2016

Real Estate Market Analyzer

Process Book

Shane Brown, ShaneJBrown@Gmail.com, u0852900
Sigmund Chow, sigmundccs@gmail.com, u0597938
My Huynh, my.huynh@utah.edu, u0729654

Project GIT repository:
https://github.com/shanejb/dataviscourse-pr-RealEstateMarketAnalyzer
Website: TBD

11/11/2016

Table of Contents

1. **BACKGROUND MOTIVATION/OVERVIEW** **2**
2. **RELATED WORKS** **2**
3. **QUESTION** **3**
4. **DATA** **4**
5. **EXPLORATORY DATA ANALYSIS** **6**
6. **DESIGN EVOLUTION** **7**
7. **IMPLEMENTATION** **9**
8. **EVALUATION** **12**

1. Background Motivation and Overview

Recently, we have been seeing many new constructions for condominiums and apartments around the city. Are millennials moving towards renting? Renting has long since been an acceptable way to pay for housing throughout the civilized world for many, many generations, but does renting make financial sense? Part of the American dream is to own your own home, but is the rising property prices pushing our young generation away from buying houses?

In a recent article in New York Times, it states "Today's young adults have not become homeowners at the same rate that earlier generations did. That probably reflects a mix of a weak economy — and thus poor job prospects during the initial aftermath of the recession — and the lack of affordable housing supply in many of the hottest markets combined with perhaps some cultural shift toward buying homes later or even not buying at all."[1] However, based on one of the census reports, there were more new homes sold in July than in nearly a decade. Buyers are purchasing single-family houses at annual rate of 654,000. This has motivated us to find out more about our current housing market and where we stand right now compared to the housing market crisis in 2008.

2. Related Works

There are several websites which show comparisons between rent and house costs but they don't visualize how fast house price increase in historical time as well as reveal the housing market crisis in 2008. Some of website allows readers to calculate the costs of renting or buying a home, they aim to answer which is better? Here is an example of rent/house cost calculator: http://www.nytimes.com/interactive/business/buy-rent-calculator.php

Figure 1: Screen shot of The York Times website

[1] Irwin, Neil. The New York Times. "The Housing Market Is Finally Starting to Look Healthy." Aug. 23, 2016. http://www.nytimes.com/2016/08/24/upshot/the-housing-market-is-finally-starting-to-look-healthy.html

In addition, many visualizations use Zillow data to compare rent and house prices between big cities, but not all states are represented.

http://www.economist.com/blogs/graphicdetail/2016/08/daily-chart-20 is an example:

