Poisoned Names Analysis

Chris Hua

10/31/2016

Motivation

Hilary Parker wrote a fantastic blog/article a few years ago, about how the name Hillary is The Most Poisoned Baby Name in U.S. History. We wanted to replicate her analysis, show off how easy and clean `dplyr` makes this analysis, and show some cool trends visible in the 2015 data.

For this analysis, we used a modified version of Hadley's `babynames` package. Our version is found on Github as well, but is updated to include 2015 data.

Poisoned Names

We borrow the "relative risk" framework that Parker uses. We define relative risk and loss as:

$$risk_{i,t} = p_{i,t}/p_{i,t-1} loss_{i,t} = 1 - risk_{i,t}$$

Where p_t is the proportion of babies born in a year t, with a given name i.

We can do this very easily using the built-in `dplyr` functions to offset each row by 1, so we can examine the changes year to year. We also add a check to ensure that we aren't picking up names dropping in and out of the census observation. We also need to group by name and sex, since some typically female names are given to men occasionally, and vice versa. [Insert Archer gif here]

Note: we choose to perform this analysis using lagged data, rather than leading data, because we want to pick up trends in 2015. Parker's analysis uses a leading approach, but this should yield us identical results except for when a given year is the first or last of that observation period.

```
names_loss <- babynames %>%
    group_by(name, sex) %>%
    arrange(year) %>%
    mutate(loss = 1 - (prop / lag(prop, 1))) %>%
    filter(year == lag(year, 1) + 1) %>%
    arrange(-loss) %>%
    ungroup
```

Then, let's only look at the female names:

```
names_loss %>%
    filter(sex == "F", n > 100) %>%
    head(10) %>%
    knitr::kable(digits = 2)
```

year	sex	name	n	prop	loss
1978	F	Farrah	332	0	0.78
1995	F	Kadijah	119	0	0.75
2002	F	Citlalli	103	0	0.75
1974	F	Catina	328	0	0.74
1990	F	Stephani	173	0	0.74
1995	F	Khadijah	438	0	0.72

year	sex	name	n	prop	loss
1965	F	Deneen	421	0	0.72
1984	F	Kaleena	101	0	0.71
2015	F	Isis	117	0	0.70
1993	F	Hilary	343	0	0.70

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
Iesha	69	1992
Minna	68	1883
Ashanti	68	2003
Celestine	67	1881
Infant	67	1991

This looks pretty similar to .

