Homework 10

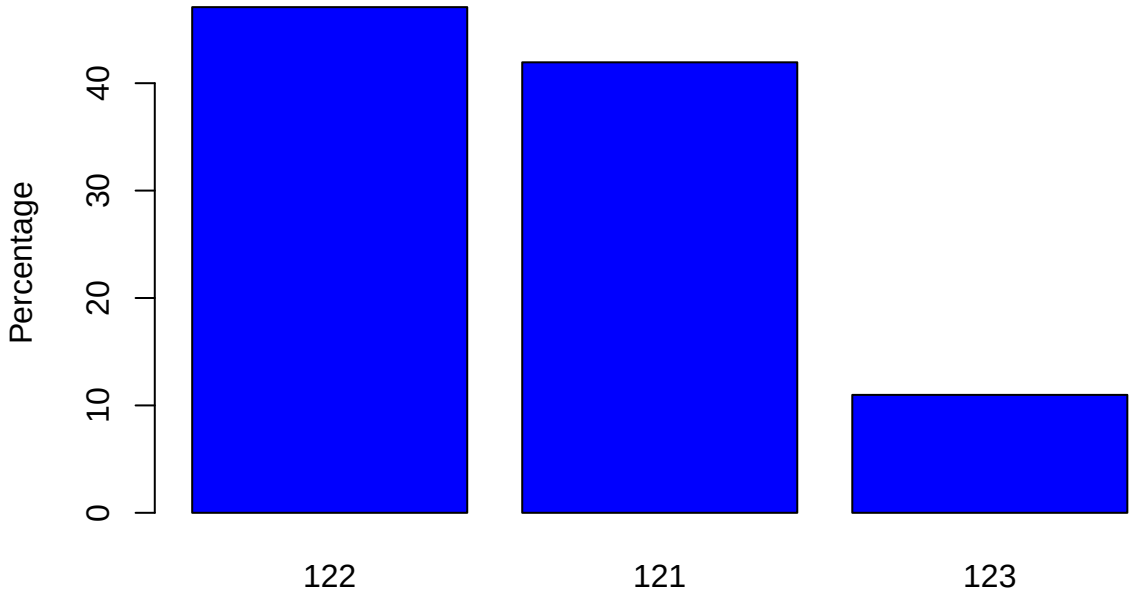
Li Zhang

11/10/2021

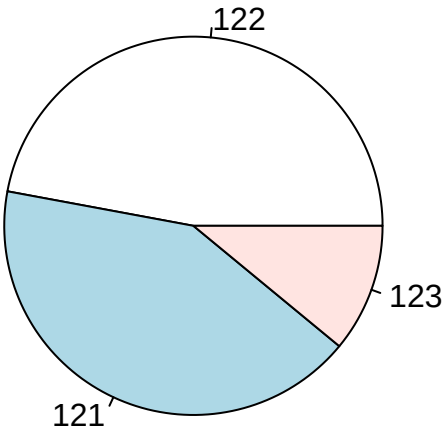
Exercise 5.2

Compute the count and percentage of each level of DRG by filling the blanks below. Use bar plot and pie chart similar to Figure 5.1 and Figure 5.2 to visualize. Briefly discuss your results.

```
# import data file heartatk4R.txt into R
heartatk4R <- read.table("~/Desktop/heartatk4R.txt",
                         header = TRUE,
                         sep = "\t",
                         colClasses = c("character", "factor", "factor", "factor",
                                        "factor", "numeric", "numeric", "numeric"))

drg <- heartatk4R$DRG #Import values of DRG
counts <- sort(table(drg), decreasing = TRUE) # tabulate & sort
percentages <- 100 * counts/length(drg) # Compute the percentage and convert to %
barplot(percentages, ylab = "Percentage", col = "blue") # barplot
```




```
pie(counts) # pie chart
```




```
## Discussion:
# DRG 121 stands for survivors with cardiovascular complication,
# DRG 122 stands for survivors without complications
# DRG 123 stands for patients who died.
# As we can see in the bar plot and pie chart,
# DRG 122 and 121 takes almost 90 % while 123 only takes about 10%.
# which means that there are small number of patients died and most of the patients are survived.
# The percentage of DRG 122-patients survive without complications is slightly greater
# than the percentage of DRG 121 patients survive with cardiovascular complications.
```

