Lab Assignment 13

Monica Buczynski

4/28/2020

```
library(tidyverse)
dat <- read_csv("https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv")
```

Question 1. The first date reported in the data is January 21, 2020. Find the latest available date reported in these data.

The last available date reported is April 28, 2020. There is a one day lag in the data. Note: the remainder of the assignment was completed yesterday, on April 27th, 2020.

```
dat %>%
arrange(desc(date))
```

```
## # A tibble: 310,256 x 6
##    date       county   state   fips  cases deaths
##    <date>     <chr>    <chr>   <chr> <dbl>  <dbl>
##  1 2020-07-08 Autauga  Alabama 01001   651     13
##  2 2020-07-08 Baldwin  Alabama 01003  1056     10
##  3 2020-07-08 Barbour  Alabama 01005   366      2
##  4 2020-07-08 Bibb     Alabama 01007   201      1
##  5 2020-07-08 Blount   Alabama 01009   262      1
##  6 2020-07-08 Bullock  Alabama 01011   374     11
##  7 2020-07-08 Butler   Alabama 01013   639     28
##  8 2020-07-08 Calhoun  Alabama 01015   411      5
##  9 2020-07-08 Chambers Alabama 01017   660     27
## 10 2020-07-08 Cherokee Alabama 01019   114      7
## # ... with 310,246 more rows
```

Question 2. Find the cumulative number of deaths reported in the U.S. to date.

The cumulative number of deaths reported in the U.S to date (April 27, 2020) is 744,217.

```
dat %>%
summarize(total_deaths = sum(deaths))
```

```
## # A tibble: 1 x 1
##   total_deaths
##          <dbl>
## 1      8244765
```

Question 3. Find the cumulative number of cases reported in the U.S. to date.

The cumulative number of cases reported in the U.S. to date (April 27, 2020) is 17,477,653.

```
dat %>%
summarize(total_cases = sum(cases))
```

```
## # A tibble: 1 x 1
##   total_cases
##         <dbl>
## 1   154536061
```

Question 4. Which state reported the most total cases on the most recent date available?

New York, New Jersey, Massachusetts, Illinois and California have reported the most cases on April 27, 2020.

```
dat %>%
group_by(state,date) %>%
summarize(total_cases = sum(cases)) %>%
filter(date =="2020-04-27") %>%
arrange(desc(total_cases))
```

```
## # A tibble: 55 x 3
## # Groups:   state [55]
##    state         date       total_cases
##    <chr>         <date>           <dbl>
##  1 New York      2020-04-27      296991
##  2 New Jersey    2020-04-27      111188
##  3 Massachusetts 2020-04-27       56462
##  4 Illinois      2020-04-27       45883
##  5 California    2020-04-27       45208
##  6 Pennsylvania  2020-04-27       43728
##  7 Michigan      2020-04-27       38457
##  8 Florida       2020-04-27       32130
##  9 Louisiana     2020-04-27       27111
## 10 Connecticut   2020-04-27       25997
## # ... with 45 more rows
```

Question 5. Which county(ies) in the U.S. has/have the fewest cumulative confirmed cases to date?

As of April 27, 2020, the counties of Arthur, Boundary, Cottle, Day, Pushmataha have the fewest cumulative confirmed cases.

```
dat %>%
group_by(county) %>%
summarize(total_cases = sum(cases)) %>%
arrange(total_cases)
```

```
## # A tibble: 1,814 x 2
##    county               total_cases
##    <chr>                      <dbl>
```

```
##  1 Hickory                        1
##  2 Towner                         2
##  3 Haakon                         3
##  4 Kingman                        3
##  5 Gilliam                        5
##  6 Ontonagon                      5
##  7 Kenedy                         6
##  8 Throckmorton                   6
##  9 Kusilvak Census Area           7
## 10 Hooker                        11
## # ... with 1,804 more rows
```

Question 6. Which county in Pennsylvania has the most total cases reported, to date? How many cases have they identified?

The top 5 counties in Pennsylvania has the most total cases reported are Philadelphia (199412), Montgomery (65494), Delaware (51802) Lehigh (47474) and Bucks (40244).

```
dat %>%
group_by(county,state) %>%
summarize(total_cases = sum(cases)) %>%
filter(state =="Pennsylvania") %>%
arrange(desc(total_cases))
```

```
## # A tibble: 68 x 3
## # Groups:   county [68]
##    county       state        total_cases
##    <chr>        <chr>              <dbl>
##  1 Philadelphia Pennsylvania     1787962
##  2 Montgomery   Pennsylvania      561942
##  3 Delaware     Pennsylvania      497698
##  4 Bucks        Pennsylvania      392237
##  5 Berks        Pennsylvania      318082
##  6 Lehigh       Pennsylvania      314784
##  7 Lancaster    Pennsylvania      257080
##  8 Northampton  Pennsylvania      245083
##  9 Luzerne      Pennsylvania      230230
## 10 Chester      Pennsylvania      219039
## # ... with 58 more rows
```

Question 7. Make a plot of the number of cases over time in Westmoreland County—where St. Vincent College is located.

```
dat %>%
    group_by(county, state, date) %>%
summarize(total_cases = sum(cases)) %>%
  filter(county == "Westmoreland", state == "Pennsylvania") %>%
ggplot(aes(x = date, y = total_cases, col = county)) +
geom_line() +
geom_point() +
  scale_y_log10() +
labs(y = "Total Cases (log scale)",
```

```
x = "Date",
title = "Total COVID-19 Cases in Westmoreland County, PA") +
guides(color = FALSE) +
theme(plot.title = element_text(hjust = 0.5)) # centers title
```