Produce Availability in the U.S.

UC Berkeley, School of Information

INFO 247

Spring 2014

Katey Basye & Anna Swigart

Project Goals

Most Americans have little understanding of where their food comes from, or could come from, when buying fresh produce at the supermarket. For consumers interested in eating produce that is at its peak in freshness and quality, there is no simple way to understand whether the eggplant or apples on their shopping list could have been grown on a farm nearby. Some regions in the United States, like the San Francisco Bay Area, seem to have a strong focus on making local foods available to consumers. When presented with data about the variety and quantity of fruits, vegetables, nuts, and berries grown in California, it becomes clear how such a local food ecosystem is able to thrive here. But how do other places in the country compare?

Our ultimate goal with this visualization is to provide users with perspective that can translate into more informed choices in food purchases. By taking advantage of produce items that are grown near them when possible, consumers can support their local economies while enjoying freshly picked ripe foods that are at their peak in flavor and nutrition. When traveling or relocating to other areas, our visualization can help users understand how fresh produce quality may differ from the county they are coming from. For example, Southern California counties may have an abundance of fresh citrus and avocados, but the west bank of Michigan is bursting with blueberry farms.

Supported Tasks

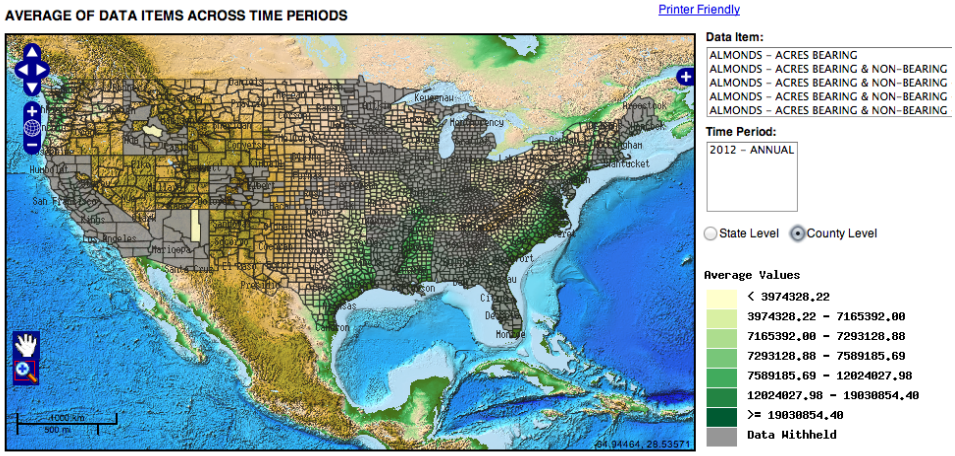
In this visualization, we provide users with a way to explore the types of produce items that are grown near where they live, compared to any other county in the United States. Users can also search for their favorite produce items and see 1) a choropleth United States map showing a representation of the number of farms in each county that produce that crop, and 2) a list of the top 10 counties that produce that crop. On the map, users can use the mouse to hover over any county and see a tooltip that displays the county and state name along with the farm count for the chosen crop. By clicking on the county in the map, it can be added it to a list of counties to be compared. County names from this list can be dragged and dropped into one of two panes, allowing for pairwise county comparisons. Each comparison pane includes a sorted bar chart showing the number of farms that produce each crop there, for every crop that is grown by at least one farm. Users can also drag counties from the “Top 10” list directly into the comparison dock, or search for specific counties by name.

