

Imiona dzieci – symulacje i fakty

Krzysztof Kułakowski

Antoni Dydejczyk

Piotr Groniek

Małgorzata Krawczyk

Piotr Kulczycki

Krzysztof Misztal

Mateusz Pomorski

oraz

Jarosław Kwapien

Marcel Ausloos



plan

- Motywacja
- Źródło danych: USA
- Moda jako proces – efekt Simmla
- Moda na mapie – struktura sieci korelacji
- Prawo Zipfa dla imion
- Prezydenci i inni – wskaźnik popularności
- Interpretacje

Motywacja – kult jednostki



Totmes III (Horus),
1481 -1425 BC



Boski August,
63 BC –14 AD



Car Symeon I, 864-927



„Przez dwa lata cały świat oddawał niemal boską cześć zwycięzcy spod Marny. Jego bagaże dosłownie ugięły się pod ciężarem pudełek, paczek i listów, które przesyłali mu nieznani ludzie w gorączkowych dowodach podziwu. Myślę, że poza generałem Joffre żaden dowódca czasu wojny nie miał okazji doznać porównywalnej chwały.”

[W. Lippmann, Public Opinion, 1921]



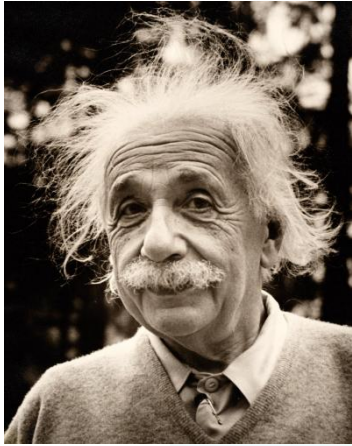
General Joffre,
Commander-in-Chief of French Army
BY UNDERWOOD & UNDERWOOD, N.Y.

Joseph Joffre, 1852-1931



Kult jednostki – jak zanika? w jakich warunkach?

Motywacja Cd: lista pomysłów na życie



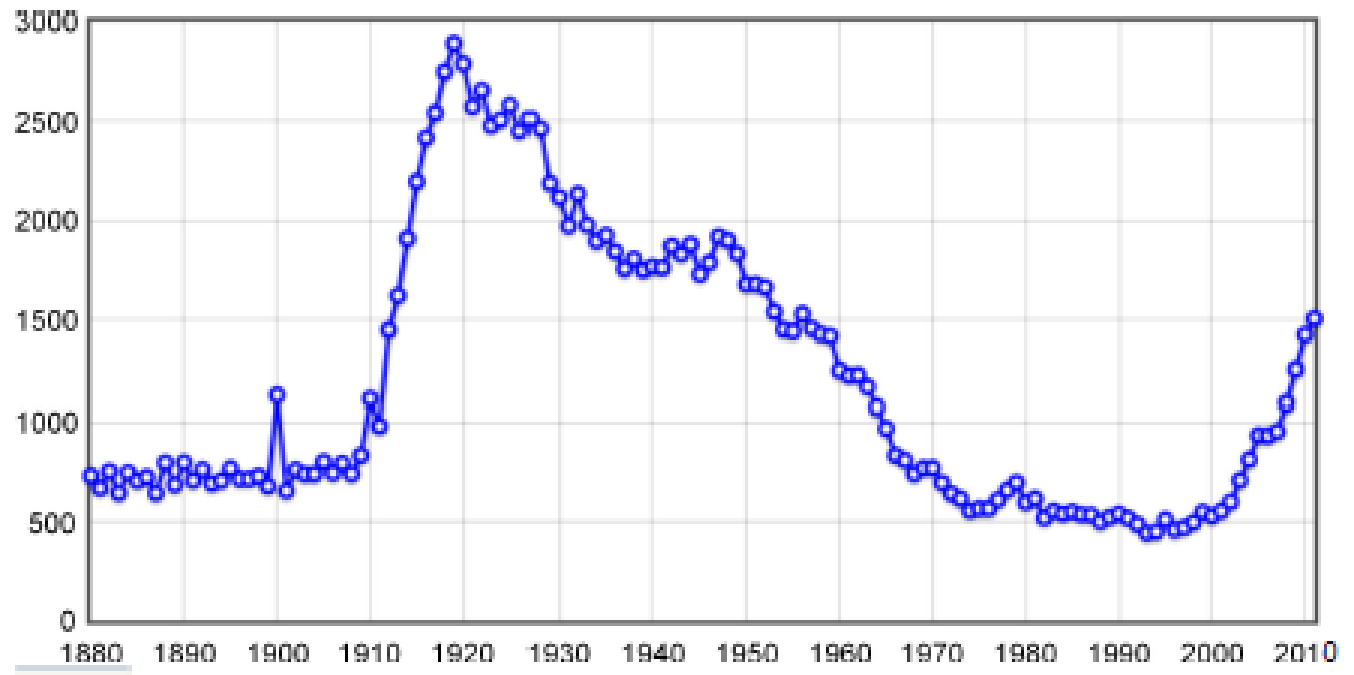
Źródło danych

Belgia: ilość osób o danym imieniu w Brukseli, Flandrii i Walonii, w rozbiciu na grupy wiekowe : <18, 18-65, >65

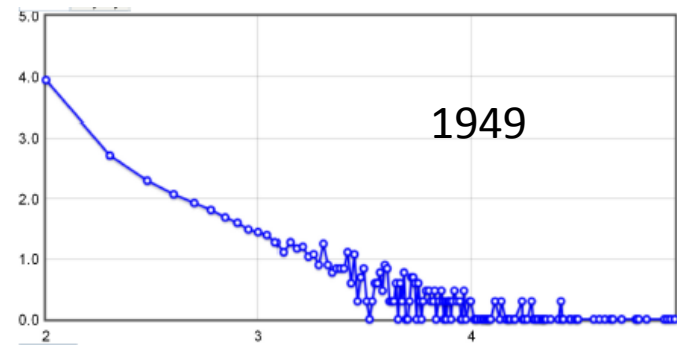
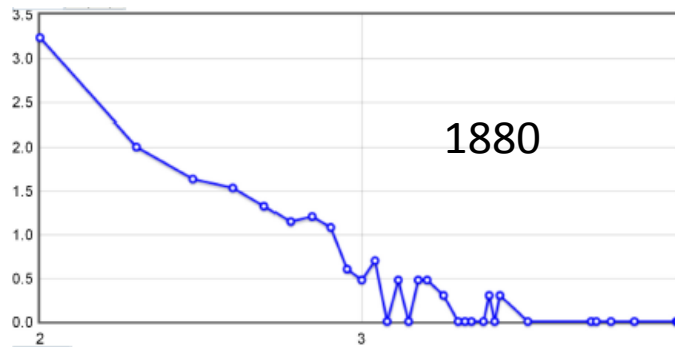
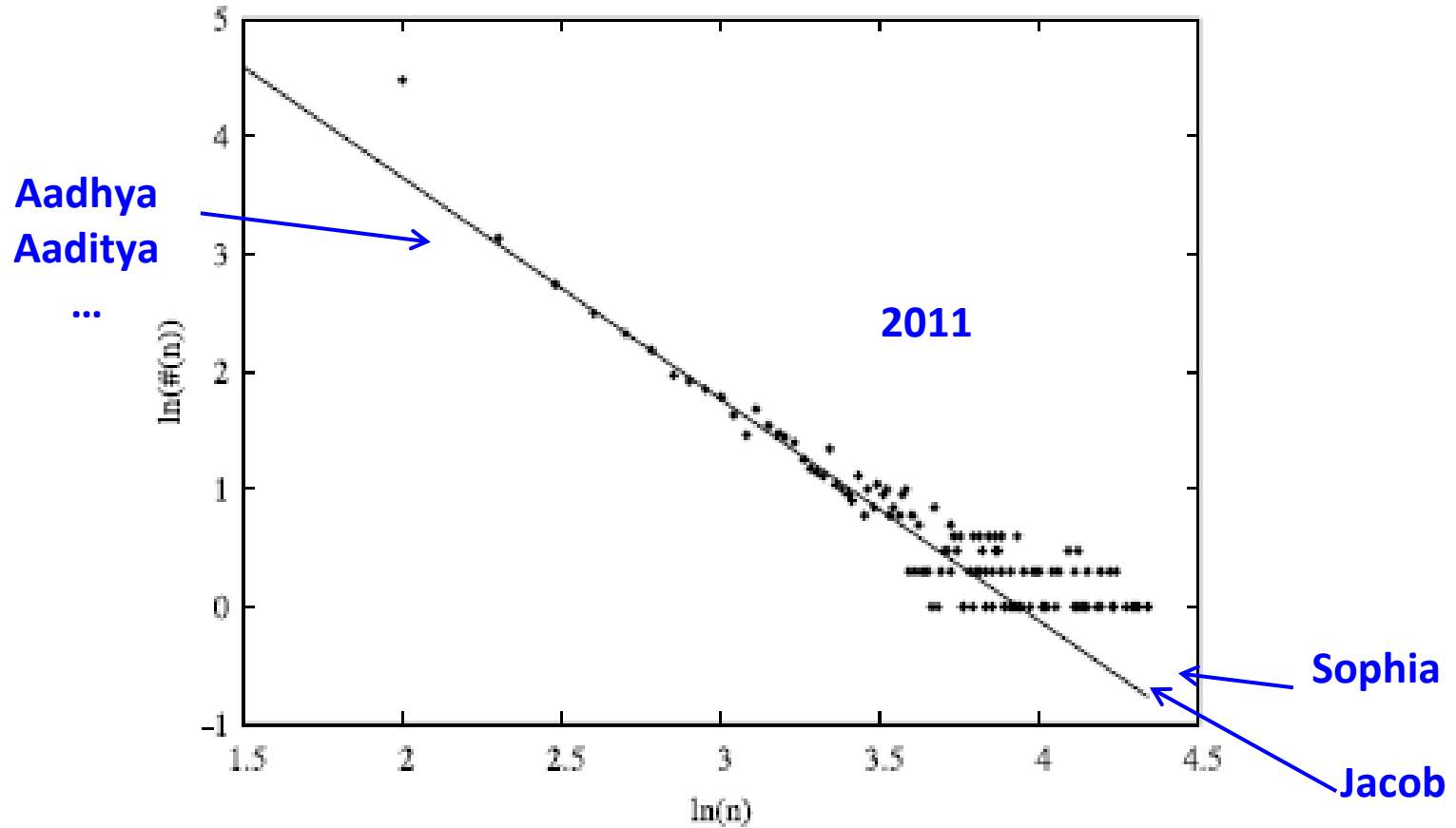
<http://statbel.fgov.be/fr/statistiques/chiffres/population/noms/>

USA: ilość dzieci którym nadano dane imię w każdym stanie (bez Alaski i Hawajów) w latach 1910-2010; bez rozbięcia na stany od 1880: www.ssa.gov

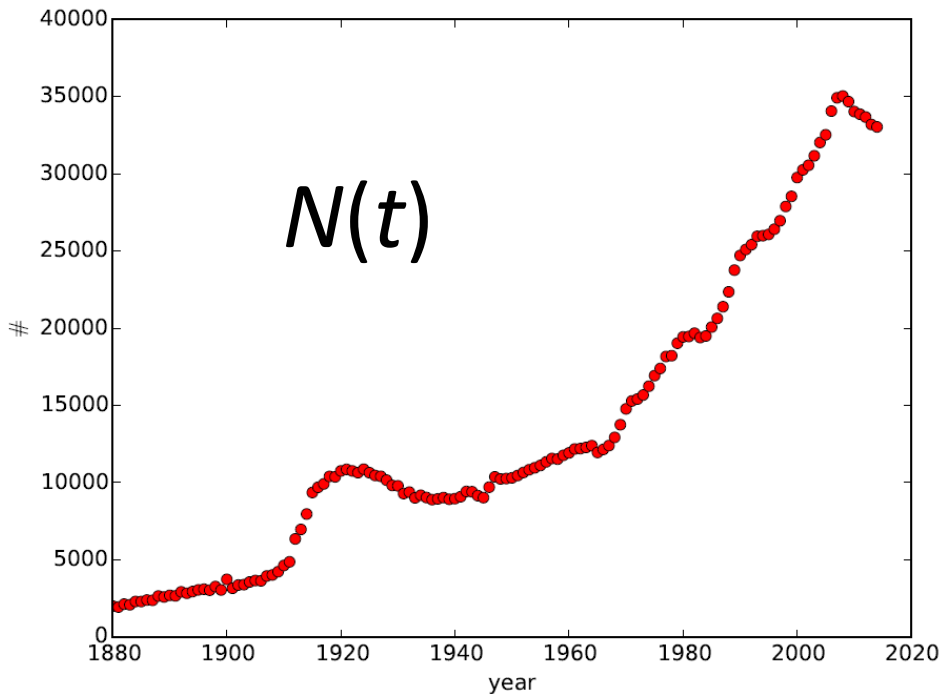
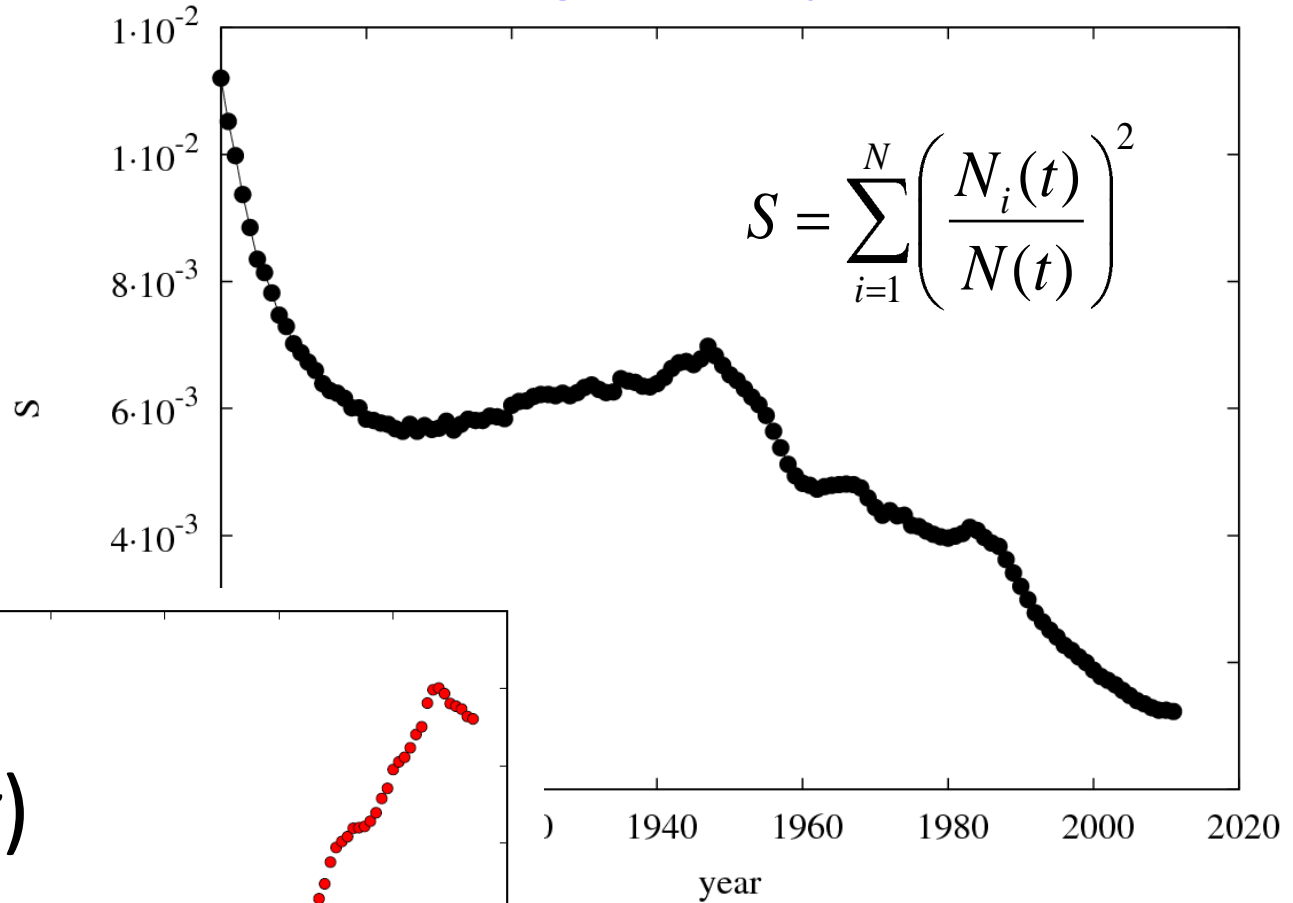
Przykład: Charlie



Dane: rozkład częstości imion



Dane: ilość imion $N(t)$, fragmentacja $S(t)$



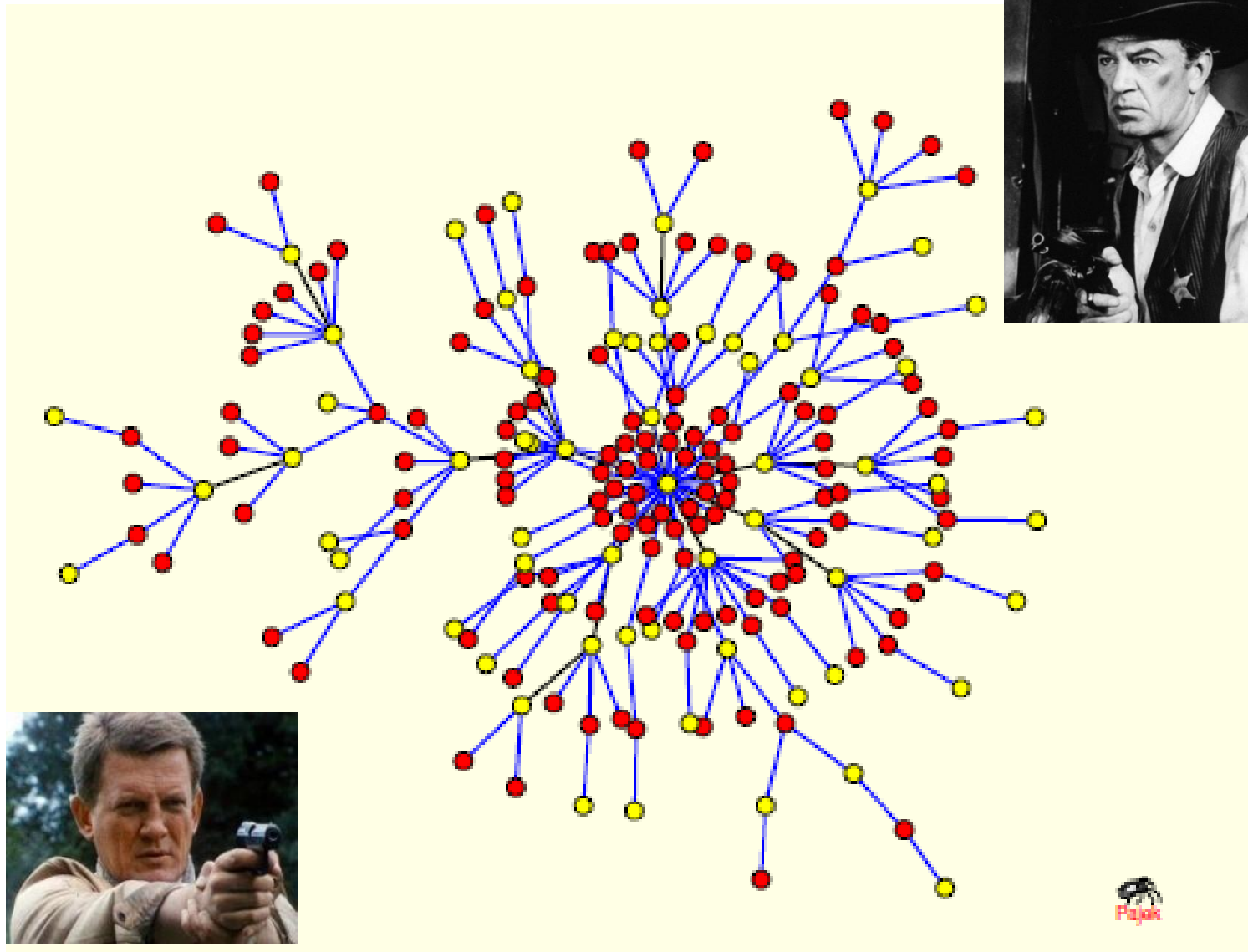


Efekt Simmla

... agenci są skłonni nie tylko do naśladowania, ale także, aby odróżnić się od innych. (Simmel) opisał modę jako efekt łącznego działania obu postaw. (...) Gdy grupy społeczne są sortowane według rangi, aktorzy naśladują symbole statusu określające wyższe poziomy hierarchii i porzucają symbole wyznaczające niższe poziomy. Na skutek tej dynamiki symbole wysokiego statusu szybko rozprzestrzeniają się w populacji, przebiegając w dół od wyższych do niższych poziomów w hierarchii społecznej. Jednak, gdy tylko się rozprzestrzeniają, symbole te zostają porzucone i zastąpione nowymi.

*[Roberto Pedone and Rosaria Conte,
Soc. Sci. Comp. Rev. 2001]*





Sekwencje naśladowania w sieci bezskalowej

Dlaczego sieci bezskalowe?

- Niektóre sieci społeczne są bezskalowe [Newman, SIAM 2003; Fu et al., Phys A 2008]
- Stopień wierzchołka jest wygodnym miernikiem statusu

„The degree of an actor is important; thus, a centrality measure for an individual actor should be the degree of the node.”

[Wasserman & Faust, Social Network Analysis 1994, p. 178]

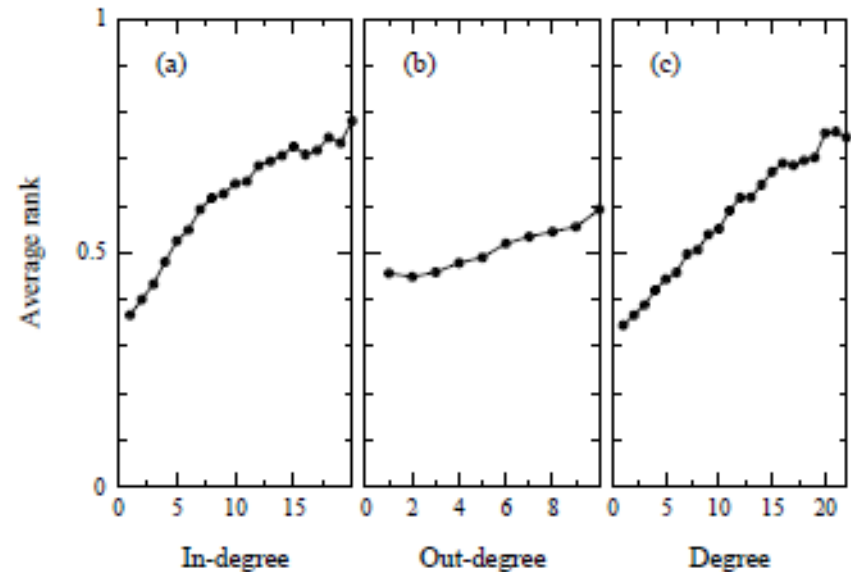
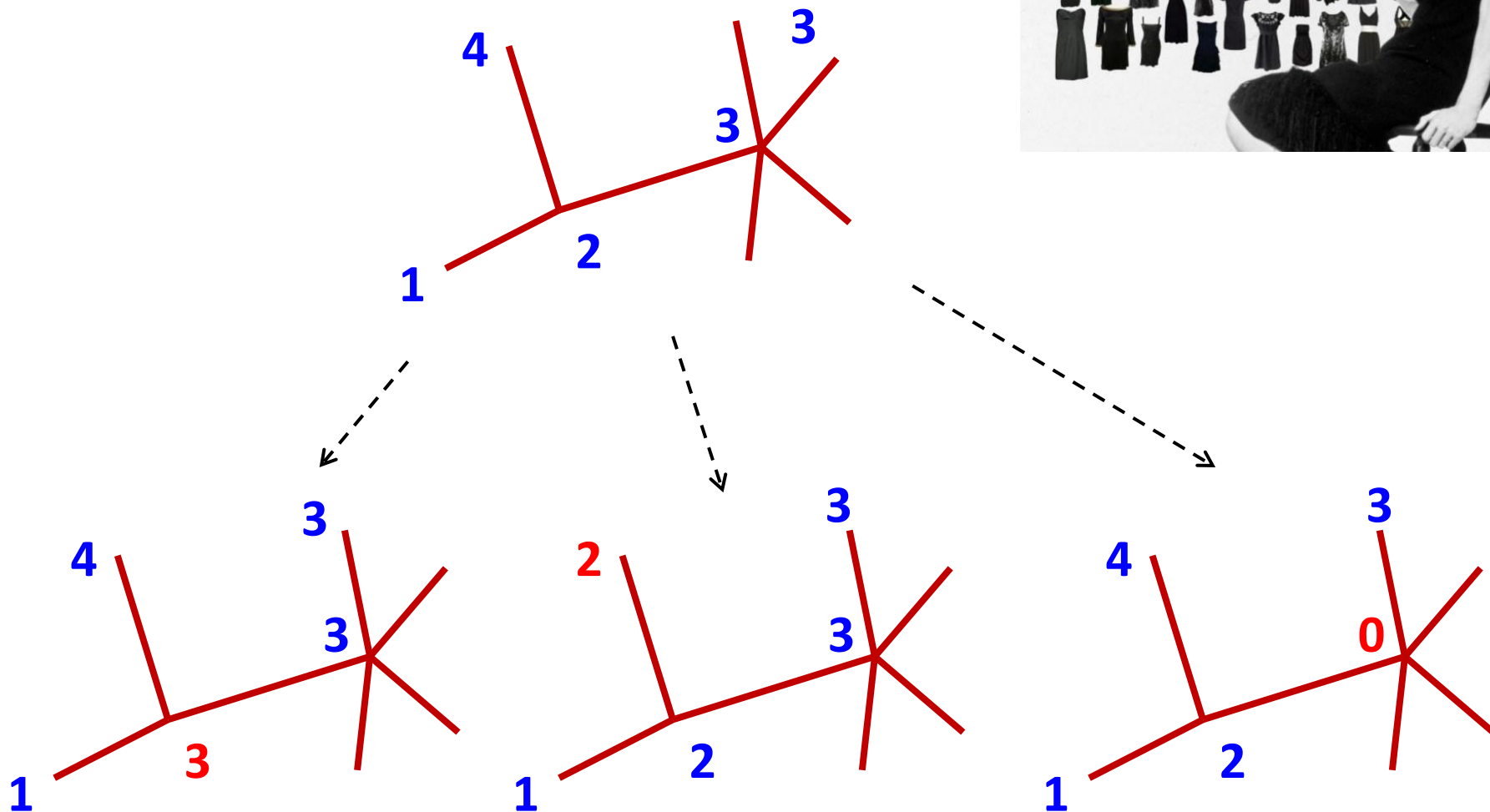
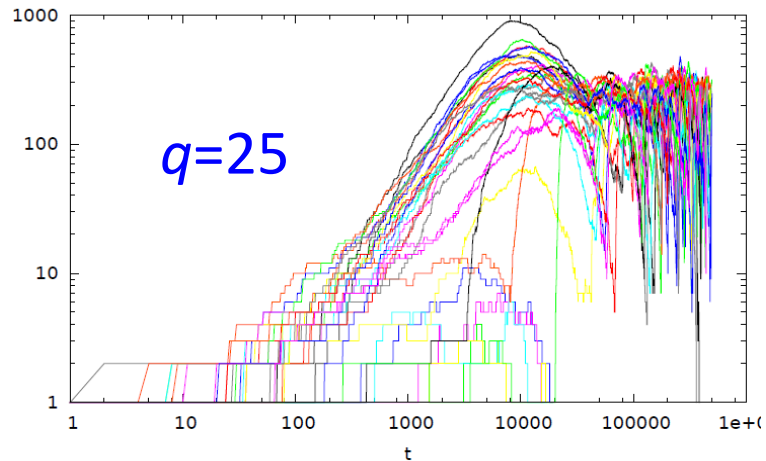
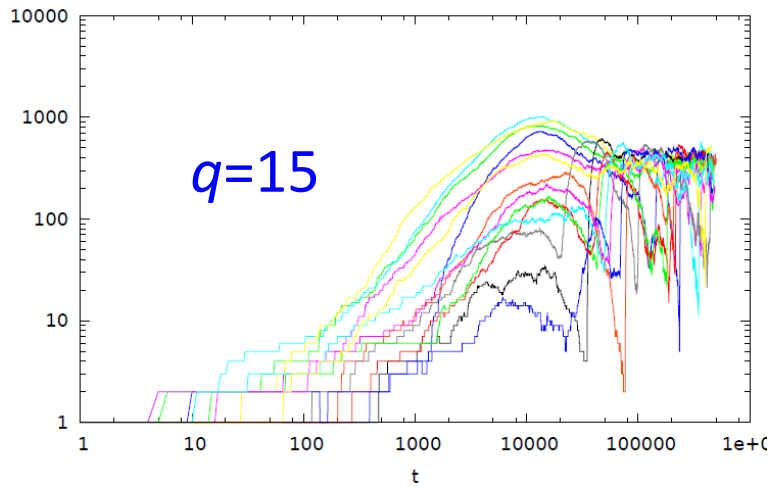
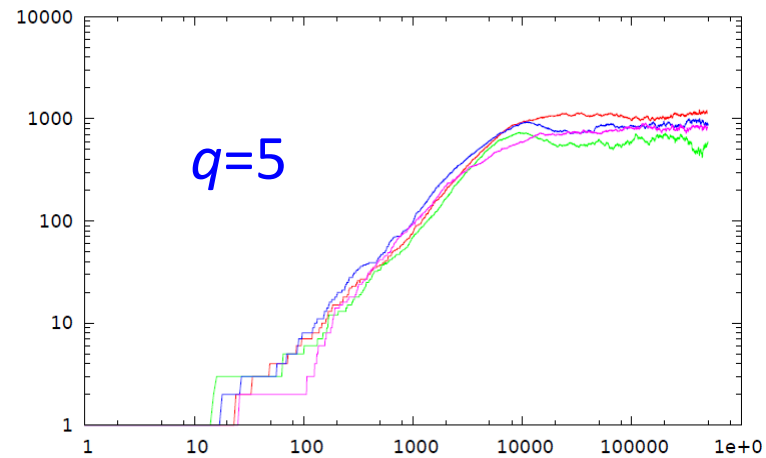


FIG. 4: Plots of rescaled rank versus degree, averaged over all individuals in all networks for (a) in-degree, (b) out-degree, and (c) the sum of degrees. Measurement errors are comparable with or smaller than the sizes of the data points and are not shown.

