Assignment 1 – Ireland Kuhn

# https://www.stats.govt.nz/large-datasets/csv-files-for-download/ #  
# Import the dataset into R #  
dataset <- read.csv("C:\\Users\\imcguire\\Downloads\\annual-enterprise-survey-2021-financial-year-provisional-csv.csv")  
head(dataset)

## Year Industry\_aggregation\_NZSIOC Industry\_code\_NZSIOC Industry\_name\_NZSIOC  
## 1 2021 Level 1 99999 All industries  
## 2 2021 Level 1 99999 All industries  
## 3 2021 Level 1 99999 All industries  
## 4 2021 Level 1 99999 All industries  
## 5 2021 Level 1 99999 All industries  
## 6 2021 Level 1 99999 All industries  
## Units Variable\_code  
## 1 Dollars (millions) H01  
## 2 Dollars (millions) H04  
## 3 Dollars (millions) H05  
## 4 Dollars (millions) H07  
## 5 Dollars (millions) H08  
## 6 Dollars (millions) H09  
## Variable\_name Variable\_category Value  
## 1 Total income Financial performance 757,504  
## 2 Sales, government funding, grants and subsidies Financial performance 674,890  
## 3 Interest, dividends and donations Financial performance 49,593  
## 4 Non-operating income Financial performance 33,020  
## 5 Total expenditure Financial performance 654,404  
## 6 Interest and donations Financial performance 26,138  
## Industry\_code\_ANZSIC06  
## 1 ANZSIC06 divisions A-S (excluding classes K6330, L6711, O7552, O760, O771, O772, S9540, S9601, S9602, and S9603)  
## 2 ANZSIC06 divisions A-S (excluding classes K6330, L6711, O7552, O760, O771, O772, S9540, S9601, S9602, and S9603)  
## 3 ANZSIC06 divisions A-S (excluding classes K6330, L6711, O7552, O760, O771, O772, S9540, S9601, S9602, and S9603)  
## 4 ANZSIC06 divisions A-S (excluding classes K6330, L6711, O7552, O760, O771, O772, S9540, S9601, S9602, and S9603)  
## 5 ANZSIC06 divisions A-S (excluding classes K6330, L6711, O7552, O760, O771, O772, S9540, S9601, S9602, and S9603)  
## 6 ANZSIC06 divisions A-S (excluding classes K6330, L6711, O7552, O760, O771, O772, S9540, S9601, S9602, and S9603)

# Print out descriptive statistics for a selection of quantitative and categorical variables. #  
# Quantitative variables #  
quant\_vars <- c("Year", "Value")  
cat\_vars <- c("Variable\_code", "Variable\_name")  
  
# Summary statistics for quantitative variables #  
summary\_quant <- summary(dataset[, quant\_vars])  
print(summary\_quant)

## Year Value   
## Min. :2013 Length:41715   
## 1st Qu.:2015 Class :character   
## Median :2017 Mode :character   
## Mean :2017   
## 3rd Qu.:2019   
## Max. :2021

# Frequency table for categorical variables #  
for (var in cat\_vars) {  
 cat\_freq <- table(dataset[, var])  
 print(paste("Frequency table for", var))  
 print(cat\_freq)  
}

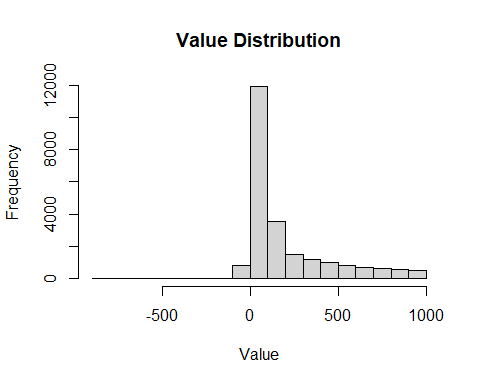
## [1] "Frequency table for Variable\_code"  
##   
## H01 H02 H03 H04 H05 H06 H07 H08 H09 H10 H11 H12 H13 H14 H17 H18   
## 1251 468 468 702 1251 1080 1251 1251 1251 1251 1251 1251 1251 153 468 468   
## H19 H20 H21 H22 H23 H24 H25 H26 H27 H28 H29 H30 H31 H32 H33 H34   
## 783 1251 1251 1251 1251 1251 1206 1206 1080 1080 1206 1251 1251 1206 1206 1143   
## H35 H36 H37 H38 H39 H40 H41   
## 1143 1206 1206 468 1251 1251 1251   
## [1] "Frequency table for Variable\_name"  
##   
## Additions to fixed assets   
## 1080   
## Closing stocks   
## 1251   
## Current assets   
## 1206   
## Current liabilities   
## 1206   
## Current ratio   
## 1206   
## Depreciation   
## 1251   
## Disposals of fixed assets   
## 1080   
## Fixed tangible assets   
## 1206   
## Government funding, grants and subsidies   
## 999   
## Indirect taxes   
## 1251   
## Interest and donations   
## 1251   
## Interest, dividends and donations   
## 1251   
## Liabilities structure   
## 1251   
## Margin on sales of goods for resale   
## 468   
## Non-operating expenses   
## 1251   
## Non-operating income   
## 1251   
## Opening stocks   
## 1251   
## Other assets   
## 1206   
## Other liabilities   
## 1206   
## Other purchases and operating expenses   
## 52   
## Other Purchases and operating expenses   
## 416   
## Purchases and other operating expenses   
## 783   
## Purchases of goods bought for resale   
## 468   
## Quick ratio   
## 1206   
## Redundancy and severance   
## 1251   
## Return on equity   
## 1251   
## Return on total assets   
## 1251   
## Salaries and wages paid   
## 1251   
## Salaries and wages to self employed commission agents   
## 153   
## Sales of goods and services   
## 531   
## Sales of goods not further processed   
## 468   
## Sales of other goods and services   
## 468   
## Sales, government funding, grants and subsidies   
## 252   
## Shareholders funds or owners equity   
## 1251   
## Surplus before income tax   
## 1251   
## Surplus per employee count   
## 1143   
## Total assets   
## 1251   
## Total equity and liabilities   
## 1251   
## Total expenditure   
## 1251   
## Total income   
## 1251   
## Total income per employee count   
## 1143

# Transform a variable #  
dataset$LogYear <- log(dataset$Year)

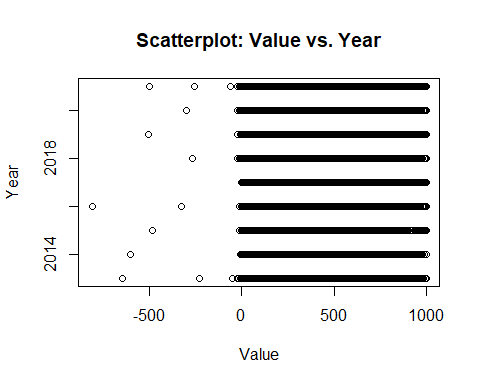
# Histogram for Value #  
dataset$Value <- as.integer(dataset$Value)

## Warning: NAs introduced by coercion

hist(dataset$Value, main="Value Distribution", xlab="Value")



# Scatterplot for Value vs. Year #  
plot(dataset$Value, dataset$Year, main="Scatterplot: Value vs. Year", xlab="Value", ylab="Year")



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.