490rt

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(funModeling)
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
## funModeling v.1.6.8 :)
## Examples and tutorials at livebook.datascienceheroes.com
```

```
library(Hmisc)
data <- read.csv("rt.csv")
#data = read.csv("/Users/haoqingchen/Desktop/sta490rt/rt.csv")
basic_eda <- function(data)
{
  glimpse(data)
  df_status(data)
  freq(data)
  profiling_num(data)
  plot_num(data)
  describe(data)
}
```

Analysis of the relationship between sleep and reaction time

For the analysis of sleep, I created 4 plot which are scatter plot, boxplot, barplot, as well as frequency plot. First of all, the scatter plot does not show a linear relationship between the amount of sleep and reaction time

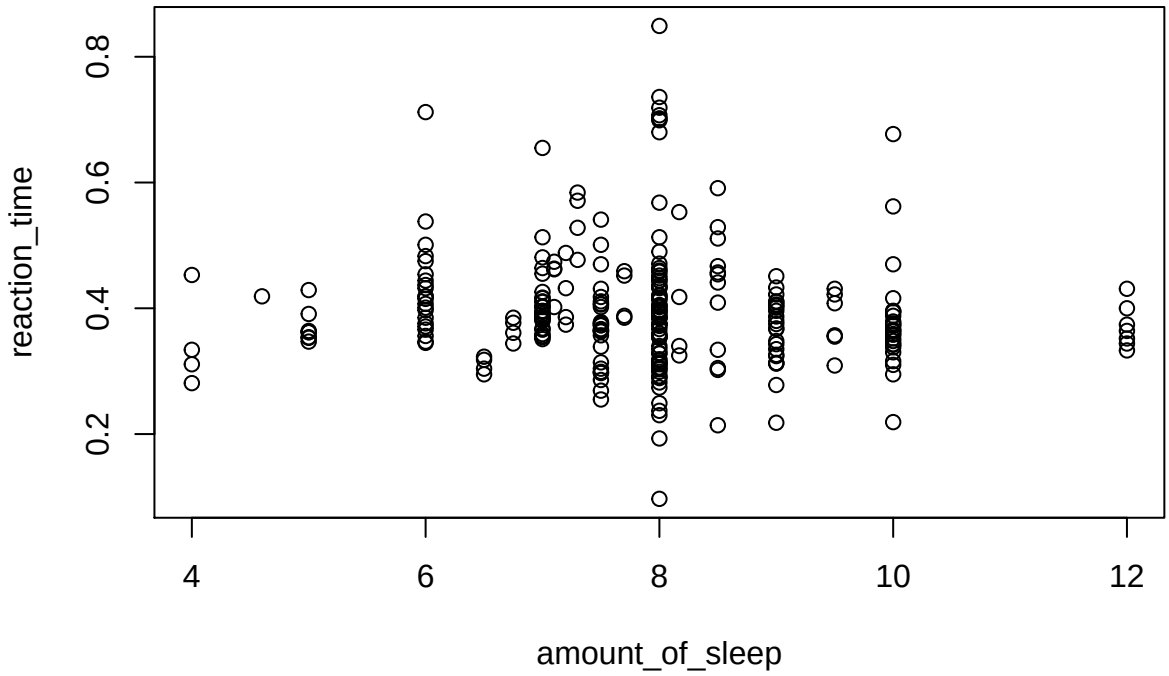
since the plots do not increase/decrease constantly. And the boxplot indicates there are three outliers while median falls in 8 hours of sleep. The next colorful graph is the average reaction time in terms of different amount of sleep. I find it interesting that the the slowest reaction time does not happen in neither shortest nor longest sleep time. The last frequency graph shows that people who sleep 8 hours per day occupy 24.68%.

why is this interesting?

```
amount_of_sleep = data$Sleep
reaction_time = data$RT
mean(data$Sleep, na.rm= T)
```

```
## [1] 7.898109
```

```
plot(amount_of_sleep, reaction_time)
```




```
summary(data$Sleep)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   4.000   7.000   8.000   7.898   8.500  12.000      37
```

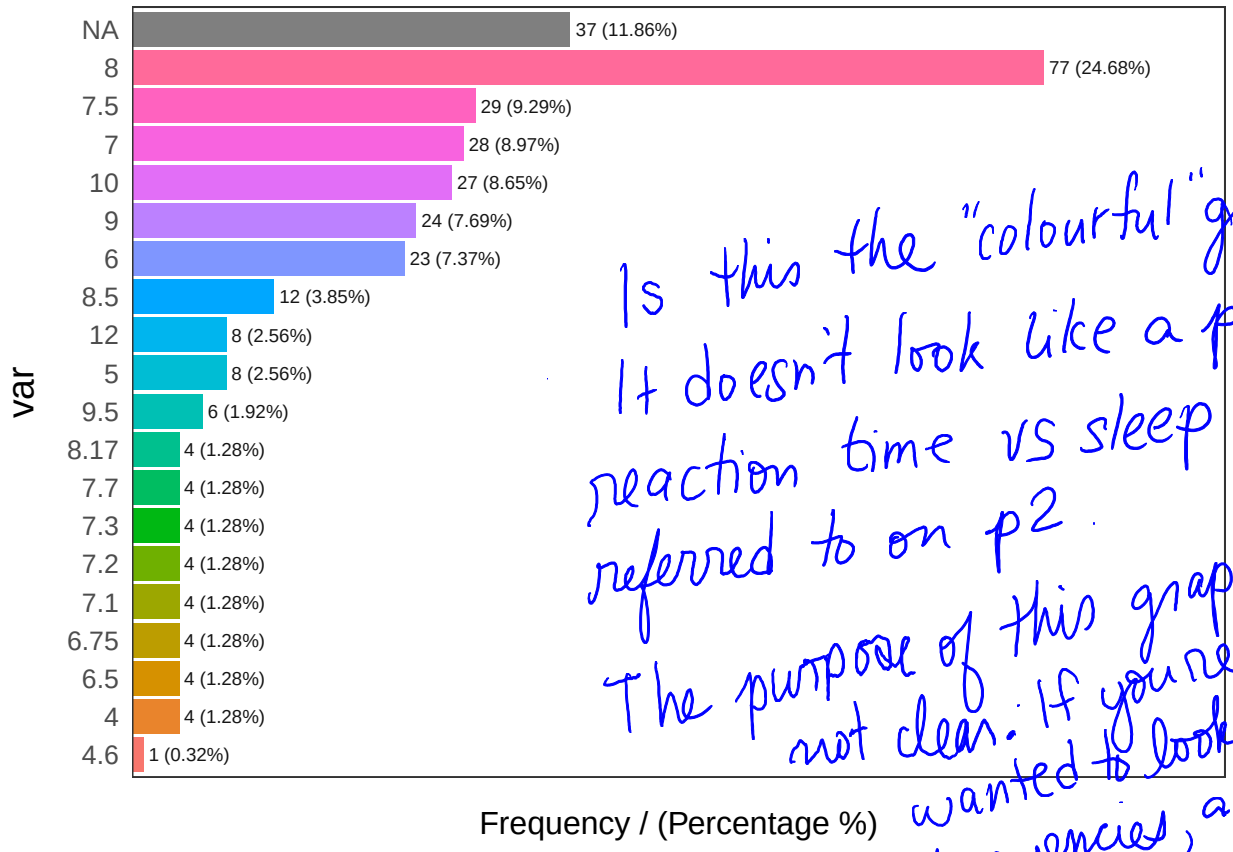
```
par(mfrow = c(1,2))
boxplot((data$Sleep), ylab = "Amount of Sleep")
af <- na.omit(data)
af$Sleep <- as.factor(af$Sleep)
ggplot(summarise(group_by(af, Sleep), RT = mean(RT)),aes(x= Sleep, y = RT)) + geom_bar(stat = "identity
summarise(group_by(af, Sleep), mean(RT))
```

```
## # A tibble: 19 x 2
##    Sleep `mean(RT)`
##    <fct>      <dbl>
##  1 4          0.345
```

```
##  2 4.6        0.419
##  3 5          0.370
##  4 6          0.431
##  5 6.5        0.31
##  6 6.75       0.367
##  7 7          0.409
##  8 7.1        0.450
##  9 7.2        0.42
## 10 7.3        0.54
## 11 7.5        0.369
## 12 7.7        0.421
## 13 8          0.405
## 14 8.17       0.409
## 15 8.5        0.418
## 16 9          0.356
## 17 9.5        0.380
## 18 10         0.378
## 19 12         0.371
```

