Patel, Yash-007-Data Assignment 3

Yash Patel

3/25/2022

Question 1

Here the frequencies of events in each of five categories

```
library(knitr)
df1 <- data.frame(
  Colors = c("Orange", "Blue", "Green", "Red", "Maroon"),
  Frequency = c(21, 21, 18, 19, 20)
)
kable(df1)
```

Colors	Frequency
Orange	21
Blue	21
Green	18
Red	19
Maroon	20

Using r, (you must show your code and output – see Handbook for code) find the appropriate measure of central tendency.

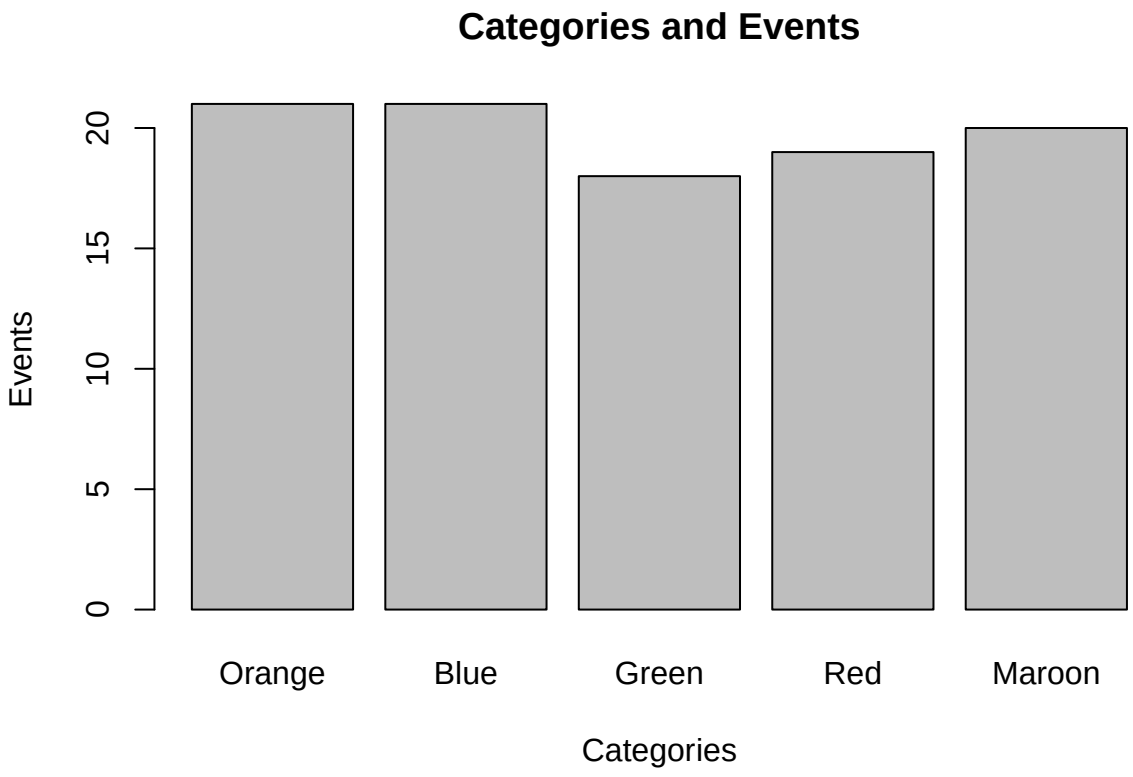
```
getmode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}

getmode(df1$Frequency)
```

```
## [1] 21
```

Create a barplot of the distribution. Be sure to label the x and y axes and the names of each team. The title of the graph should be "Categories and Events".

```r
barplot(df1$Frequency,
        main = "Categories and Events",
        xlab = "Categories",
        ylab = "Events",
        names.arg = df1$Colors
)
```