Related Work

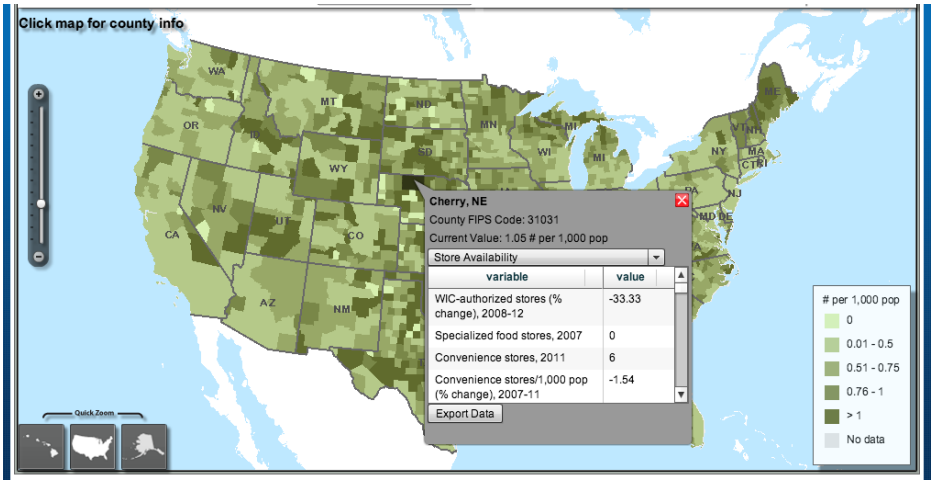
In researching for this project, we first looked at existing attempts to communicate food access and availability. We found that most resources in this category rely on maps of the U.S. to visualize their data, something that we agreed could be an effective way to highlight regional patterns for such large dataset. However, we found that the current map visualizations available in this area either suffered from inaccuracy stemming from oversimplification or else failed to provide any sense of summary narrative. Many suffered from unsatisfactory scope—usually too broad or overwhelming to be meaningful. In terms of performance and usability, these food maps were generally slow, small, and frustrating. Furthermore, we found that such food maps either afforded such exploration as zooming, mouseovers, and click events—or such comparisons like data switches on the map—but not both. All of these investigations lead us to the decision that our final product would be explorable and offer both a general overview of food availability in the U.S. and granular comparison between specific places.

The following two maps from the USDA attempt to visualize the data we used for this project:

This map is from the USDA Agricultural Census. While it offers many appealing choices—a plethora of foods, the ability to aggregate by state or county levels, and, in theory, time periods—it has several issues with usability and visualization. Most obviously, the black lines and the topographical background detract from the choropleth and defeat its purpose.



This USDA Food Environment Atlas does not allow comparing between counties or across indicators, although the choropleth does a better job of giving users an impression of quantity.

