

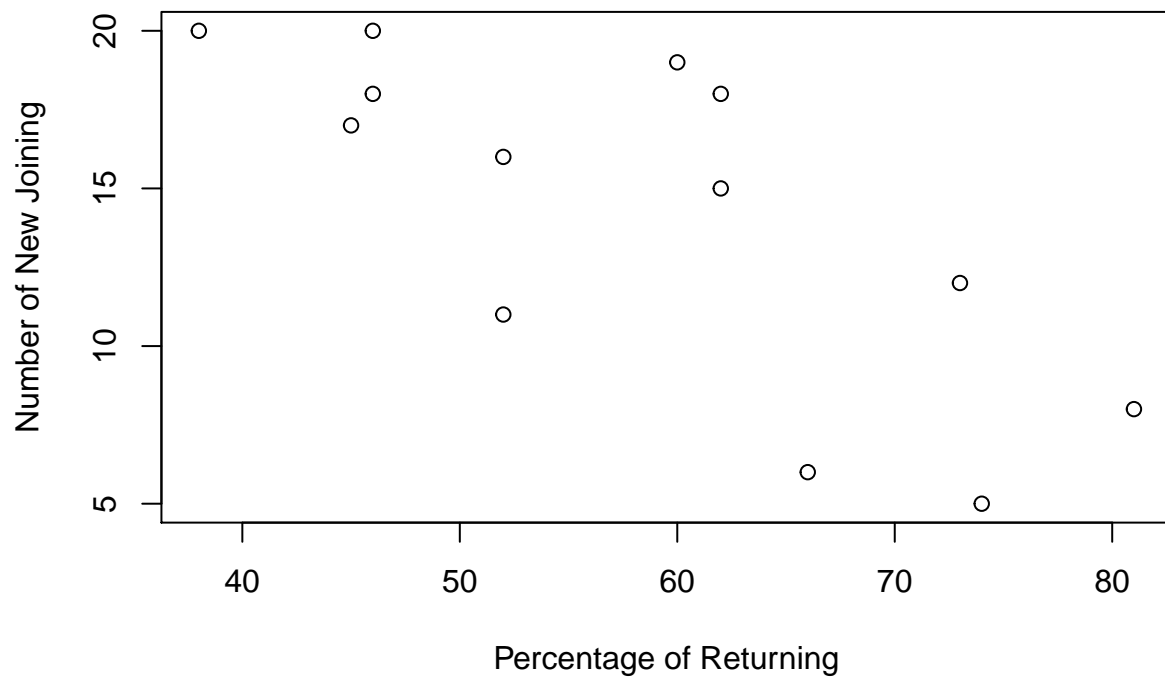
HW 1

309554005 黄睿宇

Problem 1

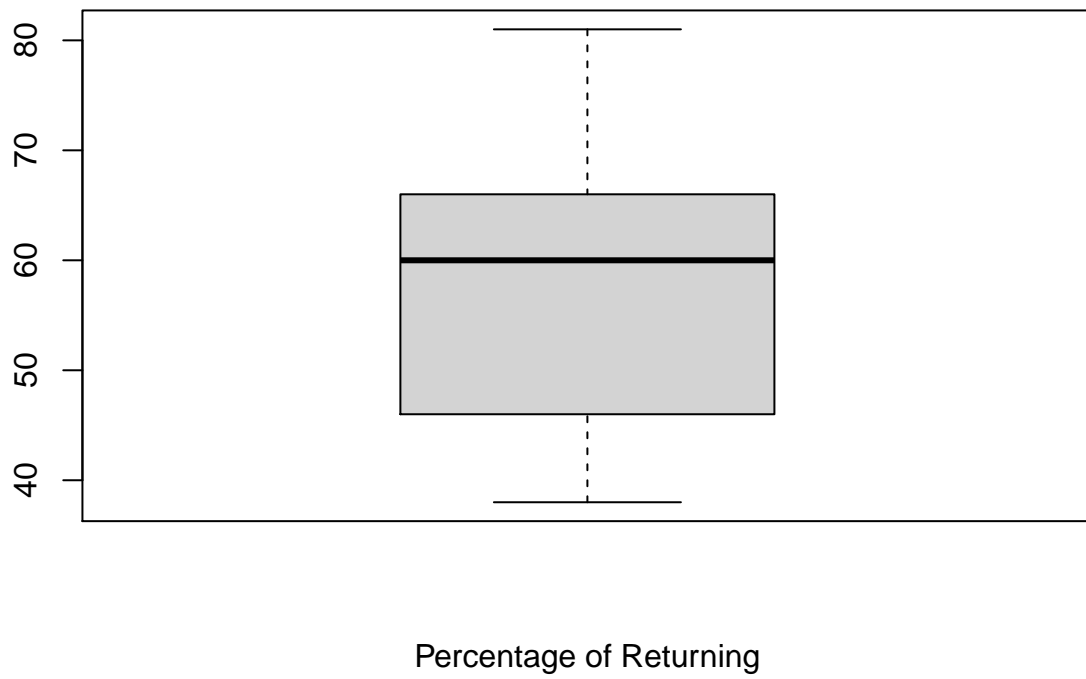
(a)

```
returning_percent <- c(74, 66, 81, 52, 73, 62, 52, 45, 62, 46, 60, 46, 38)
new_adult <- c(5, 6, 8, 11, 12, 15, 16, 17, 18, 18, 19, 20, 20)
plot(returning_percent, new_adult, xlab="Percentage of Returning", ylab="Number of New Joining")
```