```
freq(data$Sleep)
```

```
##     var frequency percentage cumulative_perc
## 1     8        77      24.68           24.68
## 2  <NA>        37      11.86           36.54
## 3   7.5        29       9.29           45.83
## 4     7        28       8.97           54.80
## 5    10        27       8.65           63.45
## 6     9        24       7.69           71.14
## 7     6        23       7.37           78.51
## 8   8.5        12       3.85           82.36
## 9     5         8       2.56           84.92
## 10   12         8       2.56           87.48
## 11  9.5         6       1.92           89.40
## 12    4         4       1.28           90.68
## 13  6.5         4       1.28           91.96
## 14 6.75         4       1.28           93.24
## 15  7.1         4       1.28           94.52
## 16  7.2         4       1.28           95.80
## 17  7.3         4       1.28           97.08
## 18  7.7         4       1.28           98.36
## 19 8.17         4       1.28           99.64
## 20  4.6         1       0.32          100.00
```



Is this the "colourful" graph? It doesn't look like a plot of reaction time vs sleep referred to on p2.

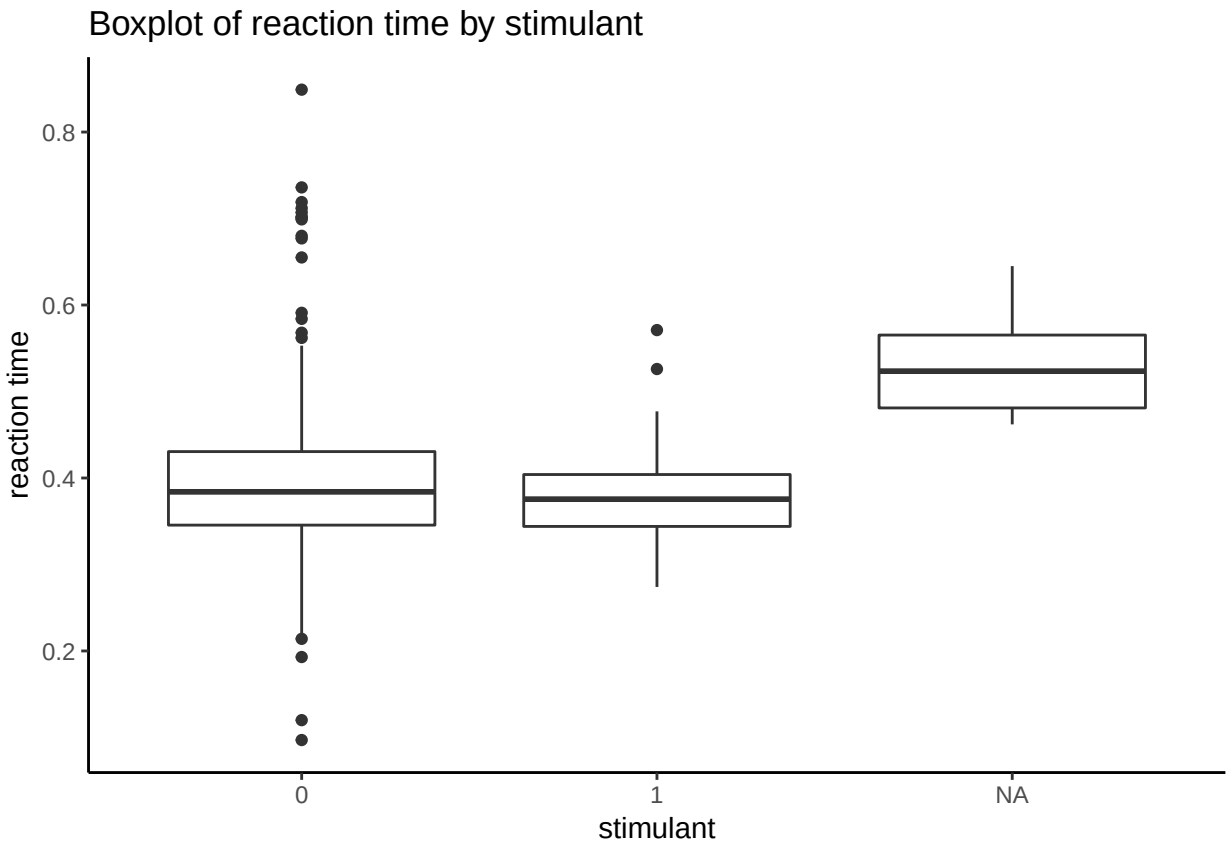
The purpose of this graph is not clear. If you really wanted to look at frequencies, a histogram would be more appropriate for continuous variables

Analysis of stimulant

For stimulant variable, I have the two-side boxplot to compare the mean reaction time value. I did not take out the missing data because they are not useless in this case. I treat them like a reference group. And in frequency plot, it is obvious that most of the people did not use stimulant during the test.

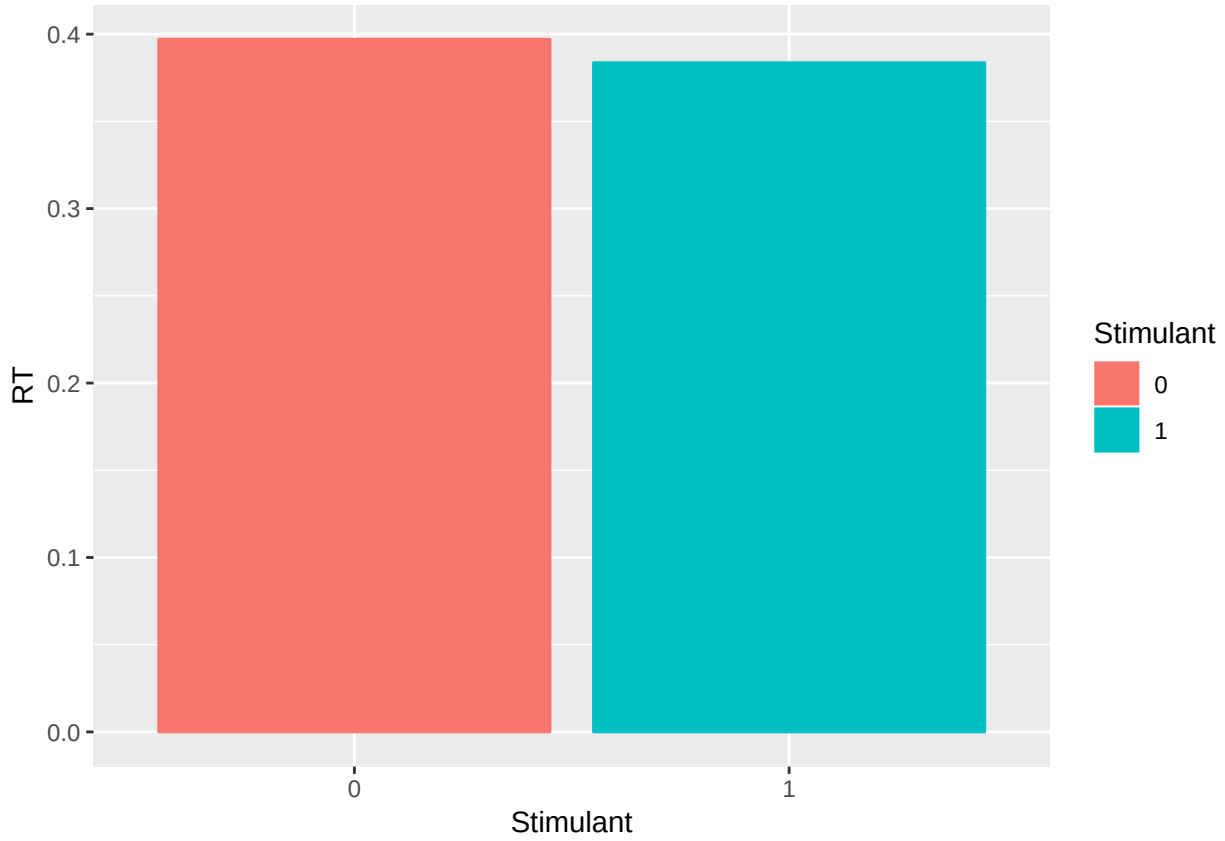
```
stimulant = data$Stimulant
ggplot(data,aes(x = factor(stimulant),y = reaction_time)) +
  theme_classic()  +
  geom_boxplot() +
  labs(title = "Boxplot of reaction time by stimulant",
       x = "stimulant",
       y = "reaction time")
```

```
## Warning: Removed 10 rows containing non-finite values (stat_boxplot).
```




```
af$Stimulant <- as.factor(af$Stimulant)
ggplot(summarise(group_by(af, Stimulant), RT = mean(RT)),aes(x= Stimulant, y = RT)) + geom_bar(stat = "
```

what did you learn from this plot?



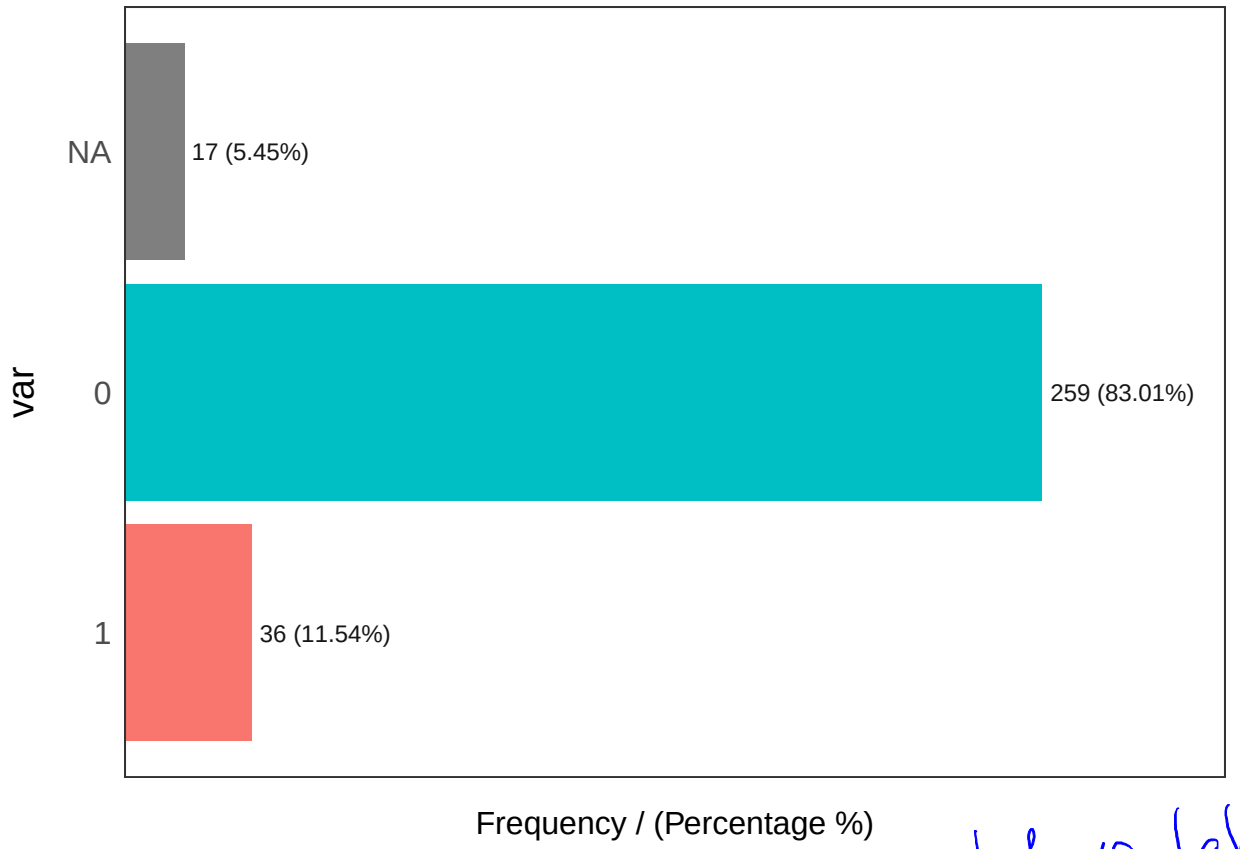

```
summarise(group_by(af, Stimulant), mean(RT))
```

```
## # A tibble: 2 x 2
##   Stimulant `mean(RT)`
##   <fct>           <dbl>
## 1 0               0.397
## 2 1               0.384
```

```
freq(stimulant)
```

Is this a good way to display this data visually?

What is the value of the plot as compared to the table?



plot vs table?

```
##    var frequency percentage cumulative_perc
## 1    0       259      83.01           83.01
## 2    1        36      11.54           94.55
## 3 <NA>        17       5.45          100.00
```

Fatigue

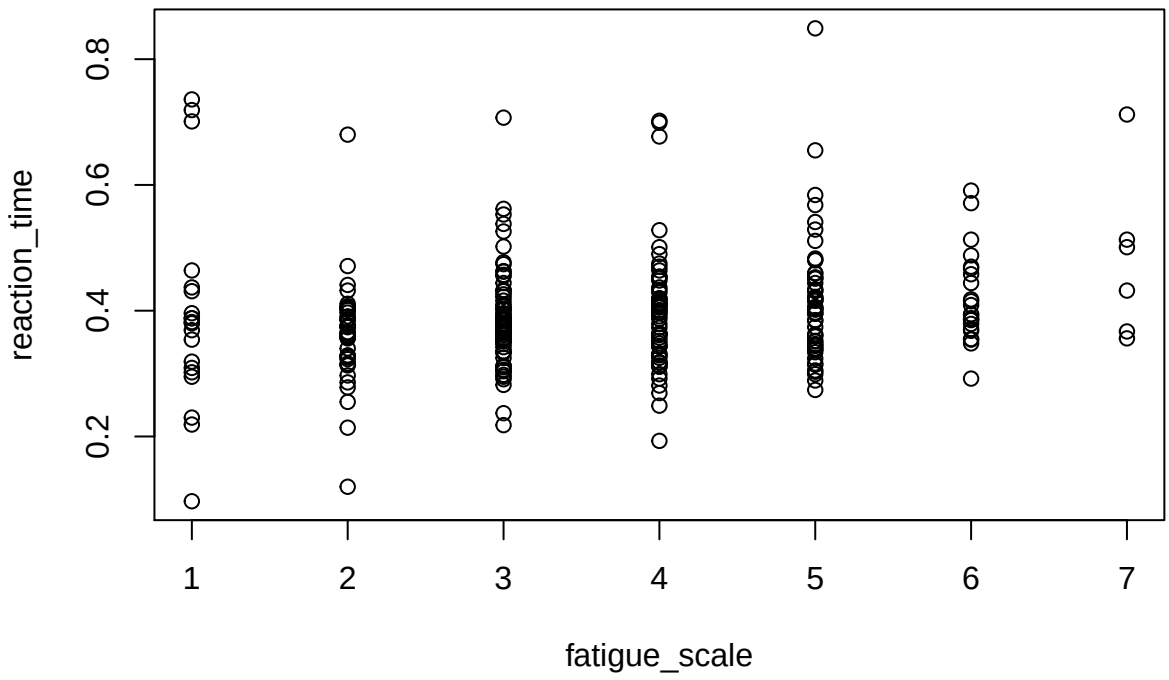
The graph below basically shows the distribution of reaction time in terms of fatigue levels, and I did not exclude the missing data since it does not affect the result very much.

