Final Report

due November 16, 2021 by 11:59 PM

Shelby Brown, Katie Lam, Kaeden Hill

10/31/21

Load Packages

```
library(tidyverse)
```

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
```

Load Data

```
heart <- readr::read_csv("heart.csv")
```

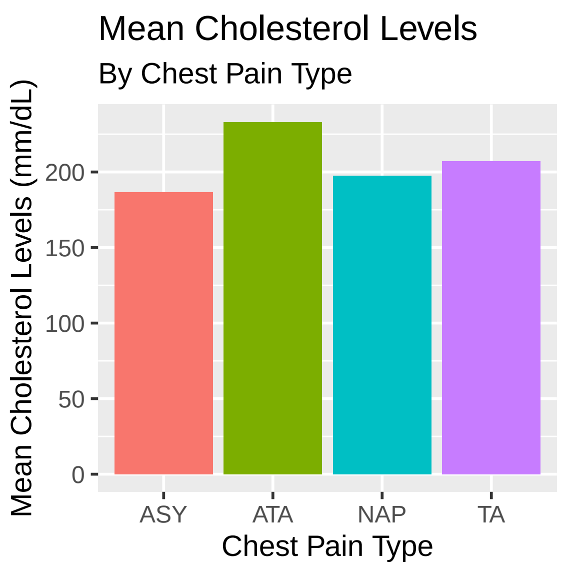
Data Analysis Plan

```
mean_cholesterol <- heart %>%
  group_by(ChestPainType) %>%
  summarize(mean_cholesterol = mean(Cholesterol))%>%
  print()
```

```
## # A tibble: 4 x 2
```

```
##   ChestPainType mean_cholesterol
##   <chr>                    <dbl>
## 1 ASY                       187.
## 2 ATA                       233.
## 3 NAP                       197.
## 4 TA                        207.
```

```
mean_cholesterol %>%
  ggplot()+
  geom_col(mapping = aes(x = ChestPainType, y = mean_cholesterol,  fill = ChestPainType), position = "d
    theme(legend.position = "none")+
    labs(title = "Mean Cholesterol Levels",
         subtitle = "By Chest Pain Type",
         x = "Chest Pain Type",
         y = "Mean Cholesterol Levels (mm/dL)")
```



Grouping Variables

```
heart_grouped <- heart %>%
  mutate(chol_level = cut(Cholesterol,
                          breaks = c(-Inf, 120, 200, 239, Inf),
                          labels = c("Low", "Normal", "Intermediate", "High"),
                          right=FALSE))
```

```
heart_grouped <- heart_grouped %>%
  mutate(press_level = cut(RestingBP,
                          breaks = c(-Inf, 120, 130, 140, Inf),
                          labels = c("Normal", "Elevated", "Hypertension 1", "Hypertension 2"),
                          right=FALSE))
```

Counting Variables

```r
heart_grouped <- heart_grouped %>%
  mutate(Sex, sex_factor=ifelse(Sex=="M", 0,1)) %>%
  mutate(ExerciseAngina, exer_factor=ifelse(ExerciseAngina=="N", 0,1))
```

