Waller SISMID 2023 Reading and Mapping Shapefiles: Alcohol, Drugs, and Crime in Houston

Lance A. Waller

7/18/2023

- **What we have** An ArcGIS shapefile for census tracts in Houston.
- A "shapefile" is actually several separate files with the same file name and different extentions.
- The .dbf file is a dbase database file containing the attribute table
- The .shp file contains information on the outlines of the tracks.
- The .shx, .sbn, and .sbx files have the connective tissue to link it all together (sometims you can have a .prj paper for the map projection). The following attributes are of interest to us: Population (census), violent crimes (police), number of illegal drug arrests (police), total alcohol sales (state alcoholic beverage license data)

- **What we want**
- *Choropleth* maps of the violent crime rate and standardized log drug arrests and standardized log alcohol sales.
- Will need to calculate these variables, choose intervals and colors, and map.

- **How we get it**
- We will use the 'rgdal' package and the function 'readOGR' function.
- We will use the 'RColorBrewer' and 'classInt' libraries to choose colors and intervals.
- For this example, we will use the basic 'plot' function to make maps.

First, load the libraries. The working directory should automatically be set to the correct one if you opened the R project (`2023-SISMID-Spatial-Epi.Rproj`) through RStudio.

```
##Load libraries

library(maptools)     # loads sp library too
library(RColorBrewer) # creates nice color schemes
library(classInt)     # finds class intervals for continuous variables
library(spgwr)        # Adds the geographically weighted regression functions
library(rgdal)        # Provides tool for reading in shapefiles
library(here)         # For constructing filepaths relative to root directory
```

Now to read in the shapefile. The following five files all constitute what is a "shapefile." If you downloaded the repository from GitHub correctly, they should all be in your `data` folder.

- HoustonENAR2012final.shp
- HoustonENAR2012final.dbf

- HoustonENAR2012final.shx
- HoustonENAR2012final.sbx
- HoustonENAR2012final.sbn

```
##Read in shapefile - Houston Census Tracts
houston = readOGR(dsn = here("data"),layer = "HoustonENAR2012final")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/home/thsiao3/research/2023-SISMID-Spatial-Epi/data", layer: "HoustonENAR2012final"
## with 439 features
## It has 133 fields
```

Plotting the map

- To get the outlines simply plot our 'houston' data object.

```
plot(houston)
```

***

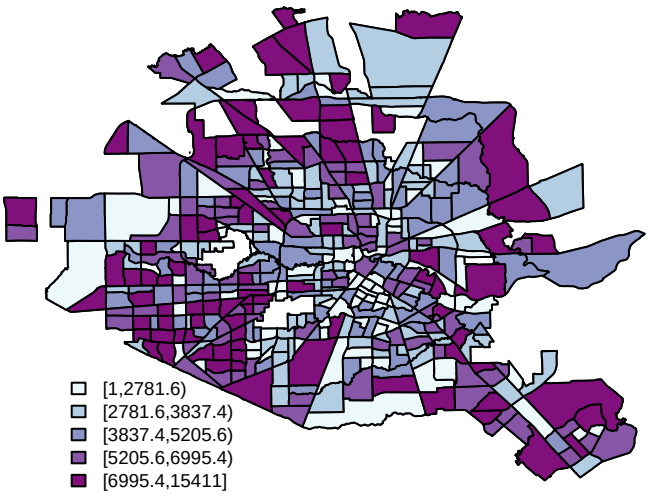
Next, we want to make choropleth maps (shading in each tract based on it attribute value).

To do this, we need to decide how many intervals (colors) we want. I like odd numbers so there is a 'middle' color, and I usually start with quintiles.

'classInt' will assign each tract to the appropriate quintile for a particular attribute.

'RColorBrewer' will assign a color scheme for the quantiles... lots of fascinating work on color choices by Cynthia Brewer, well worth reading and checking the ColorBrewer webpage (https://colorbrewer2.org/)

```
## Making choropleth maps
# Plot the outlines (we'll add color below)
plot(houston)

# Define the variable (attribute) to shade tracts by
pop2000 <- houston@data$POP2000

# Define the number of classes
nclr <- 5  # quintiles

# Use RColorBrewer to choose the colors
plotclr <- brewer.pal(nclr,"BuPu")
class <- classIntervals(pop2000, nclr, style="quantile")
colcode <- findColours(class, plotclr)

#Fill in the tracts with the colors
plot(houston, col=colcode, add=T)
#Add a title
title(main="Population 2000",
      sub="Quantile (Equal-Frequency) Class Intervals")

#Add a legend  (Coordinates are in longitude, latitude).
legend(-95.7, 29.65, legend=names(attr(colcode, "table")),
       fill=attr(colcode, "palette"), cex=0.6, bty="n")
```



OK, great, we can map the data. Let's map the main variables for our analysis.