```
fatigue_scale = data$Fatigue
mean(data$Fatigue, na.rm= T)
```

```
## [1] 3.629252
```

```
plot(fatigue_scale, reaction_time)
```

When you have a scatterplot like this, consider adding a smoothed line (e.g LOESS) to show trend.

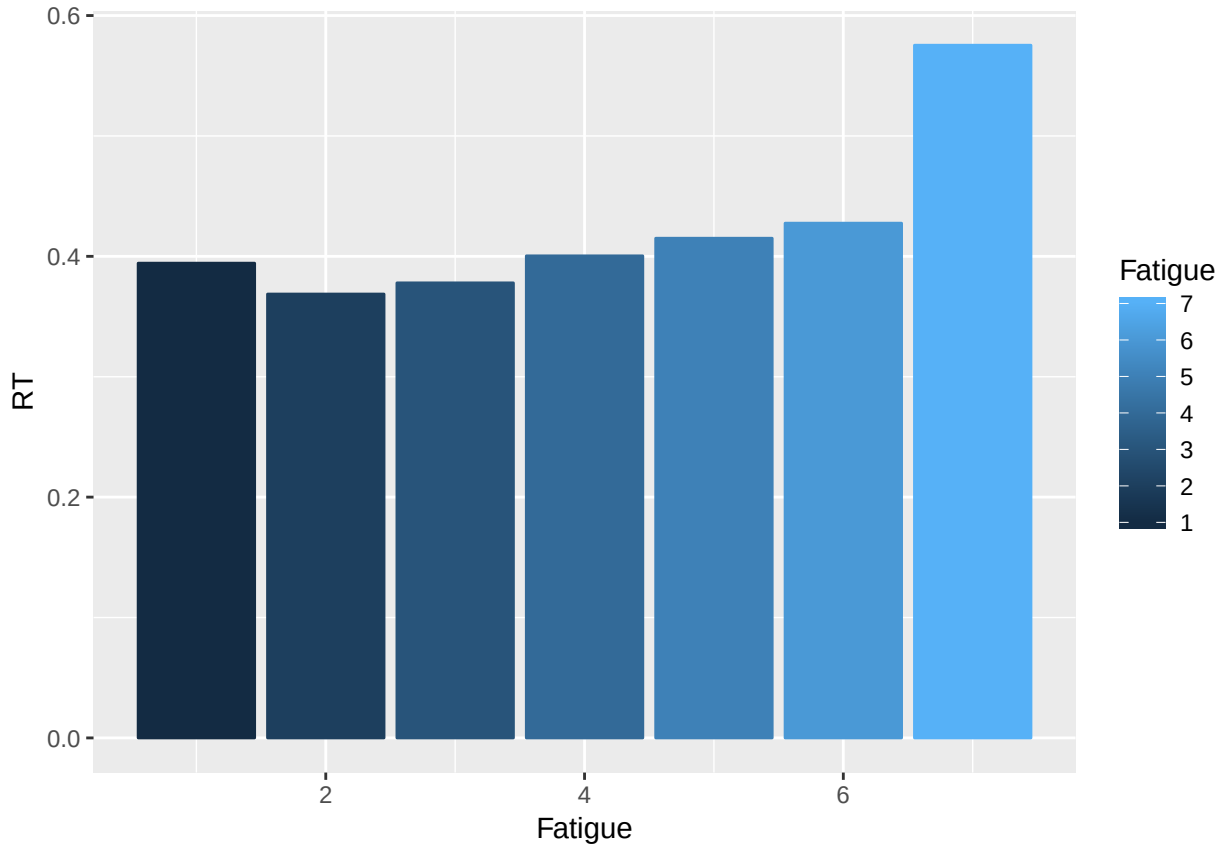



```
summary(data$Fatigue)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   3.000   3.500   3.629   5.000   7.000      18
```

```
par(mfrow = c(1,2))
af$Sleep <- as.factor(af$Fatigue)
ggplot(summarise(group_by(af, Fatigue), RT = mean(RT)),aes(x= Fatigue, y = RT)) + geom_bar(stat = "iden
```

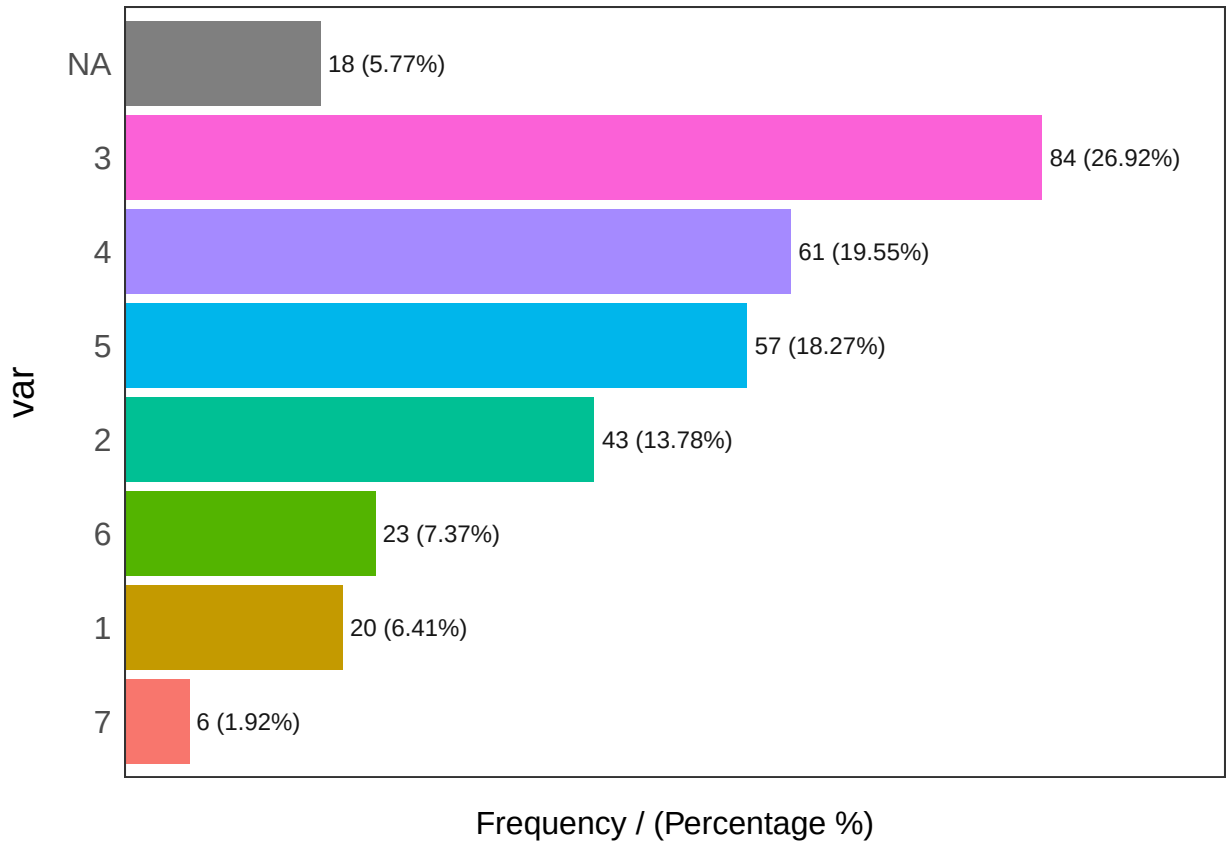
- what did you learn?




```
summarise(group_by(af, Fatigue), mean(RT))
```

```
## # A tibble: 7 x 2
##   Fatigue `mean(RT)`
##     <int>      <dbl>
## 1       1      0.394
## 2       2      0.369
## 3       3      0.378
## 4       4      0.400
## 5       5      0.415
## 6       6      0.428
## 7       7      0.575
```

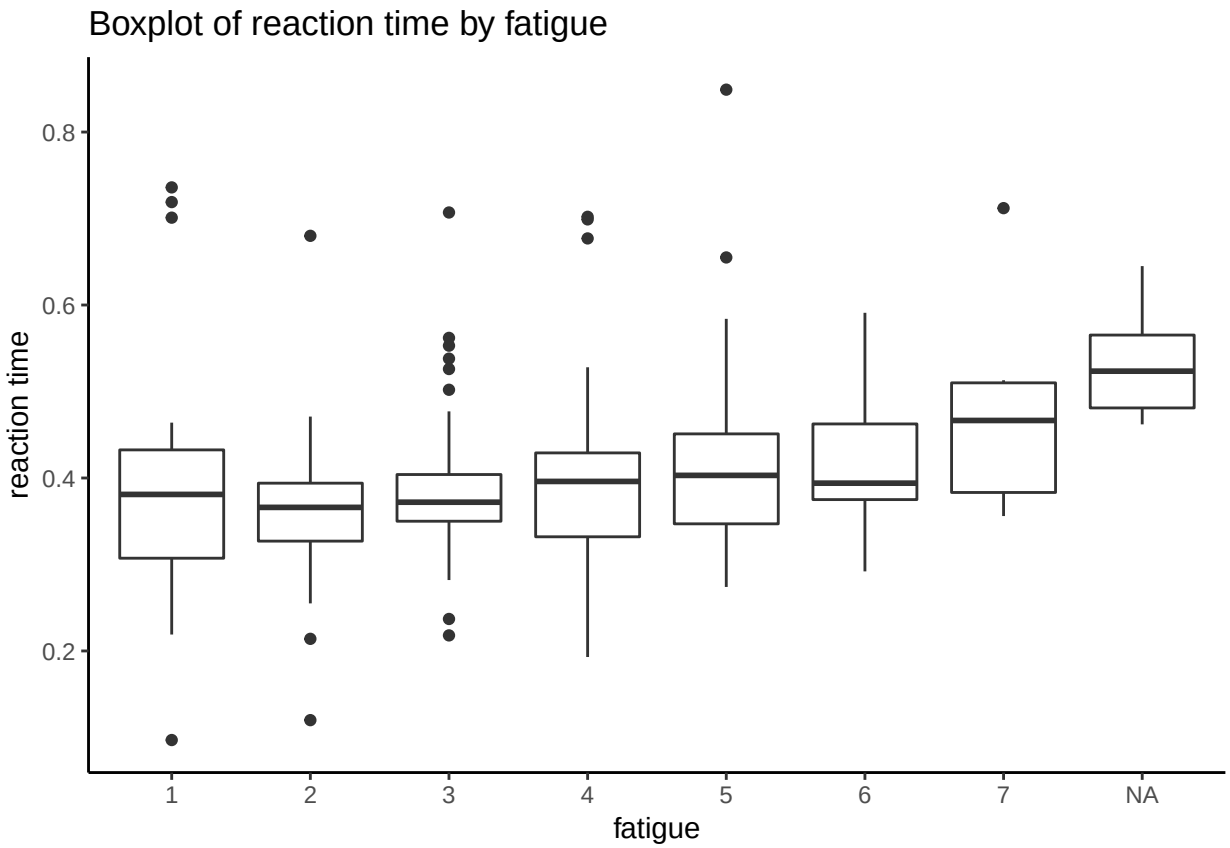
```
freq(data$Fatigue)
```




```
##    var frequency percentage cumulative_perc
## 1    3        84      26.92           26.92
## 2    4        61      19.55           46.47
## 3    5        57      18.27           64.74
## 4    2        43      13.78           78.52
## 5    6        23       7.37           85.89
## 6    1        20       6.41           92.30
## 7 <NA>        18       5.77           98.07
## 8    7         6       1.92          100.00
```

```
ggplot(data,aes(x = factor(fatigue_scale),y = reaction_time)) +
  theme_classic()  +
  geom_boxplot() +
  labs(title = "Boxplot of reaction time by fatigue",
       x = "fatigue",
       y = "reaction time")
```

