Himmicanes vs Hurricanes: Stat 472 Homework 1

Samantha Bothwell, Crystal Wu, and Charlotte Ding

February 5th, 2019

Group Name: ÷ and Conquer

Problem 1 - Study Summary

The purpose of the hurricane/himmicane study was to show the influence of gender-based expectations on the human toll of hurricanes that are assigned gendered names. Since severe hurricanes might cause a huge amount of fatalities and damage, motivating hurricane preparedness in different ways is still a challenge for authorities. This paper studied the social science factor that predict responses to natural hazards.

The study was based on data collected over six decades of death rates from US. The author of the study performed a series of negative binomial regression analyses to test the relationship between the gender of hurricanes and the number of deaths. Also, the researchers did some laboratory experiments to find whether the gender of the hurricane name affects subjective predictions of hurricane intensity. Some experiments were conducted by the researchers to make the conclusion. The first experiment used five male and five female names from the official 2014 Atlantic Hurricane names and they asked participants predicted each hurricane's intensity. They learned from this experiment that hurricanes with male names were predicted to be more intense than those with female names. In experiment 2, 108 participants judged the riskiness of the hurricanes that have different names. The results of experiment 2 further support the notion that perceived vulnerability to a hurricane depends on the gender of its assigned name. Experiment 3 tested if the gender of the hurricane name affects perceived risk. The participants were given map and scenario on Hurricane Christopher or Hurricane Christina, then reported their evacuation intentions. In this case, Hurricane Christopher was perceived to be riskier than Hurricane Christina. In experiment 4, the scenario given involves a voluntary evacuation order. This experiment indicated that people facing a hurricane with a male name reported significantly greater intentions to follow a voluntary evacuation order. In experiment 5, the researchers addressed possible differences in name familiarity by using a male name that was less popular than the female one and the hurricane has a male name still elicited greater intentions to follow the evacuation order.

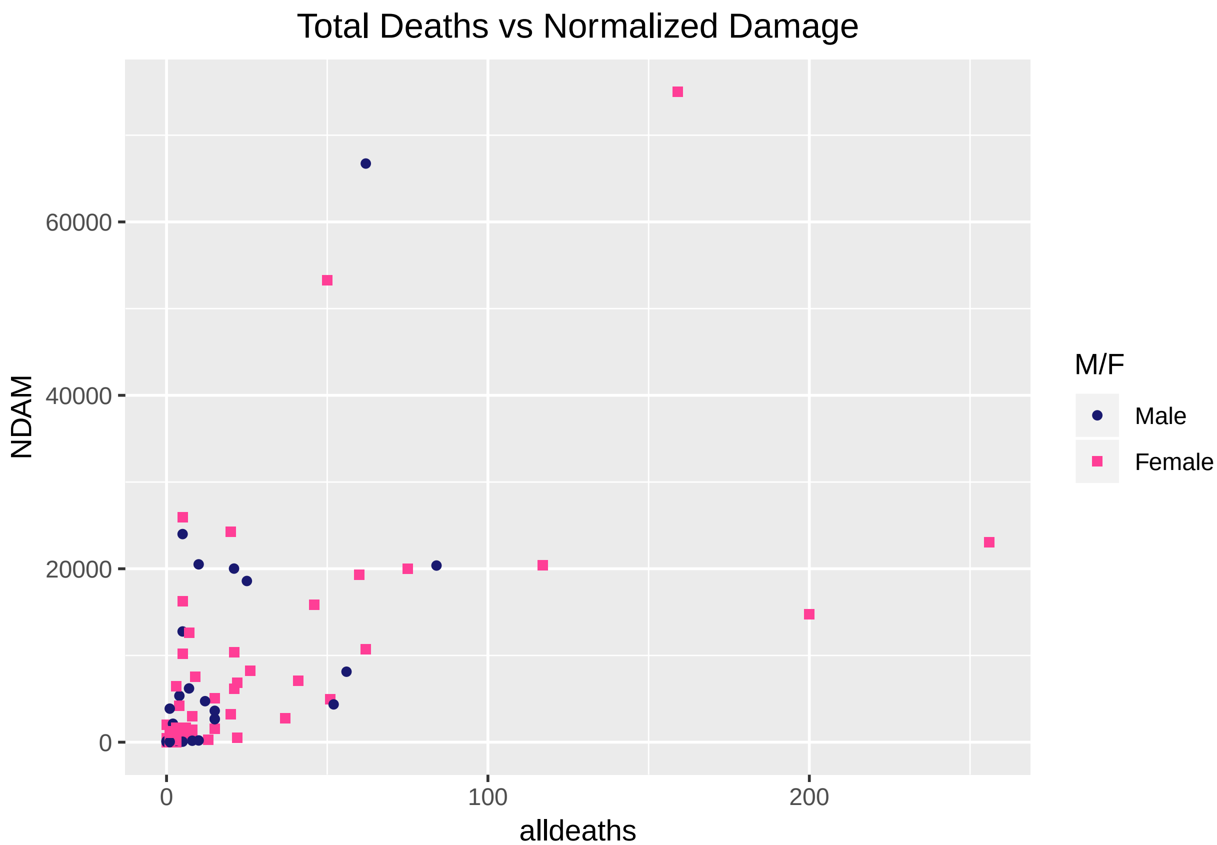
The researchers concluded that feminine-named hurricanes have a more significant effect on causing deaths than masculine-named hurricanes. The reason is that the masculinity level of hurricanes' name let people make different assumptions about severity and guide them to take protective action. Since men and women have different social roles, they generate descriptive and prescriptive expectancies about men and women. People might underestimate the severity of hurricanes with more feminine names and this leads to less preparedness and more deaths.

