takeover_time_analysis

Oct 23, 2019

Takeover time versus lead time

```
setwd("C://Doc//resume//apply//TOPS")
require(ggplot2)
```

```
## Loading required package: ggplot2
```

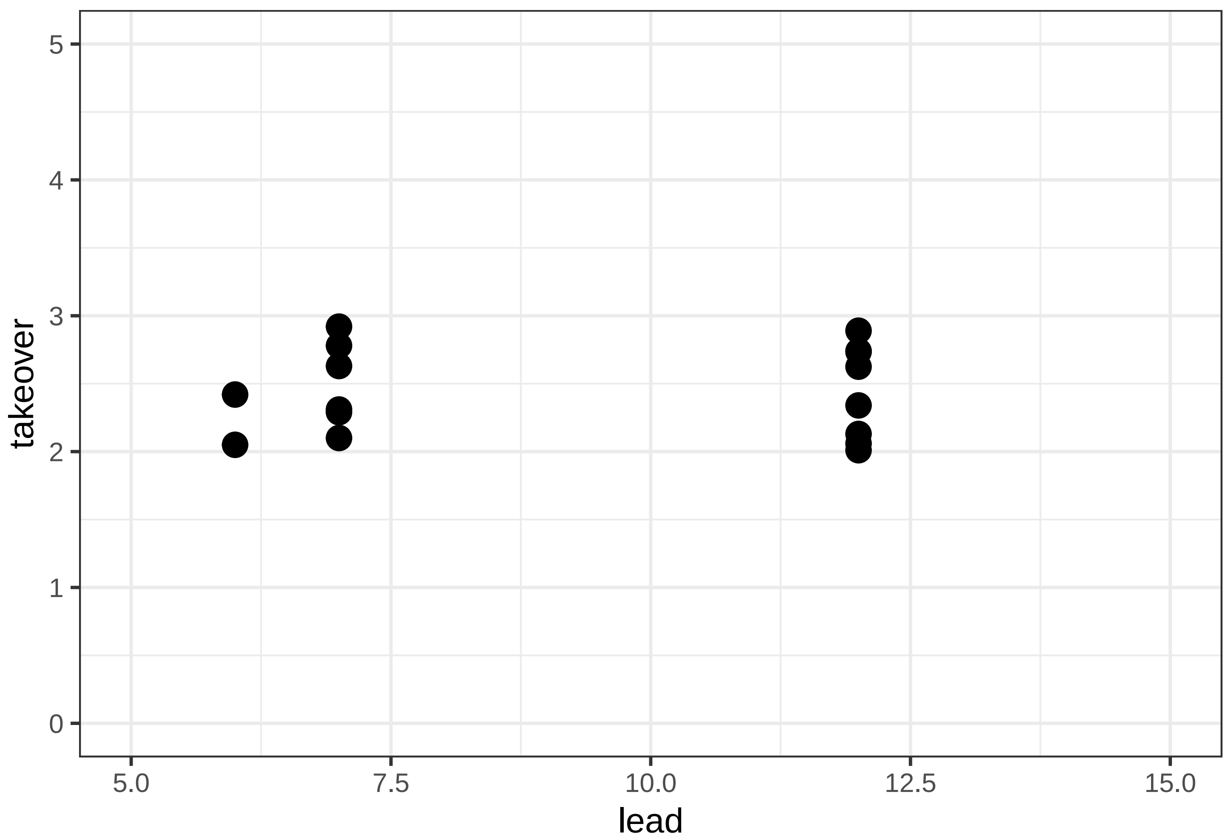
```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
raw <- read.csv("takeover_time.csv", header = T)
raw=raw[-73,] #Remove outliers
```

Here 73rd sample point with 42s takeover time seems suspecious, thus it's removed.

```
takeover_time = sapply(1:nrow(raw), function(x)
  ifelse(is.na(raw$Takeover_Time[x]), mean(unlist(raw[x,c("Braking","Steering")]), na.rm = T), raw$Takeo
ggplot(data.frame(lead=raw$Lead_time, takeover=takeover_time), aes(x = lead, y = takeover))+geom_point(s
```

