State Analysis and Recommendations using Visualizations in D3.js

1. Basic Information

- Team Member 1
  - Name : Tanveen Singh Bharaj
  - U-Id : U1126027
  - U-Mail : tanveen.singh@utah.edu
  - Course : MSIS – Fall 2017

- Team Member 2
  - Name : Vishal Pandey
  - U-Id : U1126029
  - U-Mail : vishal.pandey@utah.edu
  - Course : MSIS – Fall 2017

- GitHub Repository : https://github.com/vishalpandey2192/DV_Project.git

2. Overview and Motivation

This proposed project aims at visualizing the trends of unemployment rate, average salary, mortality rate, population and rental cost determined by price per square feet for the years 2010-2016 in all states of USA. As per our research and study, we choose these five parameters as the most important ones which directly impact the likability and habitability in a particular state. We aim at improving decision making and recommend best states to live through visualization of trends and weighted mean algorithm.

The motivation for this project was the struggle that people especially international students/employees have in evaluating that which state is best for them for studies or job. We ourselves being international students know how much research we went through, researching different websites and articles, reading reviews to know that how life could be in a particular state we choose to live. Therefore to ease this risky process of selecting the most favorable state to live we leverage the power of data visualizing to analyze the tenability of each US state.

3. Related Work

We ourselves developed a python plugin for QGIS (Quantum Geographice Information System) in collaboration with Government of India, that through that simplified the task of relocating people living in slums of Delhi,India to permanent residences. The plugin through it's pictorial representation, heat maps, visual animation and predictive analysis helped the government officials in formulating strategies for slum dwellers reallocation. Our previous work in the domain of providing recommendation through visualization laid foundation for this project.

The blogs and posts that we followed for deciding our parameters for state recommendation are:

https://career.du.edu/blog/2015/04/24/5-important-factors-to-consider-when-choosing-your-dissertation-research-topics/

https://www.moneycrashers.com/where-should-i-live-decide-best-places/

https://www.entrepreneur.com/slideshow/299849

https://lifehacker.com/top-10-ways-to-find-the-best-place-to-live-1634031154

Apart from these blogs there were many other questionnaires we followed where people used to talk about the factors they consider while relocating.
Also, we found www.zillow.com very helpful. We followed they blogs where they talked a lot about peoples preferences about states. Moreover we are also using the pricing dataset from here.

4. Project Objectives/Questions Addressed

There are three primary questions that we are trying to answer through the scope of our project:

1. First, what are the current trends in terms of jobs and salary across all states of USA?
   We answer this question by illustrating visualization of current trends across all states of USA for population, unemployement rate, and average salary, rental cost (price per square feet) and mortality rate. These current trends will allow user to get present picture of trends across states.
2. Second, I give more priority to salary as opposed to population of state, but your visualization gives more weightage to population and not salary, hence your recommendation tends to differ my preferences. How will your visualization resolve this?
   Our visualization gives user the flexibility to rank the parameters in his order of precedence. Based on the user priority, we recommend him top 3 US states to live in based on weighted mean algorithm.
3. Finally, as a user, I want to see broader view of trends and not trends limited to single year
   We give the user the flexibility to visualize and analyze the trends of population, unemployement rate, average salary, price per square feet and mortality rate among the top 3 recommended states for the years 2010-2016 as opposed to single year in first question, which will give users broader perspective to choose the best state out of three.

5. Data

We collected data sets for the years 2010-2017 for the following parameters:
population, unemployment rate, average salary, mortality rate and rental cost (price per square feet).
The data sources are:
Population - https://www2.census.gov/programs-surveys/popest/datasets/
Rental Price measured through price per square feet –
https://www.zillow.com/research/data/#rental-data
UnEmployment rate –
https://data.bls.gov/map/MapToolServlet?survey=la&map=state&seasonal=u

Mortality Rate – http://ghdx.healthdata.org/us-data
http://datacenter.kidscount.org/data/tables/6051-infant-mortality#detailed/2/2-52/false/573,869,36,868,867/any/12718,12719
Average Salary –
https://www2.census.gov/programs-surveys/acs/summary_file/2015/data/5_year_by_state/

6. Data Processing

1. The data sets we received were of county level for each state containing 15000 rows and our visualizations are state level so our first data processing involved converting the data from county level data to state level by aggregating values.
2. The rental cost measured in price per square feet was month wise for each year, which we averaged to get price per square feet data annually.
3. The data we collected was on the basis of parameter like average salary data for 2010-2016 in one csv, unemployement rate for 2010-2016 in one csv, we converted this to yearwise by aggregating all parameters under one csv file for a particular year. For example 2010_us_states_data.csv will contain data for unemployment rates for 2010, salary for 2010 etc.
4. In some datasets the value give are not standardized like for the rents data set in some files the rent was in price per square feet while at some places meter square so we need to standardize it to price per square feet.

