### CSSS508, Week 9

Mapping

Chuck Lanfear

May 27, 2020

Updated: May 25, 2020



#### Today

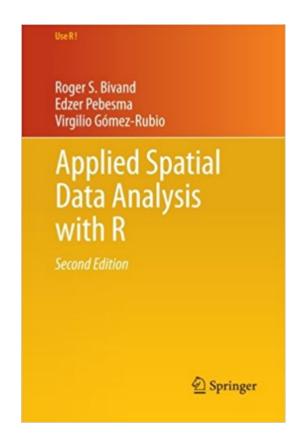
#### Basic Mapping in ggplot2

- Mapping with raw ggplot2 using coordinates
- ggmap for mashing up maps with ggplot2
- Labeling points and using ggrepel to avoid overlaps

#### Advanced Mapping

- sf: <u>Simple Features</u> geometry for R
- tidycensus for obtaining Census Bureau data

## Mapping in R: A quick plug



If you are interested in mapping, GIS, and geospatial analysis in R, acquire this book.

RSpatial.org is a great resource as well.

You may also consider taking Jon Wakefield's **CSSS 554: Statistical Methods for Spatial Data**, however it is challenging and focuses more heavily on statistics than mapping.

<u>CSDE offers workshops</u> using <u>QGIS</u> and/or ArcGIS. I recommend QGIS because it is free software with an extensive feature set and large user community.

## Basic Mapping

ggplot2 and ggmap

·UW CS&SS•

#### One Day of SPD Incidents

In Week 5, we looked at types of incidents the Seattle Police Department responded to in a single day. Now, we'll look at where those were.

library(tidyverse)

spd\_raw <- read\_csv("https://clanfear.github.io/CSSS508/Seattle\_Polic</pre>

# Taking a glimpse()

#### glimpse(spd\_raw)

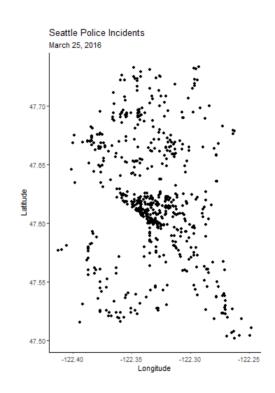
```
## Rows: 706
## Columns: 19
## $ `CAD CDW ID`
                                   <dbl> 1701856, 1701857, 1701853, ...
## $ `CAD Event Number`
                                   <dbl> 16000104006, 16000103970, 1...
## $ `General Offense Number`
                                   <dbl> 2016104006, 2016103970, 201...
                                   <chr> "063", "064", "161", "245",...
## $ `Event Clearance Code`
## $ `Event Clearance Description` <chr> "THEFT - CAR PROWL", "SHOPL...
                                   <chr> "CAR PROWL", "THEFT", "TRES...
## $ `Event Clearance SubGroup`
                                   <chr> "CAR PROWL", "SHOPLIFTING",...
## $ `Event Clearance Group`
                                   <chr> "03/25/2016 11:58:30 PM", "...
## $ `Event Clearance Date`
## $ `Hundred Block Location`
                                   <chr> "S KING ST / 8 AV S", "92XX...
                                   <chr> "K", "S", "D", "M", "M", "B...
## $ `District/Sector`
## $ `Zone/Beat`
                                   <chr> "K3", "S3", "D2", "M1", "M3...
## $ `Census Tract`
                                   <dbl> 9100.102, 11800.602, 7200.1...
## $ Longitude
                                   <dbl> -122.3225, -122.2680, -122....
## $ Latitude
                                   <dbl> 47.59835, 47.51985, 47.6142...
## $ `Incident Location`
                                   <chr> "(47.598347, -122.32245)", ...
## $ `Initial Type Description`
                                   <chr> "THEFT (DOES NOT INCLUDE SH...
## $ `Initial Type Subgroup`
                                   <chr> "OTHER PROPERTY", "SHOPLIFT...
                                   <chr> "THEFT", "THEFT", "TRESPASS...
## $ `Initial Type Group`
## $ `At Scene Time`
                                   <chr> "03/25/2016 10:25:51 PM", "...
```

# x,y as Coordinates

Coordinates, such as longitude and latitude, can be provided in aes() as x and y values.

This is ideal when you don't need to place points over some map for reference.

