به نام حق

محمد ناظمی اردکانی

**سوال1:**

در ابتدا دیتا فریم مورد نظر را در متغییر 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)**

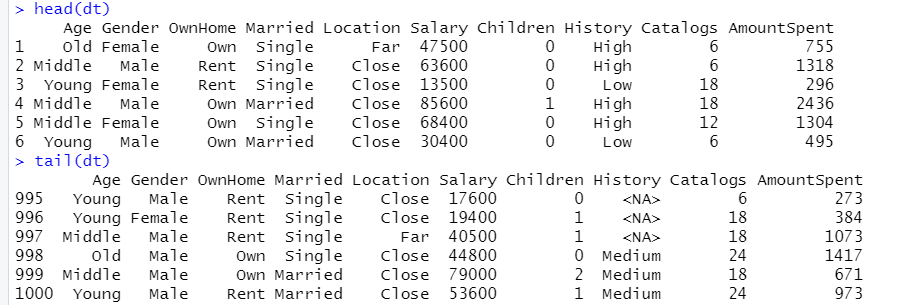
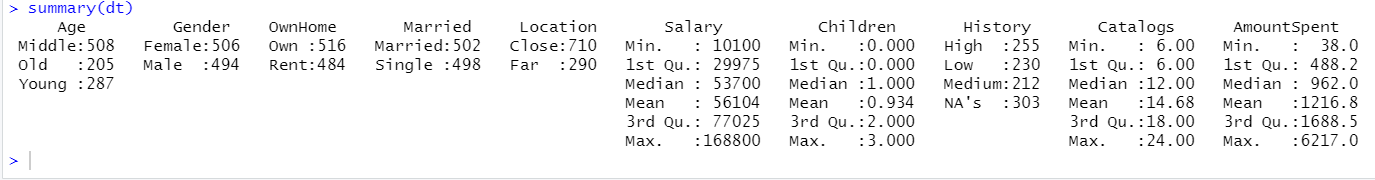
**dt$OwnHome <- factor(dt$OwnHome)**

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

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

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

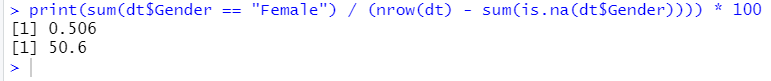
**summary(dt)**



**سوال2:**

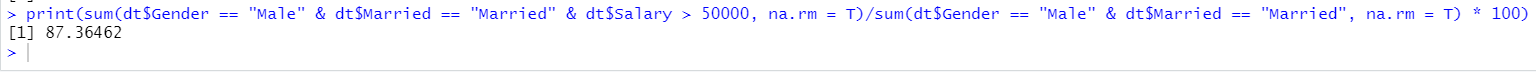
برای محسابه درصد حضور بانوان در مجموعه مشتریان از کد زیر استفاده میکنیم.که پاسخ 50.6% است.

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



**سوال3:**

ابتدا باید آقایان متأهل را بیابیم و در قدم بعد حقوقشان را بررسی کنیم. که از بین آقایان متأهل به میزان 87.36462% حقوق بالای 50 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)**

**سوال4:**

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"),**

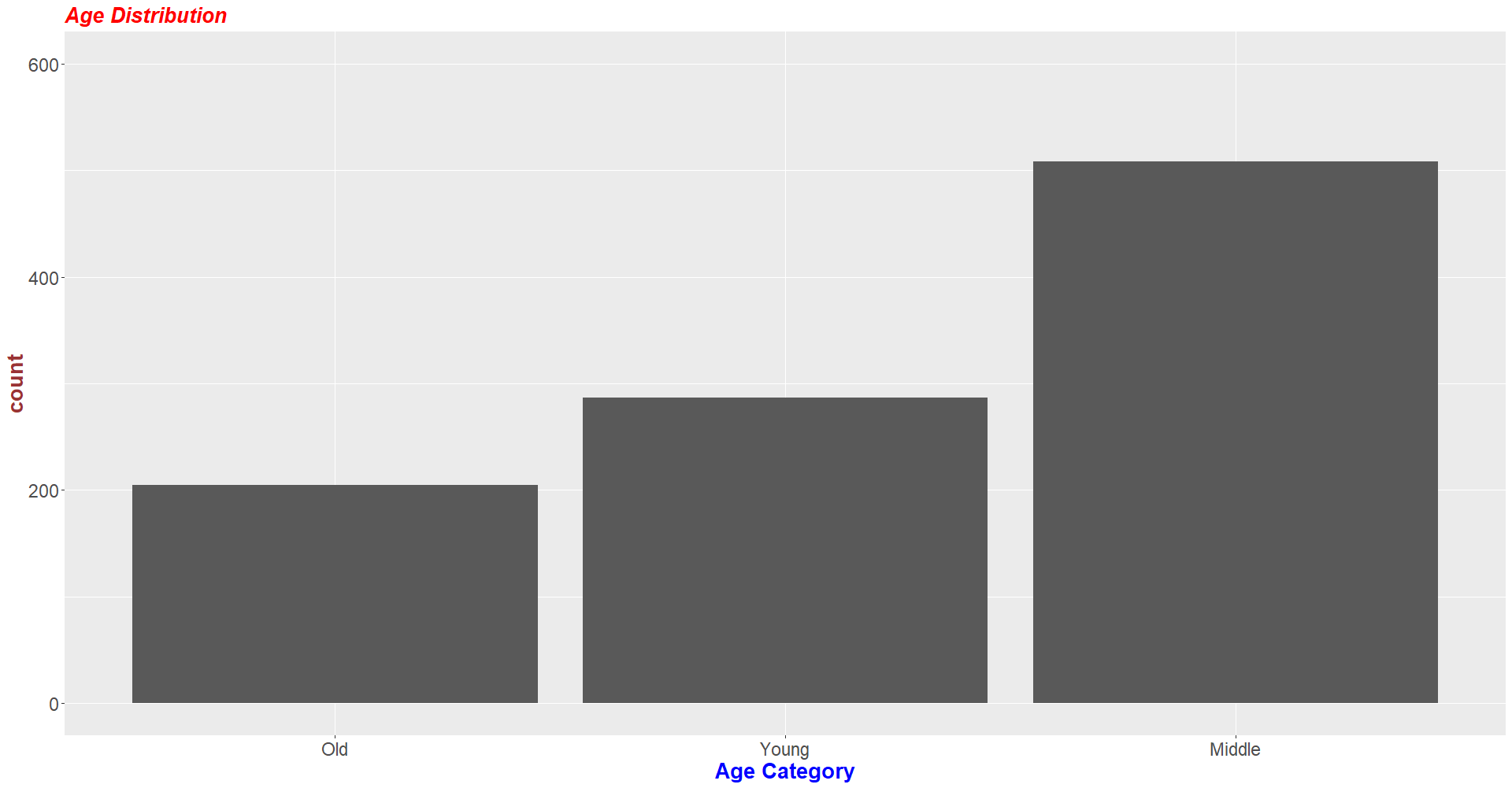
**axis.title.x = element\_text(color="blue", size=20, face="bold"),**

