

Final Report

due November 16, 2021 by 11:59 PM

Shelby Brown, Katie Lam, Kaeden Hill

10/31/21

```
## 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()

##
## Attaching package: 'janitor'

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

## Registered S3 method overwritten by 'tune':
##   method      from
##   required_pkgs.model_spec parsnip

## -- Attaching packages ----- tidymodels 0.1.4 --

## v broom      0.7.9      v rsample      0.1.0
## v dials      0.0.10     v tune         0.1.6
## v infer      1.0.0      v workflows    0.2.4
## v modeldata  0.1.1      v workflowsets 0.1.0
## v parsnip    0.1.7      v yardstick    0.0.8
## v recipes    0.1.17

## -- Conflicts ----- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Dig deeper into tidy modeling with R at https://www.tnwr.org

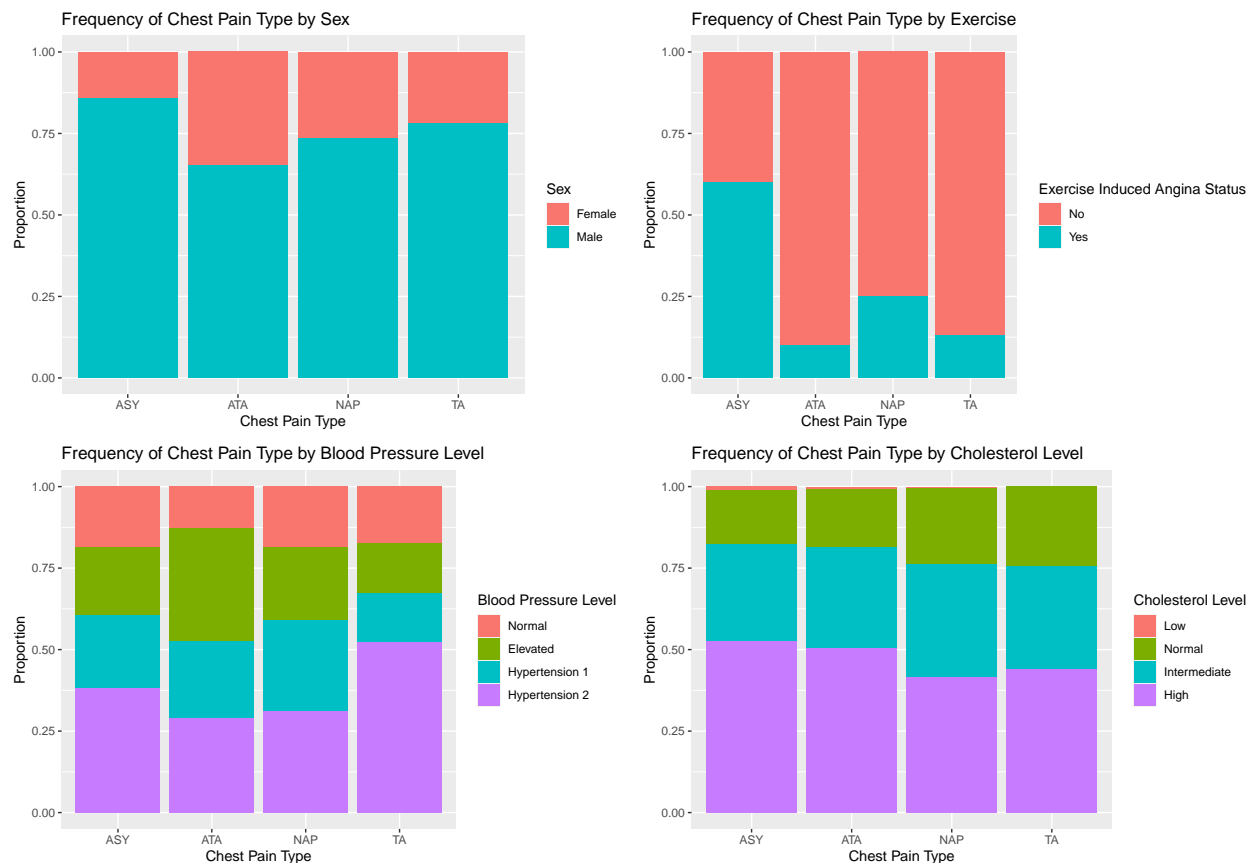
##
## Attaching package: 'gridExtra'
```

