```
rm(list=ls())
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(readr)
library(vcd)
```

```
## Loading required package: grid
```

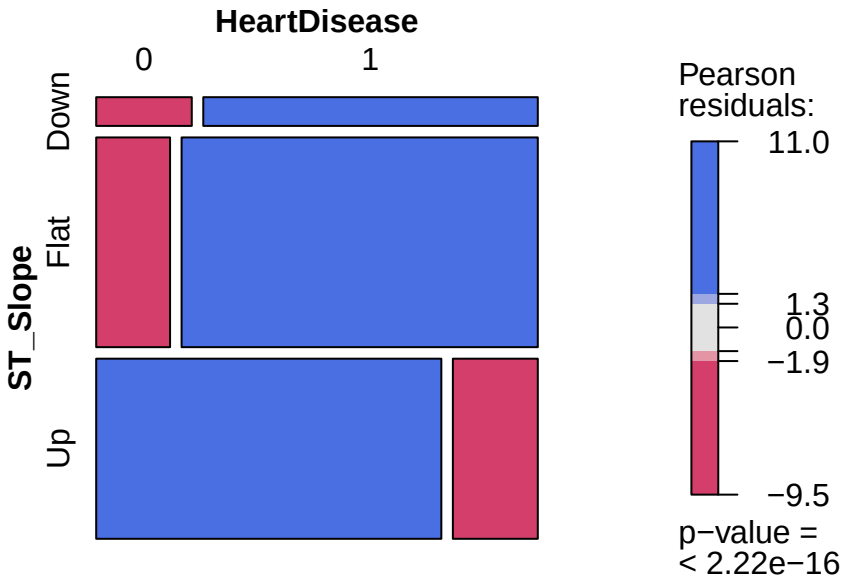
```
heart <- read_csv("heart.csv")
```

```
## Rows: 918 Columns: 12
```

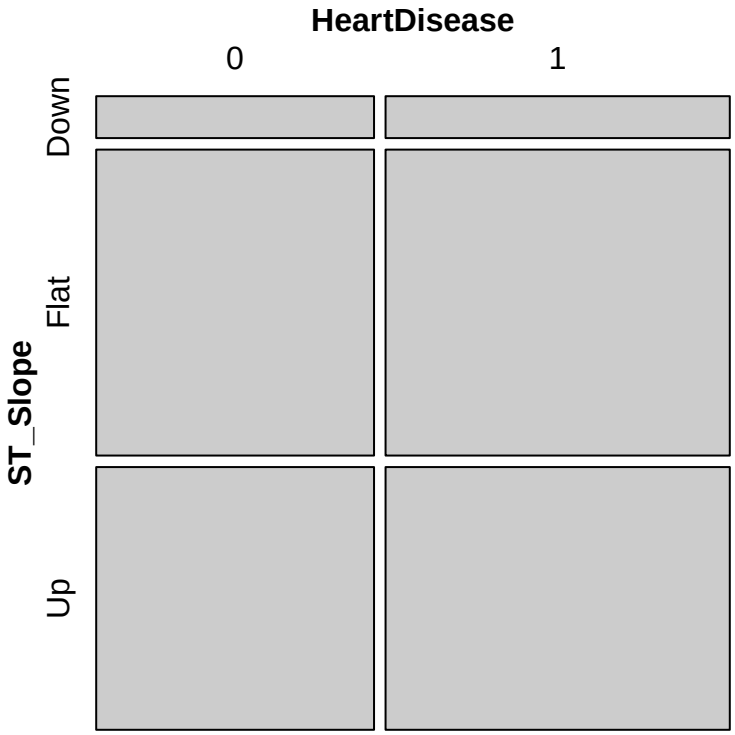
```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): Sex, ChestPainType, RestingECG, ExerciseAngina, ST_Slope
## dbl (7): Age, RestingBP, Cholesterol, FastingBS, MaxHR, Oldpeak, HeartDisease
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
heart <- heart |>
  select(HeartDisease, ST_Slope, ExerciseAngina, RestingECG)
```

