Mission Impossible II
Innovator's Summit – RootsTech 2016

Salt Lake City, Utah – 3 February 2016

Diane Loosle, AG, CG and David E. Rencher, AG, CG

Problem Statement

Novice researchers have more success with direct evidence than indirect evidence. Post-1850 U.S. census records, for example, list the names of everyone living in the household (with the exception of slaves until 1870). Pre-1850 U.S. census records and tax records simply list the name of the head of the household and then a tick mark by age and gender for the other persons living in the dwelling. As the census years developed over time, the age brackets for these categories changed, making analysis of multiple years advantageous to identify individuals by name and to narrow potential birthdates. It also allows for analysis across censuses and tax records to locate potential relatives.

Census Year	Males Including Heads of Families	Females Including Heads of Families	Free White Persons	Slaves/Free Colored	
1790	16+ (inc. heads) Under 16	16+ (inc. heads) Under 16	All others	Total number of slaves Free colored not recorded	
1800/1810	Under 10 10-16 16-26 (inc. heads) 26-45 (inc. heads) 45+ (inc. heads)	Under 10 10-16 16-26 (inc. heads) 26-45 (inc. heads) 45+ (inc. heads)	All other free persons except Indians not taxed	Total number of slaves Free colored not recorded	
1820	Under 10 10-under 16 Between 16-18 of 16-under 26 (inc. heads) 26-45 (inc. heads) 45+ (inc. heads)	Under 10 10-under 16 Between 16-18 of 16-under 26 (inc. heads) 26-45 (inc. heads) 45+ (inc. heads)	Foreigners not naturalized # persons in: Agriculture Commerce Manufacturing All others	Slave M/F Under 14 of 14-26 of 26-45 45+	Free M/F Under 14 of 14-26 of 26-45 45+
1830/1840	Under 5 of 5-under 10 of 10-under 15 of 15-under 20 of 20-under 30 of 30-under 40 of 40-under 50 of 50-under 60 of 60-under 70 of 70-under 80 of 80-under 90 of 90-under 100 of 100 & upwards	Under 5 of 5-under 10 of 10-under 15 of 15-under 20 of 20-under 30 of 30-under 40 of 40-under 50 of 50-under 60 of 60-under 70 of 70-under 80 of 80-under 90 of 90-under 100 of 100 & upwards	White & Slave Deaf & Dumb under 14 of 14-under 25 of 25 & upwards Number of: Blind Aliens – Foreigners not naturalized	Slave M/F Under 10 of 10-under 24 of 24-under 36 of 36-under 55 of 55-under 100 above 100	Free M/F Under 10 of 10-under 24 of 24-under 36 of 36-under 55 of 55-under 100 above 100
1840	Added an entry for Pensioners for Revolutionary or military services with names and ages				

Examples of the Data

Example 1: Census Comparison to identify individuals

This example illustrates the challenges of identifying by name, individuals for each of the hash marks recorded in 1840 compared to the names and ages listed in the 1850 census. Probability algorithms are challenged when pulling this data together for analysis. Since relationships are not stated in either census, other individuals in the household can easily be misapplied to known family members.

1840-1850 Census Comparison for C. A. Hatch

1850 Census	**Noxubee, Mississippi**		**1840 Census**	**Sumter County, Alabama**
C. A. Hatch	47	cal. 1803	Male 30-40	cal. 1800-1810
[Ann?]*	XX		Female 20-30	cal. 1810-1820
[——?——]	XX		Male 20-30	cal. 1810-1820
[——?——]	XX		Male 15-20	cal. 1820-1825
Elizabeth	25	cal. 1825	[——?——]	XX
Edward	21	cal. 1829	Male 10-15	cal. 1825-1830
[——?——]	XX		Male 10-15	cal. 1825-1830
Catharine	19	cal. 1831	[——?——]	XX
[William A.]**	XX	cal. 1833	Male 5-10	cal. 1830-1835
[——?——]	XX		Female Under 5	cal. 1836-1840
Christopher	14	cal. 1836	Male Under 5	cal. 1836-1840
Hambleton	13	cal. 1837	Male Under 5	cal. 1836-1840
Richard	11	cal. 1839	Male Under 5	cal. 1836-1840
Joseph†	1	cal. 1849		
Hope‡	74	cal. 1776	Female 60-70	cal. 1770-1780

