Visualizing Relocation Factors

AUTHORED BY: CHEN YE. JACK TAO. KALEN WONG. LENA YEOH

DECEMBER 2014

ABSTRACT

This paper introduces a visualization tool our team created to assist people in making relocation decisions and describes the research and design process that supported its design. Our team began with a research interest in migration patterns and the factors that people used when considering to relocate. We selected official government data sources with a focus on population demographics, weather, and economic factors. Established design principles were incorporated into design iterations, and exploratory user research and usability testing were utilized to further refine designs. As a result, we delivered a visualization tool that is supported by rich data, with a set of features that allow users to view and compare factors of cities that affect relocation.

INTRODUCTION

Deciding to relocate to a different city is a big decision for people. Deciding whether to relocate, or if it is worth relocating, is a difficult and tedious process. It often involves an individual doing a lot of research and looking for different types of information from numerous sources. The information often comes in pieces, in various formats, and from sources with different levels of trustworthiness. An alternative way is to seek information directly from government websites like the U.S. Census. However, data from government sources are often complex, unintuitive and can be overwhelming.

Our goal of this project is create a visualization tool that retrieves multiple government data sources and convert these data into more readable and consumable formats. We focused on several key economic and social areas affecting relocation choices - housing cost, wage, expenditure, demographic and weather information. Our data sources were imported from the U.S. Census Bureau, Bureau of Labor Statistics, and the National Oceanic and Atmospheric Administration. Our scope includes 11 metropolitan areas (Seattle, Los Angeles, Boston, New York, Chicago, Dallas, Detroit, Miami, Philadelphia, San Francisco, Washington D.C) in the United States.

We adopted a user centered approach for this project by conducting user research, and incorporating results of usability study into an iterative design process. We used Tableau as our visualization design tool and we referenced academic literature and related work to inform our design. In this paper, we will walk through this creation process, introduce the features & user experience of this data visualization tool, and explain the insights resulting from the process.

RELATED WORK

There are numerous attempts to visualize similar data around population, jobs, or economic information. Some of these were presented as static reports, while others are animated visualizations that support rich user interactions. In our research, we referenced both types to inform our design.

The city-level economic summary report by the U.S. Bureau of Labor Statistics is one common type of reports displaying information about the economic health of a city.

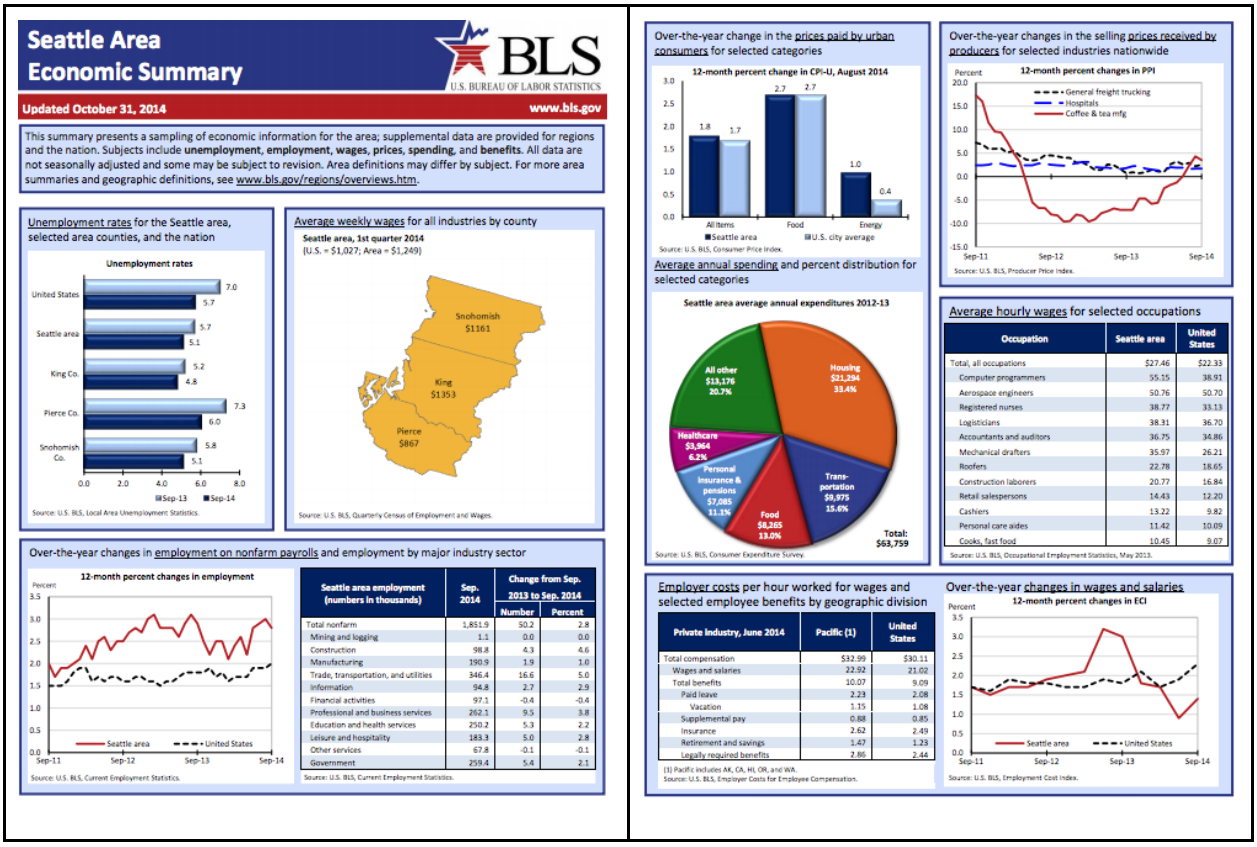

FIGURE 1 - SEATTLE AREA ECONOMIC SUMMARY - BUREAU OF LABOR STATISTICS

This type of report has the advantage of providing users with data from various categories at a glance. For example, categories of data include unemployment, employment, wages, prices, spending, and benefits. However, since the visualization is not interactive, further exploration of specific areas of interest is limited. It is difficult to draw relationships between different pieces of content. The different charts lack a cohesive storyline to assist users in understanding and relating the information to their needs.

We referenced this report (Figure 1) to primarily get an idea of what the important topics are for a similar audience, and we used this report as a framework to build upon and improve on.

We also looked at an interactive example by the U.S. Census Bureau. The Census Flow Mapper (Figure 2) is an interactive choropleth map displaying data on population flow. Users

can select on a location to display the net migration flow volume and can zoom in to see the county locations and their borders. On hovering over a location, annotations provide supplemental information for a county. Colors and values on the map were used to indicate positive or negative migration information. On the left, radio buttons enable users to choose the time frame, the migration type, characteristics, colors and the number of movies.

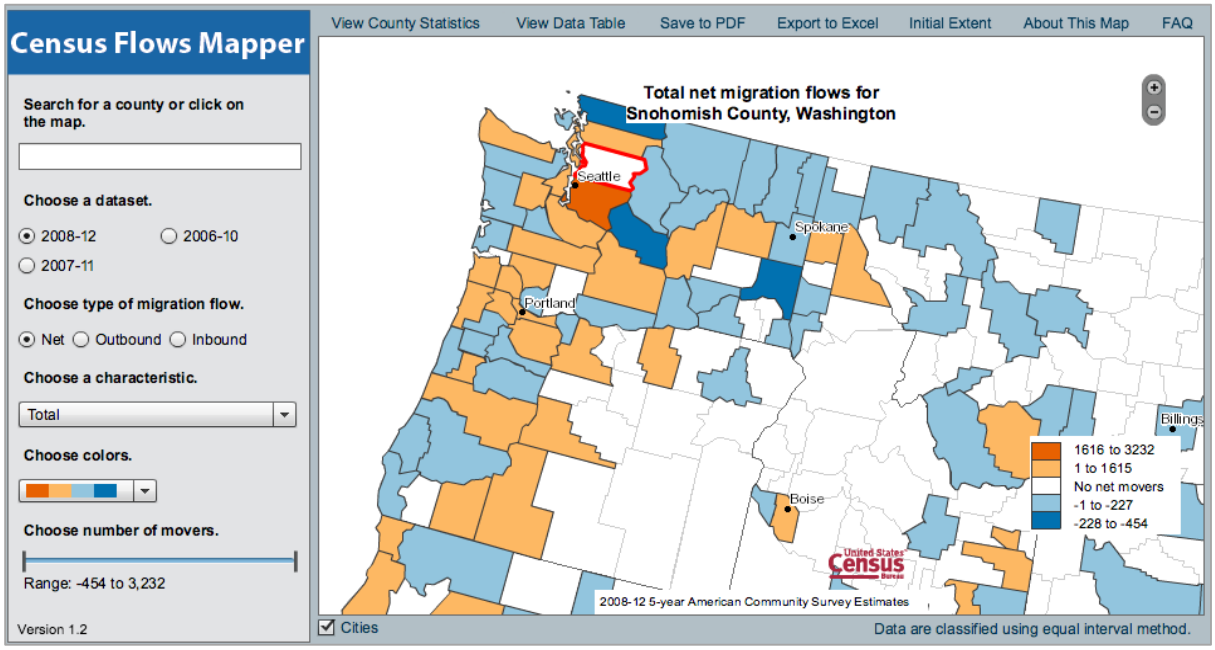

FIGURE 2 - CENSUS FLOWS MAPPER

The interactivity presented in this visualization influenced the design of our visualization. Since our topic of interest is based on location, a map with zooming functionality provides a good starting point for data overview. Filters allow users to further explore an area of interest, and selecting items on the map gives user additional details. However, we also found areas we could learn from an improvement. For example, the use of color was ineffective. The color palette selection under the dropdown list was used for aesthetic purpose only, and that the different colors used for positive and negative were arbitrary. Figure 3 illustrates an example where teal indicates negative and brown indicates positive, where darker brown and darker teal mean higher volume.

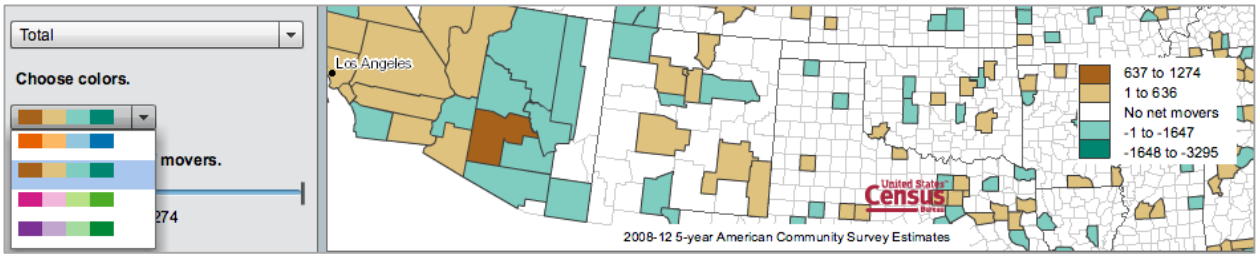

FIGURE 3 - CENSUS FLOWS MAPPER COLOR PICKER

We also observed that explanation text was inadequate in this visualization. We thought that a description of what data is visualized and explanations of how to interact with visualization would improve understandability.

There are examples from private sector we also looked into. The visualization created by New York Times for housing price is a great example (Figure 4) of an engaging interactive time series visualization.

