*The University of Texas at Austin*

*Zhaowen Fan*

*Rafael Ignacio Gonzalez Chong*

*Assignment 11*

GSI Intro to Big Data and Data Mining

Table of Contents

[Task 1: Summarize the data by each feature. Use an appropriately labelled table to show the results. Also include a graphical presentation that shows the distribution of Cholesterol  for participants vs. non-participants. Describe the shape of each distribution. Use R to calculate the quantities and generate the visual summaries. 3](#_Toc204938723)

[Task 2: Does the mean cholesterol level is less than 250? Formally test at the alpha = 0.05 level using the 5 steps outlined in the last lecture. 4](#_Toc204938724)

[Task 3: Calculate a 90% confidence interval for the mean cholesterol.  Interpret the confidence interval. 5](#_Toc204938725)

[Task 4: Formally test that resting blood pressure level is less than 130 at the alpha = 0.05 level using the 5 steps outlined in our last class. 5](#_Toc204938726)

[Task 5: Calculate a 95% confidence interval for the resting blood pressure.  Interpret the confidence interval. 6](#_Toc204938727)

[Task 6: Are the cholesterol level of the two groups with target 1 or 0 different? (Is it bigger, less or equal?) 6](#_Toc204938728)

[Task 7: Are resting blood pressure level of the two groups with target 1 or 0 different?  (Is it bigger, less or equal?) 7](#_Toc204938729)

[Task 8: Are the fasting blood sugar level of the two groups with target 1 or 0 different?  (Is it bigger, less or equal?) 7](#_Toc204938730)

[Task 9: Are the maximum heart rate level of the two groups with target 1 or 0 different?  (Is it bigger, less or equal?) 7](#_Toc204938731)

[Appendices (Code) 8](#_Toc204938732)

# **Task 1: Summarize the data by each feature. Use an appropriately labelled table to show the results. Also include a graphical presentation that shows the distribution of Cholesterol  for participants vs. non-participants. Describe the shape of each distribution. Use R to calculate the quantities and generate the visual summaries.**

