

Exploratory Data Analysis

```
followers <- read.csv("/Users/Anton/Desktop/DSU Project/Artists Compiled/initial_artists_followers.csv")
listeners <- read.csv("/Users/Anton/Desktop/DSU Project/Artists Compiled/initial_artists_monthly_listeners.csv")
popularity <- read.csv("/Users/Anton/Desktop/DSU Project/Artists Compiled/initial_artists_popularity.csv")
festivals <- read.csv("/Users/Anton/Desktop/DSU Project/Artists Compiled/songkick_data_initial_small_artist.csv")
songs <- read.csv("/Users/Anton/Desktop/DSU Project/Artists Compiled/spotify_song_data_initial_small_artist.csv")

str(followers)

## 'data.frame': 35224 obs. of 4 variables:
## $ artist : chr "junkie kid" "junkie kid" "junkie kid" "junkie kid" ...
## $ DateTime : chr "2017-01-06 16:00:00" "2017-01-07 16:00:00" "2017-01-08 16:00:00" "2017-01-09 16:00:00" ...
## $ Total.Followers: int 2580 2589 2592 2608 2618 2628 2631 2637 2644 2649 ...
## $ Change : int 0 9 3 16 10 10 3 6 7 5 ...

followers$DateTime <- as.Date(followers$DateTime)
names(followers)[names(followers) == "Change"] <- "followers_change"

str(listeners)

## 'data.frame': 20980 obs. of 4 variables:
## $ artist : chr "junkie kid" "junkie kid" "junkie kid" "junkie kid" ...
## $ DateTime : chr "2017-01-15 16:00:00" "2017-03-09 16:00:00" "2017-04-02 17:00:00" "2017-04-03 17:00:00" ...
## $ Monthly.Listeners: int NA NA NA NA NA NA NA NA 117861 116918 116481 ...
## $ Change : int NA NA NA NA NA NA NA NA 0 -943 -437 ...

listeners$DateTime <- as.Date(listeners$DateTime)
names(listeners)[names(listeners) == "Change"] <- "listeners_change"

str(popularity)

## 'data.frame': 23379 obs. of 3 variables:
## $ artist : chr "junkie kid" "junkie kid" "junkie kid" "junkie kid" ...
## $ DateTime : chr "2017-01-06 16:00:00" "2017-01-07 16:00:00" "2017-01-08 16:00:00" "2017-01-09 16:00:00" ...
## $ Popularity: int 41 41 41 41 41 41 41 41 42 ...

listeners$DateTime <- as.Date(listeners$DateTime)

artist_reach <- merge(merge(followers, listeners, all = TRUE), popularity, all = TRUE)

artist_reach <- artist_reach[!is.na(artist_reach$Total.Followers),]

str(festivals)

## 'data.frame': 2550 obs. of 5 variables:
```

```

## $ X : int 6 7 8 9 10 11 12 13 17 18 ...
## $ Artist : chr "Junkie Kid" "Junkie Kid" "Junkie Kid" "Junkie Kid" ...
## $ Event : chr "New Horizons 2019" "New Horizons 2019" "New Horizons 2019" "New Horizons 2019" ...
## $ Festival.or.Concert: int 0 0 0 0 0 0 0 0 0 0 ...
## $ Date : chr "Wednesday 21 August 2019" "Thursday 22 August 2019" "Friday 23 August 2019" ...

festivals <- festivals[,-1]

library(stringr)
festivals$Date <- as.Date(str_remove(festivals$Date, "\\w*\\s"), format = "%d %B %Y")

festivals$Artist <- str_replace_all(tolower(festivals$Artist), "\\.", "")

str(songs)

## 'data.frame': 1304 obs. of 16 variables:
## $ X : int 0 1 2 3 4 5 6 7 8 9 ...
## $ artist_name : chr "Junkie Kid" "Junkie Kid" "Junkie Kid" "Junkie Kid" ...
## $ album_group : chr "album" "album" "single" "single" ...
## $ album_type : chr "album" "album" "single" "single" ...
## $ artists : chr "[{'external_urls': {'spotify': 'https://open.spotify.com/artist/3Cj...
## $ available_markets : chr "['AD', 'AE', 'AR', 'AT', 'AU', 'BE', 'BG', 'BH', 'BO', 'BR', 'CA', ...
## $ external_urls : chr "{'spotify': 'https://open.spotify.com/album/5kjDWsciPAjBS2y9vidJK0'...
## $ href : chr "https://api.spotify.com/v1/albums/5kjDWsciPAjBS2y9vidJK0" "https://...
## $ id : chr "5kjDWsciPAjBS2y9vidJK0" "5za9nsr6ne6HQDKrKcyQNB" "5NDy2Iyi9ULnMHUvys...
## $ images : chr "[{'height': 640, 'url': 'https://i.scdn.co/image/0c5cca729feb2ca247...
## $ name : chr "Puro Pinche Hardstyle, Vol. 1" "Find Yourself" "Love Is Dead (LNY T...
## $ release_date : chr "2019-10-18" "2018-05-04" "2018-07-09" "2018-05-25" ...
## $ release_date_precision: chr "day" "day" "day" "day" ...
## $ total_tracks : int 11 16 1 1 1 1 1 1 7 1 ...
## $ type : chr "album" "album" "album" "album" ...
## $ uri : chr "spotify:album:5kjDWsciPAjBS2y9vidJK0" "spotify:album:5za9nsr6ne6HQDKrKcyQNB" ...

songs <- songs[,c("artist_name", "name", "release_date", "total_tracks")]
songs$release_date <- as.Date(songs$release_date)
songs$artist_name <- str_replace_all(tolower(songs$artist_name), "\\.", "")

library(dplyr)

##
## Attaching package: 'dplyr'

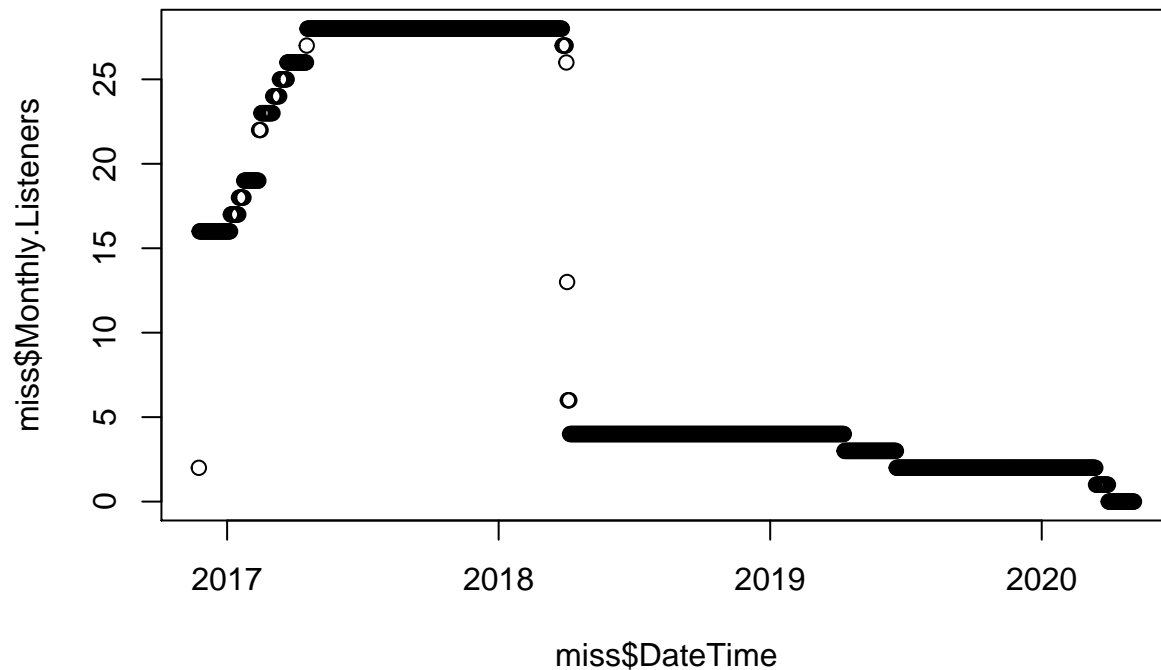
## The following objects are masked from 'package:stats':
##
## filter, lag

## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

miss <- artist_reach %>% select(-artist) %>% group_by(DateTime) %>% summarise_all(function(x) sum(is.na(x)))

plot(x = miss$DateTime, y = miss$Monthly.Listeners)

```



```

artist_reach <- artist_reach[artist_reach$DateTime > "2018-04-03",]

miss <- artist_reach %>% group_by(artist) %>% summarise_all(function(x) sum(is.na(x)))

missing_artists <- miss[miss$Monthly.Listeners > 5, ]$artist

remove_bad_artists_i <- !((artist_reach$artist %in% missing_artists & is.na(artist_reach$Monthly.Listeners)))

artist_reach <- artist_reach[remove_bad_artists_i,]

artist_reach <- artist_reach[!(is.na(artist_reach$Monthly.Listeners)),]

artist_reach <- artist_reach %>% select(-Popularity)

songs <- songs[!(is.na(songs$release_date)),]

library(xts)

```

```

## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
##
## Attaching package: 'xts'
## The following objects are masked from 'package:dplyr':
##
##   first, last

```

```

for(artist in unique(artist_reach$artist)) {
  example_artist_reach <- artist_reach[artist_reach$artist == artist, -1]

  example_festivals <- festivals[festivals$Artist == artist,]
  example_songs <- songs[songs$artist_name == artist,]

  example_ts <- xts(example_artist_reach[, -1], order.by = example_artist_reach[, 1])

  example_festivals <- example_festivals[!(duplicated(example_festivals[, 1:3])),]
  example_events_ts <- xts(example_festivals$Event, example_festivals$Date)
  example_songs_ts <- xts(example_songs$name, example_songs$release_date)

  plot(example_ts[, c(1)], legend.loc = "topleft", main = artist)
  addEventLines(example_events_ts, col = 3) # Green
  print(addEventLines(example_songs_ts, col = 4)) # Blue
  plot(example_ts[, c(3)], legend.loc = "topleft", main = artist)
  addEventLines(example_events_ts, col = 3) # Green
  print(addEventLines(example_songs_ts, col = 4)) # Blue
  plot(example_ts[, c(2)], legend.loc = "topleft", main = artist)
  addEventLines(example_events_ts, col = 3)
  print(addEventLines(example_songs_ts, col = 4))
  plot(example_ts[, c(4)], legend.loc = "topleft", main = artist)
  addEventLines(example_events_ts, col = 3)
  print(addEventLines(example_songs_ts, col = 4))
}

```

```

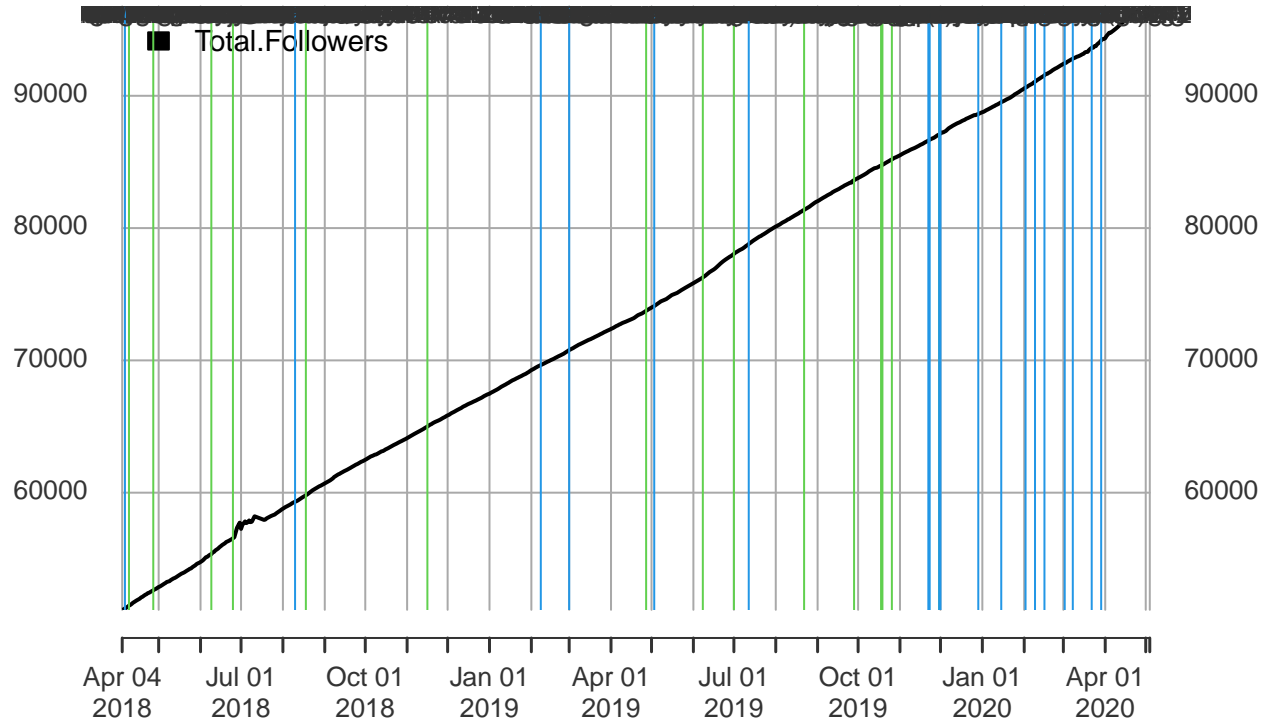
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbsToSbcs': dot substituted for <e2>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbsToSbcs': dot substituted for <80>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbsToSbcs': dot substituted for <94>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : font metrics unknown for Unicode character U+2014

```

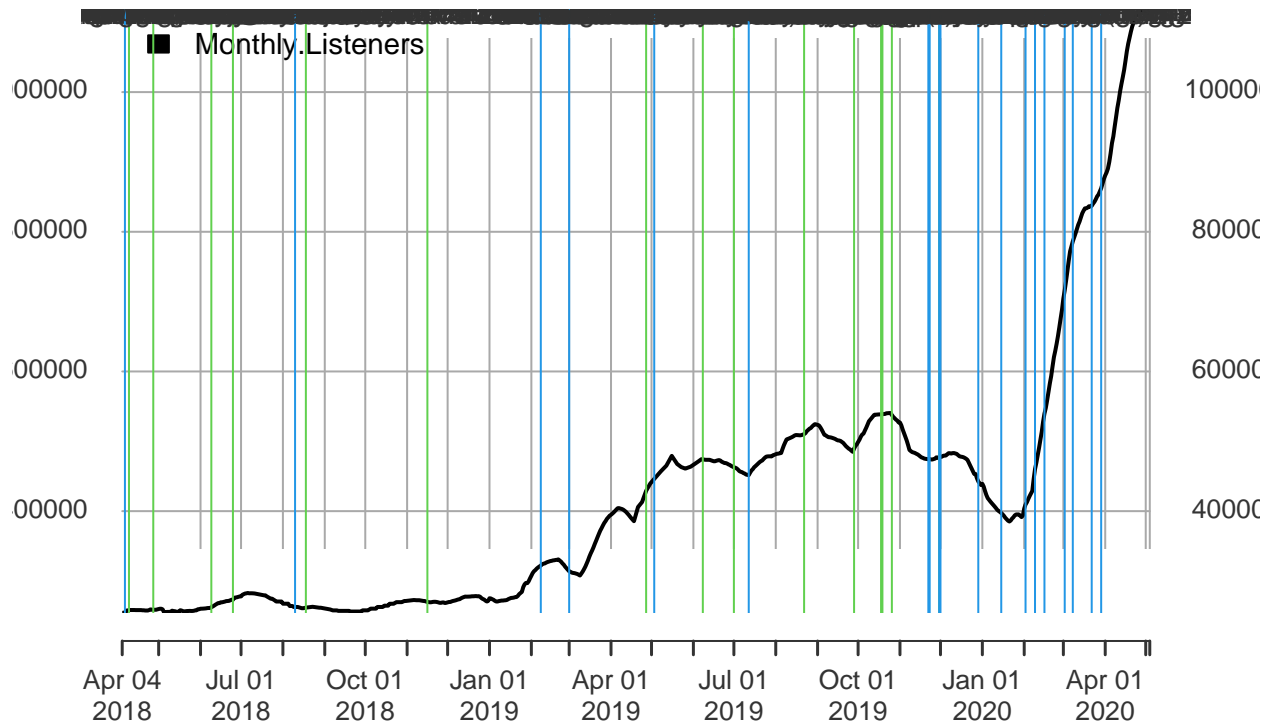


```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbcsToSbcs': dot substituted for <94>
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : font metrics unknown for Unicode character U+2014
```

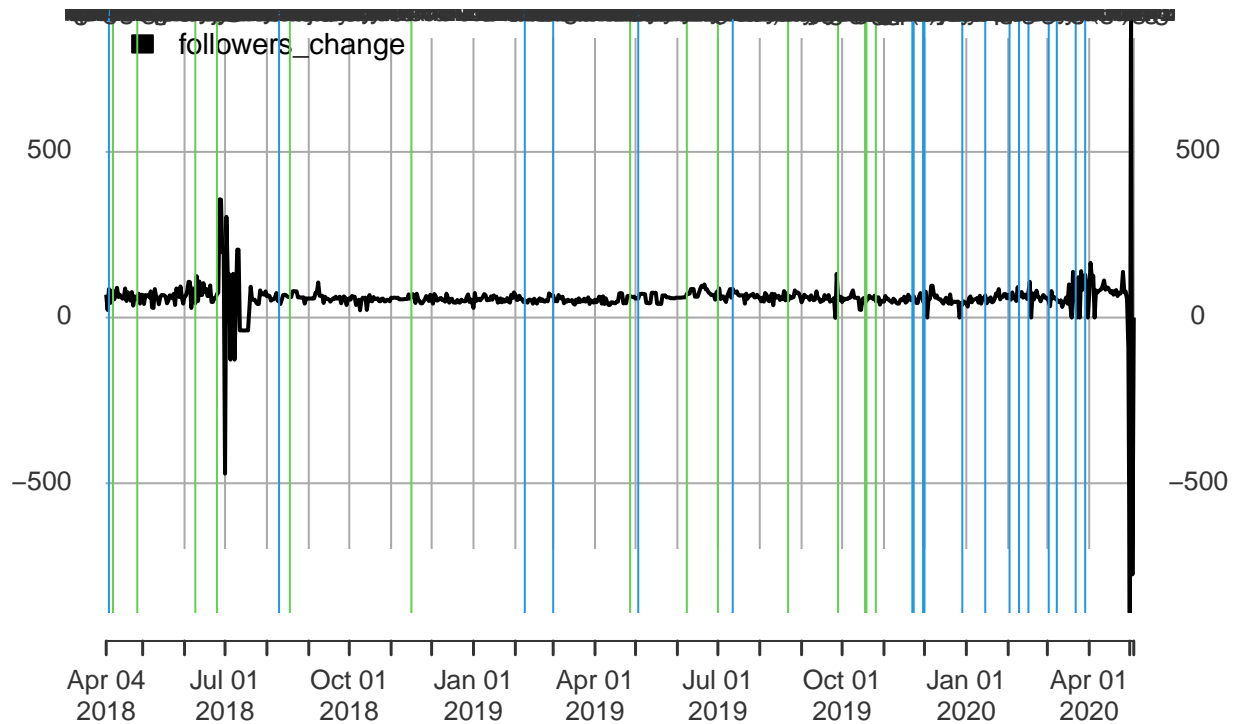


```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbcsToSbcs': dot substituted for <94>
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : font metrics unknown for Unicode character U+2014
```



```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbsToSbcs': dot substituted for <e2>
```

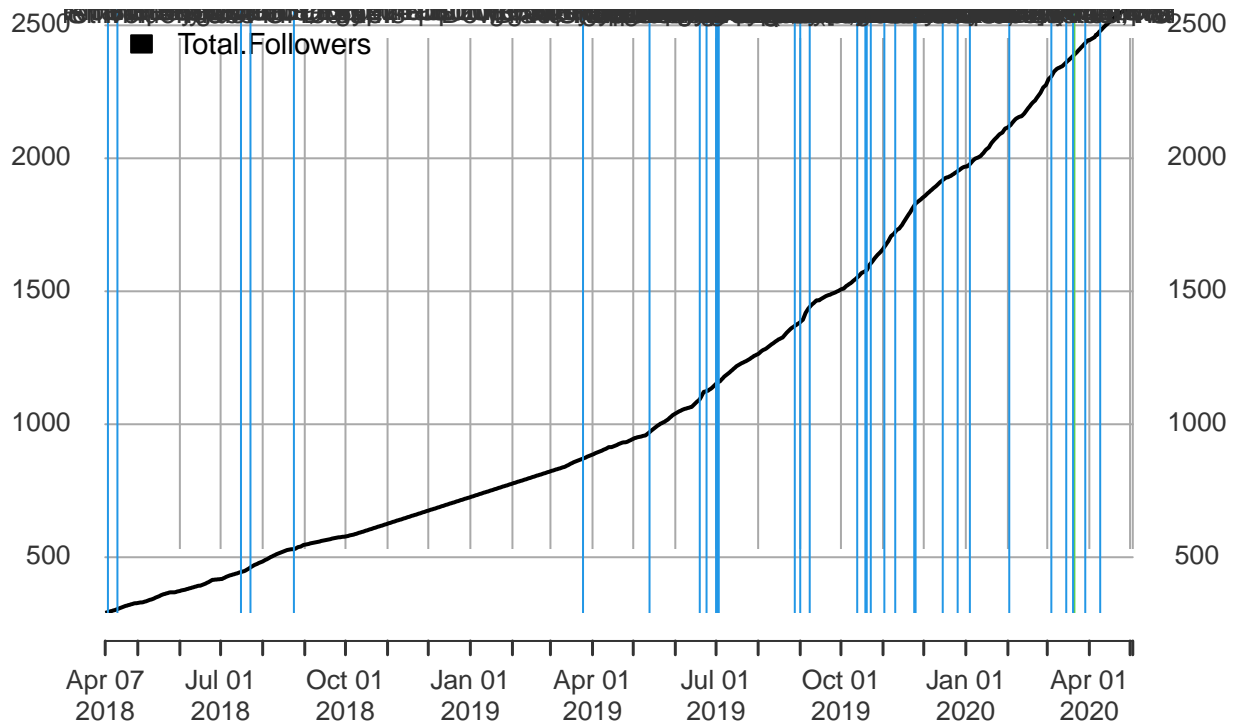
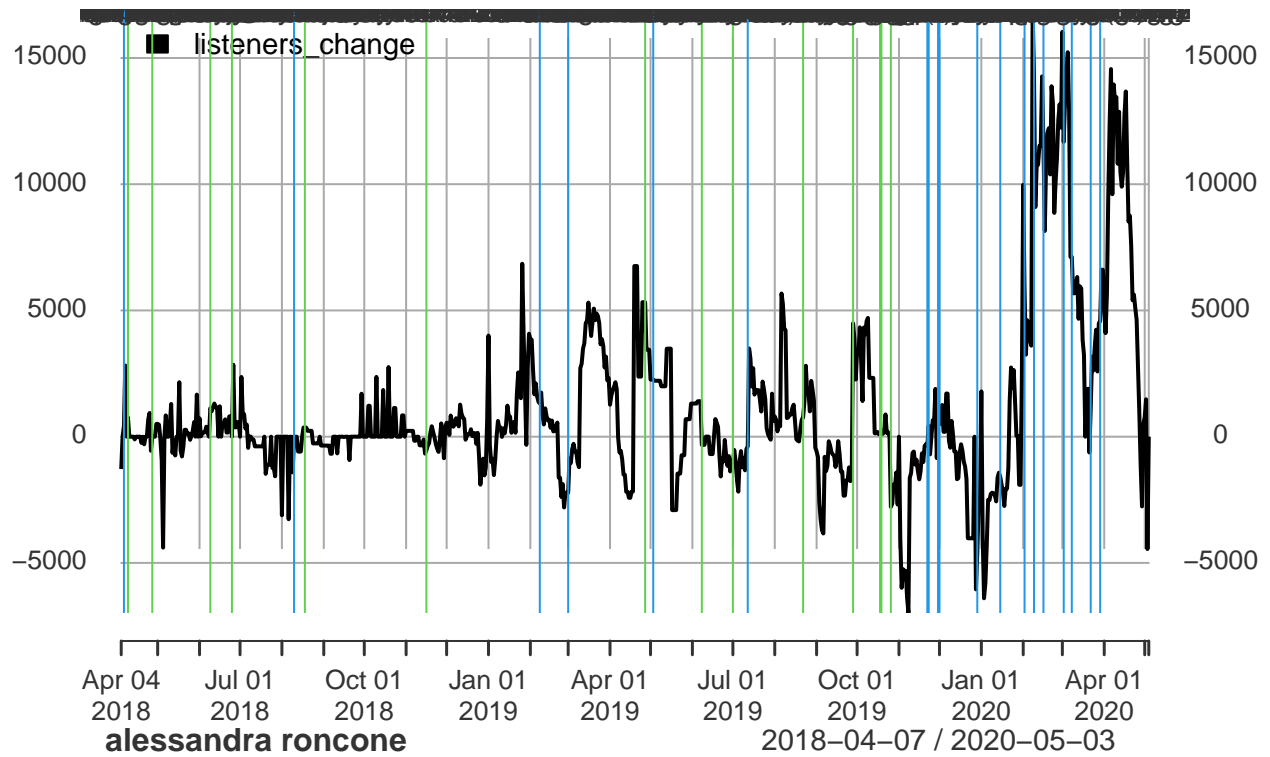
```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbsToSbcs': dot substituted for <80>
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'Awakenings: Day 2 - Drumcode
## 2012' in 'mbsToSbcs': dot substituted for <94>
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : font metrics unknown for Unicode character U+2014
```

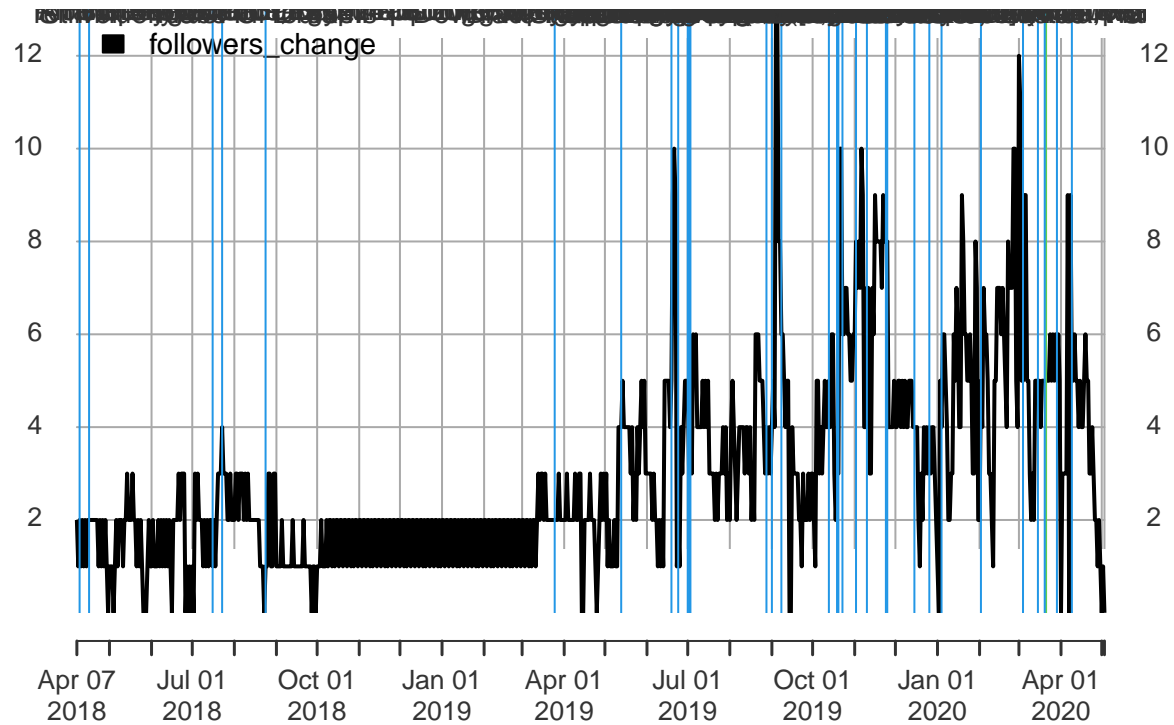
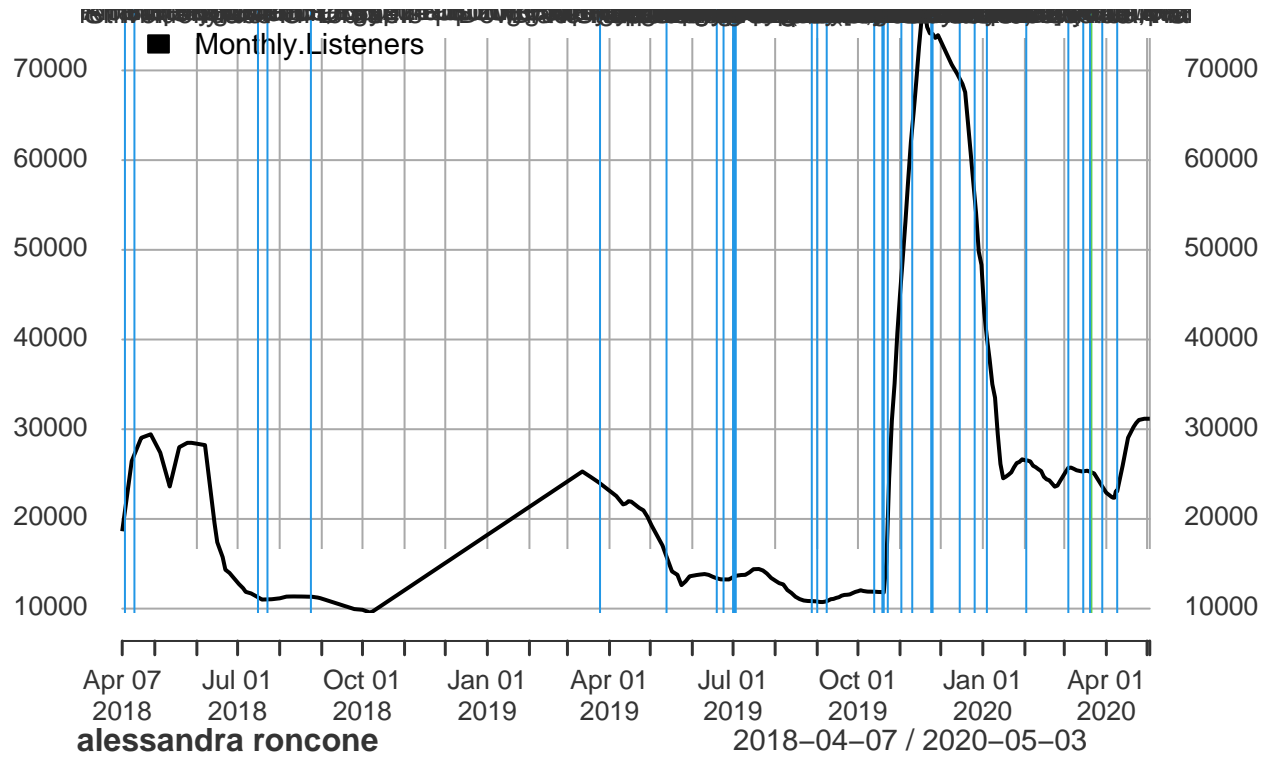
alan fitzpatrick

2018-04-04 / 2020-05-04



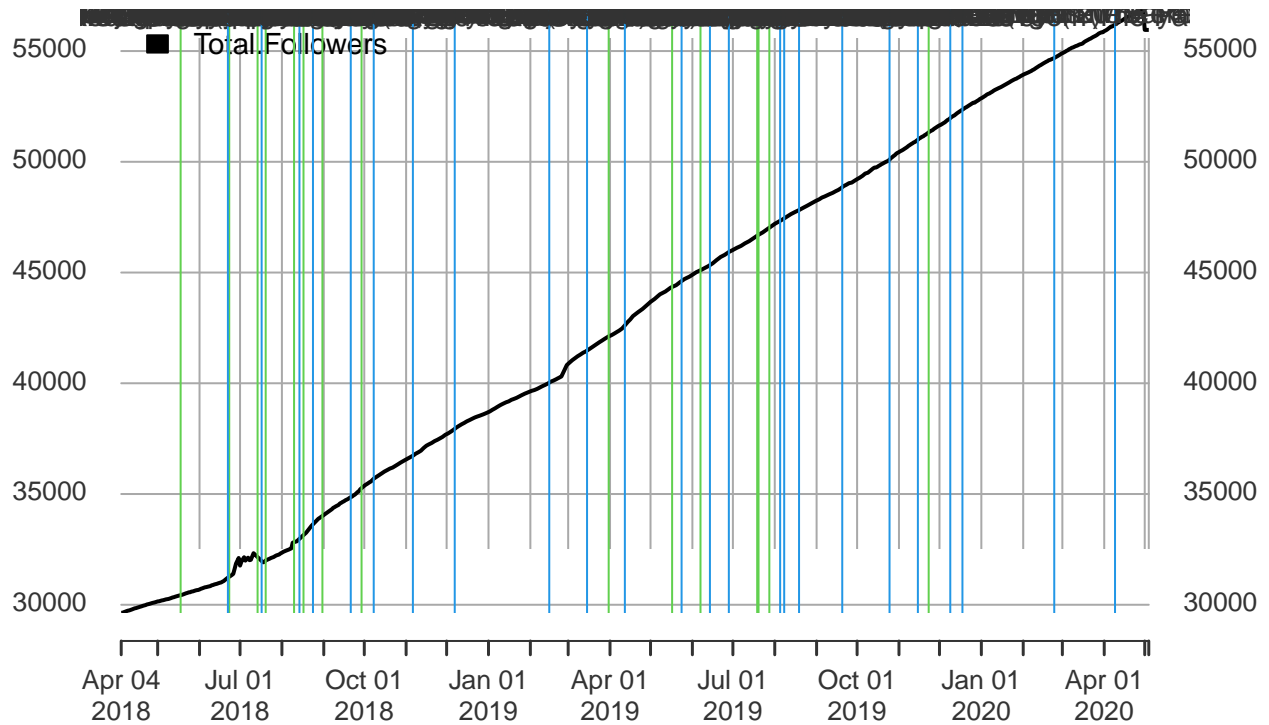
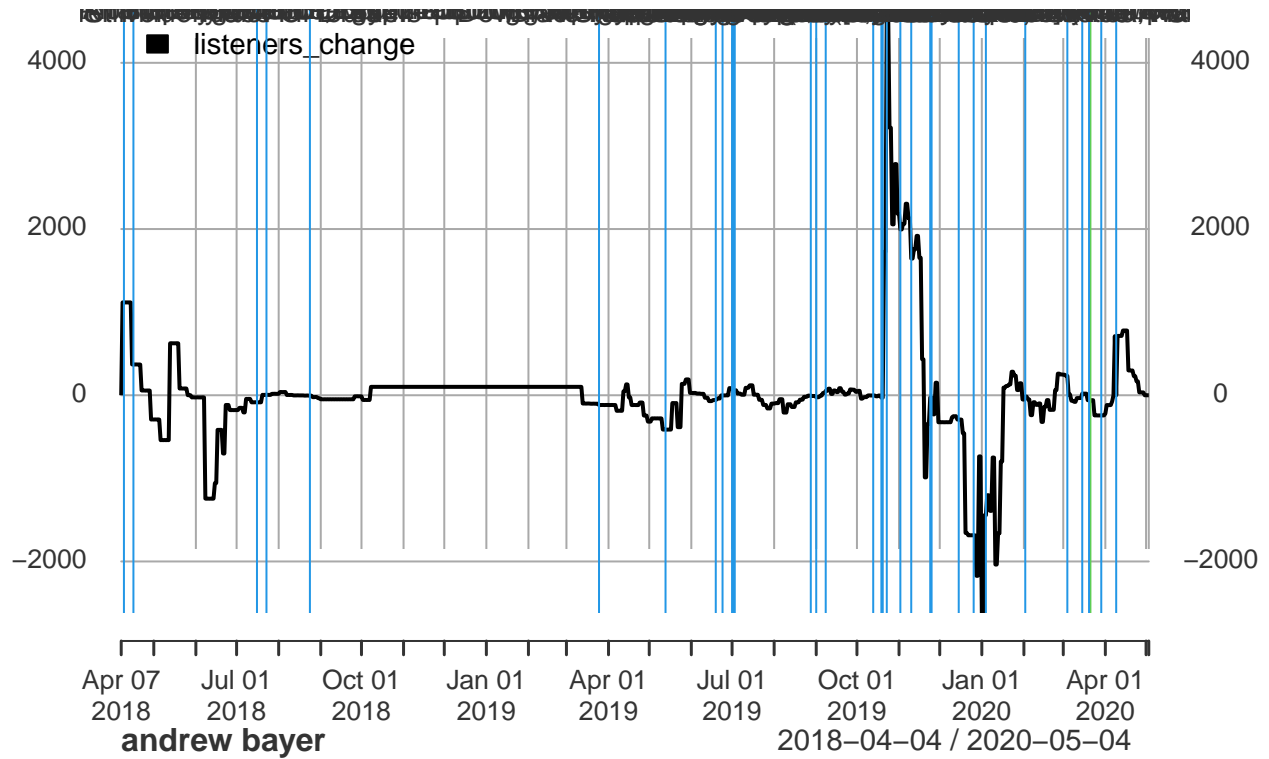
alessandra roncone

2018-04-07 / 2020-05-03



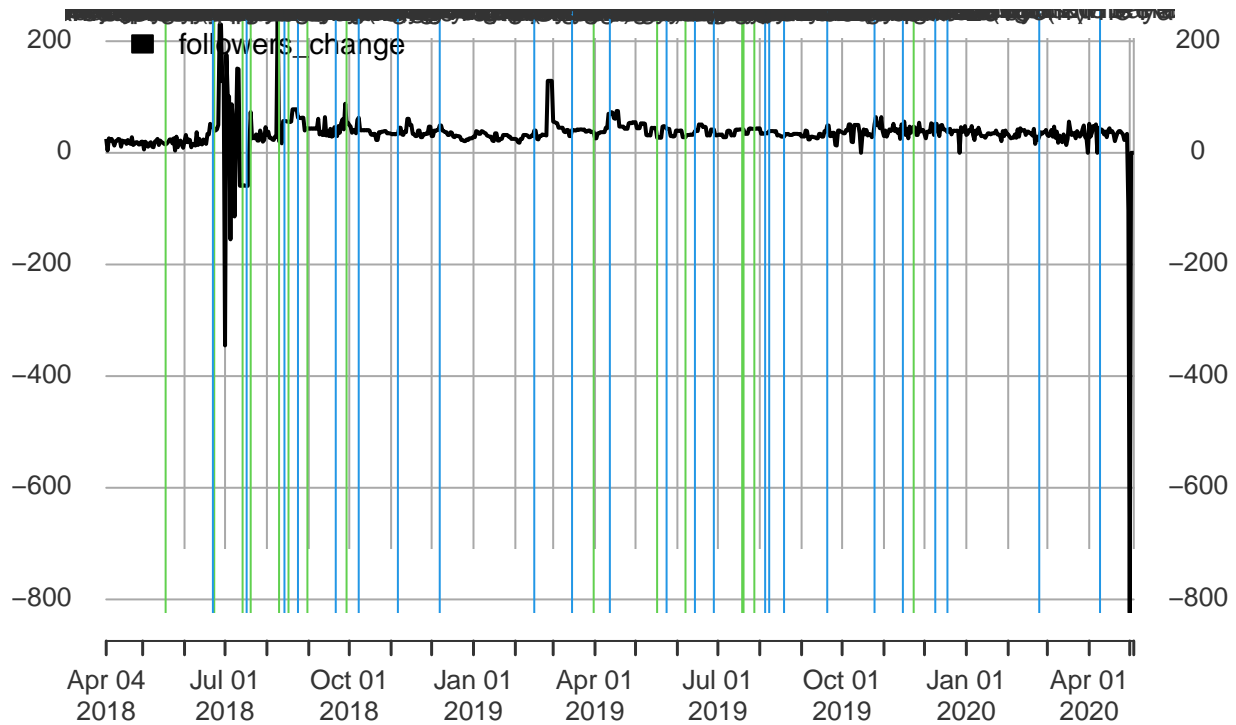
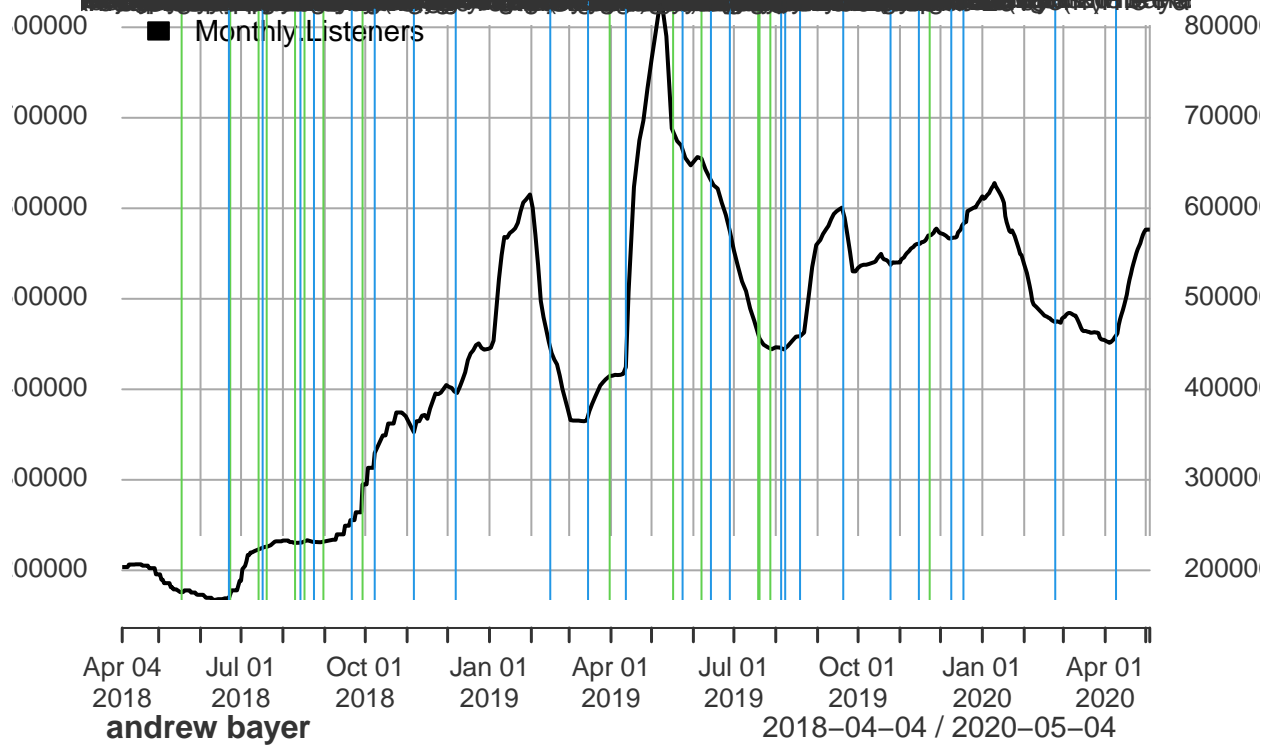
alessandra roncone

