

Visualizing Your Data: Part 1

Class #10 / GEOG 215

Intro To Spatial Data Science

Today's Class

The Fun Part

Today's Class

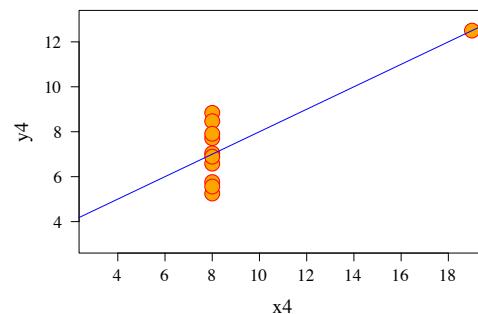
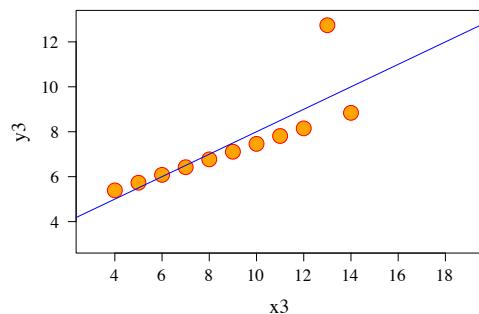
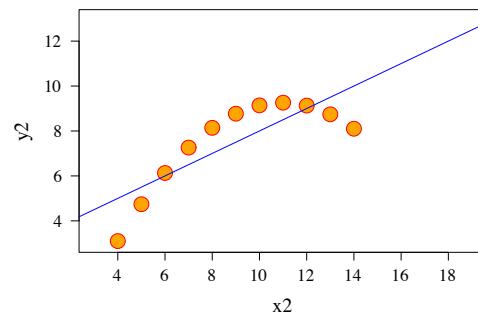
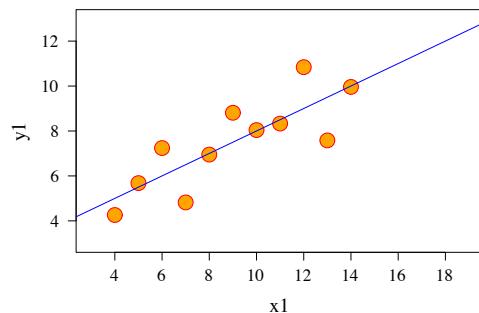
The Fun Part (For real)

- . Data Visualizations

WHY VISUALIZE DATA

A motivating Example

Anscombe's quartet



A motivating Example

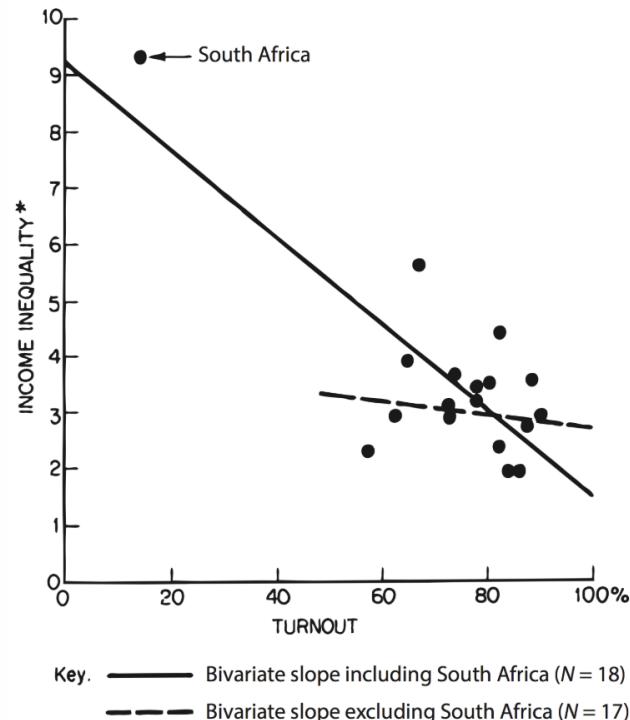
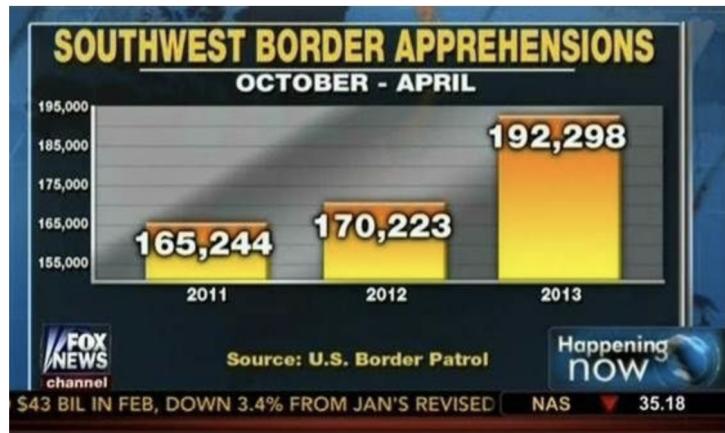


Figure 1.2: Seeing the effect of an outlier on a regression line.