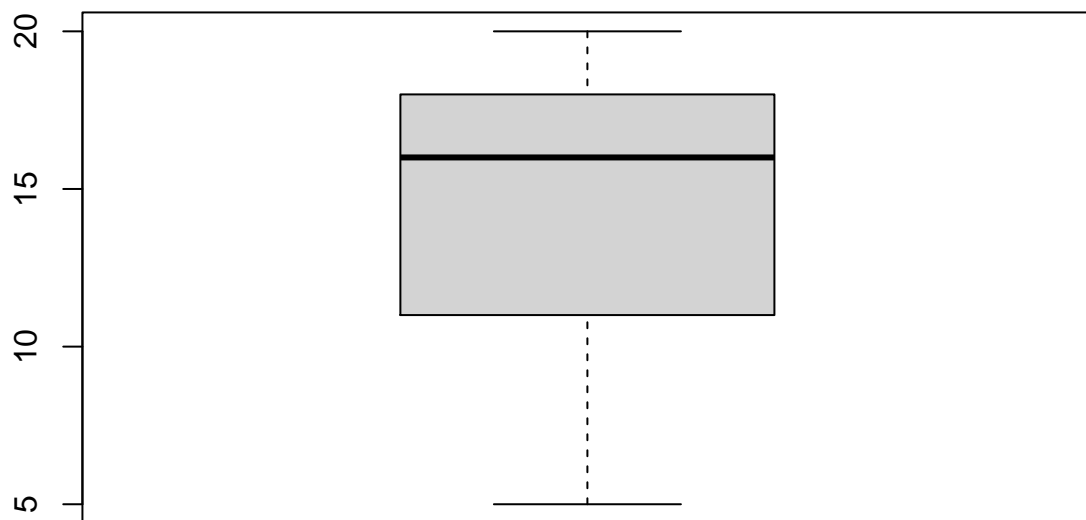


We could see that the percent of adult birds that return from the previous year is inversely proportional to the number of new adults that join.

```
boxplot(returning_percent, xlab="Percentage of Returning")
```



```
boxplot(new_adult, xlab="Number of New Joining")
```



Number of New Joining

There is no possible outliers from the boxplot charts.

