initial exploration

Stuart Barnum

8/29/2022

```r
knitr::opts_chunk$set(echo = TRUE)
```

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(funtimes)

historic <- read_csv("NYPD_Arrests_Data__Historic.csv",
                     guess_max = Inf)
```

```
## Rows: 5308876 Columns: 19
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (10): ARREST_DATE, PD_DESC, OFNS_DESC, LAW_CODE, LAW_CAT_CD, ARREST_BORO...
## dbl  (9): ARREST_KEY, PD_CD, KY_CD, ARREST_PRECINCT, JURISDICTION_CODE, X_CO...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
summary(historic)
```

```
##    ARREST_KEY        ARREST_DATE            PD_CD         PD_DESC
##  Min.   :  9926901   Length:5308876     Min.   :  0.0   Length:5308876
##  1st Qu.: 61436632   Class :character   1st Qu.:269.0   Class :character
##  Median : 85671028   Mode  :character   Median :511.0   Mode  :character
##  Mean   :102879939                      Mean   :505.8
##  3rd Qu.:150090000                      3rd Qu.:748.0
##  Max.   :238513928                      Max.   :997.0
```

```
##                                         NA's   :313
##      KY_CD         OFNS_DESC           LAW_CODE          LAW_CAT_CD
##  Min.   :101.0   Length:5308876     Length:5308876     Length:5308876
##  1st Qu.:126.0   Class :character   Class :character   Class :character
##  Median :341.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :298.4
##  3rd Qu.:348.0
##  Max.   :995.0
##  NA's   :9169
##  ARREST_BORO        ARREST_PRECINCT  JURISDICTION_CODE  AGE_GROUP
##  Length:5308876     Min.   :  1.00   Min.   : 0.000    Length:5308876
##  Class :character   1st Qu.: 33.00   1st Qu.: 0.000    Class :character
##  Mode  :character   Median : 60.00   Median : 0.000    Mode  :character
##                     Mean   : 60.76   Mean   : 1.296
##                     3rd Qu.: 84.00   3rd Qu.: 0.000
##                     Max.   :123.00   Max.   :97.000
##                                      NA's   :10
##    PERP_SEX          PERP_RACE           X_COORD_CD        Y_COORD_CD
##  Length:5308876     Length:5308876     Min.   : 913357   Min.   : 121131
##  Class :character   Class :character   1st Qu.: 993280   1st Qu.: 186857
##  Mode  :character   Mode  :character   Median :1004892   Median : 209285
##                                        Mean   :1005355   Mean   : 214587
##                                        3rd Qu.:1015924   3rd Qu.: 236614
##                                        Max.   :1067302   Max.   :8202360
##                                        NA's   :1         NA's   :1
##     Latitude       Longitude        Lon_Lat
##  Min.   :40.50   Min.   :-74.25   Length:5308876
##  1st Qu.:40.68   1st Qu.:-73.97   Class :character
##  Median :40.74   Median :-73.93   Mode  :character
##  Mean   :40.76   Mean   :-73.92
##  3rd Qu.:40.82   3rd Qu.:-73.89
##  Max.   :62.08   Max.   :-73.68
##  NA's   :1       NA's   :1
```

```r
historic_w_date <- historic %>%
  mutate(Date = as.Date(ARREST_DATE, "%m/%d/%Y"))
summary(historic_w_date)
```

```
##    ARREST_KEY         ARREST_DATE            PD_CD          PD_DESC
##  Min.   :  9926901   Length:5308876     Min.   :  0.0   Length:5308876
##  1st Qu.: 61436632   Class :character   1st Qu.:269.0   Class :character
##  Median : 85671028   Mode  :character   Median :511.0   Mode  :character
##  Mean   :102879939                      Mean   :505.8
##  3rd Qu.:150090000                      3rd Qu.:748.0
##  Max.   :238513928                      Max.   :997.0
##                                         NA's   :313
##      KY_CD         OFNS_DESC           LAW_CODE          LAW_CAT_CD
##  Min.   :101.0   Length:5308876     Length:5308876     Length:5308876
##  1st Qu.:126.0   Class :character   Class :character   Class :character
##  Median :341.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :298.4
##  3rd Qu.:348.0
##  Max.   :995.0
##  NA's   :9169
##  ARREST_BORO        ARREST_PRECINCT  JURISDICTION_CODE  AGE_GROUP
```

```
##  Length:5308876     Min.   :  1.00   Min.   : 0.000    Length:5308876
##  Class :character   1st Qu.: 33.00   1st Qu.: 0.000    Class :character
##  Mode  :character   Median : 60.00   Median : 0.000    Mode  :character
##                     Mean   : 60.76   Mean   : 1.296
##                     3rd Qu.: 84.00   3rd Qu.: 0.000
##                     Max.   :123.00   Max.   :97.000
##                                      NA's   :10
##    PERP_SEX           PERP_RACE           X_COORD_CD        Y_COORD_CD
##  Length:5308876     Length:5308876     Min.   : 913357   Min.   : 121131
##  Class :character   Class :character   1st Qu.: 993280   1st Qu.: 186857
##  Mode  :character   Mode  :character   Median :1004892   Median : 209285
##                                        Mean   :1005355   Mean   : 214587
##                                        3rd Qu.:1015924   3rd Qu.: 236614
##                                        Max.   :1067302   Max.   :8202360
##                                        NA's   :1         NA's   :1
##     Latitude       Longitude        Lon_Lat              Date
##  Min.   :40.50   Min.   :-74.25   Length:5308876     Min.   :2006-01-01
##  1st Qu.:40.68   1st Qu.:-73.97   Class :character   1st Qu.:2009-05-05
##  Median :40.74   Median :-73.93   Mode  :character   Median :2012-07-10
##  Mean   :40.76   Mean   :-73.92                      Mean   :2012-11-10
##  3rd Qu.:40.82   3rd Qu.:-73.89                      3rd Qu.:2016-02-04
##  Max.   :62.08   Max.   :-73.68                      Max.   :2021-12-31
##  NA's   :1       NA's   :1
```

```
year_to_date <- read_csv("NYPD_Arrest_Data__Year_to_Date_.csv") %>%
  mutate(Date = as.Date(ARREST_DATE, "%m/%d/%Y"))
```

```
## Rows: 93238 Columns: 19
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (10): ARREST_DATE, PD_DESC, OFNS_DESC, LAW_CODE, LAW_CAT_CD, ARREST_BORO...
## dbl  (9): ARREST_KEY, PD_CD, KY_CD, ARREST_PRECINCT, JURISDICTION_CODE, X_CO...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

