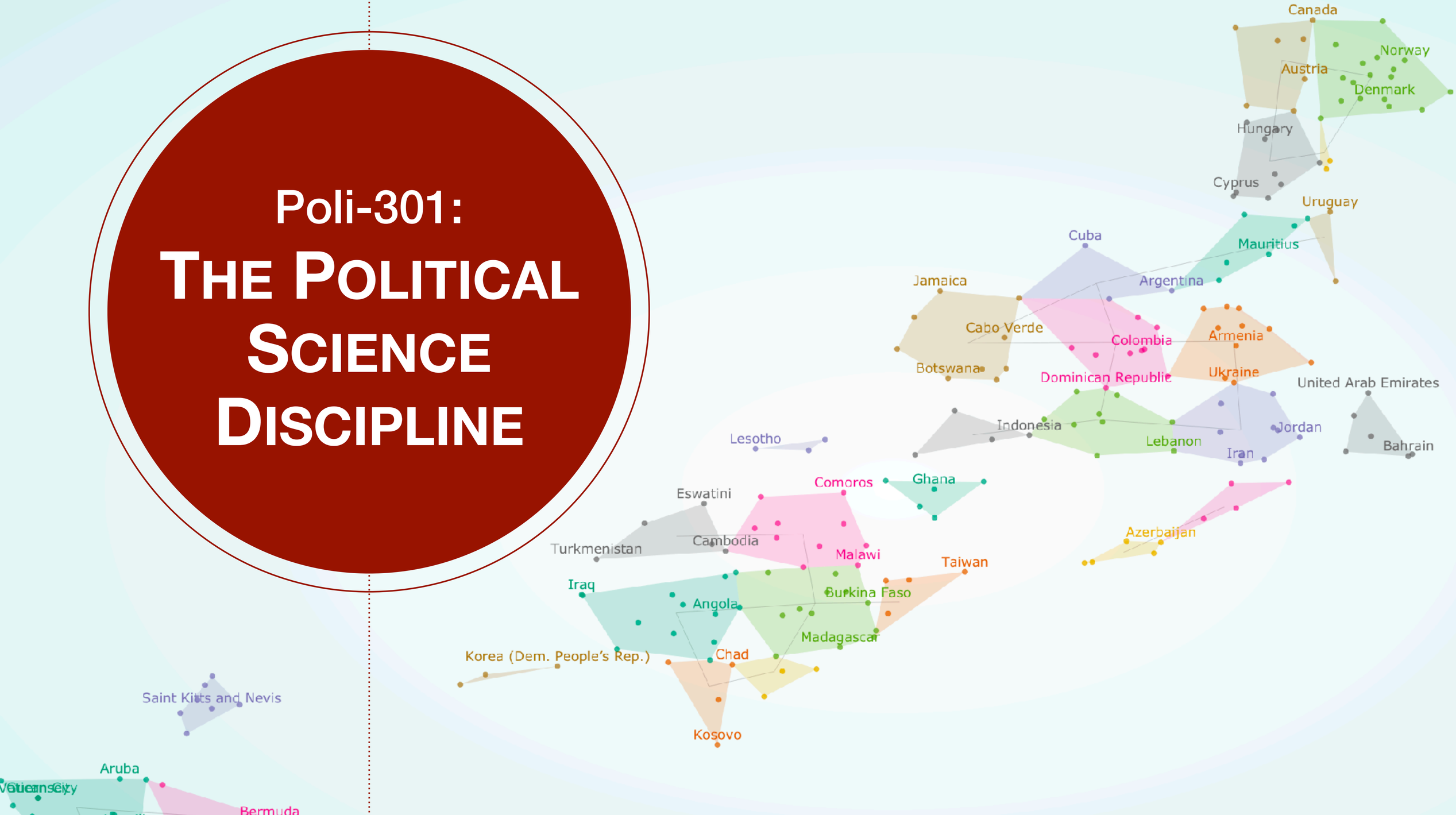
Poli-301:
The Political Science Discipline