```r
sex_grouped <- heart_grouped %>%
  group_by(ChestPainType, sex_factor) %>%
  summarize(count = n())
```

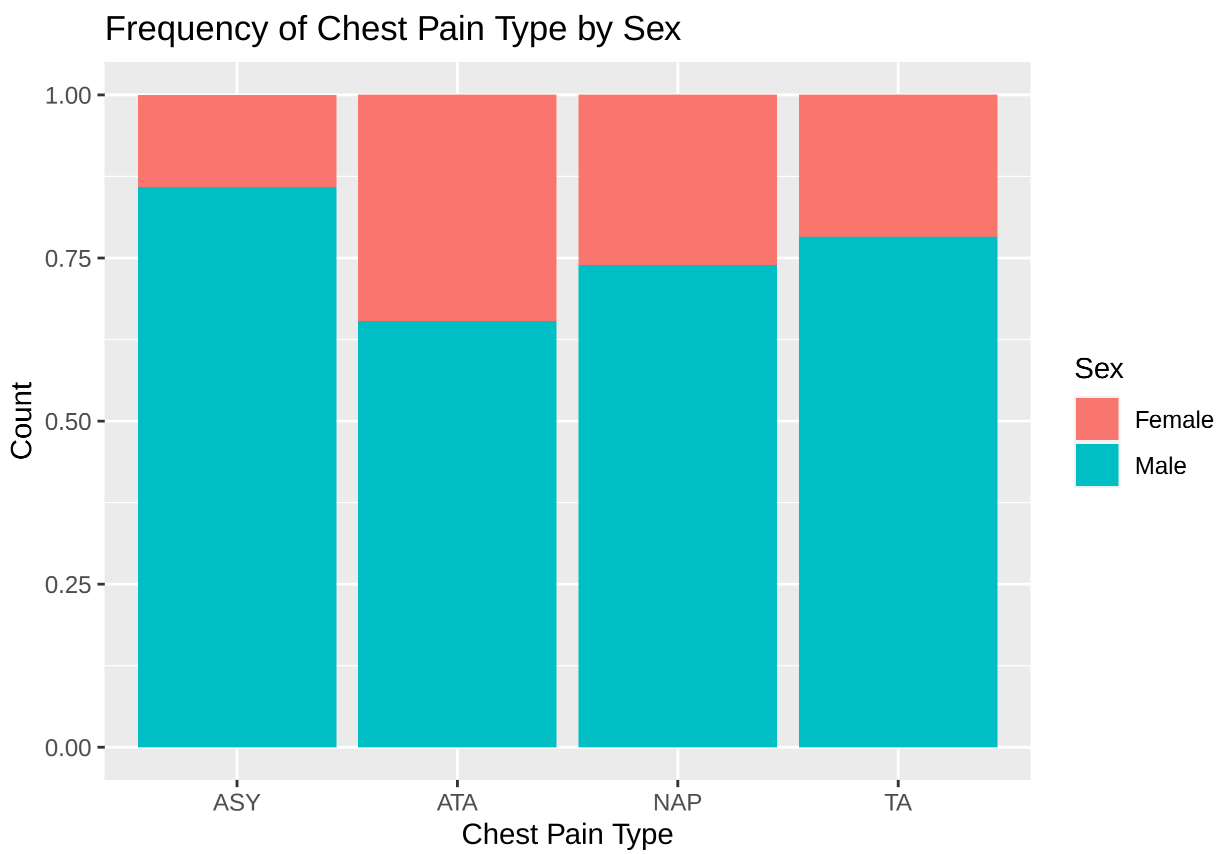
```r
exer_grouped <- heart_grouped %>%
  group_by(ChestPainType, exer_factor) %>%
  summarize(count = n())
```

