A Data Story: Cholera in 1854

School of Information Studies
Syracuse University

Jeff Hemsley, PhD
Assistant Professor
School of Information Studies

A Story of Patterns

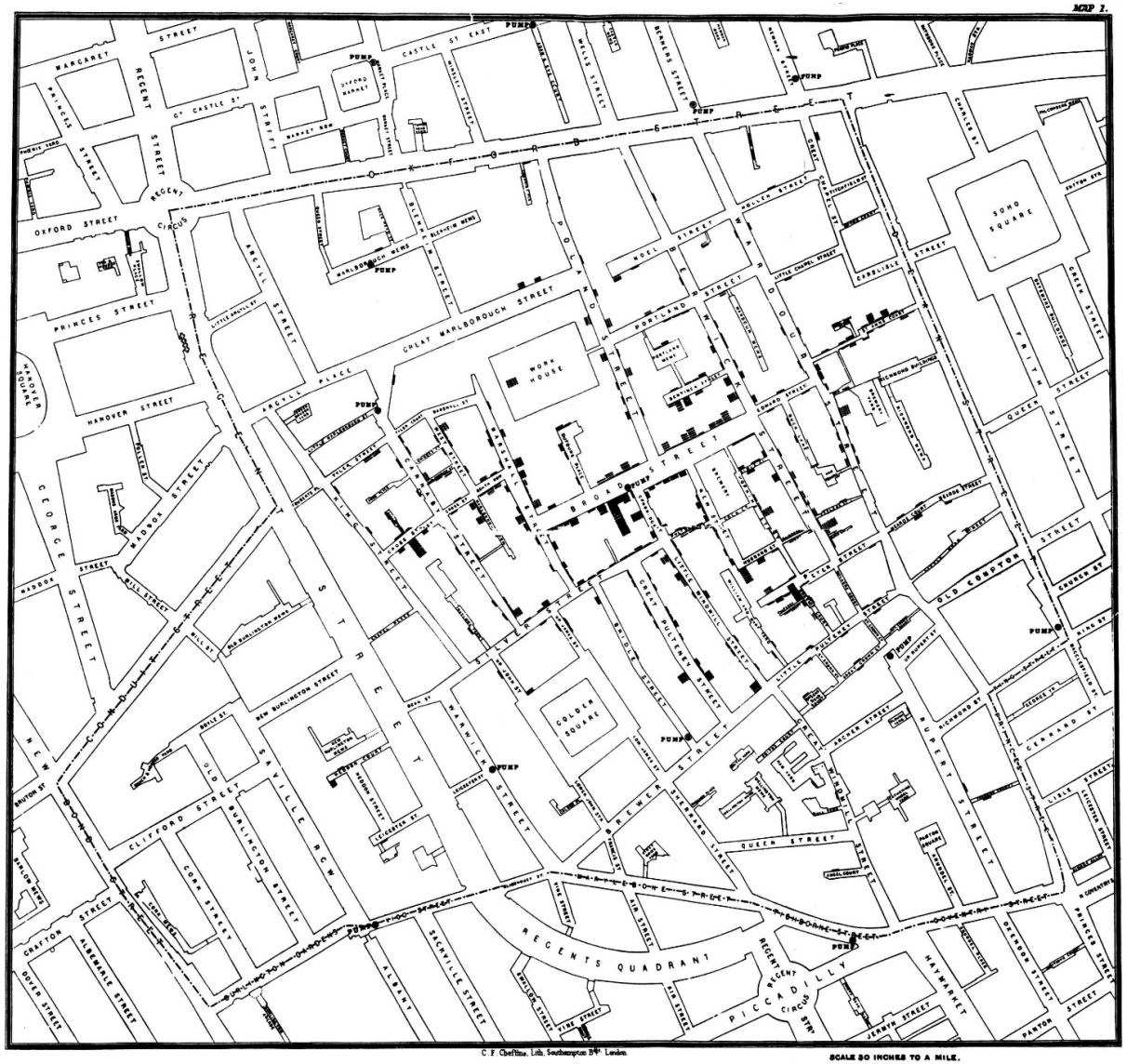
John Snow

Cholera outbreak on August 31, 1854

Data: 83 deaths from cholera

