Quality Alloys

Manish Grewal

26/07/2020

Table of Contents

[1. How many people visit the website, and how do they come to website 4](#_Toc47214464)

[Unique visits over time 4](#_Toc47214465)

[Summary of unique visits by period 4](#_Toc47214466)

[Mean unique visits by period 5](#_Toc47214467)

[How do people come to the website: 5](#_Toc47214468)

[2. Is the website generating interest, and does this interest yield actual sales? 10](#_Toc47214469)

[Summary of Lbs. Sold by Period: 10](#_Toc47214470)

[Mean Lbs. Sold by period 11](#_Toc47214471)

[Comparison of Unique Visits and Lbs. Sold after Z-score standardization: 11](#_Toc47214472)

[Scatterplot of Lbs. Sold and Unique Vists 12](#_Toc47214473)

[Coefficient of correlation of Lbs. Sold and Unique Visits 13](#_Toc47214474)

[3. Check whether the data variables (if applicable) follows the normality assumption. 13](#_Toc47214475)

[Normality assumption for Revenue 13](#_Toc47214476)

[Normality assumption for Profit 13](#_Toc47214477)

[Normality assumption for Lbs. Sold 14](#_Toc47214478)

[Normality assumption for Visits 15](#_Toc47214479)

[Normality assumption for UniqueVisits 16](#_Toc47214480)

[PART 1 - EXECUTIVE SUMMARY 17](#_Toc47214481)

[Evaluation of the goals of the company website 17](#_Toc47214482)

[(a) drive new sales - NOT MET 17](#_Toc47214483)

[(b) make product and contact information available - MAY BE MET 17](#_Toc47214484)

[(c) give or add legitimacy to its brand. - MAY BE MET 17](#_Toc47214485)

[Recommendations from demographics: 18](#_Toc47214486)

[PART 2 - DATA ANALYSIS 18](#_Toc47214487)

[1. Column Charts 18](#_Toc47214488)

[a. Unique visits over time 18](#_Toc47214489)

[b. Revenue over time, 19](#_Toc47214490)

[c. Profit over time 19](#_Toc47214491)

[d. Pounds sold over time. 20](#_Toc47214492)

[2. Summary by period 20](#_Toc47214493)

[a. Summary table of Initial Period 21](#_Toc47214494)

[b. Summary table of Pre-promotion Period 21](#_Toc47214495)

[c. Summary table of Promotion Period 22](#_Toc47214496)

[d. Summary table of Post-promotion Period 22](#_Toc47214497)

[3. Column chart of the means over the four periods 22](#_Toc47214498)

[a. Mean Visits by Period 22](#_Toc47214499)

[b. Mean Unique Visits by Period 23](#_Toc47214500)

[c. Mean Revenue by Period 23](#_Toc47214501)

[d. Mean Profit by Period 24](#_Toc47214502)

[e. Mean Lbs. Sold by Period 24](#_Toc47214503)

[4. Summary of findings thus far 25](#_Toc47214504)

[5. Scatter plot of Revenue and Lbs. Sold 25](#_Toc47214505)

[b. Coefficient of Correlation 26](#_Toc47214506)

[6. Scatter plot of Visits and Revenue 26](#_Toc47214507)

[7. Summary of results 27](#_Toc47214508)

[8. Modeling data 27](#_Toc47214509)

[a. Summary of Lbs. Sold data per week 27](#_Toc47214510)

[b. Histogram of the pounds of material sold data. 27](#_Toc47214511)

[c. Histogram description 28](#_Toc47214512)

[d. Empirical rule 28](#_Toc47214513)

[e. Refine empirical rule 29](#_Toc47214514)

[f. Is pound of material sold bell shaped? 30](#_Toc47214515)

[g. Skweness and Kurtosis 30](#_Toc47214516)

[9. Comparison of the distribution of the pounds sold data with daily visit data. 30](#_Toc47214517)

[10. Demographics 31](#_Toc47214518)

[a. All Traffic Sources 32](#_Toc47214519)

[b. Top Ten Referring Sites 32](#_Toc47214520)

[c. Top Ten Search Engine Sources of Visits 33](#_Toc47214521)

[d. Top Ten Geographic Sources by Sub Continent Region 34](#_Toc47214522)

[e. Top Ten Browsers Used 35](#_Toc47214523)

[f. Top Ten Operating Systems Used 35](#_Toc47214524)

library(readxl)  
library(ggplot2)  
library(dplyr)  
library(moments)

setwd("C:/Users/manish.grewal/emdp/R/QualityAlloys")

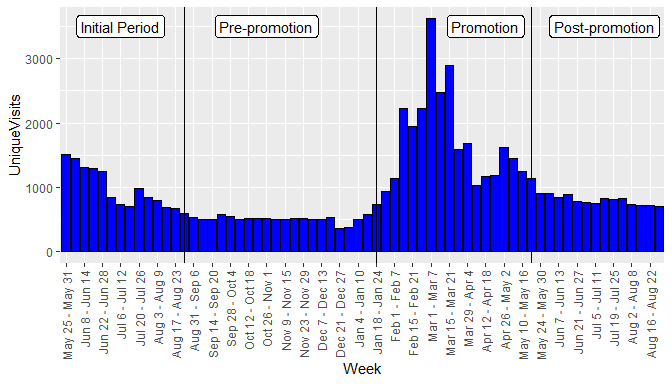
# Read the main worksheets from xls workbook   
daily\_visits <- read\_excel("CU46-XLS-ENG.xls", sheet = "Daily Visits", skip = 4)  
weekly\_visits <- read\_excel("CU46-XLS-ENG.xls", sheet = "Weekly Visits", skip = 4)  
financials <- read\_excel("CU46-XLS-ENG.xls", sheet = "Financials", skip = 4)  
lbs\_sold <- read\_excel("CU46-XLS-ENG.xls", sheet = "Lbs. Sold", skip = 4)  
  
# and remove non alphabet characters from column headers (names)  
  
names(weekly\_visits) <- gsub("[^a-zA-Z]+", "", names(weekly\_visits))  
names(financials) <- gsub("[^a-zA-Z]+", "", names(financials))  
names(lbs\_sold) <- gsub("[^a-zA-Z]+", "", names(lbs\_sold))  
  
# gather columns into weekly\_data  
weekly\_data <- data.frame(  
 Week = weekly\_visits$Week,   
 Visits = weekly\_visits$Visits,  
 UniqueVisits = weekly\_visits$UniqueVisits,   
 Revenue = financials$Revenue,   
 Profit = financials$Profit,   
 LbsSold = financials$LbsSold)  
  
# convert Week to ordered factored   
weekly\_data$Week <- factor(weekly\_data$Week, levels = weekly\_data$Week)  
  
idx\_init <- which(levels(weekly\_data$Week) == "Aug 24 - Aug 30")  
idx\_pre <- which(levels(weekly\_data$Week) == "Jan 18 - Jan 24")  
idx\_prom <- which(levels(weekly\_data$Week) == "May 17 - May 23")  
idx\_post <- nrow(weekly\_data)  
weekly\_data$Period <- c(rep("Initial Period", idx\_init),  
 rep("Pre-promotion", idx\_pre - idx\_init),  
 rep("Promotion", idx\_prom - idx\_pre),  
 rep("Post-promotion", idx\_post - idx\_prom))  
  
  
weekly\_data$Period <- factor(weekly\_data$Period, ordered = TRUE, levels = c("Initial Period", "Pre-promotion", "Promotion", "Post-promotion"))

## 1. How many people visit the website, and how do they come to website

p <- ggplot(weekly\_data, aes(x = Week)) +  
 scale\_x\_discrete(breaks = weekly\_visits$Week[seq(1, nrow(weekly\_visits), 2)]) +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))

### Unique visits over time

p + geom\_bar(aes(y = UniqueVisits), stat = "identity", fill = "blue", color = "black") +  
 geom\_vline(xintercept = c(idx\_init, idx\_pre, idx\_prom)) +  
 geom\_label(aes(x = 7, y = 3500, label = "Initial Period")) +  
 geom\_label(aes(x = 23, y = 3500, label = "Pre-promotion")) +  
 geom\_label(aes(x = 47, y = 3500, label = "Promotion")) +  
 geom\_label(aes(x = 60, y = 3500, label = "Post-promotion"))



