

Trend in First Names Foreshadowed Hillary Clinton's Electoral Defeat

Stefano Ghirlanda

Brooklyn College

Abstract

I examine trends in the popularity of first names around the years of USA presidential elections, showing that the names 'Hillary' and 'Hilary' decreased abruptly by more than 90% in popularity following the 1992 election of Hillary Clinton's husband Bill. I show that this outcome is unique to the 1992 election, and argue that it may evidence a "dislike" for Hillary Clinton's public image among both Democratic and Republican voters, which may have eventually contributed to Hillary Clinton's losing the 2016 presidential election.

Introduction

Choosing a child's name is, for most expectant parents, charged with personal, social, and cultural meaning (Edwards and Caballero, 2008). Naming decisions are usually taken only after lengthy deliberation, including the perusal of specialized books, web-sites, and even paid consultants (Alter, 2007). Despite the effort devoted to them, naming decisions are not entirely rational, and can be influenced by seemingly extraneous factors of which parents are not aware. For example, names related to those of destructive hurricanes have been found to *increase* after hurricane strikes (Berger et al., 2012). Moreover, names often undergo fashion-like popularity cycles, lasting from a few years to several decades, in which a name turns from unpopular to popular, only to become unpopular again (Hahn and Bentley, 2003; Berger and Le Mens, 2009; Acerbi et al., 2012). Relating these cycles to other cultural events may yield valuable insight into individual decision making, as well as into social and cultural dynamics (Acerbi et al., 2012; Gureckis and Goldstone, 2009; Berger and Le Mens, 2009; Berger et al., 2012).

Here I show that trends in first names foreshadowed, in the USA, the defeat of Democratic Party candidate Hillary Clinton in the 2016 presidential election. Clinton's defeat came as a surprise to many experts and laypersons, and its causes are debated (Long, 2016; Klein, 2016; Ball, 2016). Yet, it is possible to infer from naming trends that, to the extent that we name our children after people we

Corresponding author's e-mail: drghirlanda@gmail.com

Citation: Ghirlanda, Stefano. 2017. Trend in First Names Foreshadowed Hillary Clinton's Electoral Defeat. *Cliodynamics* 8: 48–58.

admire, Hillary Clinton had a very negative public image since at least 1992. In that year, Clinton's husband Bill gained the USA Presidency, resulting in extensive media coverage of Hillary in her new role as First Lady. Before 1992, the names 'Hillary' and 'Hilary' had been increasing in popularity for several decades. After 1992, however, their popularity dropped suddenly 10-fold. I show that such a sudden reversal is unique among naming trends, and is also unique to the 1992 election. In other elections between 1884 and 2012, a First Lady's name had little impact on naming trends. These considerations, and others detailed below, suggest an exceptional negative reaction to Hillary Clinton's public image, which may have ultimately affected the 2016 election. These results show that identifiable, if subtle cultural trends can foreshadow major social events.

Methods

Data Sources

USA naming data were obtained from the USA Social Security Administration website at <http://www.ssa.gov/oact/babynames>. These data are compiled from 100% of the Social Security Numbers issued between 1880 and, at the time of writing, 2015. For each year, the data provide the number of new SSNs issued to individuals with each recorded name. To protect the privacy of individuals with rare names, names with a count of 4 or less are omitted.

Names of First Ladies were obtained from https://en.wikipedia.org/wiki/List_of_First_Ladies_of_the_United_States. Electoral results of the 1992 presidential election were obtained from https://en.wikipedia.org/wiki/United_States_presidential_election,_1992.

Analysis of Naming Trends

The analysis of naming trends presented below focuses on two properties of name-popularity time series, which are defined as follows:

Trend changes. A "trend change" refers to how a popularity trajectory changes around a focal year. This quantity was used by Ghirlanda et al. (2014) to show that the popularity of a dog breed increases after the release of a movie featuring that breed. Here, I use it to investigate changes in the popularity of First Ladies' names around Presidential election years. Formally, the trend change is defined as

$$100 \times \frac{a_{\text{post}} - a_{\text{pre}}}{p} \quad (1)$$

where a_{post} and a_{pre} are the average yearly changes in popularity before and after the election, and p is average name popularity in the 9 years surrounding the election. The application of equation (1) is illustrated in Figure 1a.

