R Notebook

sn0wfree

Contents

have a try

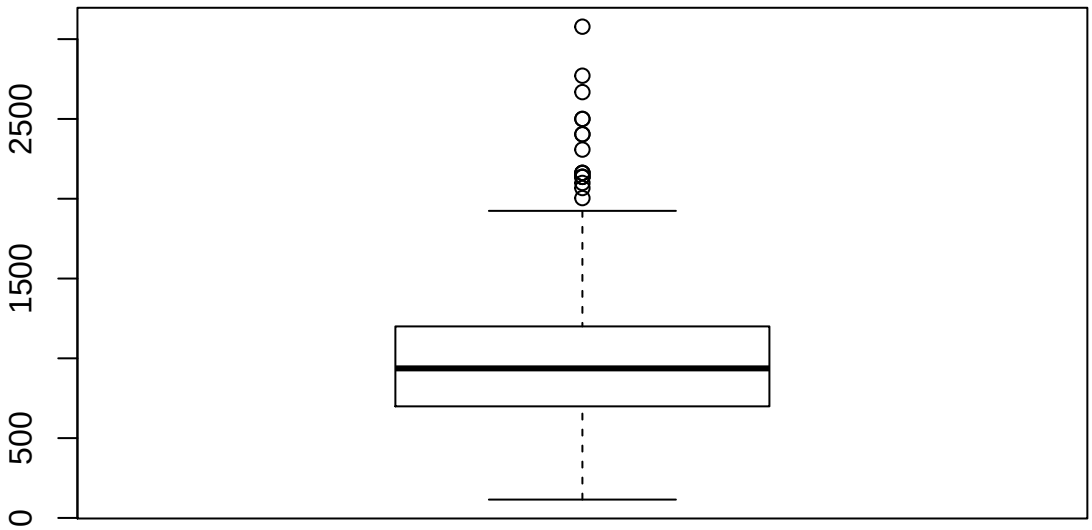
import data

```
#dd=read.table("/Users/sn0wfree/Dropbox/PhD_1st_study/BST215_Quantitative_Research_Methods_Term_1/r_cod
dd=read.csv("/Users/sn0wfree/Dropbox/PhD_1st_study/BST215_Quantitative_Research_Methods_Term_1/r_code/D
head(dd)
```

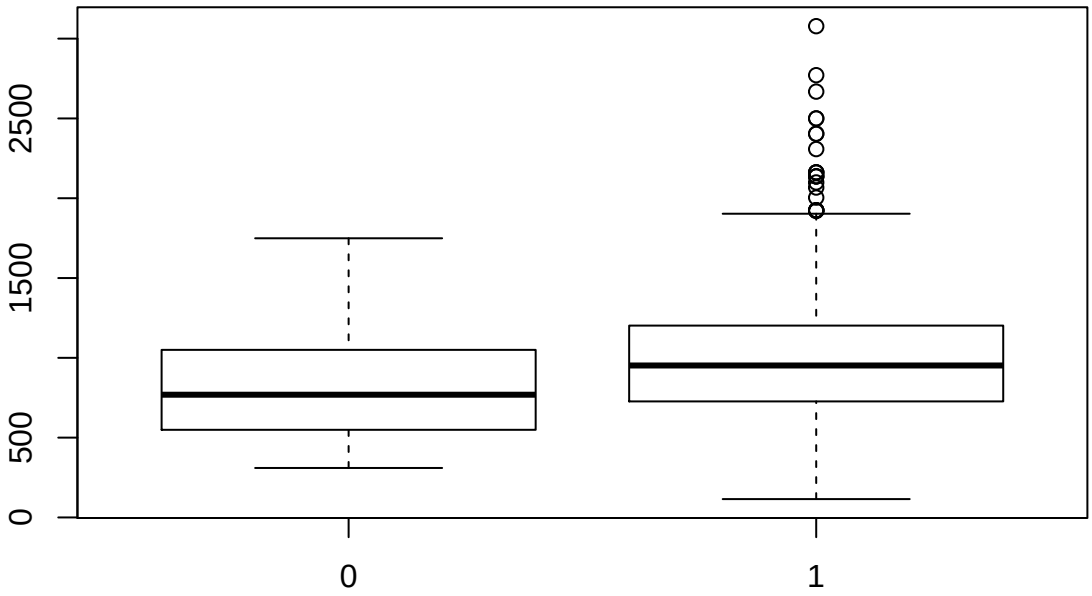
```
##   wage hours  IQ educ exper tenure age married south urban sibs brthord
## 1  866    40  85   12    11     12  30       1     1     0    8       5
## 2  926    55  64   12    18      0  38       1     0     1    4       5
## 3 1400    40  92   12    15      5  36       1     1     1    2       1
## 4  400    50  68   12     4      9  31       1     1     1    6       1
## 5  950    40  96   12    14      2  35       1     0     1    1       1
## 6  560    49 112   16    13      2  32       1     1     0    3       1
##   meduc feduc
## 1     6     0
## 2     0     2
## 3     6     2
## 4     6     2
## 5     8     2
## 6     8     2
```

