Project 5

Read in the dataset you will be working with:

```
stations <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2022/2022-03-01/stations.csv')
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

The dataset we will be using is the ‘Alternative Fuel’ data set. This dataset comes from the US Department of Energy and it provides information on alternative fuel stations in the United States.

Full data set: https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-03-01/readme.md (https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-03-01/readme.md). Full data dictionary: https://afdc.energy.gov/data_download/alt_fuel_stations_format (https://afdc.energy.gov/data_download/alt_fuel_stations_format)

Question: Where in the US (lower 48 states only) are the most common types of alternative fuel stations concentrated? Which states have the most alternative fuel stations, which have the most stations per unit area, and which have the most stations per person?

Introduction: The main dataset we will be using to answer these questions is the ‘Alternative Fuel’ data set. This dataset comes from the US Department of Energy and it provides information on the alternative fuel stations within the United States. The columns we will use include FUEL_TYPE_CODE (the type of fuel station, coded with an abbreviation), STATE (the US state in which fuel station is located), X (the latitude coordinate of the fuel station), and Y (the longitude coordinate of the fuel station). We will also use R’s built in dataset “state.x77” to get the populations and areas for the states as well as “state.center” for coordinates of the center of each state and “state.abb” for abbreviations of the state names.

Approach: We will start by wrangling all the data necessary for analysis by creating, filtering, summarizing, and joining all necessary data frames, computing the relevant density statistics, and combining data frame with sf object for use in all plots. Next, we will create a Cloropleth map and add a fill aesthetic that corresponds to the count of charging stations for each state. We will also make cloropleth maps for station density per unit area and station density per person. These plots will allow us to see which states have more total alternative charging stations, which states have more per person, and which have more per unit area. We will then create geospatial plots of the 5 most common alternative fuel stations in the lower 48 states and color by type. These plots will allow us to see which alternative fuel technologies are used in the different regions of the US.

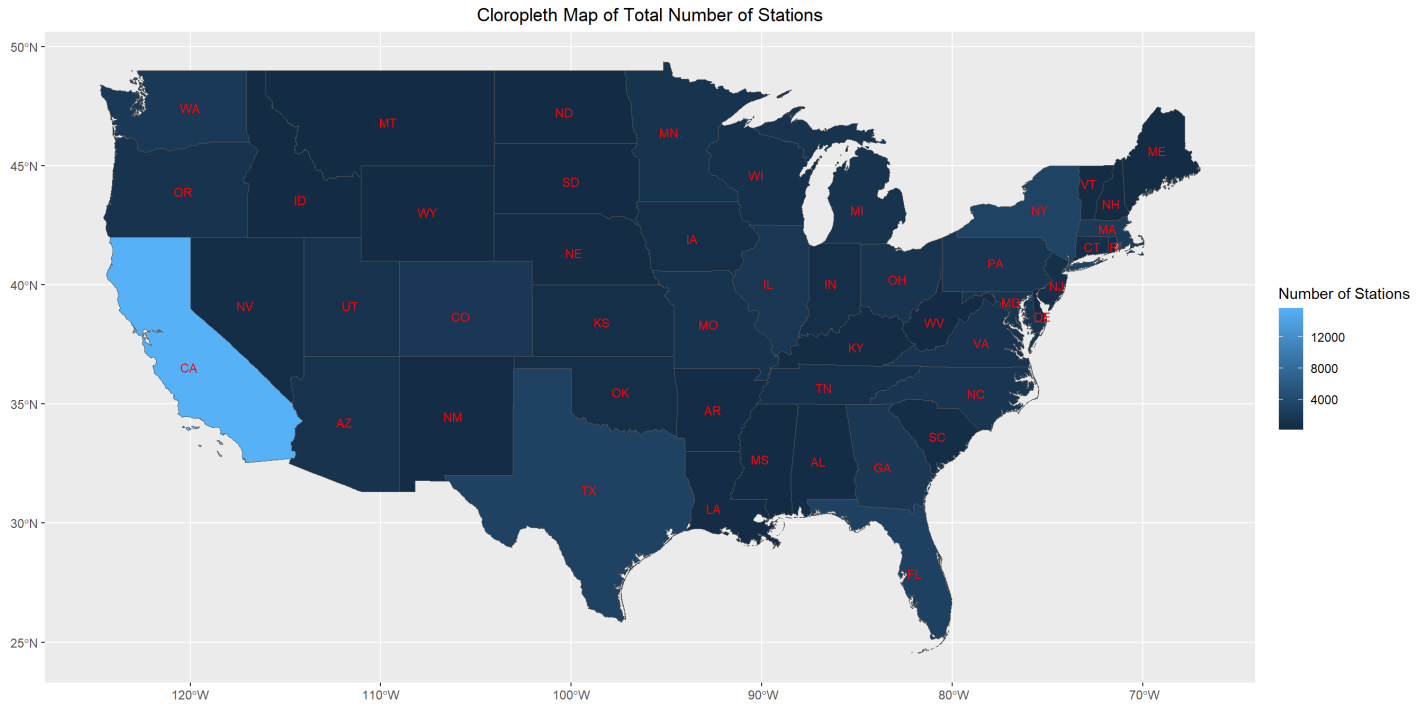
Analysis: Let’s start by wrangling all the data we will need for our plots.

