Smith SCHOOL OF BUSINESS Queen's University

Global Master of Management Analytics

GMMA 860
Acquisition and Management of Data

Alexander Scott

Assignment #1
March 22, 2020

William Chak Lim Chan

Order of files:

Filename	Pages	Comments and/or Instructions

Additional Comments:

Part 1

#Question A
sqldf('select PRCDDA, WSWORTHWPV from wealth order by WSWORTHWPV desc')
#ANSWER: PRCDDA = 5602, WSWORTHWPV = 2,365,581,172

```
> #Question A
> sqldf('select PRCDDA, WSWORTHWPV from wealth order by WSWORTHWPV desc')
    PRCDDA WSWORTHWPV
1     5602 2365581172
2     2291 1288397495
3     5889 1176844243
4     2260 1069088235
5     5391 1051627463
6     1955  972946020
```

#Question B
sqldf('SELECT STYAPT/(STYHOUS+STYAPT) AS APTPERCENTAGE from demo WHERE APTPERCENTAGE < .50')
#ANSWER: 5019 DAs are less than 50% condos

```
993    0.000000000
994    0.000000000
995    0.214285714
996    0.077966102
997    0.000000000
998    0.000000000
999    0.000000000
1000   0.000000000
 [ reached 'max' / getOption("max.print") -- omitted 4019 rows ]
```

#Question C
sqldf('SELECT wealth.PRCDDA, demo.BASPOP, wealth.WSWORTHWPV FROM wealth JOIN demo ON wealth.PRCDDA = demo.PRCDDA order by wealth.WSWORTHWPV')
#ANSWER: ID = 190, Population = 0, Total Net Worth = 0

```
> #Question C
> sqldf('SELECT wealth.PRCDDA, demo.BASPOP, wealth.WSWORTHWPV FROM wealth JOIN demo ON wealth.PRCDDA = demo.PRCDDA order by wealth.WSWORTHWPV')
    PRCDDA BASPOP WSWORTHWPV
1      190      0        0.0
2      207      0        0.0
3      422      0        0.0
4      496      0        0.0
5      862      0        0.0
6     1009      0        0.0
```

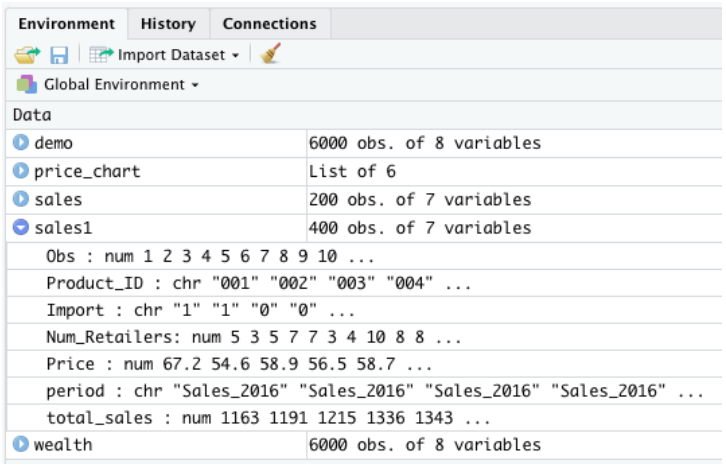
#Question D
#ANSWER: In these two datasets, wealth and demo, there would be no differences in output (on both Full Outer Join and Right Join) because of the join on ID. The ID column in both data sets has exactly 6000 unique IDs that match perfectly 1:1. In the event where demo had ID 1,2,4 and wealth had 1,2,3, there would be a difference in the output.

