

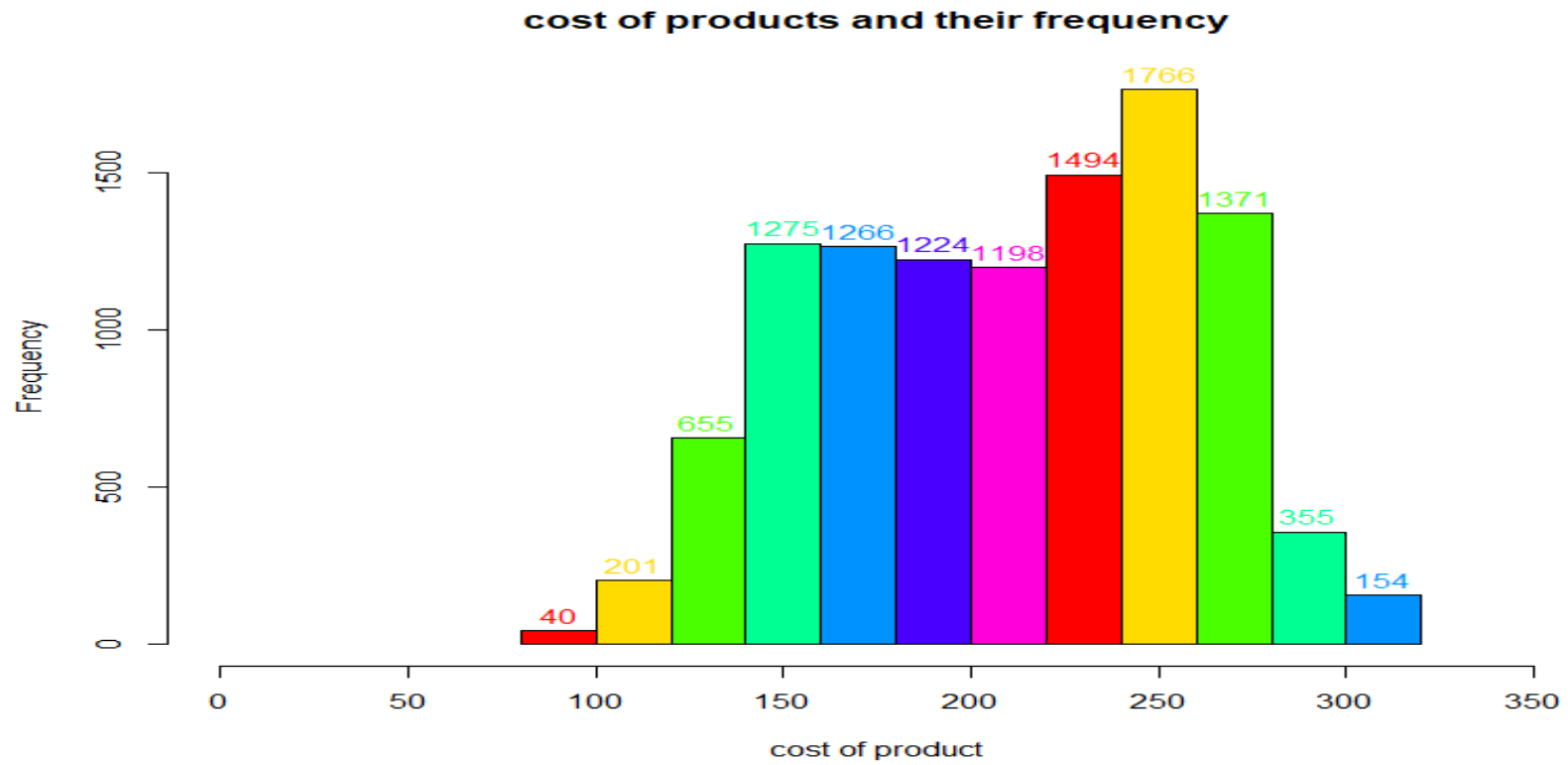
E-commerce shipping data

Rahul Yadav

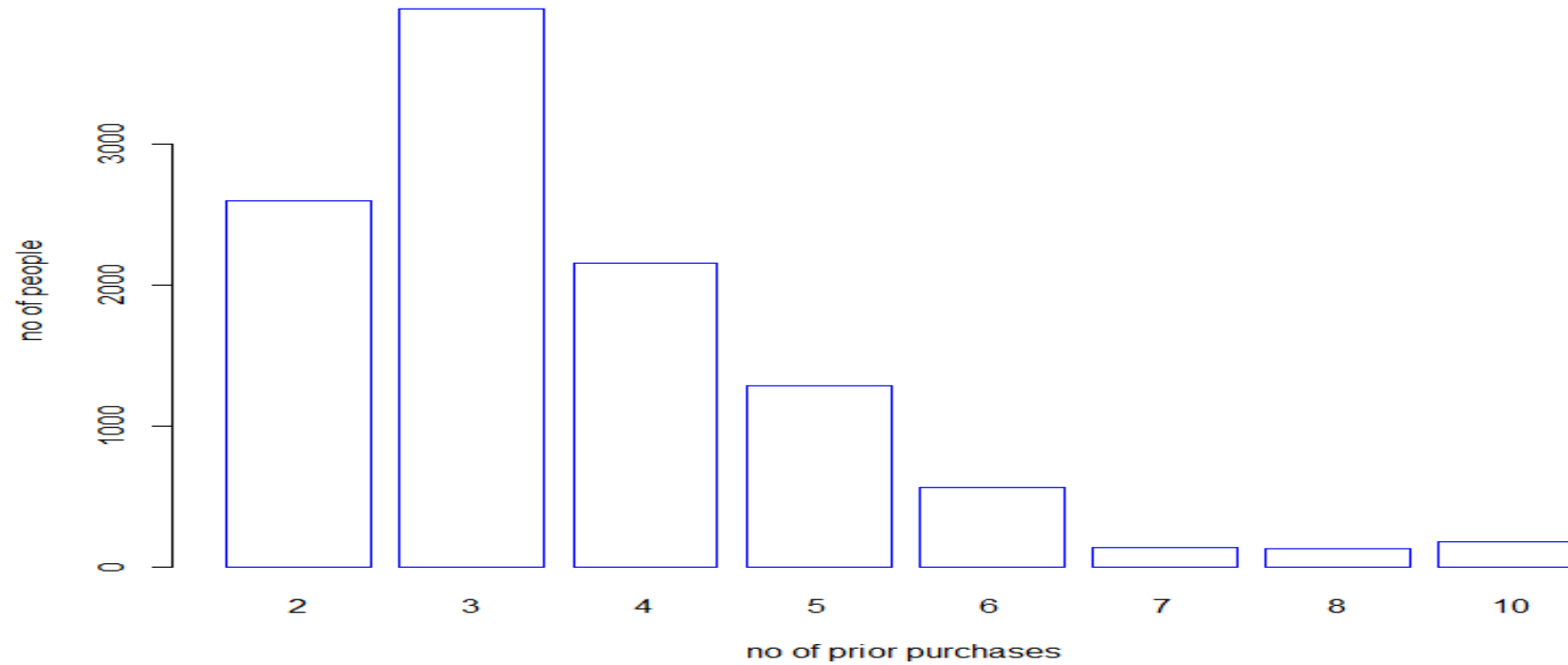
Shashwat Tiwari

Gaurav Tomar

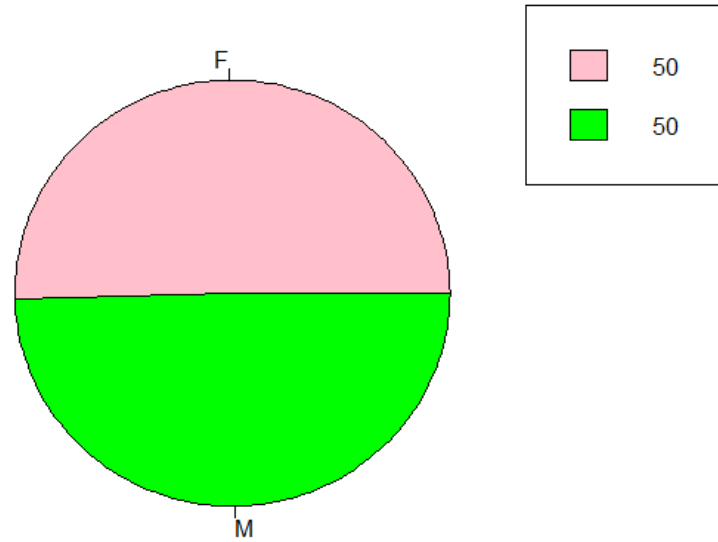