Exercise 5.3

Investigate the correlation between length of stay and charges by filling in the blanks below. Remember to include plain English interpretation of your results even your grandpa can understand.

```
LOS <- heartatk4R$LOS # import the values of LOS in the heart attack data

CHARGES <- heartatk4R$CHARGES #import the values of Charges

cor.test(LOS, CHARGES, method = "pearson") # Pearson correlation
```

```
##
##  Pearson's product-moment correlation
##
## data:  LOS and CHARGES
## t = 120.86, df = 12143, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7307703 0.7469202
## sample estimates:
##       cor
## 0.7389514
```

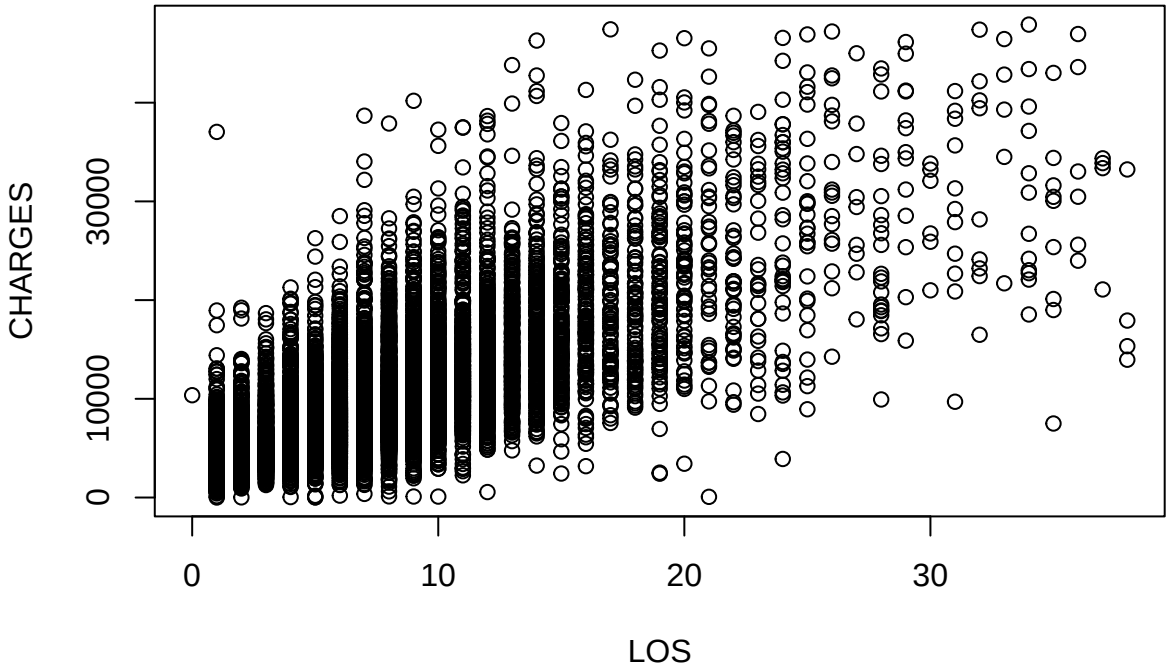
```
cor.test(LOS, CHARGES, method = "spearman") # spearman correlation
```

```
## Warning in cor.test.default(LOS, CHARGES, method = "spearman"): Cannot compute
## exact p-value with ties
```

