STAT 231: Problem Set 8B

Jack Dove

due by 5 PM on Friday, November 6

This homework assignment is designed to help you further ingest, practice, and expand upon the material covered in class over the past week(s). You are encouraged to work with other students, but all code and text must be written by you, and you must indicate below who you discussed the assignment with (if anyone).

Steps to proceed:

1. In RStudio, go to File > Open Project, navigate to the folder with the course-content repo, select the course-content project (course-content.Rproj), and click "Open"
2. Pull the course-content repo (e.g. using the blue-ish down arrow in the Git tab in upper right window)
3. Copy ps8B.Rmd from the course repo to your repo (see page 6 of the GitHub Classroom Guide for Stat231 if needed)
4. Close the course-content repo project in RStudio
5. Open YOUR repo project in RStudio
6. In the ps8B.Rmd file in YOUR repo, replace "YOUR NAME HERE" with your name
7. Add in your responses, committing and pushing to YOUR repo in appropriate places along the way
8. Run "Knit PDF"
9. Upload the pdf to Gradescope. Don't forget to select which of your pages are associated with each problem. *You will not get credit for work on unassigned pages (e.g., if you only selected the first page but your solution spans two pages, you would lose points for any part on the second page that the grader can't see).*

If you discussed this assignment with any of your peers, please list who here:

ANSWER:

1. Mapping spatial data

Reproduce the map you created for Lab08-spatial (and finish it if you didn't in class). In 2-4 sentences, interpret the visualization. What stands out as the central message?

NOTE: you do NOT need to say what colors are representing what feature (e.g, NOT: "In this map, I've colored the countries by GDP, with green representing low values and red representing high values") – this is obvious to the viewer, assuming there's an appropriate legend and title. Rather, what *information* do you extract from the visualization? (e.g., "From the choropleth below, we can see that the percent change in GDP per capita between 1957-2007 varies greatly across countries in Central America. In particular, Panama and Costa Rica stand out as having GDPs per capita that increased by over 200% across those 50 years. In contrast, Nicaragua's GDP per capita decreased by a small percentage during that same time span.")