Mayank Agnihotri



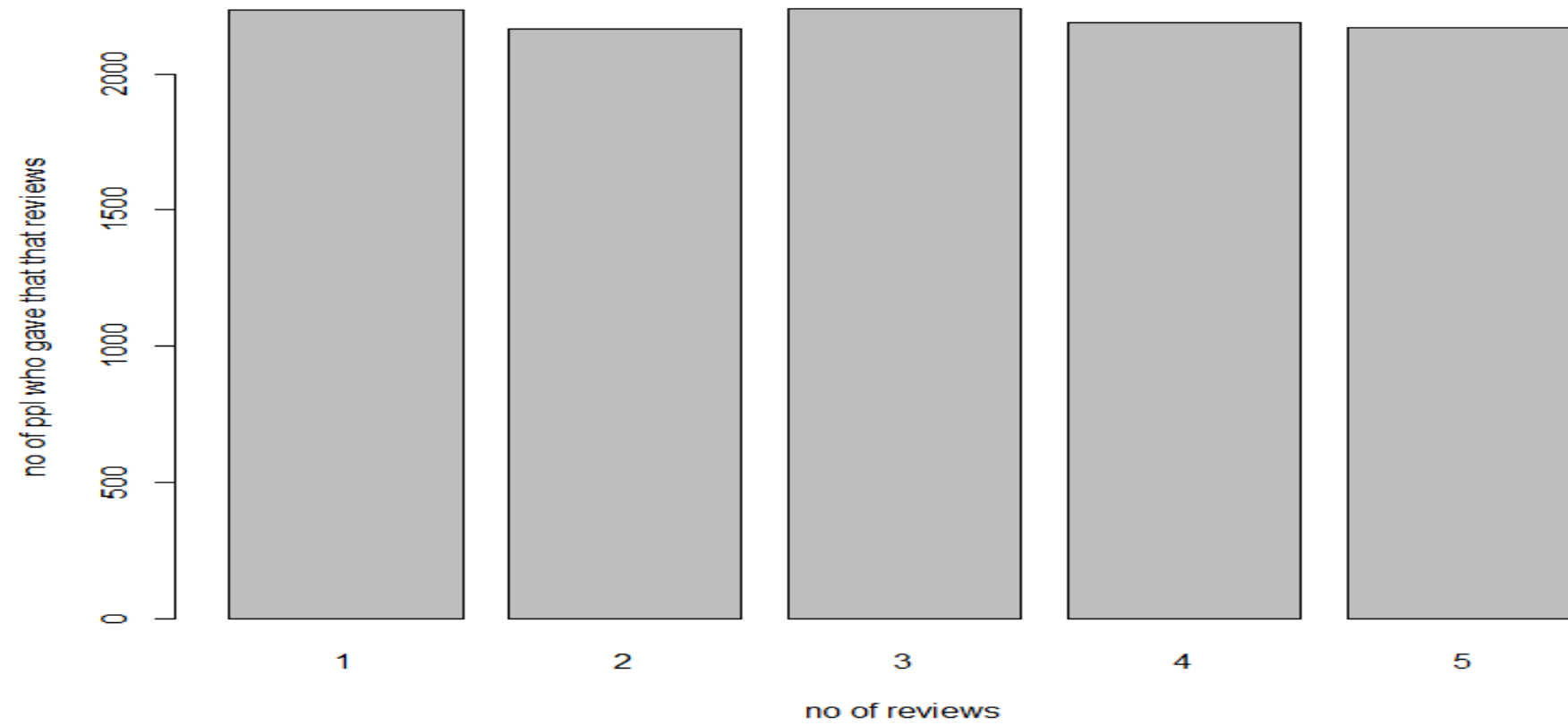
The above histogram shows the cost of product and the no of products(frequency) for each cost



