Case Study #4 - Bivariate Mapping

PUT YOUR NAME HERE

November 18, 2022

Case Study #4 - Bivariate Mapping (100 points)

This markdown document provides instructions, code, and data to create a bivariate map of environmental quality indicators and alongside demographic information. We will be using data from the 2020 5-year ACS and Texas's Drinking Water Watch website.

Getting Started

The repository's main README has instructions for installing *R* and *RStudio*, installing *Github*, and cloning the course repository. Please refer to those instructions if you have not done so.

You can run individual code sections or lines by navigating into any code section ("chunk") marked with `{r}` and hit the `Run` button at the top of the code editor, or use a keyboard shortcut and type `Ctrl + Enter` (on a Mac, type `Cmd + Return`). The Keyboard shortcuts are really handy for running the code line-by-line. Whatever line the cursor is on, regardless of where the cursor is on that line, can be run by simply typing `Ctrl + Enter` (on a Mac, type `Cmd + Return`).

Running individual lines is useful when testing the commands, but the entire markdown document (compete with code chunks) is ran by hitting the `Knit` button at the top of the page (this is called "knitting"). When you `knit` a document, it will run all the text and all the code and populate a output file (in this case, a .pdf). **This output file is what you will submit for your case study.**

The Bivariate Map

To get started with the replication (after installing *R*, *RStudio*, and cloning/downloading the repository), navigate in your file explorer (or equivalent) to `case_studies\case_study_4` and double click on (open) the markdown file `bivariate_mapping.Rmd`. This will prompt your machine to open the file in *RStudio*. This markdown document includes all the instructions and code to create a bivariate map.

Opening the data (20 points)

Let's start by opening the simplest file, an R dataframe of census information. These are census tract-level variables pulled from the 2020 5-year American Community Survey.

Important note, you will need to change the directory below to your working folder (wherever you saved the materials for this case study).

```
## load data
my_wd <- 'D:/Dropbox (Personal)/Teaching Materials/Econ 475 Environmental Economics/case_studie
load(paste0(my_wd,'tx_census.RData'))
```

1. (10 points) List one variable that might be used to characterize an indicator of social vulnerability.

- Answer:

Next , we're going to load water system service boundary shapefiles for the state of Texas. This data comes pre-merged with average concentrations of all finished water samples from 2012-2020 in a given water system for four specific water quality analytes. Note that the average concentration from 2012-2020 does not necessarily reflect current conditions, and averages are sensitive to extreme values.

```
tx_water <- st_read(paste0(my_wd,'tx_water.shp'))
```

```
## Reading layer 'tx_water' from data source
##   'D:\Dropbox (Personal)\Teaching Materials\Econ 475 Environmental Economics\case_studies\ca
##   using driver 'ESRI Shapefile'
## replacing null geometries with empty geometries
## Simple feature collection with 8830 features and 6 fields (with 4286 geometries empty)
## Geometry type: GEOMETRY
## Dimension:     XY
## Bounding box:  xmin: -795761.7 ymin: 2865031 xmax: 444453 ymax: 4075503
## Projected CRS: NAD83 / UTM zone 15N
```