TODAY'S AGENDA

1. Data Wrangling: pipes, filter, summary
2. Data Wrangling: mutate, select
3. Practice

W. E. B Du Bois

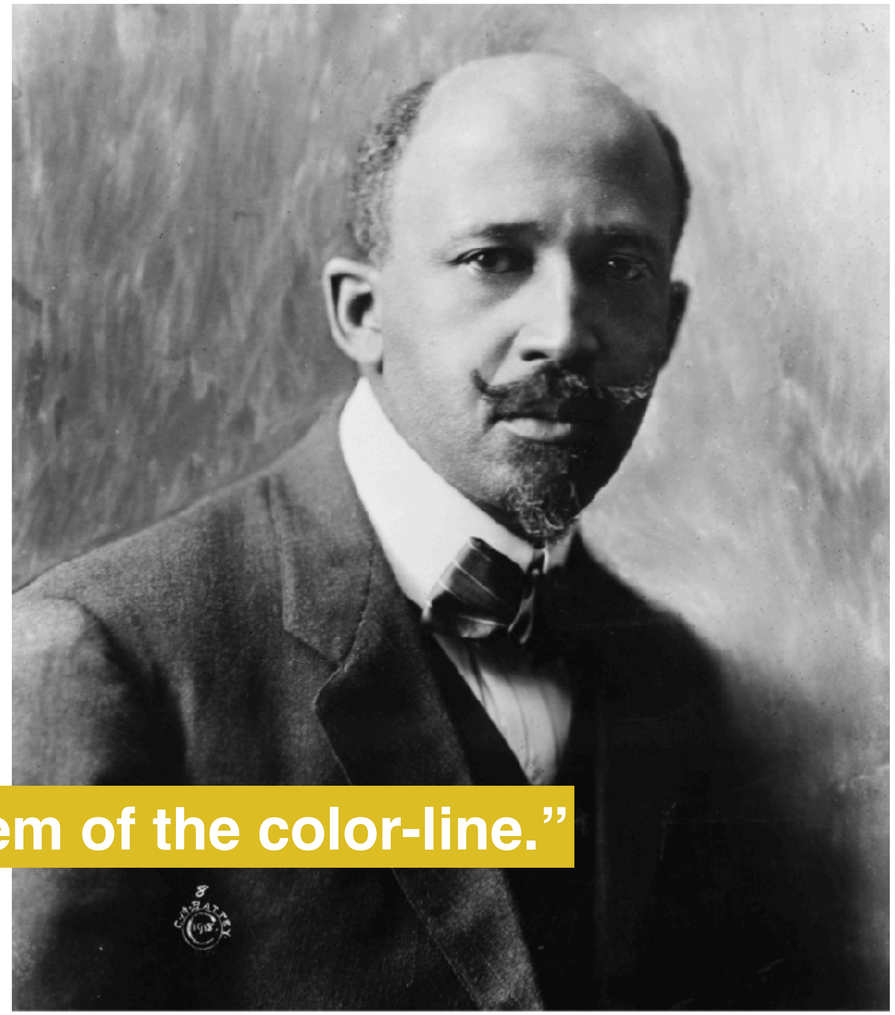
1868 — 1963

First African-American to earn PhD from Harvard

