

Naming Boys after U.S. Presidents in 20th Century

K. KULAKOWSKI^{a,*}, P. KULCZYCKI^{a,b}, K. MISZTAL^a, A. DYDEJCZYK^a, P. GRONEK^a
AND M.J. KRAWCZYK^a

^aFaculty of Physics and Applied Computer Science, AGH University of Science and Technology,
al. A. Mickiewicza 30, 30-059 Cracow, Poland

^bSystems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland

This paper deals with the popularity of given names in the United States, for the period 1885–2009. Based on the data obtained from the website of U.S. Social Security Administration, it was demonstrated that the fashion of naming babies after the incumbent American president passed away in the '60s. At the same time, however, examples were given, mainly concerning celebrities, after whom babies are still named. The above theses were strengthened with the aid of quantitative data analysis by constructing an index dedicated to the specifics of the task under investigation. The obtained results were discussed in the terms of the rally effect and of the Simmel theory of fashion.

DOI: [10.12693/APhysPolA.129.1038](https://doi.org/10.12693/APhysPolA.129.1038)

PACS/topics: 89.20.-a, 89.65.Cd

1. Introduction

In the Bogardus scale of social acceptance [1], the highest level of approval is to accept somebody as a close relative by marriage. Symbolically, a similar level of acceptance is to name a newborn child after the person we like. Tracing the fashion of names, one can learn about emotional reactions to popular personalities in a social scale. Contrary to previous research [2–5], which dealt with statistical averages, here we are going to concentrate on case studies. Namely, we intend to trace the popularity of names of U.S. presidents in the 20th century and to compare the results with those on the names of selected celebrities. We will show that this specific dataset provides results complementary to standard poll data on approval of politicians [6]. Our main goal is to show that this level of social acceptance of U.S. presidents vanished in the 1960's. Yet this thesis cannot be maintained if applied to celebrities.

The data is taken from the website of the U.S. Social Security Administration [7]. Its discussion comprises the next section. Section 3 presents a quantitative data analysis: an index is proposed, designed to measure the jump in popularity of a given name due to an appearance of a person with this name. For a president, the date of appearance is the year of the campaign and/or the first year at the office; for a celebrity, the career is often triggered by a popular movie or even a song. In Sect. 4, the results are discussed in terms of the *rally 'round the flag* effect [8] and the Simmel theory of fashion [9].

2. Data on the names of presidents

Absolute numbers of babies' given names in the USA in 1885–2009 are taken from [7]. There, names with less

than five entries are omitted. Although the data in percent is a good indication of the popularity of names, it is sensitive to the popularity of other names at the same time. The absolute numbers provide a measure of the number of hard-core believers, and this is taken into account in the obtained conclusions. The data in percentages is also monitored, in the event when it is inconsistent with the data in absolute numbers.

An interface is prepared, available at [10], with the range (m_1, m_2) as an input. As a result, we get a list of names characterized by a maximum value m , where $m_1 < m < m_2$. The value m of a name is the maximum number of babies with this name within one calendar year, between 1885 and 2009. When a name from the list is clicked, one can get the plot of the number of namesakes, again in absolute values, within the whole period. Alternatively, the data is accessible as numbers. The interface possesses other functionalities, which are not used here.

For the first presidents in the considered period, (20th) James A. Garfield (1881) and (21st) Chester A. Arthur (1881–1885), the effect is hardly visible. James is always a popular name in U.S. The number of Chesters increased from 168 in 1880, when C.A. Arthur served already as vice-president, to about 300 in subsequent years, and remained at this level with some fluctuations. The first president to be considered for the purpose of the analysis described here is (22nd and 24th)[†] Grover Cleveland (1885–1889) and (1893–1897). Figure 1 leaves no doubt, that his two elections had twice the influence on popularity of given names in the USA. For (23rd) Benjamin Harrison (1889–1893), there is only a two-point

*corresponding author; e-mail: kulakowski@fis.agh.edu.pl

[†]The twenty-second and twenty-fourth (as he was elected on two separate occasions) president of the USA. This notation will be used throughout the text.

