Homework #4

Sang Doan

10/15/2020

Problem 1

See `data_munge.R`.

Problem 2

See `data.R`.

Problem 3

```
pop <- getPop()
facts <- getFacts()
deaths <- getDeaths()
head(get_counties(pop))
```

```
##           county state
## 1 Autauga County    AL
## 2 Baldwin County    AL
## 3 Barbour County    AL
## 4    Bibb County    AL
## 5  Blount County    AL
## 6 Bullock County    AL
```

Problem 4

```
dem1 <- 'AGE775214'
head(map_counties(deaths, pop, facts, dem1))
```

```
##           county state    deathRate statistic
## 1 Autauga County    AL 0.0004295763      13.8
## 2 Baldwin County    AL 0.0002060618      18.7
## 3 Barbour County    AL 0.0002835615      16.5
## 4    Bibb County    AL 0.0004018934      14.8
## 5  Blount County    AL 0.0002075191      17.0
## 6 Bullock County    AL 0.0013860014      14.9
```

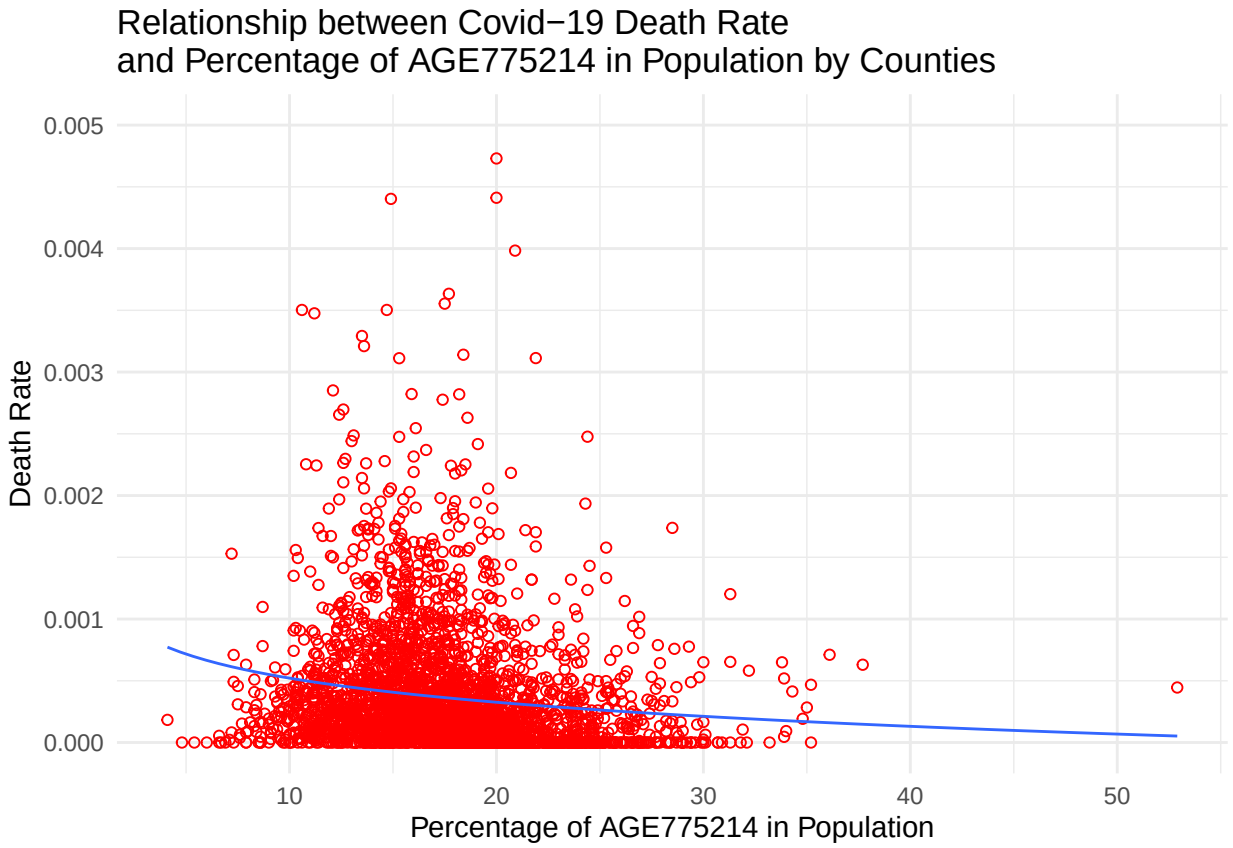
Problem 5

Five demographics I chose are:
(1) AGE775214 Persons 65 years and over, percent, 2014
(2) EDU685213 Bachelor's degree or higher, percent of persons age 25+, 2009-2013
(3) RHI225214 Black or African American alone, percent, 2014
(4) RHI325214 American Indian and Alaska Native alone, percent, 2014