boxplot

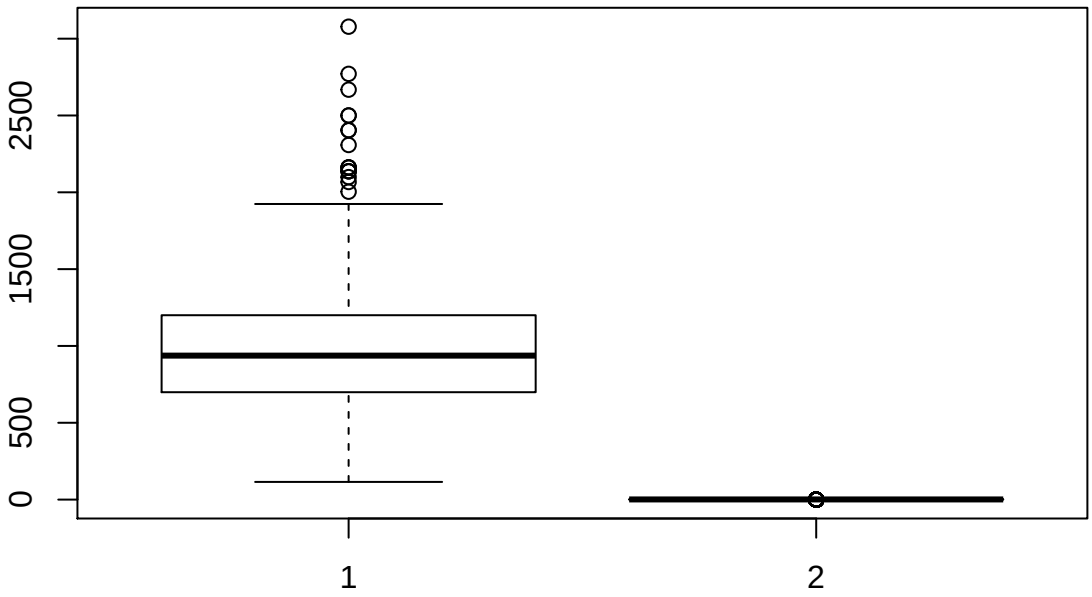
```
boxplot(dd$wage)
```




```
boxplot(dd$wage~dd$married)
```




```
boxplot(dd$wage,dd$married)
```




```
mean(dd$wage)
```

```
## [1] 988.4751
```

```
table(dd$south,dd$married)
```

```
##
##       0   1
##   0  45 404
##   1  21 193
```

```
by(dd$wage,dd$south,mean)
```

```
## dd$south: 0
## [1] 1024.993
## ------------------------------------------------------------
## dd$south: 1
## [1] 911.8551
```

