Data Analytics Assignment 5

Jing Pang, Tian Xue, Chuqian Ma, Jiaxiang Leng

1. Replicate the analysis of the correlation between Nobel prizes and chocolate consumption

a. Find data from the given website and create a new .csv file to store these data
In this assignment, choosing the data of cocoa consumption per capita by countries in 2010. Then, creating a new file to store Nobel laureates per 10 million population and chocolate population per capita by countries.

| Rank	Entity	Nobel laureates	Population	Laureates/ 10 million	2010cocoa
50	Algeria	2	42008054	0.476	NA
47	Argentina	5	44688864	1.119	NA
26	Australia	12	24772247	4.844	4.61
8	Austria	21	8751820	23.995	8.16
49	Azerbaijan	1	9923914	1.008	NA
72	Bangladesh	1	1.66E+08	0.06	NA
36	Belarus	2	9452113	2.16	NA
19	Belgium	10	11498519	8.697	8.1
24	Bosnia and Herzegovina	2	350354	5.708	NA
74	Brazil	1	2.11E+08	0.047	2.93
45	Bulgaria	1	7036848	1.421	1.26
22	Canada	24	33953765	6.495	NA
48	Chile	2	18197209	1.099	NA
71	China	9	1.42E+09	0.064	NA
53	Colombia	2	49464683	0.404	NA
38	Costa Rica	1	4953199	2.019	NA
34	Croatia	1	4164783	2.401	NA
20	Cyprus	1	1189085	8.41	NA
29	Czech Republic	5	10625250	4.706	NA
6	Denmark	14	5754356	24.329	8.49
67	DR Congo	1	84004989	0.119	NA
10	East Timor	2	1324094	15.105	NA
54	Egypt	4	99375741	0.403	NA
/	European Union	481	5.1E+08	9.437	NA
/	Faroe Islands	1	49489	202.065	NA
18	Finland	5	5542517	9.021	7.26
16	France	63	65233271	9.658	6.34
13	Germany	108	82293457	13.124	11.56

b. R programming to input data and process data
Input the data, and use function complete.case() to remove missing values.

```
which(complete.cases(combine)==F)
newcombine<-combine[complete.cases(combine),]
```

Rank	Entity	Nobel.laureates	Population.2018.	Laureates..10.million	X2010cocoa
26	Australia	12	24772247	4.844	4.61
8	Austria	21	8751820	23.995	8.16
19	Belgium	10	11498519	8.697	8.10
74	Brazil	1	210867954	0.047	2.93
45	Bulgaria	1	7036848	1.421	1.26
6	Denmark	14	5754356	24.329	8.49
18	Finland	5	5542517	9.021	7.26
16	France	63	65233271	9.658	6.34
13	Germany	108	82293457	13.124	11.56
41	Greece	2	11142161	1.795	3.51
17	Hungary	9	9688847	9.289	3.33
32	Italy	20	59260969	3.373	3.74
35	Japan	27	127185332	2.123	2.09
31	Lithuania	1	2876475	3.476	5.40
7	Norway	13	5353363	24.284	9.44
33	Poland	12	38104832	3.149	3.61
40	Porturgal	2	10291196	1.943	2.72
43	Spain	8	46397452	1.724	3.60

