GEOG 5680 Project 1: House Prices

Simon Brewer

May 14, 2020

Introduction

The excel file *homeprice.csv* contains information about homes that sold in a town of New Jersey in the year 2001. We are interested in understanding the relationship between house characteristics and the sale price. The file contains the following information:

- sale: sale price of house
- list: list price of house
- full: number of full bathrooms
- half: number of half bathrooms
- bedrooms: number of bedrooms
- rooms: number of non-bedrooms
- neighborhood: neighborhood rank (1: poor; 5: rich)

Analysis

1. Using this file, explore the relationship between the sale price and the other variables using scatterplots, histograms and/or boxplots. Identify those variables that appear to have the strongest relationship with sale price.
2. Now use these variables to build a multiple linear regression model to explain the sale price. Use the `summary()` function to find the coefficients and goodness-of-fit of the model. Use the `anova()` function to identify which variable appears to have the greatest effect on sale price. Remember to look at the distribution of residuals.
3. Build a second model using the same variables to explain the list price. Use the `anova()` function to identify which variable appears to have the greatest effect on list price. Are there differences from the sale price? Could you use this information to recommend which characteristic of a house a real estate agent should concentrate on?
4. Finally, what is the effect of neighborhood on the difference between sale price and list price? Do nicer neighborhoods mean it is more likely to have a house go over the asking price?

Report

You should submit the results of your analysis as an HTML document generated by R Markdown. In addition to the results, this document should contain a brief description of the steps you took, and what the results mean.