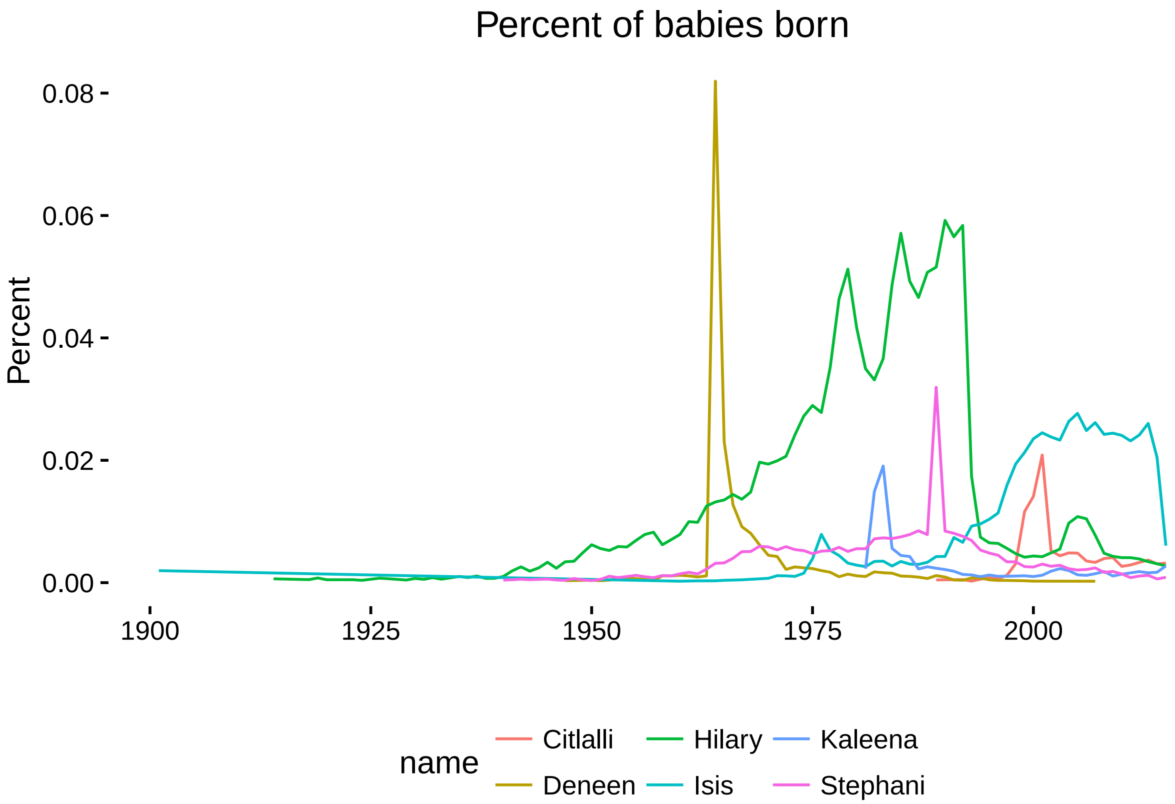
As with Parker's analysis, we will cut out a few "flash in the pan" names. I removed the ones that she covered in her analysis, but there's a few new ones, and I don't have great explanations for why these names could/should be treated as flash in the pan names.

```
flash_names <- c("Dewey", "Katina", "Catina", "Farrah", "Renata", "Infant",
                 "Iesha", "Kadijah", "Khadijah")
big_drops <- names_loss %>%
    filter(sex == "F", n > 100) %>%
```

```
    head(10) %>%
    filter(!(name %in% flash_names))
big_drops %>% knitr::kable()
```

year	sex	name	n	prop	loss
2002	F	Citlalli	103	0.0000522	0.7498452
1990	F	Stephani	173	0.0000842	0.7361881
1965	F	Deneen	421	0.0002304	0.7188817
1984	F	Kaleena	101	0.0000560	0.7060266
2015	F	Isis	117	0.0000605	0.7041397
1993	F	Hilary	343	0.0001740	0.7019090

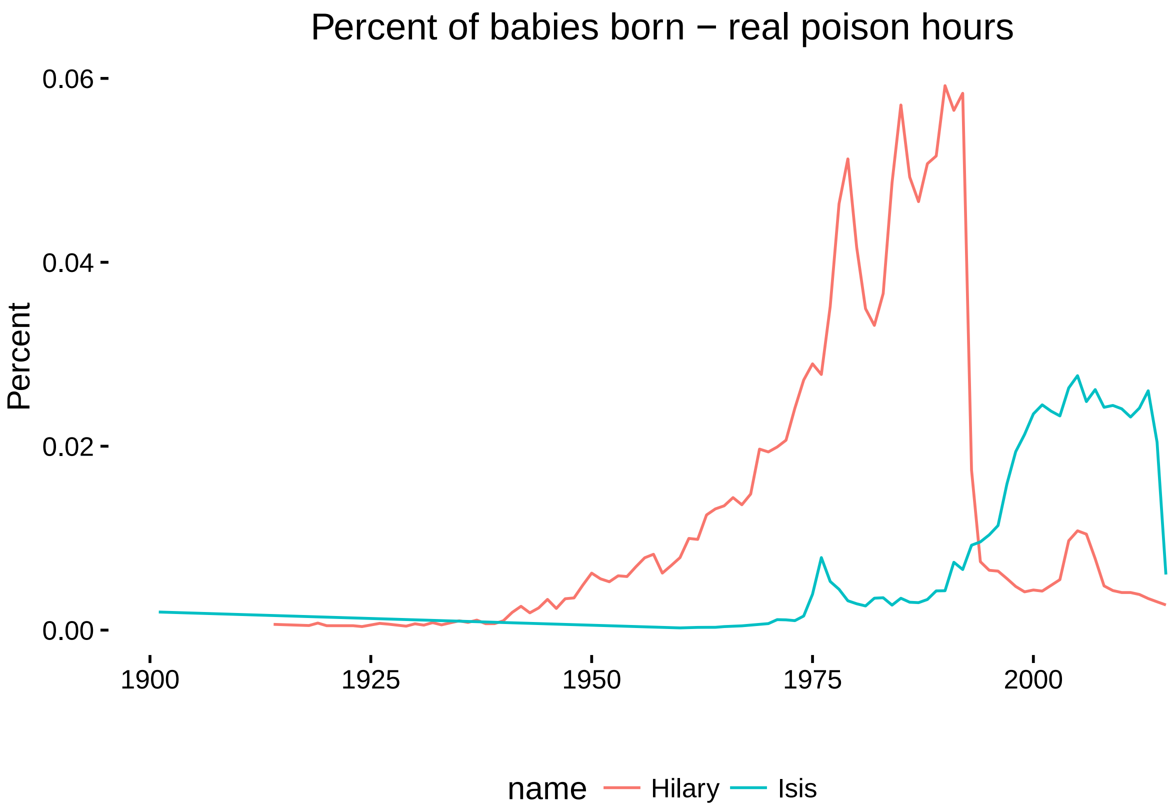
Then, we will do the same process of plotting the names and seeing trends - do these names have any sticking power?



