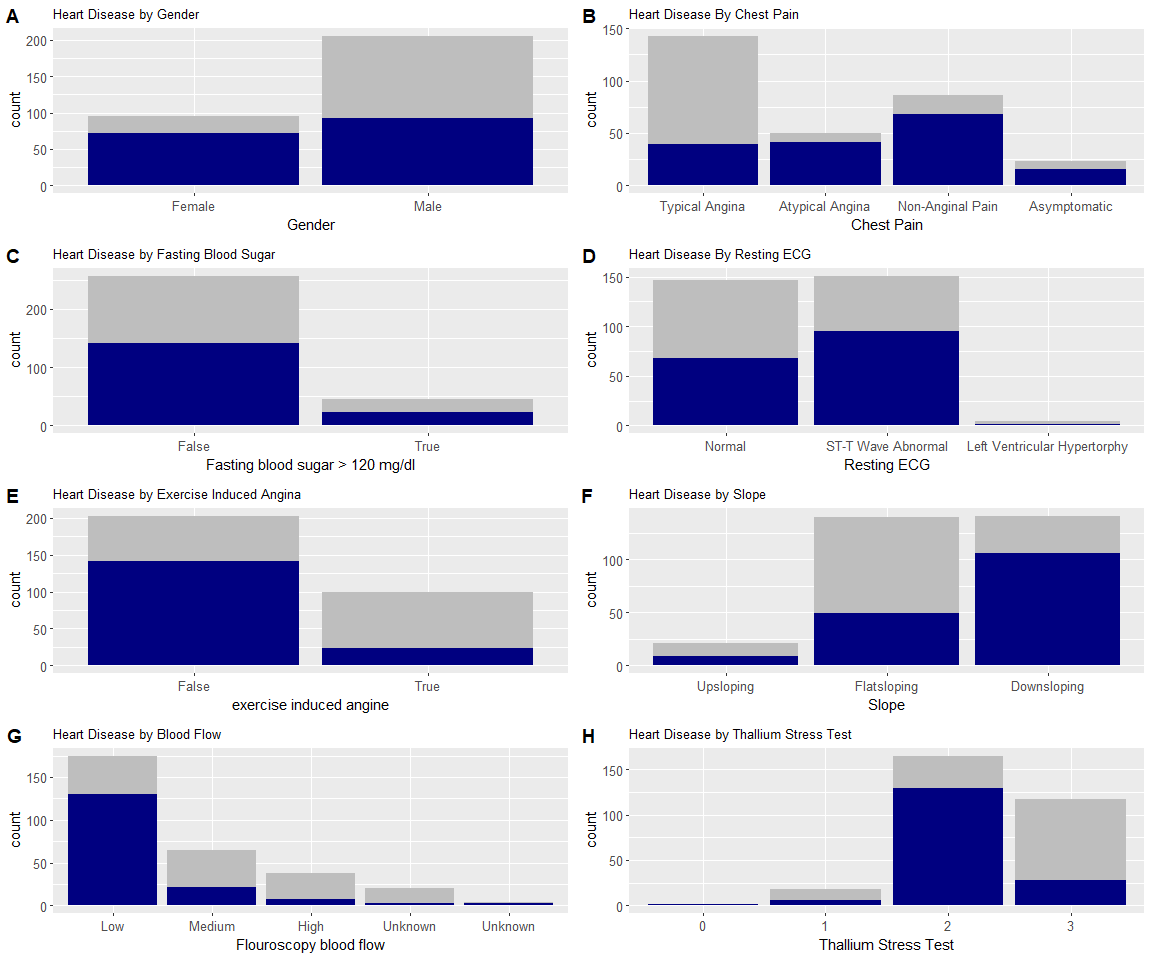
Heart Disease Project 1

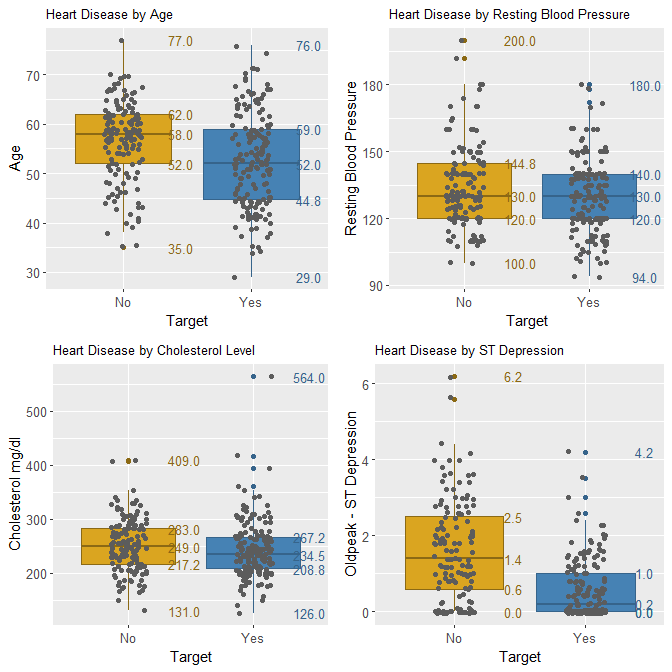
Crystal Sawtelle

2023-02-01



cowplot::plot\_grid(legend)

- The first graph(s) are exploratory analysis on the categorical variables against the - For the heart disease by gender there is ### males represented in the data and ### females. It appears that a larger percentage of the females have heart disease compared to the males. There is also twice as many males in the study then females. - For heart disease by chest pain the larger amount, ###, of subjects appear to have typical angina, meaning the subjects have chest pain related to decrease blood supply to the heart. Although, there is only ### of the subjects who have heart disease, or %%% that show typical angina and have heart disease. Atypical angina is chest pain that is not related to the heart. This group has the most subjects with heart disease making me think that there many be a -



The bar graphs are also a

## Appendix

knitr::opts\_chunk$set(echo = TRUE)  
if (!require(tidyverse)) install.packages('tidyverse')  
if (!require(readr)) install.packages('readr')  
if (!require(ggfortify)) install.packages('ggfortify')  
if (!require(gt)) install.packages('gt')  
if (!require(misty)) install.packages('misty')  
library(misty)  
library(gt)  
library(dplyr)  
library(heatmaply)  
library(ggplot2)  
library(flextable)  
library(officer)  
library(HeartDisease)  
#View(HeartDisease)  
# Heart Disease annotation   
HD\_var\_data <- data.frame(  
 Variable = c("age", "gender", "cp", "trestbps", "chol", "fbs", "restecg", "thalach", "exang", "oldpeak", "slope", "ca", "thal", "target"),  
   
