Introduction

Welcome to Menu Journeys! At the early stages of this project, we were intently discussing our shared passions and we centered on a theme that resonated strongly in the team: FOOD. There were many ways we could have gone with this theme, and through our research we stumbled upon an interesting data set that could give us perspective into the history of restaurant dining in the United States.

The New York Public Library has an interesting archive of 17,544 restaurant menus containing menu data ranging from the mid-19th century to 2008. In April 2011, the NYPL opened up image data from the archive to the general public, who performed crowdsourced transcription of the menu data to enable further research by educators and academics. We examined the transcribed data and set out to visualize it to help us learn more about the menus included in the archive and understand if there was anything we could glean from under two centuries worth of restaurant data.

Project Goals

Goal 1: Provide an overview of the data inside NYPL's "What's On The Menu?" Archive

The primary goal of the project was to help future researchers or students understand what data was contained in the restaurant menu archive, to help them have a better grasp of the distribution and scope of the information prior to them undertaking efforts to identify trends or correlations across different variables contained in the data set. To achieve this, we set out to create a website that shows the user how we analyzed, clustered, and wrangled the hand-transcribed menu data.

Selected Approach:

Create an infographic and Highcharts section that provides a bird's eye view of what is in the archive and how we created clusters that we grouped the data into to help with our analysis.

Goal 2: Visualize clustered menu data

The secondary goal after laying out the data was to clean noisy data (incorrectly transcribed, non-US related, etc.), organize the remaining information into clusters and look into what we can possibly visualize in the selected data for the top 25 curated dish categories from the data set.

Selected Approach:

The menu data contained several interesting metrics such as: date appeared, location, event, normalized position on the menu, original price, and page number. From these metrics, we had to decide what was most feasible and meaningful to visualize.

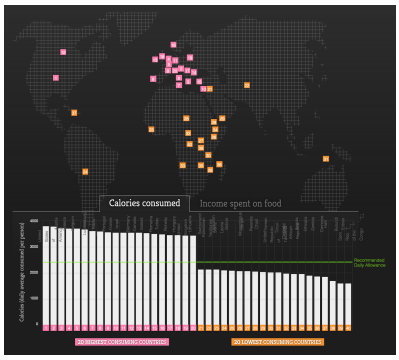
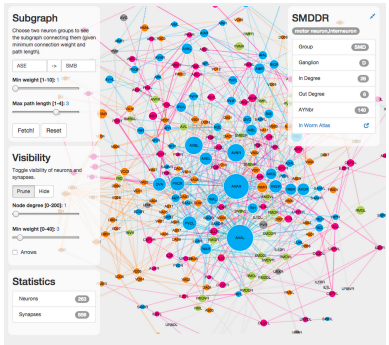
Our visualizations will enable the user to perform the following tasks:

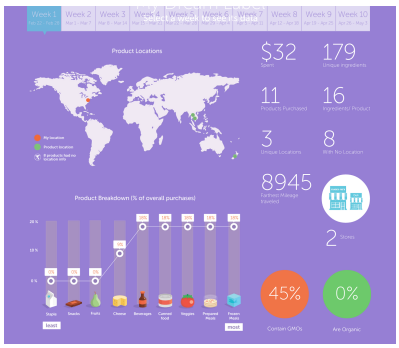
- Explore the number of menus and dishes in the original NYPL archive
- Interact with the different levels of clustering for the visualization
- Visualize trends across time such as

    - dish price
    - dish location on the menu
- Browse and find locations that serve these dishes in present-day San Francisco

Related Work

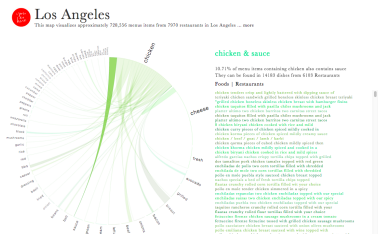
In conducting research for this project, we first looked at existing attempts to communicate food consumption related to socio-economic factors and/or food trends over time. We then looked various ways to visualize the data that was provided in the archive such as force-directed tree graphs, co-occurrence matrices, and tree maps.

Visualization Resources	
	**Visualizing the World's Food Consumption** http://www.foodservicewarehouse.com/calorie-viz/ This is a visualization of calories consumed on food and income spent on food by country. Our visualization is not a visualization of food by country but rather over time, but this visualization also relates food consumption with income, which we also hope to do by bringing in CPI data and normalizing prices.
	**C. Elegans** https://synergenz.github.io/elegans.html This is an exploration tool for visualizing the the connectome of the worm's neural network. The d3.js visualization "fetches either the whole network or a subgraph and displays it using a force-directed layout". We were inspired by the force-directed graph and its ability to show the relationship between different nodes as well as filter for and highlight certain data points. We were hoping to use this graph to show the frequency of dishes appearing on the same menu together.

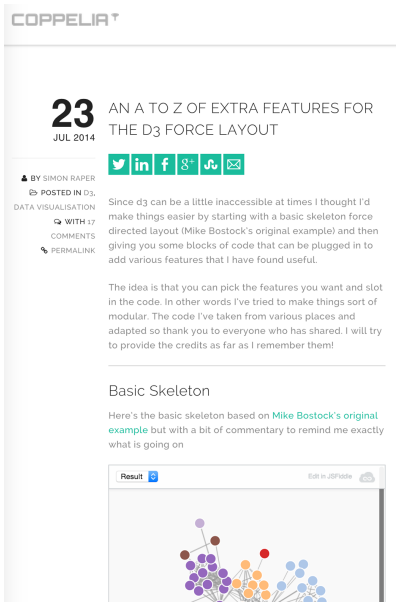


WrapGenius
http://www.wrapgenius.me/

WrapGenius is a data visualization project showing the breakdown of ingredients in the foods we eat. Its goal was to find a better way to display the often-confusing nutrition labels of our foods. We liked the way the author broke down his research into sections that ultimately created an extensive, but cohesive story.

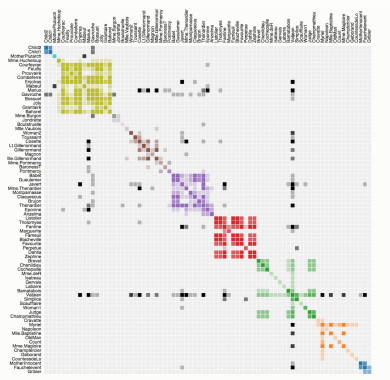


Los Angeles Menu Visualization
http://youarehere.cc/j/menu/losAngeles.html
This is a chord diagram that shows the relationships between two dishes -- the chords are thicker when the dishes occur more frequently on the same menu. This was similar to the message that we wanted to convey about dish relationships on menus, but we wanted to use a diagram that also allowed for greater interactivity.



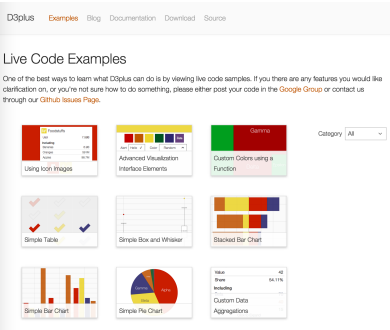
Extra Features for the D3 Force Layout
http://www.coppelia.io/2014/07/an-a-to-z-of-extra-features-for-the-d3-force-layout

When we were planning to use the force-directed graph to represent menu relationships between dishes, these were some features that inspired us. Some of the features we were hoping to implement included collision detection, highlighting, labels, search, and tooltips.



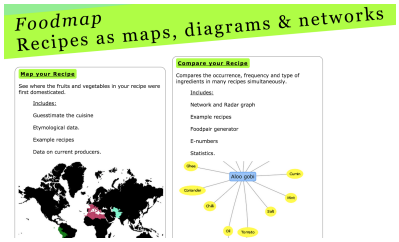
Les Miserables Co-Occurrence by Mike Bostock
http://bost.ocks.org/mike/miserables/

Mike Bostock's example of this co-occurrence matrix using *Les Misérables* data stood out to us because it helped to depict trends in interactions between objects. In this example, Bostock's data tracks the number of interactions between pairs of characters throughout the musical. The visualization shows which characters



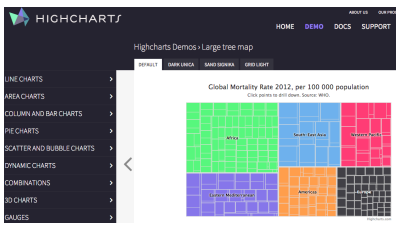
D3plus Examples
http://d3plus.org/examples/

D3plus is a library that extends D3 with pre-made visualizations, such as an animated timeline-treemap hybrid. We used considered using D3plus for this hybrid visualization to help depict time-based trends. However, the library, which is open-source and mainly developed by one person, lacked some of the features we wanted to implement, such as only depicting the time data by decades. As the library continues to develop, we look forward to hopefully implementing some of its features in the future.



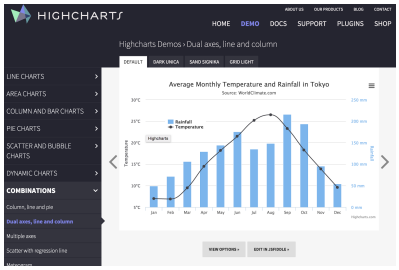
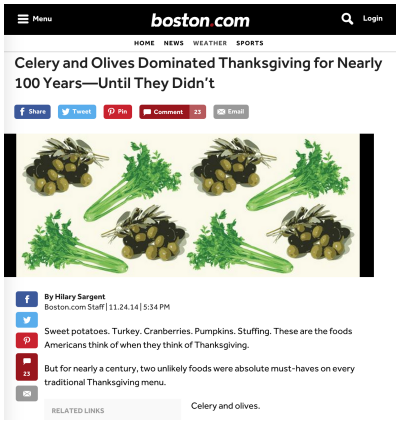
Foodmap: Recipes as maps, diagrams & networks
http://selborne.nl/foodmap/index.php

Wilfried Hou Je Bek is a data artist and cartographer who has done work in combining recipe and menu data with maps, charts, and graphs. His work in graphing food networks and matrices was particularly instructive.