It looks like my Google skills are lacking - Deneen, Kaleena, Citali, and Stephani each have huge spikes with little other sticking power. Stefani may be the artifact of a typo mess. That leaves…

The Syria Connection

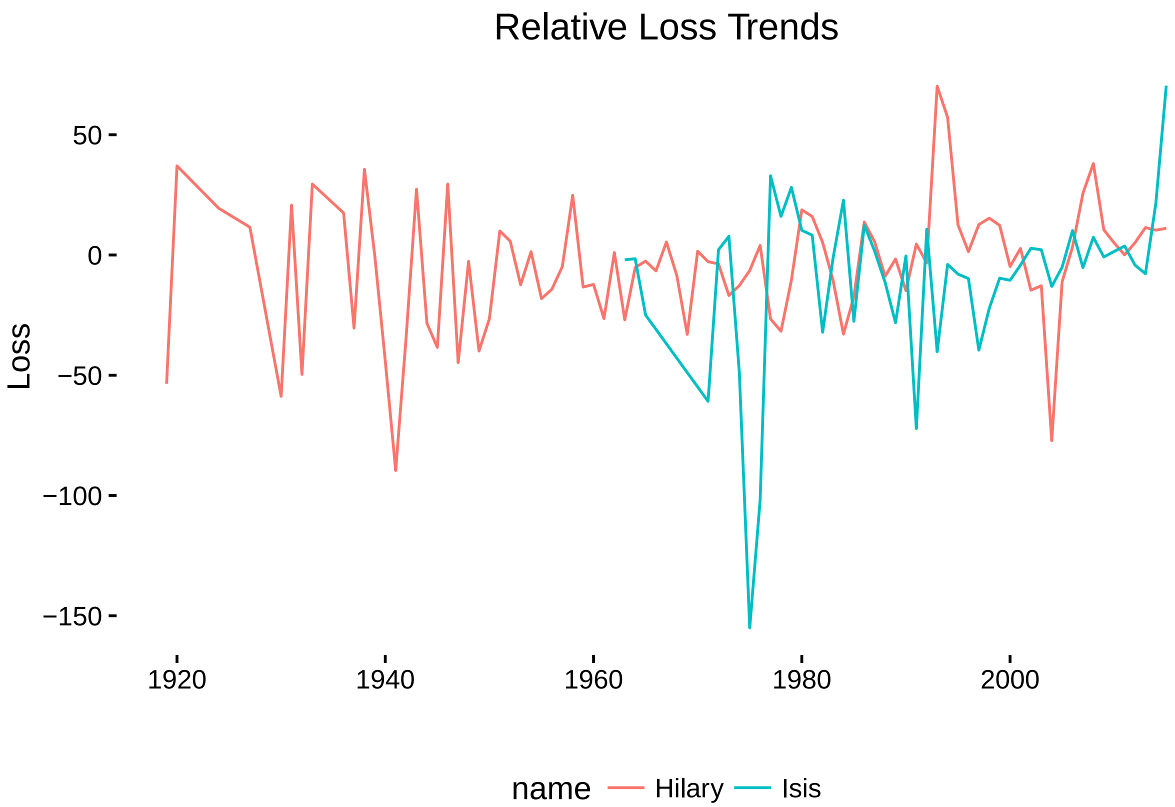
The proportional dropoff that Hilary faced in 1993 was 70%. The proportional dropoff that Isis faced in 2015 was 70%.



