Boston organized

Zhengkai Fu

14/05/2020

```
library(tidyverse)
library(tsbox) # transform data into time series
library(xts)
library(COVID19) # to get data about covid 19
library(aTSA) # adf.test
```

This document is to study the possible relationship between COVID-19 and frequency of crime committed in the city of Boston.

COVID 19

Load COVID 19 data for Massachusetts

```
# extract MA data from US data. level 3 is by cities but cannot find Boston.
covid19_MA <- covid19("USA", level = 2) %>%
  filter(state == "Massachusetts") %>%
  # filter out days when confirmed is zero
  filter(confirmed > 0)

# brief display
head(covid19_MA)
```

```
## # A tibble: 6 x 24
## # Groups:   id [1]
##   id    date       deaths confirmed tests recovered  hosp   icu  vent driving
##   <chr> <date>      <dbl>     <dbl> <dbl>     <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1 USA,~ 2020-02-01      0         1     0         0     0     0     0       0
## 2 USA,~ 2020-02-02      0         1     0         0     0     0     0       0
## 3 USA,~ 2020-02-03      0         1     0         0     0     0     0       0
## 4 USA,~ 2020-02-04      0         1     0         0     0     0     0       0
## 5 USA,~ 2020-02-05      0         1     0         0     0     0     0       0
## 6 USA,~ 2020-02-06      0         1     0         0     0     0     0       0
## # ... with 14 more variables: walking <dbl>, transit <dbl>, country <chr>,
## #   state <chr>, city <chr>, lat <dbl>, lng <dbl>, pop <int>, pop_14 <lgl>,
## #   pop_15_64 <lgl>, pop_65 <lgl>, pop_age <lgl>, pop_density <lgl>,
## #   pop_death_rate <lgl>
```

Choice explained

Scope of data

The covid 19 data related to the whole state of Massachusetts are chosen, because the auther believes that suburban area and capital city are closely related in the context of disease and crime. It is well known that in America most of the residential area, suburban or rual area, is separately identifiable from commercial zone, the cities. While the crime might have happened in the city of Boston, the suspects or victims might live outside of Boston.

