

# Data Science? Make it Spatial

DC R Conference 2018

Angela Li

University of Chicago

2018-11-08

# About me



# About me



- R Spatial Advocate at Center for Spatial Data Science, UChicago

# About me



- R Spatial Advocate at Center for Spatial Data Science, UChicago
- Founder of R-Ladies Chicago

# About me



- R Spatial Advocate at Center for Spatial Data Science, UChicago
- Founder of R-Ladies Chicago
- Wrote a thesis using spatial econometrics...

# About me



- R Spatial Advocate at Center for Spatial Data Science, UChicago
- Founder of R-Ladies Chicago
- Wrote a thesis using spatial econometrics...
- About property values in Detroit...

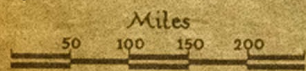
# About me



- R Spatial Advocate at Center for Spatial Data Science, UChicago
- Founder of R-Ladies Chicago
- Wrote a thesis using spatial econometrics...
- About property values in Detroit...
- Under Luc Anselin, who literally wrote the book on spatial econometrics

# Geography?







# Geography!

## Healthy Access: Medical Services

Explore the locations of medical service providers in Chicago, as provided by the Department of Public Health.

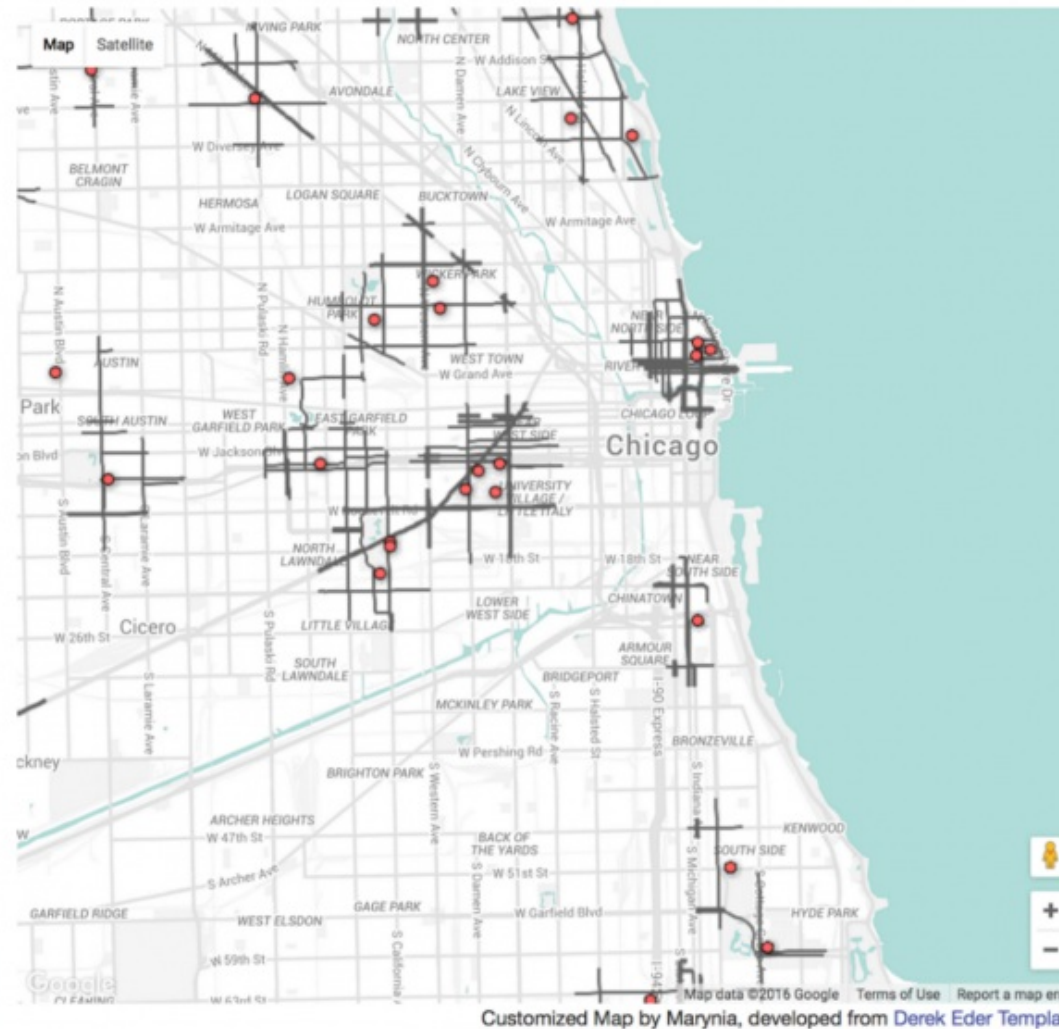
- ☐ Federally Qualified Health Centers
- ☒ Hospitals
- ☐ All Free Health Clinics
- ☐ School-Based Health Centers
- ☐ Other Health Providers

### Organization Search

### Address [\(find me\)](#)

within **1/2 mile**

51 results found



## Service Areas

Service areas are calculated along the road network. In this version, primary and secondary roads are used (and residential streets are not). Click to view service areas for selected health providers by distance.