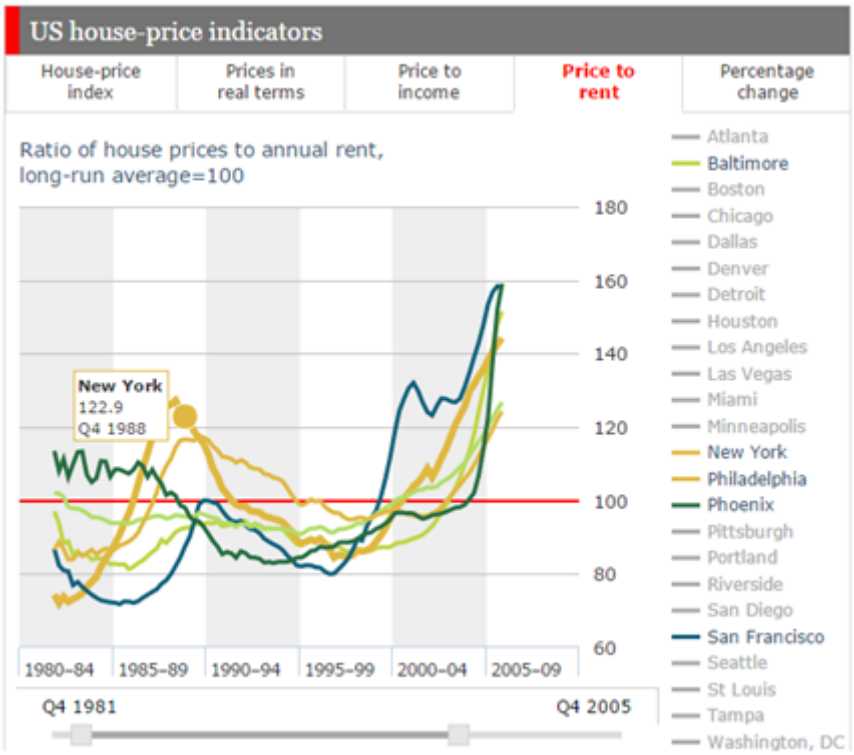


Figure 2: Screen shot of www.economist.com

Besides, we've realized that most of websites about the housing market is very heavy in text and lack of visualizations.

3. Questions

This project aims to answer the following questions. Have we recovered from the 2008 housing market crash? Are housing overpriced now? When does it make sense to buy? When does it make sense to rent? What areas in the United States are the best rental markets? What areas in the United States are the best real estate markets with the highest appreciation forecast?

The goal of this project is to see how the housing market has been affected by the 2008 housing market crash and the current state of recovery. We would like to be able to display areas, which are the best rental markets in the United States. We would also like to display which areas are the most affordable home prices.

A benefit of this project is to educate others on where the best rental markets are in comparison to other states. Another benefit will be to educate the user on the differences of the rental and home prices in different states by comparing the data of the states.

4. Data

a. Source Links

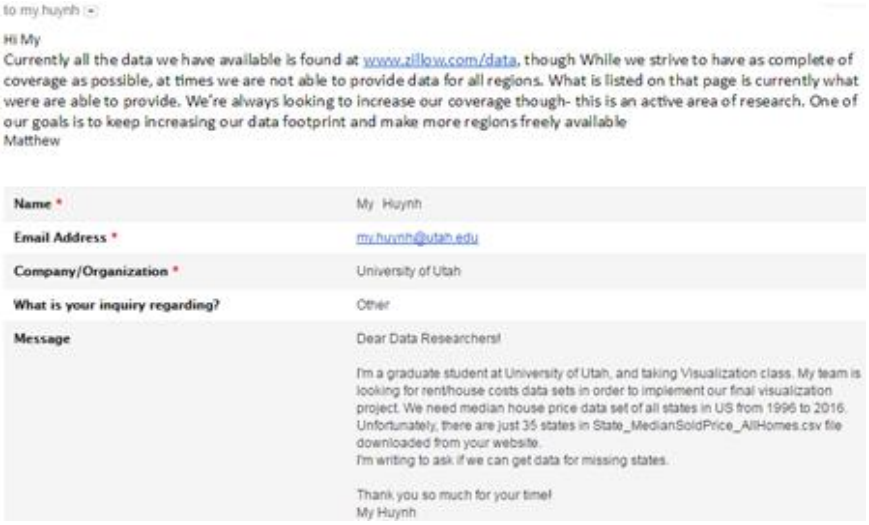
Our data sets are derived from below sources:

- 1996-2015 populations estimates - https://www.census.gov/
- 2016 population estimate - http://worldpopulationreview.com/states/
- ZHVI (Zillow Home Value Index) - http://www.zillow.com/research/data/#median-home-value
- ZRI (Zillow Rent Index) http://www.zillow.com/research/data/#rental-data
- Map boundaries data - http://bl.ocks.org/mbostock/raw/2206489/7110de3d8412433d3222c9b7e3ac6593593162b2/us-states.json

b. Processing

We planned using sold house price dataset in 20 years from 1996 to 2016. Unfortunately, Zillow doesn't provide enough data for all regions and there are 15 states missing data. Therefore, we consulted Zillow data research team about this issue.

