

Heart Disease Visualization

Mingwei Wu

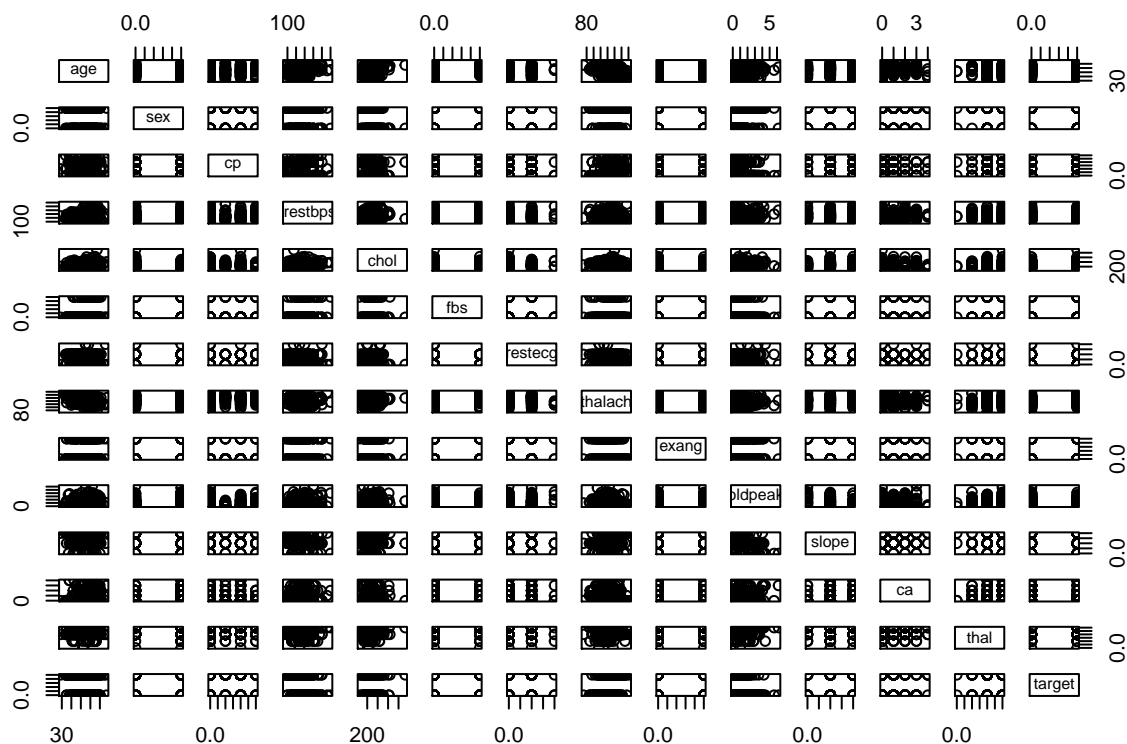
#Background ## The database contains 76 attributes, but all published experiments refer to using a subset of 14 of them, at particular, the Cleveland database is

##Data Description ###age: patients' age ###sex: 1 is male, 2 is female ###cp: chest pain type 4 level ###trestbps: resting blood pressure ###chol: serum cholestoral in mg/dl ###fbs: fasting blood sugar >120 mg/dl is value 1 ###restecg: resting electrocardiographic results (values 0,1,2) ###thalach: maximum heart rate achieved ###exang: exercise induced 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) #pairs data to see the relationship in numeric values
```