Sometimes, however, we want to plot these points over existing maps.



ggmap

LUW CS&SS-

### ggmap

ggmap is a package that works with ggplot2 to plot spatial data directly on map images downloaded from Google Maps<sup>1</sup>, OpenStreetMap, and Stamen Maps (good artistic/minimal options).

What this package does for you:

- 1. Queries servers for a map (get\_map()) at the location and scale you want
- 2. Plots the **raster** (bitmap) image as a ggplot object
- 3. Lets you add more ggplot layers like points, 2D density plots, text annotations
- 4. Additional functions for interacting with Google Maps (e.g. getting distances by bike)

[1] Requires an API Key now.

#### Installation

We can install ggmap like other packages:

```
install.packages("ggmap")
```

Because the map APIs it uses change frequently, sometimes you may need to get a newer development version of ggmap from the author's GitHub. This can be done using the remotes package.

```
if(!requireNamespace("remotes")) install.packages("remotes")
remotes::install_github("dkahle/ggmap", ref = "tidyup")
```

Note, this may require compilation on your computer.

```
library(ggmap)
```

# Quick Maps with qmplot()

qmplot will automatically set the map region based on your data:

All I provided was numeric latitude and longitude, and it placed the data points correctly on a raster map of Seattle.

I() is used here to specify *set* (constant) rather than *mapped* values.



# get\_map()

Both qmplot() and qmap() are wrappers for a function called get\_map() that retrieves a base map layer. Some options:

- location= search query or numeric vector of longitude and latitude
- zoom= a zoom level (3 = continent, 10 = city, 21 = building)
- source=
  - "google": Google Maps for general purpose maps<sup>1</sup>
  - "osm": OpenStreetMaps, general purpose but open access
  - "stamen": Aesthetically pleasing alternatives based on OpenStreetMaps
- maptype=
  - Google types: "terrain", "terrain-background", "satellite", "roadmap", "hybrid"
  - Stamen types: "watercolor", "toner", "toner-background",
    - "toner-lite"
- color= "color" or "bw"

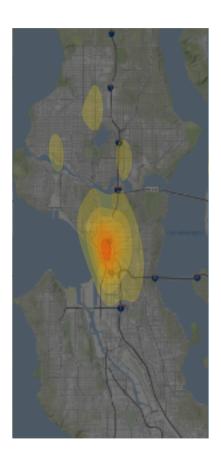
[1] Requires API key!

### Adding Density Layers

Call qmplot() with no geom(), and then add density layers:

```
qmplot(data = spd_raw, geom = "blank",
    x = Longitude, y = Latitude,
    maptype = "toner-lite",
    darken = 0.5) +
    stat_density_2d(
        aes(fill = stat(level)),
            geom = "polygon",
            alpha = .2, color = NA) +
    scale_fill_gradient2(
        "Incident\nConcentration",
            low = "white",
            mid = "yellow",
            high = "red") +
    theme(legend.position = "bottom")
```

stat(level) indicates we want
fill= to be based on level values
calculated by the layer.





## Labeling Points

Let's label the assaults and robberies specifically in downtown:

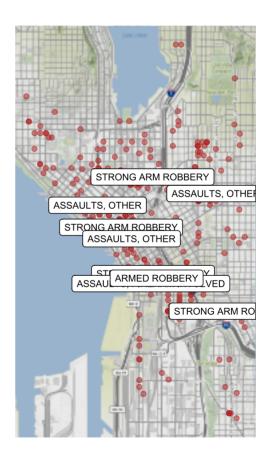
First filter to downtown based on values "eyeballed" from our earlier map:

Then make a dataframe of just assaults and robberies:

# Plotting with Labels

Now let's plot the events and label them with geom\_label() (geom\_text() without background or border):

Placing the arguments for color= and alpha= inside I() prevents them from also applying to the labels. We would get transparent red labels otherwise!



