DESIGN DOCUMENT

of

FOOD BUSINESS DISTRIBUTION ACROSS NEW YORK

Introduction

The objective of converting our opinions on a specific topic into a visual was the underlying aim of the project. Graphics that suggests a motive behind it is much more influential to society than the analysis of it. Looking back at the work done on the project "Food Business Distribution across New York", this design document conveys the choices made and enacted that had a lot to do with expressing the significance of the presence of educational institutions in a county.

There are two main visualizations in this project. The first one is titled "Food Establishments with respect to universities and tourist places" and the second one is titled "Food establishments types and their ratings". While the first one is the face of the project, the second one supports the claim by proving that the quality of food produced is, in fact, not unhealthy and hence, exposes the popular myth that "outside food is unhygienic". As we will see in the later part of the document, with most food establishments being given an 'A' grade, it stands to reason that college-going students and professors do not face adverse effects after eating from the food place. This holds a consequence in the reasoning that the more the number of 'A' grades, the better the business.

The project in its entirety was implemented in Altair, which is a visualization tool that is built on top of Vega-Lite. Each and every single aspect of both the visuals had a voluntary choice behind them, taking into account the significance it has on the overall theme.

These are the topics that will be discussed in this report, and the effect it had on making the design choices:-
- Data
- Design specifications
- Visual encodings
- Interactivity
- Conclusion
- References
- Visuals

Data

The project was built on 4 data sources:- food inspection ratings across New York state, universities and colleges present in the state, famous tourist destinations across the state, and the topojson file that was required in order to make a map. Let's look at them in brief:-

Food Inspection Rating: - This dataset was taken from the Open New York data website, [Food Safety Inspections – Current Ratings | State of New York (ny.gov)](), and later on exported into CSV. This dataset was updated on March 13, 2023. It consists of 13 columns and 183481 rows. We won't go into much detail about all the columns, however, an explanation of some columns is pivotal to understanding the visuals:-
- *County* - County of the food establishment
- *Inspection Grade* - This dataset was created based on the ratings given by the Department of Agriculture and Markets and any food business is given one of the three grades:- A: when the food business doesn't have any critical deficiency; B: when some deficiencies were found but rectified at the time of inspection; C: when there are critical deficiencies that could not be corrected at the time of inspection.
- *Establishment Type* - A food establishment can have one, two, or many types depending on the business operations conducted at the geographic location. A: Store; B: Bakery; C: Food Manufacturer; D: Food Warehouse; E: Beverage Plant; F: Feed Mill/Non-Medicated; G: Processing Plant; H: Wholesale Manufacturer; I: Refrigerated Warehouse; K: Vehicle; L: Produce Refrigerated Warehouse; N: Wholesale Produce Packer; O: Produce Grower/Packer/Broker, Storage; P: Controlled Atmosphere Room; Q: Feed Mill/Medicated; R: Pet Food Manufacturer; S: Feed Warehouse and/or Distributor; T: Disposal Plant; V: Slaughterhouse; W: Farm Winery. It is important to note that the primary focus was to display food businesses that were manufacturing food, C.
- *Trade Name* - The name of the food business.
- *Georeference* - The geographical location of the establishment.

Educational institutions: - This dataset consisted of the name of the college, along with their geographical coordinates. This dataset did not have the latitudes and longitudes of the places but had to be extrapolated by an external Python library called GeoPy. There were a total of 196 universities, extracted from the website [New York Universities and Colleges]().

Tourist places: - The dataset of the 15 most famous tourist destinations across the New York state, and its geographical coordinates.

New York State geographical specifications(topoJSON): - In order to create a geospatial representation with any visualization tool, we need a JSON file that encodes the topology of the structure. From here, we can plot the state of New York, and also, get the coordinates of counties, that   would   be   visualized   by   encoding   the   geoshape.   The   data   in

https://raw.githubusercontent.com/deldersveld/topojson/master/countries/us-states/NY-36-new-york-counties.json was used. From this data, the centroids of each county were calculated, which would be used to label the county names.

Design Specifications

There are two main visuals in this project. We will discuss both of them separately. The first is a geographical map of New York State, with areas enclosed within the white outlines depicting counties. There are in total 62 counties in New York, and outlining the map based on counties communicates much more effectively, as compared to being outlined based on cities.

Marks: - On top of the map, point marks have been used. It consists of 2 square points and 1 circular filled point. The two square markings represent the educational institutions and the tourist places respectively. The filled point represents the food establishment at that location.

Visual channels: - Color, shape, and text were the channels used here to represent the elements of the chart. Text is used here to represent the county names.

Guides: - The guides used here are the axis labels, title, and subtitles. These annotations help in presenting contextual information about the work. The name of the county is specified to the user by the label of it.

The second visual is a combination of interconnected dynamic graphs that represents the proof of said claim. Here there are 2 bar graphs, a pie chart, and a map of the New York state.

- The first bar chart, which is titled "Establishment types", is a normalized bar chart with 5 different food establishments: ABC, ABCH, ACD, ACK, and CDK. They are all encoded with different colors: yellow, black, orange, cyan, and purple, respectively. The reason for deciding on this color set was that they were quite clearly distinguishable from each other when all of them were selected, except the region of New York City, which has all five establishment types.

