HOMEWORK NO.2 EXPLORATORY DATA ANALYSIS

L.H. Thien

March 20, 2021

I. INTRODUCTION TO THE PROBLEM

In this report, I would like to introduce an improved version in term of readability for the line graph taken from CNBC broadcasted on Squawk Box| Dec. 10, 2020.

Source: https://www.youtube.com/watch?v=cxVFopLpIQY (uploaded on March 3,2021)

In general, this plot is quite good. But for an issue of readability, there are still remaining some drawbacks, such as:

1) X axis and y axis are in the lack of titles.
2) 3D graph frequently increases the time readers spend on evaluating data (for example, we cannot instantly tell the value for Dec !).
3) The way of scaling y axis is counter-intuitive. Firstly, it did not start at 0. So the ratio sometimes can be misunderstood (ex: value at DEC. is doubled value at OCT.). Secondly, all the time, we use the decimal system, we are not familiarized with the dividing y axis into multiple of 6 (in this graph 30-36-42-…!)

II. CODES IN R AND CREATED VISUALIZATION:

```r
# Loading necessary libraries:
library(readr)
library(dplyr)
library(ggplot2)
install.packages("hrbrthemes")
library(hrbrthemes)

# Loading the data set: since I downloaded the data set from: https://finance.yahoo.com/quote/UAL/history/ into my computer.
UAL <- read_csv("C:/Users/DELL/OneDrive/Desktop/Ly_HoangThien/UAL.csv")
View(UAL)

# Cleaning and trimming data set: We only keep 2 columns, namely: Date and Close.
UAL <- UAL %>% select(Date,Close)

#Due to the fact that our line graph contains information within around 3 months
#(in 2020 from mid-Sep to mid-Dec), so we will again cook up the data to prepare
#for the visualization process.
UAL <- UAL %>% filter(Date > as.Date("2020-09-8"),Date < as.Date("2020-12-10"))

# Creating plot:
library(ggtext)



c1 <- c(30, 35, 40,45,50)
c2 <- c(0,20,30,35,40,45,50)

highlight_point <- UAL %>% filter(Date== "2020-12-09")

ggplot(UAL,aes(x=Date,y=Close)) +
  geom_line(linetype="solid",size=1.5,color="#6b029c") +
  labs(x="Date",y="UAL Stock (USD)",title="UNITED AIRLINE (UAL)",
       subtitle = "47.00  <span style='color:#eb0040'>  -1.02 ↓</span>  <span style='color:#eb0040'>[-2.12%↓]</span>") +
  theme(plot.subtitle = element_markdown(),
        panel.background = element_rect(fill = "#cad3eb", colour = "#07558c",
        size = 1, linetype = "solid"),
        panel.grid.major = element_line(size = 0.3, linetype = 'dotted',colour = "#8aa5b8"),
        panel.grid.minor = element_line(size = 0.25, linetype = 'dotted',colour = "#8aa5b8"),
        axis.text.x = element_text(size=13),axis.text.y = element_text(size=11),
        axis.title.x = element_text(color="#0a0a0a", vjust= 0.75),
        axis.title.y = element_text(color="#0a0a0a", vjust= 1.75)) +
        scale_color_brewer()+
        ft_geom_defaults() + scale_y_continuous(limits = c(1.5,55),breaks=c2 ,minor_breaks = c1)+
        geom_point(data=highlight_point,color="#eb0e0e",size=3.5)
```

III. CONCLUSION:

Some enhancements are introduced in this graph respect to these aspects:

i) Axes x and y have corresponding titles “UAL Stock (USD)” and “Date”, which gives more detail that we are considering in US Dollar.

ii) In the subtitle: We used the color and the down arrow to display the current situtation that UAL Stock is decreasing -1.02 US Dollar and -2.12% in the comparision to the closed price of the previous day.

iii) With the dotted grid and breaking x-axis into multiples of 5, readers would quickly evaluate UAL Stock (rather than multiples of 6).

iv) X axis starts from 0 giving a better comprehension about the ratio of changes.

v) 2D map in this case is direct and more explicit than 3D graph.