Lab 6

Ashwin Dev

Math 241, Week 8

```r
# Put all necessary libraries here
library(tidyverse)
library(leaflet)
library(tidycensus)
library(dplyr)
library(readr)
library(lubridate)
library(tidycensus)
library(ggplot2)
library(sf)
library(tigris)
```

Due: Friday, March 22nd at 8:30am

Goals of this lab

- Practice creating static and interactive choropleth maps.

Problem 1: Mapping Bike Rides in Portland

For this problem we will return to the biketown dataset.

a. Grab the code from activity 9, Problem 1 to read the data directly from Biketown's API- make sure to keep the longitude and latitude of the start of each ride (`StartLatitude`, `StartLongitude`).

```r
print(getwd())
```

```
## [1] "/Users/ashwindev/Desktop/DataScience/math241/labs"
```

```r
X2017_01 <- read.csv("/Users/ashwindev/Desktop/DataScience/math241/activities/2017_01.csv")
X2017_07 <- read.csv("/Users/ashwindev/Desktop/DataScience/math241/activities/2017_07.csv")
X2017_11 <- read.csv("/Users/ashwindev/Desktop/DataScience/math241/activities/2017_11.csv")
```

```r
biketown_data <- bind_rows(X2017_01, X2017_07, X2017_11) %>%
  select(StartDate, StartTime, EndDate, EndTime, Distance_Miles,
         BikeID, StartLatitude, StartLongitude)
```

b. Create an interactive map of the start point of the rides using the `leaflet` package. Make sure to include a legend and a title. What do you notice about the distribution of rides?

```
biketown_data %>%
  leaflet() %>%
  addTiles() %>%
  addCircleMarkers(lng = ~StartLongitude, lat = ~StartLatitude)

#I set this code chunk to eval = FALSE since every time I tried to knit, R would time out.
```

The distribution of rides seems to be concentrated in central Portland, with a high density of bike use in the downtown area, which suggests it's a popular hub for commuting and recreational cycling. The presence of thinner clusters around the outskirts, such as Beaverton and Gresham, indicates that bike usage extends into suburban areas, though to a lesser extent. This pattern could reflect urban planning, with central areas offering more bike-friendly infrastructure or a higher population density leading to more frequent bike rides.

c. Using the `lubridate` package, create a variable, `month`, indicating the month of each variable. Add this variable to your interactive map using color. Make sure to include a legend and be mindful of your color palette choice. Do ride locations vary by months of the year?

```
biketown_data <- biketown_data %>%
  mutate(StartDate = mdy(StartDate),
         month = month(StartDate, label = TRUE))
```

```
color_palette <- colorFactor(
  palette = "viridis",
  domain = biketown_data$month
)

biketown_data %>%
  leaflet() %>%
  addTiles() %>%
  addCircleMarkers(
    lng = ~StartLongitude,
    lat = ~StartLatitude,
    color = ~color_palette(month),
    popup = ~paste("Month:", month)
  ) %>%
  addLegend(
    "bottomright",
    pal = color_palette,
    values = ~month,
    title = "Month of Ride",
```

```
    labFormat = labelFormat(prefix = "")
  )

# generates map with color-coded circle markers based on the month variable we just made

# I set eval =  FALSE again since this code chunk would also cause R to time out when I try to knit.
```

We can see a pretty distinct variation in bike ride start locations by month, suggesting seasonal trends in bike usage.

In July, a month characterized by warmer weather, rides are more dispersed, possibly indicating an increase in recreational activities and the use of bikes for exploring more distant areas. Conversely,

in November, which is typically colder, rides are concentrated closer to Portland's center. This pattern might reflect a preference for shorter, utilitarian trips possibly due to less favorable weather conditions.

This trend also could be influenced by the geographic distribution of infrastructure like bike lanes or racks, as well as cultural events and tourism patterns which vary throughout the year. It would be interesting to analyze further how these patterns correlate with weather data, public events, and changes in the urban landscape to understand the full context of these seasonal mobility patterns.

