Author: Millie Wagstaff

#To do – prepare R script of solutions

Week 3 – Tasks (teacher's perspective)

Different Breakout rooms will filter data by different house:

Kai: Hufflepuff

Asma: Gryffindor

Millie: Slytherin

Kasia & Nanaki: Ravenclaw


```r
# Load required libraries
library(RCurl)
library(dplyr)
library(magrittr)
library(ggplot2)


# Task 1: Scatter plot of data

data <- read.csv("Hogwarts_enrolment_data.csv")

slytherin.data <- data %>% filter(Hogwarts.House == "Slytherin") %>%
select(History.of.Magic, Flying)

View(slytherin.data)

ggplot(slytherin.data , aes(x=History.of.Magic, y=Flying)) +
geom_point(color = "green")


# Task 2: Calculate the r squared value and regression line equation

slytherin.lm <- lm(formula = Flying ~ History.of.Magic, data =
slytherin.data)

summary(slytherin.lm)


# Task 3: Add regression line
```

```
ggplot(slytherin.data , aes(x=History.of.Magic, y=Flying)) +
geom_point(color = "green")   +
  geom_smooth(method = "lm", formula = y ~ x  , se = FALSE)


# Task 4: Change the title , x and y  labs

ggplot(slytherin.data , aes(x=History.of.Magic, y=Flying)) +
geom_point(color = "green")   +
  geom_smooth(method = "lm", formula = y ~ x  , se = FALSE)+
 ggtitle("Scatter Plot of History of Magic against Flying") +
 xlab("History of Magic") + ylab("Flying")
```
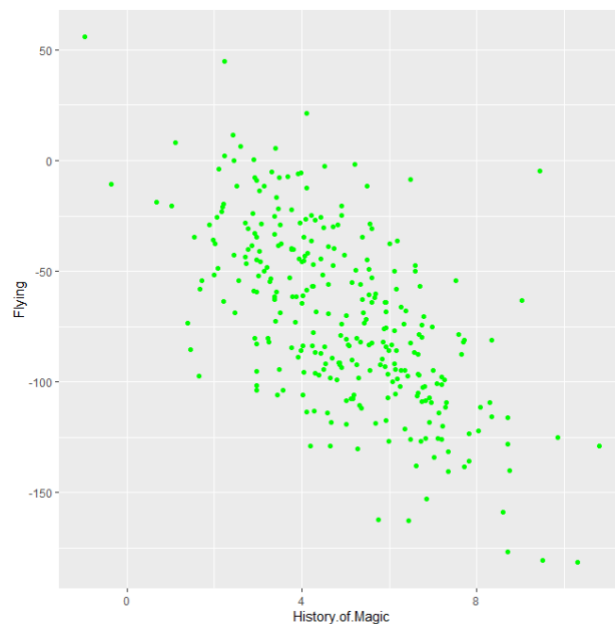
Task Outputs:

Task 1: Produce Scatter Plot of data

	History.of.Magic	Flying
1	2.2381124	44.80
2	6.9906027	-94.84
3	2.2011857	-19.50
4	6.6226550	-138.01
5	2.9138809	-7.64
6	3.9086786	-6.14
7	5.4305604	-73.55
8	1.0997702	8.12
9	4.7209361	-29.83
10	3.4539400	-38.21

Task 2: Running linear Regression

a) *Finding the R squared Value*

```
Call:
lm(formula = History.of.Magic ~ Flying, data = slytherin.data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0753 -1.0386  0.0849  0.9079  6.5020

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.804051   0.170700   16.43   <2e-16 ***
Flying      -0.030010   0.002109  -14.23   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.467 on 299 degrees of freedom
Multiple R-squared:  0.4038,    Adjusted R-squared:  0.4018
F-statistic: 202.5 on 1 and 299 DF,  p-value: < 2.2e-16
```

b) *Finding the regression equation*

Coefficient table:

Rows 1 of the table is labeled (intercept). It contains the y-intercept coefficient. It has a value of 2.804

Row 2 of the table is labeled Flying. It contains the slope coefficient and the error term. These values are -0.03 and 0.002 respectively.

So when we put these into our regression line we have

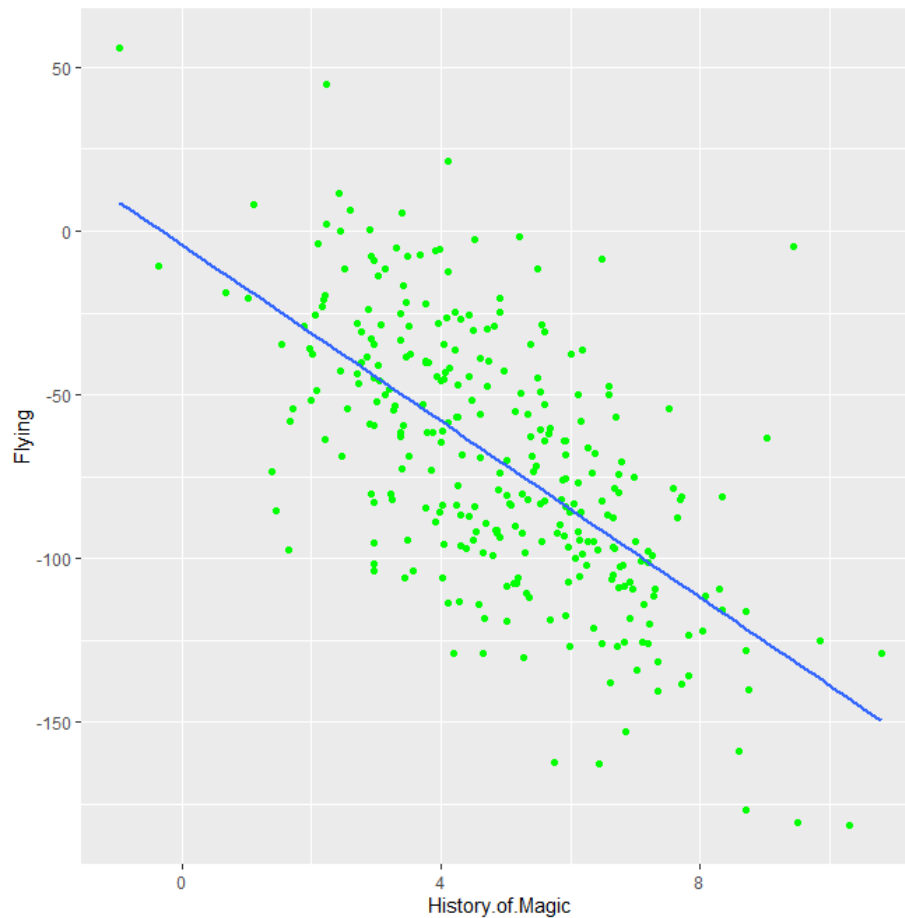*History of Magic = 2.804 + -0.03*Flying*

```
Call:
lm(formula = History.of.Magic ~ Flying, data = slytherin.data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0753 -1.0386  0.0849  0.9079  6.5020

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.804051   0.170700   16.43   <2e-16 ***
Flying      -0.030010   0.002109  -14.23   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.467 on 299 degrees of freedom
Multiple R-squared:  0.4038,	Adjusted R-squared:  0.4018
F-statistic: 202.5 on 1 and 299 DF,  p-value: < 2.2e-16
```

Task 3: Add Regression Line

Task 4:

Scatter Plot showing History of Magic against Flying