When's It A Name?

Jeanne Howard Roper and L. David Roper
(jhoward@vt.edu and roperld@vt.edu)
1001 Auburn Drive
Blacksburg VA 24060-8123

Introduction

This is a study of the frequencies of a large number of given names in the population for the last five-hundred years. We use the word "frequency" of a given name to mean its percentage of occurrence among all persons in a given time period.

There are several ways to classify given names with respect to their frequencies:
- According to their longevity:
  - Those that occur over large time periods (e.g., Mary and John)
  - Those that occur only briefly (e.g. Tiffany and Clara)
  - Those that occur with high frequencies sporadically, i.e. with several peaks in the frequencies (e.g. Francis and David)
  - New names that are now on the rise (e.g. Jessica and Jordan)
- According to their frequency magnitude

We shall classify given names according to whether they occur with significant frequencies in the time period 1500-1997. In most cases we smooth the frequencies according to a three-decades running mean.

Data Base Used

One author (LDR) has several large genealogy data bases. The one used for this work is for his father's paternal (Roper) ancestors and their descendants, and all other Ropers for which data can be found. When this work was done the data base had 75,906 individuals in it. It seemed to us that this data base is large enough to give good statistics about given-name frequencies. We will discuss some checks on this assumption later.

The Roper family is mostly English, with some German and other nationalities mixed in, as with most English families. A sizable fraction of the data are for people who never left England, but most of the data are for people living in the United States.

Procedure:

The number of names in a genealogy data base for a given time period will reflect how difficult it is gather data from that time period, as well as any "baby booms" and "baby busts" for that period. We factor out these and other time-period effects by dividing the number of individuals with a certain given name by the total number of persons in that time period, i.e. we calculate the name frequency for that time period (number of persons with the selected given name divided by the total number of persons for the selected time period). Since these frequencies are much less than one, we report them as percentages.

Of course, we cannot include people for whom birth dates are not known. We did not try to estimate birth dates for those who had no birth dates. Those people were deleted from the decades' lists.

We chose decades (e.g. 1800-1809) as our time periods. Table I. shows the number of names that are in the Roper data base for each decade from 1600-1997.

Table I. Number of people with birth dates in the Roper data base for each decade.

1600-1610	95	1700-1709	141	1800-1809	650	1900-1909	2383
1610-1619	114	1710-1719	151	1810-1819	794	1910-1919	2418
1620-1629	122	1720-1729	157	1820-1829	1066	1920-1929	2359
1630-1639	120	1730-1739	186	1830-1839	1340	1930-1939	2138
1640-1649	146	1740-1749	225	1840-1849	1658	1940-1949	2676
1650-1659	131	1750-1759	247	1850-1859	1932	1950-1959	3325
1660-1669	128	1760-1769	301	1860-1869	1956	1960-1969	2536
1670-1679	174	1770-1779	348	1870-1879	2234	1970-1979	2169
1680-1689	175	1780-1789	415	1880-1889	2287	1980-1889	1785
1690-1699	153	1790-1799	518	1890-1899	2542	1990-1997	420

Note that before 1800 there are fewer than ~500 persons. Since any given name is about 0% to 5% in frequency for a given decade, and the statistical error for numbers is the square root of the number, the error for any name is large compared to the number for such small numbers. So the statistical errors are large for the earlier dates; we shall quantify this later for a specific case (the name David). One needs a much larger data base to get reliable results for earlier decades.

The genealogy program used is Personal Ancestral File 3.0. In it the FOCUS/ALL command was used to select the given names and to write them to a text file. The file was then edited to remove all names without dates and last names that fitted the given-name search criterion. Then the file was imported into Quattro Pro for Windows 6.0, which was used to parse the text lines into columns for counting the number of given names in each decade, to calculate the name frequencies for each decade and to plot the name frequencies versus decade.

Name-Search Criteria:

The following names were searched (Names in parentheses were included with the name not in parentheses. A * indicates any number of letters. [f] or [m] indicates female and male only. [fn] means "first name only".):

Aaron, Abigail, Abraham (Abe), Adam, Agnes, Albert (Alfred, Alvin), Alex*, Alice, Allison, Amy, Andrew (Andy), Angel*[f], Ann*[fn], Anthony (Tony), Arch*, Ashl*, Barb*, Bea*, Ben*, Ben F*, Bert*[f], Beulah, Bev*, Brad*, Brenda, Br(i,y)an, Brit*[f], Calvin, Cand*, Carol*, Cass*, Cath* (Kath*), Chad, Charity, Charles, Charlot*, Chris*, Christian, Clara, Clar*[m], Clyde, Craig, Cynthia (Cind*), Dan*, Dav*, Dawn, Deb*, Don*[m], Donna, Dor*, Doug*, Dustin, Earl*, Ed*, Effie, Eileen, Elizabeth (Liz*, Betty), Elmer (Delmer), Emily, Eric, Erin, Ethel, Eugene (Gene), Eva* (Eve*), Fay*, Flo*, For*[m], Fran*(not Frank*), Frank*, Fred*, Gayle (Gail), Gary, Gay*, George, George W*, Grace, Hanna*, Harold, Harriet*, Harv*, Heather, Helen*, Henry (Harry), Isaac, Jack*, Jacob, James (Jim*), Jane, Jason, Jay, Jean*, Jef* (Geof*), Jenn*, Jess*, Jessica, Jewel, Jill*, Joel, John* (not Jonathan or Johnathan), Jordan, Joseph (Joe), Josephin*, Josh*, Jud*[f], Keith, Kell*, Ken*, Kim*, Kris*, Lawrence (Laurence, Larry), Laur*, Leo* (not Leopold), Leona, Lil*, Linda, Lois(Eloise), Louis (Lewis)[m], Louis*[f], *oyd[m], Lucil*, Luke, Luther, Lynn*, Mabel*, Mahal*, Margar* (Magg*), Marj* (Marge*), Marilyn, Mark (Marc*), Martha, Martin, Mary (Polly), Mat*[f], Mat*[m], Maud*, May*[fn], Meagan (Megan), Melan*, Mich* (Mike)[m], Michel*, Mil*[f], Morgan, Muriel, Myrtle, Nancy, Nathan, Nell*, Nicol*, Noah, Oscar, Patric* (Pat*)[f], Paul*, Pearl, Pete*, Phil*, Phoebe (Phebe), Rach*, Rand*, Ray*, Rebe* (Beck*), Reuben (Rueben, Ruben), Rich*