Problem 2 - Exploratory Data Analysis

Note 1: Hurricane Katrina (2005) and Audrey (1957) were removed as outliers in terms of number of deaths.
Note 2: All hurricanes from 1950-1979 had female names.

Histogram:

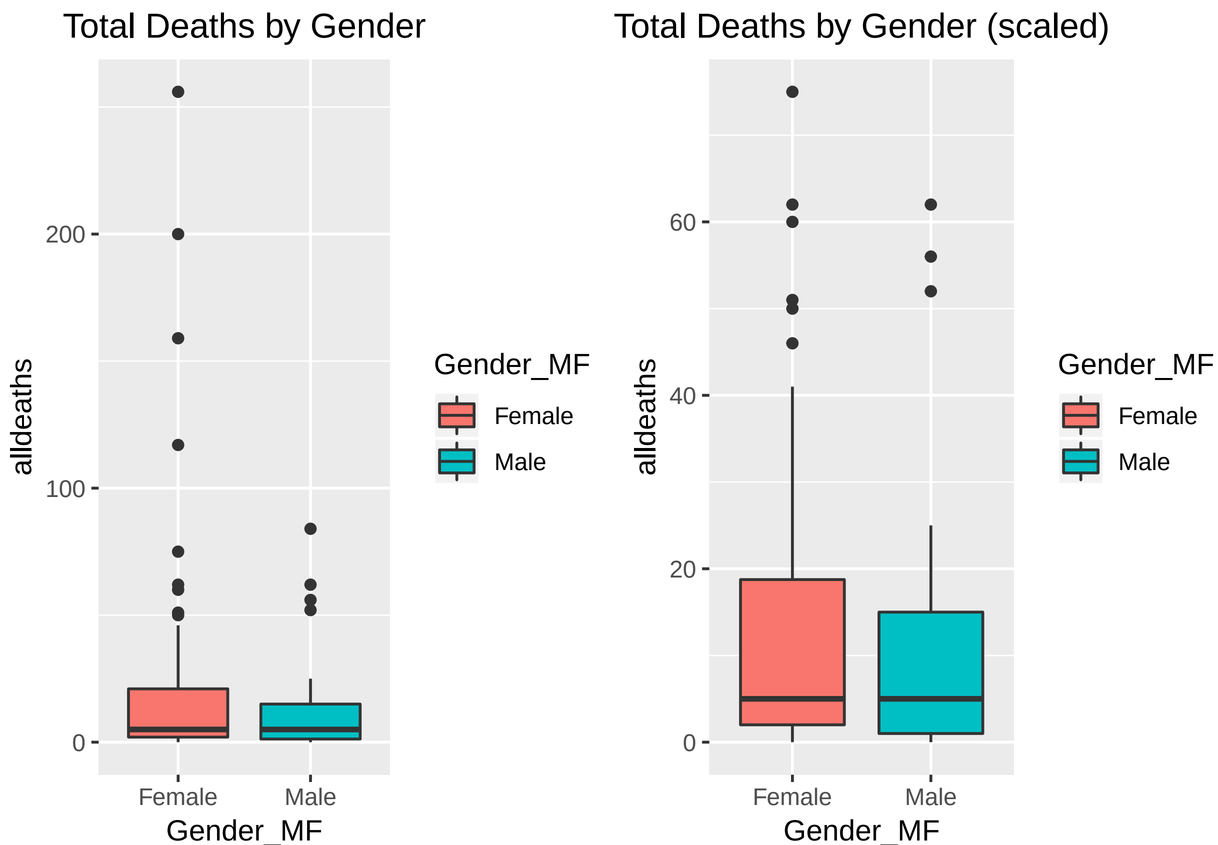
Looking at deaths in vs Normalized Damage separated by male and female:



This, however, does not provide a ton of information about the spread of the data.

Boxplot:

Let's split it up for male and female and assess for outliers with boxplots.



From these side by side boxplots there are a significant number of outliers. 1 is coded for female and 0 for male. While the boxplot for female does appear to have a larger spread, their means are similar within the boxplot, and the plot for females has more outliers than the plot for males.

Summary Statistics:

We can show that the means are approximately similar:

```
# Making a data set for female named hurricanes and male named hurricanes.
HurricanesF <- Hurricanes[Hurricanes$Gender_MF == "Female",]
HurricanesM <- Hurricanes[Hurricanes$Gender_MF == "Male",]

summary(HurricanesF$alldeaths); summary(HurricanesM$alldeaths)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    2.00    5.00   23.76   21.00  256.00

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    1.25    5.00   14.23   15.00   84.00
```

The top row is the summary for female named hurricanes and the bottom row is the summary for male named hurricanes. We can see that the means are close, however the female hurricanes have a higher mean - potentially due to outliers.

