

به نام حق  
محمد ناظمی اردکانی

سوال ۱:

در ابتدا دیتا فریم مورد نظر را در متغیر dt ذخیره میکنیم و head و tail و مختصری از اطلاعات داده ها را از طریق summary بررسی میکنیم.

```
dt <- read.table("C:/Users/Mohammad/Desktop/DirectMarketing.csv", header = T, sep = ",")
```

```
head(dt)
```

```
tail(dt)
```

```
dt$Age <- factor(dt$Age)
```

```
dt$Gender <- factor(dt$Gender)
```

```
dt$OwnHome <- factor(dt$OwnHome)
```

```
dt$Married <- factor(dt$Married)
```

```
dt$Location <- factor(dt$Location)
```

```
dt$History <- factor(dt$History)
```

```
summary(dt)
```

```
> head(dt)
  Age Gender OwnHome Married Location Salary Children History Catalogs AmountSpent
1 Old Female Own Single Far 47500 0 High 6 755
2 Middle Male Rent Single Close 63600 0 High 6 1318
3 Young Female Rent Single Close 13500 0 Low 18 296
4 Middle Male Own Married Close 85600 1 High 18 2436
5 Middle Female Own Single Close 68400 0 High 12 1304
6 Young Male Own Married Close 30400 0 Low 6 495
> tail(dt)
  Age Gender OwnHome Married Location Salary Children History Catalogs AmountSpent
995 Young Male Rent Single Close 17600 0 <NA> 6 273
996 Young Female Rent Single Close 19400 1 <NA> 18 384
997 Middle Male Rent Single Far 40500 1 <NA> 18 1073
998 Old Male Own Single Close 44800 0 Medium 24 1417
999 Middle Male Own Married Close 79000 2 Medium 18 671
1000 Young Male Rent Married Close 53600 1 Medium 24 973
```

```
> summary(dt)
  Age           Gender OwnHome Married Location Salary Children History Catalogs AmountSpent
Middle:508 Female:506 Own :516 Married:502 Close:710 Min. : 10100 Min. :0.000 High :255 Min. : 6.00 Min. : 38.0
Old :205 Male :494 Rent:484 Single :498 Far :290 1st Qu.: 29975 1st Qu.:0.000 Low :230 1st Qu.: 6.00 1st Qu.: 488.2
Young :287 Median : 53700 Median :1.000 Medium:212 Median :12.00 Median : 962.0
Mean : 56104 Mean :0.934 NA's :303 Mean :14.68 Mean :1216.8
3rd Qu.: 77025 3rd Qu.:2.000 3rd Qu.:18.00 3rd Qu.:1688.5
Max. :168800 Max. :3.000 Max. :24.00 Max. :6217.0
```

سوال ۲:

برای محاسبه درصد حضور بانوان در مجموعه مشتریان از کد زیر استفاده میکنیم. که پاسخ ۵۰٫۶٪ است.

```
print(sum(dt$Gender == "Female") / (nrow(dt) - sum(is.na(dt$Gender)))) * 100
```

```
> print(sum(dt$Gender == "Female") / (nrow(dt) - sum(is.na(dt$Gender)))) * 100  
[1] 0.506  
[1] 50.6  
> |
```

سوال ۳:

ابتدا باید آقایان متأهل را بیابیم و در قدم بعد حقوقشان را بررسی کنیم. که از بین آقایان متأهل به میزان ۸۷/۳۶۴۶۲٪ حقوق بالای ۵۰ هزار دلار دارند.

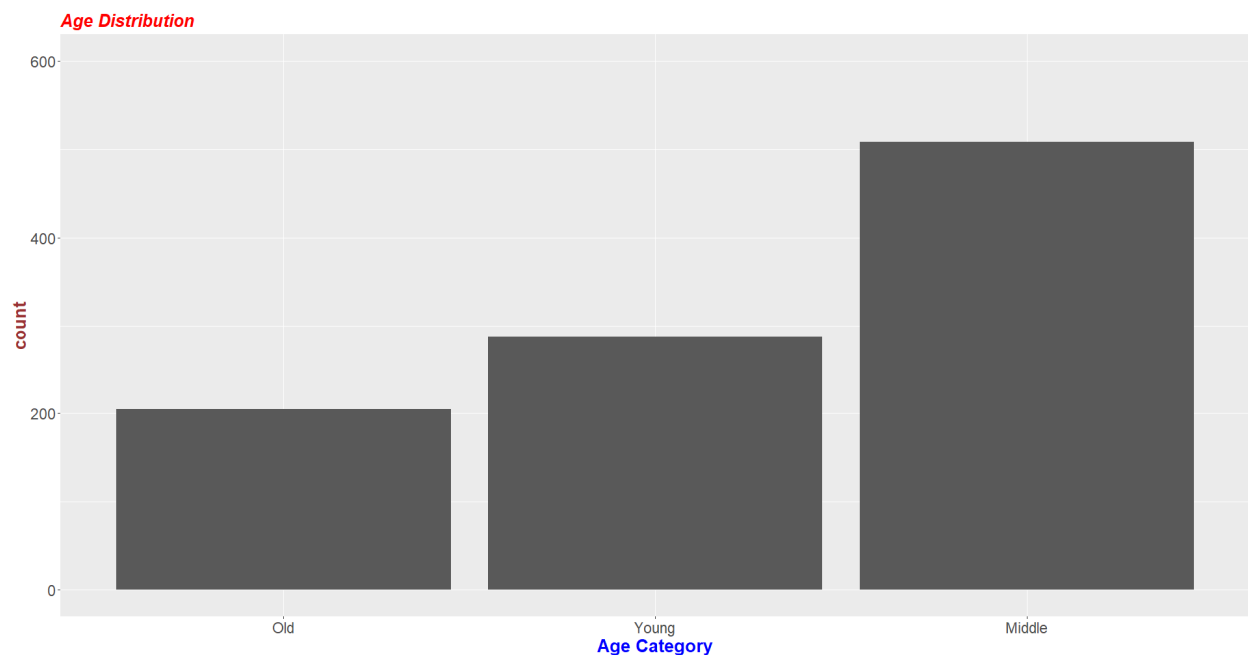
```
print(sum(dt$Gender == "Male" & dt$Married == "Married" & dt$Salary > 50000, na.rm = T)/sum(dt$Gender == "Male" &  
dt$Married == "Married", na.rm = T) * 100)
```

```
> print(sum(dt$Gender == "Male" & dt$Married == "Married" & dt$Salary > 50000, na.rm = T)/sum(dt$Gender == "Male" & dt$Married == "Married", na.rm = T) * 100)  
[1] 87.36462  
> |
```