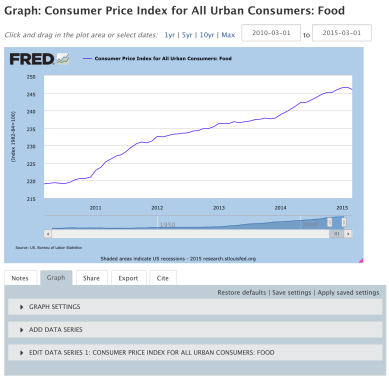


Highcharts Treemap
http://www.highcharts.com/demo/treemap-large-dataset

A drilldown treemap was a way to show the clustering methodology we performed to categorize the dishes by their common synonymous names and the by food group. It could also set the stage and create a color

	legend for the rest of the visualization.
	**Highcharts Dual Axes Chart** http://www.highcharts.com/demo/combo-dual-axes The dual axes chart was a way to show the amount of menus and dishes on the same bar chart. Since we had a large data set, allowing a zooming interaction helps users to see the values in more detail.
	**Food Timeline** http://www.foodtimeline.org/foodfaq5.html Food Timeline is an extensive website dedicated to tracking the history and trends of certain foods. We used the website to better understand some of the (often strange and unfamiliar) foods found in the data set.
	**Historical Overview of Celery & Olives** http://www.boston.com/food-dining/food/2014/11/24/celery-and-olives-dominated-thanksgiving-for-nearly-years-until-they-didn/4GFrGQyPYexAs8OuyCKVBK/story.html We found that celery and olives were (strangely, to us) very popular while we were doing our EDA on the dataset. We did some research to see if there was any historical significance about celery and olives, and came across this article from Boston.com. The article was incredibly interesting to us as it provided historical facts about, for instance, how celery represented the allure of fresh vegetables before industrialization brought us the convenience of frozen foods.

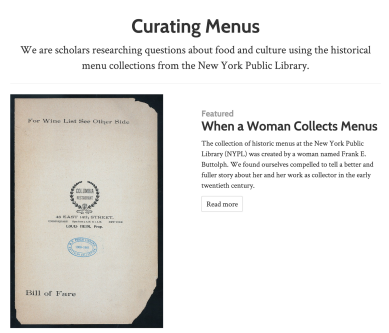
Data Resources



Consumer Price Index Inflation from 1913 to Present

http://research.stlouisfed.org/fred2/graph/?g=8l2

We were initially hoping to normalize the original price data to present day prices in order to be able to compare the dish prices with each other and find out which menus had more expensive dishes or whether the data set contained more expensive or cheaper dishes as compared to the cost of living or average household income.



CuratingMenus

http://www.curatingmenus.org/

CuratingMenus is a site run by a team of researchers, Katie Rawson and Trevor Muñoz, who are dedicated to working with the NYPL's menu dataset. The site includes a data dictionary and several blog posts by Muñoz about how to clean the data. These blog posts were instrumental for us in figuring out how to cluster the data and to remove unnecessary duplicates.

Overview of Work

To see our final product, go to:

http://people.ischool.berkeley.edu/~carlos/menujourneys/

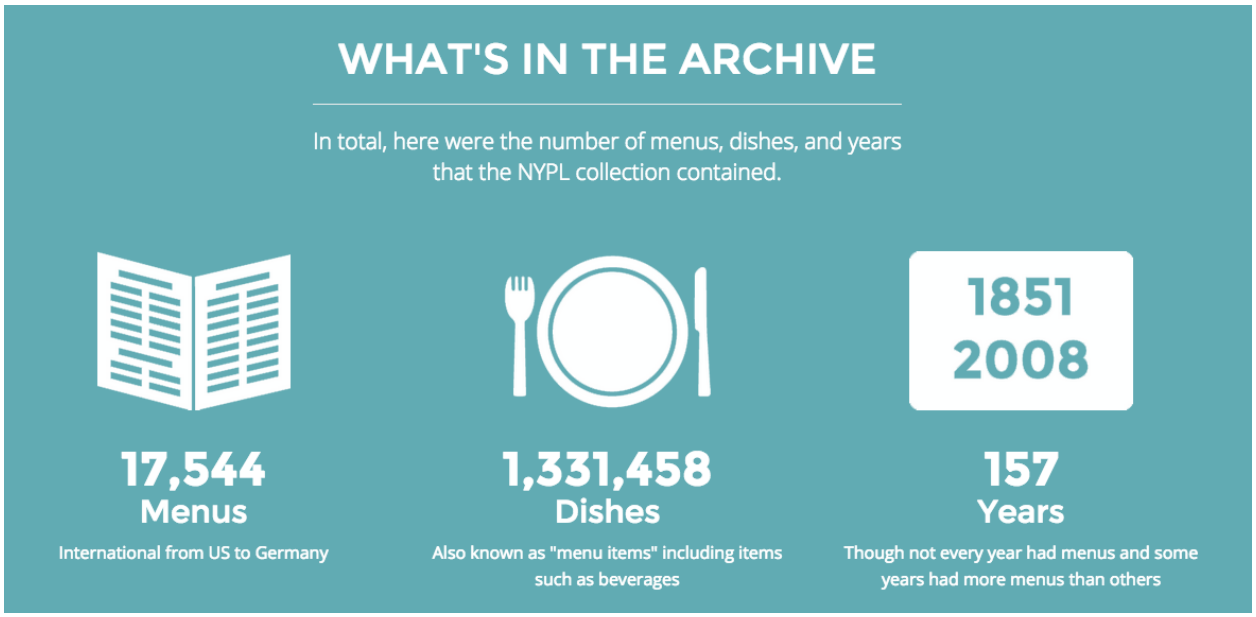
For a guided video tour of our final product, go to:

https://www.youtube.com/watch?v=K_UYdzUNRzQ

To access our code on Github, go to:

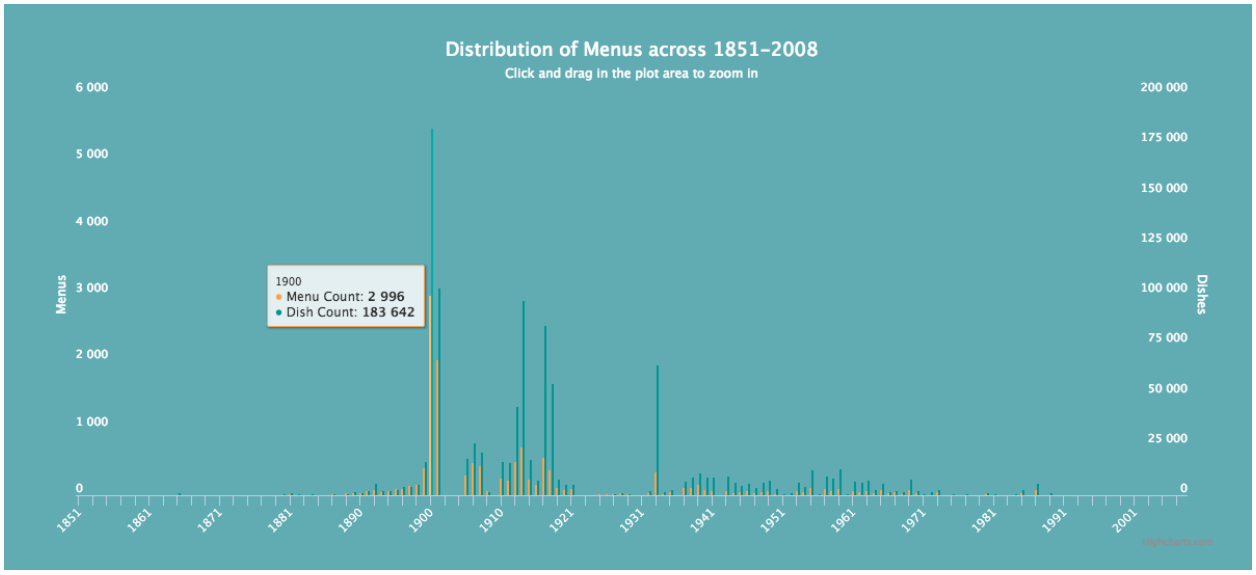
https://github.com/carlooos/menujourneys

Menu Journeys: The Visualization



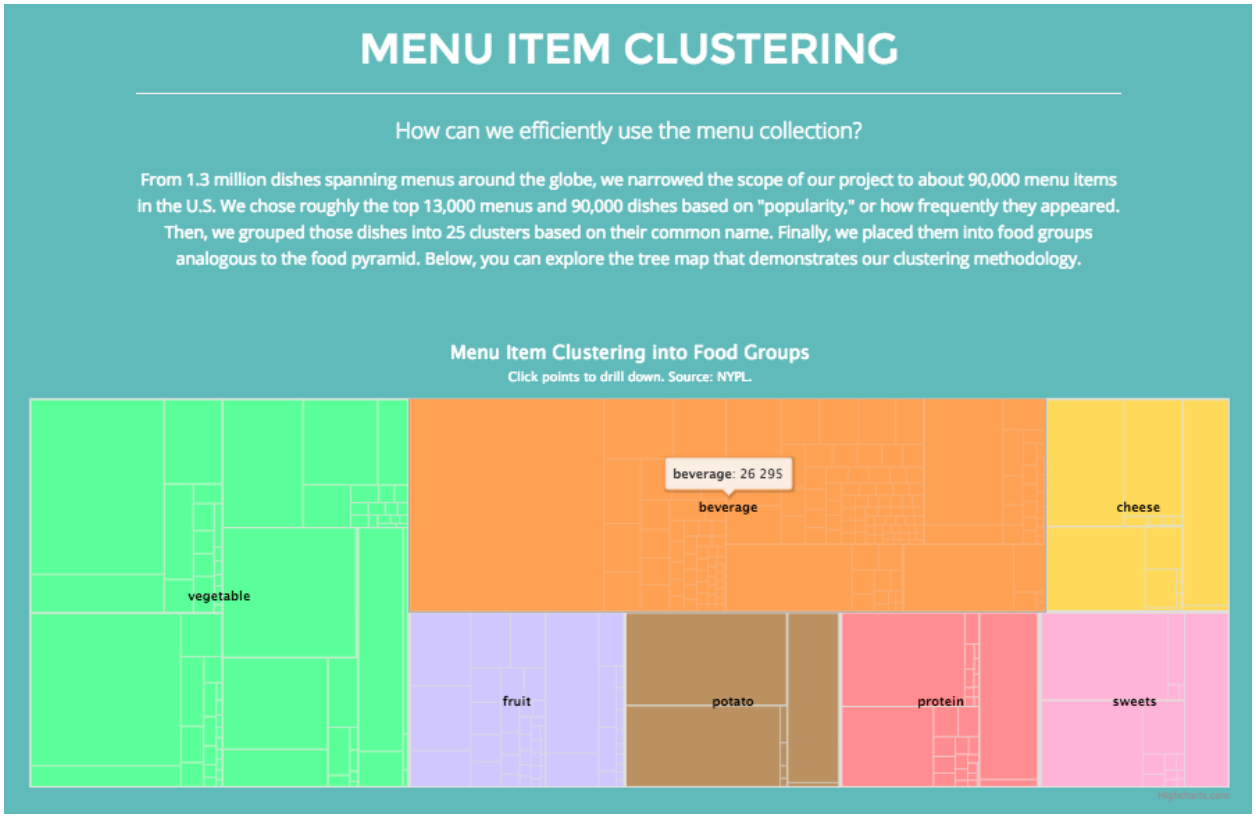
About the Archive

First, we listed some informational numbers to help describe what was in the dataset from the NYPL archive that we started with. We had thousands of menus that included international menus, over 1 million dishes, and a span of 157 years.



