Homework 5

1. (1 points) Load the Car Road Tests dataset (`data("mtcars")`, `?mtcars`), then create a new column for `mtcars` named as `mpg_2`. This new column will categorize "mpg" into four categories using the thresholds below:

mpg_2 category	Thresholds
Low	mpg < 16
Low_intermediate	$16 \leq$ mpg < 21
Intermediate_high	$21 \leq$ mpg < 26
High	$26 \leq$ mpg

```
mtcars$mpg_2 <- ifelse(mtcars$mpg < 16, "low",
                    ifelse(mtcars$mpg < 21, "low_mediate",
                        ifelse(mtcars$mpg < 26, "mediate_high", "high")))
mtcars$mpg_2
```

```
##  [1] "mediate_high" "mediate_high" "mediate_high" "mediate_high"
##  [5] "low_mediate"  "low_mediate"  "low"          "mediate_high"
##  [9] "mediate_high" "low_mediate"  "low_mediate"  "low_mediate"
## [13] "low_mediate"  "low"          "low"          "low"
## [17] "low"          "high"         "high"         "high"
## [21] "mediate_high" "low"          "low"          "low"
## [25] "low_mediate"  "high"         "high"         "high"
## [29] "low"          "low_mediate"  "low"          "mediate_high"
```

2. (1 points) Make a boxplot to show the `mpg` values for each level of `mpg_2`. Add horizontal dashed lines to depict the thresholds in question 1.

```
boxplot(mpg ~ mpg_2, data = mtcars)
abline(h = c(16, 21, 26), lty = 2)
```
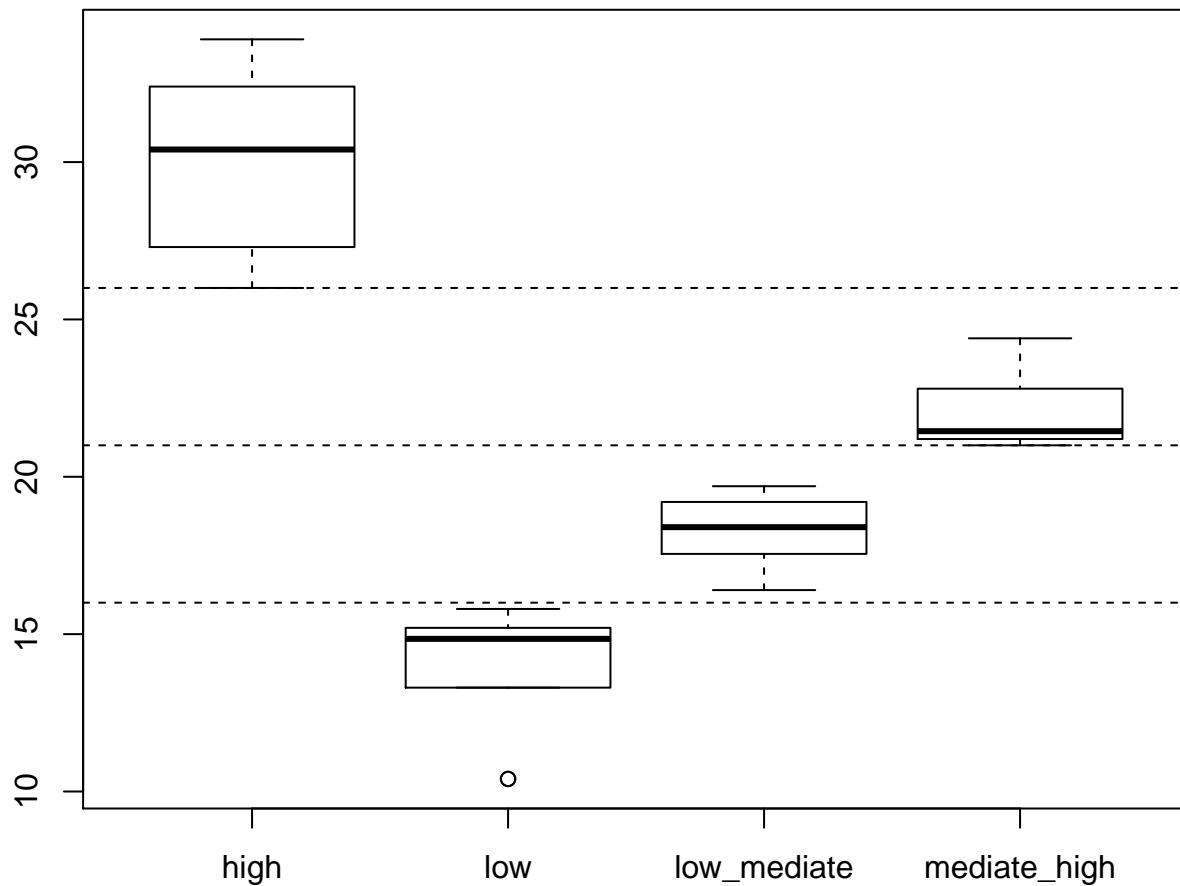
3. (3 points) Make a scatter plot to show the raw `mpg` values and use `mpg_2` to color points. (Hint: `?plot` and you will find that `y` argument is *optional*.) Again add horizontal dashed lines to depict the thresholds in question 1.

```
plot(x = mtcars$mpg, col = as.factor(mtcars$mpg_2))
abline(h = c(16, 21, 26), lty = 2)
```
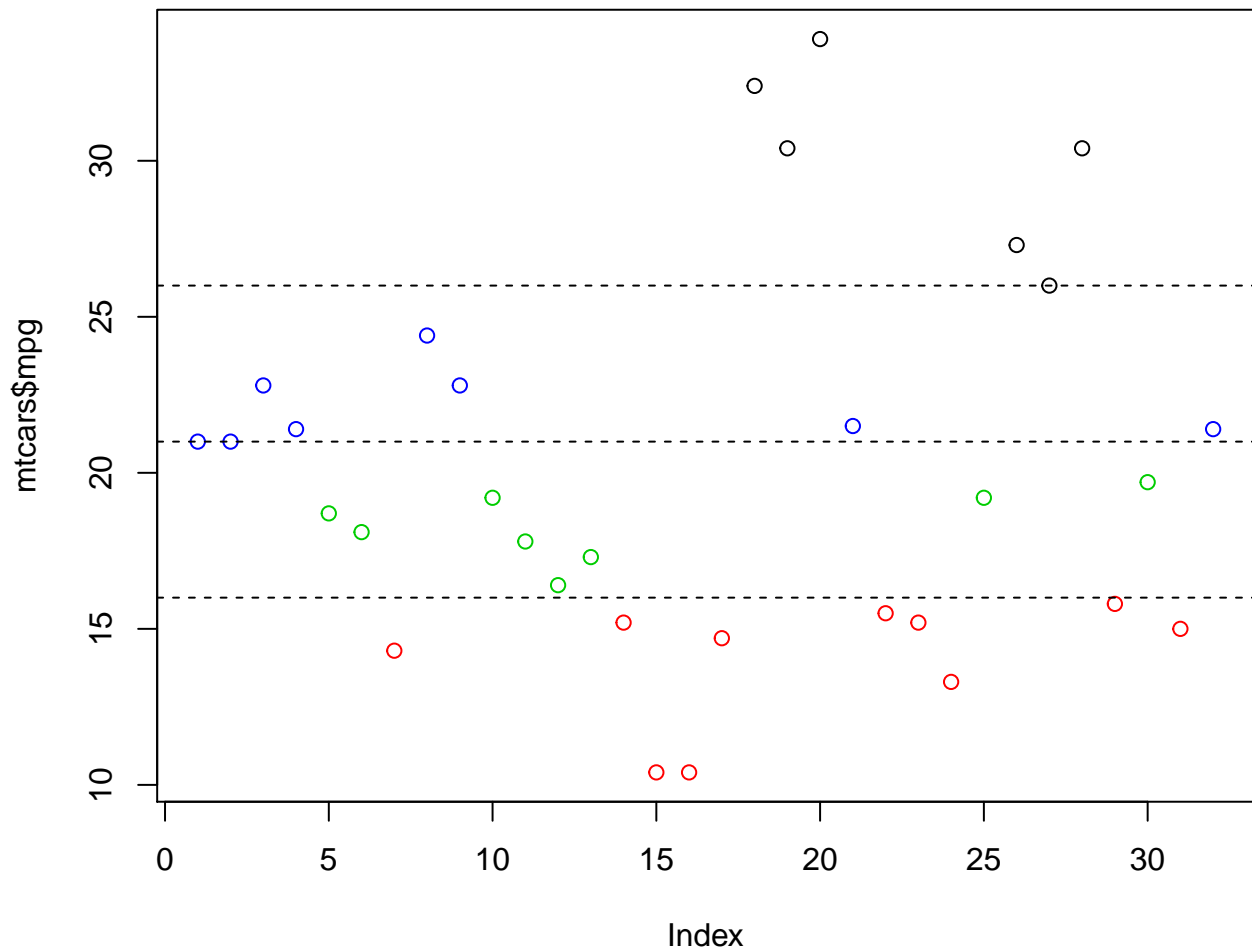
4. (1 points) Load the Earthquakes dataset (`data(quakes)`, `?quakes`), then create a new column called `depth_2`. This new column will categorize `depth` into 3 categories using the thresholds below:

