



Week 8

Social media and segregation

CATE HEINE, ARIANNA SALAZAR MIRANDA

11.S951 *Senseable City: Data and Analytics*

April 1

Lab



Use social media to study segregation

Link to access today's lab

1 Installing and loading packages

The installation process will take around 10 minutes. In the meantime, go ahead and familiarize yourself with the sections and overall structure of the lab.

```
## Spatial data analysis
system("sudo apt-get install r-base-dev r-base-doc")
system("sudo apt-get install r-cran-geojson")
system("sudo apt-get install r-cran-leaflet")
system("sudo apt-get install r-cran-leaflet")
system("sudo apt-get install r-cran-leaflet")
system("sudo apt-get install r-cran-leaflet")
system("sudo apt-get install r-cran-leaflet")
system("sudo apt-get install r-cran-leaflet")

## General data analysis
install.packages("tidy")
install.packages("dplyr")

## Cluster analysis for home location selection
install.packages("sp")

## Twitter scraping
system("sudo apt-get install r-cran-tweet")

## Mapping
install.packages("leaflet")
system("sudo apt-get install r-cran-leaflet")

## Extra leaflet support: if you're not using Ubuntu, you won't need this
if (system("dpkg-query -f='${Package} ${Version} ${Architecture}\n' | grep -i leaflet") == "") {
  install.packages("leaflet")
}
if (system("dpkg-query -f='${Package} ${Version} ${Architecture}\n' | grep -i leaflet") == "") {
  install.packages("leaflet")
}
if (system("dpkg-query -f='${Package} ${Version} ${Architecture}\n' | grep -i leaflet") == "") {
  install.packages("leaflet")
}
}
```

2 Load the libraries

```
## Spatial data analysis
library(sf)
sf::sf_use_s2(FALSE)
library(geojsonsf)
library(ropr)
library(acsc)
library(dbscan)

## General data analysis
library(tidy)
library(dplyr)
require("isbbs")

## Twitter scraping
library(rtweet)

## Mapping
library(mapsView)
library(leaflet)
library(htmlwidgets)
library(IRdisplay)

## Plotting
library(ggplot2)
library(classInt)
```

Traditional measures of segregation are based on residential data



- Black
- White
- Hispanic
- Asian/Pacific Islander
- Native American
- Multi-race and other

DIVERSITY SCALE

Less diverse

More diverse

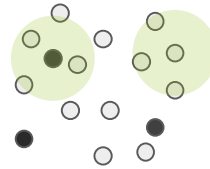
Measures of Segregation

Measures of Exposure

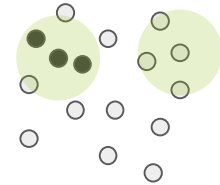
what is the probability of a minority member co-locating with another minority member (as opposed to a majority member)?

Isolation index, exposure index

low isolation



high isolation



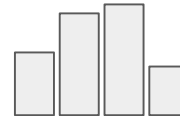
Measures of Evenness

what is the difference between the distribution of people you encounter and the distribution of people across the whole city?

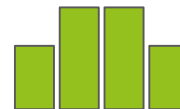
Dissimilarity index, Theil's entropy, Gini coefficient

