Demographics



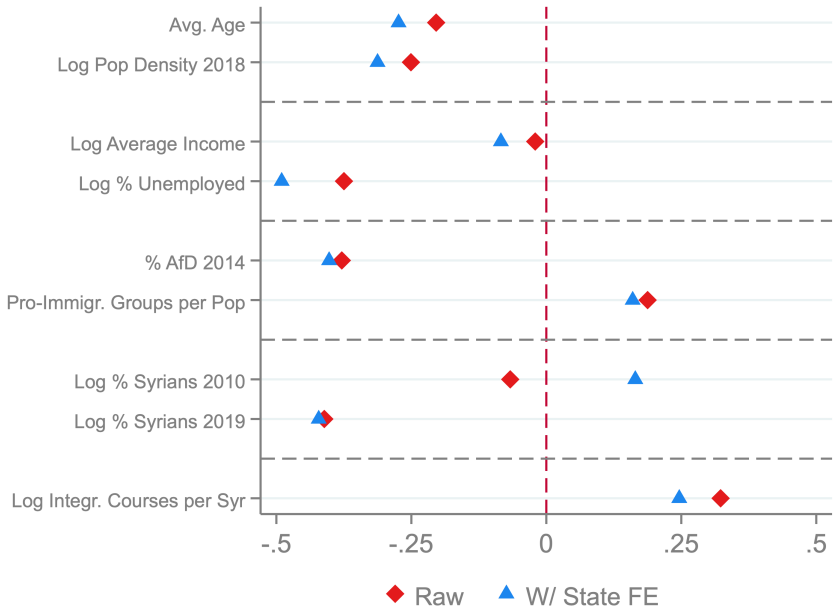
Note: Figures show slopes corresponding to versions of the respective panels in Figure 5. The coefficients in black are the slopes using the full sample of German native movers; the coefficients in red use samples of only one gender; and the coefficients in blue use samples of only one age group. Bars display 95% confidence intervals.

Overall, our results thus far indicate that general and relative friending by locals adjusts substantially around moves. This suggests that differences in friending integration across places are neither dominantly explained by persistent differences of migrants (as argued in Section 3.1) nor by persistent differences in the friending behavior of local natives. That is, friending behavior is largely influenced by place based factors not just for migrants but also for locals.

5 Correlates and Determinants of Regional Measures of Social Integration

After having established strong place-based effects for the social integration of migrants—and their determinants of general friendliness and relative friending by natives—we now explore their relationships with a variety of regional characteristics. We first focus on salient correlations with regional characteristics, before studying the causal effect of language courses as one prominent policy intervention.

Figure 7: County-Level Univariate Correlations with Friending Integration



Note: Figure presents correlations between our county-level measure of social integration and various other regional measures. Social integration is based on Syrian migrants' number of native local friends (mapped in Figure 2). Correlations are weighted by the size of the Syrian migrant sample in each county. Red diamonds depict raw, univariate correlations and blue triangles depict correlations after controlling for state fixed effects. The regional measures are average age, log 2018 population density; log average income, log employment rate; the vote share for the Alternative for Germany, demeaned by state, pro-immigration groups per capita; log of the shares of the population that were Syrian in 2010 and 2019, and log of the numbers of integration courses completed from 2015-2019 per Syrian. For more information on each measure see Appendix K.

Figure 7 presents univariate county-level correlations between migrants' social integration and various county characteristics. Red diamonds denote raw correlations, while blue triangles denote correlations with state fixed effects. Appendix Figure A14 includes similar relationships for a number of additional county-level measures, and Appendix K describes each measure in detail. Table 5 presents multivariate analyses that explore how these county-level characteristics correlate with social integration, general friendliness, relative friending, and language integration as outcomes.[15] To help with the interpretation of magnitudes, we use the log-form for some of the dependent and explanatory variables, but correlations are very similar with raw magnitudes.

Demographics & Urbanity. While the top row of Figure 7 suggests that, unconditionally, Syrians tend to be less socially integrated in places with an older population, this relationship weakens significantly in the multivariate regressions in Table 5. By contrast, in both univariate and multivariate analyses, migrants are better integrated in less densely populated areas. The results in Table 5 show that this is driven by both relative friending and general friendliness being lower in urban areas. These trends are consistent with research that finds that rural areas have higher levels of social capital and lower levels of social isolation relative to more densely populated urban areas (Putnam, 1995*b*; Rupasingha, Goetz and Freshwater, 2006; The Social Capital Project, 2018; Henning-Smith, Moscovice and Kozhimannil, 2019).

[15] In Figure 7 and Table 5, we weight all relationships by the county's Syrian migrant sample size, except when we look at general friendliness as outcome variable, in which case we weight by the county's German native sample size.

Table 5: County-level Multivariate Relationships with Friending Integration

	Friending Integration		General Friendliness		Relative Friending		Language	
Average Age	-0.032	-0.034*	-0.034***	-0.034***	0.015	0.003	-0.005	-0.011*
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Log Pop. Density 2018	-0.098*	-0.136***	-0.029	-0.071***	-0.066**	-0.058**	-0.034**	-0.016
	(0.05)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)
Log Average Income (in EUR)	-0.198	0.140	0.168	0.097	-0.296	0.054	0.070	0.035
	(0.26)	(0.18)	(0.14)	(0.10)	(0.20)	(0.15)	(0.08)	(0.06)
Log % Unemployed	-0.056	-0.291***	-0.108***	-0.065*	0.015	-0.209***	-0.129***	-0.032
	(0.09)	(0.09)	(0.04)	(0.04)	(0.07)	(0.07)	(0.03)	(0.03)
Vote Share AFD European Elections 2014	-8.953***	-6.167***	-1.939**	-1.039	-6.917***	-5.091***	-0.569	-1.289**
	(2.64)	(1.92)	(0.85)	(0.69)	(2.29)	(1.55)	(0.68)	(0.65)
Number of ProAsyl Groups per Pop	4.778*	4.286***	-1.381	-0.341	4.876***	3.167**	3.557***	1.558**
	(2.55)	(1.40)	(1.22)	(0.76)	(1.69)	(1.29)	(0.85)	(0.62)
Log Fraction of Syrians 2010	0.105***	0.150***	0.025***	0.030***	0.067***	0.114***	0.019**	0.043***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Log Fraction of Syrians 2019	-0.239***	-0.135***	-0.048*	-0.065***	-0.117**	-0.060	-0.044*	-0.103***
	(0.08)	(0.05)	(0.03)	(0.02)	(0.05)	(0.04)	(0.02)	(0.02)
Log Int. Courses Completed 2015-19 per Syrian	0.235***	0.200***	0.005	-0.013	0.222***	0.202***	0.076***	0.052***
	(0.05)	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.01)
State FE		x		x		x		x
R-squared	0.487	0.709	0.261	0.665	0.330	0.633	0.519	0.668
N	390	390	390	390	390	390	390	390

Note: Table presents results from regressions of various county-level measures on the logs of friending integration (columns 1 and 2), general friendliness (columns 3 and 4), relative friending (columns 5 and 6), and language (columns 7 and 8). The county measures are those discussed in Figure 7. Regressions are weighted by the number of Syrian migrants in the Facebook data in columns 1-2 and 5-8. Regressions in columns 3 and 4 are weighted by the number of natives in the Facebook data. Standard errors are shown in parentheses. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01)

Economic Conditions. Some prior works have explored the feedback between social and economic integration. For example, Laurentsyeva and Venturini (2017) discuss the possibility that employment contributes to migrants' social integration and Cheung and Phillimore (2014) use survey data to highlight the importance of language proficiency for employment. Figure 7 and Table 5 show that while there is no strong relationship between the average income level in a county and migrants' social integration, integration does appear to be higher in areas with lower unemployment rates, in particular when comparing counties within states. For instance, controlling for state fixed effects, we find that a 1% higher unemployment rate is associated with 0.29% lower level of social integration, an effect that is largely driven by lower relative friending rather than general friendliness.

Attitudes Towards Migrants. We explore correlations with two measures of local attitudes towards migrants: (i) the vote share for Alternative for Germany or AfD, a political party in favor of limiting immigration, in the 2014 EU Election (predating the main influx of Syrian migrants);[16] and (ii) the number of pro-immigration groups per capita. Support for the AfD has a strong negative relationship with social integration and relative friending: a one percentage point increase in AfD vote share relative to state-level averages is associated with a decrease in social integration of nearly 9% and in relative friending of 6.9%. Pro-immigration groups are independent organizations that offer a wide range of services to migrants, including help filing for asylum status, medical attention, and the provision of child care. We study groups registered with *ProAsyl*, a widely-known pro-immigration organization in the coun-

[16] Because political parties in Germany are differentially important across states, and often run with varying policy positions by state, in Figure 7 and Table 5 we always demean AfD vote share by state.

try. In both univariate and multivariate analyses, we find places with relatively more pro-immigration groups per capita tend to have higher levels of social integration. Table 5 shows this is driven entirely by variation in relative friending rather than general friendliness.

Concentration of Migrants. Several researchers have studied the relationship between local co-ethnic populations and the economic integration of migrants. For example, Edin, Fredriksson and Åslund (2003) and Damm (2009) find a positive effect on earnings for refugees living in areas with many co-ethnic individuals (so-called "ethnic enclaves"), while Cutler, Glaeser and Vigdor (2008) find negative effects if the community has low levels of average education. Our results suggest that migrants do make fewer native friends in places with more recent Syrian migrants. However, we see that social integration generally increases with the share of the population that was Syrian in 2010, largely through effects on relative friending. We find similar results when looking at the extent of German language usage. These patterns are consistent with earlier migrants providing important information or connections with natives to new arrivals to aid their social integration. It is also possible that local natives more exposed to Syrian migrants in 2010 became more friendly toward Syrians in the future, a notion we explore at the individual level in Section 6. On the other hand, large communities of migrants arriving at the same time leading to fewer migrant-native connections.

Integration Courses. The German government and other independent organizations have invested heavily in efforts to integrate recent migrants (see, e.g., Bundesregierung, 2021). Integration courses, which are intended to teach migrants the German language and other relevant information, are "at the core of the government's integration measures" (BAMF, 2015). Indeed, they have been taken by 1.13 million individuals from 2015-2019 (BAMF, 2021). In both the univariate and multivariate analyses, we find strong positive relationships between a county's social integration outcomes and the number of integration courses completed per Syrian between 2015 and 2019. The effect appears to be entirely driven by a relationship between integration course completion and relative friending. While these results are not causal, they are consistent with integration courses supporting the integration efforts of Syrian migrants. To isolate a possible causal effect of integration courses, we next use an instrumental variables approach that leverages exogenous variation in course availability across regions.

5.1 Causal Effect of Integration Policy: Integration Courses

In this section we study the causal effects of integration courses on integration outcomes. Unlike many other regional characteristics related to social integration, such as population density, policy makers can and do affect the offering of such courses, so understanding their causal effects is especially important. We use an instrumental variables (IV) approach that exploits the effect of quasi-random variation in the presence of qualified teachers across counties on the availability—and in turn completion—of integration courses. This IV approach is necessary to identify causal effects, since prior work has noted that language courses are offered more frequently in denser areas with a high share of foreigners, attributes that themselves affect migrants' social integration (Kanas and Kosyakova, 2022).

The German government required individuals teaching integration courses to either have a college degree in teaching German as a second language or, with a degree in a different pedagogical field, sig-

nificant experience teaching German as a second language (BAMF, 2018). Because of these very specific requirements, integration courses were generally taught by the small group of previously unemployed teachers with these qualifications. Indeed, in a widely-televised 2016 interview, the federal government's coordinator of refugee policy (*Flüchtlingskoordinator*) called on unemployed teachers to meet the rapid demand for integration course instructors (Tagesschau, 2016). The unemployment rate of qualified teachers in a given county at the start of the major influx of migrants thus likely affected the availability of local integration courses. We test this story using county-level data on 2014 teacher unemployment from the Federal Employment Agency. These data allow us to distinguish between four types of teachers: general, vocational, driving or sports, and other. "Other" teachers are primarily adult educators, often focused on non-native populations, and are much more likely than the other groups of teachers to meet the necessary requirements to teach integration courses. Therefore, if local teacher unemployment affects integration course availability, it should do so primarily through this particular set of teachers.

Table 6: Integration Courses and Teacher Unemployment Rates

	Log Integration Courses per Syrian 2015-19			
Log Unemp. General Schools Teachers 2014 per Syrian	0.088			
	(0.05)			
Log Unemp. Vocat. School Teachers 2014 per Syrian		0.084		
		(0.05)		
Log Unemp. Driving and Sports Teachers 2014 per Syrian			0.052	
			(0.06)	
Log Unemp. Other School Teachers 2014 per Syrian				0.229***
				(0.05)
Control Covariates	x	x	x	x
Control Log General Unemployment Rate	x	x	x	x
F-statistic	2.37	3.67	0.94	20.97
N	390	367	388	390
R-Squared	0.349	0.354	0.347	0.379

Note: Table presents results from county-level regressions between various 2014 teacher unemployment rates and integration course completion. The outcome is the log of the number of integration courses completed per Syrian between 2015 and 2019. In all regressions we control linearly for the log of the share of the population unemployed, the number of unemployed people per Syrian (as of 2014) as well as average age, log population density, log average income and log number of open training positions per applicant. Regressions are weighted by the total number of Syrians in each county as of 2019. Standard errors are shown in parentheses. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01)

Table 6 presents results that are highly consistent with the availability of teachers driving the availability and eventual completion of integration courses. Columns 1-3 show that, after controlling for general unemployment and other county-level covariates, there are no significant relationships between integration course completion and unemployed general, vocational, and driving or sports teachers per Syrian. By contrast, column 4 shows a positive and highly significant relationship for "other" teachers: a 10% increase in their unemployment per Syrian as of 2014 corresponds to a 2.3% increase in integration course completion per Syrian. With an F-statistic of just under 21, this "first stage" relationship for our IV strategy is remarkably strong given the limited number of counties.

While this evidence supports the notion that teacher unemployment meaningfully affects the completion of integration courses, for the measure to serve as a valid instrument it must also satisfy the exclusion restriction. Namely, teacher unemployment must not affect social integration other than through its effect on integration courses. To mitigate concerns that our results are driven by general economic conditions or other confounders that might affect integration, we always include a rich set of county-level controls in our regressions: general unemployment rates, the number of unemployed people per Syrian, average age, population density, average incomes, and open training positions.[17] Moreover, our use of 2014 teacher unemployment, before the large influx of migrants, allows us to rule out stories in which reverse causality violates the exclusion restriction.

Table 7: IV Estimates - Measures of Integration and Integration Courses

	Integration	General Friendliness	Relative Friending	Language	Employ. / Training
Log Integration Courses per Syrian	1.698***	0.204	1.389***	0.193***	0.891***
	(0.33)	(0.21)	(0.25)	(0.07)	(0.15)
Control Covariates	x	x	x	x	x
Control Log General Unemployment Rate	x	x	x	x	x
N	390	390	390	390	384

Note: Table presents results from county-level IV regressions of various measures related to integration on the completion of integration courses. We instrument for integration courses with the 2014 total number of unemployed "other" per Syrian. In both stages of our estimation we include the same controls as in Table 6. The outcomes are overall friending integration (column 1), general friendliness (column 2), relative friending (columns 3), the share of Syrian migrant Facebook users producing content in German (column 4), and the share of all Syrians employed or in training programs (column 5). All independent and dependent variables are specified in logs. Regressions are weighted by the total number of Syrians as of 2019 except when the outcome variable is general friendliness in which case we weight by the number of German natives in the Facebook data. Standard errors are shown in parentheses. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01)

Table 7 presents results for our IV regressions. Column 1 suggests that a 10% increase in completed integration courses increases the social integration of Syrians by nearly 17%. Quantitatively, this means that moving a migrant from a 25th percentile to a 75th percentile county in terms of the relevant teacher unemployment would result in them having about 1.7 more native friends.

This IV estimate is substantially larger than the OLS estimates in columns 1 and 2 of Table 5. At least two forces contribute to this relative size. First, our IV strategy corrects for possible downward bias due to omitted variables in the OLS estimates. Such downward bias can occur, for example, if integration courses were specifically targeted toward or advertised in areas with low integration levels. We find supporting evidence that this is indeed the case: on average, courses tend to be concentrated in urban places and places with a greater total immigrant share, both factors that are negatively correlated with integration as discussed in Section 5. Second, the IV identifies a Local Average Treatment Effect (LATE), rather than an Average Treatment Effect (ATE). If the marginal integration course participant aided by expanded course supply had higher-than-average returns from integration courses, the LATE would exceed the ATE. There are good reasons to think the marginal course participant did indeed benefit more from the course. For example, women were less likely to participate in integration courses

[17] Our controls differ from the variables used in Table 5, since we refrain from controlling for covariates that are potentially endogenous to our outcome of interest, such as the share of Syrians in 2019 or the number of pro-immigration groups.

when those courses are in short supply, but they also achieved substantially higher performance in both language and civic tests administered at the end of the course (Tissot et al., 2019; Tissot, 2021). While both LATE and ATE estimates are relevant for different applications, the LATE from our IV strategy is likely to be of particular interest for policy makers, whose primary tool to increase the completion of integration courses is to make them more easily accessible. Our LATE provides an estimate of the marginal effectiveness of precisely such relaxations of supply constraints on these courses.

Columns 2 and 3 of Table 7 present IV estimates of the effect of integration courses on general friendliness and relative friending—the two factors driving migrant integration. Because friending behavior among natives should not be impacted by integration courses, integration courses should affect overall integration only through relative friending. Highly consistent with this story, we find significant effects for relative friending, but not for general friendliness. Our IV estimates suggest that a 10% increase in integration courses completed increases friending by close to 14%.

Columns 4 and 5 measure the causal effect of integration courses on language and economic integration. In particular, our outcomes are the share of Syrian migrant Facebook users producing content in German (in column 4) and the share of all Syrians employed or in training programs (in column 5). For both, we find highly significant and positive effects of integration courses. The IV estimates suggest that a 10% increase in integration courses completed increases language integration by just under 2% and the rate of Syrians in employment or training by about 9%.

6 Exposure and Native Behaviors Toward Migrants

In the prior section, we showed that average regional differences in the propensity of natives to befriend migrants are largely driven by characteristics of the locations rather than fixed characteristics or preferences of the natives living in those locations. The lower adjustment to local friending rates for older natives, however, could be a result of long-term place based effects that become internalized over time. At the same time, there is also substantial *within-region* variation in the friending behavior of natives. In this final section of the paper, we show that differences in natives' prior exposures to Syrians explain some of these across-native differences in friending behavior, suggesting some lasting effects of prior experiences. To provide variation in natives' exposures to migrants, we exploit Germany's strict age cutoffs for school entry and show that individuals who are quasi-randomly exposed to a Syrian migrant in their high school are more likely to subsequently make Syrian friends outside of high school.

Sample Construction. We generate our sample for this analysis by subsetting our German native and Syrian migrant samples into those with a birth date between 1995 and 1999. These individuals were roughly 15 to 19 years old in 2014, at the start of the major influx of Syrian migrants. We observe 26,000 such Syrian migrant users and 2.2 million such German native users. We match individuals to their high schools using self-reports and friend-based imputations (see Appendix L). We assign 63.2% of individuals within this age group to a high school. We then sort individuals into cohorts within a school using the German system of age cutoffs for school entry. In Germany, children are eligible to enroll in school for the first time if they have turned six by a certain date that varies by state. Though students are allowed to enroll earlier or to defer enrollment at the advice of a pediatrician, the vast majority of students comply with the entry time suggested by the cutoff date (Schwandt and Wuppermann, 2015).

Research Design. Since students are disproportionately exposed to individuals in their own grade (relative to individuals in the years above and below them), variation in cohort composition can generate exogenous differences in the social networks formed by the members of each grade. Similar sources of variation in exposure and network composition have been utilized in other studies (e.g. Chetty et al., 2022*b*; Billings, Chyn and Haggag, 2021; Sacerdote, 2011). Because Syrian students are relatively uncommon in the German school system overall, we focus on how German natives are affected by having at least one Syrian migrant in their cohort. In particular, we focus on adjacent cohorts within a school where one cohort contains at least one Syrian migrant and the other does not. For instance, if the only Syrian who attends Marie Curie Gymnasium is in the class of 2016, we will study natives who fall on either side of the cutoff that divides the 2015 and 2016 cohorts.[18] We estimate equations of the form:

$$Y_i = \alpha_1 SyrianInCohort_i + \xi_{t,L} + \gamma_s + \epsilon_{i,t}. \tag{4}$$

Here, Y_i is the number of friends of a given type that user i has today, $SyrianInCohort$ is an indicator variable set to one if a user has at least one Syrian in their assigned school cohort, $\xi_{t,L}$ is a birth year-by-county fixed effect, and γ_s is a school fixed effect. Under the assumption that it is random whether a student's birth date places them into a cohort with a Syrian or into an adjacent cohort without one, α_1 identifies the effect of the additional exposure via placement into a cohort containing a Syrian. In some specifications, we include an interaction term, $SyrianInCohort_i \times CohortSize_i$, where $CohortSize_i$ is the number of students in that cohort, normalized to have mean 0 and standard deviation 1. This interaction term allows us to examine how the effects of exposure differ according to the size of the cohort.

