Assignment2

Weijie Gao

10/12/2017

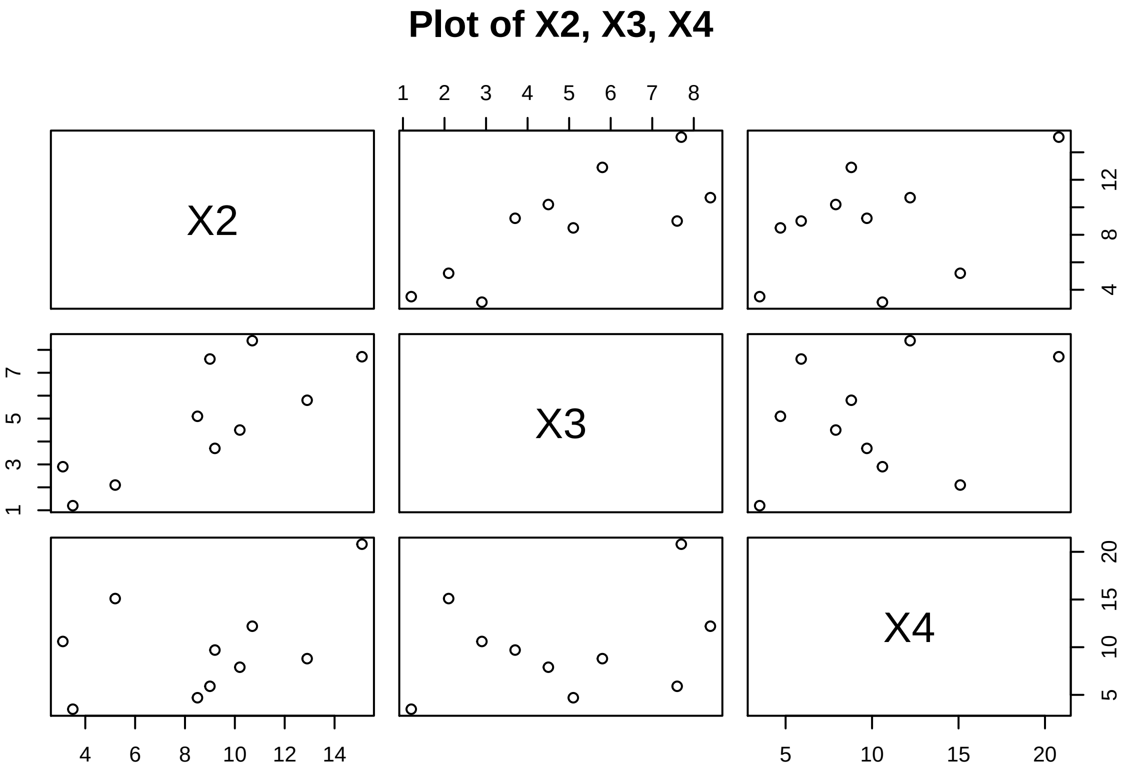
```
dataPath <- "~/Google Drive/2017 Fall/Time Series/week 2"
movie_data <- read.csv(paste(dataPath,"Hollywood movies dataset.csv",sep='/'),header=TRUE)
movie_data
```

```
##       X1   X2       X3   X4
## 1   85.1  8.5 5.100000  4.7
## 2  106.3 12.9 5.800000  8.8
## 3   50.2  5.2 2.100000 15.1
## 4  130.6 10.7 8.399999 12.2
## 5   54.8  3.1 2.900000 10.6
## 6   30.3  3.5 1.200000  3.5
## 7   79.4  9.2 3.700000  9.7
## 8   91.0  9.0 7.600000  5.9
## 9  135.4 15.1 7.700000 20.8
## 10  89.3 10.2 4.500000  7.9
```

