

Data Manipulation

Introduction to Data Visualization

The Graduate Center at CUNY | Fall 2019

Michelle A. McSweeney

Five Minute Reflection

Data visualization, from the need for precision to the very notion of “data” is flawed. From a Humanities perspective, what can be gained by continuing to do visualizations? What is lost? Do you have any suggestions for a way forward to mediate some of the flaws?

Data

Categorical
DIMENSIONS

Numeric
MEASURES

Nominal

Ordinal

Discrete
GREEN PILL

Continuous
BLUE PILL

Examples:
Occupation,
Nationality,
Labels

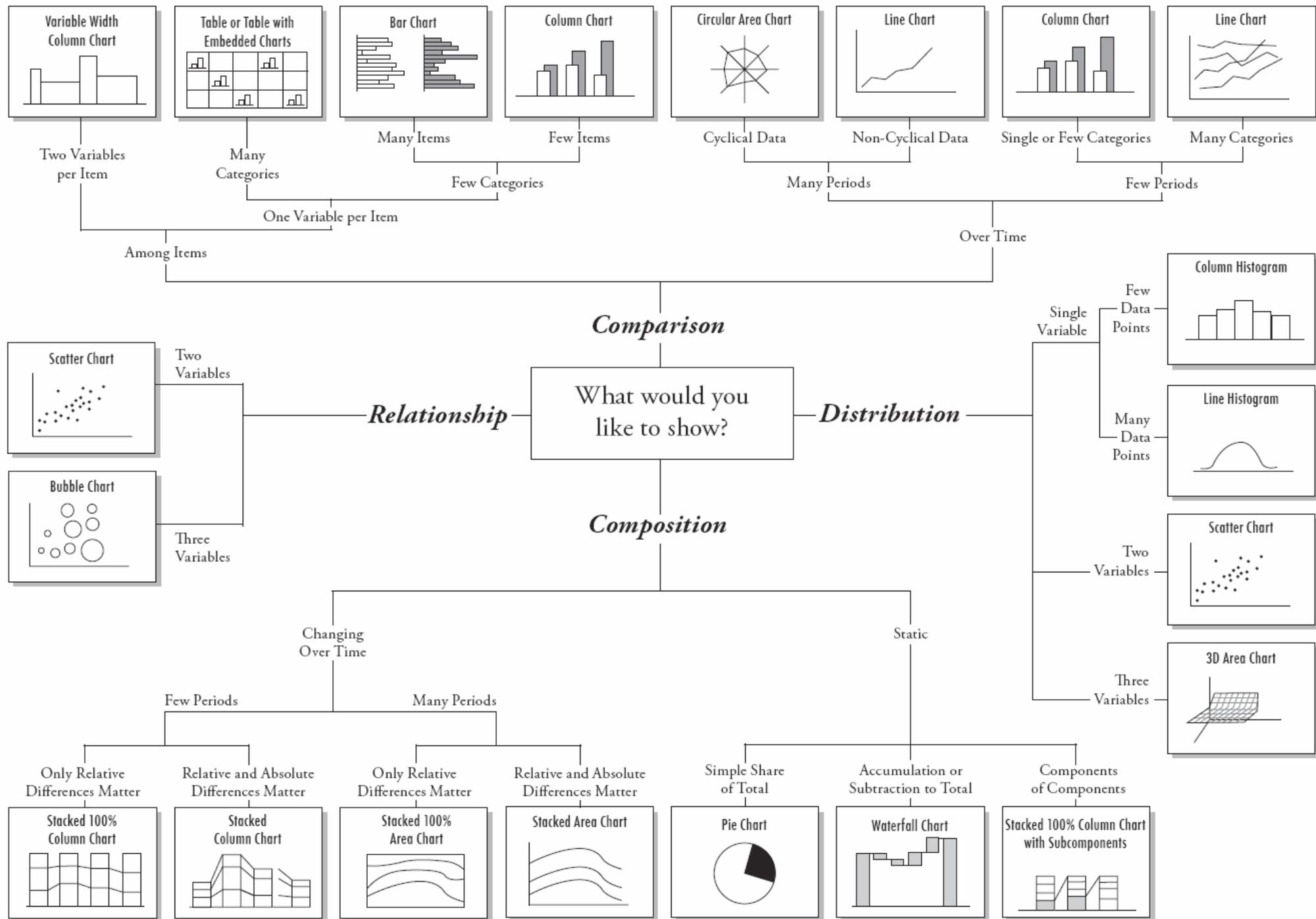
Examples:
Days of week,
Place in race,
Grades