**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)),**

**(**

**سوال5:**

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 تغییراتی در آن ایجاد کرد.

قابل ذکر است متغییر 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)),**

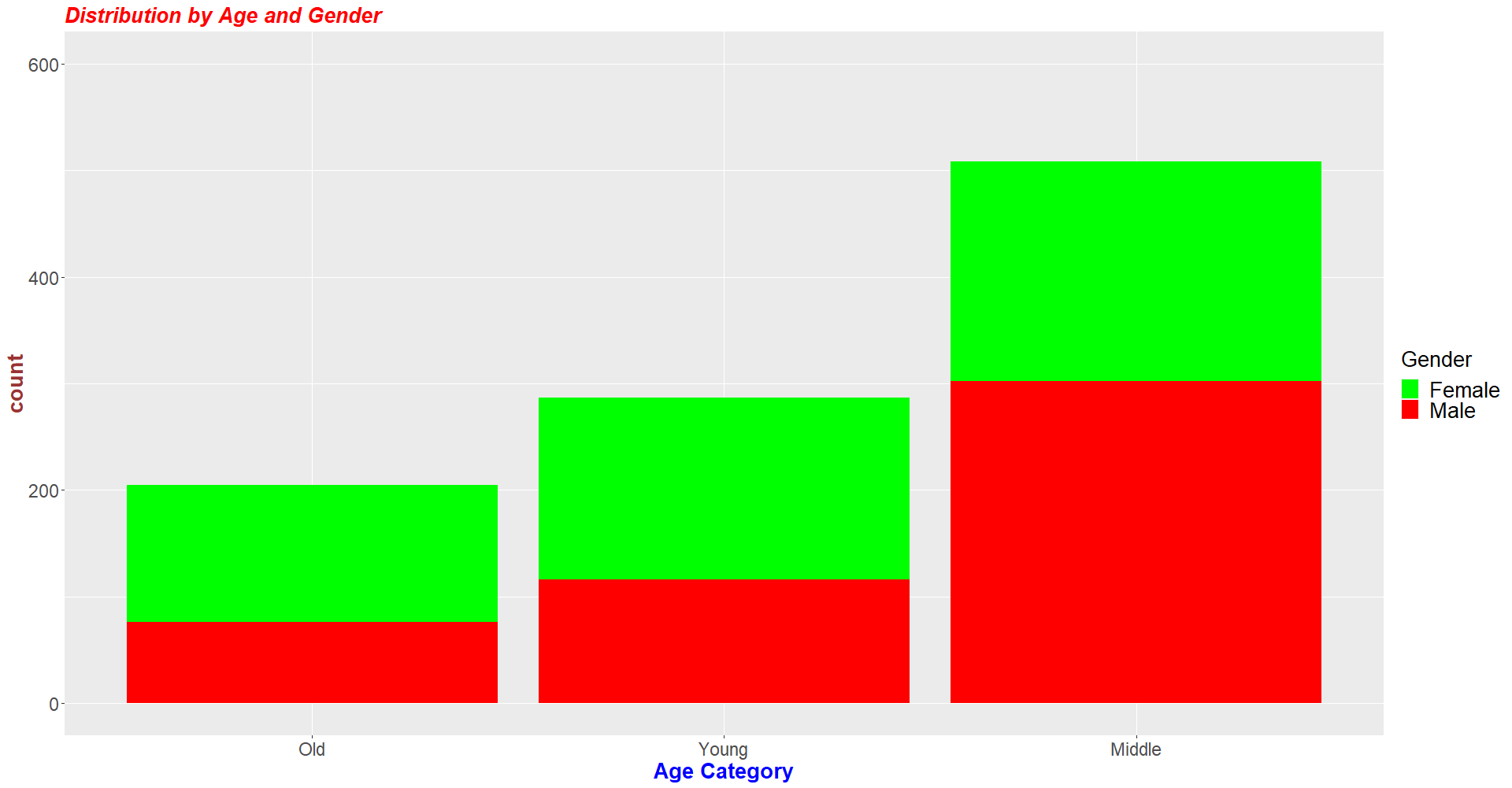
**axis.text.y=element\_text(size=rel(2)),**