### Summary of unique visits by period

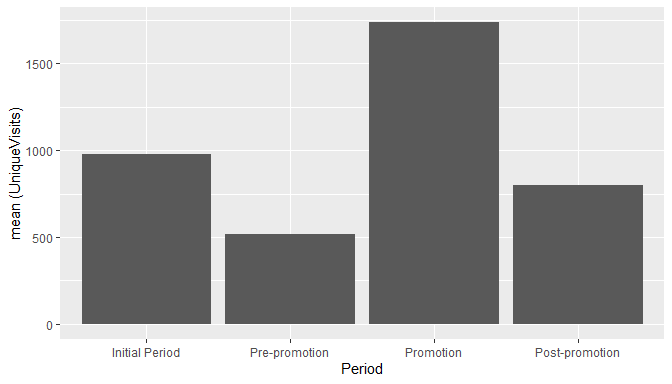
summ\_uv\_by\_period <- select(weekly\_data, Period, UniqueVisits) %>%  
 group\_by(Period) %>%  
 summarise(mean = mean(UniqueVisits), median = median(UniqueVisits), sd = sd(UniqueVisits), minimum = min(UniqueVisits), maximum = max(UniqueVisits))  
  
knitr::kable(summ\_uv\_by\_period, caption = "UNIQUE VISTS BY PERIOD")

UNIQUE VISTS BY PERIOD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Period | mean | median | sd | minimum | maximum |
| Initial Period | 975.9286 | 845.5 | 319.59721 | 594 | 1509 |
| Pre-promotion | 516.8095 | 510.0 | 70.93914 | 366 | 734 |
| Promotion | 1738.8235 | 1585.0 | 743.02534 | 930 | 3617 |
| Post-promotion | 800.7857 | 800.0 | 72.35615 | 709 | 912 |

### Mean unique visits by period

ggplot(summ\_uv\_by\_period, aes(x = Period, y = mean)) +  
 geom\_col() +  
 ylab("mean (UniqueVisits)")



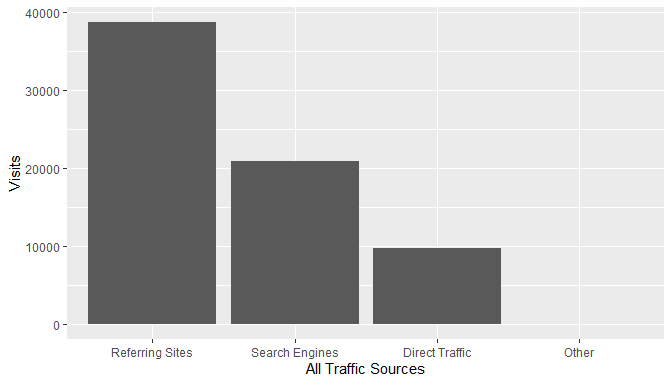
### How do people come to the website:

* Most of the people come through referring sites, search engines and direct traffic in that order.
* Google and sedoparking.com are the predominant sources of referral traffic.
* Google and yahoo are the top referring search engines.
* Bulk of the traffic is from the American continents.
* Bulk of the users use Internet Explore or Firefox on Windows.

all\_traffic <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B7:C11")  
all\_traffic$Visits <- lapply(all\_traffic$Visits, function(x) {as.numeric(gsub(",", "", x))})  
all\_traffic$Visits <- as.numeric(c(all\_traffic$Visits))  
names(all\_traffic)[1] <- 'header'  
  
top\_ten\_ref <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B14:C24")  
top\_ten\_ref$Visits <- lapply(top\_ten\_ref$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_ref$Visits <- as.numeric(c(top\_ten\_ref$Visits))  
  
top\_ten\_eng <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B27:C37")  
top\_ten\_eng$Visits <- lapply(top\_ten\_eng$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_eng$Visits <- as.numeric(c(top\_ten\_eng$Visits))  
  
top\_ten\_geo <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B40:C50")  
top\_ten\_geo$Visits <- lapply(top\_ten\_geo$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_geo$Visits <- as.numeric(c(top\_ten\_geo$Visits))  
  
top\_ten\_bro <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B54:C64")  
top\_ten\_bro$Visits <- lapply(top\_ten\_bro$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_bro$Visits <- as.numeric(c(top\_ten\_bro$Visits))  
  
top\_ten\_os <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B68:C78")  
top\_ten\_os$Visits <- lapply(top\_ten\_os$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_os$Visits <- as.numeric(c(top\_ten\_os$Visits))

#### a. All Traffic Sources

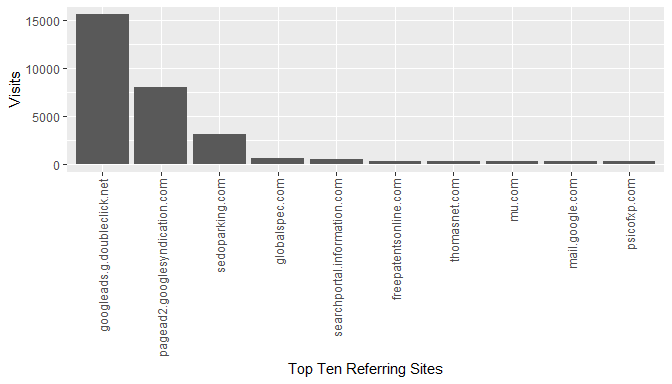
ggplot(all\_traffic, aes(x = reorder(header, -Visits), y = Visits)) +  
 geom\_col() +  
 xlab("All Traffic Sources")



There is significant Direct Traffic, which could be a result of the website link socialized in the promotion. Traffic from search engines could be increased with the help of SEO.

#### b. Top Ten Referring Sites

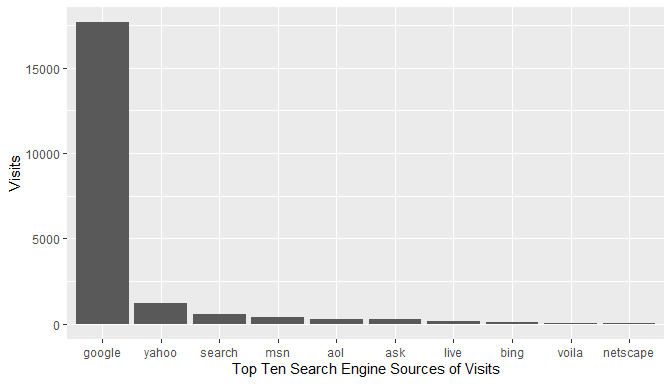
ggplot(top\_ten\_ref, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 xlab("Top Ten Referring Sites")



The paid listings on globalspec are generating more referrals than thomasnet, however, both of them together are not generating much referrals compared to google ads. The traffic from sedoparking.com may need to be investigated more as it could indicate users mistyping the QA website address and landing on a parked domain with a similar name.

#### c. Top Ten Search Engine Sources of Visits

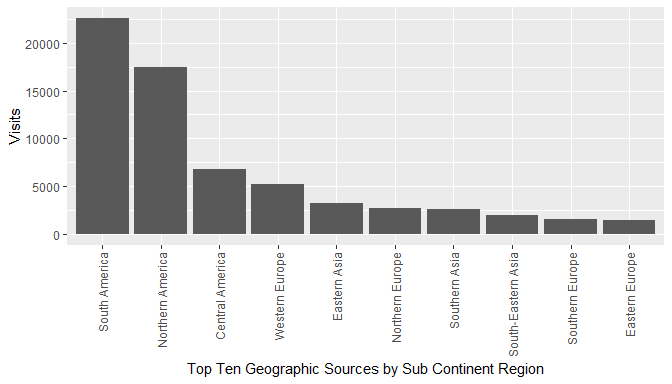
ggplot(top\_ten\_eng, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 xlab("Top Ten Search Engine Sources of Visits")



As expected, google is the top search engine source of visits. SEO should be targeted at google search engine.

#### d. Top Ten Geographic Sources by Sub Continent Region

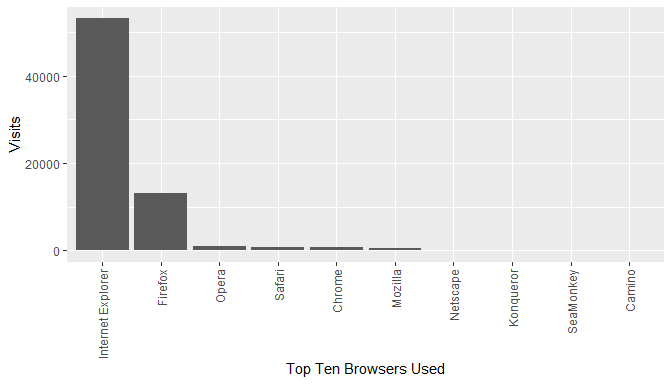
ggplot(top\_ten\_geo, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 xlab("Top Ten Geographic Sources by Sub Continent Region")



As the traffic is coming from various geographies, it may be a good idea to support multiple languages on the website (internationalization).

