

DataAnalytics_Phase3.R

2019-04-07

```
#Read data Table:
dataTable <- read.table(
  "drug_consumption.data",
  sep="," , header=FALSE)

#Viewing the original table:
View(dataTable)

#Naming columns:
colnames(dataTable) <- c("ID", "Age", "Gender", "Education", "Country", "Ethnicity", "NScore",
                        "EScore", "OScore", "AScore", "CScore", "Impulsive", "SS",
                        "Alcohol",
                        "Amphet", "Amyl", "Benzos", "Caff", "Cannabis", "Choc", "Coke",
                        "Crack", "Ecstasy", "Heroin", "Ketamine", "Legalh", "LSD",
                        "Meth",
                        "Mushrooms", "Nicotine", "Semer", "VSA")

#Summaries of important decisive attributes:

#Gives us the summary of Neurotocism.
NScore<- dataTable$NScore
summary(NScore)

##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -3.464360 -0.678250  0.042570  0.000047  0.629670  3.273930

#Gives us the summary of Extraversion.
EScore<- dataTable$EScore
summary(EScore)

##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -3.273930 -0.695090  0.003320 -0.000163  0.637790  3.273930

#Gives us the summary of Openness to experience.
OScore<- dataTable$OScore
summary(OScore)

##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -3.273930 -0.717270 -0.019280 -0.000534  0.723300  2.901610
```

#Gives us the summary of Agreeableness.

```
AScore<- dataTable$AScore
```

```
summary(AScore)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -3.464360 -0.606330 -0.017290 -0.000245  0.760960  3.464360
```

#Gives us the summary of Conscientiousness.

```
CScore<- dataTable$CScore
```

```
summary(CScore)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -3.464360 -0.652530 -0.006650 -0.000386  0.584890  3.464360
```

```
Age<- dataTable$Age
```

```
summary(Age)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -0.95197 -0.95197 -0.07854  0.03461  0.49788  2.59171
```

```
Gender <-dataTable$Gender
```

```
summary(Gender)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -0.4824600 -0.4824600 -0.4824600 -0.0002559  0.4824600  0.4824600
```

```
Education <-dataTable$Education
```

```
summary(Education)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -2.435910 -0.611130 -0.059210 -0.003806  0.454680  1.984370
```

```
Country <-dataTable$Country
```

```
summary(Country)
```

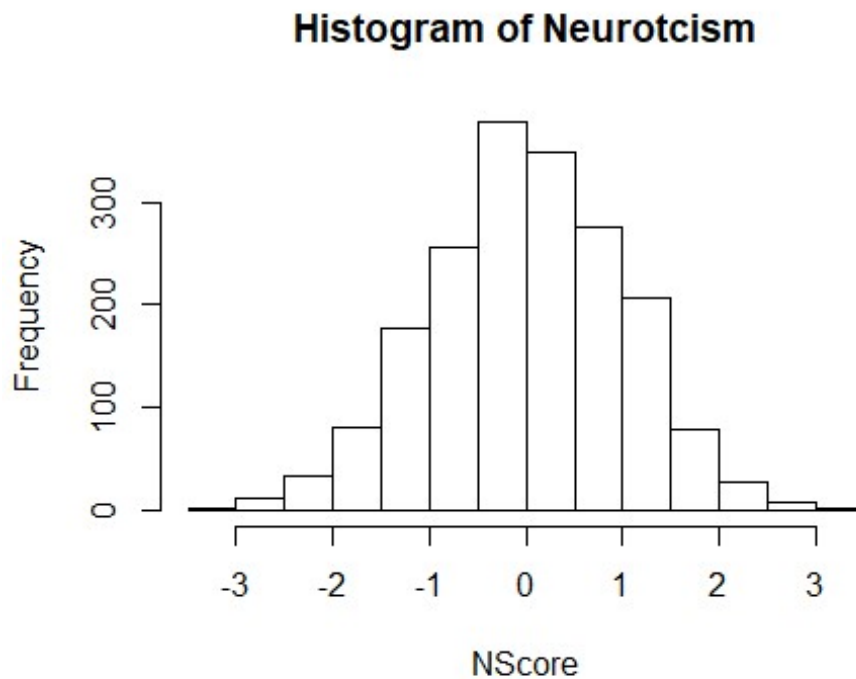
```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -0.5701 -0.5701  0.9608  0.3555  0.9608  0.9608
```

```
Ethnicity <-dataTable$Ethnicity
```

