PNIS HARD

# We are entering an unprecedented age in baby name flux

Department of American Nomenclature[1], Division of Legal Compliance[1]
1 - PNIS

## Introduction

Federal law requires that, within the first year of their establishment, new publications must publish an article utilizing the United States Social Security database on baby names[1]. As a result of this law, articles on baby names have been published in: online blogs [e.g., Prooffreader (link, link, link, link), FlowingData (link, link, link), FiveThirtyEight, Jezebel, Waitbutwhy, nymag, and Fulcrum. Then, there's Freakonomics and babynamewizard which have more baby name articles than there are baby names], academic journals (e.g., *Proceedings of the National Academy of Sciences*; *Proceedings of the Royal Society B*; *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; *Proceedings of the 3rd Workshop on Computational Linguistics for Literature*; *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*; *Proceedings of the 13th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*; *Proceedings of the 16th ACM International Conference on Multimedia*; and *Academy of Management, Proceedings*), and an academic journal that seems to have originated as a result of this database[2]. Therefore, in accordance with US Statute 72-4224[3], PNIS is proud to present our analysis of baby names in the United States.

What exactly have all these previous studies uncovered about our denomination of offspring? Well, we've found out that the popularity of top-ranked names has decreased (Prooffreader, FlowingData, Twenge et al. 2010), boys names that end in 'n' are on the rise (Proofreader), "Jennifer" used to be popular (Jezebel), girls names are more varied than boys names (Li 2012), popular names are less common on frontiers (Varnum and Kitayama 2011), and that you can't attend a data visualization or computing conference without running into this dataset.

In choosing our particular research topic, we wanted something unique that no one else had ever studied, or, at the very least, was popular many years ago but has since decreased in popularity. Mostly, this desire stems from the eminent publishability and coolness of novel research papers. Plus, many famous people are choosing really unique research topics and we kind of want to be just like them. Deep inside, though, we're hopeful that the research topic we pick will start a new hip trend.

We've observed that prospective parents are choosing names solely on their uniqueness[4], which, in our humble opinion, is not the best reason for a name. In the past, inspiration for names has come from politicians, relatives, and stars of the cinema, while current inspiration originates from compass directions, various fruits, or popular tourist destinations. The desire to pay homage to a respected and renowned person has been replaced by the desire to be unique, a most

1 This is obviously not true. However, the rest of this article is based on real data, making it a true HARD article
2 That would be *NAMES: A Journal of Onomastics*, with onomastics meaning the study of proper names (and not the study of masturbation statistics). This journal launched in 1953 and has published such articles as: "Variation in automobile naming", "Is *1069* a name?" and "The stuff of which names are made: a look at the colorful and eclectic namecraft of Lord Dunsany".
3 Well, ok, this also not true, but this is really a continuation of the joke noted in the first footnote. Plus, we're not even sure if the US does statutes (we're also not sure what a statute is).
4 Our source here is a relative of one of the authors that will remain nameless.

disturbing trend. Thus, we've decided to analyze the Social Security (SS) database to see how four different metrics have changed over time: 1) number of unique names, 2) the diversity of names, 3) the evenness of names, and 4) flux, the year-to-year change in name popularity.

## Methods

The Social Security Administration of the United States of America has a database containing the frequency of first names of US citizens from 1880 to present (we downloaded the dataset that contains data until 2012). There are several things to note about this dataset:

- SS in the US didn't start issuing cards until 1937[5], so the data from 1880 to 1937 represents older people (i.e., not newborn) that signed up for SS (probably for employment). So, for example, an entry of 500 Irene's in 1900 in the dataset doesn't mean that 500 baby girls were named Irene in 1900, but that 500 Irene's born in 1900 signed up for SS in 1937. In other words, the data before 1937 is a bit dubious.
- Even for some time after 1937, it's debatable that everyone registered their newborns with the government.
- Gender designations can be a bit wonky at times (see Prooffreader).
- For privacy reasons, the database only gives names with at least 5 entries.
- The database also includes entries such as "Unknown", "Baby", or "Infant", which are too sad to warrant further comment.