- The second bar chart, which is titled "Corresponding Bar chart representation", is a normalized stacked bar chart that shows the ratio of the count of businesses possessing 'A', 'B', and 'C' grades, of that particular establishment. The colors chosen were the primary colors, green, red, and blue, respectively. The reason for having all the bars normalized was so that we can infer the higher number of As than Cs.

- The pie chart distribution in the middle, also represents the same thing as the stacked bar chart. This chart has the same color characterization as that of the stacked bar chart.

Marks: - On top of the map, there are 5 points that are circular and filled. They represent the locations of the food establishment.

Visual channels: - Color, shape, and text were the channels used here to represent the distinction between the elements. Text is used here to represent the county names.

Guides: - The axis labels, titles of all four sub-graphs, and labels of the counties in the map, are the guides employed here.

Visual Encodings

For both visuals, data is encoded and mapped to the visual channels with respect to its color. For geometrical and spatial mapping of data, the position is encoded with it. For the visualization of "Food, Tourist, and University distribution graph", the shape is used as the encoder for representing universities and tourist places. They hold the most important role in this project since small clusters of black points are seen where there are yellow squares. This aspect of togetherness underlines the fact.

The color combination of light blue(#52b7ee), yellow, dark green(#4b5320), and black was chosen specifically to highlight the tightness of educational institutes with the food businesses. Black represents the food establishments and the colleges are on top of it. A light color like yellow was chosen due to the fact that, if by any chance, there happens to be a food business in the exact location of a college, then the food spot would be visible, and not obscured. The presence of light blue representing the map area has a causality of making the yellow marks 'pop out'. The map has a light brown(#fed0b1) background to it, that was in tribute to W.E.B DuBois's style.

In the second visual, "Food Establishments types and their ratings", all the food establishment types are encoded with different colors: yellow, black, orange, cyan, and purple. This color set doesn't exaggerate the significance of any one establishment, in spite of being clearly distinguishable from each other, especially when all are selected. The map has a light grey fill since the color set of the food establishments demands the necessity of having a background color that justifies the purpose mentioned in the previous statement. Cyan is the only color that stretches the eyesight of the viewer, but when selected one can see that it is mostly present around New York City region and only a small number of those that are scarcely visible in other parts of the state. The primary reason for this is that K(vehicles, i.e., food trucks and carts) are generally in more population in NYC.

For both the graphics, text is also used as an encoding, for the representation of the county names, titles, axis labels, and subtitles. The particularity of text annotation holds an important role because the meaning of food establishment types in the second chart and color codes in the first chart holds the key to understanding the visual.

Interactivity

Altair has the ability to make a visualization interactive by allowing the creator to include certain transformations and selections in the charts. Tooltips are used almost everywhere in this project. All the marks, squares and points, are affordances that provide the user with the indication that by hovering over it, the details of that mark are visible. The guide on how to be aware of these affordances is mentioned as the text annotation in the subtitles of each visual. Through the utilization of tooltips, the need, and also the significance of a conventional legend is diminished.

In the first visual, these are the interactivities employed:-
- On yellow squares(educational institutions): - Tooltip that describes the Name of the school
- On dark green square(tourist places): - Tooltip that describes the Name of the place
- On black points(food business): - Tooltip that describes the establishments' County, City, Inspection Grade, Establishment Type, and Trade Name
- On any empty space inside the map: - Tooltip that describes the Name of the county

In the second visual, these are the interactivities employed:-
- On the bars of "Bar chart representation": - Tooltip that describes the food establishment's Inspection Grade and its count.
- On the sectors of "Distribution of Inspection Grades for the respective food establishments": - Tooltip that describes the food establishment's Inspection Grade and its count.
- On any empty space inside the map: - Tooltip that describes the Name of the county
- On filled circular point: - Tooltip that describes the establishments' County, City, Inspection Grade, Establishment Type, and Trade Name
- With the use of click selection, the transformation in the map points is made. The bars in the "Establishment Types" can be selected, and their corresponding food inspection grade count distribution can be seen in the pie chart. Those points that are not of the selected establishment types would become invisible on the map, allowing the viewer to see the regions with more of their population.

Conclusion

In this project, the work that has been produced reflects the influence the education industry has on food businesses. It is also important to note here that only those food establishments were included that had a "C" in them. Meaning only food manufacturers were included. If we were to take into account, the entire dataset of "Food Inspection Rating", the underlying message would only be strengthened. The design choices were made, keeping in mind the aspects of what makes a visualization effective and expressive. Keeping in mind Tufte's principle of not having excess chart junk in our visualization, the project had a sense of optimality to it. Decorative features like animation and stylish gradients were purposefully not included since it does not relate to the story. By leveraging the power of visual representation, the aim of using this as a platform for communicating a story was achieved. Moving forward, let this design document reflect the choices made to make the viewer understand the design.

References

1. [Food Safety Inspections – Current Ratings | State of New York (ny.gov)](#)
2. https://college.ai/Colleges/New-York-Colleges.html
3. https://raw.githubusercontent.com/deldersveld/topojson/master/countries/us-states/NY-36-new-york-counties.json
4. [Vega-Altair: Declarative Visualization in Python — Vega-Altair 5.0.0 documentation (altair-viz.github.io)](#)
5. [Welcome to GeoPy's documentation! — GeoPy 2.3.0 documentation](#)

Visuals