*Ann is listed as the mother on the tombstone marker of William A. Hatch in T22 R3E, Sumter Co., AL
**William A. Hatch marker, son of C. A. and Ann Hatch, died October 18, 1847, aged 14
†Joseph is likely a grandchild of C. A. Hatch by either Elizabeth or Catherine, neither of whom have a corresponding entry in the 1840 census. Catherine may be the wife of Edward; Elizabeth may be the husband of the one of the missing males born 1810-1820 or 1820-1825.
‡Hope is likely the mother of C. A. Hatch, both born in North Carolina

Example 2: Determining potential parents through generation analysis

Comparison of census and tax lists assists assists the genealogist to determine generations and potential parentage. In the table below comparing the Tax Lists of Nelson County, Tennessee across several years, Ignatius and Joseph appear from the beginning and so are likely a similar generation. Charles appears later and is therefore likely a new generation (or perhaps moved into the area. However, by analyzing the numbers in the age column it is clear that Charles is likely the son of Ignatius. This is a fairly obvious example of this analysis technique. Ignatius Cissell has a white male above 16 and under 21 living with him from 1804-1808. Then that individual becomes over 21 with 2 listed in the household in that category in 1809. Then in 1810, Charles appears in the same area and there it is back to only one male over 21 in Ignatius's household. It also appears that Ignatius may have passed away between 1810 and 1811 as he no longer appears in the record (or he moved away).

	Nelson County, Tennessee Tax Lists					
	Charles Cissell Beech Fork		Ignatius Cissell Beech Fork		Joseph Cissell Beech Fork	
	White Males above 21	White Males above 16 below 21	White Males above 21	White Males above 16 below 21	White Males above 21	White Males above 16 below 21
1792			1		1	
1793			1		1	
1794			1		1	
1795			1		1	
1796			1		1	
1797			1		1	
1798						
1799			1		1	
1800			1	1	1	
1801			1		1	
1802			1		1	
1803			1		1	
1804			1	1	1	
1805			1	1	1	
1806			1	1	1	
1807			1	1	1	
1808			1	1	1	
1809			2		1	
1810	1		1		1	
1811	1				1	
1812	1				1	

Example 3: Census comparison across multiple areas of residence

As families migrated west across the United States, others with whom they had a relationship or close friendship were likely to go with them or they would join others they knew. Analysis across multiple census years is very time consuming today to find the common surnames across an ancestral family's migration path. In the example below, Cecil/Sisle in green is the ancestral family. Notice "Gray" and "Dunbar" are also found in each location and all the surnames found in 1810 in Nelson County are found at one point or another in the other census locations. These are likely relatives.

1850 Marion County, IA

357	Haymaker	378	Plank
359	Doud	378	Stall
360	Burton	379	Michal
360	Elizabeth	379	Leach
361	Says	379	Michel
362	Morgan	380	McGruder
363	Jones	381	Hendricks
364	Welson	381	Frederick
365	Johnson	382	Rogers
366	Rousseau	384	Edwards
367	Robinson	385	Shaw
367	Frederick	386	Spilman
368	Craig	387	Hiddle
369	Day	388	Hendricks
370	Gushway	389	Stilwell
370	World	390	Howard
371	Gillman	390	Scott
372	Pearson	391	Neel
373	Wilsey	392	Starr
374	Robinson	393	Hibbard
374	Hendricks	393	Barnes
374	Sappenfield	394	Simmons
375	Carr	395	Gray
376	Linn	396	Mullins
376	Benison	396	Probst
376	Hendricks	396	Clase
376	Haven	397	Brobst
377	Cecil	419	Dunbar