(5) RHI125214 White alone, percent, 2014

People 65 years and over

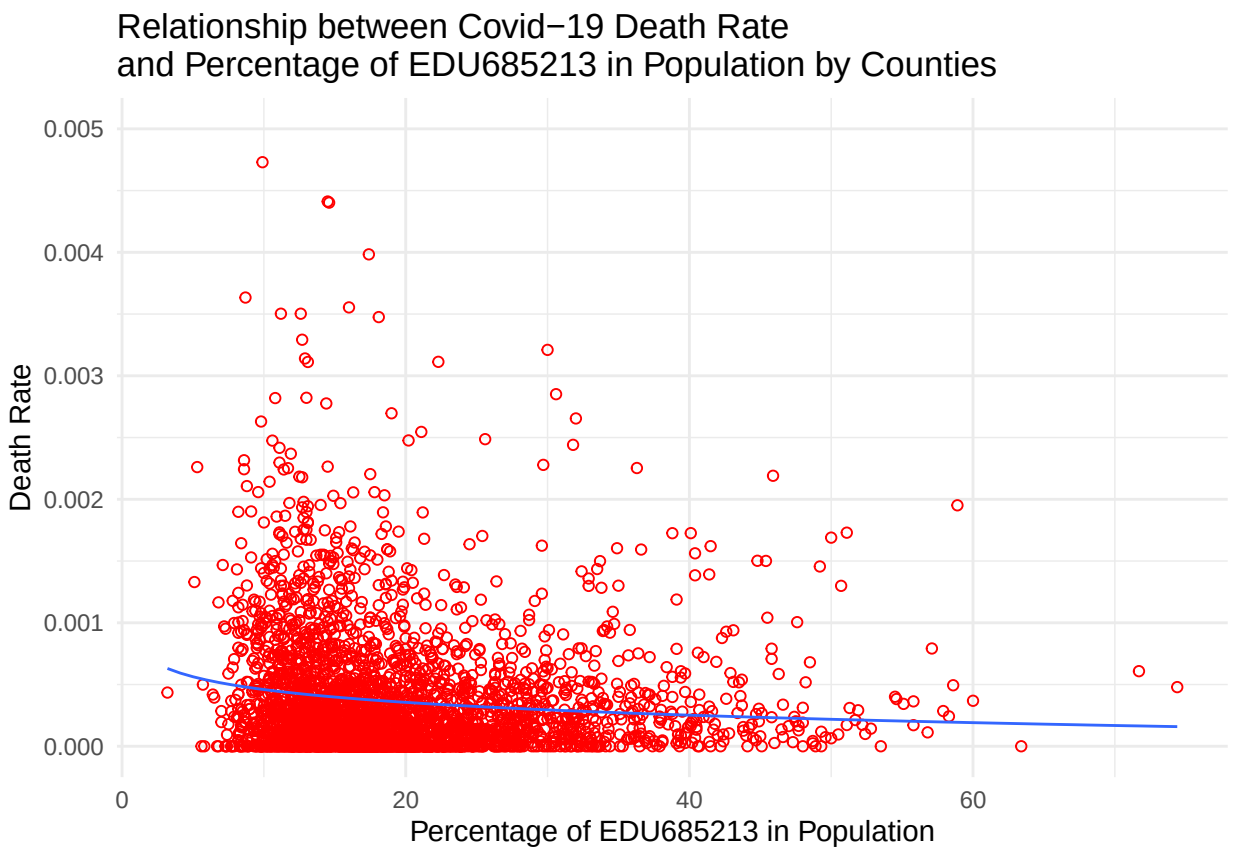
```
plotDem(deaths, pop, facts, 'AGE775214') %>% format(scientific = F)
```




```
## [1] "-0.00001802184"
```

