ASSIGNMENT 6

Probabilistic Approaches

Chunyao Wang

Wenxing Li

Jie Rong

1.Introduction

In this assignment, we replicated the analysis of NEJMon1211064.pdf. Firstly, we find the data that the author based his paper on and calculate the correlation between the two quantities. Then, we collection other data and using same method to figure out the relation between the two new quantities.

In the paper, authors want to know whether there would be a correlation between a country's level of chocolate consumption and its population's cognitive function. They used the number of people who won Nobel Prize to reflect overall cognitive function of a given country.

The aim of this assignment is not only to figure out whether there is a relation between the number of people who won Nobel Prize and chocolate consumption, but also to find some other subjects which may have relationship with the number of people who won Nobel Prize.

Programming language: R
Software: RStudio
Data source:

1. Laureates pre 10 million people, 2015, List of countries by Nobel laureates per capita https://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita
2. Amount of chocolate consumed per person by country(kg), 2011, Statista http://www.statista.com/statistics/262981/amount-of-chocolate-consumed-per-person-by-country/
3. Milk consumption per capita(kg), 2007, List of countries by milk consumption per capita (2007) https://en.wikipedia.org/wiki/List_of_countries_by_milk_consumption_per_capita
4. Pure alcohol consumption among adults in liters per capita per year, 2010, List of countries by alcohol consumption per capita https://en.wikipedia.org/wiki/List_of_countries_by_alcohol_consumption_per_capita
5. Number of cigarettes per adult per year, List of countries by cigarette consumption per capita https://en.wikipedia.org/wiki/List_of_countries_by_cigarette_consumption_per_capita

2.Chocolate Consumption

```
> chocolate <- read.csv("~/Desktop/chocolate.csv")
> library(ggplot2)
```

ggplot2 scatter plot:

```
> p3<-ggplot(data=Nobel_ch,mapping=aes(x=Chocolate,y=Nobel,label=(Nobel_ch$Country)))
> p3+geom_text_repel(alpha=0.7,size=4,segment.size =0)+geom_point(aes(colour=Country),alpha=0.6,size=6,shape=19,position='jitter')+labs(x='Chocolate Consumption(kg/yr/capita)',y='Nobel Laureates per 10 million Population')+geom_smooth(method=lm,se=FALSE,size=0.7,color="red",alpha=0.5)
```

correlation coefficient:

```
> nobel=Nobel_ch$Nobel
> chocolate=Nobel_ch$Chocolate
> cor(nobel,chocolate)
[1] 0.7557891
>
```

