

DTSA 5301 - COVID-19 Report

Ashish Bhutiani

6/20/2021

Question

Was there a bigger spike in cases in Miami/Dade county after the curfew was lifted versus other counties in Florida that didn't have the curfew?

Data Source and Summary

In order to attempt to answer this question, we will be using a dataset provided by Johns Hopkins University and is an aggregation of data from multiple sources.

The URL for the repository holding the dataset is: <https://github.com/CSSEGISandData/COVID-19>

```
base_url <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data"
file_names <- c("time_series_covid19_confirmed_US.csv", "time_series_covid19_deaths_US.csv")
urls <- str_c(base_url, file_names)
us_cases <- read.csv(urls[1])
us_deaths <- read.csv(urls[2])
```

Data Cleanup

The information isn't exactly in the format we want to work with it in. We will perform the following data cleanup operations:

- Pivot Cases and Deaths to create a vertical vs horizontal dataset
- Join cases and deaths
- Convert the date column to a date type
- Filter on only records where cases are more than 0 and the state is Florida.

```
pivoted_cases <- us_cases %>%
  pivot_longer(cols = starts_with("X"), names_to = "date",
    names_prefix = "X", values_to = "cases") %>%
  select(-c("Lat", "Long_"))
pivoted_deaths <- us_deaths %>%
  pivot_longer(cols = starts_with("X"), names_to = "date",
    names_prefix = "X", values_to = "deaths") %>%
  select(-c("Lat", "Long_", "UID", "iso2", "iso3", "code3",
    "FIPS"))
```

```
florida <- pivoted_cases %>%
  full_join(pivoted_deaths) %>%
  mutate(date = mdy(date)) %>%
  filter(cases > 0, Province_State == "Florida", Admin2 !=
         "Unassigned")

## Joining, by = c("Admin2", "Province_State", "Country_Region", "Combined_Key", "date")

summary(florida)
```

```
##      UID                iso2                iso3                code3
## Min.   :84012001  Length:30704      Length:30704      Min.   :840
## 1st Qu.:84012035  Class :character  Class :character  1st Qu.:840
## Median :84012069  Mode  :character  Mode  :character  Median :840
## Mean   :84012068                                     Mean   :840
## 3rd Qu.:84012101                                     3rd Qu.:840
## Max.   :84012133                                     Max.   :840
##      FIPS      Admin2      Province_State      Country_Region
## Min.   :12001  Length:30704      Length:30704      Length:30704
## 1st Qu.:12035  Class :character  Class :character  Class :character
## Median :12069  Mode  :character  Mode  :character  Mode  :character
## Mean   :12068
## 3rd Qu.:12101
## Max.   :12133
## Combined_Key      date                cases      Population
## Length:30704      Min.   :2020-03-02  Min.   : 1.0  Min.   : 8354
## Class :character  1st Qu.:2020-07-13  1st Qu.: 747.8  1st Qu.: 29210
## Mode  :character  Median :2020-11-04  Median : 2809.0  Median : 149657
##                                     Mean   :2020-11-04  Mean   : 15710.7  Mean   : 326348
##                                     3rd Qu.:2021-02-27  3rd Qu.: 11926.2  3rd Qu.: 375751
##                                     Max.   :2021-06-21  Max.   :506428.0  Max.   :2716940
##      deaths
## Min.   : 0.0
## 1st Qu.: 12.0
## Median : 52.0
## Mean   : 264.2
## 3rd Qu.: 270.0
## Max.   :6472.0
```

