

# Segregation and Sentiment:

## Estimating Refugee Segregation and Its Effects Using Digital Trace Data

Neal Marquez - Sociology

University of Washington

nmarquez@uw.edu



CENTER FOR STUDIES IN  
DEMOGRAPHY & ECOLOGY

### Abstract

In light of the ongoing events of the Syrian Civil War, many governments have shifted the focus of their hospitality efforts from providing temporary shelter to sustaining a long-term population. In Turkey, a heightened focus has been placed on the encouragement of integration of Syrian refugees into Turkish culture, through the dismantling of Syrian refugee-only schools in Turkey and attempts to grant refugees permanent citizenship, among other strategies. Most of the existing literature on the integration and assimilation of Syrian refugees in Turkey has taken the form of surveys assessing the degree to which Syrian refugees feel they are part of Turkish culture and the way Turkish natives view the refugee population. Our analysis leverages call detail record data, made available by the Data 4 Refugees Challenge, to assess how communication and segregation vary between Turkish natives and Syrian refugees over time and space. In addition, we test how communication and segregation vary with measures of hostility from Turkish natives using data from the social media platform Twitter. We find that measures of segregation vary significantly over time and space. We also find that measures of inter group communication positively correlate with measures of public sentiment towards refugees. Attempts to address the concerns of Turkish natives to minimize the traction of online hate movements may help to improve the integration process.

### Main Objectives

1. Build quantitative measures of segregation for Turkey's major metropolitan areas.
2. Compare segregation measures across Turkey.
3. Track changes in online conversation of Turkish natives about Syrian refugees.
4. Compare how changes in sentiment compare to changes in social

### Background and Methods

Segregation has long been seen as a mechanism that isolates individuals from accessing greater opportunities if their isolated enclave is poor in group resources. In addition, greater isolation of communities has been linked to increased xenophobic attitudes toward minority migrant groups in the global south. Previous policy research has advocated for working towards greater cohesion between groups in the form of public education campaigns as a way of combating negative opinions towards these minority groups. The extent to which segregation between native populations and refugees is an issue in Turkey is not yet well understood.

The analysis utilizes call detail records from the Turkish mobile network carrier Turk Telecom (TT), a member of the group TTG, as part of the Data for Refugees in Turkey (D4R) challenge. Data came in the form of call detail records containing the location of a cell phone tower, time of call, whether a caller was marked as a refugee, and the refugee status of the other person on the call. We estimated residential and activity space dissimilarity for a district by creating subunits within each district by way of Voronoi tessellation from the cell phone towers within the district. While traditional residential dissimilarity indexes measure differences from the perspective that individuals are situated in a single location, activity space dissimilarity measures the probability of remaining isolated from another group through daily activity or 'potential to encounter' as defined in previous research. Activity space dissimilarity scores were calculated for each district for each week of the analysis while residential dissimilarity scores were created once per district as scores did not change much over time.

To measure sentiment towards refugees we used data from the social media platform Twitter that originated in Turkey and referenced refugees in the context of their tweet. Tweets were analyzed using a Turkish translated version of the AFINN, a common sentiment analysis tool with words valence rated on a scale from -5 to 5. Each tweet is rated by the sum of individual word scores. To test the relationship between Twitter sentiment and intergroup connectivity, we run a series of logistic regressions where each outgoing call made by a refugee is the response variable. The outcome is 0 if the call/text was made to a fellow refugee or 1 if made to a non-refugee, with a total of 10,235,988 records. Call records were connected with covariates by their district of call location (for population size, urban area coverage), the biweekly time period that they occurred (for Twitter sentiment), or the combination of the two (for activity space dissimilarity index). We tested a number of covariate combinations to test the robustness of the relationships between covariates and the outcome. To account for the bias in the data from repeated calls from a single user, we ran a mixed effects model with a random intercept on individual.

