INFO 4602 Final Project – Pizza Visualizations

Hannah Weber, Trevor Buck, Clark Mousaw

University of Colorado Boulder

**Abstract**

The origins of America’s favorite delivery food can be traced back to antiquity. The first known documentation of the word “pizza” was in the year 997 AD in central Italy, but archaeologists believe that Persian and Greek citizens baked flatbreads topped with cheese as early as the 6th century BC. Since it was introduced in the United States by Italian immigrants, American culinary ingenuity has greatly influenced the many styles and flavors of pizza. To further explore this American cultural and culinary staple, the use of visualizations presents an interesting and creative opportunity to convey statistical information. Using visualizations to show data clearly and concisely led to the affirmation of some widely held pizza preconceptions such as America’s favorite toppings, but also helped to unveil some unexpected information such as regional costs.

**Keywords**: Pizza, visualization, dataset.

**Index Terms**: Information Visualization

## Introduction

This paper seeks to demonstrate the different methods, strategies, and software that can be used to create visualizations which show the varying types of information related to America’s pizza preferences and consumption.

Analysis was performed using a single dataset called, “Pizza Restaurants and the Pizza They Sell” found on Kaggle, which was provided by Datafiniti’s Business Database. This dataset lists over 3,500 pizzas, offered by 1,827 restaurants, in 1,028 cities, across 44 states. The dataset is a comma separated value (csv) file, with each line of data containing a unique pizza, along with its restaurant, the restaurant location, price, category, and classification. 

## Related Work

There are plenty of instances where food-related data has been visualized using various tools. Research related to the health effects of food has been presented using visualization types including scatter plots and line graphs [1]. Researchers have also utilized infographics to communicate multiple findings in one large visualization. These infographics are composed of text, tables, scatterplots, line graphs, and spatial visualizations [2]. Visualizations have also been used to make sense of different food items and to find similarities amongst different food groups. To do so, multivariate data was plotted in a 3D graph to determine the correlation between salads, sandwiches, and soups [3]. In addition, a researcher wanted to challenge the accuracy of a survey so he visualized the data using radar charts. He used the radar charts to plot multivariate data to see how people rated various burgers based on eight different criteria. He then compared the findings of the visualization to the prevalence of the restaurants in the US to determine the accurate of the initial findings [4].

Visualizations have also been used to compare the frequency of Google searches related to various food phrases over the past 15 years. They wanted to see the distribution of the searches through the 15 years and throughout each calendar year to find any trends that may occur in the data [5]. In a similar realm, thousands of recipes were analyzed from a website and used to create a food recommender system for users of the website. This food recommender was then visualized to communicate the information produced by the system in a user-friendly manner for users [6].

The most unique forms of graphical representation for food-related data is edible visualizations. These visualizations are made up of real food, and use food as a means of data expression. Although it seems like a simple tool, food can be used to communicate complex datasets, such as the gender inequality in kitchens and the distribution of wealth across various countries [7]. There is even a company who is devoted to hosting workshops in numerous countries related to data visualizations with food. This company attempts to visualize local data using the local cuisine [8].

## Discussion

To show the statistical information contained within the dataset in interesting and creative ways, multiple types of visualizations were utilized. The first method, which pursued a more creative approach, was a series of three pie charts that visualized counts of different toppings contained within the dataset. The second method identified the most popular pizza in each state and visualized the findings using a map of the United States. The third method was to look at pizza sold by city. It gives us insight as to the average price per pizza, as well as the amount of pizzas sampled.

## Pizza Themed Pie Charts

Three separate pie charts that were created to show the counts of different pizza toppings within the dataset. The pie charts are separated by the category of pizza topping found within the dataset. The three categories are shown in Figure 1 below:

1. Cheese 2. Veggies 3. Meat

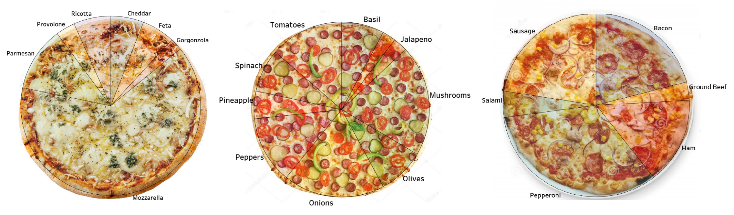


Figure 1: Pizza Pie Charts

For each pie chart, the size of the “slice” directly corresponds to the number of times the topping was found within the dataset. To ensure accurate counts of each topping, code was written in the Python programming language to iterate through a single column of the dataset where toppings were listed. The code searched for specific keywords or toppings, and returned the exact tally of keywords. With the exact counts of each topping, accurate pie charts were then created.