#Question E
#sqldf('SELECT avg(WSDEBTB) FROM wealth JOIN demo ON wealth.PRCDDA = demo.PRCDDA where ACTER < 50')
#ANSWER: 11,503,957

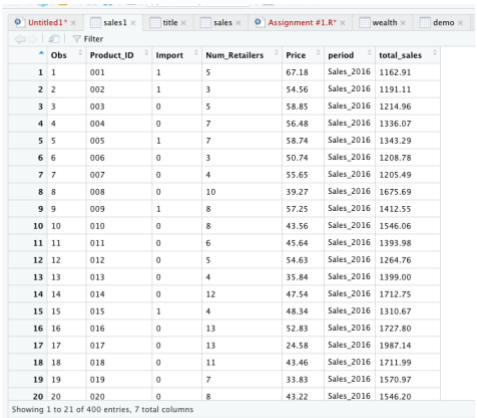
```
> sqldf('SELECT avg(WSDEBTB) FROM wealth JOIN demo ON wealth.PRCDDA = demo.PRCDDA where ACTER < 50')
  avg(WSDEBTB)
1     11503957
```

Part 2

#Question A



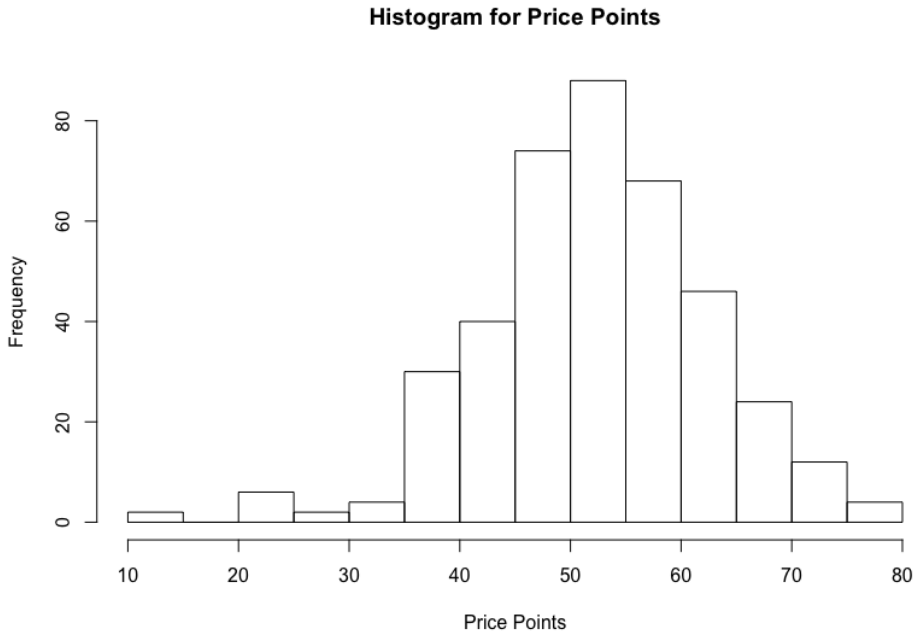
#Question B



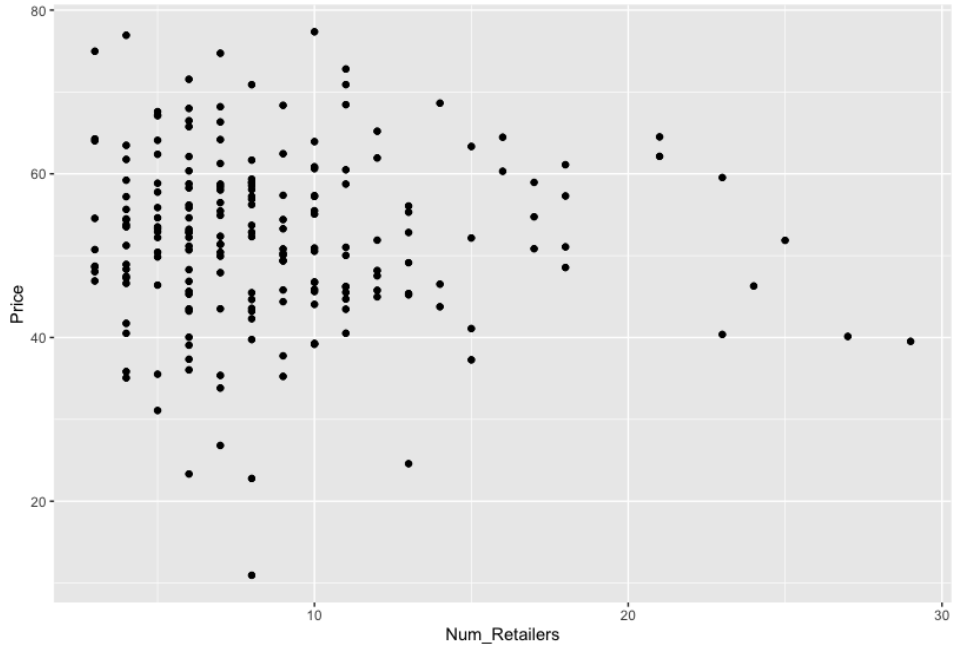
	Obs	Product_ID	Import	Num_Retailers	Price	period	total_sales
1	1	001	1	5	67.18	Sales_2016	1162.91
2	2	002	1	3	54.56	Sales_2016	1191.11
3	3	003	0	5	58.85	Sales_2016	1214.96
4	4	004	0	7	56.48	Sales_2016	1336.07
5	5	005	1	7	58.74	Sales_2016	1343.29
6	6	006	0	3	50.74	Sales_2016	1208.78
7	7	007	0	4	55.65	Sales_2016	1205.49
8	8	008	0	10	39.27	Sales_2016	1675.69
9	9	009	1	8	57.25	Sales_2016	1412.55
10	10	010	0	8	43.56	Sales_2016	1546.06
11	11	011	0	6	45.64	Sales_2016	1393.98
12	12	012	0	5	54.63	Sales_2016	1264.76
13	13	013	0	4	35.84	Sales_2016	1399.00
14	14	014	0	12	47.54	Sales_2016	1712.75
15	15	015	1	4	48.34	Sales_2016	1310.67
16	16	016	0	13	52.83	Sales_2016	1727.80
17	17	017	0	13	24.58	Sales_2016	1987.14
18	18	018	0	11	43.46	Sales_2016	1711.99
19	19	019	0	7	33.83	Sales_2016	1570.97
20	20	020	0	8	43.22	Sales_2016	1546.20

