A View of Covid Between States

Herb

2024-03-04

In the following assignment for our DataAnalysis class – I was trying to provide two visuals and an analysis of the Covid data we were provided. I did a few more as some just did not give much information. I found setting up the data to be very time consuming, but I learned a lot. My first two graphs were of California and just showed how covid increased overtime by administration site acummulative reporting data–i.e. nothing really surprising – so I moved on to look at the data at the start of 2020 for each state — so I could do a state by state comparison. Rank ordered the states by death count at the end of the data reporting period by aggregating all the reports per state.

Relooking at the data, I realized I needed to adjust for the population per state to provide information on which state had more or less deaths per person so to speak (this is done in most accident reports or medical reposts for the same reason). So I tallied the states by admin reports and used one of the last cummulative date counts provided. I then compared population totals with those found online and everything look copacetic.

To show the dramatic change in which state has the highest death rate versus just high numbers of deaths per state, I provided the first graph of JUST total death counts–then keep the ordering for the death ratio plots. The values were all over the place, showing that the states had different death rates than population rates. I followed this with regression of both variables and deaths per population were not significant but death by state temperature was. Take a look below.

Finally out of curiosity I through in some comparisons with smokers, pets, and temperature per state–which was fun. I would love to have delved further into the climate, financial, population density, altitude average and other variables across the states. Too fun. When you check the death ratio per state to the population rate there is NO correlation. However, the second you look at the ratio to other predictors such as state temperature, the correlation climbs significantly. So there are variables or predictors out there that can model how the states did once the population effect is controlled for. Research will be needed to answer further questions.

In relation to biases–there are many. Just by my looking at smoking rates per state and finding a correlation is due somewhat to bias. There are many other variables that can be attributed to smoking that makes it correlated with the death rate–as you see in my quick correlation analyzed, pet ownwership per state was the highly correlated with reduce deaths while smoking was correlated with increased deaths (as you would expect with a lung desease).

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
library(readxl)
library(writexl)
library(devtools)
```

```
## Loading required package: usethis
```

```
## Warning: package 'usethis' was built under R version 4.3.2
```

```
library(ggplot2)
library(markdown)
library(lubridate)
library(RCurl)
```

```
##
## Attaching package: 'RCurl'
##
## The following object is masked from 'package:tidyr':
##
##     complete
```

```
library(knitr)
suppressWarnings({
 # Code that generates warning messages
})
```

```
## NULL
```

```
url_in<-"https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_
```

In the follwoing code chucks I combine the two US Death databases we downloaded. Removed all. the time data except the 2020 end point count per precinct and state. I then combined the precincts per each state to get the total reported deaths at the end of January 2020.

Note: I left my first graph of the points of deaths in the database which alert me to the fact that multiple administration points were reporting the tallys of deaths. this enabled me to add the data together to get a more meaningful chart and exminationi of death BY state.

Obviously, a more detailed analysis can be achieved by looking at admin counts and location and population per thos locationis but that was not my intent in this investigatioini – besides it took me long enough too learn how to do a smaller data base, clean its errors and graphs its visualiztions – maybe next time I will dig deeper.

```
file_names<-c("time_series_covid19_confirmed_US.csv",  "time_series_covid19_confirmed_global.csv", "time
urls<-str_c(url_in,file_names)
us_cases<-read_csv(urls[1])
```

```
## Rows: 3342 Columns: 1154
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (6): iso2, iso3, Admin2, Province_State, Country_Region, Combined_Key
## dbl (1148): UID, code3, FIPS, Lat, Long_, 1/22/20, 1/23/20, 1/24/20, 1/25/20...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
us_deaths<-read_csv(urls[3])
```

```
## Rows: 3342 Columns: 1155
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (6): iso2, iso3, Admin2, Province_State, Country_Region, Combined_Key
## dbl (1149): UID, code3, FIPS, Lat, Long_, Population, 1/22/20, 1/23/20, 1/24...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(us_cases)
```

```
## # A tibble: 6 x 1,154
##        UID iso2  iso3  code3  FIPS Admin2  Province_State Country_Region   Lat
##      <dbl> <chr> <chr> <dbl> <dbl> <chr>   <chr>          <chr>          <dbl>
## 1 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
## 2 84001003 US    USA     840  1003 Baldwin Alabama        US              30.7
## 3 84001005 US    USA     840  1005 Barbour Alabama        US              31.9
## 4 84001007 US    USA     840  1007 Bibb    Alabama        US              33.0
## 5 84001009 US    USA     840  1009 Blount  Alabama        US              34.0
## 6 84001011 US    USA     840  1011 Bullock Alabama        US              32.1
## # i 1,145 more variables: Long_ <dbl>, Combined_Key <chr>, `1/22/20` <dbl>,
## #   `1/23/20` <dbl>, `1/24/20` <dbl>, `1/25/20` <dbl>, `1/26/20` <dbl>,
## #   `1/27/20` <dbl>, `1/28/20` <dbl>, `1/29/20` <dbl>, `1/30/20` <dbl>,
## #   `1/31/20` <dbl>, `2/1/20` <dbl>, `2/2/20` <dbl>, `2/3/20` <dbl>,
## #   `2/4/20` <dbl>, `2/5/20` <dbl>, `2/6/20` <dbl>, `2/7/20` <dbl>,
## #   `2/8/20` <dbl>, `2/9/20` <dbl>, `2/10/20` <dbl>, `2/11/20` <dbl>,
## #   `2/12/20` <dbl>, `2/13/20` <dbl>, `2/14/20` <dbl>, `2/15/20` <dbl>, ...
```

```
us_deaths1 = us_deaths[,-c(1:5,8,11)]

deaths <- us_deaths1 %>%
  pivot_longer(
    cols = -c(Admin2, Province_State, Lat, Long_, Population),
    names_to = "date" ,
    values_to = "deaths") %>%
    mutate(date=mdy(date))
```

```
head(deaths)
```

```
## # A tibble: 6 x 7
```

```
##   Admin2  Province_State   Lat Long_ Population date       deaths
##   <chr>   <chr>          <dbl> <dbl>      <dbl> <date>      <dbl>
## 1 Autauga Alabama         32.5 -86.6      55869 2020-01-22      0
## 2 Autauga Alabama         32.5 -86.6      55869 2020-01-23      0
## 3 Autauga Alabama         32.5 -86.6      55869 2020-01-24      0
## 4 Autauga Alabama         32.5 -86.6      55869 2020-01-25      0
## 5 Autauga Alabama         32.5 -86.6      55869 2020-01-26      0
## 6 Autauga Alabama         32.5 -86.6      55869 2020-01-27      0
```

```r
head(us_cases)
```

```
## # A tibble: 6 x 1,154
##        UID iso2  iso3  code3  FIPS Admin2  Province_State Country_Region   Lat
##      <dbl> <chr> <chr> <dbl> <dbl> <chr>   <chr>          <chr>          <dbl>
## 1 84001001 US    USA     840  1001 Autauga Alabama        US              32.5
## 2 84001003 US    USA     840  1003 Baldwin Alabama        US              30.7
## 3 84001005 US    USA     840  1005 Barbour Alabama        US              31.9
## 4 84001007 US    USA     840  1007 Bibb    Alabama        US              33.0
## 5 84001009 US    USA     840  1009 Blount  Alabama        US              34.0
## 6 84001011 US    USA     840  1011 Bullock Alabama        US              32.1
## # i 1,145 more variables: Long_ <dbl>, Combined_Key <chr>, `1/22/20` <dbl>,
## #   `1/23/20` <dbl>, `1/24/20` <dbl>, `1/25/20` <dbl>, `1/26/20` <dbl>,
## #   `1/27/20` <dbl>, `1/28/20` <dbl>, `1/29/20` <dbl>, `1/30/20` <dbl>,
## #   `1/31/20` <dbl>, `2/1/20` <dbl>, `2/2/20` <dbl>, `2/3/20` <dbl>,
## #   `2/4/20` <dbl>, `2/5/20` <dbl>, `2/6/20` <dbl>, `2/7/20` <dbl>,
## #   `2/8/20` <dbl>, `2/9/20` <dbl>, `2/10/20` <dbl>, `2/11/20` <dbl>,
## #   `2/12/20` <dbl>, `2/13/20` <dbl>, `2/14/20` <dbl>, `2/15/20` <dbl>, ...
```

```r
us_cases1 = us_cases[,-c(1:5,8,11)]

cases <- us_cases1 %>%
  pivot_longer(cols = -c(Admin2, Province_State, Lat, Long_),
    names_to = "date" ,
    values_to = "cases") %>%
    mutate(date=mdy(date))
head(cases)
```

```
## # A tibble: 6 x 6
##   Admin2  Province_State   Lat Long_ date       cases
##   <chr>   <chr>          <dbl> <dbl> <date>     <dbl>
## 1 Autauga Alabama         32.5 -86.6 2020-01-22     0
## 2 Autauga Alabama         32.5 -86.6 2020-01-23     0
## 3 Autauga Alabama         32.5 -86.6 2020-01-24     0
## 4 Autauga Alabama         32.5 -86.6 2020-01-25     0
## 5 Autauga Alabama         32.5 -86.6 2020-01-26     0
## 6 Autauga Alabama         32.5 -86.6 2020-01-27     0
```

```r
cases[,6]
```

```
## # A tibble: 3,819,906 x 1
##    cases
##    <dbl>
```

```
## 1      0
## 2      0
## 3      0
## 4      0
## 5      0
## 6      0
## 7      0
## 8      0
## 9      0
## 10     0
## # i 3,819,896 more rows
```

```r
usdata = cbind(deaths,cases[,6])
head(usdata)
```

```
##   Admin2 Province_State      Lat     Long_ Population       date deaths cases
## 1 Autauga        Alabama 32.53953 -86.64408      55869 2020-01-22      0     0
## 2 Autauga        Alabama 32.53953 -86.64408      55869 2020-01-23      0     0
## 3 Autauga        Alabama 32.53953 -86.64408      55869 2020-01-24      0     0
## 4 Autauga        Alabama 32.53953 -86.64408      55869 2020-01-25      0     0
## 5 Autauga        Alabama 32.53953 -86.64408      55869 2020-01-26      0     0
## 6 Autauga        Alabama 32.53953 -86.64408      55869 2020-01-27      0     0
```

