Data and Tools

Data

To conduct our analysis, we need data and we have the benefit of possessing an abundance. We have three categories of data: a list of delegates who participated in ratifying the Constitution, records on debt ownership before and after Hamilton's 1790 financial plan and nineteenth century census records.

Delegates

We have data on the identities of individuals who participated in the Constitutional Convention, which drafted the Constitution, and the state conventions that ratified the Constitution. These are the individuals who Beard and McDonald were interested in.

Debt Certificates

As mentioned earlier, we can divide our data on debt into debt records for individuals who received debt in exchange for goods and services provided in the 1770s and 1780s (henceforth known as pre-1790 securities), and individuals who exchanged debt for consols under Hamilton's 1790 plan to refund the national debt (henceforth known as post-1790 securities). Another way to think about the difference between the pre-1790 and post-1790 securities is that the pre-1790 securities tell us who **lent** to the government, and the post-1790 securities tell us who the government **paid back**.

The location of the data and other relevant information are described in the GitHub README.md (https://github.com/alicezg2/SPEOC-pt-1/blob/revamp/doc.md).

Pre-1790 Securities

The organization of the pre-1790 securities is somewhat complicated because the Continental Congress and state governments provided compensation for goods and services in various ways. Referencing Tom and George's 2021 Summer Proposal, the main components of this debt were
1. Loan office certificates issued by state loan officers between 1777 and 1781
2. Liquidated debt certificates issued by five military commissioners
3. Liquidated debt certificates issued by 13 state loan officers after 1783
4. Debt certificates to Continental soldiers for back-pay issued by John Pierce
5. Other miscellaneous credits.

We have records of the amounts issued to individuals of
1. Loan office certificates for 9 of 13 states
2. Liquidated debt certificates issued by 1 of the 5 military commissioners
3. Liquidated debt certificates issued by 8 of the 13 state loan officers

4. Nearly all of the debt certificates issued by John Pierce to Continental soldiers for back-pay
5. None of the other miscellaneous credits

Overall we have records on over 208,000 individual debt certificates issued by the Continental government with specie value totaling over 24 million Spanish Dollars. These represent about 85% of the domestic interest-bearing debt (prior to accounting for interest in arrears). Note that we do not have data on debt issued by states during the Revolution; we only have data on state debt that was assumed.

The [2021 Summer Proposal](#) describes the structure of the data in more detail, but here are some notes you may find helpful. First, liquidated debt and loan office certificates differ by date of issue, not issuer. Second, the military liquidated debt certificates we do have are for the Marine department, which includes the Continental Navy. Third, the Debt certificates issued by John Pierce, known as Pierce Certificates, are basically IOUs that promised to pay the wages of Continental soldiers at a later date. Some soldiers sold these IOUs at steep discounts and Hamilton's 1790 plan compensated the purchasers of these IOUs, not the soldiers, who were the original holders. Fourth, note that although the total amount of species-denominated debt was around 24 million Spanish dollars, Hamilton's plan refinanced debt based on the face value of debt holdings.

Post-1790 Securities

Tom and George's [2021 Summer Proposal Part II](#) also outlines the structure of the post-1790 securities. Owners of domestic Continental debt, which includes all of the debt types mentioned earlier, were compensated under the plan outlined in section 3.1 of Hall and Sargent 2014. Continental Debt consists of debt issued by the **Continental Congress.** These records have the suffix CD (which stands for continental debt) in their file names while the assumed state debt files have the suffix ASD (which stands for assumed state debt). Assumed State Debt was issued by state governments during the Revolution (think State of Massachusetts, or New York). We call that debt "Assumed State Debt" because the federal government, under Hamilton's plan, assumed responsibility for paying for the debt that the states had accumulated.

See the **Main Table** section for more information on how this data was processed.

Census Records

Last summer (2022), George and I had the idea that it would be fascinating to link the individuals in our data to 1790 census records. Ancestry.com's Library edition (you should check if your high school or college has subscribed to this database) contains (as far as I know) the only individual level database of 1790 census records (census websites like IPUMS contains aggregated county level data, which is still helpful). I wrote a web scraper to match individuals

from our Post-1790 securities to this database, which helped us gather useful information for debt holders such as the demographic breakdown of their household, how many slaves they owned, their occupation and their place of residence.

