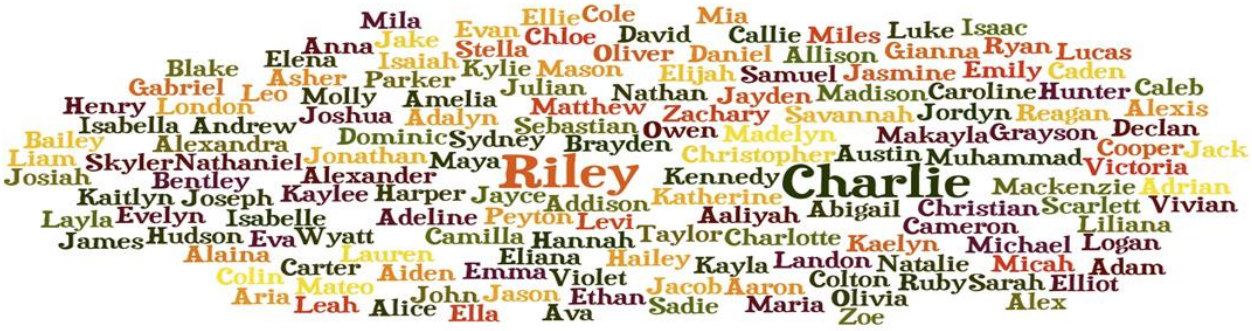
UNIQUE BABY NAMES IN WASHINGTON STATE
A NAMING TOOL FOR EXPECTANT PARENTS

Kira Cassels Valeria Gonzalez Samir Patel



ABSTRACT

This report illustrates our methodology for utilizing multiple naming datasets to create a tool that assists expectant parents in the naming process. Through a combination of name popularity data in the United States, a dataset that contains information on historical figures from early times to modern days, and a nickname data set, our objective was to provide a visualization that aids users in the name selection process. This paper provides information about: 1) initial research on our subject area, 2) requirement delineation and user goals, 3) ideation and exploration of designs, 4) usability testing, 5) explanation of our final visualization, and 6) an evaluation of our success and next steps.

INTRODUCTION

Naming of a child is of great importance to many expectant parents. Numerous websites, books, and other resources exist to provide naming information to assist parents in the naming process. However, naming is a very personal undertaking and an important decision that affects both the child and the parents. Our research suggests that uniqueness, culture, and success are key factors that influence naming decisions. Additionally, interviews and surveys, conducted with expectant parents, suggest that there are a plethora of naming resources that exist but a lack of resources that serve as a helpful tool to guide users in name selection. Competitive analysis of existing naming tools showed that many highlight name popularity, but our own user research found that uniqueness is more valuable to most users. Additionally, currently available tools neglect to address cultural importance and potential success indicators for names (two additional indicators we found users to desire from our research).

Our visualization seeks to bridge the gap between what parents desire and what they are currently afforded through naming tools. Through usage of robust datasets, which contain historical naming data in the United States, name data for historical figures, and occupations/industries associated with certain names, we will create a visualization tool that meets the user goal to find names that they would potentially like to use. Drawing on our research conducted with users, identification of where the gaps and pain points are with visualizations that already exist, and design methods proposed by significant figures in the field of information and data visualization, such as

Edward Tufte and Ben Shneiderman, we intend to create a novel, intuitive, and usable design. Through a user centered design process, which involves the initial user research, prototyping, usability testing, and iteration, we will ensure that the resultant product meets the user goals, needs, and abilities.

DESIGN QUESTION

Our design question was inspired by the needs of users that we spoke with during initial surveys and interviews, as well as a competitive analysis. A naming tool should be intuitive and aligned with the typical naming process of expectant parents, to best aid them in selection of a name for their child. Additionally, when speaking with parents, we found that they were interested in name uniqueness, cultural connection, and historical significance. This led us to ask the question, **"How can we assist expectant parents during their naming process, and afford them the capability to select a name that suits their needs through a display of information of interest?"**

GOAL

Our goal in approaching the design question is to create a visualization that would enable expecting parents the ability to search for potential suitable names for their baby. Through evaluation of previous works (in a competitive analysis) and performing surveys and interviews of our users (expectant parents), we will refine our goal to a narrower scope.

PREVIOUS WORK

In our research of related work on baby names, most works appeared to be strictly on popularity rank, and focused on the highest occurring names across the United States. These visualizations include other parameters such as date range of selection, gender, and geographic location (typically U.S. States). Typical tasks with visualizations related to this topic include filtering (by date, by location), zooming to view a specific part of the visualized information, and detail on demand, where users will hover to obtain additional information about the data by hovering or clicking.

Other visualizations reviewed focused on phonetics of names, such as the number of names selected based upon certain beginning and ending sounds.

Given the user's desires of having a resource of baby names connecting to uniqueness, we saw an opportunity there as this could be a matter of emphasizing low occurring names in a dataset. The challenge would be to find a dataset that could provide valuable information relating to success and culture.

The SSA dataset has previously been used for multiple visualizations. Most of these visualizations highlight name popularity, year of popularity, name trends, and geographic progression. Other visualizations show very specific examples of name trends, such as for pop culture figure names or simply showing the number of occurrences of a single name.

Existing Visualization 1

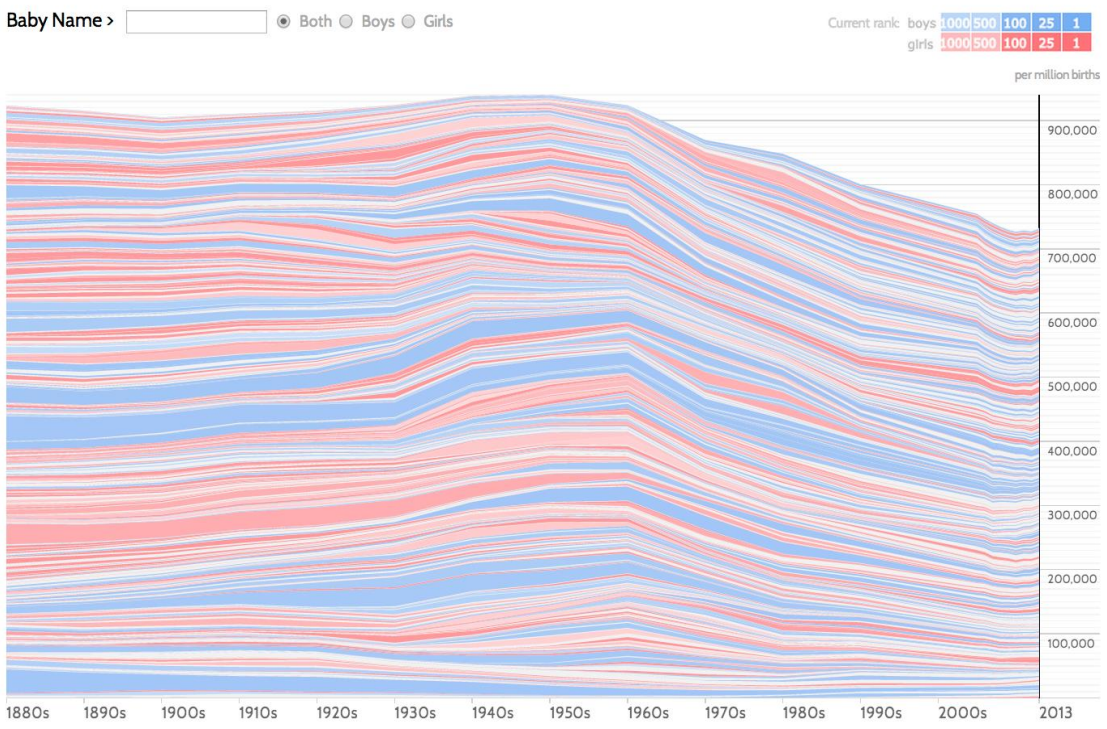


Figure 1. Baby Names Popularity - NameVoyager: Baby Name Wizard Graph of Most Popular Baby Names. (n.d.). Retrieved December 08, 2016, from http://www.babynamewizard.com/voyager

Description	This visualization displays 100 years of baby name trends. It allows you to type in a name and see a graph of popularity for that name over the past century. Additionally, you can filter by gender, see the rank of that name in a particular decade, and see the number of occurrences of that name.

Pros	• Clearly illustrates naming trends over time
	• User is able to see more and less popular names, based on the year
	• Interactive
	  ○ On hover: shows popularity rank by decade
	  ○ Ability to search for a name
	  ○ Ability to filter by gender

Cons	• Highlights popularity through color • Only displays four variables (year, occurrences, gender, and first name)

Existing Visualization 2

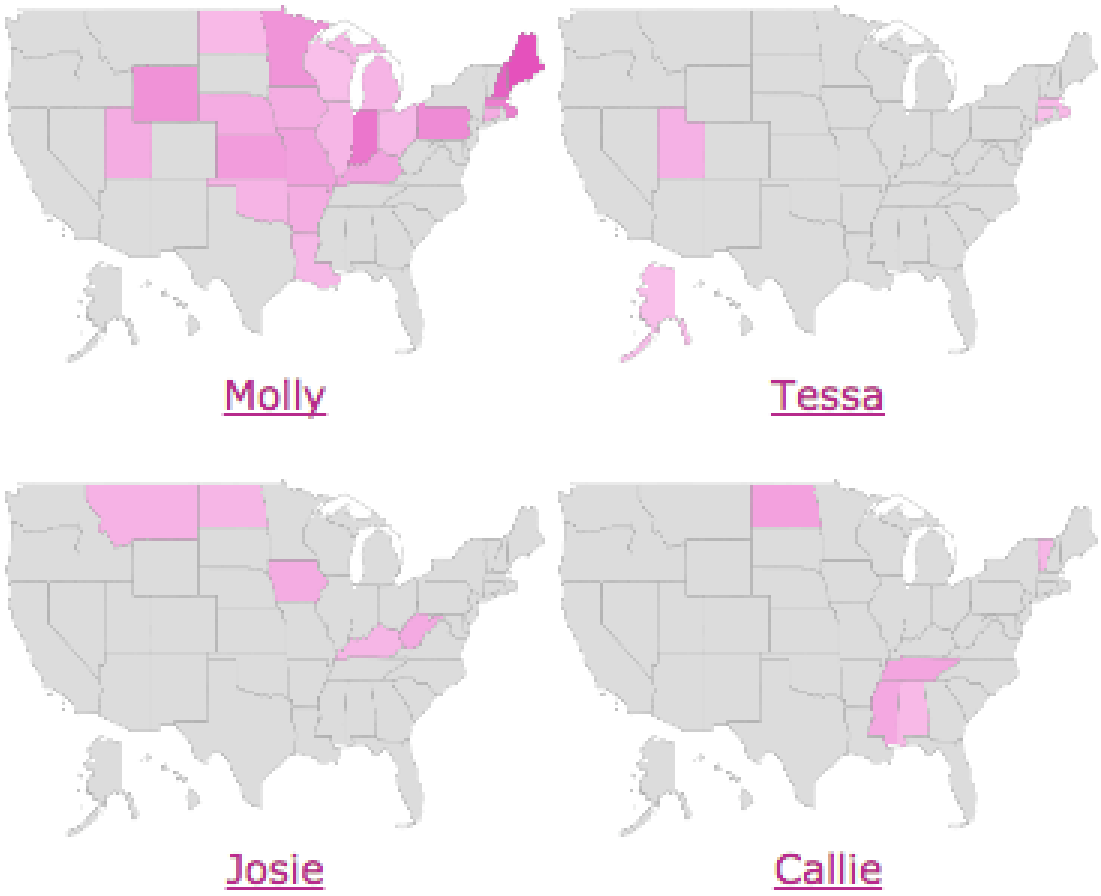


Figure 2. Baby Name Trends - NameTrends. (2016.). Retrieved December 08, 2016, from http://nametrends.net/

Description	This visualization shows name popularity over time and displays the rank of a particular name for each state in the United States in a given year between 1960 and 2015. Additionally, it provides graphs displayed as small multiples, which show names that were popular at the same time or in the same region.
Pros	• Shows geographic popularity of names • Displays historic popularity of names • Shows names that were similar at the same time • Shows names that were popular in the same region • Interesting usage of small multiples
Cons	• Fails to highlight uniqueness • More of a reporting tool than a searching tool • Does not tell a story because reporting on names is done by the individual name and is displayed on different pages (for historic popularity and geographic popularity) • Besides searching, the interactivity is not effective

Existing Visualization 3

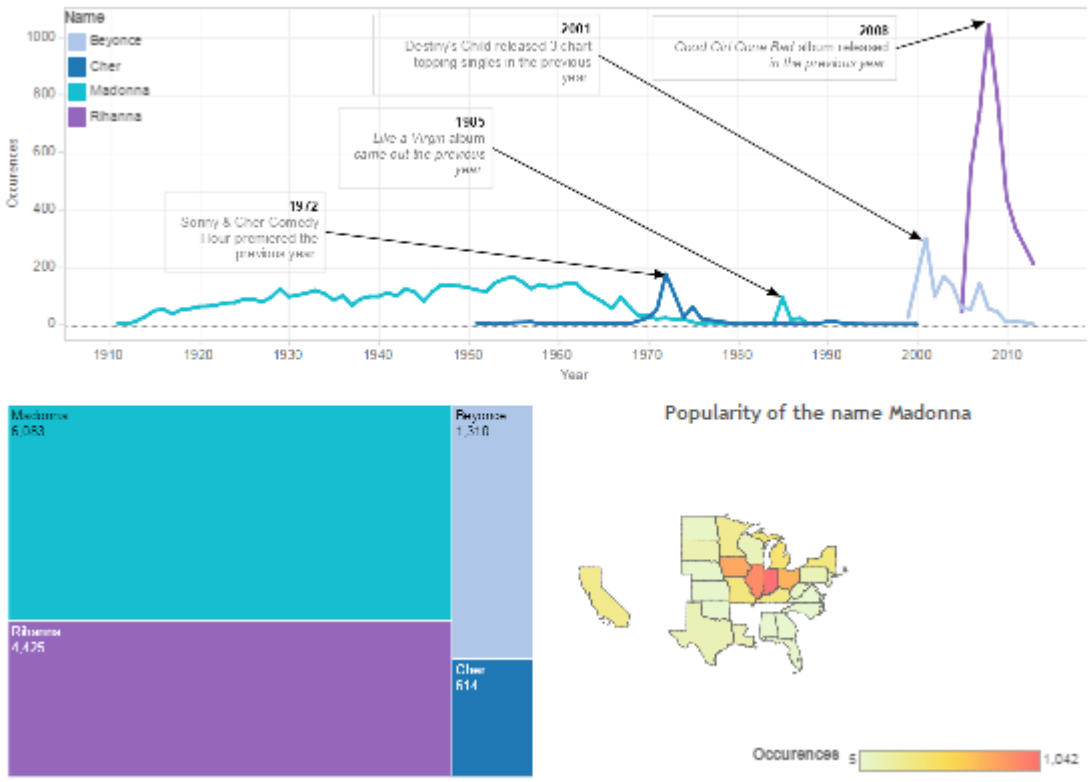


Figure 3. Comparing One Name Pop Divas -Exploring the SSA Baby Names Dataset. **(2015.). Retrieved December 08, 2016, from** https://public.tableau.com/en-us/s/gallery/exploring-ssa-baby-names-dataset