(b)

```
mean(returning_percent)
```

```
## [1] 58.23077
```

```
mean(new_adult)
```

```
## [1] 14.23077
```

```
sd(returning_percent)**2
```

```
## [1] 169.859
```

```
sd(new_adult)**2
```

```
## [1] 28.02564
```

$\bar{X} = 58.23077$

$\bar{Y} = 14.23077$

$$S_x^2 = 169.859$$

$$S_y^2 = 28.02564$$

```
sum = 0.0
for (i in 1:length(returning_percent)){
  count <- ((returning_percent[i]-mean(returning_percent)) * (new_adult[i] - mean(new_adult)))
  sum <- sum + count
}
r <- sum / ((sd(returning_percent)*sqrt(length(returning_percent)-1)) * (sd(new_adult)*sqrt(length(new_adult)-1)))
r
```

```
## [1] -0.7484673
```

Pearson's correlation = -0.7484673

(c)

```
cor(returning_percent, new_adult, method="pearson")
```

```
## [1] -0.7484673
```

```
cor(returning_percent, new_adult, method="kendall")
```

```
## [1] -0.5960396
```

```
cor(returning_percent, new_adult, method="spearman")
```

```
## [1] -0.7538043
```

Pearson's correlation = -0.7484673

Kendall's tau = -0.5960396

Spearman's rho = -0.7538043

(d)

```
xy <- data.frame(returning_percent, new_adult)
model <- lm(new_adult ~ returning_percent, data=xy)
model
```

```
##
```

```
## Call:
```

```
## lm(formula = new_adult ~ returning_percent, data = xy)
```

```
##
```

```
## Coefficients:
```

```
##      (Intercept)  returning_percent
```

```
##      31.934         -0.304
```

$$\hat{\alpha} = 31.934$$

$$\hat{\beta} = -0.304$$

```
sigma(model)
```

```
## [1] 3.666891
```

$$\hat{\sigma} = 3.666891$$

(e)

```
summary(model)
```

```
##
## Call:
## lm(formula = new_adult ~ returning_percent, data = xy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.8687 -1.2532  0.0508  2.0508  5.3071
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    31.93426     4.83762   6.601 3.86e-05 ***
## returning_percent -0.30402     0.08122  -3.743  0.00325 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.667 on 11 degrees of freedom
## Multiple R-squared:  0.5602, Adjusted R-squared:  0.5202
## F-statistic: 14.01 on 1 and 11 DF,  p-value: 0.003248
```