Showing 1 to 21 of 400 entries, 7 total columns

#Question C

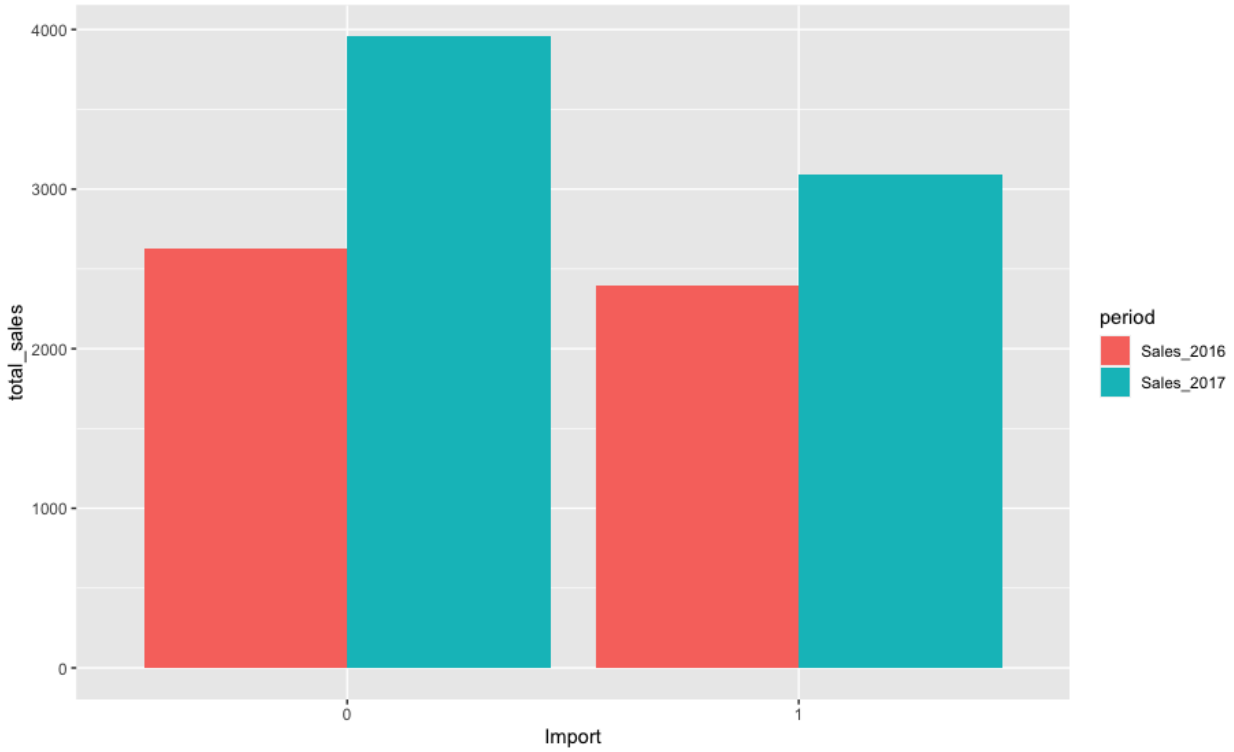


#Question D



The smaller number of retailers there are, the greater the spread and variation of price points.

#Question E



Total sales grew year-over-year for non-import and imported products. This could suggest a strength in demand for products as a whole. Keep an eye out for decreased YoY growth on imported products. Assumption (1 = Yes, 0 = No)

R Script

```
#ASSIGNMENT #1 - 860

library("tidyverse")
library("sqldf")
library("readxl")

wealth <- read_excel("OneDrive - Queen's University/Global Master of Management Analytics/GMMA 860 - Acquisition and Management of Data/Assignment #1/GMMA860_Assignment1_Data.xlsx", sheet = "Wealth")
demo <- read_excel("OneDrive - Queen's University/Global Master of Management Analytics/GMMA 860 - Acquisition and Management of Data/Assignment #1/GMMA860_Assignment1_Data.xlsx", sheet = "Demo")

#PART 1

#Question A
sqldf('select PRCDDA, WSWORTHWPV from wealth order by WSWORTHWPV desc')
#ANSWER: PRCDDA = 5602, WSWORTHWPV = 2,365,581,172

#Question B
```

```
sqldf('SELECT STYAPT/(STYHOUS+STYAPT) AS APTPERCENTAGE from demo WHERE APTPERCENTAGE <
 .50')
#ANSWER: 5019 DAs are less than 50% condos

#Question C
sqldf('SELECT wealth.PRCDDA, demo.BASPOP, wealth.WSWORTHWPV FROM wealth JOIN demo ON
 wealth.PRCDDA = demo.PRCDDA order by wealth.WSWORTHWPV')
#ANSWER: ID = 190, Population = 0, Total Net Worth = 0

#Question D
#In these two datasets, wealth and demo, there would be no differences in output (on both Full Outer
