

---

## Data Project Guidelines

---

### The Big Picture

This data project is one of the central components of the course. In class, we are covering a wide range of empirical approaches to measuring mobility and inequality for the United States as a whole over the past two centuries. We are going to examine how structural changes to the US economy and major political movements shaped mobility and inequality patterns nationally. The goal of this data project is to do the same thing, but with a narrow focus on Williamsburg. As a class, we will construct inequality and intergenerational mobility data for Williamsburg from the times of the Great Depression to the present. These data will then form the basis your final research papers.

The data project consists of two distinct components. The first is the creation of an intergenerational dataset of Williamsburg families, constructed by linking those families across federal censuses. These intergenerational data will provide insights into the geographic and occupational mobility of Williamsburg residents over time. The second is a dataset of property histories. These data will capture ownership histories and changes in covenants across the different neighborhoods of Williamsburg. Details on the construction of each of the datasets are provided below.

### An Intergenerational Dataset of Williamsburg Families

Each student will be responsible for linking a sample of approximately 50 Williamsburg residents forward from the 1920 federal census to the 1930 and 1940 federal censuses.<sup>1</sup> This methodology is similar to that used in several of the papers discussed in class, including Long & Ferrie (2007), Long & Ferrie (2013), Feigenbaum (2014) and Collins & Wanamaker (2015). This time period captures several major shocks to intergenerational mobility covered in class: the Great Depression, the High School Movement, the end of the Age of Mass Migration, the Great Migration and, locally, the construction of Colonial Williamsburg.

### General Instructions

You will be assigned a spreadsheet containing approximately 50 individuals from the 1920 federal census living in the Williamsburg area (including what is today the City of Williamsburg, James City County and York County). The spreadsheet will contain a unique id

---

<sup>1</sup>The 1940 census is the last federal census that is publicly available and therefore available for linking. Census data becomes fully public after 72 years.

number for each individual as well as basic demographic information for the individual from the 1920 census. Each individual will also have a link to the Ancestry.com page for his 1920 census record and links pointing to Ancestry.com search results for the individual in 1930 and in 1940.

1. Open the link to the Ancestry.com 1930 census search results for your first individual. Note that you may first have to log in through Swem's link to Ancestry available [here](#). These search results are based on the individual's first and last name, age, and birth state. Note that if the link to search results does not work, you can access the main 1930 federal census search page [here](#) (you may need to open this link twice, the first time you open any Ancestry link you will likely get redirected to the main Ancestry Library page).
2. Look through the results returned by Ancestry and see if there is a good match. You may also choose to refine the search results by using additional information from the 1920 census if you are having trouble finding a decent match.
3. Once you determine the best match, record the match quality in the relevant columns on your spread sheet. The first aspect of match quality is whether the match represents a good, unique match. Enter the most appropriate of the following designations in the match type column:
  - **Good and unique** - In this case, the match looks like a good match for your individual and there should be no other matches that look like a good match.
  - **Good and nonunique** - In this case, the match looks like a good match for your individual but there is at least one other individual that looks like a good match.
  - **Bad** - In this case, even the best match does not look like a good match.

The second aspect of match quality is your assessment of how good the match is on a zero to ten scale, with zero being a terrible match and ten being a perfect match. This is a subjective score and your scores may differ from those of other students. The key is that you be consistent throughout your dataset with how you assign your match quality scores. Enter your score in the match quality column.

4. For your best match, copy the link to the relevant 1930 census record (the 'View Record') link in the relevant column in your spreadsheet.
5. Now search open the link to the Ancestry.com 1940 census search results for your first individual (if you want to enter the search criteria yourself, you can access the 1940 census search page [here](#)).
6. Follow the same procedures you used to identify and record the 1930 census match to identify and record the best 1940 census match. You may use information from your 1930 match to help identify the best 1940 match if you had a good, unique 1930

match. However, if any information conflicts with the 1920 information, use the 1920 information rather than the 1930 information. For example, if you have an individual who was single in 1920 but married in 1930, the wife's name in 1930 can be helpful for identifying the best match in 1940. However, if you find a good match in 1930 but the age is off by two years, use the birth year implied by the 1920 age when searching for the best 1940 match.

7. Now repeat steps 1 through 6 for all of your remaining individuals from 1920.
8. Once you have found all of your best matches, save your spreadsheet as either an Excel spreadsheet (with a .xlsx extension) or a comma-separated-variable file (with a .csv extension) and email your file to me. Please note that it is important to keep the variable names identical to they way they were in the original template file. This ensures that all of the files can be successfully merged together.

Once everyone has completed their samples, I will combine the samples into a single dataset and transcribe additional details from all of the linked census records. I will then post the complete dataset online so that everyone can use it for their final projects.

## A History of Williamsburg Neighborhoods

The second part of the data project requires each student to explore the history of two of the older houses in Williamsburg. For each house, you will trace its price and ownership history, identify any changes in covenants and zoning regulations restricting the use of the property, and identify the house residents in 1940. These data, particularly the demographics of the homeowners and the restrictive covenants, overlap with the types of evidence explored in Aaronson et al. (2017) and Rothstein (2017).