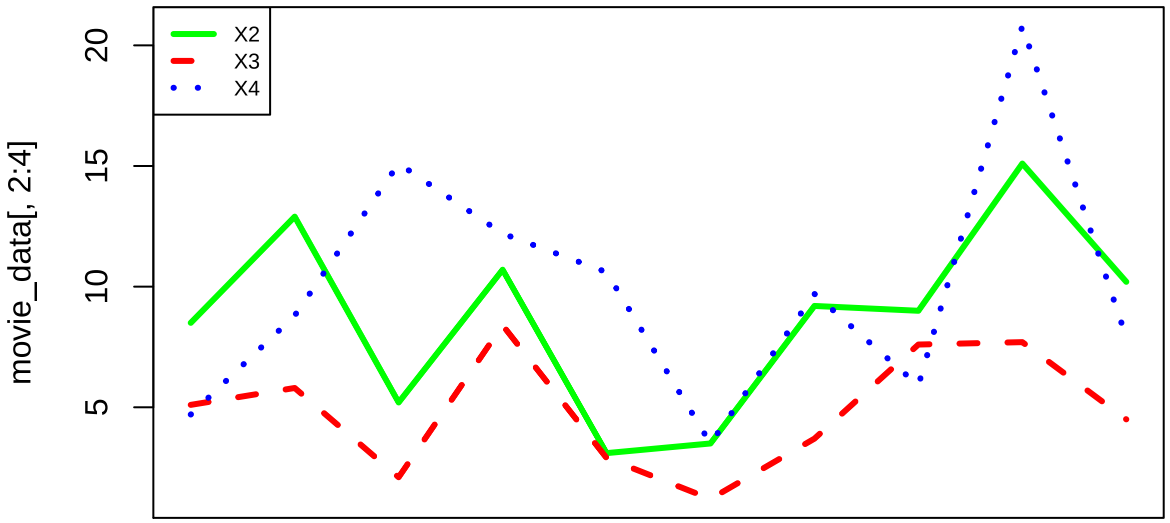
```
pairs(movie_data[,2:4],data=movie_data, main = "Plot of X2, X3, X4")
```

```
## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter
```



```r
matplot(movie_data[,2:4], xaxt="n",type="l",lwd = 3, col=c("green","red","blue"))
legend(x="topleft",c("X2","X3","X4"),lty=c(1,2,3),lwd=3,col=c("green","red","blue"),cex=0.7)
```



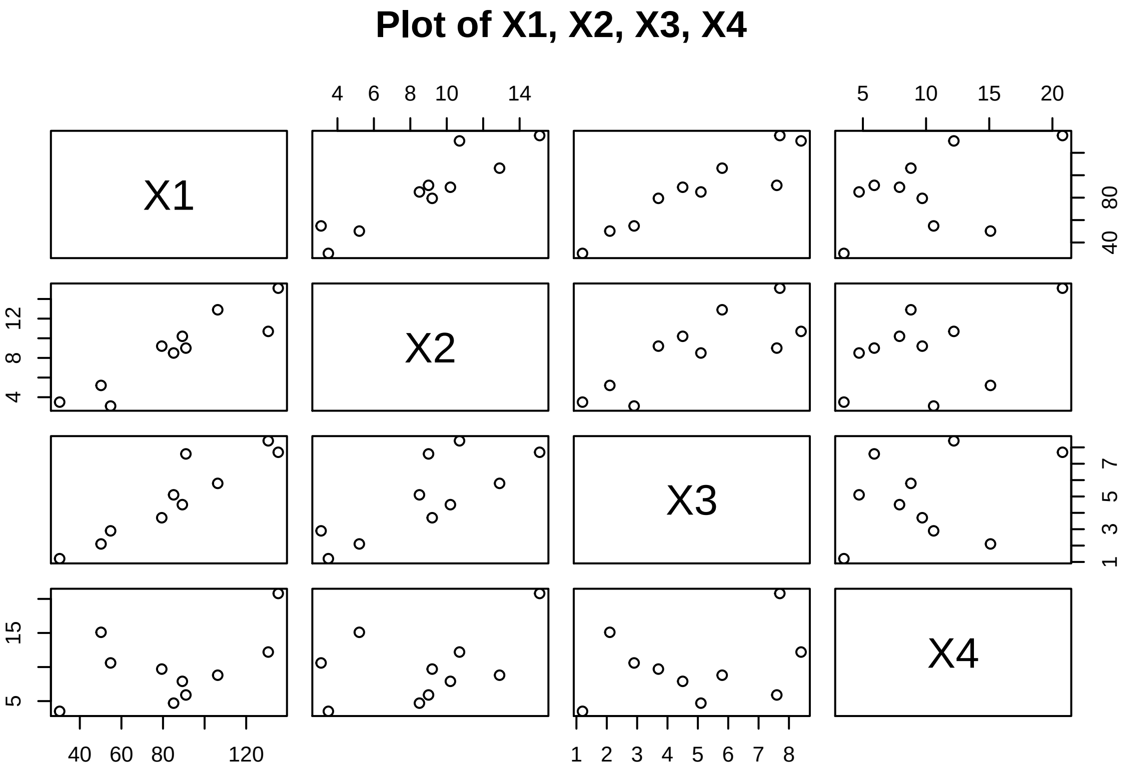
From both plots, we can see that X2 (total production costs/millions) and X3 (total promotional costs/millions) have a fairly strong positive relationship, meaning that movies with high total production costs are highly likely to have a high total promotional costs. X2 (total production costs/millions) and X4 (total book sales/millions) have a relatively weak positive relationship. X3 and X4 have a relatively weakest positive relationship. Note that the the point in the upper right corner of scatterplot between X2 and X4 is seemly an outlier, may bias the correlation. The relationship implies that movies with high production costs or high promotional costs may likely to have high book sales.

```r
pairs(movie_data,data=movie_data, main = "Plot of X1, X2, X3, X4")
```

```
## Warning in plot.window(...): "data" is not a graphical parameter
## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter
## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter
## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
## a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "data" is not a
## graphical parameter

## Warning in plot.window(...): "data" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter

## Warning in title(...): "data" is not a graphical parameter
```