[Ball & Newman, arXiv:1205.6822
sieci przyjaźni 90 000 studentów]^{11/17}

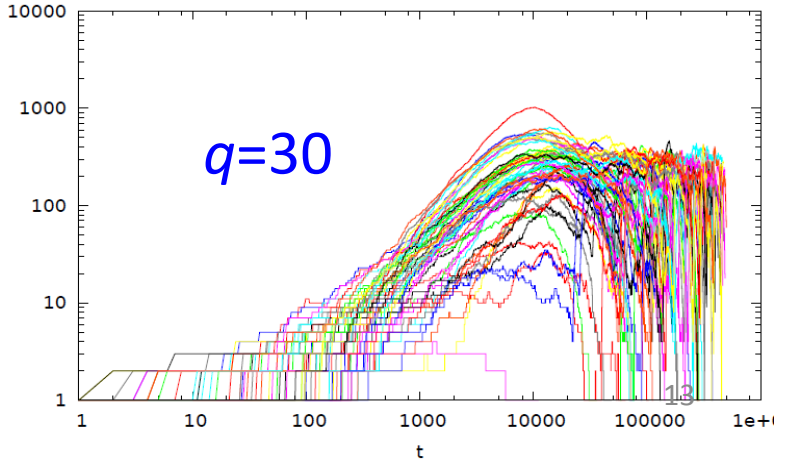
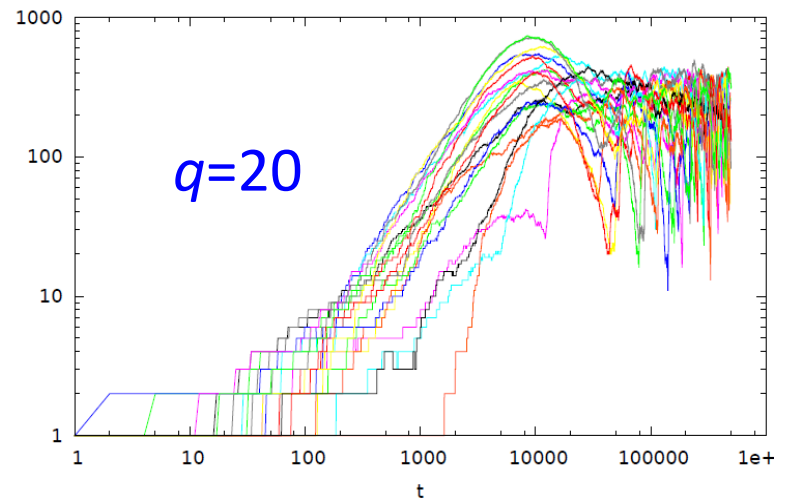
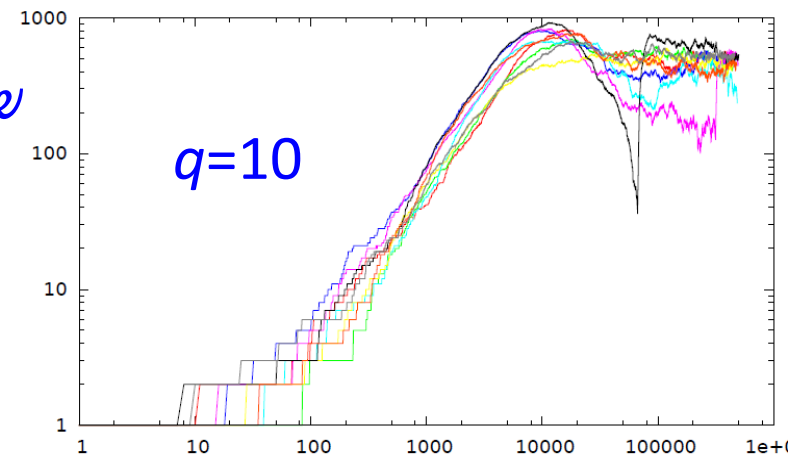
Algorytm





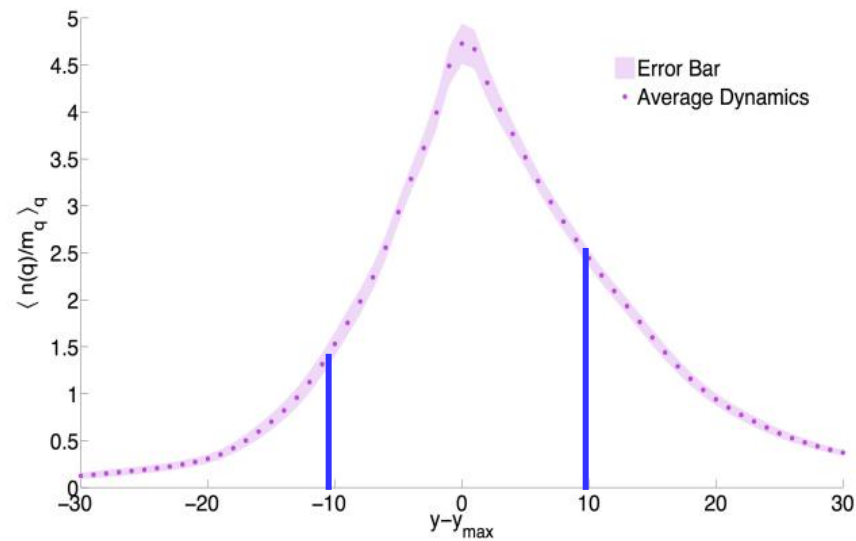
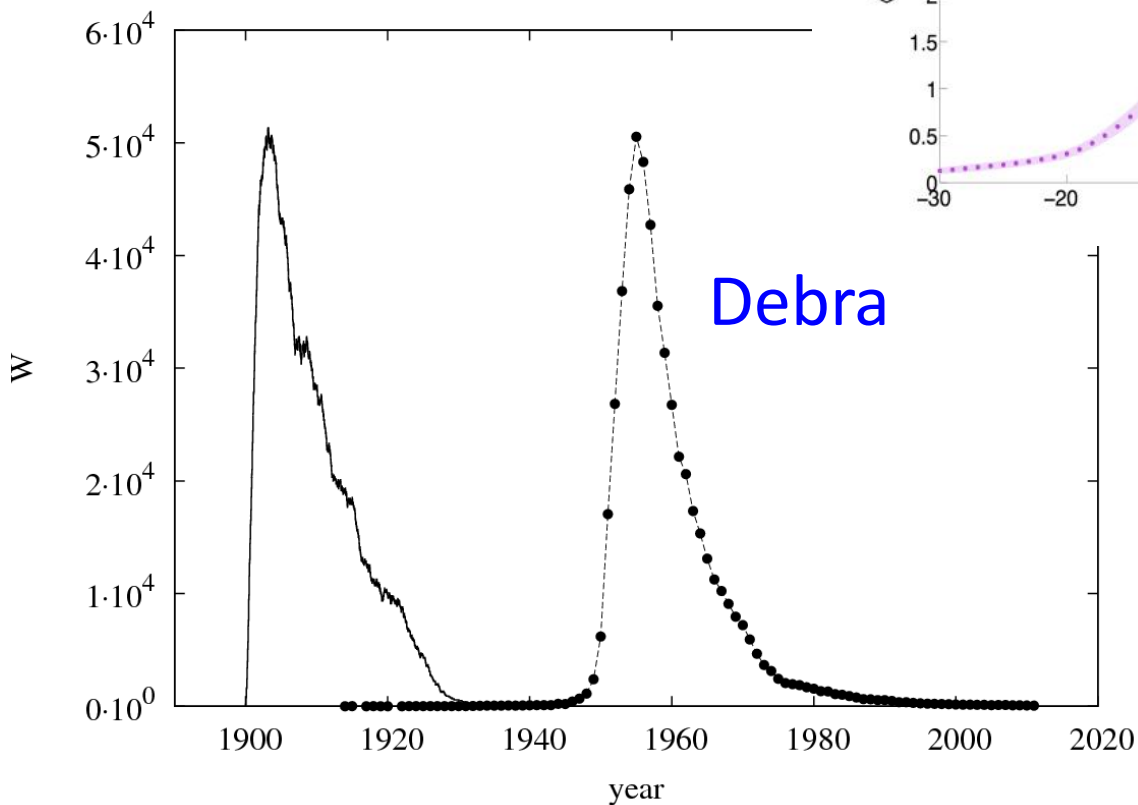
*Wyniki
numeryczne*

**Gdy ilość
 q opcji
wzrasta,
poszcze-
gólne
symbole
zaczynają
wymierać.**



Dane i wyniki numeryczne: kształt

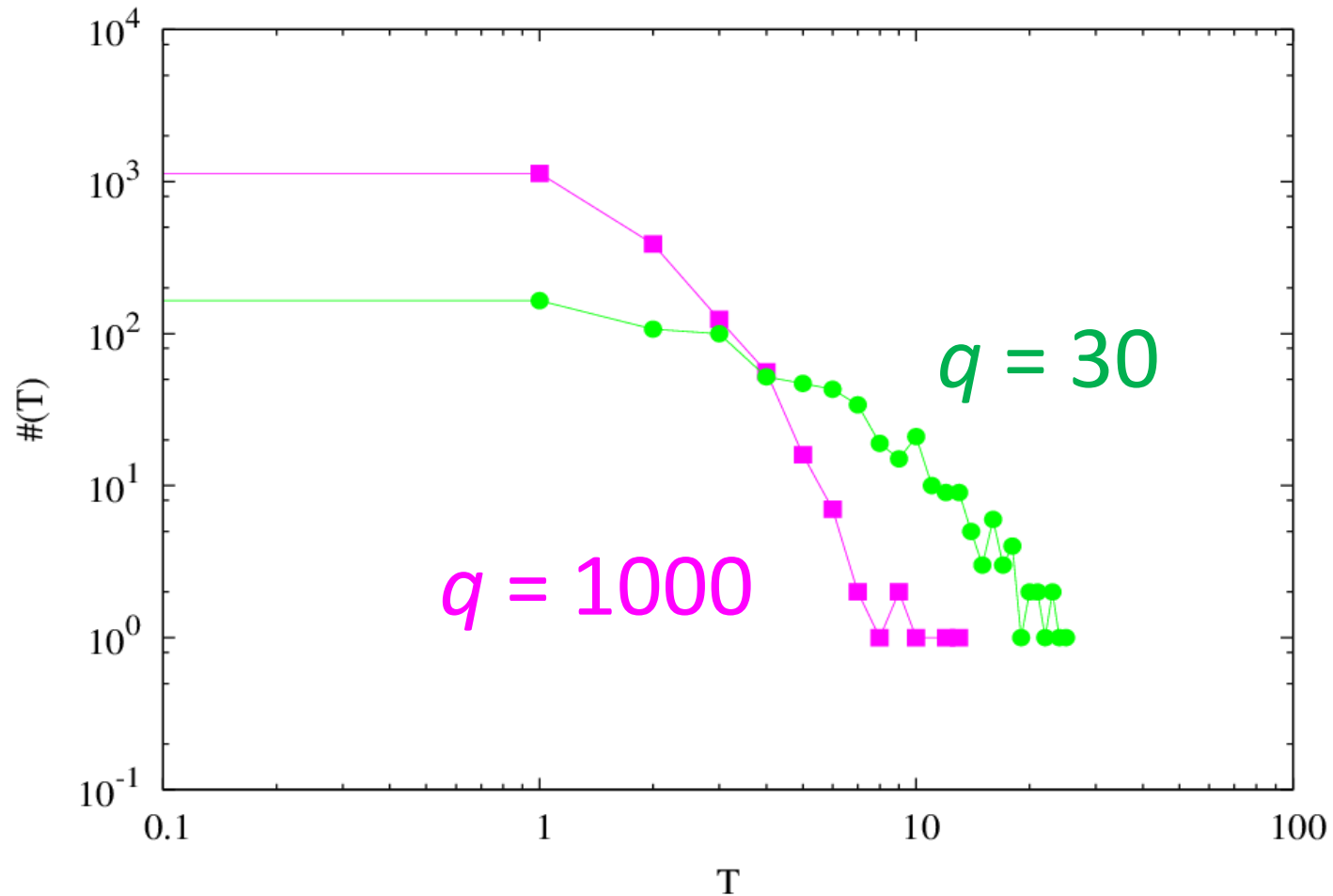
λ



P. Barucca, J. Rocchi,
E. Marinari, G. Parisi,
F. Ricci-Tersenghi ,
*Cross correlations of
American baby names*,
PNAS 112 (2015) 7943;
(max między 1940 i 1982)

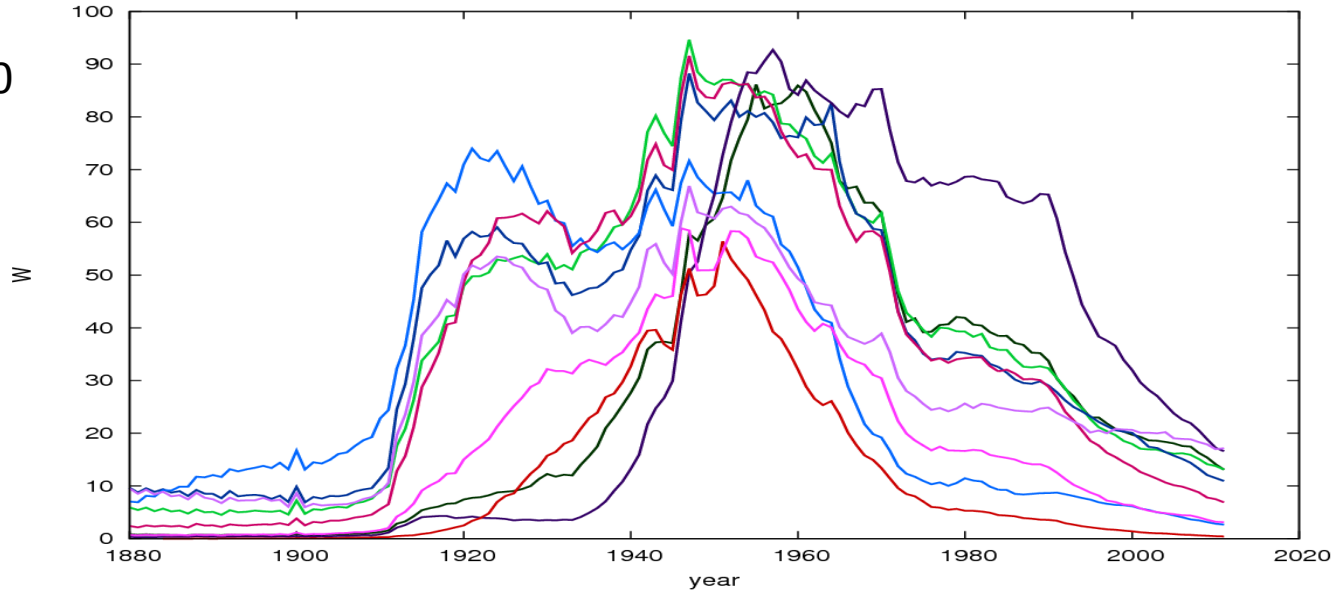
Wyniki numeryczne

Rozkład czasów trwania imienia

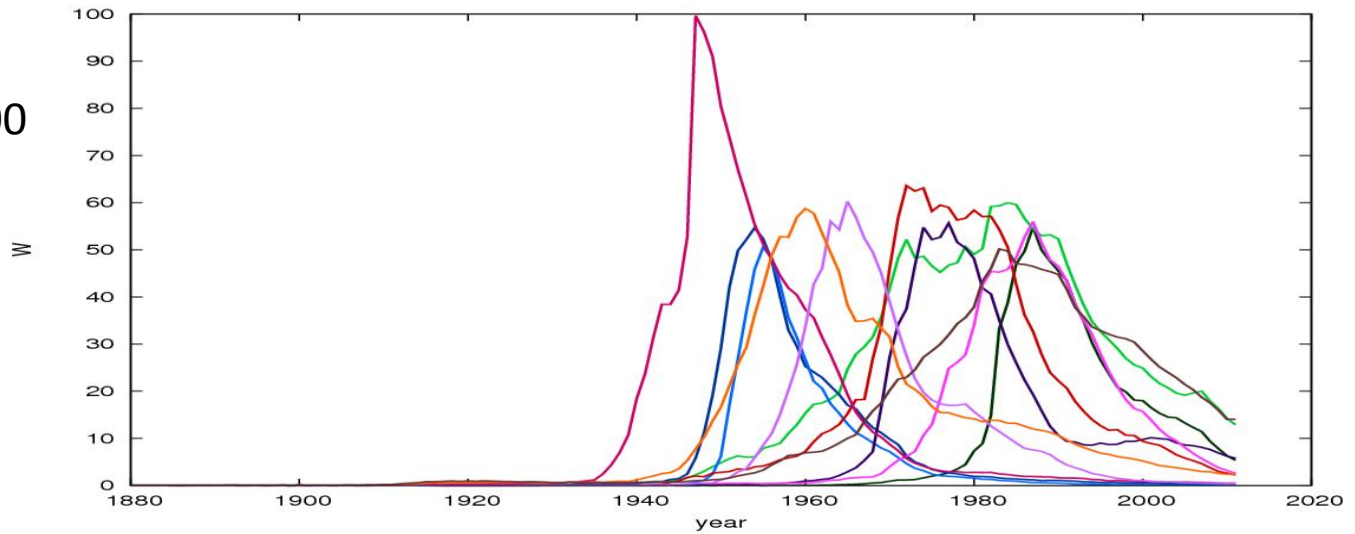


Dane -20 najpopularniejszych imion

$W(1935) > 2000$

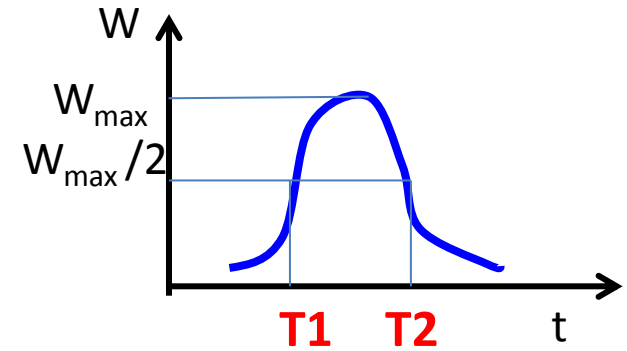
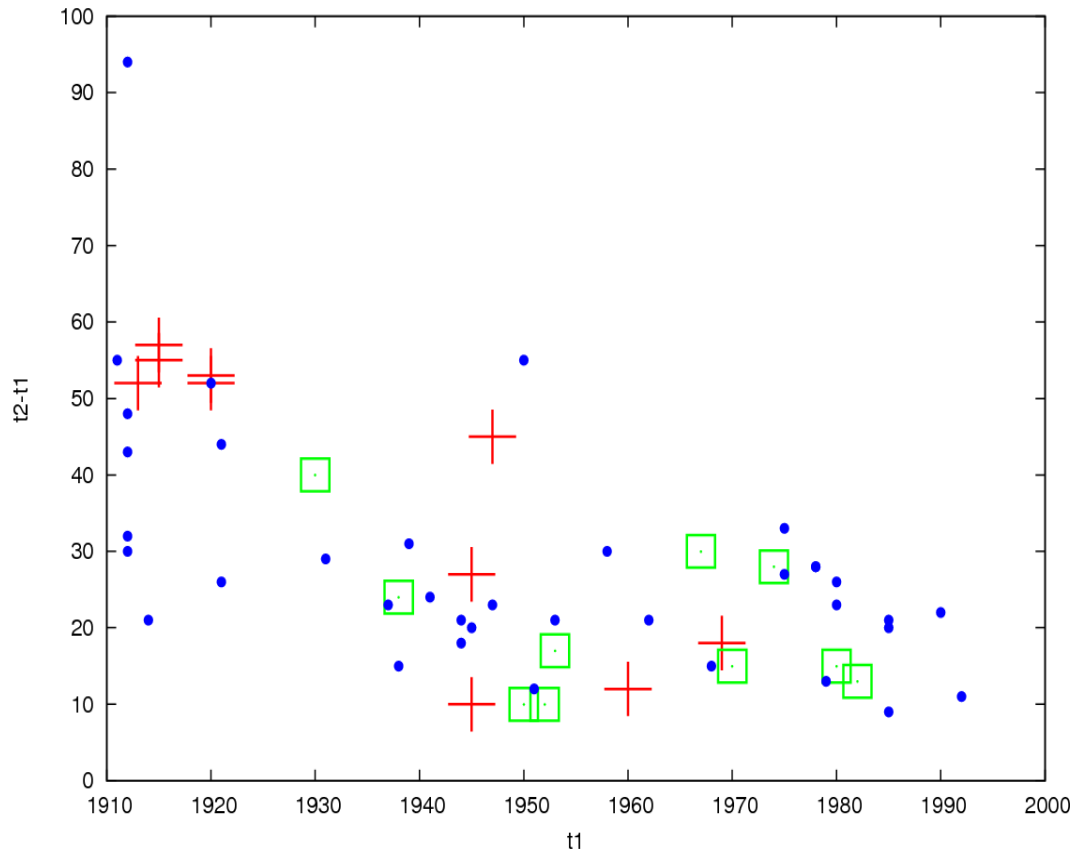


$W(1935) < 2000$



Dane: czy czas trwania imienia maleje?

