Problem 1

1)

The last digit of the zip code				

Exterior	1	2	3	4
Brick	4	10	4	30
Frame	1	1	5	1
other	1	2	7	3

Interpretation: There are a lot of brick exterior type houses when compared to others. Zip code 1 area have fewer houses.
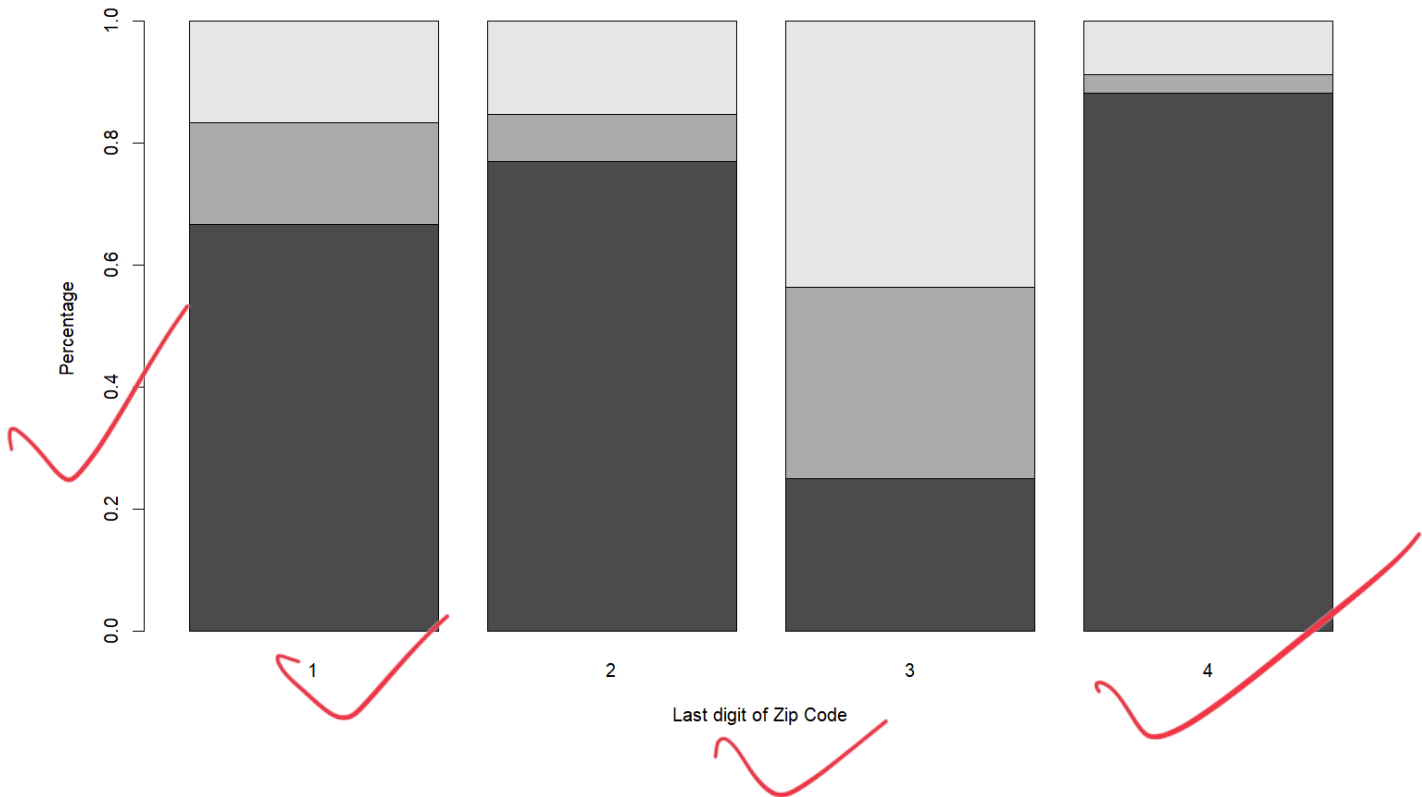
2)

