

# Visualizing Your Data: Part 2

*Class #11 | GEOG 215*

Intro To Spatial Data Science



*“The greatest value of a picture is when it forces us to notice what we never expected to see.” -John Tukey*

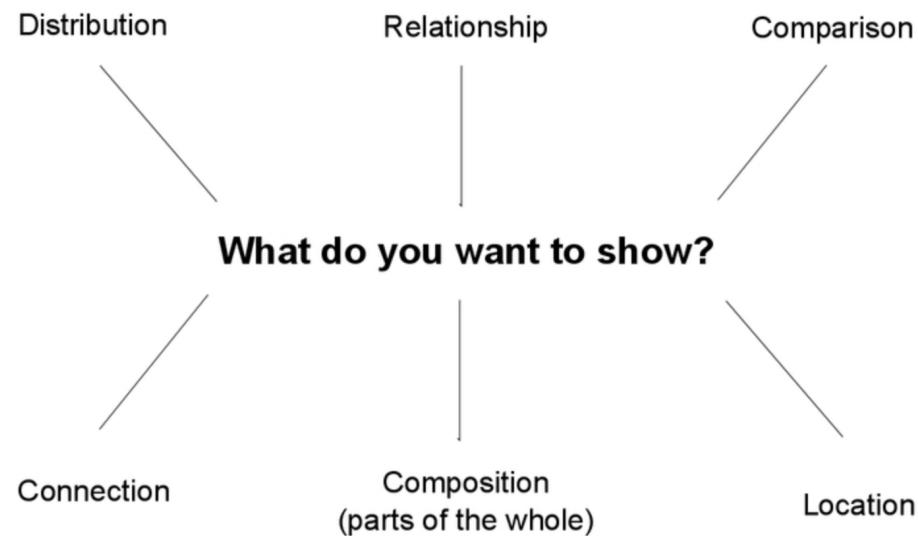
# Today's Class

## Visualizing data multiple ways

- Focus on seeing what pops out
- Develop an intuition for visualizations

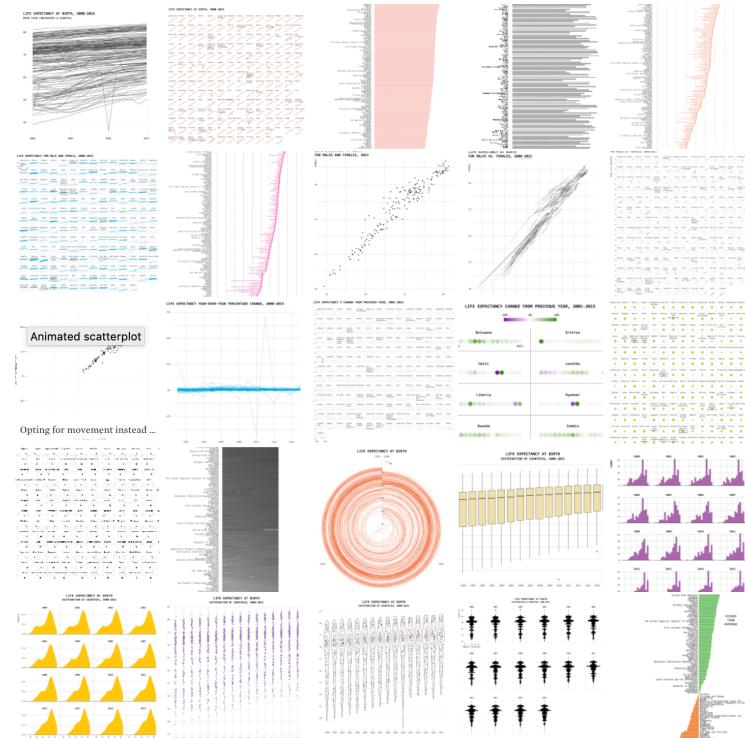
## Maps, Maps and Maps

# Recap



Source: <https://paldhous.github.io/ucb/2016/dataviz/week2.html#>

# Let the data speak



Source: flowing data

# Lets make our data speak

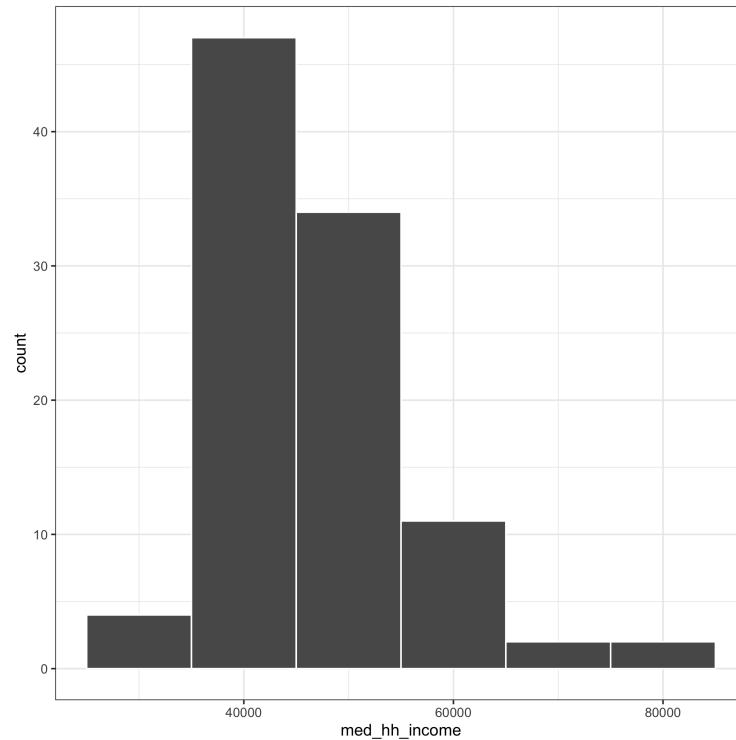