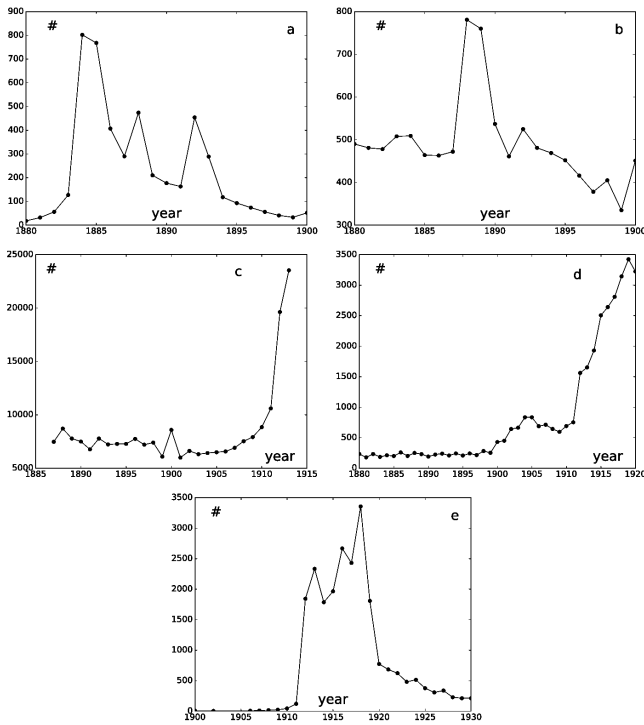


Fig. 1. The popularity of given names of: (a) Grover Cleveland (1885–1889) and (1893–1897), (b) Benjamin Harrison (1889–1893), (c) William McKinley (1897–1901) and William Howard Taft (1909–1913), (d) Theodore Roosevelt (1901–1909), (e) Woodrow Wilson (1913–1921).

peak in the data (Fig. 1b) for this name, otherwise popular in this country. For (25th) William McKinley (1897–1901), the effect is hardly visible in the applied scale (Fig. 1c); yet the one-point peak in 1900, perhaps related to his successful second term campaign, exceeds neighboring points by 2500 namesakes; the data in percent does not confirm the effect. Next is (26th) Theodore Roosevelt (1901–1909); the peak of 400 at his term seems to be small when compared with the subsequent fly-off (Fig. 1d), which could be connected to his post-term activities. Strangely enough, the data in percentages shows a large peak of popularity only at 1904. The case of (27th) William Howard Taft (1909–1913) (Fig. 1c again) should be treated with special care, as the related increase should be at least partially attributed to the wave of immigrants [11], together with Robert, John, Mary and James. While the time dependences of the absolute numbers related to these names reveal almost the same pattern [3], the related percentages do not show any increase [12].

In the case of (28th) Woodrow Wilson (1913–1921) (Fig. 1e) as well as for (29th) Warren G. Harding (1921–1923) (Fig. 1f), (30th) Calvin Coolidge (1923–1929) (Fig. 2a), (31st) Herbert Hoover (1929–1933) (Fig. 2b) and (32nd) Franklin D. Roosevelt (1933–1945) (Fig. 2c), the data is explicit. However, two terms of (33rd) Harry S. Truman (1945–1953), although clearly visible