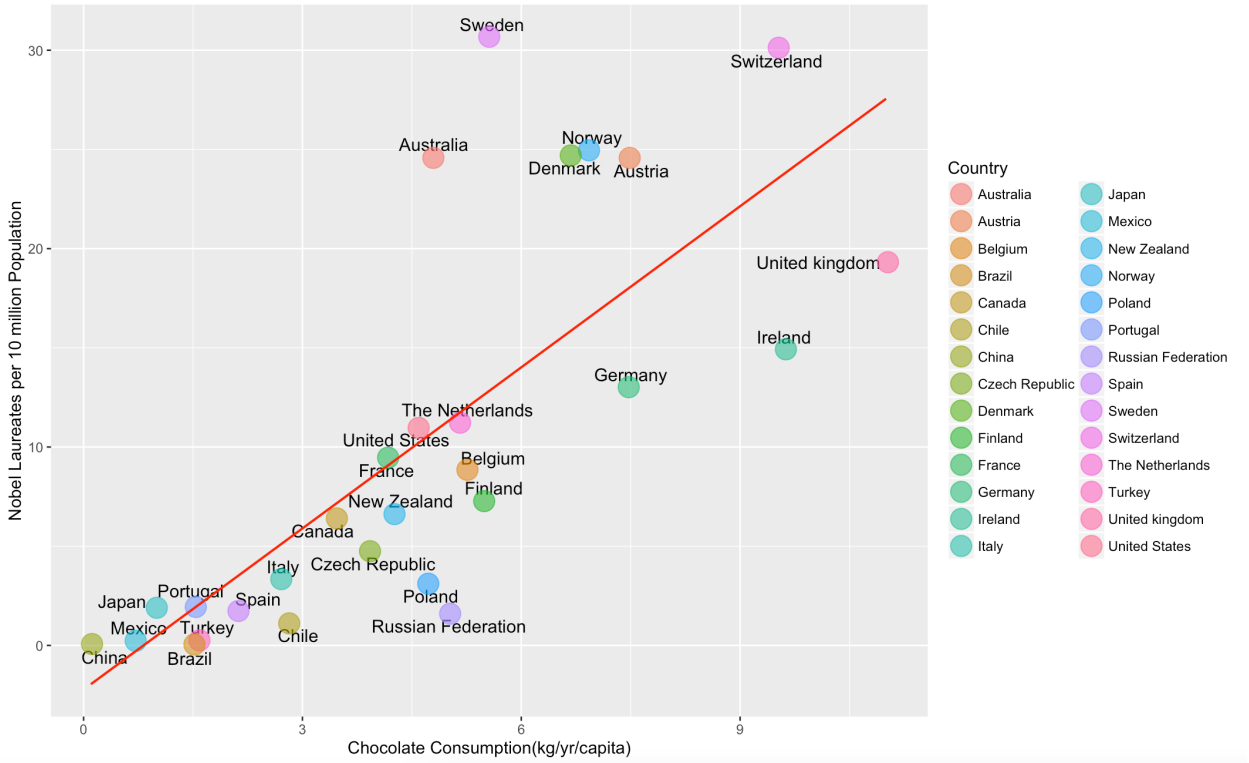

Figure 1

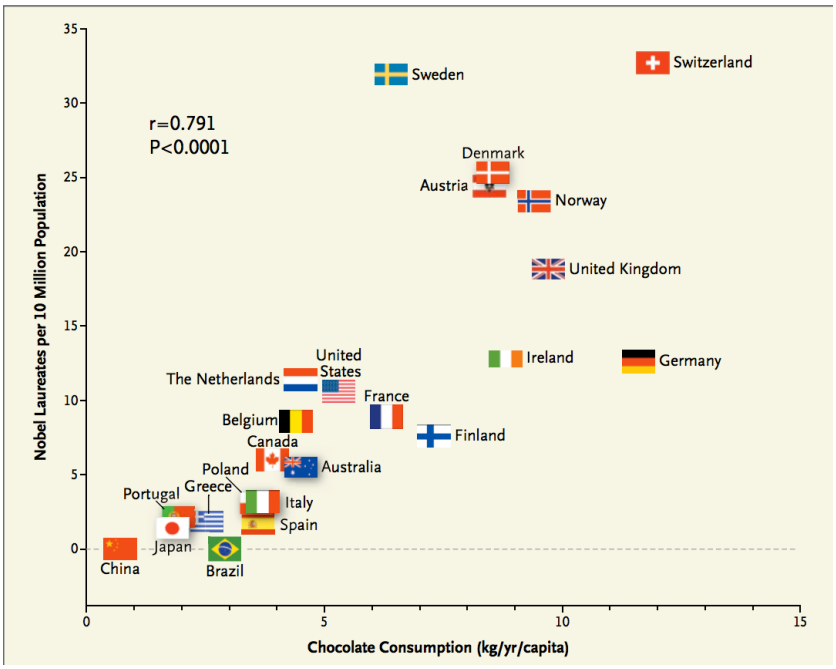

Figure 2

The result of correlation coefficient(Figure1) is slightly different from authors' result (0.791). This maybe because the year of chocolate consumption data we used was different with authors'. But the overall trend is very similar. The positons of all countries are relatively in the same place as in paper's plot (Figure 2).

3. Milk Consumption

```
> Nobel_milk <- read.csv("~/Desktop/Nobel_milk.csv")
>   View(Nobel_milk)
> library(ggplot2)
> p4<-ggplot(data=Nobel_milk,mapping=aes(x=Milk,y=Nobel,label=(Nobel_milk$Country)))
> p4+geom_text_repel(alpha=0.7,size=4,segment.size =0)+geom_point(aes(colour=Country),alpha=
0.6,size=6,shape=19,position='jitter')+labs(x='Milk Consumption(kg/yr/capita)',y='Nobel Laur
eates per 10 million Population')+geom_smooth(method=lm,se=FALSE,size=0.7,color="red",alpha=
0.5)
>
> milk=Nobel_milk$Milk
> nobel1=Nobel_milk$Nobel
> cor(nobel1,milk)
[1] 0.585751
>
```

