

HW7 – Marley Akonnor

#Step 1: Load the Data

#Access readxl function from the library

```
library(gdata)
```

```
library(usdata)
```

#Read the data

```
MedIncome <- read.xls("/Users/m/Documents/M.S Syracuse Data Science/Courses/IST 687 -  
Intro to Data Science/Week 7/MedianZIP_2_2.xlsx")
```

#Inspect the dataframe

```
head(MedIncome, 10)
```

#Clean up the dataframe

#a. Remove any info at the front of the file that's not needed

```
MedIncome <- MedIncome[-1, ]
```

```
rownames(MedIncome) <- NULL
```

#b. Update the column names (zip, median, mean, population)

```
ColumnNames <- c("zip", "median", "mean", "population")
```

```
names(MedIncome) <- ColumnNames
```

#upload the zipcode package

```
newZip <- zipcode
```

#Merge the zip code information from the two data frames (merge into one dataframe)

```
mergedZip <- merge(x = MedIncome, y = newZip, by = "zip")
```

#Remove Hawaii and Alaska (just focus on the 'lower 48' states)

```
mergedZip <- mergedZip[mergedZip$state != "AK", ]
```

```
mergedZip <- mergedZip[mergedZip$state != "HI", ]
```

```
finalZip <- mergedZip
```

```
finalZip <- na.omit(finalZip)
```

```
finalZip$statename <- abbr2state(finalZip$state)
```

```
finalZip$statename <- tolower(finalZip$statename)
```

#Step 2: Show the income & population per state

#install ggmap

```
install.packages("ggmap")
```

```
library(ggmap)
```

```
#install ggplot2
install.packages("ggplot2")
library(ggplot2)
```

```
#create numberize function
Numberize <- function(inputVector) {
  inputVector <- gsub(",", "", inputVector)
  inputVector <- gsub(" ", "", inputVector)
  return(as.numeric(inputVector))
}
```

```
#Create a simpler dataframe, with just the average median income and the the population for
each state.
```

```
SimplifiedZip <- data.frame(finalZip$median, finalZip$population)
```

```
#Add the state abbreviations and the state names as new columns (make sure the state names
are all lower case)
```

```
SimplifiedZip$state <- finalZip$state
SimplifiedZip$statename <- finalZip$statename
SimplifiedZip$statename <- tolower(SimplifiedZip$statename)
```

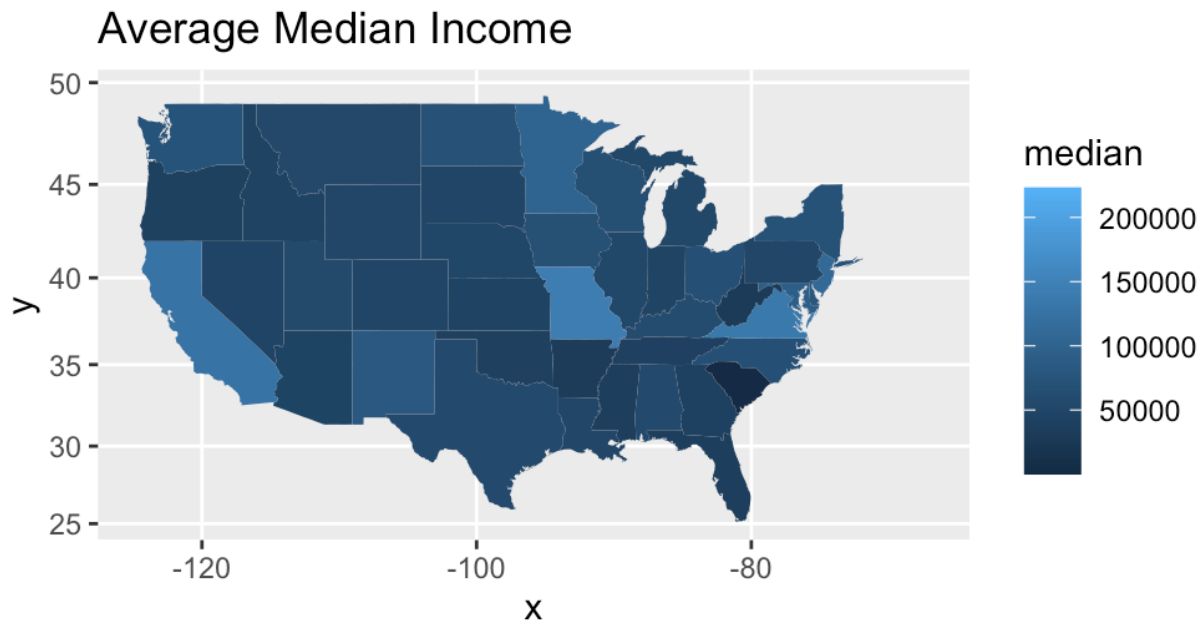
```
colnames(SimplifiedZip) <- (c('median', 'population', 'stateabbr', 'statename'))
```

```
SimplifiedZip$median <- Numberize(SimplifiedZip$median)
SimplifiedZip$population <- Numberize(SimplifiedZip$population)
```

