P8106 Midterm Project

Paula Wu

```
library(tidyverse)
library(viridis)
```

Read in data (source: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset)

13 predictors, 1 binary response variable (target: 1 - presence, 0 - absence)

```
read_csv("./data/heart.csv") %>%
  mutate(sex = as.factor(sex),  # 0 = F, 1 = M
         cp = as.factor(cp),
         fbs = as.factor(fbs), # 0 = False, 1 = True
         exang = as.factor(exang), # 0 = no, 1 = yes
         thal = as.factor(thal),  # 0 = normal, 1 = fixed defect, 2 = reversable defect
         target = as.factor(target)) # 0 = absence, 1 = presence
```

```
## # A tibble: 1,025 x 14
##      age sex   cp    trestbps  chol fbs   restecg thalach exang oldpeak slope
##    <dbl> <fct> <fct>    <dbl> <dbl> <fct>   <dbl>   <dbl> <fct>   <dbl> <dbl>
##  1    52 1     0          125   212 0           1     168 0         1       2
##  2    53 1     0          140   203 1           0     155 1         3.1     0
##  3    70 1     0          145   174 0           1     125 1         2.6     0
##  4    61 1     0          148   203 0           1     161 0         0       2
##  5    62 0     0          138   294 1           1     106 0         1.9     1
##  6    58 0     0          100   248 0           0     122 0         1       1
##  7    58 1     0          114   318 0           2     140 0         4.4     0
##  8    55 1     0          160   289 0           0     145 1         0.8     1
##  9    46 1     0          120   249 0           0     144 0         0.8     2
## 10    54 1     0          122   286 0           0     116 1         3.2     1
## # ... with 1,015 more rows, and 3 more variables: ca <dbl>, thal <fct>,
## #   target <fct>
```