Source: Jackson, 1980: *American Sociological Review*

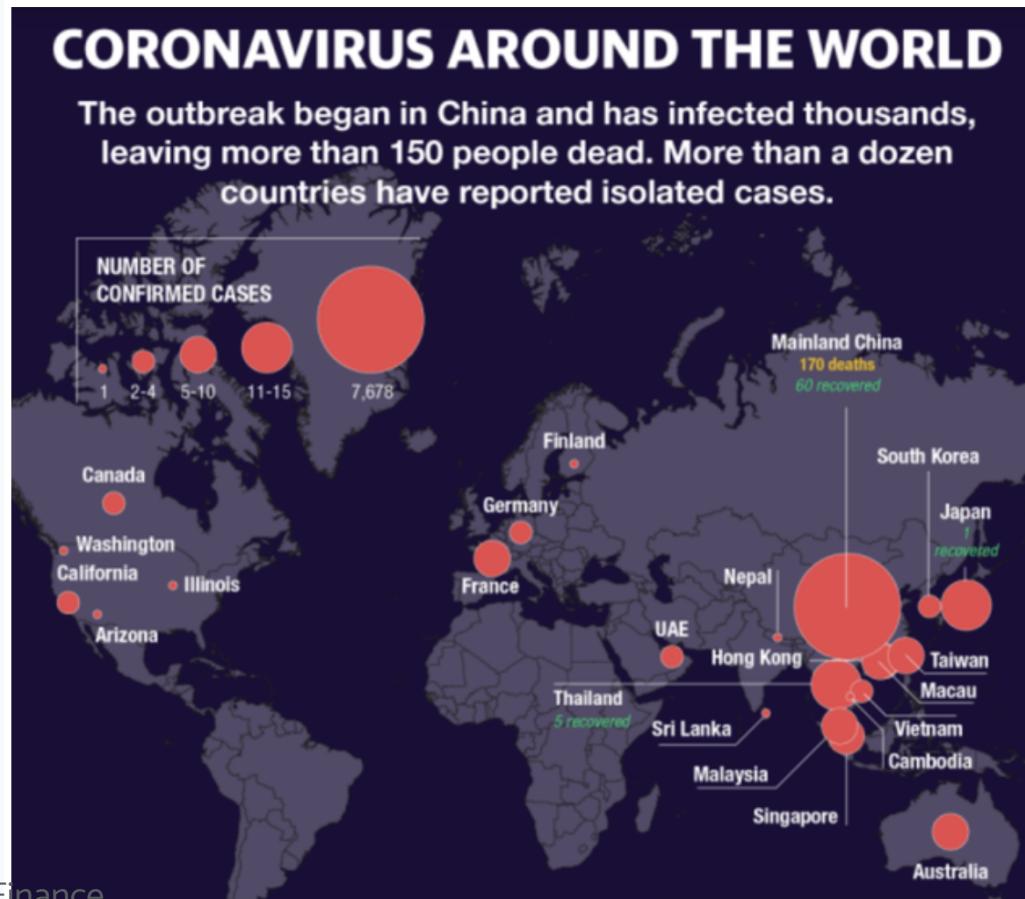
WHY VISUALIZE CORRECTLY

Bad Taste