```r
tail(usdata)
```

```
##         Admin2 Province_State      Lat     Long_ Population       date deaths
## 3819901 Weston        Wyoming 43.83961 -104.5675       6927 2023-03-04     23
## 3819902 Weston        Wyoming 43.83961 -104.5675       6927 2023-03-05     23
## 3819903 Weston        Wyoming 43.83961 -104.5675       6927 2023-03-06     23
## 3819904 Weston        Wyoming 43.83961 -104.5675       6927 2023-03-07     23
## 3819905 Weston        Wyoming 43.83961 -104.5675       6927 2023-03-08     23
## 3819906 Weston        Wyoming 43.83961 -104.5675       6927 2023-03-09     23
##         cases
## 3819901  1905
## 3819902  1905
## 3819903  1905
## 3819904  1905
## 3819905  1905
## 3819906  1905
```

```r
sum(is.na(usdata$cases))
```

```
## [1] 0
```

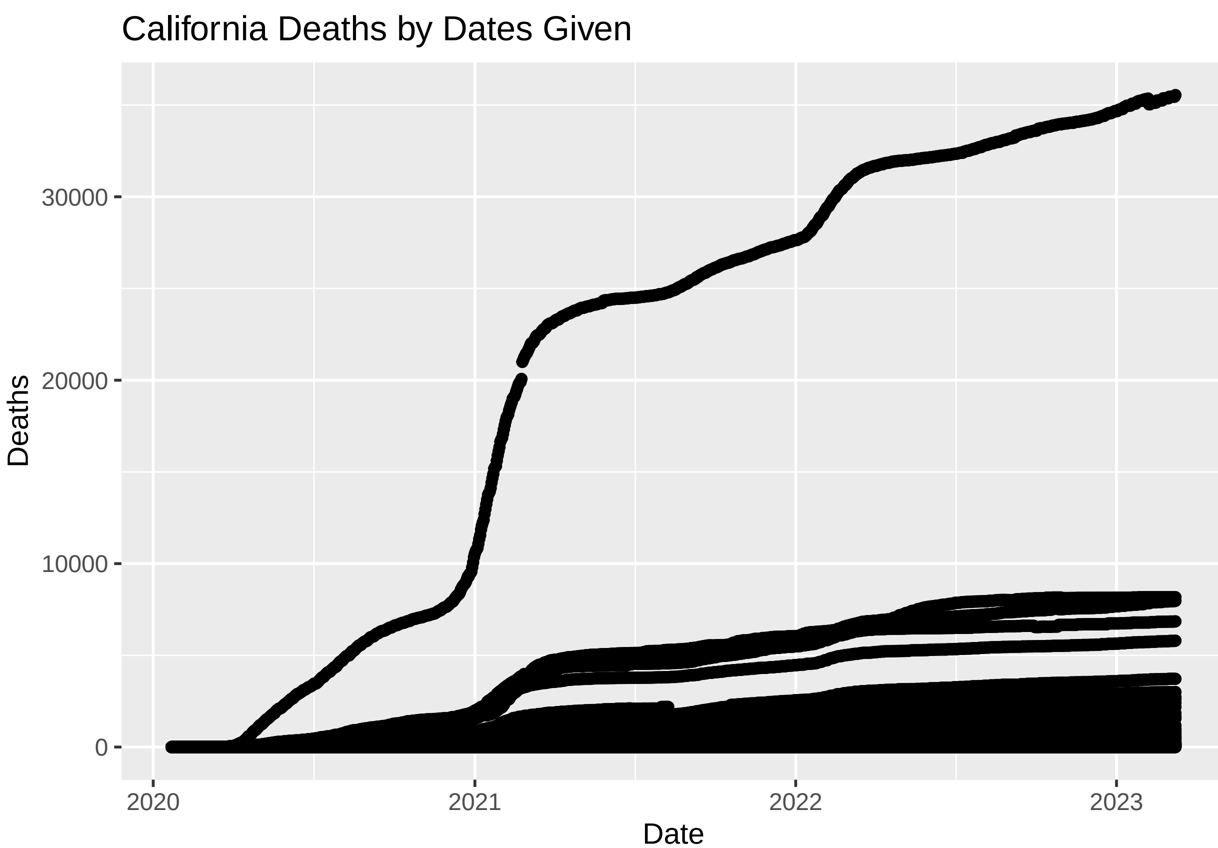
```r
sum(is.na(usdata$deaths))
```

```
## [1] 0
```

```r
cal = usdata[usdata$Province_State == "California",]
head(cal)
```

```
##          Admin2 Province_State      Lat     Long_ Population       date deaths
## 225172 Alameda     California 37.64629 -121.8929    1671329 2020-01-22      0
## 225173 Alameda     California 37.64629 -121.8929    1671329 2020-01-23      0
## 225174 Alameda     California 37.64629 -121.8929    1671329 2020-01-24      0
## 225175 Alameda     California 37.64629 -121.8929    1671329 2020-01-25      0
## 225176 Alameda     California 37.64629 -121.8929    1671329 2020-01-26      0
## 225177 Alameda     California 37.64629 -121.8929    1671329 2020-01-27      0
##        cases
## 225172     0
## 225173     0
## 225174     0
## 225175     0
## 225176     0
## 225177     0
```

```r
p <- ggplot(cal, aes(x=date, y=deaths)) +
  geom_point() +
  labs(x="Date",
  y = "Deaths",
  title = "California Deaths by Dates Given")
p
```



#This was my first plot which was done out of curiousity of what was in there. I did learn a lot. Data wwas reported by admin within the state which had to be added toegther to get an overal state average. The curves confirmed that the reports were cumulative overtime and I was not losing any end count data by looking at the last non NA data reported per admin area.

```r
yearlycases = us_cases[,c(10,1104)]
head(yearlycases)
```

```
## # A tibble: 6 x 2
##   Long_ `1/18/23`
##   <dbl>     <dbl>
## 1 -86.6     19389
## 2 -87.7     68764
## 3 -85.4      7258
## 4 -87.1      7889
## 5 -86.6     18130
## 6 -85.7      2956
```

```r
yearlydeaths = us_deaths[,c(6,7, 9,10,11,12,1105)]
head(yearlydeaths)
```

```
## # A tibble: 6 x 7
##   Admin2  Province_State   Lat Long_ Combined_Key         Population `1/18/23`
##   <chr>   <chr>          <dbl> <dbl> <chr>                     <dbl>     <dbl>
## 1 Autauga Alabama         32.5 -86.6 Autauga, Alabama, US      55869       230
## 2 Baldwin Alabama         30.7 -87.7 Baldwin, Alabama, US     223234       722
## 3 Barbour Alabama         31.9 -85.4 Barbour, Alabama, US      24686       103
## 4 Bibb    Alabama         33.0 -87.1 Bibb, Alabama, US         22394       109
## 5 Blount  Alabama         34.0 -86.6 Blount, Alabama, US       57826       261
## 6 Bullock Alabama         32.1 -85.7 Bullock, Alabama, US      10101        54
```

```r
usdata = cbind(yearlydeaths,yearlycases)
head(usdata)
```

```
##    Admin2 Province_State      Lat     Long_         Combined_Key Population
## 1 Autauga        Alabama 32.53953 -86.64408 Autauga, Alabama, US      55869
## 2 Baldwin        Alabama 30.72775 -87.72207 Baldwin, Alabama, US     223234
## 3 Barbour        Alabama 31.86826 -85.38713 Barbour, Alabama, US      24686
## 4    Bibb        Alabama 32.99642 -87.12511    Bibb, Alabama, US      22394
## 5  Blount        Alabama 33.98211 -86.56791  Blount, Alabama, US      57826
## 6 Bullock        Alabama 32.10031 -85.71266 Bullock, Alabama, US      10101
##   1/18/23     Long_ 1/18/23
## 1     230 -86.64408   19389
## 2     722 -87.72207   68764
## 3     103 -85.38713    7258
## 4     109 -87.12511    7889
## 5     261 -86.56791   18130
## 6      54 -85.71266    2956
```

```r
tail(usdata)
```

```
##          Admin2 Province_State      Lat     Long_            Combined_Key
## 3337 Sweetwater        Wyoming 41.65944 -108.8828 Sweetwater, Wyoming, US
## 3338      Teton        Wyoming 43.93522 -110.5891      Teton, Wyoming, US
## 3339      Uinta        Wyoming 41.28782 -110.5476      Uinta, Wyoming, US
## 3340 Unassigned        Wyoming  0.00000    0.0000 Unassigned, Wyoming, US
```

```
## 3341   Washakie        Wyoming 43.90452 -107.6802   Washakie, Wyoming, US
## 3342     Weston        Wyoming 43.83961 -104.5675     Weston, Wyoming, US
##      Population 1/18/23     Long_ 1/18/23
## 3337      42343     137 -108.8828   12442
## 3338      23464      16 -110.5891   12065
## 3339      20226      43 -110.5476    6346
## 3340          0       0    0.0000       0
## 3341       7805      47 -107.6802    2733
## 3342       6927      22 -104.5675    1884
```

```r
sum(is.na(usdata$cases))
```

```
## [1] 0
```

```r
sum(is.na(usdata$deaths))
```

```
## [1] 0
```

```r
#checking which values are not NA
summary(usdata)
```

```
##     Admin2          Province_State          Lat              Long_
##  Length:3342        Length:3342        Min.   :-14.27   Min.   :-174.16
##  Class :character   Class :character   1st Qu.: 33.90   1st Qu.: -97.80
##  Mode  :character   Mode  :character   Median : 38.01   Median : -89.49
##                                        Mean   : 36.72   Mean   : -88.64
##                                        3rd Qu.: 41.58   3rd Qu.: -82.31
##                                        Max.   : 69.31   Max.   : 145.67
##  Combined_Key         Population          1/18/23            Long_
##  Length:3342        Min.   :       0   Min.   :    0.0   Min.   :-174.16
##  Class :character   1st Qu.:    9917   1st Qu.:   38.0   1st Qu.: -97.80
##  Mode  :character   Median :   24892   Median :  100.0   Median : -89.49
##                     Mean   :   99604   Mean   :  329.9   Mean   : -88.64
##                     3rd Qu.:   64975   3rd Qu.:  243.0   3rd Qu.: -82.31
##                     Max.   :10039107   Max.   :35052.0   Max.   : 145.67
##     1/18/23
##  Min.   :      0
##  1st Qu.:   2852
##  Median :   7602
##  Mean   :  30480
##  3rd Qu.:  19840
##  Max.   :3663899
```

