Fundamentals of Data Science & Analytics
Rollins College
Ramon A. Mata-Toledo Ph.D.
Assignment on Linear Regression

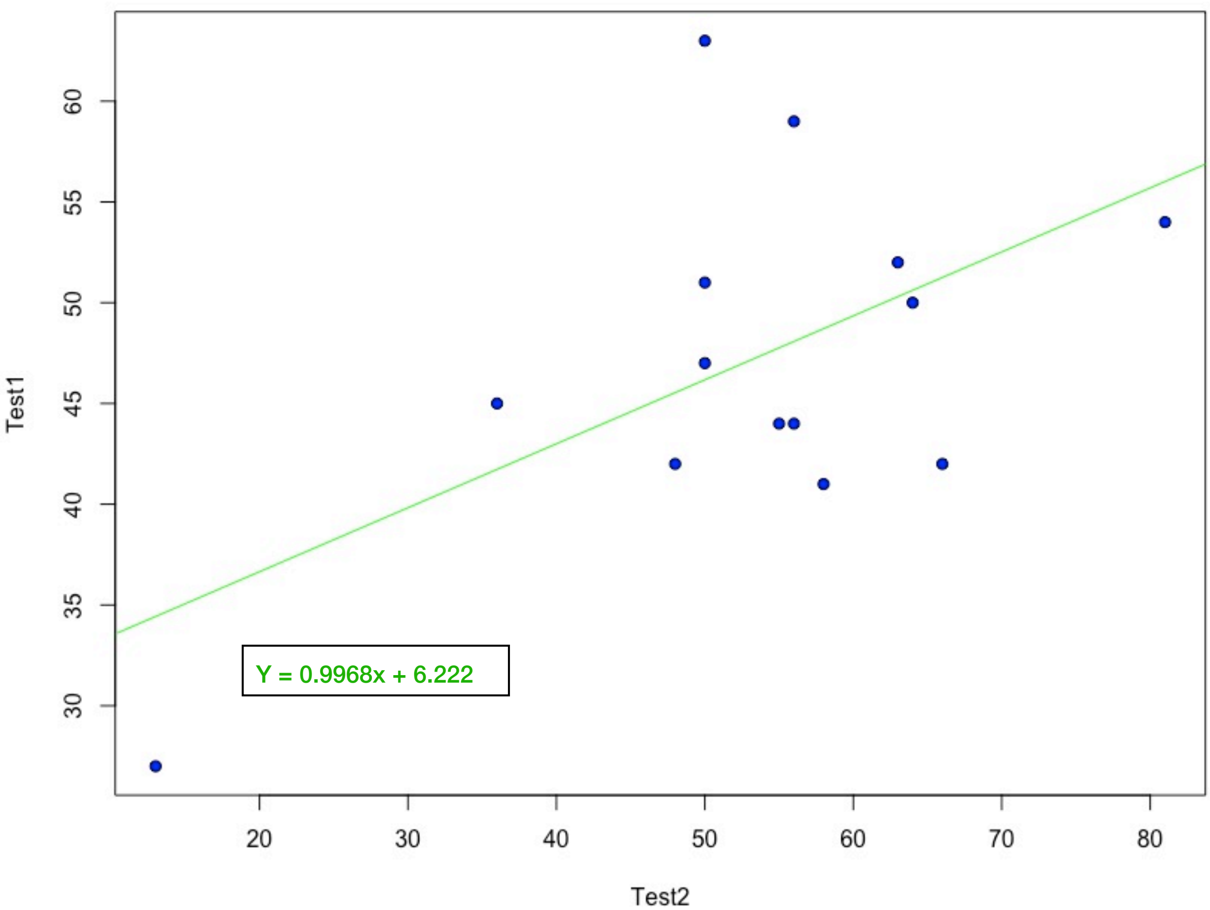
Team Member : Alejandra De Osma, Shannon Polk, Jack Gabriel

Upload a document with the R instructions for the first question and the graphs that you produce.For the second question, upload the graphs of the variables that are correlated and explain what type they may be based on the PowerPoint that we covered during the lecture. This assignment can be answered in a couple of hours. However, as usual, I give you the usual time to do it but, why wait that long?