c. Plot the correlation between the two quantities
Use function plot() to draw a diagram of the relation between nobel laureates and chocolate consumption.
Use function text() to mark the country of each point.
Use Summary (relation) get the data of linear regression.

```
x=newcombine[,6]
y=newcombine[,5]
relation <- lm(x~y)
summary(relation)
plot(x, y, abline(lm(y~x)), xlab="cocoa", ylab="nobel", col="blue")
text(x, y, newcombine[,2], cex=0.7, col="green")
```
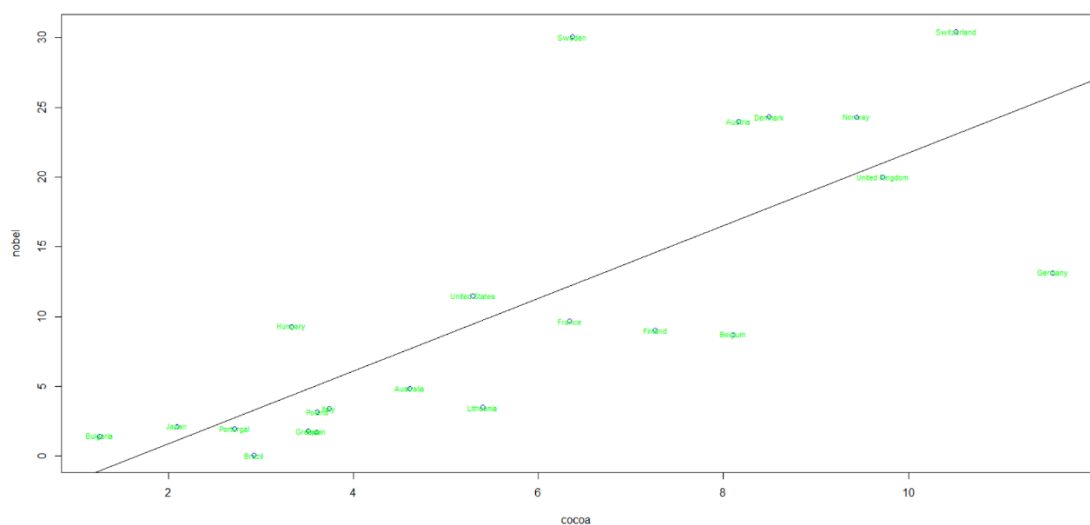
```
> summary(relation)

Call:
lm(formula = x ~ y)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8056 -0.6683 -0.3260  0.7317  5.2199

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.3664     0.6153   5.471 2.35e-05 ***
y             0.2266     0.0421   5.382 2.87e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.938 on 20 degrees of freedom
Multiple R-squared:  0.5916,  Adjusted R-squared:  0.5711
F-statistic: 28.97 on 1 and 20 DF,  p-value: 2.875e-05
```

2. **Correlation between Nobel prizes and wine consumption**

   a. Find data from websites about wine consumption per capita by country and create a new .csv file to store these data
   Website: https://www.statista.com/statistics/232754/leading-20-countries-of-wine-consumption/

6	Denmark	14	5754356	24.329	35.9
67	DR Congo	1	84004989	0.119	NA
10	East Timor	2	1324094	15.105	NA
54	Egypt	4	99375741	0.403	NA
/	European Union	481	5.1E+08	9.437	NA
/	Faroe Islands	1	49489	202.065	NA
18	Finland	5	5542517	9.021	NA
16	France	63	65233271	9.658	47.4
13	Germany	108	82293457	13.124	36.2
56	Ghana	1	29463643	0.339	NA
41	Greece	2	11142161	1.795	38.1
46	Guatemala	2	17245346	1.16	NA
/	Hong Kong	1	7428887	1.346	NA
17	Hungary	9	9688847	9.289	30.9

b. R programming to input data and process data

Rank	Entity	Nobel.laureates	Population.2018.	Laureates..10.million	X2014wine
47	Argentina	5	44688864	1.119	33.6
26	Australia	12	24772247	4.844	29.3
8	Austria	21	8751820	23.995	41.1
19	Belgium	10	11498519	8.697	30.5
6	Denmark	14	5754356	24.329	35.9
16	France	63	65233271	9.658	47.4
13	Germany	108	82293457	13.124	36.2
41	Greece	2	11142161	1.795	38.1
17	Hungary	9	9688847	9.289	30.9
32	Italy	20	59260969	3.373	48.1
14	Netherlands	20	17084459	11.707	27.0
23	New Zealand	3	4749598	6.316	27.1
40	Porturgal	2	10291196	1.943	42.6
37	Romania	4	19580634	2.043	25.5
43	Spain	8	46397452	1.724	19.8
4	Sweden	30	9982709	30.052	27.1
3	Switzerland	26	8544034	30.431	44.7
9	United Kingdom	133	66573504	19.978	23.9