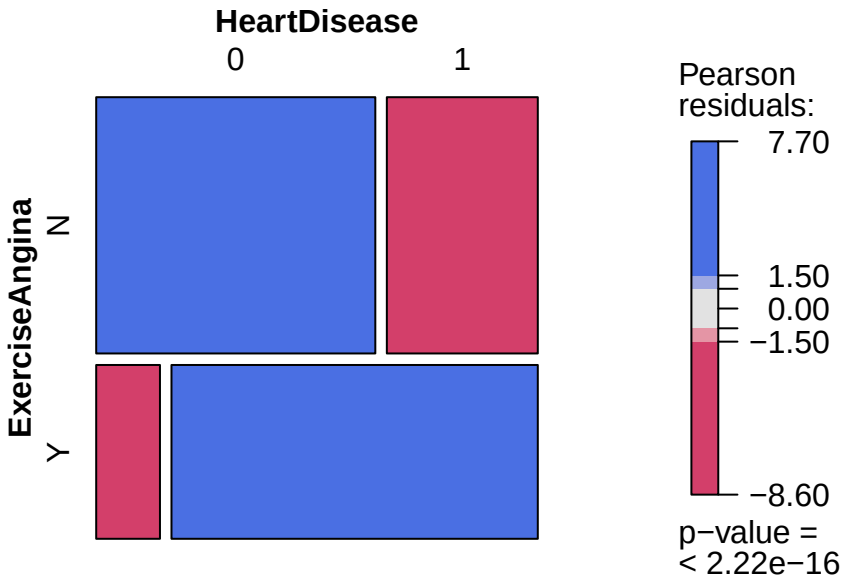
```
## 1.
mosaic(~ ST_Slope + HeartDisease, data=heart, gp=shading_max)
```



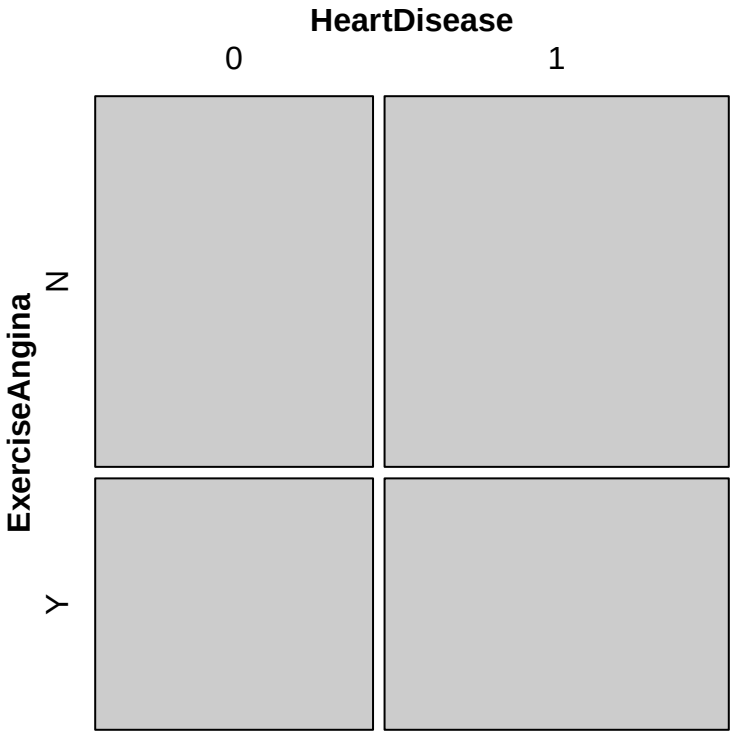

```
mosaic(independence_table(structable(HeartDisease ~ ST_Slope, data=heart)))
```



