Statistical Data Mining – I

New York Stock Exchange

ANCHAL SOJATIYA
KAVYA ANANTHA RAO
SAMRATSINH DHUMAL

AGENDA

- Motivation and Objective
- Goals
- Dataset Description
- Exploratory Data Analysis
- Holt-Winters method for Time Series
- Z-Score Calculations

MOTIVATION AND OBJECTIVE

- As the variation in the IT market is on a rise, we have focused on the data in the 'Information Technology' sector, to predict how a company might perform in the near future and determine the companies that might get bankrupt
- The analysis of the same is a sought field in finance, which could be used to protect a company by tracking its performance or to advise investors, share holders and stock market specialists to make decisions
- These might help to foresee any challenges pertaining to recession and help take measures to avoid it

GOALS

- Check for the completeness of the data and use appropriate data cleansing method
- Investigate data in the Information Technology sector using EDA and explore the trends of company's stocks over a period to analyze the performance
- Predict the future stock prices using Holt-Winters for time series data
- Determine the companies that might get bankrupt in the near future

DATASET DESCRIPTION

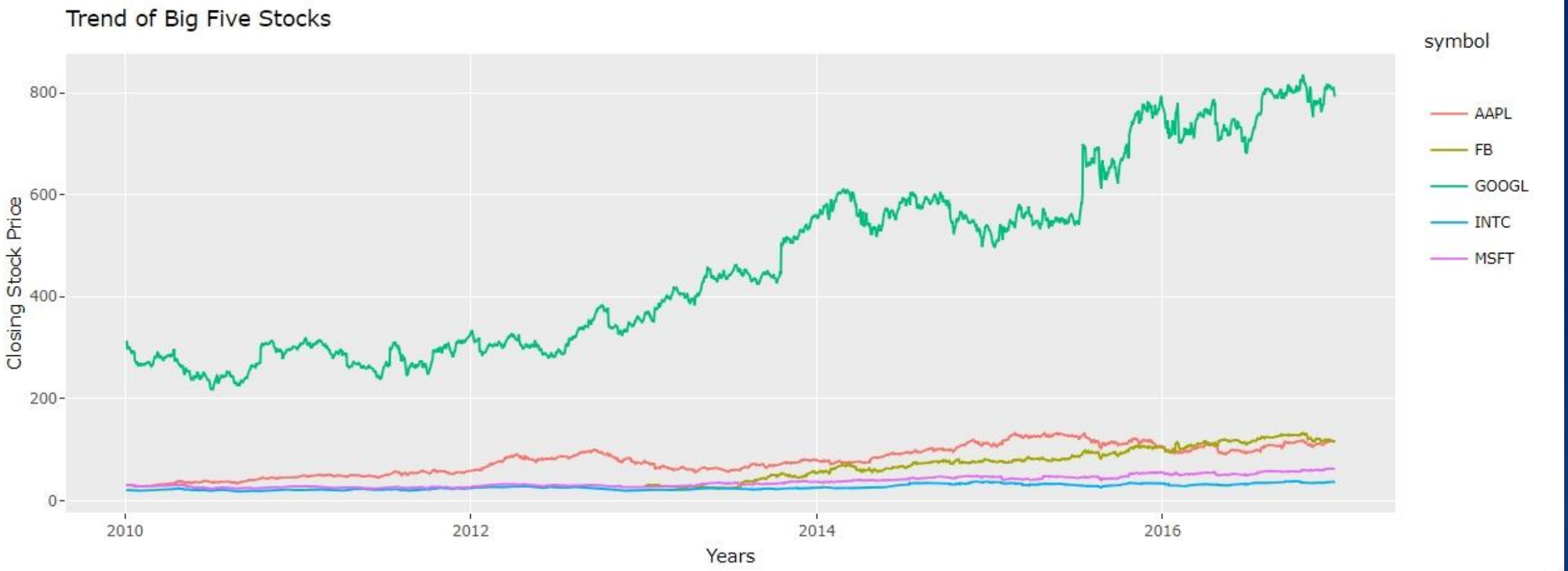
- Includes data from 2010 to 2016, for 508 companies, and has 4 files
- ***Prices.csv:*** raw data with daily stock prices, mostly from ***2010 to 2016.*** For companies that are new in stock market, the date range is shorter
- ***Prices-split-adjusted.csv:*** same as prices, with additional adjustments for splits
- ***Securities.csv:*** details of companies with respect to sectors, regions and start date
- ***Fundamentals.csv:*** metrics data from annual SEC 10K fillings with 78 predictors

