All the Queens Voices:
An Oral History, Visualized

Stephanie Shieh Yung

May 2019

Submitted in partial fulfillment of the requirements for the
Master of Science degree in Data Visualization at Parsons School of Design

Abstract

Oral history is a method of historical inquiry rooted in dialogue between an interviewer and an interviewee. The practice predates the written word, yet persists to the current internet age. Oral history bridges the gap between what is captured in history books and the individual experience. As oral history has shifted into the digital age, rich debate has ensued surrounding best practices and new questions and issues continue to arise as technologies advance. One of these touch points is on how far oral history can and should stray from its rooting in the oral and aural tradition. This project seeks to explore whether oral history, a fundamentally oral/aural experience, can be visualized. Partnering with Queens Memory, an oral history program based in the Queens Public Library system of New York City, interview sound files and interview metadata were collected, analyzed and visualized.

Table of Contents

Introduction

In July 2018, my entire family - myself, my husband and our three small children - attended an event at the Flushing Library branch of the Queens Public Library, hosted by an organization called Queens Memory. We were invited to this community event by a friend of ours who was an MFA candidate at Queens College specializing in fiber arts and creating work grounded in and in response to her community. As part of an initiative called *Memories of Migration*, community members were invited to share stories of their family's migration and design quilt blocks representing these stories. Participants learned various quilting techniques to sew individual quilt blocks and their migration stories were recorded onto Arduino LilyPad devices that were sewn into the finished quilt. I was intrigued by the work of Queens Memory -- Queens County is the locale of my birth and childhood and the borough I have lived in for the last eleven years. My childhood was spent in Flushing, where immigration beginning in the 1970s rapidly evolved the neighborhood into New York City's largest Chinatown and my post-undergraduate years were spent living a block away from the Museum of the Moving Image and the historic Kaufman-Astoria Studios. I have fond memories of learning to bike to our neighborhood park, riding the "redbird" trains along the elevated 7 subway line and vague recollections of public school yards and Flushing Meadows Park. However, I also spent my formative undergraduate years on the Upper West Side of Manhattan and Queens had always been somewhat lackluster compared to the epicenter that is "The City" and in recent years, the hipster allure of Brooklyn.

I regret now not having more thoroughly appreciating the richness of life and history in my home borough. In conducting research for this thesis, I unveiled a Queens history unbeknownst to me -- historic Dutch settlements in Newtown, the existence of the oldest dwelling in New York City that is still a dwelling (the Lent-Riker-Smith Homestead), changes in housing laws that have left so few historic traces today. Though I experienced remorse at the almost-complete eradication of Queens' rich history, the dearth of historic homes and scant number of preserved neighborhoods, I also came to realize that this dramatic and almost complete change is what paved way for the diversity that exists in Queens today.

This project seeks to visualize the richness of the oral history interview data Queens Memory has gathered, to celebrate and illuminate the richness of story Queens Country has to offer.

Background

Queens County

Queens County is New York City's largest geographical and second most populous borough: it is bordered by the East River on the west and north (across from which lie Manhattan and the Bronx, respectively), Nassau County on the east and Brooklyn on the south. It is topographically diverse, encompassing Astoria's rolling hills, Jamaica's natural bay, and Far Rockaway's sandy shoreline. Queens has the most diversified economy among the five boroughs and is home to two of the world's busiest airports - John F. Kennedy International Airport and LaGuardia Airport. Queens housing stock ranges from Queensbridge Houses, the largest housing project in America; to quaint homes in tree-lined, nationally-recognized, historic neighborhoods in Sunnyside and Jackson Heights; to large, suburban, single family homes in Little Neck and Bayside. Queens is diverse in all the aforementioned and countless other ways, but most notably, and arguably, most importantly, Queens is diverse because of the people who choose to call the borough home.

According to the U.S. Census Bureau's July 2018 population estimate, Queens County's population is 47.5% foreign born, with 56% of persons over the age of five living in a home where a language other than English is spoken. The Census Bureau reports 138 languages spoken in the borough, though some linguistic experts believe that over 800 languages are spoken in Queens, including many "endangered" languages, expected to no longer exist in 20 to 30 years. No racial or ethnic group holds a 50% majority in the borough - this ethnic, linguistic, country of origin diversity has long led Queens to be referred to as the most ethnically diverse place in the world.

Queens Public Library

The Queens Public Library (formerly known as the Queens Borough Public Library and Queens Library), the library system serving the borough, is one of the largest public library systems in the United States, serving 2.3 million people at its 62 branch locations. In April 2019, during the course of this thesis work and writing, the Queens Public Library added "Public" prominently to their name, unveiling a new brand identity and stating:

> *No matter who you are, where you're from, or where you want to go, we speak your language.*
>
> *We're excited to announce our renewed promise to the public and our new look to reflect who we are and the exceptional service you can expect from us.*
>
> *We put "public" at the center of our name to reinforce who is at the center of our work and to whom the Library belongs.*

From personal experience and through conversations conducted in the course of undertaking this research, the Queens Library strives daily to fulfill this mission statement, acting as a focal point in communities, curating programming and tailoring circulated materials to meet the unique needs of each community they serve.

The Archives at Queens Public Library is a historic, special collection begun in 1912, currently housed in the Central Library building in Jamaica, Queens, that focuses on collecting and preserving materials documenting the social, economic, and political history of Brooklyn, Queens, Nassau, and Suffolk counties. The collection consists of books, publications, current and historical newspapers, family manuscripts and genealogical material, historical maps and atlases, late 19th- and early 20th-century photographs, as well as other archival material.

Queens Memory

The Queens Memory Project is a *digital* archive program supported by Queens Public Library and Queens College, CUNY, which aims to record and preserve contemporary history across the borough of Queens. A self-described mission of the program:

> *Queens Memory empowers residents from diverse backgrounds to document the personal histories that together tell a more complete story of life in the borough.*

One meaningful way Queens Memory documents history is through the recording and archiving of oral histories. Queens Memory provides training and equipment to volunteer oral historians to document oral histories from their families, neighbors and communities.

In the guidance provided by Queens Memory, suggested topics for these interviews are outlined: Family Traditions and Lore (migration stories, how interviewees decided to settle in Queens, holiday or other traditions); Education (educational history of

interviewee and family, assessment of school quality and safety now and in the past, educational expectations); Work and Leisure (occupational history of interviewee and family members, role of work/career in interviewee's life, hobbies and interests, religion); Getting Around (modes of transportations now and and in the past, commuting experience, changes in transportation through the years); Home (reasons for moving to current home, memories of past homes, descriptions of home, special features); Neighbors and Neighborhood (level of familiarity with neighbors, identity and divisions, density and home ownership change over time, civic/political/social connections to neighborhood, memories of childhood play in neighborhood); Change Over Time (biggest change in neighborhood, population changes, changes in local shopping and services); Landscape (ways parks and public spaces have changed over time, ways community uses public spaces); The Future (predictions for change over next year, five years, decade, best or worst things that could happen to the neighborhood in the future); Links to History (to be drawn from interviewer's research of the subject's life and neighborhood)

What is Oral History?

