Heart Disease Prediction Using a Random Forrest Model in R

When I first began learning about data science, machine learning seemed unapproachable. I thought that to understand machine learning algorithms, I would need to be a mathematical genius and understand complex equations and write a lot of brilliant programs. As my craft skills improved, I began to grow more comfortable with coding and shifting and shifting I could easily build a **[illegible]** decision models in R, I had no idea where to begin as far as implementing more advanced models.

Recently, I decided to learn more about machine learning and data science by participating in a Kaggle competition and building a Random Forrest model using TidyModels. I saw this these glamorous packages make it easy to fit, test, tune and evaluate a wide variety of diverse model types without getting bogged down in disparate syntax.

After attending some hands-on tutorials using various types of decision tree models, I decided to try my hand at using a random forest classification model. The University of California at Irvine contains excellent real-world data sets, and I decided to use the Cleveland Heart Disease data set to predict heart disease. Here is the url:
https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

You will see that the UCI has data from different hospitals however a combined version was uploaded to Kaggle. In this tutorial, I will use the Kaggle version of the Cleveland Heart Disease data set which is hosted at U. C. Irvine is the ur-file and the version of the data which I used: https://www.kaggle.com/ronitf/heart-disease-uci.

Random forest models are based on decision tree models, a simple model type of model where each data point is evaluated on whether or not it passes a logical test based on one or more variables. The result of the test determines which branch of the tree that data point will continue down, determining the next logical test it will encounter. After passing through enough tests, the model will have enough information to accurately classify the data point. Decision trees are fast, easy to fit, classification models but not replicable one of the possible solutions categories. A random forest is a bootstrapped version of a decision tree model that uses a random subset of predictors at each split and the model result is better, often producing a more accurate model.

Let's start by loading the packages we need and reading the data set:

```
library(tidyverse)
library(tidymodels)
library(skimr)
library(vroom)

# dplyr is part of the core,
# so load it:
library(dplyr)
```

```
## Rows: 303
## Columns: 14
## chr [2]: sex, cp
## dbl [12]: age, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target
## ℹ Use `spec()` for the full column specifications.
## ℹ Use `cols()` to print all.
## Rows: 303
## Columns: 14
## chr [2]: sex, cp
## dbl [12]: age, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target
## ℹ Use `spec()` for the full column specifications.
## ℹ Use `cols()` to print all.
```

```
heart_df <- read_csv("./data/heart.csv")
```

```
## Rows: 303
## Columns: 14
## chr [2]: sex, cp
## dbl [12]: age, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target
## ℹ Use `spec()` for the full column specifications.
## ℹ Use `cols()` to print all.
```

```
heart_df
```

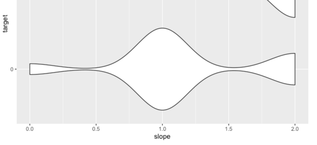
```
## Rows: 303
## Columns: 14
## chr [2]: sex, cp
## dbl [12]: age, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target
## ℹ Use `spec()` for the full column specifications.
## ℹ Use `cols()` to print all.
```

