

Urban Growth Visualization

Process Book

By Tony Zhang (u1183156@utah.edu)

and Cody Matthews (u0712564@utah.edu)

FINAL PROJECT, DS-4630 (FALL 2024)

Table of Contents

Table of Contents.....	1
Initial Project Proposal:.....	2
Background and Motivation.....	2
Project Objectives.....	2
Data.....	2
Data Processing.....	3
Visualization Design.....	3
Must-Have Features.....	4
Optional Features.....	4
Project Schedule.....	4
Feedback on Initial Meeting with TA Rifat Ara Proma.....	11
Data Process.....	12
1st Data Retrieval/Cleaning.....	13
2nd Data Retrieval.....	14
Sources for initial data:.....	14
2nd Data Cleaning.....	15
Design Evolution/Implementation.....	16
1st Code Entry.....	17
2nd Code Entry.....	18
3rd Code Entry.....	19
4th Code Entry.....	20
5th Code Entry.....	21
6th Code Entry.....	22
7th Code Entry.....	23
8th Code Entry.....	24
Generalizing Code To Other Datasets.....	25
9th Code Entry.....	26
10th Code Entry.....	27
11th Code Entry.....	28
12th Code Entry.....	29
13th Code Entry.....	30
1st Comparison Code Entry.....	31
2nd Comparison Code Entry.....	32
3rd Comparison Code Entry.....	33
4th Comparison Code Entry.....	34
Finishing Touches.....	35
Evaluation.....	37
Data Visualization Inferences:.....	37
Visualization Assessment.....	38

[Click for Table of Contents](#)

Initial Project Proposal:

Background and Motivation

We chose this topic for a number of reasons:

1. Real estate investing:
 - i. We are curious to see which places have grown in population/housing price the most/least over time and to then be able to piece together reasons for why this is and what to expect moving forward.
2. Home purchases and budgeting:
 - i. The decision to either buy a home or rent is always a topic worthy of discussion. Seeing what housing prices are for various places over time can then be used to compare rental rates to be able to decide which path is best for each person.
3. Population growth and job opportunities:
 - i. With urban growth comes opportunities for companies to relocate to fit the demand in each region or vice-a-versa. This allows for decision making of where to move that has a job that best fits your skill set. We also would like to see GDP by state to see if there are any key factors for movements.

Project Objectives

What would we like to learn and accomplish?

1. How Urban Growth has affected housing prices over time in different locations.
2. To understand which regions have experienced the fastest growth and which regions have seen the greatest increase in housing costs.
3. To better understand how market rates change based on population size.

Data

We plan on gathering the following data:

1. Urban growth data:
 - i. This will come from population and urban density from the US census bureau, and other open-source data platforms. We also plan on gathering information related to the GDP of each state to better understand population shifts and the effects from them.
 1. <https://www.census.gov/>

2. <https://www.bea.gov/data/gdp/gdp-state>
2. Housing price data:
 - i. This will come from historical selling prices and current market rates through places such as Zillow, realtor.com, and other government related tax assessment records.
 1. <https://www.zillow.com/>
 2. <https://www.realtor.com/>
 3. <https://www.bls.gov/eag/home.html>
 4. <https://www.redfin.com/news/data-center/>
3. Employment Data:
 - a.

Data Processing

We expect there to be data cleanup in order to correctly match up data that is gathered from different sources.

We plan to derive urban growth quantities and housing prices for the timeline for the regions we choose to highlight.

How will data processing be implemented?

Data will be scraped/read in using a python colab notebook. Inside the notebook is where all cleaning and formatting of the data will be done.

We will also be gathering data from the websites mentioned above as the majority of them have .csv files available for download that can easily be formatted to match each other.

Visualization Design

Develop three alternative prototype designs for your visualization

1. Have a main page with the US states and then sub plots off to the side that can be manipulated based on filters.
2. Have a main page with the US states that can be selected and then have plots off to the side reflect data from the highlighted states.
3. Have a heat map of the US states that reflects growth rate, population, etc that can be toggled.

And one final design that incorporates the best of your three designs

- A geographical map where different regions are color coded based on housing prices/population density. With multiple plots on the side.

Describe your designs and justify your choices of visual encodings.

Bubble chart for each region that represents the population size and its x/y orientation represents housing prices vs growth rate of population.

A heatmap of each region where darker colors represent higher prices/growth/density with an option to toggle between layers. Include a legend.

An urban growth map that toggles before/after views of housing prices/population density

Line graph (one for population growth and another for housing price trends)

Must-Have Features

A map-based visual that shows each location's population and housing costs.

Line or bar graph that highlights each region's population and housing costs over time.

Optional Features

A map-based visual that shows a prediction of what each region will look like in 5 years, 10 years.

A comparative analysis that allows the user to compare two regions side by side.

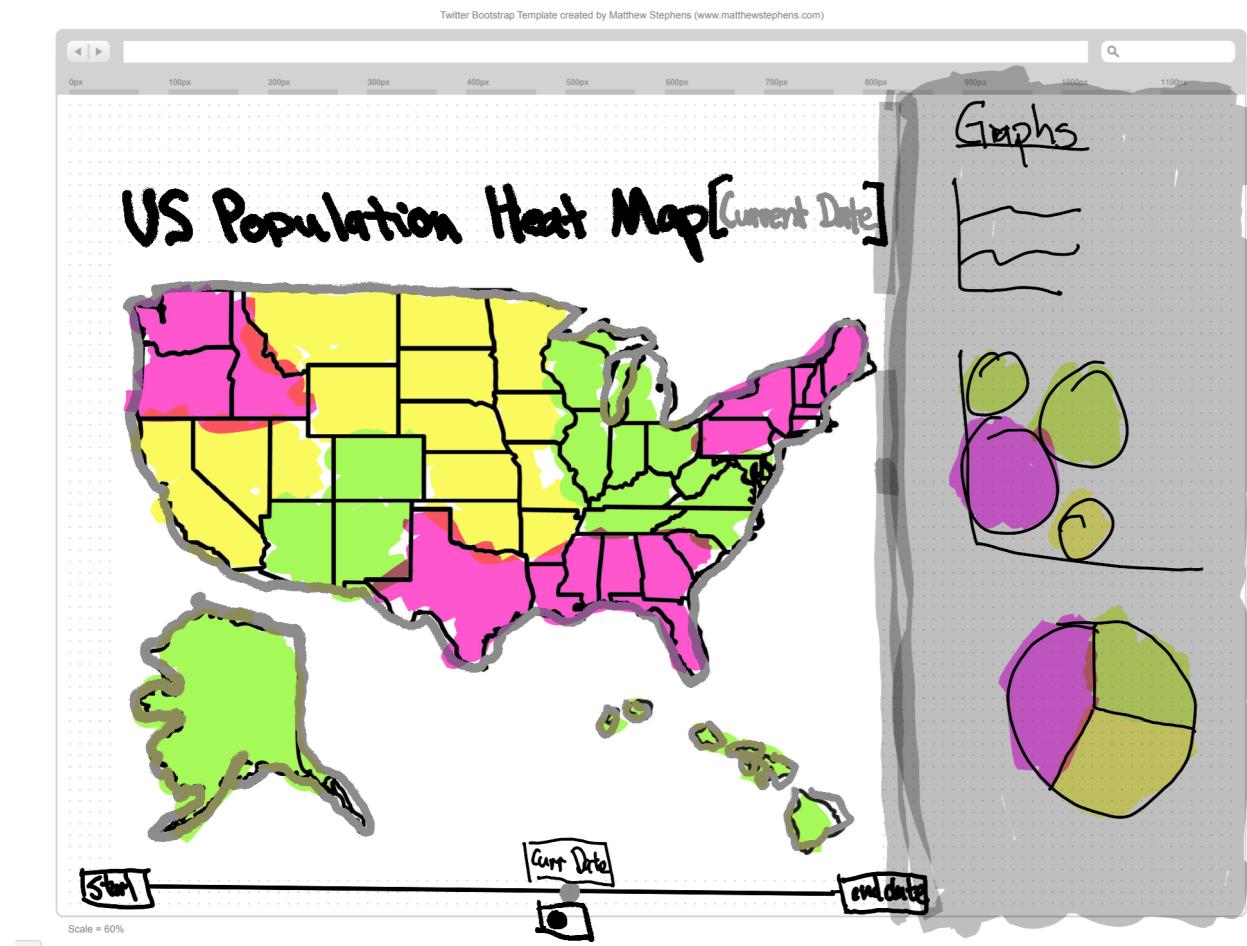
When small regions are selected the counties/cities within the state(s) will be colored individually.

Project Schedule

- Week 3 (Sep 2-6)
 - Rough draft of project proposal
 - Create group GitHub repository
- Week 4 (Sep 9-13) Project Proposal Due
 - Complete and turn in Project Proposal
 - Schedule a time to meet with staff member
- Week 5 (Sep 16-20) Project Review w/ Staff (as a group!)

