**Assignment-6:**

Team member: Lv Kaiyang, Ma Ying, Tuo Weiyi

First, we collected Nobel prizes data from Wikipedia (http://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita) and chocolate consumption data from International Cocoa Organization (http://www.icco.org/about-us/international-cocoa-agreements/cat_view/30-related-documents/45-statistics-other-statistics.html). Then, we used Rstudio to analyze the correlation between these two variables.

Since we have Nobel prizes statistic from 2015 and chocolate consumption data from 2002 to 2010, we made an assumption that chocolate consumption did not change significantly. Therefore, the chocolate consumption from 2015 should be the mean of that from 2002 to 2010.

Next, we merged these two tables as one and standardized our datasets. Finally, we plot the correlation between these two variables using R.

ggplot(output, aes(x = ChocolateConsumption, y = Laureates.10.million, color=Country)) +geom_point() +geom_text(aes(x = ChocolateConsumption, y = Laureates.10.million, label = Country))
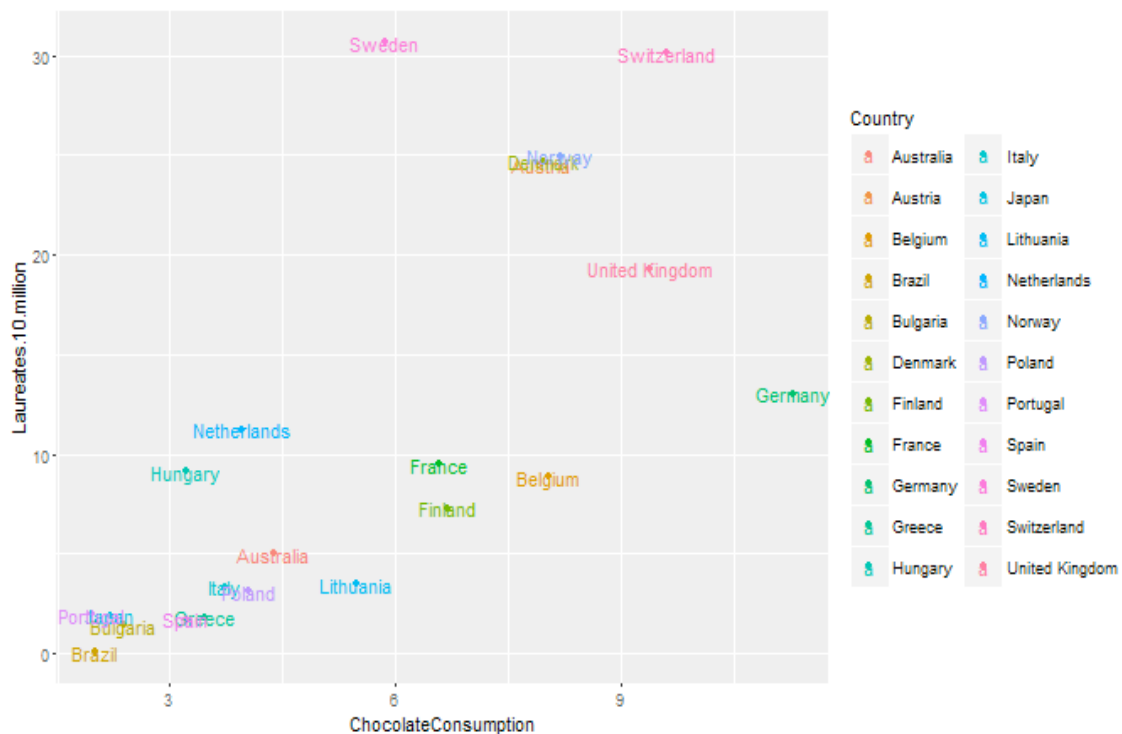
We calculated the correlation by using Pearson correlation coefficient.
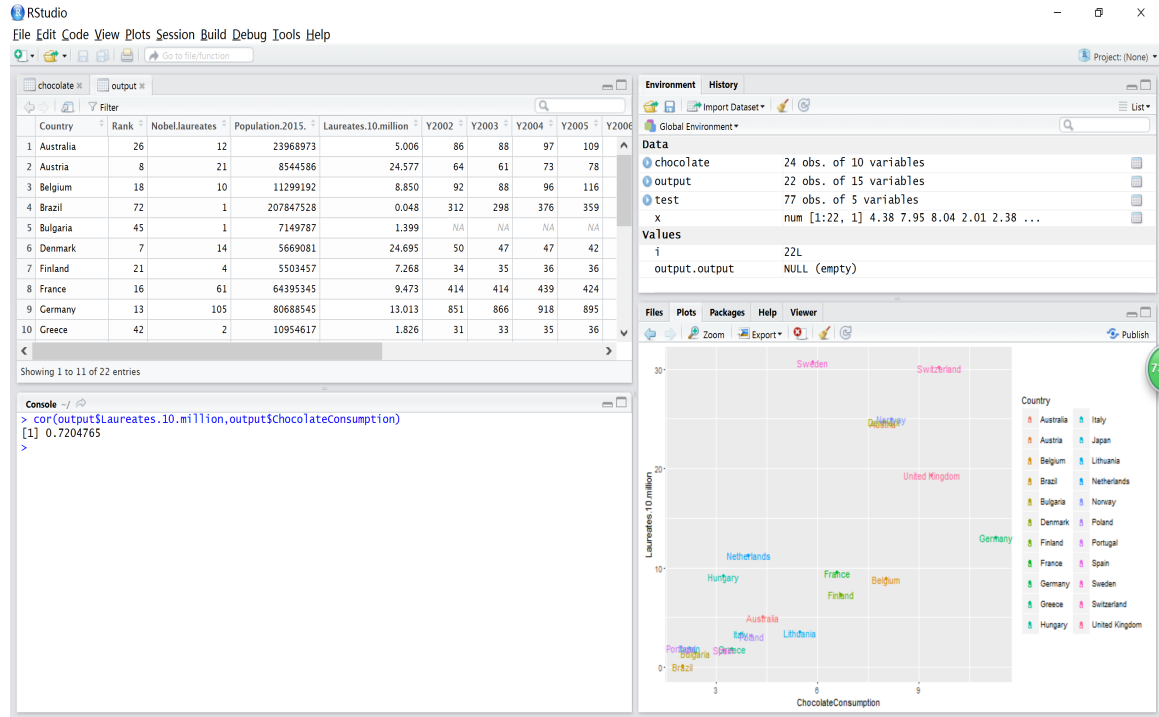
We got the correlation value which is equal to 0.7204765.

