

Assignment 7

Ojaas Hampiholi

4/4/2020

```
library(tidyverse)

library(ggplot2)
library(maps)

library(gridExtra)

library(ggthemes)

library(socviz)

library(evaluate)

library(openintro)

file = "county_data.txt"
a = read.table(file, header = TRUE, sep = "", dec = ".")

file = "county_data.txt"
countyData = read.table(file, header = TRUE, sep = "", dec = ".")
countyData = countyData[,c(1:4,28)]
countyData = countyData[countyData$state != "AK", ]
countyData = countyData[countyData$state != "HI", ]
countyData = na.omit(countyData)
countyData$winner = as.numeric(countyData$winner)
countyData$winner = countyData$winner - 1
#countyData$winner[countyData$winner == "Trump"] = 1
# countyData

stateCodes = read.table("statecodes.txt", header = FALSE, sep = "", dec = ".")
)
stateCodes = stateCodes[,c(2,3)]
names(stateCodes)[names(stateCodes) == "V2"] <- "state"
names(stateCodes)[names(stateCodes) == "V3"] <- "stateName"
head(stateCodes)

##   state  stateName
## 1    AL    ALABAMA
## 2    AK    ALASKA
## 3    AZ    ARIZONA
## 4    AR    ARKANSAS
## 5    CA CALIFORNIA
## 6    CO    COLORADO
```

```

countyDataNew = merge(countyData, stateCodes, by="state")
countyDataNew$stateName <- tolower(countyDataNew$stateName)
data.agg <- countyDataNew %>%
  group_by(stateName) %>%
  summarise(winner=mean(winner))
names(data.agg)[names(data.agg) == "stateName"] <- "region"
data.agg1 = data.agg[(data.agg$winner > 0.5) ,]
data.agg1$president = "Trump"
data.agg2 = data.agg[(data.agg$winner <= 0.5) ,]
data.agg2$president = "Clinton"
data.agg = rbind(data.agg1, data.agg2)

data.agg3 = data.agg[data.agg$region == "districtofcolumbia",]
data.agg3$region = "district of columbia"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "newhampshire",]
data.agg3$region = "new hampshire"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "newjersey",]
data.agg3$region = "new jersey"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "newmexico",]
data.agg3$region = "new mexico"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "newyork",]
data.agg3$region = "new york"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "northcarolina",]
data.agg3$region = "north carolina"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "northdakota",]
data.agg3$region = "north dakota"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "rhodeisland",]
data.agg3$region = "rhode island"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "southcarolina",]
data.agg3$region = "south carolina"
data.agg = rbind(data.agg, data.agg3)

```

```

data.agg3 = data.agg[data.agg$region == "southdakota",]
data.agg3$region = "south dakota"
data.agg = rbind(data.agg, data.agg3)

data.agg3 = data.agg[data.agg$region == "westvirgina",]
data.agg3$region = "west virginia"
data.agg = rbind(data.agg, data.agg3)

us_states = map_data("state")
us_states = us_states[, -c(6)]
head(us_states)

##           long      lat group order  region
## 1 -87.46201 30.38968     1     1 alabama
## 2 -87.48493 30.37249     1     2 alabama
## 3 -87.52503 30.37249     1     3 alabama
## 4 -87.53076 30.33239     1     4 alabama
## 5 -87.57087 30.32665     1     5 alabama
## 6 -87.58806 30.32665     1     6 alabama

usStateElection = left_join(us_states, data.agg, by = "region")
#usStateElection = na.omit(usStateElection)
head(usStateElection)

##           long      lat group order  region  winner president
## 1 -87.46201 30.38968     1     1 alabama 0.8059701      Trump
## 2 -87.48493 30.37249     1     2 alabama 0.8059701      Trump
## 3 -87.52503 30.37249     1     3 alabama 0.8059701      Trump
## 4 -87.53076 30.33239     1     4 alabama 0.8059701      Trump
## 5 -87.57087 30.32665     1     5 alabama 0.8059701      Trump
## 6 -87.58806 30.32665     1     6 alabama 0.8059701      Trump

```

What in one data frame but not in the other?

```
setdiff(us_states$region, data.agg$region)

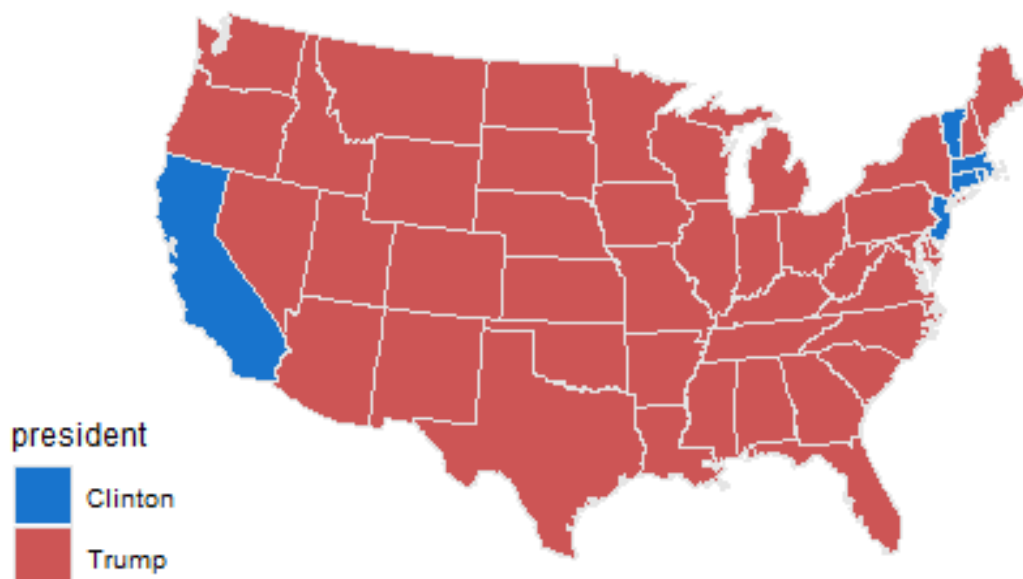
## character(0)

setdiff(data.agg$region, us_states$region)

## [1] "newhampshire"      "newmexico"          "newyork"
## [4] "northcarolina"      "northdakota"        "southcarolina"
## [7] "southdakota"        "westvirgina"        "districtofcolumbia"
## [10] "newjersey"          "rhodeisland"
```

Plotting the discrete variable showing the winner of the elections

```
ggplot(usStateElection, aes(x = long, y = lat, group = group, fill = presiden
t)) +
  geom_polygon(color = "gray90", size = 0.25) + theme_map() +
  coord_map(projection = "lambert", lat0 = 30, lat1 = 40) +
  scale_fill_manual(values = c("dodgerblue3", "indianred3"))
```



Plotting the continuous variable winner (percentage of votes received by Trump, with 1 being 100% votes to Trump and 0 being 0% votes to Clinton.)

```
ggplot(usStateElection, aes(x = long, y = lat, group = group, fill = winner))  
+  
  geom_polygon(color = "gray90", size = 0.25) + theme_map() +  
  coord_map(projection = "lambert", lat0 = 30, lat1 = 40) +  
  scale_fill_gradient2(low = "white", mid = "blue", high = "red")
```