People with a Bachelor's degree or higher

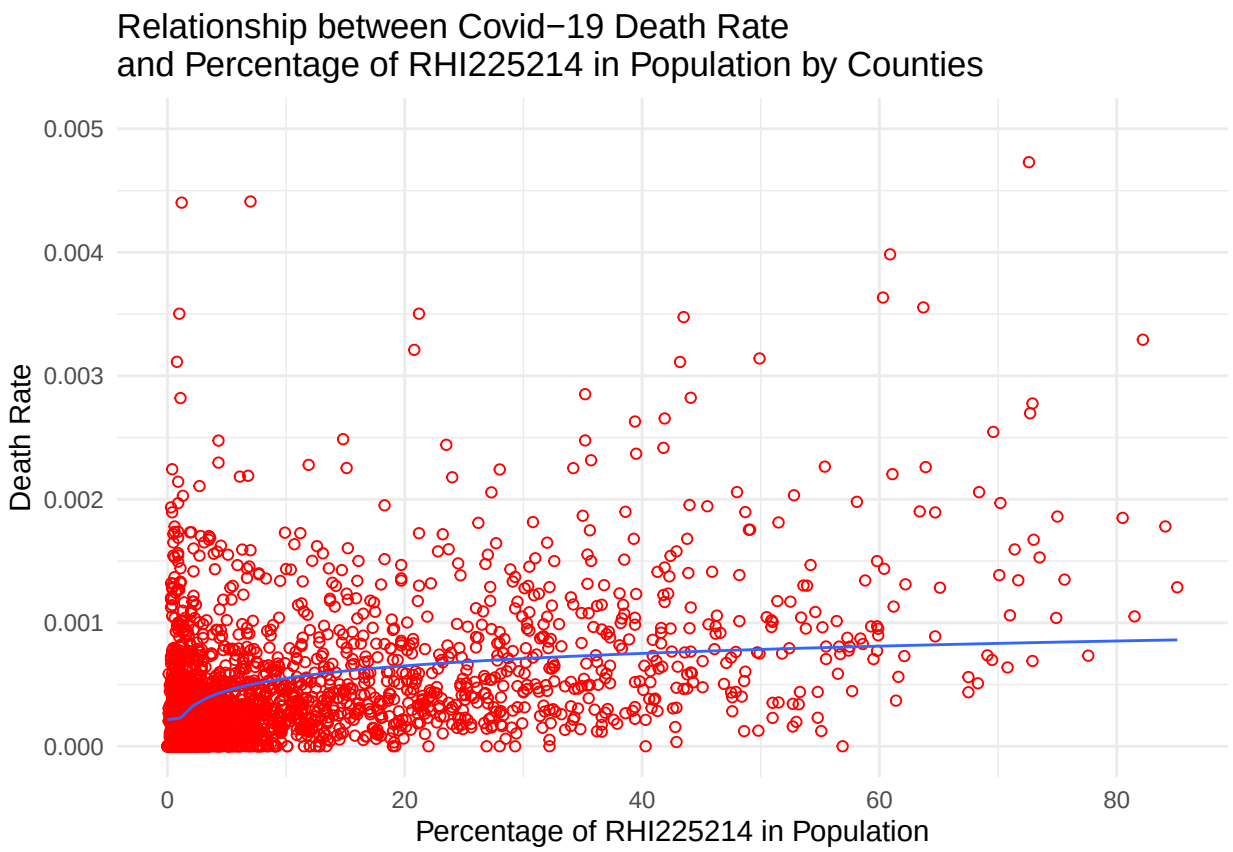
```
plotDem(deaths, pop, facts, 'EDU685213') %>% format(scientific = F)
```




```
## [1] "-0.00000465274"
```