T-test: Total Deaths

We can perform a simple t-test to determine if there is a significant difference between the death toll for male and female hurricanes.

```
t.test(HurricanesF$alldeaths, HurricanesM$alldeaths,
       alternative = "two.sided", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  HurricanesF$alldeaths and HurricanesM$alldeaths
## t = 1.3302, df = 89.605, p-value = 0.1868
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.700956 23.750418
## sample estimates:
## mean of x mean of y
##  23.75806  14.23333
```

We have a p-value of **0.2**, which indicates that there is no significant difference between the deaths for female named hurricanes and the deaths for male hurricanes.

T-test: Normalized Damage

We can also perform a t-test to determine if there is a significant difference between the normalized damage for male and female hurricanes.

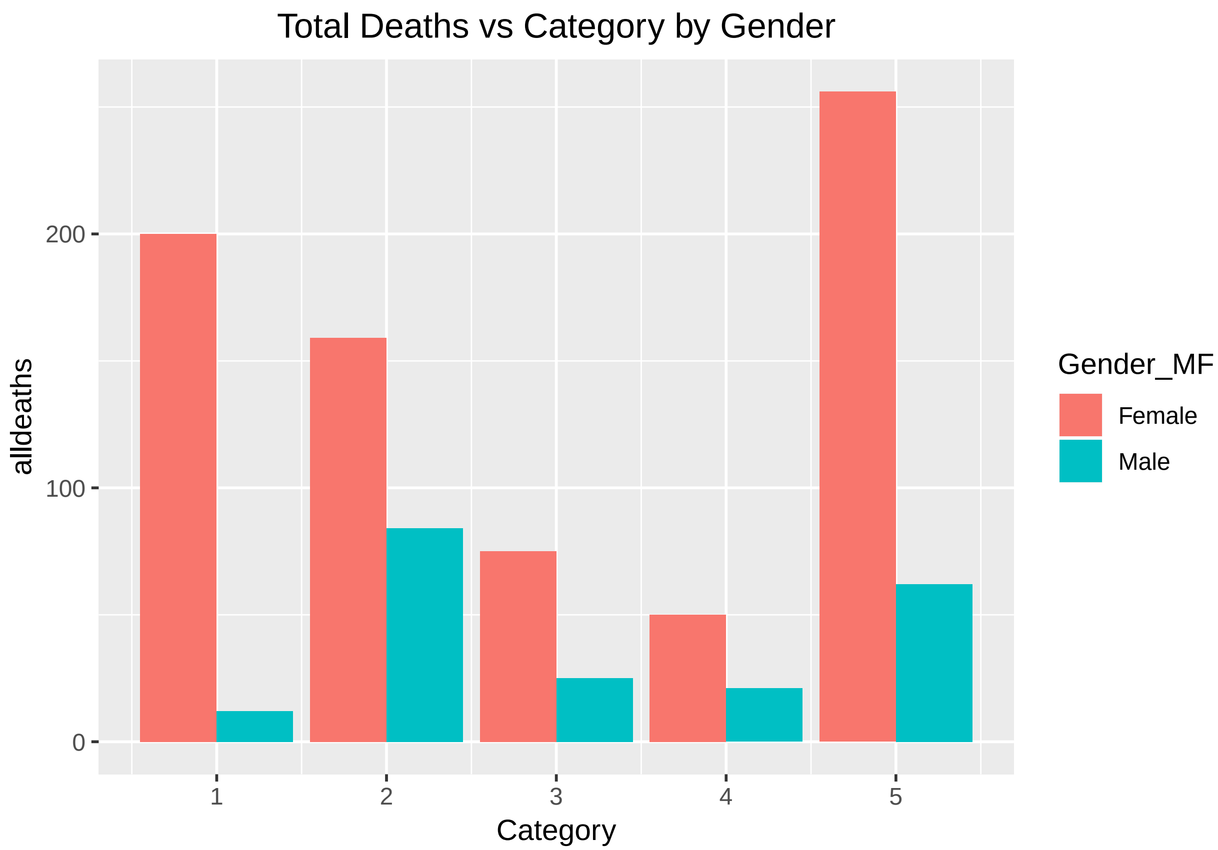
```
t.test(HurricanesF$NDAM, HurricanesM$NDAM,
       alternative = "two.sided", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  HurricanesF$NDAM and HurricanesM$NDAM
## t = -0.16984, df = 54.82, p-value = 0.8658
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6404.648  5403.973
## sample estimates:
## mean of x mean of y
##  7106.629  7606.967
```

Here we have a p-value of **0.9**, indicating no significant difference between the normalized damage for male and female hurricanes.