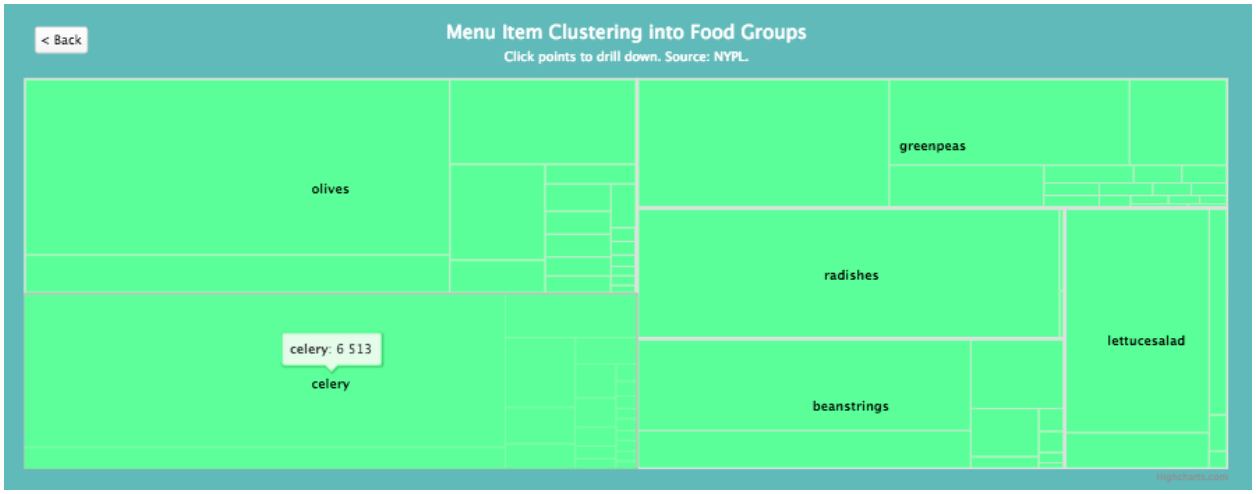
About the Archive: Distribution of Menus

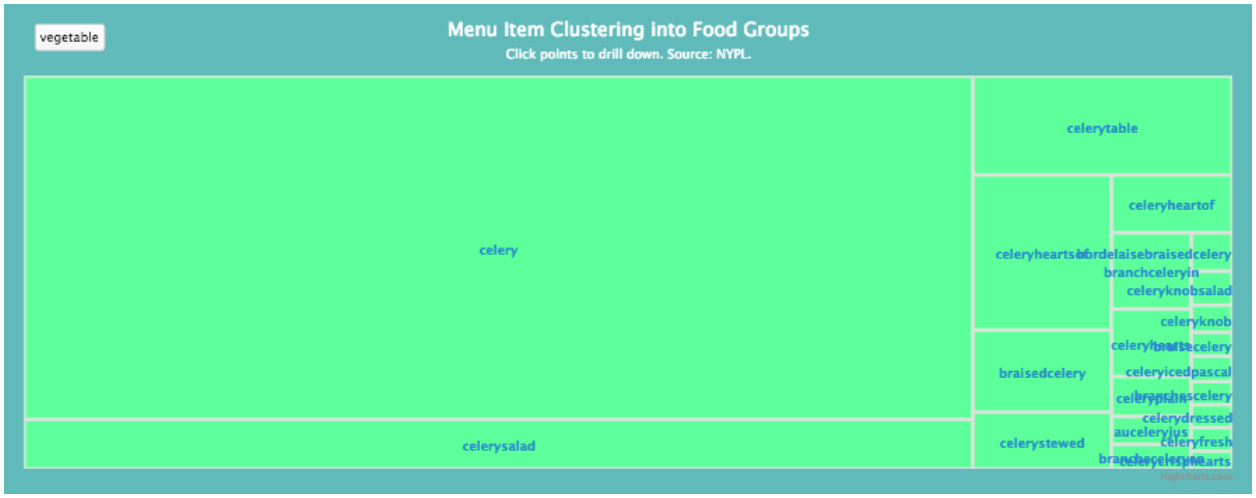
Then, we wanted to show the number of menus and dishes across time. This bar chart shows that the number of menus spiked from the late 19th-century to about the 1920s. This was due to a generous benefactor to the NYPL archive who collected menus from every restaurant she went to during that time period. Due to this skewed distribution, it was difficult to draw any conclusions about all dishes or menus. Thus, our goal was to accurately visualize what was in this archive.



Menu Item Clustering

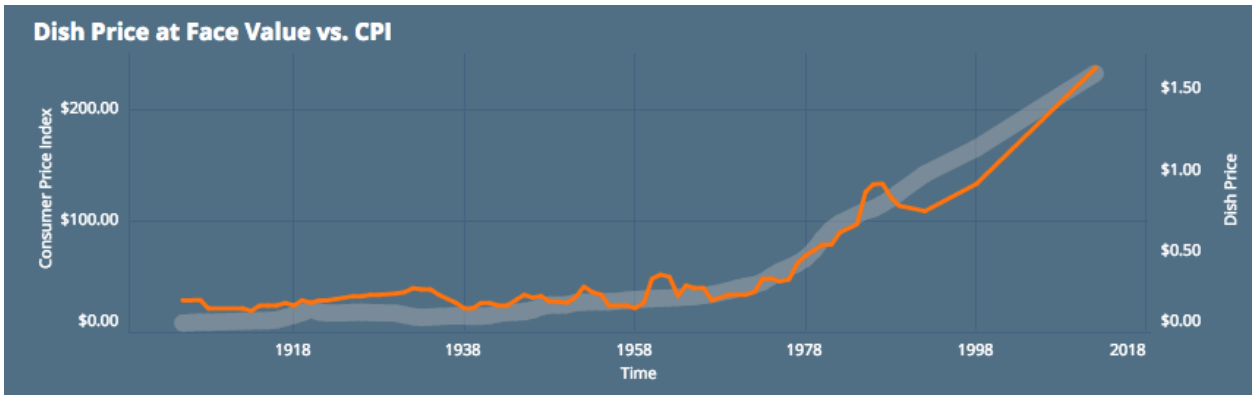
The treemap was an effective way to visualize the clustering of the 90,000 menu items down to ‘fingerprints’, then 25 ‘mega clusters’, and finally seven food groups: vegetable, beverage, fruit, potato, protein, cheese, and sweets. We also color coded the food groups to match the type of food they represented. For example, vegetables is green and cheese is a yellow-orange color. These color choices match a user’s natural mapping of colors to foods. Tool-tips also showed the number of dishes under each category. Interestingly, vegetables were the most common item on the menu, followed by beverages and cheese.





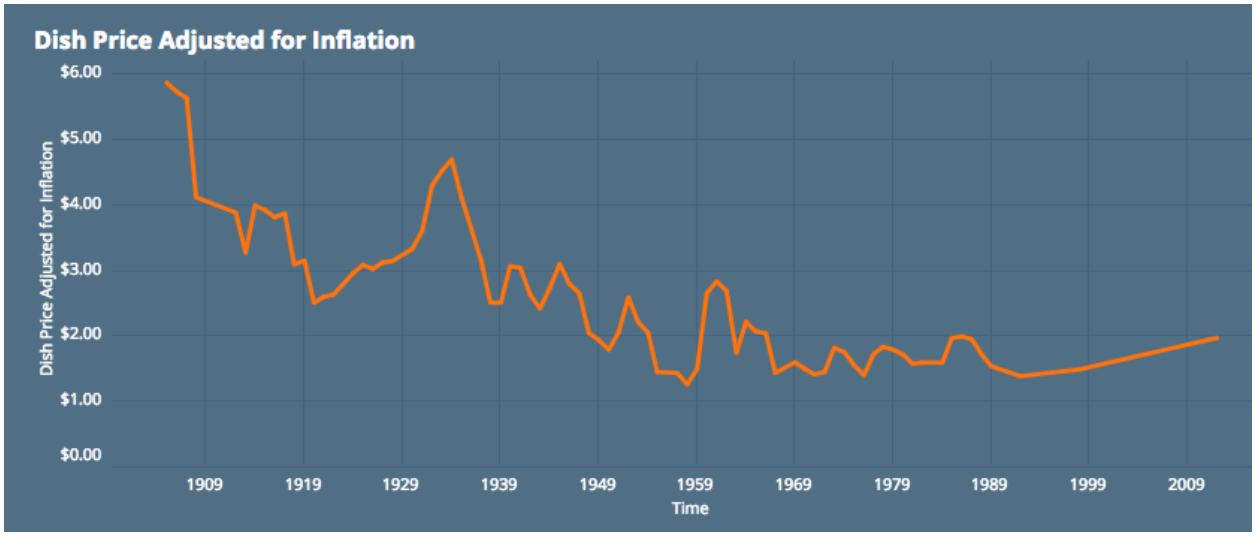
Menu Item Clustering

The treemap allows drilling down two additional layers to see the ‘mega cluster’ categorization as well as the ‘fingerprint’ clustering. On the final layer, the dish actually links back to the search for that term on the NYPL menu archive. We incorporated this change per Marti’s feedback, and we also felt that this added an additional layer of information that was very valuable for the user to continue exploring. In a next iteration, it would be even more beneficial to link the user back to the NYPL page linked to the dish id, which differs from the search for that term on the website.



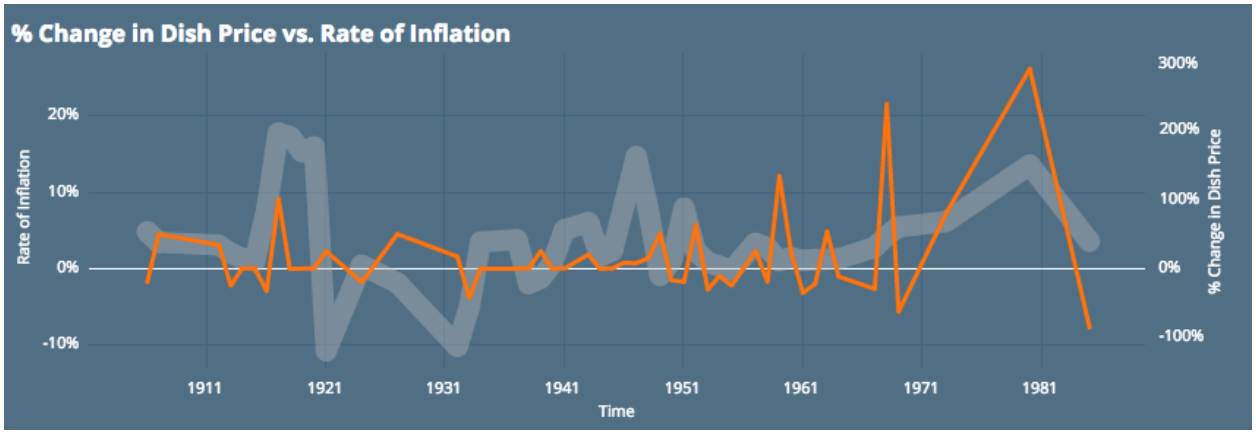
Mapping Dish Price Against Inflation

An interesting way to look at the data would be to analyze the changes in prices of dishes over time. Since we were able to group dishes into clusters, we took the average price of dishes inside each cluster and plotted it over time to see the variation in the face value of the price. This was then contrasted with the changes in the Consumer Price Index provided by the Bureau of Labor and Statistics from 1913 onward. Trends show that dish prices generally followed the inflation of the Consumer Price Index.

