STA303 PROJECT PROPOSAL

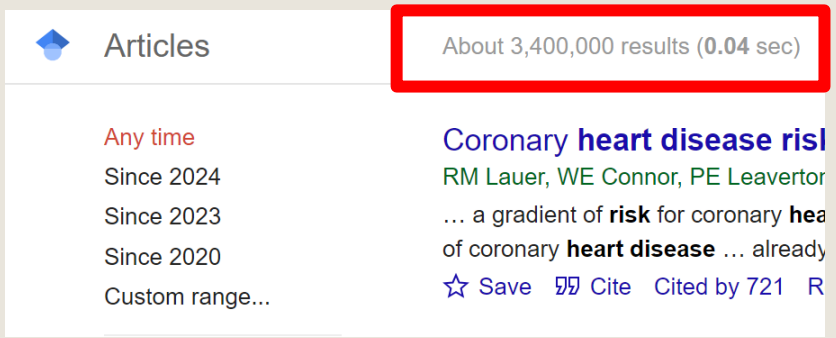
Li Quan Soh (utorID: sohli)

RESEARCH QUESTION

Does unhealthy lifestyle, characterized by risk factors:

- Resting Blood Pressure
- Cholesterol Level
- Fasting Blood Sugar Level

increase the risk of developing **heart diseases**?



Why is this interesting? How is it novel?

- Heart diseases and its risk factors has been intensively studied
- **My focus**: combination of these risk factors, along with these confounders
  - Age
  - Sex

ARTICLE #1

Blood pressure and cholesterol as coronary heart disease (CHD) risk factors

(Tyroler, 1971)

Research Question:

- How does blood pressure & cholesterol affect incidence of CHD in Georgia?

Main Findings:

- ↑ cholesterol level & ↑ blood pressure → ↑ CHD risk
- Black men more likely to get CHD than white men

How it influence my analysis:

- Exposure variable selection: Cholesterol & Blood Pressure

ARTICLE #2

Effects of lifestyle-related factors on ischemic heart disease according to body mass index and fasting blood glucose levels in Korean adults

(Shin et al., 2019)

Research Question:

- How does lifestyle-related factors affect the risk of ischemic heart disease (IHD)?

Main Findings:

- Life-style related risk factors, measured by
  - ↑ BMI
  - ↑ Fasting Blood Glucose (FBG) Level

  increases the risk of IHD
- Effect of risk factors more pronounced in women

How it influence my analysis:

- Exposure variable selection: Fasting Blood Sugar Level
- Confounder selection: Age & Sex

ARTICLE #3

Risk factors for coronary heart disease: implications of gender

(Roeters van Lennep et al., 2002)

Research Question:

- Does sex affect the impact of cardiovascular risk factors (diabetes, lipoprotein level etc)?

Main Findings:

- Most risk factors have **no** sex-specific difference
- Except
  - diabetes, HDL (cholesterol) level, triglycerides increases the risk of heart diseases in women than men
  - lipoprotein level increases the risk of heart diseases in men than women

How it influence my analysis:

- Exposure variable selection: Fasting BS & Cholesterol level
- Confounder selection: Sex