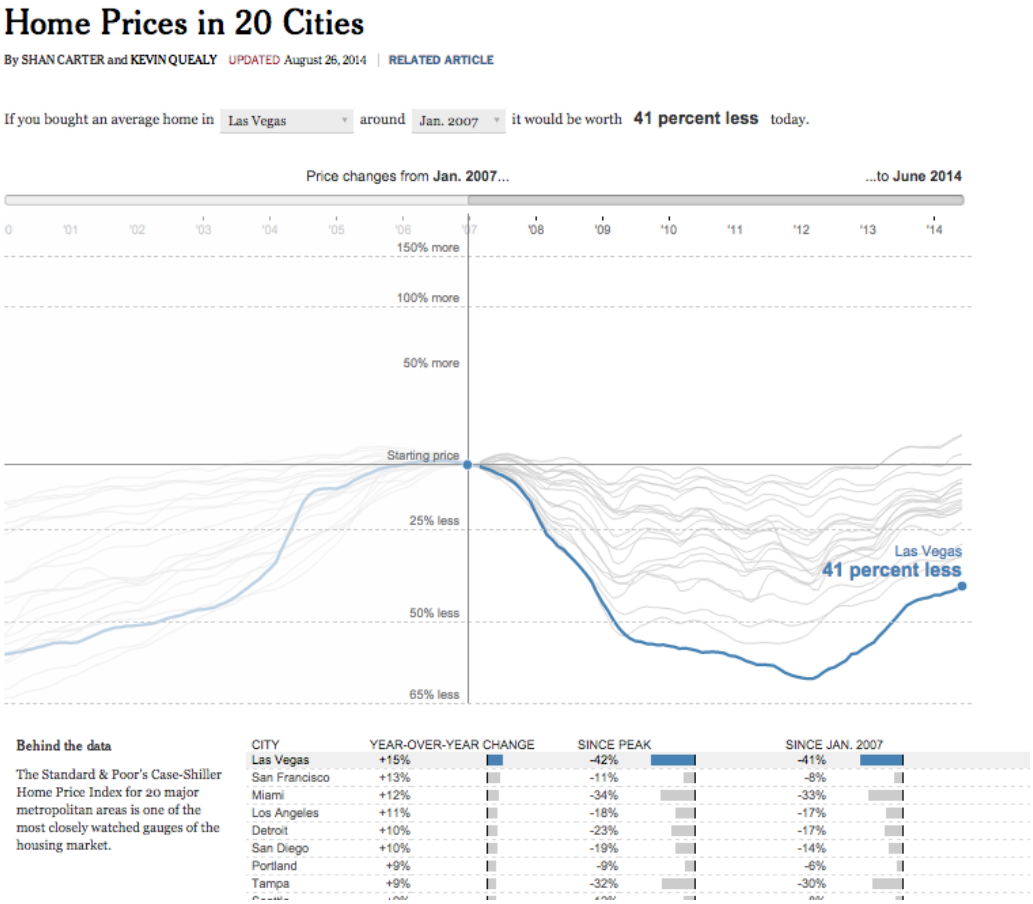

FIGURE 4 - HOUSING PRICES IN 20 CITIES. NEW YORK TIMES.

The use of single color line on a background of gray took advantage of the pre-attentive attributes of color, which established visual perception theories suggest draws the user's attention. The labels are particularly clear. Instead of using a numerical scale, a percentage-based relative scale was used that normalizes the data. On the x-axis, the scale included a description "Price changes from [Date]… to [Date]" that accurately summarizes the entire chart. Manipulating data was easy as visualization supported brushing and linking technique, where the table below highlights data corresponding to the selection on the time series above. The description under "Behind the data" on the Standard & Poor's Case-Shiller Home Price Index not only explained the source but reinforced the credibility of the data.

Our team incorporated elements that contributed to this visualization's success into our own visualization.

In our design, we utilized the same brushing and linking technique in the map displaying demographic information. We also included text explanations that established the trustworthiness of our data sources used.

DATA

At the earlier stage of the project, a significant amount of effort was devoted to data gathering and cleaning. The following describes the sources and datasets used, the gathering process, and the challenges in cleaning and getting them ready for our visualization design.

SOURCES

Initially we sought information and inspiration from sources across the Internet. These included publications like Forbes, The New York Times, Real Estate websites, etc., as well as government agency websites. However, later we focused only on government data due to its neutrality and comprehensiveness. The data sources we used were the U.S. Census, Bureau Labor of Statistics, and the National Oceanic & Atmospheric Administration.

Population: Our population data came from the results of annual American Community Survey (ACS) conducted by the Census Bureau. The survey is conducted annually, and samples 1 in every 38 households in United States. ACS provides detailed estimated count of people by marital status by sex and age group for 343 major counties in America, and also detailed estimated count of people by marital status by race for hundreds of major counties in America. The data set includes counties with population larger than 65,000 people. This data file is very large. The detailed tables in the survey we used are listed in appendix.

Weather: Weather data originated from information published in Annual Climatological Summary from the National Climatic Data Center of the National Oceanic and Atmospheric Administration. Data was from last 6 years (2008-2013) and was captured by weather stations located at the major commercial airports in each city.

Housing: The annual American Community Survey (ACS) also provided us with a breakdown of monthly housing costs for 817 major counties in the United States. The data included the estimated count of the number of housing units in a given range of costs for each county, and also the margin of error at 90% confidence level. The monthly housing costs factors in ownership and mortgage costs for homeowners and rent costs for renters.

Employment & Wage: The source of our employment and wages data came from the Occupational Employment Statistics (OES) Survey posted on the Bureau of Labor Statistics website. The latest data were released on April 1, 2014 for the month of May 2013. The data represented 24 occupational groups and 415 different occupational titles. It provided us the overall employment numbers, the types of occupation, and their respected mean wage.

Economic: Our economic data also came from the Consumer Expenditure Survey from the Bureau of Labor Statistics website. Expenditure data were manually extracted from individual data table for each city.

Job openings & Education: We also explored data on job openings from the Census Bureau and Education/School data from the Bureau of Labor Statistics. This data was not included in the final visualization due to changes in focus of the project.

PROCESS AND CHALLENGES

On the up side, having rich government data sources gave us more opportunities for analysis. On the down side, the processes of gathering, merging, and cleaning data was time-consuming and labor intensive. The volume of government data was enormous and the data tables were complex. We also realized that there was not one single source for all the data we needed. Data was distributed among different agencies with different formats. Often, the data provided by agencies overlapped and it wasn't clear how one agency differed from another and what criteria they used. For example, economic information by industry could be found on the Census Bureau website, whereas the Bureau of Labor Statistics website included major economic indicators such as price indexes. This made the selection of the appropriate data sources for our users very challenging.

We also ran into challenges in gathering consumer expenditure data, where we found data was only available for the past two years. This limited our ability to construct a time series scatter plot to reflect the relationships between average expenditure and wages. We also found that different cities exported their report formats quite differently and used different metadata in their files. This resulted in significant difference in indexed search results, and much effort was spent in searching for the right file on the Bureau of Labor Statistics website.

Merging data was a strenuous process. Due to the geographic dependency of our topic, we had to set a location boundary such as county, city, or state as our base criteria. However, since the granularity of our data sources varied significantly, our team had to manually create a custom mapping table to join different data that were coded in different location criteria such as county, metropolitan area, and FIPS county code. We realized that by doing so there are trade-offs in the precision of the data.

Though we did not observe a lot of missing data within each source, there were many challenges in the data cleaning process. One biggest challenge was the overall data volume. The huge volume not only make joining data tables difficult and inefficient, it also prevented us from publishing the visualization to Tableau public website, which has an upper limit of 1 million rows. We also had to perform a lot of manual calculation and converting units. For example, while the NOAA data reported temperature in tenths of degree Celsius (e.g. 250 = 25 °C), the data was converted to degrees Fahrenheit, a unit more familiar to our target persona. Weather data from past 6 years were averaged to improve accuracy and minimize bias due to climate variations from a specific year. The ratio of men to women in population demographics were calculated from separate population count data.

As for our graph on housing cost distribution, the original data we retrieved only included the absolute number of households. We converted the numbers into percentages, so the proportion of households paying a specific cost range in different cities are comparable. We also manually grouped inconsistent intervals that varied from $100 to $500 to a consistent interval of $500.

VISUALIZATION DESIGN PROCESS

We took a user-centered design approach in our visualization design, where we incorporated user research and usability testing in an iterative design process. Figure 5 describes the different undergoing activities we engaged in the creation of the visualization.

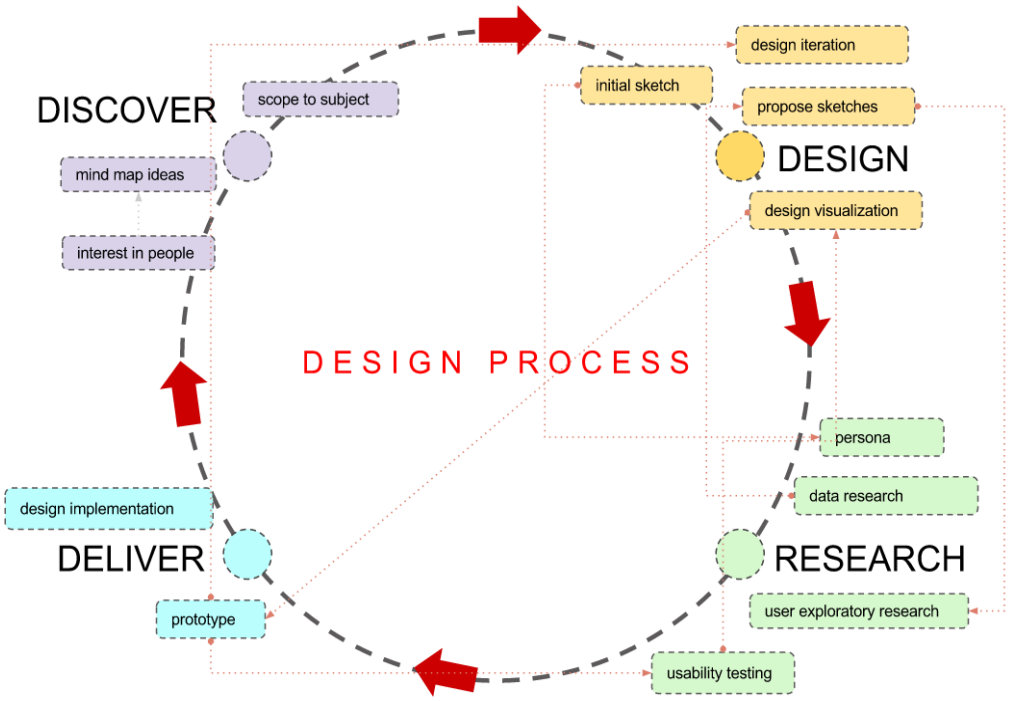

FIGURE 5 - DESIGN PROCESS DIAGRAM

BRAINSTORMING & IDEATION

We began our design process by having a team brainstorming session. We iterated on multiple mind mapping diagrams and landed with the idea of the subject. First, we split out ideas and words from all different perspectives without any limitations. We wanted to see where this process would lead us. We finalized our idea with the keywords such as marital status, reasons to move, decision factors and single segments for population which resulted in the idea of factors that impacted relocation decisions. Figure 6 illustrates the mind map for our ideation process.

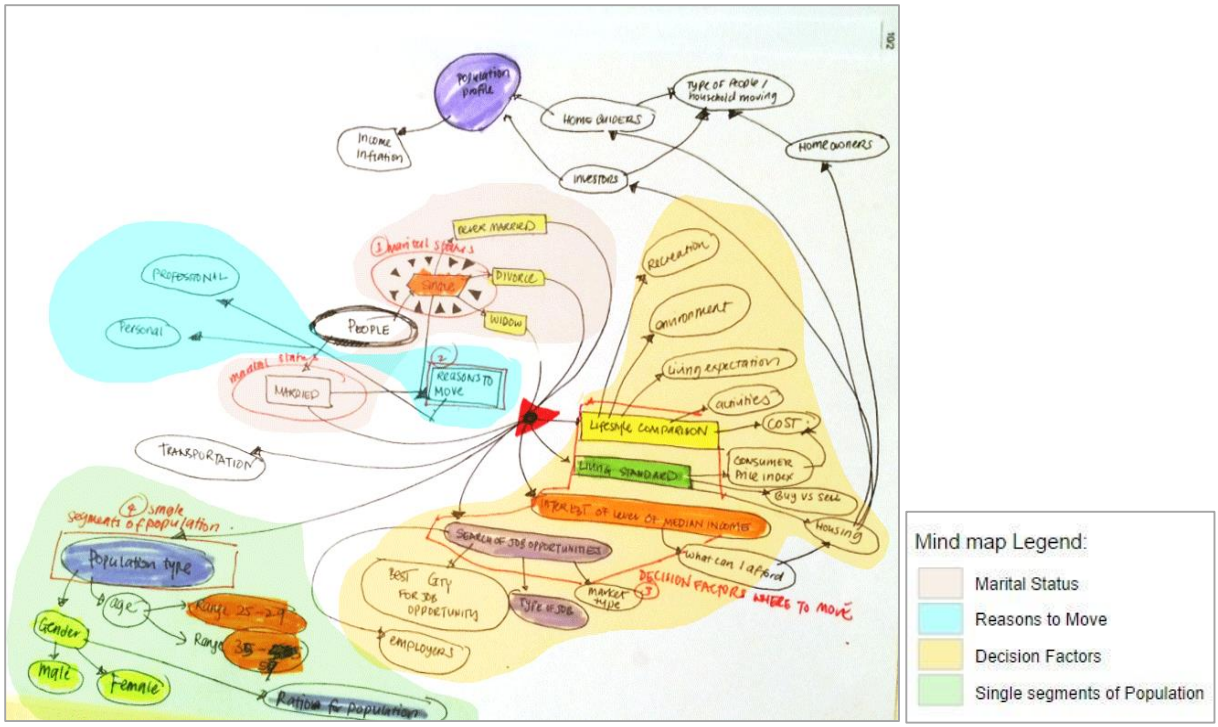

FIGURE 6 - MINDMAP

SKETCHING & PROTOTYPING

We started with the goal of enabling users to compare data on single demographics and employment. With this goal, we sketched different types of charts (Figure 7). Based on these sketches, we created charts in Tableau and then determined the most appropriate type of chart to use for the data.