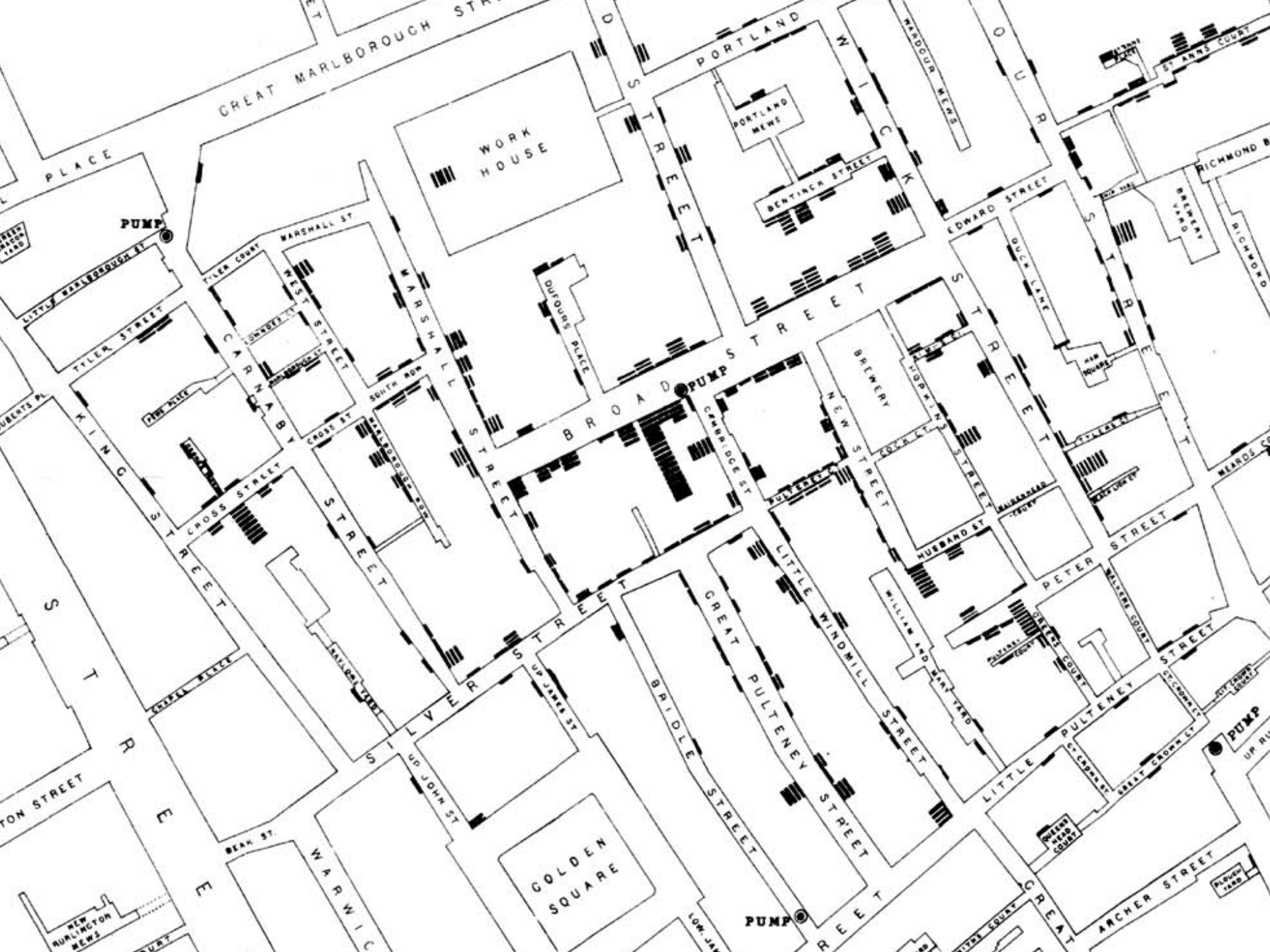
Motivating questions:

- Was there a pattern to the outbreak?
  Reduce uncertainty
- Could the pattern tell us anything about the cause of the outbreak?
  Look for exceptions
- Could impurities in water be to blame?
  Hypothesis testing