Problem 2: Choropleth Maps

For this problem, I want you to practice creating choropleth maps. Let's grab some data using `tidycensus`. Remember that you will have to set up an API key.

```
api_key <- "2810d317f3b81a8553f2386fa2e8bc0a7ceb2481"
```

```
api_key <- Sys.getenv("CENSUS_API_KEY")
print(api_key)
```

```
## [1] "2810d317f3b81a8553f2386fa2e8bc0a7ceb2481"
```

a. Let's grab data on the median gross rent (`B25064_001`) from the American Community Survey for Multnomah county, Oregon. I want you to do data pulls at three geography resolutions: county subdivision, tract, and block group.

```
census_api_key(api_key, install = TRUE, overwrite = TRUE)
```

```
## [1] "2810d317f3b81a8553f2386fa2e8bc0a7ceb2481"
```

```
variable <- "B25064_001"

state <- "OR"
county <- "Multnomah"
```

```r
subdivision_data <- get_acs(
  geography = "county subdivision",
  variables = variable,
  state = state,
  county = county,
  year = 2022,
  geometry = TRUE,
  survey = "acs5"
)
```

```
##   |                                                                              |
```

```r
# data for county subdivisions

tract_data <- get_acs(
  geography = "tract",
  variables = variable,
  state = state,
  county = county,
  year = 2022,
  geometry = TRUE,
  survey = "acs5"
)
```

```
##   |                                                                              |
```

```r
# data for tracts

block_group_data <- get_acs(
  geography = "block group",
  variables = variable,
  state = state,
  county = county,
  year = 2022,
  geometry = TRUE,
  survey = "acs5"
)
```

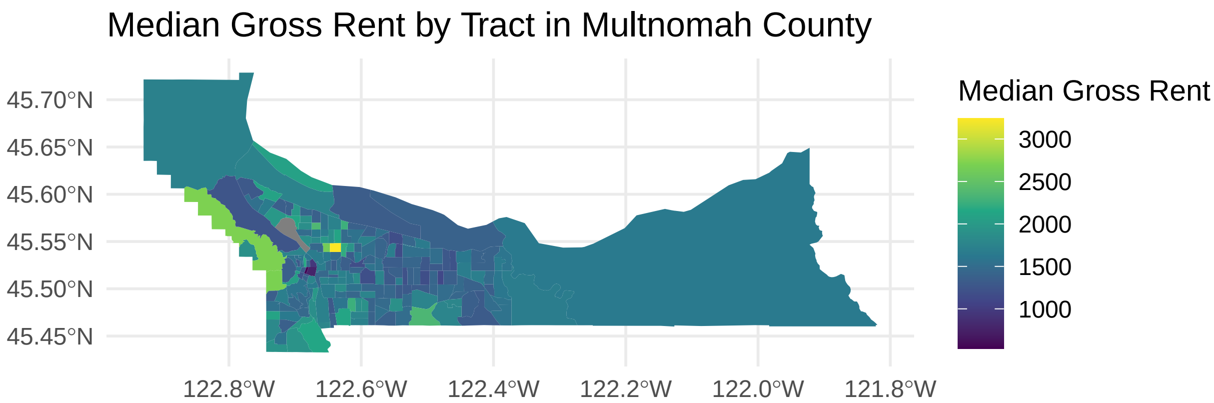
```
##   |                                                                              |
```

```r
# data for block groups
```

