Week 2_Exercise

Fenqi Guo

2023-01-17

```
library("haven")
library("car")
library("ggplot2")

data <- read_dta("Teaching_Dataset.dta")
```

Question 1A

```
summary(data$Pr_1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    1.00   14.32   12.00  100.00
```

```
sd(data$Pr_1)
```

```
## [1] 25.8003
```

```
summary(data$Pr_2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    8.00   21.46   35.00  100.00
```

```
sd(data$Pr_2)
```

```
## [1] 29.10451
```

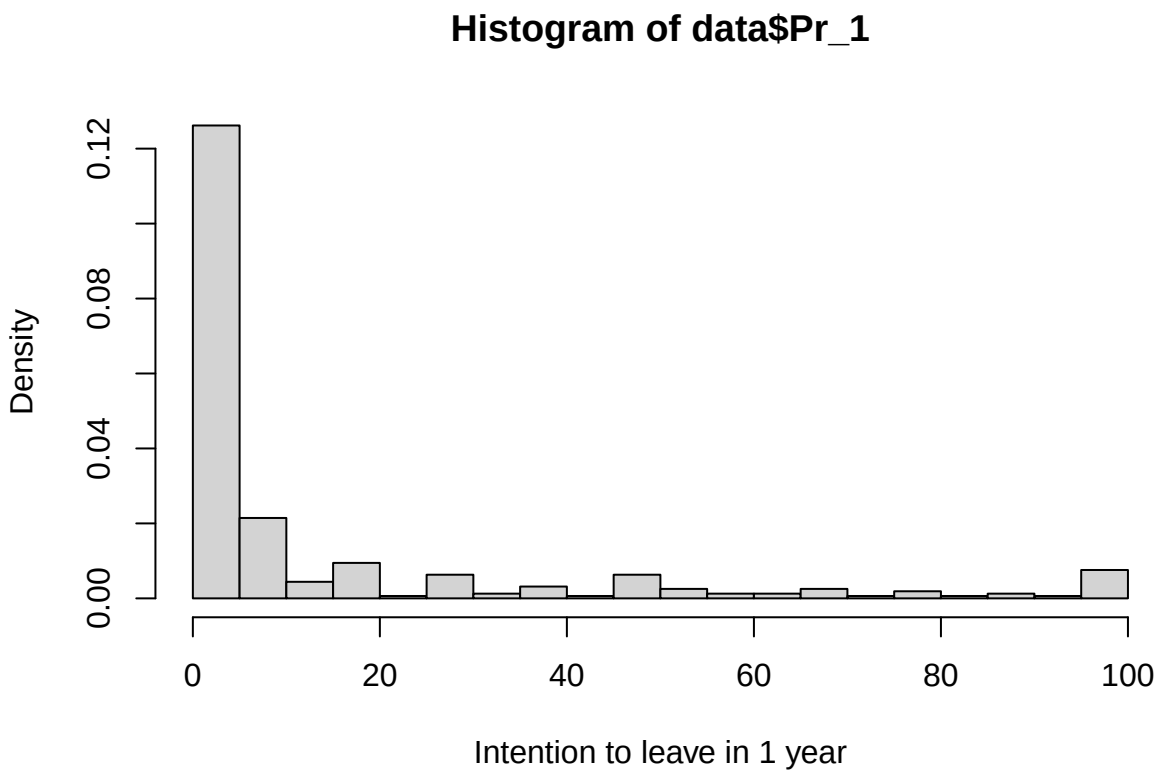
```
summary(data$Pr_3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0     4.0    30.0    37.3    61.0   100.0
```

