IST 687
Homework 7
Due Date: 11/23

Code requires the following packages to run

```
library(ggplot2) #install.packages('ggplot2')
library(ggmap) #install.packages('ggmap')
library(gdata) #install.packages('gdata')
library(readxl) #install.packages('readxl')
library(zipcodeR) #install.packages('zipcodeR')
library(dplyr) #install.packages('dplyr')
library(usdata) #install.packages('usdata')
```

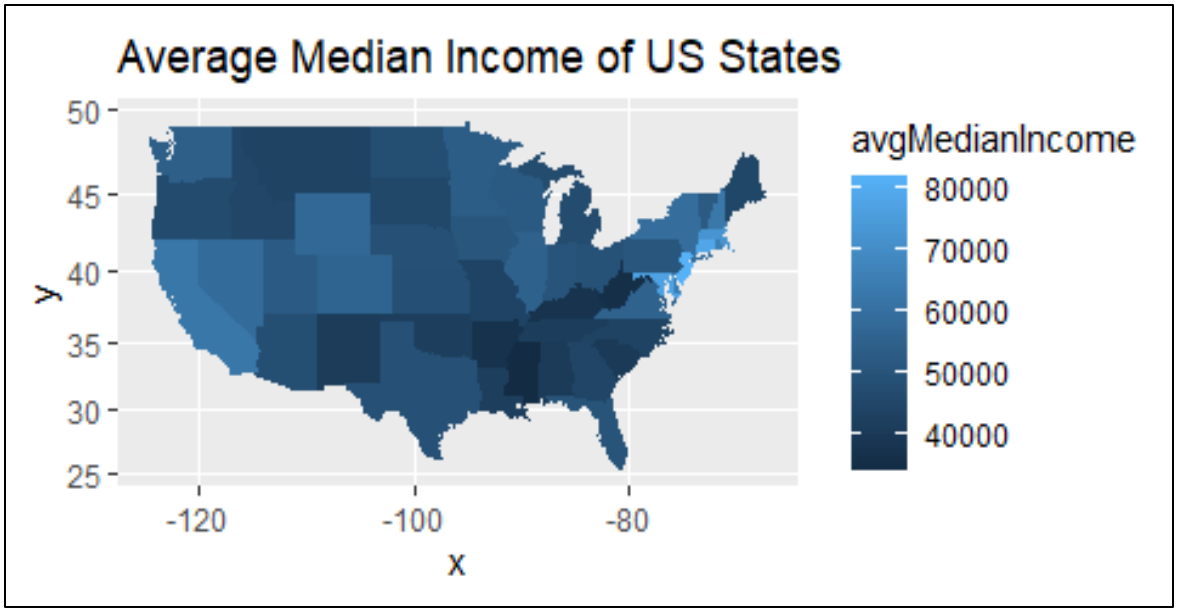
Step 1: Load the Data

```
#load income raw data
median_income_rawdata <- read_excel("MedianZIP-3.xlsx", sheet = "nation")
median_income_rawdata$Zip <- as.character(median_income_rawdata$Zip)

#function to add back leading zeros
addleadingzerosback <- function(value){
if (nchar(value) < 5) {return(paste("0", value, sep=""))}
else if (nchar(value) > 4) {return(value)}
return(value)}

#apply function to zips
median_income_rawdata$Zip <- sapply(median_income_rawdata$Zip, FUN = addleadingzerosback)

#load zip code raw data
zipcode_lookup <- zip_code_db #data from zipcodeR package
colnames(zipcode_lookup)[colnames(zipcode_lookup) == "zipcode"] <- "Zip" #change column header for l
ookup

#join tables and delete columns that are not needed
