Assignment 2: Grammar of Graphics

Szymon Bujowski - 148050, Preetam Sharma - 150504

Contents

Foreword

The following document is an assignment for Poznan University of Technology's Data Visualization course. The course is conducted by Dariusz Brzeziński during the 4th semester of Artificial Intelligence Bachelor degree.

The assignment is an implementation of the grammar of graphics, intended to create rich visualizations from the data we were provided with. The data consists of two data sets, for both of which we've chosen the upcoming visualizations. As it is stated in the assignment description:

- The data in the Sectors folder present the percentage changes of stock prices and trading volume in selected sectors
  - Some of the data sets also contain information about the media sentiment about the companies
- The Correlations.csv data set contains correlations between the stock prices of pairs of companies identified by stock symbols (tickers)

For both the following visualizations, we will provide brief descriptions and reasoning behind them.

Credits

We have to credit the lecturer, **Dariusz Brzeziński**, for the interactive tables we have used. They were built using the `DT` package.

The interactive tables should work as intended in html format of the document, however they will not be visible in pdf format. For that reason, a standard head of the data frames are displayed.

You can access the project's repository on GitHub - https://github.com/bujowskis/put-DV/tree/main/ass-2

Stocks by sectors

The data

For simplicity reasons, we are going to show two out of 8 data sets. One of them will be a representative of the sets with sentiments included, and the other one with sentiments missing.

Sentiment included

```
##    X Symbol                 Name   Volume X1dC.  X1dV.   Open   High  Close
## 1  7   ANTM               Anthem   927200  1.44  -3.25 461.80 465.03 464.86
## 2 16   CTLT             Catalent   800600  3.90 -41.72 128.03 128.26 124.49
## 3 53    SYK  Stryker Corporation  1169700 -3.48 -69.48 267.41 268.87 268.42
## 4 52    STE               Steris   351600 -1.07 -36.01 243.11 243.76 242.57
## 5 59   VTRS              Viatris 11959600 -1.13  -4.98  13.60  14.30  14.21
## 6 41    MCK McKesson Corporation   643400  0.04  -3.25 247.54 248.44 248.10
##   Volume.1 Sentiment
## 1   927200      0.68
## 2   800600      0.66
## 3  1169700      0.63
## 4   351600      0.62
## 5 11959600      0.60
## 6   643400      0.56
```

Sentiment missing

```
##    X Symbol                                Name    Volume X1dC.  X1dV.  Open
## 1  5   BBBY              Bed Bath & Beyond Inc. 105519200 -5.30  82.21 30.00
## 2 29      F                  Ford Motor Company  87711400 -0.38 -13.84 16.84
## 3 14    CCL          Carnival Corporation & plc  67608300 -2.25  -0.28 17.30
## 4 60   NCLH Norwegian Cruise Line Holdings Ltd.  39182900 -3.77   3.61 17.17
## 5 49   LCID                   Lucid Group, Inc.  35016600 -4.62   4.76 22.94
## 6 21   DKNG                     DraftKings Inc.  29355000  3.71  -6.35 20.58
##    High Close Sentiment
## 1 30.06 21.71        NA
## 2 16.90 15.97        NA
## 3 17.48 15.53        NA
## 4 17.38 15.38        NA
## 5 24.41 23.17        NA
## 6 20.89 18.05        NA
```

