Customizing Plots

We now know how to create many plot types using `ggplot2`, but often you want to customize your plots. Today, we'll discuss common tools for customizing our plots.

To begin, make sure that you have the `tidyverse` loaded.

```
library(tidyverse)
```

The Deadest Names in the U.S.

The "Deadest Names" graphics appeared on fivethirtyeight.com on 5/29/14. Today, we'll recreate this plot to illustrate additional customizations that are possible in R.

The data for this plot can be found on github and loaded into R using the following command:

```
dead_names <- read_csv("https://raw.githubusercontent.com/cmsc205/data/master/dead_names.csv")
```

```
## Parsed with column specification:
## cols(
##   name = col_character(),
##   sex = col_character(),
##   pct_dead = col_double()
## )
```

Recreating the graphic

First, we must recreate the structure of the bar chart.

```
dnplot <-
  ggplot(dead_names, aes(x = reorder(name, pct_dead), y = pct_dead, fill = sex)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = pct_dead), hjust = -.1) +
  facet_wrap(~sex, scales = "free_y", ncol = 1) +
  coord_flip() +
  ylim(0, 95) +
  labs(x = NULL, y = NULL,
       title = "Deadest Names",
       subtitle = "Estimated percentage of Americans with a given name born since\n1900 who were dead a
```

Choosing color palettes

`ggplot2` does not choose the nicest looking palettes, and the palettes are typically not colorblind friendly. Luckily, it's quite easy to change the color palettes you use.

Manual adjustments

```
dnplot +
  scale_fill_manual(values = c("goldenrod1", "dodgerblue1"))
```