Description	This visualization shows name trends for a few pop culture figures over time, as well as displays the number of occurrences for these names and a map of occurrences in the United States. The visualization also provides information about significant happenings in the industry to give a bit more context about possibly why certain trends existed.
Pros	• Shows a time plot with contextual references to events in the year of spiked popularity • Shows geographic popularity of names • Displays historic popularity of names • Effectively encodes time, occurrences, and first name • Effective interactivity • Tells a complete story
Cons	• Fails to highlight uniqueness • More of a reporting tool than a searching tool • Only reports on specific names • No way to prove correlation between spikes in names and events

Existing Visualization 4

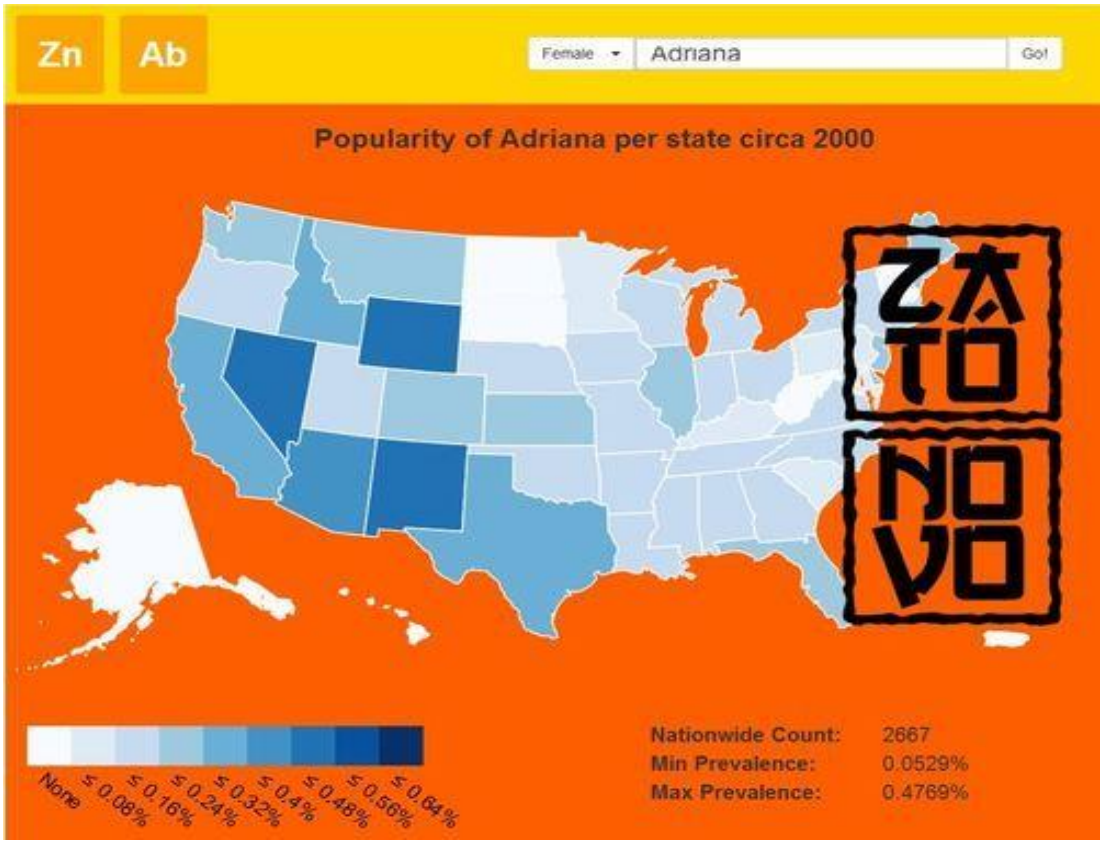


Figure 4. Zato Novo- Baby Names : Foresight from Insight. (n.d.). Retrieved December 08, 2016, from http://zatonovo.com/dataviz/baby_names

Description	This is an animated visualization, where a user is able to type in a name and the graph displays how the popularity of that name spread across the United States over the past century.
Pros	• Ability to see names based on gender • Displays progression of names over time, over geographic location
Cons	• Only highlights name popularity • Fails to use effective encoding for some variables (i.e. name prevalence) • Fails to provide additional detail, which is necessary for obtaining a complete picture • Cannot control what is being displayed, as the visualization is animated

RESEARCH

For preliminary research, we conducted interviews amongst expectant and recent parents to determine the process and factors they considered when selecting potential names for their child. Based on our user interviews, we retrieved information to help us scope our visualization project and determine the content that would make our visualization the most useful to our target user.

Prior to selecting datasets, we distributed surveys to expectant parents. The goal of this research was help us to identify potential data sets that would provide the appropriate information to our users through gaining information about the factors that users consider when naming a child. We received 11 responses to our survey. Below are the questions we asked on our survey and a summary of the answers:

Initial Survey Results

Question	Answer Summary
What factors do you feel affect how parents select their baby's name?	Uniqueness, Family, Tradition, Culture
Do you associate the naming of a child with how it may affect his or her success?	Majority say "yes"
If you were choosing a baby name, would you be interested in historical naming trends?	Majority say "yes"
How strongly would you consider a name that has a history in the pacific northwest?	"Not really"
Would you consider celebrities or more modern popular figures for names?	For the most part "no", "maybe musicians"
What factors do you consider in a middle name? Do you put much thought into a middle name?	"Family", "place names", "something that sounds good with a first name"
Are you more inclined to select a name that is less or more popular?	Majority say "less popular"
Are the names you choose influenced by popular culture? If so, please identify if you're more inclined to select a name based on books, music, movies, etc.?	"Books" and "music"

Additionally, we held 5 one-on-one interviews with expectant parents. The goal of this method of initial research was to determine the needs and goals of our target user, as well as to identify potential data sets that would provide the appropriate information to our users. Additionally, we wanted to gain more information about the naming process and the factors that users consider when naming a child. During these interviews, we asked parents the same questions as in the surveys but obtained more thorough findings.

Initial Interview Results

Question	Answer Summary
What factors do you feel affect how parents select their baby's name?	Uniqueness, Family, Tradition
Do you associate the naming of a child with how it may affect his or her success?	Majority say "yes"
If you were choosing a baby name, would you be interested in historical naming trends?	Majority say "yes"
How strongly would you consider a name that has a history in the pacific northwest?	All say "no"
Would you consider celebrities or more modern popular figures for names?	For the most part "historical figures" or "significant figures in my life"
What factors do you consider in a middle name? Do you put much thought into a middle name?	Family", "place names", "significant figures in my life"
Are you more inclined to select a name that is less or more popular?	All say "less popular"
Are the names you choose influenced by popular culture? If so, please identify if you're more inclined to select a name based on books, music, movies, etc.?	"Books" and "music" or "no"

From our surveys, most parents said the things that affected how they selected their baby's name were uniqueness, success, and family/culture. The value of uniqueness to the user was in regards to having a name that was not so common in the area amongst other children of similar age. An example where this would be a problem for parents would be if a child was in a class with 2-3 children with the same name.

Regarding a name that represents success, most parents felt the name they would choose could affect their child's future opportunities and achievements. Family and culture were also listed often by expectant parents as something they valued in the naming process, as many wanted to keep with names from their background or in their lineage.

To acquire additional insight, we also posed a question to users, asking how they felt about using popular names in the mainstream or historical culture for their child? Many parents seemed to reject the idea of names based on modern pop culture figures, but there was agreement by many users that they would consider names of respected musicians, figures or authors/characters of books. These results helped us determine the need of the user and led us to perform research of visualizations currently used, as well as searching for datasets that met a unique user need and did not replicate prior work.

DATA SETS

After the preliminary research stage was completed, the next step was to identify appropriate data sets to meet the defined goals for the user. A base dataset for name occurrences in the United States would be easy to identify given the plethora of baby name visualizations we identified in our competitive analysis. The difficulty would be in identifying a dataset to quantify or qualify "success" and "culture", which would expand the scope and variables at our disposable.

In the end, we chose to use the following datasets:

1. Birth Names dataset from the United States Social Security Administration Records
2. MIT Pantheon dataset from the Macro Connections Group at MIT Media Labs
3. Nickname and diminutive name dataset

DATA SET 1:

Data Set	United States Social Security Administration Records (ssa.gov)	
Static Link	https://www.ssa.gov/oact/babynames/limits.html	
Summary	We retrieved our primary data set from the Social Security Administration (SSA) records. This resource contains all recorded names at birth from 1910 to 2015 by geographic location (state), birth year, gender of the infant born with the name, and number of infants born with a particular name in a given year. This dataset, with its wide range and depth, is key for our study of creating a naming tool for a user, as it gives context to what historic naming trends exist and by geographical location. We were able to calculate some additional quantitative variables in order to provide users with more information and a better overall experience. Specifically, we took the inverse of the number of occurrences to create a new variable called "Uniqueness". We also parsed each name after the first letter to create a new variable called "First Letter". Lastly, we calculated the string length for each name, which resulted in a variable called "Name Length", containing a count of the letters in each name. Overall, the SSA dataset did have limitations given the limited variables (dimensions and measures). At best, alone, this dataset can provide naming trends which is something our project proposal research has shown to be exhausted in the mainstream. As a result, it will need to be joined with another dataset to provide a unique approach to a naming tool for the user.	
Size	5.7 million rows and is 6.6 MB in size.	
Variables	**Nominal**	• **First Name -** the baby's name in the SSA archive • **Gender -** M or F denoting male or female, from the SSA archive • **State -** the U.S. state in the SSA; the state that the user is from • **Birth City** - birth city of historical figure from Pantheon dataset

		• **Birth Country** - birth country of historical figure from Pantheon dataset • **Industry** - industry of historical figure from Pantheon dataset • **Occupation** - occupation of historical figure from Pantheon dataset
	Quantitative	• **# of Occurrences -** the number of babies born with that particular name in that state/year/gender • **Uniqueness** - the calculated inverse of the # of occurrences, used to emphasize the rareness of a name with a large quantitative value • **Name Length** - a calculated field, providing the count of the number of characters in a "First Name"
	Ordinal	• **Year-** year of birth recorded in the SSA archive • **First Letter** - first letter of a name, a calculated field taking the substring of the first character of the "First Name"
Quality Assessment	When we assessed the quality of the SSA dataset, it did not show any data integrity issues. Though it should be noted the data itself could be confusing in some cases, where names which are typically assumed to be for males were at times used for females in the past. Examples of this includes the rare occurrences of "Bennett", "Beau" and "Blake" for females (when searching for names which start with "B", which may confuse the user into thinking the data is erroneous when the data integrity was still intact. Another assessment of the dataset was its limited scope with nearly all nominal values and one quantitative measure, "# of Occurrences" (provided for a given name in a year and state). The "# of Occurrences" is a category we would invert and calculate a new field from for "Uniqueness". This meant a large value for "Uniqueness" represented a more unique name. There was a limitation to the upper end of the uniqueness value, as the SSA dataset kept the lower range of name occurrences at a value of 5. This meant a name had to be used 5 times in a year and U.S. State to make it in the record. While it would increase our project scope if the data set had more variables (particularly quantitative ones), the limited data domain is likely due to potential privacy issues. With that said, if it was possible to know such things as the city of birth, ethnicity, and birth month, this data could potentially provide some new insights to our users. But these data fields would not be necessary to answer the question our project has defined.	

DATA SET 2:

Data Set	MIT Pantheon (Mapping Historical Productivity)
Static Link	http://pantheon.media.mit.edu/rankings/people/all/FINE%20ARTS/-4000/2010/H15

Summary	We coupled the SSA data with a data set from MIT's Pantheon project, which considered how to map historical productivity based off web searches of popular historical figures. The MIT group's mission was to "quantify, analyze, measure and visualize global culture" (MIT Media Lab). The analyses and methodology for this project showed that cultural productivity (accomplishments) of a figure correlate positively to the historical popularity. The Pantheon dataset contains information about historical figures such as their name, place of birth, year of birth, gender, and occupation. Notably, it contained a metric called Historical Popularity Index (also referred to as HPI), which is a metric derived by the MIT group that quantifies the presence of that individual in Wikipedia page searches along with the number of languages globally. Using this dataset in conjunction with the list of names from the Social Security Administration would provide a novel dimension for our users to explore the significance of certain baby names with a connection to success and culture.	
Size	**11,000 rows and is 2 MB in size.**	
Variables	**Nominal**	• **Name -** Name of historical figure • **Occupation-** the occupation of the historical figure • **Country Name -** the home country of the historical figure
	Quantitative	• **HPI -** Historical Popularity Index, metric derived from how often this figure is searched on Wikipedia • **Latitude/Longitude -** coordinates of where the historical figure is from
	Ordinal	• **Birth Year** - birth year of historical figure from Pantheon dataset
Quality Assessment	The Pantheon dataset is robust but focuses on one quantitative measure, HPI, a measurement of popularity and success. Overall, its value is in providing a metric of historical productivity/success. We saw this as having the potential for a useful pairing with the SSA data.	

DATA SET 3:

Data Set	Nickname and diminutive name dataset
Static Link	http://www.tngenweb.org/franklin/frannick.htm
Summary	To provide additional information about names in the dataset, we located a data set that provided nicknames for first names used in the United States. During our second round of usability testing, users mentioned that they would be interested in seeing nicknames or name alternatives for each name. When prompted for more information about this, our users stated that this was a factor when selecting a name for their child. The nickname dataset was found on GitHub and contains common US given names (first name) and their associated nicknames or diminutive names. It is from a project created by the Old Dominion University - Web Science and Digital Libraries Research Group. This dataset was

	created by mining the dataset found at http://www.tngenweb.org/franklin/frannick.htm. Some of the names found in the dataset are very uncommon, as the dataset is used for genealogy purposes. Inclusion of the nickname dataset provides users with additional information, which can be accessed on-demand through the visualization, and will assist parents even further in the naming process. It serves as a secondary resource when joined with the SSA dataset.	
Size	1600 rows and is 35 KB in size.	
Variables	**Nominal**	• **First Name -** first names from a genealogical name dataset • **Nickname -** diminutive or nicknames based upon naming traditions
Quality Assessment	In terms of the data quality, issues were minor. The data values were in a .csv file that had the first name and the corresponding nicknames separated by commas. Each first name and its nicknames were on a new line. Additionally, this data set is limited to 1600 names. However, the names that it provides nicknames for are the more commonly used names in the United States.	

After selection of our datasets, an extensive data merging and cleaning process was done through use of the statistical programming language R. This allowed us to merge and join all three datasets and perform curation to put the data in a form where it was ready to be analyzed. The final merged dataset including all 50 U.S. States was 3.5 GB in size. Later in the user design process, it would be shortened to 70 MB by using a subset of only Washington State data.