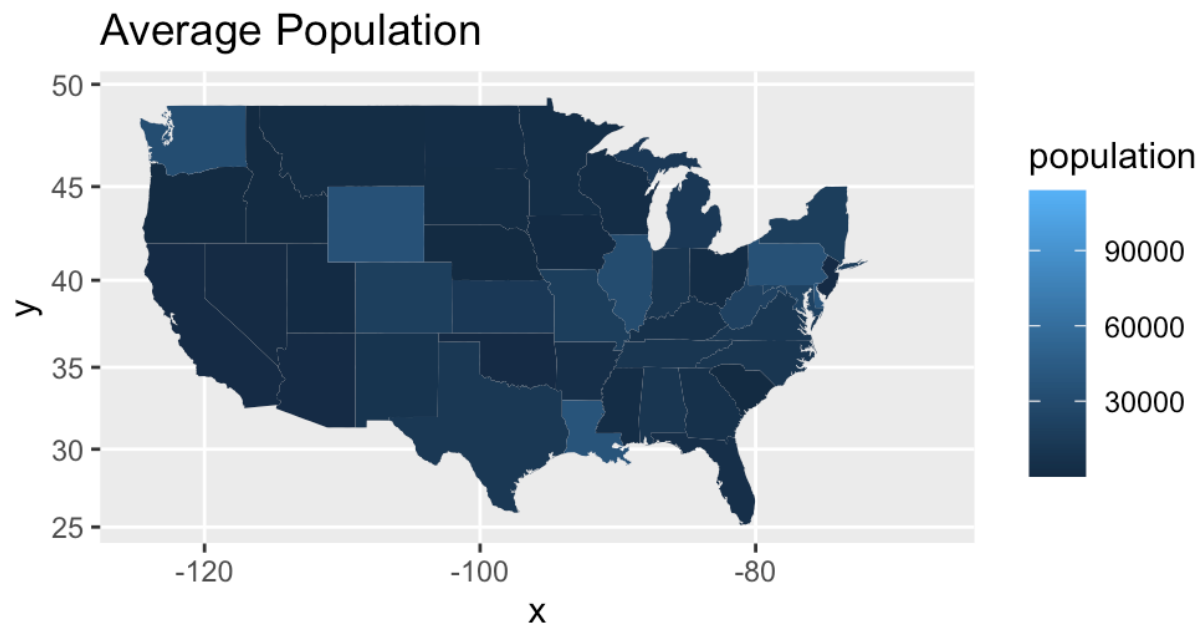
```
#Create map of the US
us <- map_data("state")
```

```
#Show the U.S. map, representing the color with the average median income of that state
```

```
map.med <- ggplot(SimplifiedZip, aes(map_id = statename))
map.med <- map.med + geom_map(map = us, aes(fill = median))
map.med <- map.med + expand_limits(x = us$long, y = us$lat)
map.med <- map.med + coord_map() + ggtitle("Average Median Income")
map.med
```