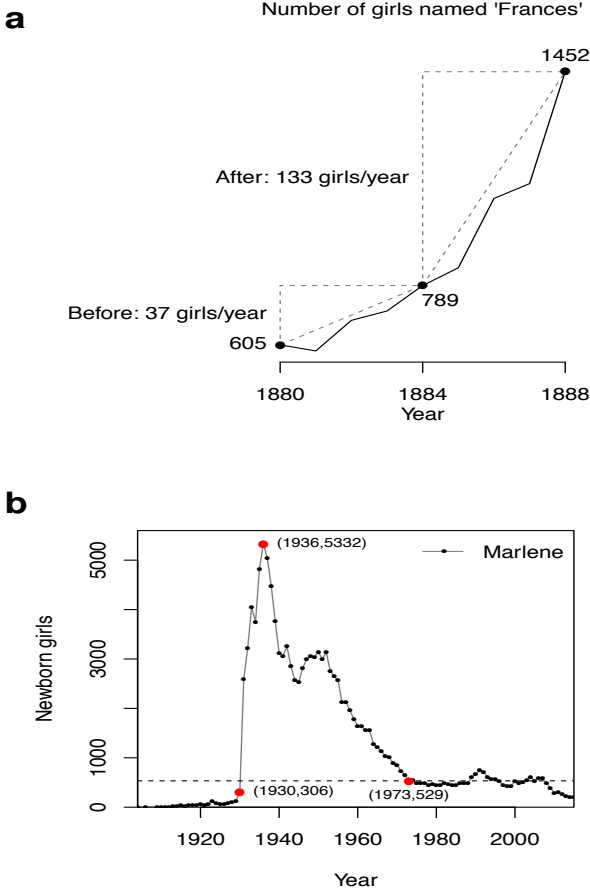


Figure 1a. Example of trend change in name data. The graph shows the popularity of the name ‘Frances’ around the time Frances Cleveland’s husband, Grover, was elected to his first presidential mandate. In the four years before the election, the number of girls named ‘Frances’ each year was increasing by 37 girls/year. In the four years after the election, the same average increase was 133. In the years between 1880 and 1888, an average of 870 girls were named ‘Frances’ each year. Thus, equation (1) quantifies the trend change as $100 \times \frac{133-37}{870}$. **b.** Calculation of name cycle parameters for the name ‘Marlene,’ according to the definitions in Table 1. The rise time is $1936-1930 = 6$ years, the fall time is $1973-1936 = 37$ years, and the cycle skew is $s = 6/37 \cong 0.16$. The dashed line indicates the level used to delimit the cycle, i.e., 10% of maximum popularity.

Equation (1) takes into account ongoing cultural trends, i.e., whether the name was already rising or falling in popularity before the election, and is comparable across names of different popularity because the trend is normalized according to name popularity (Ghirlanda et al., 2014). A positive (negative) value of equation (1) indicates that, after the election, the name was being given each year to more (fewer) newborns than would have been expected should the pre-election trend have continued unchanged.

Cycle skew. Cycle skew refers to an asymmetry that often occurs in name popularity trends. As recalled in the Introduction, many names exhibit “fashion cycles,” during which name popularity rises from a low level to a considerably higher one, and then gradually reverts to a low level. The time it takes to reach peak popularity (“rise time”) and the time it takes to revert to pre-peak popularity levels (“fall time”) are correlated (Berger and Le Mens, 2009; Acerbi et al., 2012). The fall time, however, is often longer than the rise time. I refer to a difference between rise and fall times as the “skew” of the name cycle. Formally, I define skew in Table 1 as the ratio s between rise and fall times. As it will be seen below, the average skew for name cycles is $s \approx 0.77$, with $s < 1$ for about 75% of cycles. Rise and fall times are calculated by defining a cycle as “starting” and “ending” when name frequency is at 10% of its peak value, although results are not sensitive to the precise value of this threshold (Berger and Le Mens, 2009). Figure 1b exemplifies the calculation of cycle skew.