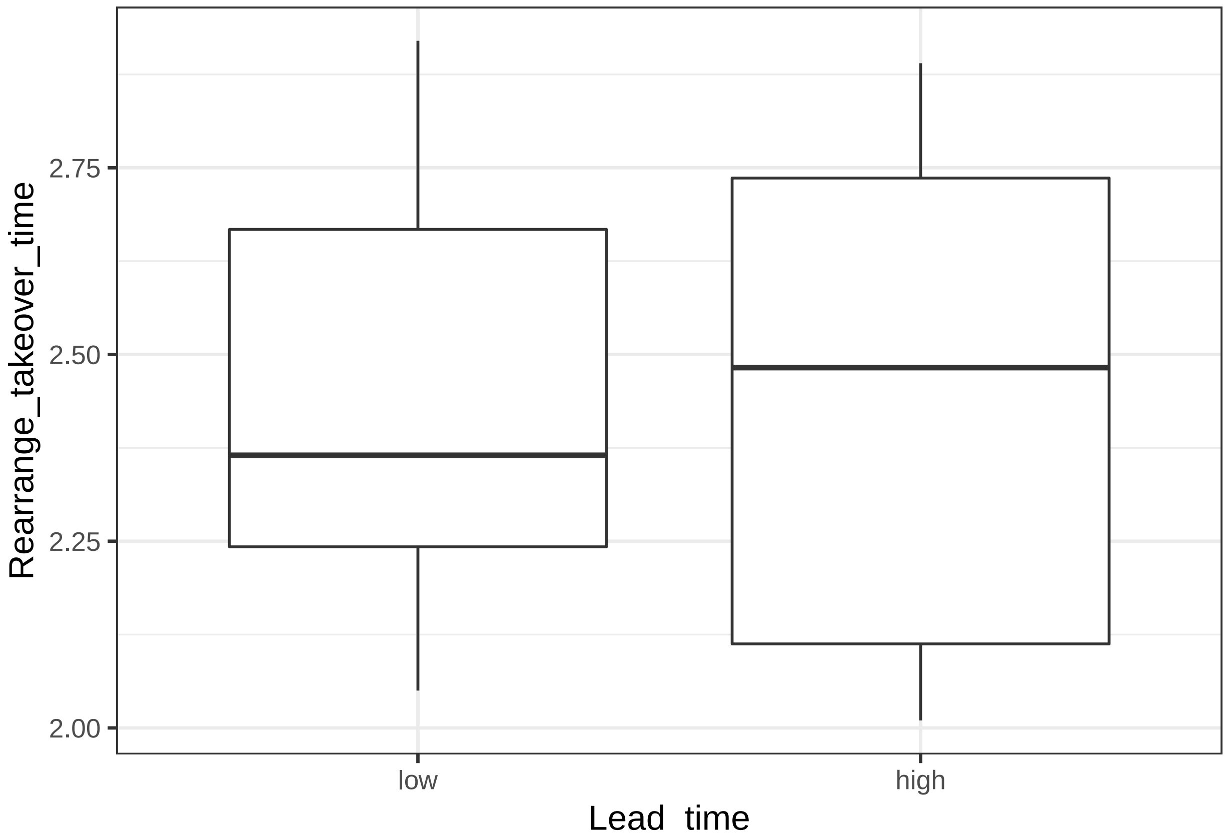
```
## Warning: Removed 106 rows containing missing values (geom_point).
```



