

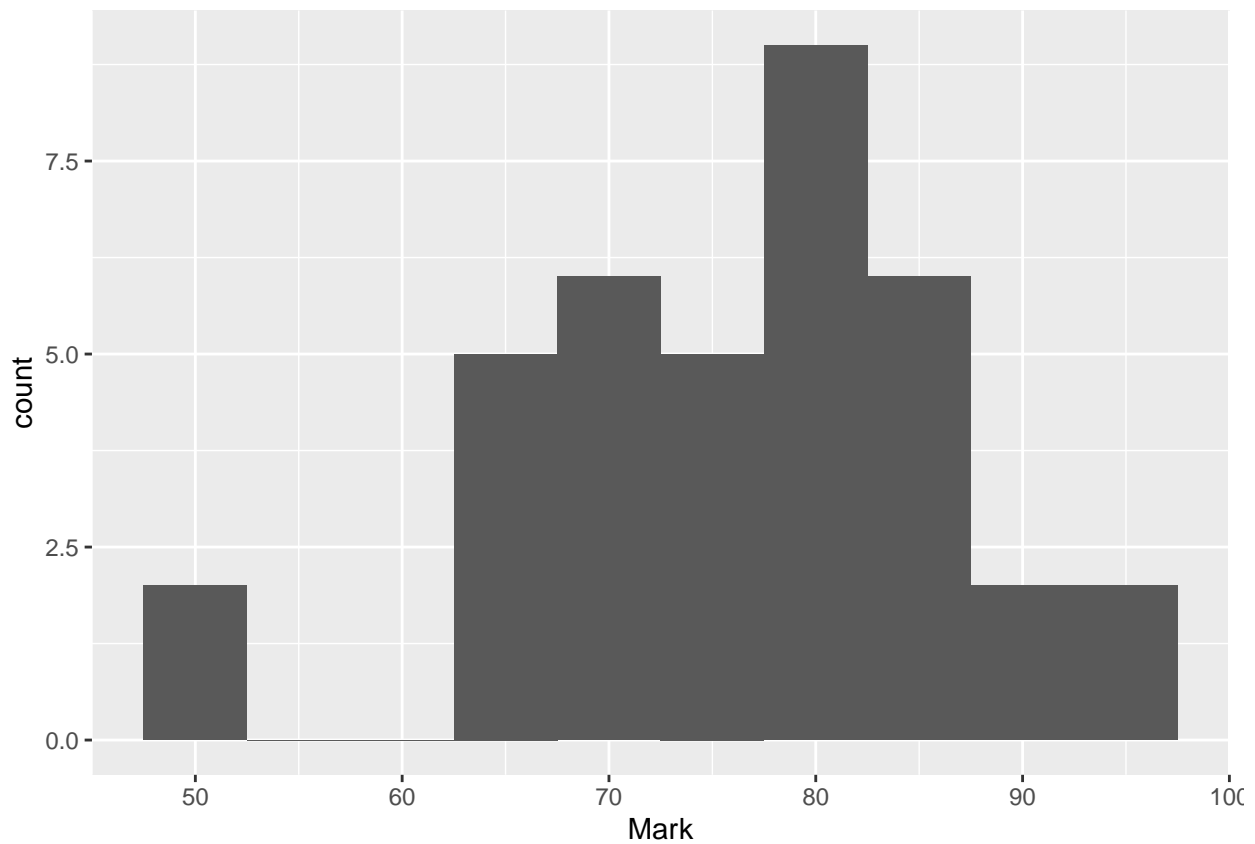
# Mark Analysis

Hanxiao Du