They're very helpful but we are not able to get data for all regions in US.



to my huynh

Hi My
Currently all the data we have available is found at www.zillow.com/data, though While we strive to have as complete of coverage as possible, at times we are not able to provide data for all regions. What is listed on that page is currently what were are able to provide. We're always looking to increase our coverage though- this is an active area of research. One of our goals is to keep increasing our data footprint and make more regions freely available
Matthew

Name *	My Huynh
Email Address *	my.huynh@utah.edu
Company/Organization *	University of Utah
What is your inquiry regarding?	Other
Message	Dear Data Researchers! I'm a graduate student at University of Utah, and taking Visualization class. My team is looking for rent/house costs data sets in order to implement our final visualization project. We need median house price data set of all states in US from 1996 to 2016. Unfortunately, there are just 35 states in State_MedianSoldPrice_AllHomes.csv file downloaded from your website. I'm writing to ask if we can get data for missing states. Thank you so much for your time! My Huynh

Figure 3: Screen shot of Zillow data research team's email

However, Zillow has come up with their own methodology for real estate data - Zillow Home Value Index (ZHVI). Each ZHVI is a time series tracking the monthly median home value in a particular geographical region. The ZHVI are available for seven geographical levels: neighborhood, ZIP code, city, congressional district, metropolitan area, county, state and the nation. Similar to ZHVI, Zillow Rent Index (ZRI) tracks the monthly median rent in particular

regions[2]. We've used their ZHVI, ZRI of states, and three other metrics of population, median rental price/house cost per square foot.

In addition, missing states data is a main cause raised a dilemma faced us. In our proposal, we wanted to display how the housing market looks like since the 2008 financial crisis. We planned to have charts to show the home prices and rental rates for years from 1996 to 2016 for US States. Since that is the dates made available to us in some of the home prices data sets we found. However, since we submitted the proposal we haven't been able to find rental data prior to 2010.

There were two options came to our team. In the first option, we had to miss our project goal and only showing house price changes since 2010. On the other hand, we would show home prices from 1996 to 2016 and rental cost from 2010 to 2016 in second option. We would like to say "THANK YOU" to our TA Vinitha Yaski. She has suggested to us implement our project goes with the second option, which is to show home prices from 1996 to 2016 and rental cost from 2010 to 2016. At least we will be able to show part of the original story we were trying to tell with our visualizations.

Zillow offers huge data sets. In order to obtain data sets, which are suitable for what we'd like to implement in our proposal, we pick house price data in 1-year intervals from 1996 to 2016 and rent cost data in 4-month interval from Dec. 2010 to Aug. 2016.

RegionName	abbr	SizeRank	2010-12	2011-04	2011-08	2011-12	2012-(
California	CA	1	1956	1940	1933	1932	1925
Texas	TX	2	1218	1215	1210	1207	1210
New York	NY	3	1720	1749	1737	1742	1771
Florida	FL	4	1257	1247	1242	1241	1247
Illinois	IL	5	1462	1443	1414	1391	1387
Pennsylvania	PA	6	1201	1195	1186	1169	1176
Ohio	OH	7	982	978	983	983	999
Michigan	MI	8	1100	1079	1071	943	913
Georgia	GA	9	1110	1103	1089	1082	1077

Data sets provided by Zillow and Census actually come in CSV format. Hence, we use R for cleaning, aggregating and reformatting the data appropriated for working with in d3.

Figure 4: Screen shot of house price dataset

[2] Bun, Yengon. Zillow. Zillow Rent Index: Methodology. Mar. 12, 2012. http://www.zillow.com/research/zillow-rent-index-methodology-2393/

RegionName	abbr	SizeRank	1996	1997	1998	1999	2000	2001	2002	2003	2004
California	CA	1	157500	165000	179000	198700	230000	258600	303200	359000	452000
Texas	TX	2		84400	87700	94200	97300	100300	105100	108400	119000
New York	NY	3									
Florida	FL	4	86900	90500	94400	100600	109600	120600	133900	150800	184500
Illinois	IL	5	114000	118400	121800	128800	137500	148500	158100	170400	181600
Pennsylvania	PA	6	83800	86100	88900	90900	95400	101400	109000	119600	130500
Ohio	OH	7	89700	94000	97300	102500	107900	111700	115400	120500	123800
Michigan	MI	8	87700	93100	100000	111600	119700	126500	131900	136400	140500