The Oral History Association states that oral history refers both to a method of recording and preserving oral testimony and to the product of that process. A simple, but encompassing understanding of oral history is "narrative construed from memories". There are several recognized benchmarks of oral history, including: evidence of thoughtful planning and research before the interview is recorded; a structured, well-researched interview format; a clear statement of who each interview participant is; a controlled interview setting; use of high-quality recording equipment; attention to ownership of interview information; adherence to careful processing techniques; preservation in a designated repository; and access to interview information. However, the field of oral history practitioners is vast - volunteer interviewers like those utilized by Queens Memory, labor historians, collectors of oral tradition, social scientists, broadcasters, museum staff, teachers and the students, and as can be expected, benchmarks are not always adhered to.

Oral history's intrinsic and unique value is the way it places individual experiences and memory within a larger social and historical context. The oral history interview becomes a record documenting past experience, individual and collective. Oral history relies on memory -- that ability of human brains to encode, store, retain and recall information and past experiences -- and captures recollections that are filtered through a lens of a changing experience.

The heart of an oral history interview is the collaborative and dynamic relationship between the interviewer, who has ideally conducted careful research and preparation of questions, and the interviewee (or narrator), who answer those questions based on what they, in the moment, deem (and remember to be) relevant, meaningful, or appropriate to share.

Though the oral history critically relies on, depends on and informs the interview, it is not the interview alone that situates oral history. The *Oral History Life Cycle* in Figure 1 (adapted from *Practicing Oral History in Historical Organizations*) shows how the interview is part of a cohesive process.

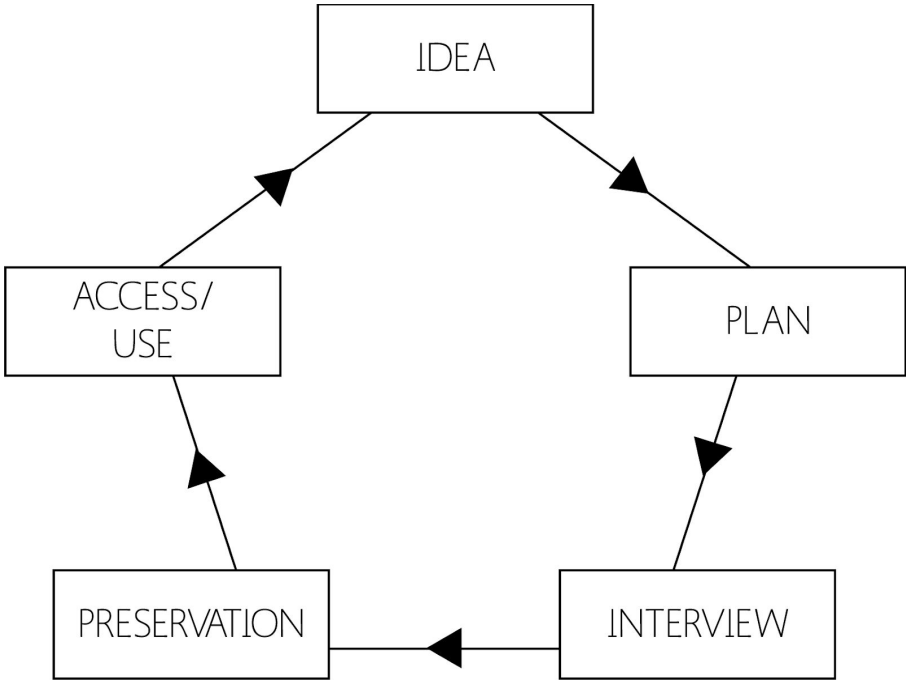


Figure 1. Oral History Life Cycle, adapted from *Practicing Oral History in Historical Organizations*

This holistic approach moves from "...conceptualization to creation to curation". The process does not start and end with solely the interview itself, rather, the interview is part of a series of actions and activities, which, when guided by appropriate benchmarks, work to create and preserve the interview, create both the interview and the information in it (metadata) and make the interview accessible on an ongoing basis.

Oral History in the Digital Age

As recording technologies rapidly evolved, the entire oral history field has had to redefine best practices in response. Critical points of debate regarding access and preservation have emerged. Technological advancements have resulted in digital video and automated transcription services becoming markedly more affordable, however, the field of oral history remains divided over its role and use, with some arguing that the visuals created by video intrinsically change the experience and purpose of oral history and others arguing that visual aids like video and transcripts allow for more greater accessibility.

The Oral History in the Digital Age project summarizes the current conundrum:

> *Overall it is a daunting challenge to reach consensus and clear directions. Museums and libraries are at the forefront of these issues as they are the world's repositories for cultural heritage, preserving and providing access to past, current, and future oral histories for the peoples of the world. Thus, there is a pressing need to build a sustainable, authoritative and collaborative framework that will put museums, libraries, and oral historians in a position to address collectively issues of video, digitization, preservation, and intellectual property and to provide both a scholarly framework and regularly updated best practices for moving forward. This site is dedicated to taking on this challenge.*

Another aspect to consider is the role of oral history in the creation and curation of information in the *information age*. Apart from it's inevitable intersect with technology, what is the meaning and purpose of oral history in the current information landscape? Stephen M. Sloan lays out the following characteristics of the "immutable core" of oral history in *Swimming in the Exaflood: Oral History as Information in the Digital Age*: firstly, that it is a linear, long-form source. "In an age where digital information is commonly characterized by its truncated delivery, the long form of oral history is increasingly important." (Sloan, 181) Oral history offers a deep experience, one that is seemingly antithetical both the current scroll-by Insta-verse and 280-character Twitterverse. Sloan urges oral history practitioners to cling to the long form, and resist segmenting and excerpting, as these can distort the original narrator and interviewee intentions.

The beauty and benefit of listening to and truly experiencing oral history - hearing first person narratives with all their inflections, emotions, subtleties and nuance - is why oral histories are recorded in the first place. However, historically, oral history projects have been stored in archives or repositories at universities or libraries, with the hope that

researchers can make meaningful use of the curation of this data. With online oral history collections now available, access is no longer limited to academia, however, it is still limited to those with the invested interest and time to engage in the source material. Visualization, whether it be video, animation, or illustration, broadens the audience and allows stories to be more accessible to a wider public.

StoryCorps, the "born digital" story-collecting and preservation project is firmly in the American zeitgeist, with nationally broadcast clips on NPR and over 75,000 interviews recorded, with a heartfelt mission to "...preserve and share humanity's stories in order to build connections between people and create a more just and compassionate world." StoryCorps was once described as "nothing less than an oral history of America", however it has faced criticism from the oral history community on several issues: the lack of training of the interviewer, the undefined roles of interviewer and interviewee (many StoryCorps interviews flow in a conversational manner, with the "narrator" role shifting), the 45-minute time limit imposed, the issue of control and ownership (StoryCorps retains all intellectual property rights to materials) and access (though clips are accessible on their website, full interviews are archived, accessible only by visiting the Library of Congress. "By prioritizing collection over curation, StoryCorps provides a model of opportunities lost." (Tebeau) StoryCorps seems to have shied away from self-describing as an oral history project - the term is no longer on their main about page, relying instead on phrases like "collecting stories" and "sharing the wisdom of humanity"; though reference to oral history can still be found in their "Fast Facts" section:

> *Since its founding in 2003, StoryCorps has collected and archived nearly 75,000 facilitated interviews from more than 150,000 participants from across the country who visit one of our recording sites. It is one of the largest oral history projects of its kind.*

StoryCorps provides an intriguing example of how storytelling and an oral history-esque project can enter the mainstream and capture the broader imagination. However, is the tradeoff of increased and sustained public engagement worth the concessions to the oral history format and methodology? Is there an alternative where an oral history project that follows the field's standards of practice, can be equally as engaging?