### ggrepel

You can also try

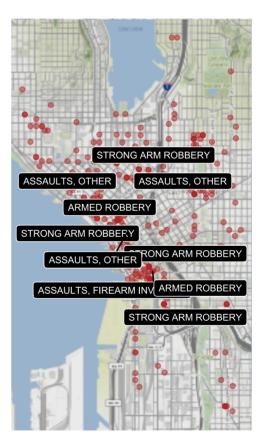
geom\_label\_repel() or

geom\_text\_repel() in the

ggrepel package to fix or reduce

overlaps (total space is limited here):

```
library(ggrepel)
qmplot(data =
          downtown,
          x = Longitude,
          y = Latitude,
          maptype = "toner-lite",
          color = I("firebrick"),
          alpha = I(0.5)) +
          geom_label_repel(
          data = assaults,
          aes(label = assault_label),
          fill = "black",
          color = "white",
          segment.color = "black",
          size=2.5)
```



## Advanced Mapping

GIS and R with sf

-UW CS&SS-

# Terminology

- Simple Features (sf)
- Coordinate Reference System (CRS)
- Shapefile

### sf

Until recently, the main way to work with geospatial data in R was through the sp package. sp works well but does not store data the same way as most GIS packages and can be bulky and complicated.

The more recent sf package implements the GIS standard of <u>Simple</u> <u>Features</u> in R.

sf is also integrated into the tidyverse: e.g. geom\_sf() in ggplot2.

The package is somewhat new but is expected to *replace* sp eventually. The principle authors and contributors to sf are the same authors as sp but with new developers from the tidyverse as well.

Because sf is the new standard, we will focus on it today.

#### library(sf)

## Simple Features

A <u>Simple Feature</u> is a single observation with some defined geospatial location(s). Features are stored in special data frames (class sf) with two properties:

- **Geometry**: Properties describing a location (usually on Earth).
  - Usually 2 dimensions, but support for up to 4.
  - Stored in a single reserved *list-column* (geom, of class sfc).<sup>1</sup>
  - Contain a defined coordinate reference system.
- Attributes: Characteristics of the location (such as population).
  - These are non-spatial measures that describe a feature.
  - Standard data frame columns.

[1] A list-column is the same length as all other columns in the data, but each element contains *sub-elements* (class sfg) with all the geometrical components.

*List-columns* require special functions to manipulate, *including removing them*.

#### Coordinate Reference Systems

**Coordinate reference systems (CRS)** specify what location on Earth geometry coordinates are *relative to* (e.g. what location is (0,0) when plotting).

The most commonly used is <u>WGS84</u>, the standard for Google Earth, the Department of Defense, and GPS satellites.

There are two common ways to define a CRS in sf:

- EPSG codes (epsg in R)
  - Numeric codes which refer to a predefined CRS
  - Example: WGS84 is 4326
- **PROJ.4 strings** (proj4string in R)
  - Text strings of parameters that *define a CRS*
  - Example: NAD83(NSRS2007) / Washington North

# Shapefiles

Geospatial data is typically stored in **shapefiles** which store geometric data as **vectors** with associated attributes (variables)

Shapefiles actually consist of multiple individual files. There are usually at least three (but up to 10+):

- .shp: The feature geometries
- .shx: Shape positional index
- .dbf: Attributes describing features<sup>1</sup>

Often there will also be a .prj file defining the coordinate system.

[2] This is just a dBase IV file which is an ancient and common database storage file format.



·UW CS&SS

### Selected sf Functions

sf is a huge, feature-rich package. Here is a sample of useful functions:

- st\_read(), st\_write(): Read and write shapefiles.
- geom\_sf(): ggplot() layer for sf objects.
- st\_as\_sf(): Convert a data frame into an sf object.
- st\_join(): Join data by spatial relationship.
- st\_transform(): Convert between CRS.
- st\_drop\_geometry(): Remove geometry from a sf data frame.
- st\_relate(): Compute relationships between geometries (like neighbor matrices).
- st\_interpolate\_aw(): Areal-weighted interpolation of polygons.

### Loading Data

We will work with the voting data from Homework 5. You can obtain a shape file of King County voting precincts from the <u>county GIS data portal</u>.

We can load the file using st\_read().

```
## Reading layer `votdst' from data source `C:\Users\cclan\OneDrive\GitHub\CS
## Simple feature collection with 2592 features and 5 fields
## geometry type: MULTIPOLYGON
```

## dimension: XY

If following along, click here to download a zip of the shapefile.