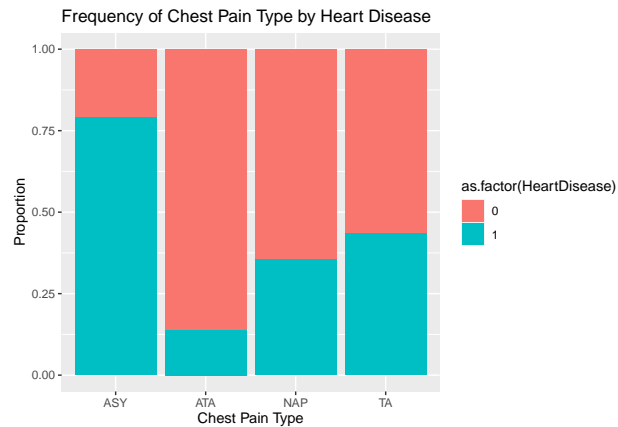
```
## The following object is masked from 'package:dplyr':
##
##     combine
```

Abstract

Introduction

For this data visualization, chest pain types and their relation to other physiological factors were analyzed to search for an association between factors, such as blood pressure, cholesterol, and exercise, and the type of chest pain a patient experiences. The dataset being analyzed is the Heart Failure Prediction Dataset, retrieved from Kaggle, and compiled from five sets with common variables. These sources are the Hungarian Institute of Cardiology. Budapest, University Hospital, Zurich, Switzerland, University Hospital, Basel, Switzerland, the V.A. Medical Center, Long Beach, and the Cleveland Clinic Foundation. The variables of interest include gender, resting blood pressure (mm Hg), serum cholesterol (mm/d), whether or not the angina was exercise induced, and whether or not the patient was diagnosed with heart disease (**diagnosis before or after the angina??). To allow for the use of statistical tests, each variable was transformed into a categorical variable.





Statistical Tests

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

##
## Pearson's Chi-squared test
##
## data: exer_table
## X-squared = 179.27, df = 3, p-value < 2.2e-16

##
## Pearson's Chi-squared test
##
## data: chol_table
## X-squared = 8.6117, df = 9, p-value = 0.4739

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

##
## Pearson's Chi-squared test
##
## data: disease_table
## X-squared = 268.07, df = 3, p-value < 2.2e-16

##
## 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

##
## 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

##
## 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

##
## 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

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

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

##
## Fisher's Exact Test for Count Data
##
## data: disease_step3

```

```
## p-value = 6.656e-07
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.1039251 0.3983253
## sample estimates:
## odds ratio
## 0.204859