Commons

School of Information Studies
Syracuse University


GREAT MARLBOROUGH STREET
PLACE
PUMP
LITTLE MARLBOROUGH ST
TYLER STREET
TYLER COURT
MARSHALL ST
WEST STREET
CHAPEL
KING STREET
CROSS STREET
CHAPEL PLACE
WORK HOUSE
DUFOURS PLACE
MARSHALL STREET
BROAD STREET
PUMP
CAMBRIDGE ST
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
HARDGUR MEWS
WICK STREET
EDWARD STREET
BREWERY
NEW STREET
HOPKINS STREET
HUSBAND ST
PETER STREET
LITTLE PULTENEY STREET
GREAT PULTENEY STREET
BRIDLE STREET
LITTLE WINDMILL STREET
SILVER STREET
UP JAMES ST
UP JOHN ST
GOLDEN SQUARE
PUMP
WARWICK
BEAK ST.
NEW BURLINGTON MEWS
ARCHER STREET
PLOUGH YARD
QUEEN HEAD COURT
GREAT CROWN CT
LITTLE PULTENEY STREET
ST ANNS COURT
RICHMOND


GREAT MARLBOROUGH STREET
PLACE
PUMP
LITTLE MARLBOROUGH ST
TYLER STREET
TYLER COURT
MARSHALL ST
WORK HOUSE
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
HARDWICK STREET
HARDWICK MEWS
EDWARD STREET
DUCK LANE
BREWERY YARD
RICHMOND
ST ANNS COURT
DUFOURS PLACE
BROAD STREET
MARSHALL STREET
SOUTH ROW
CAMBRIDGE ST
BREWERY
NEW STREET
FOUR CT
PULTENEY
HUSBAND ST
KING STREET
CROSS STREET
CHAPEL PLACE
GOLDEN SQUARE
UP JOHN ST
BRIDE STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
PETER STREET
LITTLE PULTENEY STREET
GREAT CROWN CT
QUEEN HEAD COURT
ARCHER STREET
PLOUGH YARD
NEW BURLINGTON MEWS
BEAK ST
WARWICK
PUMP


GREAT MARLBOROUGH STREET
PLACE
PUMP
LITTLE MARLBOROUGH ST
TYLER STREET
TYLER COURT
MARSHALL ST
WORK HOUSE
PORTLAND STREET
HARDOUR MEWS
BENTINCK STREET
EDWARD STREET
RICHMOND
DUFOURS PLACE
MARSHALL STREET
BROAD STREET
PUMP
BREWERY
NEW STREET
KING STREET
CROSS STREET
CHAPEL PLACE
BRIDE STREET
LITTLE WINDMILL STREET
GREAT PULTENEY STREET
PETER STREET
LITTLE PULTENEY STREET
GOLDEN SQUARE
WARWICK STREET
BEAK ST.
UP JOHN ST
ARCHER STREET
QUEEN HEAD COURT
PUMP

Finding Data

Jeff Hemsley, PhD
Assistant Professor
School of Information Studies

School of Information Studies
Syracuse University

Finding Data

Universities

- Syracuse University data resources

Government

- U.S. Government's open data
- U.S. climate data
  - Education/data (tab)
- NASA data portal