Analysis and Visualization

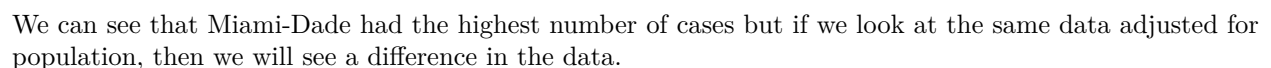
Now let's look at what values we have in our data:

```
head(florida)

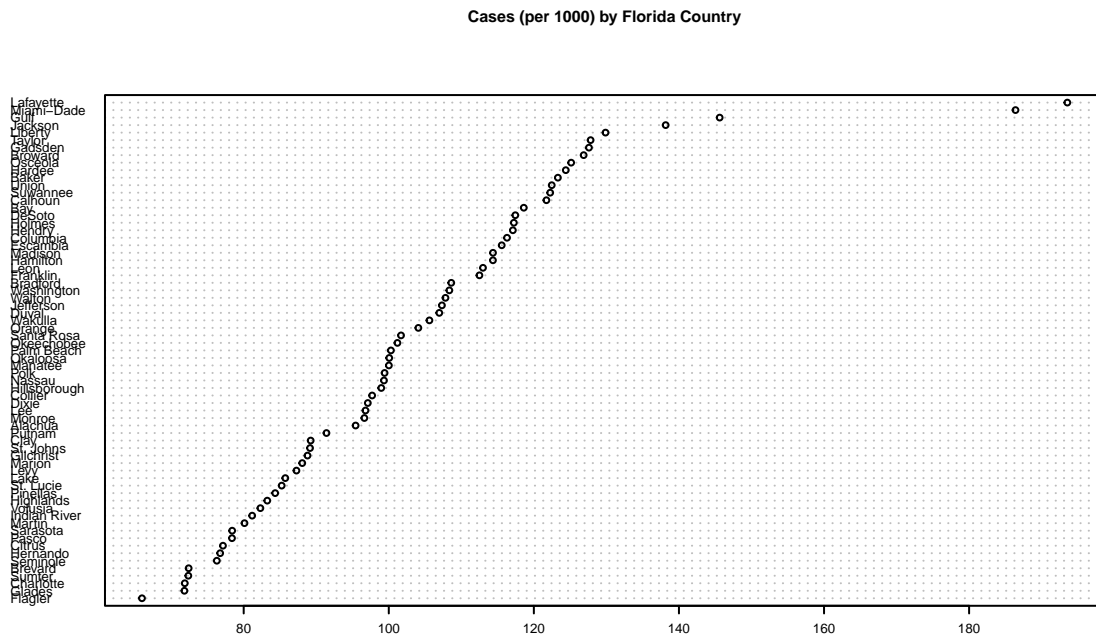
## # A tibble: 6 x 13
##      UID iso2 iso3 code3 FIPS Admin2 Province_State Country_Region
##      <int> <chr> <chr> <int> <dbl> <chr>      <chr>          <chr>
## 1 84012001 US   USA   840 12001 Alachua Florida      US
## 2 84012001 US   USA   840 12001 Alachua Florida      US
```

Let's summarize the data and see total cases/deaths by county:

Cases by Florida Country



```
ordered_county_counts <- county_counts[order(county_counts$cases_per_thousand),
]
dotchart(ordered_county_counts$cases_per_thousand, main = "Cases (per 1000) by Florida Country",
labels = ordered_county_counts$Admin2, cex = 0.4)
```



This shows that we would want to use cases per 1000 as a more representative marker to compare to the rest of the state.

Let's also build a linear model.

```
linearMod <- lm(deaths ~ cases, data = florida)

summary(linearMod)

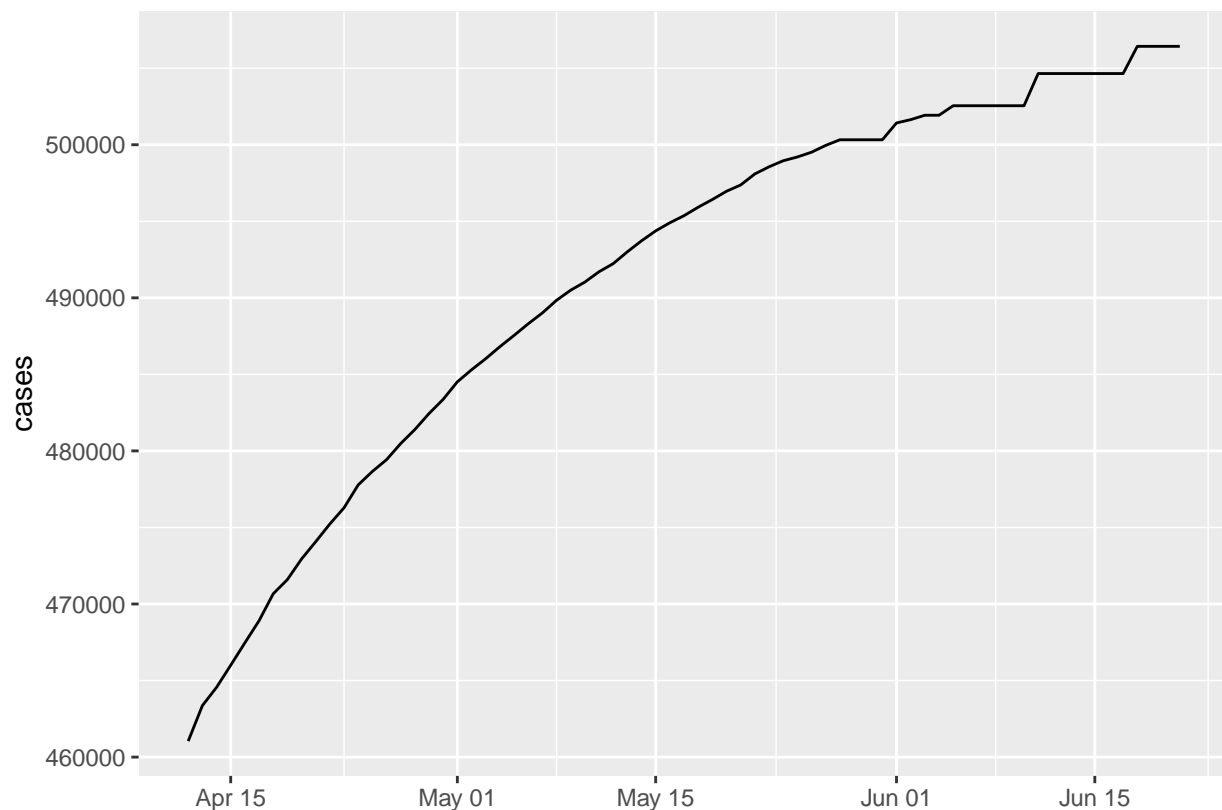
##
## Call:
## lm(formula = deaths ~ cases, data = florida)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -737.33  -48.06  -40.36   10.88 1055.61
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.755e+01  8.814e-01   53.95  <2e-16 ***
## cases        1.379e-02  1.918e-05   718.79  <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 145.1 on 30702 degrees of freedom
## Multiple R-squared:  0.9439, Adjusted R-squared:  0.9439
## F-statistic: 5.167e+05 on 1 and 30702 DF,  p-value: < 2.2e-16
```

