Hw1:Teengamb Data Exploratory data analysis

Jingxian Chen

Sep 14th, 2019

1. Numerical Summaries:

```
summary(teengamb)
```

```
##      sex         status          income          verbal
##  male  :28   Min.   :18.00   Min.   : 0.600   Min.   : 1.00
##  female:19   1st Qu.:28.00   1st Qu.: 2.000   1st Qu.: 6.00
##              Median :43.00   Median : 3.250   Median : 7.00
##              Mean   :45.23   Mean   : 4.642   Mean   : 6.66
##              3rd Qu.:61.50   3rd Qu.: 6.210   3rd Qu.: 8.00
##              Max.   :75.00   Max.   :15.000   Max.   :10.00
##      gamble
##  Min.   :  0.0
##  1st Qu.:  1.1
##  Median :  6.0
##  Mean   : 19.3
##  3rd Qu.: 19.4
##  Max.   :156.0
```

According to the summary above,

(1) we get 9 more male samples than female.

(2) the mean of "income" variable is 4.642 pounds per week, the median of it is 3.250 pounds per week.

(3) the mean of "gamble" variable is 19.3 pounds per year, the median of it is 6.0 pounds per year.

I think the reason why the mean values of these two variables are much larger than the median is that they both have outliers which get high value. I will use graphical summary to prove my point.

---

2. The number of different values of verbal:

```
length(unique(teengamb$verbal))
```

```
## [1] 9
```

So there are 9 different values of "verbal" variable.

---

3. The row numbers of the verbal outlier observations:

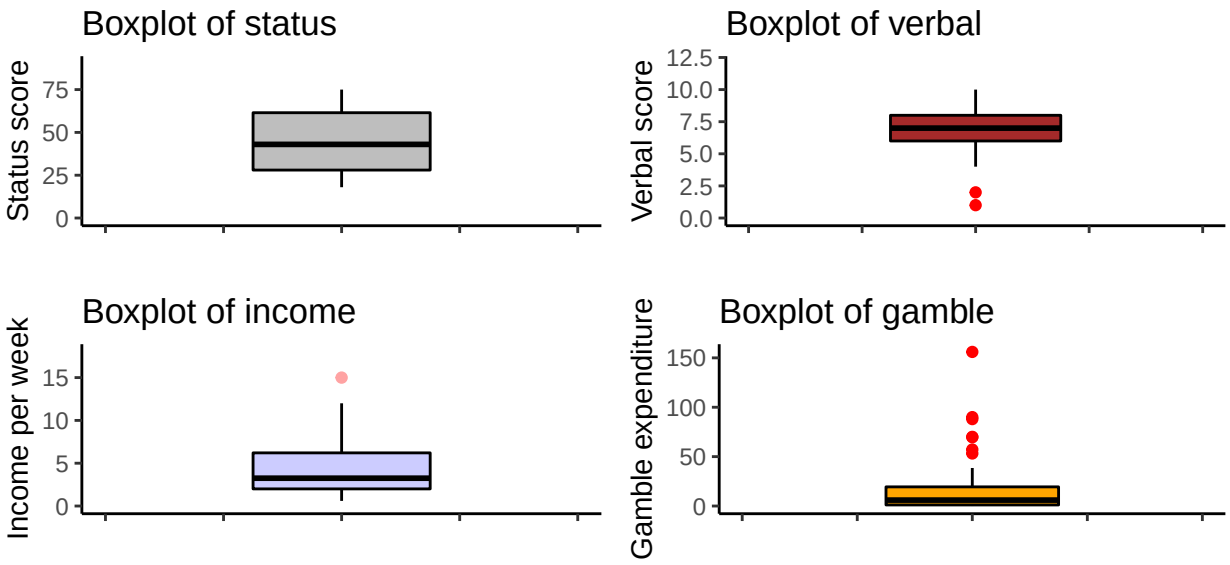
```
which(teengamb$verbal<5)
```

```
## [1]  4 13 24 27 31 35
```