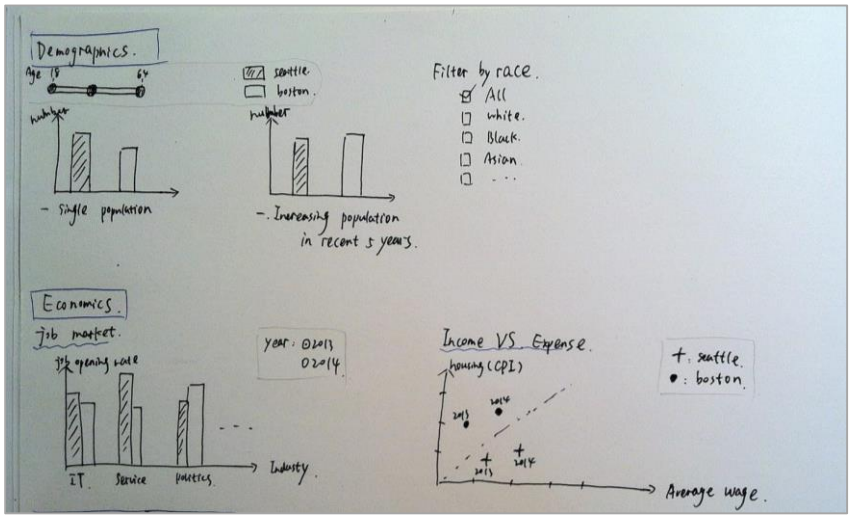

FIGURE 7 - INITIAL SKETCHES

After we incorporated weather data, we realized our data sources fell into three categories. Because our data was complex and we wanted to display numerous charts on a single page, we focused on designing a layout that combined all the charts together in a cohesive way. We considered different types of layouts and page navigation, such as creating tabs to separate categories of charts onto different pages, or putting all charts on one page. Figure 8 illustrates a mockup we made with Balsamiq application.

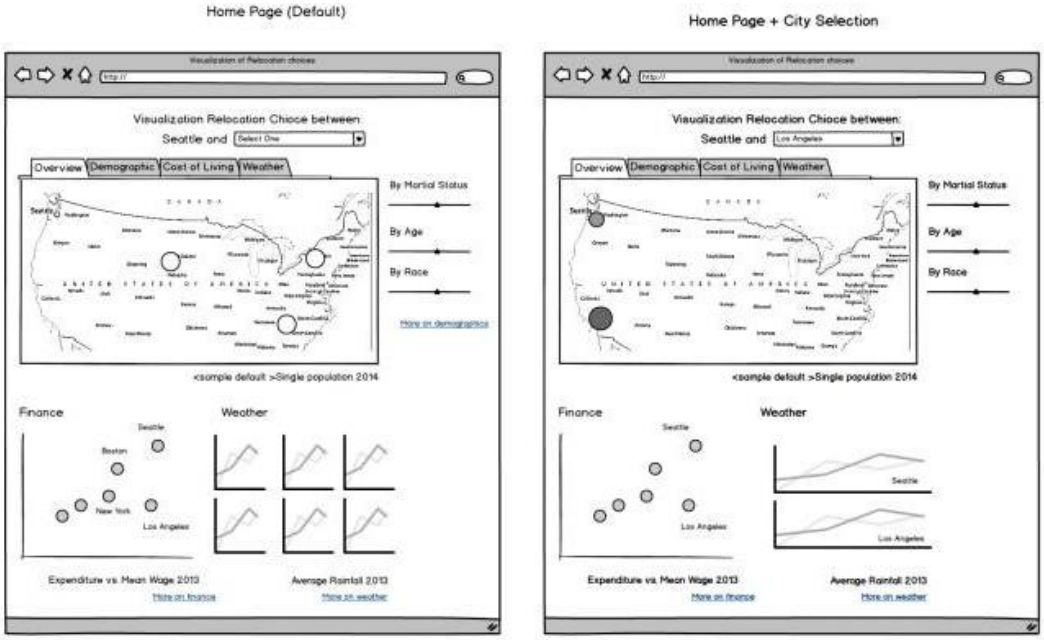

FIGURE 8 - PROTOTYPE IN BALSAMIQ

With these sketches, we created a prototype including all aspects of data we had in Tableau, which applied more interaction techniques. Then using this hi-fidelity prototype, we conducted a new rounds of usability testing, and further refined visualization.

USER RESEARCH & USABILITY TESTING

EXPLORATORY UESR RESEARCH

Our research question was: How do we present complex government data to help people visualizing relocation factors?

With this question, we wanted to figure out the main factors people considered when planning a relocation to determine the focus of our tool. We interviewed 7 participants who had relocated within the U.S. (See appendix for detailed interview questions). We also asked participants to rank 8 common relocation factors. The top five factors users cared the most about relocation were: Job Opportunities, Environment, Recreation, Lifestyle and Family-related.

Based on research results, we decided to add weather data into the visualization, which provides information on the natural environment for a city. Later rounds of usability studies

confirmed this, as two users in subsequent studies found the weather charts useful to them.

DEVELOPING PERSONAS

We were curious about the kind of information people research when they plan to relocate to another city. This led us to explore a variation of data sources related to facts about cities, and resulted in building a visualization based on that data. The visualization was targeted at users who were planning to relocate to another city, and wanted to find out facts about the city that would impact the decision to relocate. In addition, the visualization also provides interesting data to those researching facts about a city.

The characteristics below in Figure 9 and Figure 10 describe the typical users of our designed visualization tool.

Primary Persona

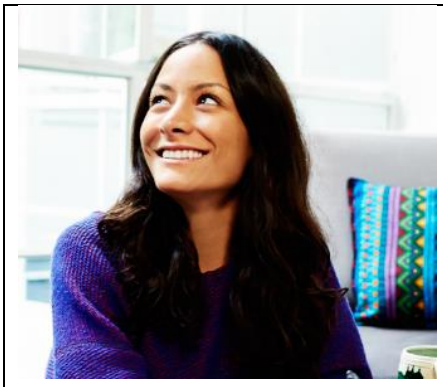	Jenny, a 43 year old married female marketing analyst.
	Scenario: Jenny has learned a lot about Seattle. She is attracted to the cityscape and enjoys outdoor activities. She is thinking of relocating to a city which offers exploration of outdoor activities and mount climbing.
	Objective: Jenny uses the visualization tool to explore general facts of Seattle's demographic, living condition, and weather. She also wants to see if she can afford living there by looking at the job market and the housing costs.

FIGURE 9 - PERSONA JENNY

Secondary Persona

	Bernard, a 30 year old single male software developer.
	Scenario: Bernard received a job offer to relocate to Seattle and is evaluating the pros and cons of a move. He is concerned about the environment, weather and lifestyle changes that may have a big impact to his life.
	Objective: Bernard uses the visualization tool to compare overall facts between his current city and Seattle. He is particularly curious on housing costs and demographics of both cities.

FIGURE 10 - PERSONA BERNARD

USER FLOW CHART

We created a user workflow chart that outlines the supported interactions in our visualization to inform design (Figure 11). We incorporated established information visualization design principles from Schneiderman, Tufte, Mackinlay and Bertin. The thought process included refining the flow, identify the actions, allow users input and interaction and understanding the process of data filtration.

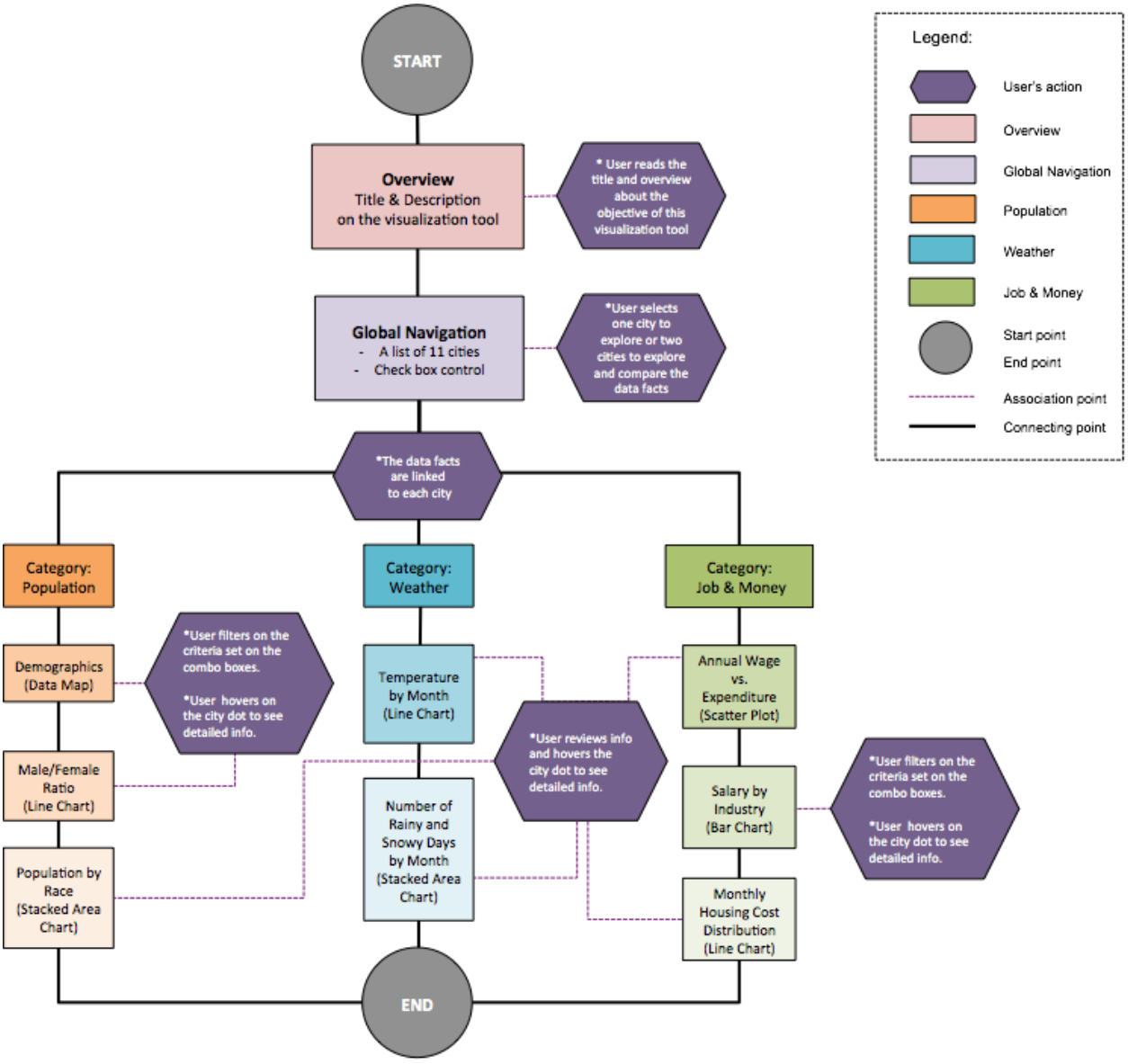

FIGURE 11 - USER FLOW CHART

INITIAL PROTOTYPE USABILITY TESTING

We created two interactive prototypes with different data and conducted the first round of usability testing to gather feedback to further guide our design. The first one (Figure 12) displayed the ratio of single women to single men aged 25 to 29 years old in different counties across U.S. The second visualization (Figure 13) displayed the total number of employment, the hourly mean wage and annual mean wage by different occupations in 11 major cities.

