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")

franklinCountyCorona <- data.frame("day" = c(seq(1,14)), "dates" = seq(as.Date("2020-03-20"), by = "days
```

##Scraping PA Tables

##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
## -0.59632 -0.07975 -0.03420  0.17465  0.46081
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5083     0.1558  -3.263  0.00679 **
## day           0.2762     0.0183  15.094 3.62e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.276 on 12 degrees of freedom
## Multiple R-squared:   0.95,  Adjusted R-squared:  0.9458
## F-statistic: 227.8 on 1 and 12 DF,  p-value: 3.62e-09
```

##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)

(finalModel)
```

```
##         dates cases.x day.x day.y cases.y
## 1  2020-03-20       1     1     1       1
## 2  2020-03-21       2     2     2       1
## 3  2020-03-22       2     3     3       1
## 4  2020-03-23       2     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       6     8     8       5
## 9  2020-03-28       8     9     9       7
## 10 2020-03-29      10    10    10      11
## 11 2020-03-30      13    11    11      12
## 12 2020-03-31      17    12    12      19
## 13 2020-04-01      22    13    13      21
## 14 2020-04-02      29    14    14      23
## 15 2020-04-03      38    15    NA      NA
## 16 2020-04-04      50    16    NA      NA
## 17 2020-04-05      66    17    NA      NA
## 18 2020-04-06      87    18    NA      NA
## 19 2020-04-07     115    19    NA      NA
## 20 2020-04-08     151    20    NA      NA
## 21 2020-04-09     199    21    NA      NA
## 22 2020-04-10     262    22    NA      NA
## 23 2020-04-11     345    23    NA      NA
## 24 2020-04-12     455    24    NA      NA
## 25 2020-04-13     600    25    NA      NA
## 26 2020-04-14     790    26    NA      NA
## 27 2020-04-15    1041    27    NA      NA
## 28 2020-04-16    1373    28    NA      NA
## 29 2020-04-17    1809    29    NA      NA
## 30 2020-04-18    2384    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 16 rows containing missing values (geom_point).
```

```
## Warning: Removed 16 row(s) containing missing values (geom_path).
```