2018-04-07 / 2020-05-03



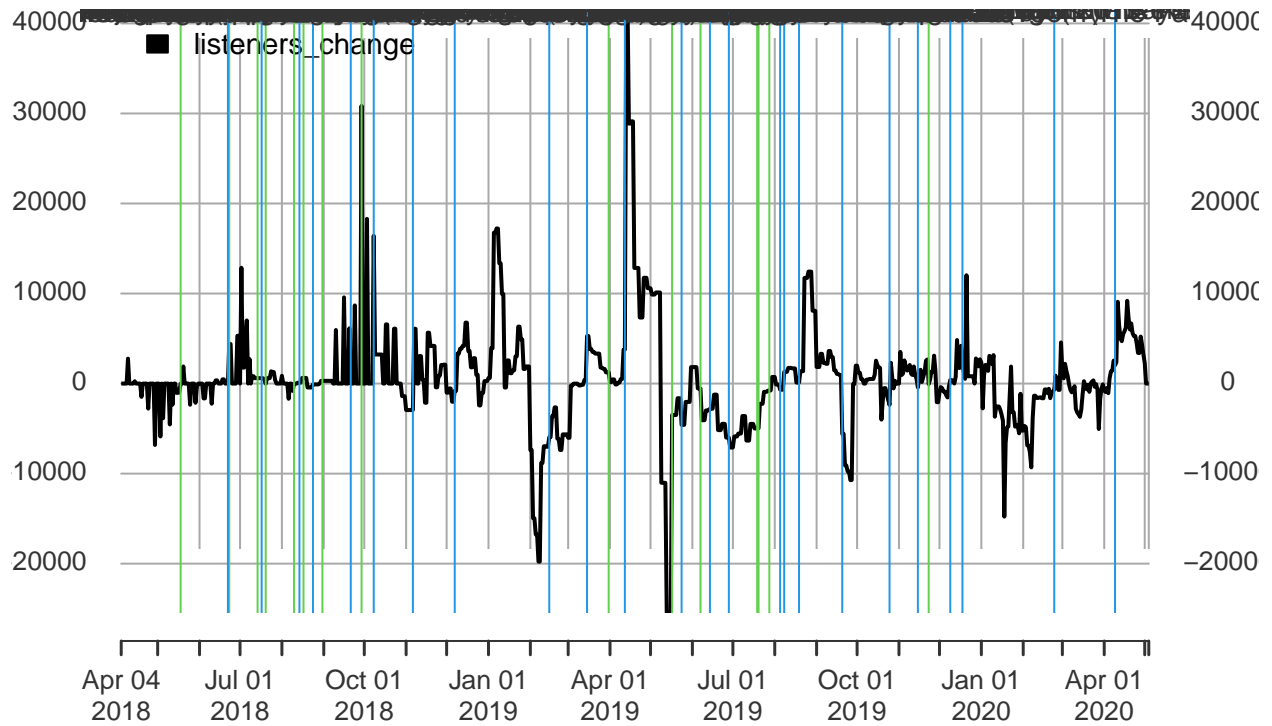
andrew bayer

2018-04-04 / 2020-05-04



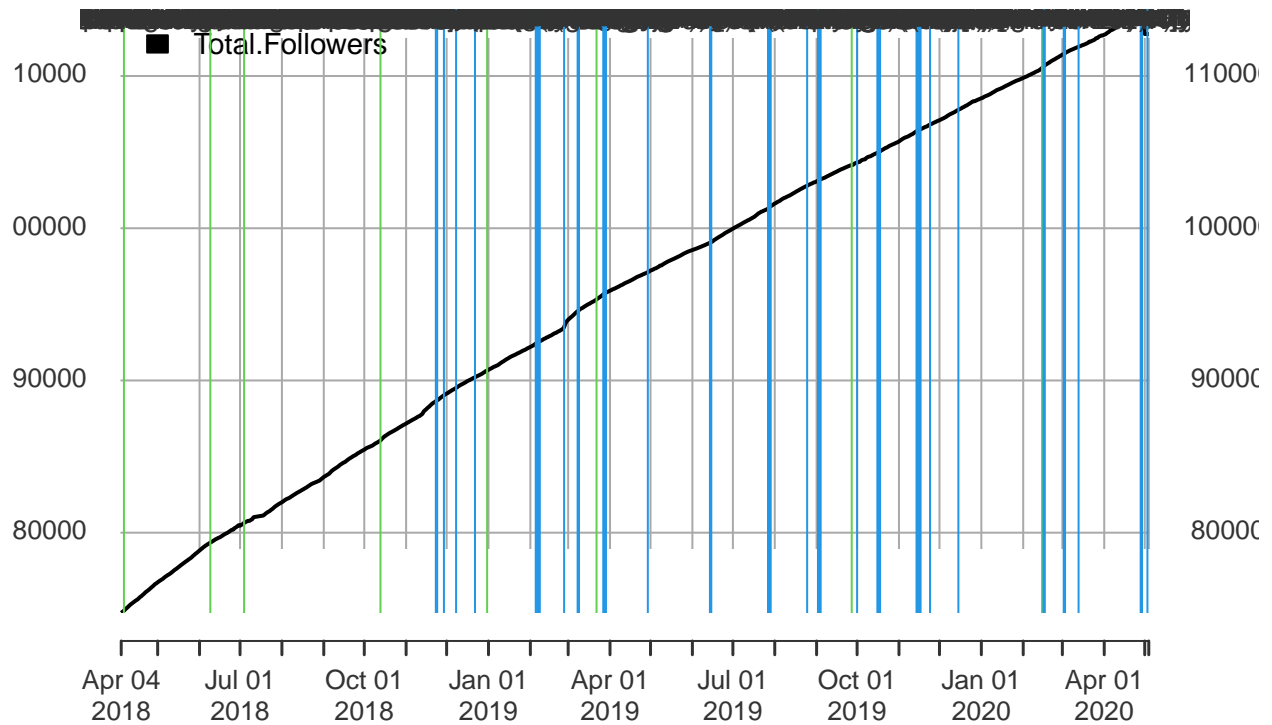
andrew bayer

2018-04-04 / 2020-05-04



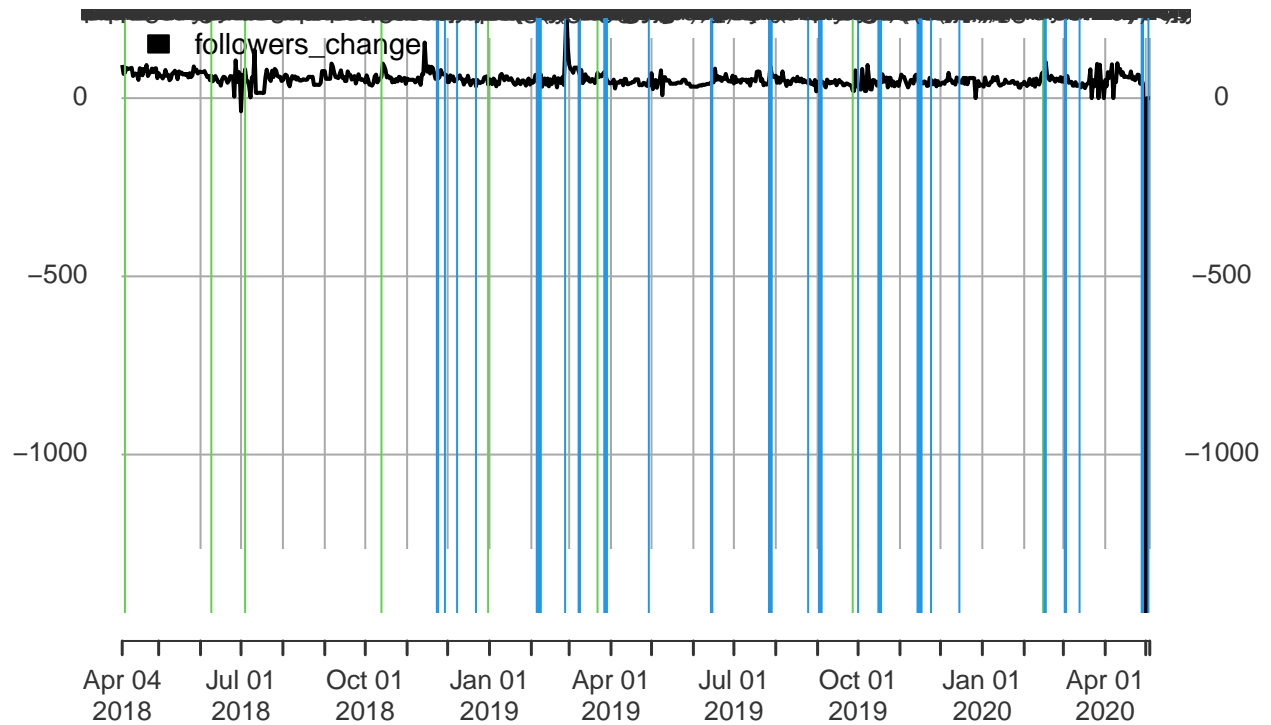
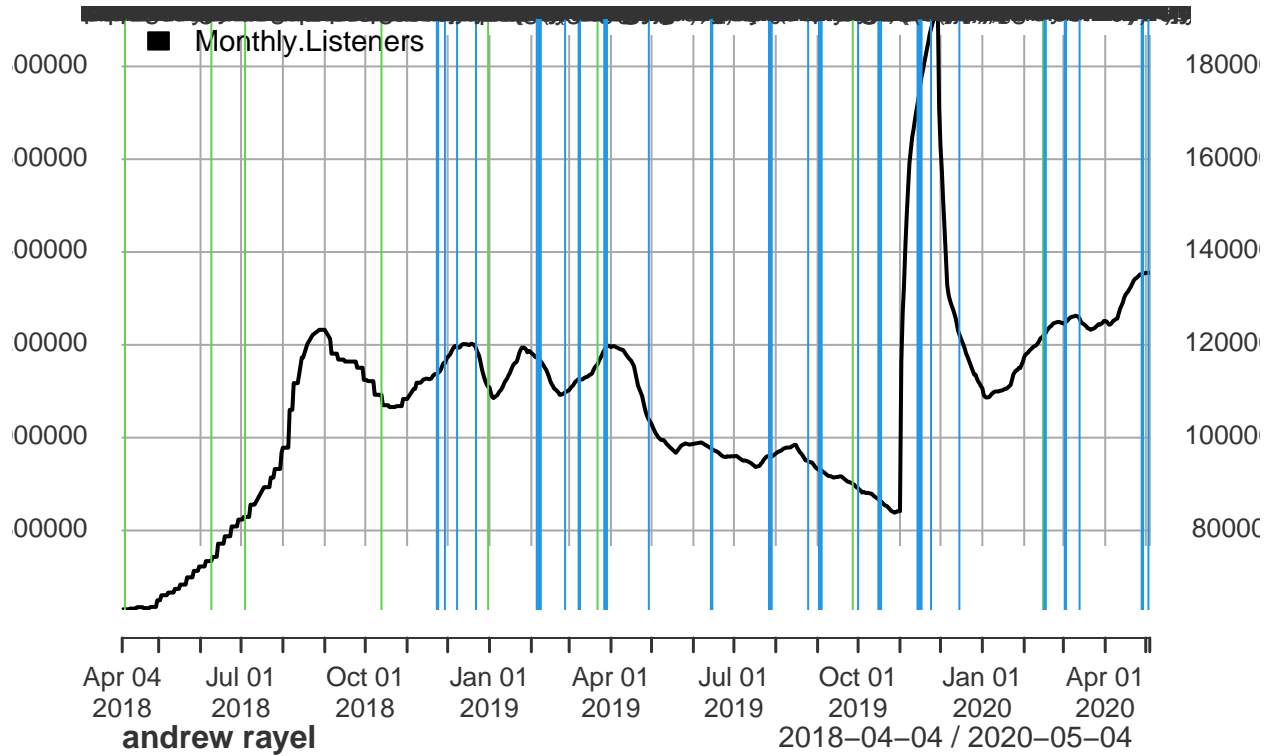
andrew rayel

2018-04-04 / 2020-05-04



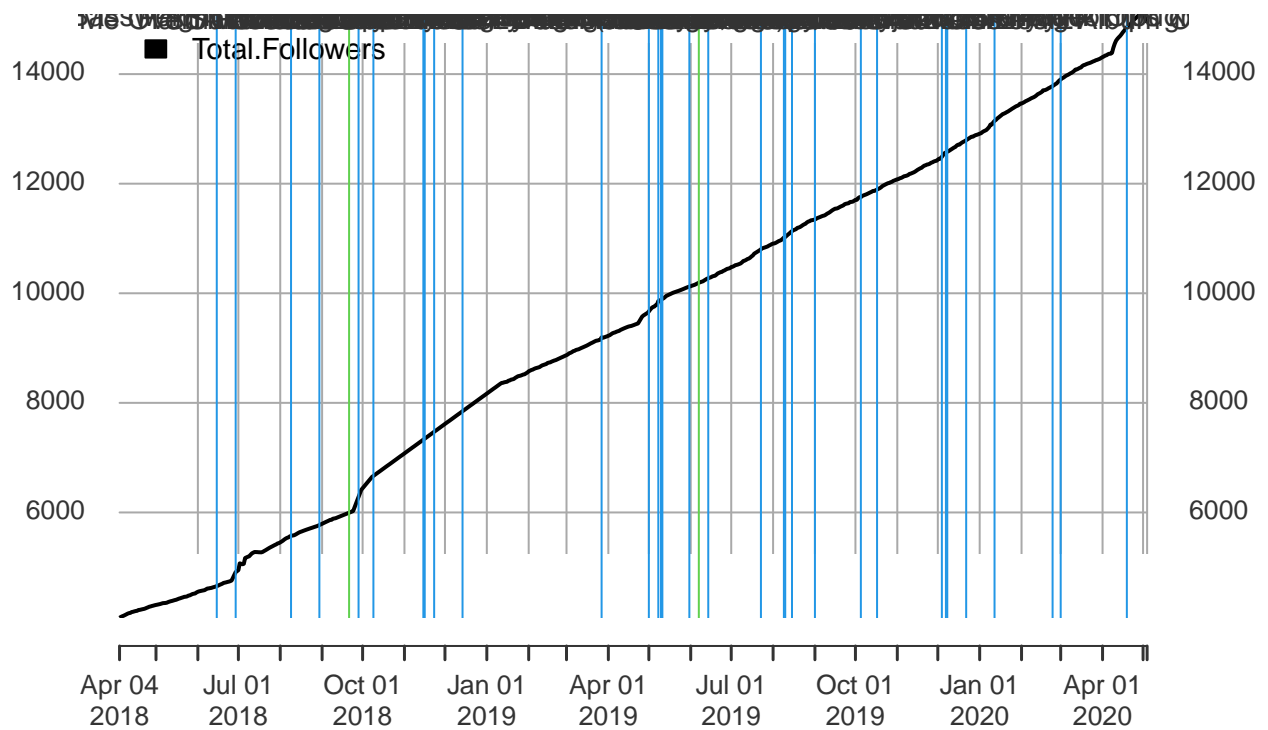
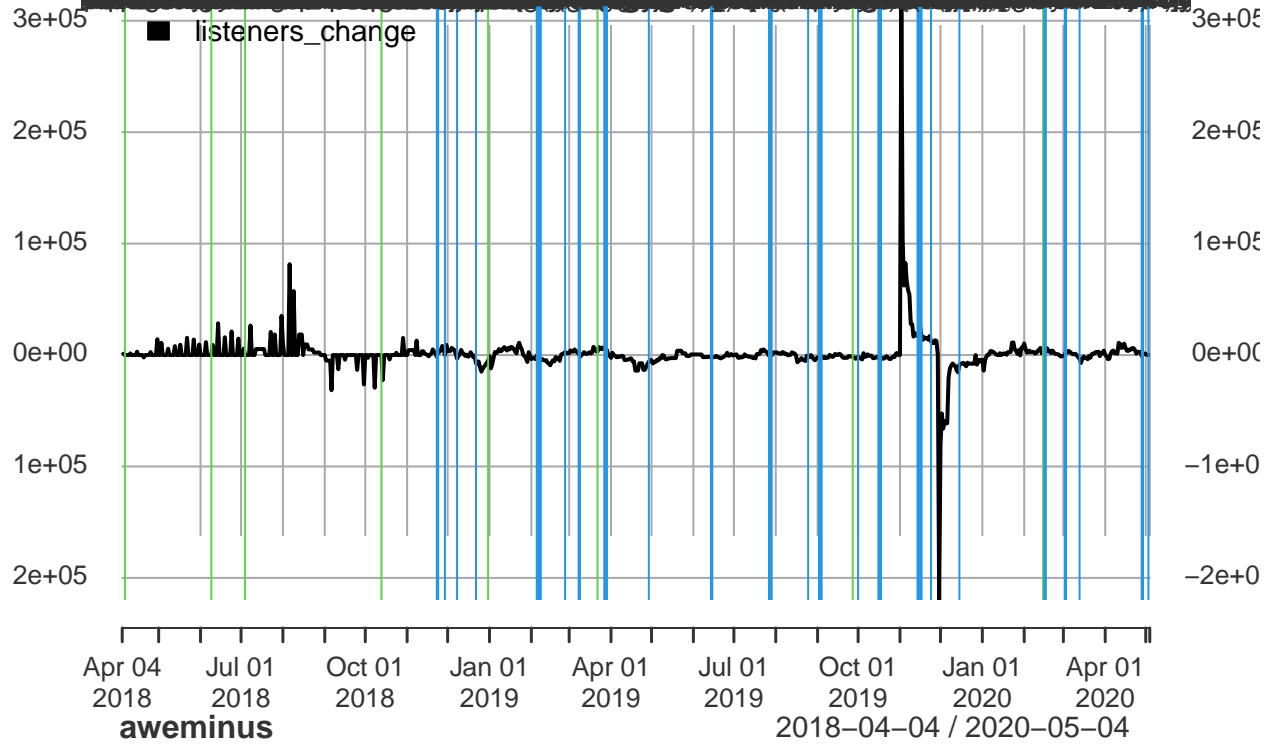
andrew rayel

2018-04-04 / 2020-05-04



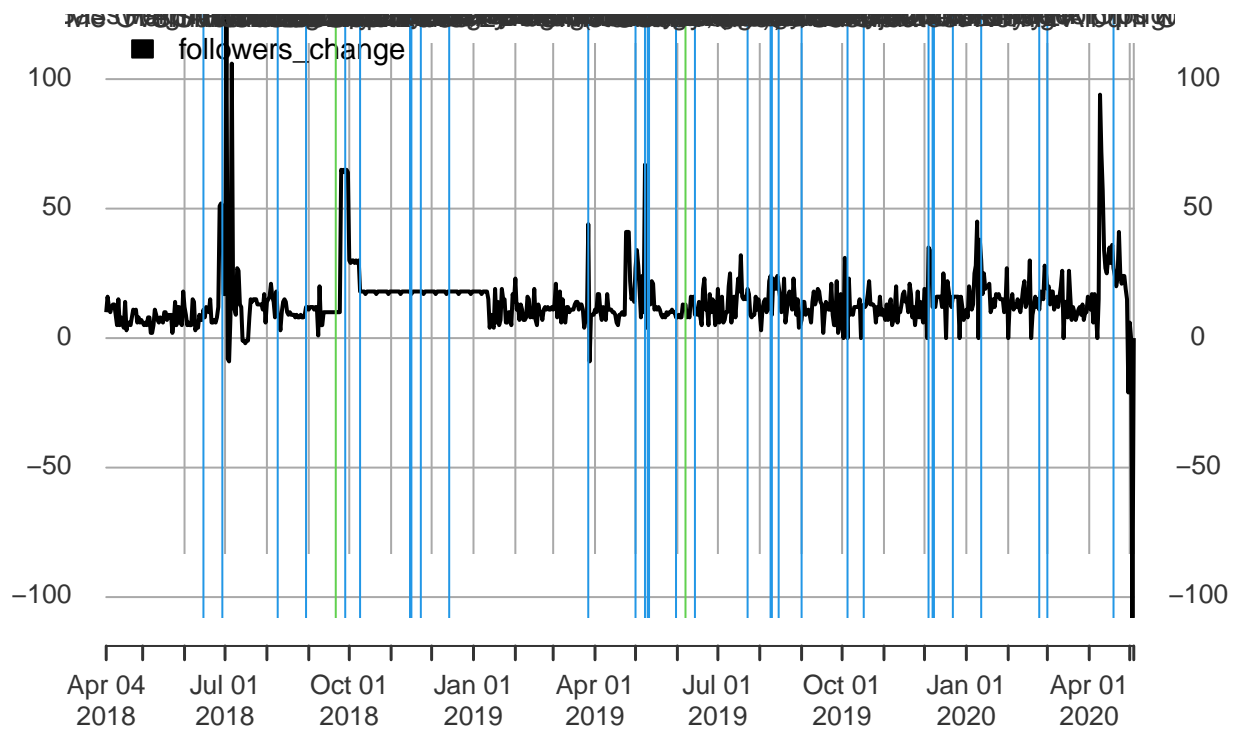
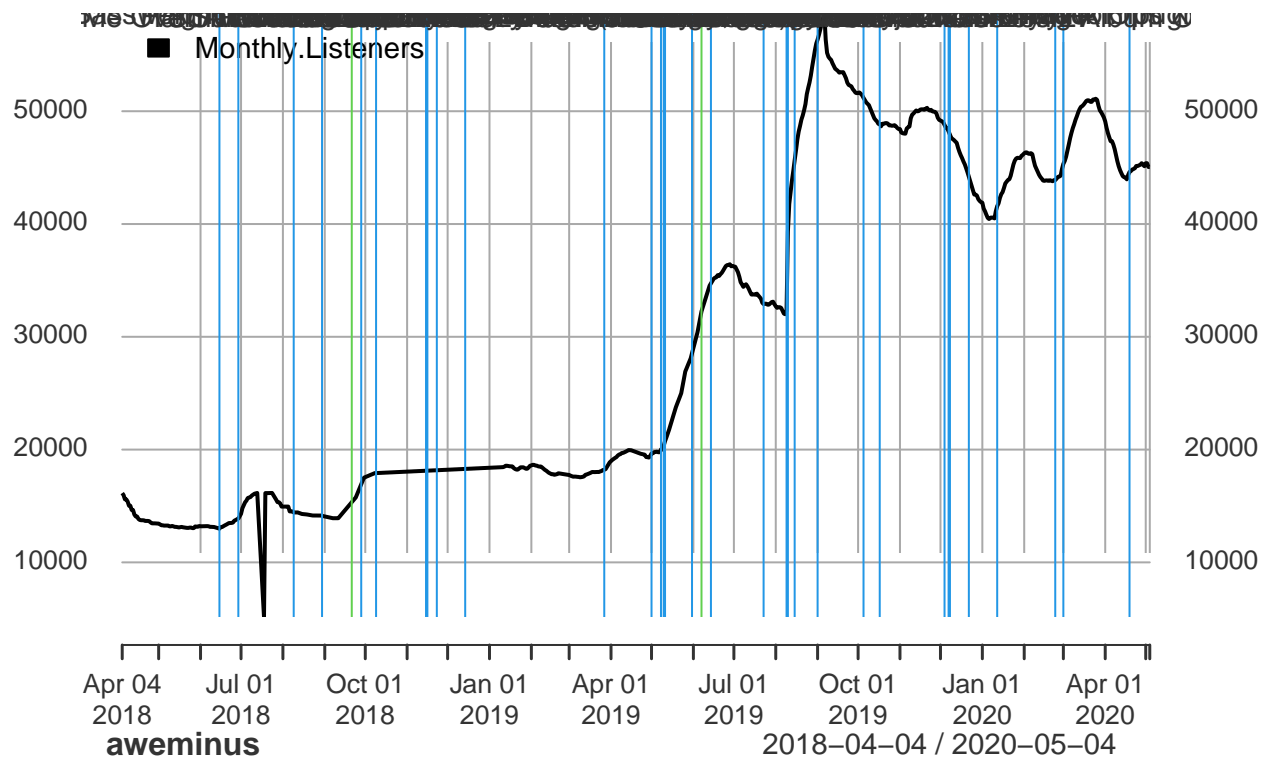
andrew rayel

2018-04-04 / 2020-05-04



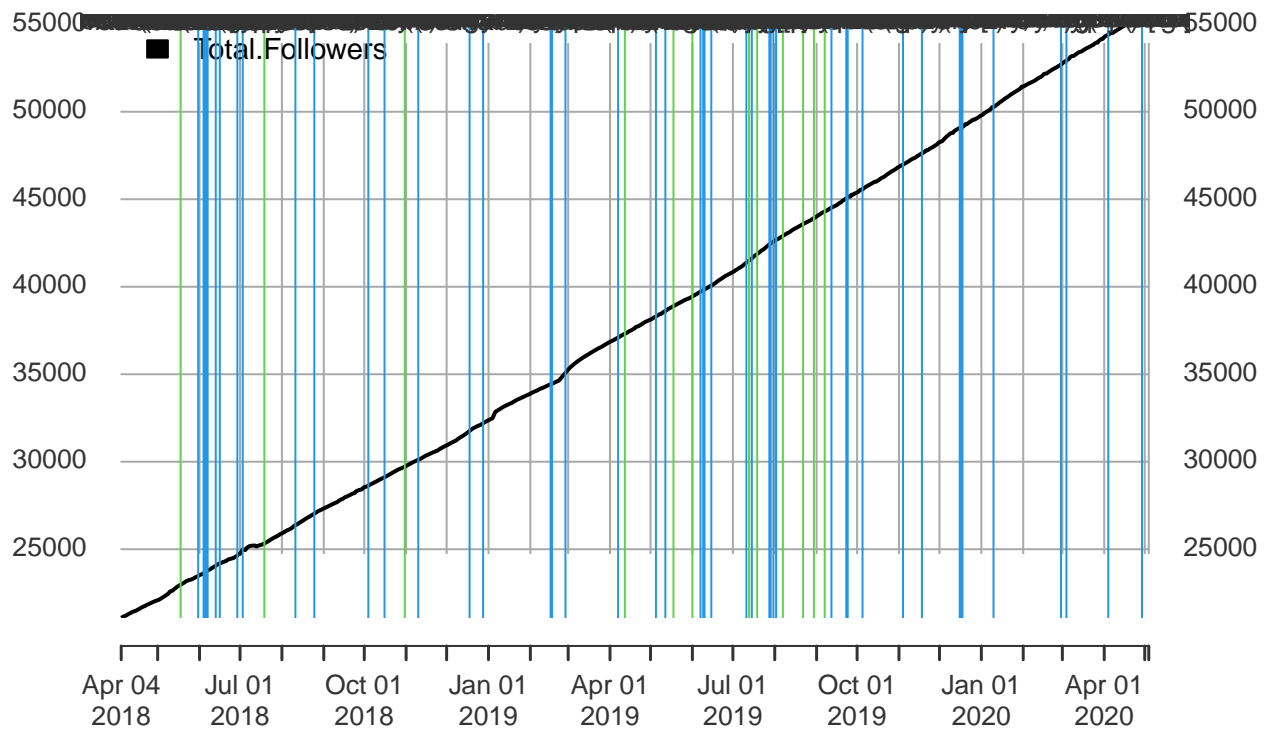
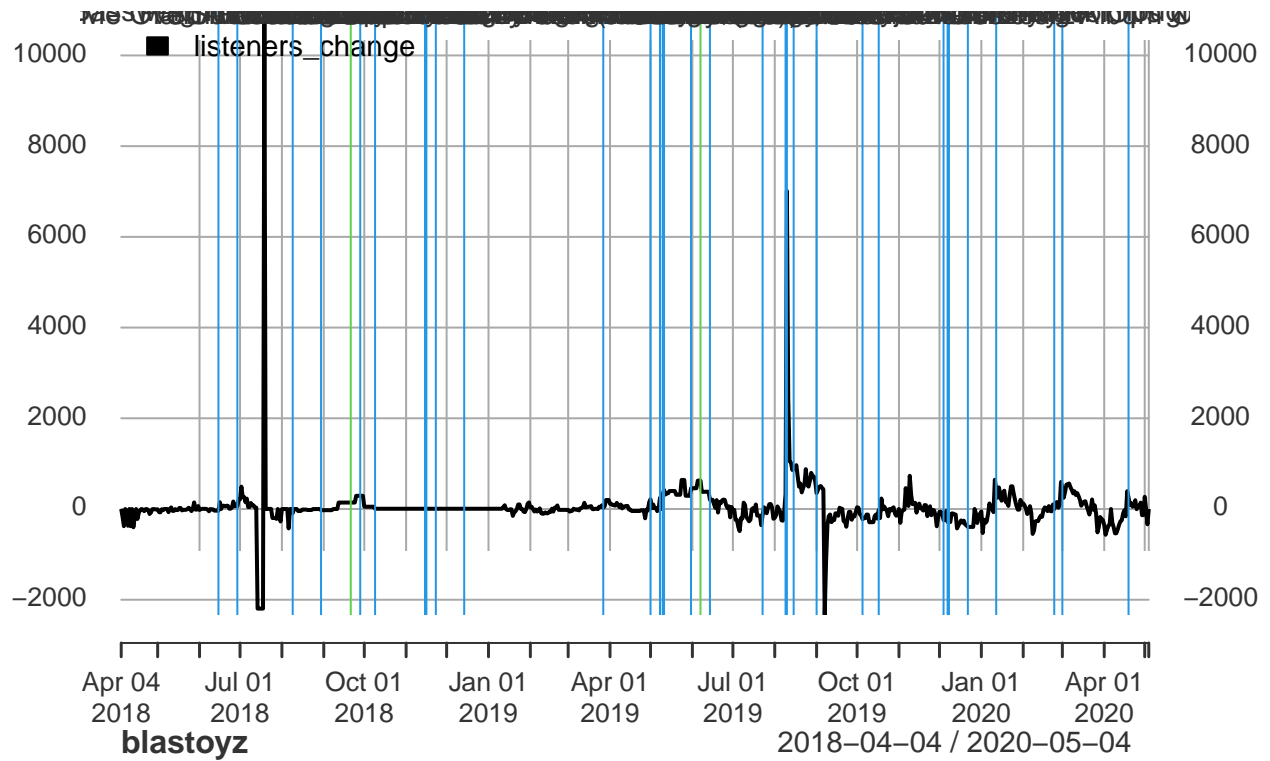
aweminus

2018-04-04 / 2020-05-04



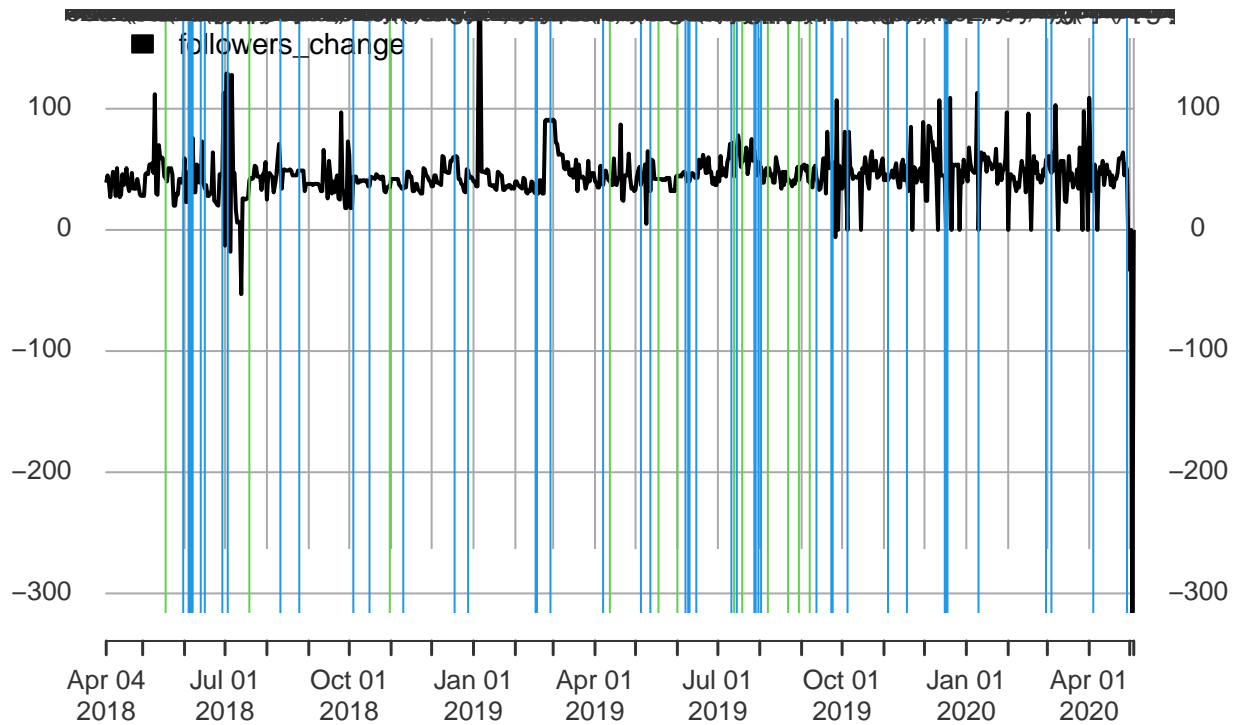
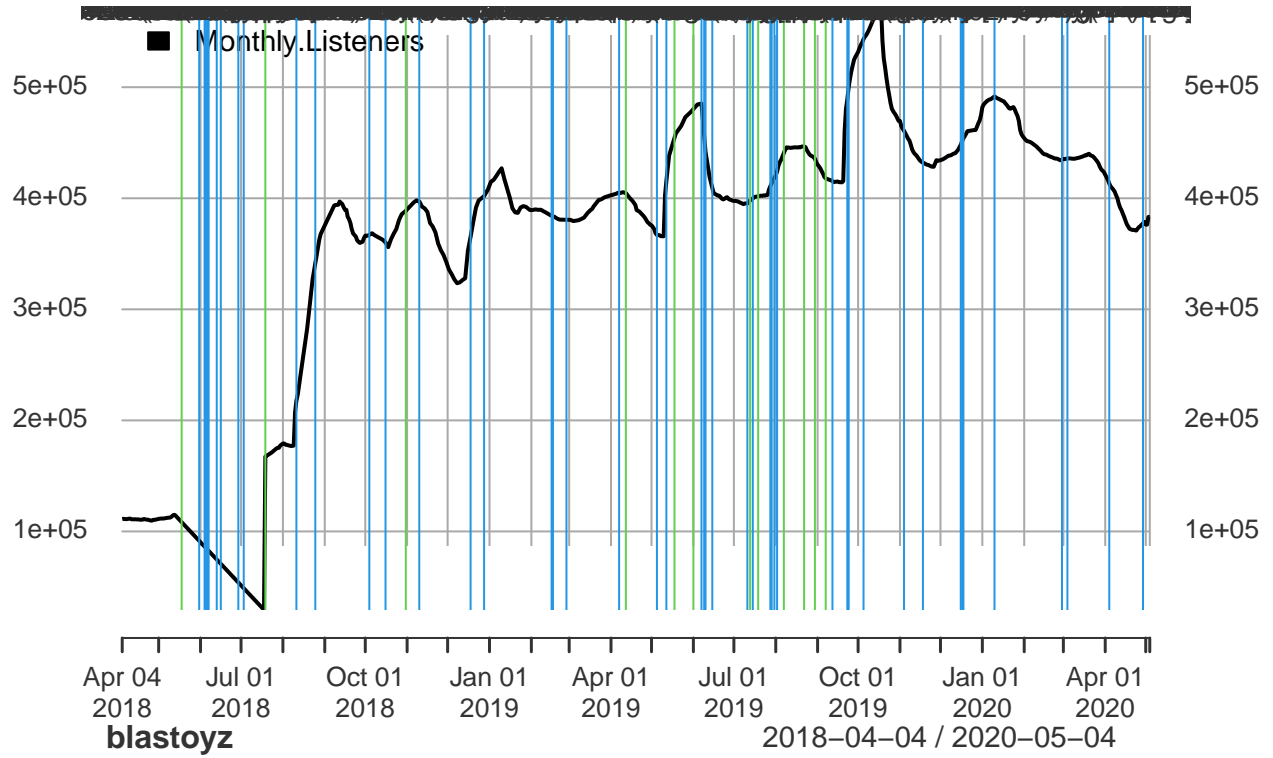
aweminus

2018-04-04 / 2020-05-04



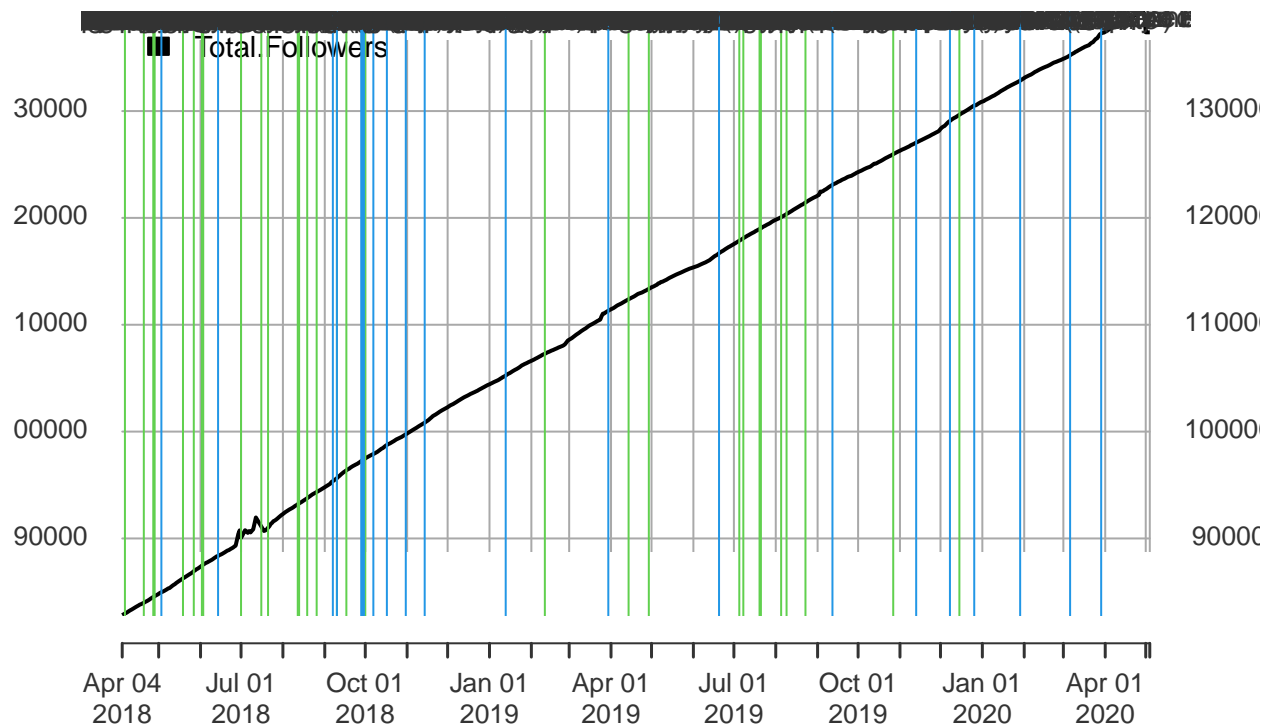
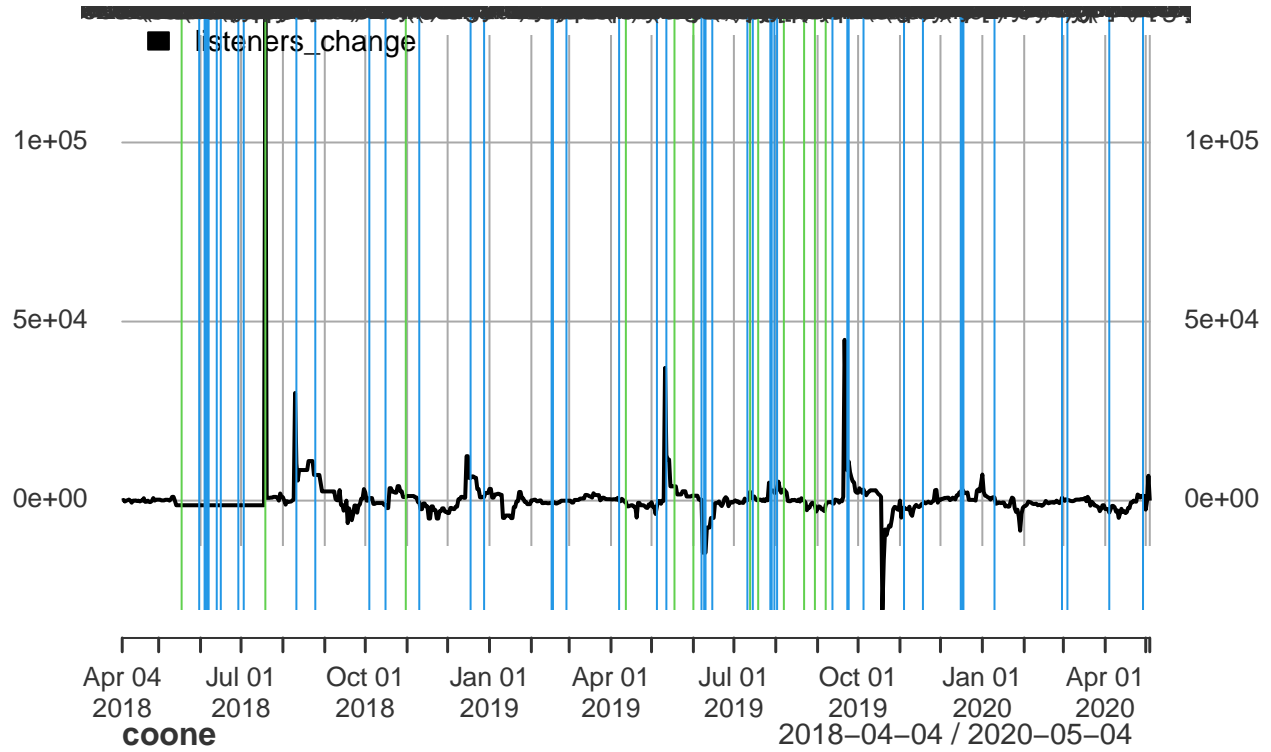
blastoyz

2018-04-04 / 2020-05-04



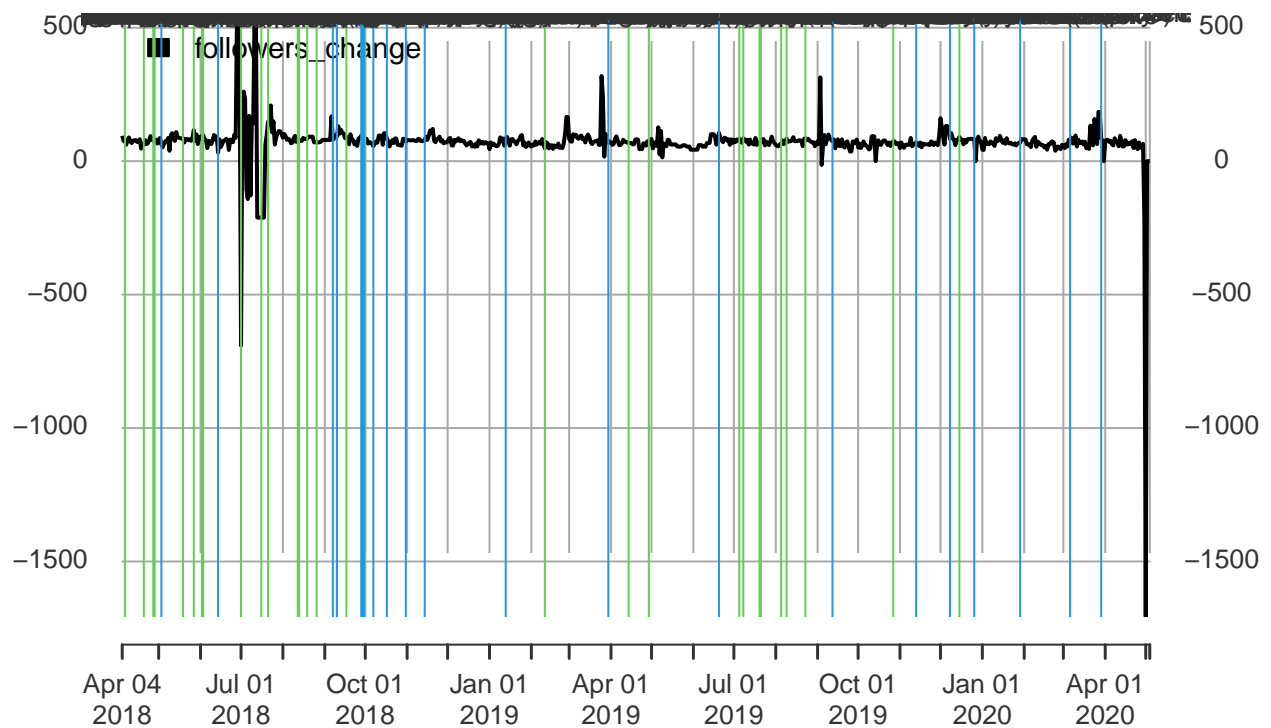
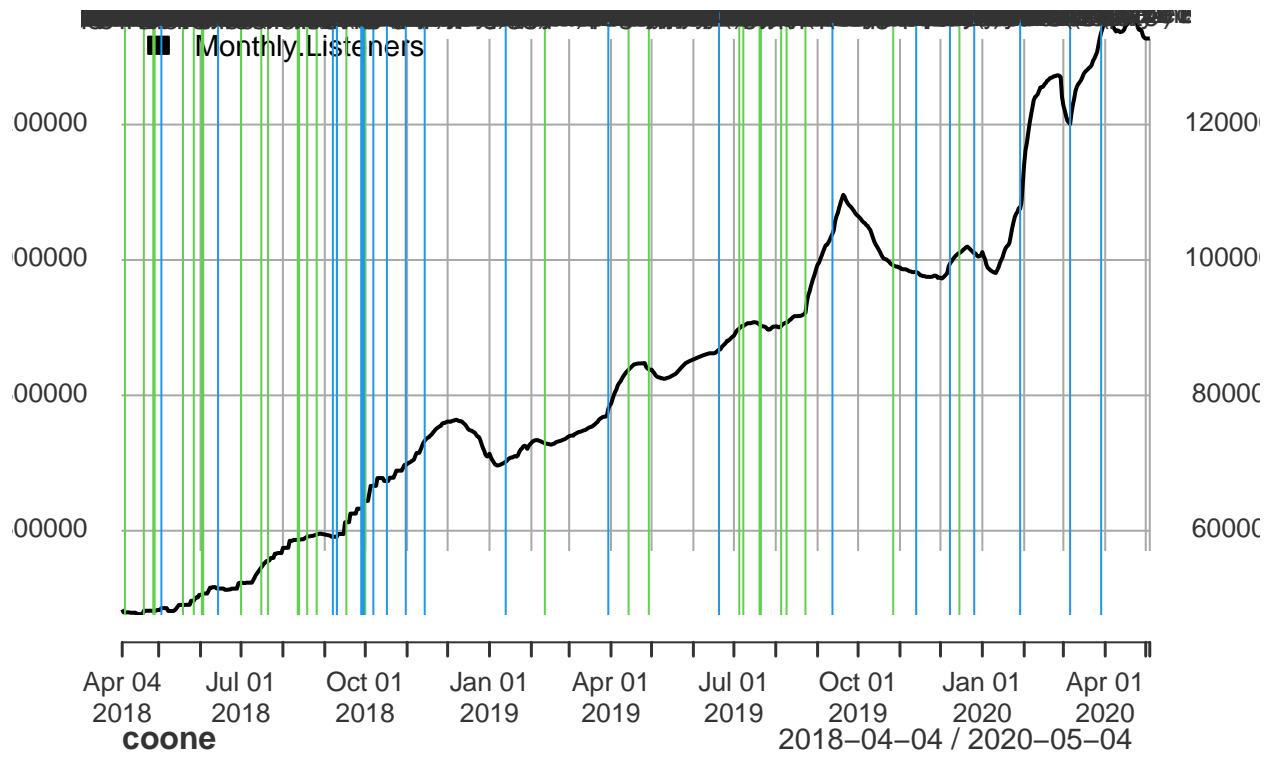
blastoyz