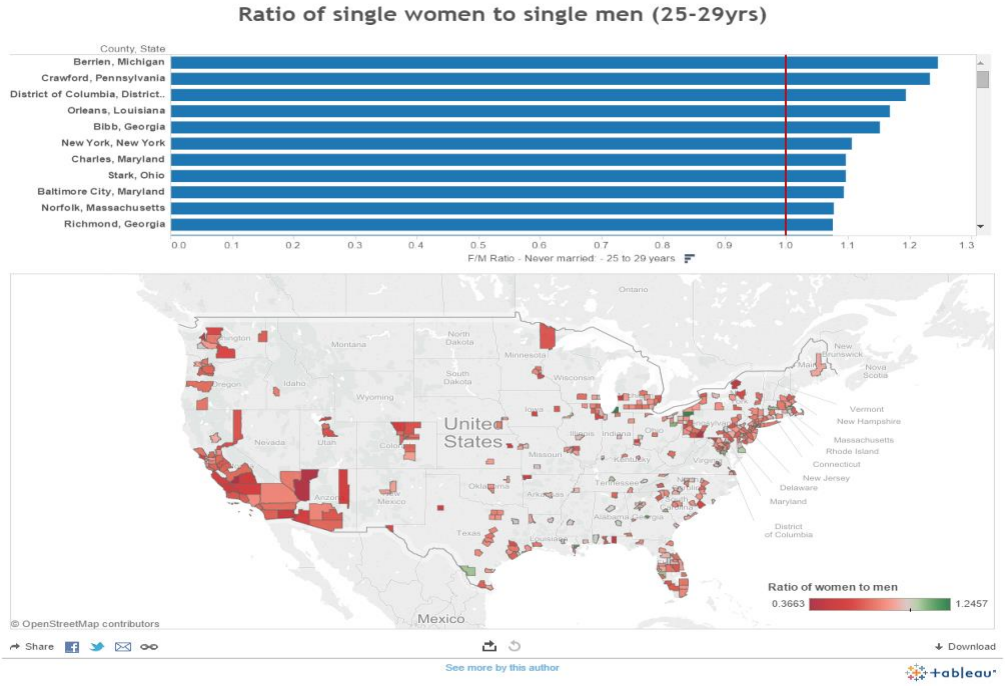

FIGURE 12 - INTERACTIVE PROTOTYPE - DEMOGRAPHICS

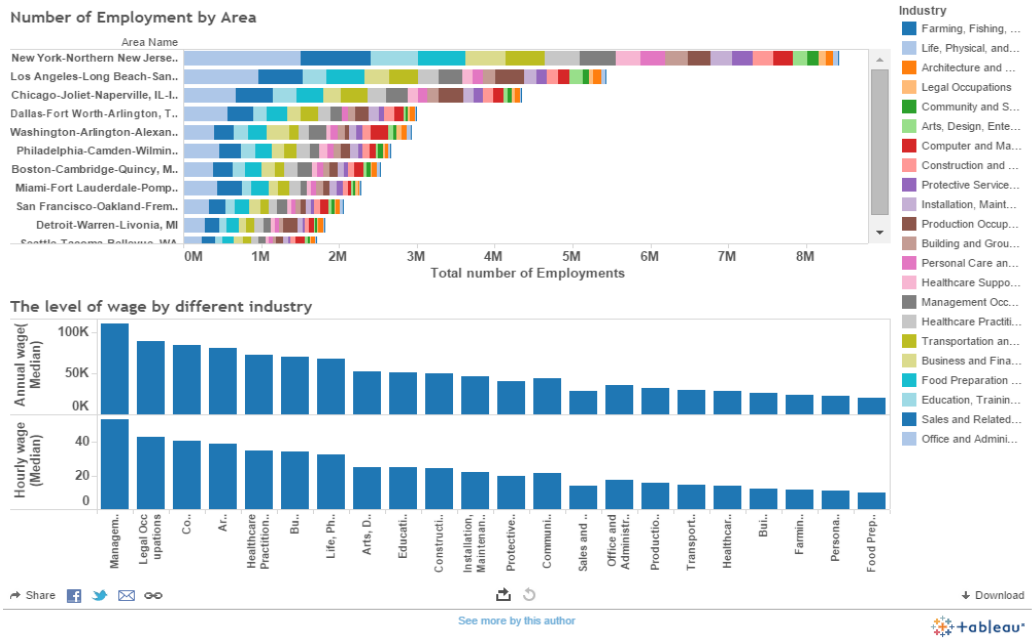

FIGURE 13 - INTERACTIVE PROTOTYPE - JOBS BY INDUSTRY

We ran the test with 2 participants, asking them to interact with the visualizations using the think-aloud protocol.

For the first visualization (Figure 12), one user continued to orient himself back to the map even after several failures. Both participants were frustrated after zooming-in on the map and then not knowing how to go back to original view. We also realized that we didn't use color-blind safe color palette.

For the second visualization (Figure 13), both users felt overwhelmed by the data and colors, and found it difficult to compare salary of two different occupations in the same city. They liked the brushing and linking functionality between different charts.

In general, major usability issues existed in the selection and navigation of the graphics. We didn't make it clear which parts of the visualization was interactive, so users struggled in selecting content they were interested in, and struggled more in clearing their selections and filters. Also, most of the chart and axis labels were missing or not clear enough for the user to understand.

HI-FIDELITY PROTOTYPE USABILITY TESTING

After we created hi-fidelity prototype that combined all graphics we had into one page, we conducted another round of both exploratory and task-based usability testing. The purpose of the testing was to review how a user performs the task effectively and efficiently. Besides, we also observed the potential issues that prevent users from accomplishing the tasks.

Target Users: We targeted participants who had relocation experience and users who were planning to relocate. Users could be anyone who was looking for information on relocating to

another city or interested in learning about their cities. Besides, we were also looking for users to share their perspective of what they have learned during their relocation experience.

Participants: We had a total of 4 participants (3 female, 1 male) with diverse professional background who have experience with relocation.

User Scenario: The usability study was conducted at participants' location: office and home environment. Each participant was given an overview of the usability testing, detailed instruction of the test and a demo to the actual visualization tool. We encouraged users to 'think out loud' during the testing walkthrough and ask questions at any point of time. We facilitated a series of exploratory and task-based scenario (see references Usability Testing: Test Plan).

RESULTS AND INSIGHTS

All 4 participants struggled to understand the structure of the visualization tool. At first, the visualization tool looks easy and straightforward. But they realized they didn't know how to begin or navigate in the visualization and needed guidance navigating through the visualization. There was a lack of structure and too much information in the layout. 2 participants commented that the interface felt congested.

The participants found the visualization tool difficult to use and understand. They didn't find the method for filtering by city and found it challenging to understand the data on the chart. By having the map as the first visualization, participants thought that was the method for filtering data to a city. Unfortunately, brushing and linking using the map was not implemented yet.

One participant struggled with zoom interaction in the population demographics map. Text labels on the weather rainfall chart was not legible. They were confused about the meaning of the different colors on the chart. Overall, participants felt overwhelmed with the structure of the layout due to charts without explanations that lacked detailed explanations.

DESIGN ITERATION

The results from our usability studies led us to iterate the design for the visualization tool based on 4 areas: design layout, color usage, labels and content, navigation.

Design Layout: We started with a flat, non-hierarchical layout on the main page of our first prototype (Figure 14). We improved the structure of the charts to show hierarchy and categorize the charts based on content (Figure 15). With this change, we were able to show a clear visualization for both exploration data and data comparison.

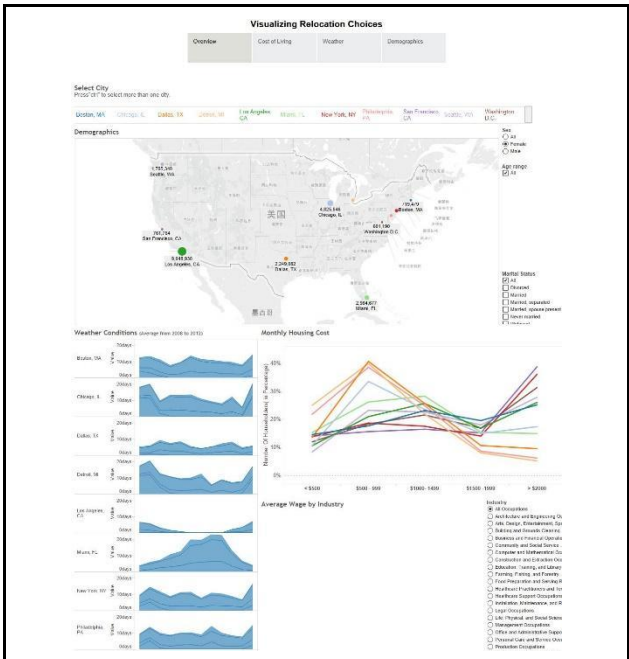

FIGURE 14 - DESIGN ITERATION/DESIGN LAYOUT - (BEFORE)

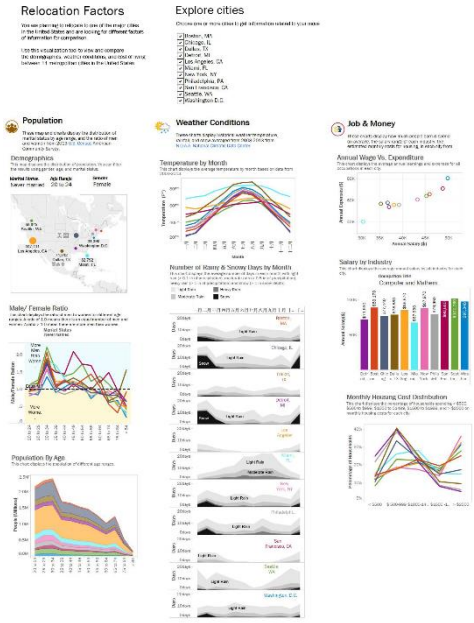

FIGURE 15 - DESIGN ITERATION/DESIGN LAYOUT - (AFTER)

Color Usage: Color is crucial for information visualization. The feedback we received from the usability studies show the colors used in the initial prototype was poor and the contrast level for readability was not sufficient (Figure 16). We took the effort to unify the colors used across charts. In addition, we also looked into how to compensate for the color blindness when selecting colors and improve contrast ratios for better legibility. We explored the city view, culture, and natural environment to improve the color palette used. We decided to use colors to associate a semantic meaning for each represented city (Figure 17).

FIGURE 16 - DESIGN ITERATION/COLOR USAGE (BEFORE)

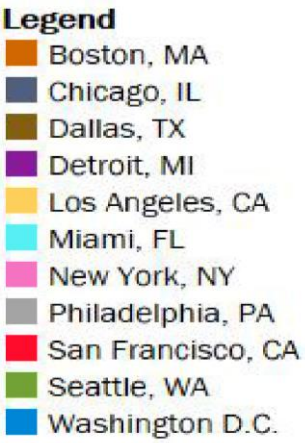

FIGURE 17 - DESIGN ITERATION/COLOR USAGE (AFTER)

Labels and Descriptions: Labels and descriptions were the main instructions provided to users. The initial prototype didn't provide any descriptions of the charts (Figure 18). We updated our prototype by introducing chart descriptions and labels to improve understandability (Figure 19). The objective was to provide a summary of how to interact with the data using the provided controls. The detailed instructions allowed users to navigate and explore the overview layout as well as get details-on-demand.