```
#keep only stations in the lower 48 states and lump to keep only top 5 alt fuel types
stations_us <- stations |>
  select(X, Y, STATE, FUEL_TYPE_CODE)|>
  filter(between(X, -125, 65)) |>   #remove Hawaii and Alaska
  filter(between(Y, 20, 50)) |>
  mutate(FUEL_TYPE_CODE_LUMPED = fct_infreq(fct_lump_n(FUEL_TYPE_CODE, 5)))  #lump least frequen
t fuel type codes into 'other' category

#get number of alt fuel stations for each state
num_stations_by_state <- stations_us |>
  count(STATE) |>
  arrange(desc(n)) |>
  rename("Number_of_stations" = "n")

#get population and area for each state
state_stats <- data.frame(state.x77)|>    #create df using R's built in data for state character
istics
  select(Population, Area) |>
  mutate(STATE = state.abb)

#combine number of stations with population and area
num_stations_with_stats <- num_stations_by_state |>
  left_join(state_stats, by = join_by(STATE))

#compute density stats and add to dataframe
num_stations_with_density_stats <- num_stations_with_stats |>
  mutate(Stations_per_sq_m = Number_of_stations/Area,
         Stations_per_person = Number_of_stations/Area)

#create sf obj for US states map
sf_lower48 <- ne_states(country = "United States of America", returnclass='sf') |>
  filter(!code_local %in% c("US02", "US15")) # exclude Alaska (US02), Hawaii (US15)

#add density stats info to sf obj
sf_lower48_with_density_stats <- sf_lower48 |>
  left_join(num_stations_with_density_stats, by = c("postal" = "STATE"))

#get coords of state centers
center_coords <- data.frame(state.center, state.abb)|>
  filter(!state.abb %in% c("AK","HI")) |>
  rename("STATE" = "state.abb",
         "x_center" = "x",
         "y_center" = "y")

#add center coords to sf obj
sf_lower48_with_density_stats_and_center <- sf_lower48_with_density_stats |>
  left_join(center_coords, by = c("postal" = "STATE"))