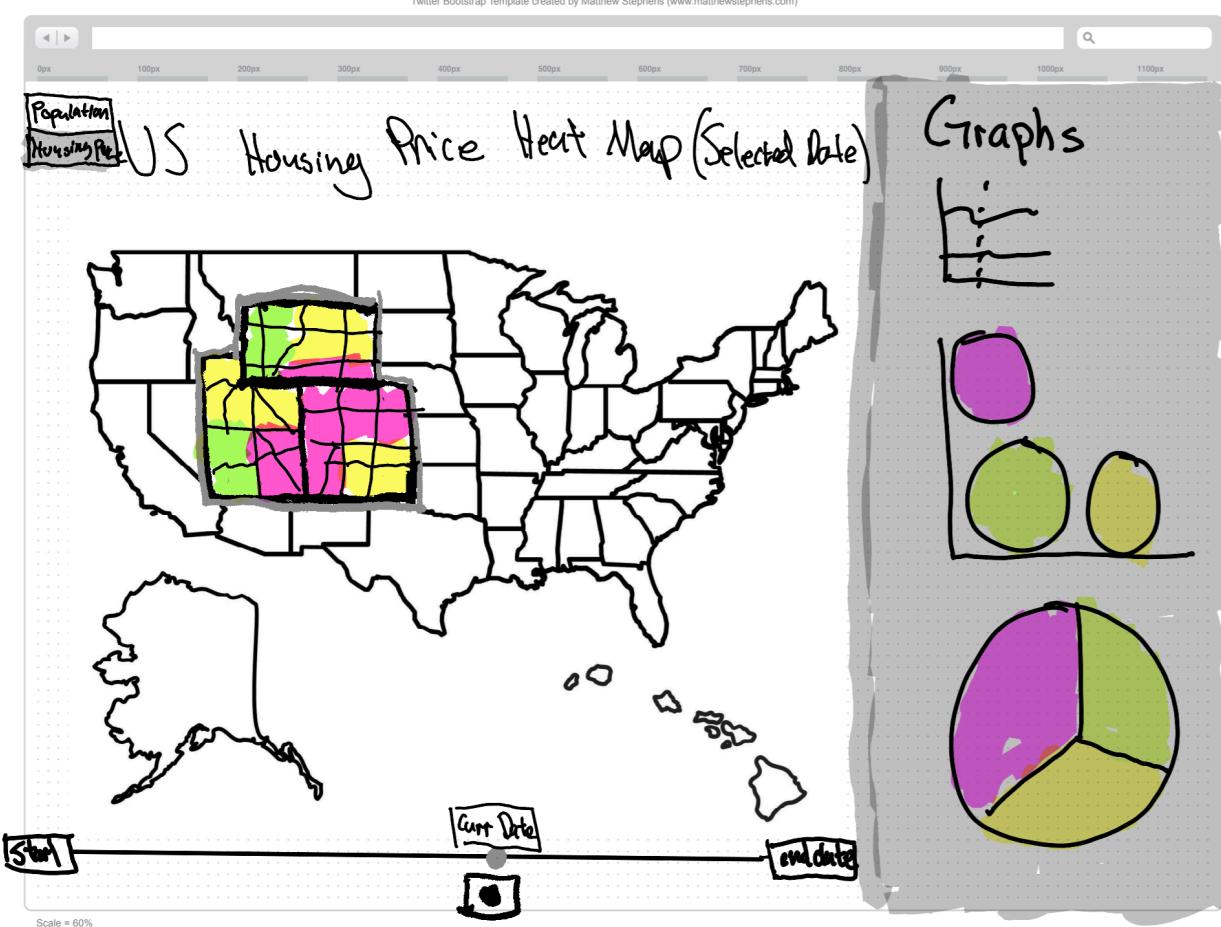
- Meet with TA for project review
- Week 6 (Sep 23-27) Project Review w/ Staff (as a group!)
 - Gather data needed to implement project
- Week 7 (Sep 30-Oct 4) Midterm Exam
- Week 8 (Oct 7-11) Fall Break
- Week 9 (Oct 14-18)
 - Finish compiling data for project milestone
- Week 10 (Oct 21-25) Project Milestone Due
 - Submit project milestone for review
 - Start coding website
- Week 11 (Oct 28-Nov 1)
 - Code website
- Week 12 (Nov 4-8)
 - Code website/fix bugs
- Week 13 (Nov 11-15)
 - Finish coding website/fix bugs
- Week 14 (Nov 18-22) Project Screencast Submission Due
 - Record and submit Project Screencast
- Week 15 (Nov 25-29)
- Week 16 (Dec 2-6) Final Project Submission / Group Member Feedback Due
- Week 17 (Dec 9-13) Final Exam

Opening View



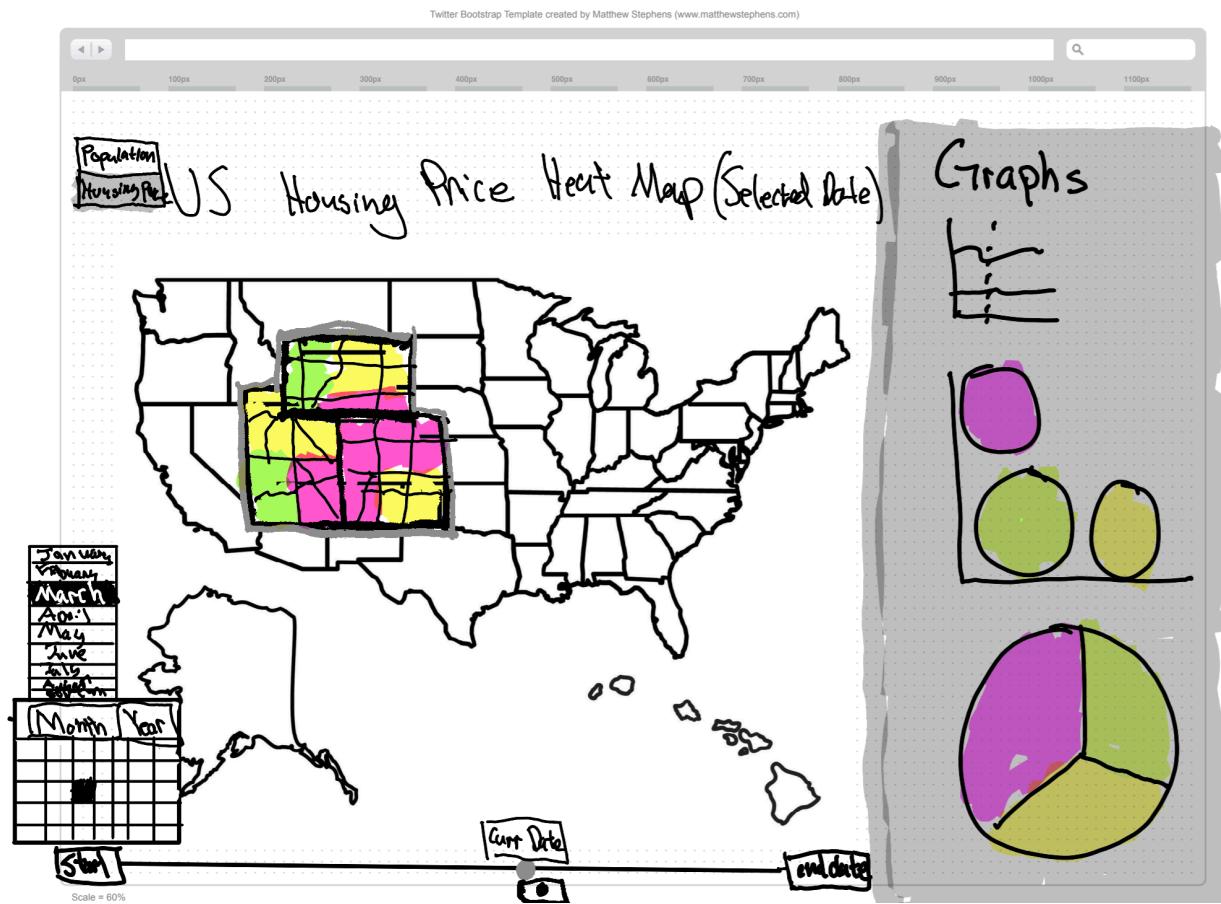
[Click for Table of Contents](#)

Region Selection



[Click for Table of Contents](#)