1840 Stewart County, TN

25	Smith	1	Barry
24	Parkins	2	Milam
23	Hill	3	Dawlett
22	Summers	4	Ford
21	Holady	5	Amandon
20	McIntyre	6	Ford
19	Parish	7	Shrognorton
18	Parish	8	Broadway
17	Broadway	9	Williams
16	Dunbar	10	Williams
15	Kenady	11	Dunbar
14	Summers?	12	Lewis
13	Broadway	13	Lackber
12	Dunbar	14	Rufty
11	Lewis	15	Abit
10	Williams	16	Kenady
9	Barns	17	Kenady
8	Randolph	18	Hamlitt
7	Lewis	19	Pater
6	Summers?	20	Hall
5	Simmons?	21	Elkins
4	Morris	22	Haward
3	Ford	23	Hall
2	Gray	24	Jeffrey
1	Sisle	25	Beasley
	Sisle	26	Lee

1820 Montgomery County, TN

25	Pike	1	Fields
24	Humphries	2	Venable
23	Boothe	3	Malone
22	Trice	4	Taylor
21	Barr	5	Hinson
20	Cane	6	Doherty
19	Killebrew	7	Oldham
18	Pettice	8	Kill
17	Allen	9	Fields
16	Pollard	10	Spied?
15	Peterson	11	Robins
14	Bunck	12	Henderson
13	Brumfield	13	Travis
12	Dycus	14	Sims
11	Shepherd	15	Legraves
10	Chipman	16	Hubbard
9	Whitfield	17	Smith
8	Herring	18	Hatcher
7	Herring	19	Sims
6	Coleman	20	Hatcher
5	Herring	21	Peterson
4	Kent	22	Gray
3	Garrett	23	Hals?
2	Whitfield	24	Dinton
1	Whitfield	25	Crouch
	Cecil	26	Dunbar

1810 Nelson County & Washington County Kentucky In alphabetical order.

Cecil/Scisle
Dunbar
Gray
Hubbard
Malone
Mullins
Sims

How experienced researchers aggregate the data – workspace

A "workspace" provides a way for the researcher to gather any number of pieces of raw data, or evidence, which may or may not pertain to the same person or persons, but which have not yet led to a firm conclusion ready to be placed in a tree. The data is then organized into "problem sets" or groupings of data which the researcher analyzes to see what conclusions can be drawn.

A challenge for the researcher is to keep track of where pieces of data or evidence are being used in either closed conclusions in a tree or open ongoing analysis of "problem sets" which has not yet reached conclusion. Often genealogists employ tools at their disposal such as spreadsheets, timelines, notetaking applications, etc. For the examples provided above, the use of spreadsheets was extensive to analysis the data. This method, however, can be very time consuming requiring data entry or is lucky "copy & paste" of massive amounts of data and then utilizing the power of the spreadsheet software to sort and analyze the data. The examples provided above do not exhaust the potential ways in which the researcher "workspace" might be employed.

Challenges that developers may face

- Due to the current costs of indexing, many of the items in a record which would allow for this kind of analysis are not being keyed and are therefore not available to the computer algorithms.
- Access to proof sets for validating if developed algorithms are accurately identifying potential candidates is not simple today. They are found in genealogical publications and conference syllabi.
- No standard data model for the industry

Conclusion

With the large datasets now available some amazing analysis is now possible—call it the next level of "hinting". Let's compare with Amazon. Amazon brings to the customer other items they may have interest in based on the item selected and what items are similar in nature or were selected by others who also selected the same item. Could this apply to family history?

Imagine a world where the beginner family historian selects the records associated with their ancestor and then the system uses the data about that ancestor to compare against the large datasets available and then "recommends" likely individuals/surnames of those who may be related. Or the system may "recommend" potential parents for an individual based on "tick mark" analysis. Many of the techniques used by genealogists to analyze data could be employed programmatically with the right data being available and the right algorithms.

We believe this world is very possible. Help us achieve it as an industry.

Bibliography

1. Lee Albright and Helen F. M. Leary, "Designing Research Strategies – Strategy for the Census." Helen F. M. Leary, editor. *North Carolina Research* 2d. Raleigh: North Carolina Genealogical Society, 1996. Chap. 2, pp. 24-27.