The last digit of the zip code				

Exterior	1	2	3	4
Brick	0.66666667	0.76923077	0.25000000	0.88235294
Frame	0.16666667	0.07692308	0.31250000	0.02941176
other	0.16666667	0.15384615	0.43750000	0.08823529

use 2 or 3 decimal digits

3) −1

Bar plot for percentage distribution of Exterior Type by Zip Code Last Digit
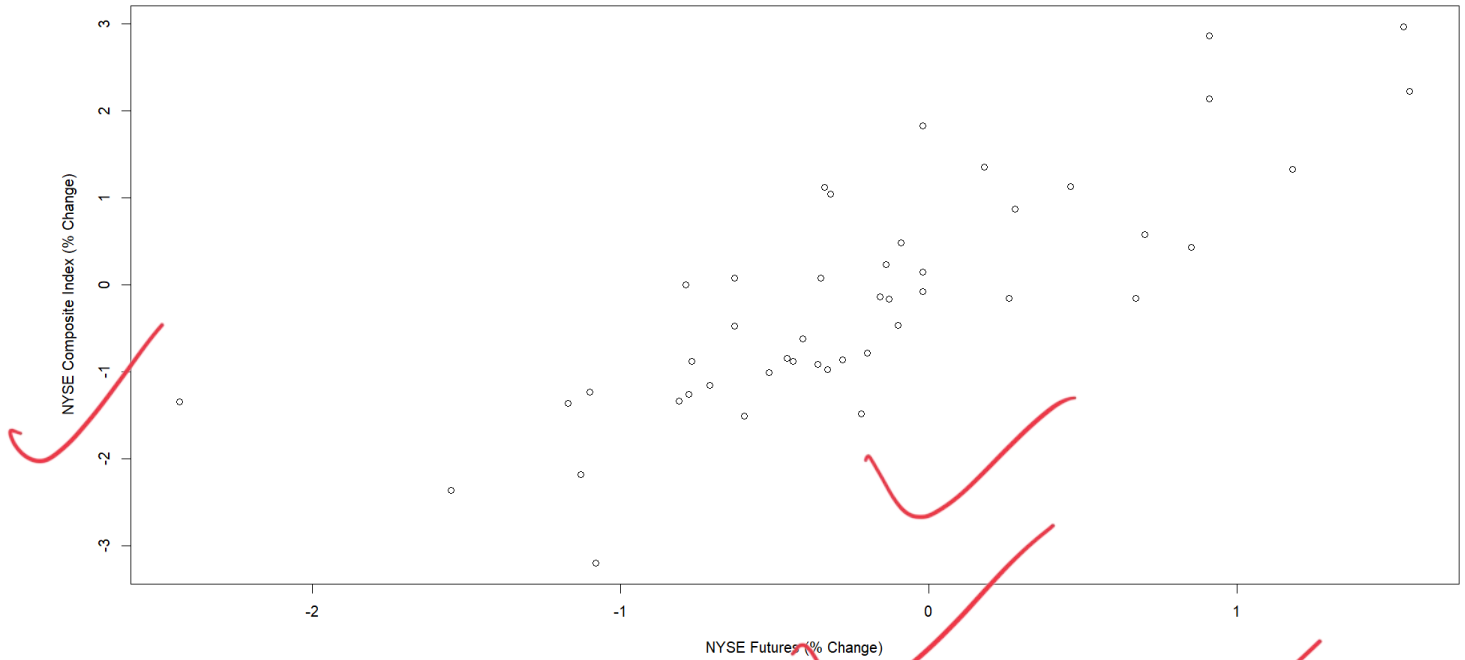

Percentage

Last digit of Zip Code

4)

Features: In all area zip codes except 3 there are more brick exterior type houses than others. In zip code 1 exterior brick type is high and frame and other are the same percentage. In zip code 2 the exterior brick type is more followed by other and then frame type. In zip code 3 other types are more when compared to brick and frame. In zip 4 exterior brick type is most of it and the other are very less.

Problem 2

Scatter Plot of NYSE Composite Index vs. NYSE Futures

The scatter plot shows a positive linear trend between NYSE Futures(% Changes) and NYSE Composite Index(% Change). This indicates that percentage change in NYSE Futures increases, the percentage change in the NYSE Composite Index also tends to increase. Most data points lie in the range between -1 and 0.