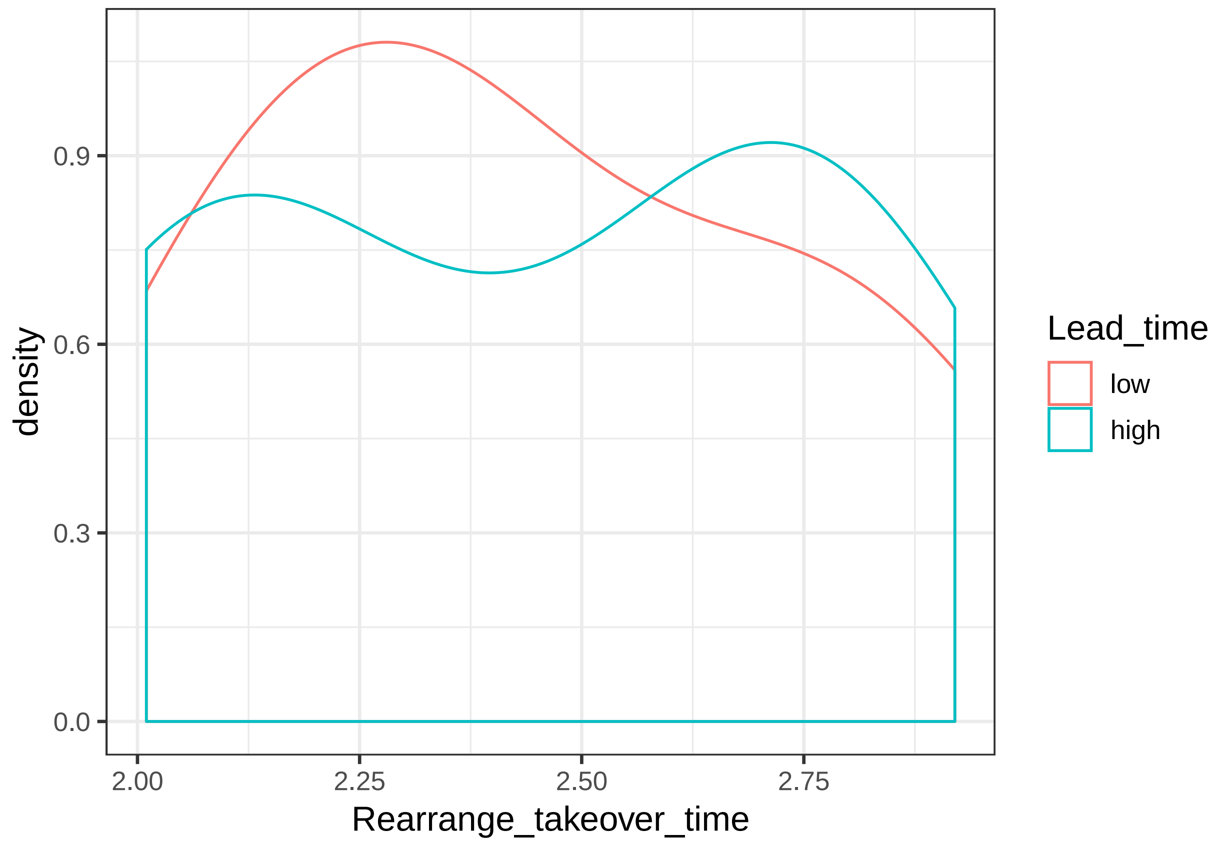

```
raw[,"Rearrange_takeover_time"] <- takeover_time
```

Here we recalculate the takeover time for more observations. If the takeover time is recorded, it won't be changed. Otherwise, replace it by the average value of steering time and Braking time, ignoring the missing value. The scatter plot shows we still does not have enough observations for regression, especially the lead time. We only have 3 different lead time. Next we categorize the lead time 6 and 7 into low-level and 12 into high-level. Check the density plot first.

```
lead_vs_takeover <- raw[,c("Lead_time","Rearrange_takeover_time")]
lead_vs_takeover <- lead_vs_takeover[apply(!is.na(lead_vs_takeover), 1,FUN = function(x) all(unlist(x))
lead_vs_takeover$Lead_time[1:2] <- c(7,7)
lead_vs_takeover$Lead_time <- factor(lead_vs_takeover$Lead_time,labels=c("low","high"))
ggplot(lead_vs_takeover, aes(x=Lead_time,y=Rearrange_takeover_time))+geom_boxplot()+theme_bw(base_size=
```




```
ggplot(lead_vs_takeover, aes(x=Rearrange_takeover_time,color=Lead_time))+geom_density()+theme_bw(base_s
```



obviously they have different peaks. It may support my assumption before. It's interesting that the box plot indicates that the high lead time group have greater deviation. I guess drivers have no choice but stopping the secondary task at once when the lead time is short. But drivers in the other group have more time to decide how to tackle with it. Consequently, they have various reaction. low level variance: 0.0985, high level variance: 0.121.

