Assignment 3

Esther Chen

2022-10-16

##Set Up

#Load packages  
library(rvest)  
library(sf)  
library(tidyverse)  
library(jsonlite)  
library(tidycensus)  
library(ggplot2)

#Part 1

#Read in data  
emergencyVisits <- fromJSON("https://data.cityofnewyork.us/resource/2nwg-uqyg.json?$limit=3000")

##Question 1

emergencyVisits$date <- as.Date(emergencyVisits$date)

emergencyVisits$date

##Question 2

head(emergencyVisits)

The number of ED visits is a character.

#Change the character to integer  
emergencyVisits$total\_ed\_visits <- strtoi(emergencyVisits$total\_ed\_visits)

##Question 3

#Order by date  
emergencyVisits <- arrange(emergencyVisits, date)

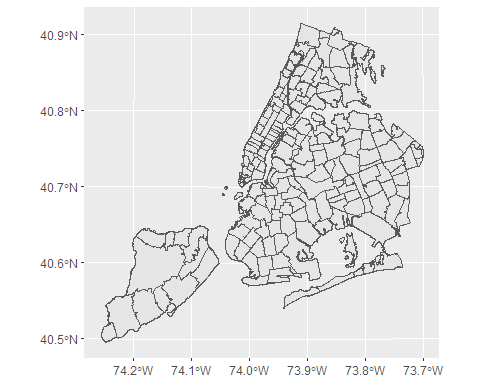
#Most recent 300 visits  
recentvisits <- emergencyVisits[2701:3000,]

##Question 4

#Visits per zip code  
totalrecentvisits <- recentvisits %>%  
 group\_by(mod\_zcta) %>%  
 summarise(Total = sum(total\_ed\_visits))  
  
totalvisits <- emergencyVisits %>%  
 group\_by(mod\_zcta) %>%  
 summarise(Total = sum(total\_ed\_visits))

#Part 2

#Map of zipcodes  
NYCmap = st\_read("C:/Users/jk/Documents/ZIP\_CODE\_040114.shp")  
  
ggplot(sample\_frac(NYCmap)) + geom\_sf()



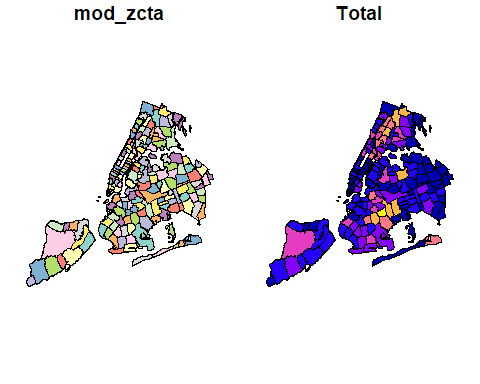
#Part 3

##Question 6

#Joining the data  
NYCmapChr = mutate(NYCmap, mod\_zcta = as.character(ZIPCODE))  
NYCmapZip = select(NYCmapChr, mod\_zcta, geometry)  
  
recentvisitsmap <- merge(NYCmapZip, recentvisits, by="mod\_zcta")  
  
totalvisitsmap <-merge(NYCmapZip, totalvisits, by="mod\_zcta")  
  
totalrecentvisitsmap <- merge(NYCmapZip, totalrecentvisits)

##Question 7

plot(totalvisitsmap)



#Part 4

#Set API key  
census\_api\_key("98e447b6bc529e5f36793dc8ff9dac0b8981f37a")  
  
#Get data from api  
censusData = load\_variables(2018, "acs5", cache=T)  
populationData = get\_acs(geography = "zcta",  
variables = 'B01003\_001',  
geometry = FALSE)  
  
#Remove the extra tag  
populationData$mod\_zcta = sub("ZCTA5 ", "", populationData$NAME)  
  
  
#Rename the column  
populationData =  
populationData %>%  
rename(population = estimate)

##Question 8

#Merge the population data with my main dataset  
  
populationvisits <- merge(totalvisitsmap, populationData, by="mod\_zcta")

##Question 9

#Calculate visits per inhabitant  
populationvisits <- populationvisits %>%  
 mutate(visits\_per\_inhabitant = Total/population)

#Part 5

#Download poverty data from api  
vars = c(poverty = 'B17001\_002')  
povertyData = get\_acs(geography = "zcta",  
variables = vars,  
geometry = FALSE)  
  
povertyData$mod\_zcta = sub("ZCTA5 ", "", povertyData$NAME)  
  
povertyData = povertyData %>% rename(povertyCount = estimate)