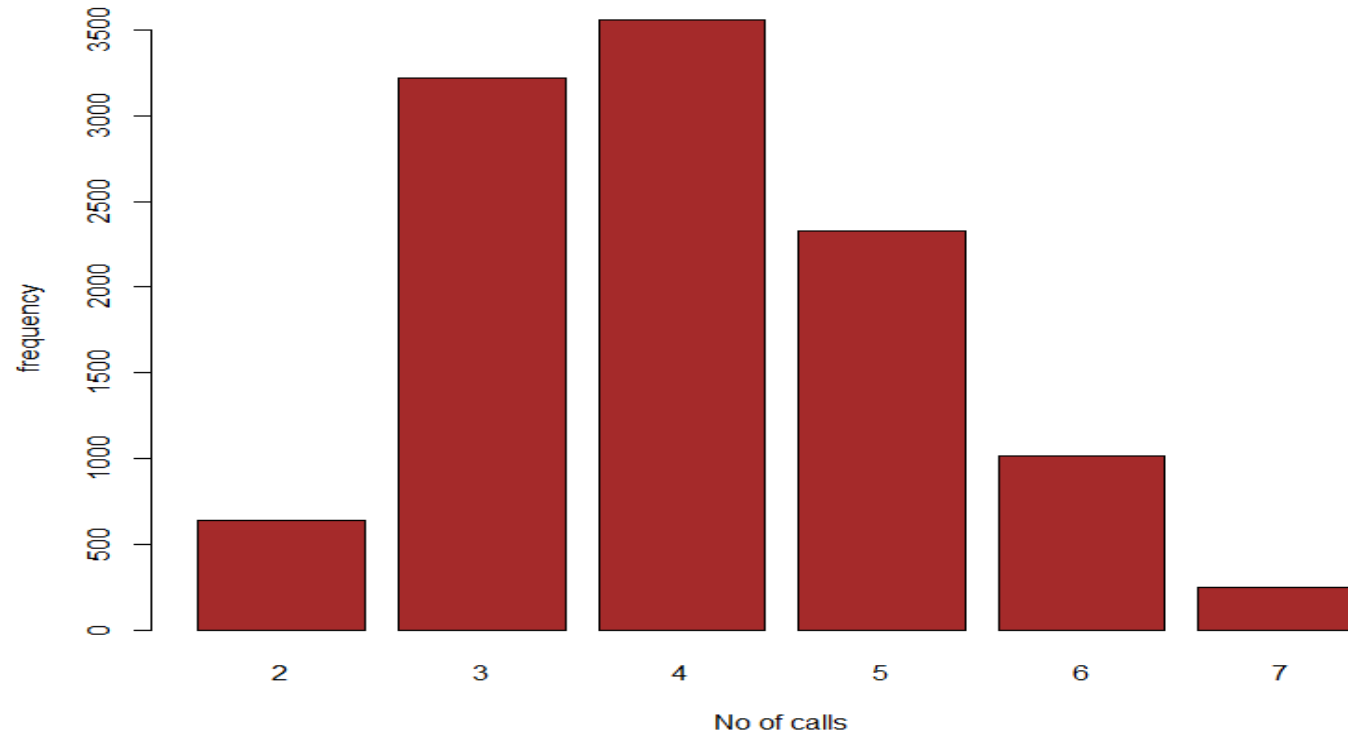
The above barplot shows the no of prior purchases and the no of ppl who bought it.



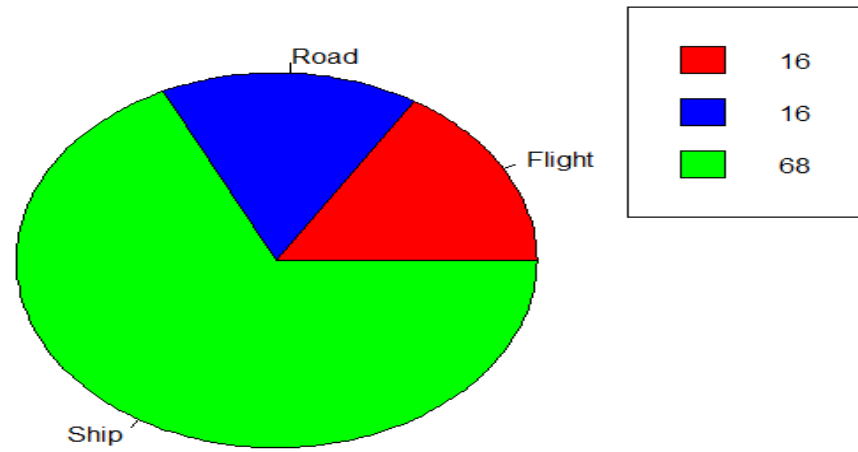
From the above pie chart we conclude that the no of females are approximately equal to the number of males who bought the products online.



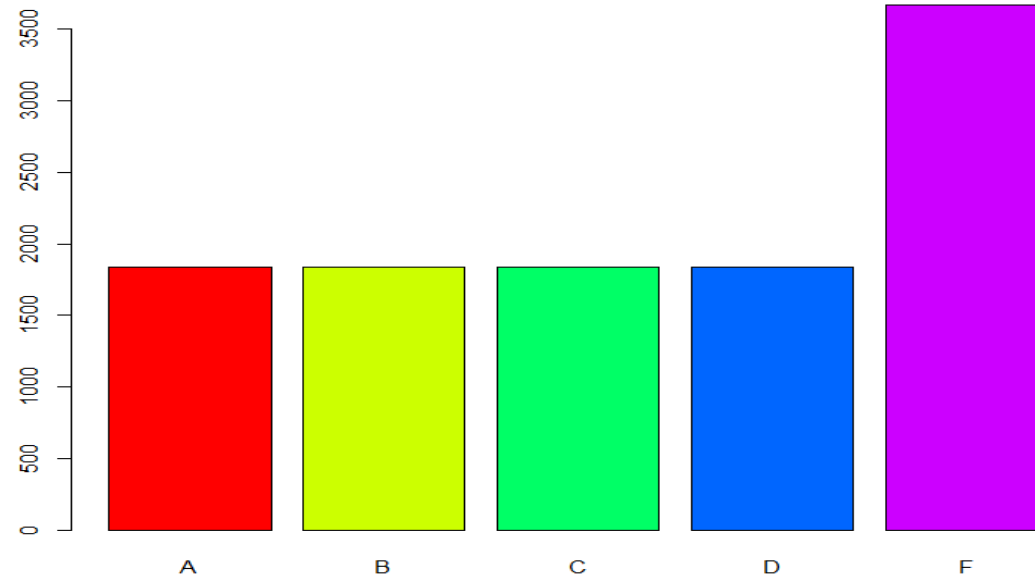
Approximately all the rating are equal



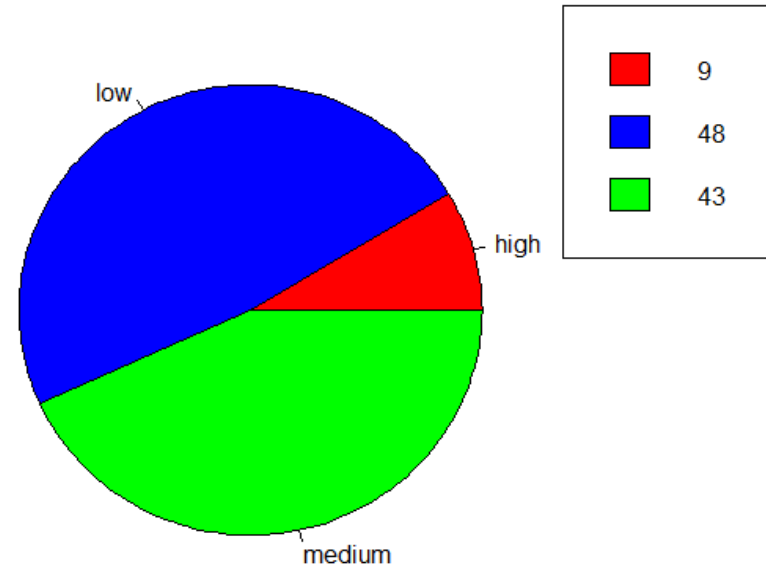
The above graph shows that most of the people made 4 calls to enquire about the product that they bought and 7 was the minimum number of calls made to enquire about the product.