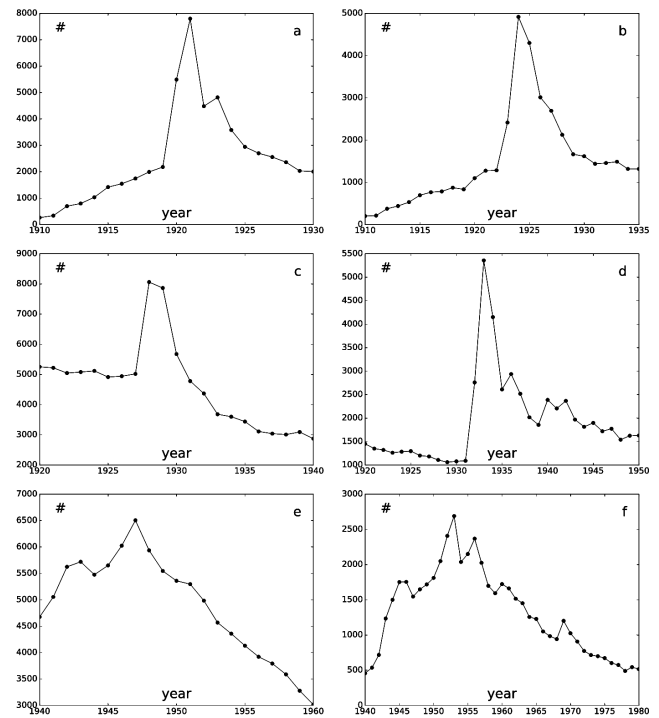


Fig. 2. The popularity of given names of: (a) Warren G. Harding (1921–1923), (b) Calvin Coolidge (1923–1929), (c) Herbert Hoover (1929–1933), (d) Franklin D. Roosevelt (1933–1945), (e) Harry S. Truman (1945–1953), and (f) Dwight D. Eisenhower (1953–1961).

in (Fig. 2d), are marked only weakly in the data in percent. For (34th) Dwight D. Eisenhower (1953–1961) (Fig. 3a), the maximum is again clear, with some precursor in the 40s', perhaps related to his military function during WWII. The term of (35th) John F. Kennedy (1961–1963), even with its dramatic end, is hardly visible in the data (Fig. 3b) because the name was so popular anyway. Yet, both increases of the popularity of the name in 1961 and 1964 reached almost 4000 each; the second peak is visible also in the data in percent. Finally, the case of (36th) Lyndon B. Johnson (1963–1969) (Fig. 3d) is again visible, together with a precursor in 1960 when he unsuccessfully competed with JFK for the Democratic nomination. Actually, the contribution of Lyndon B. Johnson to the number of Lyndons appears to be extremely high, which can be related to his social program of the Great Society.

As shown in Fig. 3e, there is no mark of any increase of popularity of Richard even at the beginning of the term of (37th) Richard M. Nixon (1969–1974). Instead of being an exception, this result appears to be rather a rule from now on. The data related to (38th) Gerald R. Ford (1974–1977) (Fig. 3f) only shows a continuous fall. An impact of the peaceful personality of (39th) James Carter (1977–1981) (Fig. 4a) is hardly visible in the data. Two terms of leadership of (40th) Ronald Reagan (1981–1989) (Fig. 4b) elapsed with no reaction of American parents. It is hard to distinguish an influence

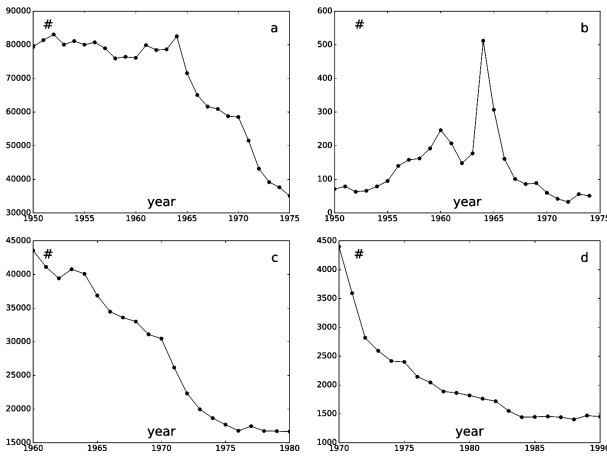


Fig. 3. The popularity of given names of: (a) John F. Kennedy (1961–1963), (b) Lyndon B. Johnson (1963–1969), (c) Richard M. Nixon (1969–1974), and (d) Gerald R. Ford (1974–1977).

of (41st) George H.W. Bush (1989–1993) (Fig. 4c) from random fluctuations, except two points in 1990; in the data in percent, nothing can be seen. Neither in the case of (42nd) William J. Clinton (1993–2001) (Fig. 4d), nor for the entirely different personality of (43rd) George W. Bush (2001–2009) (Fig. 4c again) can any mark of their influence be seen.