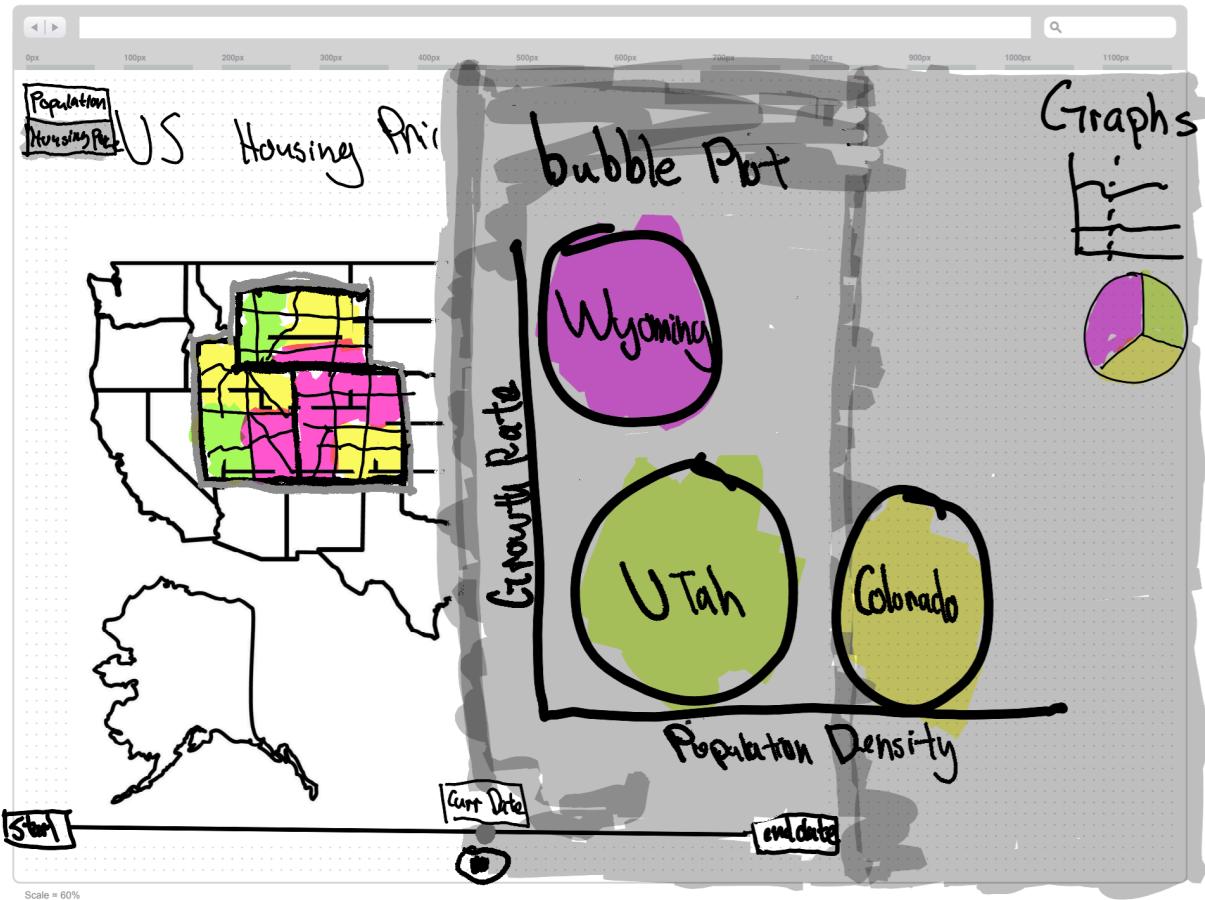
Choose starting dates for animation



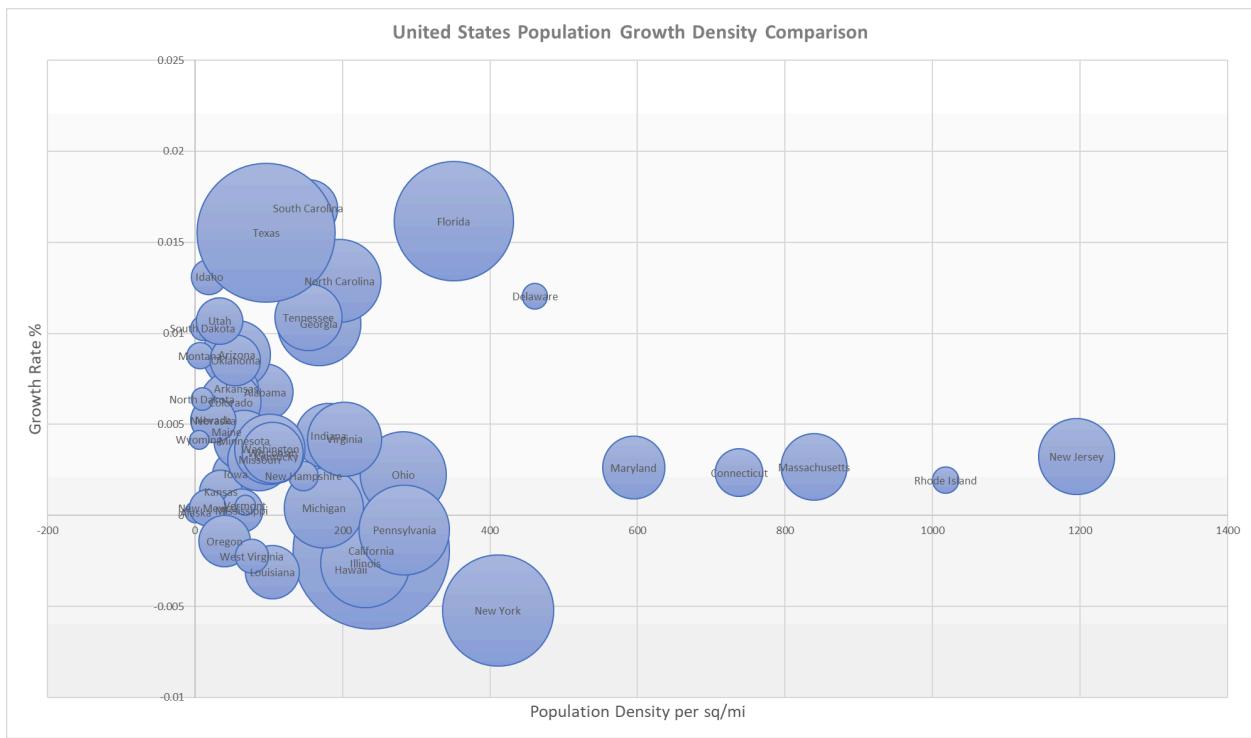
[Click for Table of Contents](#)

Enlarging a graph

Twitter Bootstrap Template created by Matthew Stephens (www.matthewstephens.com)



Bubble plot example of the states with population vs growth rate vs density



Feedback on Initial Meeting with TA Rifat Ara Proma

[Entry by Tony Zhang and Cody Mathews, 9-19]

During our meeting with Rifat, her main critique was that our proposal visualization sketches weren't very clear and didn't seem to directly accomplish our objectives in our proposal. She suggested that we look for other visualizations to get inspiration and come up with some clearer relationships that we wanted to investigate/demonstrate within our visualization.

Data Process

This section details the process of how the data was collected and processed for visualization.

1st Data Retrieval/Cleaning

[Entry by Tony Zhang, 10-19]

After our very brief overview of the potential sources for data for the proposal it was time to start looking into the sources we mentioned in the proposal more carefully. Something that was quickly noticed was that the initial data seemed to cover different periods of time as the initial median income dataset went from 1984-2018 and the population dataset went from 2019-2023. It also did not have county level data. However, I decided to clean it so my teammate could have some data to start working with for the visualization.

2nd Data Retrieval

[Entry by Tony Zhang, 10-22]

I spent a decent amount of time looking for new data sources so that all the datasets can overlap allowing for visualization that will allow for comparisons between the datasets. It seems that the data on housing prices will be the limiting dataset in terms of the years that can be analyzed as it only goes from 2012-2024. Unemployment data did not cover this time period and could not find more data on this so found a source for job growth which should still get the desired effect that this visualization is hoping for. The median and population datasets are the datasets that I cannot find monthly data for. Hopefully we will be eventually able to find the necessary datasets to find the monthly data for allowing for a more in depth visualization but we should be able to implement all of the desired features with this data. We are probably going to give up on visualizing the data on a county level as I could not find any trace of data that we could use for it.

Sources for initial data:

- **Median Income Data Sources:**
 - 2019-2023: <https://fred.stlouisfed.org/release/tables?eid=259515&rid=249>
 - 1984-2018: <https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-income-households/h08.xls>
 - 2024: <https://worldpopulationreview.com/state-rankings/median-household-income-by-state>
 -
- **Housing Data Sources:**
 - 2012-2024: <https://www.redfin.com/news/data-center/>
- **Population Data Sources:**
 - 2010-2019: <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html>
 - 2020-2023: <https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html>
 - 2024: <https://worldpopulationreview.com/states>
- **Employment Data Sources:**
 - **Unemployment:** <https://www.kaggle.com/datasets/justin2028/unemployment-in-america-per-us-state?resource=download>
 - **Job Growth:** <https://seidmaninstitute.com/job-growth/state/>

[Click for Table of Contents](#)

2nd Data Cleaning

[Entry by Tony Zhang, 10-23]

For this task, I intended to use the techniques used in the data wrangling course I took (DS-2500). After uploading these datasets to a shared google drive for our team, these techniques did prove useful and the data was cleaned and formatted within a Google Colab Document ([see Colab Document here](#)) to be used in the D3 data visualization. The bulk of the cleaning was removing unneeded columns and renaming the needed columns into a shared common one. A majority of the datasets had years listed as columns so they were all melted down into a single year column. Some of the columns had numbers listed with commas and dollar signs so those were removed as well. Finally, the 2020-2024 median income and job growth data sources didn't have data files that I could download so I simply copied the data from the website, pasted it into ChatGPT to format it to be pasted into an excel file, and used those files. Finally, for those datasets that covered different time periods I simply concatenated them together and exported them to CSV files. The Colab document is definitely messy but I am happy with the final cleaned datasets.

Design Evolution/Implementation

This section details the process of how the visualizations were implemented and how they evolved over time.

1st Code Entry

[Entry by Cody Mathews, 10-20]

For this task, I started out by creating the US map. I first searched D3's website for graphics and interactions. I came across this map:

<https://observablehq.com/@d3/zoom-to-bounding-box>

I used the file us.json and chart.svg from this site to draw and render the US map.

I then tied the cleaned data to the us.json. I did this by using Promise.all(). Within this, I linked the paths to the feature selected and tied it to the purple color scheme that adjusted its saturation fill based on the population in relation to other states.

<https://d3js.org/d3-scale-chromatic/sequential#interpolateRdPu>

I used tooltip to create a hover description tied to each state that displays the state name and data associated with it (population) based on the hover feature that was used in assignment 3 for the bar graph..

2nd Code Entry

[Entry by Cody Mathews, 10-20]

I created a drop down menu that was linked to each year (2019-2024). Based on the selected year that the user selected, the states would then adjust saturation in relation to each other and correctly display the new population for that year when hovered.

A click feature was added that highlighted the state red when clicked.

3rd Code Entry

[Entry by Cody Mathews, 10-21]

I created three separate containers to resemble the three graphs that we had shown in our original design thought.

I then implemented the first chart which was a line graph. I used the same theory from assignment 3 to draw and scale this graph. I used red for the line to match the highlighted state.

4th Code Entry

[Entry by Cody Mathews, 10-21]

I implemented the bar graph with the same theory to that of assignment 3. I realized there was a lot of color to the chart, so I changed it to the class standard of steelblue to make it easier on the eyes.

I added a transition to the states for the hover feature. Each state would transition to the highlighted color over 300ms instead of instantly to make it look smoother.

5th Code Entry

[Entry by Cody Mathews, 10-21]

There were a lot of colors still and wanting to make it simpler, I changed the color scheme of the map to grayscale and the highlights to steelblue.

6th Code Entry

[Entry by Cody Mathews, 10-22]

Now that everything is working smoothly and updates with the correct features, I decided to add transitions to the two charts as well. I was able to make the entry work, but the exit was still disappearing instantly without the smooth transition that was done in assignment 3.