T2-T1



- + : $\max \in 60-100 \times 10^3$
- \square : $50-60 \times 10^3$
- \bullet : $25-50 \times 10^3$

T1

podsumowanie 1/4

Teoria Simmla:

- kształt λ ilości imienników w czasie
- czas trwania mody maleje z ilością imion
(dane nie wykluczają takiej prawidłowości)

teraz będzie: moda na mapie

dane : graf korelacji

i, j – imię

a, b – stan

$t = 1910, \dots, 2011$

$n(i, t, a)$

www.ssa.gov

Korelacje ρ_{ij} między trochę innymi zmiennymi:

Barucca *et al*, 2015:

$$x(i, t, a) = \frac{n(i, t, a)}{\sum_j n(j, t, a)}$$

$i, j = 1, 19492$ dziewcząt

ta praca:

$$y(i, t, a) = \frac{x(i, t, a)}{\sum_b x(i, t, b)}$$

$i, j = 1, 100$

klastry stanów - obliczenia

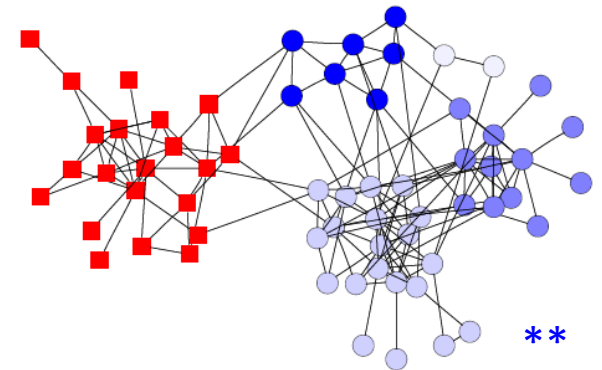
$$w_{ab} = \frac{1 + \rho_{ab}}{2}$$

$$\frac{dw_{ab}}{dt} = G(w_{ab}) \sum_c (w_{ac} w_{cb} - \beta)$$

$$G(x) = \Theta(x)\Theta(1-x)$$

$$Q = \frac{1}{m} \sum_{ab} \left(w_{ab} - \frac{k_a k_b}{m} \right) \delta(a, b)$$

$$k_a = \sum_c w_{ac} \quad m = \sum_a k_a$$



* M. J. Krawczyk, PRE 77 (2008) 065701

** M. E. J. Newman, M. Girvan, PRE 69 (2004) 026113

test metody

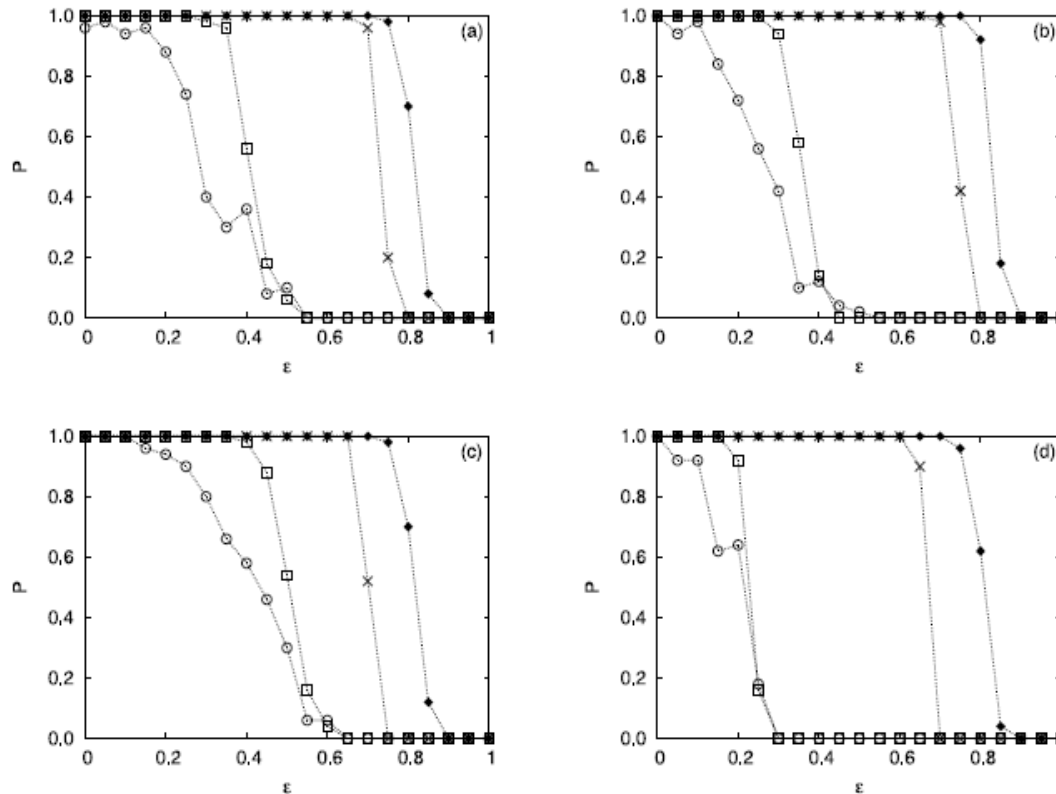
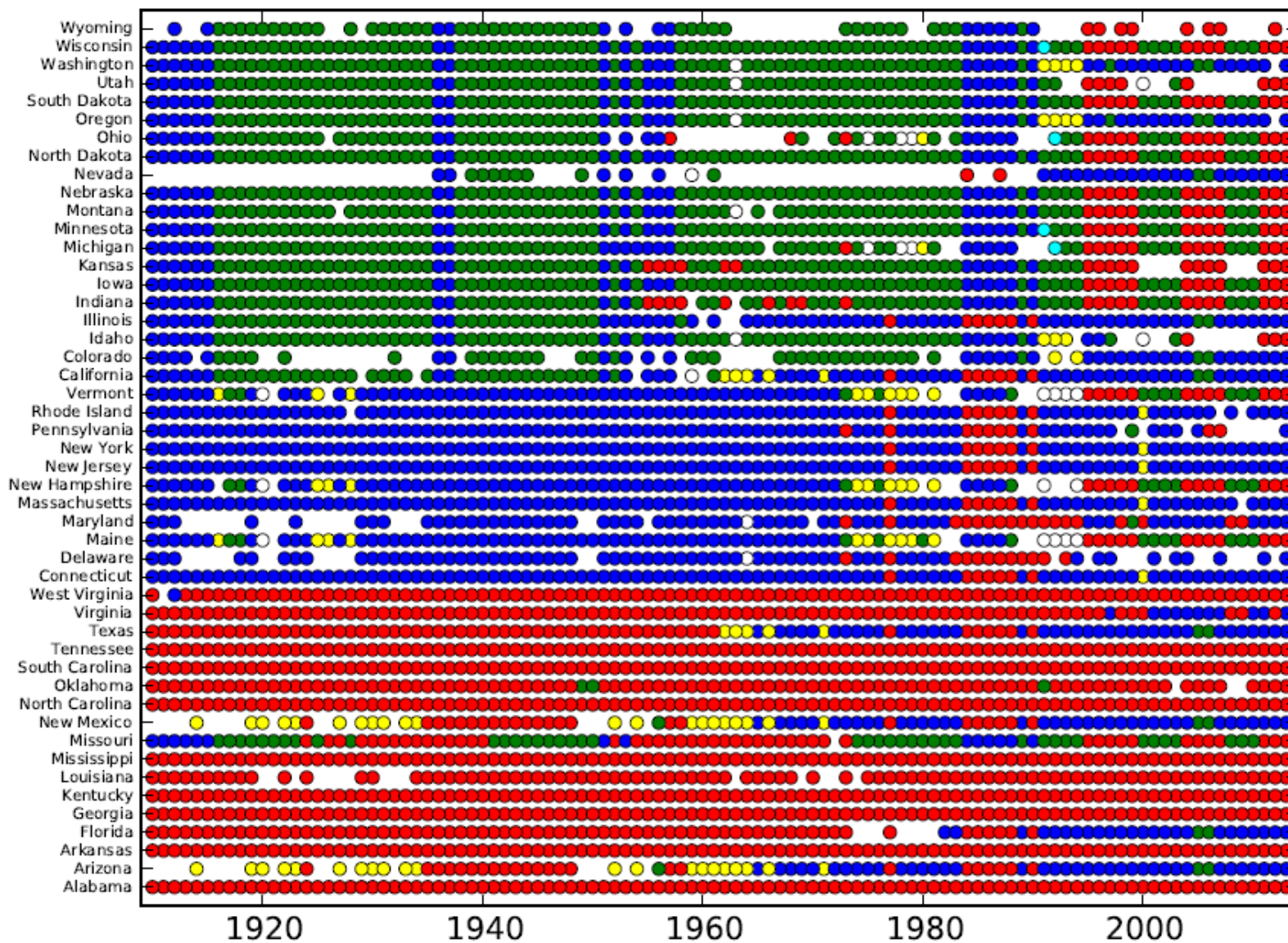


Fig. 2. Results for the network of size $N = 300$, clusters of: (a) 165, 93 and 42 nodes with $pz = 1/3$, (b) 175, 75 and 50 nodes with $pz = 1/3$, (c) 148, 117 and 35 nodes with $pz = 1/4$, (d) 191, 74 and 35 nodes with $pz = 1/4$ (symbols denote: \circ - the Duch and Arenas method, \square - the Newman method, \times - Eq. (1) with $\beta = 0.25$, \blacklozenge - Eq. (1) with $\beta = 0.4$).

[Krawczyk M.J., *PRE* 77 (2008) 065701(R); *Comp. Phys. Comm.* 181 (2010) 1702]

klastry stanów - wyniki



Klasy stanów - wyniki

Region 1: Northeast

Division 1: New England

Connecticut (09)
Maine (23)
Massachusetts (25)
New Hampshire (33)
Rhode Island (44)
Vermont (50)

9

Division 2: Middle Atlantic

New Jersey (34)
New York (36)
Pennsylvania (42)

Region 2: Midwest*

Division 3: East North Central

Indiana (18)
Illinois (17)
Michigan (26)
Ohio (39)
Wisconsin (55)

12

Division 4: West North Central

Iowa (19) Nebraska (31)
Kansas (20) North Dakota (38)
Minnesota (27) South Dakota (46)
Missouri (21)

Region 3: South

Division 5: South Atlantic

Delaware (0)
District of Columbia (11)
Florida (12)
Georgia (13)
Maryland (24)
North Carolina (37)
South Carolina (45)
Virginia (51)
West Virginia (54)

Division 6: East South Central

Alabama (01)
Kentucky (21)
Mississippi (28)
Tennessee (47)

16

Division 7: West South Central

Arkansas (05)
Louisiana (22)
Oklahoma (40)
Texas (48)

Region 4: West

Division 8: Mountain

Arizona (14) Montana (30)
Colorado (08) Utah (49)
Idaho (16) Nevada (32)
New Mexico (35) Wyoming (56)

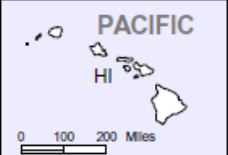
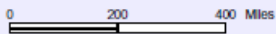
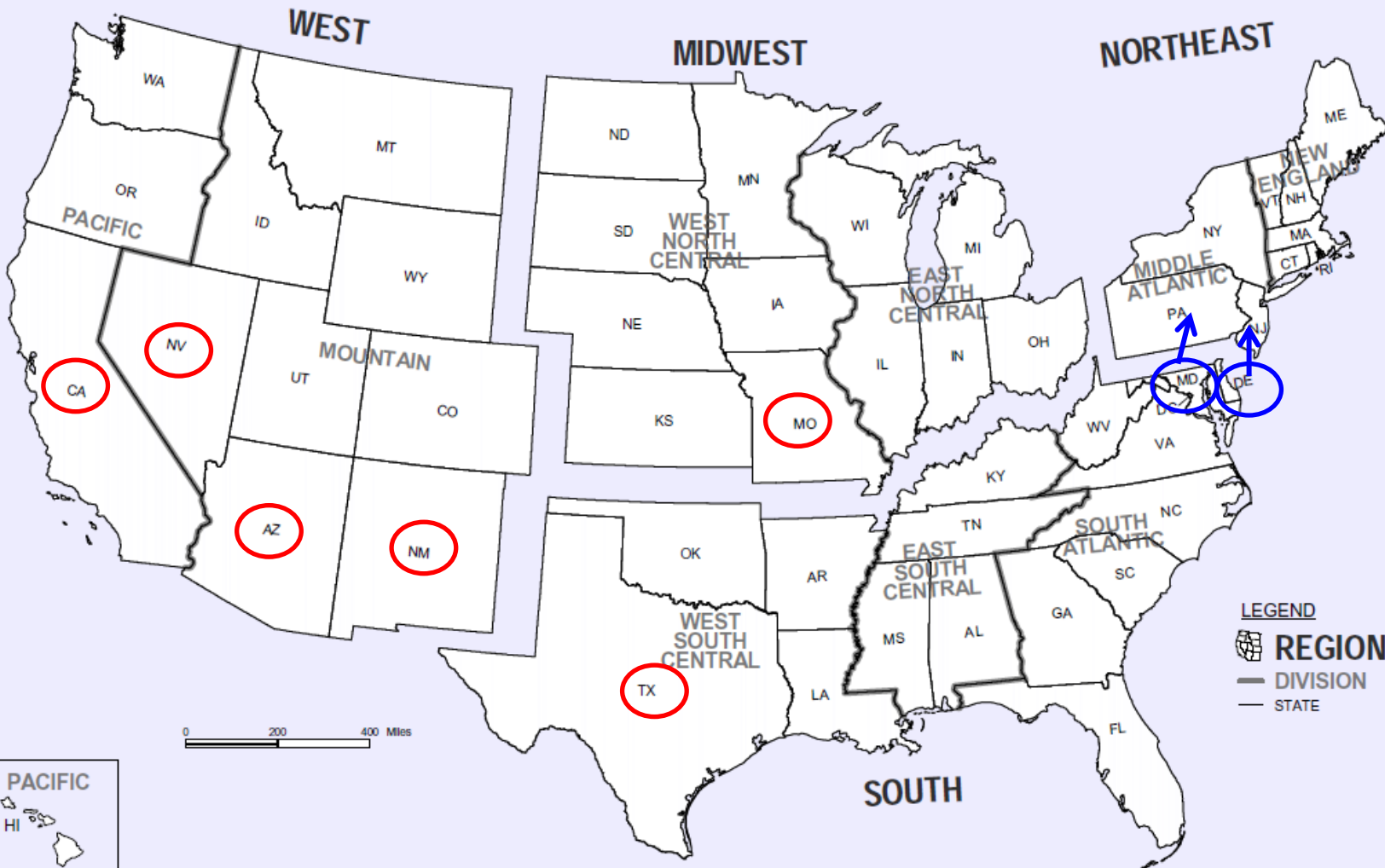
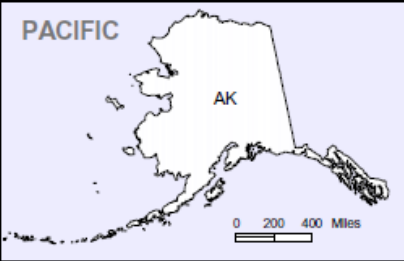
11

Division 9: Pacific

Alaska (02)
California (0)
Hawaii (45)
Oregon (41)
Washington (53)

- wszystkie 9 stanów *Northeast* są przypisane do wspólnego klastra;
- 18 stanów *Midwest + West* należą do wspólnego klastra. Jednak status Arizony, Kalifornii, Missouri, Nowego Meksyku i Nevady zmienia się w czasie, więc ich przynależność pozostaje otwarta;
- 13 stanów *South* należą do wspólnego klastra. Jednak status Teksasu zmienia się w czasie. Poza tym Delaware i Maryland przypisano do *NorthEast*.

Census Regions and Divisions of the United States

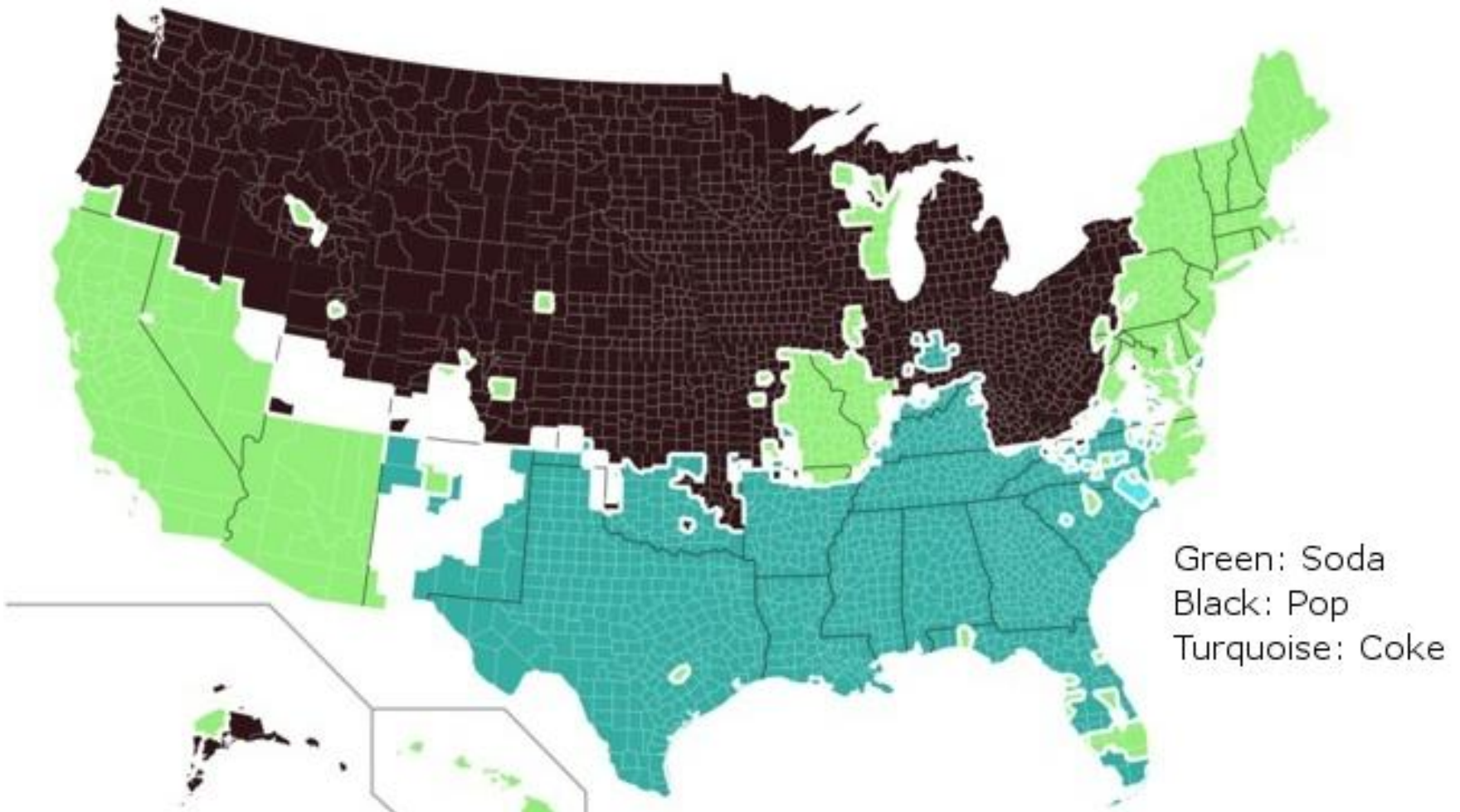


LEGEND

- REGION
- DIVISION
- STATE

S. Arbesman, The Atlantic CityLab, April 26, 2012

The Soft Drink Borders: Soda, Pop, or Coke?



www.citylab.com/politics/2012/04/invisible-borders-define-american-culture/1839/

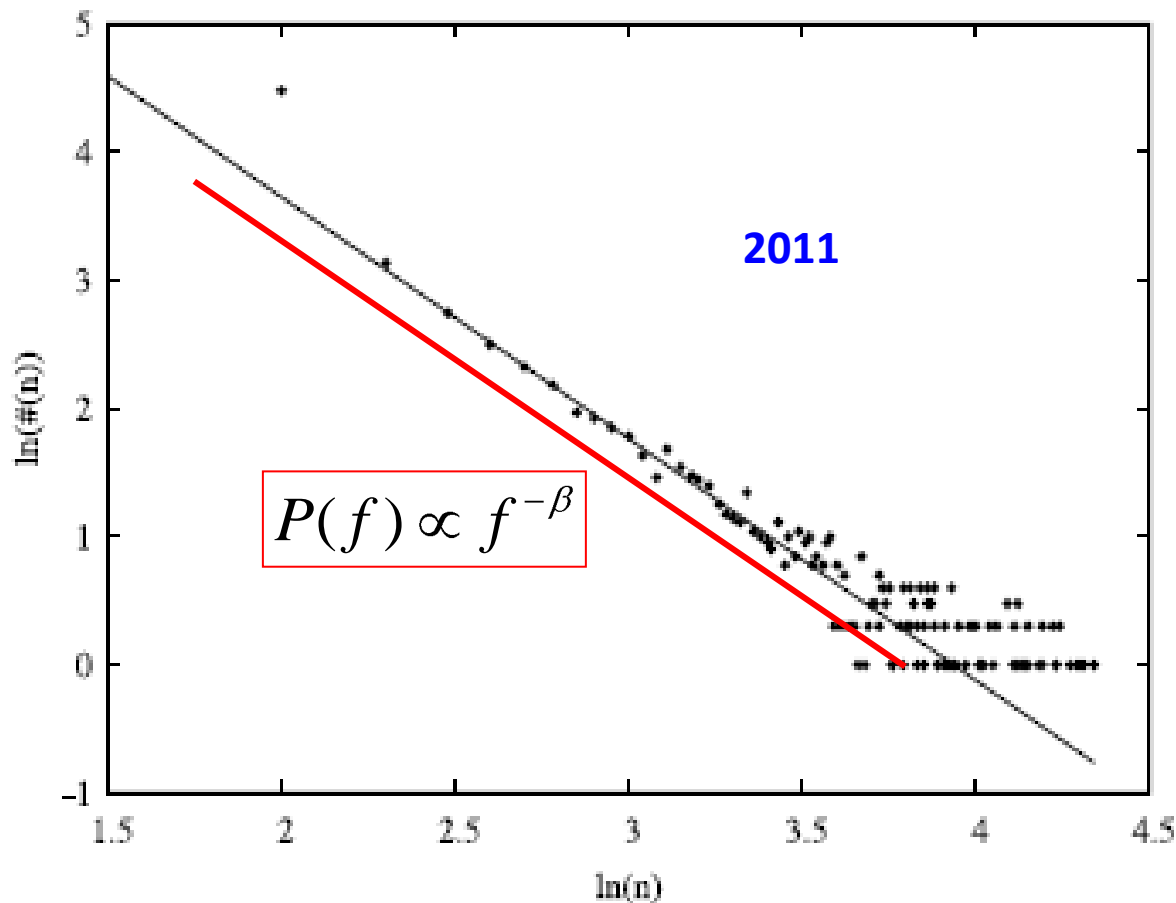
podsumowanie 2/4

Moda na mapie:

Analiza częstości imion w poszczególnych stanach
w przybliżeniu odtwarza regiony administracyjne USA

teraz będzie: prawo Zipfa

Jeszcze raz: rozkład częstości imion



Mała wartość β : dużo imion występujących często, stosunkowo mało imion egzotycznych.
Tu: w XX wieku $1.7 < \beta < 1.9$, w XXI wieku $\beta \approx 2.0$.

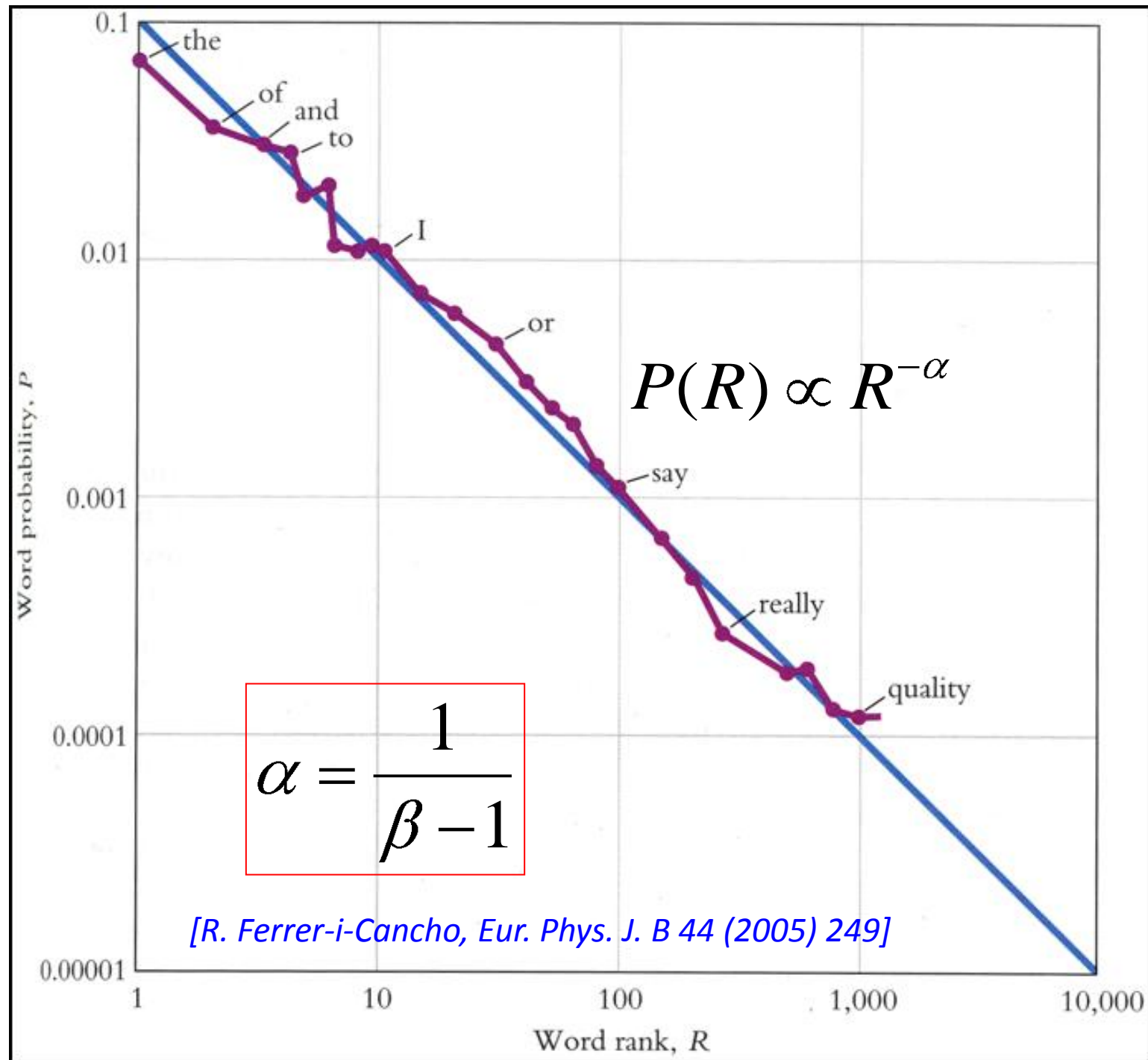
interpretacja β w tekstach

- $\beta \cong 2$ - wartość typowa; $\beta \cong 3/2$ dla rzadszych słów w tekstach wieloautorskich
- $\beta > 2$ - wiele tematów, chaotyczne wypowiedzi; także zbyt małe próbki tekstu ; może też oznaczać lekką schizofrenię
- $\beta < 2$ - małe dzieci (1.6), teksty bojowe (1.7); także cięższa schizofrenia

[R. Ferrer i Cancho, Eur. Phys. J. B 44 (2005) 249]

Prawo Zipfa

www.prismnet.com/~dierdorf/ww-27.html



Uzasadnienie związku α i β

$m(n) = N P(n)$ – ilość imion o populacji n

$$m(n) \propto n^{-\beta} \qquad n(R) \propto R^{-\alpha}$$

Pozycja R w rankingu

$$R(n) = \int_n^{\infty} m(n') dn' \propto n^{1-\beta} \qquad R \propto n^{-1/\alpha}$$

Porównując, mamy

$$\frac{1}{\alpha} = \beta - 1$$

[R. Ferrer-i-Cancho, R. V. Sole, *Journal of Quantitative Linguistics* 8, 165 (2001)]

Co oznacza α ?