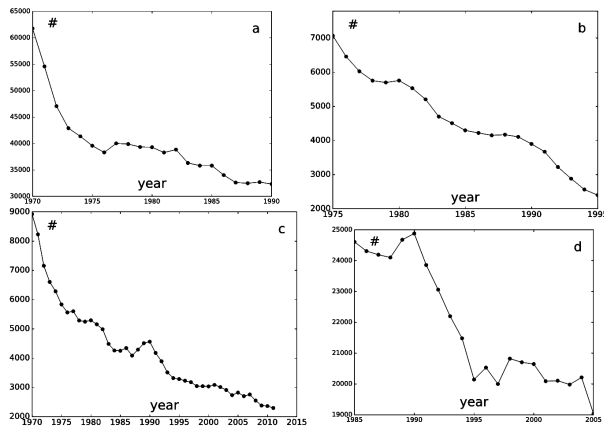


Fig. 4. The popularity of given names of: (a) James Carter (1977–1981), (b) Ronald Reagan (1981–1989), (c) George H.W. Bush (1989–1993) and George W. Bush (2001–2009), and (d) William J. Clinton (1993–2001).

As we remarked above several times, the effect is much more visible if the president’s name is less popular. Therefore for popular names, as John, George or William, our results should be taken with reservations. Making one more step into the 21st century, we have inspected also the case of Barack Obama (2009–present) — a very rare named indeed. The popularity of this name increased from the threshold 5 in 2007 to 69 in 2009, then decreased again, remaining at extremely low levels

throughout. The case of Obama cannot be, then, treated as a support of any statistical statement.

3. Quantitative data analysis

In order to investigate more accurately the problem of a newly elected president’s influence on the naming of babies in the USA, investigated in this paper, a formal mathematical apparatus was worked out, enabling exact numerical results to be obtained. Due to the specific nature of the subject, it is not possible to apply *explicitly* the standard methods used for outlier detection [13] or unit step [14].

Denote the inaugural year of a president of the United States as y_0 and — naturally — the following years as

$$y_1, y_2, y_3, \dots \quad (1)$$

and those preceding a

$$\dots, y_{-3}, y_{-2}, y_{-1}, \quad (2)$$

respectively. As can be seen from the graphs of popularity of particular presidents’ names for babies, the local maximum occurs one year before inauguration (which is obviously the influence of a spectacular election campaign) or — somewhat rarer — in the year of taking office or occasionally the year immediately following[‡]. The maximum popularity in these years can be — therefore — used to ascertain the influence a new president’s name has on baby naming trends in the USA. So, let us denote

$$y = \max\{y_{-1}, y_0, y_1\}. \quad (3)$$

This value will be compared with parameters characterizing the popularity of given names in a period of time extended by the four years (the length of a presidential term) preceding inauguration, therefore in the years

$$Y = \{y_{-5}, y_{-4}, y_{-3}, y_{-2}, y_{-1}, y_0, y_1\}. \quad (4)$$

In order to characterize the above, parameters will be used, the concept of which is based on the notion of quantiles[§]; they are more robust to atypical values than common moments, which is highly valuable in the task under investigation. In particular, consider the quantiles of orders 0.5 (median) $q_{0.5}$, as well as 0.75 (the third quartile) $q_{0.75}$, and 0.25 (the first quartile) $q_{0.25}$. The median $q_{0.5}$ represents a central point of the set Y , while the quartile deviation, that is the difference between the third and the first quartiles $q_{0.75} - q_{0.25}$, signifies dispersion around it. The value of quantile order $r \in (0, 1)$ will be guessed by using the concept of a positional estimator of the second degree [15], given by

[‡]This “year immediately following” also helps to unify the investigated methodology and to include elements other than presidents, e.g. celebrities, for which the “year previous” has no election campaign, thereby moving the phenomenon of a name’s popularity forward one year relatively.