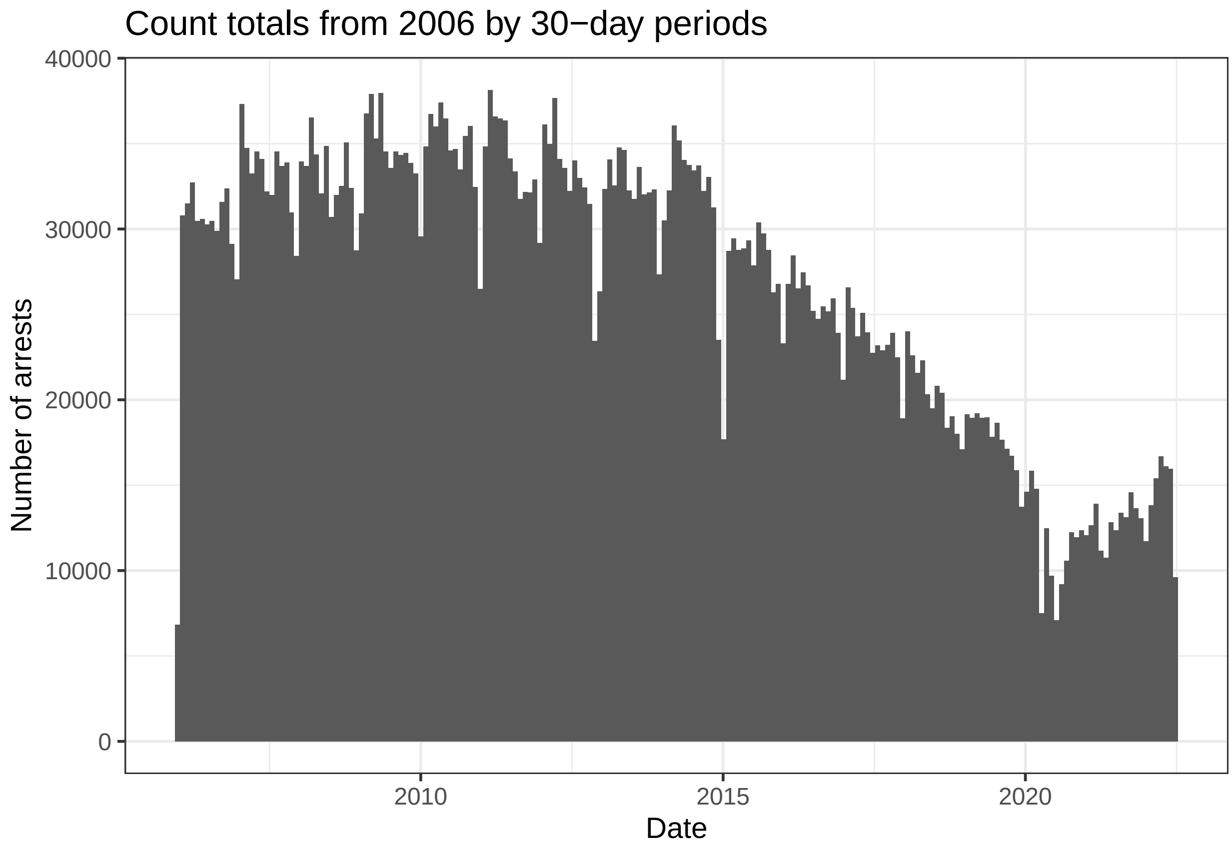
```
all_arrests <- bind_rows(historic_w_date, year_to_date)

summary(all_arrests)
```

```
##    ARREST_KEY         ARREST_DATE            PD_CD          PD_DESC
##  Min.   :  9926901   Length:5402114     Min.   :  0.0   Length:5402114
##  1st Qu.: 62053042   Class :character   1st Qu.:268.0   Class :character
##  Median : 86350562   Mode  :character   Median :510.0   Mode  :character
##  Mean   :105300015                      Mean   :504.1
##  3rd Qu.:152357173                      3rd Qu.:748.0
##  Max.   :247417454                      Max.   :997.0
##                                         NA's   :546
##      KY_CD        OFNS_DESC           LAW_CODE           LAW_CAT_CD
##  Min.   :101.0   Length:5402114     Length:5402114     Length:5402114
##  1st Qu.:126.0   Class :character   Class :character   Class :character
##  Median :341.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :297.5
##  3rd Qu.:348.0
##  Max.   :995.0
```

```
##  NA's   :9473
##  ARREST_BORO        ARREST_PRECINCT JURISDICTION_CODE  AGE_GROUP
##  Length:5402114     Min.   :  1.0   Min.   : 0.00     Length:5402114
##  Class :character   1st Qu.: 33.0   1st Qu.: 0.00     Class :character
##  Mode  :character   Median : 60.0   Median : 0.00     Mode  :character
##                     Mean   : 60.8   Mean   : 1.29
##                     3rd Qu.: 84.0   3rd Qu.: 0.00
##                     Max.   :123.0   Max.   :97.00
##                                     NA's   :10
##    PERP_SEX          PERP_RACE          X_COORD_CD        Y_COORD_CD
##  Length:5402114     Length:5402114     Min.   : 913357   Min.   : 121131
##  Class :character   Class :character   1st Qu.: 993212   1st Qu.: 186857
##  Mode  :character   Mode  :character   Median :1004892   Median : 209223
##                                        Mean   :1005347   Mean   : 214481
##                                        3rd Qu.:1015947   3rd Qu.: 236608
##                                        Max.   :1067302   Max.   :8202360
##                                        NA's   :1         NA's   :1
##     Latitude       Longitude        Lon_Lat               Date
##  Min.   :40.50   Min.   :-74.25   Length:5402114     Min.   :2006-01-01
##  1st Qu.:40.68   1st Qu.:-73.97   Class :character   1st Qu.:2009-05-22
##  Median :40.74   Median :-73.93   Mode  :character   Median :2012-08-20
##  Mean   :40.76   Mean   :-73.92                      Mean   :2013-01-09
##  3rd Qu.:40.82   3rd Qu.:-73.89                      3rd Qu.:2016-04-21
##  Max.   :62.08   Max.   :-73.68                      Max.   :2022-06-30
##  NA's   :1       NA's   :1
##  New Georeferenced Column
##  Length:5402114
##  Class :character
##  Mode  :character
##
##
##
##
```

Have a quick look at counts for 30 day periods for the entire period from 2006. I use the entire period, for now, to perhaps uncover cyclical affects (e.g. over each year) that may be distorted by COVID.