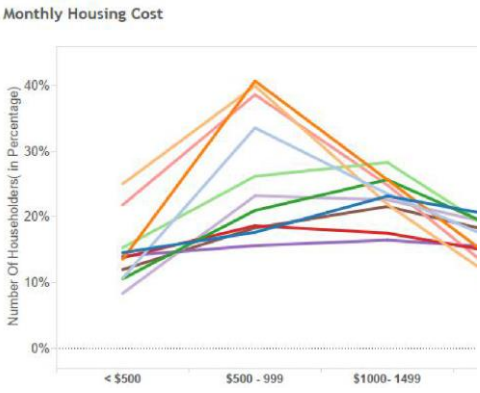

FIGURE 18 - DESIGN ITERATION/LABEL & DESCRIPTION (BEFORE)

FIGURE 19 - DESIGN ITERATION/LABEL & DESCRIPTION (AFTER)

Navigation: The initial prototype displayed charts across multiple pages (Figure 20), and participants found it difficult to relate information presented across different pages. From the feedback, we understood the value of showing all the charts on a single page. We took a step back to re-evaluate the structure of the page layout and identified the basic navigation pattern. We came up with an overview with instructions at the top, global navigation controls and the diversified and detailed city facts in three different categories (Figure 21). We also focused on

solving technical challenges associated with merging data from different sources together first.

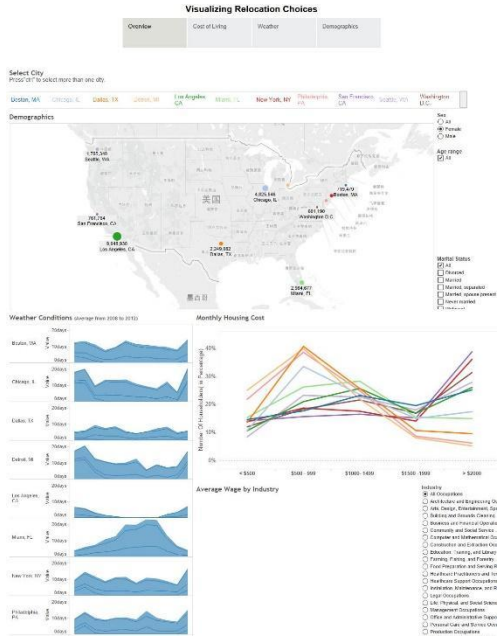

FIGURE 20 - DESIGN ITERATION/NAVIGATION (BEFORE)

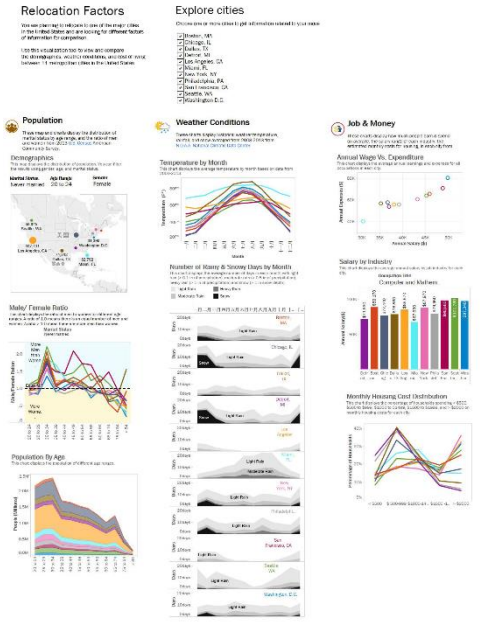

FIGURE 21 - DESIGN ITERATION/NAVIGATION (AFTER)

TOOL DESIGN

In order to support discovery of insights from data in making relocation decisions, the visualization included 8 distinct charts displaying different aspects of demographics, climate, and economy. The dashboard organization enabled users to view and manipulate different aspects at the same time and instantly identify relationships.

To improve understandability, the 8 distinct charts were arranged into 3 columns. The first column included 3 population charts displaying population count demographics data by marital status, sex, age, and race. The second column displayed historical weather and provided average temperature, rainfall, and snowfall data. The third column displayed economic data that provides information on salary and expenses by city.

Users can interact with the visualization to zoom in on items of interest and use filters to highlight or only display a specific portion of data (e.g. data from specific cities), enabling users to focus on a particular area of interest and ignore unwanted data. When details for a particular item of data are desired, hovering over the item of data with the mouse cursor displays an overlay with numerical details.

Besides an overview description of the whole visualization and descriptions of each category of data displayed in a column, every chart included a detailed explanation to help users quickly understand what data is displayed in each chart at a glance.

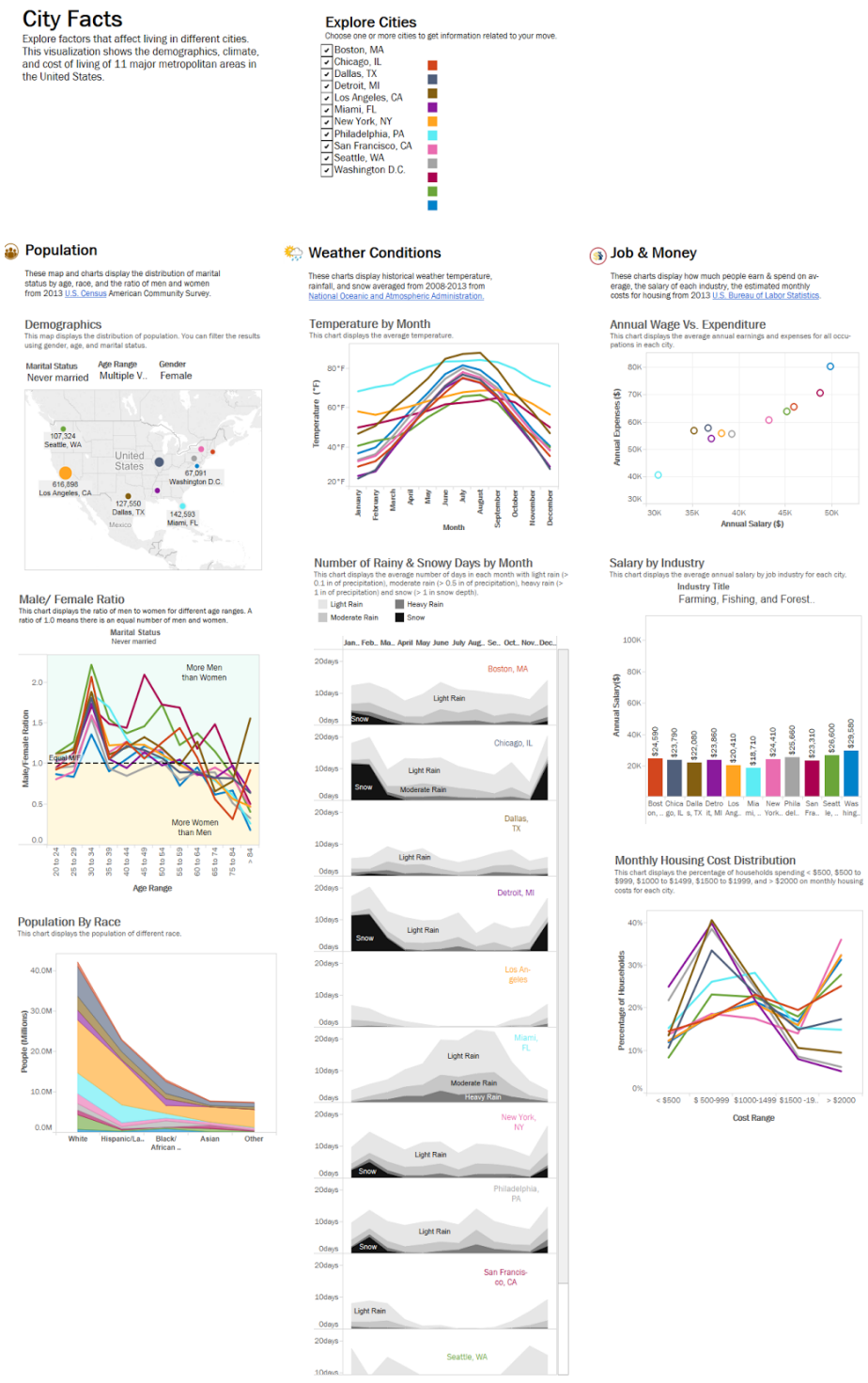

FIGURE 22 - OVERVIEW OF ALL DATA ON A CONTINUOUS PAGE

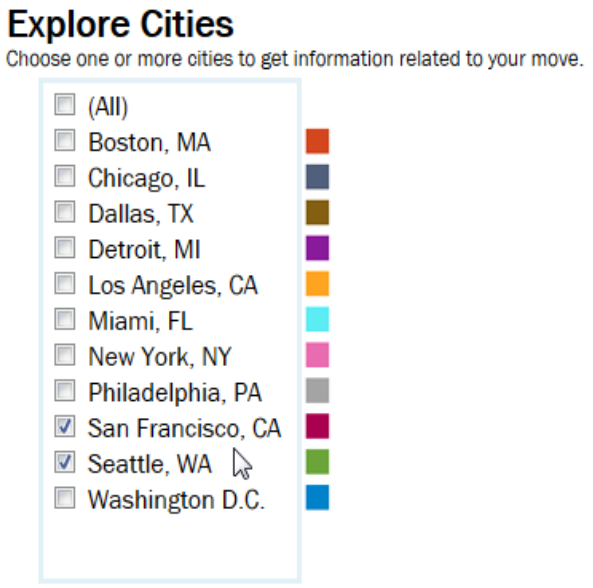

FIGURE 23 - FILTERING TO SELECT A CITY

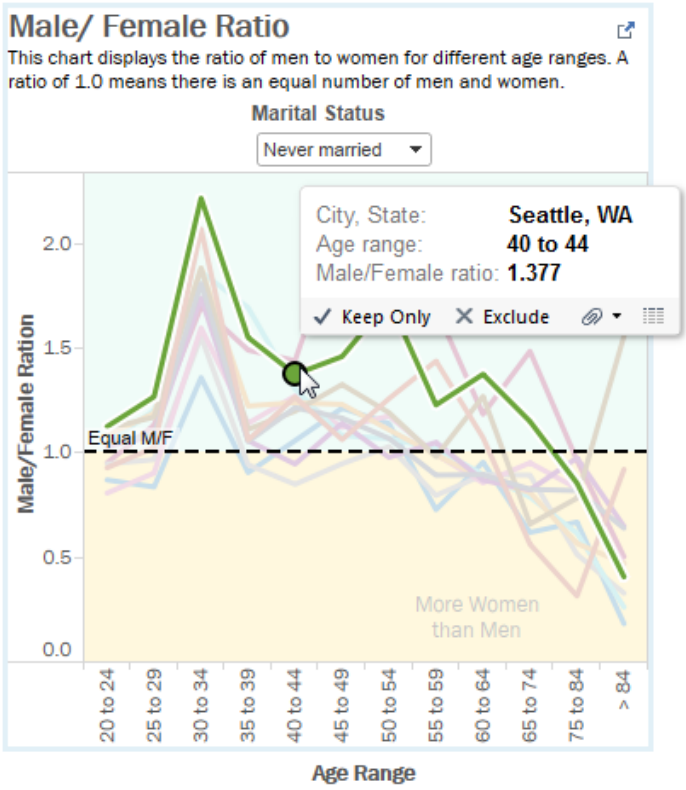

FIGURE 24 - OVERLAY WITH DATA DETAILS

In addition, the visualization supported exploring and understanding the relationship between different cities. Filters for selecting a city applied to every chart in the visualization; brushing and linking technique is used to highlight relationships between different cities while preserving the

data for all the cities. Selecting a city in any chart brushes all the charts: data for the selected city to be highlighted in all charts. Cities were universally encoded using hue (color), as the visualization includes 11 cities that can be easily differentiated using hue.

