Auto Process Monroe County Covid-19 Case Data

John Minter

Started 2020-03-30, Last modified: 2020-04-07

Table of Contents

Set up the analysis. Use the bVerbose variable as a flag to decide whether or not to print diagnostic data. The default is FALSE.

bVerbose <- FALSE  
  
if(bVerbose){  
 print(getwd())  
}

temp <- tempfile()  
download.file("https://github.com/nytimes/covid-19-data/archive/master.zip",temp)  
us\_counties <- read\_csv(unz(temp, "covid-19-data-master/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()  
)

path <- paste0(here(), "/us-counties.csv")  
write\_csv(us\_counties, path, append = FALSE, col\_names=TRUE)  
us\_states <- read\_csv(unz(temp, "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()  
)

pa\_states <- paste0(here(), "/us-states.csv")  
write\_csv(us\_states, pa\_states, append = FALSE, col\_names=TRUE)  
unlink(temp)  
tail(us\_counties)

# A tibble: 6 x 6  
 date county state fips cases deaths  
 <date> <chr> <chr> <chr> <dbl> <dbl>  
1 2020-04-07 Sheridan Wyoming 56033 12 0  
2 2020-04-07 Sublette Wyoming 56035 1 0  
3 2020-04-07 Sweetwater Wyoming 56037 6 0  
4 2020-04-07 Teton Wyoming 56039 44 0  
5 2020-04-07 Uinta Wyoming 56041 3 0  
6 2020-04-07 Washakie Wyoming 56043 4 0

tail(us\_states)

# A tibble: 6 x 5  
 date state fips cases deaths  
 <date> <chr> <chr> <dbl> <dbl>  
1 2020-04-07 Virgin Islands 78 45 1  
2 2020-04-07 Virginia 51 3333 69  
3 2020-04-07 Washington 53 8682 409  
4 2020-04-07 West Virginia 54 412 4  
5 2020-04-07 Wisconsin 55 2578 94  
6 2020-04-07 Wyoming 56 221 0

Next, load the data file and extract what we need using functions from the dplyr package to create a tibble of values (an enhanced R dataframe that works nicely with the tidyverse collection of R packages by Hadley Wickham.) We will use the kable function from the knitr package to get a nice looking table. We really only want the last few values…

pa\_cty <- paste0(here(), "/us-counties.csv")  
df <- read.csv(pa\_cty, header = TRUE, sep = ",")  
df$date <- format(as.Date(df$date), "%m-%d")  
tib <- as\_tibble(df)  
tib$date <- as.Date(df$date, "%m-%d")  
ny <- tib %>% filter(state == "New York")  
monroe\_cty <- ny %>% filter(county == "Monroe")  
monroe\_cty %>% select(date, cases, deaths) -> mc\_vals  
tail(mc\_vals)

# A tibble: 6 x 3  
 date cases deaths  
 <date> <int> <int>  
1 2020-04-02 420 10  
2 2020-04-03 464 14  
3 2020-04-04 512 17  
4 2020-04-05 548 19  
5 2020-04-06 574 23  
6 2020-04-07 596 31

tail(mc\_vals)

# A tibble: 6 x 3  
 date cases deaths  
 <date> <int> <int>  
1 2020-04-02 420 10  
2 2020-04-03 464 14  
3 2020-04-04 512 17  
4 2020-04-05 548 19  
5 2020-04-06 574 23  
6 2020-04-07 596 31

print(mc\_vals)

# A tibble: 28 x 3  
 date cases deaths  
 <date> <int> <int>  
 1 2020-03-11 1 0  
 2 2020-03-12 1 0  
 3 2020-03-13 1 0  
 4 2020-03-14 2 0  
 5 2020-03-15 2 0  
 6 2020-03-16 10 0  
 7 2020-03-17 11 0  
 8 2020-03-18 14 0  
 9 2020-03-19 28 0  
10 2020-03-20 32 0  
# … with 18 more rows

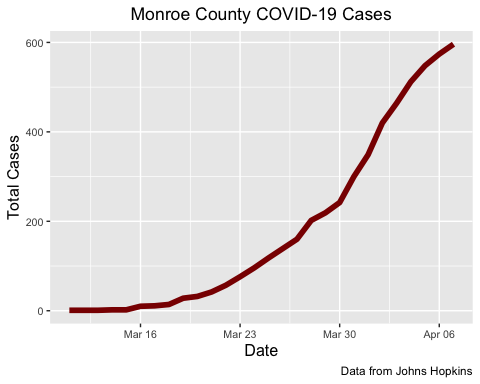
Next, we compute the total deaths for Monroe County

tot\_deaths <- sum(monroe\_cty$deaths)  
tot\_cases <- sum(monroe\_cty$cases)  
  
death\_rate\_pct <- 100\*tot\_deaths/tot\_cases  
  
death\_rate\_pct <- round(death\_rate\_pct, digits = 2)

Monroe County’s COVID-19 has 5216 cases and 156 deaths with a death rate of 2.99 percent.

Next, we plot the curve.

mc\_tot\_plt <- ggplot(mc\_vals, aes(x=date, y=cases)) +  
 geom\_line(colour='darkred', size=2) +  
 xlab("Date") +  
 ylab("Total Cases") +   
 ggtitle("Monroe County COVID-19 Cases") +  
 labs(caption = 'Data from Johns Hopkins') +  
 scale\_x\_date() +  
 theme(axis.text=element\_text(size=8),  
 axis.title=element\_text(size=12),  
 plot.title=element\_text(hjust = 0.5)) +  
 NULL  
  
print(mc\_tot\_plt)



It looks like Monroe County’s curve is starting to flatten.

We will save the plot as in both the png and jpg formats.

strOutPng <- "plt/Monroe\_Cty\_Covid-19.png"  
  
# we want the png to be close to 1024x768...  
ggsave(mc\_tot\_plt, file=strOutPng, width=9.0, height=6.0,  
 units="in", dpi=113.7778)  
  
strOutJpg <- "plt/Monroe\_Cty\_Covid-19.jpg"  
  
# we want the png to be close to 1024x768...  
ggsave(mc\_tot\_plt, file=strOutJpg, width=9.0, height=6.0,  
 units="in", dpi=113.7778)