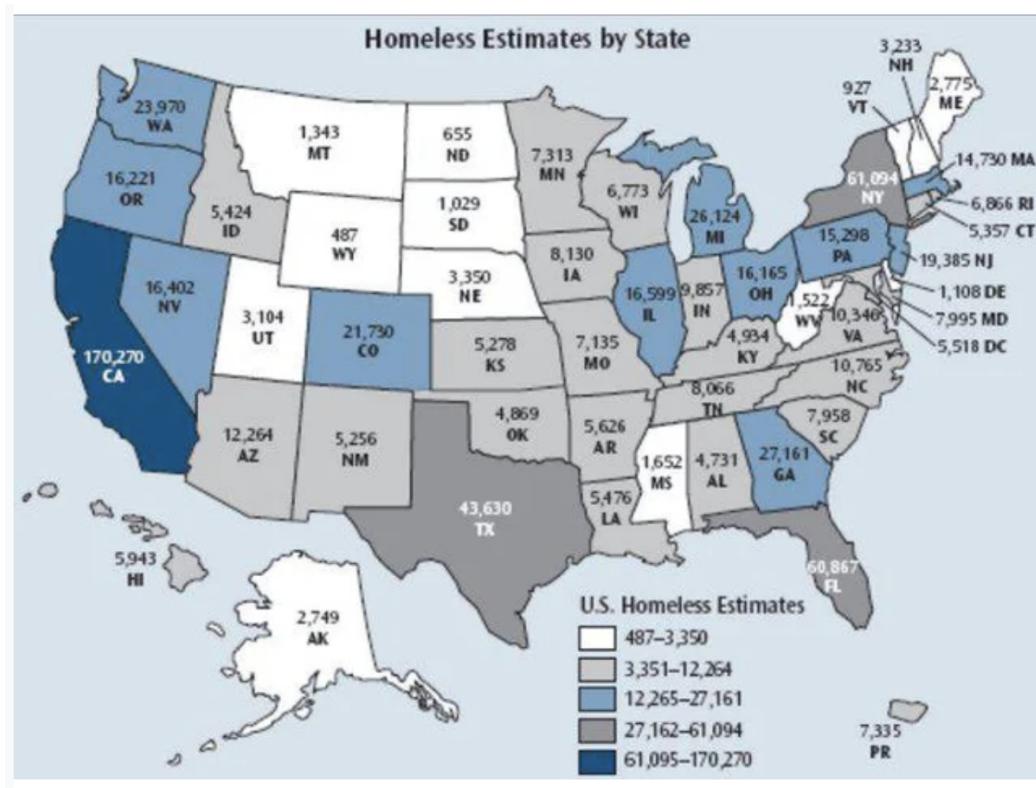


Source: Media Matters for America

Your Turn (figure 1)



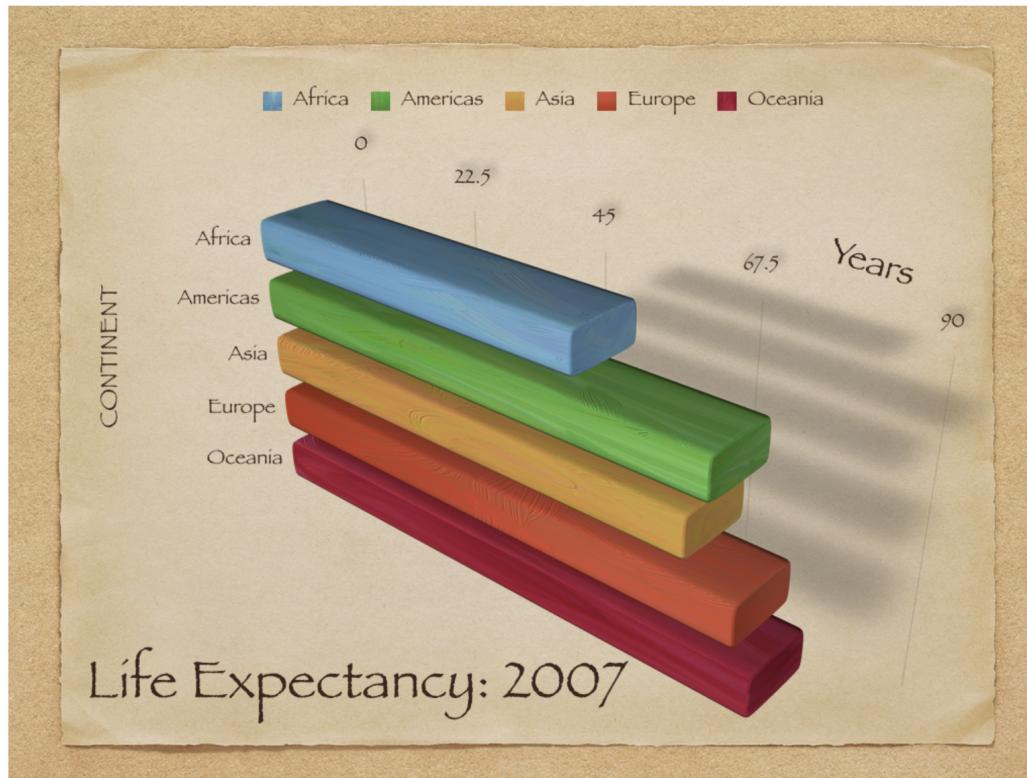
Your Turn (figure 2)



