Student Name: Henry J. Hu
Course: DSC-611-Z1 Data Visualization
Visualization Assignment

Introduction

In this assignment, we learned how to apply the concepts of Graphical Excellence, Graphical Integrity, Graphical Sophistication, how to choose suitable visualization types depending on the task and data at hand, how to assign data to appropriate chart elements to depict relationships between data items, and how to employ the tool of your choice to create basic visualizations.

Data Source

Description

A year worth of trading data belonging to the Dow Jones Index. It was originally collected for the "Brown, Pelosi & Dirska, 2013" project at the University of Maryland.

Data Temporal

The entire year of 2011.

Data Size

750 observations and 16 fields.

Data Geographic Information

Worldwide.

Source

Dr. Michael Brown, michael.brown@umuc.edu, University of Maryland University College.

Data Download

http://archive.ics.uci.edu/ml/datasets/Dow+Jones+Index

Table 1: Field Detail

Field Name	Data Type	Description	Unit
quarter	Nominal	The yearly quarter (1 = Jan-Mar; 2 = Apr=Jun)	Text
stock	Nominal	The stock symbol	Text
date	Interval	The last business day of the work (this is typically a Friday)	Date
percent_change_price	Ratio	The percentage change in price throughout the week	Percent
percent_chagne_volume_over_last_wek	Ratio	The percentage change in the number of shares of stock that traded hands for this week compared to the previous week	Percent
percent_change_next_weeks_price	Ratio	The percentage change in the price of the stock in the following week	Percent
percent_return_next_dividend	Ratio	The percentage of return on the next dividend	Percent
open	Ratio	The price of the stock at the beginning of the week	Amount in US Dollar
high	Ratio	The highest price of the stock during the week	Amount in US Dollar
low	Ratio	The lowest price of the stock during the week	Amount in US Dollar
close	Ratio	The price of the stock at the end of the week	Amount in US Dollar
next_weeks_open	Ratio	The opening price of the stock in the following week	Amount in US Dollar
next_weeks_close	Ratio	The closing price of the stock in the following week	Amount in US Dollar
volume	Ratio	The number of shares of stock that traded hands in the week	Count
previous_weeks_volume	Ratio	The number of shares of stock that traded hands in the previous week	Count
days_to_next_dividend	Ratio	The number of days until the next dividend	Count

Data Wrangling Verdict

Out of thirteen numeric fields, two have missing values. A total of 60 missing values were replaced with column averages. The remaining two text fields and one date field have no missing value.

Data Wrangling R Markdown Page

https://aaacomply.com/data_science/DSC611/Henry_Hu_Moduel_2_Programming.html

Data Visualization

IBM - Outperform The Rest

As depicted in Figure 1 below, the multi-line chart reveals the pattern of each stock's closing price for the year 2011. According to the Grubb's Outlier test results in the R Markdown output, the only variable which does not have outlier is days_to_next_dividend. Therefore, all the other variables, including closing price have noise. However, the majority of the data points do move in the same direction at each day interval. The increase and decrease in values of each group of data points at each day interval are what form the pattern. According to chapter 2 of Introduction to Data Visualization & Storytelling, our human's understanding of gravity is what helps us perceive the data points higher on the chart represent larger stock closing values and data points lower on the chart represent smaller stock closing values. The different line colors and the stock symbol at the end of each line allow the audience to quickly identify the distinct pattern of each stock. The chart gridlines, labels, and units on the y-axis and x-axis help the audience approximate the amounts of increase or decrease in the patterns. Also, as explained in Jeff Johnson's book, Designing with the Mind, the Gestalt principle of Proximity indicated

that our human brains have certain ways of perceiving different arrangements of data. These perceptions were done involuntarily by our brains without any cognitive effort. Based on the proximity of space between the data points and their arrangement, our brains automatically perceive the pattern of the data points. The plot title helps convey to the audience the narrative of the plot. Each line on the plot is a story. All the other elements of the plot are just information.