Black or African Americans

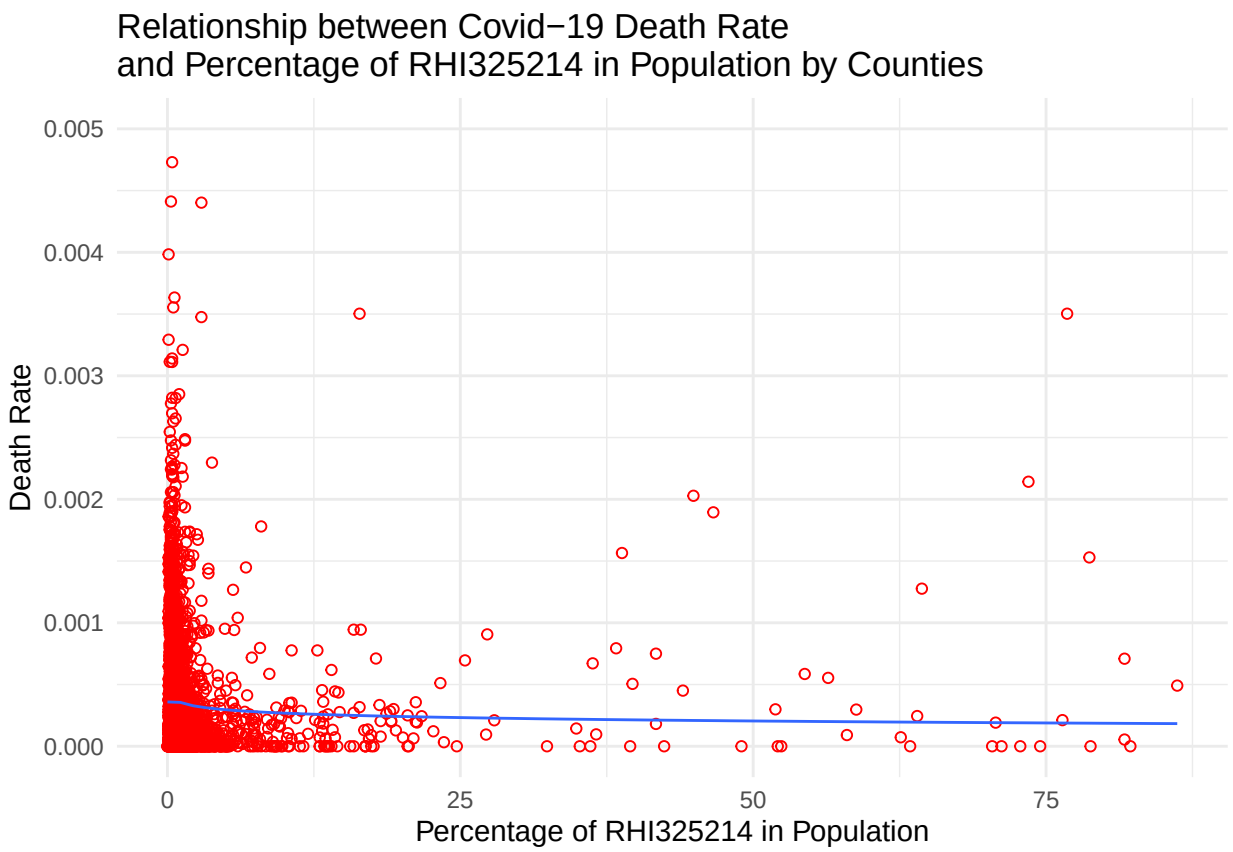
```
plotDem(deaths, pop, facts, 'RHI225214') %>% format(scientific = F)
```




```
## [1] "0.00001762311"
```

American Indians and Alaska Natives

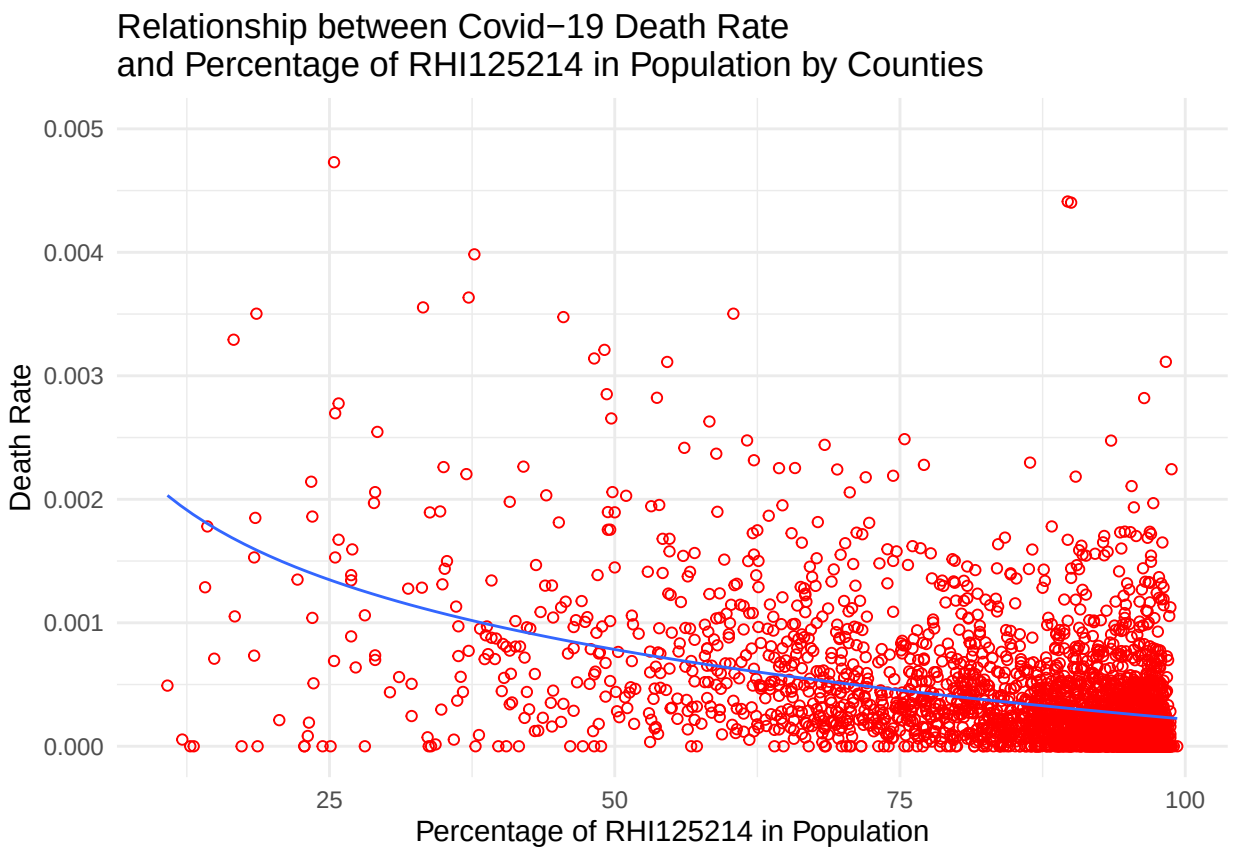
```
plotDem(deaths, pop, facts, 'RHI325214') %>% format(scientific = F)
```




```
## [1] "0.00000004150738"
```