Then we made a conclusion that the countries whose people consume a lot of chocolate have many Nobel prizes winners. The more chocolate the people consume, the more chance to win the Nobel prizes. The correlation between them is slightly strong to support our conclusion.

Besides the Nobel prizes data from Wikipedia, we collected tea/coffee/alcohol/ consumption data
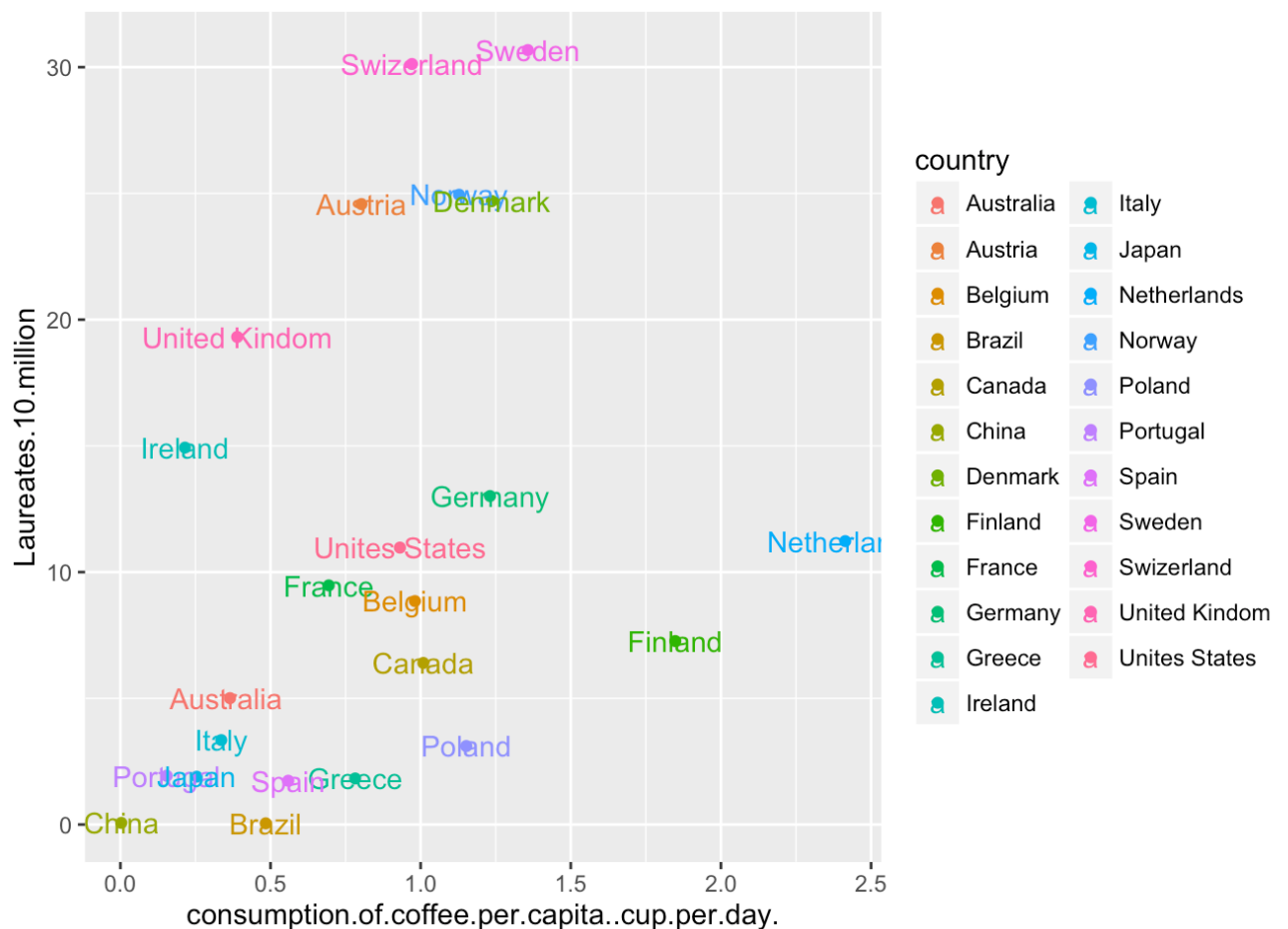
```
library(datasets)
file=read.csv("/Users/tuzigudong/Desktop/assignment.csv")
```