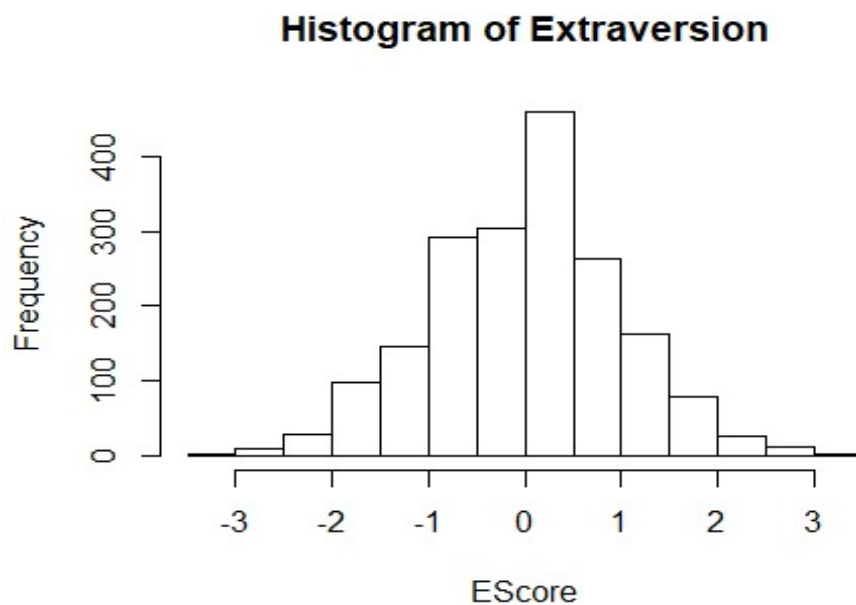
```
summary(Ethnicity)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -1.1070 -0.3169 -0.3169 -0.3096 -0.3169  1.9072
```

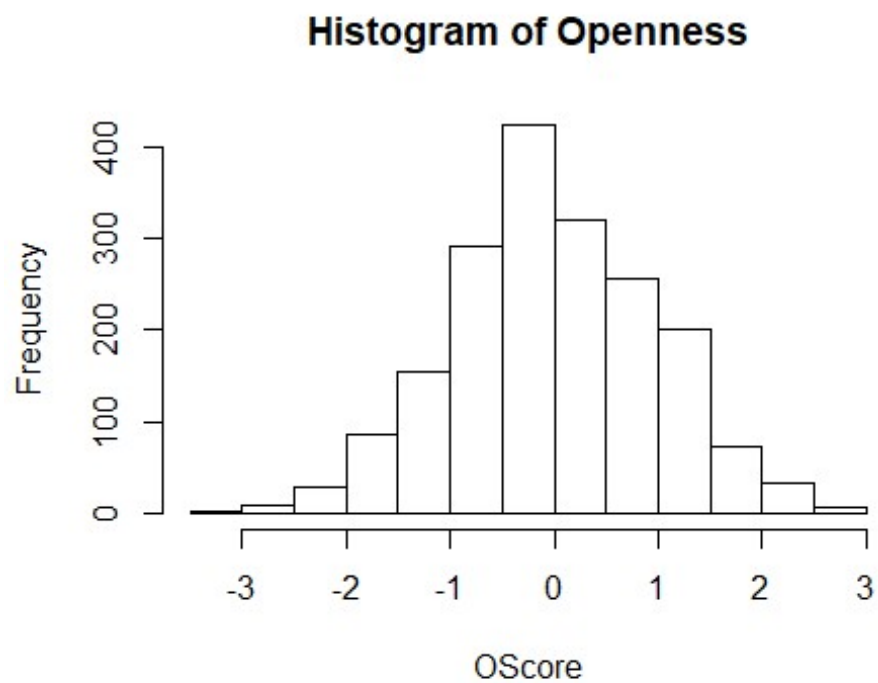
```
#Histograms of some important attributes:  
# From the below histograms we can gain an insight on distribution  
# of each attribute and their frequency.  
hist(NScore, main = 'Histogram of Neurotcism')
```



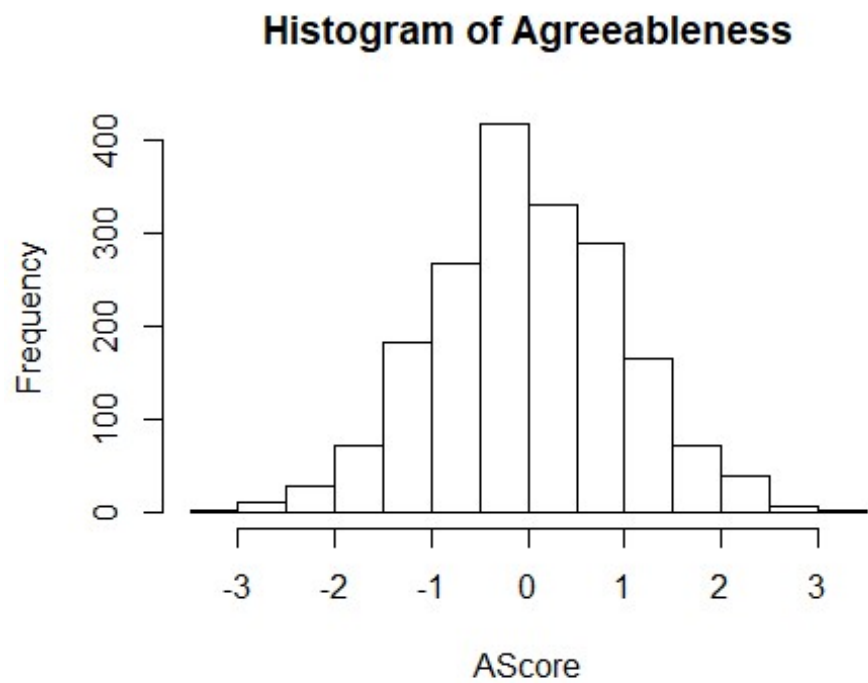
```
hist(EScore, main = 'Histogram of Extraversion')
```