Question 2

A researcher wants to know if there is a correlation between introspection and optimism. Using two questionnaires, the researcher collects the following data:

```
library(knitr)
df2 <- data.frame(
  Participants = 1:15,
  Introspection = c(51, 65, 74, 74, 75, 76, 87, 86, 51, 85, 52, 92, 97, 70, 86),
  Optimism = c(55, 80, 71, 66, 74, 68, 51, 67, 74, 67, 81, 83, 87, 74, 84)
)
kable(df2)
```

Participants	Introspection	Optimism
1	51	55
2	65	80
3	74	71
4	74	66
5	75	74
6	76	68
7	87	51
8	86	67
9	51	74
10	85	67
11	52	81
12	92	83
13	97	87
14	70	74
15	86	84

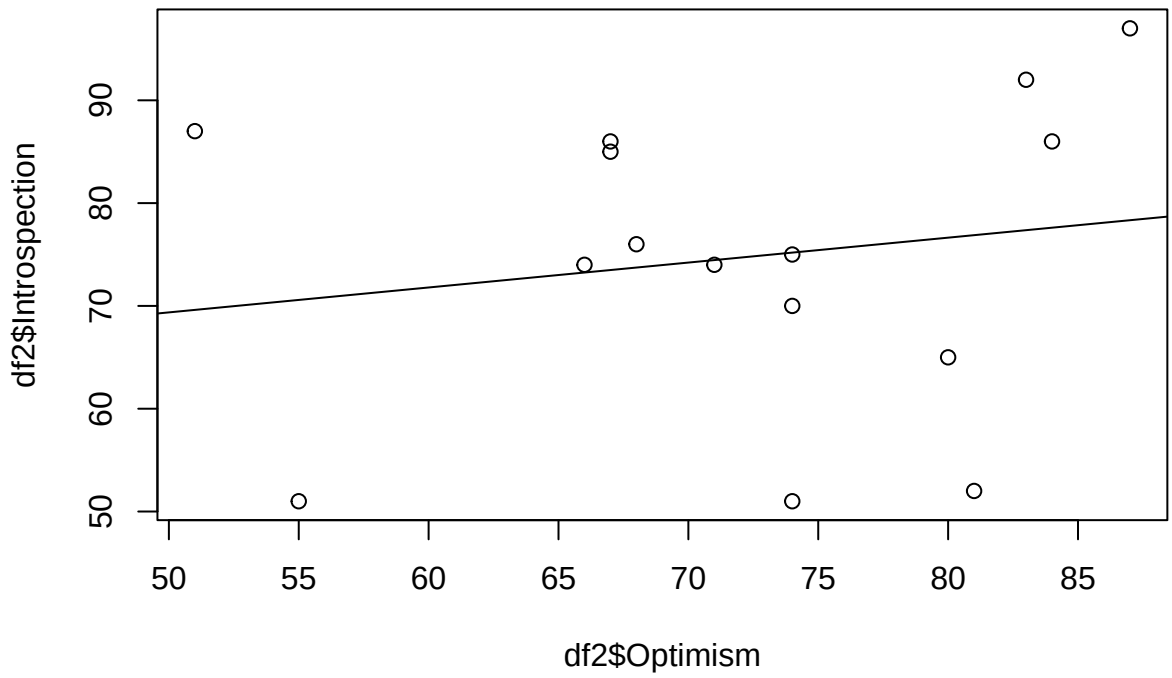
Compute the (Pearson's r) correlation for these data.

```
cor(df2$Participants, df2$Introspection, method = "pearson")
```

```
## [1] 0.4046691
```

Plot introspection against optimism.

```
plot(df2$Optimism, df2$Introspection)
lm2 <- lm(df2$Introspection ~ df2$Optimism)
abline(lm2)
```



Compute the linear model for the data. What are the slope and intercepts?

```
summary(lm2)
```

```
##
## Call:
## lm(formula = df2$Introspection ~ df2$Optimism)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.8817  -8.4125   0.7527  12.0105  18.6645
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   57.2557    28.6592   1.998   0.0671 .
## df2$Optimism   0.2423     0.3936   0.616   0.5488
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.14 on 13 degrees of freedom
## Multiple R-squared:  0.02833,    Adjusted R-squared:  -0.04642
## F-statistic: 0.379 on 1 and 13 DF,  p-value: 0.5488
```

The slope is 0.2423 and the intercept is 57.2557.