**legend.title = element\_text( size = 20),**

**legend.text = element\_text( size =20)**

**)+**

**scale\_fill\_manual(values = c("Green","Red"))**



**سوال6:**

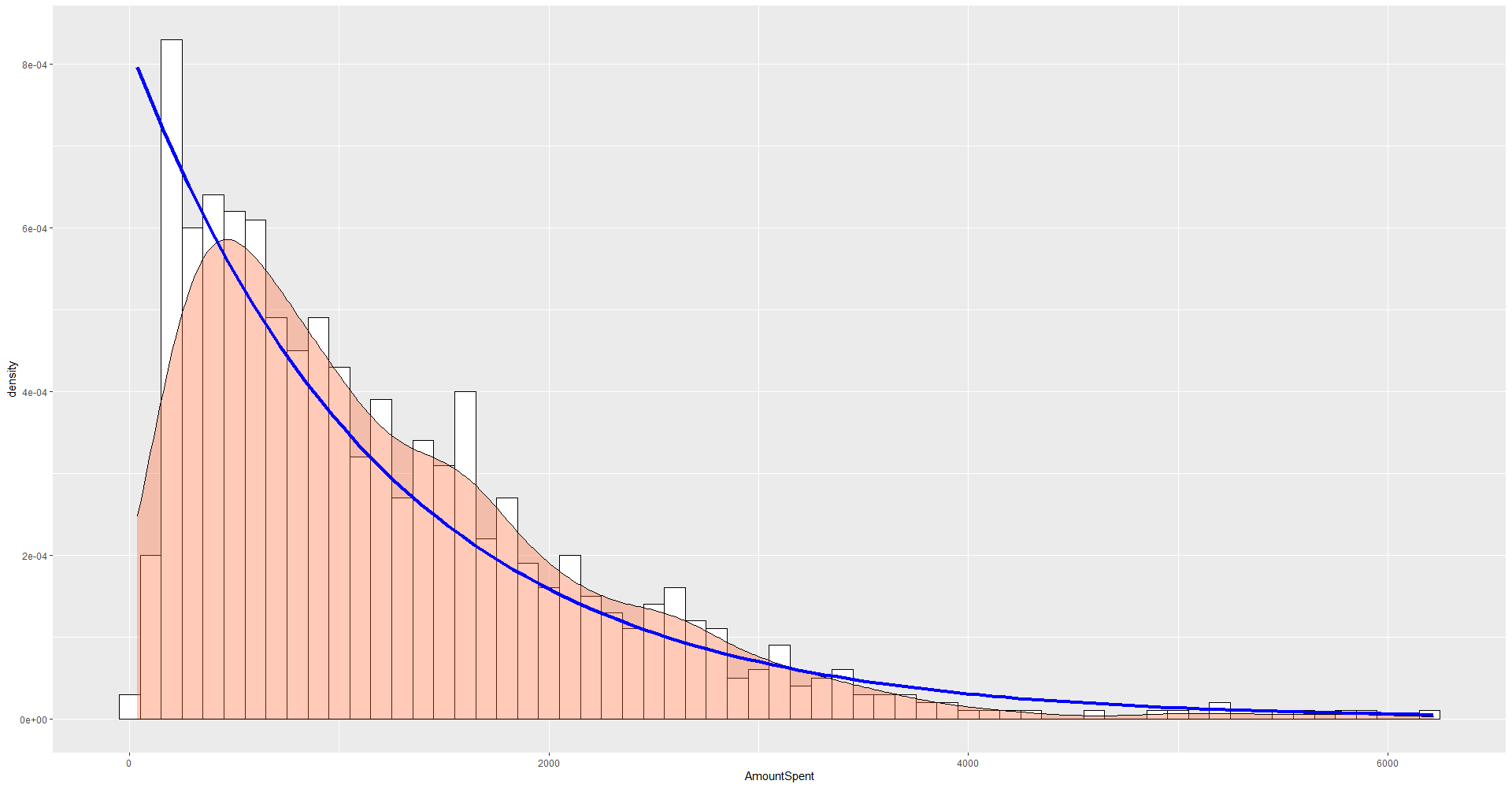
برای رسم این نمودار نیز از کتابخانه ggplot2 بهره میبریم.

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

**geom\_histogram(aes(y=..density..),binwidth = 100, color="black", fill = "white")+**

**geom\_density(alpha = 0.35, fill="#FF6633")+**

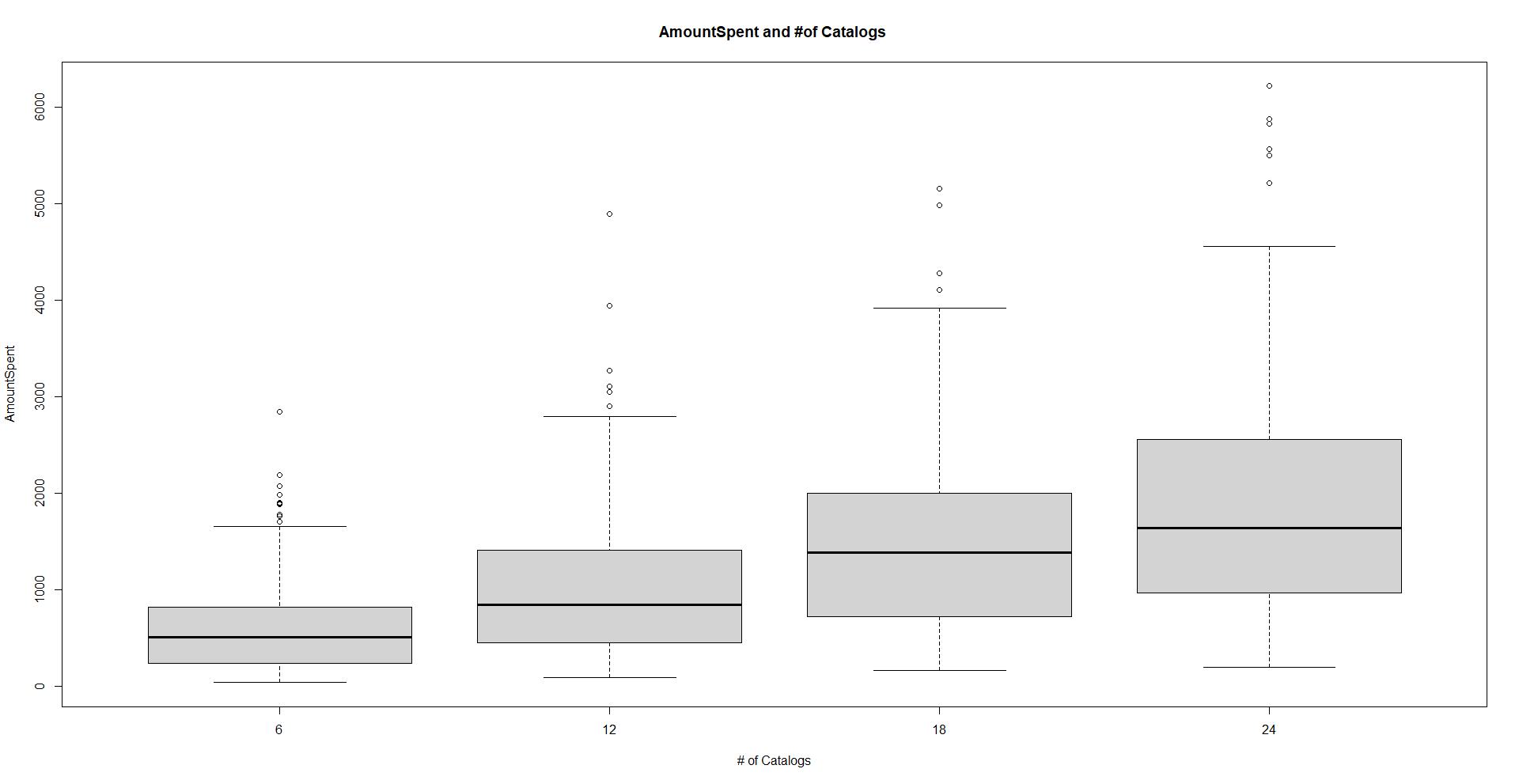
**stat\_function(fun=dexp, geom = "line", size=1.5, col="blue", args = (mean=1/mean(dt$AmountSpent)))**