Others

- Kaggle
- Reddit datasets

Data clean-up

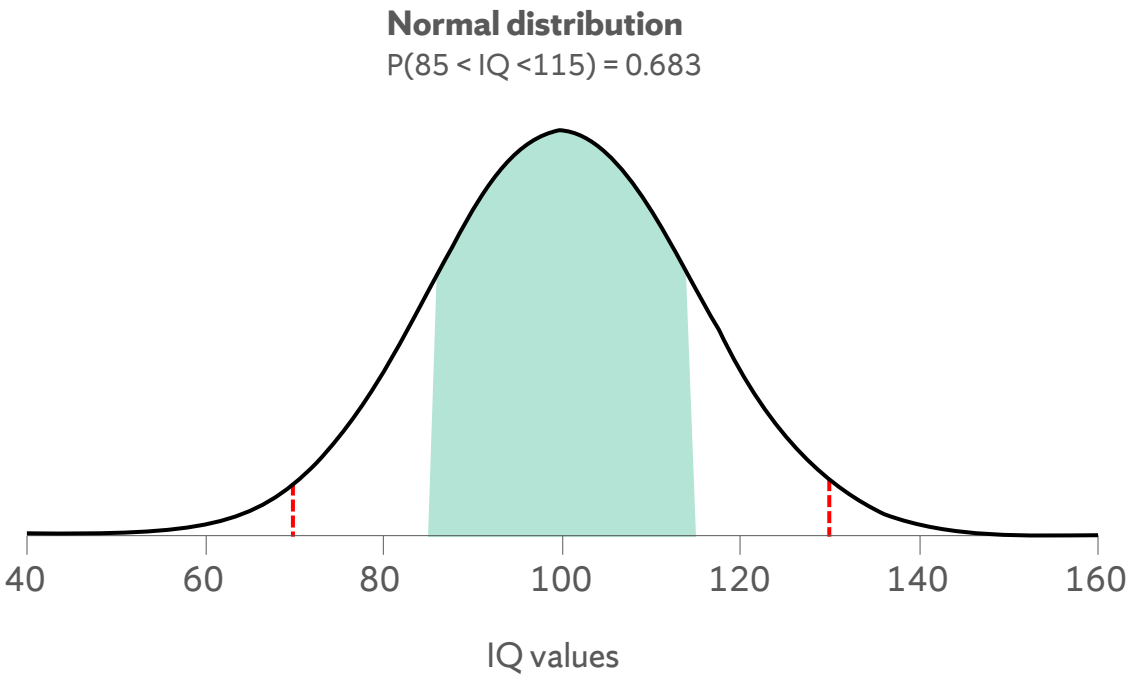
- U.S. Census Bureau (Construction Spending)

Distributions and Plot Dimensions

School of Information Studies
Syracuse University

Jeff Hemsley, PhD
Assistant Professor
School of Information Studies

Normal Distribution

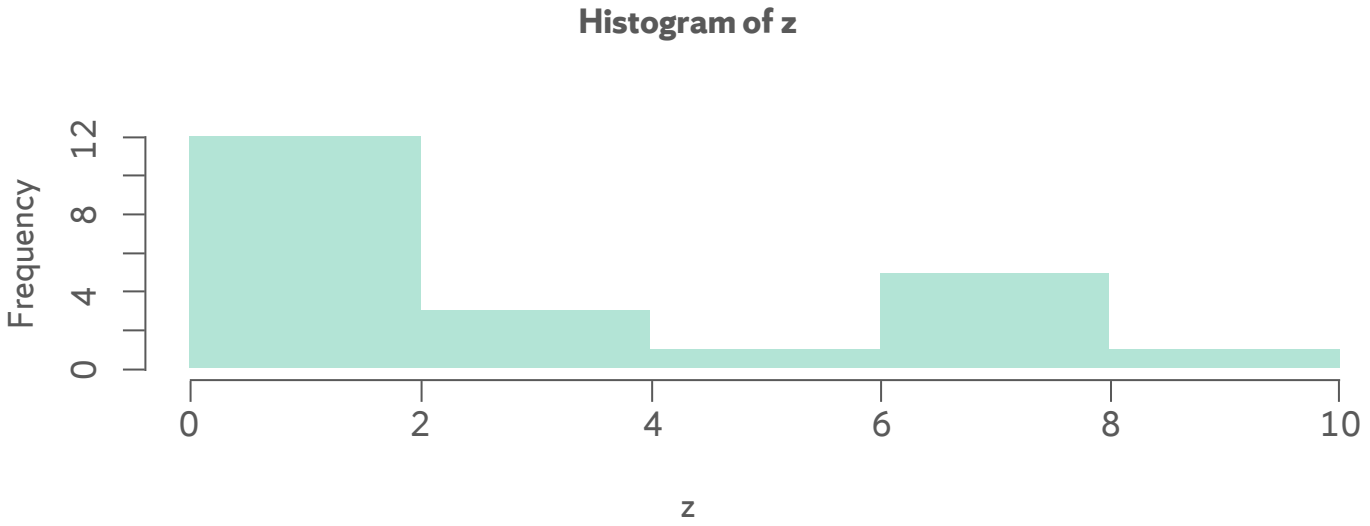


Distribution

Given a set of numbers, what is the central tendency? What is the shape?