$$R^2 = 0.5602$$

```
r**2
```

```
## [1] 0.5602033
```

$$r^2 = 0.5602033$$

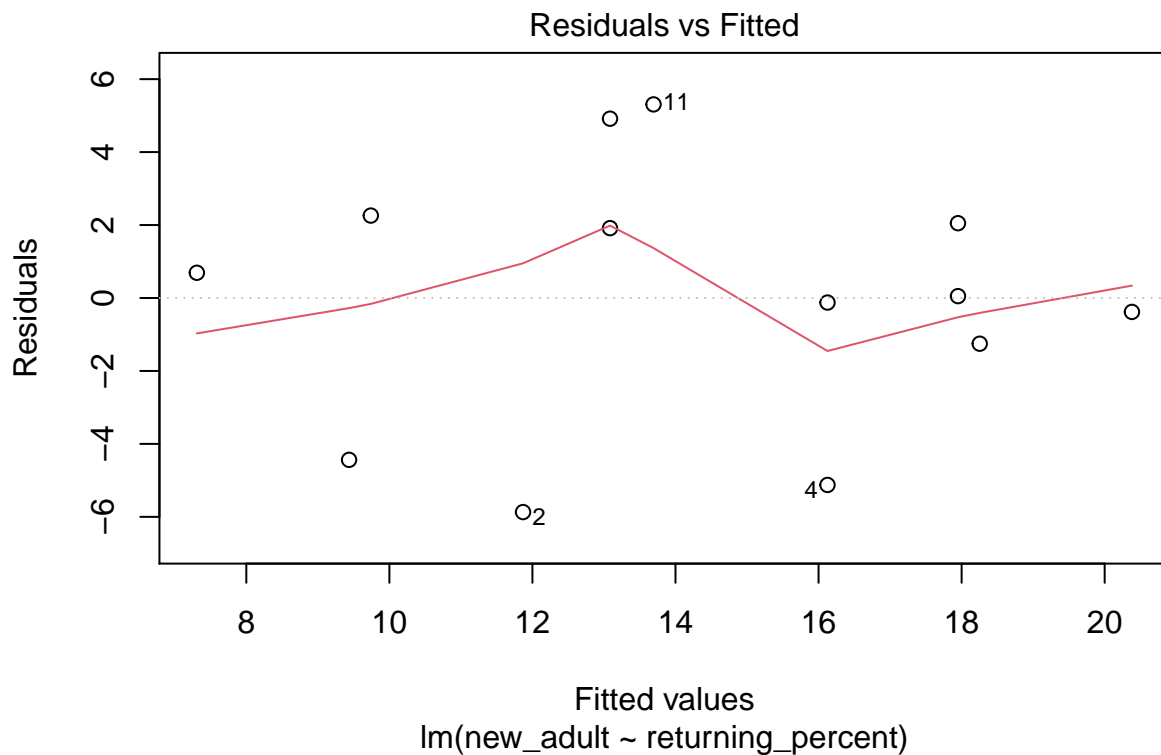
$$\Rightarrow R^2 = r^2$$

(f)

```
residuals(model)
```

```
##          1          2          3          4          5          6
## -4.43656125 -5.86874481  0.69159937 -5.12506604  2.25941580  1.91516341
##          7          8          9         10         11         12
## -0.12506604 -1.25322666  4.91516341  0.05079629  5.30711752  2.05079629
##          13
## -0.38138727
```

```
plot(model, which=1)
```

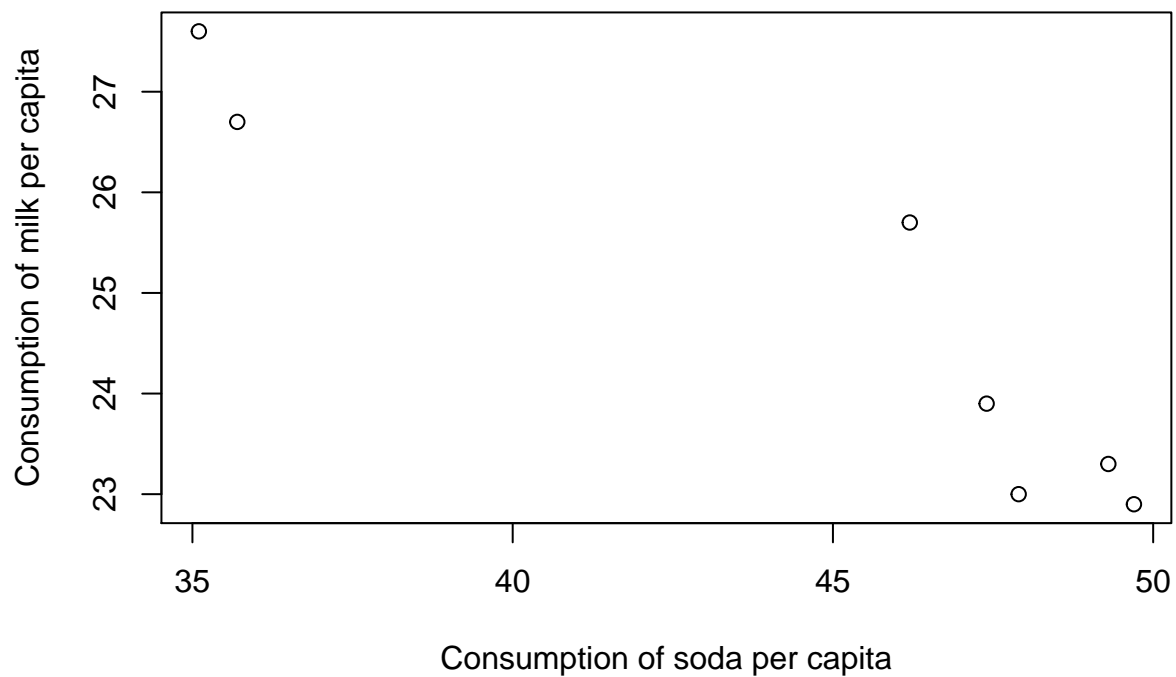


Since the line should be horizontal and close to 0 on the residual plot ideally, we could infer that residual variance does not conform to uniformity and residual mean isn't 0.