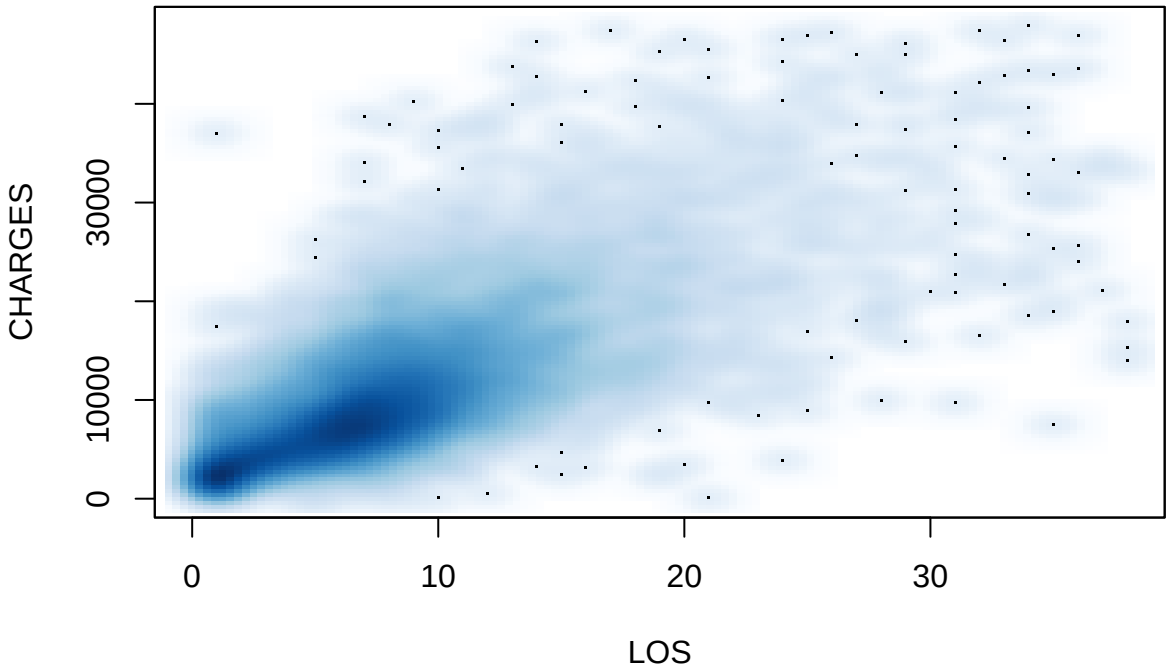
```
##
##  Spearman's rank correlation rho
##
## data:  LOS and CHARGES
```

```
## S = 7.8102e+10, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.7384085
```

```
plot(LOS, CHARGES) # standard scatter plot
```




```
smoothScatter(LOS, CHARGES) # a smoothed color density representation of a scatterplot
```



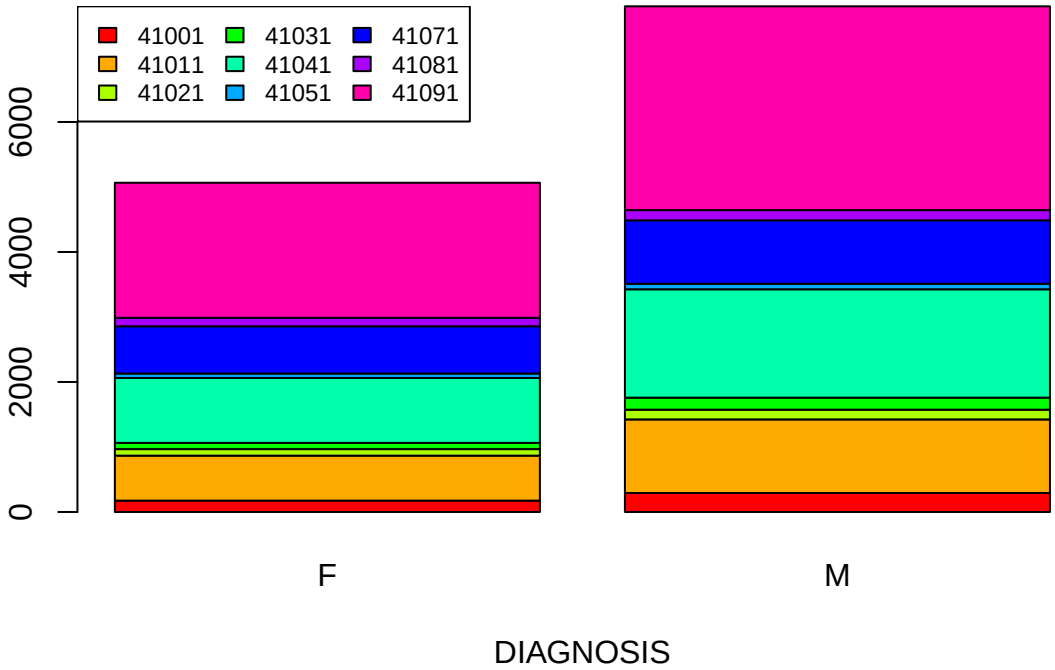

```
# There is a large, and statistically significant between Length of stay in hospital and charge.
# The plain English interpretation is this:
```

```
# the longer patients stay in the hospital, the more charge they will have.
```