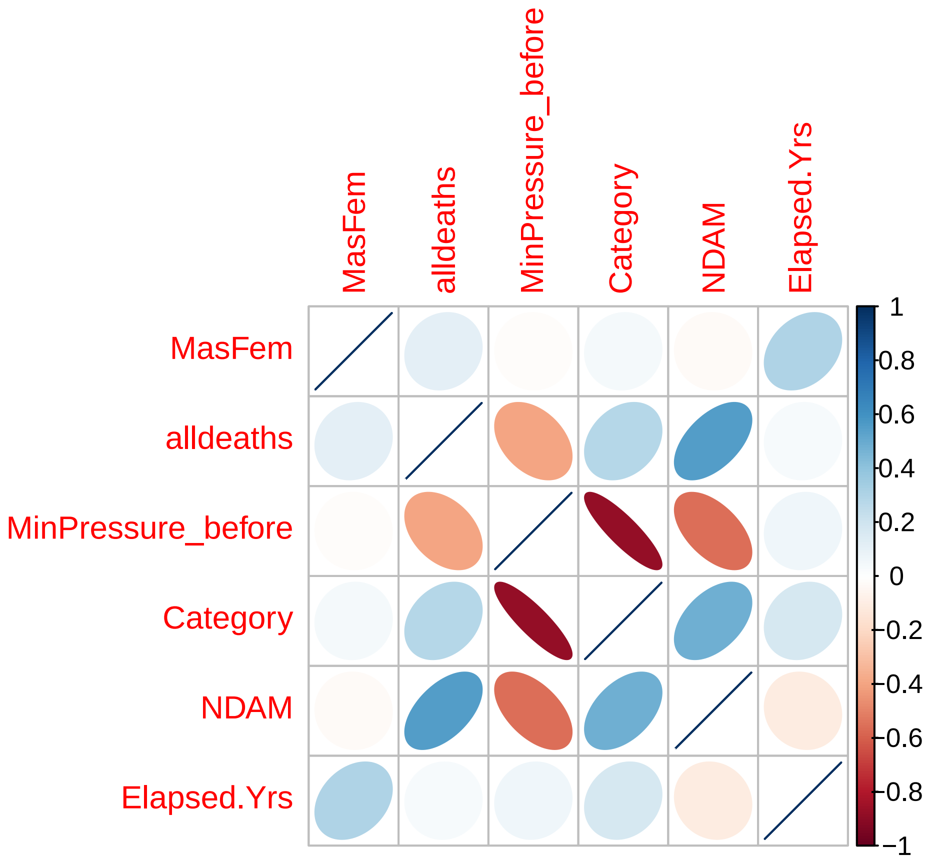
Bar Graph:

Let's also view a bar graph that breaks up hurricanes by category for male and female:



We can see that no matter the category of hurricane, there were more deaths for female named hurricanes.

Correlation Plot:



This correlation plot shows the strength of correlation between variables. We can see that Normalized Damage and Total deaths are positively correlated. So is Normalized Damage and the category of the hurricane. Category and Min Pressure are strongly negatively correlated. Normalized Damage and Min Pressure are moderately negatively correlated. Notice that the Masculine-Feminine Index does not show a strong correlation with any of the variables.

Problem 3

Table S1. Means, SDs, and intercorrelations of key variables

	Mean	SD	Mas-fem index (MFI)	Deaths (total)	Minimum pressure	Category	Normalized damage
Mas-fem index (MFI)	6.781	3.227					
Deaths (total)	20.652	40.904	0.11				
Minimum pressure	964.902	19.369	-0.016	-0.394			
Category	2.087	1.055	0.047	0.281	-0.875		
Normalized damage	7269.783	12934.087	-0.029	0.555	-0.556	0.481	
Years elapsed	30.913	18.771	0.306	0.032	0.067	0.173	-0.102

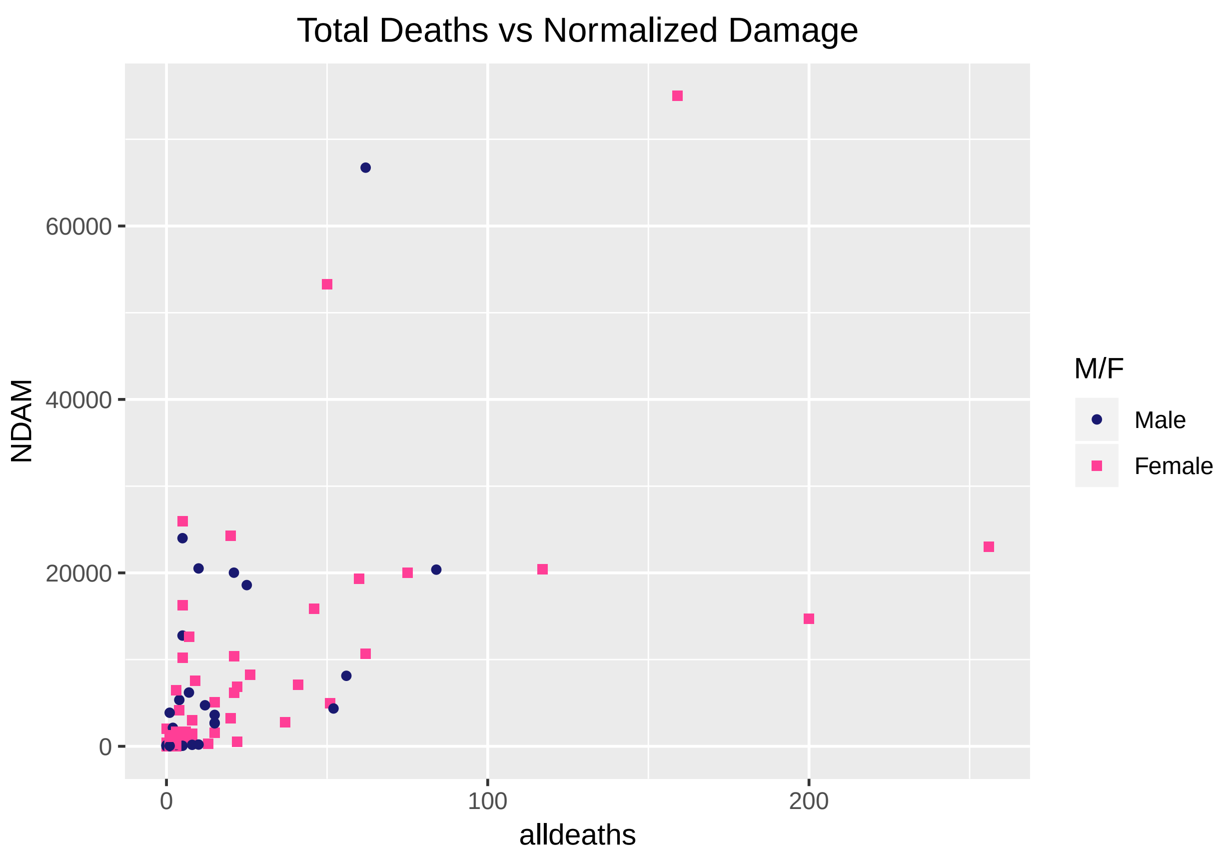
Table S2. Statistical summary of archival study (outcome variable: total deaths)

