



# Do We Tweet Where We Ride?

**Alysha Alloway and Lauren Strug**  
**University of Minnesota MGIS**

**An exploration of tweet  
locations and transit data to  
discover if they show a  
spatial relationship.**

---

**People + Place**

**What do we know about where  
people spend their time?**

---

# Traditional

## Survey Data

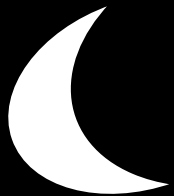
- Census, ACS, ATUS

## Population Counts

- Transit, Traffic

# Nontraditional

- Tweets
- Mobile phone calls
- IP location history
- Banking transactions
- Any app on your phone



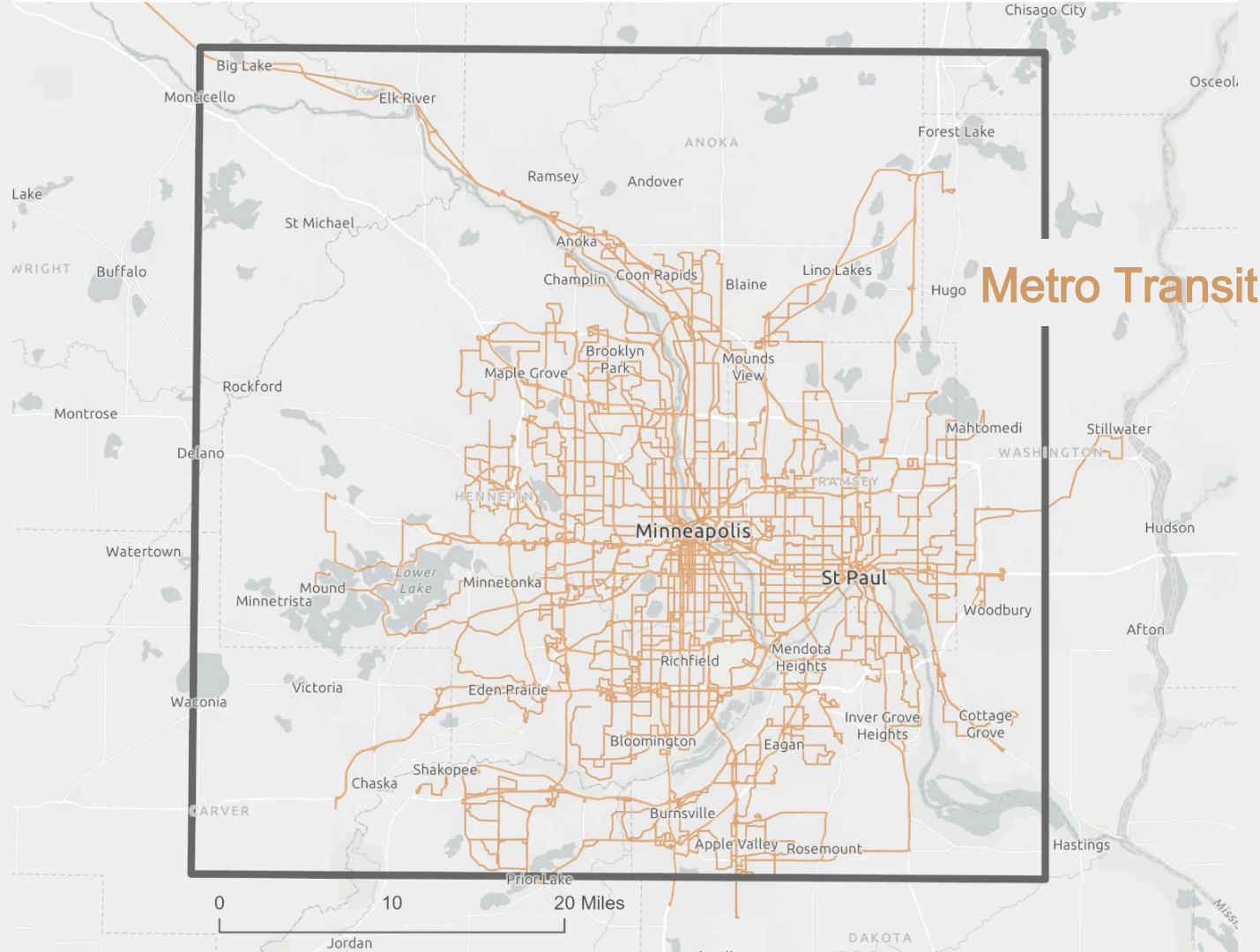
# US Census and ACS Population (Density)



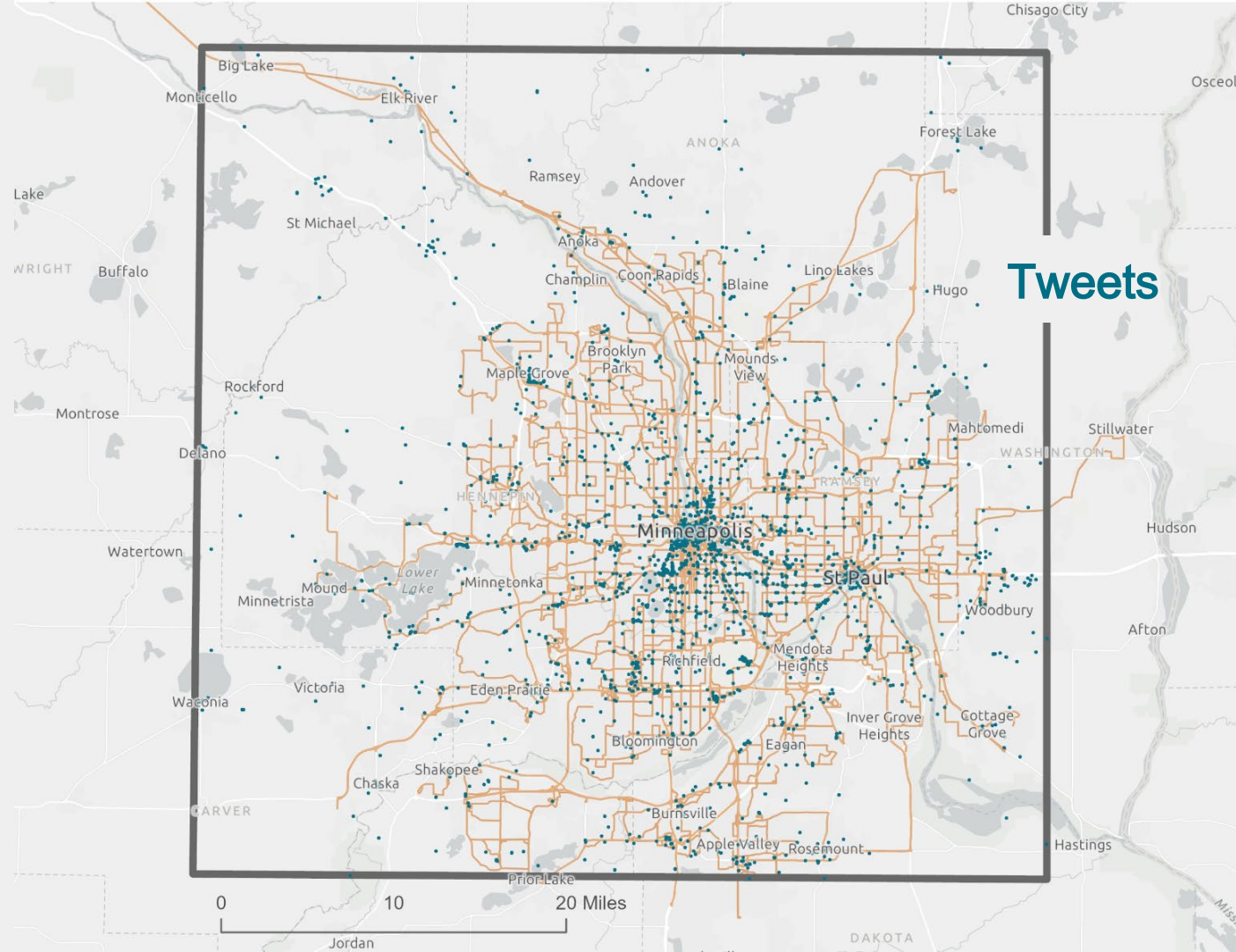
Where are people during the  
day?