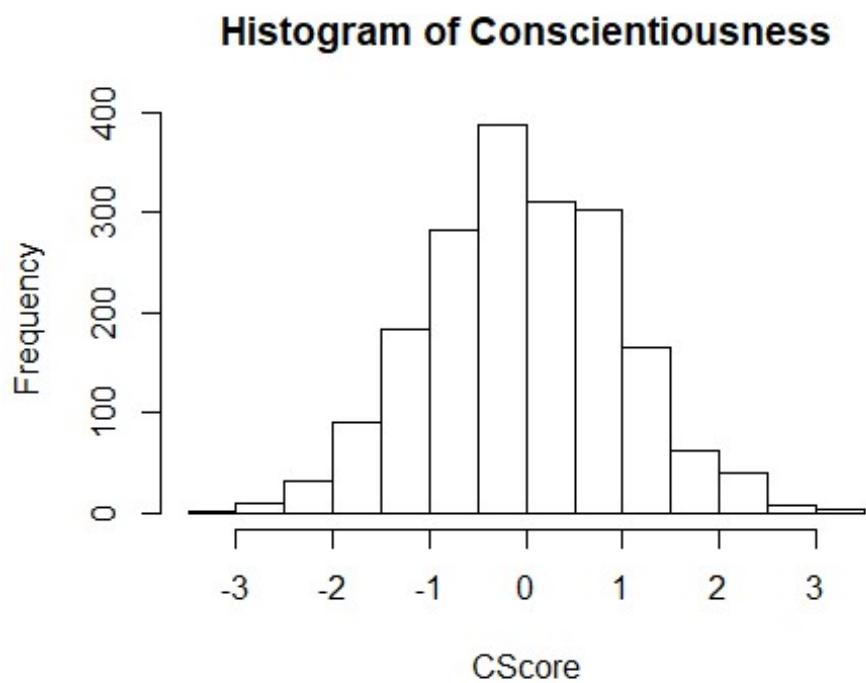
```
hist(OScore, main = 'Histogram of Openness')
```



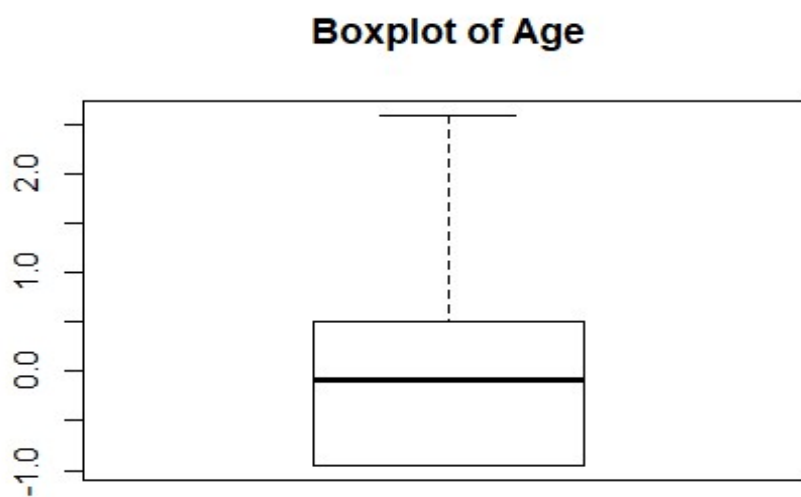
```
hist(AScore, main = 'Histogram of Agreeableness')
```



```
hist(CScore, main = 'Histogram of Conscientiousness')
```

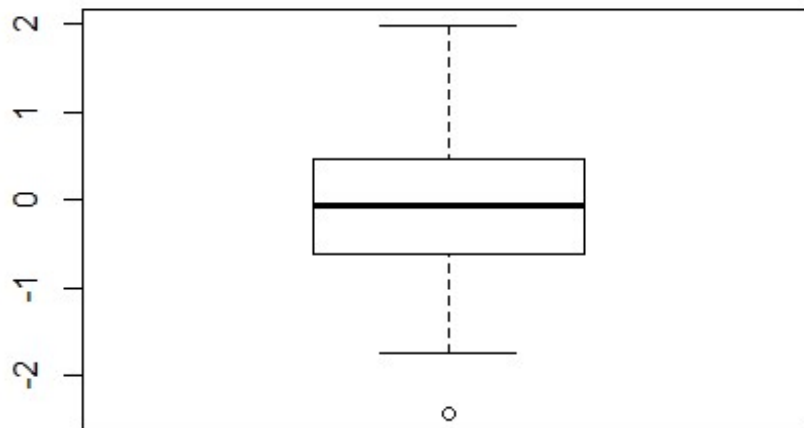


```
#Boxplots:  
boxplot(dataTable$Age, main="Boxplot of Age")
```



