Week 8

Social media and segregation

CATE HEINE, ARIANNA SALAZAR MIRANDA

Lab

Use social media to study segregation

Link to access today's lab


```
#load the libraries

## Spatial data analysis
library(sf)
sf::sf_use_s2(FALSE)
library(geojsonsf)
library(repr)
library(acid)
library(dbscan)

## General data analysis
library(tidyr)
library(dplyr)
require('tibble')

## Twitter scraping
library(rtweet)

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

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

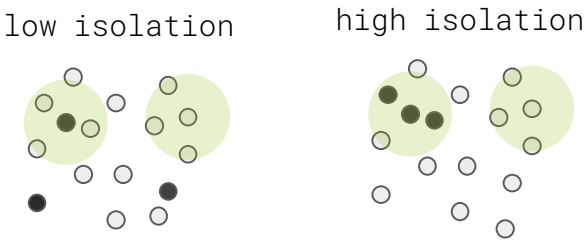
Traditional measures of segregation are based on residential data

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)?

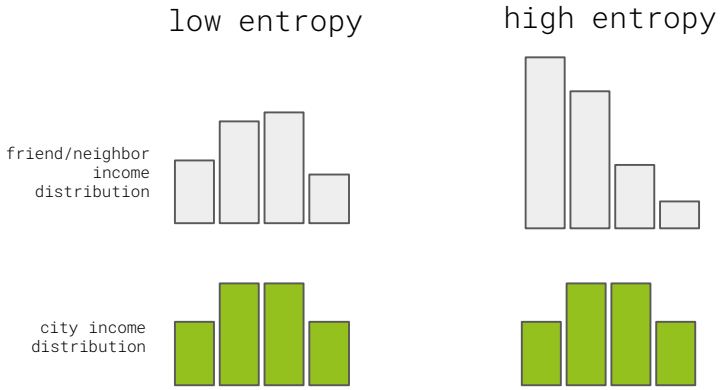
Isolation index, exposure index



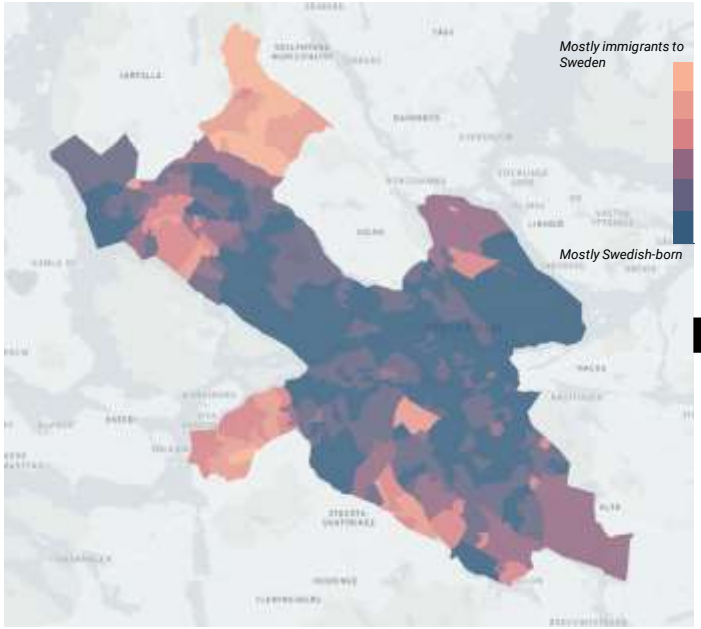
Measures of Evenness

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

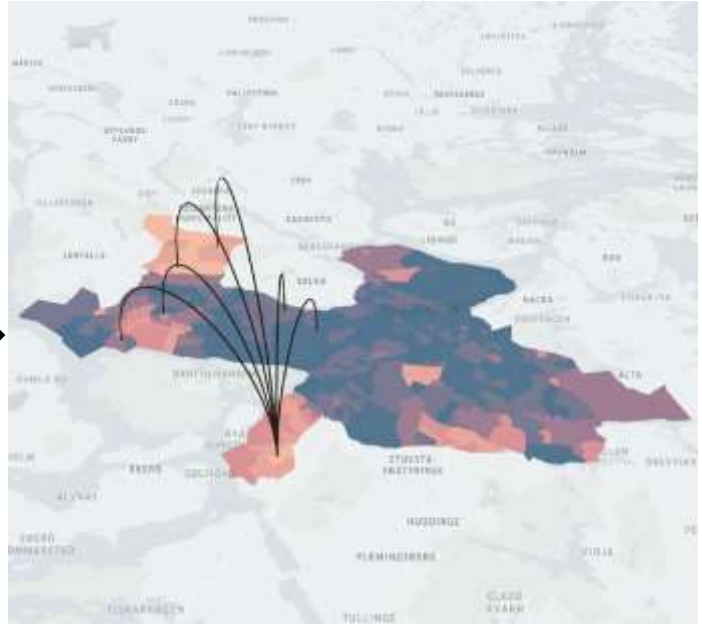
Dissimilarity index, Theil's entropy, Gini coefficient



But we experience segregation beyond our homes as well

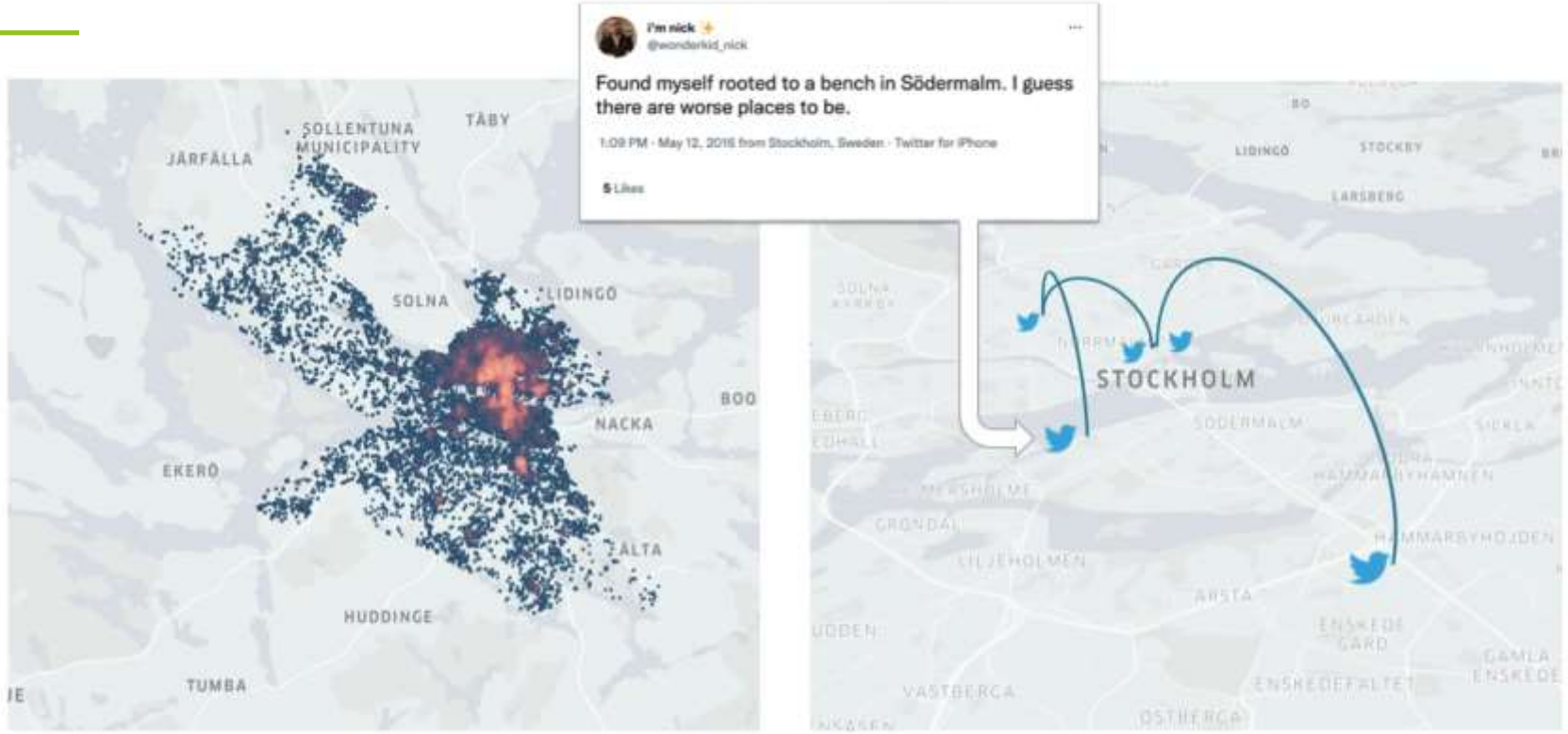


Residential immigration segregation in Stockholm.