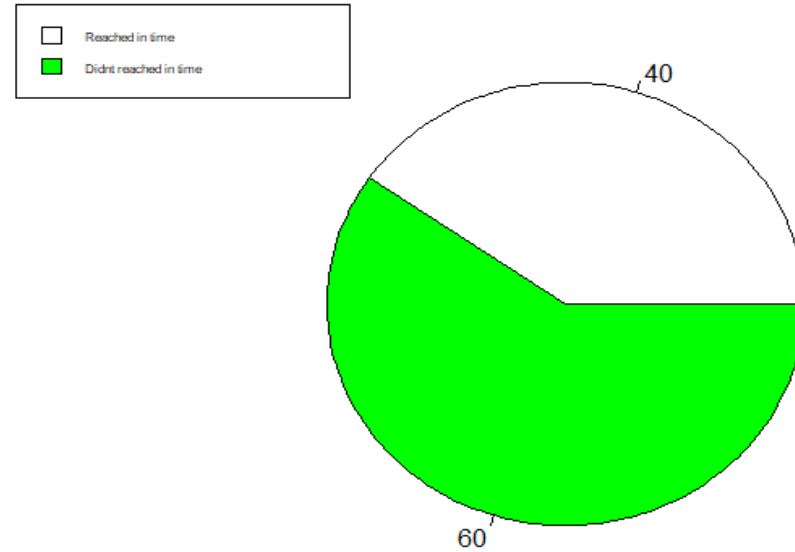
The above pie chart shows that most of the product were shipped through the ship while transports through road and flight carried approximately same no of items.



The above bar graph is plotted against the warehouse which ships the product and the amount of product shipped by each warehouse. Thus from above figure we conclude warehouse F ships most of the products while all the other warehouses ship approximately equal



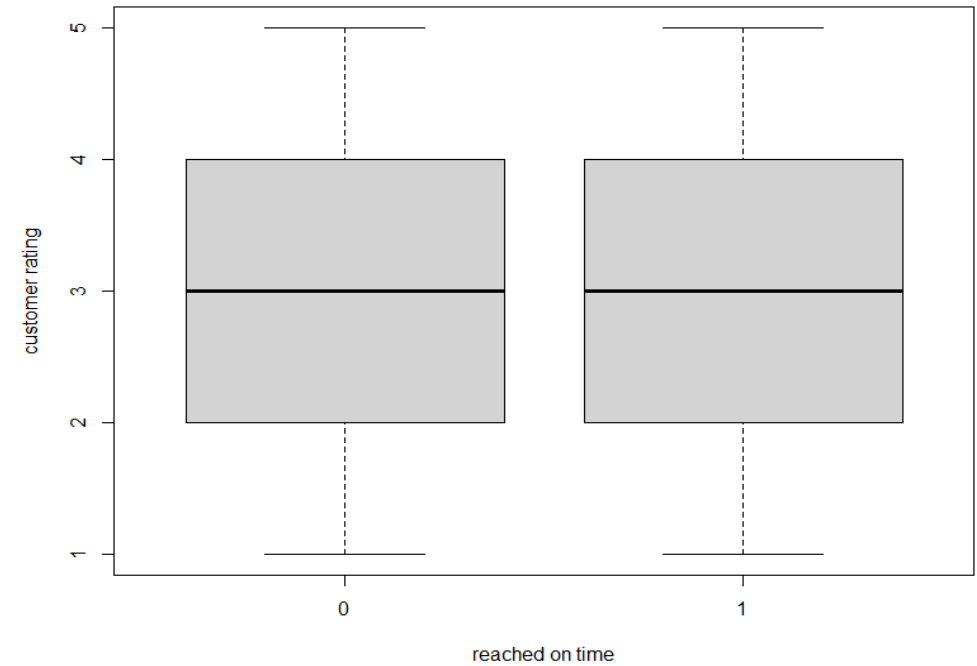
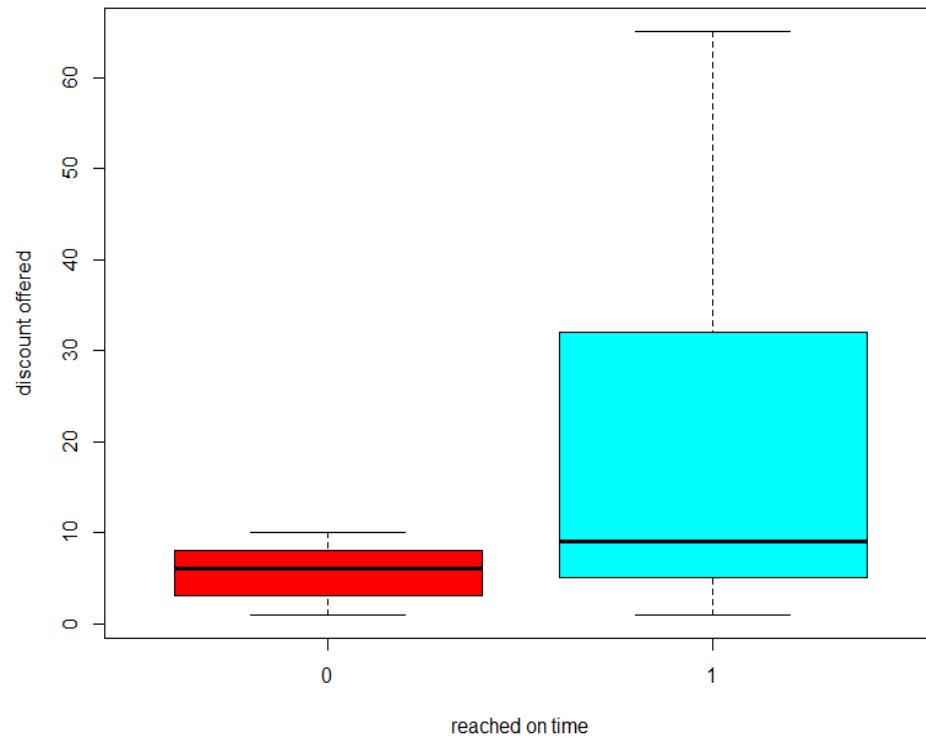
Thus above pie chart shows that max product bought was of low importance and products of high importance was bought least.