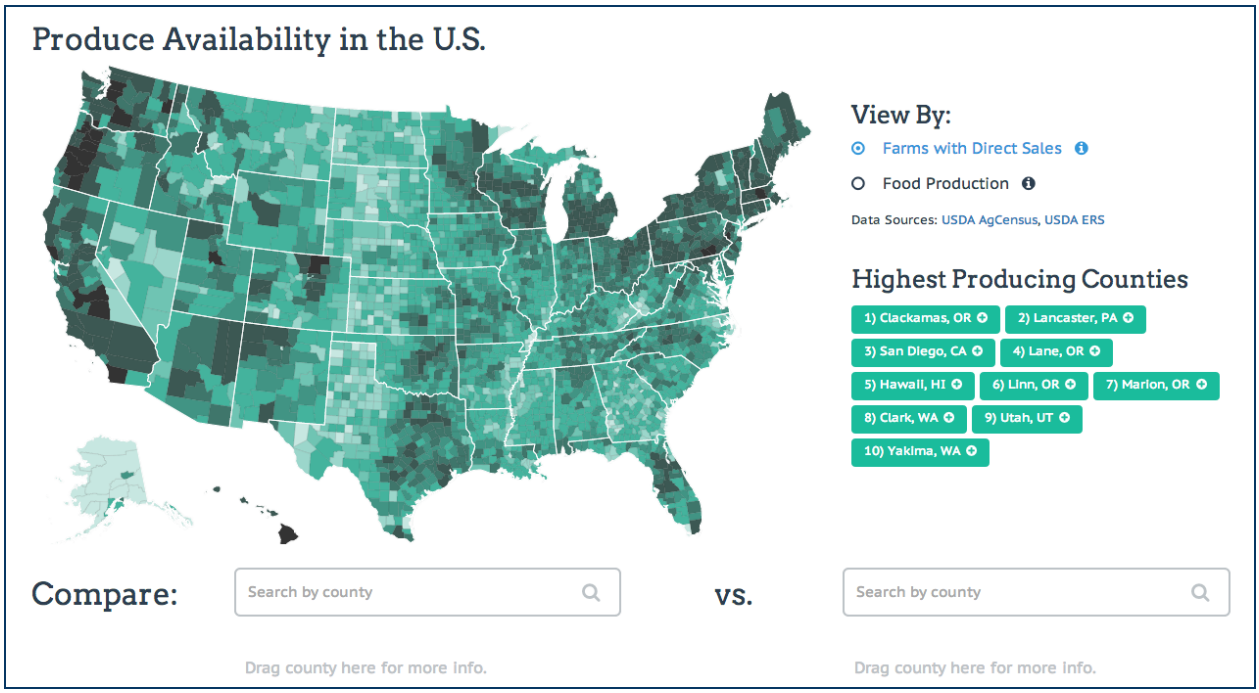


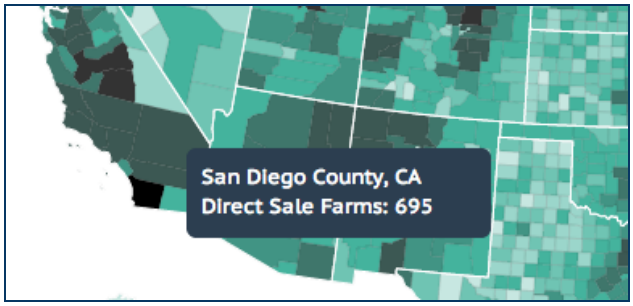
This Seasonal Ingredient Map from epicurious has a singular white dot representing an entire state and no data citation. The state lines are arbitrary in this visualization, since seasonality varies by region and climate, not by U.S. state, and generally, users do not get any sense of whether the seasonal food is actually available, “fresh in your area.”

Our Project

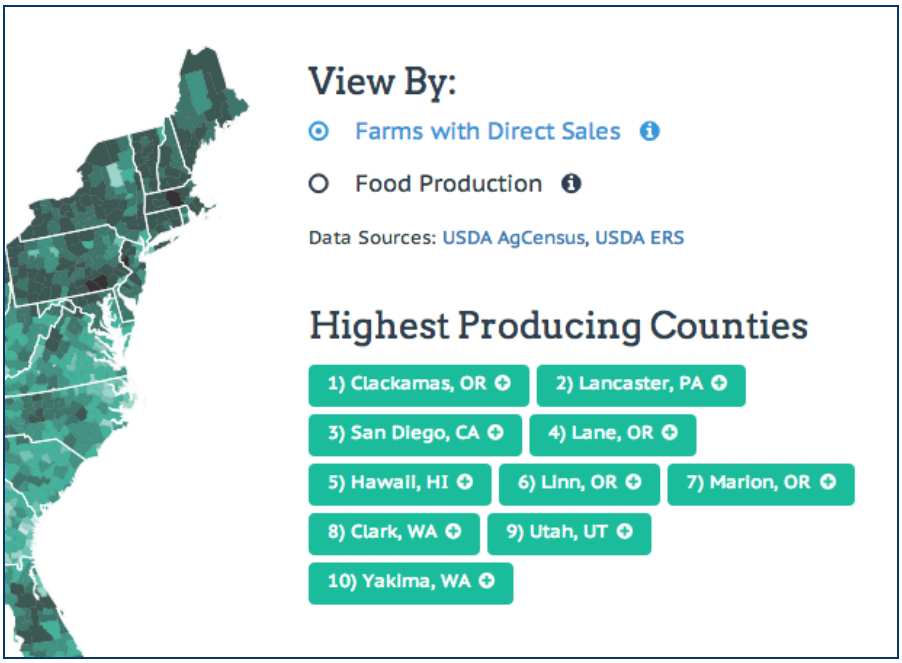
The interface we have created uses a choropleth map of the United States as its central figure, which displays data for 3,154 counties by modulating the fill color of each county based on a chosen measure.

