

Analysis

Prashan A. Welipitiya

This markdown is just an overview of the process that I used to get my information.

Data

I'm going to be pulling the covid numbers from the New York Times Github. They have total numbers of cases and deaths per county and per state.

```
#https://github.com/nytimes/covid-19-data
setwd("C:/Users/Prashan.Welipitiya/Documents/Prashan/covid-19-data")
covid_states = read.csv("us-states.csv")
covid_states$date <- ymd(covid_states$date)
covid_counties = read.csv("us-counties.csv")
covid_counties$date <- ymd(covid_counties$date)

#head(covid_counties)

td_counties <- covid_counties %>% filter(date == max(covid_counties$date))
head(td_counties)

##           date   county   state fips cases deaths
## 1 2020-07-01 Autauga Alabama 1001    553     12
## 2 2020-07-01 Baldwin Alabama 1003    703     10
## 3 2020-07-01 Barbour Alabama 1005    326      1
## 4 2020-07-01 Bibb Alabama 1007    174      1
## 5 2020-07-01 Blount Alabama 1009    218      1
## 6 2020-07-01 Bullock Alabama 1011    367     10
```

I'm going to be using a dataset that was part of a homework assignments in an old class. This dataset has a lot of important information that I am curious about on counties in the US. It includes percent populations of 2016 voting information, elderly, black, white, hispanic, asian, education and income.

```
county_votes16 <- readRDS(url("https://ericwfox.github.io/data/county_votes16.rds"))
#head(county_votes16)

# To match the New York Times data, I'm going to add a column that changes the state abbreviation to two letters.
county_votes16$state <- state.name[match(county_votes16$state, state.abb)]

# And take the word county out of the county names.
county_votes16$county <- as.character(county_votes16$county)
county_votes16$county <- substr(county_votes16$county, 1, nchar(county_votes16$county) - 7)

head(county_votes16)
```

```

##      state county clinton_pctvotes trump_pctvotes obama_pctvotes pct_pop65
## 1 Alabama Autauga          23.96        73.44       26.58     13.8
## 2 Alabama Baldwin          19.57        77.35       21.57     18.7
## 3 Alabama Barbour         46.66        52.27       51.25     16.5
## 4 Alabama Bibb            21.42        76.97       26.22     14.8
## 5 Alabama Blount           8.47        89.85       12.35     17.0
## 6 Alabama Bullock         75.09        24.23       76.31     14.9
##   pct_black pct_white pct_hispanic pct_asian highschool bachelors income
## 1     18.7    77.9        2.7      1.1      85.6    20.9 53.682
## 2      9.6    87.1        4.6      0.9      89.1    27.7 50.221
## 3     47.6    50.2        4.5      0.5      73.7    13.4 32.911
## 4     22.1    76.3        2.1      0.2      77.5    12.1 36.447
## 5      1.8    96.0        8.7      0.3      77.0    12.1 44.145
## 6     70.1    26.9        7.5      0.3      67.8    12.5 32.033
##   trump_win
## 1
## 2
## 3
## 4
## 5
## 6

```

```

covid = merge(td_counties, county_votes16)
head(covid)

```

```

##      county      state      date fips cases deaths clinton_pctvotes
## 1 Abbeville South Carolina 2020-07-01 45001 113 0 34.61
## 2 Acadia      Louisiana 2020-07-01 22001 919 37 20.59
## 3 Accomack    Virginia 2020-07-01 51001 1043 14 42.76
## 4 Ada          Idaho 2020-07-01 16001 2288 23 38.69
## 5 Adair        Iowa 2020-07-01 19001 15 0 29.98
## 6 Adair        Kentucky 2020-07-01 21001 126 19 16.08
##   trump_pctvotes obama_pctvotes pct_pop65 pct_black pct_white pct_hispanic
## 1     62.87        42.61      19.4      28.3      69.7      1.2
## 2     77.26        24.44      13.5      18.3      79.6      2.2
## 3     54.47        47.77      21.3      28.1      68.8      8.9
## 4     47.93        42.72      12.6      1.3      92.4      7.7
## 5     65.34        45.16      22.1      0.4      98.2      1.7
## 6     80.66        21.84      16.4      3.0      95.1      1.9
##   pct_asian highschool bachelors income trump_win
## 1     0.4      76.8      12.2 35.947      1
## 2     0.4      72.1      10.2 37.587      1
## 3     0.7      78.0      17.2 39.328      1
## 4     2.6      93.6      36.0 55.210      1
## 5     0.5      90.7      16.3 47.892      1
## 6     0.3      73.7      13.9 32.524      1

```

I need to add population data. I'm getting this from census.gov (https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html#par_textimage_70769902)

```

pop = read.csv('https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/counties/totals/co-e')