We conclude from the above chart that 40 % of the items reached on time while 60% of the items were delayed

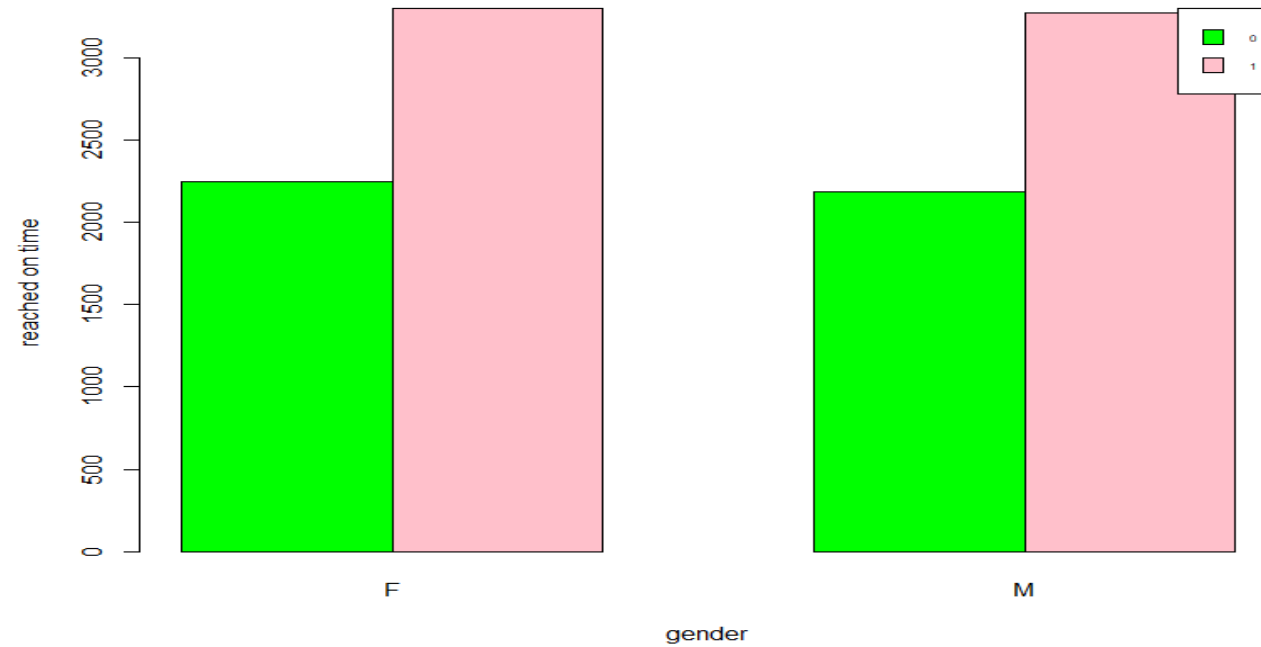
```
> cor(Train$Discount_offered,Train$Reached.on.Time_Y.N)
[1] 0.3971085
> cor(Train$weight_in_gms,Train$Reached.on.Time_Y.N)
[1] -0.2687926
> cor(Train$Discount_offered,Train$weight_in_gms)
[1] -0.3760667
> cor(Train$Customer_care_calls,Train$weight_in_gms)
[1] -0.2766152
> cor(Train$Customer_care_calls,Train$Cost_of_the_Product)
[1] 0.3231818
```

- **Discount Offered have high positive correlation with Reached on Time or Not of 40%.**
- **Weights in gram have negative correlation with Reached on Time or Not -27%.**
- **Discount Offered and weights in grams have negative correlation -38%.**
- **Customer care calls and weights in grams have negative correlation -28%.**
- **Customer care calls and cost of the product have positive correlation of 32%.**



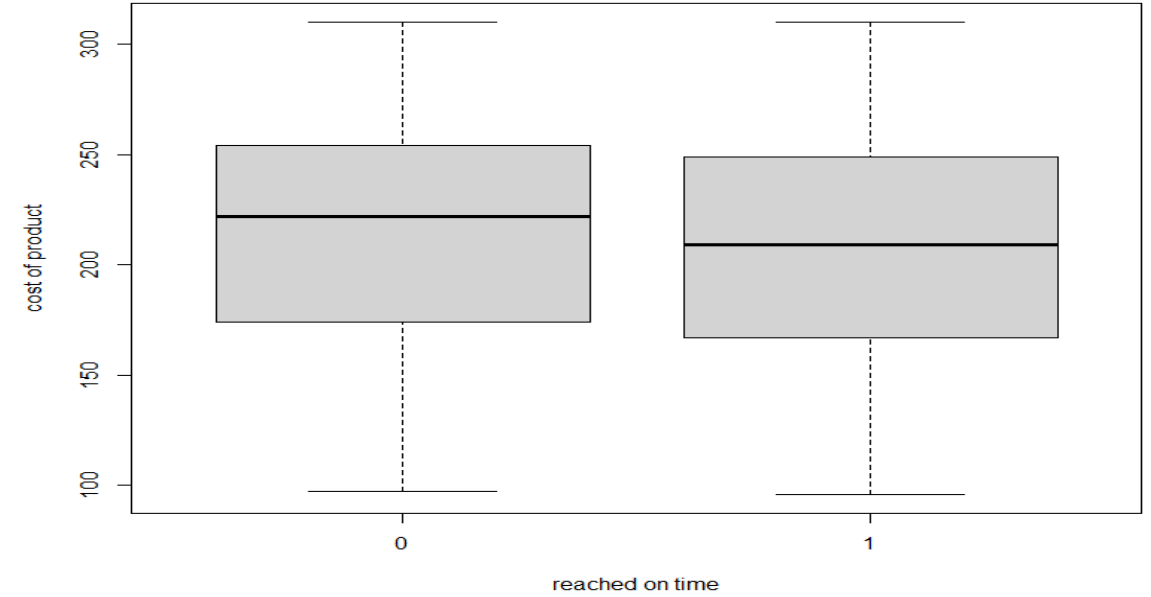
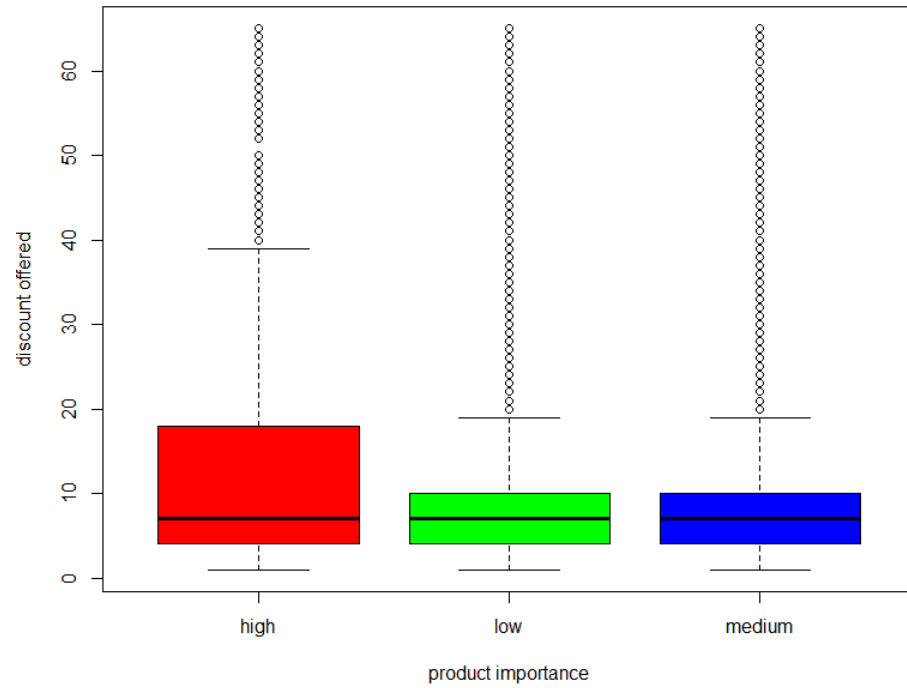
The above boxplot is plotted for product reached on time vs consumer rating. Here 0 indicates that product reached on time and 1 indicates the product didn't reach on time.

