

Shiny App For Ames Housing Sales

Jiang He, Chaohui Yuan, Haiyang Zhang

Motivation

Real estate market is booming year by year, especially for properties in big cities or for houses in some campus towns, like Ames, Iowa. Most people buying a new house without helps from real estate agents. They would search online to see the house information, including listing price, location, house style, etc. All the information listed online are for current housing lists, however many people would like to see historical data of real estate market in their city before they make decisions of buying a new house.

Current online resources are not enough to meet people's needs. There should be a platform that people can search and learn the housing sales histories in certain cities for certain years. This project makes a Shiny app to help people meeting this goal, to build interactivities between information and app users.

Why Shiny App?

We used Shiny app to show our results for the following three reasons.

Firstly, Shiny app is not an isolated display which you can only see one or a part of information from it. Instead, Shiny app collects a massive of data and graphs together into a single application. People can reach all these information by simple clicks. They don't need to spend tons of time on web surfing.

Secondly, Shiny app provides interactions with users. Users can choose what they would like to see in a minute. We can provide selections on color and size which users may like choose by themselves.

Thirdly, and especially for our project, we want to create this app for searching historical housing information in Ames for potential house buyers. They can check the sale price of house sold in Ames in 2016 which matches the features of their dream house. For example, people can locate information by selecting the building type, house style, number of bedrooms, year of built and living areas of house.

Data description

Data processing and Data variable description

Story Country, IA/City of Ames data was downloaded from the Beacon website by setting **Sale Date** to be 18 month and filtering **Bedrooms** from 0 up to 8 one by one. Table 1 shows the original variables in the data set downloaded from the website.

| Variables | Description | Range |
|--------------------|-----------------------------------|---|
| Parcel ID | Unique ID for each house | |
| Address | House Address | |
| Sale Price | House Price in dollars | \$29 ~ \$830K |
| Sale Date | Sale Data | 1/1/16 - 12/1/16 |
| Sale Condition | Sale Condition | 6 levels characters: Abnorml, AdjLand, Alloca, Family, Normal, Partial |
| Assessed Value | Assessed price | \$300 ~ \$707.3K |
| Parcel Type | House type | 4 types characters: vacant land, dwelling and interior lot, condominium unit building and interior lot, single-family dwelling |
| Year Built | House built | |
| Total Living Area | Living area | Numeric: 0 - 4615F |
| Finished bsmt Area | Finished basement area | Numeric: 0 - 2114F |
| Lot Area | Lot area | Numeric: 0 - 157600F |
| Bedrooms | Number of bedrooms | Numeric: 0-8 |
| Central Air | Weather the house has central Air | 2 levels characters: yea, no, no data |
| Fireplaces | Number of fireplaces | 0 - 3 |
| Occupancy | Building Type | 7 levels characters: single family detached(includes detached townhouses; condominium unit; townhouse:end unit; townhouse:inner unit; two-family conversion:originally built as one-family dwelling; duplex; no data |
| Style | Style | 9 levels characters: one and one-half story: 2nd level finished; one story; split foyer two story; split level; one and one-half story: 2nd level unfinished two and one half story: 3rd level unfinished; two and one-half story:3rd level finished; no data |
| Neighborhood | Neighborhood near the house | 34 levels characters |
| INSTRUMENTNUMBER | Deed book data and page | Numeric numbers |

Table 1: Variables description in the raw dataset downloaded from website

Then we processed those datasets as follows:

1. Got rid of all records before 2016 January, since we were more interested in the housing sale in 2016.
2. Created a new feature **house age** defined by using 2016 subtracting the year of house was built and then grouped this feature **house age** into 13 groups.
3. Added the longitudinal and latitudinal information using Google API into the data set.

4. Broken down Ames into Five areas(North, South, West and center) based on the **Neighborhood**. Table 2 shows how we defined this:

| Area | Neighborhood |
|--------|--|
| North | HaydnLk, Somerst, N Ames, NridgHt, StoneBr, Blmngtn, IOCondo, Veenker, Greens, Gilbert, NW Ames, Br'Dale, NPkVill, NoRidge |
| West | SawyerW, ClearCr, CollgCr, DakotaR, Edwards, Crawford, WllwCr1, S&W ISU, Blueste, Sawyer, MsCondo, Landmrk, WllwCr2 |
| South | GrnHill, Timber, Mitchel, MeadowV |
| Center | OldTown, IDOT&RR, BrkSide |

Table 2: Four areas with corresponding neighborhood

After the above processing, there are 952 observations left to the following visualization.