Confirmed cases instead of Death count

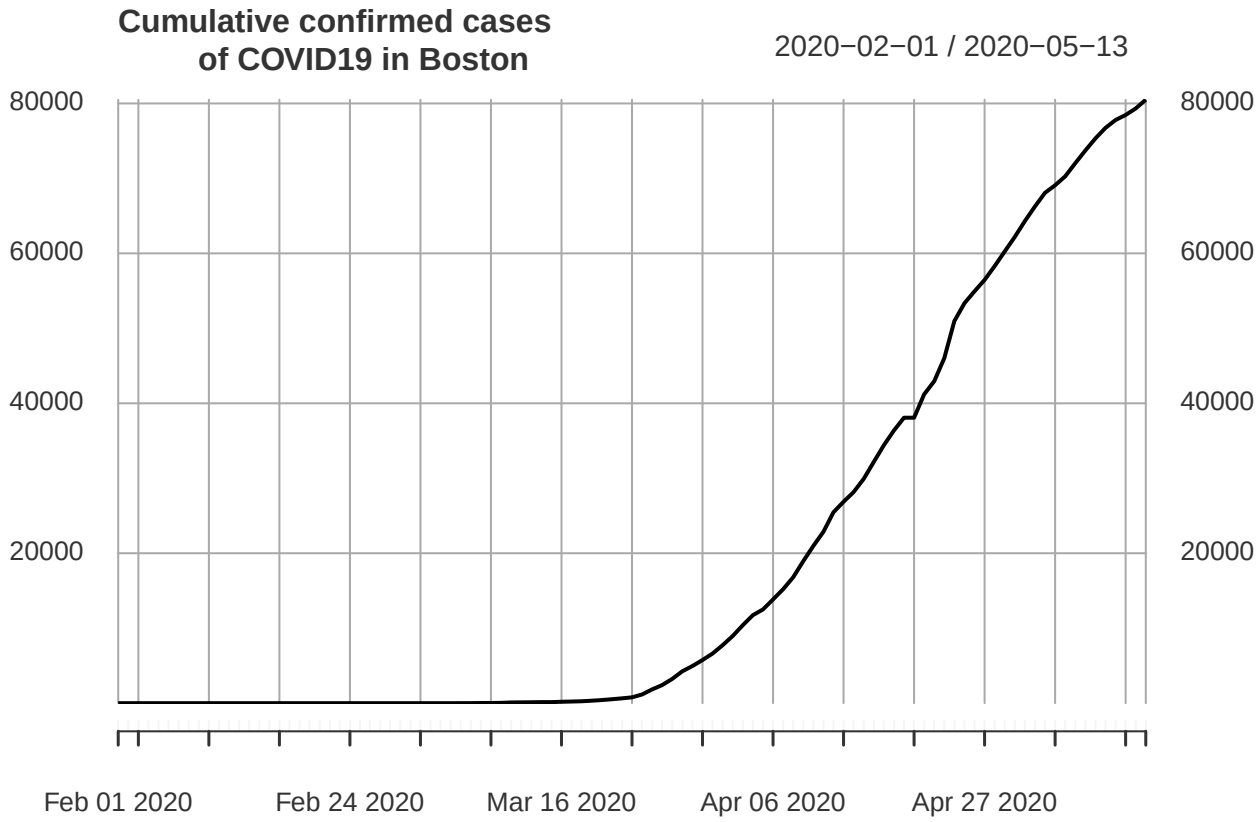
Although the number of confirmed cases can largely be influenced by the testing policy of the local government, the psychological effects of seeing confirmed cases alone might be enough to have some psychological relationship with committing crimes, which is what the author would like to explore.

Overview of the data

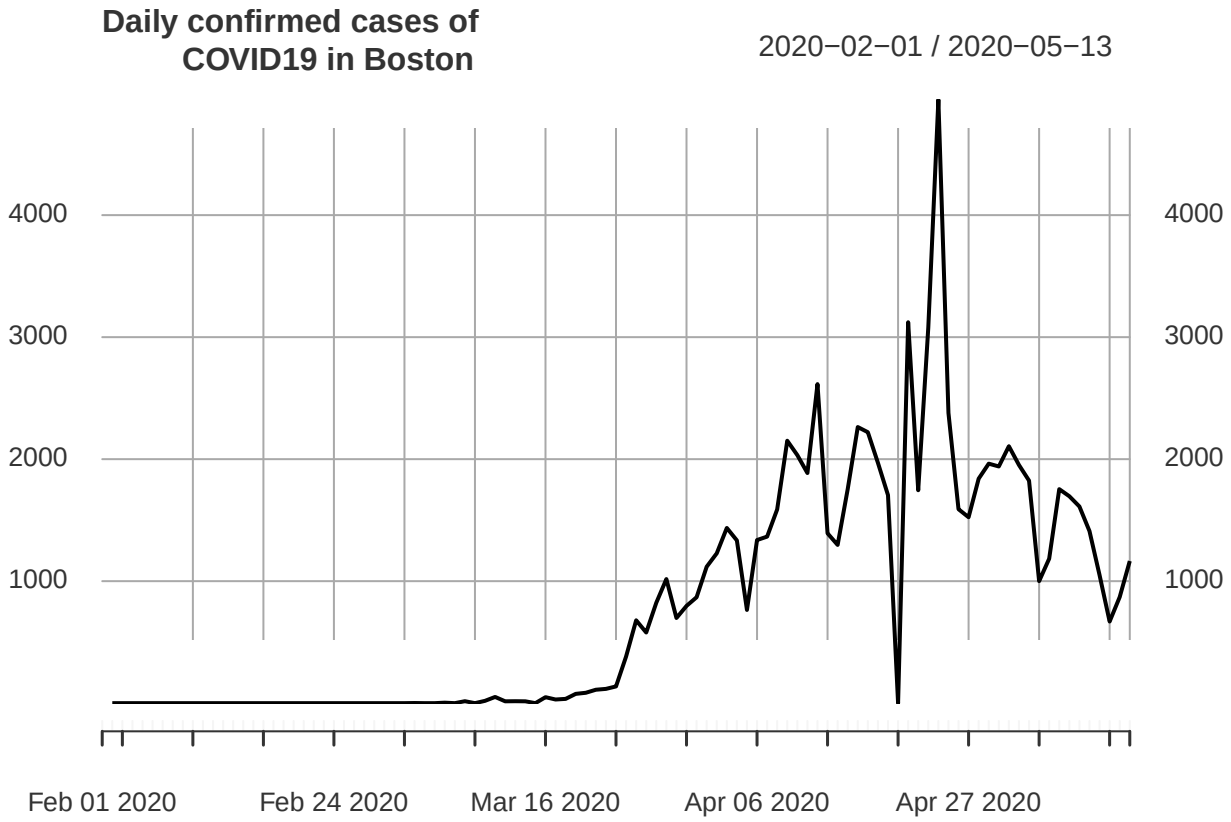
Visualization

```
# plot cumulative cases
# extract for tranforming into time series data
ts_MA <- covid19_MA %>%
  dplyr::select(date, confirmed) %>%
  ts_xts()

plot.xts(ts_MA,
         main = "Cumulative confirmed cases
         of COVID19 in Boston")
```




```
# plot daily cases
# first difference
ts_diff_MA <- diff(ts_MA)
plot.xts(ts_diff_MA,
         main = "Daily confirmed cases of
         COVID19 in Boston")
```