num_stations_by_state
```

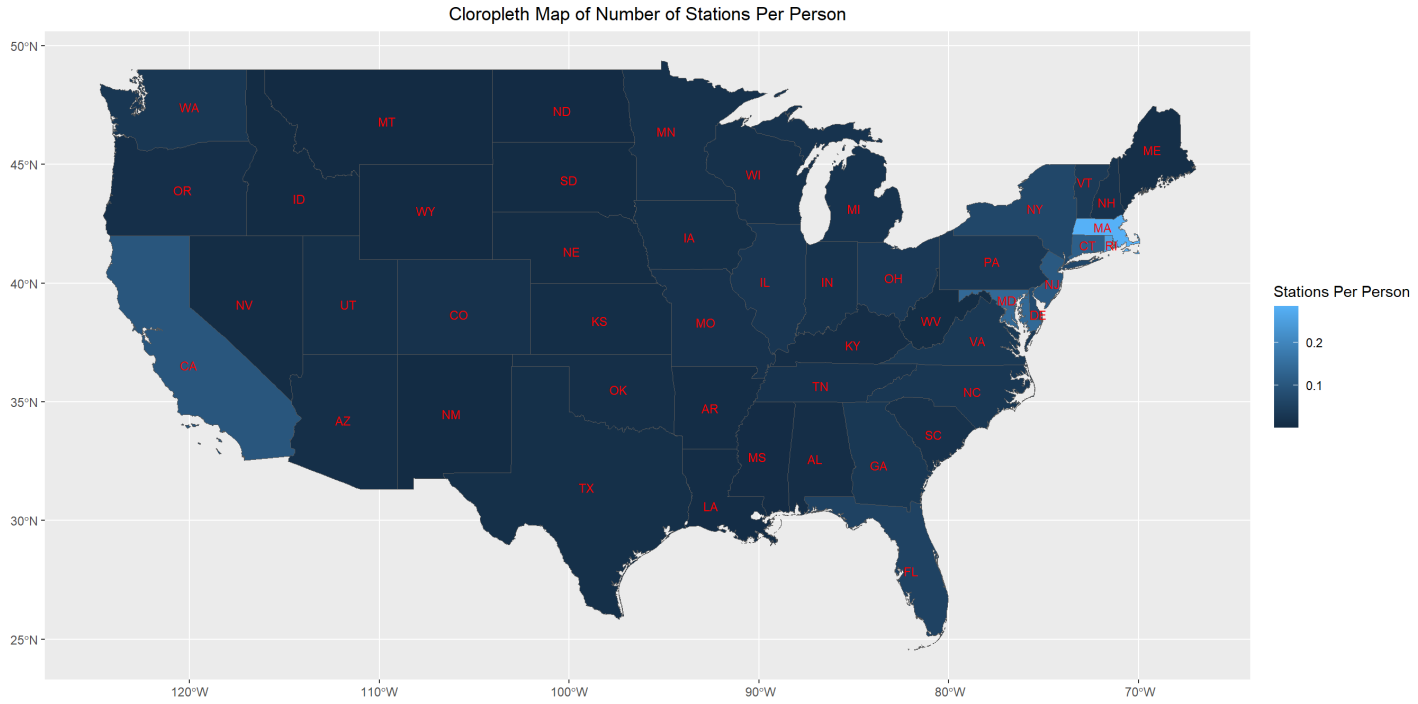
```
## # A tibble: 50 × 2
##    STATE Number_of_stations
##    <chr>              <int>
##  1 CA                 15660
##  2 NY                  3180
##  3 TX                  3128
##  4 FL                  2990
##  5 MA                  2232
##  6 WA                  1910
##  7 CO                  1820
##  8 GA                  1796
##  9 IL                  1536
## 10 PA                  1483
## # ℹ 40 more rows
```

Next, we create cloropleth maps of the states with color indicating the number (or density) of alternative fuel stations in that state.

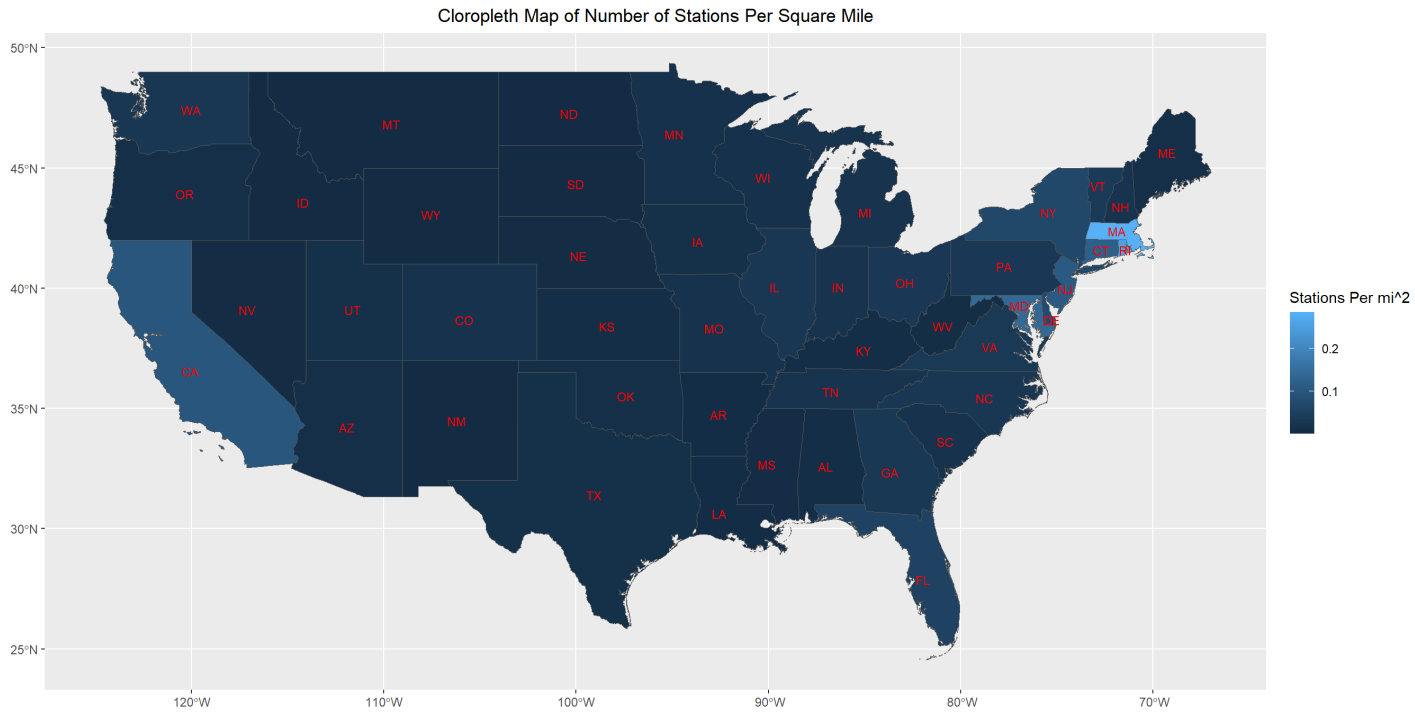
```
#plot choropleths
choropleth_plot <- ggplot(data = sf_lower48_with_density_stats_and_center) +
  geom_sf() +
  geom_text(aes(x_center, y_center, label = postal), color = "red", size = 3, nudge_x = -.3, na.
rm=TRUE) +
  theme(plot.title = element_text(hjust = 0.5)) +
  xlab(NULL) + ylab(NULL)

choropleth_plot +
  aes(fill = Number_of_stations) +
  ggtitle("Choropleth Map of Total Number of Stations") +
  scale_fill_continuous(name = "Number of Stations")
```




