Assigment_3

vinoth

September 28, 2017

Question 1 solution for Univariate analysis

```r
dataset=read.csv("C:/Users/admin/Desktop/EDA/assignment 3/ds3.csv")
dataset$score=((4*dataset$VI)+(3*dataset$SWI)+(2*dataset$NVI)+(1*dataset$Nai))/
  (dataset$VI+dataset$SWI+dataset$NVI+dataset$Nai)

dataset=replace(dataset,is.na(dataset),0)


s1=subset(dataset,dataset$question=="q1")
s2=(s1$score-mean(s1$score))/sd(s1$score)
s1$ts=s2

s3=subset(dataset,dataset$question=="q2")
s4=(s3$score-mean(s3$score))/sd(s3$score)
s3$ts=s4

s5=subset(dataset,dataset$question=="q3")
s6=(s5$score-mean(s5$score))/sd(s5$score)
#step used converted the standard score zero
s6[14]=0
s5$ts=s6

s7=subset(dataset,dataset$question=="q4")
s8=((s7$score-mean(s7$score))/sd(s7$score))
s7$ts=s8

tsum<- round(s1$ts+s3$ts+s5$ts+s7$ts,digits=4)

countries=subset(s1,select = c("country"))
univar<-cbind(countries,tsum)
colnames(univar)<-c("country","score")

library(ggplot2)
```
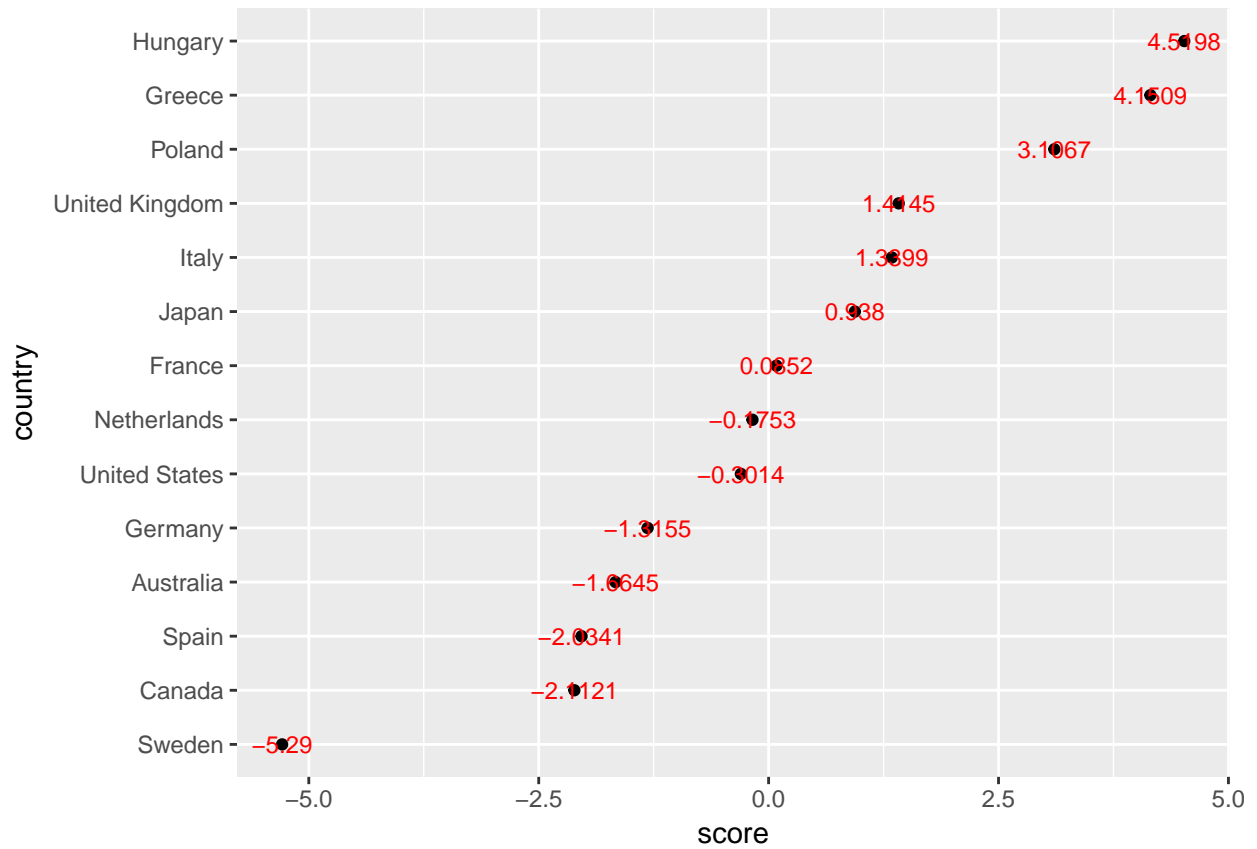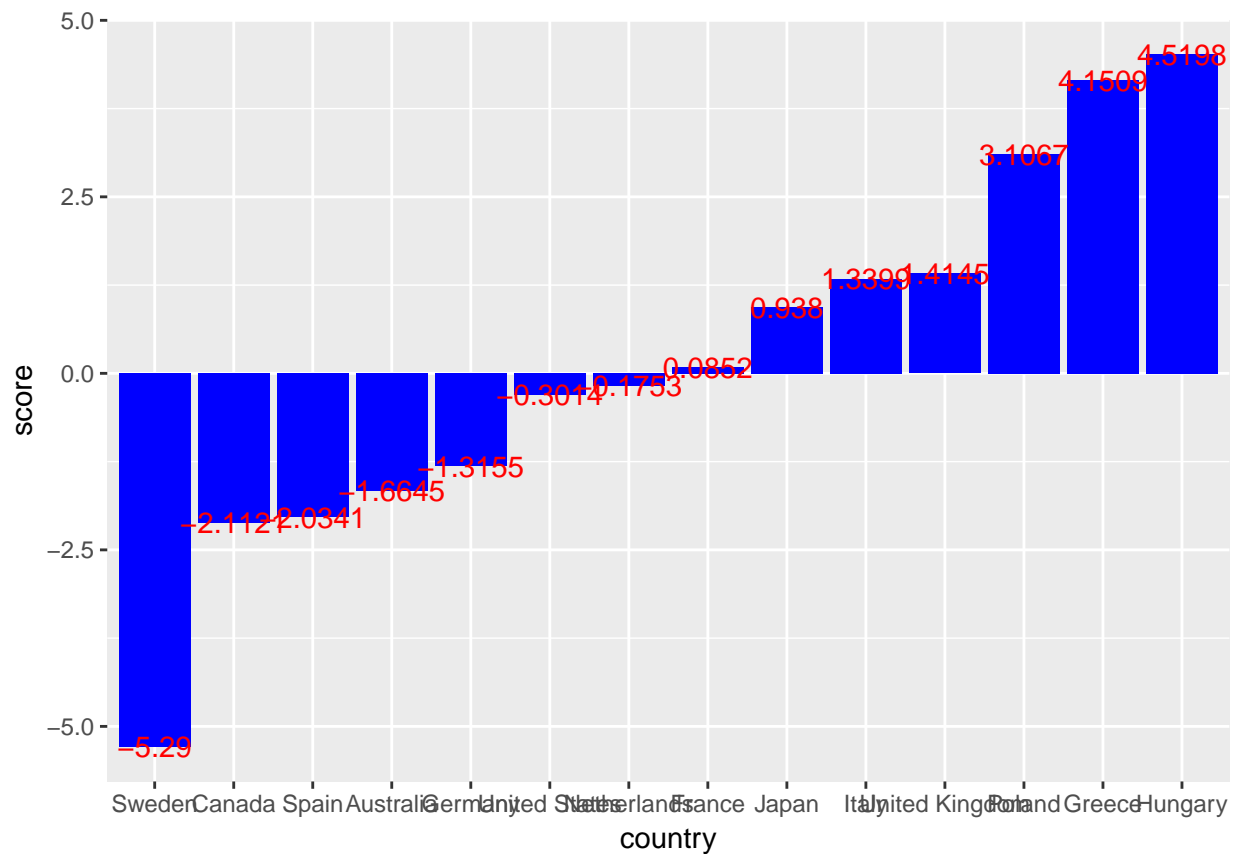
```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```r
library(grid)
univar<-univar[order(univar$score),]
univar$country<-factor(univar$country,levels=univar$country)
