Data Visualisation

Bonny Runameso

2022-11-19

```
library(readxl)
Sales <- read_excel("C:/Users/bonny/Desktop/Data Visualisation/Sales.xlsx")
```

View(Sales) replace n/a with NA

```
Sales[Sales[,] == "n/a"] <- 'NA'
```

drop those values we cannot calculate from other given variables

```
library(tidyr)
summary_Sales <- Sales %>% drop_na()
summary_Sales

## # A tibble: 49 x 11
##        ID Paid  Region  Country `Item Type` `Units Sold` `Unit Price` `Unit
Cost`
##     <dbl> <chr> <chr>   <chr>   <chr>       <chr>        <chr>        <chr>
##  1  1002 No    Centra~ Antigu~ Baby Food   552          255.28
159.419999~
##  2  1003 No    Centra~ Panama  Snacks      2167         152.5800000~ 97.44
##  3  1004 No    europe  Czech ~ Beverages   4778         47.45        31.79
##  4  1005 No    Asia    North ~ Cereal      9016         205.7
117.11
##  5  1006 No    Asia    Sri La~ Snacks      7542         152.5800000~ 97.44
##  6  1007 No    Middle~ Morocco Personal C~ 48           81.73        56.67
##  7  1008 No    Austra~ Federa~ clothees    NA           NA           NA
##  8  1009 No    Europe  Bosnia~ Clothes     927          109.28
35.8400000~
##  9  1010 No    Middle~ Afghan~ Clothes     8841         109.28
35.8400000~
## 10  1011 No    Sub-Sa~ Ethiop~ Baby Food   9817         NA
159.419999~
## # ... with 39 more rows, and 3 more variables: Total Revenue <chr>,
## #   Total Cost <chr>, Total Profit <chr>

Sales$`Total Revenue`<- as.character(as.numeric(Sales$`Unit
Price`)*as.numeric(Sales$`Units Sold`))

## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion
```

```
Sales$`Total Cost`<- as.character(as.numeric(Sales$`Unit
Cost`)*as.numeric(Sales$`Units Sold`))
```

Warning: NAs introduced by coercion
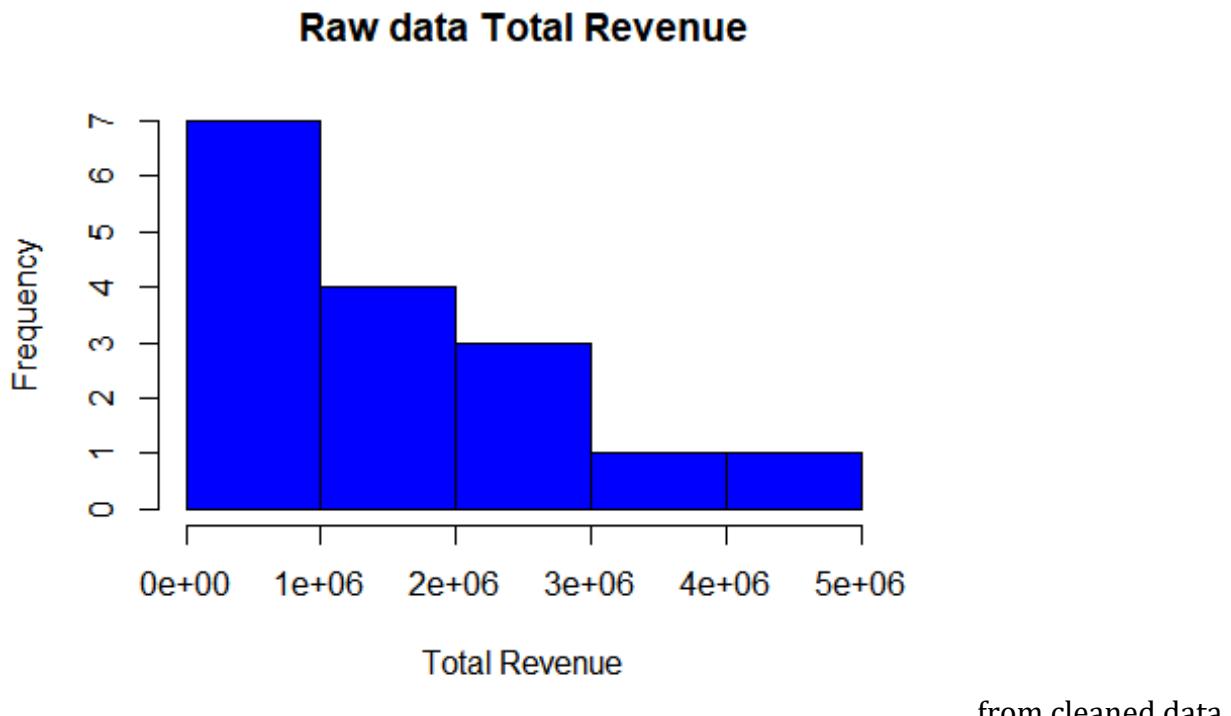
Warning: NAs introduced by coercion

```
Sales$`Total Profit`<- as.character(as.numeric(Sales$`Total
Revenue`)*as.numeric(Sales$`Total Cost`))
```