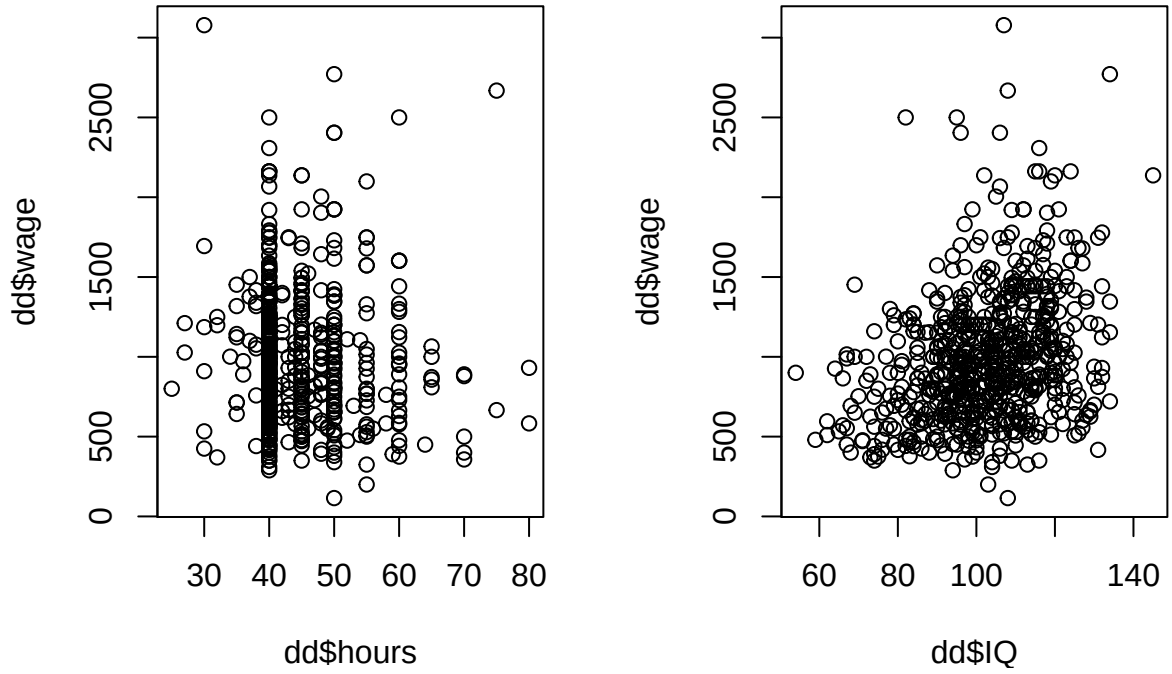
```
by(dd$wage,dd$brthord,mean)#calculate mean of wage grouping by brthord
```

```
## dd$brthord: 1
## [1] 1050.608
## ------------------------------------------------------------
## dd$brthord: 2
## [1] 959.544
## ------------------------------------------------------------
## dd$brthord: 3
## [1] 992.5773
## ------------------------------------------------------------
## dd$brthord: 4
## [1] 895.875
## ------------------------------------------------------------
## dd$brthord: 5
## [1] 913.6667
## ------------------------------------------------------------
## dd$brthord: 6
```

```
## [1] 728.8889
## ------------------------------------------------------------
## dd$brthord: 7
## [1] 878.5714
## ------------------------------------------------------------
## dd$brthord: 8
## [1] 777.25
## ------------------------------------------------------------
## dd$brthord: 10
## [1] 838
```

compare image

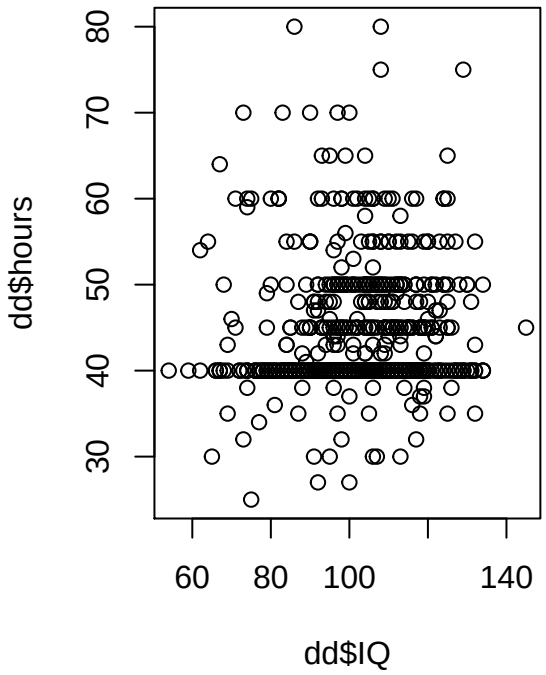
```
par(mfrow=c(1,2))
plot(dd$hours,dd$wage)
plot(dd$IQ,dd$wage)
```




```
median(x = dd$IQ)
```

```
## [1] 104
```