7th Code Entry

[Entry by Cody Mathews, 10-24]

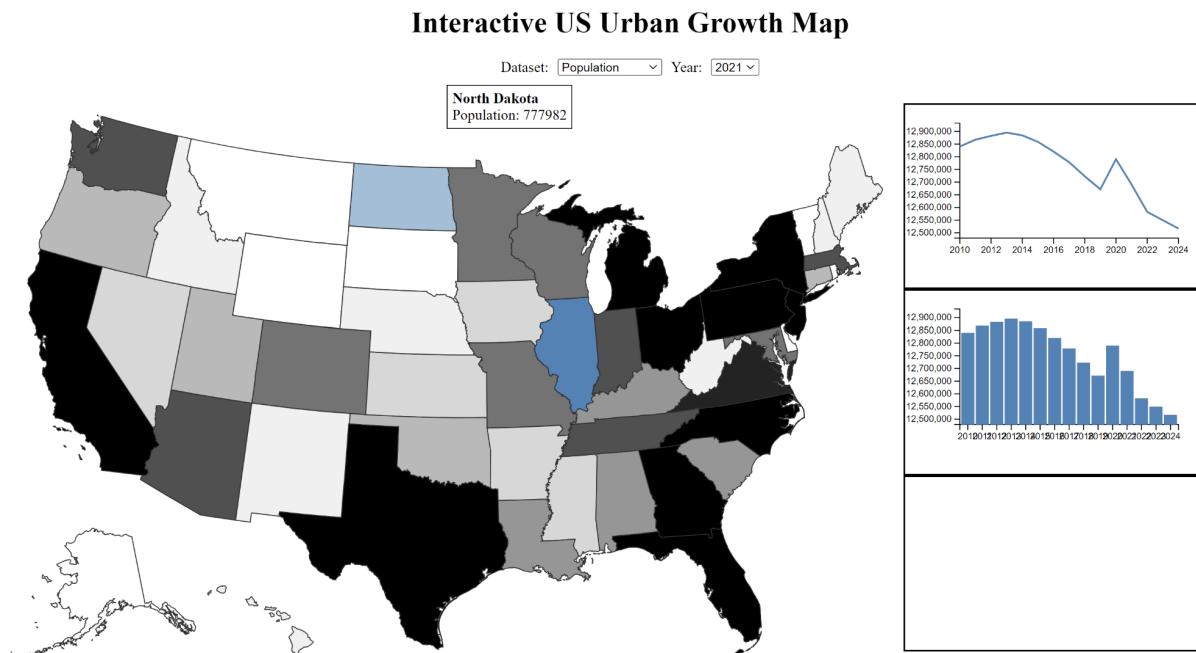
I removed the current data sets and replaced them with updated ones. The reason for this is the current population dataset was missing the year 2022. Also, we felt 5 years of information was relatively small for when we want to do our comparison later on between data sets to find correlations and patterns for growth over time in each state.

I updated both the line graph and bar chart to populate the new data with the extended years.

I was able to implement the exit transition as I found out I was removing the data at the beginning of the method in addition to and before the exit method at the end.

I added another drop down menu to choose from the datasets. I implemented the switch function to swap between datasets. The next task will be to update information in the US map, and graphs to change and reflect the different data.

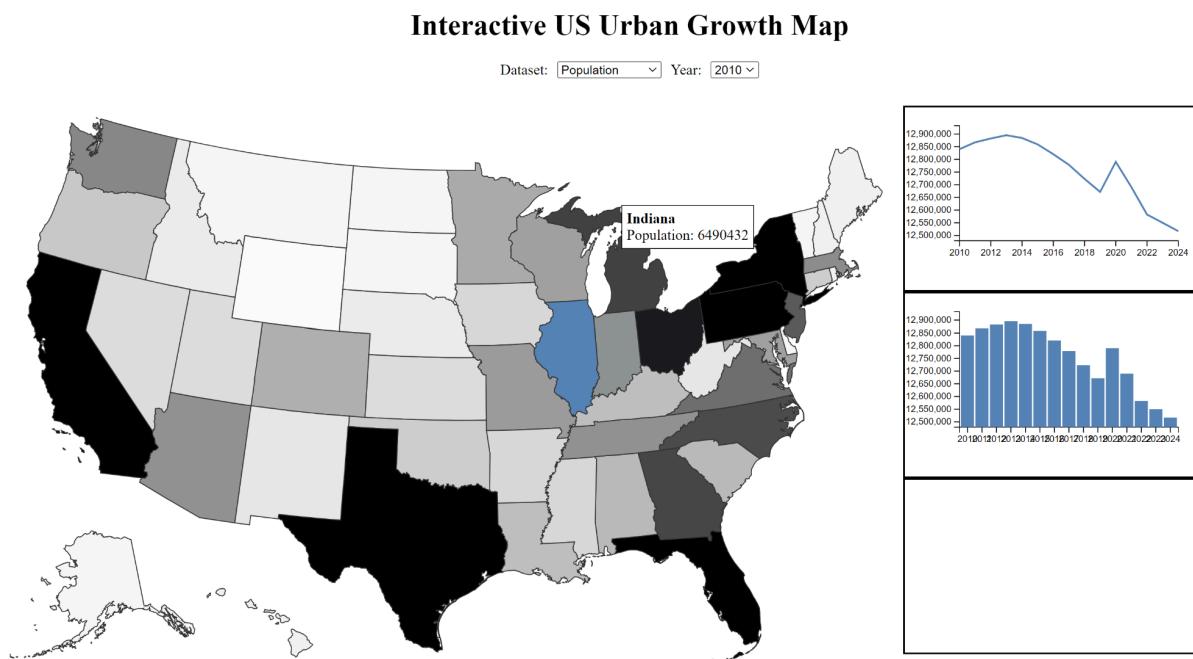
The image below shows the graphs and click feature tied to Illinois. It shows the population density amongst the states. It also depicts the hover feature for North Dakota as it's in transition to turning steel blue. The drop down menus to choose between datasets and years are also shown.



8th Code Entry

[Entry by Cody Mathews, 10-24]

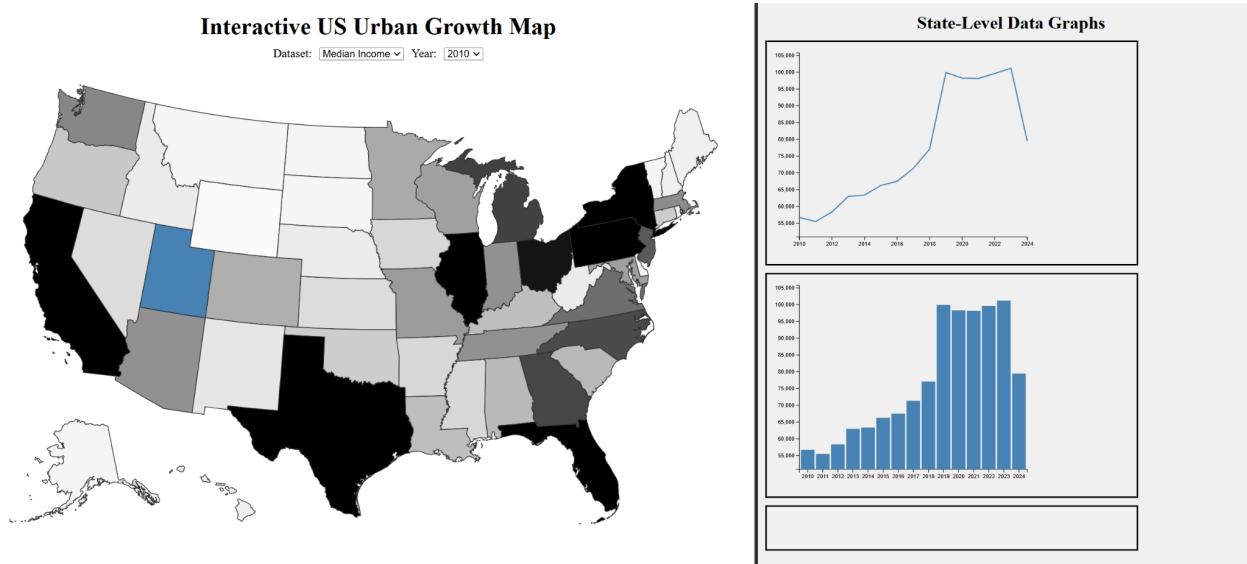
I changed the color scheme from schemeGreys[k] to interpolateGreys(t) and set the max at 2,000,000 for the population at 80%. This change allowed the dataset to go from only 9 gradients to match the scale of the population which allowed for more states such as Georgia, North Carolina and Michigan to distinguish themselves from the states that have more than double their population to give a better depiction of the overall population of each state in respect to each other.



Generalizing Code To Other Datasets

[Entry by Tony Zhang and Cody Mathews, 10-24]

Made changes to the layout of the page so graphs and the map are separated from each other on the page:

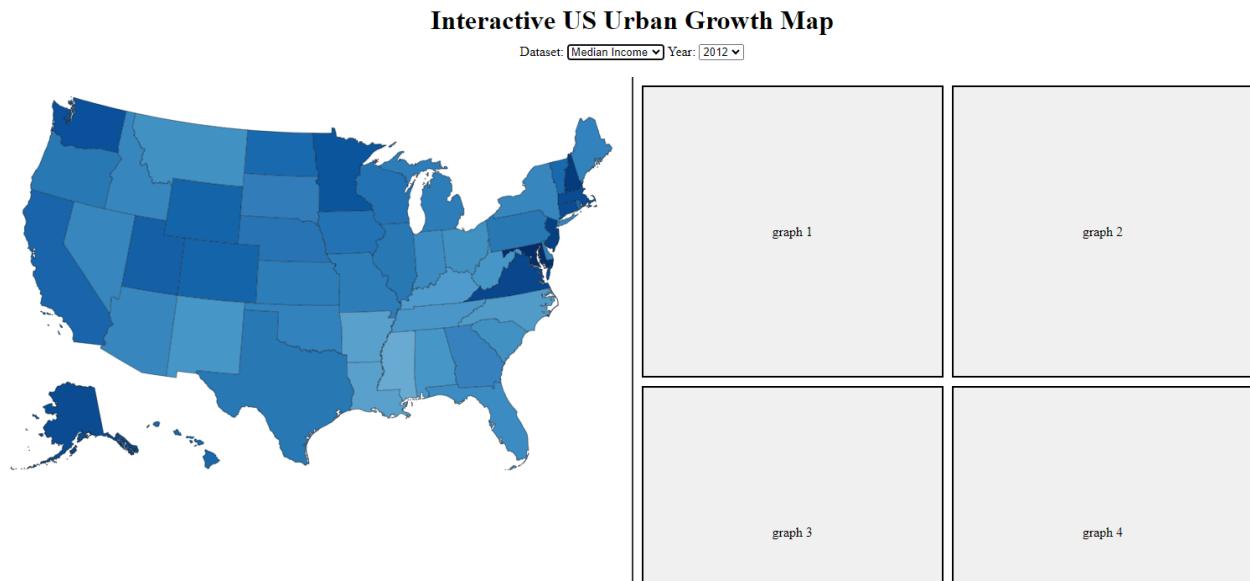


State level graphs now correctly display for other datasets but heat map does not yet currently, tried many things but did not work, will take some more time to figure out.