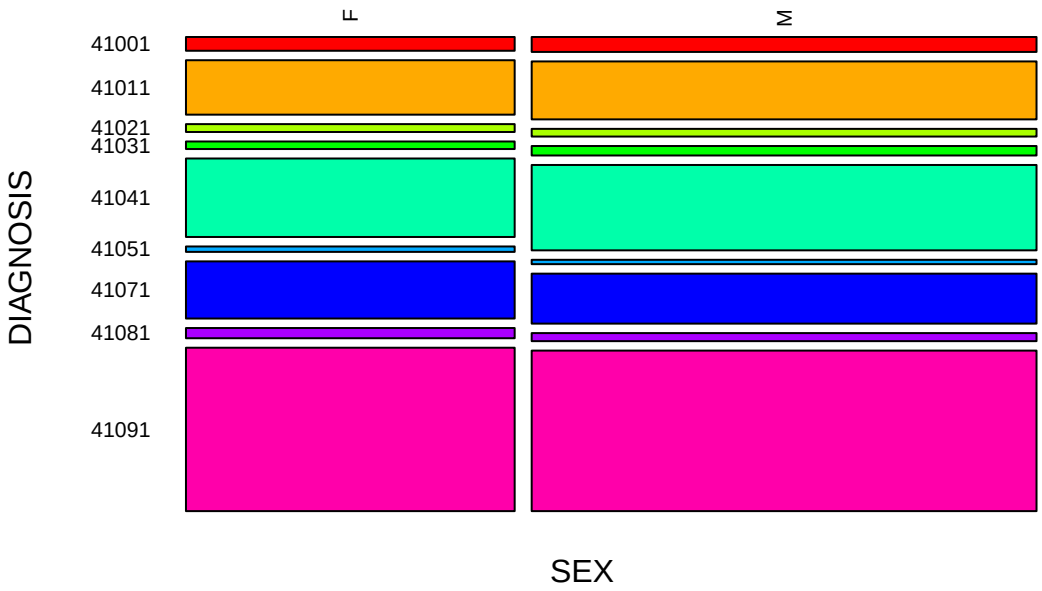
Exercise 5.4

The DIAGNOSIS column contains IDC codes that specifies the part of the heart that are affected. Use a stacked bar plot and a mosaic plot to compare the difference in frequency of DIAGNOSIS between men and women. Based on your observation, do you think men and women are equal in their frequencies of diagnoses?

```
DIAGNOSIS <- heartatk4R$DIAGNOSIS # import the values of DIAGNOSIS in the heart attack data

SEX <- heartatk4R$SEX  # import the values of SEX in the heart attack data

counts <- table(DIAGNOSIS, SEX) # Take SEX as columns

barplot(counts, xlab = "DIAGNOSIS", col = rainbow(9), beside = F)
legend("topleft", legend = rownames(counts), fill = rainbow(9), ncol = 3, cex = 0.75)
```




```
mosaicplot(table(SEX, DIAGNOSIS), color = rainbow(9),las = 2)
```