two sample t test

```
shapiro.test(lead_vs_takeover$Rearrange_takeover_time[lead_vs_takeover$Lead_time=="low"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lead_vs_takeover$Rearrange_takeover_time[lead_vs_takeover$Lead_time ==     "low"]
## W = 0.94502, p-value = 0.661
```

```
shapiro.test(lead_vs_takeover$Rearrange_takeover_time[lead_vs_takeover$Lead_time=="high"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lead_vs_takeover$Rearrange_takeover_time[lead_vs_takeover$Lead_time ==     "high"]
## W = 0.89051, p-value = 0.2366
```

```
t.test(Rearrange_takeover_time~Lead_time,data=lead_vs_takeover)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  Rearrange_takeover_time by Lead_time
## t = -0.022623, df = 13.851, p-value = 0.9823
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3596302  0.3521302
## sample estimates:
##  mean in group low mean in group high
##            2.43750            2.44125
```

The two sample t test shows there is no significant difference between them. But it doestnot contradict to the result from desity plot since t test take the variance into consideration.

Add more variables

```
ancova_data <- raw[,c("Lead_time","Rearrange_takeover_time","Mean_age","Modality","NDRT")]
ancova_data <- ancova_data[apply(!is.na(ancova_data), 1,FUN = function(x) all(unlist(x))),]
ancova_data
```

```
##    Lead_time Rearrange_takeover_time Mean_age Modality NDRT
## 2          6                   2.050    44.60        A    N
## 3          6                   2.420    44.60        V    N
## 4         12                   2.890    25.70       HV    P
## 5         12                   2.735    25.70       HV    P
## 6         12                   2.740    25.70       HV    P
## 7         12                   2.625    25.70       HV    P
## 8         12                   2.010    25.70       HV    P
## 9         12                   2.340    25.70       HV    P
## 10        12                   2.060    25.70       HV    P
## 11        12                   2.130    25.70       HV    P
## 13         7                   2.310    30.10      AHV    M
## 14         7                   2.290    30.10      AHV    M
## 15         7                   2.630    30.10      AHV    M
## 16         7                   2.100    30.10      AHV    M
## 17         7                   2.920    31.36       AV    E
## 18         7                   2.780    31.36       AV    E
```

The category of lead time and other factors are really similar. Thus we can't do regression or ANCOVA. We have to remove lead time.

```
anova_data <- raw[,c("Rearrange_takeover_time","Modality","NDRT")]
anova_data <- anova_data[apply(!(is.na(anova_data)|anova_data==""), 1,FUN = function(x) all(unlist(x)))
print(anova_data)
```

```
##    Rearrange_takeover_time Modality NDRT
## 2                    2.050        A    N
## 3                    2.420        V    N
## 4                    2.890       HV    P
## 5                    2.735       HV    P
## 6                    2.740       HV    P
## 7                    2.625       HV    P
## 8                    2.010       HV    P
## 9                    2.340       HV    P
## 10                   2.060       HV    P
```

```
## 11                    2.130       HV    P
## 13                    2.310      AHV    M
## 14                    2.290      AHV    M
## 15                    2.630      AHV    M
## 16                    2.100      AHV    M
## 17                    2.920       AV    E
## 18                    2.780       AV    E
## 19                    2.770       AV    M
## 20                    3.180       AV    M
## 21                    3.210       AV    M
## 22                    3.480       AV    M
## 23                    4.580       AV    M
## 24                    5.500       AV    M
## 25                    4.510       AV    M
## 26                    5.420       AV    M
## 27                    3.230       AV    M
## 28                    4.010       AV    M
## 29                    2.940       AV    E
## 30                    6.450       AV    E
## 31                    3.400       AV    E
## 32                    4.200       AV    E
```

Modality

Haptic

```
lapply(split(anova_data[11:30,1],anova_data[11:30,2]),function(x) {if(length(x)!=0){shapiro.test(x)$p.va
```

```
## [[1]]
## NULL
##
## $A
## NULL
##
## $AHV
## [1] 0.6098569
##
## $AV
## [1] 0.03805787
##
## $HV
## NULL
##
## $V
## NULL
```

```
bartlett.test(Rearrange_takeover_time~Modality,data=anova_data[11:30,])
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Rearrange_takeover_time by Modality
## Bartlett's K-squared = 5.9142, df = 1, p-value = 0.01502
```