Source: <http://www.pbs.org/now/shows/305/homeless-map.html>

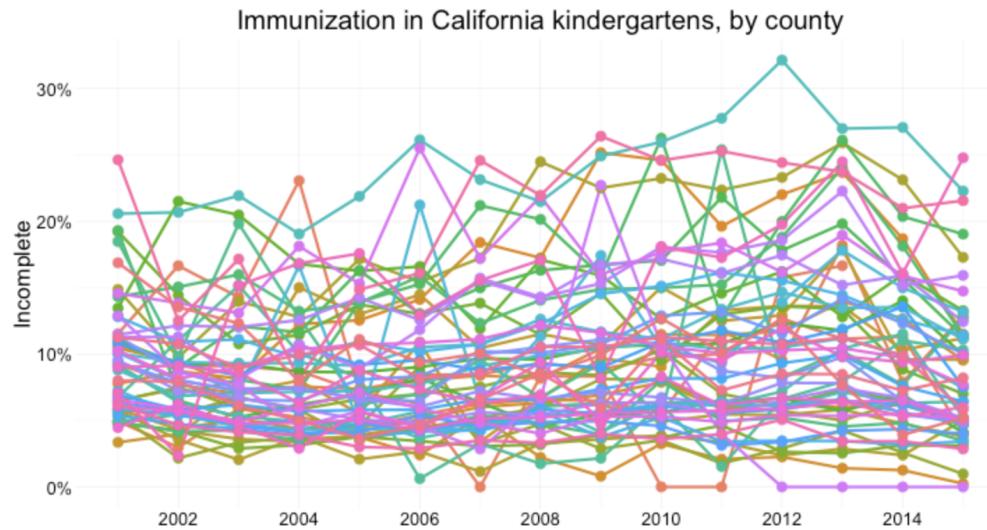
WHY VISUALIZE NICELY?

Bad Perception



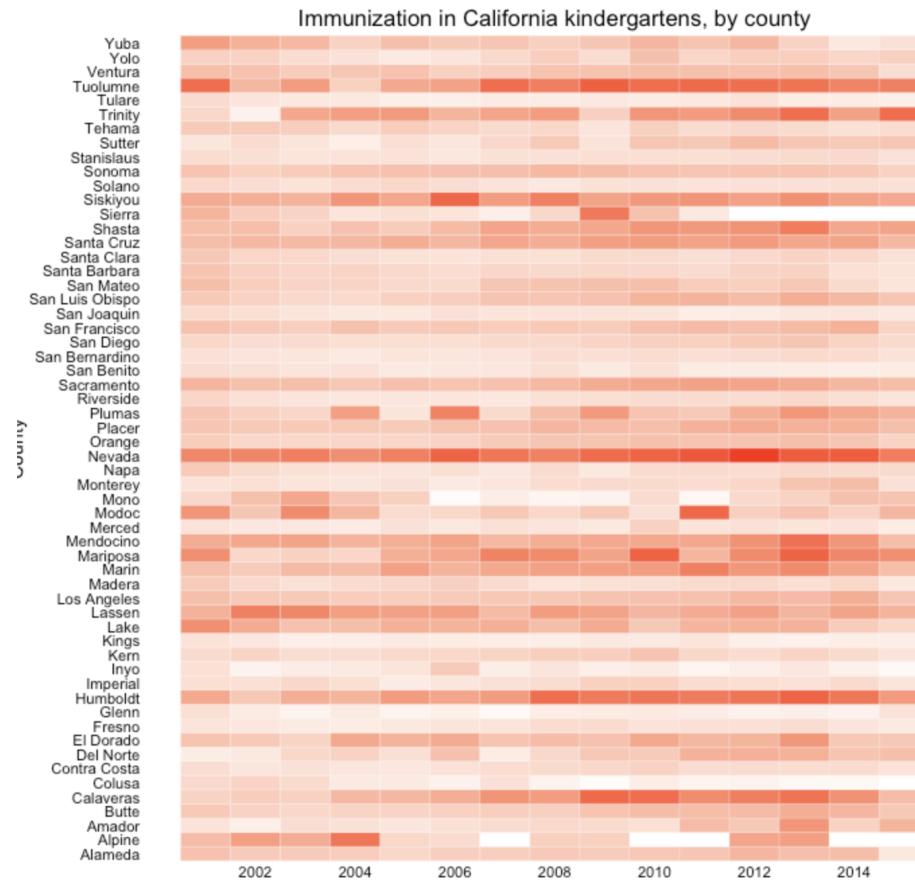
Source: <http://socviz.co/lookatdata.html#why-look-at-data>

Pick your battles or Risk Sensory Overload



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

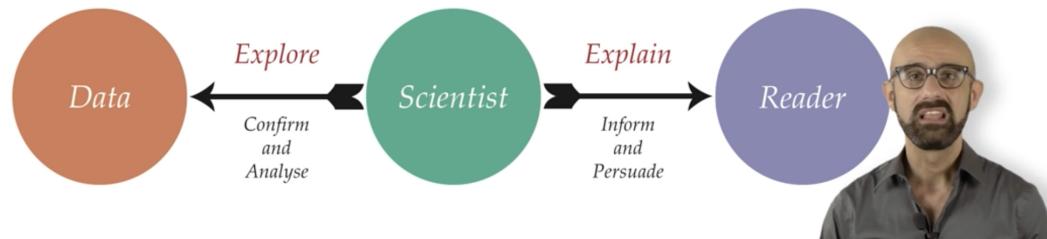
Pick your battles or Risk Sensory Overload