Given the slope, what do you expect the trend of the data to be (e.g. "as x increases, y...")

As optimism increases by 1, introspection should increase by 0.2423.

If "Introspection" has a value of 75, what would be the predicted value of "Optimism?"

```
# The formula isn't y = mx + b since we're using the 'y' value (Introspection)
# to predict the 'x' value (Optimism). Because of this, the formula is y = (x - b)/m
(75 - 57.2557)/0.2423
```

```
## [1] 73.23277
```

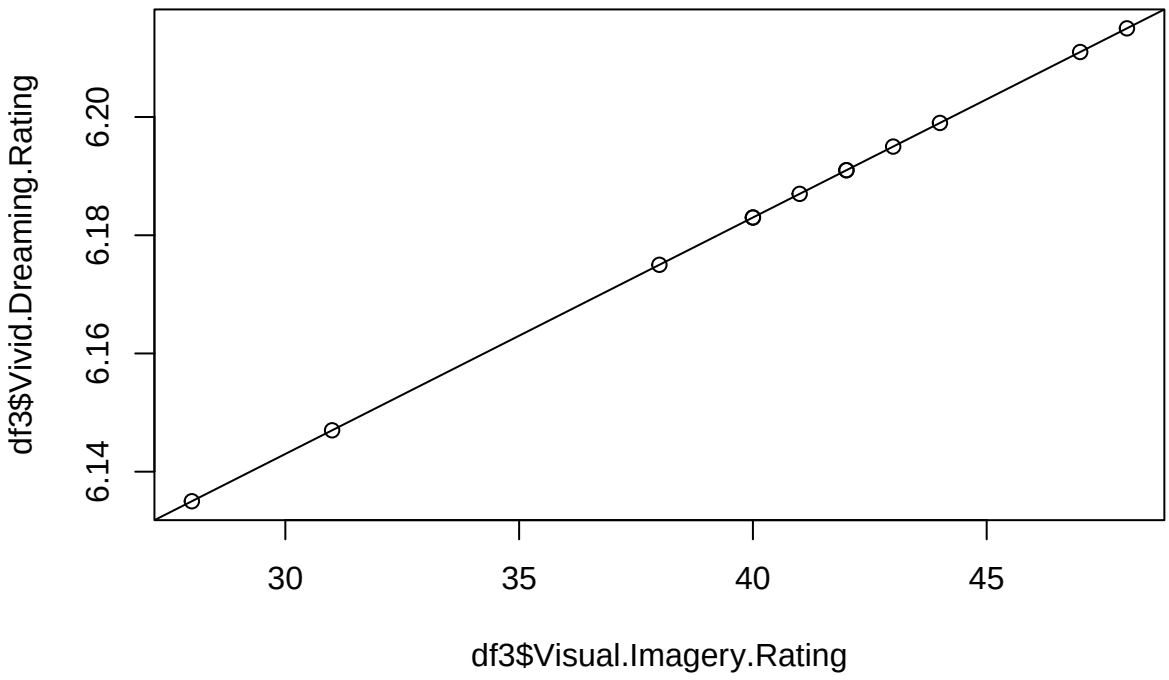
Question 3

Assume that you have a set of scores measuring visual imagery ability. You have also slope and intercept of a plot that used visual imagery data as the X values and vivid dreaming (detailed and visually-involved dreams) ratings as the Y values. However, the vivid dreaming data isn't available. Using the data below, recreate the Y values, then create the plot and trendline. (15 pts.) For the data – the intercept= 6.023, slope = 0.004

```
df3 <- data.frame(
  Participants = 1:12,
  Visual.Imagery.Rating = c(44, 38, 40, 43, 47, 31, 42, 42, 48, 28, 40, 41)
)
df3$Vivid.Dreaming.Rating <- with(df3, Visual.Imagery.Rating * 0.004 + 6.023)
kable(df3)
```