```
The following chunk is the key to simplifying the data so I could compare state to state efficiency so t
```

```
‘‘‘r
sum(usdata$deaths, na.rm = TRUE)

## [1] 0
```

```r
usdata[usdata$"Population">1000000,]
```

```
##              Admin2 Province_State      Lat      Long_
## 111        Maricopa        Arizona 33.34836 -112.49182
## 115            Pima        Arizona 32.09713 -111.78900
## 198         Alameda     California 37.64629 -121.89293
## 204    Contra Costa     California 37.91923 -121.92895
## 216     Los Angeles     California 34.30828 -118.22824
## 227          Orange     California 33.70148 -117.76460
## 231       Riverside     California 33.74315 -115.99336
## 232      Sacramento     California 38.45107 -121.34254
## 234  San Bernardino     California 34.84060 -116.17747
## 235       San Diego     California 33.03485 -116.73653
## 241     Santa Clara     California 37.23105 -121.69705
## 348         Broward        Florida 26.15185  -80.48726
## 370    Hillsborough        Florida 27.92766  -82.32013
## 385      Miami-Dade        Florida 25.61124  -80.55171
## 390          Orange        Florida 28.51368  -81.31799
## 393      Palm Beach        Florida 26.64676  -80.46536
## 471          Fulton        Georgia 33.79217  -84.46319
## 643            Cook       Illinois 41.84145  -87.81659
## 1255     Montgomery       Maryland 39.13676  -77.20358
## 1275      Middlesex  Massachusetts 42.48608  -71.39049
## 1347        Oakland       Michigan 42.66090  -83.38595
## 1368          Wayne       Michigan 42.28098  -83.28126
## 1396       Hennepin      Minnesota 45.00762  -93.47695
## 1816          Clark         Nevada 36.21459 -115.01302
## 1905          Bronx       New York 40.85209  -73.86283
## 1926          Kings       New York 40.63618  -73.94936
## 1932         Nassau       New York 40.74067  -73.58942
## 1933       New York       New York 40.76727  -73.97153
## 1944         Queens       New York 40.71088  -73.81685
## 1955        Suffolk       New York 40.88320  -72.80122
## 2026    Mecklenburg North Carolina 35.24469  -80.83177
## 2060           Wake North Carolina 35.78879  -78.65249
## 2142       Cuyahoga           Ohio 41.42412  -81.65918
## 2149       Franklin           Ohio 39.96996  -83.01116
## 2333      Allegheny   Pennsylvania 40.46810  -79.98168
## 2383   Philadelphia   Pennsylvania 40.00339  -75.13793
## 2715          Bexar          Texas 29.44929  -98.52020
## 2743         Collin          Texas 33.18820  -96.57264
## 2757         Dallas          Texas 32.76671  -96.77796
## 2801         Harris          Texas 29.85865  -95.39340
## 2921        Tarrant          Texas 32.77144  -97.29102
## 2928         Travis          Texas 30.33432  -97.78536
## 2977      Salt Lake           Utah 40.66617 -111.92160
## 3048        Fairfax       Virginia 38.83678  -77.27566
## 3162           King     Washington 47.49138 -121.83461
##                         Combined_Key Population 1/18/23      Long_ 1/18/23
## 111            Maricopa, Arizona, US    4485414   18591 -112.49182 1493595
## 115                Pima, Arizona, US    1047279    4216 -111.78900  312126
## 198          Alameda, California, US    1671329    2112 -121.89293  394694
## 204     Contra Costa, California, US    1153526    1505 -121.92895  290023
```

```
## 216      Los Angeles, California, US   10039107   35052 -118.22824 3663899
## 227           Orange, California, US    3175692    7742 -117.76460  773519
## 231        Riverside, California, US    2470546    6761 -115.99336  768374
## 232       Sacramento, California, US    1552058    3635 -121.34254  403144
## 234   San Bernardino, California, US    2180085    8146 -116.17747  737401
## 235        San Diego, California, US    3338330    5681 -116.73653 1050110
## 241      Santa Clara, California, US    1927852    2601 -121.69705  488518
## 348             Broward, Florida, US    1952778    6577  -80.48726  758025
## 370        Hillsborough, Florida, US    1471968    4302  -82.32013  469096
## 385          Miami-Dade, Florida, US    2716940   12049  -80.55171 1514363
## 390              Orange, Florida, US    1393452    3205  -81.31799  466897
## 393          Palm Beach, Florida, US    1496770    5842  -80.46536  469048
## 471              Fulton, Georgia, US    1063937    2614  -84.46319  271886
## 643               Cook, Illinois, US    5150233   15127  -87.81659 1502422
## 1255       Montgomery, Maryland, US    1050688    2312  -77.20358  240468
## 1275    Middlesex, Massachusetts, US    1611699    4590  -71.39049  429459
## 1347           Oakland, Michigan, US    1257584    4442  -83.38595  378986
## 1368             Wayne, Michigan, US    1749343    8940  -83.28126  526333
## 1396         Hennepin, Minnesota, US    1265843    2871  -93.47695  378324
## 1816               Clark, Nevada, US    2266715    9248 -115.01302  665139
## 1905             Bronx, New York, US    1418207    8431  -73.86283  541439
## 1926             Kings, New York, US    2559903   14010  -73.94936  944310
## 1932            Nassau, New York, US    1356924    4279  -73.58942  542937
## 1933          New York, New York, US    1628706    6075  -73.97153  584496
## 1944            Queens, New York, US    2253858   13204  -73.81685  887614
## 1955           Suffolk, New York, US    1476601    4888  -72.80122  561921
## 2026 Mecklenburg, North Carolina, US    1110356    1863  -80.83177  360949
## 2060        Wake, North Carolina, US    1111761    1300  -78.65249  385179
## 2142              Cuyahoga, Ohio, US    1235072    4107  -81.65918  342327
## 2149              Franklin, Ohio, US    1316756    2816  -83.01116  360018
## 2333    Allegheny, Pennsylvania, US    1216045    3746  -79.98168  334208
## 2383  Philadelphia, Pennsylvania, US    1584064    5456  -75.13793  385412
## 2715                Bexar, Texas, US    2003554    6441  -98.52020  689205
## 2743               Collin, Texas, US    1034730    1592  -96.57264  271855
## 2757               Dallas, Texas, US    2635516    7062  -96.77796  686164
## 2801               Harris, Texas, US    4713325   11495  -95.39340 1255228
## 2921              Tarrant, Texas, US    2102515    6264  -97.29102  676208
## 2928               Travis, Texas, US    1273954    1826  -97.78536  327377
## 2977             Salt Lake, Utah, US    1160437    1802 -111.92160  405923
## 3048           Fairfax, Virginia, US    1147532    1666  -77.27566  256096
## 3162             King, Washington, US    2252782    3424 -121.83461  541429
```

```r
colnames(usdata) = c("city", "state", "lat", "long", "city/state","population", "deaths", "longtocheck"
head(usdata)
```

```
##      city   state      lat      long           city/state population deaths
## 1 Autauga Alabama 32.53953 -86.64408 Autauga, Alabama, US      55869    230
## 2 Baldwin Alabama 30.72775 -87.72207 Baldwin, Alabama, US     223234    722
## 3 Barbour Alabama 31.86826 -85.38713 Barbour, Alabama, US      24686    103
## 4    Bibb Alabama 32.99642 -87.12511    Bibb, Alabama, US      22394    109
## 5  Blount Alabama 33.98211 -86.56791  Blount, Alabama, US      57826    261
## 6 Bullock Alabama 32.10031 -85.71266 Bullock, Alabama, US      10101     54
##   longtocheck cases
## 1   -86.64408 19389
```

```
## 2   -87.72207 68764
## 3   -85.38713  7258
## 4   -87.12511  7889
## 5   -86.56791 18130
## 6   -85.71266  2956
```

```
#. aggregate the data bystate summing deaths and cases and taking mean of population.
statecases=aggregate(usdata$cases, list(usdata$state), FUN=sum)
statedeaths=aggregate(usdata$deaths, list(usdata$state), FUN=sum)
statepop=aggregate(usdata$population, list(usdata$state), FUN=sum)
head(statecases)
```

```
##          Group.1        x
## 1        Alabama  1602891
## 2         Alaska   302921
## 3 American Samoa     8309
## 4        Arizona  2394646
## 5       Arkansas   992745
## 6     California 11951728
```

```
head(statedeaths)
```

```
##          Group.1     x
## 1        Alabama 20846
## 2         Alaska  1455
## 3 American Samoa    34
## 4        Arizona 32631
## 5       Arkansas 12766
## 6     California 99331
```

```
head(statepop)
```

```
##          Group.1        x
## 1        Alabama  4903185
## 2         Alaska   740995
## 3 American Samoa    55641
## 4        Arizona  7278717
## 5       Arkansas  3017804
## 6     California 39512223
```

```
bystate =data.frame(statepop,statecases$x, statedeaths$x)
names(bystate)[1] = "state"
names(bystate)[2] = "population"
names(bystate)[3] = "cases"
names(bystate)[4] = "deaths"
bystate[1:10,]
```