```
cloropleth_plot +
  aes(fill = Stations_per_person) +
  ggtitle("Cloropleth Map of Number of Stations Per Person") +
  scale_fill_continuous(name = "Stations Per Person")
```

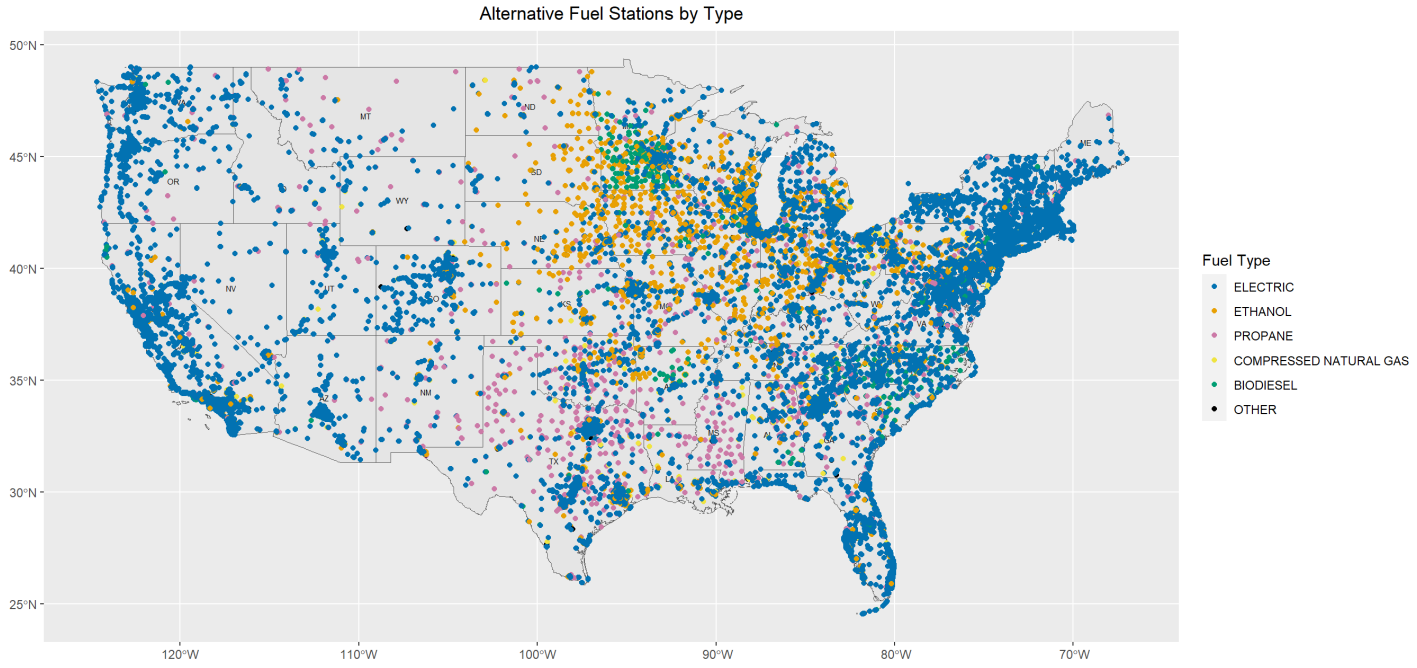