Examples:
Children,
Books read,
Complaints

Examples:
Weight,
Distance,
(Time)

Data Type Dictates Visualization Type

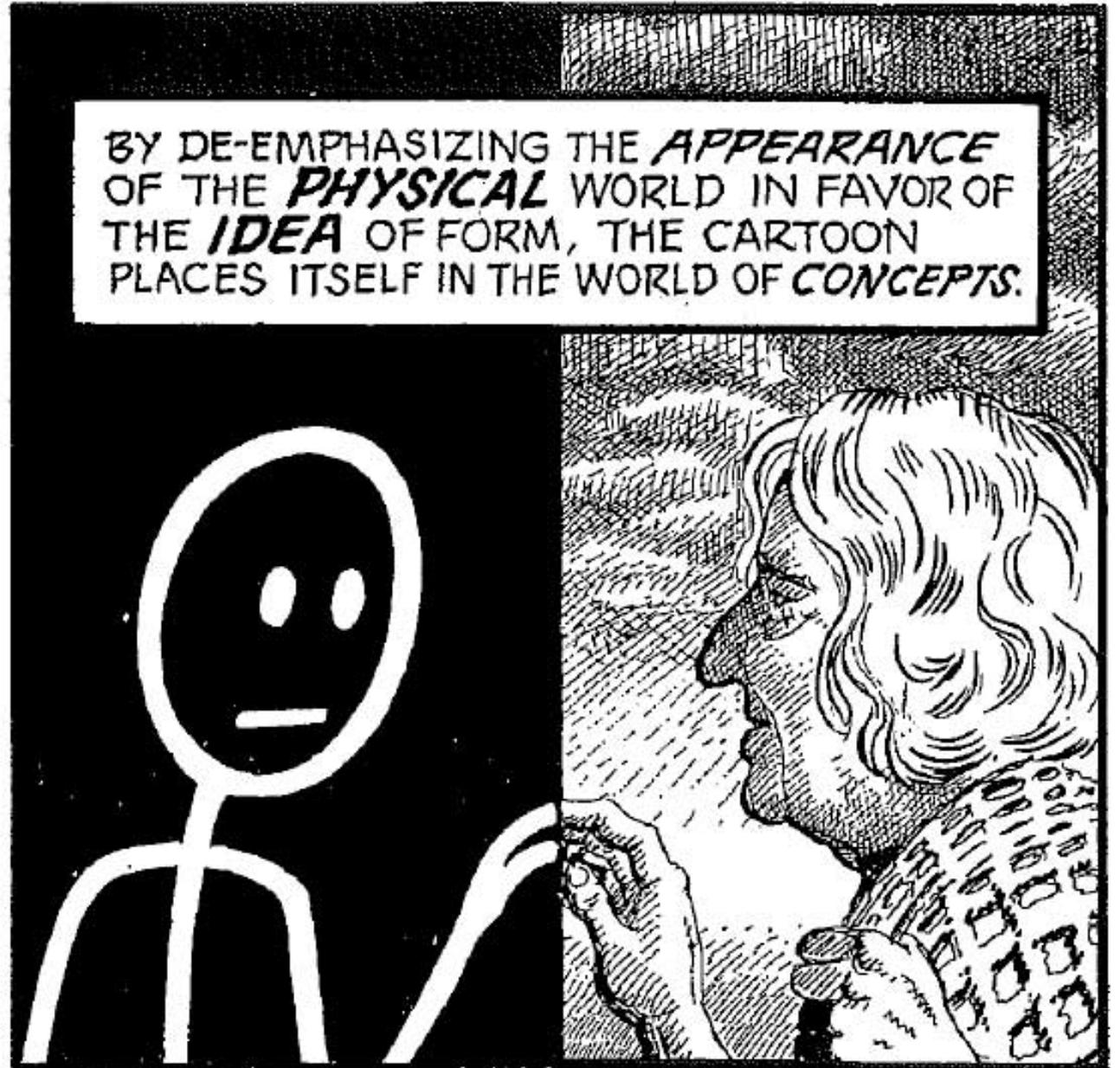
- Comparison
- Distribution
- Relationship
- Composition