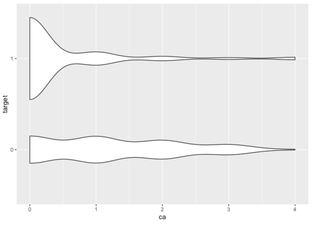
```
## # A tibble: 303 × 14
##    age  sex    cp trestbps  chol   fbs restecg thalach exang oldpeak slope
##  <dbl> <chr> <chr>    <dbl> <dbl> <dbl>   <dbl>   <dbl> <dbl>   <dbl> <dbl>
##  1    63 male  typ      145   233    1       2     150     0     2.3     3
##  2    37 male  ast      130   250    0       1     187     0     3.5     3
##  3    41 fem   ata      130   204    0       2     172     0     1.4     1
##  4    56 male  ata      120   236    0       0     178     0     0.8     1
##  5    57 fem   nap      120   354    0       1     163     1     0.6     1
##  6    57 male  asy      140   192    0       1     148     0     0.4     2
##  7    56 fem   ata      140   294    0       2     153     0     1.3     2
##  8    44 male  ast      120   263    0       1     173     0     0       1
##  9    52 male  nap      172   199    1       1     162     0     0.5     1
##  10    57 male  asy      150   168    0       1     174     0     1.6     1
##  11    54 male  asy      140   239    0       1     160     0     1.2     1
##  12    48 fem   nap      130   275    0       1     139     0     0.2     1
##  13    49 male  ast      130   266    0       1     171     0     0.6     1
##  14    64 male  typ      110   211    0       2     144     1     1.8     2
##  15    58 fem   ata      150   283    1       2     162     0     1       3
##  16    50 male  ast      120   219    0       1     158     0     1.6     2
##  17    43 male  nap      150   304    0       1     148     1     1.6     2
##  18    60 male  ast      120   232    0       1     143     1     0.7     2
##  19    67 male  asy      160   286    0       2     108     1     1.5     2
##  20    67 male  ast      120   229    0       2     129     1     2.6     2
##  21    62 fem   asy      140   268    0       2     160     0     3.6     3
##  22    65 male  nap      120   177    0       1     140     0     0.4     1
##  23    44 male  ast      112   290    0       2     153     0     0       2
##  24    60 male  ast      132   227    0       2     138     0     0.3     1
##  25    65 male  asy      120   177    0       1     140     0     0.4     1
##  26    60 male  asy      150   258    0       2     157     0     2.6     2
##  27    60 male  asy      102   318    0       1     160     0     0       1
##  28    54 male  ast      150   205    0       1     165     0     0       1
##  29    44 male  nap      112   240    0       1     138     1     0.2     2
##  30    44 male  ast      108   141    0       1     175     0     0.6     2
##  31    51 male  nap      120   295    0       2     157     0     0.6     1
##  32    59 male  asy      170   326    0       2     140     1     3.4     3
##  33    71 fem   asy      160   302    0       1     162     0     0.4     2
##  34    61 male  asy      150   243    1       1     137     1     1       2
##  35    54 male  asy      125   273    0       2     152     0     0.5     3
##  36    78 fem   asy      160   250    0       2     165     0     2.4     2
##  37    62 fem   ata      160   254    0       2     155     1     1.5     2
##  38    55 fem   nap      140   239    0       1     160     0     1.2     1
##  39    44 male  ata      130   233    0       0     179     1     0.4     3
##  40    11 fem   typ      120   265    0       1     146     0     0.4     2
##  41    43 fem   asy      115   564    0       2     160     0     1.6     2
##  42    65 fem   asy      148   960    0       2     158     1     0.8     1
##  43    70 male  asy      130   322    0       2     109     0     2.4     2
##  44    62 male  asy      115   564    0       2     160     0     1.6     2
##  45    67 fem   asy      160   286    0       2     108     1     1.5     2
##  46    57 male  asy      165   289    1       2     124     0     0.8     2
##  47    65 male  asy      148   960    0       2     158     1     0.8     1
##  48    54 male  asy      124   266    0       2     109     1     2.2     2
##  49    63 fem   typ      145   233    1       2     150     0     2.3     3
##  50    31 male  ata      140   219    0       1     162     0     1.8     1
##  51    52 male  nap      172   199    1       1     162     0     0.5     1
##  52    44 male  ast      120   263    0       1     173     0     0       1
##  53    62 fem   asy      140   268    0       2     160     0     3.6     3
##  54    54 male  asy      125   273    0       2     152     0     0.5     3
##  55    51 male  nap      120   295    0       2     157     0     0.6     1
##  56    43 male  nap      150   304    0       1     148     1     1.6     2
##  57    60 male  asy      102   318    0       1     160     0     0       1
##  58    65 male  asy      120   177    0       1     140     0     0.4     1
##  59    44 male  ast      112   290    0       2     153     0     0       2
##  60    44 male  ast      108   141    0       1     175     0     0.6     2
##  61    60 male  ast      132   227    0       2     138     0     0.3     1
##  62    67 male  asy      160   286    0       2     108     1     1.5     2
##  63    67 male  ast      120   229    0       2     129     1     2.6     2
##  64    62 fem   asy      140   268    0       2     160     0     3.6     3
##  65    65 male  nap      120   177    0       1     140     0     0.4     1
##  66    60 male  asy      150   258    0       2     157     0     2.6     2
##  67    60 male  asy      102   318    0       1     160     0     0       1
##  68    54 male  ast      150   205    0       1     165     0     0       1
##  69    44 male  nap      112   240    0       1     138     1     0.2     2
##  70    44 male  ast      108   141    0       1     175     0     0.6     2
##  71    51 male  nap      120   295    0       2     157     0     0.6     1
##  72    59 male  asy      170   326    0       2     140     1     3.4     3
##  73    71 fem   asy      160   302    0       1     162     0     0.4     2
##  74    61 male  asy      150   243    1       1     137     1     1       2
##  75    54 male  asy      