`depth_2` category	Thresholds
1	depth < 150
2	$150 \leq$ depth < 500
3	$500 \leq$ depth

```
x = quakes
x$depth_2 = ifelse(x$depth < 150, 1, ifelse(x$depth < 500, 2, 3))
x$depth_2[1:100]
```

```
##   [1] 3 3 1 3 3 2 1 2 2 3 3 2 3 3 1 2 1 3 3 3 3 2 3 1 2 3 2 2 3 2 3 1 2 2 1
##  [36] 3 2 3 3 2 2 1 3 1 2 1 2 1 2 1 3 2 1 3 2 3 2 3 3 3 3 2 3 1 3 2 3 3 3 1
##  [71] 2 2 1 3 3 3 2 2 2 2 1 2 3 2 2 2 1 3 1 1 1 2 2 2 2 2 2 2 1 2
```

5. (1 points) Create a new column called `mag_2`.
   - If the `mag` value is < 4.5, assign `mag_2` to "green";
   - If the `mag` value is $>= 4.5$ but < 5.0, assign `mag_2` to "blue";
   - If the `mag` value is $>= 5.0$, assign `mag_2` to "red".

```
x$mag_2 = ifelse(x$mag < 4.5, "green",
                 ifelse(x$mag < 5.0, "blue", "red"))
x$mag_2[1:100]
```

```
##   [1] "blue"   "green"  "red"    "green"  "green"  "green"  "blue"   "green"
##   [9] "blue"   "green"  "green"  "blue"   "green"  "green"  "red"    "green"
##  [17] "red"    "blue"   "green"  "green"  "blue"   "green"  "green"  "blue"
##  [25] "red"    "green"  "blue"   "red"    "blue"   "green"  "blue"   "blue"
##  [33] "blue"   "green"  "blue"   "green"  "blue"   "blue"   "green"  "green"
##  [41] "blue"   "blue"   "green"  "green"  "blue"   "blue"   "blue"   "green"
##  [49] "green"  "red"    "green"  "green"  "blue"   "green"  "green"  "green"
##  [57] "blue"   "green"  "green"  "green"  "blue"   "green"  "red"    "blue"
##  [65] "blue"   "blue"   "green"  "red"    "blue"   "red"    "green"  "blue"
##  [73] "green"  "red"    "green"  "green"  "green"  "blue"   "green"  "red"
##  [81] "red"    "blue"   "blue"   "green"  "green"  "green"  "blue"   "blue"
##  [89] "green"  "green"  "red"    "blue"   "red"    "green"  "green"  "green"
##  [97] "blue"   "red"    "red"    "blue"
```

