ECS 163 Project 2 Report

Visualizing Housing Prices

Jianfu Situ

Data Source: https://www.kaggle.com/c/house-prices-advanced-regression-techniques

The interactive system is based on a dataset from Kaggle, which contains various attributes of houses sold between 2006 and 2010 in Ames, Iowa. Some notable attributes include sales price, lot area, year/month of sale, etc. The overview of the system is a heatmap that encodes the distribution of sales count by year and month of sale. A tooltip of the exact sales count appears when a user mouseovers a tile. A user can interact with the heatmap by clicking on a tile to bring up two additional views, a parallel coordinates chart and a bar graph.

The parallel coordinate chart encodes a few attributes for the data entries that a user clicks on. These attributes chosen because they seem to have high correlation with sales price, the attribute of focus in this system. As such, parallel coordinates are useful at identifying trends across multiple attributes. In addition, the paths in the chart are color coded according to the sales price of the house entry.

The bar graph is a histogram of the sales prices for the tile that a user clicks on. The x-axis encodes the price bins, which is fixed so that a user can compare the distribution of prices across different tiles; the y-axis encodes the count of house sales in a price bin.

There are several interesting findings upon interacting with the system. The heatmap indicates that house sales tend to be the highest in the months of June and July. The parallel coordinates show a generally positive correlation between sales price and number of bedrooms, lot area, and year built. Finally, the histogram suggests that the most popular price range for sales is around $100,000 to $200,000, while only a few outliers have a sales price of over $500,000.

Student Feedback

Student Name: Vincent Yeh

Student Comments:

The heat map is very clear in visualizing house price. By navigating through each month, I can see both a histogram describing the cost, and a parallel coordinate describing features such as: bedroom, lot area, quality, year built, and price. The visualizations used perfectly fits what is needed to be convey. The issue I have is the lack of information on where these houses are located, and what colors mean on the parallel coordinates.

Follow-Up

Upon receiving the student's feedback, I felt that the criticisms were warranted. I followed the advice of the student to include the location of the houses contained in the dataset for better context, and a legend for the color encodings in the parallel coordinates.

After some extensive testing, I did not find any glaring bugs or unexpected behaviors in the interactive system. One limitation is that there seems to be a lack of intuition that the tiles of the heatmap can be clicked on without explicit instructions to the user. Possible future improvements to the system would be to allow user to click on a year or month to display aggregated details for that row/column, and to change the attribute being encoded by the heatmap (e.g. each tile can represent average lot area).