سوال ۴:

با استفاده از کتابخانه ggplot2 نمودار را رسم میکنیم.

```
library("ggplot2")  
  
age_groups <- table(dt$Age)  
  
age_groups <- as.data.frame(age_groups)  
  
colnames(age_groups) <- c("age", "count")  
  
age_groups <- age_groups[order(age_groups$count),]  
  
ggplot(age_groups, aes(x= reorder(age, count), count))+  
  
  geom_bar(stat = "identity")+  
  
  ggtitle("Age Distribution")+  
  
  xlab("Age Category")+  
  
  ylim(0,600)+  
  
  theme(  
  
    plot.title = element_text(color="red", size=20, face="bold.italic"),  
  
    axis.title.x = element_text(color="blue", size=20, face="bold"),  
  
    axis.title.y = element_text(color="#993333", size=20, face="bold"),  
  
    axis.text.x=element_text(size=rel(2)),  
  
    axis.text.y=element_text(size=rel(2)),  
  
  )
```



## سوال ۵:

برای نمایش از کتابخانه ggplot2 استفاده میکنیم.

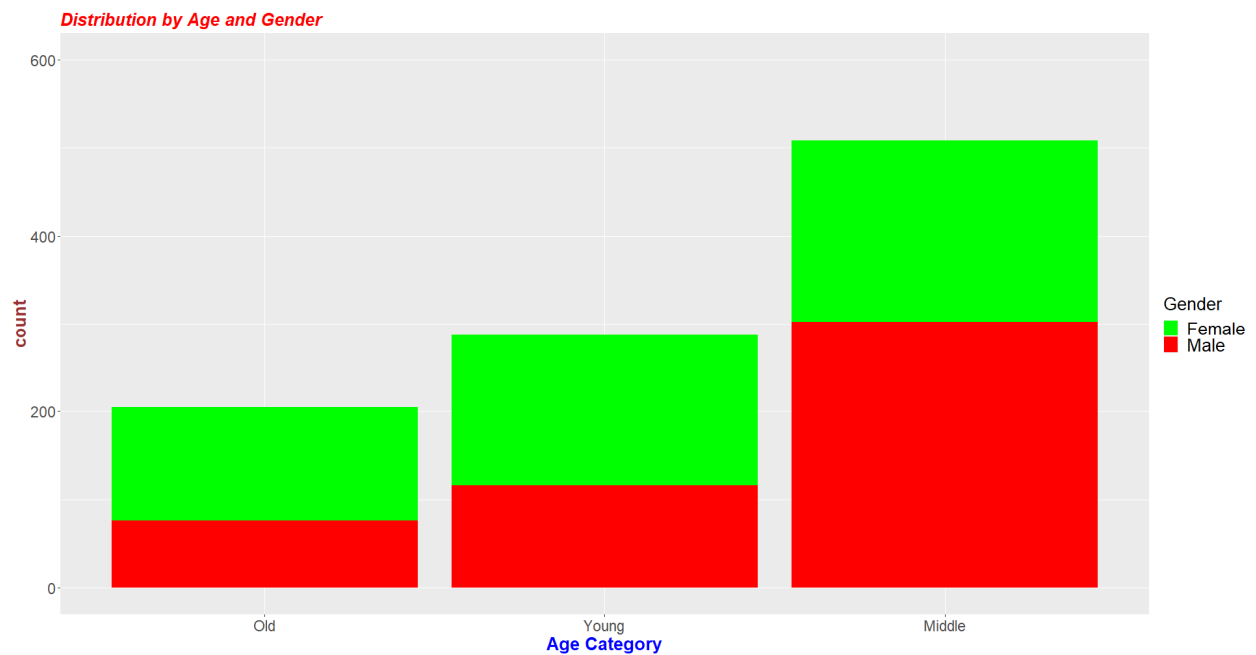
نکته‌ای که وجود دارد این است که این کتابخانه legend را ایجاد میکند اما میتوان از طریق theme تغییری در آن ایجاد کرد.

