Polishing Graphics
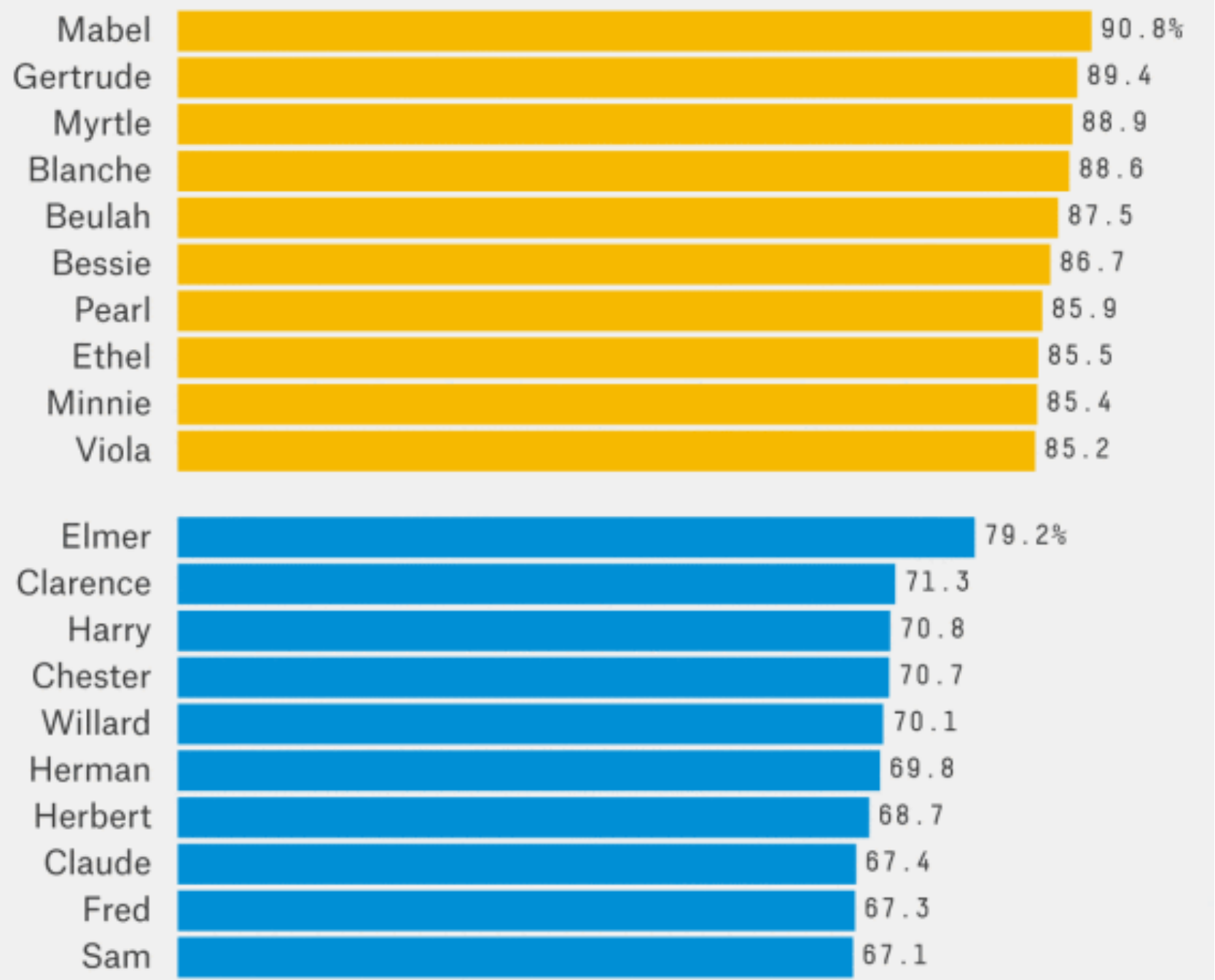
Deadest Names

Estimated percentage of Americans with a given name born since 1900 who were dead as of Jan. 1, 2014

Name	Percentage
Mabel	90.8%
Gertrude	89.4
Myrtle	88.9
Blanche	88.6
Beulah	87.5
Bessie	86.7
Pearl	85.9
Ethel	85.5
Minnie	85.4
Viola	85.2
Elmer	79.2%
Clarence	71.3
Harry	70.8
Chester	70.7
Willard	70.1
Herman	69.8
Herbert	68.7
Claude	67.4
Fred	67.3
Sam	67.1

FIVETHIRTYEIGHT   SOURCE: SOCIAL SECURITY ADMINISTRATION

Goals

- bar chart for names

- faceting and fill by sex

- flip the axes

- text giving the percentages

Changing scales

Recipe: `scale_<aes>_<method>`

Examples:

- `scale_fill_manual`

- `scale_fill_brewer`

- `scale_color_viridis`

- `scale_shape_manual`

Changing themes

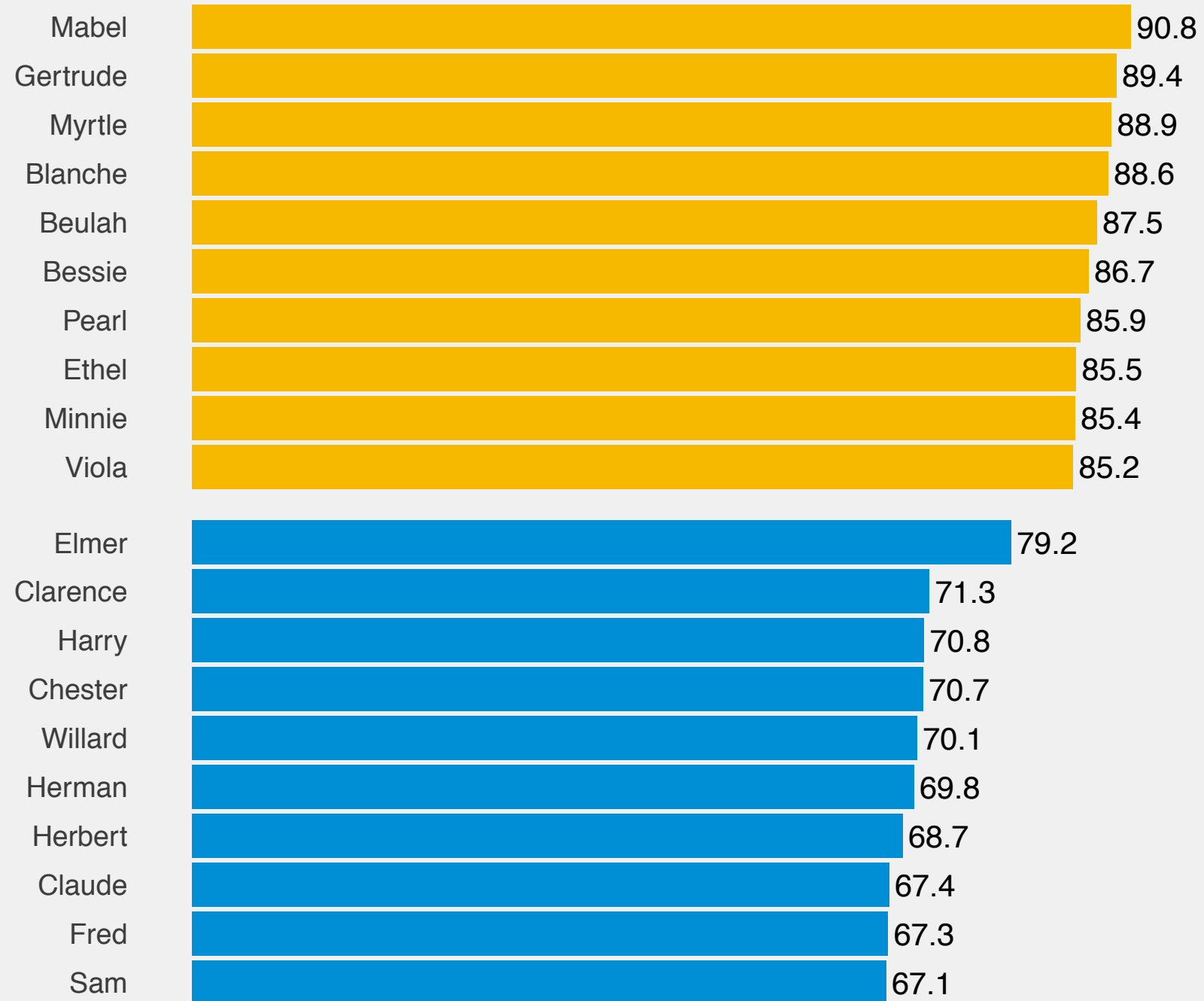Loading ggthemes allows you to access numerous <u>prebuilt themes</u>

- `theme_wsj()`

- `theme_few()`

- `theme_hc()`

- `theme_gdocs()`

Custom themes

- theme() function allows us to fine tune every aspect of our plot canvas

- A LOT of decisions (run ?theme to see more)

- A LOT of control

Deadest Names

Estimated percentage of Americans with a given name born since 1900 who were dead as of Jan. 1, 2014

Name	Percentage
Mabel	90.8
Gertrude	89.4
Myrtle	88.9
Blanche	88.6
Beulah	87.5
Bessie	86.7
Pearl	85.9
Ethel	85.5
Minnie	85.4
Viola	85.2
Elmer	79.2
Clarence	71.3
Harry	70.8
Chester	70.7
Willard	70.1
Herman	69.8
Herbert	68.7
Claude	67.4
Fred	67.3
Sam	67.1

```
dnplot2 +
  theme_fivethirtyeight() +
  theme(legend.position = "none",
       panel.grid = element_blank(),
       strip.text = element_blank(),
       axis.text.x = element_blank()
       )
```

Mini project

Assignment

1. You will work with a partner to write a short blog post that contains (at least) one data graphic. Your goal is to tell the class something interesting using a well-crafted, thoughtfully-prepared data graphic. One data graphic should suffice, but you may include more if you choose.

2. Give a lightning talk (~ 5 min) summarizing the main point(s) of your blog post.

Things to think about

- What point are you trying to make?

- What variable(s) is/are important to your?

- What visualization(s) will help you make that point effectively?

Data

- You can use any data set you want, but don't focus on data wrangling

- There are many well-curated data sets in R packages:

  - `nycflights13`: 2013 flights departing 3 NYC airports

  - `babynames`: history of baby names from the Social Security Administration

  - Lahman: comprehensive historical archive of major league baseball data

  - `fueleconomy`: fuel economy data from the EPA, 1985–2015

  - `fivethirtyeight`: provides access to data sets that drive many articles on FiveThirtyEight.

Resources

- ggplot2 cheat sheet

- DataCamp

- Google

Start brainstorming!