From thescatterplots, we can see that the dependent variable (X1: first year box office receipts) independent variables have positive relationship with X2, X3, and X4 respectively. Note that the positive relationship between X1 and X4 are relatively the weakest.

```
library(tseries)
adf.test(movie_data$X2)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  movie_data$X2
## Dickey-Fuller = -0.8326, Lag order = 2, p-value = 0.9452
## alternative hypothesis: stationary
```

```
adf.test(movie_data$X3)
```

```
##
##  Augmented Dickey-Fuller Test
```

```
##
## data:  movie_data$X3
## Dickey-Fuller = -1.454, Lag order = 2, p-value = 0.7804
## alternative hypothesis: stationary
```

```
adf.test(movie_data$X4)
```

```
## Warning in adf.test(movie_data$X4): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  movie_data$X4
## Dickey-Fuller = -10.734, Lag order = 2, p-value = 0.01
## alternative hypothesis: stationary
```

From the ADF test we see that For X2, the p-value is 0.9452, thus we fail to reject the null hypothesis and there is sufficient evidence to suggest that the X2 is a non-stationary process. For X3, the p-value is 0.7804, thus we fail to reject the null hypothesis and there is sufficient evidence to suggest that the X3 is a non-stationary process.For X4, the p-value is less than 0.01, thus we reject the null hypothesis and there is sufficient evidence to suggest that the X4 (total book sales) is a stationary process.