1,5,0,3,8,8,1,2,1,8,7,2,1,2,10,3,2,8,1,0,3,0

0,0,0,1,1,1,1,1,2,2,2,2,3,3,3,5,7,8,8,8,8,10



Dimensions of Your Plots

An Informal Way to Talk About Plots

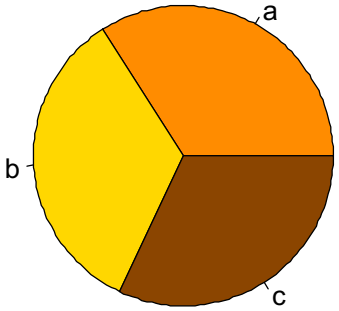
Single

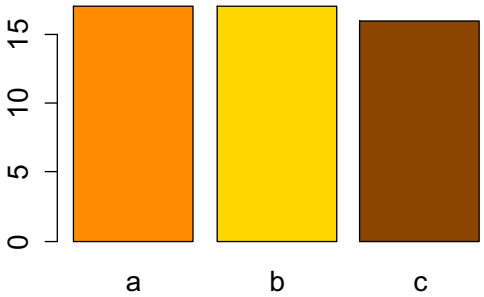
- Numeric or categorical
- Distribution within a single variable
- Frequency within a single variable

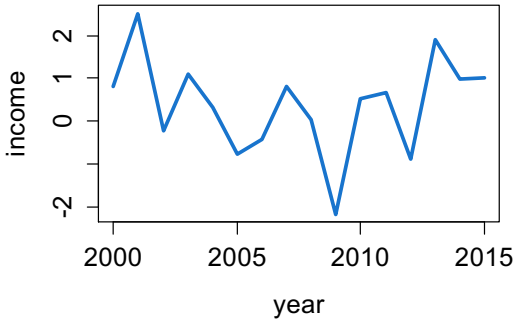
Multiple

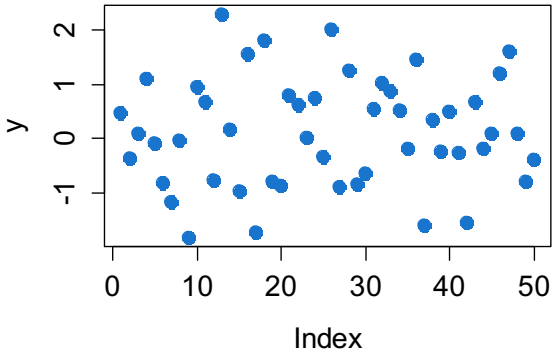
- Relating one variable to another
- Distribution of variable across categories
- Frequency of a categories across other categories