BAC - Traded The Most

As depicted in Figure 2 below, the pie chart reveals the accumulation of stock volume during the year of 2011. Outliers would not skew the data points much in this case since each data point is the result of summing up all the volume values belonging to each stock. Therefore, noise in this data does not affect the integrity of the chart much. As one could see that BAC has the largest total accumulated volume of 18B. The slices are colored and sized differently so the audience could quickly identify the proportion of each stock's total volume in the pie chart. Also, the pair of stock symbol and volume value right next to each slice allows the audience to quickly identify each stock's actual total accumulated volume. The reason the pie chart is used for depicting the accumulated volumes is because the iso-measure for this presentation is a pie of stock volumes. According to chapter 2 of Introduction to Data Visualization & Storytelling, iso-measure is the person's ability to relate to what is being presented. In this case, since we all have eaten a pie of something at a certain time in our life, we can automatically relate to the presentation and understand that the slices are the proportions of stock volume. By using a pie chart, it helps the audience quickly retrieve knowledge from the presentation. The chart title

helps convey to the audience the narrative of the chart. Each slice on the pie chart is a story. All the other elements on the pie chart are just information.

AT&T Inc. (T) - Highest Average Dividend %

As depicted in Figure 3 below, the horizontal bar plot reveals the average dividend percentage for each stock during the year 2011. The reason the horizontal bar plot is used instead of the vertical bar plot is that we are not measuring growth, but rather comparing percentage values. According to chapter 2 of Introduction to Data Visualization & Storytelling, our human's understanding of gravity will cause our brains to perceive the vertically increasing length of the vertical bar as growth. According to the Grubb's Outlier test results in the R Markdown output, the only variable which does not have outlier is days_to_next_dividend. Therefore, all the other variables, including percent_return_next_dividend have noise. In order words, either the minimum dividend percentage value or the maximum dividend percentage value of a particular stock could just be noise. That is the reason why the average dividend percentage should be used instead of the maximum dividend percentage. The stock symbols on the y-axis have been re-arranged so that the y-axis has been sorted by the horizontal bar length of each stock symbol. This allows the audience to quickly identify which stock has the largest average dividend percentage for the year. The different bar colors help the audience quickly identify which bar belongs to which stock symbol. The chart gridlines, labels, and units on the y-axis and x-axis help the audience identify the average dividend percentage for each stock. The plot title helps convey to the audience the narrative of the plot. Each horizontal bar on the plot is a story.

