Princeton at the turn of the 20th Century: Building geospatial data to tie the 12th population Census (1900) of the United States to individual addresses

Keywords: Historical Census data, Historical geocoding, Historical address data, Population Census

Summary: This project explores methodology and workflows of spatially linking the United States historical population census data to individual houses located in an urban area as well as in a rural area. For this project I used different maps of Princeton published around 1900 and georeferenced and then extracted locations that were listed in the census enumeration forms from the maps, and later linked the historical census data to the locations. In the project I also describe the challenges and opportunities of using historical census data downloaded from the Ancestry library licensed by ProQuest and the method of correcting transcription errors found in the Ancestry library database. The historical location data built for the census was later used in geocoding the Princeton Business Directory to see the possibility of filling missing data in the census as well as in the Business Directory.

Introduction

The “Princeton at the Turn of the 20th Century” project is the outcome of a few years of working with researchers who were trying to build a historical GIS database and link it to the United States historical population census data of small neighborhoods or towns. In addition, an undergraduate history class that I was involved with was learning how to build a spatial history of persons living in Princeton’s historically black neighborhood through a variety of sources such as historical United States population censuses, local oral history recordings, and others. These works led me to think of exploring the methodology and workflow of developing a geospatial database to tie the United States historical population census data to an individual house. I wanted to understand what sources might be needed and what methods might be involved to build a proper database. The census gives detailed information about a person, their household, and where they lived. While exploring the works done by others, I found the Urban Transition Historical GIS Project. This project has built historical geocoding services to tie 1880 population census data of 39 major cities in the United States to individual addresses based on address ranges. Unfortunately, they did not release the detailed individual household data with names enumerated by the census takers to the public. Their methodology could be used in tying historical census data to individual addresses if a place is located in an urban area because most of the urban areas since 1880 in the United States do have house numbers and street names. However, if a place is located in a non-urban area, there is no example to tie census data to individual household locations.

To find out what processes are involved in building a geospatial database to link the historical population census data to individual houses in urban areas, and also in rural areas, I needed to pick a place that has both urban and rural areas and is manageable to test the methodology and workflow of linking the population census data to individual house locations. I picked Princeton, New Jersey as a place to do my research. Princeton is an important place in the history of New Jersey and the United States. A decisive battle of American independence called the Battle of Princeton took place
there in 1777; and for four months Princeton became the country’s capital. Two American presidents, Grover Cleveland and Woodrow Wilson, lived in this town. Three world renowned institutions are located there: Princeton University, the Institute for Advanced Studies, and Princeton Theological Seminary. Many intellectuals, thinkers, and scientists, including Albert Einstein, made their homes in Princeton. It is also centrally located between two major American cities: Philadelphia and New York City. In addition, I happen to live in Princeton so it would be relatively easy for me to collect materials and verify the locations.

At the turn of the 20th century, Princeton was divided into two municipalities, Princeton Borough and Princeton Township (Figure 1). In fact, this division existed from the early 1800s until 2012, when it became one administrative unit again. Princeton Borough was where most of the businesses were located and it was also the oldest part of the town, where in 1900 all the houses had numbers as well as street names. This part of Princeton represented urban America. Princeton Township was more rural; it surrounded the Borough and many of the houses did not have house numbers in 1900 when the census was taken. This allowed me to research the feasibility of developing a methodology and a workflow to spatially link census data not only to houses that had street names and house numbers but also to houses located in a rural area that did not have house numbers. Understanding a methodology and a workflow to build historical geographic data and spatially link it to the historical United States population census will help others who want to replicate similar work in the future. Unlike the Urban Transition Historical GIS Project that matched an individual household address of the census to a location based on the address range, this Princeton project looked at the feasibility of developing geospatial data that can geocode the census data to individual house locations rather than an address range. The address range gives an approximate location of the address, whereas the spatial data I have developed ties census records to actual locations of houses.