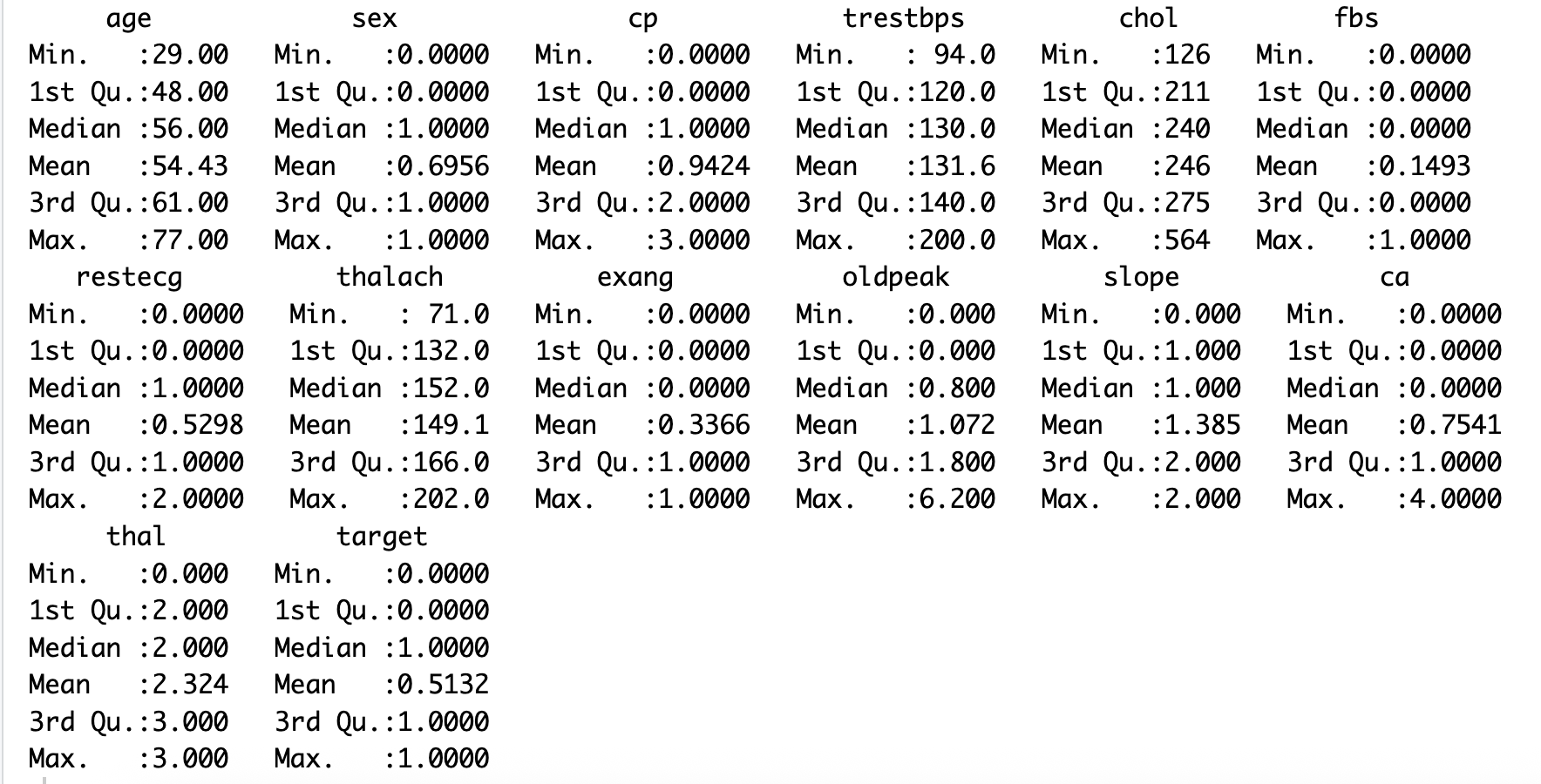
****

Fig. 1 Summary of each Feature.

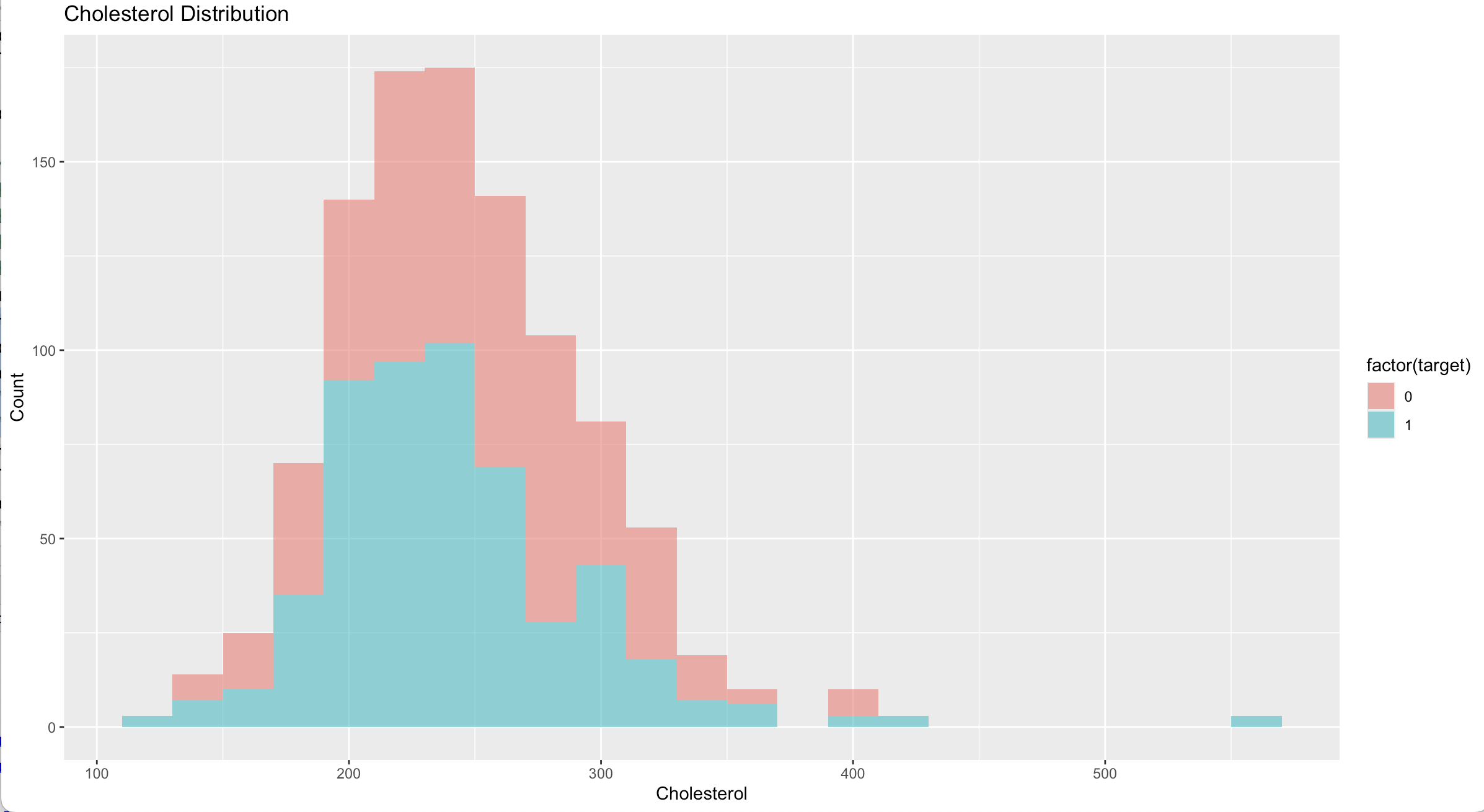


Fig. 2 Histogram of Cholesterol Distribution .