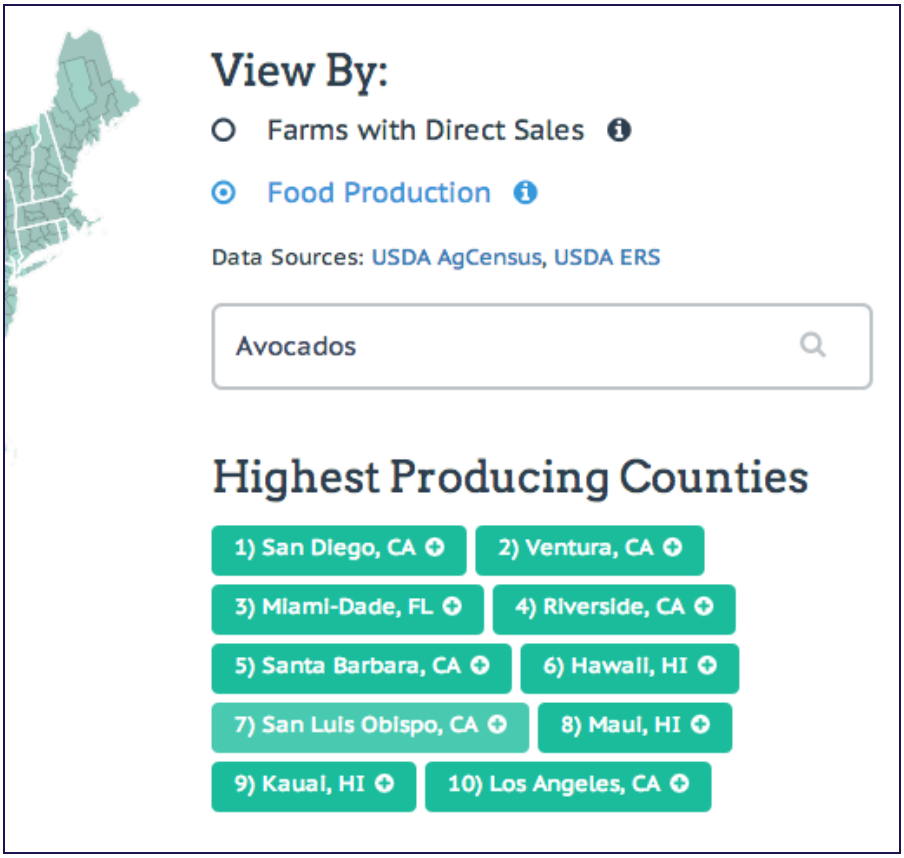
The default measure gives an overall impression of local food access by indicating the number of farms in each county that engage in selling goods directly to consumers (avenues such as farmer's markets, roadside stands, and Community Supported Agriculture (CSAs)).