# Distribution

What is the composition of my variable?

- Are values concentrated/spread out
- Are there any breaks/ outliers
- Influences the questions you want to ask/ story you want to tell

# Distribution

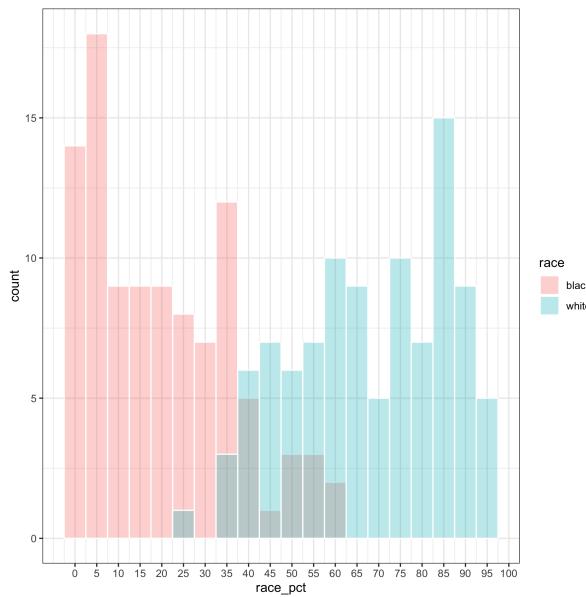
Continuous Variable



# Distribution

Continuous Variable, multiple categories

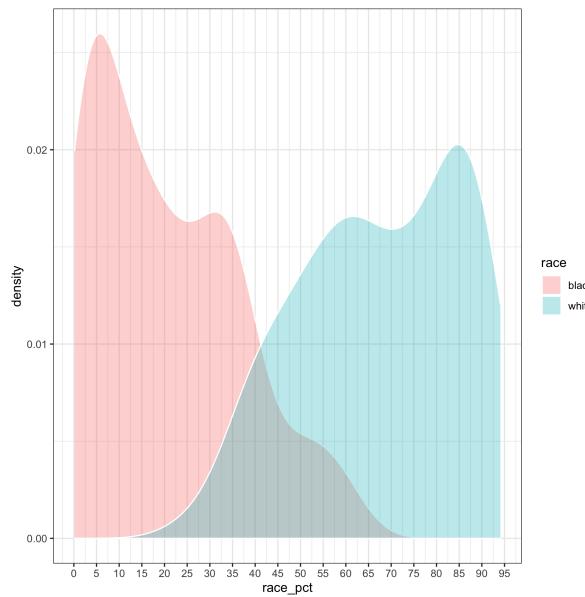
```
nc_tidy1 %>% filter(race %in% c("white", "black")) %>%
  ggplot(aes(x = race_pct, fill = race)) +
  geom_histogram(binwidth = 5 ,col = "white",
                 alpha = 0.3,position = "identity") +
  scale_x_continuous(breaks = seq(0, 100, 5))+  
  theme_bw()
```