Co-founded the NAACP

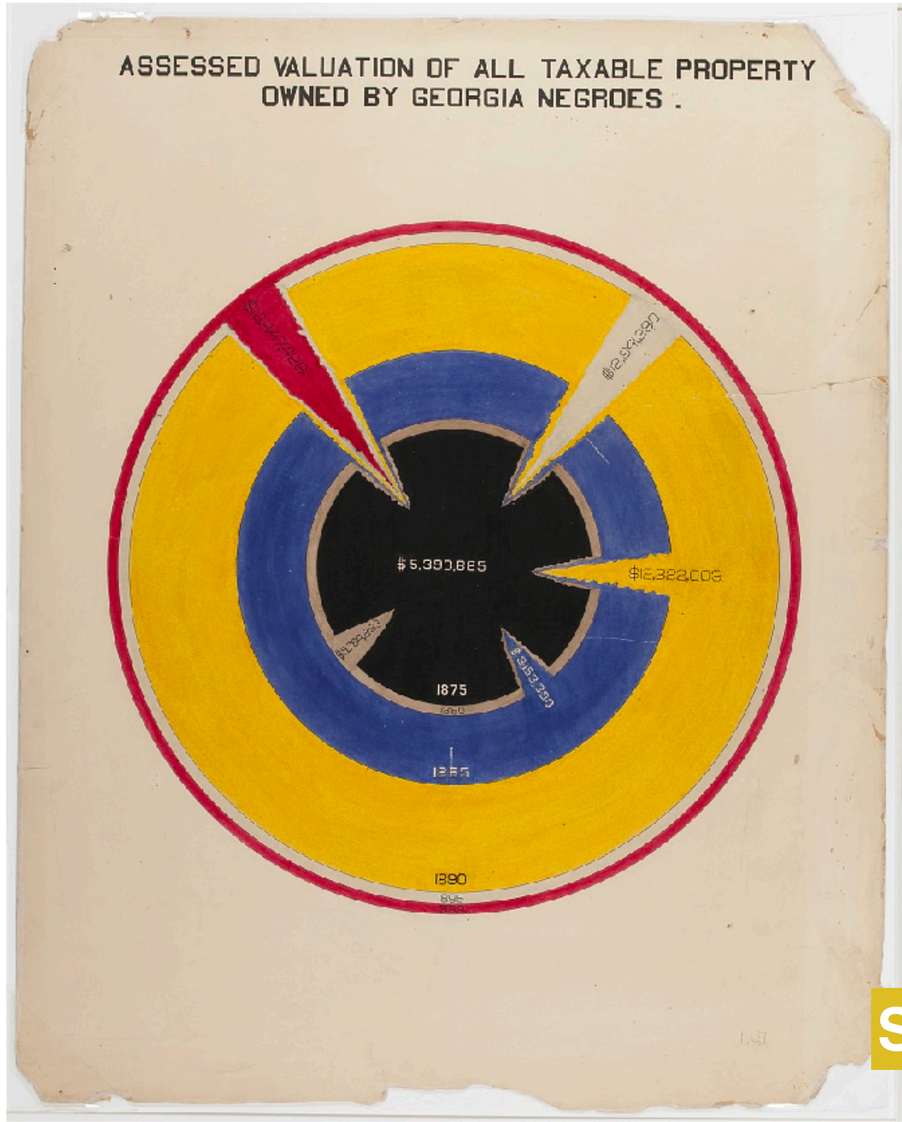
The Souls of Black Folk

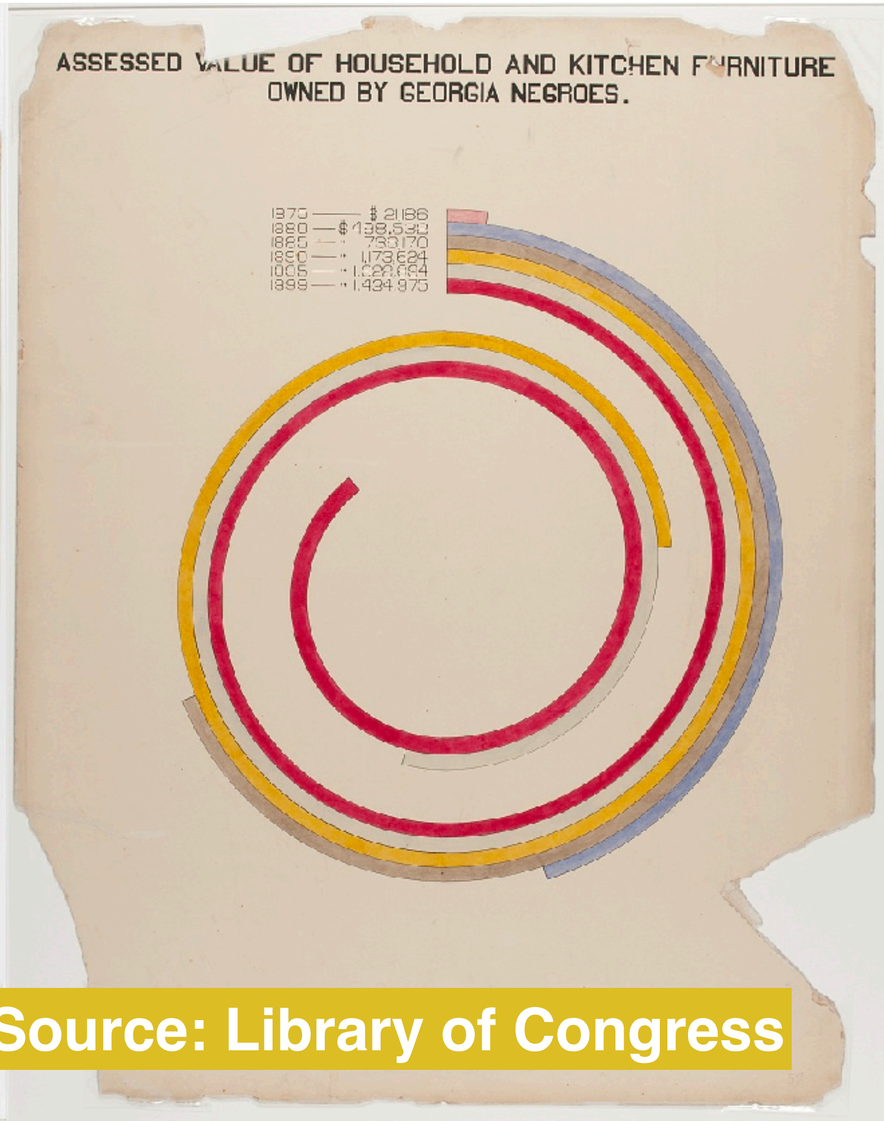
Wrote about Reconstruction, Jim Crow, civil rights