3

(Rick*,Dick), Robert (Bob), Robin, Roger, Ron*, Roy, Ruth, Saman*, Sam*[m], Sara* (Sall*), Scott, Sher*[f], Shirley, Simon, Sol*[m], Stephan*[f], Stev* (Steph*)[m], Su*, Sylv* (Silv*)[f], Tam*, Theo* (Ted*), Ter*, Thomas (Tom*), Thomas J*, Tif*, Todd, Virg*[f] (Ginger), Virg*[m], Walt*, Wanda, Wayne, Will* (Bill*), Zach* (Zack*).
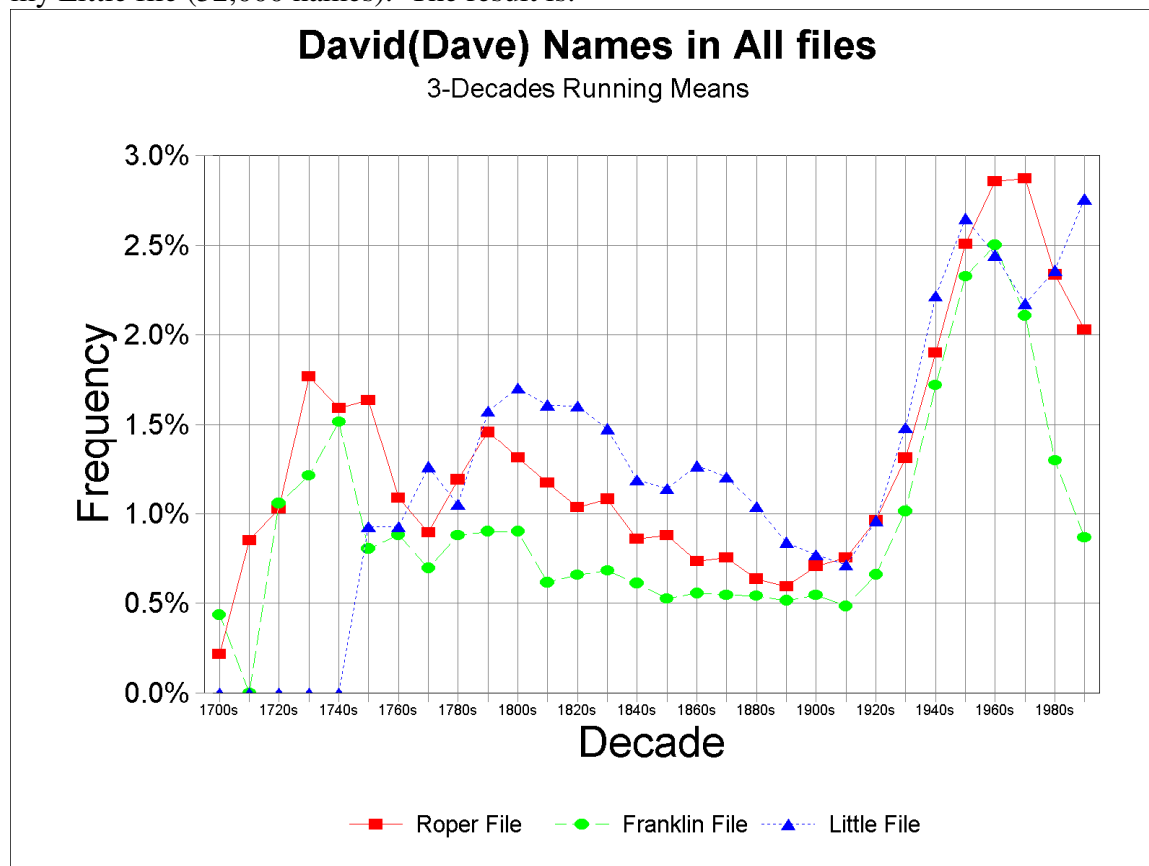
Note that most of these names actually are a collection of names; e.g. Su*=Susan, Susannah, Susanna, Susana, Susann, Sue, Susie, Suzan, Suzanne, Suzette and Suanne for the Roper file.

There are, of course, many other given names we could have studied, but we feel that this selection is a good representation of the major given names over the last five hundred years.
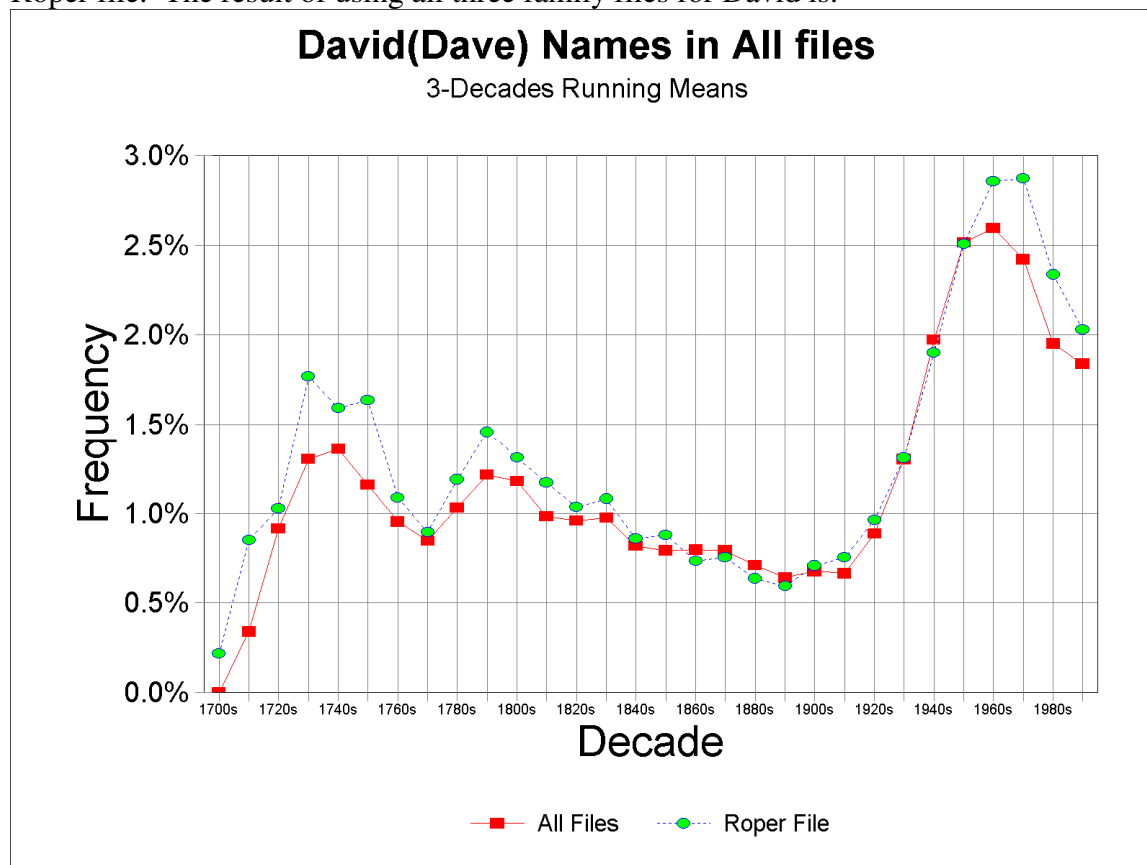
Results

Reliability of Roper Data

We could worry that certain given names are selected in some way for a given family file. For example, one would expect my Franklin family file to have more Benjamin names that most other family files would have. The Roper family file does not have any such influential given name to my knowledge. As a check on the reliability of the Roper file, I compared my second name, David, for the Roper file (~76,000 names) with my Franklin file (~64,000 names) and with my Little file (52,000 names). The result is:

Note that the main features are there for all three files.