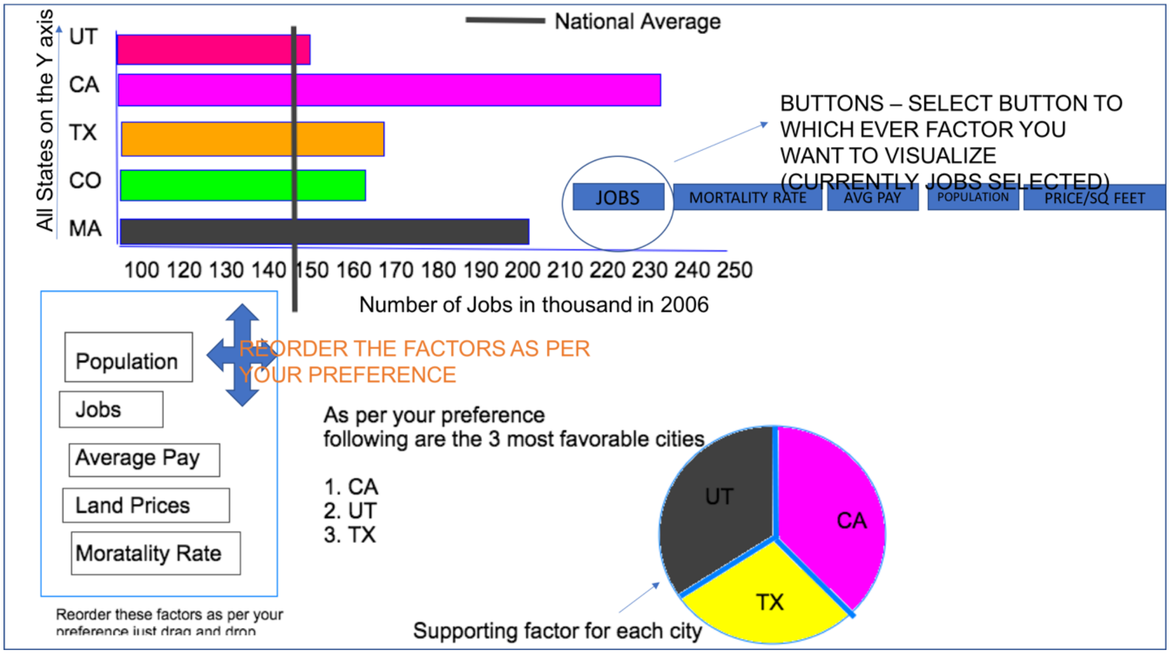
7. Exploratory Data Analysis

We used color coding and line/position channel to analyze our data. I believe having a map view of the effect of different factors over the united states , helps us understand how these factors are associated with each other. Like the population distributions is very disparate throughout the united states the major population lies within major cities like CA and NY. Similarly is the unemployment rate , the states which has more population they tend to have higher unemployment rate although the average pay are higher in those cities.
Such insights helps us more to understand the basis of the recommendations which we plan to give to the user based on his priorities in the second phase.

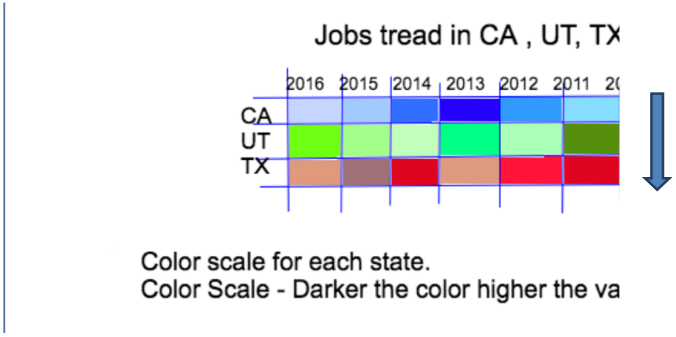
8. Visualization Design

1. Design 1

We initially thought to visualize our design in 2 phases; where in the first phase we will show the visualizations and data for all the states of USA for all parameters for the year 2016 using bar chart. This will allow user to analyze the trends across all the states at once to get a sound understanding of current trends.
And in the second phase, we will show data corresponding only to the 3 recommended states based on user preferences of parameters using color grid map.
Our main aim for the second phase is to show trends in parameters over past 10 years so the user has better understanding of those states.