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

What is the Purpose of visualization

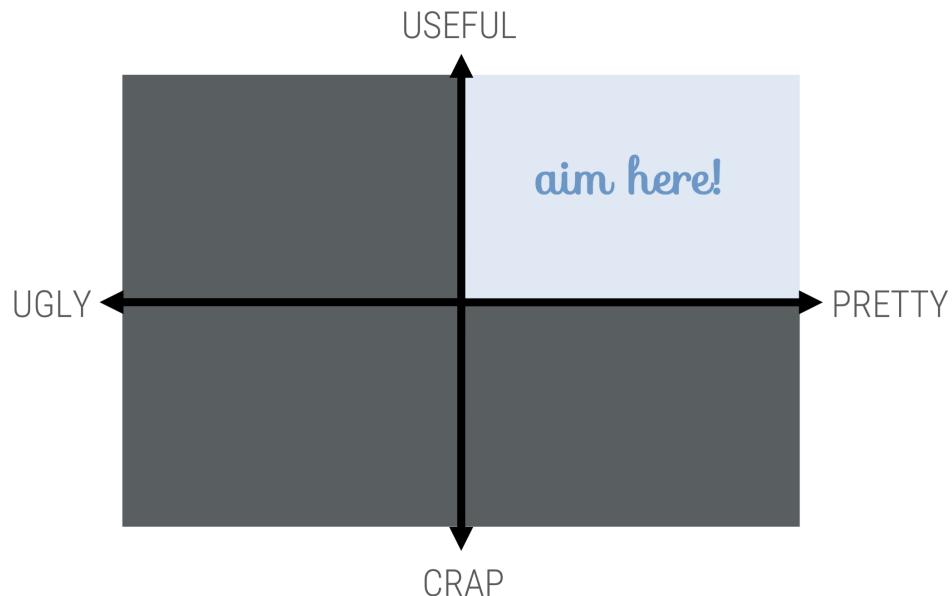
Exploratory VS Explanatory visualization



Source: DataCamp, Rick Scavetta

What is the Purpose of visualization

Exploratory VS Explanatory visualization



Source: Jackie Wirz, from <https://apreshill.github.io/data-vis-labs-2018/>

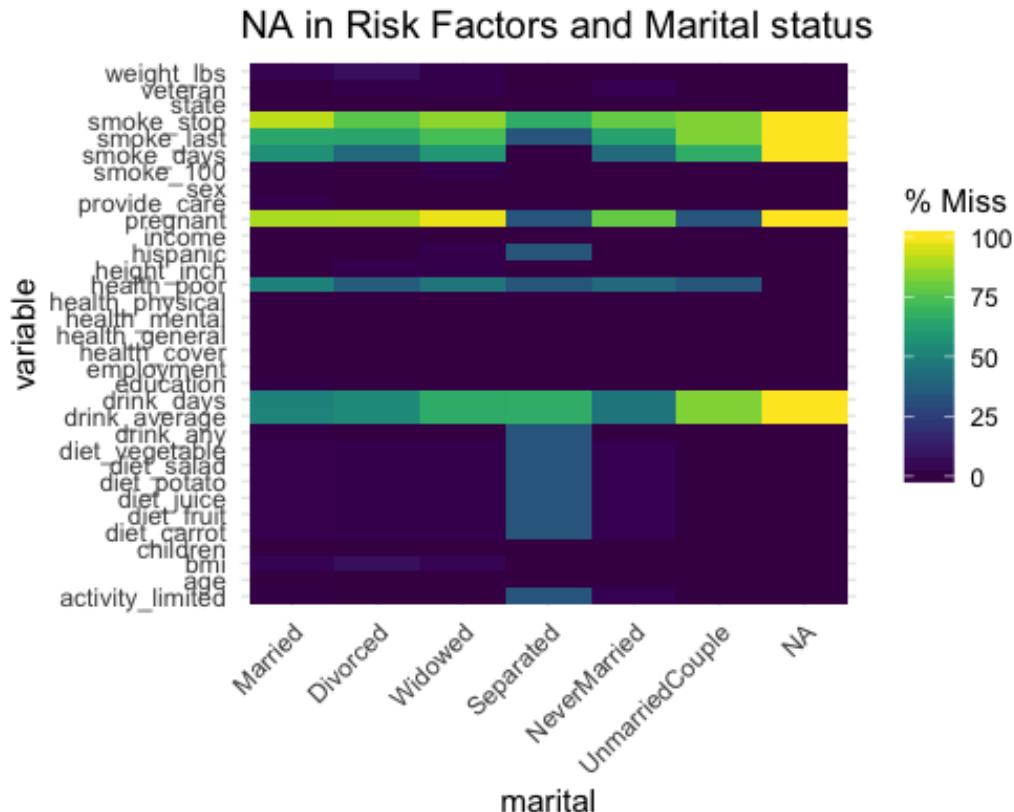
Data Visualization for Exploratory Analysis