Problem 2

(a)

```
soda <- c(35.1, 35.7, 46.2, 47.4, 47.9, 49.7, 49.3)
milk <- c(27.6, 26.7, 25.7, 23.9, 23.0, 22.9, 23.3)
plot(soda, milk, xlab="Consumption of soda per capita", ylab="Consumption of milk per capita")
```



We could see that the consumption of soda per capita is inversely proportional to those of milk, and the possible outliers are (35.1, 27.6) and (35.7, 26.7) since these two points are far from other points.

(b)

```
cor(soda, milk, method="pearson")
```

```
## [1] -0.9262881
```

```
cor(soda, milk, method="kendall")
```

```
## [1] -0.9047619
```

```
cor(soda, milk, method="spearman")
```

```
## [1] -0.9642857
```

Pearson's correlation = -0.9262881

Kendall's tau = -0.9047619

Spearman's rho = -0.9642857

(c)

```
xy <- data.frame(soda, milk)
model <- lm(milk ~ soda, data=xy)
model
```

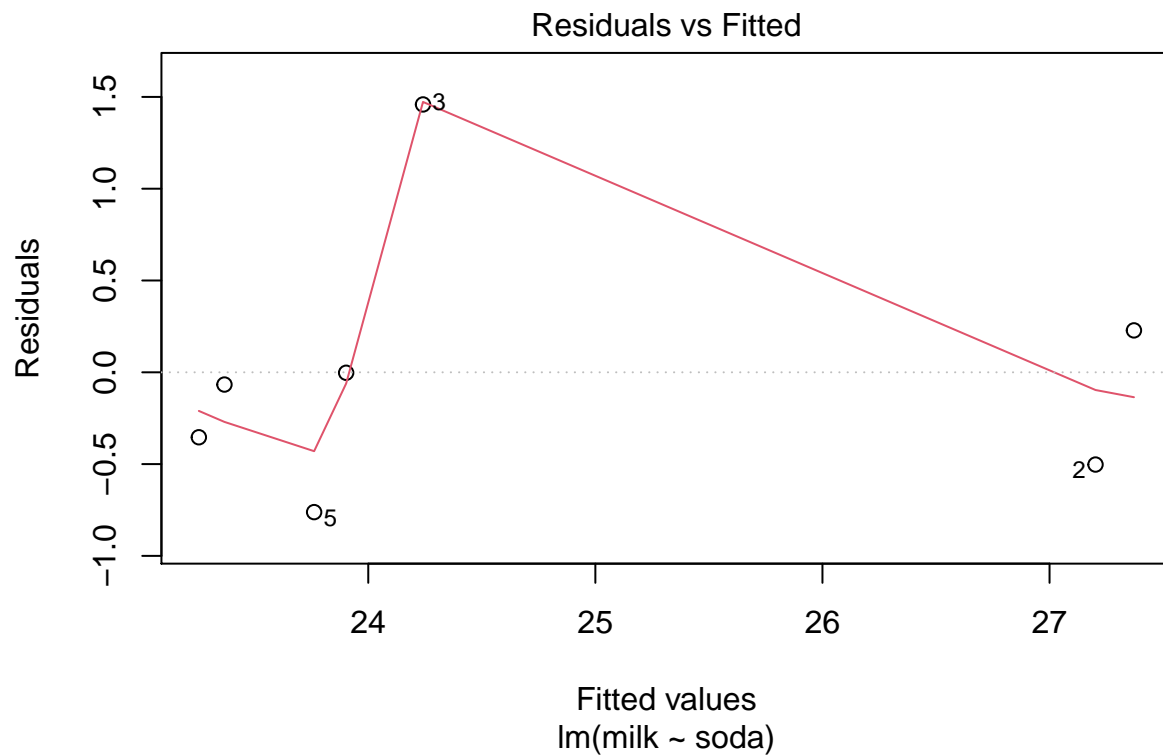
```
##
## Call:
## lm(formula = milk ~ soda, data = xy)
##
## Coefficients:
## (Intercept)      soda
##      37.272      -0.282
```