**سوال7:**

**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 داشته‌اند. 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

****

**سوال8:**

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"),**

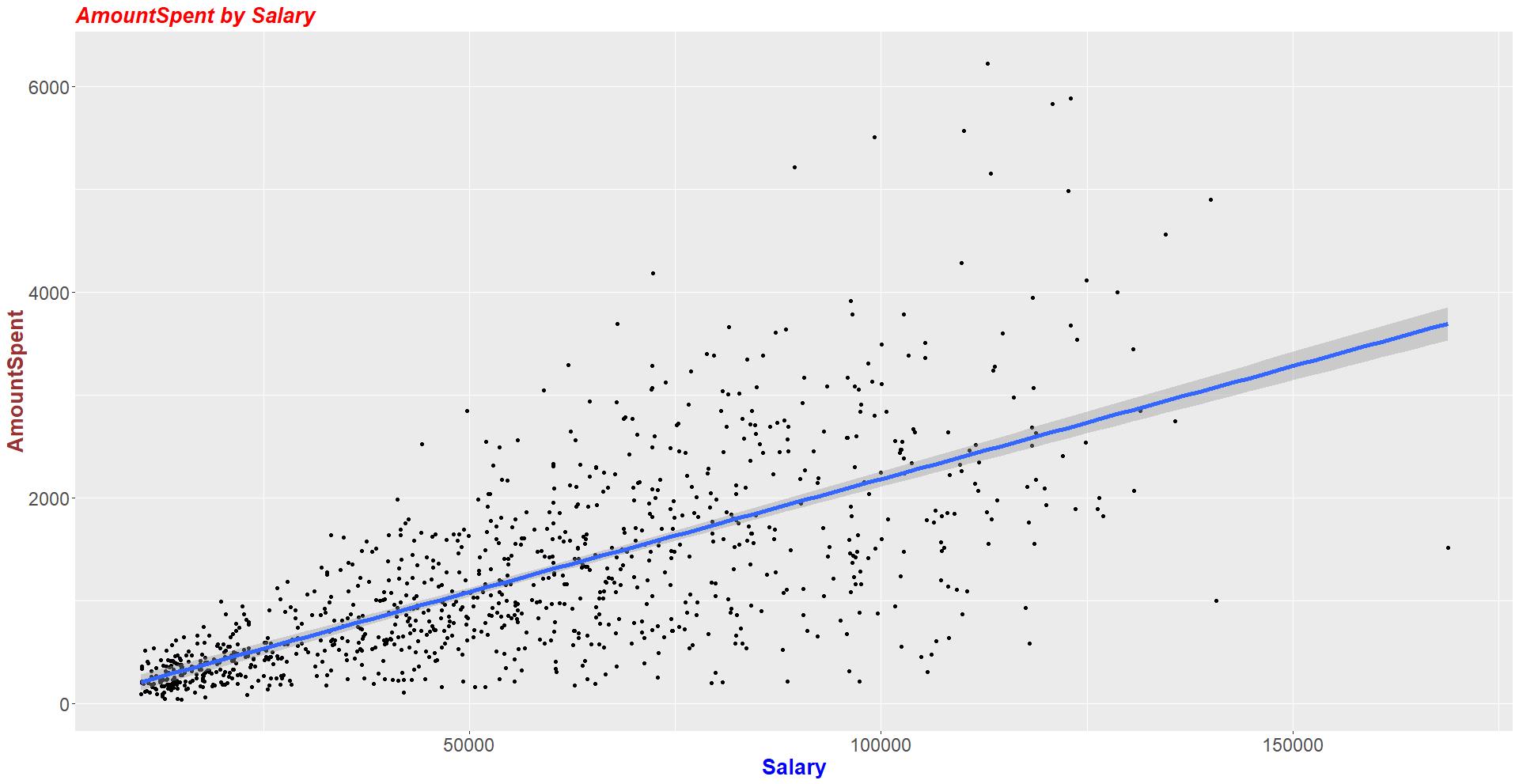
**axis.text.x=element\_text(size=rel(2)),**

**axis.text.y=element\_text(size=rel(2)),**

**legend.title = element\_text( size = 20),**

**legend.text = element\_text( size =20)**

**)**

****

**سوال9:**

بخش الف:

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

**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"),**

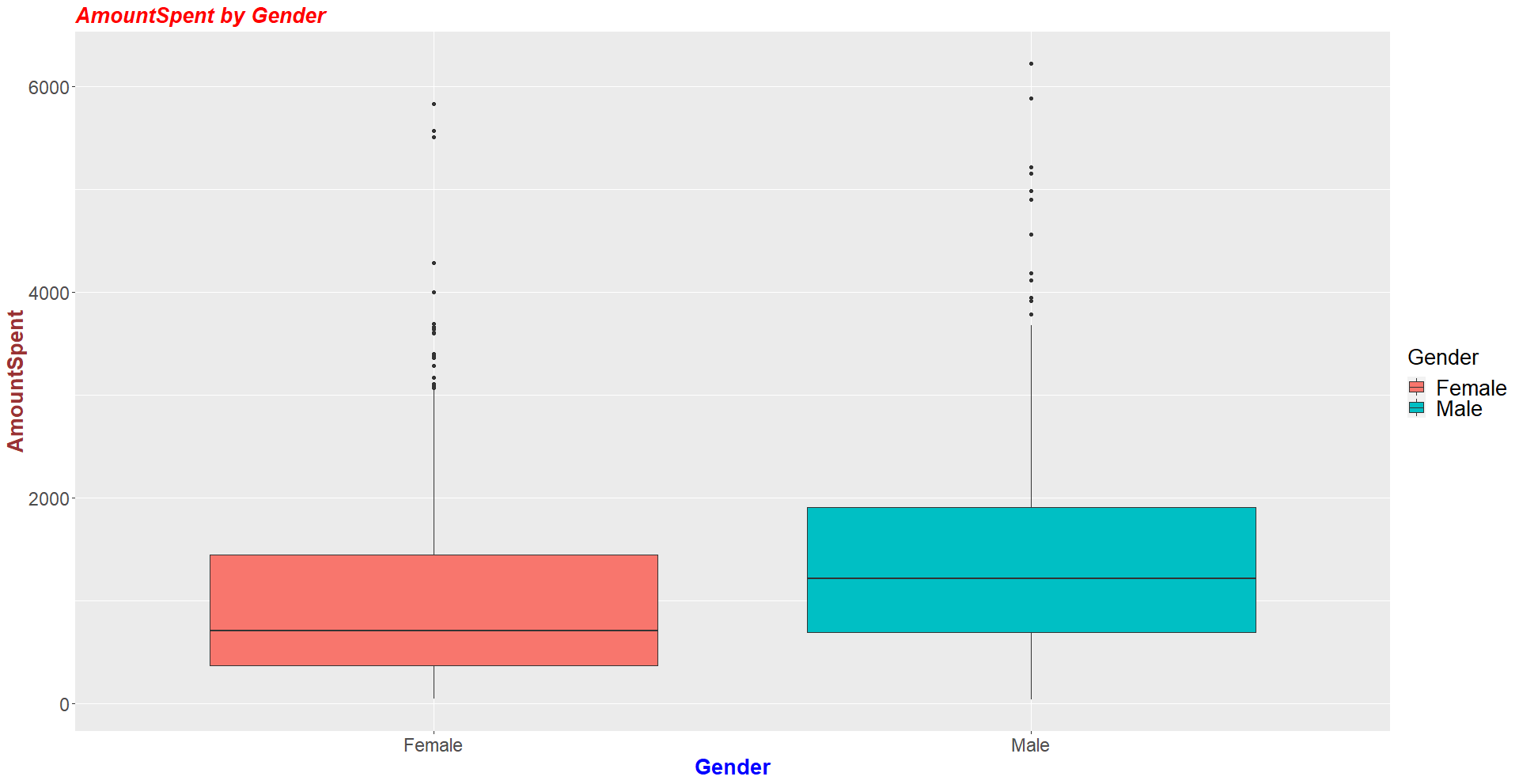
**axis.text.x=element\_text(size=rel(2)),**

**axis.text.y=element\_text(size=rel(2)),**

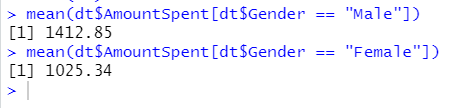
**legend.title = element\_text( size = 20),**

**legend.text = element\_text( size =20)**

**)**

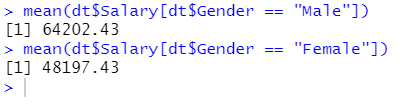
****

میانگین خرید آقایان 1412.85 است و از میانگین خرید بانوان که 1025.34 است بیشتر است.



بخش ب:

میانگین درآمد آقایان 64202.43 است و از میانگین درآمد بانوان که 48197.43 است بیشتر است.