The Shiny App

The Shiny app contains four tabs: Sales by Month, Sales Overview, House Sale History Search, and Sales vs Location.

Tab 1: Sales by Month

The purpose of this tab is to show the housing sales by time. We have observations of housing sales history in Ames for the 2016 whole year.

Users are able to choose between bar chart (Figure 1) or pie chart to display the house sale amount in each month. Also, there is a box-plot (Figure 2) showing the house sale price range by month. From the box-plot, we can see the distributions of sale price for all twelve months are right-skewed, which indicates there are a lot large numbers (high price) appeared each month. The orange dot in the middle of each box shows us the average house price in each month. Seeing this average house price trend and comparing the median house price, we can conclude that the sale price of houses sold last year (2016) in Ames is not correlated with time.

Tab 2: Sales Overview

This tab gives users a general idea about housing sales before going to the search functions.

Users can see the relationship between Year of Built and Sale Price in the first graph (Figure 3). From the scatter plot below, we can see it clearly that newer houses cost more than older houses in general.

The second graph (Figure 4) in this tab shows the relationship between Sale Price and Assessed Value. It is a strong positive linear relationship with some outliers. This means for most houses in the dataset are assessed to accurate values, while some houses assessed too low or too high. We're wondering the reason behind this. From the plots below, most outliers are with assessed value smaller than 200,000 dollars and sale price larger than 200,000 dollars. To figure out the reasons of appearance of these outliers, we assigned colors by Sale Condition and add Year of Built into the texts when users hover over to a certain point. The graph shows all the interested outliers are in Partial condition and are built in either 2015 or 2016. We guess the sale condition of house probably act an important role in house assessing. At the same time, age of house makes a lot difference on sale price. Therefore, we have these interesting outliers with lower assessed value and higher sale price. We would like to have more information and do more investigation in the future.

Tab 3: House Sale History Search

There is the tab that users can actually do search on.

There are four widgets on the left: Area gives 4 areas (there is no sales in the east part last year) of Ames, building type gives what type of houses the users may want. Also, users can specify number of bedrooms and year built of the house they desire. We intentionally chose those four widgets since we have a friend who is buying house is interested in them, and we believe they are also the main features any house buyer would have strong preference on. Then, the plot gives sale price in thousand dollars versus the total living area in sq ft. As shown in (Figure 5).

Not surprising, the bigger the total living area is, the higher the sale price was. The points are colored by the number of bedrooms and shaped by different style of house, which is also a key feature for home buyers. Since the plot is loaded with information, the various functions of plotly come handy now. For example, the users can specify certain price range and how old the house would be, and then the plot will give a very good display of how many bedrooms, what building type and style you may get. Also, the users can hover over the plots to see detailed information about the house sold. It's designed to give users an informative feel of what housing features they can get for certain price range, which will better aid them on finding their dream house.

Tab 4: Sales vs Location

In this section, we were interested in how the house sale price changes at different locations.

Overall, average sale price in story county of year 2016 was \$180K - \$190K. Houses with average sale price are mostly located near Grand Avenue, 13 street and 24 street and near Ontario street. We noticed that the highest house sales on average were tend to centered near the northridge heights park area in 2016 (Figure 6). What's more, for most houses with higher sale price were recently built (the median built year for house with sale price greater than \$300K were 2003).

How to use the shinyApp

- Download the folder `AmesHoursePrice2016`
- Open `AmesHousePrice2016.Rproj`
- Open the shiny app file in scripts folder called `amesHousing16App.R`

Future Work

There are many things we would like to work on when time is allowed.

First of all, more data for major cities and a longer time span. We only have one year of sales in Ames right now due to the limitations of the website. If we can access a longer time span and more major cities in the U.S., we want to design the first page of our app to be the place the user choose the city and time they want to research on.

Secondly, more variables need to be included. Such as, crime rate, transportation convenience, hospital or school access. I think all these are not huge issue in Ames, but they would be for home buyers in big cities, so we want to make that available for the users.

Last but not least, we will include the definitions of some housing terms used in our app, for example, partial sale condition in tab 2 and building types in tab 3. Every time user hover over the legends, the definitions will pop up since they may not know them well.

Overall, we want to make our app more informative and user-friendly.