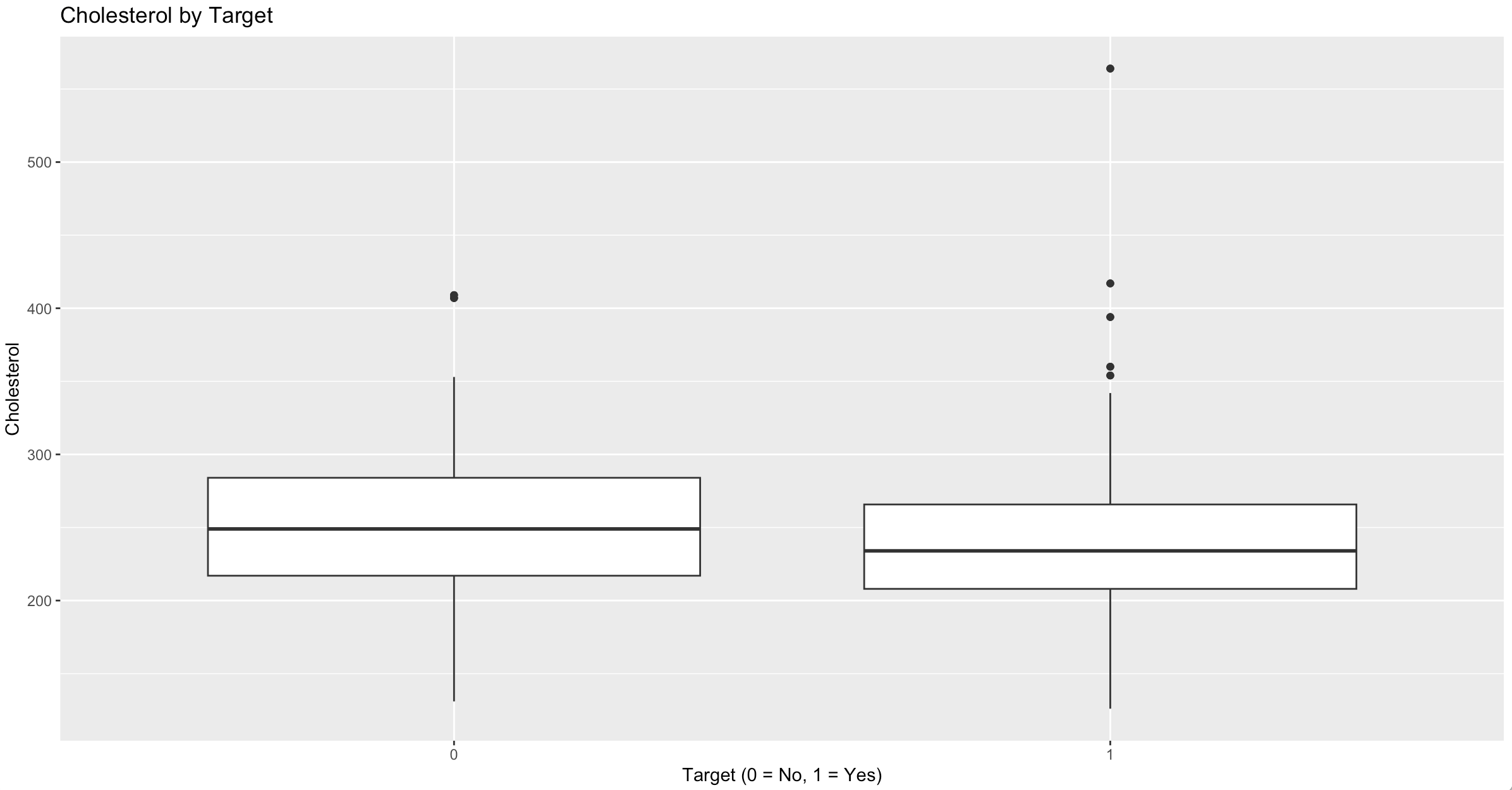


Fig. 3 Histogram of Cholesterol by Target .

# **Task 2: Does the mean cholesterol level is less than 250? Formally test at the alpha = 0.05 level using the 5 steps outlined in the last lecture.**

**A screenshot of a computer program

AI-generated content may be incorrect.**

Fig. 4 Task 2 result.

