Problem Set 4

Kristina Finley
STAT 100, SECTION 0221

PROBLEM #1

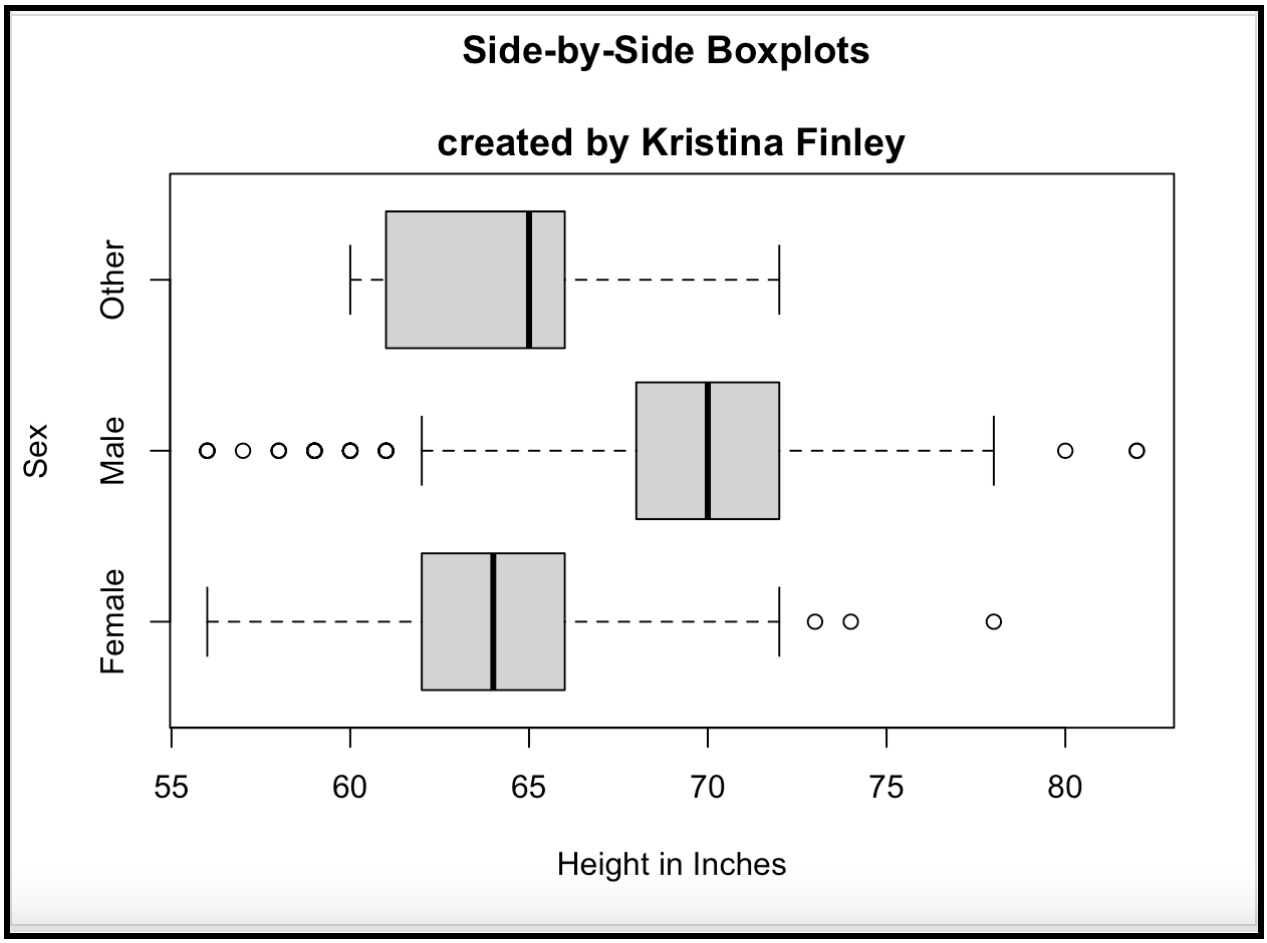
A. The explanatory variable is the person's sex(male or female).
B. The response variable is the person's height in inches.
C. The role type classification is categorical(sex) to quantitative(height in inches).

```
> #1(D-1) extract variables from data frame, by Kristina Finley
> Sex <- Course_Data_Set$Sex
> Height_inches <- Course_Data_Set$Height_inches
> #1(D-2) create descriptive statistics, by Kristina Finley
> tapply(X = Height_inches, INDEX = Sex, FUN = summary)
$Female
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  56.00   62.00   64.00   64.03   66.00   78.00       4

$Male
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  56.00   68.00   70.00   69.57   72.00   82.00       5

$Other
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  60.00   61.00   65.00   64.45   66.00   72.00

> #1(D-3) create a side-by-side boxplots, by Kristina Finley
> boxplot(formula = Height_inches ~ Sex, main = "Side-by-Side Boxplots \n
+         created by Kristina Finley", xlab = "Sex", ylab = "Height in Inches",
+         horizontal = TRUE)
>
```



E. The difference in inches between the median height for males and females is 6.
70 (male) - 64 (female) = 6
F. The Q1 for the distribution of student heights for male students is 62. The Q3 for the distribution of student heights for female students is 66.
G. The conclusion is that students who identify as male are more likely to be taller than students who identify as female based on the data shown.