[§]Because of the small size of set (4), it is difficult to treat them *explicitly* as quantile estimators.

$$\hat{q}_r = (0.5 + i - mr)z_i + (0.5 - i + mr)z_{i+1}, \quad (5)$$

where m denotes sample size, $i = [mr + 0.5]$, while $[a]$ means the integral part of the number $a \in R$, whereas z_i represents the i -th greatest value of the set Y after its ordering (another usable concept is given in the paper [16]). In practice, without reducing generality one can assume $\hat{q}_{0.75} - \hat{q}_{0.25} \neq 0$. So that $\hat{q}_{0.75} - \hat{q}_{0.25} = 0$, in the case $m = 7$ considered here, it would have to be thus that $z_2 = z_3 = z_4 = z_5 = z_6$, which is not realistic in practice. What is more, because the inequality $\hat{q}_{0.75} - \hat{q}_{0.25} \geq 0$ is always true, we are finally left with

$$\hat{q}_{0.75} - \hat{q}_{0.25} > 0. \quad (6)$$

As a measure of popularity of a newly-elected president's name, the number $c \in R$, defined as:

$$c = \frac{y - \hat{q}_{0.5}}{\hat{q}_{0.75} - \hat{q}_{0.25}} \quad (7)$$

will be used. It indicates how far the number y , given by formula (3), is removed from the median of the set Y on the scale of quartile deviation. The greater the number c , the greater the relative increase in influence of a newly-elected president on naming trends in the USA. Values close to zero indicate an insignificant or non-existent impact on baby names. Negative numbers most often show a negative influence on the choice of that name. The above concept is drawn from outlier detection procedures [17], adapted to the specifics of the task under investigation here.

It is also worth noting that from a methodological point of view, the results obtained for very small values of elements of the set Y must be treated as statistically unreliable. Such a case indicates a generally faint popularity of a name and then, in real terms, even a small — insignificant on the national scale — number of additional times a name is given after a president's election will constitute a proportionally substantial jump and in consequence a big value for the parameter c . Based on observed graphs of given names for particular presidents we can assume that this effect occurs when

$$\hat{q}_{0.5} < 100. \quad (8)$$

If the above condition is fulfilled, then — according to the language of the association rules [18] — confidence in the value of the parameter c would be connected with unacceptably low support.

The methodology presented above was applied for time analyses of changes in influence of the names of presidents of the USA on their popularity in that country. For this purpose, the data in base [7] was used. It enabled research going from president (22nd) Grover Cleveland (1885–89). With respect to existing ambiguities, it should be clarified that names of particular presidents were taken from information listed on the official website of the White House [19].

As noted previously, statistical inference for the incumbent president (44th) Barack Obama (2009–present) is statistically unreliable ($\hat{q}_{0.5} = 5$). The Arabic name Barack (his family often referred to him using the name Barry and he was even registered to school as such) was

practically unheard of in the USA — the first time 5 babies were given the name ¶ was two years before Obama's coming into office. In this instance, the name's registration 52 times in 2008, 69 in 2009 and 28 in 2010 cannot in any way attest to social phenomena in a country as large as the USA.

Figure 5 shows values of the parameter c , defined by formula (7), calculated for the presidents from (22nd) Grover Cleveland (1885–89) to (43rd) George W. Bush (2001–09).

In the first period, from president (22nd) Grover Cleveland (1885–89) to (32nd) Franklin D. Roosevelt (1933–45) a lateral trend for the values above $c = 1$ can be seen, clearly showing a steady influence of a president's name on its use in the USA. It is worth noting that a local minimum occurs for (24th) Grover Cleveland (1893–97), who was elected for the second time following a four-year intermission, and so patriotic fervor was obviously slightly muted. A sudden increase in the value of the parameter c , from 1.6 to 2.9, for president (30th) Calvin Coolidge (1923–29) shows that from a statistical point of view this should be treated as atypical (an outlier); it also has fundamental relevance as a manifestation of patriotism following the sudden death from a heart attack of president (29th) Warren G. Harding (1921–23), whom Calvin Coolidge — as vice-president — succeeded.