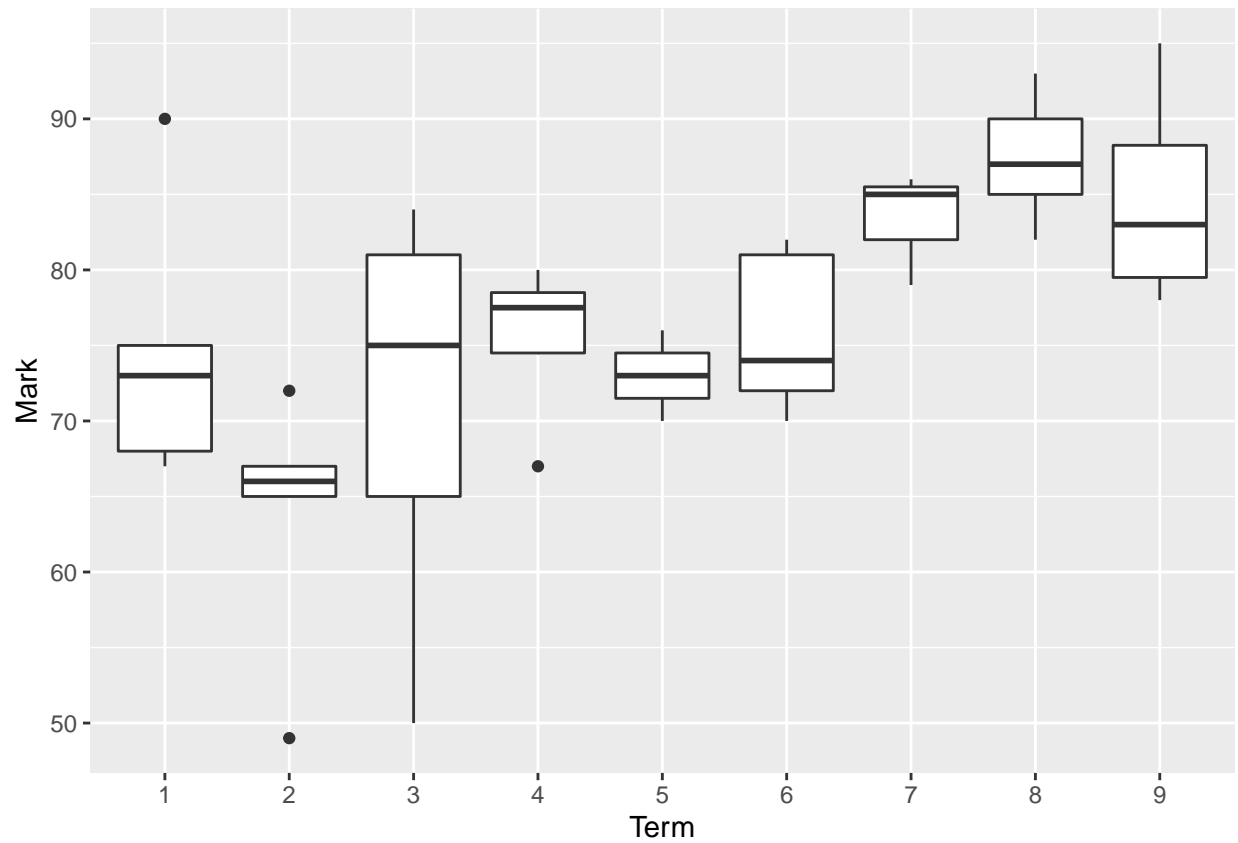
#An analysis of my marks throughout all semesters in university.