Participants	Visual.Imagery.Rating	Vivid.Dreaming.Rating
1	44	6.199
2	38	6.175
3	40	6.183
4	43	6.195
5	47	6.211
6	31	6.147
7	42	6.191
8	42	6.191
9	48	6.215
10	28	6.135
11	40	6.183
12	41	6.187

```
plot(
  df3$Visual.Imagery.Rating,
  df3$Vivid.Dreaming.Rating
)
abline(lm(df3$Vivid.Dreaming.Rating ~ df3$Visual.Imagery.Rating))
```


df3$Vivid.Dreaming.Rating
6.14
6.16
6.18
6.20
30
35
40
45
df3$Visual.Imagery.Rating

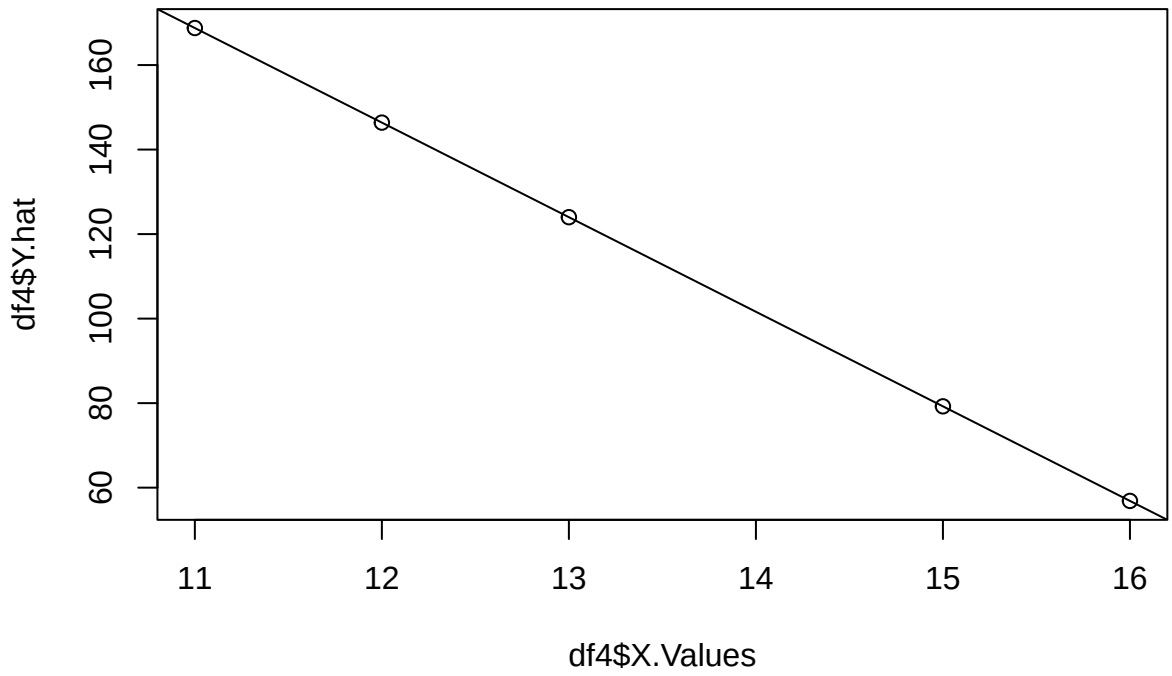
Question 4

Below is a table of X values, $\hat{Y}$, and residuals. Create a trend line and report the slope and intercept values (15 pts.)

```
df4 <- data.frame(
  X.Values = c(16, 12, 15, 13, 11),
  Y.hat = c(56.86, 146.38, 79.2415, 124, 168.76),
  Residuals = c(-21.86, -66.38, 30.76, 26, 31.24)
)
kable(df4)
```