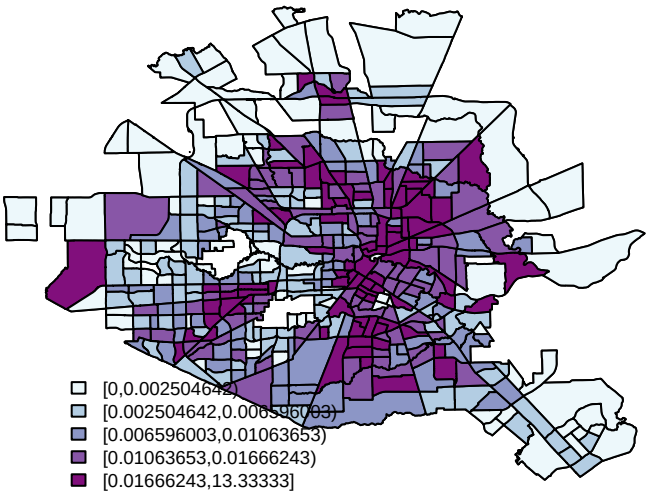
The data table has a lot of census data and various transformations of the violent crime, alcohol sales, and drug arrest data. The next section pulls the values we want. ***

```
# Outcome:  Number of violent crimes by tract
violence = houston@data$violence_2

# Divide by the 2000 population to get the rate
violence.rate = violence/houston@data$tot_pop

# Covariate 1 (log standardized total alcohol sales)
Z.log.total = houston@data$Zl_total

# Covariate 2 (log standardized illegal drug arrests)
Z.log.drug = houston@data$Zl_drug
```

Now to map the outcome. ***

```
# Plot Outcome first
plot(houston)
# Define the number of classes
nclr <- 5  # quintiles
# Use RColorBrewer to choose the colors
plotclr <- brewer.pal(nclr,"BuPu")
```

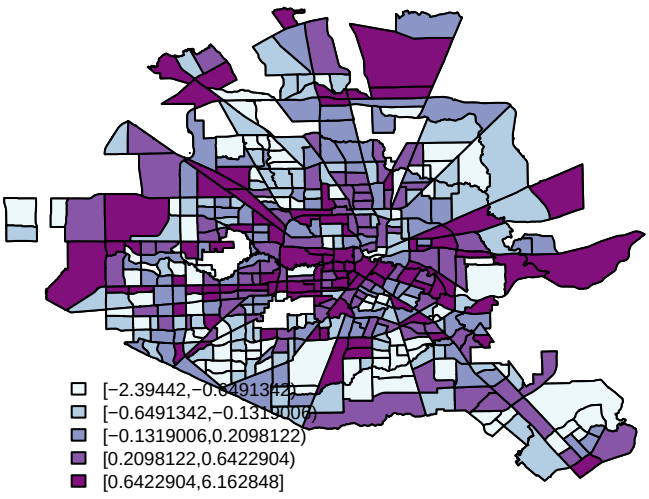
```
class <- classIntervals(violence.rate, nclr, style="quantile")
colcode <- findColours(class, plotclr)
#Fill in the tracts with the colors
plot(houston, col=colcode, add=T)
#Add a title
title(main="Violence Rate",
      sub="Quantile (Equal-Frequency) Class Intervals")
#Add a legend  (Coordinates are in longitude, latitude).
legend(-95.7, 29.65, legend=names(attr(colcode, "table")),
       fill=attr(colcode, "palette"), cex=0.6, bty="n")
```



Next, map standardized log total alcohol sales.

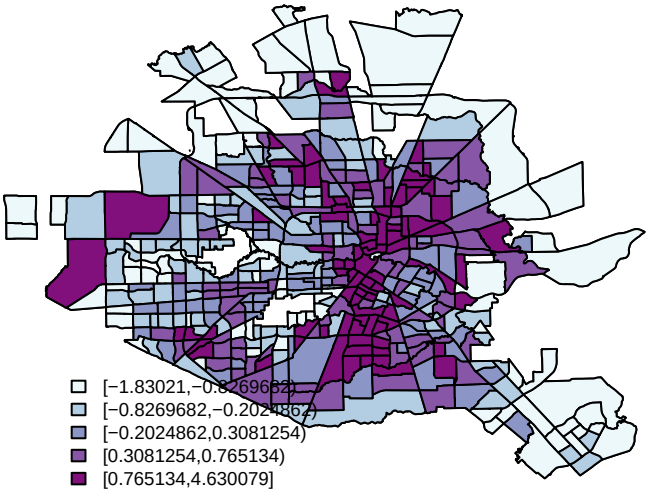
```
# Next plot stdized log total alcohol sales
plot(houston)
# Define the number of classes
nclr <- 5  # quintiles
# Use RColorBrewer to choose the colors
plotclr <- brewer.pal(nclr,"BuPu")
class <- classIntervals(Z.log.total, nclr, style="quantile")
colcode <- findColours(class, plotclr)
#Fill in the tracts with the colors
plot(houston, col=colcode, add=T)
#Add a title
title(main="Std log total sales",
      sub="Quantile (Equal-Frequency) Class Intervals")
#Add a legend  (Coordinates are in longitude, latitude).
```

```
legend(-95.7, 29.65, legend=names(attr(colcode, "table")),
       fill=attr(colcode, "palette"), cex=0.6, bty="n")
```



Finally, plot standardized log illegal arrests.

```
# Next plot stdized log illegal arrests
plot(houston)
# Define the number of classes
nclr <- 5  # quintiles
# Use RColorBrewer to choose the colors
plotclr <- brewer.pal(nclr,"BuPu")
class <- classIntervals(Z.log.drug, nclr, style="quantile")
colcode <- findColours(class, plotclr)
#Fill in the tracts with the colors
plot(houston, col=colcode, add=T)
#Add a title
title(main="Std log drug arrests",
      sub="Quantile (Equal-Frequency) Class Intervals")
#Add a legend  (Coordinates are in longitude, latitude).
legend(-95.7, 29.65, legend=names(attr(colcode, "table")),
       fill=attr(colcode, "palette"), cex=0.6, bty="n")
```



***

These three figures will match the maps in Figure 1 of:

Waller LA, Zhu L, Gotway CA, Gorman DM, and Gruenewald PJ (2007) "Quantifying geographic variations in associations between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models". *Stochastic Environmental Research and Risk Assessment.* **21**, 573-588. ***