2018-04-04 / 2020-05-04



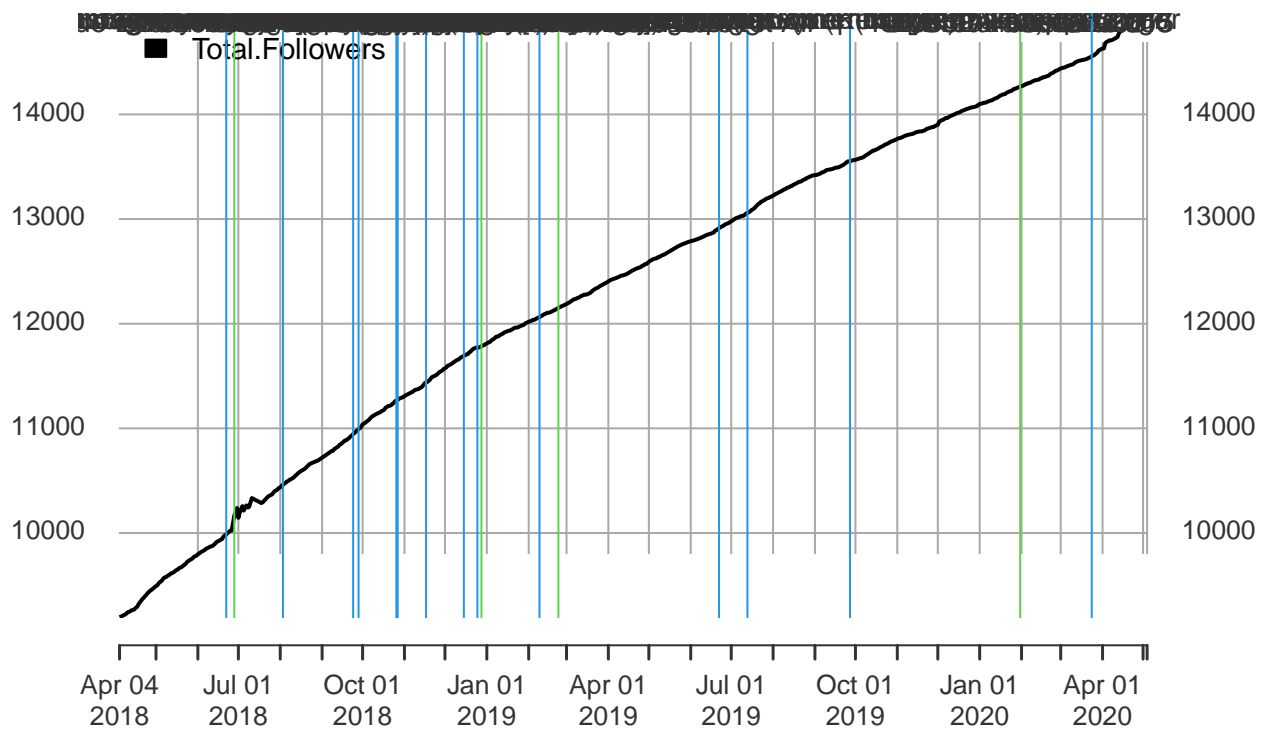
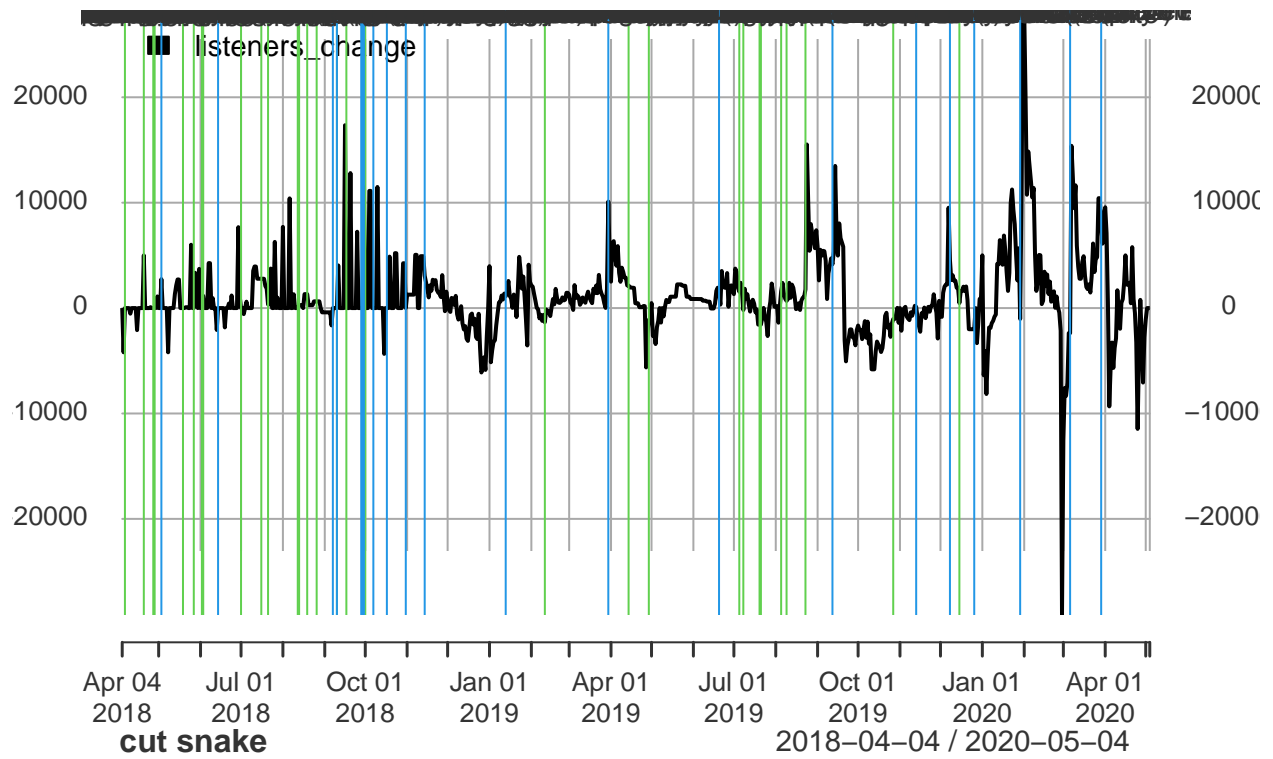
coone

2018-04-04 / 2020-05-04



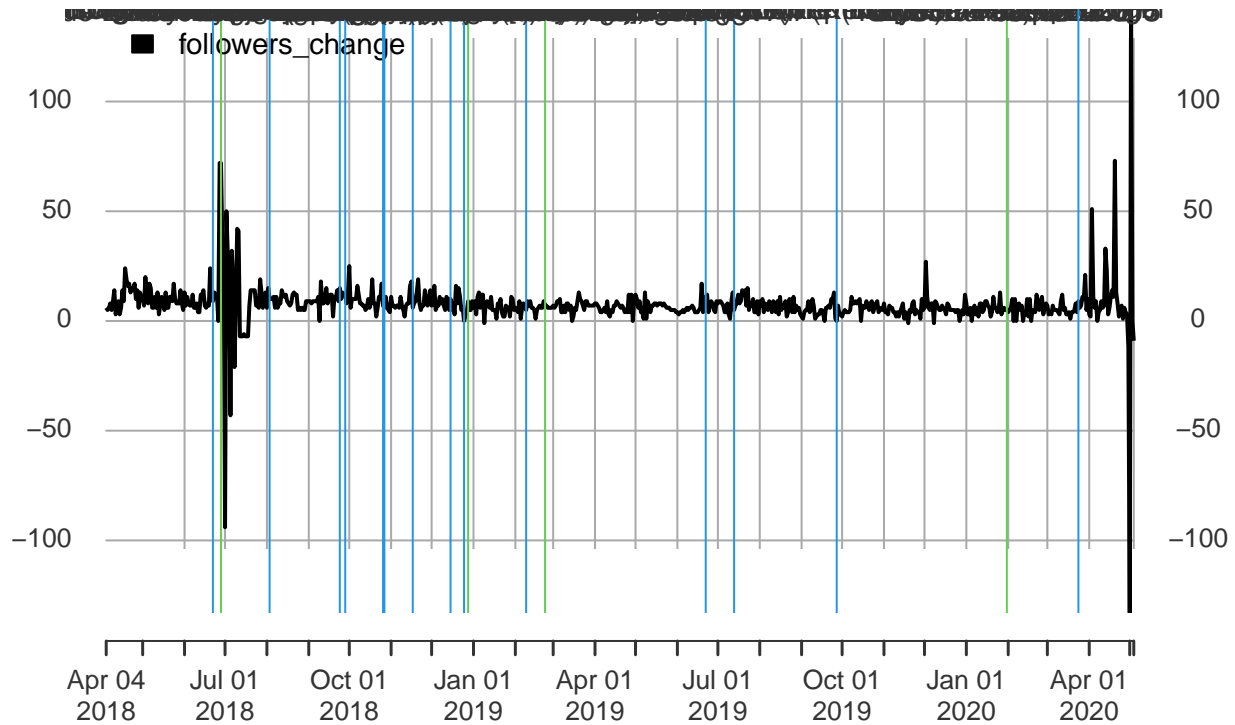
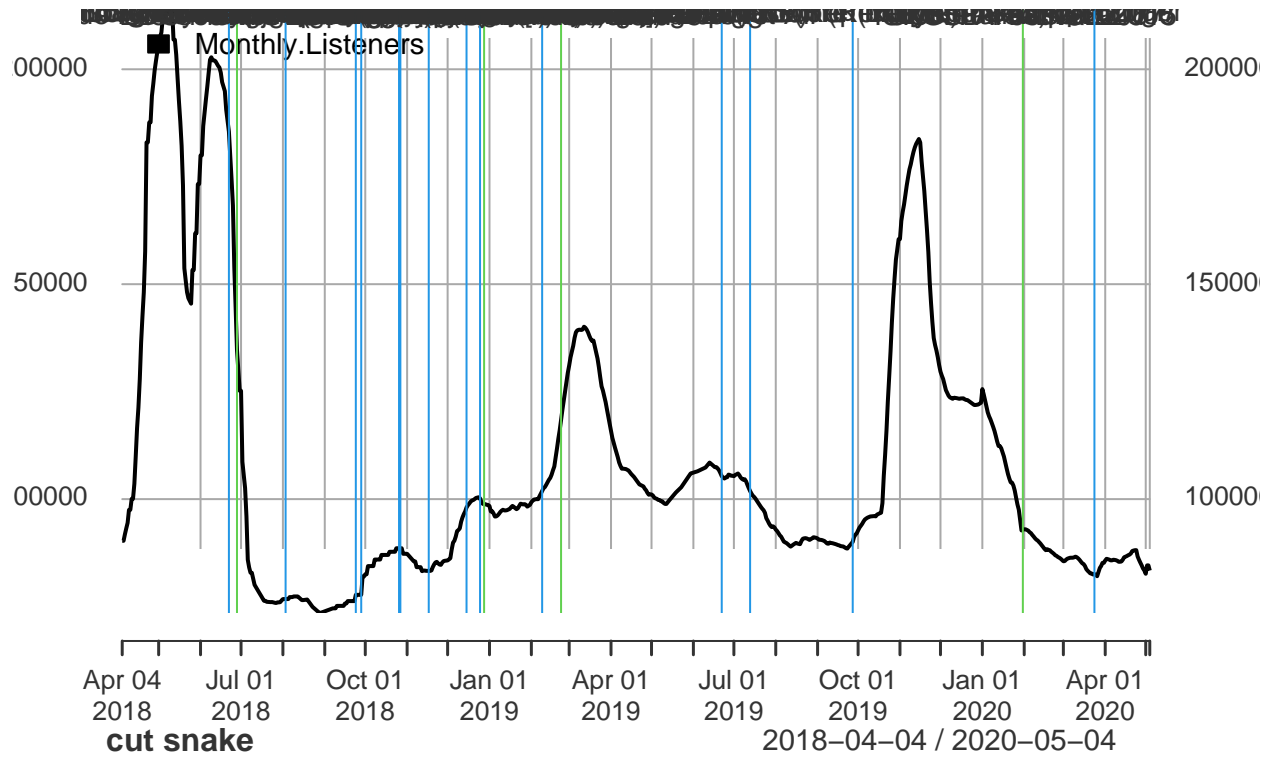
coone

2018-04-04 / 2020-05-04



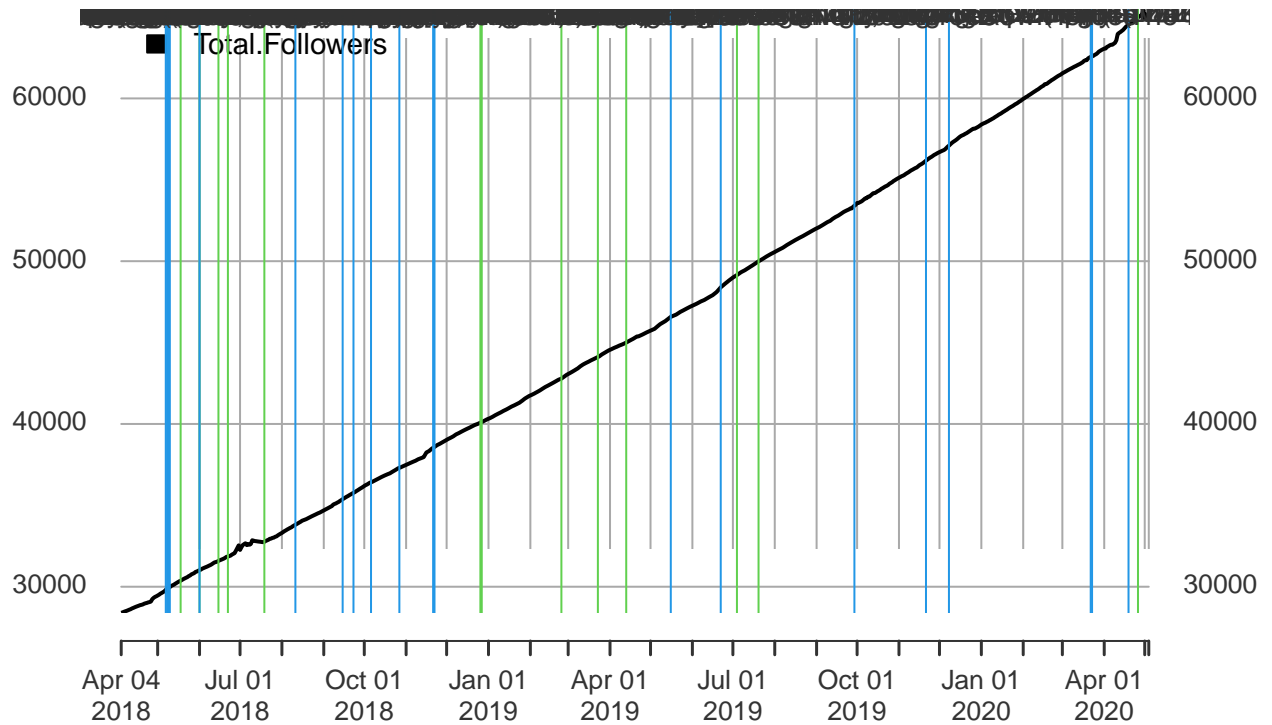
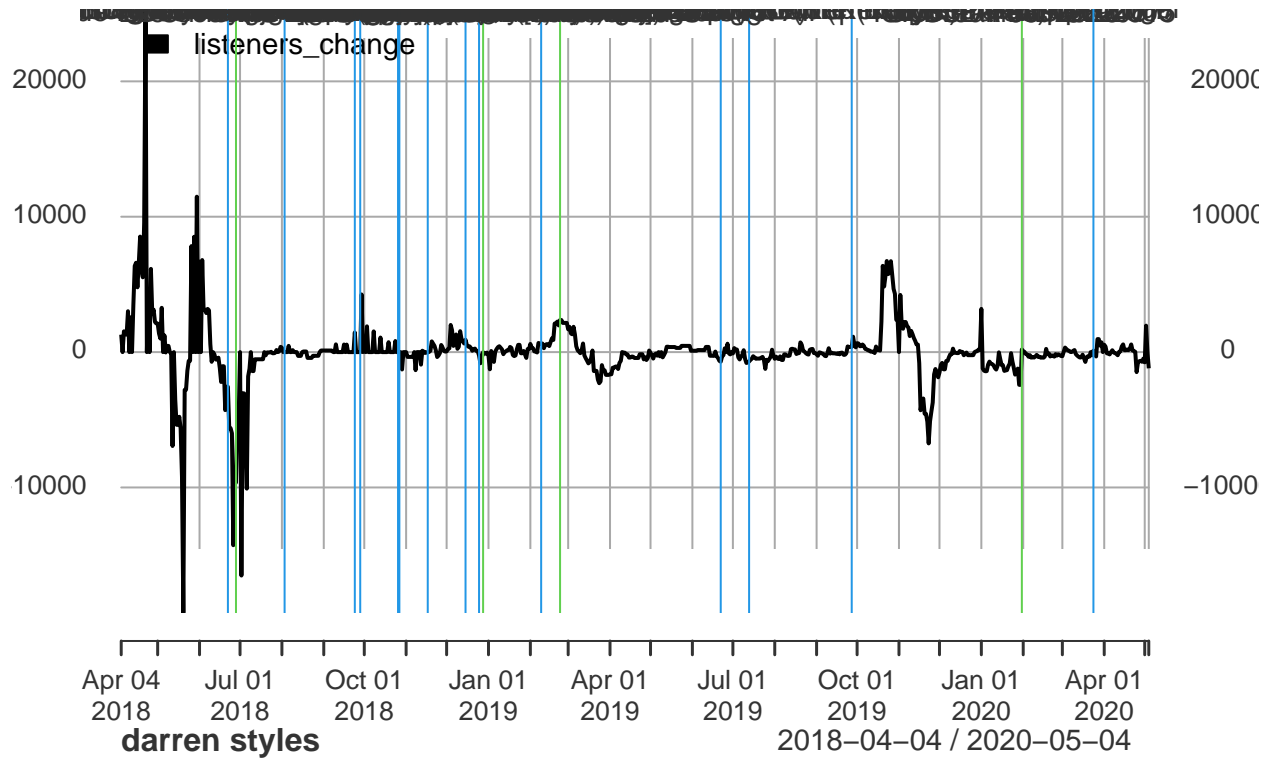
cut snake

2018-04-04 / 2020-05-04



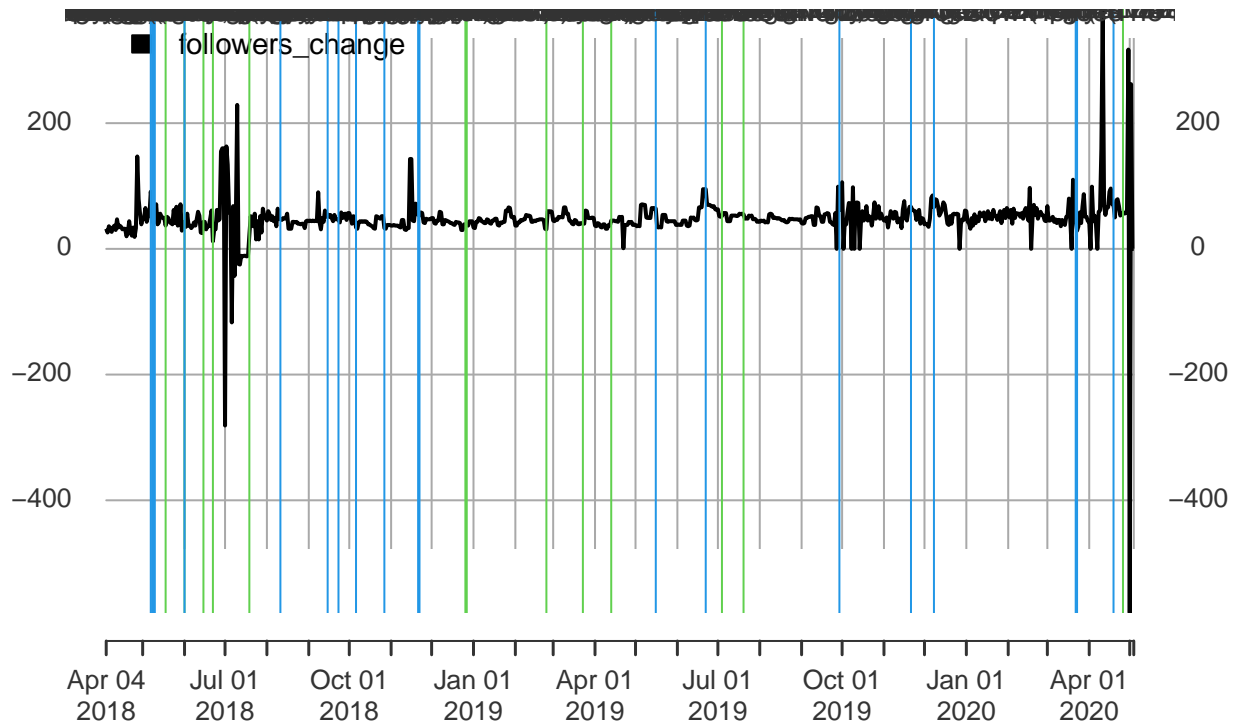
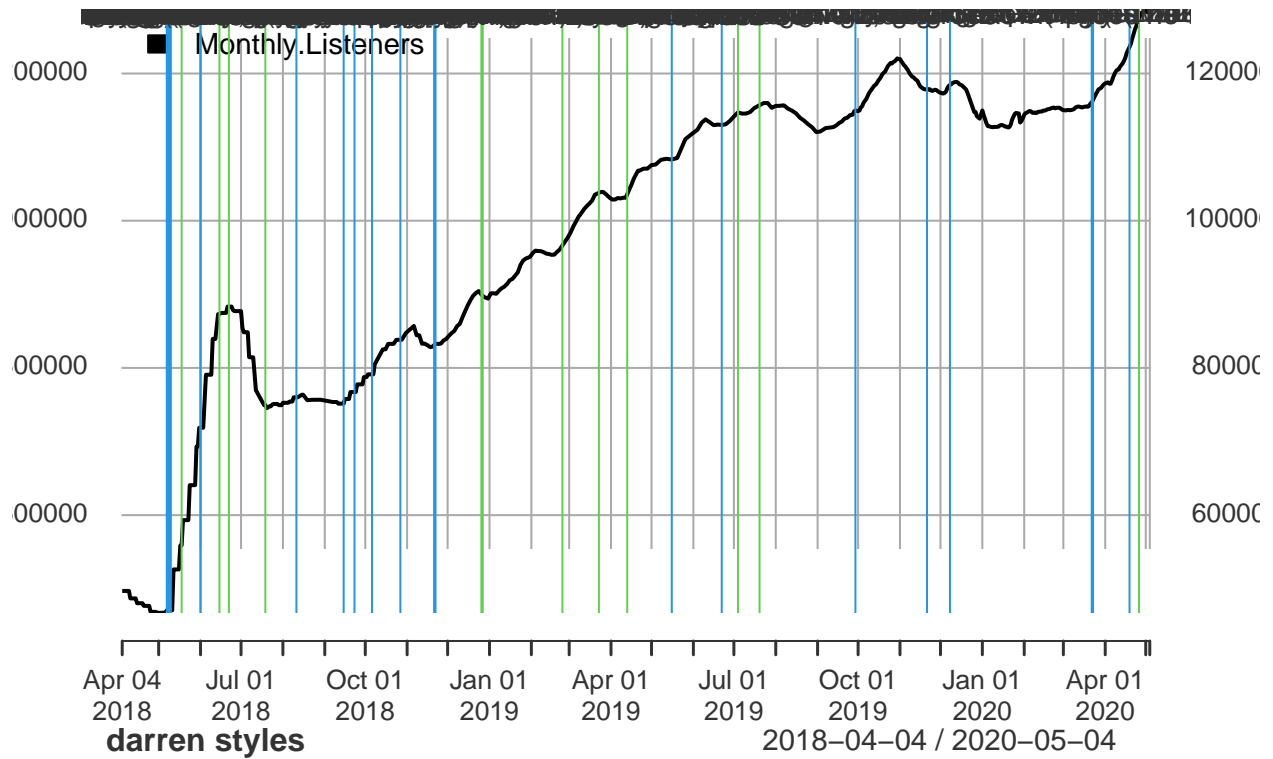
cut snake

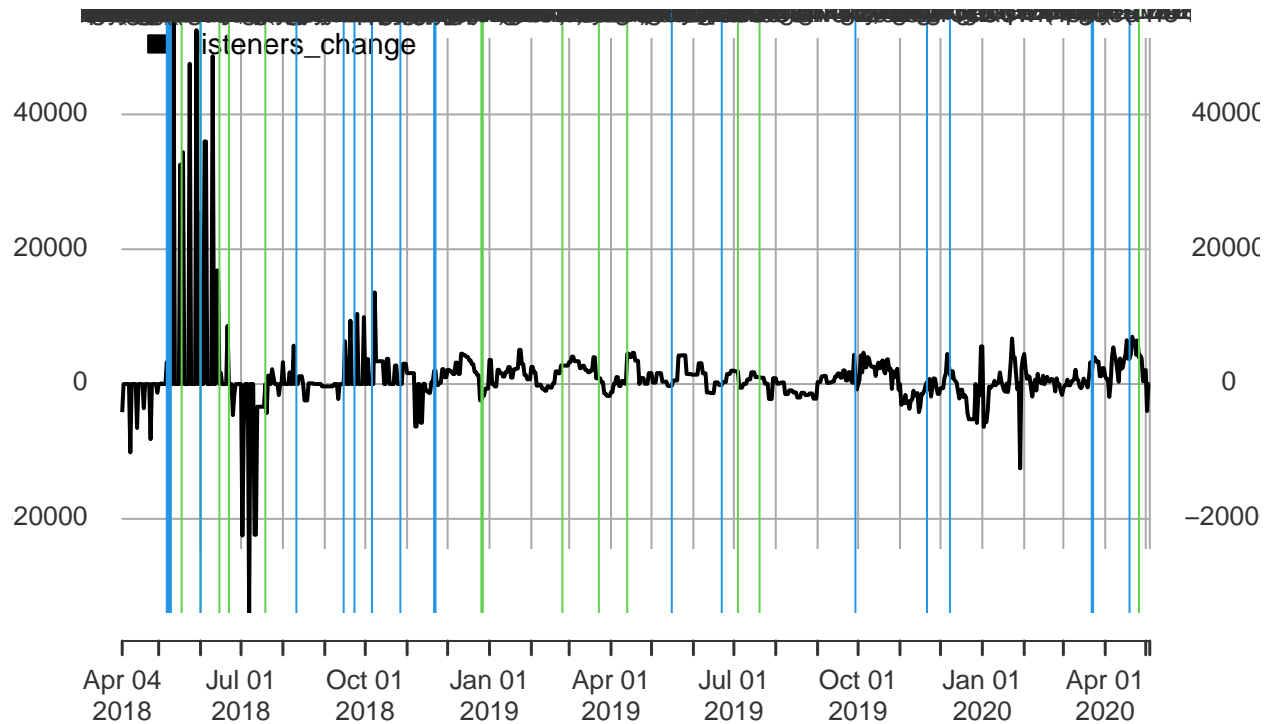
2018-04-04 / 2020-05-04



darren styles

2018-04-04 / 2020-05-04



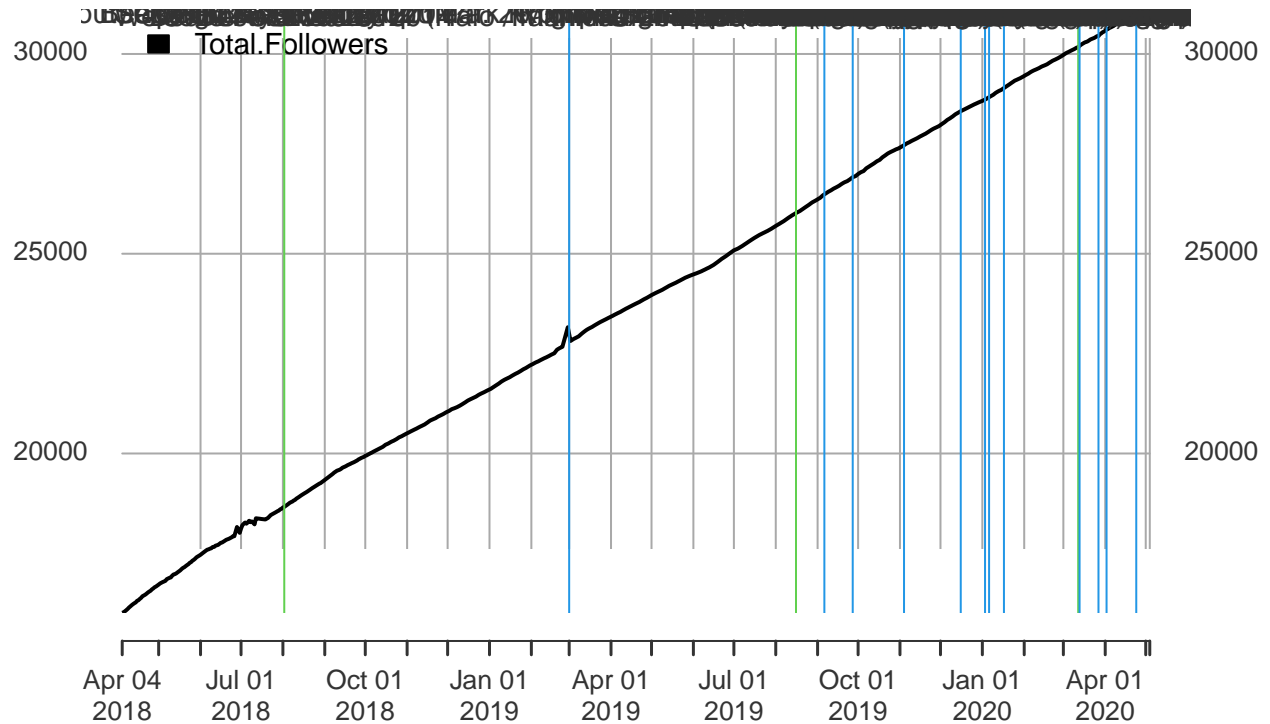


```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <80>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <a6>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : font metrics unknown for Unicode character U+2026
```

```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <e2>

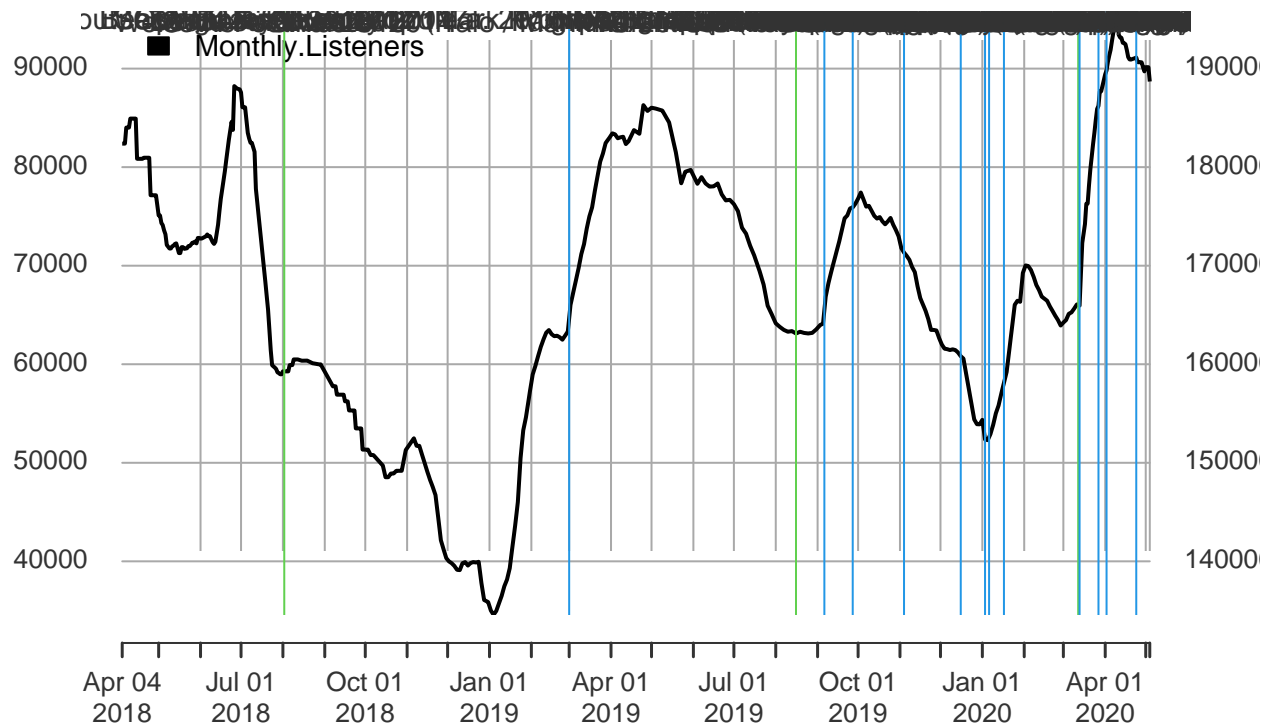
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <80>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <a6>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : font metrics unknown for Unicode character U+2026
```

dc breaks

2018-04-04 / 2020-05-04



```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <e2>

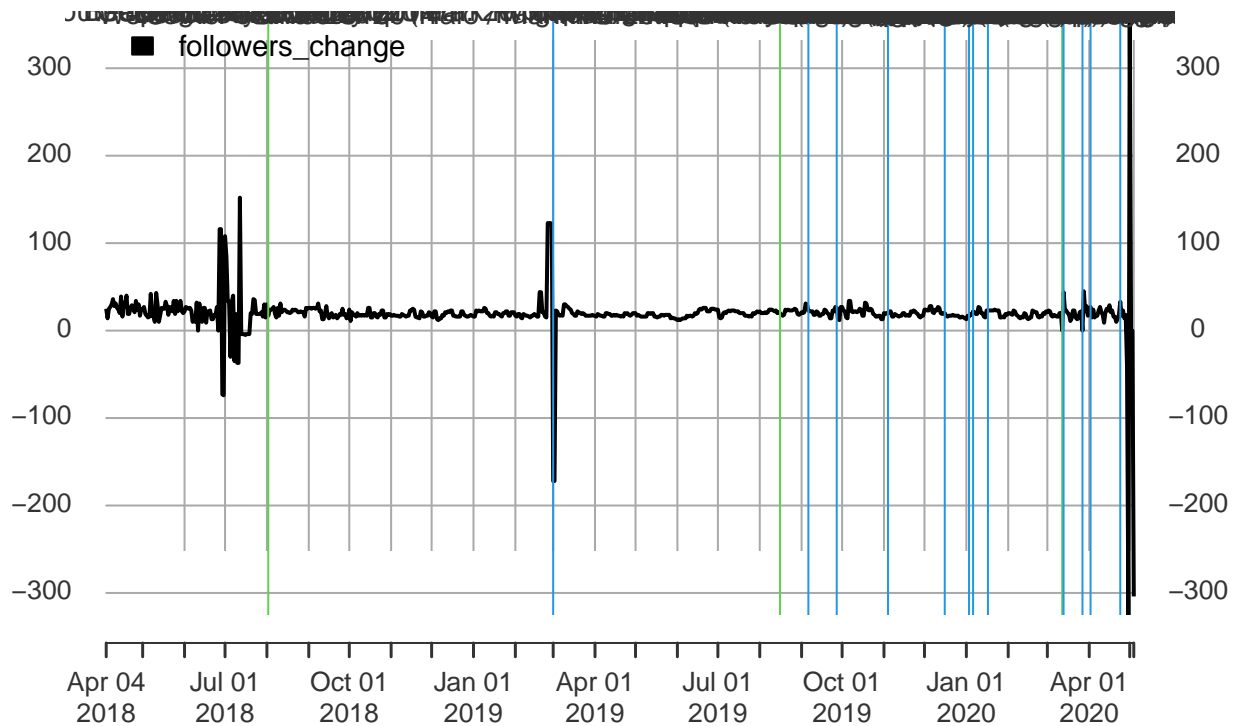
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <80>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <a6>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : font metrics unknown for Unicode character U+2026
```

dc breaks

2018-04-04 / 2020-05-04



```
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <e2>

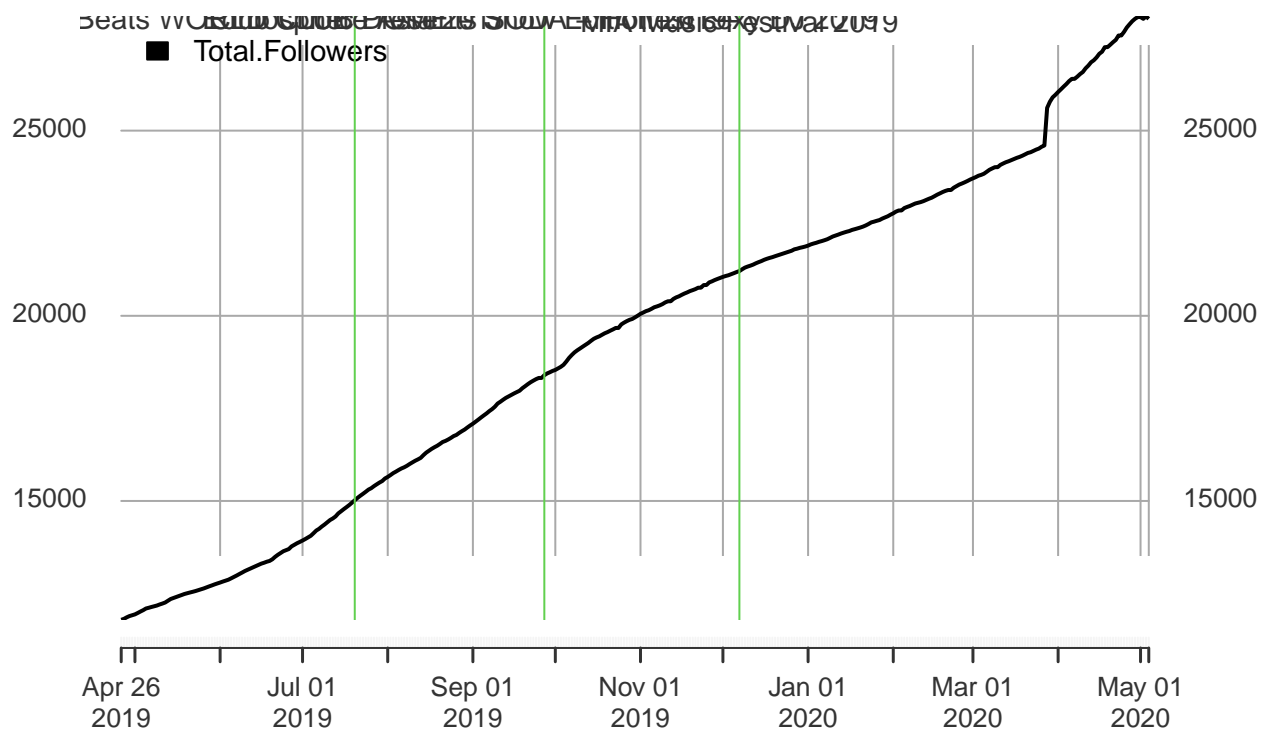
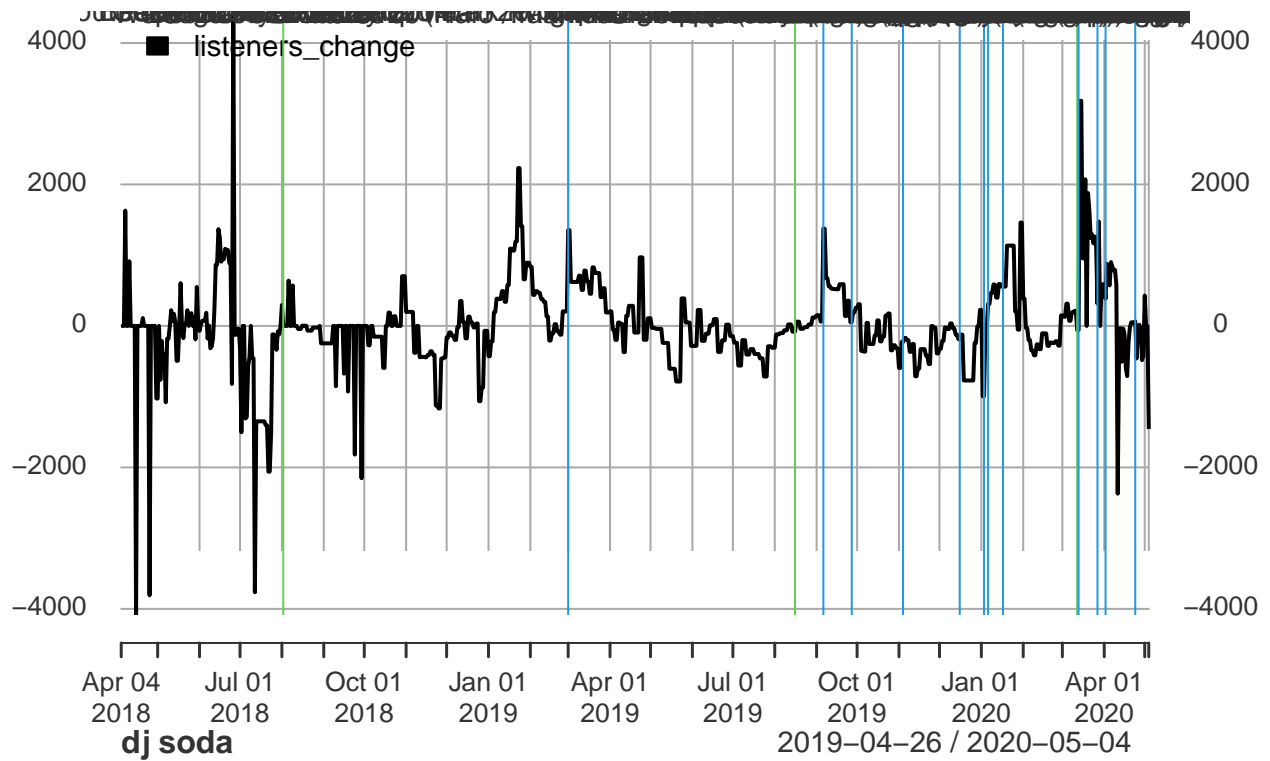
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <80>

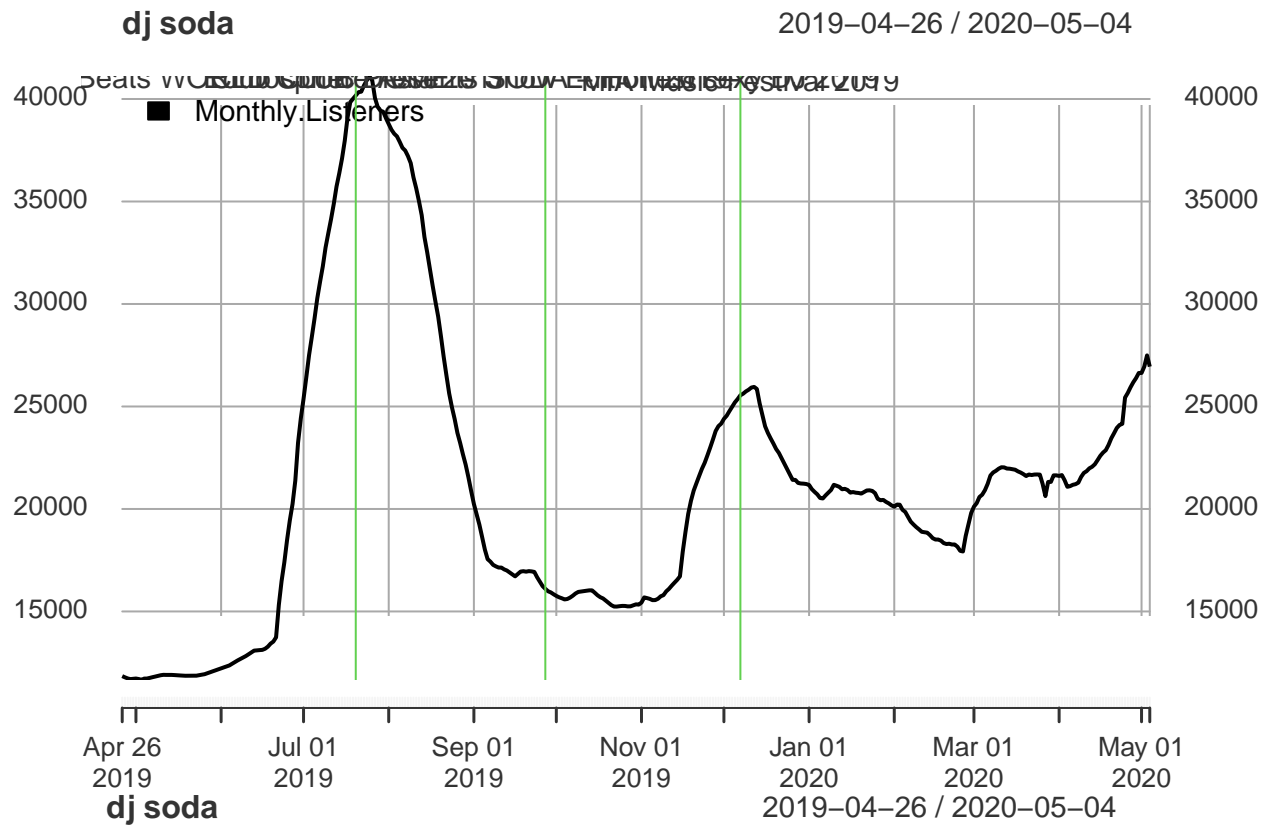
## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : conversion failure on 'We Love... Coloursound End of Term
## Festival 2011' in 'mbcsToSbcs': dot substituted for <a6>

## Warning in text.default(x = x$Env$xycoords$x[event.ind], y = ypos, labels =
## as.character(events[, : font metrics unknown for Unicode character U+2026
```