This trend was broken by (33rd) Harry S. Truman (1945–53). Despite a local growth for (34th) Dwight D. Eisenhower (1953–61) and later for (36th) Lyndon B. Johnson (1963–69), the parameter c value approached zero for (37th) Richard M. Nixon (1969–74). It is worth noting once more that the exceptionally violent increase — in the second case — from 0.4 to 4.6 should be taken as an outlier, fundamentally arising from national grief following the tragic loss of (35th) John F. Kennedy (1961–63). Moreover, the fact that the aforementioned falling trend was started by (33rd) Harry S. Truman (1945–53), who also in fact assumed office after the sudden death of (32nd) Franklin D. Roosevelt (1933–45), proves the strong fundamental reasons for breaking the previous lateral trend.

For the last group of considered presidents, from (37th) Richard M. Nixon (1969–74) to (43rd) George W. Bush (2001–09), one can see oscillations around the value $c = 0$, which ultimately points to the present lack of influence of a president's election on babies' names in the USA. The sudden temporary increase in the parameter c value, from 0.0 to 1.3, for (41st) George H.W. Bush (1989–93) can be treated as an outlier (the so-called exception proving the rule).

It is worth adding that substituting the names as given on the White House website [19] with the first names as on the birth cert (e.g. "John" for (30th) Calvin Coolidge

¶The database does not include cases with less than 5 instances of a name in one year.

(1923–29)) or nicknames (e.g. “Jimmy” for (39th) James Carter (1993–2001)) would not generally change results for the investigated analysis (as examples $c = 1.3$ for John-Coolidge or $c = 0.2$ for Jimmy-Carter).

In summary, from a quantitative point of view, based on the presented statistical inference, one can see that a

clearly formed long-term lateral trend above the level 1.0 — proving strong and stable influence of a president’s name on its popularity in the USA — was replaced in 1945 with a declining trend, which after 1969 became lateral at the level 0, indicating a lack of such effect.

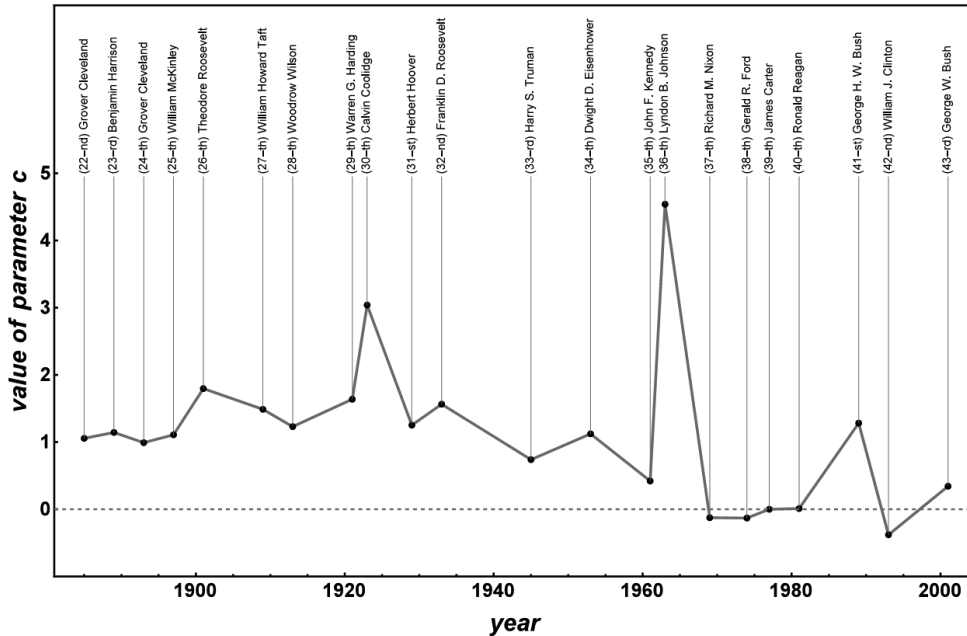


Fig. 5. Values of the parameter c for the presidents.