```

```

# I just want the county info, state name and population.
pop19 <- subset(pop, select = c(COUNTY, STNAME, CTYNAME, POPESTIMATE2019))

# County number of 0 is the state population so I'm going to take that out for now.
popCounty <- pop19 %>%
  filter(COUNTY != 0)

# Change headers to match

popCounty <- popCounty %>% rename(ID = COUNTY)
popCounty <- popCounty %>% rename(state = STNAME)
popCounty <- popCounty %>% rename(county = CTYNAME)
popCounty <- popCounty %>% rename(popEst19 = POPESTIMATE2019)

# Take the word county out again
popCounty$county <- substr(popCounty$county, 1, nchar(popCounty$county) - 7)

head(popCounty)

##   ID state county popEst19
## 1  1 Alabama Autauga    55869
## 2  3 Alabama Baldwin   223234
## 3  5 Alabama Barbour   24686
## 4  7 Alabama Bibb     22394
## 5  9 Alabama Blount    57826
## 6 11 Alabama Bullock   10101

covid <- merge(popCounty, covid)

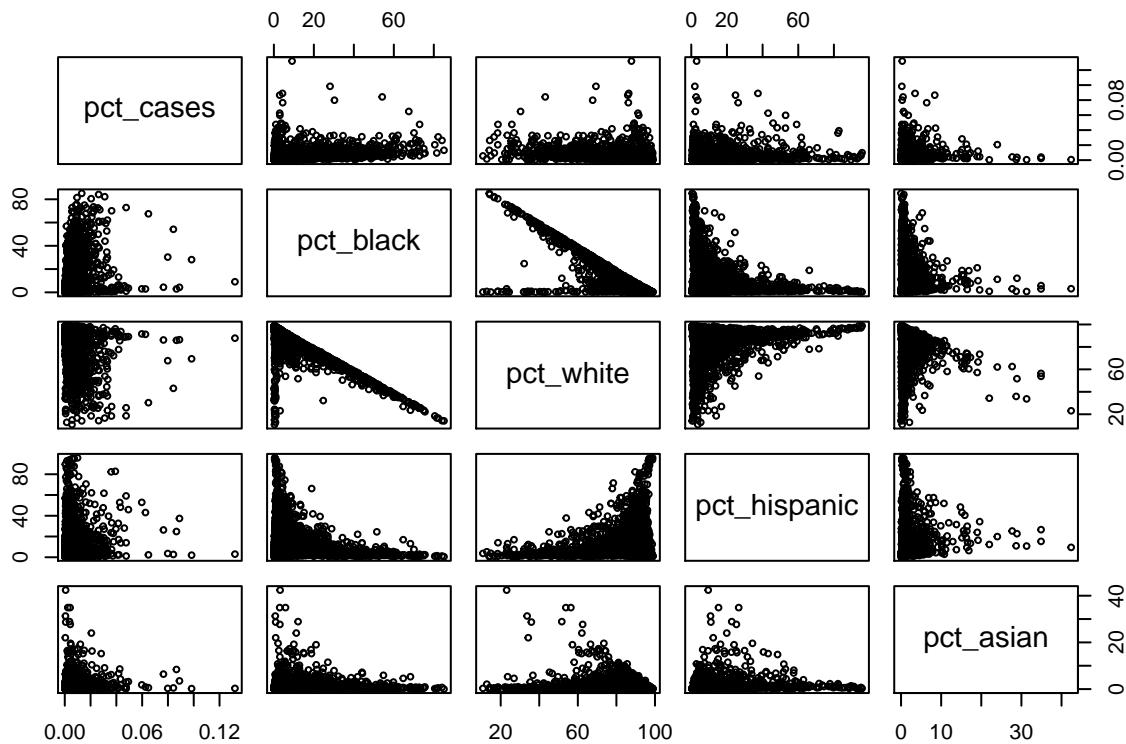
covid$pct_cases <- covid$cases / covid$popEst19
covid$pct_deaths <- covid$deaths / covid$popEst19
head(covid)

##      state county ID popEst19      date fips cases deaths clinton_pctvotes
## 1 Alabama Autauga  1    55869 2020-07-01 1001    553    12        23.96
## 2 Alabama Baldwin  3   223234 2020-07-01 1003    703    10        19.57
## 3 Alabama Barbour  5   24686 2020-07-01 1005    326     1        46.66
## 4 Alabama Bibb    7   22394 2020-07-01 1007    174     1        21.42
## 5 Alabama Blount   9   57826 2020-07-01 1009    218     1         8.47
## 6 Alabama Bullock 11  10101 2020-07-01 1011    367    10        75.09
##      trump_pctvotes obama_pctvotes pct_pop65 pct_black pct_white pct_hispanic
## 1            73.44          26.58     13.8      18.7     77.9       2.7
## 2            77.35          21.57     18.7      9.6      87.1       4.6
## 3            52.27          51.25     16.5      47.6     50.2       4.5
## 4            76.97          26.22     14.8      22.1     76.3       2.1
## 5            89.85          12.35     17.0      1.8      96.0       8.7
## 6            24.23          76.31     14.9      70.1     26.9       7.5
##      pct_asian highschool bachelors income trump_win  pct_cases  pct_deaths
## 1            1.1          85.6      20.9  53.682      1 0.009898155 2.147882e-04
## 2            0.9          89.1      27.7  50.221      1 0.003149162 4.479604e-05
## 3            0.5          73.7      13.4  32.911      1 0.013205866 4.050879e-05
## 4            0.2          77.5      12.1  36.447      1 0.007769938 4.465482e-05

```

```
## 5      0.3      77.0     12.1 44.145      1 0.003769930 1.729326e-05
## 6      0.3      67.8     12.5 32.033      0 0.036333036 9.900010e-04
```

```
pairs(pct_cases~pct_black+pct_white+pct_hispanic+pct_asian, data = covid, cex = 0.6)
```



```
pairs(pct_cases~highschool+bachelors+income+pct_pop65+trump_pctvotes, data = covid, cex = 0.6)
```

