

HW 1 Solutions

Question 1

(a)

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

##create data frame by criteria
latest<-Covid[which(Covid$date == "2021-06-03"
                    & Covid$county != "Unknown"),
              -c(1,4)]

latest<-latest[order(latest$county, latest$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)

```
##calculate case fatality rate, convert to %, round to 2dp
death.rate<-round(latest$deaths/latest$cases * 100, 2)

##add case fatality rate to data frame
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)

```
##find counties with 10 highest number of cases
```

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

```
latest[1:10,]
```

##	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)

```
##find counties with 10 highest number of deaths
```

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

```
latest[1:10,]
```

##	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)

```
##find counties with 10 highest case fatality rates
```

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

```
latest[1:10,]
```

##	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	Glascok	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

These counties have small case numbers.

(f)

```
##consider counties with at least 100,000 cases
most.cases<-latest[which(latest$cases >= 100000),]

##find counties with 10 highest case fatality rates,
##with at least 100,000 cases
most.cases<-most.cases[order(-most.cases$death.rate),]
most.cases[1:10,]
```

##	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)

```
##find numbers for Albemarle county, VA
latest[which(latest$county=="Albemarle"
             & latest$state=="Virginia"),]
```

##	county	state	cases	deaths	death.rate
----	--------	-------	-------	--------	------------

```
## 1384363 Albemarle Virginia 5801      83      1.43
##find numbers for Charlottesville, VA
latest[which(latest$county=="Charlottesville city"
             & latest$state=="Virginia"),]

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

Question 2

(a)

Since we are looking at state level data, we can include counties that are unknown, since their states are known, unlike question 1. We just want data from June 3 2021, and the number of cases, deaths, and name of state.

```
counties.latest<-Covid[which(Covid$date == "2021-06-03"), -c(1,4)]
```

This data frame still has data at the county level. To find data at the state level, we can add up the cases and deaths for all counties in each state.

```
##total cases by state
cases.state<-tapply(counties.latest$cases,
                    counties.latest$state, sum, na.rm=T)

state1<-data.frame(State=names(cases.state),
                  Cases=cases.state)

##total deaths by state.
deaths.state<-tapply(counties.latest$deaths,
                     counties.latest$state, sum, na.rm=T)

state2<-data.frame(State=names(deaths.state),
                  Deaths=deaths.state)

##merge cases and deaths into a data frame
state.level<-merge(state1,state2, by="State")

state.level<-state.level[order(state.level$State),]

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)

```
##calculate case fatality rate
state.rate<-round(state.level$Deaths/state.level$Cases * 100, 2)

##add case fatality rate to data frame
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)

```
state.level[which(state.level$State=="Virginia"),4]
```

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

The case fatality rate in Virginia is 1.66%.

(d)

```
state.level[which(state.level$State=="Puerto Rico"),4]
```

```
## [1] 1.46
```

The case fatality rate in Puerto Rico is 1.46%.

(e)

```
##10 highest case fatality rates
state.level<-state.level[order(-state.level$state.rate),]
state.level[1:10,]
```

```
##      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)

```
##10 lowest case fatality rates
```

```
state.level<-state.level[order(state.level$state.rate),]
state.level[1:10,]
```

##	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)

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