low entropy

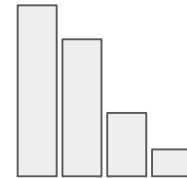
friend/neighbor
income
distribution



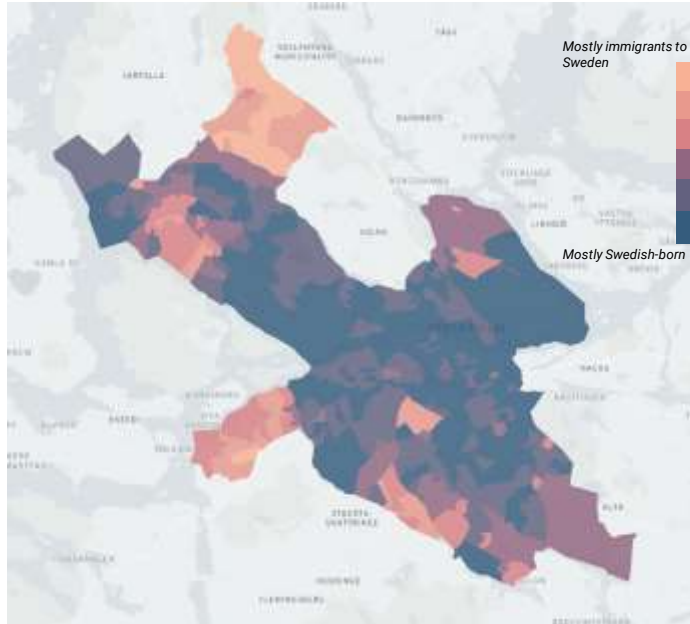
city income
distribution



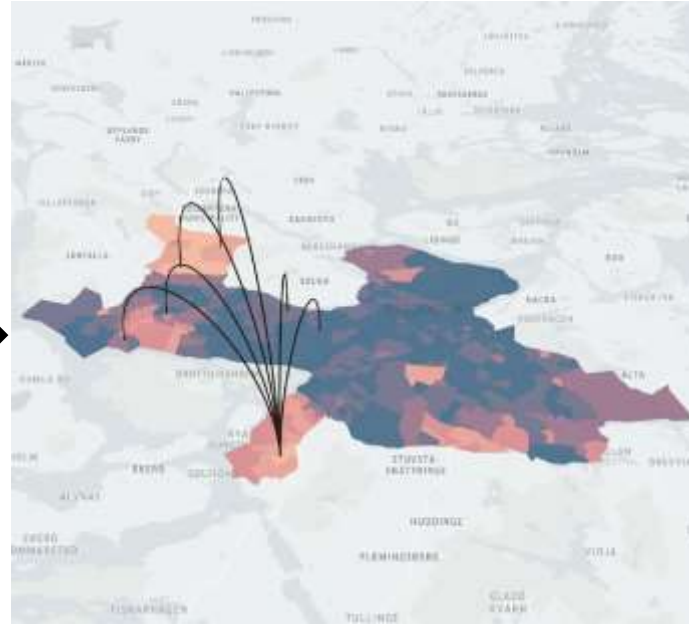
high entropy



But we experience segregation beyond our homes as well



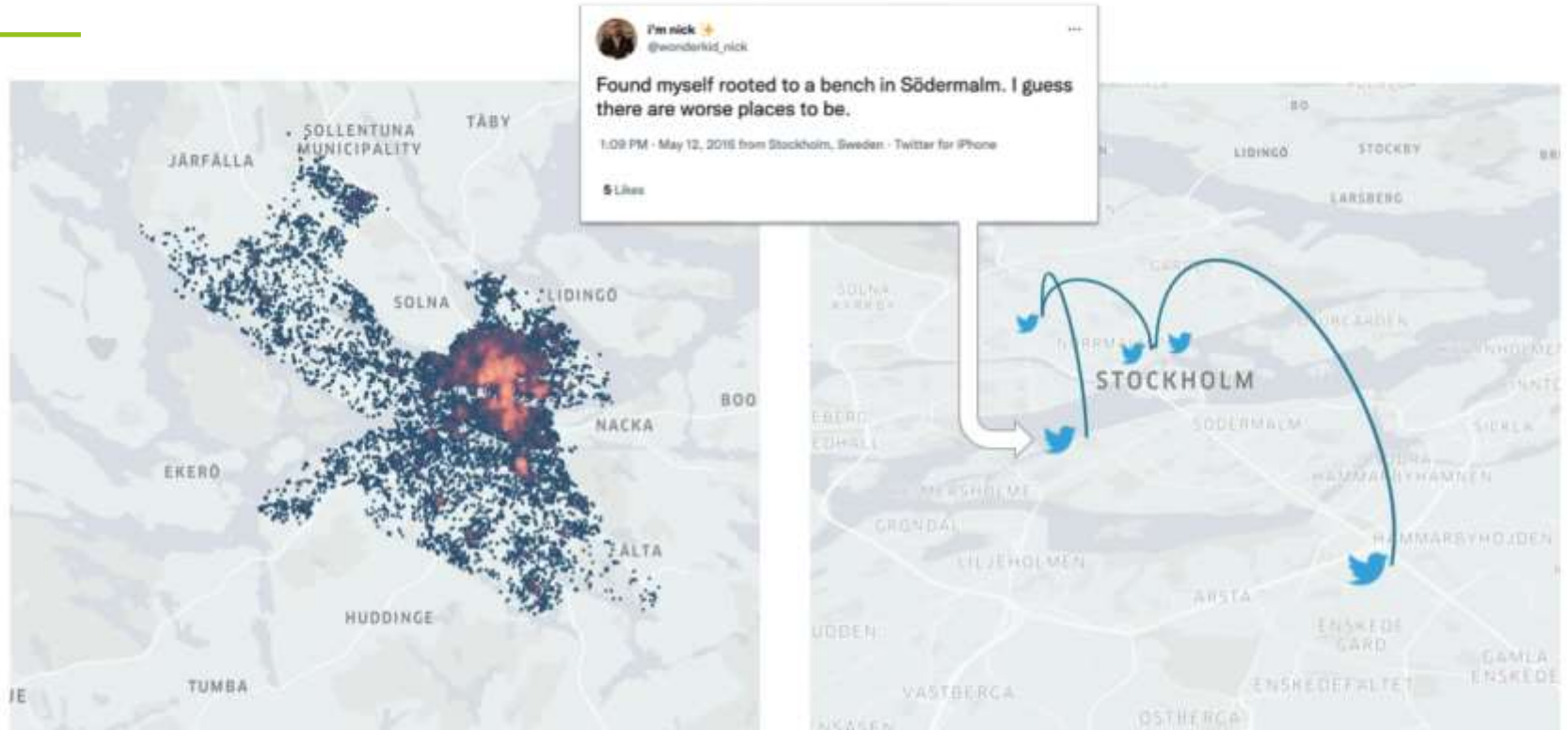
Residential immigration segregation in Stockholm.



