Heart Disease Visualization

Mingwei Wu

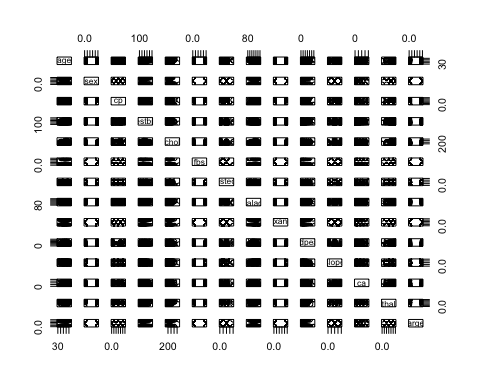
#Background ## The database contains 76 attributes, but all published experiments refer to using a subset of 14 of them, at particular, the Cleveland database is

##Data Description ###age: patients’ age ###sex: 1 is male, 2 is female ###cp: chest pain type 4 level ###trestbps: resting blood presure ###chol: serum cholestoral in mg/dl ###fbs: fasting blood sugar >120 mg/dl is value 1 ###restecg: resting electrocardiographic reults (values 0,1,2) ###thalach: maximum heart rate achieved ###exang: exercise incuced angina ###oldpeak: ST depression induced by exercise relative to rest ###slope: the slope of the peak exercise ST segment ###ca: number of major vessels (0-3) colored by flourosopy ###thal: 3 is normal; 6 fixed defect; 7 reversable defect ###target: 1 has heart disease, 0 not

data<-read.csv("heart.csv",header = TRUE)  
head(data)

## age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal  
## 1 63 1 3 145 233 1 0 150 0 2.3 0 0 1  
## 2 37 1 2 130 250 0 1 187 0 3.5 0 0 2  
## 3 41 0 1 130 204 0 0 172 0 1.4 2 0 2  
## 4 56 1 1 120 236 0 1 178 0 0.8 2 0 2  
## 5 57 0 0 120 354 0 1 163 1 0.6 2 0 2  
## 6 57 1 0 140 192 0 1 148 0 0.4 1 0 1  
## target  
## 1 1  
## 2 1  
## 3 1  
## 4 1  
## 5 1  
## 6 1

pairs(data) #paris data to see the relationship in numeric values



library(ggplot2)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ tibble 3.0.4 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.0  
## ✓ purrr 0.3.4

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

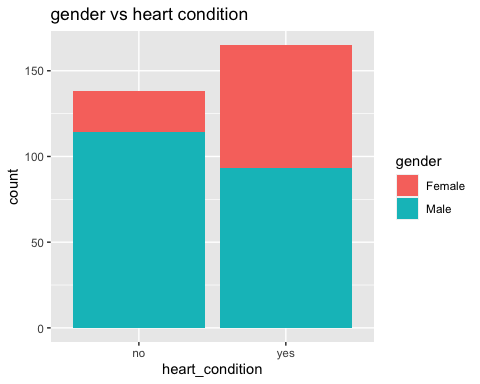
library(dplyr)  
heart<-data%>%  
 mutate(gender= ifelse(sex==1,"Male","Female"),  
 chest\_pain\_level= ifelse(cp==0,"normal",  
 ifelse(cp==1,"mild",  
 ifelse(cp==2,"moderate","severe"))),  
 fblood\_sugar=ifelse(fbs==1,">120","<=120"),  
 rest\_electrocardigoraphic= ifelse(restecg==0,"normal",  
 ifelse(restecg==1,"abnormalily","definite")),  
 exercise=ifelse(exang==1,"yes","no"),  
 heart\_condition=ifelse(target==1,"yes","no")) # rebuild the column to the data frame

heart%>%  
 group\_by(gender)%>%  
 summarise(gender\_rate=mean(target)) # calculate the heart disease rate of the gender

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 2 x 2  
## gender gender\_rate  
## <chr> <dbl>  
## 1 Female 0.75   
## 2 Male 0.449

heart%>%  
 ggplot(aes(heart\_condition,fill=gender))+geom\_bar()+ggtitle("gender vs heart condition") #data visualization to display the relationship