Effects of Exposure. In Table 8, we quantify the effects of being randomly assigned to a cohort including a Syrian migrant. The first column presents baseline results: students placed into a cohort containing a Syrian have 0.02 more Syrian friends by age 21, an increase of around 40% relative to the 0.054 Syrian friends that Germans in the adjacent cohort have on average. In the second column, we interact the treatment term with the z-score of cohort size. We find that treated students in a cohort one standard deviation larger than the mean make one-third fewer Syrian friends.

We next turn our attention to the mechanisms through which these friendships can be formed. Broadly speaking, there are three possible mechanisms. First, and most trivially, German natives can befriend the Syrian in their cohort. Second, the Syrian can play a direct role in mediating connections between native Germans and other Syrians by introducing previously disconnected individuals across groups. Third, the presence of the Syrian can play a role in shaping the preferences of native Germans for contact with other Syrians. This last mechanism could play a role in future network formation if stereotypes about individuals outside one's own group inhibit friendship formation.

[18] Conceptually, we could also study Germans around the assignment cutoff for the 2016 and 2017 cohorts. However, since many Syrians enter the German school system with low levels of German proficiency, some are assigned to a cohort younger than would be suggested by the assignment rule (though we find that most Syrians have the plurality of their friends in the cohort they would be assigned into under the allocation rules used for Germans). As a result, if we use this second design (where the Syrian is supposed to be in the older cohort), we will swap the treatment and control groups of Germans when the Syrian is assigned to a *younger* cohort. We also exclude pairs of years where there is a cohort without Syrians that is flanked by cohorts with Syrians. Since Syrians from the older cohort is sometimes mis-assigned, these configurations can lead us to inadvertently compare two cohorts that both contain Syrians, which would attenuate our results.

Table 8: Impacts of High School Exposure on Friendship

	Syrian Friends		Syrian Friends (Excluding Classmates)		Syrian Friends (Excluding Syrian Classmates and their Friends)	
Syrian in Cohort	0.020***	0.020***	0.005***	0.005***	0.005***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Syrian in Cohort x Standardized Cohort Size		-0.007***		-0.003***		-0.003***
		(0.001)		(0.001)		(0.001)
School FE	X	X	X	X	X	X
Birth Year x County FE	X	X	X	X	X	X
N	115,625	115,625	115,625	115,625	115,625	115,625
Mean in Control Cohort	0.054	0.054	0.029	0.029	0.027	0.027

Note: Table presents results from regressions of the form outlined in Equation 4. The sample includes Germans who were assigned to one high school cohort where the younger cohort contains a Syrian and the older cohort does not. The treatment years include students who entered kindergarten between 2001 and 2004, while students in the paired control cohorts entered kindergarten between 2002 and 2005. In columns 1-2, we include all Syrian friends that a user makes; in columns 3-4, we only include Syrian friends who did not attend the user's high school; and in column 5-6 we only include Syrian friends who did not attend the user's high school and who did not have a prior friendship with a Syrian that attended the user's high school. In all columns, we include only Syrian friends made in the first 21 years of a person's life, in order to avoid mechanically calculating larger treatment effects for older users. All users in our sample have already turned 21. In all columns, we cluster standard errors at the school and cohort level. *(p<0.10), **(p<0.05), ***(p<0.01)

In columns 3 and 4, we repeat the regressions in columns 1 and 2, but now include only Syrian friends who did not attend the German's high school in our outcome measure. This allows us to isolate friends made through the second and third mechanisms above. We find that Germans in the treated cohorts make 0.005 more friends of this type, about 17% more than the average number of such friends in the control group. As in column 2, these effects are larger for students whose cohorts are smaller. These friendships outside of one's school comprise about one quarter of the overall effect of exposure.

In column 5, we exclude from the dependent variable both Syrians who attended the German's high school (as in column 3) as well as any friends of those Syrians. The estimate is similar to that in column 3, indicating that many of the new friendships were made in new social contexts and do not correspond to connections directly facilitated by the Syrians in one's school. This finding is consistent with quasi-random exposure to Syrian migrants shifting natives' propensities to befriend migrants across settings (see Bursztyn et al., 2021).

7 Conclusion

The challenge of harmoniously integrating immigrants into new communities has become central for policymakers around the world. In the coming decades, climate change could displace as many as one billion individuals, increasing the flow of international migrants and further raising the importance of these challenges (Kamal, 2017). However, due to the difficulty of measuring social networks using traditional data sources, understanding the drivers of international migrants' social integration has historically proven to be challenging. Are there environments where newly arriving migrants are relatively better integrated, and why? What can governments do to foster the social integration of migrants?

We use de-identified data from Facebook to document sizable spatial variation in the social integration of Syrian migrants in Germany that is driven by causal place-based factors rather than unobserved migrant characteristics. We show that regional variation in migrants' social integration outcomes is shaped by both the rate at which local natives befriend other locals in general (*general friendliness*) and the relative rate at which they form friendships with Syrian migrants in particular (*relative friending*). Natives' friending behavior adjusts substantially along both margins when they move between locations, suggesting that local institutions and environments are more important than fixed individual preferences of natives in determining whether a native makes migrant friends (although both play some role).

We then describe several characteristics of communities where migrants are better integrated. For example, our results suggest that while large numbers of migrants arriving at the same time may lead to fewer migrant-native connections, when migrants arrive in a place with many *earlier arriving* migrants they make more native connections. We also show that government-sponsored integration courses have a substantial positive *causal* effect on relative friending, helping to close the gap between the rates at which German natives befriend Syrian migrants vs. other natives. This finding highlights that integration outcomes are not immutable, but can be shaped by government policies.

We hope that the increasing availability of data sources similar to the ones used in this paper—as well as other digital trace data discussed in Kuchler and Stroebel (2022)—will help researchers better understand the forces that shape social integration and help policymakers develop programs that effectively foster interconnected communities.

References

Abdelgadir, Aala, and Vasiliki Fouka. 2020. "Political Secularism and Muslim Integration in the West: Assessing the Effects of the French Headscarf Ban." *American Political Science Review*, 114(3): 707–723.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2020. "Do immigrants assimilate more slowly today than in the past?" *American Economic Review: Insights*, 2(1): 125–41.

Aksoy, Cevat Giray, Panu Poutvaara, and Felicitas Schikora. 2020. "First Time Around: Local Conditions and Multi-dimensional Integration of Refugees."

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public goods and ethnic divisions." *The Quarterly Journal of Economics*, 114(4): 1243–1284.

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew. 1954. "The nature of prejudice."

Arendt, Jacob Nielsen, Iben Bolvig, Mette Foged, Linea Hasager, and Giovanni Peri. 2020. "Language Training and Refugees' Integration." National Bureau of Economic Research.

Arendt, Jacob Nielsen, Iben Bolvig, Mette Foged, Linea Hasager, and Giovanni Peri. 2022. "Intergenerational Spillover Effects of Language Training for Refugees." National Bureau of Economic Research.

Åslund, Olof, and Dan-Olof Rooth. 2007. "Do when and where matter? Initial labour market conditions and immigrant earnings." *The Economic Journal*, 117(518): 422–448.

Auer, Daniel, Dennis Egger, and Johannes Kunz. 2022. "Effects of Migrant Networks on Labor Market Integration, Local Firms and Employees."

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel. 2021. "International trade and social connectedness." *Journal of International Economics*, 129: 103418.

Bailey, Michael, Drew M Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel. 2020*a*. "Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing During the COVID-19 Pandemic." National Bureau of Economic Research.

Bailey, Michael, Drew M Johnston, Theresa Kuchler, Dominic Russel, Johannes Stroebel, and Bogdan State. 2020*b*. "The determinants of social connectedness in europe." *International Conference on Social Informatics*, 1–14.

Bailey, Michael, Drew M Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2019*a*. "Peer effects in product adoption." National Bureau of Economic Research.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. 2019*b*. "House price beliefs and mortgage leverage choice." *The Review of Economic Studies*, 86(6): 2403–2452.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel. 2020*c*. "Social connectedness in urban areas." *Journal of Urban Economics*, 118: 103264.

Bailey, Michael, Rachel Cao, Theresa Kuchler, and Johannes Stroebel. 2018*a*. "The economic effects of social networks: Evidence from the housing market." *Journal of Political Economy*, 126(6): 2224–2276.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018*b*. "Social connectedness: Measurements, determinants, and effects." *Journal of Economic Perspectives*, 32(3): 259–80.

BAMF. 2015. "Konzept für einen bundesweiten Integrationskurs." Available at https://www.bamf.de/SharedDocs/Anlagen/DE/Integration/Integrationskurse/Kurstraeger/KonzepteLeitfaeden/konz-f-bundesw-integrationskurs.pdf?__blob=publicationFile&v=7, (accessed: 07.30.2021).

BAMF. 2018. "Matrix: Zulassungskriterien für Lehrkräfte zum Integrationskurs bis 30.09.2020." Available at https://www.bamf.de/SharedDocs/Anlagen/DE/Integration/Integrationskurse/Lehrkraefte/matrix-zulassung-lehrkraefte-integrationskurse-bis-30092020.pdf, (accessed: 07.27.2021).

BAMF. 2021. "Geschäftsstatistik zum Integrationskurs 2020." Available at https://www.bamf.de/DE/Themen/Statistik/Integrationskurszahlen/integrationskurszahlen-node.html;jsessionid=75DC03EF1E61BC40EABEA96CA2FD8930.internet561, (accessed: 07.30.2021).

Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo. 2019. "Nation-building through compulsory schooling during the age of mass migration." *The Economic Journal*, 129(617): 62–109.

Battisti, Michele, Yvonne Giesing, and Nadzeya Laurentsyeva. 2019. "Can job search assistance improve the labour market integration of refugees? Evidence from a field experiment." *Labour Economics*, 61: 101745.

Beaman, Lori A. 2012. "Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the US." *The Review of Economic Studies*, 79(1): 128–161.

Becker, Sascha O, and Andreas Ferrara. 2019. "Consequences of forced migration: A survey of recent findings." *Labour Economics*, 59: 1–16.

Berry, John W. 1997. "Immigration, acculturation, and adaptation." *Applied psychology*, 46(1): 5–34.

Billings, Stephen B., Eric Chyn, and Kareem Haggag. 2021. "The Long-Run Effects of School Racial Diversity on Political Identity." *American Economic Review: Insights*, 3(3): 267–84.

Boisjoly, Johanne, Greg J Duncan, Michael Kremer, Dan M Levy, and Jacque Eccles. 2006. "Empathy or antipathy? The impact of diversity." *American Economic Review*, 96(5): 1890–1905.

Braun, Sebastian, and Nadja Dwenger. 2017. "The local environment shapes refugee integration: Evidence from post-war Germany." Hohenheim Discussion Papers in Business, Economics and Social Sciences.

Brell, Courtney, Christian Dustmann, and Ian Preston. 2020. "The Labor Market Integration of Refugee Migrants in High-Income Countries." *Journal of Economic Perspectives*, 34(1): 94–121.

Bundesregierung. 2021. "Migration and integration - what is the German government doing?" Available at https://www.bundesregierung.de/breg-en/news/migration-und-integration-1659206, (accessed: 06.18.2021).

Bursztyn, Leonardo, Thomas Chaney, Tarek Alexander Hassan, and Aakaash Rao. 2021. "The Immigrant Next Door: Exposure, Prejudice, and Altruism." National Bureau of Economic Research.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace heterogeneity and the rise of West German wage inequality." *The Quarterly Journal of Economics*, 128(3): 967–1015.

Carrell, Scott E, Mark Hoekstra, and James E West. 2015. "The impact of intergroup contact on racial attitudes and revealed preferences." National Bureau of Economic Research.

Chetty, Raj, and Nathaniel Hendren. 2018*a*. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *The Quarterly Journal of Economics*, 133(3): 1107–1162.

Chetty, Raj, and Nathaniel Hendren. 2018*b*. "The impacts of neighborhoods on intergenerational mobility II: County-level estimates." *The Quarterly Journal of Economics*, 133(3): 1163–1228.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, et al. 2022*a*. "Social capital I: measurement and associations with economic mobility." *Nature*, 1–14.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, et al. 2022*b*. "Social capital II: determinants of economic connectedness." *Nature*, 1–13.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." *American Economic Review*, 106(4): 855–902.

Cheung, Sin Yi, and Jenny Phillimore. 2014. "Refugees, social capital, and labour market integration in the UK." *Sociology*, 48(3): 518–536.

Chi, Guanghua, Joshua E Blumenstock, Lada Adamic, et al. 2019. "Who Ties the World Together? Evidence from a Large Online Social Network." 451–465, Springer.

Chyn, Eric. 2018. "Moved to opportunity: The long-run effects of public housing demolition on children." *American Economic Review*, 108(10): 3028–56.

Chyn, Eric, and Lawrence F Katz. 2021. "Neighborhoods Matter: Assessing the Evidence for Place Effects." National Bureau of Economic Research.

Coleman, James S. 1988. "Social capital in the creation of human capital." *American journal of sociology*, 94: S95–S120.

Corno, Lucia, Eliana La Ferrara, and Justine Burns. 2022. "Interaction, stereotypes, and performance: Evidence from South Africa." *American Economic Review*, 112(12): 3848–75.

Cutler, David M, Edward L Glaeser, and Jacob L Vigdor. 2008. "When are ghettos bad? Lessons from immigrant segregation in the United States." *Journal of Urban Economics*, 63(3): 759–774.

Damm, Anna Piil. 2009. "Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence." *Journal of Labor Economics*, 27(2): 281–314.

Damm, Anna Piil. 2014. "Neighborhood quality and labor market outcomes: Evidence from quasi-random neighborhood assignment of immigrants." *Journal of Urban Economics*, 79: 139–166.

Dekker, Rianne, Godfried Engbersen, Jeanine Klaver, and Hanna Vonk. 2018. "Smart refugees: How Syrian asylum migrants use social media information in migration decision-making." *Social Media+ Society*, 4(1): 2056305118764439.

Deutscher Bundestag. 2020. "Verteilungsschlüssel bei Bund-Länder-Finanzierungen."

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. "Ethnic enclaves and the economic success of immigrantsEvidence from a natural experiment." *The Quarterly Journal of Economics*, 118(1): 329–357.

European Commission. 2020. "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Action plan on Integration and Inclusion 2021-2027."

Facebook. 2021. "Facebook Form 10-Q, Q1 2021."

Finkelstein, Amy, Matthew Gentzkow, and Heidi L Williams. 2019. "Place-based drivers of mortality: Evidence from migration." National Bureau of Economic Research.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of geographic variation in health care: Evidence from patient migration." *The Quarterly Journal of Economics*, 131(4): 1681–1726.

Fouka, Vasiliki. 2020. "Backlash: The unintended effects of language prohibition in US schools after World War I." *The Review of Economic Studies*, 87(1): 204–239.

Gauci, Jean-Pierre. 2020. *Integration of migrants in middle and small cities and in rural areas in Europe.* European Committee of the Regions.

Hammoud, Ryan, Stefania Tognin, Ioannis Bakolis, Daniela Ivanova, Naomi Fitzpatrick, Lucie Burgess, Michael Smythe, Johanna Gibbons, Neil Davidson, and Andrea Mechelli. 2021. "Lonely in a crowd: investigating the association between overcrowding and loneliness using smartphone technologies." *Scientific reports*, 11(1): 1–11.

Heller, Blake, and Kirsten Slungaard Mumma. 2020. "Immigrant Integration in the United States: The Role of Adult English Language Training."

Henning-Smith, Carrie, Ira Moscovice, and Katy Kozhimannil. 2019. "Differences in social isolation and its relationship to health by rurality." *The Journal of Rural Health*, 35(4): 540–549.

Herdağdelen, Amaç, Bogdan State, Lada Adamic, and Winter Mason. 2016. "The social ties of immigrant communities in the United States." *Proceedings of the 8th ACM Conference on Web Science*, 78–84.

Hilbig, Hanno, and Sascha Riaz. 2020. "Freedom of Movement Restrictions Inhibit the Social Integration of Refugees."

Jaschke, Philipp, Sulin Sardoschau, and Marco Tabellini. 2021. "Scared Straight? Threat and Assimilation of Refugees in Germany."

Jones, Jason J., Jaime E. Settle, Robert M. Bond, Christopher J. Fariss, Cameron Marlow, and James H. Fowler. 2013. "Inferring Tie Strength from Online Directed Behavior." *PLOS ONE*, 8(1): 1–6.

Kamal, Baher. 2017. "Climate Migrants Might Reach One Billion by 2050." *United Nations OCHA*.

Kanas, Agnieszka, and Yuliya Kosyakova. 2022. "Greater local supply of language courses improves refugees labor market integration." *European Societies*, 1–36.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica*, 75(1): 83–119.

Kuchler, Theresa, and Johannes Stroebel. 2022. "Social interactions, resilience, and access to economic opportunity: A research agenda for the field of computational social science."

Kuchler, Theresa, Dominic Russel, and Johannes Stroebel. 2021. "JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook." *Journal of Urban Economics*, 103314.

Kuchler, Theresa, Yan Li, Lin Peng, Johannes Stroebel, and Dexin Zhou. 2020. "Social proximity to capital: Implications for investors and firms." National Bureau of Economic Research.

Laurentsyeva, Nadzeya, and Alessandra Venturini. 2017. "The social integration of immigrants and the role of policy–A literature review." *Intereconomics*, 52(5): 285–292.

Lazear, Edward P. 1999. "Culture and language." *Journal of political Economy*, 107(S6): S95–S126.

Lleras-Muney, Adriana, and Allison Shertzer. 2015. "Did the Americanization movement succeed? An evaluation of the effect of English-only and compulsory schooling laws on immigrants." *American Economic Journal: Economic Policy*, 7(3): 258–90.

Mall, Naline, Ayla Jean Yackley, Marton Dunai, and Aleksander Vasovic. 2015. "For many refugees, journey to Europe begins on Facebook." *Reuters*.

Martén, Linna, Jens Hainmueller, and Dominik Hangartner. 2019. "Ethnic networks can foster the economic integration of refugees." *Proceedings of the National Academy of Sciences*, 116(33): 16280–16285.

Mustafa, Nujeen, and Christina Lamb. 2017. *The Girl from Aleppo: Nujeen's Escape from War to Freedom.* New York, NY:Harper Wave. Pg. 111.

Nawras, Nastaran. 2017. "Heimweh nach der deutschen Kleinstadt." *Die Zeit*.

Niehues, Wenke, Nina Rother, and Manuel Siegert. 2021. "Language acquisition and social contacts progress among refugees: fourth wave of the IAB-BAMF-SOEP survey of refugees." Bundesamt für Migration und Flüchtlinge.

Paluck, Elizabeth Levy, Seth A Green, and Donald P Green. 2019. "The contact hypothesis re-evaluated." *Behavioural Public Policy*, 3(2): 129–158.

Phillips, Nolan E, Brian L Levy, Robert J Sampson, Mario L Small, and Ryan Q Wang. 2019. "The social integration of American cities: Network measures of connectedness based on everyday mobility across neighborhoods." *Sociological Methods & Research*, 0049124119852386.

Putnam, Robert D. 1995*a*. "Bowling Alone: America's Declining Social Capital." *Journal of Democracy*, 6: 65–78.

Putnam, Robert D. 1995*b*. "Tuning in, tuning out: The strange disappearance of social capital in America." *PS: Political science & politics*, 28(4): 664–684.

Rao, Gautam. 2019. "Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools." *American Economic Review*, 109(3): 774–809.

Rehbein, Oliver, and Simon Rother. 2020. "Distance in bank lending: The role of social networks." University of Bonn and University of Mannheim, Germany.

Ritscher, Annia. 2016. "Mehr als blosser Klicktivismus." *Der Bund*.

Rupasingha, Anil, Stephan J Goetz, and David Freshwater. 2006. "The production of social capital in US counties." *The journal of socio-economics*, 35(1): 83–101.

Sacerdote, Bruce. 2011. "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In *Handbook of the Economics of Education*. Vol. 3, 249–277. Elsevier.

Sale, Laura. 2021. "The social determinants of refugee integration and optimal placement: a study in Denmark." *Working Paper*.

Scheibe, Katrin, Franziska Zimmer, and WG Stock. 2019. "Social media usage of asylum seekers in germany." 263–272.

Schmidt, Katja, Jannes Jacobsen, and Magdalena Krieger. 2020. "Social integration of refugees is improving." *DIW Weekly Report*, 10(34): 355–363.

Schwandt, Hannes, and Amelie C Wuppermann. 2015. "The Youngest Get the Pill: ADHD Misdiagnosis and the production of education in Germany."

Srole, Leo. 1956. "Social integration and certain corollaries: An exploratory study." *American sociological review*, 21(6): 709–716.

Tagesschau. 2016. "tagesthemen 22:15 Uhr, 14.04.2016." Available at https://www.youtube.com/watch?v=f9r7ijIr9Tk, (accessed: 07.27.2021).

The Social Capital Project. 2018. "The Geography of Social Capital in America."

Tissot, Anna. 2021. "Hürden beim Zugang zum Integrationskurs."

Tissot, Anna, Johannes Croisier, Giuseppe Pietrantuono, Andreea Baier, Lars Ninke, Nina Rother, and Christian Babka von Gostomski. 2019. "Zwischenbericht I zum Forschungsprojekt" Evaluation der Integrationskurse (EvIk)": erste Analysen und Erkenntnisse."

