```r
# Part A:
covid_dataset <- read.csv("UScovid.csv")

# filter covid_dataset
covid_dataset <- covid_dataset[covid_dataset$date == "2021-06-03" &
covid_dataset$county != "Unknown", ]

# removes "date" and "fips"
covid_dataset <- covid_dataset[, !(names(covid_dataset) %in% c("date",
"fips"))]

# alphebectial order
covid_dataset <- covid_dataset[order(covid_dataset$state,
covid_dataset$county), ]


head(covid_dataset)
```

```
##          county   state cases deaths
## 1381437 Autauga Alabama  7172    111
## 1381438 Baldwin Alabama 21684    312
## 1381439 Barbour Alabama  2343     59
## 1381440    Bibb Alabama  2665     64
## 1381441  Blount Alabama  6894    139
## 1381442 Bullock Alabama  1236     42
```

```r
# Part B: Death Rate column calulated from latest
covid_dataset$death.rate <- round((covid_dataset$deaths /
covid_dataset$cases) * 100, 2)

head(covid_dataset)
```

```
##          county   state cases deaths death.rate
## 1381437 Autauga Alabama  7172    111       1.55
## 1381438 Baldwin Alabama 21684    312       1.44
## 1381439 Barbour Alabama  2343     59       2.52
## 1381440    Bibb Alabama  2665     64       2.40
## 1381441  Blount Alabama  6894    139       2.02
## 1381442 Bullock Alabama  1236     42       3.40
```

```r
# Part C: Top 10 countries with largest number of cases
top_cases <- covid_dataset[order(covid_dataset$cases, decreasing = TRUE),
][1:10, c("state", "county", "cases", "deaths", "death.rate")]
print(top_cases)
```

```
##              state         county   cases deaths death.rate
## 1381641 California    Los Angeles 1245127  24375       1.96
## 1383311   New York  New York City  949986  33257       3.50
## 1382052   Illinois           Cook  554390  10893       1.96
## 1381539    Arizona       Maricopa  551509  10084       1.83
```

```
## 1381801    Florida     Miami-Dade  501925   6472       1.29
## 1384160      Texas         Harris  401345   6462       1.61
## 1384116      Texas         Dallas  303533   4082       1.34
## 1381655 California      Riverside  300879   4614       1.53
## 1381658 California San Bernardino  298599   4760       1.59
## 1381659 California      San Diego  280410   3760       1.34

# Part D: Top 10 counties with the largest number of deaths
top_deaths <- covid_dataset[order(covid_dataset$deaths, decreasing = TRUE),
][1:10, c("state", "county", "cases", "deaths", "death.rate")]
print(top_deaths)

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

#Part E : Top 10 counties with the highest case fatality rates

top_fatality <- covid_dataset[order(covid_dataset$death.rate, decreasing =
TRUE), ][1:10, c("state", "county", "cases", "deaths", "death.rate")]

print(top_fatality)

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

#Part F: Top 10 counties with the highest case fatality rates and at least
100,000 cases

top_100k_cases <- covid_dataset[covid_dataset$cases >= 100000, ]

top_fatality_100k <- top_100k_cases[order(top_100k_cases$death.rate,
decreasing = TRUE), ][1:10, c("state", "county", "cases", "deaths",
"death.rate")]
```

```
print(top_fatality_100k)
```

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

```
# Part G: Albemarle, Virginia

print(covid_dataset[covid_dataset$county == "Albemarle" & covid_dataset$state
== "Virginia", c("county", "state", "cases", "deaths", "death.rate")])
```

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

```
# Part G: Charlottesville city, Virgnia

print(covid_dataset[covid_dataset$county == "Charlottesville city" &
covid_dataset$state == "Virginia", c("county", "state", "cases", "deaths",
"death.rate")])
```

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

```
#Part A: State Level Anaylsis
state_leveldata <- aggregate(cbind(cases, deaths) ~ state, data =
covid_dataset, sum)


state_leveldata <- state_leveldata[order(state_leveldata$state), ]


head(state_leveldata)
```

```
##        state   cases deaths
## 1    Alabama  545028  11188
## 2     Alaska   69534    352
## 3    Arizona  882691  17653
## 4   Arkansas  338986   5842
## 5 California 3793055  63345
## 6   Colorado  547961   6746
```

```r
# Part B: Calculate Case Fatality Rate at the state level
state_leveldata$state.rate <- round((state_leveldata$deaths /
state_leveldata$cases) * 100, 2)

# Display the first 6 rows of the state-level data with the new death rate
column
head(state_leveldata)
```

```
##        state   cases deaths state.rate
## 1    Alabama  545028  11188       2.05
## 2     Alaska   69534    352       0.51
## 3    Arizona  882691  17653       2.00
## 4   Arkansas  338986   5842       1.72
## 5 California 3793055  63345       1.67
## 6   Colorado  547961   6746       1.23
```

```r
#Part (c) - Case Fatality Rate for Virginia
virginia_rate <- state_leveldata[state_leveldata$state == "Virginia",
"state.rate"]
print(virginia_rate)
```

```
## [1] 1.66
```

```r
# 3. Part D: Case Fatality Rate for Puerto Rico
puerto_rico_rate <- state_leveldata[state_leveldata$state == "Puerto Rico",
"state.rate"]
print(puerto_rico_rate)
```

```
## numeric(0)
```

```r
# Part E: The highest top 10 fatality rate
top_states_fatality <- state_leveldata[order(state_leveldata$state.rate,
decreasing = TRUE), ][1:10, ]
print(top_states_fatality)
```

```
##                   state   cases deaths state.rate
## 22        Massachusetts  660563  17881       2.71
## 31           New Jersey 1016219  26253       2.58
## 33             New York 2102003  52811       2.51
## 7           Connecticut  346564   8244       2.38
## 9  District of Columbia   49041   1136       2.32
## 25          Mississippi  318048   7324       2.30
## 40         Pennsylvania 1208879  27349       2.26
## 19            Louisiana  472222  10605       2.25
## 32           New Mexico  203330   4275       2.10
## 21             Maryland  460406   9587       2.08
```

```r
# Part F: States with the 10 lowest fatality
bottom_states_fatality <- state_leveldata[order(state_leveldata$state.rate,
decreasing = FALSE), ][1:10, ]
print(bottom_states_fatality)
```

```
##                         state  cases deaths state.rate
## 2                      Alaska  69534    352       0.51
## 46                       Utah 405721   2286       0.56
## 48             Virgin Islands   3512     28       0.80
## 47                    Vermont  24218    255       1.05
## 28                   Nebraska 222317   2385       1.07
## 13                      Idaho 192704   2103       1.09
## 36 Northern Mariana Islands    183      2       1.09
## 52                  Wisconsin 675152   7923       1.17
## 53                    Wyoming  60543    720       1.19
## 6                     Colorado 547961   6746       1.23

#CSV export
write.csv(state_leveldata, "stateCovid.csv", row.names = FALSE)
```

```

1.What is the case fatality rate in Virginia? the case fertility rate was 1.66

2.What is the case fatality rate in Puerto Rico? The case fertility rate for Puerto Rico was 0. This was not in the dataset.

3. Which states have the 10 highest case fatality rate? Massachusetts
   New Jersey New York Connecticut District of Columbia Mississippi Pennsylvania Louisiana
   New Mexico Maryland

4.Which states have the 10 lowest case fatality rate? Alaska
Utah Virgin Islands
Vermont Nebraska
Idaho
Northern Mariana Islands
Wisconsin
Wyoming Colorado
```