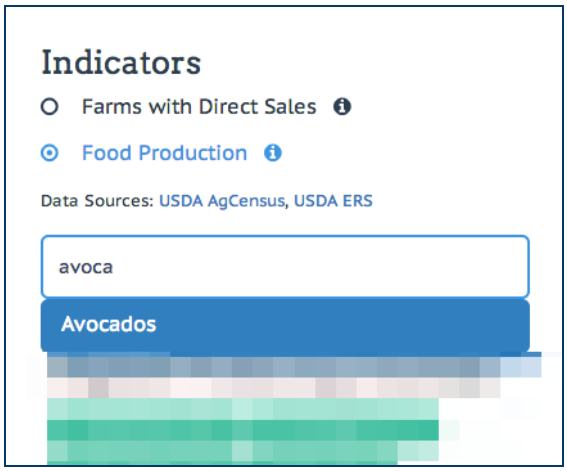


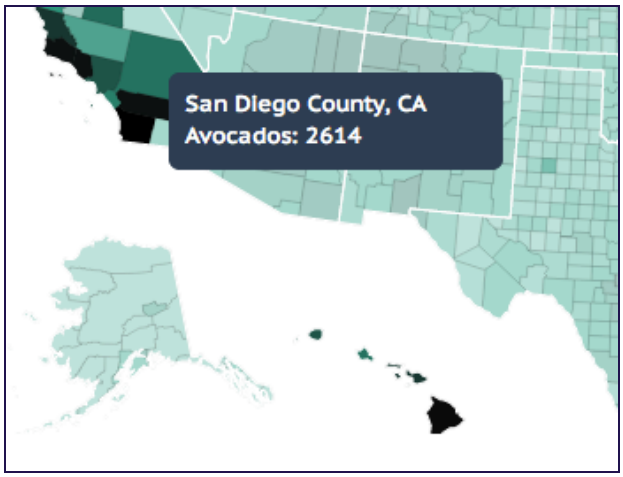
Users are also able to see the Top 10 producing counties for each measure.