### Voting Data: Processing

```
precincts votes sf <-
  read csv("./data/king county elections 2016.txt") %>%
  filter(Race=="US President & Vice President",
         str detect(Precinct, "SEA ")) %>%
  select(Precinct, CounterType, SumOfCount) %>%
  group bv(Precinct) %>%
  filter(CounterType %in%
           c("Donald J. Trump & Michael R. Pence".
             "Hillary Clinton & Tim Kaine",
             "Registered Voters".
             "Times Counted")) %>%
 mutate(CounterType =
           recode(CounterType,
                  `Donald J. Trump & Michael R. Pence` = "Trump",
                  `Hillary Clinton & Tim Kaine` = "Clinton",
                  `Registered Voters`="RegisteredVoters",
                  `Times Counted` = "TotalVotes")) %>%
  spread(CounterType, SumOfCount) %>%
 mutate(P Dem = Clinton / TotalVotes,
         P Rep = Trump / TotalVotes,
        Turnout = TotalVotes / RegisteredVoters) %>%
  select(Precinct, P Dem, P Rep, Turnout) %>%
  filter(!is.na(P Dem)) %>%
 left join(precinct shape) %>%
  st as sf() # Makes sure resulting object is an sf dataframe
```

# Taking a glimpse()

#### glimpse(precincts\_votes\_sf)

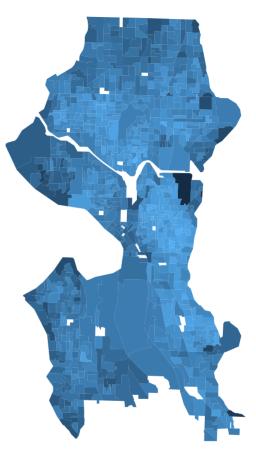
Notice the geometry column and its unusual class: MULTIPOLYGON

A single observation (row) has a geometry which may consist of multiple polygons.

### Voting Map

We can plot sf geometry using geom\_sf().

- fill=P\_Dem maps color inside precincts to P\_Dem.
- size=NA removes precinct outlines.
- theme\_void() removes axes and background.







-UW CS&SS

## tidycensus

tidycensus can be used to search the American Community Survey (ACS) and Dicennial Census for variables, then download them and automatically format them as tidy dataframes.

#### These dataframes include geographical boundaries such as tracts!

This package utilizes the Census API, so you will need to obtain a <u>Census API</u> <u>key</u>. I talk more about APIs in the Social Media Data and Text Mining lecture.

**Application Program Interface (API)**: A type of computer interface that exists as the "native" method of communication between computers, often via http (usable via httr package).

- R packages that interface with websites and databases typically use APIs.
- APIs make accessing data easy while allowing websites to control access.

See the developer's GitHub page for detailed instructions.

30 / 47

-UW CS&SS

## Key tidycensus Functions

- census\_api\_key() Install a census api key.
  - Note you will need to run this prior to using any tidycensus functions.
- load\_variables() Load searchable variable lists.
  - year =: Sets census year or endyear of 5-year ACS
  - o dataset =: Sets dataset (see ?load\_variables)
- get\_decennial() Load Census variables and geographical boundaries.
  - variables =: Provide vector of variable IDs
  - geography =: Sets unit of analysis (e.g. state, tract, block)
  - year =: Census year (1990, 2000, or 2010)
  - geometry = TRUE: Returns sf geometry
- get\_acs() Load ACS variables and boundaries.

## Searching for Variables

```
library(tidycensus)
# census_api_key("PUT YOUR KEY HERE", install=TRUE)
acs_2015_vars <- load_variables(2015, "acs5")
acs_2015_vars[10:18,] %>% print()
```

```
## # A tibble: 9 x 3
##
               label
                                                     concept
    name
##
    <chr>
               <chr>
                                                     <chr>
## 1 B01001 008 Estimate!!Total!!Male!!20 years
                                                SEX BY AGE
## 2 B01001 009 Estimate!!Total!!Male!!21 years
                                                SEX BY AGE
    B01001 010 Estimate!!Total!!Male!!22 to 24 years SEX BY AGE
## 3
## 4 B01001 011 Estimate!!Total!!Male!!25 to 29 years SEX BY AGE
## 5 B01001 012 Estimate!!Total!!Male!!30 to 34 years SEX BY AGE
## 6 B01001 013 Estimate!!Total!!Male!!35 to 39 years SEX BY AGE
## 7 B01001 014 Estimate!!Total!!Male!!40 to 44 years SEX BY AGE
## 8 B01001 015 Estimate!!Total!!Male!!45 to 49 years SEX BY AGE
## 9 B01001 016 Estimate!!Total!!Male!!50 to 54 years SEX BY AGE
```

### Getting Data

What do these look like?

```
glimpse(king_county)
```

Variable names are in variables and values are in estimate.