Table 1. Definition of cycle skew s and related quantities

| Quantity | Symbol | Definition |
|------------|--------|---|
| Peak year | y_P | Year of maximum popularity |
| Frequency | $f(y)$ | Name frequency in year y |
| Rise year | y_R | Year $y < y_P$ closest to y_P such that $f(y) \leq 0.1f(y_P)$ |
| Fall year | y_F | Year $y > y_P$ closest to y_P such that $f(y) \leq 0.1f(y_P)$ |
| Rise time | t_R | $y_P - y_R$ |
| Fall time | t_F | $y_F - y_P$ |
| Cycle skew | s | t_R/t_F |

Statistical Analysis

In the following, I will ask whether the popularity cycles of the name ‘Hillary’ and its variant ‘Hilary’ differ from the popularity cycles of other names. Assume we are considering n names. Measuring a property of their popularity cycles results in n values, and we want to estimate the probability that the two values pertaining to ‘Hillary’ and ‘Hilary’ are in some sense “unusual.” Let h_1 and h_2 denote these values. The strictest definition of “unusual” is that h_1 and h_2 are the most extreme of the n values. However, even such an unusual arrangement could

be the product of chance. To evaluate whether the event “ h_1 and h_2 are the most extreme of n values” is exceptional or not, we must calculate its probability under the null hypothesis that h_1 and h_2 have been placed at random among the n values. Such a probability is:

$$P_n = \frac{2 \times 2 \times (n-2)!}{n!} \quad (2)$$

where $n!$ is the factorial function: $n! = 1 \times 2 \times \dots \times n$. The denominator of equation (2) is the number of possible arrangements of n items. The numerator is the number of arrangements in which two specific items, in our case h_1 and h_2 , have the two most extreme values. The latter is arrived at by noticing that there are 2×2 ways to arrange h_1 and h_2 to be most extreme: they can be either the two lowest or the two highest values. Moreover, in both cases either of h_1 or h_2 can be the largest value. In other words, if X denotes the remaining $n - 2$ values, we have the four valid arrangements: $h_1 h_2 X$, $h_2 h_1 X$, $X h_1 h_2$ and $X h_2 h_1$. Lastly, the factor $(n - 2)!$ at the numerator of equation (2) is the number of ways in which the remaining $n - 2$ items can be arranged.

There may be a concern that the above reasoning is incorrect because ‘Hillary’ and ‘Hilary’ have similar popularity cycles (Figure 2), hence the values h_1 and h_2 are not independent. This is true, but note that the probability that a single randomly selected value is the most extreme one among n is $2(n - 1)!/n!$, which is always smaller than the value in equation (2) for $n > 2$. Hence considering two values is always more compatible with the null hypothesis than considering one value, even if the two values are strictly correlated.

Equation (2) can be used to compare ‘Hillary’ and ‘Hilary’ to other names non-parametrically, i.e., without being concerned with estimating the probability distribution of the measured properties. For large n , however, the latter is also possible, as exemplified below.

Software

All analyses were conducted with R, version 3.3.1 (R Core Team, 2016). The gamma distribution in Figure 2 was fit with function `fitdist` in the `fitdistrplus` package, version 1.0-8 (Delignette-Muller and Dutang, 2015). Factorials of large numbers arising in equation (2) were calculated using $\frac{a!}{b!} = \exp[l(a) - l(b)]$ where $l(a)$ gives the logarithm of $a!$ and is implemented in R as the `lfactorial` function.