☐ FQHC Clinics

☐ Hospitals

☐ Free Clinics

☐ School Clinics

☐ Other Providers

## Data

This dynamic data is provided from the Chicago Department of Public Health. It will be updated on a regular basis. Explore the data table in more depth:

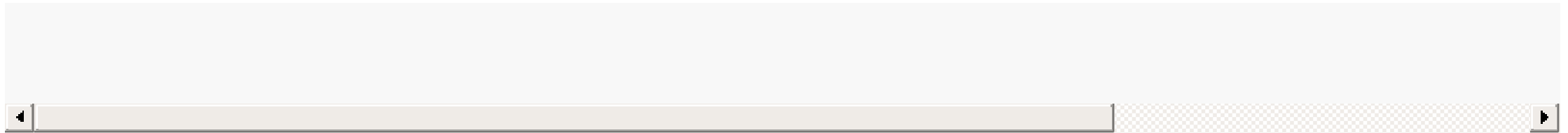
[Medical Provider Dataset](#)

What's exploratory  
data analysis (ESDA)?

# Making lots of maps!



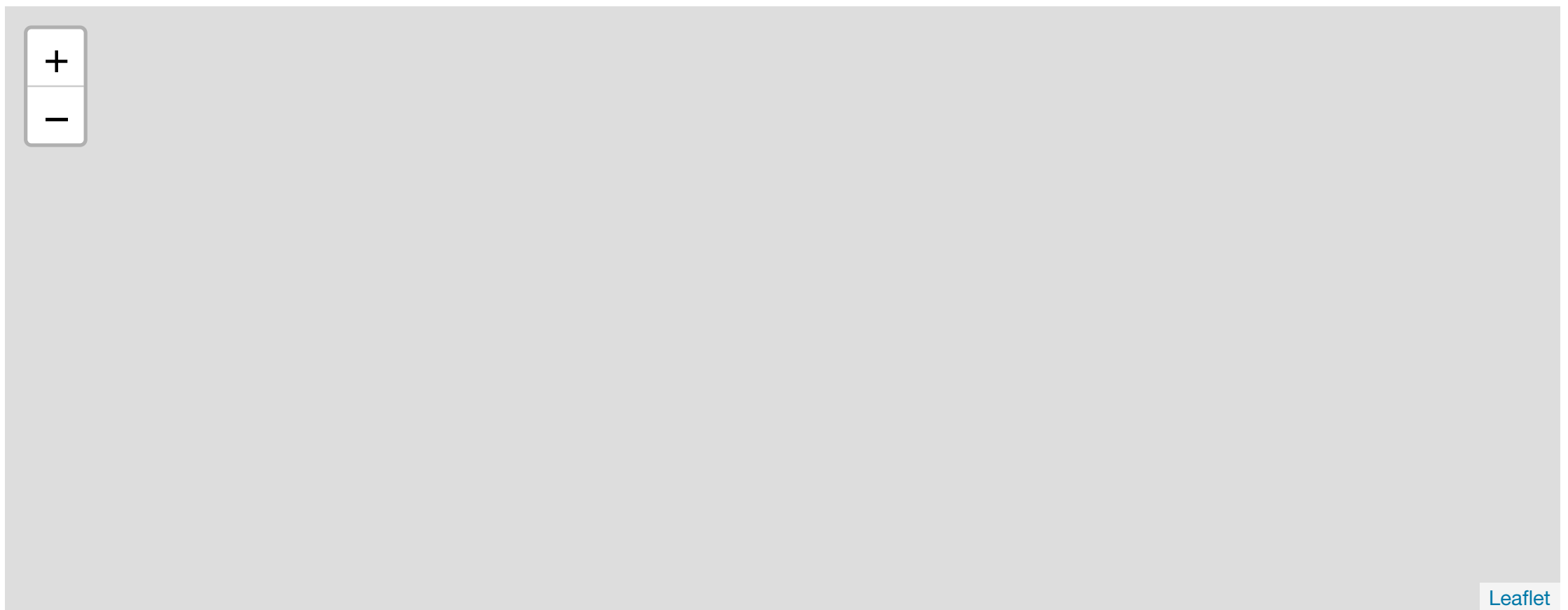
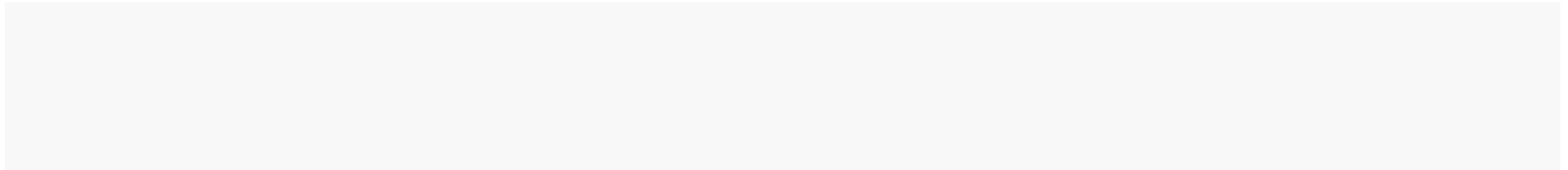
# Data: crimes in DC in the last 30 days



