

Waller SISMID 2024 Reading and Mapping Shapefiles: Alcohol, Drugs, and Crime in Houston

Lance A. Waller

7/15/2024

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 extensions.
 - 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 (sometimes 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 ‘sf’ package and the function ‘st_read’
 - We will use the ‘RColorBrewer’ library 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 (2024-SISMID-Spatial-Epi.Rproj) through RStudio.

```

##Load libraries
#install.packages("pacman")
pacman::p_load(sf, #replaces "maptools", "rgdal" and other deprecated packages
  tmap, #helps with plotting your map
  RColorBrewer, # creates nice color schemes
  ClassInt, # finds class intervals for continuous variables
  spgwr, # Adds the geographically weighted regression functions
  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 = st_read(dsn = here("data"), layer = "HoustonENAR2012final")

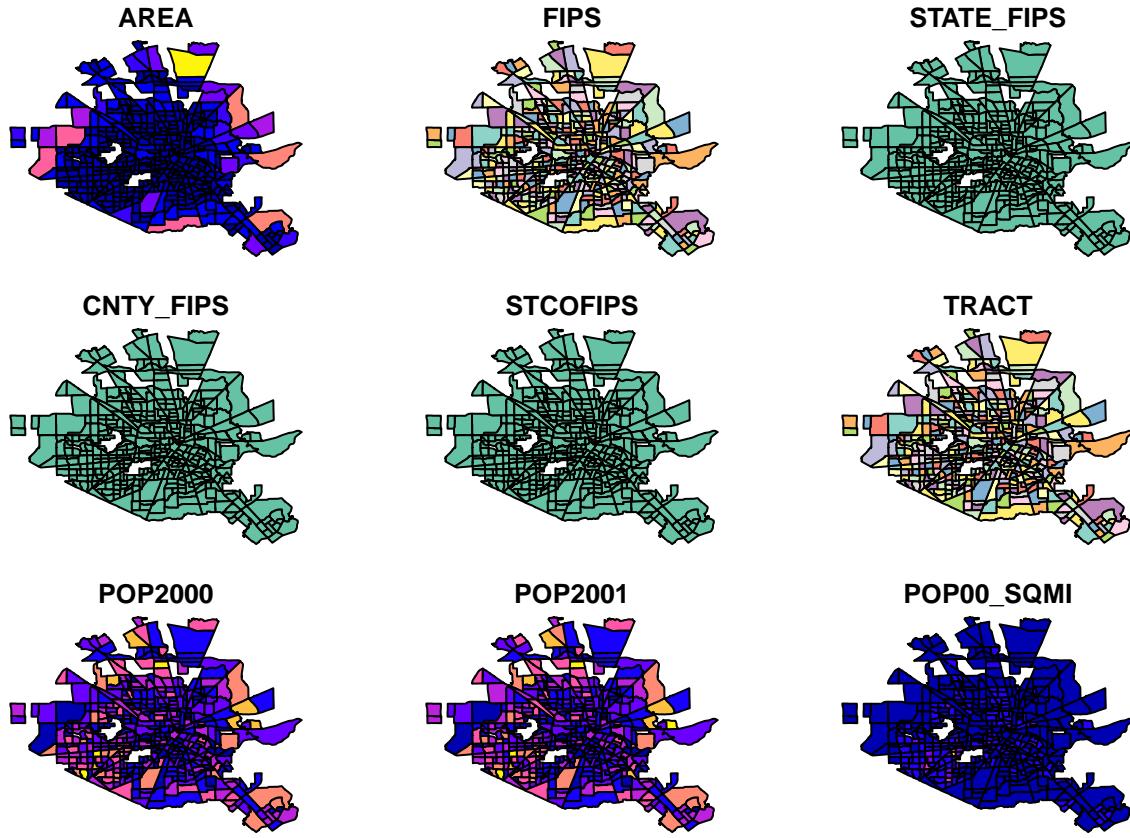
## Reading layer 'HoustonENAR2012final' from data source
##   '/Users/nmolson/Desktop/SISMID/SISMID/data' using driver 'ESRI Shapefile'
## Simple feature collection with 439 features and 133 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:  xmin: -95.75434 ymin: 29.52563 xmax: -95.06086 ymax: 30.03774
## CRS:            NA

```

Plotting the map

- To get the outlines simply plot our ‘`houston`’ data object.

```
plot(houston)
```