6. (3 points) Plot the earthquakes, with longitude in the x-axis and latitude in the y-axis. Use `pch = 16` for *shape* of points; use the `depth_2` variable to define the *size* of the points; and use the `mag_2` variable to define the *color* of the points. Does the magnitude of earthquakes appear to be larger for deeper ones?

```
plot(x = x$long, y = x$lat, pch = 16, cex = x$depth_2, col = x$mag_2)
```
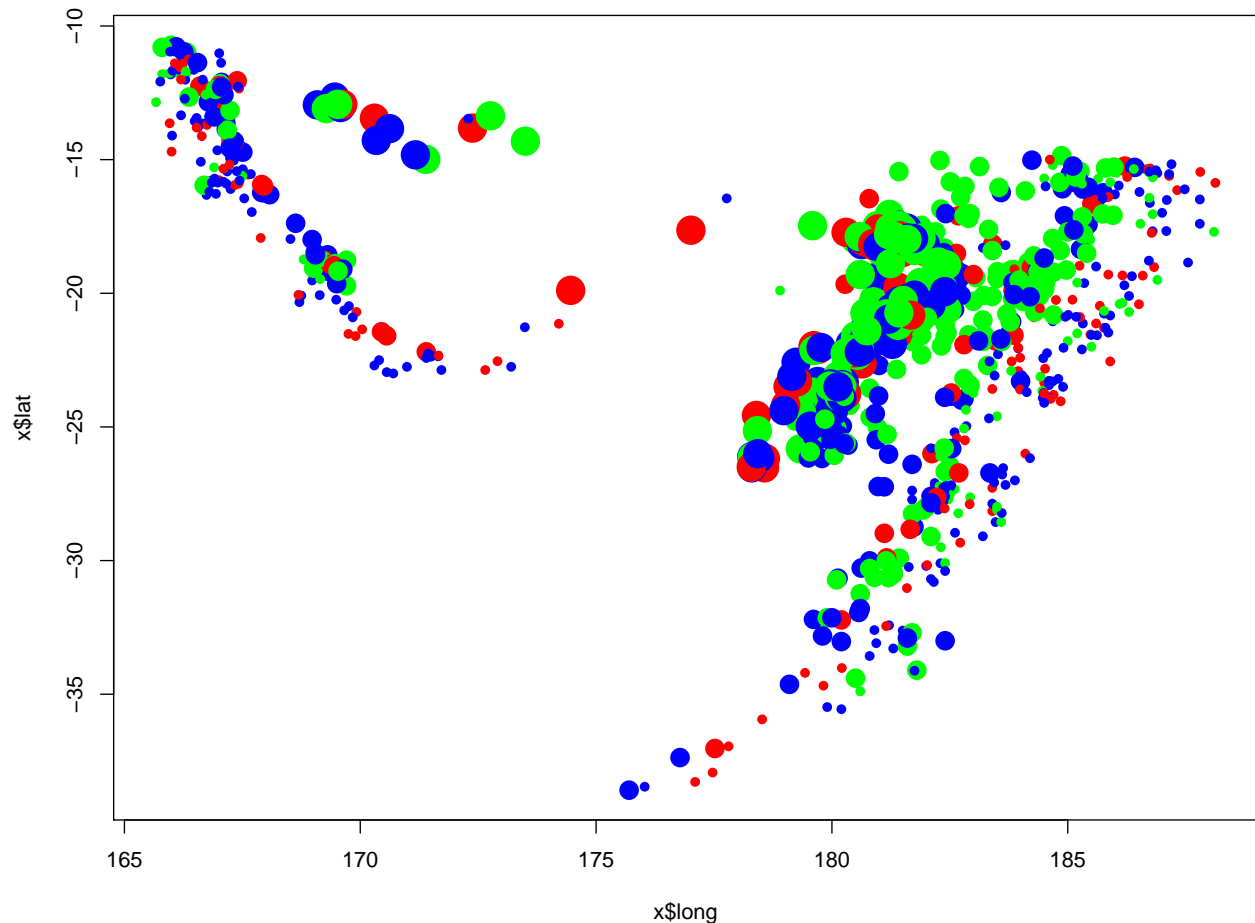

If the depth and magnitude of earthquakes are correlated, we should see that larger points will be mostly red. However, this seems not the case based on the above plot. Let's test the

4

correlation between them.

```
cor.test(x$depth, x$mag)
```

```
##
##  Pearson's product-moment correlation
##
## data:  x$depth and x$mag
## t = -7.488, df = 998, p-value = 1.535e-13
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2885057 -0.1710909
## sample estimates:
##        cor
## -0.2306377
```

We can see desipite that the p value is significant, the correlation between depth and magtitue is fairly small (**-0.23**). The small p value is not surprising given the large sample size (**1000**) we have. Therefore, we confirmed the pattern from the plot that the magtitude and depth of earthquakes in this dataset is weakly correlated.