United Nations. 2019. "The number of international migrants reaches 272 million, continuing an upward trend in all world regions, says UN." *United Nations Department of Economic and Social Affairs*.

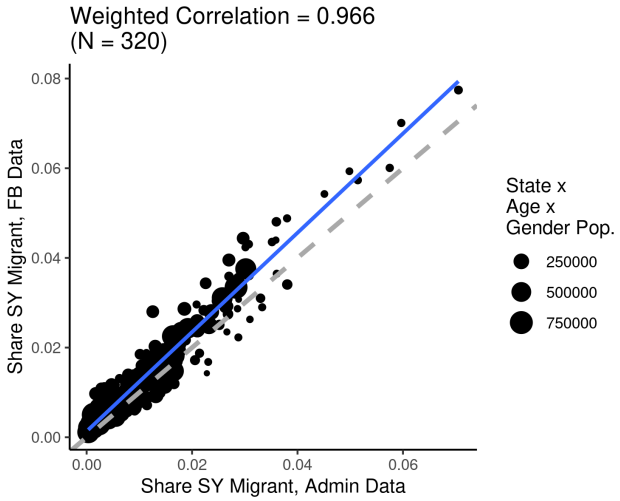
Wilson, Riley. 2019. "The impact of social networks on eitc claiming behavior." *Review of Economics and Statistics*, 1–45.

Appendices

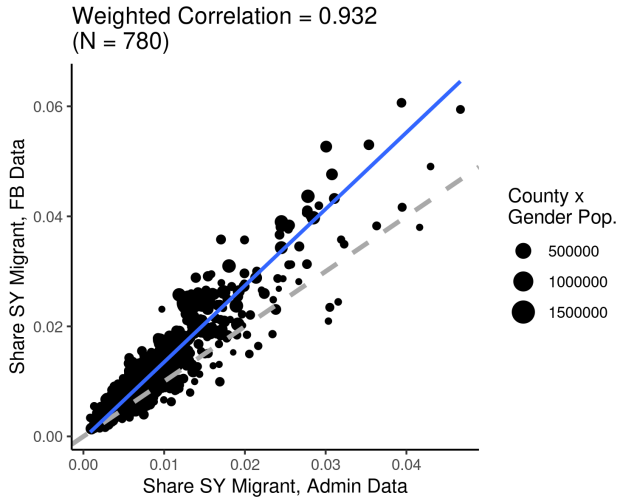
A Additional Figures and Tables

Figure A1: Syrian Migrant Sample vs. Admin Data

(a) By State × Age × Gender



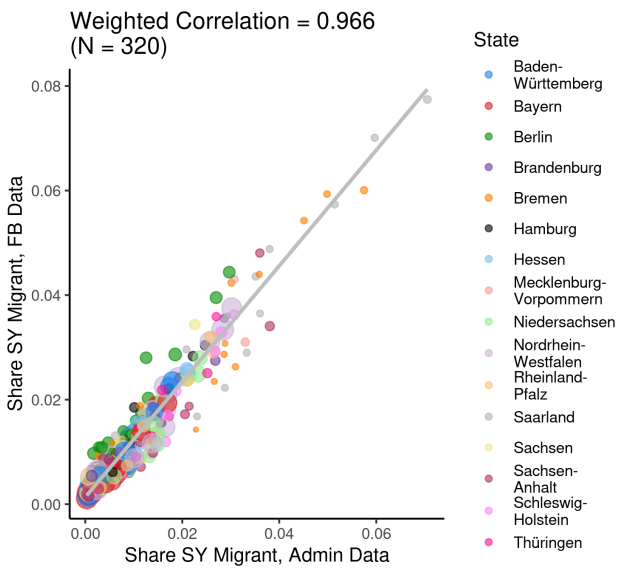
(b) By County × Gender



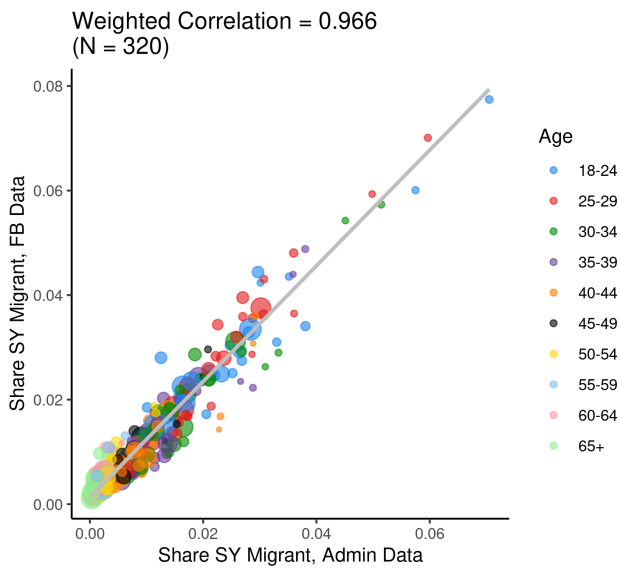
Note: Figures show the shares of the primary sample of Facebook users that are also in the Syrian migrant sample (on the y-axis), against shares of the population that are Syrian from administrative data (on the x-axis). The size of each dot is proportional to the true population it represents. The solid blue lines are from weighted linear regressions. The dashed grey line is the line $y = x$. Panel (a) plots these shares by state × age × gender. The age groups are 18-24, 25-29, 30-24, 35-39, 40-44, 45-49, 50-54, 55-50, 60-64, and 65+. There are 16 states X 10 age groups X 2 genders = 320 observations. Panel (b) plots these shares by county × gender. Admin data is unavailable for 11 counties. There are 390 counties X 2 genders = 780 observations.

Figure A2: Syrian Migrant Sample vs. Admin Data

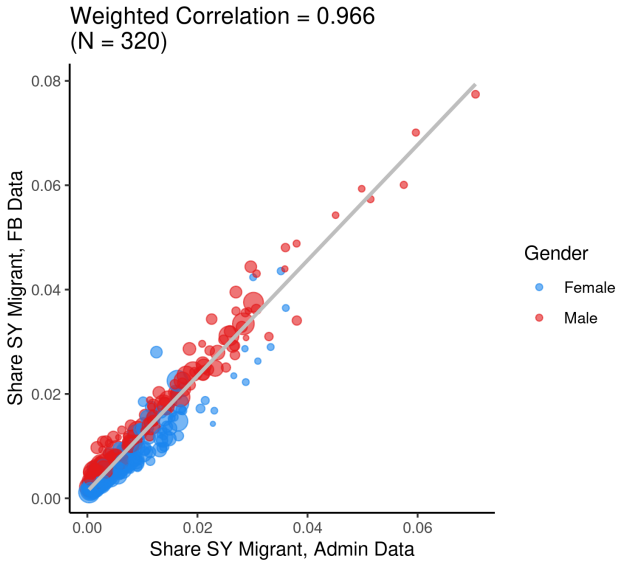
(a) By State × Age × Gender – Color by State



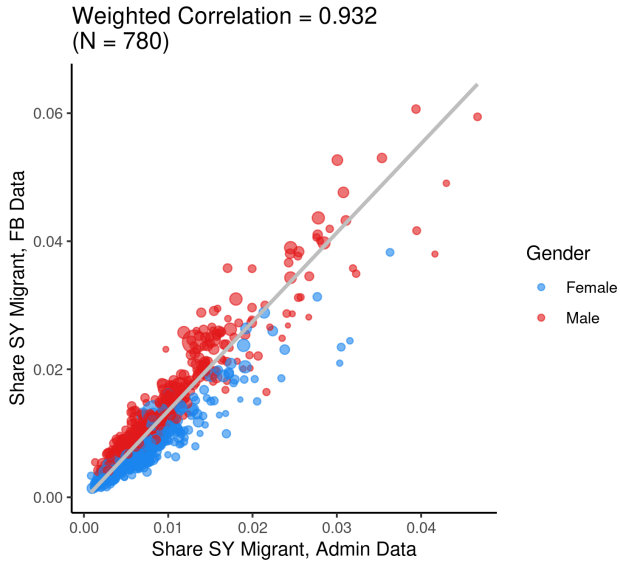
(b) By State × Age × Gender – Color by Age



(c) By State × Age × Gender – Color by Gender



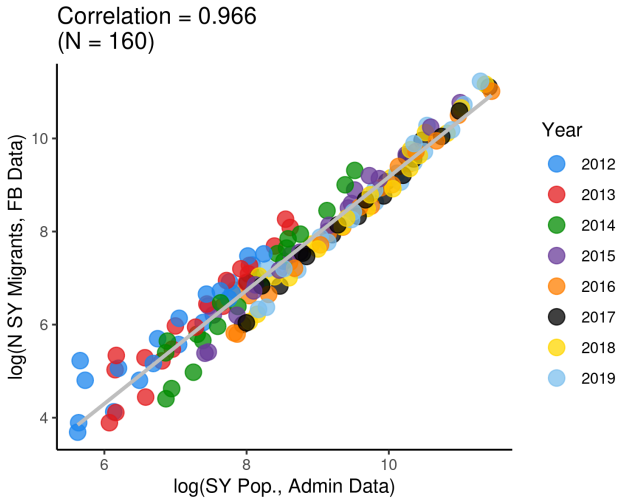
(d) By County × Gender – Color by Gender



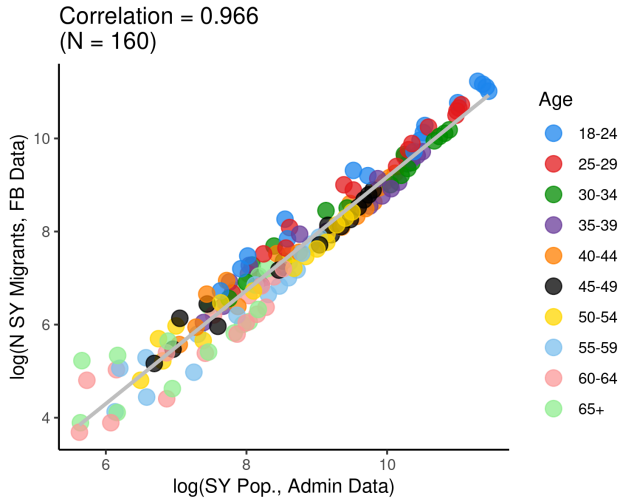
Note: Figures show the shares of the primary sample of Facebook users that are also in the Syrian migrant sample (on the y-axis), against shares of the population that are Syrian from administrative data (on the x-axis). The size of each dot is proportional to the size of the population it represents. The solid grey lines are from weighted linear regressions. Panels (a), (b), and (c) plot these shares by state, age, and gender. The age groups are 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-50, 60-64, and 65+. There are 16 states × 10 age groups × 2 genders = 320 observations. Panel (d) plots these shares by county and gender. Administrative data is unavailable for 11 counties. There are 390 counties × 2 genders = 780 observations. Panel (a) colors observations by state; panel (b) colors by age; and panels (c) and (d) color by gender.

Figure A3: Syrian Migrant Sample vs. Admin Data – By Age × Gender X Year

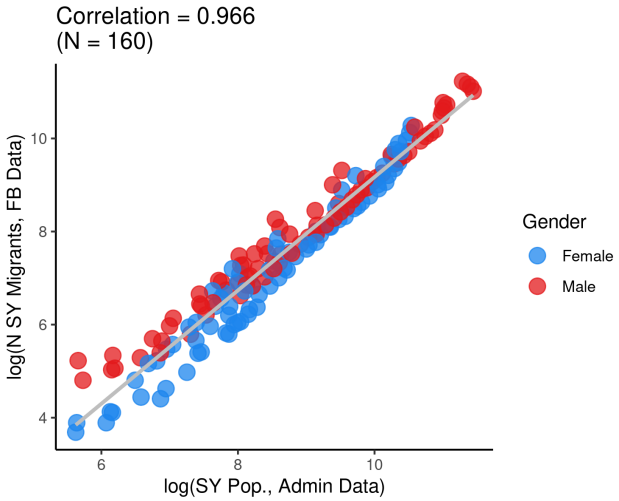
(a) Color by Year



(b) Color by Age

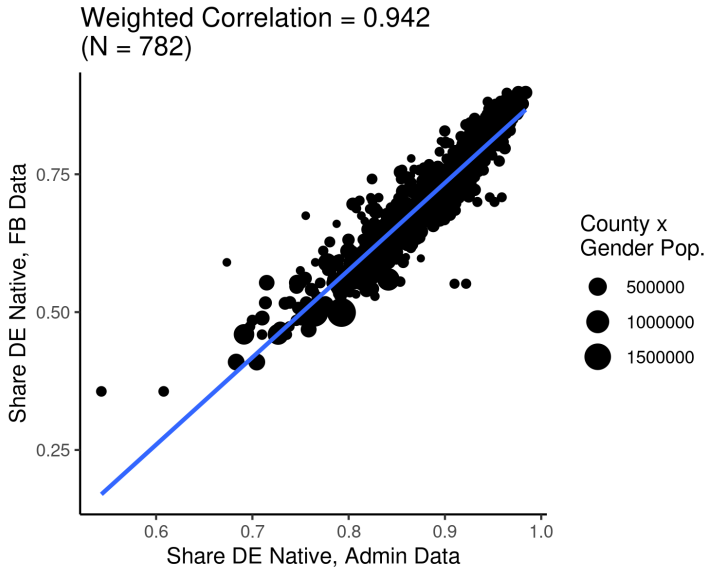


(c) Color by Gender



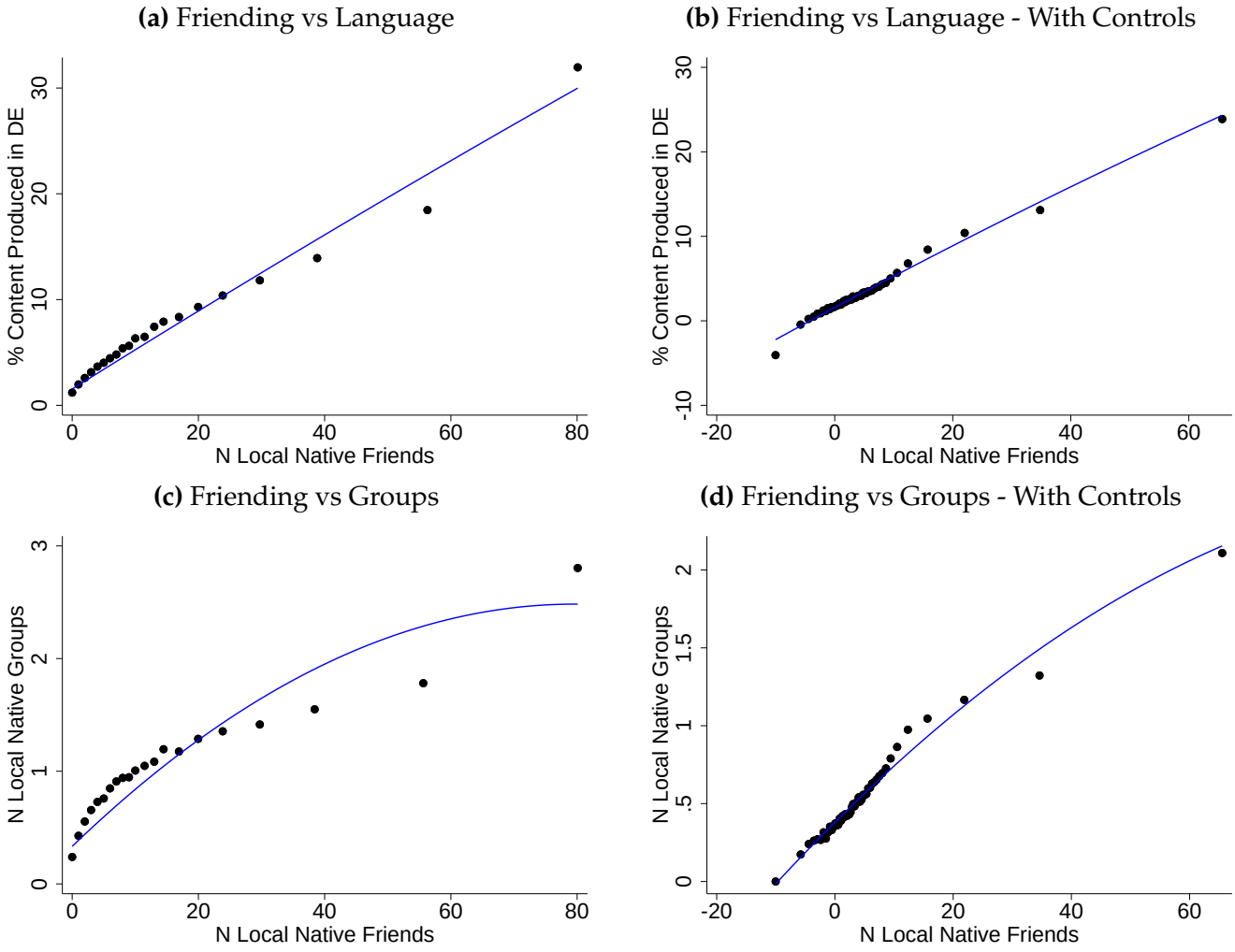
Note: Figure shows the number of users in our Syrian migrant sample using Facebook in Germany by the end of each year from 2012 to 2019 (on the y-axis), against analogous measures of Syrian migrant population from German administrative data (on the x-axis). Each observation is an age by gender by year group. The age groups are the same as those used in Figure A1. Both axes are transformed by the natural logarithm. The solid grey line is from a linear regression. Observations are colored by year in panel (a), age in panel (b), and gender in panel (c).

Figure A4: Native German Sample vs Admin Data



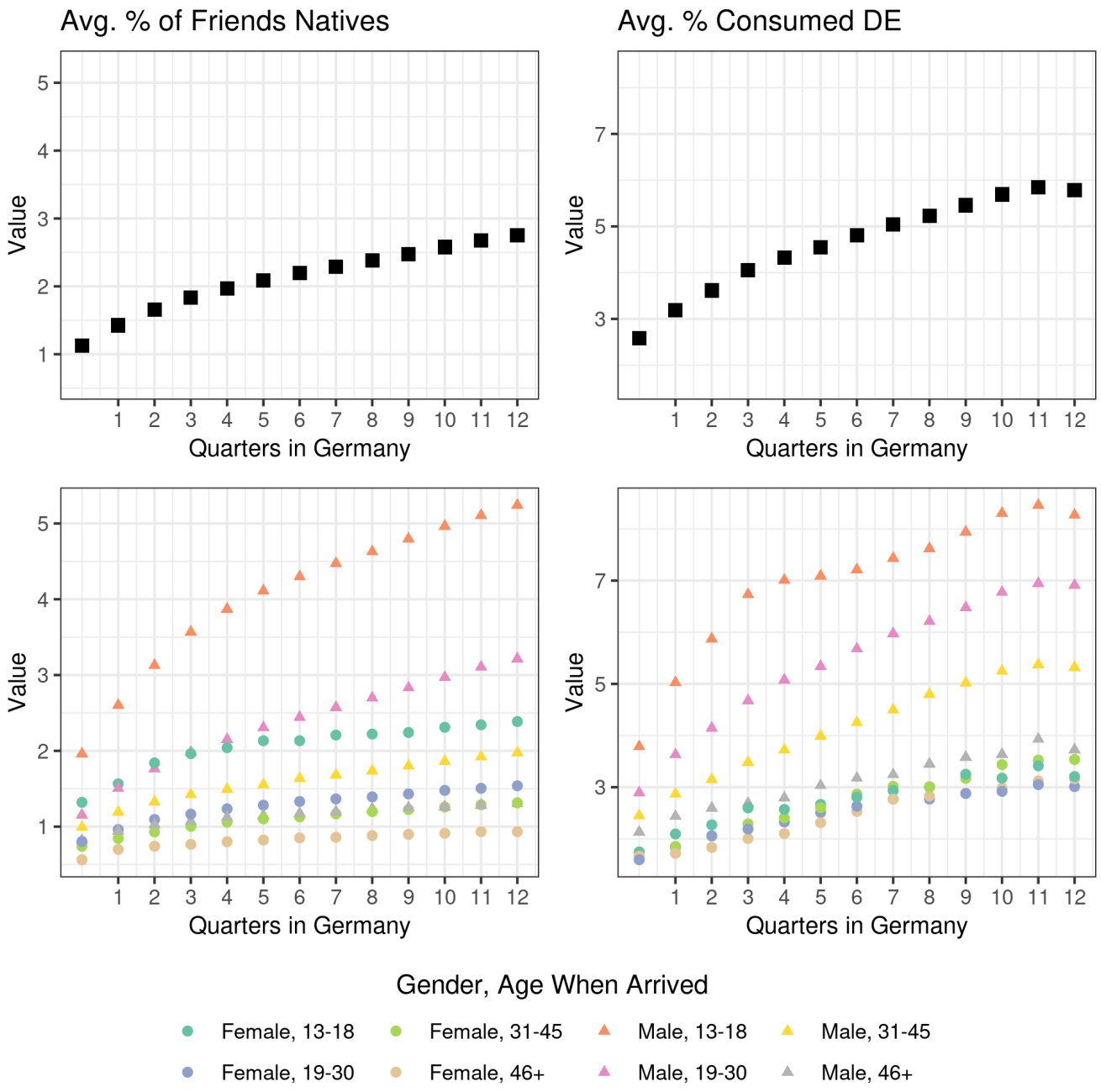
Note: Figure shows the shares of the primary sample of Facebook users that are also in the German native sample (on the y-axis), against shares of the population that are native from administrative data (on the x-axis). Each observation is a county by gender group. The size of each dot is proportional to the "true" population it represents. The solid blue lines are from weighted linear regressions. Admin data is unavailable for 10 counties. There are 391 counties X 2 genders = 782 observations.