The dataset for any particular year is just a list with three columns: Name, Gender (M or F), and Number (corresponding to the number of people having that first name). We first separated the dataset by gender, and then calculated the proportion of each specific name (for instance, in 2012, the proportion of newborn girls named Sophia was 0.0127, or 1.27%). Using this proportion we were able to calculate:

1. **Diversity**. Diversity could just be the number of unique baby names. But, when you have information on the abundance of each name, you can calculate diversity indices, which consider the frequency of each type (i.e., for us, a "type" is any particular name). A high diversity index value indicates that the names are evenly spread out (i.e., one name doesn't dominate over all the others). We used the Shannon Index (abbreviated as $H'$), which is calculated as: $-\sum p_i \ln(p_i)$, where $p_i$ is the proportion of the population given a particular name (so, in 2012, $p_{SOPHIA} = 0.0127$).
2. **Evenness**. This is related to diversity, but is perhaps a bit easier to comprehend because it can only take a value between 0 and 1 and is often thus expressed as a percentage (e.g., a value of 0.75 means that baby names for that particular year were 75% even). Higher numbers indicate names are more spread out, while lower numbers indicate that several names are more dominant. We calculated evenness (abbreviated as $J'$) as: $H'/H'_{MAX}$, where $H'_{MAX}$ is the natural log of the total number of unique names in that year (so, as you can see, evenness also takes into account the number of baby names per year).
3. **Flux**. This one's easy: take the proportion of any name in one year, subtract from it the proportion of that name in the very next year, and obtain the absolute value (in other words, don't have any negative numbers). Then just add up all the values for all names to get an overall measure of how much every name changed from the previous year. Large values indicate a lot of change from the previous year.

We calculated these three metrics (and some others self-explanatory ones like number of unique names) for each gender and each year, and then made some cool graphs[6].

## Results

We made six cool graphs, which we'll call Figure 1 (note the shading from 1880 to 1937 to reiterate the dubiousness of that data). Here are some trends that we saw:

**Gender**: 1) girls consistently have more unique names than boys (top right panel), 2) girls names are more spread out than boys names, or, in other words, more boys share names than do girls (mid left and both bottom panels), 3) girls names consistently go through more year-to-year change than boys (mid right panel).

**Historical**: 1) the number of unique names has gone up nearly 4-fold (girls) or 3-fold (boys) since the 1930's (top right panel), 2) names of both genders have gotten more spread out, or, in other words,

5 Actually, in November 1936 they started requesting applications from postal workers.

6 US Statute 72-4224 also requires that in any graph made using baby names, all data concerning boys must be represented as light blue, and all data concerning girls must be represented as pink.