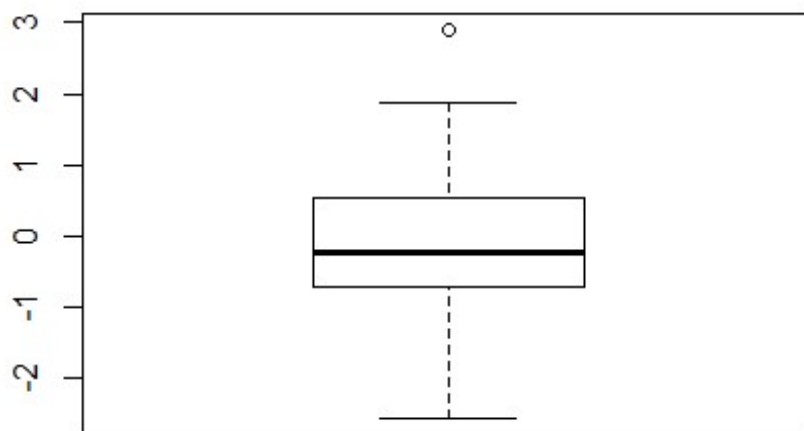
```
boxplot(dataTable$Education, main="Boxplot of Education")
```

Boxplot of Education

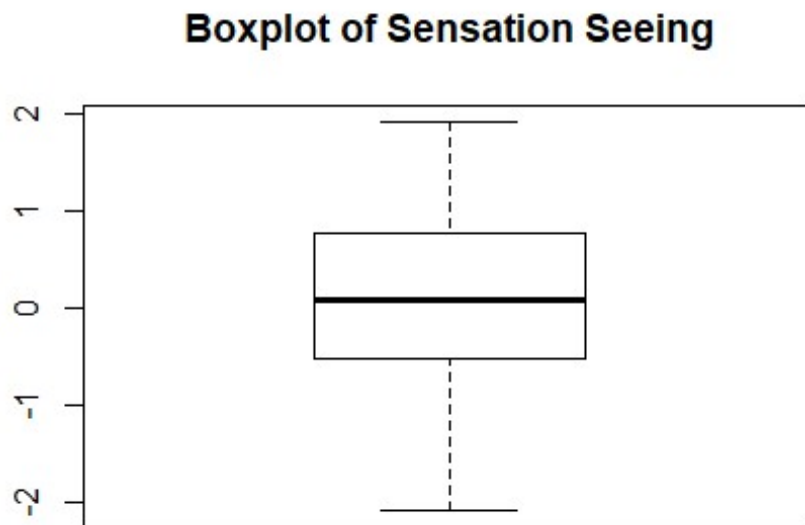


```
boxplot(dataTable$Impulsive,main="Boxplot of Impulsiveness")
```

Boxplot of Impulsiveness

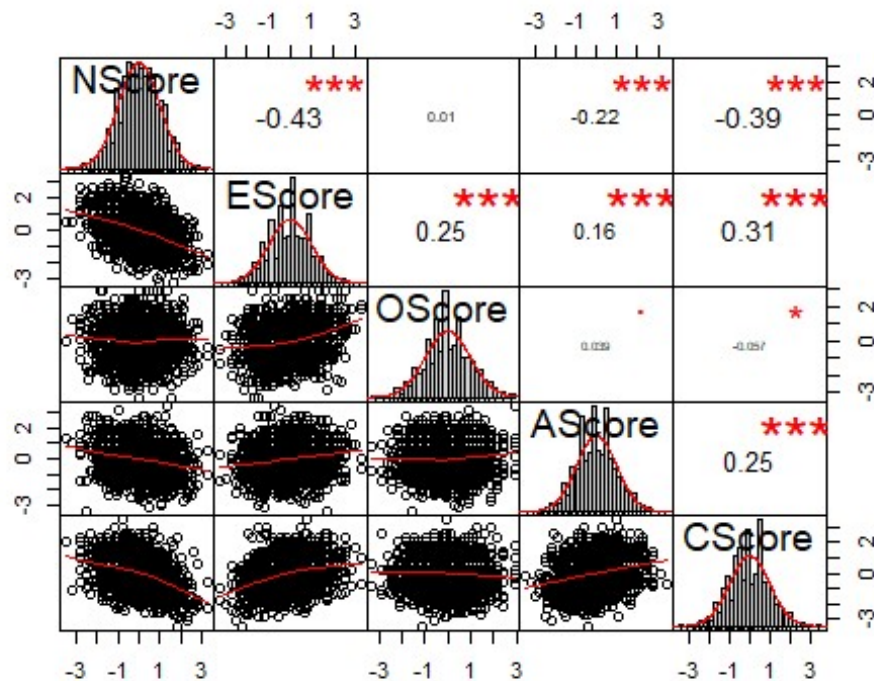


```
boxplot(dataTable$SS,main="Boxplot of Sensation Seeing")
```



```
#To display chart of correlation  
library("PerformanceAnalytics")
```

```
DrugCorrelation <-dataTable[,c(7,8,9,10,11)]  
chart.Correlation(DrugCorrelation,histogram = TRUE,pch=19)
```



#In the above plot:

The distribution of each variable is shown on the diagonal.

On the bottom of the diagonal: the bivariate scatterplots with a fitted line are displayed.