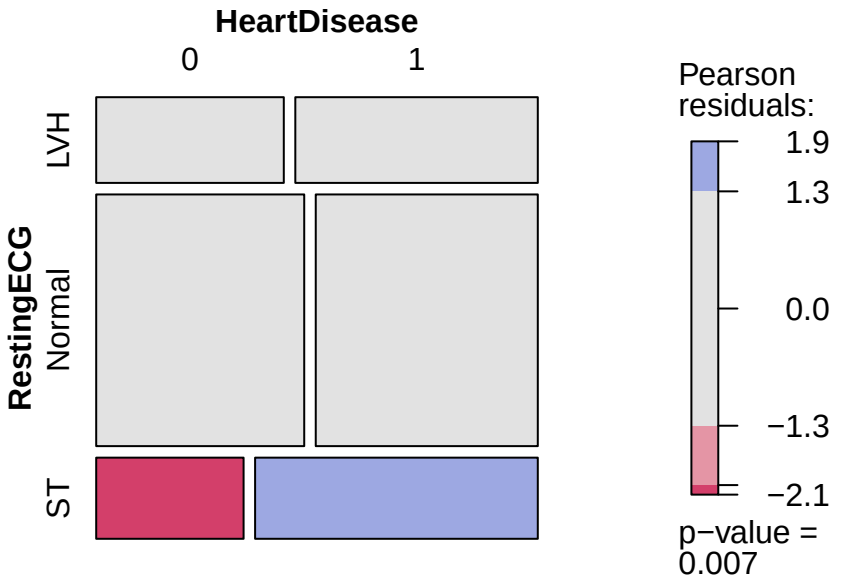

```
mosaic(~ ExerciseAngina + HeartDisease, data=heart, gp=shading_max)
```



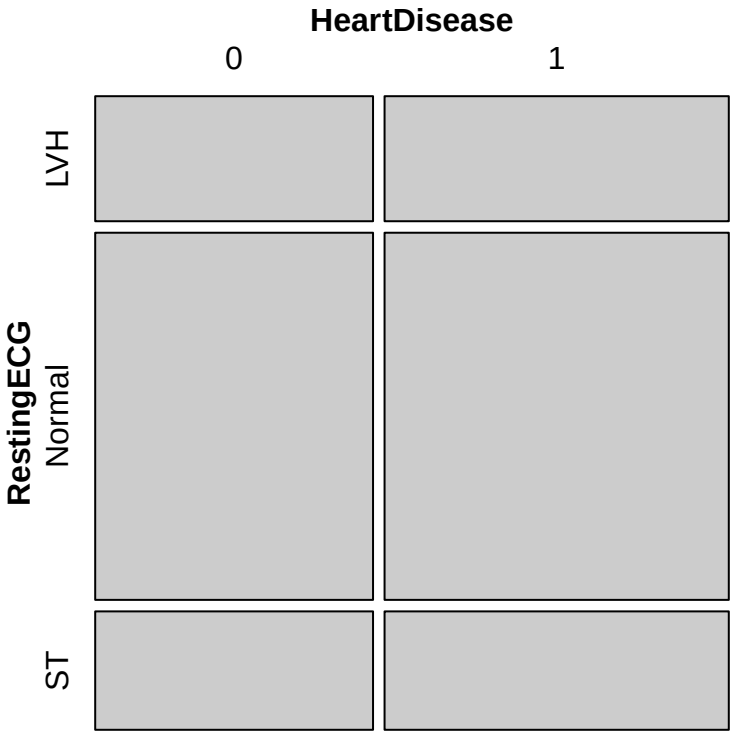

```
mosaic(independence_table(structable(HeartDisease ~ ExerciseAngina, data=heart)))
```




```
mosaic(~ RestingECG + HeartDisease, data=heart, gp=shading_max)
```