```
##                state population    cases deaths
## 1            Alabama    4903185  1602891  20846
## 2             Alaska     740995   302921   1455
## 3     American Samoa      55641     8309     34
```

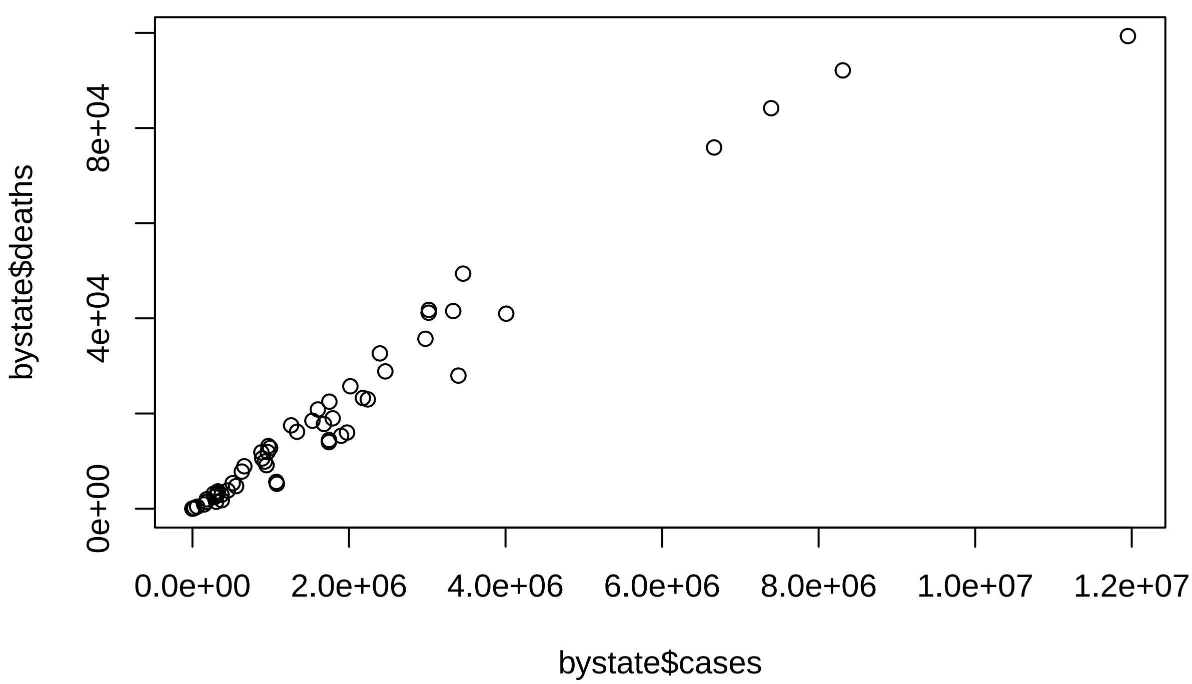
```
## 4           Arizona    7278717  2394646  32631
## 5          Arkansas    3017804   992745  12766
## 6        California   39512223 11951728  99331
## 7          Colorado    5758736  1743671  13985
## 8       Connecticut    3565287   960940  11895
## 9          Delaware     973764   324137   3220
## 10 Diamond Princess          0       49      0
```

```r
bystate = na.omit(bystate)
bystate
```

```
##                       state population    cases deaths
## 1                   Alabama    4903185  1602891  20846
## 2                    Alaska     740995   302921   1455
## 3            American Samoa      55641     8309     34
## 4                   Arizona    7278717  2394646  32631
## 5                  Arkansas    3017804   992745  12766
## 6                California   39512223 11951728  99331
## 7                  Colorado    5758736  1743671  13985
## 8               Connecticut    3565287   960940  11895
## 9                  Delaware     973764   324137   3220
## 10         Diamond Princess          0       49      0
## 11     District of Columbia     705749   175014   1415
## 12                  Florida   21477737  7393712  84176
## 13                  Georgia   10617423  3020166  41772
## 14           Grand Princess          0      103      3
## 15                     Guam     164229    60526    415
## 16                   Hawaii    1415872   375925   1775
## 17                    Idaho    1787065   514326   5344
## 18                 Illinois   12671821  4008843  40980
## 19                  Indiana    6732219  2017978  25722
## 20                     Iowa    3155070   892558  10538
## 21                   Kansas    2913314   924193   9903
## 22                 Kentucky    4467673  1680601  17793
## 23                Louisiana    4648794  1533257  18479
## 24                    Maine    1344212   309680   2853
## 25                 Maryland    6045680  1336429  16156
## 26            Massachusetts    6892503  2178027  23259
## 27                 Michigan    9986857  3017948  41185
## 28                Minnesota    5639632  1745105  14421
## 29              Mississippi    2976149   970585  13151
## 30                 Missouri    6626371  1749656  22490
## 31                  Montana    1068778   324726   3630
## 32                 Nebraska    1934408   558003   4730
## 33                   Nevada    3080156   881498  11834
## 34            New Hampshire    1359711   371710   2908
## 35               New Jersey    8882190  2976788  35699
## 36               New Mexico    2096829   662967   8902
## 37                 New York   19453561  6664854  75913
## 38           North Carolina   10488084  3398161  27967
## 39             North Dakota     762062   282222   2428
## 40 Northern Mariana Islands      55144    13430     41
## 41                     Ohio   11689100  3331651  41530
## 42                 Oklahoma    3956971  1261310  17502
```

```
## 43                       Oregon    4217737    946727    9141
## 44                 Pennsylvania   12801989   3458136   49397
## 45                  Puerto Rico    3754939   1071990    5623
## 46                 Rhode Island    1059361    450559    3798
## 47               South Carolina    5148714   1791933   18983
## 48                 South Dakota     884659    273354    3145
## 49                    Tennessee    6829174   2464488   28853
## 50                        Texas   28995881   8308895   92118
## 51                         Utah    3205958   1079001    5222
## 52                      Vermont     623989    149687     884
## 53               Virgin Islands     107268     24176     129
## 54                     Virginia    8535519   2240431   22962
## 55                   Washington    7614893   1899401   15312
## 56                West Virginia    1792147    631197    7790
## 57                    Wisconsin    5822434   1975535   15989
## 58                      Wyoming     578759    183586    1970
```

```
plot(bystate$cases,bystate$deaths)
```




```
summary(bystate)
```

```
##      state             population           cases              deaths
##  Length:58          Min.   :       0   Min.   :      49   Min.   :    0
##  Class :character   1st Qu.: 1137636   1st Qu.:  336472   1st Qu.: 3164
##  Mode  :character   Median : 3660113   Median : 1032368   Median :12330
##                     Mean   : 5739226   Mean   : 1756260   Mean   :19007
##                     3rd Qu.: 6876671   3rd Qu.: 2138015   3rd Qu.:23185
##                     Max.   :39512223   Max.   :11951728   Max.   :99331
```

```
bystate
```

```
##                        state population    cases deaths
## 1                    Alabama    4903185  1602891  20846
## 2                     Alaska     740995   302921   1455
## 3             American Samoa      55641     8309     34
## 4                    Arizona    7278717  2394646  32631
## 5                   Arkansas    3017804   992745  12766
## 6                 California   39512223 11951728  99331
## 7                   Colorado    5758736  1743671  13985
## 8                Connecticut    3565287   960940  11895
## 9                   Delaware     973764   324137   3220
## 10          Diamond Princess          0       49      0
## 11      District of Columbia     705749   175014   1415
## 12                   Florida   21477737  7393712  84176
## 13                   Georgia   10617423  3020166  41772
## 14            Grand Princess          0      103      3
## 15                      Guam     164229    60526    415
## 16                    Hawaii    1415872   375925   1775
## 17                     Idaho    1787065   514326   5344
## 18                  Illinois   12671821  4008843  40980
## 19                   Indiana    6732219  2017978  25722
## 20                      Iowa    3155070   892558  10538
## 21                    Kansas    2913314   924193   9903
## 22                  Kentucky    4467673  1680601  17793
## 23                 Louisiana    4648794  1533257  18479
## 24                     Maine    1344212   309680   2853
## 25                  Maryland    6045680  1336429  16156
## 26             Massachusetts    6892503  2178027  23259
## 27                  Michigan    9986857  3017948  41185
## 28                 Minnesota    5639632  1745105  14421
## 29               Mississippi    2976149   970585  13151
## 30                  Missouri    6626371  1749656  22490
## 31                   Montana    1068778   324726   3630
## 32                  Nebraska    1934408   558003   4730
## 33                    Nevada    3080156   881498  11834
## 34             New Hampshire    1359711   371710   2908
## 35                New Jersey    8882190  2976788  35699
## 36                New Mexico    2096829   662967   8902
## 37                  New York   19453561  6664854  75913
## 38            North Carolina   10488084  3398161  27967
## 39              North Dakota     762062   282222   2428
## 40  Northern Mariana Islands      55144    13430     41
## 41                      Ohio   11689100  3331651  41530
## 42                  Oklahoma    3956971  1261310  17502
## 43                    Oregon    4217737   946727   9141
## 44              Pennsylvania   12801989  3458136  49397
## 45               Puerto Rico    3754939  1071990   5623
## 46              Rhode Island    1059361   450559   3798
## 47            South Carolina    5148714  1791933  18983
## 48              South Dakota     884659   273354   3145
## 49                 Tennessee    6829174  2464488  28853
## 50                     Texas   28995881  8308895  92118
```

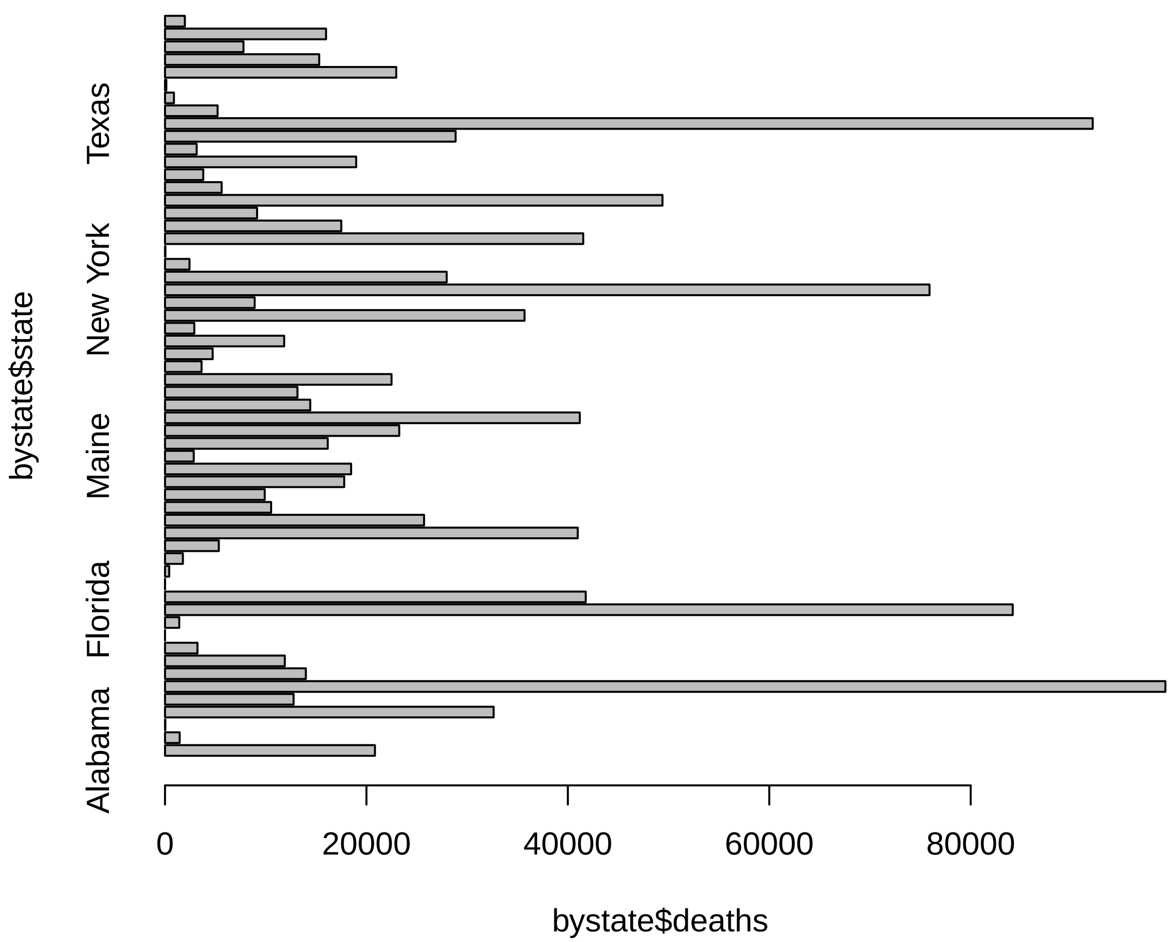
```
## 51                   Utah    3205958  1079001   5222
## 52                Vermont     623989   149687    884
## 53         Virgin Islands     107268    24176    129
## 54               Virginia    8535519  2240431  22962
## 55             Washington    7614893  1899401  15312
## 56          West Virginia    1792147   631197   7790
## 57              Wisconsin    5822434  1975535  15989
## 58                Wyoming     578759   183586   1970
```