```r
RBP_grouped <- heart_grouped %>%
  group_by(ChestPainType, press_level) %>%
  summarize(count = n())
```

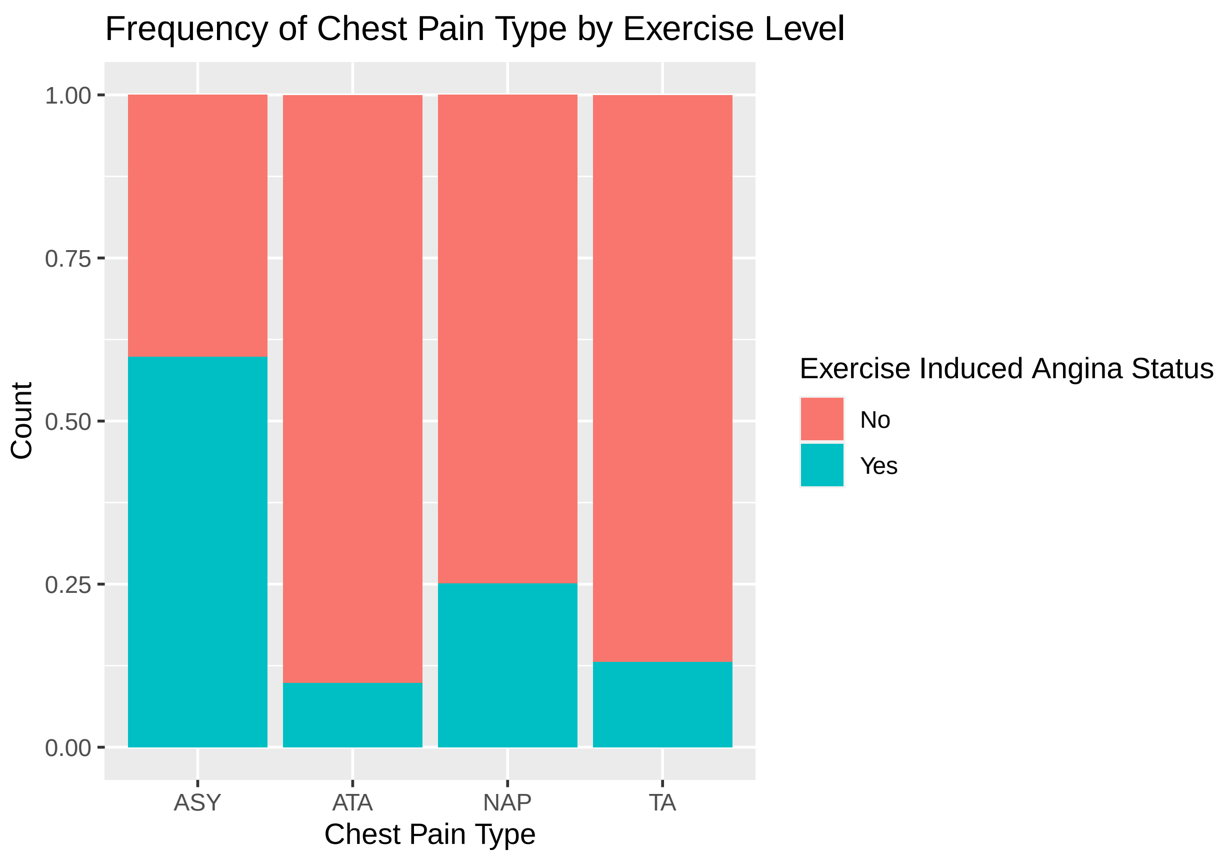
```r
chol_grouped <- heart_grouped %>%
  group_by(ChestPainType, chol_level)%>%
  summarize(count = n())
```

Visualizing Data

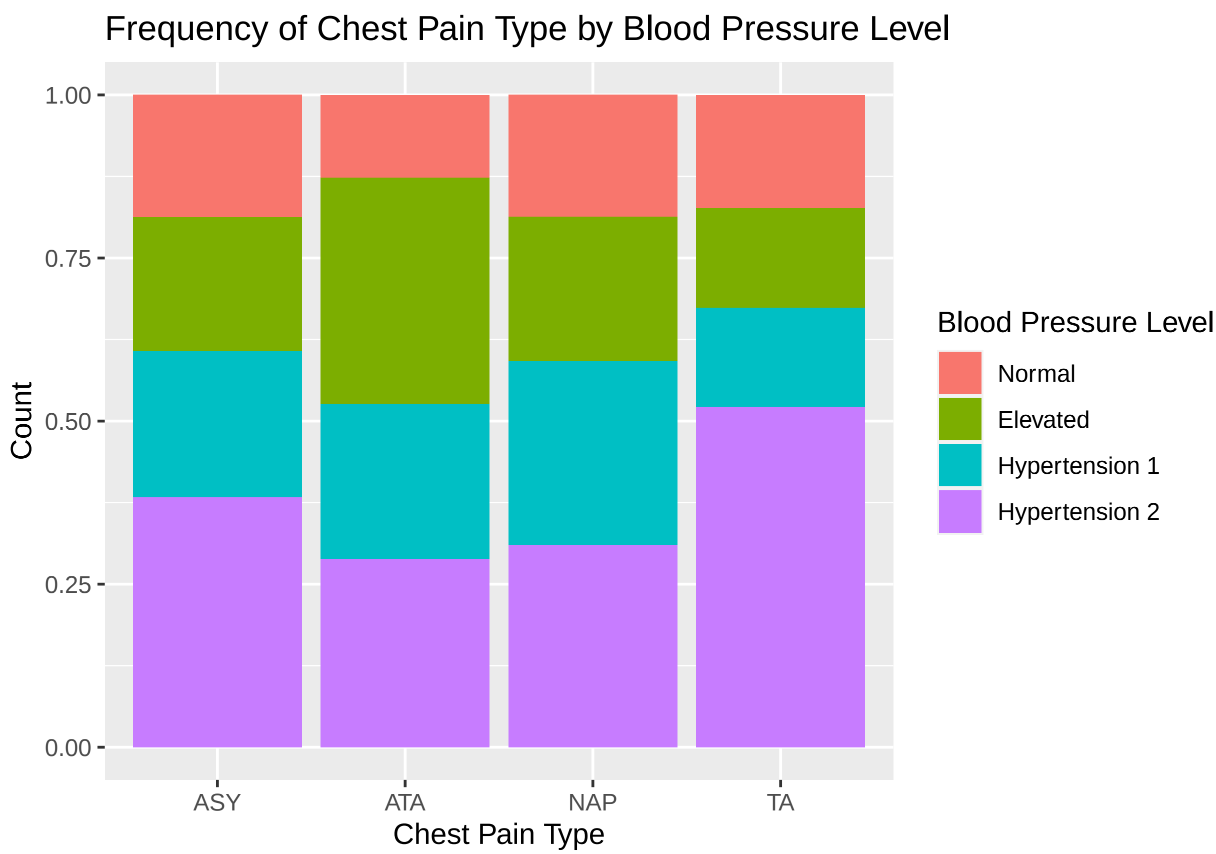
```r
sex_grouped %>%