Results

Naming Trends for 'Hillary' and 'Hilary'

Figure 2a shows a sharp drop in female newborns named 'Hillary' or 'Hilary' occurring in the USA after 1992. The popularity of these names dropped by over 90% in just 5–7 years, after having increased for more than 20 years. As recalled in the Methods section, this is uncommon: in name popularity cycles, falls are typically slower than rises. Figure 2b illustrates this fact and shows that 'Hillary' and 'Hilary' are outliers in the joint distribution of rise and fall times (cyan circles are names of First Ladies other than Hillary Clinton). Indeed, Figure 2c shows that the popularity cycles of 'Hillary' and 'Hilary' are the two most skewed ones among the 630 names included in the analysis (the latter are all names that exhibit a popularity cycle among the 1500 most common female names). An estimate of how unlikely these extreme values are can be obtained based on equation (2), which yields $P_{630} = 10^{-5}$. A further estimate may be based on the fact that a gamma distribution fits well the distribution of cycle skew (black line in Figure 2c). According to the fitted gamma distribution, there is a probability of 7.5×10^{-4} to observe a cycle skew at least as extreme as that of 'Hillary.' The probability for 'Hilary' is 7.5×10^{-7} .

The above conclusions hold also when comparing 'Hillary' and 'Hilary' to specific sets of names, which may be deemed more appropriate comparison groups than the set of all names. For example, 'Hillary' and 'Hilary' exhibit the most skewed cycle among all names that peaked around 1992, and also among all names that took a similar amount of time to reach their peak. Indeed, if 'Hillary' and 'Hilary' are the most extreme of all names, they are also, necessarily, the most extreme of any smaller set. As a pertinent example, Figure 2c also shows the skew values for all First Ladies names (cyan circles). These are unremarkable, with the only exception of the name 'Lou' (the closest point to 'Hillary' in the figure). Lou Hoover, however, was first Lady in 1929–1933, while the popularity cycle giving rise to the data point in the graph occurred in the 1950's.