**Tweets**  
**Transit Origin**  
**Transit Destination**

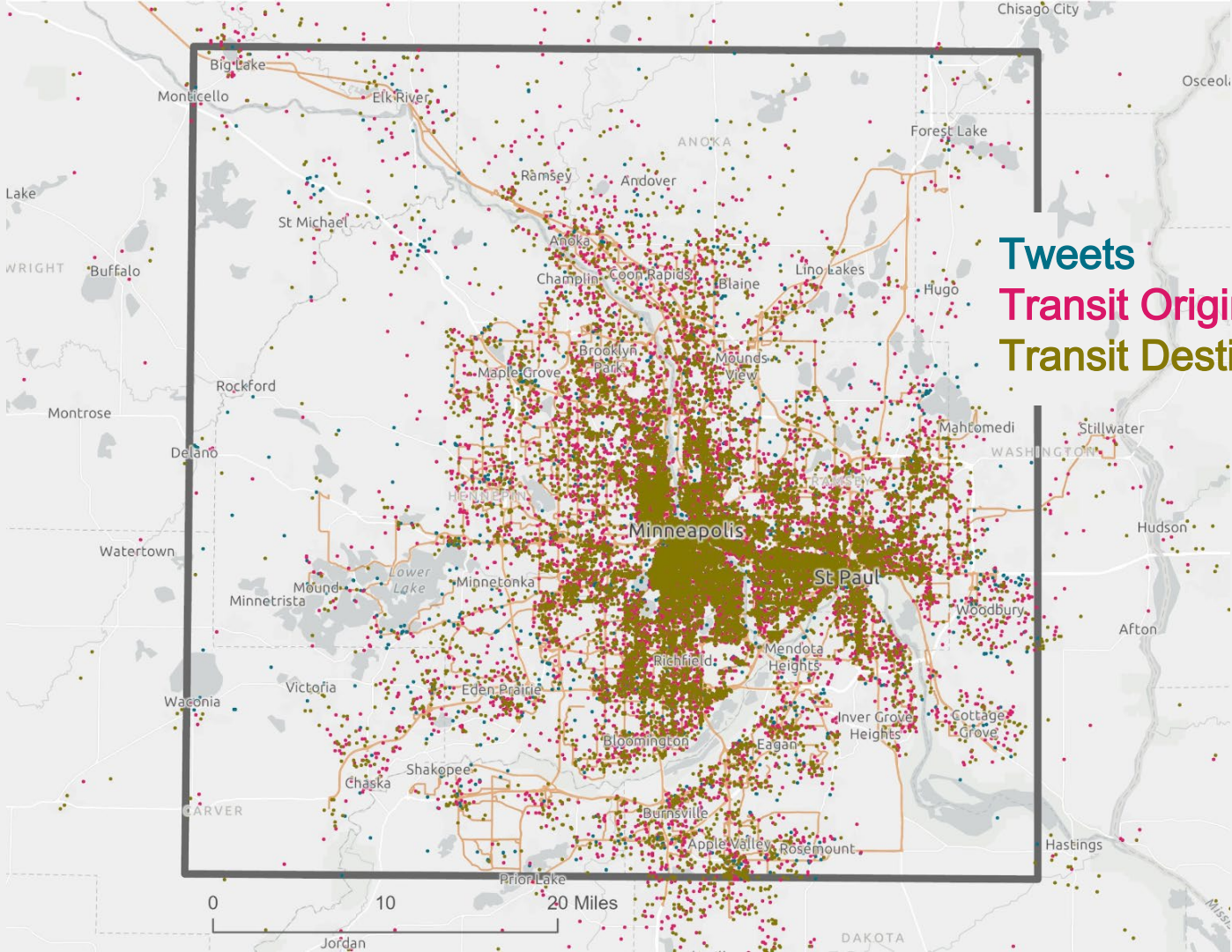
---







Tweets



**Tweets**  
**Transit Origin**  
**Transit Destination**

# Approach

**Exploratory, Python -based project**

# Approach

Exploratory, Python - based project

**Jupyter Notebook and Google Colab to work with python spatial libraries and the Twitter API**

# Approach

Exploratory, Python - based project

Jupyter Notebook and Google Colab to work with Python spatial libraries and the Twitter API

**ArcPro for visualizations**

**Tweets**

**Transit Origin**

**Transit Destination**

---

**“Geolocated tweets can capture  
features of human mobility for  
individuals within cities”**

**Jurdak et al 2015**

---

# Gathering Tweets

- List of users with **geolocated tweets** in our bounding box from SEDE (The Socio-Environmental Data Explorer) database
- Twitter REST API, pandas, csv, json

```
""" This is a function to get all of the tweets from a users timeline that have
coordinates and create a JSON file as output"""
def get_all_tweets(screen_name):
    #List to store tweets
    tweet_list = []

    r = api.request('statuses/user_timeline', {'screen_name': screen_name})
    for tweet in r:
        # Only grabs tweets with coordinates to append to the list
        if tweet['coordinates'] is not None:
            tweet_list.append(tweet)

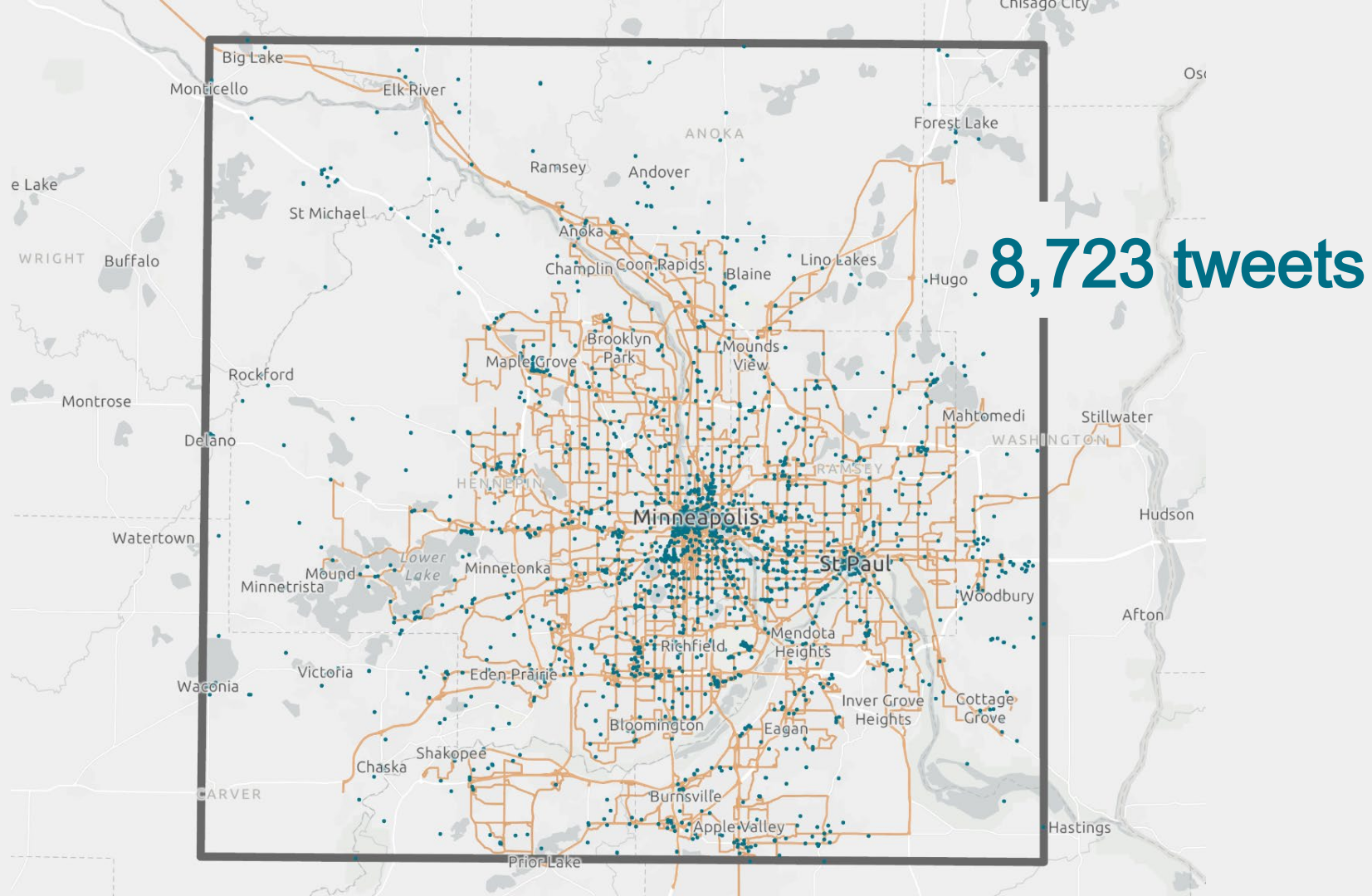
    # Convert our tweet list to JSON
    jsonlist = json.dumps(tweet_list)

    # Set filename we will use to store our tweets, so we have multiple files
    tweets_filename = "C:/Users/laure/Desktop/tweets/" + screen_name + ".json"

    # Open tweets.json to write contents (all of our tweets)
    with open(tweets_filename, 'w') as outfile:
        json.dump(jsonlist, outfile)

    #files.download('tweets.json')
    return tweet_list
```





