in_class_27th_jan

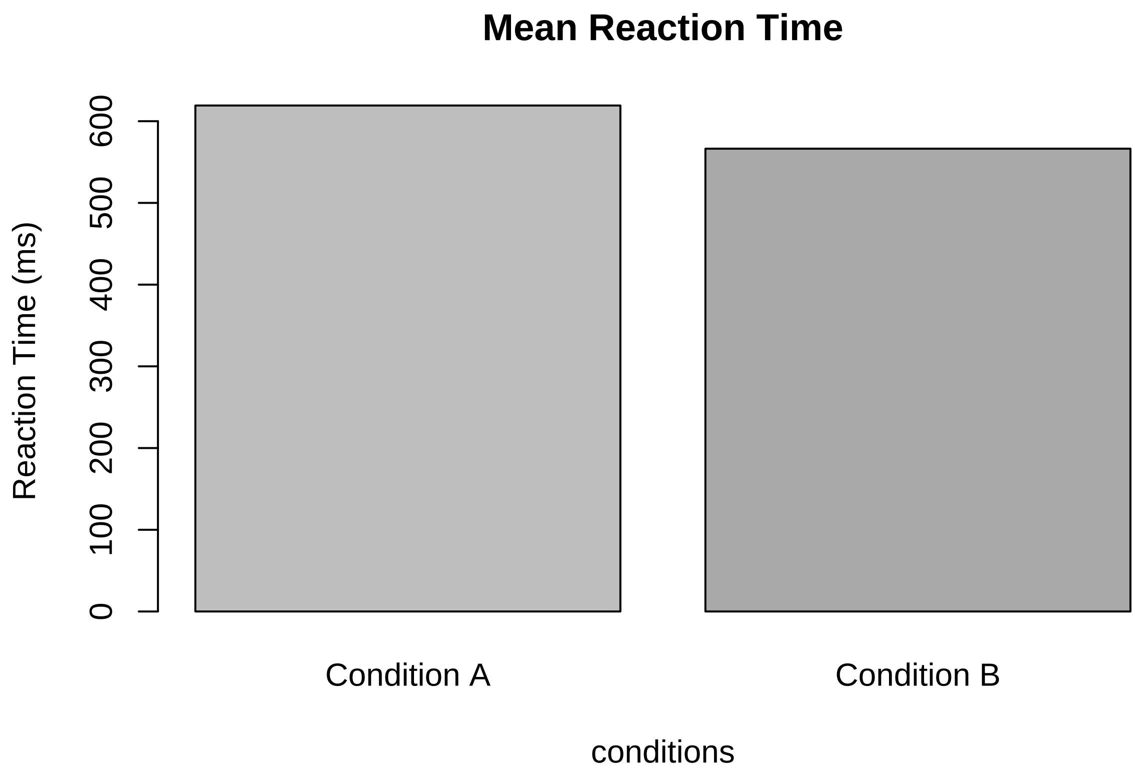
2026-01-27

set environment

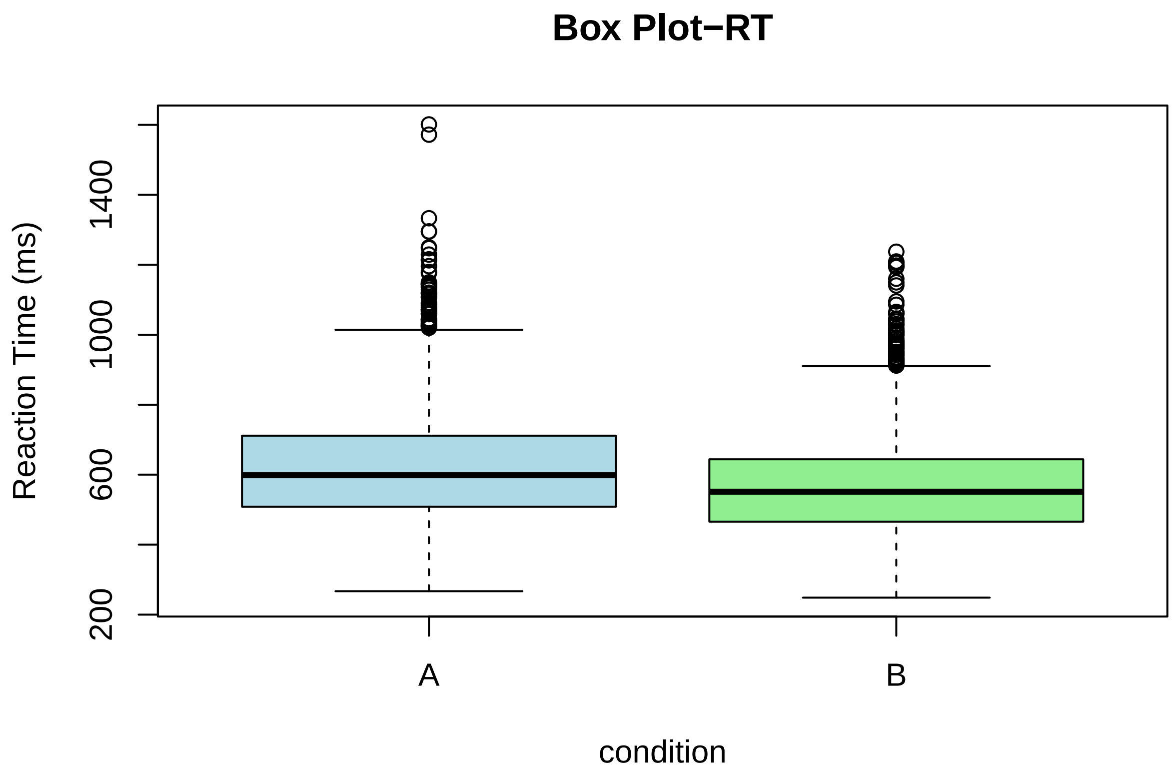
```
# Set rendering parameters
knitr::opts_chunk$set(message = FALSE, warning = FALSE)