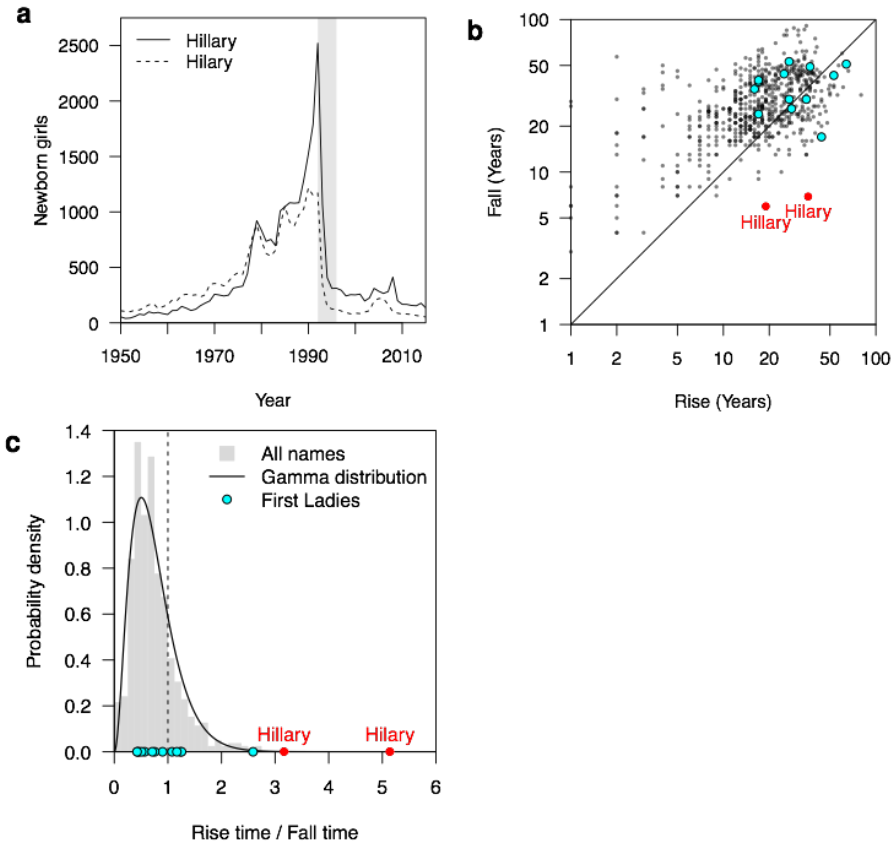


Figure 2a. Newborn girls given the names ‘Hillary’ and ‘Hilary’ between 1950 and 2015. The popularity of these names before 1950 is at or below the 1950 level. The shaded area indicates Bill Clinton’s first term as U.S. President. **b.** Relationship between rise and fall time of 630 names, i.e., all names among the 1500 most popular female names that underwent at least one complete popularity cycle between 1910 and 2015. ‘Hillary’ and ‘Hilary’ are highlighted. Cyan circles pertain to the names of First Ladies other than Hillary Clinton. See Methods and Table 1 for a definition of popularity cycle. The diagonal line indicates equal rise and fall times. Rise and fall times are correlated ($r = 0.41, df = 628, p < 10^{-15}$). In about 75% of cases (points above the diagonal), rises are faster than falls. **c.** Distribution of cycle skew values. The grey bars are a histogram of cycle skew values for the data plotted in the center panel. The dashed line indicates absence of skew. The solid line is a gamma distribution with shape parameter ≈ 3.12 and scale parameter ≈ 0.24 .

Naming Trends for First Ladies of the USA

The above analysis does not exclude that 'Hillary' and 'Hilary,' while exhibiting atypically skewed popularity cycles, may be in line with other First Ladies' names in terms of trend changes around election years. First Ladies' names, however, have typically negligible impact on naming trends, as shown in Figure 3. The negative changes for 'Hillary' and 'Hilary' are the two most extreme ones among 25, which according to equation (2) would occur by chance with a probability of $P_{25} = 0.007$.

The only other strongly negative trend change is the one for 'Rosalynn,' during the unpopular presidency of Rosalynn Carter's husband Jimmy. 'Rosalynn' is a very uncommon name (74/year at its highest point), so that relatively few naming decisions can yield a large impact (cf. peak values of 2521 and 1216 per year for 'Hillary' and 'Hilary,' respectively).

To gauge what a "strongly" negative or positive change may be, note that 'Jackie' and 'Jacqueline' change in opposite directions following the 1960 election, in which Jacqueline Kennedy's husband John was elected. Because these two names relate to the same person, the fact that the observed change is in opposite directions suggests that trend changes between 20% and 20% (approximately) can be expected from factors that are unrelated to presidential elections. Only trends for 'Hillary,' 'Hilary,' and 'Rosalynn' fall outside of this interval. (Detailed inspection of name popularity data reveals also that 'Rosalynn,' 'Jackie,' and 'Jacqueline' all spike for one year following the respective elections, with the latter two also spiking following President Kennedy's 1962 assassination. Trends for 'Hillary' and 'Hilary,' in contrast, show an immediate post-electoral drop. See Figure 3).

'Hillary' May be Disliked Regardless of Political Opinion

Data also suggest that, surprisingly, 'Hillary' became less popular among Democratic and Republican voters alike. Figure 4 displays the trend change for 'Hillary' in each of the 50 United States, following the 1992 election. Apart from Wyoming, where the name was absent, 'Hillary' decreased in *all* states. These trend changes correlate neither with the percentage of Republican votes in each state (two-tailed Pearson's correlation test, $r = 0.11$, d.f. = 47, $p = 0.45$), nor with the difference between Republican and Democratic votes ($r = 0.07$, d.f. = 47, $p = 0.65$). 'Hilary' shows a weak effect of electoral results, with a larger drop in states with more Republican votes ($r = 0.35$, d.f. = 47, $p = 0.01$), but no correlation between the size of the drop and the difference between Republican and Democratic votes ($r = 0.23$, d.f. = 47, $p = 0.11$).