```
cloropleth_plot +
  aes(fill = Stations_per_sq_m) +
  ggtitle("Cloropleth Map of Number of Stations Per Square Mile") +
  scale_fill_continuous(name = "Stations Per mi^2")
```

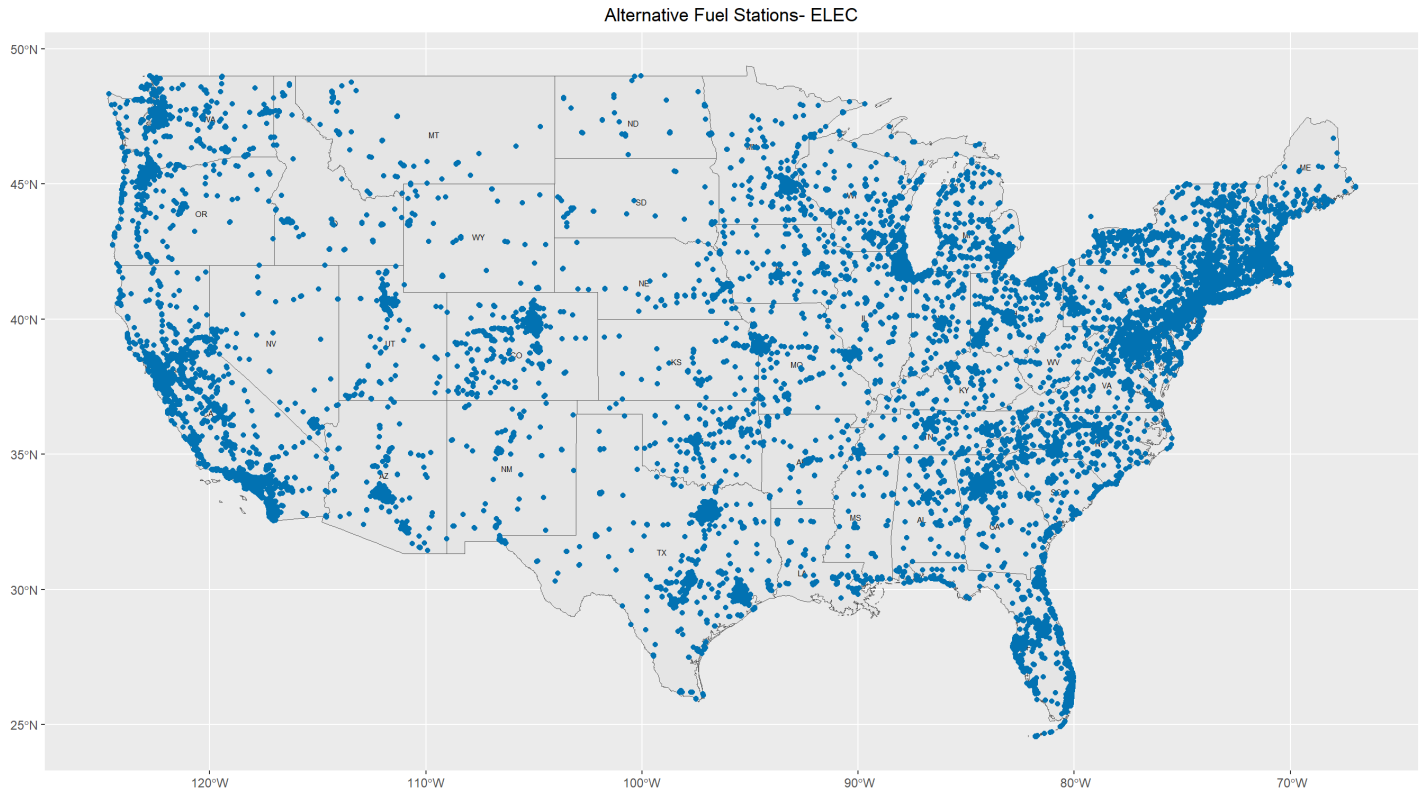


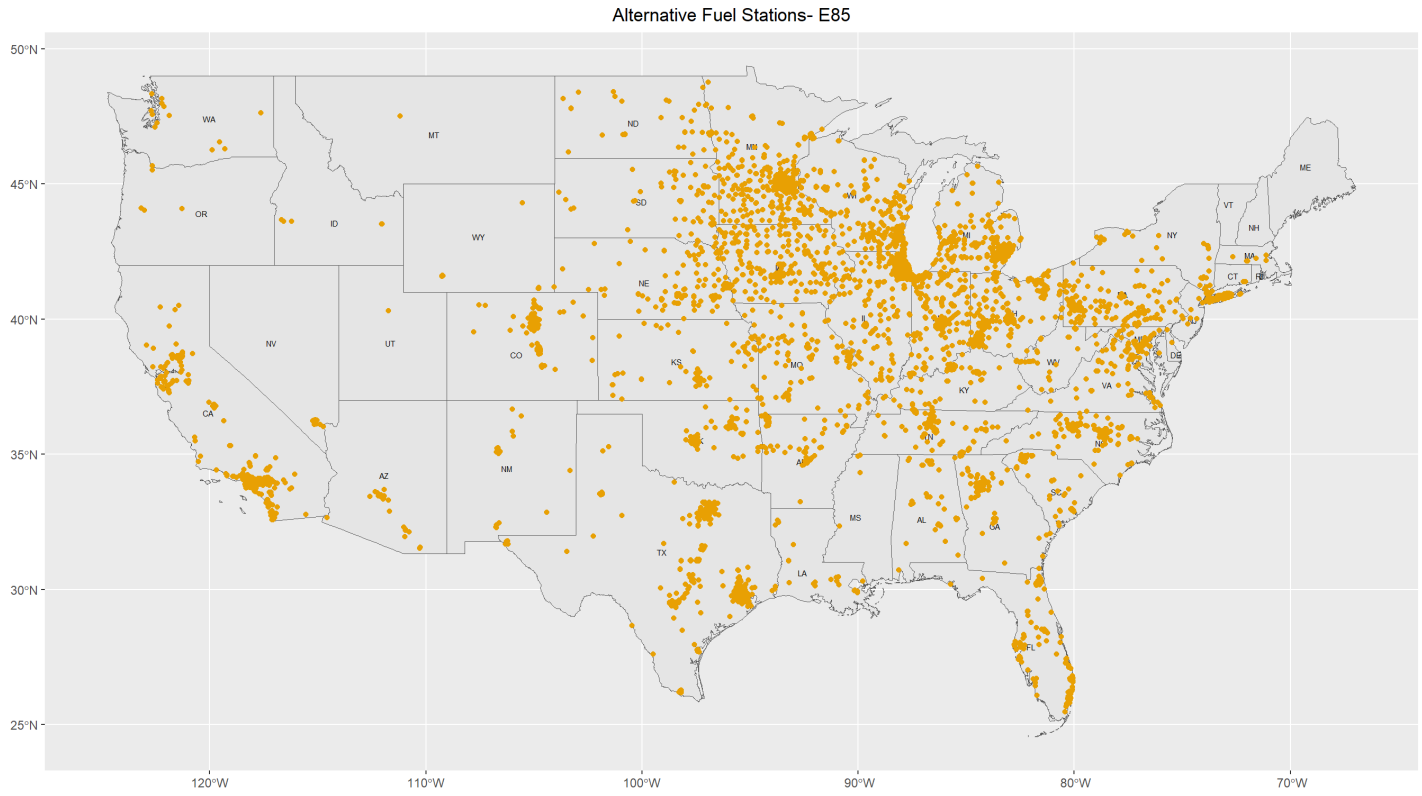
Finally, we create geospatial plots of all the alternative fuel stations in the lower 48 states and color by type.