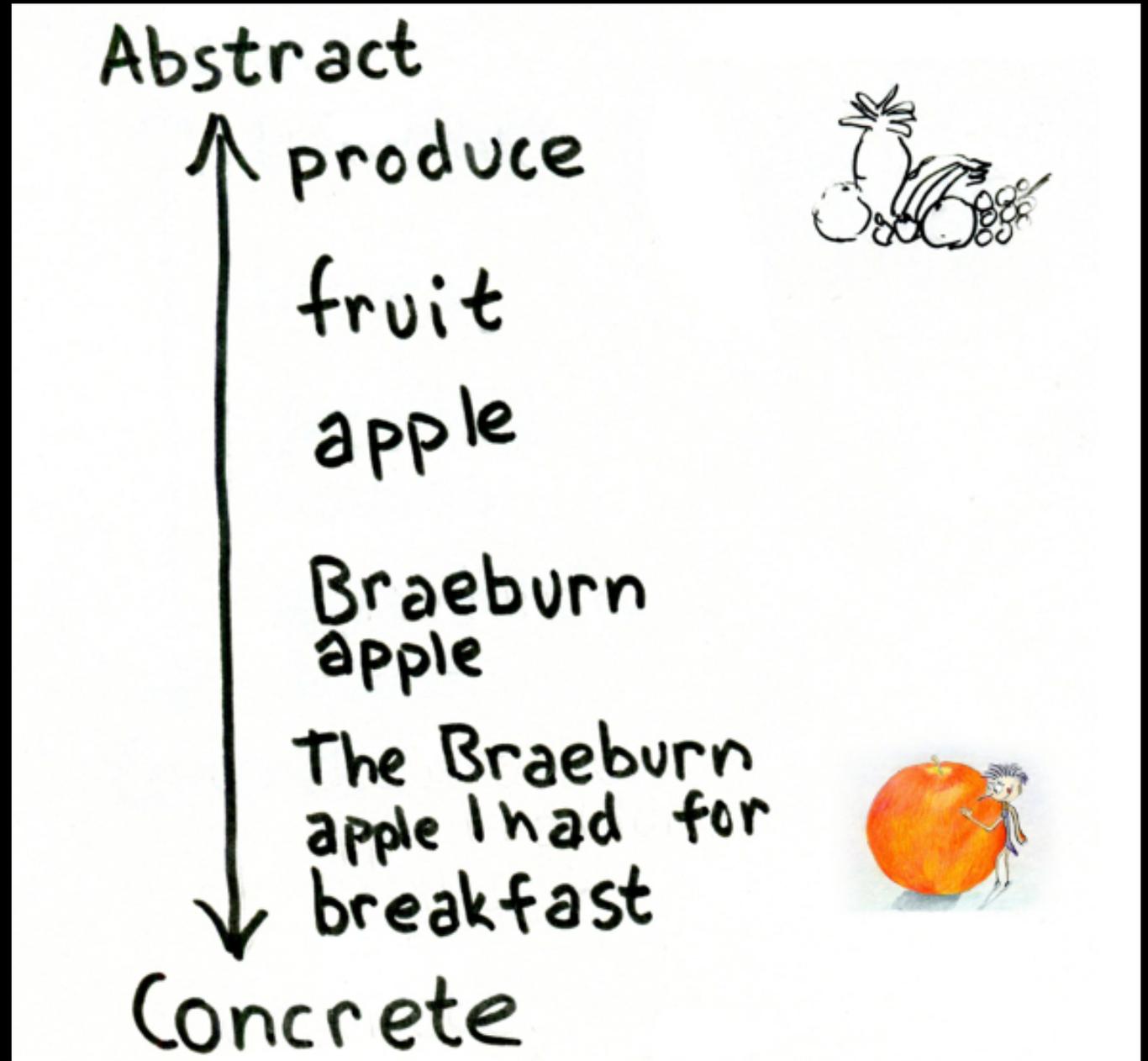
Tools to help you pick

- annkemery.com/essentials
- datavizcatalogue.com

Data Visualization: Concepts or Realities



Hiyakawa's Ladder of Abstraction



Mathematical Representations

Data itself is an abstraction from reality, but in order to make comparisons between different abstractions, they must be transformed into complimentary formats

Normalization

- Adjusting data to a “common” scale