X.Values	Y.hat	Residuals
16	56.8600	-21.86
12	146.3800	-66.38
15	79.2415	30.76
13	124.0000	26.00
11	168.7600	31.24

```
plot(df4$X.Values, df4$Y.hat)
lm4 <- lm(df4$Y.hat ~ df4$X.Values)
abline(lm4)
```



```
summary(lm4)
```

```
##
## Call:
## lm(formula = df4$Y.hat ~ df4$X.Values)
##
## Residuals:
##          1          2          3          4          5
## -6.628e-04 -1.047e-04  9.767e-04 -2.442e-04  3.488e-05
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.149e+02  2.279e-03  182032 3.66e-16 ***
## df4$X.Values -2.238e+01  1.685e-04 -132815 9.41e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0006988 on 3 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 1.764e+10 on 1 and 3 DF,  p-value: 9.413e-16
```

The slope is -22.38 and the intercept is 414.9.

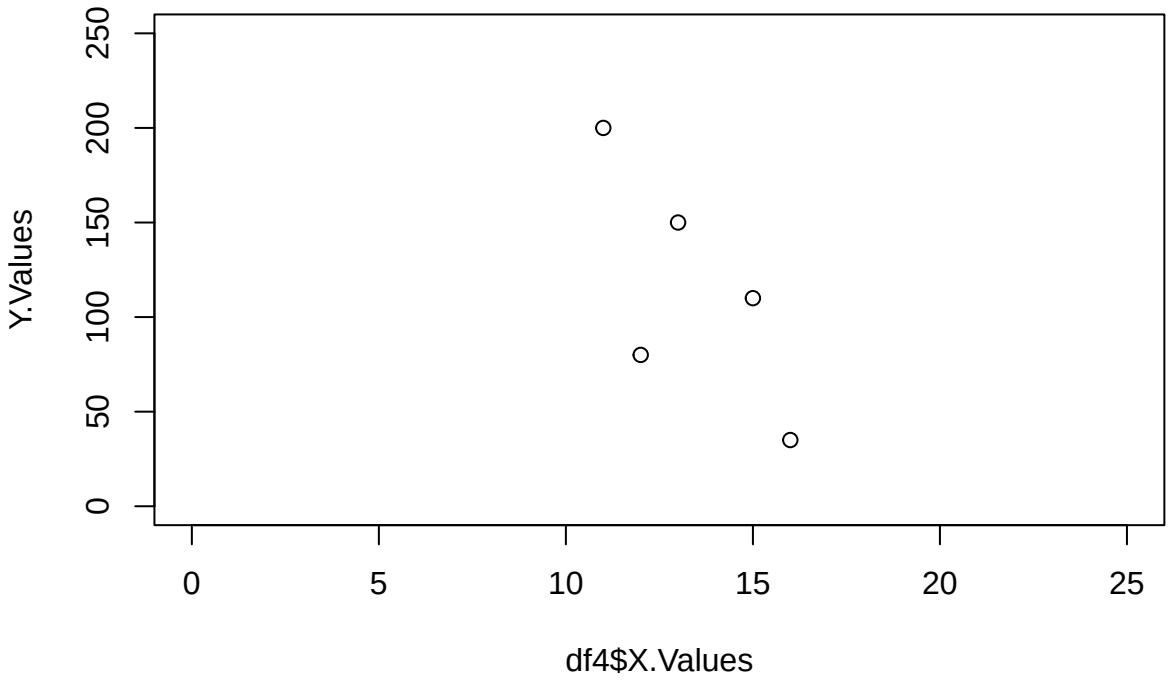
Find the y values. (10 pts.)

```
Y.Values <- with(df4, Y.hat + Residuals)
kable(Y.Values)
```