National Average
All States on the Y axis
UT
CA
TX
CO
MA
100 120 130 140 150 160 170 180 190 200 210 220 230 240 250
Number of Jobs in thousand in 2006
BUTTONS – SELECT BUTTON TO WHICH EVER FACTOR YOU WANT TO VISUALIZE (CURRENTLY JOBS SELECTED)
JOBS
MORTALITY RATE
AVG PAY
POPULATION
PRICE/SQ FEET
Population
Jobs
Average Pay
Land Prices
Moratality Rate
REORDER THE FACTORS AS PER YOUR PREFERENCE
As per your preference following are the 3 most favorable cities
1. CA
2. UT
3. TX
UT
CA
TX
Supporting factor for each city
Reorder these factors as per your preference just drag and drop

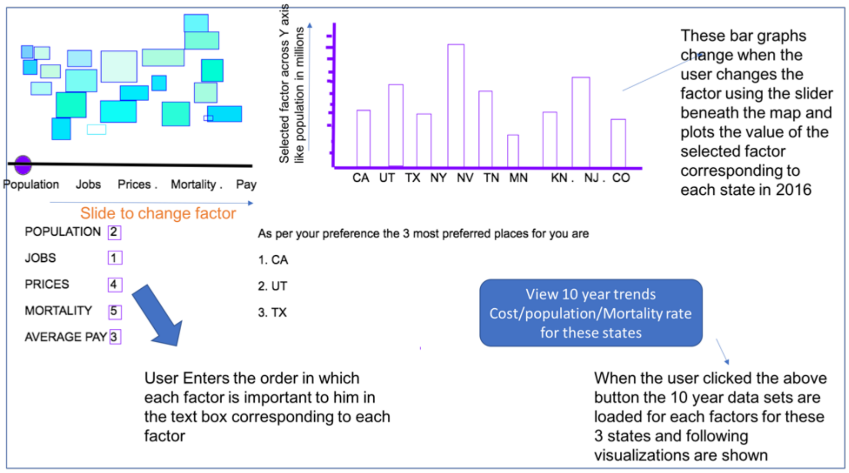
Once the user receives the top 3 recommendation, he can select to view the yearly trends in those 3 states for each of the parameter to get deeper insights of those states


Jobs tread in CA , UT, TX
2016
2015
2014
2013
2012
2011
20
CA
UT
TX
Color scale for each state.
Color Scale - Darker the color higher the va

In this frame the user sees color grid map for each parameter he chooses to view and analyze.
As in the above visualization we plot the job trends from past 10 years and their number (denoted by the color channel) in each of the 3 states.
We will have 4 similar color grid map for each of the other parameter which are population, average salary, mortality rate and rental cost in price per square feet.

2. Design 2

In design 2, in our first phase illustration, we planned to use tile chart of the United States and follow intensity based color coding to mark the intensity of value for the selected parameter on each state.
The parameter to view is selected from horizontal line chart below the map by moving the marker towards the parameter.


Selected factor across Y axis like population in millions
CA UT TX NY NV TN MN KN . NJ . CO
Population Jobs Prices . Mortality . Pay
Slide to change factor
POPULATION 2
JOBS 1
PRICES 4
MORTALITY 5
AVERAGE PAY 3
View 10 year trends Cost/population/Mortality rate for these states

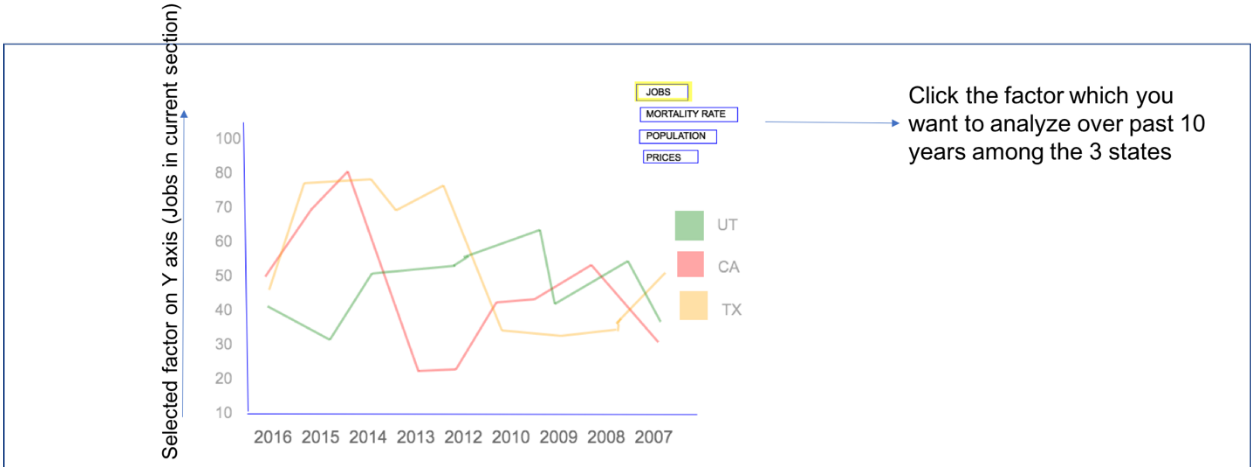
These bar graphs change when the user changes the factor using the slider beneath the map and plots the value of the selected factor corresponding to each state in 2016