# Distribution

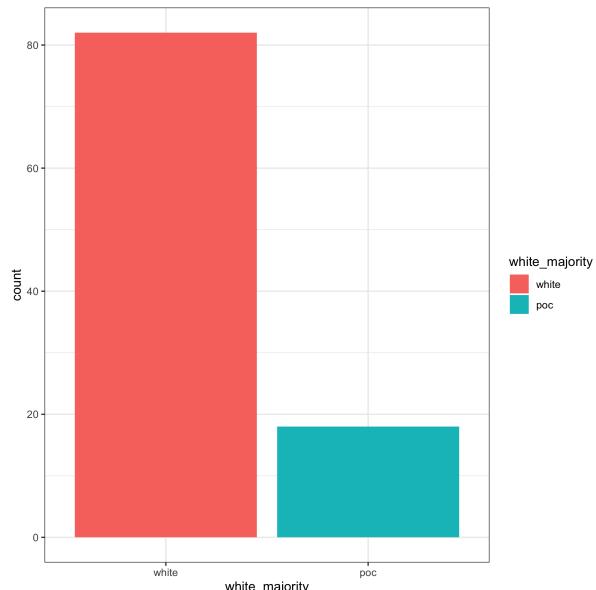
Continuous Variable, multiple categories - density

```
nc_tidy1 %>% filter(race %in% c("white", "black")) %>%
  ggplot(aes(x = race_pct, fill = race)) +
  geom_density(binwidth = 5, col = "white",
               alpha = 0.3, position = "identity") +
  scale_x_continuous(breaks = seq(0, 100, 5)) +
  theme_bw()
```



# Distribution

Categorical variable



```
nc_wide %>%
  ggplot() + geom_bar(aes(x = white_majority, fill = white_majority)) +
  theme_bw()
```

# Distribution

## Continuous variable, categorical variable

```
nc_tidy1 %>%  
  ggplot(aes(x = white_majority,y=med_hh_income)) +  
  geom_boxplot() +  
  theme_minimal()
```

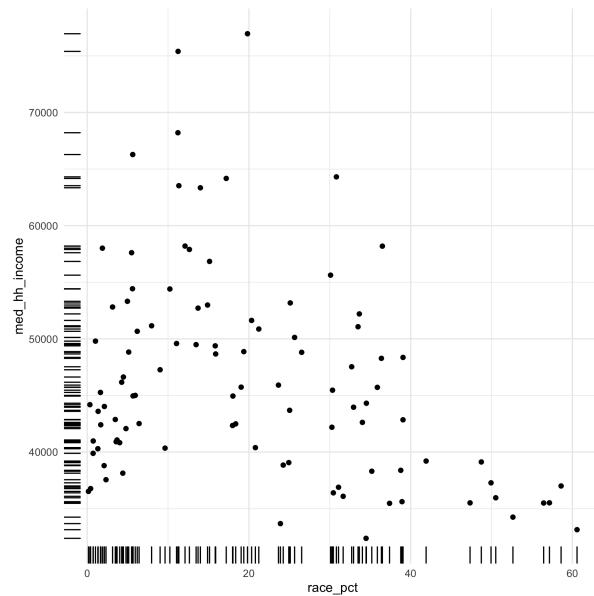
# Relationship

Focus on co-variation

- How do x and y vary together
- Is there a trend?
- *Correlation is not equal to Causation*