# Python Workflow

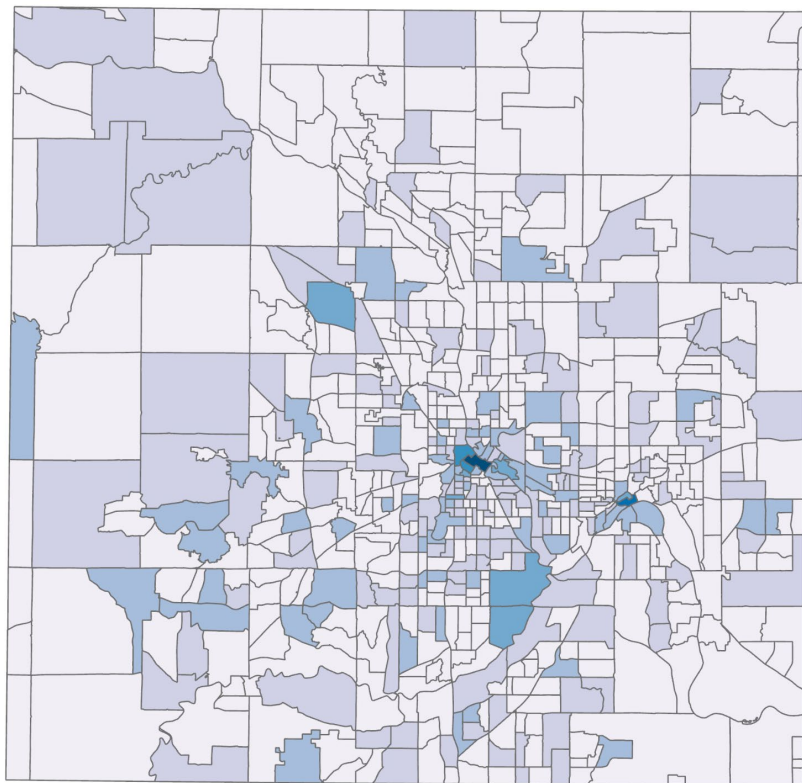
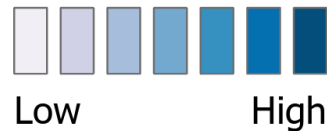


# What The Literature Tells Us About Tweeting

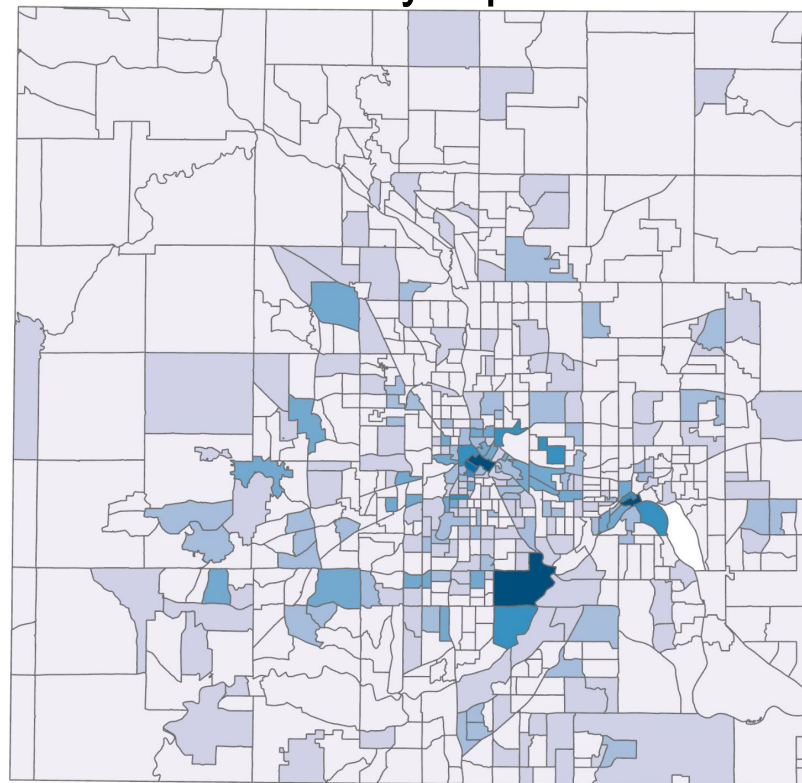
**People mainly tweet from home, work, or  
touristic locations/sports events** (Soliman et al, 2017)

**What do we see in the Twin Cities?**

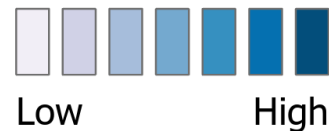
# Tweets Aggregated to Census Tract



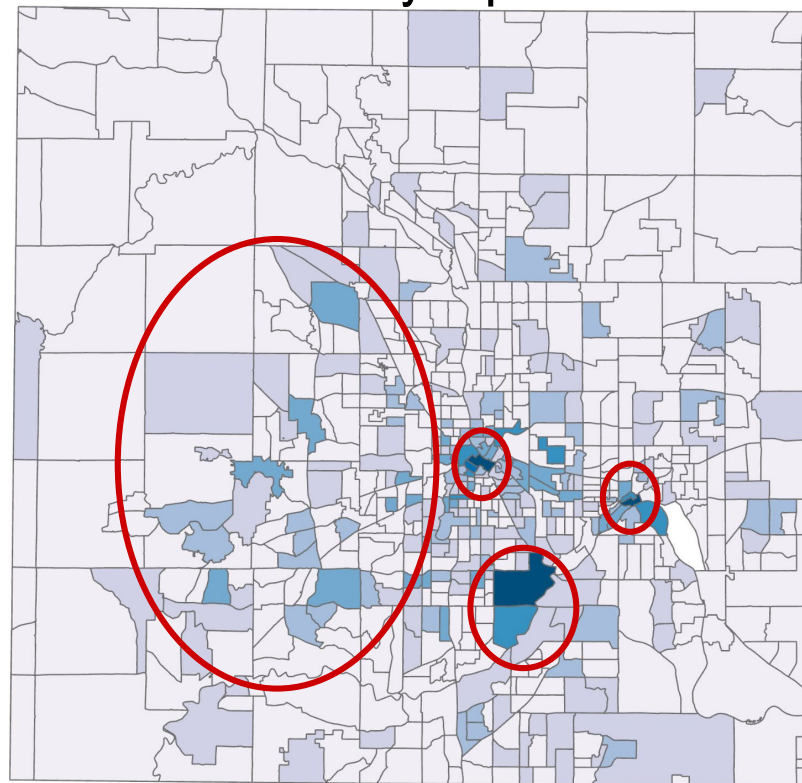
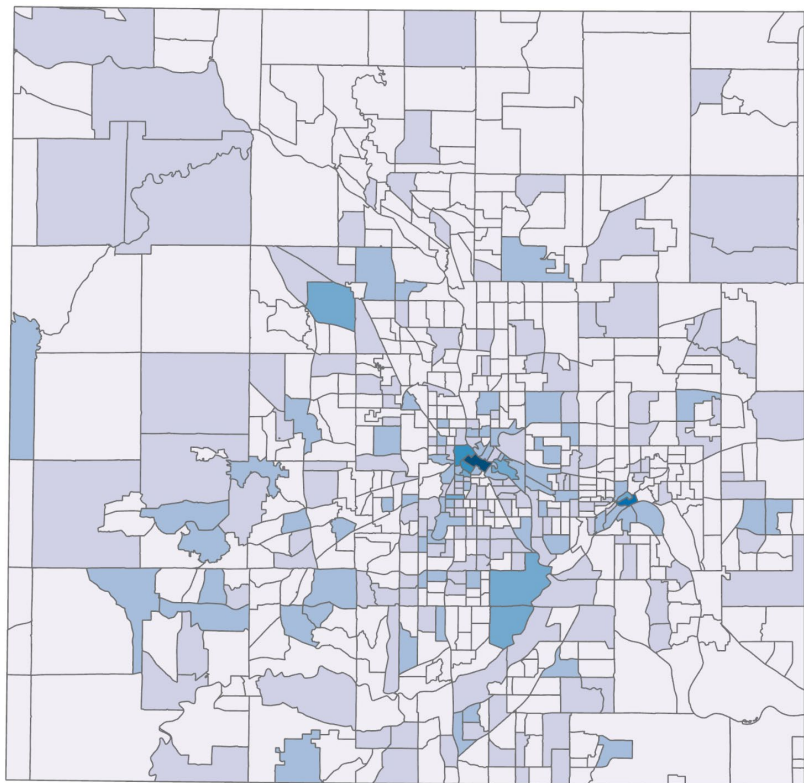
Normalized by Population