```
## Warning: Removed 10 rows containing non-finite values (stat_boxplot).
```



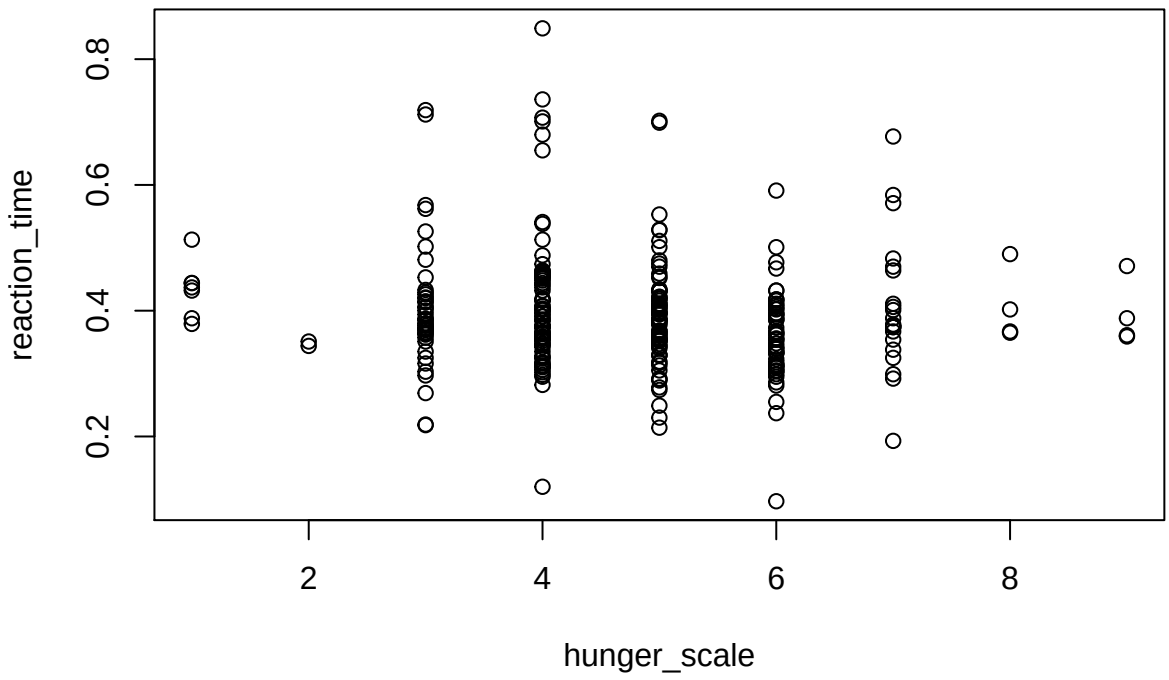
Hunger

In the scatter plot below, it presents that most of the people test their reaction time near a hunger level of 4. When hunger level reaches 6, it is more likely that the person would have the fastest reaction speed.

```
hunger_scale = data$Hunger
mean(data$Hunger, na.rm= T)
```

```
## [1] 4.744898
```

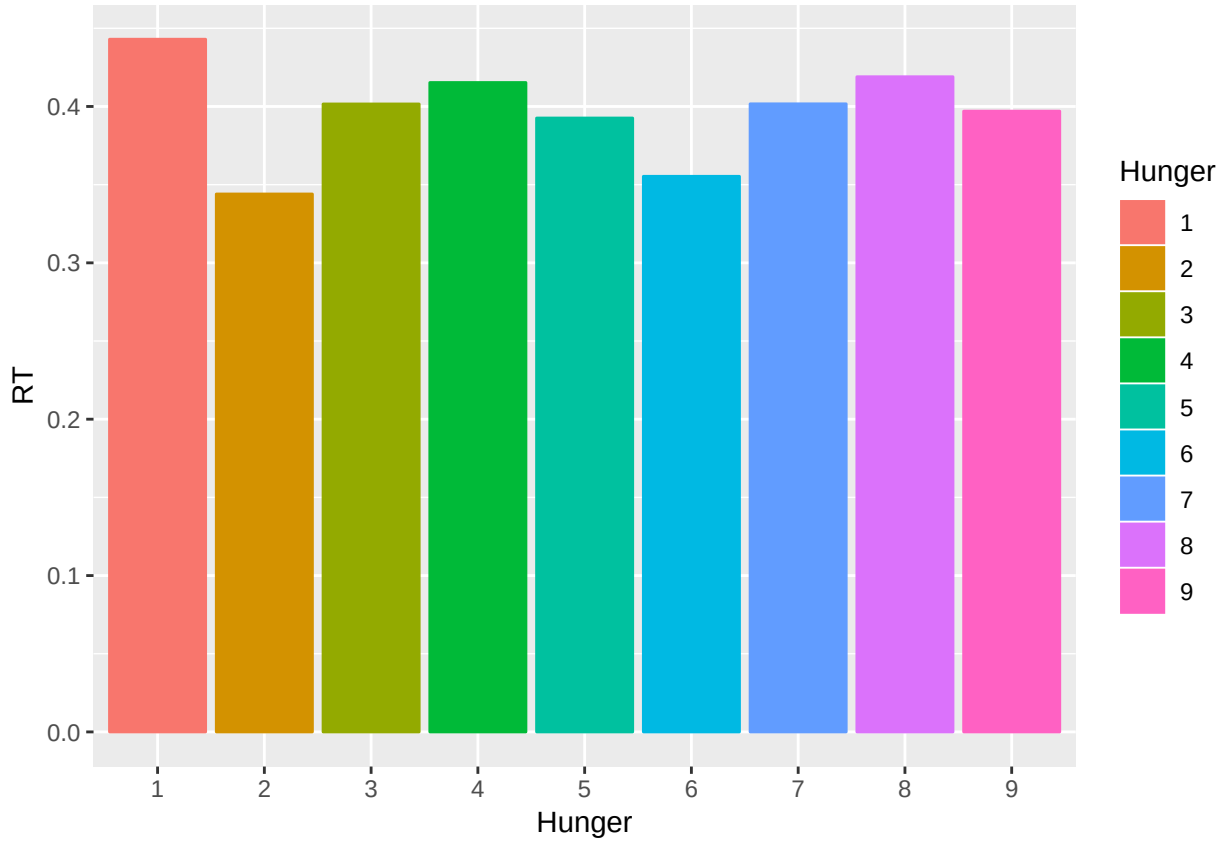
```
plot(hunger_scale, reaction_time)
```




```
summary(data$Hunger)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   4.000   5.000   4.745   6.000   9.000      18
```

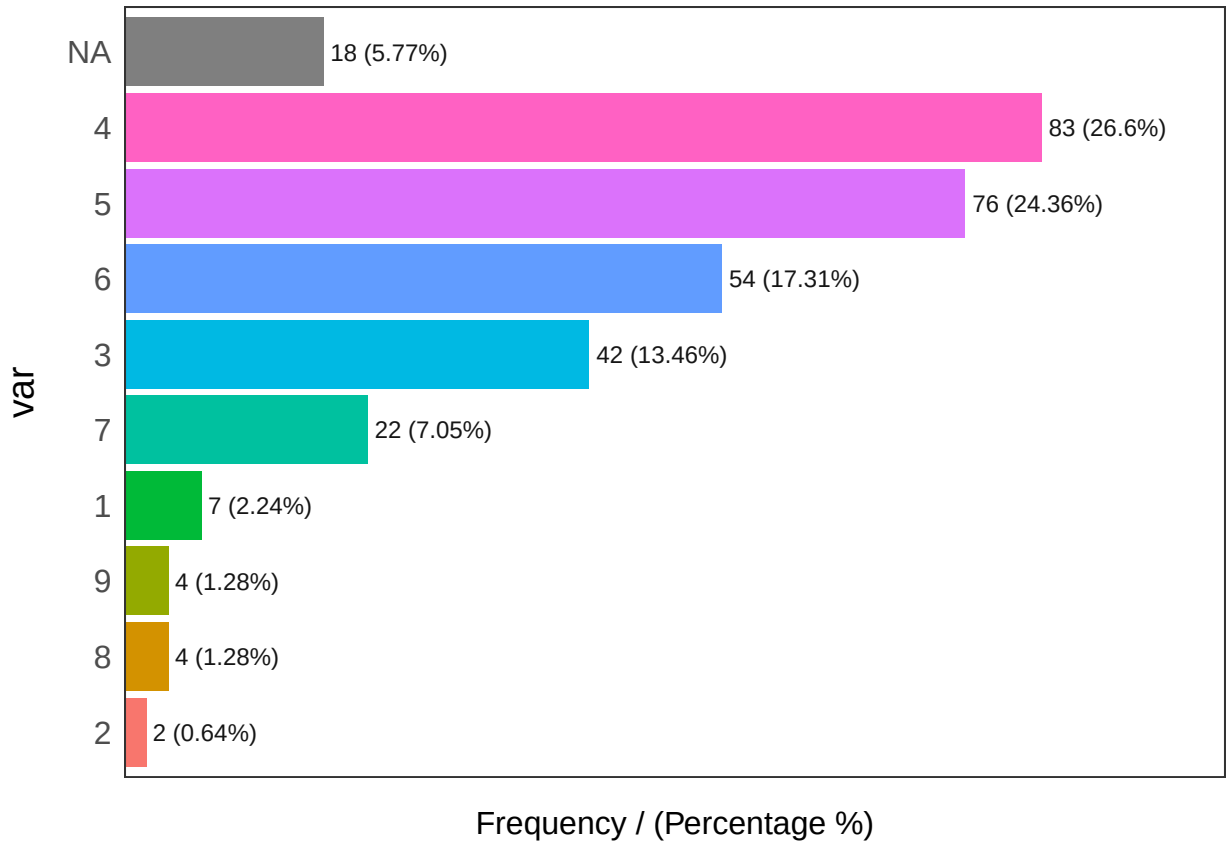
```
par(mfrow = c(1,2))
af$Hunger <- as.factor(af$Hunger)
ggplot(summarise(group_by(af, Hunger), RT = mean(RT)),aes(x= Hunger, y = RT)) + geom_bar(stat = "identi
```




```
summarise(group_by(af, Hunger), mean(RT))
```

```
## # A tibble: 9 x 2
##   Hunger `mean(RT)`
##   <fct>       <dbl>
## 1 1           0.443
## 2 2           0.344
## 3 3           0.402
## 4 4           0.415
## 5 5           0.393
## 6 6           0.355
## 7 7           0.402
## 8 8           0.419
## 9 9           0.397
```

```
freq(data$Hunger)
```




```
##     var frequency percentage cumulative_perc
## 1     4        83      26.60           26.60
## 2     5        76      24.36           50.96
## 3     6        54      17.31           68.27
## 4     3        42      13.46           81.73
## 5     7        22       7.05           88.78
## 6  <NA>        18       5.77           94.55
## 7     1         7       2.24           96.79
## 8     8         4       1.28           98.07
## 9     9         4       1.28           99.35
## 10    2         2       0.64          100.00
```

```
ggplot(data,aes(x = factor(hunger_scale),y = reaction_time)) +
  theme_classic()  +
  geom_boxplot() +
  labs(title = "Boxplot of reaction time by hunger",
       x = "hunger",
       y = "reaction time")
```

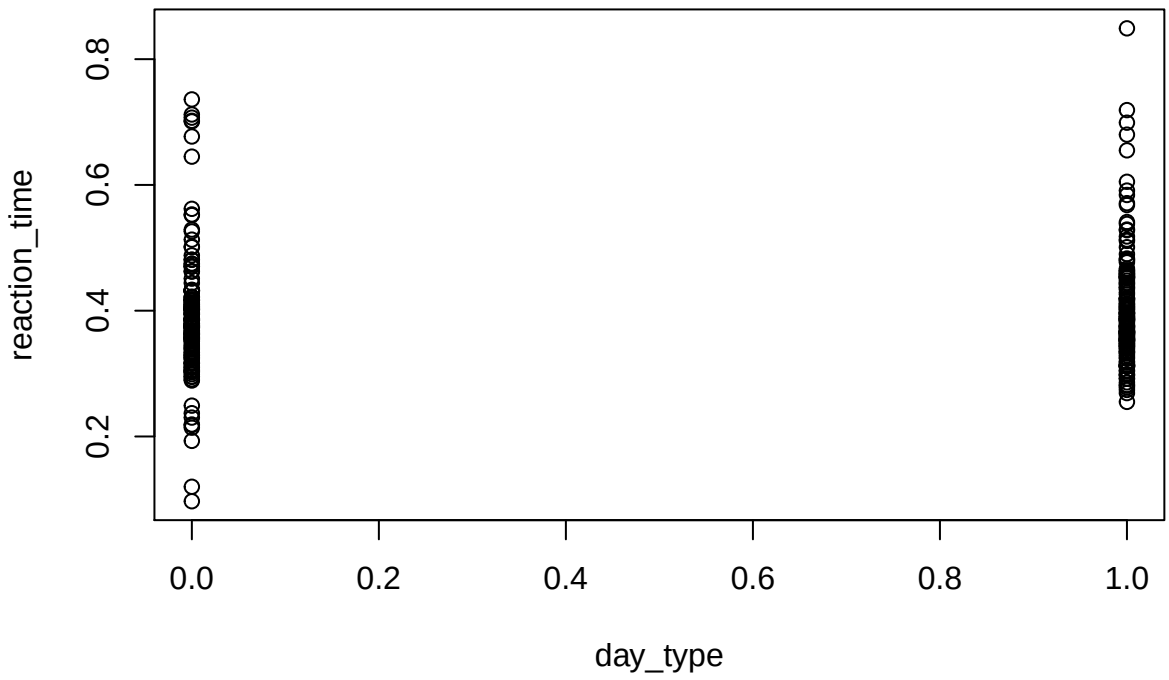
```
## Warning: Removed 10 rows containing non-finite values (stat_boxplot).
```

#heavy/light day I used number 0 to represent light day and 1 for busy day. According to the plots below, it is easy to find that normally people react faster in light days.