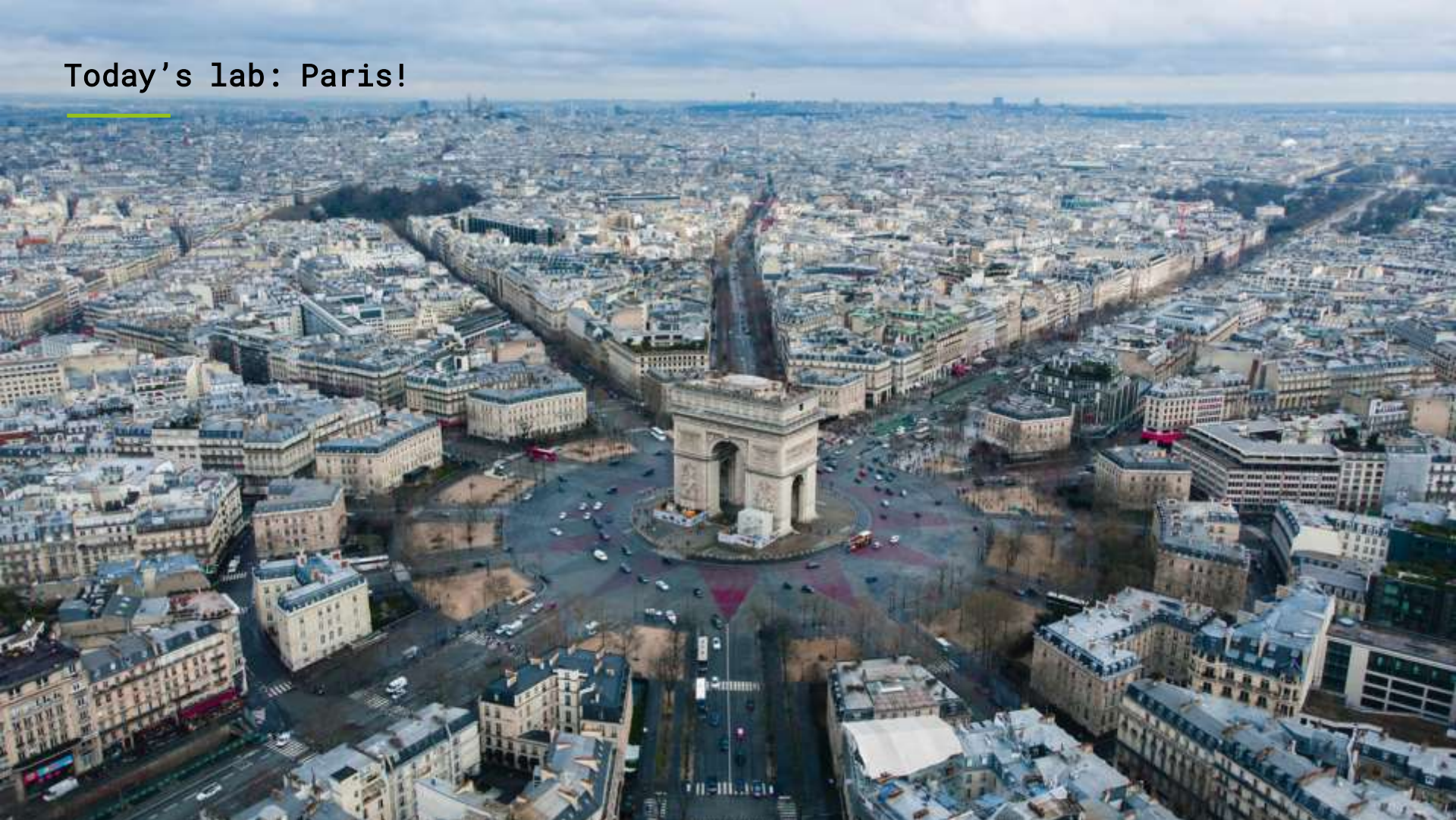


Daily movements.