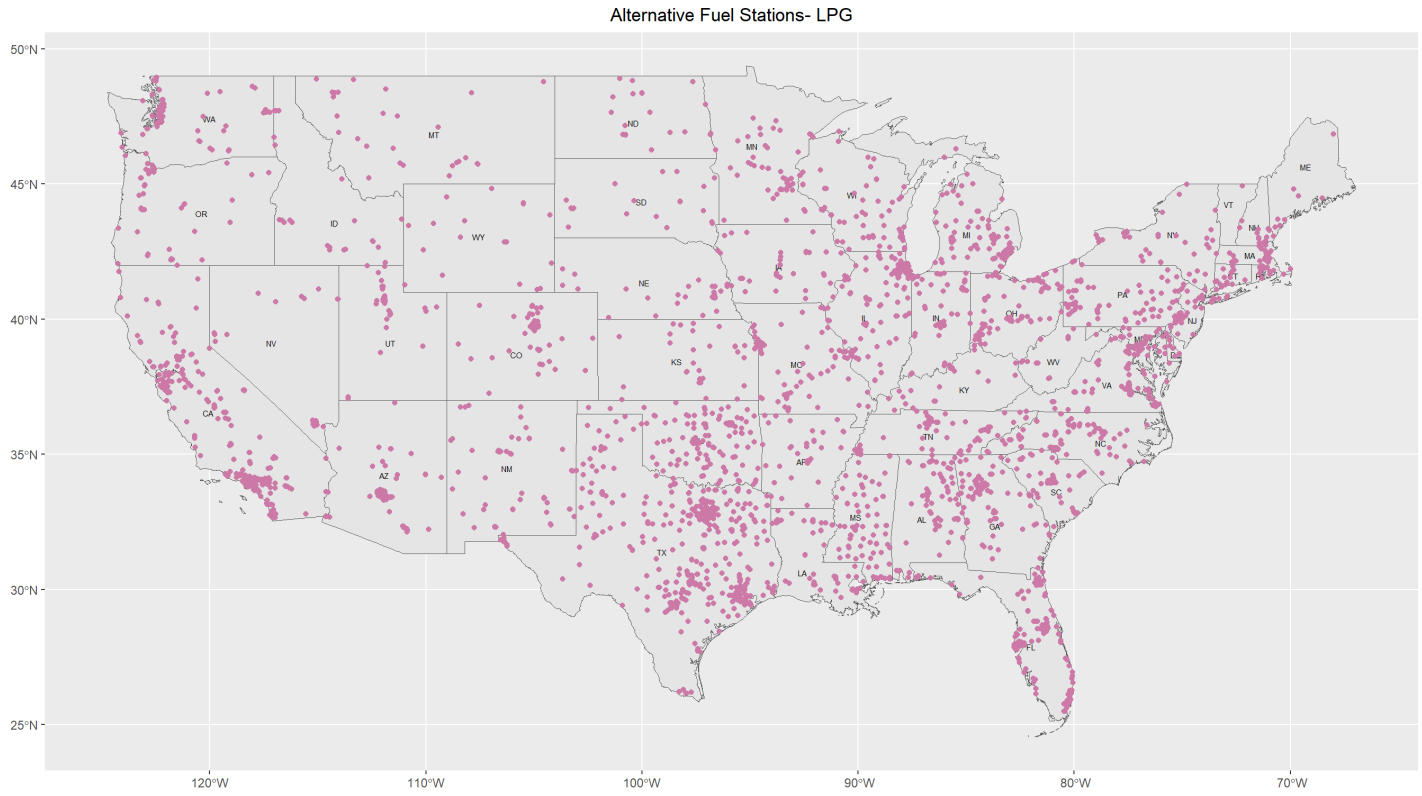
```
#custom color scale for mapping onto fill aesthetic and corresponding labels
color_scale <- c('#0272B2','#E8A003', '#CC79A7','#EFE441', '#009E73','#000000')
labels_ <- c("ELECTRIC","ETHANOL", "PROPANE", "COMPRESSED NATURAL GAS", "BIODIESEL", "OTHER")