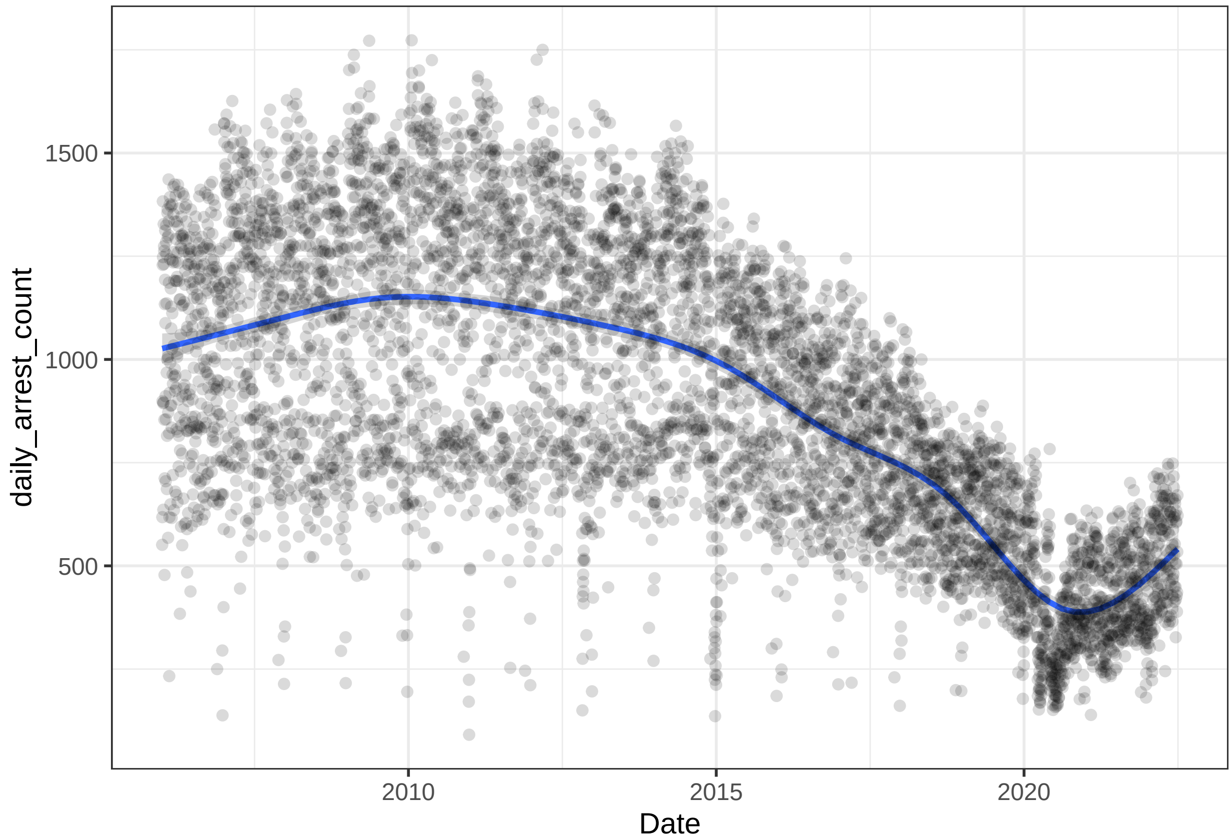
```r
ggplot(all_arrests) + geom_histogram(aes(Date), binwidth=30) +
  labs(y = "Number of arrests", title="Count totals from 2006 by 30-day periods") +
  theme_bw()
```



Find arrest counts for each day, to facilitate modeling. Have a quick look, noting that that the seeming cyclical effects, though somewhat evident for earlier years, are now more difficult to discern.

```
arrest_day_counts <-
  all_arrests %>%
  group_by(Date) %>%
  summarize(daily_arrest_count = n())

#plot the daily counts together with a loess-smoothed curve
ggplot(arrest_day_counts, aes(x=Date, y = daily_arrest_count)) +
  geom_smooth() + geom_point(alpha = .15)+
  theme_bw()
```

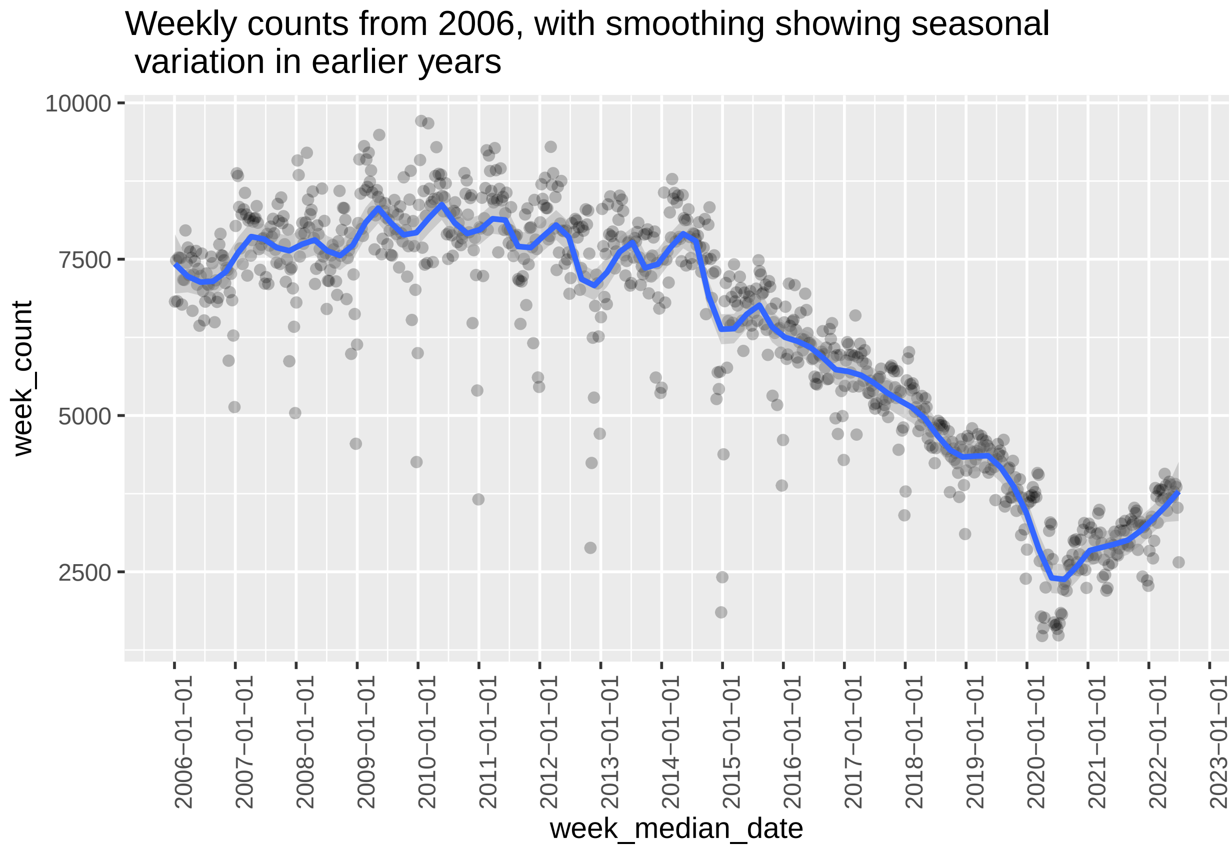
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



I might expect to see differences in week days, with e.g. more arrests of certain types on weekends. I will bin the weeks, obtaining the count for each week. Assign these counts to the median day of the week. This will also facilitate satisfaction (or approximate satisfaction) of the regression assumptions if I do simple regression to address Question 1.