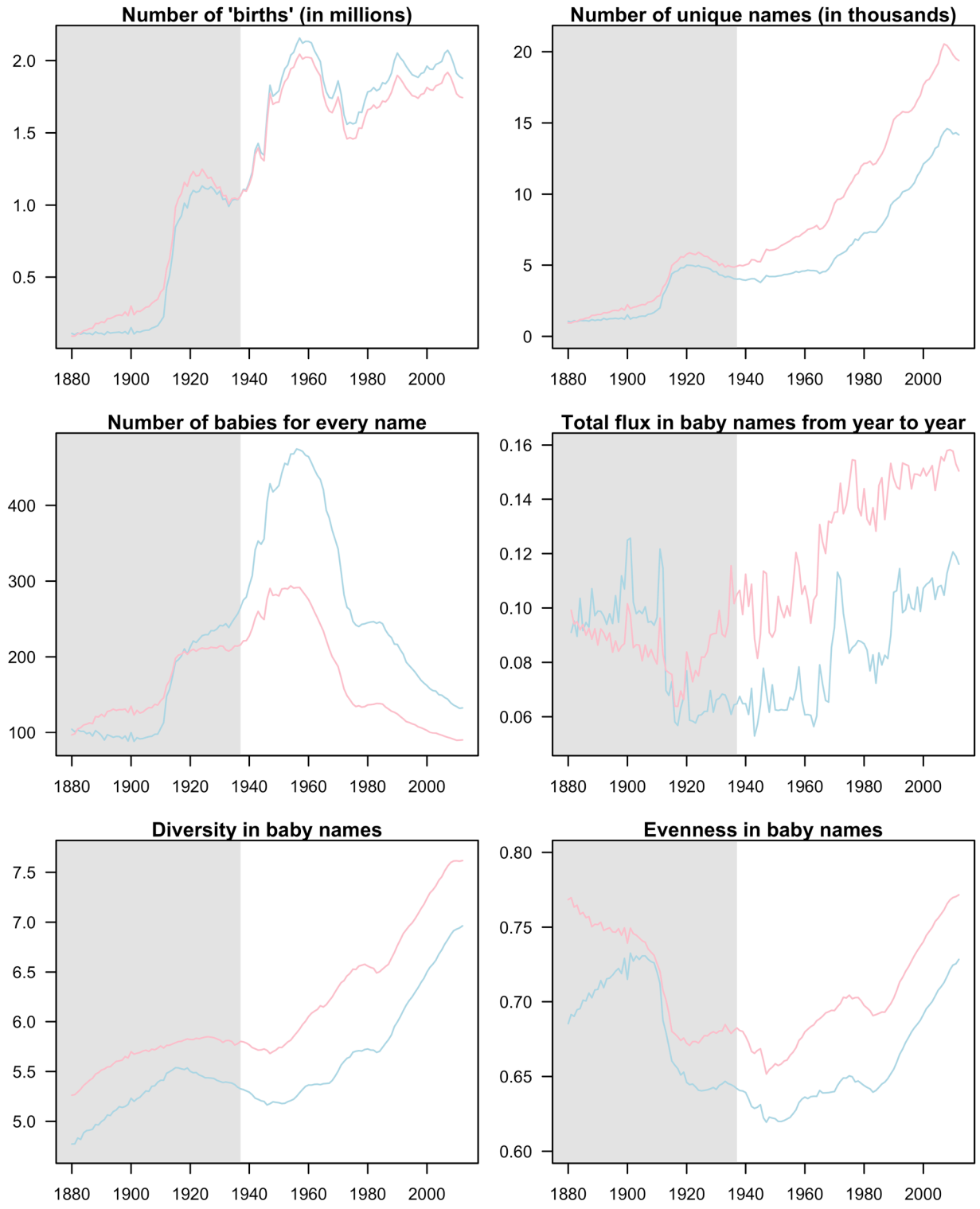


**Figure 1.** Various metrics of American baby names over time. Light blue represents boys and pink represents girls. X-axis denotes year. Gray shading from 1880 to 1937 denotes that SS applications did not begin until 1937.

less names have been dominant over time (mid left and both bottom panels), 3) flux in baby names has steadily increased since the 1930's (mid right panel)

We also put together a table of the names with the biggest gains and losses in proportion for every year and gender (see Appendix). Mostly this table exists so we can marvel about the silly names of the past and see if our parents were hip to the then-current naming trends.

## Discussion

Well, first we supported what just about everyone else has found out with this dataset: namely that the distribution of baby names has evened out over the years. In other words, the top-ranked names are losing popularity. In other, other words, fewer babies are being called the same name; there's much more diversity out there. This trend has been recognized by researchers publishing in both academic journals (e.g., Twenge et al. 2010, Li 2012) and blogs (e.g., FlowingData[7], Fulcrum, Prooffreader). Interestingly, different calculation methods have been used to arrive at the same conclusion: for example, Li (2012) used the Gini index, which is typically used to measure wealth distribution, and Fulcrum used the Herfindahl-Hirschman Index, which is typically used to measure how concentrated a market is (i.e., is there perfect competition or is it a complete monopoly?). We used Shannon's Diversity and Evenness, which are typically used to assess species diversity in different habitats.