Regression model: $Y = 37.272 - 0.282X$

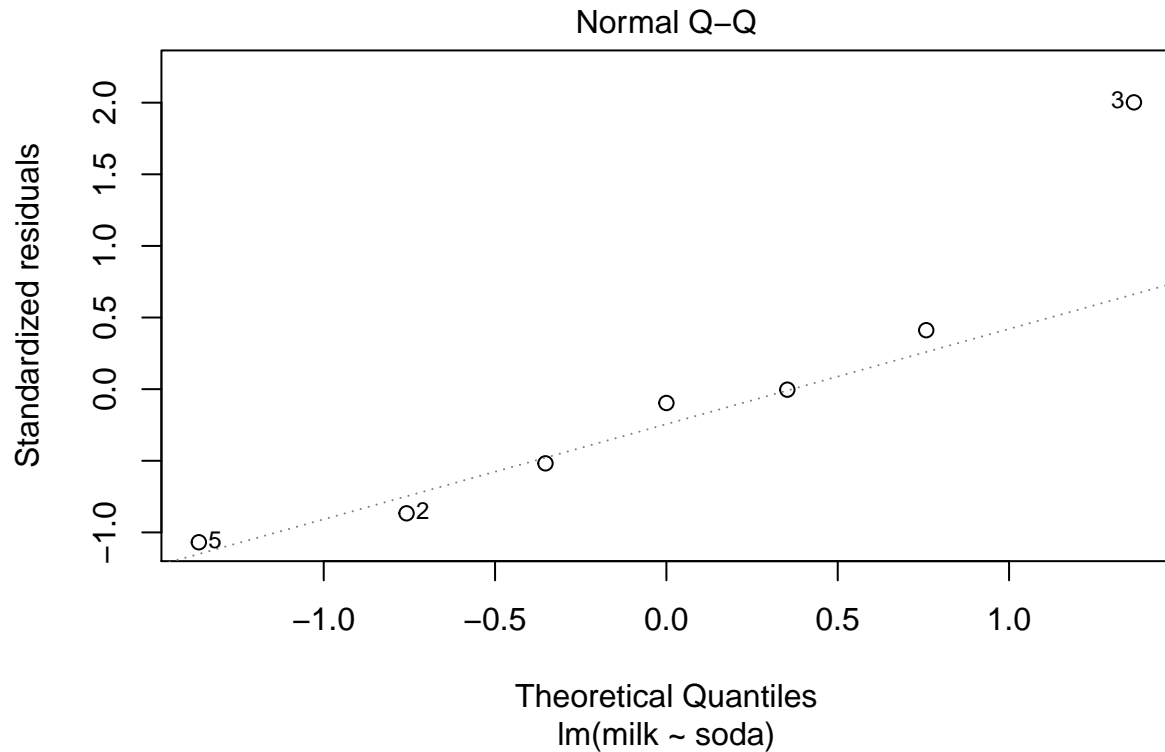
The meaning of β is the amount of change in the y coordinate(milk) when the corresponding x coordinate(soda) change.

(d)

```
plot(model, which=1)
```



```
plot(model, which=2)
```

The residual variance does not conform to uniformity since the line in residual plot should be horizontal, and residual mean isn't 0 as well since the average of the points doesn't close to 0. In addition, since the line in Normal Q-Q plot doesn't fit 45° line and curves down, which means the amount of small values in y coordinate is more than the amount of large values in y coordinate, we could infer that the distribution of residual is not a Normal distribution.