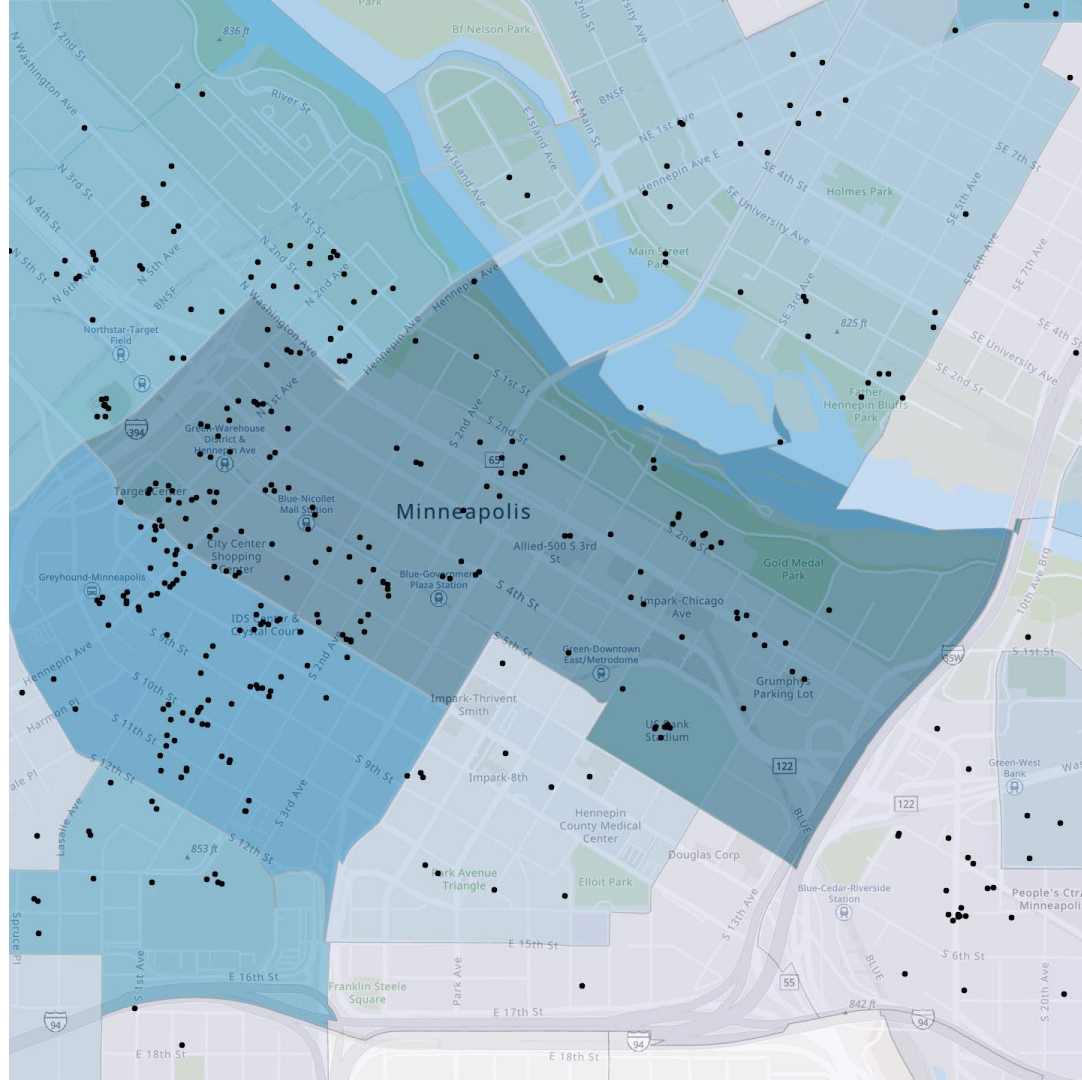


# Tweets Aggregated to Census Tract



Normalized by Population

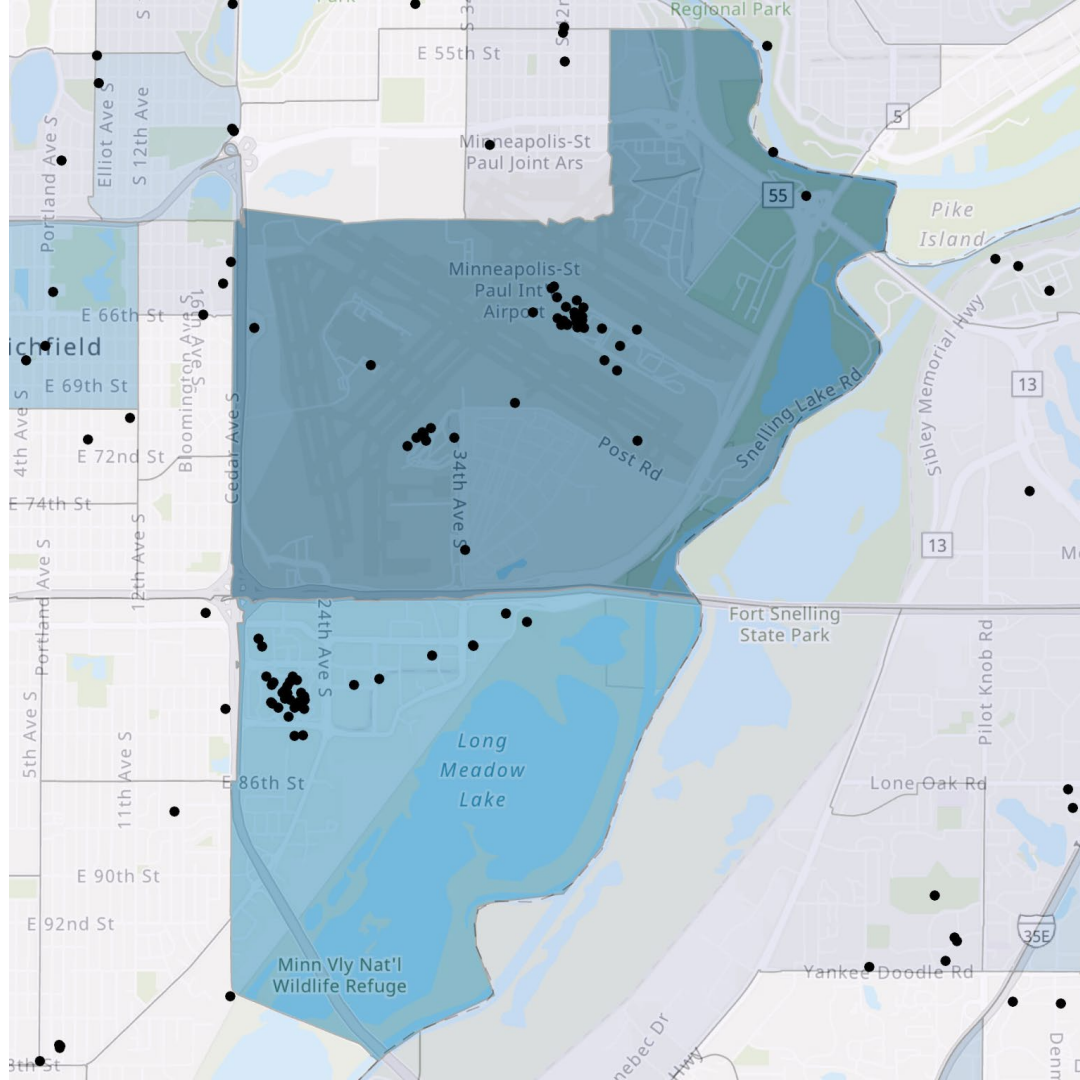




## 1421 Tweets from Downtown Minneapolis

- US Bank Stadium
- Target Center
- Target Field
- Government Plaza





213 Tweets

- Minneapolis/St. Paul International Airport (MSP)
- Mall of America

**Our twitter data supports  
work - home - special event  
location representation.**

---



**Tweets**

**Transit Origin**

**Transit Destination**

---

