BST 270 Final Project

Corriene Sept

1/19/2021

- 1. New cases as a function of time with a rolling average plot
- 2. Table of cases, hospitalizations, and deaths
- 4. Table of cases by state

Load data:

```
us <- read_csv("~/R/BST270-Winter2021/covid-19-data-master/us.csv")
```

```
## Parsed with column specification:
## cols(
##   date = col_date(format = ""),
##   cases = col_double(),
##   deaths = col_double()
## )
```

```
us_states <- read_csv("~/R/BST270-Winter2021/covid-19-data-master/us-states.csv")
```

```
## Parsed with column specification:
## cols(
##   date = col_date(format = ""),
##   state = col_character(),
##   fips = col_character(),
##   cases = col_double(),
##   deaths = col_double()
## )
```

```
national_history <- read_csv("~/R/BST270-Winter2021/national-history.csv")
```

```
## Parsed with column specification:
## cols(
##   date = col_date(format = ""),
##   death = col_double(),
##   deathIncrease = col_double(),
##   inIcuCumulative = col_double(),
##   inIcuCurrently = col_double(),
##   hospitalizedIncrease = col_double(),
##   hospitalizedCurrently = col_double(),
##   hospitalizedCumulative = col_double(),
##   negative = col_double(),
##   negativeIncrease = col_double(),
##   onVentilatorCumulative = col_double(),
##   onVentilatorCurrently = col_double(),
##   positive = col_double(),
##   positiveIncrease = col_double(),
##   states = col_double(),
##   totalTestResults = col_double(),
##   totalTestResultsIncrease = col_double()
## )
```

Note that all below analyses are for **January 17, 2021.**

1. New cases as a function of time with a rolling average plot

First, compute new cases and new deaths on each day.

```
us$new_cases <- us$cases - lag(us$cases)
us$new_deaths <- us$deaths - lag(us$deaths)
```

Compute 7 day average. Note that the previous 7 days are averaged (this is NOT center aligned.) The first 7 days (chronologically) in the dataset are then missing.

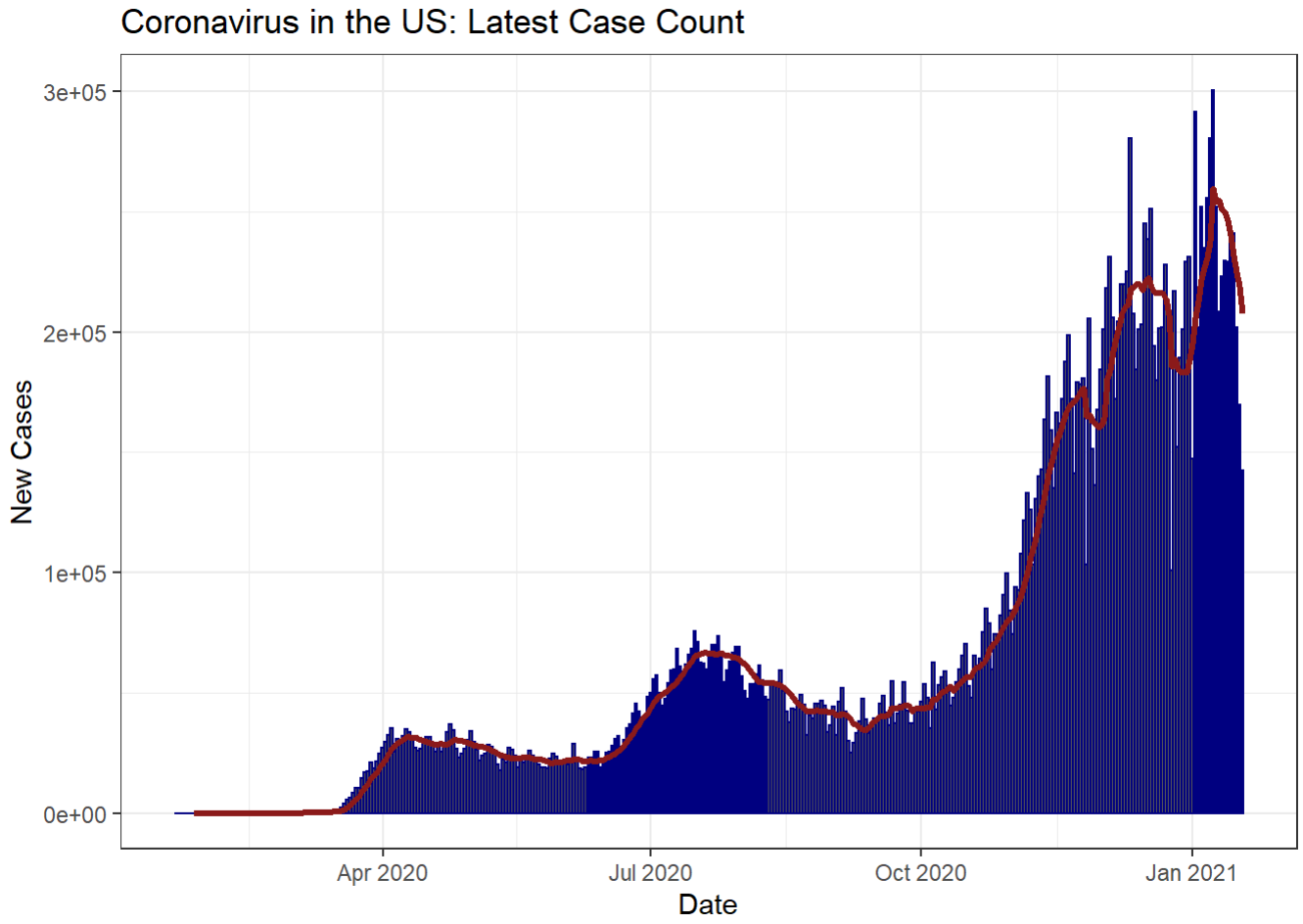
```
us$new_cases_7dayavg <- rollmean(us$new_cases, k = 7, fill = NA, align = "right")
us$new_deaths_7dayavg <- rollmean(us$new_deaths, k = 7, fill = NA, align = "right")
us$deaths_7dayavg <- rollmean(us$deaths, k = 7, fill = NA, align = "right")
us$cases_7dayavg <- rollmean(us$cases, k = 7, fill = NA, align = "right")
national_history$hospitalizedCurrently_7dayavg <- rollmean(national_history$hospitalizedCurrentl
y, k = 7, fill = NA, align = "left")
```