2. (10 points) This dataset includes information on four water quality parameters. List the parameters.
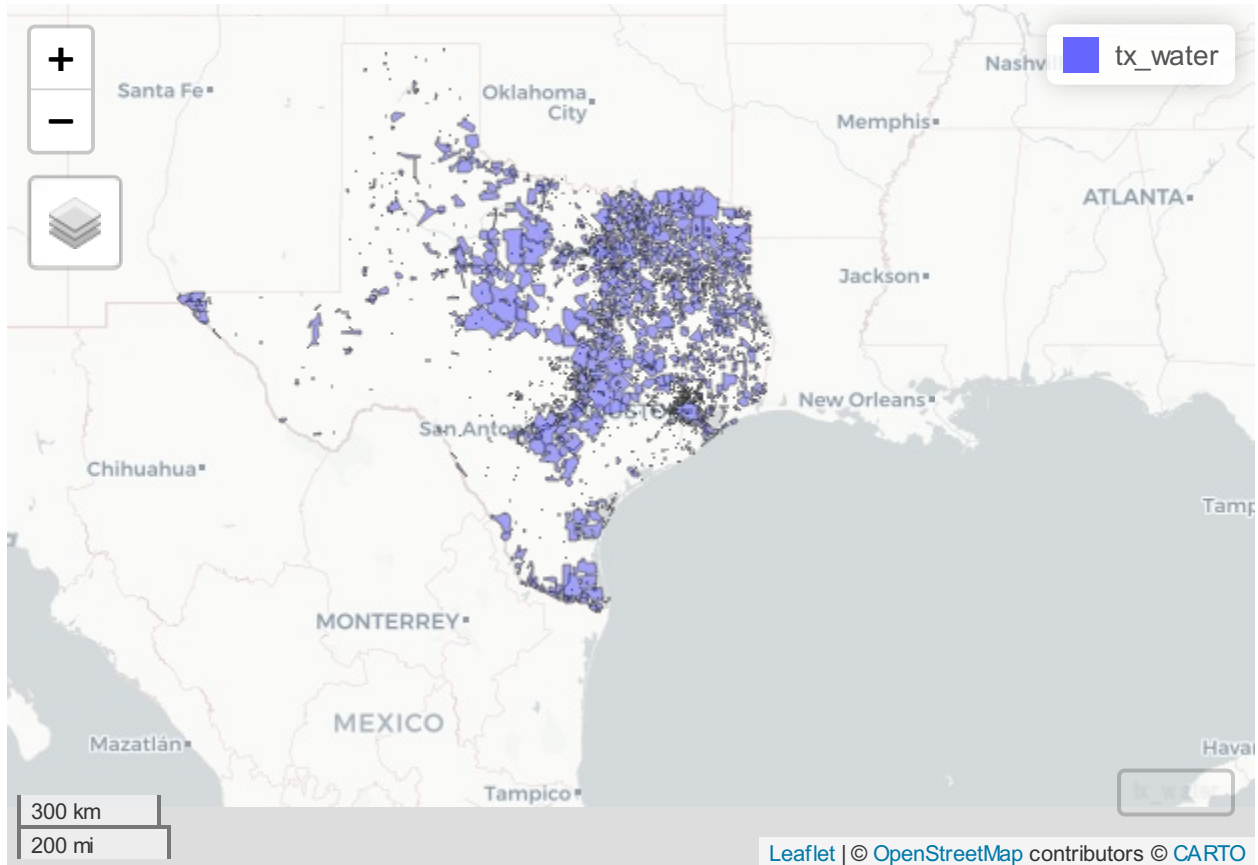
- Answer:

Inspecting the Data (20 points)

Before we can make a bi-variate map, we first need to combine water system service boundaries with census information.

3. (10 points) First, let's take a look at the service boundaries to ensure they were imported correctly.

```
mapview(tx_water)
```


Looks like these were imported successfully.