dc breaks

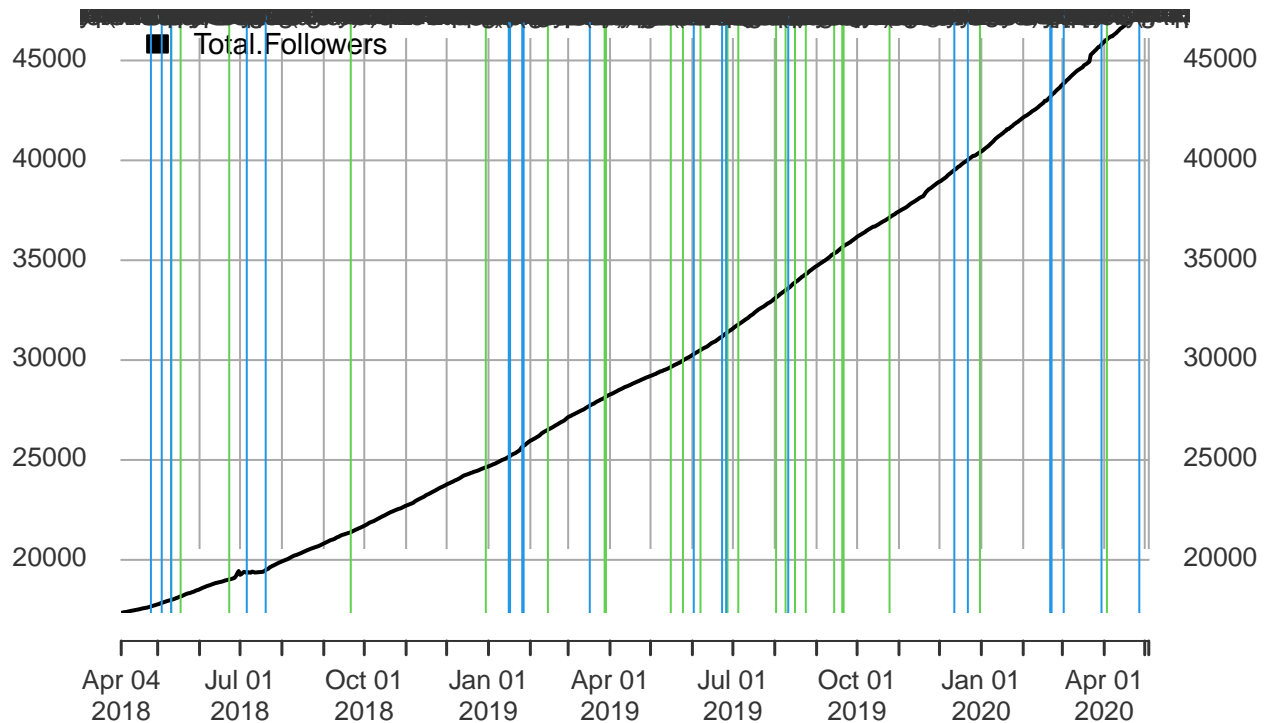
2018-04-04 / 2020-05-04





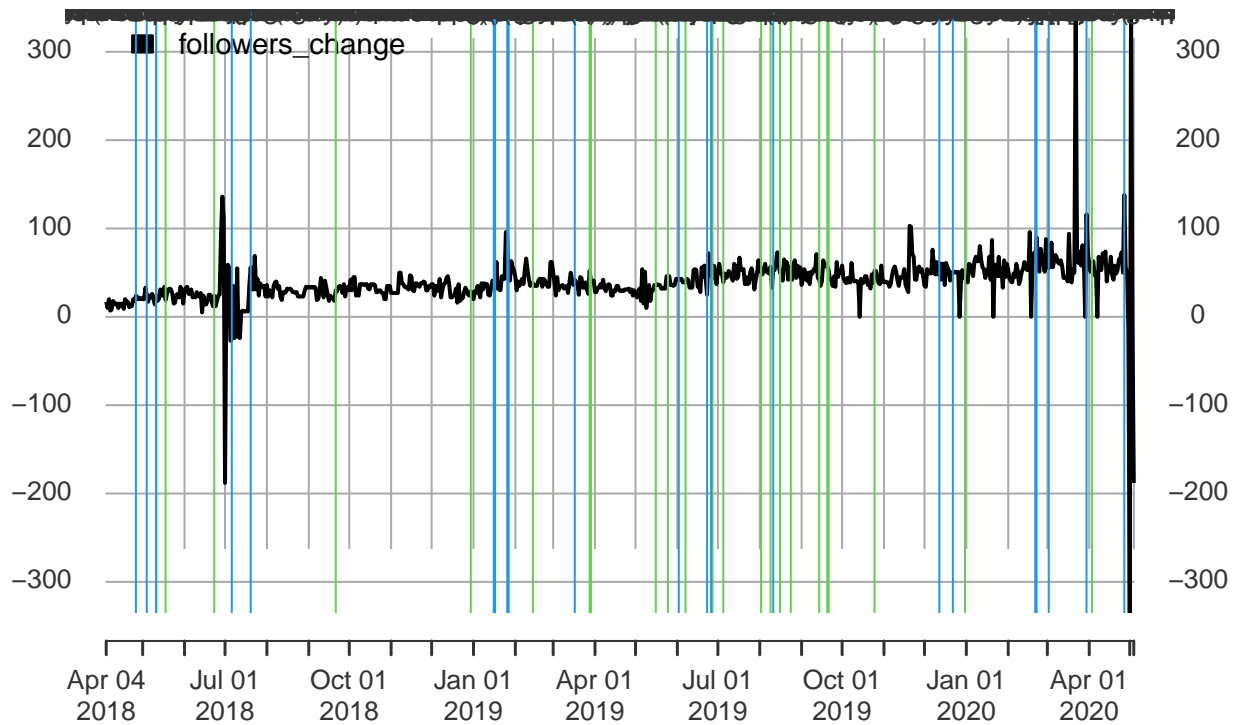
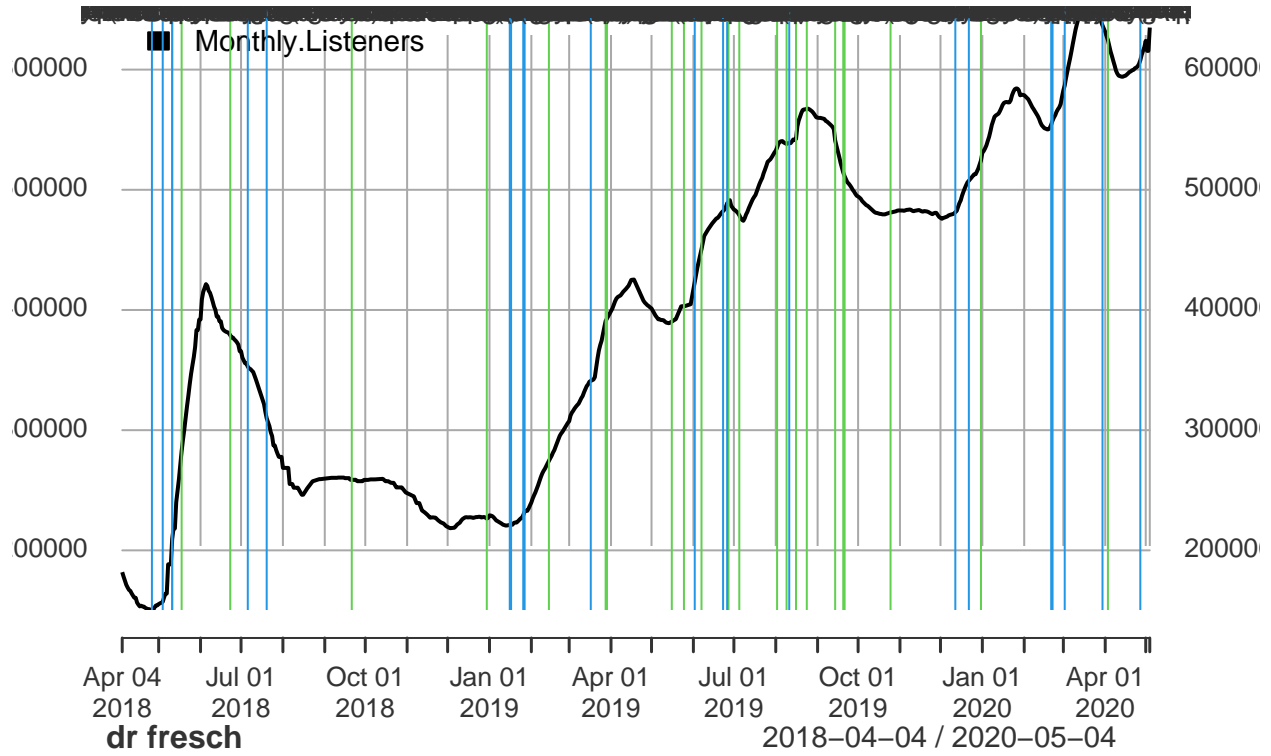
dj soda

2019-04-26 / 2020-05-04



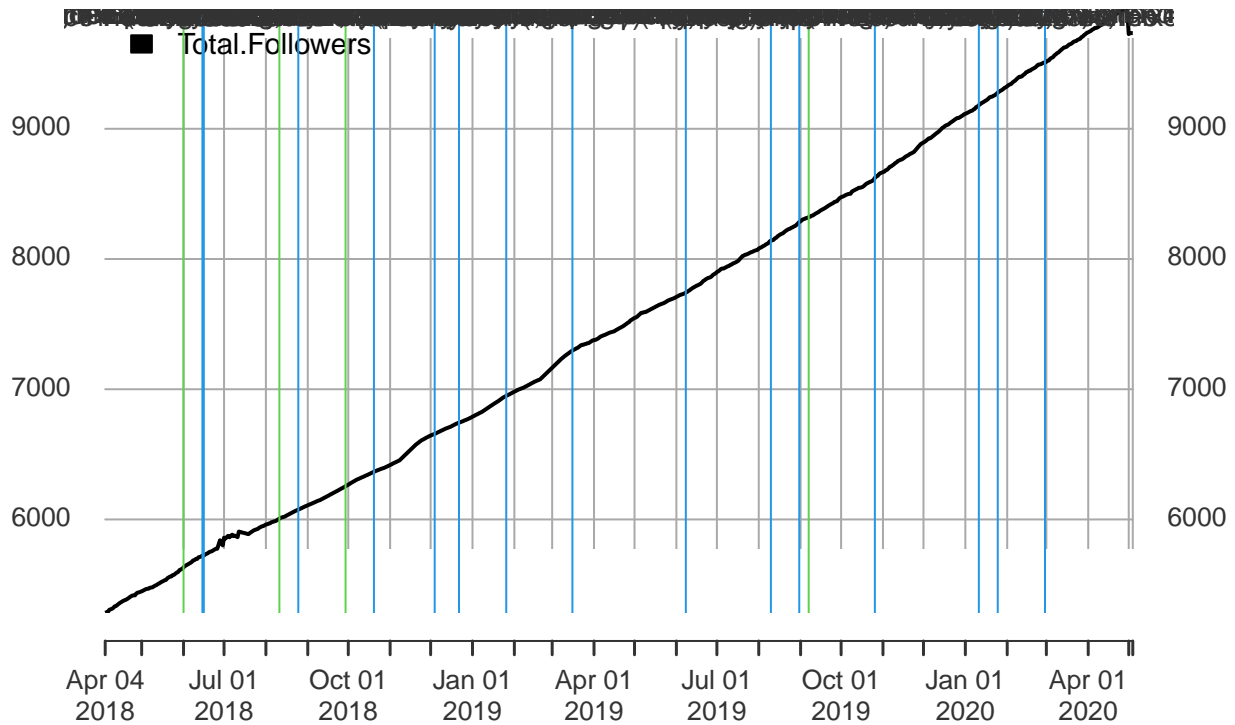
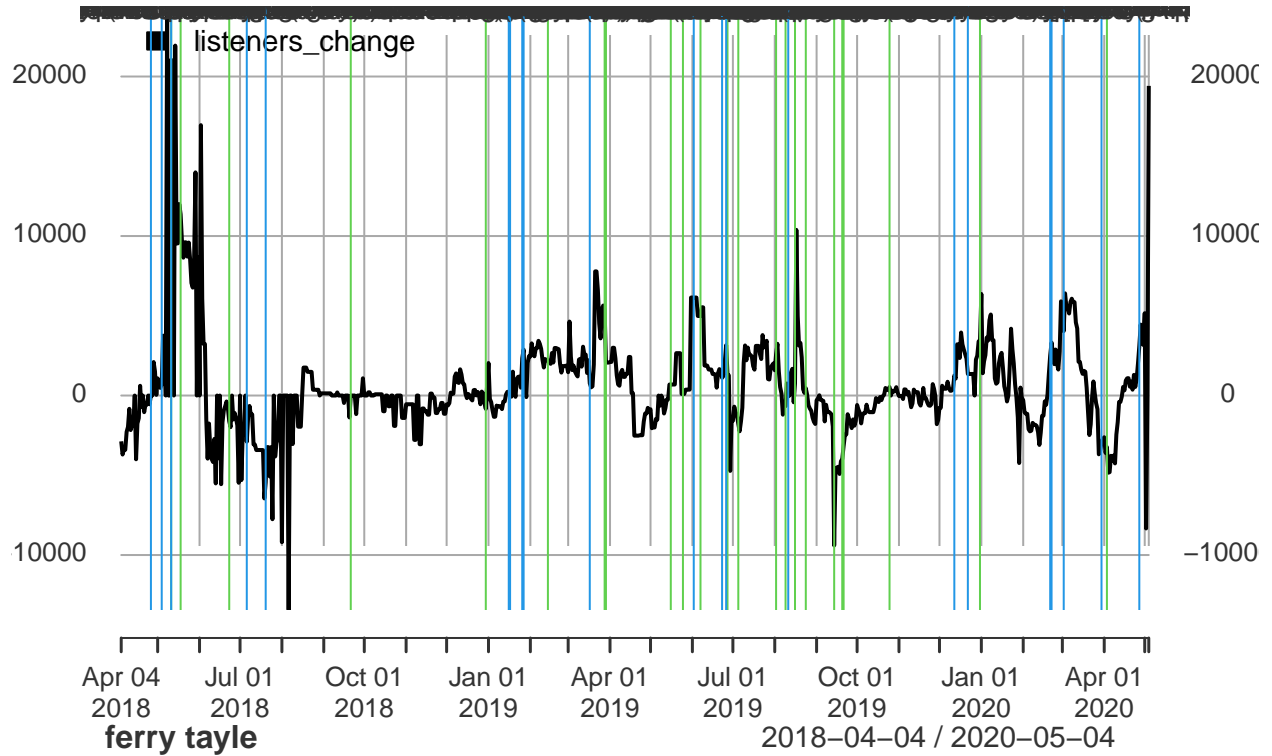
dr fresch

2018-04-04 / 2020-05-04



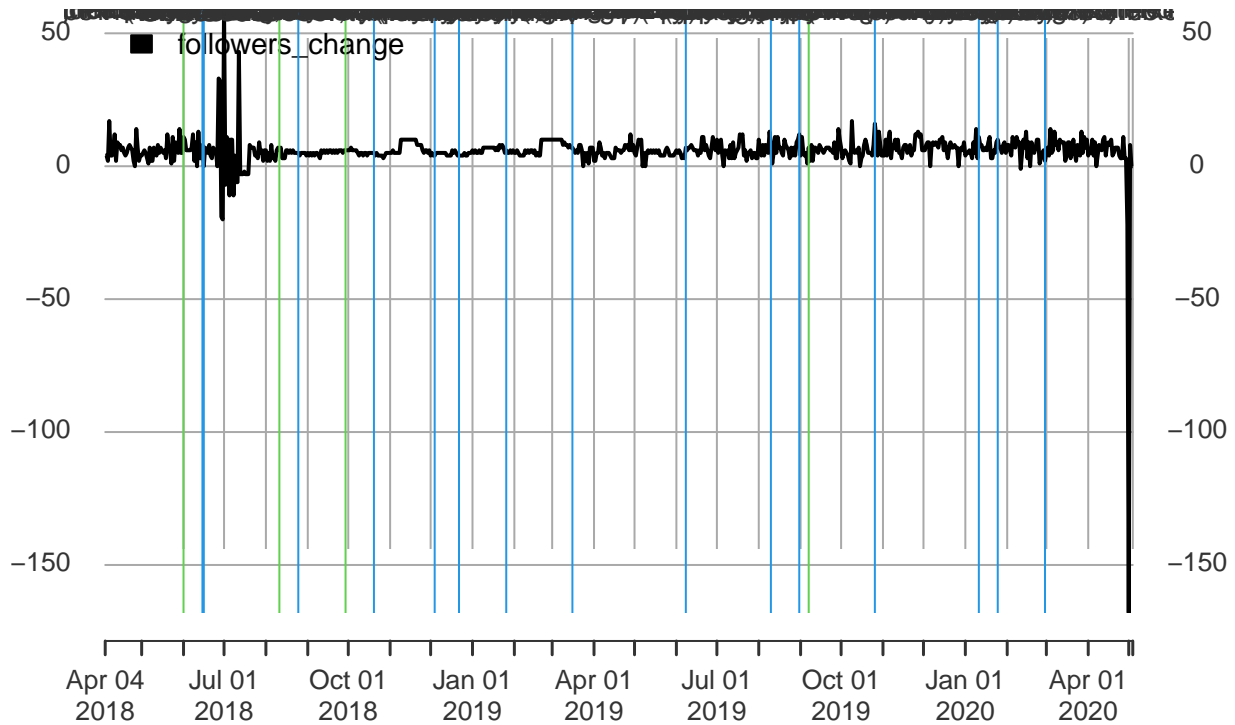
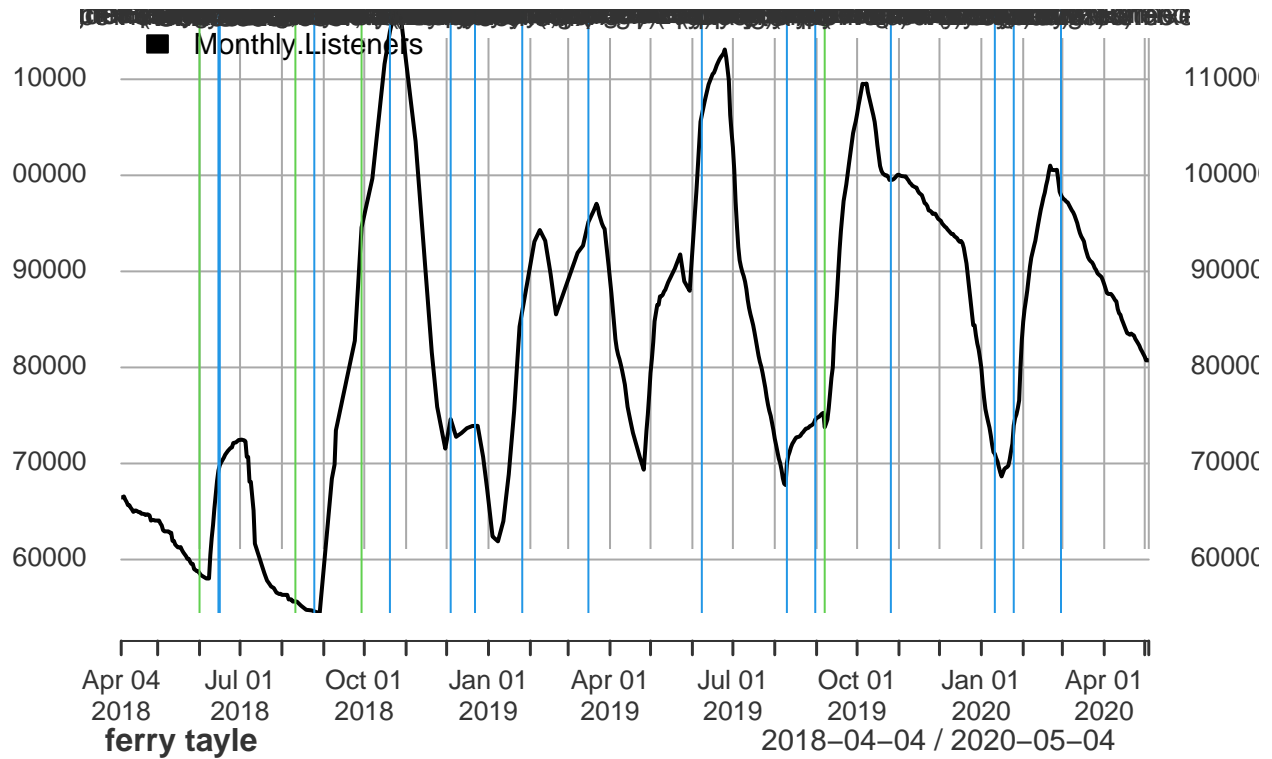
dr fresch

2018-04-04 / 2020-05-04



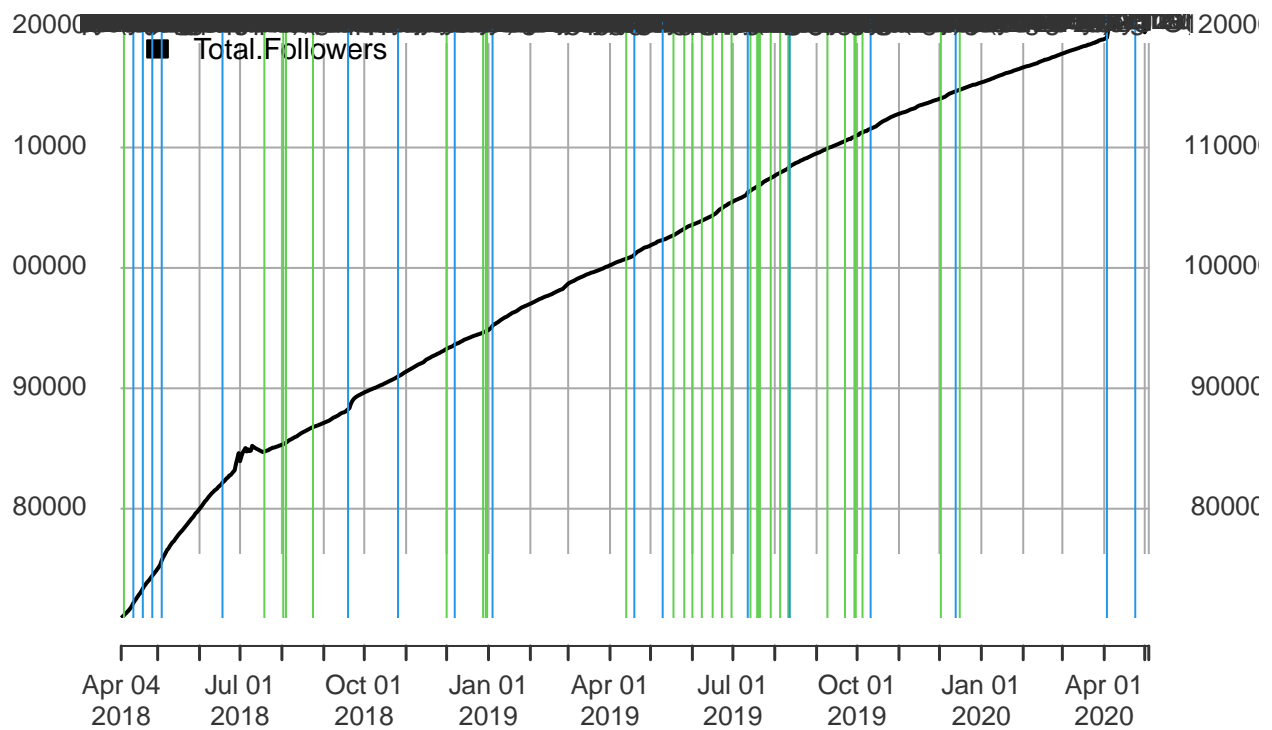
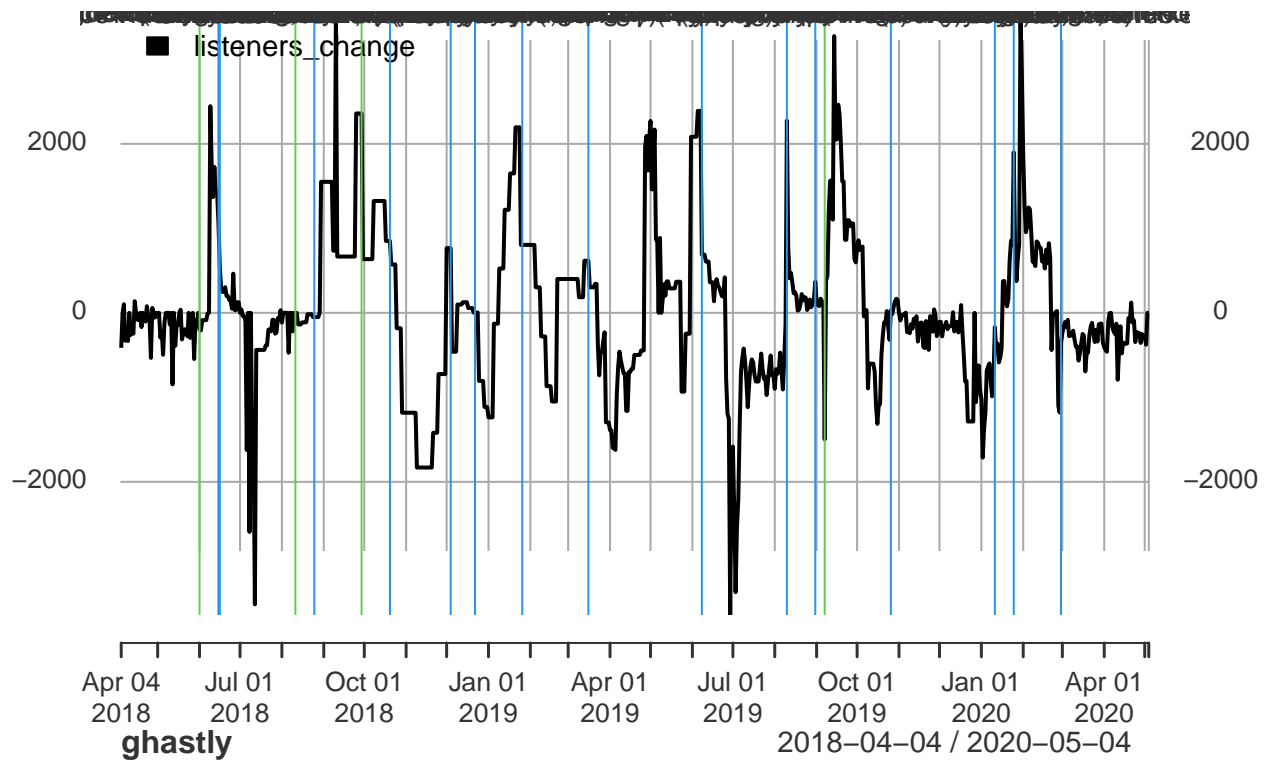
ferry taylor

2018-04-04 / 2020-05-04



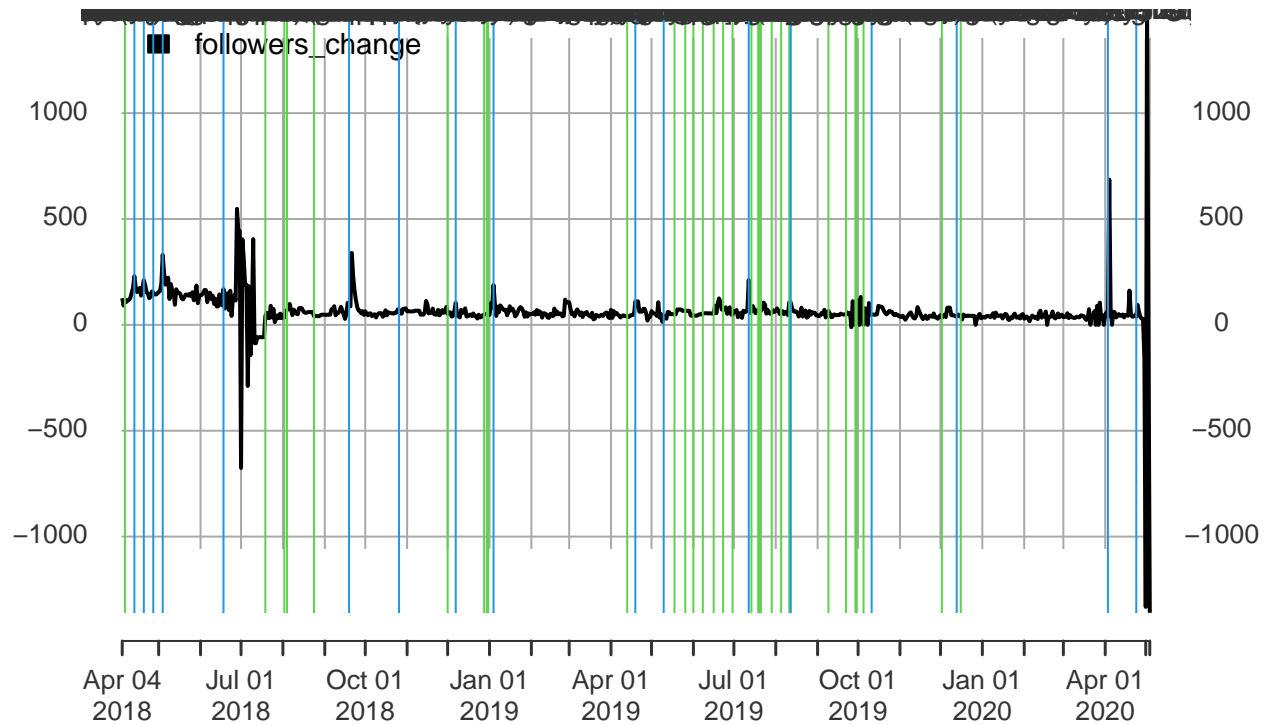
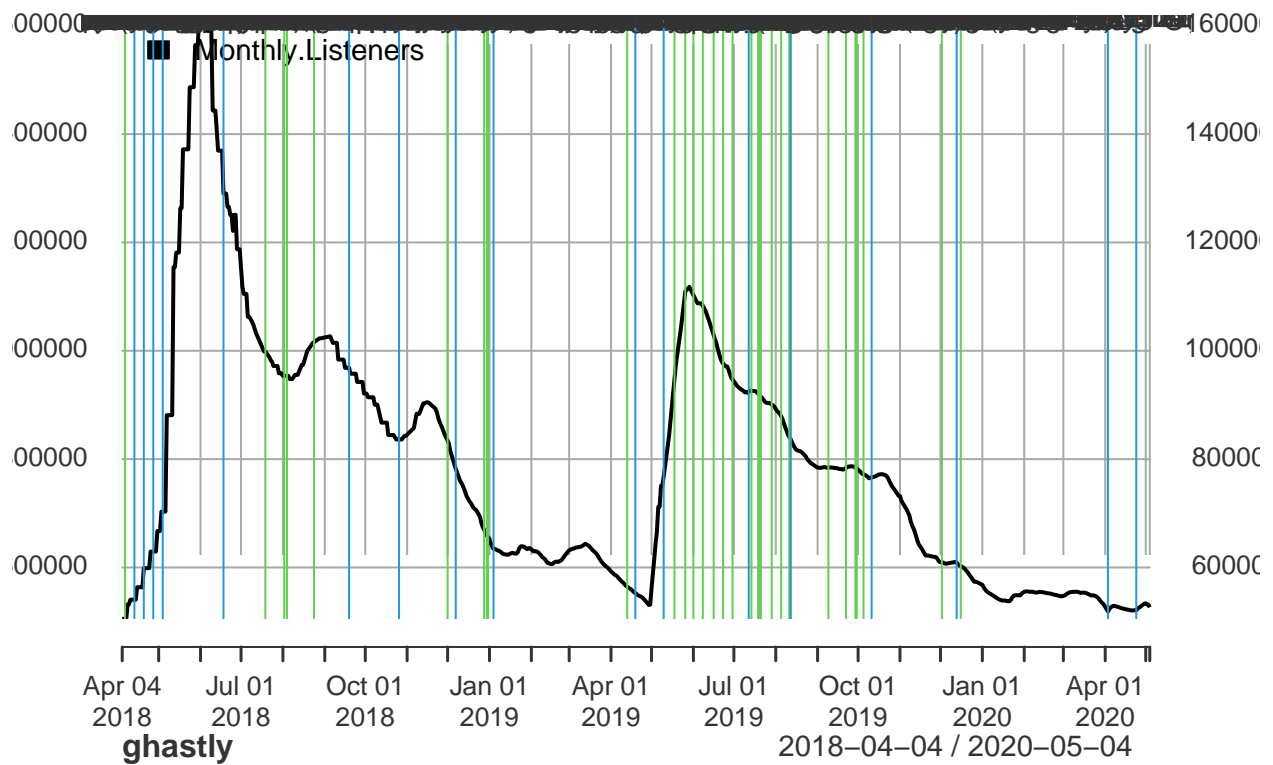
ferry taylor

2018-04-04 / 2020-05-04



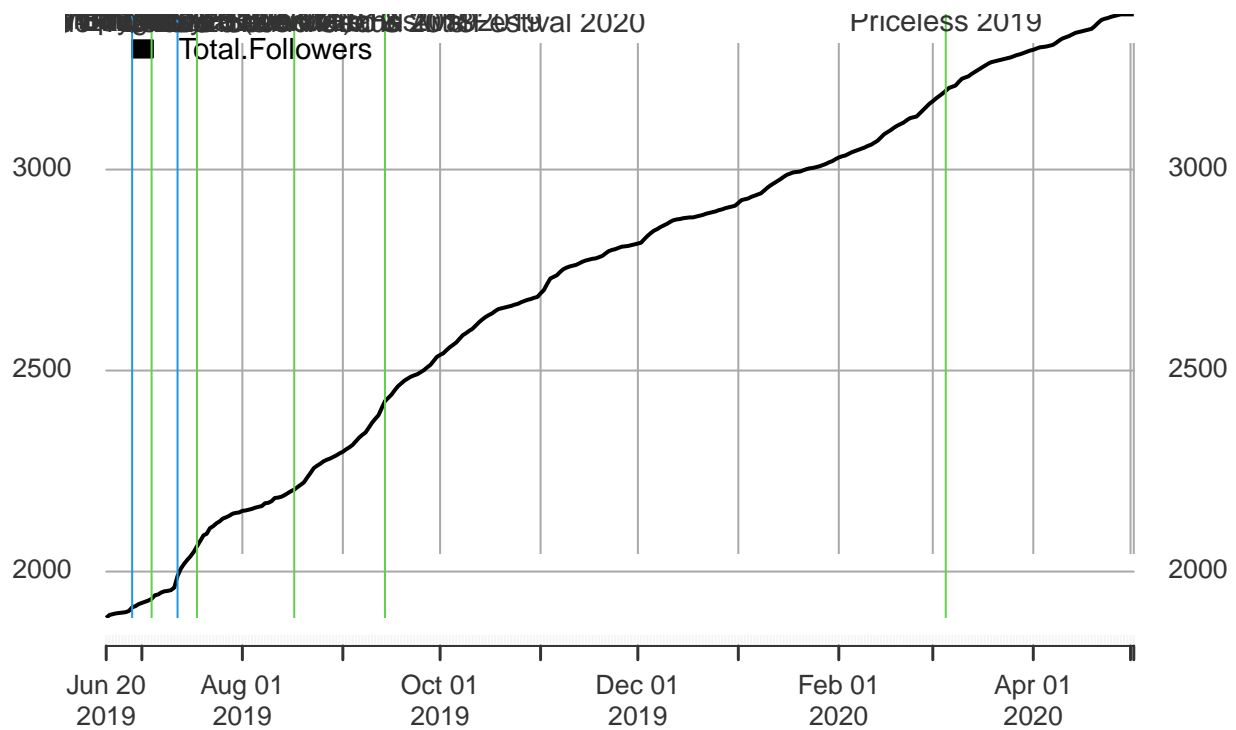
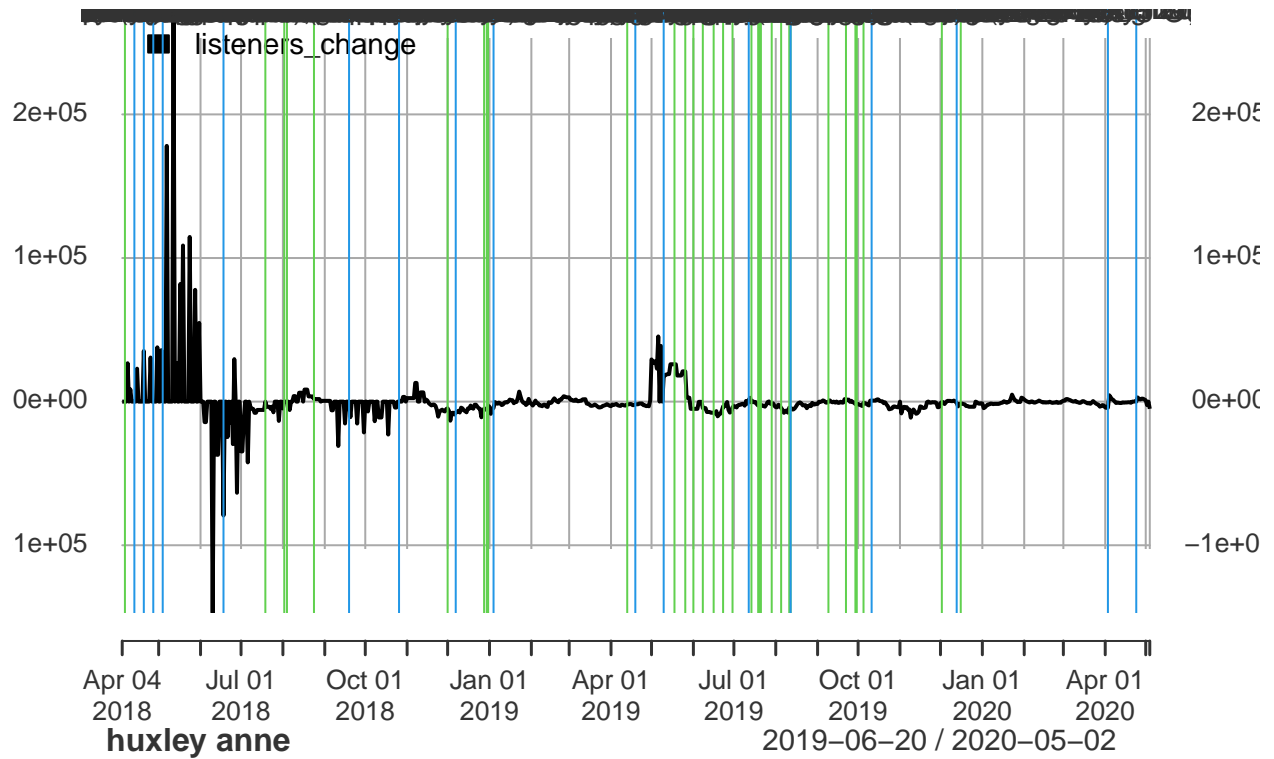
ghastly