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

rt plot

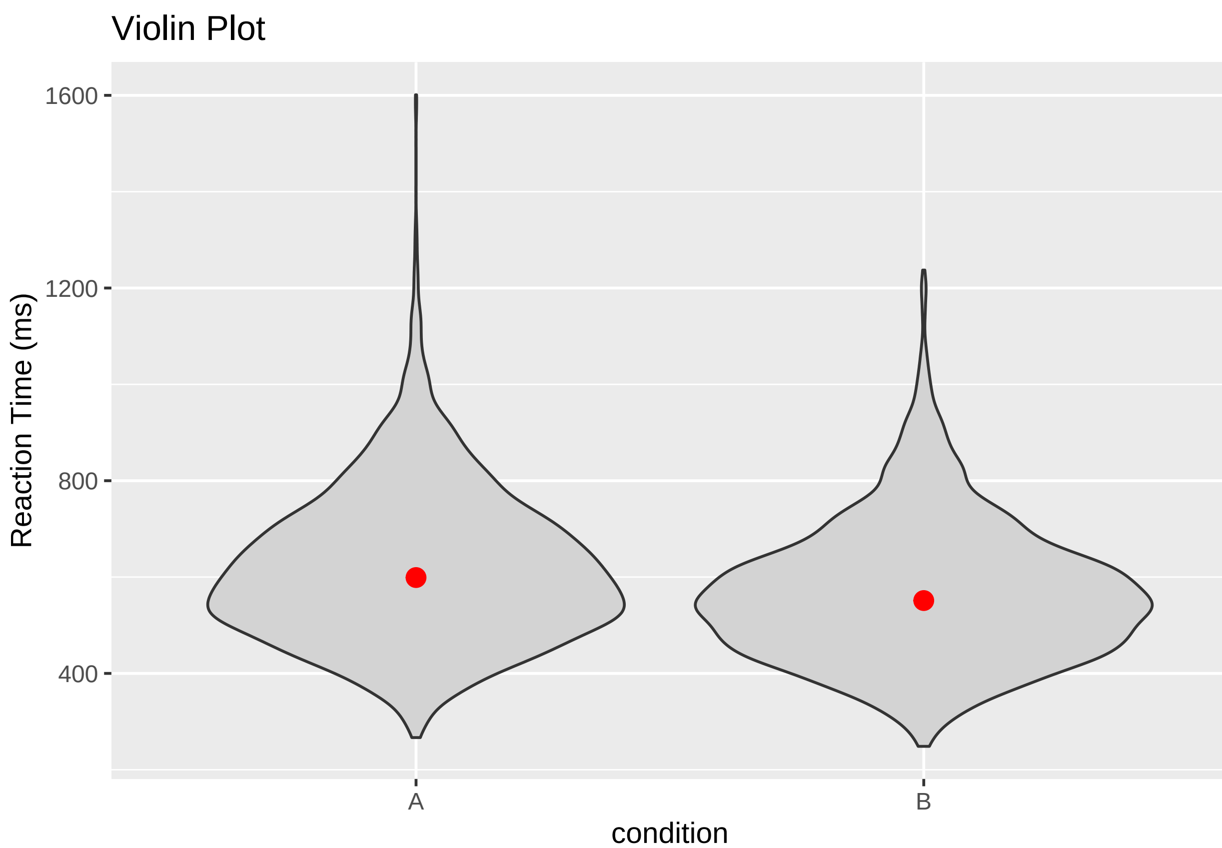
```
rt <- read.csv("C:/Users/rahma/OneDrive/Desktop/brsm/assignments/in_class_gargi_27th_jan/RT dataset.csv