	Model 1		Model 2		Model 3		Model 4	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Minimum pressure	-0.028	0.0076	-0.017	0.008	-0.071	0.0192	-0.552	0.1503
Normalized damage	-	-	1e-04	1e-05	-5e-05	3e-05	0.863	0.1445
Mas-fem index(MFI)	-	-	0.04	0.04	-6.163	2.3629	0.172	0.1238
MFI x minimum pressure	-	-	-	-	0.006	0.0024	0.395	0.1521
MFI x normaliazed damage	-	-	-	-	2e-05	4e-06	0.705	0.1501
Goodness of fit (Pearson chi^2/df)	3.41	NA	1.53	NA	1.094	NA	1.094	NA
Liklehood ratio chi^2	17.529	NA	49.31	NA	60.565	NA	60.565	NA

Appendix: Code

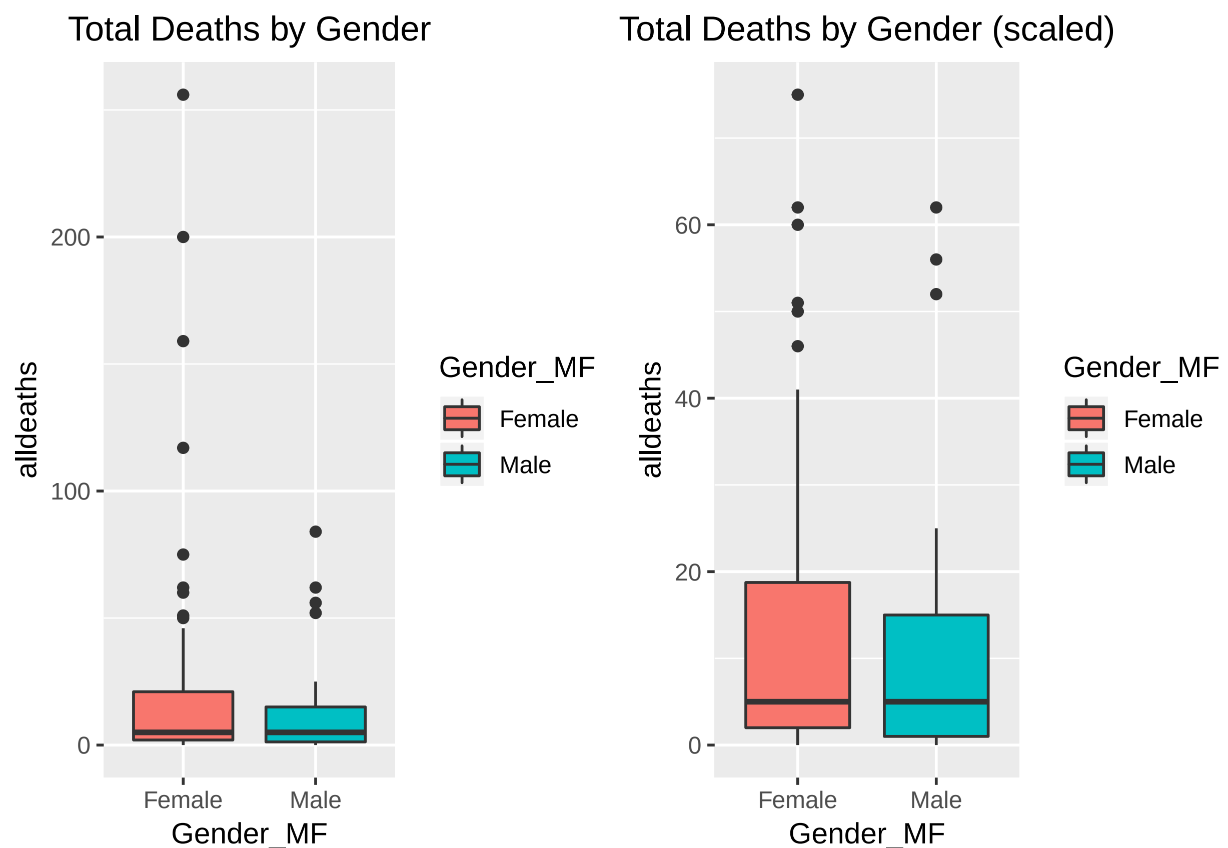
Problem 2 Code

```r
Hurricanes <- read.csv("D:/School/Spring 2019/Stat 472/Hurricanes/HurricaneDat.csv",
                       header=TRUE)
Hurricanes <- Hurricanes[,-11] # Unnecessary column
Hurricanes$Gender_MF <- as.character(Hurricanes$Gender_MF)
library(ggplot2)
library(MASS)

ggplot(Hurricanes, aes(x=alldeaths, y=NDAM, color=Gender_MF, shape=Gender_MF)) +
  geom_point() +
  scale_color_manual(name= "M/F", labels = c("Male", "Female"),
                     values = c("0" = "midnightblue", "1" = "violetred1")) +
  scale_shape_manual(name="M/F", labels=c("Male", "Female"), values=c("0"=16,"1"=15)) +
  ggtitle("Total Deaths vs Normalized Damage")+
  theme(plot.title = element_text(hjust = 0.5))
```