The pie charts were created in the Tableau data visualization software. In Tableau, there are distinct differences between *dashboards* and *sheets*. *Dashboard* views have a functionality where images can be inserted as backgrounds. This allows for more rich and customizable visualizations. *Sheets* are where raw data is typically visualized. For these three specific pizza pie charts, an image of a whole pizza was uploaded as the background of a dashboard, and the raw data was visualized in a sheet as a transparent pie chart. The sheet containing the pie chart was then overlaid onto the dashboard, and aligned with the image of the pizza.

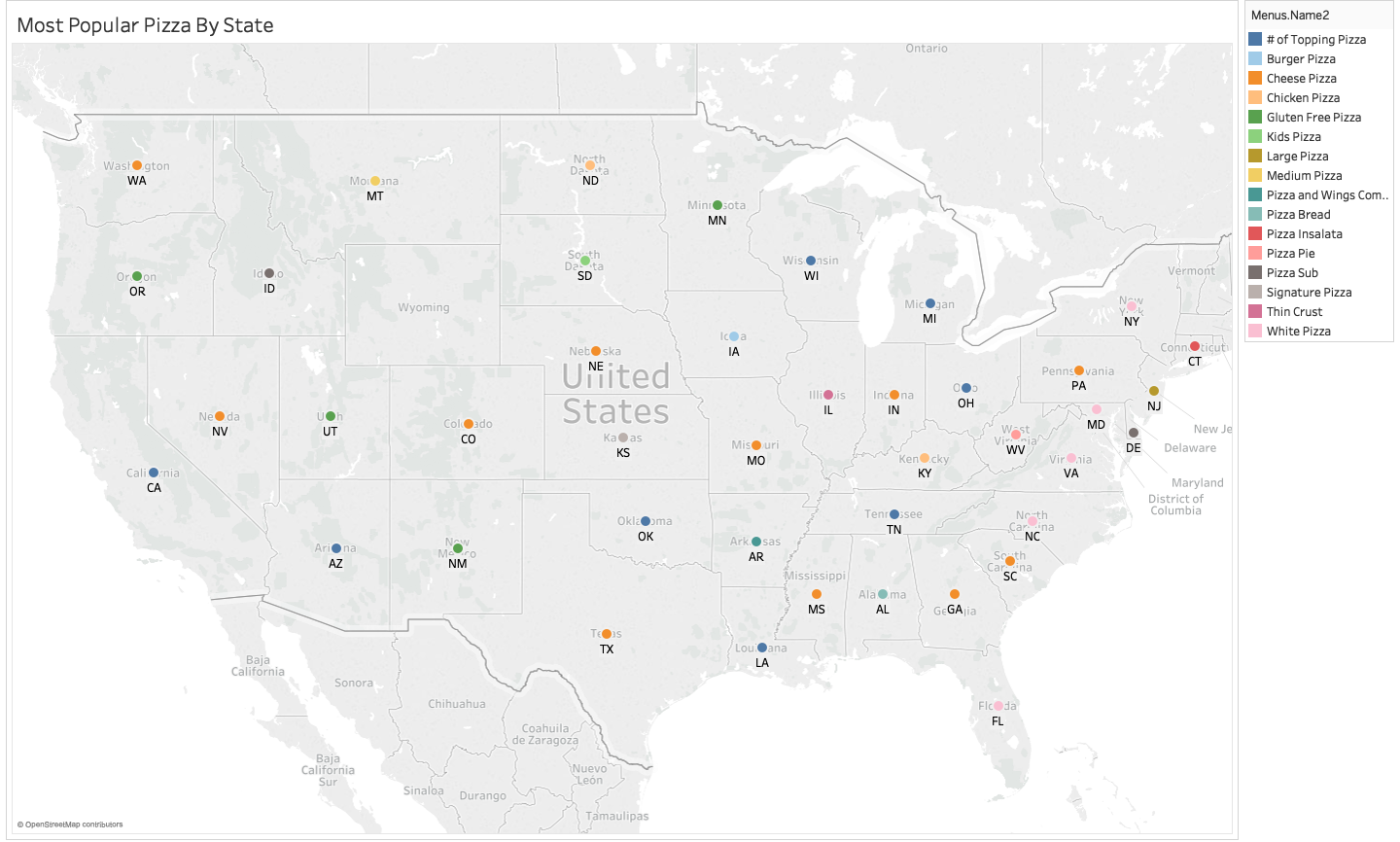
For the category of “Cheese” the most popular topping was Mozzarella, as expected. The least popular cheese was Gorgonzola.