Deadest Names

Estimated percentage of Americans with a given name born since 1900 who were dead as of Jan. 1, 2014
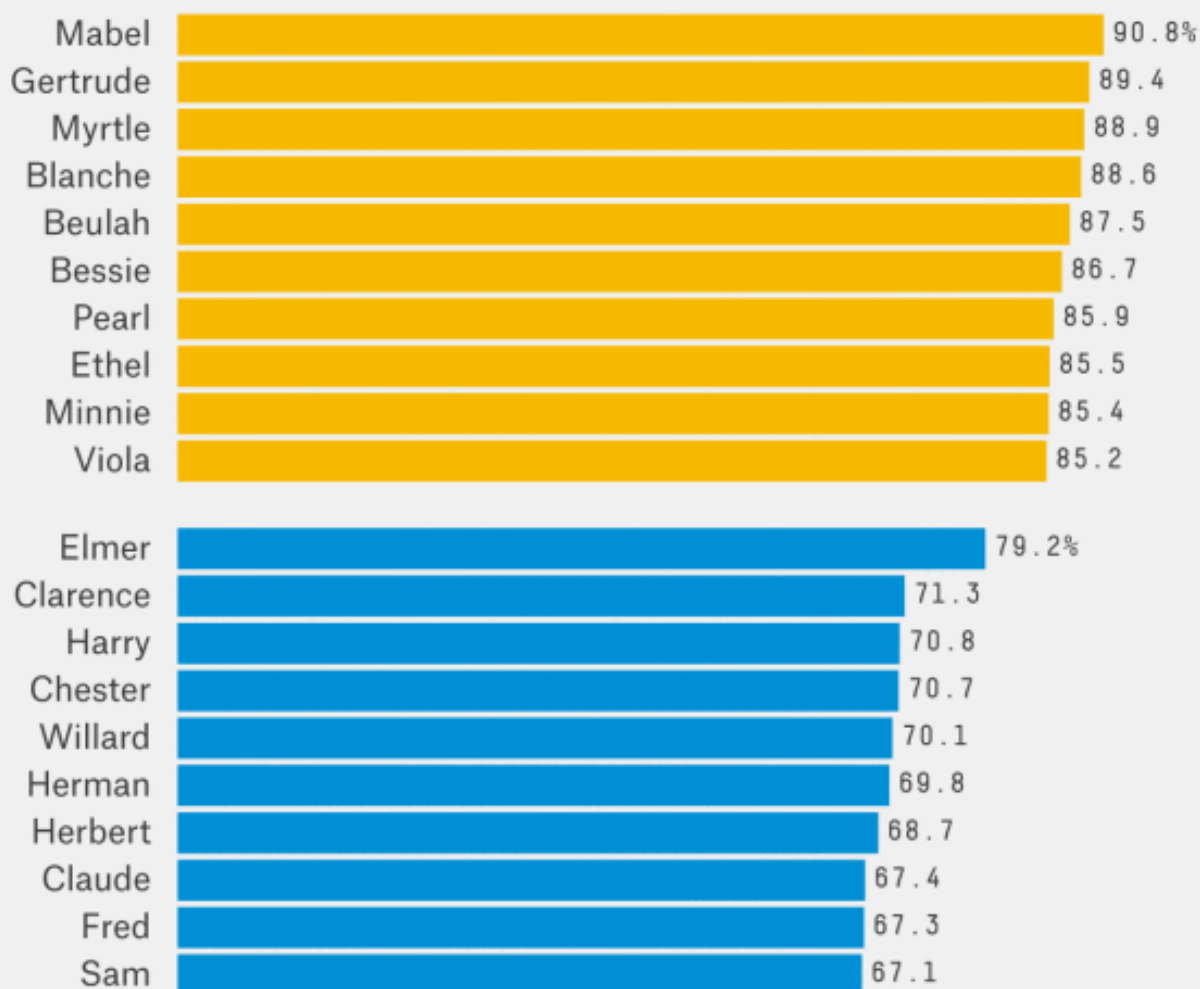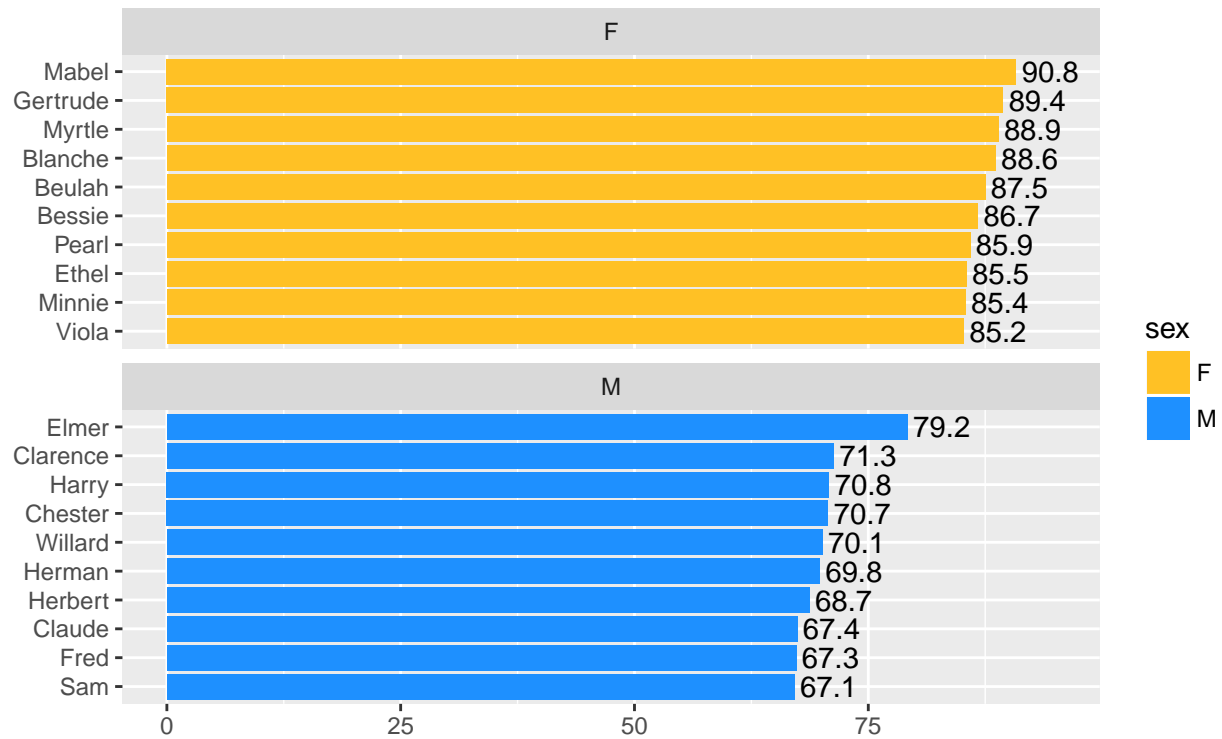
Name	Percentage
Mabel	90.8%
Gertrude	89.4
Myrtle	88.9
Blanche	88.6
Beulah	87.5
Bessie	86.7
Pearl	85.9
Ethel	85.5
Minnie	85.4
Viola	85.2
Elmer	79.2%
Clarence	71.3
Harry	70.8
Chester	70.7
Willard	70.1
Herman	69.8
Herbert	68.7
Claude	67.4
Fred	67.3
Sam	67.1