```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

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

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

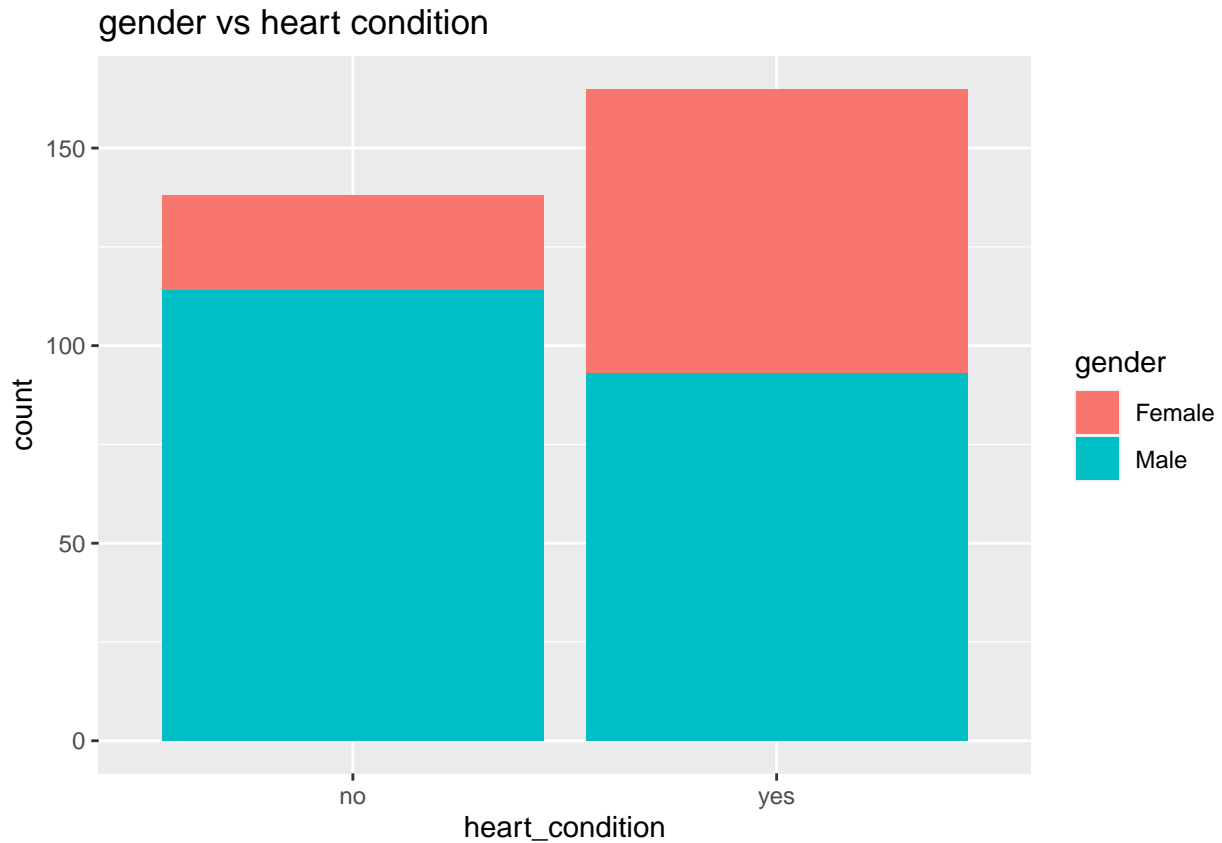
```
library(dplyr)
heart<-data%>%
  mutate(gender= ifelse(sex==1,"Male","Female"),
         chest_pain_level= ifelse(cp==0,"normal",
                                   ifelse(cp==1,"mild",
                                           ifelse(cp==2,"moderate","severe"))),
         fblood_sugar=ifelse(fbs==1,">120","<=120"),
         rest_electrocardiographic= 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 visual
```