HEART DISEASE PREDICTION DATASET

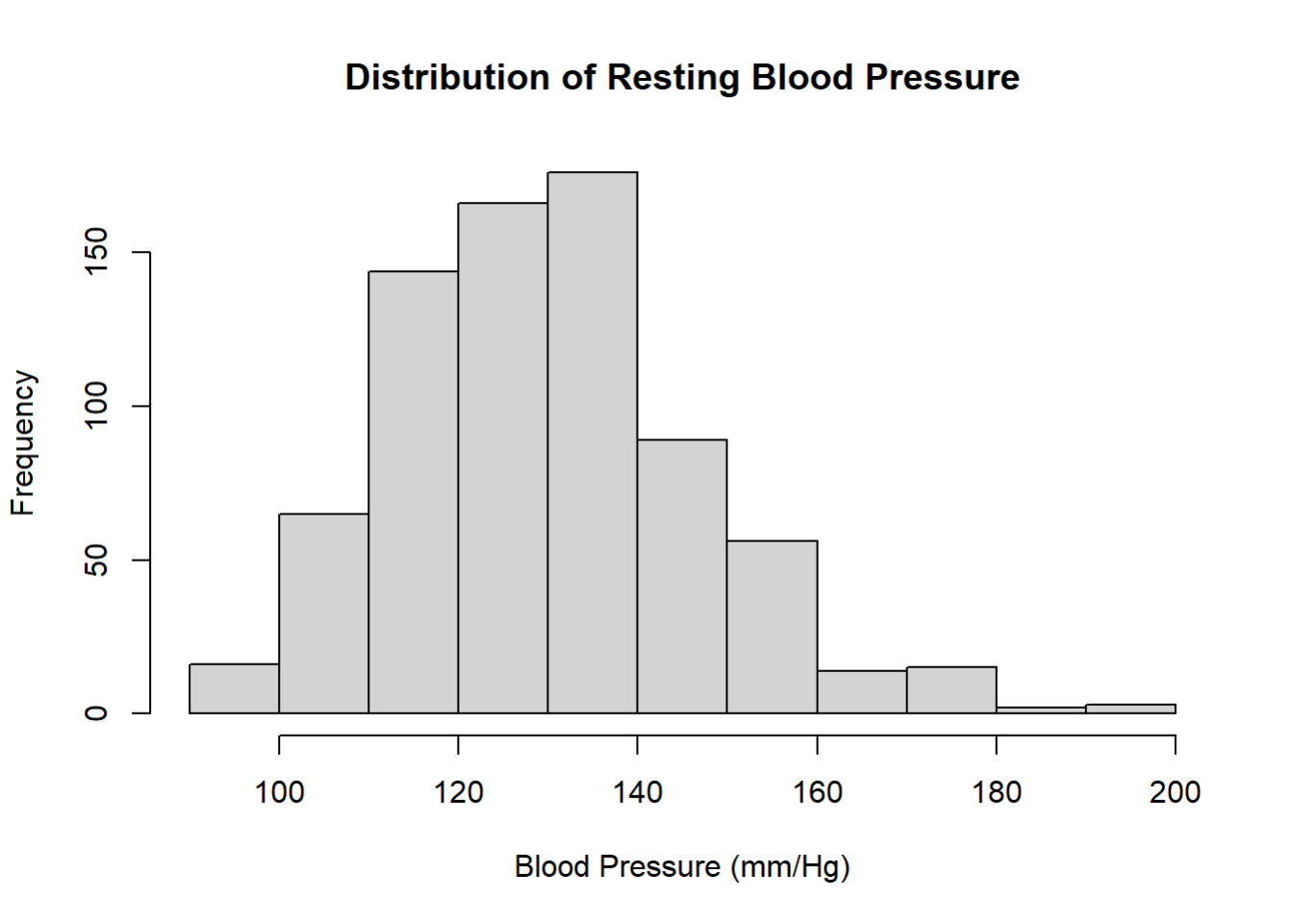
Key Features

- Dataset curated from 5 datasets
- 918 observation (human individuals)
  - After cleaning: 746 observations
- No missing values or duplicated observations
- 12 variables:

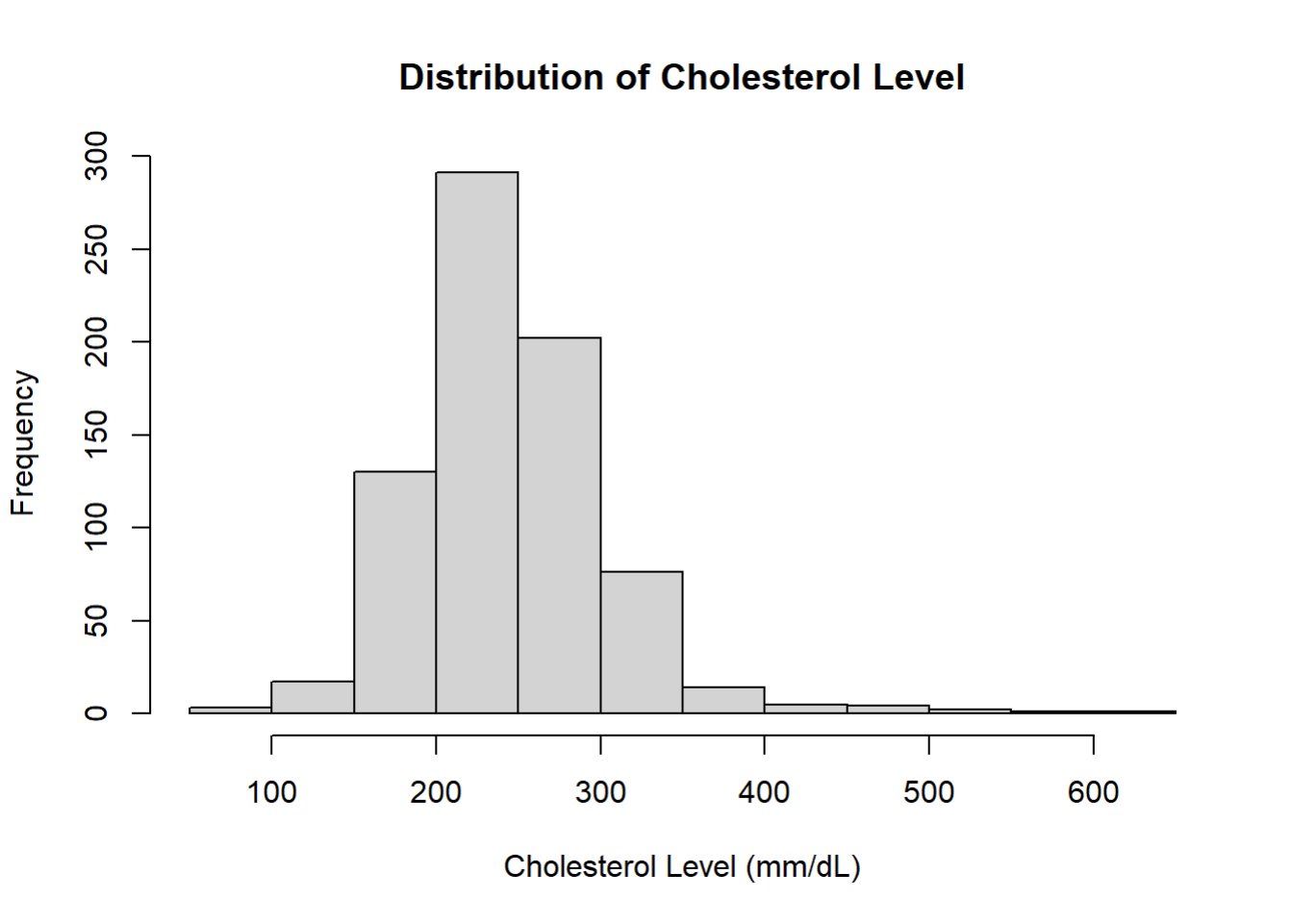
  - Age
  - Sex
  - Chest Pain Type
  - Resting Blood Pressure
  - Cholesterol Level
  - Fasting Blood Sugar Level
  - Resting ECG results
  - Max Heart Rate
  - Exercised-induced Angina
  - Oldpeak
  - ST Slope
  - Heart Disease

Color Code
Red: Outcome
Blue: Risk factors/ Predictors of Interest
Green: Confounders

Predictors of Interest: Resting Blood Pressure



- Unit: mmHg
- Continuous variable
- Mean: 133 mmHg
- Min: 92 mmHg
- Max: 200 mmHg
- Note: 1 observation with 0 mmHg. Removed as it's invalid.



Predictor of Interest: Cholesterol Level

- Unit: mm/dL
- Continuous variable
- Mean: 244.6 mm/dL
- Min: 85 mm/dL
- Max: 603 mm/dL
- Note: 172 observation with 0mm/dL. Removed as this value is invalid.