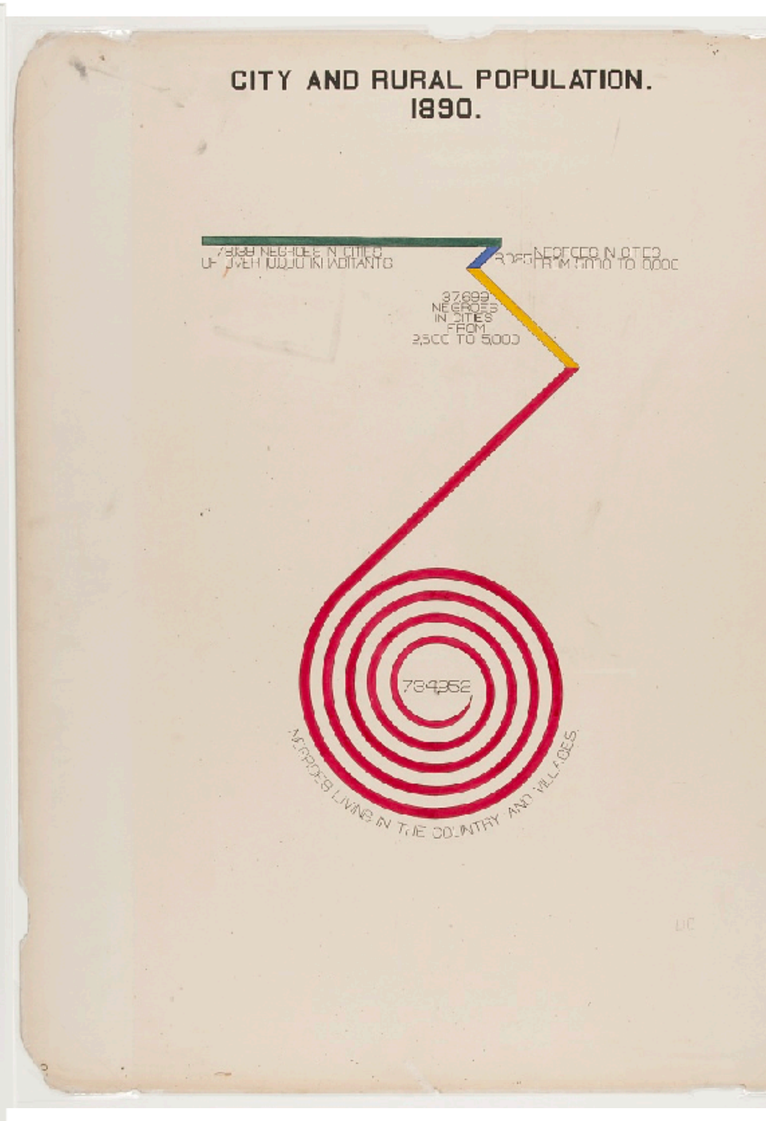


"The problem of the twentieth century is the problem of the color-line."