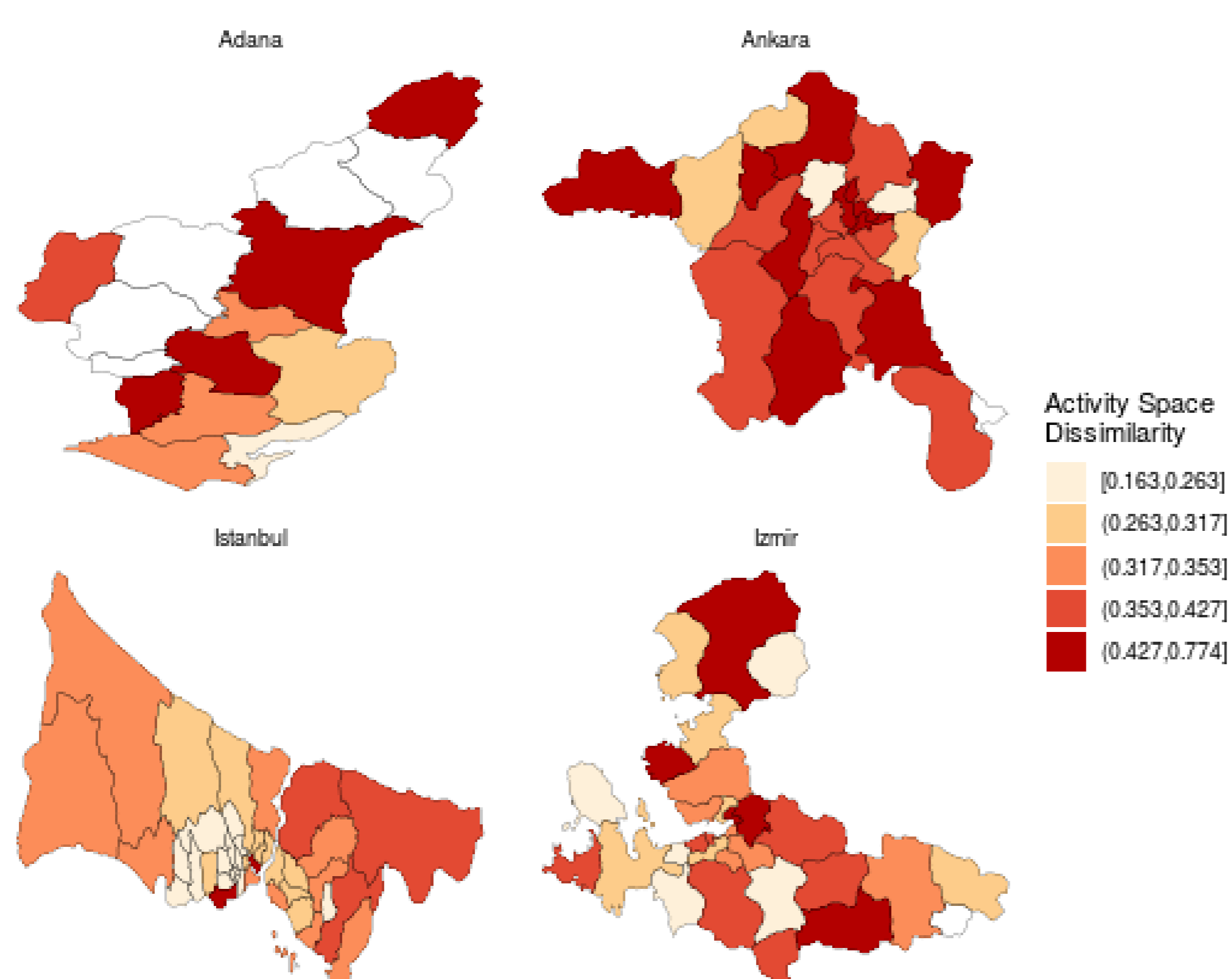


Figure 1: Activity Space Dissimilarity Scores for selected provinces. Results with observed dissimilarity less than 4 standard deviations away from mean are whitened out.

### Results

Urban land coverage of a district was inversely correlated with dissimilarity although the effect was non-significant when running a simple linear model. Using bootstrapped estimates of the uncertainty of our calculations for activity space dissimilarity, we found that there were significant differences over time and space at both the district and province level. We also found that residential dissimilarity was strongly correlated with activity space dissimilarity with a correlation coefficient of 79.96 at the district level and 83.83 at the

province level. In line with previous literature, we found that activity space dissimilarity was more often less than residential dissimilarity. Twitter sentiment was also found to change significantly over time but not over locations. Because our province level analysis required that users tweets be geocoded at least to the provincial level, our sample size was dramatically reduced when examining geographic differences in tweets.