The Miami Dade curfew ended on April 12, 2021 so we are interested in data after that point.

Let's just graph the total cases in Miami Dade after that date.

```
miami_data <- florida %>%
  filter(Admin2 == "Miami-Dade", date >= "2021-04-12")
ggplot(miami_data, aes(x = date, y = cases)) + geom_line() +
  xlab("")
```



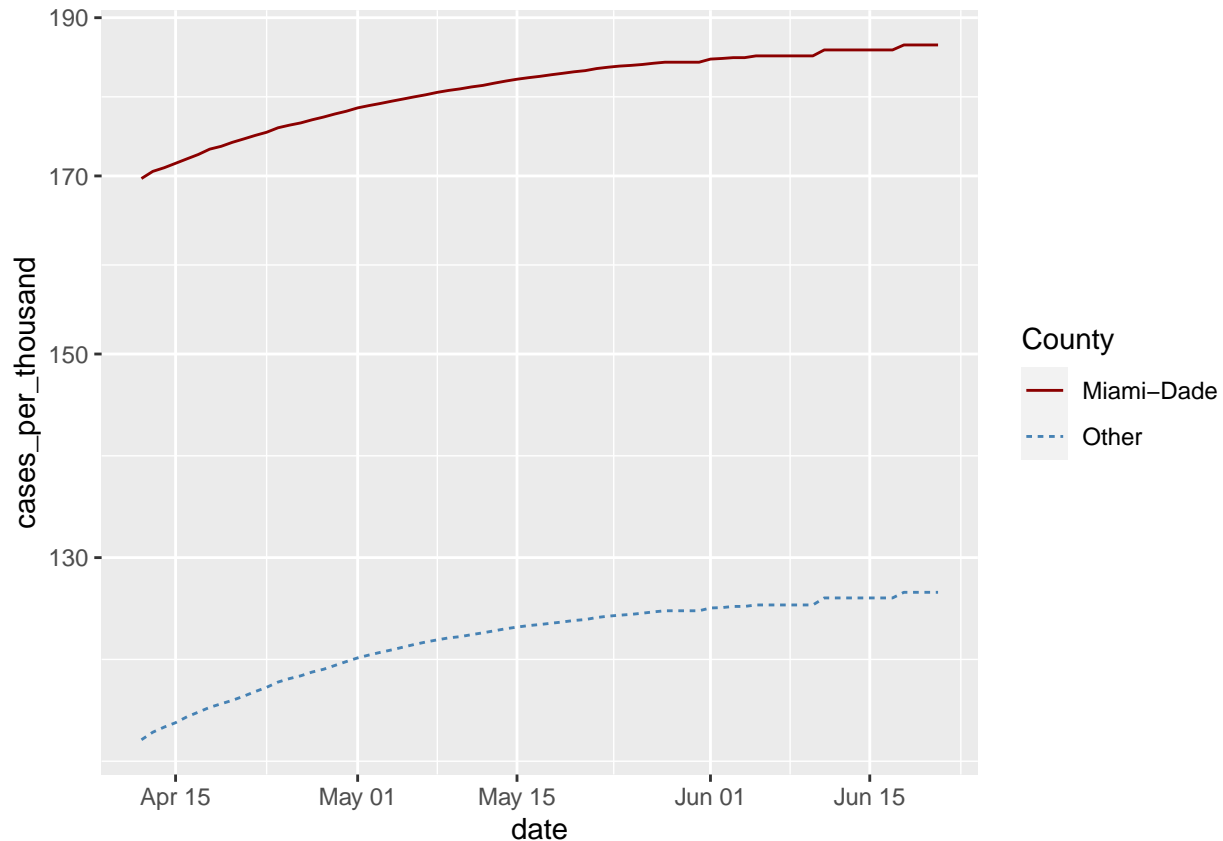
Now let's add the rest of Florida to the graph:

```
chart_data <- florida %>%
  filter(date >= "2021-04-12") %>%
  mutate(County = ifelse(Admin2 == "Miami-Dade", "Miami-Dade",
    "Other")) %>%
  group_by(County, date) %>%
  summarize(cases = max(cases), deaths = max(deaths), population = max(Population)) %>%
  mutate(deaths_per_thousand = deaths * 1000/population) %>%
  mutate(cases_per_thousand = cases * 1000/population) %>%
```

```
select(County, date, cases, deaths, cases_per_thousand, deaths_per_thousand,
       population)
```