  mutate(sex_factor, sex_name=ifelse(sex_factor==0, "Male","Female")) %>%
  ggplot()+
  geom_col(aes(x = ChestPainType, y = count, fill = as.factor(sex_name)), position = "fill") +
  labs(title = "Frequency of Chest Pain Type by Sex",
       x = "Chest Pain Type",
       y = "Count",
       fill = "Sex")
```



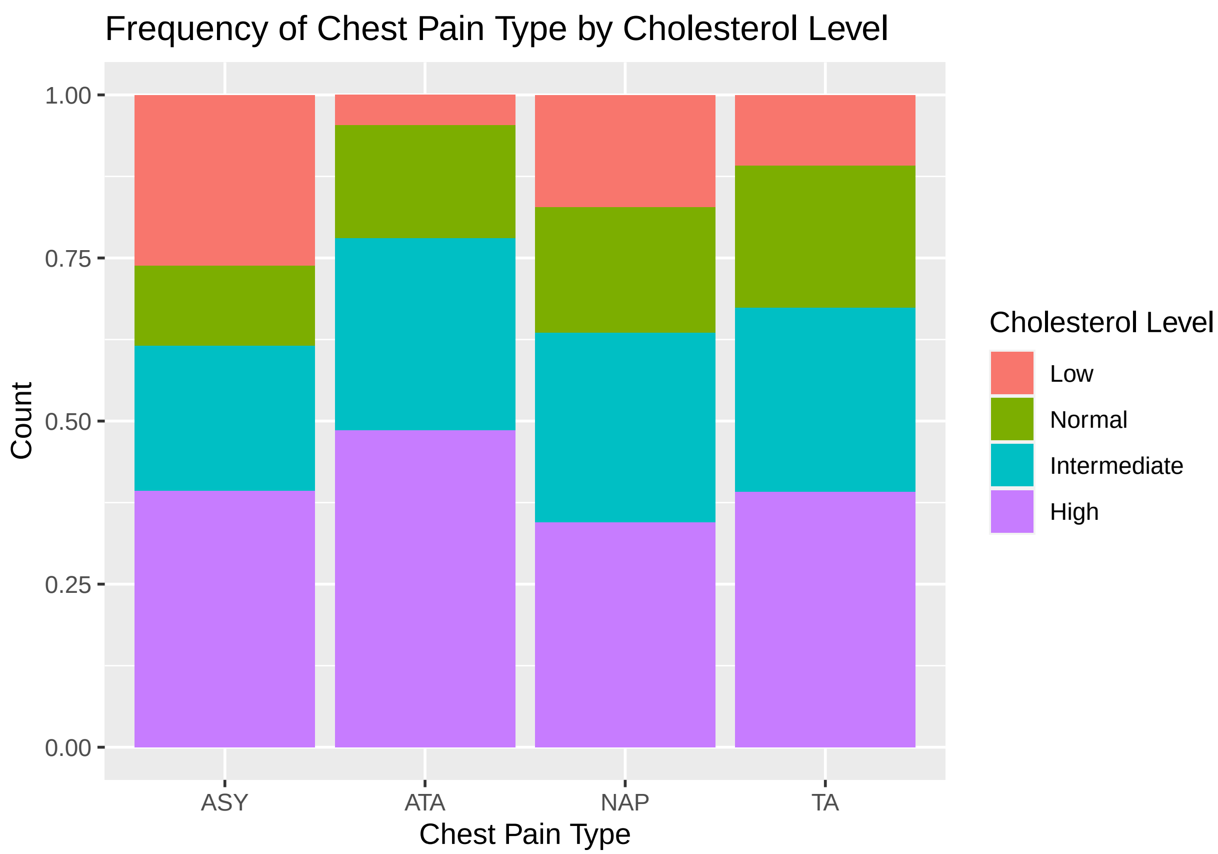

