uwa6xv_M01_HW

Alanna Hazlett

2024-02-01

Problem 1

(a) Create latest Read in UScovid.csv.

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

Get subset of data from June 3, 2021 and remove date and flip columns.

```
USCovid1.df<-USCovid.df[c(1381437:1384683),c(2,3,5,6)]
```

Remove rows with Unknown counties.

```
UnknownCounty<-which(USCovid1.df$county=="Unknown")
USCovid2.df<-USCovid1.df[-c(91,181,316,321,386,535,551,692,884,1177,1199,1224,1241,1322,1405,1757,1791,
```

Order by county then state alphabetically. Name 'latest' and head() first 6 rows.

```
latest<-USCovid2.df[order(USCovid2.df$county,USCovid2.df$state),]
head(latest)
```

```
##            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
```

(b) County case fatality rate (numeric variable to numeric variable).

```
death.rate<-(latest$deaths / latest$cases) * 100
death.rate<-round(death.rate,2)
```

Add death.rate to latest and display first 6 rows.

```
latest<-data.frame(latest,death.rate)
head(latest)
```

```
##            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
```

(c) Display the 10 largest cases by county.

```
LargeCases<-latest[order(-latest$cases),]
head(LargeCases,10)
```

```
##                 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
```

(d) Display the 10 largest deaths by county.

```
LargeDeaths<-latest[order(-latest$deaths),]
head(LargeDeaths,10)
```

```
##                 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
```

(e) Display the 10 largest case fatality rates by county.

```
LargeDeathRate<-latest[order(-latest$death.rate),]
head(LargeDeathRate,10)
```

```
##               county      state cases deaths death.rate
## 1383143        Grant   Nebraska    41      4       9.76
## 1384261       Sabine      Texas   524     45       8.59
## 1383261      Harding New Mexico    12      1       8.33
## 1383084    Petroleum    Montana    12      1       8.33
## 1384137        Foard      Texas   124     10       8.06
## 1381896      Hancock    Georgia   928     68       7.33
## 1381888     Glascock    Georgia   269     19       7.06
## 1384232       Motley      Texas   116      8       6.90
## 1381847      Candler    Georgia   978     67       6.85
## 1384283 Throckmorton      Texas    73      5       6.85
```

(f) Display the counties with the 10 highest case fatality rates among counties with at least 100,000 cases.

```
LargeCountyLargeRate<-latest[which(latest$cases>99999),]
LargeCountyLargeRate<-LargeCountyLargeRate[order(-LargeCountyLargeRate$death.rate),]
head(LargeCountyLargeRate,10)
```

```
##                 county         state  cases deaths death.rate
## 1383311 New York City      New York 949986  33257       3.50
```

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

(g) Display the number of cases, deaths, and case fatality rates for Albemarle, Virginia and Charlottesville city, Virginia.

```
Albemarle<-latest[c(35),]
head(Albemarle)
```

```
##            county    state cases deaths death.rate
## 1384363 Albemarle Virginia  5801     83       1.43
```

```
Charlottesville<-latest[c(474),]
head(Charlottesville)
```

```
##                        county    state cases deaths death.rate
## 1384385 Charlottesville city Virginia  4014     57       1.42
```

Problem 2

(a) Create state.level Select rows with June 3, 2021 date; select columns state, number of cases, and number of deaths.

```
USCovid3.df<-USCovid.df[c(1381437:1384683),c(3,5,6)]
```

Create state.level with sums of cases and sums of deaths by state. *Received help from Karunya Iyappan on Piazza to add state names column.*

```
state.names<-unique(USCovid3.df$state)
state.cases<-tapply(USCovid3.df$cases,USCovid3.df$state,sum)
state.deaths<-tapply(USCovid3.df$deaths,USCovid3.df$state,sum)
state.level<-data.frame(state.names,state.cases,state.deaths)
```

Remove row labels (non-working column). *Received help from Karunya Iyappan on Piazza to remove state row names.*

```
rownames(state.level)<-c()
```

Renaming columns of state.level.

```
names(state.level)[c(1,2,3)]<-c("state","cases","deaths")
head(state.level)
```

```
##        state   cases deaths
## 1    Alabama  545028  11188
## 2     Alaska   69826    352
## 3    Arizona  882691  17653
## 4   Arkansas  341889   5842
## 5 California 3793055  63345
## 6   Colorado  547961   6746
```

(b) State case fatality rate (numeric variable to numeric variable).

```
state.rate<-(state.level$deaths / state.level$cases) * 100
state.rate<-round(state.rate,2)
```

Add state.rate to state.level and display the first 6 rows.

```
state.level<-data.frame(state.level,state.rate)
head(state.level)
```

```
##        state   cases 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
```

(c) The case fatality rate for Virginia is 1.66%.

```
Virginia<-state.level[c(51),]
head(Virginia)
```

```
##       state  cases deaths state.rate
## 51 Virginia 676041  11216       1.66
```

(d) The case fatality rate for Puerto Rico is N/A. The fatality rate can not be calculated, because there are no recorded deaths for Puerto Rico.

```
Puerto_Rico<-state.level[c(42),]
head(Puerto_Rico)
```

```
##          state  cases deaths state.rate
## 42 Puerto Rico 172414     NA         NA
```

(e) Display the 10 largest state case fatality rates. The 10 states with the highest fatality rates (from highest to lowest) are New Jersey, Massachusetts, New York, Connecticut, District of Columbia, Mississippi, Pennsylvania, Lousiana, New Mexico, and Maryland.

```
LargeStateRate<-state.level[order(-state.level$state.rate),]
head(LargeStateRate,10)
```

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

(f) Display the 10 lowest state case fatality rates. The 10 states with the lowest case fatality rates (from lowest to highest) are Alaska, Utah, Virgin Islands, Vermont, Nebraska, Idaho, Northern Mariana Islands, Wisconsin, Wyoming, and Colorado.

```
SmallStateRate<-state.level[order(state.level$state.rate),]
head(SmallStateRate,10)
```

```
##                      state  cases 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
```

(g) Create csv called stateCovid.csv.

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