Figure 5: Screen shot of rent cost dataset

Since Mike Bostock's website provides raw U.S. map boundaries data set, in order to implement US map, we merely need to add states abbreviations to JSON file to look up quickly the data we need for those other files based on the states selected.


```
{"type":"FeatureCollection","features":[
{"type":"Feature","id":"01","properties":{"name":"Alabama", "abbr":"AL"},'
[-85.184951,32.859696],[-85.069935,32.580372],[-84.960397,32.421541],[-85.
[-85.042551,31.539753],[-85.113751,31.27686],[-85.004212,31.003013],[-85.4
[-87.37025,30.427934],[-87.518128,30.280057],[-87.655051,30.247195],[-87.9
[-88.394438,30.367688],[-88.471115,31.895754],[-88.241084,33.796253],[-88.
{"type":"Feature","id":"02","properties":{"name":"Alaska", "abbr":"AK"},"g
[-131.355558,55.183705],[-131.38842,55.01392],[-131.645836,55.035827],[-13
[-131.832052,55.189182],[-131.832052,55.42469]]],[[[-132.976733,56.437924]
```


Figure 6: Screen shot of map boundaries dataset

5. Exploratory Data Analysis

We have many reasons to use Tableau in order to explore our data sets. Tableau is one of easy ways people represent data fast, beautiful and useful.

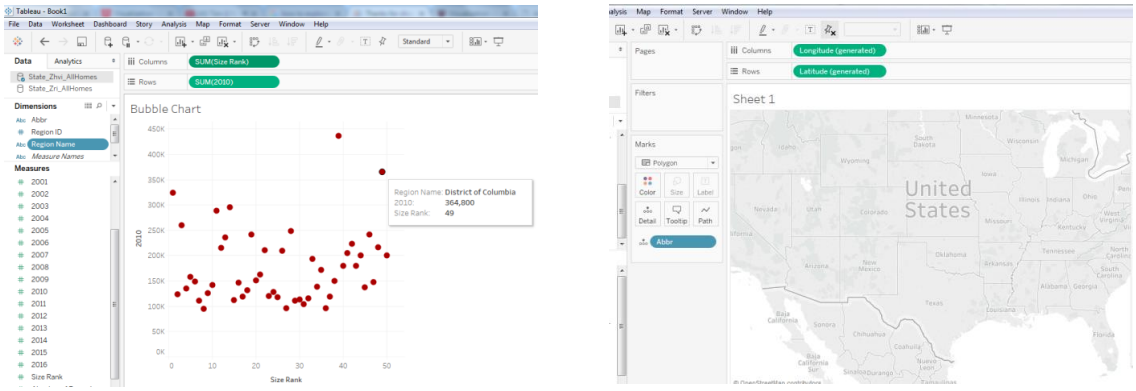


Figure 7: Initial bubble chart and map using Tableau

After getting the initial map, we tried to color our map but Tableau showed a message notified there is not any related attribute between map and rent value datasets. The important thing we gain here is that we have to add states abbreviations in datasets. We've learnt that exploratory data analysis is significant step that helps to detect mistakes, figure out necessary attributes, and discover relationships between features in order to prepare for a successful visualization.