Visualizing Oral History

Given the history and current debate, it is worthwhile to engage the questions: *should* and *could* oral histories be visualized?

On the first question of ethical responsibility and authenticity to the field, the debate may never end. There will always be purist practitioners who believe oral history should remain strictly an auditory experience and that visual interpretations and elements of an oral history interview detract from the aural form.

On the second point, oral histories have historically been visualized by elements such as post-interview transcriptions or long-form video interviews. Mark Tebeau's argues in Case Study: "Visualizing Oral History", that these typical accompaniments to oral history prioritize sight above all other senses. Tebeau also states:

> *Digital practice has uncritically adopted that trope in discussing and valorizing evidence for historical curation. And, arguably, our efforts to deal with visual evidence in archives, online exhibits, and other presentation have far outstripped our work in dealing with sound (much less the other sensory-based) materials. To a degree, this is ironic because some of the premier digital tools that have so altered daily life–cell phones and portable music players–have their origins in transforming voice and sound experiences, making them pervasive and deeply individualized.* (Tebeau)

The idea of elevating the aural sense in the digital realm, particularly as many digital tools were born out of improving voice and sound experiences, is compelling and most of Tebeau's case study addresses digital curation in a way that connects the aural experience via metadata tags, creating a richer experience of sound. He offers further guidance on points of commonality that will allow practitioners to pursue a "richer perspective on sound":

> a) recognition that oral history is fundamentally an aural experience and not just a text experience;
> b) oral history should be evaluated for meaning at the clip or segment level, not just at the level of the 60-minute or 90-minute interview;
> c) clips and segments should be connectable across interviews, collections, or archives;
> d) our metadata schemes, as well as our work in representing oral history to public audiences, have to account such rich metadata schemes (think linked open data);
> e) collecting oral history is one thing, making it open and accessible is another goal—and one that the oral history community should embrace;
> f) efforts at linked open data have to account for segment- and clip-level metadata; and

g) it is vital that we involve communities in processing and connecting to oral histories together, just as we are learning to involve those same crowds in collecting oral interviews. (Tebeau)

Oral histories *can* be visualized, and this next section provides a sampling of what others have created in this realm.

Visualization Examples

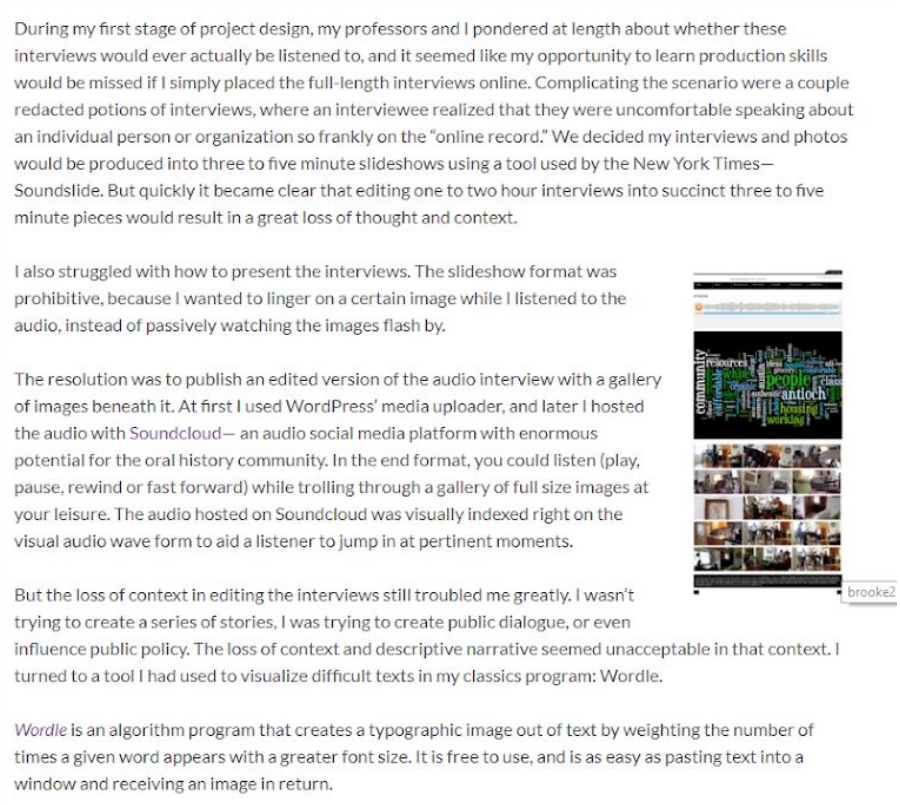


Figure 2. Why Here-Why Now, Oral History in the Digital Age

On the Oral History in the Digital Age site, Brooke Bryan provides an overview of her (now defunct) project chronicling "sound, sight & sentiment in yellow springs, ohio" (Figure 2), and of her lessons learned in curating a web-based digital oral history project. She describes the visual aspects of her project:

> *The Why Here/Why Now Project was always a work in progress… My overarching goal was to illustrate a living moment in my community by curating sense of place photography, semantic typography, and edited audio interview together in a convergent website. As I collected and created content, it fell into distinct categories: interview audio, interview images, my attempts to visualize the content of the narratives, along with images of my town, notes on my process, and other written pieces which were required for class and set the tone for the undertaking.*

She utilized photography and Wordle to curate visuals that accompanied audio interviews. Bryan describes how creating word clouds from one, then two, then three

combined interview transcripts shed light on themes and led to important discoveries of what was being discussed by her community members.