EXPLORATORY DATA ANALYSIS

To understand our cleaned joined dataset, initially composed of the U.S. SSA baby names dataset and the MIT Pantheon datasets after our research stage, we performed an Exploratory Data Analysis (EDA).

The EDA was an exercise which helped us evaluate our dataset, the best way to visualize it and answer our users' questions. Some of the visualization forms we created included a tree map, U.S. map, a bubble plot, time plot, bar graph, and tables. These examples allowed us to see the full range of our data end to end and find some outliers. For example, we found the names "Beau" and "Bennett" as apparent female names in the record (which would later be confirmed as valid). We also found that given how big our dataset was, any functionality would require applying filters so that users could narrow down to their needs. Our datasets were very nominal heavy and we didn't have a lot quantitative variables, which also informed us there may be an opportunity to create/calculate new variables (which would later be done in Stage 2 of our User Testing). We also played around with layout options, comparing the option of creating a storyboard vs. a single interactive visualization dashboard.

USER AND VISUALIZATION DESIGN PROCESS

INTRODUCTION

As mentioned, a competitive analysis and research in the form of surveys was done to help determine user needs which were not currently being addressed. This was followed by identification and exploration of datasets to meet the requirements of visualization tool to help our users (expectant parents).

The next step was to move from the research stage into the "Ideation", "Prototyping" and "Evaluation" stages of the User-Centered Design (UCD) process (**see Figure 5 below**). This structured process would provide our team the required iterations in idea formation (ideation), development of prototypes, and evaluation via direct user testing to make timely revisions and optimize the final product.

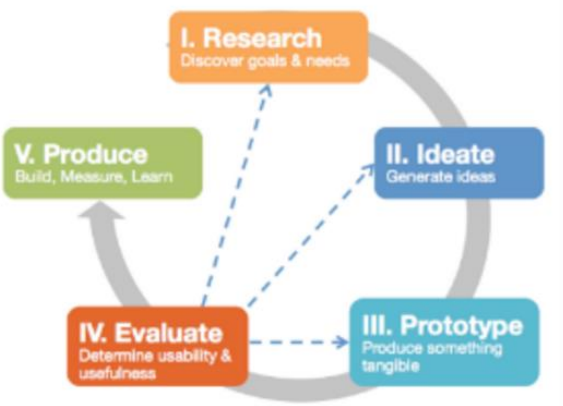


Image from Prof. Julie Kientz, HCDE, University of Washington,

Figure 5. UCD Process flow, outlined by Professor Julie Kientz, University of Washington

STAGE 1: LOW-FIDELITY PROTOTYPE

Ideation, Sketching and Persona

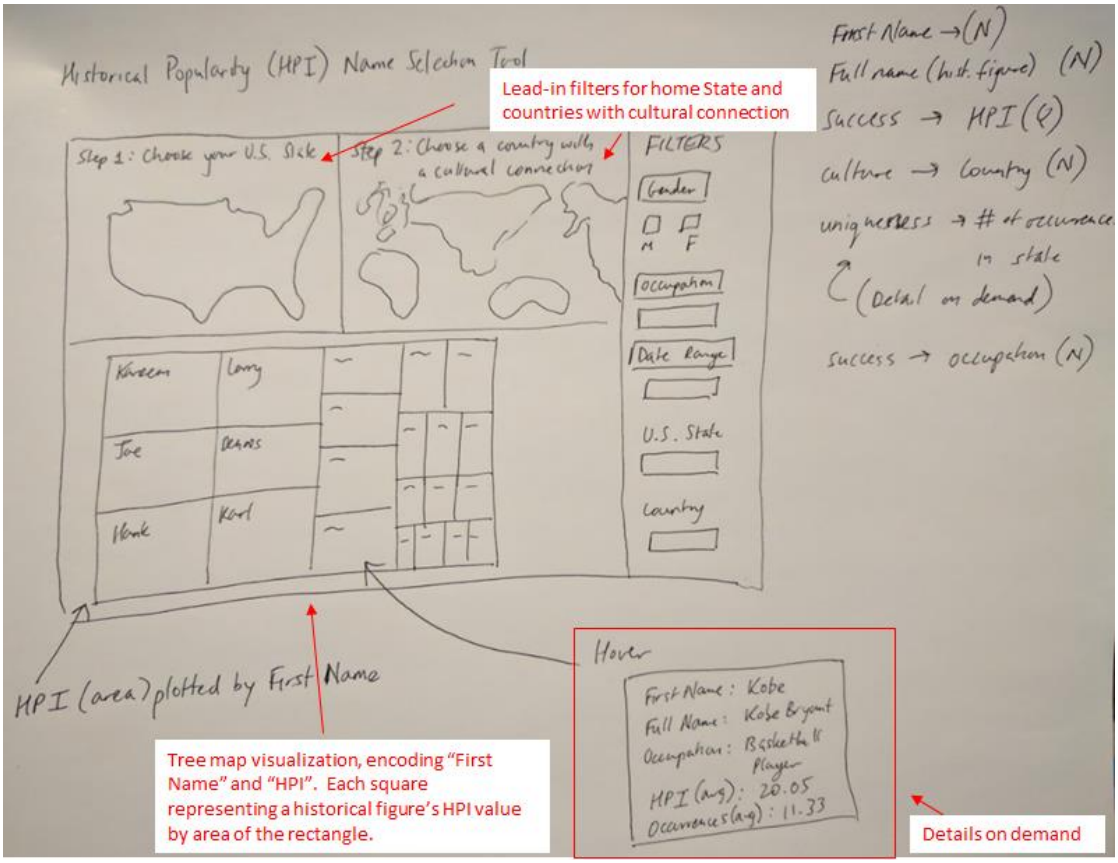
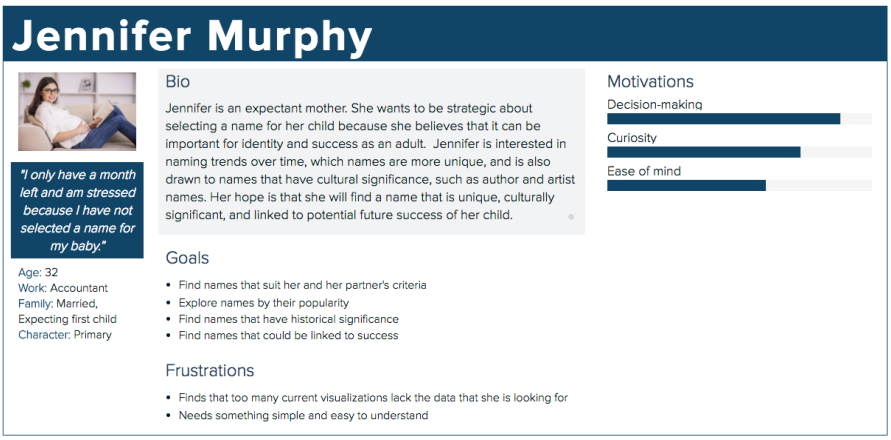


Figure 6. An initial prototype sketch for a naming tool dashboard for expectant parents. The primary metric encoded on the visualization was HPI (Historical Popularity Index, a quantitative measure) from the MIT Pantheon dataset, along with "First Name" from the U.S. SSA dataset.

Ideation and Sketching	When building our first naming tool prototype, our initial focus was on how we wanted to lead the user into the name searching process. Because our tool would be providing an interface for a random name search, the visualization would require the user to have lots of interactive and dynamic filters, depending on their respective needs. In our initial prototype sketch, we wanted to provide the user these filters in steps leading into the main visualization. As a first step, a map of the United States would be shown where the user would select their state of residence (**see Figure 6 above**). This would narrow down the list of names in the U.S. SSA dataset to just in their home region, since "uniqueness" is primarily relevant to the user's geographic location. As a second interactive step before leading into the main visualization, we provided a map of the world, which allowed the user(s) to select countries where they had cultural connection.

	These two filtration steps would then lead to the main visualization. In the design of the visualization, we chose "HPI" (Historical Popularity Index) as the primary metric to encode in a tree map visualization alongside "First Name", where each rectangle in the tree map represented a Historical Figure and the area encoded the quantitative value from the "Historical Popularity Index". The thought rationale in choosing "HPI" was our design team thought it would be the most interesting and useful to parents. And to fulfill other naming metric interest, they would the ability to filter for "# of Occurrences" (for uniqueness) to the right of the visualization. With this feature, the user would be able to search for a unique name while the output on the tree map dynamically adjusted with historical figures that represented success (through HPI) and a connection to culture, based off the countries the selected in Step 2. Another feature of the tree map visualization was a details-on-demand interaction integrated into the tree map (**see Figure 6)**. A pop-up appears when the user hovered over a particular name on the tree map, providing values such as the "# of Occurrences", the "Occupation" of the historical figure, and a quantitative value for the HPI (since visually, HPI is only represented by area is encoded on the tree map).
Persona	Our persona, Jennifer Murphy, was informed directly by the information we obtained through all of the research conducted thus far. This persona reflected the user goals that we found early on in our research process, to find a name that was unique and indicated potential success and cultural relevance. The persona below is the one created during the first stage of our research process: 

Usability Testing - Stage 1

Summary

In our first round of usability testing, we found that users were unfamiliar with tree map visualizations **(see Figure 9)** and were confused by them. They were also unable to understand the color application (each "First Name" was a distinct color). Our primary encoding of Historical Popularity Index (HPI), provided to reflect the user need of an indicator of future "success" for the baby name, was not clear to the users.

Overall, in our reflection, we saw that our main goal of helping users name their child was not met. The tree map was intended to help encode HPI as a primary variable for "success", and provide additional information for "culture" and "uniqueness" via details-on-demand. But, because tree maps are not a commonly recognized visualization form to our user base (expectant parents), it became clear that the encodings needed to be simplified. Along the same lines, the use of the HPI metric to represent "success" turned out be a hard sell to users. Since "success" is subjective depending on the user, having the metric as a primary encoding would not work.

As a result, we decided to iterate based on our user feedback. Our main take-away from the first stage of development was that our tool would need even more emphasis on usability and less on novelty. We would need to reconsider the priority levels of the three metrics that users identified ("uniqueness", "success" and "culture") based on how effectively we could encode them to their benefit of the user.

Methodology

The methodology we utilized for our first round of testing was a Contextual Task Analysis, in which we informed users a role that they would adopt during testing. We gave users a scenario and asked them to execute a task for the given scenario. During this first round of testing, we wished to obtain information about what parts of our visualization (encodings methods, graph type, etc.) made sense and which needed improvement or alteration. Additionally, it was our goal to see which variables and measures were useful to our users and which were not. All of the users we tested were expectant parents or had a child within the last 6 months. Four participants were female and two were male.

After each task, we asked users to rate the difficulty of the task (on a scale of 1-10) and asked the user to report on if the visualization provided the information that they expected. Additionally, we made notes on their comments during each task completion, success, time on task, number of errors, and any guidance requested. After the task completion portion, we conducted a short qualitative interview.

Low Fidelity Prototype and User Test Results

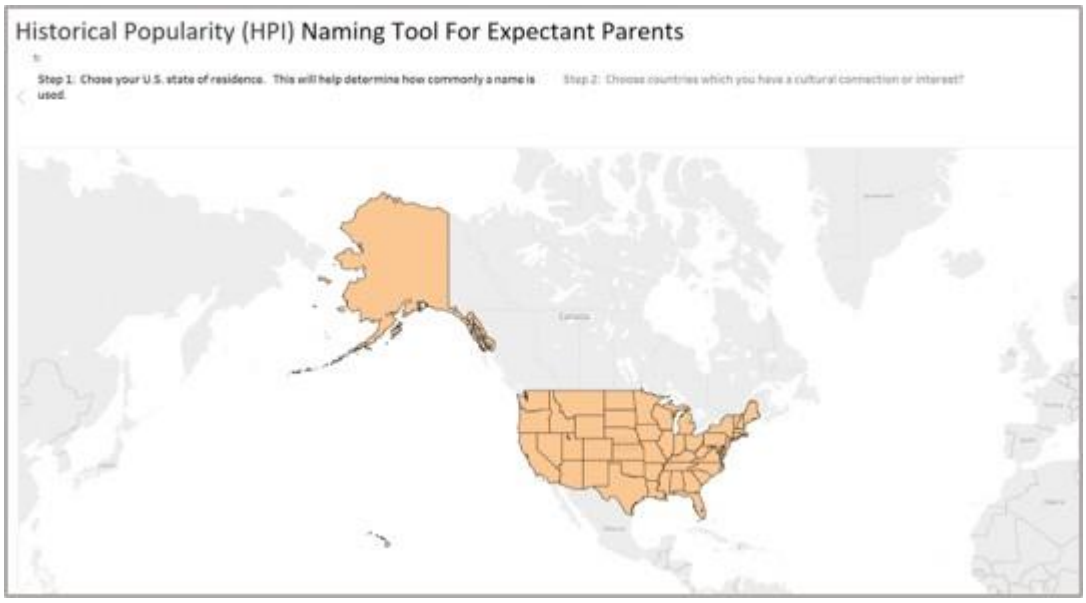


Figure 7. Low-Fidelity Prototype for Naming Tool: This is the first slide in Tableau story mode to help the user filter by their U.S. state of residence.

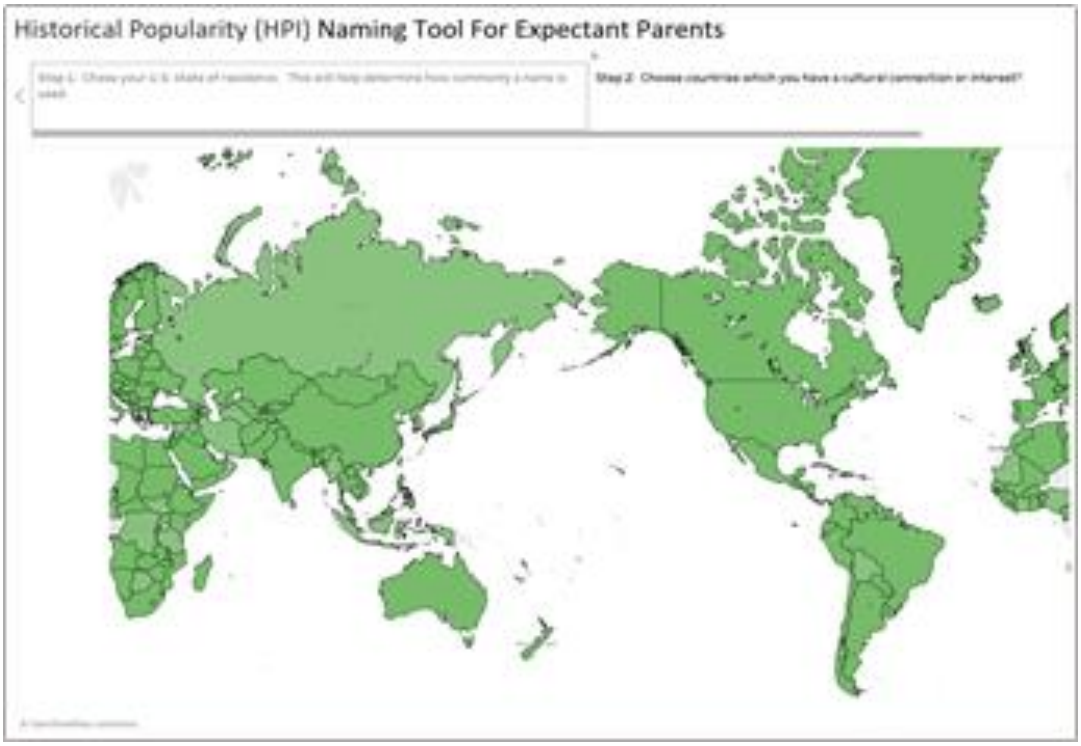


Figure 8. Low-Fidelity Prototype for Naming Tool: This is the second slide in Tableau story mode featuring a world map to help the user filter by a country where they have a cultural connection.

