

Lecture : The W's of

Data Visualization and basic definitions

DATA ANALYSIS AND VISUALIZATION
Fall 2021

Dr. Muhammad Faisal Cheema
FASTNU

WHAT IS DATA VISUALIZATION

- Definition
 - **Communicates non-visual data visually**
 - Result should be readable and recognizable*
- Two Subfields
 - Scientific visualization
 - Information visualization
- Transforms raw **data** into **information**

- **CS related definition**
 - “computer-based visualization systems providing visual representations of datasets intended to help people carry out some task more effectively.” (T Munzner)

GOALS OF VISUALIZATION RESEARCH

Understand how visualizations convey information

What do people perceive/comprehend?

How do visualizations correspond with mental models?

Develop principles and techniques for creating effective visualizations and supporting analysis

Amplify perception and cognition

Strengthen tie between visualization and mental models

WHY DATA VISUALIZATION

- Scale of DATA
- Insights into DATA

DATA

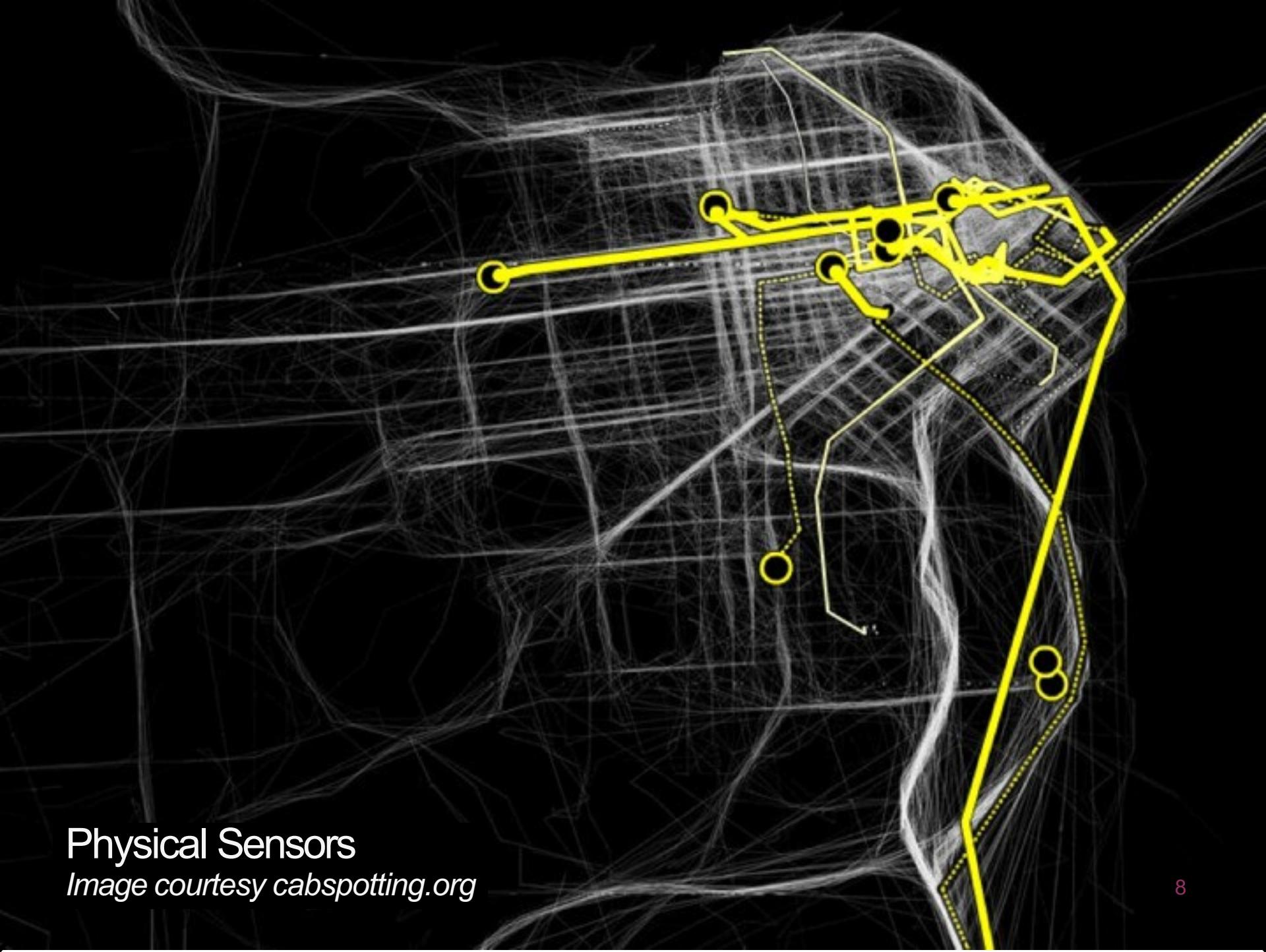
- Data is a set of values of qualitative or quantitative variables.
Pieces of data are individual pieces of information



**How much data (bytes)
did we produce in 2020?**

SCALE OF DATA

By 2020: 44 zetabytes
(that's 10^{21})