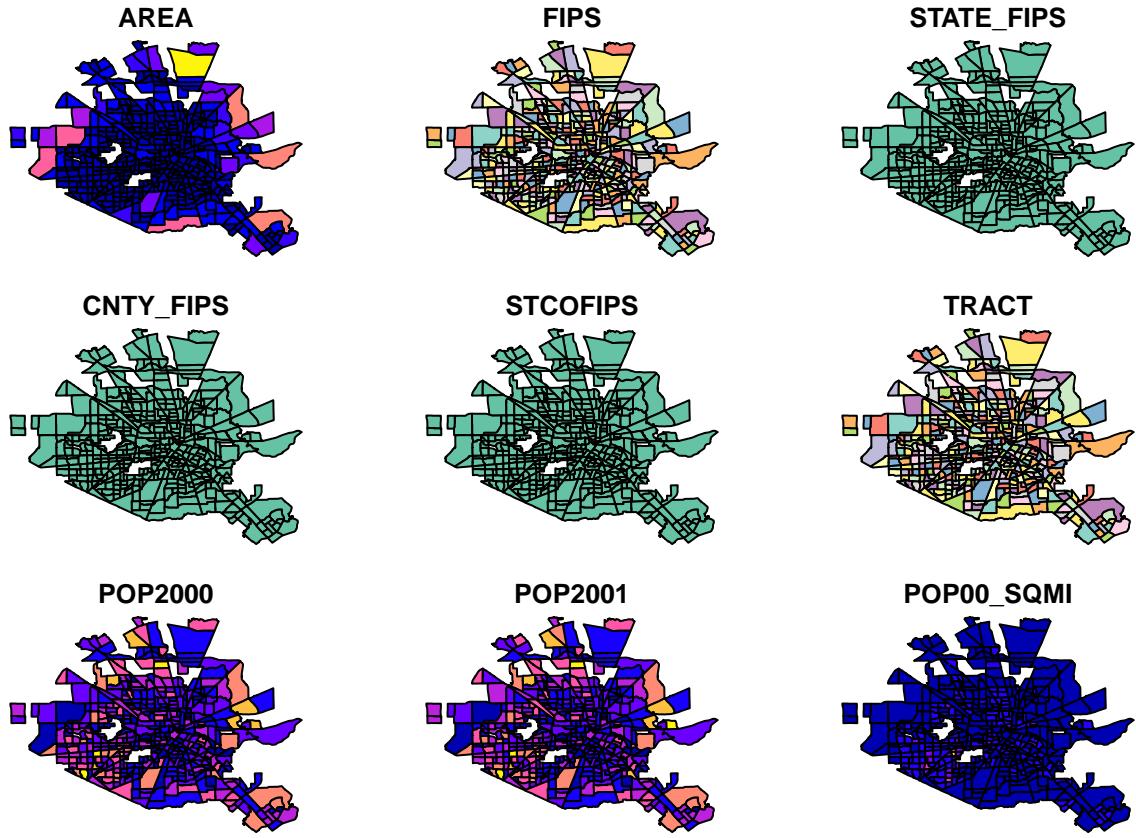
Next, we want to make choropleth maps (shading in each tract based on its attribute value).

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/>)