1) Given the table shown below, crate a regression model to predict the scores of test No. 2 based upon the scores of test No. 1 score. What would your model predict for someone who got 46 in test No. 1?What are the slope and intercepts of your model? Plot the graph of this linear correlation. (Question taken from Data Analytics by A. Maheswari. McGraw-Hill 2017)

Test1	Test2	Test1 X Test2	Test1^2	Test2^2
59	56	3304	3481	3136
52	63	3276	2704	3969
44	55	2420	1936	3025
51	50	2550	2601	2500
42	66	2772	1764	4356
42	48	2016	1764	2304
41	58	2378	1681	3364
45	36	1620	2025	1296
27	13	351	729	169
63	50	3150	3969	2500
54	81	4374	2916	6561
44	56	2464	1936	3136
50	64	3200	2500	4096
47	50	2350	2209	2500

Test 1 vs Test




```
   Test1 Test2
1     59    56
2     52    63
3     44    55
4     51    50
5     42    66
6     42    48
7     41    58
8     45    36
9     27    13
10    63    50
11    54    81
12    44    56
13    50    64
14    47    50
> lm(formula = Test1~Test2)

Call:
lm(formula = Test1 ~ Test2)

Coefficients:
(Intercept)        Test2
    30.3033       0.3174

> abline(lm(formula = Test1~Test2),col = "green")
> plot(Test1~Test2,pch=21,bg="blue")
> abline(lm(formula = Test1~Test2),col = "green")
```

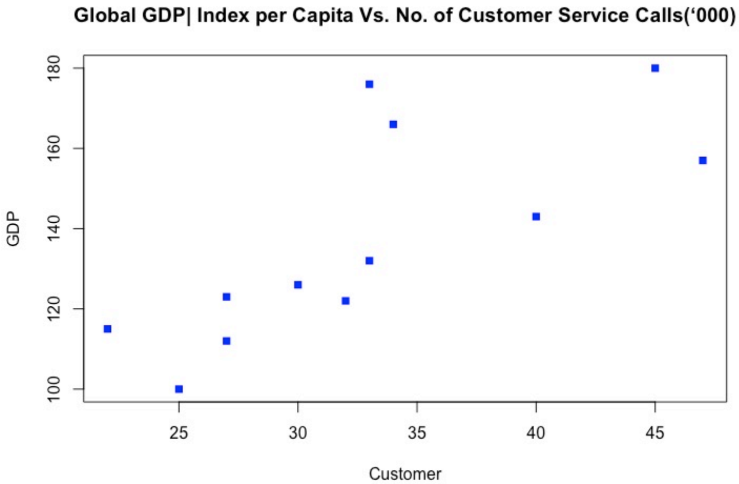
Equation Of The Line

Y = 0.9968x + 6.222 , X = 46
Y = 0.9968 (46) + 6.222
Y = 45.8528 + 6.222
Y = 52.0748

If the student got a 46 in the first exam, through a linear regression model analysis, the student is predicted to receive a 52.0748 in Test number 2.

Given the table shown below, which variables are strongly correlated. How can you tell that? Create a regression model that best predicts the revenue.
What are the slope and intercepts of your model? Plot the graph of this linear correlation.
(Question taken from Data Analytics by A. Maheswari. McGraw-Hill 2017)

Year	GDP	Customer	Employee	Items	Revenue
1	100	25	45	11	2000
2	112	27	53	11	2400
3	115	22	54	12	2700
4	123	27	58	14	2900
5	122	32	60	14	3200
6	132	33	65	15	3500
7	143	40	72	16	4000
8	126	30	65	16	4200
9	166	34	85	17	4500
10	157	47	97	18	4700
11	176	33	98	18	4900
12	180	45	100	20	5000

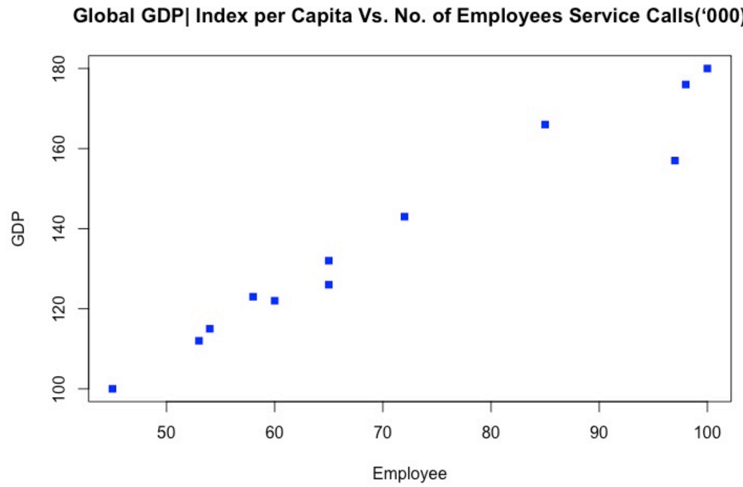



```
> lm(formula = GDP~Customer)

Call:
lm(formula = GDP ~ Customer)

Coefficients:
(Intercept)     Customer
     53.101        2.569
```

