Homework 3

DATA 202 - Alexander - Fall 2023

Please submit **Homework 3** responses as a `.pdf` file on Canvas here.

Exercise 1.1

Is the relationship between the x and y variables in the below model significant?

If so, explain. If not, explain why.

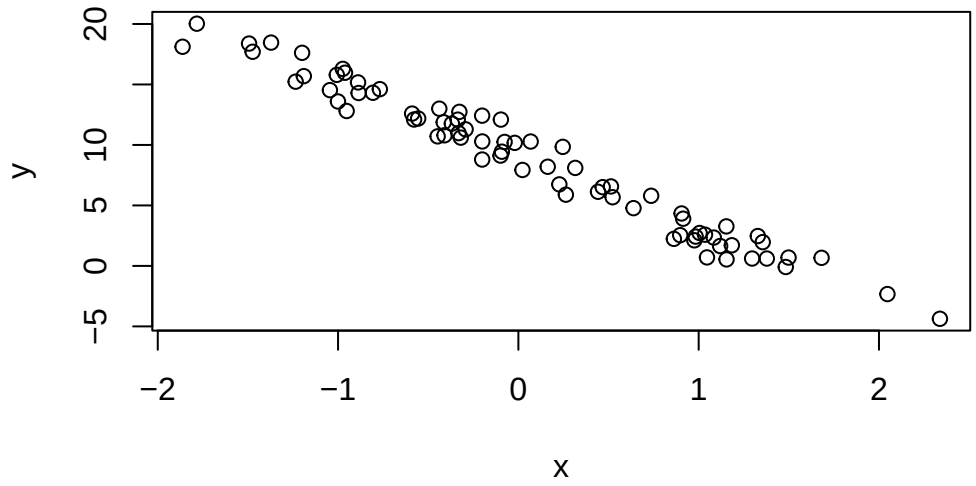
```
model <- lm(y ~ x)
summary(model)
```

```
Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-2.19086 -0.70179 -0.07264  0.79898  2.37303

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.1441     0.1231   74.28   <2e-16 ***
x            -5.9740     0.1277  -46.77   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.062 on 73 degrees of freedom
Multiple R-squared:  0.9677,    Adjusted R-squared:  0.9673
F-statistic:  2187 on 1 and 73 DF,  p-value: < 2.2e-16
```

Exercise 1.2

Examine the plot below. Estimate the correlation coefficient for the plot.

```
plot(x, y)
```

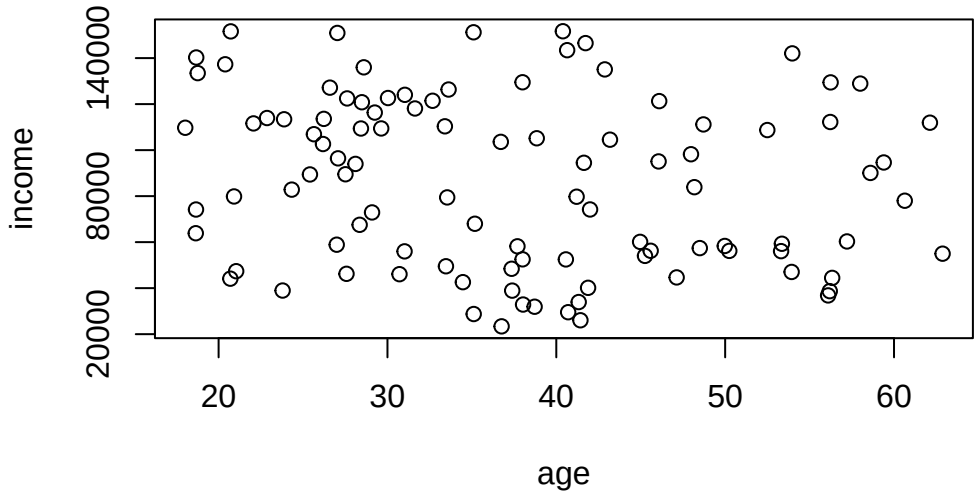


Exercise 1.3

Examine the plot below. Estimate the correlation coefficient for the plot.

Based on your estimate, should we move forward with our analysis? If so, why? If no, why not?

```
plot(age, income)
```



Exercise 1.4

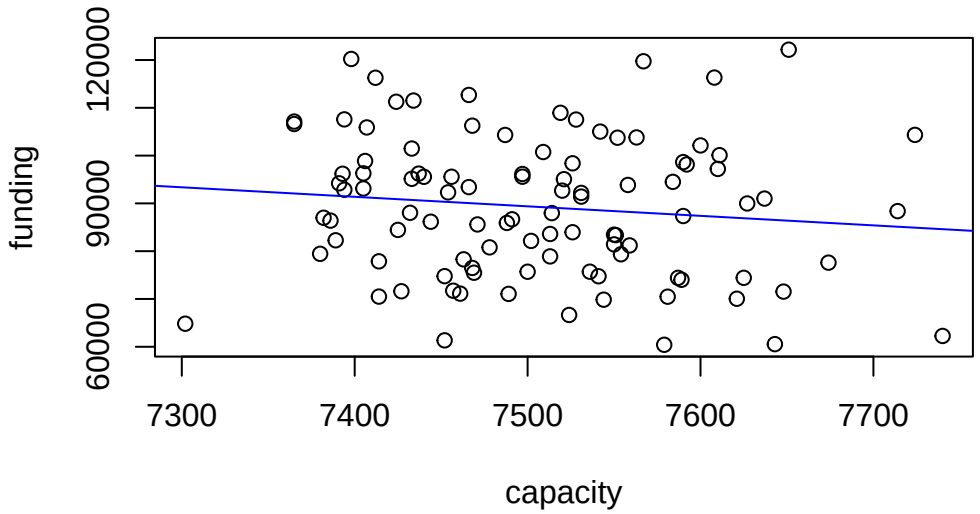
In a few sentences, summarize the relationship between the variables based on the output.

Is there a significant relationship?

```
model2 <- lm(funding ~ capacity)
summary(model2)
```

```
Call:
lm(formula = funding ~ capacity)

Residuals:
   Min     1Q Median     3Q    Max
-28999 -12361   1602  10632  35789

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 238752.68  124897.76   1.912   0.0589 .
capacity       -19.92      16.64  -1.197   0.2342
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14590 on 98 degrees of freedom
Multiple R-squared:  0.01441,   Adjusted R-squared:  0.004349
F-statistic: 1.432 on 1 and 98 DF,  p-value: 0.2342
```

```
plot(capacity, funding)
abline(model2, col="blue")
```



Exercise 1.5

Using the model outlined above and the plot shown below, explain the function of a residual plot.

Does the residual plot represent a "healthy" or "problematic" pattern?

```
resids2 <- residuals(model2)
plot(funding, resids2)
```