median_income_combined <- left_join(median_income_rawdata, zipcode_lookup, on = "Zip")
median_income_combined <- median_income_combined[,-5]
median_income_combined <- median_income_combined[,-6:-8]
median_income_combined <- median_income_combined[,-9:-11]
median_income_combined <- median_income_combined[,-10:-20]

#confirm no nulls
sum(as.numeric(is.null(median_income_combined$lat)))

## [1] 0
```

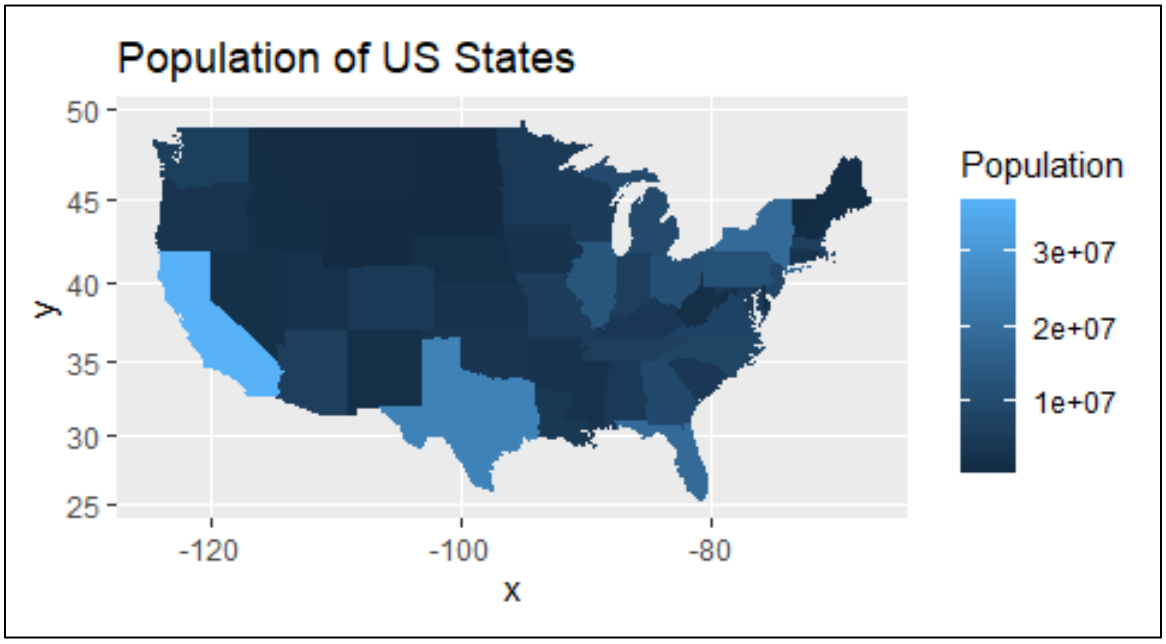
```
#remove Hawaii and Alaska
median_income_combined <- median_income_combined[-which(
median_income_combined$state=="AK" |
median_income_combined$state=="HI" |
median_income_combined$state=="DC"),]
"HI" %in% median_income_combined$state #validate
## [1] FALSE
"AK" %in% median_income_combined$state #validate
## [1] FALSE
"DC" %in% median_income_combined$state #validate
## [1] FALSE
```

Step 2: Show the income & population per state

```
#create avgMedianIncome dataframe
avgMedianIncomeByState <- data.frame(unlist(tapply(median_income_combined$Median, median_inco
me_combined$state, mean)))
avgMedianIncomeByState$state <- rownames(avgMedianIncomeByState)
rownames(avgMedianIncomeByState) <- NULL
colnames(avgMedianIncomeByState) <- c("avgMedianIncome", "state")