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"),

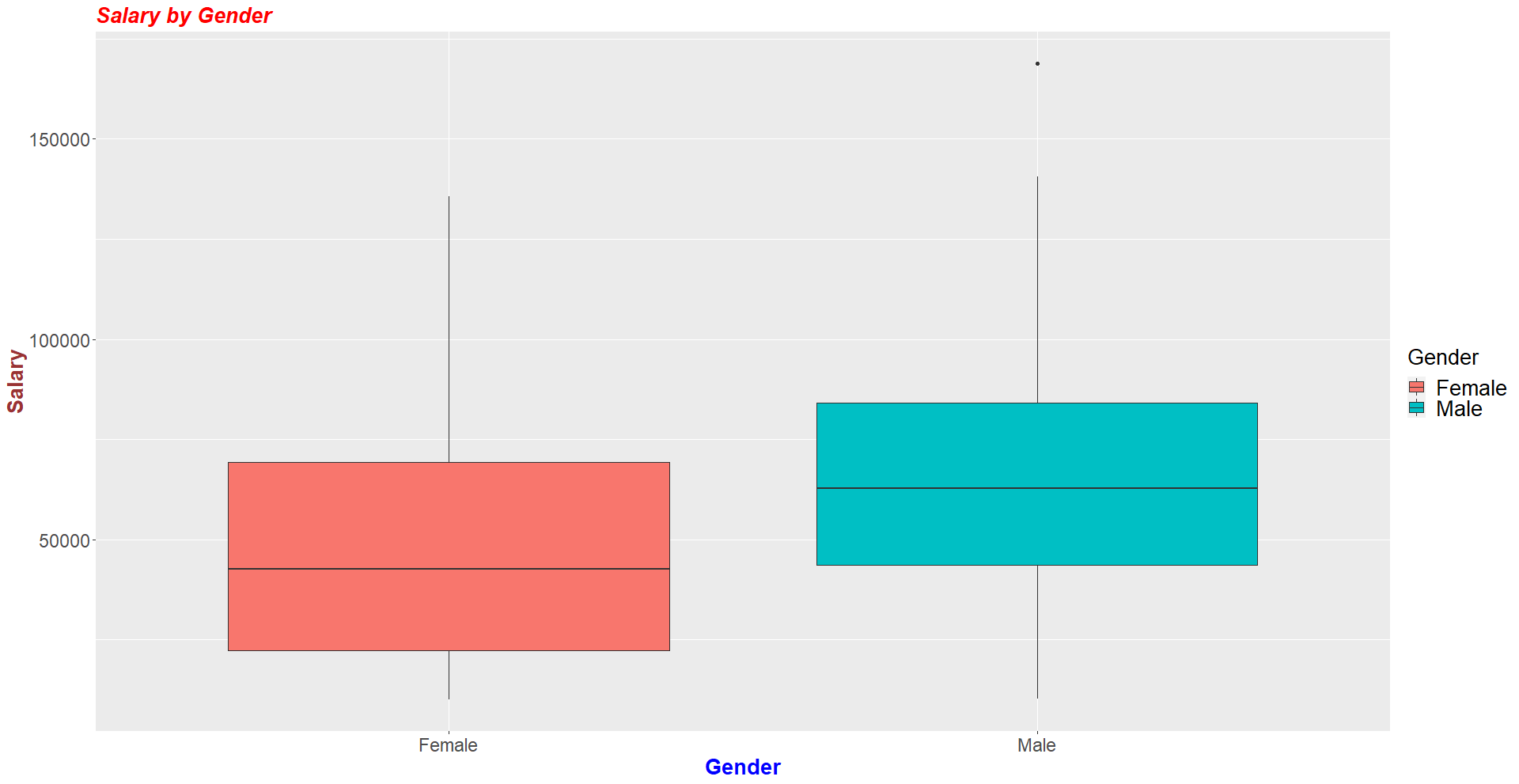
axis.text.x=element\_text(size=rel(2)),

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 داشته اند و جنسیت تاثیری در خرید نداشته.

بخش ج:

پس از آنکه دو دسته کم درآمد و پردرآمد را به دست اوردیم 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 داشته انده. 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 نشده اند 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")**

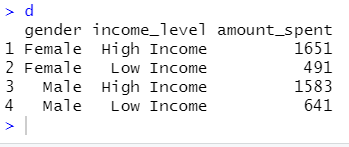
**amount\_spent <- c(f\_high[2],**

**f\_low[2],**

**m\_high[2],**

**m\_low[2])**

**d <- data.frame(gender, income\_level, amount\_spent)**

****

**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"),**

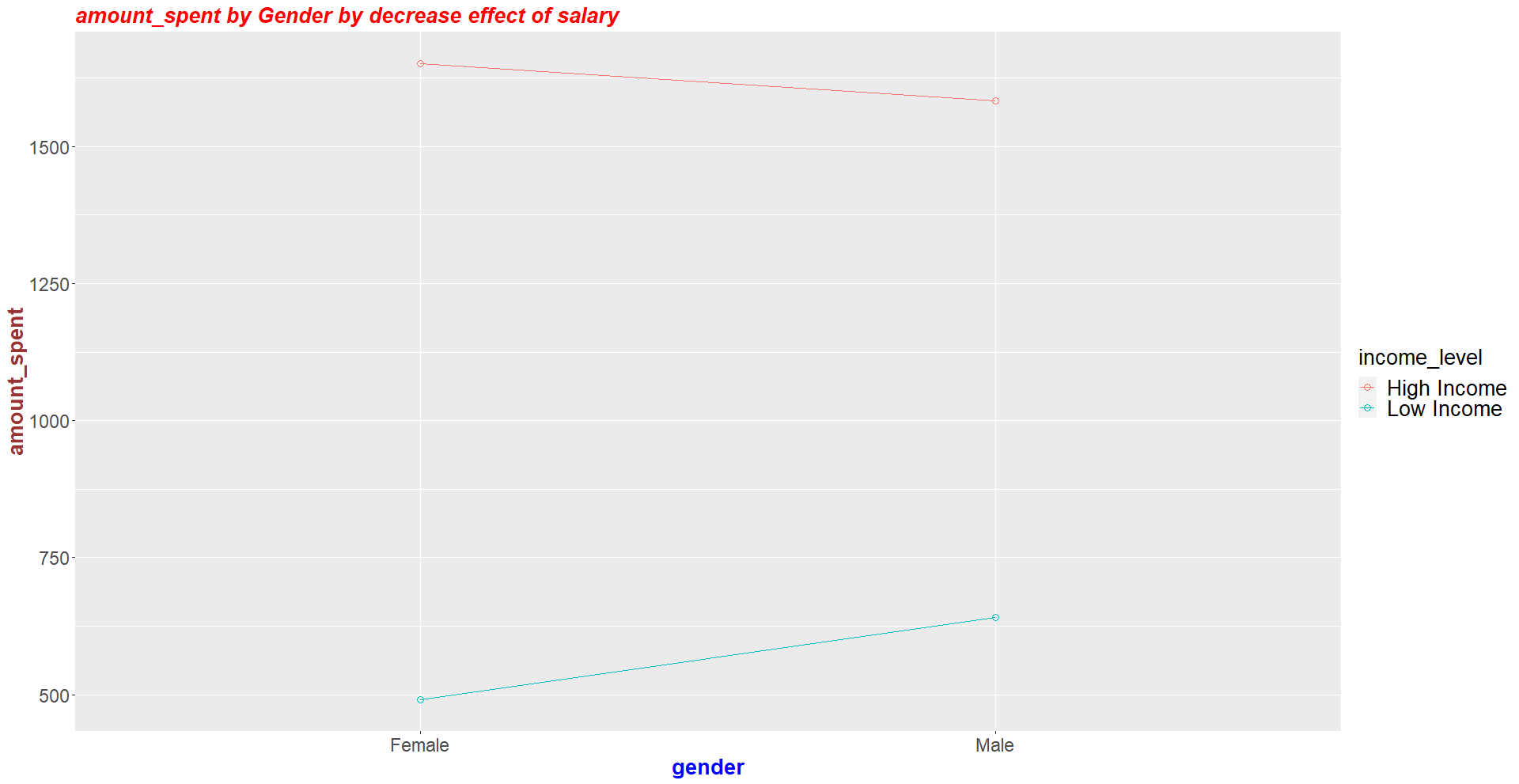
**axis.text.x=element\_text(size=rel(2)),**