```r
bystatesort = bystate[order(bystate$death, decreasing = TRUE),]
print(bystatesort)
```

```
##                     state population    cases deaths
## 6              California   39512223 11951728  99331
## 50                  Texas   28995881  8308895  92118
## 12                Florida   21477737  7393712  84176
## 37               New York   19453561  6664854  75913
## 44           Pennsylvania   12801989  3458136  49397
## 13                Georgia   10617423  3020166  41772
## 41                   Ohio   11689100  3331651  41530
## 27               Michigan    9986857  3017948  41185
## 18               Illinois   12671821  4008843  40980
## 35             New Jersey    8882190  2976788  35699
## 4                 Arizona    7278717  2394646  32631
## 49              Tennessee    6829174  2464488  28853
## 38         North Carolina   10488084  3398161  27967
## 19                Indiana    6732219  2017978  25722
## 26          Massachusetts    6892503  2178027  23259
## 54               Virginia    8535519  2240431  22962
## 30               Missouri    6626371  1749656  22490
## 1                 Alabama    4903185  1602891  20846
## 47         South Carolina    5148714  1791933  18983
## 23              Louisiana    4648794  1533257  18479
## 22               Kentucky    4467673  1680601  17793
## 42               Oklahoma    3956971  1261310  17502
## 25               Maryland    6045680  1336429  16156
## 57              Wisconsin    5822434  1975535  15989
## 55             Washington    7614893  1899401  15312
## 28              Minnesota    5639632  1745105  14421
## 7                Colorado    5758736  1743671  13985
## 29            Mississippi    2976149   970585  13151
## 5                Arkansas    3017804   992745  12766
## 8             Connecticut    3565287   960940  11895
## 33                 Nevada    3080156   881498  11834
## 20                   Iowa    3155070   892558  10538
## 21                 Kansas    2913314   924193   9903
## 43                 Oregon    4217737   946727   9141
## 36             New Mexico    2096829   662967   8902
## 56          West Virginia    1792147   631197   7790
## 45            Puerto Rico    3754939  1071990   5623
## 17                  Idaho    1787065   514326   5344
## 51                   Utah    3205958  1079001   5222
## 32               Nebraska    1934408   558003   4730
## 46           Rhode Island    1059361   450559   3798
```

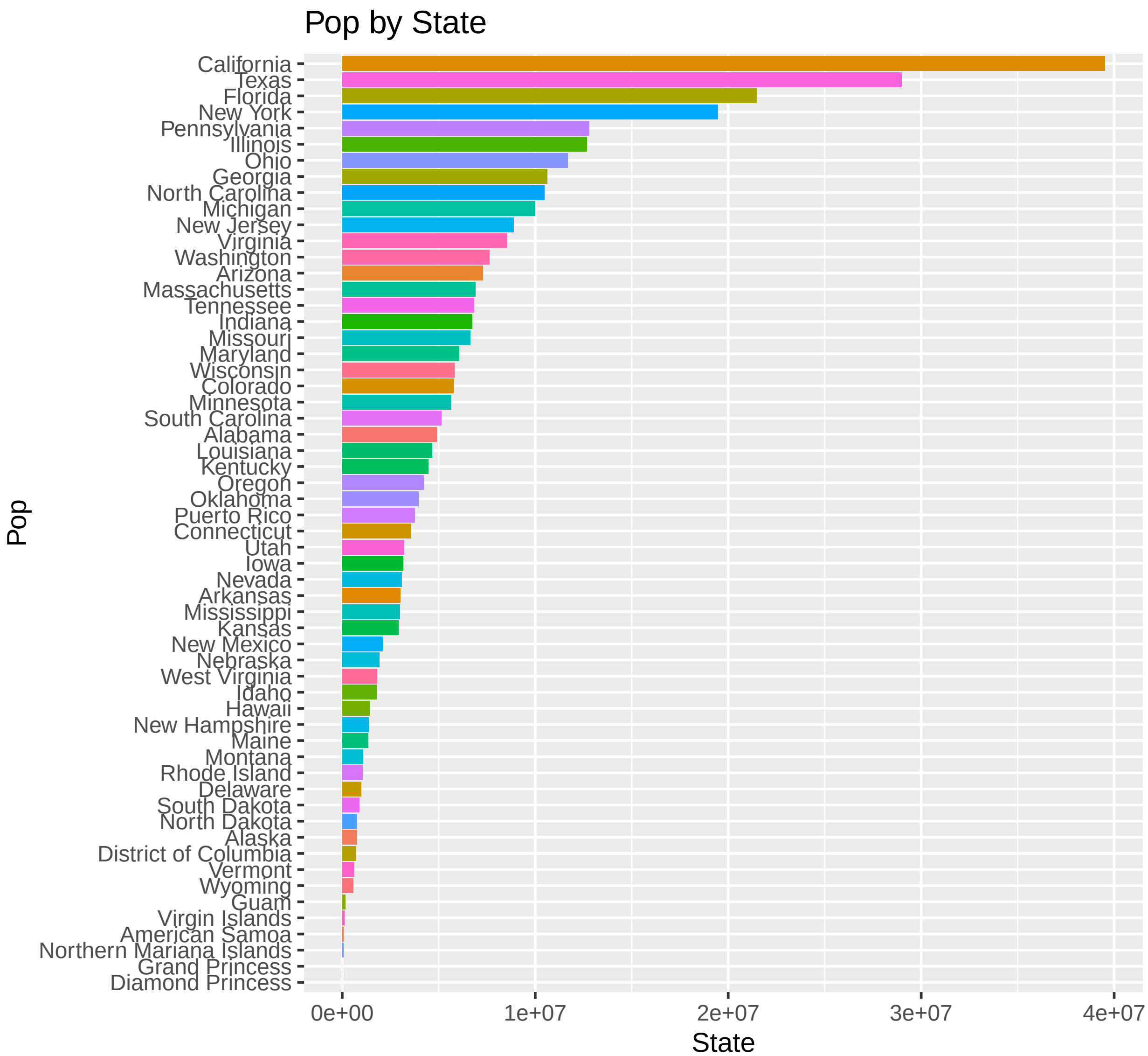
```
## 31                    Montana    1068778    324726    3630
## 9                    Delaware     973764    324137    3220
## 48               South Dakota     884659    273354    3145
## 34              New Hampshire    1359711    371710    2908
## 24                      Maine    1344212    309680    2853
## 39               North Dakota     762062    282222    2428
## 58                    Wyoming     578759    183586    1970
## 16                     Hawaii    1415872    375925    1775
## 2                      Alaska     740995    302921    1455
## 11       District of Columbia     705749    175014    1415
## 52                    Vermont     623989    149687     884
## 15                       Guam     164229     60526     415
## 53             Virgin Islands     107268     24176     129
## 40 Northern Mariana Islands      55144     13430      41
## 3              American Samoa      55641      8309      34
## 14             Grand Princess          0       103       3
## 10           Diamond Princess          0        49       0
```

