NTU Exercise0

Yu Tian

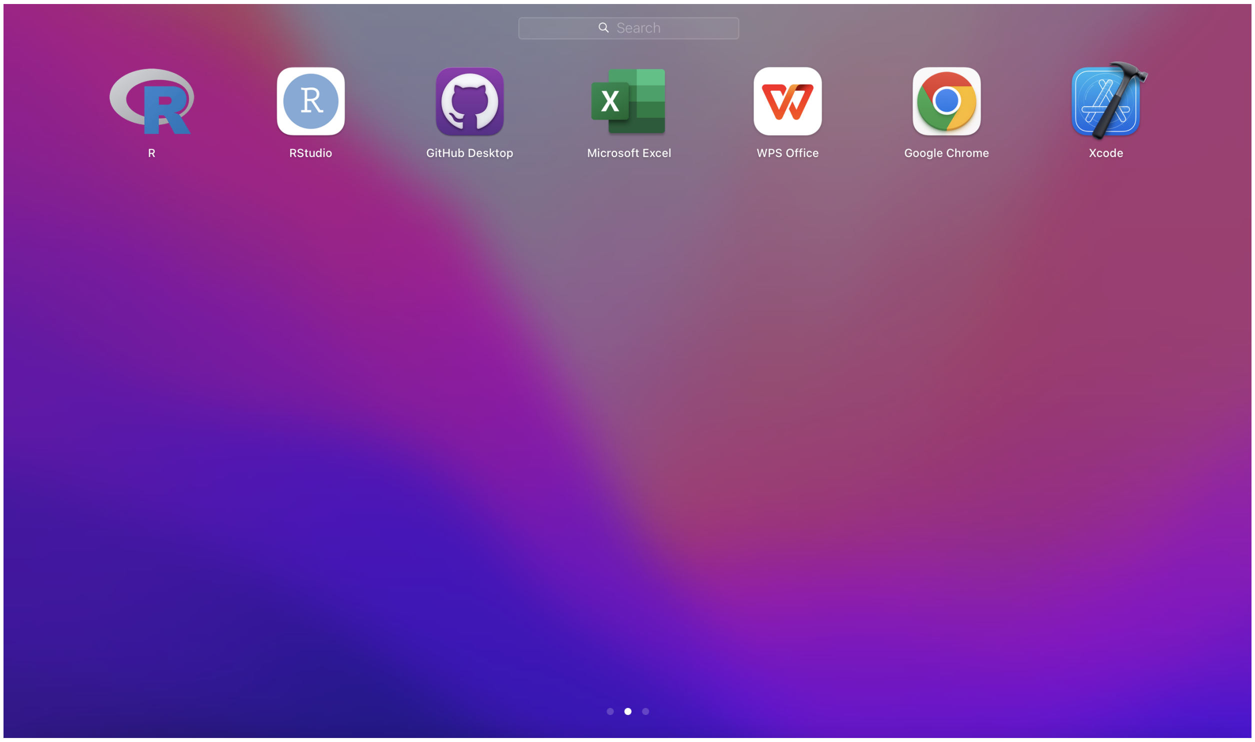
2022-07-28

Package

```
# package
library(dplyr)

# set seed
set.seed(0623)
```

1.Successfully Install R and RStudio.

```
# screenshot: the proof of install R and Rstudio
knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Install R and RStudio.png")
```



