COVIDMINDER Bootcamp Notebook

Alex Chan

4/28/2020

Contents

Final Project: Guidance for Students (DELETE)

DELETE THIS SECTION IN FINAL NOTEBOOK!

GENERAL INSTRUCTIONS (DELETE)

- This notebook serves as a written presentation of your work so it should be written like a research paper.
- The rubric for evaluating this project is on LMS.
- The R code that executes the results should be embedded in this notebook if possible. It's also okay to "source" external scripts from within your notebook, including those provided by your instructors.
- If there is work that is not appropriate to be embedded on your final notebook, then you can describe the work in the notebook and provide figures generated elsewhere (e.g. screen shots, graphs). Indicate if that work has been committed to github. If necessary put details in Appendix.

- Your wording should be suitable for sharing with external partners/mentors and useful to future contributors.
- Please don't summarize everything you did; discuss only the *most important* aspects of your work. Ask yourself *what really matters?*
- IMPORTANT: Discuss any insights you found regarding COVID-19.

- If there are limitations to your work, discuss.
- Include any background or support for your work. For example, mention any relevant research articles you found – and be sure to include references!
- Include your recommendations for extending the COVIDMINDER app:
  - Data utilization. . .
  - Analytics. . .
  - Visualizations
  - User interface design. . .
- **Expect internet problems!** Commit versions of this notebook to the IDEA github *early and often*. You can continue to refine and update; that is the point of github!

SUBMISSION CHECKLIST: Things to check before you submit (DELETE)

- Have you given your file(s) a clear, sensible name? `my_notebook.Rmd` or `joey.Rmd` are **not* acceptable!
- Is every figure clearly labeled and titled?
- Does every figure serve a purpose?
  - What question does it answer?
  - You can put extra (non-essential) figues in the Appendix.
- Are your tables clearly labelled and legible?
  - We recommend using `kable()` (built into `knitr`)
  - `xtable()` might also work for you (not as easy as `kable()`)
- **CRITICAL:** Have you given enough information for someone to reproduce, understand and extend your results?
  - Where can they *find* the data that you used?
  - Have you *described* the data that used?
  - Have you *documented* your code?
  - Are your figures *clearly labeled*?
  - Did you *discuss your findings*?
  - Did you use good grammar and *proofread* your results?
  - Finally, have you *committed* your work to github and made a *pull request*?

github Submission Instructions (DELETE!)

Fill out the bullets at the beginning of this notebook. . .

- Provide your github ID.
- Indicate the issues associated with this work on github.
- Commit the Rmd and html versions of this notebook to the `MATP-4400-FINAL` directory on github and make a pull request.

- Provide the link to your github submission on LMS.
  - **NOTE:** You can post your link any time after it exists

  - We will grade the latest version available on Github starting at 3:00 on Thursday, 7 May.

- **NOTE:** RStudio Server will be unavailable due to maintenance for up to a week beginning the weekend of 9 May, so there can be no extensions or late submissions.

github help (DELETE!)

- DO THIS FIRST:
  - `git clone https://github.com/TheRensselaerIDEA/COVID-Notebooks` (if you haven't already)
    * NOTE: We may have added to the master since your last `git pull` so make sure to do a `git pull origin master`

Final Project: Submission Links

This should be the first section of your final project notebook. Fill out the following according to how you submitted your notebook!

- github repository: https://github.com/TheRensselaerIDEA/COVID-Notebooks (for reference)
- My github ID: *alexchan18* (example)
- github issues addressed by this work: `#27` (example)
- Github branch name of my submitted notebook: *erickson_final* (example)
- link to merged notebook (post these to LMS!:
  - https://github.com/TheRensselaerIDEA/COVID-Notebooks/blob/feature-27/MATP-4400-FINAL/covid-notebook-Chana6.Rmd
  - https://github.com/TheRensselaerIDEA/COVID-Notebooks/blob/master/MATP-4400-FINAL/COVID_FINAL_2020.html (example; HTML version)

Overview & Problems Tackled

On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic. The number of cases reached a point where the entire world needed to take action. In the United States, state governors began closing down states to reduce the rate of infection. This meant that all citizens had to remain in their homes until further notice. Fast forward 6 weeks, the citizens of the United States are eager to leave their homes and return to normal life. The decision to do open up the country will be made based on the rate of infection and the increase in new cases. I looked at the data for this in New York State to see if Governor Cuomo should make the decision to open.

Data Description

Include data sources/locations, versions/dates, etc.

Results

Break out your results by each problem you attacked

Problem 1

Describe the problem you are examining. If there is background that is necessary for this problem, then put it here. Include any references.

