Module 1 HW

Matt Scheffel

2022-09-05

#Worked with Dora Eskridge

```
UScovid<-read.csv("UScovid.csv", header=TRUE)
```

```
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("scales")
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

Question 1

1A

```
latest<-UScovid[which(UScovid$date=="2021-06-03"),]
latest<-latest[,c(0,2,3,5,6)]
latest<-latest[order(latest$county,latest$state),]

head(latest, n=6L)
```

```
##            county          state cases deaths
## 1383852 Abbeville South Carolina  2599     41
## 1382557    Acadia      Louisiana  6703    195
## 1384362  Accomack       Virginia  2862     43
## 1381993       Ada          Idaho 52964    475
```

```
## 1382232     Adair           Iowa   873     32
## 1382437     Adair       Kentucky  1944     54
```

1B

```
death.rate<-latest$deaths/latest$cases
death.rate<-percent(death.rate, accuracy=.01)
latest<-data.frame(latest,death.rate)
latest<-latest[,c(1,2,3,4,5)]
head(latest, n=6L)
```

```
##            county          state cases deaths death.rate
## 1383852 Abbeville South Carolina  2599     41      1.58%
## 1382557    Acadia      Louisiana  6703    195      2.91%
## 1384362  Accomack       Virginia  2862     43      1.50%
## 1381993       Ada          Idaho 52964    475      0.90%
## 1382232     Adair           Iowa   873     32      3.67%
## 1382437     Adair       Kentucky  1944     54      2.78%
```

1C

```
top10cases<-latest[with(latest,order(-cases)),]
head(top10cases, n=10L)
```

```
##                county      state   cases deaths death.rate
## 1381641    Los Angeles California 1245127  24375      1.96%
## 1383311  New York City   New York  949986  33257      3.50%
## 1382052           Cook   Illinois  554390  10893      1.96%
## 1381539       Maricopa    Arizona  551509  10084      1.83%
## 1381801     Miami-Dade    Florida  501925   6472      1.29%
## 1384160         Harris      Texas  401345   6462      1.61%
## 1384116         Dallas      Texas  303533   4082      1.34%
## 1381655      Riverside California  300879   4614      1.53%
## 1381658 San Bernardino California  298599   4760      1.59%
## 1381659      San Diego California  280410   3760      1.34%
```

1D

```
top10deaths<-latest[with(latest,order(-deaths)),]
head(top10deaths, n=10L)
```

```
##                county      state   cases deaths death.rate
## 1383311  New York City   New York  949986  33257      3.50%
## 1381641    Los Angeles California 1245127  24375      1.96%
## 1382052           Cook   Illinois  554390  10893      1.96%
## 1381539       Maricopa    Arizona  551509  10084      1.83%
## 1381801     Miami-Dade    Florida  501925   6472      1.29%
## 1384160         Harris      Texas  401345   6462      1.61%
## 1381652         Orange California  272242   5070      1.86%
## 1382761          Wayne   Michigan  164612   5048      3.07%
## 1381658 San Bernardino California  298599   4760      1.59%
## 1381655      Riverside California  300879   4614      1.53%
```

1E

```
top10deathrates<-latest[with(latest,order(death.rate)),]
top10deathrates<-na.omit(top10deathrates)
top10deathrates<-filter(top10deathrates, county!="Unknown")
tail(top10deathrates,n=10L)
```

```
##            county      state cases deaths death.rate
## 3134      Candler    Georgia   978     67      6.85%
## 3135 Throckmorton      Texas    73      5      6.85%
## 3136       Motley      Texas   116      8      6.90%
## 3137     Glascock    Georgia   269     19      7.06%
## 3138      Hancock    Georgia   928     68      7.33%
## 3139        Foard      Texas   124     10      8.06%
## 3140      Harding New Mexico    12      1      8.33%
## 3141    Petroleum    Montana    12      1      8.33%
## 3142       Sabine      Texas   524     45      8.59%
## 3143        Grant   Nebraska    41      4      9.76%
```

These counties all appear to have relatively low total number of cases and deaths.

1F

```
hundredkdeaths<-latest[which(latest$cases>99999),]
hundredkdeaths<-hundredkdeaths[with(hundredkdeaths,order(death.rate)),]
hundredkdeaths<-na.omit(hundredkdeaths)
tail(hundredkdeaths,n=10L)
```