```
day_type = data$Type
mean(data$Hunger, na.rm= T)
```

```
## [1] 4.744898
```

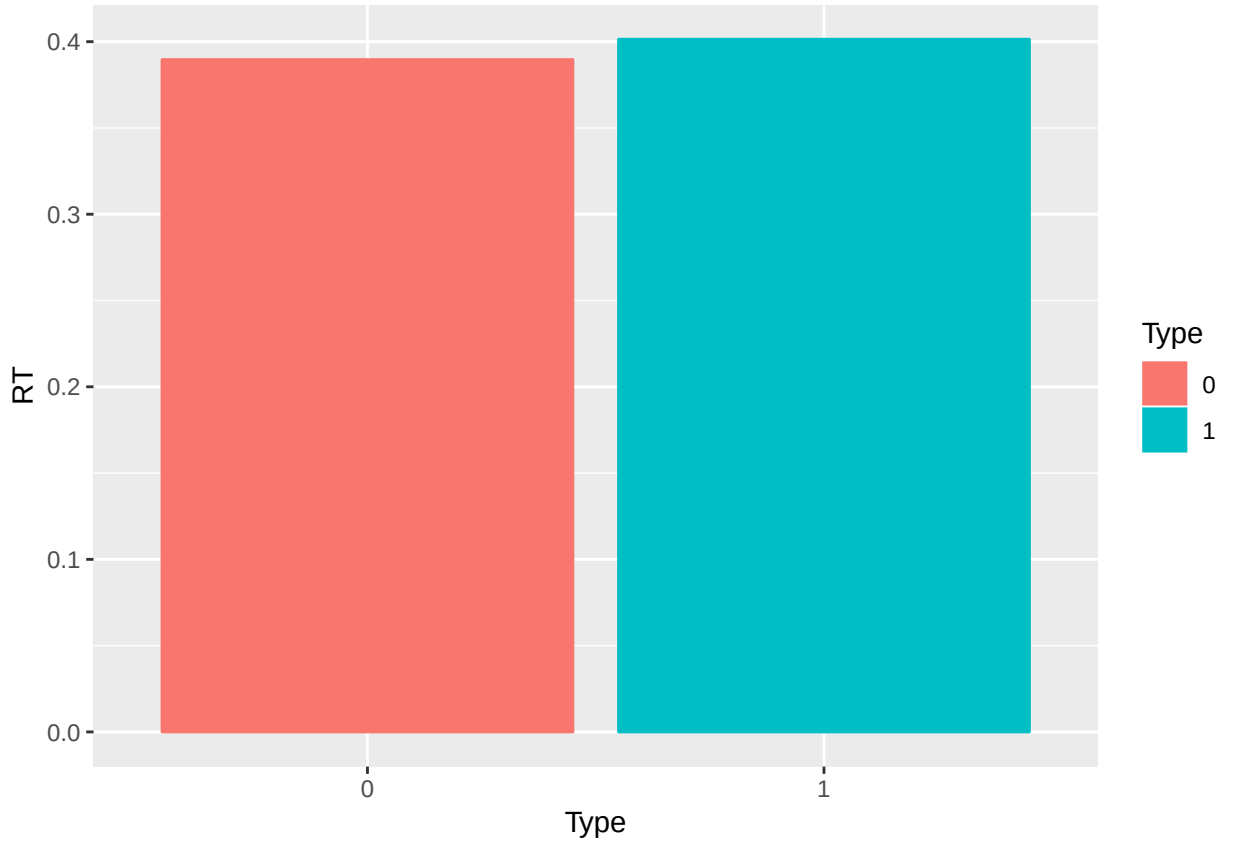
```
plot(day_type, reaction_time)
```




```
summary(data$Type)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.0000  0.0000  1.0000  0.5065  1.0000  1.0000       4
```

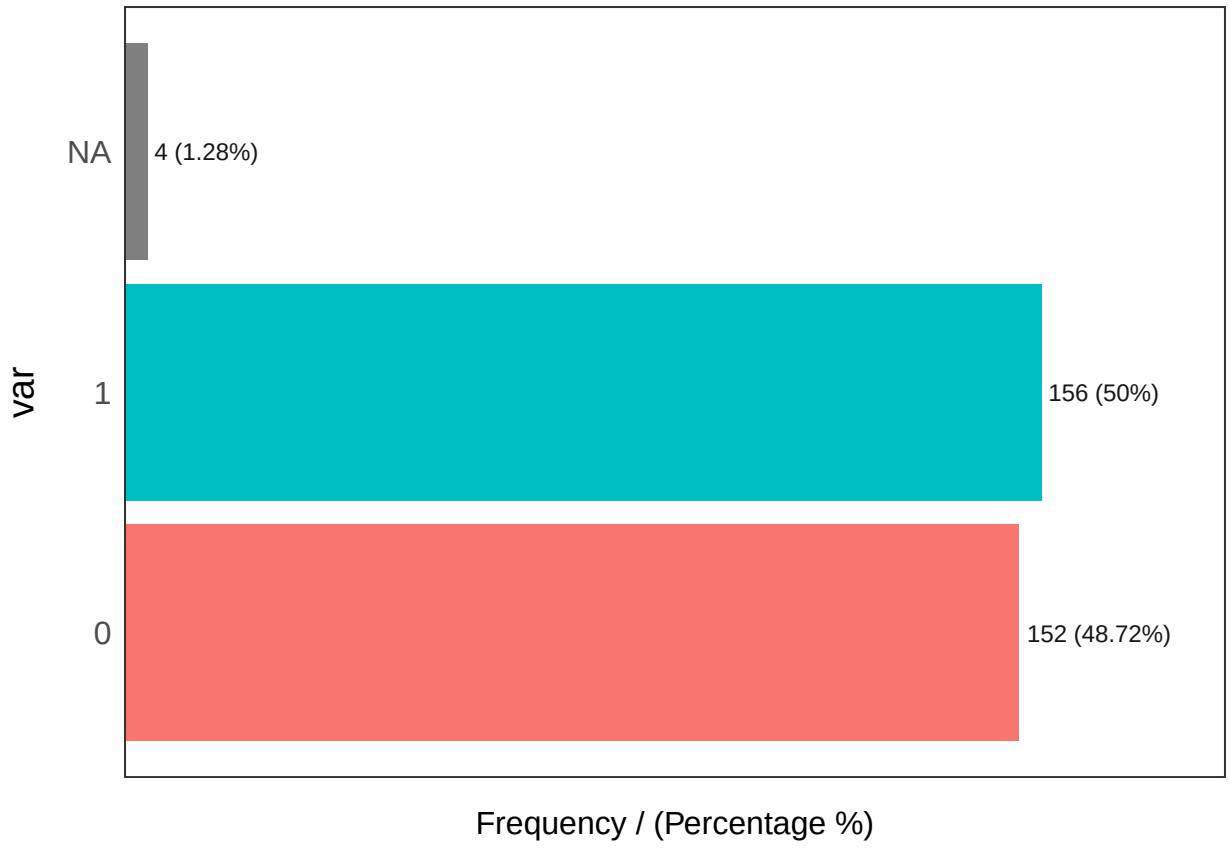
```
par(mfrow = c(1,2))
af$Type <- as.factor(af$Type)
ggplot(summarise(group_by(af, Type), RT = mean(RT)),aes(x= Type, y = RT)) + geom_bar(stat = "identity",
```




```
summarise(group_by(af, Type), mean(RT))
```

```
## # A tibble: 2 x 2
##   Type  `mean(RT)`
##   <fct>      <dbl>
## 1 0          0.390
## 2 1          0.401
```

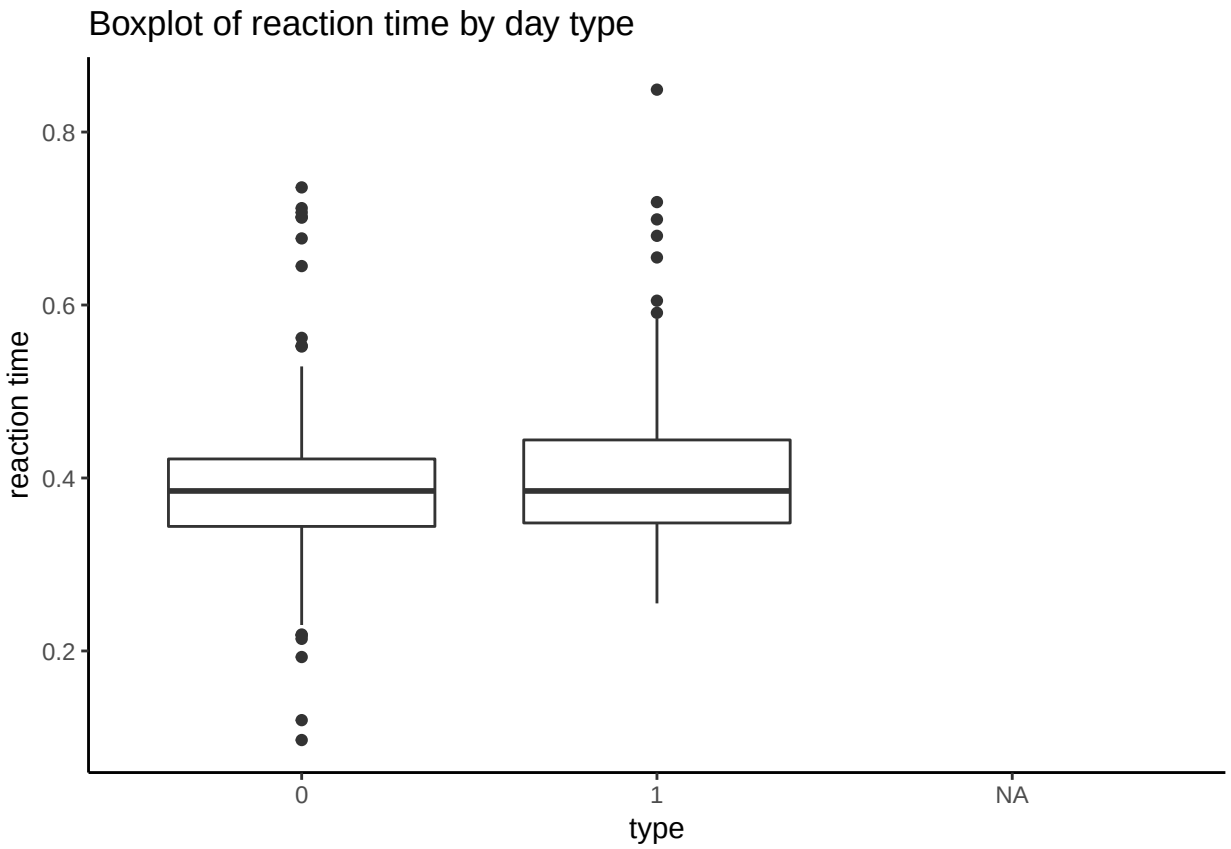
```
freq(data$Type)
```




```
##    var frequency percentage cumulative_perc
## 1    1       156      50.00           50.00
## 2    0       152      48.72           98.72
## 3 <NA>         4       1.28          100.00
```

```
ggplot(data,aes(x = factor(day_type),y = reaction_time)) +
  theme_classic()  +
  geom_boxplot() +
  labs(title = "Boxplot of reaction time by day type",
       x = "type",
       y = "reaction time")
```