c. Plot the correlation between the two quantities

```
x=newcombine[,6]
y=newcombine[,5]
relation <- lm(x~y)
summary(relation)
plot(x, y, abline(lm(y~x)), xlab="wine", ylab="nobel", col="blue")
text(x, y, newcombine[,2], cex=0.7, col="red")
```
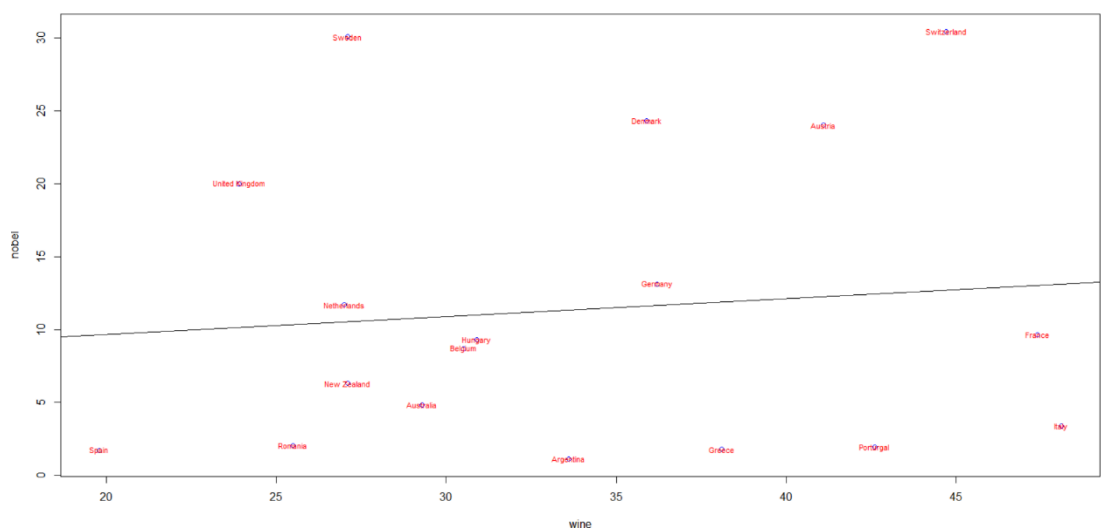
```
> summary(relation)

Call:
lm(formula = x ~ y)

Residuals:
    Min      1Q  Median      3Q     Max
-13.189  -6.711  -1.040   5.915  14.968

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 32.84034    3.11539  10.541 1.31e-08 ***
y            0.08646    0.20768   0.416    0.683
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.636 on 16 degrees of freedom
Multiple R-squared:  0.01072,  Adjusted R-squared:  -0.05111
F-statistic: 0.1733 on 1 and 16 DF,  p-value: 0.6827
```

3. Correlation between Nobel prizes and tea consumption

a. Find data from websites about wine consumption per capita by country and create a new .csv file to store these data

Website: https://www.statista.com/statistics/507950/global-per-capita-tea-consumption-by-country/

59	Morocco	1	36191805	0.276	2.68
66	Myanmar	1	53855735	0.186	NA
14	Netherlands	20	17084459	11.707	1.72
23	New Zealand	3	4749598	6.316	2.63
73	Nigeria	1	1.96E+08	0.051	NA
7	Norway	13	5353363	24.284	NA
69	Pakistan	2	2.01E+08	0.1	NA
39	Palestine	1	5052776	1.979	NA
58	Peru	1	32551815	0.307	NA
33	Poland	12	38104832	3.149	2.2
40	Porturgal	2	10291196	1.943	NA
37	Romania	4	19580634	2.043	NA
44	Russia	23	1.44E+08	1.598	3.05

