Portfolio

Summation

Both projects were based on the idea of speaker data on News/Opinion broadcast media. This data source was chosen early as a focus, though quite a bit of the designs were in flux until the very end.

Challenges Encountered

Data collection was a huge issue. Hand coding transcripts was a much more intensive project than initially predicted. Once the data was collected and scored, and stored in multiple data files. It then had to be joined, which also was more difficult than anticipated.

Exploratory Project

Overview

The basic idea of this project was to visualize the turn taking and order of speech in broadcast media. This would be visualized across multiple diverse types of shows and networks.

Data

The data was sourced from transcripts from EBSCO, Factiva, and NPR from News Shows that aired in September 2019. It was then processed by a custom python script to resolve it down to a reduced data format. This reduced data format consisted of the speaker tag, the length of the utterance, the longest word in the utterance. Each speaker was then looked up and categorized as a “Recorded,” “In-House,” “Guest,” or “Anchor” speaker. If a speaker was in one segment a recorded clip, and in another segment a guest this was accounted for by subtyping the speaker tag. Gender was inferred by nearby pronoun use, or if needed a web search (there were no non binary genders, though some were indeterminate, mostly because they were in recorded snippets and tagged as “Speaker #1” or similar).

Design Process

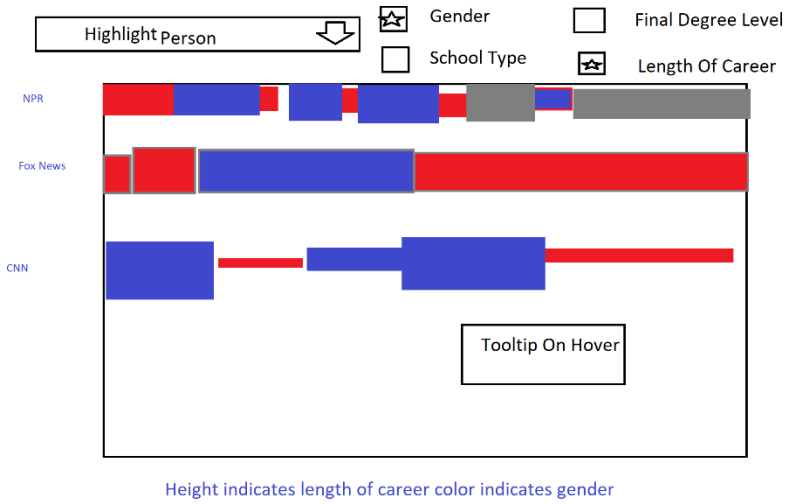


Fig. 1 an early prototype

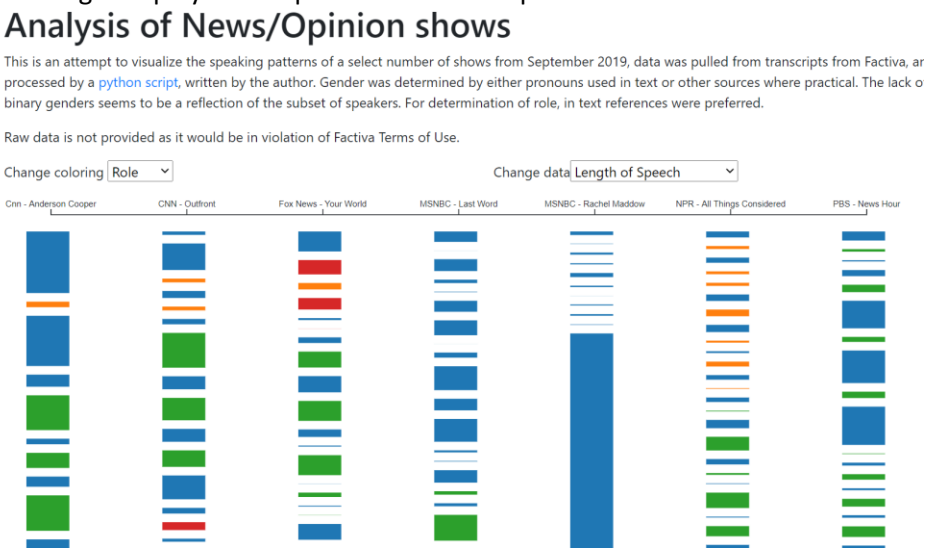
Early on the idea was that more aspects would be presented in the graph and it would be oriented horizontally with an ability to highlight individuals. Initially, coloring would be determined by a selectable factor, with horizontal distance reflecting a scale of utterance and there would be an alternate selectable scale based on the height of the bar (such as length of career). In addition, there would be a tooltip with speaker information.

This swiftly resolved itself in workshops to being over worked for the task at hand. The multiple x and y scales combined with hue-based information and extreme flexibility certainly would have allowed a depth of interaction, but at the cost of accessibility.

In the next iteration, the option to highlight a person was dropped as was the alternate scales. This simplified the presentation. Workshopping this version went better, and people wanted to see the results.

In actual development, it was found that the horizontal format did not allow enough data on the page without a large amount of horizontal scrolling which is not a preferred UX modality. It was also decided to focus on speaker type, because the initially planned default of gender was not terribly

interesting a display from a presentation standpoint.