#Part 6 ##Question 10

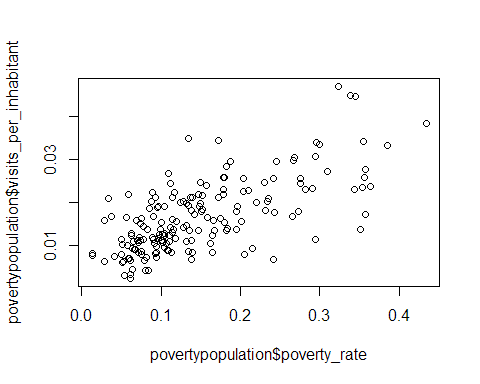
#Inner join with population data  
povertypopulation <- merge(populationvisits, povertyData, by="mod\_zcta")

##Question 11

#Calculate the poverty rate  
povertypopulation = povertypopulation %>%  
 mutate(poverty\_rate = povertyCount/population)

#Part 7 ##Question 12-14

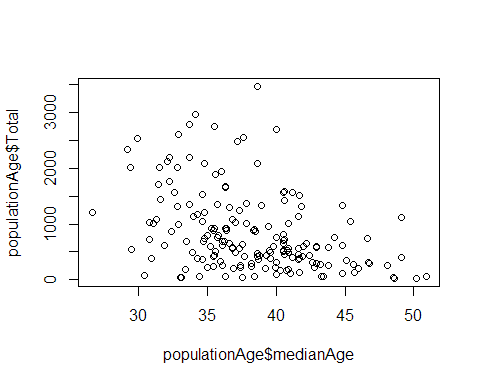
#Analyze the relationship between poverty rate and the number of visits  
  
povertyVisits <- lm(visits\_per\_inhabitant ~ poverty\_rate, data = povertypopulation)  
  
summary(povertyVisits)  
  
plot(x = povertypopulation$poverty\_rate, y = povertypopulation$visits\_per\_inhabitant)



There is a slight positive relationship between the poverty rate of a zip code and the number of emergency room visits per person.

#Part 8 #Question 15 and 16

#Get data from API  
vars = c(age = 'B01002\_001')  
ageData = get\_acs(geography = "zcta", variables = vars, geometry = FALSE)  
  
ageData$mod\_zcta = sub("ZCTA5 ","",ageData$NAME)  
  
ageData = ageData %>%  
 rename(medianAge = estimate)  
  
#Merge data  
populationAge <- merge(populationvisits, ageData, by="mod\_zcta")  
  
#Analyze relationship  
ageVisits <- lm(Total ~ medianAge, data = populationAge)  
  
plot(populationAge$medianAge, populationAge$Total)



summary(ageVisits)

There is a negative association between the median age of a zip code and the total number of hospital visits in that zip code. #Part 9 ##Question 17 and 18

vars = c('B02001\_002')  
ethnicityData = get\_acs(geography = "zcta", variables = vars, geometry = FALSE)  
ethnicityData = ethnicityData %>% rename(whiteAloneCount = estimate)  
ethnicityData$mod\_zcta = sub("ZCTA5 ","",ethnicityData$NAME)  
  
#Merge data and calculate percentage of population that stated they are "White Alone".  
  
populationEthnicity <- merge(populationvisits, ethnicityData, by="mod\_zcta")   
  
populationEthnicity <- populationEthnicity %>%  
 mutate(percentWhite = whiteAloneCount/population)

populationDat <- populationvisits %>%  
 mutate(percentWhite = populationEthnicity$percentWhite)  
  
populationDat <- populationDat %>%  
 mutate(medianAge = populationAge$medianAge)  
  
populationDat <- populationDat %>%  
 mutate(povertyRate = povertypopulation$poverty\_rate)

#Part 10 ##Question 19 and 20

#Multilinear regression and analysis  
mlr <- lm(formula = visits\_per\_inhabitant ~ percentWhite + medianAge + povertyRate, data= populationDat)  
  
summary(mlr)

By taking a look at the coefficients, we can see that being white and having a higher median age actually has slightly negative association with the rate of emergency room visits. And we can see that poverty has a positive relationship with emergency room visits. It looks like the factor with the largest association is poverty. I would say this makes sense because a person’s ability to get preventative care is deeply related to their economic status. Therefore it could lead to more emergency and severe cases of illness before they finally seek treatment.