Simple Regression Project

1. Load housing data (house_data.csv) from moodle into R.
2. Explore the data
3. Make sure to check if there are any NA's
4. Remove the columns – lat, long and view which do not contribute to the model.
5. Take a subset of data if you can't load everything may be 5000 rows
6. Use the glimpse function to explore the data
7. Get a summary of the data
8. Plot a histogram of price distribution
9. Plot a histogram of number of bedroom distribution
10. Get a count of frequency of houses with the number of bedrooms. Basically how many houses with 1 bedroom, 2 bedrooms and so on.
11. Get a box plot to see how the price and number of bedrooms are associated
12. Get a box plot to see how the price and number of bathrooms are associated
13. Plot price against squrefeet
14. Plot price against number of bedrooms
15. Create a linear regression model with all the variables
16. Identify the significant variables.
17. Explain the coefficients. Which variable has high impact
18. Plot the correlations using ggcorrplot.