125   273    0       2     152     0     0.5     3
##  76    78 fem   asy      160   250    0       2     165     0     2.4     2
##  77    62 fem   ata      160   254    0       2     155     1     1.5     2
##  78    55 fem   nap      140   239    0       1     160     0     1.2     1
##  79    44 male  ata      130   233    0       0     179     1     0.4     3
##  80    11 fem   typ      120   265    0       1     146     0     0.4     2
##  81    43 fem   asy      115   564    0       2     160     0     1.6     2
##  82    65 fem   asy      148   960    0       2     158     1     0.8     1
##  83    70 male  asy      130   322    0       2     109     0     2.4     2
##  84    62 male  asy      115   564    0       2     160     0     1.6     2
##  85    67 fem   asy      160   286    0       2     108     1     1.5     2
##  86    57 male  asy      165   289    1       2     124     0     0.8     2
##  87    65 male  asy      148   960    0       2     158     1     0.8     1
##  88    54 male  asy      124   266    0       2     109     1     2.2     2
##  89    63 fem   typ      145   233    1       2     150     0     2.3     3
##  90    31 male  ata      140   219    0       1     162     0     1.8     1
##  91    52 male  nap      172   199    1       1     162     0     0.5     1
##  92    44 male  ast      120   263    0       1     173     0     0       1
##  93    62 fem   asy      140   268    0       2     160     0     3.6     3
##  94    54 male  asy      125   273    0       2     152     0     0.5     3
##  95    51 male  nap      120   295    0       2     157     0     0.6     1
##  96    43 male  nap      150   304    0       1     148     1     1.6     2
##  97    60 male  asy      102   318    0       1     160     0     0       1
##  98    65 male  asy      120   177    0       1     140     0     0.4     1
##  99    44 male  ast      112   290    0       2     153     0     0       2
##  100    44 male  ast      108   141    0       1     175     0     0.6     2
##  101    60 male  ast      132   227    0       2     138     0     0.3     1
##  102    67 male  asy      160   286    0       2     108     1     1.5     2
##  103    67 male  ast      120   229    0       2     129     1     2.6     2
##  104    62 fem   asy      140   268    0       2     160     0     3.6     3
##  105    65 male  nap      120   177    0       1     140     0     0.4     1
##  106    60 male  asy      150   258    0       2     157     0     2.6     2
##  107    60 male  asy      102   318    0       1     160     0     0       1
##  108    54 male  ast      150   205    0       1     165     0     0       1
##  109    44 male  nap      112   240    0       1     138     1     0.2     2
##  110    44 male  ast      108   141    0       1     175     0     0.6     2
##  111    51 male  nap      120   295    0       2     157     0     0.6     1
##  112    59 male  asy      170   326    0       2     140     1     3.4     3
##  113    71 fem   asy      160   302    0       1     162     0     0.4     2
##  114    61 male  asy      150   243    1       1     137     1     1       2
##  115    54 male  asy      125   273    0       2     152     0     0.5     3
##  116    78 fem   asy      160   250    0       2     165     0     2.4     2
##  117    62 fem   ata      160   254    0       2     155     1     1.5     2
##  118    55 fem   nap      140   239    0       1     160     0     1.2     1
##  119    44 male  ata      130   233    0       0     179     1     0.4     3
##  120    11 fem   typ      120   265    0       1     146     0     0.4     2
##  121    43 fem   asy      115   564    0       2     160     0     1.6     2
##  122    65 fem   asy      148   960    0       2     158     1     0.8     1
##  123    70 male  asy      130   322    0       2     109     0     2.4     2
##  124    62 male  asy      115   564    0       2     160     0     1.6     2
##  125    67 fem   asy      160   286    0       2     108     1     1.5     2
##  126    57 male  asy      165   289    1       2     124     0     0.8     2
##  127    65 male  asy      148   960    0       2     158     1     0.8     1
##  128    54 male  asy      124   266    0       2     109     1     2.2     2
##  129    63 fem   typ      145   233    1       2     150     0     2.3     3
##  130    31 male  ata      140   219    0       1     162     0     1.8     1
##  131    52 male  nap      172   199    1       1     162     0     0.5     1
##  132    44 male  ast      120   263    0       1     173     0     0       1
##  133    62 fem   asy      140   268    0       2     160     0     3.6     3
##  134    54 male  asy      125   273    0       2     152     0     0.5     3
##  135    51 male  nap      120   295    0       2     157     0     0.6     1
##  136    43 male  nap      150   304    0       1     148     1     1.6     2
##  137    60 male  asy      102   318    0       1     160     0     0       1
##  138    65 male  asy      120   177    0       1     140     0     0.4     1
##  139    44 male  ast      112   290    0       2     153     0     0       2
##  140    44 male  ast      108   141    0       1     175     0     0.6     2
##  141    60 male  ast      132   227    0       2     138     0     0.3     1
##  142    67 male  asy      160   286    0       2     108     1     1.5     2
##  143    67 male  ast      120   229    0       2     129     1     2.6     2
##  144    62 fem   asy      140   268    0       2     160     0     3.6     3
##  145    65 male  nap      120   177    0       1     140     0     0.4     1
##  146    60 male  asy      150   258    0       2     157     0     2.6     2
##  147    60 male  asy      102   318    0       1     160     0     0       1
##  148    54 male  ast      150   205    0       1     165     0     0       1
##  149    44 male  nap      112   240    0       1     138     1     0.2     2
##  150    44 male  ast      108   141    0       1     175     0     0.6     2
##  151    51 male  nap      120   295    0       2     157     