9th Code Entry

[Entry by Cody Mathews, 11-16]

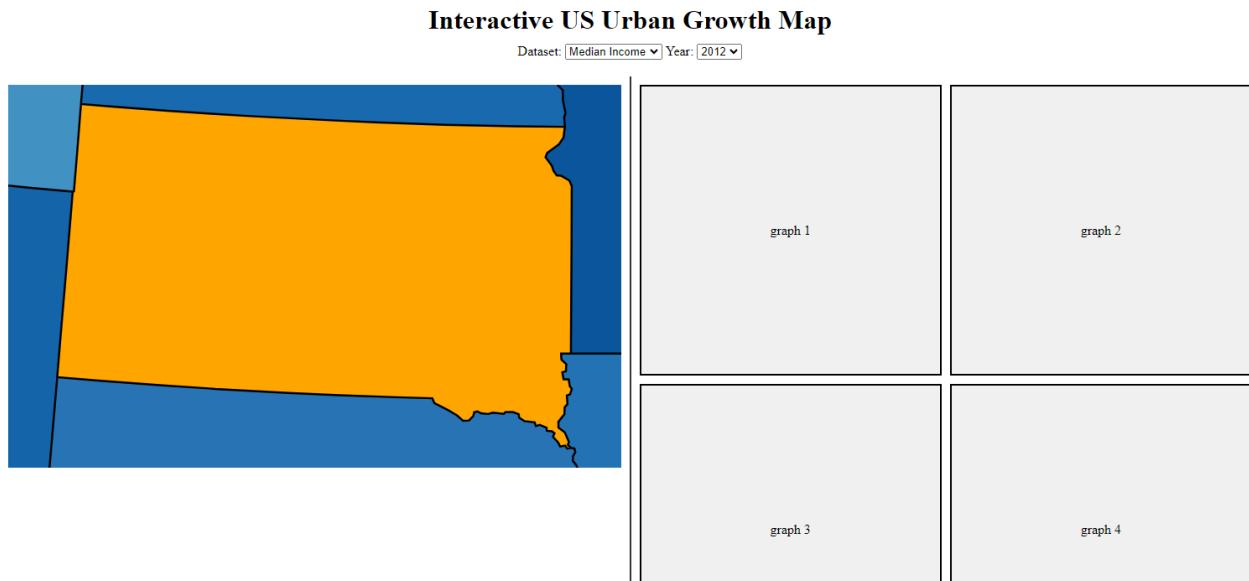
I remade the files to be their own based on their container space to avoid having hundreds of lines of code to scroll through.



10th Code Entry

[Entry by Cody Mathews, 11-17]

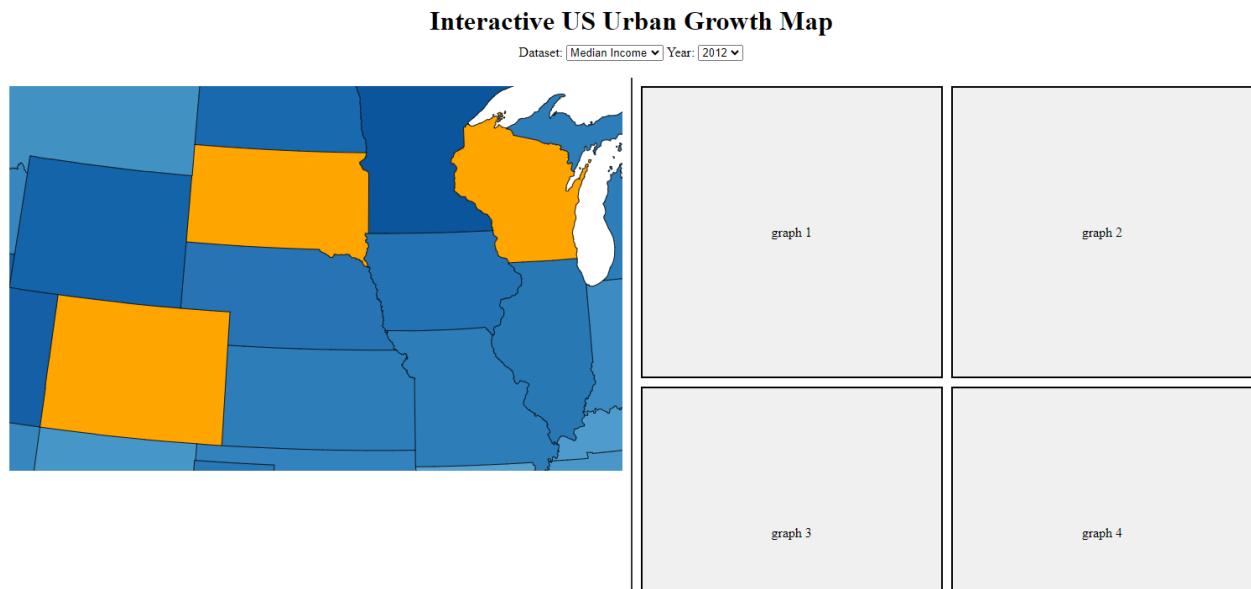
I added a zoom functionality that focuses on the state that is clicked.



11th Code Entry

[Entry by Cody Mathews, 11-17]

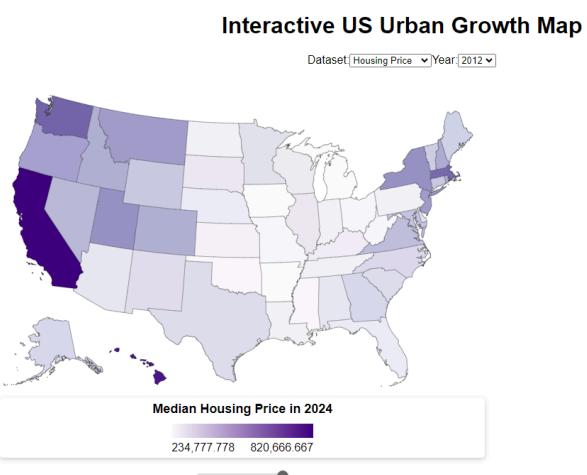
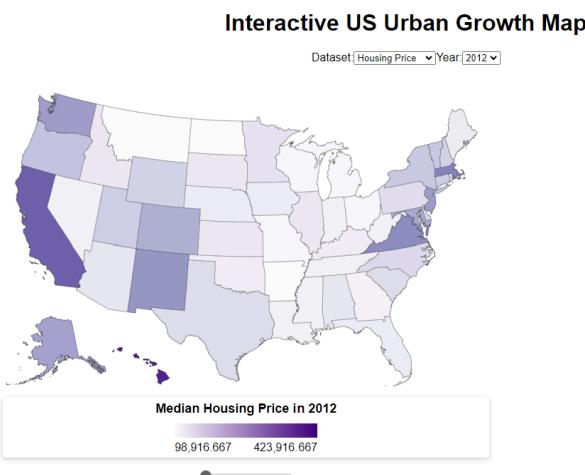
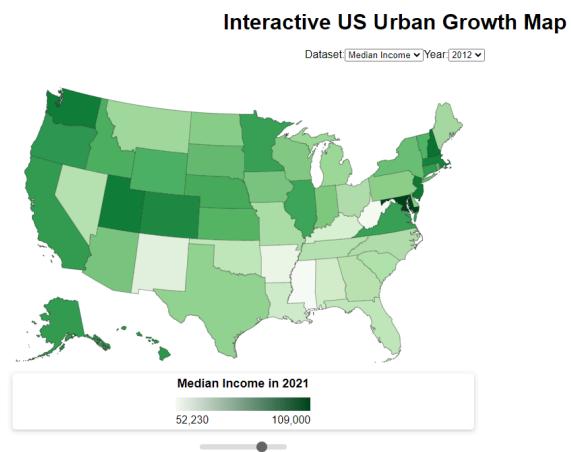
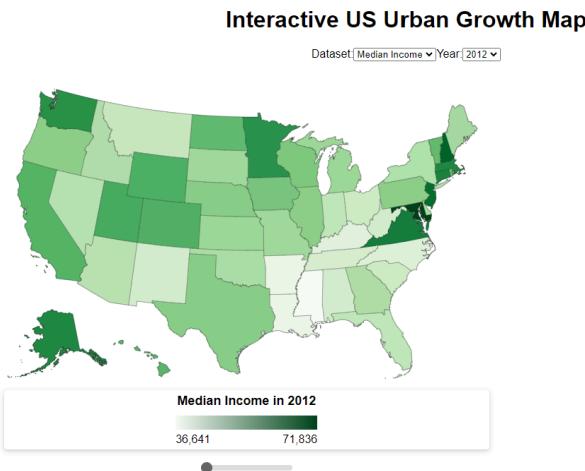
I updated the zoom window and states to click to include multiple states for us to later compare for analysis and assumptions.



12th Code Entry

[Entry by Cody Mathews, 11-21]

Added a legend and scroll bar that updates the map as it slides. I still can't get the globalApplicationState to send data over to the graph(s).

