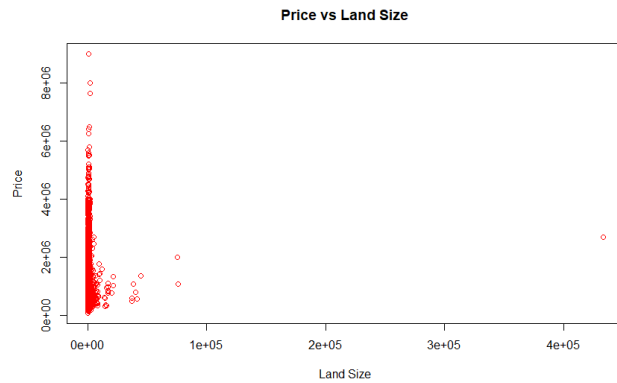
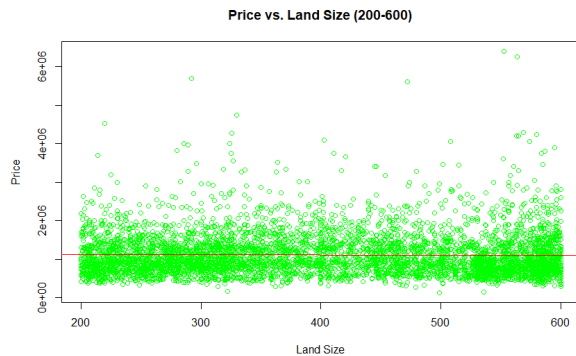


1.



I observe that there is no visible correlation between land size and price at this scope. It is noticeable that there are some outliers in price and land size.

2.



Surprisingly, landsize is not a good predictor of land size as seen by this graph and the r^2 value of 0.000201

3. I found that "Southern Metropolitan"

Has the houses with the highest mean price. From the previous data I can only conclude that the Southern Metropolitan area must have other factors that are not captured by the regression model that could affect price, such as being a beachfront property or the area it is located.

4. `CleanHousing$PricePerArea <- CleanHousing$Price/CleanHousing$LandSize`

5. One Sample t-test

data: cleaned data

t = 5.151, df = 11640, p-value = 2.633e-07

alternative hypothesis: true mean is not equal to 2884.718

90 percent confidence interval:

```
3211.768 3518.662
sample estimates:
mean of x
3365.215
```

6.

T.test results:

Results in a table:

	Region	Result
1	Northern Metropolitan	2.060909e-01
2	Southern Metropolitan	1.969583e-04
3	South-Eastern Metropolitan	1.285897e-66
4	Northern Metropolitan	2.060909e-01
5	Southern Metropolitan	1.969583e-04
6	South-Eastern Metropolitan	1.285897e-66
7	South-Western Metropolitan	NA
8	Western Metropolitan	1.937368e-67
9	Eastern Metropolitan	4.588998e-01
10	Northern Victoria	3.930230e-31
11	Western Victoria	1.533473e-34
12	Eastern Victoria	6.594132e-38
13	Southern Victoria	NA

What we can learn from these results:

Several regions, including Southern Metropolitan, South-Eastern Metropolitan, and Northern Victoria, exhibit significant differences in their mean PricePerArea compared to the overall mean, with p-values considerably smaller than 0.05. However, some regions, such as Northern Metropolitan, do not show significant differences. Overall, the aggregated result indicates that the PricePerArea across the entire dataset is significantly higher than the overall mean. This finding is further supported by the individual regional tests, where the majority of regions show statistically significant differences in their means in the previous set of data.

7. `CleanHousing$PropertyTax1 <- sapply(CleanHousing$PricePerArea, compute_tax)`

8.

```
> # Tax based on Price Per Area
> compute_tax2 <- function(PPA) {
+   if (PPA <= 3500) {
+     return(PPA * 0.01) # 1% tax for PPA <= 3500
+   } else if (PPA > 3500 && PPA <= 15000) {
+     return(PPA * 0.0115) # 1.15% tax for 3500 < PPA <= 15000
+   } else {
+     return(PPA * 0.015) # 1.5% tax for PPA > 15000
+   }
+ }
```

9. It seems that you will be losing money with the new tax scheme, about -145434.3 AUD.