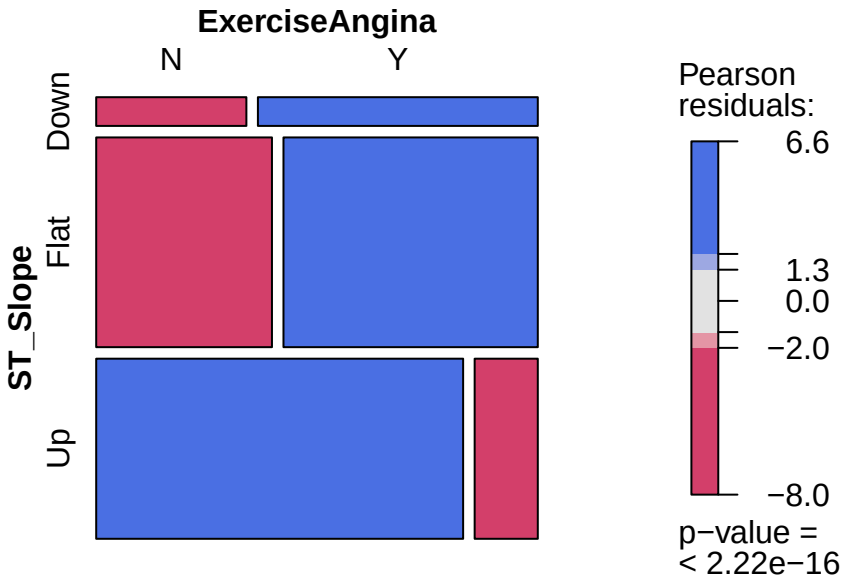



```
mosaic(independence_table(structable(HeartDisease ~ RestingECG, data=heart)))
```

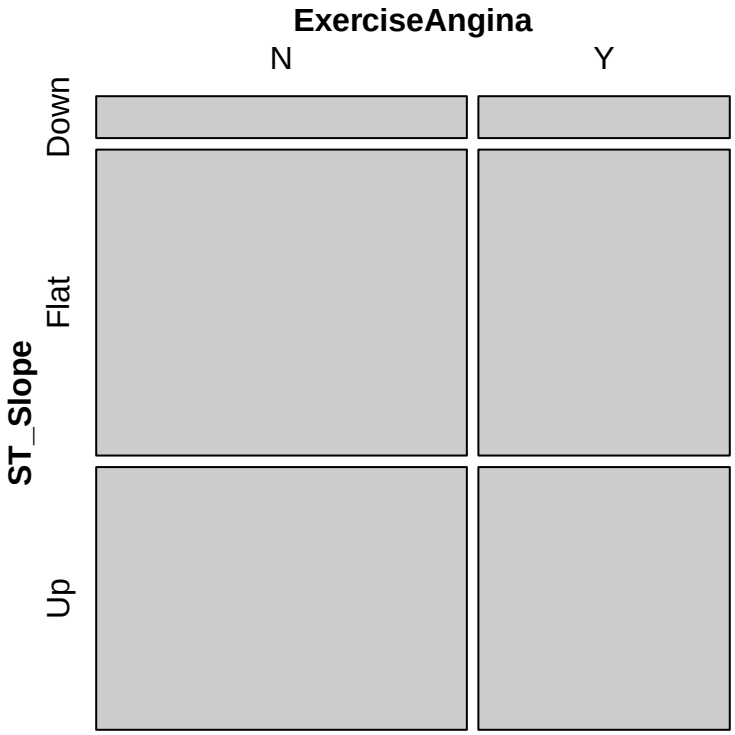


Answer: ST_Slope and ExerciseAngina seem useful for predicting HeartDisease while RestingECG does not seem useful for predicting HeartDisease. Among those that have heart disease, a greater proportion have a "flat" or "down" ST_slope and a smaller proportion have an "up" ST_slope than would be expected under independence. Among those that have heart disease, a greater proportion have exercise angina than expected under independence. The proportion of those that have heart disease and that don't have heart disease are distributed among the levels of RestingECG in a way roughly expected if the two variables were independent.