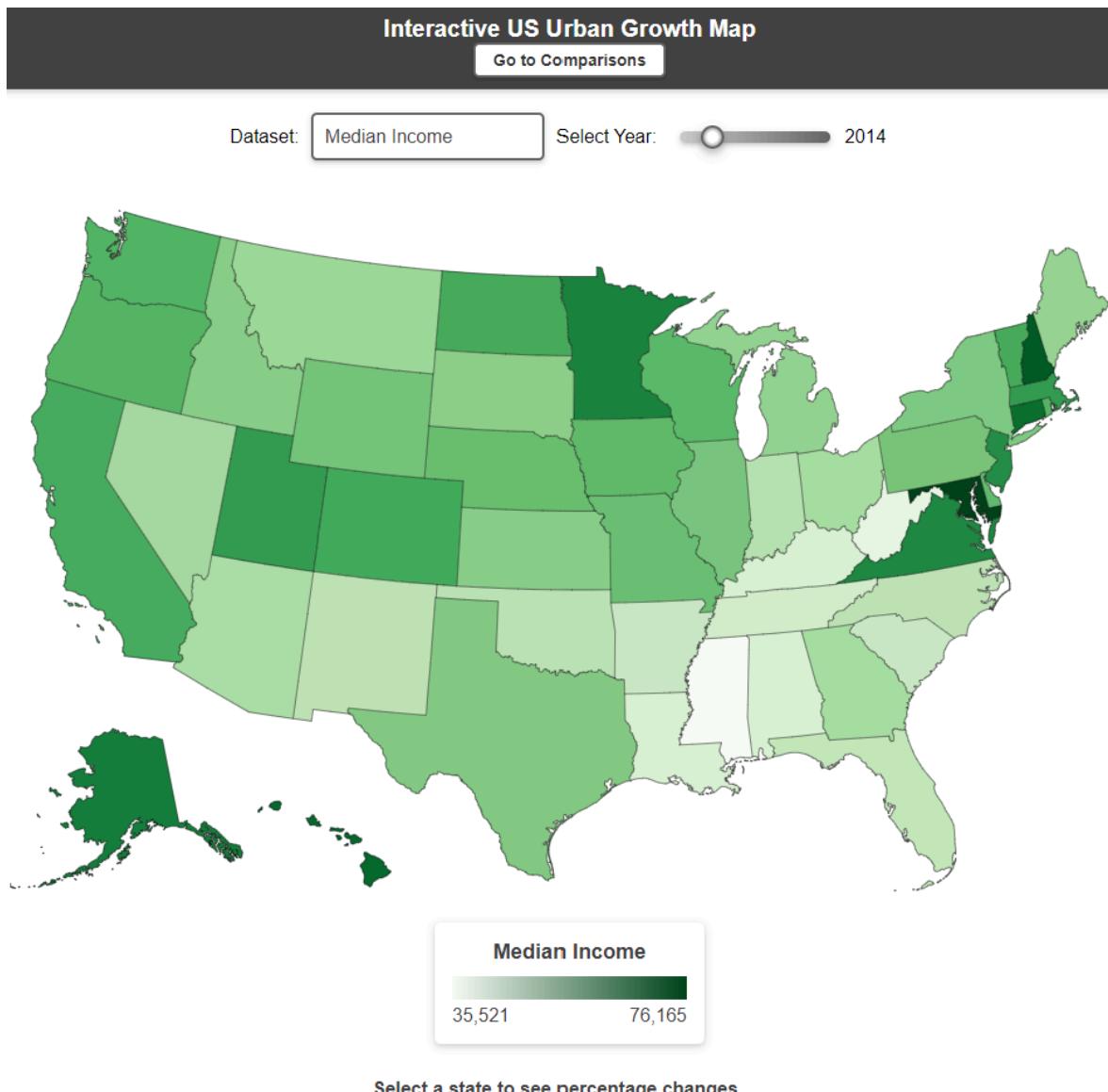


[Click for Table of Contents](#)

13th Code Entry

[Entry by Cody Mathews and Tony Zhang, 11-22]

We could not figure out the errors that were being thrown when attempting to share data between the map and the graphs so we decided that we are going to create separate pages for the map and the graphs that act independently from each other. Also updated .css properties to make the functionality and look more appealing. I also added a comparative text section below the legend for when states are selected it will give a synopsis of the change for that state over time.

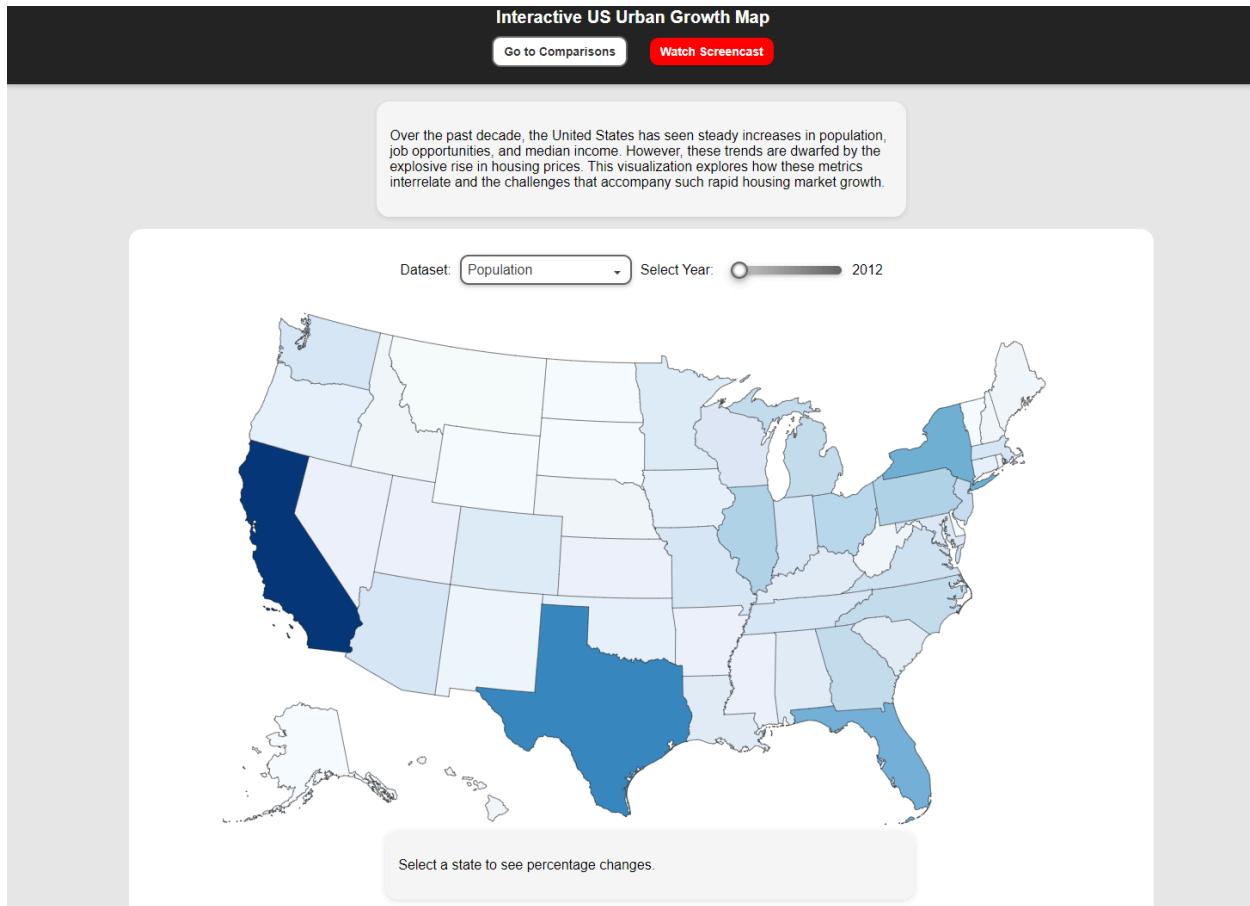


[Click for Table of Contents](#)

14th Code Entry

[Entry by Cody Mathews, 12-4]

Cleaned up the css properties for alignment and size. Made css properties the same for both pages for the drop down menu, buttons and slider.



[**Click for Table of Contents**](#)

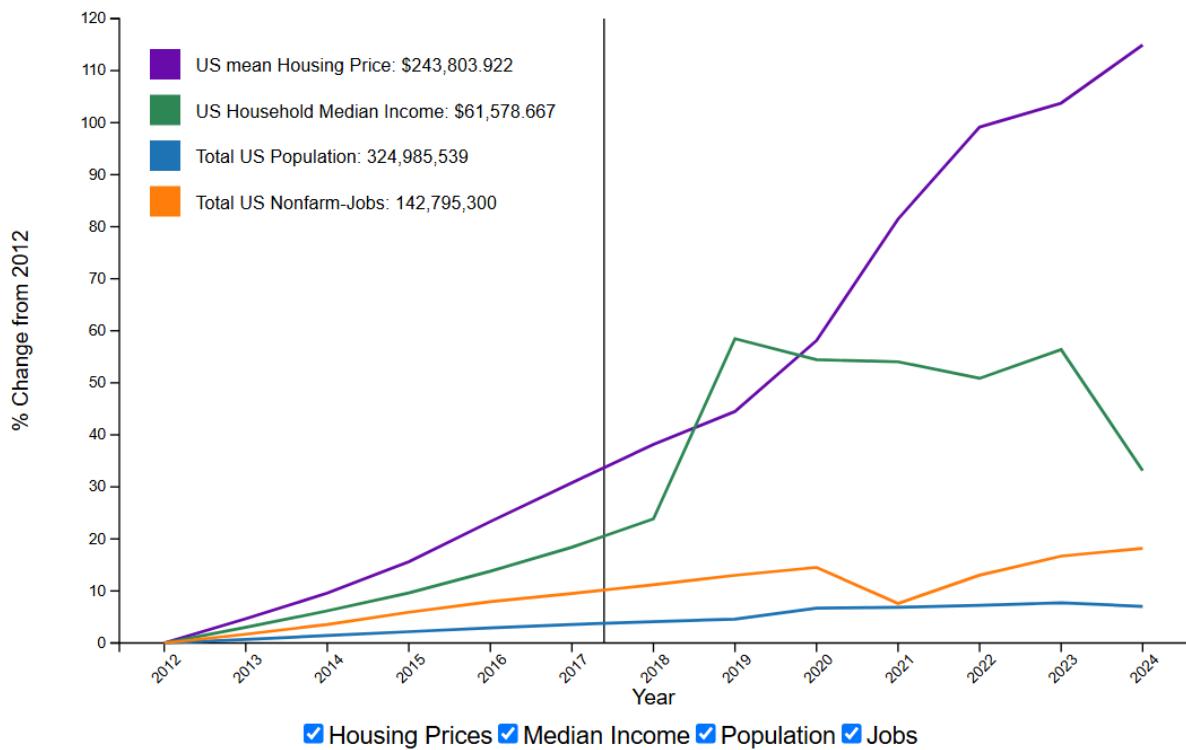


[Click for Table of Contents](#)

1st Comparison Code Entry

[Entry by Tony Zhang, 11-20]

The main goal with this other html page is to show the relationship between the datasets by having all datasets in each visualization. The first visualization took inspiration from [homework 4](#), we initially wanted to have this visualization be the exact same as the map visualization but couldn't figure out a very strange error. Decided to improvise by instead of selecting states we can select datasets to view. Also chose to graph the percentage change from 2012 as all of the datasets have different distributions, normalizing it or having multiple axes would've made it more complicated to understand. This approach is the most natural and easy to interpret.