EXPLORATORY DATA ANALYSIS

- Data was loaded from CSV
- Column names were modified
- Checked for NA values and eliminated them where appropriate
- Filtered the dataset to include data from Information Technology sector
- Merged the data from fundamentals and securities tables by ‘index’

TREND OF BIG FIVE STOCKS

- Considered big five stocks namely – Google, Apple, FB, Microsoft, Intel Corporation
- Used the ‘closing price’ data to study the trend over the past years


Trend of Big Five Stocks
symbol
AAPL
FB
GOOGL
INTC
MSFT
Closing Stock Price
800
600
400
200
0
2010
2012
2014
2016
Years

HOLT-WINTERS METHOD

- Model of time series behavior allows to forecast
- Allows to model three aspects of the time series: a typical value (average), a slope (trend) over time, and a cyclical repeating pattern (seasonality)
- Triple exponential smoothening to encode values from past and use them to predict values for present and future

Predicted value of Stocks using Holt-Winters Process for Google
graphset.me
Actual
Fitted
1000
800
600
400
200
Value
2010
2012
2014
2016
Time

THE Z-SCORE VALUE

- The Altman Z Score is used to predict the likelihood that a business will go bankrupt within the next two years.
- The formula is based on information found in the income/balance sheet of an organization.
- It is given by:

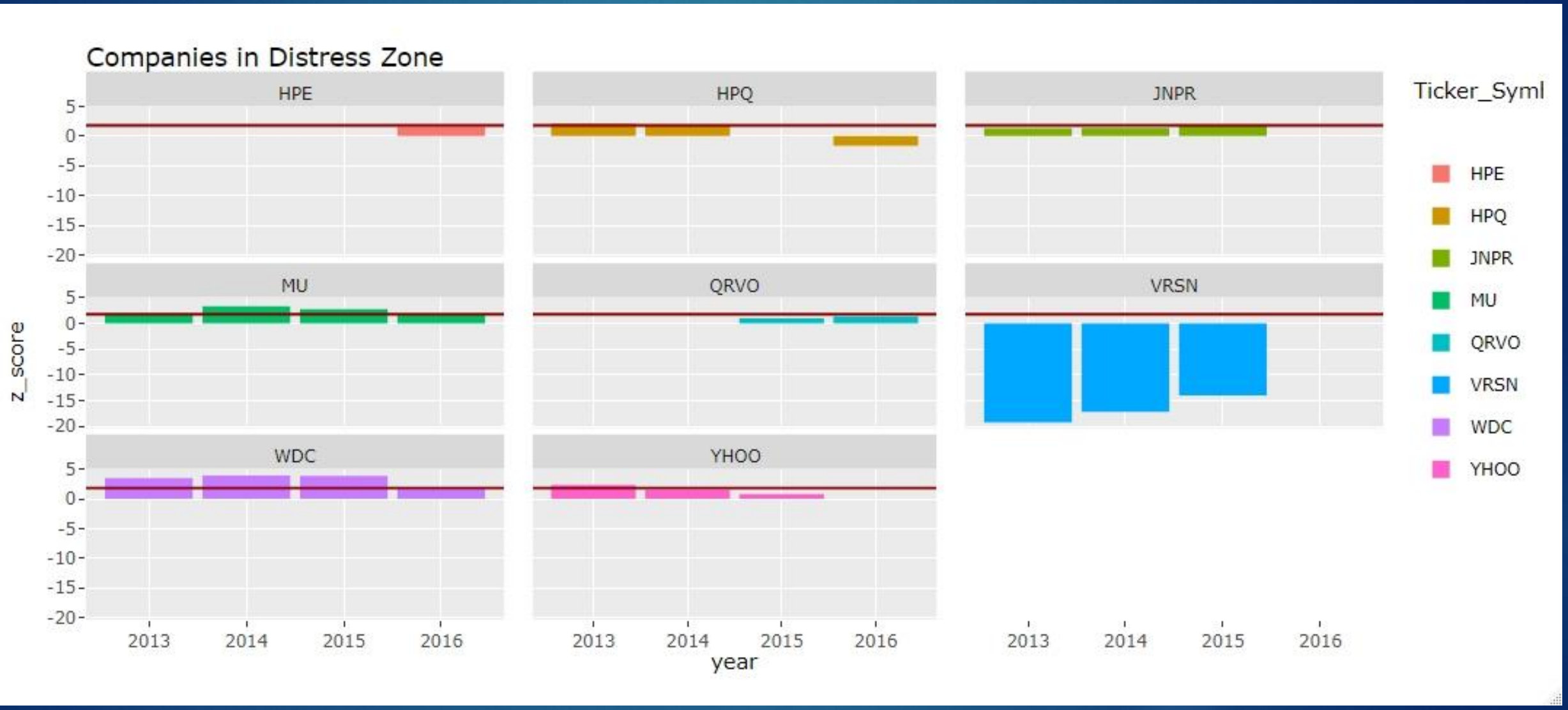
$$Z = 1.2*A + 1.4*B + 3.3*C + 0.6*D + 1*E$$