Data Cleaning

- Testing for obvious inconsistencies and errors in your data
- Think about what your data represents
 - Are there any missing values
 - Values out of range
 - Are there any outliers

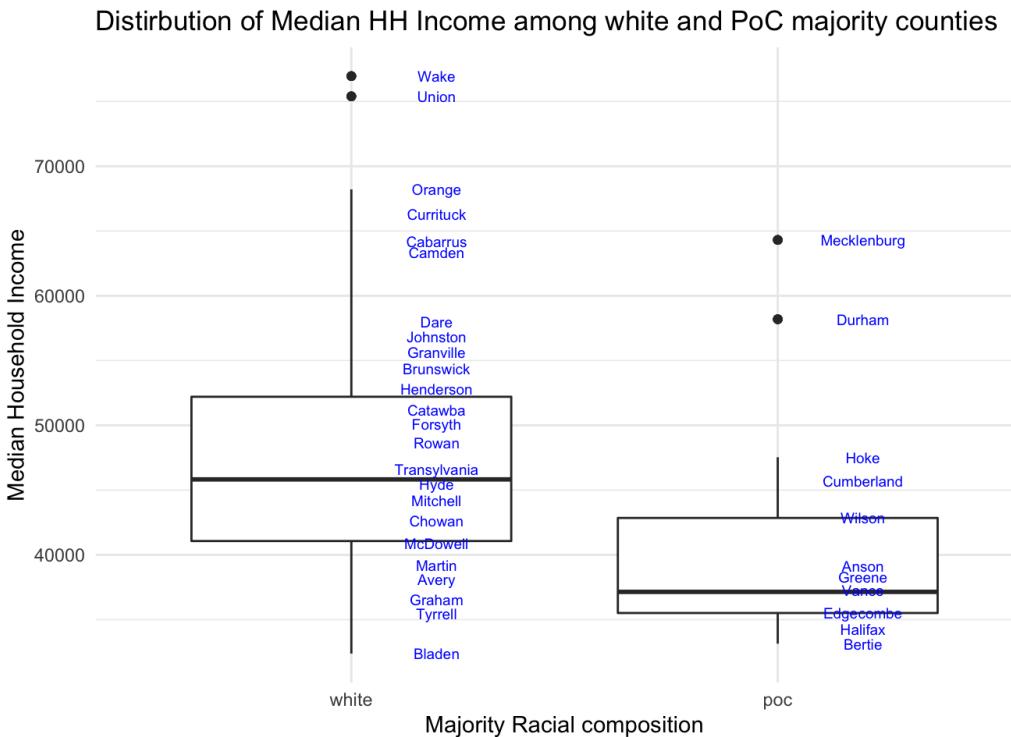
Data Cleaning

Detect Missing Data



Data Cleaning

Detect Outliers



Data Visualization for Exploratory Analysis

Looking for 'structure'/'signal' in the data/'noise'

- What "pops" out?
- Are there obvious patterns or is the data random
- Visualize data multiple ways to generate possible explanations to explore
- Iterate/Repeat/Update

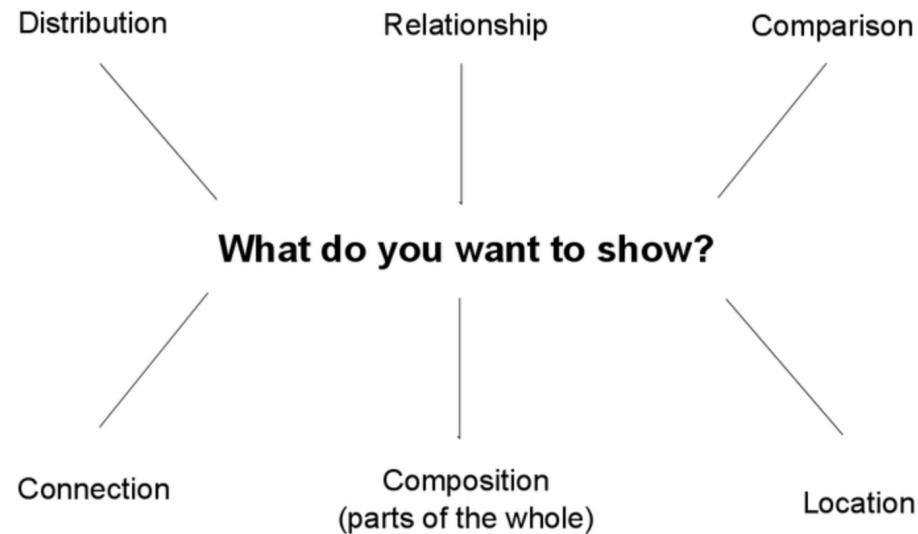
Think about the visualization based on the type of data/variable

Table 2.1: Types of variables encountered in typical data visualization scenarios.

