INFSCI 2725: Data Analytics

Assignment 6

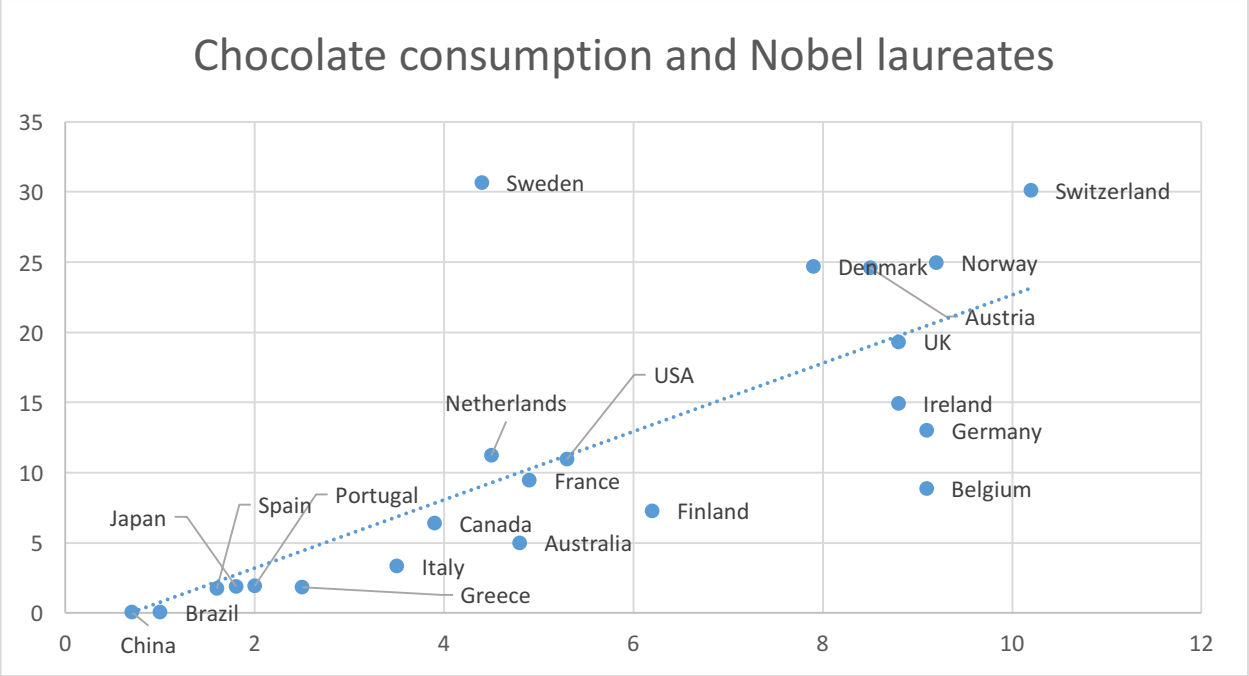
Tong Wei

TOW6

1. Chocolate consumption, cognitive function, and Nobel laureates

We replicate the analysis from the article. Data on chocolate consumption in kg per capita per year in 22 countries was obtained from Caobisco (http://www.theobroma-cacao.de/wissen/wirtschaft/international/konsum). Data on the number of Nobel Laureates per 10 million population in 22 countries was obtained from Wikipedia (https://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita).

We generated the scatter plot and did a linear regression data analysis.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.75089642
R Square	0.56384543
Adjusted R Sq	0.5420377
Standard Erro	6.79046732
Observations	22

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1192.19956	1192.19956	25.855303	5.6574E-05
Residual	20	922.208928	46.1104464		
Total	21	2114.40849			