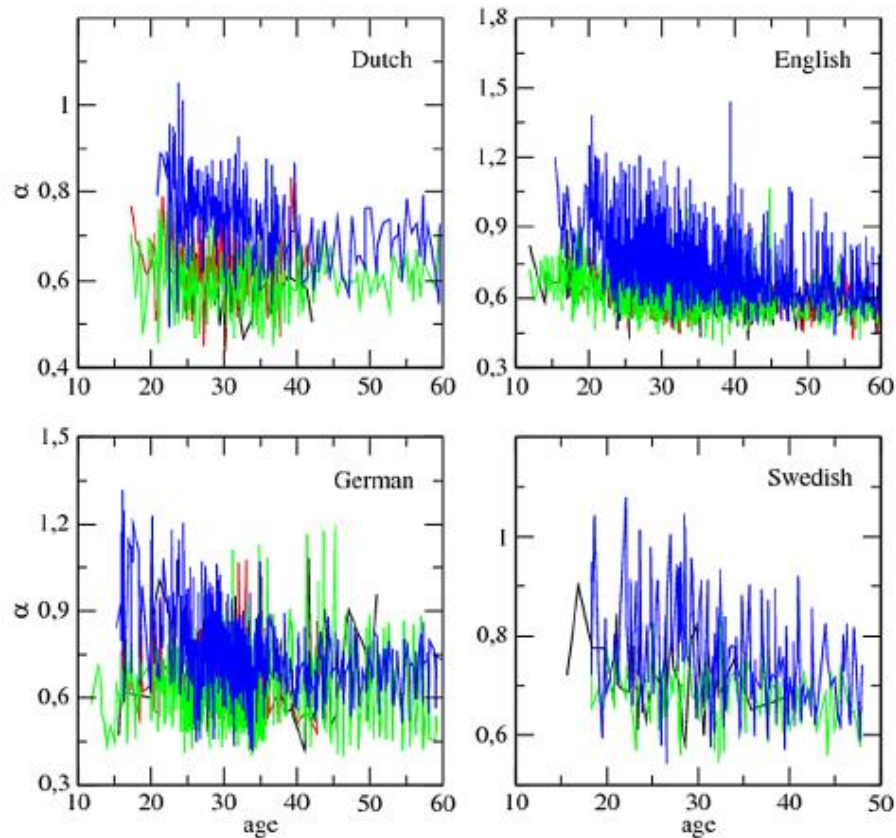


Figure 2. The evolution of the exponent α versus child age (in months): $n^* = 100$. The major classes of roles, i.e. target children (blue), mothers (green), investigators (red) and fathers (black), are shown. Length normalization by prefix with $n^* = 100$ is used. Swedish lacks the class 'investigator'.
doi:10.1371/journal.pone.0053227.g002

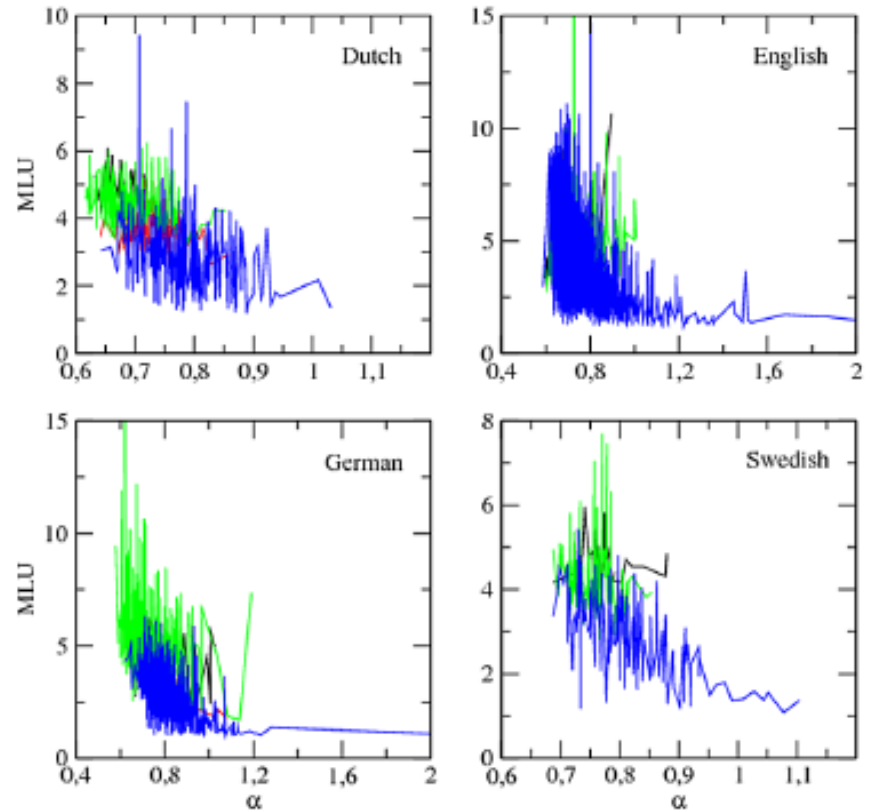
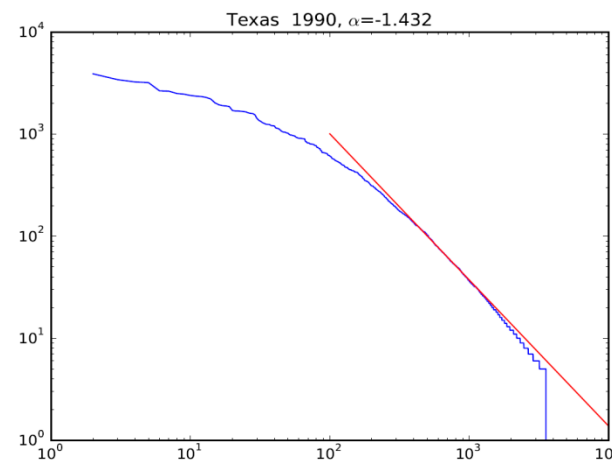
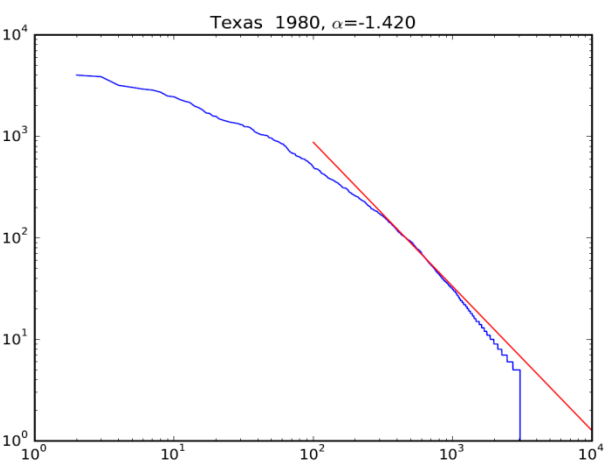
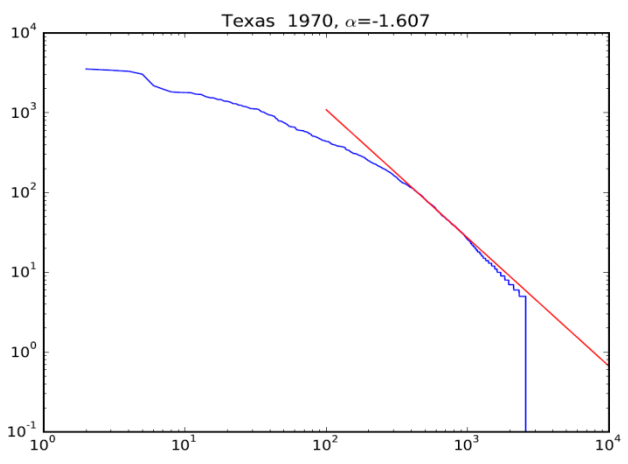
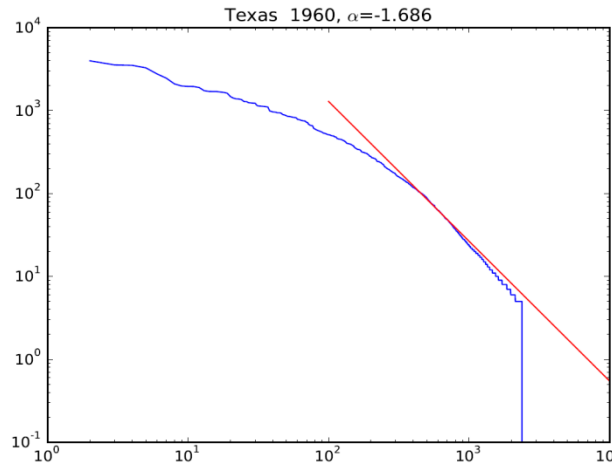
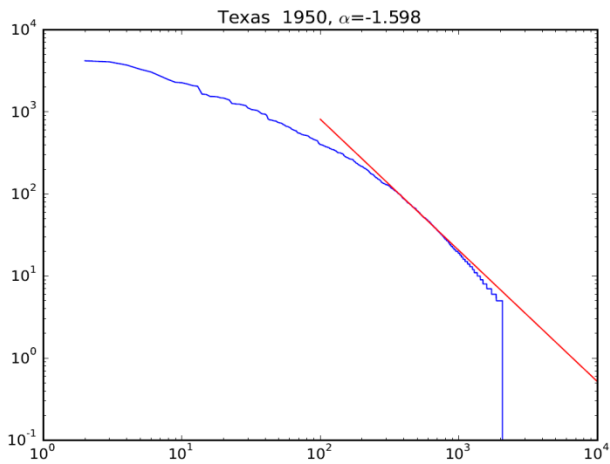
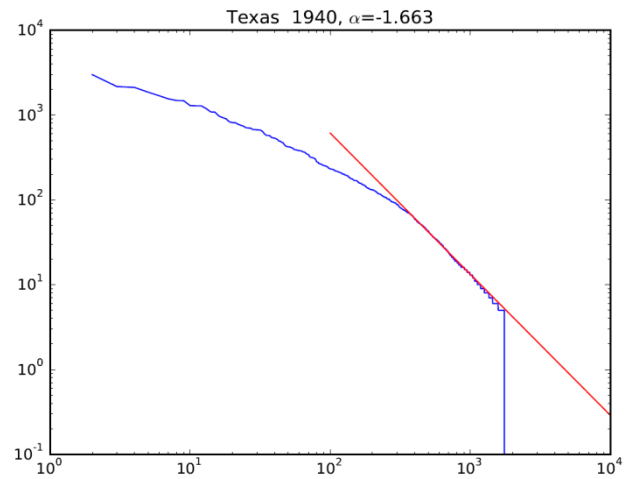
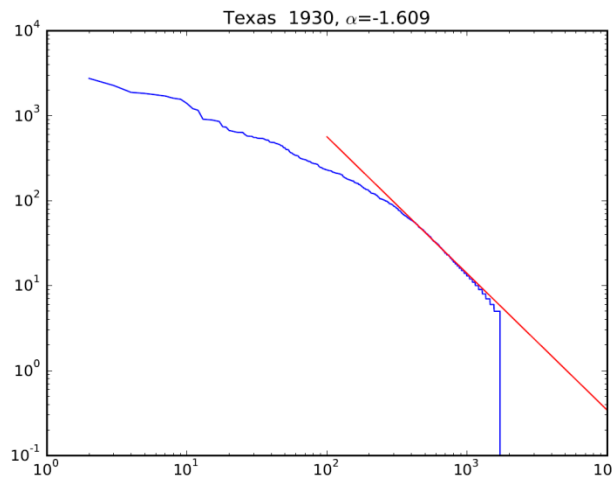
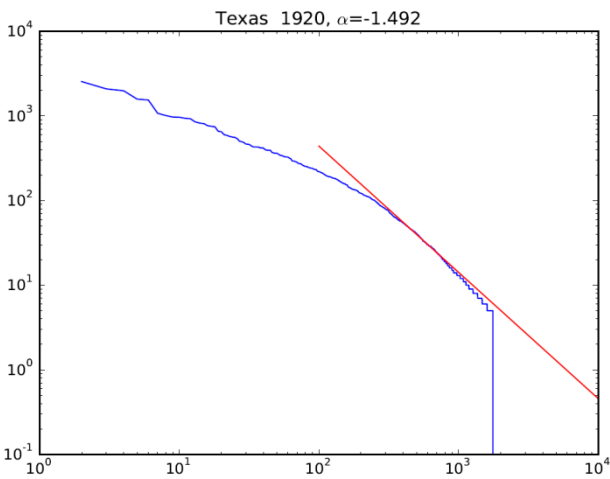


Figure 4. The MLU (in words) versus α : $T^* = 500$. The major classes of roles, i.e. target children (blue), mothers (green), investigators (red) and fathers (black), are shown. Length normalization by prefix with $T^* = 500$ is used. Swedish lacks the class 'investigator'. In order to facilitate the visual inspection of the series, the few points with MLU above 15 or α above 2 are not shown (this concerns English and German).
doi:10.1371/journal.pone.0053227.g004



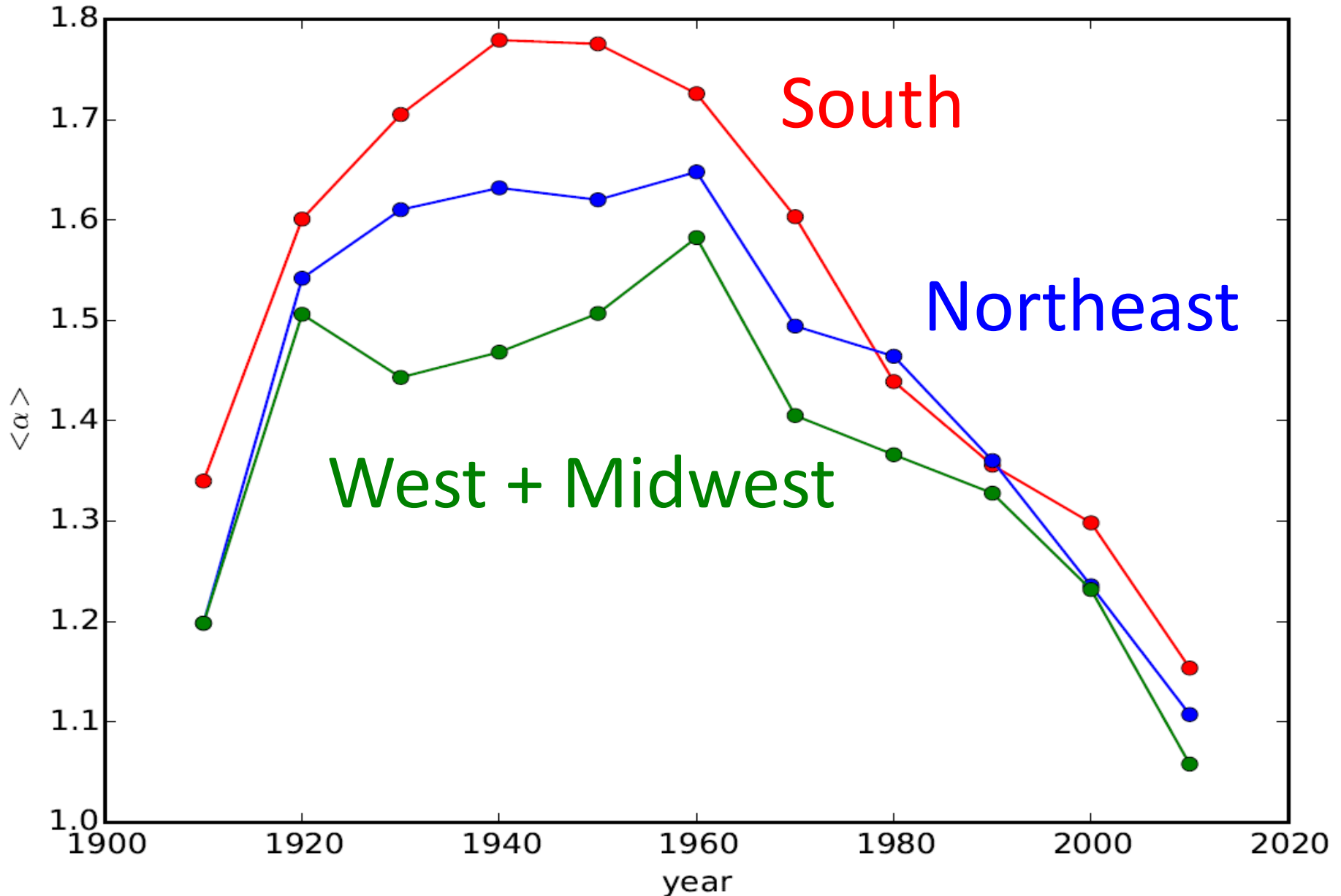
Dopasowanie danych
do prawa Zipfa
dla stanu Texas
1910-1990
home.agh.edu.pl/~gos/ZIPF/

Dane z Belgii – wykładnik α

	males	females
Belgium	1.62	1.67
Flanders	1.67	1.57
Wallonia	1.60	1.69
Brussels	1.36	1.37
Belgium < 18	1.42	1.54
Belgium 18 – 64	1.65	1.72
Belgium > 65	1.85	1.83
Flanders < 18	1.41	1.33
Flanders 18 – 64	1.65	1.60
Flanders > 65	2.18*	2.03
Wallonia < 18	1.53	1.42
Wallonia 18 – 64	1.56	1.64
Wallonia > 65	1.58	1.66
Brussels < 18	1.41*	1.30*
Brussels 18 – 64	1.37	1.33
Brussels > 65	1.32	1.40

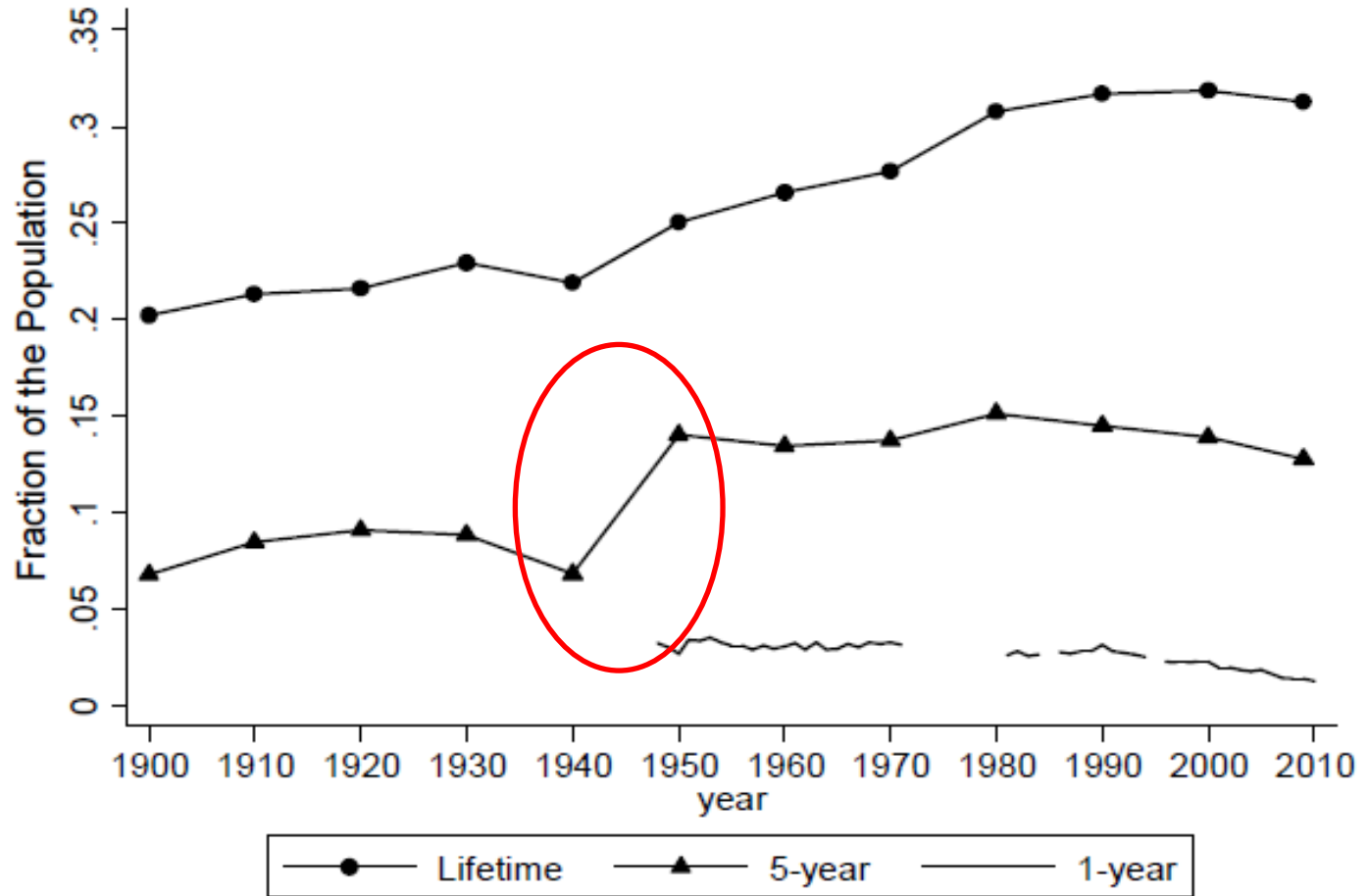


Ewolucja α w klastrach stanów USA



Przyczyny : migracja?

Figure 2
Inter-State Migration Rates Since 1900



Note. Lifetime and 5-year migration rates are from the decennial Census 1900-2000 and from the ACS for 2009. Five-year migration rates are estimated from the fraction of households with a 4- or 5-year old residing outside of their birth state (Rosenbloom and Sundstrom 2004). Annual migration rates are from the Current Population Survey.

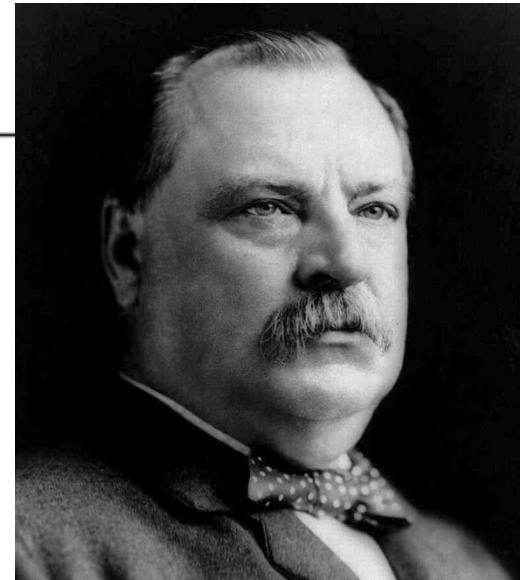
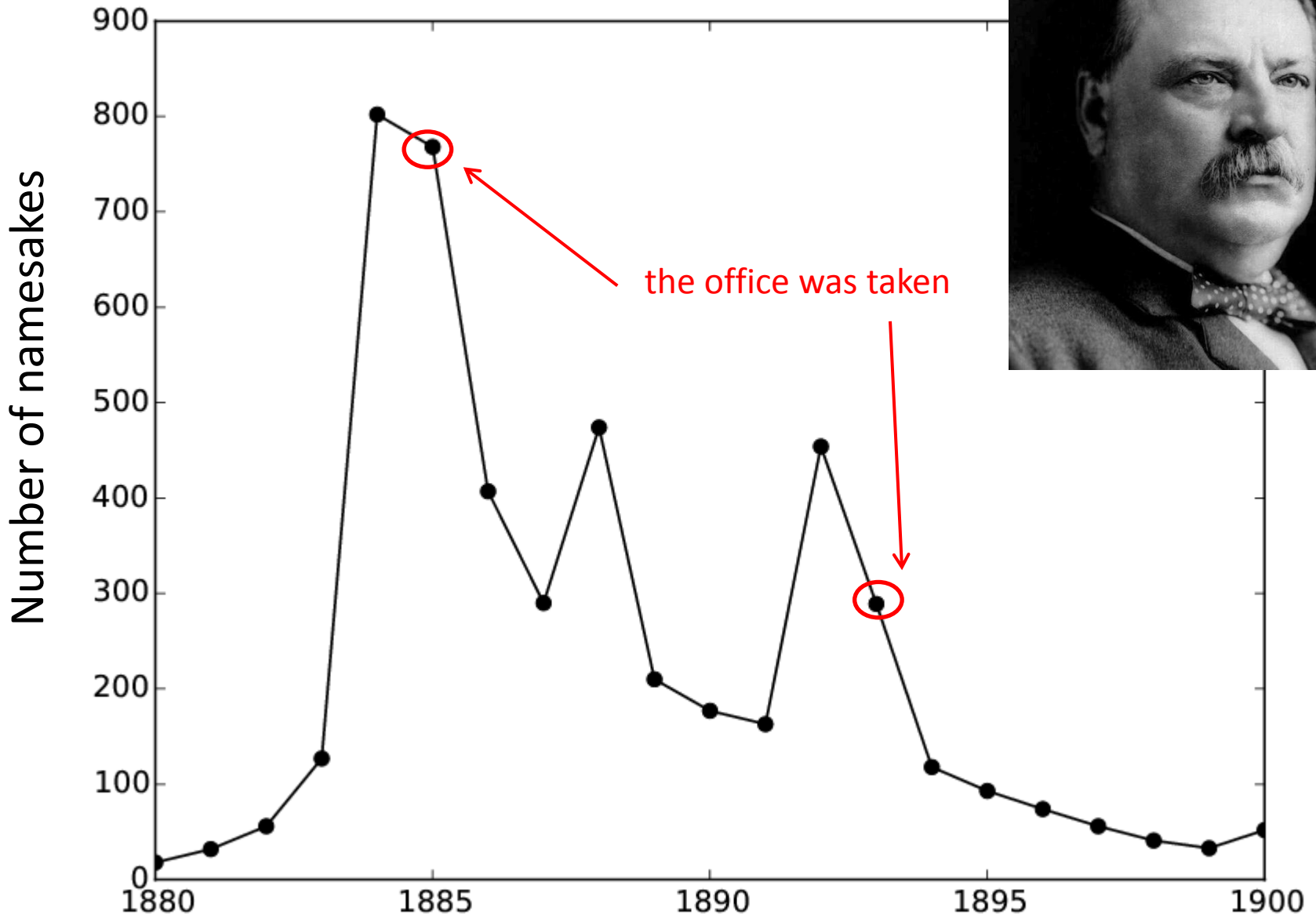
podsumowanie 3/4

Prawo Zipfa:

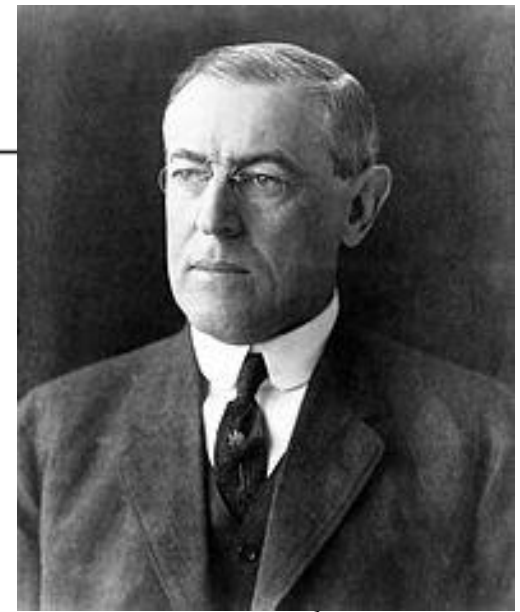
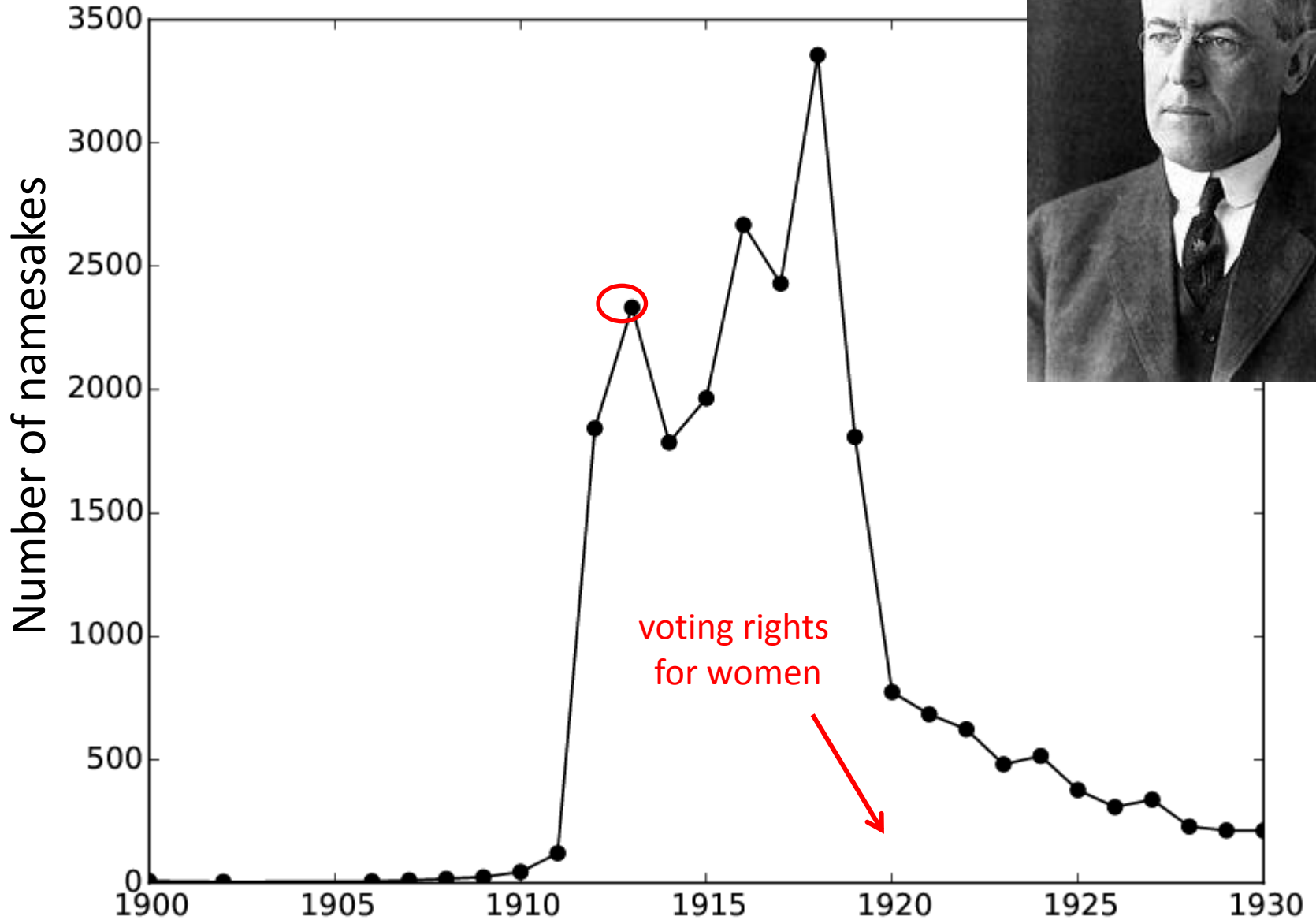
Zmierzone wskaźniki Zipfa wykazują systematyczną zależność od regionu: $\alpha(\text{West}) < \alpha(\text{Northeast}) < \alpha(\text{South})$

teraz będzie: prezydenci i inni

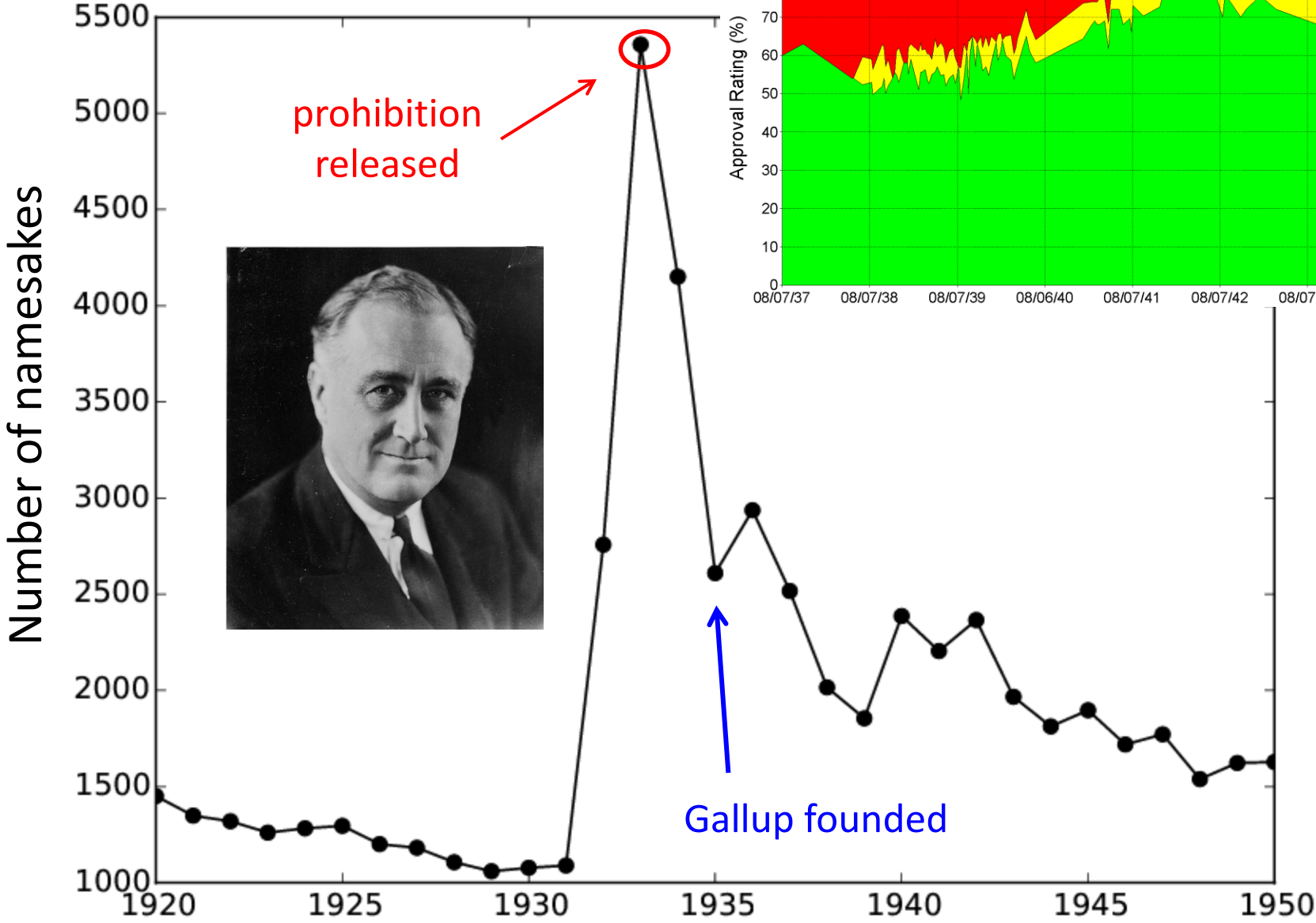
Grover Cleveland, 1885-89, 1893-97



Woodrow Wilson, 1913-21

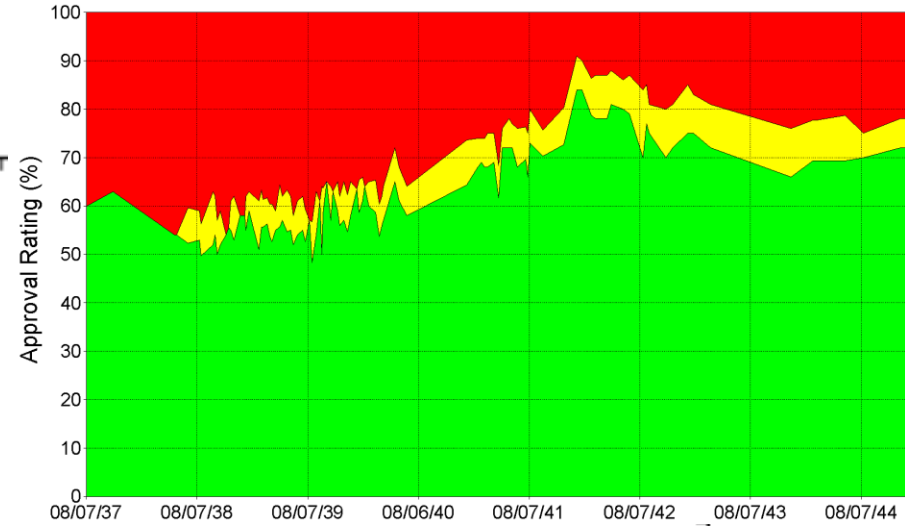


Franklin D. Roosevelt, 1933-45

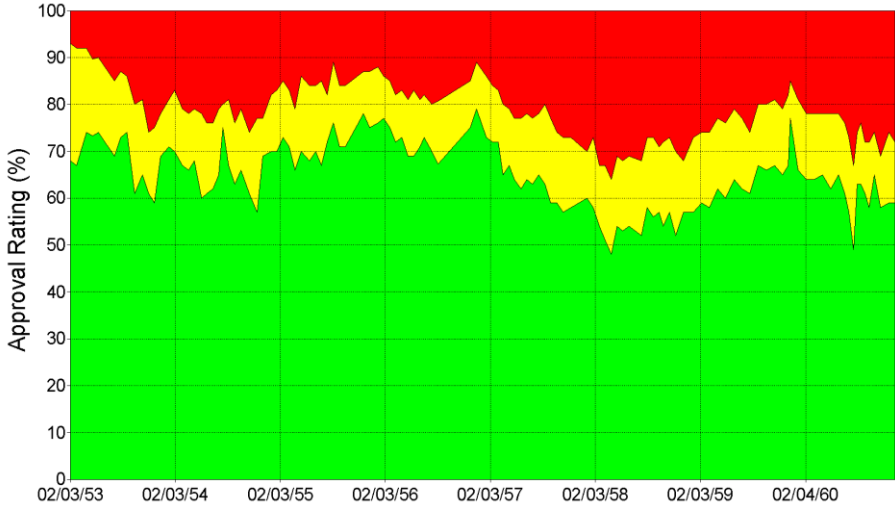
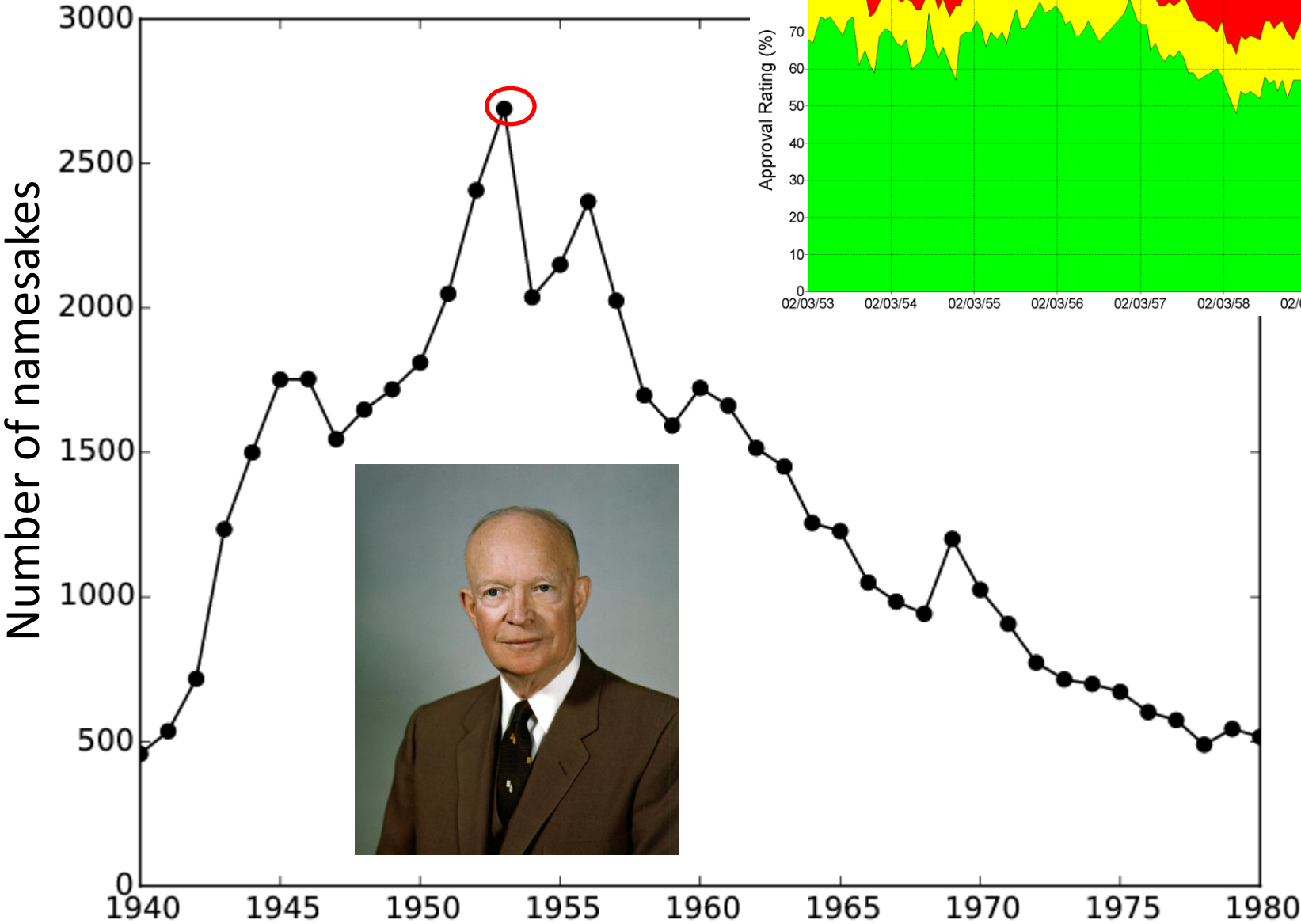


prohibition released

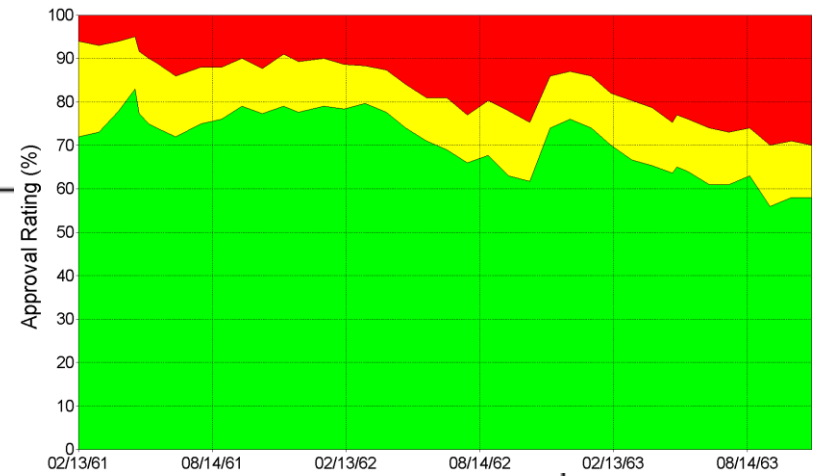
Gallup founded



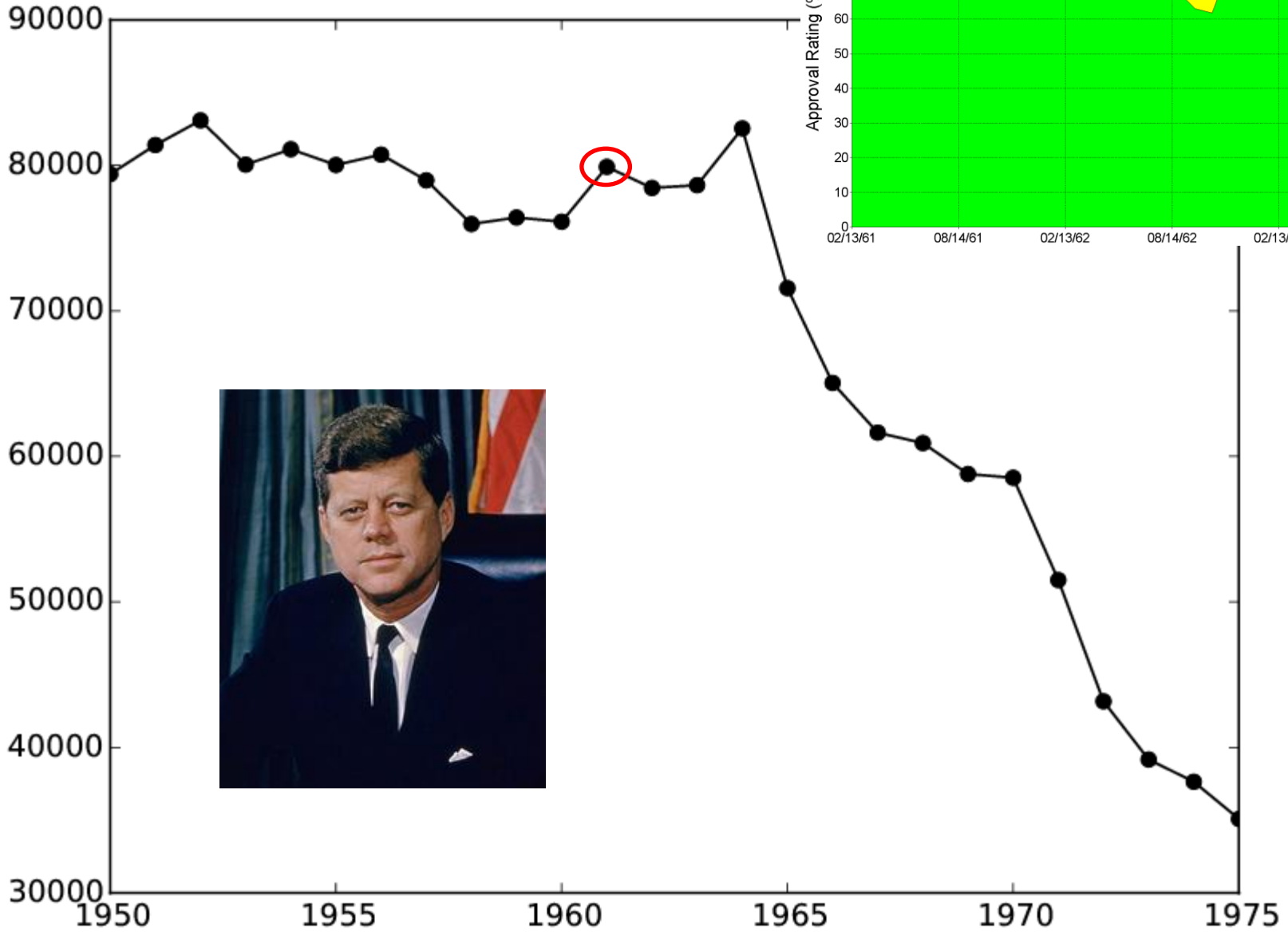
Dwight Eisenhower, 1953-61



John F. Kennedy, 1961-63



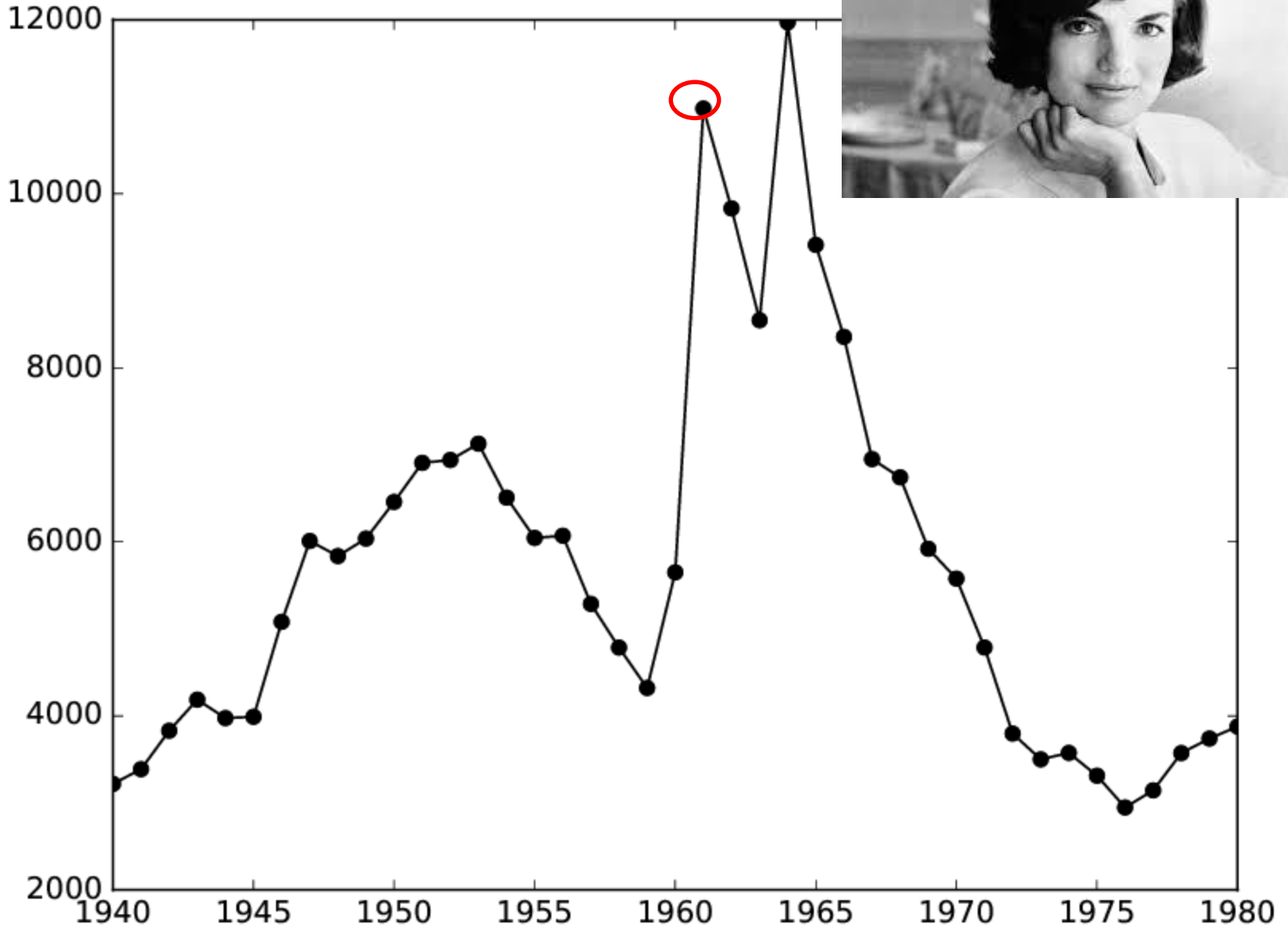
Number of namesakes



Jacqueline

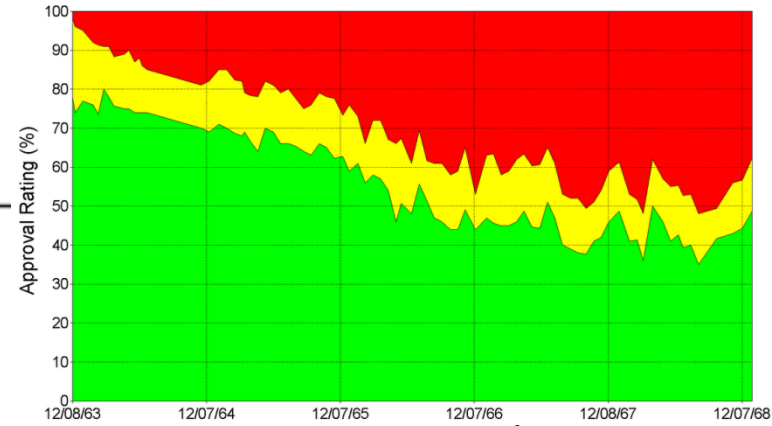


Number of namesakes



Lyndon B. Johnson, 1963-69

Number of namesakes



Johnson Cult

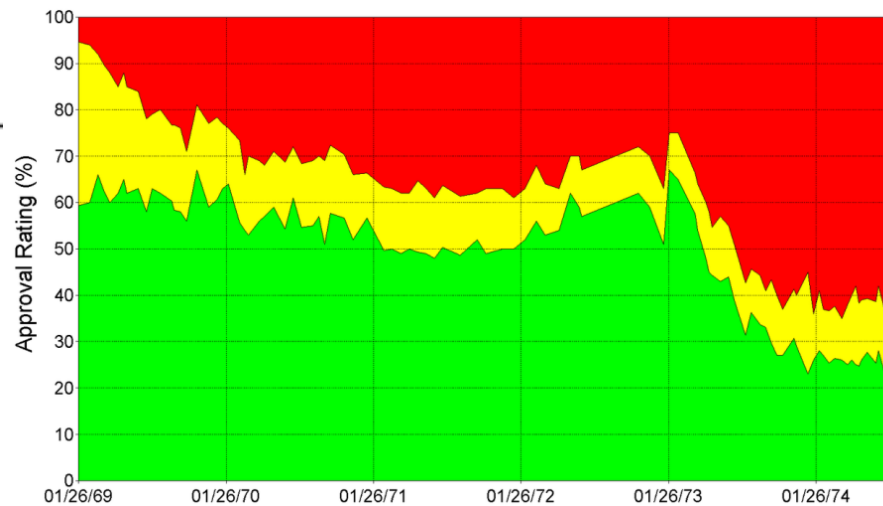
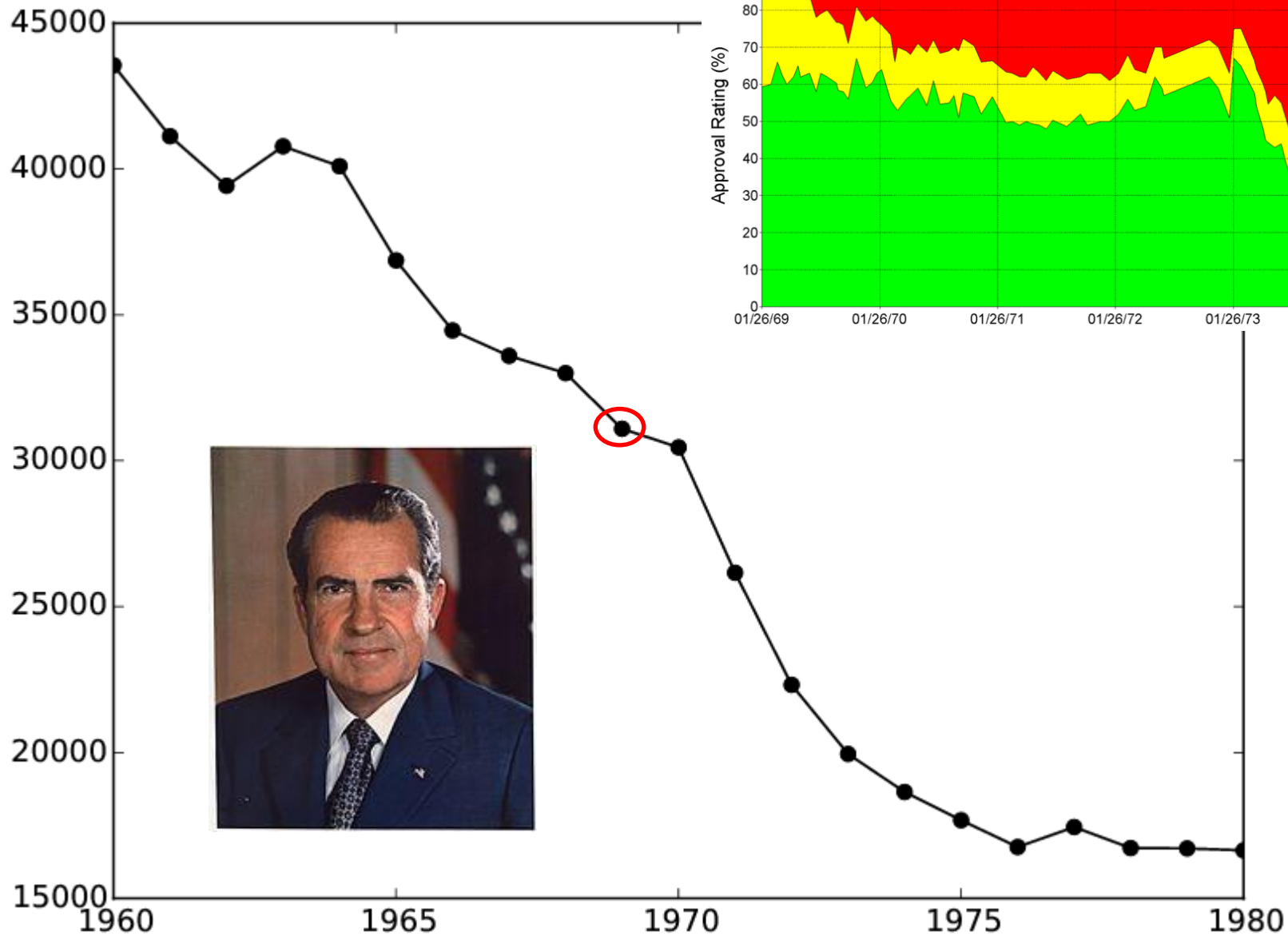
„In February, 1964, the territory of Papua and New Guinea held elections to choose representatives to the newly formed House of Assembly. In New Hanover about half the population of 7,000 refused to follow the prescribed balloting procedure and instead proclaimed, through spokesmen and blackboard announcements, that they all wanted to vote for only one man: Johnson of America.”