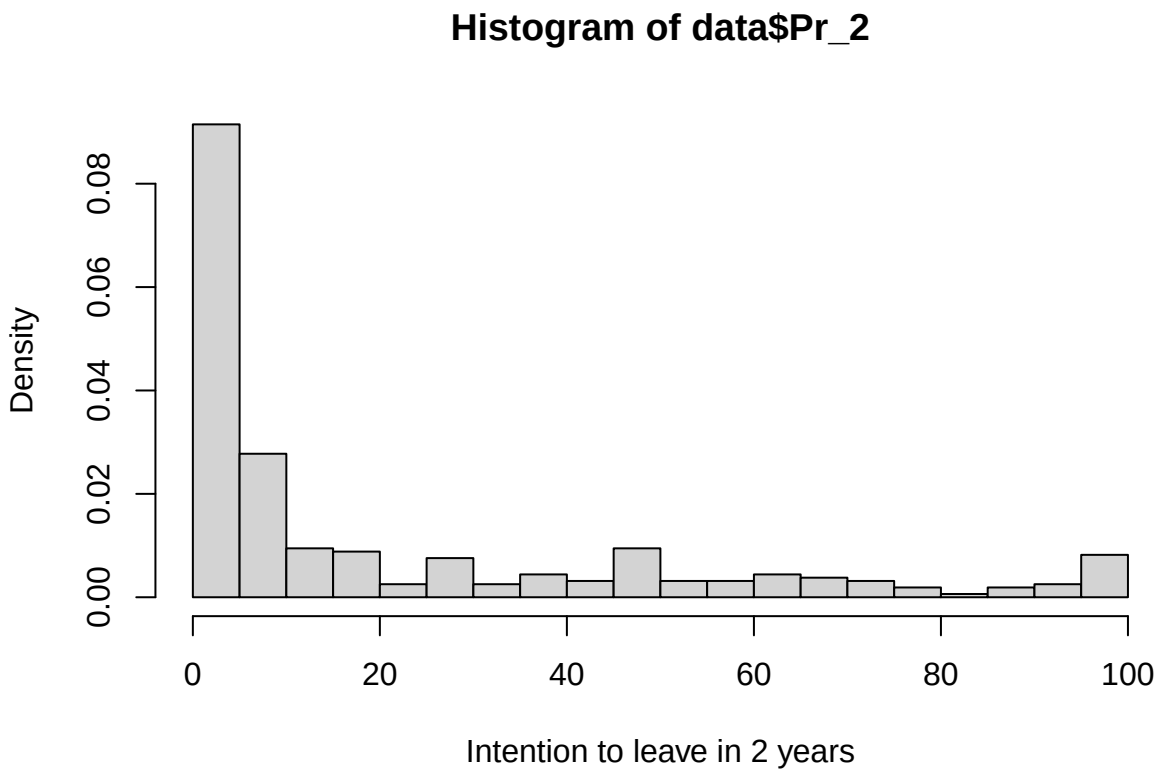
```
sd(data$Pr_3)
```

```
## [1] 34.09407
```

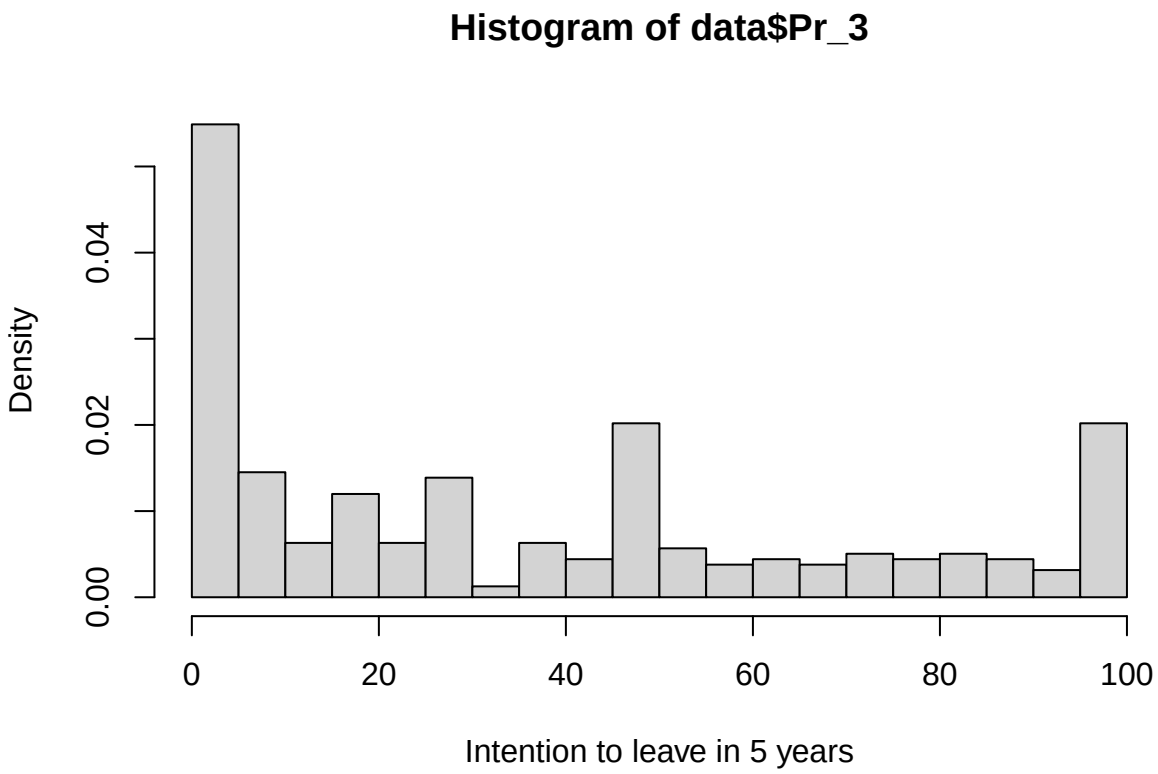
```r
#histogram:
hist(data$Pr_1, freq = FALSE, breaks = 20, xlab = "Intention to leave in 1 year")
```




```r
hist(data$Pr_2, freq = FALSE, breaks = 20, xlab = "Intention to leave in 2 years")
```