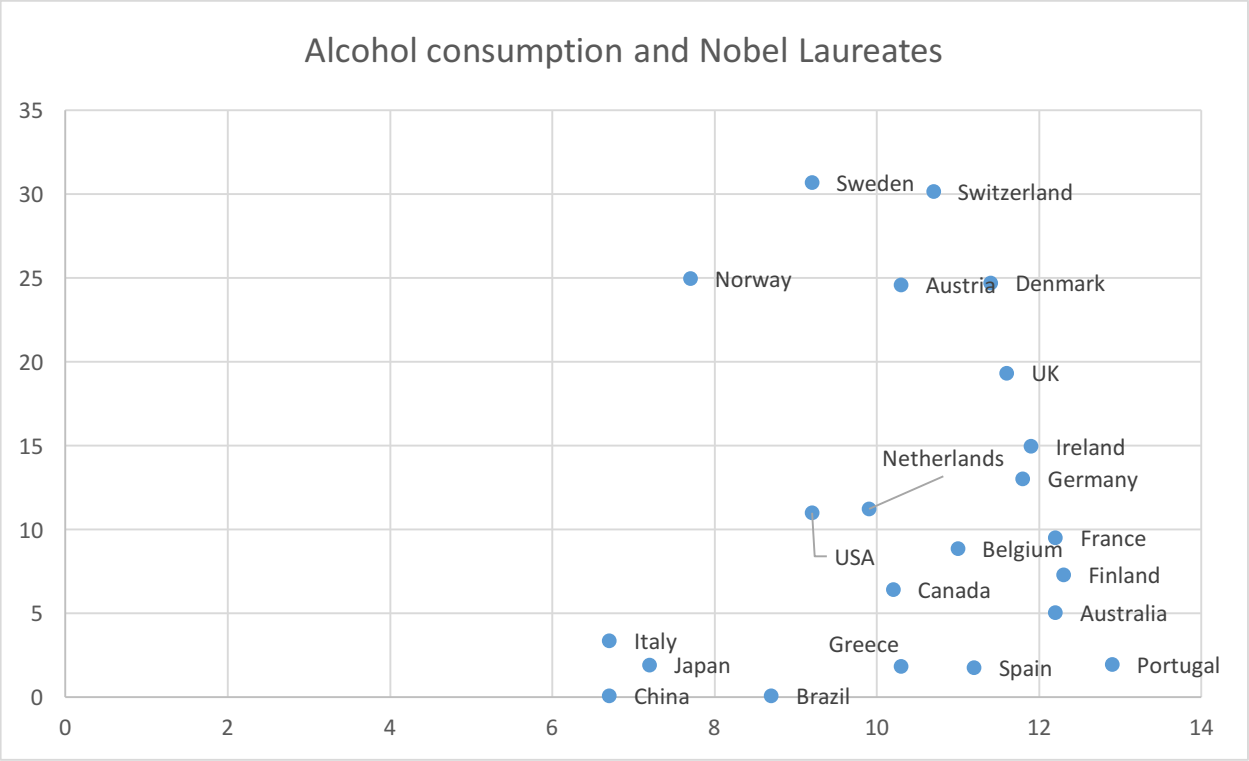
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.6520022	2.95881698	-0.558332	0.58281499	-7.8239863	4.51998182	-7.8239863	4.51998182
chocolate	2.43187068	0.47826176	5.08481101	5.6574E-05	1.43423412	3.42950723	1.43423412	3.42950723

The result indicates that the there was a close, significant linear correlation between chocolate consumption per capita and the number of Nobel laureates per 10 million persons in a total of 22 countries. Switzerland had the highest chocolate consumption and had the second most Nobel laureates (the first is Sweden, which is only 0.552 more than Switzerland in average). And we notice that although Sweden had the most Nobel laureates, it merely had an amount of 4.4kg chocolate consumption, which is less than half of the chocolate consumption of Switzerland. We agreed with the opinion that was came up with from the article, which is that the Nobel Committee in Stockholm has some inherent patriotic bias when assessing the candidates for these awards, or perhaps that Swedes are particularly sensitive to chocolate so that even minuscule amounts greatly enhance their cognition.

Besides the opinion from the article, we think that there might be other factors that can influence the result such as alcohol consumption and coffee consumption, so we did an analysis in terms of these two factors as well.