As per your preference the 3 most preferred places for you are

1. CA
2. UT
3. TX

User Enters the order in which each factor is important to him in the text box corresponding to each factor

When the user clicked the above button the 10 year data sets are loaded for each factors for these 3 states and following visualizations are shown



Click the factor which you want to analyze over past 10 years among the 3 states

The above visualization is for the 3 recommended states which the user will get after applying weighted mean algorithm. In this visualization the user can select the parameter he wants to analyze, view and compare among the 3 states over 10 years using line chart.

3. Design – 3 (Final Design)

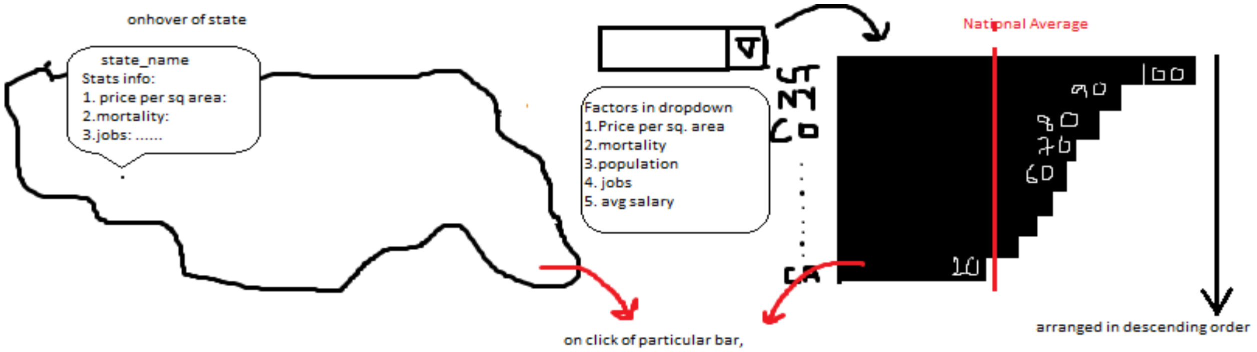
In design 3, we divided our layout to three sections.

1. In this first section, we will create map of USA using geoJSON.
    - This is an interactive map where the visualizations change when you choose another parameter from dropdown or when you hover over or click on any particular state.

- When you change the parameter from the dropdown, visualization changes on map depending on value for that parameter for a particular state for the year 2016.
- Also, a bar chart for all states of USA is plotted in descending order of values for that parameter.
- When you hover over particular state on map, tooltip appears showing additional information about that state, along with highlighting the corresponding bar on the bar chart
- When you click on particular state on map or bar, a table is displayed showing additional information about that bar and comparing it with national average.

2. In the second section, there is a table which expects user input to rank the parameters; 5 being the highest and 1 being the lowest. Based on user rankings and applying weighted mean algorithm, we recommend top 3 states to user to live in.