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



```

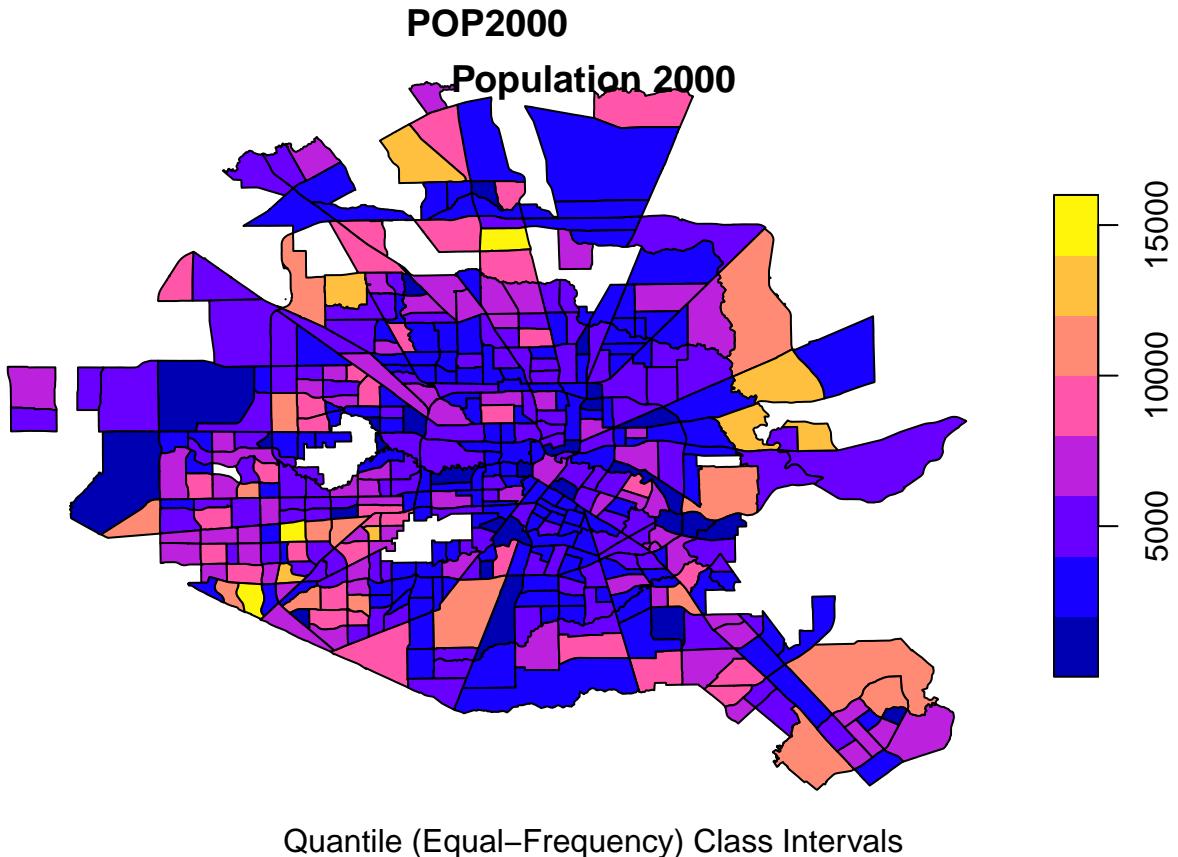
# Define the variable (attribute) to shade tracts by
pop2000 <- houston$POP2000

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

class <- classInt::classIntervals(pop2000, nclr, style="quantile")
colcode <- classInt::findColours(class, plotclr)

#Fill in the tracts with the colors
plot(houston['POP2000'])
#Add a title
title(main="Population 2000",
      sub="Quantile (Equal-Frequency) Class Intervals")