#### e. Top Ten Browsers Used

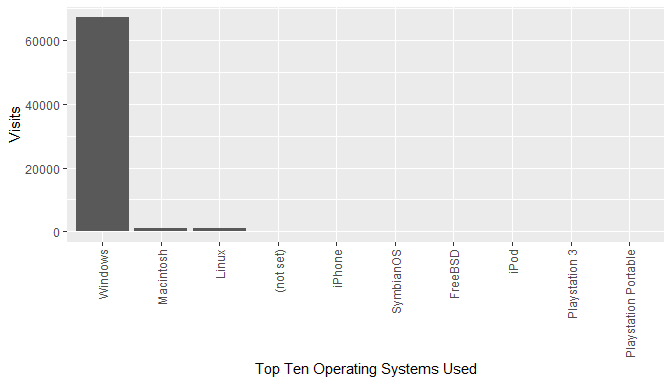
ggplot(top\_ten\_bro, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 xlab("Top Ten Browsers Used")



Website development / testing should focus on Internet Explorer and Firefox compatibility / support.

#### f. Top Ten Operating Systems Used

ggplot(top\_ten\_os, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 xlab("Top Ten Operating Systems Used")



Website development / testing should focus on Windows platform, followed by Mac and Linux.

## 2. Is the website generating interest, and does this interest yield actual sales?

The website generated interest in the Promotion period which can be seen above by the increased number of Unique Visits over the promotion period. However, the interest did not yield actual sales as we see a decreasing Lbs. Sold over the four periods.

### Summary of Lbs. Sold by Period:

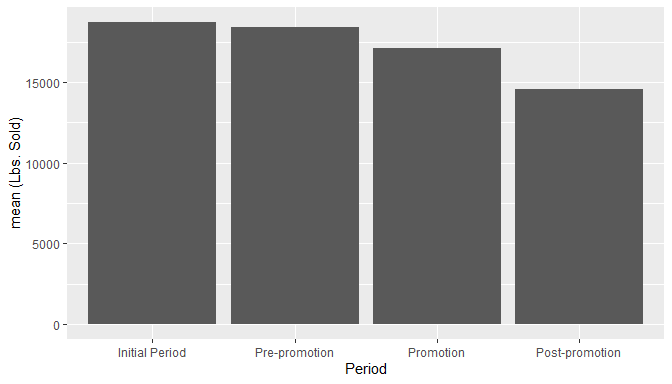
summ\_ls\_by\_period <- select(weekly\_data, Period, LbsSold) %>%  
 group\_by(Period) %>%  
 summarise(mean = mean(LbsSold), median = median(LbsSold), sd = sd(LbsSold), minimum = min(LbsSold), maximum = max(LbsSold))  
  
knitr::kable(summ\_ls\_by\_period, caption = "LBS. SOLD BY PERIOD")

LBS. SOLD BY PERIOD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Period | mean | median | sd | minimum | maximum |
| Initial Period | 18736.73 | 17269.70 | 5427.393 | 8633.059 | 28052.92 |
| Pre-promotion | 18440.77 | 17215.12 | 5965.630 | 8992.422 | 31968.98 |
| Promotion | 17112.92 | 17299.12 | 6519.071 | 7814.050 | 31496.26 |
| Post-promotion | 14577.79 | 13646.89 | 5941.556 | 3825.748 | 23761.61 |

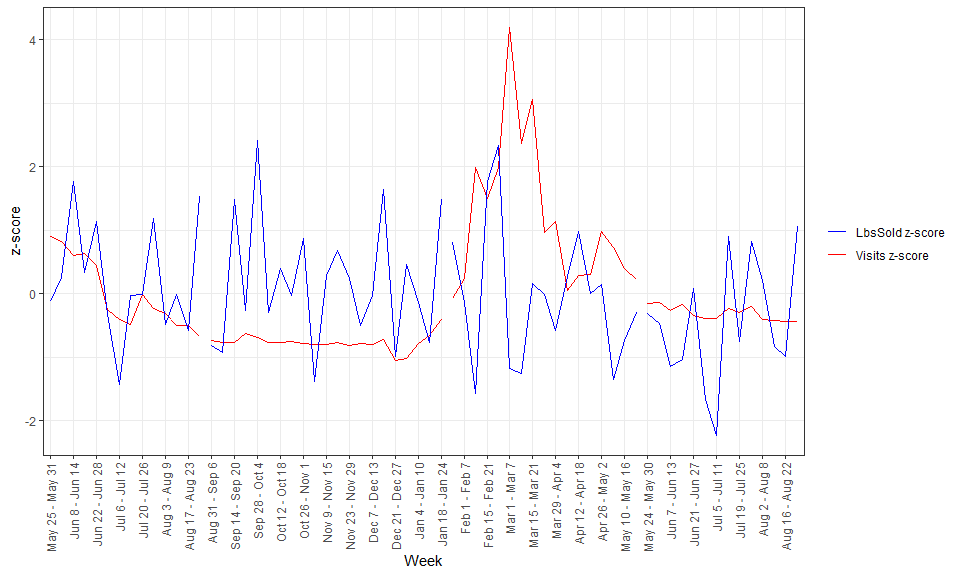
### Mean Lbs. Sold by period

ggplot(summ\_ls\_by\_period, aes(x = Period, y = mean)) +  
 geom\_col() +  
 ylab("mean (Lbs. Sold)")



### Comparison of Unique Visits and Lbs. Sold after Z-score standardization:

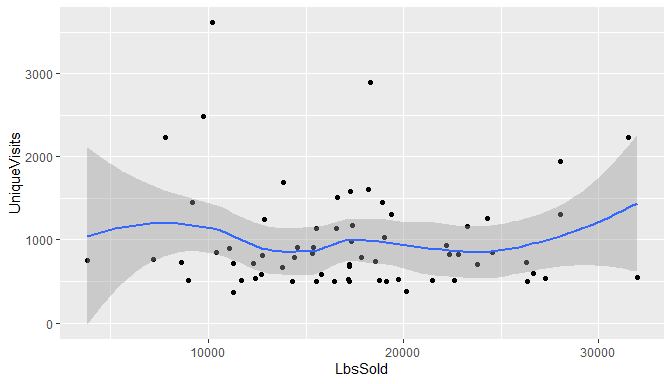
weekly\_data$Visits\_z = scale(weekly\_data$Visits)  
weekly\_data$LbsSold\_z = scale(weekly\_data$LbsSold)  
  
my\_colors <- c("Visits z-score" = "Red", "LbsSold z-score" = "blue", "Visits\_smooth" = "blue", "LbsSold\_smooth" = "red")  
  
ggplot(weekly\_data, aes(Week)) +  
 geom\_line(aes(y = Visits\_z, group = Period, color="Visits z-score")) +  
 #geom\_smooth(aes(y = Visits\_z, group = Period, color="Visits\_smooth"), method = "loess") +  
 geom\_line(aes(y = LbsSold\_z, group = Period, color = "LbsSold z-score")) +  
 #geom\_smooth(aes(y = LbsSold\_z, group = Period, color = "LbsSold\_smooth", method = "loess")) +  
 scale\_color\_manual(name = "", values = my\_colors) + # legend  
 scale\_x\_discrete(breaks = weekly\_visits$Week[seq(1, nrow(weekly\_visits), 2)]) +  
 theme\_bw() +  
 ylab("z-score") +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))



#geom\_col(aes(y = LbsSold, position = "dodge"), stat = "identity", color = "red")

### Scatterplot of Lbs. Sold and Unique Vists

ggplot(weekly\_data, aes(x = LbsSold, y = UniqueVisits)) +  
 geom\_point() +   
 geom\_smooth()



### Coefficient of correlation of Lbs. Sold and Unique Visits

cor\_coef <- cor(weekly\_data$UniqueVisits, weekly\_data$LbsSold, method = "pearson")  
cor\_coef

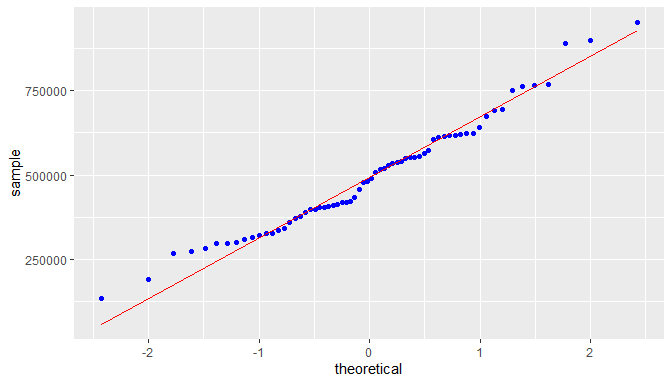
## [1] -0.05622245

Small negative value shows that Lbs. Sold does not increase or decrease with Unique Visits.