قابل ذکر است متغیر result\_1 به این منظور ایجاد شده تا بتوان برای رده سنی ترتیب ایجاد کرد.

```
result_1 <- as.character(age_groups$age)
dt$Age <- factor(dt$Age, levels = result_1)
ggplot(dt, aes(Age, fill = Gender))+
  geom_bar()+
  ggtitle("Distribution by Age and Gender")+
  xlab("Age Category")+
  ylim(0,600)+
  theme(
    plot.title = element_text(color="red", size=20, face="bold.italic"),
    axis.title.x = element_text(color="blue", size=20, face="bold"),
    axis.title.y = element_text(color="#993333", size=20, face="bold"),
    axis.text.x=element_text(size=rel(2)),
    axis.text.y=element_text(size=rel(2)),
    legend.title = element_text( size = 20),
    legend.text = element_text( size =20)
```

)+

```
scale_fill_manual(values = c("Green","Red"))
```



سوال ۶:

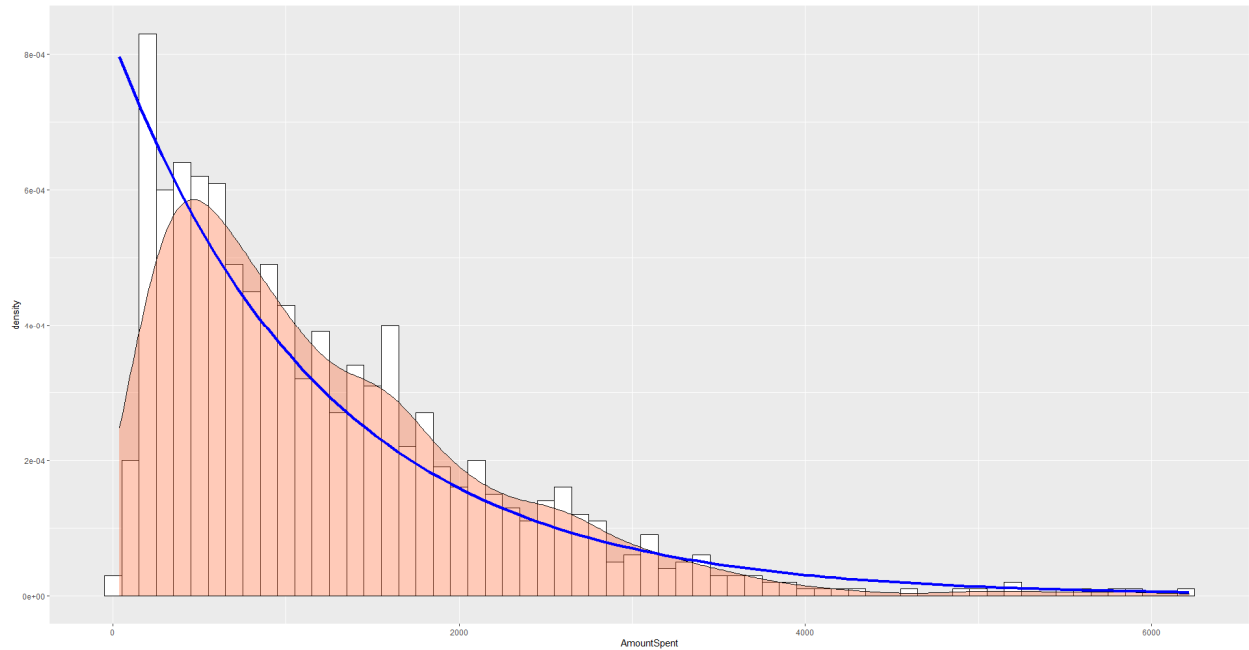
برای رسم این نمودار نیز از کتابخانه ggplot2 بهره میبریم.

```
ggplot(dt, aes(x=AmountSpent))+
```

```
  geom_histogram(aes(y=..density..),binwidth = 100, color="black", fill = "white")+
```

```
  geom_density(alpha = 0.35, fill="#FF6633")+
```