2. Alcohol consumption and Nobel laureates.

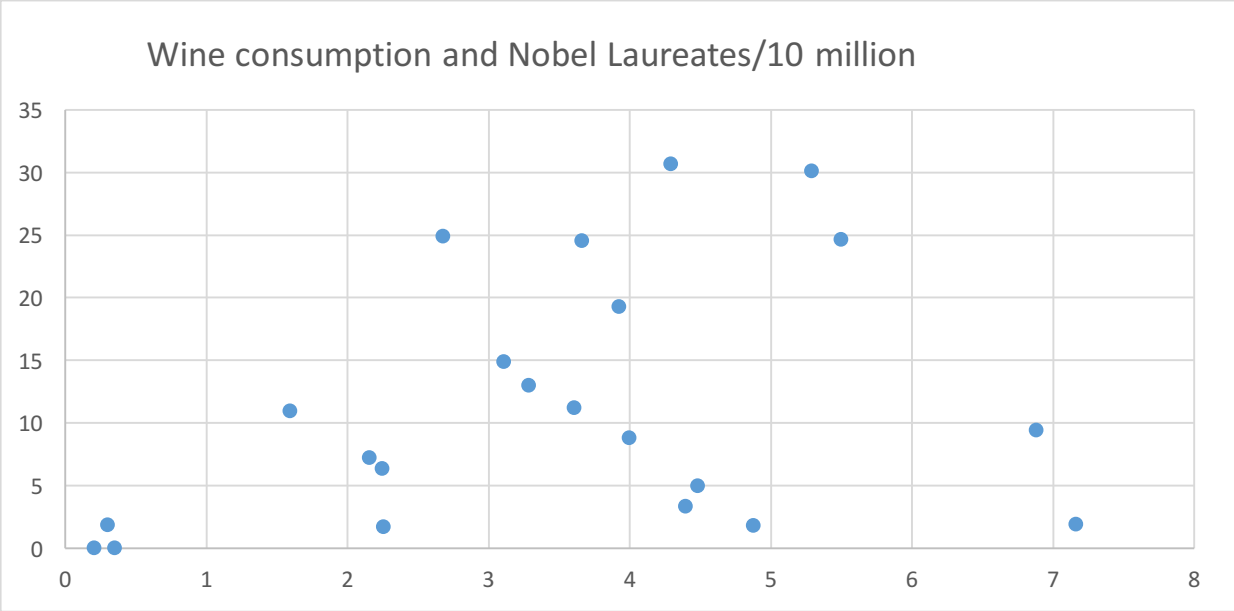
Data on pure alcohol consumption among adults (age 15+) in liters per capita per year (2010) in 22 countries was obtained from Wikipedia (https://en.wikipedia.org/wiki/List_of_countries_by_alcohol_consumption_per_capita).



SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.10942144							
R Square	0.01197305							
Adjusted R Sq	-0.0374283							
Standard Erro	10.2203047							
Observations	22							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	25.3159209	25.3159209	0.24236284	0.62786538			
Residual	20	2089.09257	104.454629					
Total	21	2114.40849						
	Coefficients	Standard Erro	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	5.49980847	12.3193469	0.44643669	0.66007806	-20.197899	31.1975157	-20.197899	31.1975157
alcohol	0.58288155	1.18398797	0.49230361	0.62786538	-1.8868741	3.05263718	-1.8868741	3.05263718

Because dietary flavonoids are abundant in red wine, according to the research about the chocolate consumption, we assume that higher alcohol consumption will be corresponding to more Nobel laureates, but the scatter plot and the regression data analytic result above shows a different result. There was only a weak linear correlation ($r = 0.1094$, $p = 0.6278$) between alcohol consumption and Nobel laureates. We notice that the data also has more detailed information, there are percentage

about beer, wine, and spirits consumption separately. What we really interested in should be the wine consumption. So we did another analysis base only on the wine consumption.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.31415016
R Square	0.09869032
Adjusted R Sq	0.05362484
Standard Erro	1.86577536
Observations	22