# Relationship

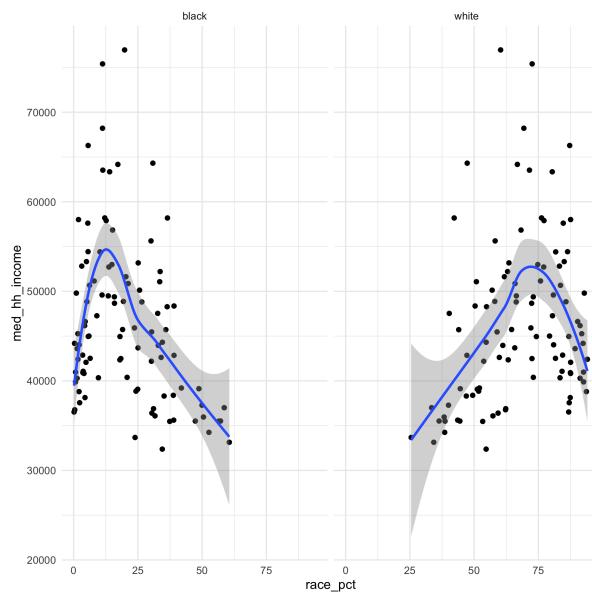
```
nc_tidy1 %>% filter(race == "black") %>%
  ggplot(aes(x = race_pct, y=med_hh_income)) +
  geom_point() + geom_rug() +
  theme_minimal()
```



# Relationship

Comparison between 2 categories

```
nc_tidy1 %>% filter(race %in% c("black", "white")) %>%
  ggplot(aes(x = race_pct, y=med_hh_income)) +
  geom_point() + geom_smooth() + facet_wrap(~race) +
  theme_minimal()
```



# Comparison

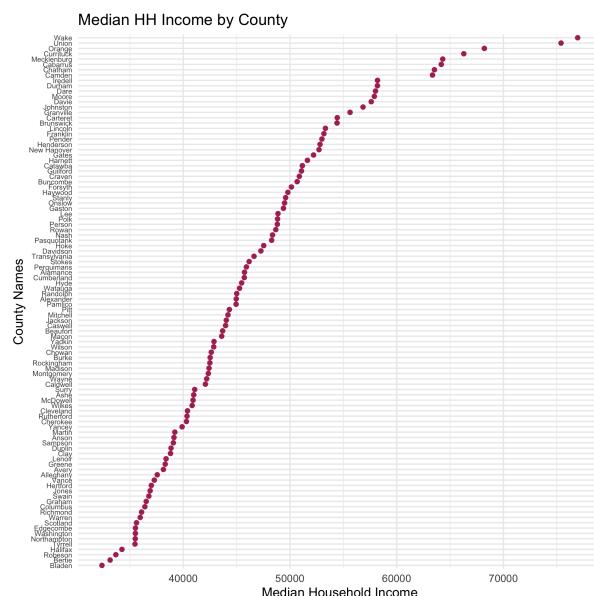
Focus on spotting differences

- Differences across categories
- Differences across time
- Differences across units

# Comparison

## Differences across units

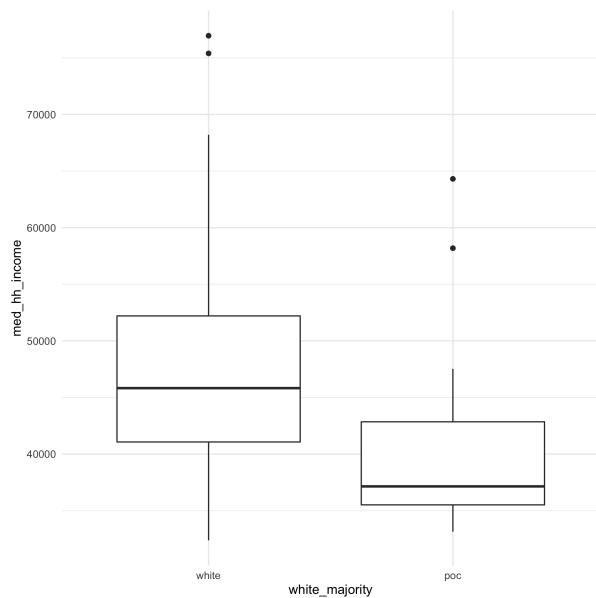
```
nc_wide %>% mutate(name = name %>% str_replace(" County, North Carolina", "")) %>%
  ggplot(aes(y = reorder(name, med_hh_income), x=med_hh_income)) +
  geom_point(color = "maroon") + labs(x = "Median Household Income",
                                       y = "County Names",
                                       title = "Median HH Income by County") +
  theme_minimal() + theme(axis.text.y = element_text(size = 6))
```



# Comparison

Differences across categories

```
nc_tidy1 %>%  
  ggplot(aes(x = white_majority,y=med_hh_income)) +  
  geom_boxplot() +  
  theme_minimal()
```