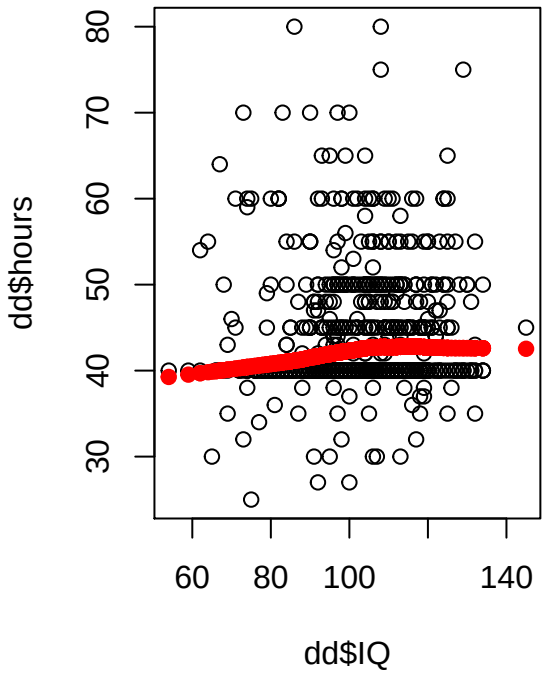
```
plot(dd$IQ,dd$hours)
par(mfrow=c(1,1))
```

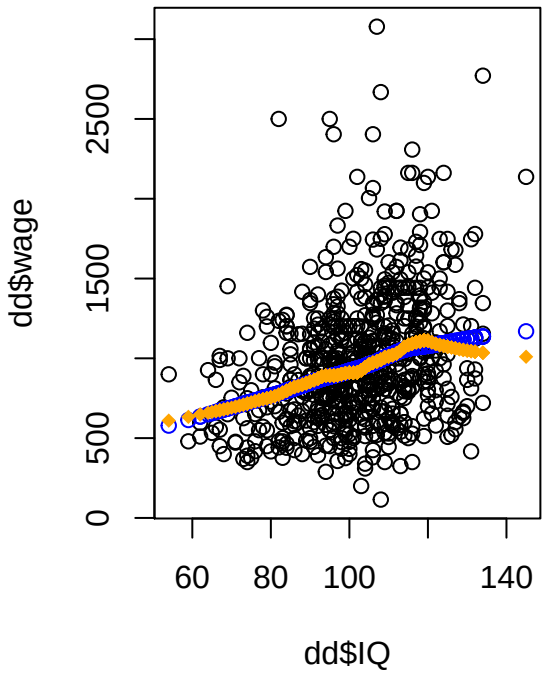



```
par(mfrow=c(1,2))
plot(dd$IQ,dd$hours)
points(lowess(dd$IQ,dd$hours),col="red",pch=19)
cor.test(dd$IQ,dd$hours)
```

```
##
##  Pearson's product-moment correlation
##
## data:  dd$IQ and dd$hours
## t = 1.0562, df = 661, p-value = 0.2913
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.03520676  0.11682578
## sample estimates:
##       cor
## 0.0410471
```

```
#
plot(dd$IQ,dd$wage)
points(lowess(dd$IQ,dd$wage),col="blue",pch=21)
points(lowess(dd$IQ,dd$wage,f=1/3),col="orange",pch=18)
```






```
cor.test(dd$IQ,dd$wage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  dd$IQ and dd$wage
## t = 8.3218, df = 661, p-value = 4.966e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2373721 0.3752936
## sample estimates:
##       cor
## 0.3079499
```