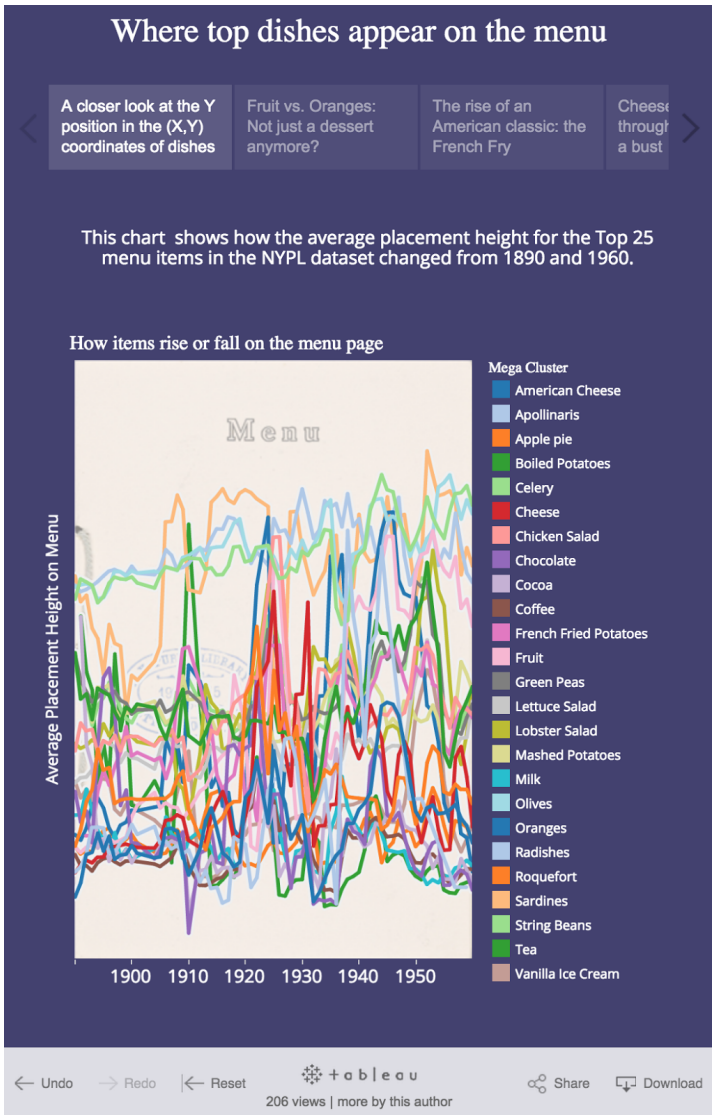


An additional chart was then created to show the dish prices of the clusters adjusted for inflation.

It was interesting to note that there were decreases in dish price over time for certain mega clusters, perhaps due to decreasing costs involved in sourcing and producing those dishes.



Lastly, a final chart was created to contrast year on year changes in average dish prices against the rate of inflation.



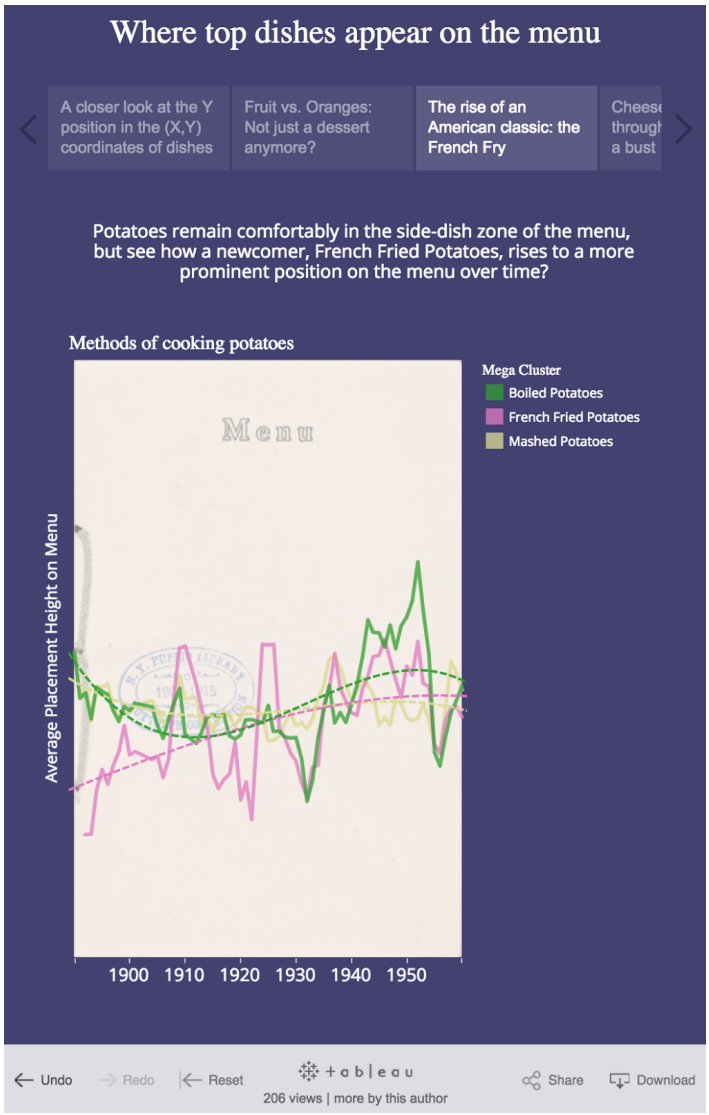
Placement of Dishes on Menus

Here we see all 25 Mega Cluster dishes' average placement height changing over time. Users can click on a particular menu item to highlight that trend line.

Notice how there's a section in the middle-left of the Menu without much activity -- that's the "main course" zone of a menu, and most of our Top 25 dishes are starters, sides, beverages, and desserts.

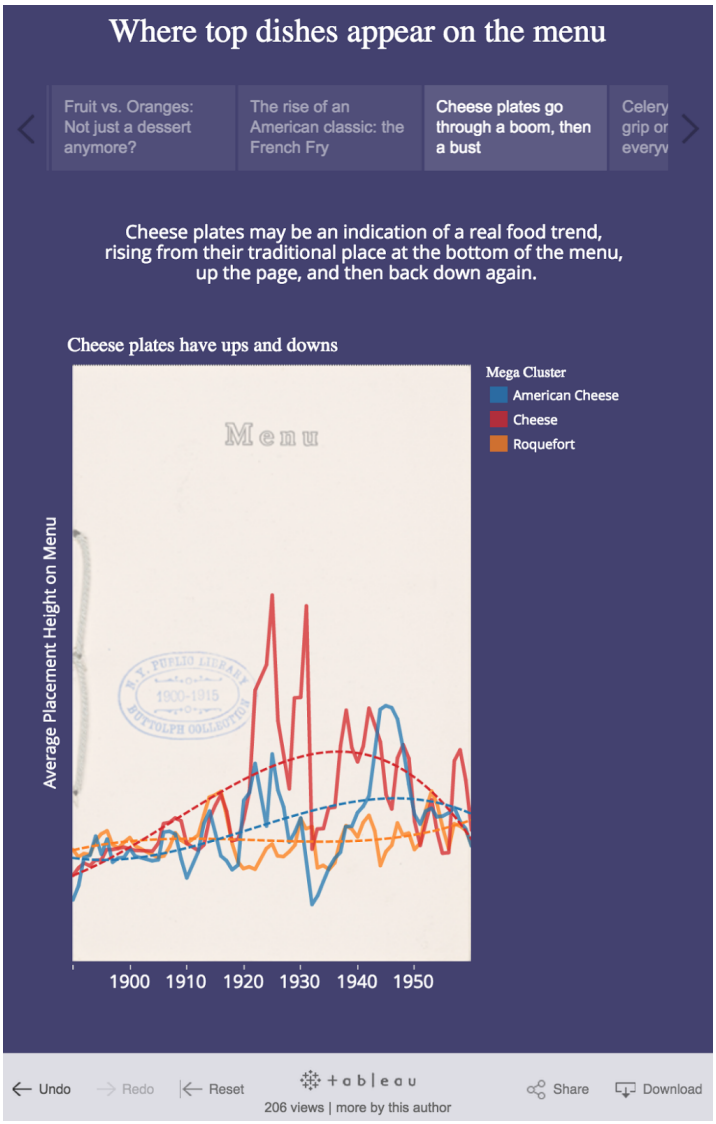
Any time a menu item crosses that "main course" threshold is pretty interesting and may be indicative of changing perceptions of a dish.

We think these visualizations in general get at the heart of the feedback we received from Marti to show how dishes become more or less popular over time.



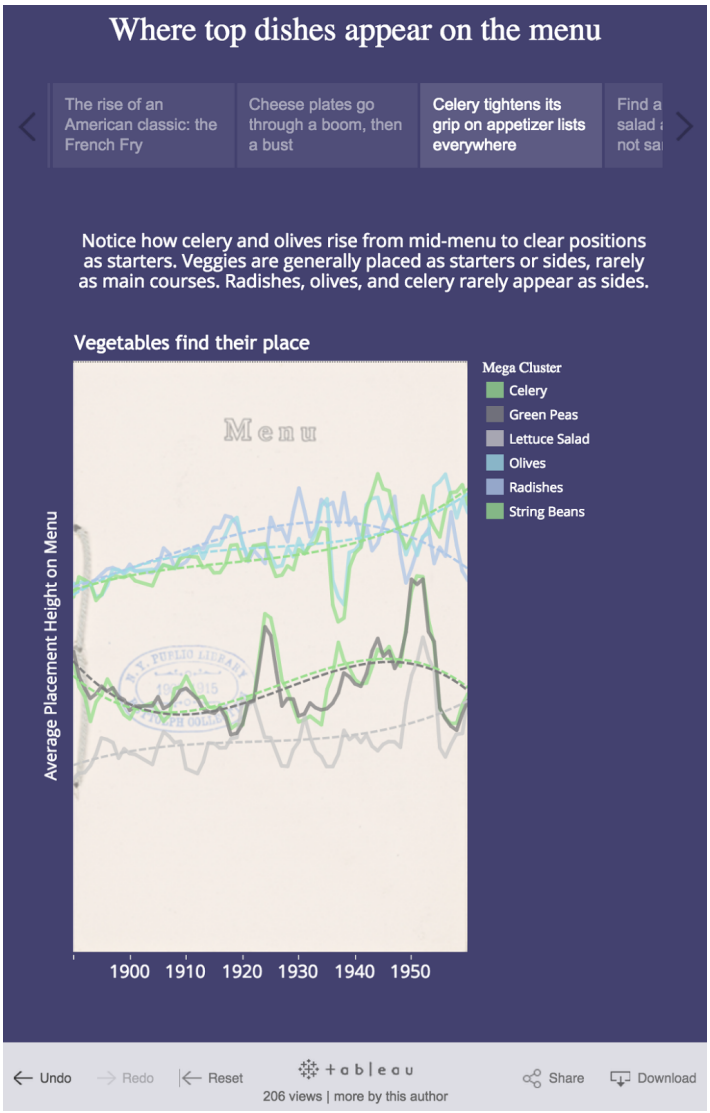
Placement of Dishes on Menus

Here we see the three top cooking methods for potatoes. French fries rise in stature while boiled and mashed preparations taper off a bit. None of the three preparations venture much from the side-dish zone, though.