## 3. Check whether the data variables (if applicable) follows the normality assumption.

### Normality assumption for Revenue

ggplot(weekly\_data, aes(sample = Revenue)) +  
 stat\_qq(color = "blue") +  
 stat\_qq\_line(col = "red")



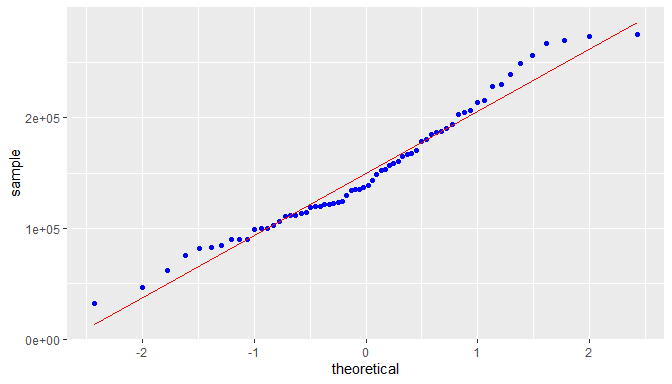
Revenue is approximately normal but slightly right skewed. Skewness:

skew <- 3 \* (mean(weekly\_data$Revenue) - median(weekly\_data$Revenue)) / sd(weekly\_data$Revenue)  
skew

## [1] 0.1850349

### Normality assumption for Profit

ggplot(weekly\_data, aes(sample = Profit)) +  
 stat\_qq(color = "blue") +  
 stat\_qq\_line(col = "red")



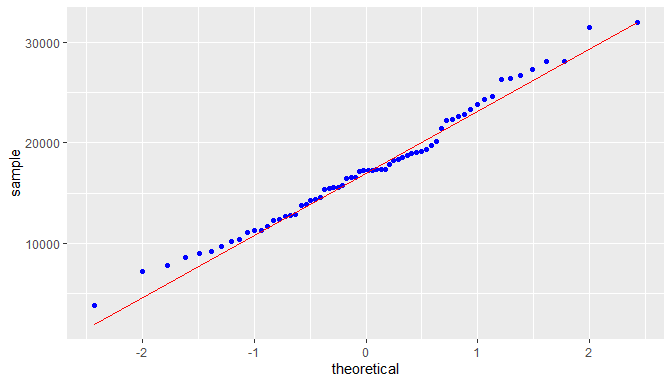
Profit is approximately normal but significantly right skewed. Skewness:

skew <- 3 \* (mean(weekly\_data$Profit) - median(weekly\_data$Profit)) / sd(weekly\_data$Profit)  
skew

## [1] 0.6745705

### Normality assumption for Lbs. Sold

ggplot(weekly\_data, aes(sample = LbsSold)) +  
 stat\_qq(color = "blue") +  
 stat\_qq\_line(col = "red")



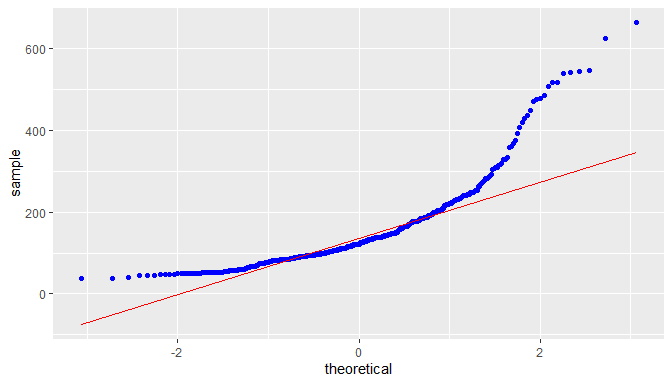
Lbs. Sold is normal with a very low right skewness. Skewness:

skew <- 3 \* (mean(weekly\_data$LbsSold) - median(weekly\_data$LbsSold)) / sd(weekly\_data$LbsSold)  
skew

## [1] 0.0624725

### Normality assumption for Visits

ggplot(daily\_visits, aes(sample = Visits)) +  
 stat\_qq(color = "blue") +  
 stat\_qq\_line(col = "red")



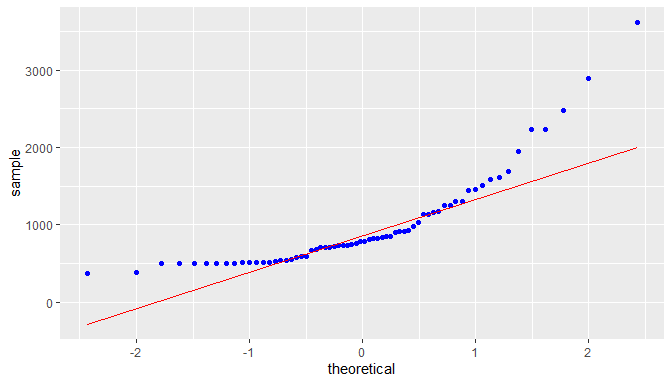
Visits shows a high degree of right skewness. Skewness:

skew <- 3 \* (mean(weekly\_data$Visits) - median(weekly\_data$Visits)) / sd(weekly\_data$Visits)  
skew

## [1] 0.987206

### Normality assumption for UniqueVisits

ggplot(weekly\_data, aes(sample = UniqueVisits)) +  
 stat\_qq(color = "blue") +  
 stat\_qq\_line(col = "red")



UniqueVisits shows a high degree of right skewness. Skewness:

skew <- 3 \* (mean(weekly\_data$UniqueVisits) - median(weekly\_data$UniqueVisits)) / sd(weekly\_data$UniqueVisits)  
skew

## [1] 0.9622686

# PART 1 - EXECUTIVE SUMMARY

## Evaluation of the goals of the company website

### (a) drive new sales - NOT MET

The data analysis presented in PART B of this document clearly shows that revenue and quantity of material sold have been reducing over the period of analysis. Thus, we can say that the website has not been able to drive new sales even during the promotion period.

### (b) make product and contact information available - MAY BE MET

This goal can be considered as met if we assume that the contact information has been presented well on the website, .

### (c) give or add legitimacy to its brand. - MAY BE MET

Just the existence of an official company website does add legitimacy to the branch. However, a survey given to the right audience, customers, potential customers and general public, may be able to shed more light on this. The available data is not sufficient to answer whether this goal has been met.

## Recommendations from demographics:

##### **Note:** Refer to Q10 below for more details on demographics.

There is significant Direct Traffic, which could be a result of the website link socialized in the promotion. Traffic from search engines could be increased with the help of SEO (Search engine optimization).

The paid listings on globalspec.com are generating more referrals than thomasnet.com, however, both of them together are not generating much referrals compared to google ads.

The traffic from sedoparking.com may need to be investigated more as it could indicate users mistyping the QA website address and landing on a parked domain with a similar name.

As expected, google is the top search engine source of visits. SEO should be targeted at google search engine.

As the traffic is coming from various geographies, it may be a good idea to support multiple languages on the website (internationalization).

Website development / testing should focus on Internet Explorer and Firefox compatibility / support on Windows platform, followed by Mac and Linux as the majority of users are using this stack.