One could combine all three files for a total number of names of 192,000 names. The result for any given name should be a smoother curve. This doubles the amount of work of just using the Roper file. The result of using all three family files for David is:

David(Dave) Names in All files
3-Decades Running Means

Indeed, the curve for all files is smoother than the curve for the Roper file alone, but not enough to make the extra work important. In fact, the main features are shown in both the Roper-file result and the all-files result.
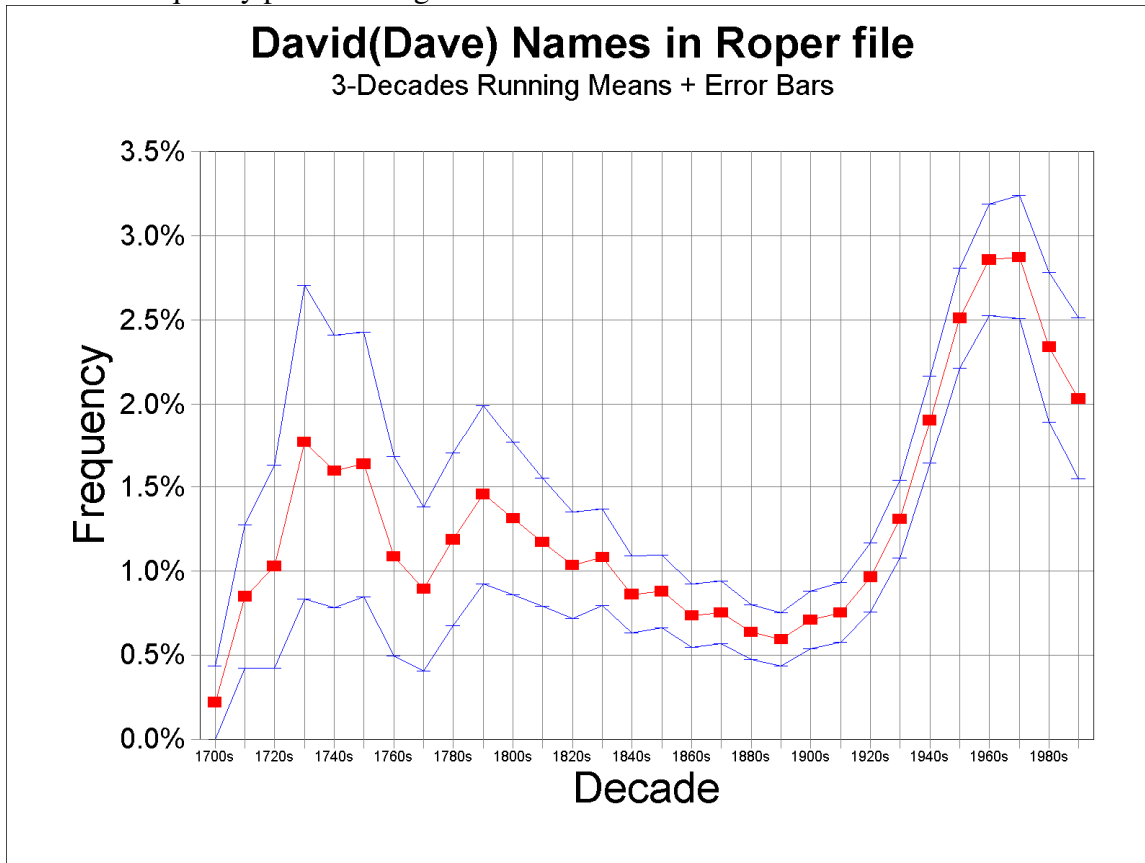
An interesting fact about all of these three files is that they have almost the same ratios of females to males and almost the same ratios of marriages per person. In the following table the first column is the ratio of females to number of persons and the second is the ratio of recorded marriages to number of persons.

File	Females/person	Marriages/person
Roper	47.0%	35.8%
Franklin	47.3%	36.2%
Little	46.8%	36.2%

The fact that all three files give about the same value for these two ratios is an indication that all of these files are a good random collection of data. The amount of overlap among these files is not very large according to my knowledge; the exact amount of overlap is not easy to determine.

Statistical Errors

The standard-deviation errors for counting data are given by the square root of the numbers for randomly distributed data. To get an idea of how big the errors are for the data used here, we show the frequency plot for the given name David with statistical errors:

David(Dave) Names in Roper file
3-Decades Running Means + Error Bars

Note that the errors are large for early times because the data numbers are small. So one should not give much credence to the data below 1600 and little more credence for data in the 1700s.
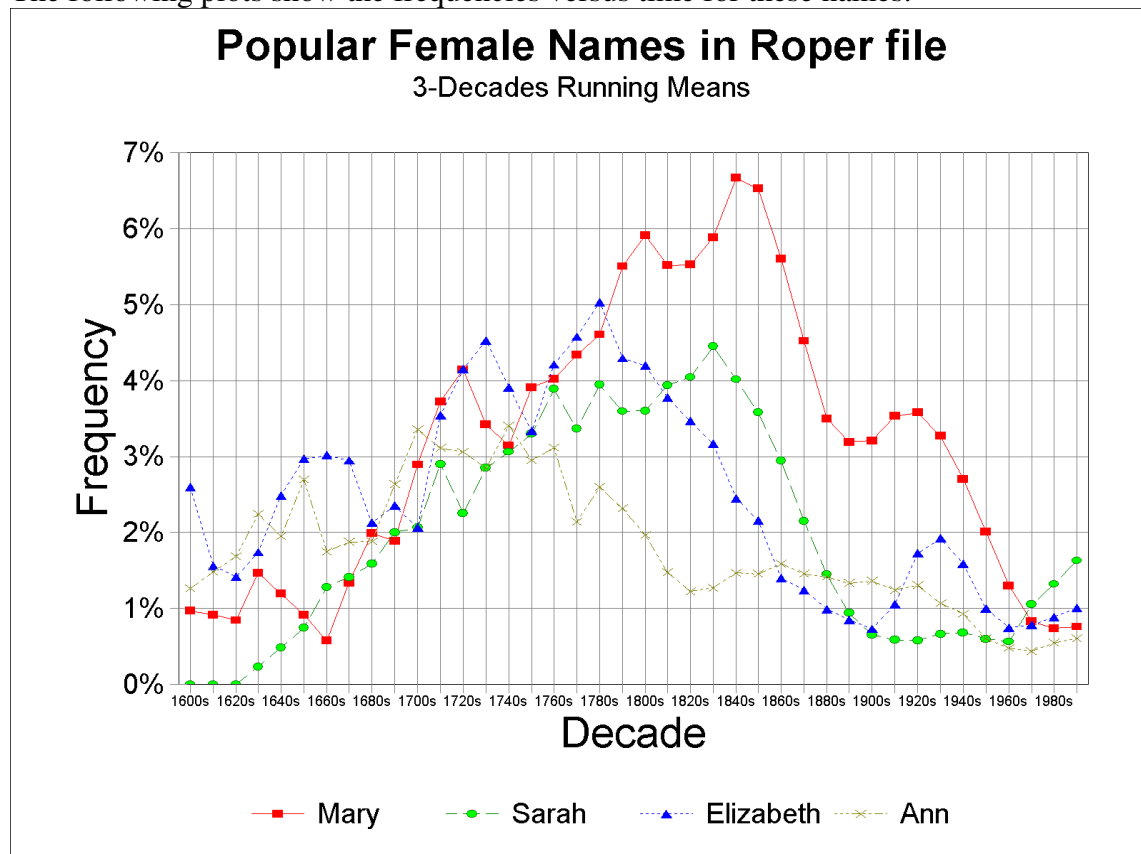
Names Categories

We shall divide the given names into the following categories:
- Most popular names over long time spans
- Briefly popular names
- Long-duration single-peaked names
- Double-peaked names
- Multiply-peaked names
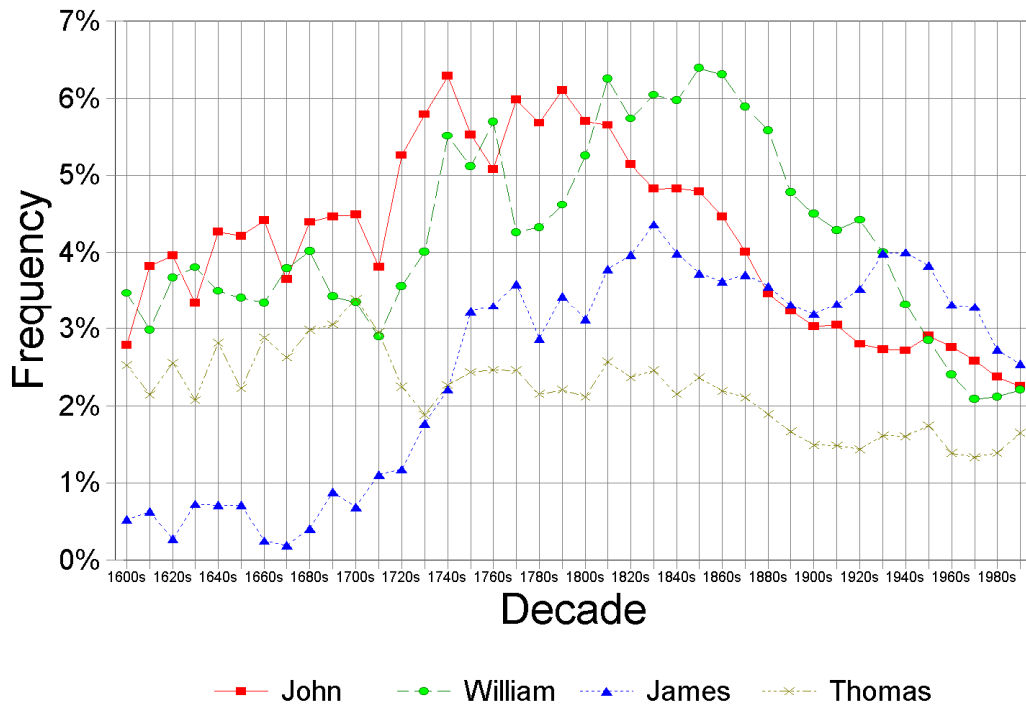- Resurgent names
- New names rising

Most Popular Names over Long Time Spans

We restrict this category to those names that exceeded 3% frequency at some time and that extended at a high percentage over at least one hundred years. The female names in this category are Mary, Sarah, Elizabeth and Ann. The male names are John, William, James, and Thomas. The following plots show the frequencies versus time for these names:

Popular Female Names in Roper file
3-Decades Running Means

Popular Male Names in Roper file
3-Decades Running Means

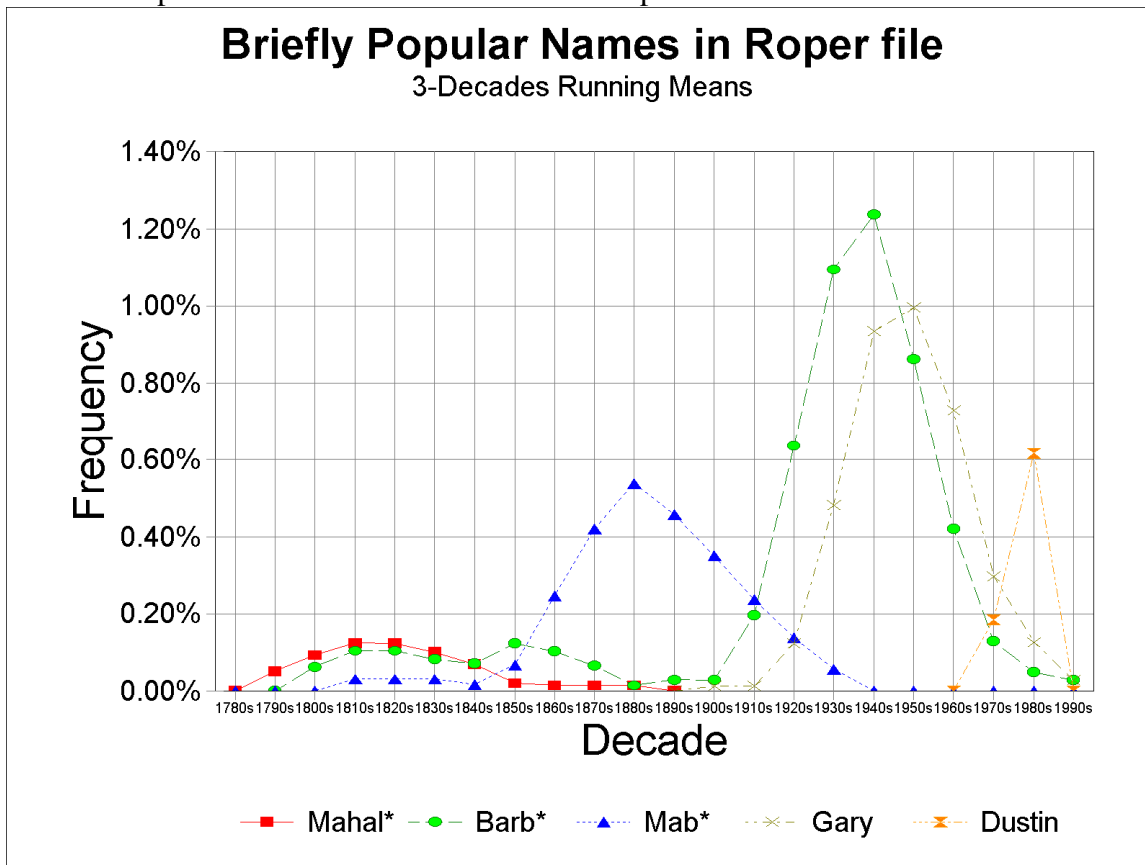Note that Thomas has remained between about 1.5% and 3% over the last four hundred years.

