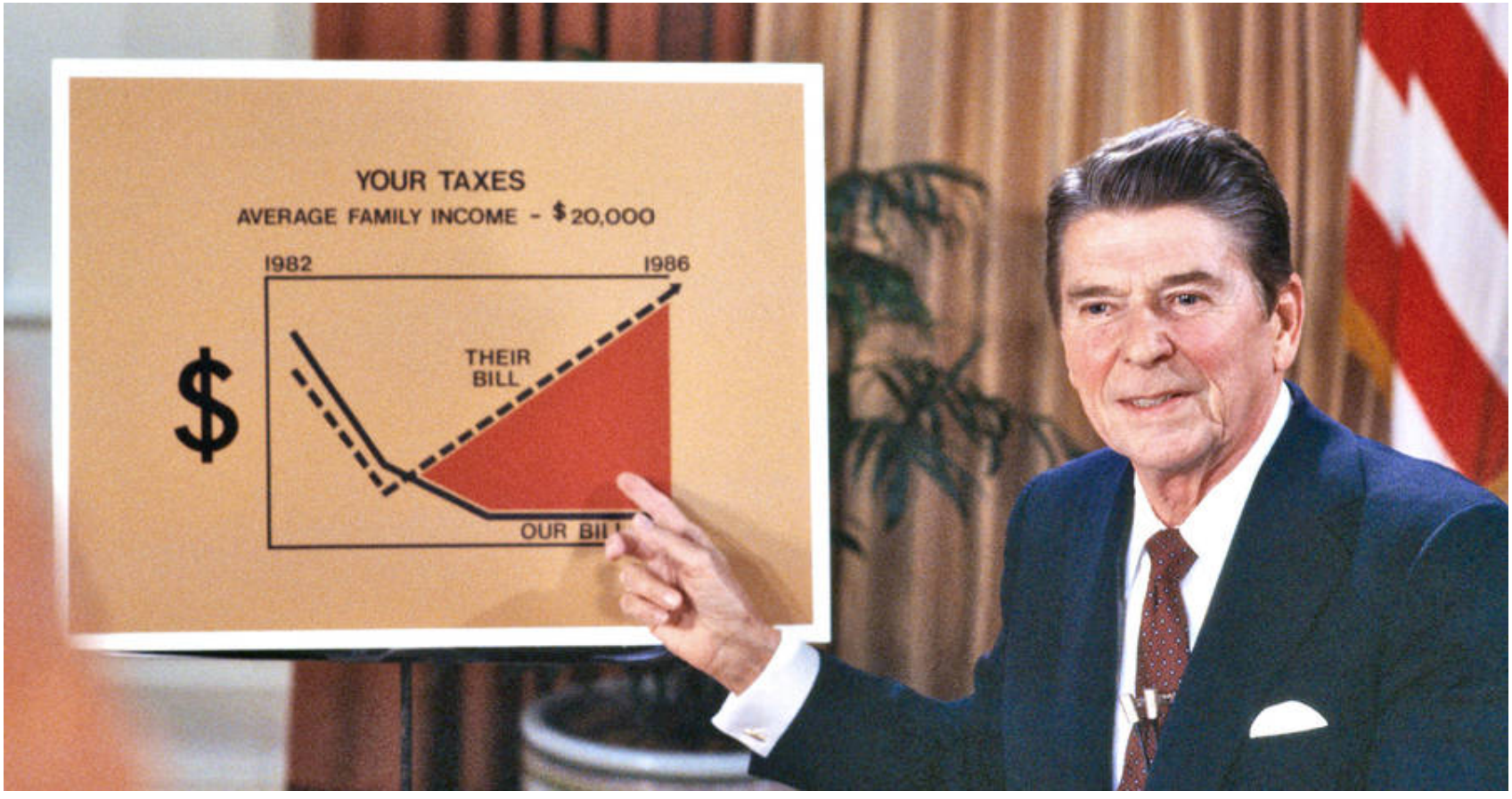
Basic Data Viz Theory and `ggplot2basics`

Sarah Moore

2022-10-04


YOUR TAXES
AVERAGE FAMILY INCOME - $20,000
1982
1986
$
THEIR BILL
OUR BIL

What do we know *already?*

- How can we think through mapping out what we already know about data analysis and graphics and how that translates into a "theory" of data viz?
- We can even get into thinking technically along the lines of probability and statistical distributions (i.e. in terms of random variables).
- Should also think through what *unit of analysis* our data are in and how we might have to transform that.

Let's try to dissect these:

What are the *variables* displayed?