Comparing the medians for 2 boxplots, we see that medians for both of the plots are equal to 3. the lower as well as upper quartile are also equal to each other.

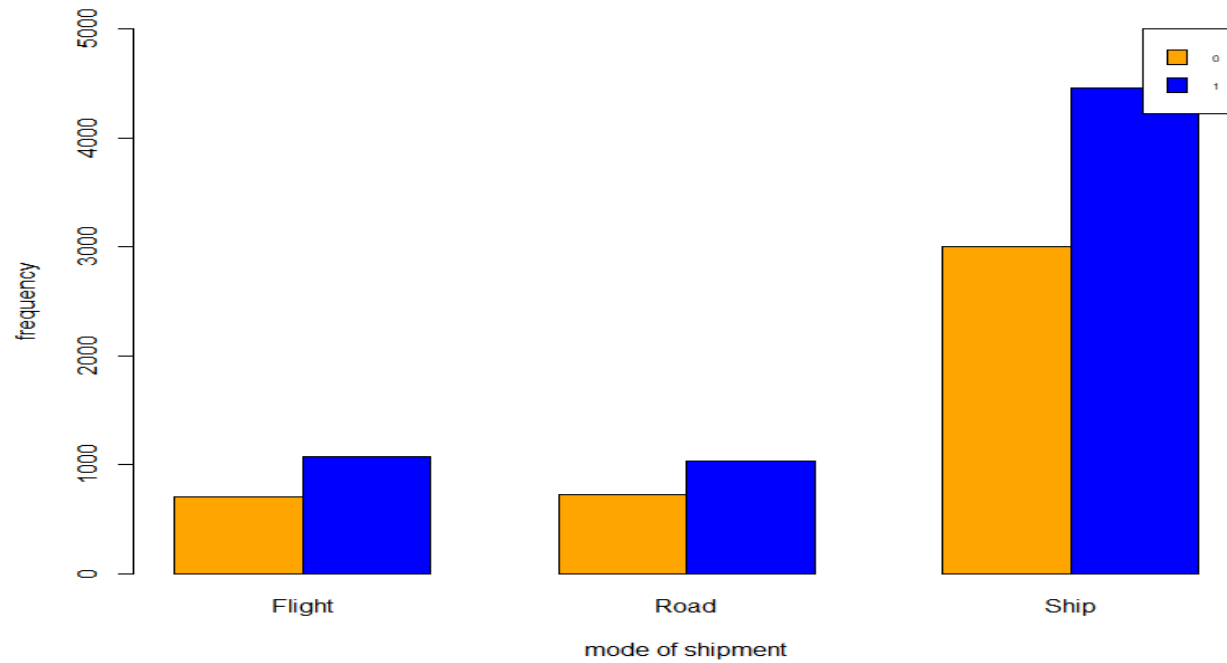


Here green represents the no of items reached on time for a particular item and pink represents the no of items that didn't reach on time for a given gender

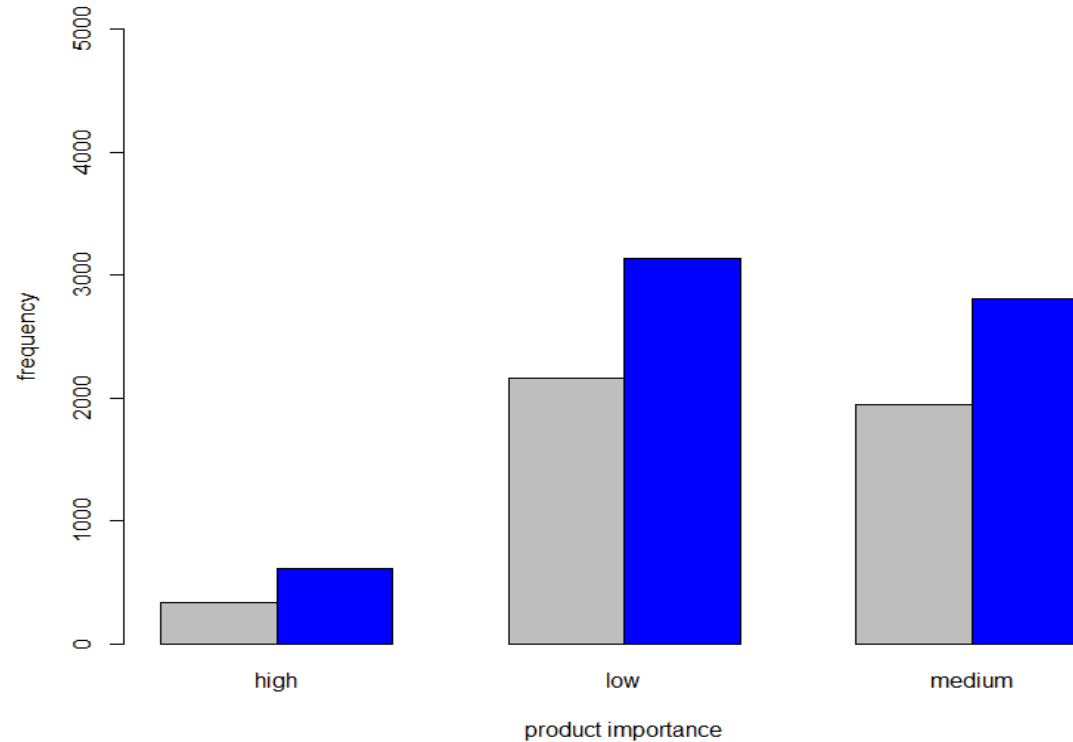
We see no of items reached on time are approximately equal for both the genders, similar case for products that didn't reach on time.