So, what's behind this trend? Fulcrum identifies three different possibilities: 1) increased immigration, especially from areas that have little first name overlap with the US, 2) a desire for uniqueness, and 3) the fact that older names die hard (for example, there were still 28 girls named Bertha in 2012). Twenge et al. (2010) also point out increased individualism as a driver of this trend, but question immigration as an important factor, as their models controlled for immigration rate.

All of these explanations probably have some merit, but we would also like to mention that baby name diversity also seems to have risen with the increasing annual temperature of the US (i.e., climate change). In fact, there is a significant positive relationship between average annual temperature of the US (since 1895) and the number of unique male and female names[8]. In fact, for every increase in one degree Fahrenheit, there are 2,191 new male names and 3,127 new female names introduced in the US (Figure 2). Its unknown why hotter temperatures are causing parents to go crazy with baby names[9], but as the US is expected to warm about 5° F this century, you can expect about another 10 to 15 thousand more baby names that are even dumber than Even, Ever, Zaydin, and Kyler.

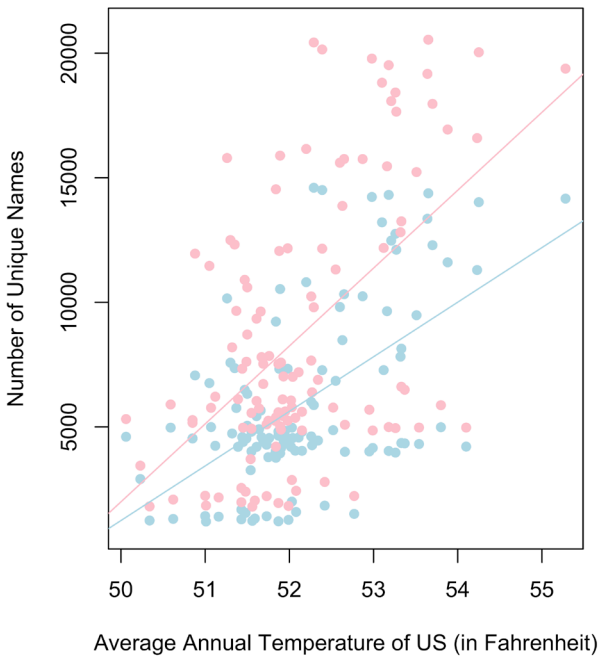


**Figure 2.** Warmer temperatures in the US are positively related to the number of unique names chosen by American parents. Each dot represents one year from 1895 to 2012 (Light blue = boys, pink = girls)

## References

Li W (2012) Analyses of baby name popularity distribution in U.S. for the last 131 years. Complexity 18:44-50.

Twenge J, Abebe EM, Campbell WK (2010) Fitting in or standing out: trends in American parents' choices for children's names, 1880-2007. Social Psychological & Personality Science 1:19-25.

Varnum MEW, Kitayama S (2010) What's in a name? Popular names are less common on frontiers. Psychological Science 22:176-183.

## Appendix (next page)

Boy and girls names that increased and decreased the most from the previous year.

7 Note that there is a graph in this post that is extremely similar, if not identical, to our Figure 1 mid left panel.

8 *P*-values for both the male and female correlation are <0.001. $R^2$ values are not given because they are very weak.