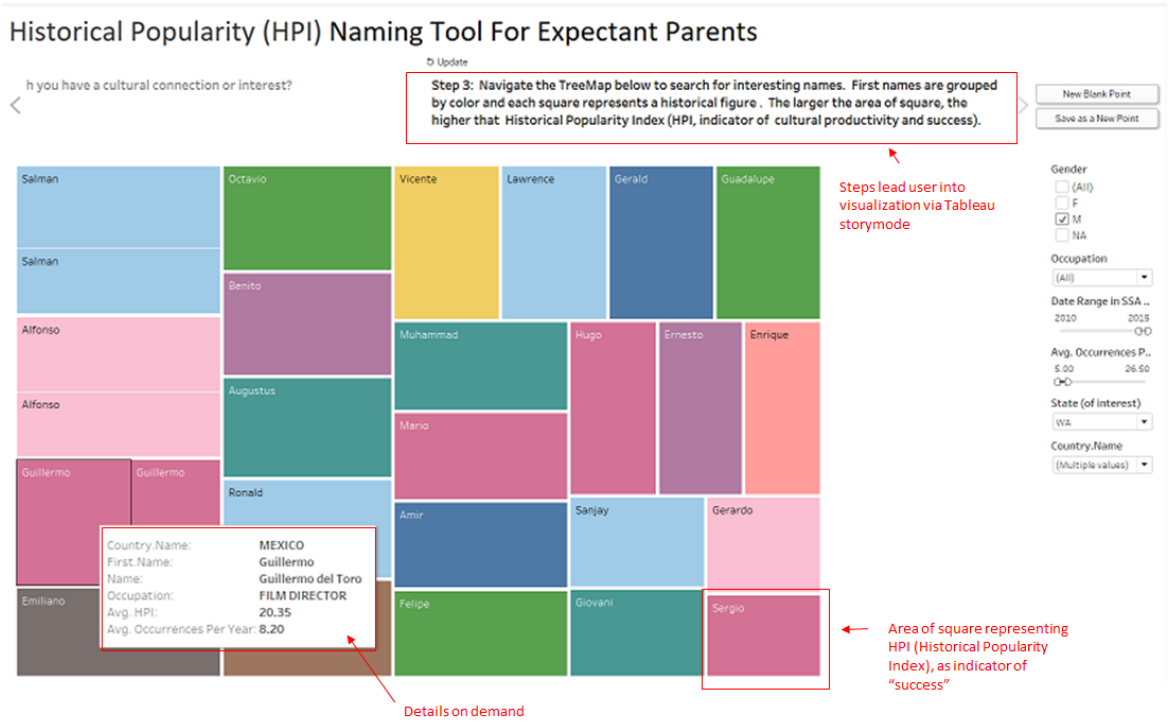


Figure 9. Low-Fidelity Prototype for Naming Tool: This is the third slide in Tableau story mode featuring a tree map visualization to help the user filter to a name of a historical figure. The size of each rectangle on the tree map encodes the Historical Popularity Index value (a metric representing "success").

Visualization Description	The low-fidelity prototype we presented to users, incorporated the same narrative, filters and visualization as our initial sketch but used Tableau story mode to ease the user into the visualization. The idea was to help the user by providing some context on how to use the two maps provided to filter for their state of residence and country of cultural connection.
Research Goals	1. To learn what expectant parents find useful. 2. To see if users can understand the measures and data that we are encoding. 3. To see if parents could effectively apply filters and gain feedback about the visualization, based on our use case.
Participants	6 Participants (4 female, 2 male), Expectant Parents or parents of an infant aged 6 months or younger
Method	Contextual Task Analysis- a user research method that applies ethnographic observation and one-on-one interviewing to understand the task procedures that users follow to reach their goals.
Scenario	Maria and Abhishek met as students at Cambridge University in London. They are now happily married and expecting a baby boy. They moved to Seattle for Abhishek's new job and were researching baby names that would be unique but have some historical significance in India, Mexico, and the U.K.

<table>
<tr><td>Tasks</td><td colspan="2">Task 1: You would like to find names used for males. Please filter the data to see only male names.
Task 2: You would like to see male names from only India, Mexico, and the UK. Please filter the data to see only these names.
Task 3: You would like to see which of these names have been used in WA state. Please filter the data to see only these names.
Task 4: You would like to find additional information about a name presented on the screen.
Please go ahead and find more information about any name in the visualization.
Please answer the following questions:
What is the name that you selected?
What is a full name for a historic figure with that same first name?
What is the historical figure’s occupation?
How many occurrences of this name exist?
Task 5: You and your partner have decided that you’re only interested in the least popular names used between 2005 and 2015. Please find less popular names between 2005 and 2015 used less than 60 times and select one that you might like to use.</td></tr>
<tr><td>Qualitative Interview</td><td colspan="2">Answer questions:
Did this visualization help answer the question in the scenario? Why or why not?
What is helpful in this visualization?
What is distracting or not value-added?
In order, what are the steps you take when selecting a name for your child and what factors do you consider?</td></tr>
<tr><td rowspan="2">Results</td><td>Usability Issues</td><td>Labels needed more clarity
HPI (Historical Popularity Index) was a confusing parameter to the users
The tree map visualization form is confusing to the expectant parent user audience</td></tr>
<tr><td>Positive Feedback</td><td>Liked the ability to see names based on different criteria
Liked the concept of a naming tool
Ability to search for historic figures is interesting</td></tr>
</table>

		• Uniqueness is more important than culture or success, but all are important • Like the idea of a search for a name based on different criteria
	Negative Feedback	• Typing text to filter is annoying (can search for WA, but not Washington) • Meaning of HPI is not clear • Tree map is confusing • Column labeling issues (duplicates and abbreviations), which makes it confusing • Want a tool that actively helps them in the process of naming
Improvements	• Reduce the scope of our visualization to primarily focus on uniqueness, and de-emphasize culture/success but still include it • Only use the dataset for WA state (our user base) for easier implementation • Remove HPI as a variable, as it is too confusing • Change from a tree map to a dashboard of more effective visualizations, which will be ordered in a way that serves as a tool and resource for people naming a child • Tell a more complete story of the data through visualization layout and variable inclusion	
Milestones	• Analysis of usability testing • Alternate visualizations to the tree map • Remove extraneous data from the combined dataset • Map of the name selection process	

Additional feedback (Mid-term)

The feedback that we obtained from our classmates and professors on our Mid-term presentation was similar to what we heard from users. The most common feedback was as follows:

Classmate Feedback	• Need to provide more information to the user about HPI, as there is confusion about the context of and variables related to Historical Popularity Index • Not convinced that such a tool can help users to make decisions • Tree-map is confusing • Other visualizations suggested are: scatterplots or line graphs related to trends of names

	• Confusion about why a user would be interested in HPI, instead of uniqueness of name or cultural significance of name • Would like to see more on trends of name popularity over time based on success, or based counts
Instructor Feedback	• Unclear how the motivations and tasks of the users translated into actionable items in the UI • Tree map likely not the best visualization for these datasets • Focus on a clearer way for users to identify uniqueness • Not clear if the intent is to walk users through a very specific task (as opposed to enabling) exploratory analysis

STAGE 2: RE-DESIGN AND HIGH-FIDELITY PROTOTYPE

Ideation, Sketching and Persona

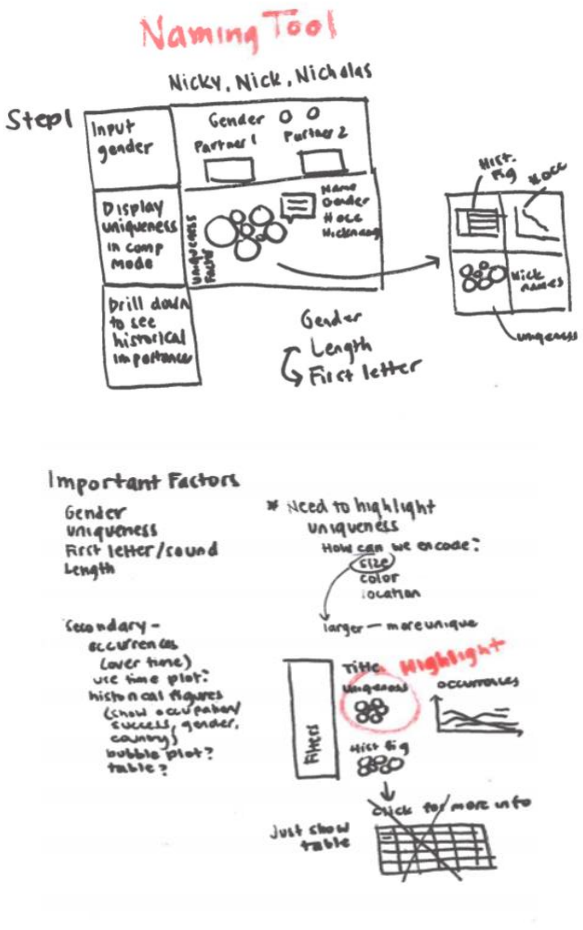
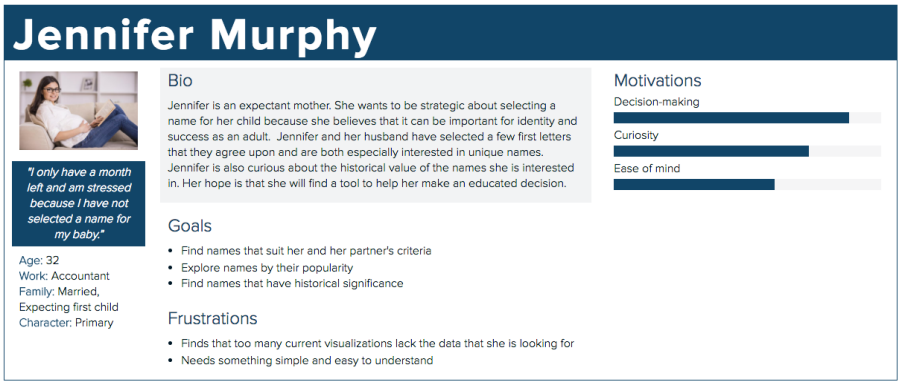


Figure 10. Second round of ideation to resolve issues from stage 1 of usability testing.

Ideation and Sketching	During development of our revised naming tool prototype, we utilized research findings from the first round of usability testing, as well as the feedback from classmates and instructors. As mentioned, our takeaway was to focus less on novelty and more on usability. In applying this to our visualizations, we looked at simpler and more effective visualizations for the user. Additionally, we sought to make sure that the visualization was scoped appropriately and that the correct variables were highlighted in the visualization. Through a better understanding of the goals users had with our visualization, which were to find names that they were interested in, we decided that we needed to include more filters that addressed actual steps that users took when finding a name (gender, first letter(s) of a name, uniqueness, etc.). We explored different types of plots, such as a bubble plot, bar chart, and time plots to encode our data points, and determined that we should include multiple visualizations (**see Figure 10**) so that the data could tell the users a complete story about a name. At this stage, we determined that the steps a user would take when using our visualization would be to apply filters to find a name and obtain details about names by hovering. We also decided that details-on-demand would be included when hovering over a name. After reviewing the usability testing and survey data again, we determined that uniqueness was the measure that we should highlight on our visualization since it was a measure that had little ambiguity while also being a user desire. We would include historical figure and success data as a supplementary visualization. We did a subsequent brainstorm, where we looked at how different types of visualizations could best tell a story about this data and suit the user needs. It was decided that we would utilize a dashboard, which would include a bubble plot to show the names and their uniqueness, a table to display historical figures with the same name and their occupations, and a time-series plot that showed the number of occurrences. This would be coupled with a set of filters to the left of the dashboard to help the random search process. By filtering names based on gender, first letter of a name, name uniqueness, name length (in characters), number of occurrences (by year), the new encodings highlighted the uniqueness of a name by displaying more unique names as larger bubbles in a bubble plot. Upon hover over a bubble, the user can see the name and a quantitative value for the "uniqueness". The user can then scroll through names on the historical figures table to see any figures with the same name and their occupation, gender, and country of origin. On hover over a line on the occurrences time-series plot, a user can see the number of occurrences for that name in a given year to look at any interesting trends.
Personas	Given the narrowing of scope and re-design for the project, we needed to update our user persona. Through additional research, we obtained a clearer structure of the process parents follow when selecting a name for their child. We found that parents are often interested in names that begin with a particular letter, are used for a certain gender, and are unique. The new persona had the same demographic information, but with this updated needs to reflect our findings.



Usability Testing - Stage 2

Summary

In our second round of usability testing for our visualization prototype (see Figure 11 below), we received a more positive response from users. The visualizations were easier for the users to understand due to the simpler encodings. The constructive feedback did come in the area of interactions, where participants said that labels for categories were somewhat confusing (ex. "First Name" from the baby names dataset and “Name" of the historical figure from the Pantheon dataset, representing different values but sounding the same).
This round of usability tests also showed 2 of 5 users clicking outside of the visualization to look a historical figure's name up on Google. Two participants mentioned wanting to be able to search for a specific name or a few names to compare, as they already have a few names selected. Overall, all users cited that the ability to search by name uniqueness was useful and 3/5 mentioned that having names of historical figures was interesting and would influence their decision for a baby name.

Methodology

The methodology we utilized for this round of testing was Unstructured Exploration, in which we gave users a brief explanation of what the tool would be used for and allowed them to explore the functionality without any prompting or tasks. During this second round of testing, we wished to see how users would interact with our revised visualization and obtain information about their process, usage of filters, and interactions with the visualization, as well as ideas for improvement. Additionally, it was our goal to see if the new variables and measures that we included were useful to our users and which were not. All the users that we tested were expectant parents or had a child within the last 6 months. Three participants were female and two were male.

For each participant, we recorded their process, for searching for a name using the tool, as well as any comments they made during usage. Because of this feedback, we decided to make changes to our visualization.