forloop

```
t.test(dd$wage,mu=950)
```

```
##
##  One Sample t-test
##
## data:  dd$wage
## t = 2.437, df = 662, p-value = 0.01507
## alternative hypothesis: true mean is not equal to 950
## 95 percent confidence interval:
##   957.4753 1019.4749
## sample estimates:
## mean of x
##  988.4751
```

```
t.test(dd$wage,dd$hours)
```

```
##
##  Welch Two Sample t-test
##
## data:  dd$wage and dd$hours
## t = 59.811, df = 662.41, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  913.4087 975.4178
## sample estimates:
## mean of x mean of y
## 988.47511  44.06184
```

```
t.test(dd$wage[dd$brthord==1],mu=900)
```

```
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == 1]
## t = 5.9302, df = 277, p-value = 8.991e-09
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  1000.612 1100.603
## sample estimates:
## mean of x
##  1050.608
```

```
max(dd$brthord)
```

```
## [1] 10
```

```
i=1
for(i in c(1,2,3,4,5,6,7,8,10)){print (t.test(dd$wage[dd$brthord==i],mu=900))}
```

```
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = 5.9302, df = 277, p-value = 8.991e-09
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  1000.612 1100.603
## sample estimates:
## mean of x
##  1050.608
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = 2.114, df = 192, p-value = 0.03581
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   903.9876 1015.1005
## sample estimates:
## mean of x
##   959.544
```

```
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = 2.1487, df = 96, p-value = 0.03417
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   907.0552 1078.0995
## sample estimates:
## mean of x
##  992.5773
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -0.082374, df = 39, p-value = 0.9348
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  794.5859 997.1641
## sample estimates:
## mean of x
##   895.875
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = 0.16787, df = 23, p-value = 0.8681
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   745.2565 1082.0768
## sample estimates:
## mean of x
##  913.6667
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -2.4888, df = 17, p-value = 0.02348
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  583.8315 873.9463
## sample estimates:
## mean of x
##  728.8889
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -0.1805, df = 6, p-value = 0.8627
```

```
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   588.0757 1169.0672
## sample estimates:
## mean of x
##  878.5714
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -1.2179, df = 3, p-value = 0.3103
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   456.4869 1098.0131
## sample estimates:
## mean of x
##    777.25
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -0.31633, df = 1, p-value = 0.805
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  -1652.416  3328.416
## sample estimates:
## mean of x
##       838
```

```
for(i in c(1,2,3,4,5,6,7,8,10)){temp=t.test(dd$wage[dd$brthord==i],mu=900)
print (temp)}
```

```
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = 5.9302, df = 277, p-value = 8.991e-09
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  1000.612 1100.603
## sample estimates:
## mean of x
##  1050.608
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = 2.114, df = 192, p-value = 0.03581
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   903.9876 1015.1005
## sample estimates:
```