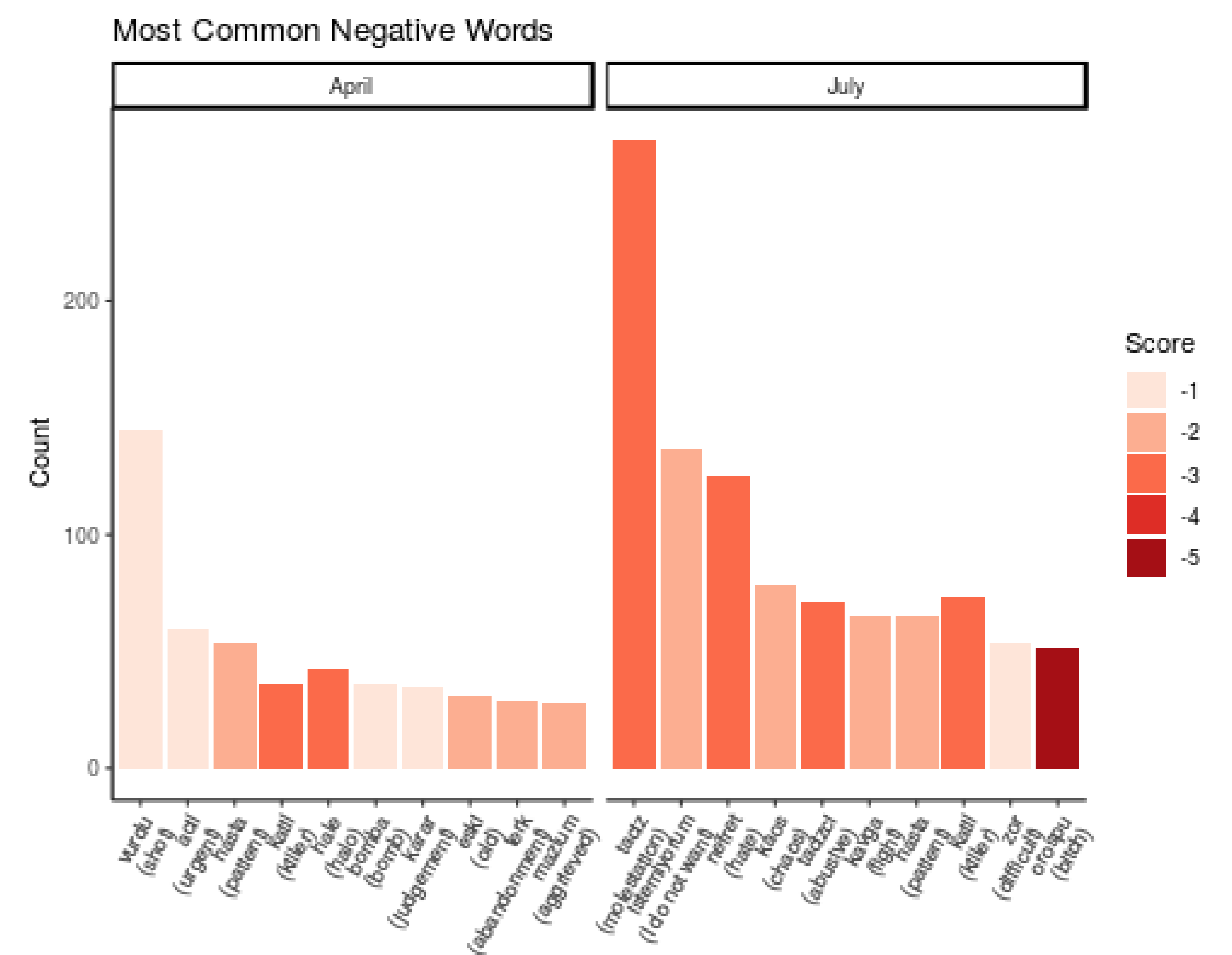


Figure 2: Comparison of most common negative words in our data set of tweets about refugees for selected months.