1b) from raw data

```
hist(as.numeric(Sales[1:20,]$`Total Revenue`),main = "Raw data Total
Revenue",xlab="Total Revenue",col = "blue")
```


from cleaned data
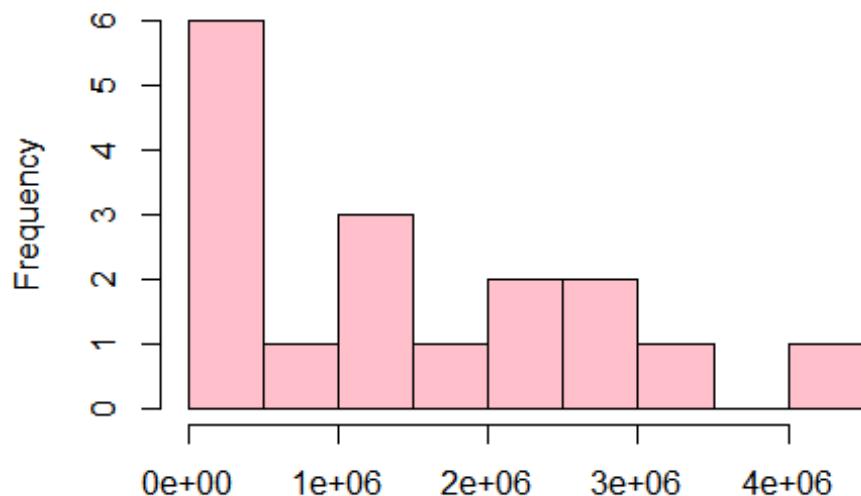
```
hist(as.numeric(summary_Sales[1:20,]$`Total Revenue`),main = "Cleaned data
Total Revenue",col = "pink")
```

Warning in hist(as.numeric(summary_Sales[1:20,]$`Total Revenue`), main =
"Cleaned data Total Revenue", : NAs introduced by coercion
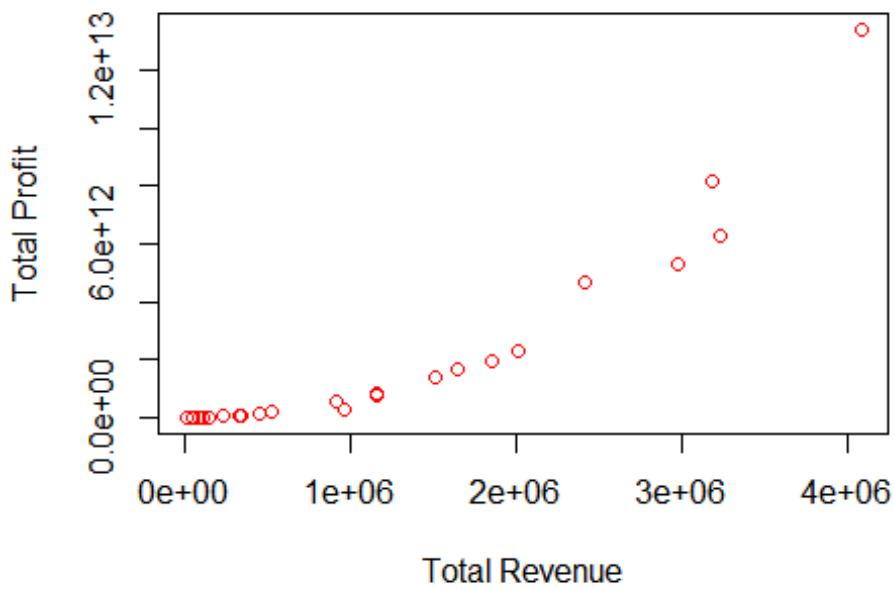
Cleaned data Total Revenue