DEMOGRAPHICS BY SEX, MARITAL STATUS, AGE

The properties of visuals were matched with the underlying properties of the data. Demographic data on the count of people by gender, marital status, and age range was displayed using circle symbols on a map of the United States (Figure 25). A map was used in this visualization for the ease of perceptual recognition of the geographic nature of the data. The map included zoom functionality, enabling users to control and view a subset of data from a particular geographic area. Since 2-D position spatial attributes were used for encoding geographic position, the next most precise attribute, area, was used to encode the number of people. Selection of filters was enabled through dropdown selection control that supports simultaneous filtering of single or multiple items (Figure 26).

The government data provided information on population counts of marital status by age and marital status by race, but not by marital status by age AND race. It wasn't possible to restrict Tableau to hide selection combinations where data was not available.

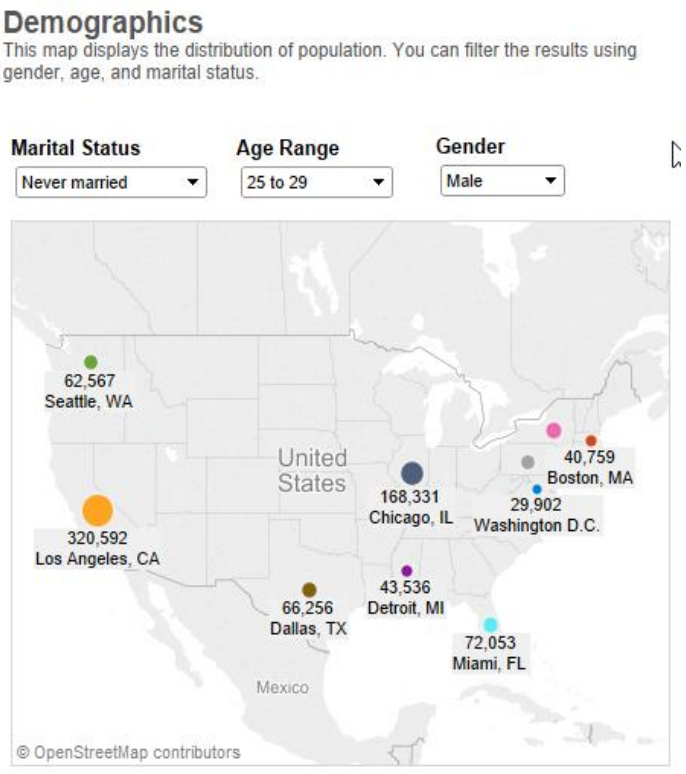

FIGURE 25 - DEMOGRAPHICS

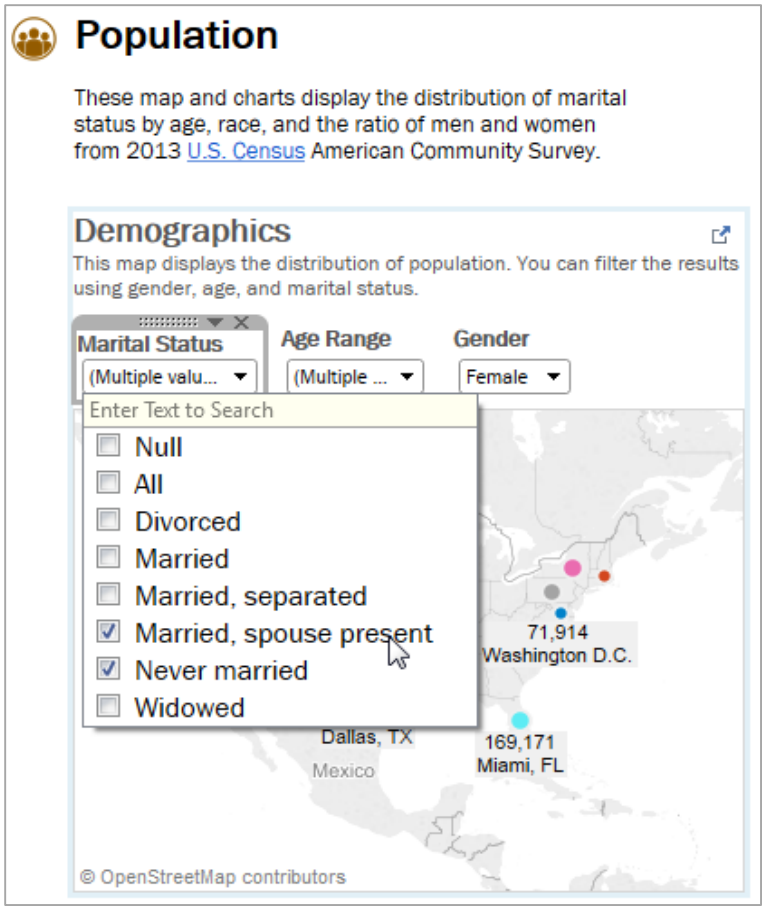

FIGURE 26 - SIMULTANEOUS FILTERING BY SINGLE OR MULTIPLE ITEMS

Visual Encoding Attributes

Data Type	Data	Visual Attributes
Nominal	Cities (geographic location)	2-D Position, Hue (color)
Quantitative	Count of # of people	Size

MALE/FEMALE RATIO

The ratio of count of men to women by marital status and age was visualized as a time series line chart (Figure 27). The quantitative value of the male/female ratio was visualized as vertical position on y-axis, while age range was visualized as position on x-axis. Male/female ratio was visualized using the most accurately perceived visual quality of spatial position. A reference value line at male/female ratio = 1.0, with a different hue (color) and text labels, were added for improved legibility at understanding whether a city had a larger population of men or women.

The time series line chart visualization used did not accommodate displaying a 4th nominal variable, so filtering data by marital status is enabled through usage of a dropdown selection control. Cities are encoded using the same hue for consistency.

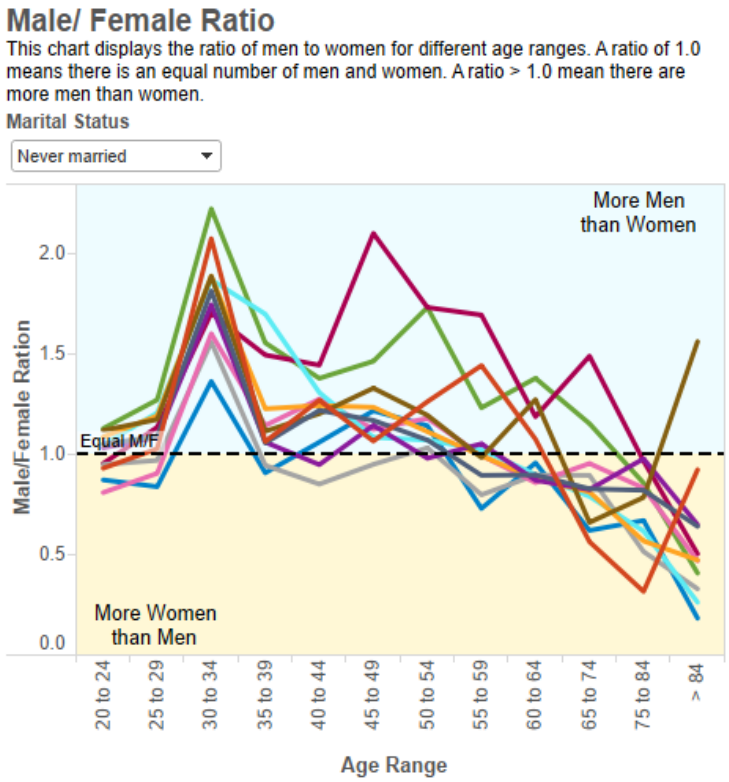

FIGURE 27 - MALE/FEMALE RATIO

Visual Encoding Attributes

Data Type	Data	Visual Attributes
Nominal	Cities (geographic location)	Hue (color)
Quantitative	Male/Female ratio	position (on y axis)

Ordinal	Age range	position (on x axis)

POPULATION BY RACE

The number of people in each metropolitan city was visualized using a stacked area chart (Figure 28). Spatial position was used to encode for population count and race. Combined with encoding the city using hue (color), the total population of all races is perceived by area.

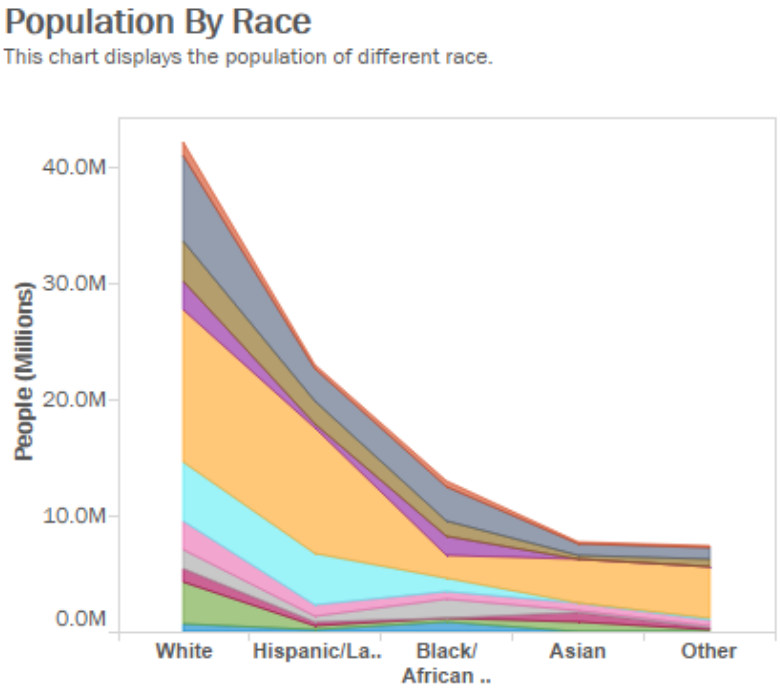

FIGURE 28 - POPULATION BY RACE

Visual Encoding Attributes

Data Type	**Data**	**Visual Attributes**
Quantitative	Population count	Position (on y axis)
Nominal	Race	Position (on x axis)
Nominal	Cities (geographic location)	Hue (color)

TEMPERATURE BY MONTH

Average monthly temperature was visualized using a time series line chart to match spatial position with the most important variable, temperature (Figure 29). The chart enables users to see how average temperature varies over the course of a year, and enables users to quickly compare how temperature varies in different months in a city and also how temperature varies in different cities in the same month. Cities are encoded using the same hue for consistency.

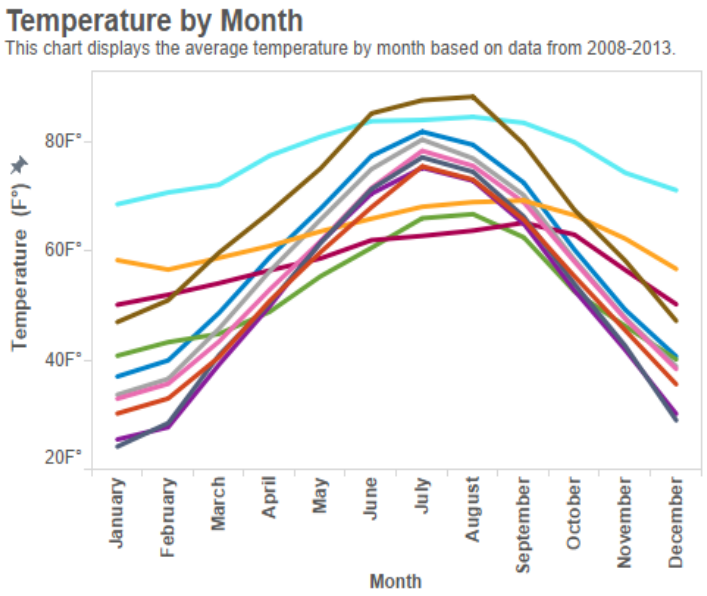

FIGURE 29 - TEMPERATURE BY MONTH

Visual Encoding Attributes

Data Type	Data	Visual Attributes
Quantitative	Temperature	position (on y axis)
Ordinal	Month	position (on x axis)
Nominal	Cities (geographic location)	Hue (color)

NUMBER OF RAINY AND SNOWY DAYS BY MONTH

A novel way of representing climate of each city was realized using small multiples of area charts (Figure 30). Each stacked area chart displayed the average number of days with precipitation in each month, creating a chart that visualizes the average annual precipitation profile for each city. Users can identify the type of precipitation a city experiences, and what time of year the precipitation occurs. The small multiples layout enabled easy side-by-side

comparison of different cities while simultaneously displaying 3 different variables of precipitation intensity, month, and number of days in each month.