As shown above, cumulative cases and daily cases have been plotted above.

Model the infection

Construct the model

```
# construct GAMM model from prof. Brown's work

# construct data frame of difference, not time series
covid19_MA_diff <- data.frame(diff(covid19_MA$confirmed))
colnames(covid19_MA_diff)[1] = "confirmed"
covid19_MA_diff$date = covid19_MA$date[2:length(covid19_MA$date)]

# time as integer
covid19_MA_diff$timeInt = as.numeric(covid19_MA_diff$date)
# make a copy to avoid perfect collinearity for mixed effect
covid19_MA_diff$timeIid = covid19_MA_diff$timeInt

# GAMM model
gamMA <- gamm4::gamm4(confirmed ~ s(timeInt, k = 100),
                      random = ~(1|timeIid),
                      data = covid19_MA_diff,
                      family = poisson(link = 'log'))
# currently 100 is the max due to length of data
```

In order to study covid19's impact on Boston, its trend needs to be modeled first to have a better understanding of the situation. A Generalized Additive Mixed Model is used here, which is a direct copy from prof. Patrick Brown's work in STA303 Assignment 3.

Visuzalization of the model

```
# plot fitted value
toPredict = data.frame(time = seq(covid19_MA_diff$date[1],
                                  covid19_MA_diff$date[length(covid19_MA_diff$date)],
                                  by = '1 day'))
toPredict$timeInt = as.numeric(toPredict$time)

# plot

matplot(toPredict$time,
        exp(do.call(cbind, mgcv::predict.gam(gamMA$gam, toPredict, se.fit=TRUE)) %*%
              Pmisc::ciMat()),
        col='red', lty=c(1,2,2), type='l', xaxt='n', xlab='', ylab='Daliy Confirmed cases',
        ylim = c(0.5, 3500), xlim = as.Date(c(covid19_MA$date[1], covid19_MA$date[length(covid19_MA$dat
title("Daily confirmed cases of COVID 19 in Boston")

matpoints(toPredict$time, covid19_MA_diff$confirmed,
          col = 'black',
          type = 'l')
```