x
35.0000
80.0000
110.0015
150.0000
200.0000

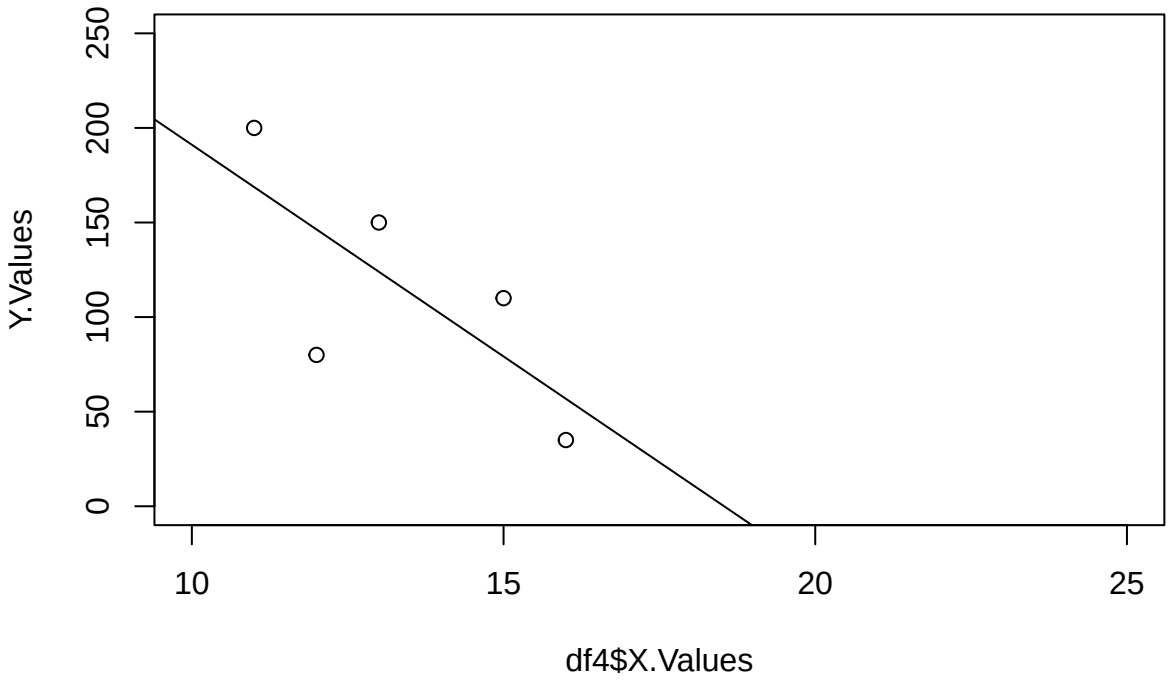
Create a plot (plot the x and y values) with the on the x-axis from 0 to 25 and the y-axis from 0 to 250 (7 pts.)

```
plot(df4$X.Values, Y.Values,
  xlim = c(0, 25),
  ylim = c(0, 250)
)
```



Create a plot (plot the x and y values) with the on the x-axis from 10 to 25 and the y-axis from 0 to 250 (7 pts.)

```
plot(df4$X.Values, Y.Values,
  xlim = c(10, 25),
  ylim = c(0, 250)
)
abline(lm(Y.Values ~ df4$X.Values))
```



Create a trend line based only on the y mean

See above.