No Correlation

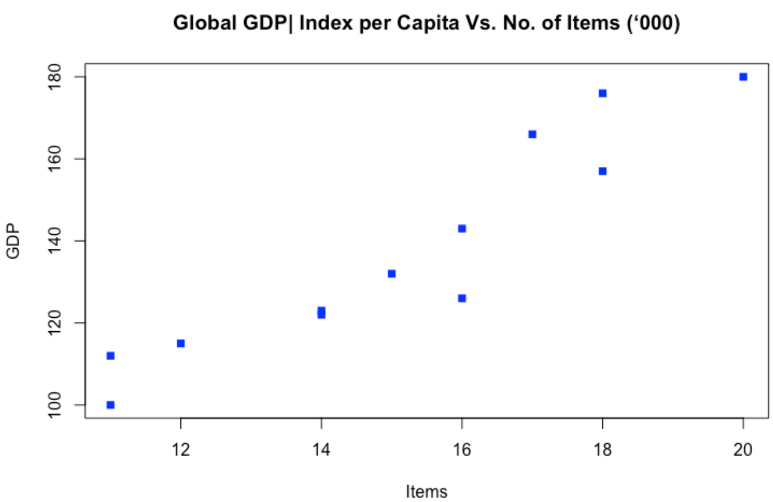


Strongly Correlated

```
> lm(formula = GDP~Employee)

Call:
lm(formula = GDP ~ Employee)

Coefficients:
(Intercept)     Employee
     42.960        1.334
```

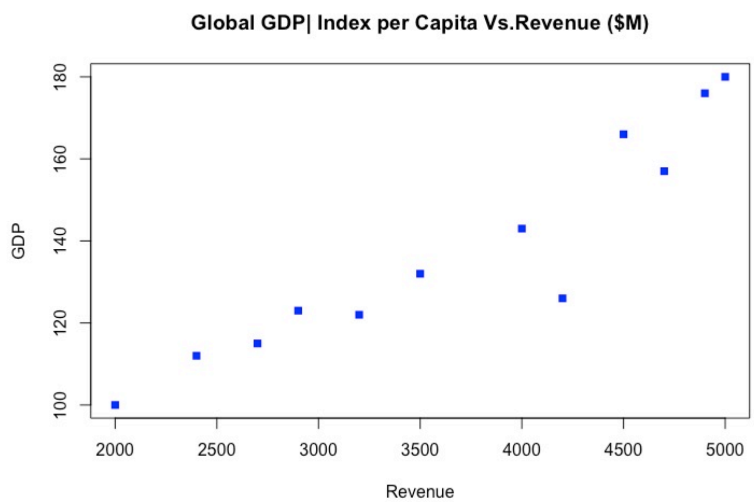



```
> lm(formula = GDP~ Items)

Call:
lm(formula = GDP ~ Items)

Coefficients:
(Intercept)        Items
      7.509        8.582
```

Correlated



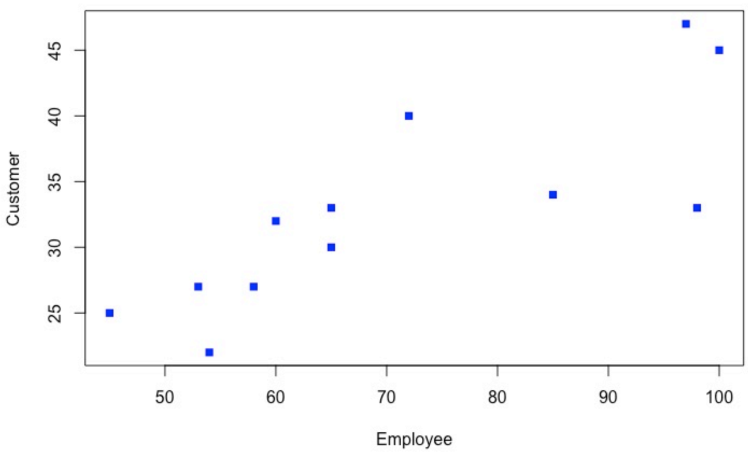
Strongly Correlated

```
> lm(formula = GDP~ Revenue)

Call:
lm(formula = GDP ~ Revenue)

Coefficients:
(Intercept)      Revenue
   49.34865      0.02409
```