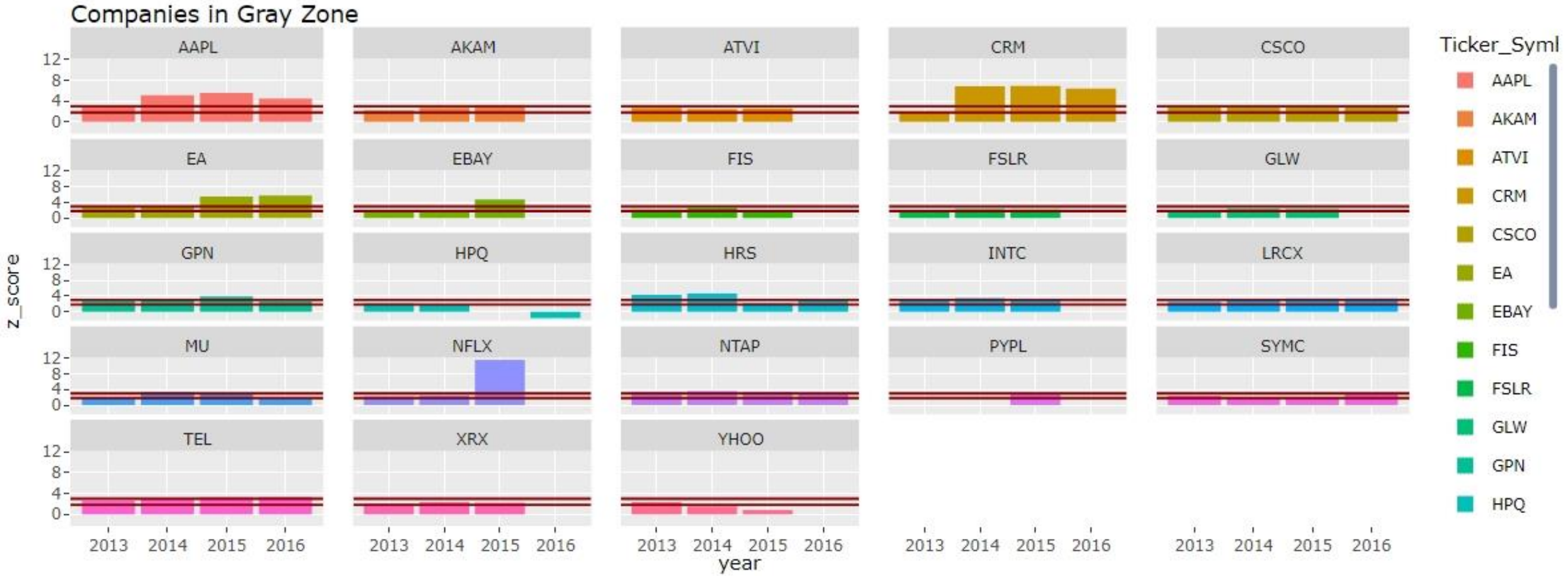
where,

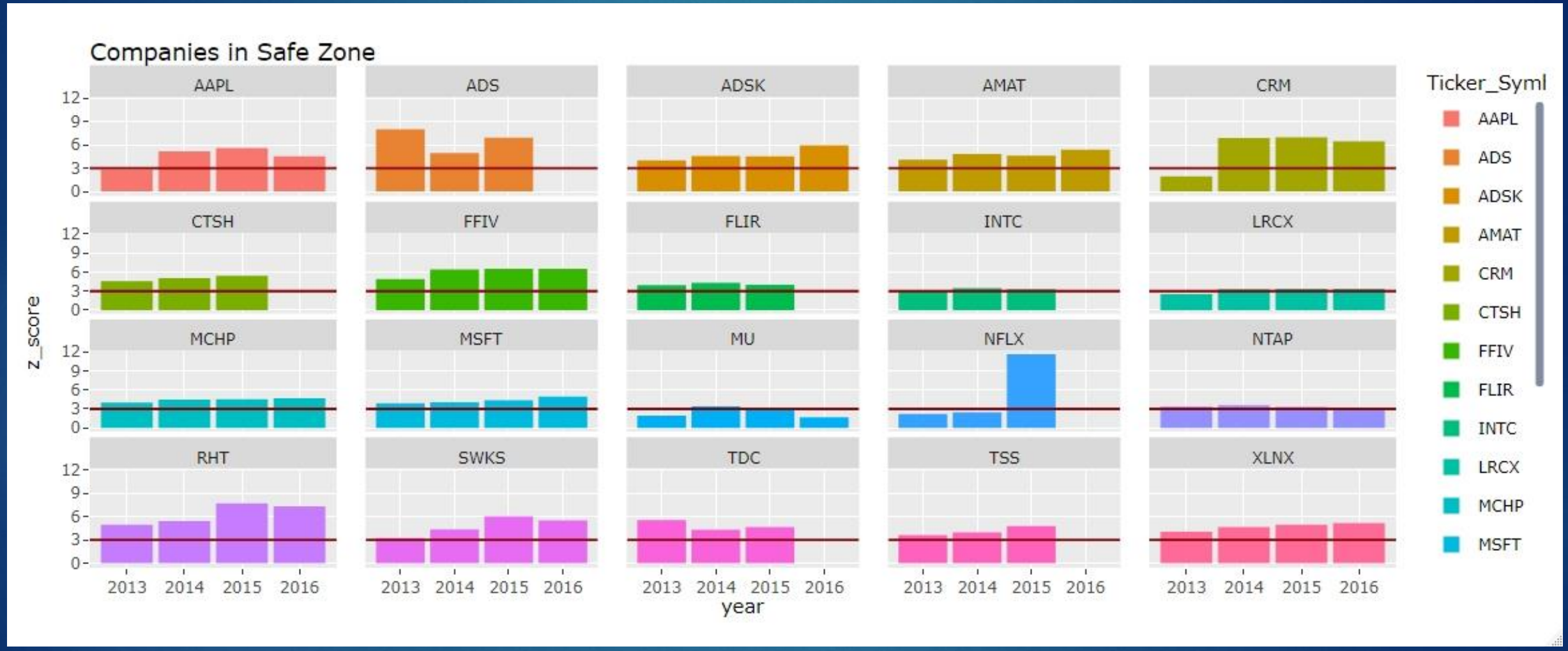
- A = working capital / total assets
- B = retained earnings / total assets
- C = earnings before interest and tax / total assets
- D = market value of equity / book value of equity
- E = sales / total assets

DISTRIBUTION OF Z-SCORES

Z-score range	Zone of discrimination	Distress level
Negative	Distress zones	Severely distressed
Between 0 and 9		Moderately distressed
Between 1.00 and 1.81		Mildly distressed
Between 1.81 and 2.40	Gray zones	Slightly distressed
Between 2.41 and 2.99		Stable, but weak
Between 3 and 4.99	Safe zones	Strong