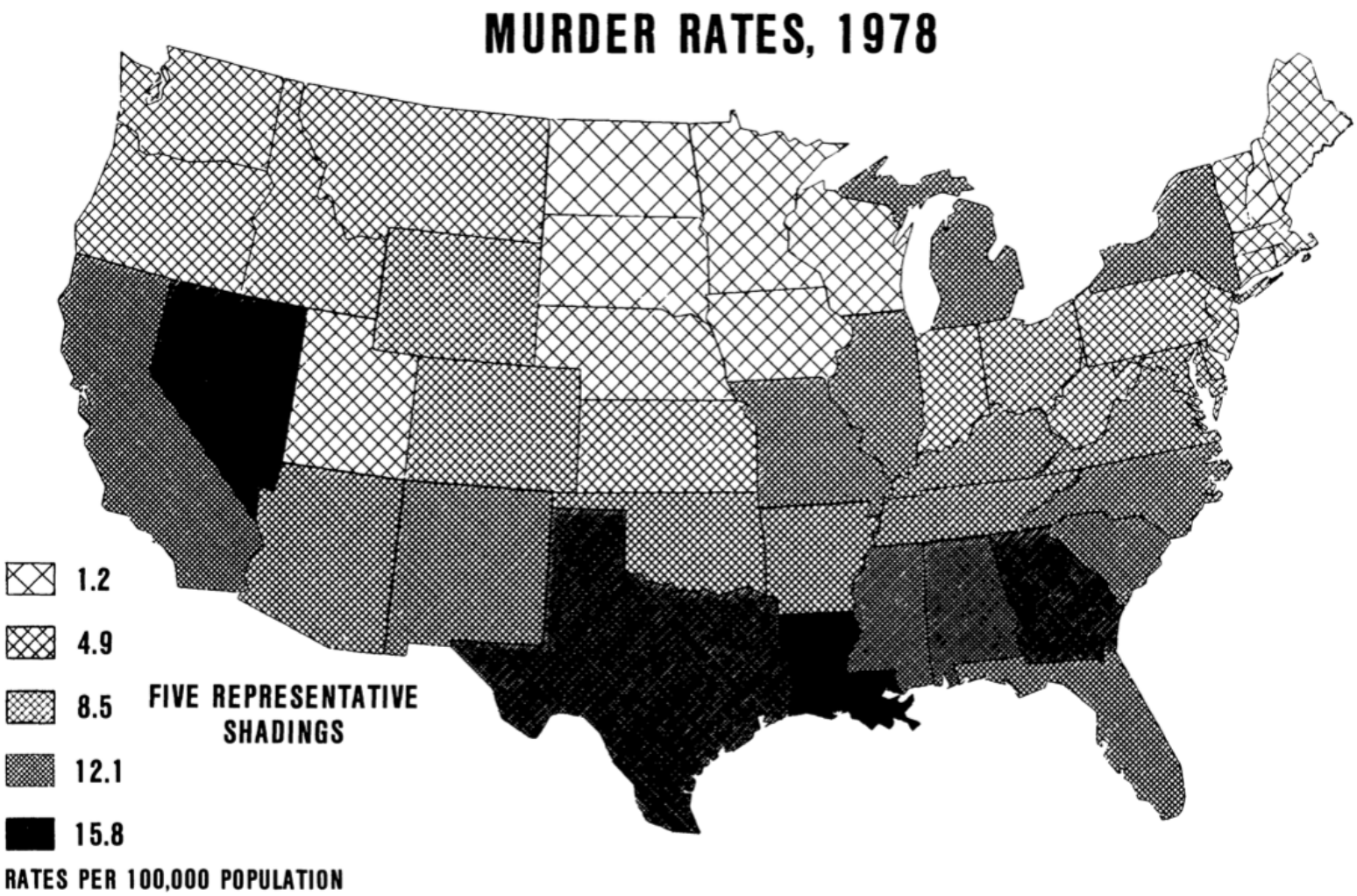


Figure 5. Statistical map with shading.

And this one?

And this one?

Tables are also a form of data viz, why not use them?

Why does this matter?

- Some graphics packages in other more viz forward platforms make it incredibly easy to just point and click through what you want to put together in a graphic. (With limitations!!!)
- `ggplot2` requires that we think a little more carefully about what we put together. Because of the way that our code layers... But this is for the better.

Anatomy of a ggplot2 call

- Any visualization starting out in ggplot2 mirrors the same way we were thinking of dplyr. Our data are the base.

Codetalk

Additive code create exactly what we want our graph to be. It's called *declarative design*.

```
some_data %>% # feed in the data
  ggplot(aes(x = x_var, y = y_var)) +
  # make the initial aesthetic spec
  geom_function() +
  # specify the geometric design
  facet_wrap(~z_var)
# optional specification of a facet

#... and so on
```

Now let's make this concrete.