Briefly Popular Names

The following given names have a single frequency peak of less than a century duration. (Peak decades are in the left column and peak frequencies are given in parentheses for each name.)

1810s	Mahal*(0.12%)
1820s	Sol*(0.25%)
1880s	Eff*(0.3%), Flo*(1.13%), Leona(0.25%), Mab*(0.53%), Maud*(0.68%), Oscar(0.4%)
1890s	Clara*(0.5%), Bert*[f](0.6%), Beulah(0.3%),Ethel*(1.0%), Lil*(1.6%), Luther(0.42%), Myrtle(0.5%), Pearl(0.75%),
1900s	Clarence(0.66%), Earl*(0.8%), Elmer(0.35%), Fay*(0.11%), Muriel(0.17%), Virgil(0.2%)
1910s	Bea*(0.27), Clyde(1.08%), Leo*(0.8%), Leroy(1.3%), Lois(0.68%), (L,Ll,Fl,Cl)oyd(0.9%), Mil*(0.65%)
1920s	Harold(0.67%), Helen(1.1%), Jewel(0.25%), Lucil*(0.65%), Nell*(0.48%)
1930s	Eileen(0.18%), Eugene(1.52%), L(i,y)nda(2.6%)
1940s	Barb*(1.22%), Bev*(0.39%), Brenda(1.2%), Don*[m](1.75%),Jean*(3.1%), Jud*(1.55%), Ken*(1.03%), Larry(1.3%), Marilyn(0.4%), Pat*[f](1.12%), Roger(0.5%), Shirley(0.6%), Wanda(0.25%)
1950s	Donna(0.7%), Gary(1.0%), Gay(e)(0.25%), Gayle(1.0%), Ron*(0.95%), Wayne(2.0%)
1960s	Deb*(1.2%), Doug(1.08%), Jef*(1.17%), Keith(0.72%), Lynn*(4%), Mark(1.25%), Melan*(0.27%), Mike(2.0%), Sher(0.7%), Tam*(1.0%)
1970s	Amy(0.7%), Br(i,y)an(1.47%), Chad(0.87%), Craig(0.45%), Dawn(1.3%), Jason(1.6%), Jenn*(1.3%), Jill*(0.27%), Kell*[f](0.65%), Kim*(0.65%), Scott(1.7%), Todd(0.52%)
1980s	Dustin(0.6%), Heather(0.6%), Kris(1.05%), Michel*(1.8%), Stephan*(0.6%)

Some of these names have small, usually early, side lobes: Barb*, Bea*, Bev*, Br(i,y)an, Clarence (late side lobe), Craig, Deb*, Doug*, Elmer (late side lobe), Eugene, Flo* (late side lobe), Helen (early and late side lobes), Jason, Jef*, Jud*, Keith, Leo*, Leona, Lil* (late side lobe), L(i,y)nda, (L,Ll,Fl,Cl)oyd (late side lobe), Lois, Luther, Lynn*, Mab*, Marilyn (late side lobe), Mark, Mike, Mil*, Nell* (late side lobe), Oscar (late side lobe), Pat*, Roger, Scott, Sol*, Virgil
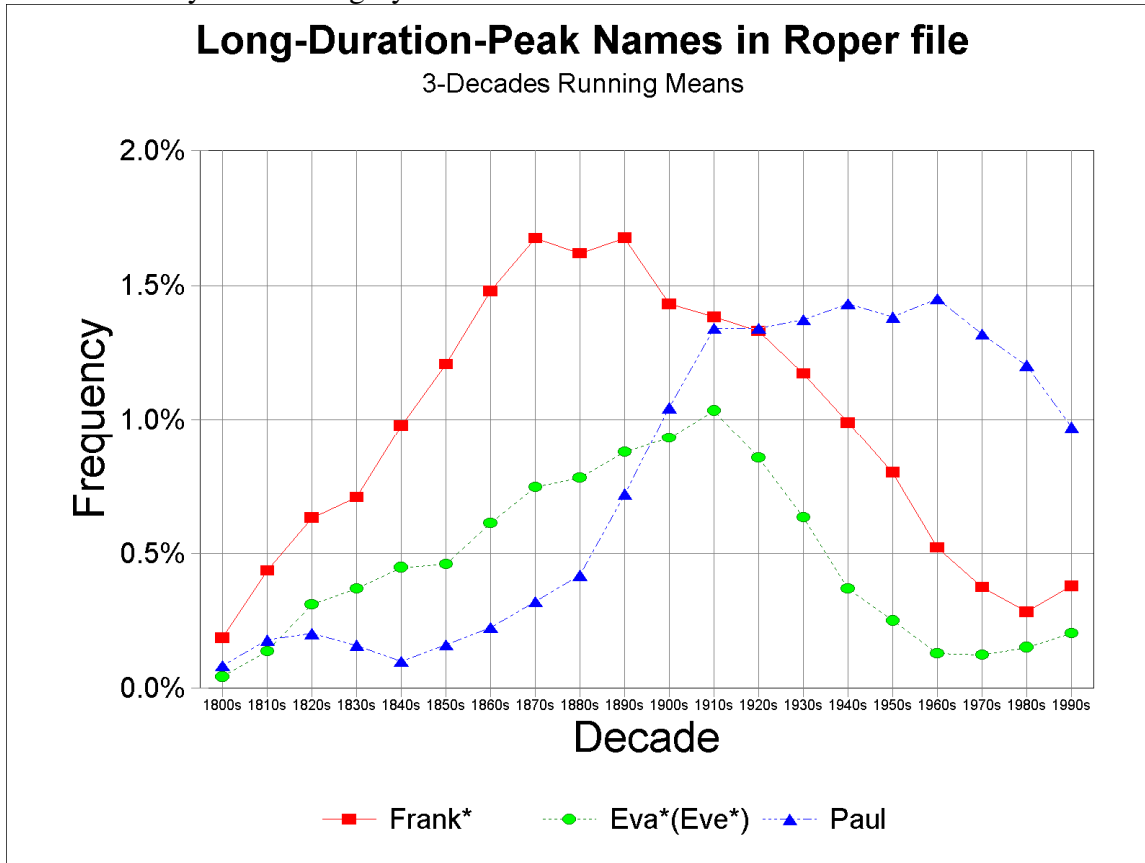
Some examples of these names are shown in this plot:

Briefly Popular Names in Roper file
3-Decades Running Means

Note that Barb* and Mab* have earlier values that are dominated by the later peak.

Long-Duration Single-Peaked Names

There are a few given names whose single peak extends over a time duration longer than one hundred years: Abr*(1790s,0.31%), Al(b,v,f)*(1860s,1.5%),   The three given names that are most obviously in this category are:
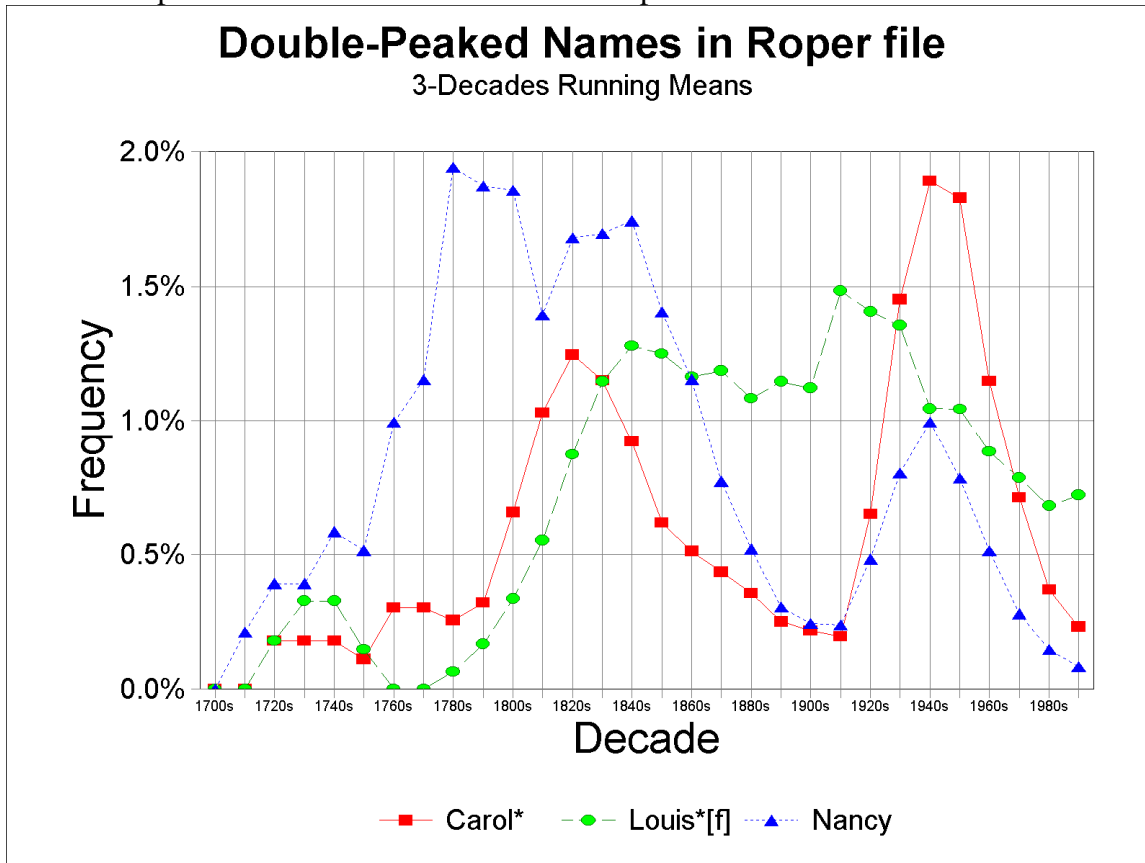

Double-Peaked Names

Some names have peaked twice over the last four hundred years. In this list we show in parentheses the following: (first-peak decade, first-peak frequency; valley frequency; second-peak frequency, second-peak decade)

Charlot* (1660,0.5%; 0%; 0.7%,1810)
Charles (1670,2.25%; 1.0%; 2.75%,1870)
Dorot* (1680,1.42%; 0.02%; 1.38%,1920)
Isaac (1700,0.43%; 0%; 0.62%,1780)
Abr* (1710,0.22%; 0%; 0.31%,1790)
Richard (1720,0.9%; 0.1%; 1.7%,1940)
Chris*[m] (1740,0.45%; 0%; 1.3%,1980)
Phoebe (1740,0.15%; 0%; 0.11%,1850)
Ter* (1740,0.42%; 0%; 0.7%,1950)
Fred* (1780,0.5%; 0.35%; 1.05%,1870)
Nancy (1780,1.9%; 0.2%; 1.0%,1940)
Jack* (1800,0.5%; 0.25%; 0.95%,1930)
Cynthia (1810,0.31%; 0.02%; 0.71%,1950)
For*[m] (1810,0.08%; 0.04%; 0.155%,1910)