6. Design Evolution

a. Final design in proposal

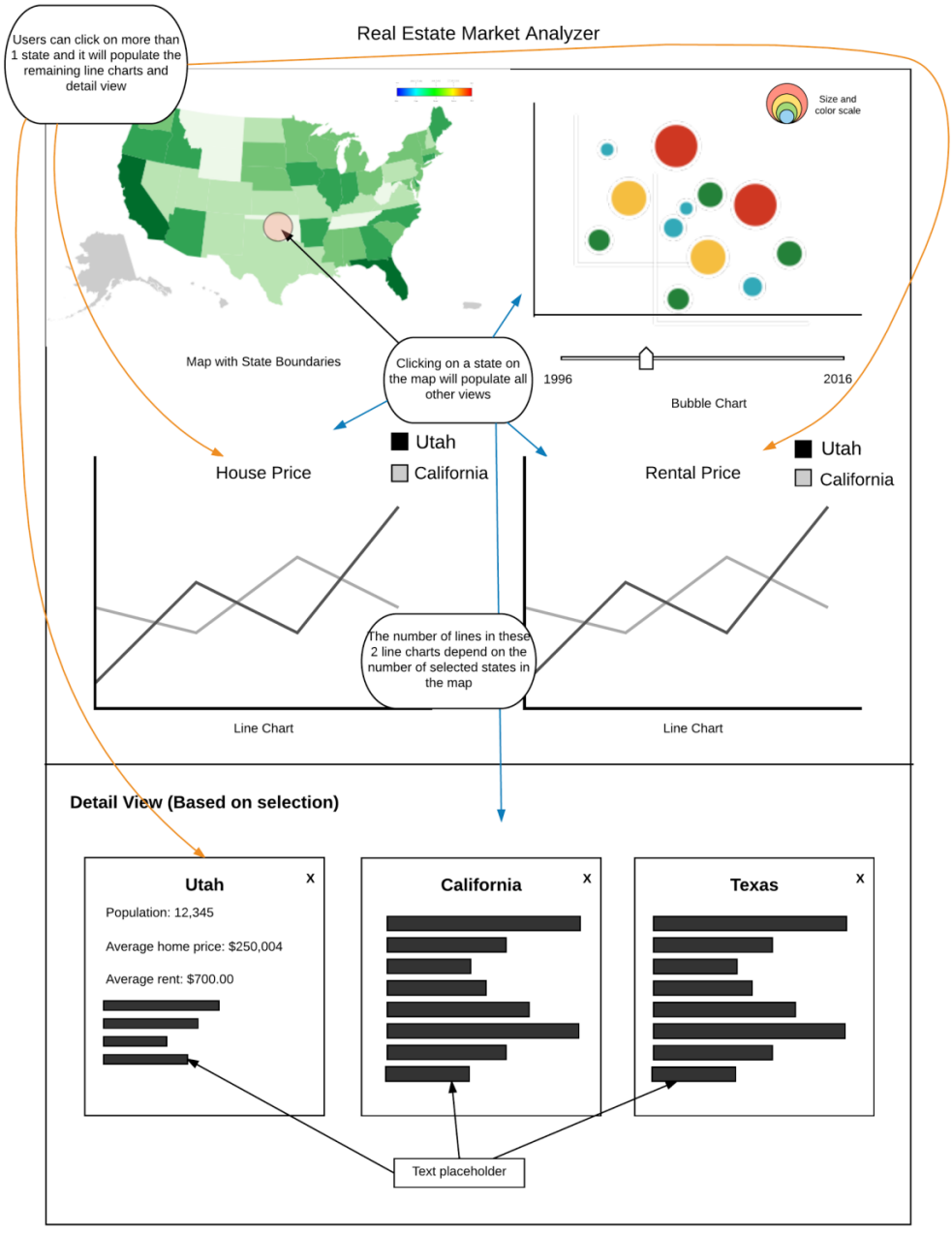


Figure 8: Our final sketch in proposal

b. Description

Slider View: This will be a time slider, as in design 2, for the years 1996 to 2016, in one year increments. This will be interactive by the user to update the rest of the views.

Map View: The map section, as in design 1, will display the entire United States of America with the states outlines and names shown. The map will be interactive where the user is able to click on and select one or more states. Selected states data will be shown in the line charts view and detail view. The map will also have a color scale which will represent the rent statistics for the year selected by the slider view.

Bubble Chart View: This will be as in design 2. Chart view has a function that shows how fast housing price increases over time. As the map, bubble chart will be interactive with time slider in order to show average house price of all states have changed in 20 years. One circle represents for one state. Size and color of each circle shows US state size rank by population. The more state's population is, the more size of circle is.

Line Charts View: This will represent historical rent and home price data for each of the selected states in either the map view or bubble chart view.

Detail View: This will represent more specific details of each of the selected states. Some details it may include, but not limited to, are population, average home price, average rent rate and so forth.

Must-Have Features:

- A map that displays all states with boundaries
- Rent throughout the years
- Median home price throughout the years
- Bubble chart shows appropriate data
- Bubble chart and map interact with time slider
- Detailed view of the selected states

Optional Features:

- Include visualizations for crime rate per capita of locations
- Include visualizations for school ratings of locations
- Include data on the cost of living of locations
- Include interest rates since 2011
- Add pan and zoom to the map

c. Design changes

We didn't deviate from our proposal. We've decided to not change anything on the final design above but adding more optional features if we have time. Since we've use two different time scales for map and bubble chart, we will add one more slider into our visualization.