# Look at observations

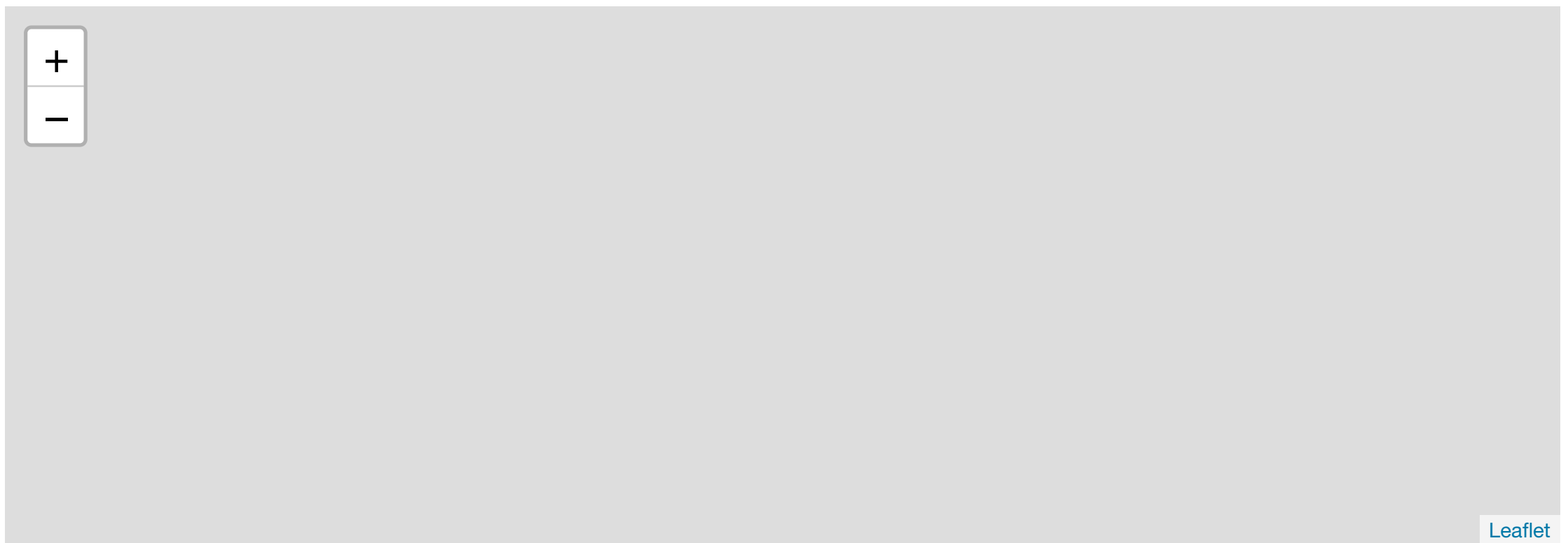
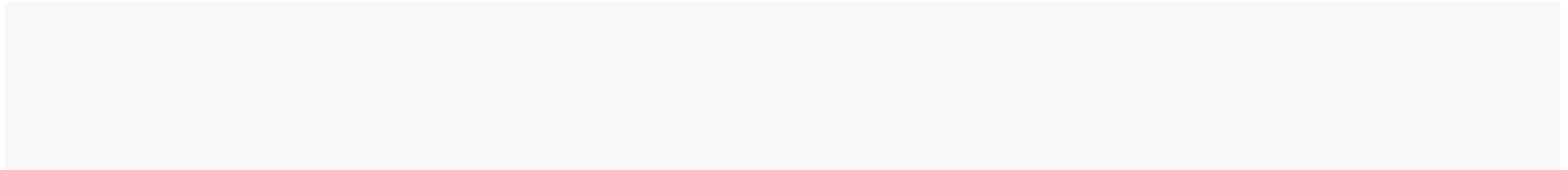