FIVETHIRTYEIGHT   SOURCE: SOCIAL SECURITY ADMINISTRATION

Figure 1:

Deadest Names

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

F

Name	Value
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

M

| Name	Value
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

sex

- F
- M

R has numerous named colors, but they are limited. For more freedom in your color specification, you may wish to switch to hexdecimal color specification.

We can match the colors used exactly using a hexadecimal specification determined with a digital color meter.

```
dnplot2 <-
  dnplot +
  scale_fill_manual(values = c("#f6b900", "#008fd5"))
dnplot2
```

Deadest Names

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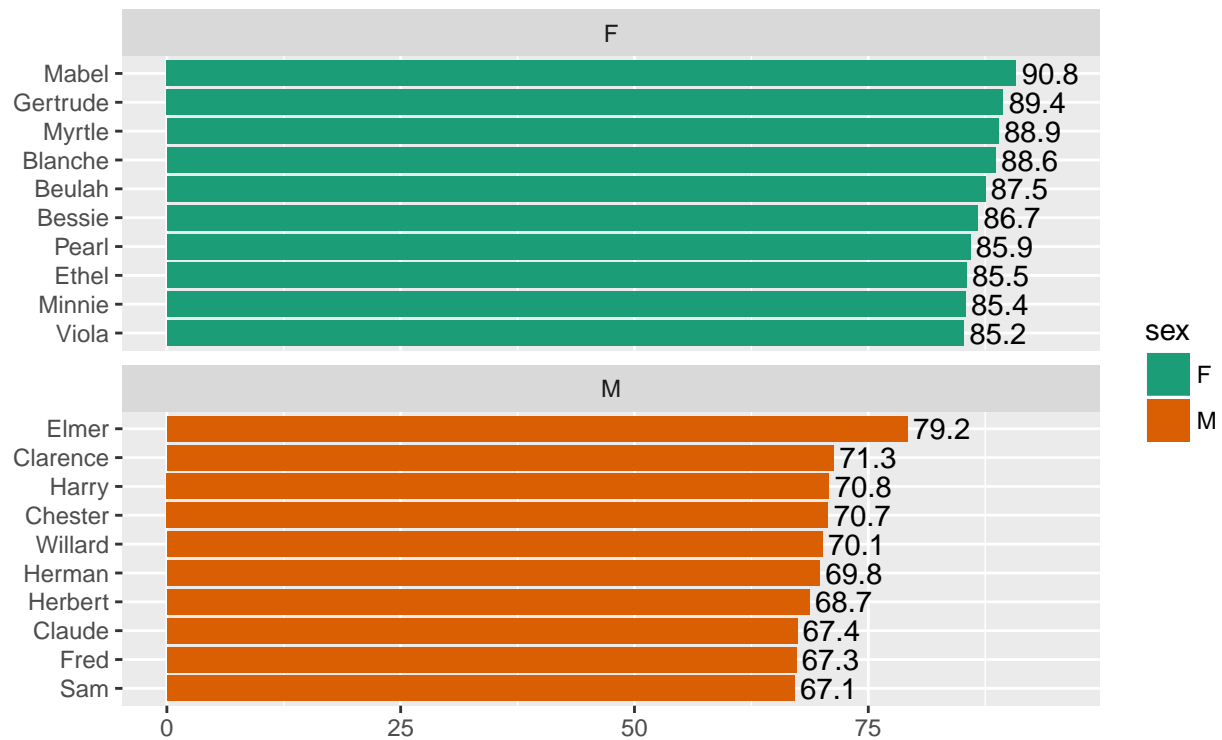1900 who were dead as of Jan. 1, 2014
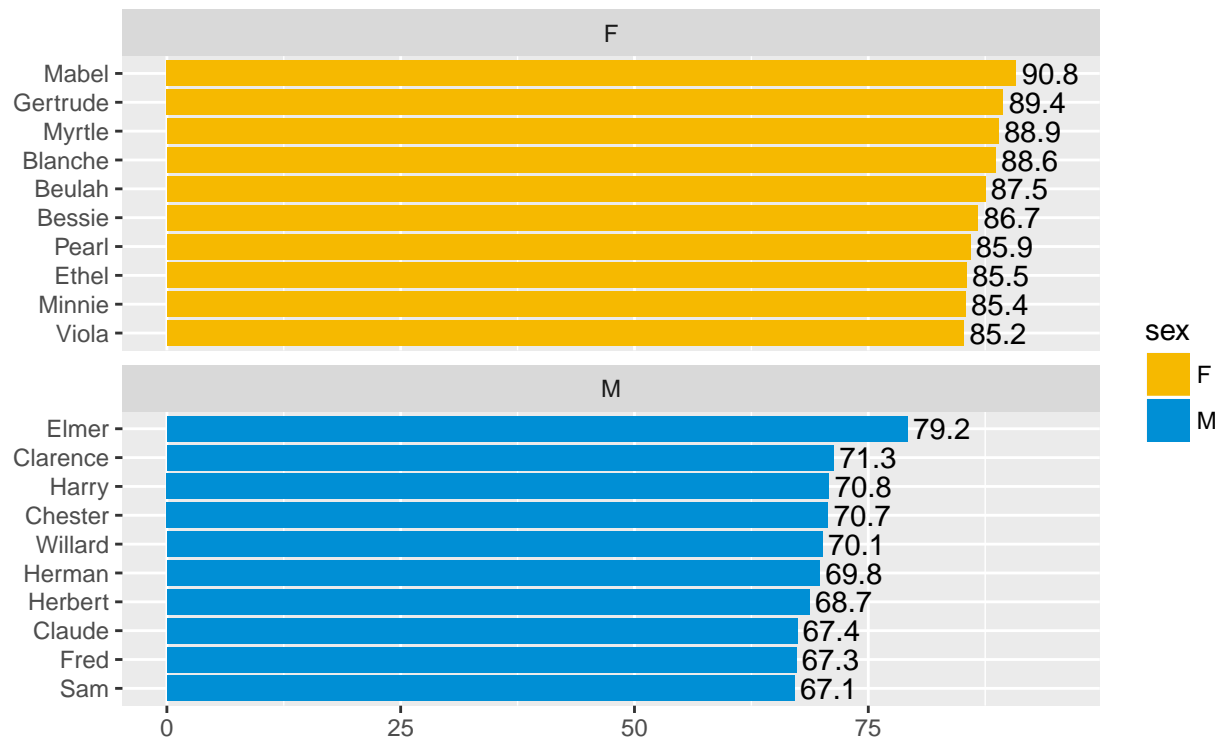
Using RColorBrewer

While it was relatively easy to pick two colors, it can be much harder to select color palettes. Luckily, we can use other vetted palettes.

```
dnplot +
  scale_fill_brewer(palette = "Dark2")
```