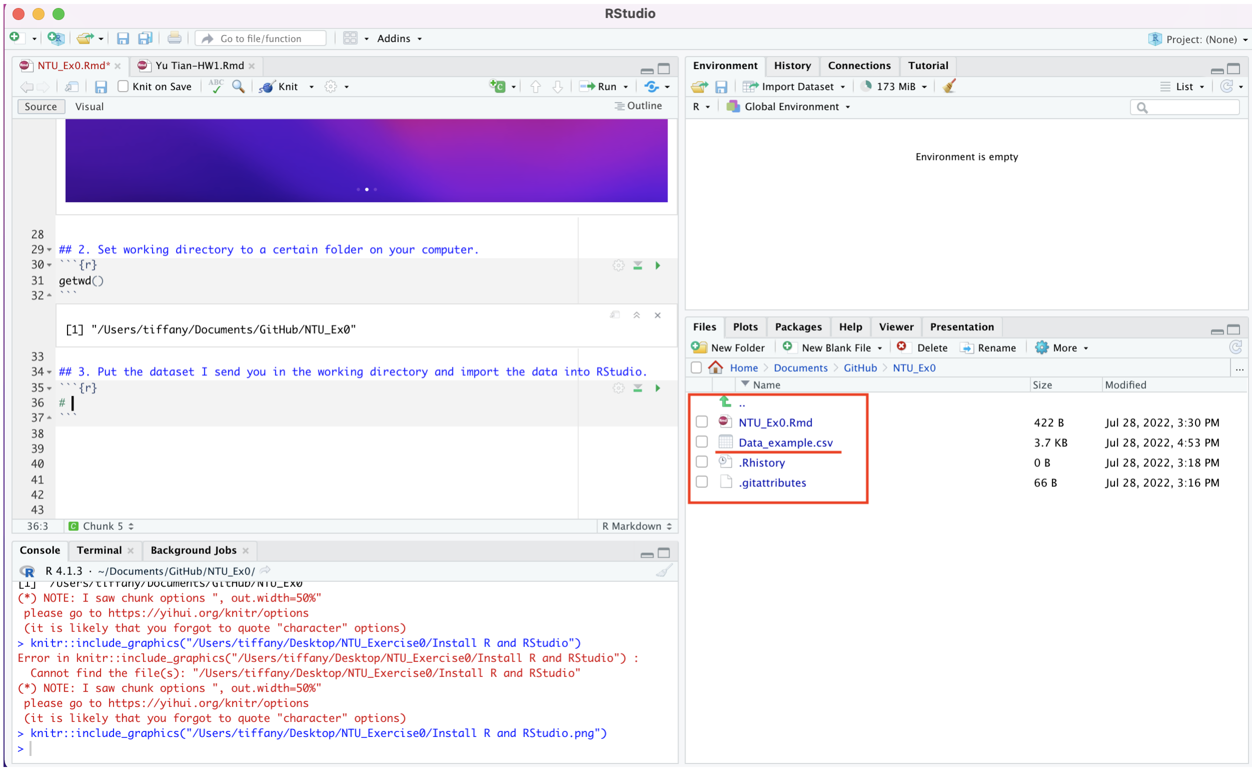
2.Set working directory to a certain folder on your computer.

```
getwd()
```

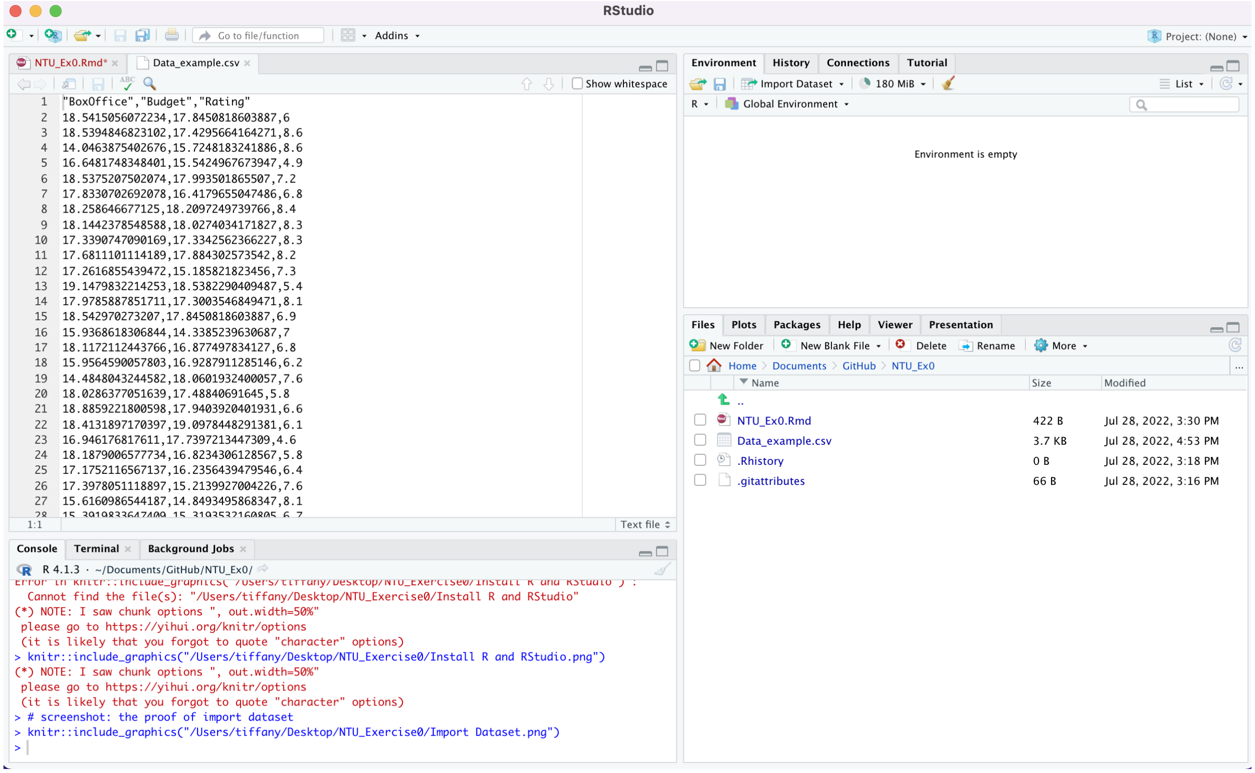
```
## [1] "/Users/tiffany/Documents/GitHub/NTU_Ex0"
```