Type of variable	Examples	Appropriate scale	Description
quantitative/numerical continuous	1.3, 5.7, 83, 1.5×10^{-2}	continuous	Arbitrary numerical values. These can be integers, rational numbers, or real numbers.
quantitative/numerical discrete	1, 2, 3, 4	discrete	Numbers in discrete units. These are most commonly but not necessarily integers. For example, the numbers 0.5, 1.0, 1.5 could also be treated as discrete if intermediate values cannot exist in the given dataset.
qualitative/categorical unordered	dog, cat, fish	discrete	Categories without order. These are discrete and unique categories that have no inherent order. These variables are also called <i>factors</i> .
qualitative/categorical ordered	good, fair, poor	discrete	Categories with order. These are discrete and unique categories with an order. For example, "fair" always lies between "good" and "poor". These variables are also called <i>ordered factors</i> .
date or time	Jan. 5 2018, 8:03am	continuous or discrete	Specific days and/or times. Also generic dates, such as July 4 or Dec. 25 (without year).
text	The quick brown fox jumps over the lazy dog.	none, or discrete	Free-form text. Can be treated as categorical if needed.

Source: <https://serialmentor.com/>

Think about the visualization based on what you want to show



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

Visualizing using ggplot2()

ggplot2()

Based on the Grammar of Graphics

- A graph is made of several **layers**
- Grammar (i.e mathematical and aesthetic rules define layers)

ggplot2 layers

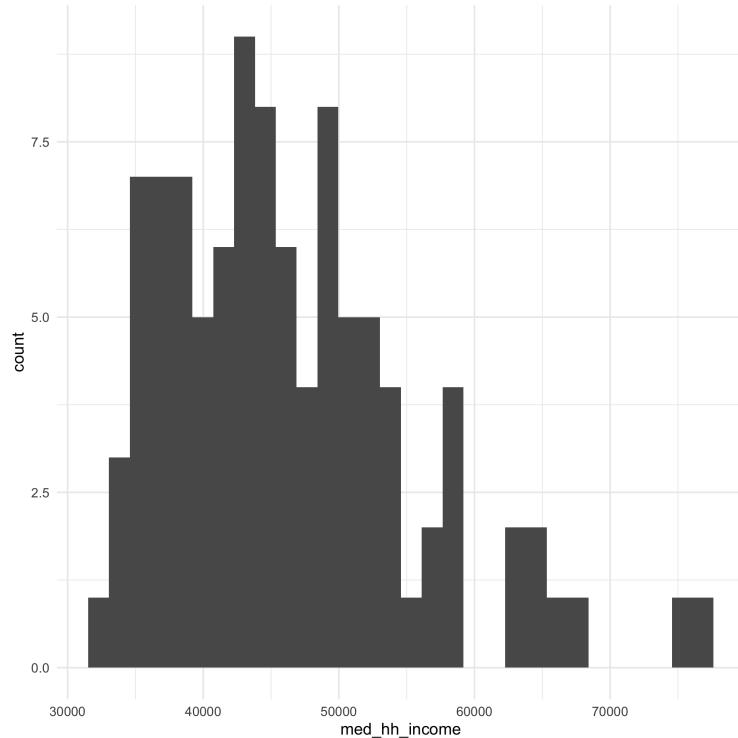


For this class, we will mainly focus on:

- Data
- Aesthetics
- Geometries
- Facets

Example:

Histogram



```
ggplot(data = nc_wide) +  
  geom_histogram(aes(x = med_hh_income)) + theme_minimal()
```