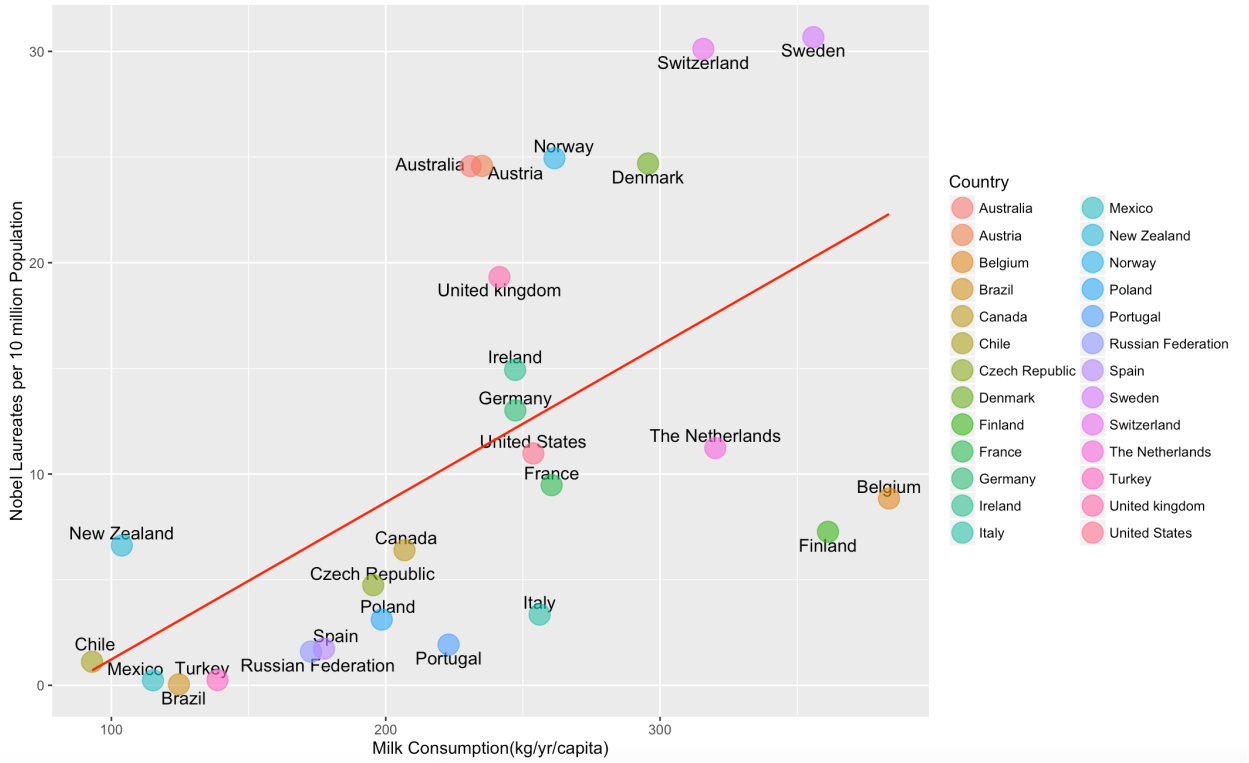

Figure 3

The reason why we choose milk consumption per capita is that milk can supply animal protein and Ca^{+} to human being. And these are very essential for brain and body development. Thus, our assumption is that the number of Nobel laureates has relation with milk consumption
As a result, we can see there is a positive correlation between milk consumption and Nobel Laureates. Although the value of correlation is not that high as chocolate, we can see a clearly trend in the plot (Figure 3).