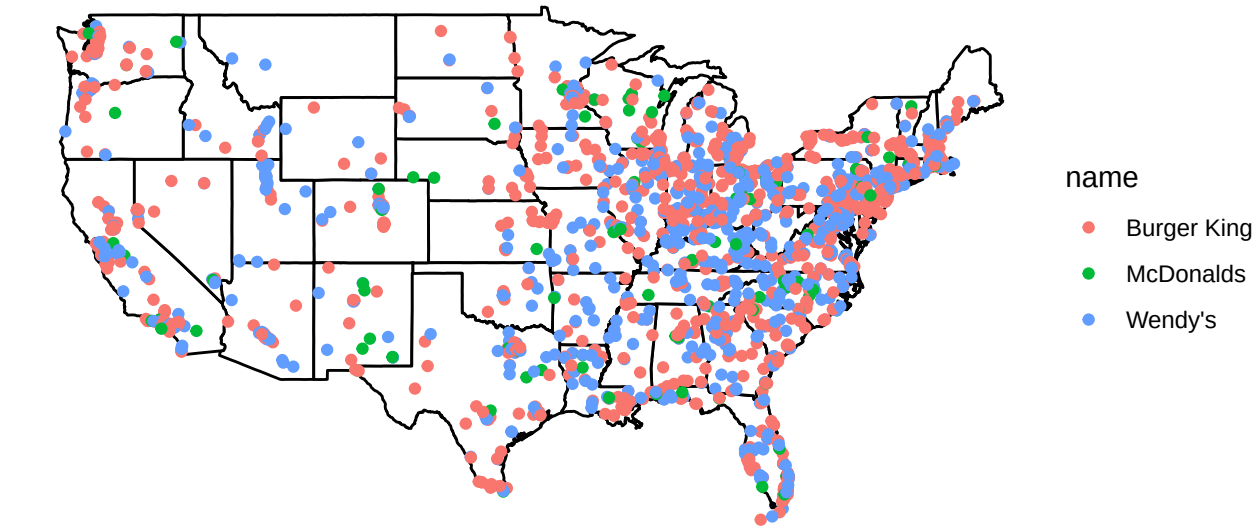
> ANSWER: While my map is not a random sample nor does it cover every location of every fast food restaurant, it does display a strong prevalence of fast food restaurants on the east coast. Especially dense in rust belt states and Northeast cities, these restaurants pop up at similar frequencies. Burger King appears to be the dominant chain on this map, followed by Wendy's and then McDonalds. After an analysis of this plot, I conclude that a possible trend to research would be the disparity between fast food restaurant densities on the east vs the west coast, as well as if McDonalds is truly third fiddle to Wendy's and Burger King.

```
#Preliminary Wrangling
restaurants <- read_csv("~/Desktop/Data Science/Stat231JackDove/Labs/FastFoodRestaurants.csv")
```

```
## Parsed with column specification:
## cols(
##   address = col_character(),
##   city = col_character(),
##   country = col_character(),
##   keys = col_character(),
##   latitude = col_double(),
##   longitude = col_double(),
##   name = col_character(),
##   postalCode = col_character(),
##   province = col_character(),
##   websites = col_character()
## )
```

```
restaurants1 <- restaurants %>%
  rename(state=province)

#Selecting three main restaurants, filtering out non-continental states
restaurants2 <- restaurants1 %>%
  filter(name == "McDonalds" | name == "Burger King" | name == "Wendy's")  %>%
  filter(state != "AK") %>%
  filter(state != "Co Spgs") %>%
  filter(state != "HI") %>%
  filter(longitude > -125 && longitude < -67) %>%
  filter(latitude >25 && latitude < 50)

#USA states map
usa_states <- map_data(map = "state"
```

```
                        , region = ".")

#Plot
ggplot() +
  geom_polygon(data = usa_states, aes(x = long, y = lat, group = group)
               , color = "black", fill = "white") +
  geom_point(data = restaurants2, aes(x = longitude, y = latitude
                                      , color = name)) +
  coord_quickmap(xlim = c(-125, -67), ylim=c(25, 50)) +
  theme_void() +
  labs(title = "Fast Food Restaurant Locations"
       , subtitle = "US McDonalds, Wendy's, and Burger King Stores"
       , fill = "")
```



2. Mapping spatial data at a different level

Create a map at the world, country, or county level based on the choices provided in lab08-spatial, that is at a DIFFERENT level than the map you created for the lab (and included above). For instance, if you created a map of US counties for the lab, then choose a country or world map to create here.

Note: While I recommend using one of the datasets provided in the lab so you don't spend a lot of time searching for data, you are not strictly required to use one of those datasets. You could, for instance, create a static map that might be relevant to your project (so long as it's at a different level than your map above).

Describe one challenge you encountered (if any) while creating this map.

> ANSWER: While I found it rather simple to overlay restaurant locations on the counties of CT, when I wanted to add context of my home's location, it took me a bit to figure out how to add a singular point. I then learned that I could simply add another geom_point function with my home's coordinates.