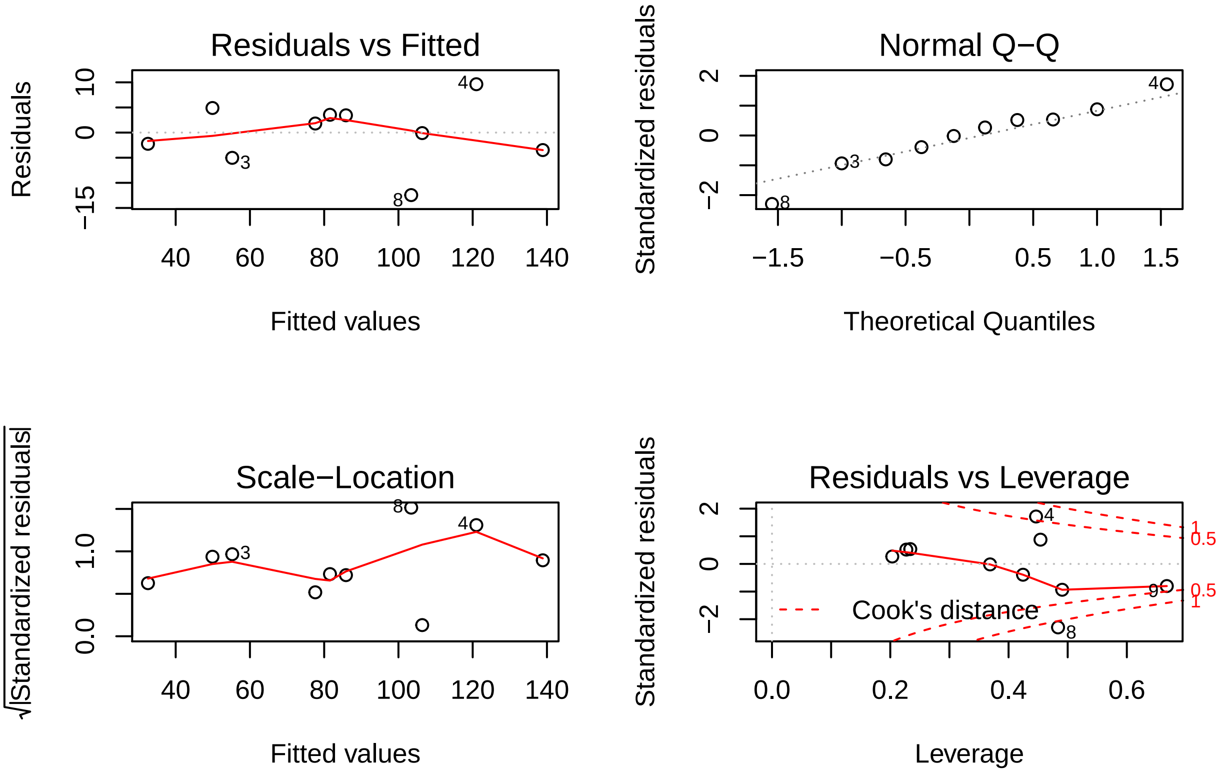
```
m <- lm(X1~., data=movie_data)
summary(m)
```

```
##
## Call:
## lm(formula = X1 ~ ., data = movie_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4384  -3.1695   0.8499   3.5134   9.6207
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.6760     6.7602   1.135   0.2995
## X2            3.6616     1.1178   3.276   0.0169 *
## X3            7.6211     1.6573   4.598   0.0037 **
## X4            0.8285     0.5394   1.536   0.1754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.541 on 6 degrees of freedom
## Multiple R-squared:  0.9668, Adjusted R-squared:  0.9502
## F-statistic: 58.22 on 3 and 6 DF,  p-value: 7.913e-05
```

```
summary(m)$r.squared
```

```
## [1] 0.9667888
```

```
par(mfrow=c(2,2))
plot(m)
```



In summary results, the model has a high R-squared of 0.9668, meaning that our model explain about 96.8% variation in the response variable. Adjust R-squared of 0.9502 is also pretty high. The F-stats of 58.22 and p-value suggests that the model overall is statistically significant, and this model is a better fit than the intercept-only model. Looking at diagnostic plots.