#### Processing Data

To get one row per precinct, we want to spread()¹ key (variable) and value (estimate) columns into separate columns for each variable. Then we mutate() Any Black into a proportion measure.

[1] pivot\_wider() is currently broken for sf objects, so we must spread. I remove the *margin of error* (moe) column before spreading.

#### The Data

#### head(king\_county, 10)

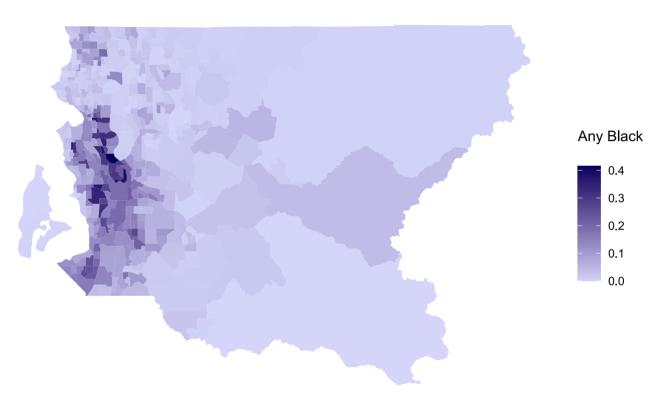
```
## Simple feature collection with 10 features and 4 fields
## geometry type: MULTIPOLYGON
## dimension:
                  XY
                  xmin: -122.3799 ymin: 47.69503 xmax: -122.2551 ymax: 47.73416
## bbox:
## CRS:
                  4269
## # A tibble: 10 x 5
## # Groups: GEOID [10]
##
     GEOID NAME
                                    geometry `Total Populati~ `Any Black`
                         <MULTIPOLYGON [°]>
                                                        <dbl>
     <chr> <chr>
                                                                    <dbl>
##
## 1 53033~ Cens~ (((-122.2965 47.73378, -~
                                                         7985
                                                                  0.196
## 2 53033~ Cens~ (((-122.3236 47.72309, -~
                                                         8498
                                                                  0.0938
   3 53033~ Cens~ (((-122.3452 47.73413, -~
                                                         2991
                                                                  0.0475
##
   4 53033~ Cens~ (((-122.3609 47.73414, -~
                                                                  0.128
                                                         6560
##
## 5 53033~ Cens~ (((-122.3609 47.72507, -~
                                                         5419
                                                                  0.0627
  6 53033~ Cens~ (((-122.3799 47.71327, -~
##
                                                         3143
                                                                  0.00700
## 7 53033~ Cens~ (((-122.345 47.72868, -1~
                                                                  0.100
                                                         8124
## 8 53033~ Cens~ (((-122.3127 47.71944, -~
                                                         5217
                                                                  0.141
## 9 53033~ Cens~ (((-122.2915 47.7155, -1~
                                                         2710
                                                                  0.0365
## 10 53033~ Cens~ (((-122.285 47.7192, -12~
                                                         2009
                                                                  0.0164
```

### Mapping Code

#### New functions:

- geom\_sf() draws Simple Features coordinate data.
- coord\_sf() is used here with these arguments:
  - o crs: Modifies the coordinate reference system (CRS); WGS84 is possibly the most commonly used CRS.
  - datum=NA: Removes graticule lines, which are geographical lines such as meridians and parallels.

#### Proportion Any Black



#### Removing Water

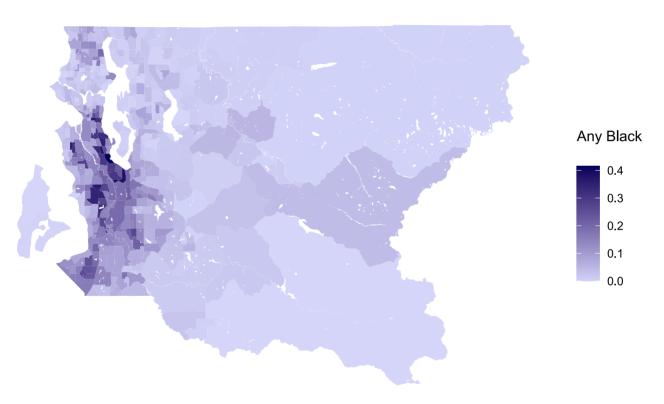
With a simple function and boundaries of water bodies in King County, we can replace water with empty space.

```
st_erase <- function(x, y) {
   st_difference(x, st_make_valid(st_union(st_combine(y))))
}
kc_water <- tigris::area_water("WA", "King", class = "sf")
kc_nowater <- king_county %>%
   st_erase(kc_water)
```

- st\_combine() merges all geometries into one
- st\_union() resolves internal boundaries
- st\_difference() subtracts y geometry from x
- lwgeom::st\_make\_valid() fixes geometry errors from subtraction
- area\_water() obtains sf geometry of water bodies.