On the top of the diagonal:the value of the correlation plus the significance level

as stars.

#Each significance level is associated to a symbol: p-values(0,0.001,0.01,0.05,0.1,1)

*# <=>symbols("****", "***", "**", ".", " ")*

*#To better understand the data, we decoded the original numerical
#numbers to ranges and values using following R commands:
#Decoding the data*

#Before decoding, our dataset looked like this:

ID	Age	Gender	Education	Country
1	0.49788	0.48246	-0.05921	0.96082
2	-0.07854	-0.48246	1.98437	0.96082
3	0.49788	-0.48246	-0.05921	0.96082
4	-0.95197	0.48246	1.16365	0.96082
5	0.49788	0.48246	1.98437	0.96082
6	2.59171	0.48246	-1.22751	0.24923
7	1.09449	-0.48246	1.16365	-0.57009
8	0.49788	-0.48246	-1.73790	0.96082
9	0.49788	0.48246	-0.05921	0.24923

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.4.3
```

```
dataTable$Age <- as.factor(dataTable$Age)
dataTable$Age <- revalue(dataTable$Age, c("-0.95197"="18-24"))
dataTable$Age <- revalue(dataTable$Age, c("-0.07854"="25-34"))
dataTable$Age <- revalue(dataTable$Age, c("0.49788"="35-44"))
dataTable$Age <- revalue(dataTable$Age, c("1.09449"="45-54"))
dataTable$Age <- revalue(dataTable$Age, c("1.82213"="55-64"))
dataTable$Age <- revalue(dataTable$Age, c("2.59171"="65-74"))
```

```
dataTable$Gender <- as.factor(dataTable$Gender)
dataTable$Gender <- revalue(dataTable$Gender, c("0.48246"="Female"))
dataTable$Gender <- revalue(dataTable$Gender, c("-0.48246"="Male"))
```

```
dataTable$Education <- as.factor(dataTable$Education)
dataTable$Education <- revalue(dataTable$Education, c("-2.43591"="Left school  
before 16 years"))
dataTable$Education <- revalue(dataTable$Education, c("-1.7379"="Left school  
at 16 years"))
dataTable$Education <- revalue(dataTable$Education, c("-1.43719"="Left school  
at 17 years"))
```

```

dataTable$Education <- revalue(dataTable$Education, c("-1.22751"="Left school
at 18 years"))
dataTable$Education <- revalue(dataTable$Education, c("-0.61113"="Some colleg
e or university, no certificate or degree"))
dataTable$Education <- revalue(dataTable$Education, c("-0.05921"="Professiona
l certificate/ diploma"))
dataTable$Education <- revalue(dataTable$Education, c("0.45468"="University d
egree"))
dataTable$Education <- revalue(dataTable$Education, c("1.16365"="Masters degr
ee"))
dataTable$Education <- revalue(dataTable$Education, c("1.98437"="Doctorate de
gree"))

```

```

dataTable$Country <- as.factor(dataTable$Country)
dataTable$Country <- revalue(dataTable$Country, c("-0.09765"="Australia"))
dataTable$Country <- revalue(dataTable$Country, c("0.24923"="Canada"))
dataTable$Country <- revalue(dataTable$Country, c("-0.46841"="New Zealand"))
dataTable$Country <- revalue(dataTable$Country, c("-0.28519"="Other"))
dataTable$Country <- revalue(dataTable$Country, c("0.21128"="Republic of Irel
and"))
dataTable$Country <- revalue(dataTable$Country, c("0.96082"="UK"))
dataTable$Country <- revalue(dataTable$Country, c("-0.57009"="USA"))

```

```

dataTable$Ethnicity <- as.factor(dataTable$Ethnicity)
dataTable$Ethnicity <- revalue(dataTable$Ethnicity, c("-0.50212"="Asian"))
dataTable$Ethnicity <- revalue(dataTable$Ethnicity, c("-1.10702"="Black"))
dataTable$Ethnicity <- revalue(dataTable$Ethnicity, c("1.90725"="Mixed-Black/
Asian"))
dataTable$Ethnicity <- revalue(dataTable$Ethnicity, c("0.126"="Mixed-White/As
ian"))
dataTable$Ethnicity <- revalue(dataTable$Ethnicity, c("-0.22166"="Mixed-White
/Black"))
dataTable$Ethnicity <- revalue(dataTable$Ethnicity, c("0.1144"="Other"))
dataTable$Ethnicity <- revalue(dataTable$Ethnicity, c("-0.31685"="White"))

```

```

dataTable$VSA <- revalue(dataTable$VSA, c("CL0"="Never Used"))
dataTable$VSA <- revalue(dataTable$VSA, c("CL1"="Used over a Decade Ago"))
dataTable$VSA <- revalue(dataTable$VSA, c("CL2"="Used in Last Decade"))
dataTable$VSA <- revalue(dataTable$VSA, c("CL3"="Used in Last Year"))
dataTable$VSA <- revalue(dataTable$VSA, c("CL4"="Used in Last Month"))
dataTable$VSA <- revalue(dataTable$VSA, c("CL5"="Used in Last Week"))
dataTable$VSA <- revalue(dataTable$VSA, c("CL6"="Used in Last Day"))
dataTable$VSA <- as.factor(dataTable$VSA)

```