Figures

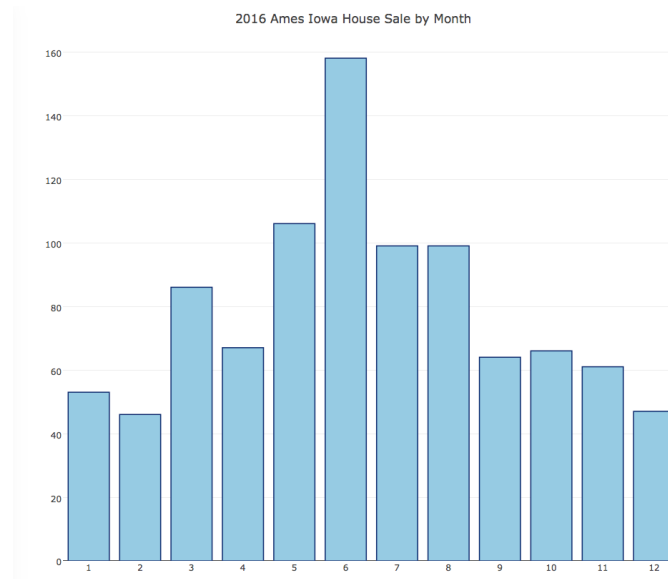


Figure 1: Sale Amount by month in 2016

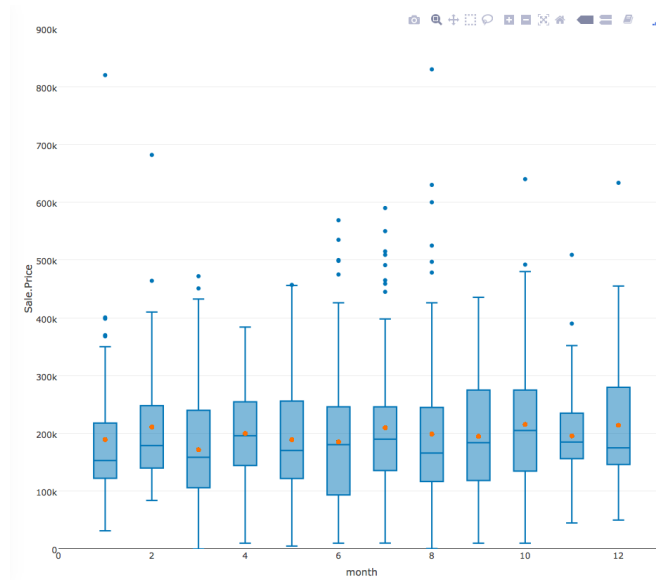


Figure 2: Price Range by month in 2016. Orange points are average price in each month