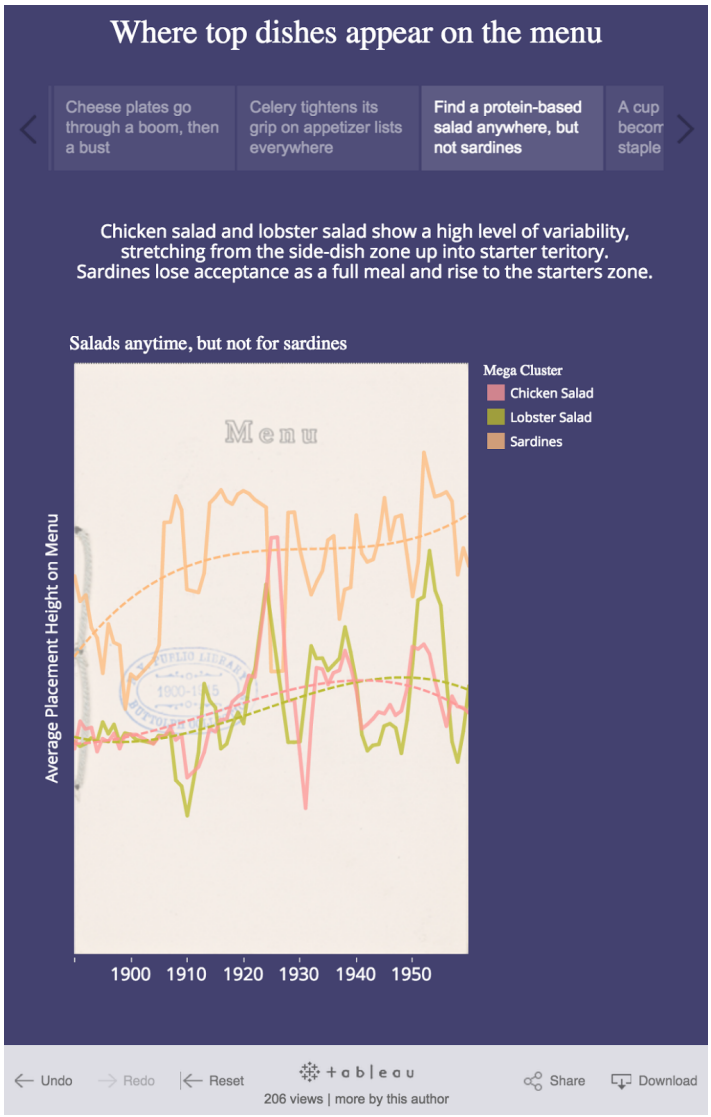
Placement of Dishes on Menus

In fancy restaurants, cheese is often served at the end of the meal, alongside dessert. But comparing the Cheese Mega Cluster, which comprises cheese plates and cheese boards, to Roquefort (a popular variety of cheese), one notices a bit of a trend. Those cheese plates slowly rise up the menu page throughout the first half of the 20th century before falling back down to Roquefort levels once again.



Placement of Dishes on Menus

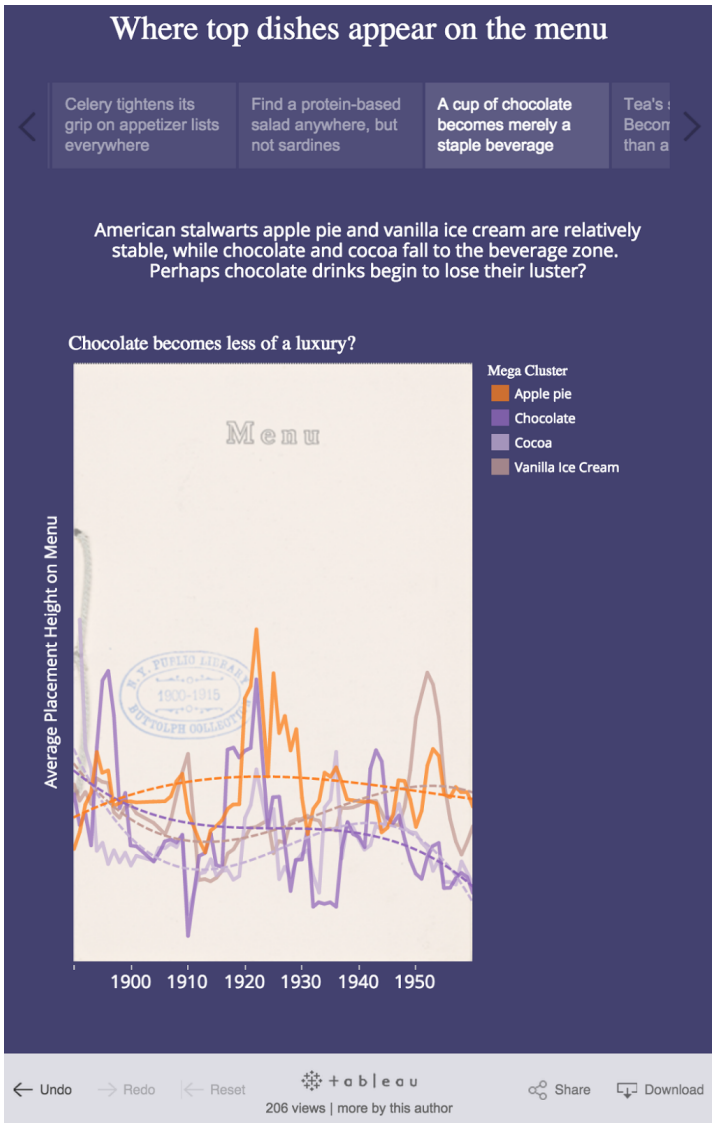
This bifurcation of vegetableness clearly shows how some veggies are perceived only as sides while others, like celery, cement themselves as purely an appetizer option. Vegetables rarely breach main course territory in the center of the menu.



Placement of Dishes on Menus

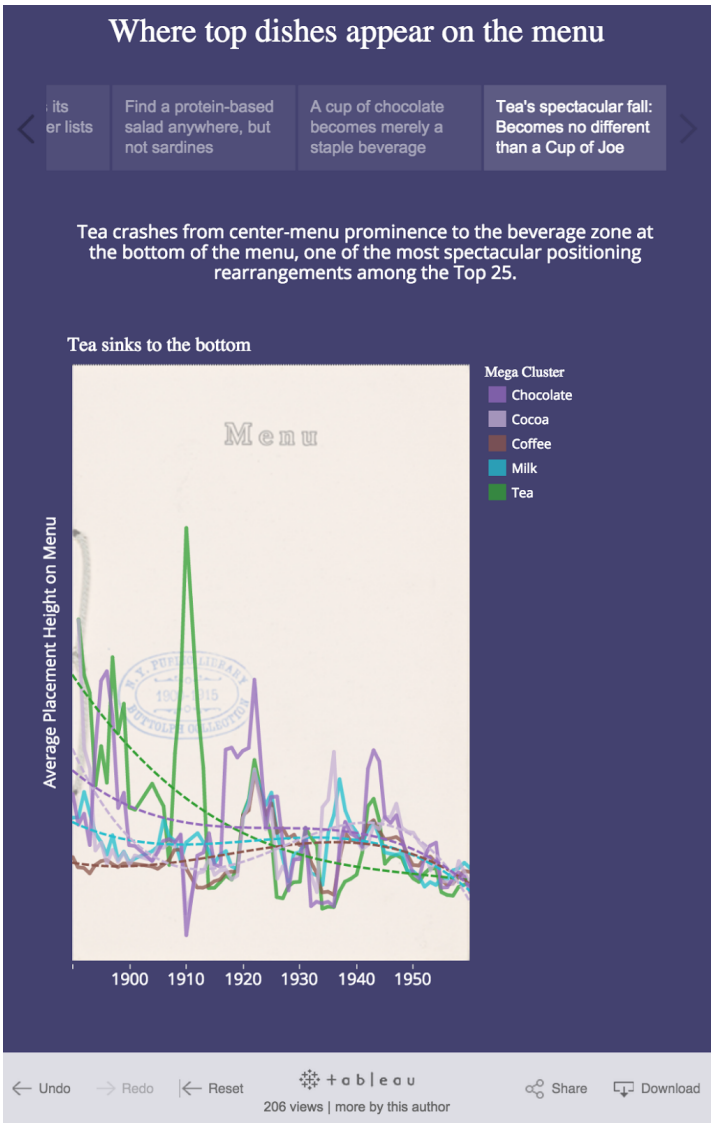
Here we see big swings up and down the menu for chicken salad and lobster salad. Is they sides? Main courses? Starters?

The case of sardines is a bit different, though. They start smack dab in the middle of the main course zone but soon rise to the starters section of the menu. This may be an example of a changing perception of a menu item.