# Map it!

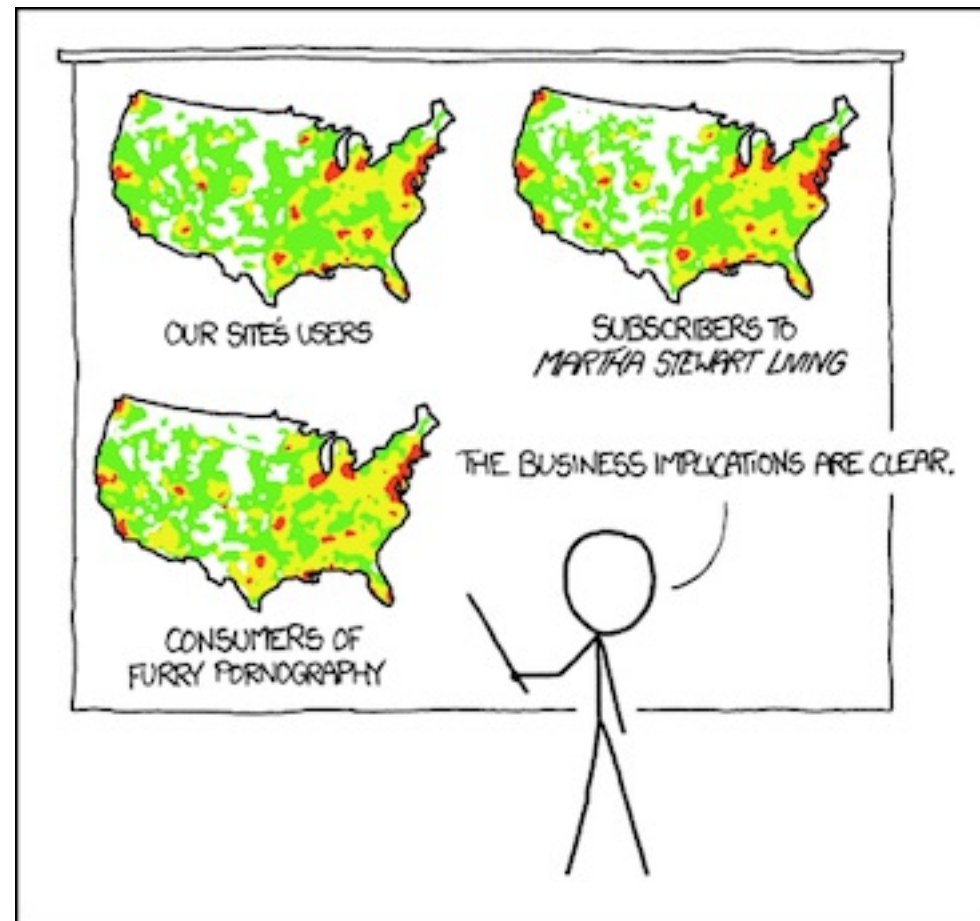




# Change into a heatmap



# The problem with heat maps



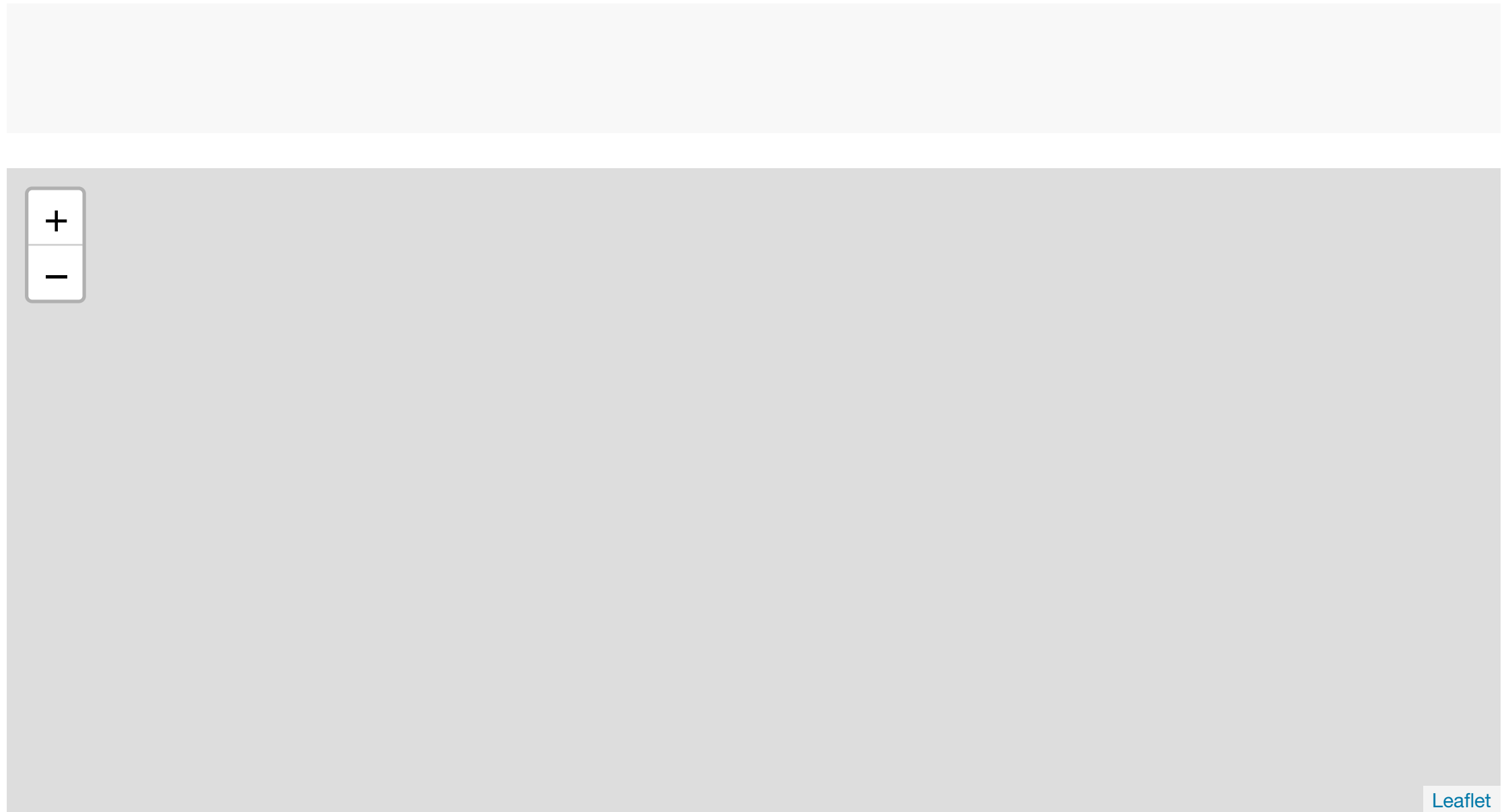
PET PEEVE #208:  
GEOGRAPHIC PROFILE MAPS WHICH ARE  
BASICALLY JUST POPULATION MAPS