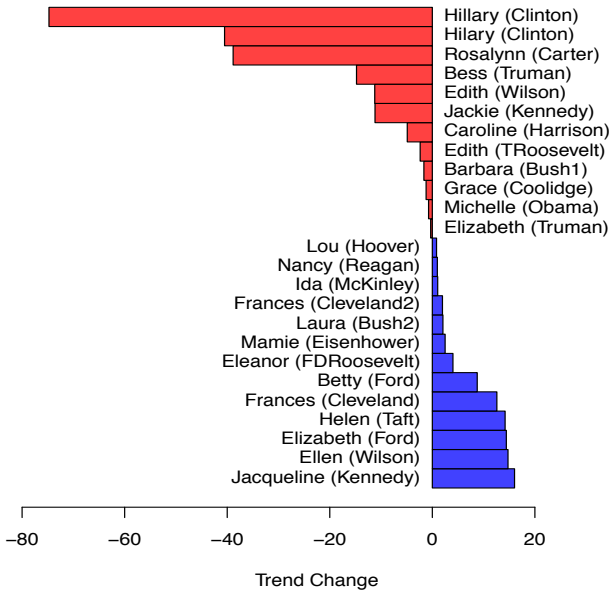


Figure 3. Trend changes in the popularity of First Ladies’ names, calculated according to equation (1). Presidents’ last names are indicated alongside First Lady’s names. Bush1 refers to George H. W. Bush; Bush2 refers to George W. Bush. Grover Cleveland was President in two non-consecutive terms, indicated here as Cleveland and Cleveland2. Consecutive presidential terms have been analyzed as a single term.

Discussion

I presented several lines of evidence supporting the conclusion that the sudden drop in popularity of names ‘Hillary’ and ‘Hilary’ in the years following 1992 is exceptional. Names, in general, rarely decrease so abruptly. Moreover, First Ladies’ names have typically a negligible effect on naming trends. Although my analysis was prompted by the outcome of the 2016 presidential election, results presented above could have been derived entirely from data available before the election.

In the Introduction, I hinted to a tentative explanation of the abrupt drop in popularity of ‘Hillary’ and ‘Hilary’ as due to Hillary Clinton acquiring a negative public image as a consequence of extensive media exposure as First Lady. While this interpretation requires careful evaluation, it is plausible, in general, that naming trends are influenced by exposure effects of which we are not necessarily aware (Berger et al., 2012). In particular, it can be conjectured that attitudes

toward public figures who bear a certain name may be important determinants of that name's popularity. It is relatively easy to find suggestive evidence that greatly admired figures can boost name popularity. The number of boys named Elvis, for example, spiked dramatically after the release, in 1956, of Elvis Presley's first single (*Heartbreak Hotel*), as well as after his death in 1977. Similarly, the number of girls named Marlene skyrocketed after 1930, the year Marlene Dietrich starred in the successful movies *The Blue Angel* and *Morocco* (see Figure 1b). Apart from the case at hand, however, I have not found convincing examples that public figures can influence name popularity negatively. Further research into identifying positive and negative influences on naming trends promises to be highly rewarding to understand the forces that shape popular culture, and, potentially, important events such as US presidential elections.