0     0.6     1
##  152    59 male  asy      170   326    0       2     140     1     3.4     3
##  153    71 fem   asy      160   302    0       1     162     0     0.4     2
##  154    61 male  asy      150   243    1       1     137     1     1       2
##  155    54 male  asy      125   273    0       2     152     0     0.5     3
##  156    78 fem   asy      160   250    0       2     165     0     2.4     2
##  157    62 fem   ata      160   254    0       2     155     1     1.5     2
##  158    55 fem   nap      140   239    0       1     160     0     1.2     1
##  159    44 male  ata      130   233    0       0     179     1     0.4     3
##  160    11 fem   typ      120   265    0       1     146     0     0.4     2
##  161    43 fem   asy      115   564    0       2     160     0     1.6     2
##  162    65 fem   asy      148   960    0       2     158     1     0.8     1
##  163    70 male  asy      130   322    0       2     109     0     2.4     2
##  164    62 male  asy      115   564    0       2     160     0     1.6     2
##  165    67 fem   asy      160   286    0       2     108     1     1.5     2
##  166    57 male  asy      165   289    1       2     124     0     0.8     2
##  167    65 male  asy      148   960    0       2     158     1     0.8     1
##  168    54 male  asy      124   266    0       2     109     1     2.2     2
##  169    63 fem   typ      145   233    1       2     150     0     2.3     3
##  170    31 male  ata      140   219    0       1     162     0     1.8     1
##  171    52 male  nap      172   199    1       1     162     0     0.5     1
##  172    44 male  ast      120   263    0       1     173     0     0       1
##  173    62 fem   asy      140   268    0       2     160     0     3.6     3
##  174    54 male  asy      125   273    0       2     152     0     0.5     3
##  175    51 male  nap      120   295    0       2     157     0     0.6     1
##  176    43 male  nap      150   304    0       1     148     1     1.6     2
##  177    60 male  asy      102   318    0       1     160     0     0       1
##  178    65 male  asy      120   177    0       1     140     0     0.4     1
##  179    44 male  ast      112   290    0       2     153     0     0       2
##  180    44 male  ast      108   141    0       1     175     0     0.6     2
##  181    60 male  ast      132   227    0       2     138     0     0.3     1
##  182    67 male  asy      160   286    0       2     108     1     1.5     2
##  183    67 male  ast      120   229    0       2     129     1     2.6     2
##  184    62 fem   asy      140   268    0       2     160     0     3.6     3
##  185    65 male  nap      120   177    0       1     140     0     0.4     1
##  186    60 male  asy      150   258    0       2     157     0     2.6     2
##  187    60 male  asy      102   318    0       1     160     0     0       1
##  188    54 male  ast      150   205    0       1     165     0     0       1
##  189    44 male  nap      112   240    0       1     138     1     0.2     2
##  190    44 male  ast      108   141    0       1     175     0     0.6     2
##  191    51 male  nap      120   295    0       2     157     0     0.6     1
##  192    59 male  asy      170   326    0       2     140     1     3.4     3
##  193    71 fem   asy      160   302    0       1     162     0     0.4     2
##  194    61 male  asy      150   243    1       1     137     1     1       2
##  195    54 male  asy      125   273    0       2     152     0     0.5     3
##  196    78 fem   asy      160   250    0       2     165     0     2.4     2
##  197    62 fem   ata      160   254    0       2     155     1     1.5     2
##  198    55 fem   nap      140   239    0       1     160     0     1.2     1
##  199    44 male  ata      130   233    0       0     179     1     0.4     3
##  200    11 fem   typ      120   265    0       1     146     0     0.4     2
##  201    43 fem   asy      115   564    0       2     160     0     1.6     2
##  202    65 fem   asy      148   960    0       2     158     1     0.8     1
##  203    70 male  asy      130   322    0       2     109     0     2.4     2
##  204    62 male  asy      115   564    0       2     160     0     1.6     2
##  205    67 fem   asy      160   286    0       2     108     1     1.5     2
##  206    57 male  asy      165   289    1       2     124     0     0.8     2
##  207    65 male  asy      148   960    0       2     158     1     0.8     1
##  208    54 male  asy      124   266    0       2     109     1     2.2     2
##  209    63 fem   typ      145   233    1       2     150     0     2.3     3
##  210    31 male  ata      140   219    0       1     162     0     1.8     1
##  211    52 male  nap      172   199    1       1     162     0     0.5     1
##  212    44 male  ast      120   263    0       1     173     0     0       1
##  213    62 fem   asy      140   268    0       2     160     0     3.6     3
##  214    54 male  asy      125   273    0       2     152     0     0.5     3
##  215    51 male  nap      120   295    0       2     157     0     0.6     1
##  216    43 male  nap      150   304    0       1     148     1     1.6     2
##  217    60 male  asy      102   318    0       1     160     0     0       1
##  218    65 male  asy      120   177    0       1     140     0     0.4     1
##  219    44 male  ast      112   290    0       2     153     0     0       2
##  220    44 male  ast      108   141    0       1     175     0     0.6     2
##  221    60 male  ast      132   227    0       2     138     0     0.3     1
##  222    67 male  asy      160   286    0       2     108     1     1.5     2
##  223    67 male  ast      120   229    0       2     129     1     2.6     2
##  224    62 fem   asy      140   268    0       2     160     0     3.6     3
##  225    65 male  nap      120   177    0       1     140     0     0.4     1
##  226    60 male  asy      150   258    0       2     157     0     2.6     2
##  227    60 male  asy      102   318    0       1     160     0     0       