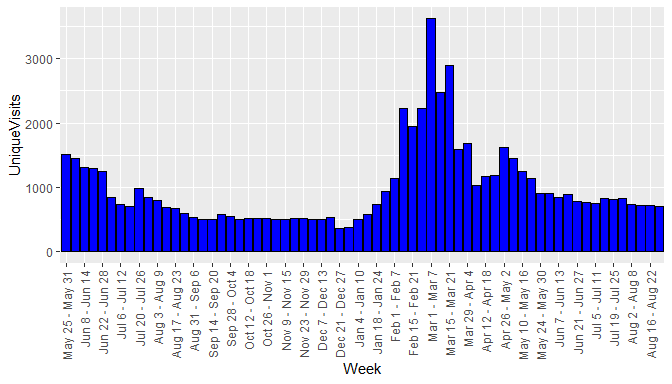
# PART 2 - DATA ANALYSIS

## 1. Column Charts

p <- ggplot(weekly\_data, aes(x = Week)) +  
 scale\_x\_discrete(breaks = weekly\_visits$Week[seq(1, nrow(weekly\_visits), 2)]) +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1))

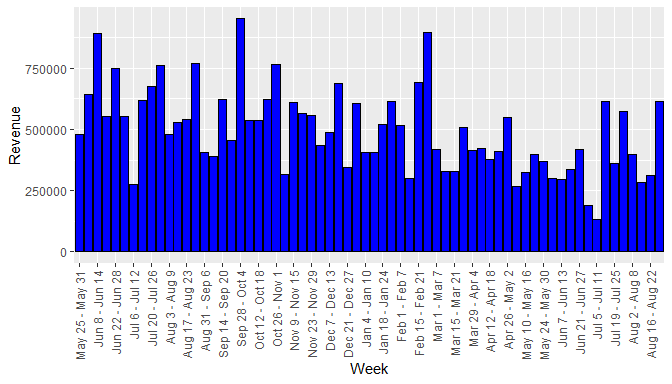
### a. Unique visits over time

p + geom\_bar(aes(y = UniqueVisits), stat = "identity", fill = "blue", color = "black")



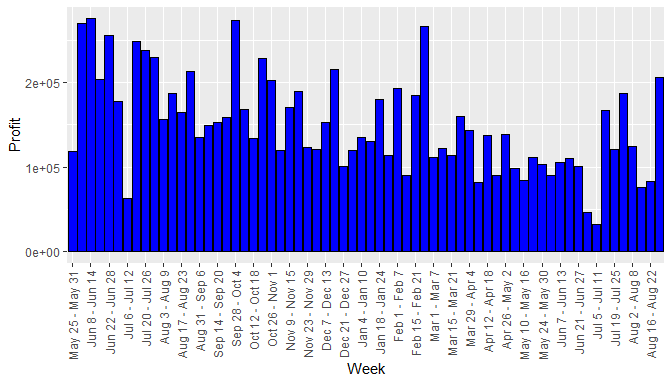
### b. Revenue over time,

p + geom\_bar(aes(y = Revenue), stat = "identity", fill = "blue", color = "black")



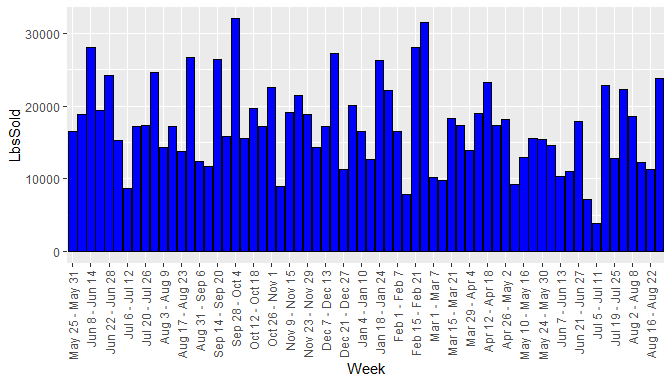
### c. Profit over time

p + geom\_bar(aes(y = Profit), stat = "identity", fill = "blue", color = "black")



### d. Pounds sold over time.

p + geom\_bar(aes(y = LbsSold), stat = "identity", fill = "blue", color = "black")



## 2. Summary by period

Below are the summaries of the variables for each period.

by\_period <- group\_by(weekly\_data, Period)  
  
summ <- by\_period %>%   
 select\_if(is.numeric) %>%   
 summarise(across(everything(), list(mean, median, sd, min, max)   
 #, .names = c("{col}.{fn}")  
 ))  
  
summ\_init <- as.data.frame(  
 matrix(as.numeric(summ[1, 2:26]), nrow = 5),   
 row.names = c('mean', 'median', 'std. dev.', 'minimum', 'maximum'))  
names(summ\_init) <- names(by\_period)[2:6]  
  
summ\_pre <- as.data.frame(  
 matrix(as.numeric(summ[2, 2:26]), nrow = 5),  
 row.names = c('mean', 'median', 'std. dev.', 'minimum', 'maximum'))  
names(summ\_pre) <- names(by\_period)[2:6]  
  
summ\_prom <- as.data.frame(  
 matrix(as.numeric(summ[3, 2:26]), nrow = 5),  
 row.names = c('mean', 'median', 'std. dev.', 'minimum', 'maximum'))  
names(summ\_prom) <- names(by\_period)[2:6]  
  
summ\_post <- as.data.frame(  
 matrix(as.numeric(summ[4, 2:26]), nrow = 5),  
 row.names = c('mean', 'median', 'std. dev.', 'minimum', 'maximum'))  
names(summ\_post) <- names(by\_period)[2:6]

### a. Summary table of Initial Period

knitr::kable(summ\_init, caption = "VISIT AND FINANCIAL SUMMARY MEASURES - INITIAL PERIOD")

VISIT AND FINANCIAL SUMMARY MEASURES - INITIAL PERIOD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Visits | UniqueVisits | Revenue | Profit | LbsSold |
| mean | 1055.2143 | 975.9286 | 608250.1 | 200233.41 | 18736.729 |
| median | 899.0000 | 845.5000 | 586169.7 | 208913.00 | 17269.696 |
| std. dev. | 355.0333 | 319.5972 | 155930.4 | 60691.55 | 5427.393 |
| minimum | 626.0000 | 594.0000 | 274567.6 | 62580.40 | 8633.059 |
| maximum | 1632.0000 | 1509.0000 | 890076.7 | 275218.10 | 28052.924 |

### b. Summary table of Pre-promotion Period

knitr::kable(summ\_pre, caption = "VISIT AND FINANCIAL SUMMARY MEASURES - PRE PROMOTION PERIOD")

VISIT AND FINANCIAL SUMMARY MEASURES - PRE PROMOTION PERIOD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Visits | UniqueVisits | Revenue | Profit | LbsSold |
| mean | 562.95238 | 516.80952 | 534313.5 | 159932.03 | 18440.774 |
| median | 558.00000 | 510.00000 | 534541.7 | 152476.40 | 17215.118 |
| std. dev. | 80.87118 | 70.93914 | 150502.8 | 42682.68 | 5965.630 |
| minimum | 383.00000 | 366.00000 | 315647.1 | 100388.40 | 8992.422 |
| maximum | 795.00000 | 734.00000 | 951216.2 | 273174.70 | 31968.976 |

### c. Summary table of Promotion Period

knitr::kable(summ\_prom, caption = "VISIT AND FINANCIAL SUMMARY MEASURES - PROMOTION PERIOD")

VISIT AND FINANCIAL SUMMARY MEASURES - PROMOTION PERIOD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Visits | UniqueVisits | Revenue | Profit | LbsSold |
| mean | 1814.3529 | 1738.8235 | 456398.8 | 131929.90 | 17112.923 |
| median | 1663.0000 | 1585.0000 | 413937.1 | 114328.40 | 17299.115 |
| std. dev. | 758.1017 | 743.0253 | 161741.1 | 47776.85 | 6519.071 |
| minimum | 1000.0000 | 930.0000 | 268159.5 | 81841.40 | 7814.050 |
| maximum | 3726.0000 | 3617.0000 | 897163.7 | 266476.70 | 31496.257 |

### d. Summary table of Post-promotion Period

knitr::kable(summ\_post, caption = "VISIT AND FINANCIAL SUMMARY MEASURES - POST PROMOTION PERIOD")