To emphasize small differences in the most important variable of rainfall and snowfall between cities, spatial position (height) was used to encode the quantitative data on the number of days per month for each category of rainfall and snowfall. In addition, value (lightness) was selected to encode for the ordinal values of intensity of precipitation. Hue (color) was not used in the visualization to encode city to minimize ambiguity due to use of value (lightness). Hue (color) was subtly used in the text label for each city.

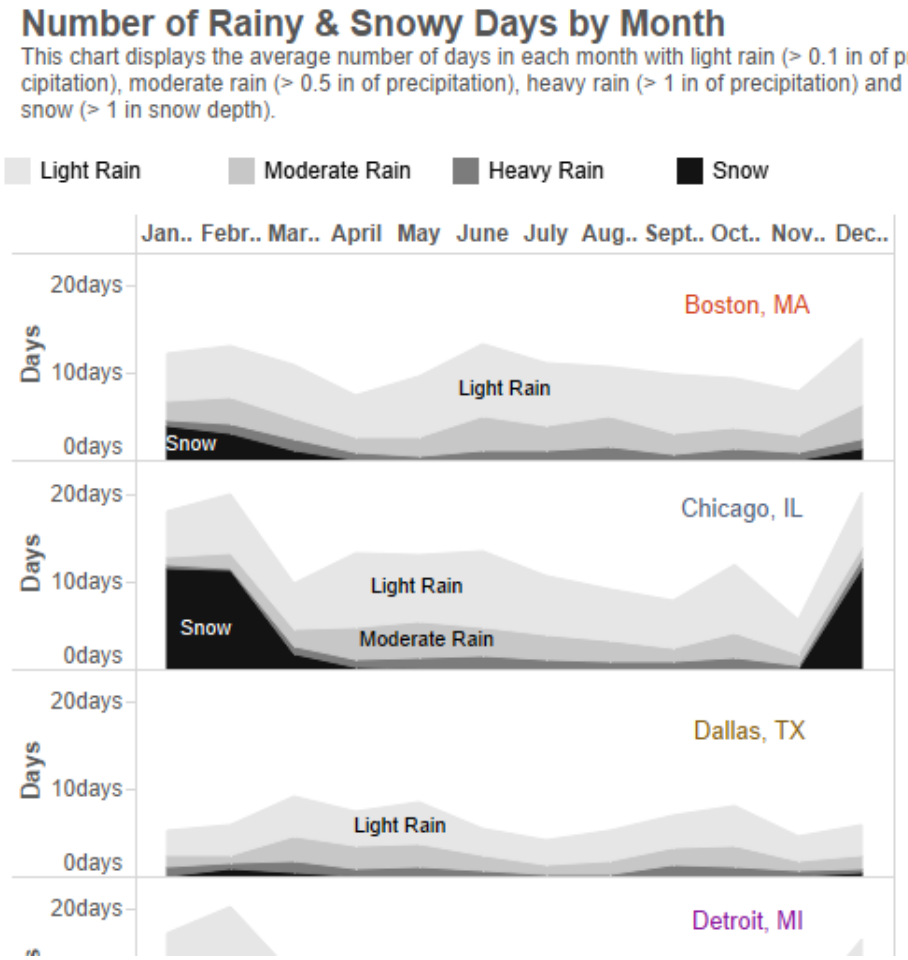

FIGURE 30 - NUMBER OF RAINY & SNOWY DAYS

Visual Encoding Attributes

Data Type	Data	Visual Attributes
Quantitative	Count of days	position (on y axis)

Ordinal	Month	position (on x axis)
Ordinal	Precipitation intensity	Value (lightness)
Nominal	Cities (geographic location)	spatial position, color of text label

ANNUAL AVERAGE WAGE VS. EXPENDITURE

Bivariate quantitative data on average annual wages and expenses per city were visualized using a scatterplot (Figure 31). The chart enables users to simultaneously compare average annual wages and average annual expenses between different cities, and also comparing wages and expenses for a particular city. For consistency with all other charts, cities is represented using hue (color) with the same color palette.

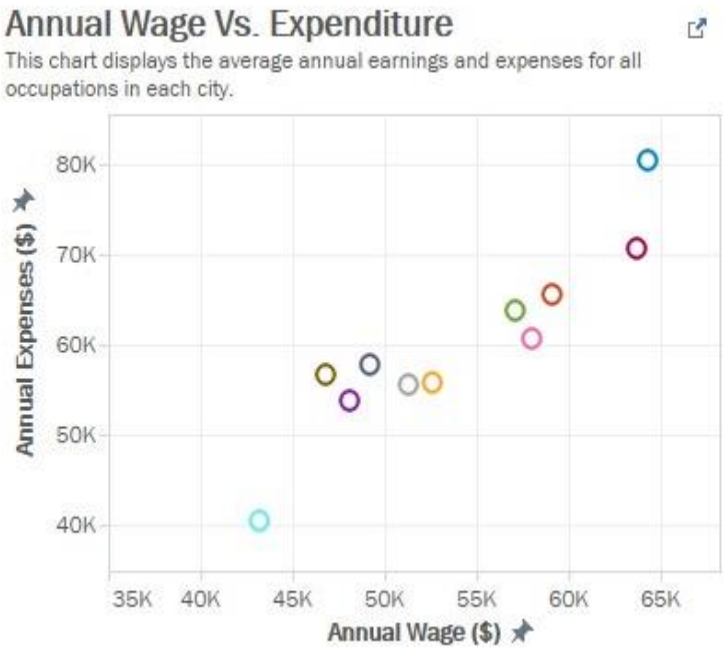

FIGURE 31 - ANNUAL WAGES VS. EXPENSES

Visual Encoding Attributes

Data Type	Data	Visual Attributes
Quantitative	Annual expenses (in $)	position (on y axis)

Quantitative	Annual salary (in $)	position (on x axis)
Nominal	Cities (geographic location)	Hue (color)

SALARY BY INDUSTRY

Quantitative data on annual salary by city was displayed using a bar chart (Figure 32). Combined with the filter to select from a list of 22 industries, the data provides a comparison of how annual salary in a specific industry differs between cities. The scale for the bar chart started at $0, ensuring differences in height of bars are proportional to the differences in salary. City is encoded using position and hue (color), with consistent ordering alphabetically and color palette shared across other charts.

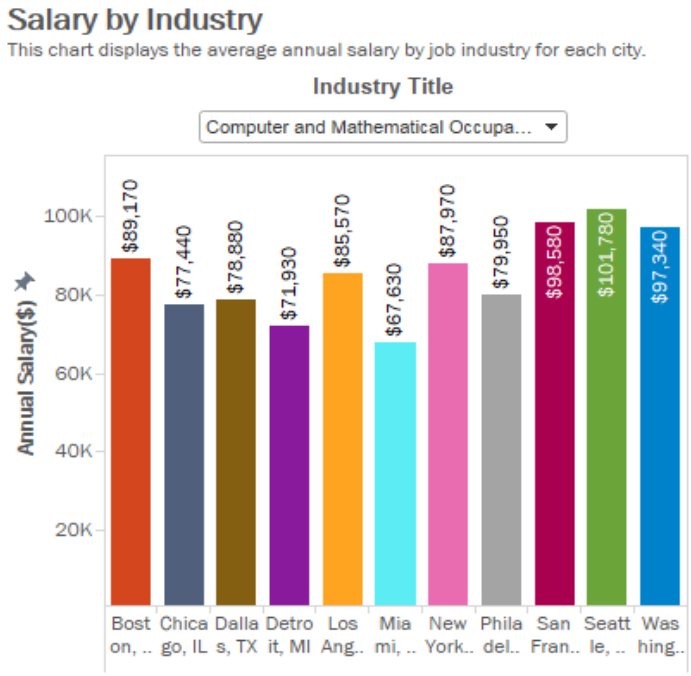

FIGURE 32 - SALARY BY INDUSTRY

Visual Encoding Attributes

Data Type	Data	Visual Attributes
Quantitative	Annual salary (in $)	position (on y axis)
Nominal	Cities (geographic location)	spatial position on x axis, hue (color)

MONTHLY HOUSING COST DISTRIBUTION

Quantitative data on housing costs was displayed using a line chart (Figure 33). The chart provided a comparison of the distribution of spending on housing cost within a specific city, and a comparison between different cities of the percentage of households with housing costs that fall within a particular range. The most important variables, percentage of households and housing cost range, were displayed using position, the most accurately perceived visual attribute.

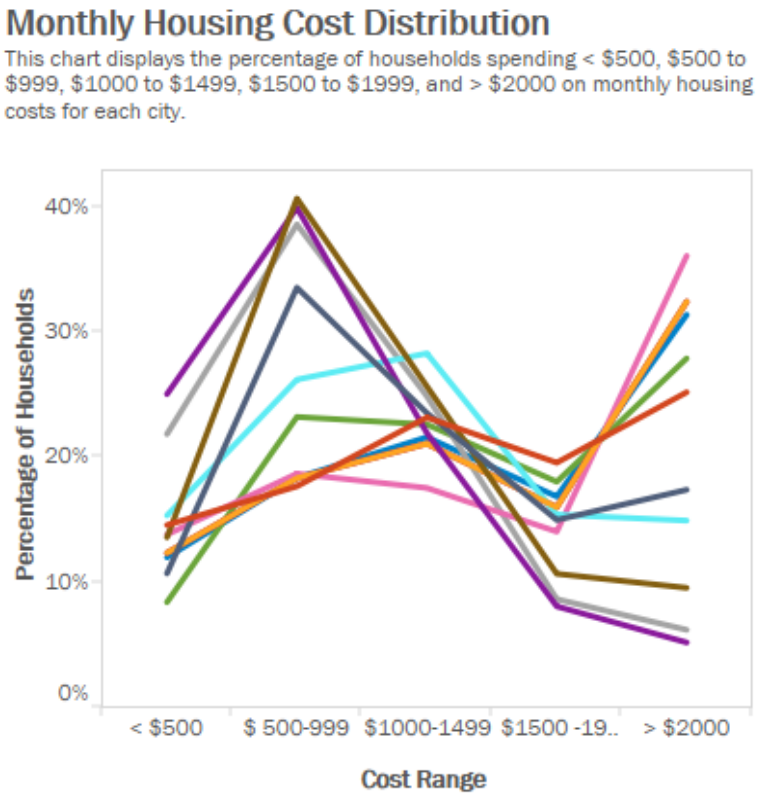

FIGURE 33 - MONTHLY HOUSING COST DISTRIBUTION

Visual Encoding Attributes

Data Type	Data	Visual Attributes
Quantitative	Percentage of households (%)	position (on y axis)
Quantitative	Monthly housing costs (in $)	position (on x axis)
Nominal	Cities (geographic location)	Hue (color)

EVALUATION & DISCUSSION

In general, although we didn't have chance to conduct usability test with our final visualization, we got some positive comments from the previous testing, especially on the content and goal of this tool. Two users noted that they appreciated that we provided a lot of official and trustworthy information from government sources. Some of them are interesting such as single demographics; others are useful such as weather conditions. One user spent much time digging in different charts and seemed interested in data.

On the other hand, based on their negative feedback, we have made redesign changes accordingly towards major usability issues. Finally, we reviewed all of our design decisions, and made sure that we have tried our best to follow the principles. Our final visualizations were designed and optimized to be visually and aesthetically appealing, immediately understandable, and enables analysis of new insights without requiring excessive cognitive effort.

VISUAL INFORMATION SEEKING PRINCIPLES

We designed the visualization based on Schneiderman's visual information seeking principles of presenting an overview first, then providing zoom and filtering to a specific set of cities, then providing details on demand, and then viewing relationships among items.

Users are initially presented with a single continuous page with an overview of the entire collection, with zoomed out views with data from the entire data set of 11 major metropolitan areas in United States. Filters for selecting a city apply to every chart in the visualization; brushing and linking technique is used to highlight relationships between different cities while preserving the data for all the cities. The map displaying demographic data includes zoom functionality, enabling users to control and view a subset of data from a particular geographic area. Selecting a city or multiple cities in any chart brushes all the charts; data for the selected city or cities is highlighted in all charts.