The shapiro test shows it doesnot satisfy the normality assumption. And bartlett test shows it doesnot satisfy the homoskedastcity assumption. Thus we tried the logrithm transformation.

```
anova_data$Rearrange_takeover_time <- log(anova_data$Rearrange_takeover_time)
lapply(split(anova_data[11:30,1],anova_data[11:30,2]),function(x) {if(length(x)!=0){shapiro.test(x)$p.va
```

```
## [[1]]
## NULL
##
## $A
## NULL
##
## $AHV
## [1] 0.6769054
##
## $AV
## [1] 0.1425158
##
## $HV
## NULL
##
## $V
## NULL
```

```
bartlett.test(Rearrange_takeover_time~Modality,data=anova_data[11:30,])
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Rearrange_takeover_time by Modality
## Bartlett's K-squared = 3.1077, df = 1, p-value = 0.07792
```

```
m1 <- aov(Rearrange_takeover_time~Modality,data=anova_data[11:30,])
summary(m1)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Modality     1 0.7546  0.7546   12.52 0.00235 **
## Residuals   18 1.0847  0.0603
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

And the ANOVA table shows haptic warning is significant under 0.99 confidence level. It indicates that Haptic warning will decrease the takeover time significantly.

Audio

```
lapply(split(anova_data[3:14,1],anova_data[3:14,2]),function(x) {if(length(x)!=0){shapiro.test(x)$p.valu
```

```
## [[1]]
## NULL
##
## $A
## NULL
##
## $AHV
## [1] 0.6769054
```

```
##
## $AV
## NULL
##
## $HV
## [1] 0.2178439
##
## $V
## NULL
```

```
bartlett.test(Rearrange_takeover_time~Modality,data=anova_data[3:14,])
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Rearrange_takeover_time by Modality
## Bartlett's K-squared = 0.63998, df = 1, p-value = 0.4237
```

```
m2 <- aov(Rearrange_takeover_time~Modality,data=anova_data[3:14,])
summary(m2)
```

```
##             Df  Sum Sq  Mean Sq F value Pr(>F)
## Modality     1 0.00422 0.004216   0.245  0.631
## Residuals   10 0.17216 0.017216
```

The audio warning is not significant under 0.95 confidence level. Thus it won't have great influence on the takeover time.

NDRT

```
t.test(anova_data$Rearrange_takeover_time[anova_data$NDRT=="N"],anova_data$Rearrange_takeover_time[anova
```

```
##
##  Welch Two Sample t-test
##
## data:  anova_data$Rearrange_takeover_time[anova_data$NDRT == "N"] and anova_data$Rearrange_takeover_t
## t = -0.84803, df = 1.865, p-value = 0.4912
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5323895  0.3671117
## sample estimates:
## mean of x mean of y
## 0.8008037 0.8834425
```

```
t.test(anova_data$Rearrange_takeover_time[anova_data$NDRT=="N"],anova_data$Rearrange_takeover_time[anova
```

```
##
##  Welch Two Sample t-test
##
## data:  anova_data$Rearrange_takeover_time[anova_data$NDRT == "N"] and anova_data$Rearrange_takeover_t
## t = -3.0985, df = 5.4528, p-value = 0.02398
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.87165552 -0.09186522
## sample estimates:
```