High Fidelity Prototype and User Test Results

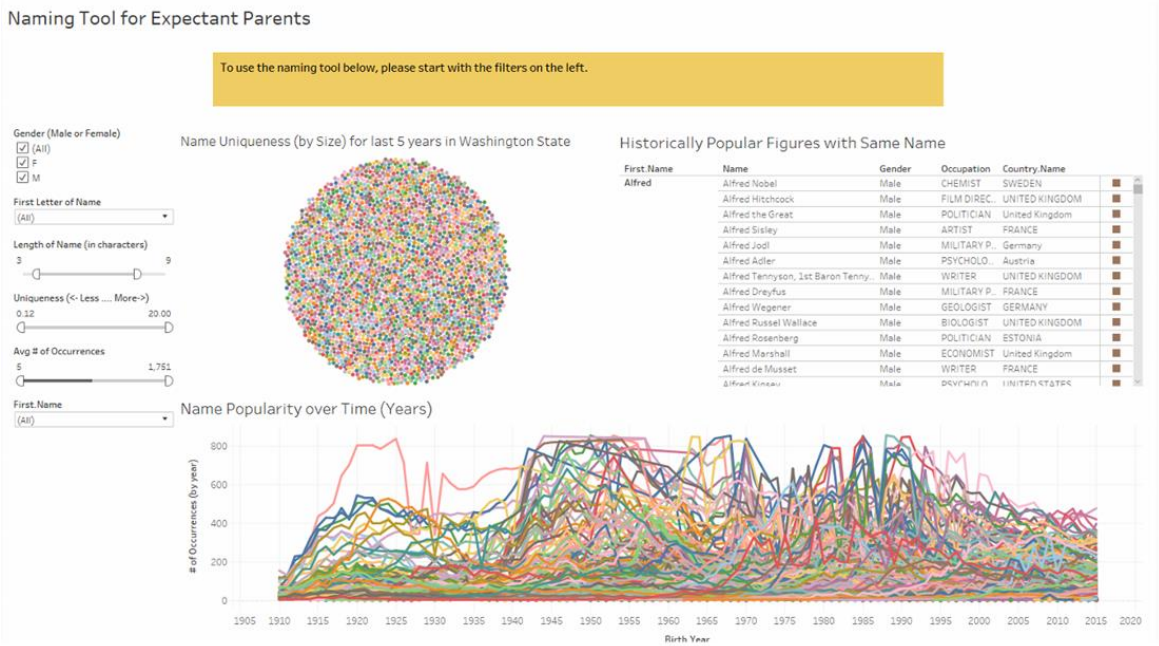


Figure 11. High-Fidelity Prototype of visualization dashboard used for testing.

Research Goal	To see how users interact with our revised visualization and obtain information about their process, their usage of filters, and their interactions with the visualization. See if the new variables and measures that we included were useful to our users and which were not, as well as elicit ideas for improvement.
Participants	5 Participants (3 female, 2 male), Expectant Parents or parents of an infant aged 6 months or younger
Method	Unstructured Exploration- users given a brief description of what the visualization intends to do and users are instructed to explore the functionality freely without any prompting or tasks
Prompt	You are an expectant parent and would like to find a name for your new baby. Please use this tool to search for and find a name for your child. I'm going to ask you to think aloud as you use this visualization, making sure to tell me what you are thinking as you explore.
User Process	• Filter gender (5/5) • Select letter(s) for first name (5/5) • Hover over names on bubble chart (4/5) • Changed slider for uniqueness (5/5) • Clicked on name on bubble chart (5/5) • Scrolled through names of historical figures (5/5) • Conduct search for historical figure, if unfamiliar (3/5) • Hover over information on occurrences plot (5/5)

Comments	• "Titles of the graphs (column names) are confusing" • "I want to know who this person is" (historical figures chart) • "The beginning letter is really important to my husband and I" • "I really like being able to see names that are less common" • "I've only ever seen the name rankings. Looking based on uniqueness is really helpful!" • "It would be interesting to see nicknames or alternatives for these names." • "I'd like to see names by theme. Nameberry has that information and I've been using that site." • "Selecting less common names is important to us." • "I want to search for names that my wife and I have talked about." • "These titles don't seem right." • "I'd like to see more information about this name when I hover over it."
Feedback	• Labeling wans confusing and needs to be improved, so that it is clear and concise • Users would like the ability to search for a particular name • Add additional information about the name with details on demand principle • Users would like an easier way to search for historical figures with which they are unfamiliar • Users would like to see nicknames or name alternatives
Improvements	Change labels to make them clearer • Provide contextual aids for the user to understand filters and how to use the visualization • Add ability to search for historical figures on Google • Add additional information to detail for each name, including occurrences and nicknames (from a new dataset) • Set a default startup view to avoid having the users overwhelmed (or underwhelmed) by the unfiltered overview of the large dataset
Milestones	• Changes to original visualization to reflect tasks in updated persona • Address issues and comments from second round of usability testing • Locate dataset for nicknames • Add functionality for Google search • Improve interactivity guides and functionality • Create a "home" state for the visualization

STAGE 3: Final set of testing and iteration

Changes applied from Stage 2 Feedback

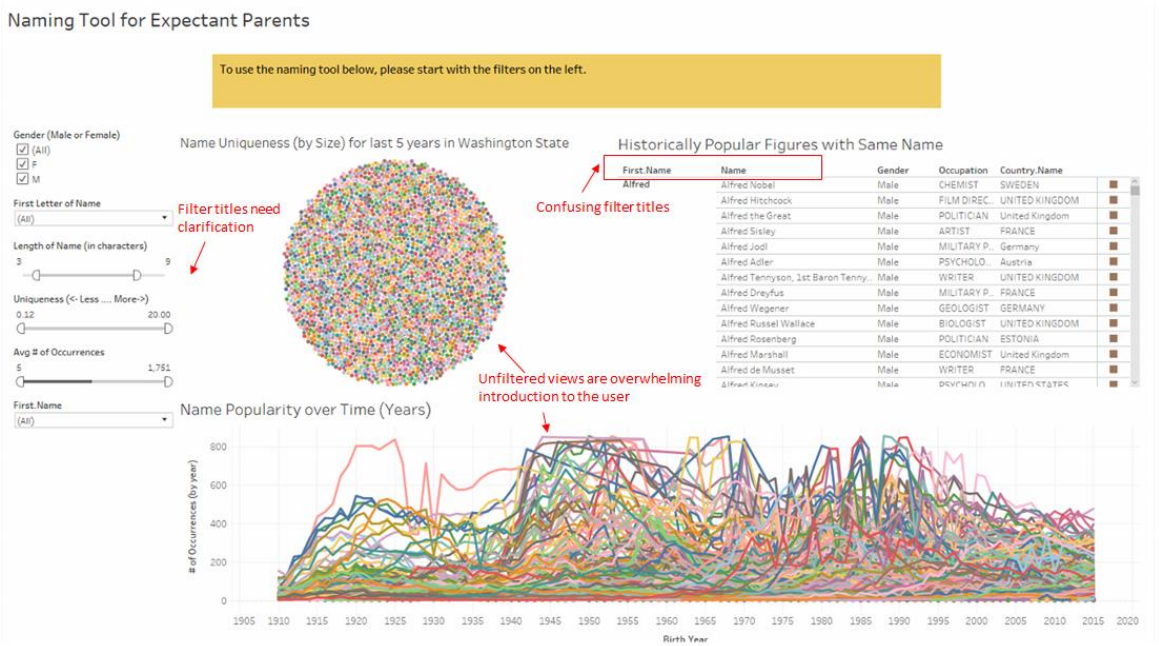


Figure 12. The High-Fidelity Prototype from Stage 2 shown above with areas of improvement highlighted in red.

Changes incorporated	During this stage, we refined our visualization through re-arrangement and clarification of filters and incorporating a specific name search feature for users who did not have need for a random name search (**see Figure 12**). Also, since users showed interest in Round 2 of usability testing in finding and searching more information on names, we integrated a nicknames dataset as a "detail-on-demand" in the bubble plot when hovering over a name. In addition, a Google search feature was included when clicking on a historical figure in the table visualization. We used Tableau to make a few versions of the visualization, with filters and the search boxes tested in different orders. And to ensure the filters met the thought process of a user during random searches, we compared the order of the filters with the (average) steps that users told us they use to find a name. To aid the filtration and generally usability of the visualization, we supplemented it with text aids for filters and to the right of the dashboard. Finally, since the prior prototype showed an unfiltered view initially where all our data is displayed, we wanted to set up a default view with some names that were pre-set. This would help the user understand the value of the visualization by seeing 4-5 names, as opposed to an overwhelming (or underwhelming, depending on the perspective) view of all the data.

Usability Testing - Stage 3

Summary

During the third round of usability testing, our team sought to find out any additional changes to the visualization that we might need to make and to see if the revisions we had made after our second round of testing were helpful to users completing tasks with the tool (**see Figure 13**). We found some minor bugs that we needed to address, specifically interference between the search functionality and the filters. The distinct tasks of performing a random name search and performing a specific name search was not clear by the organization of the filters. If a specific name was entered in the "First Name" field, this would prevent the user from being able to use the random search filters and vice versa.

Additionally, given this round was our first test with our visualization uploaded online to Tableau Public, users found that the visualization was skewed or too large on certain computer screens. At times, they could not read certain titles or needed to scroll to view data.

Methodology

The methodology we utilized for the third round of testing was Contextual Task Analysis, in which we informed users a role that they would adopt during testing. We gave users a scenario and asked them to execute a task for the given scenario. During this third round of testing, we wished to obtain feedback about the overall usability of our visualization, including interactions and understandability. Additionally, we sought to find out if there were any additional changes that needed to be made. All the users that we tested were expectant parents or had a child within the last 6 months. Three participants were female and two were male.

After each task, we asked users to rate the difficulty of the task (on a scale of 1-10) and asked the user to report on if the visualization provided the information that they expected. Additionally, we made notes on their comments during each task completion, success, time on task, number of errors, and any guidance requested. After the task completion portion, we conducted a short qualitative interview.

As a result of this feedback we found some minor changes that we needed to make, in order to meet the needs and capabilities of our users.

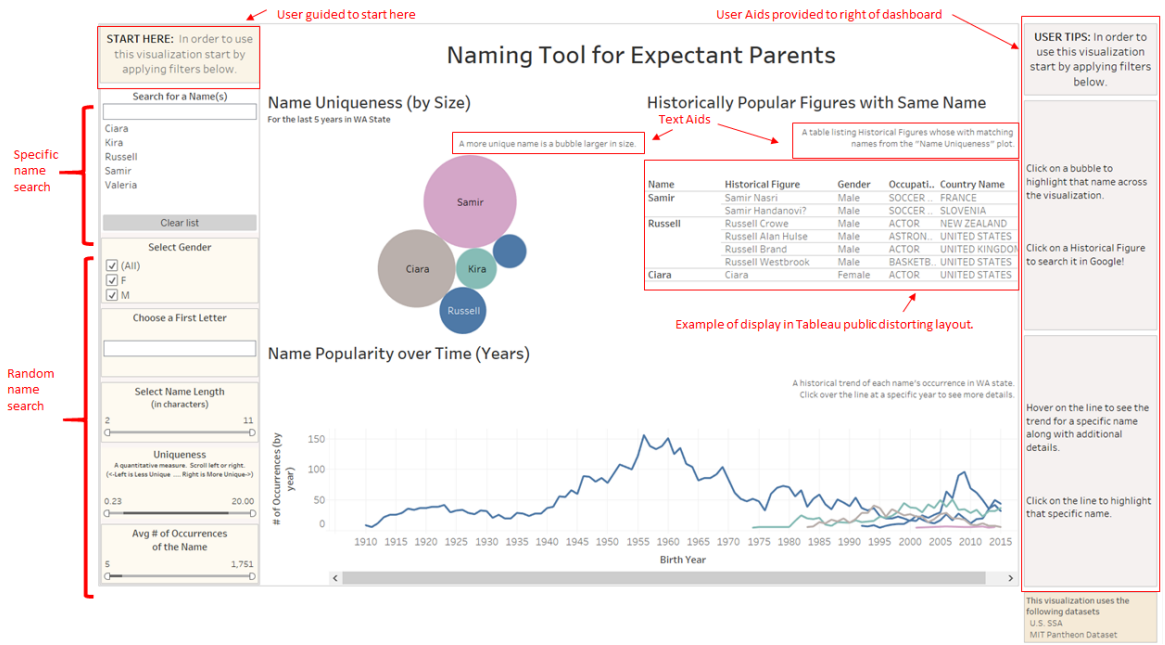


Figure 13. Stage 3 visualization dashboard used for testing with notable features and issues noted.

Research Goal	To determine the overall usability of our visualization, including interactions and understandability. Also, to find out if there were any additional changes that needed to be made.
Participants	5 Participants (3 female, 2 male), Expectant Parents or parents of an infant aged 6 months or younger
Method	Contextual Task Analysis- a user research method that applies ethnographic observation and one-on-one interviewing to understand the task procedures that users follow to reach their goals.
Scenario	Susan and Mark Sammy are expectant parents of a baby boy. They want to name their baby boy a name that starts with an 'M' so that all the boys in the house have names with M. They have a few ideas about possible names, such as Michael and Matthew, but want more options and information about this names. They are looking for names that are unique, are easy to pronounce and if the name has some historical importance that would be a plus.
Tasks	• You would like to find names used for males. Please filter the data to see only male names. • You would like to see the male names "Matthew", "Michael". Please filter the data to see only these names. • You would like to see additional male names that begin with the letter "M". Please filter the data to see only these names.

	• You would like to see names that have a uniqueness score of 14.00-20.00. Please filter for names that have this uniqueness. • You would like to get more information about a historical figure that has a name that you are interested in. Select any name that you like and find a historical figure with that same name. • You would like to know how many times the name was used in 2015. Please find this number.	
Results	**Usability Issues**	• Users could not see the entire visualization on the screen • Users were unable to search for names and apply filters at the same time
	Positive Feedback	• Benefitted from uniqueness measure • Appreciated the tips on the side of the visualization • Historical figure table remains an interesting and helpful feature
	Negative Feedback	• The visualization is too large for the computer screen and the user does not want to zoom out or scroll • Did not like not being able to search for particular names and apply filters- there is interference between searching and filtering
Improvements	• Optimize the visualization layout for display, so that users can see the entire visualization regardless of the computer they are using • Changed order of filters to highlight ability to use this for a random name search as well as a specific name search • De-emphasize on exact name search to the bottom of the filters panel since majority of users conducted random searches	
Milestones	• Final visualization with adjustments made per findings of the second round of usability testing • Make visualization more compact (no scrolling or zooming out required) • Final visualization creation	

Updates Leading to Final Visualization

The main takeaways from the last round of usability testing were to fix filters to avoid confusion in the name search options provided and improve layout/display settings to avoid complication on Tableau Public. We found that the search issues could be addressed by improving filter labeling and organization to clarify the ability to conduct a random name search and specific name search (**see Figure 14**). The specific name search filter was prioritized at the top earlier thinking users would gravitate towards preconceived names; however, the last

usability testing provided feedback otherwise, so this was moved to the bottom of the layout. Also, in consideration of the confusion with user interaction and how to apply/clear filters, we provided additional text aids to help the user.