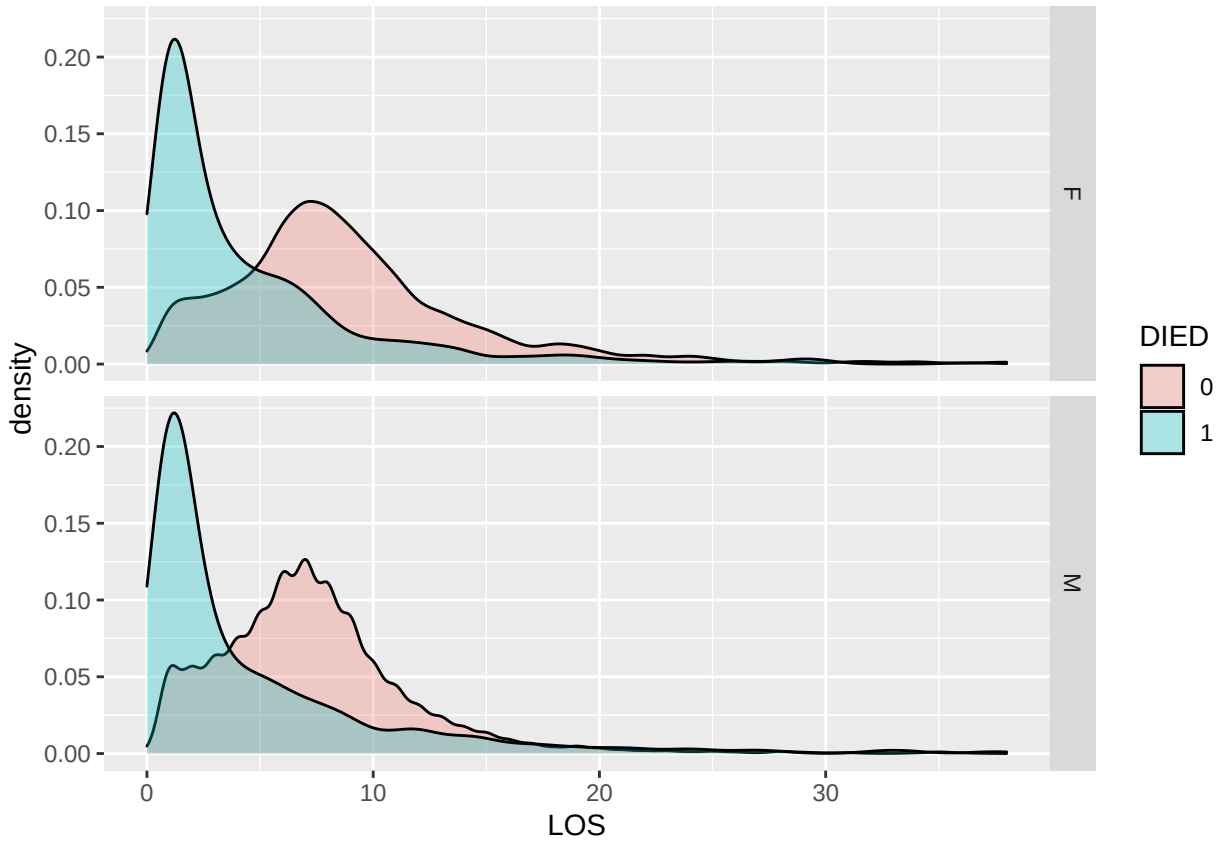



```
# Yes, I think that men and woman are equal in their frequencies of diagnoses.
```

Exercise 5.6

Use the ggplot2 package to compare the distribution of lengths of stay among patients who survived and those who did not, but compare men and women separately (similar to Figure 5.11).

```
library(ggplot2)    # use the ggplot package

ggplot(heartatk4R, aes(x = LOS, fill = DIED)) + geom_density(alpha = .3) + facet_grid(SEX~ .)
```




```
# Female and male patients who did not survive heart attack tend to have shorter LOS.
```