No. of Customer Service Calls('000) Vs.No.of Employees Calls('000)



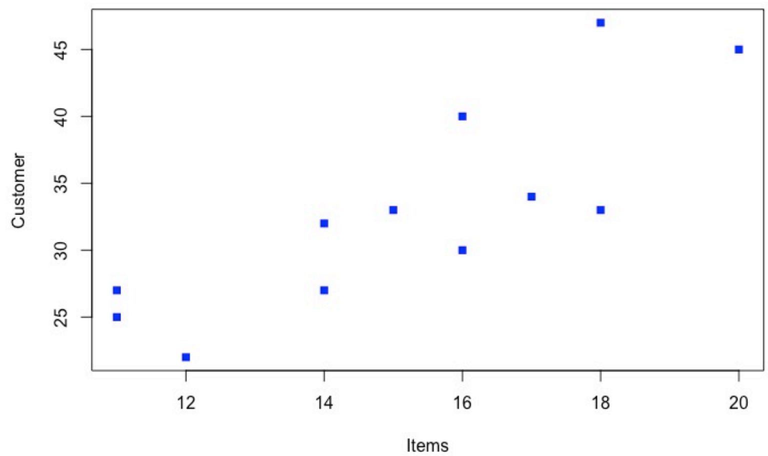

```
> lm(formula = Customer ~ Employee)

Call:
lm(formula = Customer ~ Employee)

Coefficients:
(Intercept)     Employee
      9.626        0.328
```

No Correlation

No. of Customer Service Calls('000) Vs. No. of Items Calls ('000)



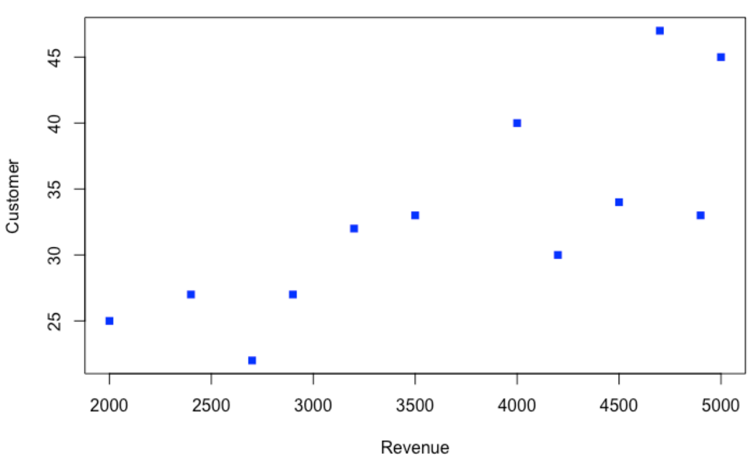
No Correlation

```
> lm(formula = Customer~ Items)

Call:
lm(formula = Customer ~ Items)

Coefficients:
(Intercept)        Items
    -0.8636       2.2273
```

No. of Customer Service Calls('000) Vs.Revenue ($M)



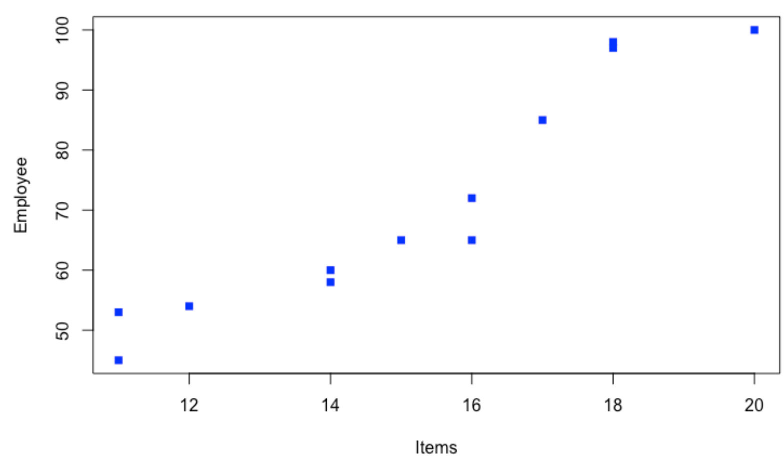

```
> lm(formula = Customer~Revenue)

Call:
lm(formula = Customer ~ Revenue)

Coefficients:
(Intercept)      Revenue
  11.224153     0.005916
```