```

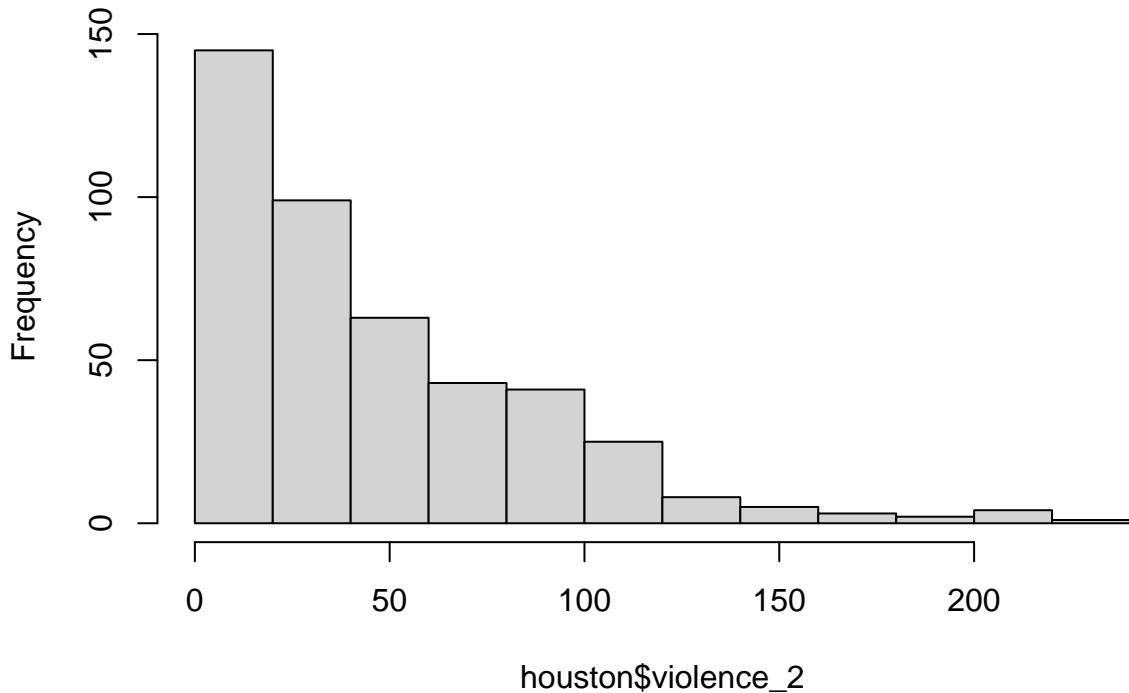


OK, great, we can map the data. Let's map the main variables for our analysis.

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
hist(houston$violence_2)
```

Histogram of houston\$violence_2



```
# Divide by the 2000 population to get the rate
houston$violence.rate = houston$violence_2/houston$tot_pop
```

```
#Summarize key variables
summary(houston$violence.rate)
```

```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## 0.000000  0.003351  0.008368  0.052432  0.014706 13.333333
```

```
summary(houston$violence_2)
```

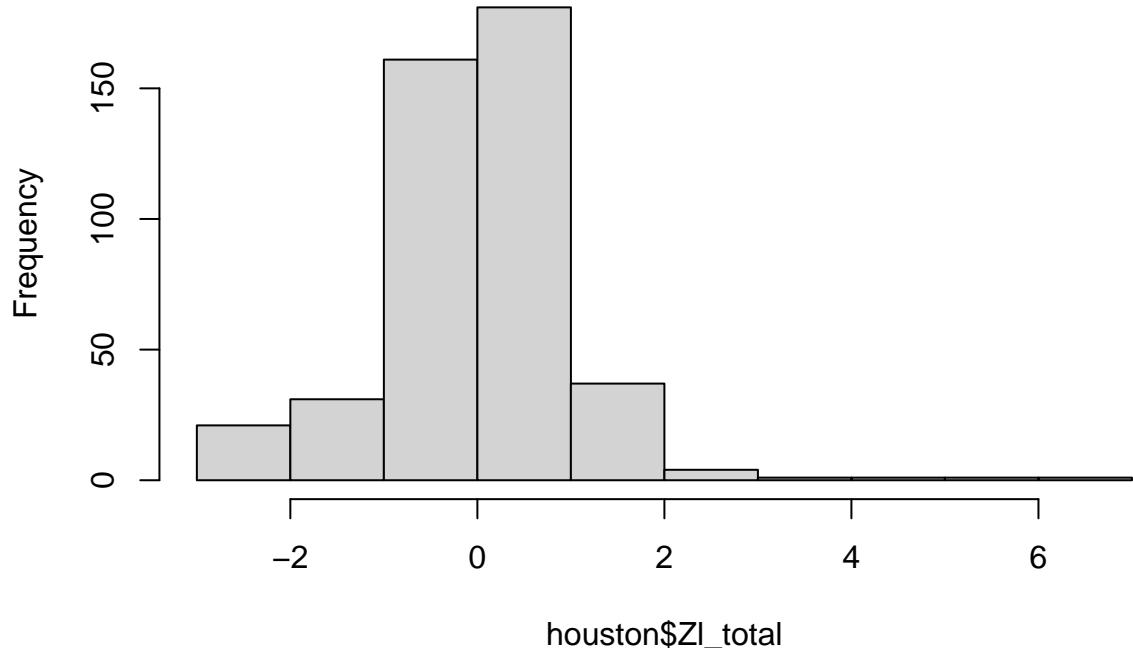
```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## 0.00    14.50    36.00   46.97    69.00  234.00
```