#mean RT per condition
mean_A <- mean(rt$reaction_time_ms[rt$condition == "A"])
mean_B <- mean(rt$reaction_time_ms[rt$condition == "B"])
means <- c(mean_A, mean_B)
names(means) <- c("Condition A", "Condition B")

# Bar plot
barplot(means,
        ylab = "Reaction Time (ms)",
        xlab = "conditions",
        main = "Mean Reaction Time",
        col = c("gray", "darkgray"))
```



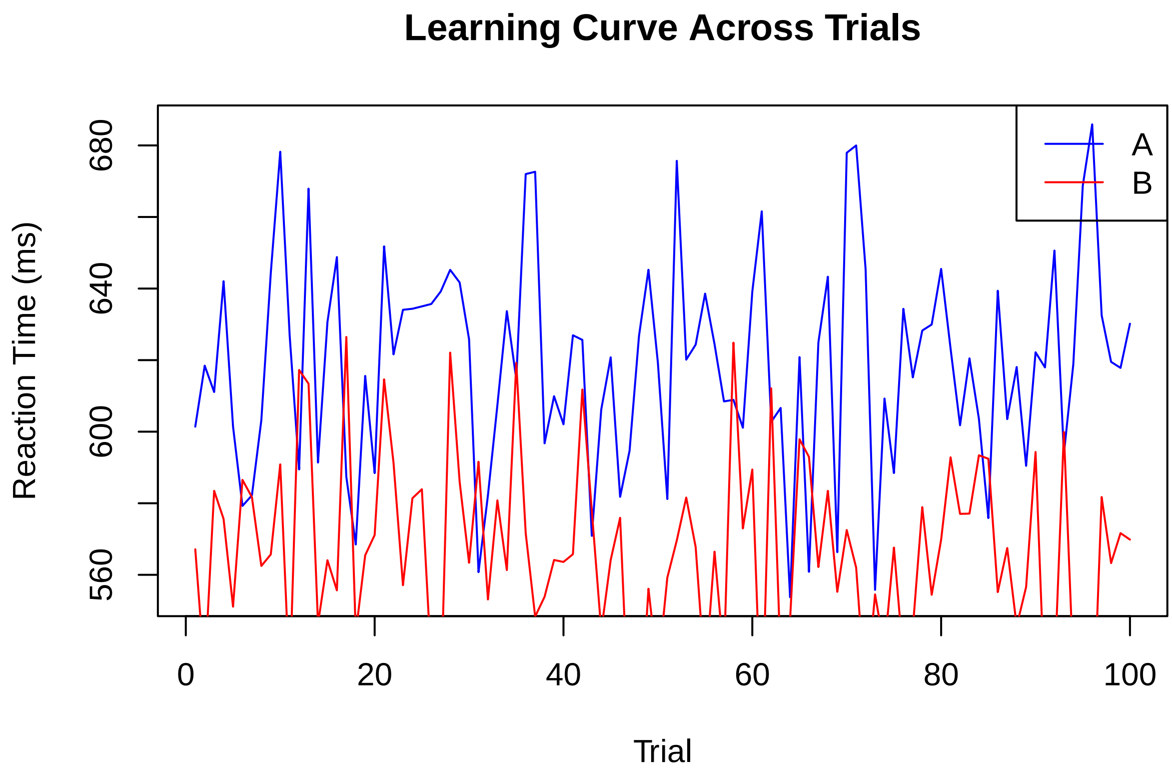