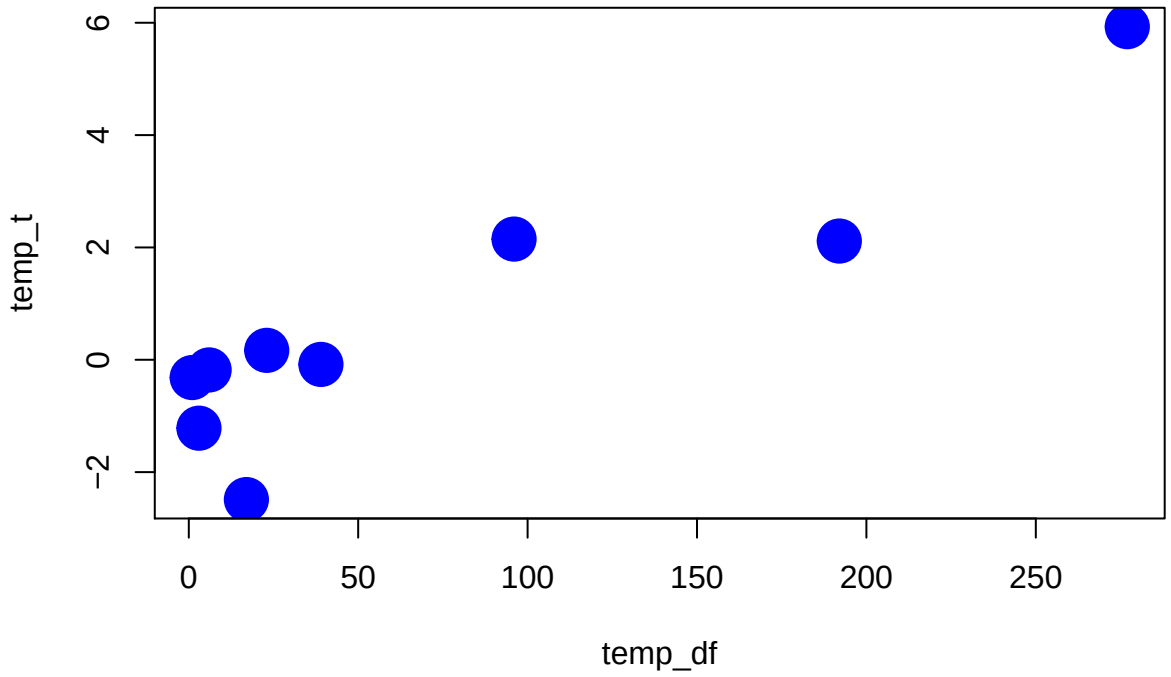
```
## mean of x
##   959.544
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = 2.1487, df = 96, p-value = 0.03417
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   907.0552 1078.0995
## sample estimates:
## mean of x
##  992.5773
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -0.082374, df = 39, p-value = 0.9348
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  794.5859 997.1641
## sample estimates:
## mean of x
##   895.875
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = 0.16787, df = 23, p-value = 0.8681
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   745.2565 1082.0768
## sample estimates:
## mean of x
##  913.6667
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -2.4888, df = 17, p-value = 0.02348
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  583.8315 873.9463
## sample estimates:
## mean of x
##  728.8889
##
##
##  One Sample t-test
##
```

```
## data:  dd$wage[dd$brthord == i]
## t = -0.1805, df = 6, p-value = 0.8627
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   588.0757 1169.0672
## sample estimates:
## mean of x
##  878.5714
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -1.2179, df = 3, p-value = 0.3103
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##   456.4869 1098.0131
## sample estimates:
## mean of x
##    777.25
##
##
##  One Sample t-test
##
## data:  dd$wage[dd$brthord == i]
## t = -0.31633, df = 1, p-value = 0.805
## alternative hypothesis: true mean is not equal to 900
## 95 percent confidence interval:
##  -1652.416  3328.416
## sample estimates:
## mean of x
##       838
```

```
temp_t=0
temp_estimate=0
temp_df=0
temp_pvalue=0

for(i in c(1,2,3,4,5,6,7,8,10)){temp=t.test(dd$wage[dd$brthord==i],mu=900)

temp_t[i]=temp$statistic
temp_estimate[i]=temp$estimate
temp_df[i]=temp$parameter
temp_pvalue[i]=temp$p.value}

table(dd$brthord)
```

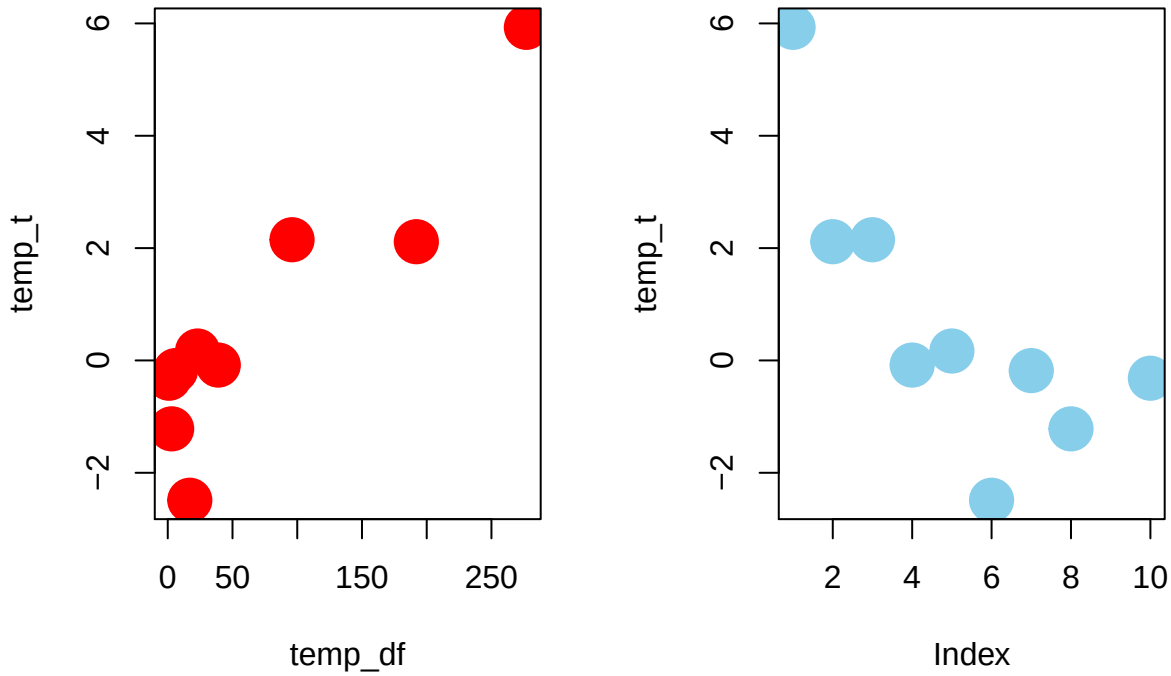
```
##
##   1   2   3   4   5   6   7   8  10
## 278 193  97  40  24  18   7   4   2
```

```
plot(temp_df,temp_t,pch=19,cex=3,col="blue")
```