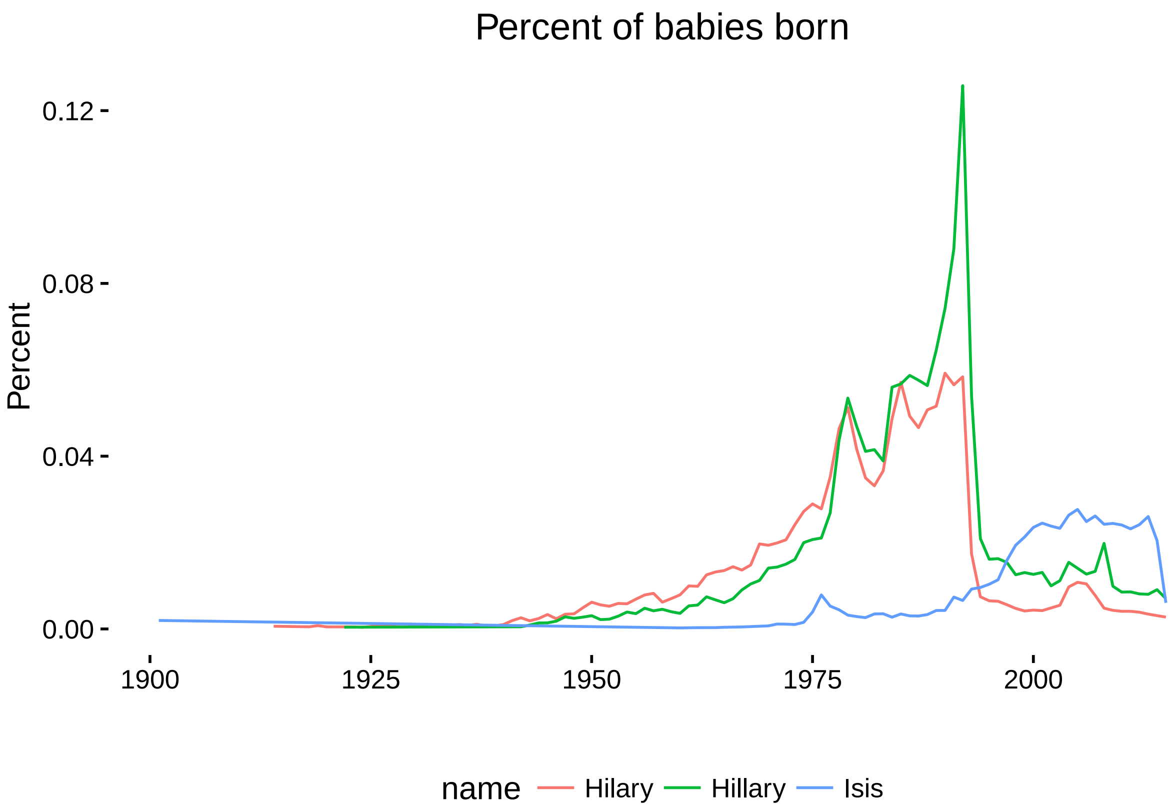
I can't think of any better metaphor for how people thought about Hillary Clinton than that her name had as much dropoff as the name of a group that's despicable by even terrorist standards.

Appendix

We can also plot the loss trends for these two names. We see that the two names have similar peaks at 70% loss, which is what we expect. Note that in this chart, higher points mean more relative loss.



We also plot the effect on the name of Hillary. It's probably worth questioning if some of the dropoff in "Hilary" is due to parents who would have named their daughters "Hilary" naming their daughters "Hillary" instead. The two names are roughly equal in popularity until the Clintons come to the public eye, but now Hillary is twice as popular as Hilary. Somewhat amusingly, Isis was more popular than both of these names until recently…


Percent of babies born
Percent
0.12
0.08
0.04
0.00
1900
1925
1950
1975
2000
name
Hilary
Hillary
Isis