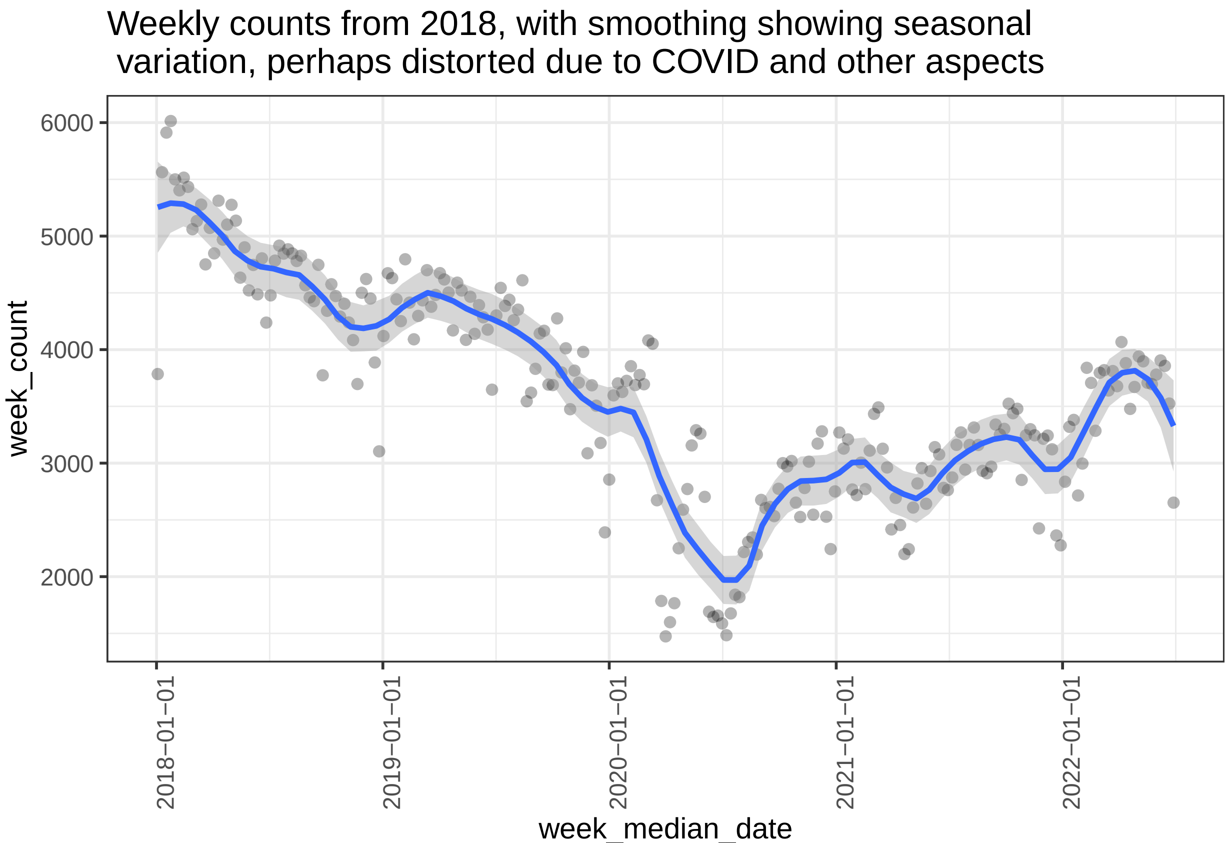
```
week_counts <- all_arrests %>%
  mutate(week_floor = floor_date(Date, unit="weeks"),
         week_median_date = week_floor + 3) %>%
  group_by(week_median_date) %>%
  summarize(week_count = n())

ggplot(week_counts, aes(week_median_date, week_count)) +
  geom_point(alpha = .25) +
  geom_smooth(method = "loess", span = .1) +
  scale_x_date(breaks = '1 year') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Weekly counts from 2006, with smoothing showing seasonal\n variation in earlier years")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
week_counts_fm_2018 <- week_counts %>%
  filter(week_median_date >= ymd("2018-01-01"))

ggplot(week_counts_fm_2018, aes(week_median_date, week_count)) +
  geom_point(alpha = .3) +
  geom_smooth(method = "loess", span = .15) +
  theme_bw() +
  scale_x_date(breaks = '1 year') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Weekly counts from 2018, with smoothing showing seasonal\n variation, perhaps distorted
```

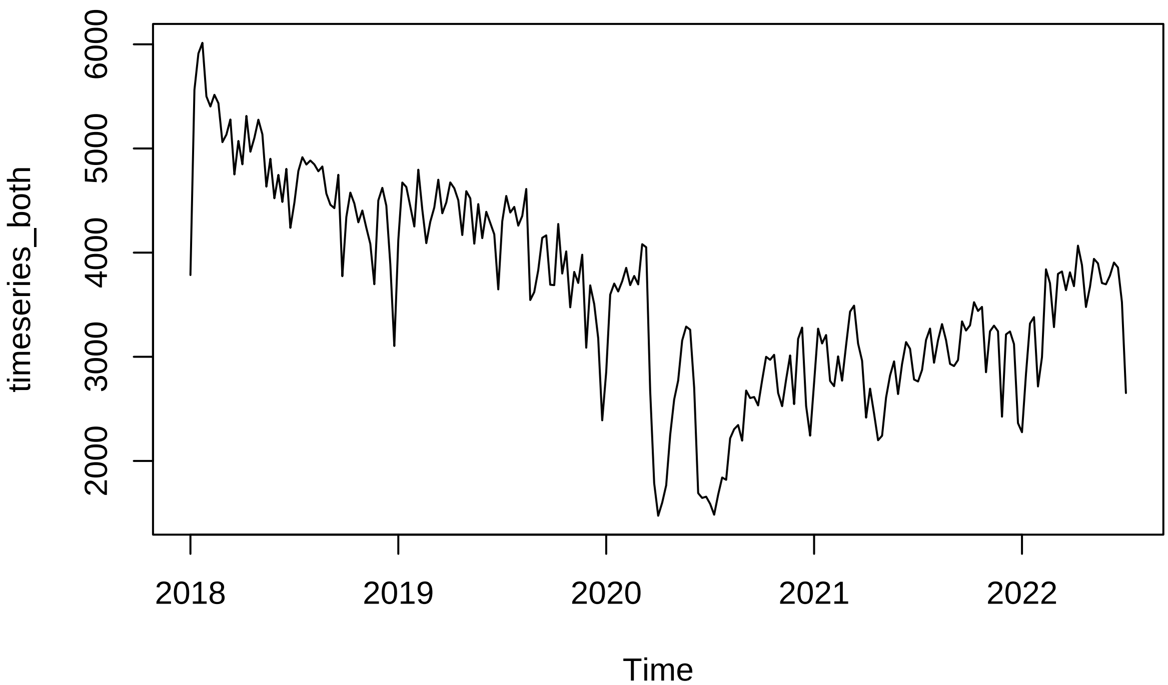
```
## `geom_smooth()` using formula 'y ~ x'
```