3. In the third phase, as similar to second design, we show the trends in the parameters considered over the past 10 years for the recommended states but in addition to previous designs, we will use the brush to select the years you want to visualise.



on click of particular bar,
we will show a table depicting comparision of each factor with national average for that state

	Utah	National Average
Prices per sq. area	172	500
Mortality	4.2	7
Population	193744	120502
Jobs	70000	150000
Salary Data	72000	95000

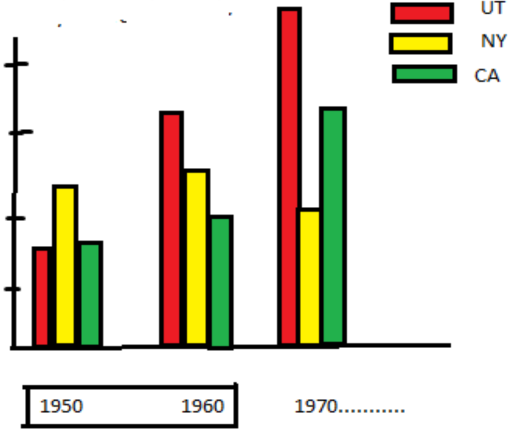
S.No	Factors	Input Weights (1-5: 1 lowest, 5 highest)
1.	Prices per sq. area	2
2.	Mortality	1
3.	Population	3
4.	Jobs	5
5.	Salary Data	4

Top 3 USA states based on your weights (based on weighted mean algorithm)

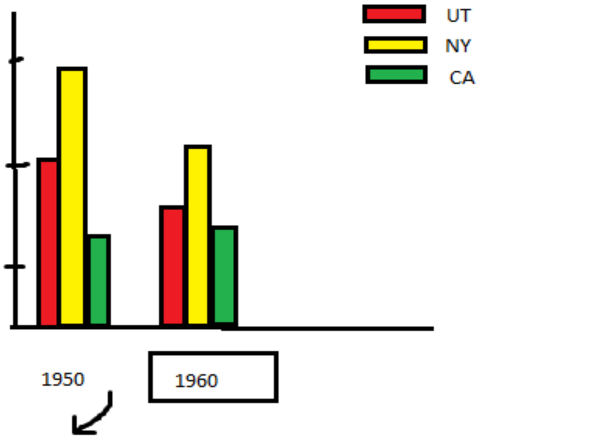
1. UT   click to see salary trends   click to see employement trend
2. NY
3. CA

} similar buttons for other 2 states

Comparing Avg Salary trends of past ten years for your top States



Comparing Employement trends of past 10 years for your top 3 states



Here we will use Brush - which will allow to select range of years whose data you want to focus. The data out of this brush range will be dimmed, so that focus is on selected year range

We finalized on the Design 3 based on multiple factors:

1. The map of USA added in 3rd design, will make the visualization very easy to grasp especially when used with color coding to show the intensity of each factor in each state.
2. Giving the user the flexibility to rank his preferences in a convenient and user friendly manner.

3. Also, the visualizations showing the trend of parameters for the years 2010-2016 for each of the 3 recommended states is very crisp and graspable which helps the user further select the one state which suits him the most out of three.

9. Implementation

As discussed above throughout, project scope is divided into 3 sections:

1. Visualization of trends of different parameters which are:
    i. Unemployement Rate
    ii. Mortality Rate
    iii. Population
    iv. Average Salary
    v. Rental cost determined by : price per square feet

For all states of US for the year 2016.
In this section we analyze in all states of the United States the trend of the chosen parameter out of five from the dropdown.

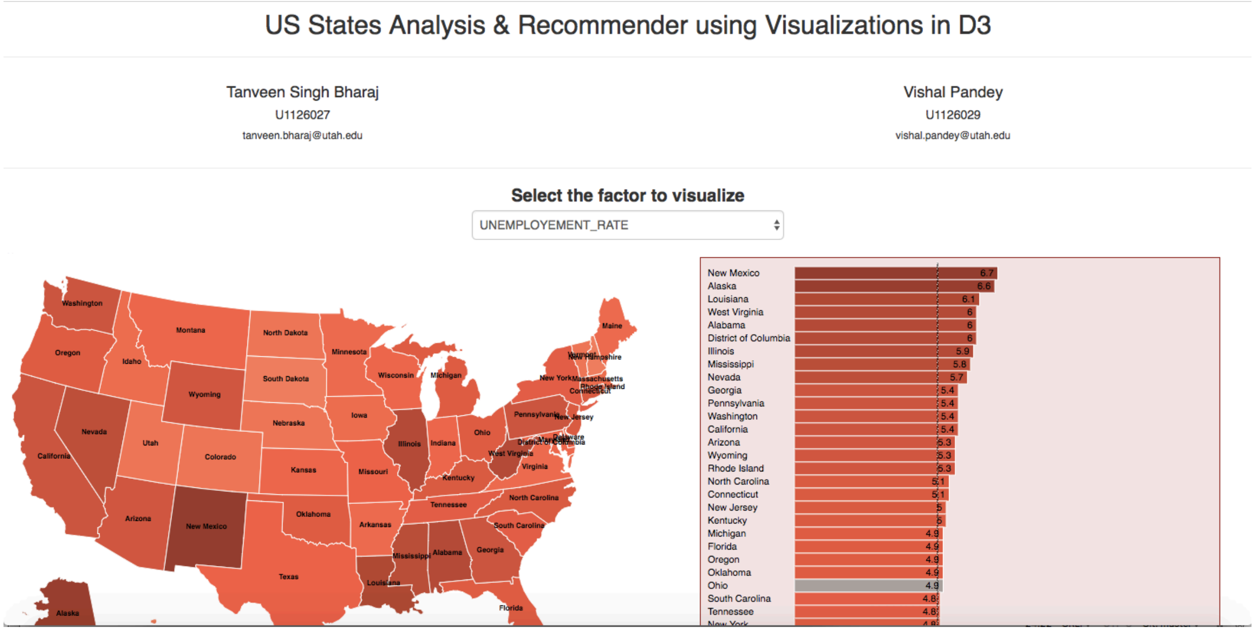
In this section we create visualization of US Map using geoJSON where we show the impact of each chosen parameter on each state for year 2016 using color coding channel/color scale that makes it easier to visualize that how big or small is the impact of each factor on each state of US.