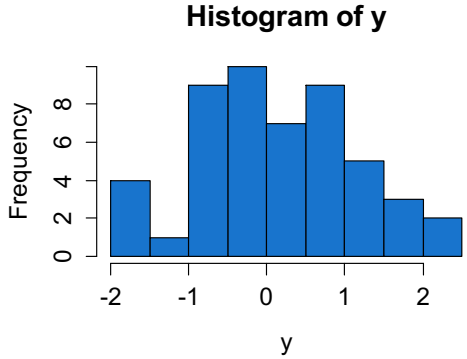
Single Dimension Plots

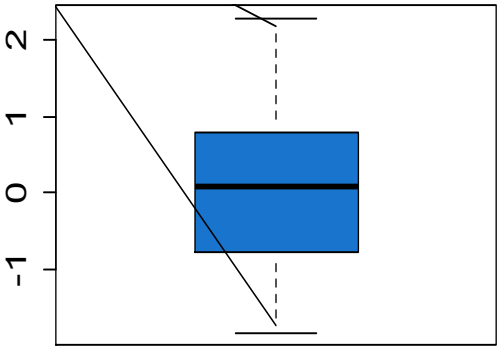




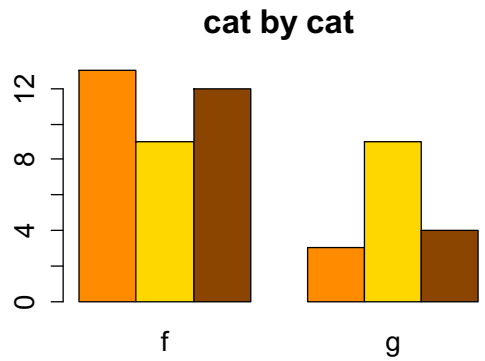


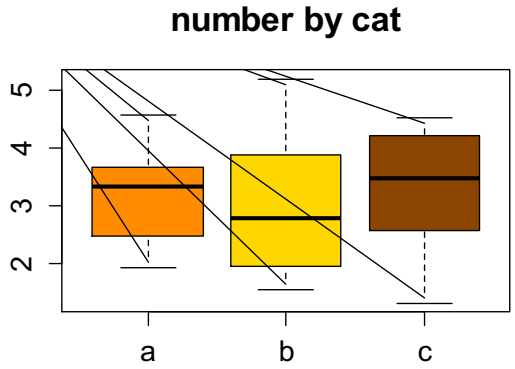


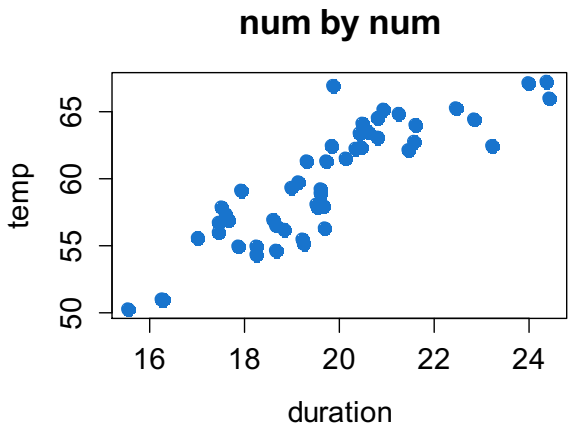


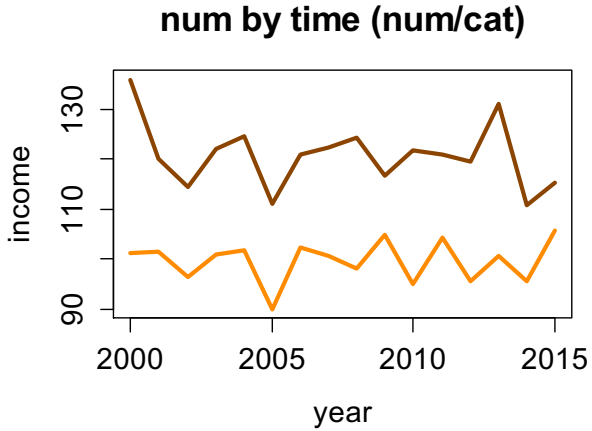


Multi-Dimension Plots