3.Put the dataset I send you in the working directory and import the data into RStudio.

```
# screenshot: the proof of import dataset
knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Import Dataset.png")
```



~~~
## 2. Set working directory to a certain folder on your computer.
```{r}
getwd()
~~~
```

```
[1] "/Users/tiffany/Documents/GitHub/NTU_Ex0"
```

```
## 3. Put the dataset I send you in the working directory and import the data into RStudio.
```{r}
# 
```
```

36:3

```
R 4.1.3 · ~/Documents/GitHub/NTU_Ex0/
[1] "/Users/tiffany/Documents/GitHub/NTU_Ex0"
(*) NOTE: I saw chunk options ", out.width=50%"
 please go to https://yihui.org/knitr/options
 (it is likely that you forgot to quote "character" options)
> knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Install R and RStudio")
Error in knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Install R and RStudio") :
  Cannot find the file(s): "/Users/tiffany/Desktop/NTU_Exercise0/Install R and RStudio"
(*) NOTE: I saw chunk options ", out.width=50%"
 please go to https://yihui.org/knitr/options
 (it is likely that you forgot to quote "character" options)
> knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Install R and RStudio.png")
> 
```

Name	Size	Modified
..		
NTU_Ex0.Rmd	422 B	Jul 28, 2022, 3:30 PM
Data_example.csv	3.7 KB	Jul 28, 2022, 4:53 PM
.Rhistory	0 B	Jul 28, 2022, 3:18 PM
.gitattributes	66 B	Jul 28, 2022, 3:16 PM

```
knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Dataset.png")
```
```