##
## Fisher's Exact Test for Count Data
##
## data: disease_step5
## p-value = 3.929e-05
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 2.15486 10.44700
## sample estimates:
## odds ratio
## 4.73165
```

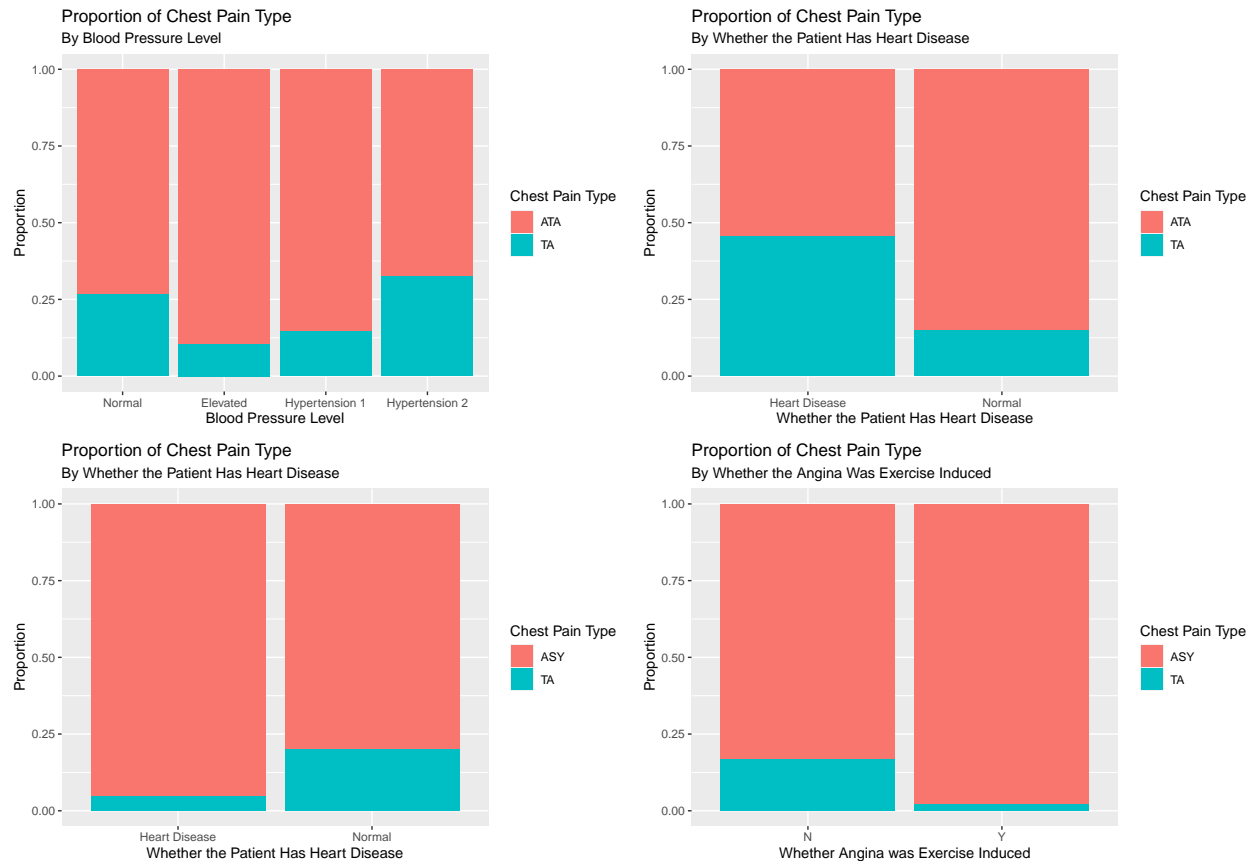
Results

Table 1: χ^2 Test Results - P Values for Preliminary χ^2 Tests

<u>Predicting Variable</u>	<u>P Value for χ^2 Statistic</u>
<i>Sex</i>	4.88e-8
<i>ExerciseAngina</i>	<2.2e-16
<i>chol_level</i>	0.4738
<i>press_level</i>	0.001493
<i>HeartDisease</i>	<2.2e-16

Table 2: Results From Fisher's Exact Step Down Tests

<u>Predicting Variable</u>	<u>TA & ATA Outcome P Values</u>	<u>TA & ASY Outcome P Values</u>
<i>Sex</i>	0.1107	0.1896
<i>ExerciseAngina</i>	0.5883	4.23e-10
<i>press_level</i>	0.006839	0.34
<i>HeartDisease</i>	3.929e-05	6.656e-07



```
## # A tibble: 5 x 7
##   term                                estimate std.error statistic p.value conf.low conf.high
##   <chr>                                <dbl>     <dbl>     <dbl>   <dbl>   <dbl>    <dbl>
## 1 (Intercept)                        3.76      0.440      3.01 2.62e-3  1.66    9.51
## 2 press_levelElevated                 3.38      0.597      2.04 4.15e-2  1.05   11.2
## 3 press_levelHypertension 1           2.27      0.605      1.35 1.76e-1  0.691   7.63
## 4 press_levelHypertension 2           0.889     0.506     -0.233 8.16e-1  0.317   2.35
## 5 HeartDisease                       0.219     0.384     -3.95 7.68e-5  0.102   0.464
```