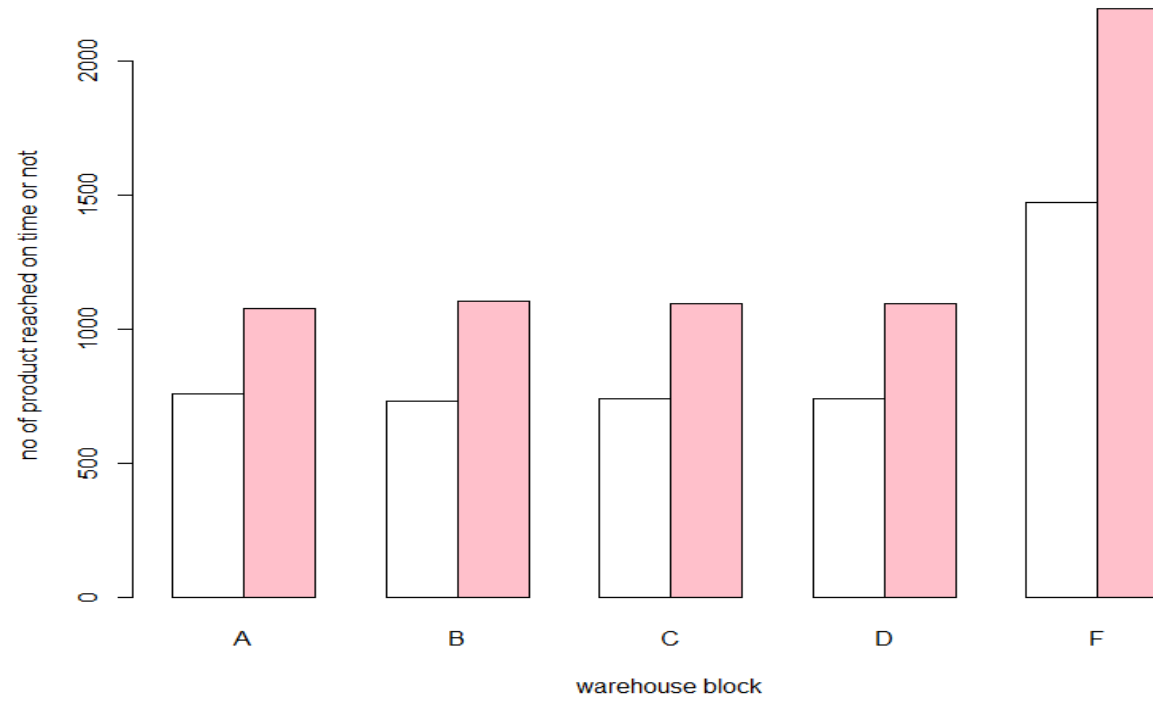
Here we see that median of product reached on time is > that median of product that didn't reach on time.



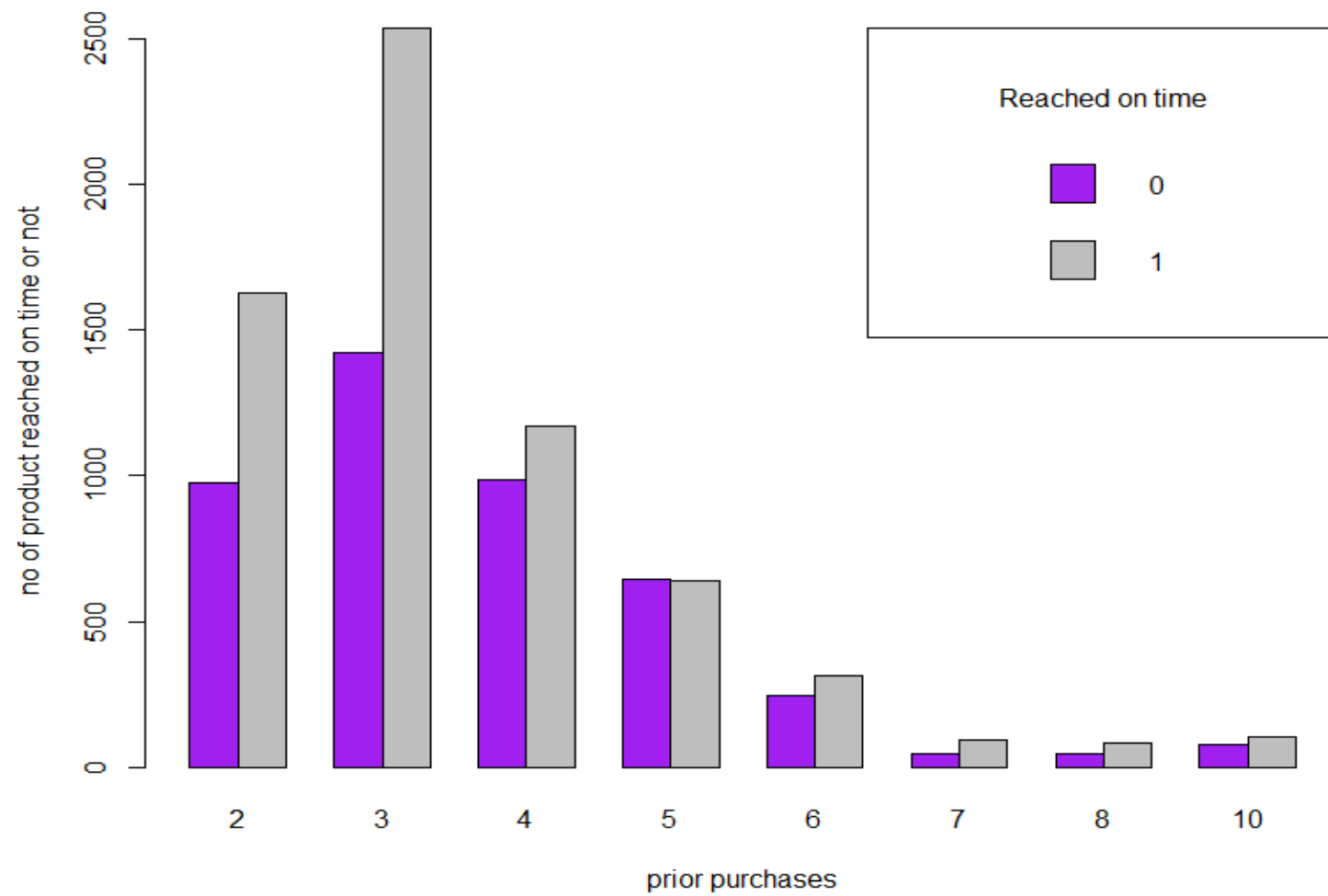
The above bar chart represents the mode of shipment and the frequency of items that reach on time or not for each shipment .Here we see that ship had the max no of items in both the sections i.e the products that reached on time and that didn't reach on time.