[D. K. Billings, Oceania 40/1 (1969) 13]

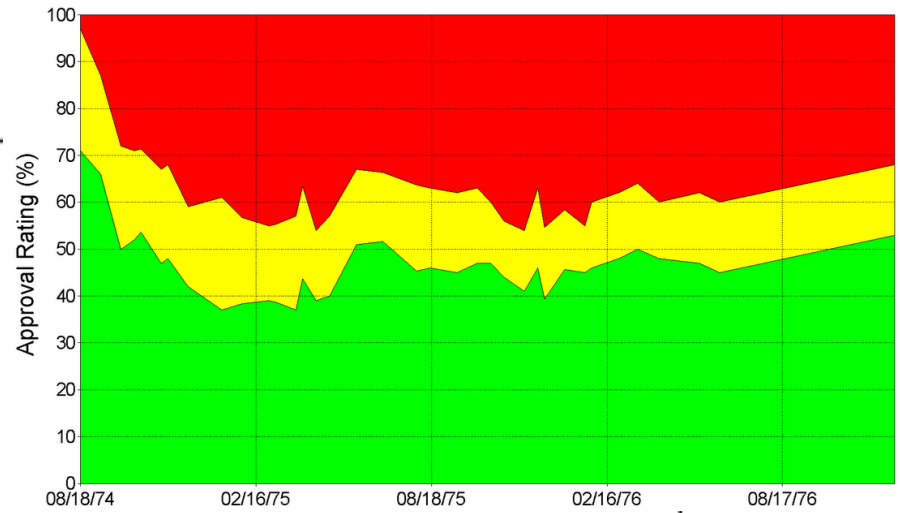
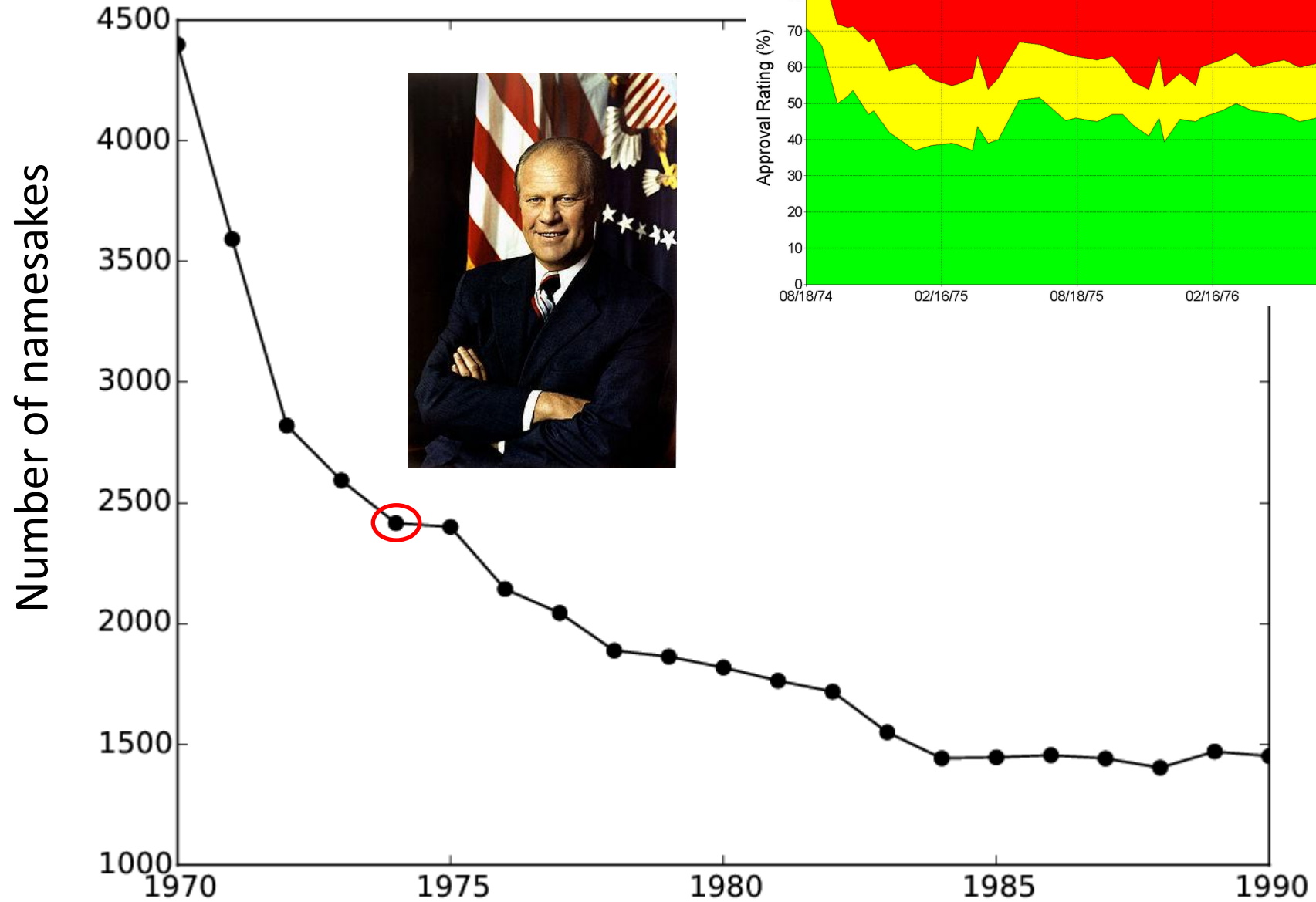


Richard Nixon, 1969-74

Number of namesakes

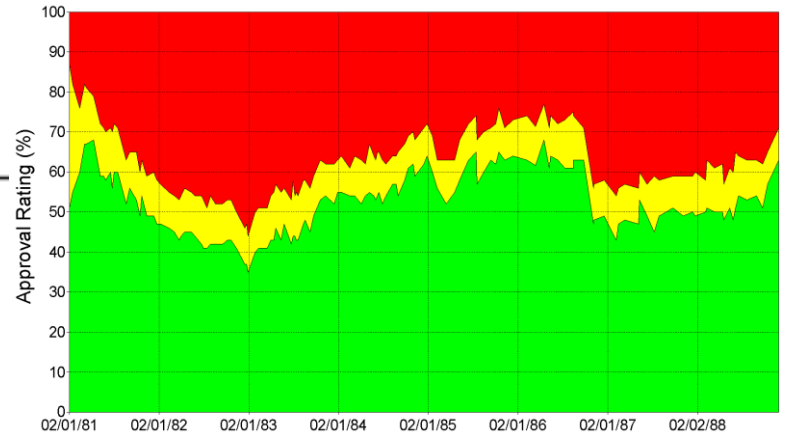
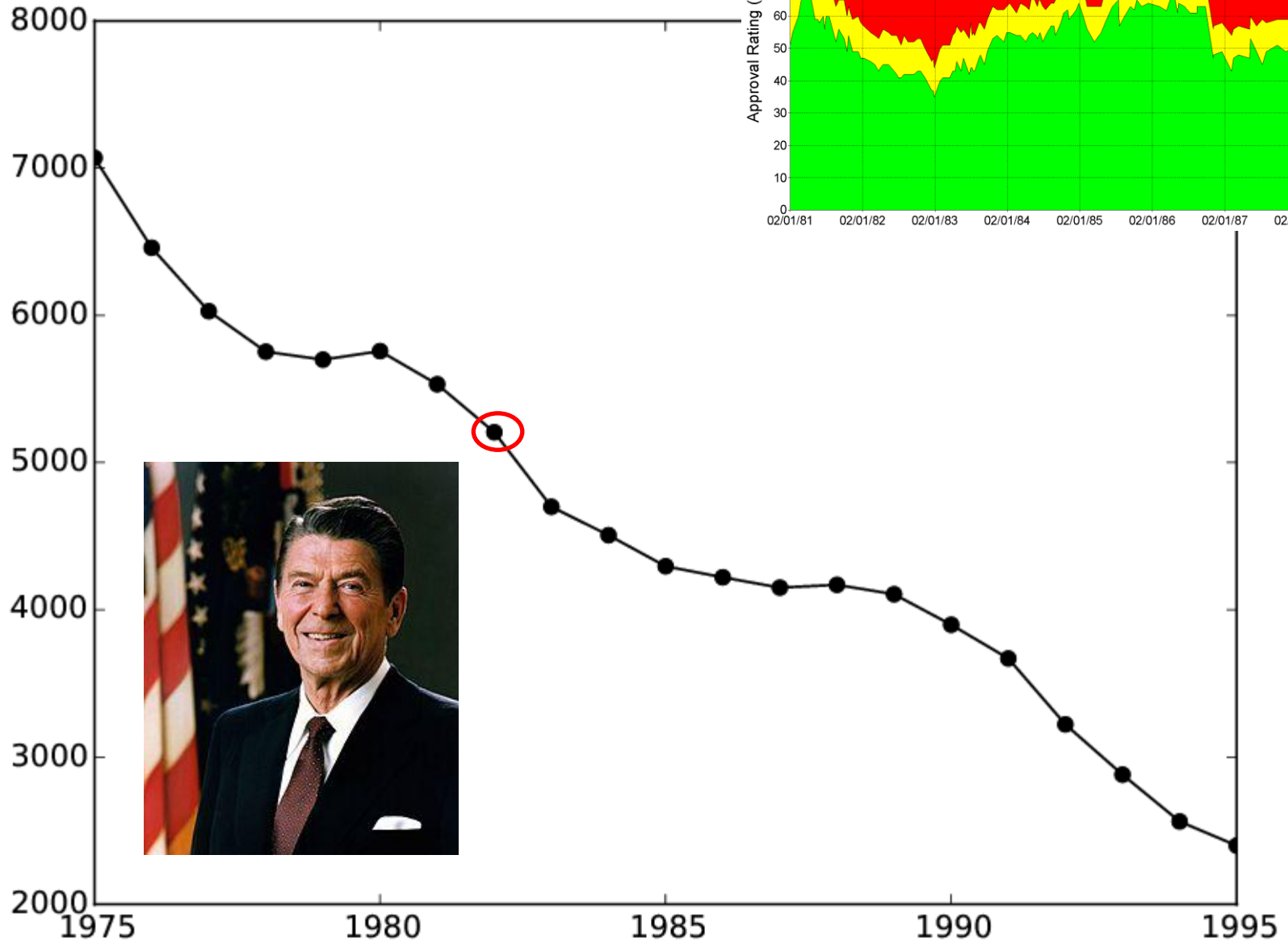


Gerald Ford, 1974-77



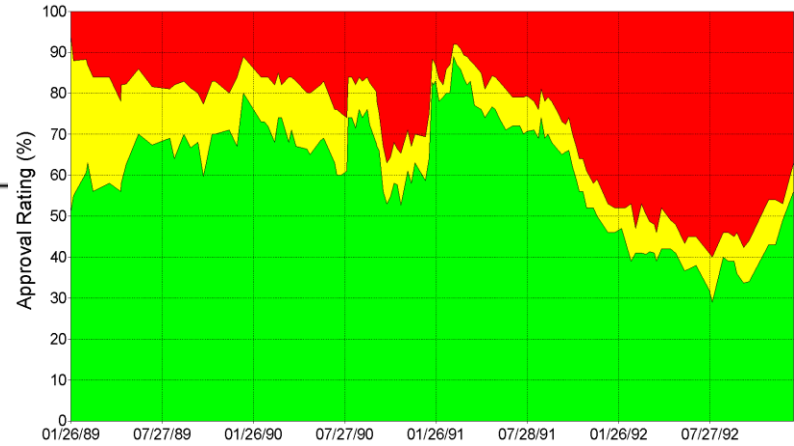
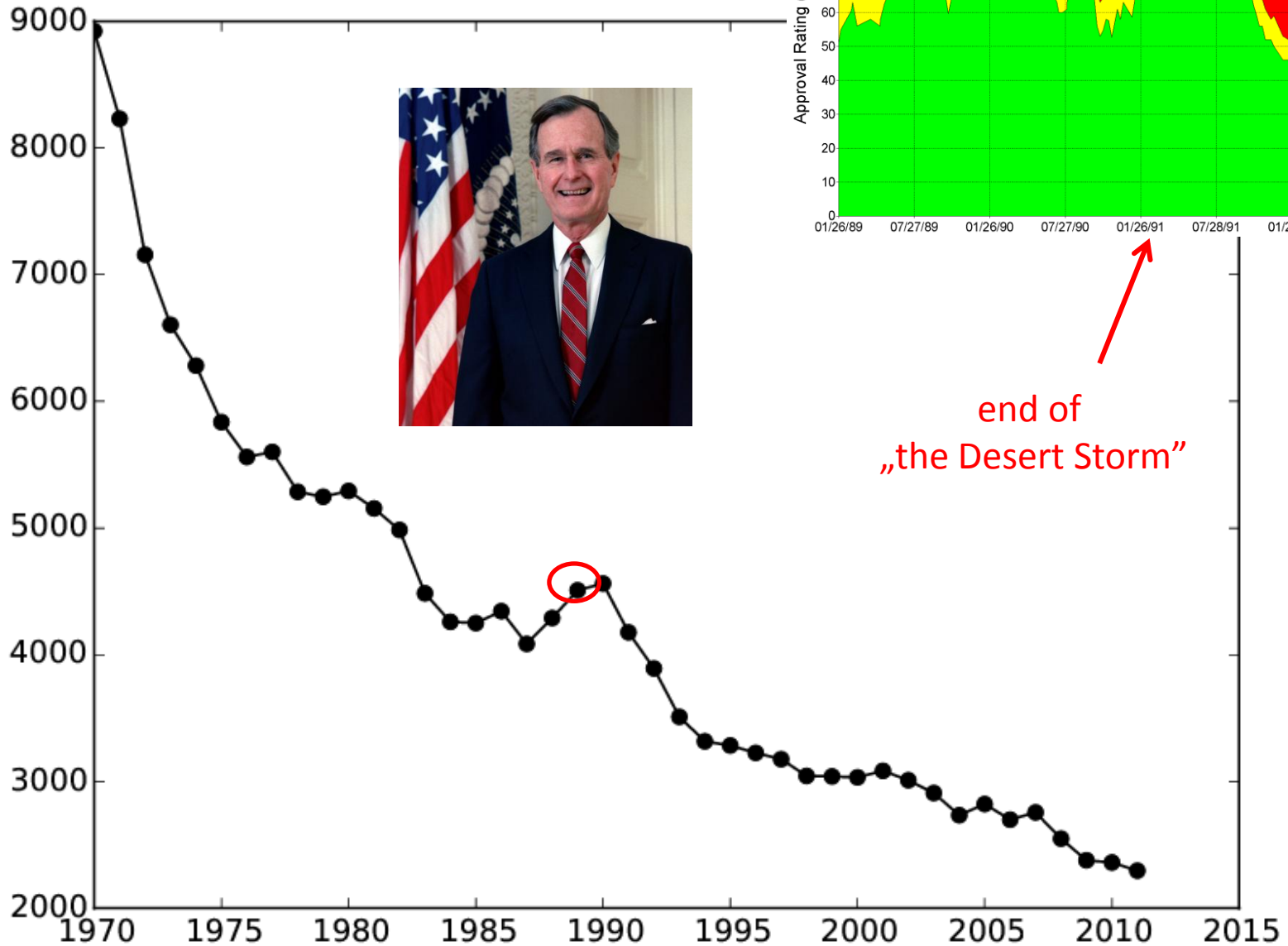
Ronald Reagan, 1981-89

Number of namesakes



George H. W. Bush, 1989-93

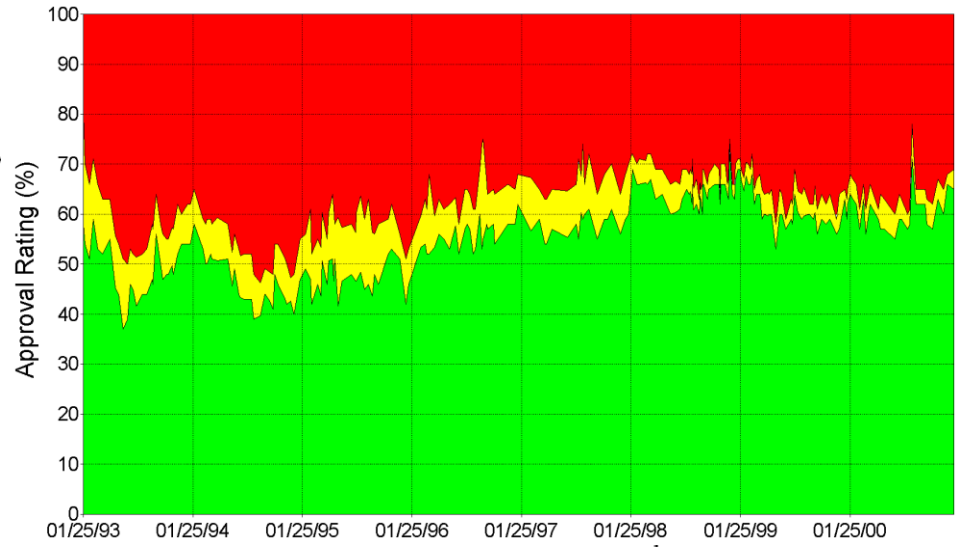
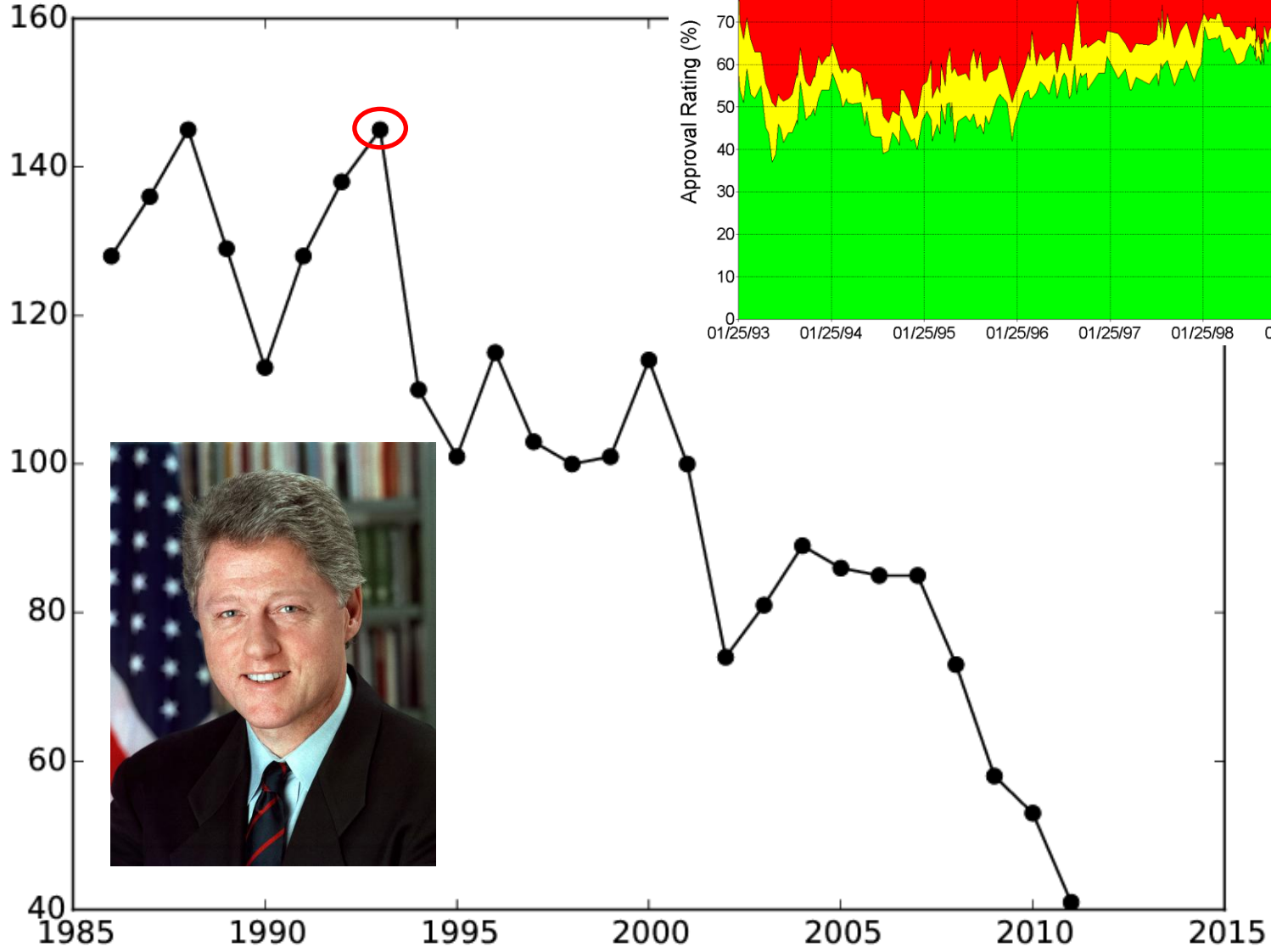
Number of namesakes



end of „the Desert Storm“

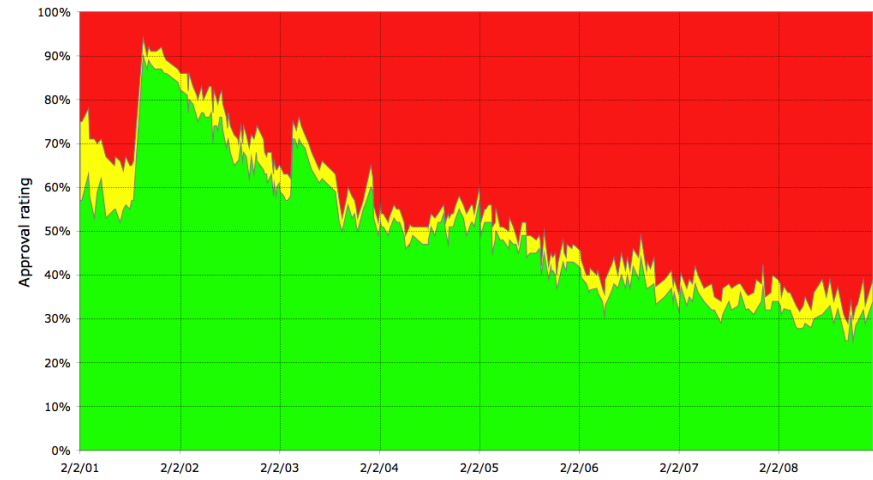
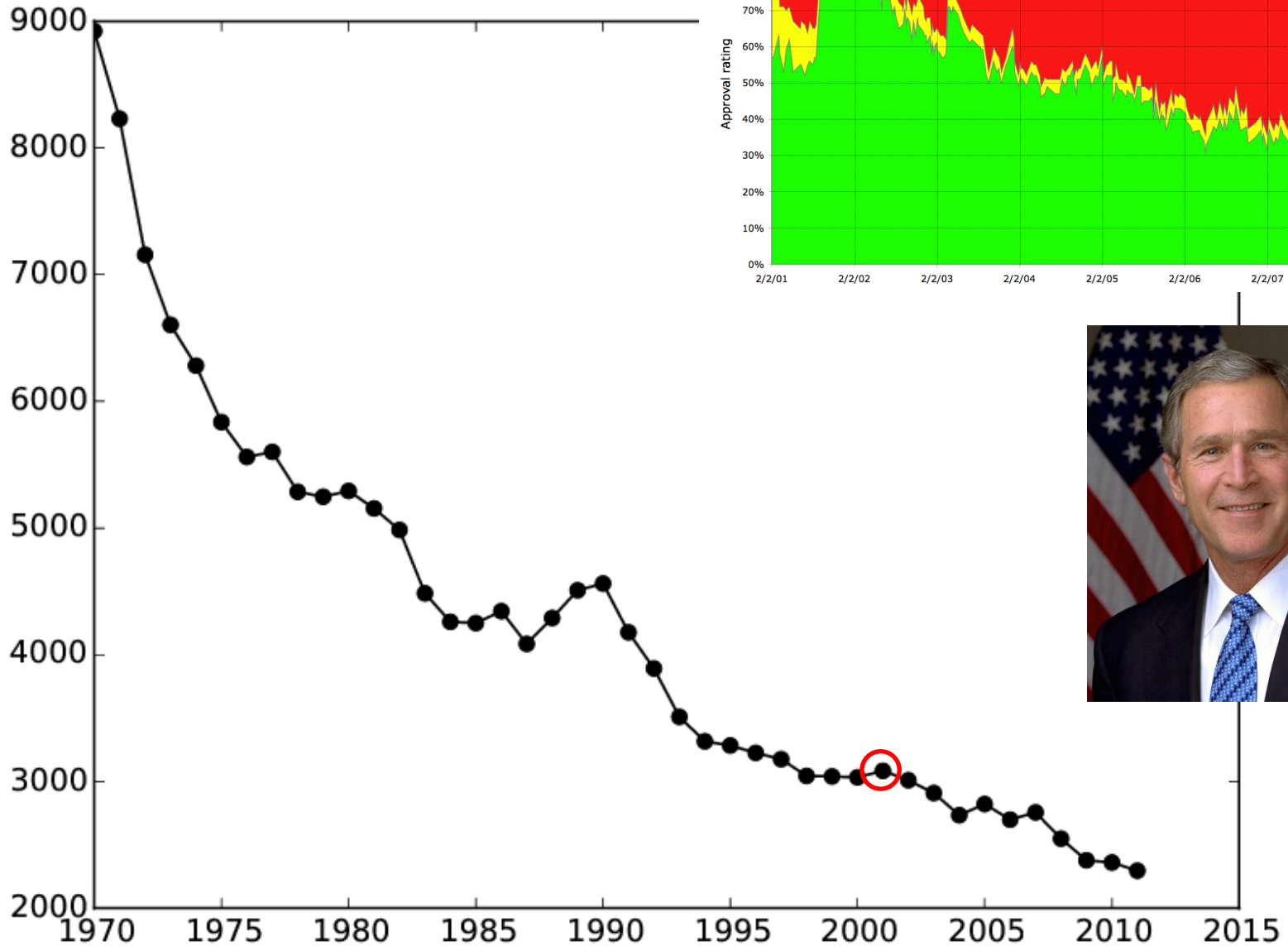
Bill Clinton, 1993-2001