```

dataTable$Semer <- revalue(dataTable$Semer, c("CL0"="Never Used"))
dataTable$Semer <- as.factor(dataTable$Semer)

```

```

dataTable$Nicotine <- revalue(dataTable$Nicotine, c("CL0"="Never Used"))
dataTable$Nicotine <- revalue(dataTable$Nicotine, c("CL1"="Used over a Decade Ago"))
dataTable$Nicotine <- revalue(dataTable$Nicotine, c("CL2"="Used in Last Decade"))
dataTable$Nicotine <- revalue(dataTable$Nicotine, c("CL3"="Used in Last Year"))
dataTable$Nicotine <- revalue(dataTable$Nicotine, c("CL4"="Used in Last Month"))
dataTable$Nicotine <- revalue(dataTable$Nicotine, c("CL5"="Used in Last Week"))
dataTable$Nicotine <- revalue(dataTable$Nicotine, c("CL6"="Used in Last Day"))
dataTable$Nicotine <- as.factor(dataTable$Nicotine)

dataTable$Mushrooms <- revalue(dataTable$Mushrooms, c("CL0"="Never Used"))
dataTable$Mushrooms <- revalue(dataTable$Mushrooms, c("CL1"="Used over a Decade Ago"))
dataTable$Mushrooms <- revalue(dataTable$Mushrooms, c("CL2"="Used in Last Decade"))
dataTable$Mushrooms <- revalue(dataTable$Mushrooms, c("CL3"="Used in Last Year"))
dataTable$Mushrooms <- revalue(dataTable$Mushrooms, c("CL4"="Used in Last Month"))
dataTable$Mushrooms <- revalue(dataTable$Mushrooms, c("CL5"="Used in Last Week"))
dataTable$Mushrooms <- revalue(dataTable$Mushrooms, c("CL6"="Used in Last Day"))
dataTable$Mushrooms <- as.factor(dataTable$Mushrooms)

dataTable$Meth <- revalue(dataTable$Meth, c("CL0"="Never Used"))
dataTable$Meth <- revalue(dataTable$Meth, c("CL1"="Used over a Decade Ago"))
dataTable$Meth <- revalue(dataTable$Meth, c("CL2"="Used in Last Decade"))
dataTable$Meth <- revalue(dataTable$Meth, c("CL3"="Used in Last Year"))
dataTable$Meth <- revalue(dataTable$Meth, c("CL4"="Used in Last Month"))
dataTable$Meth <- revalue(dataTable$Meth, c("CL5"="Used in Last Week"))
dataTable$Meth <- revalue(dataTable$Meth, c("CL6"="Used in Last Day"))
dataTable$Meth <- as.factor(dataTable$Meth)

dataTable$LSD <- revalue(dataTable$LSD, c("CL0"="Never Used"))
dataTable$LSD <- revalue(dataTable$LSD, c("CL1"="Used over a Decade Ago"))
dataTable$LSD <- revalue(dataTable$LSD, c("CL2"="Used in Last Decade"))
dataTable$LSD <- revalue(dataTable$LSD, c("CL3"="Used in Last Year"))
dataTable$LSD <- revalue(dataTable$LSD, c("CL4"="Used in Last Month"))
dataTable$LSD <- revalue(dataTable$LSD, c("CL5"="Used in Last Week"))
dataTable$LSD <- revalue(dataTable$LSD, c("CL6"="Used in Last Day"))
dataTable$LSD <- as.factor(dataTable$LSD)

```

```

dataTable$Legalh <- revalue(dataTable$Legalh, c("CL0"="Never Used"))
dataTable$Legalh <- revalue(dataTable$Legalh, c("CL1"="Used over a Decade Ago"))
dataTable$Legalh <- revalue(dataTable$Legalh, c("CL2"="Used in Last Decade"))
dataTable$Legalh <- revalue(dataTable$Legalh, c("CL3"="Used in Last Year"))
dataTable$Legalh <- revalue(dataTable$Legalh, c("CL4"="Used in Last Month"))
dataTable$Legalh <- revalue(dataTable$Legalh, c("CL5"="Used in Last Week"))
dataTable$Legalh <- revalue(dataTable$Legalh, c("CL6"="Used in Last Day"))
dataTable$Legalh <- as.factor(dataTable$Legalh)

dataTable$Ketamine <- revalue(dataTable$Ketamine, c("CL0"="Never Used"))
dataTable$Ketamine <- revalue(dataTable$Ketamine, c("CL1"="Used over a Decade Ago"))
dataTable$Ketamine <- revalue(dataTable$Ketamine, c("CL2"="Used in Last Decade"))
dataTable$Ketamine <- revalue(dataTable$Ketamine, c("CL3"="Used in Last Year"))
dataTable$Ketamine <- revalue(dataTable$Ketamine, c("CL4"="Used in Last Month"))
dataTable$Ketamine <- revalue(dataTable$Ketamine, c("CL5"="Used in Last Week"))
dataTable$Ketamine <- revalue(dataTable$Ketamine, c("CL6"="Used in Last Day"))
dataTable$Ketamine <- as.factor(dataTable$Ketamine)