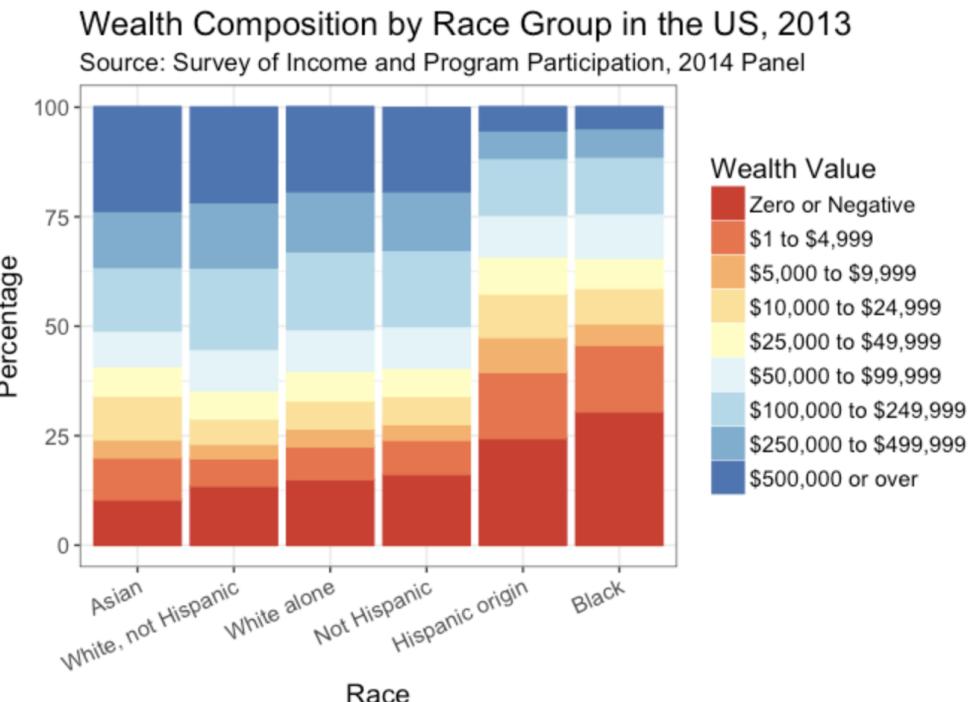


# Composition

Focus on parts of a whole

- Focus on relative percentages
- Generally not on counts

# Composition



Source: Simon Halliday

# Composition

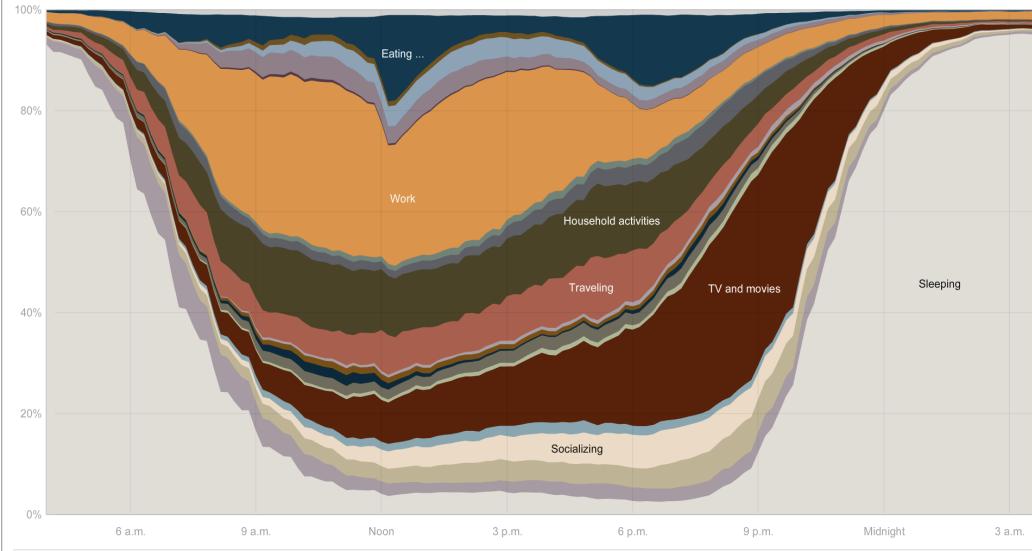
## How Different Groups Spend Their Day

The American Time Use Survey asks thousands of American residents to recall every minute of a day. Here is how people over age 15 spent their time in 2008. [Related article](#)

### Everyone

Sleeping, eating, working and watching television take up about two-thirds of the average day.

Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children



Source: NYtimes

# Composition

Save your pies for Dessert, Please!!!!!!

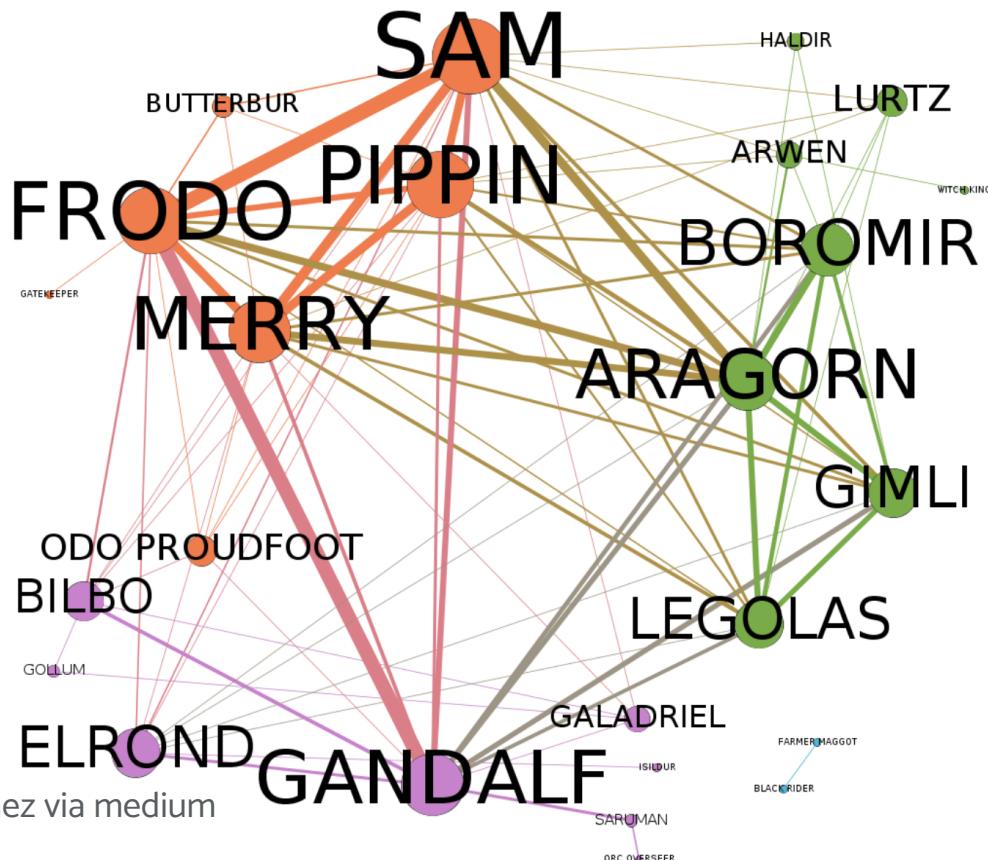
*Source:* Stephen Few

# Connection

Focus on connections

- Who is connected to who?
- Focus on the edge, rather than the vertex

# Connection



Source: Tiago Paez via medium

# *Location*

# Why do we create maps

Allows for visual interpretation of geographic patterns of phenomena

- Location of objects/events
  - Themselves (e.g., clustered? where?)
  - Or, in relation to other things
- Variation of a type or value
  - Categories
  - Quantities

# Map Types

- Point map
- Proportional Symbol map
- Choropleth map

# Point Map

Variable of interest is represented as set of point features

- e.g., disease events, facilities, wells

Symbology is held constant (all points have the same shape, color, and size)

- Visual interpretation of the density of events (per area) via the locations of points on the map

# Point Map

```
ggplot() + geom_sf(dat = nc_shp) +  
  geom_sf(data = cafo_shp,  
          col="red",alpha=0.4) + coord_sf(datum = NA) +  
  theme_minimal()
```