```
exer_grouped %>%
  mutate(exer_factor, exer_cat=ifelse(exer_factor==0, "No","Yes")) %>%
  ggplot()+
  geom_col(aes(x = ChestPainType, y = count, fill = as.factor(exer_cat)), position = "fill") +
  labs(title = "Frequency of Chest Pain Type by Exercise Level",
       x = "Chest Pain Type",
       y = "Count",
       fill = "Exercise Induced Angina Status")
```




```r
RBP_grouped %>%
  ggplot()+
  geom_col(aes(x = ChestPainType, y = count, fill = as.factor(press_level)), position = "fill") +
  labs(title = "Frequency of Chest Pain Type by Blood Pressure Level",
       x = "Chest Pain Type",
       y = "Count",
       fill = "Blood Pressure Level")
```




```
chol_grouped %>%
  ggplot()+
  geom_col(aes(x = ChestPainType, y = count, fill = as.factor(chol_level)), position = "fill") +
  labs(title = "Frequency of Chest Pain Type by Cholesterol Level",
       x = "Chest Pain Type",
       y = "Count",
       fill = "Cholesterol Level")
```



Statistical Tests

```
sex_table <- heart_grouped %>%
  tabyl(Sex, ChestPainType)

exer_table <- heart_grouped %>%
  tabyl(ExerciseAngina, ChestPainType)

RBP_table <- heart_grouped %>%
  tabyl(press_level, ChestPainType)

chol_table <- heart_grouped %>%
  tabyl(chol_level, ChestPainType)
```

```
chisq.test(sex_table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  sex_table
## X-squared = 36.879, df = 3, p-value = 4.88e-08
```

```
chisq.test(exer_table)
```

```
##
##  Pearson's Chi-squared test
```

```
##
## data:  exer_table
## X-squared = 179.27, df = 3, p-value < 2.2e-16
```

```
chisq.test(chol_table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  chol_table
## X-squared = 49.409, df = 9, p-value = 1.391e-07
```

```
chisq.test(RBP_table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  RBP_table
## X-squared = 26.829, df = 9, p-value = 0.001493
```

```
heart_grouped2 <- heart_grouped %>%
  filter(ChestPainType %in% c("ATA", "ASY"))

chol_step <- heart_grouped2 %>%
  tabyl(chol_level, ChestPainType)
fisher.test(chol_step)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  chol_step
## p-value = 8.166e-10
## alternative hypothesis: two.sided
```

```
heart_grouped3 <- heart_grouped %>%
  filter(ChestPainType %in% c("ASY", "NAP"))

chol_step2 <- heart_grouped3 %>%
  tabyl(chol_level, ChestPainType)
fisher.test(chol_step2)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  chol_step2
## p-value = 0.003813
## alternative hypothesis: two.sided
```

```
heart_grouped4 <- heart_grouped %>%
  filter(ChestPainType %in% c("TA", "ASY"))

chol_step3 <- heart_grouped4 %>%
  tabyl(chol_level, ChestPainType)
fisher.test(chol_step3)
```

```
##
##  Fisher's Exact Test for Count Data
##
```

```
## data:  chol_step3
## p-value = 0.04389
## alternative hypothesis: two.sided
```

```
heart_grouped5 <- heart_grouped %>%
  filter(ChestPainType %in% c("ATA", "NAP"))

chol_step4 <- heart_grouped5 %>%
  tabyl(chol_level, ChestPainType)
fisher.test(chol_step4)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  chol_step4
## p-value = 0.0003404
## alternative hypothesis: two.sided
```

```
heart_grouped6 <- heart_grouped %>%
  filter(ChestPainType %in% c("ATA", "TA"))

chol_step5 <- heart_grouped6 %>% #NOT SIGNIFICANT; used fisher bc count was <10 for low:TA
  tabyl(chol_level, ChestPainType)
fisher.test(chol_step5)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  chol_step5
## p-value = 0.3106
## alternative hypothesis: two.sided
```

```
heart_grouped7 <- heart_grouped %>%
  filter(ChestPainType %in% c("NAP", "TA"))

chol_step6 <- heart_grouped7 %>% #NOT SIGNIFICANT
  tabyl(chol_level, ChestPainType)
fisher.test(chol_step6)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  chol_step6
## p-value = 0.7484
## alternative hypothesis: two.sided
```