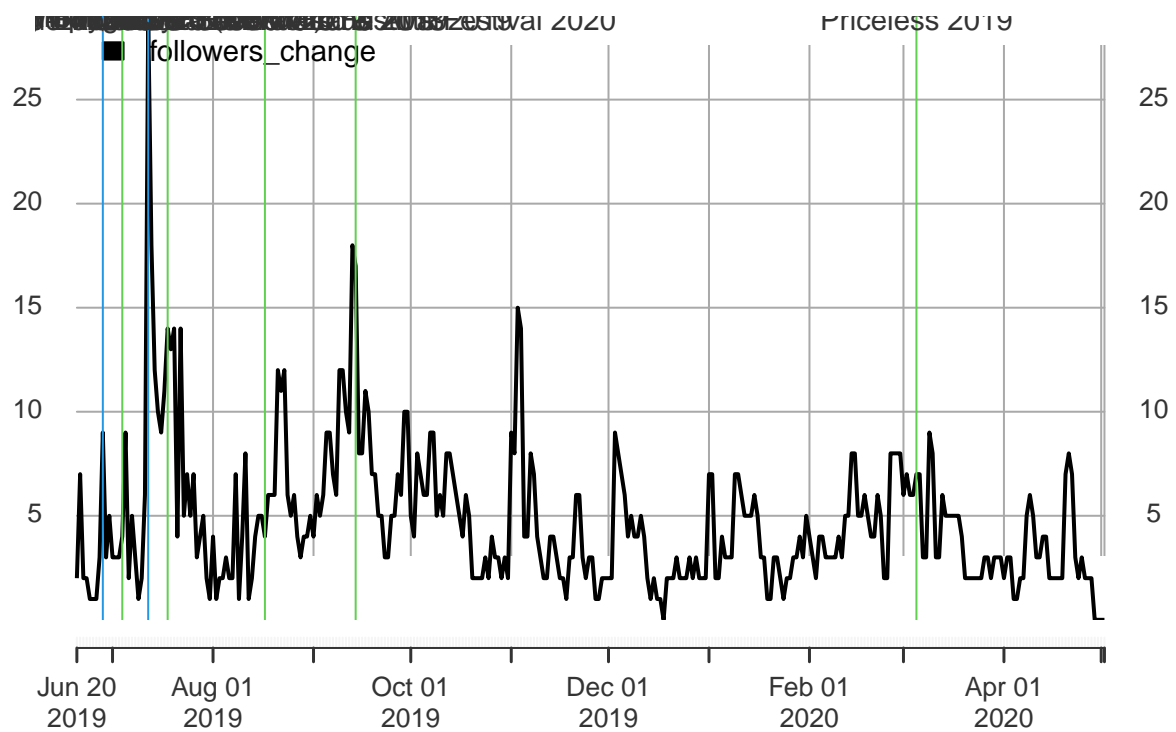
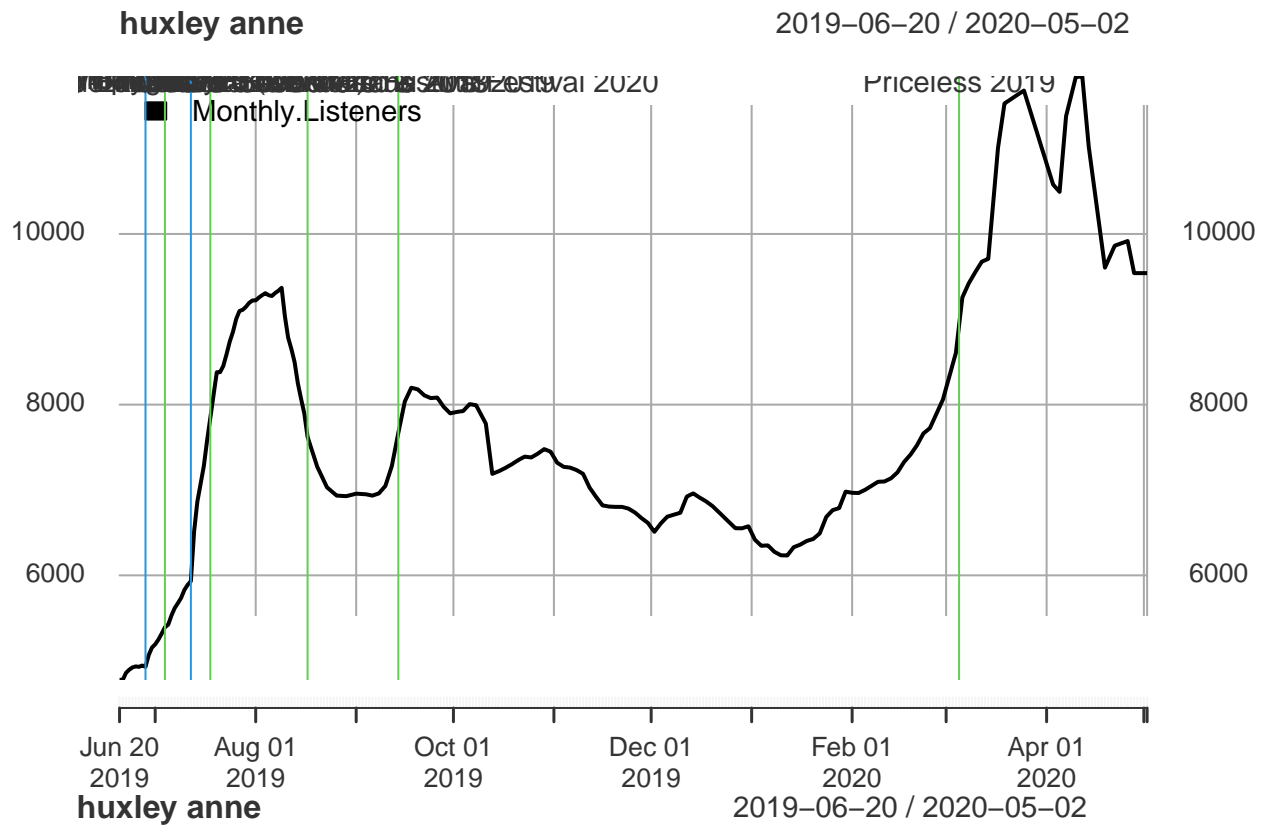
2018-04-04 / 2020-05-04



ghastly

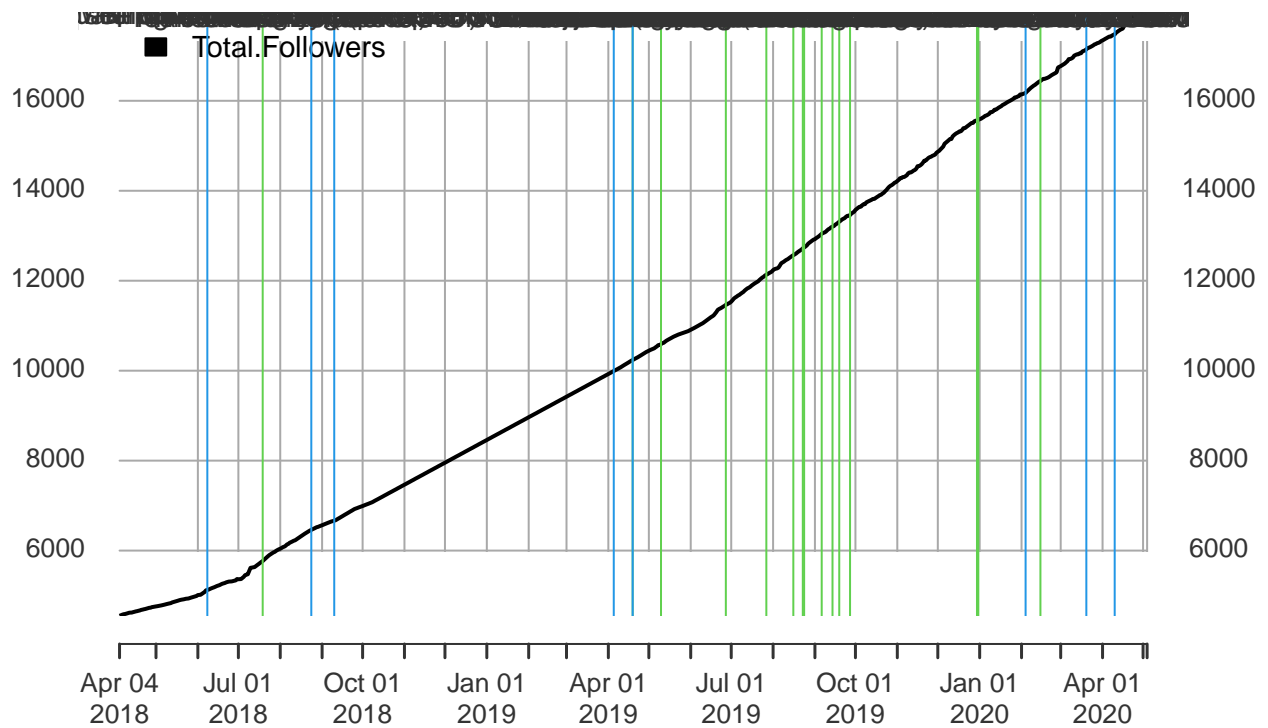
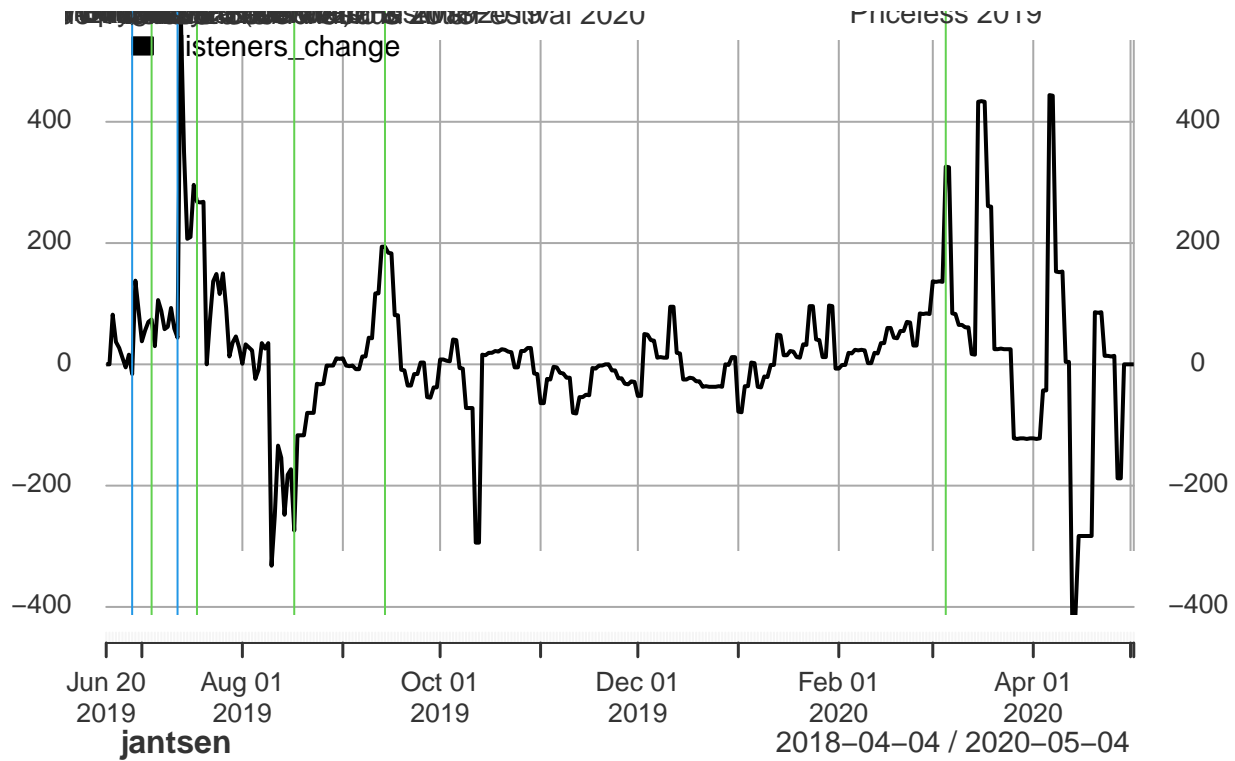
2018-04-04 / 2020-05-04

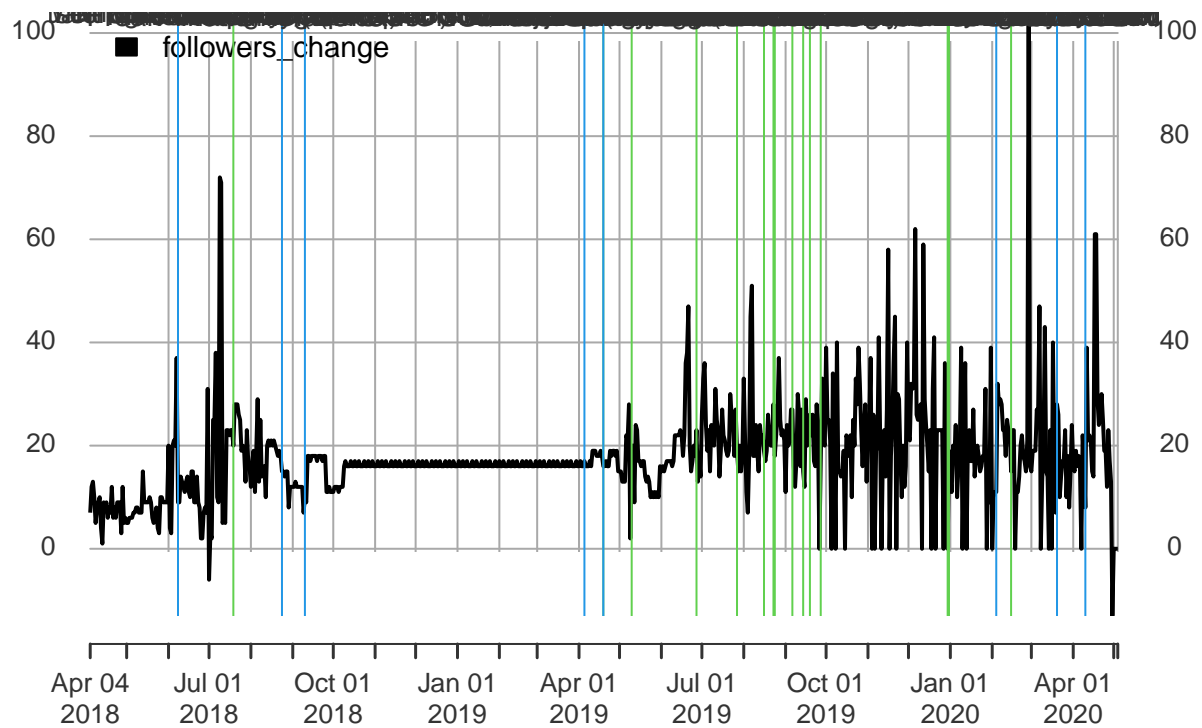
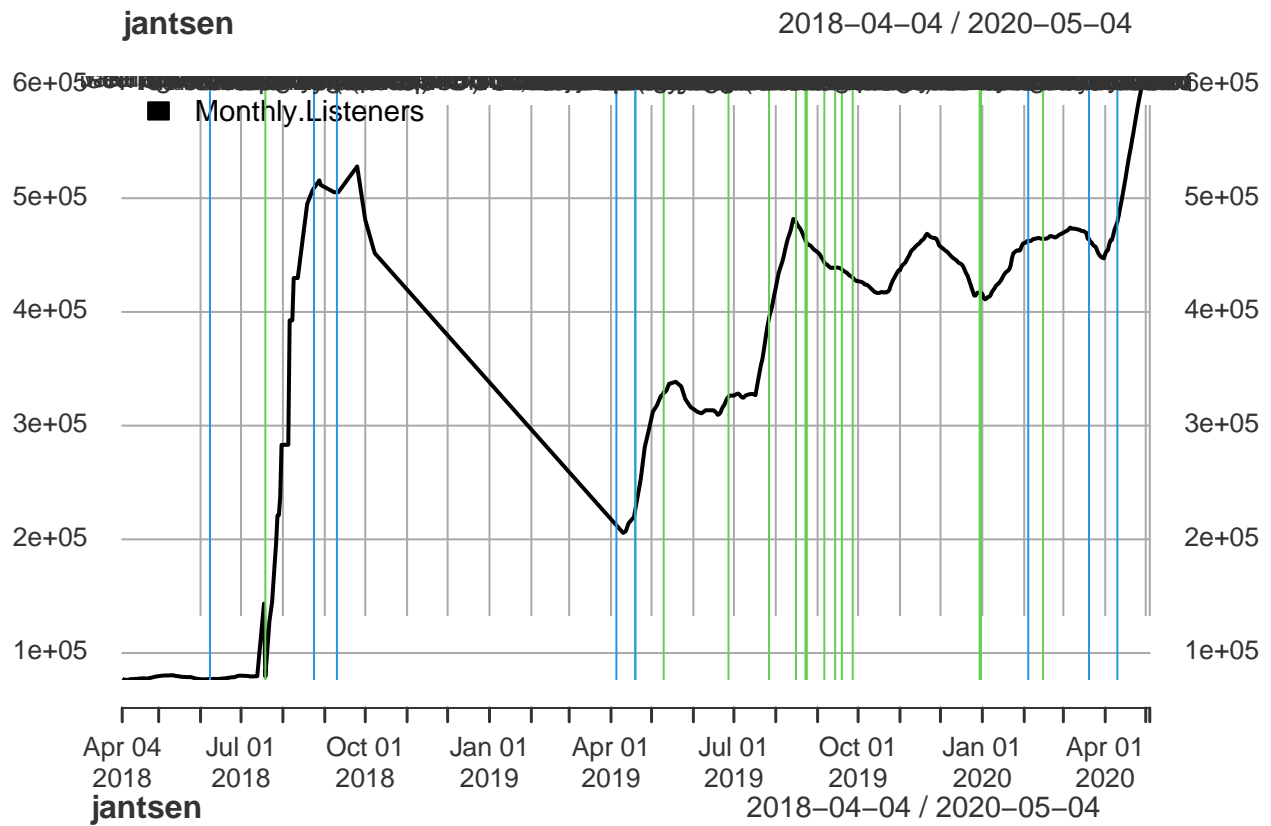


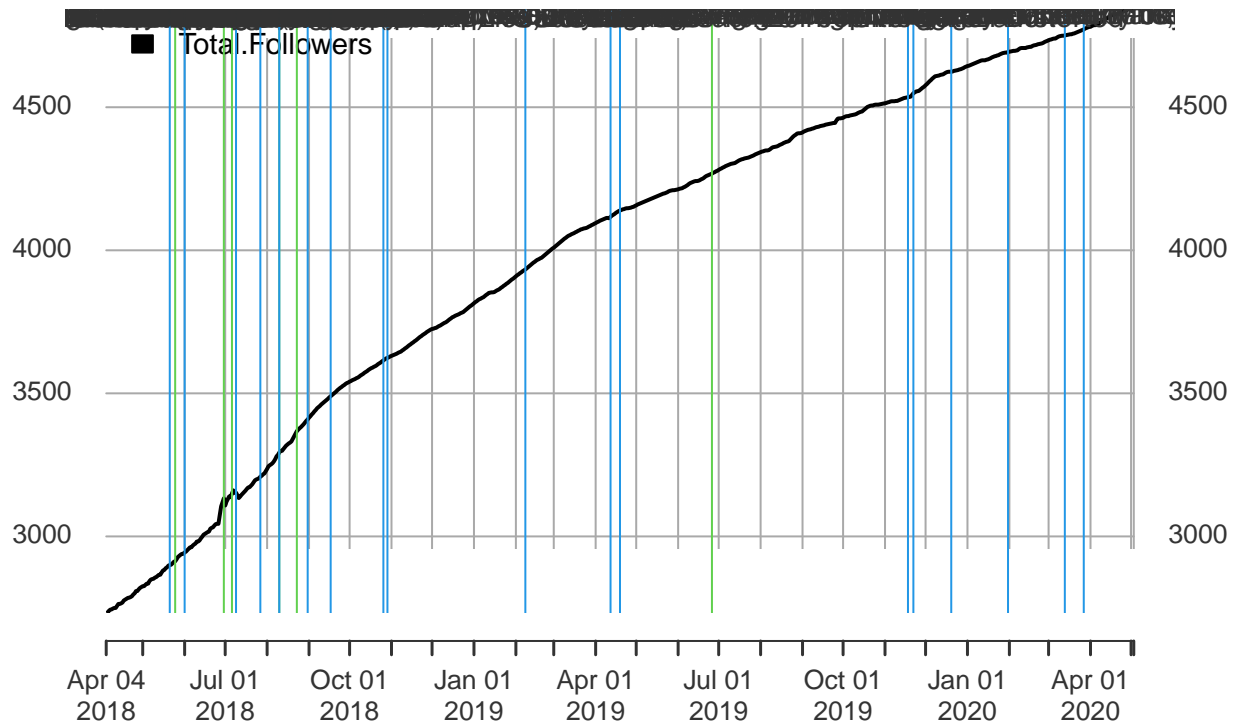
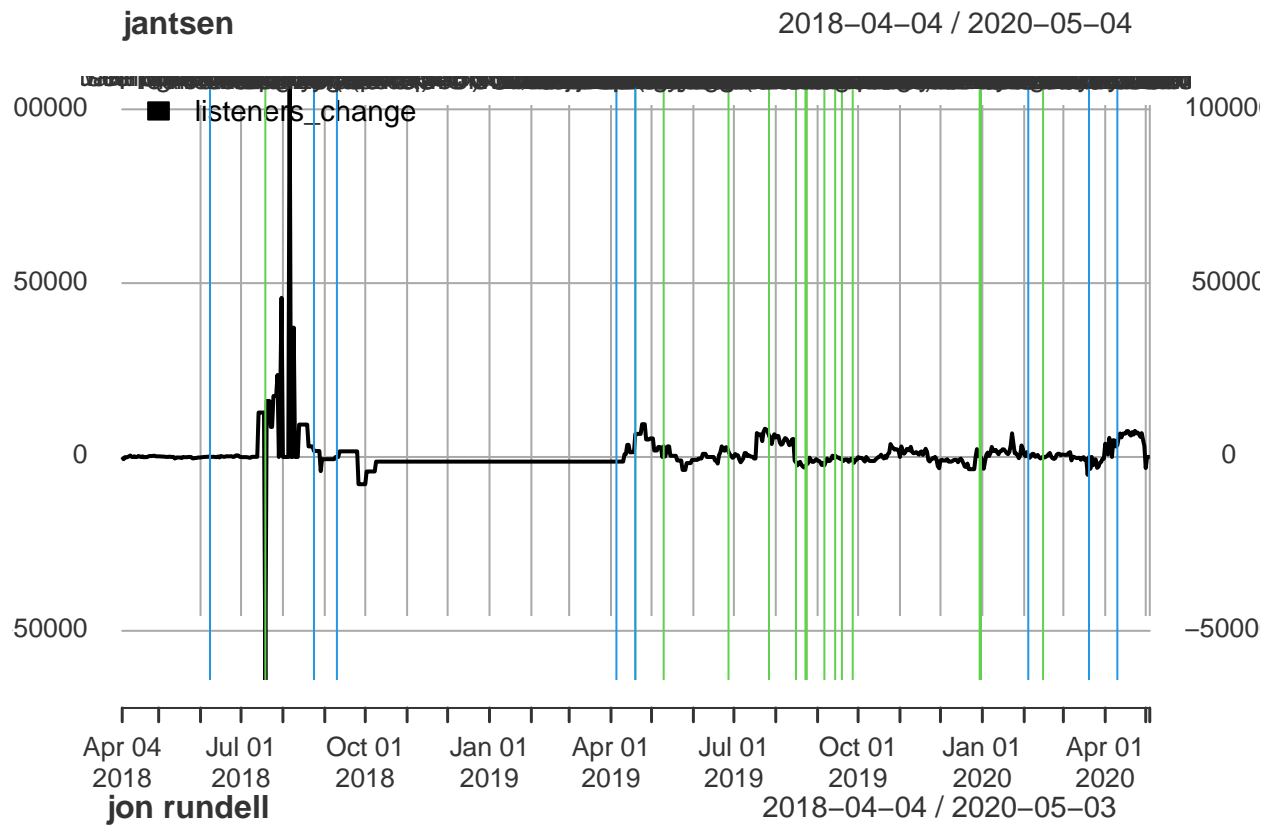


huxley anne

2019-06-20 / 2020-05-02

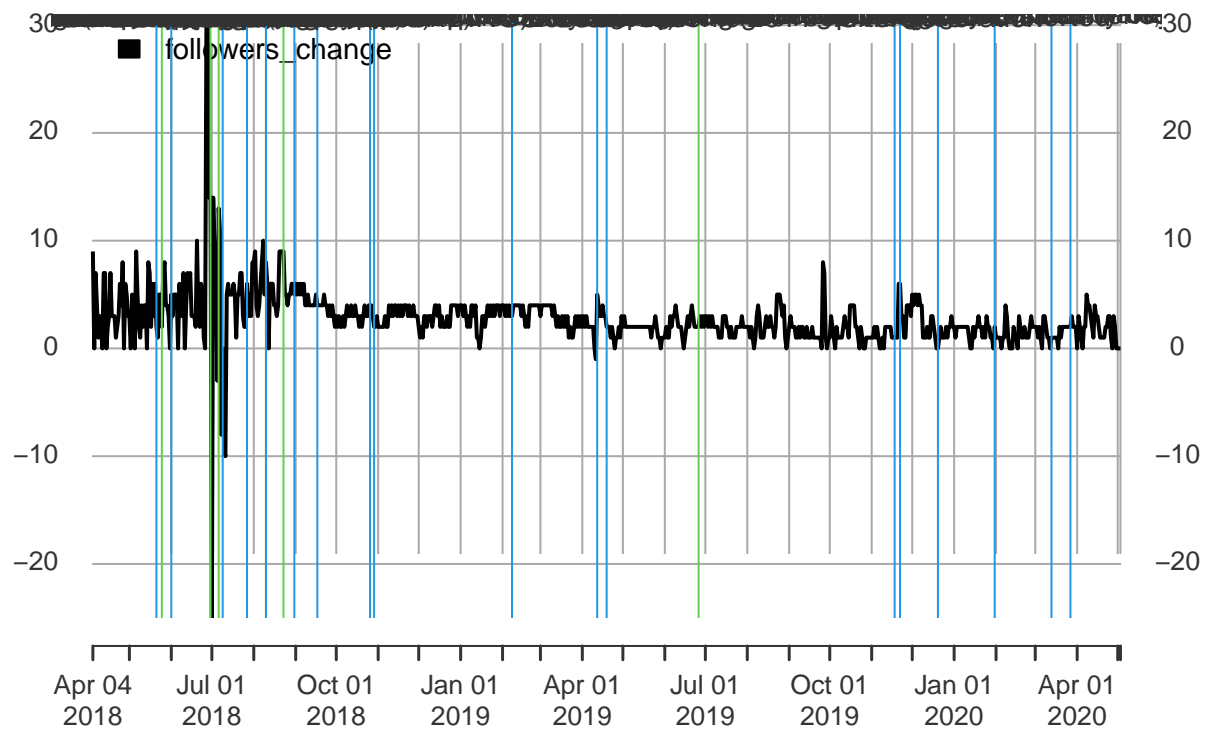
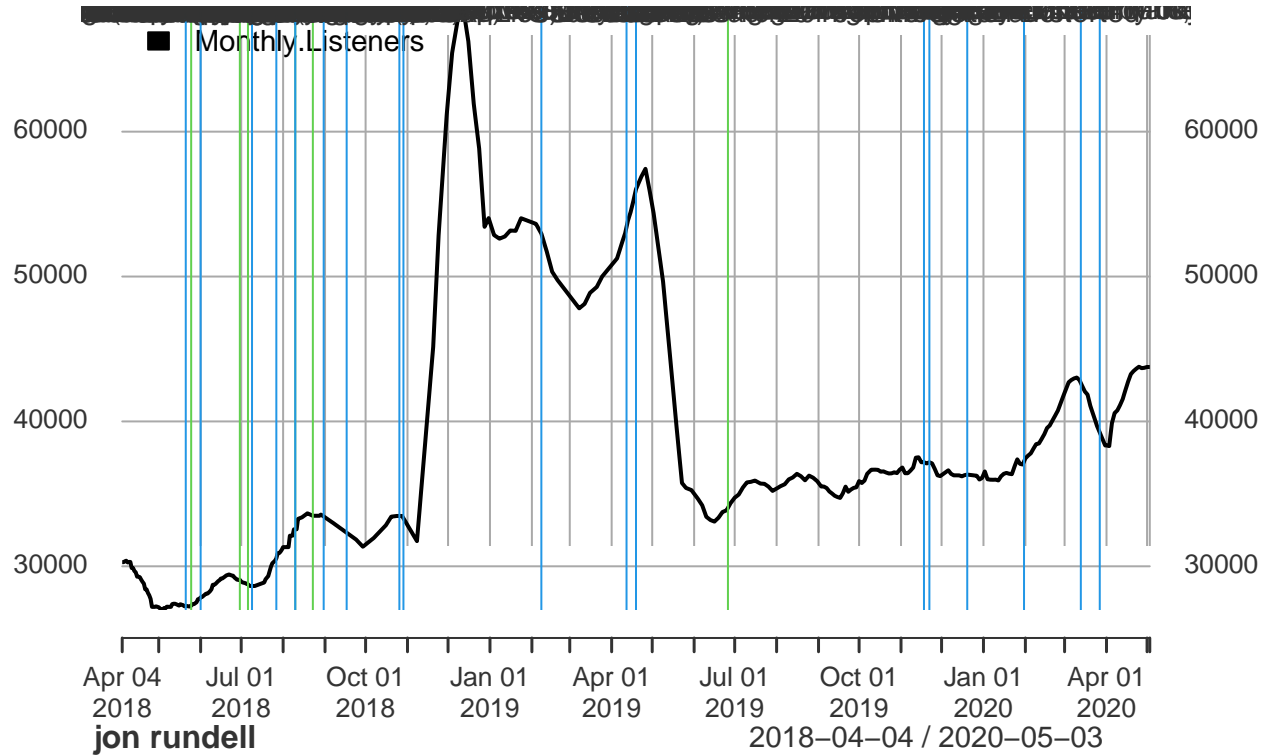






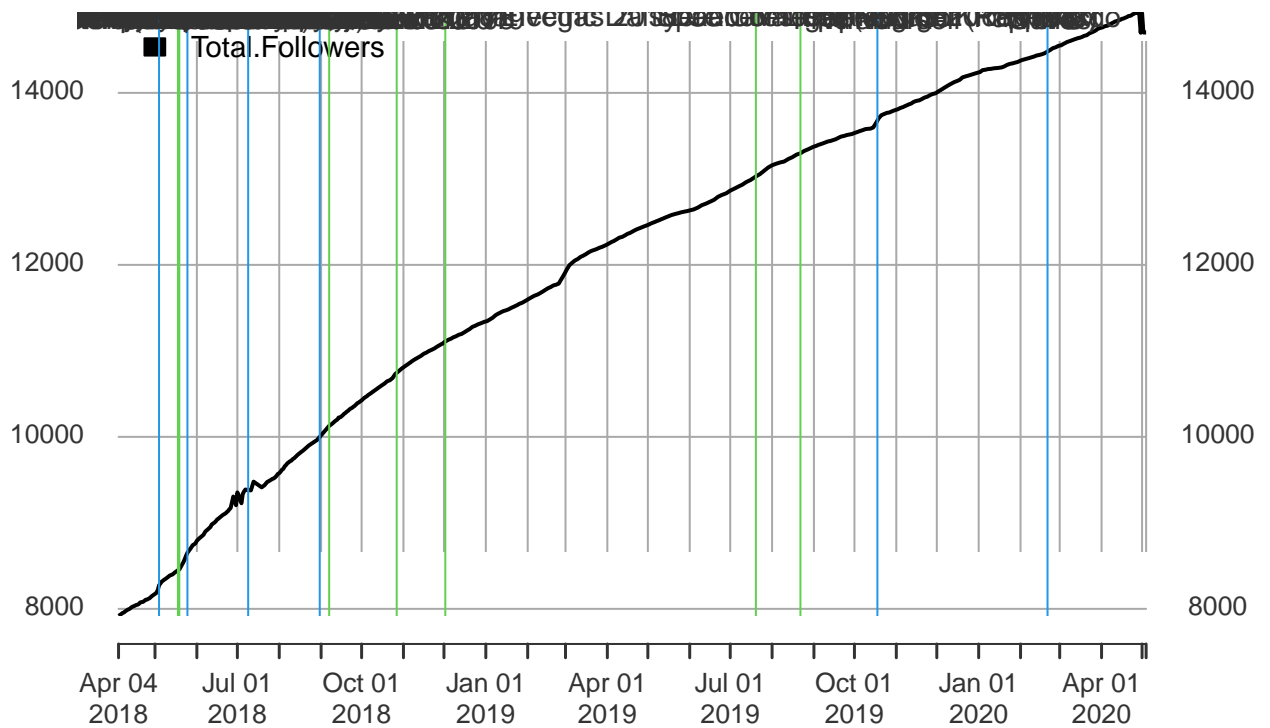
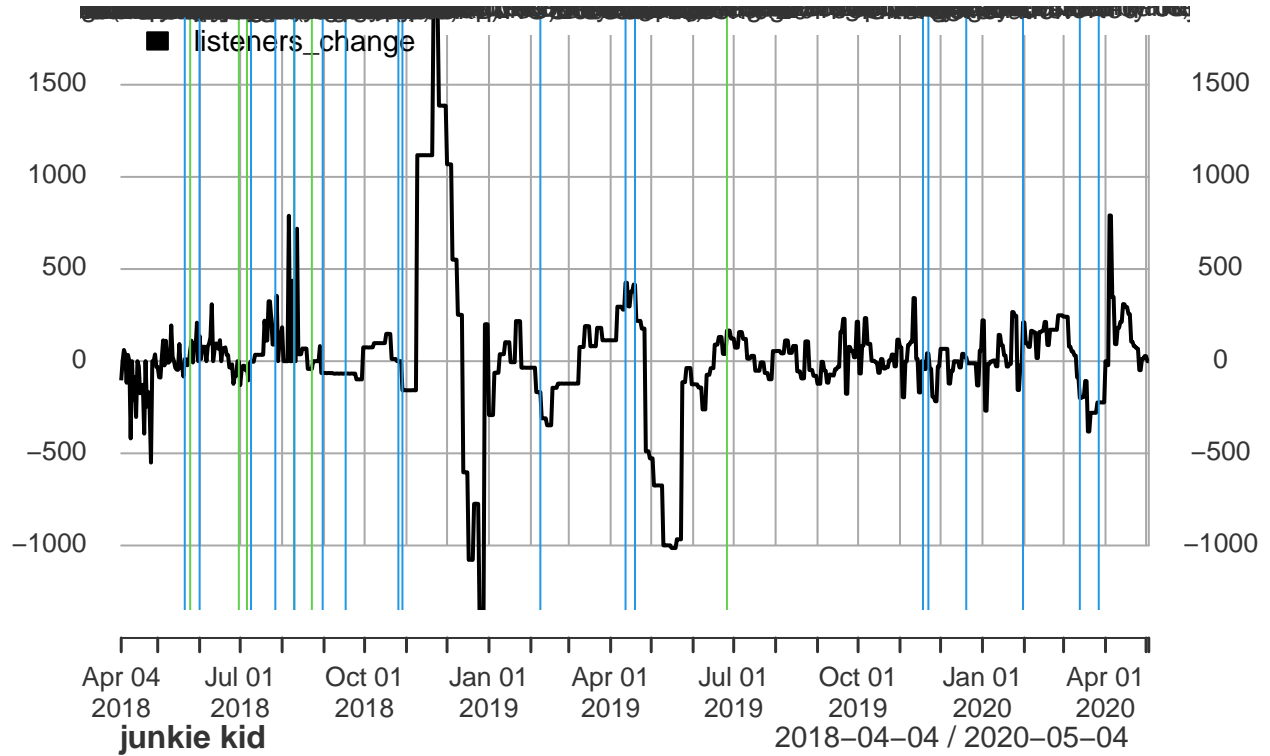
jon rundell

2018-04-04 / 2020-05-03



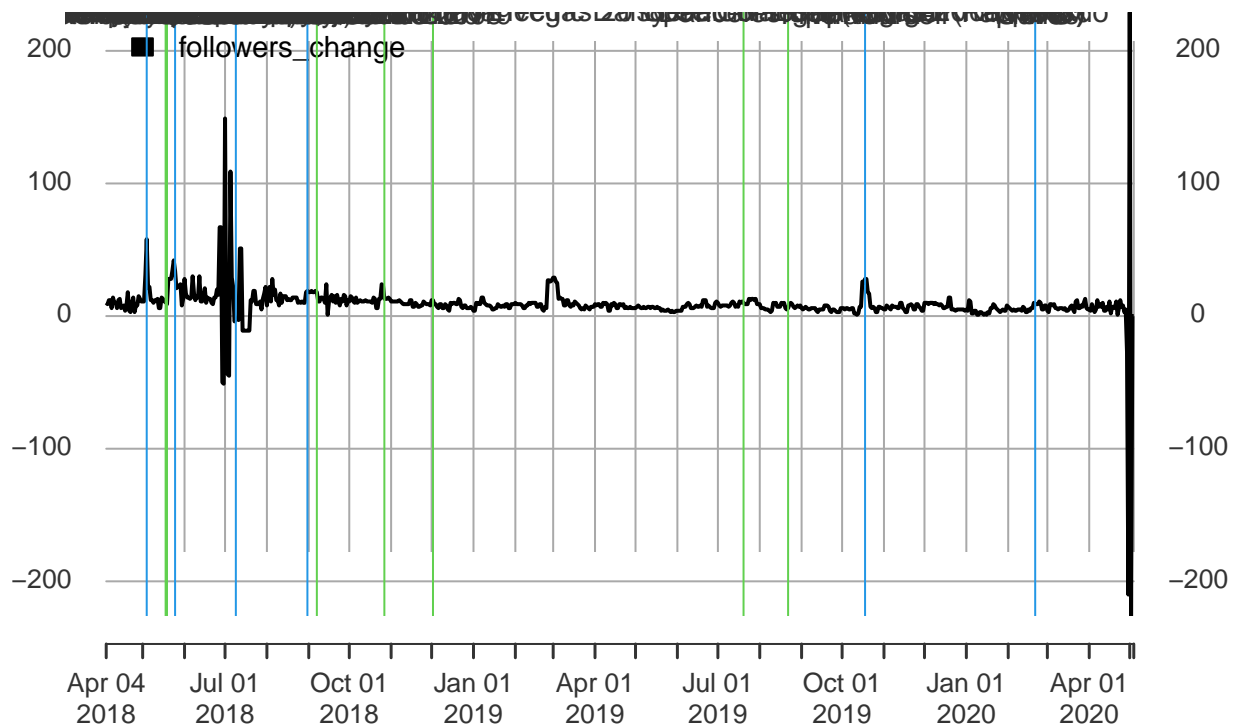
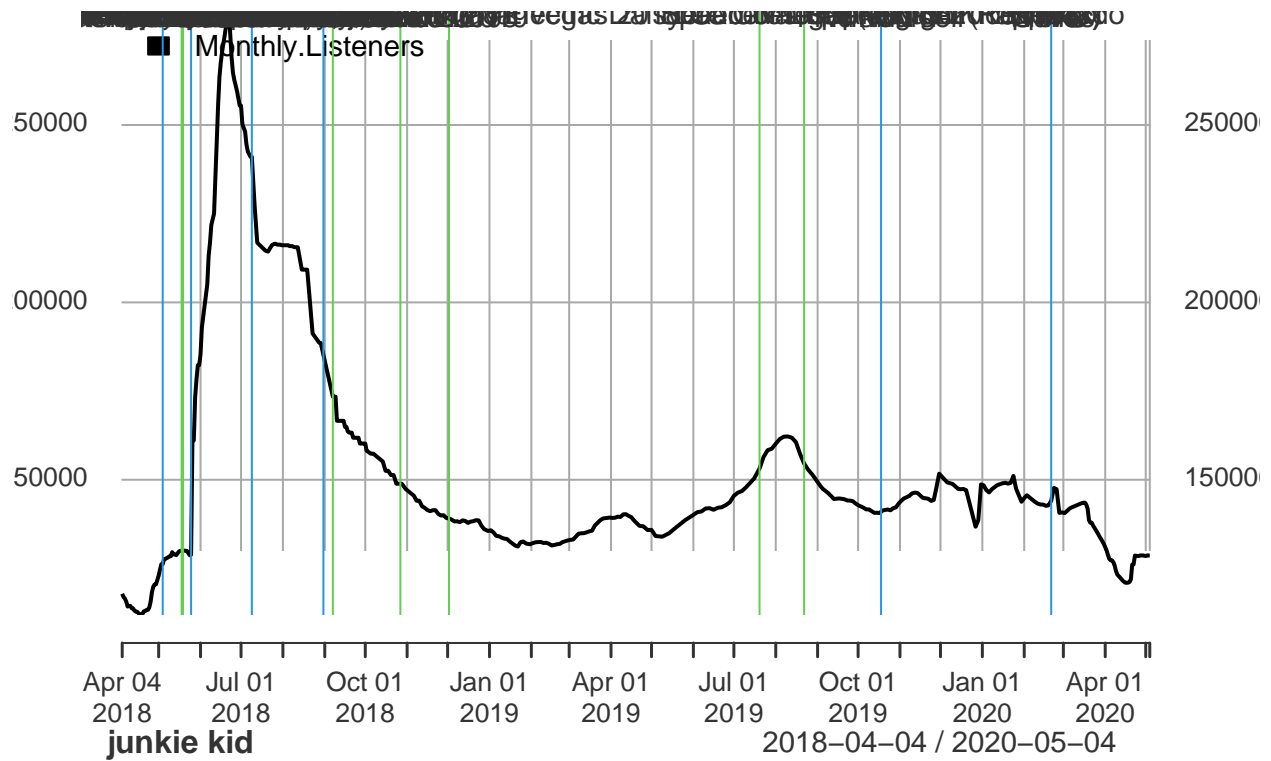
jon rundell

2018-04-04 / 2020-05-03



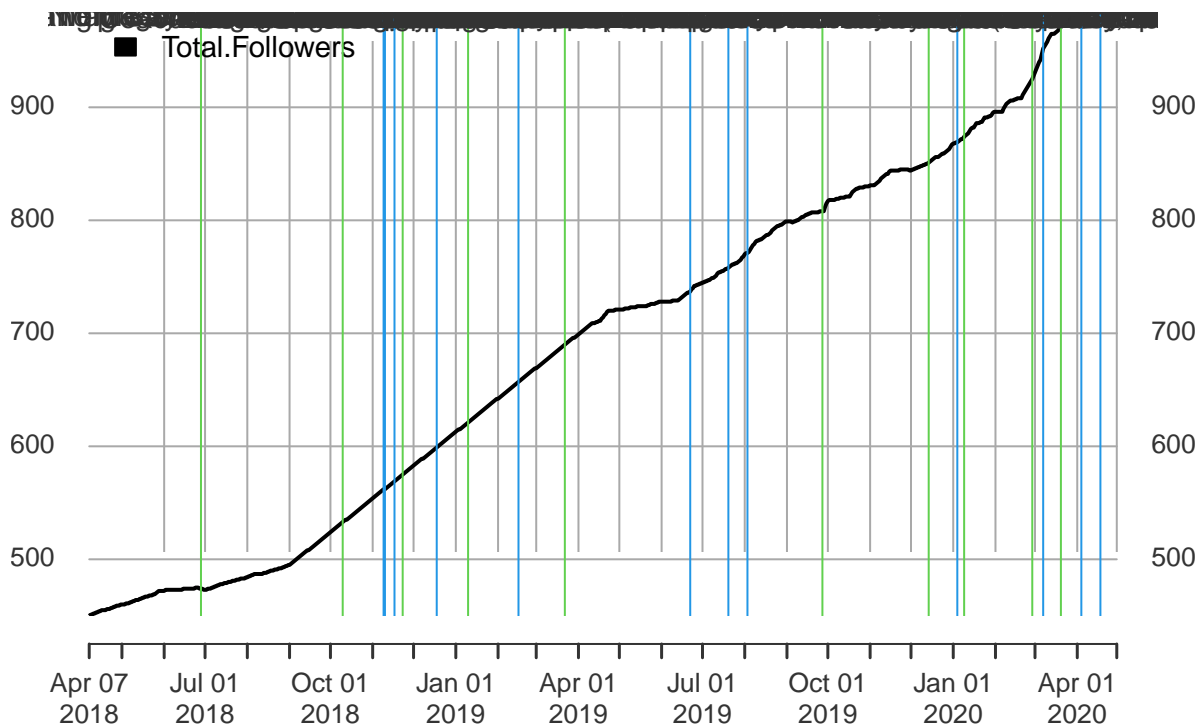
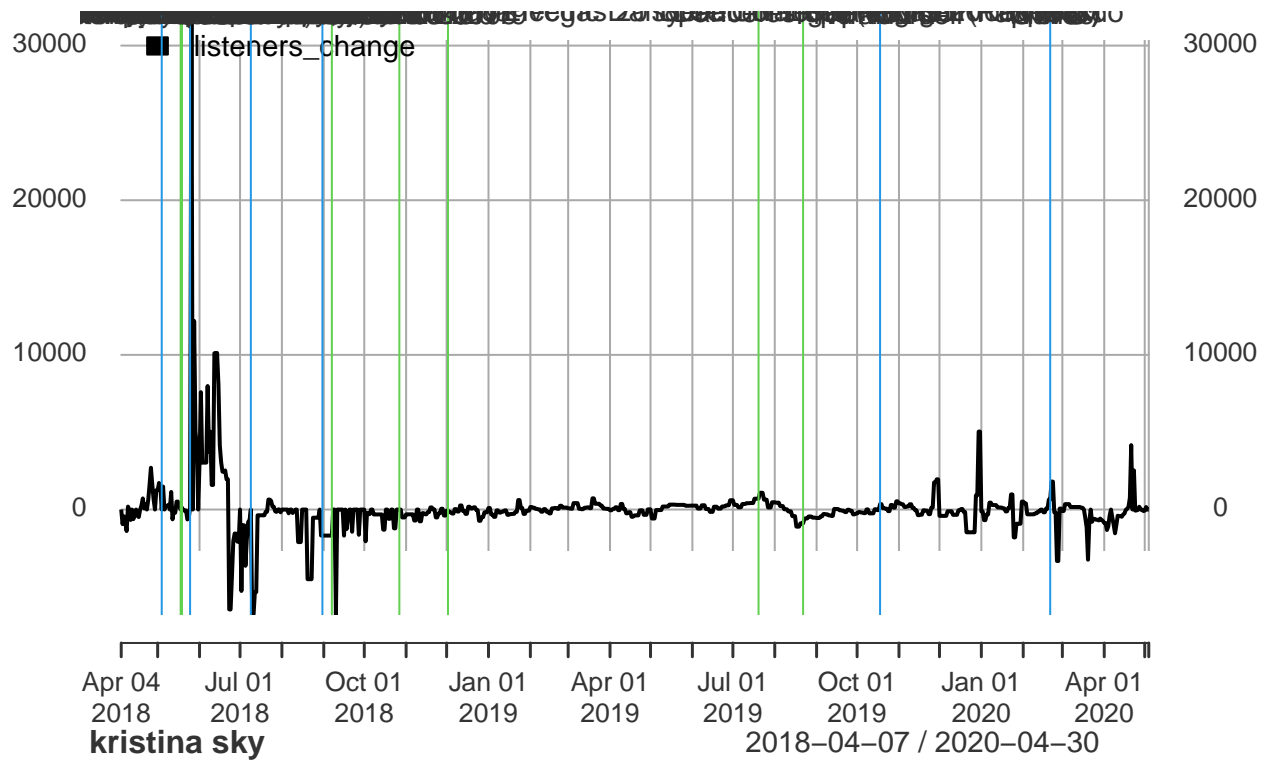
junkie kid