# Travel Behavior Inventory, 2016

## Transit On - Board Survey

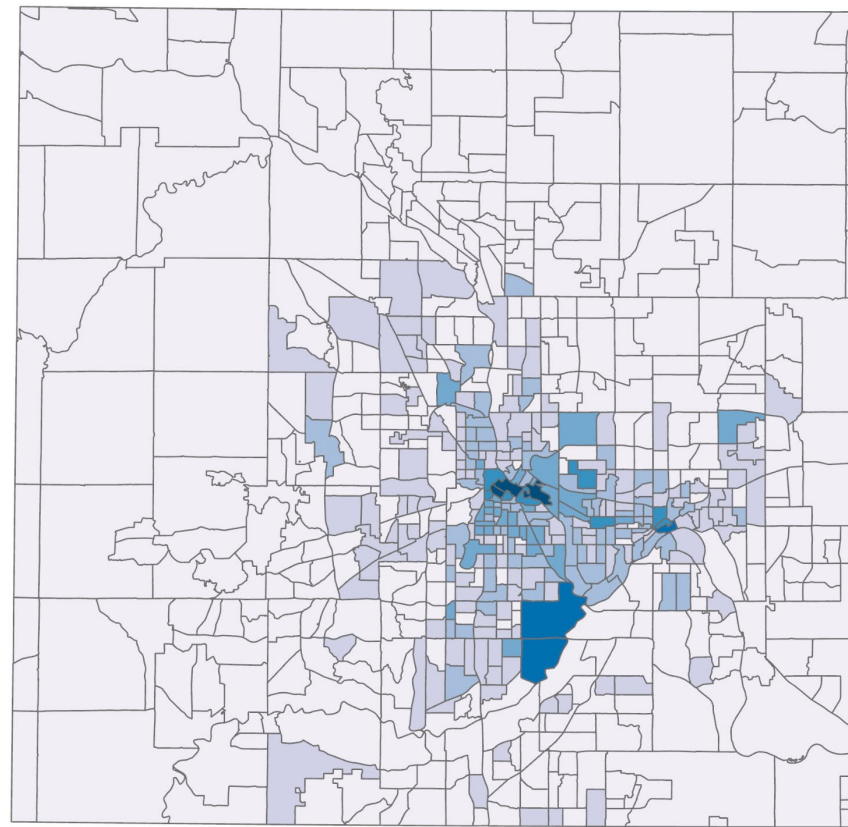
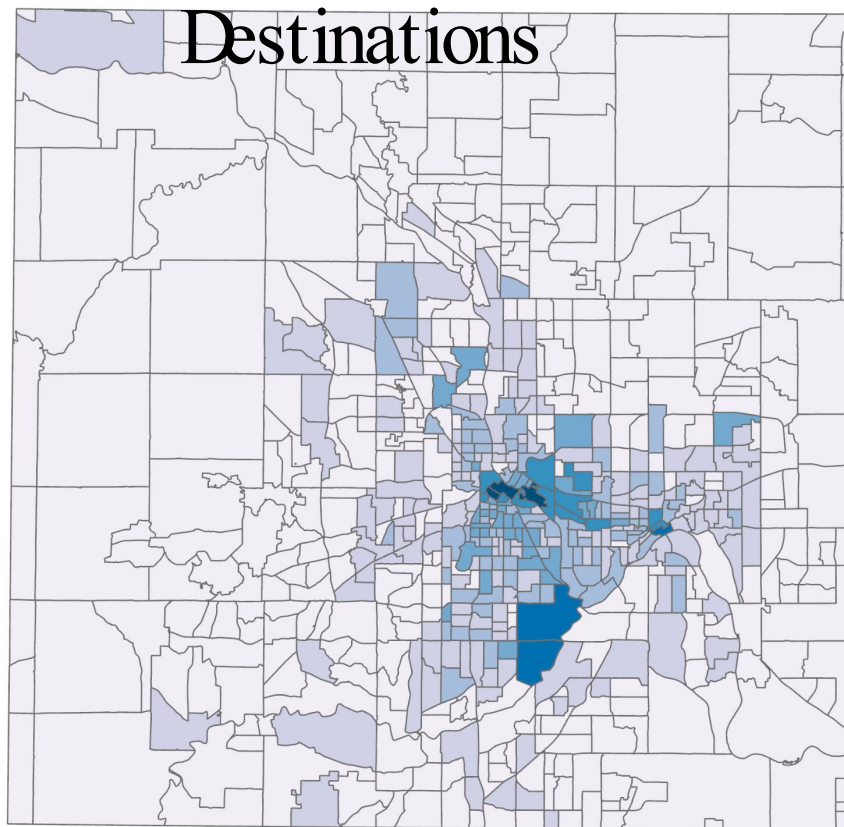
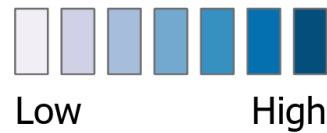
- 30,605 transit trips across all regional routes/providers
- Coordinates for origin and destination + more

**For transit, the major assumption is that people start from home and go to work/shopping/major destinations.**

**What do we see?**

---

# Origins

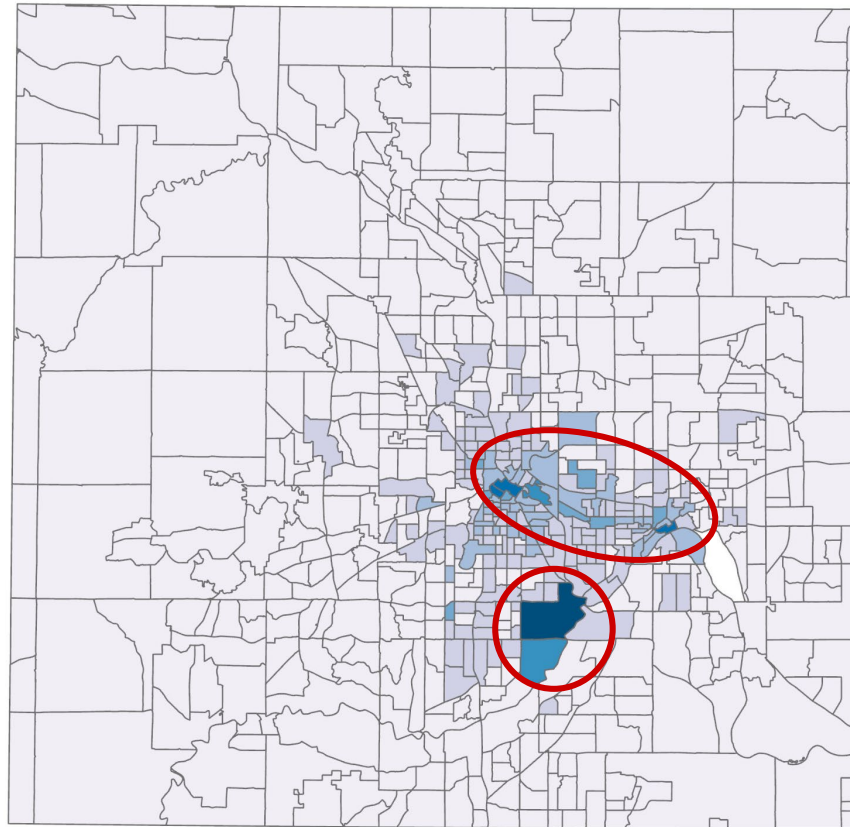
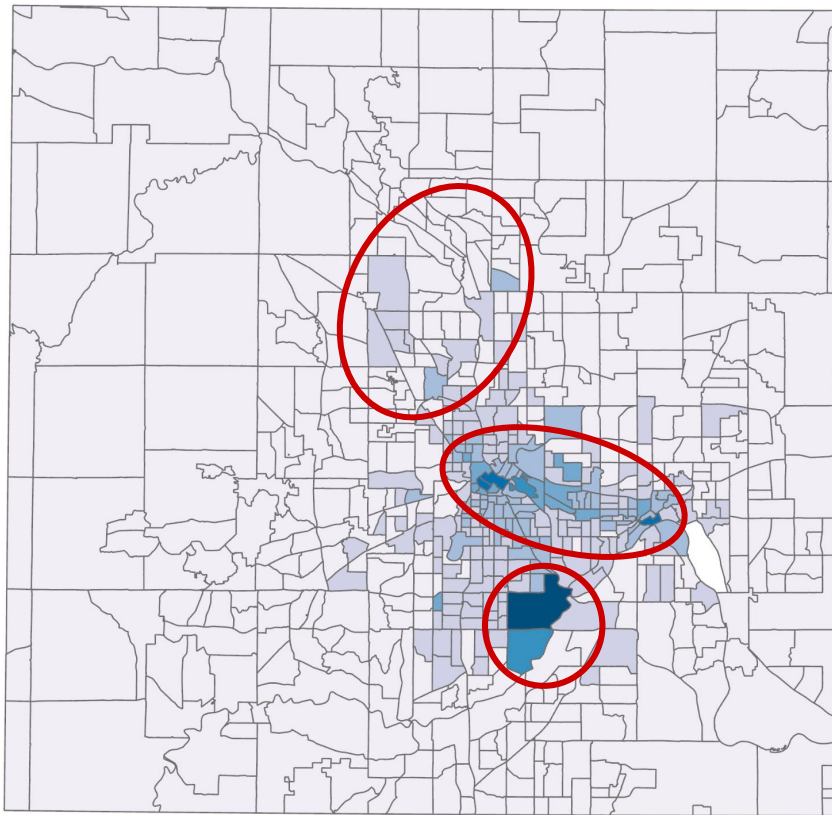