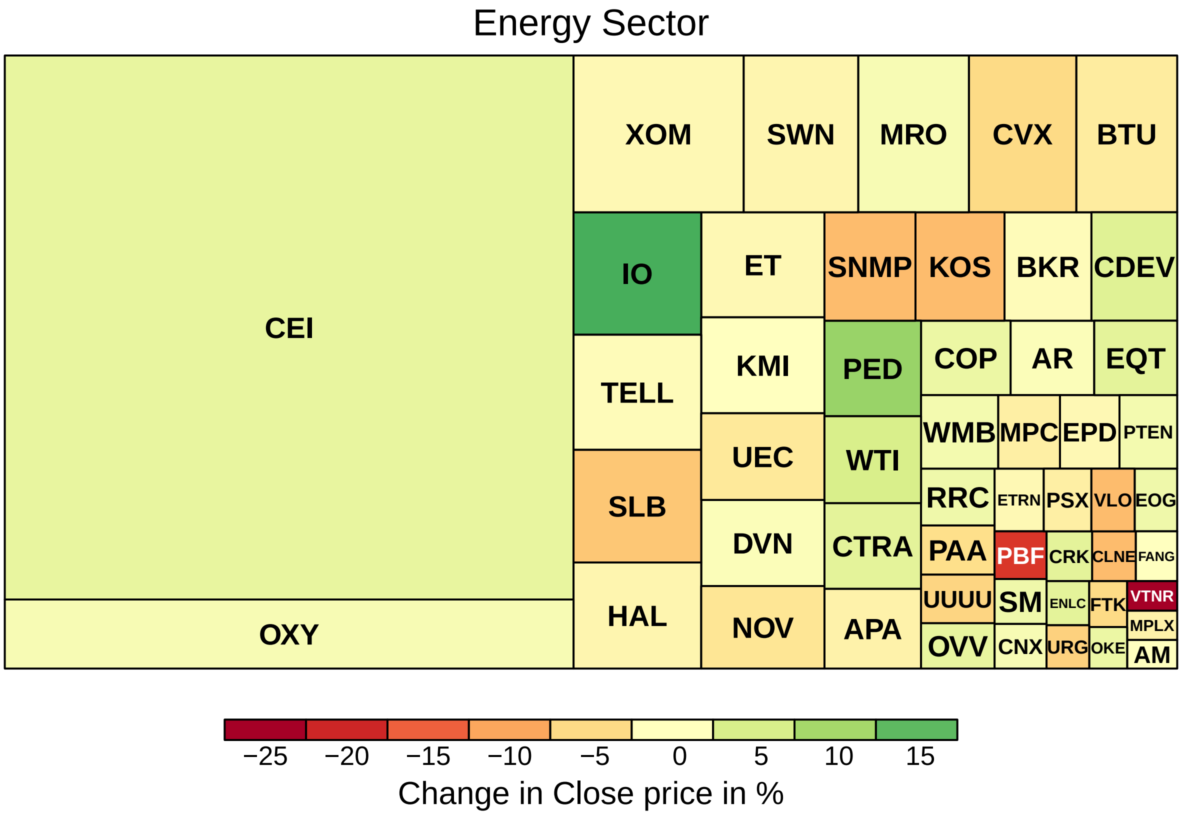
Sketch

TODO

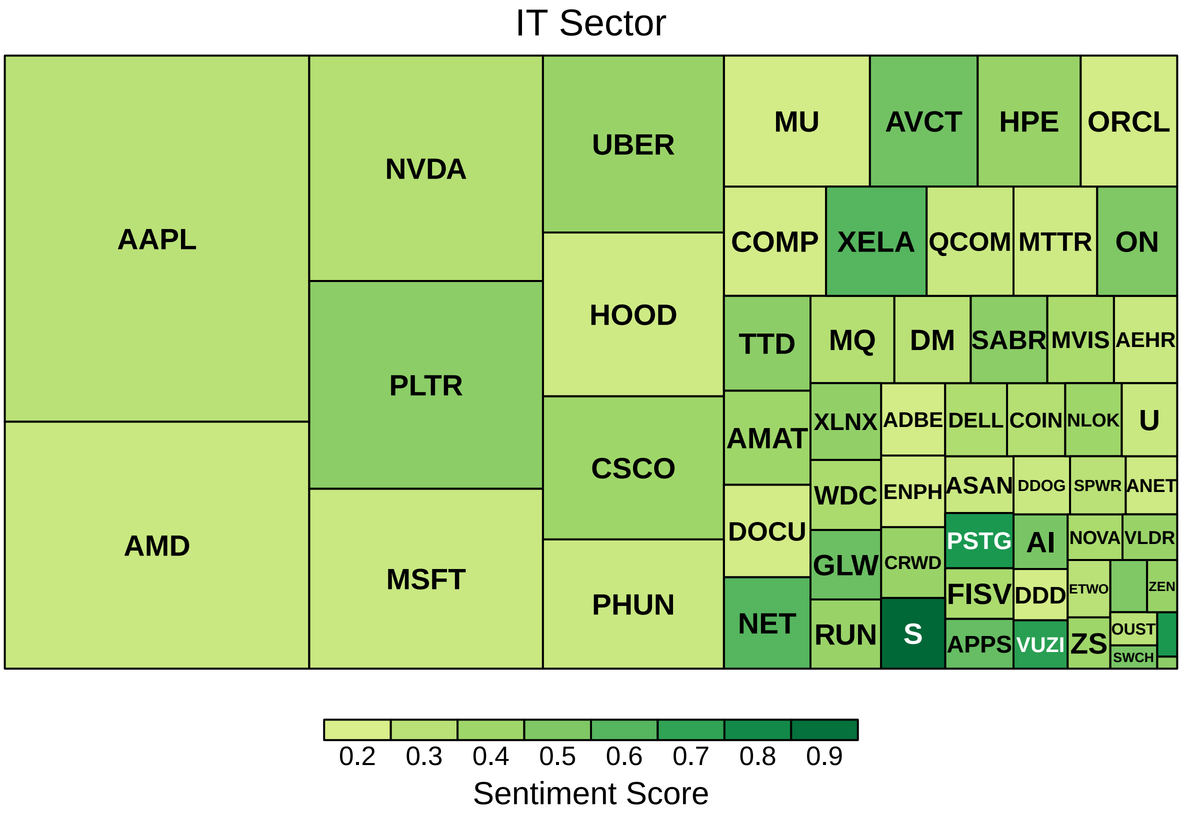
The visualization

```r
# So the idea here is to show the change in Closing price of all the stocks of particular sector in one
# for that we decided to use Treemap and color of treemap shows the chnage in close price

library(treemap)
# Here we are going to show one from including Sentiment and without sentiment
energy = read.csv("Dataset/Sectors/energy.csv")
IT = read.csv("Dataset/Sectors/it.csv")
```

```
treemap(energy,index=c("Symbol"),vSize = "Volume", vColor = "X1dC.",type="value",border.col = "black",
        border.lwds = 1,title = "Energy Sector",title.legend = "Change in Close price in %")
```



```
treemap(IT,index=c("Symbol"),vSize = "Volume", vColor = "Sentiment",type="value",border.col = "black",
        border.lwds = 1,title = "IT Sector",title.legend = "Sentiment Score")
```



