Stats500 Homework 1

Di Lu, Sept.20, 2017

1. Chapter 1, problem 2, "uswages"

Dataset:

Dataset: The uswages data frame has 2000 subjects and 10 variables. Weekly Wages for US male workers sampled from the Current Population Survey in 1988.

Descriptive analysis

First, we look at the distribution of weekly wages of the 2000 participants in the survey. Secondly, we look at the factors that may have relationship or influence on weekly wages.
From the figure below, we can see:

(1) wage: wage has a wide range from min 50.39 to max 7716.05, characterizing a dramatic inquality of wealth.

(2) wage: wage has a long tail. In the boxplot, the people whose wages are higher than third quantile + 1.5IQR are not few. They are above average. The median of population 522.32 is lower than the 608.12, so the distribution is skewed towards right, as indicated from the histogram.

```
attach(uswages)
summary(wage)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   50.39  308.64  522.32  608.12  783.48 7716.05

par(mfrow=c(1,2))
boxplot(uswages$wage, xlab = "weekly wages ($)")
hist(uswages$wage,breaks = 30, xlim = range(0,2000),  xlab = "weekly wages ($)",main="
")
```
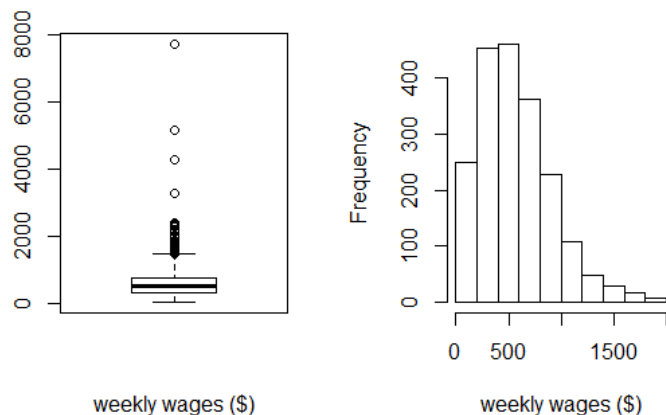

(3) education: As we may suppose, people with higher education will get better jobs and have higher salary, from data plot, we can tell people with higher education have higher wages on average and they can reach higher range compared to poeple with lower edcuation.It is true especially before people reach 12 years of education.

(4) experiences: we cannot tell experience has positive relationship with wages. For people whose wages over 2000, they seem have same years of experiences. Meanwhile, we can see

a clear decreasing trend in wage after 35 years of experiences. That is probably old workers cannot contribute as much as young workers even if their experiences may give an edge on work efficiency.
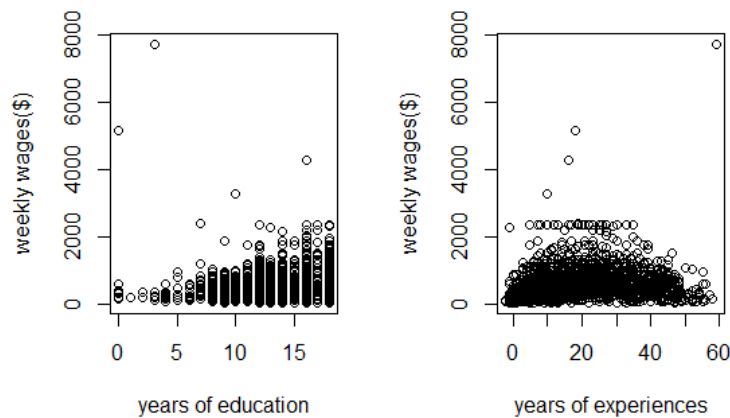

(5) race: among 2000 participants, there are 1844 white people and 156 people. White people receive higher wages on average and reach higher ranges. Almost all the most wealthiest people in this population are white. This may indicate the glass ceiling of races.

(6) metro area: among 2000 participants, there are 1512 people who live in an Standard Metropolitan Statistical Area and 488 people who do not. People living in metro areas receive higher wages on average ad reach higher ranges. This may indicate there are better job opportunities in the city and those who are richest are from the metro areas.