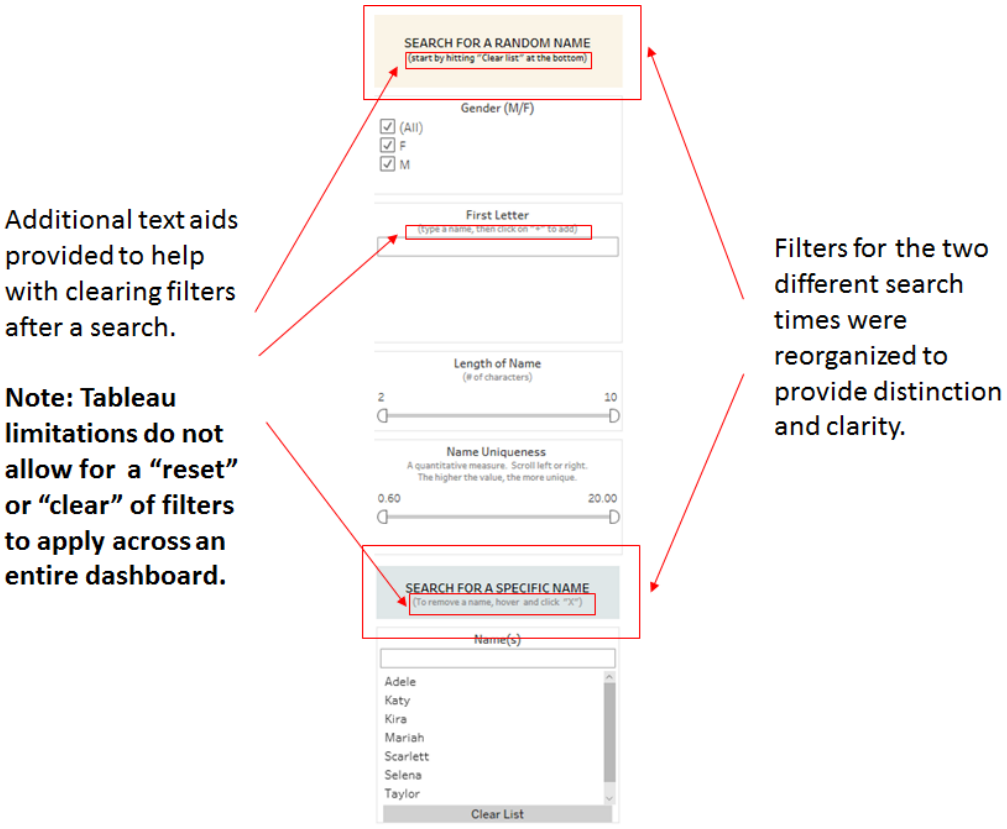


Figure 14. Reorganization of the filter layout helped distinguish the two functional filtration modes (for a random search and specific search of names). In addition, text aids for filter modes were included to help the user clear a search to avoid confusion from the two separate search functions.

The display issues would be addressed by resizing Tableau layout containers along with setup of an "Automatic" display configuration. While the "Automatic" display settings do a reasonable job of adjusting the visualization to fit varying displays, it is still prone to error and can distort layouts. In our test process, we often saw a distortion in the display once uploaded to Tableau Public. To cover the most common types of displays, we took additional measures to enable Apples and Windows laptops to be compatible in viewing the visualization in its entirety.

FINAL VISUALIZATION

Tableau Public Link:
https://public.tableau.com/profile/publish/Naming_Tool_rev15/Dashboard2#!/publish-confirm

***Note**: For optimal viewing, if you have Tableau Desktop installed on your computer, it is recommended that you download the Tableau workbook from the Tableau Public view and open it remotely. The download option is found in the bottom right corner. Otherwise, it is also recommended when using Tableau Public to try different interest browsers (such as Chrome, Firefox) if issues appear in your default browser.*

The sections below will present and outline the final visualization (**shown in Figure 15 below**) after the comprehensive design and user testing stages. In these sections, we will go through a breakdown of the visualizations, including discussion of interactions and layouts that comprise the entire dashboard view. This analysis is done in respect to the best-known principles and application of information visualization.

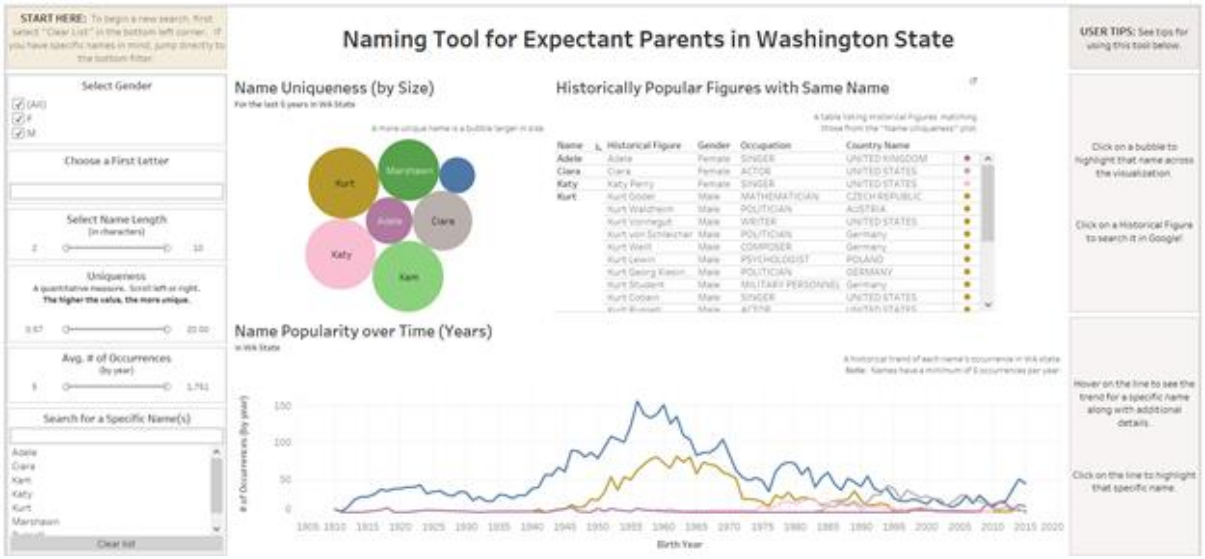


Figure 15. The final version of the Naming Tool Dashboard shown in its entirety. Filters are featured on the left-hand side, the dashboard with visualizations in the center, and textual aids for users on the right-hand side.

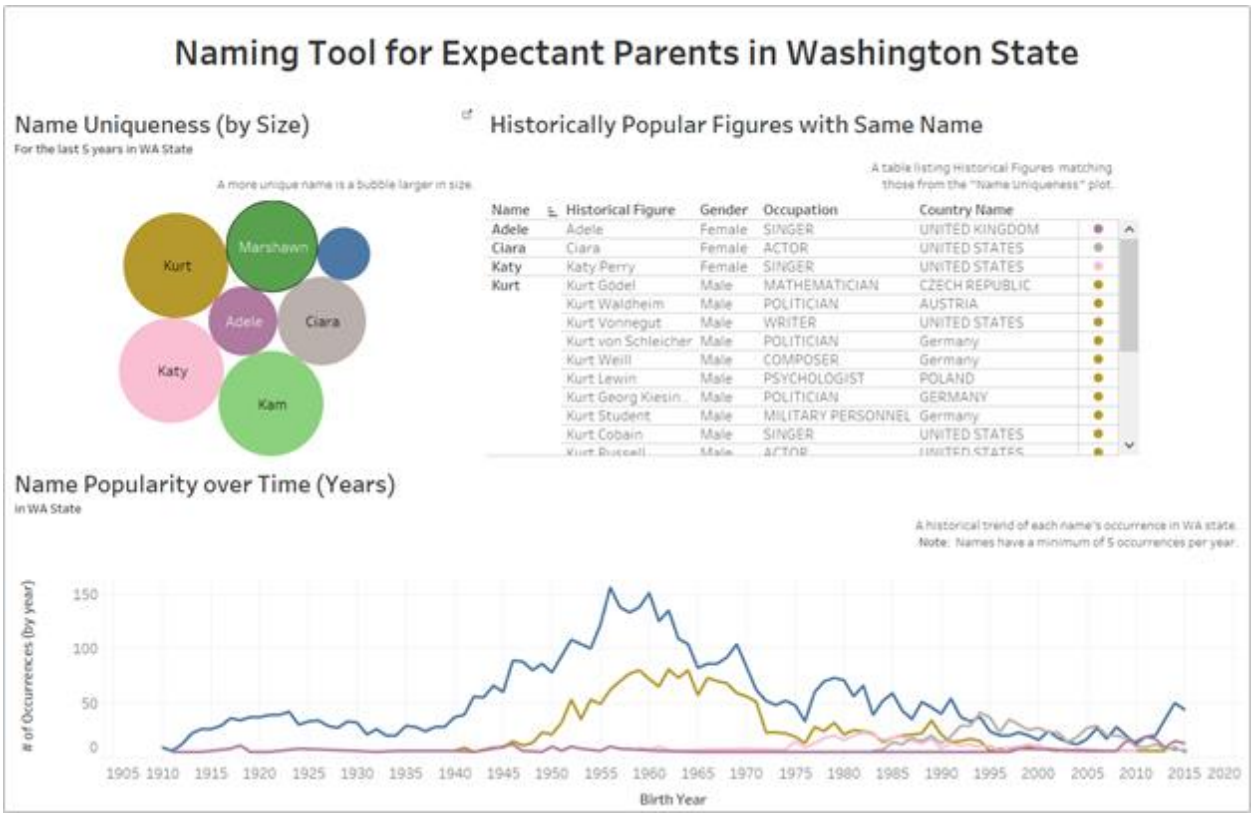


Figure 16. A zoom on the central visualization section of the naming tool dashboard.

Table A. Name Uniqueness Bubble Plot

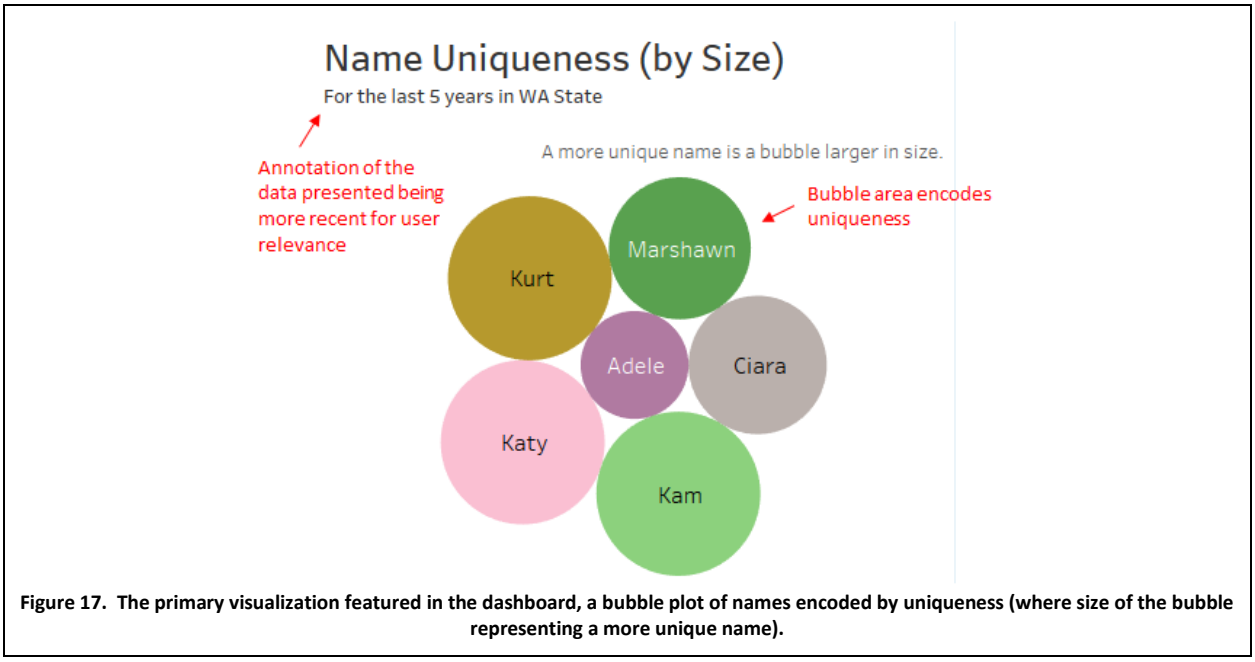


Figure 17. The primary visualization featured in the dashboard, a bubble plot of names encoded by uniqueness (where size of the bubble representing a more unique name).

Description	A bubble plot showing "Name Uniqueness" as a function of size – the larger the bubble, the more unique the name. This data is an average over the last 5 years in the WA state record (from 2010-2015).	
Dimensions and Measures	**Name Uniqueness (Quantitative)** – A value for name uniqueness, calculated (by inverting the "# of Occurrences") and averaged over the last 5 years of data in the WA state records from 2010-2015. **Name (Nominal)** – The first name of a baby born in the time frame between 2010-2015.	
Encodings	The bubble plot (**see Figure 17**) encodes "Name Uniqueness" as a function of the bubble's *area*. The greater the size of the bubble, the more unique a name is in the WA state records. Name, a nominal value, is encoded on each separate bubble with a text label.	
Expressiveness	In respect to Bertin's visual variables[2], color and size are the retinal variables are used in this bubble plot. Color for bubbles for each name to distinguish the qualitative values and the size of the bubbles provide quantitative comparison for "Name Uniqueness".	
Effectiveness	In Mackinlay's effectiveness rankings, Area is generally a mid-tier encoding in terms of rank[4]. However, for the context of this tool, where users need a holistic view of all names (bubbles) and need to quickly compare names by uniqueness, we found area to be the most effective. Furthermore, we determined users did not need a precise quantitative value for "Name Uniqueness", so the bubble plot was chosen to optimize the user's needs. It could be argued that a bar plot could be potentially more effective encoding than a bar graph. And we do agree in certain circumstances. Since the data view is dynamic and there could be lots of bubbles to display, the bar plot would fail to show the holistic view. However, in the case of bubbles that are hard to compare in respect to each other, a bar plot may show such granularity in the quantitative value much better. Overall, given the majority use cases require a holistic view and a lack of need for the user to infer the quantitative value of "Uniqueness", we selected the bubble plot over the bar graph.	
Interactions (evaluated in respect to Shneiderman's mantra)	**Overview**	Upon clearing the filters of the visualization or when reducing the large dataset, an overview is apparent to the user following the first principle of Shneiderman's mantra[5]. With this, they can easily see standout names by large bubbles, which represent a more unique name.
	Zoom	Hovering over a bubble will allow the user to focus in on details on a name. This is done through brushing and linking.
	Filter	Clicking on the bubble itself acts as a filter and displays only that selected name for clear viewing across the dashboard. In addition, applying filters for "Gender", "First Letter", "Length of Name", "Name Uniqueness" and a specific "Name" help the user navigate from an overview to a desired output of fewer, specific bubbles.
	Details-on-demand	Hovering over the bubbles also provides additional details-on-demand of "Nicknames" related to the name as well as the quantitative value for the "Name Uniqueness", encoded by the bubble size.
	Relate	Brushing and linking has been incorporated when hovering over the bubble. This will highlight that specific name across the dashboard (**see Figure 18 shown at bottom**).