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	7.62340909	7.62340909	2.18993144	0.15449603
Residual	20	69.6223538	3.48111769		
Total	21	77.2457629			

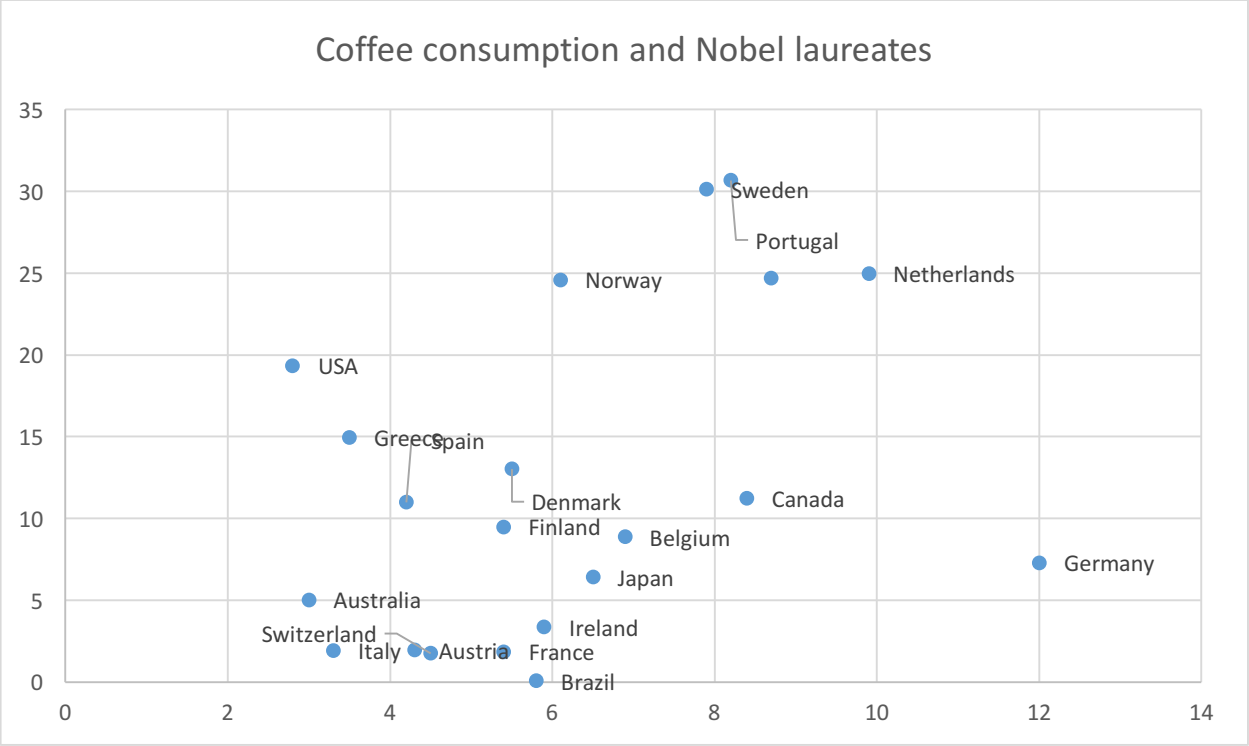
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2.77352682	0.61220558	4.53038473	0.00020382	1.49648835	4.05056528	1.49648835	4.05056528
Nobel Laureat	0.06004546	0.0405756	1.4798417	0.15449603	-0.0245938	0.14468467	-0.0245938	0.14468467

According to the regression analysis, we find that here r = 0.3141 and p = 0.1545, which also indicates that wine consumption does not have a necessary relationship with the number of Nobel laureates. Considering this, one reason might be that there are hundreds of wine types in the world, some types of wine may rich in dietary flavonoids while others not, and the type of wines consumption may vary in different countries.

To do some further research, we find the data of coffee consumption as well and want to find out if there is a linear correlation between coffee consumption and the number of Nobel laureates.