GRAPHICAL EXPRESSIVENESS

We followed Mackinlay principles on effectiveness of human perception, attempting to encode the most important information in the most effective way so that data is more readily perceived. For example, cities were universally encoded using hue (color), as the visualization includes a small set of nominal values for cities. A map was used in the visualization of demographics for the ease of perceptual recognition of the geographic nature of the data.

By utilizing perceptual properties for quantitative/ordinal/nominal data described by Mackinlay, we targeted expressiveness and ensured all the facts were presented, and only the facts were presented. Below are examples of how expressiveness principles were applied in creating visualizations.

- In the population demographics chart, since 2-D position spatial attributes is already used for encoding geographic position on a map, the next most precise attribute, size/area, was used to encode the number of people. A choropleth visualization was also considered, but not used because visual perception of hue is not as precise as size,

and representing each city with a circle symbol enabled reinforcing the encoding of city using hue (color).

- In the population male/female ratio chart, the most relevant and important data, male/female ratio, is visualized using the most accurately perceived visual quality of spatial position.
- In the average annual wages and expenses chart, bivariate quantitative data was visualized using a scatterplot to maximize expressiveness by using the most accurately perceived attribute, spatial position.

GRAPHICAL EXCELLENCE

We followed Tufte's principles of graphical excellence, maximizing the "data to ink" ratio and displaying the greatest number of ideas in the shortest time, with least ink, in smallest space.

We chose to display 8 different charts on a single page instead of using multiple pages in order to enable identification of patterns and trends.

In the rainfall and snowfall stacked chart, 1 quantitative, 2 ordinal, and 1 nominal data types were visualized in a single chart. Spatial position was used to encode the quantitative data on the number of days per month for each category of rainfall and snowfall and the nominal data on the month of the year. Value (different shades of grey) was selected to encode for the ordinal values of intensity of precipitation. Hue (color) was subtly used in the text label for each city.

GRAPHICAL INTEGRITY

We also followed Tufte's principles of graphical integrity. The scales in charts represented numbers proportionally. For example, in the salary by industry and housing costs charts, scales of the vertical axis started at 0 to ensure differences in height of bars are proportional to differences in underlying numerical value. The intentional re-use of hue (color) to always encode city minimizes design variations eases comprehension. The text in the overview description of the whole visualization, descriptions of the data displayed in each column, and the detailed explanation of each chart provide context for the data, and were reviewed in multiple rounds of user testing to optimize for quick and accurate comprehension.

FUTURE WORK

Advanced Interactivity: One future direction of this tool is to create more advanced interactivity, so that the user could have a smaller view to start with. In our current tool, in order to put all content in one view, we have been limited in the global filter of city.

We can use features of 'Overview + Detail' or custom 'Semantic Zoom' to display more data for a city based on user interaction. Or using a fisheye distortion to magnify and provide additional detail would better support Schneiderman's visual information seeking mantra. We can also show multivariate variables on the map, so that the user can choose which aspect they can view on the map. Also the below charts would change.

Display related pictures of city: We think importing the city pictures and landscape related to the city would be interesting. In the figure 34, we can let the user start with a map selection. Once they select a city on the map, they would go to a subpage (see figure35), with different aspects of information about that city. We can also include city images or connect with google map data source.

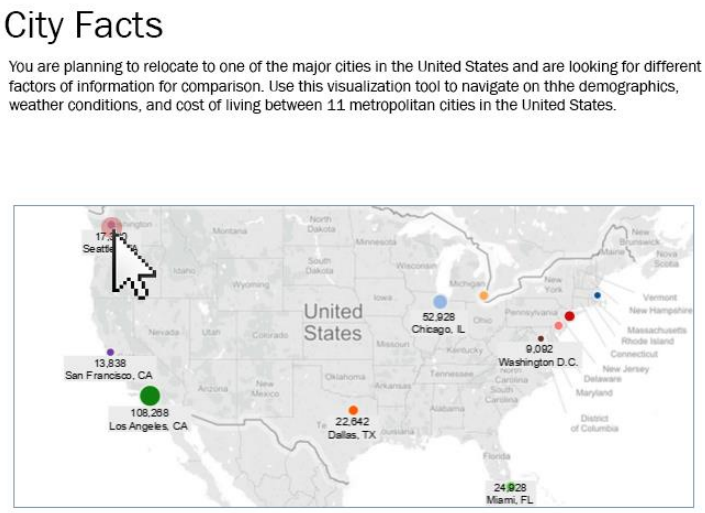

FIGURE 34 - FUTURE WORK - MAP AS NAVIGATION

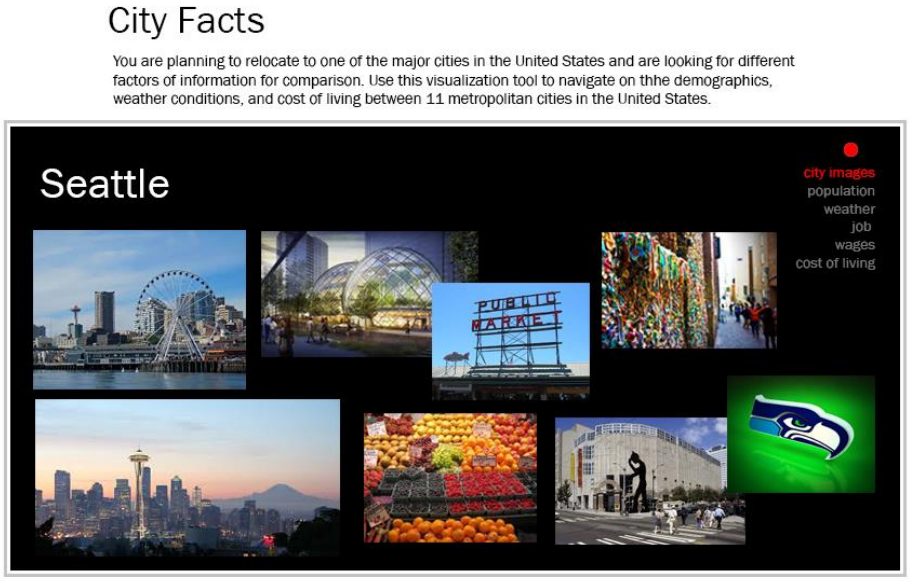

FIGURE 35 - FUTURE WORK - SEMANTIC ZOOM & GOOGLE MAP DATA

Fix Tableau's limitation of data filters: We cannot customize the quick filter to decide which items to show easily in tableau. After we joined and combined more than 5 of our data tables into Tableau, the null data showed up in some quick filters (e.g, the demographics map) because the unmatched data format of primary and secondary data sources. We will need to

figure out advanced function if any to realize this in Tableau, or transfer to other tools such as d3.

Historical data about Job & Money: It should be interesting to look into the historical trend about annual wage/expenditure as a time-series plot, or introduce another dimension of data on the scatterplot to pass more information to the user. We could not find this data for all the cities in the BLS website.

CONCLUSION

We realized that making a relocation decision involves considering many more factors than the ones we included within the scope of this project. Other data such as employment rates, property prices, crime rates, and neighborhood data are other important factors per findings from our user research. We believe that the creation of this visualization tool laid a foundation for future possibilities and help us realize the depth and complexity of this subject. In the process, we learned that creating an effective visualization requires having reliable data sources, a good understanding of the data, and a substantial amount of data cleaning and manipulation work. It is also essential to have a good handle of the tools and to understand the limitations. In the end, we think that we created a tool that balances between user needs, data availability, and resources limitations.

REFERENCES

National Oceanic and Atmospheric Administration, national climatic data center, Annual Climatological Summary, 2008-2013. http://www.ncdc.noaa.gov/cdo-web/

Census Bureau, American Community Survey(1-Year Estimated), 2012. http://old.socialexplorer.com/pub/reportdata/MetaBrowser.aspx?survey=ACS2012&header=True

Bureau of Labor Statistics, BLS regional Information Offices. Economic Summaries, 2014. http://www.bls.gov/regions/economic-summaries.htm

Bureau of Labor Statistics, Metropolitan and Nonmetropolitan Area Occupational Employment and Wage Estimates, May 2013. http://www.bls.gov/oes/current/oessrcma.htm

Tableau Desktop online Help System. Tableau Software. http://onlinehelp.tableausoftware.com/current/pro/online/en-us/help.htm#default.html

Edward Tufte. Visual Display of Quantitative Information, 1993.

Ben Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proc. IEEE Conference on Visual Languages, Boulder 1996.

Maureen Stone, StoneSoup Consulting. Expert Color Choices for presenting Data, 2006.

APPENDIX

Data Tables from ACS

For data of population, we used the following tables from ACS 2012,

(The exact number of counties included differs by each race.) The margin of error based on a 90% confidence level is included for each estimate. For our project, we used the following datasets from 2013:

- Sex By Marital Status By Age For The Population 15 Years And Over (B12002)
- Sex By Marital Status By Age For The Population 15 Years And Over By Race (B12002A, B12002B, B12002C, B12002D, B12002E, B12002F, B12002G, B12002H, B12002I)

User Research: Survey and Questionnaire

Research Questions

Have you ever relocated to another city in the U.S.?

What was the main reason *(if you don't mind sharing)* ? Personal / Professional

What factors did you consider relocating? (Please rank)

- Job opportunities
- Cost of living
- Family related
- Recreation
- Real Estate
- Transportation
- Lifestyle
- Environment
- Other

Where did you look for relocation information (resources, websites, etc.)? Were they helpful? Why?

If you have to relocate again, what kind of resources would help you?

Usability Testing: Test Plan

Introduction Script:

Thank you for taking the time to meet with me today. We will be exploring a data visualization tool for my group final project, called Visualizing Relocation Factors. The intent of this usability test is to evaluate users' exploration on obtaining the interactive data of city facts. There are no judgment and no right or wrong answers to any of these following questions. We are looking to improve our visualization tooling and your feedback is valuable to us. Thank you for your participation.

Research Questions:

- Content

- Do the users have a good understanding of the potential change or difference if they were to relocate?
- Is the information sufficient?
- What is missing / confusing?
- Are the text & graphics meaningful to them?

UI:

- Is the tool easy to use?
- Can they navigate well and can locate the information they need?
- Is the content layout logical (e.g. the section arrangement) to them?
- Do the labels and legends make sense?
- Are the checkboxes, sliders, etc. clear?
- Emotion/Attitude
- Do they like the tool?
- Is it effective?
- Does this tool encourage them to explore or analyze the data?
- Does the tool allow you to obtain detail information related to relocation?

Scenario 1

Imagine you live in Boston and you just got a job offer in Seattle. You want to compare the two cities to find out whether relocation is for you.

User Tasks for Scenario 1:

(Overview) On this page, what do you know about the two cities?
Observation: Overall interest in different elements
Answer: N/A. Open-ended

How many singles are in the age range of 30-34 for the two cities?
Answer: Boston: 60,437; Seattle: 90,757
Verify: If user understands the label “Never married” = Singles
Observation: Use of slider, checkboxes, labels and the map

(Cost of living) Let’s look at some potential changes in your cost of living.
How much does someone who work in the computer industry make on average in both cities?
Answer: Boston: 89,170; Seattle: 101,780;
Observation: Use of checkboxes & labels, colors and interactivity

How many percentages of people who pays around $1000-1499 housing cost per month?
Answer: Boston: 23%; Seattle: 23%
Observation: Interactivity - hover over, line graph

Can you compare the two cities side by side?

Observation: Discoverability of functionality (e.g. Ctrl+click)

Scenario 2

You have a friend from in Miami and have been telling you she has been hearing much about the outdoor lifestyle and the music scenes in Seattle and is thinking of relocating also. Use this tool to help her make a comparison.

User Tasks for Scenario 2

- How many days of light rain are there Miami compared to Seattle?
- Is there more rainfall in winter than in summer in Seattle?
- What about Miami?
- How much more rainfall is there in Miami?

Post-Test Questions

- Is the tool easy to use?
- Is the information sufficient?
- Are they clear to you?
- Do you think you have some basic information to help make a relocation decision?