ggplot(univar,aes(y=country,x=score,label=score))+geom_point()+geom_text(color="red", size=3)
```

```
ggplot(univar,aes(y=score,x=country,label=score))+geom_bar(stat='identity', fill="blue", size=10)+geom_
```
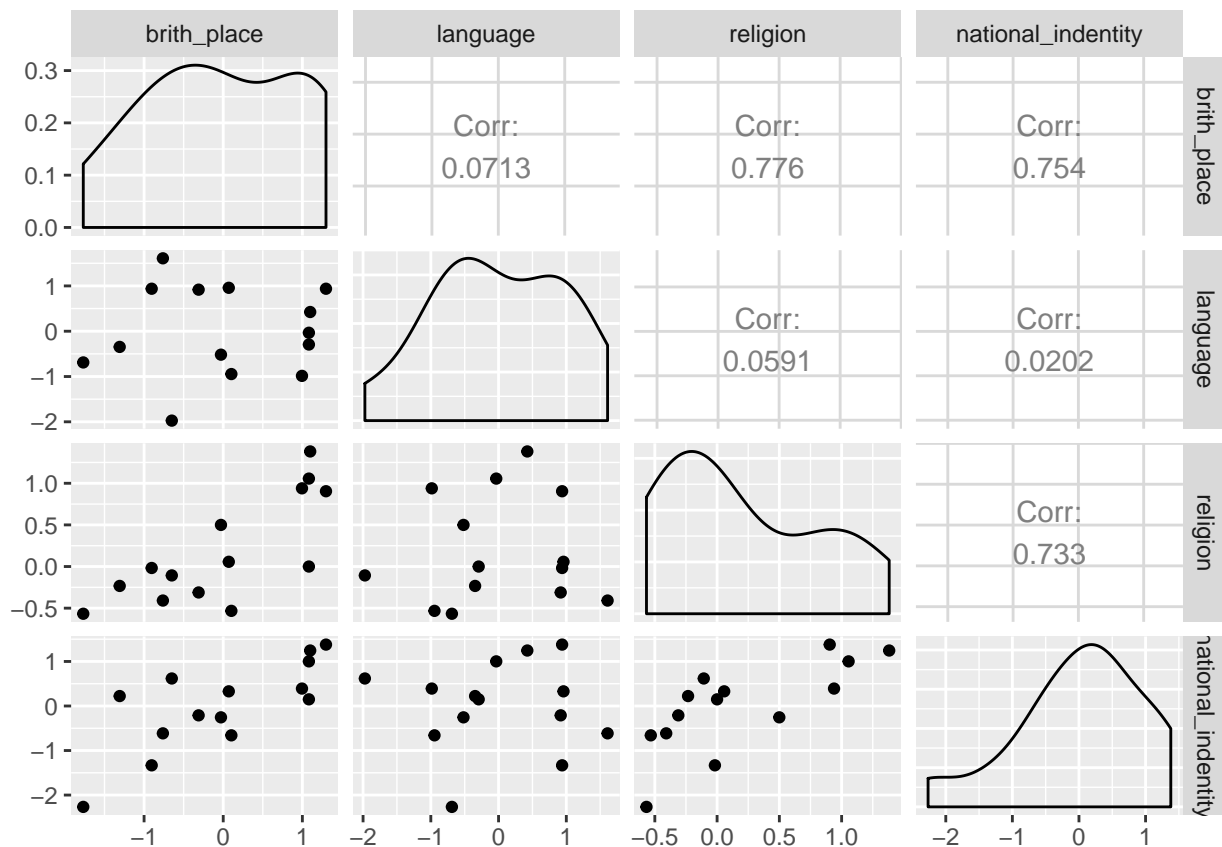
2 solution

```
bivar<-cbind.data.frame(s1$ts,s3$ts,s5$ts,s7$ts)
colnames(bivar)<-c("brith_place","language","religion","national_indentity")
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.3.3
```

```
ggpairs(bivar)
```

if look at the graph given above its clear that language is weekly related to others. And we can also know that birthplace and religion and national_indentity are strong related with other.