#Create plot object for alt fuel stations
p <- ggplot(data = sf_lower48) +
  geom_sf() +
  xlab(NULL) +
  ylab(NULL) +
  theme(plot.title = element_text(hjust = 0.5)) +
  geom_text(data = center_coords,
            aes(x_center, y_center, label = STATE),
            color = "black", size = 2, nudge_x = -.3, na.rm=TRUE)

#plot alt fuel stations colored by fuel type
p + geom_point(data = stations_us, aes(x = X, y = Y, color = FUEL_TYPE_CODE_LUMPED)) +
  scale_color_manual(name = "Fuel Type", values = color_scale, labels=labels_) +
  ggtitle("Alternative Fuel Stations by Type")
```



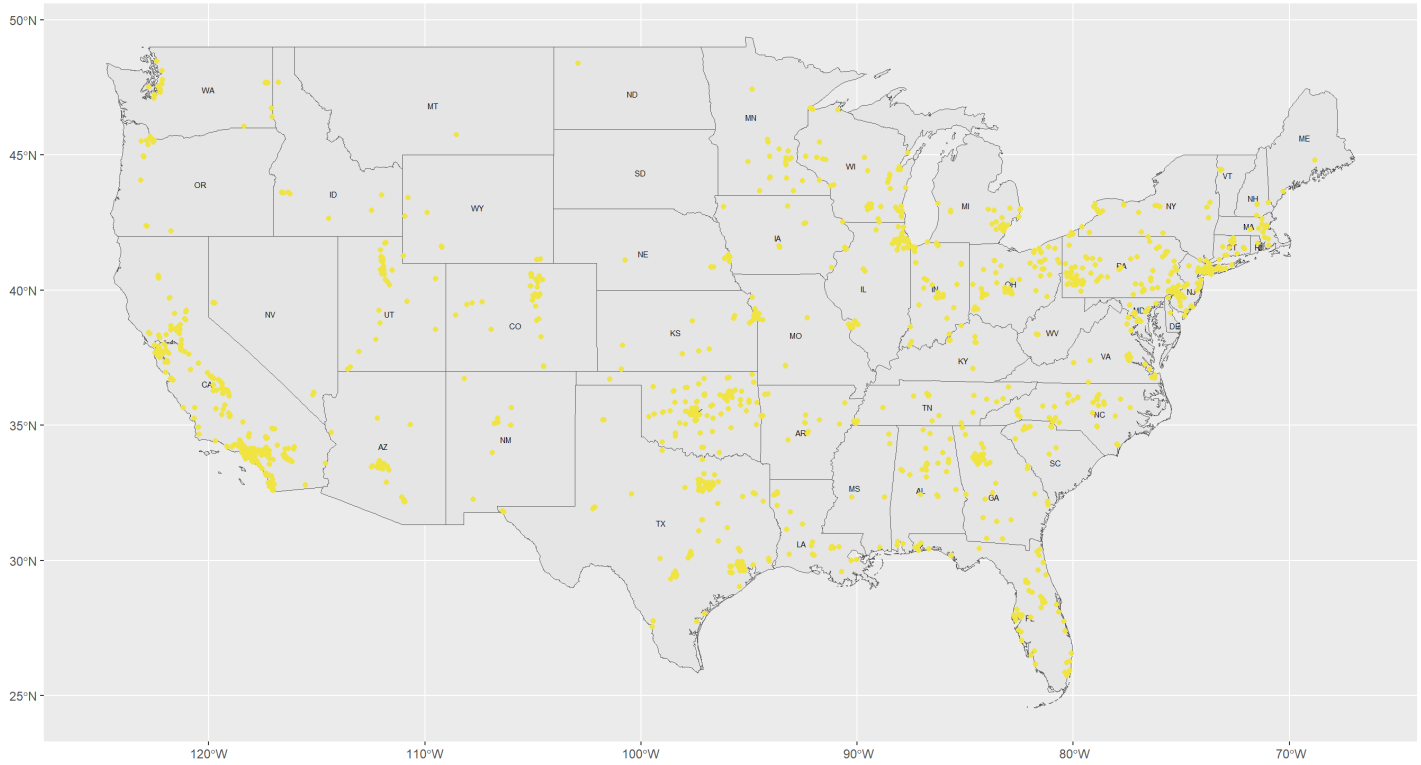
```
#plot each fuel type separately
codes = sort(unique(stations_us$FUEL_TYPE_CODE_LUMPED))
for(i in 1:5){
  df <- filter(stations_us, FUEL_TYPE_CODE == codes[i])
  print(p + geom_point(data = df, aes(x = X, y = Y), color = color_scale[i]) +
    ggtitle(glue("Alternative Fuel Stations- {codes[i]}")))
  }
```

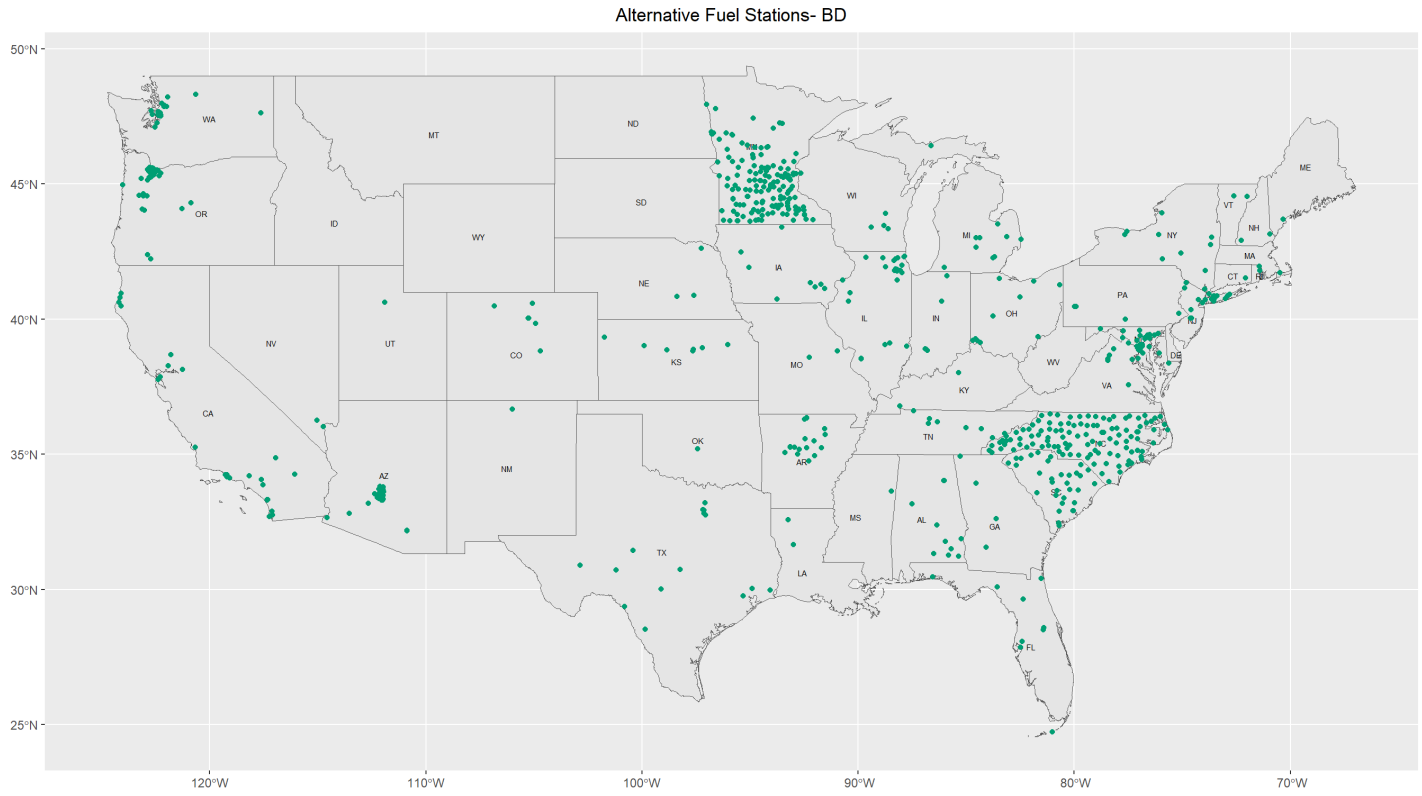

Alternative Fuel Stations- ELEC
50°N
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
70°W
WA
OR
CA
NV
ID
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
GA
FL
SC
NC
VA
WV
PA
NY
ME


Alternative Fuel Stations- E85
50°N
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
70°W
WA
OR
CA
NV
ID
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
GA
FL
SC
NC
VA
WV
OH
PA
NY
ME
VT
NH
MA
CT
RI
NJ
DE
MD


Alternative Fuel Stations- LPG
50°N
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
70°W
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
MD
NJ
CT
RI
MA
VT
NH
ME

Alternative Fuel Stations- CNG





Discussion: From the cloropleth plots, we can see that California has by far the most alternative fuel stations. Maine and Rhode Island have the most stations per person and they also have the most stations per square mile.

From the geospatial plots, we can see that there is generally a high density of alternative fuel stations in the coastal regions and much lower density in the Western states. Electric stations make up the vast majority of all alternative fuel stations are most heavily concentrated in the Eastern states as well as the West coast. Ethanol stations are the next most common and are heavily concentrated in the Midwest. Propane stations and Compressed Natural Gas stations are the next most common types and they both tend to be more evenly distributed throughout the US. Biodiesel stations are the least common of the types we have chosen to study* and are heavily concentrated in Minnesota and the Carolinas.

*We have left out Hydrogen and Liquid Natural Gas stations because they were very uncommon relative to the other station types.