2018-04-04 / 2020-05-04



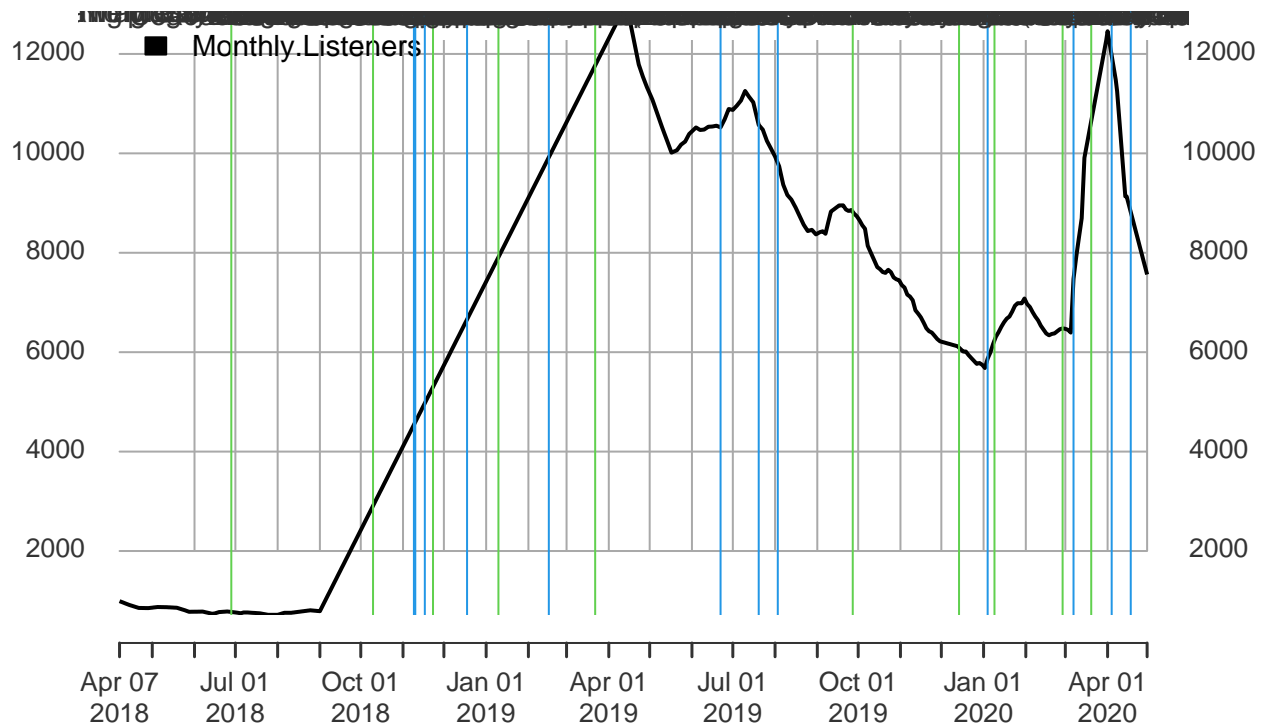
junkie kid

2018-04-04 / 2020-05-04



kristina sky

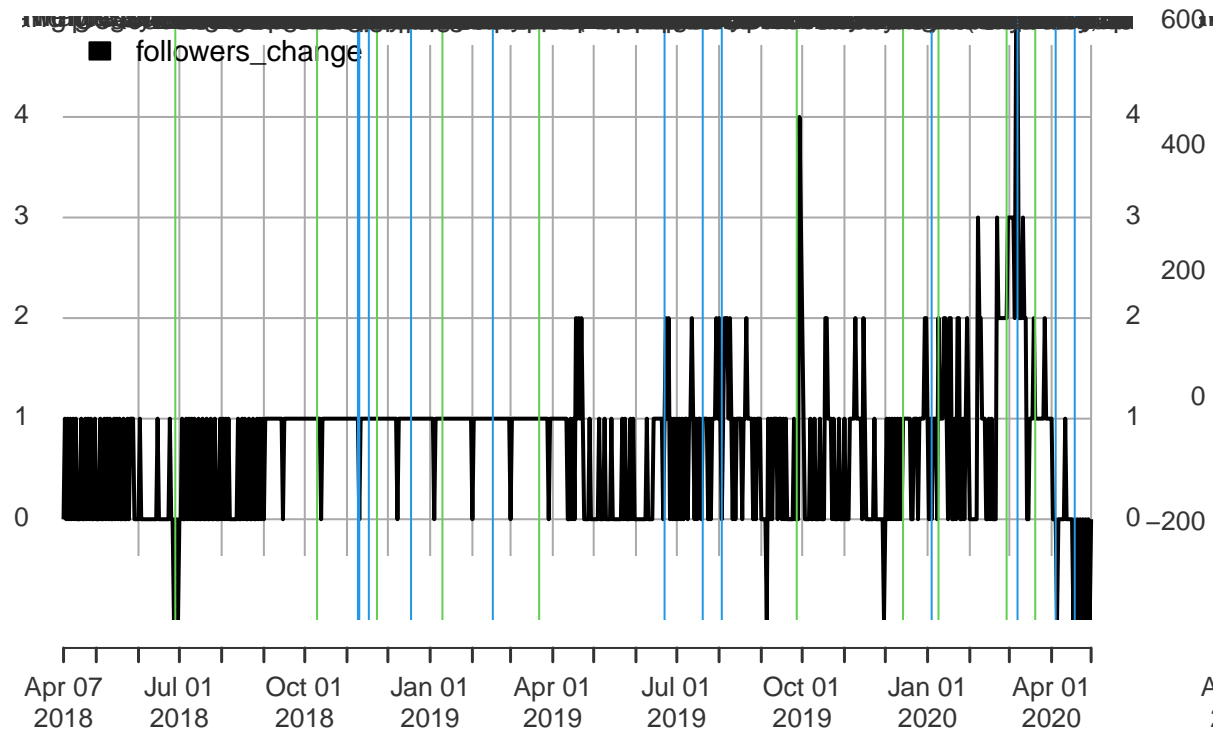
2018-04-07 / 2020-04-30



kristina sky

2018-04-07 / 2020-04-30

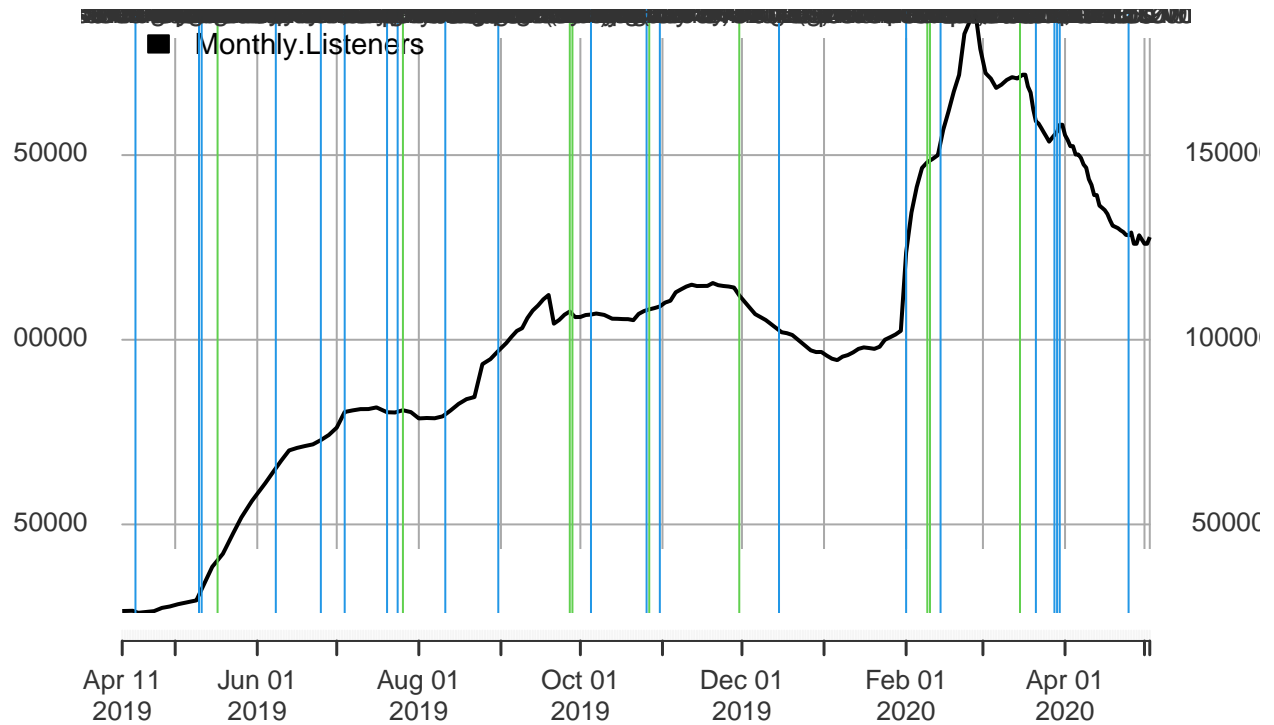
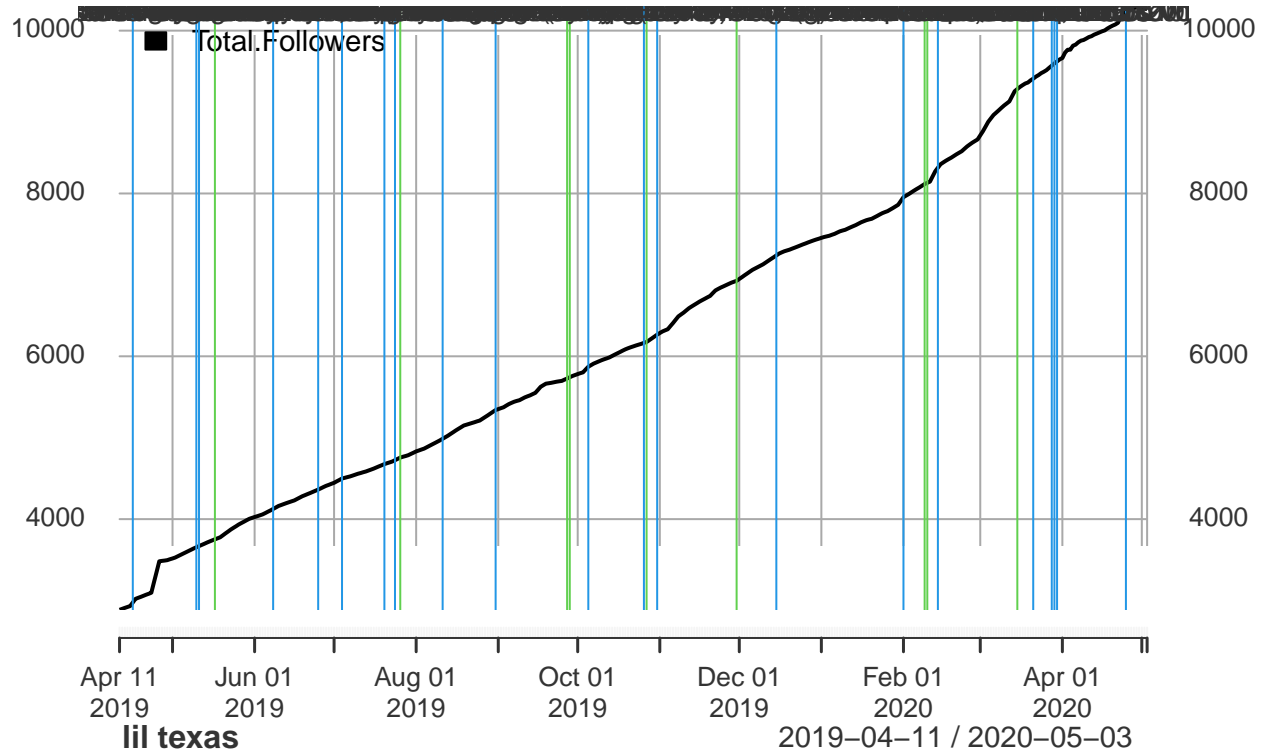
kristina sky



listeners_c

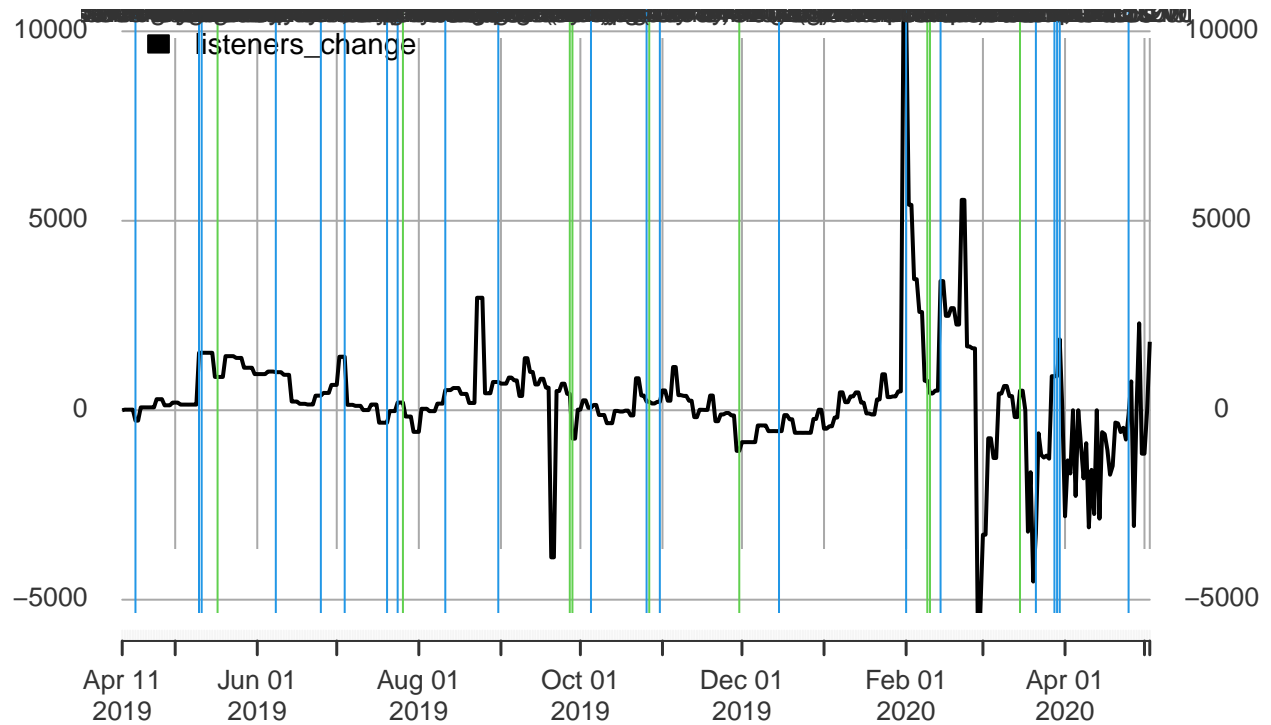
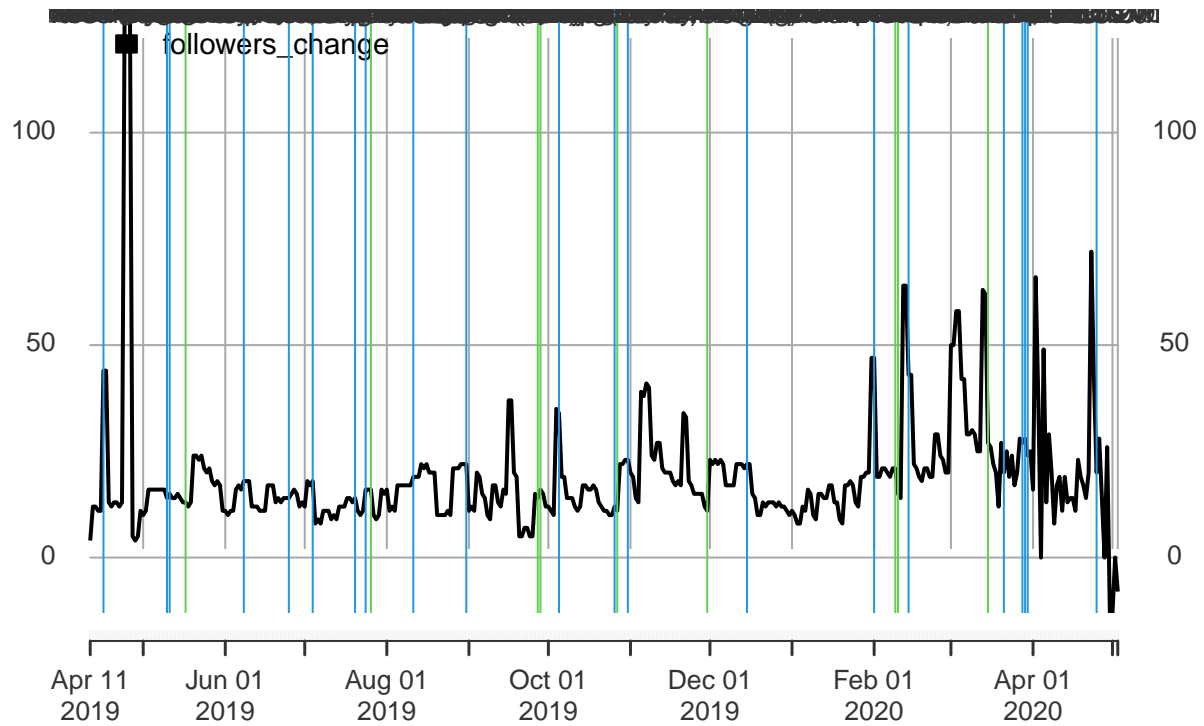
lil texas

2019-04-11 / 2020-05-03



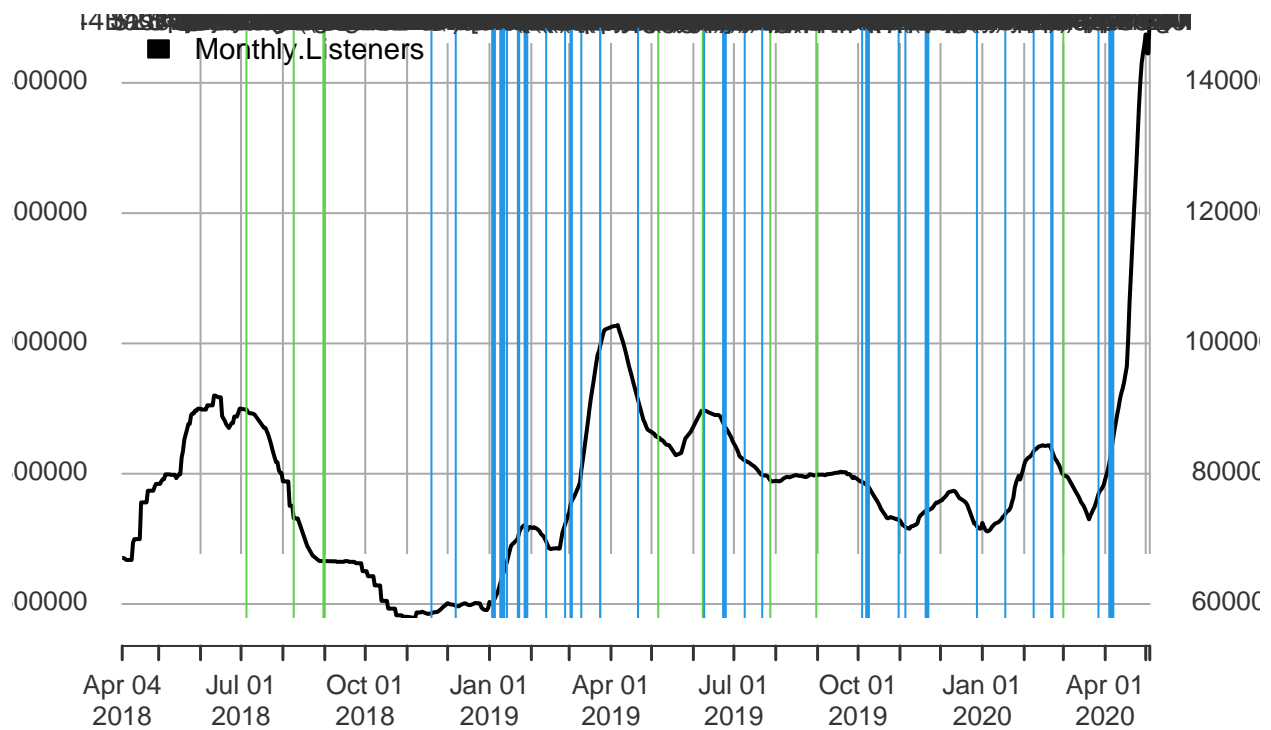
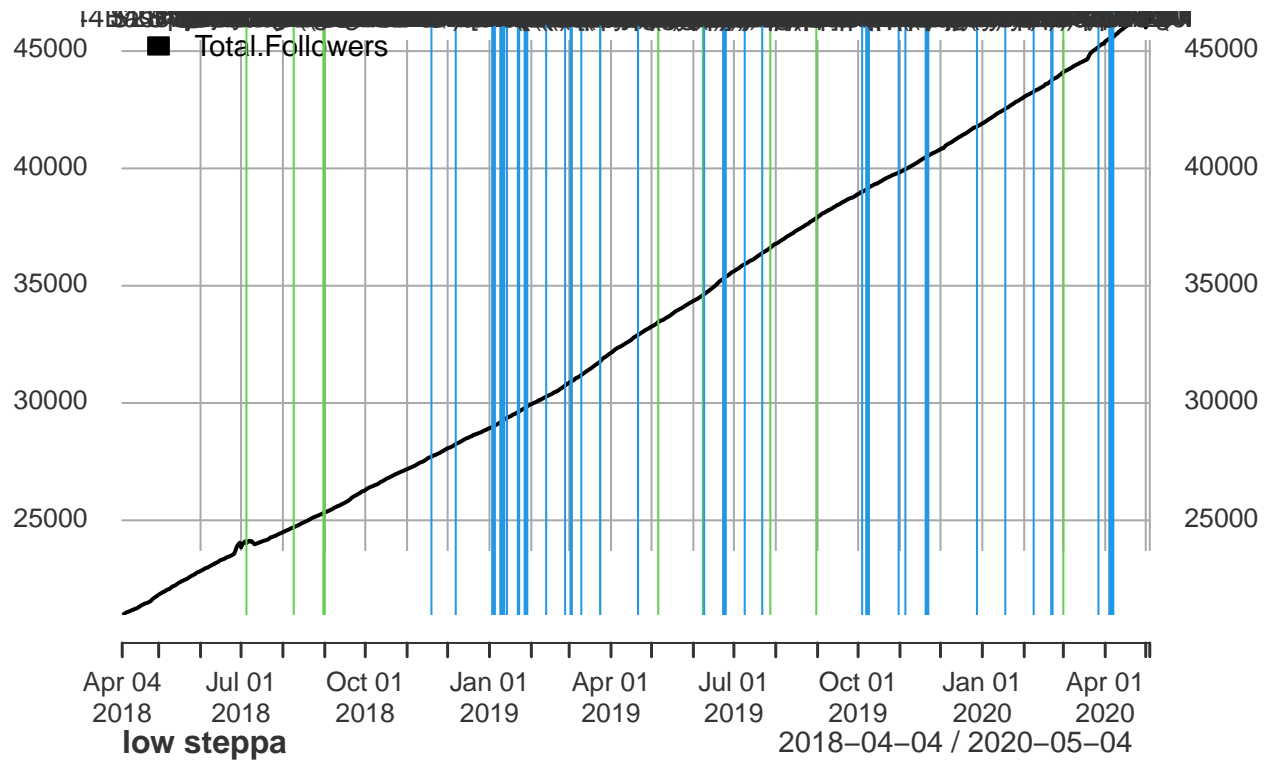
lil texas

2019-04-11 / 2020-05-03



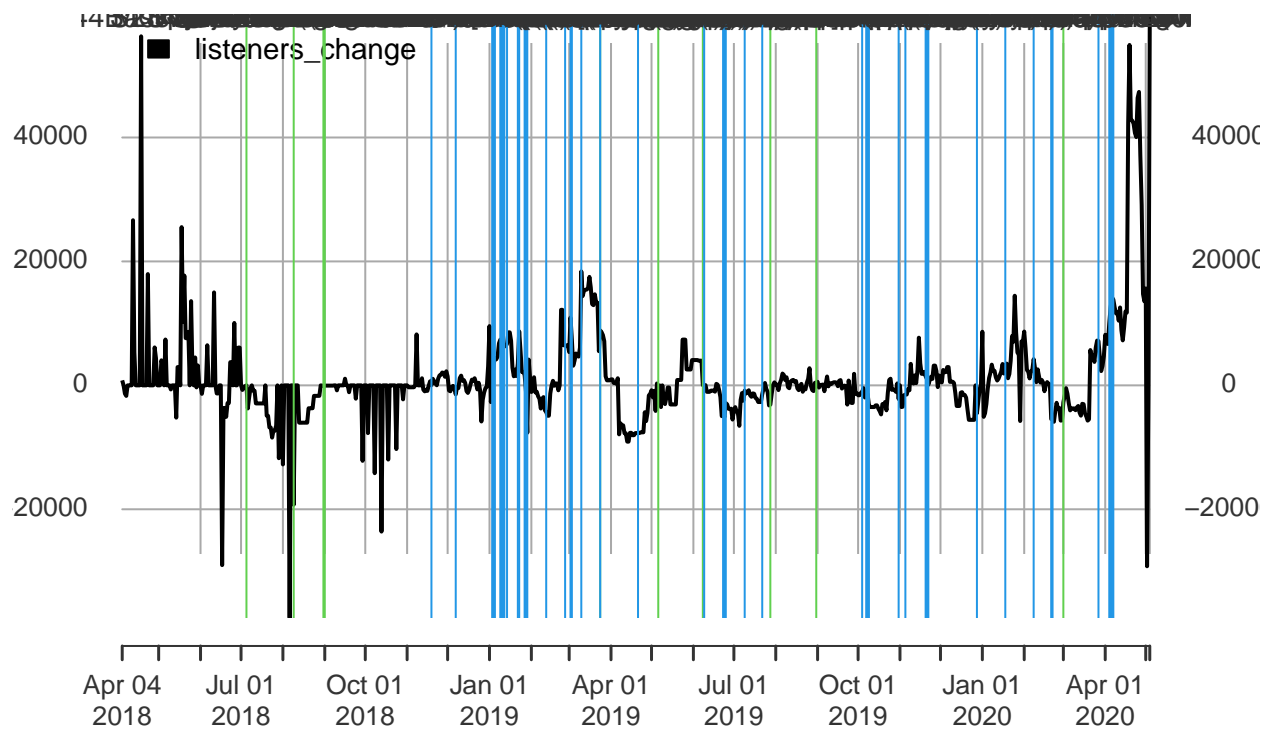
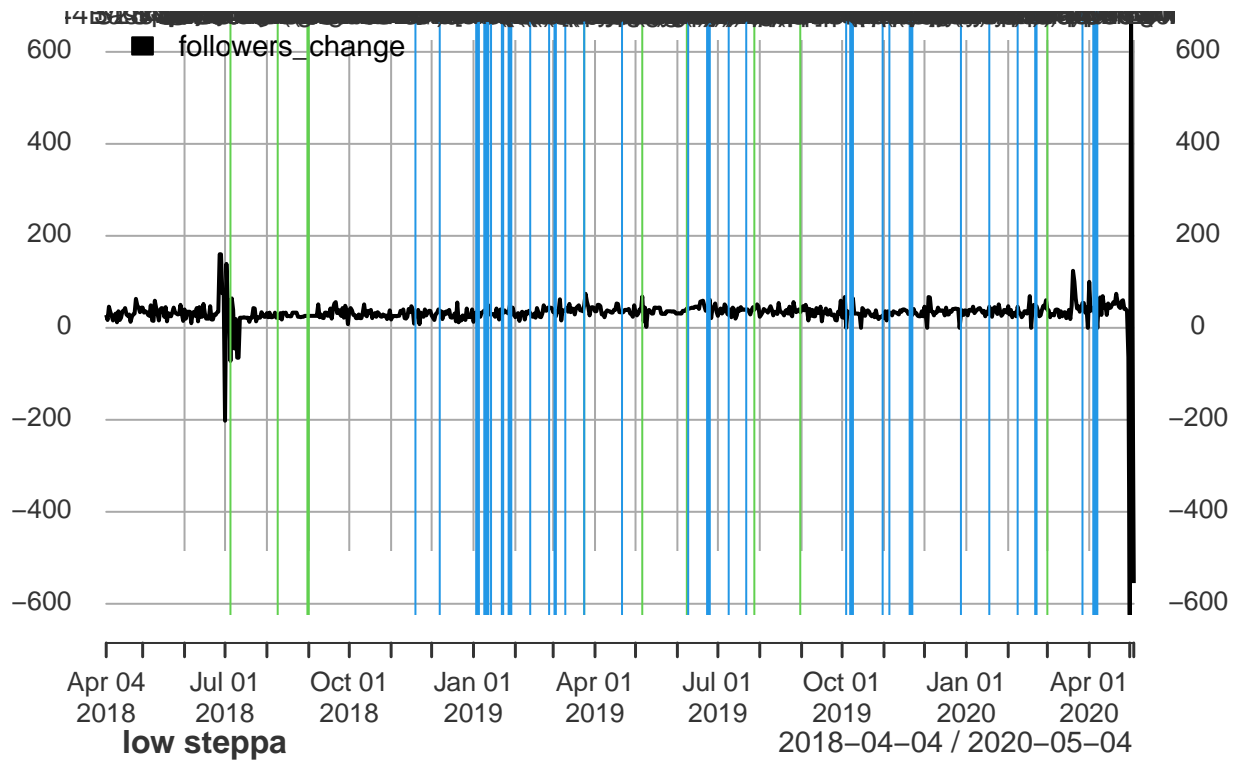
low steppa

2018-04-04 / 2020-05-04



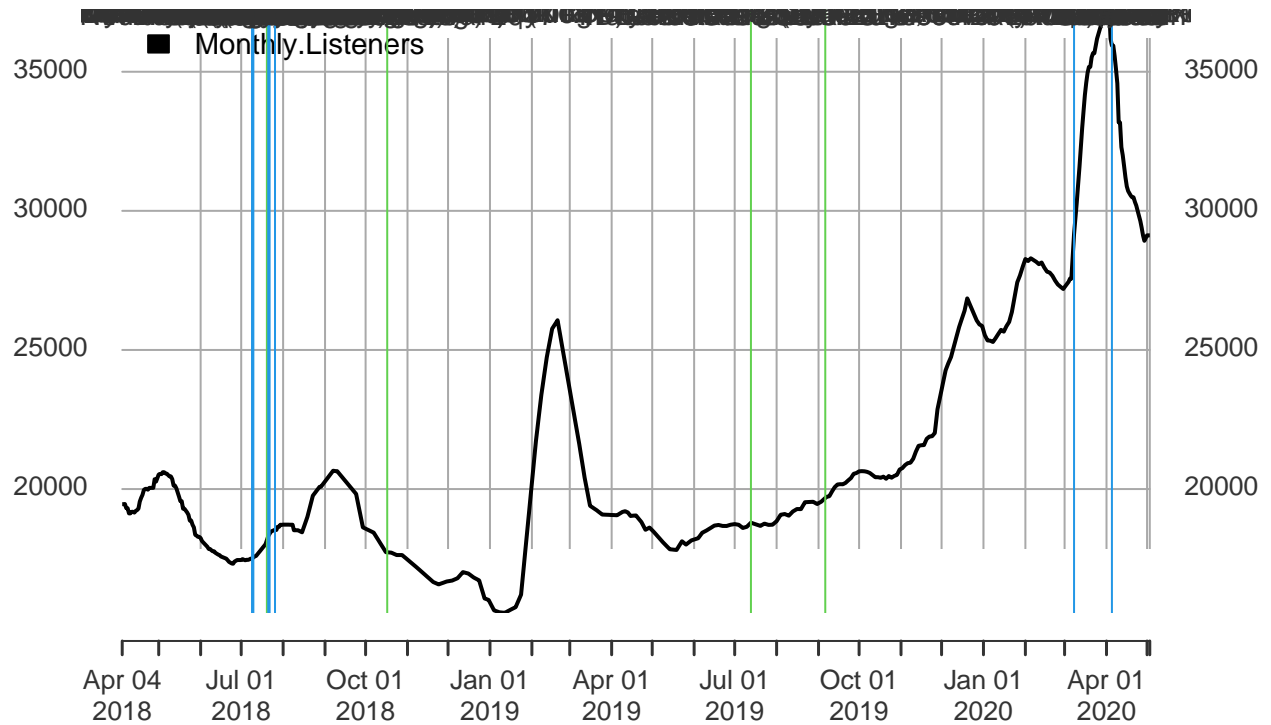
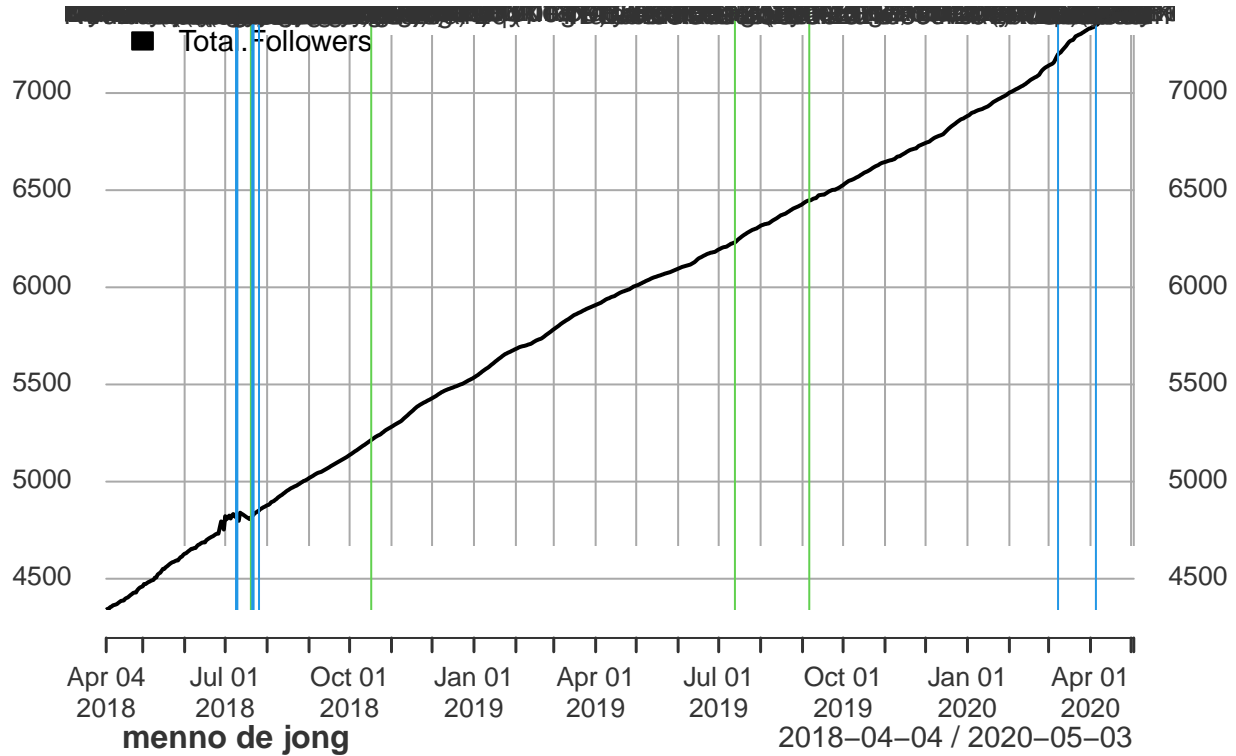
low steppa

2018-04-04 / 2020-05-04



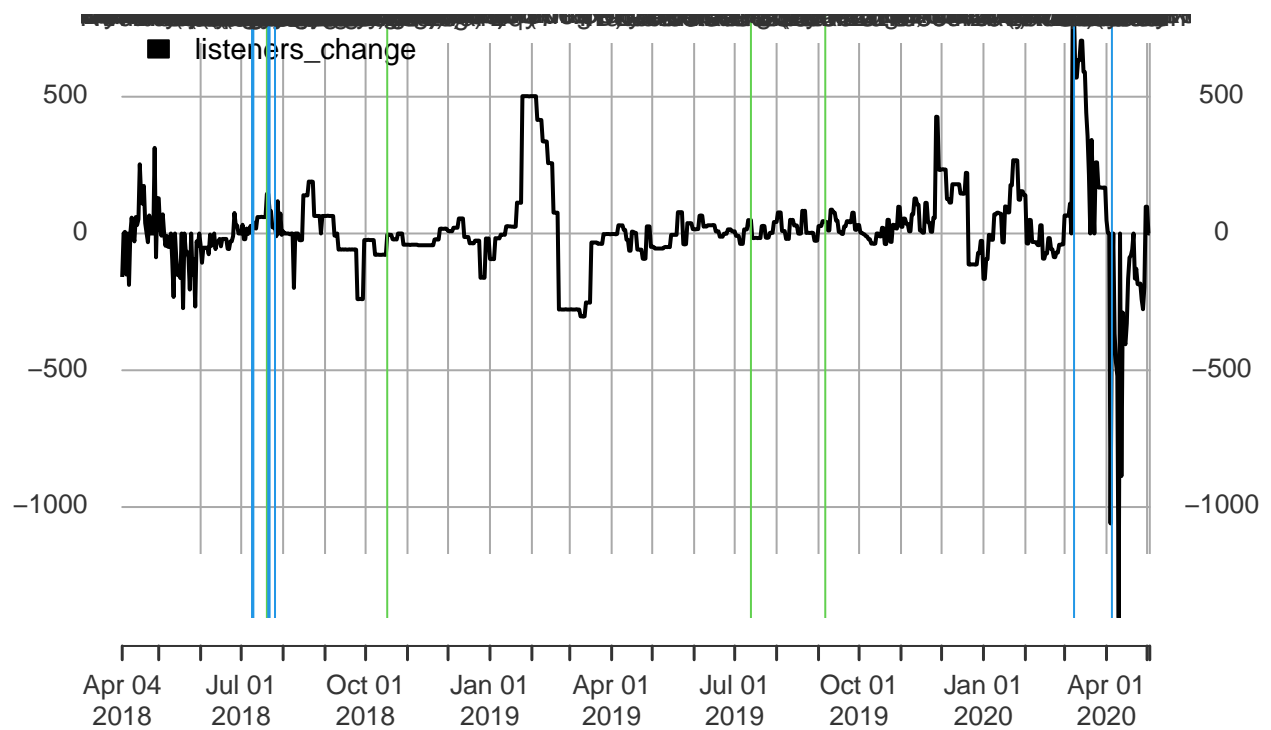
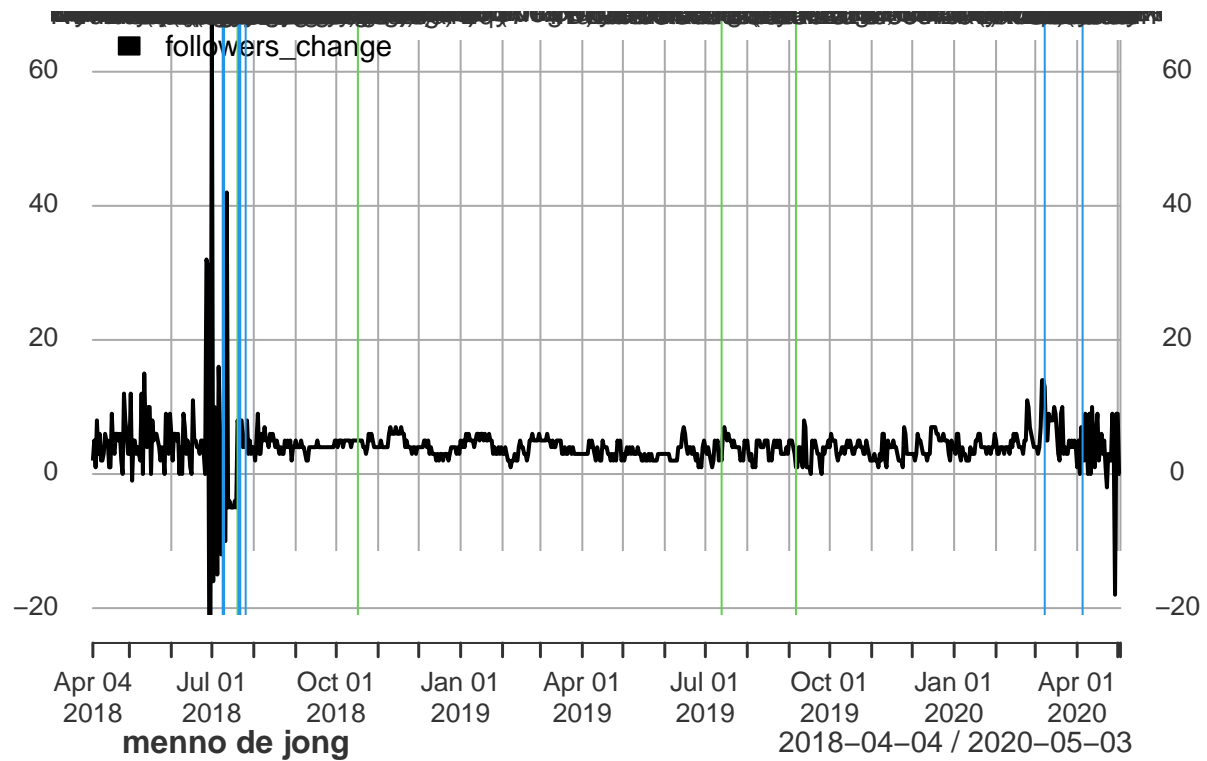
menno de jong

2018-04-04 / 2020-05-03



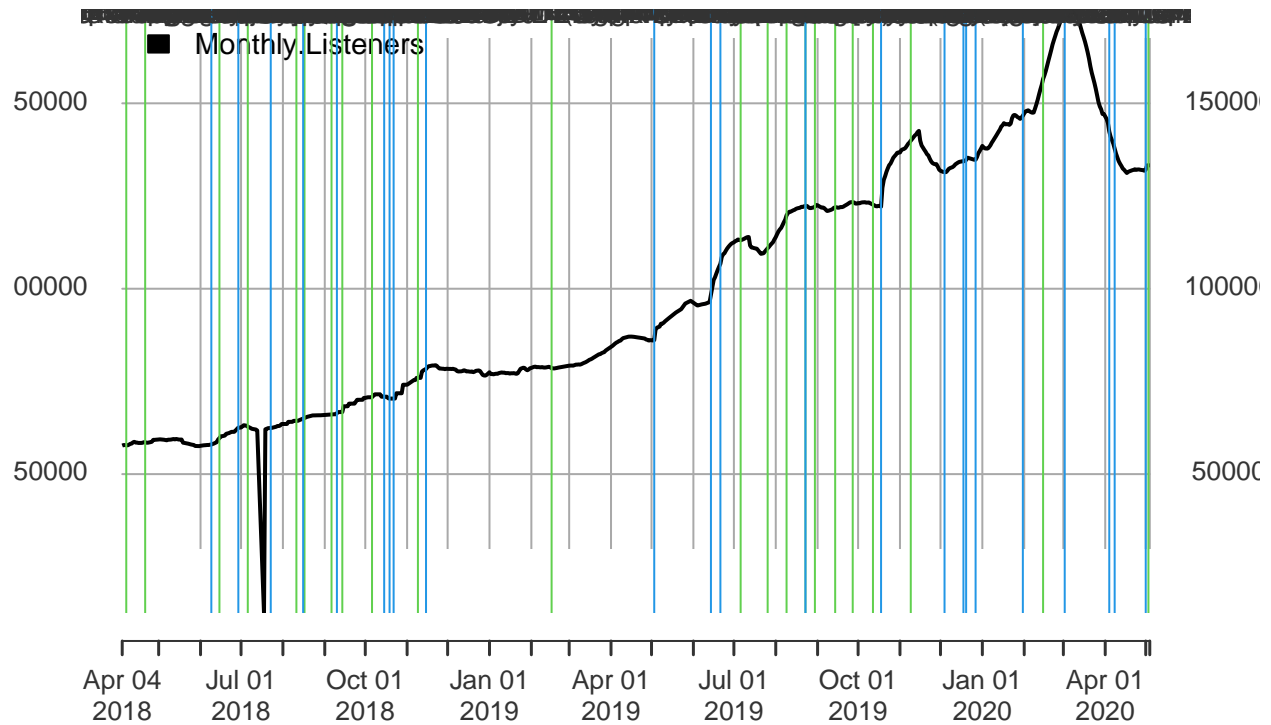
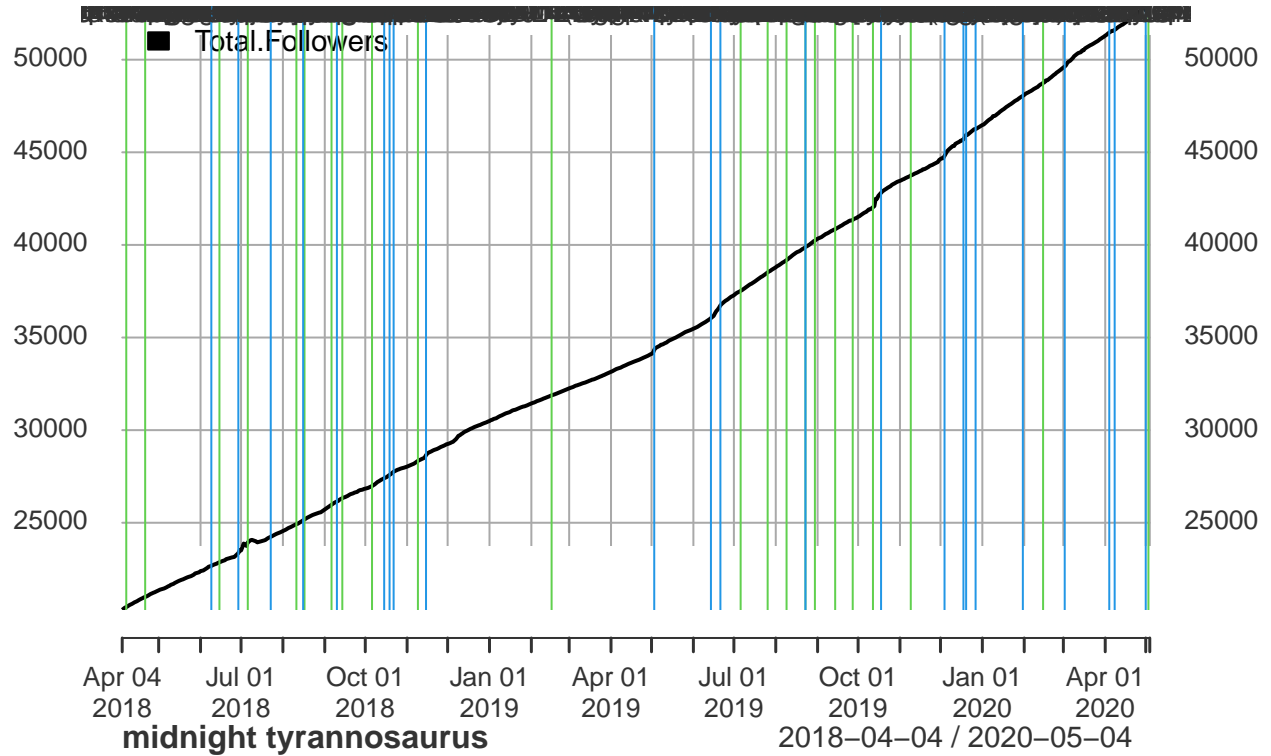
menno de jong