## Visualization

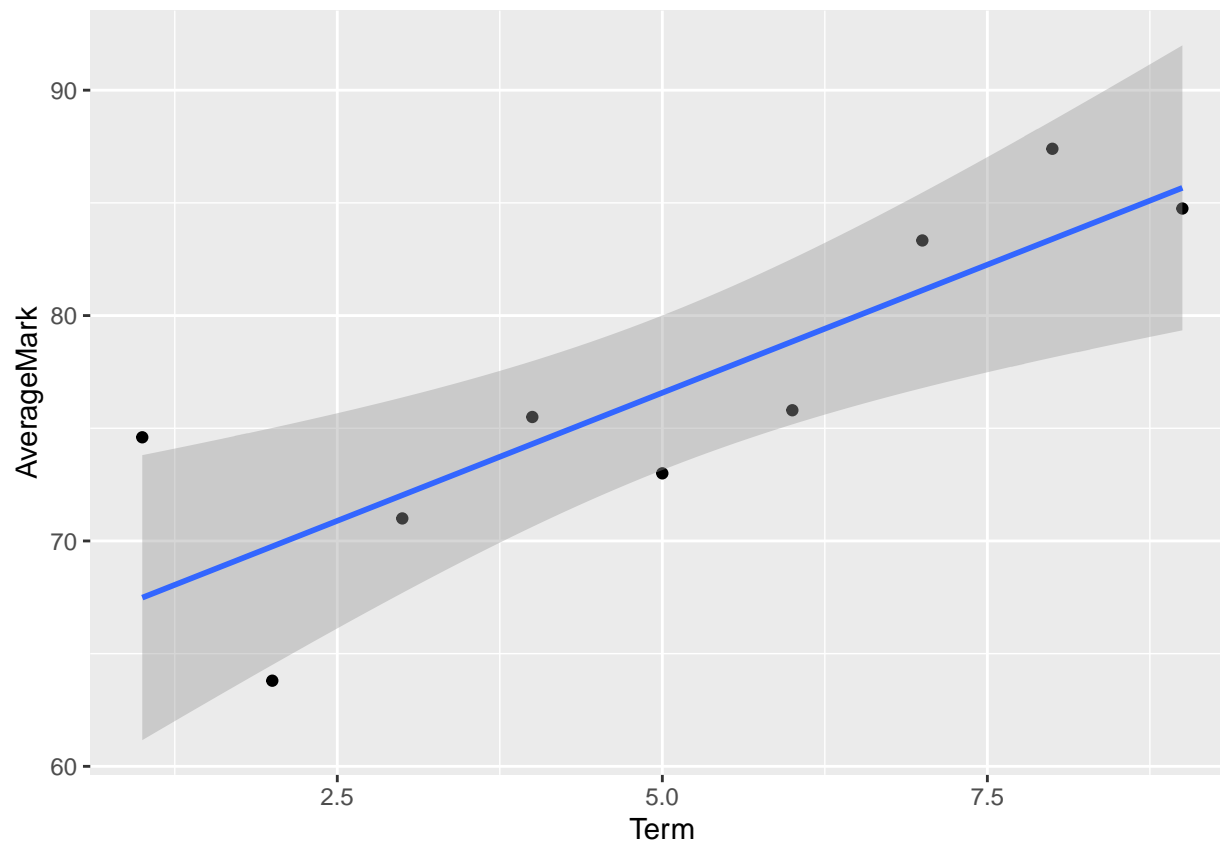
```
library(dplyr)
library(ggplot2)
# read marks from csv
all_courses = read.csv("marks.csv")
# remove courses with unknown marks
marks = all_courses[!is.na(all_courses$Mark),]
# histogram of all marks
ggplot(marks, aes(x=Mark)) + geom_histogram(binwidth = 5)
```



```
# boxplot for each term
ggplot(marks, aes(x=as.factor(Term), y=Mark)) + geom_boxplot() + xlab("Term") + ylab("Mark")
```



```
terms = marks %>% group_by(Term) %>% summarise(AverageMark=mean(Mark))  
# scatterplot of average mark of terms  
ggplot(terms, aes(x=Term, y=AverageMark)) + geom_point() +  
  geom_smooth(formula=y~x, method="lm")
```



```
cor(terms$AverageMark, terms$Term)
```

```
## [1] 0.8369944
```

## Regression Analysis

```
linear_model = lm(AverageMark~Term, data=terms)
summary(linear_model)
```

```
##
## Call:
## lm(formula = AverageMark ~ Term, data = terms)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.958 -3.049 -0.917  2.212  7.115
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   65.2120     3.1604  20.634 1.58e-07 ***
## Term           2.2728     0.5616   4.047 0.00489 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.35 on 7 degrees of freedom
## Multiple R-squared:  0.7006, Adjusted R-squared:  0.6578
```

## F-statistic: 16.38 on 1 and 7 DF, p-value: 0.00489