Further for each parameter, we will use area channel encoding using bar chart to represent the trend of this parameter in the year 2016 across all states. The bars of bar chart will be arranged in descending order going from state having maximum value to state having minimum value. The aim is to make information grasping easier for users by arranging data in a particular order. The area chart will also have line marker which would indicate the national average of that parameter so we can compare the value for each states with the national average and draw better inferences.

There will be interactions involved on hover and click of state on map as well as click on bar of bar chart.

Since, in the first project milestone we have completed the first section mentioned above and the following are the screenshots

Screen 1 – The homepage



US States Analysis & Recommender using Visualizations in D3

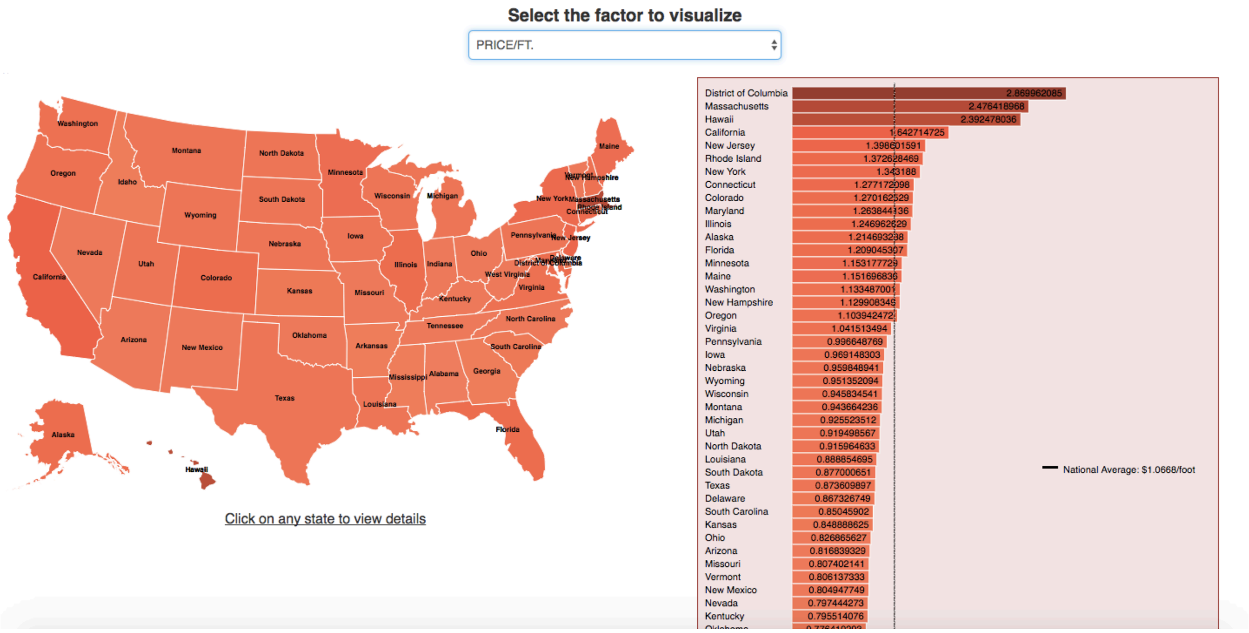
Tanveen Singh Bharaj
U1126027
tanveen.bharaj@utah.edu

Vishal Pandey
U1126029
vishal.pandey@utah.edu

Select the factor to visualize

UNEMPLOYEMENT_RATE

Screen 2 – On parameter change from drop down the , map and bar charts refreshes as per the data corresponding to the parameter selected.