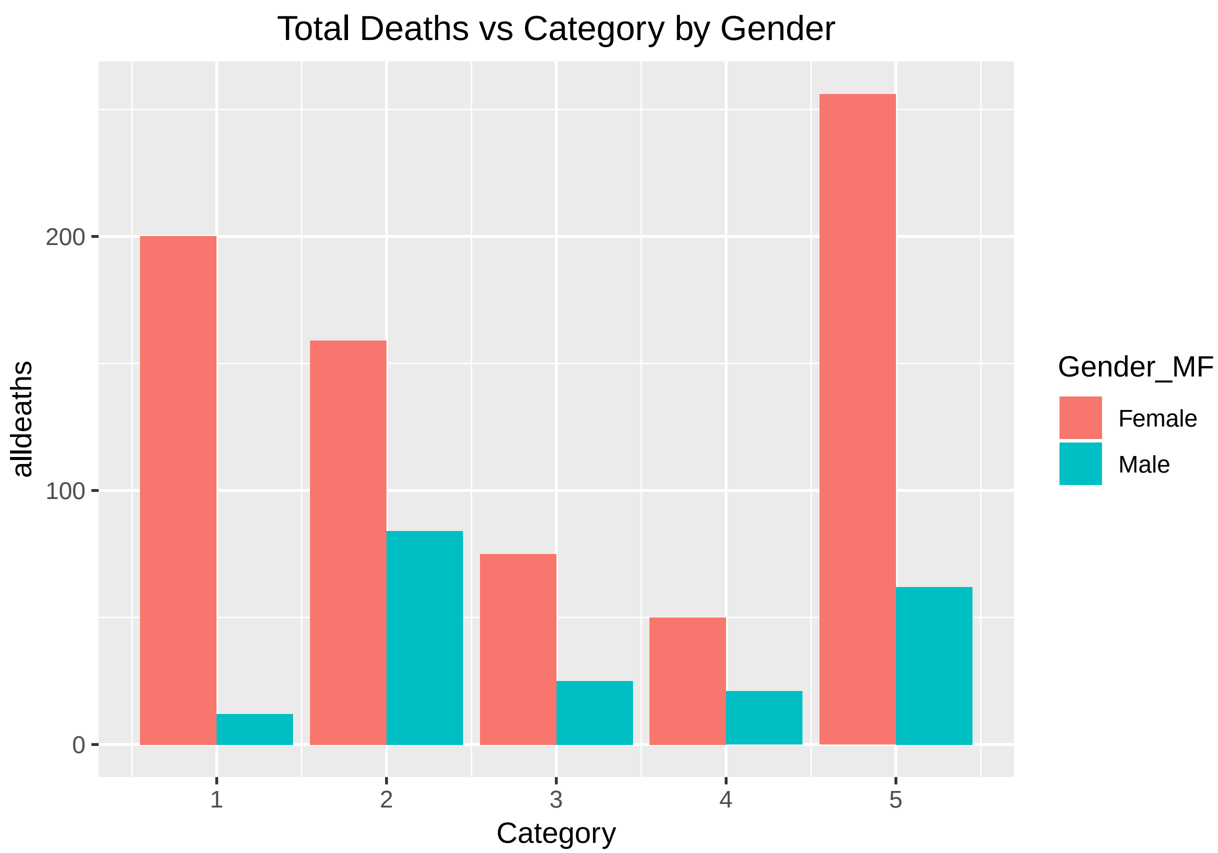



```r
Hurricanes$Gender_MF <- ifelse(Hurricanes$Gender_MF == "0", "Male", "Female")
require(gridExtra)
plot1 <- ggplot(Hurricanes, aes(x=Gender_MF, y=alldeaths, fill=Gender_MF)) +
  geom_boxplot() +
  ggtitle("Total Deaths by Gender")+
  theme(plot.title = element_text(hjust = 0.5))
```