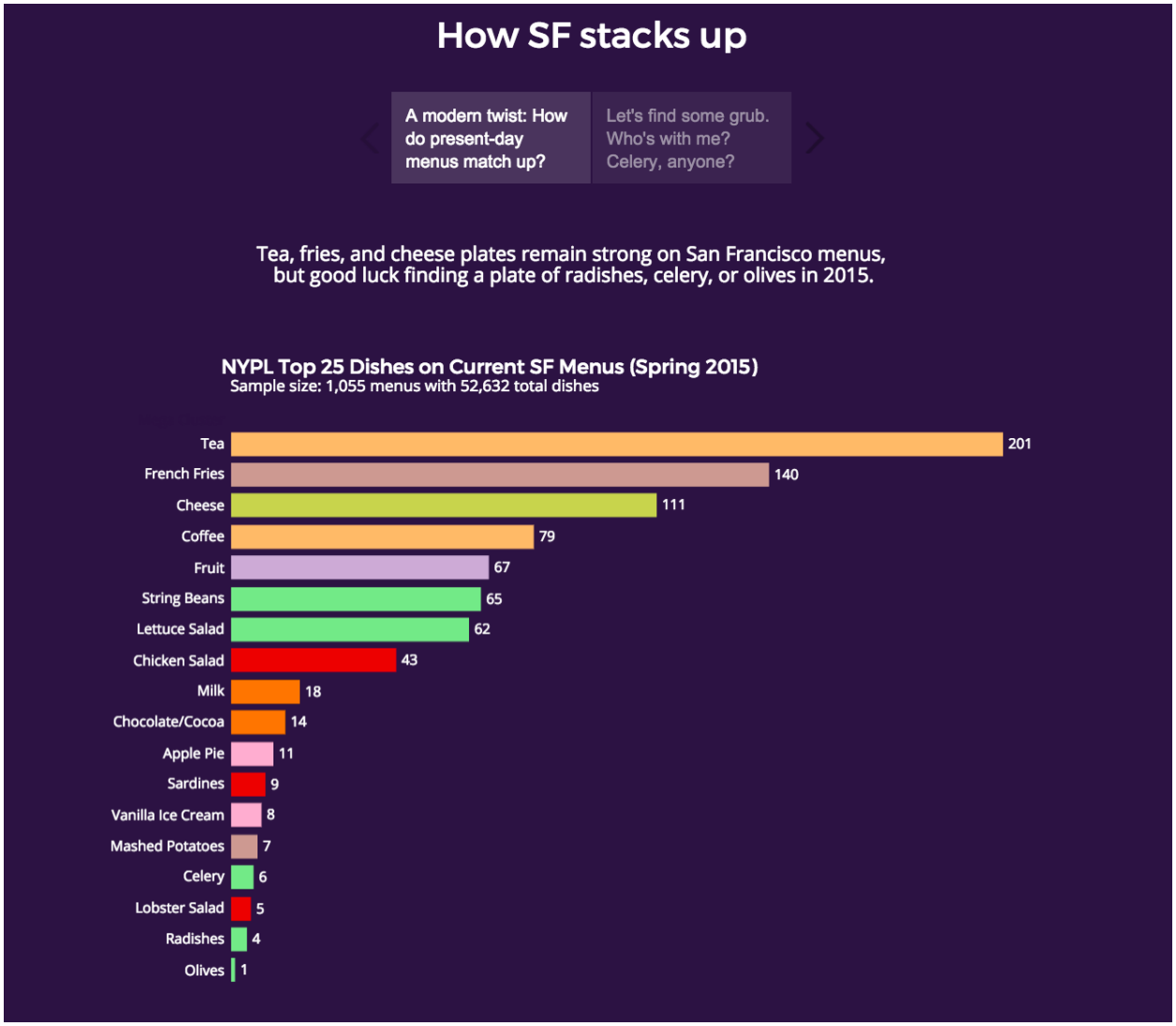
Placement of Dishes on Menus

Trends in these desserts are a bit harder to suss out, but there does seem to be a drop in prominence for chocolate desserts compared with fairly steady placement of apple pie a la mode.

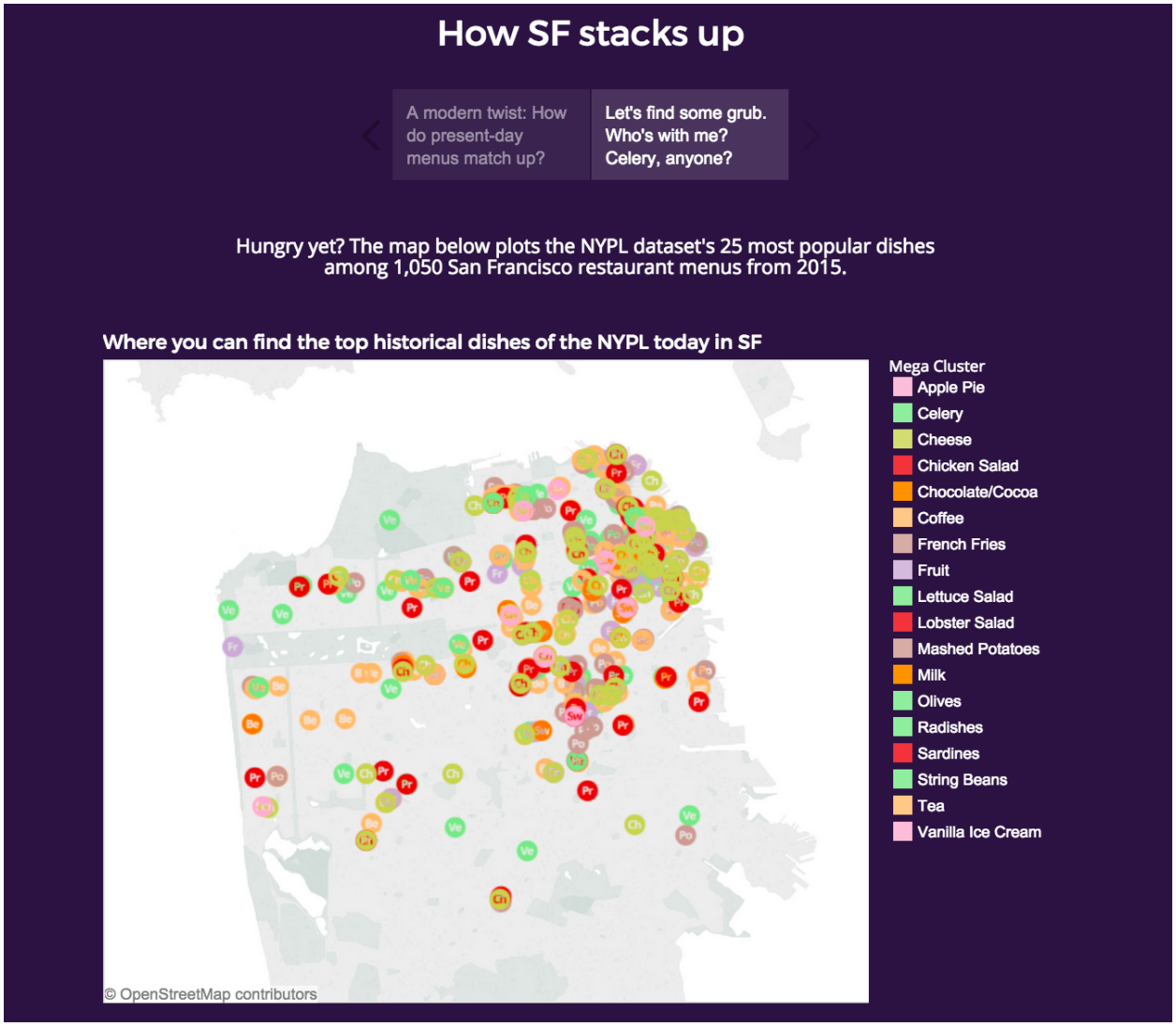


Placement of Dishes on Menus

Tea's average Y coordinate showed a really big drop over time. Its slide down the menu, where it ends up coalescing with other beverages like milk, coffee, and cocoa, might too be indicative of a change in stature for tea as a luxury item.



Here we show a simple horizontal bar chart to indicate how, using the same methodology as we searched the NYPL dataset, those Top 25 dishes would appear in a set of modern menus. The bars are color coded to match the seven categories depicted in the previous tree map.



Each menu item in the SF dataset included the restaurants' geo-coordinates, so we thought it might be fun to plot them on a map so our users could discover where these time-honored dishes can be found in San Francisco.

Data

“What’s On The Menu?” by NYPL Labs
http://menus.nypl.org/data

With approximately 17,544 transcribed menus dating from the 1840s to the present, The New York Public Library’s restaurant menu collection is one of the largest in the world, used by historians, chefs, novelists and everyday food enthusiasts. The menus contain specific information about dishes, prices, the organization of meals.

The New York Public Library’s menu dataset is available for free on their website in two formats: a set of CSVs that are updated every few weeks and an API.

Historical CPI and Inflation in the United States
http://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/

To account and adjust for inflation, we found a historical dataset tracking the consumer price index and annual percent changes of price.

San Francisco Menu Data Acquisition
In the initial stages of the project, we considered focusing solely on San Francisco menu data from the present day. Two companies, Locu and SinglePlatform, are the leading providers of menu information services to companies like Yelp and FourSquare. Through a series of APIs, we were able to slowly but surely begin the process of scraping Locu and SinglePlatform restaurant and menu data for more than 5,000 restaurants in San Francisco. About 1,000 restaurants had menu data robust enough for consideration in our project.

That said, the scraping process took more than two weeks to complete because of API call limits. On top of that, we had to parse the relevant data from a series of JSON objects that weren’t always uniform, which was quite time consuming. Also, our group was quite interested in tracking food trends over the years, and the SF menu data would only be a snapshot in time from April 2015. Because of these constraints, we decided to focus most

of our visualization work mainly on the New York Public Library dataset. We were happy, though, to bring the project full circle in the end and map our insights from the historical dataset to the modern SF data.

Tools

Data Cleaning & Clustering

The dataset from the NYPL includes images of its menu collection that have been run through OCR in order to identify where text is. Text fields appear on the sides of the images to help prompt visitors where they can help provide transcription. Visitors can also participate in manually geo-tagging where menus are from, based on textual information from the images. This ultimately created an incredibly rich and extensive dataset that is powerful due to its size, but also difficult to work with due to the uniqueness of many of the records and potentially inaccurate crowdsourced transcriptions.

To combat these issues, we first used a clustering algorithm from Trevor Muñoz of CuratingMenus.org to create a "fingerprint" for each record of data and to cluster together similar fingerprints. This helped us group together similar items with slightly different names or spellings, eliminating as many duplicate items as possible.