2018-04-04 / 2020-05-03



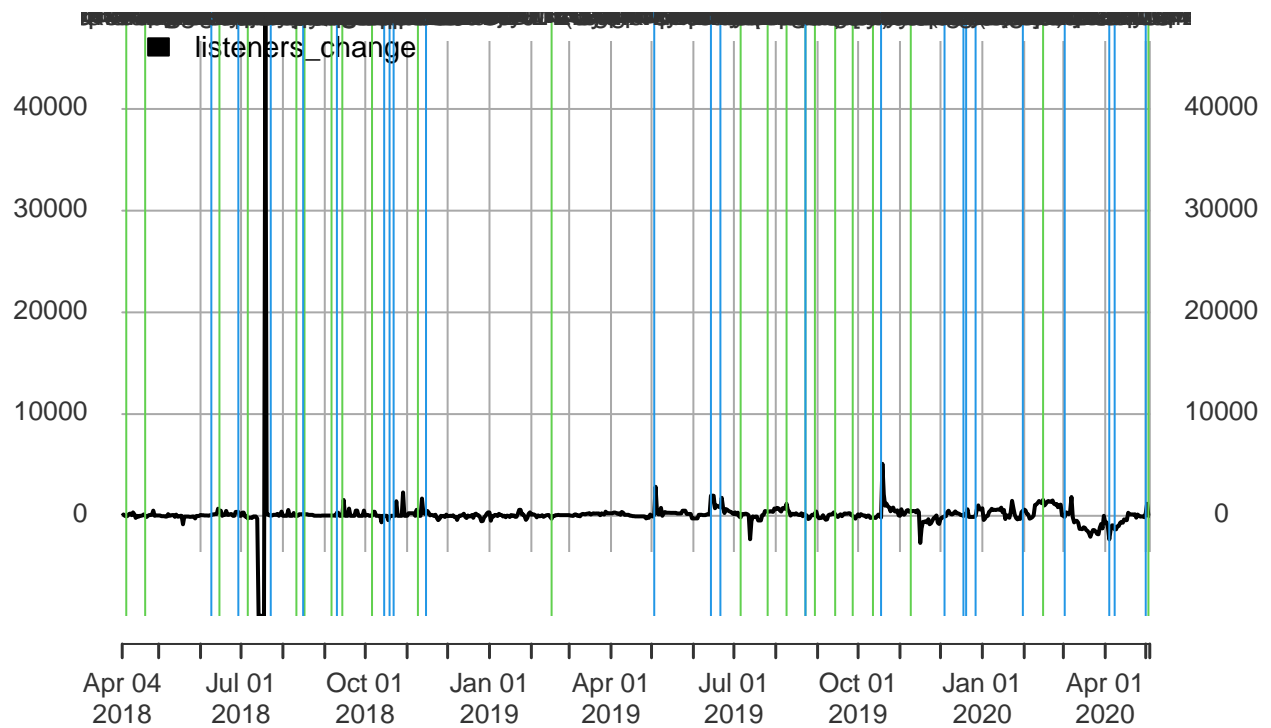
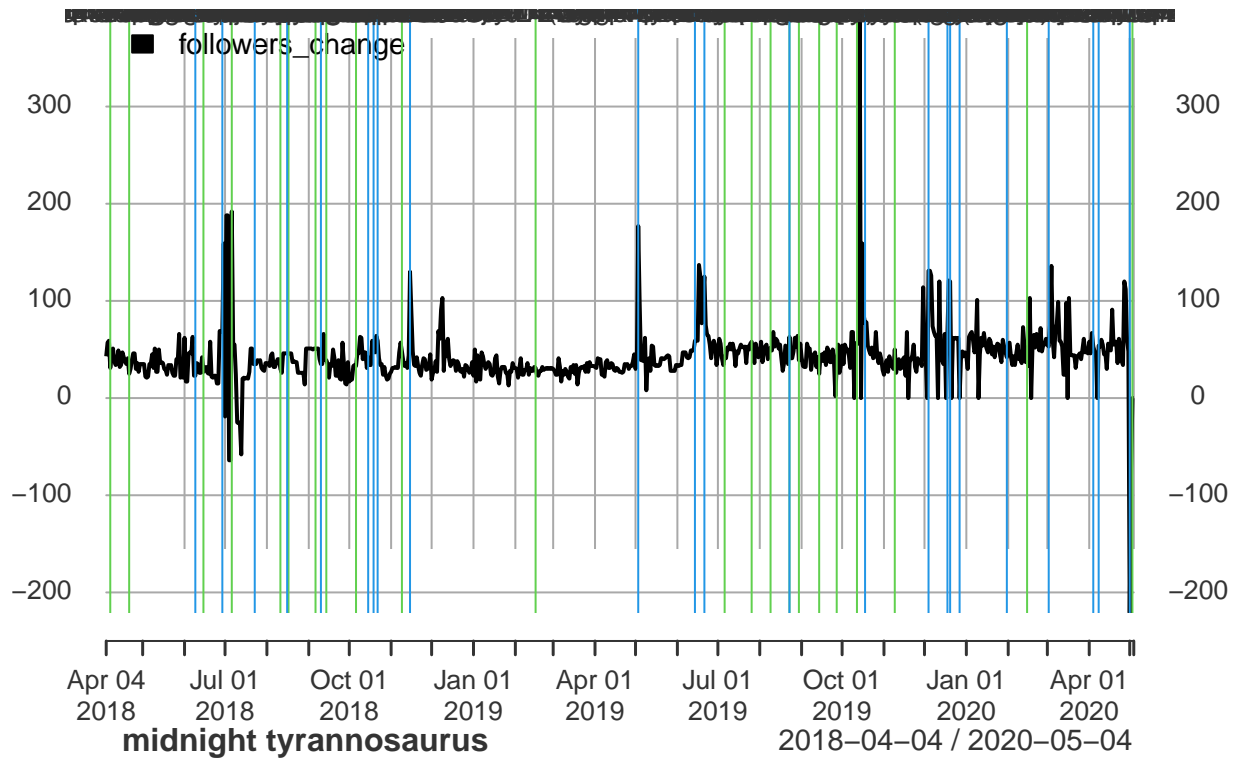
midnight tyrannosaurus

2018-04-04 / 2020-05-04



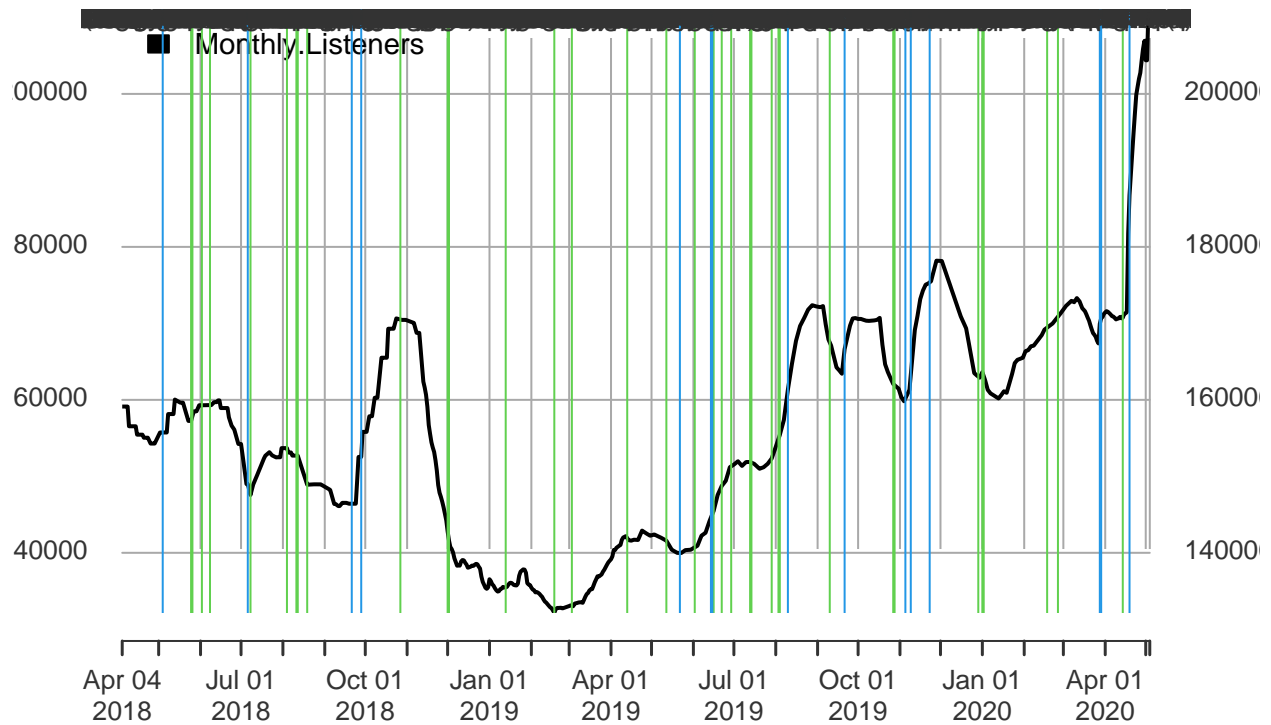
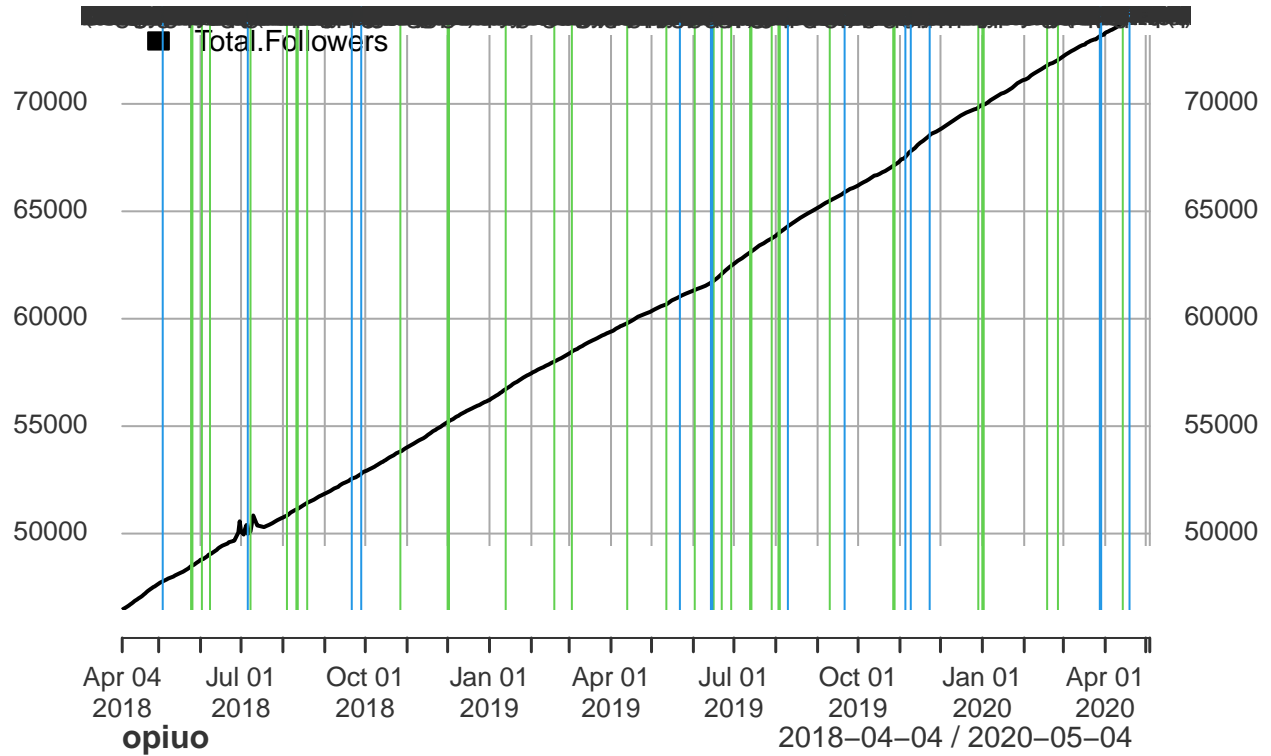
midnight tyrannosaurus

2018-04-04 / 2020-05-04



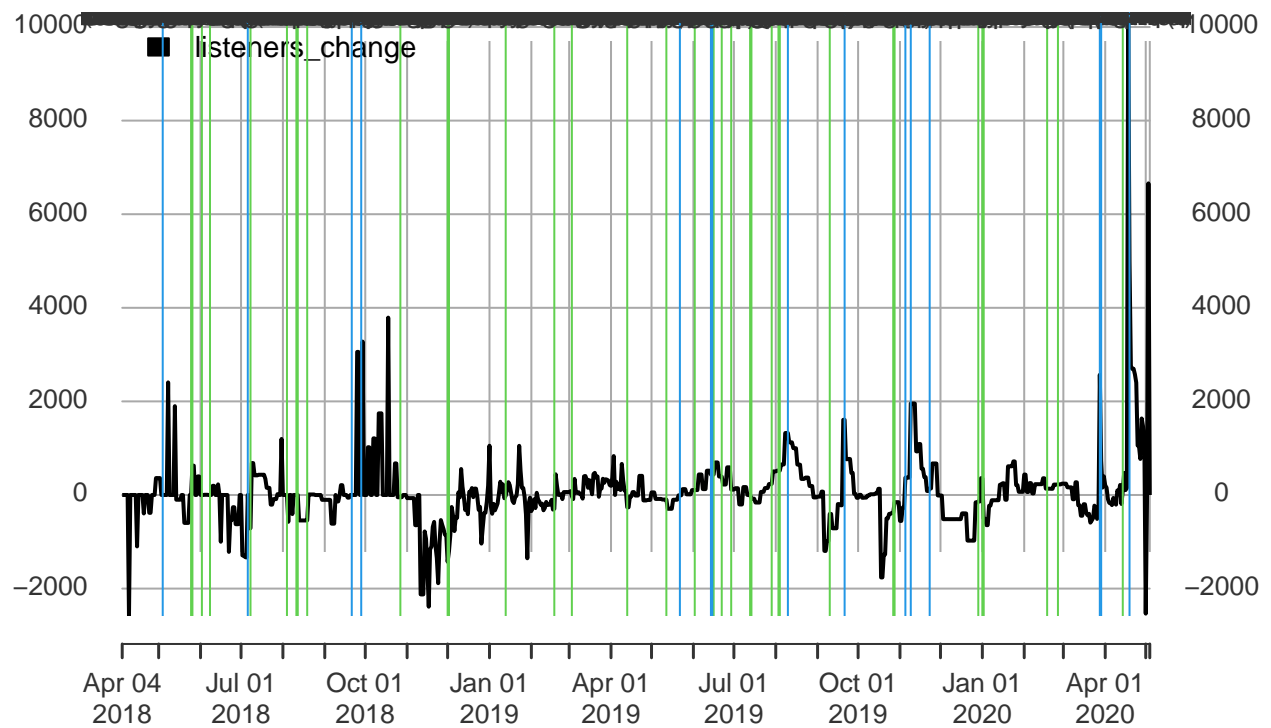
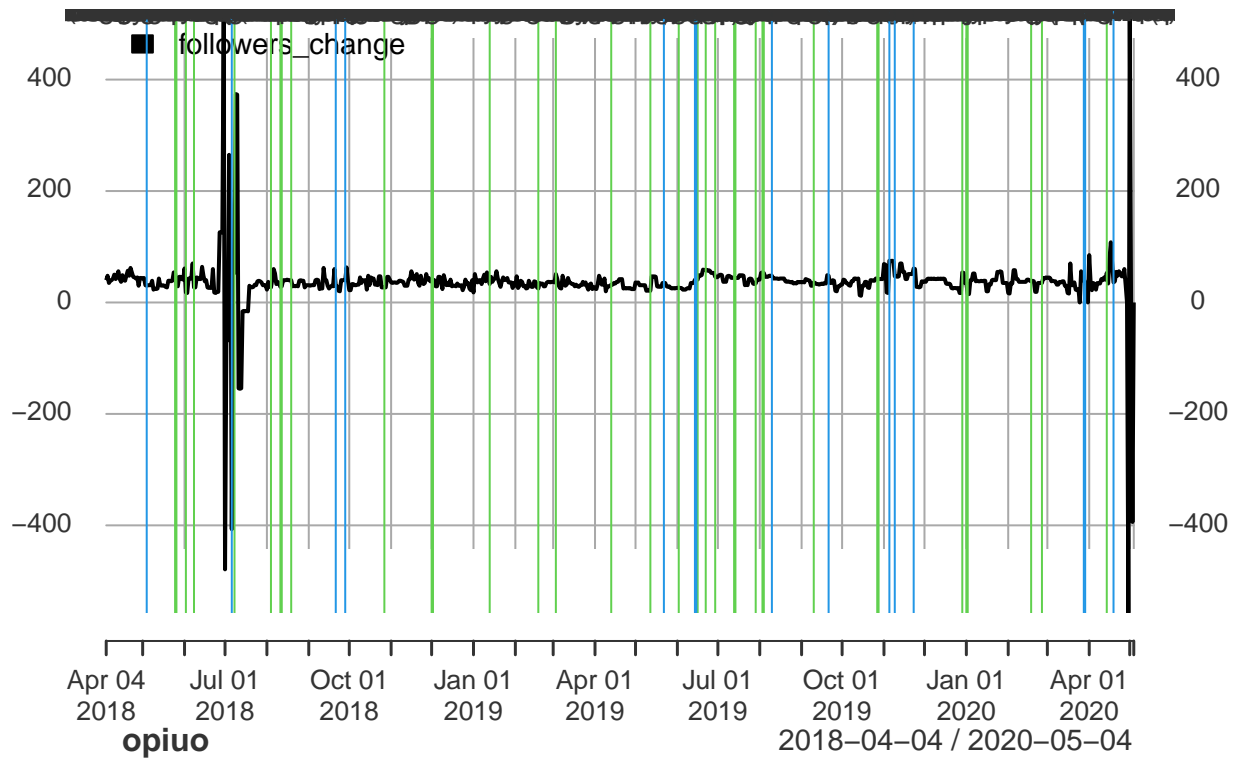
opiuo

2018-04-04 / 2020-05-04



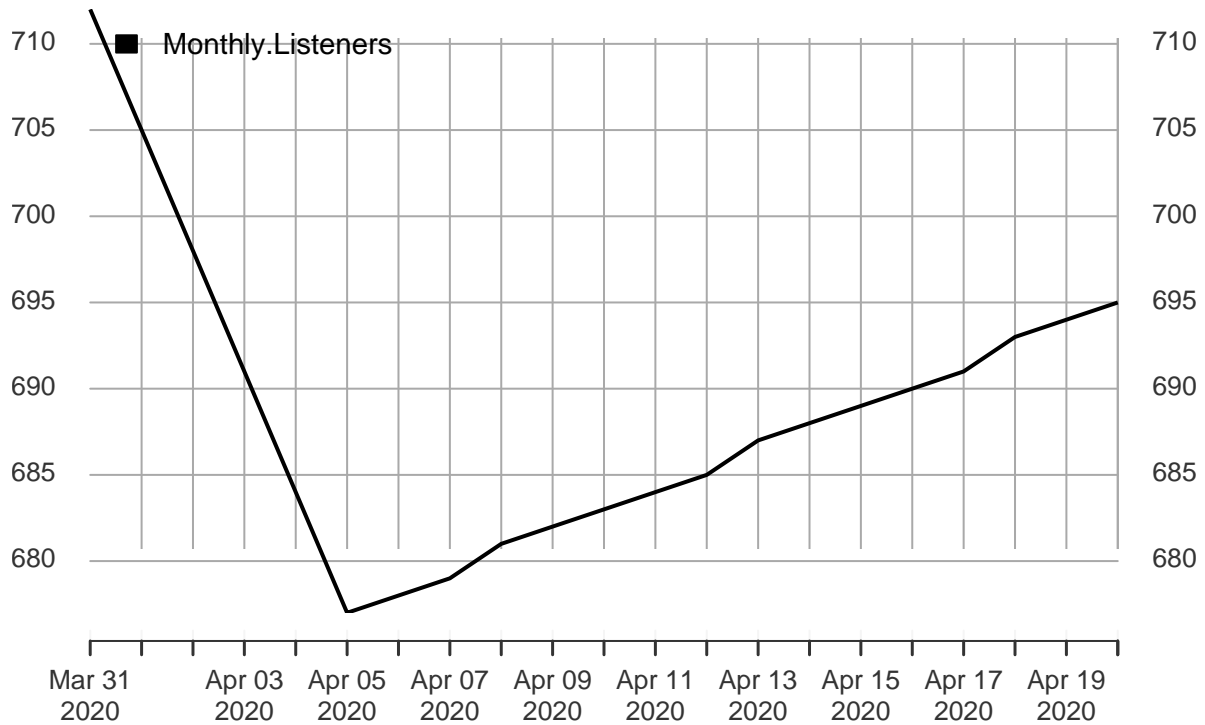
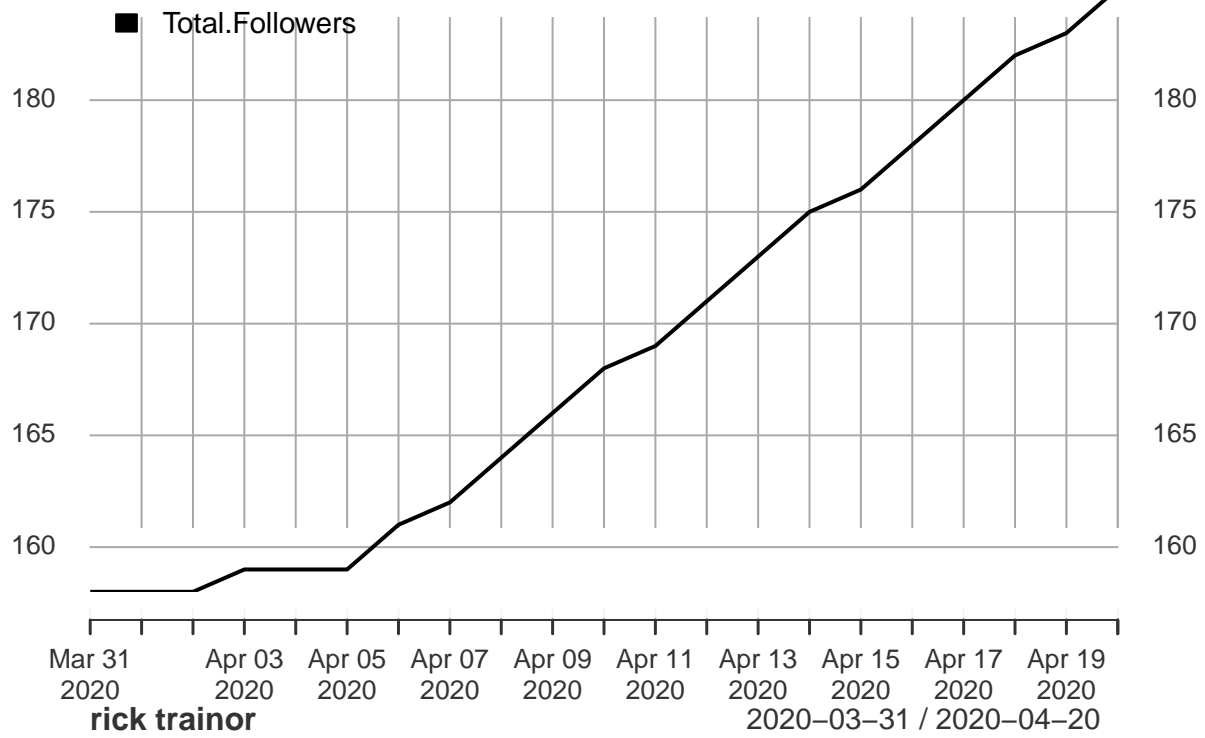
opiuo

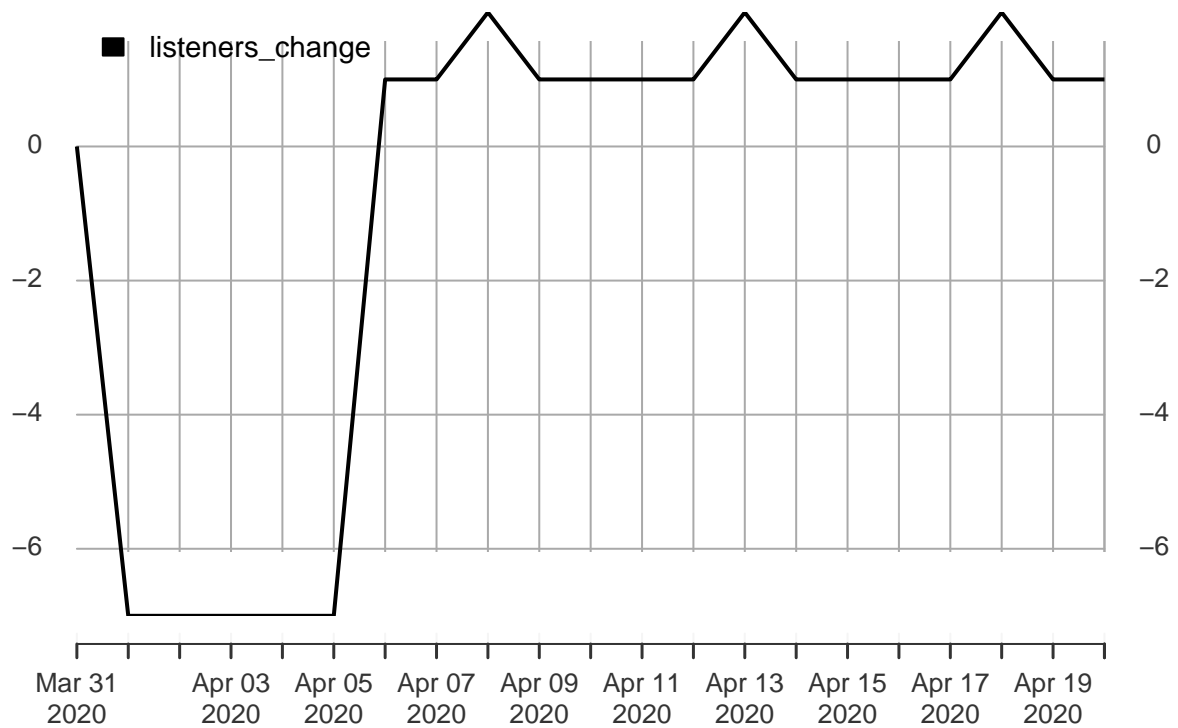
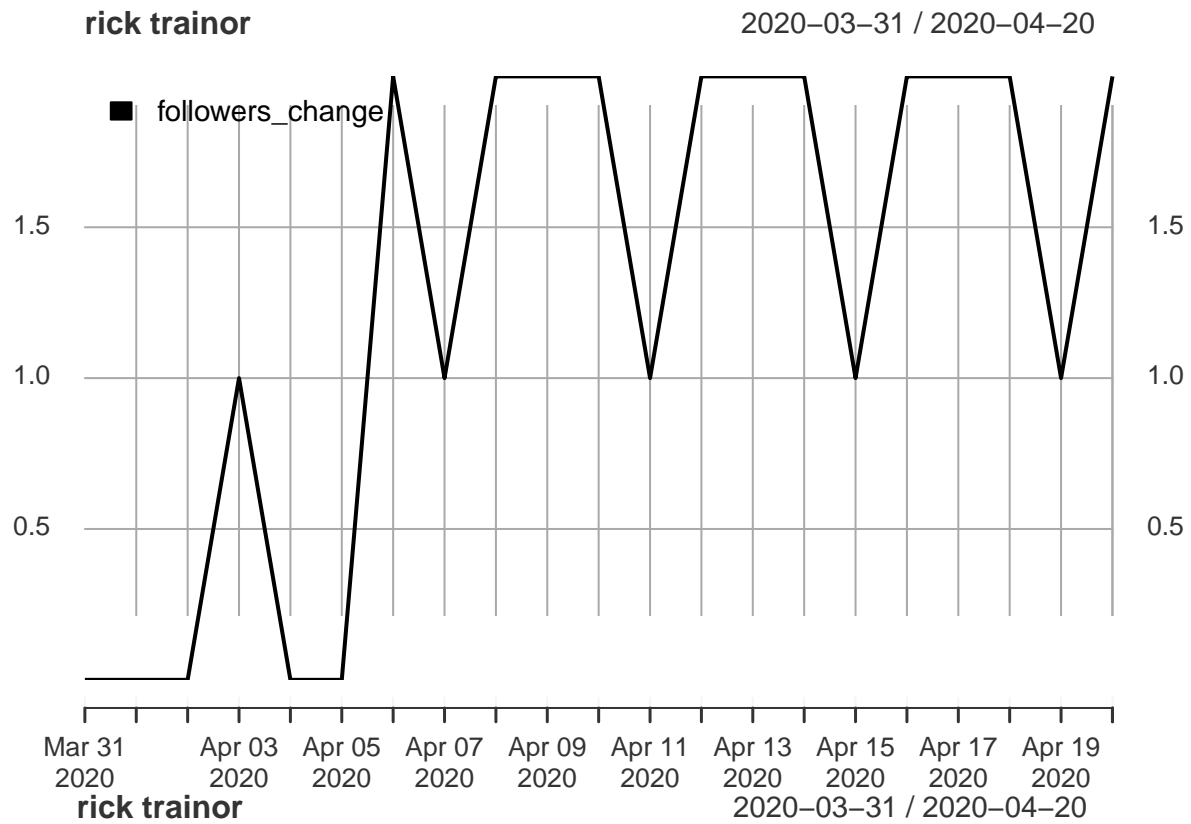
2018-04-04 / 2020-05-04



rick trainor

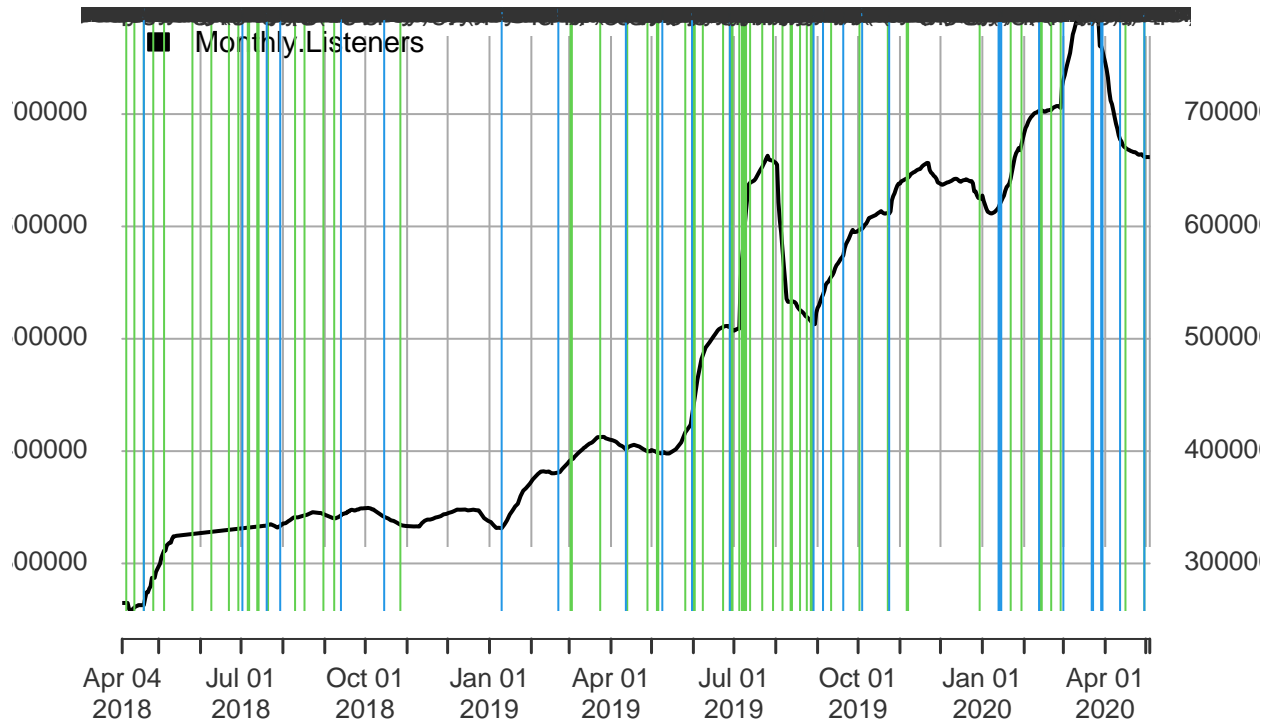
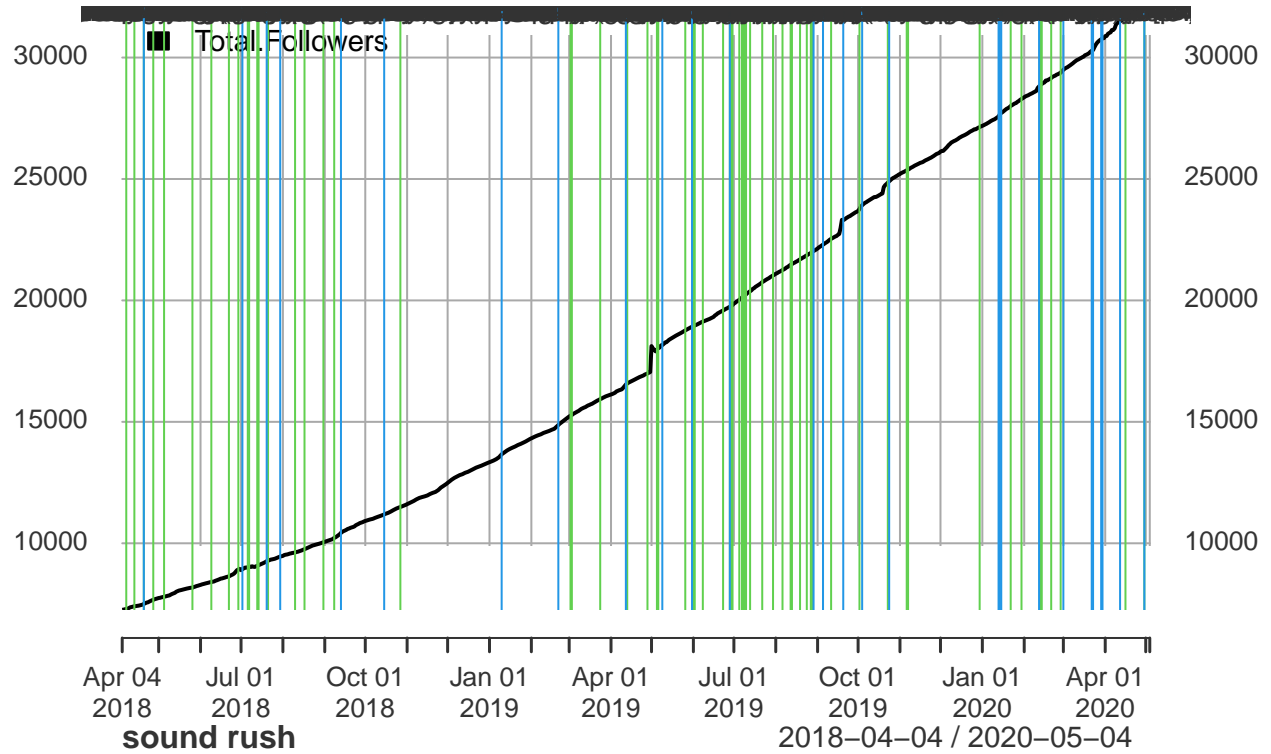
2020-03-31 / 2020-04-20





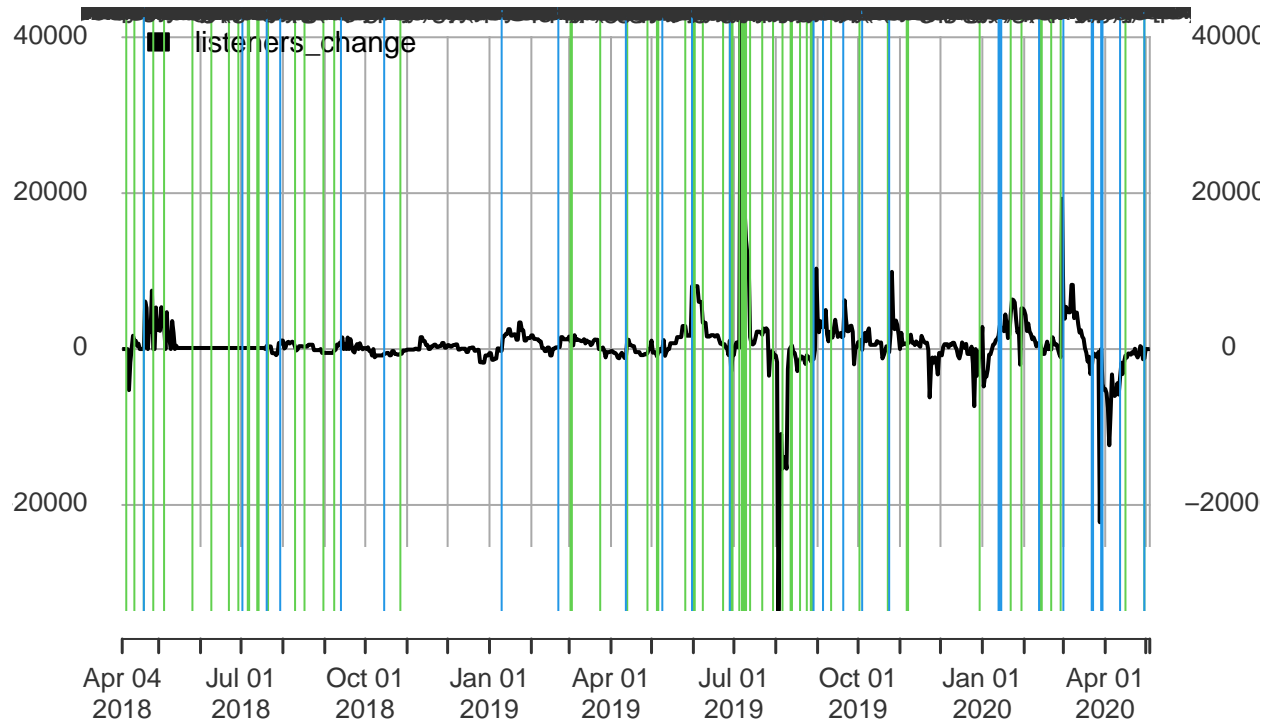
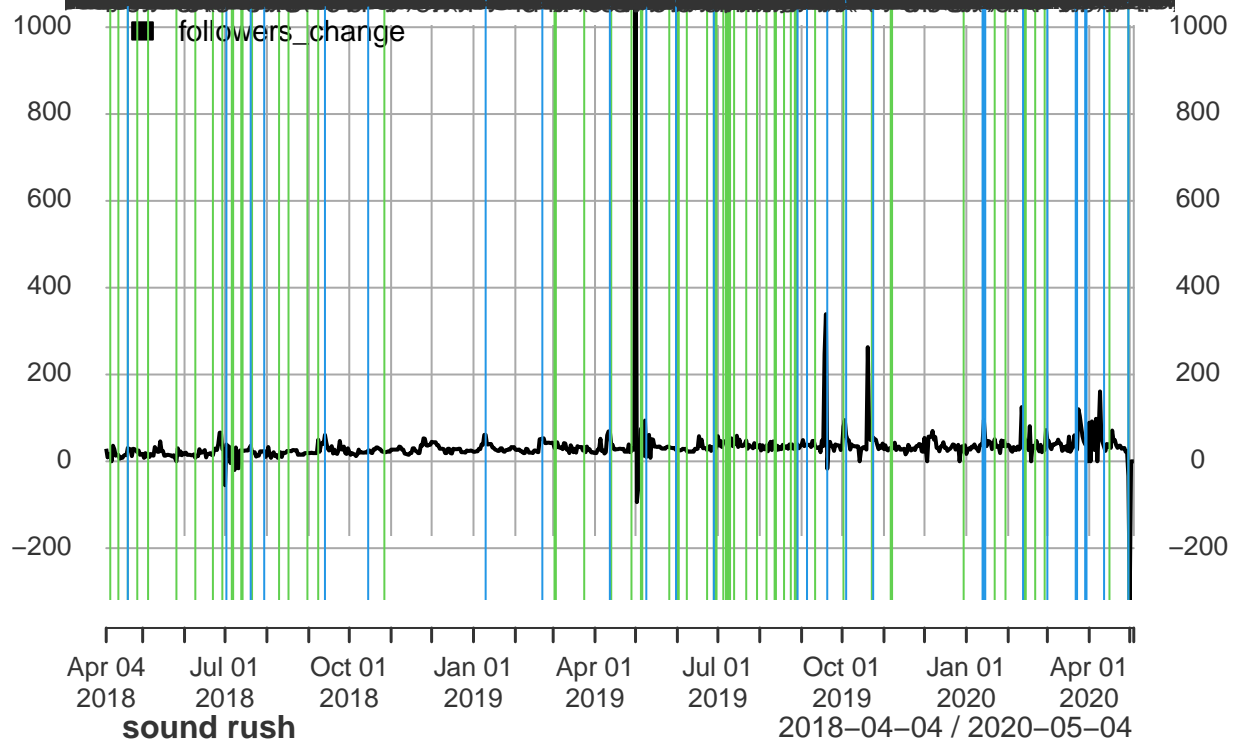
sound rush