dataTable$Heroin <- revalue(dataTable$Heroin, c("CL0"="Never Used"))
dataTable$Heroin <- revalue(dataTable$Heroin, c("CL1"="Used over a Decade Ago"))
dataTable$Heroin <- revalue(dataTable$Heroin, c("CL2"="Used in Last Decade"))
dataTable$Heroin <- revalue(dataTable$Heroin, c("CL3"="Used in Last Year"))
dataTable$Heroin <- revalue(dataTable$Heroin, c("CL4"="Used in Last Month"))
dataTable$Heroin <- revalue(dataTable$Heroin, c("CL5"="Used in Last Week"))
dataTable$Heroin <- revalue(dataTable$Heroin, c("CL6"="Used in Last Day"))
dataTable$Heroin <- as.factor(dataTable$Heroin)

dataTable$Ecstasy <- revalue(dataTable$Ecstasy, c("CL0"="Never Used"))
dataTable$Ecstasy <- revalue(dataTable$Ecstasy, c("CL1"="Used over a Decade Ago"))
dataTable$Ecstasy <- revalue(dataTable$Ecstasy, c("CL2"="Used in Last Decade"))
dataTable$Ecstasy <- revalue(dataTable$Ecstasy, c("CL3"="Used in Last Year"))
dataTable$Ecstasy <- revalue(dataTable$Ecstasy, c("CL4"="Used in Last Month"))
dataTable$Ecstasy <- revalue(dataTable$Ecstasy, c("CL5"="Used in Last Week"))
dataTable$Ecstasy <- revalue(dataTable$Ecstasy, c("CL6"="Used in Last Day"))
dataTable$Ecstasy <- as.factor(dataTable$Ecstasy)

dataTable$Crack <- revalue(dataTable$Crack, c("CL0"="Never Used"))
dataTable$Crack <- revalue(dataTable$Crack, c("CL1"="Used over a Decade Ago"))

```

```

)
dataTable$Crack <- revalue(dataTable$Crack, c("CL2"="Used in Last Decade"))
dataTable$Crack <- revalue(dataTable$Crack, c("CL3"="Used in Last Year"))
dataTable$Crack <- revalue(dataTable$Crack, c("CL4"="Used in Last Month"))
dataTable$Crack <- revalue(dataTable$Crack, c("CL5"="Used in Last Week"))
dataTable$Crack <- revalue(dataTable$Crack, c("CL6"="Used in Last Day"))
dataTable$Crack <- as.factor(dataTable$Crack)

dataTable$Coke <- revalue(dataTable$Coke, c("CL0"="Never Used"))
dataTable$Coke <- revalue(dataTable$Coke, c("CL1"="Used over a Decade Ago"))
dataTable$Coke <- revalue(dataTable$Coke, c("CL2"="Used in Last Decade"))
dataTable$Coke <- revalue(dataTable$Coke, c("CL3"="Used in Last Year"))
dataTable$Coke <- revalue(dataTable$Coke, c("CL4"="Used in Last Month"))
dataTable$Coke <- revalue(dataTable$Coke, c("CL5"="Used in Last Week"))
dataTable$Coke <- revalue(dataTable$Coke, c("CL6"="Used in Last Day"))
dataTable$Coke <- as.factor(dataTable$Coke)

dataTable$Choc <- revalue(dataTable$Choc, c("CL0"="Never Used"))
dataTable$Choc <- revalue(dataTable$Choc, c("CL1"="Used over a Decade Ago"))
dataTable$Choc <- revalue(dataTable$Choc, c("CL2"="Used in Last Decade"))
dataTable$Choc <- revalue(dataTable$Choc, c("CL3"="Used in Last Year"))
dataTable$Choc <- revalue(dataTable$Choc, c("CL4"="Used in Last Month"))
dataTable$Choc <- revalue(dataTable$Choc, c("CL5"="Used in Last Week"))
dataTable$Choc <- revalue(dataTable$Choc, c("CL6"="Used in Last Day"))
dataTable$Choc <- as.factor(dataTable$Choc)

dataTable$Cannabis <- revalue(dataTable$Cannabis, c("CL0"="Never Used"))
dataTable$Cannabis <- revalue(dataTable$Cannabis, c("CL1"="Used over a Decade Ago"))
dataTable$Cannabis <- revalue(dataTable$Cannabis, c("CL2"="Used in Last Decade"))
dataTable$Cannabis <- revalue(dataTable$Cannabis, c("CL3"="Used in Last Year"))
dataTable$Cannabis <- revalue(dataTable$Cannabis, c("CL4"="Used in Last Month"))
dataTable$Cannabis <- revalue(dataTable$Cannabis, c("CL5"="Used in Last Week"))
dataTable$Cannabis <- revalue(dataTable$Cannabis, c("CL6"="Used in Last Day"))
dataTable$Cannabis <- as.factor(dataTable$Cannabis)

dataTable$Caff <- revalue(dataTable$Caff, c("CL0"="Never Used"))
dataTable$Caff <- revalue(dataTable$Caff, c("CL1"="Used over a Decade Ago"))
dataTable$Caff <- revalue(dataTable$Caff, c("CL2"="Used in Last Decade"))
dataTable$Caff <- revalue(dataTable$Caff, c("CL3"="Used in Last Year"))
dataTable$Caff <- revalue(dataTable$Caff, c("CL4"="Used in Last Month"))
dataTable$Caff <- revalue(dataTable$Caff, c("CL5"="Used in Last Week"))
dataTable$Caff <- revalue(dataTable$Caff, c("CL6"="Used in Last Day"))
dataTable$Caff <- as.factor(dataTable$Caff)

```