	History	By uploading the dashboard to Tableau Public (online), the user can undo a change by using the "Back" button, found in the bottom left corner. Also in that same area, a "reset" button is found to allow the user to restart a name search.
Graphical Excellence	In the bubble plot, we looked to adhere to Tufte's Principles of Graphical Excellence[7], which put in usability terms was to help the user identify a name of interest. We found the size of a bubble to reflect "Name Uniqueness" did this well. Since the users will be expectant	
Graphical Integrity	In adherence to graphical integrity, this visualization avoids any misconstruing of the data by avoiding non-uniform representations and distortion.[7] "Name Uniqueness" bubbles are encoded accurately in respect to their quantitative values. Additional labeling was included as a detail to let the users know the name data was for Washington state for the past 5 years.	
Use of Color	Color has been applied on the bubbles to help with distinction. Tableau's default palette has been used to optimize contrast (value). Highlighting features provide feature contrast by lowering the value of unselected bubbles. Due to the sheer number of names in the dataset, there are names with color redundancies, which can occur and admittedly can take away from optimizing contrast. In respect to borders on the visualization, per Maureen Stone's paper on "Expert Color Choices for Presenting Data"[5], light grey was used to help with visibility and frame the visualization from the rest of the dashboard and avoid distraction.	
Layout	The bubble plot was placed in the upper left quadrant of the visualization dashboard. Since the narrative flow will start from the upper left[4], the user will recognize this bubble plot directly after acknowledging the filters. It is the primary visualization in the dashboard for the user's purpose of recognizing an interesting name, so this positioning give it the appropriate level of priority.	
Labels and Legends	Labels for the "Name" are found directly on the bubbles, also serving a dual purpose as a legend for the entire visualization. The title of the plot itself helps denote what the user is looking at with "Name Uniqueness by Size" which a light grey caption to convey what size represents in respect to uniqueness.	

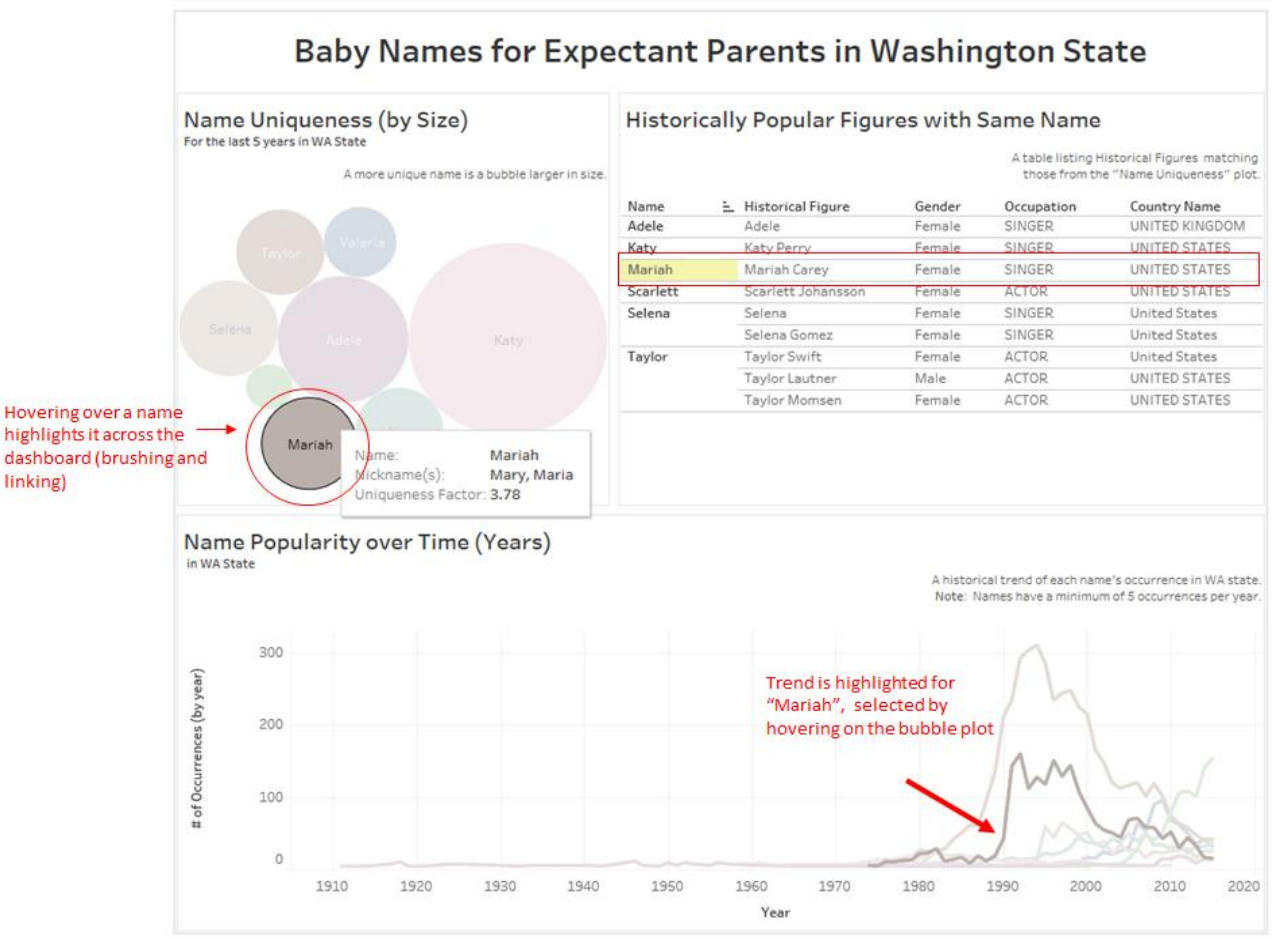


Figure 18. The interactive feature of brushing of linking among the three visualizations, increased the ability of the user to explore better and make inferences on data.

Table B. Historical Figures Table

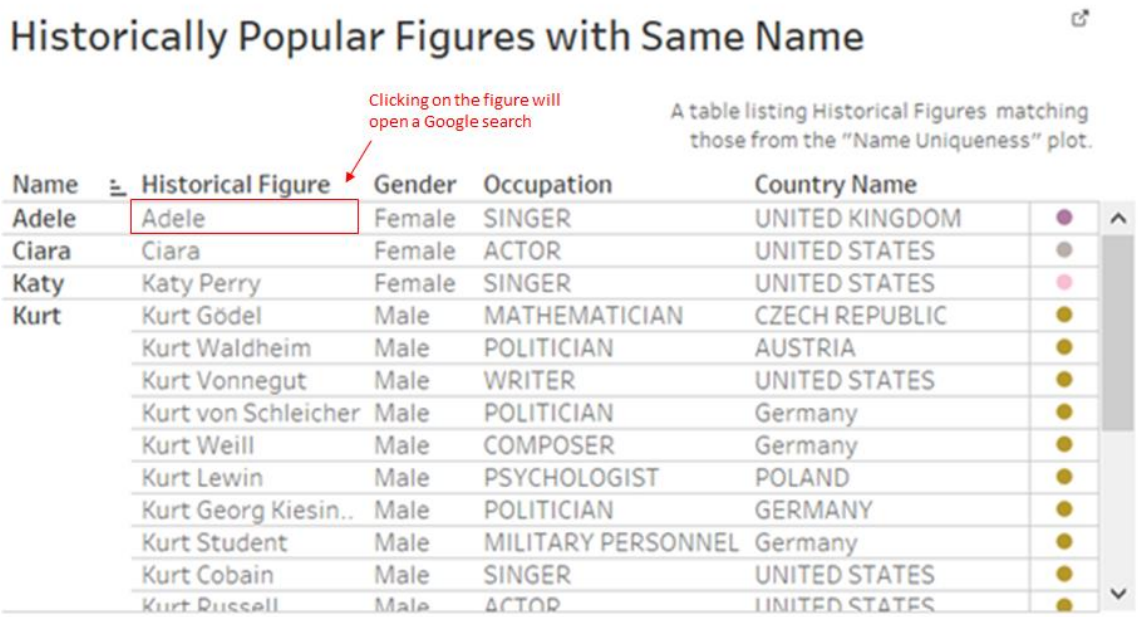


Figure 19. The table of historical figures above appears for selected names in the dashboard. Given the entirely nominal nature of this data, we determined it best to leave in that form for the user.

Description	The table above (**Figure 19**) shows any historically popular figures, which share the same name as the bubble plot. This visualization is created with aid of the MIT Pantheon dataset.	
Dimensions	**Name (Nominal)** – First name (of a baby name) **Historical Figure (Nominal)** – First and last name of a historical figure **Gender (Nominal)** - gender of the historical figure **Occupation (Nominal)** – occupation of the historical figure **Country Name (Nominal)** – native country of the historical figure	
Encodings	The data is left in tabular format since it is all nominal. The user studies show a list is more useful for this detail, given this is a supplementary feature to the bubble plot's name.	
Expressiveness	The table, while being the most basic form of visualization, takes the nominal values and organizes and displays them in the most efficient organized way for its purpose. Instead of use as a textual detail on demand, which would be more cumbersome for the user, names are listed adjacent to the bubble plot and time-series plots in the dashboard.	
Effectiveness	Nominal values are considered the least effective encodings generally, but that is taken in respect to this data being entirely nominal. One potential for encoding was the option to include and sort the names listed by "Name Uniqueness". However, the bubble plot serves that purpose and the user will naturally think in terms of alphabetical order. As a result, the names were listed from A to Z.	
Interactions	**Overview**	Without any user interactions, the name list has a scroll bar feature on the right side to allow for a user overview.

(evaluated in respect to Shneiderman's mantra)	**Zoom**	The application of zoom does not really apply on this table visualization. However, clicking on a specific "Name" does make use of brushing and linking, highlighting that name across the dashboard.
	Filter	The interactions provided with the bubble plot click feature, performs a filter on this table. It will provide the historical figures matched the name selected. Also, the general filters provided on the left side of the plot also serve as filtration for the table.
	Details-on-demand	Clicking on a "Historical Figure" will provide details-on-demand through a web browser performing a Google search of that figure.
	Relate	As mentioned, hovering on a figure's name performs a highlight of the bubble plot and time series plot.
	History	By uploading the dashboard to Tableau Public (online), the user can undo a change by using the "Back" button, found in the bottom left corner. Also in that same area, a "reset" button is found to allow the user to restart a name search.
Graphical Excellence	The table conveys minimalism in minimizing ink and space while maximizing information. It achieves the purpose of conveying the information of nominal text detail to the user to complement the bubble and time-series plots.	
Graphical Integrity	Since the table is the most basic of data representations, it naturally presents the data accurately and without any complex encodings, which could result in distortion of data. Clear labeling is in place for the headers to organize and represent the data encoding.	
Use of Color	Black text, white background and light grey borders are details, which convey the minimalist approach. In respect to borders on the visualization, per Maureen Stone's paper on "Expert Color Choices for Presenting Data"[6], light grey was used to help with visibility and frame the visualization from the rest of the dashboard and not to distract.	
Layout	Since the narrative flow is left to right, and top to bottom, the table has been placed on the right of the bubble. It helps denote its secondary importance to the bubble plot.	
Labels and Legends	Clear labeling is in place for the headers to organize and represent the data encoding.	

Table C. Time-Series plot for Name Popularity over Time

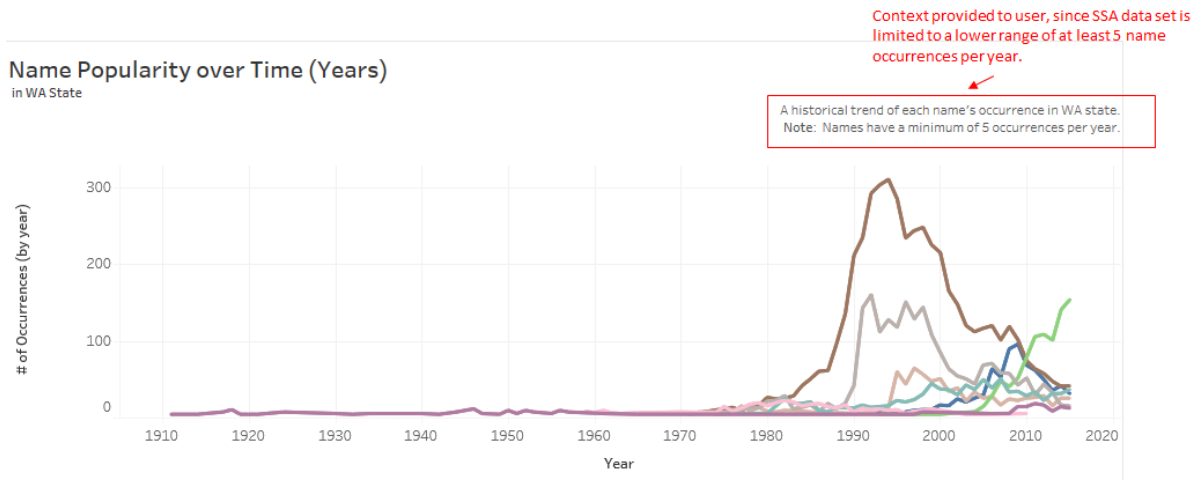


Figure 20. The time-series plot for trends of selected names appears at the bottom of the dashboard. While it provides no listed names, in the dashboard setting the user can use the bubble plot as a very clear legend to match by color. In addition, hovering over the trend line will provide this information.

Description	A time plot showing the historical trend of name occurrences in Washington state.	
Dimensions	# of Occurrences (Quantitative)	
Encodings	A time-series is used to encode the "# of Occurrences" on the y-axis versus the year on the x-axis.	
Expressiveness	The encoding of the bivariate plot for the two quantitative values, "# of Occurrences" and "Year" is done by use of position and connected lines (shown from year to year). This makes use of two of the most expressive visual variables from Bertin's list of visual variables[2].	
Effectiveness	The use of position and lines for visual mappings are accurate and appropriate use of two quantitative values. By Mackinlay's Effectiveness rankings for encodings, these are two of the most precise encodings.[4]	
Interactions (evaluated in respect to Shneiderman's mantra)	**Overview**	Before any applied filtration, the time-series plot shows trends for several names in an overview where the user can pick up interesting trends by hovering.
	Zoom	The hover feature will make use of brush and linking[5], highlighting a specific trend in the time-series plot as well as the bubble plot and historical figures table.
	Filter	The interactions provided with the bubble plot click feature, acts as a filter on this table. It will provide the historical figures matched the name selected. Also, the general filters provided on the left side of the plot also serve as filtration for the table.