Figure A5: Relationship Between Integration Outcomes, Individual Level



Note: Figures show binned scatter plots of individual Syrian migrants' number of local German native friends on the x-axis, against their share of content produced in German in panels (a) and (b), and the number of local native groups they are in panels (c) and (d). Appendix C provides more details on each measure. The measures in panels (b) and (d) are first residualized on the individual-level controls used in column 3 of Table A11. Lines are fit from quadratic regressions.

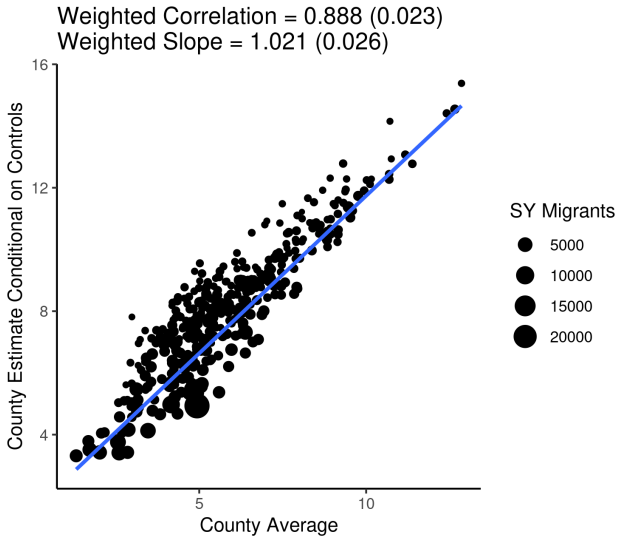
Figure A6: Integration Over Time For 2015-16 Cohort — Additional Measures



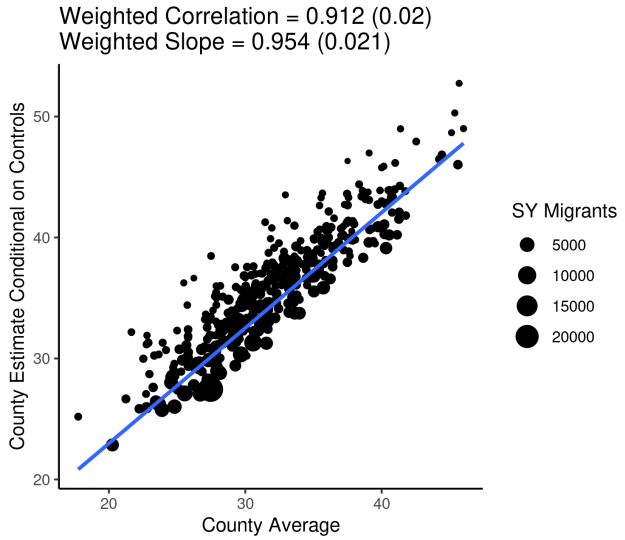
Note: Figures show the average values, by quarter, of integration measures for users in the Syrian migrant sample with an observed arrival in 2015 or 2016. The measures are share of friends native (left column) and the share of content consumed in German (right column). Appendix C provides more details on each measure. The top row shows overall trends. In the bottom row each observation's shape and color represents a gender-by-age group.

Figure A7: Regional Estimates With and Without Controls

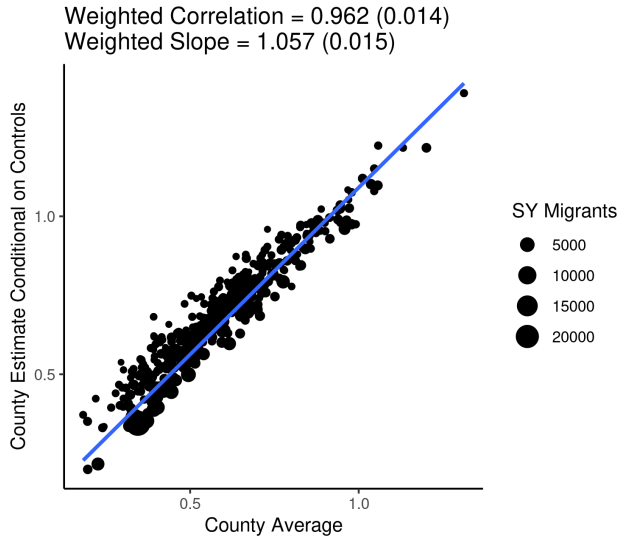
(a) Friending



(b) Language

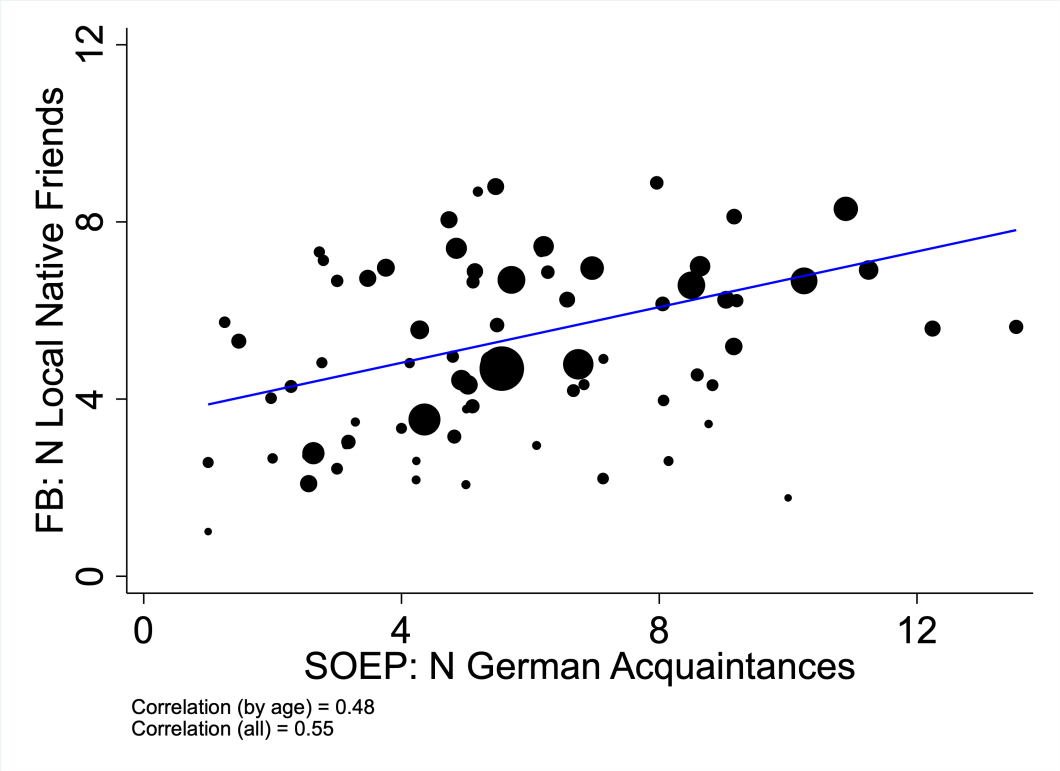


(c) Groups



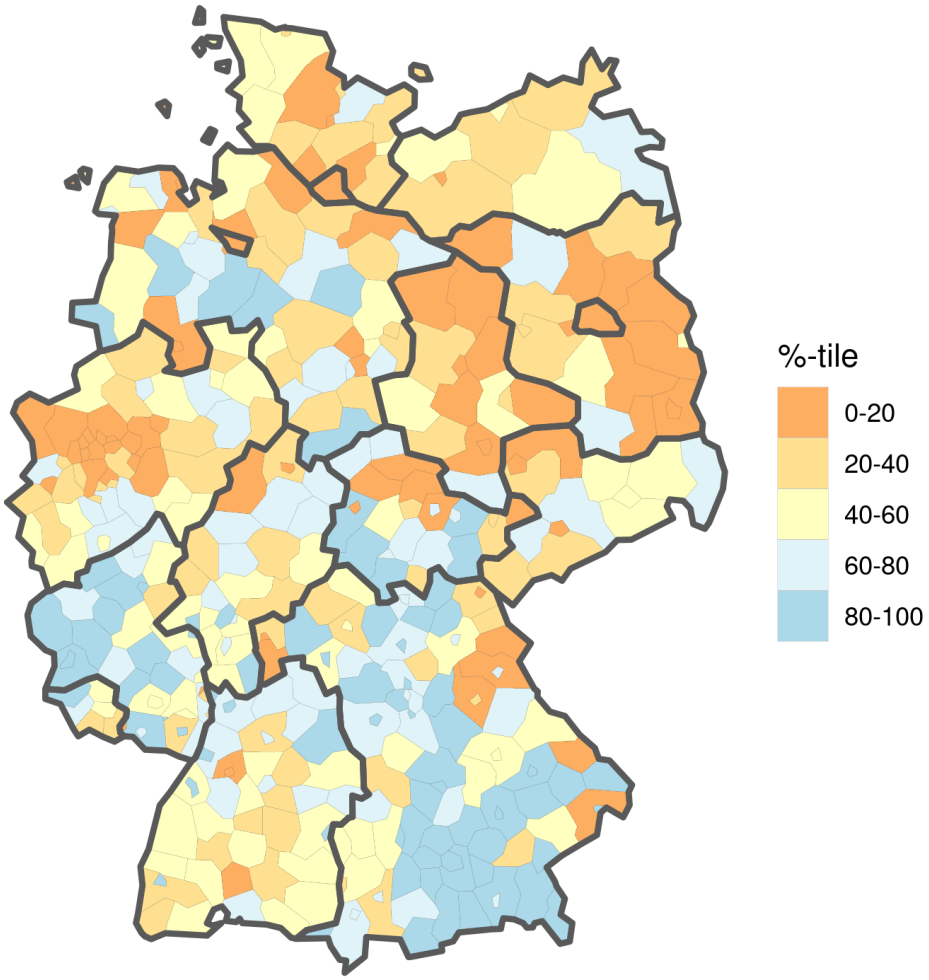
Note: Figures show the relationship between county averages of integration outcomes among Syrian migrants vs county-level fixed effect estimates constructed from versions of equation 5. The outcomes are a user's number of local German native friends in panel (a), whether the user produces content in German in panel (b), and the number of local native groups a user is in in panel (c). Appendix C provides more details on each measures. The controls in the fixed effect regressions are those used in column 3 of Table A11.

Figure A8: Comparing Regional Estimates of Integration - Facebook vs. SOEP



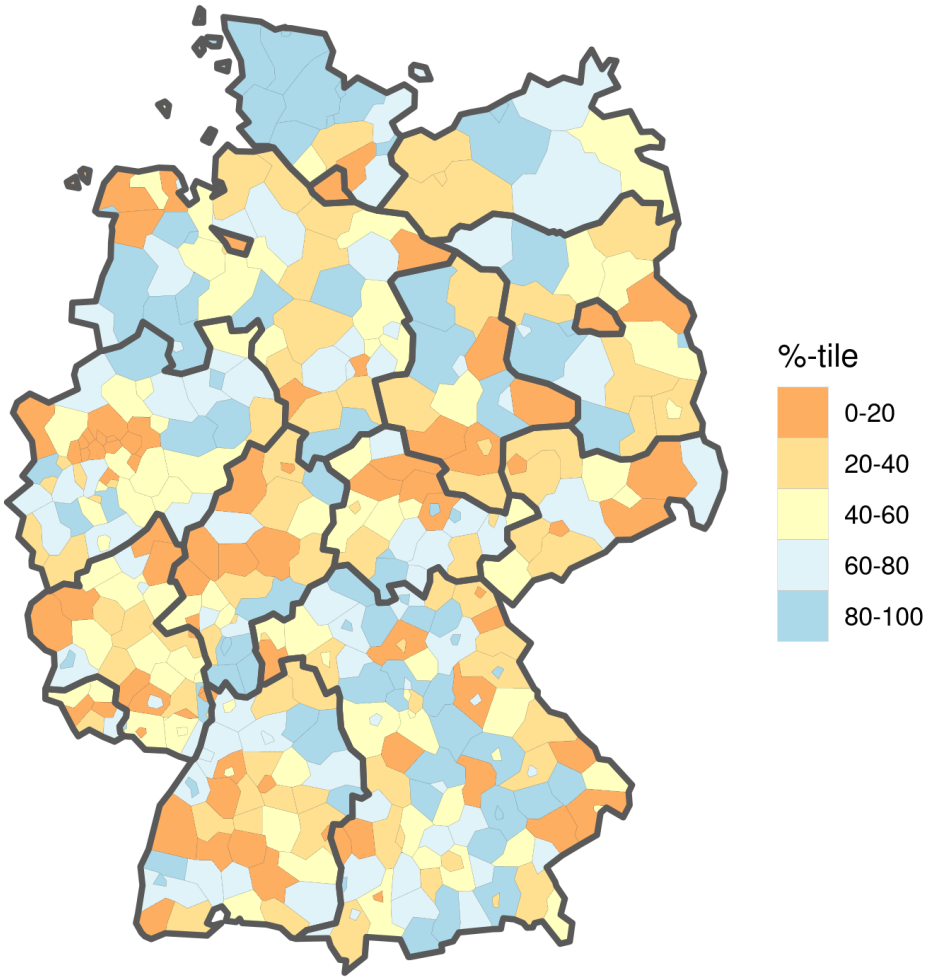
Note: Figure compares estimates of social integration based on our Facebook sample with the average number of acquaintances made by recent Syrian migrants in Germany in the SOEP data. The SOEP question is "How many German people have you met since your arrival in Germany with whom you have regular contact?". Each observation in the Figure is a state-by-age-group combination. The size of each dot corresponds to the number of Syrian migrants in the Facebook data. At the bottom of the figure, we report two correlations. The first is a correlation at the state by age-group level, i.e., the same level of aggregation as shown in the plot. The second is a correlation estimated at the state-level, i.e., we further aggregate observations to the state-level and then correlate the two data sources. Both correlations are weighted by the number of Syrian migrants in our Facebook sample.

Figure A9: Regional Estimates of Integration - German Language Usage



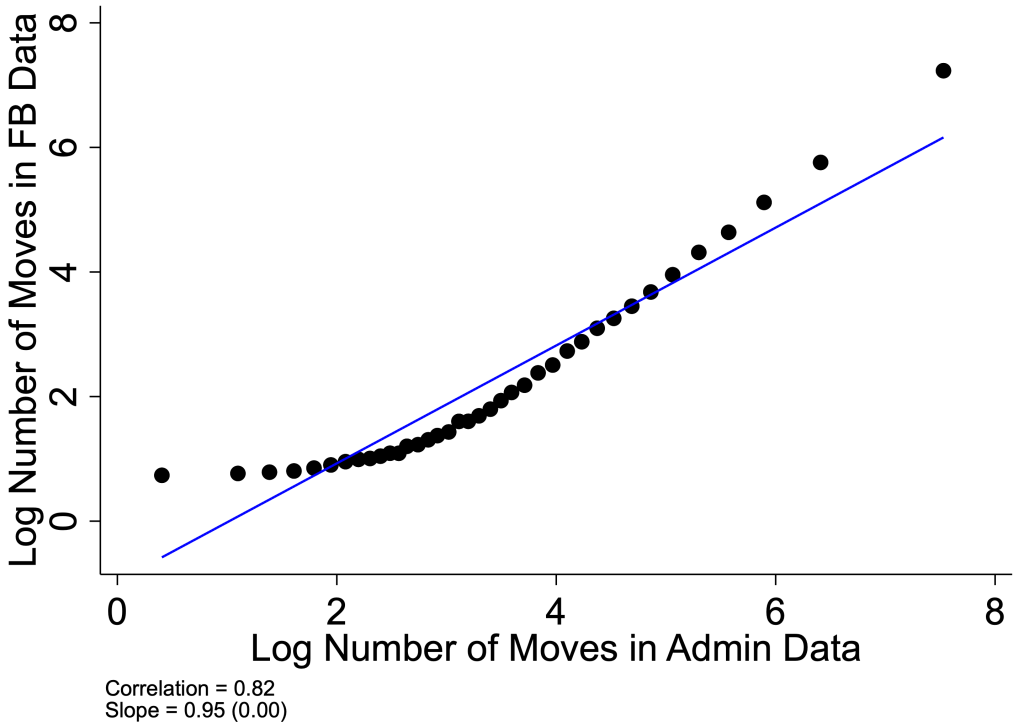
Note: Figure shows county-level estimates of Syrian migrant integration based on the share that produce content in the German language (residualized on regional patterns of Facebook usage). Colors correspond to measure ventiles. Darker orange and blue areas indicate the lowest and highest integration counties, respectively.

Figure A10: Regional Estimates of Integration - Local Native Group Joining



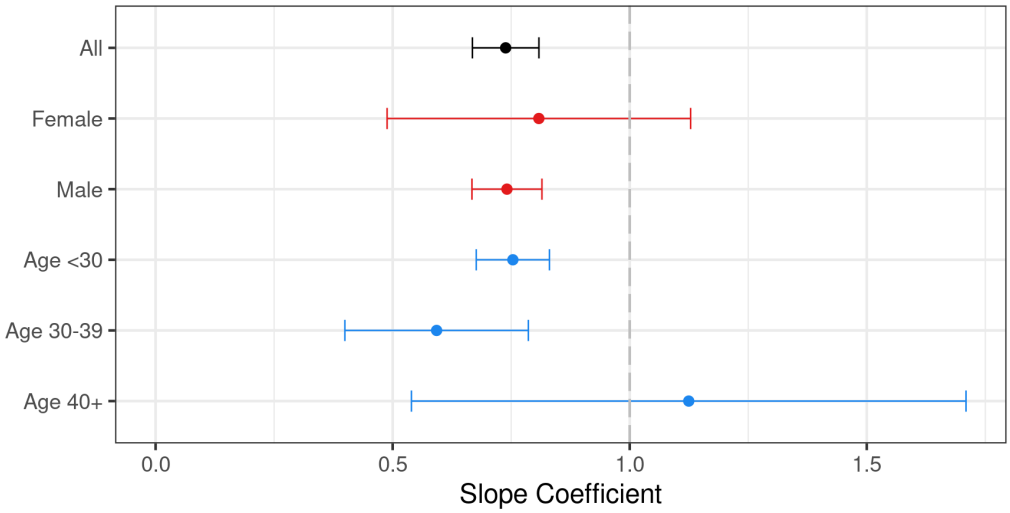
Note: Figure shows county-level estimates of Syrian migrant integration based on the average number of native local groups joined (residualized on regional patterns of Facebook usage. This includes the average number of total groups natives in the region have joined, allowing us to account for variation driven by differential usage of the groups feature in general). Colors correspond to measure ventiles. Darker orange and blue areas indicate the lowest and highest integration counties, respectively.

Figure A11: Comparing Movers in Facebook and Administrative Data



Note: Figure compares the number of moves between counties made by individuals between the ages of 18-64 in the years 2016 and 2017 in Facebook and administrative data. We obtained the administrative date from the German Statistical Office. Each observation in this analysis is a county to county combination. The Figure is a binned scatter plot with 40 equally sized bins. The Figure is weighted by the the total number of individuals living in origin and destination county.

Figure A12: Syrian Migrant Movers - Slope by Demographics

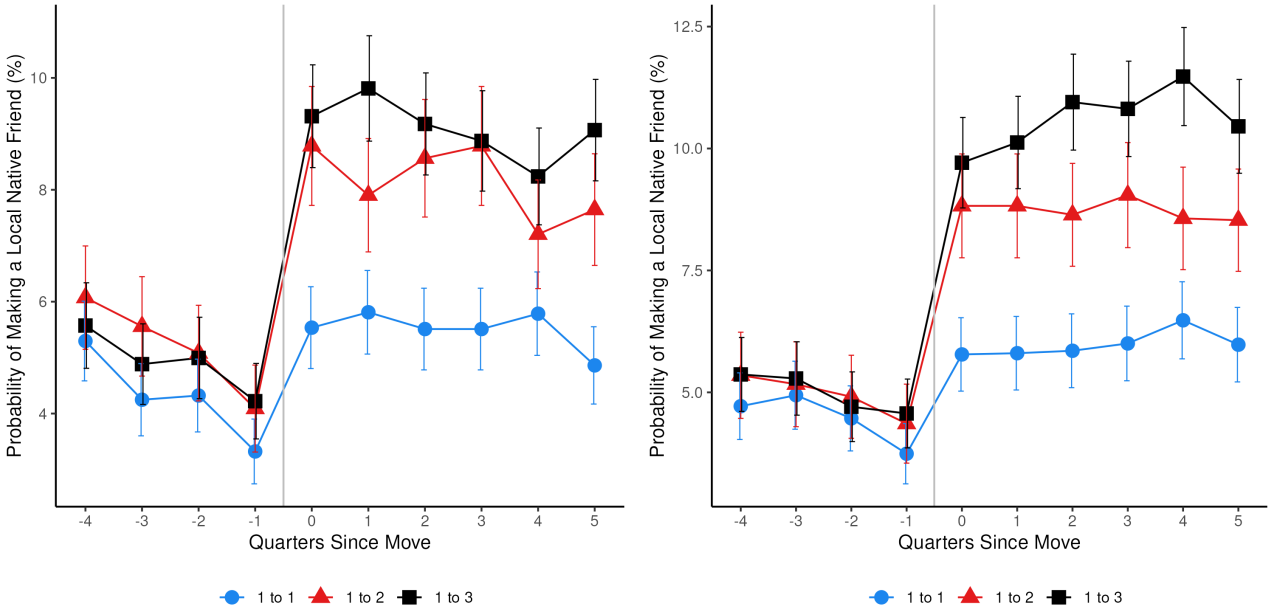


Note: Figure shows slopes corresponding to versions of Figure A15 over certain sub-samples. The coefficient in black corresponds to the slope using the full sample of Syrian migrant movers; the coefficients in red use samples of only one gender; and the coefficients in blue use samples of only one age group. Bars display 95% confidence intervals. The sample sizes used to generate each coefficient are (from top to bottom) 32,853, 6,144, 26,709, 20,796, 8,623, and 3,434.