Create the plot: new cases as a function of time with a rolling average. The trendine is the 7-day average computed above, and the bar chart is *new cases.*

```
p <- ggplot(us, aes(date, new_cases)) +
           geom_bar(stat="identity", na.rm = TRUE, color = "navy") +
  ggtitle("Coronavirus in the US: Latest Case Count") +
           xlab("Date") + ylab("New Cases")

p + geom_line(aes(date, new_cases_7dayavg), na.rm = TRUE, color = "firebrick4", size = 1.1) + th
eme_bw()
```

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



Brief critique of reproducibility:

This plot was easily reproduced using the above code. I didn't face any issues trying to reproduce this; it was pretty self explanatory.

2. Table of cases, hospitalizations, and deaths

The first two columns of the table, total reported cases & deaths and Jan 17 new cases, new deaths, and currently hospitalized is already contained in the datasets us and national_history. The new cases and new deaths were computed previously. 14-day change was computed using 7-day averages for new cases, new deaths, and currently hospitalized patients. This is the percent change between the 7 day averages of new cases, new deaths, and currently hospitalized on January 3 and January 17.

Filter to January 3rd and 17th to compute the 3rd column of the table.

```
Jan03 <- us %>%
  filter(date == "2021-01-03")
Jan17 <- us %>%
  filter(date == "2021-01-17")
```

14 day % increase: hospitalized currently 7 day average, Jan 17 and Jan 03

```
round((national_history[2,]$hospitalizedCurrently_7dayavg - national_history[16,]$hospitalizedCu
rrently_7dayavg) / national_history[16,]$hospitalizedCurrently_7dayavg * 100)
```

```
## [1] 3
```

14 day % increase: NEW cases 7 day average, Jan 17 and Jan 03

```
round((Jan17$new_cases_7dayavg - Jan03$new_cases_7dayavg) / Jan03$new_cases_7dayavg * 100)
```

```
## [1] 3
```

14 day % increase: NEW deaths 7 day average, Jan 17 and Jan 03

```
round(((Jan17$new_deaths_7dayavg - Jan03$new_deaths_7dayavg) / Jan03$new_deaths_7dayavg)*100)
```

```
## [1] 26
```