Methods

```
#April22 <- read_csv("data/csv/04-22-2020.csv")
#Ap22 <- April22[April22$Province_State == "New York",]
#A22 <- data.frame(Ap22$Admin2, Ap22$Confirmed, Ap22$Deaths)
#A22 <- A22[rowSums(is.na(Ap22)) == 0,] #April 22nd NY Counties Confirmed and Deaths
#A22Deaths = sum(A22$Ap22.Deaths)

Deaths = 19413

April23 <- read_csv("data/csv/04-23-2020.csv")
```

```
## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
```

```
Ap23 <- April23[April23$Province_State == "New York",]
A23 <- data.frame(Ap23$Admin2, Ap23$Confirmed, Ap23$Deaths)
A23 <- A23[rowSums(is.na(Ap23)) == 0,] #April 23rd NY Counties Confirmed and Deaths
A23Deaths = sum(A23$Ap23.Deaths) - Deaths

April26 <- read_csv("data/csv/04-26-2020.csv")
```

```
## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
```

```
Ap26 <- April26[April26$Province_State == "New York",]
A26 <- data.frame(Ap26$Admin2, Ap26$Confirmed, Ap26$Deaths)
A26 <- A26[rowSums(is.na(A26)) == 0,] #April 26th NY Counties Confirmed and Deaths
A26Deaths = sum(A26$Ap26.Deaths) - Deaths
```

```
April29 <- read_csv("data/csv/04-29-2020.csv")

## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
Ap29 <- April29[April29$Province_State == "New York",]
A29 <- data.frame(Ap29$Admin2, Ap29$Confirmed, Ap29$Deaths)
A29 <- A29[rowSums(is.na(A29)) == 0,] #April 29th NY Counties Confirmed and Deaths
A29Deaths = sum(A29$Ap29.Deaths) - Deaths

May2 <- read_csv("data/csv/05-02-2020.csv")

## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
Ma2 <- May2[May2$Province_State == "New York",]
M2 <- data.frame(Ma2$Admin2, Ma2$Confirmed, Ma2$Deaths)
M2 <-M2[rowSums(is.na(M2)) == 0,] #May 2nd NY Counties Confirmed and Deaths
M2Deaths = sum(M2$Ma2.Deaths) - Deaths

May5 <- read_csv("data/csv/05-05-2020.csv")

## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
```

```
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
```

```r
Ma5 <- May5[May5$Province_State == "New York",]
M5 <- data.frame(Ma5$Admin2, Ma5$Confirmed, Ma5$Deaths)
M5 <- M5[rowSums(is.na(M5)) == 0,] #May 5th NY Counties Confirmed and Deaths
M5Deaths = sum(M5$Ma5.Deaths) - Deaths

#Put the number of new deaths for each day into a csv file
DaysDeaths <- read_csv("data/csv/DaysDeaths.csv")
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   Deaths = col_double()
## )
```

```r
barplot(DaysDeaths$Deaths, main="Deaths Per Day Every 3 Days Since April 23rd",
    xlab="Date",
    ylab="# of deaths",
    names.arg=c(DaysDeaths$Date))
```

Deaths Per Day Every 3 Days Since April 23rd
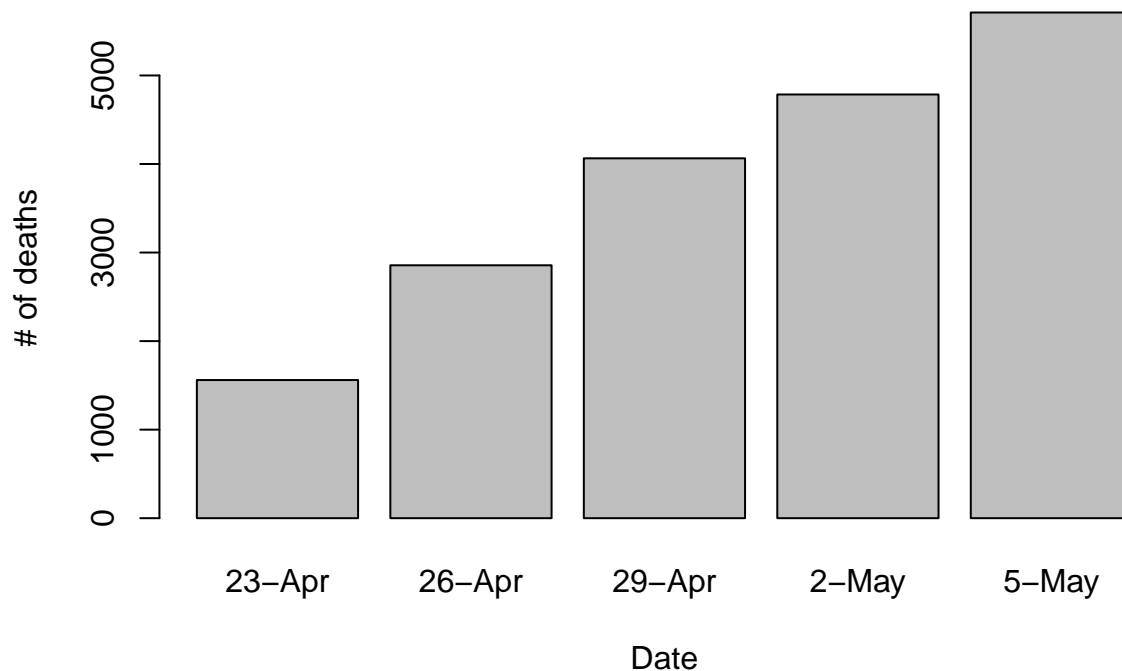
Results

What were the results on this problem?

Discussion

Interpret results. What were your findings? What do they say about COVID-19? What are the strengths and limitations of these results? Is there support for your findings from other sources? Include references as appropriate.

Problem 2

Methods

Results

Discussion

Do at least 2 problems. Add more as necessary # Summary and COVIDMINDER Recommendations *
Overall, what insights did you find about the COVID-19 epidemic in your analysis?
* What recommendations do you have for COVIDMINDER for Data utilization, Analytics, Visualizations, User interface design, etc. + Would you recommend that your analysis be included in COVIDMINDER? Why or Why not?
+ If not, is there additional work that might improve the results? Note that it is perfectly okay for you to recommend to not to include your analysis. Research doesn't always work out. As a team we need to try a lot of different ideas in hopes of finding a few good ones to pursue. Think of yourself as a consultant reporting back on a particular aspect of the analysis and application design!*

References

Include any references to literature in support of your work You might also wish to include references to unusual R packages essential to your work

Appendix

Include here whatever you think is relevant to support the main content of your notebook. For example, you may have only include example figures above in your main text but include additional ones here