	Details-on-demand	Hovering over a trend line for a name will provide additional details-on-demand, such as “Name”, Birth Year” and “# of Occurrences” of the name.
	Relate	Hovering on a trend line (representing a specific name) performs a highlight of the name in both the bubble plot and historical figures table.
	History	By uploading the dashboard to Tableau Public (online), the user can undo a change by using the “Back” button, found in the bottom left corner. Also in that same area, a “reset” button is found to allow the user to restart a name search.
Graphical Excellence	The time series plot is in line with Tufte’s Principles of Graphical Excellence, conveying information in an efficient use of ink and space[7].	
Graphical Integrity	The time series plot maintains graphical integrity by using a very accurate encoding for plotting the “# of Occurrences” and “Year” in a bivariate plot. Legends are included on the axis to define context and units and there are no complex encodings, which could result in data distortion.	
Use of Color	In the time series plot, Tableau’s default palette has been used to optimize contrast (value) between trend lines. Like the bubble plot, highlighting features provide feature contrast by lowering the value of unselected trend lines. Due to the sheer number of names in the dataset, there are names with color redundancies, which can occur and admittedly can take away from optimizing contrast. In respect to borders on the visualization, per Maureen Stone’s paper on “Expert Color Choices for Presenting Data”[4], light grey was used to help with visibility and frame the visualization from the rest of the dashboard and not to distract.	
Layout	The time-series plot is placed in the third and fourth quadrant space below the bubble plot and historical figures table. This position gives the plot a secondary importance to the bubble plot (the primary name identifier for the user) while taking up enough space to emphasize its secondary value. Positioning was more a matter of practicality in this case, as placing the historical figures table at the bottom would be a poor use of space. This would result in the time series plot.	
Labels and Legends	The time series plot is labeled with “# of Occurrences” on the y-axis versus the year on the x-axis to provide clarity on what is to be inferred from the visualization. A light grey caption in the top right helps the user understand the data is for Washington State and the minimum value for name occurrences is a value of 5 (an inherent property of the dataset).	

More on Filters and Interactive Aids

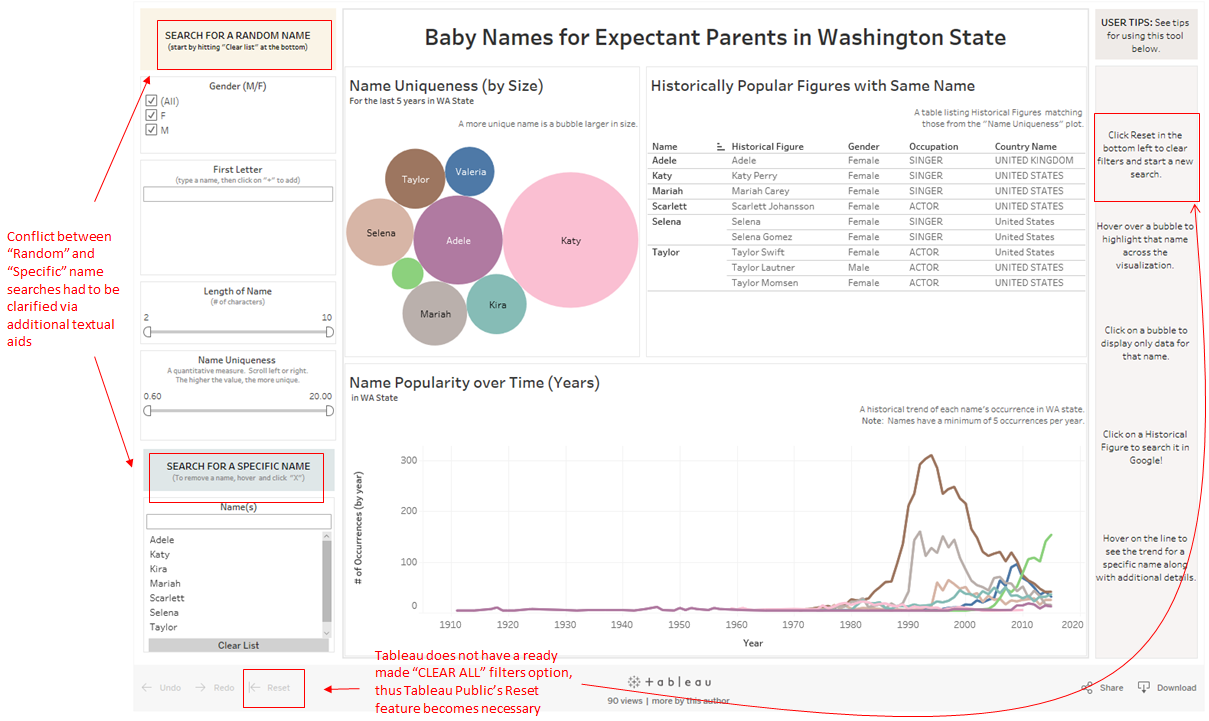


Figure 21. Since Tableau does not provide a "Clear All" filters function that can be integrated directly into the dashboard, extra measures were taken to help the user understand how to clear filters and to make use of the "Reset" button (only available when a workbook is uploaded to Tableau Public)

Due to the varying data outputs in the visualization, given the needs of different users in their name search, several modes of filtration were necessary to include in the visualization. To fit within the narrative flow of the layout, these filters were placed on the left-hand side, as the user will naturally work from left to right. When viewing from top to bottom, the filters cover two search purposes - a "Search for a Random Name" and a "Search for a Specific Name." These are tasks that need to be distinguished separately (**see Figure 21**), to prevent the user from unnecessary confusion by mixing and matching. In order to do this, colors are used to provide contrast to help the titles for these separate search sections stand out.

Furthermore, on the subject of color, a subtle application of grey color for the border was applied to help distinguish the filters as separate from the dashboard. This was also done for each filter. A very light grey was chosen as to not distract from the visualization itself.

One important user feature was the ability to reset (clear) filters applied to restart a search. Tableau has no direct feature to do this and filters generically need to be cleared one by one in the top right corner. However, in Tableau Public, when a visualization is published, there is a "Reset" button found in the bottom left-hand corner.

Due to the heavy need of user interactivity to customize a name search, explanations would be needed to clarify how to properly use the visualization dashboard. To aid in this, the right side of the dashboard contains text

detailing “User Tips” in how to maximize value out of the dashboard. This was placed on the right side to de-emphasize while still being fixed and apparent to the user if needed.

EVALUATION OF VISUALIZATION SUCCESS

SUMMARY

As we stated at the beginning of the paper, our goal was to create a visualization that would enable expectant parents the ability to search for potential suitable names for their baby. After a process of performing research to evaluate users’ needs, formulating designs and strengthening our visualizations through user testing in multiple development cycles, we were successful in creating a visualization with the following features:

- The ability for a user to filter information in the visualization for both a random name search (filters by gender, first letter, name length, and name uniqueness) and a specific name (searching a name directly)
- The ability for a user to see a visualization of the most unique baby names for the last 5 years in Washington State
- The ability for a user to identify selected names to historically popular figures and their occupations in the same viewed visualization (a supplement to identify “success” and tie to “culture”)
- Ability to see name popularity throughout the years for more in-depth analysis (by looking at trends and patterns in more detail)

The above features provide a tool to our users that can assist them throughout the name-selection process for their baby. Throughout our design process, we found that while innovation is key in creating a novel visualization it could not come at the expense of the users’ needs. Our users initially defined goals for us were to provide a search tool to find names that are unique, culturally significant, and indicate success. While we initially focused fully on all three of our users’ needs, our first round of helping informed us it would help to narrow the scope to focus on the need for “uniqueness” with secondary emphasis on “success” and “culture” in order to improve the usability of the tool. After making this decision to pivot, we immediately started seeing more positive results in our usability testing. Our users were able to use our tool to search through baby names with their partner and found our tool to be a great resource for them. Despite the narrowing of scope, we feel like we were able to touch upon the three categories that the users showed a demand. Our project helped set a foundation for future opportunities to continue further research and develop improvements to meet the user’s ideal needs.

HOLISTIC ASSESSMENT

The following table provides a holistic assessment of successes of the visualization project and the areas where we think we could have improved:

Category	Successes	Improvement Opportunities
User Needs	• Narrowed scope, focused on uniqueness, secondary priority on historical figures as supplement	• Find a way to put more emphasis user desires for success and cultural connection

	• Provided user random search filters (gender, letter, length) • Provide the user a specific name search filter	
Visualization Principles	• Application of best known "infovisualization" principles • Simple visuals with effective encodings • User has minimal cognitive effort • Less emphasis on novelty (eliminated tree map) • Lots of data, Minimize ink where possible	• Bubble plot vs. bar plot - assess possibility to have options for both for an optimal encoding • With additional time, improve Tableau layout issues by publishing on the web or evaluating option of using D3 • Account for color redundancies which can confuse users (an issue due to the vast quantity of names)
Interactions	• Exploration through multiple visualizations and filters • Brushing and linking • Followed Shneiderman's mantra • Provide text aids for interactions	• Improve interactive features • Further improve integration of separate needs for random and specific name searches • Improve Clear Filters interaction In Tableau (not provided by default) • Find a way to notify users (pop-up) why where is no viewable data (i.e. historical figures table) with certain filters
Dataset	• U.S. Baby Names - massive data set and user potential • MIT Pantheon Dataset - added breadth of scope with a connection to historically popular figures	• Find a complementary dataset to better identify success and culture • Increase dataset scope to beyond Washington State to include all 50 U.S. States (requires overcoming cloud storage limitations) • Minor data cleaning opens with accent marks not translating properly in Tableau and "Country Name" values are listed in consistent formatting (non-capital letters)

CHALLENGES AND FUTURE IMPROVEMENTS

Below are further details of the primary challenges we faced in this project along with future improvements to be pursued to help address them.

A Dataset for Success and Culture

In our project scope, we ran into the aforementioned issue of fulfilling all the users' desires with a visualization that could effectively identify unique baby names and at the same time help predict "success" and an association to "culture". While the MIT Pantheon dataset was very robust supplement to the U.S. SSA dataset for baby names, for our applications the metric it provided in the form of Historical Popularity Index (HPI) was difficult for some users to decipher. For other users, it was irrelevant as a metric to identify success. Given Pantheon was one of the more meaningful open-source datasets we could find on "success" and "culture", it showed us how hard it is to quantify and identify "success", given it is something which has different meanings to different users.
For the users who could associate "success" to a historical figure/occupation they admired, then this tool could serve to fulfill that need. If that figure came from a country where the user had a connection culturally, then "culture" could be fulfilled as well. It is likely that more testing and increased sample size would help more definitively determine how effective these applications are in helping our user base.

In the future, we would like to do further research to try to find datasets that would encode "success" of a baby name and/or "culture" better. Washington state and the entire United States, being the melting pot it is, really values the ability to see cultural background for a name as well as the "success" of it. This would also add more quantitative variables to our visualization, expanding our visualization options.

Tableau Layout and Filter Reset Limitations

While it must be said that Tableau was excellent in enabling us to meet the scope and timeline of our project, it did provide some challenges. Notable issues we encountered ranged from layout distortion when uploading our dashboard online to Tableau Public along with the inability to easily integrate a universal clear/reset filter that applied to the entire dashboard.

The issues with layout could be mitigated through usage of the "Automatic" display setting to auto-adjust the dashboard size depending on the display. However, depending on the resolution or display type (whether desktop or laptop, Apple or Windows, projection or monitor), the final output in Tableau Public could still be distorted. This was particularly a problem with our textual aids partially obstructed or illegible on filters or the "User Aids" sections.

Given the high-dependency of our name search tool on filters, integrating a clear/reset filters button into the dashboard itself was another difficulty. We found workarounds by improving layout and by making note in our "User Aids" that there is a default "Reset" button found in Tableau Public, though our desire was for it to be more prominent to our users and on the dashboard itself.

In our future revisions, we would like to figure out a more seamless experience for user to be able apply filters seamlessly and view the visualization without layout distortion.

Lack of cloud server storage

In our project, the U.S. SSA dataset was trimmed down to just WA state, which served beneficial in the reduction of scope. This was especially justified considering all of our usability testing and immediate user base was with residents of Washington state. Reduction also helped removed the need for users to sort through even more data clutter and removed an unnecessary filter.

However, in a more forward-looking view, we did want our project scope to include data for names across the entire United States. The primary issue was lack of cloud storage for our dataset, which reached 3.5 GB once the U.S. SSA dataset and MIT Pantheon datasets were merged. As a result of this size, we could connect easily work with the dataset in Tableau remotely or even upload the dashboard into Tableau Public due to size limitations. At the moment, it forced limitations in accessibility and application outside of Washington State.

In a future revision, we are looking to having cloud storage capabilities to extend the scope of the visualization build beyond Washington State.

ACKNOWLEDGEMENTS

We would like to acknowledge and thank the people that had a huge influence on this project. First, we would like to thank our Data Visualization course professors from the University of Washington, Brock Craft and Robin Martin-Emerson, who provided us guidance and constructive feedback throughout the whole process. Second, our user participants, whose feedback during helped us re-think our visualization as a whole and make an emphasis on usability, less on novelty. This helped improve our visualization to something they would actually use. Finally, we'd like to thank Amanda Pype, who not only taught us a lot of Tableau tricks in a single session, but also provided us guidance when we ran into roadblocks.

REFERENCES

1. Perin, C., Dragicevic, P., & Fekete, J. (2014, December). Revisiting Bertin Matrices: New Interactions for Crafting Tabular Visualizations. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 2082-2091. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6875988

2. Roth, R. (2015). Visual Variables. 1-12. Retrieved from https://www.geography.wisc.edu/faculty/roth/publications/Roth_2015_EG.pdf

3. Mackinlay, J. (n.d.). Designing Great Visualizations. Retrieved December 8, 2016, from https://www.tableau.com/sites/default/files/media/designing-great-visualizations.pdf

4. Kosara, R., & Mackinlay, J. (2013). Storytelling: The Next Step for Visualization. Computer, 46(5), 44-50. doi:10.1109/mc.2013.36

5. Shneiderman, B. (1996) "The eyes have it: A task by data type taxonomy for information visualizations." *Visual Languages., IEEE Symposium.*

6. Stone, M. (n.d.). Expert Color Choices for Presenting Data. Retrieved October 30, 2016, from http://www.stonesc.com/pubs/Expert Color Choices.pdf

7. Tufte, E. R. (2004). The visual display of quantitative information. Cheshire, CT: Graphics Press.

8. Yu, A. Z., et al. (2016). Pantheon 1.0, a manually verified dataset of globally famous biographies. Scientific Data2:150075. doi: 10.1038/sdata.2015.75