Make the rows of the table:

```
Cases <- c(us[363,2],us[363,4], "+3%")
Deaths <- c(us[363,3], us[363,3] - us[362,3], "+26%")
Hospitalized <- c("", national_history$hospitalizedCurrently[2], "+3%")
```

Print the table:

```
data <- as.matrix(rbind(Cases, Deaths, Hospitalized))
colnames(data) <- c("Total Reported","On Jan 17", "14-day Change")
data%>%
  kbl() %>%
  kable_styling()
```

	Total Reported	On Jan 17	14-day Change
Cases	23983607	169641	+3%
Deaths	397612	1730	+26%
Hospitalized		124387	+3%

Brief critique of reproducibility: This was more challenging to reproduce, mostly because the 14-day change is computed using 7-day averages and at first it wasn't terribly clear how this was done. For one, it needs to be aligned so that the previous 7 days are being averaged so the current date isn't missing. Also, it uses *NEW* deaths, *NEW* cases, and currently hospitalized patients to compute these metrics, not cumulative deaths and cases, and this was not explicitly stated.

4. Table of cases by state

Compute new cases by state

```
us_states <- us_states %>%
  group_by(state) %>%
  mutate(new_cases = cases - lag(cases))
```

Compute 7 day averages of new cases (aligned so the previous 7 days are being averaged) for each state.

```
us_states <- us_states %>%
  group_by(state) %>%
  mutate(new_cases_7dayavg = rollmean(new_cases, k = 7, fill = NA, align = "right"))
```

Filter to January 17, 2021 and display the table:

```
us_states %>%
  filter(date == "2021-01-17") %>%
  select(c(cases, new_cases_7dayavg)) %>%
  kbl() %>%
  kable_styling()
```

```
## Adding missing grouping variables: `state`
```

state	cases	new_cases_7dayavg
Alabama	422598	2.956857e+03
Alaska	51630	2.421429e+02
Arizona	673882	7.905143e+03
Arkansas	271154	2.296857e+03
California	3006583	3.957971e+04
Colorado	376921	1.986000e+03
Connecticut	223422	2.489714e+03
Delaware	70294	7.172857e+02
District of Columbia	33851	2.942857e+02

state	cases	new_cases_7dayavg
Florida	1571271	1.346700e+04
Georgia	791322	8.457286e+03
Guam	8453	1.028571e+01
Hawaii	24309	1.441429e+02
Idaho	155617	8.392857e+02
Illinois	1071307	5.717000e+03
Indiana	593262	3.798857e+03
Iowa	305151	1.209857e+03
Kansas	259226	1.956714e+03
Kentucky	329816	3.284714e+03
Louisiana	368980	3.164429e+03
Maine	33559	6.087143e+02
Maryland	326648	2.827857e+03
Massachusetts	470140	5.335571e+03
Michigan	580394	2.845571e+03
Minnesota	446448	1.401143e+03
Mississippi	252475	1.913286e+03
Missouri	468106	2.460000e+03
Montana	89393	4.384286e+02
Nebraska	182139	8.484286e+02
Nevada	261847	1.869857e+03
New Hampshire	56864	7.520000e+02
New Jersey	627221	6.056143e+03
New Mexico	163637	1.068571e+03
New York	1242818	1.528129e+04

state	cases	new_cases_7dayavg
North Carolina	675272	7.395571e+03
North Dakota	95886	1.660000e+02
Northern Mariana Islands	128	4.285714e-01
Ohio	826754	7.098429e+03
Oklahoma	354979	3.373857e+03
Oregon	133205	1.074571e+03
Pennsylvania	772747	6.748714e+03
Puerto Rico	120996	6.088571e+02
Rhode Island	104443	9.755714e+02
South Carolina	388184	4.808429e+03
South Dakota	105544	3.180000e+02
Tennessee	672292	4.452714e+03
Texas	2127334	2.278214e+04
Utah	323837	2.548286e+03
Vermont	10057	1.557143e+02
Virgin Islands	2260	3.114286e+01
Virginia	439305	5.778429e+03
Washington	293296	2.423286e+03
West Virginia	108821	1.087000e+03
Wisconsin	568166	2.558000e+03
Wyoming	49363	3.615714e+02

Brief critique of reproducibility: This wasn't too difficult to reproduce, but including the information that the 7 day average was computed using new cases and was aligned so the previous 7 days are being averaged would have been helpful.