4. (10 points) Next, let's check the census population information to ensure the geographic layer loaded correctly and the population information appears roughly accurate.

```r
# Define a color pallette for the map
pal <- colorNumeric(palette = "viridis",
                    domain = tx_census$pop,
                    na.color = 'whitesmoke' )
# Map Texas population across census tracts.
tx_census %>%
  st_transform(crs = "+init=epsg:4326") %>%
  leaflet(width = "100%") %>%
  addProviderTiles(provider = "CartoDB.Positron") %>%
  addPolygons(popup = ~ str_extract(GEOID, "^([^,]*)"),
              stroke = FALSE,
              smoothFactor = 0,
```

```
            fillOpacity = 0.7,
            color = ~ pal(pop)) %>%
  addLegend("bottomright",
          pal = pal,
          values = ~ pop,
          title = "Population",
          opacity = 1)
```
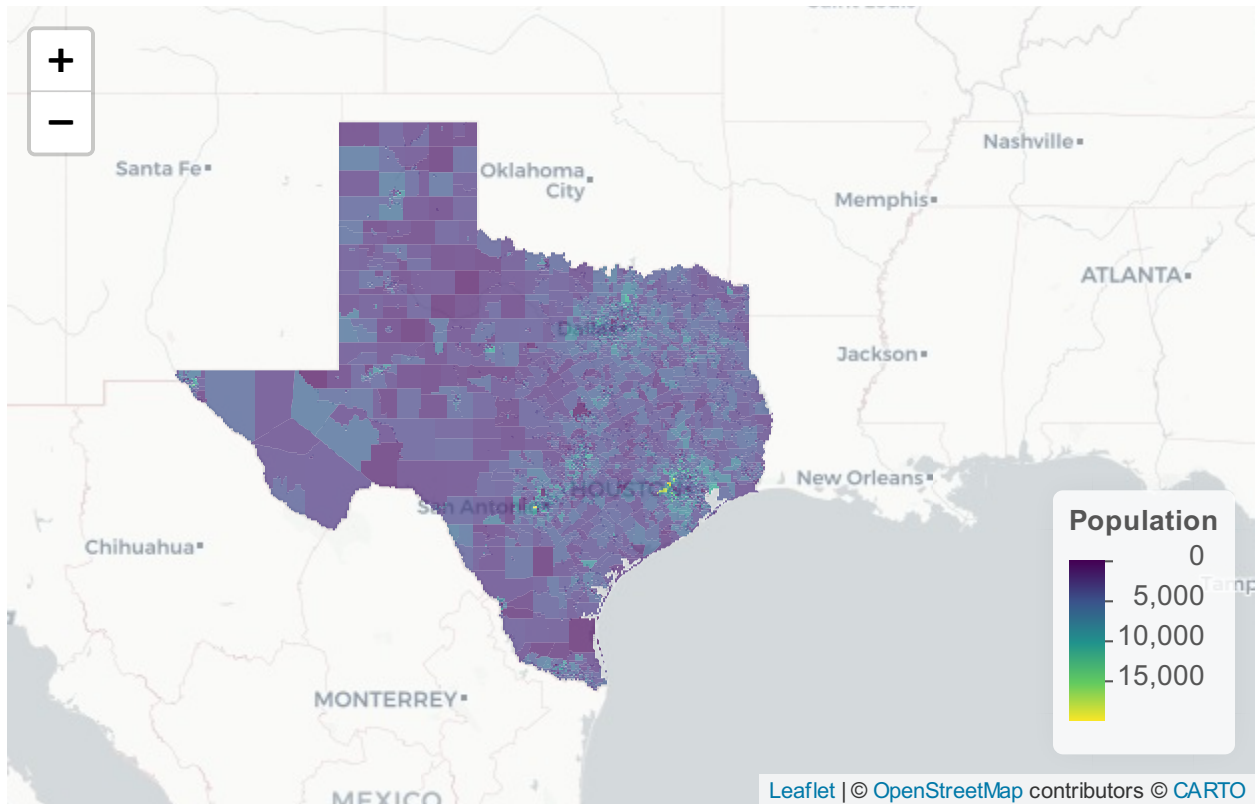

Areal apportionment (25 points)

5. (25 points) Now that we've inspected the data and everything checks out, we need a way to combine service boundaries and census tracts to create a bivariate map of water quality and demographic information.

We're going to use an areal apportionment procedure to combine these two unique sets of data. Note that this step may take a few minutes.

```
#Areally interpolate water system information to census tracts
tx_tract_interpolated <- aw_interpolate(tx_census,
                                        tid = GEOID,
                                        source = tx_water,
                                        sid = PWSID,
                                        weight = "sum",
                                        output = "sf",
                                        intensive = c("arsenic", "lead",
                                                      "nitrate", "tthm"))

# Check how many census tracts have missing tthm information
sum(is.na(tx_tract_interpolated$tthm))
```

[1] 36

(a) (5 points) Some census tracts do not have water quality samples. Given the output from the above command, how many census tracts do not have any total trihalomethane (TTHM) quality information?

- Answer:

(b) (10 points) We used the intensive option to join water system information to tracts. This contrasts with the extensive option in the same command. Use the ?$aw_interpolate$ command to view the help file for this command, and then explain why the intensive option was the correct approach.

