Annotating Plots

INFO 526 Data Analysis and Visualization

Adriana Picoral

We will focus on how to use text on this tutorial to annotate a timeseries. We will be getting data from two sources: annotation data will be scraped from a news website using the `revest` package, and stock closing prices for GameStop will be downloaded from Yahoo! Finance.

Our visualization will be of GameStop closing prices for their stock between December 2020 and March 2021.

Scraping text data from a website

We will start with scraping data from a website which contains information on the timeline for the changes in GameStop stock prices.

The first step is to read the html page in.

```
# load libraries
library(tidyverse)
library(rvest)

# scrape text from news website
# set URL
article_url <- "https://abcnews.go.com/Business/gamestop-timeline-closer-saga-upended-wall-street/story

# read the URL into an object that holds the html code for the page
article_html <- read_html(article_url)
```

Once we have the HTML page, we can parse for HTML nodes. For this specific page, the information we want is in `<h3>` tags.

```
article_html %>%
  html_nodes("h3")
```

```
## {xml_nodeset (13)}
##  [1] <h3>Dec. 8, 2020: GameStop reports dismal earnings, stock takes a tumble ...
##  [2] <h3>Jan. 11, 2021: GameStop appoints 3 new directors to its board, inclu ...
##  [3] <h3>Jan 13, 2021: Stock surges more than 50%</h3>\n
##  [4] <h3>Jan. 19, 2021: Citron Research calls GameStop buyers 'suckers'</h3>\n
##  [5] <h3>Jan. 22, 2021: GameStop surges 50%</h3>\n
##  [6] <h3>Jan. 26, 2021: 'GameStonk' gets celeb backing from Elon Musk</h3>\n
##  [7] <h3>Jan. 27, 2021: Major short sellers close -- at a significant loss</h ...
##  [8] <h3>Jan. 28, 2021: Robinhood and other platforms restrict transactions f ...
##  [9] <h3>Jan. 29, 2021: SEC weighs in, trading platforms re-allow most GME tr ...
## [10] <h3>Jan. 31, 2021: 'Deadline' reports a Hollywood movie deal</h3>\n
```

```
## [11] <h3>Feb. 2, 2021: GME falls, all eyes on what comes next</h3>\n
## [12] <h3>Feb. 4, 2021: Janet Yellen meets with regulators</h3>\n
## [13] <h3 class="LivePromotion--title">ABC News Live</h3>
```

We can add to this by parsing the text in these HTML tags.

```
article_html %>%
  html_nodes("h3") %>%
  html_text()
```

```
##  [1] "Dec. 8, 2020: GameStop reports dismal earnings, stock takes a tumble"
##  [2] "Jan. 11, 2021: GameStop appoints 3 new directors to its board, including co-founder of e-comme
##  [3] "Jan 13, 2021: Stock surges more than 50%"
##  [4] "Jan. 19, 2021: Citron Research calls GameStop buyers 'suckers'"
##  [5] "Jan. 22, 2021: GameStop surges 50%"
##  [6] "Jan. 26, 2021: 'GameStonk' gets celeb backing from Elon Musk"
##  [7] "Jan. 27, 2021: Major short sellers close -- at a significant loss"
##  [8] "Jan. 28, 2021: Robinhood and other platforms restrict transactions for GME, lawmakers react"
##  [9] "Jan. 29, 2021: SEC weighs in, trading platforms re-allow most GME transactions"
## [10] "Jan. 31, 2021: 'Deadline' reports a Hollywood movie deal"
## [11] "Feb. 2, 2021: GME falls, all eyes on what comes next"
## [12] "Feb. 4, 2021: Janet Yellen meets with regulators"
## [13] "ABC News Live"
```

If we transform the output into a dataframe, we can then transform it.

```
# save character vector to an object
raw_text <- article_html %>%
  html_nodes("h3") %>%
  html_text()

# create a dataframe with raw text
gme_annotation <- data.frame(original_text = raw_text)
```

We can split the date from the event text.

```
# split date and annotation
gme_annotation <- gme_annotation %>%
  separate(original_text, sep = ":",
           into = c("date", "event"))

gme_annotation
```

```
##             date
## 1   Dec. 8, 2020
## 2  Jan. 11, 2021
## 3   Jan 13, 2021
## 4  Jan. 19, 2021
## 5  Jan. 22, 2021
## 6  Jan. 26, 2021
## 7  Jan. 27, 2021
## 8  Jan. 28, 2021
```

```
## 9  Jan. 29, 2021
## 10 Jan. 31, 2021
## 11  Feb. 2, 2021
## 12  Feb. 4, 2021
## 13 ABC News Live
##                                                                                    event
## 1                                              GameStop reports dismal earnings, stock takes a tumble
## 2   GameStop appoints 3 new directors to its board, including co-founder of e-commerce giant Chewy
## 3                                                                   Stock surges more than 50%
## 4                                                     Citron Research calls GameStop buyers 'suckers'
## 5                                                                          GameStop surges 50%
## 6                                                      'GameStonk' gets celeb backing from Elon Musk
## 7                                                  Major short sellers close -- at a significant loss
## 8                          Robinhood and other platforms restrict transactions for GME, lawmakers react
## 9                                     SEC weighs in, trading platforms re-allow most GME transactions
## 10                                                           'Deadline' reports a Hollywood movie deal
## 11                                                              GME falls, all eyes on what comes next
## 12                                                                  Janet Yellen meets with regulators
## 13                                                                                               <NA>
```