7. Implementation

a. Project structure

We organize our working prototype follow HW5 layout.

index.html
data/
- **State_MedianRentalPricePerSqft_AllHomes.csv**
- **State_MedianValuePerSqft_AllHomes.csv**
- **State_Zhvi_AllHomes.csv**
- **State_Zri_AllHomes.csv**
- **US_Population_AllStates.csv**
- **Us_states.json**

public/
- **css/**
  - **style.css**
- **js/**
  - **bubbleChart.js** *visualizes bubble chart*
  - **detailCards.js** *displays detailed view*
  - **houseChart.js** *shows house-price line chart*
  - **main.js** *handles script files*
  - **mapView.js** *displays U.S map*
  - **rentChart.js** *shows rent-cost line chart*

b. Implementation

❖ Map

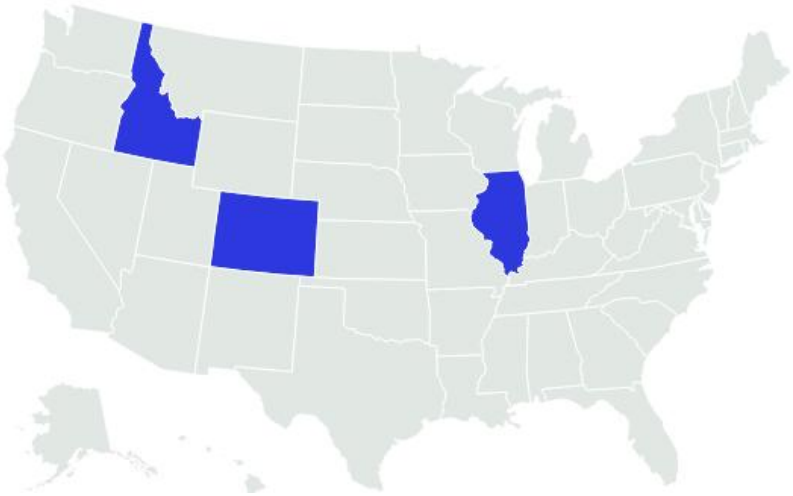


Figure 9: Initial U.S map

In this milestone, we've implemented initial U.S map allows users to select multiple states by changing the color when you click on them. The map also changes color on hover over.

Things to do:

- Add slider, set color scales, add color legend
- Implement map interacts with slider.
- Implement click event shows appropriate line charts, information views.