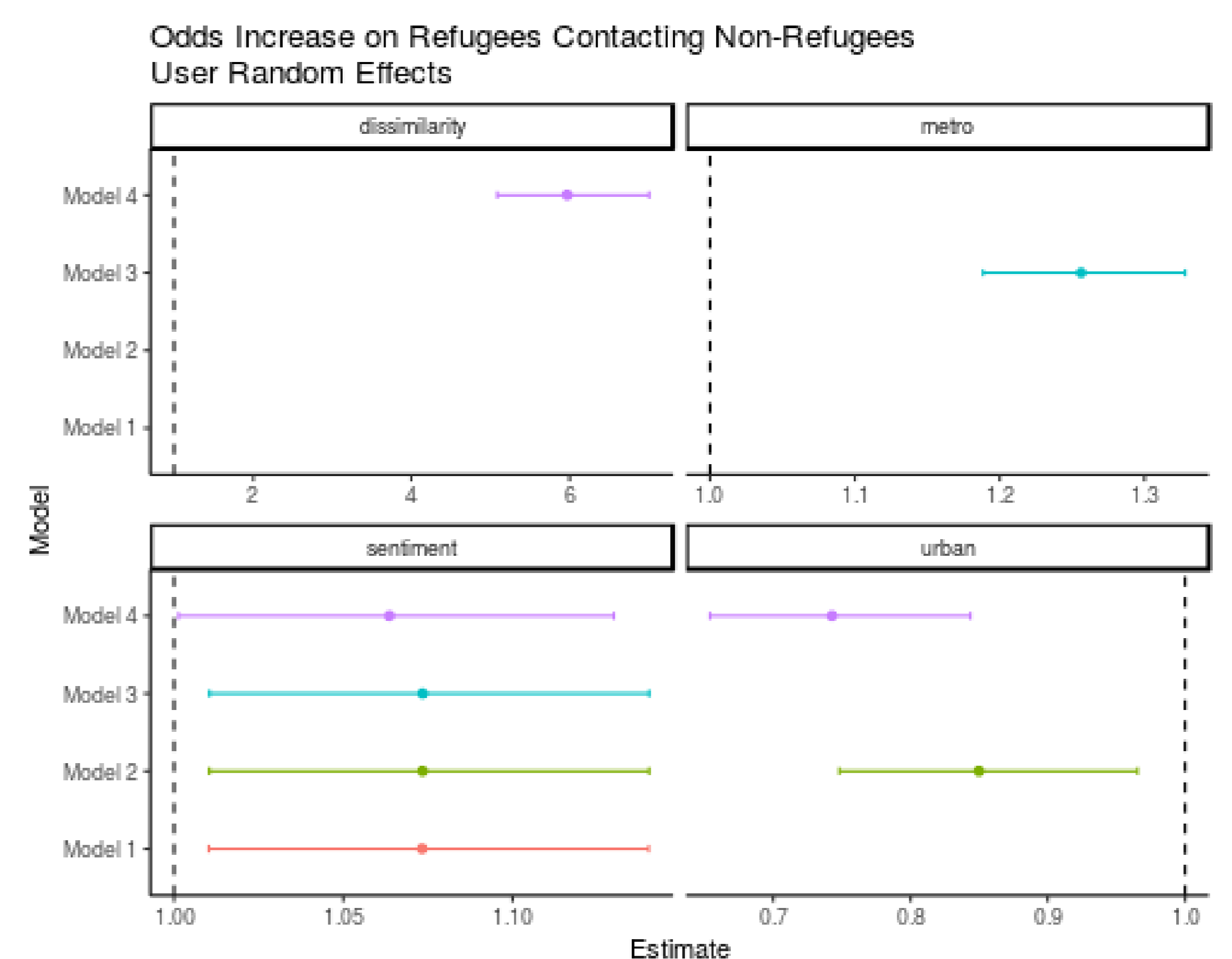


Figure 3: Model Odds Ratios Coefficient Estimates for Select Covariates. Error bars not overlapping with dotted line indicate significant result. Four models are presented in the figure on the y axis and coefficients are placed in separate panels.

Models for predicting calls and texts from refugees to non-refugees showed a significant positive relationship between sentiment and connectivity. As weekly Twitter sentiment scores increased, i.e. more positive text occurred in tweets about refugees, we observe higher probabilities of refugees contacting non-refugees. To evaluate the robustness of the relationship and remove potential confounding effects, we constructed a number of models with additional covariates. The effect was consistent across all models, and robust to the inclusion of other variables. The probability between cross group connections was larger in urban areas than non-urban, and higher when dissimilarity was higher. However, this pattern is sensitive to the definition of urban area.

### Conclusions

- Significant differences in district level segregation exists in both residence and activity.
- We observe time varying patterns of sentiment as measured by Social Media text analysis.
- Calls from non-Refugees to refugees fall in periods of low sentiment.
- We observe that refugees may respond to negative public sentiment by becoming more isolated, potentially exacerbating negative attitudes

### Acknowledgements

This research would have not been possible without the resources, advice and guidance of the people of the Center for Studies in Demography and Ecology, the Sociology Department at the University of Washington, the Data for Refugees Project, and the Max Planck Institute for Demography. Specifically I would like to thank the others who contributed to the manuscript of this paper; Kiran Garimella, Ott Toomet, Ingmar G. Weber, and Emilio Zagheni.