```
sex_step <- heart_grouped2 %>%
  tabyl(Sex, ChestPainType)
fisher.test(sex_step)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  sex_step
## p-value = 2.369e-08
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
```

```
##  0.2031800 0.4733313
## sample estimates:
## odds ratio
##  0.3100921
```

```r
sex_step2 <- heart_grouped3 %>%
  tabyl(Sex, ChestPainType)
fisher.test(sex_step2)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  sex_step2
## p-value = 0.0002766
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.3054468 0.7119609
## sample estimates:
## odds ratio
##  0.4656136
```

```r
sex_step3 <- heart_grouped4 %>% #NOT SIGNIFICANT
  tabyl(Sex, ChestPainType)
fisher.test(sex_step3)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  sex_step3
## p-value = 0.1896
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.2722762 1.4001345
## sample estimates:
## odds ratio
##  0.5922146
```

```r
sex_step4 <- heart_grouped5 %>% #NOT SIGNIFICANT
  tabyl(Sex, ChestPainType)
fisher.test(sex_step4)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  sex_step4
## p-value = 0.07251
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.9418551 2.3984968
## sample estimates:
## odds ratio
##   1.501076
```

```r
sex_step5 <- heart_grouped6 %>% #NOT SIGNIFICANT
  tabyl(Sex, ChestPainType)
fisher.test(sex_step5)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  sex_step5
## p-value = 0.1107
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8524745 4.6148967
## sample estimates:
## odds ratio
##   1.906248
```

```
sex_step6 <- heart_grouped7 %>% #NOT SIGNIFICANT
  tabyl(Sex, ChestPainType)
fisher.test(sex_step6)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  sex_step6
## p-value = 0.7073
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.5679163 3.0759273
## sample estimates:
## odds ratio
##   1.270772
```

```
exer_step <- heart_grouped2 %>%
  tabyl(ExerciseAngina, ChestPainType)
fisher.test(exer_step)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  exer_step
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.04032086 0.12583163
## sample estimates:
## odds ratio
## 0.07331366
```

```
exer_step2 <- heart_grouped3 %>%
  tabyl(ExerciseAngina, ChestPainType)
fisher.test(exer_step2)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  exer_step2
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.1529646 0.3280057
```

```
## sample estimates:
## odds ratio
##  0.2253116
```

```
exer_step3 <- heart_grouped4 %>%
  tabyl(ExerciseAngina, ChestPainType)
fisher.test(exer_step3)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  exer_step3
## p-value = 4.234e-10
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.03430591 0.24552456
## sample estimates:
## odds ratio
##   0.100903
```

```
exer_step4 <- heart_grouped5 %>%
  tabyl(ExerciseAngina, ChestPainType)
fisher.test(exer_step4)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  exer_step4
## p-value = 0.0001411
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.656297 5.935882
## sample estimates:
## odds ratio
##   3.070166
```

```
exer_step5 <- heart_grouped6 %>% #NOT SIGNIFICANT
  tabyl(ExerciseAngina, ChestPainType)
fisher.test(exer_step5)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  exer_step5
## p-value = 0.5883
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.4161352 3.9597702
## sample estimates:
## odds ratio
##   1.374316
```

```
exer_step6 <- heart_grouped7 %>% #NOT SIGNIFICANT
  tabyl(ExerciseAngina, ChestPainType)
fisher.test(exer_step6)
```

```
##
```

```
##  Fisher's Exact Test for Count Data
##
## data:  exer_step6
## p-value = 0.0836
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.1466831 1.1500616
## sample estimates:
## odds ratio
##  0.4483045
```

```
RBP_step <- heart_grouped2 %>%
  tabyl(press_level, ChestPainType)
fisher.test(RBP_step)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  RBP_step
## p-value = 0.001125
## alternative hypothesis: two.sided
```

```
RBP_step2 <- heart_grouped3 %>% #NOT SIG
  tabyl(press_level, ChestPainType)
fisher.test(RBP_step2)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  RBP_step2
## p-value = 0.2331
## alternative hypothesis: two.sided
```

```
RBP_step3 <- heart_grouped4 %>% #NOT SIG
  tabyl(press_level, ChestPainType)
fisher.test(RBP_step3)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  RBP_step3
## p-value = 0.34
## alternative hypothesis: two.sided
```