Figure A13: Change in Syrian Migrants' Friending of Local Natives Around a Move—Split by Friendship Initiator

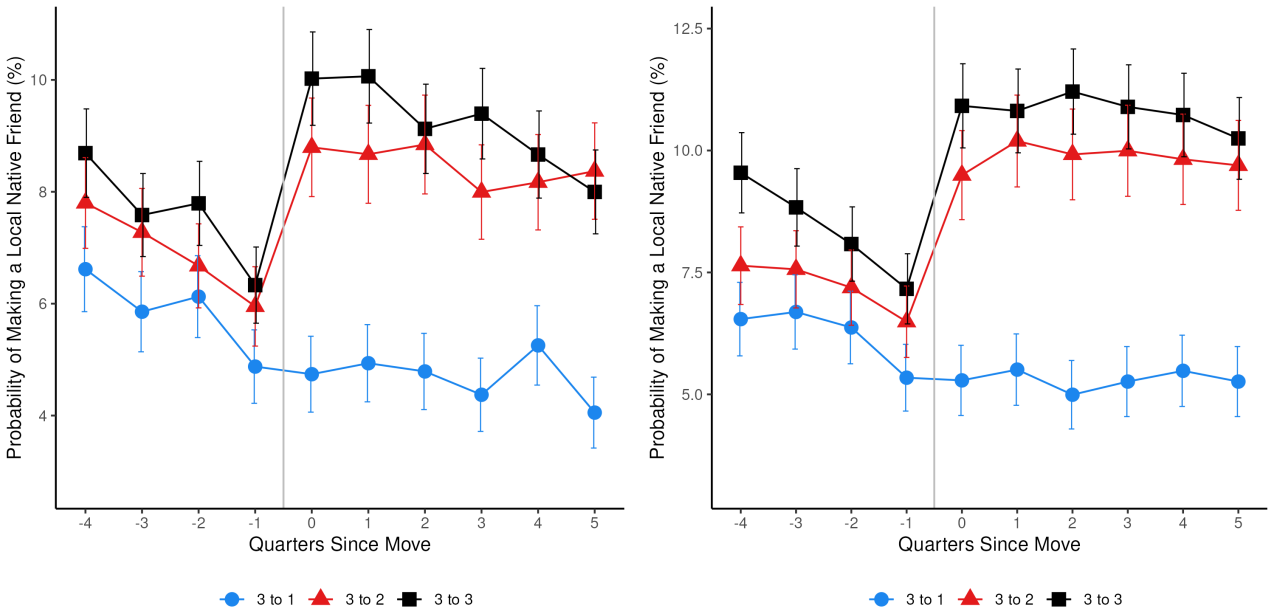
(a) Moving From Bottom Integration Tercile (Only Friendships Initiated by Syrian Migrants)

(b) Moving From Bottom Integration Tercile (Only Friendships Initiated by Native Germans)



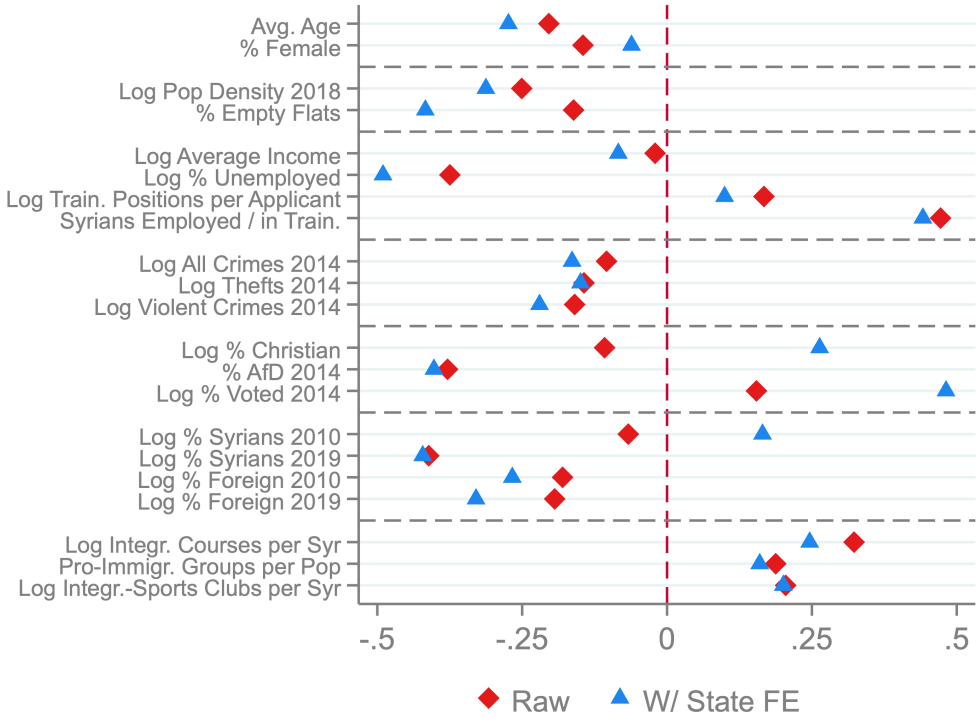
(c) Moving From Top Integration Tercile (Only Friendships Initiated by Syrian Migrants)

(d) Moving From Top Integration Tercile (Only Friendships Initiated by Native Germans)



Note: This figure reproduces the analyses presented in Figure 3. Panels (a) and (b) disaggregate the results of panel (a) of Figure 3, splitting the friendships formed into two groups according to whether it was the Syrian migrant or the local German native who sent the friendship request on Facebook. Panels (c) and (d) repeat the same exercise for panel (b) of Figure 3.

Figure A14: County-Level Univariate Correlations with Friending Integration - Long Version



Note: Figure replicates analysis conducted in Figure 7 using an extended set of covariates. For more information regarding covariates, see Appendix K.

Table A1: Syrian Migrant and German Native Sample Summaries - Additional Measures

Panel (a): Syrian Migrant Sample

	Mean	SD	P10	P25	P50	P75	P90	P99
N Native Friends	9.09	20.54	0	0	2	8	24	151
N Top 50 Native Friends	1.02	2.46	0	0	0	1	3	16
% of Friends Native	3.04	6.19	0.00	0.00	0.80	2.99	8.19	40.25
N Local Other Refugee Country Friends	2.04	3.63	0	0	1	2	6	21
N Local Recent Other Refugee Country Friends	1.04	1.87	0	0	0	1	3	10
% Content Produced in DE	3.39	9.89	0.00	0.00	0.00	2.31	8.48	70.00
% Content Consumed in DE	3.48	8.64	0.00	0.00	0.00	2.91	9.09	60.00
Consumes DE Content (0/100)	41.81	49.32	0	0	0	100	100	100
Account in DE	14.90	35.61	0	0	0	0	100	100
% Groups Local Native	0.88	3.55	0.00	0.00	0.00	0.00	2.22	15.38
Avg. % Native in DE Groups	31.09	30.21	0.15	0.52	25.06	56.44	77.84	92.91

Panel (b): German Native Sample

	Mean	SD	P10	P25	P50	P75	P90	P99
N Native Friends	204.73	189.58	40	74	148	269	443	1151
N Top 50 Native Friends	36.87	8.76	25	33	39	43	46	49
% of Friends Native	82.09	14.70	63.75	77.84	86.67	91.61	94.52	98.16
N Local Other Refugee Country Friends	1.12	2.58	0	0	0	1	3	17
N Local Recent Other Refugee Country Friends	0.05	0.22	0	0	0	0	0	1
% Content Produced in DE	94.49	9.70	81.19	92.90	100.00	100.00	100.00	100.00
% Content Consumed in DE	88.60	16.55	65.84	84.06	95.90	100.00	100.00	100.00
Consumes DE Content (0/100)	97.69	15.02	100	100	100	100	100	100
Account in DE	98.61	11.69	100	100	100	100	100	100
% Groups Local Native	22.07	22.34	0.00	4.55	16.67	33.33	50.00	100.00
Avg. % Native in DE Groups	90.42	5.88	83.52	88.16	91.70	94.15	95.95	100.00

Note: Table presents summary statistics describing users in our Facebook samples. Panel (a) shows users in the Syrian migrant sample. Panel (b) shows users in the German native sample. Each measure is winsorized at the 99% level. Section 1.1 describes sample construction. Appendix C provides more information on how individual-level outcomes are defined.

Table A2: Correlation Between Integration Outcomes, Individual Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) N Local Native Friends	1.00														
(2) N Native Friends	0.64	1.00													
(3) N Top 50 Native Friends	0.61	0.54	1.00												
(4) % of Friends Native	0.69	0.61	0.88	1.00											
(5) N Local SY Friends	0.29	0.14	0.01	0.02	1.00										
(6) N Local Other Refugee Country Friends	0.47	0.30	0.19	0.23	0.54	1.00									
(7) N Local Recent Other Refugee Country Friends	0.28	0.15	0.06	0.07	0.53	0.85	1.00								
(8) % Content Produced in DE	0.45	0.40	0.65	0.67	-0.02	0.17	0.03	1.00							
(9) % Content Consumed in DE	0.46	0.40	0.67	0.68	-0.01	0.18	0.05	0.80	1.00						
(10) Produces DE Content	0.24	0.19	0.27	0.31	0.04	0.11	0.03	0.33	0.33	1.00					
(11) Consumes DE Content	0.37	0.25	0.37	0.40	0.09	0.19	0.12	0.52	0.45	0.27	1.00				
(12) Account in DE	0.32	0.21	0.32	0.34	0.11	0.19	0.13	0.34	0.47	0.25	0.57	1.00			
(13) N Local Native Groups	0.29	0.25	0.25	0.27	0.12	0.14	0.09	0.23	0.25	0.14	0.26	0.24	1.00		
(14) % Groups Local Native	0.33	0.26	0.37	0.40	0.03	0.13	0.05	0.36	0.37	0.19	0.27	0.24	0.61	1.00	
(15) Avg. % Native in DE Groups	0.32	0.23	0.32	0.36	0.01	0.14	0.08	0.33	0.35	0.26	0.38	0.36	0.43	0.47	1.00

Note: Table presents correlations at the user level across outcome measures for the Syrian migrant sample. Each measure is winsorized at the 99% level. Appendix C provides more information on how outcomes are defined.

Table A3: Correlation Between Integration Outcomes, Individual Level - With Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) N Local Native Friends	1.00														
(2) N Native Friends	0.61	1.00													
(3) N Top 50 Native Friends	0.60	0.54	1.00												
(4) % of Friends Native	0.69	0.62	0.86	1.00											
(5) N Local SY Friends	0.20	0.07	0.02	0.03	1.00										
(6) N Local Other Refugee Country Friends	0.39	0.24	0.17	0.20	0.46	1.00									
(7) N Local Recent Other Refugee Country Friends	0.20	0.09	0.05	0.06	0.45	0.83	1.00								
(8) % Content Produced in DE	0.43	0.38	0.61	0.63	-0.01	0.15	0.03	1.00							
(9) % Content Consumed in DE	0.44	0.39	0.63	0.63	0.00	0.16	0.04	0.77	1.00						
(10) Produces DE Content	0.19	0.15	0.21	0.24	0.02	0.06	-0.00	0.27	0.27	1.00					
(11) Consumes DE Content	0.31	0.21	0.32	0.33	0.02	0.12	0.05	0.48	0.40	0.21	1.00				
(12) Account in DE	0.25	0.17	0.26	0.27	0.04	0.12	0.07	0.28	0.42	0.19	0.51	1.00			
(13) N Local Native Groups	0.25	0.22	0.24	0.26	0.03	0.08	0.03	0.23	0.25	0.13	0.22	0.18	1.00		
(14) % Groups Local Native	0.28	0.23	0.33	0.36	0.02	0.09	0.02	0.31	0.32	0.14	0.22	0.18	0.63	1.00	
(15) Avg. % Native in DE Groups	0.23	0.17	0.26	0.29	-0.04	0.05	-0.00	0.27	0.29	0.19	0.30	0.27	0.43	0.42	1.00

Note: Table presents correlations at the user level across outcome measures for the Syrian migrant sample. Each measure is first winsorized at the 99% level. Appendix C provides more information on how outcomes are defined. Before constructing the correlations, each measure is residualized on the individual-level controls used in column 3 of Table A11.

Table A4: Syrian Migrant Integration by Demographics - Language and Groups

	Produces Content in German (0/100)				N Local Native Groups			
Age 25 - 34	-2.407***	-2.241***	-2.275***	-3.312***	0.167***	0.171***	0.136***	0.140***
	(0.204)	(0.203)	(0.203)	(0.596)	(0.006)	(0.006)	(0.006)	(0.019)
Age 35 - 44	-7.133***	-7.161***	-6.875***	-6.615***	-0.002***	-0.007***	0.039*	0.072**
	(0.238)	(0.237)	(0.237)	(0.733)	(0.007)	(0.007)	(0.007)	(0.023)
Age 45 - 54	-13.651***	-13.798***	-12.553***	-16.243***	-0.184***	-0.189***	-0.064***	-0.070***
	(0.306)	(0.305)	(0.307)	(0.854)	(0.010)	(0.010)	(0.009)	(0.027)
Age 55+	-18.045***	-18.134***	-16.451***	-24.395***	-0.298***	-0.300***	-0.088***	-0.228***
	(0.382)	(0.380)	(0.384)	(1.116)	(0.012)	(0.012)	(0.012)	(0.035)
Female	-15.767***	-15.560***	-16.725***	-18.765***	-0.202***	-0.200***	-0.372***	-0.447***
	(0.164)	(0.164)	(0.173)	(0.418)	(0.005)	(0.005)	(0.005)	(0.013)
Household Member in DE 1+ Year Prior	-2.420***	-2.298***	-2.113***		-0.057***	-0.058***	-0.060***	
	(0.384)	(0.383)	(0.382)		(0.012)	(0.012)	(0.012)	
Non-Household Family in DE 1+ Year Prior	3.418***	3.451***	4.045***		0.023***	0.025***	0.030***	
	(0.347)	(0.345)	(0.345)		(0.011)	(0.011)	(0.010)	
Quarters Since DE FEs	X	X	X	X	X	X	X	X
Prev Quarters in NUTS3 FEs	X	X	X	X	X	X	X	X
Personal Usage Controls	X	X	X	X	X	X	X	X
County FEs		X	X	X		X	X	X
Log (1 + Total Outside Germany Friends)			X	X			X	X
Log (1 + Total Other Groups)			X	X			X	X
Log (1 + Total Content Produced Past Year)			X	X			X	X
Household FE				X				X
N	349,072	349,072	349,072	84,216	349,072	349,072	349,072	84,216
R-Squared	0.098	0.108	0.113	0.590	0.059	0.076	0.133	0.606
Sample Mean	30.401	30.401	30.401	27.215	0.545	0.545	0.545	0.574

Note: Table shows results from regressing various measures on language- and groups-based measures of integration. Each observation in every column is a user in the Syrian migrant Facebook sample. Columns 1 and 5 include controls for age and gender, as well as fixed effects for the number of quarters on Facebook in their current county and the number of quarters since arrival in Germany. For the latter fixed effect, we use a single dummy value for those for which we do not observe arrival, but obtain nearly identical results if we instead drop these users. We also include dummies for whether the user has another Syrian migrant household member or non-household family member in Germany more than year prior to their arrival. For all users not in the "observe arrival timing" sample, these two dummies are set to 0. Columns 2 and 6 add county fixed effects. Columns 3 and 7 add controls for each user's total number of friends outside Germany, total number of non-local/native groups joined, and total amount of content produced in the last year. Columns 4 and 8 add a household fixed effect, limiting to households for which we observe more than one Syrian migrant. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

Table A5: Syrian Migrant Integration by Demographics - Other Measures

	N Native Friends	N Top 50 Native Friends	% of Friends Native	% Content Produced in DE	% Content Consumed in DE	Account in DE	% Groups Local Native	Avg. % Native in DE Groups
Age 25 - 34	-0.894***	0.004***	-0.467***	0.076**	0.078***	-2.683***	0.197***	-0.136***
	(0.184)	(0.014)	(0.032)	(0.044)	(0.038)	(0.160)	(0.010)	(0.160)
Age 35 - 44	-4.728***	-0.263***	-1.446***	-0.694***	-0.749***	-7.099***	0.043	-4.347***
	(0.216)	(0.016)	(0.038)	(0.051)	(0.044)	(0.187)	(0.012)	(0.187)
Age 45 - 54	-6.928***	-0.454***	-1.927***	-1.245***	-1.298***	-7.676***	-0.164***	-6.940***
	(0.279)	(0.021)	(0.049)	(0.066)	(0.057)	(0.241)	(0.015)	(0.254)
Age 55+	-8.157***	-0.421***	-1.862***	-1.221***	-1.327***	-6.151***	-0.350***	-7.334***
	(0.349)	(0.026)	(0.061)	(0.083)	(0.072)	(0.302)	(0.019)	(0.360)
Female	-7.188***	-0.787***	-2.334***	-2.339***	-2.154***	-5.377***	-0.485***	-11.601***
	(0.157)	(0.012)	(0.027)	(0.037)	(0.032)	(0.136)	(0.009)	(0.137)
Household Member in DE 1+ Year Prior	-0.610	-0.030	0.013	0.146	-0.057	0.182	-0.014	-0.875***
	(0.347)	(0.026)	(0.061)	(0.082)	(0.071)	(0.300)	(0.019)	(0.295)
Non-Household Family in DE 1+ Year Prior	0.667***	0.075***	0.360***	0.535***	0.404***	3.659***	0.098***	2.649***
	(0.314)	(0.023)	(0.055)	(0.074)	(0.064)	(0.271)	(0.017)	(0.257)
Quarters Since DE FEs	X	X	X	X	X	X	X	X
Prev Quarters in County FEs	X	X	X	X	X	X	X	X
Personal Usage Controls	X	X	X	X	X	X	X	X
County FEs	X	X	X	X	X	X	X	X
Log (1 + Total Outside Germany Friends)	X	X	X	X	X	X	X	X
Log (1 + Total Other Groups)	X	X	X	X	X	X	X	X
Log (1 + Total Content Produced Past Year)	X	X	X	X	X	X	X	X
N	349,072	349,072	349,072	345,814	346,367	349,072	345,162	237,563
R-Squared	0.064	0.111	0.163	0.121	0.125	0.083	0.077	0.171
Sample Mean	10.592	1.101	3.221	3.388	3.474	14.896	0.754	31.091

Note: Table shows results from regressing various measures on outcomes for Syrian migrants in the Facebook sample. All columns include controls for age, gender, time spent on Facebook, number of friends outside Germany, total number of non-local/native groups joined, and total amount of content produced in the last year. They include fixed effects for county, the number of quarters since arrival in Germany (with a single dummy for those for which we do not observe arrival) and the number of quarters on Facebook in their current county. They also include dummies for whether the user has another Syrian migrant household member or non-household family member in Germany more than year prior to their arrival. Column 8 limits to migrants who are members of at least one group of majority users in Germany. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

Table A6: Signal Correlation Between Outcomes, Regional Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline Integration Measures							
(1) SY Migrants - N Local Native Friends	X						
(2) SY Migrants - Produced Content in DE	0.65	X					
(3) SY Migrants - N Local Native Groups	0.27	0.55	X				
(4) SY Migrants - N Local SY Friends	-0.04	-0.55	-0.42	X			
Panel B: Decomposition of Integration Measures							
(5) General Friendliness	0.64	0.31	-0.04	0.11	X		
(6) Relative Friending	0.77	0.56	0.43	-0.16	-0.05	X	
Panel C: Labor Market Integration Measure							
(7) Share Syrians in Employment or Training	0.46	0.63	0.14	-0.36	0.29	0.34	X

Note: Table presents signal-adjusted correlations between county-level estimates. The outcomes in panel (a) are the regional averages of Syrian migrants after residualizing on local German natives' Facebook usage, as described in Section 3. The outcomes in panel (b) are the regional decomposition measures described in Section 4.1. Row 5 is general friendliness, generated as a regional average of German natives after residualizing on local German natives' Facebook usage. Row 6 is relative friending, generated as the quotient from dividing the measure in row 1 by the measure in row 5. The outcome in panel C is an external county-level measure of the share of all Syrians that are employed or in training programs as described in Section 5.1. Correlations are weighted by the number of Syrian migrant users in each county. Our methodology for adjusting correlations to remove sampling error is described in Appendix E.