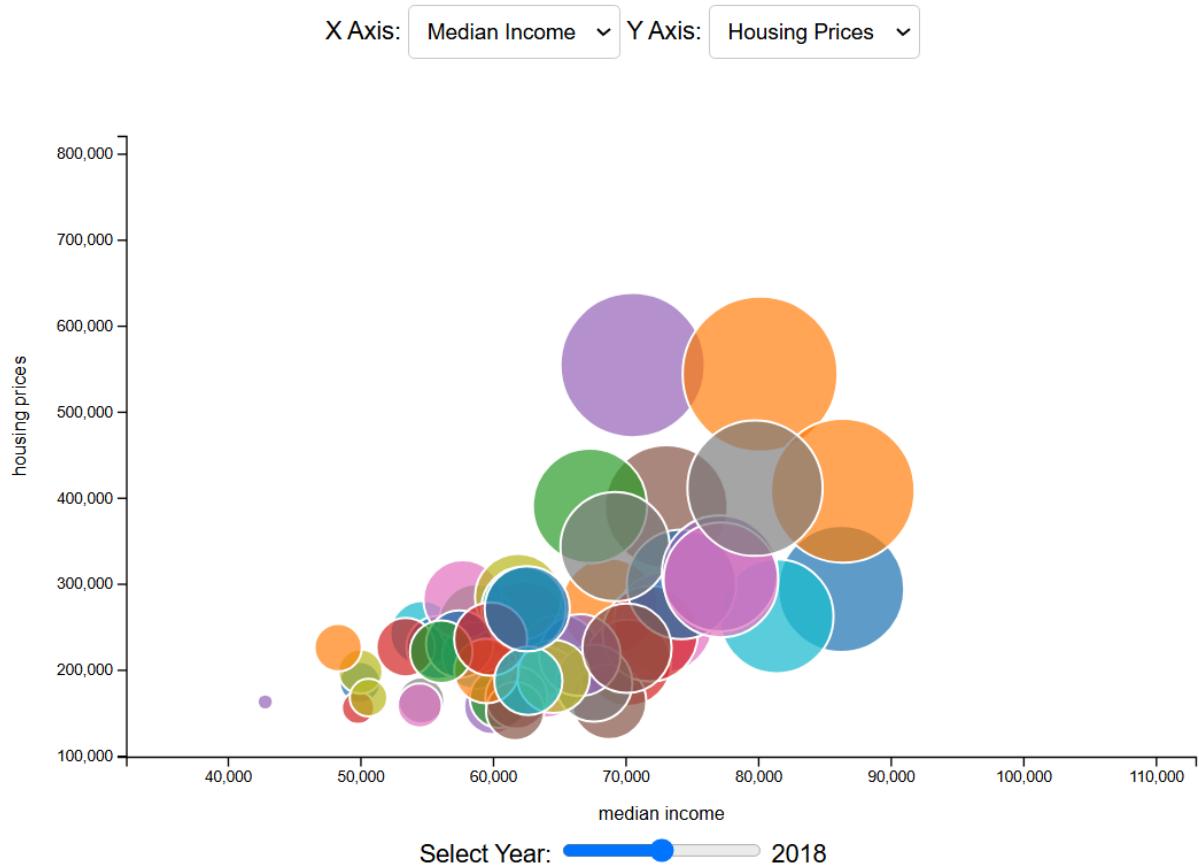


Also included animations to redraw the lines when the selected datasets are modified.

2nd Comparison Code Entry

[Entry by Tony Zhang, 11-21]

While the line chart does show the overall trends, I wanted to make a more interactive chart that allowed for a wider range of data tasks to be performed. Which is why I chose to make a bubble plot as my second visualization in the comparisons page. Typically in a bubble plot data points are represented using 3 variables, x position, y position, and size of bubble. I chose the X and Y axes to be able to be chosen by the user where the size of the bubble was a combination of the 2 variables. This way the states that have the largest values in the dataset will be the largest bubbles so as to be proportional to their overall impact on the entire US for the selected datasets. Also made it so the year could be chosen instead of the overall dataset to allow users to see how these datasets changed over time.

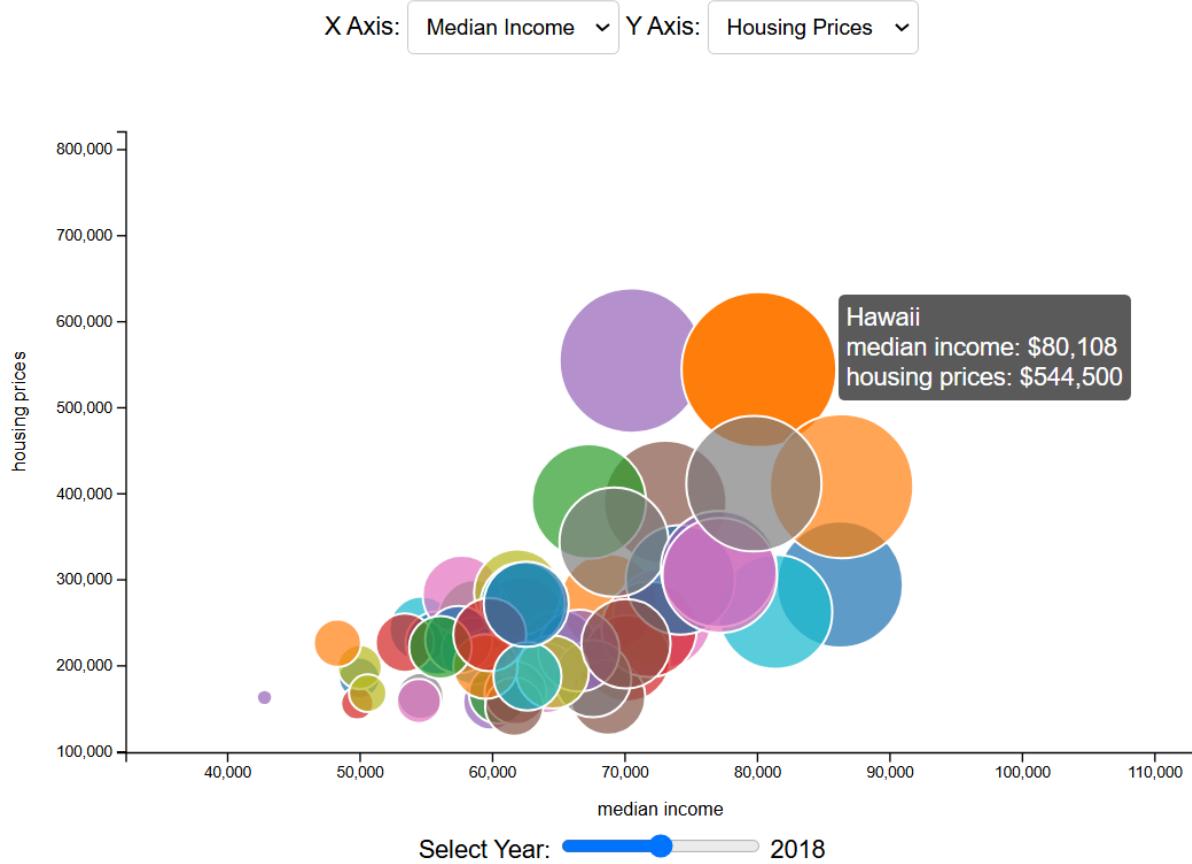


[Click for Table of Contents](#)

3rd Comparison Code Entry

[Entry by Tony Zhang, 11-21]

Added some additional features to the bubble chart. Added a hover feature that shows the name of the state and the values of the selected dataset for whichever bubble your mouse is on.

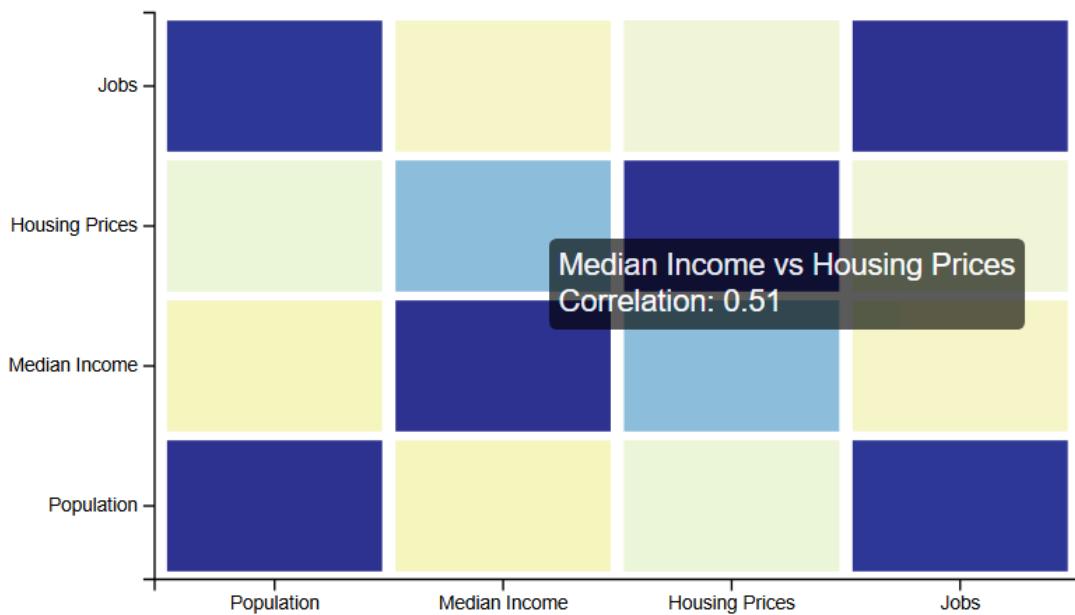


Also added a smooth transition animation when changing the year or dataset. The transition animation for the year made it much more aesthetic and easier to interpret how the states values for the dataset change over time.