1
##  228    54 male  ast      150   205    0       1     165     0     0       1
##  229    44 male  nap      112   240    0       1     138     1     0.2     2
##  230    44 male  ast      108   141    0       1     175     0     0.6     2
##  231    51 male  nap      120   295    0       2     157     0     0.6     1
##  232    59 male  asy      170   326    0       2     140     1     3.4     3
##  233    71 fem   asy      160   302    0       1     162     0     0.4     2
##  234    61 male  asy      150   243    1       1     137     1     1       2
##  235    54 male  asy      125   273    0       2     152     0     0.5     3
##  236    78 fem   asy      160   250    0       2     165     0     2.4     2
##  237    62 fem   ata      160   254    0       2     155     1     1.5     2
##  238    55 fem   nap      140   239    0       1     160     0     1.2     1
##  239    44 male  ata      130   233    0       0     179     1     0.4     3
##  240    11 fem   typ      120   265    0       1     146     0     0.4     2
##  241    43 fem   asy      115   564    0       2     160     0     1.6     2
##  242    65 fem   asy      148   960    0       2     158     1     0.8     1
##  243    70 male  asy      130   322    0       2     109     0     2.4     2
##  244    62 male  asy      115   564    0       2     160     0     1.6     2
##  245    67 fem   asy      160   286    0       2     108     1     1.5     2
##  246    57 male  asy      165   289    1       2     124     0     0.8     2
##  247    65 male  asy      148   960    0       2     158     1     0.8     1
##  248    54 male  asy      124   266    0       2     109     1     2.2     2
##  249    63 fem   typ      145   233    1       2     150     0     2.3     3
##  250    31 male  ata      140   219    0       1     162     0     1.8     1
##  251    52 male  nap      172   199    1       1     162     0     0.5     1
##  252    44 male  ast      120   263    0       1     173     0     0       1
##  253    62 fem   asy      140   268    0       2     160     0     3.6     3
##  254    54 male  asy      125   273    0       2     152     0     0.5     3
##  255    51 male  nap      120   295    0       2     157     0     0.6     1
##  256    43 male  nap      150   304    0       1     148     1     1.6     2
##  257    60 male  asy      102   318    0       1     160     0     0       1
##  258    65 male  asy      120   177    0       1     140     0     0.4     1
##  259    44 male  ast      112   290    0       2     153     0     0       2
##  260    44 male  ast      108   141    0       1     175     0     0.6     2
##  261    60 male  ast      132   227    0       2     138     0     0.3     1
##  262    67 male  asy      160   286    0       2     108     1     1.5     2
##  263    67 male  ast      120   229    0       2     129     1     2.6     2
##  264    62 fem   asy      140   268    0       2     160     0     3.6     3
##  265    65 male  nap      120   177    0       1     140     0     0.4     1
##  266    60 male  asy      150   258    0       2     157     0     2.6     2
##  267    60 male  asy      102   318    0       1     160     0     0       1
##  268    54 male  ast      150   205    0       1     165     0     0       1
##  269    44 male  nap      112   240    0       1     138     1     0.2     2
##  270    44 male  ast      108   141    0       1     175     0     0.6     2
##  271    51 male  nap      120   295    0       2     157     0     0.6     1
##  272    59 male  asy      170   326    0       2     140     1     3.4     3
##  273    71 fem   asy      160   302    0       1     162     0     0.4     2
##  274    61 male  asy      150   243    1       1     137     1     1       2
##  275    54 male  asy      125   273    0       2     152     0     0.5     3
##  276    78 fem   asy      160   250    0       2     165     0     2.4     2
##  277    62 fem   ata      160   254    0       2     155     1     1.5     2
##  278    55 fem   nap      140   239    0       1     160     0     1.2     1
##  279    44 male  ata      130   233    0       0     179     1     0.4     3
##  280    11 fem   typ      120   265    0       1     146     0     0.4     2
##  281    43 fem   asy      115   564    0       2     160     0     1.6     2
##  282    65 fem   asy      148   960    0       2     158     1     0.8     1
##  283    70 male  asy      130   322    0       2     109     0     2.4     2
##  284    62 male  asy      115   564    0       2     160     0     1.6     2
##  285    67 fem   asy      160   286    0       2     108     1     1.5     2
##  286    57 male  asy      165   289    1       2     124     0     0.8     2
##  287    65 male  asy      148   960    0       2     158     1     0.8     1
##  288    54 male  asy      124   266    0       2     109     1     2.2     2
##  289    63 fem   typ      145   233    1       2     150     0     2.3     3
##  290    31 male  ata      140   219    0       1     162     0     1.8     1
##  291    52 male  nap      172   199    1       1     162     0     0.5     1
##  292    44 male  ast      120   263    0       1     173     0     0       1
##  293    62 fem   asy      140   268    0       2     160     0     3.6     3
##  294    54 male  asy      125   273    0       2     152     0     0.5     3
##  295    51 male  nap      120   295    0       2     157     0     0.6     1
##  296    43 male  nap      150   304    0       1     148     1     1.6     2
##  297    60 male  asy      102   318    0       1     160     0     0       1
##  298    65 male  asy      120   177    0       1     140     0     0.4     1
##  299    44 male  ast      112   290    0       2     153     0     0       2
##  300    44 male  ast      108   141    0       1     175     0     0.6     2
##  301    60 male  ast      132   227    0       2     138     0     0.3     1
##  302    67 male  asy      160   286    0       2     108     1     1.5     2
##  303    67 male  ast      120   229    0       2     129     1     2.6     2
```