Deadest Names

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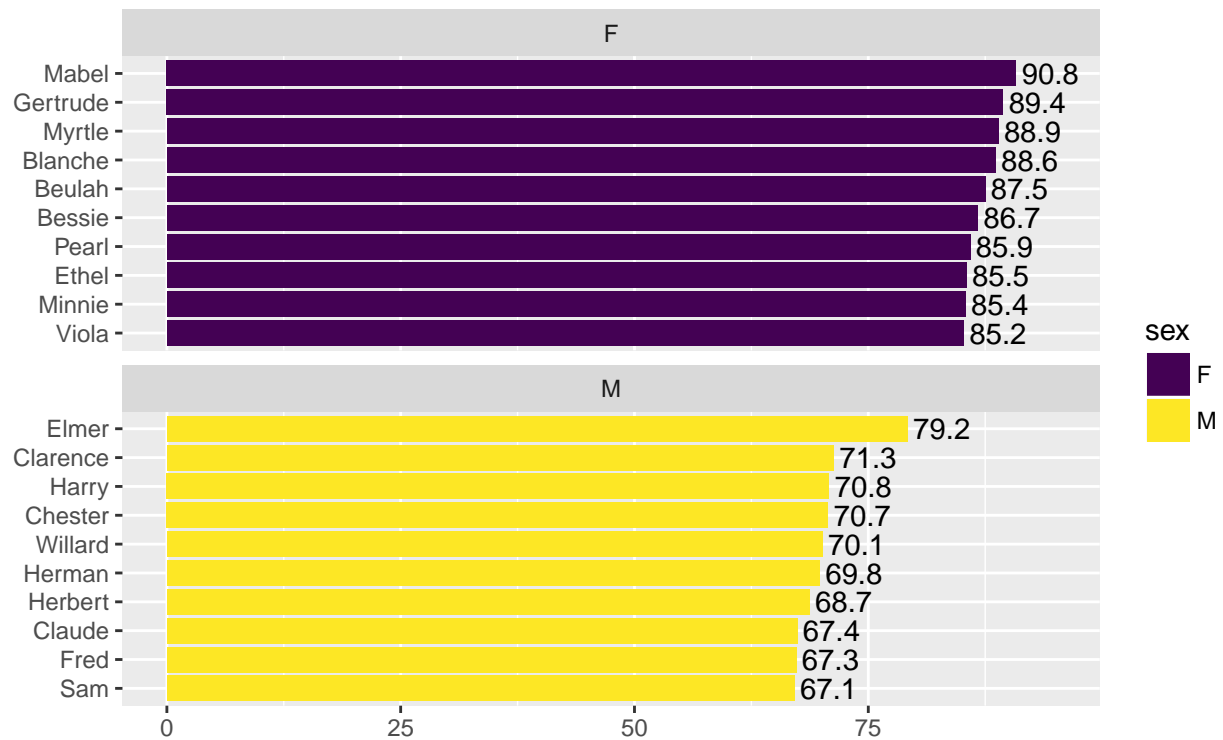1900 who were dead as of Jan. 1, 2014

Using viridis

```
library(viridis)
dnplot +
  scale_fill_viridis(discrete = TRUE)
```

