Sharat\_Sripada\_HW7.R

ssharat

2020-03-01

#  
# Course: IST-687  
# Name: Sharat Sripada  
# Homework #7  
# Due Date: 3/1/2020  
# Date Submitted: 3/1/2020  
# Topic: Map mashup!  
  
# install.packages("gdata")  
# install.packages("zipcode")  
# install.packages("openintro")  
# install.packages("ggmap")  
# install.packages("maps")  
# install.packages("mapproj")  
# install.packages("tmaptools")  
  
library("gdata")  
library("ggplot2")  
library("openintro")  
library("ggmap")  
library("maps")  
library("mapproj")  
library("tmaptools")  
  
# Read the xls  
medianzip <- read.xls("/Users/ssharat/Downloads/MedianZIP.xlsx")  
  
# Rename colnames  
colnames(medianzip) <- c("zip", "Median", "Mean", "Population")  
  
# Remove the first row  
medianzip <- medianzip[-1,]  
  
# Remove the commas in the Median, Mean & population columns  
medianzip$Median <- gsub(",","",medianzip$Median)  
medianzip$Mean <- gsub(",","",medianzip$Mean)  
medianzip$Population <- gsub(",","",medianzip$Population)  
  
# NOTE - zipcode has been archived in the CRAN repository  
# Download the package & install it via the .tar.gz  
library(zipcode)  
medianzip$zip <- clean.zipcodes(medianzip$zip)  
head(medianzip)

## zip Median Mean Population  
## 2 01001 56663 66688 16445  
## 3 01002 49853 75063 28069  
## 4 01003 28462 35121 8491  
## 5 01005 75423 82442 4798  
## 6 01007 79076 85802 12962  
## 7 01008 63980 78391 1244

head(zipcode)

## zip city state latitude longitude  
## 1 00210 Portsmouth NH 43.0059 -71.0132  
## 2 00211 Portsmouth NH 43.0059 -71.0132  
## 3 00212 Portsmouth NH 43.0059 -71.0132  
## 4 00213 Portsmouth NH 43.0059 -71.0132  
## 5 00214 Portsmouth NH 43.0059 -71.0132  
## 6 00215 Portsmouth NH 43.0059 -71.0132

df <- merge(medianzip, zipcode, by="zip")  
df$Median <- as.numeric(df$Median)  
df$Population <- as.numeric(df$Population)  
  
# Step-2: Create simpler data-frame  
# Data-frame 'dfmedian' <- Average median income by state  
income <- tapply(df$Median, df$state, mean)  
state <- rownames(income)  
dfmedian <- data.frame(state, income)  
  
# Data-frame 'dfpop' <- Population by state  
pop <- tapply(df$Population, df$state, sum)  
state <- rownames(income)  
dfpop <- data.frame(state, pop)  
  
# Create dfsimple merging the two DFs above by state  
dfSimple <- merge(dfmedian, dfpop, by="state")  
str(dfSimple)

## 'data.frame': 51 obs. of 3 variables:  
## $ state : Factor w/ 51 levels "AK","AL","AR",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ income: num 50451 40550 36961 48132 62629 ...  
## $ pop : num 703159 4770242 2936699 6360679 36927999 ...

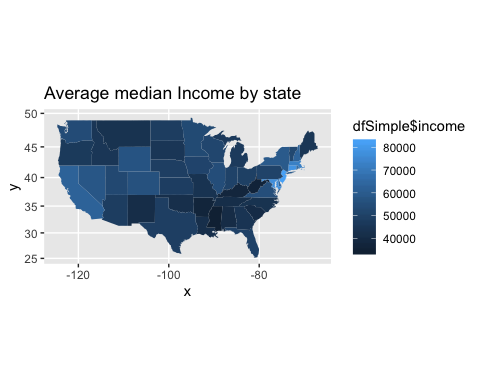
head(dfSimple)

## state income pop  
## 1 AK 50450.88 703159  
## 2 AL 40549.90 4770242  
## 3 AR 36960.95 2936699  
## 4 AZ 48132.07 6360679  
## 5 CA 62628.72 36927999  
## 6 CO 56303.02 4979279

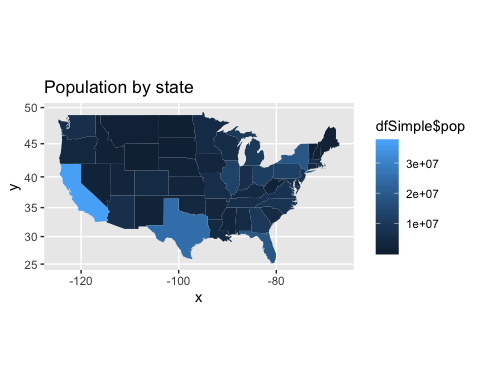
# Create a new column stateName in dfSimple  
dfSimple$stateName <- state.name[match(dfSimple$state, state.abb)]  
head(dfSimple)