Distribution of FBS Level

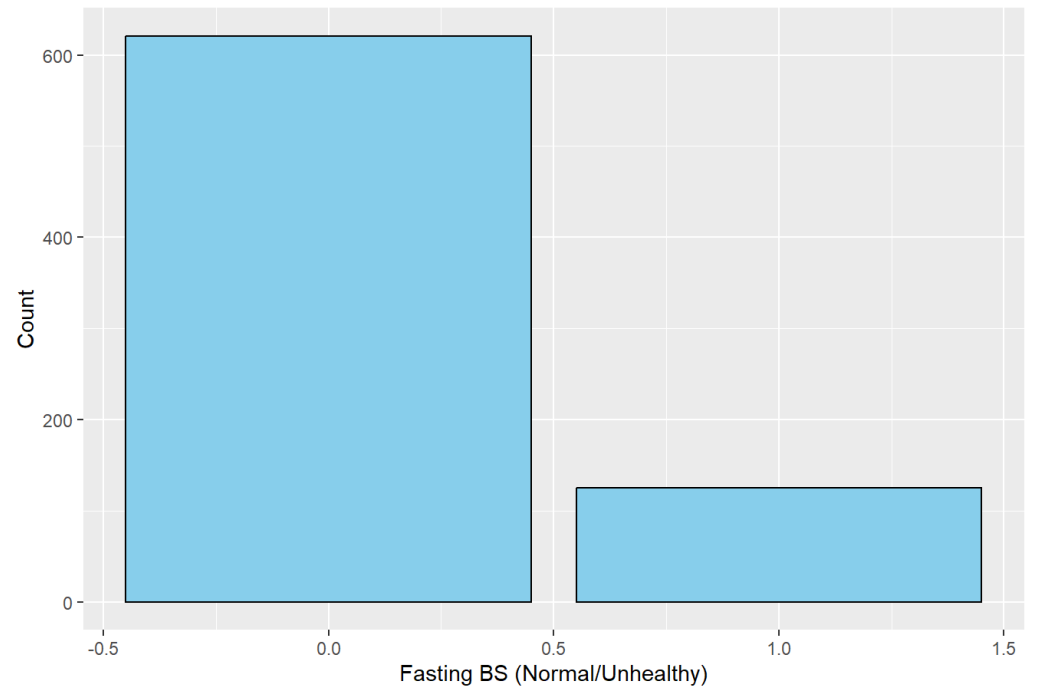
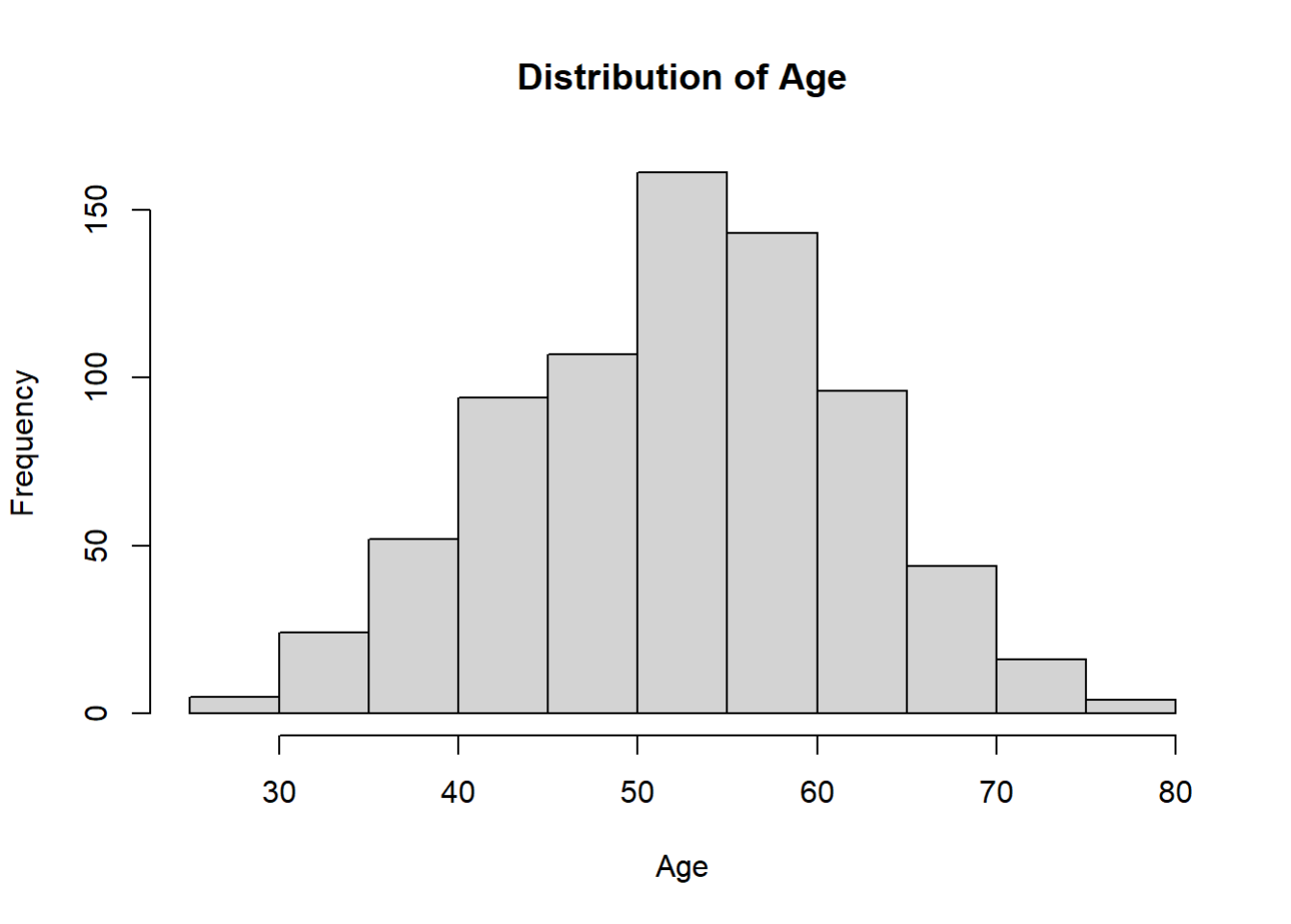