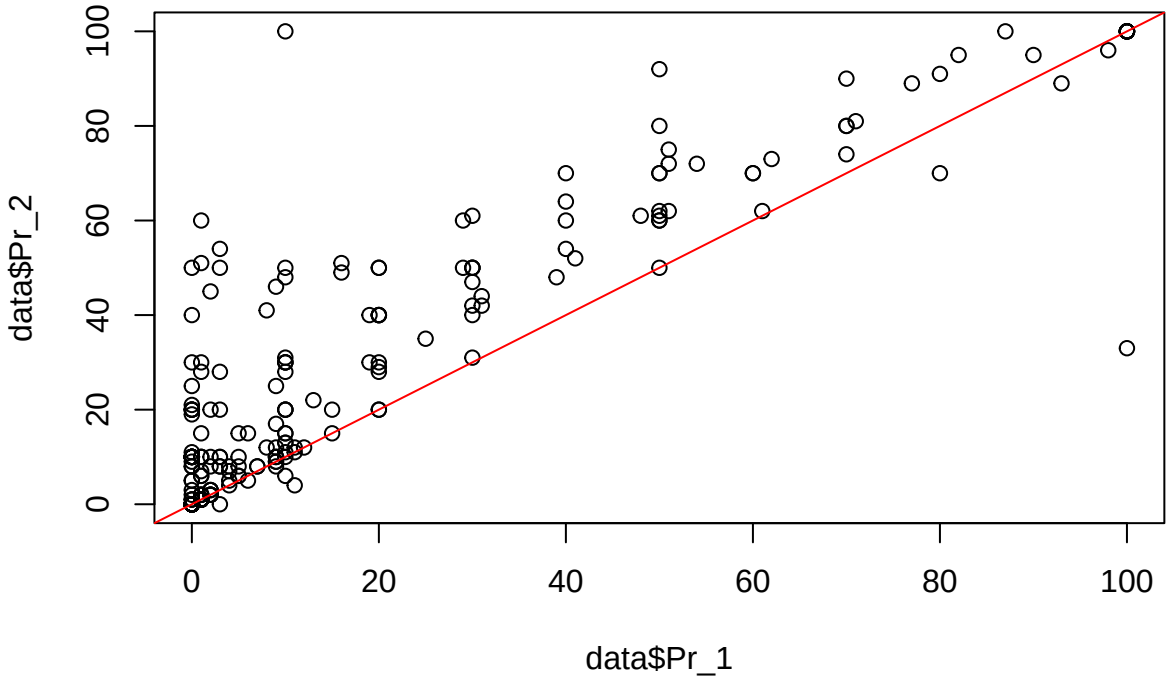



```
hist(data$Pr_3, freq = FALSE, breaks = 20, xlab = "Intention to leave in 5 years")
```