 Description = c("age of patient",   
 "0 = Female\n1 = Male",   
 "chest pain:\n 0 = Typical Angina: chest pain related to decrease blood supply to the heart\n 1 = Atypical Angina: chest pain not related to the heart\n 2 = non-Anginal Pain: typical esophageal spasms (not heart related)\n 3 = Asymptomatic: chest pain not showing signs of heart disease",   
 "Resting blood pressure (in mm Hg on admission to the hospital) anything above 130-140 is typically cause for concern",   
 "Cholesterol: serum cholesterol in mg/dl (milligrams per deciliter)\n serum = LDL + HDL + .2\*triglycerides",   
 "Fasting blood sugar: > 120 mg/dl (milligrams per deciliter)\n 1 = true\n 0 = false\n (> 126 mg/dl signals diabetes)",   
 "Resting electrocardiographic (EKG or ECG):\n 0 = nothing to note\n 1 = ST-T Wave abnormality (can range from mile symptoms to severe problems, signals non-normal heart beat)\n 2 = Possible or definite left ventricular hypertrophy (enlarged heart's main pumping chamber)",   
 "Maximum heart rate achieved",   
 "Exercise induced angina\n 1 = true\n 0 = false",   
 "ST depression induced by exercise relative to rest looks at stress of heart during exercise (unhealthy heart will stress more)",   
 "The slope of the peak exercise ST segment\n 0 = Upsloping: better heart rate with exercise (uncommon)\n 1 = Flatsloping: minimal change (typical healthy heart)\n 2 = Downsloping: signs of unhealthy heart",   
 "Number of major vessels (0-3) colored by fluoroscopy (procedure to see blood flow)",   
 "Thallium stress test: (a nuclear imaging test that shows how well blood flows into the heart while you are exercising or at rest)\n 1 = normal\n 2 = fixed defect: used to be defect but now is okay\n 3 = reversable defect: no proper blood movement when exercising",   
 " 0 = do not have heart disease\n 1 = have heart disease")  
)  
# Print annoation  
ft <- qflextable(HD\_var\_data)  
read\_docx() %>%  
 body\_add\_flextable(value = ft) %>%  
 print(target = "C:/Users/dutch/OneDrive/Desktop/HD\_var\_data.docx")  
ft  
# Refactor heart disease target from 0/1 to No/Yes and rename target to Heart\_Disease  
HD\_target <- HeartDisease %>%  
 mutate(target = ifelse(target == 0, "No", "Yes")) %>%  
 rename\_at("target", ~ "Heart\_Disease")  
# Heart Disease target count  
target\_graph <- ggplot(HeartDisease, aes(target)) +  
 geom\_bar() +  
 ggtitle("Count of people with heart disease")  
#target\_graph %>% plotly::ggplotly()  
# Heart\_Disease? vs gender  
gender\_graph <- ggplot(HD\_target, aes(gender)) +  
 geom\_bar(aes(fill=Heart\_Disease)) +  
 scale\_fill\_manual(values = c("gray", 'navy')) +  
 scale\_x\_discrete(labels=c("Female", "Male" )) +  
 ggtitle("Heart Disease by Gender") +  
 xlab("Gender") +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none")  
# Heart\_Disease? vs chest pain  
cp\_graph <- ggplot(HD\_target, aes(cp)) +  
 geom\_bar(aes(fill = Heart\_Disease)) +  
 scale\_fill\_manual(values = c("gray", 'navy')) +  
 scale\_x\_discrete(labels=c("Typical Angina", "Atypical Angina", "Non-Anginal Pain", "Asymptomatic" )) +  
 ggtitle("Heart Disease By Chest Pain") +  
 xlab("Chest Pain") +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none")  
# Heart\_Disease? vs fasting blood sugar  
fbs\_graph <- ggplot(HD\_target, aes(fbs)) +  
 geom\_bar(aes(fill = Heart\_Disease)) +  
 scale\_fill\_manual(values = c("gray", 'navy')) +  
 scale\_x\_discrete(labels=c("False", "True" )) +  
 ggtitle("Heart Disease by Fasting Blood Sugar") +  
 xlab("Fasting blood sugar > 120 mg/dl") +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none")  
# Heart\_Disease? vs resting ecg  
restecg\_graph <- ggplot(HD\_target, aes(restecg)) +  
 geom\_bar(aes(fill = Heart\_Disease)) +  
 scale\_fill\_manual(values = c("gray", 'navy')) +  
 scale\_x\_discrete(labels=c("Normal", "ST-T Wave Abnormal", "Left Ventricular Hypertorphy" )) +  
 ggtitle("Heart Disease By Resting ECG") +  
 xlab("Resting ECG") +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none")  
# Heart\_Disease? vs exercise induced angina  
exang\_graph <- ggplot(HD\_target, aes(exang)) +  
 geom\_bar(aes(fill = Heart\_Disease)) +  
 scale\_fill\_manual(values = c("gray", 'navy')) +  
 scale\_x\_discrete(labels=c("False", "True" )) +  
 ggtitle("Heart Disease by Exercise Induced Angina") +  
 xlab("exercise induced angine") +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none")  
# Heart\_Disease? vs heart rate slope  
slope\_graph <- ggplot(HD\_target, aes(slope)) +  
 geom\_bar(aes(fill = Heart\_Disease)) +  
 scale\_fill\_manual(values = c("gray", 'navy')) +  
 scale\_x\_discrete(labels=c("Upsloping", "Flatsloping", "Downsloping" )) +  
 ggtitle("Heart Disease by Slope") +  
 xlab("Slope") +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none")  
# Heart\_Disease? vs blood flow  
ca\_graph <- ggplot(HD\_target, aes(ca)) +  
 geom\_bar(aes(fill = Heart\_Disease)) +  
 scale\_fill\_manual(values = c("gray", 'navy')) +  
 scale\_x\_discrete(labels=c("Low", "Medium", "High", "Unknown", "Unknown" )) +  
 ggtitle("Heart Disease by Blood Flow") +  
 xlab("Flouroscopy blood flow") +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none")  
# Heart\_Disease? vs Thallium Stress Test  
thal\_graph <- ggplot(HD\_target, aes(thal)) +  
 geom\_bar(aes(fill = Heart\_Disease)) +  
 labs(title = "Heart Disease by Thallium Stress Test", x = "Thallium Stress Test") +  
 scale\_fill\_manual(values = c("gray", "navy")) +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10))  
  