```
"BoxOffice","Budget","Rating"
18.5415056072234,17.8450818603887,6
18.5394846823102,17.4295664164271,8.6
14.0463875402676,15.7248183241886,8.6
16.6481748348401,15.5424967673947,4.9
18.5375207502074,17.993501865507,7.2
17.8330702692078,16.4179655047486,6.8
18.258646677125,18.2097249739766,8.4
18.1442378548588,18.0274034171827,8.3
17.3390747090169,17.3342562366227,8.3
17.6811101114189,17.884302573542,8.2
17.2616855439472,15.185821823456,7.3
19.1479832214253,18.5382290409487,5.4
17.9785887851711,17.3003546849471,8.1
18.542970273207,17.8450818603887,6.9
15.9368618306844,14.3385239630687,7
18.1172112443766,16.877497834127,6.8
15.9564590057803,16.9287911285146,6.2
14.4848043244582,18.0601932400057,7.6
18.0286377051639,17.48840691645,5.8
18.8859221800598,17.9403920401931,6.6
18.4131897170397,19.0978448291381,6.1
16.946176817611,17.7397213447309,4.6
18.1879006577734,16.8234306128567,5.8
17.1752116567137,16.2356439479546,6.4
17.3978051118897,15.2139927004226,7.6
15.6160986544187,14.8493495868347,8.1
15.3919833647409,15.3193532160805,6.7
```

Environment is empty

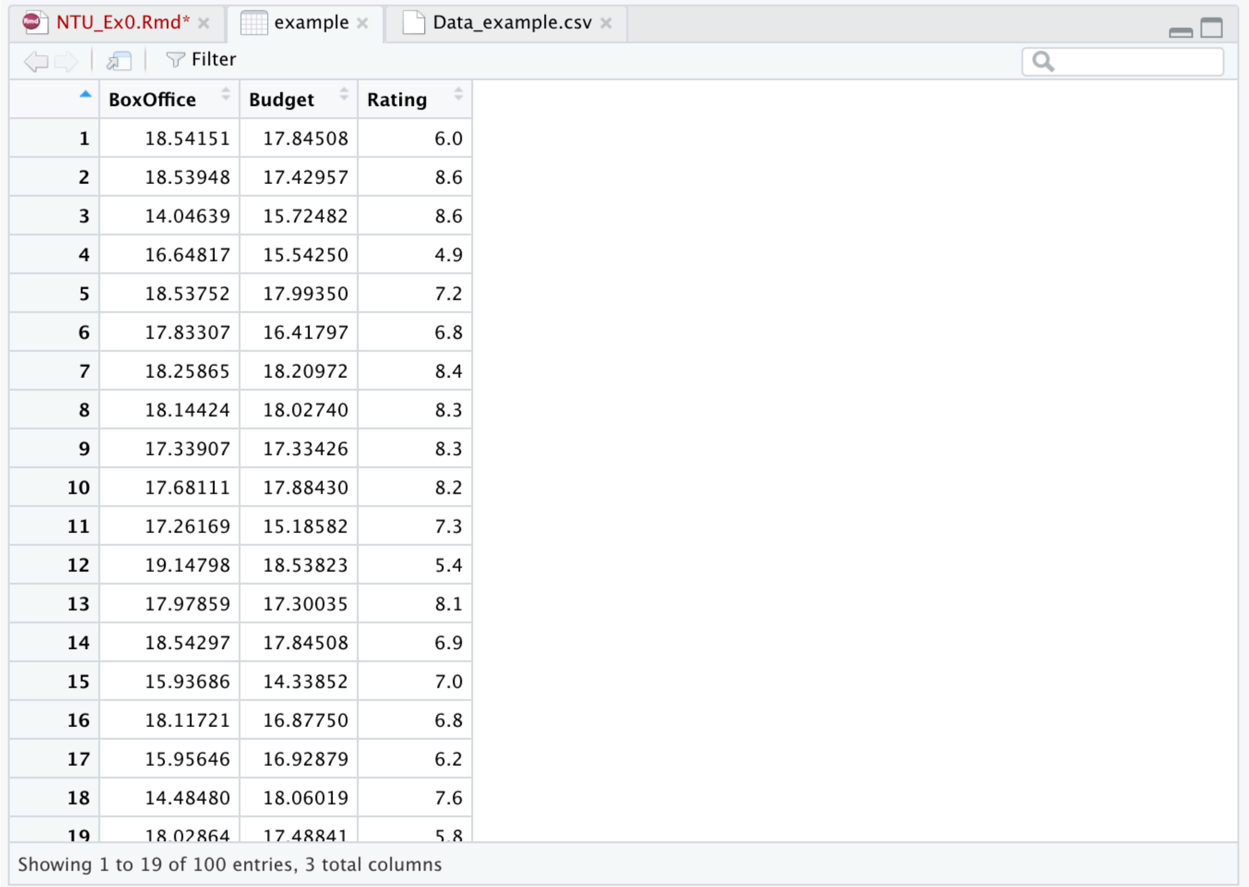
```
  Name                  Size     Modified
  ..
□ NTU_Ex0.Rmd           422 B    Jul 28, 2022, 3:30 PM
□ Data_example.csv      3.7 KB   Jul 28, 2022, 4:53 PM
□ .Rhistory             0 B      Jul 28, 2022, 3:18 PM
□ .gitattributes        66 B     Jul 28, 2022, 3:16 PM
```