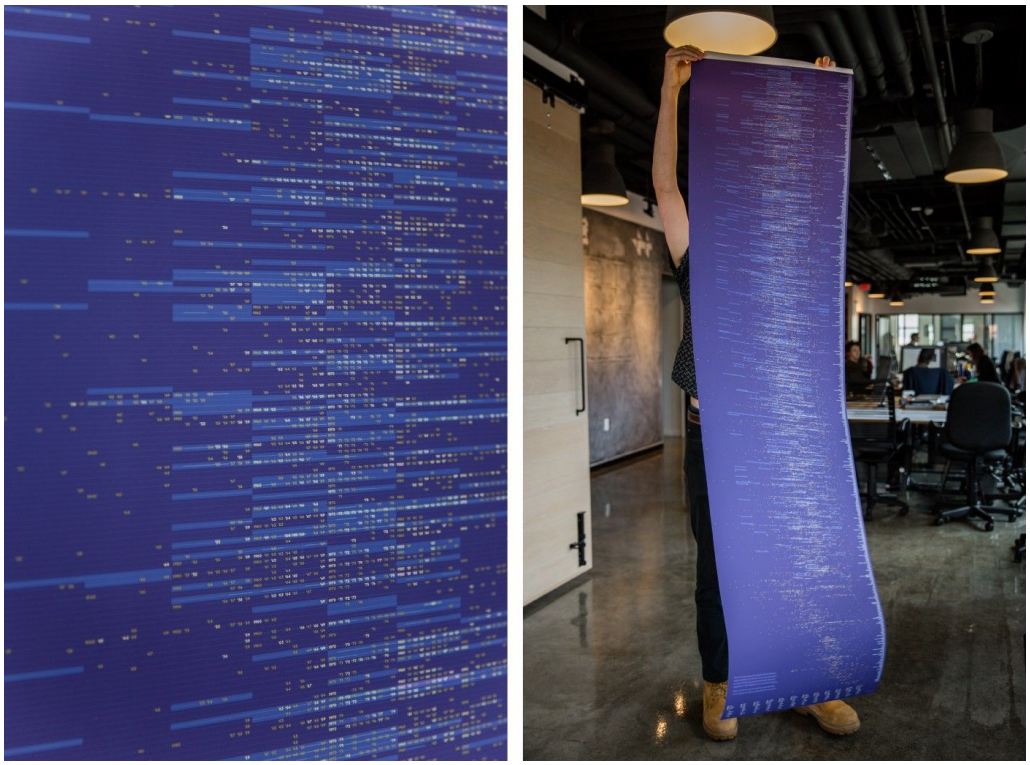


Figure 3. Fathom Information Design, Computer Memory: Visualizing a Century of Oral History

This visualization (Figure 3) from Fathom Information Design visualizes the Computer History Museum's interview archive, which includes 800+ oral histories of individuals involved in all aspects of computing over the last century. According to Fathom, "...our goal was to design an initial landscape overview of the documents. We wanted to find a way to see the nature, volume, and density of the content and connections within the archive."

The interview archive had full transcripts of interviews and using a combination of regular expressions and natural language processing, Fathom extracted the dates and entities mentioned in each interview, creating a tool that would allow the user to explore the relationships between the years mentioned in interviews as well as technologies, companies, and locations. Users would be able to filter by the top terms, pinpoint specific mentions, and dig into the surrounding context from the interview.

In addition to the digital tool, Fathom designed a poster that depicts a more "cartographic" view of the years mentioned in the interviews (Figure 3).

This project is beautiful and impressive in it's data-based methods, but it critically leaves out the "oral" component of the oral histories, mining data from solely text and resulting in an end product that is visually stunning, but lacking any aurality whatsoever.

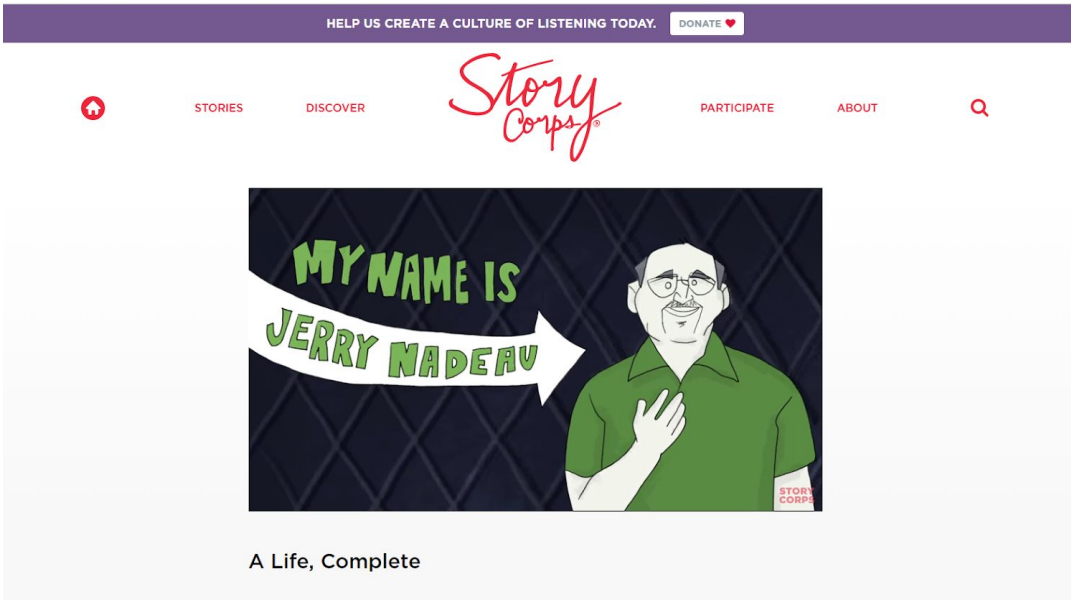


Figure 4. StoryCorps Animation, https://storycorps.org/animation/

In 2013, StoryCorps celebrated its tenth anniversary with its very first TV special, broadcast on PBS. "Listening Is An Act of Love," is an animated documentary featuring stories culled from the StoryCorps archives. In a CityLab interview, Rauch Brothers animators Tim and Mike Rauch spoke about how animations can add to the experience of the recorded stories. Mike Rauch spoke about how some stories were edited for radio and perfect in their form, but others had gaps to fill - additional information or background that would make the experience of the story more profound. For example, by showing the physical relationship between the persons recording the interview, or by enhancing the mood or the emotions being expressed. The animators craft "larger than life" characters from these real stories and with their cartoon-like artistry, magnify the experiences being shared in the StoryCorps recording.

This type of visualization is attention-grabbing and extremely engaging, but speaks more to the artistic capability of talented animators like the Rauch Brothers than illuminates the oral history form.

Figure 5. New Dimensions in Testimony, Virtual Conversation with Holocaust Survivors

