Happiness Analysis

Zhihan Fang

Import happiness data

week2.script.R × | README.md × | Code.R × | week4.script.r × | week5.R × | happiness.data ×

Filter

	IDN	AGE	COUNTRY	GENDER	IMMIGRANT	HAPPINESS
1	73390	38	Cyprus	Male	0	5.67
2	37692	41	Afghanistan	Male	0	3.03
3	92065	55	Finland	Male	0	8.55
4	38549	42	Swaziland	Female	0	5.67
5	33615	47	Kosovo	Female	0	7.12
6	73833	54	Yemen	Female	0	6.16
7	78284	87	Japan	Male	0	7.82
8	30519	54	Malawi	Male	0	4.70
9	86688	60	Laos	Female	0	3.87
10	63978	71	United Arab Emirates	Female	0	8.65
11	80570	38	Angola	Male	0	4.52
12	83683	74	Spain	Male	0	7.78
13	98614	73	Swaziland	Female	0	6.37

Showing 1 to 13 of 10,600 entries

Environment Panel → Import dataset → select your csv file

Check data -- summary

```
> summary(happiness.data)
      IDN              AGE                      COUNTRY          GENDER       IMMIGRANT        HAPPINESS
 Min.   :10001   Min.   :   18   Bolivia                :   89   Female:5277   Min.   :0.0000   Min.   : 1.000
 1st Qu.:32022   1st Qu.:   41   Bosnia and Herzegovina:   88   Male  :5323   1st Qu.:0.0000   1st Qu.: 4.500
 Median :54560   Median :   51   Spain                  :   86                 Median :0.0000   Median : 5.690
 Mean   :54785   Mean   : 2868   Georgia                :   83                 Mean   :0.1114   Mean   : 5.639
 3rd Qu.:77630   3rd Qu.:   73   Jordan                 :   82                 3rd Qu.:0.0000   3rd Qu.: 6.830
 Max.   :99991   Max.   :99841   Denmark                :   81                 Max.   :1.0000   Max.   :10.720
                                 (Other)                :10091
```

AGE mean is 2868 and max 99841, the data is meaningless and not clean. Clean the data by subset function to remove unclean data.

```
happiness.data.subset <- subset(happiness.data,happiness.data$AGE>0 & happiness.data$AGE<120)
```

lapply & tapply

```
> lapply(list(happiness.data.subset$AGE,happiness.data.subset$HAPPINESS),mean)
[[1]]
[1] 54.0132

[[2]]
[1] 5.847105
```

mean value of AGE and HAPPINESS

```
> tapply(happiness.data.subset$AGE,happiness.data.subset$GENDER,mean)
  Female     Male
53.76314 54.26176
```

lapply return a list or vector, each element of which the result of applying a function to the each element of input vector of list.
lapply(list(happiness.data.subset$AGE,happiness.data.subset$HAPPINESS),mean)

tapply to get the average age of male and female separately.
tapply(happiness.data.subset$AGE,happiness.data.subset$GENDER,mean)

plot subset of categorical data

Plot the gender distribution of Russia and India

1. subset data, select people from Russia and India

```
russian.indian <- subset(happiness.data.subset,happiness.data.subset$COUNTRY=='Russia'| happiness.data.subset$COUNTRY=='India')
```

2.create two new factors containing gender and country, new factors will rebuild the levels of categorical data.

```
russian.indian.gender <- factor(russian.indian$GENDER)
russian.indian.country <- factor(russian.indian$COUNTRY)
```

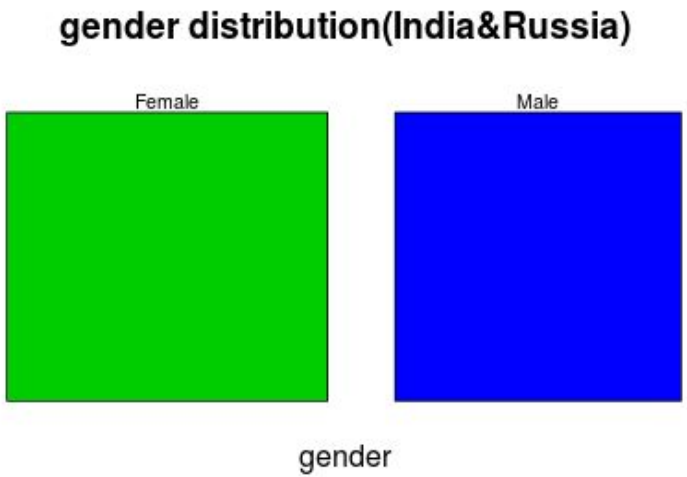
3. mosaicplot the gender distribution of Russian and India

```
mosaicplot(table(russian.indian.country,russian.indian.gender),col=c(3:4),main="gender distribution",xlab="country",ylab = "gender")
```