It is also worth analyzing whether the above scenario concerns only presidents, or also other elements which may have an influence on naming trends. Firstly there is the difficult problem of defining the character of the subject matter and the moment of occurrence: among models perceived universally in American culture, there is no given, clearly identified group (like presidents) or

other phenomenon, additionally with a strict moment of “activation” (similar to presidential elections). One can however select celebrities or songs with names in the titles, for which the analysis presented in this paper shows a contemporary influence on names given in the USA. As an example, the following were obtained:

Shirley	Shirley Temple, film “Stand Up and Cheer”	1934	$c = 4.3$
Linda	song “Linda”, Buddy Clark with Ray Noble’s Orchestra	1946	$c = 3.8$
Jacqueline	Jacqueline Kennedy, husband’s inauguration	1961	$c = 5.0$
Michelle	song “Michelle”, The Beatles	1965	$c = 3.4$
Elvis	Elvis Presley, death	1977	$c = 3.5$
Dustin	Dustin Hoffmann, film “Kramer vs. Kramer”	1979	$c = 2.5$
Whitney	Whitney Houston, album “Whitney Houston”	1985	$c = 3.9$
Selena	Selena Quantanilla-Perez, death	1994	$c = 6.4$
Leonardo	Leonardo di Caprio, film “Titanic”	1997	$c = 5.1$
Britney	Britney Spears, album “... Baby One More Time”	1999	$c = 2.0$
Malia	Malia Obama, father’s inauguration	2009	$c = 3.4$

It is interesting that, in the last case $\hat{q}_{0.5} = 782$, and so condition (8) was not fulfilled here, as in the case of her father, the president himself. For the name Dustin, a jump in the year 1967, relating to the film “The Graduate” fulfilled condition (8), as $\hat{q}_{0.5} = 52$, and in consequence was omitted. Due to the arbitrary selection of celebrities and songs, it should not be surprising that such calculated values of the parameter c are significantly greater than in the case of presidents.

The above examples — owing to the arbitrary nature of the selection — definitely do not point to the existence of an identifiable tendency, but do indicate the influence of celebrities’ names, or arising phenomena associated with specific names, still exists, while previous investigations clearly show that this is no longer the case for presidents.

4. Discussion and summary

According to the concept named *rally ’round the flag* [8], the popularity of a leader increases when the country faces a crisis in external relations. If the idea is valid, we should be able to detect the changes of popularity of a head of state in the years of military conflicts. Yet, the data does not confirm this. The short war with Spain in 1898 is not visible in the popularity of the name William. The popularity of Woodrow increased in 1917, when the States entered WWI, but these variations are weaker than in preceding years, when Woodrow Wilson entered his second term with the slogan “he kept us out of war”. Similarly, the height of popularity of the name of Franklin happens at the time of the New Deal and legalized beer (1933), and not 1941, when the United States entered WWII. The consequences of the Cuban Missile Crisis in 1962, expected in the data in 1962–63, are smaller than the changes in popularity for the name John in 1961 and, posthumously, in 1964. Neither “Desert Shield”, nor “Desert Storm” inverted the decrease of frequency of naming boys George in 1991. The same can be stated for the terrorist attack on the Twin Towers, although the approval rating trend of George W. Bush showed a jump from 55 to 90 percent at the end of 2001 (according to Gallup [6]). Apparently, these two kinds of data are related to different phenomena in American society.