3)

Boxplot of Exam Scores by Student Status
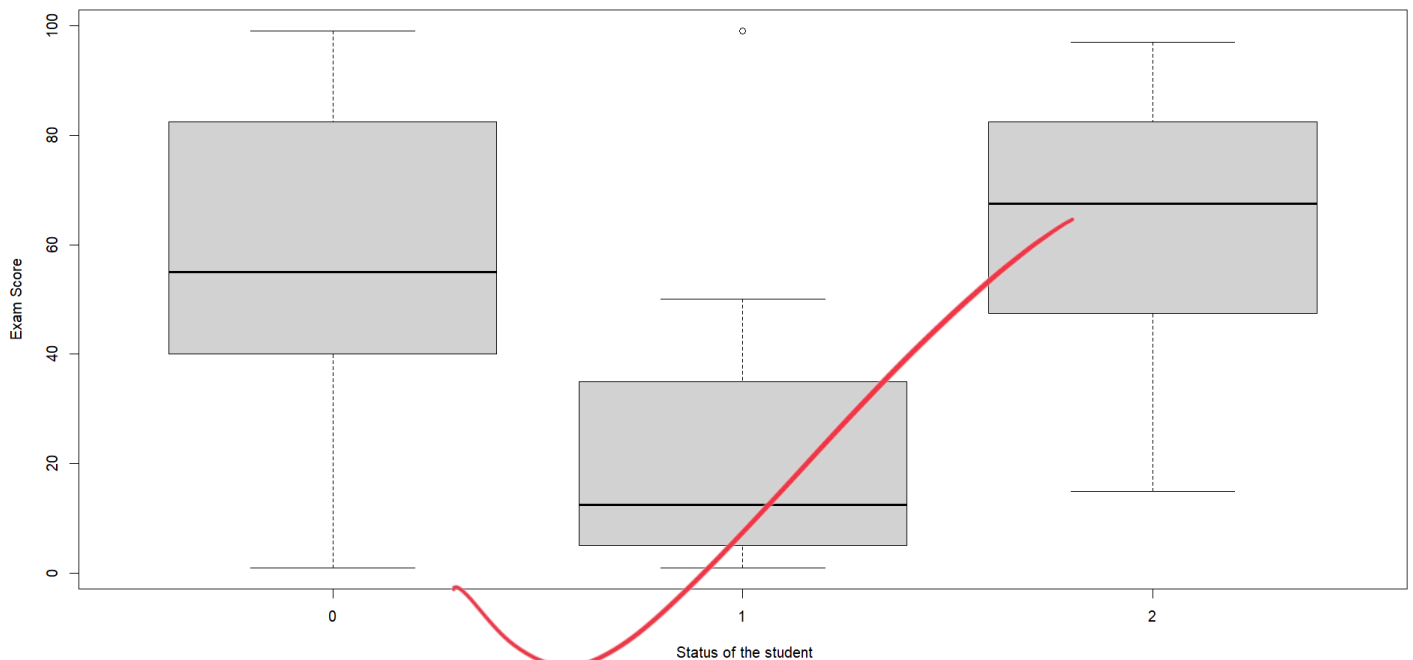
the status of the student:
coded as 0 = passed the course
1 = failed the course

2 = dropped out before the semester was over

Features:

Status 0
Median is around close to 60, IQR roughly 40 to 80
students who passed have the highest scores, with their median and upper quartiles being well above the other group 1. High variance when compared to others.

Status 1
Median around 20, IQR roughly above zero to below 40. One outlier is present.
Students who failed performed poorly, with scores concentrated at the lower end.

Status 2
Median around above 60, IQR 40 to 80
Students who dropped out have similar scores compared to students who passed. There might be several reasons students drop out.If they continued, there may be a chance that they would have passed given the scores in the test.

Given the test scores if the student has a high score they will pass or might drop out but given the student score is less they might fail.
R Code:

```
# problem 1
# 1.1
texas_house_data <- read.csv("texas-house.csv")
a <- table(texas_house_data$exter, texas_house_data$zip)
a


# 1.2
b <- prop.table(a, margin = 2)
b


# 1.3
barplot(b, main = "Bar plot for percentage distribution of Exterior Type by Last Digit of Zip Code", xlab = "Last digit of Zip Code", ylab = "Percentage")


# Problem 2
nyse_data <- read.csv("nyse.csv")
plot(nyse_data$future, nyse_data$index,
    main = "Scatter Plot of NYSE Composite Index vs. NYSE Futures",
    xlab = "NYSE Futures (% Change)",
    ylab = "NYSE Composite Index (% Change)")


# Problem 3
placement_test_data <- read.csv("exam.csv")
boxplot(
  placement_test_data$score ~ placement_test_data$status,
  main = "Boxplot of Exam Scores by Student Status",
```

```
  xlab = " Status of the student",
  ylab = "Exam Score"
)
```