```
##                county         state  cases deaths death.rate
## 1382741       Oakland      Michigan 118035   2368      2.01%
## 1381542          Pima       Arizona 116997   2406      2.06%
## 1381745     Fairfield   Connecticut 100093   2198      2.20%
## 1383035     St. Louis      Missouri 100195   2249      2.24%
## 1383750  Philadelphia  Pennsylvania 153521   3692      2.40%
## 1382728        Macomb      Michigan 100190   2441      2.44%
## 1383229        Bergen    New Jersey 104301   2868      2.75%
## 1382672     Middlesex Massachusetts 134980   3761      2.79%
## 1382761         Wayne      Michigan 164612   5048      3.07%
## 1383311 New York City      New York 949986  33257      3.50%
```

1G

```
CvilleData<-latest[which(latest$county=="Albemarle" | latest$county=="Charlottesville city"),]
head(CvilleData, n=2L)
```

```
##                       county    state cases deaths death.rate
## 1384363            Albemarle Virginia  5801     83      1.43%
## 1384385 Charlottesville city Virginia  4014     57      1.42%
```

Question 2

2A

```
state.level<-UScovid[which(UScovid$date=="2021-06-03"),]
state.level<-state.level[,c(3,5,6)]
state.level<-state.level[order(state.level$state),]
state.level<-state.level%>%
  group_by(state)%>%
  summarize(state_total_cases=sum(cases), state_total_deaths=sum(deaths))
head(state.level, n=6L)
```

```
## # A tibble: 6 x 3
##   state      state_total_cases state_total_deaths
##   <chr>                  <int>              <int>
## 1 Alabama               545028              11188
## 2 Alaska                 69826                352
## 3 Arizona               882691              17653
## 4 Arkansas              341889               5842
## 5 California           3793055              63345
## 6 Colorado              547961               6746
```

2B

```
state.rate<-state.level$state_total_deaths/state.level$state_total_cases
state.rate<-percent(state.rate, accuracy=.01)
state.level<-data.frame(state.level,state.rate)
head(state.level, n=6L)
```

```
##        state state_total_cases state_total_deaths state.rate
## 1    Alabama            545028              11188      2.05%
## 2     Alaska             69826                352      0.50%
## 3    Arizona            882691              17653      2.00%
## 4   Arkansas            341889               5842      1.71%
## 5 California           3793055              63345      1.67%
## 6   Colorado            547961               6746      1.23%
```

2C

The Virginia death rate is 1.66%.

2D

The Puerto Rico death rate is N/A due to missing data.

2E

These states have the 10 highest death rates:

```
top10staterates<-state.level[with(state.level,order(state.rate)),]
tail(top10staterates,n=11L)
```

```
##                   state state_total_cases state_total_deaths state.rate
## 22             Maryland            460406                9626      2.09%
## 33           New Mexico            203330                4275      2.10%
## 20            Louisiana            472617               10605      2.24%
## 41         Pennsylvania           1208879               27349      2.26%
## 26          Mississippi            318048                7324      2.30%
## 9  District of Columbia             49041                1136      2.32%
## 7           Connecticut            347748                8245      2.37%
## 34             New York           2102003               52811      2.51%
## 23        Massachusetts            707523               17893      2.53%
## 32           New Jersey           1017044               26253      2.58%
## 42          Puerto Rico            172414                  NA       <NA>
```

2F

These states have the 10 lowest death rates:

```
low10staterates<-state.level[with(state.level,order(state.rate)),]
head(low10staterates,n=10L)
```

```
##                       state state_total_cases state_total_deaths state.rate
## 2                    Alaska             69826                352      0.50%
## 48                     Utah            406895               2308      0.57%
## 50           Virgin Islands              3512                 28      0.80%
## 49                  Vermont             24240                255      1.05%
## 29                 Nebraska            223517               2385      1.07%
## 14                    Idaho            192704               2103      1.09%
## 37 Northern Mariana Islands               183                  2      1.09%
## 54                Wisconsin            675152               7923      1.17%
## 55                  Wyoming             60543                720      1.19%
## 6                  Colorado            547961               6746      1.23%
```

```
write.csv(state.level,"stateCovid.csv",row.names=TRUE)
```

Fin