VISIT AND FINANCIAL SUMMARY MEASURES - POST PROMOTION PERIOD

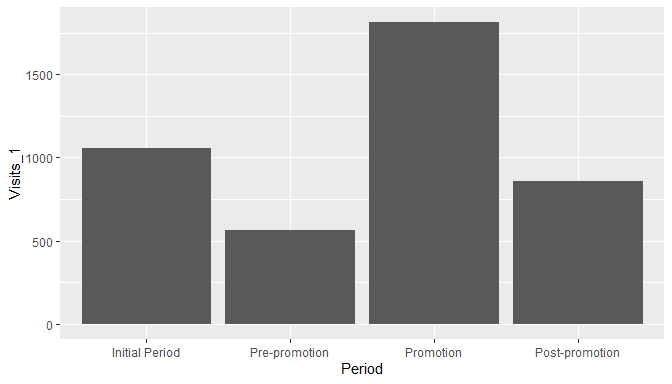
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Visits | UniqueVisits | Revenue | Profit | LbsSold |
| mean | 856.57143 | 800.78571 | 371728.0 | 111045.82 | 14577.794 |
| median | 847.50000 | 800.00000 | 348397.1 | 104530.45 | 13646.895 |
| std. dev. | 70.88855 | 72.35615 | 145728.3 | 49065.27 | 5941.556 |
| minimum | 772.00000 | 709.00000 | 133966.9 | 32825.30 | 3825.748 |
| maximum | 963.00000 | 912.00000 | 615950.2 | 206441.20 | 23761.614 |

## 3. Column chart of the means over the four periods

p <- ggplot(summ, aes(x = Period))

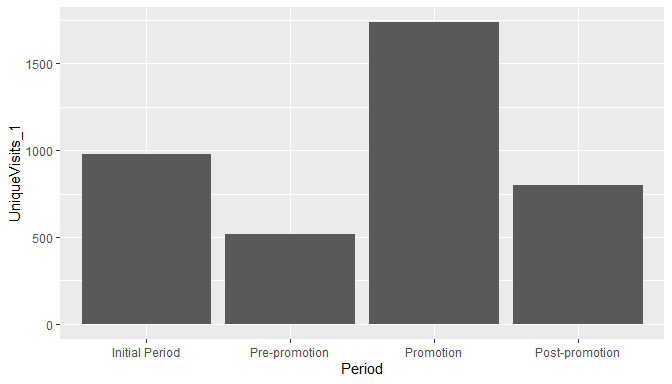
### a. Mean Visits by Period

p + geom\_col(aes(y = Visits\_1))



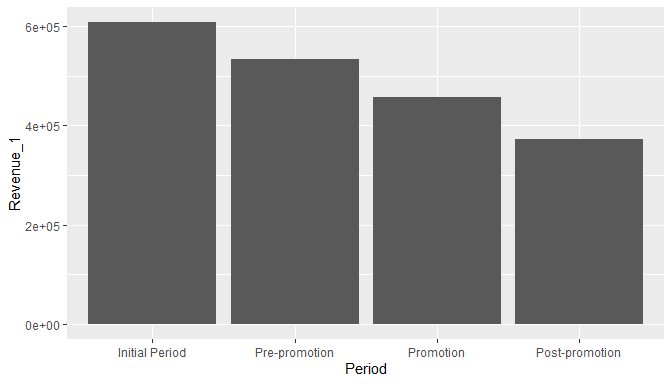
### b. Mean Unique Visits by Period

# r, fig.show="hold", fig.width=3  
p + geom\_col(aes(y = UniqueVisits\_1))



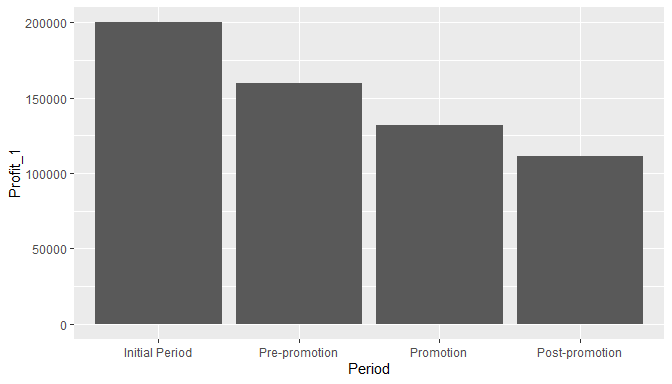
### c. Mean Revenue by Period

p + geom\_col(aes(y = Revenue\_1))



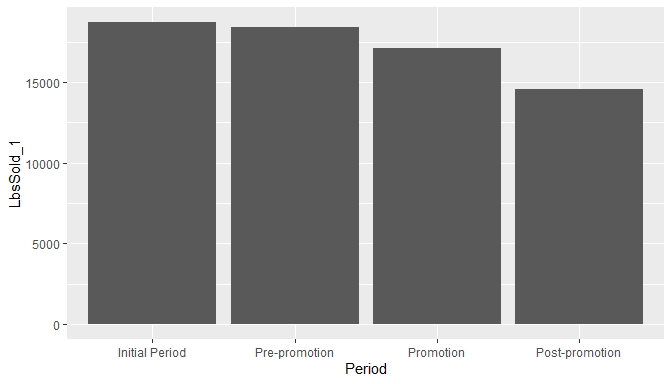
### d. Mean Profit by Period

p + geom\_col(aes(y = Profit\_1))



### e. Mean Lbs. Sold by Period

p + geom\_col(aes(y = LbsSold\_1))

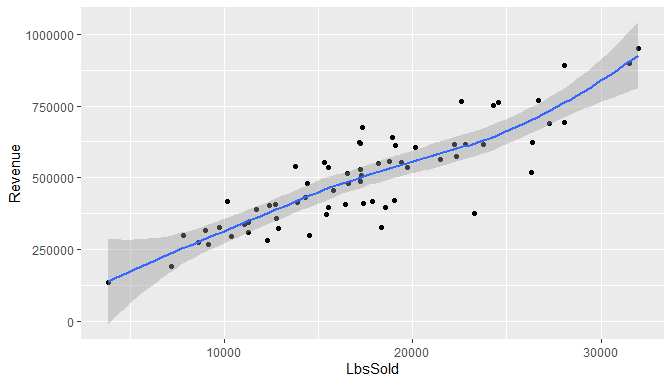


## 4. Summary of findings thus far

Average visits and average unique visits have shown a marked increase in promotion period. However, profit, revenue and lbs. sold show a steady declining trend over the four periods. This suggests that the website promotion did not have a positive effect on financials.

## 5. Scatter plot of Revenue and Lbs. Sold

ggplot(weekly\_data, aes(LbsSold, Revenue)) +  
 geom\_point() +  
 geom\_smooth()



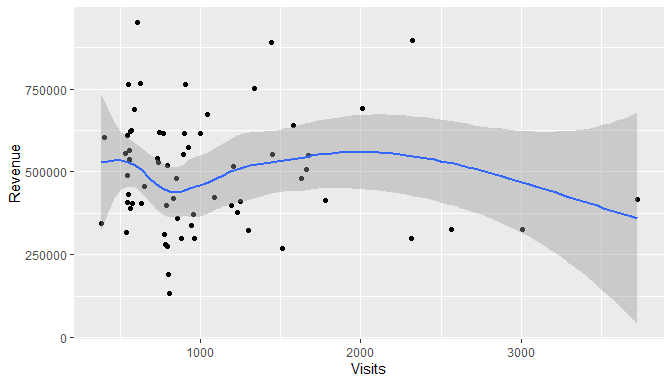
### b. Coefficient of Correlation

cor\_coef <- cor(weekly\_data$LbsSold, weekly\_data$Revenue, method = "pearson")  
cor\_coef

## [1] 0.8689297

## 6. Scatter plot of Visits and Revenue

ggplot(weekly\_data, aes(Visits, Revenue)) +  
 geom\_point() +  
 geom\_smooth()



#### b. Coefficient of Correlation

cor\_coef <- cor(weekly\_data$Visits, weekly\_data$Revenue, method = "pearson")  
cor\_coef

## [1] -0.05939183

## 7. Summary of results

The scatter plot of Revenue and Lbs. Sold shows a high degree of correlation.

The scatter plot of Revenue and Visits mirrors the earlier observations from the column chart - revenue does not increase or decrease with visits. Thus the website did not meet the goal of driving sales.