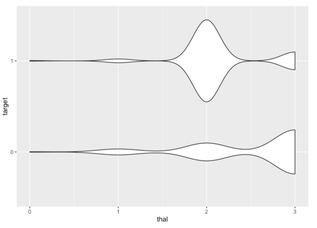
```
skimr::skim()
```

```
## Skim summary statistics
## Character variables                  Numeric variables
## ─────────────────────────────────────────────────────────────────────────────
## ℹ Use `spec()` for the full column specifications.
## ℹ Use `cols()` to print all.
## Rows: 303
## Columns: 14
## chr [2]: sex, cp
## dbl [12]: age, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target
## ℹ Use `spec()` for the full column specifications.
## ℹ Use `cols()` to print all.
```

```
## Skim summary statistics
## Character variables                  Numeric variables
## ─────────────────────────────────────────────────────────────────────────────
```







These fancy violin plots shed some light on the distribution of the predictor variables in the two outcome groups. Variables like number of major vessels colored by flourosopy (ca), chest pain type (cp), max heart rate (thalach), thal and slope of the peak exercise ST segment (slope) seem to vary significantly and will likely be important in our model. Resting heart rate, fbs, restecg, and sex seem to matter less.

Now lets get ready to build our model. First we need to split the data into training and test sets.

```
##splitting the data up into training and test sets
set.seed(123)
split_dat <- initial_split(dat)
train_dat <- training(split_dat)
test_dat <- testing(split_dat)

split_dat
```

```
## <Analysis/Assess/Total>
## <228/75/303>
```