```
Error in knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Install R and RStudio") :
  Cannot find the file(s): "/Users/tiffany/Desktop/NTU_Exercise0/Install R and RStudio"
(*) NOTE: I saw chunk options ", out.width=50%"
 please go to https://yihui.org/knitr/options
 (it is likely that you forgot to quote "character" options)
> knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Install R and RStudio.png")
(*) NOTE: I saw chunk options ", out.width=50%"
 please go to https://yihui.org/knitr/options
 (it is likely that you forgot to quote "character" options)
> # screenshot: the proof of import dataset
> knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Import Dataset.png")
> 
```

```r
# Read (and import) the full example data set into R using read.csv()
example <- read.csv(file = 'Data_example.csv')
example
```

```
##      BoxOffice   Budget Rating
## 1     18.54151 17.84508    6.0
## 2     18.53948 17.42957    8.6
## 3     14.04639 15.72482    8.6
## 4     16.64817 15.54250    4.9
## 5     18.53752 17.99350    7.2
## 6     17.83307 16.41797    6.8
## 7     18.25865 18.20972    8.4
## 8     18.14424 18.02740    8.3
## 9     17.33907 17.33426    8.3
## 10    17.68111 17.88430    8.2
## 11    17.26169 15.18582    7.3
## 12    19.14798 18.53823    5.4
## 13    17.97859 17.30035    8.1
## 14    18.54297 17.84508    6.9
## 15    15.93686 14.33852    7.0
## 16    18.11721 16.87750    6.8
## 17    15.95646 16.92879    6.2
## 18    14.48480 18.06019    7.6
## 19    18.02864 17.48841    5.8
## 20    18.88592 17.94039    6.6
## 21    18.41319 19.09784    6.1
## 22    16.94618 17.73972    4.6
## 23    18.18790 16.82343    5.8
## 24    17.17521 16.23564    6.4
## 25   17.39781 15.21399    7.6
## 26   15.61610 14.84935    8.1
## 27   15.39198 15.31935    6.7
## 28   17.17663 15.87897    7.2
## 29   17.69850 16.73642    7.7
## 30   16.23757 16.76627    7.7
## 31   18.53826 18.43287    7.6
## 32   16.71286 16.01250    6.6
## 33   15.33531 17.15193    5.2
## 34   17.56445 16.64111    5.0
## 35   16.98179 16.49801    6.7
## 36   17.59726 16.33095    7.6
## 37   18.32821 16.76627    6.6
## 38   18.13176 17.33426    6.8
## 39   18.44830 17.90335    7.3
## 40   13.95823 16.13028    7.7
## 41   17.47945 18.18155    5.4
## 42   14.68568 15.72482    7.2
## 43   18.26104 17.84508    5.0
## 44   18.21079 17.34256    6.9
## 45   17.11941 16.23564    6.8
## 46   17.36259 16.18435    6.5
## 47   17.91595 17.99350    6.2
## 48   16.85818 17.39879    6.4
## 49   17.77415 16.41797    5.9
## 50   18.61875 17.33426    6.3
## 51   16.19213 15.80486    5.7
## 52   18.17970 17.15193    6.6
## 53   18.20299 17.48841    7.5
## 54   18.10745 18.06019    6.8
## 55   17.84556 18.02740    5.4
## 56   18.14997 17.84508    6.5
## 57   16.17564 17.02410    5.6
## 58   16.00232 17.67073    7.2
## 59   16.30271 17.33426    5.8
## 60   16.43460 16.64111    5.3
## 61   17.84533 16.92879    6.6
## 62   17.56007 17.94039    5.9
## 63   16.99104 16.92879    6.8
## 64   16.27486 17.84508    5.1
## 65   17.77391 17.02410    5.8
## 66   17.38622 17.26526    6.9
## 67   16.88149 17.15193    5.5
## 68   16.80806 16.41797    5.4
## 69   17.50590 17.73972    5.9
## 70   16.23737 18.51803    5.6
## 71   18.13853 16.70565    5.6
## 72   17.34441 17.84508    6.3
## 73   16.91630 16.92879    4.4
## 74   16.52030 16.41797    6.0
## 75   16.40098 15.72482    6.1
## 76   16.79630 16.92879    6.6
## 77   16.70753 17.22890    6.7
## 78   13.29774 15.87897    6.8
## 79   12.20767 16.33095    5.9
## 80   14.68114 16.33095    7.0
## 81   17.40042 16.92879    6.6
## 82   16.04298 17.39879    5.8
## 83   16.54025 17.84508    7.1
## 84   15.42123 13.70992    7.5
## 85   17.49763 18.02740    5.6
## 86   15.83517 16.64111    7.1
## 87   15.68779 16.92879    6.4
## 88   18.27158 17.33426    5.9
## 89   17.11311 17.33426    5.5
## 90   16.37840 17.84508    4.1
## 91   17.09321 16.23564    6.2
## 92   16.05525 15.87897    7.4
## 93   16.21493 17.84508    5.1
## 94   14.88857 16.18113    7.5
## 95   15.34405 16.41797    6.6
## 96   16.33814 15.63781    7.3
## 97   14.17679 14.62621    6.7
## 98   16.94615 17.48841    5.6
## 99   16.63726 17.62194    5.9
## 100  16.50113 17.15193    6.9
```