```
summary(houston$tot_pop)
```

```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## 1      2946     4475     5038     6566    15411
```

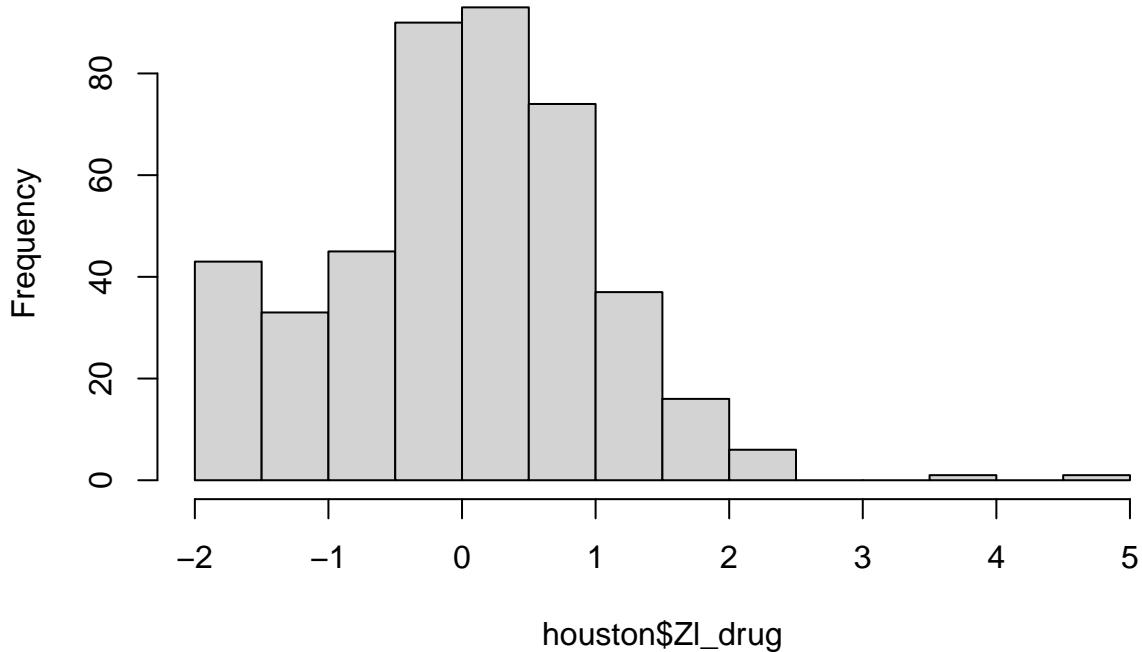
```
# Covariate 1 (log standardized total alcohol sales)
hist(houston$Zl_total)
```

Histogram of houston\$ZI_total



```
# Covariate 2 (log standardized illegal drug arrests)
hist(houston$ZI_drug)
```

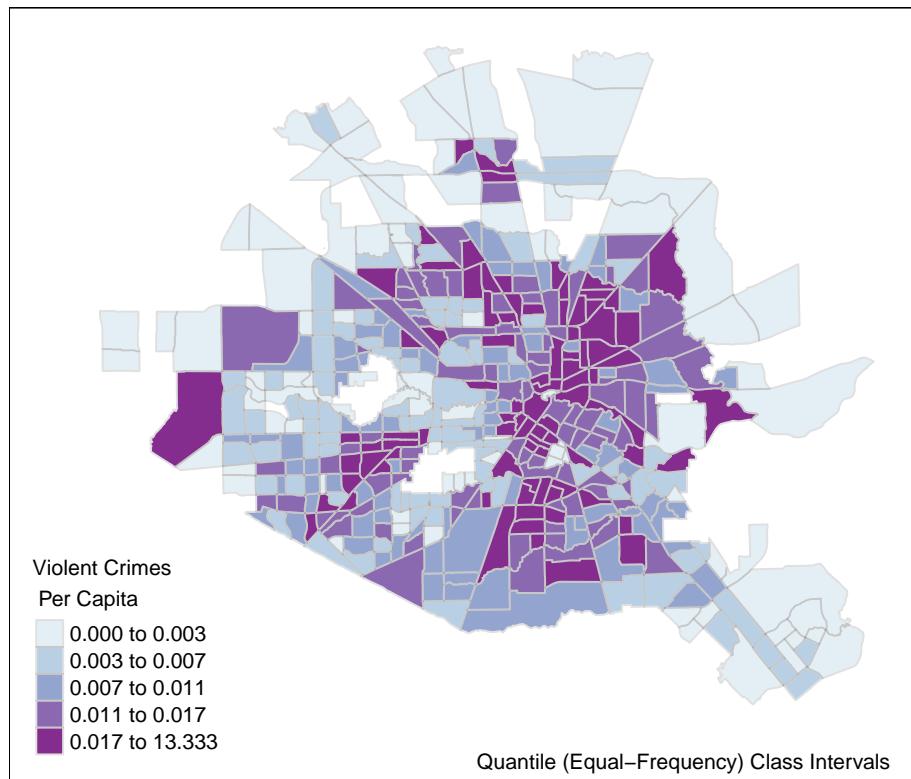
Histogram of houston\$ZI_drug



Now to map the outcome.