Narrative Project

Overview

This project's focus was a deep inspection of speaker data and their aspects and relationships to their education. This would be visualized with detail by network, and speaker type. Text would support and ground the user in the context of the wider US, as many people do not have a good grasp as to education characteristics of the United States as a whole.

Data

The data was both a subset of the data from the exploratory project but also an expansion upon it. Recorded speakers were removed as several of them were either not public figures or were figures that might not be considered affirmatively selected by the networks as they were "in the news". This data was supplemented by data on each speaker's educational background, including highest educational attainment, undergraduate college, undergraduate major, graduate college, graduate major, and age. If multiple degrees were earned the "highest" one was selected for graduate degree using the ranking Doctorate, Professional, Masters, as that will generally align to the years needed to receive the degree.

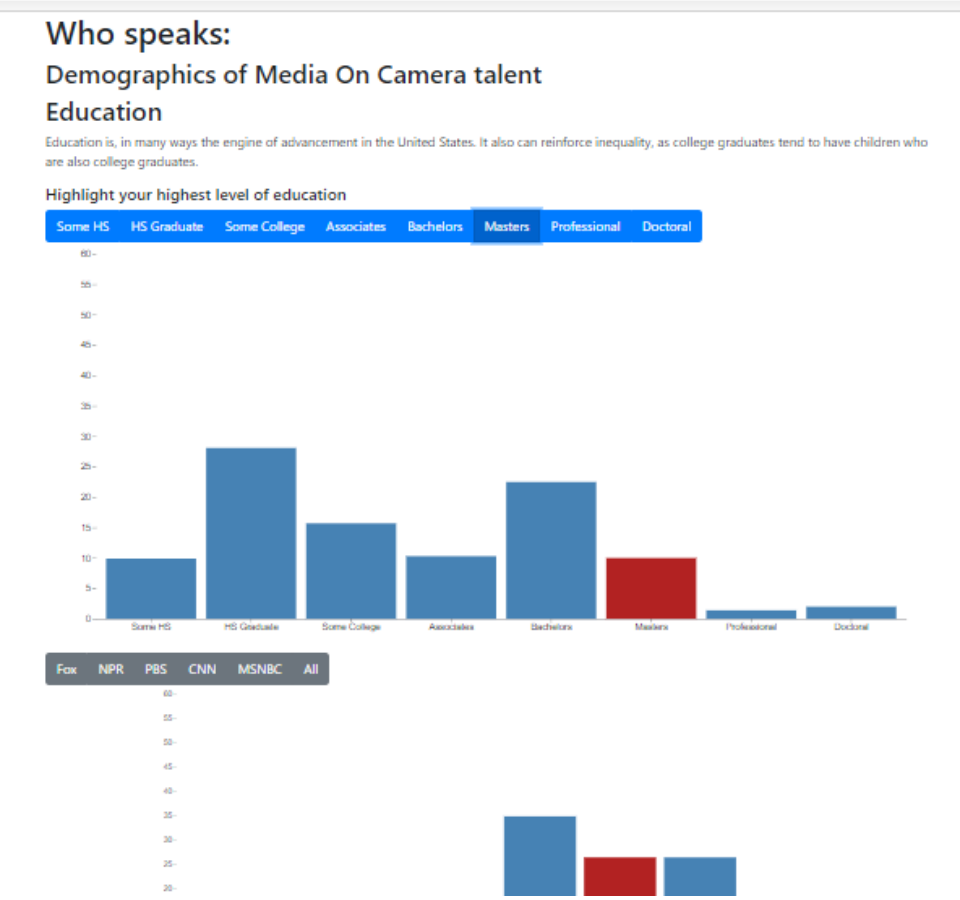
This data was then supplemented with Census data and data on college admissions. This was in comparison a breeze to source and deal with.

Design Process

Broadly the plan from the start was to "locate" the user in the data, so that they have a better understanding of where they are, where the country is, and where the target population is. This was to be done based on simple questions at the start of each section. In the initial mockups the structure was a series of accordions with each panel locked until the prior panel had been answered. Under each question would populate text tailored to the response (e.g., While it can feel lonely to have **some college without a degree,** but in fact this is not an uncommon situation, especially as some of the Associates degree holders were from four-year programs and ended up with a two-year degree because of life events)

In the end it was felt that this enforced interaction was simply too heavy a preference to be placed upon the user with too little gain. Also, the sheer quantity of text to write, much of which would never be seen was daunting. So, it was moved to a more traditional article layout with headings, some of which still have grounding questions.

Simple single level categorical data was displayed on bar charts in line. In line was chosen as it is more responsive and keeps the same data in the same line. Scales were kept the same within graph pairs to avoid confusing the reader. Multi-level data such as college qualia was displayed in a tree map with only coloring based on the first most relevant split.



Further Development

Obviously in a perfect world many more speakers and programs (and possibly even networks) would be added. However, this turned out to be very time consuming to collect. It might also be interesting to see time slices showing differences between programs that are say a decade apart.

For the exploratory project it would be helpful to display the actual text of each utterance. This was technically doable, but would have run into licensing and copyright concerns. Also, more detailed information on the speakers would of course allow a deeper dive into the data. One datum that had to be dropped for time was length of career, which might prove very interesting.