#create population dataframe
populationByState <- data.frame(unlist(tapply(median_income_combined$Pop, median_income_combin
ed$state, sum)))
populationByState$state <- rownames(populationByState)
rownames(populationByState) <- NULL
colnames(populationByState) <- c("Population", "state")

#join the two dataframes
simplerdataframe <- inner_join(populationByState, avgMedianIncomeByState, on="state")

#add the state name lower case
simplerdataframe$statename <- sapply(simplerdataframe$state, FUN = abbr2state)
simplerdataframe$statename <- tolower(simplerdataframe$statename)

#US map representing color with avg median income
USgeographicaldata <- map_data("state")
map1 <- ggplot(simplerdataframe, aes(map_id = statename))
map1 <- map1 + geom_map(map = USgeographicaldata, aes(fill=avgMedianIncome))
map1 <- map1 + expand_limits(x = USgeographicaldata$long, y=USgeographicaldata$lat)
map1 <- map1 + coord_map() + ggtitle("Average Median Income of US States")
map1
```

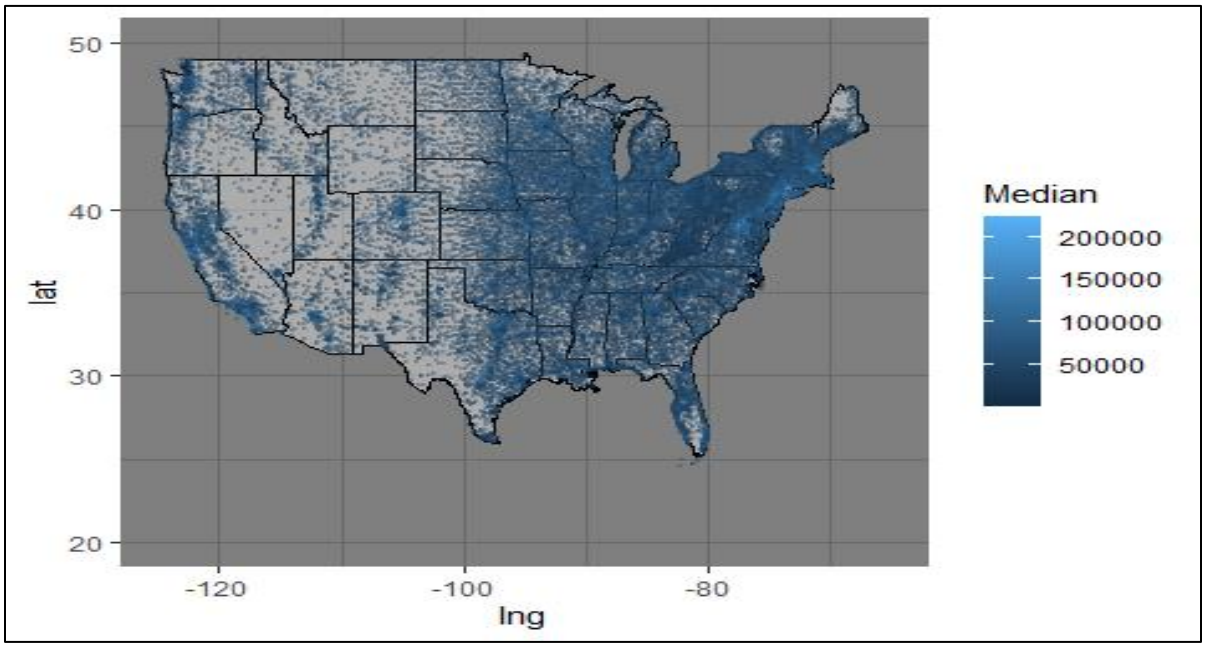



```
#US map representing color with population
USgeographicaldata <- map_data("state")
map2 <- ggplot(simplerdataframe, aes(map_id = statename))