```
barplot(bystate$deaths ~ bystate$state, horiz = TRUE)
```

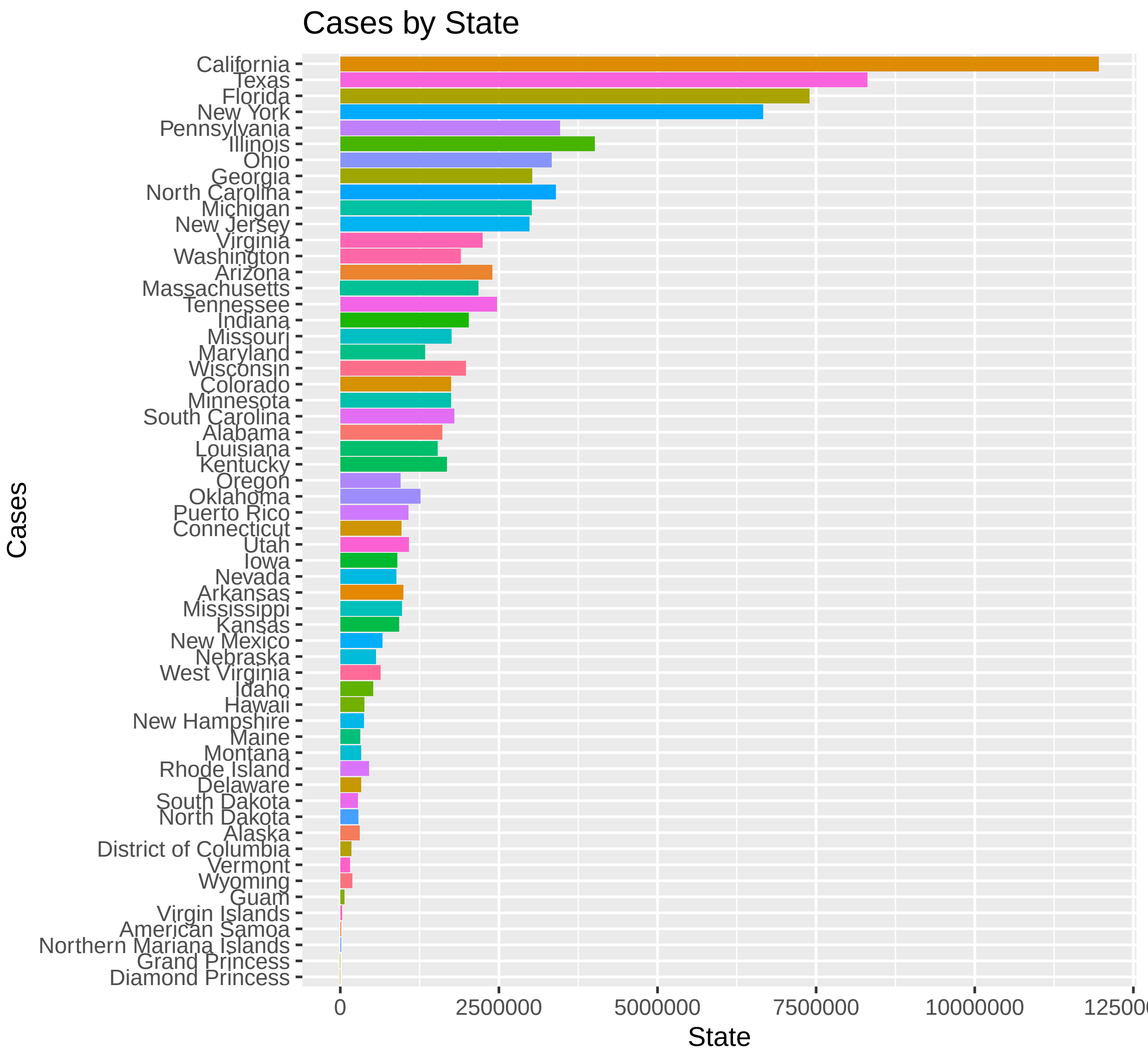



```
plotpop = bystatesort %>%
  ggplot() +
labs(title = "Pop by State",
```

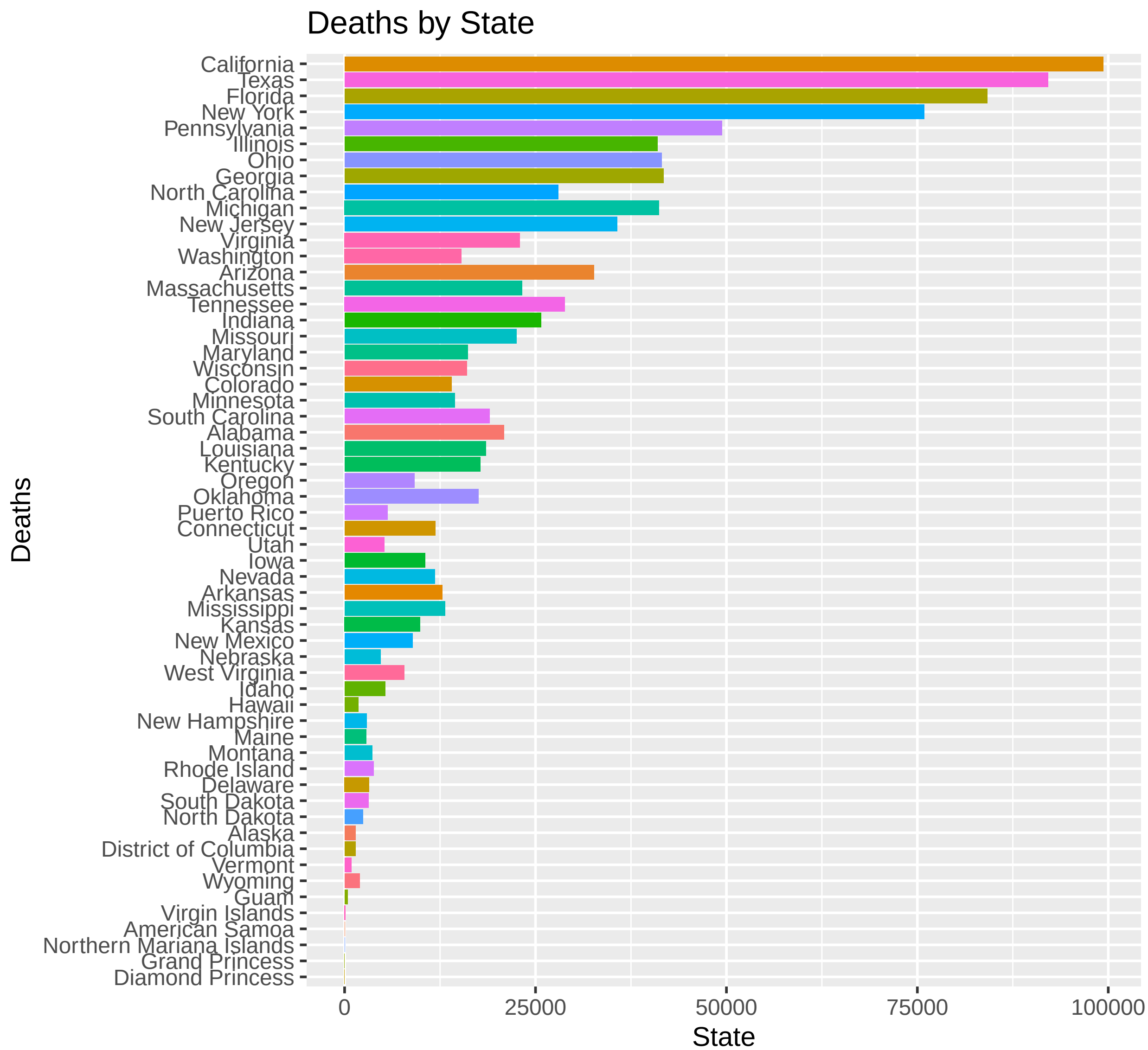
```r
       x = "Pop",
       y = "State") +
  geom_bar(aes(x = reorder(state, population), y = population,
               fill = state), stat = "identity", show.legend = FALSE) +

  coord_flip()
plotpop
```




```r
plotcases = bystatesort %>%
  ggplot() +
  labs(title = "Cases by State",
       x = "Cases",
       y = "State") +
  geom_bar(aes(x = reorder(state, population), y = cases,
               fill = state), stat = "identity", show.legend = FALSE) +
  coord_flip()
plotcases
```

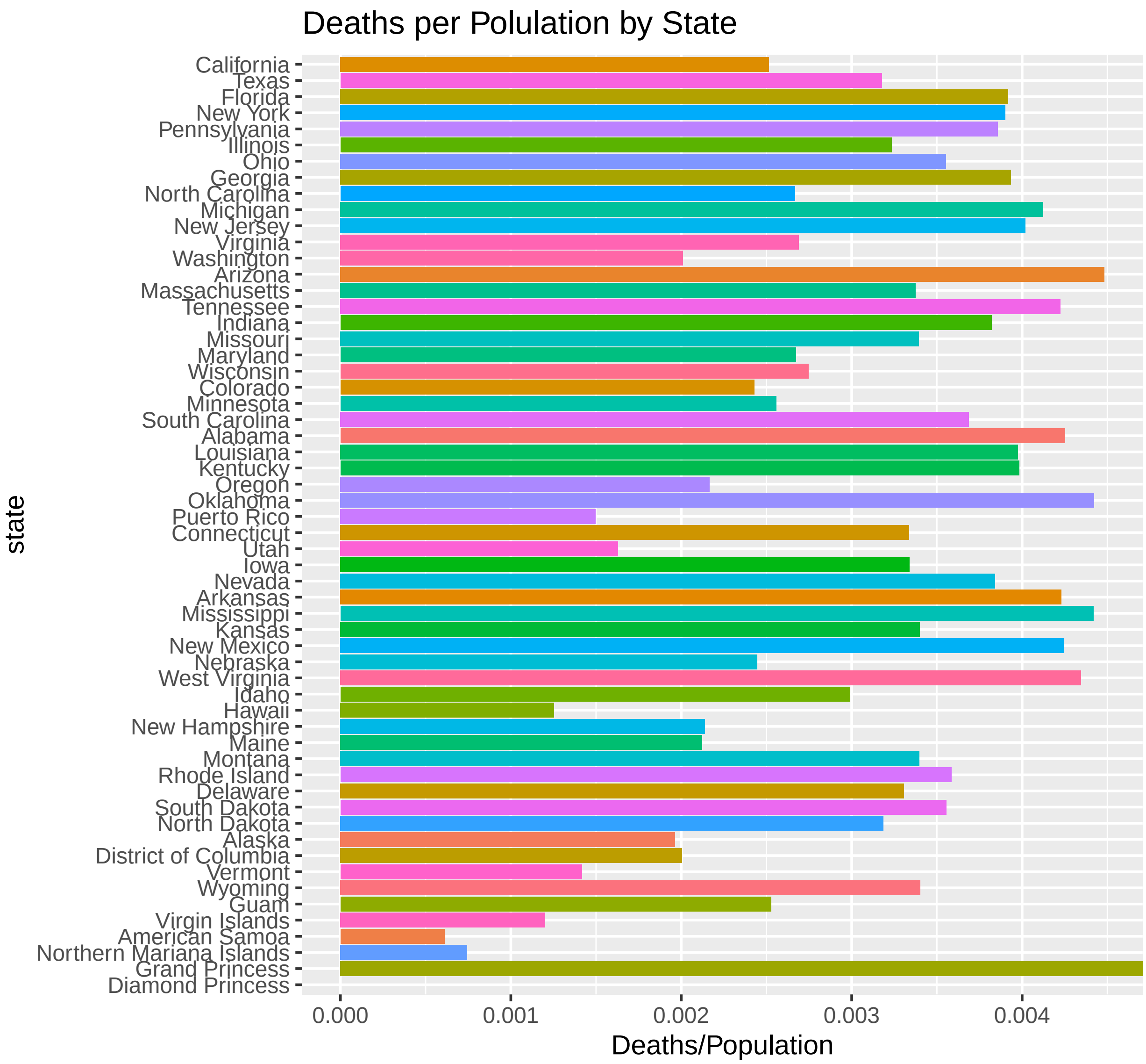