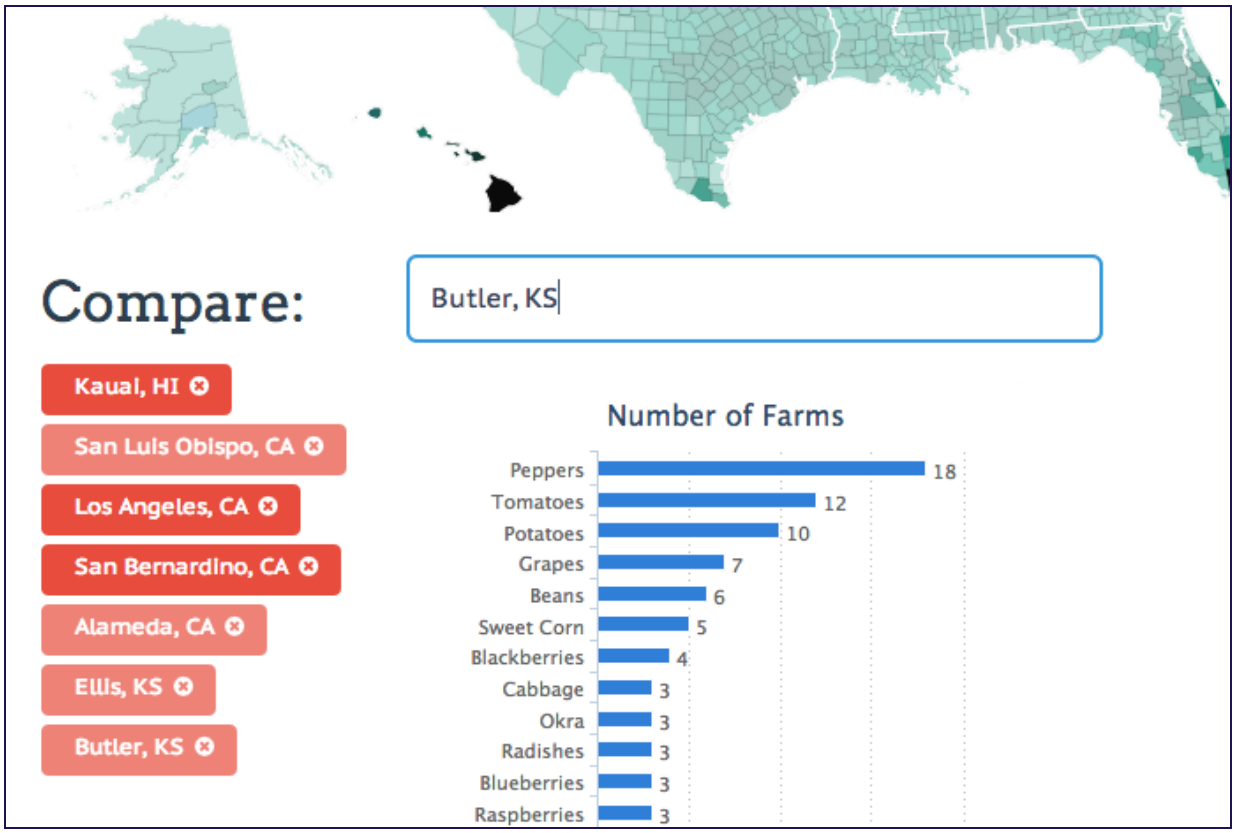


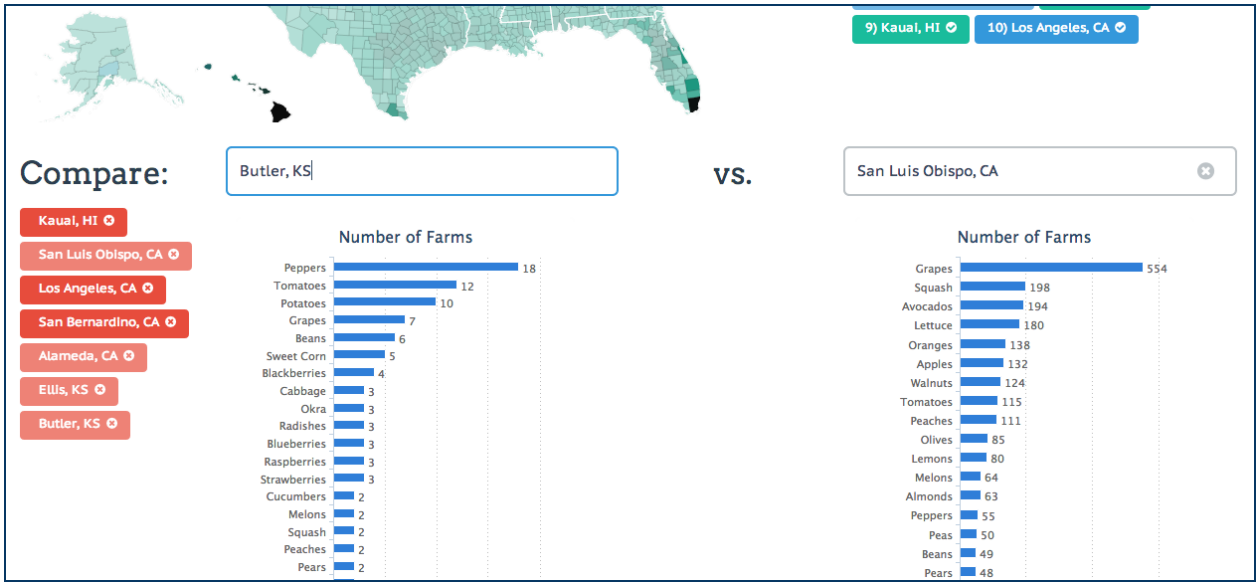
Users are then able to search for specific produce items to understand the national distribution of where that particular item is grown for fresh market, with the option to hover over any county on the map to see an exact number of farms.





Finally, the county comparison portion of the tool allows for direct comparisons between two counties at a time, across a sorted chart of produce items that are grown in each county.





Data

We ultimately used two datasets for this project:

1. The Food Access Research Atlas, provided by the United States Department of Agriculture, Economic Research Service (USDA, ERS)
2. The USDA 2012 Census of Agriculture

Tools

- For initial exploration to identify possible datasets and attributes to incorporate, we used Tableau.
- For the data wrangling portion of this project, we used Python, embedded in iPython notebooks and pandas dataframes.
- The final tool was created with HTML, CSS, and Javascript. The D3.js library was used to construct the map visualization, and highcharts.js was used for the bar charts. We also harnessed jQuery for the interaction design.

Our Process

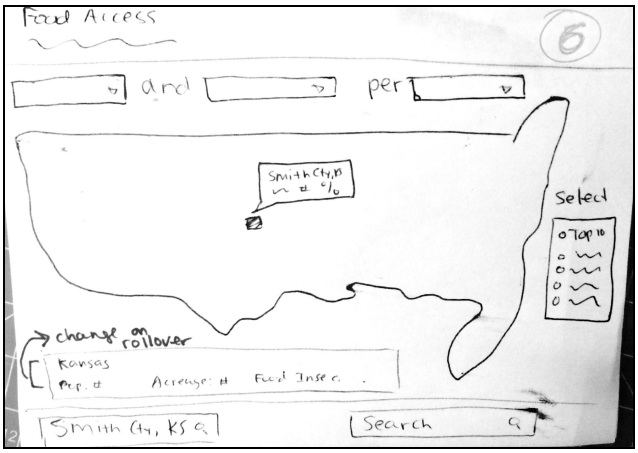
1. **Identify goals**. We decided early on that the primary goal of this visualization would be to give consumers insight that would provide both perspective and actionable information that could be used to make better healthy food decisions.