9 Although see here for an interesting hypotheses

| | Boys | | Girls | |
|---|---|---|---|---|
| **Year** | ***Increaser*** | ***Decreaser*** | ***Increaser*** | ***Decreaser*** |
| 1881 | Chester | Charles | Ethel | Mary |
| 1882 | Robert | John | Helen | Nellie |
| 1883 | Charles | James | Bertha | Mary |
| 1884 | Grover | Charles | Pearl | Bertha |
| 1885 | Albert | William | Elsie | Mary |
| 1886 | Arthur | Grover | Edna | Emma |
| 1887 | Joseph | John | Ethel | Minnie |
| 1888 | Harry | John | Hazel | Mary |
| 1889 | Roy | Grover | Ruth | Emma |
| 1890 | Charlie | William | Helen | Mary |
| 1891 | Joseph | Frank | Ruth | Alice |
| 1892 | Grover | William | Ruth | Anna |
| 1893 | Joseph | John | Esther | Mary |
| 1894 | Harold | Edward | Marie | Ruth |
| 1895 | James | Charles | Marion | Mary |
| 1896 | William | John | Helen | Jessie |
| 1897 | James | John | Helen | Anna |
| 1898 | Dewey | William | Helen | Mary |
| 1899 | Willie | Dewey | Mildred | Ethel |
| 1900 | James | Dewey | Annie | Anna |
| 1901 | Charles | James | Gladys | Annie |
| 1902 | Robert | Wiliam | Dorothy | Elsie |
| 1903 | Joseph | Henry | Dorothy | Minnie |
| 1904 | Alton | William | Mildred | Grace |
| 1905 | Samuel | John | Alice | Florence |
| 1906 | Joseph | James | Alice | Clara |
| 1907 | James | William | Evelyn | Alice |
| 1908 | William | Charles | Dorothy | Evelyn |
| 1909 | James | John | Helen | Anna |
| 1910 | Willie | William | Mary | Edna |
| 1911 | Joseph | Willie | Dorothy | Annie |
| 1912 | Woodrow | James | Dorothy | Ethel |
| 1913 | Robert | Willie | Helen | Bertha |
| 1914 | Robert | Woodrow | Dorothy | Edna |
| 1915 | Robert | John | Ruth | Ethel |
| 1916 | Robert | Willard | Virginia | Anna |
| 1917 | Robert | Charles | Virginia | Anna |
| 1918 | Robert | William | Lorraine | Mary |
| 1919 | Robert | Woodrow | Betty | Helen |
| 1920 | Robert | John | Betty | Anna |
| 1921 | Robert | Walter | Betty | Helen |
| 1922 | Robert | Warren | Betty | Ruth |
| 1923 | Robert | George | Betty | Helen |
| 1924 | Robert | Warren | Betty | Helen |
| 1925 | Richard | John | Gloria | Margaret |
| 1926 | Richard | Calvin | Barbara | Helen |
| 1927 | Donald | William | Mary | Helen |
| 1928 | Herbert | William | Dolores | Helen |
| 1929 | Donald | Alfred | Joan | Ruth |
| 1930 | Richard | Herbert | Joan | Dorothy |
| 1931 | Robert | William | Barbara | Dorothy |
| 1932 | Ronald | Robert | Barbara | Betty |
| 1933 | Franklin | Robert | Barbara | Dorothy |
| 1934 | Ronald | Franklin | Shirley | Betty |
| 1935 | David | Franklin | Shirley | Betty |
| 1936 | Robert | Donald | Carol | Shirley |
| 1937 | David | Donald | Barbara | Shirley |
| 1938 | Gary | Donald | Judith | Shirley |
| 1939 | David | Donald | Judith | Shirley |
| 1940 | Gary | Donald | Linda | Shirley |
| 1941 | Ronald | Donald | Linda | Betty |
| 1942 | Douglas | Donald | Linda | Patricia |
| 1943 | Michael | Douglas | Linda | Barbara |
| 1944 | Michael | Ronald | Cheryl | Barbara |
| 1945 | Michael | Charles | Linda | Mary |
| 1946 | Michael | James | Susan | Mary |
| 1947 | David | Richard | Linda | Carol |
| 1948 | Michael | Richard | Deborah | Judith |
| 1949 | Michael | James | Deborah | Linda |
| 1950 | Michael | John | Deborah | Linda |
| 1951 | Michael | James | Deborah | Linda |
| 1952 | David | Larry | Debra | Linda |
| 1953 | Michael | John | Debra | Linda |
| 1954 | Mark | Dennis | Debra | Linda |
| 1955 | David | Richard | Debra | Mary |
| 1956 | Mark | David | Karen | Deborah |