Using these fingerprints, we formed 25 "mega clusters" to further group together similar items and 7 categories to group together the "mega clusters." Here is a sample table of how we grouped our data:

Category	Mega Cluster	Sample Fingerprints
Beverages	Tea	ceylonpottea; chinatea; darjeelingtea; earlgreytea; breakfastenglishtea; camomiletea
Beverages	Coffee	amabassadorcoffeecream with; andcoffeecreamfleishmans

		special; andcoffeemilk; coffee; blackcoffee; coffeecreamcupofwith
Beverages	Milk	milk; bottlemilk; freshmilk; glassmilk; agrademilk; agrademilkpint; freshglassmilk
Beverages	Cocoa	1cocoaportionpot; cocoa; cocoacup; cocoapot; cocoaofpot; cocoacupper
Beverages	Apollinaris (German Mineral Water)	apollinaris; apollinarismineralwater; apollinariswater
Potatoes	French Fried Potatoes	frenchfriedorderpotatoesto; frenchfriedpotatoes; frenchfries
Potatoes	Mashed Potatoes	mashedpotato; mashedpotatoes
Potatoes	Boiled Potatoes	2boiledpotatoes; bermudaboiledpotatoes; boiledpotatoes; boiledhotpotatoes
Cheese	Cheese	cheese; cheeses; assortedcheese; andcheesecrackers
Cheese	American Cheese	americancheese; americancheeseyoung
Cheese	Roquefort (Cheese)	cheeseimportedroquefort; cheeseroquefort; defromageroquefort; roquefort
Sweets	Apple Pie	appledeepdishpie; appledutchpie; applegreenpie; applehomemadepie; applepie
Sweets	Vanilla Ice Cream	creamicevanilla;

		creamfrenchicevanilla; americancreamicevanilla
Sweets	Chocolate	chocolate
Fruit	Fruit	assortedfreshfruit; assortedfruits; cocktailfreshfruit; cupfreshfruit; fraisfruits; defruitsalade; assortedfruitstewed
Fruit	Oranges	1orange; 2oranges; californiaorange; floridaoranges; orangesliced; orangewhole; eachoranges
Protein	Sardines	4frenchsardines; bonelesssardines; ahuilelsardines; inoilsardines; sardines; sardinessmoked; fumeesardines; frenchsardines
Protein	Chicken Salad	allchickenmeatsaladwhite; chickendarkmeatsalad; chickenmayonnaisesalad; chickenmayonnaisesaladwith; chickensalad; chickensaladsmall
Protein	Lobster Salad	freshlobstersalad; lobstermayonnaisesalad; lobstersalad; lobstersaladsmall
Vegetables	String Beans	beansfreshstring; beansnewstring; beansstring; beansbutteredstring; aubeansbeurrestring
Vegetables	Green Peas	americanpeas; freshgreenpeas; freshpeas; greennewpeas; greenpeas

Vegetables	Lettuce Salad	lettucesalad; lettucesalade; heartslettuceofsalad; headlettucesalad
Vegetables	Celery	braisedcelery; celery; branchescelery; celeryfresh; celeryheartsof; celeryknob
Vegetables	Radishes	newradishes; radishes; radishesrose
Vegetables	Olives	olive; olives; olivesqueen; olivesripe; olivesqueenspanish; greenolives; frencholives

We also used R to calculate the Jaccard similarity between pairs of items, and we created a matrix of these Jaccard values which we used for our co-occurrence matrix (*please see "Co-Occurrence Matrix & Force-Directed Graph" in "Results & Feedback" for a more detailed explanation of this visualization*).

- Python (Pandas) and R to clean and optimize the data set
- Adobe Illustrator and the Noun Project to create icons
- Bootstrap for the web page template
- Highcharts.js to visualize menu and dish collection along with clustering methodology
- Tableau for exploratory data analysis and visualization of dish price and location over time as well as mapping current locations of where dishes can be found in SF
- d3.js for the force-directed graph and co-occurrence matrix (both later discarded from the final visualization - *please see "Co-Occurrence Matrix & Force-Directed Graph" in "Results & Feedback"*)

Our Process

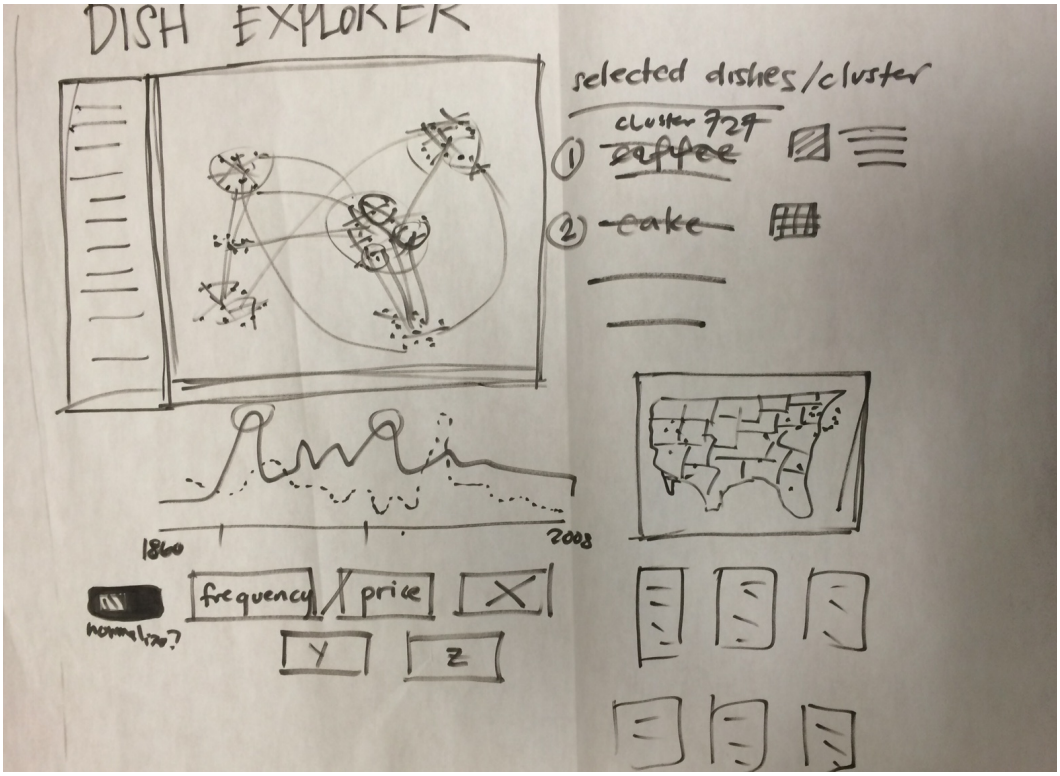
1. **Identify goals to accomplish with the data set**. Initially, we were hoping to show trends that reflected socio-economic patterns or historical events that would change people's perception of the history of certain dishes or foods. However, we found that the data set was quite skewed toward fancier menus during a specific period of time due to a generous donation of menus by a single benefactor, and it would have been misleading to draw overarching conclusions from this data set about menus or foods in general. Thus, we decided to create an exploratory visualization of the NYPL menu archive.

2. **Exploring the data with Tableau.** Each of us explored a different aspect of the data set using Tableau. One aspect that we explored was what we could visualize using location data such as menus by location, dish by location, or frequency of a location appearing. From this initial data exploration, we realized that the location data was not very uniform and locations could be transcribed many different ways or not at all. Another aspect we explored was comparative price of the dishes after adjusting for inflation. From the initial exploration, we found that the dish prices could occur in many different currencies and there were some outliers due to ranging locations of decimal places. For example, $1.00 may have been transcribed as 100 in the data set. We also found that prices could vary widely based on the way the menu item was served. For example, a pot of coffee was much more expensive than a single cup of coffee, so it was difficult to compare the prices of these two items. We also explored plotting the dish location on the menu using the x-y coordinates provides in the data.

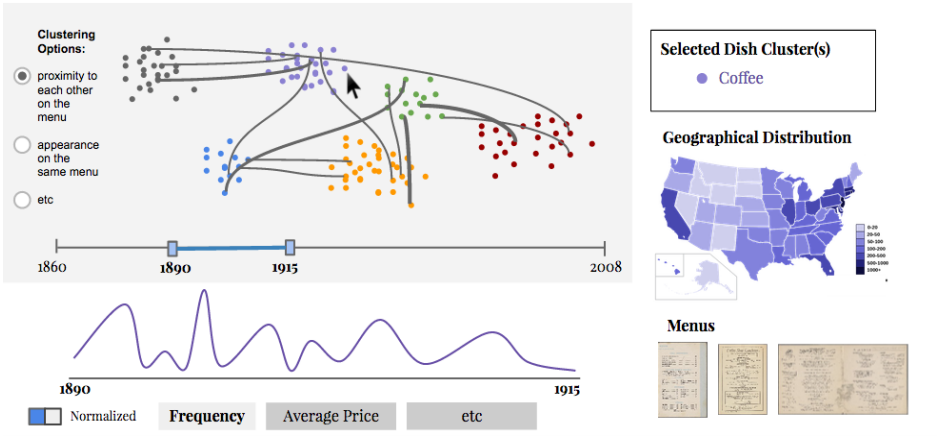
3. **Data wrangling**. We clustered or grouped duplicate dishes with the same name but spelled slightly differently together. In order to do that, we used an algorithm created by Trevo Munoz, followed by a manual clustering of the dishes that appeared most frequently into 25 mega clusters, narrowing down the data set to 13,000 dishes. We also planned to show co-occurrence on menus between dishes, so we needed to compute the intersection of dishes. In order to do this, we ran an algorithm using R on the remaining 13,000 dishes to compute the number of

menus that each dish occurred with another dish. For our project, we focused only on menus within the United States.

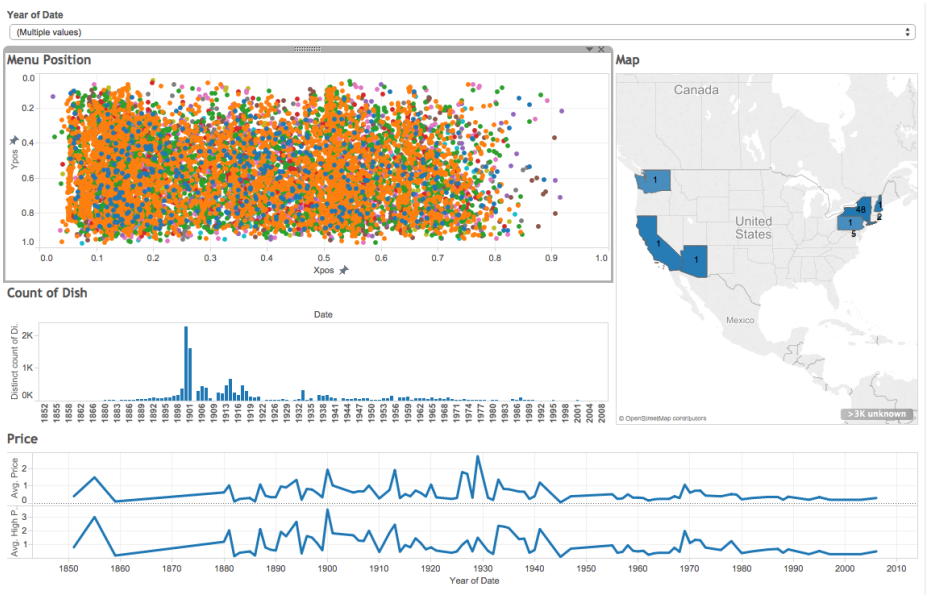
4. **Prototype layouts of the data along with planned interactions**.



Above was an initial paper prototype of our exploratory visualzation.



Above is a mockup in PowerPoint of our visualization of menu item location, price, frequency, geographical distribution, and menus.



Above is a prototype of an exploratory dashboard visualization using Tableau showing menu item location, number of dishes in the dataset, price fluctuations, and location of the menus in the US.