We'll train our data on the training set of 228 and then test and evaluate on the test set of 75.

In order to optimize our model using the tune package, we need to create a sample of the data to test different parameter values on. In this example I use vfold_cv() to create a dataframe composed of 10 randomly selected groups of data points (known as folds) from our training set. We then test each model with different parameter values on each of the 10 folds. Then we can evaluate each model's performance and select the optimal parameters. Testing different parameter values allows for a more accurate model and conducting these test on 10 different folds helps to prevent overfitting as performance is evaluated on 10 different samples. This process is known as cross validation.

```
##creating a cross validated version of the training set to tune parameters
dat_cv <- vfold_cv(train_dat)
```

Now its time to start preprocessing our model! We'll use the recipes package to specify our outcome variable target, and our predictor variables (all 13 others). Since there are no missing values, we don't need to add an imputation step, but we could easily with something like step_impute_linear() or step_knnimpute(). Since many of our variables have vastly different ranges, it's a good idea to normalize our data which I do hear with step_normalize().

```
rand_for_recipe <-
  recipe(target ~ ., data = dat) %>%
  step_normalize(all_numeric())
```

Now we'll specify the type of model, that we will tune the mtry paremeter, that we'll use the ranger engine, our method of variable importance, and that our mode is classification.

```
##specifying the use of a random forest model
rand_for_model <- rand_forest() %>%
##specifying that we will be tuning mtry
##(number of variables that can be split on at each tree node)
  set_args(mtry = tune()) %>%
##specifying the engine and the mode of variable importance
  set_engine("ranger", importance = "impurity") %>%
##selecting the mode, for ranger the options are "classification" and "regression"
  set_mode("classification")
```

Next we combine the recipe and model into a workflow for easy tuning and fitting.

```
##adding the the recipe and model to a workflow
rand_for_workflow <-
  workflow() %>%
  add_recipe(rand_for_recipe) %>%
  add_model(rand_for_model)
```

Now that our workflow is set up we can tune the mtry parameter. We'll use expand_grid to create a dataframe of possible values, select relevant metrics - in this case auc and accuracy, and then calculate.

```
##creating a dataframe of mtry values to test
rand_for_tune_grid <- expand_grid(mtry = c(2,3,4,5))
##calculating the tune results
rand_for_tune_results <- rand_for_workflow %>%
  tune_grid(resamples = dat_cv,
            grid = rand_for_tune_grid,
            ##selecting relevent metrics to assess fit
            metrics = metric_set(roc_auc, accuracy))
```

