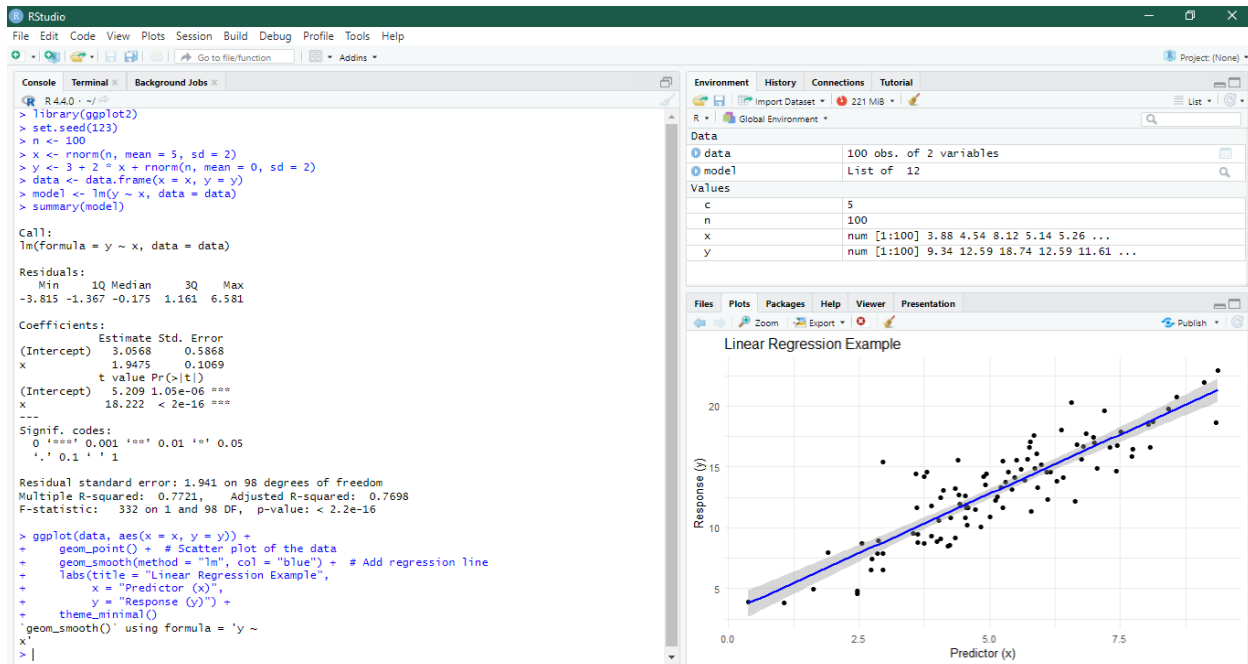


Taaha Hussain Khan

L1F21BSCS0917

F7

Linear Regression



Multiple Regression

```
> set.seed(123)
> n <- 100
> dataset <- data.frame(
+   predictor1 = rnorm(n, mean = 50, sd = 10),
+   predictor2 = rnorm(n, mean = 30, sd = 5),
+   predictor3 = rnorm(n, mean = 20, sd = 3),
+   response = rnorm(n, mean = 100, sd = 15)
+ )
>
> dataset$response <- 50 + 0.5 * dataset$predictor1 + 0.3 * dataset$predictor2 - 0.2 * dataset$predictor3 + rnorm(n, mean = 0, sd = 5)
>
>
> model <- lm(response ~ predictor1 + predictor2 + predictor3, data = dataset)
>
> summary(model)
```

Call:
lm(formula = response ~ predictor1 + predictor2 + predictor3,
 data = dataset)

Residuals:

	Min	1Q	Median	3Q
	-12.4187	-2.8411	0.4316	3.4958
	Max			
	12.3330			

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	62.07179	5.65381	10.979
predictor1	0.38852	0.05416	7.173
predictor2	0.15741	0.10145	1.552
predictor3	-0.28186	0.17337	-1.626

Pr(>|t|)