b. R programming to input data and process data

26	Australia	12	24772247	4.844	1.65
71	China	9	1415045928	0.064	1.25
54	Egypt	4	99375741	0.403	2.23
13	Germany	108	82293457	13.124	1.52
11	Ireland	7	4803748	14.572	4.83
35	Japan	27	127185332	2.123	2.13
59	Morocco	1	36191805	0.276	2.68
14	Netherlands	20	17084459	11.707	1.72
23	New Zealand	3	4749598	6.316	2.63
33	Poland	12	38104832	3.149	2.20
44	Russia	23	143964709	1.598	3.05
61	Turkey	2	81916871	0.244	6.96
9	United Kingdom	133	66573504	19.978	4.28

c. Plot the correlation between the two quantities

```
x=newcombine[,6]
y=newcombine[,5]
relation <- lm(x~y)
summary(relation)
plot(x, y, abline(lm(y~x)), xlab="tea", ylab="nobel", col="blue")
text(x, y, newcombine[,2], cex=0.7, col="blue")
```
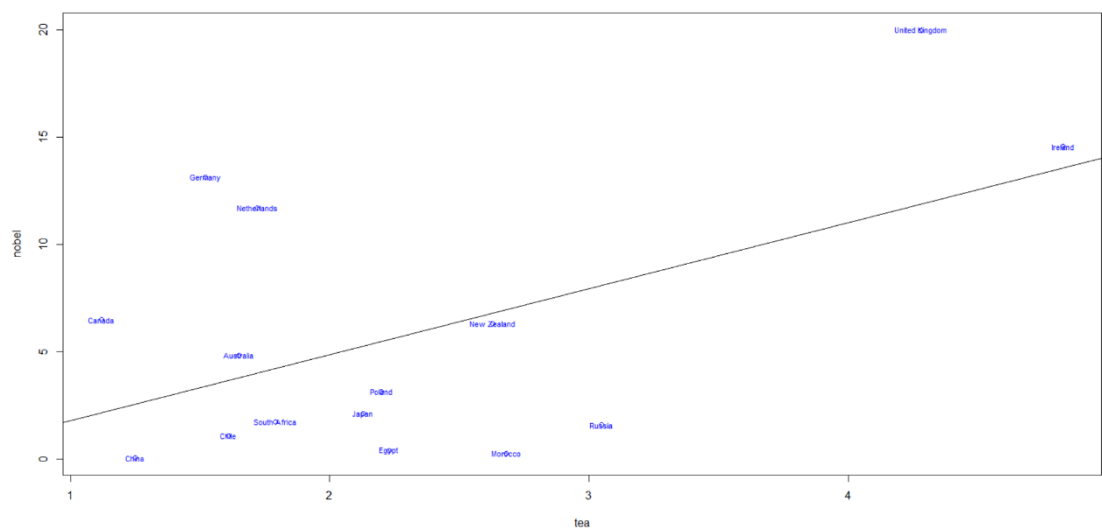
```
> summary(relation)

Call:
lm(formula = x ~ y)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4499 -0.5581  0.1292  0.5495  1.7296

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.78692    0.33774   5.291 0.000146 ***
y            0.09014    0.04041   2.231 0.043958 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9369 on 13 degrees of freedom
Multiple R-squared:  0.2768, Adjusted R-squared:  0.2212
F-statistic: 4.975 on 1 and 13 DF,  p-value: 0.04396
```

4. Result

In the three pairs of correlations, the first one replicate the correlation between Nobel prize and chocolate consumption in the lecture *"Chocolate Consumption, Cognitive Function, and Nobel Laureates"*, which shows that there is a linear regression between the two characteristics – the amount of Nobel Prizes of each country increases as the chocolate consumption increases.

The following two pairs relation shows that the wine consumption does not have a great influence on Nobel laureates and tea consumption of each country has a same positive influence as chocolate consumption on the amount of Nobel Prize.