Also brilliant at visualizations

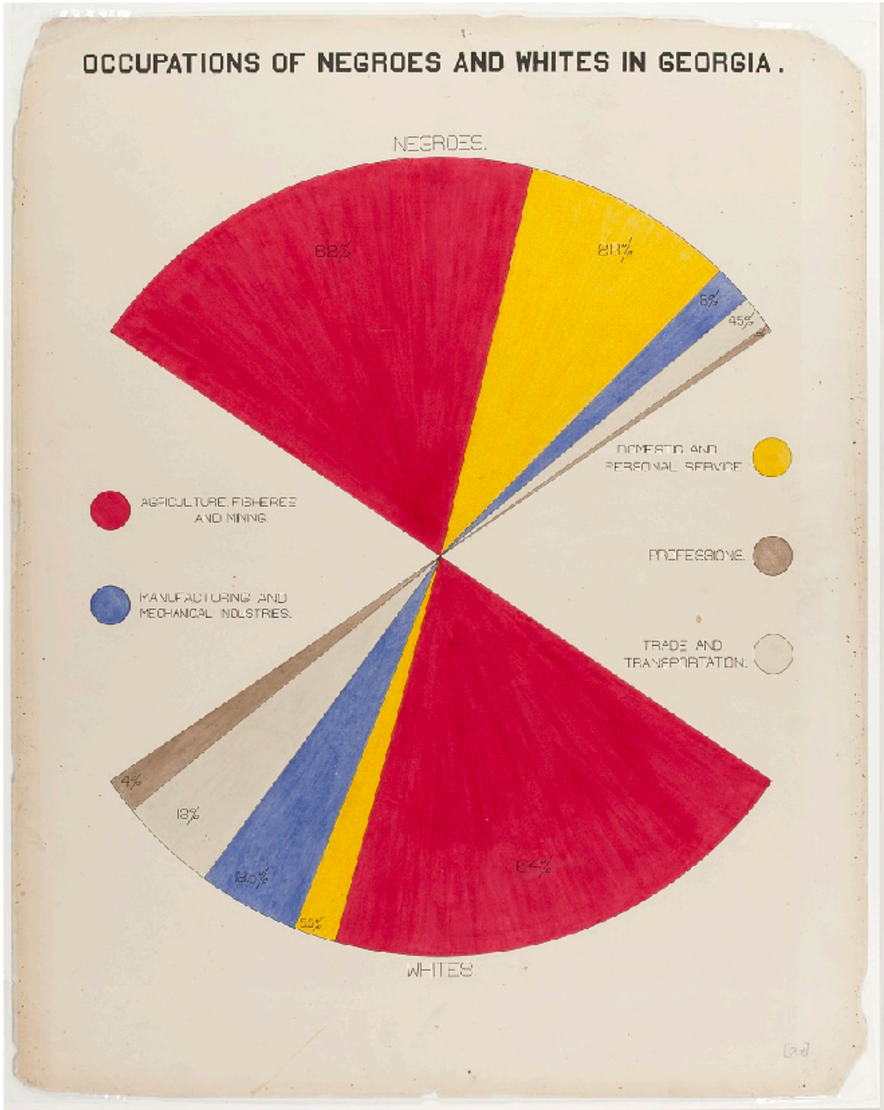


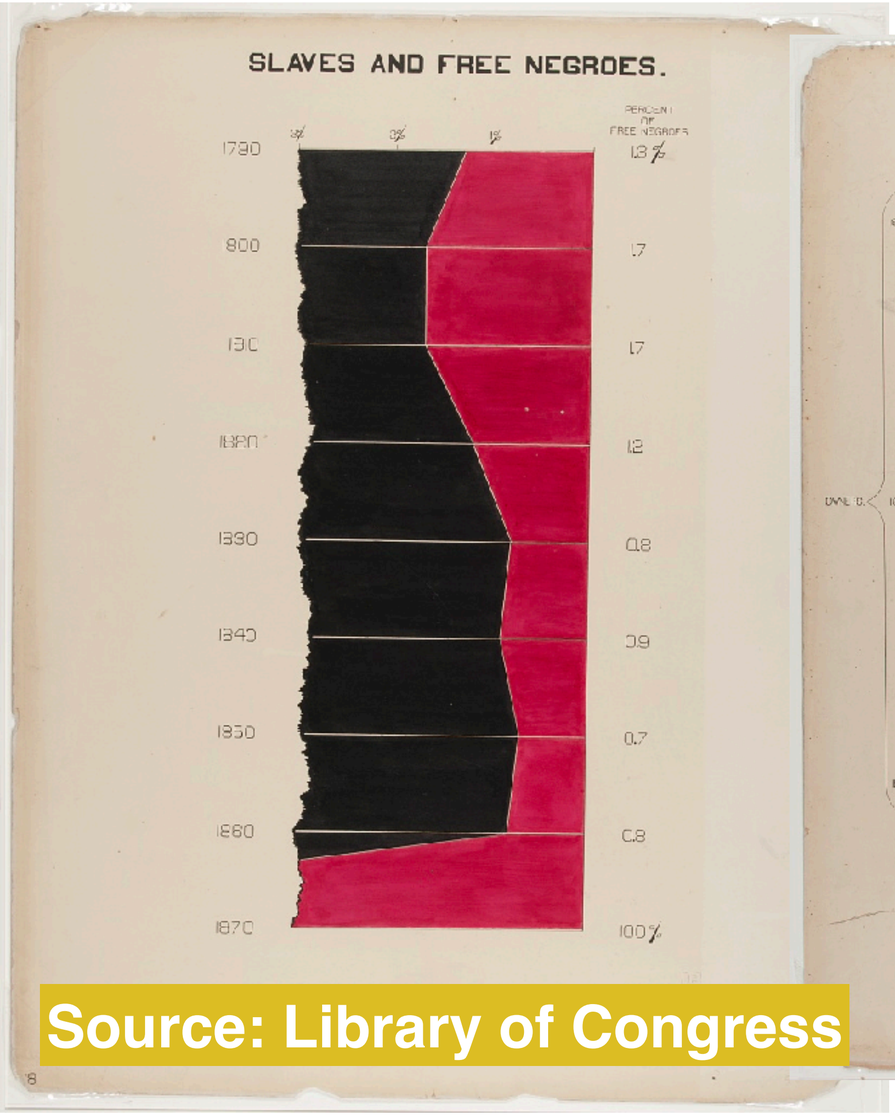


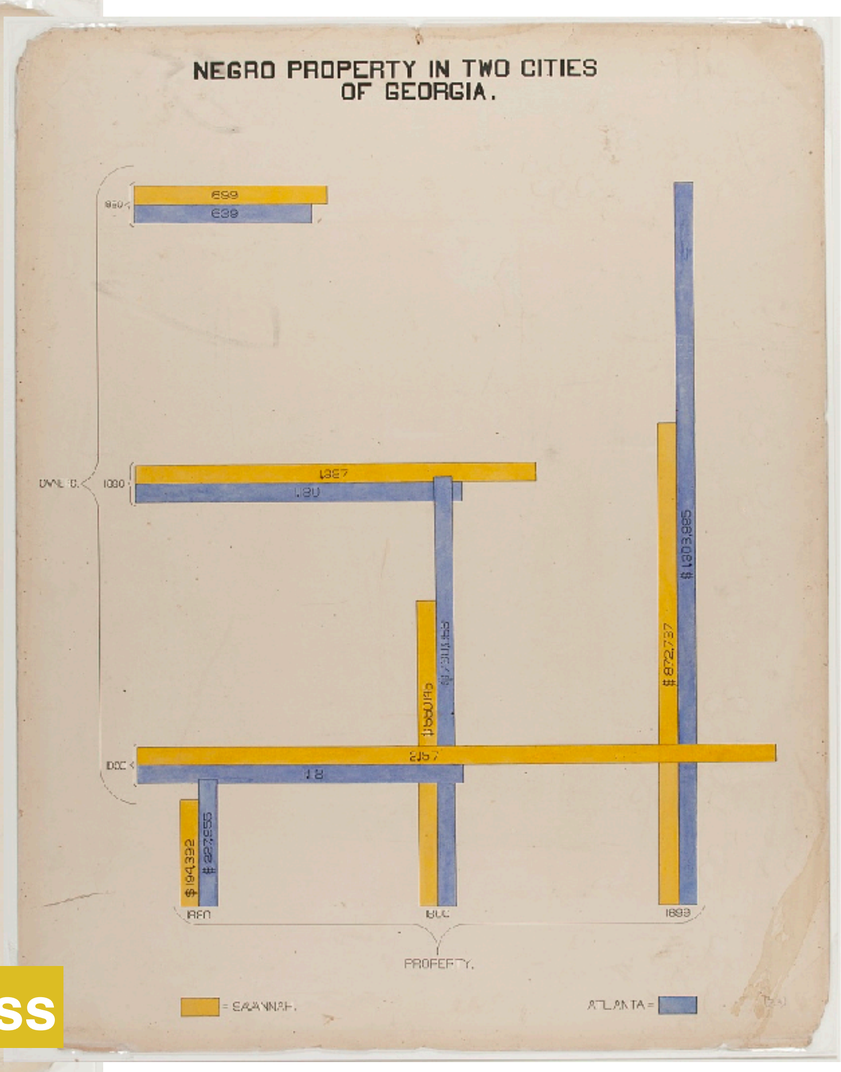


Source: Library of Congress

Also brilliant at visualizations

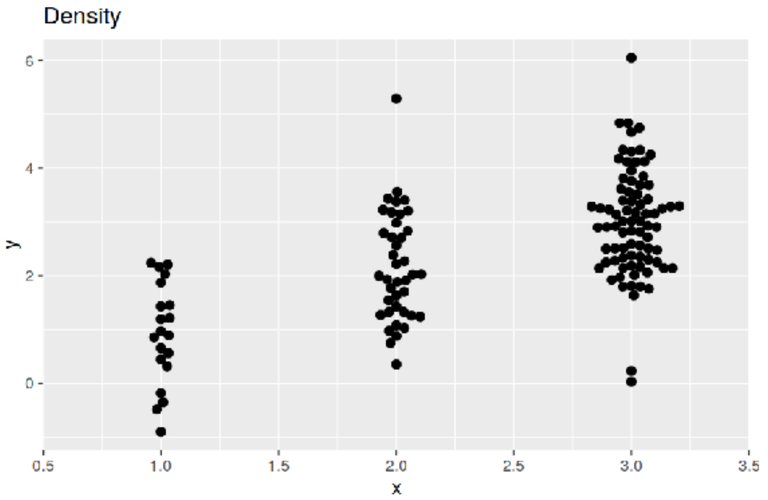


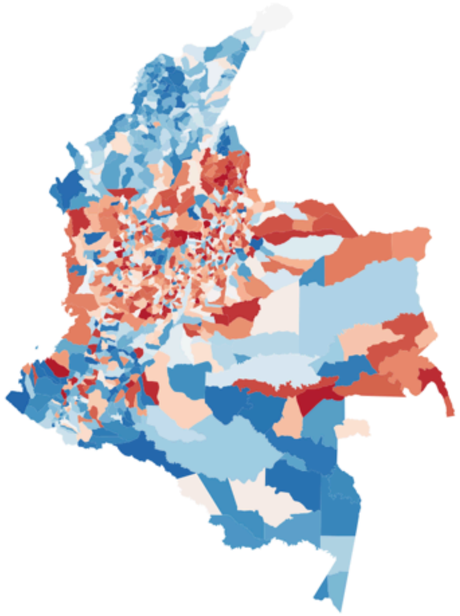




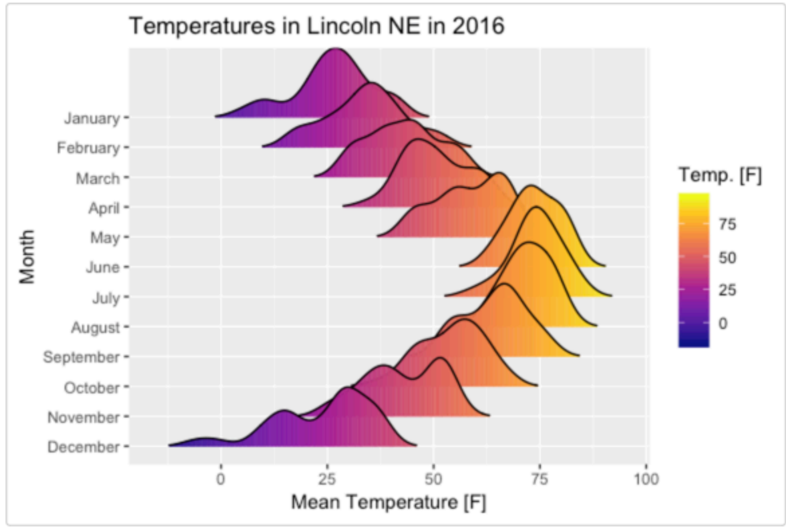
Source: Library of Congress

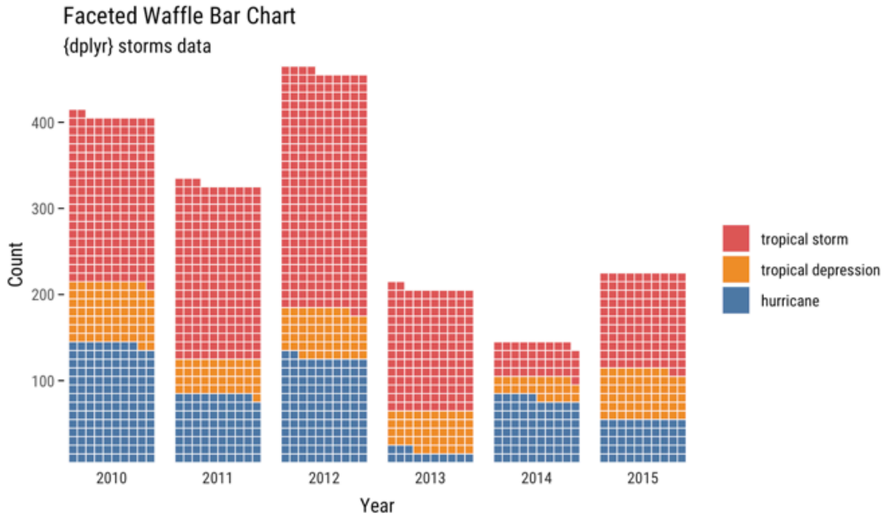
More ggplot





```
ggplot(lincoln_weather, aes(x = `Mean Temperature [F]`, y = `Month`, fill = ..x..)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01) +
  scale_fill_viridis(name = "Temp. [F]", option = "C") +
  labs(title = 'Temperatures in Lincoln NE in 2016')
```





So far...

I do stuff to data

```
# remove the outlier
shootings_no_outlier =
  mass_shootings %>%
  filter(date != "2017-10-01") # "!=" means "not equal to"


# SCATTERPLOT: plot again
ggplot(shootings_no_outlier, aes(x = date, y = injured, label = case)) +
  geom_point()
```

Ask you to plot

Now...

You do stuff to data

`join_()`

`select()`

`mutate()`

`filter()`

`group_by()`

`summarize()`

Pipes: $ > $

```
plot(select(mutate(filter(data, country = "India"), gdp_capita = gpd/pop), gdp_capita))
```

What's more readable?

```
data %>%
  filter(country == "India") %>%
  mutate(gdp_capita = gdp/pop) %>%
  select(gdp_capita) %>%
  plot()
```

windows: Ctrl + shift + m
mac: Command + shift + m

filter()

```
babynames_riley_only <- babynames %>%
  filter(name == "Riley")
babynames_riley_only
```

FIGURE 3.1: Diagram of filter() rows operation.