Number of namesakes



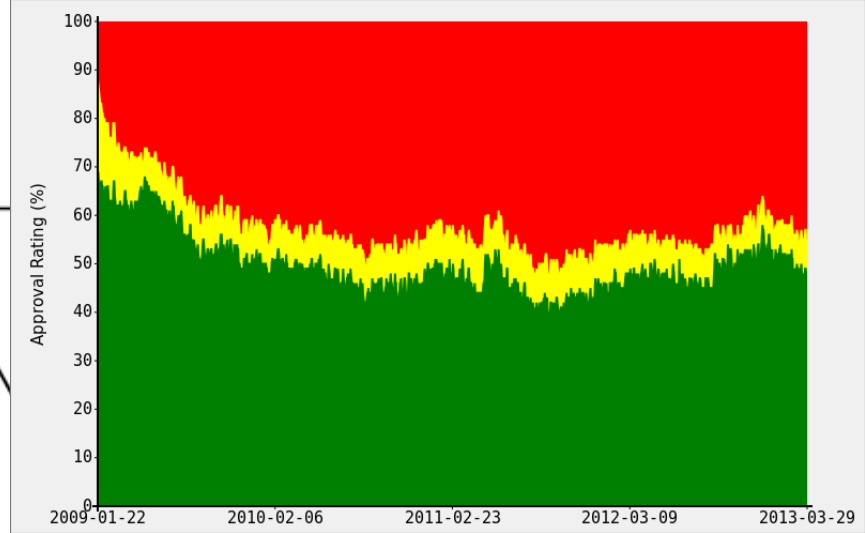
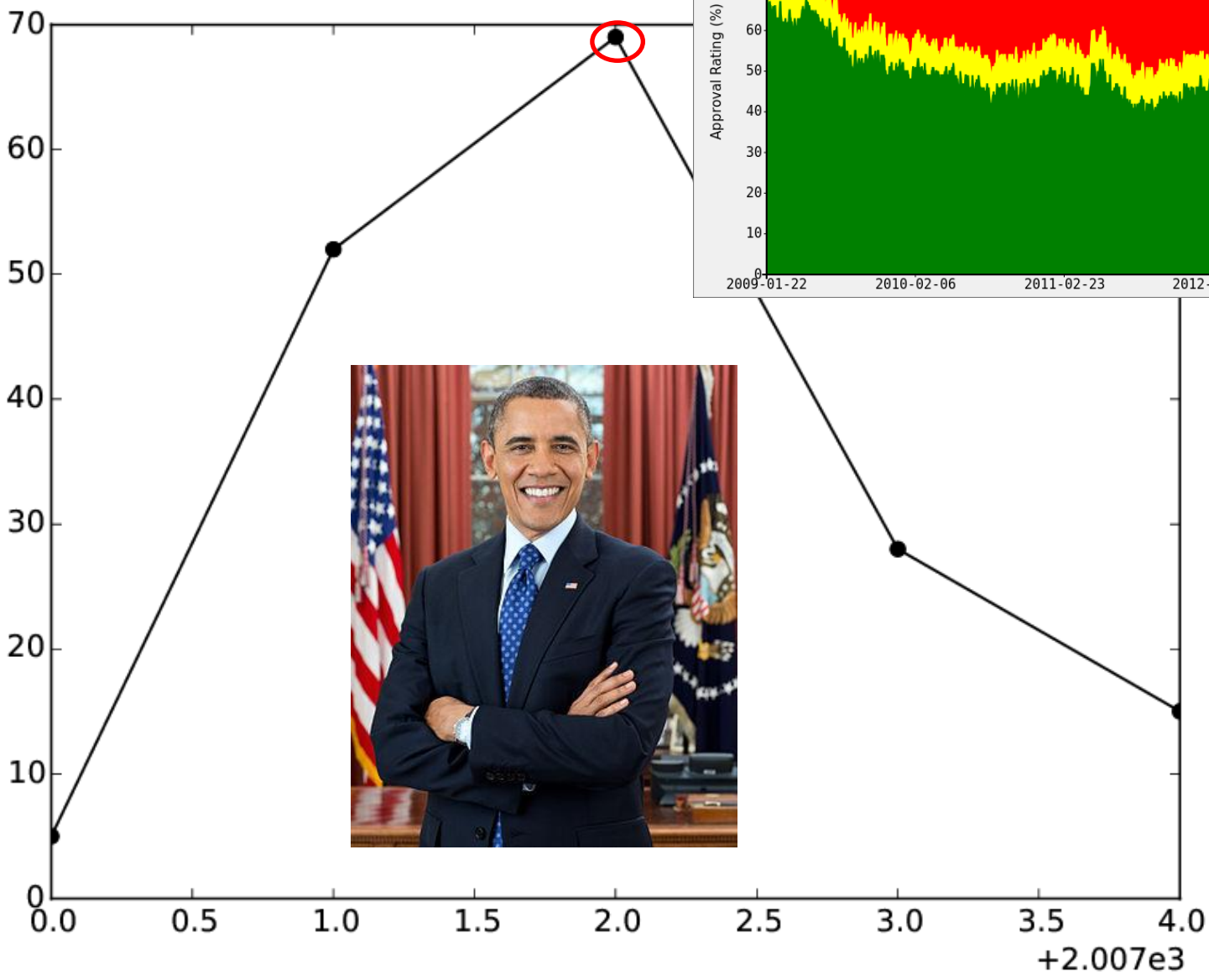
George W. Bush, 2001-09

Number of namesakes

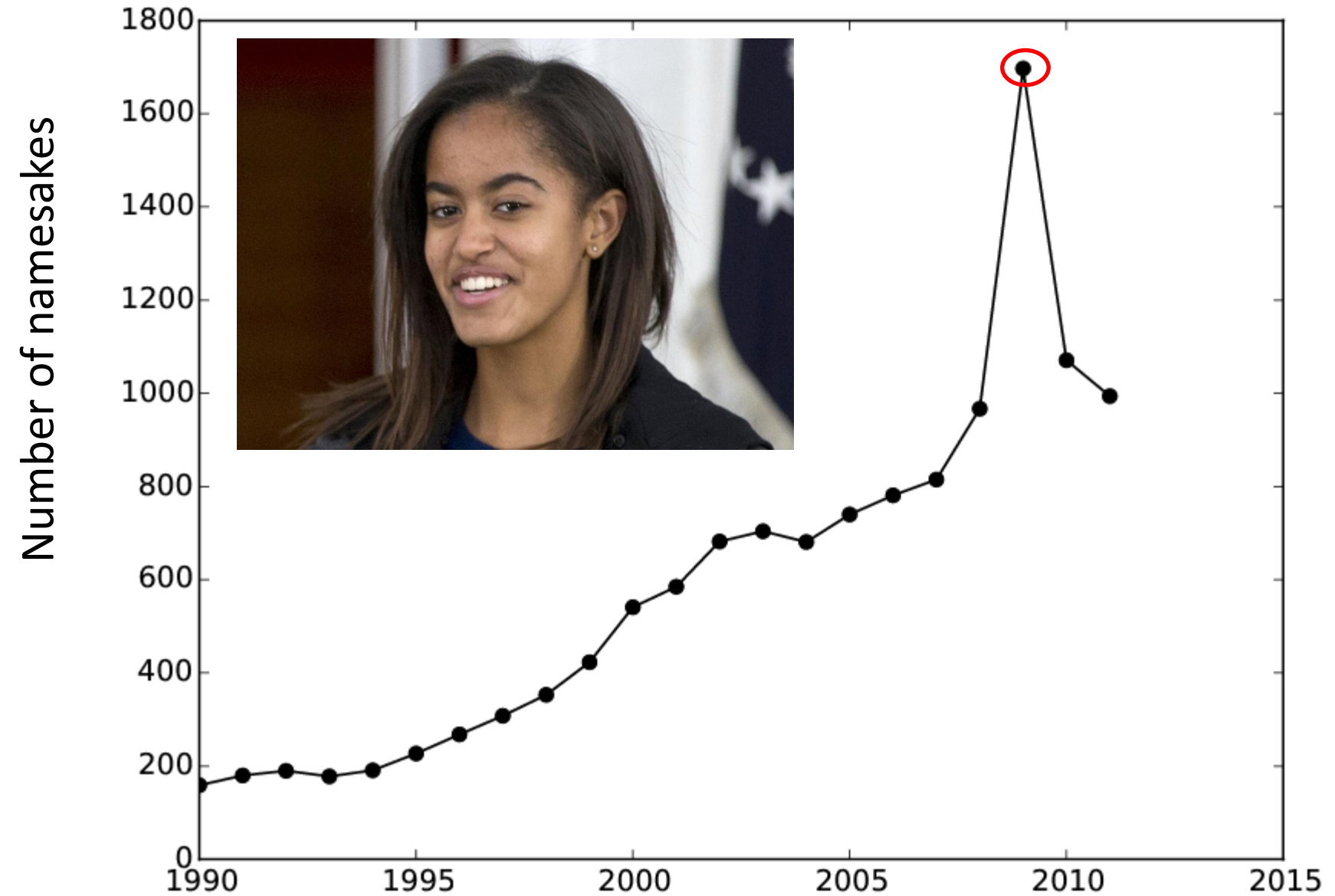


Barack Obama, 2009-

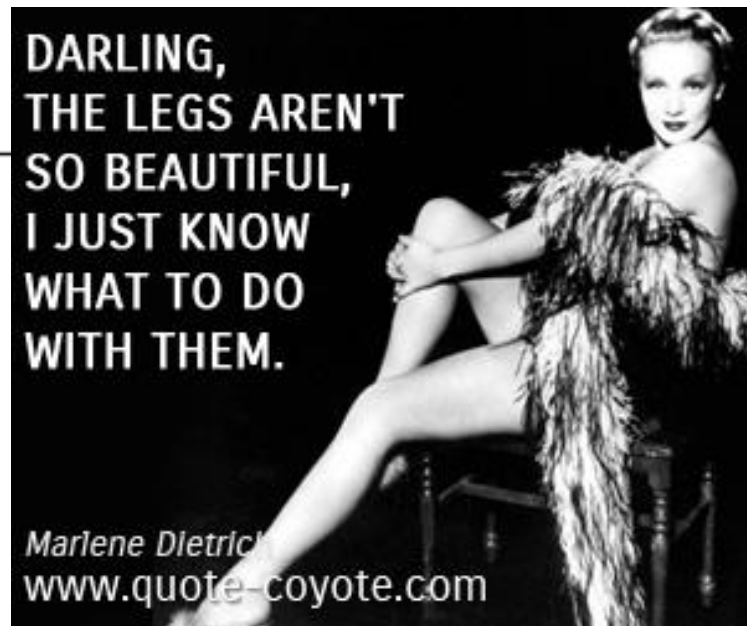
Number of namesakes



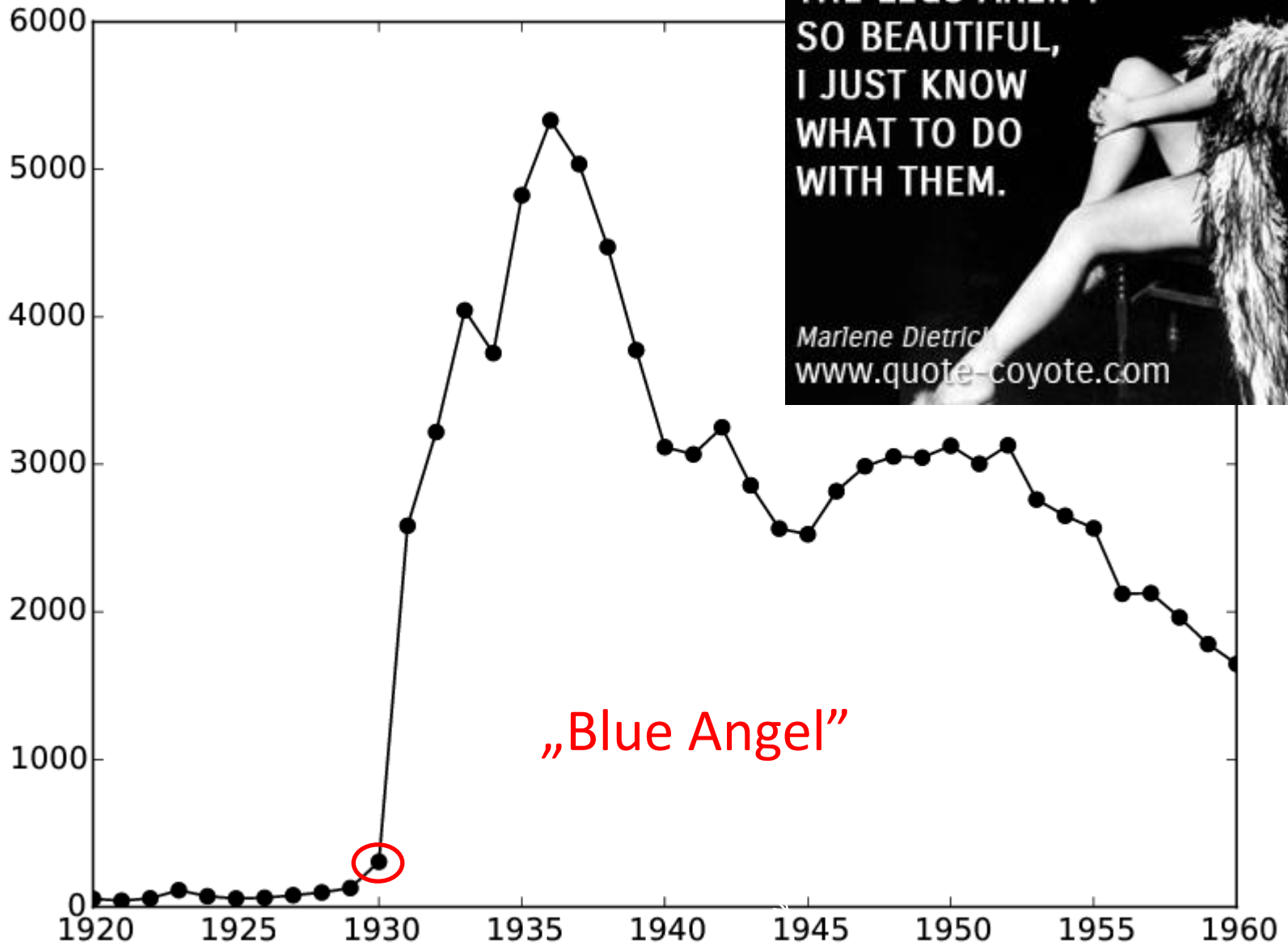
Malia Obama



Marlene Dietrich



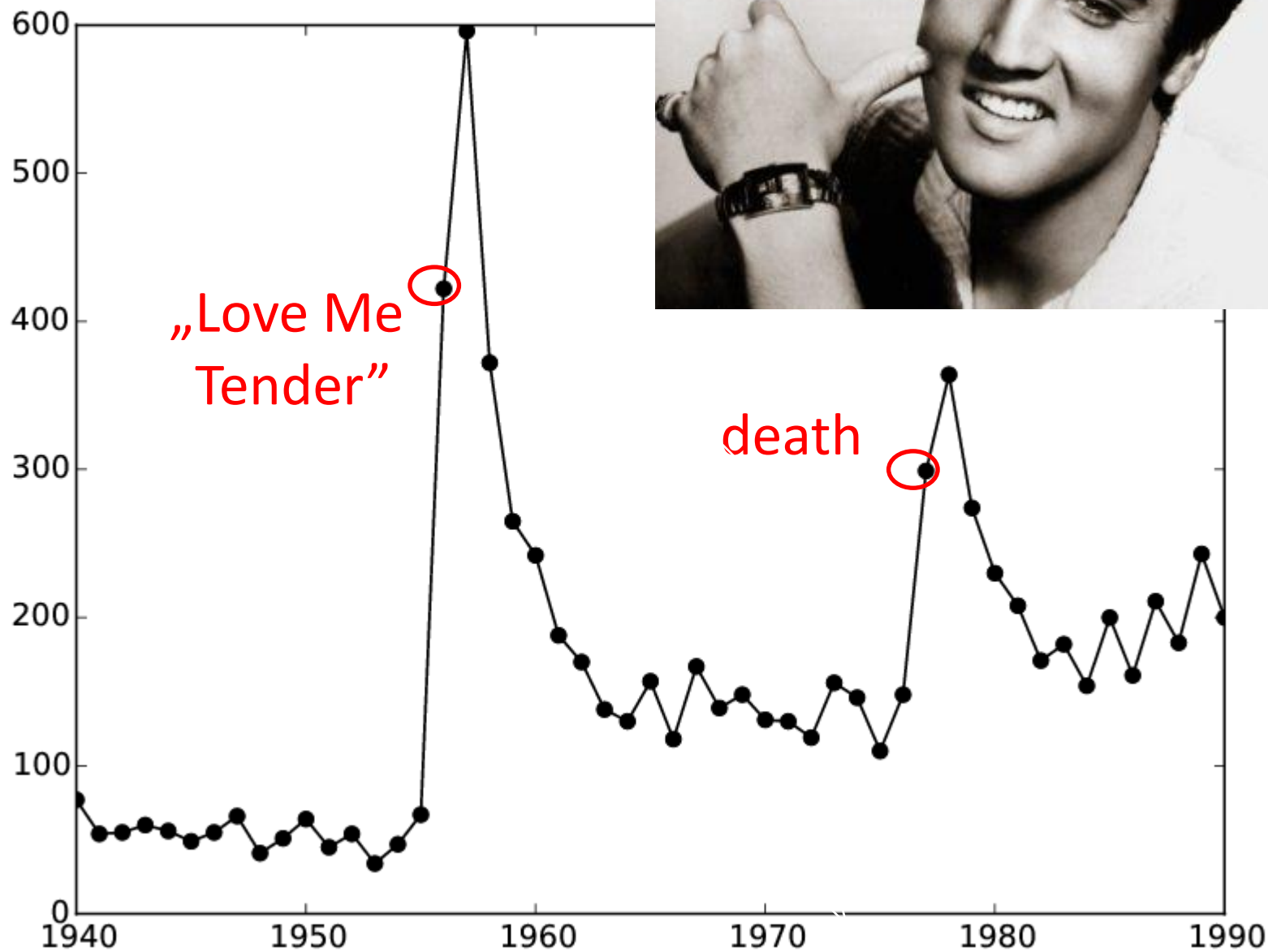
Number of namesakes



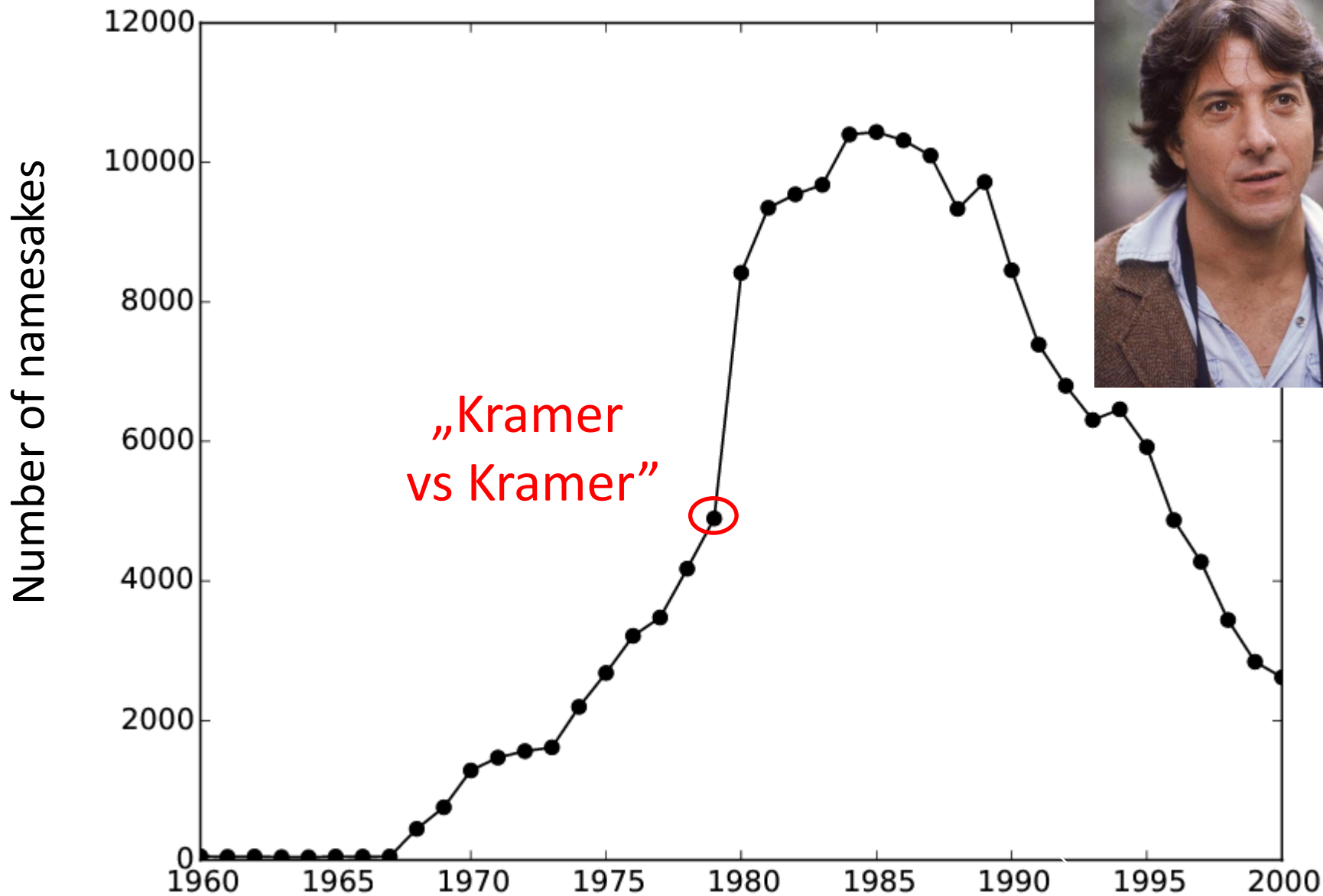
Elvis Presley



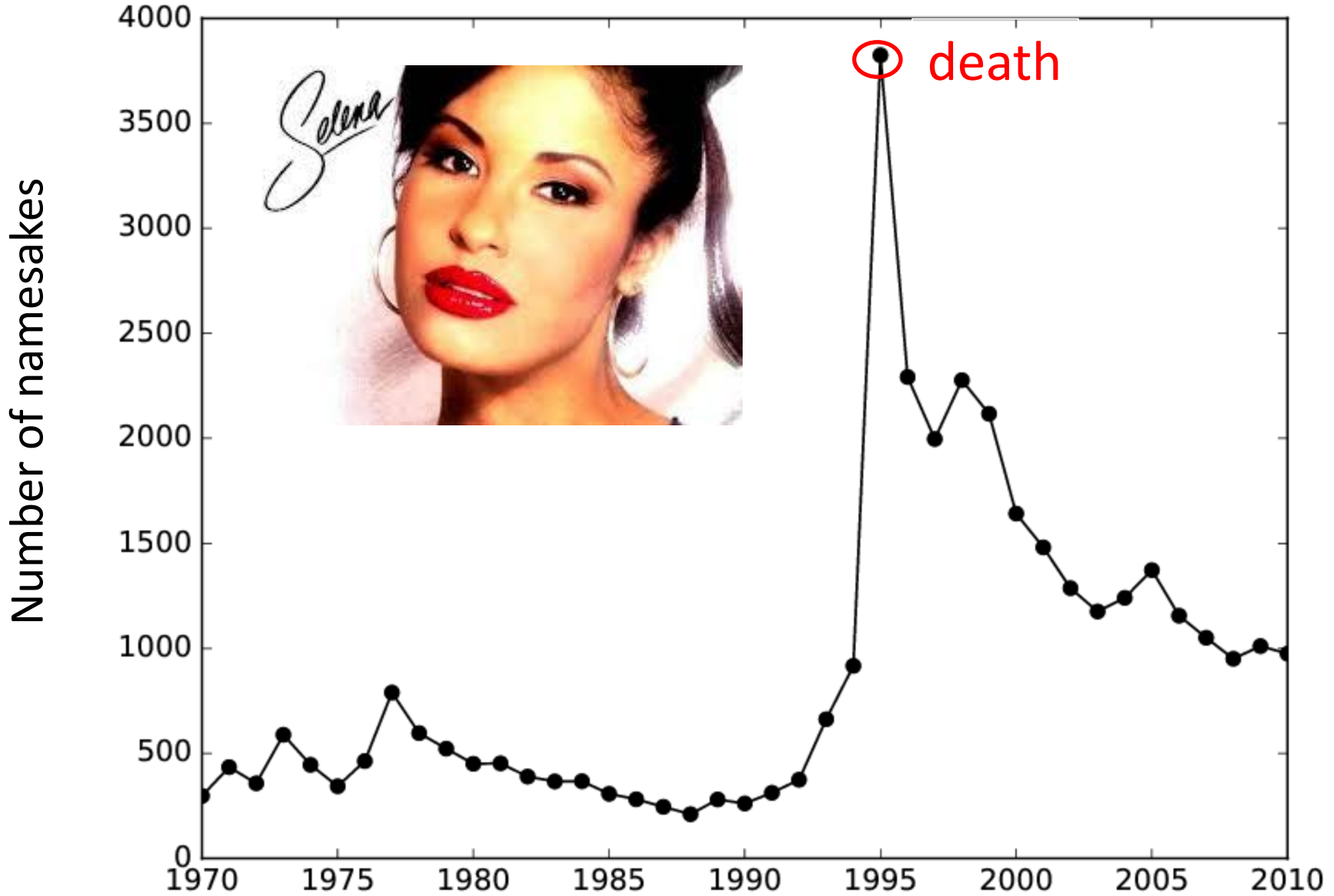
Number of namesakes



Dustin Hoffman



Selena Quintanilla Pérez



Propozycja indeksu

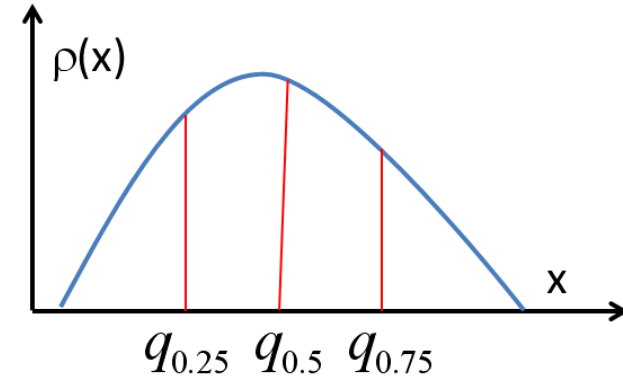
y_0 - dane w roku objęcia urzędu

$$y = \max\{y_{-1}, y_0, y_{+1}\}$$

Dla danych

$$\{y_{-5}, y_{-4}, y_{-3}, y_{-2}, y_{-1}, y_0, y_{+1}\}$$

obliczamy kwartyle q_r



Wskaźnik c definiujemy jako
$$c = \frac{y - q_{0.5}}{q_{0.75} - q_{0.25}}$$

Przykład: Lyndon B. Johnson

Dane $\{y_i\}$:

1958; 162

1959; 192

1960; 246

1961; 207

1962; 148

1963; 177

1964; 512

Dane uporządkowane $\{z_i\}$:

148

162

177

192

207

246

512

$$y = \max\{148, 177, 512\} = 512$$

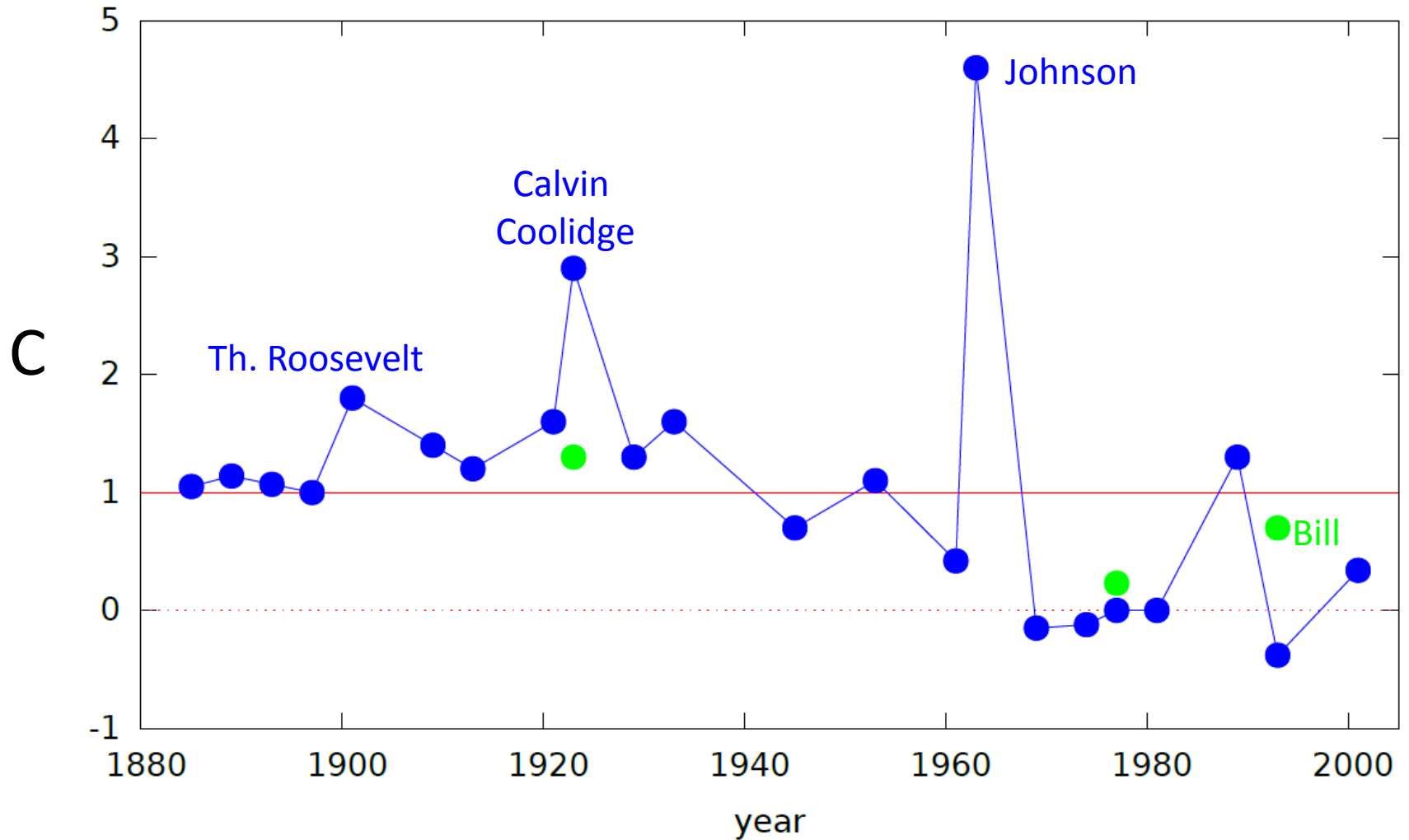
$$q_{0.5} = z_4 = 192$$

$$q_{0.25} = (3z_2 + z_3) / 4 = 166$$

$$q_{0.75} = (z_5 + 3z_6) / 4 = 236$$

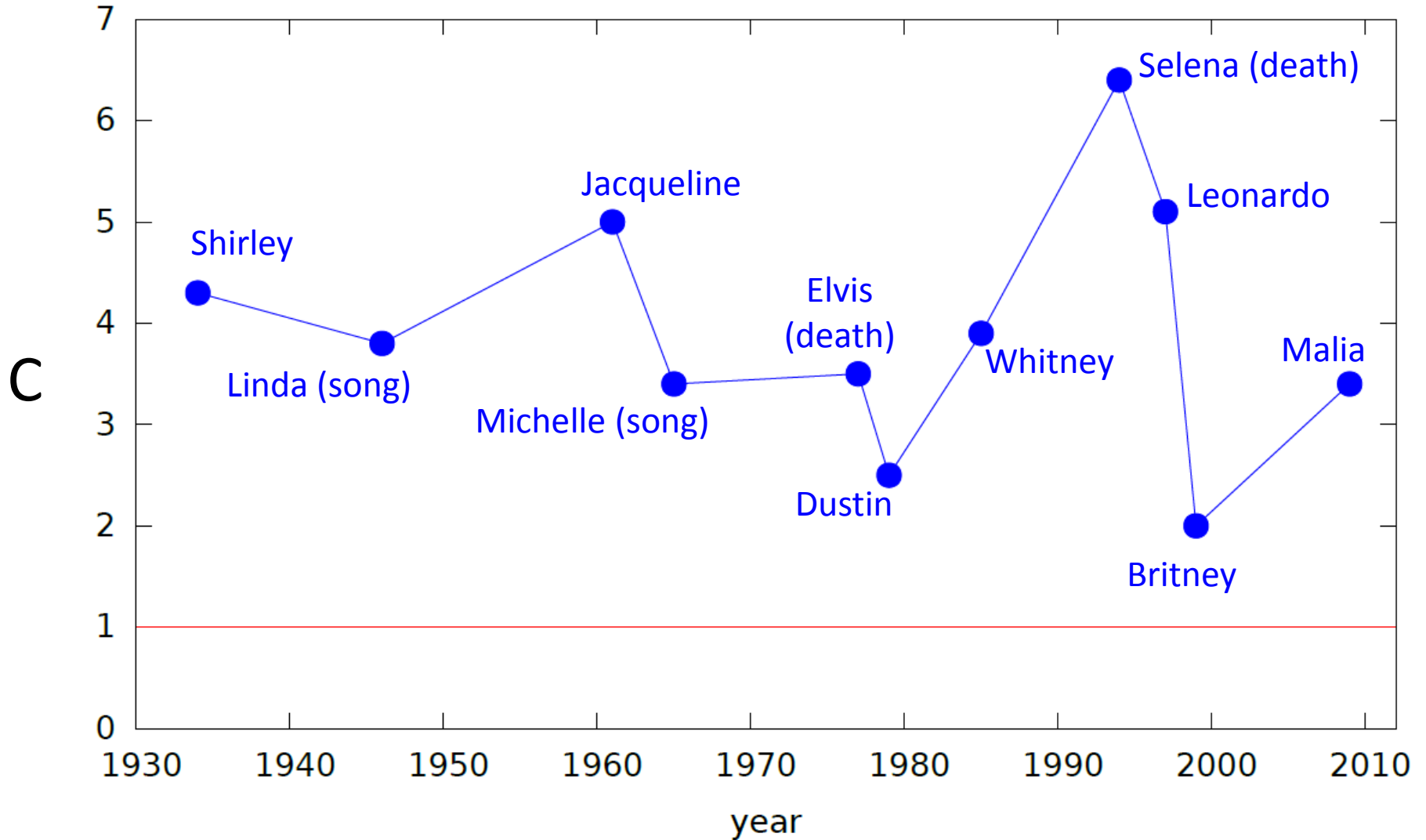
$$c = \frac{512 - 192}{236 - 166} = 4.6$$

Wyniki – prezydenci

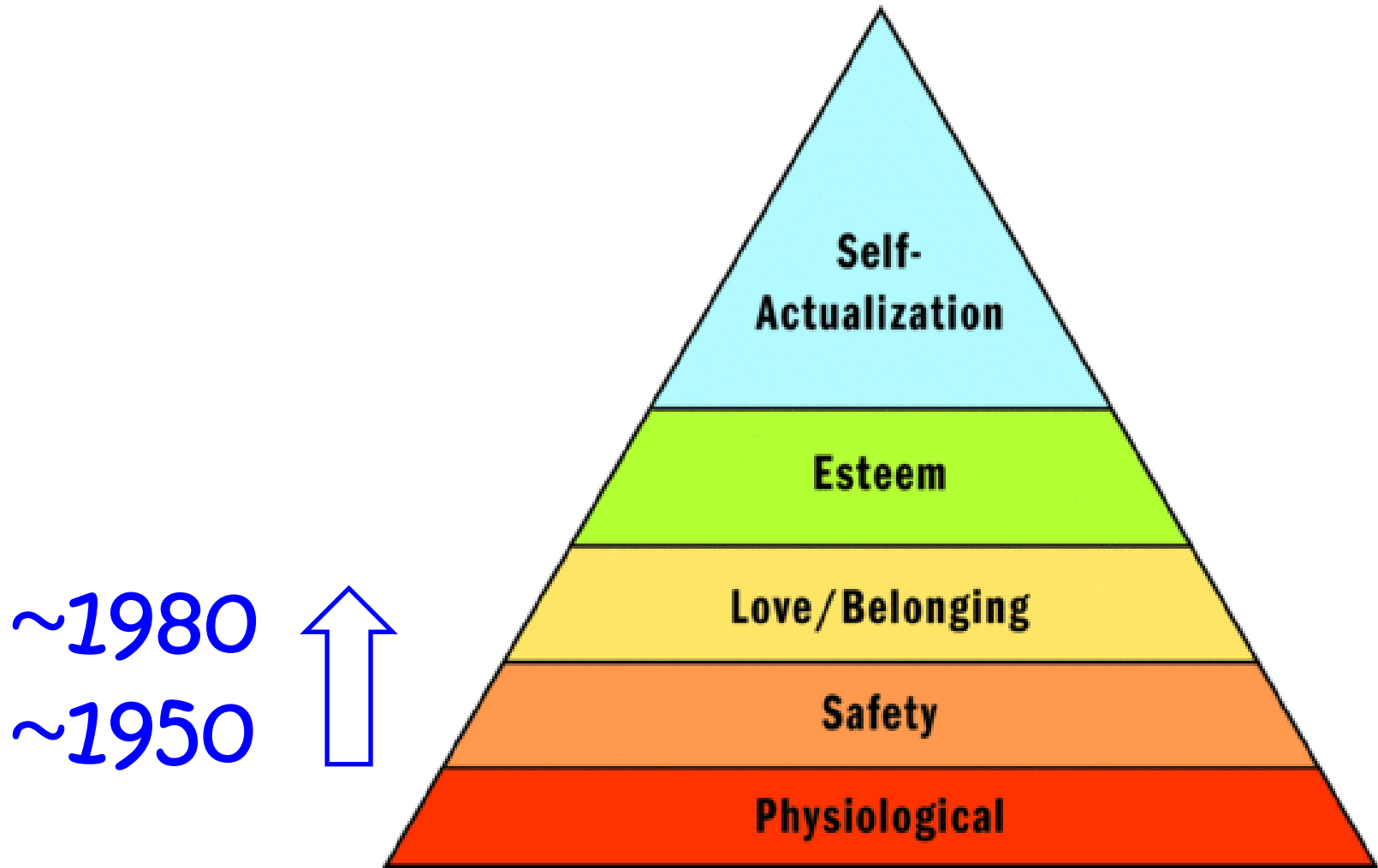


● Imiona przyjęte jak www.whitehouse.gov/1600/Presidents

Wyniki- inni



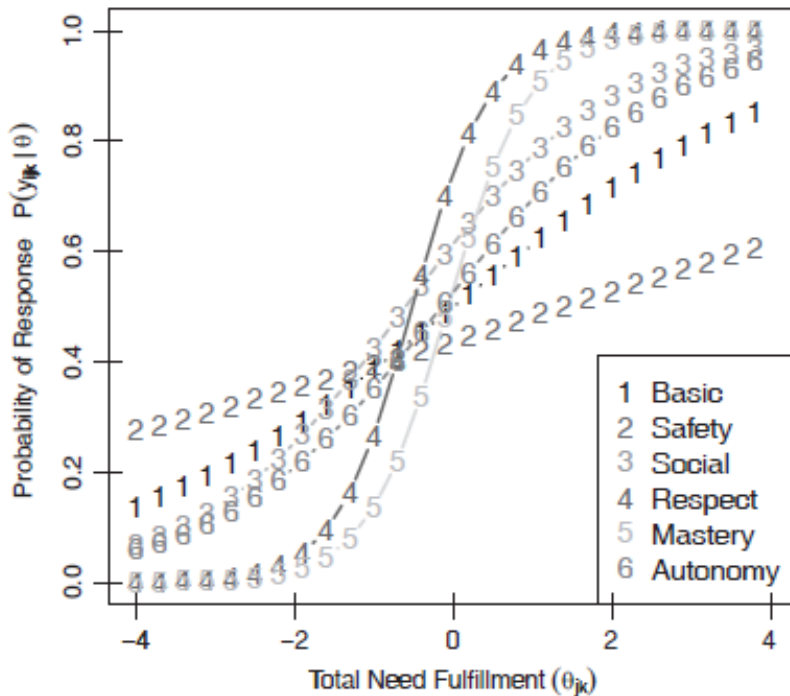
Propozycja interpretacji :



[A. H. Maslow, *A theory of human motivation*, *Psychol. Review* 50 (1943) 370]

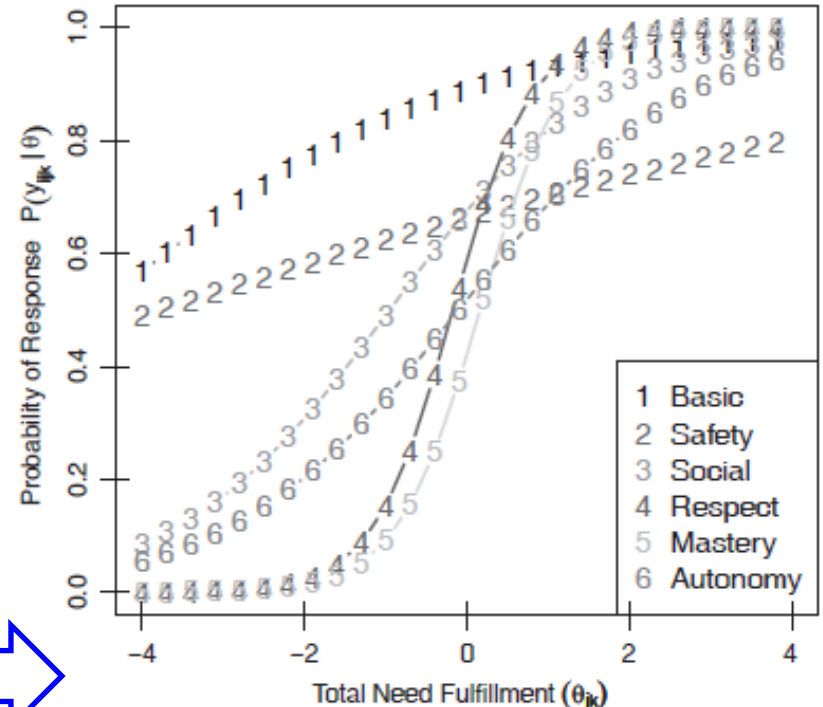
Założmy że przyszłość jest lepsza:

Kraje gdzie potrzeby są zaspokojone gorzej



Indywidualny stopień zaspokojenia potrzeb

Kraje gdzie potrzeby są zaspokojone lepiej



Indywidualny stopień zaspokojenia potrzeb

[L. Tay, E. Diener, *Needs and Subjective Well-Being around the world*, *J. of Personality and Social Psychology* 101 (2011) 354.]

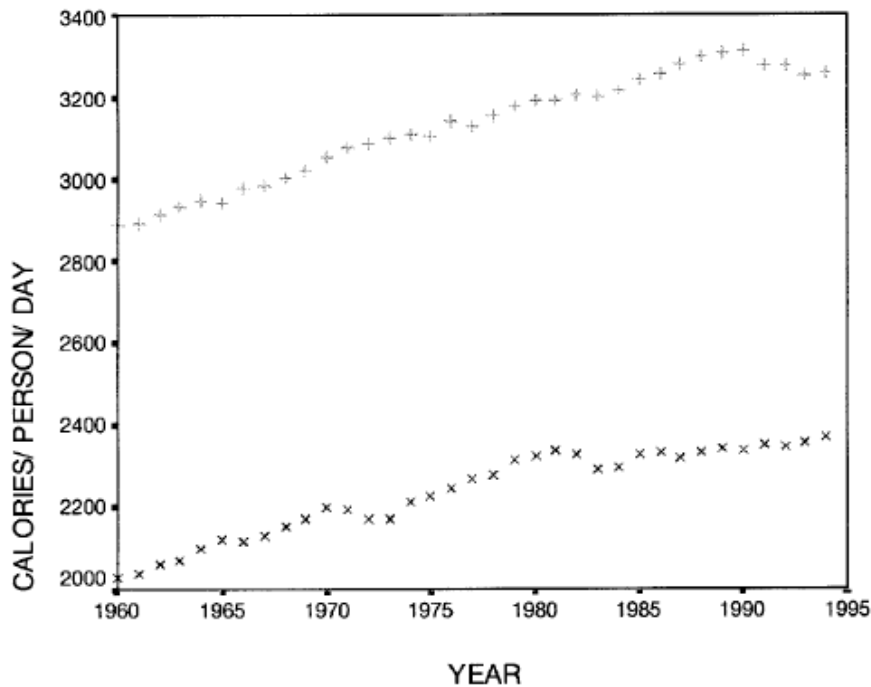


Figure 2. Calories per person per day since 1960, by top 50% (plotted as "+") and bottom 50% (plotted as "x") of nations.

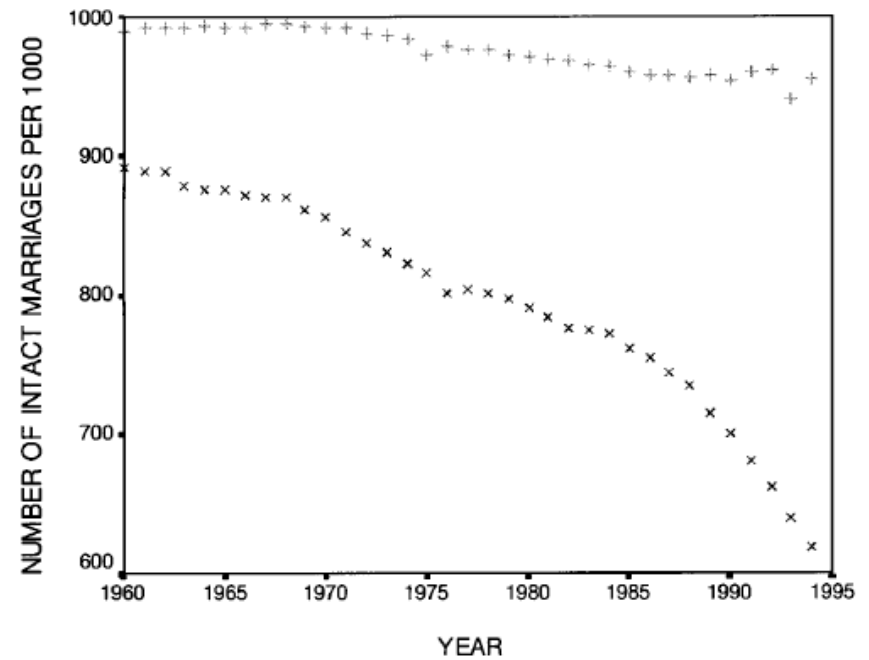


Figure 7. Number of intact marriages per 1000 population (1-Divorce Rate) since 1960, by top 50% (plotted as "+") and bottom 50% (plotted as "x") of nations.

CZY to jest tak?

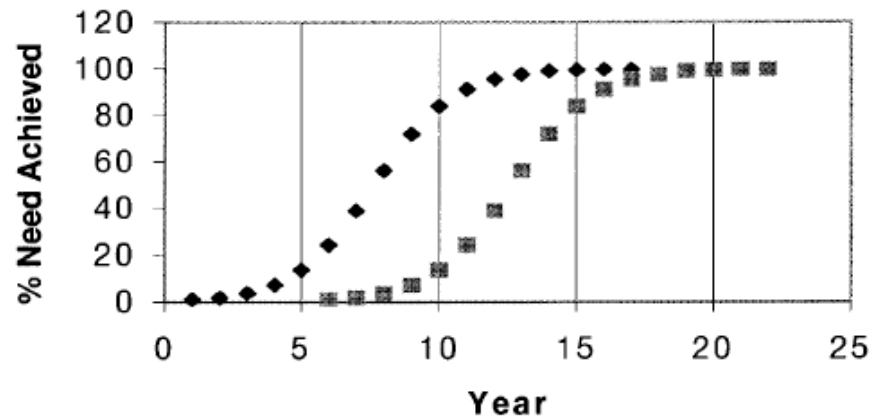
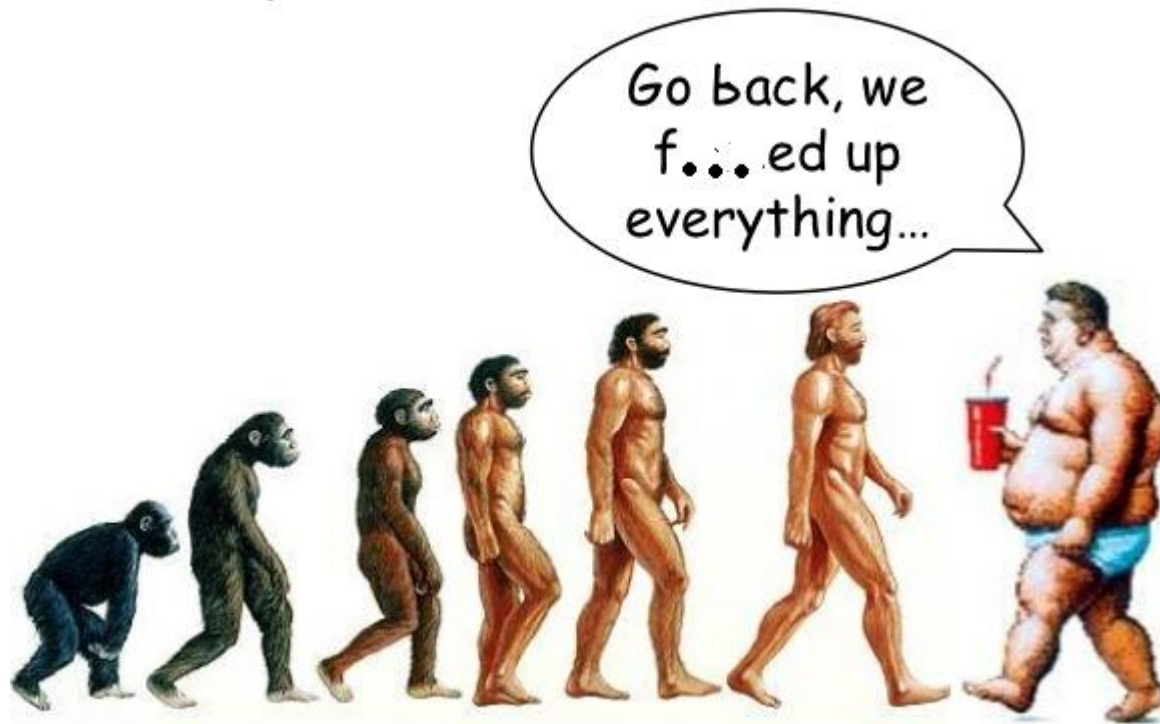


Figure 1. Hypothetical S-shaped curves predicting time path of fulfillment for two needs.

Nie.

Czyżby bezpieczeństwo wyłączało potrzebę bliskości?
Czy potrzeby wyższe są tylko wysublimowanymi formami potrzeb niższych?



podsumowanie 4/4

Zaproponowano wskaźnik do pomiaru wpływu pojawienia się danej osoby na ilość jej imienników.

Dane wskazują, że moda na nadawanie dziecku imienia aktualnego prezydenta USA wygasła w latach 60-tych. Nie odnosi się to do imion przynajmniej niektórych osobistości świata rozrywki.



literatura

M. J. Krawczyk, A. Dydejczyk and K. Kułakowski,
The Simmel effect and babies' names,
Physica A 395 (2014) 384

M. Pomorski, M. J. Krawczyk, K. Kułakowski, J. Kwapien and M. Ausloos,
Inferring cultural regions from correlation networks of given names,
Physica A (2016), w druku

K. Kułakowski, P. Kulczycki, K. Misztal, A. Dydejczyk, P. Gronek, and M. J. Krawczyk,
No more presidents in my family,
w przygotowaniu dla Acta Phys. Pol. A

Dziękuję za uwagę