```r
#boxplot
boxplot(reaction_time_ms ~ condition,
        data = rt,
        ylab = "Reaction Time (ms)",
        main = "Box Plot-RT",
        col = c("lightblue", "lightgreen"))
```




```r
#violin plot


ggplot(rt, aes(x = condition, y = reaction_time_ms)) +
  geom_violin(fill = "lightgray") +
  stat_summary(fun = median, geom = "point", size = 3, color = "red") +
  ylab("Reaction Time (ms)") +
  ggtitle("Violin Plot")
```




```r
#learning curve graph across trials

#trial means per condition

trials <- unique(rt$trial)
mean_trial_A <- numeric(length(trials))
mean_trial_B <- numeric(length(trials))
for (t in trials) {
  mean_trial_A[t] <- mean(rt$reaction_time_ms[rt$trial==t & rt$condition=="A"])
  mean_trial_B[t] <- mean(rt$reaction_time_ms[rt$trial==t & rt$condition=="B"])
}
# Plot
plot(trials, mean_trial_A, type="l", col="blue",
     xlab="Trial", ylab="Reaction Time (ms)",
     main="Learning Curve Across Trials")
lines(trials, mean_trial_B, col="red")
legend("topright", legend=c("A","B"), col=c("blue","red"), lty=1)
```




```
# The mean reaction time plots are clean visually, but it plots only the mean
# and provides no other information about the data whatsoever . It doesn't
# indicate how the data is distributed, or anything about the outliers.

# The Box Plot-RT does an okay job indicating the outliers , the inter-quantile
# range and the median but does not give any information regarding the
# distribution of the data .

# Now , the violin plot does a great job indicating the median and the outliers.
# Also , one can interpret how the data is distributed across conditions .
# One minor imrpovement would be to plot the inter-quantile ranges as well .