```

dataTable$Benzos <- revalue(dataTable$Benzos, c("CL0"="Never Used"))
dataTable$Benzos <- revalue(dataTable$Benzos, c("CL1"="Used over a Decade Ago
"))
dataTable$Benzos <- revalue(dataTable$Benzos, c("CL2"="Used in Last Decade"))
dataTable$Benzos <- revalue(dataTable$Benzos, c("CL3"="Used in Last Year"))
dataTable$Benzos <- revalue(dataTable$Benzos, c("CL4"="Used in Last Month"))
dataTable$Benzos <- revalue(dataTable$Benzos, c("CL5"="Used in Last Week"))
dataTable$Benzos <- revalue(dataTable$Benzos, c("CL6"="Used in Last Day"))
dataTable$Benzos <- as.factor(dataTable$Benzos)

dataTable$Amyl <- revalue(dataTable$Amyl, c("CL0"="Never Used"))
dataTable$Amyl <- revalue(dataTable$Amyl, c("CL1"="Used over a Decade Ago"))
dataTable$Amyl <- revalue(dataTable$Amyl, c("CL2"="Used in Last Decade"))
dataTable$Amyl <- revalue(dataTable$Amyl, c("CL3"="Used in Last Year"))
dataTable$Amyl <- revalue(dataTable$Amyl, c("CL4"="Used in Last Month"))
dataTable$Amyl <- revalue(dataTable$Amyl, c("CL5"="Used in Last Week"))
dataTable$Amyl <- revalue(dataTable$Amyl, c("CL6"="Used in Last Day"))
dataTable$Amyl <- as.factor(dataTable$Amyl)

dataTable$Amphet <- revalue(dataTable$Amphet, c("CL0"="Never Used"))
dataTable$Amphet <- revalue(dataTable$Amphet, c("CL1"="Used over a Decade Ago
"))
dataTable$Amphet <- revalue(dataTable$Amphet, c("CL2"="Used in Last Decade"))
dataTable$Amphet <- revalue(dataTable$Amphet, c("CL3"="Used in Last Year"))
dataTable$Amphet <- revalue(dataTable$Amphet, c("CL4"="Used in Last Month"))
dataTable$Amphet <- revalue(dataTable$Amphet, c("CL5"="Used in Last Week"))
dataTable$Amphet <- revalue(dataTable$Amphet, c("CL6"="Used in Last Day"))
dataTable$Amphet <- as.factor(dataTable$Amphet)

dataTable$Alcohol <- revalue(dataTable$Alcohol, c("CL0"="Never Used"))
dataTable$Alcohol <- revalue(dataTable$Alcohol, c("CL1"="Used over a Decade A
go"))
dataTable$Alcohol <- revalue(dataTable$Alcohol, c("CL2"="Used in Last Decade"
))
dataTable$Alcohol <- revalue(dataTable$Alcohol, c("CL3"="Used in Last Year"))
dataTable$Alcohol <- revalue(dataTable$Alcohol, c("CL4"="Used in Last Month"
))
dataTable$Alcohol <- revalue(dataTable$Alcohol, c("CL5"="Used in Last Week"))
dataTable$Alcohol <- revalue(dataTable$Alcohol, c("CL6"="Used in Last Day"))
dataTable$Alcohol <- as.factor(dataTable$Alcohol)

```


#After decoding, it looks like:

ID	Age	Gender	Education	Country	Ethnicity
1	35-44	Female	Professional certificate/ diploma	UK	Mixed-White/Asian
2	25-34	Male	Doctorate degree	UK	White
3	35-44	Male	Professional certificate/ diploma	UK	White
4	18-24	Female	Masters degree	UK	White
5	35-44	Female	Doctorate degree	UK	White
6	65-74	Female	Left school at 18 years	Canada	White
7	45-54	Male	Masters degree	USA	White
8	35-44	Male	Left school at 16 years	UK	White
9	35-44	Female	Professional certificate/ diploma	Canada	White
10	55-64	Male	Masters degree	UK	White
11	25-34	Female	University degree	UK	White
12	45-54	Male	Some college or university, no certificate or degree	Other	White
13	55-64	Female	University degree	UK	White
14	55-64	Female	Professional certificate/ diploma	Canada	White
15	55-64	Female	Professional certificate/ diploma	UK	White
16	55-64	Male	University degree	UK	White
17	35-44	Female	Some college or university, no certificate or degree	UK	White
18	45-54	Male	Left school at 16 years	UK	White