 - Divide by total for each unit
 - i.e., Population reported as percentages is comparable, counts are not
- Shifted and scaled for comparison (log, exponent)
- Normal Distribution
- Quantiles

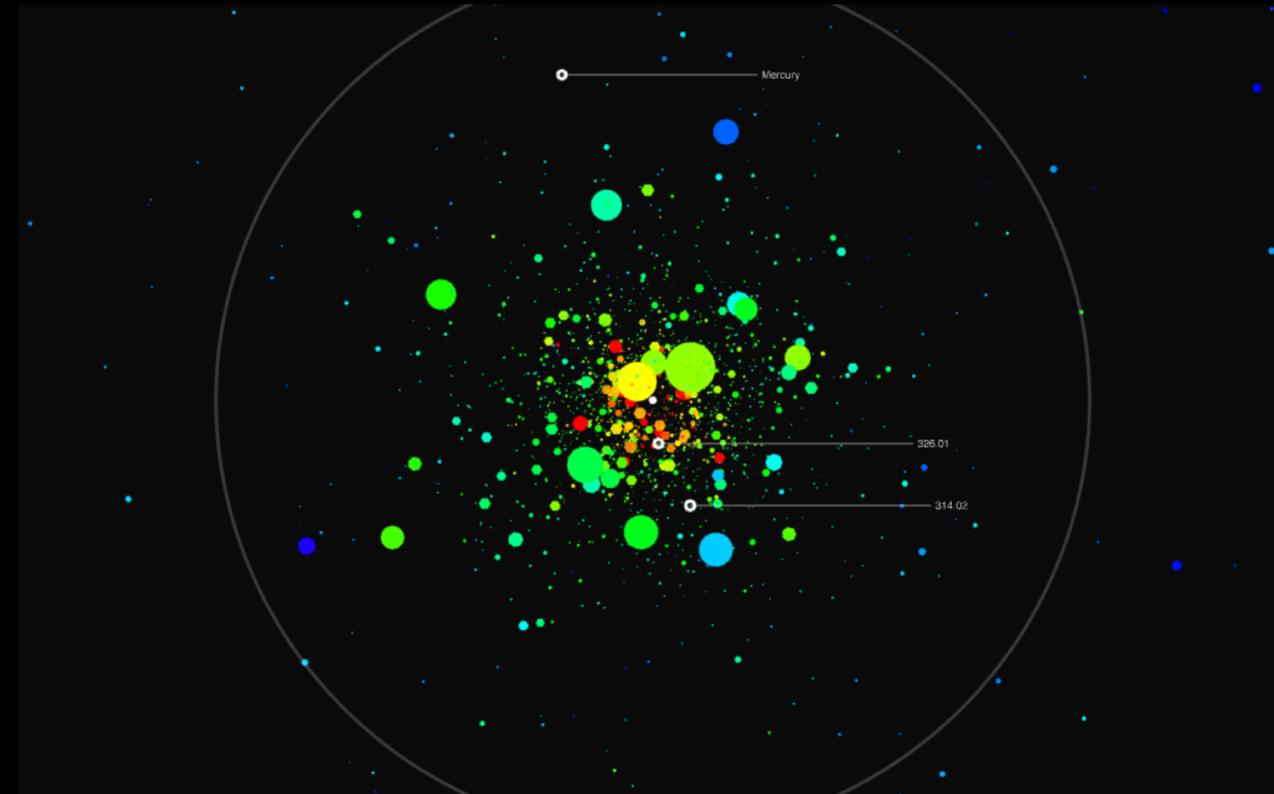
Averages

1, 3, 4, 8, 11, 11, 15

- **Mean:** simple average (7.57)
 - sum of all items \div total number of items
- **Median:** number in the middle (8)
- **Mode:** most common number (11)
- **Range:** 1 to 15 (14)