```
## Warning: Removed 10 rows containing non-finite values (stat_boxplot).
```



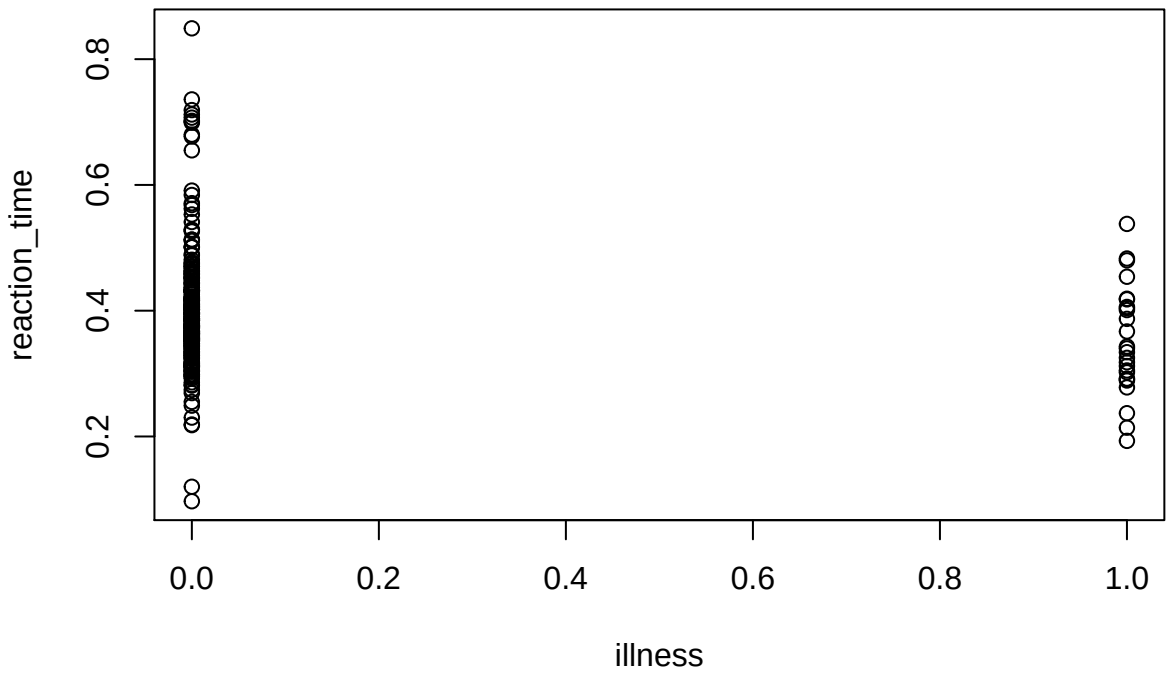
illness

Most of the people test their reaction time without illness, and I did not exclude the missing data because I personally believe that the data is gathered when healthy. Normally people would record illness when sick. And surprisingly that reaction time did not increase when people are sick, maybe it is because our sample size is not large enough.

```
illness = data$illness
mean(data$illness, na.rm= T)
```

```
## [1] 0.09246575
```

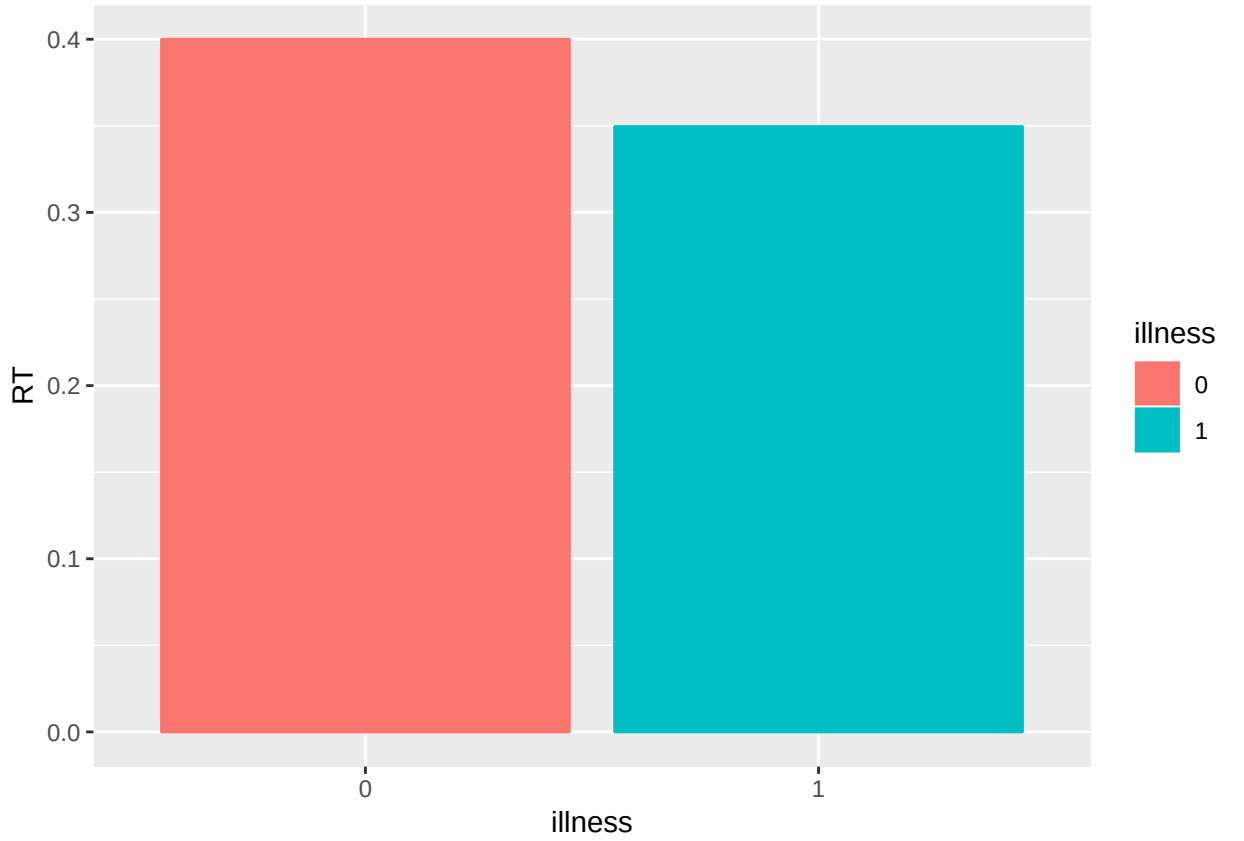
```
plot(illness, reaction_time)
```




```
summary(data$illness)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.00000 0.00000 0.00000 0.09247 0.00000 1.00000      20
```

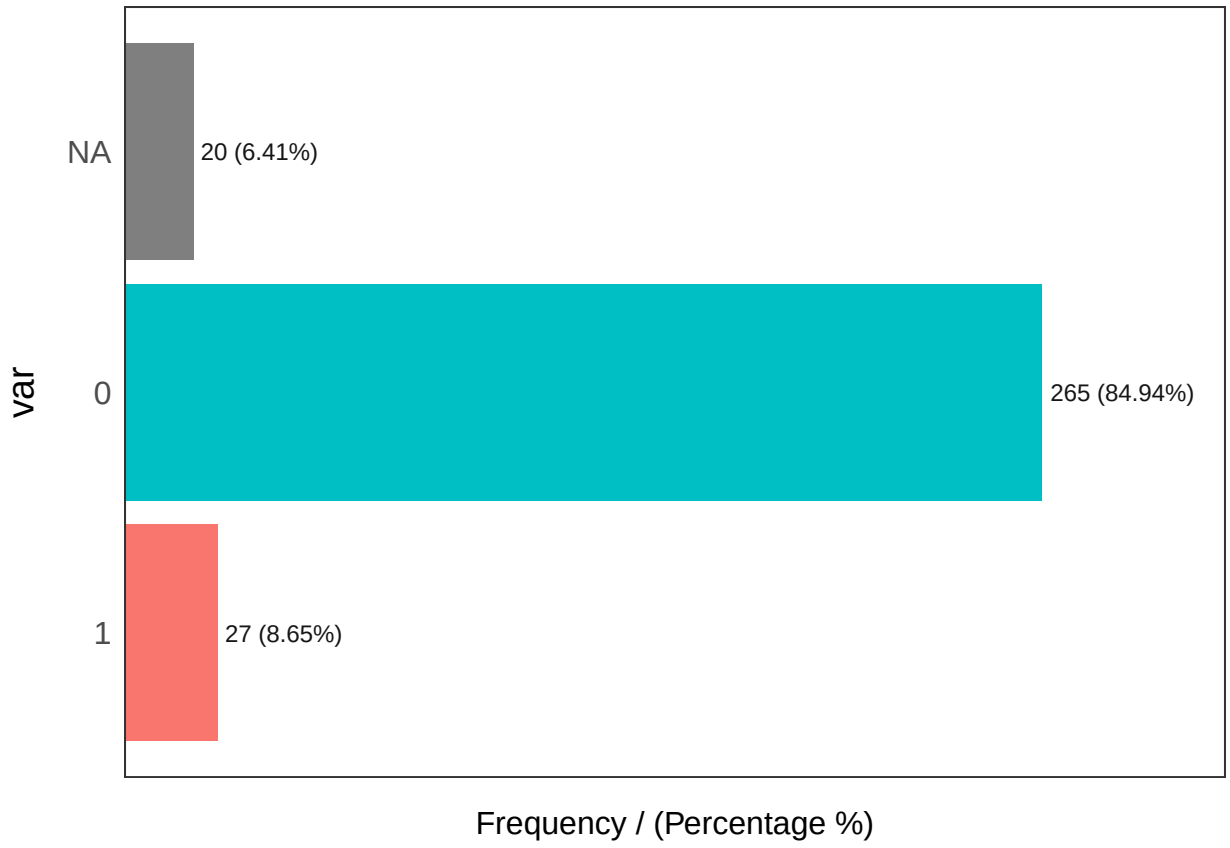
```
par(mfrow = c(1,2))
af$illness <- as.factor(af$illness)
ggplot(summarise(group_by(af, illness), RT = mean(RT)),aes(x= illness, y = RT)) + geom_bar(stat = "iden
```




```
summarise(group_by(af, illness), mean(RT))
```

```
## # A tibble: 2 x 2
##   illness `mean(RT)`
##   <fct>        <dbl>
## 1 0            0.400
## 2 1            0.350
```

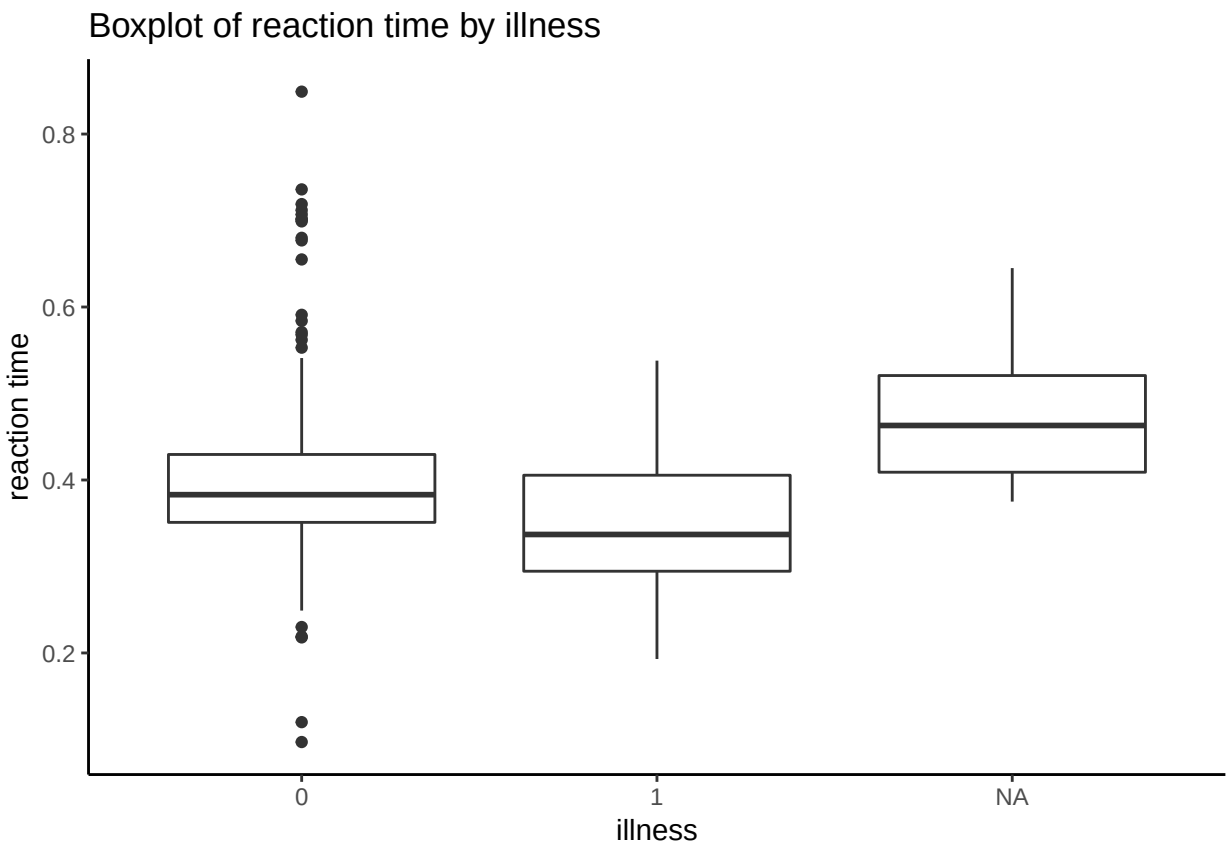
```
freq(data$illness)
```