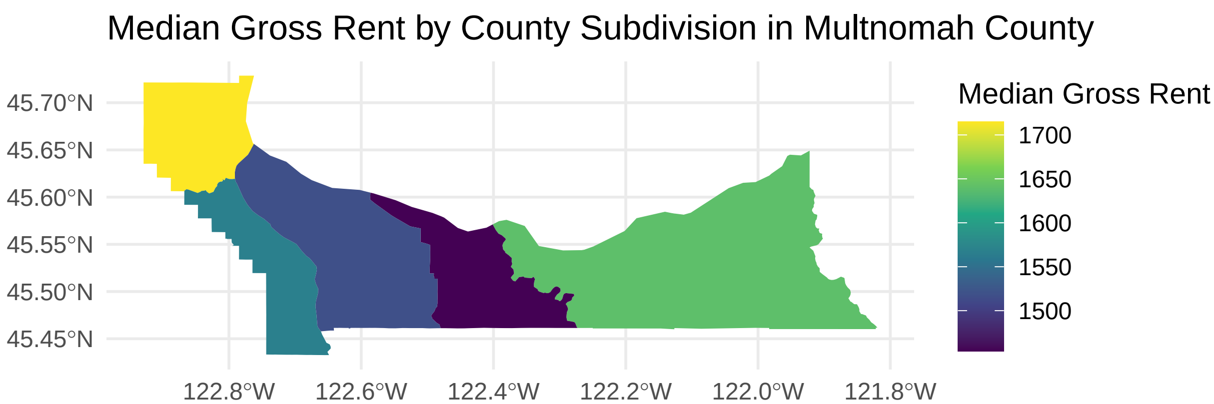
---

b. Create three choropleth maps of gross rent, one for each geography resolution. What information can we glean from these maps? Also, which resolution seems most useful for this variable? Justify your answer.

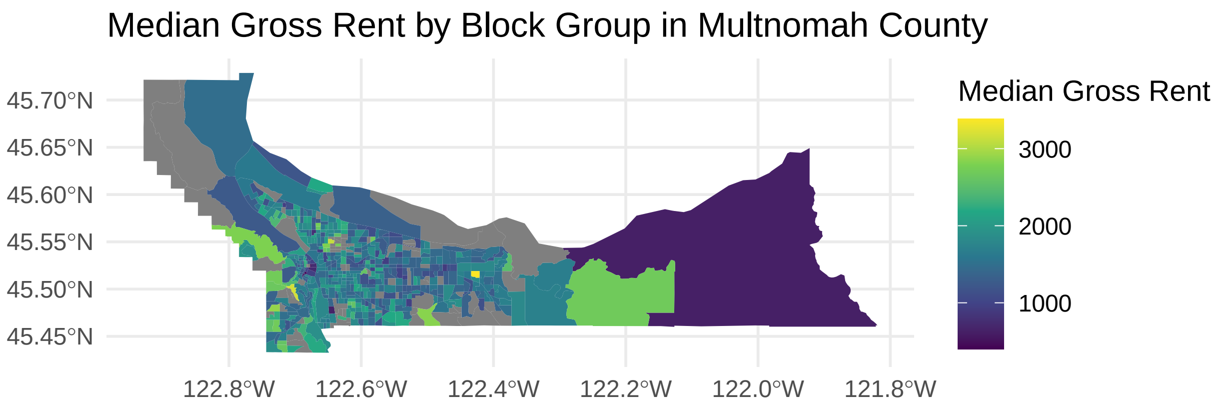
```r
ggplot(tract_data) +
  geom_sf(aes(fill = estimate), color = NA) +
  scale_fill_viridis_c() +
  labs(title = "Median Gross Rent by Tract in Multnomah County",
       fill = "Median Gross Rent") +
  theme_minimal()
```




```r
ggplot(subdivision_data) +
  geom_sf(aes(fill = estimate), color = NA) +
  scale_fill_viridis_c() +
  labs(title = "Median Gross Rent by County Subdivision in Multnomah County",
       fill = "Median Gross Rent") +
  theme_minimal()
```




```r
ggplot(block_group_data) +
  geom_sf(aes(fill = estimate), color = NA) +
  scale_fill_viridis_c() +
  labs(title = "Median Gross Rent by Block Group in Multnomah County",
       fill = "Median Gross Rent") +
  theme_minimal()
```



The choropleth maps above show the median gross rent across Multnomah County at varying levels of granularity. I think the tract-level map offers a balanced view that is detailed enough to identify neighborhood patterns without being too cluttered (like the block group map), making it a versatile choice for both general insights and some detailed analyses. With that being said, I also think the county subdivision map presents a broader overview. More specifically, it's ideal for high-level strategic planning, while the block group map, with its fine detail, is useful for in-depth analysis of specific areas but might be too complex for a quick assessment. The choice of resolution ultimately depends on the intended use; for most practical applications, the tract-level detail tends to provide a good compromise between readability and detail, suitable for both public presentation and informed decision-making.