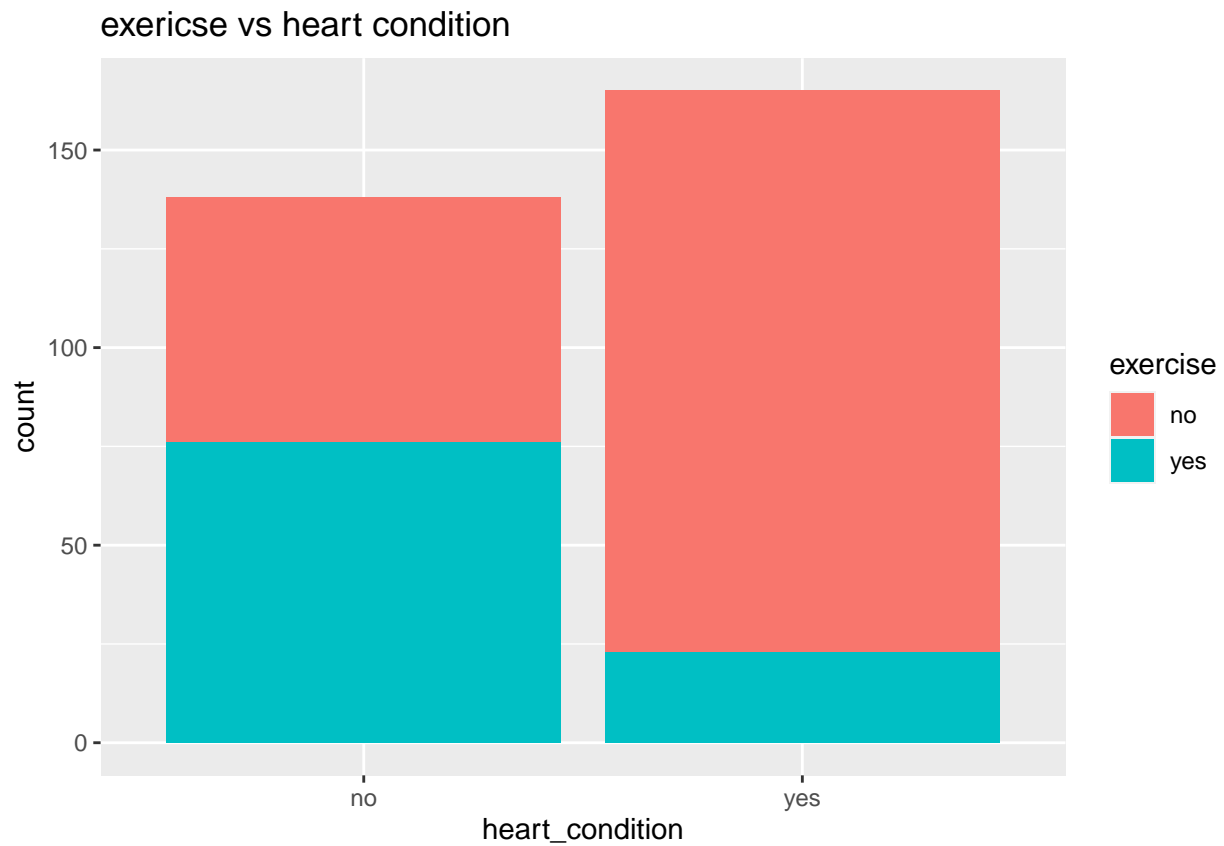
```
heart%>%
  group_by(exercise)%>%
  summarise(exercise_rate=mean(target))
```

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

```
## # A tibble: 2 x 2
##   exercise exercise_rate
```

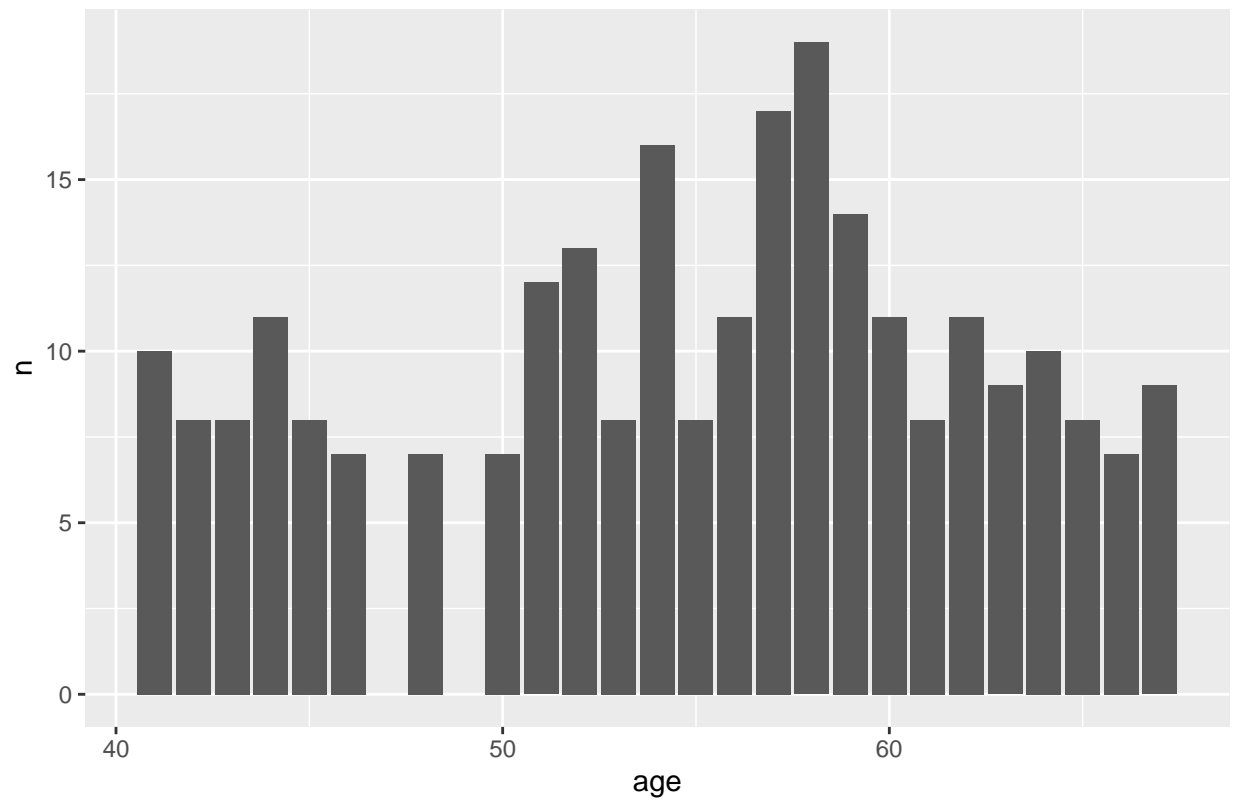
```
##   <chr>          <dbl>
## 1 no            0.696
## 2 yes           0.232
```

```
heart%>%
  ggplot(aes(heart_condition,fill=exercise))+geom_bar()+ggtitle("exericse vs heart condition")
```