No Correlation

No. of Employees Calls('000) Vs. No. of Items ('000)

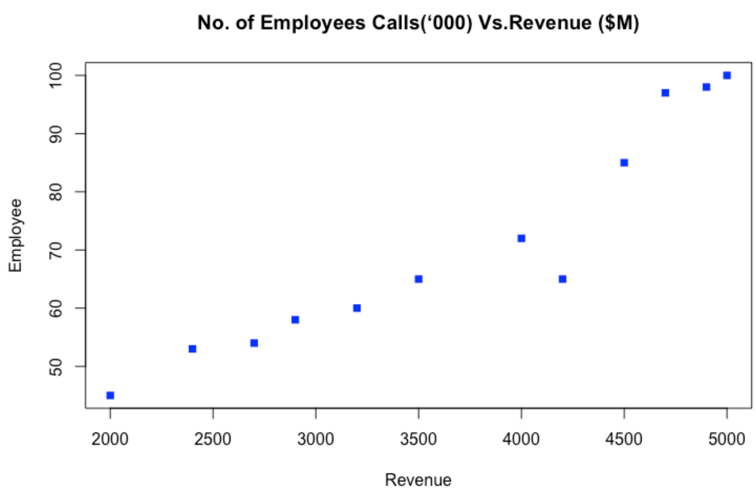


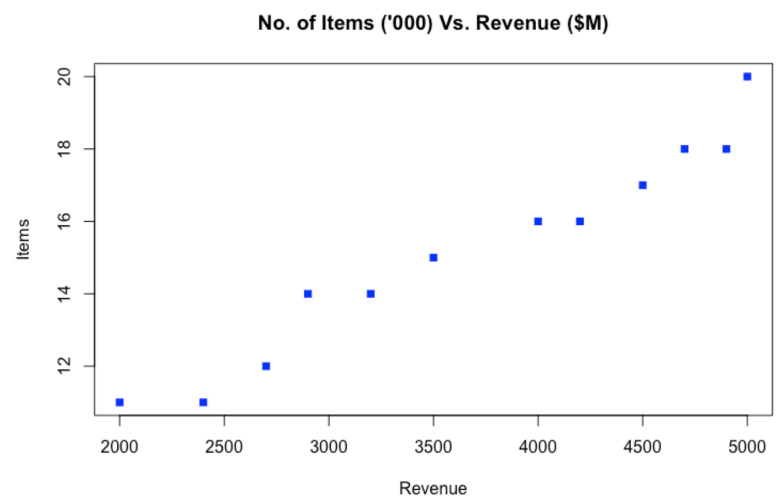
Correlated

```
> lm(formula= Employee~Items)

Call:
lm(formula = Employee ~ Items)

Coefficients:
(Intercept)         Items
    -24.633         6.305
```






```
> lm(formula= Employee~Revenue)

Call:
lm(formula = Employee ~ Revenue)

Coefficients:
(Intercept)      Revenue
    5.82769      0.01777
```