Deadest Names

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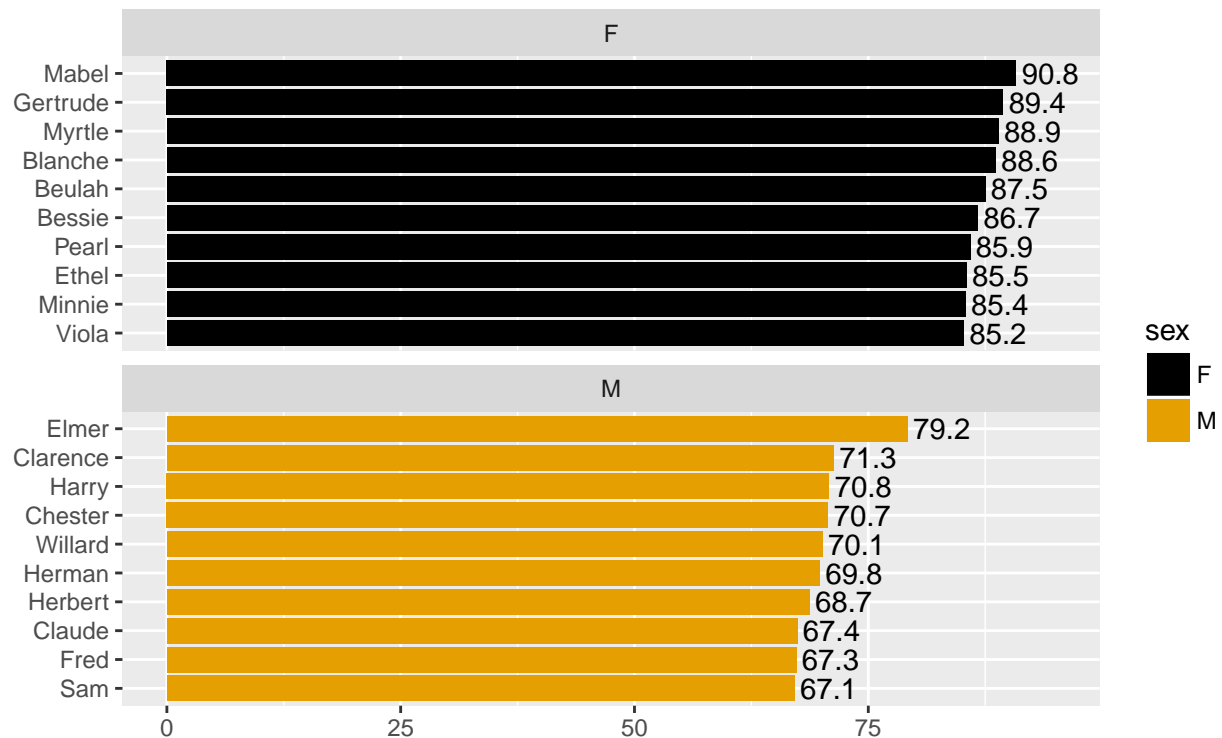1900 who were dead as of Jan. 1, 2014

A colorblind friendly palette

```
library(ggthemes)
dnplot +
  scale_fill_colorblind()
```

Deadest Names

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

	F	
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

	M	
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

sex
- F
- M

Choosing themes

By default `ggplot2` uses a grey background and guidlines on a plot; however, we don't always want our plots to appear this way. To change this the appearance, we need to work with themes.
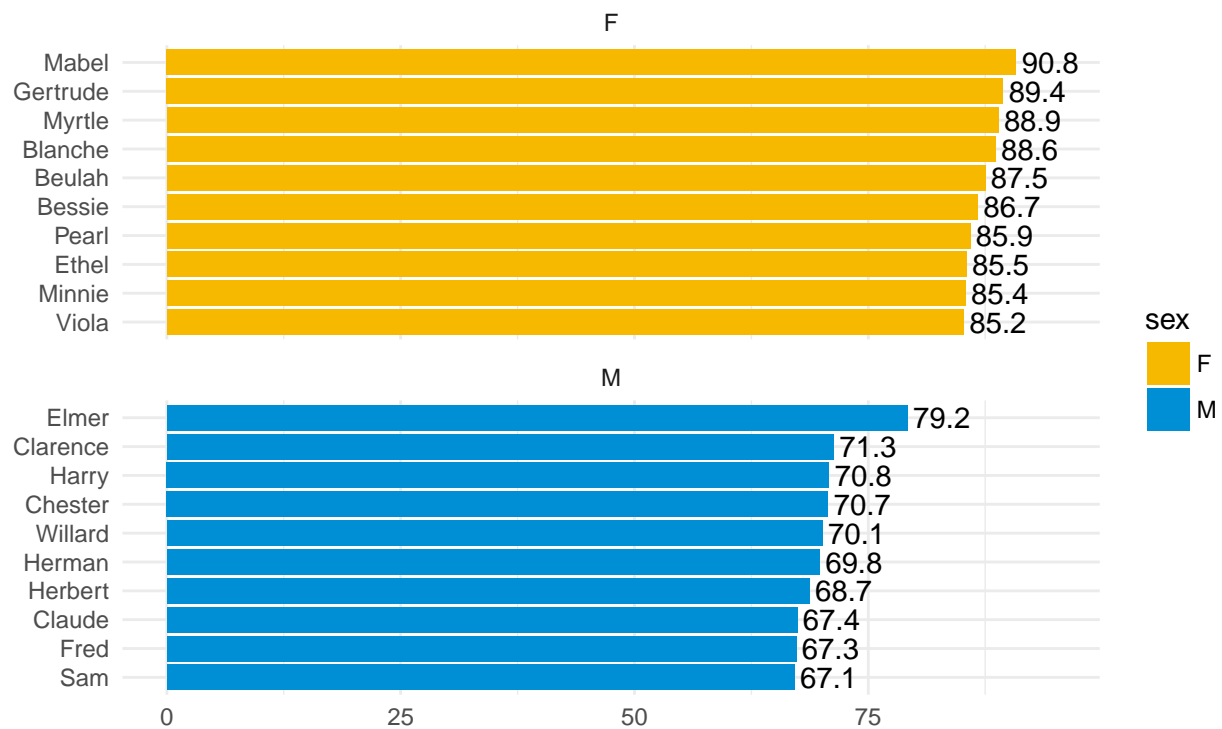
The `ggthemes` package offers many additional choices.