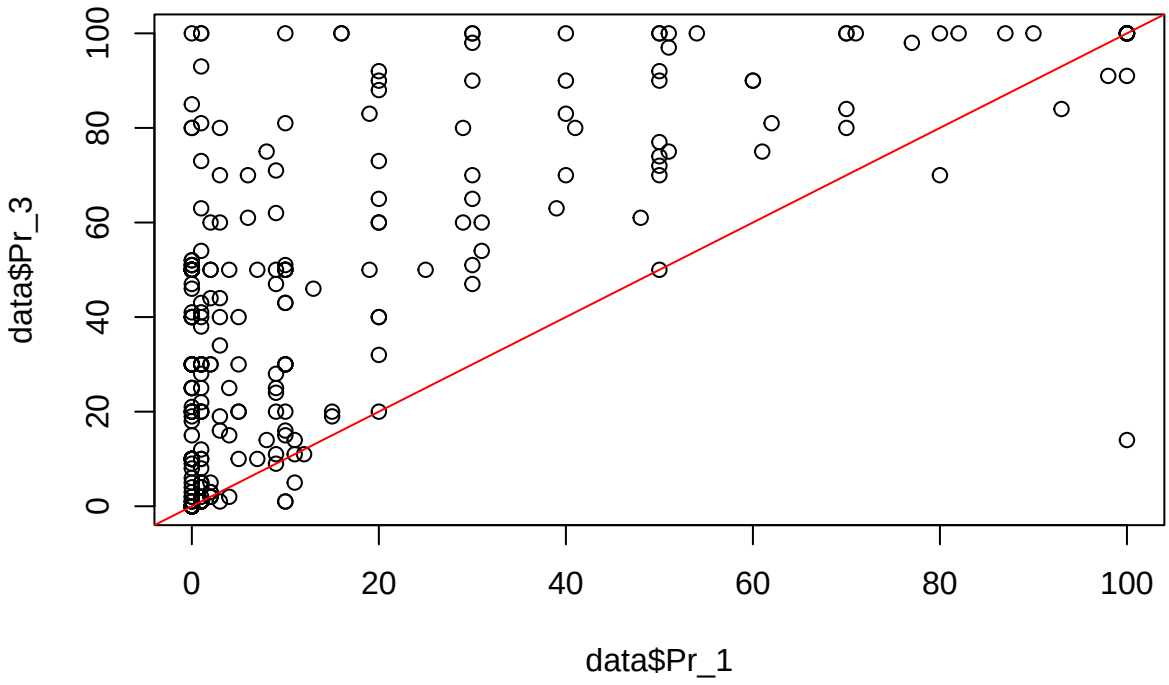


Question 1B and 1C

```
#scatter plot:
plot(data$Pr_1, data$Pr_2)
abline(coef = c(0,1), col = "red")
```




```
plot(data$Pr_1, data$Pr_3)
abline(coef = c(0,1), col = "red")
```




```
#correlation:
cor(data$Pr_1, data$Pr_2)
```

```
## [1] 0.8985178
```

```
cor(data$Pr_1, data$Pr_3)
```

```
## [1] 0.681897
```

Question 2

```
summary(data$Not_In_Teaching)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.08833 0.00000 1.00000
```

```
sd(data$Not_In_Teaching)
```

```
## [1] 0.2842201
```

```
summary(data$Not_In_Teaching_2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.2145  0.0000  1.0000
```

```
sd(data$Not_In_Teaching_2)
```

```
## [1] 0.4111317
```

```
#correlation:
cor(data$Pr_1, data$Not_In_Teaching)
```

```
## [1] 0.5510881
```

```
cor(data$Pr_2, data$Not_In_Teaching_2)
```

```
## [1] 0.5655542
```

Question 3A

```
summary(data$Teacher_Pay)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9540   26298   36000   35641   42300  150000
```

```
sd(data$Teacher_Pay)
```

```
## [1] 12396.61
```

```
tapply(data$Teacher_Pay, data$PT, summary)
```

```
## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   19739   30000   38000   37713   44000   70000
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9540   22000   26649   30076   36475  150000
```

```
summary(data$Non_Teacher_Wage_Own)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9360   28000   35000   36448   40000  166400
```

```
sd(data$Non_Teacher_Wage_Own)
```

```
## [1] 15568.36
```

```
tapply(data$Non_Teacher_Wage_Own, data$PT, summary)
```

```
## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10000   30000   35000   38299   45000  166400
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9360   22750   30000   31478   37087   99450
```

```
summary(data$Non_Teacher_Wage_Ave)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9360   29000   35000   36041   40000  166400
```

```
sd(data$Non_Teacher_Wage_Ave)
```