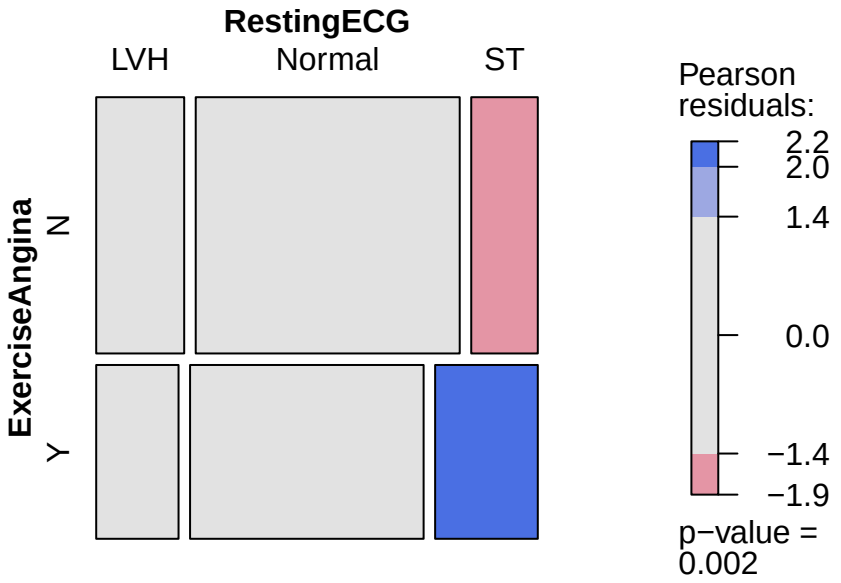
```
## 2
mosaic(~ ST_Slope + ExerciseAngina, data=heart, gp=shading_max)
```



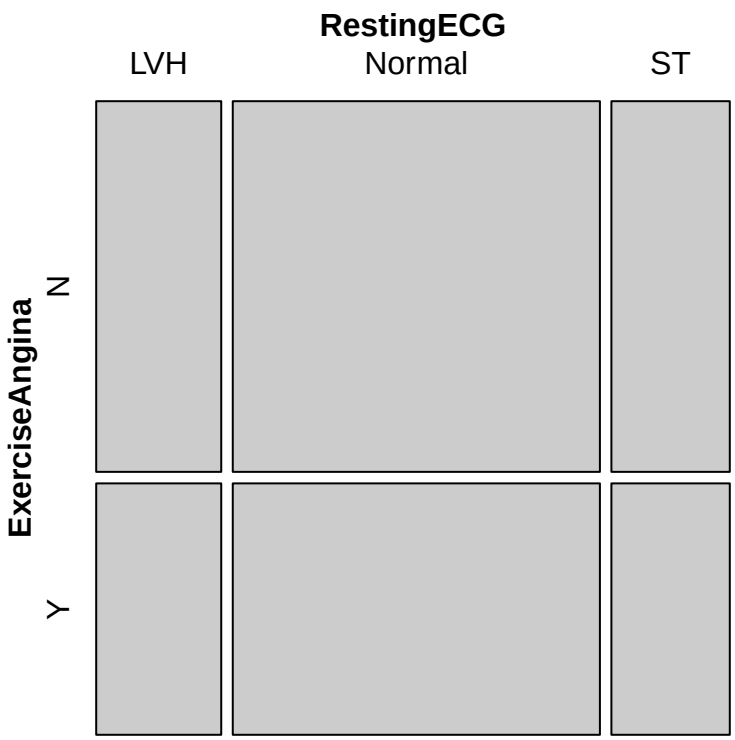

```
mosaic(independence_table(structable(ExerciseAngina ~ ST_Slope, data=heart)))
```