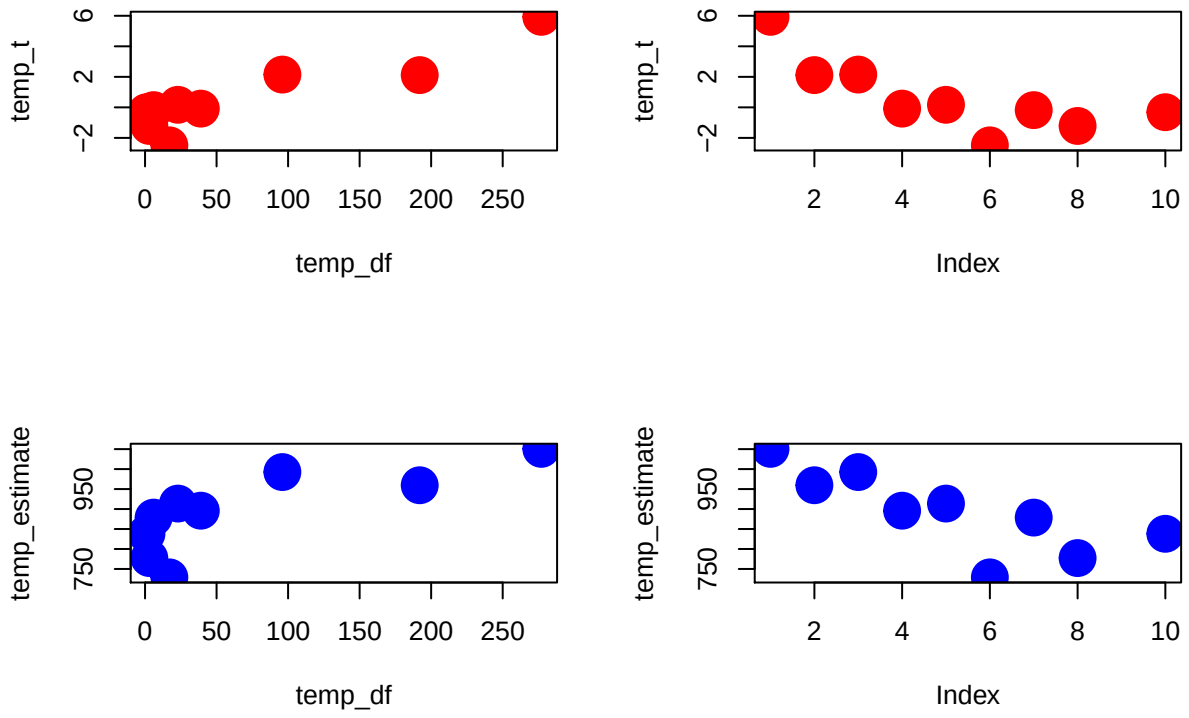
trials

```
par(mfrow=c(1,2))
plot(temp_df,temp_t,pch=19,cex=3,col="red")
plot(temp_t,pch=19,cex=3,col="skyblue")
```




```
mean_birthord=0
for(i in 1:max(dd$brthord)){mean_birthord[i]=mean(dd$brthord[dd$brthord==i])}
```

```
par(mfrow=c(2,2))
plot(temp_df,temp_t,pch=19,cex=3,col="red")
plot(temp_t,pch=19,cex=3,col="red")
plot(temp_df,temp_estimate,pch=19,cex=3,col="blue")
plot(temp_estimate,pch=19,cex=3,col="blue")
```




```
plot(jitter(dd$brthord,factor=2),dd$wage)
points(lowess(dd$brthord,dd$wage),pch=19,col="pink",cex=3)

hist(dd$brthord)

plot(dd$brthord,dd$wage)
cor.test(dd$brthord,dd$wage)
```