Daily movements.



Using social media data to study social segregation



Today's lab: Paris!



Objectives:

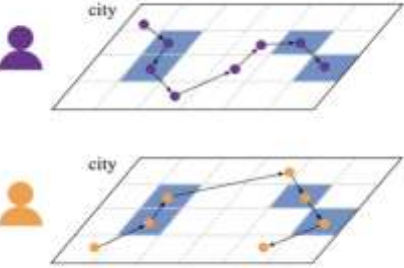
1. Scrape data from Twitter API

user_id	screen_name	location	followers_count	statuses_count	profile_image_url
150079981	111007498	2010-01-26 00:54:54	2107498	48481848	48481848
150079982	111007499	2010-01-26 00:54:54	2107499	48481849	48481849
150079983	111007500	2010-01-26 00:54:54	2107500	48481850	48481850
150079984	111007501	2010-01-26 00:54:54	2107501	48481851	48481851
150079985	111007502	2010-01-26 00:54:54	2107502	48481852	48481852
150079986	111007503	2010-01-26 00:54:54	2107503	48481853	48481853
150079987	111007504	2010-01-26 00:54:54	2107504	48481854	48481854
150079988	111007505	2010-01-26 00:54:54	2107505	48481855	48481855
150079989	111007506	2010-01-26 00:54:54	2107506	48481856	48481856
150079990	111007507	2010-01-26 00:54:54	2107507	48481857	48481857
150079991	111007508	2010-01-26 00:54:54	2107508	48481858	48481858
150079992	111007509	2010-01-26 00:54:54	2107509	48481859	48481859
150079993	111007510	2010-01-26 00:54:54	2107510	48481860	48481860
150079994	111007511	2010-01-26 00:54:54	2107511	48481861	48481861
150079995	111007512	2010-01-26 00:54:54	2107512	48481862	48481862
150079996	111007513	2010-01-26 00:54:54	2107513	48481863	48481863
150079997	111007514	2010-01-26 00:54:54	2107514	48481864	48481864
150079998	111007515	2010-01-26 00:54:54	2107515	48481865	48481865
150079999	111007516	2010-01-26 00:54:54	2107516	48481866	48481866
150080000	111007517	2010-01-26 00:54:54	2107517	48481867	48481867