Context

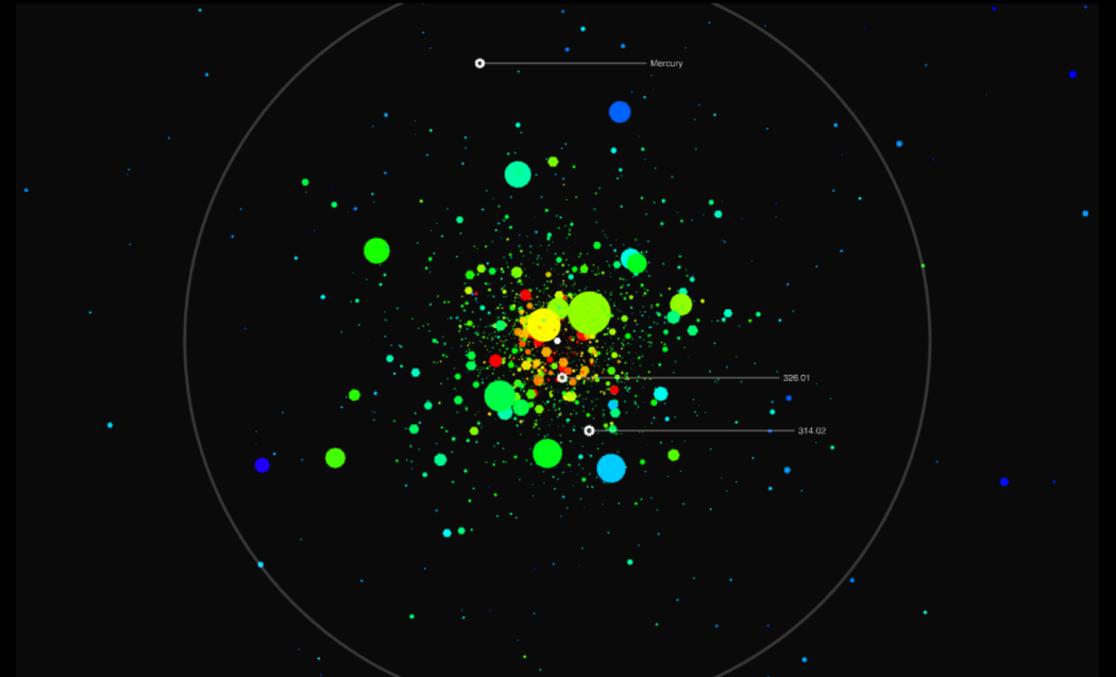
- Narrative around the data is as important as the data itself
- Allows reader to participate in your visualization
- Situates the reader so they know what to look for



Context

- Narrative around the data is as important as the data itself
- Allows reader to participate in your visualization
- Situates the reader so they know what to look for

1236 exoplanets identified by the NASA's Kepler mission.