```
plotdeaths = bystatesort %>%
  ggplot() +
  labs(title = "Deaths by State",
       x = "Deaths",
       y = "State") +
  geom_bar(aes(x = reorder(state, population), y = deaths,
                fill = state), stat = "identity", show.legend = FALSE) +
  coord_flip()
plotdeaths
```




```
plotdeaths = bystatesort %>%
  ggplot() +
  labs(title = "Deaths per Polulation by State",
       x = "state",
       y = "Deaths/Population") +
  geom_bar(aes(x = reorder(state, population), y = deaths/population,
                  fill = state), stat = "identity", show.legend = FALSE) +
  coord_flip()
plotdeaths
```

```
## Warning: Removed 1 rows containing missing values (‘position_stack()‘).
```



#I had to convert my collected data because I had trouble linking it in git..I will learn to do it later. # temp data from: https://wisevoter.com/state-rankings/average-temperature-by-state/
hospital bed data https://ceoworld.biz/2020/03/16/these-are-the-u-s-states-with-the-most-and-least-hospital-beds/
Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System 2017, analysed by the American Lung Association Epidemiology and Statistics Unit # SMOKING DATA FROM https://www.statista.com/statistics/261595/us-states-with-highest-smoking-rates-among-adults/

```
state = c("AL","AK","AZ","AR","AZ","CA","CP","DE","FL","GA","GU","ID","IL","IN","IS","KS","KY","LA","ME

pop_in_thousands = c(4903.185,  740.995,    7278.717,   3017.804,   39512.223,  5758.736,   3565.287,

deaths_per_pop = c(4.25152222483957,    1.96357600253713,   4.48307029933984,   4.2302283382221,    2.5

temp_rank = c(6,    49, 8,  9,  11, 37, 28, 15, 1,  4,  42, 22, 23, 33, 18, 14, 2,  44, 17, 30, 40, 47,

medbed_per_thou = c(3.1,    2.2,    1.9,    3.2,    1.8,    1.9,    2,  2.2,    2.6,    2.4,    1.9,
```

```r
smoker_rate = c(14, 14, 11, 17, 9,  10, 9,  12, 10, 12, 11, 12, 15, 4,  13, 17, 15, 12, 9,  9,  14, 11,

pet_own_rate = c(59.8,  59.3,   58, 69, 57.2,   64.7,   49.9,   57.9,   56, 51.1,   69.9,   48.6,   69.


my_data <- data.frame(pop_in_thousands,deaths_per_pop, temp_rank, medbed_per_thou, smoker_rate, pet_own

summary(my_data)
```

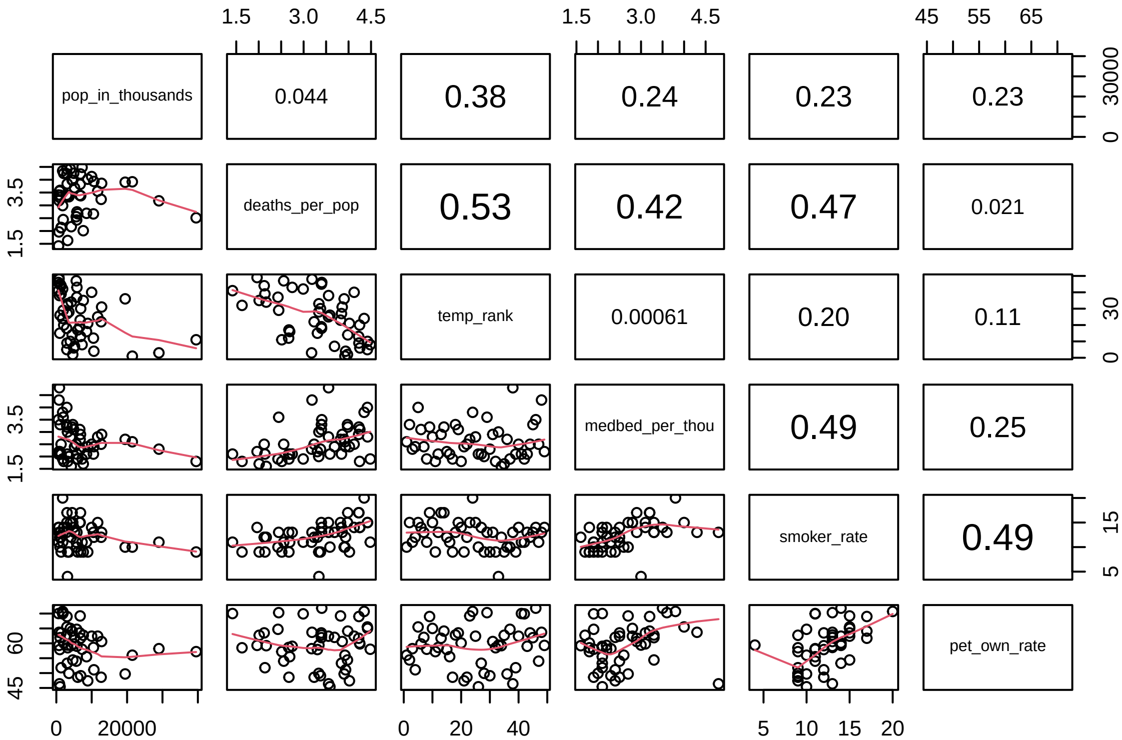
```
##  pop_in_thousands  deaths_per_pop    temp_rank  medbed_per_thou  smoker_rate
##  Min.   :  578.8   Min.   :1.417   Min.   : 1   Min.   :1.600   Min.   : 4.00
##  1st Qu.: 1934.4   1st Qu.:2.672   1st Qu.:13   1st Qu.:2.100   1st Qu.:10.00
##  Median : 4648.8   Median :3.396   Median :25   Median :2.500   Median :13.00
##  Mean   : 6665.6   Mean   :3.308   Mean   :25   Mean   :2.614   Mean   :12.33
##  3rd Qu.: 7614.9   3rd Qu.:3.934   3rd Qu.:37   3rd Qu.:3.100   3rd Qu.:14.00
##  Max.   :39512.2   Max.   :4.483   Max.   :49   Max.   :4.800   Max.   :20.00
##   pet_own_rate
##  Min.   :45.40
##  1st Qu.:54.40
##  Median :59.40
##  Mean   :59.26
##  3rd Qu.:63.50
##  Max.   :71.80
```

```r
# Function to add correlation coefficients
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...) {
    usr <- par("usr")
    on.exit(par(usr))
    par(usr = c(0, 1, 0, 1))
    Cor <- abs(cor(x, y)) # Remove abs function if desired
    txt <- paste0(prefix, format(c(Cor, 0.123456789), digits = digits)[1])
    if(missing(cex.cor)) {
        cex.cor <- 0.4 / strwidth(txt)
    }
    text(0.5, 0.5, txt,
         cex = 1 + cex.cor * Cor) # Resize the text by level of correlation
}
```

```r
# Plotting the correlation matrix

pairs(my_data,
      upper.panel = panel.cor,    # Correlation panel
      lower.panel = panel.smooth) # Smoothed regression lines
```

```
## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter
```

```
## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter

## Warning in par(usr): argument 1 does not name a graphical parameter
```



I found this one interesting.

```
model = lm(deaths_per_pop ~ temp_rank + medbed_per_thou + smoker_rate + pet_own_rate)
summary(model)
```

```
##
## Call:
## lm(formula = deaths_per_pop ~ temp_rank + medbed_per_thou + smoker_rate +
```

```
##     pet_own_rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92302 -0.51131  0.01131  0.40241  1.14655
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.304408   0.735862   4.491 5.08e-05 ***
## temp_rank       -0.024741   0.006228  -3.973  0.00026 ***
## medbed_per_thou  0.352183   0.135556   2.598  0.01271 *
## smoker_rate      0.094688   0.040129   2.360  0.02280 *
## pet_own_rate    -0.024735   0.014266  -1.734  0.08994 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5821 on 44 degrees of freedom
## Multiple R-squared:  0.5213, Adjusted R-squared:  0.4778
## F-statistic: 11.98 on 4 and 44 DF,  p-value: 1.139e-06
```

##Just a quick check to see if there was anything to the random variables I choose. THe main thing I took from this is that the state temperature is likely the highes predictor of deaths. And sure enough if we jsut do the one variable it gives an R sqyuat of 67% not bad for a cold day in May. So based on the model the colder the state the LESS Deaths – hum the virus didn't like the cold – or people stayed home - or … the list could go on. Smoking and pets did still look a bit promising. Weird huh?