```
## # A tibble: 3 x 7
##   term                estimate std.error statistic    p.value conf.low conf.high
##   <chr>                <dbl>     <dbl>     <dbl>    <dbl>   <dbl>    <dbl>
## 1 (Intercept)          2.99      0.227      4.82 0.00000144  1.94    4.75
## 2 ExerciseAnginaY       7.16      0.463      4.25 0.0000215  3.08   19.6
## 3 HeartDisease          2.72      0.335      2.99 0.00278    1.42    5.30
```

```
## # A tibble: 2 x 2
##   Sex    count
##   <chr> <int>
## 1 F      80
## 2 M     462
```

```
## # A tibble: 4 x 3
## # Groups:   Sex [2]
##   Sex ChestPainType count
##   <chr> <chr>         <int>
## 1 F    ASY          70
## 2 F    TA           10
```

```

## 3 M      ASY      426
## 4 M      TA       36

## # A tibble: 2 x 2
##   Sex      count
##   <chr> <int>
## 1 F      70
## 2 M     149

## # A tibble: 4 x 3
## # Groups:   Sex [2]
##   Sex      ChestPainType count
##   <chr> <chr>      <int>
## 1 F      ATA          60
## 2 F      TA           10
## 3 M      ATA         113
## 4 M      TA          36

## # A tibble: 2 x 2
##   ExerciseAngina count
##   <chr>      <int>
## 1 N          239
## 2 Y          303

## # A tibble: 4 x 3
## # Groups:   ExerciseAngina [2]
##   ExerciseAngina ChestPainType count
##   <chr>      <chr>      <int>
## 1 N          ASY          199
## 2 N          TA           40
## 3 Y          ASY          297
## 4 Y          TA           6

## # A tibble: 2 x 2
##   ExerciseAngina count
##   <chr>      <int>
## 1 N          196
## 2 Y           23

## # A tibble: 4 x 3
## # Groups:   ExerciseAngina [2]
##   ExerciseAngina ChestPainType count
##   <chr>      <chr>      <int>
## 1 N          ATA          156
## 2 N          TA           40
## 3 Y          ATA           17
## 4 Y          TA           6

## # A tibble: 16 x 3
## # Groups:   press_level [4]
##   press_level      ChestPainType count
##   <fct>      <chr>      <int>
## 1 Normal      ASY          93
## 2 Normal      ATA          22
## 3 Normal      NAP          38
## 4 Normal      TA           8
## 5 Elevated    ASY         102

```

```

## 6 Elevated      ATA      60
## 7 Elevated      NAP      45
## 8 Elevated      TA       7
## 9 Hypertension 1 ASY     111
## 10 Hypertension 1 ATA     41
## 11 Hypertension 1 NAP     57
## 12 Hypertension 1 TA      7
## 13 Hypertension 2 ASY     190
## 14 Hypertension 2 ATA     50
## 15 Hypertension 2 NAP     63
## 16 Hypertension 2 TA     24

## # A tibble: 8 x 3
## # Groups:   press_level [4]
##   press_level ChestPainType count
##   <fct>      <chr>      <int>
## 1 Normal      ASY          93
## 2 Normal      TA           8
## 3 Elevated    ASY         102
## 4 Elevated    TA           7
## 5 Hypertension 1 ASY     111
## 6 Hypertension 1 TA      7
## 7 Hypertension 2 ASY     190
## 8 Hypertension 2 TA     24

## # A tibble: 8 x 3
## # Groups:   press_level [4]
##   press_level ChestPainType count
##   <fct>      <chr>      <int>
## 1 Normal      ATA          22
## 2 Normal      TA           8
## 3 Elevated    ATA          60
## 4 Elevated    TA           7
## 5 Hypertension 1 ATA     41
## 6 Hypertension 1 TA      7
## 7 Hypertension 2 ATA     50
## 8 Hypertension 2 TA     24

```

Discussion