❖ Bubble chart

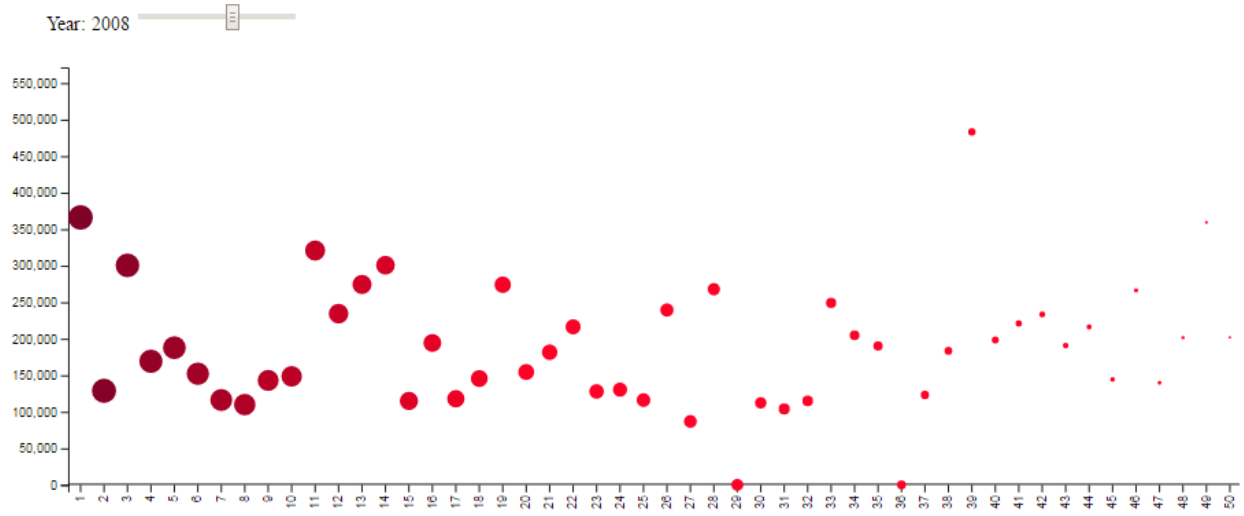


Figure 10: Initial bubble chart

So far, we have initial bubble chart and slider with time scale from 1996 to 2016 in 1-year increment. In the chart, X-axis represents state size rank by population while Y-axis represents housing cost. We were worry that bubbles would be overlapped so that we set it interactive with slider in order to see what happens and think of a better way.

Things to do:

- Add color, circle size legend.
- Radius, 'cx' attribute and color of circles depend on size rank value. We are thinking of how to set the radius and color base on state population.
- Add tooltip show state information

❖ Line charts

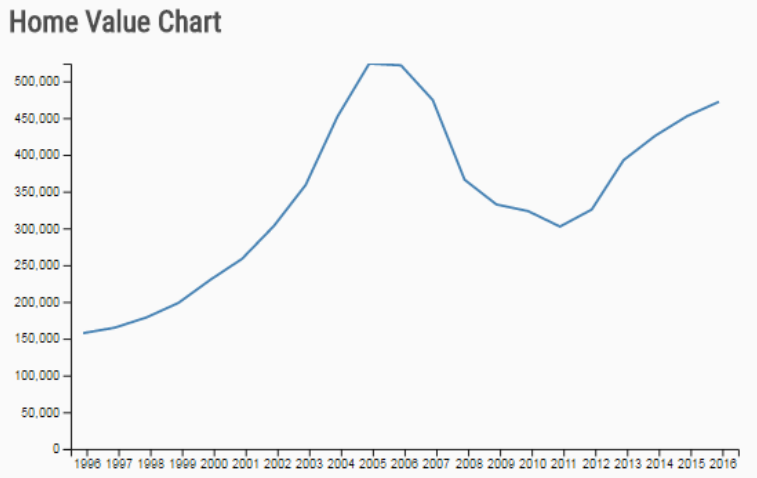


Figure 11: Initial home-value line chart

We haven't implemented interaction between map and line charts but they look great so far. Figure 11 is an example how home-value chart looks like when user selects a state.

Things to do

- Add tooltips show highest and lowest rent/house price in each for selected states
- Line charts should response on states click events

❖ Detail View

Since we is not required to have all views up and running in milestone, we decide to add images for placeholders so that users can figure out what we will show in detail views.

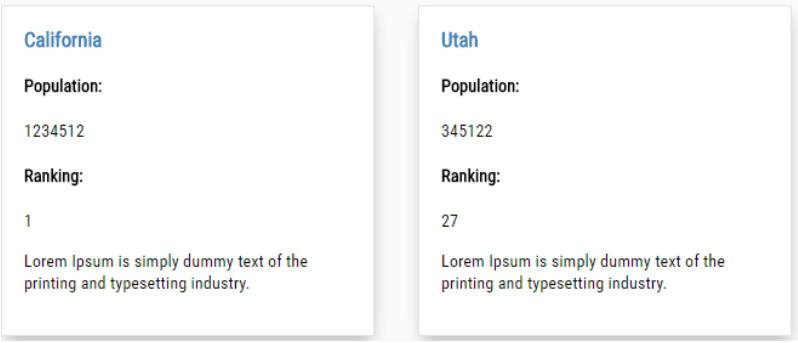


Figure 12: Example of detail views

Things to do:

- Set views response on states click events
- Views should display appropriate states information.

8. Evaluation

Our datasets are ready, visualizations and working prototype are organized and we feel that we're on the right way to get our project done. We think our visualization will be much better when we add more interactions and option features. We are also thinking of adding interaction between map view and bubble chart.

Moreover, teamwork is very important and vital to success the final project. Our teamwork is great and we work well together.