```
violence_map <- tm_shape(houston) +
  tm_fill('violence.rate',
  style='quantile',
  palette='BuPu',
  title='Violent Crimes \n Per Capita') +
  tm_borders(alpha=0.2) +
  tm_credits('Quantile (Equal-Frequency) Class Intervals',
             position=c('RIGHT', 'BOTTOM')) +
  tm_layout(main.title="Violent Crime Rate in Houston, TX",
            inner.margins = c(0.1, 0.1, 0.05, 0.05),
            main.title.size=1.2, legend.title.size=0.8)
violence_map
```

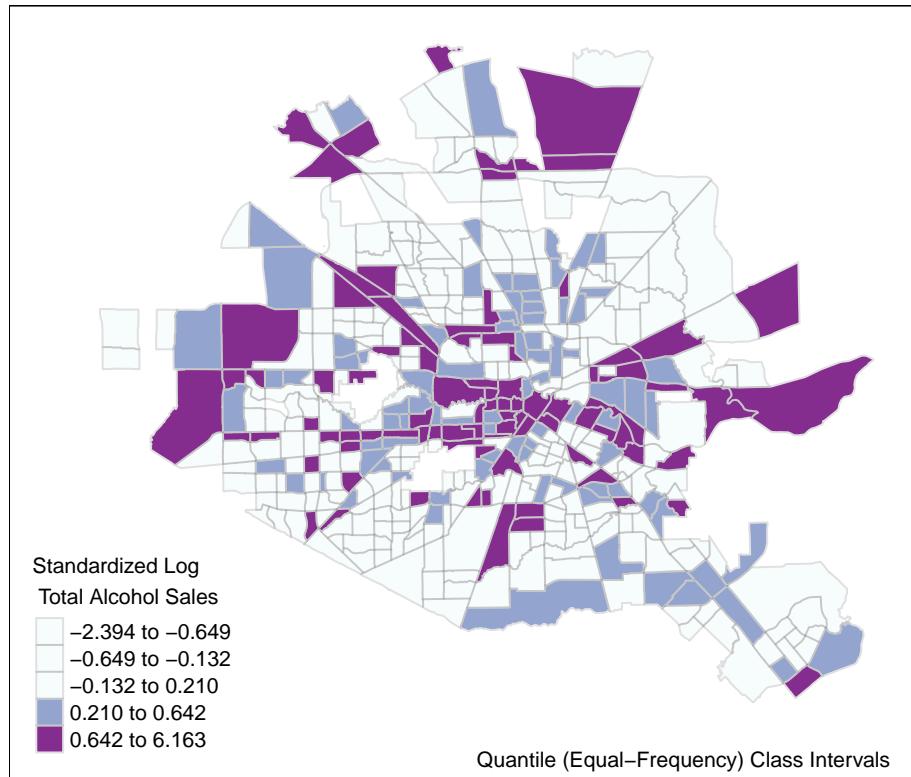
Violent Crime Rate in Houston, TX



Next, map standardized log total alcohol sales.