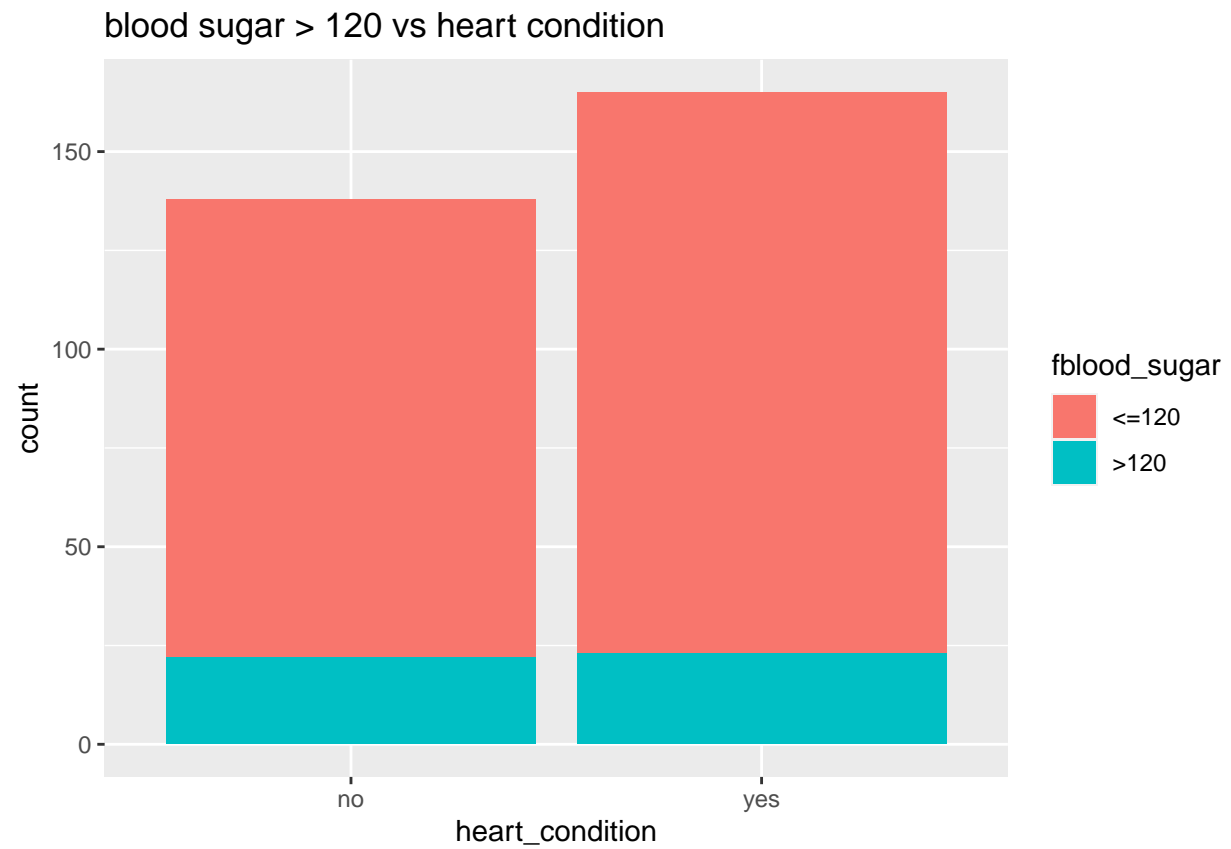


```
heart%>%
  group_by(age)%>%
  count()%>%
  filter(n>5)%>%
  ggplot(aes(age,n))+geom_col()+ggtitle("age analysis") # checking a proportion of heart condition with
```