## state income pop stateName  
## 1 AK 50450.88 703159 Alaska  
## 2 AL 40549.90 4770242 Alabama  
## 3 AR 36960.95 2936699 Arkansas  
## 4 AZ 48132.07 6360679 Arizona  
## 5 CA 62628.72 36927999 California  
## 6 CO 56303.02 4979279 Colorado

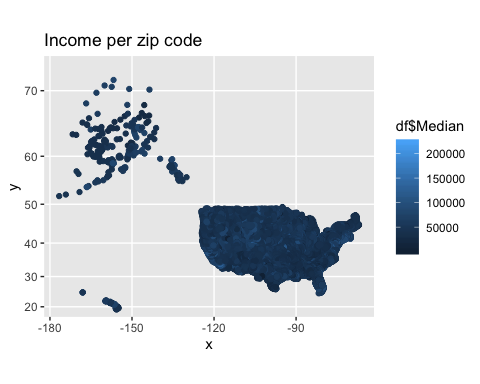
# Use tolower() on stateNames, since ggplot needs it that way   
dfSimple$stateName <- tolower(dfSimple$stateName)  
  
us <- map\_data('state')  
  
# Map average median income by states  
mapIncome <- ggplot(dfSimple, aes(map\_id = stateName))  
mapIncome <- mapIncome + geom\_map(map = us, aes(fill = dfSimple$income))  
mapIncome <- mapIncome + expand\_limits(x = us$long, y = us$lat)  
mapIncome <- mapIncome + coord\_map()  
mapIncome <- mapIncome + ggtitle("Average median Income by state")  
mapIncome



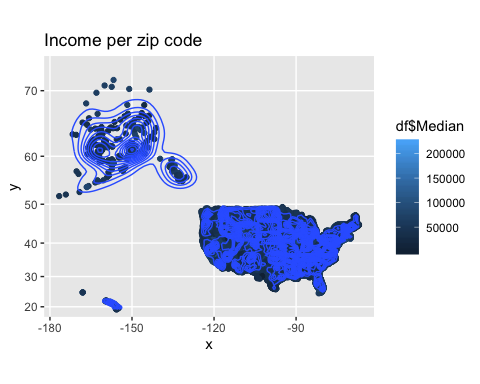
# Map population by states  
mapPop <- ggplot(dfSimple, aes(map\_id = stateName))  
mapPop <- mapPop + geom\_map(map = us, aes(fill = dfSimple$pop))  
mapPop <- mapPop + expand\_limits(x = us$long, y = us$lat)  
mapPop <- mapPop + coord\_map()  
mapPop <- mapPop + ggtitle("Population by state")  
mapPop



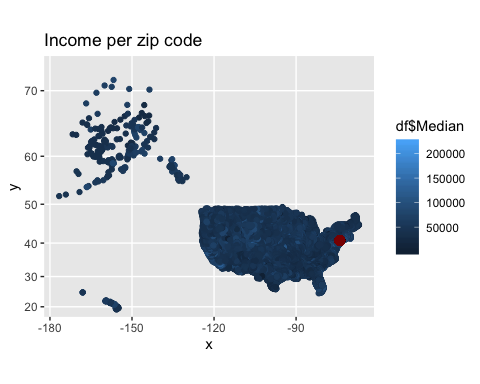
# Step-3: Income by zipcode  
df$stateName <- state.name[match(df$state,state.abb)]  
df$stateName <- tolower(df$stateName)  
mapZip <- ggplot(df, aes(map\_id = stateName))  
mapZip <- mapZip + geom\_map(map=us, fill="black", color="white")  
mapZip <- mapZip + expand\_limits(x = us$long, y = us$lat)  
mapZip <- mapZip + geom\_point(data = df, aes(x = df$longitude, y = df$latitude, color=df$Median))  
mapZip <- mapZip + coord\_map() + ggtitle("Income per zip code")  
mapZip



# Step-4: Zip-code density  
mapD <- mapZip + geom\_density\_2d(data = df, aes(x = df$longitude, y = df$latitude))  
mapD



# Step-5: Zoom-In - NYC  
Newlatlon <- function(address) {  
 raw\_latlon <- geocode\_OSM(address,   
 return.first.only=T,   
 server = "http://nominatim.openstreetmap.org"  
 )  
 # Create a new df  
 my\_df <- data.frame(raw\_latlon$coords[1], raw\_latlon$coords[2])  
 colnames(my\_df) <- c("lon", "lat")  
 return(my\_df)  
}  
  
addresses <- c("NYC, ny")  
latlon <- Newlatlon(addresses)  
  
mapZipZoomed <- mapZip + geom\_point(aes(x = latlon$lon, y = latlon$lat), color="darkred", size = 3)   
mapZipZoomed



mapDZoomed <- mapD + geom\_point(aes(x = latlon$lon, y = latlon$lat), color="darkred", size = 3)  
mapDZoomed