(7) part time or not:among 2000 participants, 1815 work full time while 185 work part time.Clearly, part time workers receive much lower weekly wages bacause of both fewer working hours and lower hour wage.

```
## white black
##  1844   156

##   not in metro live in metro
##            488          1512

## full time part time
##      1815          185
```
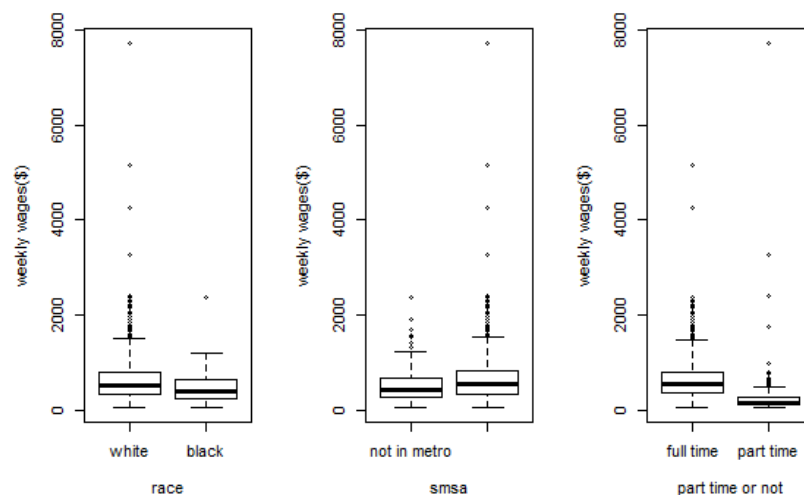
(8) region:this survey has divided the map into four regions, they are ne: North East; mw: Midwest; we: West; so: South.The dataset is well documented, because each participate has assigned to a region, no missing value occurred. Wages across different shows no significnat difference from the boxplot. The relationship between experience and wages are approximately same across regions.
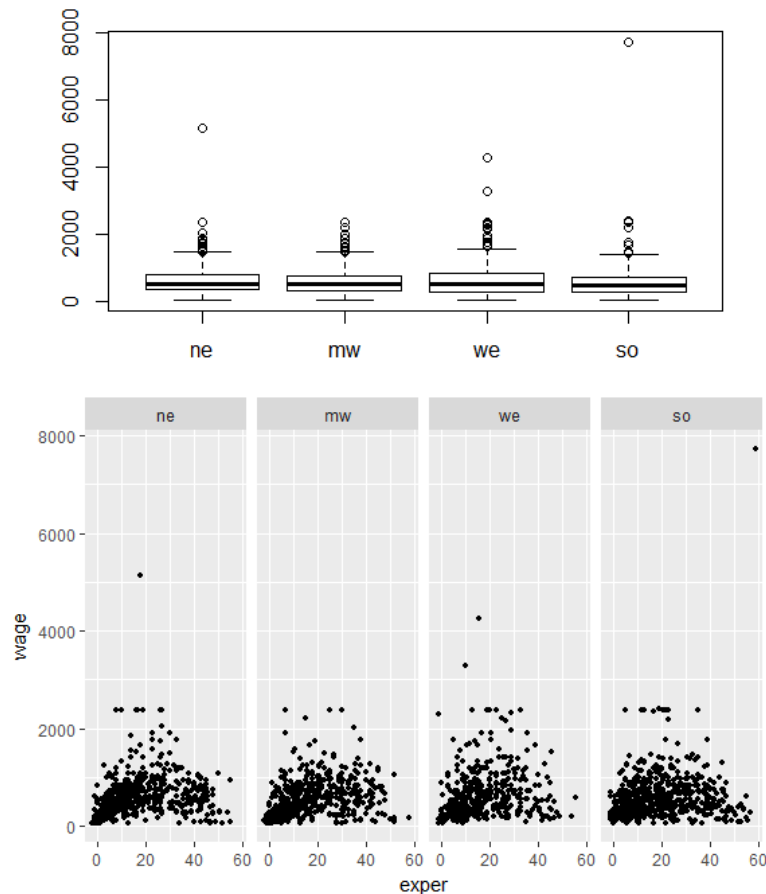
Summary:

(1) wage has a wide range from min 50.39 to max 7716.05, characterizing a dramatic inquality of wealth.Also the median of wages 522.32 is lower than the 608.12, which corresponds to a long tail and right skewness.
(2) People with higher education receive higher wages, it is true especially before people reach 12 years of education.
(3) People live in metro areas receive higher wages on average; white people receive higher wages than black people on average; full time workers earn more than part time workers.
(4) Wages and its relationship with other factors show no obious difference across regions.