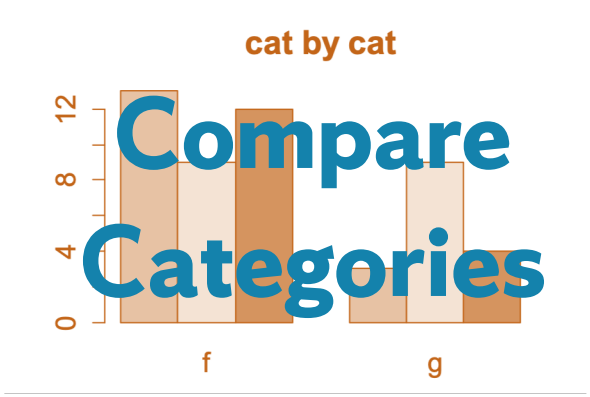


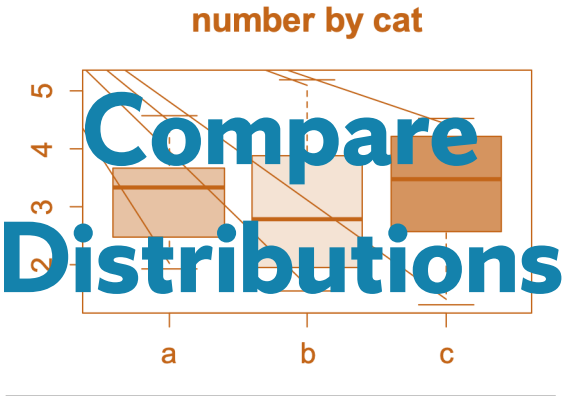


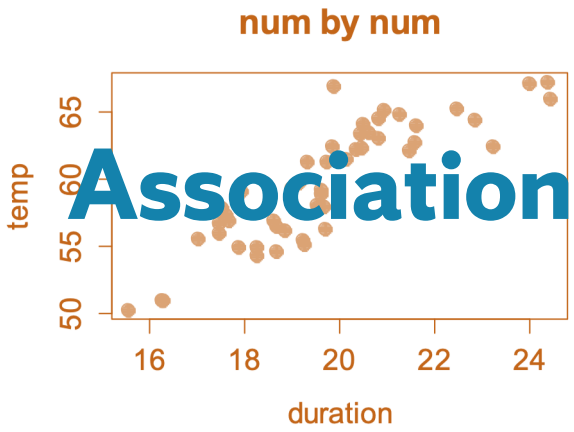


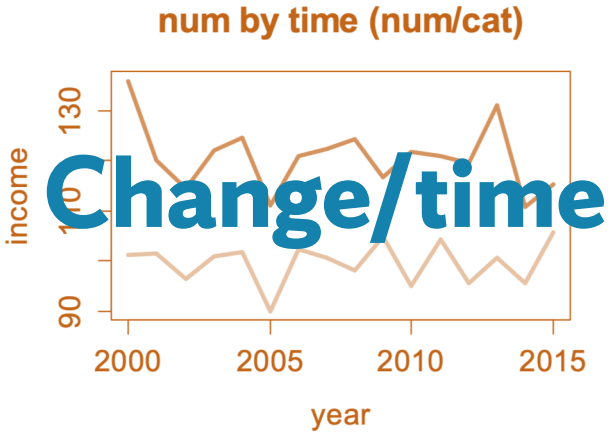


Multi-Dimension Plots

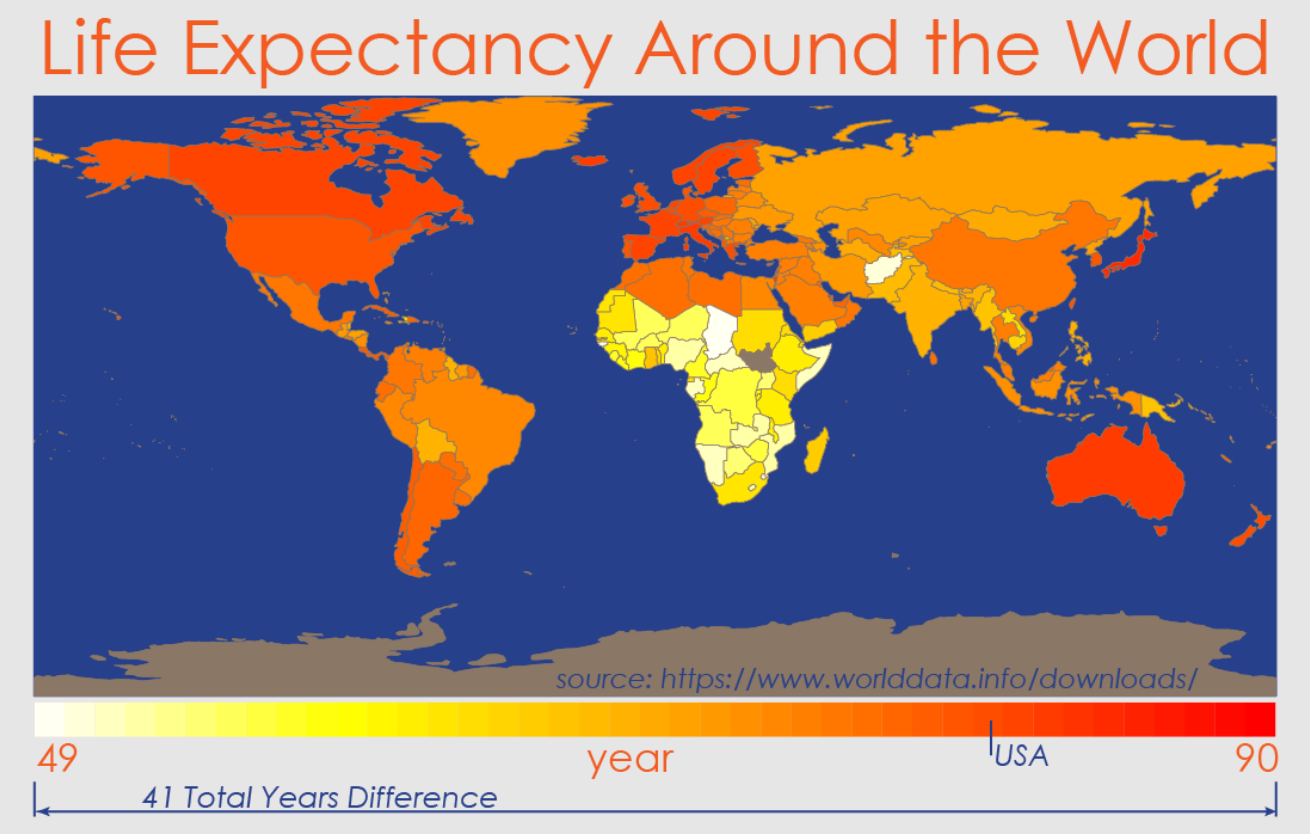




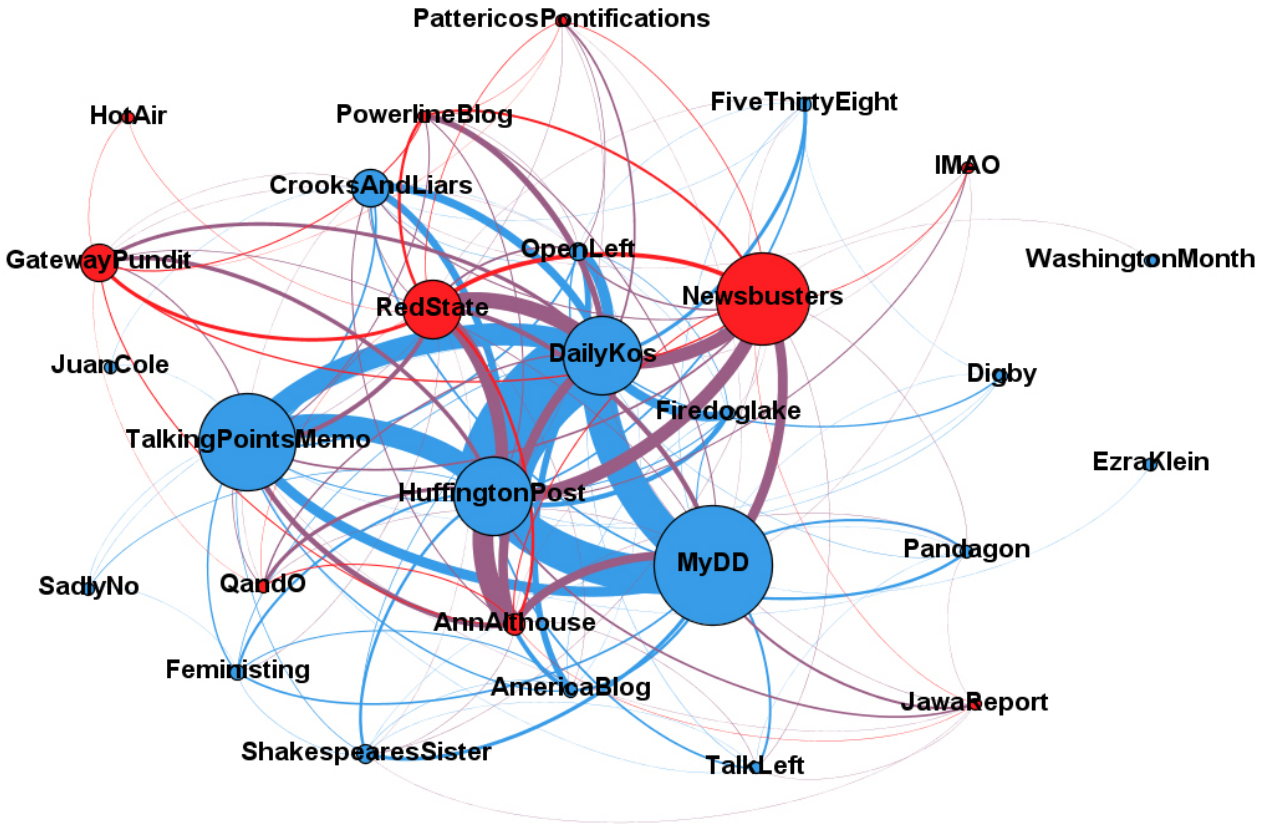




Multi-Dimension Plots



Multi-Dimension Plots



Comparisons

What kinds of comparisons?

Magnitude

- Nominal: no particular order, frequency
- Ranking: magnitude
- Parts-to-whole: proportions
- Deviation: differences between sets
- Time-series: change over time

Sorting

- Low to high

Position

- Maps

Strength of relationships

Data Story