# Proportional Symbol Map

Variable of interest is represented as set of

point features, with interval/ratio information

- e.g., number of disease cases

Size of symbol is scaled proportionally to the value of the attribute

# Proportional Symbol Map

```
ggplot() + geom_sf(dat = nc_shp) +  
  geom_sf(data = cafo_shp %>% mutate(allowable_count > 0), aes(size = allowable_co  
    col="maroon",alpha=0.7) + coord_sf(datum = NA) +  
    scale_size_area() +  
    theme_minimal()
```

# Choropleth Map

Variable of interest represented by

ratios/rates/proportions located within areal features (polygons)

- We generally do not use counts
- Areas are shaded with different colors, patterns, or intensities
  - Shows geographic variation
  - Requires classifying or binning the data values

# Choropleth Map

Lets look at examples from lab 3

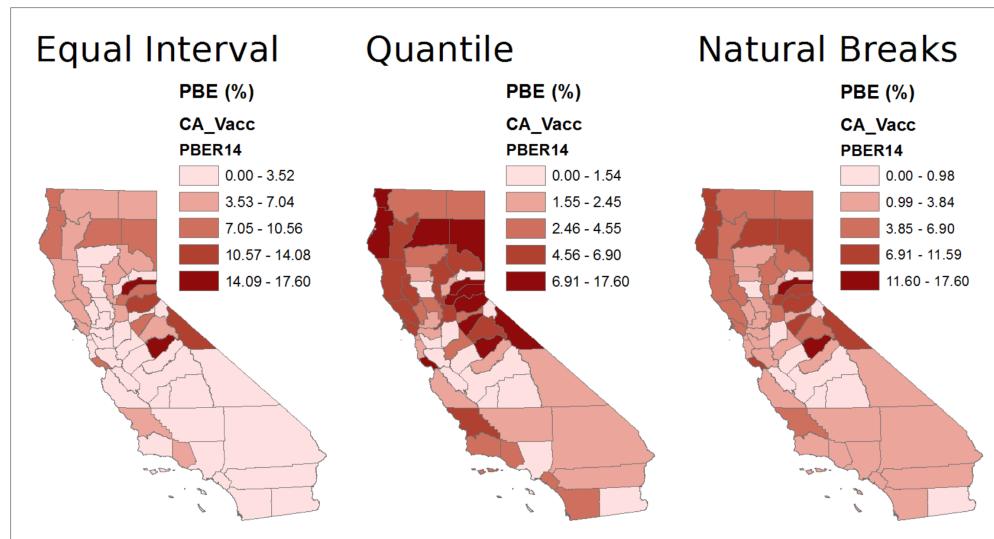
# Choropleth Map

## Data Classification

- Our eyes are not good at discerning more than 6- 8 different shades of color in one map
- Values (attribute that is being mapped) are classified or binned into classes
- Each class is assigned a specific color along a color ramp - Value or Saturation sequence