4th Comparison Code Entry

[Entry by Tony Zhang, 11-21]

For a final visualization, I chose to make a correlation heatmap plot to show the strength of the relationships between the datasets.

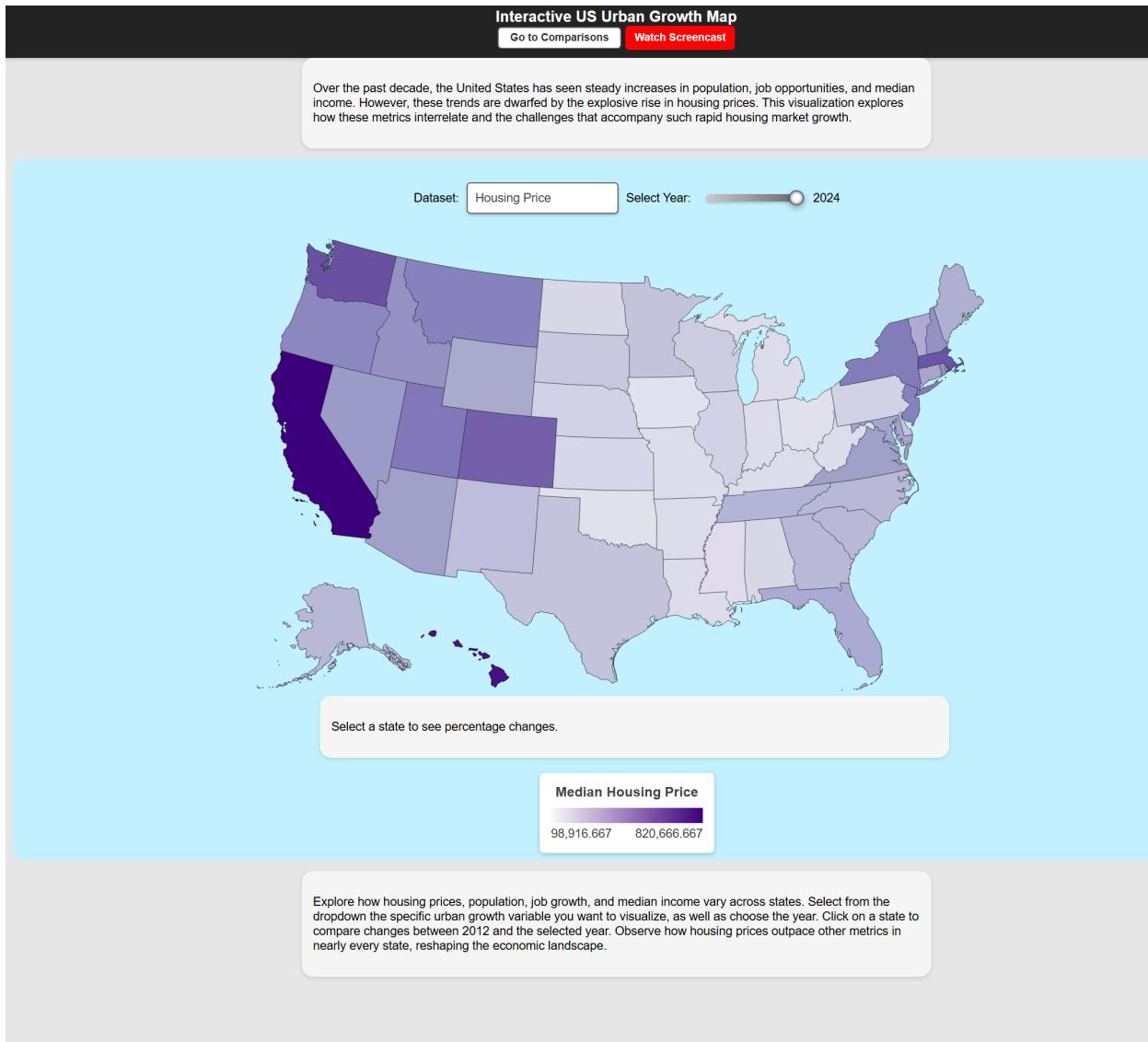


Also added a hover feature that allows the user to see the exact correlation value between the datasets for the specific cell.

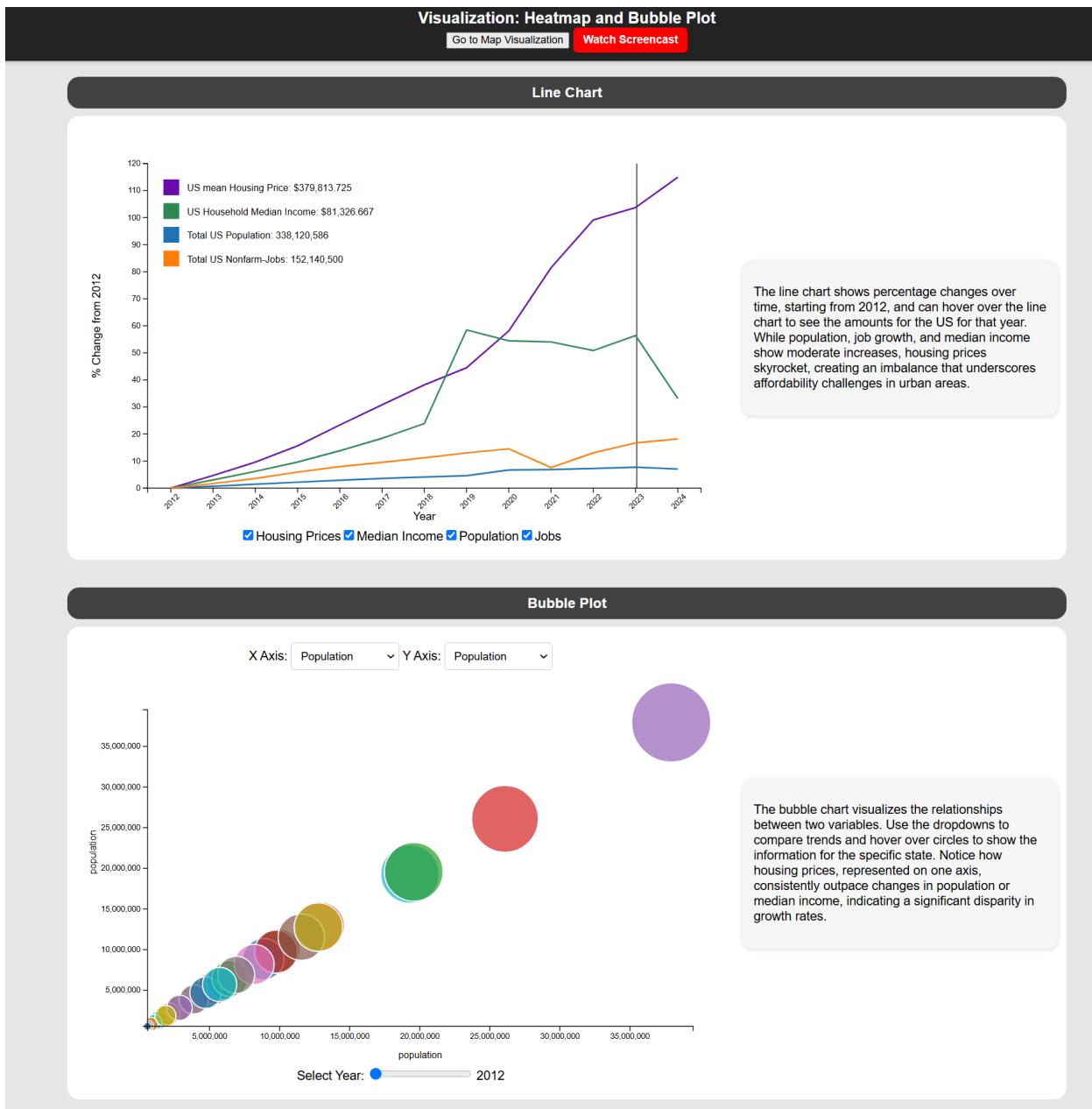
Finishing Touches

[Entry by Tony Zhang, 11-22]

Added some text blocks in all visualizations that describe them, how to use them and the key takeaways from the visualizations. Also added the hyperlink to the screencast video and updated styling to make the page more aesthetically pleasing.



[Click for Table of Contents](#)



[Click for Table of Contents](#)

Evaluation

Data Visualization Inferences:

[Entry by Tony Zhang, 12-3]

The key inferences I took from these visualizations were that housing prices have drastically increased in the past 12 years compared to other urban growth factors, as highlighted in the comparison visualizations. The map visualization also showed the distribution of the increases in housing prices across the US, as it seemed the Mideast didn't increase as much as the states on the coasts.

Visualization Assessment

[Entry by Tony Zhang, 12-4]

Despite the initial hiccup of trying to share the data from the map visualization with other visualizations, I am very happy with the end result. I believe all the visualizations are very functional and informative.

For improvements, I believe that having additional functionality with the map visualization would allow for a broader range of visualization tasks. I also believe in adding some comparison-based visualizations that allow the user to directly compare the data from certain states/regions to other states/regions. Finally, I believe that the inclusion of more datasets such as including political affiliation, crime rates, states with rent control, cost of living, etc. could further improve the visualization even more.

THE END...for now