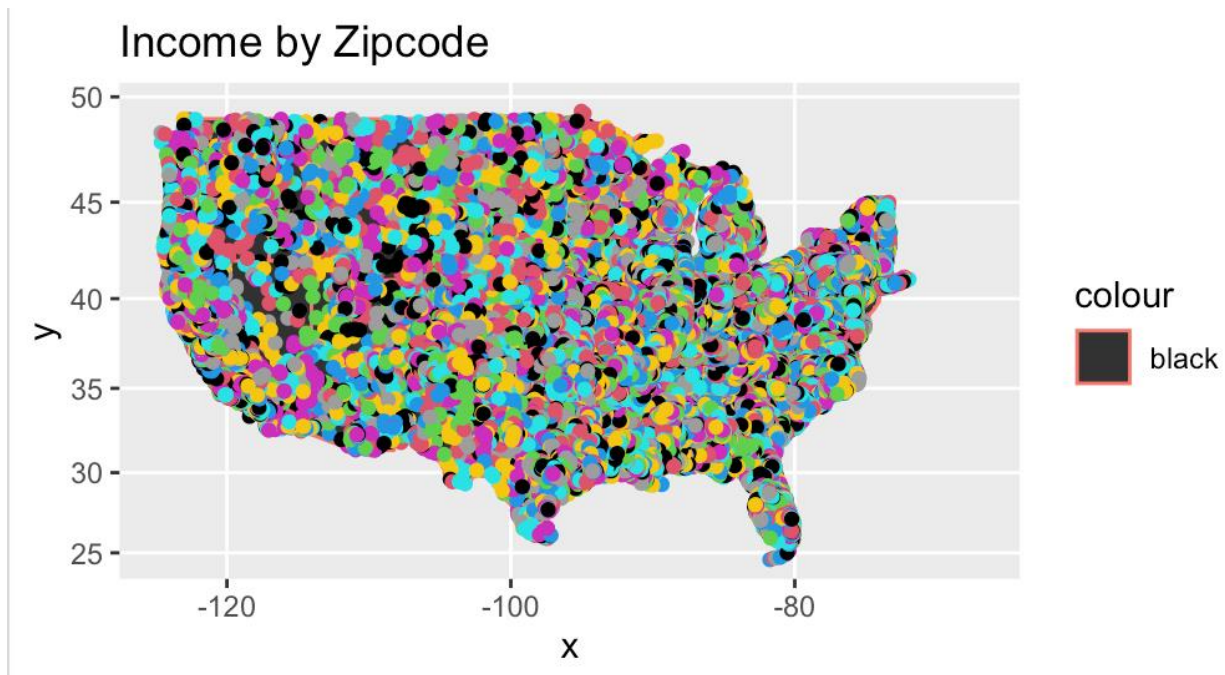
```
#Create a second map with color representing the population of the state
map.pop <- ggplot(SimplifiedZip, aes(map_id = statename))
map.pop <- map.pop + geom_map(map = us, aes(fill = population))
map.pop <- map.pop + expand_limits(x = us$long, y = us$lat)
map.pop <- map.pop + coord_map() + ggtitle("Average Population")
map.pop
```



#Step 3: Show the income per zip code

#Have draw each zip code on the map, where the color of the 'dot' is based on the median income. To make the map look appealing, have the background of the map be black.

```
map.zip <- ggplot(finalZip, aes(map_id = statename))
map.zip <- map.zip + geom_map(map = us, aes(color="black"))
map.zip <- map.zip + expand_limits(x = us$long, y = us$lat)
map.zip <- map.zip + geom_point(data=finalZip, aes(x=longitude, y=latitude),
color=finalZip$median)
map.zip <- map.zip + coord_map() + ggtitle("Income by Zipcode")
map.zip
```



#Step 4: Show Zip Code Density

#1) Now generate a different map, one where we can easily see where there are lots of zip codes, and where there are few (using the 'stat_density2d' function).

```
newDf <- data.frame(state.name, stringsAsFactors = FALSE)
newDf$state <- tolower(newDf$state.name)
map.zipDens <- ggplot(newDf, aes(map_id=state))
map.zipDens <- map.zipDens + geom_map(map=us, fill='white', color='black')
map.zipDens <- map.zipDens + expand_limits(x = us$long, y = us$lat)
```

```
NewDensMap <- map.zipDens + stat_density2d(aes(x=longitude, y=latitude), data=finalZip,
geom="polygon", na.rm = TRUE, contour = FALSE, inherit.aes = TRUE) +
  scale_fill_gradient(low="black",high="green")+
  scale_alpha(range=c(0.00,0.25))+
  ggtitle("Zipcode Density")
NewDensMap
```

#Step 5: Zoom in to the region around NYC

#Repeat steps 3 & 4, but have the image / map be of the northeast U.S. (centered around New York).

```
nycDf <- finalZip[finalZip$state=='NY',]
```

```
map.ny <- ggplot(nycDf, aes(map_id=state))
```

```
map.ny <- map.ny + expand_limits(x = us$long, y = us$lat)
```

```
map.ny <- map.ny + coord_map() + ggtitle('Income Northeast')
```

```
map.ny <- map.ny + geom_point(data=nycDf, aes(x=1, y=1), color=nycDf$Median)
```

```
map.ny
```

```
map.zipDens <- map.zipDens + expand_limits(x=xlimit, y=ylimit)
```

```
map.density <- map.zipDens + stat_density2d(aes(x = longitude, y = latitude, fill = ..level.., alpha  
= ..level..),
```

```
data = nycDf,
```

```
geom = 'polygon')+
```

```
scale_fill_gradient(low = 'blue', high = 'white') +
```

```
scale_alpha(range = c(0.05,1)) +
```

```
ggtitle('Zipcode Density') +
```

```
theme(plot.title = element_text(lineheight = 3.5, face = 'bold'))
```

```
map.density
```