PROBLEM #2

A. The explanatory variable is the age group.
B. The response variable is the tattoos.
C. The role type classification is categorical(age group) to categorical(tattoos).

```
> #2(D-1) extract variables from data frame, by Kristina Finley
> Age_group <- Course_Data_Set$Age_group
> Tattoos <- Course_Data_Set$Tatoos
> #2(D-2) create a two-way table with conditional percentages, by Kristina Finley
> Two_Way_Freq_Table <- table(Age_group,Tattoos)
> Two_Way_Prop_Table <- prop.table(x = Two_Way_Freq_Table, margin = 1)
> Two_Way_Percent_Table <- Two_Way_Prop_Table * 100
> Two_Way_Percent_Table
                Tattoos
Age_group              No      Yes
  22 or younger 75.79787 24.20213
  23 - 28       56.46259 43.53741
  29 -35        45.00000 55.00000
  Over 35       48.83721 51.16279
>
```

E. The percentage of students in the “22 or younger” age group that have tattoos are 24.20%.
F. The percent of students in the “29-35” age group that have tattoos is 55.00%.
G. The conclusion is that students in higher age groups are more likely to have tattoos based on the data shown.

PROBLEM #3

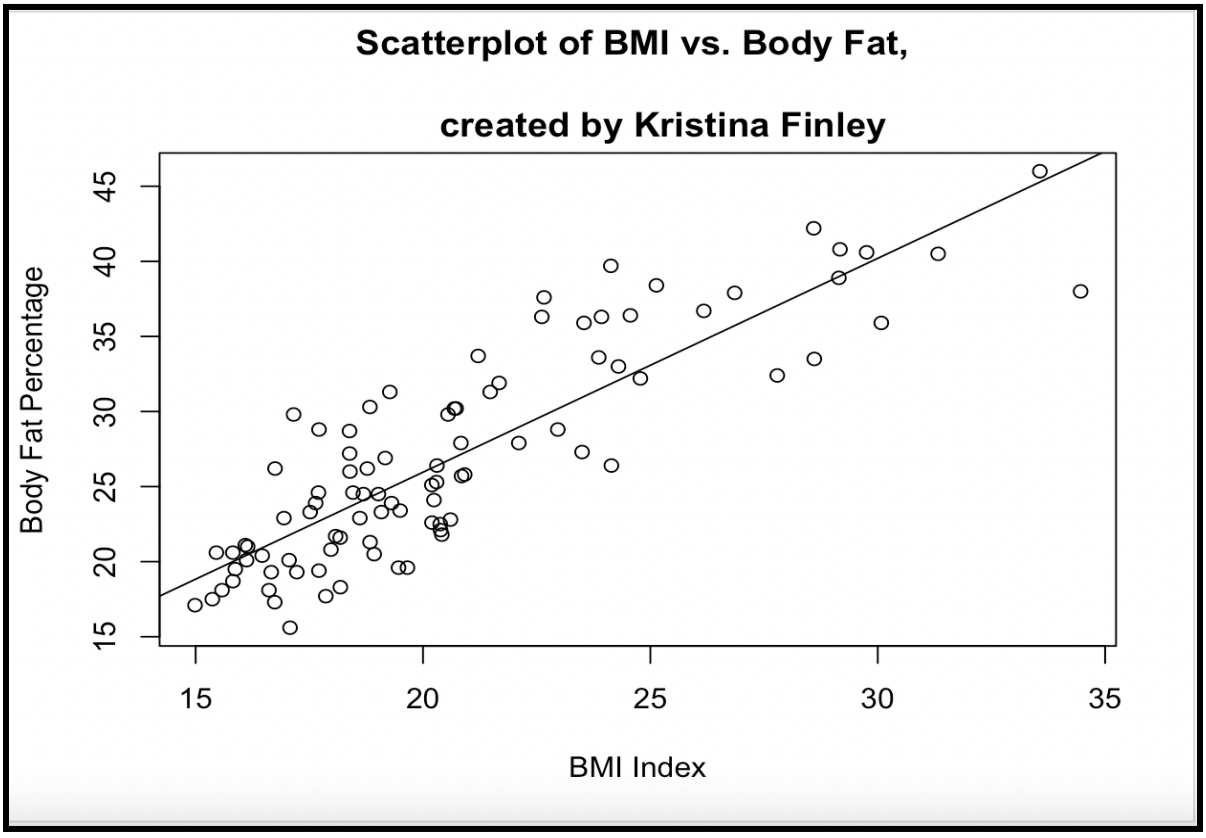
A. The explanatory variable is the BMI(Body Mass Index).
B. The response variable is the Pct_Fat(Percent Fat).
C. The role type classification is quantitative(BMI) to quantitative(Pct_Fat).

```
> #3(D-1) extract variables from data frame, by Kristina Finley
> BMI <- BodyFatPercentage$BMI
> Pct_Fat <- BodyFatPercentage$Pct_Fat
> #3(D-2) create a scatterplot for variables BMI & Pct_Fat, by Kristina Finley
> plot(x = BMI, y = Pct_Fat, main = "Scatterplot of BMI vs. Body Fat, \n
+      created by Kristina Finley", xlab = "BMI Index",
+      ylab = "Body Fat Percentage")
> #3(E) calculate the correlation coefficient, by Kristina Finley
> cor(BMI,Pct_Fat)
[1] 0.8625927
> #3(G) provide a least square regression, by Kristina Finley
> L <- lm(Pct_Fat ~ BMI)
> L

Call:
lm(formula = Pct_Fat ~ BMI)

Coefficients:
(Intercept)          BMI
     -2.535        1.425

> abline(L)
>
```



F. The strength and direction of the scatterplot is a strong positive linear relationship.
H. -2.535 + 1.425(20.50) = 26.68%
I. In conclusion, I would say there is a relationship between a girl's BMI and body fat percentage. The relationship is a strong positive relationship, meaning the BMI and body fat percentage correlate well.