Table: Heart Disease vs FBS Level

	Healthy	Unhealthy	Total
No HD	347	43	390
Have HD	274	82	356
Total	621	125	746

PREDICTOR OF INTEREST: FASTING BLOOD SUGAR (FBS) LEVEL

- Nominal categorical variable
- Level:
  - Healthy FBS (≤ 120 mg/dL)
  - Unhealthy FBS (> 120 mg/dL)
- Unbalanced
  - 83.24% of observations have healthy FBS level
- 44% of individuals with healthy FBS level have HD
- 66% of individuals with unhealthy FBS level have HD



CONFOUNDER: AGE

- Discrete variable
- Mean: ~ 53.5 yo
- Min: 28 yo
- Max: 77 yo

Distribution of Sex

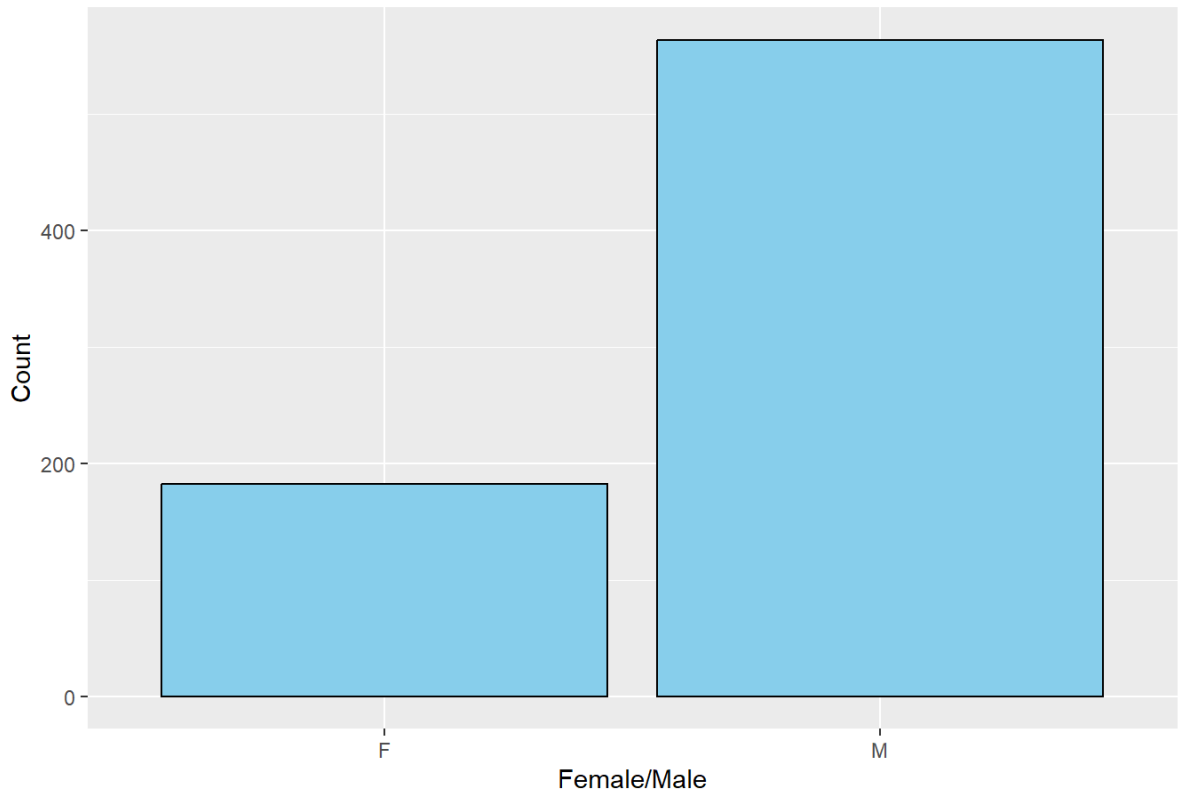
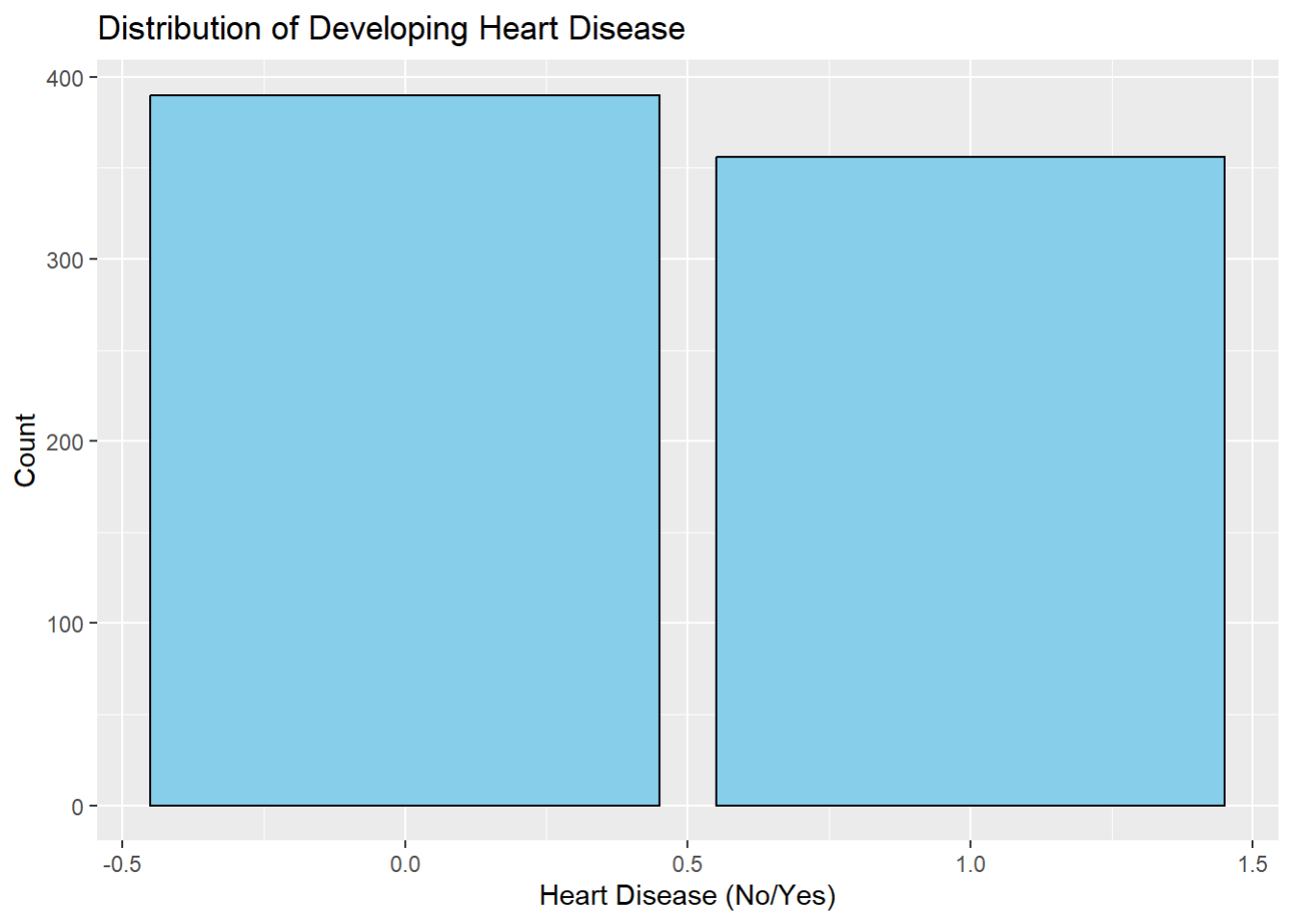