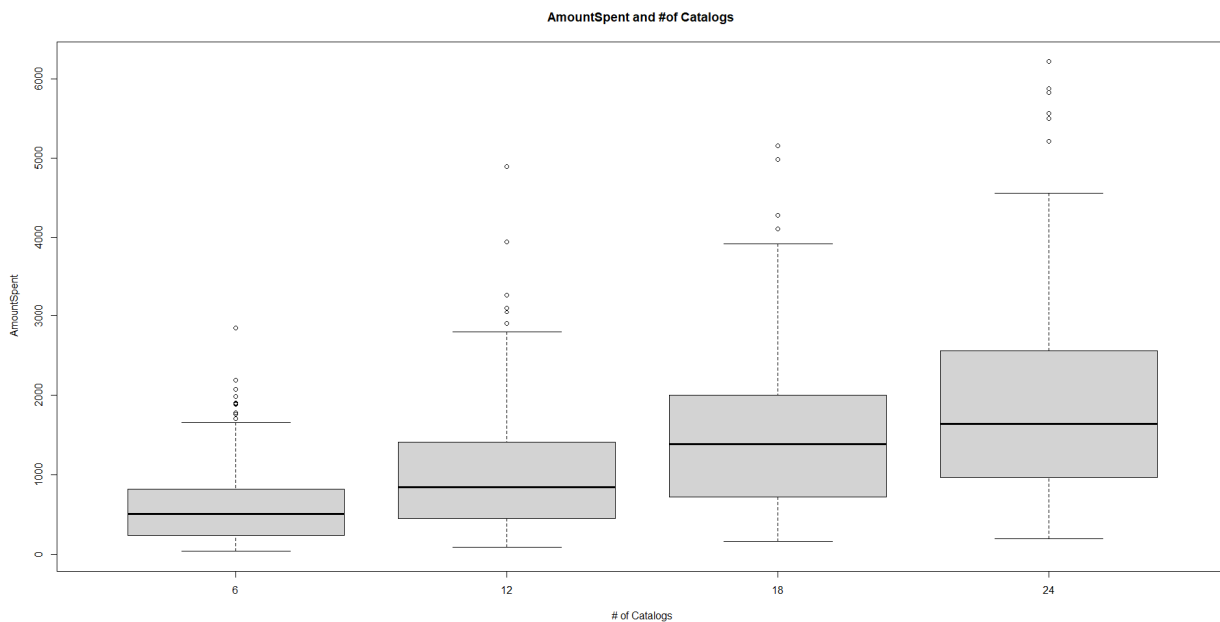
```
  stat_function(fun=dexp, geom = "line", size=1.5, col="blue", args =  
(mean=1/mean(dt$AmountSpent)))
```



سوال ۷:

**boxplot(AmountSpent~Catalogs, data= dt, main="AmountSpent and #of Catalogs ", xlab = "# of Catalogs", ylab = "AmountSpent")**

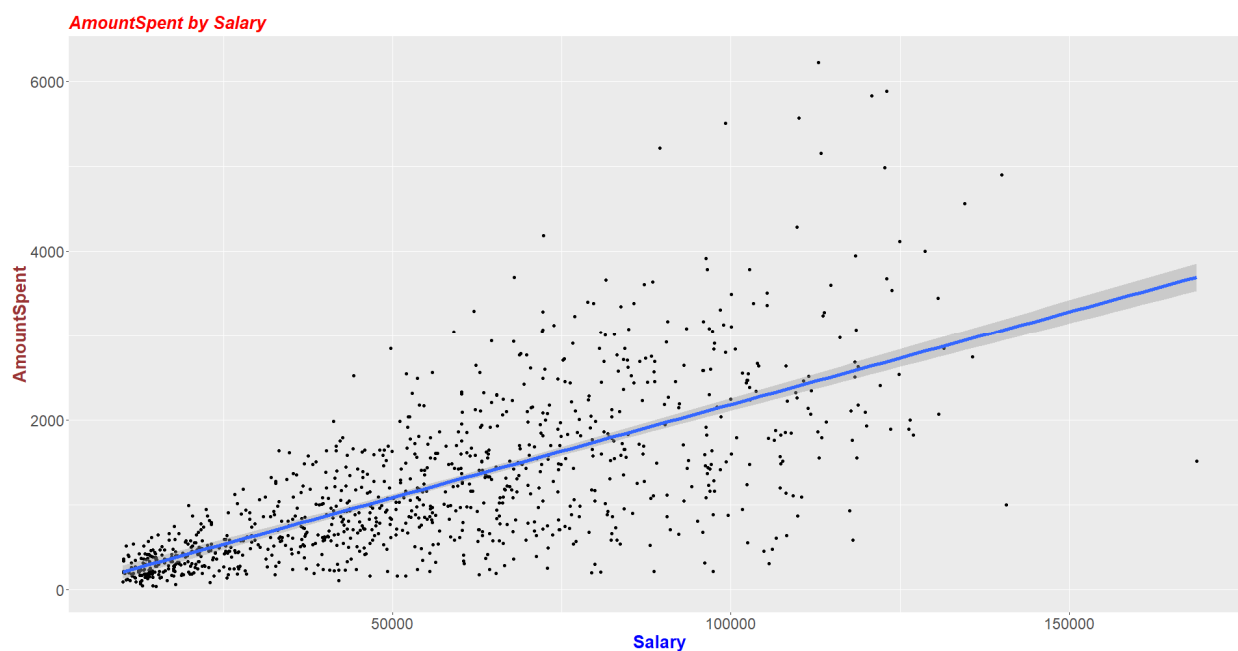
با توجه نمودار افرادی که تعداد کاتالوگ بیشتری دریافت کرده اند خرید بیشتری داشته اند. و میانگین خرید بیشتری هم داشته اند. و به نسبت چارک اول و سوم هم بیشتر بوده.. اما در اینجا شاید اگر کمی هم به تعداد افراد توجه میکردیم نتایج شفاف تر بود اما این امر که با ارسال کاتالوگ بیشتر خرید بیشتری هم صورت گرفته قابل پذیرش است



## سوال ۸:

طبق اطلاعات و نمودار و خط رگرسیون رسم شده بین حجم خرید و حقوق مشتری رابط مثبت وجود دارد و درآمد بالاتر موجب خرید بیشتر میشود.

```
ggplot(dt,aes(x=Salary,y=AmountSpent))+  
  geom_point()+  
  geom_smooth(method = lm,size=2)+  
  ggtitle("AmountSpent by Salary")+  
  theme(  
    plot.title = element_text(color="red", size=20, face="bold.italic"),  
    axis.title.x = element_text(color="blue", size=20, face="bold"),  
    axis.title.y = element_text(color="#993333", size=20, face="bold"),  
    axis.text.x=element_text(size=rel(2)),  
    axis.text.y=element_text(size=rel(2)),  
    legend.title = element_text( size = 20),  
    legend.text = element_text( size =20)  
  )
```