```
# view the data example in R
# View(example)
knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/View Example.png")
```



	BoxOffice	Budget	Rating
1	18.54151	17.84508	6.0
2	18.53948	17.42957	8.6
3	14.04639	15.72482	8.6
4	16.64817	15.54250	4.9
5	18.53752	17.99350	7.2
6	17.83307	16.41797	6.8
7	18.25865	18.20972	8.4
8	18.14424	18.02740	8.3
9	17.33907	17.33426	8.3
10	17.68111	17.88430	8.2
11	17.26169	15.18582	7.3
12	19.14798	18.53823	5.4
13	17.97859	17.30035	8.1
14	18.54297	17.84508	6.9
15	15.93686	14.33852	7.0
16	18.11721	16.87750	6.8
17	15.95646	16.92879	6.2
18	14.48480	18.06019	7.6
19	18.02864	17.48841	5.8

Showing 1 to 19 of 100 entries, 3 total columns

4.Calculate Mean, Variance, SD, Max, Min, Median of the variable Rating

```
# Summary
example %>%
  summarize(Mean = mean(Rating),
            Variance = var(Rating),
            SD = sd(Rating),
            Max = max(Rating),
            Min = min(Rating),
            Median = median(Rating))
```

```
##    Mean  Variance        SD Max Min Median
## 1 6.507 0.9265162 0.9625571 8.6 4.1    6.6
```

5.For the first 10 values of Rating, do logarithm, exponential, divide the vector by a number, multiplication by a number.

```
# x represents a vector for the first 10 values of Rating
# subtract the first 10 values of Rating
x <- example$Rating %>%
  head(10)

# logarithm
log_x <- log(x)
log_x
```

```
##  [1] 1.791759 2.151762 2.151762 1.589235 1.974081 1.916923 2.128232 2.116256
##  [9] 2.116256 2.104134
```

```
# exponential
exp_x <- exp(x)
exp_x
```

```
##  [1]  403.4288 5431.6596 5431.6596  134.2898 1339.4308  897.8473 4447.0667
##  [8] 4023.8724 4023.8724 3640.9503
```

```
# divide the vector by a number (2)
division_x <- x/2
division_x
```

```
##  [1] 3.00 4.30 4.30 2.45 3.60 3.40 4.20 4.15 4.15 4.10
```

```
# multiplication by a number (5)
multiply_x <- x*5
multiply_x
```

```
##  [1] 30.0 43.0 43.0 24.5 36.0 34.0 42.0 41.5 41.5 41.0
```

6.Vector, matrix, entries in a matrix, multiplication of vectors, matrices.