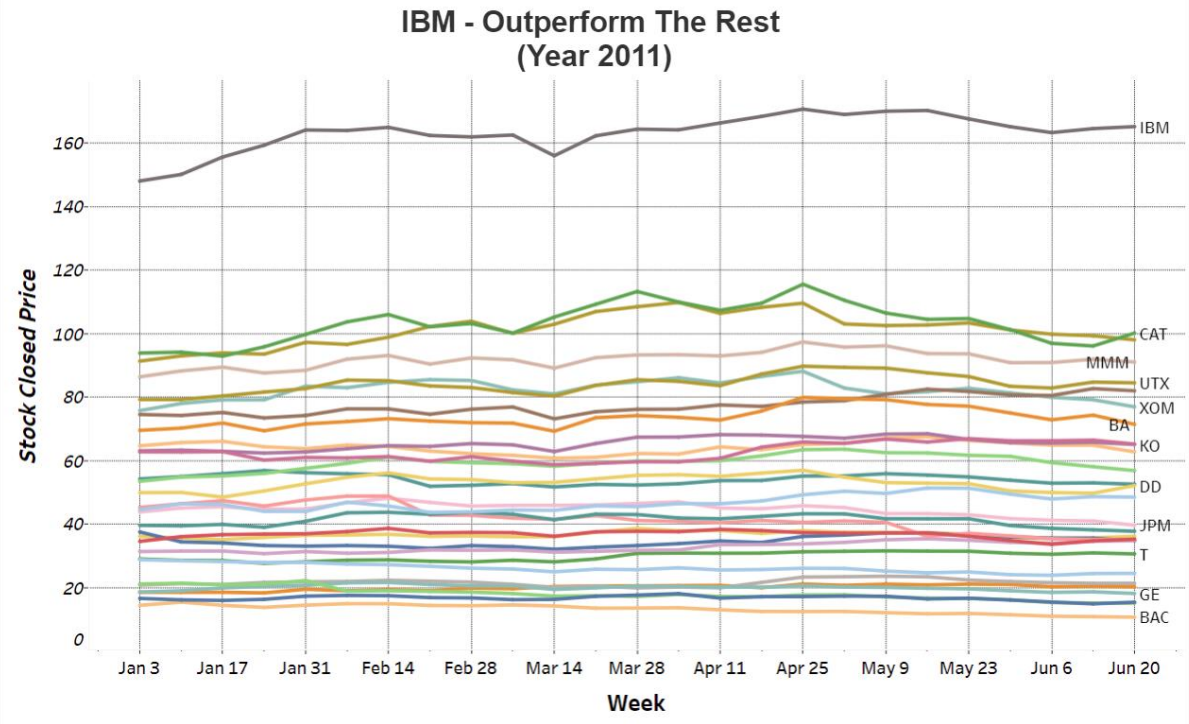

Figure 1: IBM - Outperform The Rest

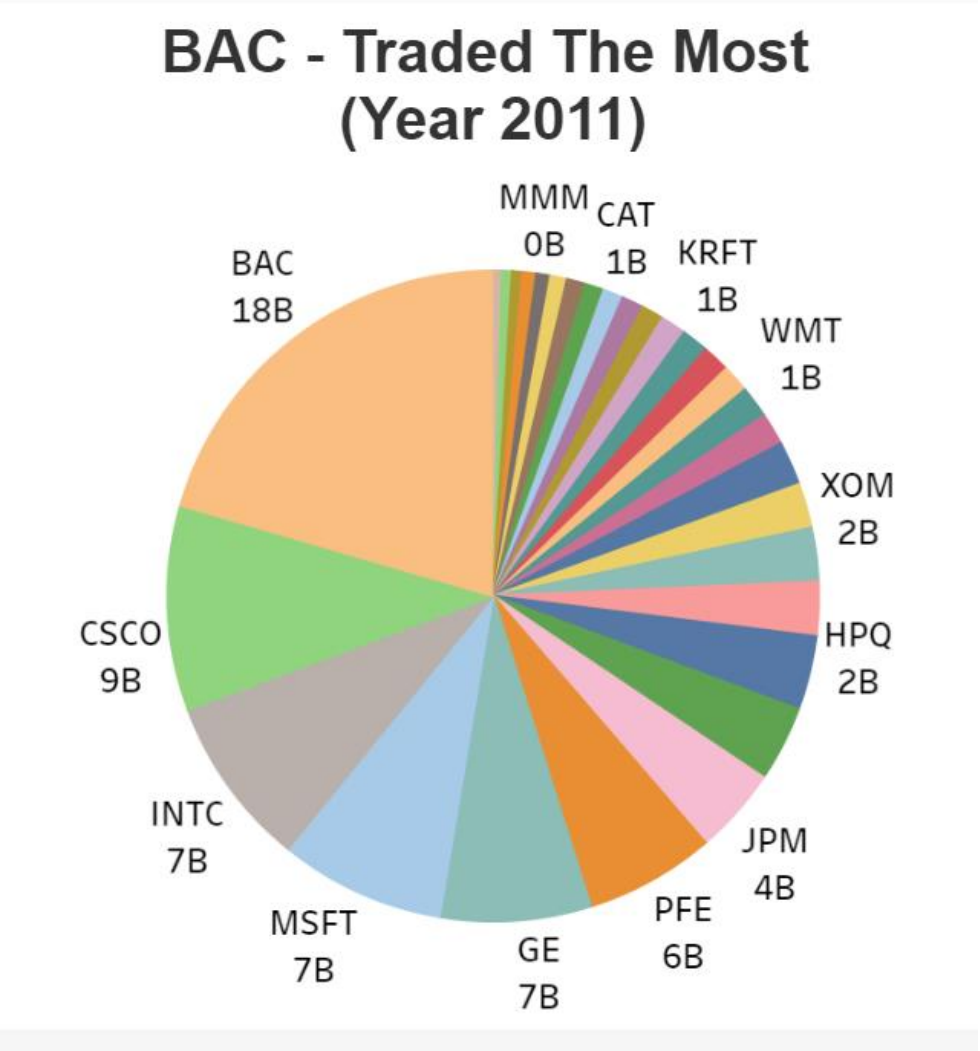

Figure 2: BAC - Traded The Most

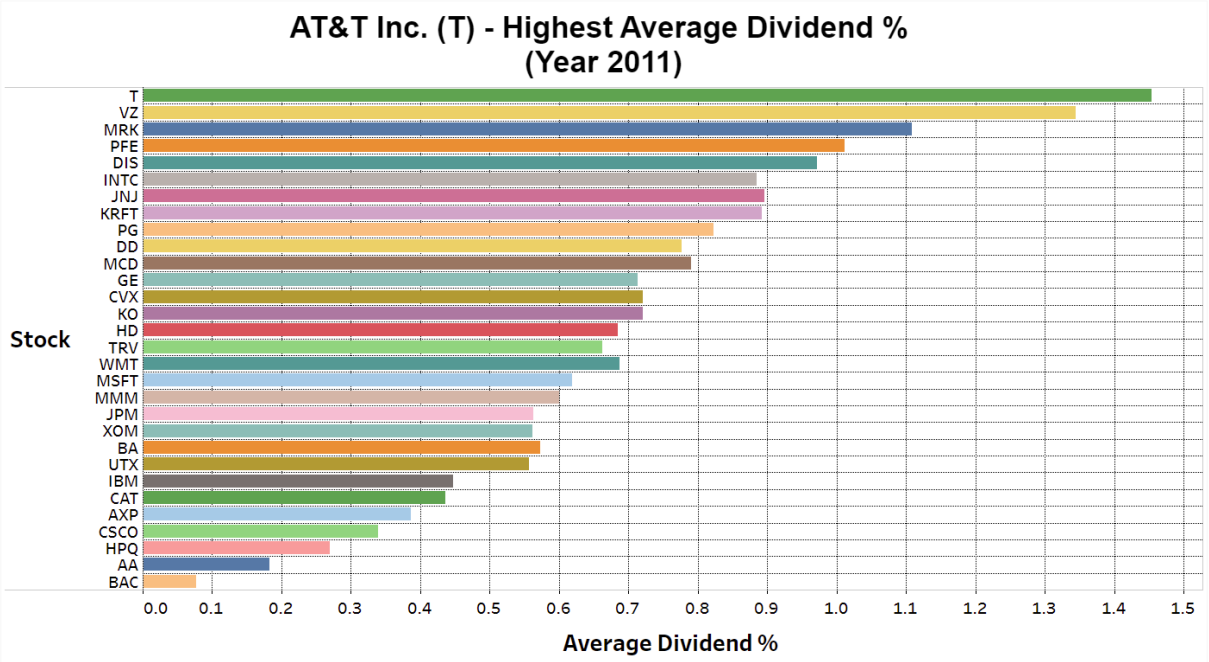

Figure 3: AT&T Inc. (T) - Highest Average Dividend %

Plot & Chart Hyperlinks

IBM - Outperform The Rest
https://public.tableau.com/views/plots_2_15998597381950/IBM-OutperformTheRest?:language=en&:display_count=y&:origin=viz_share_link

BAC - Traded The Most
https://public.tableau.com/views/plots_2_15998597381950/BAC-TradedTheMost?:language=en&:display_count=y&:origin=viz_share_link

AT&T Inc. (T) - Highest Dividend Payment %
https://public.tableau.com/views/plots_2_15998597381950/ATTInc_T-HighestAverageDividend?:language=en&:retry=yes&:display_count=y&:origin=viz_share_link

Work Cited/References

Berengueres, J., Fenwick, A., Sandell, M. (2019). Introduction to Data Visualization & Storytelling: A Guide For The Data Scientist (First Edition).
Independently published

Johnson, J., (2014). Designing with the Mind in Mind (Second Edition).
ScienceDirect. https://doi.org/10.1016/C2012-0-07128-1

Scott, B. M. (July 2013). Dynamic-Radius Species-Conserving Genetic Algorithm for the Financial Forecasting of Dow Jones Index Stocks. Retrieved from
https://www.researchgate.net/publication/236085839_Dynamic-Radius_Species-Conserving_Genetic_Algorithm_for_the_Financial_Forecasting_of_Dow_Jones_Index_Stocks (Accessed September 11th, 2020)