Table A7: Syrian Migrant Mover and Comparable Non-Mover Sample Summaries

	All		To Below Median Place		To Above Median Place	
	Movers	Matched	Movers	Matched	Movers	Matched
% Female	18.70	18.70	19.54	19.54	17.95	17.95
Avg Age	27.97	27.49	27.98	27.51	27.97	27.47
Avg Qs in DE	6.47	6.42	6.54	6.50	6.40	6.36
Avg Friends Made (total in year)	44.72	43.97	44.78	44.07	44.66	43.87
% of Qs Produ in DE	45.77	45.01	44.31	44.01	47.09	45.90
% of Qs Makes Native Local Friend	11.80	17.18	10.51	16.72	12.96	17.60

Note: Table presents summary statistics describing the movers underlying Figure A15 and their matched non-movers in their origin. Movers are matched to non-movers on county, time, age group (18-29, 30-39, 40+), gender, and the year we first observed the user on Facebook in Germany. To be in the final sample, a mover must be matched to five or more non-movers in both the origin and destination. Measures are constructed using the movers' information in the year prior to the move and their matched users in the origin location and time. Matched non-mover summaries are generated by first constructing measures within each mover's set of matched movers, then averaging across these measures. "Avg Friends Made" is constructed from summing quarterly measures that are winsorized at the 99% level across all migrant user-by-quarter observations. "% of Qs Makes Native Local Friend" is residualized by local natives' Facebook usage.

Table A8: Δ Migrant Mover Friending Integration vs. Matched Non-Movers: Robustness

	Change Quarterly Prob of Making Native Local Friend			
Dest-Origin Quarterly Prob of SY Making Native Local Friend	0.738***		0.758***	0.724***
	(0.036)		(0.051)	(0.053)
Origin Quarterly Prob of SY Making Native Local Friend		-0.712***		
		(0.037)		
Dest Quarterly Prob of SY Making Native Local Friend		0.773***		
		(0.037)		
Quarter FEs	X	X	X	X
Origin County FEs			X	
Dest County FEs				X
N	32,853	32,853	32,849	32,845
Sample Mean	0.934	0.934	0.933	0.938

Note: Table shows results from regressions exploring the change in friending of Syrian migrants to German natives, before and after a move within Germany. Column 1 corresponds to the relationship depicted in Figure A15. Column 2 regresses each component of the difference in the right-hand side measure in column 1 separately on the outcome. Columns 3 and 4 repeat column 1 with origin and destination fixed effects, respectively. We correct for sampling error in the right-hand side measures by randomly splitting the individual-level non-mover data into two halves and instrumenting for one set of averages with the other. See Appendix E for more information this procedure. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

Table A9: Native Mover and Comparable Non-Mover Sample Summaries

Panel A: Yearly General Friendliness Sample

	All		To Below Median Place		To Above Median Place	
	Movers	Matched	Movers	Matched	Movers	Matched
% Female	51.95	51.95	51.74	51.74	52.07	52.07
Avg Age	33.70	33.34	34.21	33.87	33.39	33.03
Avg Friends Made (total in year)	21.22	20.11	19.71	19.68	22.12	20.36
Yearly General Friendliness	5.33	9.74	4.81	9.49	5.63	9.89

Panel B: Yearly Relative Friending Sample

	All		To Below Median Place		To Above Median Place	
	Movers	Matched	Movers	Matched	Movers	Matched
% Female	52.75	52.75	52.48	52.48	52.90	52.90
Avg Age	31.90	31.86	32.35	32.35	31.65	31.58
Avg Friends Made (total in year)	28.19	20.70	26.41	20.20	29.21	20.99
Yearly Relative Friending	0.20	0.23	0.17	0.22	0.21	0.23

Note: Table presents summary statistics describing the users underlying Figure 5. Panels (a) and (b) show summaries for movers and matched non-movers in panels (a) and (b) of Figure 5, respectively. Measures are constructed using movers' information in the year prior to the move and their matched users in the origin location and time. Matched non-mover summaries are generated by first constructing measures within each mover's set of matched movers, then averaging across these measures. "Avg Friends Made" is constructed from summing quarterly measures winsorized at the 99% level across all native user-by-quarter observations. The final outcome in each panel is residualized by local natives' Facebook usage.

Table A10: Change in Native Mover SY Migrant Friending vs Matched Non-Movers

	Change in Mover Yearly General Friendliness				Change in Mover Yearly Relative Friending			
Dest-Origin Yearly General Friendliness	0.685***		0.711***	0.602***				
	(0.004)		(0.005)	(0.005)				
Origin Yearly General Friendliness		-0.636***						
		(0.005)						
Dest Yearly General Friendliness		0.739***						
		(0.005)						
Dest-Origin Yearly Relative Friending					0.959***		0.926***	0.988***
					(0.064)		(0.094)	(0.086)
Origin Yearly Relative Friending						-0.988***		
						(0.071)		
Dest Yearly Relative Friending						0.926***		
						(0.071)		
Quarter FEs	X	X	X	X	X	X	X	X
Origin County FEs			X				X	
Dest County FEs				X				X
N	1,771,041	1,771,041	1,771,041	1,771,041	1,096,874	1,096,874	1,096,874	1,096,874
Sample Mean	3.160	3.160	3.160	3.160	0.005	0.005	0.005	0.005

Note: Table shows results from regressions exploring the change in friending of natives, before and after a move within Germany. Columns 1 and 5 correspond to the relationships depicted in panels (a) and (b) of Figure 5. Columns 2 and 6 regress each component of the difference in the right-hand side measure in columns 1 and 5 separately on the outcome. Columns 3 and 7 repeat columns 1 and 5 with origin fixed effects; columns 4 and 8 repeat columns 1 and 5 with destination fixed effects. We correct for sampling error in the right-hand side measures by randomly splitting the individual-level non-mover data into two halves and instrumenting for one set of averages with the other. See Appendix E for more information this procedure. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

B Construction of "Native German" Sample

For many of our analyses we use a sample of Facebook users, which we refer to as "German natives", that meet *both* criteria 1 and 2 described below (as well as the primary sample inclusion criteria described in Section 1.1). Our methodology is not intended to proxy for citizenship status or ethnicity; rather it generates a sample of users who generally use the German language and—according to self-reported profile information and home region predictions—appear to have lived in Germany for a substantial amount of time. This will include, for example, individuals of Syrian descent who report a German hometown and primarily use the German language on Facebook. For more details, see footnote 3.

- **Criteria 1:** The user meets *one* of the following
  - The user produces $\geq$ 75% of their content in German
  - The user produces $\geq$ 50% of their content in German, AND lists a German hometown or high school on their profile
- **Criteria 2:** The user meets *all* of the following
  - Does not list a hometown in a "top migration country
  - Does not list a high school in a "top migration country
  - Did not first have a predicted home region in a "top migration country

The top migration countries are the 15 countries outside of the European Union and within Eastern Europe, the Middle East, or Africa with the most foreign nationals in Germany (for the full list, see: https://tinyurl.com/8jk2d4yd).

C Individual-Level Outcomes

We consider three dimensions of social integration; within each dimension, we construct a number of measures, though we focus on a primary measure within each dimension, which is noted in **bold**.

1. Friendship Measures
   (a) ***N Local Native Friends:* The number of friends a user has in the same county or a bordering county that are in the German native sample.**
   (b) *N Native Friends:* The number of friends a user has in the German native sample.
   (c) *N Top 50 Native Friends:* The number of a user's closest 50 friends that are in the German native sample.
   (d) *% of Friends Native:* The percent a user's total friends that are in the German native sample.
2. Language Measures
   (a) *% Content Produced in DE*: The share of content a user produces (e.g., in posts, comments) that is in German. "Half-life" of 30 days (i.e., a post 30 days ago is weighted as half a post today).
   (b) *% Content Consumed in DE*: The share of the content a user engages with by using the "react" and "comment" features that is in German. 1 comment = 7 reactions. "Half-life" of 30 days.
   (c) ***Produces Any DE Content* : An indicator for "% Content Produced in DE" is >1%.**
   (d) *Consumes Any DE Content*: An indicator for "% Content Consumed in DE" is >1%.
   (e) *Account in DE*: Whether a user selected German as their language in their account settings.
3. Local Group Participation Measures
   (a) ***N Local Native Groups*: The number of groups a user is in that have 5 - 5,000 users; $\geq$ 90% of users in Germany and $\geq$ 75% of users in one NUTS2 region; and $\geq$ 50% of users in the German native sample.**
   (b) *% Groups Local Native*: The share of groups a user is in that match the criteria in "N Local Native Groups."
   (c) *Avg. % Native in DE Groups*: Among groups a user is in which have >90% of users in Germany, the average share of users that are German natives.

We also observe the following additional measures at the individual level:

- *N Local Syrian Friends*: The number of friends a user has in the same county or a bordering county that are in the Syrian migrant sample
- *N Local Other Migrant Country Friends*: The number of friends a user has in the same or bordering county that are migrants (determined by hometown, high school, or past usage) from one of the five countries with the most asylum applicants in Germany in 2020 other than Syria: Turkey, Afghanistan, Iraq, Nigeria, and Iran.
- *N Local Recent Other Migrant Country Friends*: The number of friends a user has matching the "N Local Other Migrant Country Friends" criteria with observed arrival in Germany 2015 or later.[19]

[19] As described in Section 1.1, users with an "observed arrival timing" are those who first used Facebook outside of Germany.

D Syrian Migrant Integration by Demographics

We explore the heterogeneity in integration outcomes by demographics formally using the the following multivariate regression model:

$$Y_{i,j} = \alpha_0 + \alpha_1 Z_i + \psi_{j(i)} + \epsilon_i. \tag{5}$$

For the results in columns 1-4 of Table A11, $Y_{i,j}$ is the number of native local friends of individual i has. All specifications include various controls Z_i for the amount of time users spend on Facebook, ensuring that differences in observed integration outcomes are not driven by variation in the intensity of Facebook usage. We also include fixed effects for the user's number of quarters since arrival in Germany and the number of quarters living in their current county.

In column 1, Z_i also includes dummies for age, gender, and whether the user has another Syrian migrant household member or non-household family member who was in Germany more than a year prior to their arrival.[20] Consistent with the univariate patterns in Figure 1, we find that younger and male Syrians befriend disproportionately many local German natives. All else equal, a female Syrian migrant has 3.7 fewer local native friends than a male does. Similarly, a Syrian migrant aged 55 or older has 4.6 fewer native local friends than a comparable individual under the age of 25. Column 1 also shows that, while migrants with a family member who arrived earlier in Germany *outside* of the household have more local native friends, individuals with an earlier arriving Syrian migrant *inside* their household have fewer local native friends. This result adds to prior findings that connections to other migrants support integration in some settings and hinder it in others (e.g., Lazear, 1999; Edin, Fredriksson and Åslund, 2003; Cutler, Glaeser and Vigdor, 2008; Damm, 2009; Beaman, 2012; Martén, Hainmueller and Hangartner, 2019). In our context, the results suggest that somewhat-distant familial connections might provide support and guidance to help the social integration of newly arriving migrants, whereas the presence of close household connections might reduce the need to form connections with local natives.

Column 2 adds fixed effects for the Syrian migrants' current county of residence, $\psi_{j(i)}$, to the regression. The R^2 increases by 21% from 0.132 to 0.160, consistent with the presence of important regional differences in the social integration of Syrian migrants. The coefficients on the demographic characteristics in Z_i are largely unaffected by the addition of county fixed effects, suggesting there is a little selection based on these characteristics into more or less integrated places.

Column 3 adds controls for each user's total number of friends outside Germany, total number of groups joined, and total amount of recent content produced. These controls absorb additional variation in individuals' Facebook usage patterns beyond those in column 1, but could also remove variation in the true sociability of individuals that might influence their ability and desire to socially integrate with natives. While most coefficients remain largely unchanged, the gender coefficient falls somewhat in absolute terms, from -3.6 to -3.2. A possible interpretation is that Syrian migrant men generally have larger social networks, but, even conditional on overall network size, also make more German friends.

In column 4 of Table A11 we add household fixed effects while dropping individuals without additional household members from the sample. Even within the same household, and conditional on general Facebook usage patterns, younger and male Syrian migrants are better socially integrated.

Appendix Table A4 presents results analogous to column 1-4 of Table A11 for our key language- and group-based measures of social integration, and Table A5 presents results analogous to column 3 of Table A11 for a number of other outcomes. Across all measures, we find highly consistent relationships between age, gender, and family connections and the social integration of Syrian migrants.

One concern with this analysis may be that, despite our strict controls for Facebook usage and the consistency of our results across outcome, the observed differences in integration outcomes across demographic groups may still be driven by patterns of Facebook usage, rather than reflecting true de-

[20] Family and household information is determined through self-reports and model-based imputations. Similar data are used in Bailey et al. (2019*a*) and Chetty et al. (2022*a*,*b*).

Table A11: Syrian Migrant Integration by Demographics - Friending to Natives

	Facebook Sample N Local Native Friends				SOEP Sample N German Acquaintances	
Age 25 - 34	-1.012***	-0.894***	-0.873***	-1.148***	-0.839*	-1.089**
	(0.053)	(0.052)	(0.052)	(0.129)	(0.47)	(0.47)
Age 35 - 44	-2.963***	-3.019***	-2.941***	-2.375***	-1.116*	-1.070*
	(0.062)	(0.061)	(0.061)	(0.158)	(0.58)	(0.58)
Age 45 - 54	-4.012***	-4.102***	-4.147***	-4.765***	-2.362***	-2.238***
	(0.080)	(0.079)	(0.079)	(0.184)	(0.78)	(0.77)
Age 55+	-4.548***	-4.531***	-4.586***	-7.226***	-3.378***	-3.594***
	(0.100)	(0.098)	(0.099)	(0.241)	(1.24)	(1.23)
Female	-3.676***	-3.610***	-3.225***	-3.267***	-1.421***	-1.512***
	(0.043)	(0.042)	(0.045)	(0.090)	(0.47)	(0.48)
Household Member in DE 1+ Year Prior	-0.377***	-0.290**	-0.352***			
	(0.100)	(0.099)	(0.099)			
Non-Household Family in DE 1+ Year Prior	0.524***	0.621***	0.421***			
	(0.091)	(0.089)	(0.089)			
Quarters Since DE FEs	X	X	X	X	X	X
Prev Quarters in NUTS3 FEs	X	X	X	X		
Personal Usage Controls	X	X	X	X		
County / State FEs		X	X	X		X
Log (1 + Total Outside Germany Friends)			X	X		
Log (1 + Total Other Groups)			X	X		
Log (1 + Total Content Produced Past Year)			X	X		
Household FE				X		
N	349,072	349,072	349,072	84,216	1,095	1,095
R-Squared	0.132	0.160	0.165	0.658	0.048	0.093
Sample Mean	5.029	5.029	5.029	4.195	6.232	6.232

Note: Table explores variation in migrants' social integration. Each observation in columns 1-4 is a user in the Syrian migrant Facebook sample. Column 1 includes (i) controls for age and gender; (ii) fixed effects for the number of quarters on Facebook in their current county and the number of quarters since arrival in Germany (we use a single dummy value for those for which we do not observe arrival, but obtain nearly identical results if we instead drop these users); (iii) dummies for whether the user has another Syrian migrant household member or non-household family member in Germany prior to their arrival. (For all users not in the "observe arrival timing" sample, these two dummies are set to 0); and (iv) the following measures of the Facebook usage intensity: linear controls for log(0.5 + minutes on FB in the last 28 days), log(91 - days on Facebook out of the last 90), log(1081 - days on Facebook out of the last 1080). Column 2 adds county fixed effects. Column 3 adds controls for each user's total number of friends outside Germany, total number of non-local/native groups joined, and total amount of content produced in the last year. Column 4 adds a household fixed effect, limiting to households for which we observe more than one Syrian migrant. Columns 5 and 6 use data from the Socio-Economic Panel in 2016. The dependent variable in these columns is the number of new acquaintances made in Germany (see footnote 6). Each observation is a recent migrant from Syria living in Germany as of the date of the survey. Both columns 5 and 6 include controls for the number of quarters in Germany. Column 6 also controls for state fixed-effects. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

mographic variation in social integration. To address this concern, we also look at related outcomes in the Socio-Economic Panel data, namely the number of native acquaintances made in Germany among a sample of recent Syrian migrants. In 2016, the SOEP administered a survey specifically targeted at recent migrants to Germany. We focus on the 1,095 Syrian migrants in the data that are 18+ years old.

Columns 5 and 6 show that the patterns of friending across demographics in the SOEP data mirror those we observe in the Facebook data in columns 1-4. Female and older migrants have fewer local acquaintances than male and younger migrants, respectively, on average. This holds with state fixed effects in column 6. Indeed, even the coefficient estimates using the Facebook and SOEP data are generally quite similar. We interpret this as reassuring as it shows that the patterns of social integration we identify in the Facebook data align closely with available survey evidence. The Facebook data, however, is much larger and more detailed, allowing us to more precisely explore the spatial variation in integration and to better understand the determinants of this variation.

E Assessing the Reliability of Regional Estimates

A potential concern with our regional estimates of integration outcomes is that the differences we observe might be due to sampling error, instead of capturing actual differences in the parameters of interest. In this appendix we explore this concern and describe the methods used to address it.[21]

To assess the degree to which our variation is driven by sampling error, we seek an estimate of:

$$r = \frac{Var(\delta_j)}{Var(\delta_j) + Var(\epsilon_j)} \tag{6}$$

Here δ_j is the true (un-observable) parameter for county j, $Var(\delta_j)$ is the variance of that parameter across all counties, and $Var(\epsilon_j)$ is the variance due to sampling error (noise) when we measure our estimate $Var(\hat{\delta}_j)$, such that $Var(\hat{\delta}_j) = Var(\delta_j) + Var(\epsilon_j)$. Our outcome of interest is the reliability, r.

We estimate r in two ways: (i) a "split sample" estimate generated by randomly splitting the individual-level data in half (within counties) and comparing the resulting estimates; and (ii) a "standard error-based" estimate generated by comparing the magnitudes of the standard error squared of each estimate with the variance of the estimates across counties.

Formally, our "split sample" estimates are given by:

$$\hat{r} = Corr(\hat{\delta_j^1}, \hat{\delta_j^2}) \cdot \frac{\sqrt{Var(\hat{\delta_j^1}) Var(\hat{\delta_j^2})}}{Var(\hat{\delta_j})} \tag{7}$$

Where $\hat{\delta_j}$ is the county-level estimate of δ in county j, the average of individual-level measures across users in the county; $Var(\hat{\delta_j^1})$ and $Var(\hat{\delta_j^2})$ are the population-weighted variances of these measures in the first and second split samples; $Var(\hat{\delta_j})$ is the population-weighted variance in the full sample; and $Corr(\hat{\delta_j^1}, \hat{\delta_j^2})$ is the population-weighted correlation.

Our "standard error-based" estimates are given by:

$$\hat{r} = \frac{Var(\hat{\delta_j}) - E[s^2_{\hat{\delta_j}}]}{Var(\hat{\delta_j})} \tag{8}$$

Where $s_{\hat{\delta_j}}$ is the standard error of the county level average $\hat{\delta_j}$ for county j.

The first two columns of Appendix Table A12 show that the reliability of each of our regional averages is around 0.9 or above regardless of the method used. This suggests that 90% or more of the variance in a given regional measure reflects true latent differences rather than sampling error.

As noted in Section 3, there are moderate differences in the Facebook usage of natives across space (largely at the intensive margin) which could affect the raw regional averages we measure. To account for this, our estimates in Figure 2 and Appendix Figures A9 and A10 are constructed after residualizing by differences in natives' Facebook usage. Column 3 of Appendix Table A12 shows split-sample reliability estimates using $\hat{\delta_j^1}$ and $\hat{\delta_j^2}$ that have been residualized in this same manner. The reliability estimates are largely unchanged, suggesting they are not driven by regional differences in usage.

In Section 4.1, we construct regional measures of *general friendliness* using the German native sample. The sample size for these measures is very large and, accordingly, the reliability estimates using both methods is greater than 0.995. Therefore, essentially all of the sampling error present in our measures of *relative friending* (generated by dividing the Syrian migrant integration outcomes by general friendliness) is driven by the Syrian migrant integration outcomes.

[21] The methods described in this appendix are similar to procedures used in Chetty and Hendren (2018*b*), Chetty et al. (2022*a*), and Chetty et al. (2022*b*).

Table A12: Reliability of County-Level Measures, Syrian Migrant Sample

	Reliability		
	Split-Sample	SE-Based	Split-Sample, Usage Control
N Local Native Friends	0.962	0.961	0.938
Produced Any DE Content	0.909	0.901	0.883
N Local Native Groups	0.948	0.946	0.934
N Local Syrian Friends	0.989	0.989	0.989

Note: Table shows the reliability of county-level measures. In columns 1 and 2 the measures are averages across Syrian migrant users. In column 3 these measures are residualized on extensive and intensive measures of local natives' Facebook usage, as described in Section 3. Reliability is defined by equation 6. The spilt sample reliability estimates are generated using equation 7. The standard error-based reliability estimates are generated using equation 8.

