**Week 12 Assignment**

Below steps are used to fetch compensation table from HTML and store it into table.

Step 1:

Download all html files or directly point to webpage link.

Step 2:

select summary compensation table out of all tables comparing specific keywords

Step 3:

Select required rows and columns (As of now range is hardcoded so same range will not work for other company's file)

Step 4:

Rename column names

Analysis(Coding is not yet done) -

Step 1:

For each company, multiple tables will be present for all years' of filing. Merge all tables and keep latest year's values.(Remove other rows)

So table will have single row for each year and each company.

Step 2:

Display charts performing analysis.

**R Code -**

#Installation of packages

install.packages("devtools")

devtools::install\_github("crubba/htmltable")

library(htmltab)

#MASCO CORPORATION

url <- "E:/ITMD529/masco.html"

#htmltab function with xpath is used to select "Compensation table" from file on the basis of below filters.

compensation.2015 <- htmltab(doc = url, which = "//\*[text()[contains(.,'Principal')] and text()[contains(.,'Position')]]/ancestor::table")

#Below code is used to select required columns and rows

compensation.table.2015 <- compensation.2015[3:4, c("V1", "V4","V8","V12","V16")]

#Renamed column names

compensation.table.2015 <-rename(compensation.table.2015, c("V1"="Name and Principal Position", "V4"="Year", "V8"="Salary", "V12"="Stock\_Awards", "V16"="Option\_Awards"))

compensation.table.2015