The above bar chart represents the product importance and the frequency of items that reach on time or not for each section of importance .Here we see that low importance products had the max no of items in both the sections i.e the products that reached on time and that didn't reach on time



We see from the above graph that max products were dispatched from the warehouse f



	Group.1	x
1	Flight	13.16095
2	Road	13.08409
3	Ship	13.49196

Here we observe that the discount of product offered on the product for each mode of transportation of product was approx. to 13.

T tests

Here we see the case for the mean discounts offered by the warehouse

A

Null hypothesis : H1 : mean = 13

Alt. hypothesis H2 : mean != 13

One sample t-test

```
data: q1$Discount_offered
t = 0.58774, df = 1832, p-value = 0.5568
alternative hypothesis: true mean is not equal to 13
95 percent confidence interval:
 12.47983 13.96534
sample estimates:
mean of x
 13.22259
```

**Since the p value > 0.05 we accept the null hypothesis at 95%
lvl of significance and the confidence interval for given mean
is 12.47983 13.96534**

Here we see the case for the mean discounts offered by the warehouse B

Null hypothesis : H_1 : mean = 13

Alt. hypothesis H_2 : mean \neq 13

```
One Sample t-test  
data: q2$Discount_offered  
t = 0.50648, df = 1832, p-value = 0.6126  
alternative hypothesis: true mean is not equal to 13  
95 percent confidence interval:  
 12.46251 13.91174  
sample estimates:  
mean of x  
 13.18712
```

Since the p value >0.05 we accept the null hypothesis at 95% lvl of significance and the confidence interval for given mean is 12.46251 13.91174

Here we see the case for the mean discounts offered by the warehouse
C

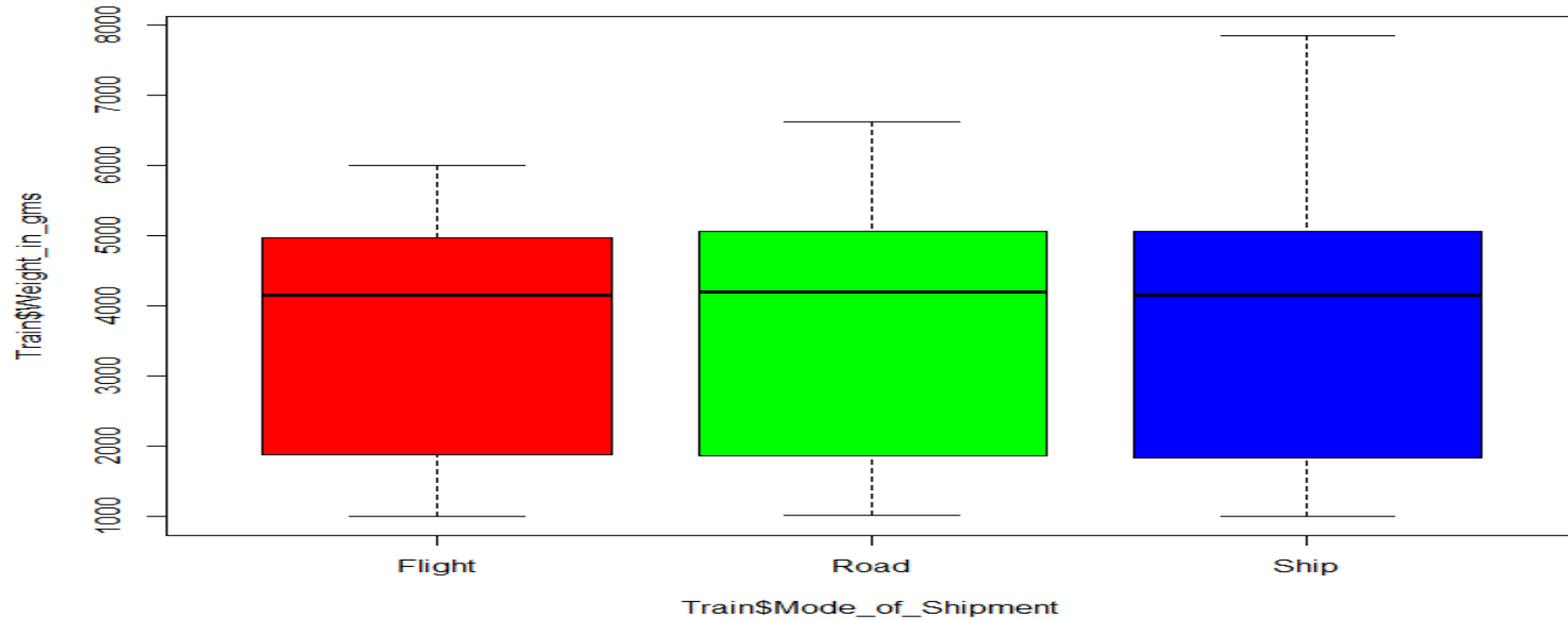
Null hypothesis : H1 : mean = 13

Alt. hypothesis H2 : mean != 13

One sample t-test

```
data: q3$Discount_offered
t = 1.0517, df = 1832, p-value = 0.2931
alternative hypothesis: true mean is not equal to 13
95 percent confidence interval:
 12.65415 14.14564
sample estimates:
mean of x
 13.39989
```

Since the p value > 0.05 we accept the null hypothesis at 95% lvl of significance and the confidence interval for given mean is 12.65415
14.14564



From the above boxplot we see that medians of all the mode of shipment is equal

Anova

Here we carryout anova for mode of shipment and the mean weight of the product carried by each mode of shipment.

Null hypothesis: H1 :

mean_(flight)=mean_weight(ship)=mean_weight(road)

Alternate hypothesis H2:

Atleast one of the mean is not equal

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	5.118e+05	255898	0.096	0.909
Residuals	10996	2.941e+10	2674899		

Since the p value (0.909) is greater than 0.05,we accept the null hypothesis at 95 percent lvl of significance and thus conclude that there is no significant difference in the mean weight of the products carried by different mode of transportation.

Here we carryout anova for different warehouse and the mean weight of the product dispatched by each warehouse.

Null hypothesis: H1 :

mean_weight(A)=mean_weight(B)=mean_weight(C)=mean_weight(D)=mean_weight(F)

Alternate hypothesis H2:

Atleast one of the mean is not equal

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group1	4	9.834e+05	245852	0.092	0.985
Residuals	10994	2.941e+10	2675342		

Since the p value (0.985) is greater than 0.05,we accept the null hypothesis at 95 percent lvl of significance and thus conclude that there is no significant difference in the mean weight of the products dispatched by each mode of transportation.