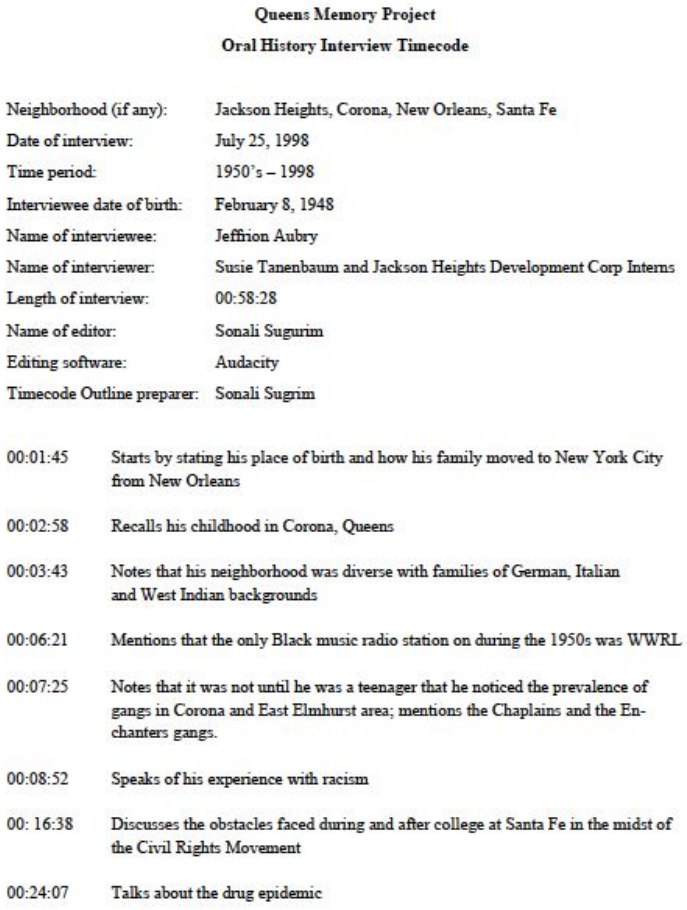
Of the visualized oral history examples I came across, New Dimensions in Testimony was the most boundary-pushing and technologically innovative. New Dimensions in Testimony allows current and future generations the opportunity to interact with Holocaust survivors through a virtual conversation. An interactive testimony platform created by USC Shoah Foundation and Conscience Display, Holocaust survivors are interviewed, answering thousands of questions, while being recorded in 360 degrees by seven cameras, enabling the testimonies to be later projected in 3D. The project uses natural language technology, which allows people to engage with the testimonies conversationally by asking questions that trigger relevant, spoken responses. The Newsweek who visited Pinchas Gutter's hologram at the Museum of Jewish Heritage in Toronto described his experience:

> *His image responded with answers—speech quirks, pauses and gestures included. He spoke to me about religion and sports; he shared his favorite Yiddish joke; I hear he sometimes sings. Gutter also told me that he was a happy child until September 1, 1939, when Hitler's armies invaded Poland and World War II began…The effect of talking to these virtual people was startling and eerie. I didn't for a minute forget that they were images, but I found myself deeply moved by both.*

This type of visualization project, so thoroughly impressive in scope and scale, is also time, labor and cost intensive, leagues beyond what the overwhelming majority of current oral history organizations and projects are able to achieve.

Data Collection

Data was collected over the course of several in-person visits to the Metadata Services Division at the Central Branch of the Queens Public Library in Jamaica, Queens. The data consists of audio files of complete interviews (generally in .wav format) and timecode outlines - documents prepared post-interview that include metadata of the interview (such as interviewee's self-identified neighborhood, date and location of interview, interviewee date of birth), provides a summary of subjects discussed during the interview and contains timestamped overviews of different portions of the interview (generally in .pdf, .docx or .doc formats).



Queens Memory Project
Oral History Interview Timecode

Neighborhood (if any):	Jackson Heights, Corona, New Orleans, Santa Fe
Date of interview:	July 25, 1998
Time period:	1950's – 1998
Interviewee date of birth:	February 8, 1948
Name of interviewee:	Jeffrion Aubry
Name of interviewer:	Susie Tanenbaum and Jackson Heights Development Corp Interns
Length of interview:	00:58:28
Name of editor:	Sonali Sugurim
Editing software:	Audacity
Timecode Outline preparer:	Sonali Sugrim

00:01:45	Starts by stating his place of birth and how his family moved to New York City from New Orleans
00:02:58	Recalls his childhood in Corona, Queens
00:03:43	Notes that his neighborhood was diverse with families of German, Italian and West Indian backgrounds
00:06:21	Mentions that the only Black music radio station on during the 1950s was WWRL
00:07:25	Notes that it was not until he was a teenager that he noticed the prevalence of gangs in Corona and East Elmhurst area; mentions the Chaplains and the Enchanters gangs.
00:08:52	Speaks of his experience with racism
00: 16:38	Discusses the obstacles faced during and after college at Santa Fe in the midst of the Civil Rights Movement
00:24:07	Talks about the drug epidemic

Figure 6. Sample Timecode Outline, Queens Memory Project

The Queens Memory data is stored on a shared network drive, accessible only to certain Queens Public Library staff and volunteers. Each interviewee generally has a folder with their name, where the sound and text files associated with their interviews are stored,

though some interviews were grouped into subcategories based on specific projects associated with those interviews. There also exist separate folders for the various stages of the archiving and processing life cycle. There is no standardized naming convention utilized. Much of the preparation work for this project consisted of data wrangling - locating the full set of interview sound files and timecode outlines; and data cleaning - removing duplicate files, renaming files based on the convention LastnameFirstname_filetype, and converting .wav files to .mp3 for ease of storage and use. All timecode outlines were standardized to .txt format with UTF-8 encoding.

Queens Memory staff explained that they had not sought to standardize file naming or file storage locations due to the varied number and types of staff and volunteers working throughout the oral history life cycle - short-term volunteers who conduct interviews in the field, other volunteers or staff preparing timecode outline documents and several permanent staff preparing files for archiving and website upload and providing oversight. Because a large portion of their dataset is also provided by third parties, such as community-based oral history projects, there is a great variety of interview styles, lengths, and timecode outline qualities. For instance, some of the timecode outlines do not follow the format provided by Queens Memory, but are instead drafted as editorial pieces, with a heavier emphasis on the writer's voice and subjective lens rather than the interviewee's.

A brief analysis of the Queens Memory Project, through the lens of the Oral History Lifecycle:

> Idea: the idea and mission of the Queens Memory Project is clear: to diversify and democratize the historic record of Queens. "Life in the borough of Queens" is a large subject area and the idea can become a bit unfocused without a more narrow research question, particularly as Queens Memory incorporates various sub-projects, which each have an individual idea and aim (for example, recording weekly interviews along a particular street in Queens or collecting museum attendees thoughts on the New York Panorama exhibit at the Queens Museum)
>
> Plan: Queens Memory has a thorough methodology covering team roles, community participation, content development, repository, ethical and legal standards, and recording equipment. A concise project design statement would be a helpful tool for Queens Memory, particularly as there is frequent turnover from volunteer staff. A authoritative document outlining general information, project content and project management could be provided as part of volunteer training.

Interview: The Queens Public Library itself is a repository of research, the Archives an important historical record of the borough. Volunteers, interviewers, staff come to the table with different understandings of the historical context of Queens - providing reading materials would be a helpful first step in situating those who will be involved in the oral history interview process. Interviewees are selected by interviewers themselves and there are also Queens residents who request an interviewer and get placed on a waiting list. An impetus of this thesis project stemmed from the Library's desire to better understand the demographic makeup of their current interviewee pool and identify any gaps that may exist. In this way, Queens Memory is taking an important step in interviewee selection. A question guide has been curated by staff for interview purposes. Interviewers include both Queens Library volunteers who undergo training and community-based practitioners who develop self-led projects and later submit these to the Queens Memory archives. It is critical to the interview process that that interviewers ask open-ended and neutral questions - though this is undoubtedly covered in training, because the interviewer roster is entirely volunteer-based, consisting also of young students and one-off interviewers, they may lack the discretion and finesse of a more seasoned interviewer. This can be seen as both a strength and weakness of the Queens Memory interviews.

Preservation: Interviews that are conducted go through several stages before they are preserved tin the Archive. Timecodes outlines are drafted, sound clips are selected for posting on the web and consent forms are double-checked. Queens Memory preserves all interview audio files in the Archive.