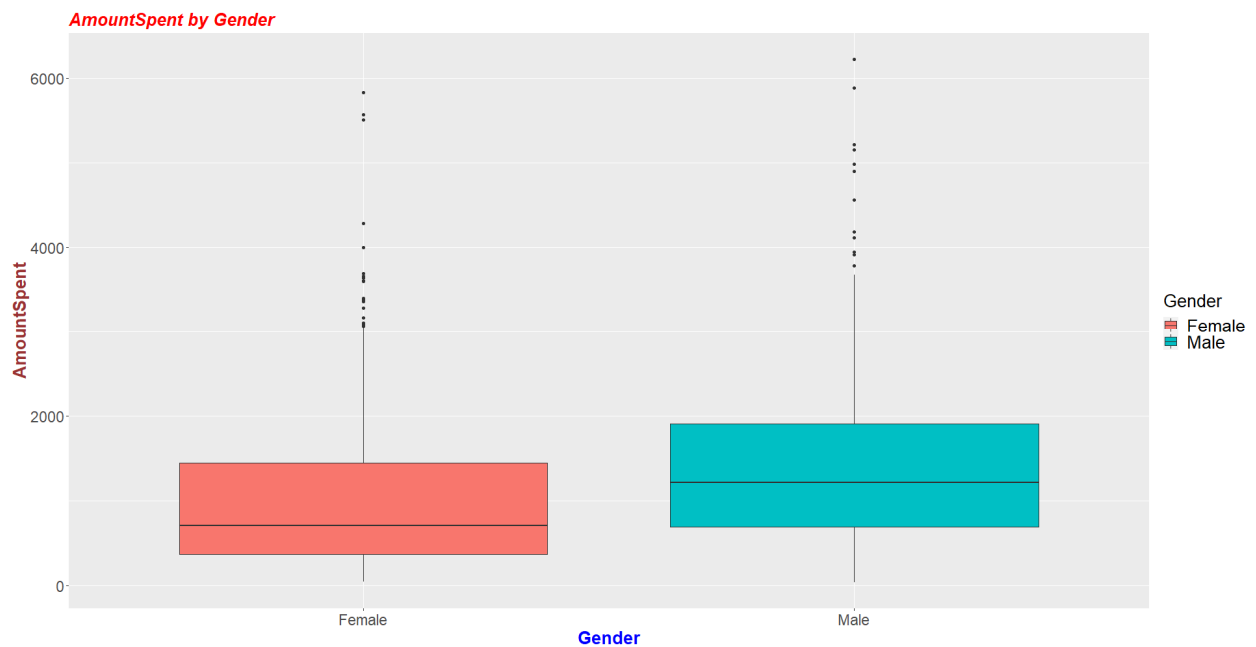


سوال ۹:

بخش الف:

با توجه به نمودار به نظر میرسد آقایان حجم خرید بیشتری داشته اند

```
ggplot(dt, aes(Gender, AmountSpent, fill=Gender))+  
  geom_boxplot()+  
  ggtitle("AmountSpent by Gender")+  
  theme(  
    plot.title = element_text(color="red", size=20, face="bold.italic"),  
    axis.title.x = element_text(color="blue", size=20, face="bold"),  
    axis.title.y = element_text(color="#993333", size=20, face="bold"),  
    axis.text.x=element_text(size=rel(2)),  
    axis.text.y=element_text(size=rel(2)),  
    legend.title = element_text( size = 20),  
    legend.text = element_text( size =20)  
  )
```



میانگین خرید آقایان ۱۴۱۲/۸۵ است و از میانگین خرید بانوان که ۱۰۲۵/۳۴ است بیشتر است.

```
> mean(dt$AmountSpent[dt$Gender == "Male"])
[1] 1412.85
> mean(dt$AmountSpent[dt$Gender == "Female"])
[1] 1025.34
> |
```