```
## mean of x mean of y
## 0.8008037 1.2825640
```

```
t.test(anova_data$Rearrange_takeover_time[anova_data$NDRT=="N"],anova_data$Rearrange_takeover_time[anova
```

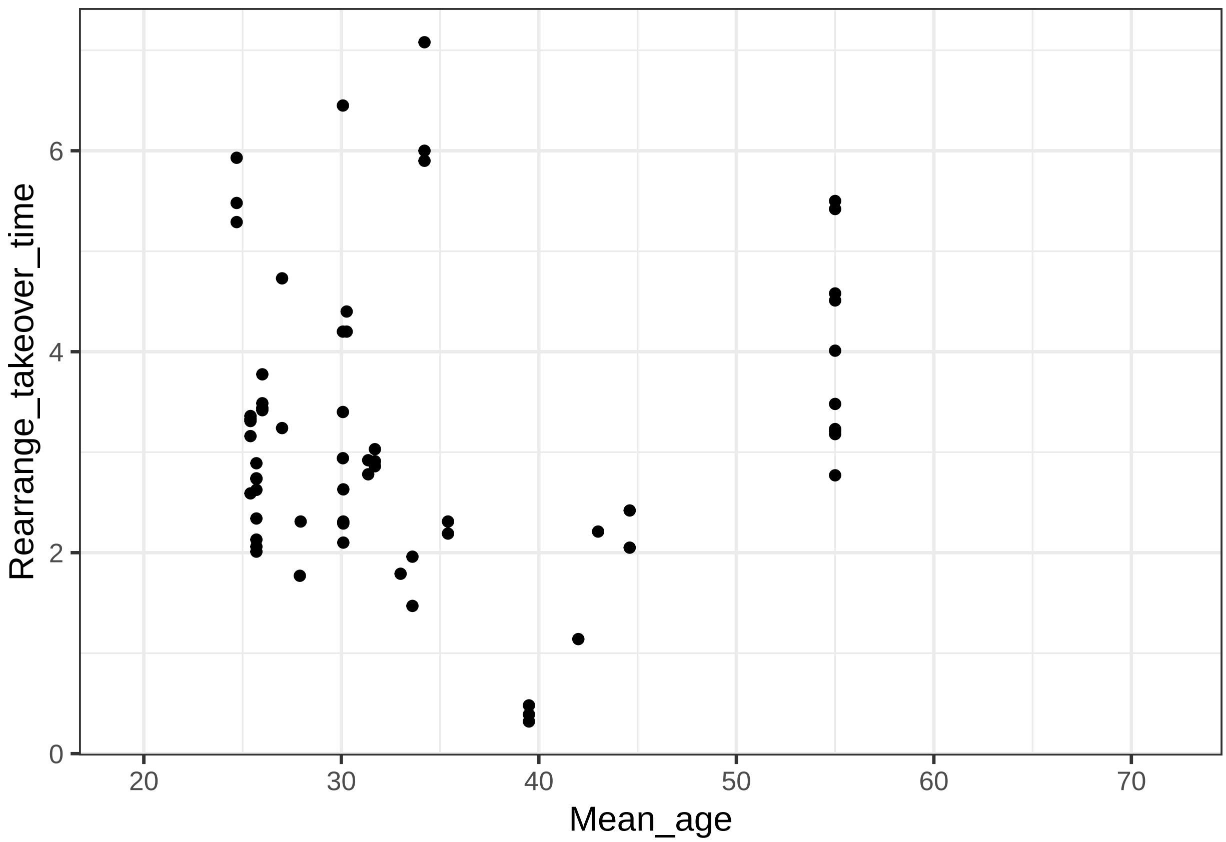
```
##
##  Welch Two Sample t-test
##
## data:  anova_data$Rearrange_takeover_time[anova_data$NDRT == "N"] and anova_data$Rearrange_takeover_t
## t = -3.4607, df = 3.8317, p-value = 0.02763
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.74427080 -0.07524554