For the category of “Veggies” the most popular topping was Mushrooms, with Onions as a close 2nd. The least popular was Broccoli.

For the category of “Meat” the most popular topping was Pepperoni, and the least popular was Ground Beef.

## Most Popular Pizza By State

Included in the dataset was information related to the location of each restaurant where the pizza was sold. It also provided the menu name for each pizza item in the dataset. I performed some data preprocessing to clean up the menu name column by grouping the names into larger categories. This preprocessing allowed me to identify the most popular pizzas in each of the 44 states where data was collected.

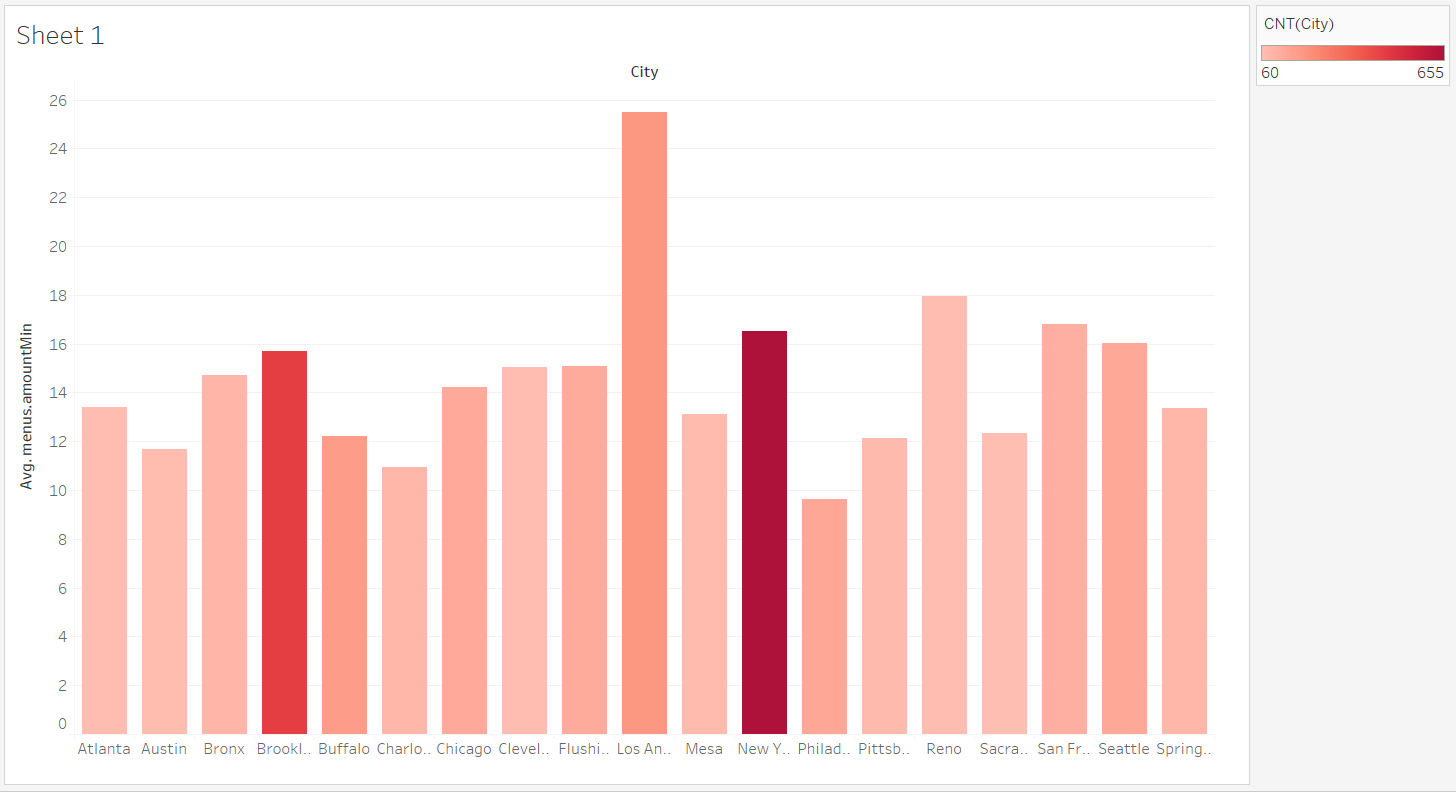


A visualization was then created using Tableau to show the most popular pizzas on a map of the United States. The data was encoded using color to represent the type of pizza, and position to show which state the pizza was popular in. There is an interactive element to the map so when the viewer hovers over the mark, a tooltip will appear. This tooltip shows the state name, the name of the most popular pizza, and a count of the number of times the pizza was on a menu in that state.

The most popular pizza in the United States was cheese pizza, and it was the top choice for 12 states. In addition to cheese, other popular pizzas included chicken pizza, white pizza, and thin crust.

## Price and Count by City

The dataset that we looked at had lots of information concerning the price of pizza. We decided that it would be interesting to look at how much the pizzas cost all across the country. So we took the information and made a simple bar chart in Tableau.



The columns each represent a different city. The length of the column tells us how much the average pizza cost in that city. As you can see from the graph, Los Angeles charges more per pizza than anywhere else. You probably could have guessed that, but now we have data to prove it!

We also wanted to include information concerning the amount of pizzas consumed in each city. In the top right hand corner of the chart, there is a scale that explains the color of the bars. Because the data set was so big, we only wanted to include the cities that had at least 60 pizzas sampled. This can be seen in cities like Atlanta and Sacramento. On the flip side, the darkest two bars represent the cities of New York and Brooklyn. New York City had over 600 pizzas samples.