Correlated

Strongly Correlated

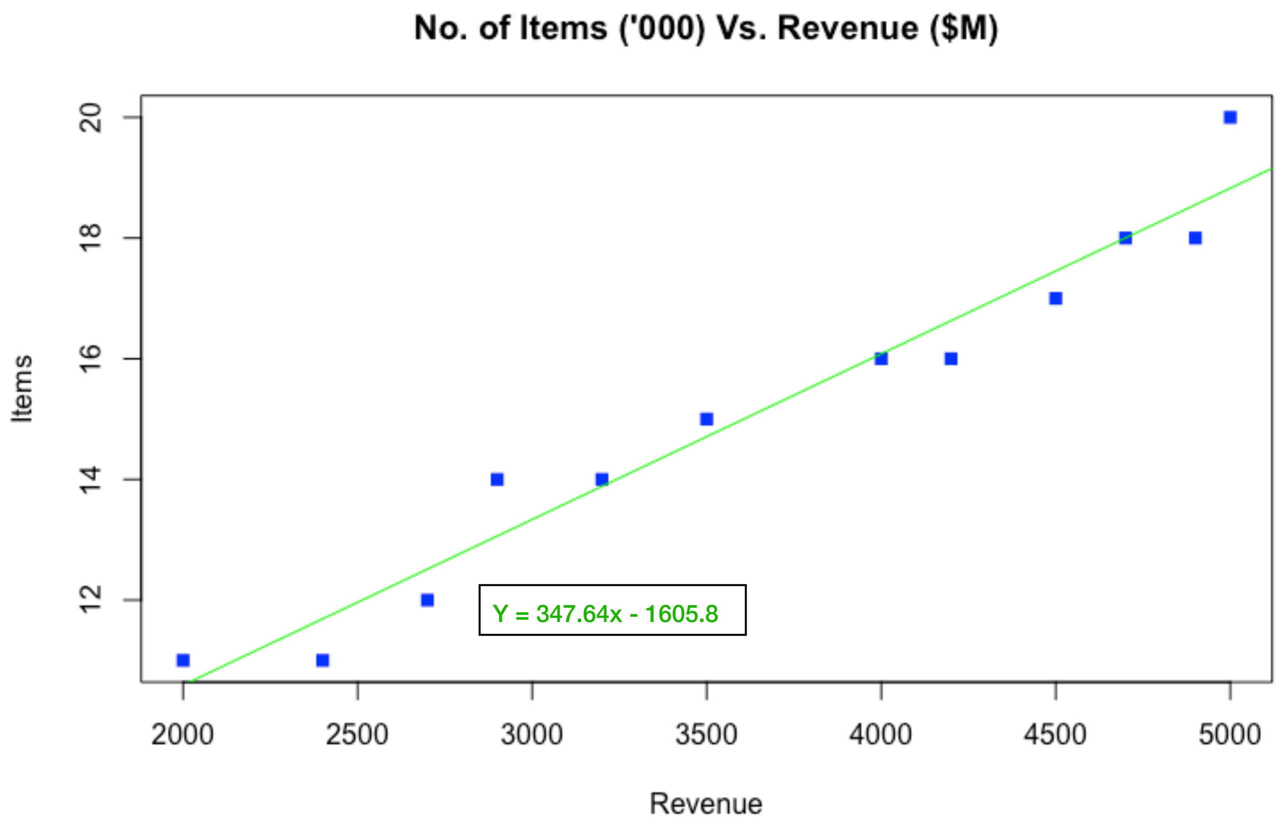
```
> lm(formula = Items~Revenue)

Call:
lm(formula = Items ~ Revenue)

Coefficients:
(Intercept)      Revenue
   5.099655     0.002746
```

Identifying correlation:

Linear correlation: We can identify correlation through the visual analysis of the data points. If they follow a linear pattern, we can predict that both variables are correlated. On the other had if the data does not visually show a linear trend we can predict that the data is not correlated.




```
> select(data_lr,Items,Revenue)
   Items Revenue
1     11    2000
2     11    2400
3     12    2700
4     14    2900
5     14    3200
6     15    3500
7     16    4000
8     16    4200
9     17    4500
10    18    4700
11    18    4900
12    20    5000
> lm(formula = Items~Revenue)

Call:
lm(formula = Items ~ Revenue)

Coefficients:
(Intercept)      Revenue
   5.099655     0.002746
```

Equation Of The Line-1

Y= 347.64x - 1605.8	**Estimations:**	**Difference**
X = 14	3261.21	61.21
X = 16	3956.44	43.56
X =18	4651.72	48.28

No. Of Items, Is the most accurate estimator for Revenue from the given data set. As shown above we where able to predict revenue with a margin of error of approximately ± 50. We tested predictions with values, x = 14, x = 16 and x = 18.

```
> plot(Items~Revenue, pch = 15,col = "blue")
> abline(lm(Items~Revenue), pch = 15, col = "green")
> title("No. of Items ('000) Vs. Revenue ($M)")
```