## 8. Modeling data

### a. Summary of Lbs. Sold data per week

paste("Mean:", mean(lbs\_sold$LbsSold))

## [1] "Mean: 18681.5551724138"

paste("Median:", median(lbs\_sold$LbsSold))

## [1] "Median: 17673"

paste("Std. Dev.:", sd(lbs\_sold$LbsSold))

## [1] "Std. Dev.: 6840.50793952086"

paste("Minimum:", min(lbs\_sold$LbsSold))

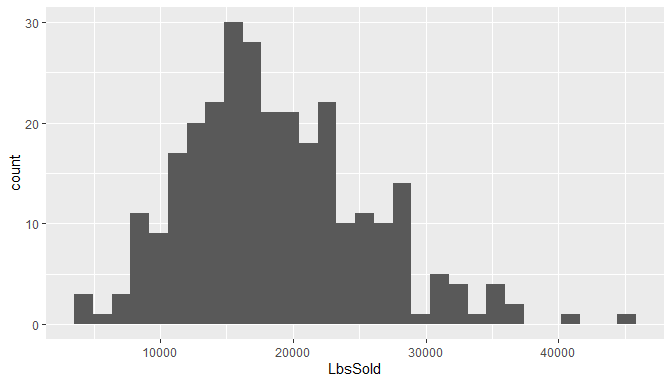
## [1] "Minimum: 3826"

paste("Maximum:", max(lbs\_sold$LbsSold))

## [1] "Maximum: 44740"

### b. Histogram of the pounds of material sold data.

ggplot(lbs\_sold, aes(LbsSold), color = "blue") +  
 geom\_histogram()



### c. Histogram description

Histogram appears bell shaped with a few outliers on the right side. Also, mean is greater than median, so the data is right skewed.

### d. Empirical rule

lbs\_sold$zscore <- scale(lbs\_sold$LbsSold)  
  
emp\_rule <- data.frame(  
 "Interval" = c("mean B1 1 std. dev.", "mean B1 2 std. dev.", "mean B1 3 std. dev. 99%"),  
 "TheoreticalPctOfData" = c(68, 95, 99)  
 )  
  
emp\_rule$TheoreticalNoOfObs <- emp\_rule$TheoreticalPctOfData \* nrow(lbs\_sold) / 100  
  
emp\_rule$ActualNoOfObs <- c(  
 nrow(filter(lbs\_sold, zscore >= -1, zscore <= 1)),  
 nrow(filter(lbs\_sold, zscore >= -2, zscore <= 2)),  
 nrow(filter(lbs\_sold, zscore >= -3, zscore <= 3)))  
knitr::kable(emp\_rule)

|  |  |  |  |
| --- | --- | --- | --- |
| Interval | TheoreticalPctOfData | TheoreticalNoOfObs | ActualNoOfObs |
| mean B1 1 std. dev. | 68 | 197.2 | 201 |
| mean B1 2 std. dev. | 95 | 275.5 | 276 |
| mean B1 3 std. dev. 99% | 99 | 287.1 | 288 |

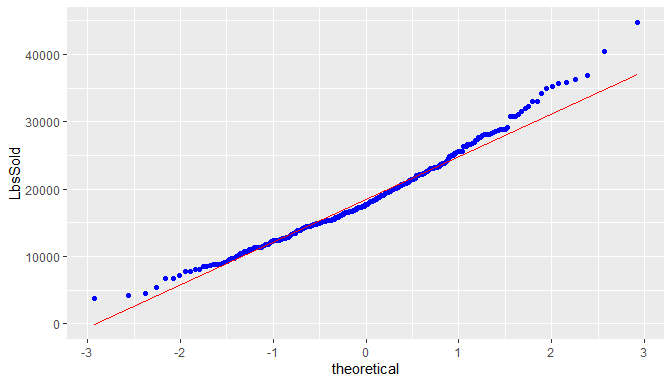
### e. Refine empirical rule

emp\_rule <- data.frame(  
 "Interval" = c(  
 "mean + 1 std. dev.",   
 "mean - 1 std. dev.",   
 "1 std. dev. to 2 std. dev.",   
 "-1 std. dev. to -2 std. dev.",   
 "2 std. dev. to 3 std. dev.",   
 "-2 std. dev. to -3 std. dev.")  
 )  
  
n <- 100000 # sufficently high number to get theoretical pct  
set.seed(123)  
rn <- data.frame(rn = rnorm(n))  
  
emp\_rule$rn <- c(  
 nrow(filter(rn, rn >= 0, rn <= 1)),  
 nrow(filter(rn, rn >= -1, rn <= 0)),  
 nrow(filter(rn, rn >= 1, rn <= 2)),  
 nrow(filter(rn, rn >= -2, rn <= -1)),  
 nrow(filter(rn, rn >= 2, rn <= 3)),  
 nrow(filter(rn, rn >= -3, rn <= -2)))  
  
emp\_rule$TheoreticalPctOfData <- emp\_rule$rn \* 100 / n  
emp\_rule$TheoreticalNoOfObs <- round(emp\_rule$TheoreticalPctOfData \* nrow(lbs\_sold) / 100)  
  
emp\_rule$ActualNoOfObs <- c(  
 nrow(filter(lbs\_sold, zscore >= 0, zscore <= 1)),  
 nrow(filter(lbs\_sold, zscore >= -1, zscore <= 0)),  
 nrow(filter(lbs\_sold, zscore >= 1, zscore <= 2)),  
 nrow(filter(lbs\_sold, zscore >= -2, zscore <= -1)),  
 nrow(filter(lbs\_sold, zscore >= 2, zscore <= 3)),  
 nrow(filter(lbs\_sold, zscore >= -3, zscore <= -2)))  
  
emp\_rule$rn <- NULL  
knitr::kable(emp\_rule)

|  |  |  |  |
| --- | --- | --- | --- |
| Interval | TheoreticalPctOfData | TheoreticalNoOfObs | ActualNoOfObs |
| mean + 1 std. dev. | 34.101 | 99 | 84 |
| mean - 1 std. dev. | 34.065 | 99 | 117 |
| 1 std. dev. to 2 std. dev. | 13.669 | 40 | 35 |
| -1 std. dev. to -2 std. dev. | 13.631 | 40 | 40 |
| 2 std. dev. to 3 std. dev. | 2.139 | 6 | 9 |
| -2 std. dev. to -3 std. dev. | 2.131 | 6 | 3 |

### f. Is pound of material sold bell shaped?

ggplot(lbs\_sold, aes(sample = LbsSold)) +  
 stat\_qq(color = "blue") +  
 stat\_qq\_line(col = "red") +  
 ylab("LbsSold")



From the normal probability plot, we can see the LbsSold data approximates the bell shape but not exactly.

### g. Skweness and Kurtosis

skew <- 3 \* (mean(lbs\_sold$LbsSold) - median(lbs\_sold$LbsSold)) / sd(lbs\_sold$LbsSold)  
paste("Skewness using formula 3\*(mean-median)/sd:", skew)

## [1] "Skewness using formula 3\*(mean-median)/sd: 0.442315913378401"

paste("Skewness using library:", skewness(lbs\_sold$LbsSold))

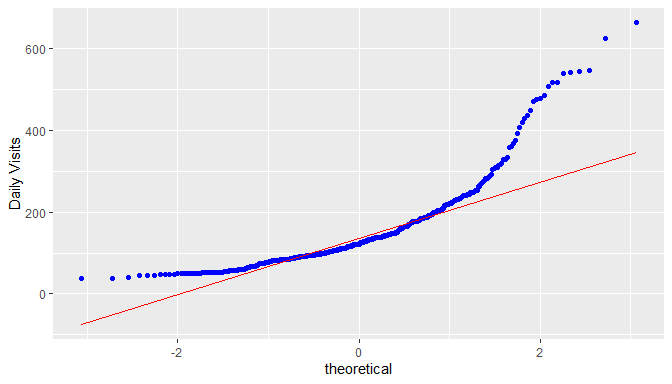
## [1] "Skewness using library: 0.629032218484559"

paste("Kurtosis:", kurtosis(lbs\_sold$LbsSold))

## [1] "Kurtosis: 3.53337138531912"

## 9. Comparison of the distribution of the pounds sold data with daily visit data.

ggplot(daily\_visits, aes(sample = Visits)) +  
 stat\_qq(color = "blue") +  
 stat\_qq\_line(col = "red") +  
 ylab("Daily Visits")



Daily visits data is less bell shaped compared to the Lbs sold data.

The empirical rule shows the actual number of observations matches the theoretical number of observations within 2nd and 3rd sd, but there is huge difference in the first sd - 314 (68%) theoretical vs 392 (85%) actual).