Using social media data to study social segregation

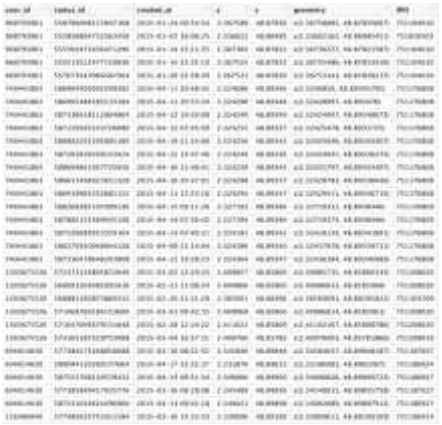


Today’s lab: Paris!



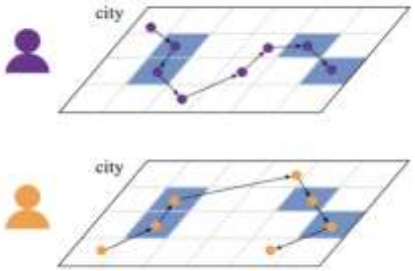
Objectives:

1. Scrape data from Twitter API

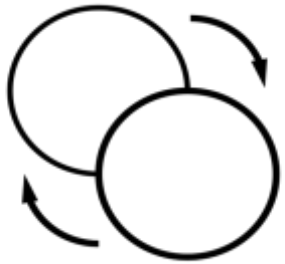


2. Maps and graphs using Twitter

3. Calculate 2 measures of experienced segregation



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

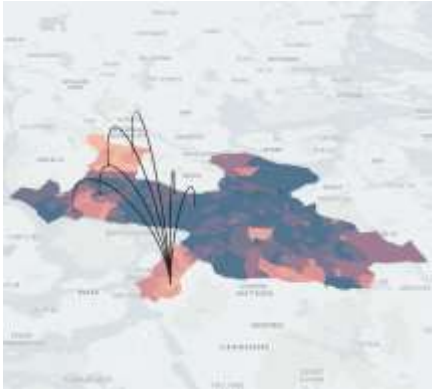


Part 1 & 2: Scrape data from Twitter API

- ❑ 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**

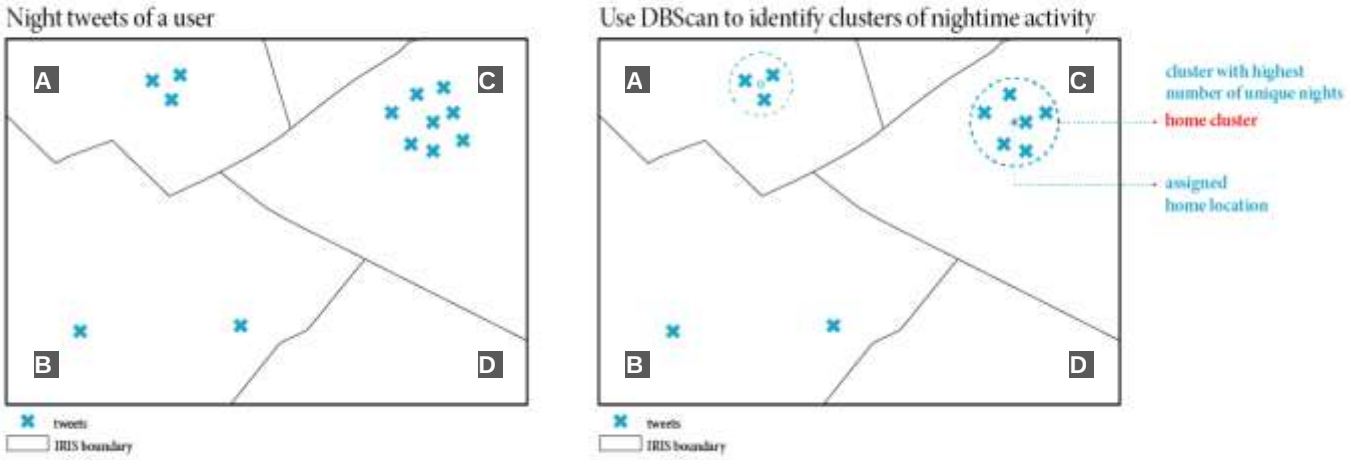
Part 3: Calculate measures of segregation using Twitter data