# Origins and Destinations Normalized by $F$



Low

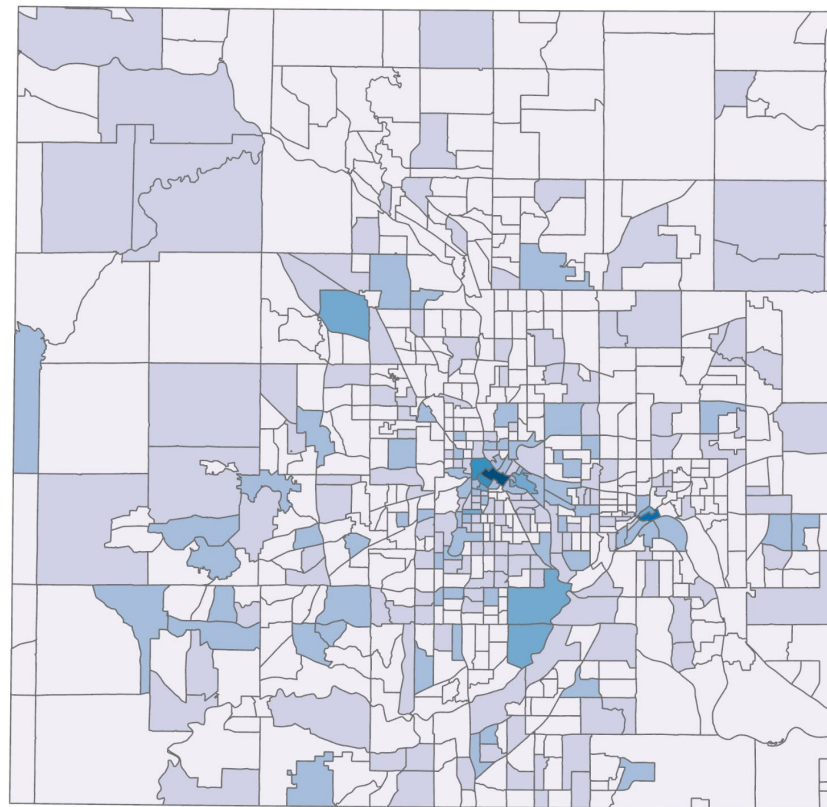
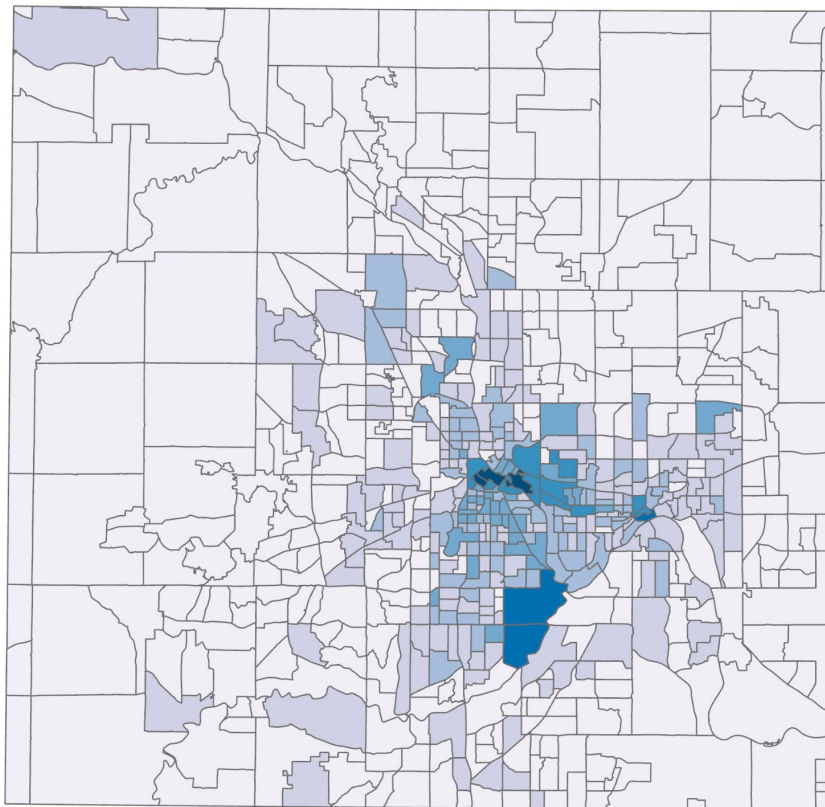
High



# Origins

vs.