---

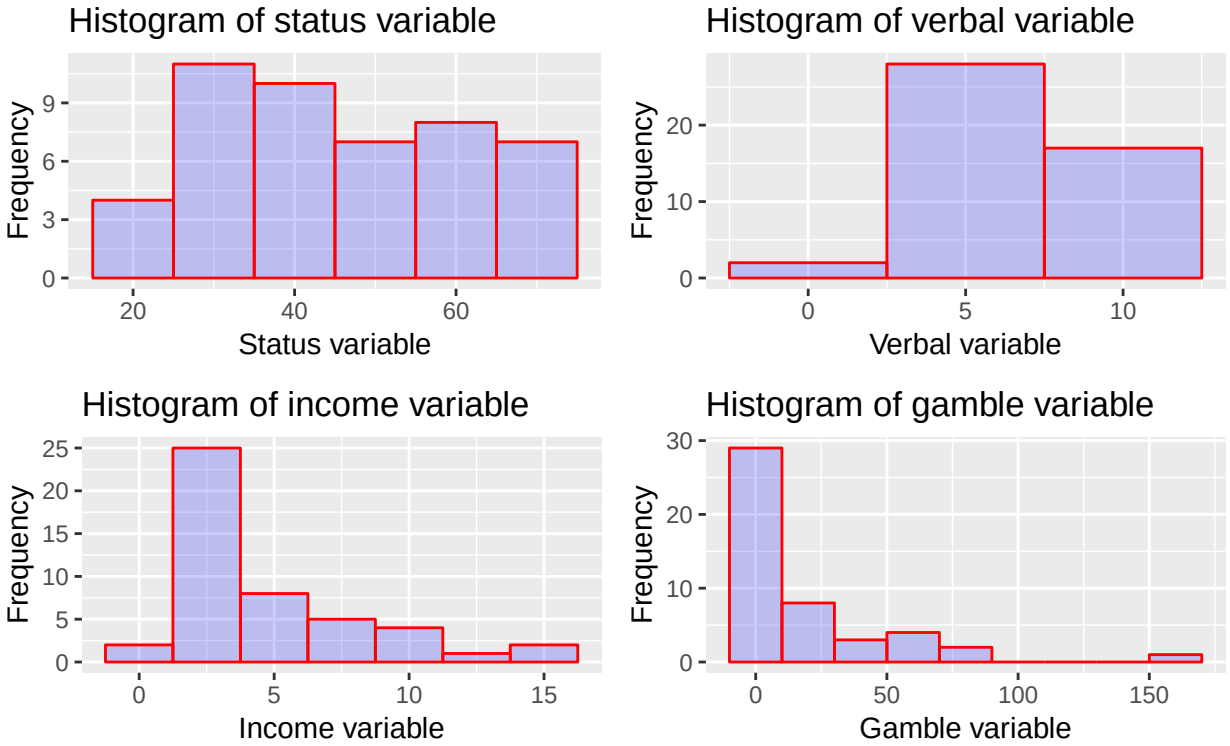
4. Graphical Summaries:

4.1 Boxplots of 4 variables(except "sex")

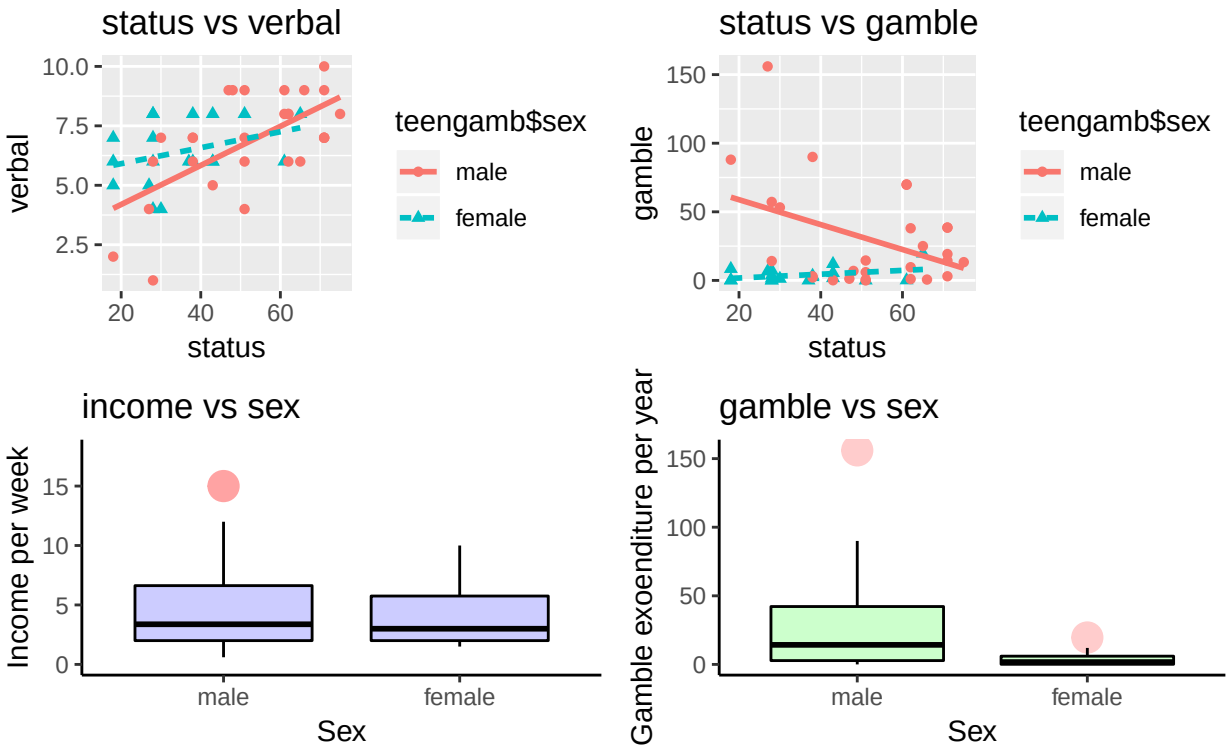


As we can see from the boxlpots above, there are some outliers exist in three variables: verbal, income and gamble. Based on these, we can conclude that if a person' s verbal score is less than 5, we can consider this sample is an outlier.

4.2 Histograms of 4 variables(except "sex")



4.3 Compare two variables



Therefore, when comparing two variable simultaneously, we can get some deduction from the above plot:

(1) Whatever male or female observations, people get a high status score tend to also get a high verbal score.

(2) Consider the male observations, people get a high status score tend to spend less money on gambling. While female observations always spend few money on gambling whatever their status scores high or low.

(3) Male and female observations tend to get similar income amount per week.

(4) Male observations tend to spend more money on gambling per year than female observations.