Explore the Heart disease data

Huy N. Pham

Install helpful library for analysis.

```
# Install required library, need not to re-install if you already have
# install.packages("dplyr")
# install.packages("Hmisc")
# install.packages("ggplot2")

# Load the require library
# library(dplyr)
# library(Hmisc)
# library(ggplot2)
```

Data source: https://raw.githubusercontent.com/pnhuy/datasets/master/heart_uci/heart.csv

1 Data Loading & Exploratory data analysis

- Load the data in csv file and store to variable name `data`.
- Show some rows of data to get some insight of data.
- Remove the first columns because it is not useful for analysis.
- What was the average age?
- From `sex` column, create new variable `gender` which only have 'male', 'female'?
- How many percent of patient who was male were there in the data?
- How many percent of male patient who suffered heart disease were there in the data? What about female?
- What was the range of RestBP?
- What was the distribution of AHD?
- Calculate some basic desciptive statistics
- Plot the distribution of AHD?
- Plot the distribution of RestBP per Sex?
- Illustrate the relationship between sex and AHD?
- Illustrate the relationship between MaxHR and AHD?
- Illustrate the relationship between Age and MaxHR?
- Illustrate the relationship among the continuous variables and AHD?

2 Hypothesis testing

- Compare the mean RestBP with normal BP (120)?
- Compare mean RestBP by AHD?
- Test the independence between AHD and Sex?
- Compare mean MaxHR by Thal

3 Linear Regression

- Build a model to predict MaxHR by Age?
- Build a model to predict MaxHR by Age & RestBP & Thal?