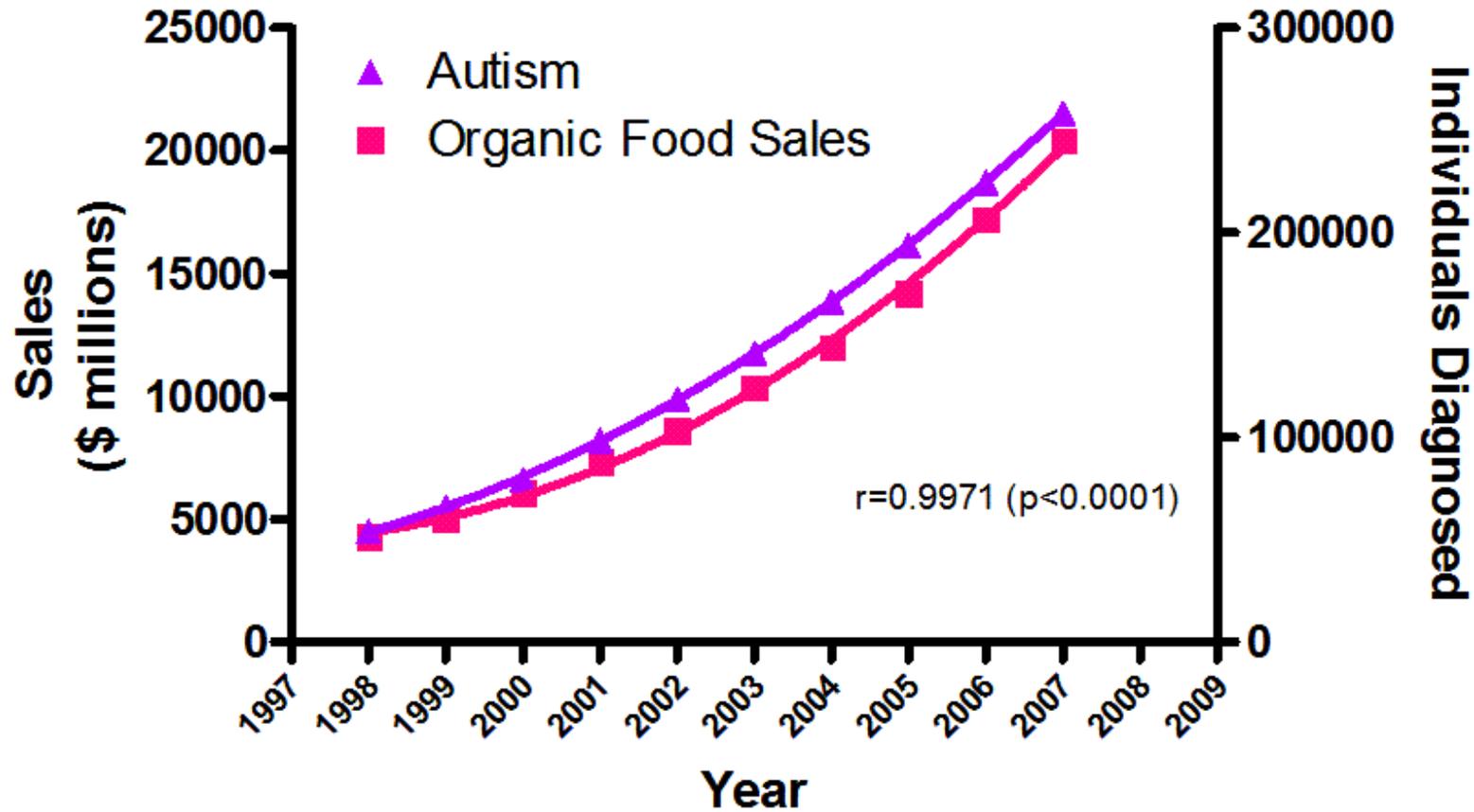
<http://blog.blprnt.com/blog/blprnt/data-in-an-alien-context-kepler-visualization-source-code>

Correlation is not Causation

Correlation/Causation

- Correlation:
 - two variables change at the same rate
- Causation:
 - One causes the other
 - Both are caused by an outside variable

The real cause of increasing autism prevalence?



Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act"

Correlation/Causation Traps

- Using Proxies (particularly with social phenomena)
- Looking at data without an experiment
- Allowing pre-existing biases to guide outcome

Ways to Avoid Causation from Correlation

- Situate visualizations in context
- Think about claims from both directions
- Support claims with articles and research

I USED TO THINK
CORRELATION IMPLIED
CAUSATION.



THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



SOUNDS LIKE THE
CLASS HELPED.

WELL, MAYBE.

