aaa

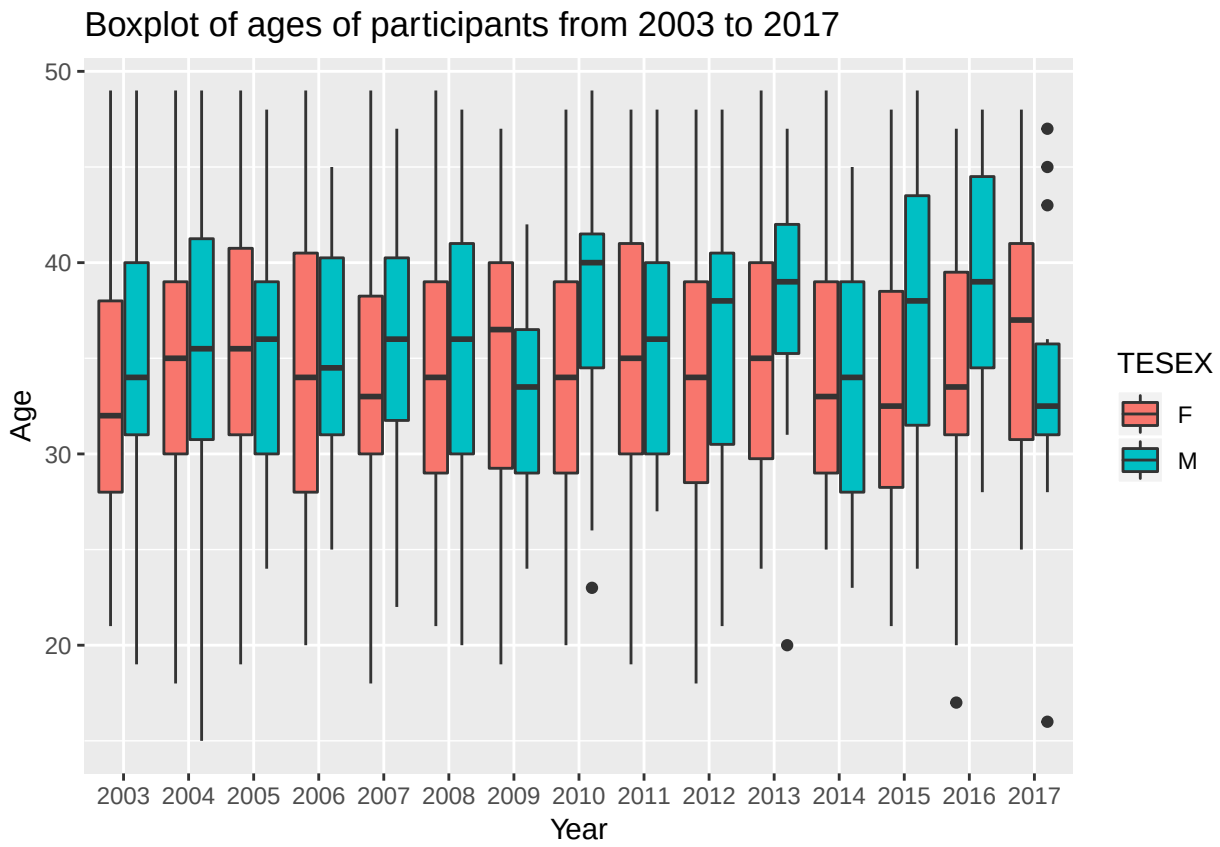
aaa

11/27/2018

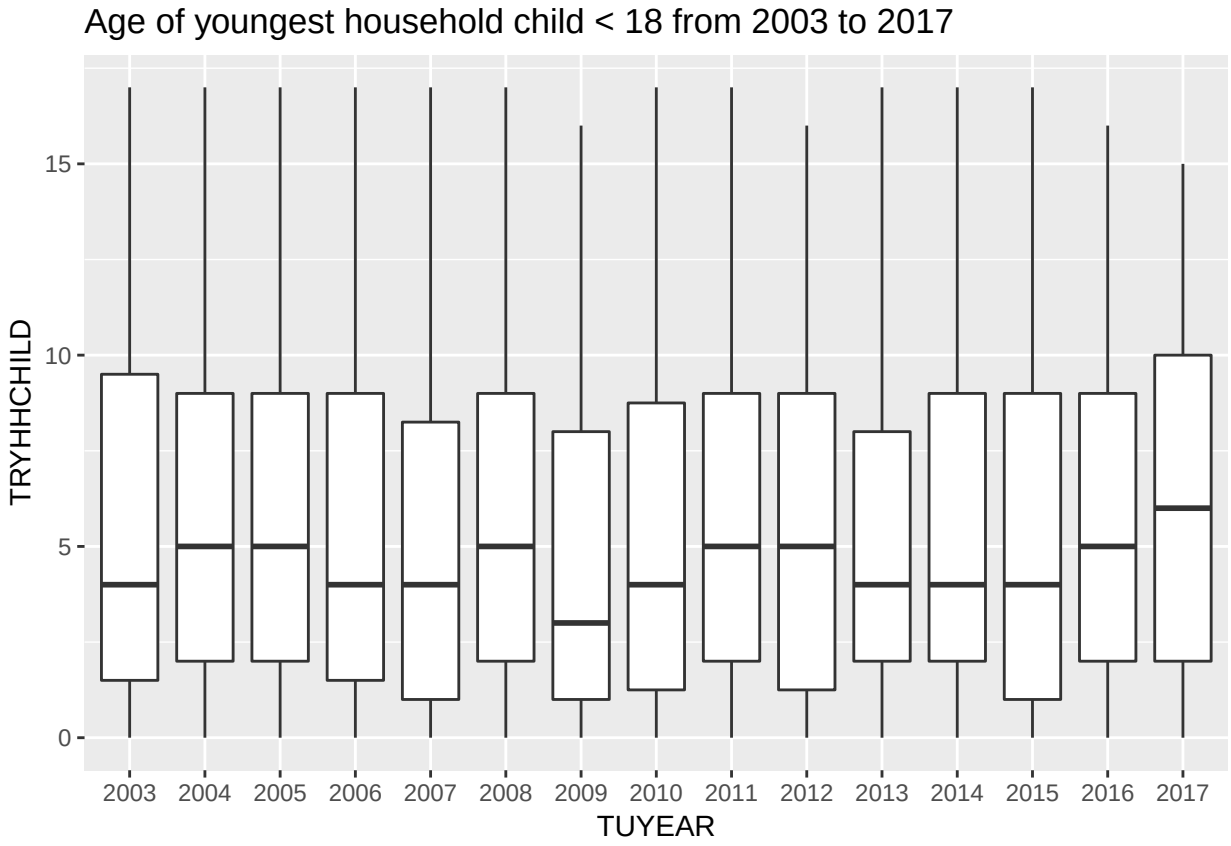
The code below is to check the distribution of age group in those 15 years. It would not be a problem if there is a participant who have joined the survey several times, as her/his data would be regarded as independent samples in different years.

```
#write.csv(Data_minus_rows,'Dataset.csv')
suppressMessages(library(rio))
Dataset <- import('Dataset.csv')

suppressMessages(library(dplyr))
suppressMessages(require(ggplot2))
suppressMessages(library(dplyr))
age <- select(Dataset, TUYEAR, TEAGE,TESEX)
age$TUYEAR <- as.factor(age$TUYEAR)
p <- ggplot(data = age , aes(x=TUYEAR, y=TEAGE,fill=TESEX)) + geom_boxplot(width=0.8)
p + ggtitle('Boxplot of ages of participants from 2003 to 2017') + xlab('Year') + ylab('Age')
```




```
youngest.child <- select(Dataset, TUYEAR, TRYHHCHILD)
youngest.child$TUYEAR <- as.factor(youngest.child$TUYEAR)
p <- ggplot(data = youngest.child , aes(x=TUYEAR, y=TRYHHCHILD)) + geom_boxplot()
p + ggtitle("Age of youngest household child < 18 from 2003 to 2017")
```




```
no.kid <- select(Dataset, TUYEAR, TRCHILDNUM,tu04)
plot(x=no.kid$TRCHILDNUM,y=no.kid$tu04)
```