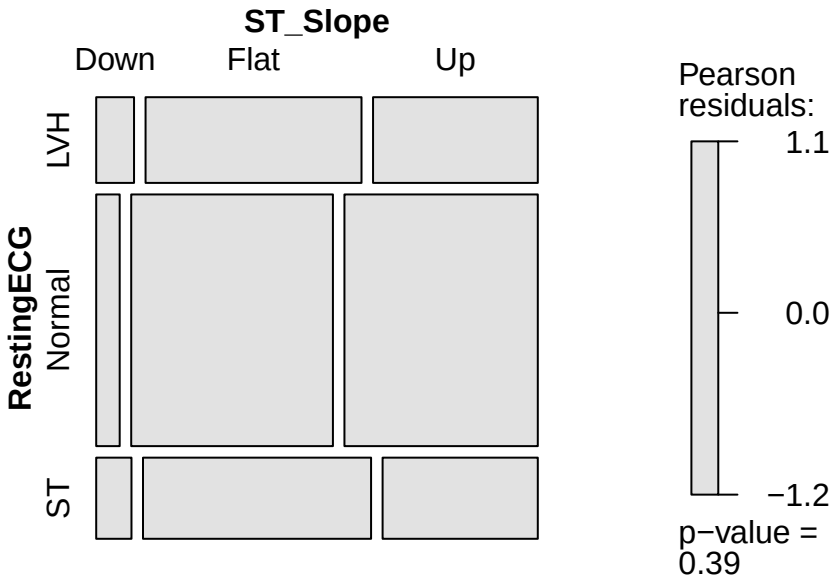

```
mosaic(~ ExerciseAngina + RestingECG, data=heart, gp=shading_max)
```



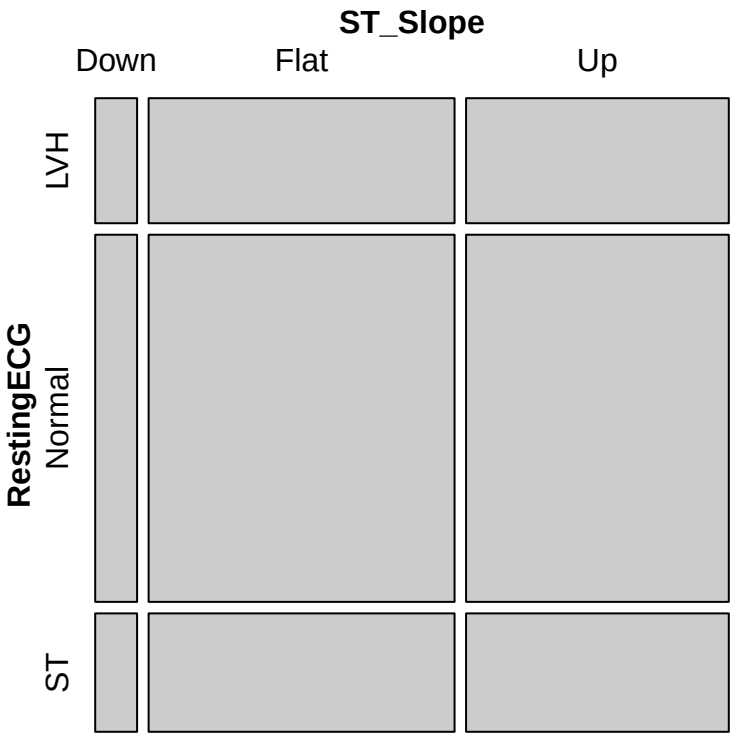

```
mosaic(independence_table(structable(RestingECG ~ ExerciseAngina, data=heart)))
```




```
mosaic(~ RestingECG + ST_Slope, data=heart, gp=shading_max)
```