plots plot for raw data

```
plot(Sales[1:30,]$`Total Revenue`,Sales[1:30,]$`Total Profit`,xlab = "Total
Revenue",
     ylab = "Total Profit",col="red",main = "Total Revenue vs Total profit
Raw Data")
```

Total Revenue vs Total profit Raw Data
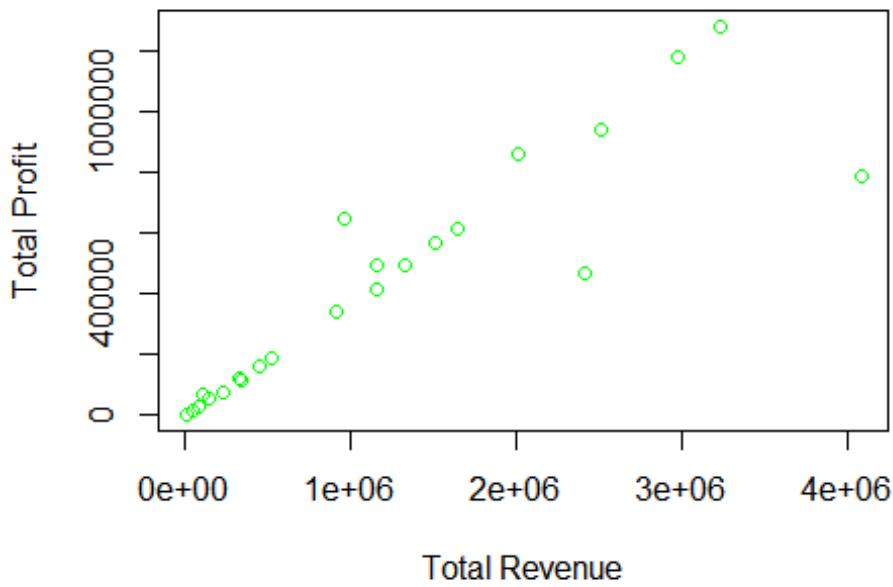

plots for cleaned

```
plot(summary_Sales[1:30,]$`Total Revenue`,summary_Sales[1:30,]$`Total
Profit`,xlab = "Total Revenue",
     ylab = "Total Profit",col="green",main = "Total Revenue vs Total profit
Raw Data")

## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduced by
coercion

## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduced by
coercion
```
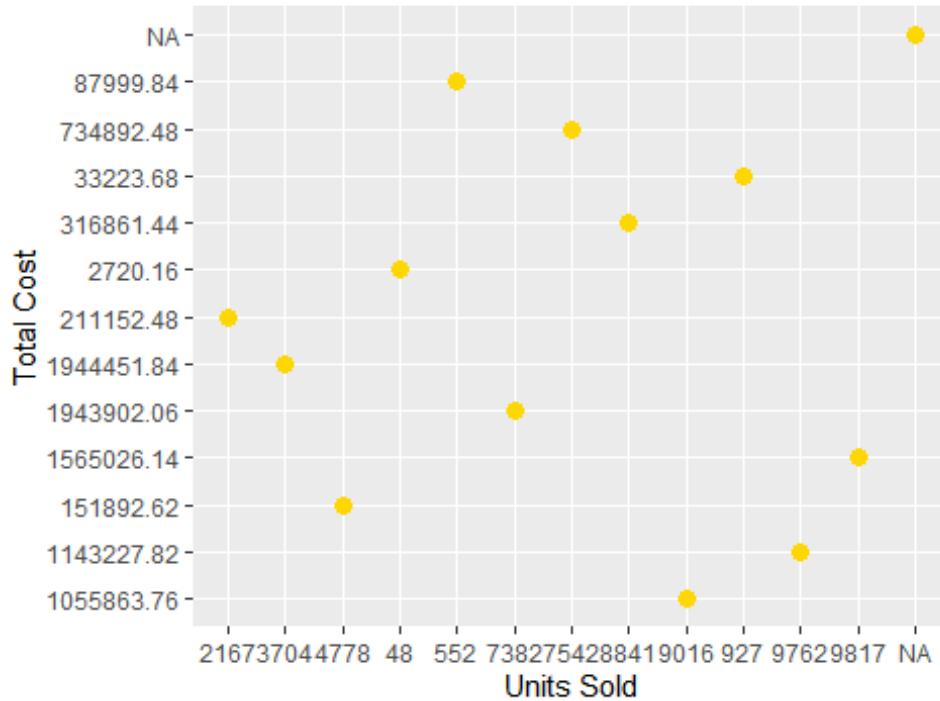
Total Revenue vs Total profit Raw Data