```
alc_map <- tm_shape(houston) +
  tm_fill('Zl_total',
  style='quantile',
  palette='BuPu',
  title='Standardized Log \n Total Alcohol Sales') + # "\n" moves text to the next line
  tm_borders(alpha=0.2) +
  tm_credits('Quantile (Equal-Frequency) Class Intervals',
             position=c('RIGHT', 'BOTTOM')) +
  tm_layout(main.title="Alcohol Sales in Houston, TX",
            inner.margins = c(0.1, 0.1, 0.05, 0.05),
            main.title.size=1.2, legend.title.size=0.8)
alc_map
```

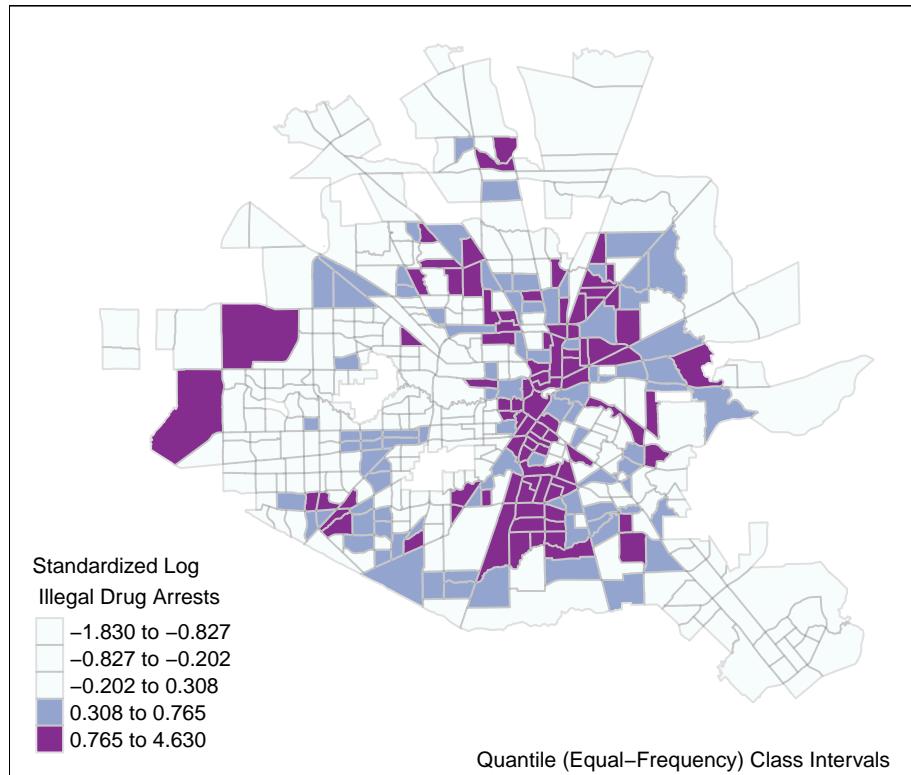
Alcohol Sales in Houston, TX



Finally, plot standardized log illegal drug arrests.

```
drug_map <- tm_shape(houston) +
  tm_fill('Z1_drug',
  style='quantile',
  palette='BuPu',
  title='Standardized Log \n Illegal Drug Arrests') +
  tm_borders(alpha=0.2) +
  tm_credits('Quantile (Equal-Frequency) Class Intervals',
             position=c('RIGHT', 'BOTTOM')) +
  tm_layout(main.title="Illegal Drug Arrests in Houston, TX",
            inner.margins = c(0.1, 0.1, 0.05, 0.05),
            main.title.size=1.2, legend.title.size=0.8)
drug_map
```

Illegal Drug Arrests in Houston, TX

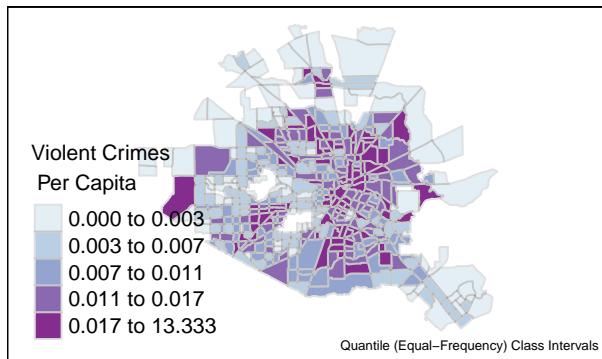


These three figures will match the maps in Figure 1 of:

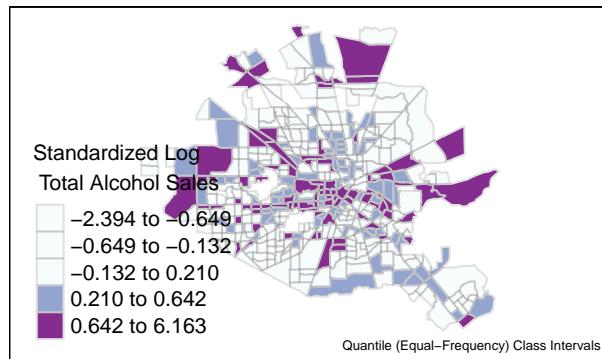
Waller LA, Zhu L, Gotway CA, Gorman DM, and Gruenewald PJ (2007) “Quantifying geographic variations in associations between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models”. *Stochastic Environmental Research and Risk Assessment*.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

```
tmap_arrange(violence_map, alc_map, drug_map)
```

Violent Crime Rate in Houston, TX



Alcohol Sales in Houston, TX



Illegal Drug Arrests in Houston, TX