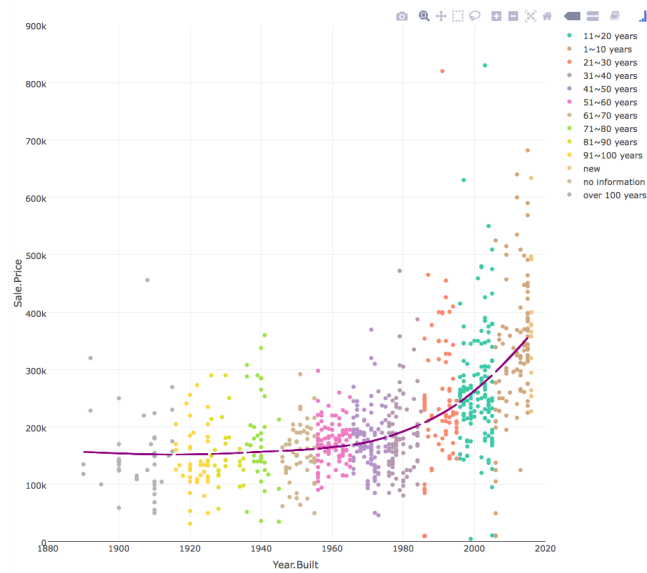


Figure 3: Sale Price by Year of Built

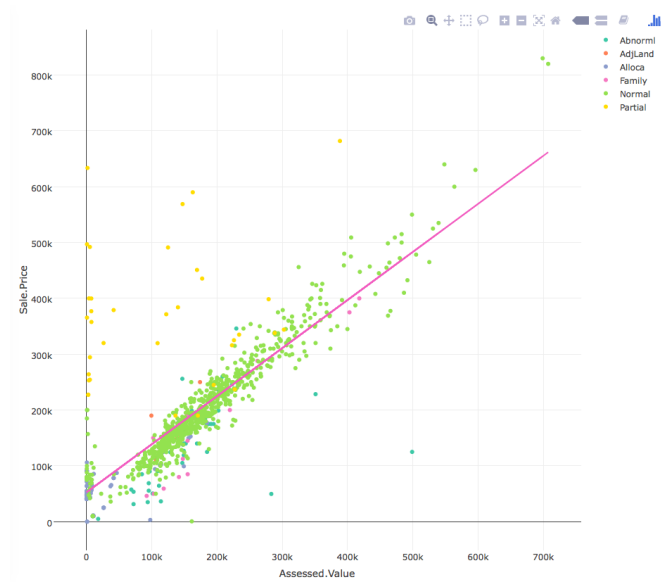


Figure 4: Sale Price by Assessed Value

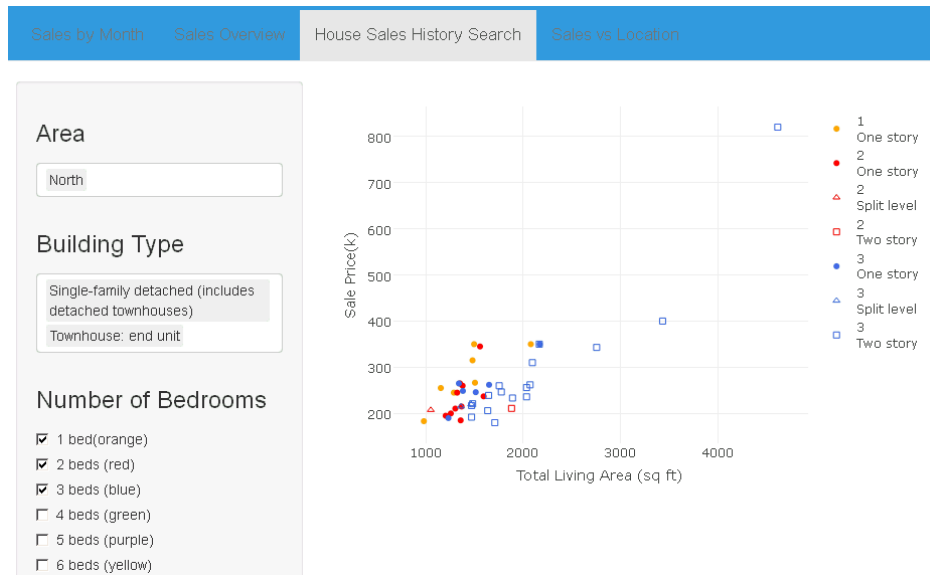


Figure 5: Search Page for House Features

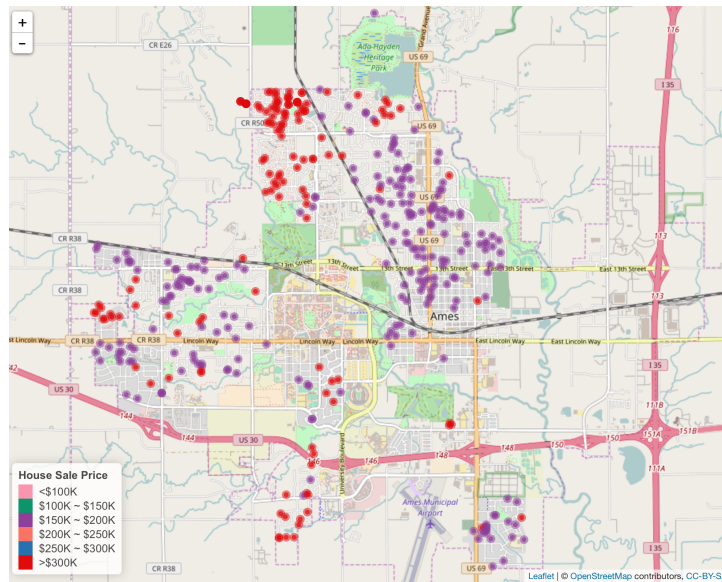


Figure 6: House sale price in 2016. Red dots represents the price larger than \$300K while the purple dots denotes the average price around \$200K in story county, IA.