# **Task 3: Calculate a 90% confidence interval for the mean cholesterol.  Interpret the confidence interval.**

A white background with black text

AI-generated content may be incorrect.

Fig. 5 Task 3 result.

# **Task 4: Formally test that resting blood pressure level is less than 130 at the alpha = 0.05 level using the 5 steps outlined in our last class.**

A screenshot of a computer code

AI-generated content may be incorrect.

Fig. 6 Task 4 result.

# **Task 5: Calculate a 95% confidence interval for the resting blood pressure.  Interpret the confidence interval.**

A white background with black text

AI-generated content may be incorrect.

Fig. 7 Task 5 result.

# **Task 6: Are the cholesterol level of the two groups with target 1 or 0 different? (Is it bigger, less or equal?)**

A black text on a white background

AI-generated content may be incorrect.

Fig. 8 Task 6 result.

# **Task 7: Are resting blood pressure level of the two groups with target 1 or 0 different?  (Is it bigger, less or equal?)**

A white background with black text

AI-generated content may be incorrect.

Fig. 9 Task 7 result.

# **Task 8: Are the fasting blood sugar level of the two groups with target 1 or 0 different?  (Is it bigger, less or equal?)**

A white background with black text

AI-generated content may be incorrect.

Fig. 9 Task 8 result.

# **Task 9: Are the maximum heart rate level of the two groups with target 1 or 0 different?  (Is it bigger, less or equal?)**

A black text on a white background

AI-generated content may be incorrect.

Fig. 10 Task 9 result.

# **Appendices (Code)**

#ASSIGNMENT 11

#GSI Intro to Big Data and Data Mining

#Zhaowen Fan

#Rafael Ignacio Gonzalez Chong

library(readr)

library(ggplot2)

heart <- read\_csv("heart.csv")

#Task 1: Summarize the data by each feature. Use an appropriately labelled

#table to show the results. Also include a graphical presentation that shows

#the distribution of Cholesterol for participants vs. non-participants.

#Describe the shape of each distribution. Use R to calculate the quantities and

#generate the visual summaries. (2 points)

summary(heart)

ggplot(heart, aes(x = chol, fill = factor(target))) +

geom\_histogram(binwidth = 20, alpha = 0.6) +

labs(title = "Cholesterol Distribution",

x = "Cholesterol",

y = "Count")

ggplot(heart, aes(x = factor(target), y = chol)) +

geom\_boxplot() +

labs(title = "Cholesterol by Target",

x = "Target (0 = No, 1 = Yes)",

y = "Cholesterol")

#Task 2: Does the mean cholestoral level is less than 250? Formally test at the

#alpha = 0.05 level using the 5 steps outlined in the last lecture. (6 points)

# Hypotheses:

# H0: mean = 250

# H1: mean < 250

t.test(heart$chol, mu = 250, alternative = "less", conf.level = 0.95)

#Task 3: Calculate a 90% confidence interval for the mean cholestoral.

#Interpret the confidence interval. (4 points)

t.test(heart$chol, conf.level = 0.90)

#Task 4: Formally test that resting blood pressure level is less than 130 at

#the alpha = 0.05 level using the 5 steps outlined in our last class. (6 points)

# Hypotheses:

# H0: mean trestbps = 130

# H1: mean trestbps < 130

t.test(heart$trestbps, mu = 130, alternative = "less", conf.level = 0.95)

#Task 5: Calculate a 95% confidence interval for the resting blood pressure.

#Interpret the confidence interval. (4 points)

t.test(heart$trestbps, conf.level = 0.95)

#Task 6: Are the cholesterol level of the two groups with target 1 or 0

#different? (Is it bigger, less or equal?)

# Hypotheses:

# H0: The mean cholesterol is the same for both groups

# H1: The means are different

t.test(heart$chol ~ heart$target, alternative = "two.sided", conf.level = 0.95)

#Task 7: Are resting blood pressure level of the two groups with target 1 or 0

#different? (Is it bigger, less or equal?)

# Hypotheses:

# H0: The mean resting blood pressure is the same for both groups

# H1: The means are different

t.test(heart$trestbps ~ heart$target, alternative = "two.sided", conf.level = 0.95)

#Task 8: Are the fasting blood sugar level of the two groups with target 1 or 0

#different? (Is it bigger, less or equal?)

# Hypotheses:

# H0: The mean fasting blood sugar is the same for both groups

# H1: The means are different

t.test(heart$fbs ~ heart$target, alternative = "two.sided", conf.level = 0.95)

#Task 9: Are the maximum heart rate level of the two groups with target 1 or 0

#different? (Is it bigger, less or equal?)

# Hypotheses:

# H0: The mean maximum heart rate is the same for both groups

# H1: The means are different

t.test(heart$thalach ~ heart$target, alternative = "two.sided", conf.level = 0.95)