We can filter out the line that has `NA` for event.

```
gme_annotation <- gme_annotation %>%
  filter(!is.na(event))
```

Then we focus on the date. We first inspect it and try to convert it to a date data format.

```
gme_annotation %>%
  pull(date)
```

```
##  [1] "Dec. 8, 2020"  "Jan. 11, 2021" "Jan 13, 2021"  "Jan. 19, 2021"
##  [5] "Jan. 22, 2021" "Jan. 26, 2021" "Jan. 27, 2021" "Jan. 28, 2021"
##  [9] "Jan. 29, 2021" "Jan. 31, 2021" "Feb. 2, 2021"  "Feb. 4, 2021"
```

```
gme_annotation %>%
  mutate(date = gsub("\\.", "", date),
         updated_date = parse_datetime(date,
                              format = "%b %d, %Y")) %>%
  select(date, updated_date)
```

```
##            date updated_date
## 1   Dec 8, 2020   2020-12-08
## 2  Jan 11, 2021   2021-01-11
## 3  Jan 13, 2021   2021-01-13
## 4  Jan 19, 2021   2021-01-19
## 5  Jan 22, 2021   2021-01-22
## 6  Jan 26, 2021   2021-01-26
## 7  Jan 27, 2021   2021-01-27
## 8  Jan 28, 2021   2021-01-28
## 9  Jan 29, 2021   2021-01-29
## 10 Jan 31, 2021   2021-01-31
## 11  Feb 2, 2021   2021-02-02
## 12  Feb 4, 2021   2021-02-04
```

Looks good, so we can make the change to the actual dataframe.

```
gme_annotation <- gme_annotation %>%
  mutate(date = gsub("\\.", "", date),
         date = parse_datetime(date,
                               format = "%b %d, %Y"))
```

We have everything we need for our annotation.

Getting data from stock prices

For the stock prices, we will get data from Yahoo. We will use `janitor` to `clean_names()` and lowercase the column names from the original data.

```
library(janitor)
gme_data <- read_csv("data/GME.csv") %>%
  clean_names()
```

We can now merge the stock prices dataframe with annotation dataframe, since both have a column called `date` that is a date format.

```
# merge dataframes, using the largest dataframe as the reference (i.e. left)
gme_data <- gme_data %>%
  left_join(gme_annotation)
```

Visualization

We will plot our stock price data on a line using `geom_line()`, with points only for the data we have annotations for. We will annotate the timeseries using `geom_label()` but from the `ggrepel` package. We will also use `annotate()`. Because of the number of annotations labels (11 total), we need to create a custom color scheme. The x scale is adjusted with `filter` and the y axis is set for a dollar format.

```
library(ggthemes)
library(ggrepel)
library(RColorBrewer)
library(scales)

my_color_scheme <- c(brewer.pal(n = 8, name = "Dark2"),
                     rev(brewer.pal(n = 3, name = "Dark2")))

gme_data %>%
  filter(date > as.Date("2020-12-01") &
           date < as.Date("2021-03-01")) %>%
  ggplot(aes(x = date,
             y = close,
             label = event,
             color = event)) +
  geom_line(color = "black") +
  geom_point(data = gme_data %>% filter(!is.na(event))) +
```

```
geom_label_repel(size = 1.8,
                 nudge_y = 200) +
scale_y_continuous(limits = c(0, 600),
                   labels = scales::dollar_format(),
                   breaks = c(0, 100, 200, 300, 350)) +
#scale_color_brewer(type = "qual", palette = "Paired") +
scale_color_manual(values = my_color_scheme) +
geom_hline(yintercept = 347.51,
           color = "dimgrey",
           linetype = "dashed") +
annotate("text",
         x = as.POSIXct("2020-12-20"),
         y = 375,
         color = "dimgrey",
         label = "Stock Closing High: $347.51") +
theme_linedraw() +
theme(panel.grid = element_blank(),
      legend.position = "none") +
labs(title = "GameStop Stock Closing Price Fluctuation",
     subtitle = "between December 2020 and March 2021",
     caption = "data from finance.yahoo.com and abcnews.go.com")
```