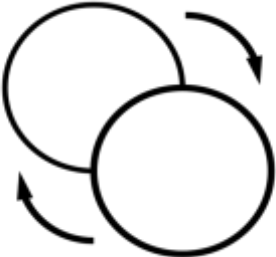
2. Maps and graphs using Twitter



3. Calculate 2 measures of experienced segregation



4. Learn about the benefits and limitations of using social media

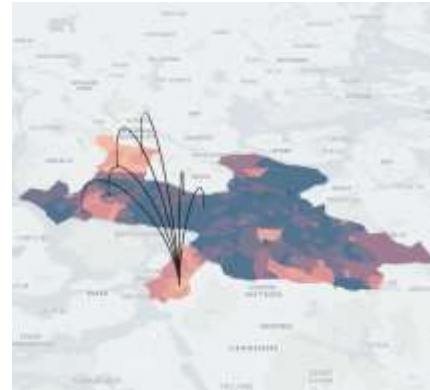


Part 1 & 2: Scrape data from Twitter API

- ❑ Create a Twitter developer account
- ❑ Obtain access keys
- ❑ Download tweets using the rtweet package in R
- ❑ Clean data and join it to Paris neighborhoods
- ❑ Visualize tweet usage patterns using graphs and maps

Part 3: Calculate measures of segregation using Twitter data

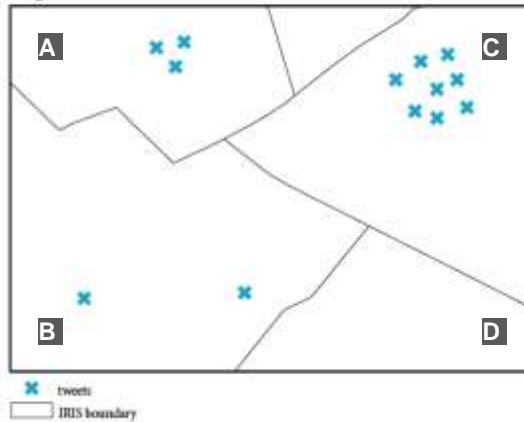
- 2 approaches to calculate experienced segregation:
 - **Home-Location Segregation:** Use home-location to infer the socio-economic characteristics of each person and use it to measure the diversity of users that visit each neighborhood.
 - **Visitation Segregation:** Measure based on the socio-economic characteristics of the neighborhoods you visit and remains agnostic of where you live



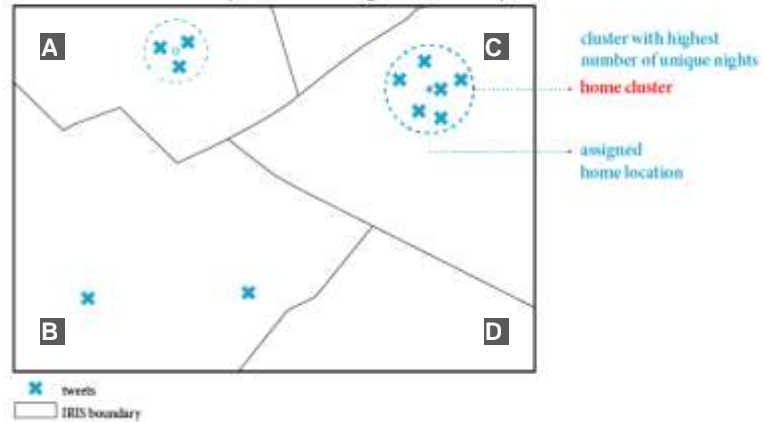
Home-Location Segregation

- ❑ Subselect users who tweet more than a given threshold (5 times at night).
- ❑ Use the DBScan algorithm to identify clusters of nighttime activity.
- ❑ Identify the cluster where the user spent the highest number of unique nights (home cluster) and assign it as the home location