2c)
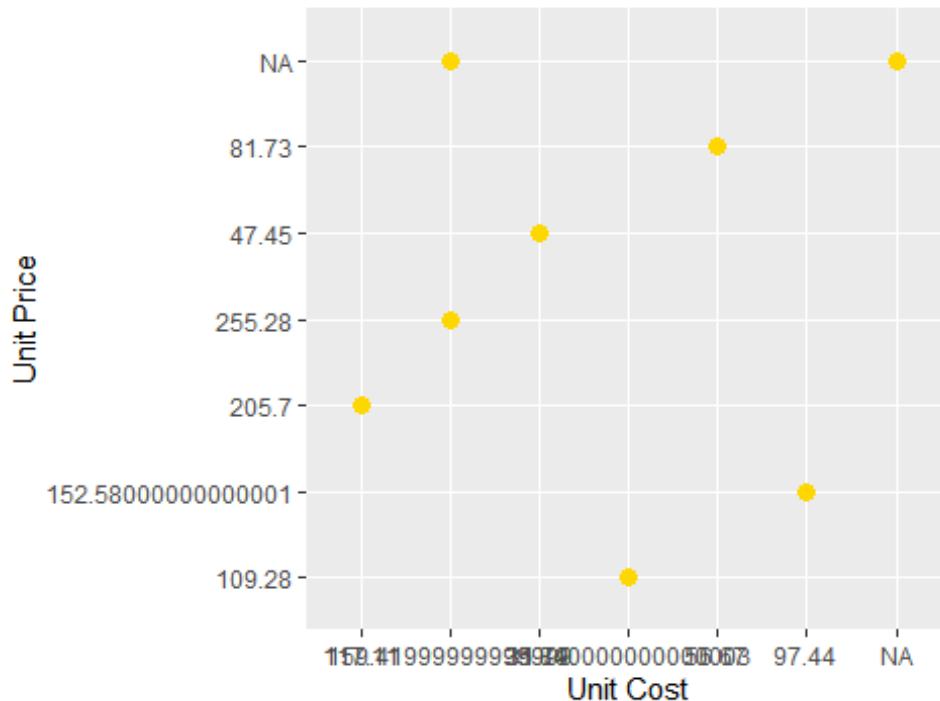
```
library(ggplot2)
ggplot(summary_Sales[1:13,])+geom_point(aes(y =`Total Cost`, x = `Units
Sold`), col = 'gold', size = 3) + ggtitle(" Unit Sold vs Total Cost") +
theme(plot.title = element_text(hjust = 0.5))
```