Remember, a new theme is added just like a layer.

```
dnplot2 +
  theme_minimal()
```

Deadest Names

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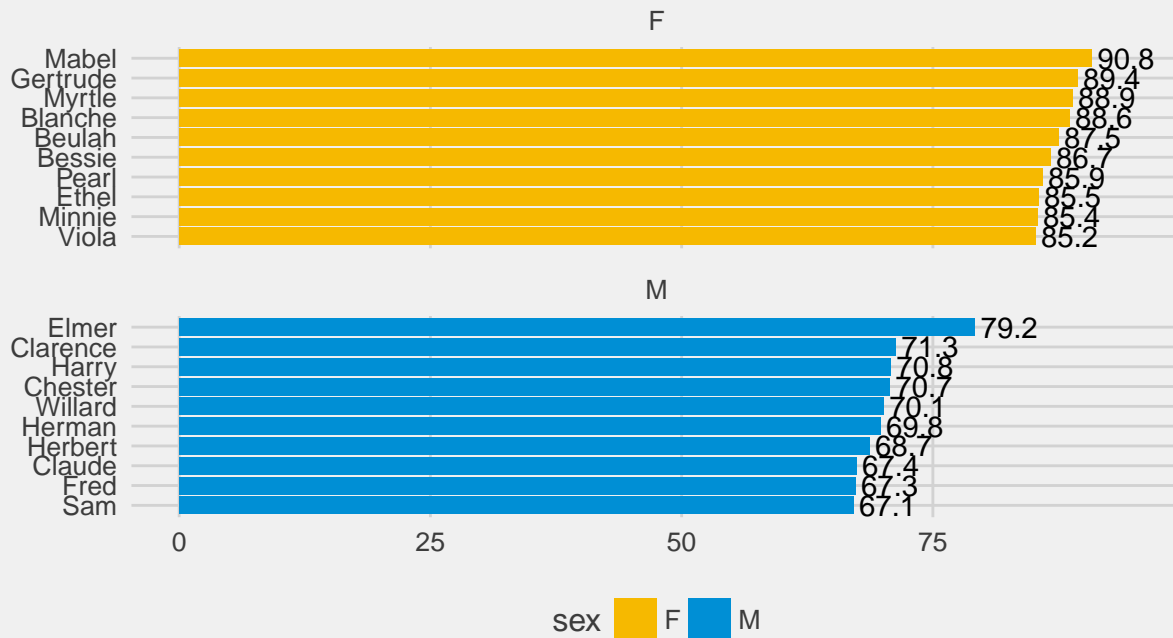1900 who were dead as of Jan. 1, 2014

What other themes could we try?

```
dnplot2 +
  theme_fivethirtyeight()
```

Customizing themes

The `theme()` function allows us to fine tune every aspect of our plot canvas.

```
dnplot2 +
  theme_fivethirtyeight() +
  theme(legend.position = "none", # get rid of legend
        panel.grid = element_blank(), # get rid of grid lines
        strip.text = element_blank(),  # get rid of facet labels
        axis.text.x = element_blank() # no scale for bar length
        )
```

Deadest Names

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

Name	Value
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