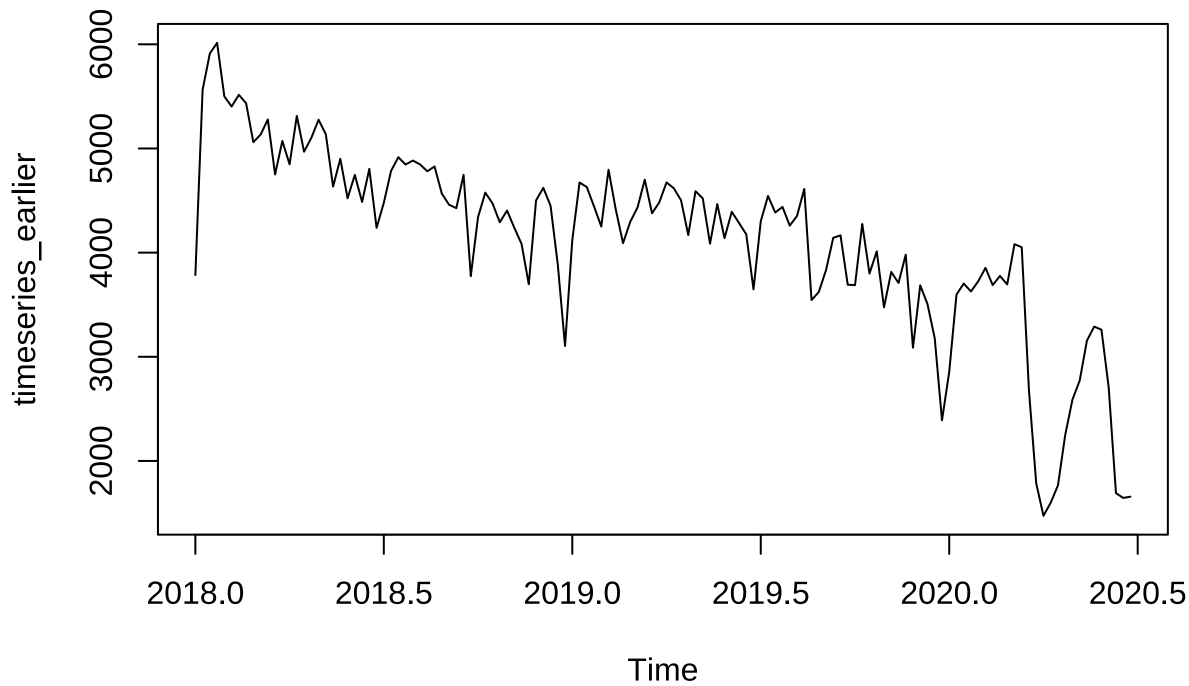
It appears that arrests decreased until the middle of 2020 and then started to increase. I will divide the data into these two time periods and apply a Mann-Kendall test with sieve bootstrap two the two results time series. This test allows that data be autocorrelated and that there be periodicity, which seems to exist, even if it is rather irregular.

As a first step, I put the relevant data into a time series R object and create a quick plot of the results

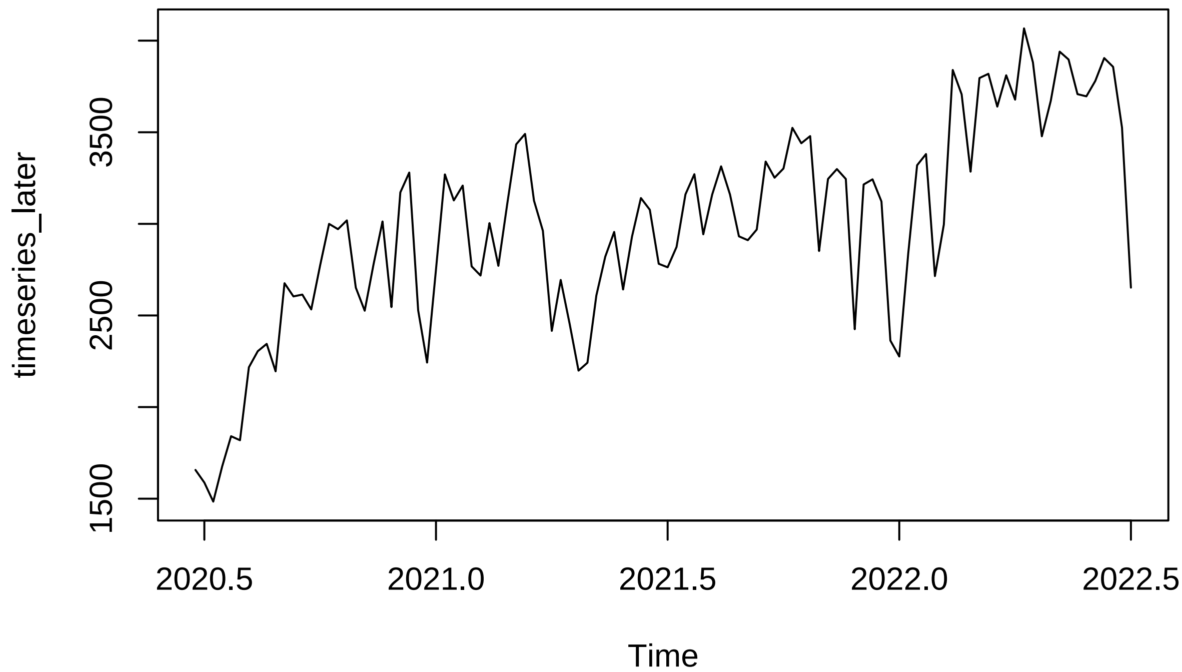
```
#these two results might be used with daily counts
first_obs <- week_counts_fm_2018 %>%
  filter(week_median_date==min(week_median_date))
last_obs <- week_counts_fm_2018 %>%
  filter(week_median_date==max(week_median_date))

#the times series object for weekly counts
timeseries_both <- ts(week_counts_fm_2018$week_count,
                  start = c(2018, 1),
                  frequency = 52)
plot(timeseries_both)
```




```
timeseries_earlier <- window(timeseries_both,
                  end = c(2020, 26))
plot(timeseries_earlier)
```




```
timeseries_later <- window(timeseries_both,
                    start = c(2020, 26))
plot(timeseries_later)
```