I want to adjust for  
population

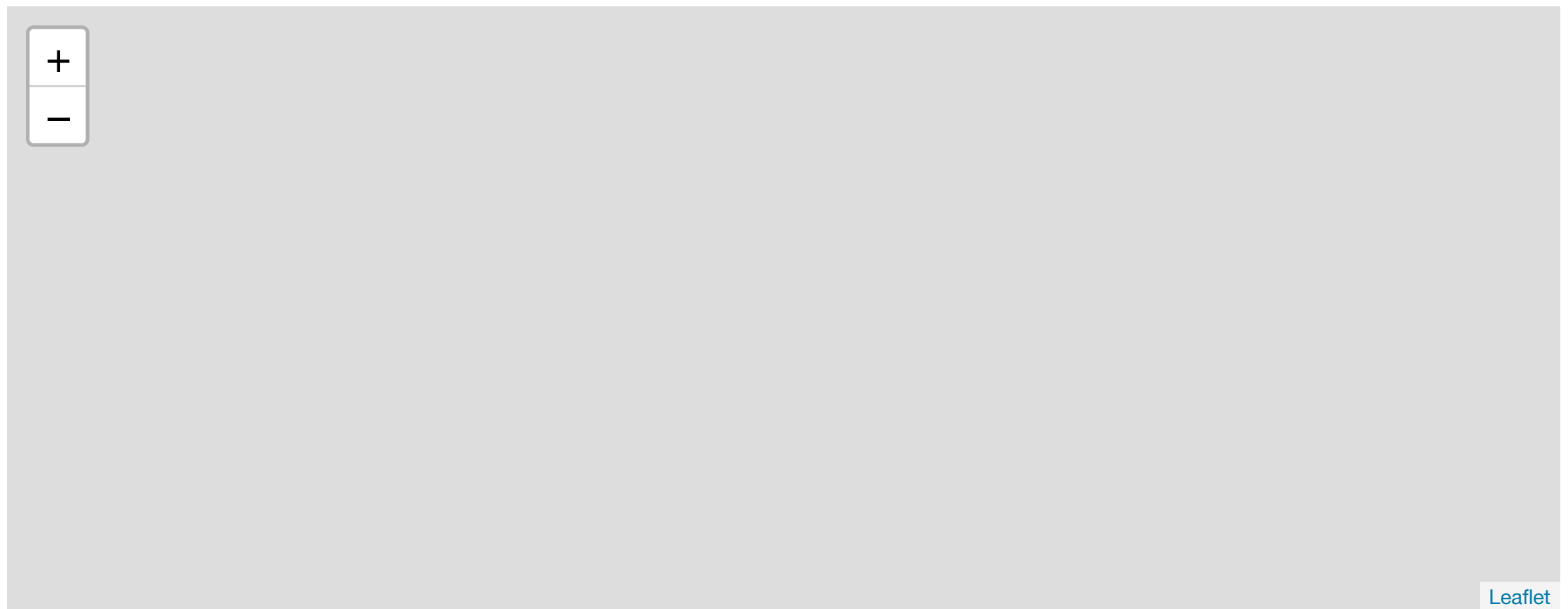
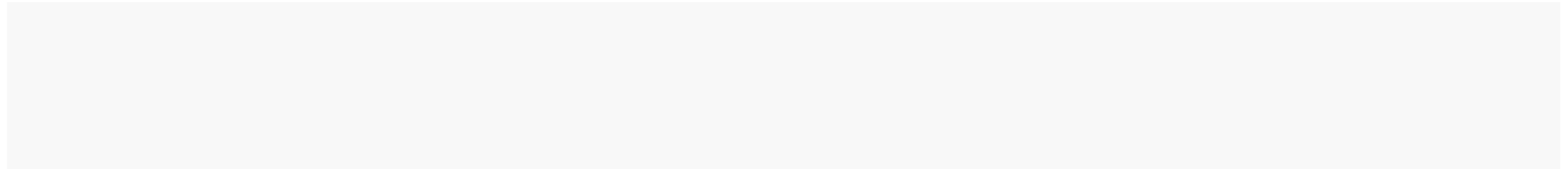
I want to adjust for  
population

But, population data is collected in  
specific geographic units

# Census tract polygons!



# Add crimes as a layer



How many crimes happened  
in each tract?

How many crimes happened  
in each tract?

Spatial aggregation

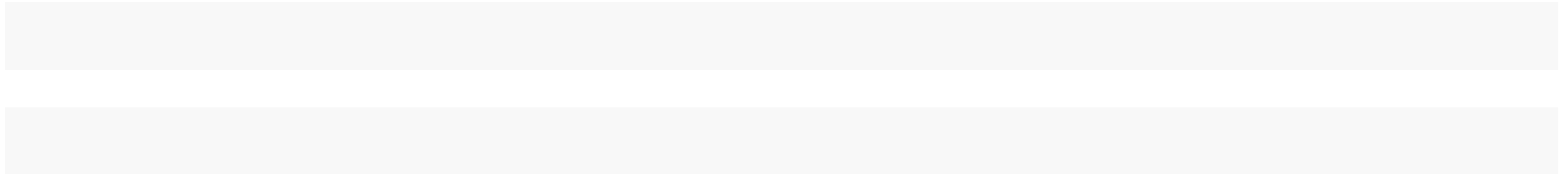


How many crimes happened  
in each tract?

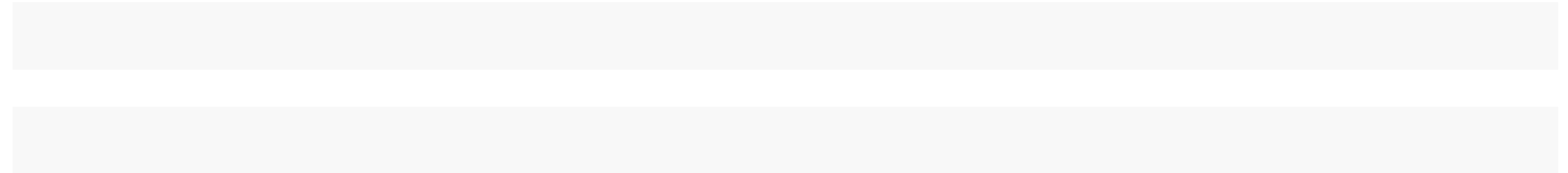
Spatial aggregation

Point-in-polygon analysis

# Count points in polygons



# Count points in polygons



What's the error???