Correlations

The data

```
##   Ticker.1 Ticker.2 Correlation.Value
## 1       GS      JPM         0.7955952
## 2     AAPL     MSFT         0.7069591
## 3      AXP      JPM         0.6833357
## 4       KO       PG         0.6553540
## 5      CRM     MSFT         0.6464821
## 6      HON      MMM         0.6289362
```

Sketch

Static visualization choice for the correlations was pretty obvious from the beginning - a heat map correlation matrix. For that reason, there was really no sketch here.

Regarding handling situations in which there is some correlation value missing, it sufficed to use `NA` value, which would result in a missing tile in the visualization.

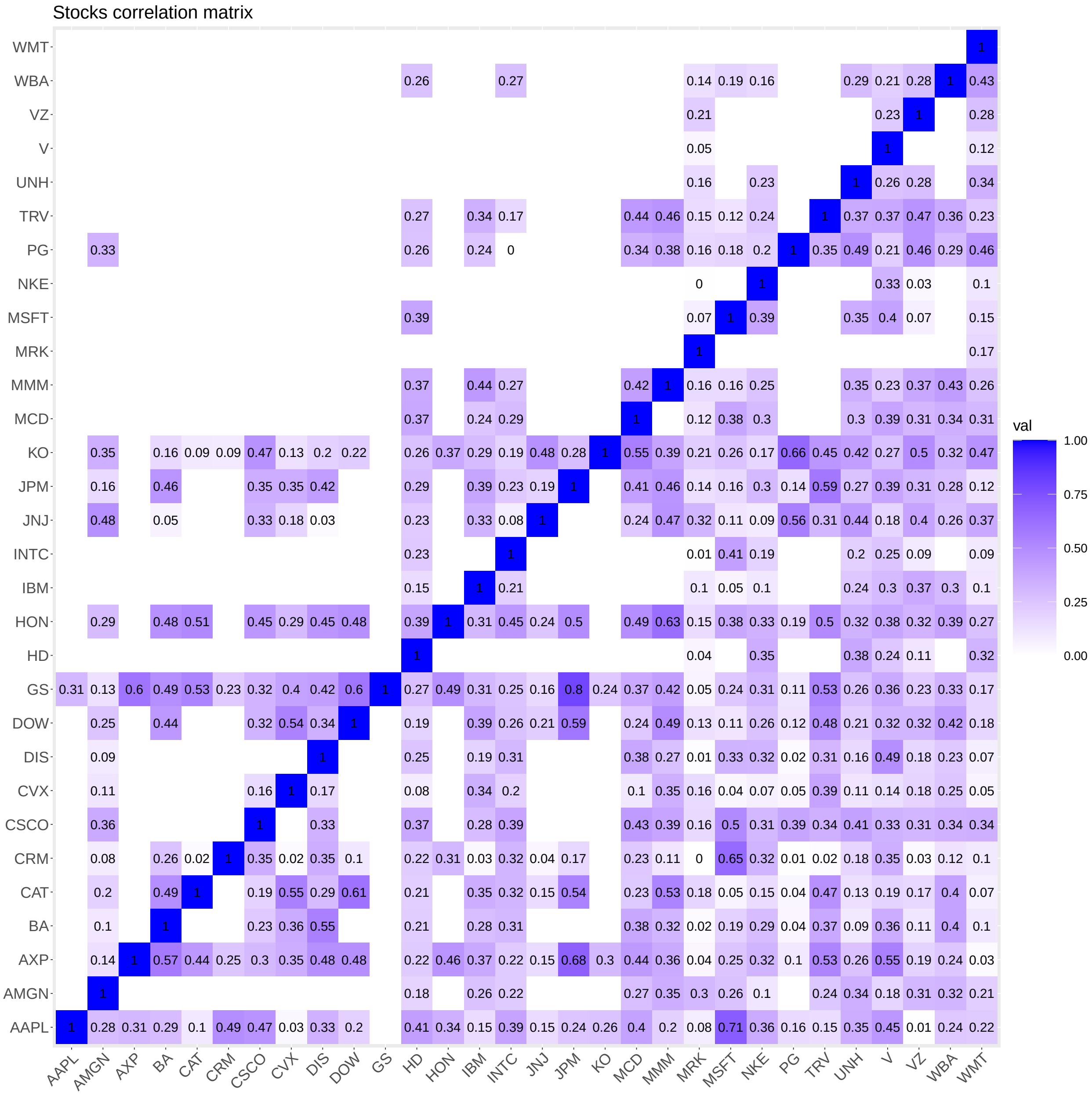
However, there was no such situations in this case, and thus this feature cannot be seen.