# get legend for Heart\_Disease?  
legend <- cowplot::get\_legend(thal\_graph)  
# categorical variables cowplot  
cowplot::plot\_grid(gender\_graph, cp\_graph, fbs\_graph, restecg\_graph, exang\_graph, slope\_graph, ca\_graph, thal\_graph + theme(legend.position = "none"), ncol = 2, labels = c("A", "B", "C", "D", "E", "F", "G", "H") )  
cowplot::plot\_grid(legend)  
# HeartDisease1 <- HeartDisease %>%  
# select(age, trestbps, chol, oldpeak)  
# HeartDisease2 <- HeartDisease %>%  
# select(gender, cp, fbs, restecg, exang, slope, ca, thal, target)  
# logistic <- glm(target ~ ., HeartDisease, family = "binomial")  
# summary(logistic)  
# caret::varImp(logistic)  
# pscl::pR2(logistic)["McFadden"]  
# car::vif(logistic)  
# smallest <- I(target == 1) ~ 1  
# largest <- I(target == 1) ~ age + gender + cp + trestbps + chol + fbs + restecg + thalach + exang + oldpeak + slope + ca + thal  
#   
# stepwise.glm.target = glm(I( target== 1) ~ 1, data=HeartDisease, family='binomial')  
# step(stepwise.glm.target, scope=list(lower=smallest, upper=largest))  
# hd\_logistic <- glm(target ~ thal + ca + cp + oldpeak + slope + gender + trestbps + exang, data = HeartDisease, family = "binomial")  
# summary(hd\_logistic)  
# caret::varImp(hd\_logistic)  
# pscl::pR2(hd\_logistic)["McFadden"]  
# car::vif(hd\_logistic)  
# log\_model <- glm(target ~ age, data = HeartDisease, family = binomial)  
# pred\_data <- data.frame(age = seq(min(HeartDisease$age), max(HeartDisease$age), len = 300))  
# pred\_data$target <- predict(log\_model, pred\_data, type = 'response')  
# plot(target ~ age, data = HeartDisease)  
# lines(target ~ age, pred\_data, lwd = 2, col = "navy")  
# ggplot(HeartDisease, aes(x=age, y=target)) +  
# geom\_point() +  
# geom\_smooth(method = 'glm', se=TRUE, method.args = list(family = 'binomial'))  
  