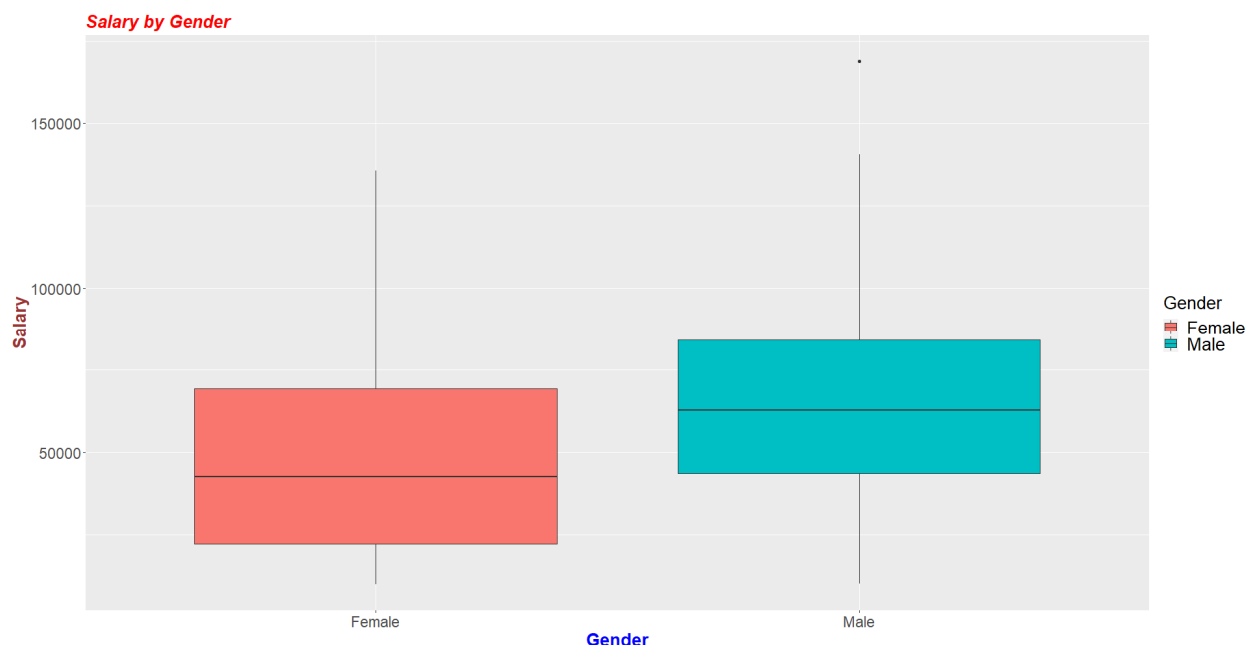
بخش ب:

میانگین درآمد آقایان ۶۴۲۰۲/۴۳ است و از میانگین درآمد بانوان که ۴۸۱۹۷/۴۳ است بیشتر است.

```
> mean(dt$Salary[dt$Gender == "Male"])
[1] 64202.43
> mean(dt$Salary[dt$Gender == "Female"])
[1] 48197.43
> |
```

```
ggplot(dt, aes(Gender, Salary, fill=Gender))+
  geom_boxplot()+
  ggtitle("Salary by Gender")+
  theme(
    plot.title = element_text(color="red", size=20, face="bold.italic"),
    axis.title.x = element_text(color="blue", size=20, face="bold"),
    axis.title.y = element_text(color="#993333", size=20, face="bold"),
    axis.text.x=element_text(size=rel(2)),
    axis.text.y=element_text(size=rel(2)),
    legend.title = element_text( size = 20),
    legend.text = element_text( size =20)
  )
```





با توجه به این دو بخش که میانگین درآمد و خرید آقایان از بانوان بیشتر است شک ما پشتوانه قویتری پیدا میکند که شاید چون آقایان درآمد بیشتری داشته‌اند خرید بیشتری هم داشته‌اند و جنسیت تأثیری در خرید نداشته.

بخش ج:

پس از آنکه دو دسته کم درآمد و پردرآمد را به دست آوردیم چارک دوم خرید هر گروه آقایان و بانوان را در هر دسته درآمدی محاسبه کردیم و به رفتار جالبی رسیدیم. بانوانی که درآمد بالایی دارند نسبت به آقایان بیشتر خرید میکنند. اما در بین کم درآمدها آقایان خرید بیشتری نسبت به بانوان داشته‌اند. برای بررسی بیشتر از نمودار جعبه ای هم استفاده کردیم و سعی شده اثر داده های پرت کم بشود و با دید بهتری مسئله بررسی شود. البته هنوز عواملی مثل فرزند و فاصله از مرکزی که محصولات مشابه بفروشد و ... بررسی نشده‌اند و برای یک تحلیل و نتیجه گیری نهایی زود است.

```
f_high <- quantile(dt$AmountSpent[dt$Gender == "Female" & dt$Salary > median(dt$Salary)], c(0.25, 0.5, 0.75))
```

```
f_low <- quantile(dt$AmountSpent[dt$Gender == "Female" & dt$Salary <= median(dt$Salary)], c(0.25, 0.5, 0.75))
```

```
m_high <- quantile(dt$AmountSpent[dt$Gender == "Male" & dt$Salary > median(dt$Salary)], c(0.25, 0.5, 0.75))
```

```
m_low <- quantile(dt$AmountSpent[dt$Gender == "Male" & dt$Salary <= median(dt$Salary)], c(0.25, 0.5, 0.75))
```

```
gender <- c("Female", "Female", "Male", "Male")
```

```
income_level <- c("High Income", "Low Income", "High Income", "Low Income")
```

```
amount_spent <- c(f_high[2],
```

```
  f_low[2],
```