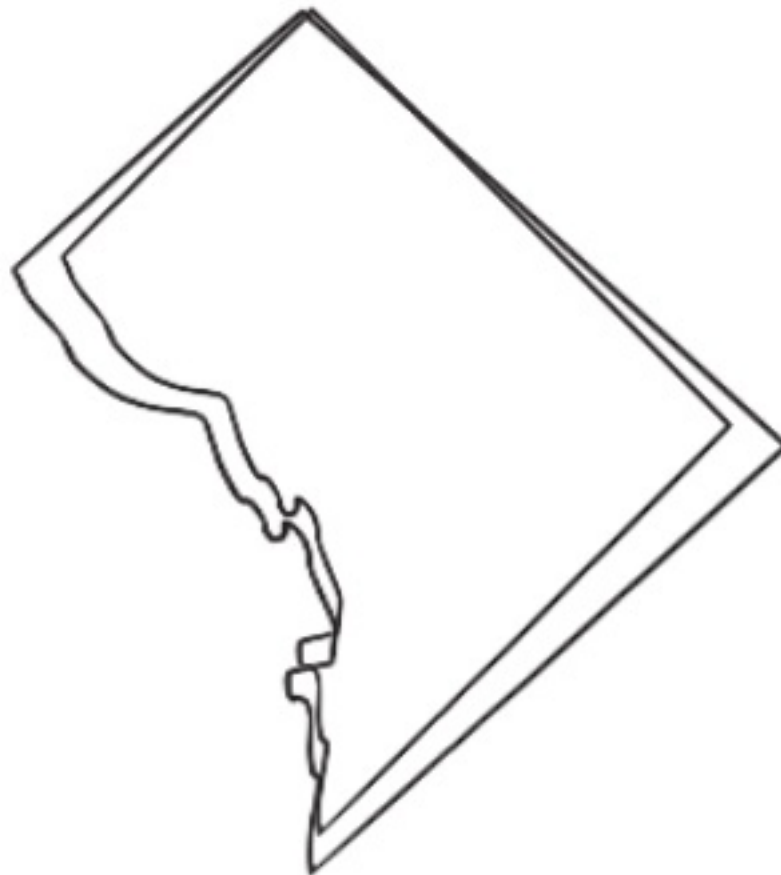
Warning: worst part of  
working with spatial data  
ahead

# Projections!



Dealing with projections =  
spatial data munging

# What's actually going on



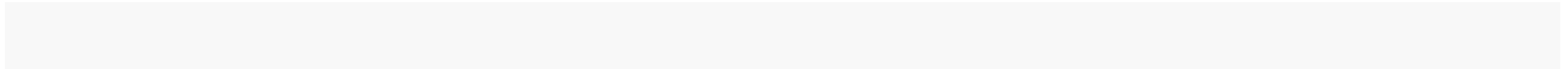
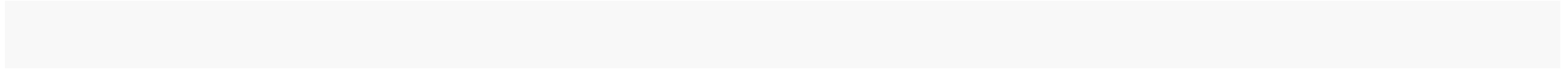
# Check the projections!



# Check the projections!



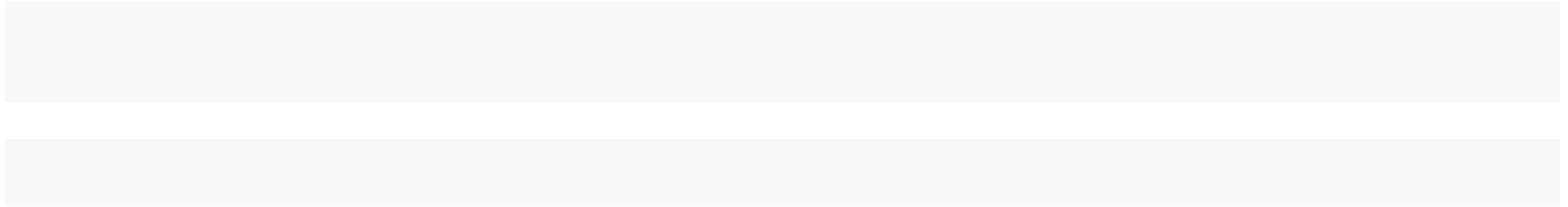
# Check the projections!



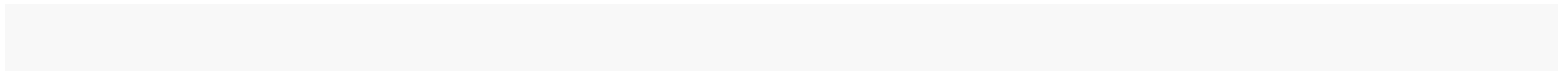
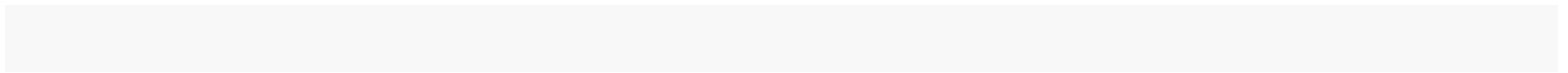
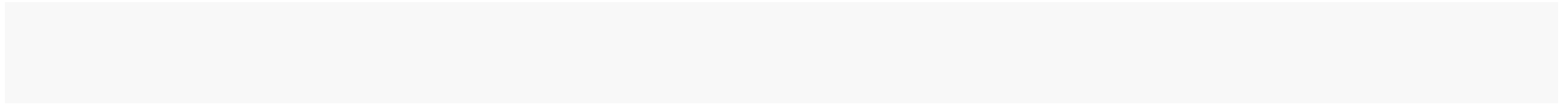
# Make them the same



