Assignment 1

2025-06-01

library(readxl)

## Warning: package 'readxl' was built under R version 4.4.3

bma07\_e1\_23a\_v7 <- read\_excel("D:/College/Business Analytics/bma07.e1.23a-v7.xlsx")  
View(bma07\_e1\_23a\_v7)

##Code was taken from class Fall 2024 INTRODUCTION TO FINANCIAL ACCOUNTING (ACCT-23020-006) 2024 & the data seems to have been extracted by Wendy Tietz initially.

##To create a set of quantitative data, we would concatenate variables such as total, day of sale, year of sales, underwriter in order to produce mean, median, mode, min, max including the quadrants

summary(bma07\_e1\_23a\_v7[, c("Total", "Day of sale", "Year of sale", "Underwriter")])

## Total Day of sale Year of sale Underwriter   
## Min. : 118.9 Min. : 1.00 Min. :2017 Min. :101.0   
## 1st Qu.: 455.7 1st Qu.: 7.00 1st Qu.:2018 1st Qu.:323.0   
## Median : 598.7 Median :15.00 Median :2019 Median :546.0   
## Mean : 660.9 Mean :14.51 Mean :2019 Mean :547.8   
## 3rd Qu.: 853.6 3rd Qu.:22.00 3rd Qu.:2020 3rd Qu.:773.0   
## Max. :1600.1 Max. :28.00 Max. :2020 Max. :999.0

##For the following 3 codes, we will simply create a table for varibles such as types of insurance, sales/total by sales people along with the different states the total may originate from

table(bma07\_e1\_23a\_v7$`Insurance type`)

##   
## Auto Disability Life   
## 24954 15438 6

table(bma07\_e1\_23a\_v7$Salesperson)

##   
## Anh Anna Audrey Caleb Cami Curtis Henry Isaac   
## 2614 3280 2506 3420 3294 2646 3174 2406   
## Leah Nate Oliver Owen Riley Scarlett Tran   
## 2392 2448 2424 2478 2436 2586 2294

table(bma07\_e1\_23a\_v7$`State or Province`)

##   
## AB AL AR AZ BC CA CO CT DC DE FL GA IA ID IL IN   
## 3901 368 411 383 990 413 421 421 439 452 429 426 421 422 403 402   
## KS KY LA MA MB MD ME MI MN MO MS MT NB NC ND NE   
## 388 394 386 379 965 425 475 425 386 389 413 419 928 414 389 401   
## NH NJ NL NM NS NV NY OH OK ON OR PA PE QC RI SC   
## 419 413 953 411 942 432 447 443 396 6655 452 401 1005 2898 395 404   
## SD SK TN TX UT VA VT WA WI WV WY   
## 414 962 400 364 421 437 415 416 407 397 421

##One method to create a new variable is to simply take log of an existing one. Conversely we can display the new variable by using summary since the data is easier to read via that method, for comparison, summary of total has been included directly below it

bma07\_e1\_23a\_v7$logtotal <- log(bma07\_e1\_23a\_v7$Total)

summary(bma07\_e1\_23a\_v7$logtotal)

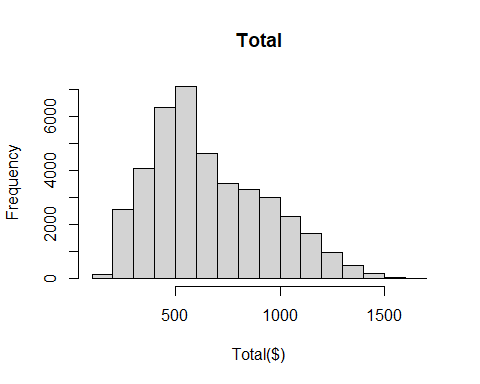
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.779 6.122 6.395 6.405 6.750 7.378

summary(bma07\_e1\_23a\_v7$Total)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 118.9 455.7 598.7 660.9 853.6 1600.1

##To create a histogram we only require one variable in which case we’ll use Total

hist(bma07\_e1\_23a\_v7$Total, main = "Total", xlab = "Total($)")



##Since the data chosen is extremely dense, we’ll create a new category for totals that are above 1250 for our scatterplot to make it easier to read.

high\_total <- bma07\_e1\_23a\_v7[bma07\_e1\_23a\_v7$Total > 1250, ]

plot(high\_total$`Day of sale`, high\_total$Total, main = "Scatterplot of Day of Sale vs Total (>$1250)", xlab = "Day of Sale", ylab = "Total ($)")