Population Census Data

The United States decennial population censuses of individuals are released to the public after 72 years. This release gives researchers rich historical records about people living in a particular place
at a certain time; if the data is spatially tied to individual houses, it will become much richer, with the ability to explore spatial distribution and patterns of people living in the area. The scanned copies of forms filled by the census enumerators are made accessible to the public by the United States National Archives and Record Administration through their website. In addition, the Church of Latter Day Saints has transcribed the forms and built spreadsheets of each form that can be downloaded by users through the Ancestry Library edition database licensed by ProQuest, the global information-content and Technology Company to the libraries in the United States. The interface to view the data is very intuitive and easy to use; however, there is no easy way to download the whole spreadsheet of place data at once. To download a transcribed part of the census records, a person has to view each form one at a time and then click on the “show index” icon which will show the transcription of records in the spreadsheet style below the scanned form, then copy each spreadsheet record and paste it on the Excel spreadsheet. There is a total of 100 scanned forms for Princeton, and each form has 50 entries, except a few forms at the end.

Princeton was divided into three Enumeration Districts (ED) in the 1900 census: two EDs covering Princeton Borough and one for Princeton Township. There is no ED boundary map to guide how the census data was geographically distributed. The EDs were numbered 0053, 0054, and 0055. The Census Enumerators were selected from the local area, one from each ED. This can be verified by their names on the census form and the places where they lived. The Census enumerators started taking the census on June 1st but they ended on different dates. ED 0053 ended on June 26, ED 0054 on June 18, and ED 0055 on June 14. ED 0053 had the largest population followed by ED 0054, and ED 0055 had the lowest.

The copying of their transcribed census data was time consuming and organizing them on one single spreadsheet required some adjustments on how the rows and the columns should align. The 1900 population census asked the following questions: name, relation to head of family, race, sex, birth month, birth year, age, marital status, years of marriage, number of children, place of birth, father’s birthplace, mother’s birthplace, immigration year, years in the United States, naturalization, occupation, months not employed, attended school, can read, can write, can speak English, house owned or rented, house owned free or mortgaged, and farm or house. The instructions given to Enumerators for taking the census are explained in detail in *Measuring America: The Decennial Censuses from 1790 to 2000*, issued on April 2002 by the United States Census (https://www.census.gov/prod/2002pubs/pol02-ma.pdf).
Developing a Historical Geocoding Database

Building a historical geocoding database needs historical spatial data that have either line features with associated street names and address ranges or point or polygon features with attribute tables associated with a single address. In this project, I planned to build a single address geocoding database. To build this kind of geocoding database, I needed a map that showed the location of each address from that period. Finding maps of Princeton Borough that showed either property boundaries or building footprints with house numbers was a bit easier than finding similar maps of Princeton Township.

I managed to find three detailed maps of Princeton Borough published around 1900. The first map was published in 1898 and shows building footprints with house numbers. The second map is part of an atlas; it has four sheets covering Princeton Borough showing parcel boundaries, street names, house numbers, and owners’ names. The atlas was published in 1905. The third map I used was the Sanborn map of Princeton published in 1902. This map shows detailed information about houses including house numbers and it comes in 12 sheets. It is the best map to use for this project, but it does not cover all of Princeton Borough. All three maps do have street names and house numbers, but some of the house numbers do not match with the 1900 census house numbers. Sometime the census did not provide a clear address of a person’s residence. However, to build a location database, I verified the addresses from other sources and created a location for each household. For example, in the census, Princeton University President Francis Patton was listed as living at 65 B Place (Figure 3) in Princeton although he was living in the Princeton University’s President residence, called Prospect House. None of the Catalogues of Princeton University list Prospect House’s address at 65 B Place, and I didn’t see any maps which showed that street name. In a situation like this, I created an address based on the house location.
I have also found errors in the Census where a person was supposed to be living at one address but in another official record the same person was said to be living at a different address. An example here is the Princeton University Professor of Physics, Cyrus Fogg Brackett, listed in the census as living at 4 Washington Road, but in the 1900 Catalogue of Princeton University, he was listed as living at 4 Prospect Ave., which is on the corner of Prospect Ave. and Washington Road. The closest address on the Washington Road side from Prospect Ave. is 42. There is no 4 Washington Road. Also, the same Catalogue a few years later listed Professor Brackett as still living at 4 Prospect Ave. This show the census record is probably incorrect. See the census form and transcription spreadsheet as well as the Catalogue of Princeton University record (Figure 4 and 5).
In the Census record another household was living at 4 Prospect Ave. (Figure 6) and they do not seem to have any blood relationship. See the copy of the census form and spreadsheet data below.

