COVID-19 in Cook County

Emily Dobson

5/10/2020

COVID-19 in Cook County, Illinois

This document draws upon data from the New York Times on COVID-19 cases and deaths. The source data is publicly available on GitHub. The NYT provides counts of total cases and deaths to date for every county in the United States. For the visualization I would like to construct, I will need data on *new cases by day* for Cook County, Illinois.

Data Transformations

```
nyt <- read_csv("us-counties.csv")
```

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

In the first section of this code, I `mutate()` the data into a more usable format. First, I filter the data to only include rows with information for Cook County, IL. Next, I format the date to be in a more easily readable format. I also calculate the growth rate of total COVID cases along with the number of new cases and deaths reported each day.

```
cook_cnty <- nyt %>%
  filter(state == "Illinois", county == "Cook") %>%
  complete(date) %>%
  mutate(date = format(date, "%m-%d"),
         cases_pct_increase = ((cases - lag(cases))/lag(cases))*100,
         deaths_pct_increase = ((deaths - lag(deaths))/lag(deaths))*100,
         new_cases = cases - lag(cases),
         new_deaths = deaths - lag(deaths))
```

Once these transformations have been performed, a few rows have unusual values because of my use of the `lag()` function and division by zero. My next step is to clean up these values.

```
filter(cook_cnty, is.na(cases_pct_increase))
```

```
## # A tibble: 1 x 10
##   date  county state fips   cases deaths cases_pct_incre~ deaths_pct_incr~
##   <chr> <chr>  <chr> <chr>  <dbl>  <dbl>            <dbl>            <dbl>
## 1 01-24 Cook   Illi~ 17031      1      0               NA               NA
## # ... with 2 more variables: new_cases <dbl>, new_deaths <dbl>
```

```
filter(cook_cnty, is.na(new_cases))
```

```
## # A tibble: 1 x 10
##   date  county state fips   cases deaths cases_pct_incre~ deaths_pct_incr~
##   <chr> <chr>  <chr> <chr>  <dbl>  <dbl>            <dbl>            <dbl>
## 1 01-24 Cook   Illi~ 17031      1      0               NA               NA
## # ... with 2 more variables: new_cases <dbl>, new_deaths <dbl>
```

The only `NA` values in the `cases_pct_increase` and `new_cases` columns are found in the first row. I repopulate these `NA` values with values of 100 and 1, respectively, since only one case was reported on January 24th and this case represents a 100% increase over the previous day.

```
cook_cnty <- cook_cnty %>%
  mutate(cases_pct_increase = replace(cases_pct_increase,
                                      which(is.na(cases_pct_increase)), 100),
         new_cases = replace(new_cases, which(is.na(new_cases)), 1))

filter(cook_cnty, is.na(deaths_pct_increase))
```

```
## # A tibble: 53 x 10
##    date  county state fips   cases deaths cases_pct_incre~ deaths_pct_incr~
##    <chr> <chr>  <chr> <chr>  <dbl>  <dbl>            <dbl>            <dbl>
##  1 01-24 Cook   Illi~ 17031      1      0              100               NA
##  2 01-25 Cook   Illi~ 17031      1      0                0              NaN
##  3 01-26 Cook   Illi~ 17031      1      0                0              NaN
##  4 01-27 Cook   Illi~ 17031      1      0                0              NaN
##  5 01-28 Cook   Illi~ 17031      1      0                0              NaN
##  6 01-29 Cook   Illi~ 17031      1      0                0              NaN
##  7 01-30 Cook   Illi~ 17031      2      0              100              NaN
##  8 01-31 Cook   Illi~ 17031      2      0                0              NaN
##  9 02-01 Cook   Illi~ 17031      2      0                0              NaN
## 10 02-02 Cook   Illi~ 17031      2      0                0              NaN
## # ... with 43 more rows, and 2 more variables: new_cases <dbl>,
## #   new_deaths <dbl>
```

```
filter(cook_cnty, is.na(new_deaths))
```

```
## # A tibble: 1 x 10
##   date  county state fips   cases deaths cases_pct_incre~ deaths_pct_incr~
##   <chr> <chr>  <chr> <chr>  <dbl>  <dbl>            <dbl>            <dbl>
## 1 01-24 Cook   Illi~ 17031      1      0              100               NA
## # ... with 2 more variables: new_cases <dbl>, new_deaths <dbl>
```

Similarly, the first row contains an `NA` value in the `new_deaths` column, and there are 53 rows with `NA` values in the `deaths_pct_increase` column. These `NA` values are the result of division by zero, so to clean the data, I replace them with 0.

```
cook_cnty <- cook_cnty %>%
  mutate(deaths_pct_increase =
           replace(deaths_pct_increase, which(
             is.na(deaths_pct_increase)), 0))

filter(cook_cnty, deaths_pct_increase == "Inf")
```

```
## # A tibble: 1 x 10
##   date  county state fips  cases deaths cases_pct_incre~ deaths_pct_incr~
##   <chr> <chr>  <chr> <chr> <dbl>  <dbl>            <dbl>            <dbl>
## 1 03-17 Cook   Illi~ 17031   107      1             40.8              Inf
## # ... with 2 more variables: new_cases <dbl>, new_deaths <dbl>
```

Finally, on the first day with an observed COVID death in Cook County, the `deaths_pct_increase` column populated with a value of infinity. While technically correct, I modify this value to 100%.

```
cook_cnty <- cook_cnty %>%
  mutate(deaths_pct_increase =
           replace(deaths_pct_increase, which(
             deaths_pct_increase == "Inf"), 100))

cook_cnty <- cook_cnty %>%
  mutate(new_cases_per_100 =
           new_cases/100)
```

Data Visualization

Now that the data is both cleaned and in the desired format, I create a simple data visualization of new COVID-19 cases and deaths by day from the first observed case in Cook County on January 24th to present. Each bar's height represents the number of new cases reported that day, while its color displays the number of new deaths reported. The dotted vertical line indicates March 21st, the date Illinois' state-wide shelter-in-place order went into effect. The increasing red line shows the total number of deaths to date.

```
ggplot(cook_cnty) +
  geom_bar(mapping = aes(x = date, y = new_cases,
                         fill = new_deaths), stat = "identity") +
   geom_line(mapping = aes(x = date, y = deaths),
             color = "red", group = 1) +
  geom_vline(xintercept = "03-21", linetype = 2, color = "red") +
  scale_x_discrete(breaks = c("01-24", "03-01", "04-01", "05-01")) +
  labs(title = "Daily COVID-19 Cases and Deaths in Cook County",
       x = "Date",
       y = "New Cases per Day",
       fill = "New Deaths per Day") +
  annotate(geom = "text", x = "02-28", y = 1500,
           label = "March 21 State-Wide\n Shelter in Place Order") +
  annotate(geom = "text", x = "04-27", y = 2300, label = "Total Deaths") +
  theme_minimal()
```

Daily COVID−19 Cases and Deaths in Cook County