- 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



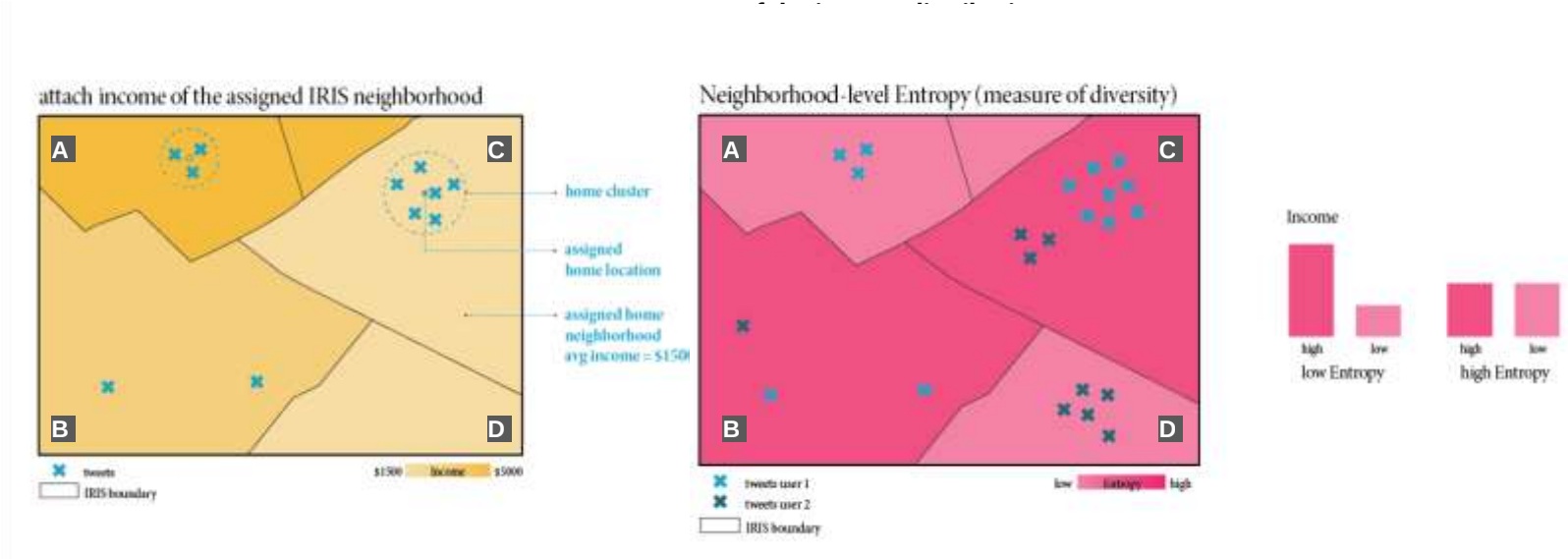
Home-Location Segregation

- 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



Home-Location Segregation

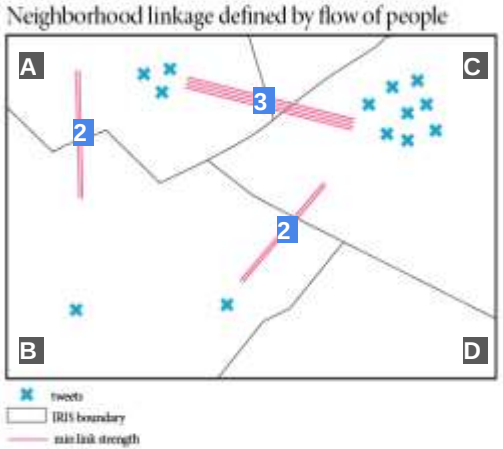
- ❏ Find the IRIS zone that contains this home cluster, and attach the user to the IRIS zone's mean income.
- ❏ Bin the user's income by deciles in each IRIS zone.
- ❏ Calculate the entropy of this distribution
  - ○

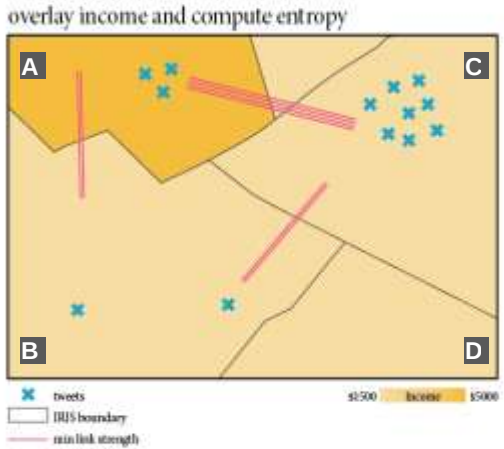


Visitation Segregation

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





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?