The visualization

```
library(ggplot2)
library(plotly)
```

```r
# get all unique tickers
ut <- data.frame(tickers=union(cor_data$Ticker.1, cor_data$Ticker.2))
rut <- data.frame(tickers=rev(ut$tickers))  # save a reversed copy for later

# create dataframe of all combinations
df <- expand.grid(ticker1=rut$tickers, ticker2=ut$tickers)

# read the correlation values
df$val <- NA # correlation not specified, cell will be colored black
for (i in 1:nrow(cor_data)) {
  # read from the dataset
  df$val[length(ut$tickers)*(match(cor_data$Ticker.1[i], ut$tickers) - 1) +
             match(cor_data$Ticker.2[i], rut$tickers)] = cor_data$Correlation.Value[i]
  # it's bidirectional
  df$val[length(ut$tickers)*(match(cor_data$Ticker.2[i], ut$tickers) - 1) +
             match(cor_data$Ticker.1[i], rut$tickers)] = cor_data$Correlation.Value[i]
}
j = length(ut$tickers)
for (i in 0:(length(ut$tickers) - 1)) {
  # remove upper triangle
  for (k in 0:i) {
    df$val[j - k] = NA
  }
  j = j + length(ut$tickers)
}
for (i in 0:(length(ut$tickers) - 1)) {
  # correlation = 1 between the same stock
  df$val[length(ut$tickers) + i*(length(ut$tickers) - 1)] = 1
}

# text for tooltip
df <- df %>%
  mutate(text = paste0(df$ticker1, "\n", df$ticker2, "\n", "Val: ", df$val))

# Heatmap
p = ggplot(df, aes(ticker1, ticker2, fill=val)) +
  geom_tile() +
  geom_text(aes(label=round(val, 2)),
            size=6
            ) +
  #scale_x_discrete(guide=guide_axis(n.dodge=2)) +
  theme(axis.title.x=element_blank(),  # remove x axis title
        axis.title.y=element_blank(),  # remove y axis title,
        text=element_text(size=20),
        axis.text=element_text(size=20),
        legend.key.size = unit(2, 'cm'),
        legend.key.height = unit(2, 'cm'),
        legend.key.width = unit(2, 'cm'),
        axis.text.x=element_text(angle=45, hjust=1)
        ) +
  scale_fill_gradient2(low="white", high="blue",
                       limits=c(c(0, 1)),
                       na.value="white"
```

```
                        ) +
  ggtitle("Stocks correlation matrix")
p
```