Subset observations based on rules

Logical operators

`>`
“greater than”

`<`
“less than”

`|`
“OR”

`>=`
“greater than or equal to”

`&`
“AND”

`<=`
“less than or equal to”

`==`
“equals to”

`!=`
“not equals to”

Practice

Install `nycflights` package

Load nycflights package

Let’s filter different observations

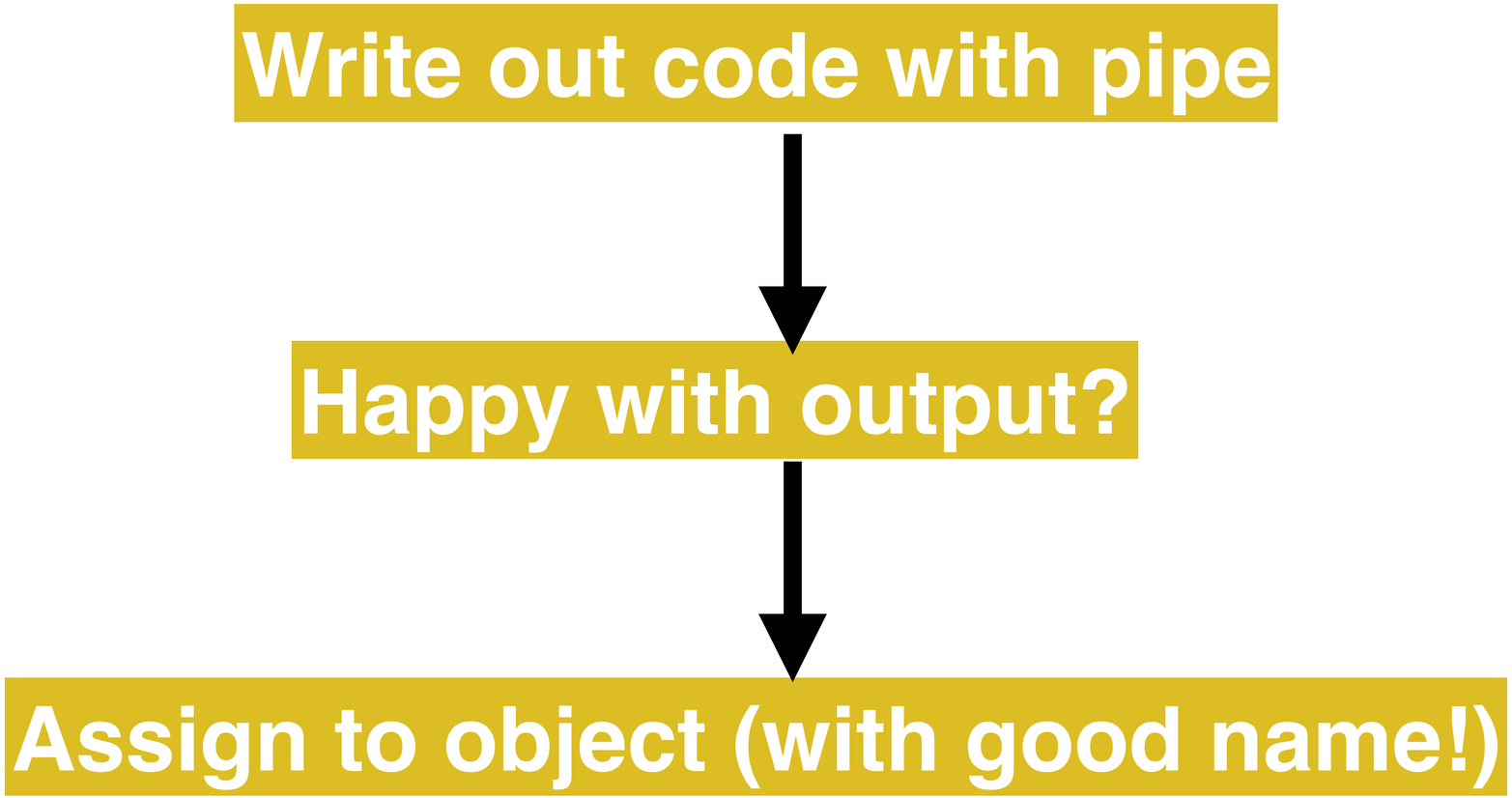
mutate()

Make New Variables

FIGURE 3.5: Diagram of mutate() columns.

```
# clean up movie data
movies_clean =
  movies %>%
  filter(year >= 1980) %>%
  mutate(profit = gross - budget) %>%
  mutate(profit = profit/1000000)
movies_clean
```

Workflow


Write out code with pipe
Happy with output?
Assign to object (with good name!)

Mutate + case_when()

Create more complex variables

```
> df
# A tibble: 4 x 1
  numbers
    <int>
1       1
2       2
3       3
4       4
```

```
df2 = df %>%
  mutate(even_odd = case_when(numbers == 1 ~ "Odd",
                              numbers == 2 ~ "Even",
                              numbers == 3 ~ "Odd",
                              numbers == 4 ~ "Even"))
```

```
# A tibble: 4 x 2
  numbers even_odd
    <int> <chr>
1       1 Odd
2       2 Even
3       3 Odd
4       4 Even
```

You *will* forget; look it up!