# The graph for learning curve across trials feels quite cluttered visually .
# The starting value on the y-axis itself does not start from a base zero .
# So, this could lead to misinterpretation .And also , one can see that the plot
# for condition B is sometimes not viewed properly because of the aforementioned
# starting value of the y-axis .
```

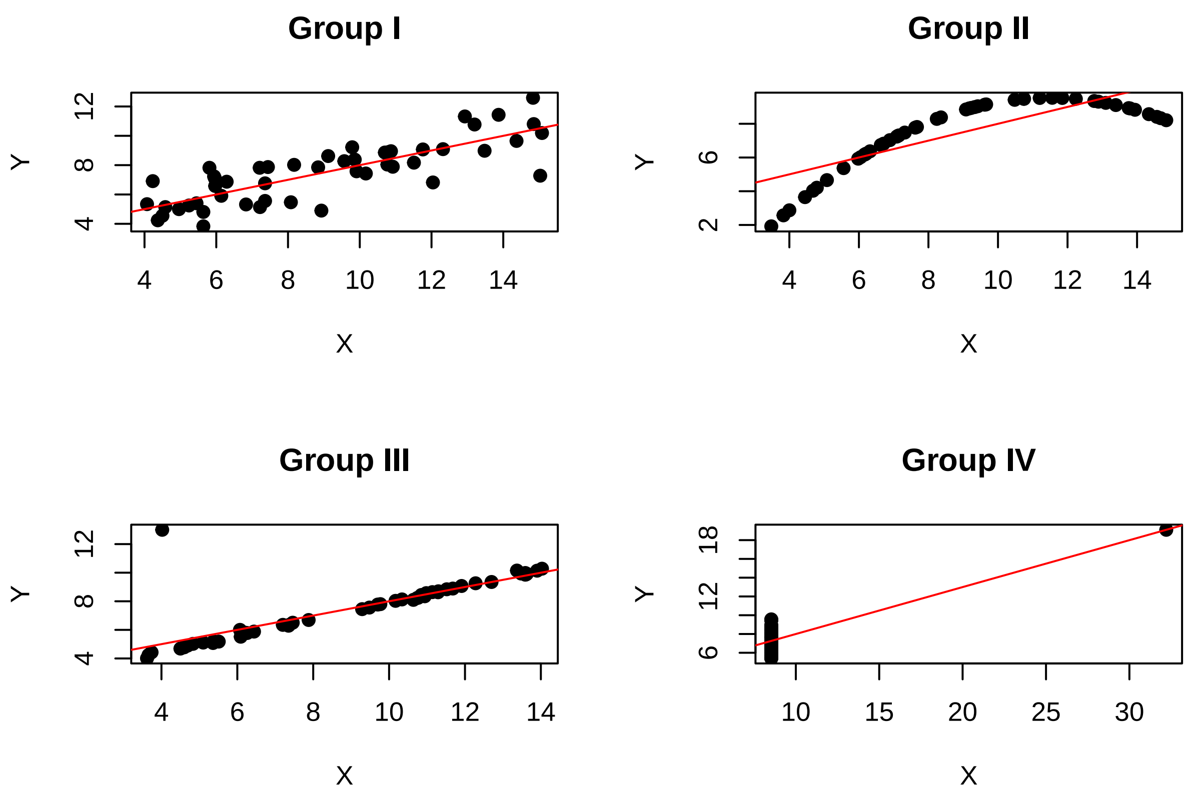
```
data <- read.csv("C:/Users/rahma/OneDrive/Desktop/brsm/assignments/in_class_gargi_27th_jan/Anscombe data
str(data)
```

```
## 'data.frame':    200 obs. of  3 variables:
##  $ Group: chr  "I" "I" "I" "I" ...
##  $ x    : num  8.17 14.85 12.32 10.77 5.64 ...
##  $ y    : num  8.02 10.8 9.09 8.04 3.83 ...
```

```r
data %>%
  group_by(Group) %>%
  summarise(
    mean_x = mean(x),
    mean_y = mean(y),
    sd_x = sd(x),
    sd_y = sd(y),
    cor_xy = cor(x, y)
  )
```

```
## # A tibble: 4 x 6
##   Group mean_x mean_y  sd_x  sd_y cor_xy
##   <chr>  <dbl>  <dbl> <dbl> <dbl>  <dbl>
## 1 I       9.00   7.50  3.35  2.05  0.816
## 2 II      9.00   7.50  3.35  2.05  0.816
## 3 III     9.00   7.5   3.35  2.05  0.816
## 4 IV      9.00   7.50  3.35  2.05  0.816
```

```r
#plots

par(mfrow = c(2,2))
groups <- unique(data$Group)

for(g in groups){
  subset_data <- subset(data, Group == g)
  plot(subset_data$x, subset_data$y,
       main = paste("Group", g),
       xlab = "X", ylab = "Y",
       pch = 19)
  abline(lm(y ~ x, subset_data), col = "red")
}
```




```
# Though the coefficient of correlation for all the groups are almost equal ,
# but the relationship in itself is very different . One cannot jump to
# conclusions just by reading the statistic value, in this case, the coefficient
# of correlation . So, these plot do a great job to see how different effects
# could result to the same statistical value .

# For group I , everything is good , the correlation is well induced and the
# line is a great fit .

# For group II though , one cannot say the same thing , the relationship is
# non-linear and yet it is being fit to a straight line which forces the
# correlation coefficient to be 0.81

# In Group III , the outlier doesn't seem to drive the outcome too much . So, it
# is acceptable to a certain level . But , the outlier must be inspected and
# dealt with accordingly.

# In Group IV though , the outlier drives a huge effect . Even though there is
# no relationship , it is being fit to the red line due to the outlier and the
# resulting coefficient value .
```