```
##    var frequency percentage cumulative_perc
## 1    0       265      84.94           84.94
## 2    1        27       8.65           93.59
## 3 <NA>        20       6.41          100.00
```

```
ggplot(data,aes(x = factor(illness),y = reaction_time)) +
  theme_classic()  +
  geom_boxplot() +
  labs(title = "Boxplot of reaction time by illness",
       x = "illness",
       y = "reaction time")
```

```
## Warning: Removed 10 rows containing non-finite values (stat_boxplot).
```



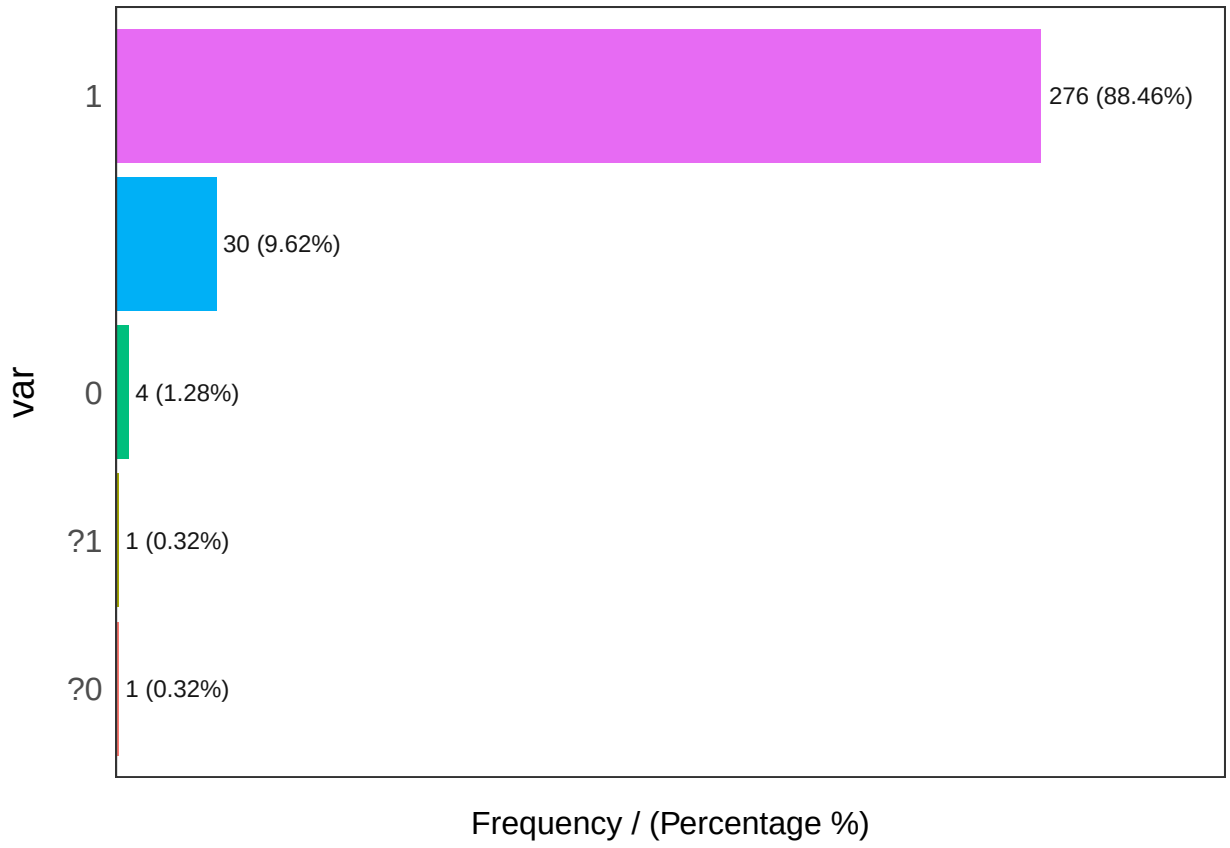
Protocol

Most of the people record themself following the protocol when testing reaction time.

```
protocol = data$Protocol
summary(data$Protocol)
```

```
##      ?0  ?1   0   1
##  30   1   1   4 276
```

```
freq(data$Protocol)
```




```
##   var frequency percentage cumulative_perc
## 1   1       276      88.46           88.46
## 2             30       9.62           98.08
## 3   0          4       1.28           99.36
## 4  ?0          1       0.32           99.68
## 5  ?1          1       0.32          100.00
```

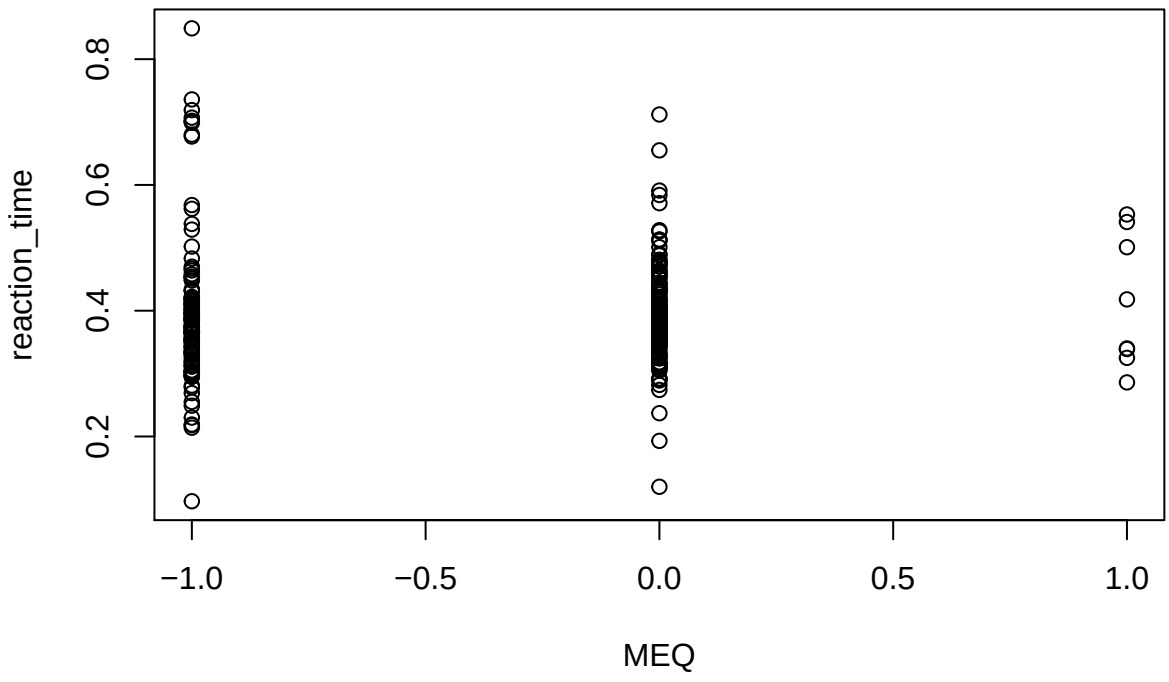
MEQ

The majority people are neither morning type nor night type, and their reaction time is slightly shorter than the other two kinds.

```
MEQ = data$MEQ
mean(data$MEQ, na.rm= T)
```

```
## [1] -0.3733333
```

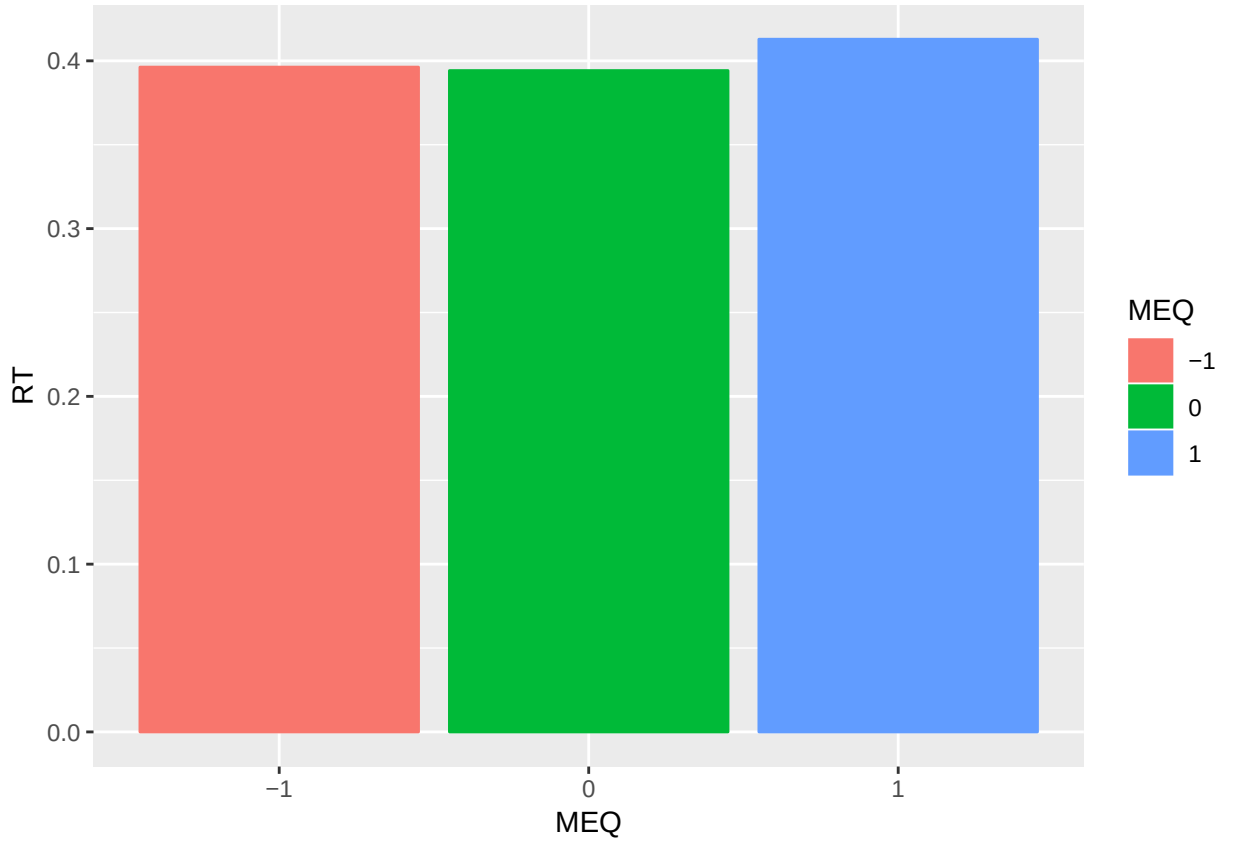
```
plot(MEQ, reaction_time)
```