Question 5

A developmental psychologist wants to look at the number of different toys a toddler touches in an hour of playing and see if it correlates to language usage (as measured on a scale of 1 to 100). Below is a table of the number toys each toddler plays with

```
df5 <- data.frame(
  Participant = 1:5,
  Number.Of.Toys.Used = c(15, 26, 27, 17, 19)
)
kable(df5)
```

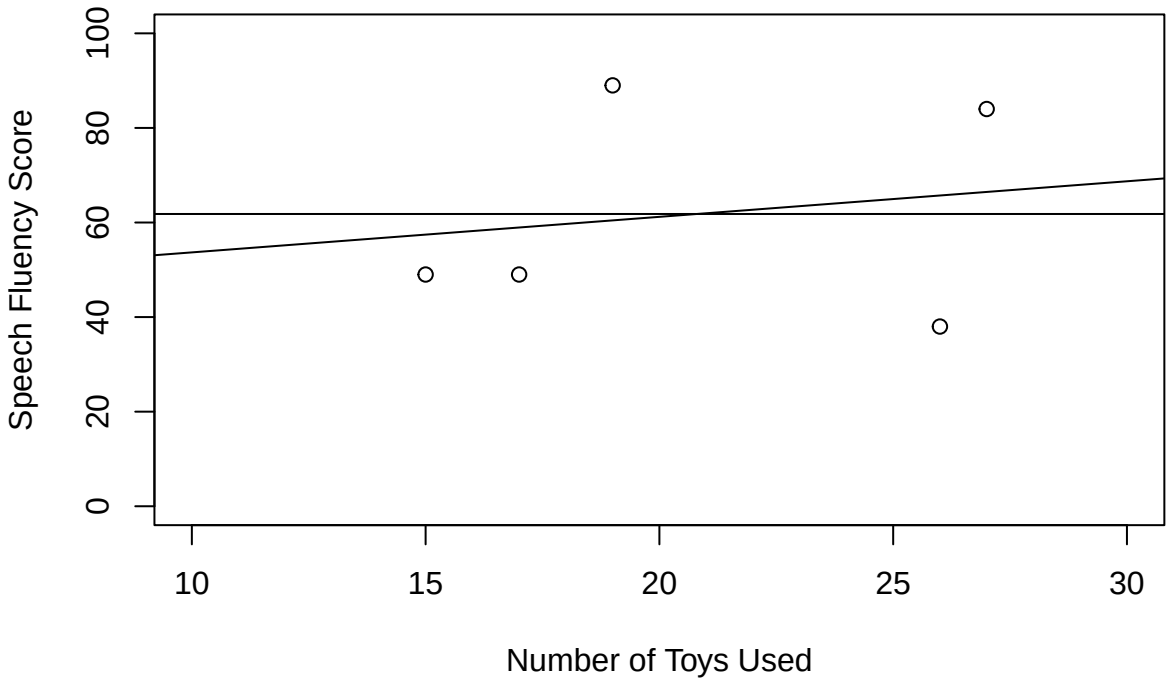
Participant	Number.Of.Toys.Used
1	15
2	26
3	27
4	17
5	19

The mean of the speech fluency scores is 61.8.

Create a plot that lists place a "best guess" trend line given the information provided. (Hint – the mean of speech fluency is 61.8). 7 pts.)

The x-axis should include at least the range of x-values. The y-axis should range from 0 to 100.

```
plot(df5$Number.Of.Toys.Used,
     c(49, 38, 84, 49, 89),
     ylab = "Speech Fluency Score",
     xlab = "Number of Toys Used",
     xlim = c(10, 30),
     ylim = c(0, 100)
)
abline(h = 61.8)
abline(lm(c(49, 38, 84, 49, 89) ~ df5$Number.Of.Toys.Used))
```



Using the values for as 49, 38, 84, 49, 89, Create a trend line using the lm() command and add it to the figure.

See above.

(In the r-command abline(), the "a" stands for the intercept value and the "b" stands for the slope.) (7 pts.) – An alternative to values for slope and intercept, in abline(), h stands for height (or intercept). So, abline(h =) would do what? (4 pts.)

It would create a horizontal line with the equation y = h.

So, abline(v =) would do what? (4 pts.)

It would create a vertical line with the equation x = v.