Unit Sold vs Total Cost
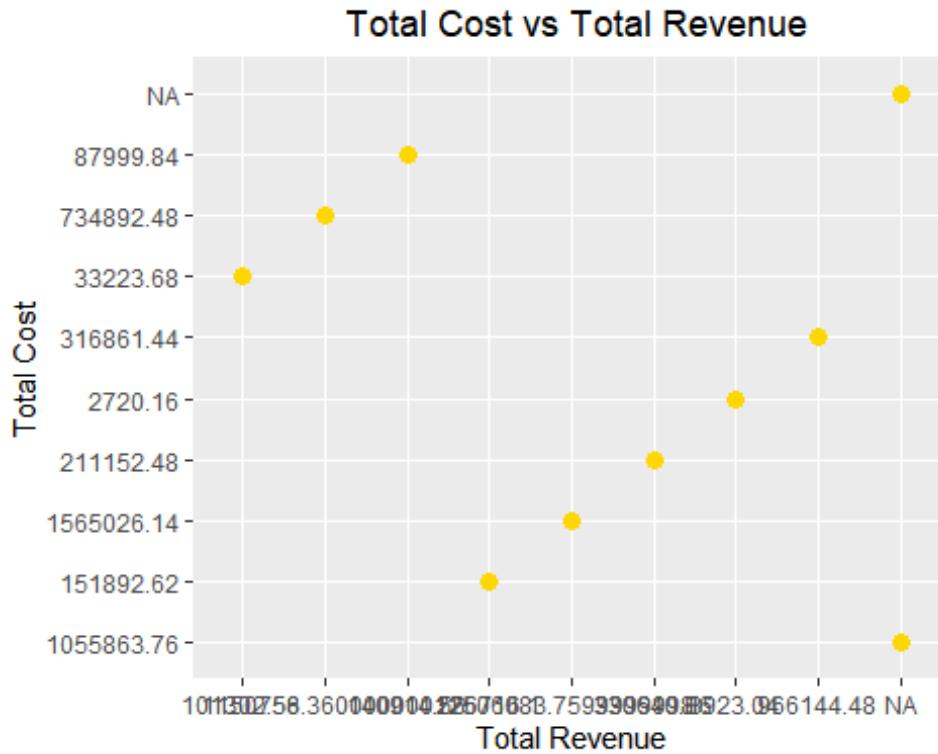

```
ggplot(summary_Sales[1:10,])+geom_point(aes(y =`Unit Price`, x = `Unit
Cost`), col = 'gold', size = 3) + ggtitle(" Unit Price vs Unit Cost") +
theme(plot.title = element_text(hjust = 0.5))
```

Unit Price vs Unit Cost

```r
ggplot(summary_Sales[1:10,])+geom_point(aes(y =`Total Cost`, x = `Total
Revenue`), col = 'gold', size = 3) + ggtitle(" Total Cost vs Total Revenue")
+ theme(plot.title = element_text(hjust = 0.5))
```


Total Cost vs Total Revenue

2d)

```r
library(tidyr)
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.3

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

summary_Sales <- summary_Sales %>%
arrange(desc(as.numeric(summary_Sales$`Total Cost`)))

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

summary_Sales$`Total Cost`
```

```
##  [1] "4842231.04"          "3289399.36"          "2564429.6"
##  [4] "2032380.94"          "2028405.78"          "2013818.18"
##  [7] "1944451.84"          "1943902.06"          "1786957.38"
## [10] "1672145.5"           "1565026.14"          "1273990.54"
## [13] "1143227.82"          "1055863.76"          "1024751.76"
## [16] "942172.2"            "824520.24"           "734892.48"
## [19] "678323.4"            "662970.63"           "657689.76"
## [22] "569448.24"           "329152.32"           "316861.44"
## [25] "283258.08"           "240999.99"           "226726.28"
## [28] "211152.48"           "193074.69"           "175903.68"
## [31] "165795.84"           "159898.26"           "151892.62"
## [34] "134184.95999999999"  "87999.84"            "55378.18"
## [37] "35638"               "34309.360000000001"  "33223.68"
## [40] "9618.7999999999993"  "8184.96"             "2720.16"
## [43] "NA"                  "NA"                  "NA"
## [46] "NA"                  "NA"                  "NA"
## [49] "NA"
```

2e)

TYPES OF EXPLORATORY DATA ANALYSIS:

Univariate Non-graphical Multivariate Non-graphical Univariate graphical Multivariate graphical 2f)

```
summary_Sales$Country[1] <- "Angola"
summary_Sales$`Item Type`[1] <- "Chicken Slice"
summary_Sales$`Units Sold`[1] <- "233"
```

2g)

```
#summary_Sales[summary_Sales[,"Paid"] == "No"] <- "Yes"
summary_Sales$Paid[summary_Sales$Paid == 'No'] <- 'Yes'
```