In residual vs.fitted plot, the residuals are not quite randomly distributed, which is not good. In normal qq plot, some points like point 4,8 are off the line, suggesting the normality assumption is not satisfied. The scale-location shows pattern, suggestion the constant variance assumption may not hold.

```
summary(m)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 7.6760285  6.7602276 1.135469 0.299491477
## X2          3.6616040  1.1177514 3.275866 0.016909724
## X3          7.6210513  1.6573172 4.598426 0.003698129
## X4          0.8284681  0.5393591 1.536023 0.175439839
```

The regression coefficient for X2 is 3.66, meaning that holding all other variables constant, on average, for every additional unit increase in total production costs, first year box office receipts tends to increase by 3.66. The regression coefficient for X3 is 7.62, meaning that holding all other variables constant, on average, for every additional unit increase in total promotional costs, first year box office receipts tends to increase by 7.62. The regression coefficient for X4 is 0.828, meaning that holding all other variables constant, on average, for every additional unit increase in total book sales, first year box office receipts tends to increase by 0.828. When total production cost, total promotional costs, and total book sales are equal to 0, the estimated office receipts on average is 7.676.

```
summary(m)$coefficients[,4]
```

```
## (Intercept)          X2          X3          X4
## 0.299491477 0.016909724 0.003698129 0.175439839
```

The variable X2, X3 have p-values smaller than 0.05, meaning they are statistically significant. There is sufficient evidence to suggest that parameter estimates for X2, X3 are different from 0, thus X2, X3 are important in explaining variations in X1. On the other hand, intercept and X4 have p-values greater than 0.05, meaning they are statistically insignificant. The parameter estimates for X4 is not different from 0.

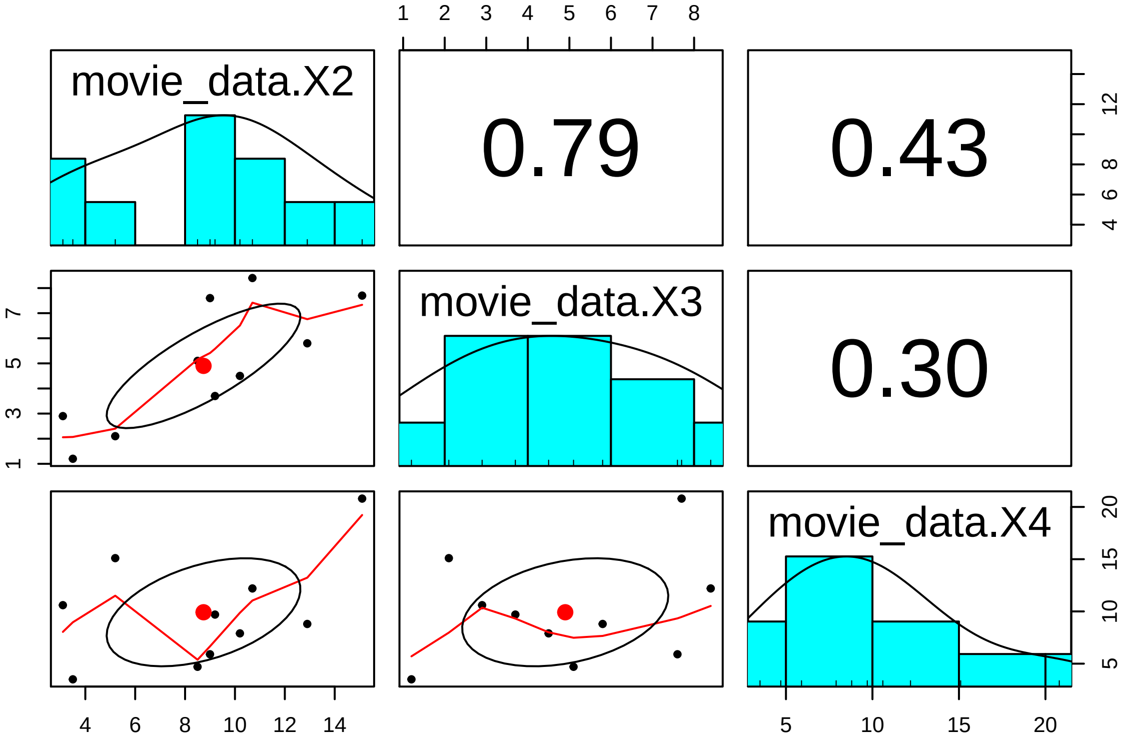
```
# install.packages("usdm")
library(usdm)
```

```
## Loading required package: sp

## Loading required package: raster
```

```
library(psych)
library(car)
```

```
##
## Attaching package: 'car'

## The following object is masked from 'package:psych':
##
##     logit

## The following object is masked from 'package:usdm':
##
##     vif
```

```
df = data.frame(movie_data$X2,movie_data$X3,movie_data$X4)
pairs.panels(df)
```