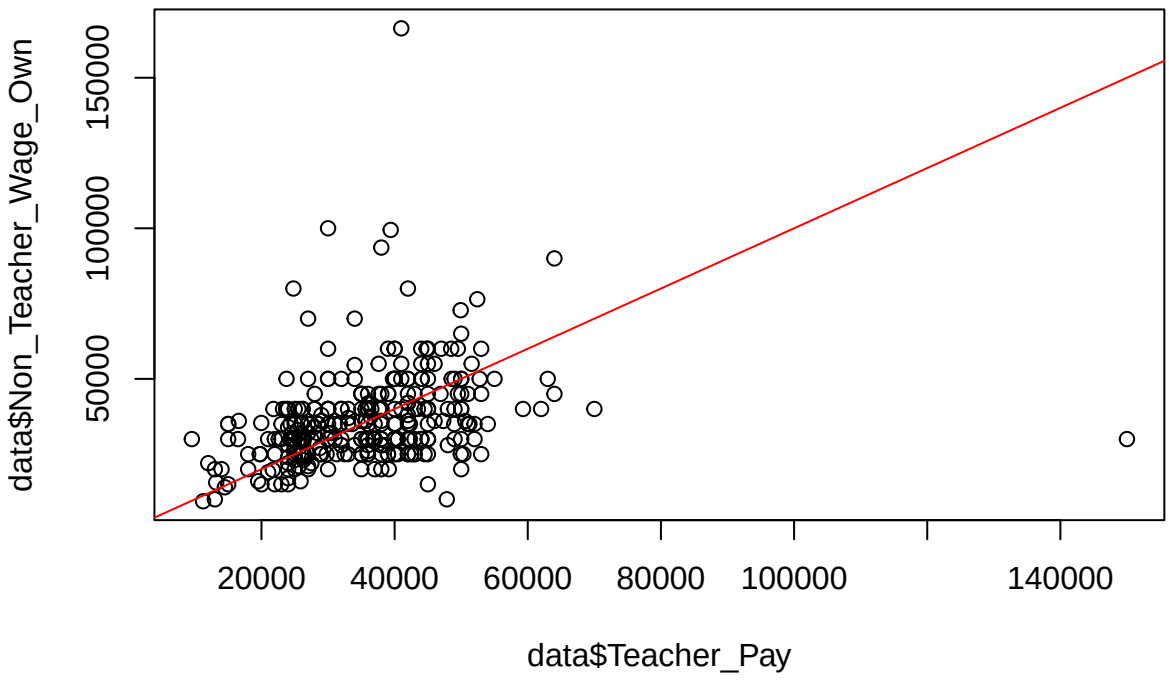
```
## [1] 14384.03
```

```
tapply(data$Non_Teacher_Wage_Ave, data$PT, summary)
```

```
## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10000   30000   35000   37280   41000  166400
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9360   25000   30000   32711   40000   88400
```

Question 3B

```
plot(data$Teacher_Pay, data$Non_Teacher_Wage_Own)
abline(coef = c(0,1), col = "red")
```



Question 3C

```
cor(data$Non_Teacher_Wage_Own, data$Pr_1)
```

```
## [1] -0.03955138
```

```
cor(data$Non_Teacher_Wage_Own, data$Pr_2)
```

```
## [1] -0.01799893
```

```
cor(data$Non_Teacher_Wage_Own, data$Pr_3)
```

```
## [1] 0.0155226
```

Question 4

```
#Are teachers who are secondary earners in their household more or less
#likely to leave? What does this say about the role of financial constrains?
tapply(data$Not_In_Teaching, data$Partner_Earn_More, summary)
```

```
## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    0.00    0.08    0.00    1.00
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1026  0.0000  1.0000
```

```
t.test(data$Not_In_Teaching~data$Partner_Earn_More)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$Not_In_Teaching by data$Partner_Earn_More
## t = -0.66155, df = 221.31, p-value = 0.5089
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.08978163  0.04465343
## sample estimates:
## mean in group 0 mean in group 1
##       0.0800000       0.1025641
```

```
#Do teachers earnings differ by school sector (Independent schools). Might
#this explain why pupil sectors schools struggle to recruit?
tapply(data$Teacher_Pay, data$Independent_School_Dummy, summary)
```

```
## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9540   26298   35806   35418   42000  150000
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13000   29000   38000   37146   44500   59280
```

```
t.test(data$Teacher_Pay~data$Independent_School_Dummy)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$Teacher_Pay by data$Independent_School_Dummy
## t = -0.93134, df = 57.356, p-value = 0.3556
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -5444.614  1987.491
## sample estimates:
## mean in group 0 mean in group 1
##        35417.83        37146.39
```