2h)

```
mean(as.numeric(summary_Sales$`Unit Cost`),na.rm = T)
```

```
## [1] 161.02
```

```
mean(as.numeric(summary_Sales$`Unit Price`),na.rm = T)
```

```
## [1] 236.7228
```

```
mean(as.numeric(summary_Sales$`Total Cost`),na.rm = T)
```

```
## [1] 907308.9
```

```
mean(as.numeric(summary_Sales$`Total Profit`),na.rm = T)
```

```
## [1] 415339.7
```

```
mean(as.numeric(summary_Sales$`Total Revenue`),na.rm = T)
```

```
## [1] 1303415

x <- quantile(as.numeric(summary_Sales$`Unit Cost`,na.rm = F),0.25,na.rm = T)

## Warning in quantile(as.numeric(summary_Sales$`Unit Cost`, na.rm = F),
0.25, :
## NAs introduced by coercion

x

##    25%
## 35.84

y <- quantile(as.numeric(summary_Sales$`Unit Cost`),0.75,na.rm = T)
y

##     75%
## 263.33

inter_quantile <- y-x
inter_quantile

##     75%
## 227.49

x <- quantile(as.numeric(summary_Sales$`Unit Price`),0.25,na.rm = T)

x

##     25%
## 95.505

y <- quantile(as.numeric(summary_Sales$`Unit Price`),0.75,na.rm = T)
y

##     75%
## 421.89

inter_quantile <- y-x
inter_quantile

##      75%
## 326.385

#quantile
x <- quantile(as.numeric(summary_Sales$`Total Cost`),0.25,na.rm = T)
x

##        25%
## 161372.7

y <- quantile(as.numeric(summary_Sales$`Total Cost`),0.75,na.rm = T)
y
```

```
##      75%
## 1492267

inter_quantile <- y-x
inter_quantile

##      75%
## 1330895

#quantile
x <- quantile(as.numeric(summary_Sales$`Total Revenue`),0.25,na.rm = T)
x

##     25%
## 259881

y <- quantile(as.numeric(summary_Sales$`Total Revenue`),0.75,na.rm = T)
y

##      75%
## 2276051

inter_quantile <- y-x
inter_quantile

##      75%
## 2016170

#quantile
x <- quantile(as.numeric(summary_Sales$`Total Profit`),0.25,na.rm = T)
x

##       25%
## 81582.83

y <- quantile(as.numeric(summary_Sales$`Total Profit`),0.75,na.rm = T)
y

##       75%
## 632735.6

inter_quantile <- y-x
inter_quantile

##       75%
## 551152.7

#quantile
x <- quantile(as.numeric(summary_Sales$`Units Sold`),0.25,na.rm = T)
x

##       25%
## 2956.25
```

```
y <- quantile(as.numeric(summary_Sales$`Unit Sold`),0.75,na.rm = T)
y
```

```
## 75%
##  NA
```

```
inter_quantile <- y-x
inter_quantile
```

```
## 75%
##  NA
```

2i) dataset before

```
summary_Sales[1,]
```

```
## # A tibble: 1 x 11
##      ID Paid  Region   Country `Item Type` `Units Sold` `Unit Price` `Unit
Cost`
##   <dbl> <chr> <chr>    <chr>   <chr>       <chr>        <chr>        <chr>
## 1  1047 Yes   Sub-Sah~ Angola  Chicken Sl~ 233          651.21
524.96
## # ... with 3 more variables: Total Revenue <chr>, Total Cost <chr>,
## #   Total Profit <chr>
```

```
summary_Sales[1,2] <- "Paid_Type"
```

after

```
summary_Sales[1,]
```

```
## # A tibble: 1 x 11
##      ID Paid    Region  Country `Item Type` `Units Sold` `Unit Price` `Unit
Cost`
##   <dbl> <chr>   <chr>   <chr>   <chr>       <chr>        <chr>        <chr>
## 1  1047 Paid_~  Sub-Sa~ Angola  Chicken Sl~ 233          651.21
524.96
## # ... with 3 more variables: Total Revenue <chr>, Total Cost <chr>,
## #   Total Profit <chr>
```