3. Coffee consumption and Nobel laureates.

Data on the coffee consumption in kg per capita per year in 21 countries was obtained from Wikipedia (https://www.google.com/fusiontables/DataSource?docid=1C-fn6nSe21acP0xJIO1T1x0wohqfMYCQyJjbqdk#rows:id=1). We eliminate China here because we did not find the data of China.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.3927326
R Square	0.15423889
Adjusted R Sq	0.10972515
Standard Erro	9.38378657
Observations	21

ANOVA

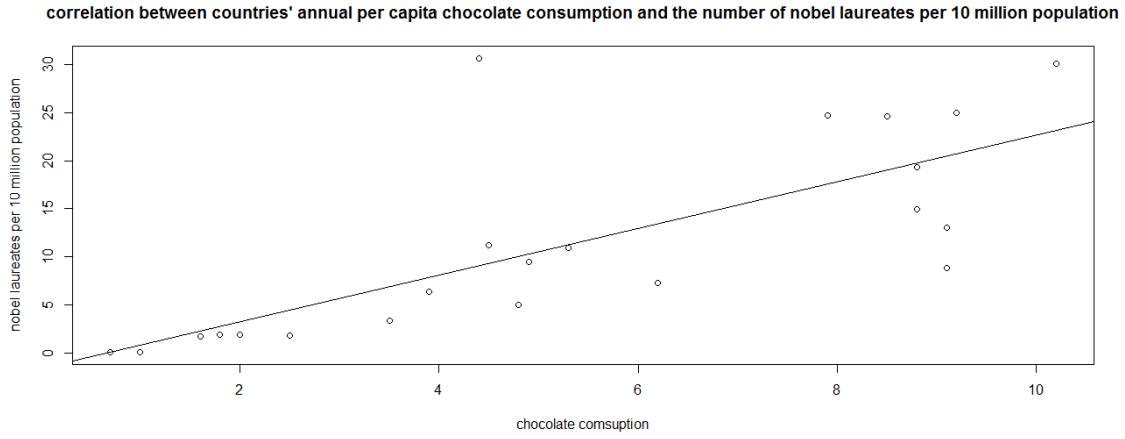
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	305.109716	305.109716	3.46497252	0.07822566
Residual	19	1673.05356	88.0554504		
Total	20	1978.16327			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2.09979003	5.70520947	0.36804784	0.71690596	-9.8413506	14.0409307	-9.8413506	14.0409307
coffee	1.62370054	0.87228025	1.86144366	0.07822566	-0.202003	3.44940408	-0.202003	3.44940408

From the regression analysis we can see that r = 0.3927 and p = 0.0782, we cannot say that there exists a significant linear correlation between coffee consumption and the number of Nobel laureates. But why does chocolate consumption and the number of Nobel laureates have a such close linear correlation while others not? According to that article, we should know that a correlation between X and Y does not prove causation but indicates that either X influences Y, Y influences X, or X and Y are influenced by a common underlying mechanism. We think it is true that it should remain to be determined that whether chocolate consumption enhances cognitive function, and that the cognitive function is a common underlying mechanism between chocolate consumption and the number of Nobel laureates.