11

Harv* (1810,0.26%; 0.11%; 0.03%,1880)
Marjorie (1810,0.09%; 0.02%; 0.4%,1920)
Simon (1810,0.08%; 0%; 0.05%,1960)
Calvin (1820,0.24%; 0.05%; 0.25%,1930)
Carol (1820,1.2%; 0.2%; 1.9%,1940)
Theodore (1820,0.12%; 0.09%; 0.28%,1920)
Angel* (1850,0.23%; 0.03%; 0.72%,1960)
Virginia (1850,0.48%; 0.33%; 1.04%,1920)
Joseph*[f] (1860,0.57%; 0.22%; 0.32%,1920)
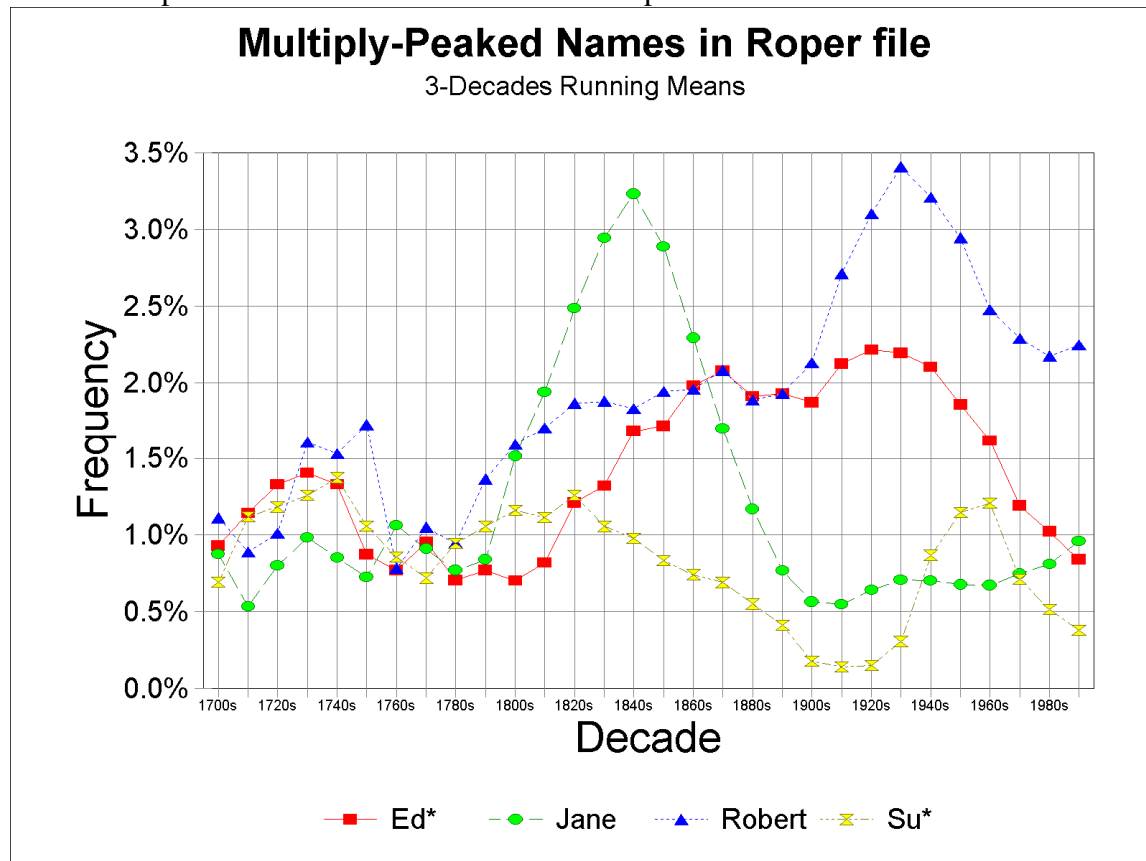Tif* (1870,0.09%; 0%; 0.18%,1960)

Some examples of these names are shown in this plot:

Multiply-Peaked Names

Several given names have had three or more peaks over the last several hundred years. Such names are Alice, Ann*, Cath*(Kath*), Dav*, Ed*, Franc*, George, Henry, Jess*, Jane, Louis*[f], Louis[m], Margaret, Martha, Rach*, Robert, Ruth, Su* and Walt*.

Some examples of these names are shown in this plot:

Multiply-Peaked Names in Roper file
3-Decades Running Means

Resurgent Names

 Some given names that were prominent in the past have recently been on the rise:

Andrew peaked at 0.9% in 1850s and rose to 2.2% in 1990s.
Mat*[m] peaked at 0.2% in 1820s and rose to 2.0% in 1990s.
Joseph peaked at about 2.2% over 1740s-1820s and rose to 1.9% in 1990s.
Dan* peaked at 1.9% in 1780s and rose to 1.8% in 1990s.
Alex* peaked at 0.75% in 1840s and rose to 1.5% in 1990s.
Emily peaked at 0.5% in 1850s and rose to 1.41% in 1990s.
Laura peaked at 0.8 in 1880s and rose to 1.35% in 1990s.
Ben* peaked at 1.2% in 1820s and rose to 1.25% in 1990s.
Jacob peaked at 0.67% in 1680s and rose to 1.2% in 1990s.
Rebe* peaked at 1.55% in 1810s and rose to 1.1% in 1990s.
Zac* peaked at 0.15% in 1840s and rose to 0.95 in 1990s.
Ray* peaked at 1.42% in 1930s and rose to 0.95% in 1990s.
Nathan peaked at 1.15% in 1740s and rose to 0.82% in 1990s.
Sam* peaked at 1.7% in 1820s and rose to 0.85% in 1990s.
Josh* peaked at 0.57% in 1810s and rose to 0.7% in 1990s.
Grace peaked at 0.65% in 1890s and rose to 0.51% in 1990s.
Hanna* peaked at 1.7% in 1780s and rose to 0.4% in 1990s.
Samantha peaked at 0.25% in 1850 and rose to 0.38% in 1990s.
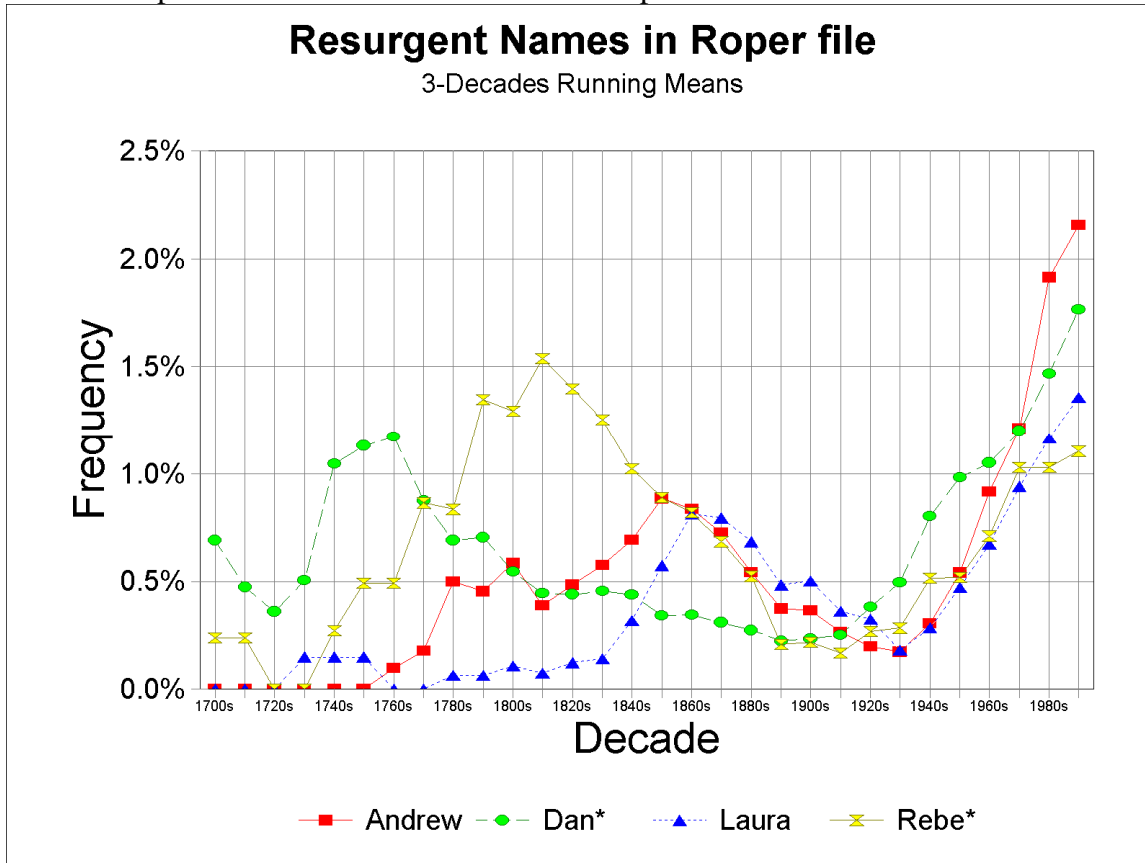Abigail peaked at 1.05% in 1720s and rose to 0.35% in 1990s.