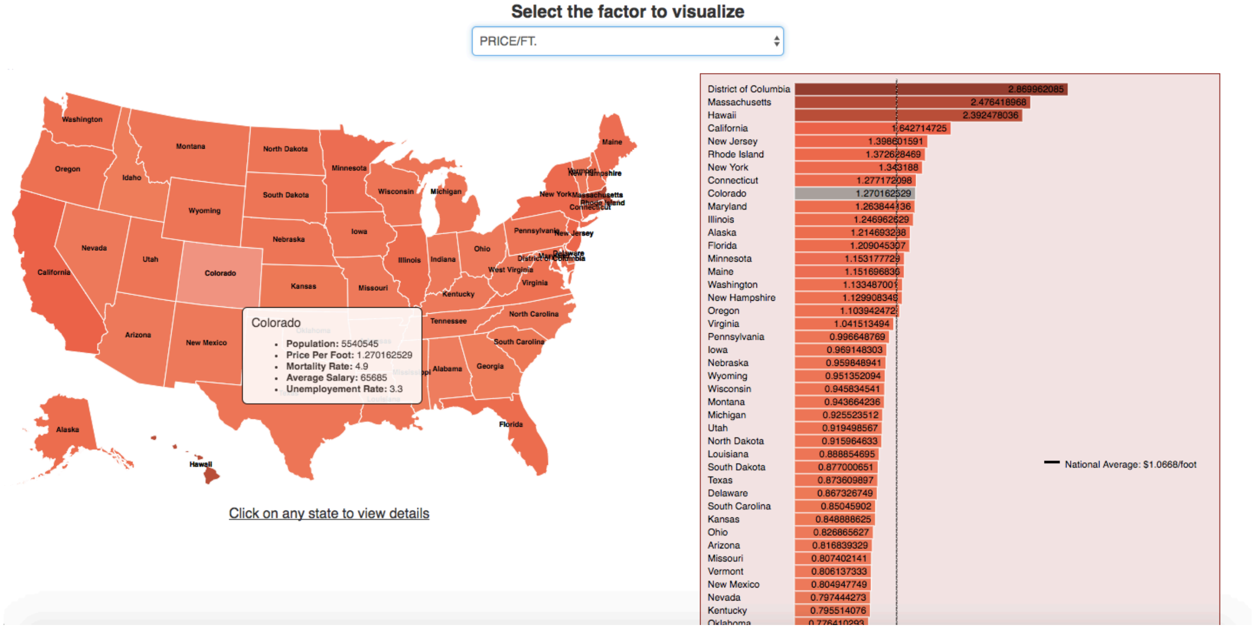


Select the factor to visualize

PRICE/FT.

Click on any state to view details

Screen 3 – When you hover on any state on the map – more details corresponding to the state appear in a tooltip and the same state is highlighted in the bar chart



Select the factor to visualize

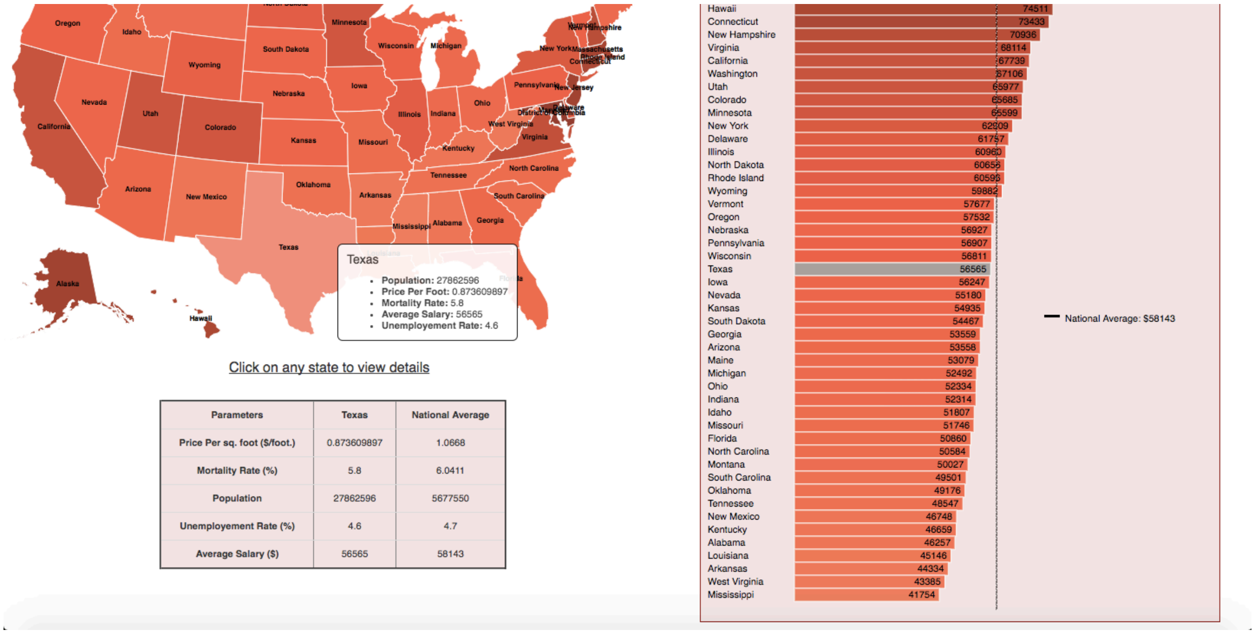
PRICE/FT.