| | | | | |
|---|---|---|---|---|
| 1957 | Mike | Robert | Cindy | Deborah |
| 1958 | Timothy | James | Tammy | Cynthia |
| 1959 | Mark | Michael | Donna | Debra |
| 1960 | Jeffrey | Richard | Lisa | Debra |
| 1961 | John | Richard | Lisa | Donna |
| 1962 | Scott | Mark | Lisa | Linda |
| 1963 | Paul | Mark | Lisa | Linda |
| 1964 | John | Mark | Dawn | Lori |
| 1965 | Christopher | John | Lisa | Mary |
| 1966 | Christopher | John | Michelle | Karen |
| 1967 | Michael | James | Melissa | Karen |
| 1968 | Matthew | David | Jennifer | Mary |
| 1969 | Jason | Timothy | Jennifer | Lisa |
| 1970 | Jason | David | Jennifer | Lisa |
| 1971 | Christopher | David | Jennifer | Lisa |
| 1972 | Christopher | Scott | Jennifer | Lisa |
| 1973 | Jason | David | Heather | Lisa |
| 1974 | Jason | Robert | Heather | Lisa |
| 1975 | Joshua | Brian | Amanda | Jennifer |
| 1976 | Jeremy | Michael | Jamie | Michelle |
| 1977 | Joshua | Michael | Kelly | Amy |
| 1978 | Nicholas | Jason | Crystal | Amy |
| 1979 | Joshua | Jason | Amanda | Amy |
| 1980 | Justin | Jason | Tiffany | Melissa |
| 1981 | Brandon | Jason | Jessica | Melissa |
| 1982 | Christopher | Jeremy | Ashley | Melissa |
| 1983 | Matthew | Jason | Ashley | Melissa |
| 1984 | Joshua | Jason | Ashley | Jennifer |
| 1985 | Kyle | Jason | Brittany | Jennifer |
| 1986 | Andrew | Joshua | Whitney | Jennifer |
| 1987 | Justin | Jason | Kayla | Jennifer |
| 1988 | Justin | Jason | Brittany | Ashley |
| 1989 | Ethan | Jason | Brittany | Jessica |
| 1990 | Jordan | Justin | Taylor | Amanda |
| 1991 | Dylan | Christopher | Shelby | Brittany |
| 1992 | Dylan | Michael | Taylor | Jessica |
| 1993 | Austin | Michael | Taylor | Chelsea |
| 1994 | Austin | Michael | Alexis | Ashley |
| 1995 | Austin | Michael | Madison | Jessica |
| 1996 | Noah | Michael | Madison | Jessica |
| 1997 | Jacob | Cody | Hannah | Jessica |
| 1998 | Noah | Christopher | Emma | Jessica |
| 1999 | Seth | Michael | Grace | Jessica |
| 2000 | Ethan | Austin | Trinity | Brittany |
| 2001 | Logan | Brandon | Isabella | Brittany |
| 2002 | Ethan | Jacob | Isabella | Taylor |
| 2003 | Aidan | Austin | Emma | Ashanti |
| 2004 | Aiden | Zachary | Ava | Hannah |
| 2005 | Landon | Jacob | Ava | Alexis |
| 2006 | Landon | Ryan | Addison | Emily |
| 2007 | Jayden | Joshua | Addison | Emily |
| 2008 | Aiden | Christopher | Peyton | Hannah |
| 2009 | Liam | Aidan | Isabella | Hannah |
| 2010 | Mason | Joshua | Sophia | Madison |
| 2011 | Mason | Jacob | Harper | Isabella |
| 2012 | Liam | Jacob | Harper | Chloe |