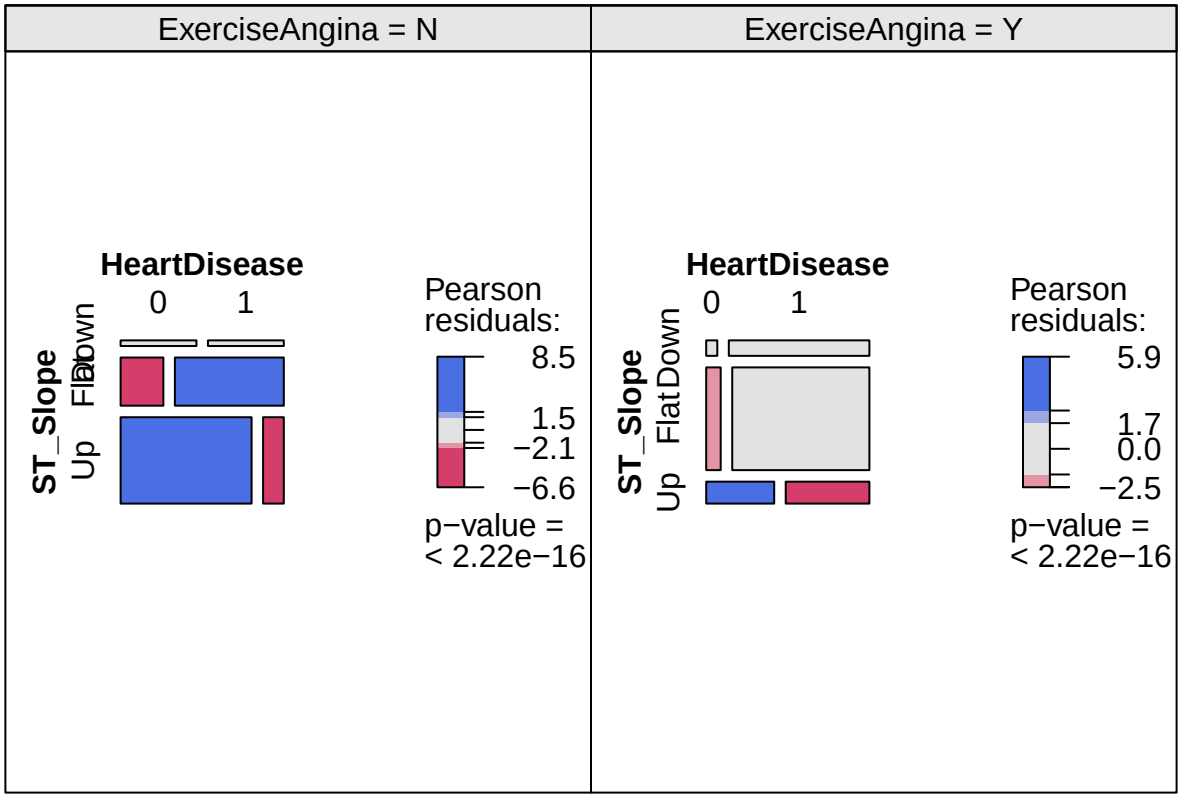



```
mosaic(independence_table(structable(ST_Slope ~ RestingECG, data=heart)))
```



Answer: ExerciseAngina appears to be redundant. The mosaic plots of RestingECG vs. ExerciseAngina and ExerciseAngina vs. ST_Slope demonstrates, both graphically and through a statistical test for independence, that these two pairs of predictors are not independent. The mosaic plot of ST_Slope vs. RestingECG demonstrates, both graphically and through a statistical test for independence, that this pair of predictors is independent. Since ExerciseAngina is related to the other two predictors, it captures the same information. The pairs of ExerciseAngina with the other two predictors is likely redundant.

```
## 3. cotabplot(~ Gender + Admit | Dept, data = UCBAdmissions, gp = shading_max, margins = rep(0, 4))
cotabplot(~ ST_Slope + HeartDisease | ExerciseAngina, data=heart, gp=shading_max)
```



Answer: The relationship between ST_Slope and HeartDisease varies conditional on the value of ExerciseAngina. With no exercise angina, those with heart disease are over represented at ST_Slope equals "Flat" and underrepresented at ST_Slope equals "Up". With exercise angina, the over representation for those with heart disease at the "Flat" level of ST_Slope and the the under representation for those with heart disease at the "Up" level of ST_Slope lessens.