## Conclusion

Based on our visualizations, we can see that pizza is pretty popular all around the country. From veggie pizzas, to meatzzas, and everything in between, everyone likes a good old slice of pizza pie. And nothing is more iconic than the pepperoni pizza.

One of the major discoveries we found was that pizza preference is impacted by where you live. Your city not only has an influence on how much pizza you eat, but which toppings you like as well. Isn’t it interesting to think that your environment has an effect on what you eat? Not only that, but your city determines the price you’re going to have to pay for a pizza. Meaning that Los Angeles natives are going to have to fork over almost twice as much dough for their dough.

The next time you’re eating a piece of pizza, stop and consider some of the visualizations that we talked about. Do you fall in the norm and stick to the tried and true pepperoni and cheese pizza? Or do you push the boundaries and try crazy toppings?

**References**

1. Data Stories: Delicious Visualizations About Food | Transforming Data with Intelligence. https://tdwi.org/articles/2016/10/19/data-visualization-food.aspx. Accessed 14 Dec. 2019.
2. Food+Tech Connect Top 5 Food & Health Infographics of 2012 | Food+Tech Connect. https://foodtechconnect.com/2012/12/28/top-5-food-health-infographics-of-2012/. Accessed 14 Dec. 2019.
3. Yau, Nathan. “Soup-Salad-Sandwich Space.” FlowingData, 17 Dec. 2018, https://flowingdata.com/2018/12/17/soup-salad-sandwich-space/.
4. Yau, Nathan. “Best Burger Ranks.” FlowingData, 6 Aug. 2019, https://flowingdata.com/2019/08/06/best-burger-ranks/.
5. “The Rhythm of Food — by Google News Lab and Truth & Beauty.” The Rhythm of Food, http://rhythm-of-food.net. Accessed 14 Dec. 2019.
6. Transforming Data Visualization into a Food Recommender. https://medium.com/tokopedia-data/transforming-data-visualization-into-food-recommender-c67e4c3db4b. Accessed 14 Dec. 2019.
7. Metcalfe, John. “A Sumptuous, Surreal Dinner of Edible Data Visualizations.” CityLab, http://www.citylab.com/design/2015/09/a-sumptuous-surreal-dinner-of-edible-data-visualizations/404353/. Accessed 14 Dec. 2019.
8. Data Cuisine | Exploring Food as a Form of Data Expression. http://data-cuisine.net/. Accessed 14 Dec. 2019.