- Answer:

(c) (10 points) Areal apportionment is valuable when a specific geographic unit does not perfectly correspond to census divisions other geographic units over which population information is available. This works well in our case because water service boundaries do not correspond to census tract boundaries. Provide one other example where this procedure might be useful.

- Answer:

Bivariate Map (35 points)

Finally, we can prepare to make our bivariate map. A bivariate map jointly displays two variables in a way that facilitates easy location of areas that are tend to have high (or low) values for both indicators simultaneously. This map generation will make use of the biscale package, which has been used to generate visuals in your class readings.

In this case, we're going to plot one of our drinking water indicators alongside a population metric. You can pick other combinations of variables here, but the code below generates a map of % minority plotted against average total trihalomethane (TTHM) concentrations. Recall that TTHM is the most common class of disinfectant byproducts in finished drinking water.

First, create the dimensions of the categories for each bin. This command splits up all census tracts into 9 categories corresponding to the lowest third, middle third, and highest third of the TTHM distribution and does the same for the % minority distribution. 3 categories times 3 categories is 9 total combinations of TTHM and percent low income.

```r
data_biscale <- bi_class(tx_tract_interpolated,
                         x = tthm,
                         y = pct_minority,
                         style = "quantile",
                         dim = 3) %>%
  filter(!str_detect(bi_class, 'NA'))
# Note there are missing values in the x and y fields,
#hence the extra %>% step. This is fine.
```

And, finally, we can generate the bivariate map for the state of Texas.

```r
# This step generates the map's information.
map <- ggplot(data_biscale) +
  geom_sf(data = data_biscale, mapping = aes(fill = bi_class),
          color = NA , size = 0.1, show.legend = FALSE) +
  bi_scale_fill(pal = "DkBlue", dim = 3) +
  labs(
    title = "TTHM and % Minority",
    subtitle = "",
    size = 2
  ) +
  bi_theme()

# This step generates the legend.
legend <- bi_legend(pal = "DkBlue",
                    dim = 3,
                    xlab = "TTHM",
                    ylab = "% Minority",
                    size = 8)

# This step draws the legend and the map in one figure.
finalPlot <- ggdraw() +
  draw_plot(map, 0, 0, 1, 1) +
  draw_plot(legend, 0.4, .08, 0.9, 0.3)

finalPlot
```

TTHM and % Minority
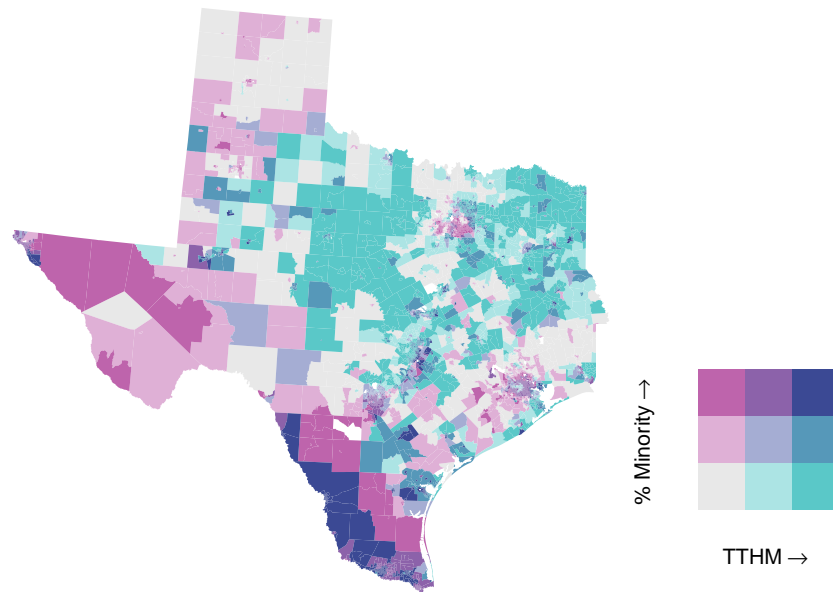
Great, there's a bivariate map! (20 points)

6. (15 points) Describe any patterns that seem apparent from the data.