

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 <- 22
```

```
franklinCountyCorona <- data.frame("day" = c(seq(1,daySinceFirstCase)), "dates" = seq(as.Date("2020-03-20"), as.Date("2020-04-10"), 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

```
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	44	1
## 3	Allegheny	836	19
## 4	Armstrong	26	1
## 5	Beaver	143	13
## 6	Bedford	5	1
## 7	Berks	930	19
## 8	Blair	10	
## 9	Bradford	18	
## 10	Bucks	1051	29
## 11	Butler	128	3
## 12	Cambria	13	1
## 13	Cameron	1	
## 14	Carbon	98	3
## 15	Centre	69	
## 16	Chester	532	15
## 17	Clarion	15	
## 18	Clearfield	9	
## 19	Clinton	7	
## 20	Columbia	99	2
## 21	Crawford	15	
## 22	Cumberland	105	3
## 23	Dauphin	213	3
## 24	Delaware	1510	39
## 25	Elk	2	
## 26	Erie	39	
## 27	Fayette	50	3
## 28	Forest	5	
## 29	Franklin	59	
## 30	Fulton	1	
## 31	Greene	23	
## 32	Huntingdon	10	
## 33	Indiana	40	
## 34	Jefferson	1	
## 35	Juniata	38	
## 36	Lackawanna	392	20
## 37	Lancaster	698	23
## 38	Lawrence	46	4
## 39	Lebanon	232	1
## 40	Lehigh	1620	16
## 41	Luzerne	1372	17
## 42	Lycoming	20	
## 43	McKean	2	
## 44	Mercer	38	
## 45	Mifflin	10	
## 46	Monroe	774	22
## 47	Montgomery	2053	60
## 48	Montour	29	
## 49	Northampton	1039	23

```
## 50 Northumberland      31
## 51      Perry          16      1
## 52 Philadelphia      6022     130
## 53      Pike          208      6
## 54      Potter         4
## 55      Schuylkill     179      2
## 56      Snyder        16      1
## 57      Somerset      10
## 58      Sullivan       1
## 59      Susquehanna    23      2
## 60      Tioga          12      1
## 61      Union          14
## 62      Venango         6
## 63      Warren         1
## 64      Washington     66
## 65      Wayne          57      1
## 66      Westmoreland   218      6
## 67      Wyoming        8
## 68      York           293      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
## -0.74592 -0.19959  0.05405  0.24111  0.49317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.10675    0.14946  -0.714    0.483
## day          0.21317    0.01138  18.733 3.75e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3386 on 20 degrees of freedom
## Multiple R-squared:  0.9461, Adjusted R-squared:  0.9434
## F-statistic: 350.9 on 1 and 20 DF,  p-value: 3.754e-14
```

##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      4     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     15    13    13     21
## 14 2020-04-02     18    14    14     23
## 15 2020-04-03     22    15    15     26
## 16 2020-04-04     28    16    16     30
## 17 2020-04-05     34    17    17     32
## 18 2020-04-06     42    18    18     39
## 19 2020-04-07     52    19    19     43
## 20 2020-04-08     64    20    20     52
## 21 2020-04-09     80    21    21     57
## 22 2020-04-10     98    22    22     64
## 23 2020-04-11    122    23    NA     NA
## 24 2020-04-12    150    24    NA     NA
## 25 2020-04-13    186    25    NA     NA
## 26 2020-04-14    230    26    NA     NA
## 27 2020-04-15    284    27    NA     NA
## 28 2020-04-16    352    28    NA     NA
## 29 2020-04-17    435    29    NA     NA
## 30 2020-04-18    539    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 8 rows containing missing values (geom_point).

Warning: Removed 8 row(s) containing missing values (geom_path).

Franklin County PA Confirmed Covid19 Cases Model 30 Days