4. Alcohol Consumption

```
> Nobel_alcohol <- read.csv("~/Desktop/Nobel_alcohol.csv")
>   View(Nobel_alcohol)
> p5<-ggplot(data=Nobel_alcohol,mapping=aes(x=Alcohol,y=Nobel,label=(Nobel_alcohol$Country))
)
> p5+geom_text_repel(alpha=0.7,size=4,segment.size =0)+geom_point(aes(colour=Country),alpha=
0.6,size=6,shape=19,position='jitter')+labs(x='Alcohol Consumption(liters/yr/capita)',y='Nob
el Laureates per 10 million Population')+geom_smooth(method=lm,se=FALSE,size=0.7,color="red"
,alpha=0.5)
>
> alcohol=Nobel_alcohol$Alcohol
> nobel2=Nobel_alcohol$Nobel
> cor(nobel2,alcohol)
[1] 0.1752293
>
```

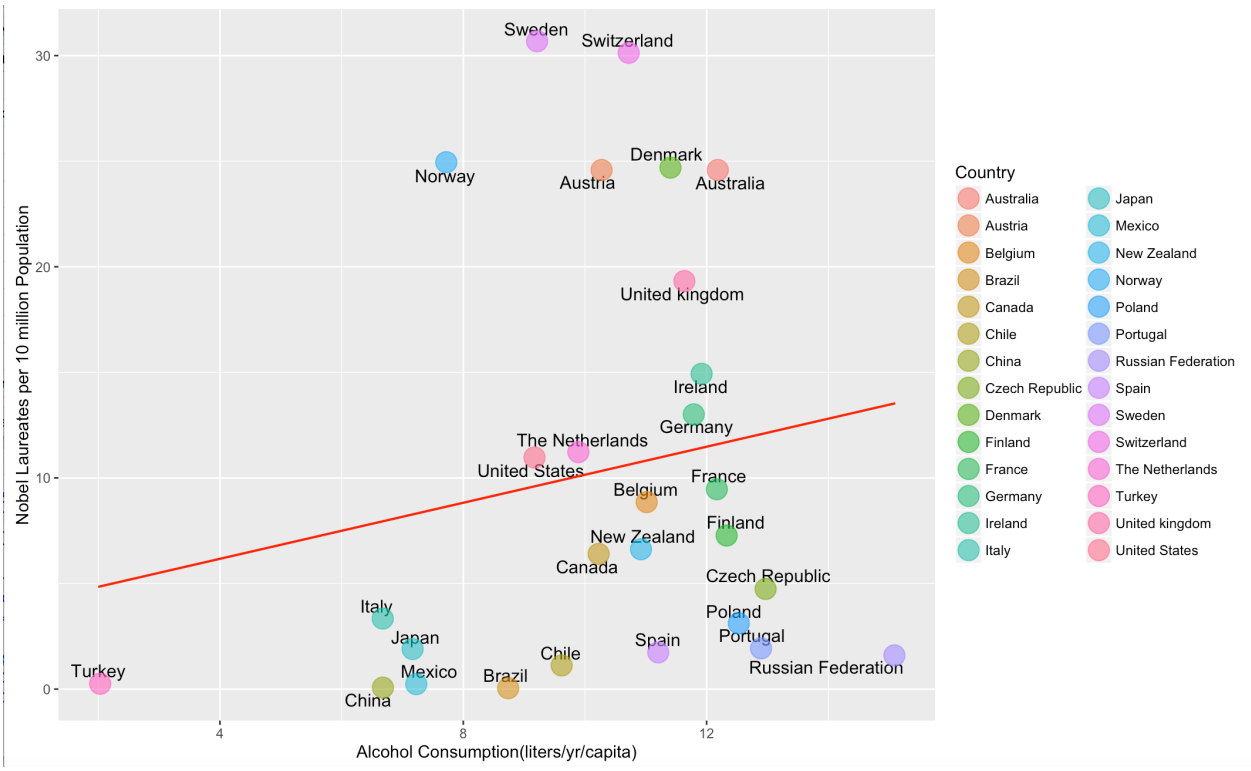

Figure 4

Alcohol have positive and negative effects on people. Moderate drinking might give people better health, such as drinking wine. Or it may give people inspirations after drinking. However, alcohol abusing will cause mental or physical illness, such as dizzy, even liver disease. It is hard to say the overall effects on human cognitive function. Just as the result we got, the correlation coefficient is very small and the scatter plot is not regular.