2. **Find data that aligns with our goals**. We scoured numerous government health and agriculture websites to identify open datasets that could be used to address our goals. We were drawn to the granularity of the USDA ERS dataset and experimented with arranging and combining many of its variables before ultimately deciding to focus on a small subset of it. It was important to limit the data we included to a scope that our users would be able to interpret and connect to in meaningful ways (see Food Environmental Atlas screenshot in *Related Work* above). The USDA Census measures by produce type, combined with the geographical nature of the data collected, created a personal space for users to connect what they were eating to where they lived.

3. **Data wrangling**. The data wrangling completed and decisions made in order to convert and merge the produce item-wise USDA Census data with the USDA ERS data are documented in the following iPython notebooks:

- Compiling Vegetable Data
- Compiling Fruit and Tree Nut Data
- Compiling Berry Data
- Compiling Master Data

4. **Plan visualization interactions and prototype layouts**. We considered many different interaction models and layouts before arriving at our current design.



5. **Program charts, layout, and supporting interactions**. During this step we put our web development tools to work and continued to define the details of our design.

6. **Evaluate the visualization with interested users**. We shared our visualization with others and evaluated how well they understood the purpose of the tool and whether it provided new and useful insights with ease.

Results and Feedback

The cross critique in class was the first feedback we received on our project. Occurring at an early stage in the process, we were somewhat overwhelmed by the amount of data we had available to us and were struggling to clearly define a narrative within our overall goals. At the time, we had a begun to develop our vision for the current layout—a choropleth map on the top that gives users a general overview and a space in the bottom for granular comparison—and came equipped with several paper sketches and an interactive map prototype to communicate our ideas. We found the pair-to-pair discussion, as well as the subsequent feedback and questions from our classmates, to be helpful because we learned more about what potential

users wanted to learn and which measures they found most interesting or confusing. From this, we were able to narrow down our scope and firm up our narrative.

Through a series of unstructured usability tests with interested consumers, we were later able to evaluate a more complete version of our visualization. The feedback we got confirmed our hypothesis that county-wise, product-specific data was of high interest to users. The map was effective in revealing robust regional patterns in food production, especially for fruit and tree nuts. We found that users enjoyed comparing the counties where they have lived in the past and reasoning out loud about why some particular produce item might be thriving in one place but not another. Many users took note of the frequency with which California counties showed up in the "Top 10" list, sometimes dominating all of it. These remarks reflect a gained awareness of possible factors that may underlie the strong local produce culture that is prevalent in this state. Overall, these results suggest that users were generally successful in generating new insights and engaging in self-reflection while interacting with our visualization.

We also received valuable feedback about portions of the design and information that needed improvement in order to fully accomplish our goals. For example, almost every user we talked to wanted to know where the data we used was from. Since the testing was conducted in person, we were able to explain verbally. To the current design, we have added prominent links to the original data sources for users who might access this visualization online. The version of the design we tested at this stage also bore the title "Local Food Access in the US". We encountered many questions about what this really meant, and ultimately changed our title to be more specific about the purpose of the tool. Several interaction details regarding how counties are selected for comparison have also been altered based on user observations and feedback. For example, county names are now draggable from either the Top 10 list or the Compare dock since there is a different comparison pane that is close to each of them. We plan to later integrate additional produce comparison features to better accommodate users who sought a more fluid interaction between the produce filter and county comparison portions of the tool.

Links

Github: https://github.com/annaswigart/local-food-access

Demo: http://people.ischool.berkeley.edu/~agswigart/local-food-access/

Contribution

	Data Wrangling	**Charts**	**Layout/Styling**	**Interactions**
Anna	90%	80%	10%	10%
Katey	10%	20%	90%	90%