4. R plot

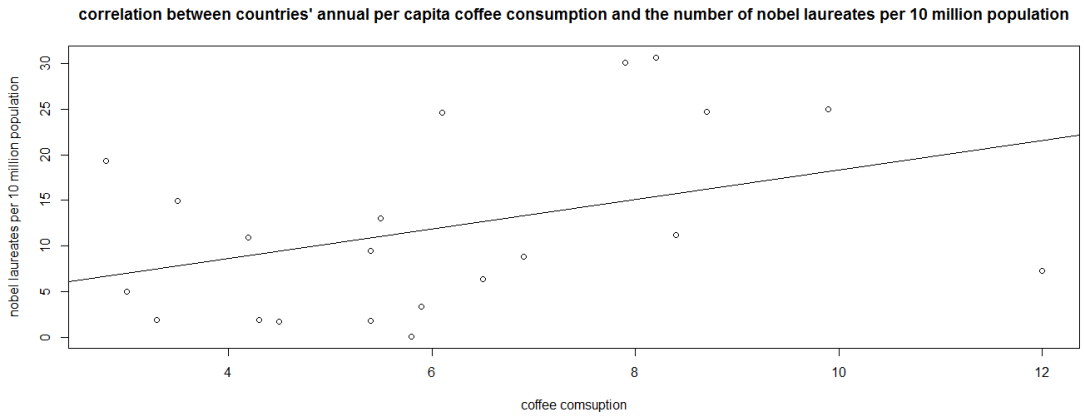
We are trying to analyze these factors using R as well.



ERE:

```
(Intercept)    chocolate
  -1.652002     2.431871
```

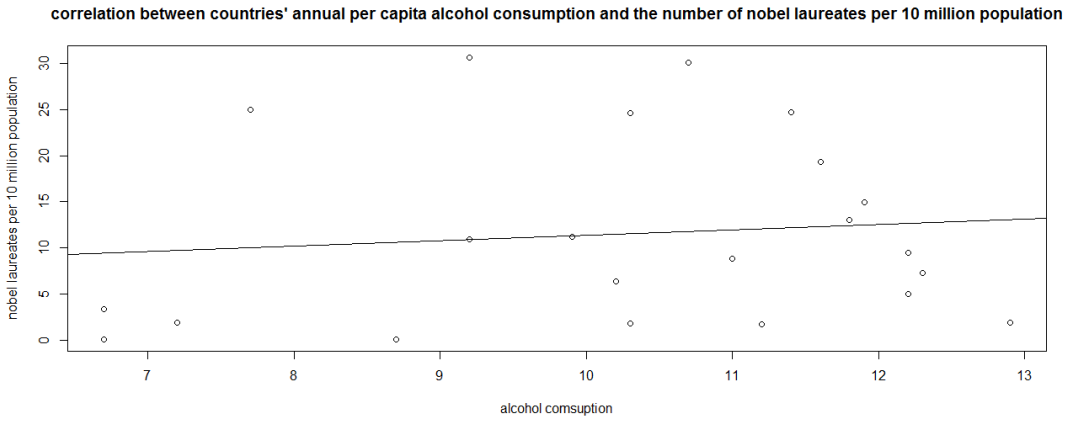
Y=-1.652002+2.431871*x



ERE:

```
(Intercept)       coffee
   2.099790     1.623701
```

Y=2.099790+1.623701*x



EME:

```
(Intercept)     alcohol
  5.4998085   0.5828816
```

Y=5.4998085+0.5828816*x

R script:

1.

```
plot(data=nobel,Nobel.Laureates.10.million~chocolate,main="correlation between countries' annual per capita chocolate consumption and the number of nobel laureates per 10 million population",xlab="chocolate comsuption",ylab="nobel laureates per 10 million population")
lm.reg<-lm(data=nobel,Nobel.Laureates.10.million~chocolate)
abline(lm.reg)
coefficients(lm.reg)
```

2.

```
plot(data=nobel,Nobel.Laureates.10.million~coffee,main="correlation between countries' annual per capita coffee consumption and the number of nobel laureates per 10 million population",xlab="coffee comsuption",ylab="nobel laureates per 10 million population")
lm.reg2<-lm(data=nobel,Nobel.Laureates.10.million~coffee)
abline(lm.reg2)
coefficients(lm.reg2)
```

3.

```
plot(data=nobel,Nobel.Laureates.10.million~alcohol,main="correlation between countries' annual per capita alcohol consumption and the number of nobel laureates per 10 million population",xlab="alcohol comsuption",ylab="nobel laureates per 10 million population")
lm.reg3<-lm(data=nobel,Nobel.Laureates.10.million~alcohol)
abline(lm.reg3)
coefficients(lm.reg3)
```