# Choropleth Map

## Data Classification



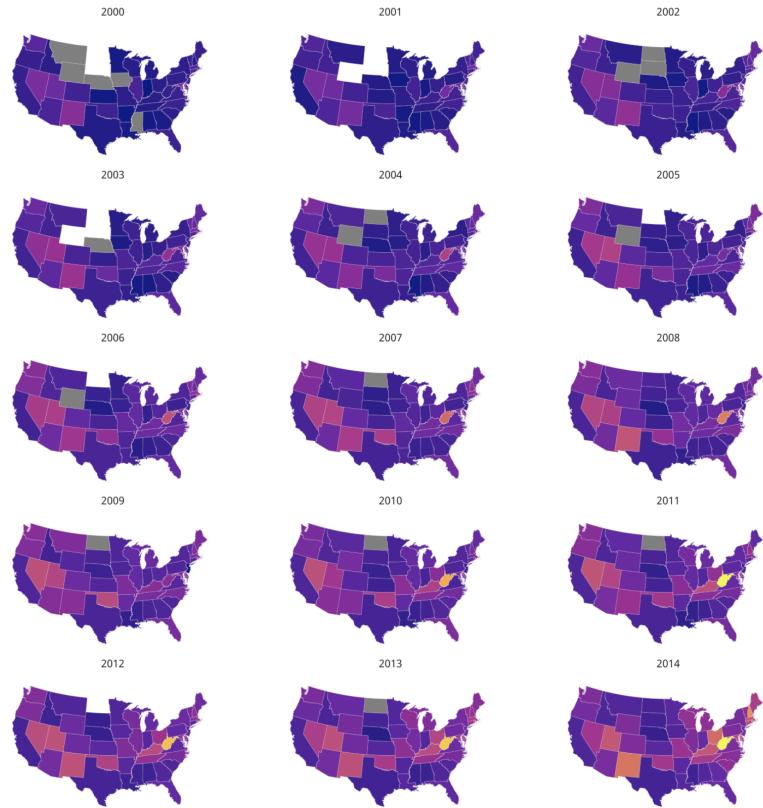
Source: Paul Delamater

# Is your data really spatial?

- What are you interested in?
  - Are you interested in the units of interest?
  - Or, are interested in the geographic distribution of those units of interest

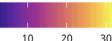
# Is your data really spatial?

Opiate Related Deaths by State, 2000-2014



Source: Kieran Healy

Death rate per 100,000 population



# Is your data really spatial?



*Source: Kieran Healy*

# Choropleth Map

*Source: xkcd*

# Choropleth Map

## Ur-choropleth Maps

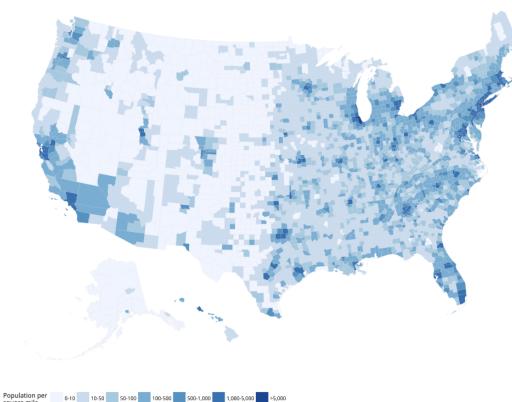


Figure 7.11: US population density by county.

# Choropleth Map

## Ur-choropleth Maps

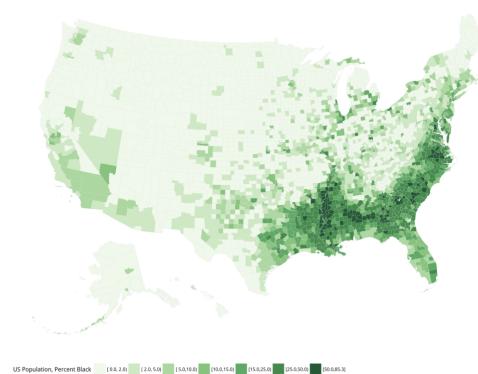
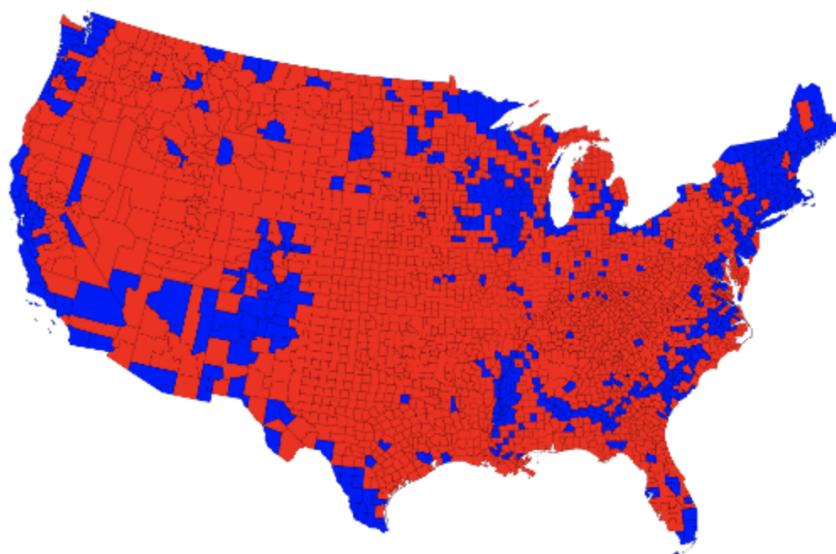


Figure 7.12: Percent Black population by county.

# Choropleth Map

Example



# Cartograms