The Simmel theory of fashion [9] is more general and less precise, as it deals with symbols of status. This could be of different origin; classically it can be assigned to “property, prestige or power” [20]. Being a famous actor seems to evade this classification, being a president of the United States does not. In the latter case, a baby named “after our President” can be a fashion, related to a desired prestige, especially for immigrants. On the other hand, to name a baby “after my favorite actor” seems to be more of a family rather than a social choice, and its collective character comes from an influence of media rather than of friends or neighbors. If this supposition is to be accepted, the jumps in popularity should be larger if a personality represented by an actor is more suitable to be “a member of family”, the highest level of acceptance in the Bogardus scale.

This kind of evaluation is subjective, but — when attempted — it produces results which are consistent with the analysis performed on politicians. Namely, the “family” character became most accepted when the fashion of “strong men” passed out. Going beyond the persons and phenomena dealt with in Sect. 2, the position of Gary Cooper in his “boom”, slowly built since 1930, reached 40 thousand in 1952, when he restored law as a retiring sheriff in “High Noon”. Less than thirty years later, Clint Eastwood’s top record was only 1,250. In the same time, in 1980, Dustin Hoffman jumped in our statistics by 3,500 in one year for being a lonely father. We may wonder how Marlene Dietrich got 5,000 namesakes in a few years after “The Blue Angel”, where she played *la femme fatale*. Perhaps the solution is that in the ’30s, she was only a requisite in hands of Fate. In the case of Marilyn 20 years later, this excuse was not adequate in the eyes of American parents.

Summarizing, the data of American names in 1880–2011 are analyzed. The designed interface is publicly available [10] and it can be used for modeling and simulations — obviously, citations are welcome. We have shown that the fashion of naming babies after the actual American president passed away in 60’s of XX century, while the fashion of naming babies after a favourite celebrity has remained. These results provoke a conclusion that the society of the United States is politically more mature now than 60 years ago, because their images of politicians are more realistic. Yet, their favourite figures are equally often taken from the world of fantasy. We hope that this research can contribute in the promotion of applications of data analysis in historical case studies.

References

- [1] *The Concise Oxford Dictionary of Sociology*, Oxford UP, Oxford 1994.
- [2] Wentian Li, *Complexity* **18**, 44 (2012).
- [3] M.J. Krawczyk, A. Dydejczyk, K. Kułakowski, *Physica A* **395**, 384 (2014).
- [4] Ning Xi, Zi-Ke Zhang, Yi-Cheng Zhang, Zehui Gee, Li Shec, Kui Zhang, *Physica A* **406**, 139 (2014).
- [5] P. Barucca, J. Rocchi, E. Marinari, G. Parisi, F. Ricci-Tersenghi, *Proc. Natl. Acad. Sci.* **112**, 7943 (2015).
- [6] www.gallup.com/poll/116677/presidential-approval-ratings-gallup-historicalstatistics-trends.aspx.
- [7] The Official Website of the U.S. Social Security Administration (www.ssa.gov).
- [8] J.T. Chatagnier, *J. Peace Res.* **49**, 631 (2012).
- [9] G. Simmel, *Int. Quarterly* **10**, 13 (1904).
- [10] <http://newton.fis.agh.edu.pl/names/>.
- [11] <http://www.immigrationeis.org/about-ieis/us-immigration-history>.
- [12] <http://www.behindthename.com/top/name/william/unitedstates>.
- [13] V.J. Hodge, J. Austin, *Artif. Intellig. Rev.* **22**, 85 (2004).

- [14] M. Basseville, I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J. 1993.
- [15] R. Parrish, *Biometrics* **46**, 247 (1990).
- [16] P. Kulczycki, A.L. Dawidowicz, *Universit. Jagellon. Acta Math.* **XXXVII**, 325 (1999).
- [17] D.T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley, 2005.
- [18] C. Zhang, S. Zhang, *Association Rule Mining: Models and Algorithms*, Springer-Verlag, Hoboken, N.J. 2002.
- [19] <https://www.whitehouse.gov/1600/Presidents>.
- [20] M. Weber, in: *From Max Weber: Essays in Sociology*, Eds. H.H. Gerth, C. Wright Mills, Oxford UP, Oxford 1946, p. 181.