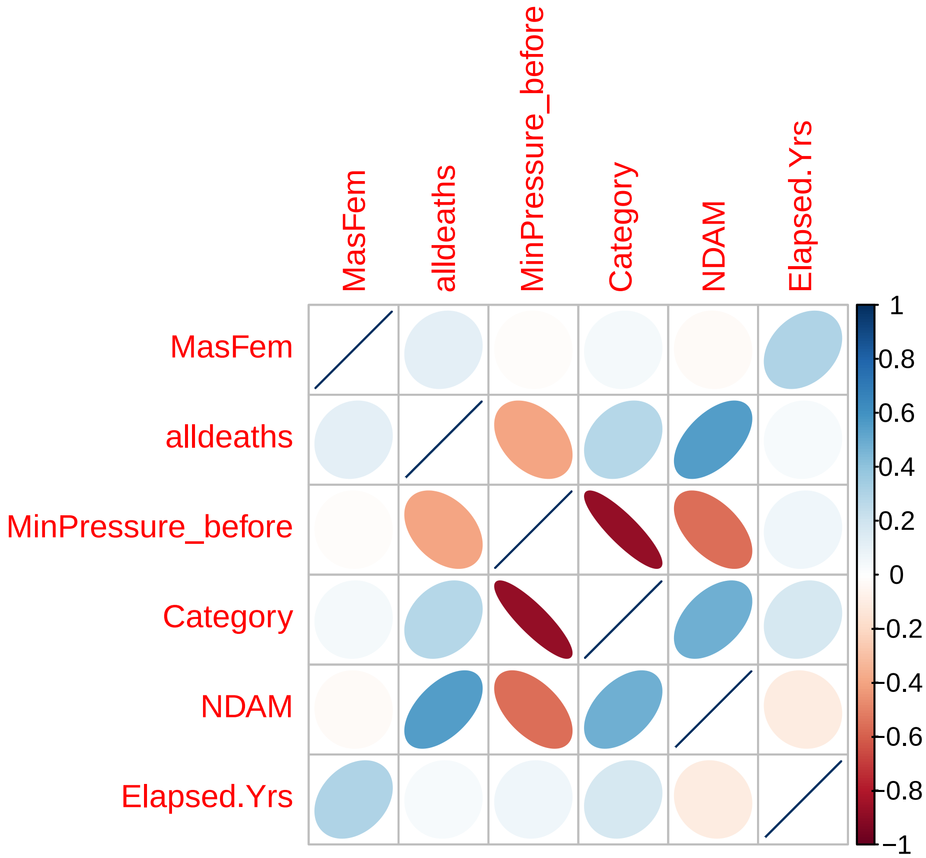
```r
plot2 <- ggplot(Hurricanes, aes(x=Gender_MF, y=alldeaths, fill=Gender_MF)) +
  geom_boxplot() + ylim(0, 75) +
  ggtitle("Total Deaths by Gender (scaled)") +
  theme(plot.title = element_text(hjust = 0.5))
grid.arrange(plot1, plot2, ncol=2)
```




```r
# Making a data set for female named hurricanes and male named hurricanes.
HurricanesF <- Hurricanes[Hurricanes$Gender_MF == "Female",]
HurricanesM <- Hurricanes[Hurricanes$Gender_MF == "Male",]

summary(HurricanesF$alldeaths); summary(HurricanesM$alldeaths)

t.test(HurricanesF$alldeaths, HurricanesM$alldeaths,
       alternative = "two.sided", var.equal = FALSE)

t.test(HurricanesF$NDAM, HurricanesM$NDAM,
       alternative = "two.sided", var.equal = FALSE)

ggplot(Hurricanes, aes(x=Category, y=alldeaths,
                       fill = Gender_MF)) +
  geom_bar(stat="identity",position='dodge') +
  ggtitle("Total Deaths vs Category by Gender")+
  theme(plot.title = element_text(hjust = 0.5))
```




```r
S1dat <- Hurricanes[,c(3,8,4,7,9,10)]
res <- cor(S1dat)

library(corrplot)
corrplot(res, method = "ellipse")
```



Problem 3 Code

Figure 1

```
S1dat <- Hurricanes[,c(3,8,4,7,9,10)]
sumdat <- data.frame(matrix(ncol=2,nrow=6))
for(i in 1:length(S1dat)){
  sumdat[i,1] <- round(mean(S1dat[,i]),3)
  sumdat[i,2] <- round(sd(S1dat[,i]),3)
}

res <- round(as.data.frame(res),3)
S1copy <- cbind(sumdat, res); S1copy <- S1copy[,-8]
S1copy[1,3:7] <- ""; S1copy[2,4:7] <- ""; S1copy[3,5:7] <- ""; S1copy[4,6:7] <- ""
S1copy[5,7] <- ""

rownames(S1copy) <- c("Mas-fem index (MFI)", "Deaths (total)", "Minimum pressure",
                      "Category", "Normalized damage", "Years elapsed")
colnames(S1copy) <- c("Mean", "SD", "Mas-fem index (MFI)", "Deaths (total)",
                      "Minimum pressure", "Category", "Normalized damage")

library(kableExtra)
library(dplyr)
```

```r
kable(S1copy, "latex", booktabs = T) %>%
  kable_styling(full_width = T, latex_options = "striped") %>%
  column_spec(1, width = "30mm")
```