# Tweets

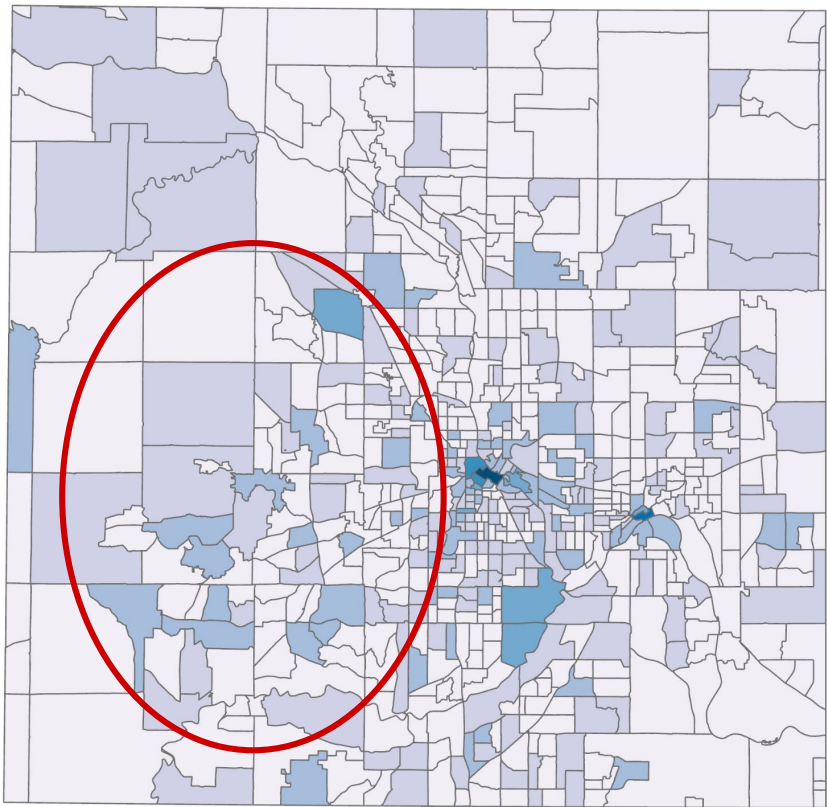
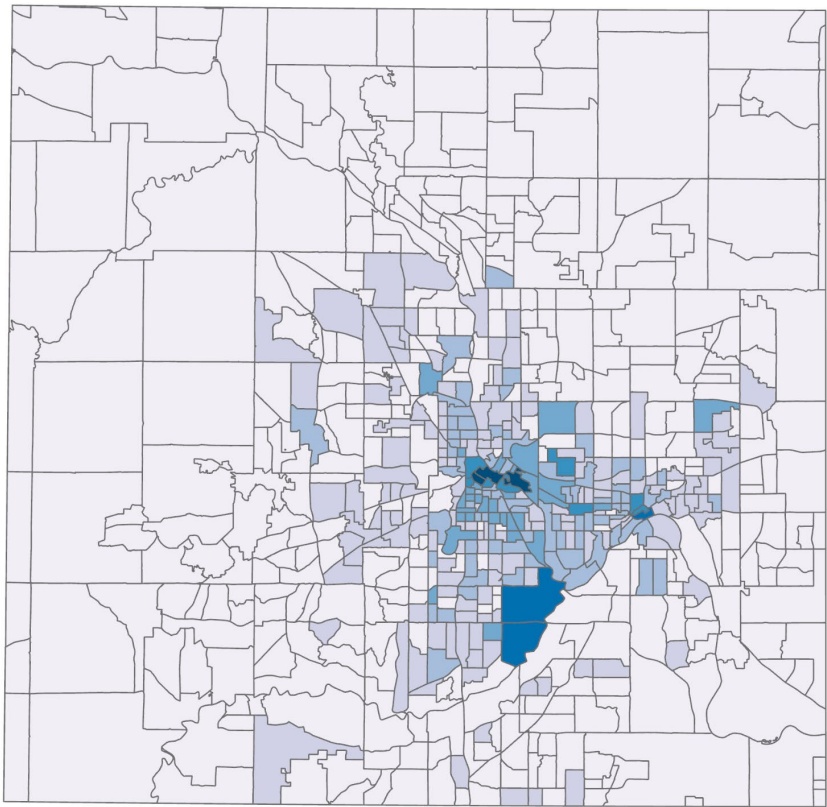




# Destinations

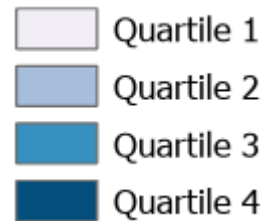
vs.

# Two

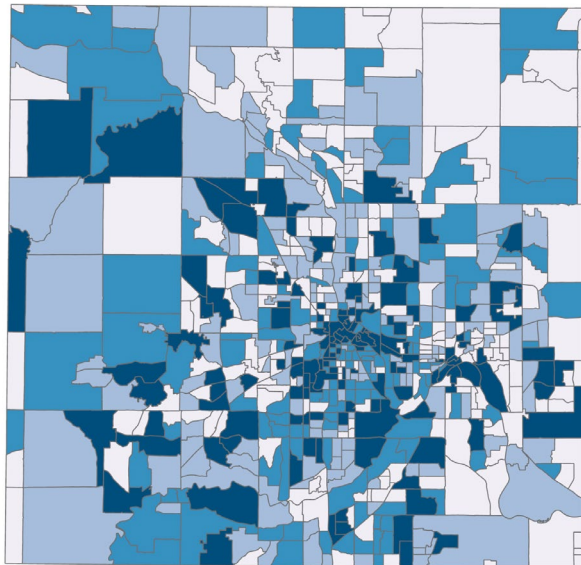


# Z-Scores - Quartiles

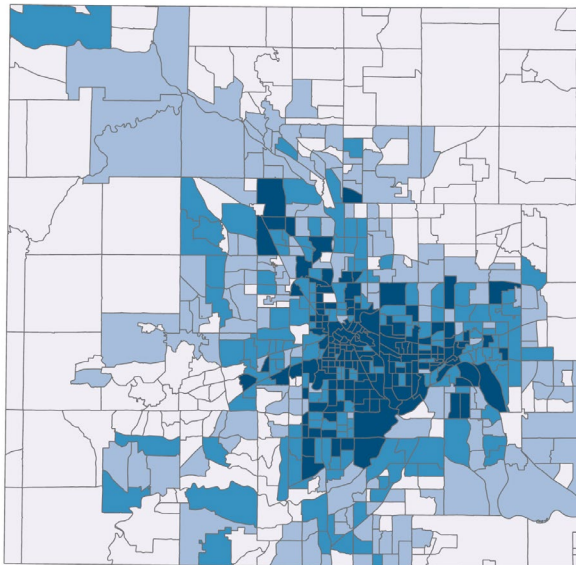
How many standard deviations each census tract deviates from the mean count of tweets? Origins? Destinations?



