

# Franklin County PA Covid

Matthew Angle

## Current Data

The data within this model is limited. There exists no easy package in R for PA Coronavirus cases by county. I've entered in this data manually.

```
knitr::opts_chunk$set(error = TRUE)
#load libs
library("tidyverse")
library("ggplot2")
library("httr")
library("rvest")
```

```
##Scraping PA Tables Making DF's
```

```
daySinceFirstCase <- 40
```

```
franklinCountyCorona <- data.frame("day" = c(seq(1,daySinceFirstCase)), "dates" = seq(as.Date("2020-03-20"), as.Date("2020-04-12"), by="day"))
```

```
franklinCountyCorona
```

##	day	dates	cases
## 1	1	2020-03-20	1
## 2	2	2020-03-21	1
## 3	3	2020-03-22	1
## 4	4	2020-03-23	1
## 5	5	2020-03-24	3
## 6	6	2020-03-25	5
## 7	7	2020-03-26	5
## 8	8	2020-03-27	5
## 9	9	2020-03-28	7
## 10	10	2020-03-29	11
## 11	11	2020-03-30	12
## 12	12	2020-03-31	19
## 13	13	2020-04-01	21
## 14	14	2020-04-02	23
## 15	15	2020-04-03	26
## 16	16	2020-04-04	30
## 17	17	2020-04-05	32
## 18	18	2020-04-06	39
## 19	19	2020-04-07	43
## 20	20	2020-04-08	52
## 21	21	2020-04-09	57
## 22	22	2020-04-10	64
## 23	23	2020-04-11	69
## 24	24	2020-04-12	78

```
## 25 25 2020-04-13 80
## 26 26 2020-04-14 89
## 27 27 2020-04-15 106
## 28 28 2020-04-16 111
## 29 29 2020-04-17 115
## 30 30 2020-04-18 143
## 31 31 2020-04-19 152
## 32 32 2020-04-20 151
## 33 33 2020-04-21 164
## 34 34 2020-04-22 181
## 35 35 2020-04-23 205
## 36 36 2020-04-24 227
## 37 37 2020-04-25 237
## 38 38 2020-04-26 264
## 39 39 2020-04-27 284
## 40 40 2020-04-28 313
```

```
url <- 'https://www.health.pa.gov/topics/disease/coronavirus/Pages/Archives.aspx'
ws <- GET(url)
```

```
tbls <- html_nodes(content(ws), "table")
```

##Fit

The fit model

```
fit <- lm(formula = log(cases) ~ day , data = franklinCountyCorona)
```

```
summary(fit)
```

```
##
## Call:
## lm(formula = log(cases) ~ day, data = franklinCountyCorona)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19516 -0.26679  0.08627  0.40875  0.61176
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.626403   0.152803   4.099  0.00021 ***
## day          0.142190   0.006495  21.893 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4742 on 38 degrees of freedom
## Multiple R-squared:  0.9265, Adjusted R-squared:  0.9246
## F-statistic: 479.3 on 1 and 38 DF,  p-value: < 2.2e-16
```

##Using the model

Using the model to generate data for an additional amount of time. Placed in final model

```
newDay <-data.frame("day" = c(seq(1, 40)))
nextTwentyDays <-predict(fit, newDay)
nextTwentyDays <- as.data.frame(nextTwentyDays)
```

```

tmp <- seq(as.Date("2020-03-20"), by = "days", length.out = 40)
names(nextTwentyDays)[1] <- "cases"
#has a null value assume model starts at 1
#nextTwentyDays[1,1] <- 1
nextTwentyDays <- mutate(nextTwentyDays,
  "day" = c(seq(1, 40)),
  "cases" = ceiling(exp(nextTwentyDays$cases)),
  "dates" = tmp)
finalModel<- merge(nextTwentyDays, franklinCountyCorona, by = "dates", all = TRUE)
#nextTwentyDays <- mutate(nextTwentyDays,
#
  "day" = c(seq(1, 30)),
#
  "cases" = ceiling(nextTwentyDays$cases),
#
  "dates" = tmp)
#finalModel<- merge(nextTwentyDays, franklinCountyCorona, by = "dates", all = TRUE)

(finalModel)

```

##		dates	cases.x	day.x	day.y	cases.y
## 1	2020-03-20	3	1	1	1	
## 2	2020-03-21	3	2	2	1	
## 3	2020-03-22	3	3	3	1	
## 4	2020-03-23	4	4	4	1	
## 5	2020-03-24	4	5	5	3	
## 6	2020-03-25	5	6	6	5	
## 7	2020-03-26	6	7	7	5	
## 8	2020-03-27	6	8	8	5	
## 9	2020-03-28	7	9	9	7	
## 10	2020-03-29	8	10	10	11	
## 11	2020-03-30	9	11	11	12	
## 12	2020-03-31	11	12	12	19	
## 13	2020-04-01	12	13	13	21	
## 14	2020-04-02	14	14	14	23	
## 15	2020-04-03	16	15	15	26	
## 16	2020-04-04	19	16	16	30	
## 17	2020-04-05	21	17	17	32	
## 18	2020-04-06	25	18	18	39	
## 19	2020-04-07	28	19	19	43	
## 20	2020-04-08	33	20	20	52	
## 21	2020-04-09	38	21	21	57	
## 22	2020-04-10	43	22	22	64	
## 23	2020-04-11	50	23	23	69	
## 24	2020-04-12	57	24	24	78	
## 25	2020-04-13	66	25	25	80	
## 26	2020-04-14	76	26	26	89	
## 27	2020-04-15	87	27	27	106	
## 28	2020-04-16	101	28	28	111	
## 29	2020-04-17	116	29	29	115	
## 30	2020-04-18	134	30	30	143	
## 31	2020-04-19	154	31	31	152	
## 32	2020-04-20	178	32	32	151	
## 33	2020-04-21	205	33	33	164	
## 34	2020-04-22	236	34	34	181	
## 35	2020-04-23	272	35	35	205	
## 36	2020-04-24	313	36	36	227	

```
## 37 2020-04-25    361    37    37    237
## 38 2020-04-26    416    38    38    264
## 39 2020-04-27    480    39    39    284
## 40 2020-04-28    553    40    40    313
```

##Plot the data

Used the data from the model to plot

```
ggplot(finalModel, aes(x = dates)) +
  geom_point(aes(y = cases.y), color = "red") +
  geom_path(aes(y = cases.y), color = "black") +
  labs(x = "Dates", y = "Cases") +
  ggtitle("Franklin County PA Confirmed Covid19 Cases Model Days") +
  theme_bw()
```