```r
mydatalm = lm(deaths_per_pop ~ temp_rank, data = my_data)
summary(mydatalm)
```

```
##
## Call:
## lm(formula = deaths_per_pop ~ temp_rank, data = my_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4707 -0.6199  0.1177  0.5628  1.2625
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.052079   0.200577  20.202  < 2e-16 ***
## temp_rank   -0.029766   0.006983  -4.263 9.65e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6913 on 47 degrees of freedom
## Multiple R-squared:  0.2788, Adjusted R-squared:  0.2635
## F-statistic: 18.17 on 1 and 47 DF,  p-value: 9.653e-05
```

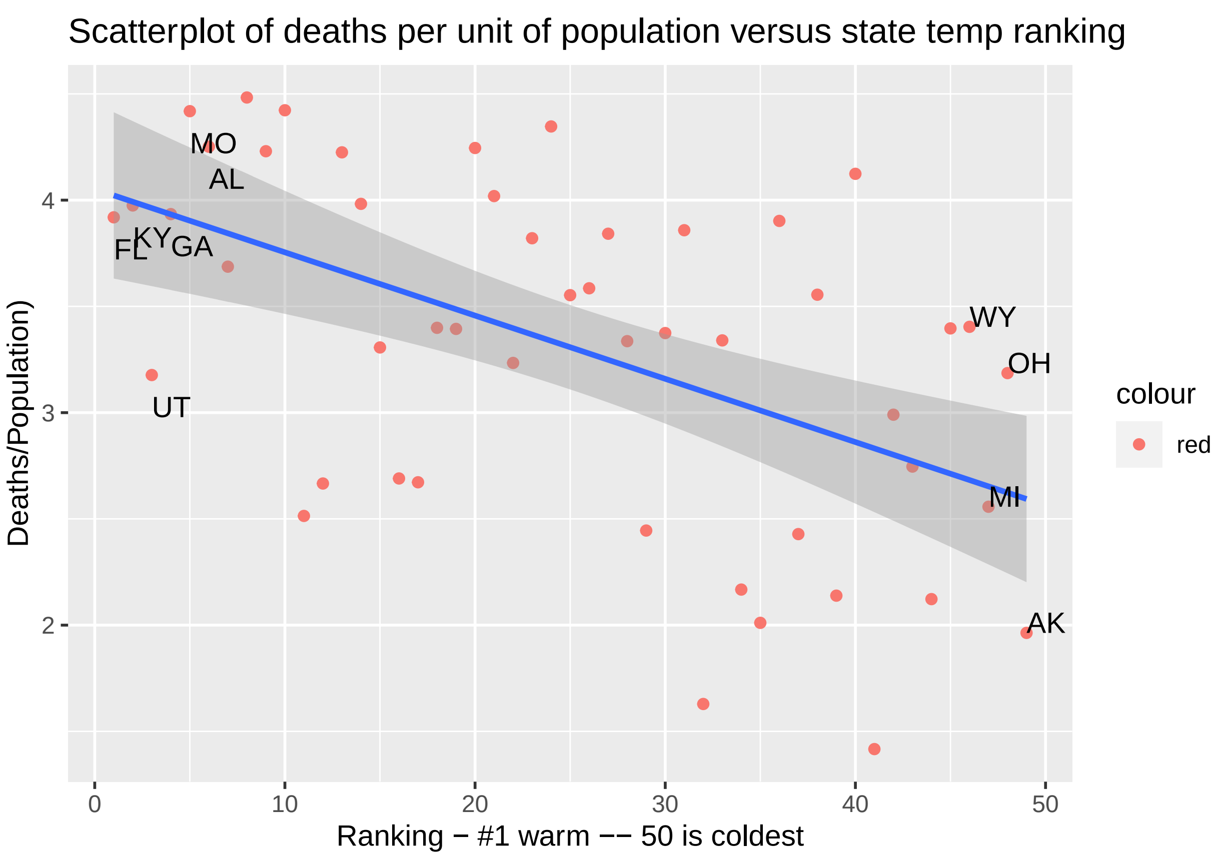
```r
my_data1 <- data.frame(state, pop_in_thousands,deaths_per_pop, temp_rank, medbed_per_thou, smoker_rate,
summary(my_data1)
```

```
##     state           pop_in_thousands  deaths_per_pop    temp_rank
```

```
##  Length:49          Min.   :  578.8   Min.   :1.417   Min.   : 1
##  Class :character   1st Qu.: 1934.4   1st Qu.:2.672   1st Qu.:13
##  Mode  :character   Median : 4648.8   Median :3.396   Median :25
##                     Mean   : 6665.6   Mean   :3.308   Mean   :25
##                     3rd Qu.: 7614.9   3rd Qu.:3.934   3rd Qu.:37
##                     Max.   :39512.2   Max.   :4.483   Max.   :49
##  medbed_per_thou  smoker_rate     pet_own_rate
##  Min.   :1.600   Min.   : 4.00   Min.   :45.40
##  1st Qu.:2.100   1st Qu.:10.00   1st Qu.:54.40
##  Median :2.500   Median :13.00   Median :59.40
##  Mean   :2.614   Mean   :12.33   Mean   :59.26
##  3rd Qu.:3.100   3rd Qu.:14.00   3rd Qu.:63.50
##  Max.   :4.800   Max.   :20.00   Max.   :71.80
```

```
ggplot(data = my_data1, aes(y = deaths_per_pop, x = temp_rank)) +
        geom_point(aes(color = "red")) +
        geom_smooth(method = "lm") +
        geom_text(aes(label=ifelse(temp_rank<7,as.character(state),'')),hjust=0,vjust=2) +
        geom_text(aes(label=ifelse(temp_rank>45,as.character(state),'')),hjust=0,vjust=0) +
        labs(title = "Scatterplot of deaths per unit of population versus state temp ranking",
             y = "Deaths/Population)",
             x = "Ranking - #1 warm -- 50 is coldest")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

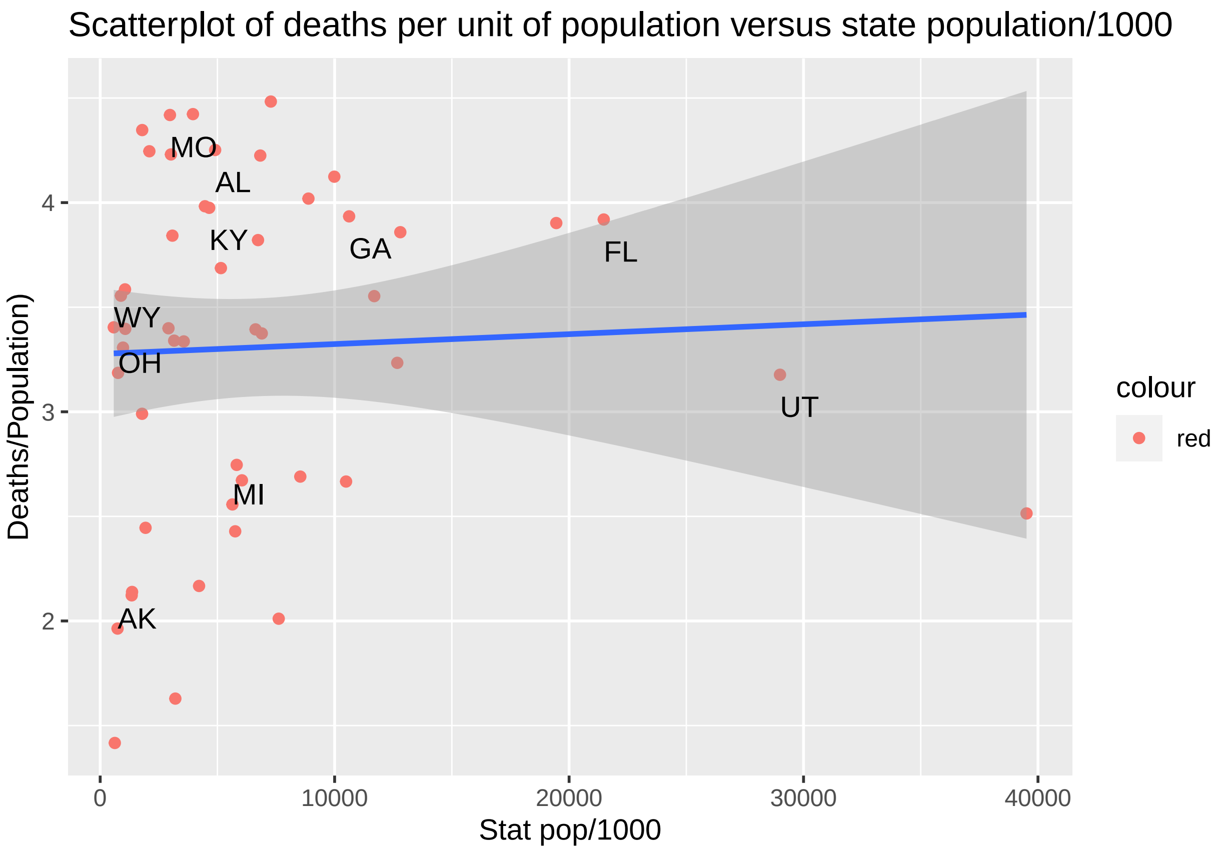


```r
mydatalm = lm(deaths_per_pop ~ pop_in_thousands, data = my_data)
summary(mydatalm)
```

```
##
## Call:
## lm(formula = deaths_per_pop ~ pop_in_thousands, data = my_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8626 -0.6327  0.1091  0.6077  1.1722
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.276e+00  1.568e-01   20.89   <2e-16 ***
## pop_in_thousands 4.734e-06  1.580e-05    0.30    0.766
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8133 on 47 degrees of freedom
## Multiple R-squared:  0.001905,   Adjusted R-squared:  -0.01933
## F-statistic: 0.08973 on 1 and 47 DF,  p-value: 0.7658
```

```r
ggplot(data = my_data1, aes(y = deaths_per_pop, x = pop_in_thousands)) +
        geom_point(aes(color = "red")) +
        geom_smooth(method = "lm") +
        geom_text(aes(label=ifelse(temp_rank<7,as.character(state),'')),hjust=0,vjust=2) +
        geom_text(aes(label=ifelse(temp_rank>45,as.character(state),'')),hjust=0,vjust=0) +
        labs(title = "Scatterplot of deaths per unit of population versus state population/1000",
             y = "Deaths/Population)",
             x = "Stat pop/1000")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



##In conclusion, or recap, I started the exploration as started in class then decided to look into DID OUR STATES VARY IN COVID DEATHS. the first look was always California had the most daeaths. Made me feel bad as this is home for now. So after sizing the data down to the total count of deaths by the year 2020 and adjusting it to cover the ratio of deaths per population–taking the shear number issue out of the equation. And sure enough the death rate ratios were all over the place. So a gleamed so data for other sources such as smoking, temperature, pet ownership (my favorite – as my pups saved me during our lock down). And sure enough starte temp had the highest correlation to death rate ratio…opposite of what I would have expected. The lowest death ratio states were the coldest states out there. We can have some fun trying to figure out why–altitude, pop density, snow drift removal exercise, etc.…I will leave that to another class or maybe my next course and some more downlads from the CDC.

#All in all I enjoyed our journey to explore how to manipulate a database and pull sosme visualizations/models from it. I found the tools covered in the classes amazing. I have a lot to learn but enjoy what we have covered so far. Thank you to all the staff and your help. Herb