### General Instructions

There will be a Doodle poll posted online listing several Williamsburg houses all built before 1940. You will sign up for two different houses. There is also a spreadsheet on our Blackboard site providing a template for recording the house details. For each of your houses, you will need to do the following:

1. First, to familiarize yourself with the house, look up the property on a real estate search website like [Zillow](#) or [Trulia](#). Take a look at the house values in the neighborhood, the assigned schools, the sizes of nearby houses and so on to get a sense of the type of neighborhood the house is situated in. One goal of this part of the data project is to assess the extent to which neighborhoods stay the same or change over time. This gives you a good benchmark for what the neighborhood looks like today. Record the current square footage and estimated value for your house in your spreadsheet and the url for the webpage on which you found those details.

2. Now establish what the neighborhood looked like in 1940 using the 1940 federal population census. Start by going to the [Ancestry search page](#) for the 1940 census. You can go about searching for individuals in the neighborhood in a couple of ways. You could try bringing up the street by putting the street name in the ‘Keyword’ field on the search screen. Alternatively, you can browse the manuscript pages corresponding to the neighborhood by using the ‘Browse this collection’ menu on the right side of the search screen. To find the correct enumeration district, refer to the enumeration district maps on Blackboard or navigate to the appropriate maps through Stephen Morse’s enumeration district [map finder tool](#). Note that if you use the map finder tool (or search for the maps on Ancestry), you need to look at York County. For 1940, James City County and Williamsburg are both included in the maps associated with York County.
3. Once you have found a manuscript page covering the street your house is on, record the url for the manuscript page image in your spreadsheet. You only need one manuscript page containing houses on your street. If there are additional pages, you do not need to include links to them. Once everyone submits their spreadsheets, I will compile data on the street residents in 1940. For now, take a look over the households in the 1940 federal census. Think about how the residents’ characteristics in 1940 suggest the neighborhood has changed in terms of demographics over the past half century.
4. Now you will trace the property history of your house. If your property is located in the City of Williamsburg, use the city’s online property lookup tool, available [here](#), to find details on the ownership history and assessed value of the property. If your property is located in James City County, the same information can be obtained [here](#). Record any and all relevant information about the property and note the parcel ID number. This number will be useful for searching earlier records at the clerk’s office.
5. Once you have exhausted the information available through the online property information systems, you need to go to the district court clerk’s office to trace the property’s history further back. The clerk’s office is located in the courthouse at 5201 Monticello Avenue, Williamsburg VA 23188.<sup>2</sup> At the clerk’s office, you can use the computer terminals to search for earlier deeds related to your property. Trace the property’s deeds back as early as you can, reading through each deed to uncover who was buying and selling the property and whether any restrictive covenants were attached to the property.<sup>3</sup>
6. Record the details of the property history in the spreadsheet. Be certain to include any interesting observations in the notes column (for example, observations about how the

---

<sup>2</sup>Note that the clerk’s office is open weekdays during business hours (exact hours can be found on their website). If you are unable to go to the courthouse due to schedule constraints or some other reason, please contact me and we will work out an alternative assignment for you.

<sup>3</sup>The clerk’s office charges 50 cents per page for any printouts or copies of deeds. I will gladly reimburse you for any pages you print at the clerk’s office.

character of the neighborhood has changed or about particularly interesting zoning or deed restrictions). If there are particularly interesting deed restrictions, make copies of the relevant pages to include with your project.

7. Repeat steps 1 through 6 for your other property, creating a separate spreadsheet for it. Once you have completed your research on both properties, submit your two Excel files to me by email along with pdf versions of any interesting excerpts from the deeds.

Once everyone has submitted their property summaries to me, I will combine them into a single dataset for use by the entire class when working on the final project.

## Due Date

Both components of the data project, the linked individuals and the property histories, are due **April 10th** at **5pm**. Please submit both parts to me as attachments by email. The project will be graded out of 20 points. Projects turned in late will incur a one-point penalty that will increase by one point every 24 hours. I strongly advise you to get started early so that I can help you with any problems that you run into.

## References

- Aaronson, D., Hartley, D., & Mazumder, B. (2017). The effects of the 1930s HOLC redlining maps. Tech. rep., Federal Reserve Bank of Chicago.
- Collins, W. J., & Wanamaker, M. H. (2015). The Great Migration in black and white: New evidence on the geographic mobility of American Southerners. *Journal of Economic History*, 75, 947–992.
- Feigenbaum, J. J. (2014). A new old measure of intergenerational mobility: Iowa 1915 to 1940. *Unpublished Manuscript*.
- Long, J., & Ferrie, J. (2007). The path to convergence: Intergenerational occupational mobility in Britain and the US in three eras. *The Economic Journal*, 117(519).
- Long, J., & Ferrie, J. (2013). Intergenerational occupational mobility in Britain and the US since 1850. *American Economic Review*, 103(4), 1109–1137.
- Rothstein, R. (2017). The color of law: A forgotten history of how our government segregated America. *New York, NY: Liveright*.