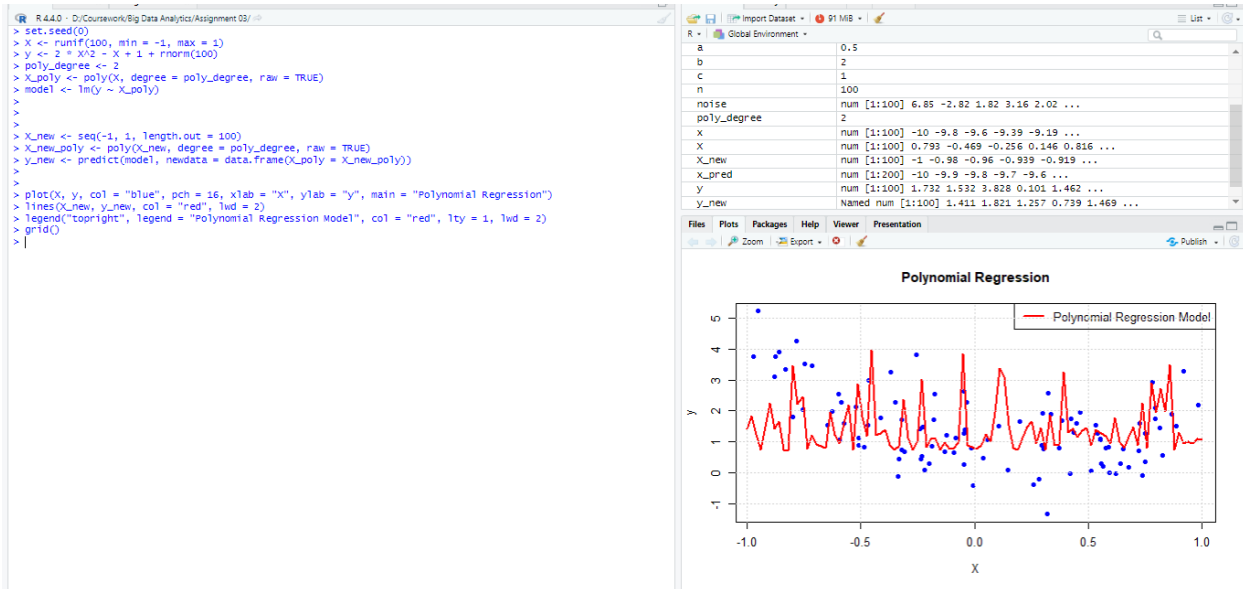
(Intercept)	< 2e-16	***
predictor1	1.53e-10	***
predictor2	0.124	
predictor3	0.107	

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

Residual standard error: 4.873 on 96 degrees of freedom
Multiple R-squared: 0.382, Adjusted R-squared: 0.3627
F-statistic: 19.78 on 3 and 96 DF, p-value: 4.589e-10

```
> |
```

Polynomial Regression



Logistic Regression

```
Console Terminal Background Jobs
R 4.4.0 - D:/Coursework/Big Data Analytics/Assignment 03/
> set.seed(0)
> n <- 100
> X <- runif(n, min = 0, max = 10)
> prob <- 1 / (1 + exp(-(-1 + 0.5 * X)))
> y <- rbinom(n, size = 1, prob = prob)
>
>
> model <- glm(y ~ X, family = binomial)
>
> summary(model)

Call:
glm(formula = y ~ X, family = binomial)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.1407      0.5291  -2.156   0.0311 *
X              0.5104      0.1228   4.157 3.22e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 112.467  on 99  degrees of freedom
Residual deviance:  87.597  on 98  degrees of freedom
AIC: 91.597

Number of Fisher Scoring iterations: 5

> |
```

K means Clustering

```
R 4.4.0 · D:/Coursework/Big Data Analytics/Assignment 03/
> library(ggplot2)
>
> set.seed(0)
> n <- 300
>
>
> cluster1 <- data.frame(
+   x = rnorm(n, mean = 0, sd = 1),
+   y = rnorm(n, mean = 0, sd = 1)
+ )
>
> cluster2 <- data.frame(
+   x = rnorm(n, mean = 4, sd = 1),
+   y = rnorm(n, mean = 4, sd = 1)
+ )
>
> data <- rbind(cluster1, cluster2)
>
> k <- 2
>
> kmeans_result <- kmeans(data, centers = k)
>
>
> ggplot(data, aes(x = x, y = y)) +
+   geom_point(aes(color = factor(kmeans_result$cluster)), size = 3) +
+   geom_point(data = kmeans_result$centers, aes(x = x, y = y), color = "black", size = 5, shape = 17) +
+   labs(title = "K-Means Clustering", color = "Cluster") +
+   theme_minimal()
```