Finding maps of Princeton Township published around 1900 that show either property boundaries or building footprints with house numbers was much more challenging. In fact, most of the households located in Princeton Township that were enumerated in the 1900 census did not have house numbers. However, I found a few land ownership maps published around 1900 for Mercer County, New Jersey, where Princeton Township and Borough were located, and a map of Princeton that included both Borough and Township. These maps show the general location of a house and the name of the person who owned the property. These maps were very helpful in linking a person to a particular location if the person owned property. If the person was not a land owner but renting a house or a farm then it was difficult to tie a household to a location. However, in the census form
there is a column called “Number of dwelling in order of visitation”. This house visitation number was very helpful in approximating the location of a household to a property. In order to approximate the location of households, I used a 1920 tax map, the oldest tax map in the Princeton Municipal office that I acquired for this project, and the first aerial photography that covers all of New Jersey State taken in the 1930s. This georeferenced aerial photography helped to identify possible locations of households based on street names and dwelling visitation numbers and the 1920 tax map helped me verify some household names who were renting in 1900 but on the 1920 tax map they were listed as property owners along the same roads where they were listed as renters in the census. On a few occasions I had to use 1920 census data to verify a person’s location which was not clearly listed in the census.

Once all the maps were acquired, I scanned and georeferenced them. Initially I thought of building an address range database, which would be easier to build than a single house address location database but I found that many streets’ house numbers were changed and many streets were built or removed since 1900, and therefore verifying each address on the street and rebuilding an address range on a street database might take more or less the same amount of time as building a single house address database. For example, Woodrow Wilson who was a Professor at Princeton University before he became the President of the United States, was listed at 50 Library Place (Figure 2) in 1900 but the same house located in the same place is now changed to 82 Library Place. Also, well-known African-American personality Paul Leroy Robeson, who was born in Princeton in 1898, was listed in the 1900 census as living at 72 Witherspoon Street, but the same house address is now 108 Witherspoon Street. Because of the above reasons, I decided to build a geographic database of individual households by extracting a point for each location. Once that decision was made, then I had to go through each street name and house number that was listed on the census and see whether those street names and house numbers existed on the map. If they existed then I extracted the point and wrote down the street name and house number. I didn’t want to extract all the house numbers displayed on all the referenced maps because some house numbers are changed from one map to another. I think that has been done because new houses were added and house numbers needed to be added or changed. I wanted to make sure I was extracting house numbers that were in the census rather than extracting every house number shown on maps which might spatially link a household to the wrong location.

While going through the process of creating a geographic database, I found many transcription errors of street names, house numbers, people’s names, occupation labels, etc. For example, Trenton Turnpike was transcribed as Turton Tumeric, Alexander was written AlenAuller, Lawrence Road was written Launi Road, 20 Baker Street was labelled 30 Boker Street, Silvester was Siliester, etc. Almost all these errors could be verified and corrected by using the maps that show property owners’ names, house numbers and street names, and sometime going through the house visitation numbers. Most of the house visitation numbers are in sequential order, which helps to verify how the Enumerator might have walked from one house to the next. Checking each census household record was very time consuming but also very helpful in correcting many of the transcription errors. In order to verify the transcription errors of each enumeration form, all the copied spreadsheet records from enumeration forms were numbered by scanned image number and ED numbers. Since the scanned enumeration forms were organized by ED number, having image and ED numbers next to the record helped me in verifying any transcription problem in the census data by looking at the enumeration scanned form and verifying the transcription. If image and ED numbers were not marked on the copied spreadsheet, then it would have taken me a lot more time to find a scanned form to verify and correct the transcription error. Cleaning the data took lots of time. The data is a
bit messy, however, it is better than someone transcribing them from the census form. It is apparent from creating the single address database that I have to create two separate databases: one that has address and street names, and another with street and household names only. It is also important to distinguish between the locations that were verified on a map versus the locations that were approximate based on visitation numbers and other sources. All the census records do have a note differentiating those two locations (Figure 7). This will help researchers to understand the spatial accuracy of the historical GIS database when I release the data to the public.

![Figure 7](image_url)

For geographic data that had address and street names, I created a geocoding database. To create a geocoding database, I used Esri’s ArcMap tool called Create Address Locator. While using the Create Address Locator tool, I selected the address locator style as Single House and reference data as geographic data I extracted from maps, and then ran the tool. After creating the Address Locator, I then changed the geocoding options to “Match without house number” and “Match with no zones” to Yes. After changing those options, I rebuilt the Address Locator to reflect the changes I made on the geocoding options. Once the Address Locator was rebuilt, then the census data were ready to geocode. Before geocoding the census data, I needed to create a new column called address. In this column, I concatenated house number and street name to create the address. This address column was used in geocoding the census data to the Address Locator.

