R_Code_Prob_3

Syam Sundar Herle

September 29, 2017

R code for the Problem 3

```r
bp=read.csv("Q1.csv",header = FALSE,fileEncoding="UTF-8-BOM")
ln=read.csv("Q2.csv",header = FALSE,fileEncoding="UTF-8-BOM")
rn = read.csv("Q3.csv",header = FALSE,fileEncoding="UTF-8-BOM")
ni =read.csv("Q4.csv",header = FALSE,fileEncoding="UTF-8-BOM")

bp <- transform(bp, score = (4 * bp[2] + 3*bp[3]+2*bp[4]+1*bp[5])/(bp[2]+bp[3]+bp[4]+bp[5]) )
ln <- transform(ln, score = (4 * ln[2] + 3*ln[3]+2*ln[4]+1*ln[5])/(ln[2]+ln[3]+ln[4]+ln[5]) )
rn <- transform(rn, score = (4 * rn[2] + 3*rn[3]+2*rn[4]+1*rn[5])/(rn[2]+rn[3]+rn[4]+rn[5]) )

ni <- transform(ni, score = (4 * ni[2] + 3*ni[3]+2*ni[4]+1*ni[5])/(ni[2]+ni[3]+ni[4]+ni[5]) )
countries <- bp[1]

bp$V2.1 <- round((bp$V2.1 - mean(bp$V2.1))/sd(bp$V2.1),digits = 3)

ln$V2.1 <- round((ln$V2.1 - mean(ln$V2.1))/sd(ln$V2.1),digits = 3)


rn$V2.1 <- round((rn$V2.1 - mean(rn[1:13,"V2.1"]))/sd(rn[1:13,"V2.1"]),digits = 3)

ni$V2.1 <- round((ni$V2.1 - mean(ni$V2.1))/sd(ni$V2.1),digits = 3)

rn[14,8] <- 0
library(ggplot2)
library(grid)
#Rounding all the values to three digit decimal place
vec <- round(bp$V2.1 + ln$V2.1 +rn$V2.1 +ni$V2.1,digits = 3)
univariate <- cbind(countries,vec)
colnames(univariate) <- c("country", "score")
univariate <- univariate[order(univariate$score), ]
univariate$`country` <- factor(univariate$`country`, levels = univariate$`country`)
ggplot(univariate, aes(x=`country`, y=score, label=score)) +
geom_point(stat='identity', fill="black", size=8) +
geom_text(color="white", size=2) +
labs(title="Dot plot for population data",subtitle="22 cities")+coord_flip()

#Binding the data to form a dataframe
bivariate <- cbind.data.frame(bp$V2.1,ln$V2.1,rn$V2.1,ni$V2.1)
colnames(bivariate) <- c("Birth Place","Language","Religion","National Identity")
library(GGally)
ggpairs(bivariate)

language.cat = cut_number(bivariate$Language, n = 2)
ggpairs(data.frame(bivariate, language.cat), aes(color = language.cat))
```