```
RBP_step4 <- heart_grouped5 %>%
  tabyl(press_level, ChestPainType)
fisher.test(RBP_step4)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  RBP_step4
## p-value = 0.0436
## alternative hypothesis: two.sided
```

```
RBP_step5 <- heart_grouped6 %>%
  tabyl(press_level, ChestPainType)
```

```
fisher.test(RBP_step5)
```

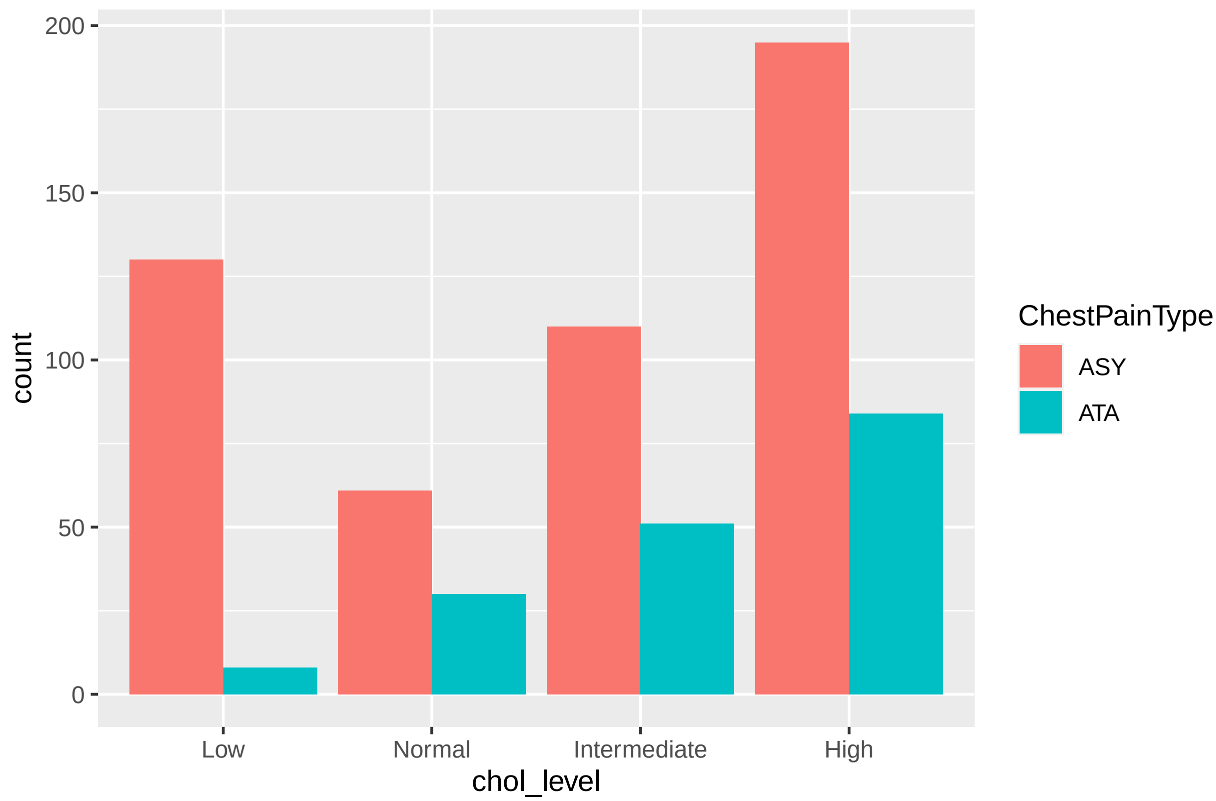
```
##
##  Fisher's Exact Test for Count Data
##
## data:  RBP_step5
## p-value = 0.006839
## alternative hypothesis: two.sided
```

```
RBP_step6 <- heart_grouped7 %>% #NOT SIG
  tabyl(press_level, ChestPainType)
fisher.test(RBP_step6)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  RBP_step6
## p-value = 0.05117
## alternative hypothesis: two.sided
```

Visualized Results

```
heart_grouped2 %>%
  ggplot()+
  geom_bar(mapping = aes(x = chol_level, fill = ChestPainType), position = "dodge")+
  labs(title = "")
```


200
150
100
50
0
count
Low
Normal
Intermediate
High
chol_level
ChestPainType
ASY
ATA