Colorado

- **Population:** 5540545
- **Price Per Foot:** 1.270162529
- **Mortality Rate:** 4.9
- **Average Salary:** 65685
- **Unemployement Rate:** 3.3

Click on any state to view details

Screen 4 – When you click on a country a table is displayed which shows the value of different parameters for the current state with the average country participation.



Texas

- **Population:** 27862596
- **Price Per Foot:** 0.873609897
- **Mortality Rate:** 5.8
- **Average Salary:** 56565
- **Unemployement Rate:** 4.6

Click on any state to view details

Parameters	Texas	National Average
Price Per sq. foot ($/foot.)	0.873609897	1.0668
Mortality Rate (%)	5.8	6.0411
Population	27862596	5677550
Unemployement Rate (%)	4.6	4.7
Average Salary ($)	56565	58143

2. In the second part of visualization, we aim to create a visual recommendation system where we recommend the top 3 states to users based on the user preference.
   We achieve this by using the concept of weighted mean algorithms.
   In this approach we provide list of major parameters to users which he should consider while selecting a state, and the user specifies the degree of importance of each paramter to him by assigning a number from 1-5 to them, 1 being the least important and 5 being the most important.
   Using the weight assigned by the user to each factor we compute a variable called the suitability factor which is calculated by the weighted mean algorithm.

The suitability factor is calculated by multiplying the weight assigned by the user to each factor with the value of factor and then dividing number obtained for each state by the minimum value obtained.

S.No	Factors	Assumed Weights (1-5: 1 lowest, 5 highest)
1.	Price per sq. square feet	2
2.	Mortality	1
3.	Population	3
4.	Jobs	5
5.	Salary Data	4

Sample Data for Year 2016

S.No	State	Prices per sq. area ($)	Mortality Rate	Population	Jobs	Average Salary Data($)
1.	Utah	172	4.2	193744	70000	70000
2.	New York	400	6.2	1059852	200000	100000
3.	Nevada	250	4.8	452351	30000	60000
4.	Kansas	150	7	337896	40000	65000
5.	Arizona	300	5	478952	150000	80000

On Applying weighted mean:

S.No	State	Summation of values*weight	Calculated Sum	Final value= sum/min(sum)
1.	Utah	1*4.2+2*172+3*193744+4*70000+5*70000	1211580.2	1
2.	New York	1*6.2+2*400+3*1059852+4*100000+5*200000	4580362.2	3.78
3.	Nevada	1*4.8+2*250+3*452351+4*60000+5*30000	1747557.8	1.44
4.	Kansas	1*7+2*150+3*337896+4*65000+5*40000	1473995	1.22
5.	Arizona	1*5+2*300+3*478952+4*80000+5*150000	2507461	2.07

Top 3 states:

1. New York
2. Arizona
3. Nevada

Now, the top 3 states with the maximum value of suitability factor are suggested to the user.

3. As, an add-on once we have given the top 3 suggestions to the user, we further use more visualizations to give him a clear picture of his choice.
   We will use line chart to show the trends of the any chosen parameter by user over the past years from 2010 to 2016 for the top 3 suggested states. Such visualizations helps us to

understand the changes which happened to the state over the past couple of years and also visualize that seeing the past what could happened in the future.