Then we can reproduce the same plot using kc\_nowater...

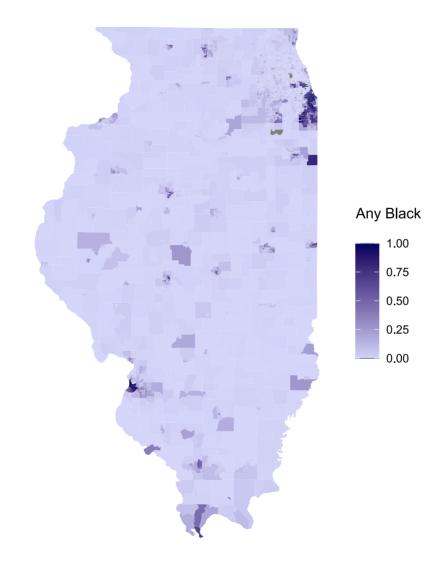
#### Proportion Any Black



### State Example Data

Let's do this again, but for the entire state of Illinois.

### State Example Plot



# Multiple geom\_sf Layers

As with other ggplot2 layers, we can add additional geom\_sf() layers using new data.

This is useful for...

- Adding points
  - Cities in states
  - Crimes in police beats
- Adding lines
  - Street grids over tracts
- Adding outlines or highlights
  - Elevation contours
  - Showing urban boundaries

#### Add Urban Outlines

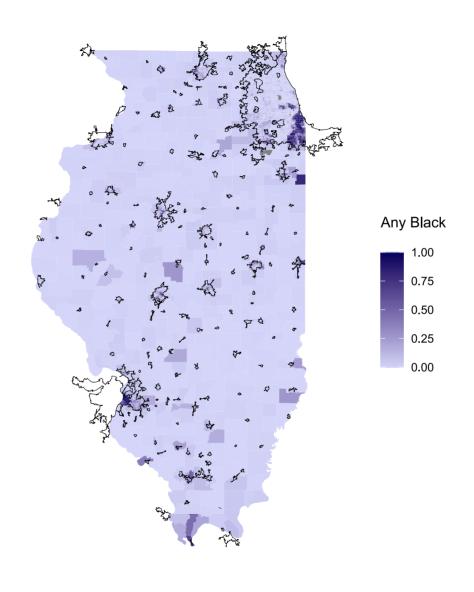
We can use tigris to download urban boundaries and add them to our prior map.

```
urbans <- tigris::urban areas(cb = TRUE, class = "sf")</pre>
 glimpse(urbans)
## Rows: 3,601
## Columns: 9
## $ UACE10
              <chr> "88732", "01819", "22366", "93322", "02548", "...
## $ AFFGEOID10 <chr> "400C100US88732", "400C100US01819", "400C100US...
## $ GEOID10
              <chr> "88732", "01819", "22366", "93322", "02548", "...
              <chr> "Tucson, AZ", "Alturas, CA", "Davenport, IA--I...
## $ NAME10
              <chr> "75", "76", "75", "76", "76", "76", "75", "75"...
## $ LSAD10
              ## $ UATYP10
## $ ALAND10
              <dbl> 915276150, 4933312, 357345121, 45455957, 23646...
## $ AWATER10
              <dbl> 2078695, 16517, 21444164, 88872, 3913803, 1510...
              <MULTIPOLYGON [°]> MULTIPOLYGON (((-110.8135 3..., M...
## $ geometry
urban il <- urbans %>% filter(str detect(NAME10, "IL"))
```

#### With Urban Outlines

We add the urban\_il data as a new layer:

- fill=NA removes the polygon fill
- size=0.1 and color="black" give a thin outline



#### Optional Exercise

Use the HW 7 template to practice making maps of the restaurant inspection data.

If you wish to submit it for bonus points, turn it in via Canvas by 11:59 PM next Tuesday.

UW CS&SS.