Now do the statistical signficance tests on upward and downward trends for the respective time period

```
notrend_test(timeseries_earlier, B=1000, test='MK')
```

```
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  timeseries_earlier
## Mann--Kendall's tau = -0.6877, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 1
##
## $AR_coefficients
##     phi_1
## 0.6165258
```

```
notrend_test(timeseries_later, B=1000, test='MK')
```

```
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  timeseries_later
## Mann--Kendall's tau = 0.54822, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
```

```
## [1] 4
##
## $AR_coefficients
##       phi_1       phi_2       phi_3       phi_4
##  0.66380318 -0.24681672  0.09182166  0.16597143
```

Moving on the Question 2, we count numbers of arrests by ‘pd_desc, and then subset the arrests dataset to list only the arrests within pd_desc categories with the top 5 counts.

```r
pd_desc_counts_top5 <- all_arrests %>%
  filter(Date >= ymd('2018-01-01')) %>%
  count(PD_DESC) %>%
  arrange(desc(n)) %>%
  top_n(5)
```

```
## Selecting by n
```

```r
print(pd_desc_counts_top5)
```

```
## # A tibble: 5 x 2
##   PD_DESC                             n
##   <chr>                           <int>
## 1 ASSAULT 3                       99757
## 2 LARCENY,PETIT FROM OPEN AREAS, 56001
## 3 ASSAULT 2,1,UNCLASSIFIED        51694
## 4 TRAFFIC,UNCLASSIFIED MISDEMEAN 40527
## 5 ROBBERY,OPEN AREA UNCLASSIFIED 29773
```

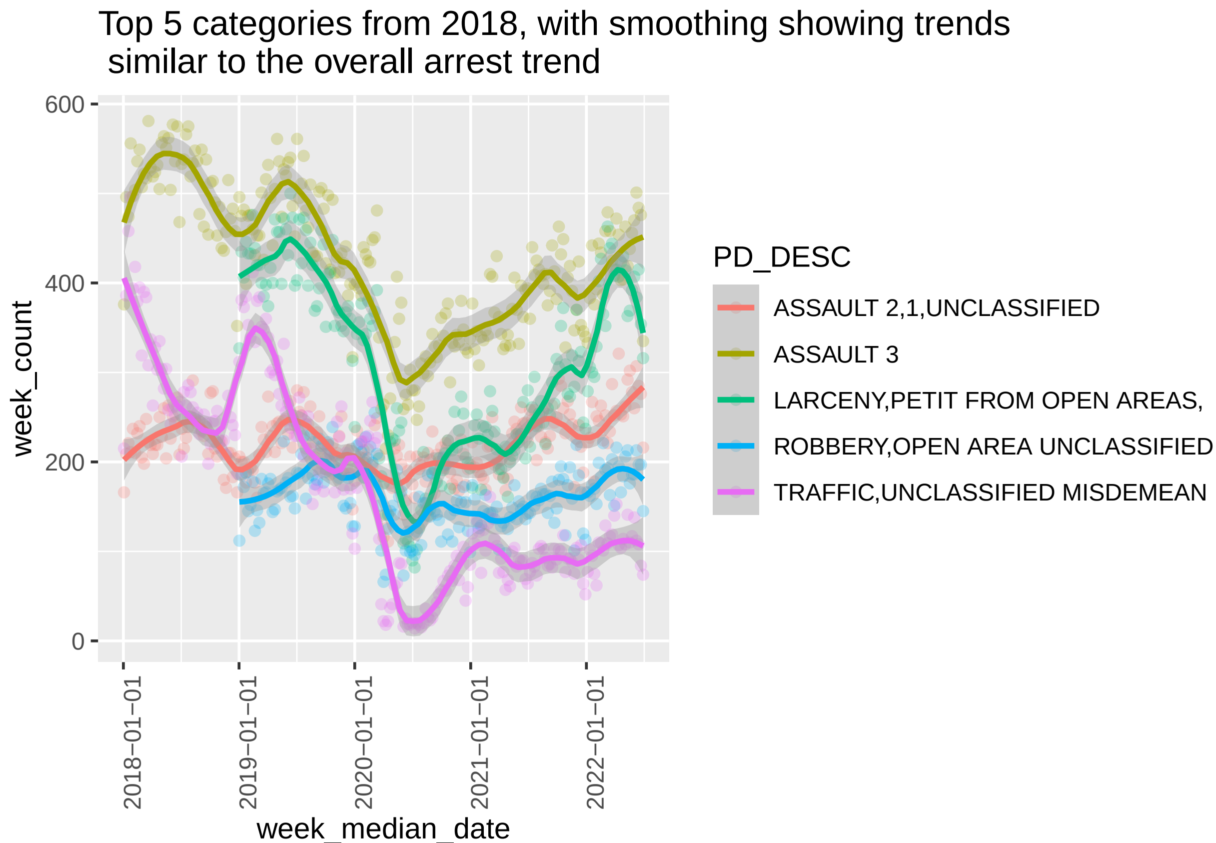
```r
pd_desc_subset <- all_arrests %>%
  filter(Date >= ymd('2018-01-01')) %>%
  inner_join(pd_desc_counts_top5, by="PD_DESC")

pd_desc_counts <- pd_desc_subset %>%
  mutate(week_floor = floor_date(Date, unit="weeks"),
         week_median_date = week_floor + 3) %>%
  group_by(PD_DESC, week_median_date) %>%
  summarize(week_count = n())
```

```
## `summarise()` has grouped output by 'PD_DESC'. You can override using the
## `.groups` argument.
```

Visualizing the apparent trend, we for the 5 categories with the highest count, we a similar count as with the older counts. Note thought that two o the categories do not have counts until the start of 2019.

```r
ggplot(pd_desc_counts, aes(x=week_median_date, y=week_count, color=PD_DESC)) +
  geom_point(alpha = .25) +
  geom_smooth(method = "loess", span = .2) +
  scale_x_date(breaks = '1 year') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Top 5 categories from 2018, with smoothing showing trends\n similar to the overall arre
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Put the counts for the 5 categories into separate variables.