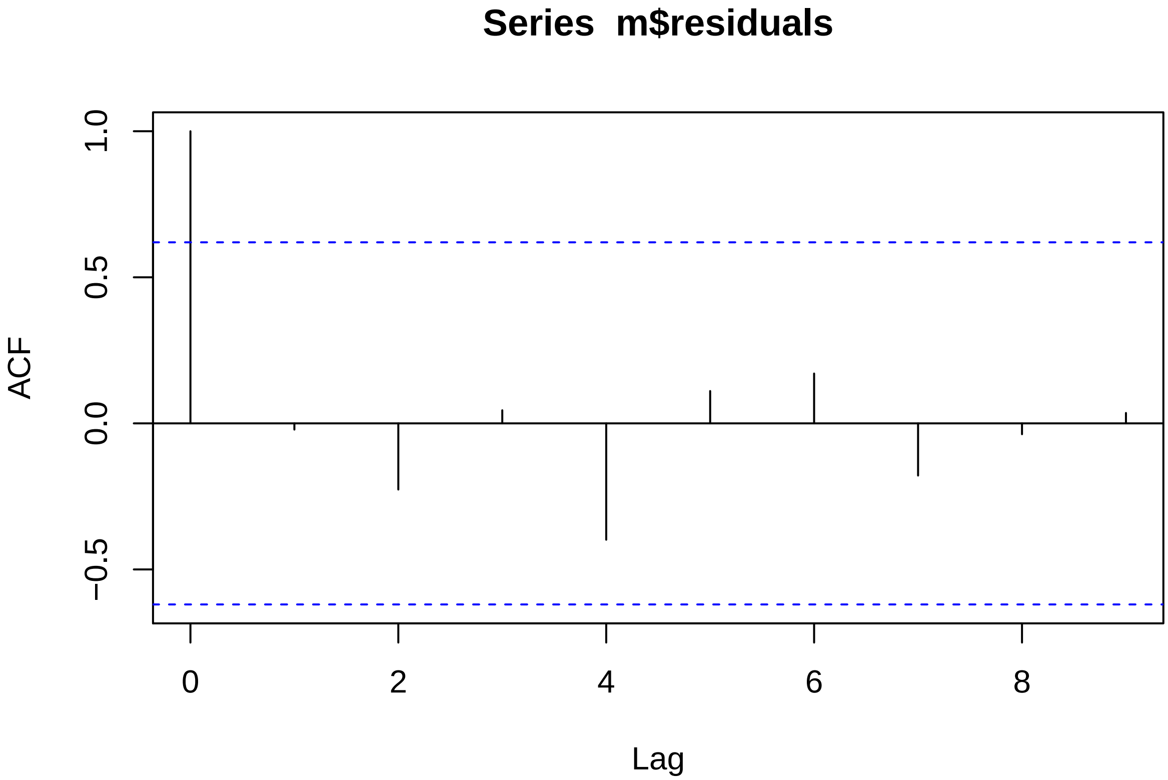
```
vif(m)
```

```
##       X2       X3       X4
## 2.984943 2.673920 1.232227
```

From the pairs plot, we see that the correlation between X2 and X3 of 0.79 is pretty strong. Also, the correlation between X2 and X4 of 0.43, so it seems like there might be some multicollineartiy in the model.

However, looking at vif results, all vif for X2, X3, X4 are smaller than 5, suggesting there is no multicollinearity problem in our model.

```
par(mfrow=c(1,1))
acf(m$residuals)
```



From the ACF plot, we see that all spikes for different lags (except the first one which should have ACF = 1) are all within the boundary and proves to be statistically insignificant. It suggests that the autocorrelation among residuals are small, which is a good sign. Most variation in the response variables could be explained by our predictors, which suggests our model fits well.