# Make them the same



# Make them the same

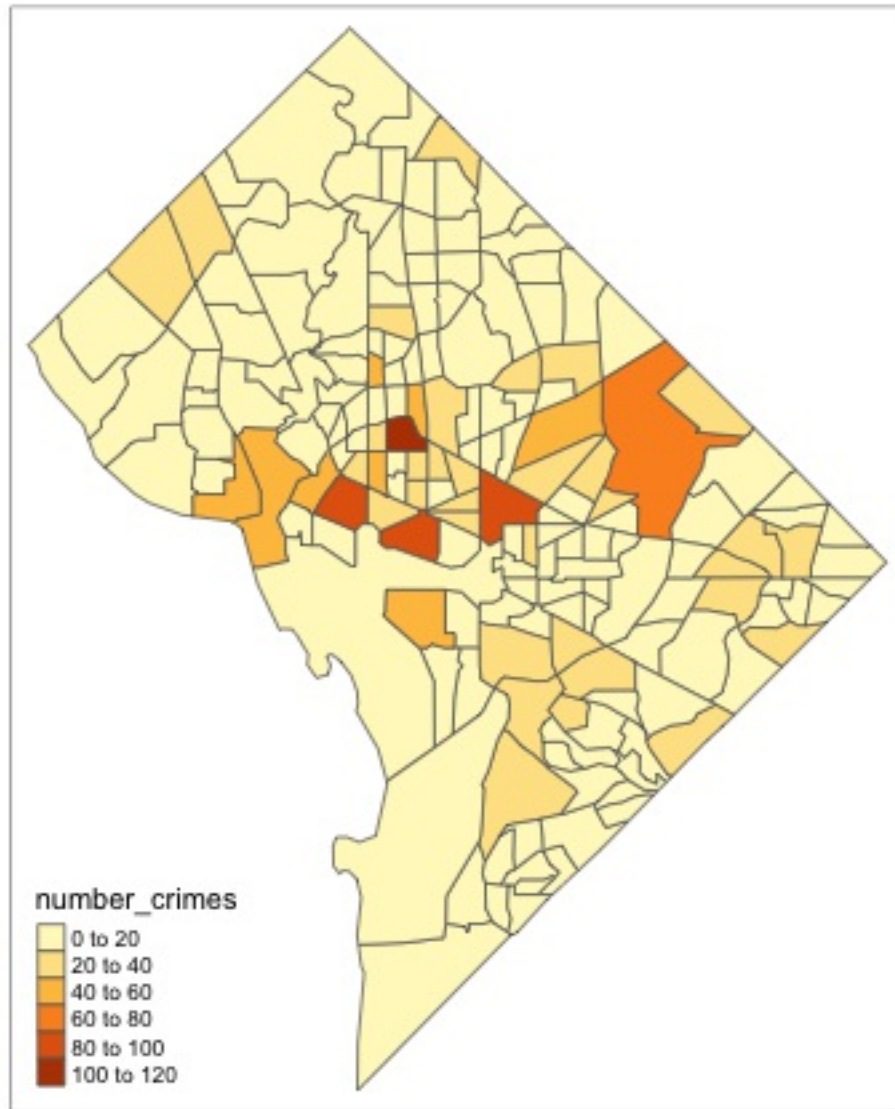


# Try the spatial join again



It worked!

# Map crime counts





This map makes a  
cartographer sad :(

This map makes a  
cartographer sad :(

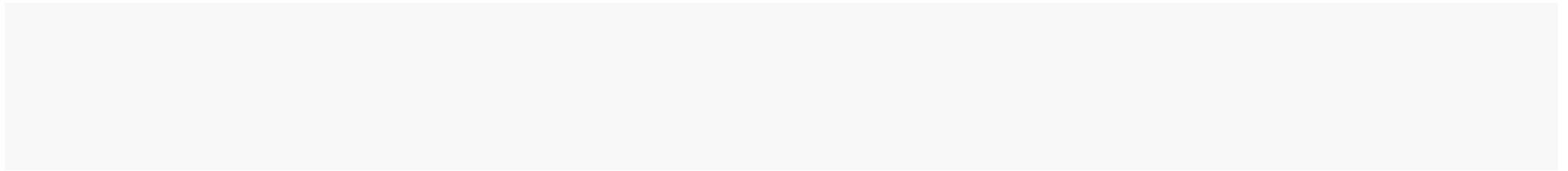
We're mapping a spatially extensive  
variable

This map makes a  
cartographer sad :(

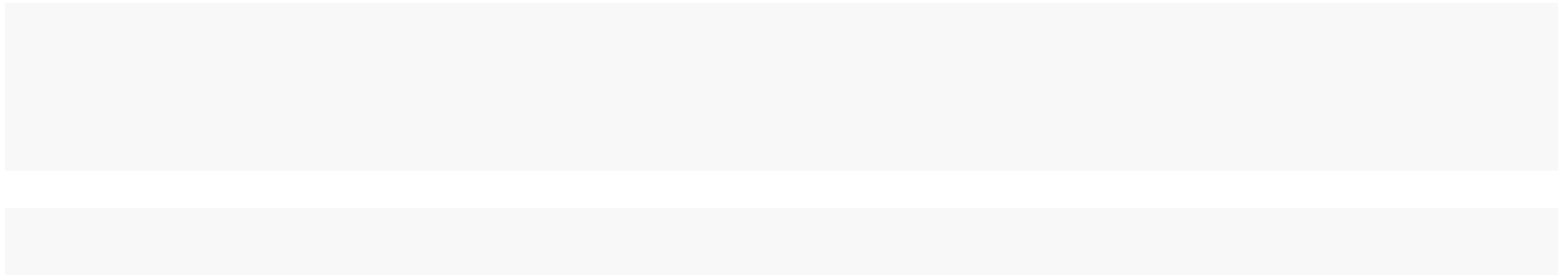
We're mapping a spatially extensive  
variable