## sample estimates:
## mean of x mean of y
## 0.8008037 1.2105618
```

The two sample t-test indicates that `eye distraction task` and `miscellaneous task` will significantly increase the takeover time compared with the no-secondary task test. But `hysical movement related` seems has little influence on it.

Others

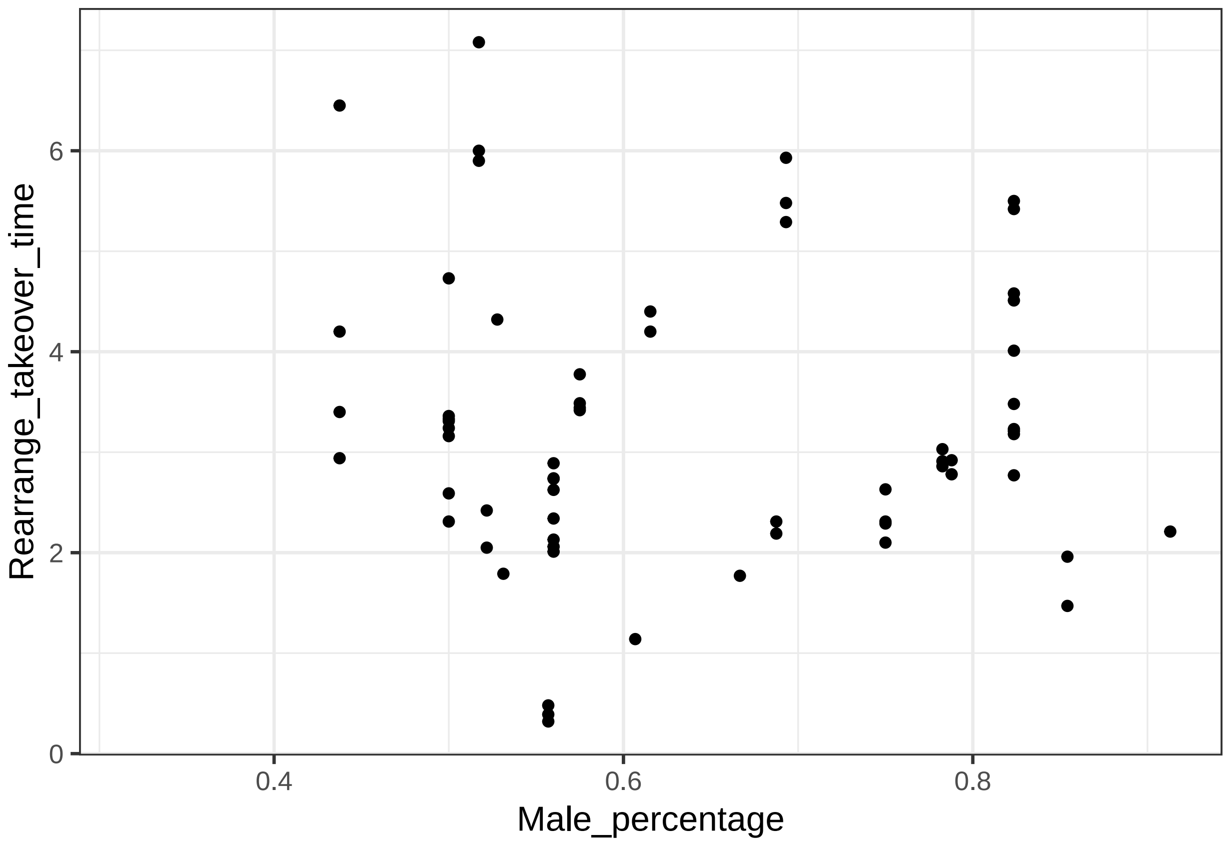
```
raw[,"Male_percentage"] <- raw[,"Male"]/(raw[,'Female']+raw[,'Male'])
ggplot(data=raw,aes(x=Mean_age,y=Rearrange_takeover_time))+geom_point()+theme_bw(base_size=13)+theme(te
```