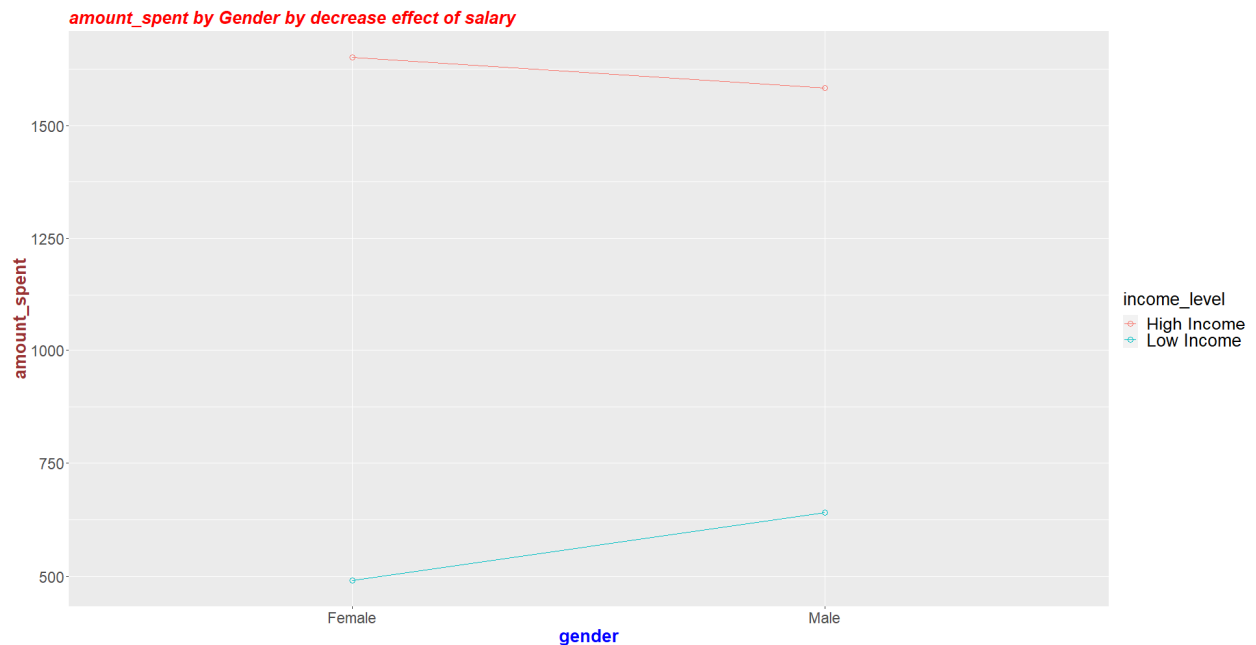
```
  m_high[2],
```

```
  m_low[2])
```

```
d <- data.frame(gender, income_level, amount_spent)
```

```
> d
  gender income_level amount_spent
1 Female   High Income       1651
2 Female   Low Income        491
3 Male    High Income       1583
4 Male    Low Income        641
> |
```

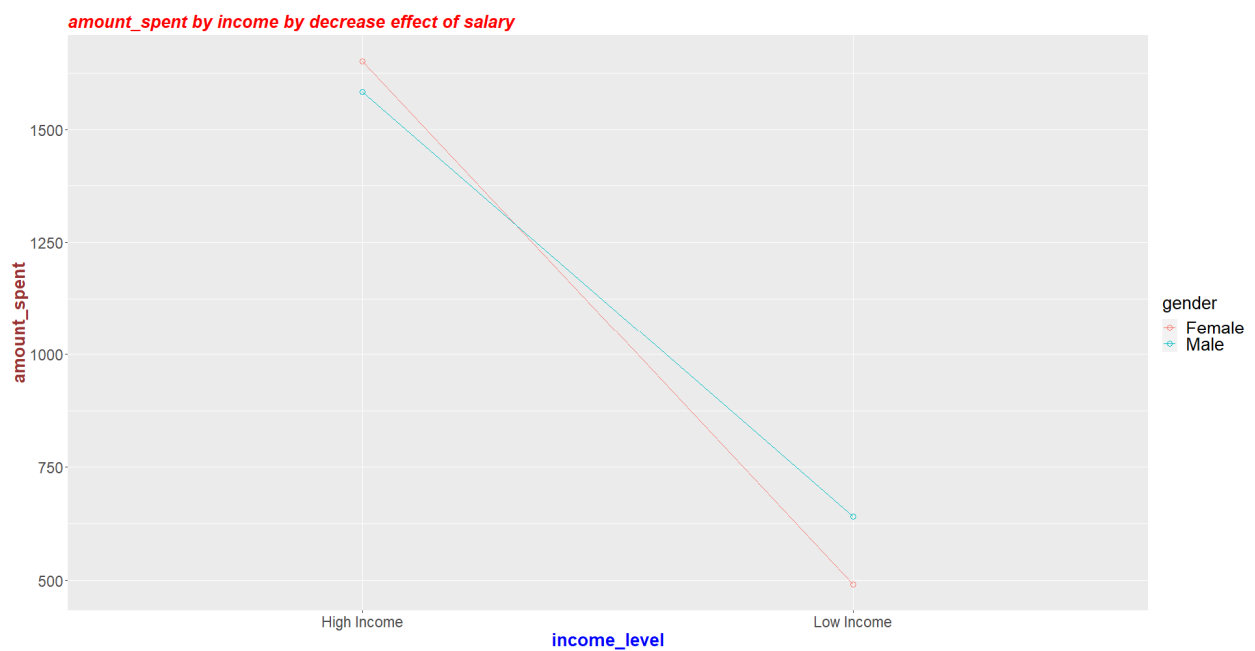
```
ggplot(d, aes(x= income_level, y=amount_spent, group=gender, colour = gender)) +
  geom_line() +
  geom_point(shape = 1 , size =3)+
  ggtitle("amount_spent by Gender by decrease effect of salary")+
  theme(
    plot.title = element_text(color="red", size=20, face="bold.italic"),
    axis.title.x = element_text(color="blue", size=20, face="bold"),
    axis.title.y = element_text(color="#993333", size=20, face="bold"),
    axis.text.x=element_text(size=rel(2)),
    axis.text.y=element_text(size=rel(2)),
    legend.title = element_text( size = 20),
    legend.text = element_text( size =20)
  )
```



```

ggplot(d, aes(x= gender, y=amount_spent, group=income_level, colour = income_level)) +
  geom_line() +
  geom_point(shape = 1 , size =3)+
  ggtitle("amount_spent by income by decrease effect of salary")+
  theme(
    plot.title = element_text(color="red" , size=20, face="bold.italic"),
    axis.title.x = element_text(color="blue", size=20, face="bold"),
    axis.title.y = element_text(color="#993333", size=20, face="bold"),
    axis.text.x=element_text(size=rel(2)),
    axis.text.y=element_text(size=rel(2)),
    legend.title = element_text( size = 20),
    legend.text = element_text( size =20)
  )