Between 5 and 7.99		Highly strong
Greater than 8		Almost no distress


Companies in Distress Zone
HPE
HPQ
JNPR
MU
QRVO
VRSN
WDC
YHOO
z_score
5
0
-5
-10
-15
-20
2013
2014
2015
2016
year
Ticker_Syml
HPE
HPQ
JNPR
MU
QRVO
VRSN
WDC
YHOO


Companies in Gray Zone
AAPL
AKAM
ATVI
CRM
CSCO
EA
EBAY
FIS
FSLR
GLW
GPN
HPQ
HRS
INTC
LRCX
MU
NFLX
NTAP
PYPL
SYMC
TEL
XRX
YHOO
z_score
year
Ticker_Syml
AAPL
AKAM
ATVI
CRM
CSCO
EA
EBAY
FIS
FSLR
GLW
GPN
HPQ


Companies in Safe Zone
AAPL
ADS
ADSK
AMAT
CRM
CTSH
FFIV
FLIR
INTC
LRCX
MCHP
MSFT
MU
NFLX
NTAP
RHT
SWKS
TDC
TSS
XLNX
z_score
year
Ticker_Syml
AAPL
ADS
ADSK
AMAT
CRM
CTSH
FFIV
FLIR
INTC
LRCX
MCHP
MSFT

THANK YOU