# 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 <- 24

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

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
```

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

tbls <- html_nodes(content(ws), "table")
print((html_table(tbls[[4]])))
```

```
##                X1              X2     X3
## 1          County Number of Cases Deaths
## 2           Adams              63      1
## 3       Allegheny             893     24
## 4       Armstrong              28      1
## 5          Beaver             156     14
## 6         Bedford               5      1
## 7           Berks            1247     27
## 8           Blair              11
## 9        Bradford              19
## 10          Bucks            1222     40
## 11         Butler             143      5
## 12        Cambria              14      1
## 13        Cameron               1
## 14         Carbon             103      3
## 15         Centre              70
## 16        Chester             621     20
## 17        Clarion              16
## 18     Clearfield               9
## 19        Clinton               8
## 20       Columbia             125      3
## 21       Crawford              16
## 22     Cumberland             124      4
## 23        Dauphin             249      5
## 24       Delaware            1806     45
## 25            Elk               2
## 26           Erie              41
## 27        Fayette              58      3
## 28         Forest               5
## 29       Franklin              69
## 30         Fulton               2
## 31         Greene              23
## 32     Huntingdon              11
## 33        Indiana              43
## 34      Jefferson               2
## 35        Juniata              43
## 36     Lackawanna             501     24
## 37      Lancaster             865     26
## 38       Lawrence              51      4
## 39        Lebanon             328      2
## 40         Lehigh            1803     23
## 41        Luzerne            1523     26
## 42       Lycoming              29
## 43         McKean               4
## 44         Mercer              44
## 45        Mifflin              16
## 46         Monroe             847     27
## 47     Montgomery            2354     76
```

```
## 48        Montour              44
## 49    Northampton            1176     25
## 50 Northumberland              48
## 51          Perry              17      1
## 52   Philadelphia            7121    131
## 53           Pike             256      6
## 54         Potter               4
## 55     Schuylkill             200      2
## 56         Snyder              24      1
## 57       Somerset              13
## 58       Sullivan               1
## 59    Susquehanna              32      1
## 60          Tioga              13      1
## 61          Union              23
## 62        Venango               6
## 63         Warren               1
## 64     Washington              70      1
## 65          Wayne              70      1
## 66   Westmoreland             231      6
## 67        Wyoming              11
## 68           York             371      3
```

##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
## -0.79440 -0.20808  0.04985  0.25529  0.53795
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.01164    0.15078  -0.077    0.939
## day          0.20151    0.01055  19.096  3.5e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3579 on 22 degrees of freedom
## Multiple R-squared:  0.9431, Adjusted R-squared:  0.9405
## F-statistic: 364.7 on 1 and 22 DF,  p-value: 3.495e-15
```

##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, 30)))
nextTwentyDays <-predict(fit, newDay)
nextTwentyDays <- as.data.frame(nextTwentyDays)
```

```
tmp <- seq(as.Date("2020-03-20"), by = "days", length.out = 30)
names(nextTwentyDays)[1] <- "cases"
#has a null value assume model starts at 1
#nextTwentyDays[1,1] <- 1
nextTwentyDays <- mutate(nextTwentyDays,
                         "day"   = c(seq(1, 30)),
                         "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       2     1     1       1
## 2  2020-03-21       2     2     2       1
## 3  2020-03-22       2     3     3       1
## 4  2020-03-23       3     4     4       1
## 5  2020-03-24       3     5     5       3
## 6  2020-03-25       4     6     6       5
## 7  2020-03-26       5     7     7       5
## 8  2020-03-27       5     8     8       5
## 9  2020-03-28       7     9     9       7
## 10 2020-03-29       8    10    10      11
## 11 2020-03-30      10    11    11      12
## 12 2020-03-31      12    12    12      19
## 13 2020-04-01      14    13    13      21
## 14 2020-04-02      17    14    14      23
## 15 2020-04-03      21    15    15      26
## 16 2020-04-04      25    16    16      30
## 17 2020-04-05      31    17    17      32
## 18 2020-04-06      38    18    18      39
## 19 2020-04-07      46    19    19      43
## 20 2020-04-08      56    20    20      52
## 21 2020-04-09      69    21    21      57
## 22 2020-04-10      84    22    22      64
## 23 2020-04-11     102    23    23      69
## 24 2020-04-12     125    24    24      78
## 25 2020-04-13     153    25    NA      NA
## 26 2020-04-14     187    26    NA      NA
## 27 2020-04-15     228    27    NA      NA
## 28 2020-04-16     279    28    NA      NA
## 29 2020-04-17     342    29    NA      NA
## 30 2020-04-18     418    30    NA      NA
```

##Plot the data

Used the data from the model to plot

```
ggplot(finalModel, aes(x = dates)) +
  geom_point(aes(y = cases.x), color = "darkgrey") +
```

```
  geom_point(aes(y = cases.y), color = "red") +
  geom_path(aes(y = cases.x), color = "grey") +
  geom_path(aes(y = cases.y), color = "black") +
  labs(x = "Dates", y = "Cases") +
  ggtitle("Franklin County PA Confirmed Covid19 Cases Model 30 Days") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

```
## Warning: Removed 6 row(s) containing missing values (geom_path).
```