```



```

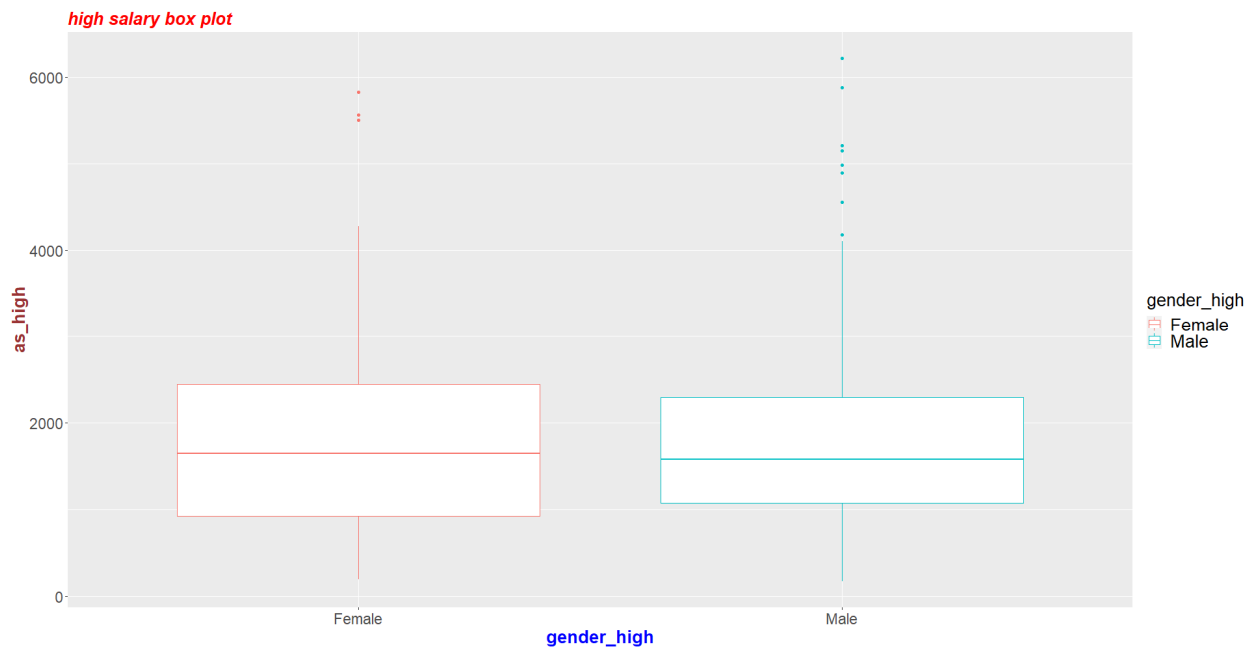
gender_high <- dt$Gender[dt$Salary >median(dt$Salary)]
as_high <- dt$AmountSpent[dt$Salary >median(dt$Salary)]
d_3 <- data.frame(gender_high,as_high)
ggplot(d_3, aes(x= gender_high, y= as_high, colour = gender_high))+

```

```

geom_boxplot()+
ggtitle("high salary box plot")+
theme(
  plot.title = element_text(color="red", size=20, face="bold.italic"),
  axis.title.x = element_text(color="blue", size=20, face="bold"),
  axis.title.y = element_text(color="#993333", size=20, face="bold"),
  axis.text.x=element_text(size=rel(2)),
  axis.text.y=element_text(size=rel(2)),
  legend.title = element_text( size = 20),
  legend.text = element_text( size =20)
)

```



```

gender_low <- dt$Gender[dt$Salary <=median(dt$Salary)]
as_low <- dt$AmountSpent[dt$Salary <=median(dt$Salary)]
d_3 <- data.frame(gender_low,as_low)
ggplot(d_3, aes(x= gender_low, y= as_low, colour = gender_low))+
  geom_boxplot()+
  ggtitle("low salary box plot")+
  theme(
    plot.title = element_text(color="red", size=20, face="bold.italic"),

```

```

axis.title.x = element_text(color="blue", size=20, face="bold"),
axis.title.y = element_text(color="#993333", size=20, face="bold"),
axis.text.x=element_text(size=rel(2)),
axis.text.y=element_text(size=rel(2)),
legend.title = element_text( size = 20),
legend.text = element_text( size =20)
)

```