In Table 3 we correlate regional measures against each other across counties. In these cases, the correlations between the estimates may understate the true correlations between parameters because of noise introduced by the sampling error. To recover estimates of the correlation between the true parameters we calculate:

$$\hat{Corr}(\psi_j, \mu_j) = Corr(\hat{\psi}_j, \hat{\mu}_j)\sqrt{\frac{1}{\hat{r}_\psi}}\sqrt{\frac{1}{\hat{r}_\mu}}. \tag{9}$$

Where $Corr(\hat{\psi}_j, \hat{\mu}_j)$ is the correlation between estimates $\hat{\psi}_j$ and $\hat{\mu}_j$ (of parameters ψ_j and μ_j) across all counties j, and $\hat{r}_\psi$ are $\hat{r}_\mu$ are their reliability estimates from equation 8. We present these "signal correlations" in Appendix Table A6.

In Section 3.1 and 4.2, we use certain regional (and region-by-demographics) measures as right-hand side variables in our movers specifications. The sampling error in these estimates will attenuate their regression coefficients. To see this, take the simple regression $Y = \beta \cdot X + \omega$ where we observe $\hat{X}$, an estimate of X with independent sampling error ϵ. Then when estimating $Y = \hat{\beta} \cdot \hat{X} + \nu$ we have:

$$\begin{aligned}\hat{\beta} &= \frac{Cov(Y, \hat{X})}{Var(\hat{X})} \\ &= \frac{Cov(Y, X+\epsilon)}{Var(X+\epsilon)} \\ &= \frac{Cov(Y, X)}{Var(X) + Var(\epsilon)} < \frac{Cov(Y, X)}{Var(X)} = \beta.\end{aligned} \tag{10}$$

To account for this, in our movers analyses we first randomly split the individual-level data used to construct the relevant right-hand side measures in two halves. We then instrument for the value constructed by one half with the other. To see the intuition behind this procedure, let $\hat{X}_1$ and $\hat{X}_2$ be the split sample estimates. Then the first stage of a two-stage least squares estimate is given by $\hat{X}_1 = \phi_1 \cdot \hat{X}_2 + \nu_1$, where $\phi_1 = \hat{r} = \frac{Var(X)}{Var(X)+Var(\epsilon_2)}$. The reduced form is given by $Y = \phi_2 \cdot \hat{X}_2 + \nu_2$, where $\phi_2 = \frac{Cov(Y,X)}{Var(X)+Var(\epsilon_2)}$. Then the resulting estimate is:

$$\hat{\beta} = \frac{\phi_2}{\phi_1} = \phi_2 \cdot \frac{1}{\hat{r}} \approx \frac{Cov(Y, X)}{Var(X)} = \beta. \tag{11}$$

F Königsteiner Schlüssel and the Assignment of Refugees to Place

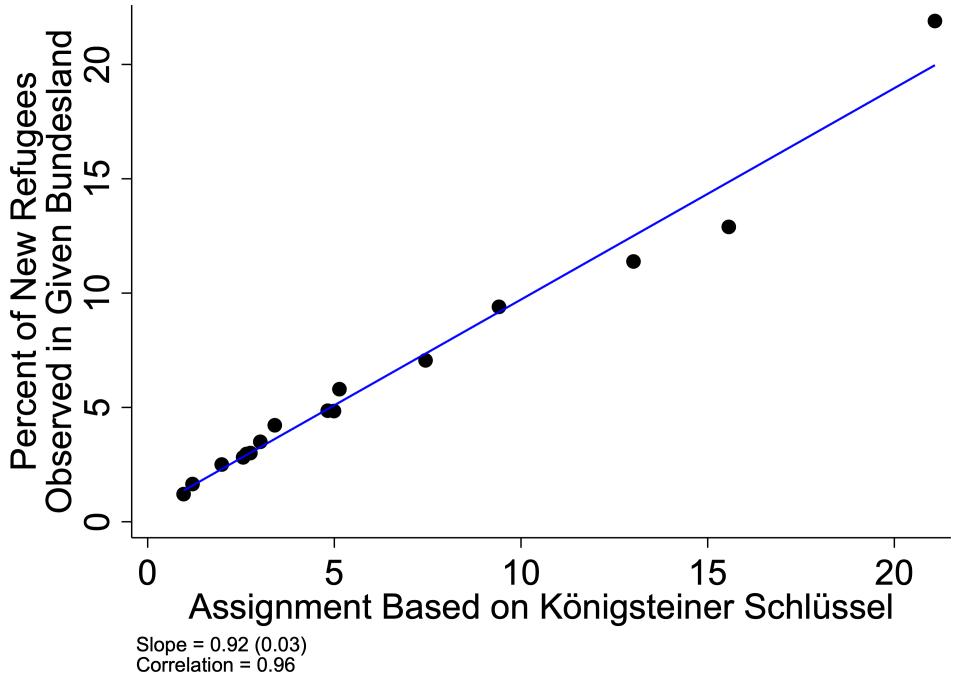
In this section, we attempt to compare the official refugee allocation rule—the so-called Königsteiner Schlüssel—to observed administrative data on refugee assignment.

The Königsteiner Key is an allocation rule which was designed in the 1940s to assign refugees to the sixteen different German states. It takes as input a state's population and tax income and weights these two factors with 1/3 and 2/3, respectively (Deutscher Bundestag, 2020). The key is updated annually, but given the slow-moving nature of its inputs, it is stable over time.

To infer to what extent the key has been abided to during the time period of interest for our study, we compare the 2019 assignment key (for data availability reasons) to the percentage of the total number of refugees that live in a given state and have been in Germany for less than 1 year, for each year from 2015 to 2019. The latter measure is intended to approximate for new-arrivals in the absence of direct data on this and the data for this approximate measure is obtained from the German Statistical Office.

Figure A13 shows the result of our comparison. The correlation of 0.96 and a slope of 0.92 indicates that the observed assignment lines up very closely with the official assignment rule. We find this reassuring, as it suggests that despite the large influx of migrants during these year, refugee assignment largely followed the official assignment key. In turn, this is suggestive that once controlling for the Königsteiner key, assignment to place is somewhat random.

Table A13: Comparison Königsteiner Key and Assignment of Refugees to Place



Note: Figure compares assignment of recent refugees to place with the official assignment key, i.e. the Königsteiner Schlüssel from 2019. The Königsteiner Schlüssel is compromised of a state's total population and a state's tax income where the former is weighted with one third and the latter is weighted two thirds. Assignment of recent refugees is approximated by the percentage of the total number of refugees that live in a given state and have been in Germany for less than 1 year, for each year from 2015 to 2019. The data comes from the German Statistical Office.

G Identifying Place Based Effects with Movers

To quantify the contribution of place-based effects to the spatial variation in migrants' integration outcomes, we propose a simple model in which the rate of friendships between migrants and a local natives is determined by the sum of place-based effects—which we allow to vary across time and with observable migrant characteristics—and other *unobservable* individual-level factors of the individuals involved. Since only place-based factors change around a move, this model allows us to estimate the share of regional variation in the social integration of migrants that can be attributed to place-based effects. We describe here the friending model and identifying assumptions in the context of the migrant mover design from Section 3.1. These features carry over to the native mover design in Section 4.2.

Friending model. We consider the following basic model of friending between migrants and locals which is similar to Finkelstein, Gentzkow and Williams (2016). We let each individual's friending outcome be the sum of their county's effect ($\text{PlaceEffect}^{(p)}$) and their personal individual effect (IndivEffect_i). Let $\text{AvgIndivEffect}^{(p)}$ be the average of IndivEffects for individuals in county p. Then the difference between the average outcomes, x, in two regions, (2) and (1), is the sum of differences between the place-based effect and the average of individual-effects.

$$x^{(2)} - x^{(1)} = (\text{PlaceEffect}^{(2)} - \text{PlaceEffect}^{(1)}) + (\text{AvgIndivEffect}^{(2)} - \text{AvgIndivEffect}^{(1)}). \tag{12}$$

We want to know the share of $x^{(2)} - x^{(1)}$ that is due to place-based effects, formally:

$$\frac{\text{PlaceEffect}^{(2)} - \text{PlaceEffect}^{(1)}}{(\text{PlaceEffect}^{(2)} - \text{PlaceEffect}^{(1)}) + (\text{AvgIndivEffect}^{(2)} - \text{AvgIndivEffect}^{(1)})}. \tag{13}$$

We cannot observe any of these parameters directly. At the individual level, however, we know that when a mover moves from (1) to (2), only the place-based factors should change. Her individual level effects are constant, so any change in friending outcomes must be driven by place based effects. So for mover i who moves from (1) to (2) at time t:

$$y^{\Delta}_{i,t} = (\text{PlaceEffect}^{(2)} - \text{PlaceEffect}^{(1)}). \tag{14}$$

Where $y^{\Delta}_{i,t}$ is the change in outcome before and after the move for mover i. Then α, below, is equivalent to equation 13, our outcome of interest.

$$y^{\Delta}_{i,t} = \alpha \cdot (x^{(2)} - x^{(1)}). \tag{15}$$

In addition to this baseline logic, we allow for separate place effects across certain observable demographics such as age and gender, as well as time. The AvgIndivEffect is then the average of the remaining unobservable individual effects. When estimating α we remove the variation in $y^{\Delta}_{i,t}$ explained by overall time trends (e.g., if throughout Germany Syrian migrants make more native friends over time) by adding quarter of move fixed effects, ζ_t.

Taking model to the data. We bring this model to the data by comparing the rate at which movers make friends in the year before and after their move to the difference in the average friending rates of otherwise similar non-movers in each location.[22] Focusing on migrant movers (rather than on native movers as in section 4.2), for each user i moving in quarter t, the outcome of interest is the change in the quarterly probability of making at least one local German friend, $y^{\Delta}_{i,t}$, defined as:

[22] In this analysis we limit to movers who were in their origin and destination counties for four or more consecutive quarters each, less stringent than the prior analysis which required six quarters in the destination. In addition, we only include observations for which there are at least five "matched" non-movers in both the origin and destination.

$$y_{i,t}^{\Delta} = 0.25 \left[\sum_{\tau=t}^{t+3} Y_{i,\tau} - \sum_{\tau=t-4}^{t-1} Y_{i,\tau} \right]. \tag{16}$$

Here, $Y_{i,t}$ is an indicator for whether Syrian migrant i makes at least one local German friend in quarter t. Similar to before, we residualize each side of the difference on regional measures of natives' Facebook usage. To compare $y_{i,t}^{\Delta}$ to differences in the average integration rates of observably similar non-movers in each place, we construct sets of users who match each mover on the important determinants of social integration in Section 2: gender, age group, and time spent in Germany. Formally, for user i moving in quarter t, we let $O(i,t)$ and $D(i,t)$ be the sets of similar non-movers in the origin at time $t-4$ and in the destination at time t, respectively. We then define the differences in their average outcomes, $x_{i,t}^{\Delta}$, as:

$$x_{i,t}^{\Delta} = 0.25 \left[\frac{1}{|D(i,t)|} \sum_{j \in D(i,t)} \sum_{\tau=t}^{t+3} Y_{j,\tau} - \frac{1}{|O(i,t)|} \sum_{j \in O(i,t)} \sum_{\tau=t-4}^{t-1} Y_{j,\tau} \right]. \tag{17}$$

The set cardinalities $|O(i,t)|$ and $|D(i,t)|$ are the number of non-movers in the matched comparison groups for each mover. Intuitively, $x_{i,t}^{\Delta}$ is the difference in the average quarterly probability of a non-mover migrant making a native local friend between the destination location in the year after the move and the origin location in the year before the move. Time-specific measures allow for changes in the differences between regions over time. Again, we residualize each side of the difference on regional measures of natives' Facebook usage. We then estimate:

$$y_{i,t}^{\Delta} = \alpha_0 + \alpha_1 x_{i,t}^{\Delta} + \xi_t + \epsilon_{i,t}, \tag{18}$$

where slope α_1 is our outcome of interest. An estimate of α_1 close to 1 would suggest that, within the first year of moving, migrant movers' friending behavior fully adjusts to the level of local non-movers' friending behavior. An α_1 close to 0 would suggest that migrants do not adjust their friending rates systematically toward the level of local non-movers. Because migrant observables do not differ significantly across space, under the relatively weak identification assumptions discussed below, α_1 estimates the share of the observed differences in the social integration of migrants across locations that are due to causal place-based effects rather than unobservable individual characteristics. The quarter of move fixed effect, ξ_T, remove variation in overall time trends in the rates of befriending local natives.

One challenge with our estimation is that we only observe a sample estimate of each mover's $x_{i,t}^{\Delta}$, denoted by $\hat{x}_{i,t}^{\Delta}$. Measurement error in the true differences in friending probabilities of non-movers across locations would thus lead to attenuation bias in α_1. To account for this sampling error, when estimating equation 18, we randomly split the individual-level data of the friending behavior of non-movers used to construct $\hat{x}_{i,t}^{\Delta}$ into two sub-samples and instrument for the value constructed in one sub-sample with the value constructed in the other sub-sample (see Appendix E for details).

Identification Assumptions. Our interpretation of α_1 relies on the identifying assumption that place-based effects are additive and additively separable from any unobservable individual-level factors. This additivity allows us to aggregate the level of within-migrant differences across migrants to identify α. It implies, for example, that a move from place A to place B should have the same effect as a move from place B to place A. This is supported by Figure 3, as well as the results in Figure A15 and Table A8. Additive separability also implies that migrants' friending rates between locations will vary by the same *absolute amount* across unobservables. (The model does, however, allow for non-additive relationships between our key observables—gender, age, and time in Germany—and migrants' friending rates). Our identification also relies on there being no systematic shocks to unobservable factors that coincide exactly with the move quarter and affect native friending differentially by origin and destination.

These identifying assumptions are relatively weak and allow for movers to differ from non-movers

on observable and unobservable characteristics, and for these differences to correlate with origin and destination characteristics. For example, our model allows for "better integrating migrants" to be more likely to move to "better places." Intuitively, this is because our estimates come from *within-migrant* differences in integration over time, and "better" integrating migrants will make more friends both before and after the move. This differs from designs used in papers such as Chetty and Hendren (2018*a*) and Chetty and Hendren (2018*b*). These papers, which rely on cross-sectional outcomes, use within-family designs to rule out selection effects. Our data allow us to measure the outcome in the panel context (as in Finkelstein, Gentzkow and Williams, 2016), mitigating these concerns.

Our research design allows the level of movers' pre-move friending within an origin county to correlate with destination friending levels due to differences in individual characteristics. Movers' native friending around a move can also differ from the trends of non-movers. This could occur if, as suggested by Figure 3, all movers make fewer local connections in anticipation of a move or more connections immediately after a move. Each of these would increase α_0, but leave α_1 unaffected. Our model would be affected if these downward trends in movers' propensity to make friends before relocating differed systematically by the integration levels in the movers' destinations.[23] Figure 3 provides evidence that such differential trends do not exist. As an additional test, in Figure A13, we decompose our results from Figure 3 into friendships initiated by the mover and those initiated by the Germans in their destination. We find that, following a move, both migrant-initiated and *native-initiated friendships* change in the predicted direction. This provides more evidence that our results are not driven by changes in migrant friending preferences around the time of the move that correlate with the characteristics of the destination.

Results for Migrant Movers. Figure A15 displays a binned scatter plot of $y^{\Delta}_{i,t}$ against $x^{\Delta}_{i,t}$, with the slope corresponding to α_1 in equation 18.[24] The relationship is symmetric around zero and linear, consistent with additive effects of place. The fact that the scatter plot is horizontally centered around zero also suggests that, conditional on demographics, migrants do not systematically move to places with higher or lower levels of integration. The slope estimate is 0.738: nearly three quarters of the observed regional variation in Syrian migrants' friendship formation with local natives is directly attributable to place-based effects that occur within the first year of after their move, rather than individual characteristics. In Appendix Figure A12 we plot the slope estimates separately for samples of users that are male, female, younger than 30 years old, 30 to 39 years old, and over 40 years old. For each group, the estimates are similar, suggesting our results are not driven by any particular demographic group of Syrian migrants.

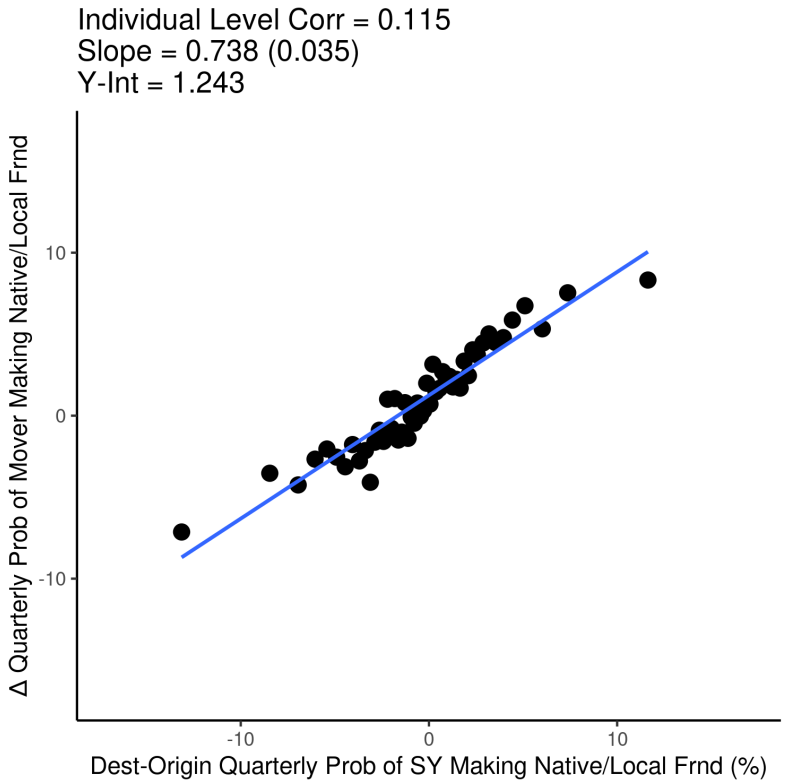
While this section focuses on measures of social integration based on migrants' friending patterns, Appendix H explores our language-based measure of integration. Whereas our prior analysis could use panel data on quarterly friending rates, our language outcome—whether the user produces content in German—is only observable at high quality in the cross section. We thus study how a mover's language use *today* is shaped by the set of places they have lived, following similar analyses in Chetty and Hendren (2018*a*) and Finkelstein, Gentzkow and Williams (2019). Our results suggest that place-based effects drive much of the cross-sectional variation in Syrian migrants' German language usage.

The prior results have documented that when Syrian migrants move between German counties, their social integration patterns quickly adjust from those of their origin towards those of their destination county. Our results thus show that most of the observed regional differences in social integration are explained by the effect of places—either due to institutional factors associated with the location, or due to local native characteristics—rather than by the characteristics of the migrants. In this context, it

[23] Put differently, our model allows for migrants' individual characteristics to change around a move so long as they do not differ systematically by destination location. For example, our estimates of α_1 would be biased upward if movers to better places became differentially less sociable before a move.

[24] Appendix Table A7 summarizes the sample of movers and the corresponding matched sample of otherwise similar non-movers in the origin location.

Figure A15: Δ Syrian Migrant Mover Friending Integration vs. Matched Non-Movers



Note: Figure shows a binned scatter plot describing the change in the friending of Syrian migrants to German natives before and after a move within Germany. The population is Syrian migrant users who moved between two non-neighboring counties and were in the first and second county for 4+ consecutive quarters each. The y-axis displays $y^{\Delta}_{i,t}$, movers' change in the quarterly probability of making a native local friend the year before to after the move. The x-axis displays $\hat{x}^{\Delta}_{i,t}$, the difference in average outcomes for comparable non-movers at the same time. We match each mover to a set of non-movers who lived in the origin location a year before the move and to a set who lived in the destination location at the move. In addition we also match movers to non-movers of the same gender and age bucket (18-29, 30-39, 40+), and whom we first observed on Facebook in Germany in the same year. We include observations for which there is at least 5 non-movers in both the origin and destination match group. We control for quarter of move fixed effects. We correct for sampling error in the x-axis measures by randomly splitting the individual-level non-mover data into two halves and instrumenting for one set of averages with the other. See Appendix E for more information this procedure. Standard errors are shown in parentheses. Appendix Table A8 presents formal regression results on the relationships in this figure.

is important to note that a mover design will not even capture the full extent to which individual integration is shaped by place-based effects. For example, Syrian migrants who learn the German language in high-integration places (possibly in local integration courses) might then use these skills to make German friends more quickly after moving to a low-integration place. This effect might be considered "place-based" in the sense that it is shaped by features of the mover's origin location, but will not be captured by our estimates. To the extent that such additional long-term place-based effects are important, our estimates of α_1 will even *understate* the extent to which places truly shape migration outcomes.

