Lorena Karine Sousa dos Santos, ID 100404805

Link: https://github.com/lorenaksantos/HousePrices

Summary of House Price Dataset

This is a data set of about 1,460 houses. Each row is a different house and each column tells you something about the house, like how many bedrooms it has or how much it sold for. The information includes things like the size of the property, how many floors the house has, and if it has a pool or fence. There are also some categories like the type of street and what kind of zoning the house is in. The last column tells you how much each house sold for.

It has data from 2006 to 2010

Exploring 1: Checking the correlation between the size of the house and its price

The scatterplot shows that there's a 0.70 coefficient of correlation between the size of the house and its price, which indicates a strong correlationship between then, assuming that the bigger a house is, the most expensive it gets

Exploring 2: Evolution of the price of the houses according to the year

The housing market got its peak on 2007, having a avarege of $$186k. Then it decresed to an average of $177k in 2008. The house prices went from the year with the highest avarage (2007) to the year with the lowest average (2008).

Exploring 3: Price Distribution

The distribution of the housing prices represents a right skewer graphic, meaning that the prices are more concentrated left. It means that the prices were not evenly distributed and most of the prices were in the oldest years (especially 2007) and it got cheaper in recent years (2008-2010). There are needed of more data to find out why it happened.