Data @ Night - What's in a name?

WUDAC

26 October 2016

About us

- Chris Hua
- Jack Soslow
- Yoonduk Kim

Goal: spend time getting familiar with R and basic data manipulation

Babies!

Figure 1: Hillary

Install packages and load data

Install our easy to use babynames package:

```
if(!require(devtools)) {
  install.packages(devtools)
}
devtools::install_github(
    "wharton-data-analytics/babynames")
```
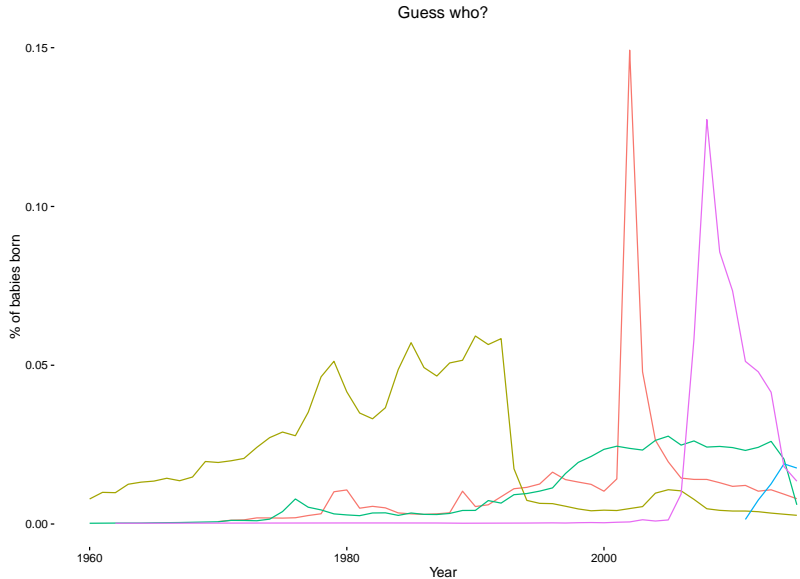
or look in the R code for a work around. Also, install a few more packages. . .
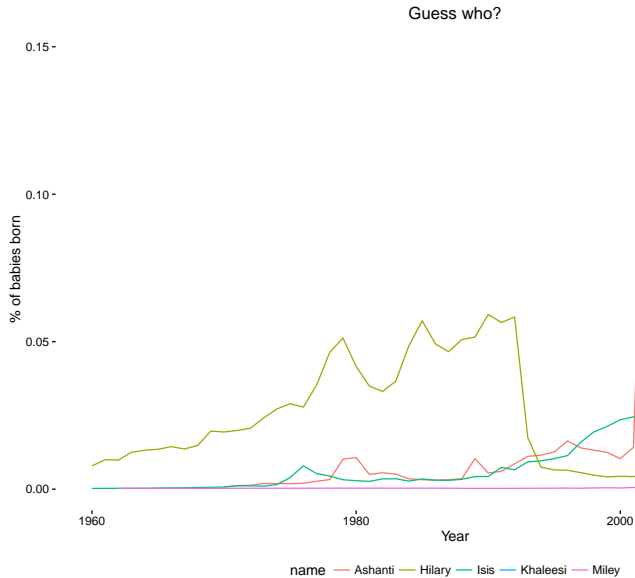
Explore the data!

Useful commands:

```
babynames %>% head()
babynames %>% str()
babynames %>% filter(name == "Christopher")
babynames %>% filter(sex == "M")
babynames %>% filter(name == "Jack") %>%
    select(year, n) %>% plot()
```

Trends

Guess who?

Trends 2

Guess who?

Poisoned names

Names that drop off - a lot - between years

How to detect?

"Relative Risk": current proportion / proportion a year before

Output

year	name	loss
1978	Farrah	0.78
1915	Meta	0.75
1995	Kadijah	0.75
1974	Catina	0.74
1990	Stephani	0.74
1915	Hulda	0.73
1965	Deneen	0.72
1995	Khadijah	0.72
1913	Mina	0.70
1914	Nelle	0.70
1993	Hilary	0.70
2015	Isis	0.70
1974	Katina	0.69
1981	Renata	0.69
1992	Iesha	0.69

Process

For each name - calculate "loss"

```
babynames %>%
    filter(sex == "F", n >= 115) %>%
    group_by(name) %>%
    mutate(loss = 1 - round((prop / lag(prop, 1)), 2)) %>%
    arrange(-loss)
```

Plot

```
babynames %>%
    filter(name %in% c("Adolph", "Adolf") ) %>%
    group_by(year) %>%
    summarize(n = sum(n)) %>%
    ggplot(aes(x = year, y = n)) + geom_line() +
    theme_classic() + ggtitle("Adolf") +
    xlab("Year") + ylab("Babies")
```


Adolf