Table: Heart Disease vs Sex

	Female	Male	Total
No HD	142	248	390
Have HD	40	316	356
Total	182	564	746

CONFOUNDER: SEX

- Nominal categorical variable
- Level:
  - Female
  - Male
- Unbalanced
  - 75.6% of observations are male
- 22% of women have HD
- 56% of men have HD



OUTCOME VARIABLE: HEART DISEASE

- Nominal categorical variable
- Level:
  - No Heart Disease (Healthy)
  - Have Heart Disease
- Balanced
  - 47.7% of the individuals have heart disease

ASSUMPTIONS OF GLM

1. Independence of Y
   - ✓ Each individual are independent from another → risk of heart disease is independent
2. Distribution of Y
   - ✓ Categorical Variable: Binar Outcome → binomial distribution
   - ✓ Binomial Logistic Regression
3. Correct Link Function
   - ✓ Logit Link
   - ✓ Correct link function for binomial logistic regression

REFERENCES

fedesoriano. (2021). Heart Failure Prediction Dataset [Data set].Kaggle. https://www.kaggle.com/fedesoriano/heart-failure-prediction

Roeters van Lennep, J., Westerveld, H., Erkelens, D., & E van der Wall, E. (2002). Risk factors for coronary heart disease: Implications of gender. Cardiovascular Research, 53(3), 538–549. https://doi.org/10.1016/s0008-6363(01)00388-1

Shin, J., Ham, D., Shin, S., Choi, S. K., Paik, H.-Y., & Joung, H. (2019). Effects of lifestyle-related factors on ischemic heart disease according to body mass index and fasting blood glucose levels in Korean adults. PLOS ONE, 14(5). https://doi.org/10.1371/journal.pone.0216534

Tyroler, H. A. (1971). Blood pressure and cholesterol as coronary heart disease risk factors. Archives of Internal Medicine, 128(6), 907–914. https://doi.org/10.1001/archinte.1971.00310240061007