Although the decennial federal census only began in 1790, Ancestry.com also has a variety of state-level census that contain records from the 1770s and 1780s that we can match to the individuals in our pre-1790 certificates. Some examples include

1. Collection 2702, All Pennsylvania, U.S., Septennial Census, 1779-1863
2. Collection 2324, All U.S., Census Reconstructed Records, 1660-1820
3. Collection 3562, All New Jersey, U.S., Compiled Census and Census Substitutes Index, 1643-1890
4. Collection 3081, All Rhode Island Census, 1774
5. Collection 3552, All Maryland, U.S., Compiled Census and Census Substitutes Index, 1776-1890

The "Main" Table

The "Main" Table is a table, indexed by identity, that describes all debt holdings held by each individual. When available, information such as town of residence, occupation or other demographic characteristics are included. The idea is simple in practice, but alas, life is never that easy. The GitHub README.md (https://github.com/alicezg2/SPEOC-pt-1/blob/revamp/doc.md) outlines how I created this table for just the post-1790 Continental Debt securities. My code can definitely be improved upon, but it outlines the challenges you will encounter during the process.

Tools

GitHub

GitHub is kind of like Google Drive for code. Since we're all going to be working on the same project together, using GitHub will make our workflow much easier. If you are unfamiliar with GitHub or the terminal/command line, you should go through the following tutorials.
Github: https://product.hubspot.com/blog/git-and-github-tutorial-for-beginners
Terminal: https://ubuntu.com/tutorials/command-line-for-beginners#1-overview

Python and Jupyter

We use Python to analyze our data, and we write our code in jupyter notebooks (which are convenient because they allow for an iterative process of code writing). Here is a guide to installing Python and JupyterLab, which allows you to write jupyter notebooks.

Python Installation:
https://melaniewalsh.github.io/Intro-Cultural-Analytics/02-Python/01-Install-Python.html

Python is convenient to use because frequently, other people have helped write code, in the form of libraries, that make it easier for us to perform certain tasks such as analyze data. The majority of our analysis will be done using pandas. Here are some pandas tutorials:

Basic 10 minute pandas tutorial: https://pandas.pydata.org/docs/user_guide/10min.html
Pandas guide collection (for more specialized tasks):
https://pandas.pydata.org/docs/user_guide/

Name Matching Algorithms

Before Noah Webster's standardization of American English, there would often be various spellings of the same name. This makes our job more difficult because it means that, for example, one individual might have records listed under multiple different names. Identifying whether two people are the same or not is one of the major challenges you will tackle this summer. Here are a few techniques you might use to determine this.

1. Levenshtein Distance: Used in fuzzy string matching, Levenshtein Distance tells you how similar two strings are. Names that are spelled differently are still likely to be very similar, with differences of no more than a few letters.
   a. The python package **rapidfuzz** allows you to calculate the Levenshtein Distance and other variations of the algorithm for two strings.
2. Phonetic Similarity: Soundex algorithms, like the American Soundex Algorithm allow you to determine how similar two words are phonetically.
3. Census Records: If two names are similar, but census records show that they match to different people, then it is likely that those two names belong to different individuals
4. Data Characteristics: There may also be specific data features that can provide valuable information. Here are two examples from our data
   a. Location: If two similar names are both recorded as being from the same town, then that increases our confidence that the two names are the same
   b. Name Groups: In our post-1790 debt redemption records, there are sometimes names listed for the same entry (see example). If we have two similar names that correspond to

These are a few examples of techniques that can increase our confidence about whether two names are spelling variations of the same name, or two different identities, although they are by no means exhaustive. When used in tandem, they can help us form more authoritative judgements. Even when two names match exactly (as in they are the exact same), using technique 3 and 4 is also very important in helping us determine whether that name match corresponds to an identity match.

When I cleaned our continental debt redemption data, I used these techniques in my decision making process but I didn't create a formal algorithm to do so. I will describe this more in the project section, but one goal for the summer should be to create an algorithm that incorporates these steps (and other ones you may think of - this list is not exhaustive!) to determine whether two names represent the same or different identities in a systematized way.