

Visualization of Complex Data Final Project

Shiyun Mai, Hao Tang
G42940068, G34199044

Environment set up:

Software applied: mongoDB, Highcharts, Express, Ejs, body-parser, Node js, Javascript

What we done:

1. set up Nodejs server. Connect to MongoDB
2. Set up express as router
3. Obtained data from mongodb in nodejs, applied Ejs to pass data to specific html page.
4. Create specific html page

Dataset and Graphs description:

Source:

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

Dataset description:

This application presents a data set describing the sale of individual residential property in Ames, Iowa from 2006 to 2010. The data set contains 2930 observations and a large number of explanatory variables involved in assessing home values, such as sale price, the condition of sale, the house style and the area of house. In order to directly exploring this house data set, several statistical plots will be shown in the future analysis.

Graphs description:

The style of dwelling is significant feature for house price. Eight house styles are exploring in this data set. Bar plot for House Style by Year shows the number of house for each house style from 2006 to 2010. This bar plot indicates that the most two popular house style in Ames are one story house and tow story house. By clicking each year on the legend, the distribution of each house style for each year will be shown. The same conclusion that one story and two story house are popular in Ames will be drawn after you see the house style bar plot for each year.

To be in a further exploration, bar plot Average Sales Price for House Styles is displayed below. The highest average sale price is the two and a half stories with finished second level house. And the second and third highest average house price are two story house and one story house. It is interesting that the average sales price for two and a half story with unfinished second level house is lower than the one story house. By clicking the columns, the time series for average sales price for each month from 2006 to 2010 will be displayed.

The total home area is also an important feature affecting the house sale price. The scatter plot Total Area versus Sale Price by Sale Condition implies that there is a positive correlation between house sale price and house total area but not very significant. By clicking each sale condition on legend, the distribution of sale cases for different sale conditions will be shown.