Cand* peaked at 0.06% in 1850s and rose to 0.28% in 1990s.
Joel peaked at 0.5% in 1790s and rose to 0.23% in 1990s.
Noah peaked at 0.11% in 1870s and rose to 0.12% in 1990s,

Note that most of these names are Biblical names.

Some examples of these names are shown in this plot:

Resurgent Names in Roper file
3-Decades Running Means

New Names Rising

Some new given names have arisen over the last few decades and are still rising in frequency:

Nicole was near 0% in 1950s and rose to 2.3% in 1990s.
Jessica was near 0% in 1940s and rose to 1.2% in 1990s.
Jordan was near 0% in 1960s and rose to 1.19% in 1990s.
Ashl* was near 0% in 1900s and rose to 1.05% in 1990s.
Meagan was near 0% in 1940s and rose to 0.95% in 1990s.
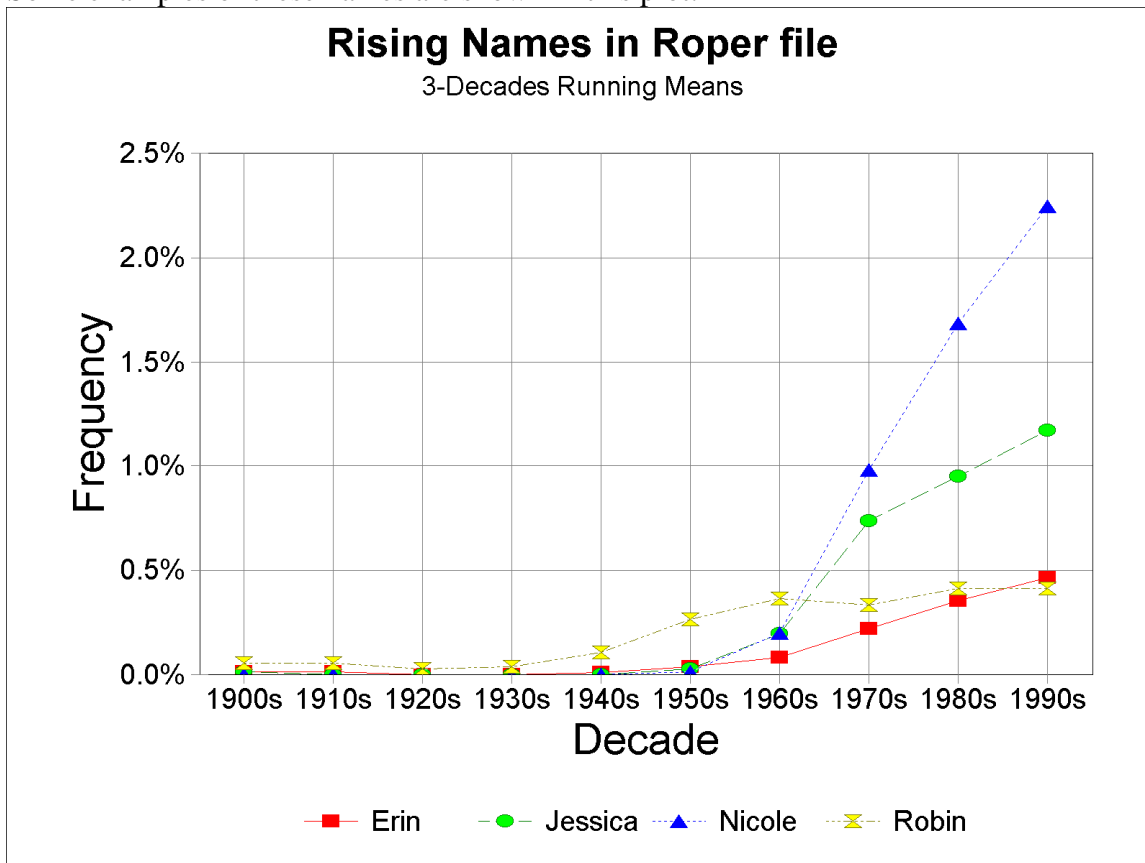Eric* was near 0% in 1870s and rose to 0.8% in 1990s.
Brit* was near 0% in 1950s and rose to 0.7% in 1990s.
Erin was near 0% in 1930s and rose to 0.465 in 1990s.
Robin was less than 0.05% in 1930s and rose to 0.4% in 1990s
Jay was near 0% in 1870s and rose slowly to 0.35% in 1990s.

Some examples of these names are shown in this plot:

Rising Names in Roper file
3-Decades Running Means

Conclusion

We believe that the Roper-family data file used in this study is a good representation of the given-names distributions in mostly English families in the United States.

Some of our guesses about when given names were popular were corroborated by this study, but there were some surprises. Some of the names we thought were very old (e.g. Bertha) were popular only for a brief time period back in the 1800s.

Many of the old long-term given names (e.g. John and William) have had rapid declines in the 1900s decades.