Aka the count is related to the area  
of the tract

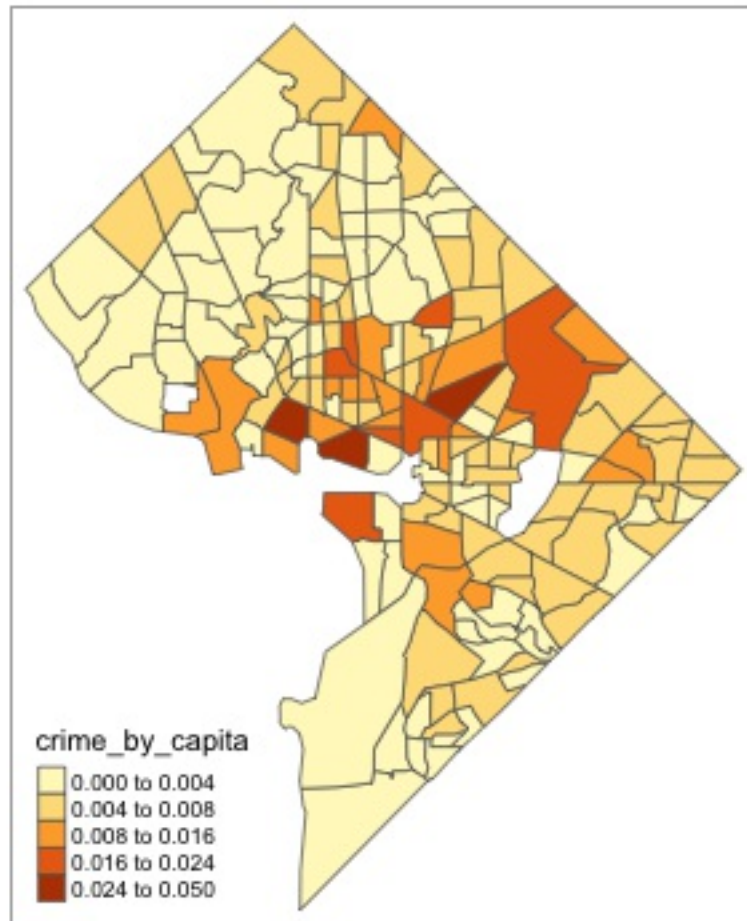
# Fix this by dividing by population



# Fix this by dividing by population

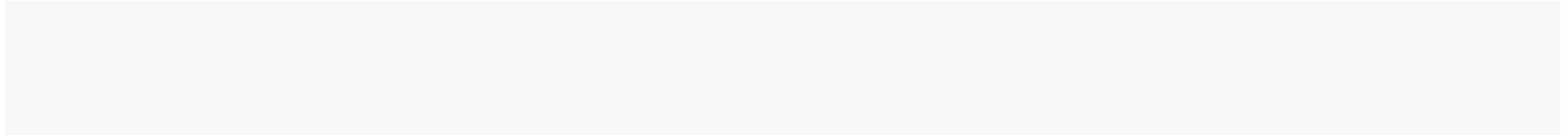


# Map crime rates per capita



Are crime rates and median  
income related?

# Get median income by tract

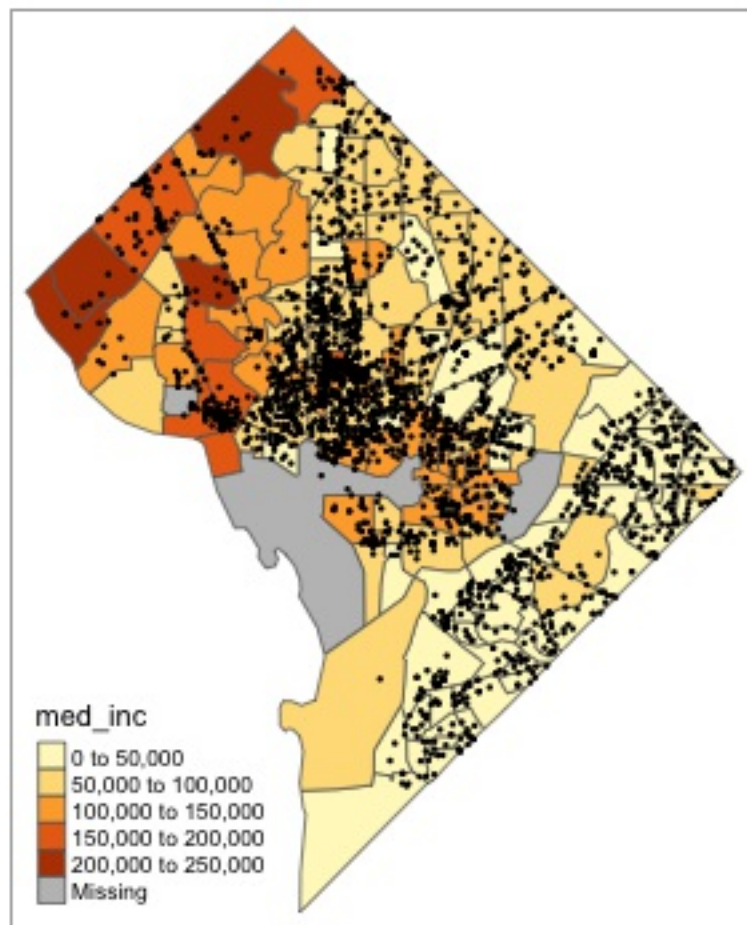




# Get median income by tract



# Map crimes and median income



I'll end it here, but...

There might be other local  
demographic & socioeconomic  
factors that affect crime rates

(spatial regression, modeling)

Crime in one area may be linked to  
crime in another area  
(spatial autocorrelation)

I might be interested in spots where  
crime is abnormally high or low  
(spatial clustering)

# Thanks!