For geographic data that had only street and household names, I did not create geocoding address locators because none of them had house numbers, so I joined them to the census data. In order to join geographic data to the census data, I created unique IDs for each household in the census data and used the same unique ID for each location data where a household lived. Later the census data were joined to location data using the unique ID as a common field. After matching each household to a location, I managed to create an Enumeration District boundary map that I did not have before. Now I have the ED boundaries of the 1900 census (Figure 8).
The population size of Princeton at the turn of the 20th century was 4,854. After a century, in the 2000 census, the population had grown almost 8 fold (30,230). In 1900, 80.33% of the population lived in Princeton Borough (3,900) and 19.67% of the people were living in Princeton Township (955). However, a century later, 55.15% of Princetonians were living in Princeton Township and only 44.85% were living in Princeton Borough. Princeton also saw a shift in the percentage of Black people living in the town. In 1900, 20.60% (1,000) of the population were defined as Black, whereas a century later, it was reduced to 5.82% (1,760). However, the percentage of White population living in Princeton remained more or less the same. In 1900, it was 79.27% (3,848) whereas in 2000 it was 80% (24,206). Princeton attracted a large immigrant population in 1900 mostly from European countries (612) (Figure 11); however, there were people who immigrated from Canada, the Caribbean Islands, South America, and Asia (35) (Figure 12). The share of the United States born population was 86.48% (4,156) and the immigrant population was 13.56% (647). The immigrant population was distributed all over Princeton Borough and Township and there was no clustering of immigrants from one particular country in any part of Princeton (Figure 7). However, most of the Black population was living in the northern part of Princeton Borough. There were a few Black households who lived outside the Black neighborhood (Figure 9). The largest number of immigrants came from Ireland, followed by Germany, then England, and then Scotland. The population born in the United States came from 38 different states including the District of Columbia and New Jersey. Most of the people were born in New Jersey state (3044), followed by Pennsylvania (257), New York (245), Virginia (215), and North Carolina (109). The rest of the people came from other states (Figure 10).

There were 1100 households living in Princeton in 1900 and the size of the households varied from 1 to 14 people. For example, Mr. Philip Golden who was born in New Jersey, and whose parents came from Ireland, was living at 94 Bayard Lane, and 14 people were living at this residence. They include his wife, 5 daughters, 3 sons, and 4 boarders. The household count statistic was created by summarizing family numbers which were listed on the enumeration form.
Figure 9.

Figure 10.
The historical geographic data I created helped me to build a geocoding database that can be used in geocoding any historical address of Princeton from 1900. In order to test this methodology, we scanned the Princeton Business Directory published in 1900 from microfilm and tried to use OCR to convert the scanned directory to text, but it did not work well because of the quality of the microfilm as well as advertisements printed in the directory. So, we transcribed the directory that did not OCR properly, and created a spreadsheet file. This spreadsheet was geocoded using the historical geographic data created for the 1900 census; all the listings in the directory that had addresses were matched and geocoded with the business listings’ locations (Figure 13).
This geocoded data could be used with the census data to fill missing gaps either in the census or in the business directory. For example, in the 1900 census data, Ms. Emily Noyes, who lived at 108 Mercer Street, did not have an occupation written on the census form. However, in the 1900 Business Directory, her occupation was listed as Music Teacher (Figure 14).

In another situation, Mr. Leavitt Howe’s address was not listed in the Business Directory, but you can find his address in the census data (Figure 15).
The methodology and workflow template I developed could be used for building any historical geographic data to link historical population census data. The method could be replicated to any urban or rural areas in the United States. There is no easy solution to spatially link the historical census data to a location without first building a house location database which needs to be based on maps published at or around the same time as the census was taken, then link the census data to a location. Once the geographic data for one decade is built, then that database can be used in building another decade’s location database. The Princeton 1900 census data and address location database will be released to the public. This will allow researchers to analyze and visualize socio-demographic characteristics of Princeton at the turn of the 20th century.

References


Irving C. Hicks. Map of Mercer County, New Jersey. Philadelphia, Penna.: Irving C. Hicks, publisher, 1903