```
top5wide <- pd_desc_counts %>%
  pivot_wider(names_from = PD_DESC,
              values_from = week_count)

top5wide %>% head()
```

```
## # A tibble: 6 x 6
##   week_median_date `ASSAULT 2,1,UNCLASSIFIED` ASSAULT ~1 LARCE~2 ROBBE~3 TRAFF~4
##   <date>                                <int>      <int>   <int>   <int>   <int>
## 1 2018-01-03                              166        376      NA      NA     215
## 2 2018-01-10                              209        496      NA      NA     386
## 3 2018-01-17                              221        474      NA      NA     458
## 4 2018-01-24                              211        556      NA      NA     496
## 5 2018-01-31                              237        497      NA      NA     394
## 6 2018-02-07                              210        504      NA      NA     418
## # ... with abbreviated variable names 1: `ASSAULT 3`,
## #   2: `LARCENY,PETIT FROM OPEN AREAS,`, 3: `ROBBERY,OPEN AREA UNCLASSIFIED`,
## #   4: `TRAFFIC,UNCLASSIFIED MISDEMEAN`
```

```
for (i in 2:6){

  print(str_c("Column ", i))

  ts <- ts(top5wide[,i],
           start = c(2018, 1),
           frequency = 52)
```

```r
  if (i %in% 4:5) {
    ts_early <- window(ts,
                       start = c(2019, 1),
                       end = c(2020, 26))
  } else {
    ts_early <- window(ts,
                       end = c(2020, 26))
  }

  print("Earlier Period")
  print(notrend_test(ts_early, B=1000, test='MK'))

  print("Later Period")
  ts_late <- window(ts,
                    start = c(2020, 26))
  print(notrend_test(ts_late, B=1000, test='MK'))
}
```

```
## [1] "Column 2"
## [1] "Earlier Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_early
## Mann--Kendall's tau = -0.20504, p-value = 0.027
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 1
##
## $AR_coefficients
##     phi_1
## 0.4007805
##
##
## [1] "Later Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_late
## Mann--Kendall's tau = 0.47345, p-value = 0.001
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 4
##
## $AR_coefficients
##       phi_1       phi_2       phi_3       phi_4
##  0.29067871  0.08557285 -0.13506349  0.30474540
##
##
## [1] "Column 3"
## [1] "Earlier Period"
```

```
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_early
## Mann--Kendall's tau = -0.51921, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 2
##
## $AR_coefficients
##     phi_1     phi_2
## 0.3195167 0.2159243
##
##
## [1] "Later Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_late
## Mann--Kendall's tau = 0.58399, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 4
##
## $AR_coefficients
##       phi_1       phi_2       phi_3       phi_4
## 0.287532895 0.005090214 0.080753378 0.115167169
##
##
## [1] "Column 4"
## [1] "Earlier Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_early
## Mann--Kendall's tau = -0.57969, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 2
##
## $AR_coefficients
##      phi_1      phi_2
## 0.56170486 0.04507668
##
##
## [1] "Later Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_late
## Mann--Kendall's tau = 0.7002, p-value < 2.2e-16
```

```
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 4
##
## $AR_coefficients
##      phi_1      phi_2      phi_3      phi_4
## 0.52260456 0.04355166 0.02173111 0.03941950
##
##
## [1] "Column 5"
## [1] "Earlier Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_early
## Mann--Kendall's tau = -0.0096958, p-value = 0.951
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 1
##
## $AR_coefficients
##     phi_1
## 0.5378248
##
##
## [1] "Later Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_late
## Mann--Kendall's tau = 0.41491, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 1
##
## $AR_coefficients
##     phi_1
## 0.1379164
##
##
## [1] "Column 6"
## [1] "Earlier Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_early
## Mann--Kendall's tau = -0.5449, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 4
```

```
##
## $AR_coefficients
##      phi_1      phi_2      phi_3      phi_4
##  0.65124765  0.12845373 -0.13398382  0.07014356
##
##
## [1] "Later Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_late
## Mann--Kendall's tau = 0.43043, p-value = 0.002
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 3
##
## $AR_coefficients
##      phi_1      phi_2      phi_3
## 0.520254352 0.002511877 0.060611831
```

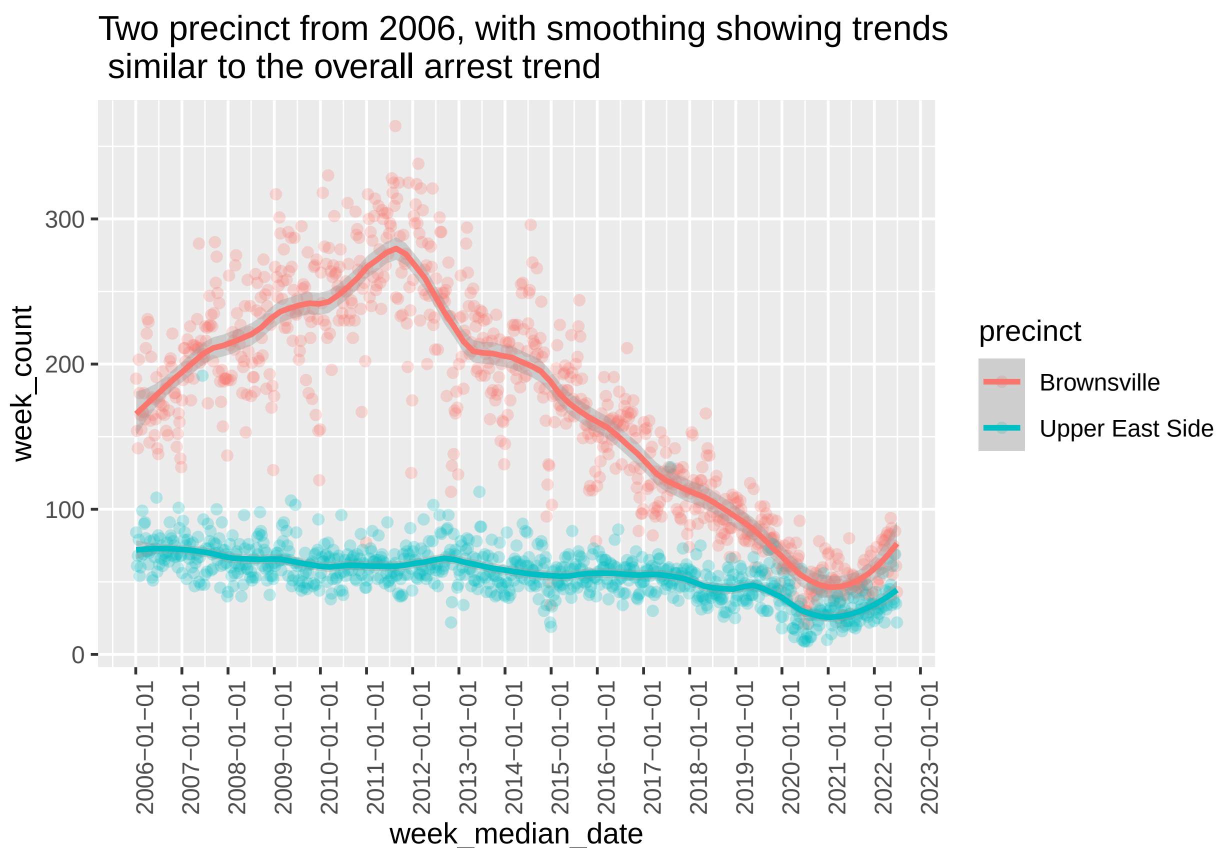
Turn to question 3 about the precincts