```
# define three 4*1 vectors
v1=c(1,2,3,4)
v2=c(6,2,8,4)
v3=c(9,3,6,1)
```

```
# add the number 3 to v1
v1+3
```

```
## [1] 4 5 6 7
```

```
# add v1 to v2
v1+v2
```

```
## [1]  7  4 11  8
```

```
# product of v1 and v2
v1*v2
```

```
## [1]  6  4 24 16
```

```
# create a matrix, using v1,v2,v3 as the columns
cnames <- c("v1", "v2","v3")
matrix1 <- matrix(c(v1,v2,v3), nrow=4, ncol=3)
colnames(matrix1) = cnames
matrix1
```

```
##      v1 v2 v3
## [1,]  1  6  9
## [2,]  2  2  3
## [3,]  3  8  6
## [4,]  4  4  1
```

```
# print the element in row 1 and columns 1
matrix1[1,1]
```

```
## v1
##  1
```

```
# print the element in row 1 and columns 3
matrix1[1,3]
```

```
## v3
##  9
```

```
# print the first two elements in row 1
matrix1[1,1:2]
```

```
## v1 v2
##  1  6
```

```
# print the first three elements in row 1
matrix1[1,1:3]
```

```
## v1 v2 v3
##  1  6  9
```

```
# print the elements in the first row
matrix1[1,]
```

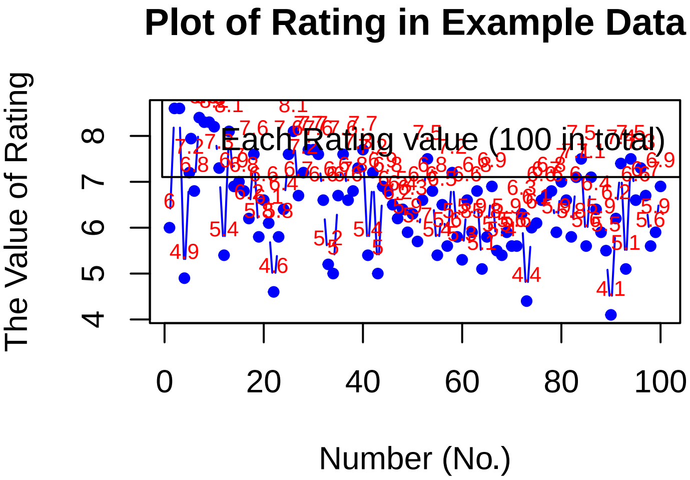
```
## v1 v2 v3
##  1  6  9
```

```
# print the elements in the first column
matrix1[,1]
```

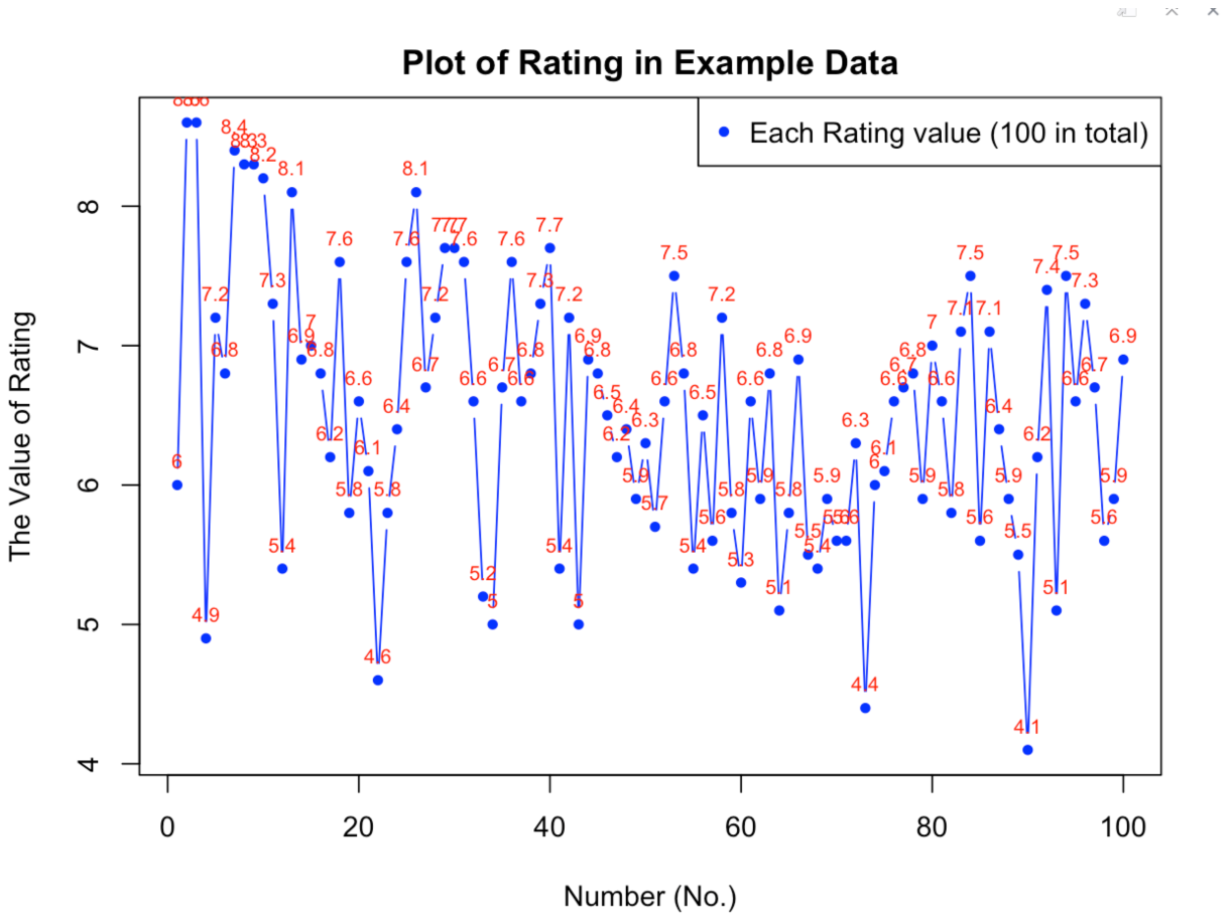
```
## [1] 1 2 3 4
```

7. Make a plot using the variable Rating. Try to make the plot better looking and more informative.