coffee consumption—Nobel prize(per 10 million people):

```
library(ggplot2)
#####correlation between Countries'Annual per Capita coffee consumption and the Number of Nobel Laureates per 10 million population
ggplot(file,aes(x=consumption.of.coffee.per.capita..cup.per.day.,y=Laureates.10.million,color=country))+geom_point()+geom_text(aes(x=consumption.of.coffee.per.capita..cup.per.day.,y=Laureates.10.million,label=country))
```
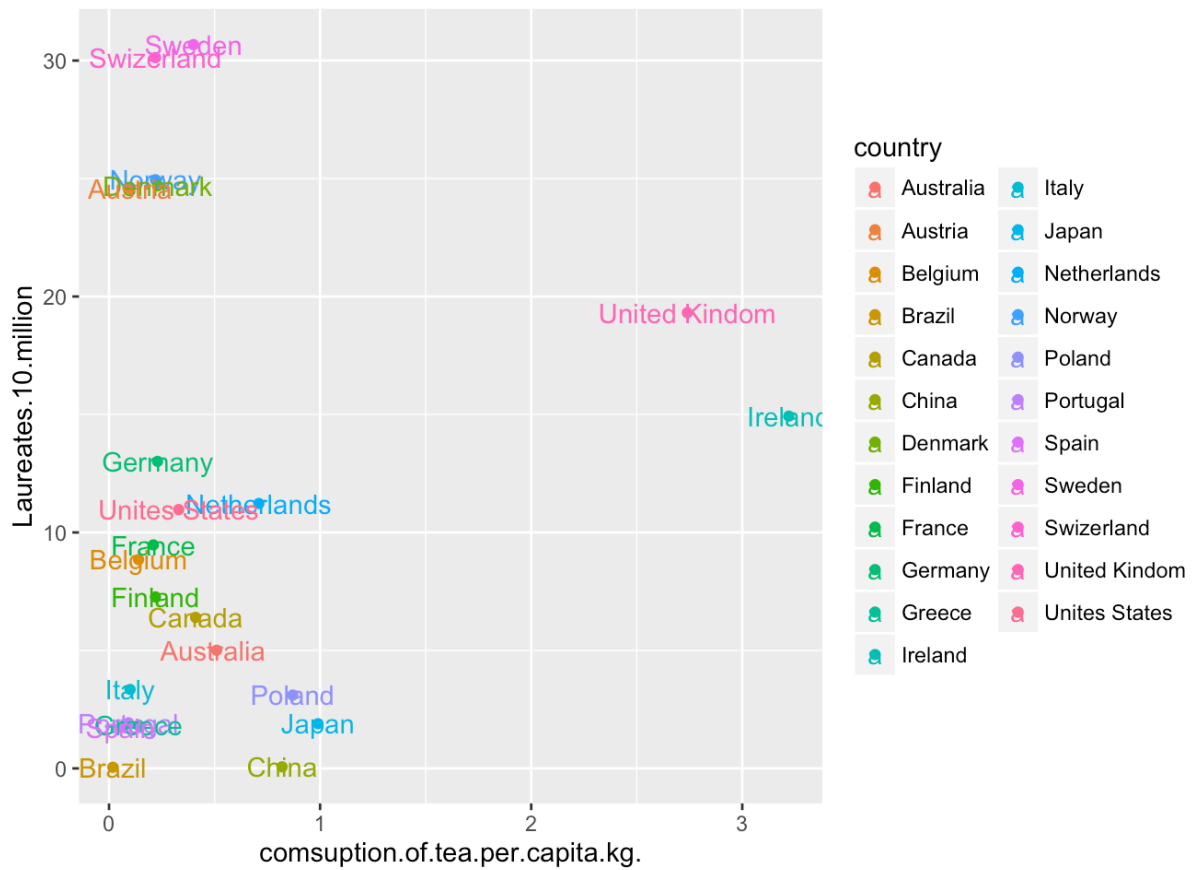
tea consumption—Nobel prize(per 10 million people):

```
ggplot(file,aes(x=comsuption.of.tea.per.capita.kg.,y=Laureates.10.million,color=country))+geom_point()+geom_text(aes(x=comsuption.of.tea.per.capita.kg.,y=Laureates.10.million,label=country))
```

```
ggplot(file,aes(x=alcohol.consumption.per.year,y=Laureates.10.million,c
olor=Country))+geom_text(aes(x=  alcohol.consumption.per.year,y=Laureate
s.10.million,label=Country))
```
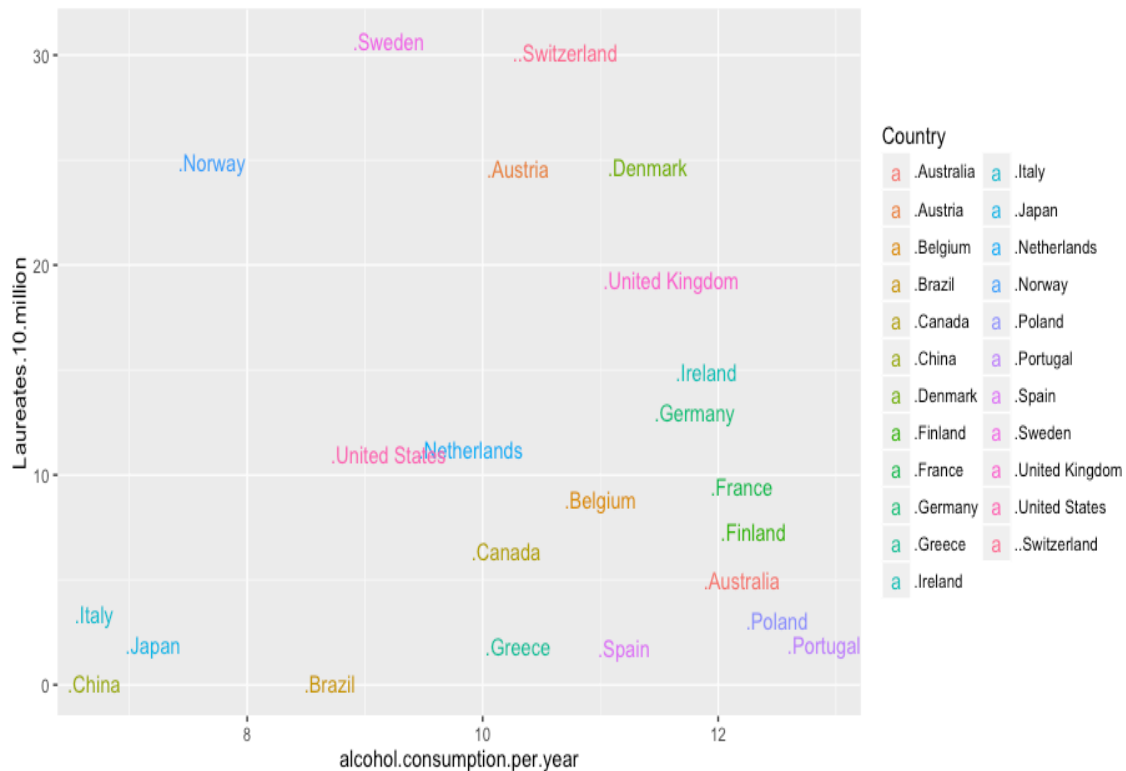

calculate the linear correlation

cor(file$comsuption.of.tea.per.capita.kg.,file$Laureates.10.million)

[1] 0.1096009

cor(file$consumption.of.coffee.per.capita..cup.per.day.,file$Laureates.10.million)

[1] 0.3518936

cor(file$alcohol.consumption.per.year,file$Laureates.10.million)

[1] 0.06095194

Final conclusion:

1. As we can see we choose four items as our influential factors, only chocolate has big influence on Nobel laureates per 10 million people. According to the reading materials this is because coca could decrease cognitive function degeneration.

2. For coffee, the correlation coefficient is 0.35. In certain degree, we can say that coffee helps to get Nobel Prize. May it is because caffeine can keep people excited, so researchers have more energy to focus on their research.

3. At the beginning, we think alcohol has bad effect on scientific research, because alcohol can make people become delirious. However according to the scatter plots, we found no relationship between the two stuffs. To make conclusion we need more data and research.