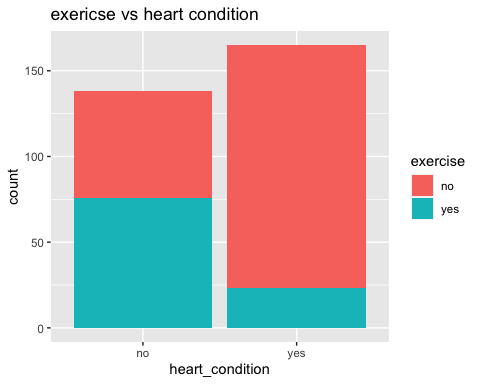


heart%>%  
 group\_by(exercise)%>%  
 summarise(exercise\_rate=mean(target))

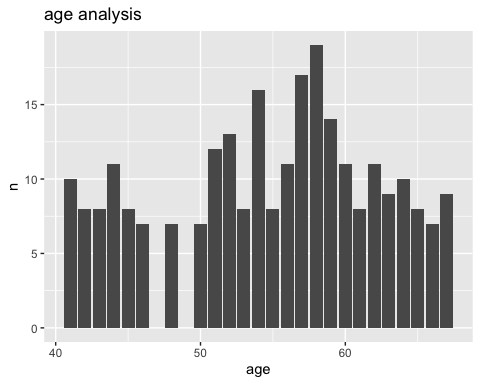
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 2 x 2  
## exercise exercise\_rate  
## <chr> <dbl>  
## 1 no 0.696  
## 2 yes 0.232

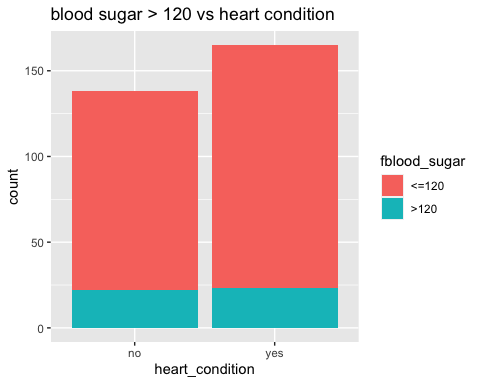
heart%>%  
 ggplot(aes(heart\_condition,fill=exercise))+geom\_bar()+ggtitle("exericse vs heart condition")



heart%>%  
 group\_by(age)%>%  
 count()%>%  
 filter(n>5)%>%  
 ggplot(aes(age,n))+geom\_col()+ggtitle("age analysis") # checking a proportion of heart condition with age

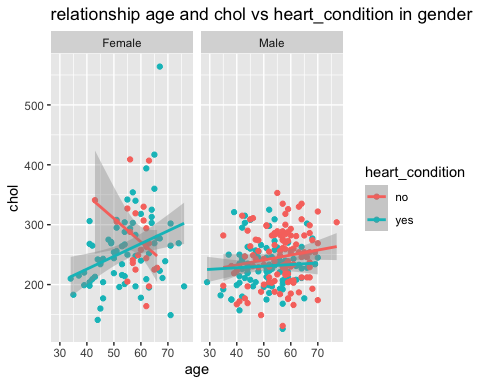


heart%>%  
 ggplot(aes(heart\_condition,fill=fblood\_sugar))+geom\_bar()+ggtitle("blood sugar > 120 vs heart condition")



heart%>%  
 ggplot(aes(age,chol,color=heart\_condition))+geom\_point()+geom\_smooth(method="lm")+ggtitle("relationship age and chol vs heart\_condition in gender")+ facet\_wrap(~gender) #plot the point and line for data visualization

## `geom\_smooth()` using formula 'y ~ x'



heart%>%  
 ggplot(aes(gender,trestbps))+geom\_boxplot()+xlab("Sex")+ylab("resting blood pressure")+facet\_grid(~chest\_pain\_level)