map2 <- map2 + geom_map(map = USgeographicaldata, aes(fill=Population))
map2 <- map2 + expand_limits(x = USgeographicaldata$long, y=USgeographicaldata$lat)
map2 <- map2 + coord_map() + ggtitle("Population of US States")
map2
```



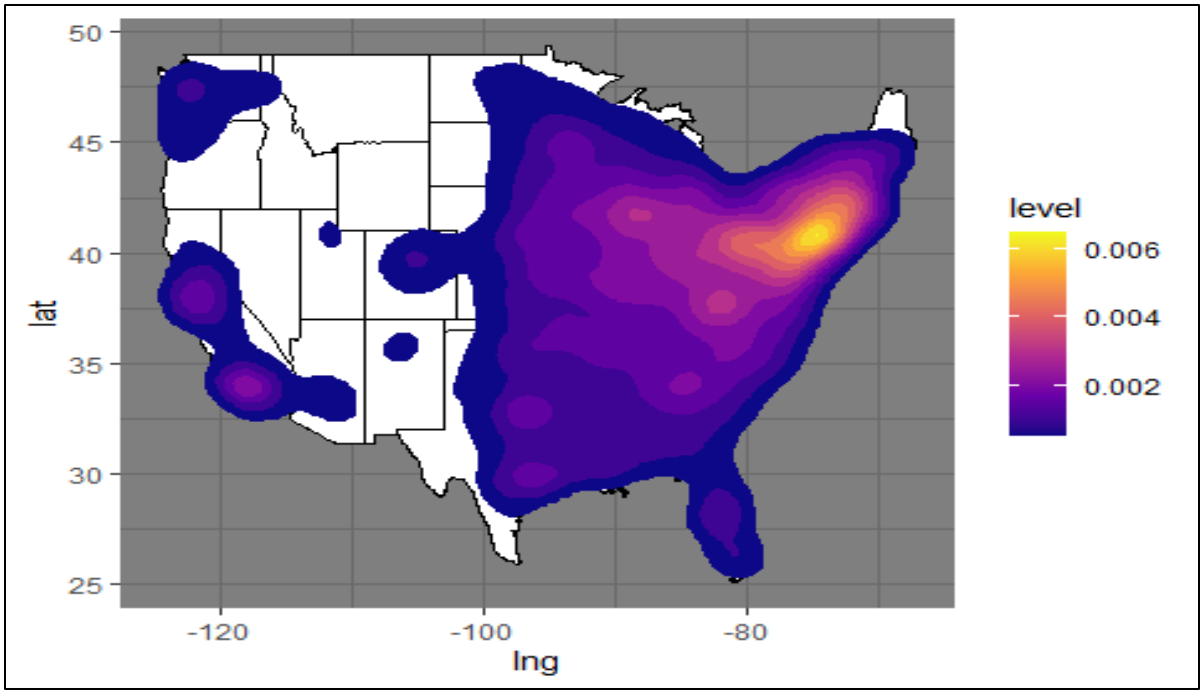
Step 3: Show the income per zip code

```
#Show the median income per zip code
USgeographicaldata <- map_data("state")
ggplot(median_income_combined, aes(lng, lat)) +
geom_polygon(data=USgeographicaldata, aes(x=long, y=lat, group=group), color='black', fill='white', alpha=.35) +
theme_dark() +
geom_point(aes(color=Median), size=.15, alpha=.25) +
xlim(-125, -65) + ylim(20,50)
```



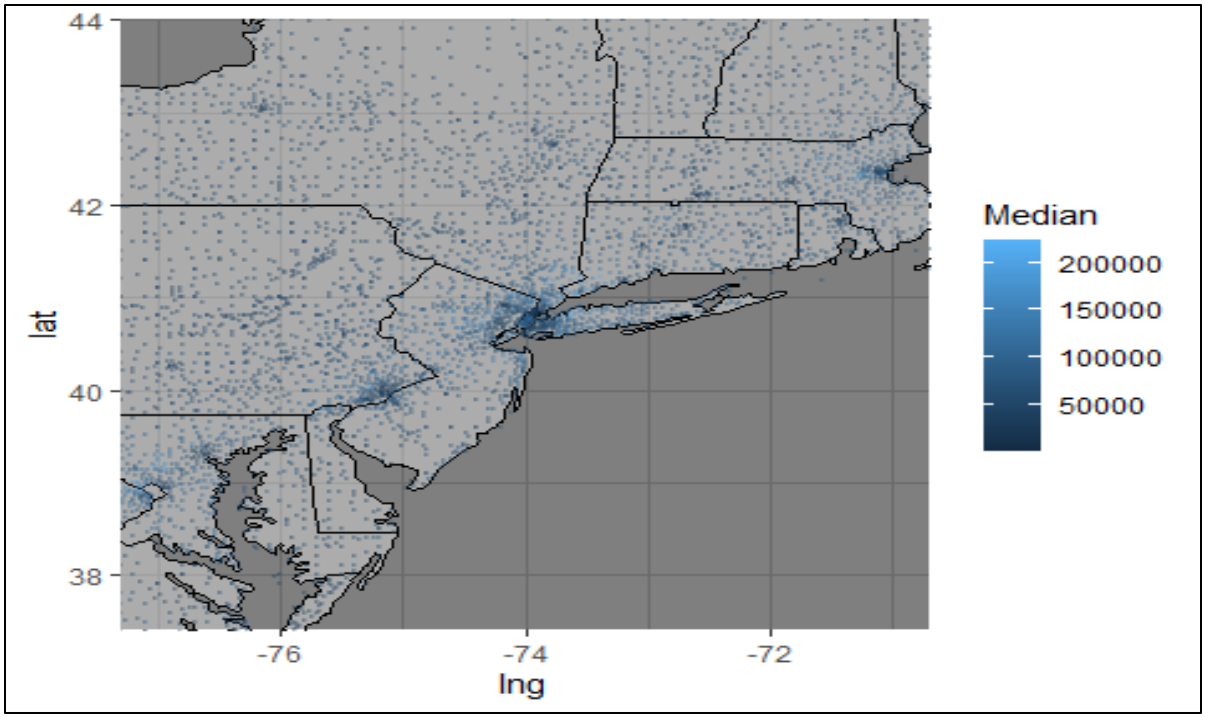
Step 4: Show the Zip Code Density

```
#Show the zip code density
USgeographicaldata <- map_data("state")
map3 <- ggplot(median_income_combined, aes(lng, lat))
map3 <- map3 + geom_polygon(data=USgeographicaldata, aes(x=long, y=lat, group=group), color='black', fill='white')
map3 <- map3 + stat_density2d(aes(fill = stat(level)), geom = "polygon")
map3 <- map3 + scale_fill_viridis_c(option = "plasma")
map3 <- map3 + theme_dark()
map3
```



Step 5: Zoom in to the region around NYC