solution 3

```
language.cat=cut_number(bivar$language,n=2)
bivar2<-cbind.data.frame(bivar$brith_place,bivar$religion,bivar$national_indentity)
bivar2
```

```
##    bivar$brith_place bivar$religion bivar$national_indentity
## 1        -0.02846888     0.49968117               -0.2549215
## 2        -0.64964149    -0.10686600                0.6177590
## 3        -0.31082007    -0.31128530               -0.2100277
## 4        -0.90375756    -0.01743255               -1.3323728
## 5         1.09910058     1.38207307                1.2454293
## 6         1.29858169     0.90439008                1.3787639
## 7         0.99533652     0.93920235                0.3909995
## 8        -0.76215416    -0.40838446               -0.6140719
## 9         1.08254025     1.05615592                1.0021049
## 10        0.10296559    -0.53230049               -0.6589657
## 11       -1.76835112    -0.56706242               -2.2657149
## 12        0.07035403     0.05703079                0.3286979
## 13       -1.30822563    -0.23309492                0.2231975
## 14        1.08254025     0.00000000                0.1491227
```

```
ggpairs(data.frame(bivar2,language.cat),aes(color=language.cat))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

4

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
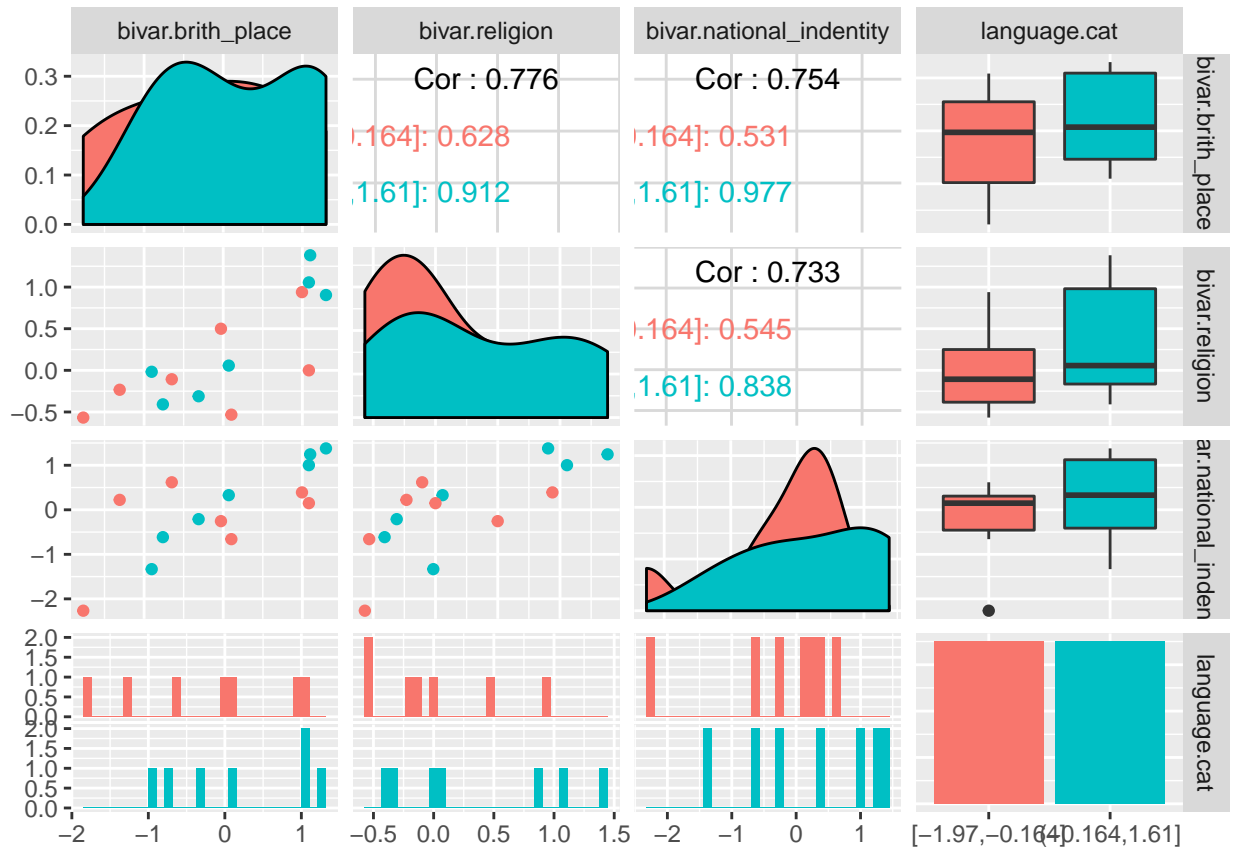


```
ggpairs(data.frame(bivar,language.cat),aes(color=language.cat))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
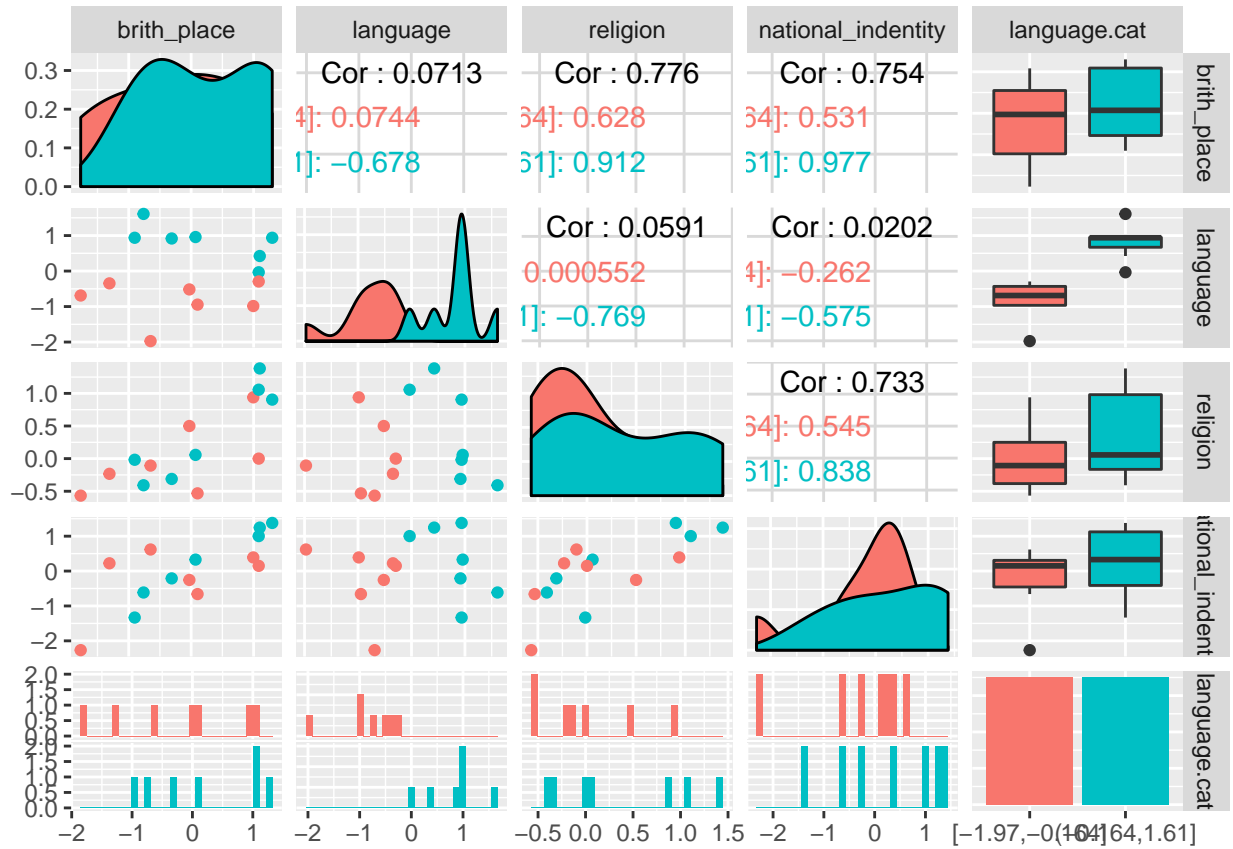
If look at above given graph that the weakly related variable in language are average the same values with others. average is same with all distribution.