# Heart\_Disease? vs age  
age\_hist <- ggplot(HeartDisease, aes(age, fill = target)) +  
 geom\_histogram(bins = 15) +  
 scale\_fill\_manual(values = c("gray", 'navy'))  
# Heart\_Disease? vs age  
age\_box <- ggplot(HD\_target, aes(x=Heart\_Disease, y = age, col = Heart\_Disease)) +  
 geom\_boxplot(fill = c("goldenrod", "steelblue")) +   
 geom\_jitter(width = 0.15, col = "gray36") +  
 scale\_x\_discrete(labels=c("No", "Yes" )) +  
 scale\_color\_manual(values = c("goldenrod4", "steelblue4"))+  
 stat\_summary(geom = "text", fun = quantile,  
 aes(label = sprintf("%1.1f", after\_stat(y)), color = Heart\_Disease),  
 position = position\_nudge(x = 0.45), size = 3.5) +  
 ggtitle("Heart Disease by Age") +  
 xlab("Target") +  
 ylab("Age")+  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none" )  
bps\_hist <- ggplot(HeartDisease, aes(trestbps, fill = target)) +  
 geom\_histogram(bins = 15) +  
 scale\_fill\_manual(values = c("gray", 'navy'))  
  
#bps\_hist %>% ggplotly()  
# Heart\_Disease? vs resting blood pressure  
bps\_box <- ggplot(HD\_target, aes(x=Heart\_Disease, y = trestbps, col = Heart\_Disease)) +  
 geom\_boxplot(fill = c("goldenrod", "steelblue")) +   
 geom\_jitter(width = 0.15, col = "gray36") +  
 scale\_x\_discrete(labels=c("No", "Yes" )) +  
 scale\_color\_manual(values = c("goldenrod4", "steelblue4"))+  
 stat\_summary(geom = "text", fun = quantile,  
 aes(label = sprintf("%1.1f", after\_stat(y)), color = Heart\_Disease),  
 position = position\_nudge(x = 0.45), size = 3.5) +  
 ggtitle("Heart Disease by Resting Blood Pressure") +  
 xlab("Target") +  
 ylab("Resting Blood Pressure") +  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none" )  
chol\_hist <- ggplot(HeartDisease, aes(chol, fill = target)) +  
 geom\_histogram(bins = 15) +  
 scale\_fill\_manual(values = c("gray", 'navy'))  
  
#chol\_hist %>% ggplotly()  
# Heart\_Disease? vs cholesterol levels  
chol\_box <- ggplot(HD\_target, aes(x=Heart\_Disease, y = chol, col = Heart\_Disease)) +  
 geom\_boxplot(fill = c("goldenrod", "steelblue")) +   
 geom\_jitter(width = 0.15, col = "gray36") +  
 scale\_x\_discrete(labels=c("No", "Yes" )) +  
 scale\_color\_manual(values = c("goldenrod4", "steelblue4"))+  
 stat\_summary(geom = "text", fun = quantile,  
 aes(label = sprintf("%1.1f", after\_stat(y)), color = Heart\_Disease),  
 position = position\_nudge(x = 0.45), size = 3.5) +  
 ggtitle("Heart Disease by Cholesterol Level") +  
 xlab("Target") +  
 ylab("Cholesterol mg/dl")+  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none" )  
old\_hist <- ggplot(HeartDisease, aes(oldpeak, fill = target)) +  
 geom\_histogram(bins = 15) +  
 scale\_fill\_manual(values = c("gray", 'navy'))  
  
#old\_hist %>% ggplotly()  
# Heart\_Disease? vs oldpeak  
old\_box <- ggplot(HD\_target, aes(x=Heart\_Disease, y = oldpeak, col = Heart\_Disease)) +  
 geom\_boxplot(fill = c("goldenrod", "steelblue")) +   
 geom\_jitter(width = 0.15, col = "gray36") +  
 scale\_x\_discrete(labels=c("No", "Yes" )) +  
 scale\_color\_manual(values = c("goldenrod4", "steelblue4"))+  
 stat\_summary(geom = "text", fun = quantile,  
 aes(label = sprintf("%1.1f", after\_stat(y)), color = Heart\_Disease),  
 position = position\_nudge(x = 0.45), size = 3.5) +  
 ggtitle("Heart Disease by ST Depression") +  
 xlab("Target") +  
 ylab("Oldpeak - ST Depression")+  
 theme(plot.title = element\_text(size = 10), axis.text = element\_text(size = 10), legend.position = "none")  
# quantitative variables cowplot  
cowplot::plot\_grid(age\_box, bps\_box, chol\_box, old\_box)