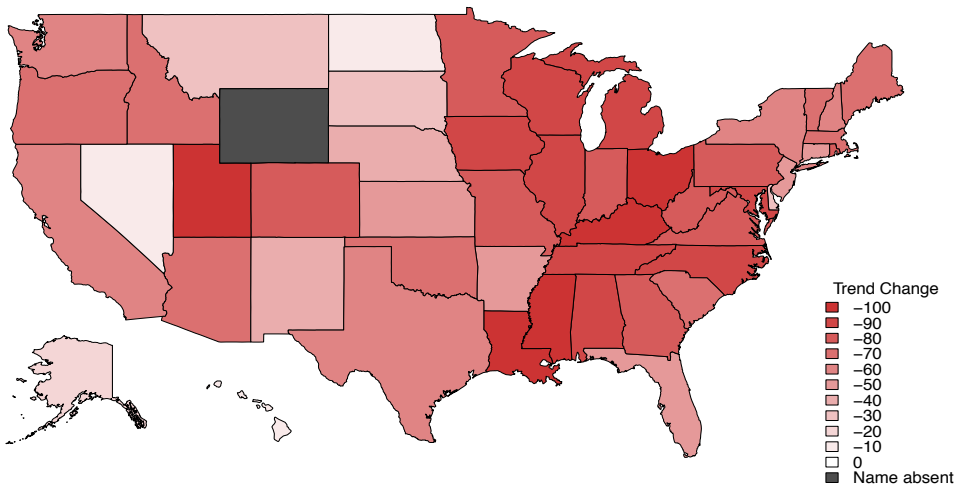


Figure 4. Trend changes in the popularity of 'Hillary' around 1992, across the United States. Trend changes calculated separately for each state using equation (1). The name 'Hillary' was not recorded in Wyoming (gray).

Acknowledgment

I am grateful to Magnus Enquist and an anonymous reviewer for comments on a previous version of the manuscript. This research was partly supported by grant 2015-0005 from the Knut and Alice Wallenberg Foundation.

References

- Acerbi, A., S. Ghirlanda, and M. Enquist. 2012. "The logic of fashion cycles." *PLoS ONE* 7(3): e32541. doi: 10.1371/journal.pone.0032541.
- Alter, A. 2007. "The baby-name business." *The Wall Street Journal*. June 22.
- Ball, M. 2016. "Why Hillary Clinton lost." *The Atlantic*. November 15.
- Berger, J., E. T. Bradlow, A. Braunstein, and Y. Zhang. 2012. "From Karen to Katie: Using Baby Names to Understand Cultural Evolution." *Psychological Science*. <http://pss.sagepubs.com>. doi: 10.1177/0956797612443371.
- Berger, J. and G. Le Mens. 2009. "How adoption speed affects the abandonment of cultural tastes." *PNAS* 106(20): 8146–8150. doi: 10.1073/pnas.0812647106.
- Delignette-Muller, M. and C. Dutang. 2015. "fitdistrplus: An r package for fitting distributions." *Journal of Statistical Software* 64(1): 1–34. doi: 10.18637/jss.v064.i04.
- Edwards, R. and C. Caballero. 2008. "What's in a name? An exploration of the significance of personal naming of 'mixed' children for parents from different racial, ethnic and faith backgrounds." *The Sociological Review* 56(1): 39–60. doi: 10.1111/j.1467-954x.2008.00776.x.
- Ghirlanda, S., A. Acerbi, and H. Herzog. 2014. "Dog movie stars and dog breed popularity: A case study in media influence on choice." *PLoS ONE* 9(9): e106565. doi: 10.1371/journal.pone.0106565.
- Gureckis, T. M. and R. L. Goldstone. 2009. "How you named your child: Understanding the relationship between individual decision making and collective outcomes." *Topics in Cognitive Science* 1(4): 651–674. doi: 10.1111/j.1756-8765.2009.01046.x.
- Hahn, M. and R. A. Bentley. 2003. "Drift as a mechanism for cultural change: an example from baby names." *Proceedings Royal Society London B* 270(Suppl_1): S120– S123. doi: 10.1098/rsbl.2003.0045.
- Klein, G. 2016. "Why did Hillary lose the election?" *Psychology Today*. December 6.
- Long, R. 2016. "Why did Hillary Clinton lose?" *The Chicago Tribune*. November 14.
- R Core Team. 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.