

# Project1

Yahui Li

2020/1/13

#Step1 Read Data

We will only examine the math scores in 1st grade in this project.

```
data <- data.frame(star1 = STAR$star1, math1 = STAR$math1)
```

```
sapply(data,class)
```

```
##      star1      math1  
## "factor" "integer"
```

```
sapply(data,summary)
```

```
## $star1  
##      regular      small regular+aide      NA's  
##      2584      1925      2320      4769  
##  
## $math1  
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's  
##      404.0  500.0  529.0  530.5  557.0  676.0  4998
```

```
data.star1.na <- data[is.na(data$star1),]  
all(is.na(data.star1.na$math1))
```

```
## [1] TRUE
```

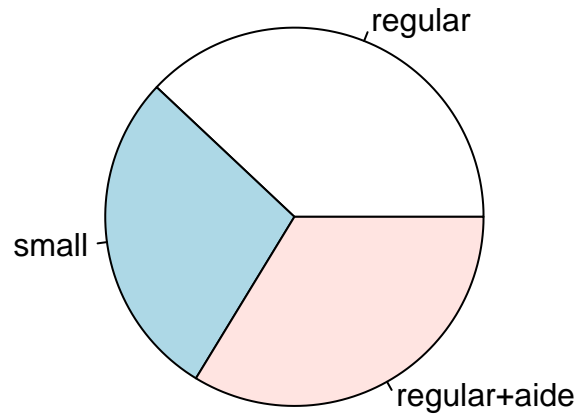
Which shows that the math score has not been recorded if class type is not recorded. So we can remove the data where star1 is NA.

One of the way to deal with NA in math1 is to remove them

```
data_remove_na <- na.omit(data[-is.na(data$star1),])
```

```
pie(table(data_remove_na$star1),main = "pie chart of STAR class type")
```

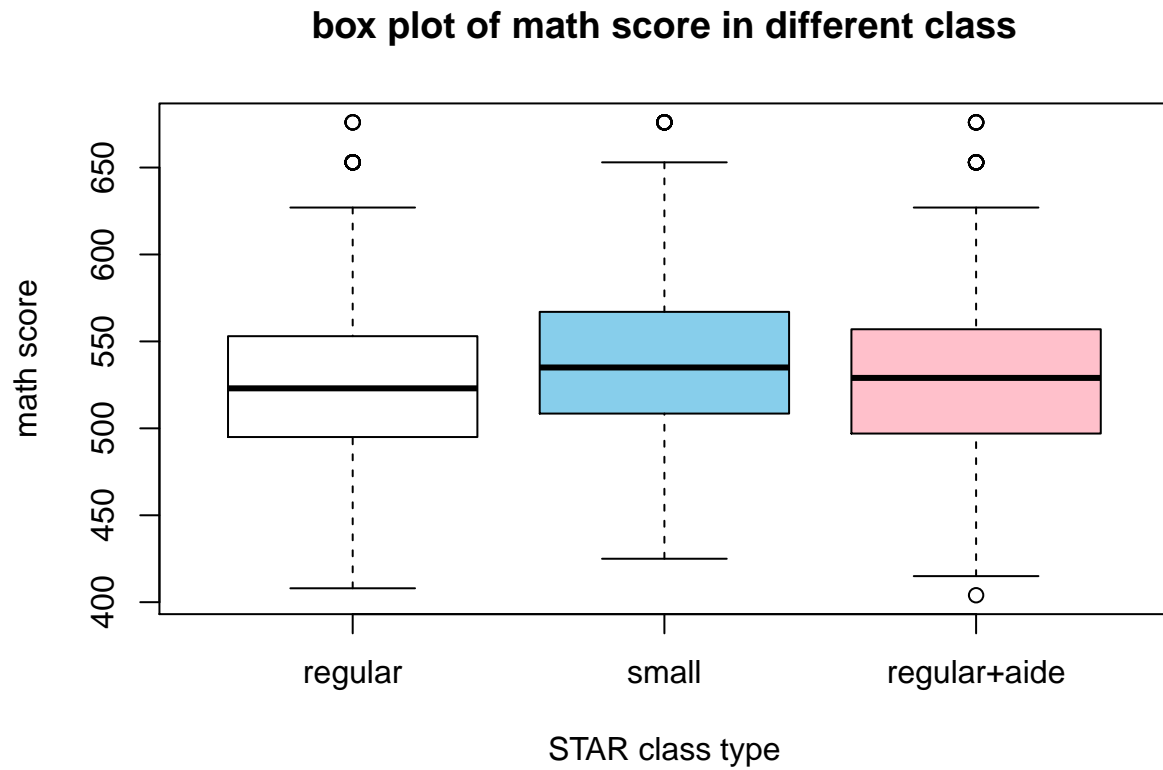
## pie chart of STAR class type



```
tapply(data_remove_na$math1, data_remove_na$star1,summary)
```

```
## $regular
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  408.0  495.0   523.0   525.3  553.0   676.0
##
## $small
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  425.0  509.2   535.0   538.7  567.0   676.0
##
## $`regular+aide`
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  404.0  497.0   529.0   529.6  557.0   676.0
```

```
boxplot(data$math1~data$star1,main = "box plot of math score in different class",
        xlab = "STAR class type", ylab = "math score", col = c("white", "skyblue", "pink"))
```



From the result For mean,  $\text{small} > \text{regular+aide} > \text{regular}$  From all quantile information,  $\text{small} > \text{regular+aide} > \text{regular}$  For min,  $\text{small} > \text{other two}$ ; For max, they are the same

something interesting: there are only some certain scores like 601 612 627 653 676.