Figure 2

```r
library(MASS)
BaseModel = glm.nb(alldeaths ~ 1, data=Hurricanes)

model1 = glm.nb(alldeaths ~ MinPressure_before, data=Hurricanes)
pres <- residuals(model1, "pearson")
# sum(pres^2)/model1$df.residual ## GOF
s1 = summary(model1)
library(lmtest)
y1 = lrtest(BaseModel, model1)

model2 <- glm.nb(alldeaths ~ MinPressure_before + NDAM + MasFem, data=Hurricanes)
s2 = summary(model2)
pres2 <- residuals(model2, "pearson")
# sum(pres2^2)/model2$df.residual ## GOF
y2 = lrtest(BaseModel, model2)

model3 <- glm.nb(alldeaths ~ MinPressure_before + NDAM + MasFem +
                   MasFem * MinPressure_before + MasFem * NDAM, data=Hurricanes)
s3 = summary(model3)
pres3 <- residuals(model3, "pearson")
# sum(pres3^2)/model3$df.residual ## GOF
y3 = lrtest(BaseModel, model3)

model4 <- glm.nb(alldeaths ~ ZMinPressure_A + ZNDAM + ZMasFem +
                   ZMasFem * ZMinPressure_A + ZMasFem * ZNDAM, data=Hurricanes)
s4 = summary(model4)
pres4 <- residuals(model4, "pearson")
# sum(pres4^2)/model4$df.residual ## GOF
y4 = lrtest(BaseModel, model4)

m11 = round(s1$coefficients[2,1],3)
m21 = round(s2$coefficients[2,1],3)
m31 = round(s3$coefficients[2,1],3)
m41 = round(s4$coefficients[2,1],3)
m22 = round(s2$coefficients[3,1],4)
m32 = round(s3$coefficients[3,1],5)
m42 = round(s4$coefficients[3,1],3)
m23 = round(s2$coefficients[4,1],3)
m33 = round(s3$coefficients[4,1],3)
m43 = round(s4$coefficients[4,1],3)
m34 = round(s3$coefficients[5,1],3)
m44 = round(s4$coefficients[5,1],3)
m35 = round(s3$coefficients[6,1],5)
m45 = round(s4$coefficients[6,1],3)

s11 = round(s1$coefficients[2,2],4)
```

```r
s21 = round(s2$coefficients[2,2],4)
s31 = round(s3$coefficients[2,2],4)
s41 = round(s4$coefficients[2,2],4)
s22 = round(s2$coefficients[3,2],5)
s32 = round(s3$coefficients[3,2],5)
s42 = round(s4$coefficients[3,2],4)
s23 = round(s2$coefficients[4,2],4)
s33 = round(s3$coefficients[4,2],4)
s43 = round(s4$coefficients[4,2],4)
s34 = round(s3$coefficients[5,2],4)
s44 = round(s4$coefficients[5,2],4)
s35 = round(s3$coefficients[6,2],6)
s45 = round(s4$coefficients[6,2],4)


sumdat2 <- matrix(ncol=8,nrow=7)
sumdat2[1,] <- c(m11,s11, m21,s21,m31,s31,m41,s41)
sumdat2[2,] <- c("-","-",m22,s22,m32,s32,m42,s42)
sumdat2[3,] <- c("-","-",m23,s23,m33,s33,m43,s43)
sumdat2[4,] <- c("-","-","-","-",m34,s34,m44,s44)
sumdat2[5,] <- c("-","-","-","-",m35,s35,m45,s45)
sumdat2[6,] <- c(round(sum(pres^2)/model1$df.residual,3),"NA",round(sum(pres2^2)/model2$df.residual,3),
                 "NA",round(sum(pres3^2)/model3$df.residual,3),"NA",round(sum(pres4^2)/model4$df.residua
sumdat2[7,] <- c(round(y1$Chisq[2],3),"NA", round(y2$Chisq[2],3),
                 "NA",round(y3$Chisq[2],3),"NA",round(y4$Chisq[2],3),"NA")
colnames(sumdat2) <- c("Mean","SE","Mean","SE","Mean","SE", "Mean","SE")
rownames(sumdat2) <- c("Minimum pressure","Normalized damage","Mas-fem index(MFI)",
                       "MFI x minimum pressure","MFI x normaliazed damage","Goodness of fit (Pearson chi
                       "Liklehood ratio chi^2")

kable(sumdat2, "latex", booktabs = T) %>%
  kable_styling(full_width = T, latex_options = "striped") %>%
  column_spec(1, width = "30mm") %>%
  add_header_above(c(" " = 1, "Model 1" = 2, "Model 2" = 2, "Model 3" = 2, "Model 4" = 2))
```

```r
model4 <- glm.nb(alldeaths ~ ZMinPressure_A + ZNDAM + ZMasFem + ZMasFem * ZMinPressure_A + ZMasFem * ZN
s4 = summary(model4)
pres4 <- residuals(model4, "pearson")
# sum(pres4^2)/model4$df.residual ## GOF
y4 = lrtest(BaseModel, model4)
```