age analysis

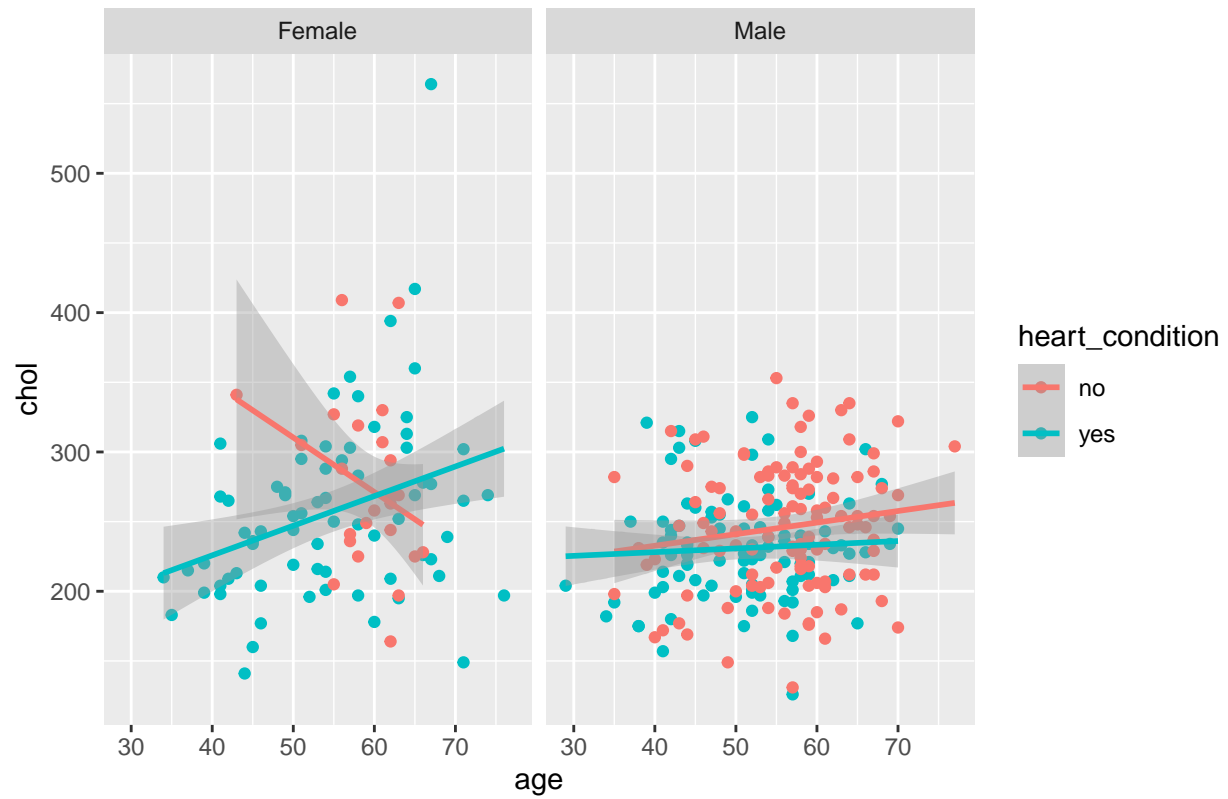


```
heart%>%  
  ggplot(aes(heart_condition,fill=fblood_sugar))+geom_bar()+ggtitle("blood sugar > 120 vs heart condition")
```



```
heart%>%  
  ggplot(aes(x=age,y=chol,color=heart_condition))+geom_point()+geom_smooth(method="lm")+ggtitle("relation between age and cholesterol")  
  
## `geom_smooth()` using formula 'y ~ x'
```

relationship age and chol vs heart_condition in gender



```
heart%>%
  ggplot(aes(gender,trestbps))+geom_boxplot()+xlab("Sex")+ylab("resting blood pressure")+facet_grid(~chol)
```