Night tweets of a user



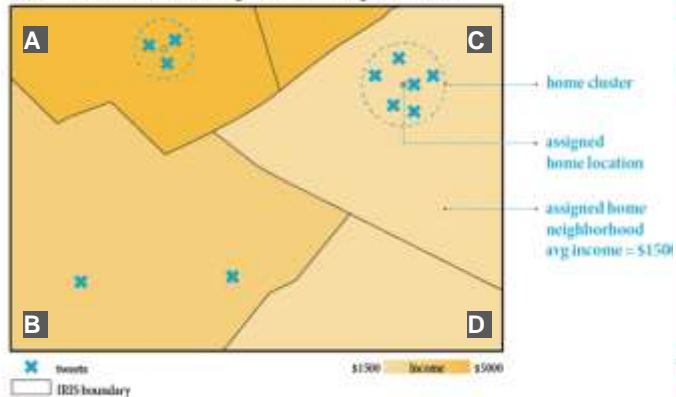
Use DBScan to identify clusters of nighttime activity



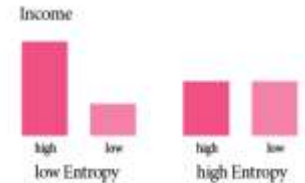
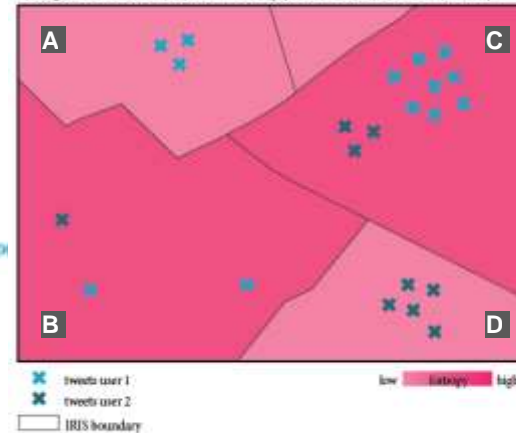
Home-Location Segregation

- ❑ Find the IRIS zone that contains this home cluster, and attach the user to the IRIS zone's mean income.
- ❑ Bin the user's income by deciles in each IRIS zone.
- ❑ Calculate the entropy of this distribution

attach income of the assigned IRIS neighborhood



Neighborhood-level Entropy (measure of diversity)

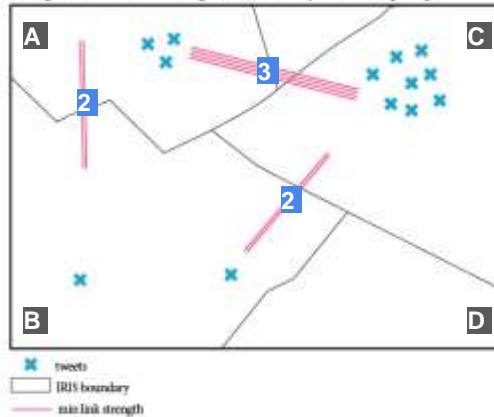


Visitation Segregation

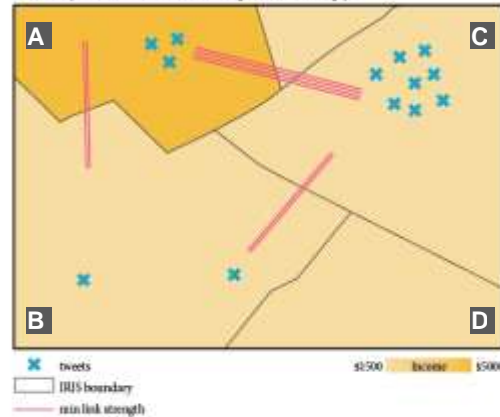
By moving between two neighborhoods, an individual creates a social or economic connection between them

- ❑ Calculate linkage strength between neighborhoods:
 - For each pair of neighborhoods, count the number of times a user tweets from each one and take the minimum.
- ❑ Minimum gives weight to heavy activity in both locations without being skewed by extreme activity in one of them.
- ❑ Assign income to each neighborhood
- ❑ Compute the entropy across the connection between neighborhoods

Neighborhood linkage defined by flow of people



overlay income and compute entropy



Part 4: Discussion

- Which measure would you use if you wanted to know which neighborhoods experience more social mixing? Can you explain why?
- Can you name one advantage of using the visitation segregation measure instead of using one that uses home-location estimation?
- What other datasets could be used to construct similar measures of social mixing?