```
## Warning: Removed 58 rows containing missing values (geom_point).
```




```
ggplot(data=raw,aes(x=Male_percentage,y=Rearrange_takeover_time))+geom_point()+theme_bw(base_size=13)+tl
```

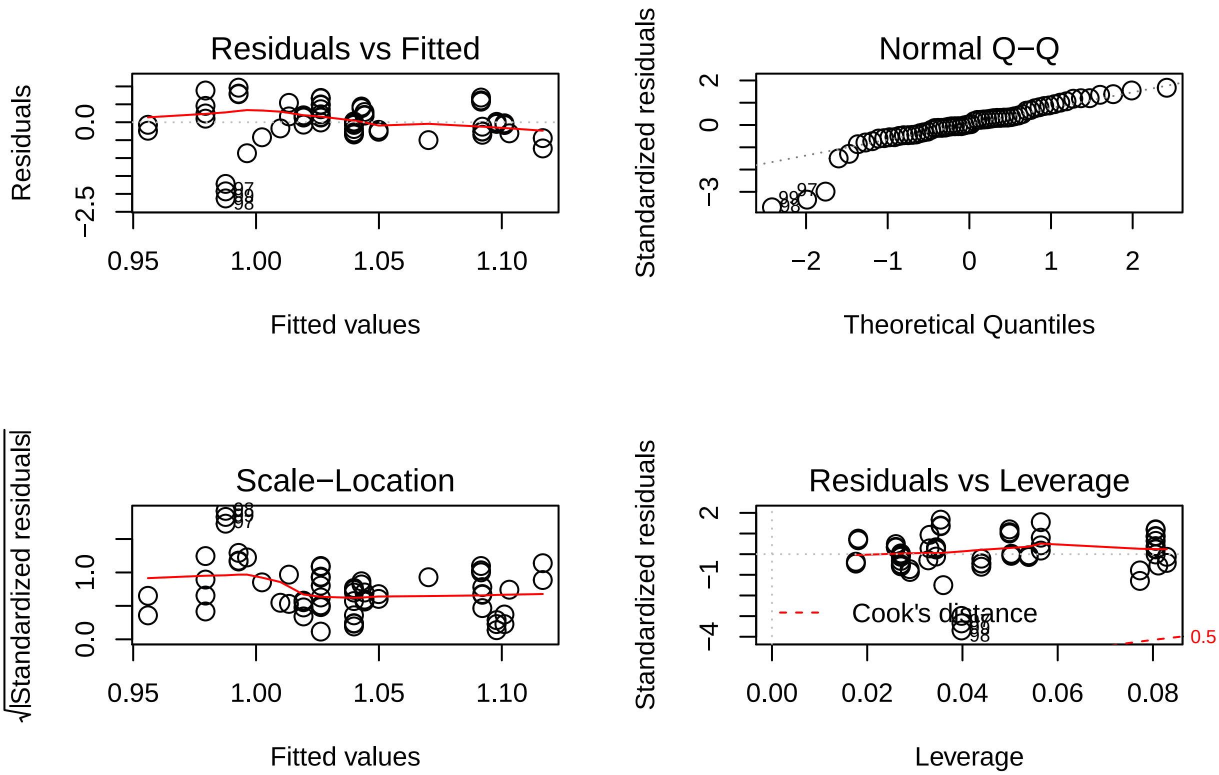
```
## Warning: Removed 57 rows containing missing values (geom_point).
```




```
m<-lm(log(Rearrange_takeover_time)~Male_percentage+Mean_age,data=raw)
summary(m)
```

```
##
## Call:
## lm(formula = log(Rearrange_takeover_time) ~ Male_percentage +
##     Mean_age, data = raw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.12704 -0.24233  0.01605  0.30382  0.96439
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.933076   0.360756   2.586   0.0121 *
## Male_percentage  0.360664   0.668750   0.539   0.5916
## Mean_age        -0.003705   0.008853  -0.418   0.6771
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5875 on 61 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.005015,   Adjusted R-squared:  -0.02761
## F-statistic: 0.1537 on 2 and 61 DF,  p-value: 0.8578
```

```
par(mfrow=c(2,2))
plot(m,cex=1.5)
```



Actually I didnot do much in this section. From the scatter plot we could find that they have no obvious pattern. But I still did the regression. All p-values are insignificant after removing the three suspecious outliers and trasforming the takeover time.

One thing to be concerned

```
anova_data
```

```
##    Rearrange_takeover_time Modality NDRT
## 2                0.7178398        A    N
## 3                0.8837675        V    N
## 4                1.0612565       HV    P
## 5                1.0061314       HV    P
## 6                1.0079579       HV    P
## 7                0.9650809       HV    P
## 8                0.6981347       HV    P
## 9                0.8501509       HV    P
## 10               0.7227060       HV    P
## 11               0.7561220       HV    P
## 13               0.8372475      AHV    M
## 14               0.8285518      AHV    M
## 15               0.9669838      AHV    M
## 16               0.7419373      AHV    M
## 17               1.0715836       AV    E
## 18               1.0224509       AV    E
```

```
## 19                 1.0188473       AV    M
## 20                 1.1568812       AV    M
## 21                 1.1662709       AV    M
## 22                 1.2470323       AV    M
## 23                 1.5216990       AV    M
## 24                 1.7047481       AV    M
## 25                 1.5062972       AV    M
## 26                 1.6900958       AV    M
## 27                 1.1724821       AV    M
## 28                 1.3887912       AV    M
## 29                 1.0784096       AV    E
## 30                 1.8640801       AV    E
## 31                 1.2237754       AV    E
## 32                 1.4350845       AV    E
```

Let's go back to the dataset. The NDRT and Modality has a similar part from No.4 to No.11. When we compare the difference of takeover time with one factor, the result may be distorted by the other one. That is to say, if Audio warning decrease the takeover time while the physical task increase the time, it may just offset the infulence. Here I just ignore this problem. I searched but have no ideal how to deal with it.