Hw1 Discussion

CS100

Famous people influence a name's popularity



- The number of male babies named Martin first peaks in 1957, Martin Luther King's first prominent demonstration year.
- The second peak is in 1963, when he delivers his "I Have a Dream" speech

John vs. William



Activity Time!

Why do the plots look similar?





Measuring a name's popularity based on absolute counts may be misleading, because the population changes over time

- Important to **normalize** the data, to understand a name's popularity relative to the size of the total population at the time
- The curves traced by the names John and William likely follow the same arc as the total population from year to year

Many of you noted these kinds of influences!

Normalized John vs. William



"The Most Poisoned Name in U.S. History"

--Hilary Parker

Hilary Parker's famous blog post

- Hilary defines a name to have been poisoned if it has ever declined dramatically and rapidly in popularity
- Obvious examples of poisoned names are:
 - Ebenezer (Scrooge): A Christmas Carol
 - Adolph (Hitler)
- Hilary Parker loves her name, but believes it to be poisoned, so she sets out to investigate

Data

- Scraped from the social security website
- 1000 most popular names in all the years from 1880 to 2011
- 4110 girls names, in total

Measurements

- The relative number of babies named a certain name is a better measure of its popularity than the absolute number
- Absolute change in percentages is also not that informative
 - A change from 15% to 14% is much less interesting than a change from 2% to 1%, but both absolute changes are 1%
- Instead we should look at relative changes in percentages, or the percent change in percentage
- This is called the relative risk

Relative Risk

For example, let's say the percentage of babies named "Jane" is 1% of the population in 1990, and 1.2% of the population in 1991. The relative risk of being named "Jane" in 1991 versus 1990 is 1.2 (that is, it's (1.2/1)=1.2 times as probable).

Findings

- For each name in her data set, and from each year to the next, if the name dropped in relative popularity, Hilary calculated the name's relative risk
- She then pulled out the names with the biggest losses

Hilary is only #6! Not good enough! And the surrounding names are kinda peculiar. Who calls their newborn "Infant"?

Name	Loss (%)	Year
Farrah	78	1978
Dewey	74	1899
Catina	74	1974
Deneen	72	1965
Khadijah	72	1995
Hilary	70	1993
Clementine	69	1881
Katina	69	1974
Renata	69	1981
lesha	69	1992
Minna	68	1883
Ashanti	68	2003
Celestine	67	1881
Infant	67	1991



Percent of baby girls given a name over time for the 14 most poisoned names

Year





Percent of baby girls given a name over time for the 14 most poisoned names, controlling for fads

Year