**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"),**

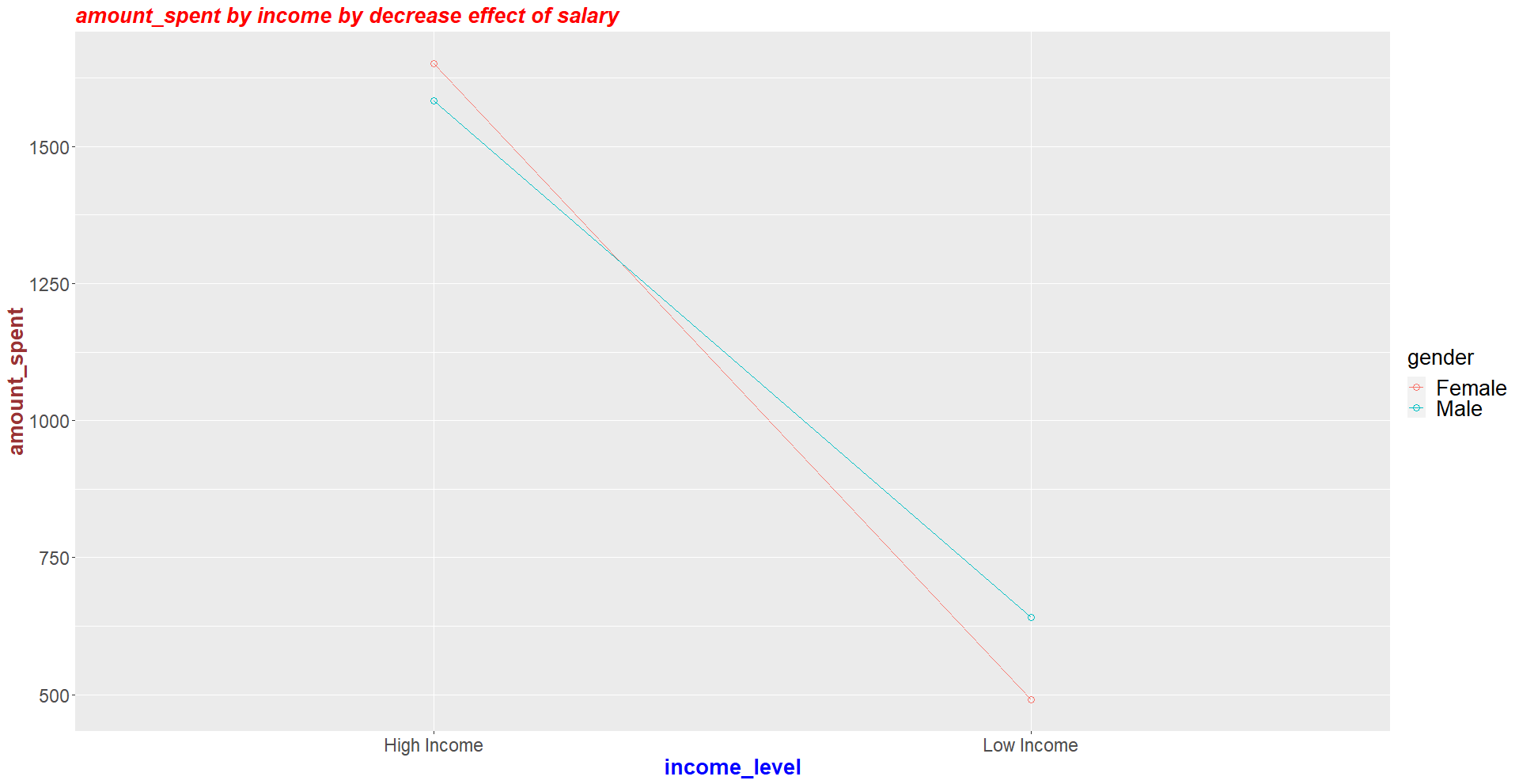
**axis.text.x=element\_text(size=rel(2)),**