Let's take a look at the results:

```
collect_metrics(rand_for_tune_results)
```

```
## # A tibble: 8 x 7
##    mtry .metric  .estimator  mean     n std_err .config
##   <dbl> <chr>    <chr>      <dbl> <int>   <dbl> <chr>
## 1     2 accuracy binary     0.820    10  0.0202 Preprocessor1_Model1
## 2     2 roc_auc  binary     0.914    10  0.0195 Preprocessor1_Model1
## 3     3 accuracy binary     0.824    10  0.0208 Preprocessor1_Model2
## 4     3 roc_auc  binary     0.914    10  0.0199 Preprocessor1_Model2
## 5     4 accuracy binary     0.833    10  0.0193 Preprocessor1_Model3
## 6     4 roc_auc  binary     0.910    10  0.0223 Preprocessor1_Model3
## 7     5 accuracy binary     0.829    10  0.0200 Preprocessor1_Model4
## 8     5 roc_auc  binary     0.902    10  0.0217 Preprocessor1_Model4
```

Now we'll select the value with the best AUC and add it to our workflow.

```
##using select_best() to choose the mtry value with the best auc
rand_for_final <-
  rand_for_tune_results %>%
  select_best(metric = "roc_auc")

##adding the final(tuned) parameter to the workflow to finalize
rand_for_workflow <-
  rand_for_workflow %>%
  finalize_workflow(rand_for_final)
```

Now we are ready to fit our model finally! The handy last_fit() function allows us to fit the model on the training set, evaluate it on the test set and returns a tibble of predicted values and evaluation metrics.

```
rand_for_fit <-
  last_fit(rand_for_workflow, split_dat)

rand_for_fit
```

```
## # Resampling results
## # Manual resampling
## # A tibble: 1 x 6
##   splits       id         .metrics       .notes         .predictions     .workflow
##   <list>       <chr>      <list>         <list>         <list>           <list>
## 1 <split [22… train/test… <tibble[,4] […  <tibble[,1]… <tibble[,6] [75… <workflo…
```

Let's see how our model performed.

```
##collecting metrics to evaluate the model
collect_metrics(rand_for_fit)
```

```
## # A tibble: 2 x 4
##   .metric  .estimator .estimate .config
##   <chr>    <chr>          <dbl> <chr>
## 1 accuracy binary         0.84  Preprocessor1_Model1
## 2 roc_auc  binary         0.920 Preprocessor1_Model1
```

Those performance metrics look excellent! Let's look at a breakdown of predicted values with actual values.

```
##collecting the model's prediction for each patient
model_predictions <-
  collect_predictions(rand_for_fit)

##returning a confusion matrix of the model's predictions and actual results
model_predictions %>%
  conf_mat(target, .pred_class)
##          Truth
## Prediction  0  1
##          0 26  6
##          1  6 37
```

Now lets try to figure out how we got there. First we'll train our model on the entire data, and save that model object which we can use on future data.

```
final_model <-
  fit(rand_for_workflow, dat)
```

Now we can extract the fit object from the data and examine the variable importance.

```
##extracting the fit object
model_obj <-
  pull_workflow_fit(final_model)$fit

model_obj
```

```
## Ranger result
##
## Call:
##  ranger::ranger(x = maybe_data_frame(x), y = y, mtry = min_cols(~2,      x), importance = ~"impurity"
##
## Type:                             Probability estimation
## Number of trees:                  500
## Sample size:                      303
## Number of independent variables:  13
## Mtry:                             2
## Target node size:                 10
## Variable importance mode:         impurity
## Splitrule:                        gini
## OOB prediction error (Brier s.):  0.1251956
```

```
##examining variable importance
model_obj$variable.importance
```

```
##       age       sex        cp  trestbps      chol       fbs   restecg   thalach
##  8.390093  4.562836 16.611129  7.434131  7.057686  1.034021  1.963367 13.516701
##     exang   oldpeak     slope        ca      thal
##  7.517539 12.800440  6.314910 13.228183 12.975627
```