	AAPL	AMGN	AXP	BA	CAT	CRM	CSCO	CVX	DIS	DOW	GS	HD	HON	IBM	INTC	JNJ	JPM	KO	MCD	MMM	MRK	MSFT	NKE	PG	TRV	UNH	V	VZ	WBA	WMT
WMT																														1
WBA												0.26			0.27						0.14	0.19	0.16			0.29	0.21	0.28	1	0.43
VZ																					0.21						0.23	1		0.28
V																					0.05						1			0.12
UNH																					0.16		0.23			1	0.26	0.28		0.34
TRV												0.27		0.34	0.17				0.44	0.46	0.15	0.12	0.24		1	0.37	0.37	0.47	0.36	0.23
PG		0.33										0.26		0.24	0				0.34	0.38	0.16	0.18	0.2	1	0.35	0.49	0.21	0.46	0.29	0.46
NKE																					0		1				0.33	0.03		0.1
MSFT												0.39									0.07	1	0.39			0.35	0.4	0.07		0.15
MRK																					1									0.17
MMM												0.37		0.44	0.27				0.42	1	0.16	0.16	0.25			0.35	0.23	0.37	0.43	0.26
MCD												0.37		0.24	0.29				1		0.12	0.38	0.3			0.3	0.39	0.31	0.34	0.31
KO		0.35		0.16	0.09	0.09	0.47	0.13	0.2	0.22		0.26	0.37	0.29	0.19	0.48	0.28	1	0.55	0.39	0.21	0.26	0.17	0.66	0.45	0.42	0.27	0.5	0.32	0.47
JPM		0.16		0.46			0.35	0.35	0.42			0.29		0.39	0.23	0.19	1		0.41	0.46	0.14	0.16	0.3	0.14	0.59	0.27	0.39	0.31	0.28	0.12
JNJ		0.48		0.05			0.33	0.18	0.03			0.23		0.33	0.08	1			0.24	0.47	0.32	0.11	0.09	0.56	0.31	0.44	0.18	0.4	0.26	0.37
INTC												0.23			1						0.01	0.41	0.19			0.2	0.25	0.09		0.09
IBM												0.15		1	0.21						0.1	0.05	0.1			0.24	0.3	0.37	0.3	0.1
HON		0.29		0.48	0.51		0.45	0.29	0.45	0.48		0.39	1	0.31	0.45	0.24	0.5		0.49	0.63	0.15	0.38	0.33	0.19	0.5	0.32	0.38	0.32	0.39	0.27
HD												1									0.04		0.35			0.38	0.24	0.11		0.32
GS	0.31	0.13	0.6	0.49	0.53	0.23	0.32	0.4	0.42	0.6	1	0.27	0.49	0.31	0.25	0.16	0.8	0.24	0.37	0.42	0.05	0.24	0.31	0.11	0.53	0.26	0.36	0.23	0.33	0.17
DOW		0.25		0.44			0.32	0.54	0.34	1		0.19		0.39	0.26	0.21	0.59		0.24	0.49	0.13	0.11	0.26	0.12	0.48	0.21	0.32	0.32	0.42	0.18
DIS		0.09							1			0.25		0.19	0.31				0.38	0.27	0.01	0.33	0.32	0.02	0.31	0.16	0.49	0.18	0.23	0.07
CVX		0.11					0.16	1	0.17			0.08		0.34	0.2				0.1	0.35	0.16	0.04	0.07	0.05	0.39	0.11	0.14	0.18	0.25	0.05
CSCO		0.36					1		0.33			0.37		0.28	0.39				0.43	0.39	0.16	0.5	0.31	0.39	0.34	0.41	0.33	0.31	0.34	0.34
CRM		0.08		0.26	0.02	1	0.35	0.02	0.35	0.1		0.22	0.31	0.03	0.32	0.04	0.17		0.23	0.11	0	0.65	0.32	0.01	0.02	0.18	0.35	0.03	0.12	0.1
CAT		0.2		0.49	1		0.19	0.55	0.29	0.61		0.21		0.35	0.32	0.15	0.54		0.23	0.53	0.18	0.05	0.15	0.04	0.47	0.13	0.19	0.17	0.4	0.07
BA		0.1		1			0.23	0.36	0.55			0.21		0.28	0.31				0.38	0.32	0.02	0.19	0.29	0.04	0.37	0.09	0.36	0.11	0.4	0.1
AXP		0.14	1	0.57	0.44	0.25	0.3	0.35	0.48	0.48		0.22	0.46	0.37	0.22	0.15	0.68	0.3	0.44	0.36	0.04	0.25	0.32	0.1	0.53	0.26	0.55	0.19	0.24	0.03
AMGN		1										0.18		0.26	0.22				0.27	0.35	0.3	0.26	0.1		0.24	0.34	0.18	0.31	0.32	0.21
AAPL	1	0.28	0.31	0.29	0.1	0.49	0.47	0.03	0.33	0.2		0.41	0.34	0.15	0.39	0.15	0.24	0.26	0.4	0.2	0.08	0.71	0.36	0.16	0.15	0.35	0.45	0.01	0.24	0.22