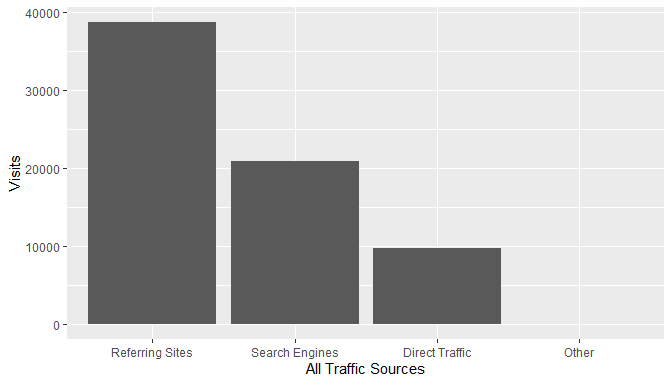
The mean is greater than median, so the data is right skewed. Skewness and Kurtosis are also high compared to Lbs Sold data.

## 10. Demographics

all\_traffic <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B7:C11")  
all\_traffic$Visits <- lapply(all\_traffic$Visits, function(x) {as.numeric(gsub(",", "", x))})  
all\_traffic$Visits <- as.numeric(c(all\_traffic$Visits))  
names(all\_traffic)[1] <- 'header'  
  
top\_ten\_ref <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B14:C24")  
top\_ten\_ref$Visits <- lapply(top\_ten\_ref$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_ref$Visits <- as.numeric(c(top\_ten\_ref$Visits))  
  
top\_ten\_eng <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B27:C37")  
top\_ten\_eng$Visits <- lapply(top\_ten\_eng$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_eng$Visits <- as.numeric(c(top\_ten\_eng$Visits))  
  
top\_ten\_geo <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B40:C50")  
top\_ten\_geo$Visits <- lapply(top\_ten\_geo$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_geo$Visits <- as.numeric(c(top\_ten\_geo$Visits))  
  
top\_ten\_bro <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B54:C64")  
top\_ten\_bro$Visits <- lapply(top\_ten\_bro$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_bro$Visits <- as.numeric(c(top\_ten\_bro$Visits))  
  
top\_ten\_os <- read\_excel("CU46-XLS-ENG.xls", sheet = "Demographics", range = "B68:C78")  
top\_ten\_os$Visits <- lapply(top\_ten\_os$Visits, function(x) {as.numeric(gsub(",", "", x))})  
top\_ten\_os$Visits <- as.numeric(c(top\_ten\_os$Visits))

## a. All Traffic Sources

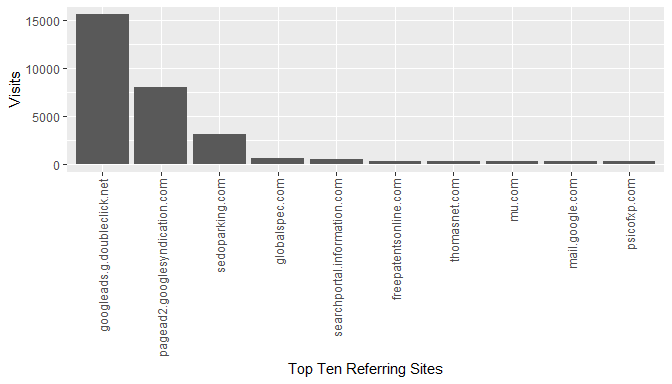
ggplot(all\_traffic, aes(x = reorder(header, -Visits), y = Visits)) +  
 geom\_col() +  
 xlab("All Traffic Sources")



There is significant Direct Traffic, which could be a result of the website link socialized in the promotion. Traffic from search engines could be increased with the help of SEO (Search engine optimization).

## b. Top Ten Referring Sites

ggplot(top\_ten\_ref, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 xlab("Top Ten Referring Sites")

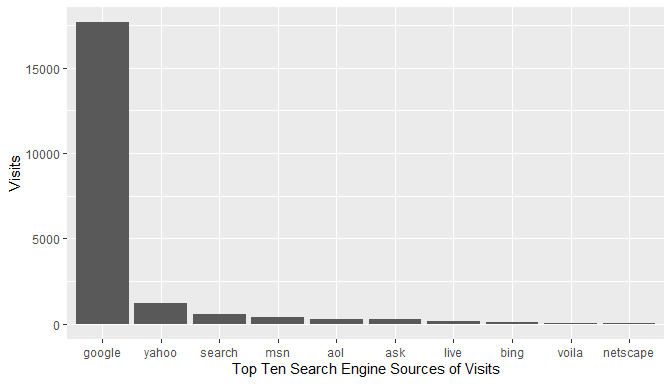


The paid listings on globalspec are generating more referrals than thomasnet, however, both of them together are not generating much referrals compared to google ads.

The traffic from sedoparking.com may need to be investigated more as it could indicate users mistyping the QA website address and landing on a parked domain with a similar name.

## c. Top Ten Search Engine Sources of Visits

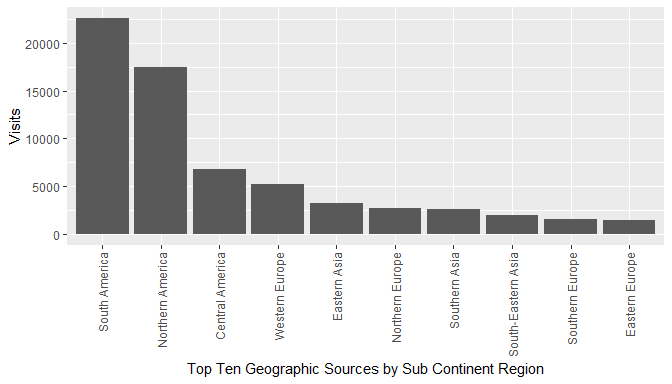
ggplot(top\_ten\_eng, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 xlab("Top Ten Search Engine Sources of Visits")



As expected, google is the top search engine source of visits. SEO should be targeted at google search engine.

## d. Top Ten Geographic Sources by Sub Continent Region

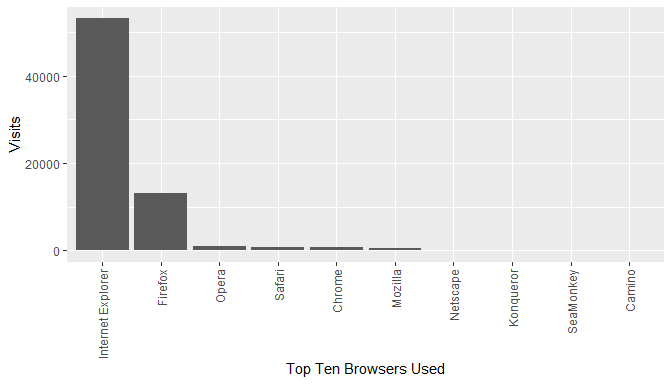
ggplot(top\_ten\_geo, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 xlab("Top Ten Geographic Sources by Sub Continent Region")



As the traffic is coming from various geographies, it may be a good idea to support multiple languages on the website (internationalization).

## e. Top Ten Browsers Used

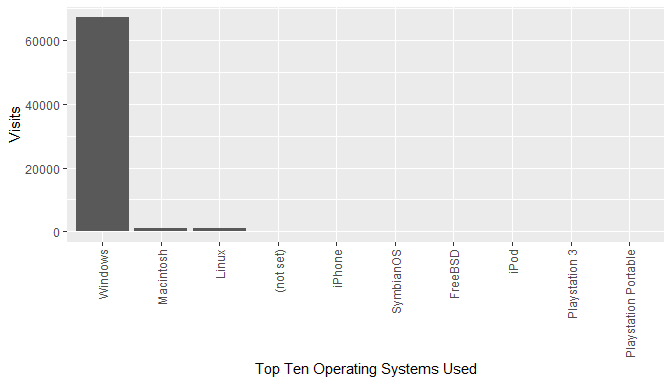
ggplot(top\_ten\_bro, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 xlab("Top Ten Browsers Used")



Website development / testing should focus on Internet Explorer and Firefox compatibility / support.

## f. Top Ten Operating Systems Used

ggplot(top\_ten\_os, aes(reorder(...1, -Visits), Visits)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 xlab("Top Ten Operating Systems Used")



Website development / testing should focus on Windows platform, followed by Mac and Linux.