```
precinct_counts <- all_arrests %>%
  filter(ARREST_PRECINCT %in% c(19, 73)) %>%
  mutate(precinct = case_when(ARREST_PRECINCT == 19 ~ "Upper East Side",
                              ARREST_PRECINCT == 73 ~ "Brownsville")) %>%
  mutate(week_floor = floor_date(Date, unit="weeks"),
         week_median_date = week_floor + 3) %>%
  group_by(precinct, week_median_date) %>%
  summarize(week_count = n())
```

```
## `summarise()` has grouped output by 'precinct'. You can override using the
## `.groups` argument.
```

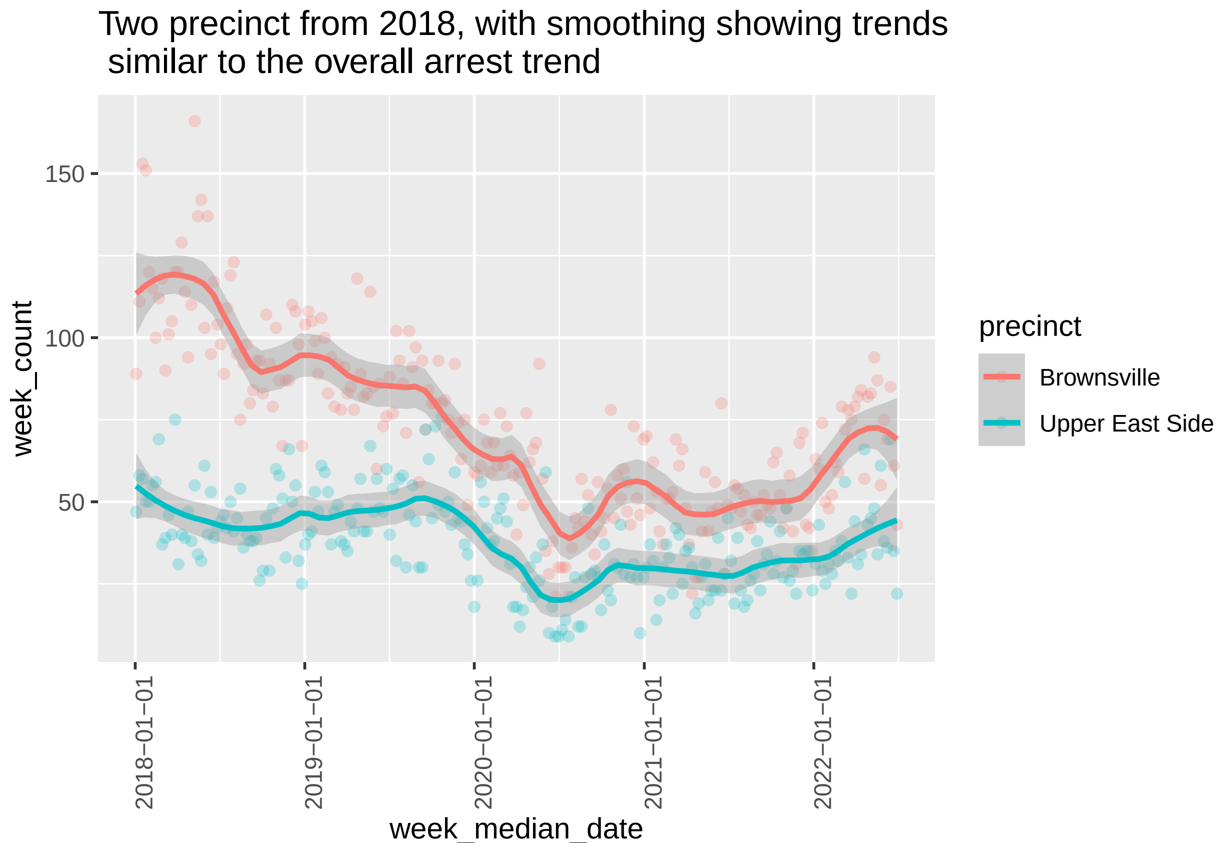
```
ggplot(precinct_counts, aes(x=week_median_date, y=week_count, color=precinct)) +
  geom_point(alpha = .25) +
  geom_smooth(method = "loess", span = .2) +
  scale_x_date(breaks = '1 year') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Two precinct from 2006, with smoothing showing trends\n similar to the overall arrest t
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
precinct_counts_fm_2018 <- precinct_counts %>%
  filter(week_median_date >= ymd("2018-01-01"))

ggplot(precinct_counts_fm_2018, aes(x=week_median_date, y=week_count, color=precinct)) +
  geom_point(alpha = .25) +
  geom_smooth(method = "loess", span = .2) +
  scale_x_date(breaks = '1 year') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Two precinct from 2018, with smoothing showing trends\n similar to the overall arrest t
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
precinct_wide <- precinct_counts_fm_2018 %>%
  pivot_wider(names_from = precinct,
              values_from = week_count)

precinct_wide %>% head()
```

```
## # A tibble: 6 x 3
##   week_median_date Brownsville `Upper East Side`
##   <date>                 <int>             <int>
## 1 2018-01-03                89                47
## 2 2018-01-10               111                58
## 3 2018-01-17               153                57
## 4 2018-01-24               151                50
## 5 2018-01-31               120                50
## 6 2018-02-07               115                55
```

```
for (i in 2:3){

  print(str_c("Column ", i))

  ts <- ts(precinct_wide[,i],
             start = c(2018, 1),
             frequency = 52)

  if (i %in% 4:5) {
    ts_early <- window(ts,
                    start = c(2019, 1),
```

```r
                    end = c(2020, 26))
  } else {
    ts_early <- window(ts,
                    end = c(2020, 26))
  }

  print("Earlier Period")
  print(notrend_test(ts_early, B=1000, test='MK'))

  print("Later Period")
  ts_late <- window(ts,
                     start = c(2020, 26))
  print(notrend_test(ts_late, B=1000, test='MK'))
}
```

```
## [1] "Column 2"
## [1] "Earlier Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_early
## Mann--Kendall's tau = -0.60437, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 4
##
## $AR_coefficients
##       phi_1       phi_2       phi_3       phi_4
##  0.27209432  0.02793462 -0.13415858  0.17738374
##
##
## [1] "Later Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_late
## Mann--Kendall's tau = 0.35488, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 1
##
## $AR_coefficients
##     phi_1
## 0.2825638
##
##
## [1] "Column 3"
## [1] "Earlier Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_early
```

```
## Mann--Kendall's tau = -0.17689, p-value = 0.033
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 2
##
## $AR_coefficients
##     phi_1     phi_2
## 0.2056698 0.1416509
##
##
## [1] "Later Period"
##
##  Sieve-bootstrap Mann--Kendall's trend test
##
## data:  ts_late
## Mann--Kendall's tau = 0.33838, p-value < 2.2e-16
## alternative hypothesis: monotonic trend.
## sample estimates:
## $AR_order
## [1] 3
##
## $AR_coefficients
##       phi_1       phi_2       phi_3
##  0.08875647 -0.07324782  0.23561994
```