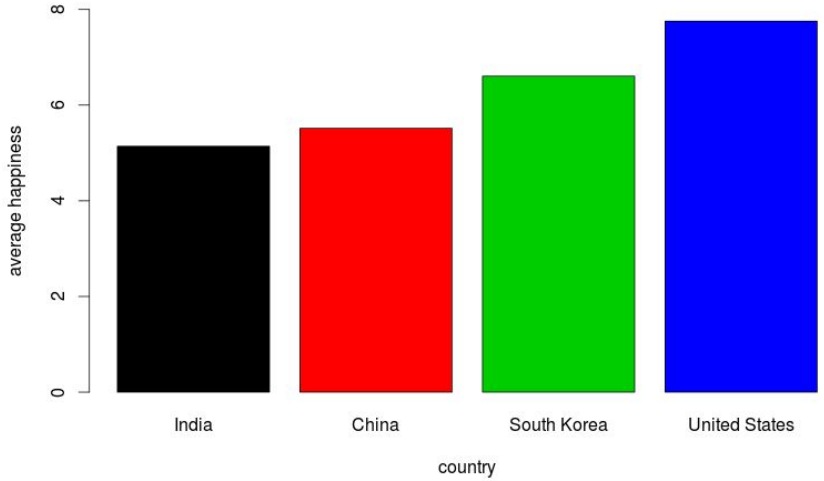
gender distribution(Russia and India)

gender distribution(Russia and India)

let’s put Russia and India together.



average happiness of country



calculate the average happiness of country

```
country.happiness <- tapply(happiness.data.subset$HAPPINESS,happiness.data.subset$COUNTRY, mean)
```

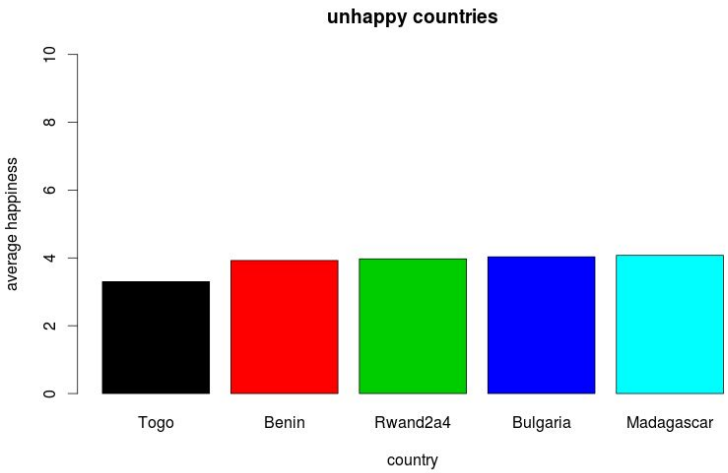
plot happiness of selected countries

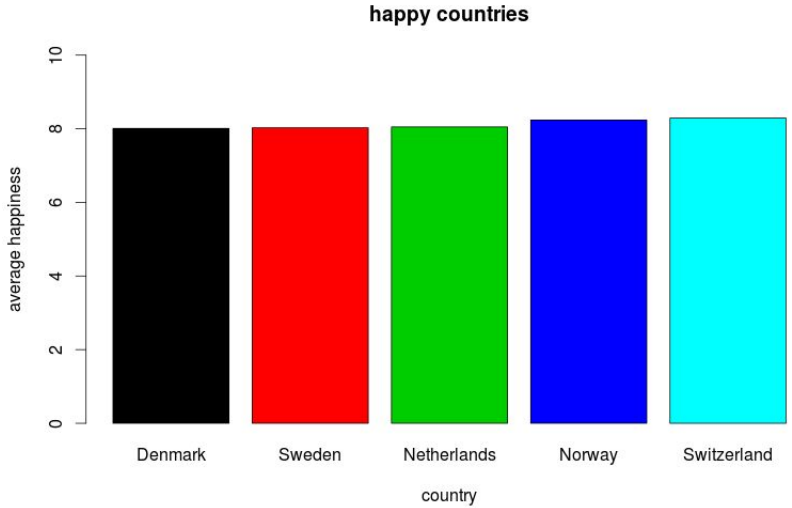
```
barplot(c(country.happiness['India'],country.happiness['China'],country.happiness['South Korea'],country.happiness['United States']),col=c(1:4),ylim=c(0,8),xlab="country",ylab="average happiness")
```

plot top 5 happy and unhappy country

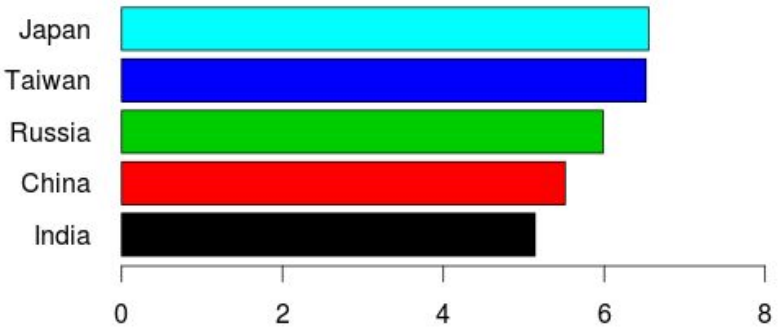
```
#sort average happiness
sorted.country.happiness <- sort(country.happiness)
#plot top 5 unhappy country
barplot(sorted.country.happiness[1:5],col=c(1:5),ylim=c(0,10),xlab="country",ylab="average happiness",main="unhappy countries")
#plot top 5 happy country
l <- length(sorted.country.happiness)
barplot(sorted.country.happiness[(l-4):l],col=c(1:5),ylim=c(0,10),xlab="country",ylab="average happiness",main="happy countries")
```

plot top 5 happy and unhappy country





flip x y



las to control the axis label
las=1 always horizontal
horiz = TRUE means to flip x y

```
barplot(c(country.happiness['India'],country.happiness
['China'],country.happiness['Russia'],country.happiness
['Taiwan'],country.happiness['Japan']),col=c(1:5),xlim=c
(0,8),las=1,horiz = TRUE)
```