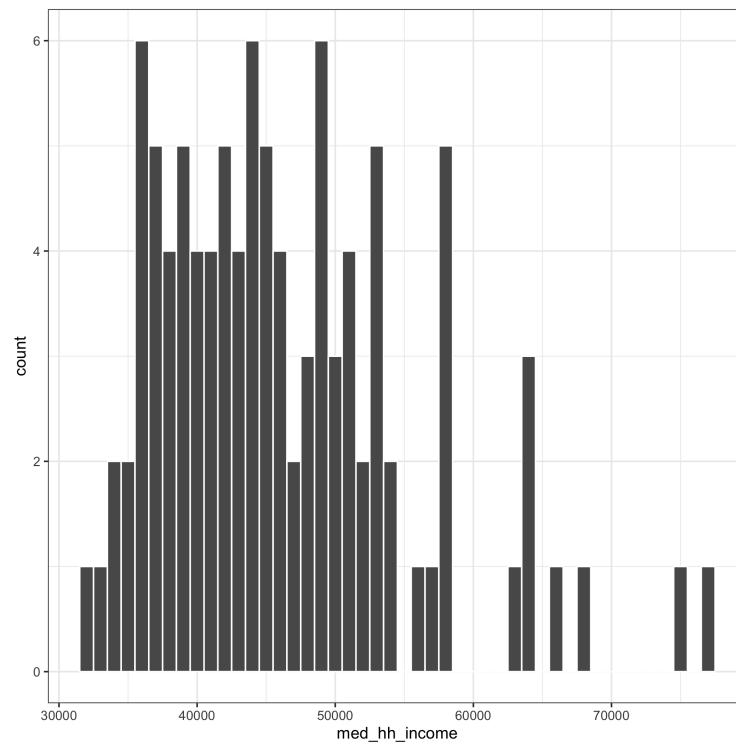
Constructing a ggplot2

```
ggplot(data = <DATA>) +  
<GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

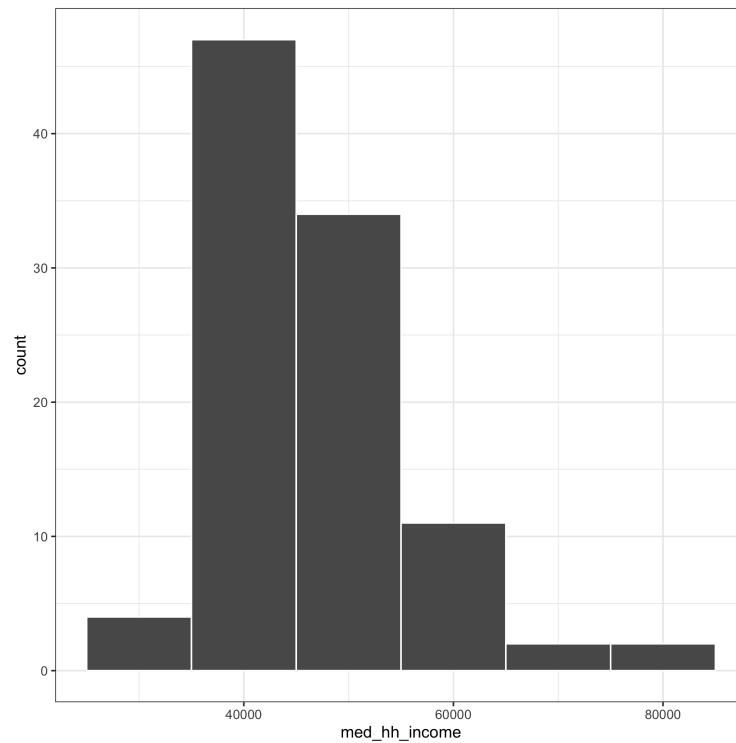
1. Initialize the data

1. add (at least) one geometry – with +
2. specify required Aesthetics – with aes()

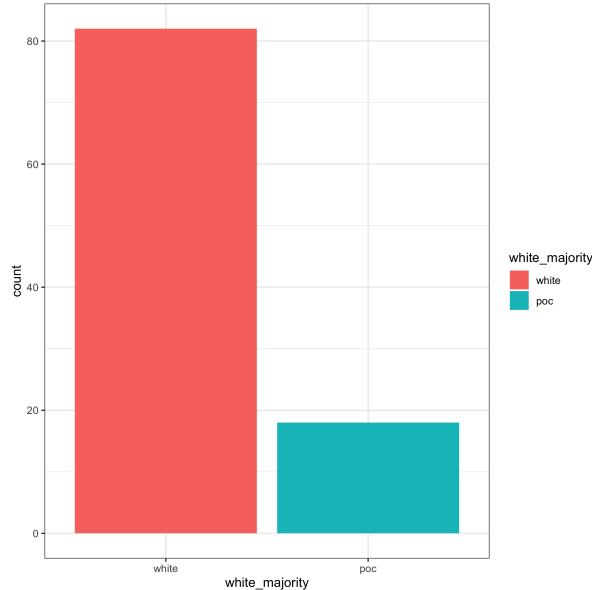
Lets look at our histogram again



Lets look at our histogram again



Distributing categorical data



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

Aesthetics in ggplot2

- x
- y
- colour
- fill
- size
- alpha
- linetype
- labels
- shape

Next Class

- Data Visualization 2 (including maps)
 - email a few visualizations, we will scrutinize them
- HW 1 doubts