'summarise()' has grouped output by 'County'. You can override using the '.groups' argument.

```
ggplot(chart_data, aes(x = date, y = cases_per_thousand)) + geom_line(aes(color = County,
  linetype = County)) + scale_color_manual(values = c("darkred",
  "steelblue")) + scale_y_log10()
```



Conclusion and Bias

When we look at the data graphed together, we don't see a huge difference in the curves, so we can answer the question with No.

In terms of Bias, I thought that we would see a jump going in, but I didn't let it impact where I was going with the data. I think further analysis on the rate of change in cases might be warranted to see if there are any other anomalies.

Session Info

```
sessionInfo()
```

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-apple-darwin20.4.0 (64-bit)
## Running under: macOS Big Sur 11.4
##
## Matrix products: default
## BLAS:   /usr/local/Cellar/openblas/0.3.15_1/lib/libopenblas-r0.3.15.dylib
## LAPACK: /usr/local/Cellar/r/4.1.0/lib/R/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] lubridate_1.7.10 forcats_0.5.1  stringr_1.4.0  dplyr_1.0.7
## [5] purrr_0.3.4      readr_1.4.0    tidyr_1.1.3    tibble_3.1.2
## [9] ggplot2_3.3.4    tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
## [1] tidymodels_1.1.1  xfun_0.24      haven_2.4.1    colorspace_2.0-1
## [5] vctrs_0.3.8       generics_0.1.0  htmltools_0.5.1.1 yaml_2.2.1
## [9] utf8_1.2.1        rlang_0.4.11   pillar_1.6.1    glue_1.4.2
## [13] withr_2.4.2       DBI_1.1.1      dbplyr_2.1.1    modelr_0.1.8
## [17] readxl_1.3.1      lifecycle_1.0.0 munsell_0.5.0    gtable_0.3.0
## [21] cellranger_1.1.0  rvest_1.0.0    evaluate_0.14    labeling_0.4.2
## [25] knitr_1.33        fansi_0.5.0    highr_0.9        broom_0.7.7
## [29] Rcpp_1.0.6        formatR_1.11    scales_1.1.1     backports_1.2.1
## [33] jsonlite_1.7.2    farver_2.1.0    fs_1.5.0         hms_1.1.0
## [37] digest_0.6.27     stringi_1.6.2   grid_4.1.0       cli_2.5.0
## [41] tools_4.1.0       magrittr_2.0.1  crayon_1.4.1     pkgconfig_2.0.3
## [45] ellipsis_0.3.2    xml2_1.3.2      reprex_2.0.0     assertthat_0.2.1
## [49] rmarkdown_2.9     httr_1.4.2      rstudioapi_0.13  R6_2.5.0
## [53] compiler_4.1.0
```