**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"),**

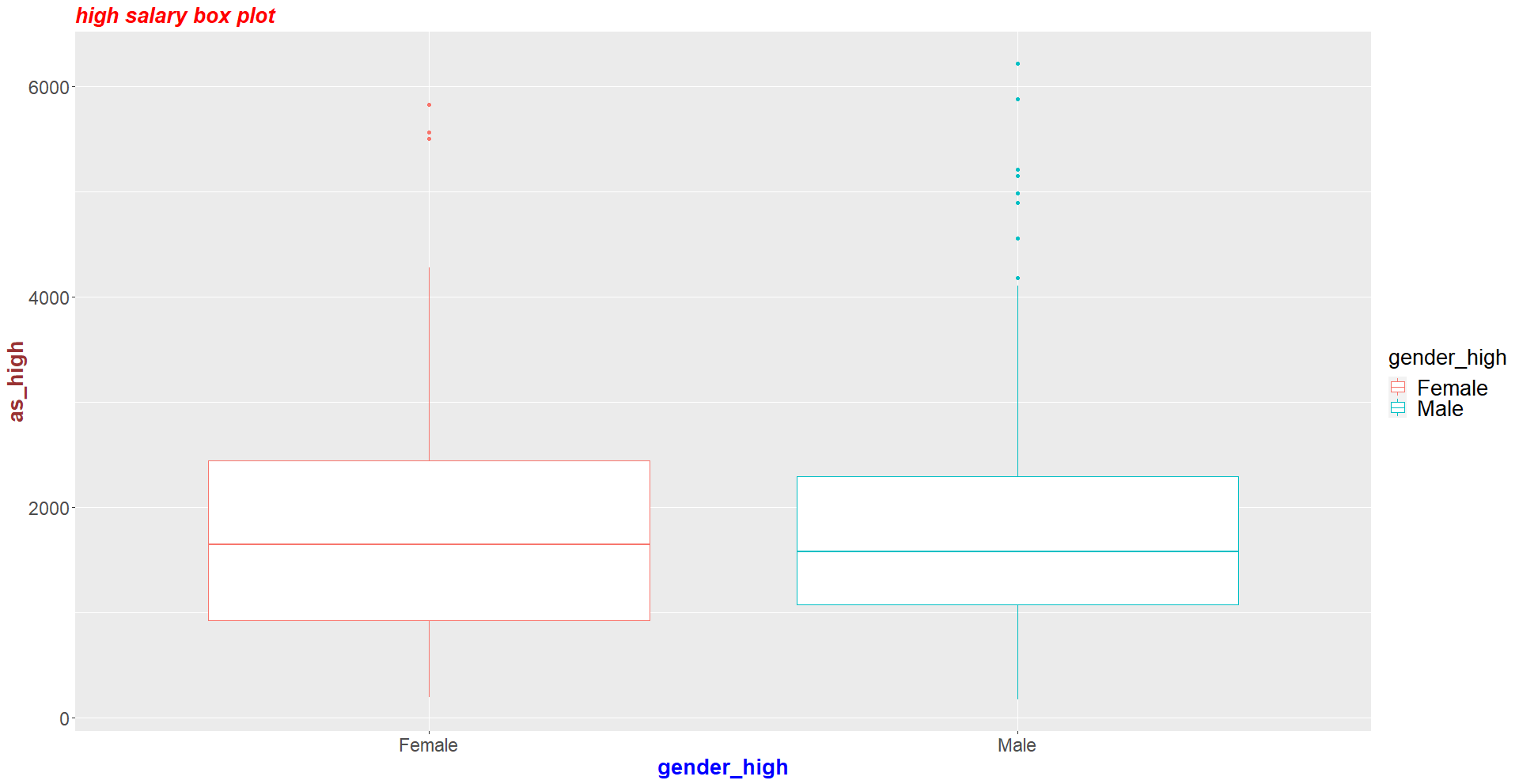
**axis.text.x=element\_text(size=rel(2)),**

**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"),**

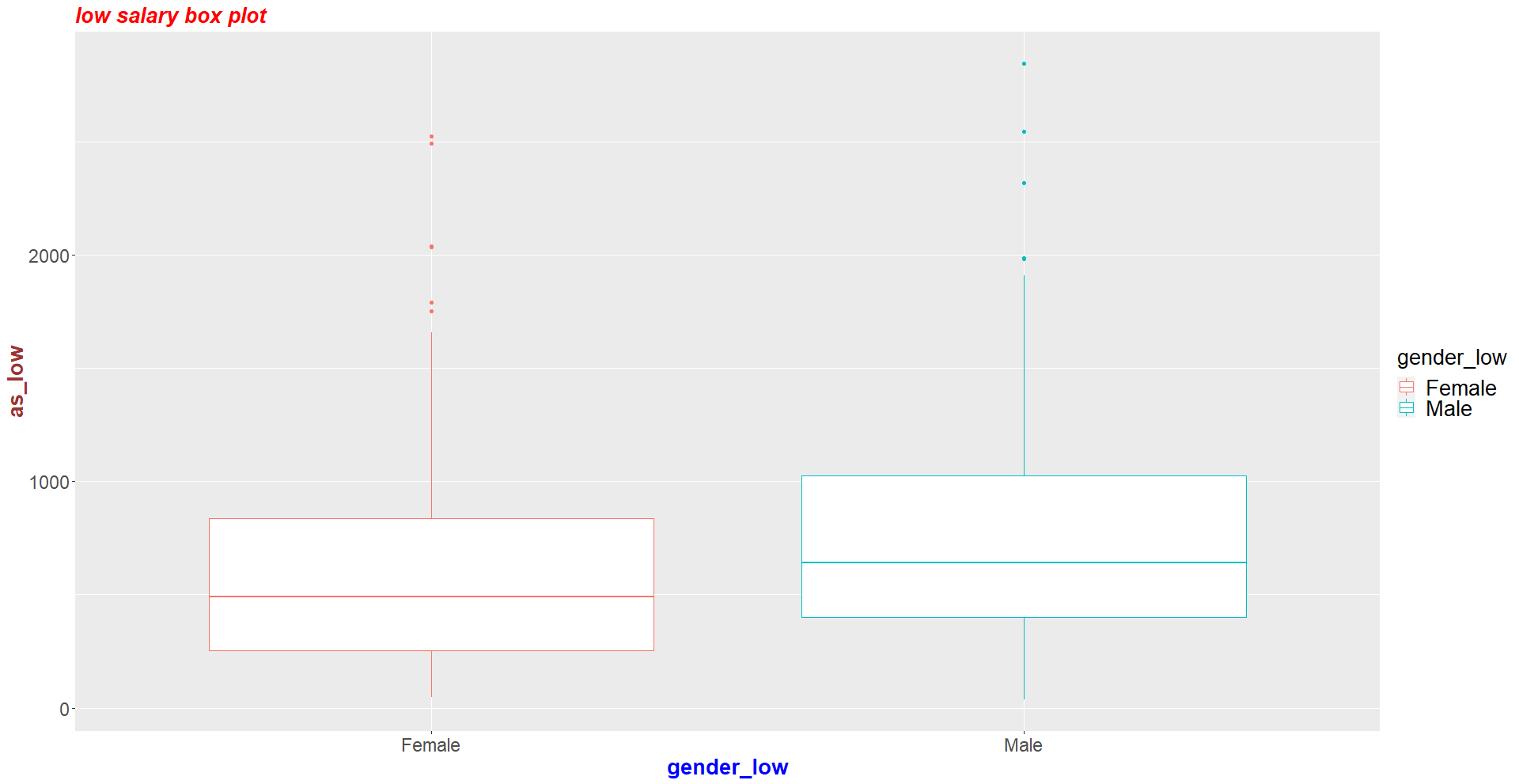
**axis.text.x=element\_text(size=rel(2)),**

**axis.text.y=element\_text(size=rel(2)),**

**legend.title = element\_text( size = 20),**

**legend.text = element\_text( size =20)**

**)**

****