Access/use: Selected clips and stories are posted on the Queens Memory website and the Queens Public Library Digital Archives site, but when I tried to access or search for oral history interviews, I was unable to find sound files to play. Though recently re-designed, the Queens Library website has long been a cumbersome eyesore to navigate and the Queens Memory and Digital Archives site are also not particularly user-friendly.

Of the Oral History Lifecycle stages, access/use is the stage Queens Memory could most improve in and also where this data visualization project is given shape and life. The oral history benchmark of preservation and access are part of what sets oral histories apart from other kinds of historical information. When an oral history is created, a new set of data and information is born. If Queens Memory's aim includes sharing a more

"complete" story of life in the borough, there needs to be more meaningful engagement and ease of access between the public and the oral histories.

Data Methods and Treatment

Initial ideas of how to visualize and create an engaging platform for the Queens Memory data included: tone/inflection analysis of audio data, investigating neighborhood "sound patterns", creating a relational map of interviewees and subjects discussed, a subway map-inspired visualization of each interview, with intersecting "stations" representing intersecting themes. Many of these were struck from the drawing board due to the limits in my time and abilities or because it would result in too busy a visualization, with users unable to glean anything meaningful. In an early sketch of the subway-inspired visualization (Figure 7), with only four interview "subway lines" mapped based off of neighborhood mentions in their interviews, the criss-crossing lines do not provide much by way of deep analysis, only an overall sense of which neighborhoods are mentioned more frequently.

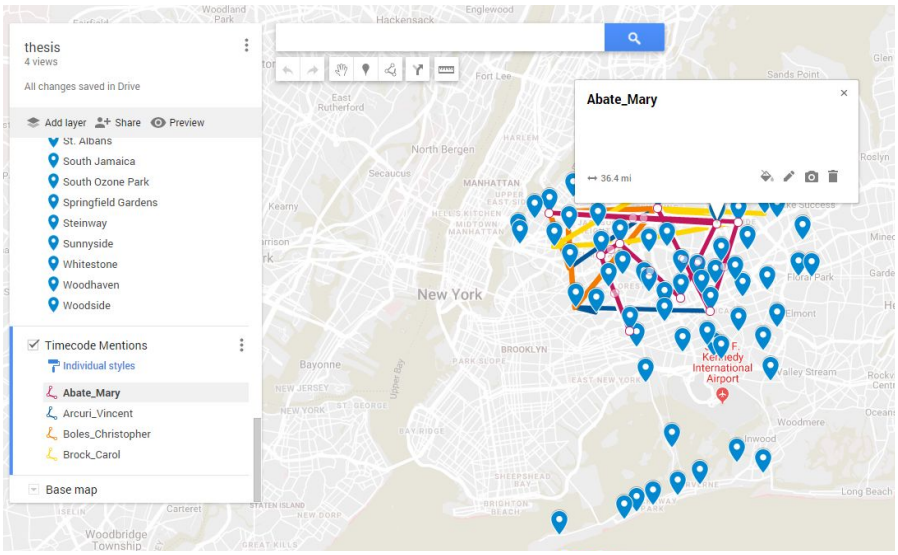


Figure 7. Early sketch of what a subway-inspired interview map might look like

I decided to visualize the most complete set of data available to me - the timecode outlines. I hoped that text mining this data would reveal patterns or provide insight into the oral histories in the Queens Memory archive and that a well-designed visual could become a critical proponent of the "access/use" stage of the Oral History Lifecycle, drawing in users and allowing them to engage with this historical record.

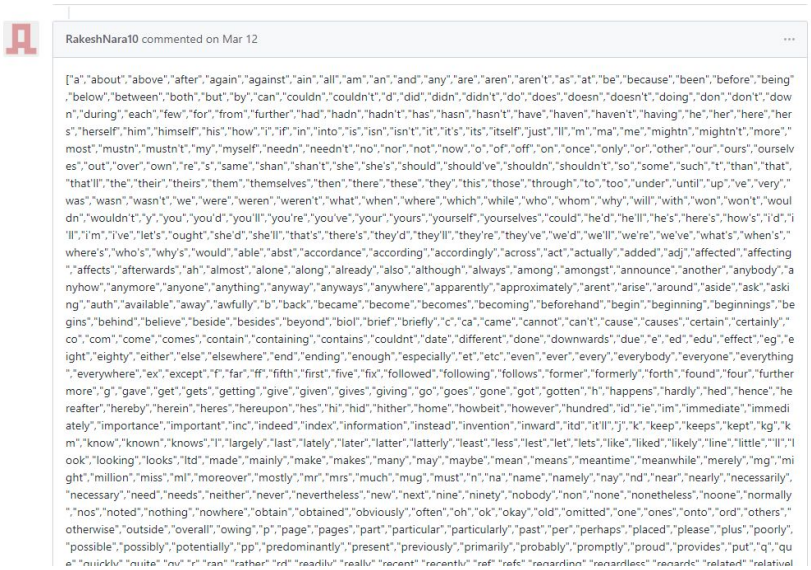


Figure 8. An array of NLTK stopwords, https://gist.github.com/sebleier/554280

Metadata sections were removed from the text files, keeping only the text related to the actual interview. Stopwords (identified in the python Natural Language Toolkit) were removed from text data, with the exception of words identified as important to analyzing this specific text data, such as “change” and “changes”, which were relevant as many people spoke about changes witnessed in their neighborhood; or “Long”, part of the Queens neighborhood “Long Island City” and “Long Island”. White space, inconsistent formatting and special characters were all issues that showed up in the first several passes through the data. (In the initial test run, a blank space was the most frequent “word” identified.) Cleaning the data and getting it to a place where it could be visualized in a meaningful way - the data janitorial work - took repeated trials and tests. These are documented on github: https://github.com/ssyung/thesis/tree/data-processing

Without machine learning training, I looked for a software solution to clustering and natural language analysis. RapidMiner Studio’s free software included clustering and sentiment analysis capabilities. Running the cleaned text data showed that the bulk of the interviews clustered together - this makes sense as stylistically and content-wise, most of the interviews are similar. After all, interviewers are working from a similar set of interview questions and all the interviews are dialogue-based. Some of the outliers were the very brief interviews with non-resident museum goers who were visiting the Panorama of New York City exhibit at the Queens Museum.

The New York Times’ interactive comparison of words used at the Republican and Democratic presidential nominating conventions was a source of inspiration for the

final visualization and I relied heavily on Mike Bostock's clustered force layout bubble chart code.