```
#set zoom variables
zoomAmount <- 3
centerx <- -74.00594
centery <- 40.71278
ylimit <- c(centery-zoomAmount, centery+zoomAmount)
xlimit <- c(centerx-zoomAmount, centerx+zoomAmount)

#Step 3 zoomed
USgeographicaldata <- map_data("state")
step3zoomed <- ggplot(median_income_combined, aes(lng, lat))
step3zoomed <- step3zoomed + geom_polygon(data=USgeographicaldata, aes(x=long, y=lat, group=grou
p), color='black', fill='white', alpha=.35)
step3zoomed <- step3zoomed + geom_point(aes(color=Median), size=.15, alpha=.25)
step3zoomed <- step3zoomed + coord_cartesian(xlim=xlimit, ylim=ylimit)
step3zoomed <- step3zoomed + theme_dark()
step3zoomed
```




```
#Step 4 zoomed
step4zoomed <- ggplot(median_income_combined, aes(lng, lat))
step4zoomed <- step4zoomed + geom_polygon(data=USgeographicaldata, aes(x=long, y=lat, group=grou
p), color='black', fill='white')
step4zoomed <- step4zoomed + stat_density2d(bins=150, color="red")
step4zoomed <- step4zoomed + geom_point()
step4zoomed <- step4zoomed + coord_cartesian(xlim=xlimit, ylim=ylimit)
step4zoomed <- step4zoomed + scale_fill_viridis_c(option = "plasma")
step4zoomed <- step4zoomed + theme_dark()
step4zoomed
```