H Cross-Sectional Analysis of Movers and German Language Usage

We assess the degree to which selection drives our regional estimates of German language integration using a cross-sectional movers design. This follows similar designs in Chetty and Hendren (2018*a*) and Finkelstein, Gentzkow and Williams (2019), and differs from the design used in Sections 3.1 and 4.2 which utilize panel data on movers' friending. In particular, we model German language usage as a linear combination of the outcomes of non-movers in each of the mover's locations. Then, using the same mover criteria as in Figure A15, we estimate:

$$y_i = \alpha_0 + \alpha_1 \sum_p q(i,p) * x_{p,d(i)} + \kappa_{d(i)} + \epsilon_i \tag{19}$$

Here, y_i is an indicator for whether individual i produces German content on Facebook and $q(i,p)$ is the share of their quarters in Germany spent in place p. The notation $d(i)$ represents a set of demographics used to match movers to similarly situated non-movers. $x_{p,d}$ is the share of users in place p and demographic group d that produces German content, and $\kappa_{d(i)}$ are demographic group fixed effects, which remove variation driven by the demographic matching from our slope estimates. In our strictest specifications, we also add fixed effects for users' first and current county in Germany.

Table A14: Syrian Migrant Mover Language Integration vs Weighted Average of Places

	Produces Content in German (0/100)				
Predicted Prob. Of Using German (Weighted Avg. of Places Lived)	0.863***	0.857***	0.863***	0.813***	0.816***
	(0.037)	(0.043)	(0.058)	(0.042)	(0.058)
FEs	Cohort	Cohort	Cohort	Cohort X Curr. Cnty.	Cohort X Curr. Cnty. X First Cnty.
Sample		< 75% in Max County	< 60% in Max County		
N	23,249	18,233	10,172	23,069	14,474
Sample Mean	38.075	37.959	38.252	38.099	36.977

Note: Table shows results for comparisons between the German language usage of Syrian migrants who moved between counties and their predicted language usage based on the outcomes of non-movers in the places they lived. For each location, movers are matched non-movers by age, gender, and the first year they used Facebook in Germany (cohort). Column 1 shows our baseline specification from equation 19, which includes cohort fixed effects. Column 2 limits to only users who spent < 75% of their quarters in Germany in one county. Column 3 limits to those who spent < 60%. Column 4 repeats column 1 with cohort-by-current county fixed effects; column 5 repeats column 1 with cohort-by-current county-by-first county in Germany fixed effects. We correct for sampling error in the right-hand side measures by randomly splitting the individual-level non-mover data into two halves and instrumenting for one set of averages with the other. See Appendix E for more information this procedure. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

In contrast to equation 18, our unit of observation is a mover, not a move, and we use movers' location for every quarter they have been in Germany. As in our panel analyses, we cannot observe $x_{p,c(i)}$, but instead account for sampling error by constructing estimates $\hat{x}_{p,c(i)}$ from random halves of the data and instrumenting for one with the other. We also again relax the assumption of fully additive-seperability between individual-level factors and place-based effects by matching movers to similarly situated non-movers on gender, age group, and year of arrival in Germany. This allows for non-additive interactions with these demographics. We enforce that each mover must have 20 matched non-movers.[25]

[25] This threshold is higher than the five user minimum in Section 3.1. Our sample in this analysis, however, will remain larger because we (mechanically) do not enforce temporal matching.

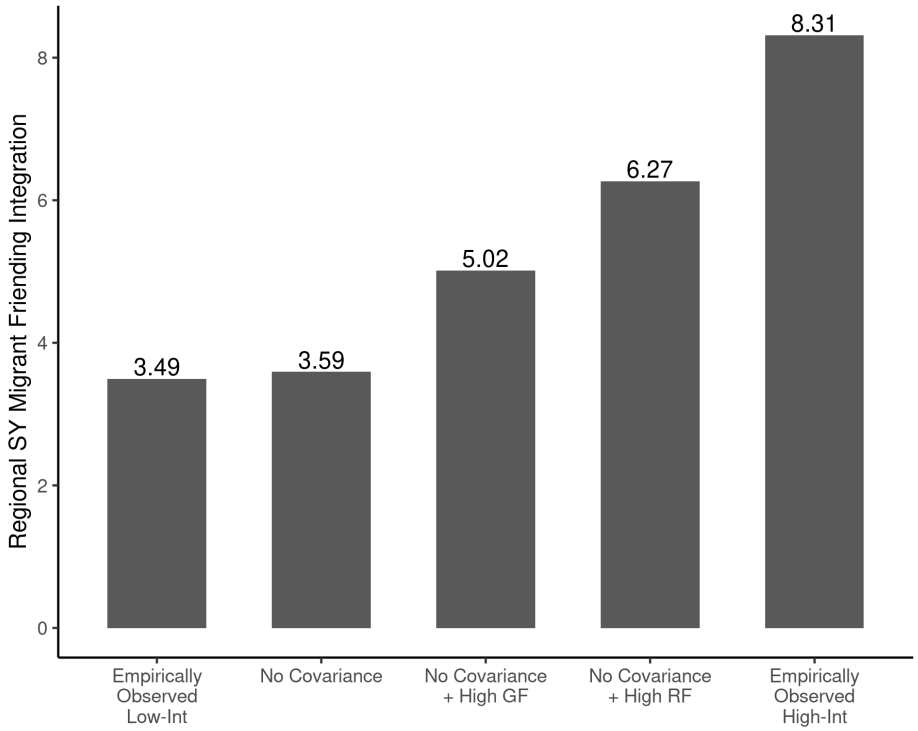
Table A14 presents results from our analysis. In column 1, an estimate of α_1 close to 1 would suggest that a Syrian migrant's likelihood of using German on Facebook is close to the averages of migrants in each location they have lived, weighted by the amount of time they lived in each location. The resulting slope estimate of 0.86 shows that this is the case. While this evidence is consistent with places having an *effect* on migrants' German language integration, it does not rule out alternative explanations. For example, it is possible that our sample includes many users who have spent a long time in a single location, and that the right hand side weighted averages are often dominated by a single region. If this were the case, our estimates could be largely driven by movers behaving similarly to local non-movers in general, rather than by place-based effects in particular. Columns 2 and 3 provide evidence that this story does not drive our overall results, as our estimates of α_1 remain similar when limiting our sample to users who spent <75% or <60% of their time in Germany in one county, respectively.

In column 4 we take another approach to testing whether our results are indicative of causal effects of place. In particular, we control for each user's current county, thereby identifying our slope estimates from variation in the user's origin counties. The slope estimate decrease slightly, but remains around 0.81. This suggests that much of the variation in language outcomes amongst movers across regions today is determined by where they *originally* lived in Germany, providing evidence against selection effects. In the final column, we control for both first county and final county fixed effects. Our identification, therefore, comes from the amount of *time* users' spend in each particular place. The slope estimates remains at 0.82, providing more evidence that a migrant's probability of using the German language scales linearly in proportion to the time they spend in high- and low-integration places.

I Decomposition of High- vs Low-Integration Regional Differences

In Figure A16, we conduct counterfactual exercises to explore the degree to which each of our two components explain the differences between counties with high- and low-friending integration. This follows a similar exercise in Chetty et al. (2022*b*). The first and fifth bars show the average integration of migrants in top and bottom quintile counties, respectively. Syrian migrants in top quintile counties make 8.31 native local friends on average, versus 3.49 in bottom quintile counties. In the second bar we multiply the bottom quintile averages of general friendliness and relative friending, thereby removing any within-quintile covariance. Doing so somewhat increases the value from the first bar, consistent with the small negative correlation between the two components in Table 3. The third and fourth bars replace the bottom-quintile averages of general friendliness and relative friending with the corresponding top-quintile averages, respectively. We view this as a counterfactual in which we hold one of the two integration components of low-integration regions fixed and adjust the other to the levels of high-integration regions. We interpret the difference between the second and fourth bars (2.68), compared to the second and third bars (1.43), as relative friending explaining about 1.9x as much of the difference between high and low-integration places as general friendliness.

Figure A16: Decomposition of Difference Between High- and Low-Integration Regions



Note: Figure shows how much of the difference between high and low friending integration counties is driven by general friendliness versus relative friending. The first and fifth bars show the average friending integration of Syrian migrants in top and bottom quintile counties, respectively. The second bar replaces each county observation from the first bar with the bottom quintile averages of general friendliness and relative friending. The third and fourth bars replace the bottom-quintile averages of general friendliness and relative friending with the corresponding top-quintile averages, respectively.

J Individual-level Correlates of Natives Behavior Towards Migrants

This appendix explores the relationship between observable native characteristics and behaviors toward Syrian migrants. In particular we focus on their (i) friending of local Syrian migrants; (ii) general friendliness; (iii) relative friending; and (iv) joining of pro-immigration organizations on Facebook.

Table A15: Natives - Measures of Friending

	N Local SY Friends		General Friendliness		Relative Friending		In Pro Imm. Group (0/100)	
Age 25 - 34	-0.073***	-0.073***	-19.097***	-14.407***	-0.059***	-0.061***	0.359***	0.146***
	(0.000)	(0.000)	(0.098)	(0.092)	(0.001)	(0.001)	(0.018)	(0.018)
Age 35 - 44	-0.116***	-0.114***	-55.586***	-52.328***	-0.081***	-0.080***	0.951***	0.858***
	(0.000)	(0.000)	(0.103)	(0.097)	(0.001)	(0.001)	(0.018)	(0.018)
Age 45 - 54	-0.132***	-0.131***	-62.533***	-62.415***	-0.098***	-0.095***	1.116***	1.152***
	(0.000)	(0.000)	(0.108)	(0.102)	(0.001)	(0.001)	(0.019)	(0.019)
Age 55+	-0.139***	-0.141***	-82.666***	-84.728***	-0.098***	-0.095***	2.105***	2.157***
	(0.000)	(0.000)	(0.108)	(0.102)	(0.001)	(0.001)	(0.020)	(0.020)
Female	-0.015***	-0.015***	-19.519***	-18.725***	-0.008***	-0.009***	0.882***	0.843***
	(0.000)	(0.000)	(0.056)	(0.053)	(0.001)	(0.001)	(0.010)	(0.010)
Has College	0.006***	0.006***	4.131***	7.619***	-0.000	-0.002***	1.931***	1.788***
	(0.000)	(0.000)	(0.060)	(0.056)	(0.001)	(0.001)	(0.011)	(0.011)
Prev Quarters in NUTS3 FEs	X	X	X	X	X	X	X	X
Personal Usage Controls	X	X	X	X	X	X	X	X
County FEs		X		X		X		X
N	17,768,822	17,768,822	17,768,822	17,768,822	17,515,164	17,515,164	17,768,141	17,768,141
R-Squared	0.020	0.031	0.170	0.263	0.001	0.002	0.035	0.042
Sample Mean	0.086	0.086	122.510	122.510	0.074	0.074	4.835	4.835

Note: Table shows results from regressing various outcomes on the demographics of users in the German native Facebook sample. The outcome is their number of local friends in the Syrian migrant sample in columns 1 and 2; their number of local friends in the German native sample in columns 3 and 4; their relative friending to Syrians and Germans defined by equation 2 in columns 5 and 6; and the number of groups registered with *ProAsyl* they are in in columns 7 and 8. Columns 1, 3, 5, and 7 include controls for age, gender, and whether they list a college on Facebook, as well as fixed effects the number of quarters on Facebook in their current county. They also include linear controls for log(0.5 + minutes on FB in the last 28 days), log(91 - days on Facebook out of the last 90), log(1081 - days on Facebook out of the last 1080). Columns 2, 4, 6, and 8 add county fixed effects. In columns 7 and 8 the personal usage controls also include fixed effects for each number of Facebook groups a user is in. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

Equation 5 is our multivariate regression of interest. Each observation is a German native user. In all specifications we include controls for the amount of time each user spends on Facebook and for the number of quarters they have been on Facebook in their current county. In certain specifications we also include county fixed effects. Y_i represents measures of the four outcomes listed above. Friending of local Syrian migrants is measured by the user's number of local Syrian migrant friends. Individual-level general friendliness is measured by the user's number of local native friends. We construct individual-level relative friending by replacing each term in the numerator of equation 2—$NLocalFriends_c^{DE \to SY}$ and $NLocalFriends_c^{DE \to DE}$—with its individual-level analog.[26] We identify pro-immigration Facebook pages and groups using a combination of string, url, and manual matching. Our outcome measure is whether a user "likes" one of these page or is in one of these groups. In total, we identify 8,171 groups and pages, and measure 2.1 million user-page or user-group connections.

Table A15 presents results. Columns 1 and 2 show that younger natives and male natives are more

[26] A user must have at least one local native friend for this individual-level measure. The county-level average of this measure will equal the county-level measure in equation 2 if each observation in the former is weighted by the user's number of local native friends.

likely to befriend migrants than older and female natives, respectively. Columns 3 and 4 show that these patterns are driven in part by general friendliness: a native being younger, male, or college educated is associated with having a larger network of local native friends. Columns 5 and 6 show that our individual-level measure of relative friending is also higher for younger and male German natives, while it is somewhat lower for college educated Germans compared to college educated Germans. Because Syrian migrants in Germany are more likely to be young and male than the average German native (see Table 1), one possible explanation for this finding is that homophily plays a strong role in shaping which natives befriend Syrian migrants. For example, younger German natives might be more likely to connect with younger Syrian migrants because younger people in general are more likely to connect, rather than because of particular behaviors toward migrants.

Columns 7 and 8 show that older, female, and college-educated natives are more likely than others to join pro-immigration groups on Facebook, conditional on Facebook usage. (For these analyses we include fixed effects for each number of total Facebook groups as user is in, holding constant a user's overall propensity to join Facebook groups. Our results remain qualitatively unchanged without this control.) These are *opposite* the relationships presented for relative friending in columns 5 and 6, suggesting that is not necessarily those who are most supportive of pro-immigration groups that are most likely to disproportionately befriend Syrian migrants. This is again consistent with a story in which homophily, above specific attitudes or behaviors toward migrants, contribute to the demographic differences we observe in prior columns.

K Data Description of County-Level Covariates

Variable	Description	Data Source	Link to Data
Average Age	Average age of populaton, 2014	German Statistical Office	https://www-genesis.destatis.de/genesis/online?operation=sprachwechsel&language=en
% Female Age	Share of population that is female, 2014	German Statistical Office	https://www-genesis.destatis.de/genesis/online?operation=sprachwechsel&language=en
Pop. Density 2018	Population density, 2018.	Regionalatlas Deutschland	https://www-genesis.destatis.de/gis/genView?GenMLURL=https://www-genesis.destatis.de/regatlas/AI002-1.xml&CONTEXT=REGATLAS01
% Empty Flats	Share of flats that that are vacant, 2017	Thünen-Landatlas	https://karten.landatlas.de/app/landatlas/
Average Income	Average income, 2018	Statistische Ämter des Bundes und der Länder *(Federal and state statistical offices)*	https://www.statistikportal.de/de/vgrdl/ergebnisse-kreisebene
% Unemployed	Unemployment rate, 2014	Bundesagentur für Arbeit *(Federal Employment Agency)*	https://statistik.arbeitsagentur.de/SiteGlobals/Forms/Suche/Einzelheftsuche_Formular.html?topic_f=beschaeftigung-eu-heft-eu-heft
Train. Positions per Applicant	Number of training positions (Lehrstellen) per applicant (Auszubildender)	Bundesagentur für Arbeit *(Federal Employment Agency)*	https://statistik.arbeitsagentur.de/SiteGlobals/Forms/Suche/Einzelheftsuche_Formular.html?gtp=15084_list%253D4&topic_f=analyse
Syrians Employed / in Train.	Number of Syrians employed or in training divided by Syrian population	Bundesagentur für Arbeit *(Federal Employment Agency)*	https://statistik.arbeitsagentur.de/SiteGlobals/Forms/Suche/Einzelheftsuche_Formular.html?topic_f=beschaeftigung-eu-heft-eu-heft
All Crimes 2014	Reported crimes (total) per population, 2014	Polizeiliche Kriminalstatistik *(Police Crime Statistics)*	https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/pks_node.html
Thefts 2014	Theft crimes per population, 2014	Polizeiliche Kriminalstatistik *(Police Crime Statistics)*	https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/pks_node.html
Violent crimes 2014	Violent crimes per population, 2014	Polizeiliche Kriminalstatistik *(Police Crime Statistics)*	https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/pks_node.html
% Christian	Number of Christians per population, 2011	Zensus Datenbank *(Census Results)*	https://ergebnisse2011.zensus2022.de/datenbank/online/
% AfD 2014	Vote share Alternative für Deutschland (AfD), European elections, 2014, demeaned by state	Der Bundeswahlleiter *(Federal Returning Officer)*	https://www.bundeswahlleiter.de/europawahlen/2014/ergebnisse.html
% Voted 2014	Log turnout, European elections, 2014	Der Bundeswahlleiter *(Federal Returning Officer)*	https://www.bundeswahlleiter.de/europawahlen/2014/ergebnisse.html

% Syrians 2010	Number of Syrians divided by population, 2010	German Statistical Office	https://www-genesis.destatis.de/genesis/online?operation=sprachwechsel&language=en
% Syrians 2019	Number of Syrians divided by population, 2019	German Statistical Office	https://www-genesis.destatis.de/genesis/online?operation=sprachwechsel&language=en
% Foreign 2010	Number of foreigners divided by population, 2010	German Statistical Office	https://www-genesis.destatis.de/genesis/online?operation=sprachwechsel&language=en
% Foreign 2019	Number of foreigners divided by population, 2019	German Statistical Office	https://www-genesis.destatis.de/genesis/online?operation=sprachwechsel&language=en
Integr. Courses per Syrian	Number of integration courses completed 2015-2019 per Syrian	Federal Office for Migration and Refugees	https://www.bamf.de/DE/Themen/Statistik/Integrationskurszahlen/integrationskurszahlen-node.html
Pro-Immigr. Groups per Syrian	Number of groups affiliated with ProAsyl activist group per Syrian	ProAsyl	not publicly available, data received directly from organisation
Integr. Sports Clubs per Syrian	Number of sports clubs that are part of Integration through Sport initiative	German Olympic Sports Confederation	https://integration.dosb.de
Unemp. General Schools Teachers per Pop. 2014	Number of unemployed general school teachers divided by population, 2014	Bundesagentur für Arbeit *(Federal Employment Agency)*	not publicly available, data received directly from organisation
Unemp. Higher Ed. School Teachers per Pop. 2014	Number of unemployed university and research institute teachers divided by population, 2014	Bundesagentur für Arbeit *(Federal Employment Agency)*	not publicly available, data received directly from organisation
Unemp. Driving and Sports Teachers per Pop. 2014	Number of driving and sports teachers divided by population, 2014	Bundesagentur für Arbeit *(Federal Employment Agency)*	not publicly available, data received directly from organisation
Unemp. Other School Teachers per Pop. 2014	Number of teachers in other education centers divided by population, 2014	Bundesagentur für Arbeit *(Federal Employment Agency)*	not publicly available, data received directly from organisation

L High School Matching Procedure

We assign users to high schools using a three-step process. On Facebook, users can provide the high school that they attended in their profile. Some of these high schools (such as "Hogwarts" and "the School of Hard Knocks") are obviously incorrect, so we begin by filtering out such schools. We are left with a list of plausible high school names, which we then need to disambiguate, since many high schools share the same name. For this, we use a listing of high schools from the websites of German state governments (see DatenSchule Project.) For each user in our sample, we are able to observe the counties in which they lived during high school age. We use this information and their self-reported high school name to match them to a high school in the administrative data. To do this, we make use of a fuzzy string matching algorithm, applied to the list of high schools that are in the regions in which they lived between the ages of 13 and 18.[27] Using this methodology, we are able to match 1.2 million of the 2.2 million users to high schools from the administrative data.

In the second step, we consider the users who report a high school that we are unable to find in the administrative data. In some cases, simple misspellings or inconsistencies in the school's name prevent a match from being formed between the two data sets. In other cases, these discrepancies are due to variations in states' criteria for including schools in the lists provided on their websites (e.g., states differ in their inclusion of vocational high schools in the lists we use). For this reason, we create a listing of school names that are reported by 50 or more users in a single county, but which are not included in the administrative data. We allow users to be assigned to these well-attested schools as we would any other. We call these schools the "non-canonical schools", and include them in all regressions, though our results are robust to excluding them. This process adds another 81 thousand users to our sample. For users who attend a school which we cannot find in the administrative data, and which appears in the self-reported data fewer than 50 times in the same county, we discard their self-reported school.

Finally, for users without a validated self-reported high school, we attempt to impute the school they attended using information on their social network. Intuitively, this approach takes advantage of the fact that most users will attend the same school as their friends who live in the same area and are the same age. To do this, we find the modal high school among a user's friends in the county they live in (as well as counties bordering it) and who are no more than 3 years different in age from the user. If this modal high school is attended by at least 10 friends, and there are at least 5 times as many friends attending this high school as the next most common school, we assign the user to this high school. We repeat this process 10 times, adding 137 thousand more users to our sample.[28]

We are able to assign 63% of native users to high schools using this methodology. In the cohorts we use for our regression, the median cohort has 31 students, with an inter-quartile range of 15 to 52 students. The match rate is lower (24%) for Syrian migrant students, since they have relatively few local friends and are less likely to list a high school on their profile. Any mistakes we make in assigning Syrians to high schools are likely to bias our analyses away from finding an effect of exposure.

[27] If we are unable to find a high school that matches in one of the regions that they lived in, we consider the regions that neighbor the regions the user lived in.

[28] To get a sense for the predictive power of the above imputation methodology, we can examine how accurate it is in determining the high school attended by users who self-report the school they attended. The imputation method is able to assign a school to 25% of such users, agreeing with the self-reported school in more than 90% of cases.