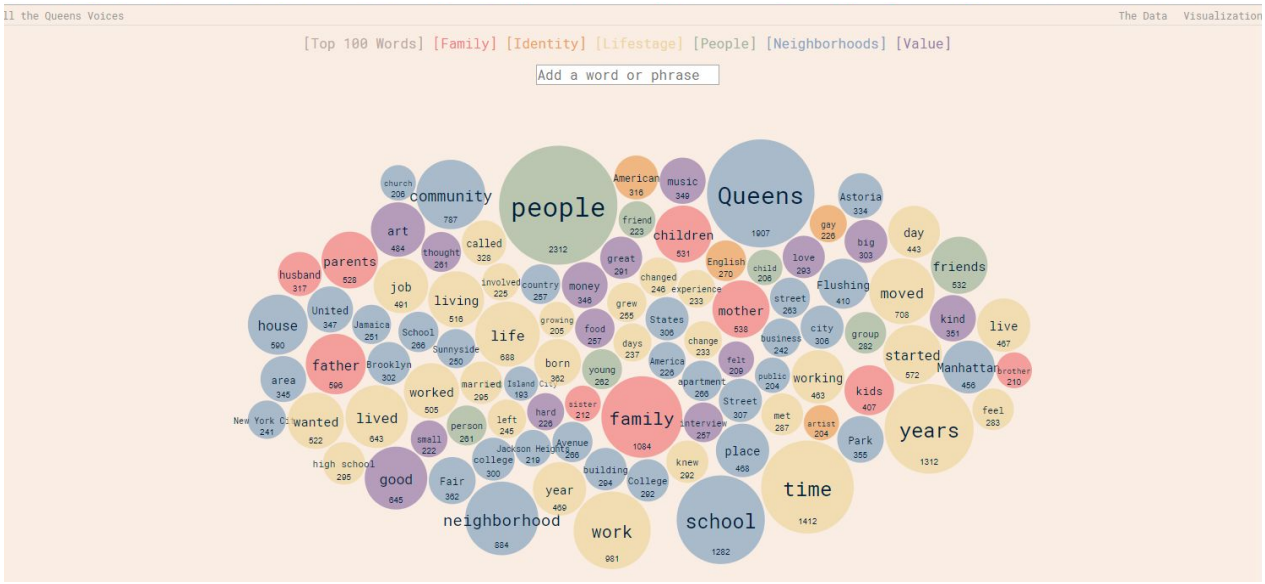


Figure 9. Visualization of Queens Memory oral history interviews, https://ssyung.github.io/thesis/

Analyzing the top 100 words, I categorized each word into five "buckets": Family (words identifying familial relationships like father, mother, parent, daughter, son, sister, brother); Identity (words that describe someone's cultural, linguistic, religious, gender or sexual identity, like Italian, immigrant, Catholic, Asian); Lifestage (generally verbs that described the action in a interviewee's life, words like worked, lived, moved, died, growing, experience); People (words identifying persons other than family such as neighbor, friend, or the generic "person"); and Value (a bit of a catch all for adjectives and other words that capture the stuff in an interviewee's life, words like art, great, money, food).

This text analysis is not a 1:1, perfect analysis of mentions, as it is based on timecode outlines of interviews and not transcripts. Timecode outlines are prepared by either the interviewer or a third party and are subjective in how the preparer chooses to summarize an interviewee's words. This was a helpful lesson in working with imperfect data, as is often the case in the "real world". Despite the degrees of removal the timecode text is from the actual spoken words of the interviewee, there is still value to be found, particularly when individual words are shown in the context of how they were utilized.

Per the request of Queens Memory, I also identified neighborhoods mentioned in interviews (see Appendix 1), based on a list of neighborhoods utilized by the Queens

Public Library Flushing - a combination of community board-designated neighorboods (http://home2.nyc.gov/html/cau/html/cb/queens.shtml), historically recognized neighborhoods, and neighborhood names per the customers they serve. Neighborhoods with over 300 mentions: Flushing, Astoria, Sunnyside, Jamaica and those neighborhoods that were not mentioned at all: Arverne, Holliswood, Malba, Neponsit, Saint Albans. The neighborhood count can be used by Queens Memory in their considerations of where to send their oral history interviewer volunteers next - neighborhoods that are underrepresented can be prioritized.

Another important aspect of this visualization project was that it not be limited to the visual realm. An oral history project would not be true to its core and intentions without including sound files and creating an aural experience. It was important to provide access to sound files so users could, after following threads in the text analysis, return to the actual data source to hear and experience actual interviews.

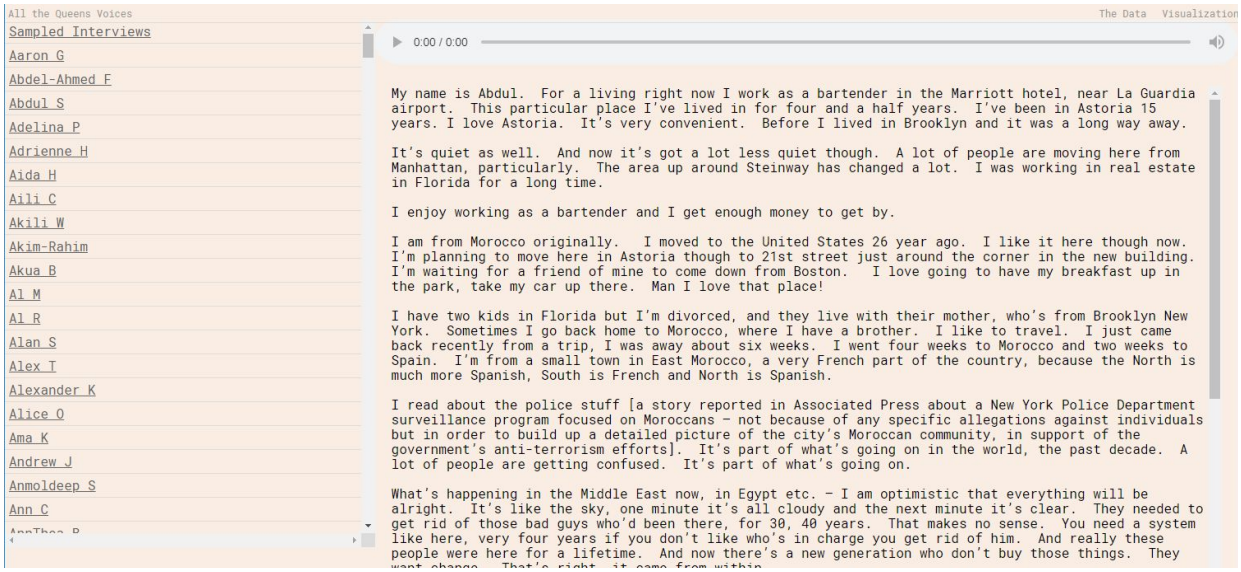


Figure 10. Interview sound files and cleaned text data made accessible to the user

Overall, the data visualization creates a space where users can access and use the data, even interacting with it by typing keywords of interest and seeing when and how they show up in interviews.

Conclusions, Considerations, Future Directions

This thesis project was a valuable experience working with real data - data that has not been standardized in any way, qualitative data that is far removed from a clean, tabulated spreadsheet. Though frustrating at times to collect, standardize and make sense of the data, it was rewarding to hear Queens Memory staff's hopes for what their data could tell them and to design solutions to some of the questions they had.