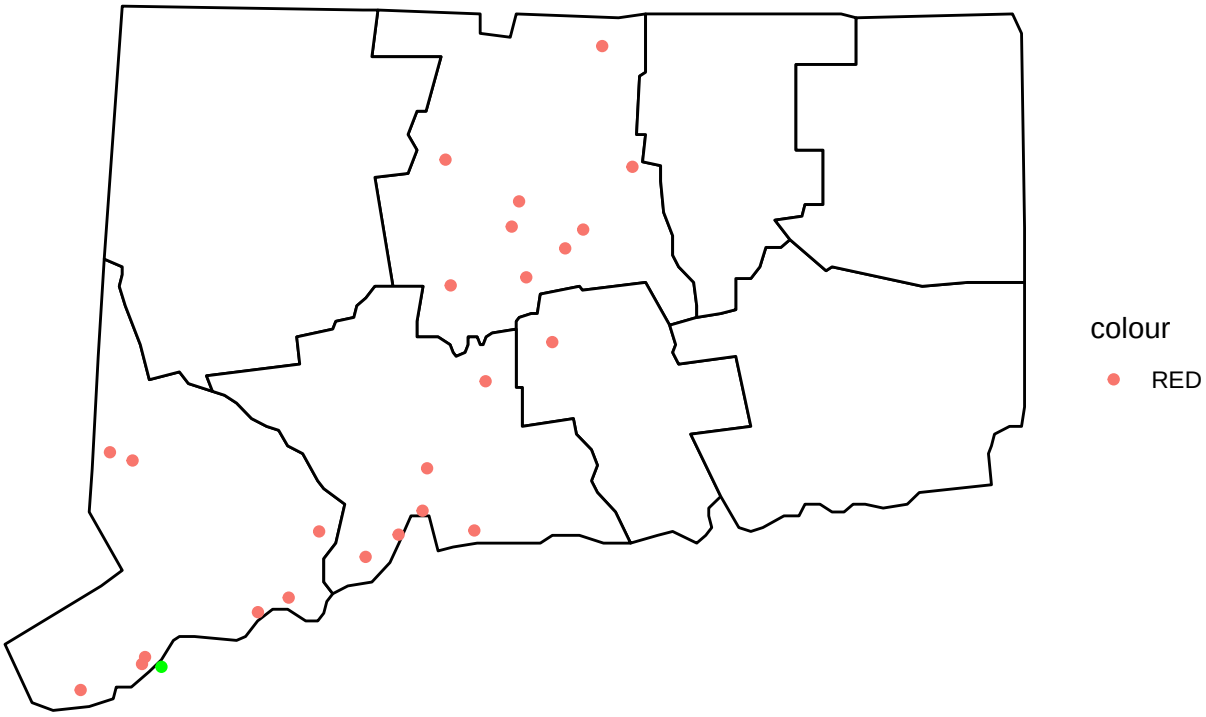
```
#Dataset Load
chipotle_stores <- read_csv("~/Desktop/Data Science/Stat231JackDove/Homeworks/chipotle_stores.csv")
```

```
## Parsed with column specification:
## cols(
##   state = col_character(),
##   location = col_character(),
##   address = col_character(),
##   latitude = col_double(),
##   longitude = col_double()
## )
```

```
#Chipotle locations in CT
chipotle_stores1 <- chipotle_stores %>%
  filter(state=="Connecticut") %>%
  select(latitude, longitude)

#CT counties data
ct_counties <- map_data(map = "county"
                        , region = "connecticut")

homelong <- -73.424380
homelat <- 41.076560
home<- data.frame(homelat, homelong)

ggplot() +
  geom_polygon(data = ct_counties, aes(x = long, y = lat, group = group)
               , color = "black", fill = "white") +
   theme(legend.position = "none") +
  geom_point(data = chipotle_stores1, aes(x = longitude, y = latitude
                                          , color = "RED")) + geom_point(data=home, aes(x=homelong, y=homela
  theme_void() +
  labs(title = "Connecticut Chipotle Locations"
       , subtitle = "Red = Stores, Green = My Home"
       , fill = "")
```

Connecticut Chipotle Locations

Red = Stores, Green = My Home



3. Ethics follow-up

(a) Thinking about the discussion you had with the first group you were with during class last Thursday (focused on either "Predicting Policing & Recidivism" or "Predicting Financial Risk"), did your perspective on, or understanding of, any of the questions shift? If so, please describe. If not, was there anything you found surprising in the resources or your first group discussion?

ANSWER: My perspective on the question of why many algorithms aren't trained on diverse images changed when someone explained the programming process of many alogirthms. She mentioned how, in Amherst's Intro to Computer Science class, students write programs, test their own programs, and then submit them. If programmers follow this method, they will test their algorithms off of themselves (easier), and base their algorithm on those tests. This idea helped me understand why many algorithms are solely trained on white faces, as the programming world is dominated by white and asian programmers.

(b) Thinking about the discussion you had with the second group you were with during class last Thursday (focused on considering the use of algorithms in the college admissions processs), did your perspective on, or understanding of, the use of algorithms in these contexts shift? If not, was there anything you found surprising in the resources or your second group discussion?

ANSWER: I was surprised that my second group was very supportive of an algorithm deciding on admissions decisions. I thought that, especially at an affirmative action-supporting school, students would be merit of a solely-merit based, machine decicision. Students in my discussion felt very comfortable taking human subjectivity out of the equation, replacing it with the objectivity of an algorithm.