 Join and Right Join) because of the join on ID. The ID column in both data sets has exactly 6000 unique
 IDs that match perfectly 1:1. In the event where demo had ID 1,2,4 and wealth had 1,2,3, there would
 be a difference in the output.

#Question E
sqldf('SELECT avg(WSDEBTB) FROM wealth JOIN demo ON wealth.PRCDDA = demo.PRCDDA where
 ACTER < 50')
#ANSWER: 11,503,957

#Part 2

library(stringr)

sales <- read_excel(file.choose())

str(sales)

sales$Product_ID <- as.character(as.numeric(sales$Product_ID))  # Convert Product_ID to character
sales$Import <- as.character(as.numeric(sales$Import))  # Convert Import to character
sales$Num_Retailers <- as.numeric(as.character(sales$Num_Retailers))  # Convert Num_Retailers to
 character
sales$Price <- as.numeric(gsub("\\$","",sales$Price)) #Remove $ sign and convert to numeric
sales$Product_ID <- str_pad(sales$Product_ID, 3, pad = "0") #Pad Product ID with zeros
sales1 <- gather(sales, period, total_sales, Sales_2016, Sales_2017) #Gather the data for 2016 and 2017

price_chart <- hist(sales1$Price, main = "Histogram for Price Points",xlab ="Price Points") #Create Price
 Point Histogram

price_plot <- ggplot(sales1, aes(y=Price, x=Num_Retailers)) + geom_point() #Create Scatterplot of price
 vs. number of retailers
price_plot #View the Plot

price_bar <- ggplot(sales1, aes(x=Import, y=total_sales, fill=period)) + geom_bar(stat = "identity",
 position=position_dodge()) #Create a sales bar chart of import vs. non-imported products
price_bar #View the Chart
```