5. **Choosing the graphic forms and principles for visualizing the data**. In our final visualization, we used bar charts to show the collection of menus and dishes in the archive, similar to the prototype; we used line graphs to show the change in dish prices. Instead of a scatter plot, we used a line graph to show the movement in menu item location over time. The map was used to show location of the menu items using current San Francisco menus, rather than the locations in the NYPL dataset. We also added a treemap to demonstrate the layers of clustering performed.

6. **Code the layout and interactions**. We used a bootstrap template to structure our web page and to tell the story of the archive, our menu clustering work, and the final visualization through a one-page, scrolling website. We used photoshop

to create icons that would visualize the data from the archive that we had to work with.

During this process, we encountered a few challenges and obstacles along the way and had to shift our course as a result. This is explained in detail in the "Co-Occurrence & Force-Directed Graph" section below.

7. **Evaluate the visualization with interested researchers**. We evaluated the visualization with three users who were interested in menu archives to walk through our visualization and evaluated how well they understood what the visualizations represented and whether it provided new insights or takeaways.

Results and Feedback

We went through several rounds of user testing with researchers throughout the development process to clarify our visualization strategy and our overall narrative in Menu Journeys.

One researcher had access to early iterations of the force-directed graph and called it quite confusing, to say the least, which was apparent to everyone else on our team as well. She suggested exploring alternative methods to visualize co-occurrence, such as a matrix, which also ultimately proved to be too challenging to implement under the constraints of our dataset and the deadline at hand.

Another researcher suggested a multi-pronged approach to our design -- using many visualizations woven together to tell a story -- which is the direction our team ultimately took. She cautioned against the use of an all-encompassing dashboard tool and instead wanted to be guided toward inspiring insights.

Perhaps the most valuable user-testing session was with a former researcher for the federal Food and Drug Administration. He particularly liked the feature in our cluster tree map that allowed him to directly search the NYPL menu archive based on dish names via hyperlink, which we implemented following feedback from Marti.

He also appreciated the narrative structure of Menu Journeys. He said he liked scrolling from visualization to visualization and being told a story by each one. The supplemental text and story tiles from Tableau made it possible to guide the discussion. “These write-ups by the visualizations really highlight and guide the analysis. It’s like something you would see in a museum,” he said, by pointing out things “beyond the obvious.”

Co-Occurrence & Force-Directed Graph

We initially created two visualizations that we thought would work well for our project, but they turned out to not be appropriate for the type of data we were working with. These two visualizations included a Co-Occurrence Matrix (adapted from Mike Bostock’s example here) and a Force-Directed Graph (also adapted from Mike Bostock).

The problems with these visualizations was they displayed single nodes or columns and rows for *each* record of data, which was a problem considering that our dataset was massive and included either hundreds or thousands of records depending on how we decided to segment the data.

Co-Occurrence Graph

For our Co-Occurrence graph, while it was flashy and very visually appealing due to its use of animation and color, based on the feedback that we received, we did not feel that it was good enough to include in our final project files. Our viewers liked the animation, but they could not immediately understand what was going on, which was a problem. The use of color was confusing to them, and they could not understand the clustering that was happening in the matrix.

We realized that the data we were using was not ideal for this type of visualization. Of a 356x356-sized matrix of Jaccard similarity values for each pair of data, we sliced and used only the top 100 records based on these values. Unfortunately, what we realized this meant was that we were displaying data that was already similar to each other by calculation and pre-clustered.

In Mike Bostock's example, which is based on a dataset of interactions between *Les Miserables* characters, the matrix lists every character who appears in the musical. The visualization helps the viewer see how certain characters may have grouped together. The use of color in his example is helpful because it differentiates between these different clusters.

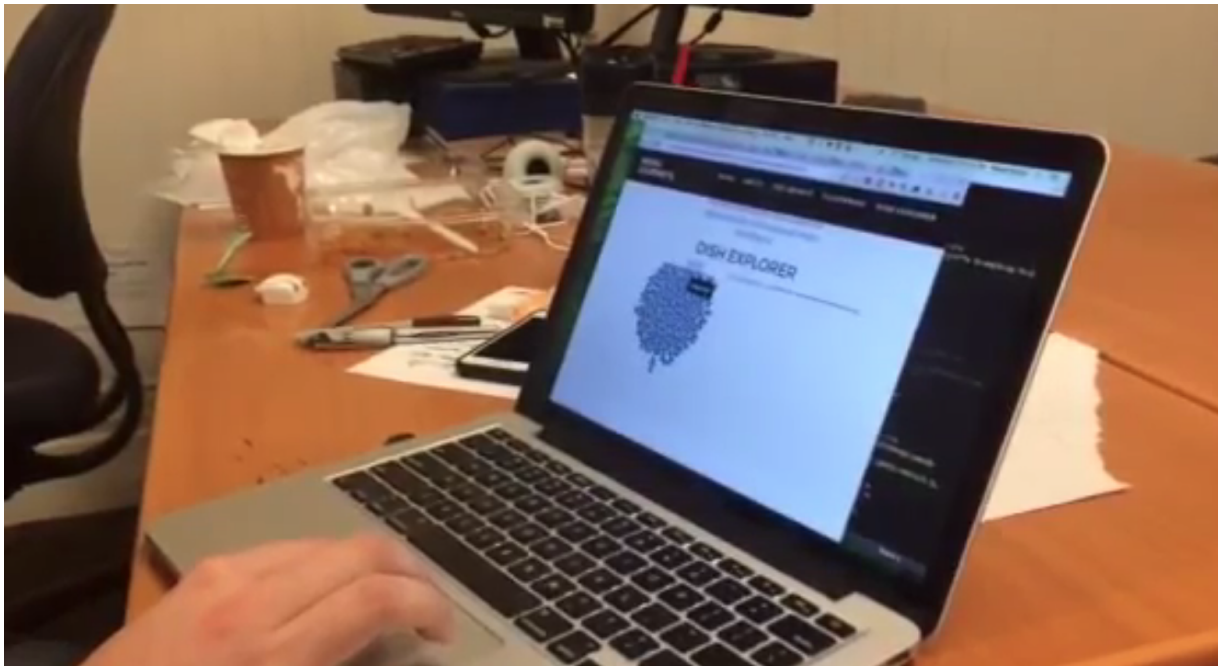
For our example (no longer linked to our final submission though still available for review), the data we used was already in clusters, such as coffee, cocoa, milk, apollinaris, apple pie, chocolate, and tea. When sorting "By Cluster" through the drop-down menu available at the top, what was depicted in the end visualization was therefore unhelpful; it merely showed which records were part of the same cluster.

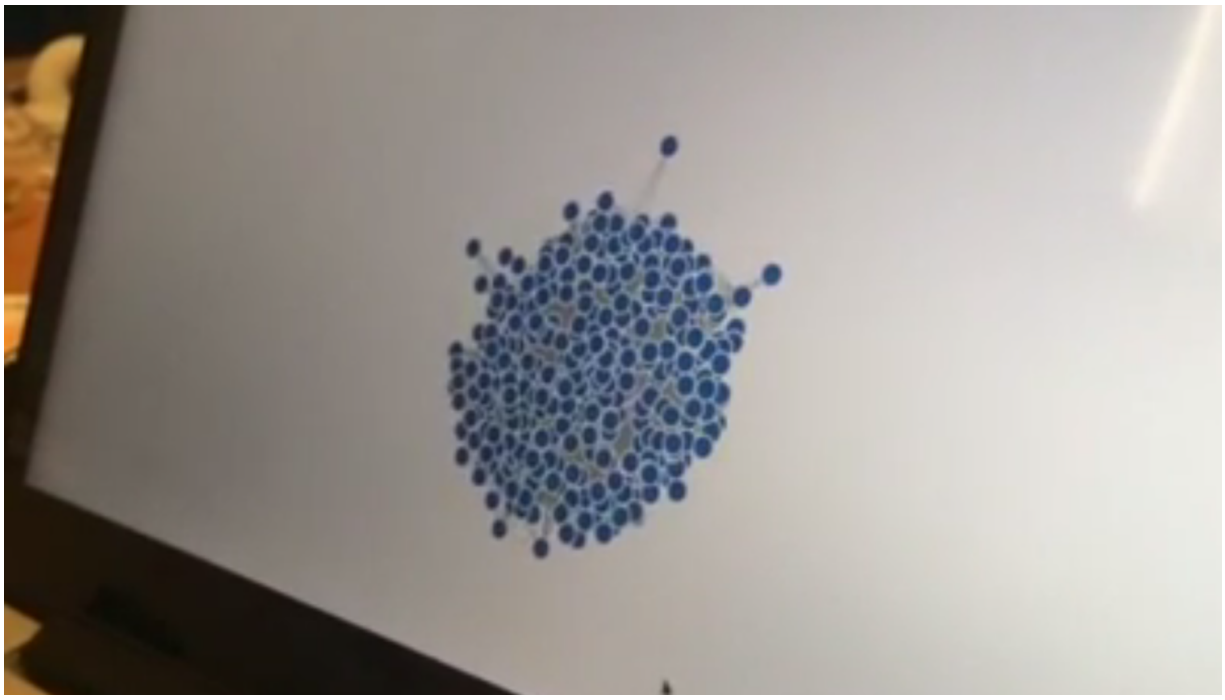
When sorting "By Name," this was slightly more helpful as it showed that coffee matched with other non-coffee items, such as tea, milk, and apple pie. However, looking at the visualization on an overall basis, we realized that the visualization did not do our dataset justice and was not the best form of visualization we should use for our project.

Force Directed Graph

We initially tried to use a force-directed graph to depict relationships between items, similar to our attempt with the co-occurrence matrix. However, like the matrix, this did not work well because of the sheer number of records of data in our dataset. Our force-directed graph attempt included 356 nodes with over 12,000 edges.

We tested the force-directed graph and captured the effect on video here (please click the image to visit our YouTube video):





Links

Final Visualization: http://people.ischool.berkeley.edu/~carlos/menujourneys/

Guided video tour of final visualization: www.youtube.com/watch?v=K_UYdzUNRzQ
Github: https://github.com/carlooos/menujourneys
YouTube demo of force-directed graph: www.youtube.com/watch?v=bPdfbyMobBw

Contribution

We each contributed to the project in different ways, often contributing in the ways that we had the most inspiration or areas where we had the most interest. Below, we estimated our individual contribution to each part of the project on a scale from 1 - 5:

1= “low” 2 =”sufficient” 3 = “supporting” 4=“significant” 5=”outstanding”

	Audrey	**Brian**	**Carlos**	**Stephanie**
Concept Development	4	4	4	4
Data Wrangling	2	4	2	4
Research / EDA	3	3	3	3
Web Design	3	2	4	3
Tableau Visualization	2	5	4	2
d3 Matrix Visualization	2	2	2	3
Highcharts Visualization	5	1	2	2
Total	**21**	**21**	**21**	**21**