While analyzing word counts and clustering words into groups, a thread of discussion that caught my attention was the topic of change. Here is a selection from the visualization of mentions of "changed" and "changing":

Isabel C : at the CUNY Graduate Center, which slowly changed her feelings towards Sunnyside. Isabel now

Terry K : 00:30:36 Terry explains how much the world has changed since his childhood. He says that people are

Elinor S : a great neighborhood at the time although it's changed a lot by now. And my one of the kids I was

EunYoung K : Ms. Kim feels that physical Flushing has not changed in the six years she has resided in Queens. She

Julian W : time, the Jackson Heights neighborhoods were changing. Many Indians moved in and operated jewelry

Tom P : news is important. 01:36:06* Tom speaks to changing demographics in district, in Bayside. Says that

Alan S : the diversity. The complexion of the borough is changing, getting browner. There are people from

Panayiotis M : people here. Now there is not, everything is changing. A lot of Spanish people, Mexicans,

Maria S : that is going on? M - No. Queens has always been changing since I grew up here. It's amazing how different

I was interested in the tone and texture of the way change was discussed. Some interviewees felt that their neighborhood had not changed much and others felt it had changed drastically - was this because of the specific neighborhood they lived in, or their socio-economic status, race, gender, age? Change was sometimes discussed with a negative undertone - lamenting the way “things used to be” and a fear for what may be lost in the future, other times change was spoken of with awe - that even within neighborhoods where demographics drastically changed, neighbors got along, children played together.

In a similar way, I hope users who interact with this visualization will be drawn into this rich body of interviews -- by searching keywords of interest (say, the country their grandparents immigrated from, or the primary language they speak), they can identify interviews that may be of interest to them and hear a story that paints a fuller picture of Queens.

I plan to continue to work with Queens Memory to build out their database of interviewee metadata - gender, age, race, ethnicity, neighborhood, languages spoken, number of generations they have been in America. This information, paired with more robust natural language processing such as sentiment analysis, is ripe with possibility for deeper analyzing and understanding.

References

Bostock, Mike, Shan Carter and Matthew Ericson. “At the National Conventions, the Words They Used.” *New York Times,* updated September 6, 2012, https://archive.nytimes.com/www.nytimes.com/interactive/2012/09/06/us/politics/convention-word-counts.html. Accessed March 1, 2019.

Boyd, Douglas A. *Oral History and Digital Humanities: Voice, Access, and Engagement*. Palgrave Macmillan, 2014.

Bryan, Brooke. “Why Here/Why Now: Using Websites to Power Community Projects.” *Oral History in the Digital Age*, http://ohda.matrix.msu.edu/2012/06/why-here-why-now/. Accessed May 9, 2019.

Fathom Information Design. “Computer Memory: Visualizing a Century of Oral History.” *Medium*, May 21, 2018, https://medium.com/fathominfo/computer-memory-visualizing-a-century-of-oral-history-3d004e27dbdb. Accessed March 1, 2019.

Lakeram, Scott. “The Geography Of Queens.” *Queens Gazette*, 17 September 2014, m.qgazette.com/news/2014-09-17/Features/The_Geography_Of_Queens.html. Accessed (Cached) March 28, 2019.

Lieberman, Janet E., and Richard K. Lieberman. *City Limits: a Social History of Queens.* Kendall/Hunt, 1983.

Morrison, Pat. “Dave Isay on StoryCorps' DIY oral history project.” LA Times, 7 April 2015, https://www.latimes.com/opinion/op-ed/la-oe-morrison-isay-20150408-column.html. Accessed May 1, 2019.

Pozzi-Thanner, Elisabeth. “Reviews of Storycorps.” The Oral History Review, Volume 32, Issue 2, 2005, Pages 103–104, Published by the University of California Press for the Oral History Association.

“QuickFacts: Queens County (Queens Borough), New York.” *United States Census Bureau*,

https://www.census.gov/quickfacts/fact/table/queenscountyqueensboroughnewyork,newyorkcitynewyork,US/PST045218. Accessed March 28. 2019.

Roberts, Sam. “Listening to (and Saving) the World’s Languages.” *The New York Times*, 29 April 2010, https://www.nytimes.com/2010/04/29/nyregion/29lost.html. Accessed March 28, 2019.

Sanjek, Roger. *The Future of Us All: Race and Neighborhood Politics in New York City.* Cornell University Press, 1998.

Sloan, Stephen M. “Swimming in the Exaflood: Oral History as Information in the Digital Age.” Oral History and Digital Humanities, edited by Douglas A. Boyd and Mary A. Larson. Palgrave Macmillan, 2014, pp 175-186.

Sommer, Barbara W. *Practicing Oral History in Historical Organizations.* Routledge, 2016.

Tebeau, Mark. “Case Study: ‘Visualizing Oral History’.” *Oral History in the Digital Age*, http://ohda.matrix.msu.edu/2012/06/visualizing-oral-history/. Accessed May 9, 2019.

Thompson, Paul, and Joanna Bornat. *The Voice of the Past: Oral History*. Oxford University Press, 2017.

Tsui, Bonnie. “How Our Most Intimate Stories Become Animated Shorts.” Citylab, Dec 3, 2013, https://www.citylab.com/design/2013/12/how-our-most-intimate-stories-become-animated-shorts/7751/. Accessed May 1, 2019.

Weber, Andrew. "Queens." NewYork.com, April 30, 2013, https://web.archive.org/web/20150513065643/http://www.newyork.com/articles/neighborhoods/queens-72876/. Archived from the original on May 13, 2015. Accessed March 28, 2019.

Ziv, Stav. “How Technology is Keeping Holocaust Survivor Stories Alive Forever.” Newsweek Magazine, October 18, 2017, https://www.newsweek.com/2017/10/27/how-technology-keeping-holocaust-survivor-stories-alive-forever-687946.html. Accessed May 2, 2019.

Appendix

Appendix 1. Neighborhood count

Neighborhood	Number of Mentions		Neighborhood (cont)	Number of Mentions (cont)
Arverne	0		Jamaica Estates	9
Astoria	375		Jamaica Hill	1
Auburndale	3		Kew Gardens	48
Baisley Park	8		Kew Gardens Hills	2
Bayside	71		Laurelton	46
Bayswater	1		Little Neck	9
Belle Harbor	2		Long Island City	216
Bellerose	28		Malba	0
Briarwood	41		Maspeth	12
Broad Channel	2		Middle Village	6
Cambria Heights	19		Neponsit	0
College Point	24		Oakland Gardens	3
Corona	117		Ozone Park	105
Douglaston	19		Queens Village	34
East Elmhurst	27		Queensboro Hill	3
Elmhurst	79		Ravenswood	4
Far Rockaway	20		Rego Park	25
Floral Park	9		Richmond Hill	79
Flushing	550		Ridgewood	107
Flushing Meadows-Corona Park	43		Rockaway Beach	24
Forest Hills	152		Rockaway Park	4
Forest Hills Gardens	10		Rockaway Point	3
Fresh Meadows	17		Rosedale	15
Glen Oaks	8		Saint Albans	0
Glendale	21		South Jamaica	21
Hillcrest	6		South Ozone Park	25
Hollis	44		Springfield Gardens	15
Holliswood	0		Steinway	32
Howard Beach	18		Sunnyside	338
Hunters Point	5		Whitestone	19
Jackson Heights	269		Woodhaven	39
Jamaica	315		Woodside	165