```
summary(data$MEQ)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -1.0000 -1.0000  0.0000 -0.3733  0.0000  1.0000      12
```

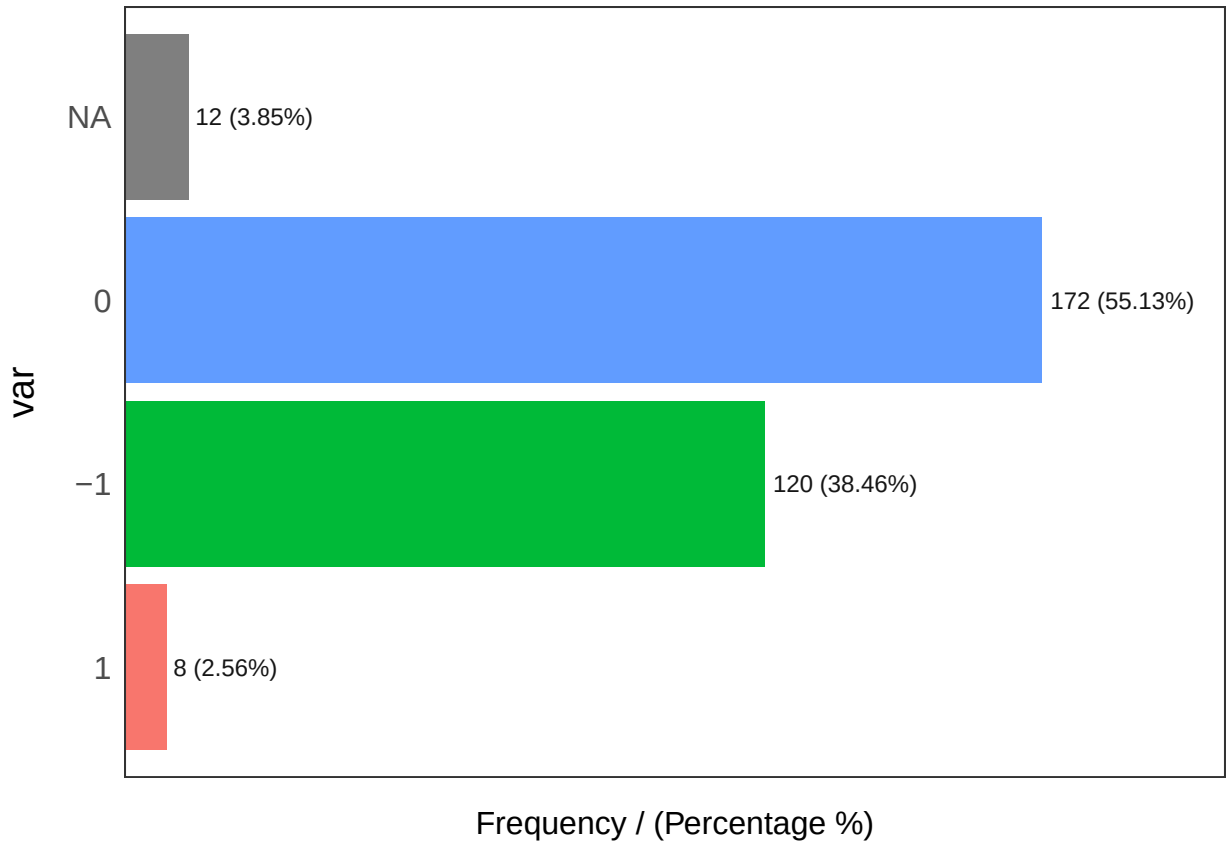
```
par(mfrow = c(1,2))
af$MEQ <- as.factor(af$MEQ)
ggplot(summarise(group_by(af, MEQ), RT = mean(RT)),aes(x= MEQ, y = RT)) + geom_bar(stat = "identity", p
```




```
summarise(group_by(af, illness), mean(RT))
```

```
## # A tibble: 2 x 2
##   illness `mean(RT)`
##   <fct>        <dbl>
## 1 0            0.400
## 2 1            0.350
```

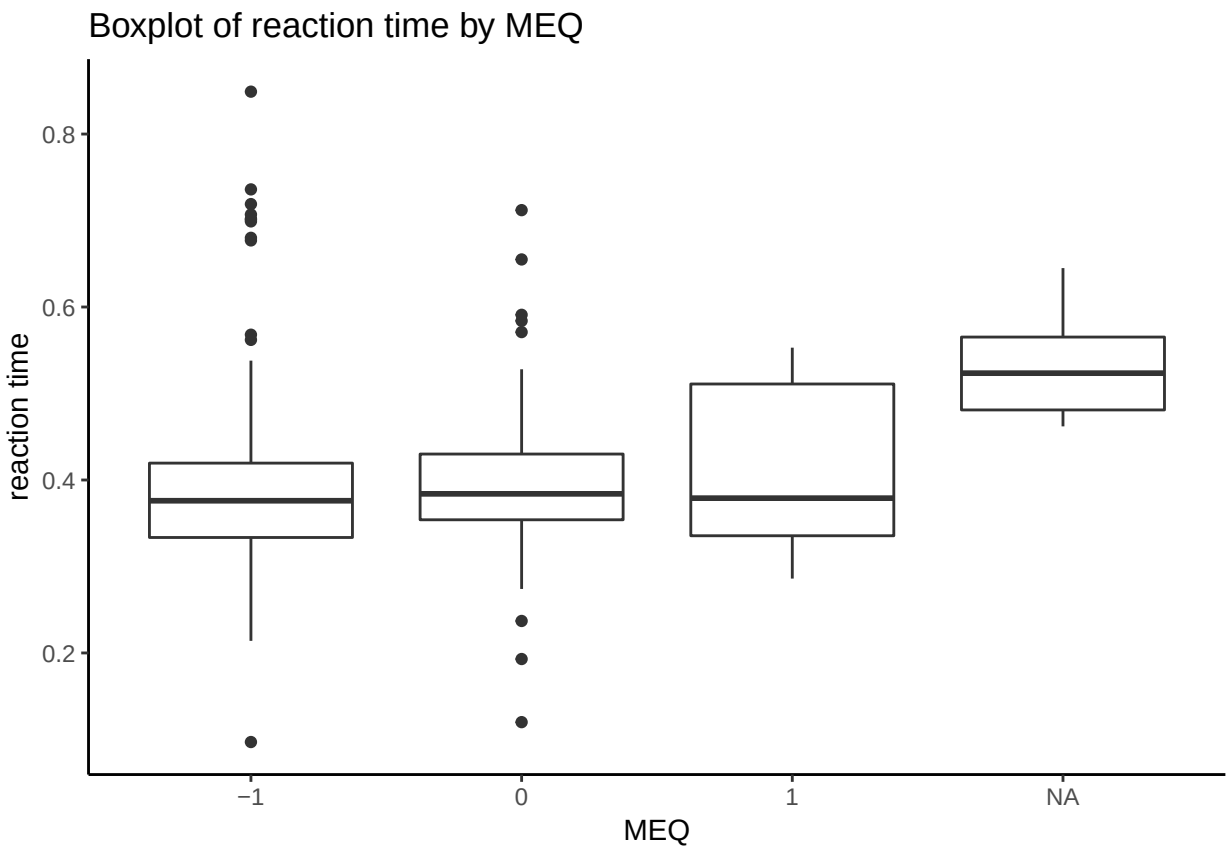
```
freq(data$MEQ)
```




```
##    var frequency percentage cumulative_perc
## 1    0       172      55.13           55.13
## 2   -1       120      38.46           93.59
## 3 <NA>        12       3.85           97.44
## 4    1         8       2.56          100.00
```

```
ggplot(data,aes(x = factor(MEQ),y = reaction_time)) +
  theme_classic()  +
  geom_boxplot() +
  labs(title = "Boxplot of reaction time by MEQ",
       x = "MEQ",
       y = "reaction time")
```

```
## Warning: Removed 10 rows containing non-finite values (stat_boxplot).
```



Record

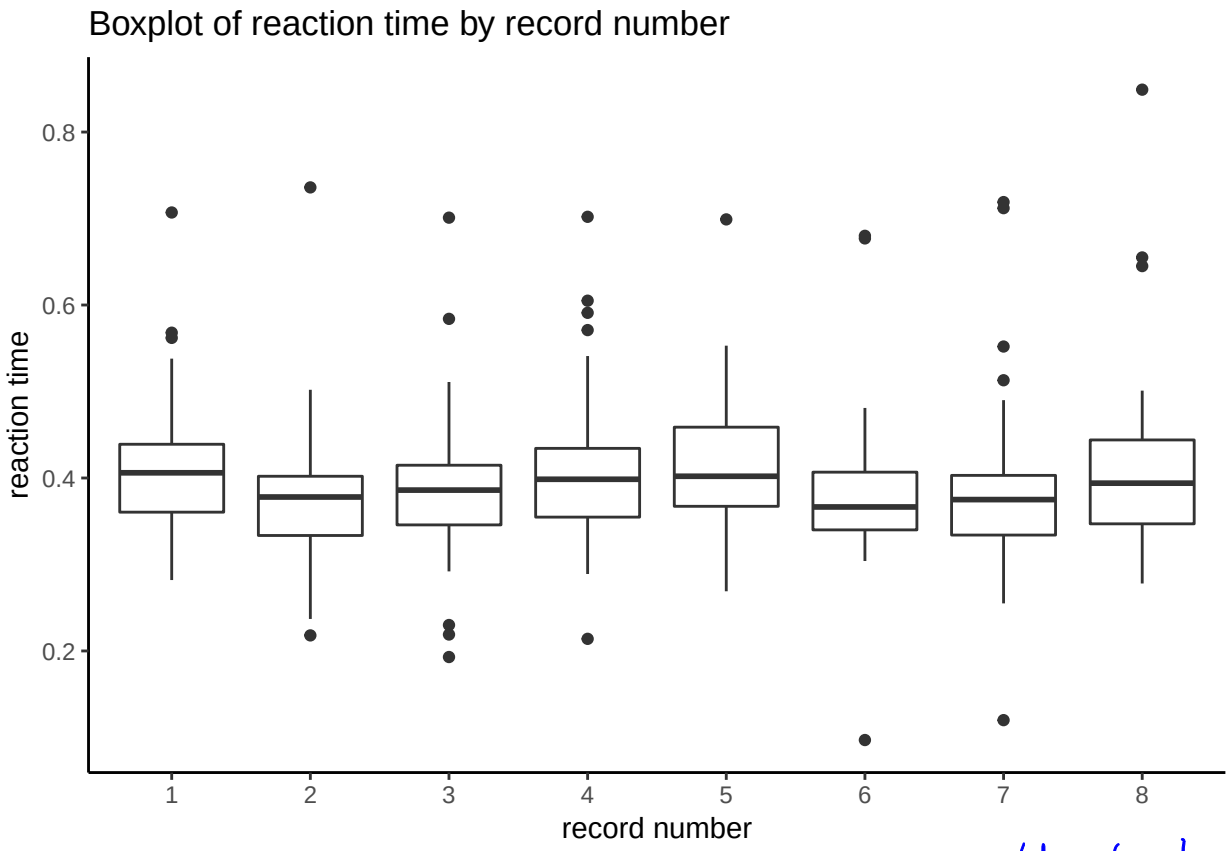
Here is the boxplot of reaction time for 8 trials, the first four tests are from the first day, and the other four are from the second day people choose. There is a pattern that the middle two tests have shorter reaction time.

```
record = data$Record
ggplot(data,aes(x = factor(record),y = reaction_time)) +
  theme_classic()  +
  geom_boxplot() +
  labs(title = "Boxplot of reaction time by record number",
       x = "record number",
       y = "reaction time")
```

```
## Warning: Removed 10 rows containing non-finite values (stat_boxplot).
```

busy vs light days?

population-level trend vs pattern within individuals?



This is the first of your plots that directly relates to the primary question of this project.

- What did you learn from this plot?
- Are there any other ways you could visualize the data to help you understand the diurnal pattern of reaction time?