5.Cigarettes Consumption

```
> Nobel_cig <- read.csv("~/Desktop/Nobel_cig.csv")
>   View(Nobel_cig)
> p6<-ggplot(data=Nobel_cig,mapping=aes(x=Cigarettes,y=Nobel,label=(Nobel_cig$Country)))
> p6+geom_text_repel(alpha=0.7,size=4,segment.size =0)+geom_point(aes(colour=Country),alpha=
0.6,size=6,shape=19,position='jitter')+labs(x='Cigarettes Consumption(number/yr/capita)',y='
Nobel Laureates per 10 million Population')+geom_smooth(method=lm,se=FALSE,size=0.7,color="r
ed",alpha=0.5)
> cigarette=Nobel_cig$Cigarettes
> nobel3=Nobel_cig$Nobel
> cor(nobel3,cigarette)
[1] -0.1599062
>
```

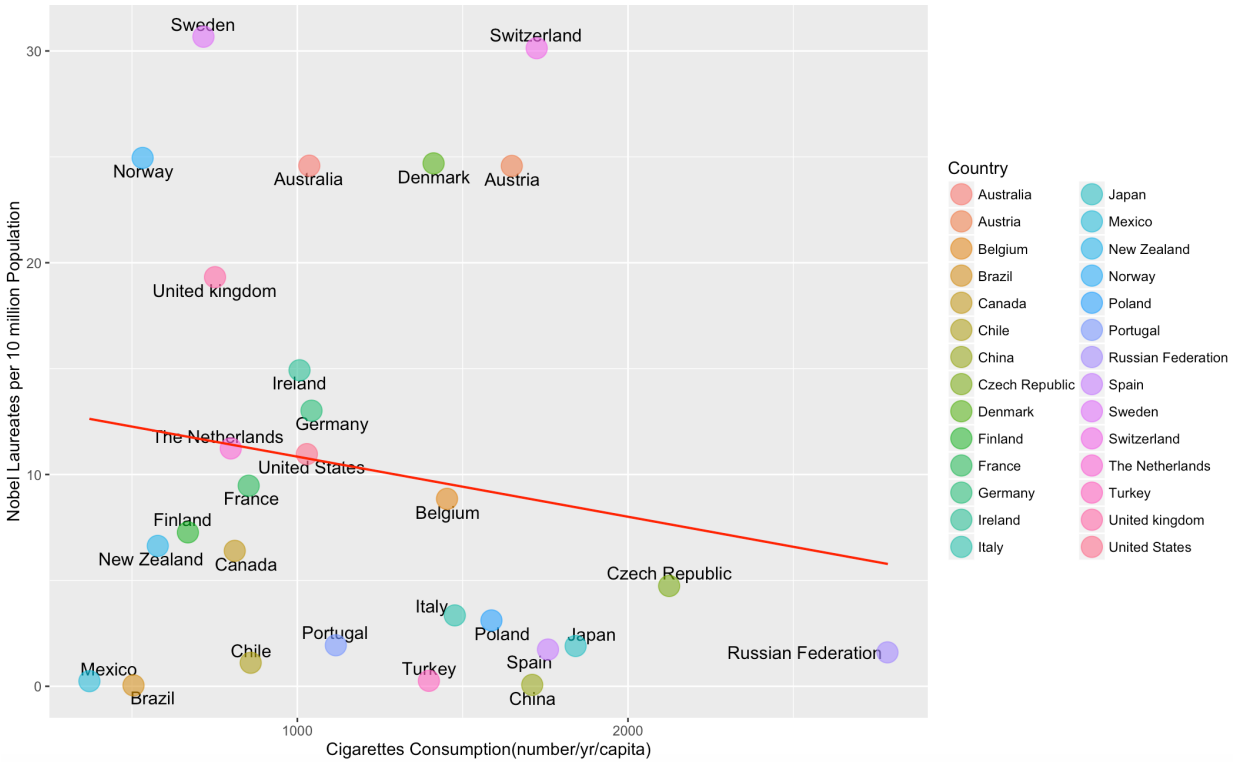

Figure 5

Our assumption of this part is that the correlation coefficient will be negative and the absolute value will be very high. Because we have seen a research which shows smoking will damage the brain. As a result, the correlation coefficient is negative indeed. However, the absolute value doesn't fit. We find that there are some exceptions, like Norway, Australia, Denmark and United Kingdom. The reason might be that there are lots of people smoking, but few people smoking excessively.