```r
# make a plot by Rating
plot(example$Rating, type = "b", pch = 20, col = "blue",
     main = "Plot of Rating in Example Data",
     xlab = "Number (No.)", ylab = "The Value of Rating")
text(example$Rating, labels=example$Rating, cex=0.7, pos=3, col="red")
legend("topright", "Each Rating value (100 in total)", pch=20, col="blue")
```



```r
knitr::include_graphics("/Users/tiffany/Desktop/NTU_Exercise0/Plot of Rating.png")
```



8. Linear regression and interpret the results.

```
# use BoxOffice as dependent variable; Budget and Rating as independent variables to run the regression
lr_example <- lm(BoxOffice ~ Budget + Rating, data = example)
```

```
# show the regression results
summary(lr_example)
```

```
##
## Call:
## lm(formula = BoxOffice ~ Budget + Rating, data = example)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2438 -0.6313  0.2161  0.8380  1.5671
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.53866    2.31246   1.963   0.0525 .
## Budget       0.69615    0.11827   5.886 5.66e-08 ***
## Rating       0.09221    0.12074   0.764   0.4469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.128 on 97 degrees of freedom
## Multiple R-squared:  0.2649, Adjusted R-squared:  0.2497
## F-statistic: 17.47 on 2 and 97 DF,  p-value: 3.303e-07
```

Interpretation:

From the summary above, we can find that this linear regression model fits the data NOT well.

The p-value for Intercept is around 0.0525 with the one period signify (.) and p-value for Rating is around 0.4469 with a blank, which means the coefficients are not very significant. Thus, the model fit not well. However, the p-value for Budget is around 5.66e-08 with the three asterisks signify (***), which means this coefficient is very significant.

Besides, the Multiple R-squared value is about 0.2649 and Adjusted R-square value is about 0.2497, which are small. It means 25 percentage of the variation within our dependent variable that all predictors are explaining. Thus, the model is not fitting the data very well.

9. Install a package: stargazer

```
# install.packages('stargazer')
```

10. Use functions from this package to output the regression results

```
library(stargazer)

# output the regression results
stargazer(lr_example, type = "text", title = "Linear regression model of data example")
```

```
##
## Linear regression model of data example
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                              BoxOffice
## -----------------------------------------------
## Budget                       0.696***
##                               (0.118)
##
## Rating                         0.092
##                               (0.121)
##
## Constant                      4.539*
##                               (2.312)
##
## -----------------------------------------------
## Observations                    100
## R2                             0.265
## Adjusted R2                    0.250
## Residual Std. Error       1.128 (df = 97)
## F Statistic           17.474*** (df = 2; 97)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```