Whites

```
plotDem(deaths, pop, facts, 'RHI125214') %>% format(scientific = F)
```




```
## [1] "-0.00001406819"
```

Old age is the best predictor of Covid-19 deaths among the five demographics, but all correlations assessed above are insignificant.

Code

data_munge.R

```
deaths <- read.csv('../raw_data/covid_deaths.csv', header=T, stringsAsFactors = F)
facts <- read.csv('../raw_data/county_facts.csv', header=T, stringsAsFactors = F)
pop <- read.csv('../raw_data/county_population.csv', header=T, stringsAsFactors = F)

names(deaths)[1:3] <- c('FIPS', 'county', 'state')
names(facts)[1:3] <- c('FIPS', 'county', 'state')
names(pop)[1:3] <- c('FIPS', 'county', 'state')

write.csv(deaths, '../processed_data/processed_covid_deaths.csv', row.names = F)
write.csv(facts, '../processed_data/processed_county_facts.csv', row.names = F)
write.csv(pop, '../processed_data/processed_county_population.csv', row.names = F)
```

data.R

```
getFacts <- function() {
  return(read.csv('../processed_data/processed_county_facts.csv', header = T,
                  stringsAsFactors = F))
}

getDeaths <- function() {
  return(read.csv('../processed_data/processed_covid_deaths.csv', header = T,
                  stringsAsFactors = F))
}

getPop <- function() {
  return(read.csv('../processed_data/processed_county_population.csv',
                  header = T, stringsAsFactors = F))
}
```

analysis.R

```
get_counties <- function(d) {
  return(dplyr::select(d, county, state))
}

map_counties <- function(cd, cp, cf, demographic) {

  #It appears that all data frames are already arranged in an ascending order by FIPS.
  #In such case, the following three lines are unnecessary.
  cd <- cd %>% arrange(FIPS)
  cf <- cf %>% arrange(FIPS)
  cp <- cp %>% arrange(FIPS)

  counties <- get_counties(cd)
  counties$deathRate <- rowSums(select(cd, -FIPS, -county, -state, -stateFIPS)) / cp$population
  counties$statistic <- cf[[demographic]]

  return(counties)
}

plotDem <- function(cd, cp, cf, dem) {
  thisDem <- select(map_counties(cd, cp, cf, dem), deathRate, statistic)
```

```
  thisPlot <- ggplot(data = thisDem, aes(x = statistic, y = deathRate)) +
    ggtitle(paste('Relationship between Covid-19 Death Rate \nand Percentage of', dem,
                  'in Population by Counties')) +
    geom_point(color = 'red', size = 1.5, fill = 'blue', shape = 1, stroke = 0.5) +
    xlab(paste('Percentage of', dem, 'in Population')) +
    geom_smooth(method = 'lm', formula = y ~ ifelse(is.finite(log(x)), log(x), 0),
                se = F, size = 0.5) +
    scale_y_continuous(name = 'Death Rate', limits = c(0, 0.005)) +
    theme_minimal()

  print(thisPlot)

  return(coef(lm(deathRate ~ statistic, thisDem))[[2]])

}
```

config.R

```
source('data.R')
source('analysis.R')

library(tidyverse)
library(magrittr)
```