Analysis Of Our Own Generated Data

Create a data set with two independent variables (X1 and X2) and one dependent variable (Y).

```
set.seed(333) #for reproducibility
X1 = rnorm(100)
X2 = rnorm(100)
data <- data.frame(
  X1 ,               # independent variable X1
  X2 ,               # independent variable X2
  Y = 3*X1 + 2*X2 + rnorm(100))     # dependent variable Y (to x1 and x2)
```

View the first 6 observations:

```
head(data)
```

```
           X1          X2         Y
1 -0.08281164 -0.15610909 -1.472033
2  1.93468099  0.51556746  6.823355
3 -2.05128979 -0.65300932 -8.055534
4  0.27773897  2.13594973  6.291255
5 -1.52596060 -0.08955971 -4.538625
6 -0.26916362 -1.28985922 -2.530421
```

Get summary statistics:

```
summary(data)
```

```
       X1                  X2                 Y          
 Min.   :-2.18785   Min.   :-2.4536   Min.   :-9.2215  
 1st Qu.:-0.61050   1st Qu.:-0.5105   1st Qu.:-2.1541  
 Median : 0.04878   Median : 0.1919   Median : 0.5038  
 Mean   :-0.02278   Mean   : 0.1730   Mean   : 0.3019  
 3rd Qu.: 0.42316   3rd Qu.: 0.7195   3rd Qu.: 2.8542  
 Max.   : 1.93468   Max.   : 2.4662   Max.   : 7.1113  
```

Get the correlation matrix:

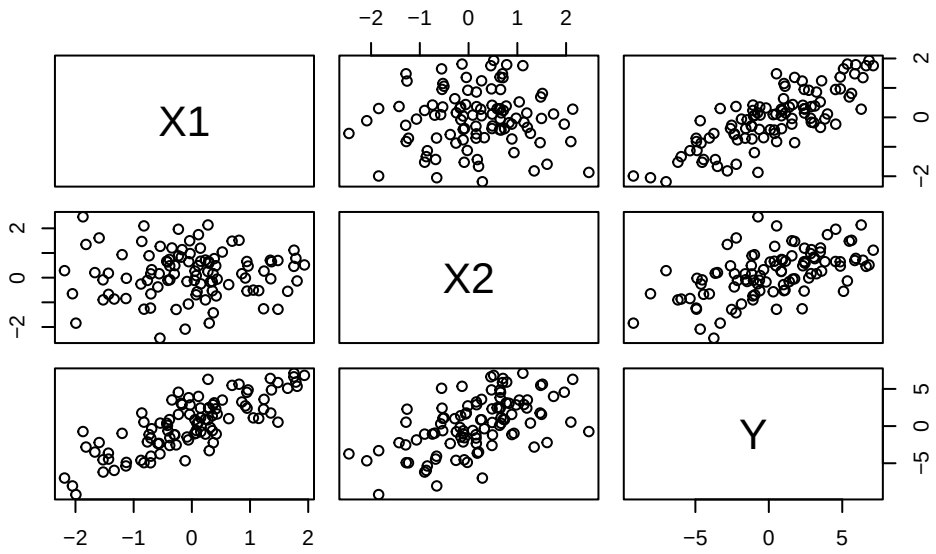
```
cor(data)
```

```
           X1         X2         Y
X1 1.00000000 0.02012909 0.8004960
X2 0.02012909 1.00000000 0.5546317
Y  0.80049603 0.55463175 1.0000000
```

- The correlation between x1 and x2 is 0.02012909
- The correlation between x1 and y is 0.8004960
- The correlation between x1 and y is 0.55463175

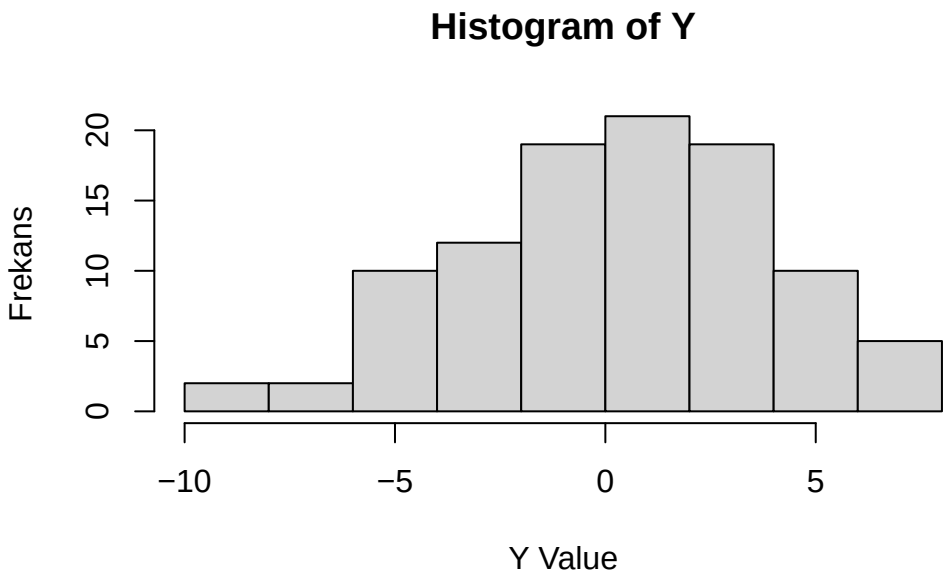
Get the scatter plot matrix:

```
pairs(data)
```



Histogram of Y:

```
hist(data$Y, main = " Histogram of Y ", xlab = "Y Value", ylab = "Frekans")
```



Conducting the regression model:

```
lm_model <- lm(Y ~ X1 + X2, data = data)
summary(lm_model)
```

```
Call:
lm(formula = Y ~ X1 + X2, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.47642 -0.66059  0.06449  0.47086  2.84289

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.01778    0.09690   0.183    0.855
X1           2.95669    0.09995  29.583   <2e-16 ***
X2           2.03130    0.10065  20.183   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9529 on 97 degrees of freedom
Multiple R-squared:  0.9309,    Adjusted R-squared:  0.9295
F-statistic: 653.5 on 2 and 97 DF,  p-value: < 2.2e-16
```

Creating a categorical variable:

```r
data$Z <- factor(rep(c("A", "B"), each = 50))
```

ANOVA:

```r
anova_model <- aov(Y ~ Z, data = data)
summary(anova_model)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Z            1      0   0.001       0  0.994
Residuals   98   1275  13.009
```

ANCOVA:

```r
ancova_model <- lm(Y ~ X1 + X2 + Z, data = data)
summary(ancova_model)
```

```
Call:
lm(formula = Y ~ X1 + X2 + Z, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.26008 -0.59642 -0.07848  0.51431  3.10118

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.2022     0.1337  -1.512   0.1339
X1            2.9647     0.0978  30.314   <2e-16 ***
X2            2.0451     0.0986  20.742   <2e-16 ***
ZB            0.4355     0.1868   2.331   0.0219 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9318 on 96 degrees of freedom
Multiple R-squared:  0.9346,    Adjusted R-squared:  0.9326
F-statistic: 457.4 on 3 and 96 DF,  p-value: < 2.2e-16
```