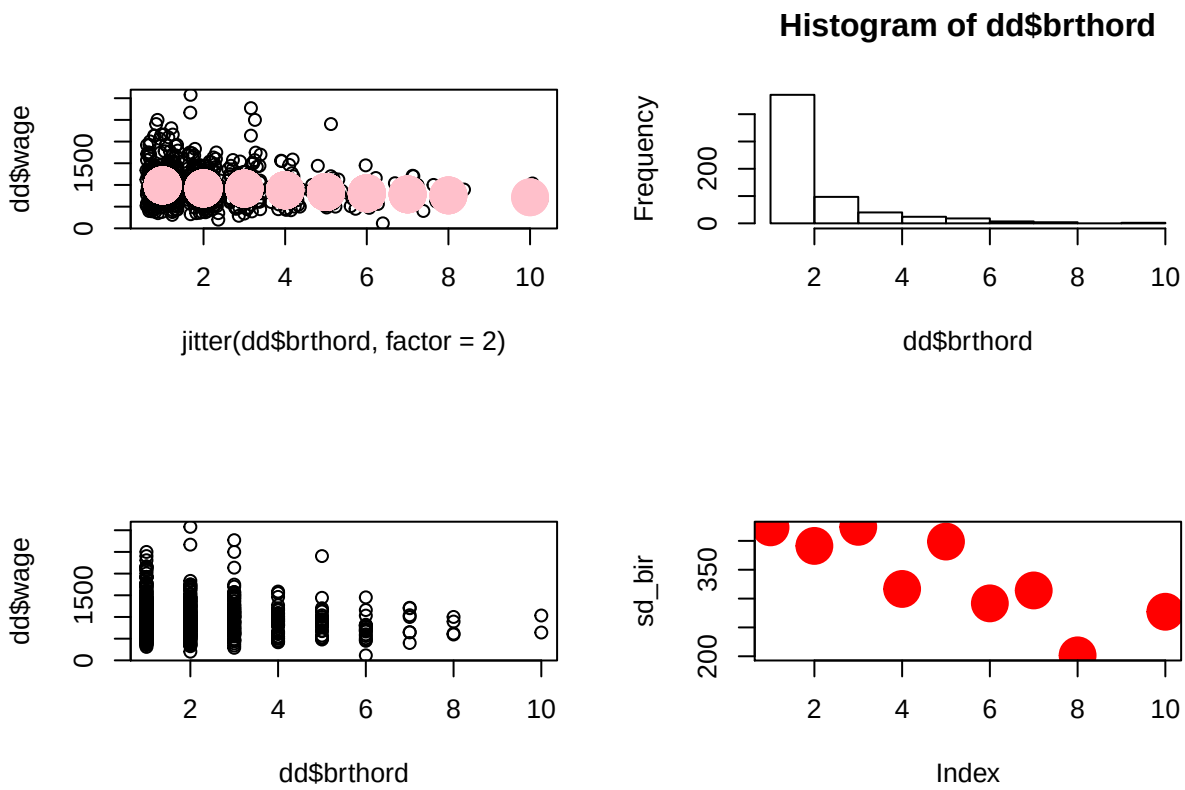
```
##
##  Pearson's product-moment correlation
##
## data:  dd$brthord and dd$wage
## t = -3.8763, df = 661, p-value = 0.0001167
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.22269965 -0.07377756
## sample estimates:
##        cor
## -0.1490839
```

```
sd_bir=0
for(i in c(1,2,3,4,5,6,7,8,10)){sd_bir[i]=sd(dd$wage[dd$brthord==i])
print (sd_bir[i])}
```

```
## [1] 423.4507
## [1] 391.3079
```

```
## [1] 424.3337
## [1] 316.7112
## [1] 398.8276
## [1] 291.6968
## [1] 314.1018
## [1] 201.5827
## [1] 277.1859
```

```
plot(sd_bir,pch=19,cex=3,col="red")
```




```
plot(dd$IQ,dd$brthord)
```