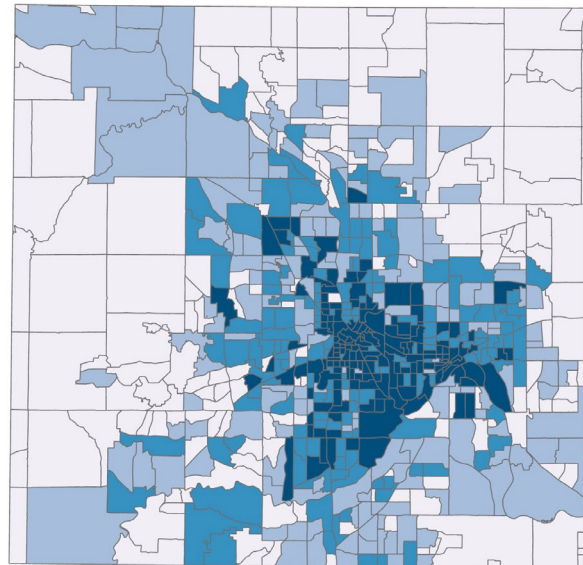
tweets



origins



destinations

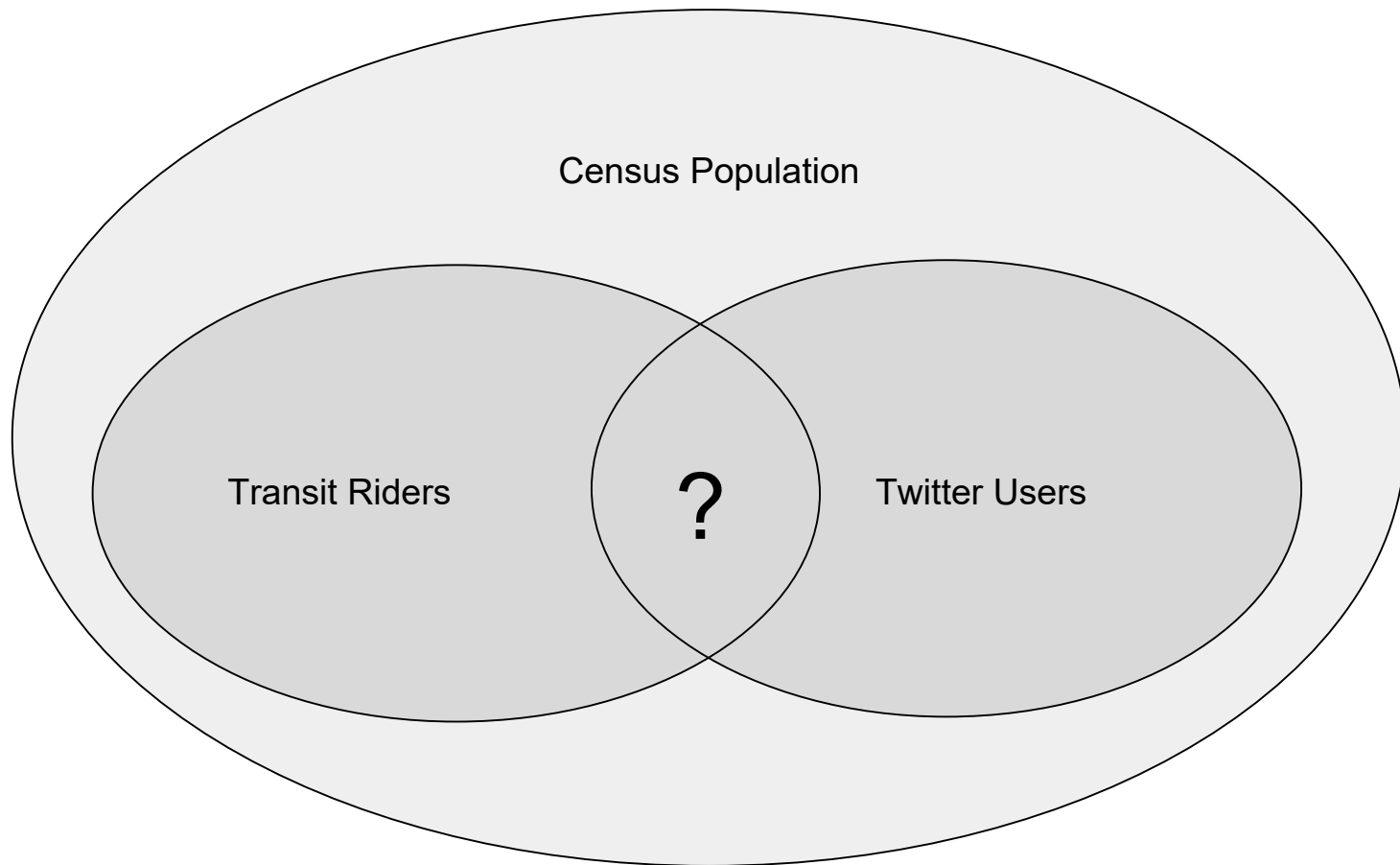




What we see:

Tweets come from places not represented well by transit users - people who use Twitter in the Twin Cities may not be in the same places as people who ride transit.

# The demographic breakdown



# Transit Users: Demographics



are people of color



have no working  
automobile available



are under age 35



have annual household  
income below \$25,000

(MetroTransit)

# Twitter Users: Demographics



66% of users are men



22% of adults in the US use Twitter



37% of Twitter users are between 18 and 29 years old, 25% between 30-49



56% of Twitter users earn \$50,000 or more in a year

(Omnicores, Salman Aslam)

**Do we tweet where we ride?**

---

# Takeaways

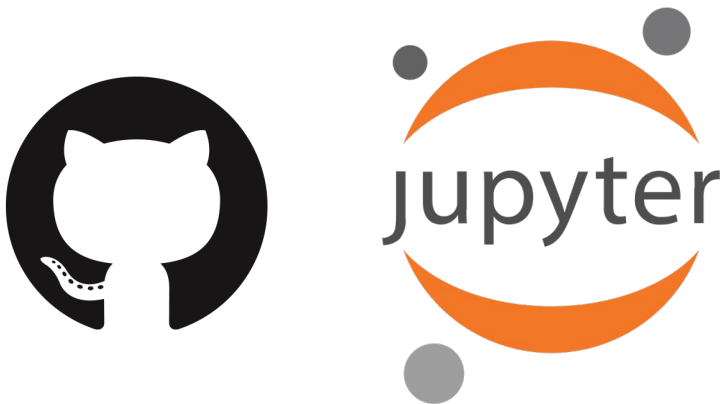
- We tweet where we ride - to a certain extent
- Data sources only tell us about the people they capture well

# Future Research

- Gather more tweets!
- Use tweet text to understand context
- Compare user location history with linked origin and destination of transit users
- Incorporate traffic data

# If you want to run our code:

<https://github.com/laurenstrugewald/Dweet-where-we-ride.git>





**Alysha Alloway**

**allow001@umn.edu**

**Lauren Strug**

**strug005@umn.edu**

---

# References

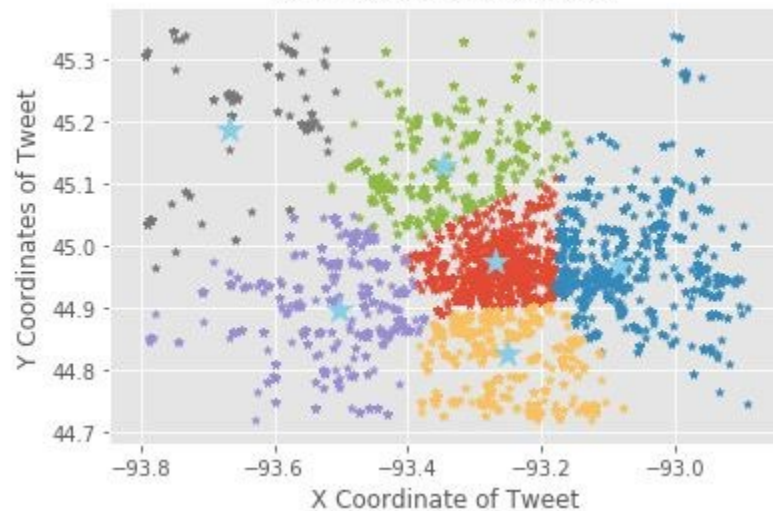
1. Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D., & Wu, Y. (2015). Understanding Human Mobility from Twitter. *PLoS ONE*, 10(7), E0131469.
2. Lee, J.H., Gao, S., Janowicz, K., Goulias, K.G., 2015. Can Twitter data be used to validate travel demand models?, In: IATBR., WINDSOR.
3. Lenormand M, Picornell M, Cantu'-Ros OG, Tugores A, Louail T, et al. (2014) Cross-Checking Different Sources of Mobility Information. *PLoS ONE* 9(8): e105184. doi:10.1371/journal.pone.0105184
4. Li, Yue, Li, Qinghua, & Shan, Jie. (2017). Discover Patterns and Mobility of Twitter Users-A Study of Four US College Cities. *ISPRS International Journal of Geo-Information*, 6(2), 42.
5. Liu J, Zhao K, Khan S, Cameron M, Jurdak R. Multi-scale population and mobility estimation with geo-tagged tweets. In: Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference on. IEEE; 2015. p. 83–86.
6. Longley, P. A., Adnan, M., & Lansley, G. (2015). The Geotemporal Demographics of Twitter Usage. *Environment and Planning A: Economy and Space*, 47(2), 465–484.
7. Molinari, M. E. , Oxoli, D., Kilsedar, C. E., & Brovelli, M. A. . (2018). User Geolocated Content Analysis for Urban Studies: Investigating Mobility Perception and Hubs Using Twitter. *The International Archives of the Photogrammetry*, 42(4), 439-442.
8. Perrin, A., & Anderson, M., (2019). Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. *Pew Research Center*. <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/>

# References con't

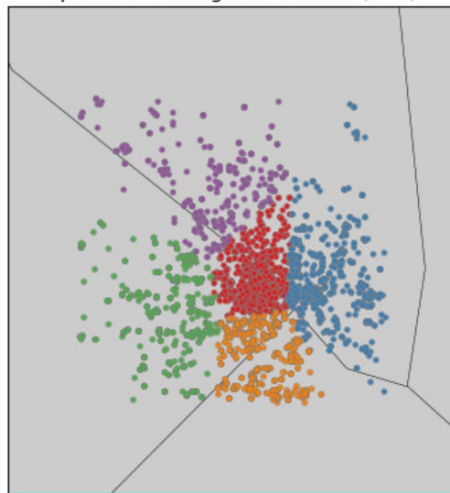
9. Rashidi, Abbasi, Maghrebi, Hasan, & Waller. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C*, 75(C), 197-211.
10. Shook, E. & Turner, V. (2015). The socio-environmental data explorer (SEDE): a social media-enhanced decision support system to explore risk perception to hazard events. *Cartography and Geographic Information Science*, 43(5), 427-441.
11. Soliman, A., Soltani, K., Yin, J., Padmanabhan, A. and Wang, S., (2017) Social sensing of urban land use based on analysis of Twitter users mobility patterns. PLoS One.
12. Yin, Junjun, Soliman, Aiman, Yin, Dandong, & Wang, Shaowen. (2017). Depicting urban boundaries from a mobility network of spatial interactions: A case study of Great Britain with geo-located Twitter data. *International Journal of Geographical Information Science*, 31(7), 1293-1313.

## Appendix - K Means

K number of clusters=6



Spatial Clustering with KMeans (k=5)



Spatial Clustering with KMeans (k=6)