2018-04-04 / 2020-05-04



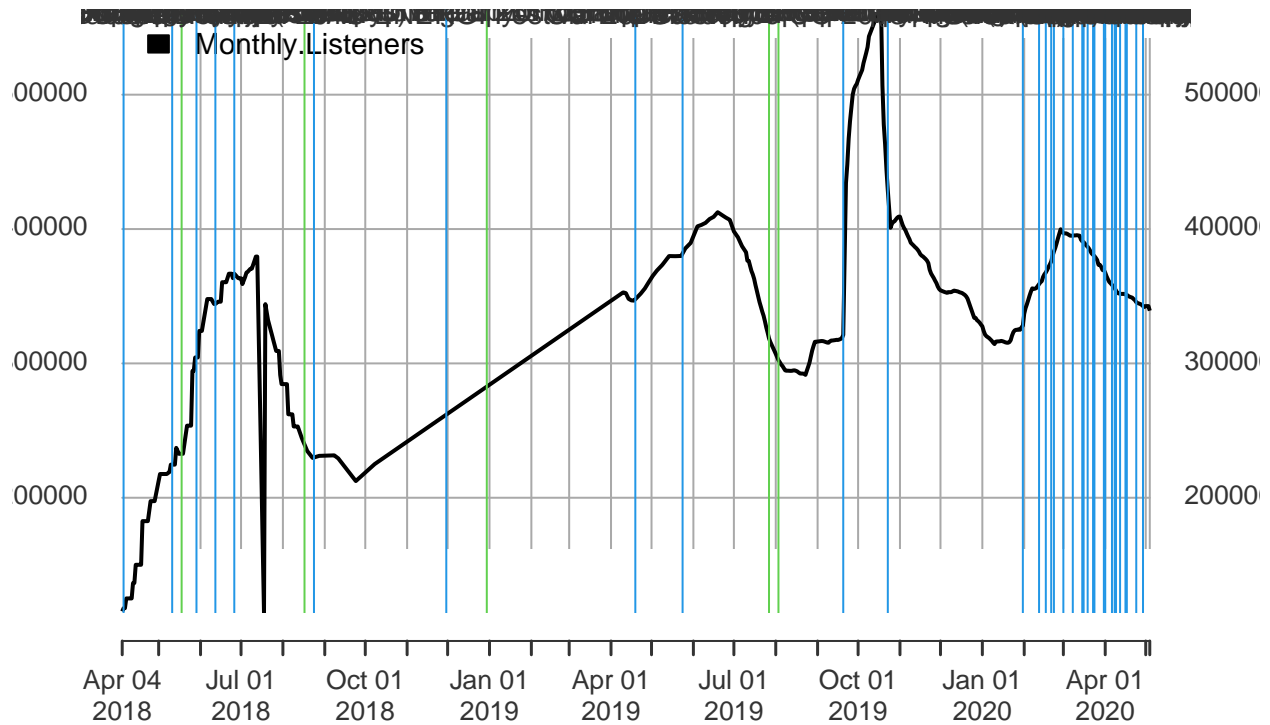
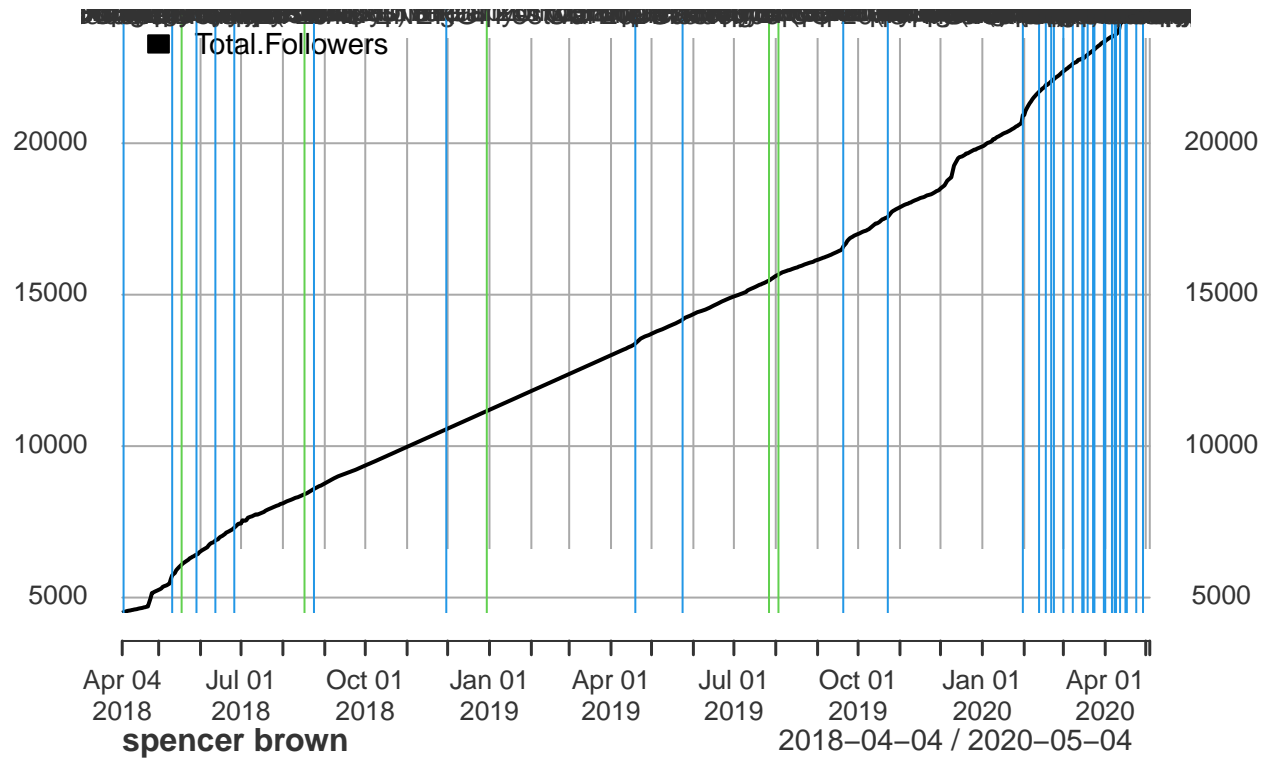
sound rush

2018-04-04 / 2020-05-04



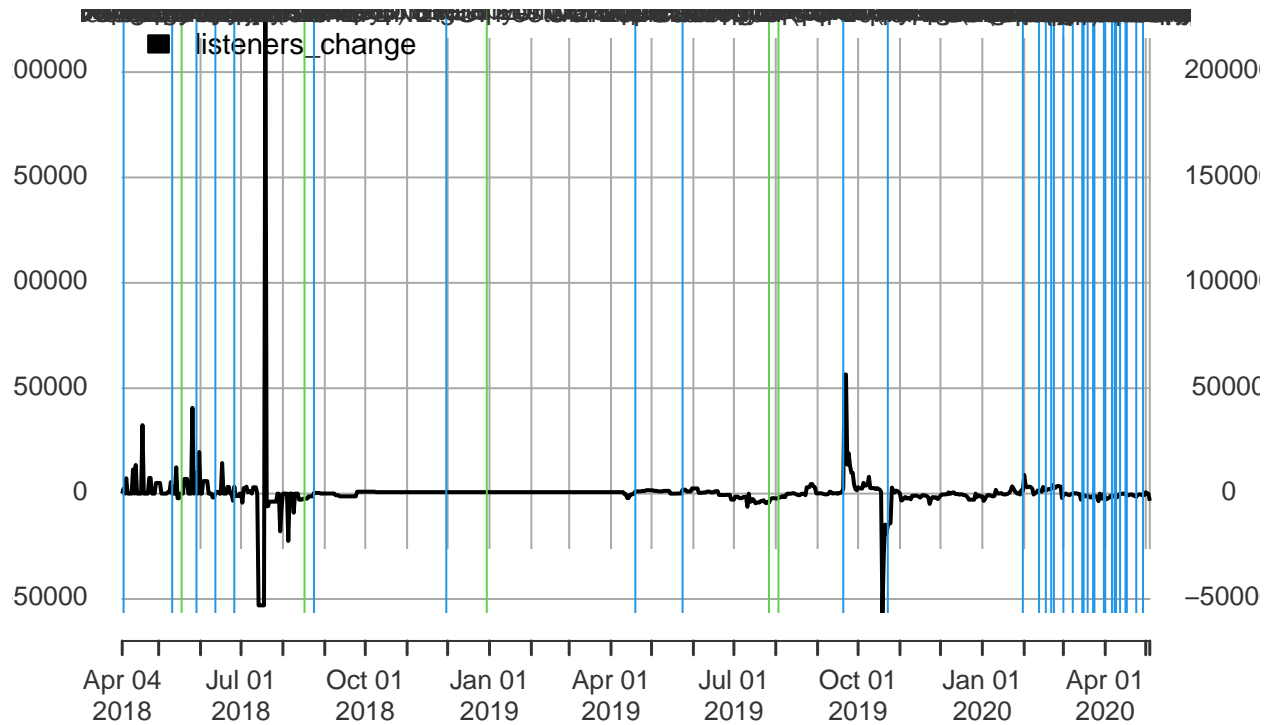
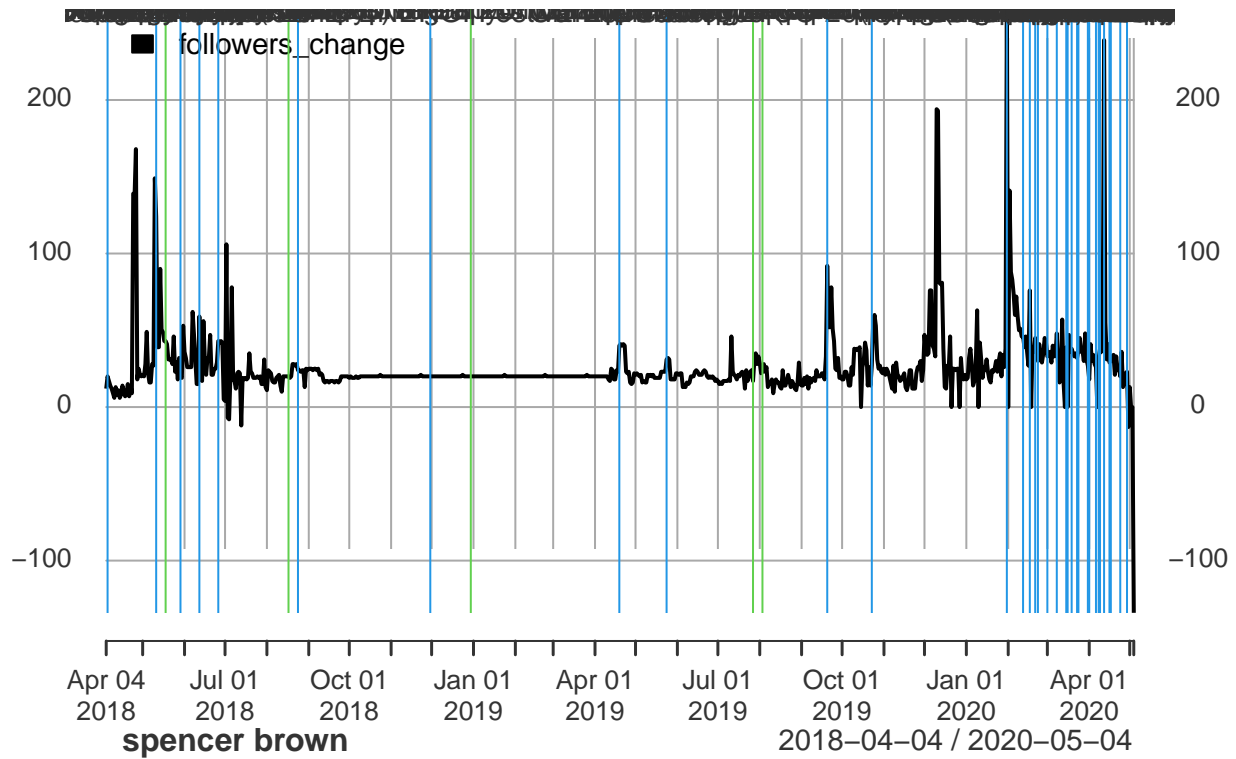
spencer brown

2018-04-04 / 2020-05-04



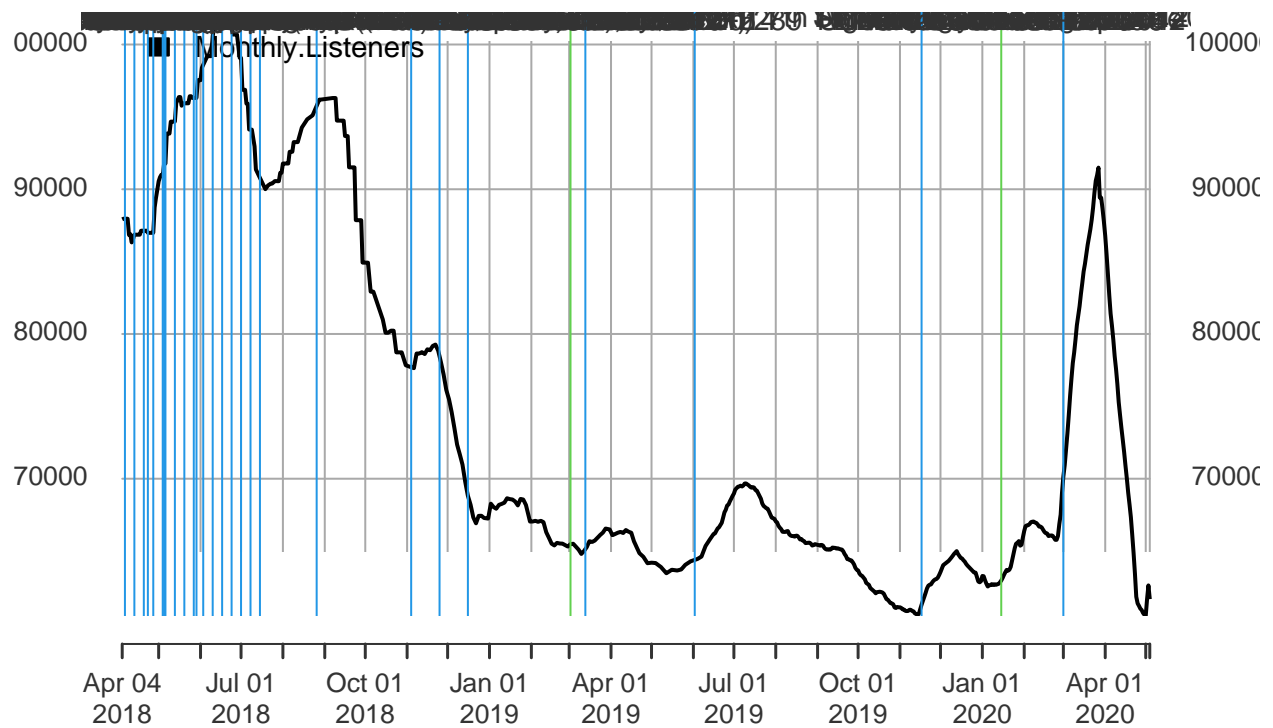
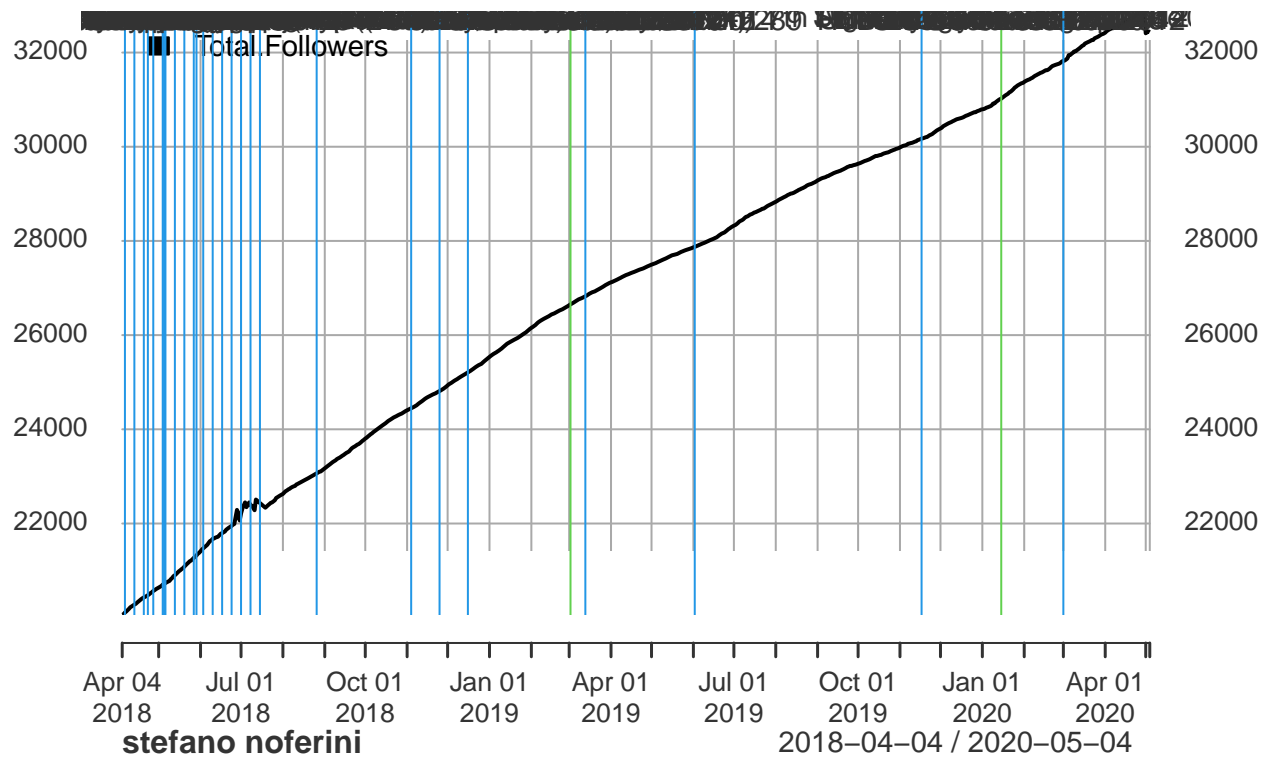
spencer brown

2018-04-04 / 2020-05-04



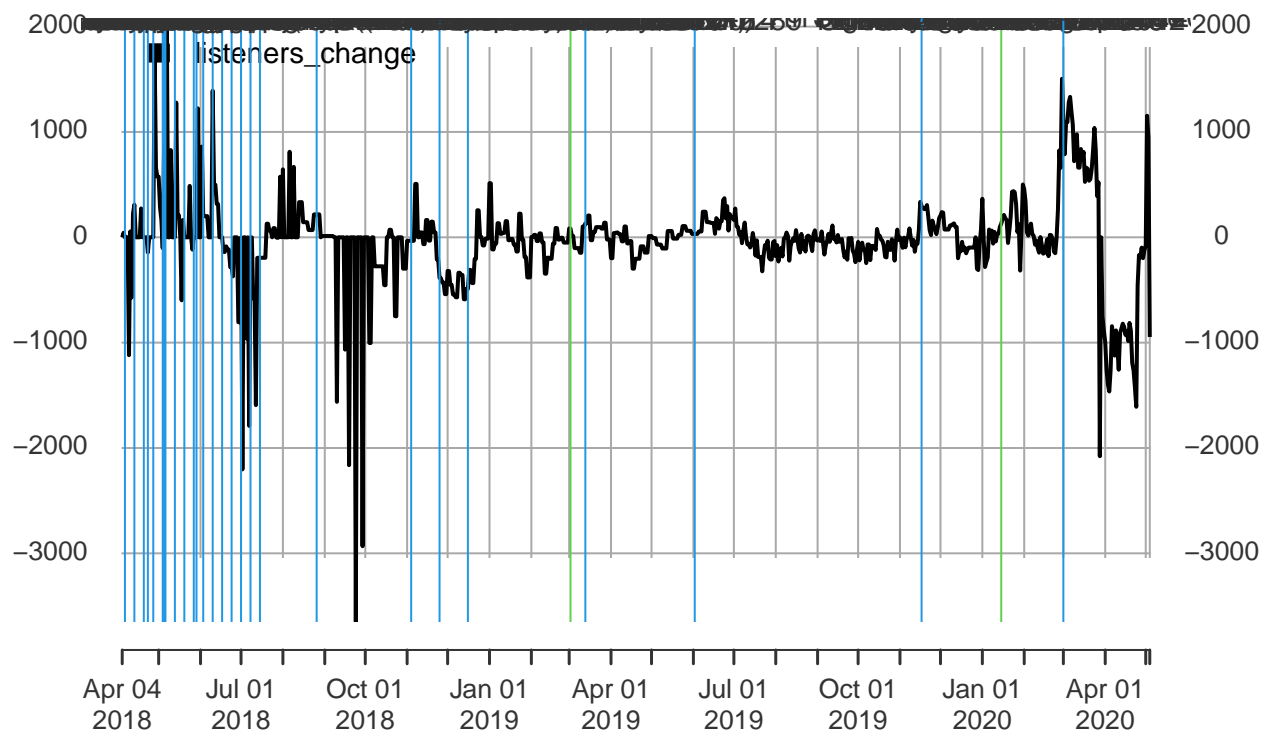
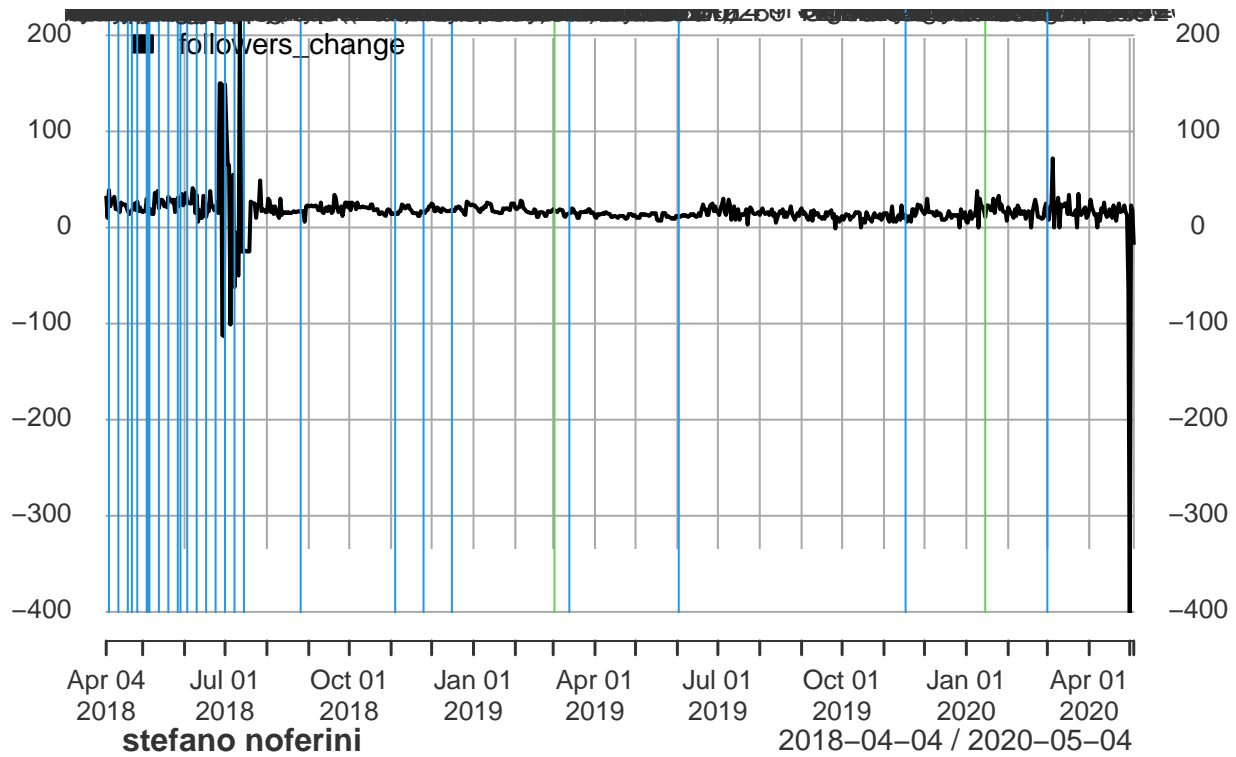
stefano noferini

2018-04-04 / 2020-05-04



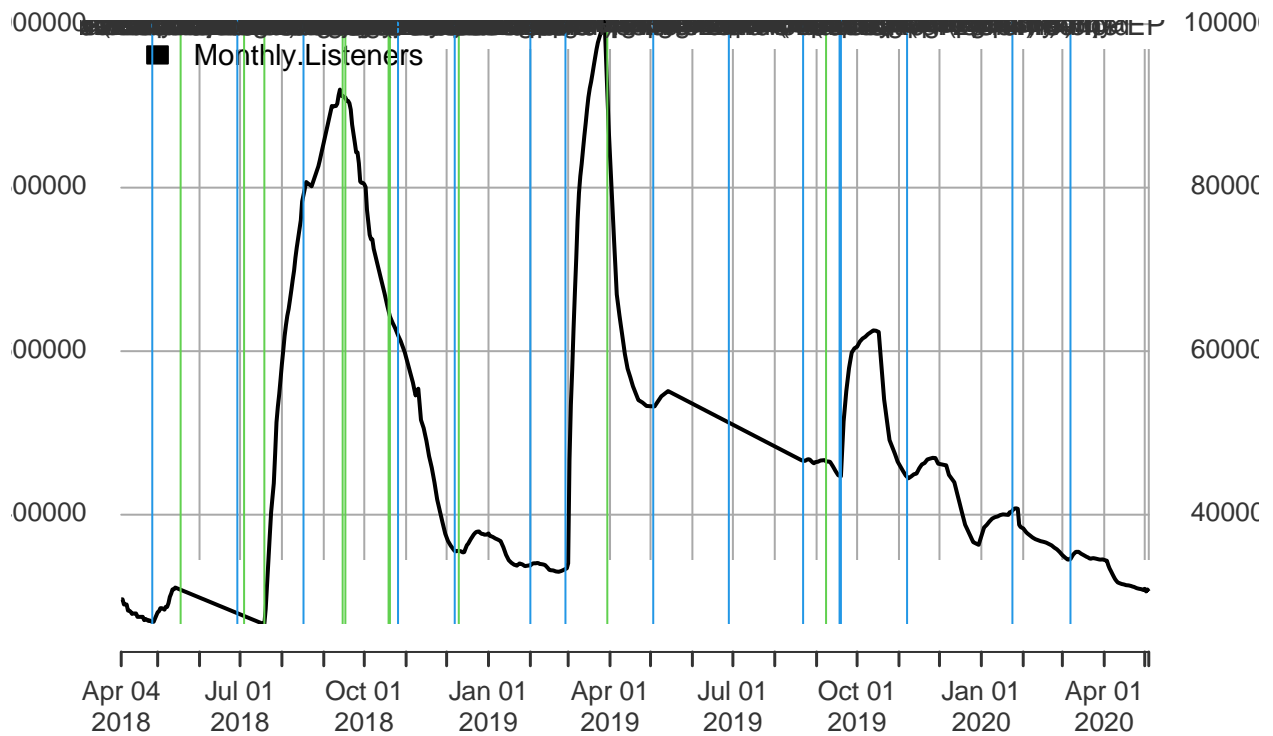
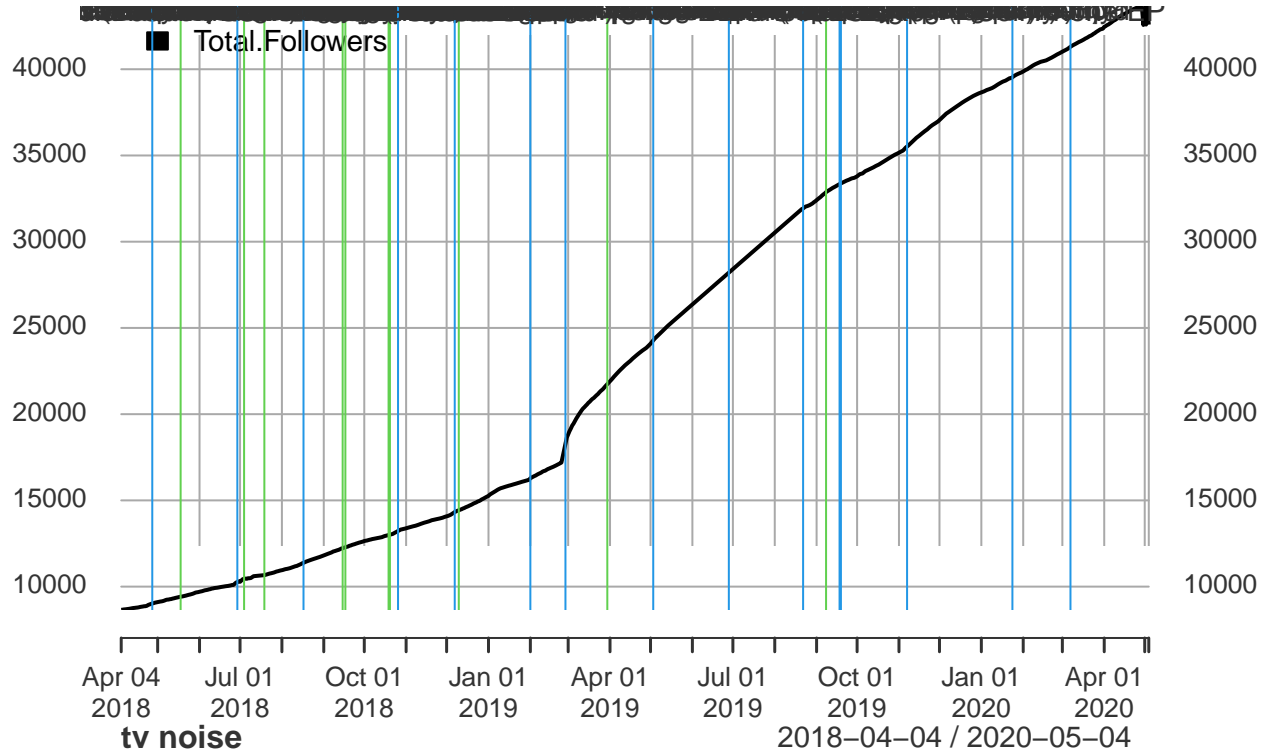
stefano noferini

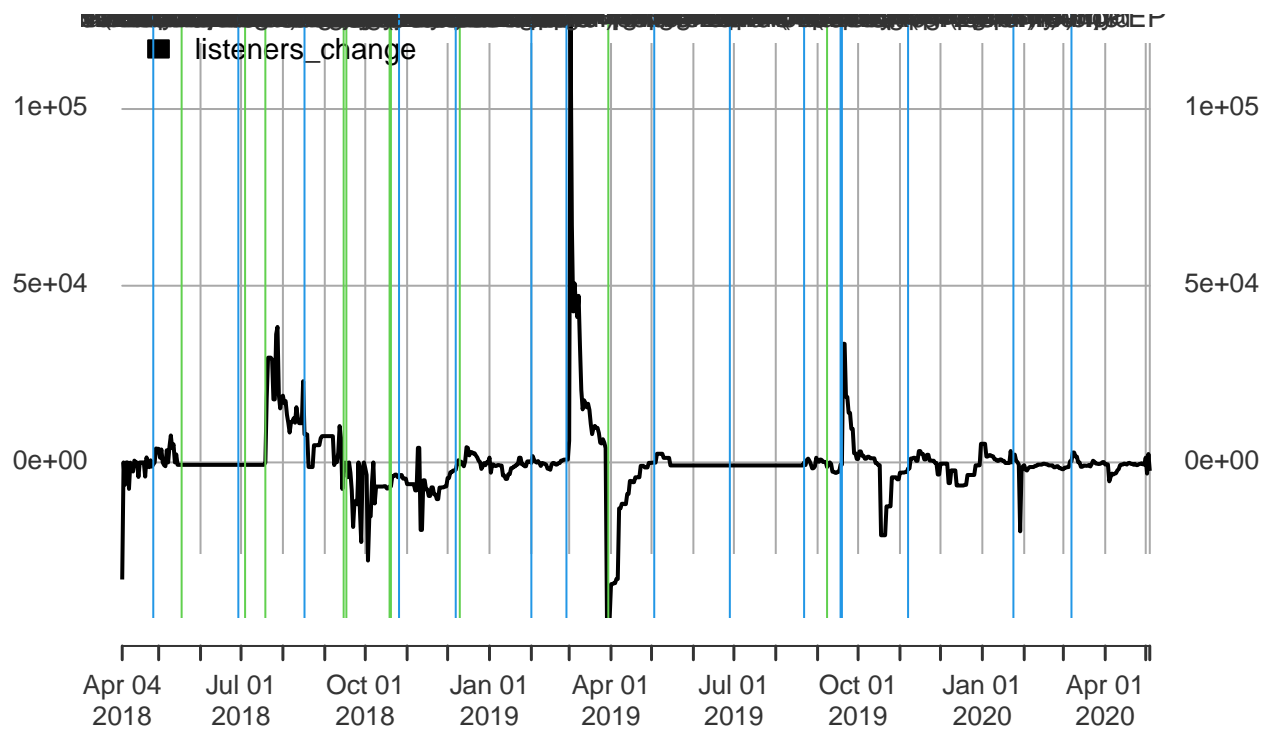
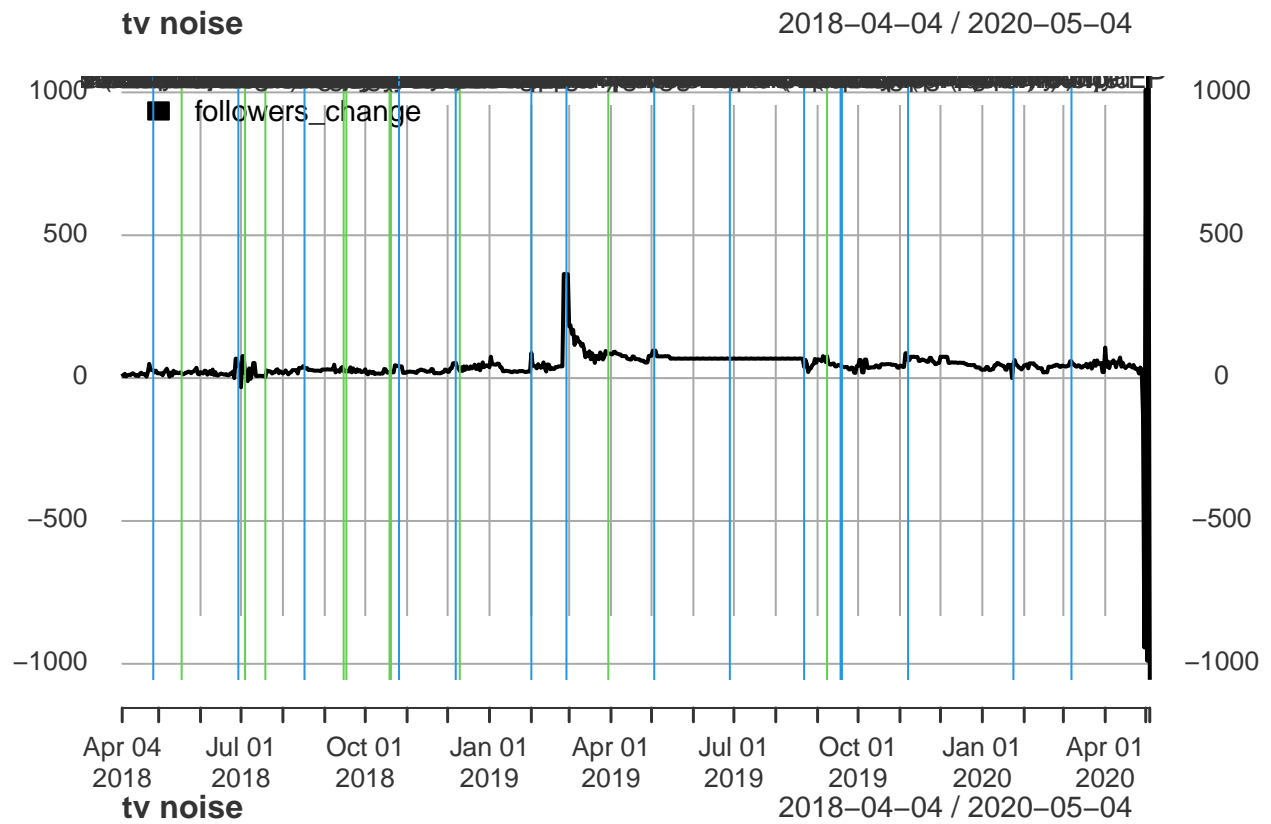
2018-04-04 / 2020-05-04



tv noise

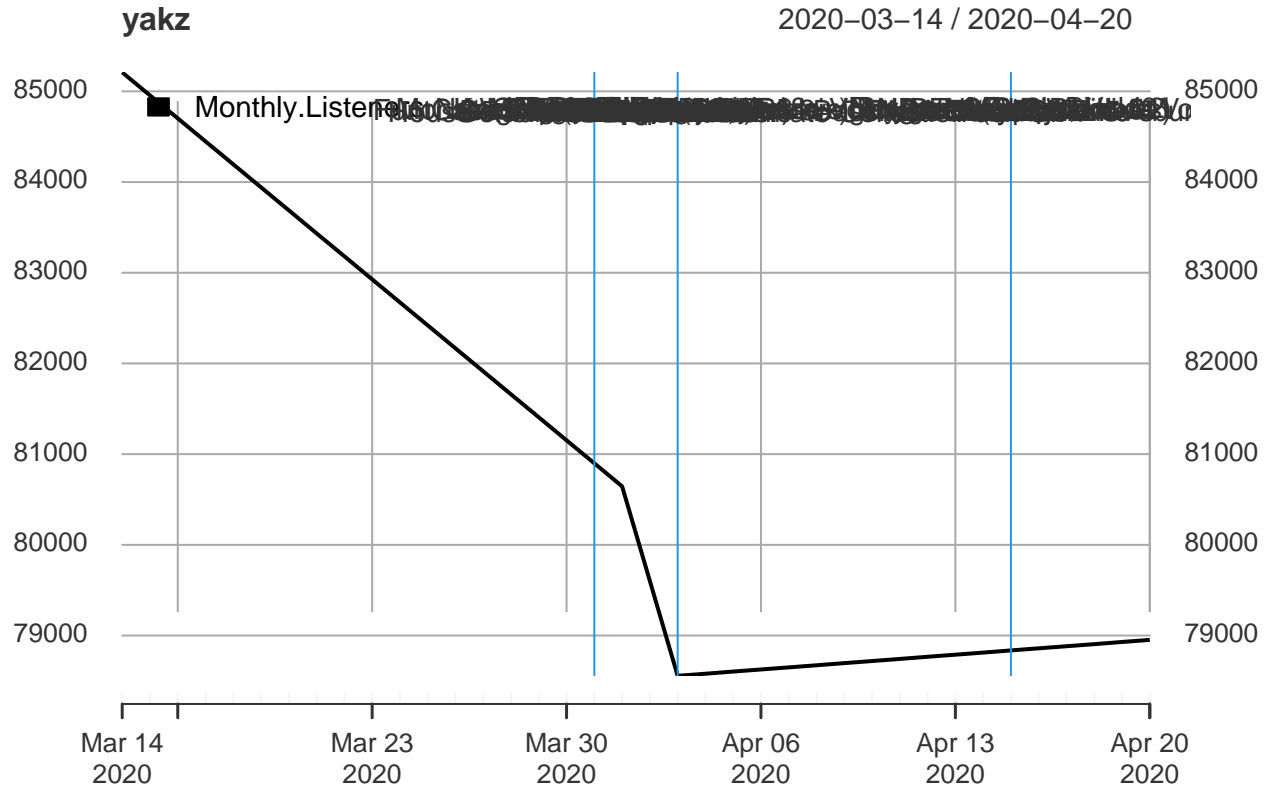
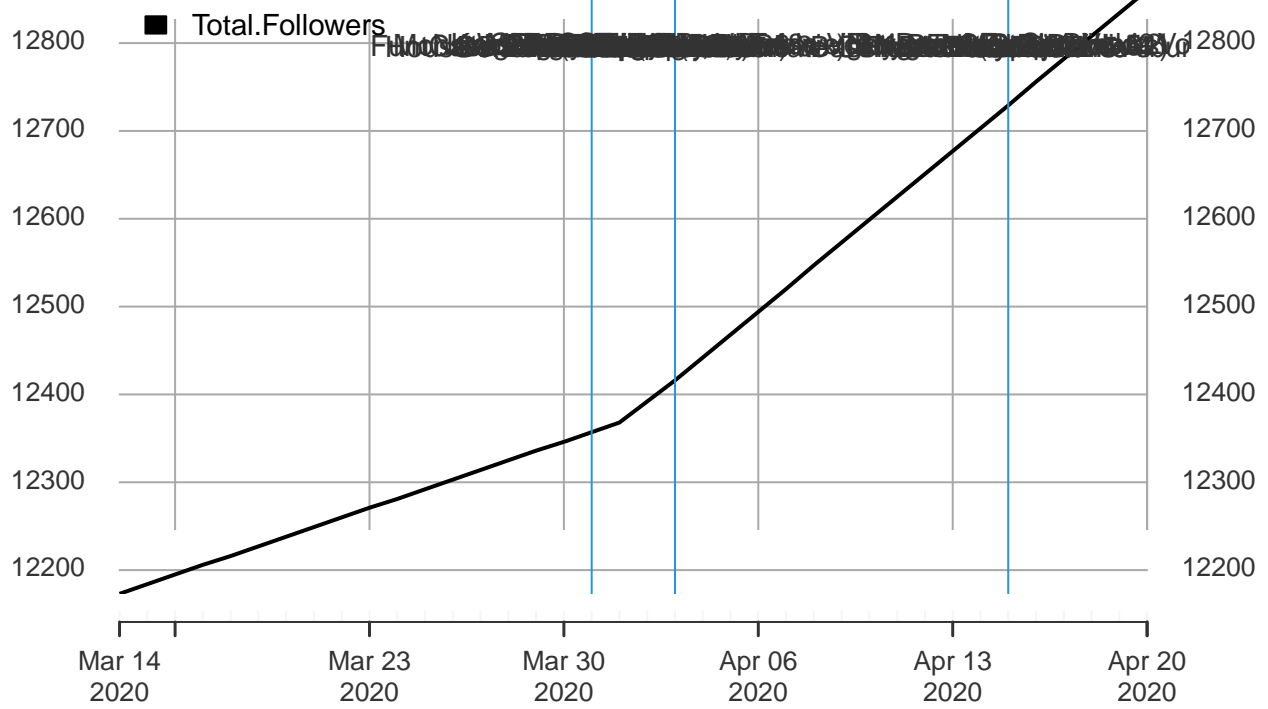
2018-04-04 / 2020-05-04





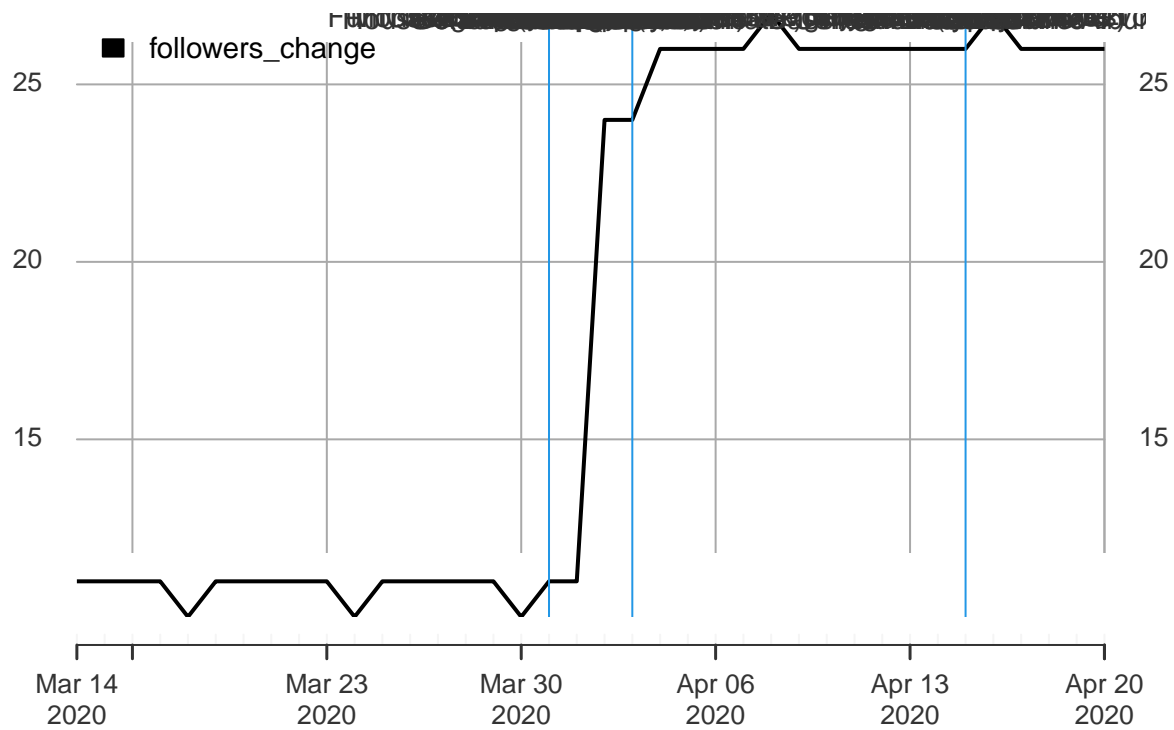
yakz

2020-03-14 / 2020-04-20



yakz

2020-03-14 / 2020-04-20



yakz

2020-03-14 / 2020-04-20