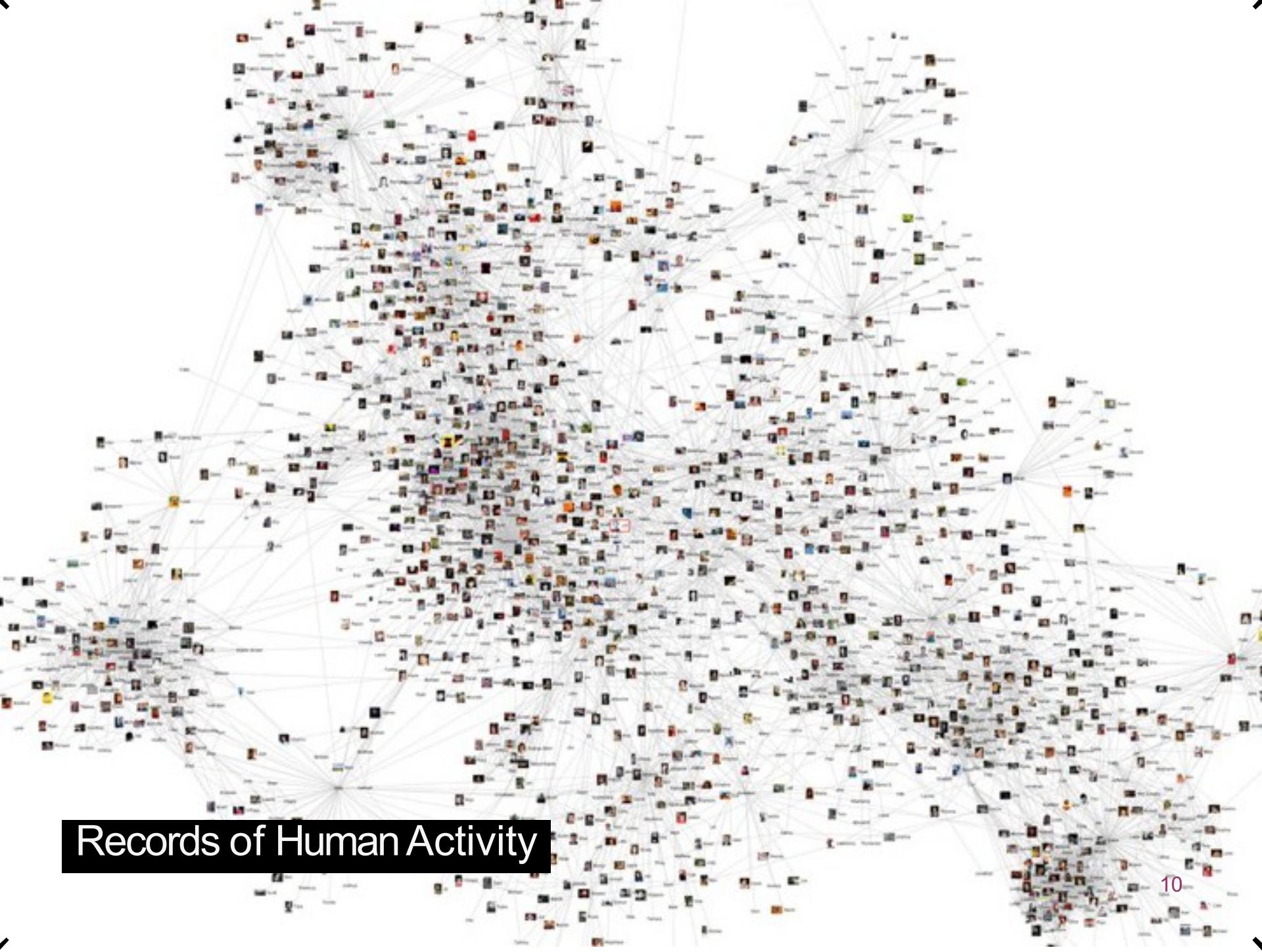


Physical Sensors

Image courtesy cabspotting.org



Health & Medicine



Records of Human Activity

Abortion

(Revision as of 22:56 4 Jan 2009)

"**Abortion**," in its most commonly used sense, refers to the deliberate early termination of a pregnancy, resulting in the death of the fetus. [1] Medically, the term also refers to early termination of a pregnancy by natural ("spontaneous abortion" or *miscarriage*, which occurs in 1 in 5 of all pregnancies, usually within the first 12 weeks) or to the cessation of normal growth of a body part or organ. What follows is a discussion of the issues related to deliberate or "induced" abortion.

Methods

Depending on the stage of pregnancy an abortion can be performed by a number of different methods. In the earliest terminations (before nine weeks) a **chemical abortion** is the usual method, though **methotrexate** is usually the only legal method, although research has uncovered similar effects from **methotrexate and misoprostol**. Combined with chemical abortion and exceeding up to about the fifteenth week **suction-aspiration** and **vacuum abortion** is the most common approach, replacing the more risky **dilation and curettage** (D & C). From the fifteenth week up until around the eighteenth week a **surgical dilation and extraction** (D & X) or a **hysterotomy abortion**, similar to a **c-section**, is used.

As the fetus size increases other techniques can be used to secure abortion in the third trimester. Premature expulsion of the fetus can be induced with **prostaglandin**, this can be coupled with injecting the amniotic fluid with saline or urea solution. Very late abortions can be brought about by the controversial **postabortal dilation and extraction** (P&X) or a **hysterotomy abortion**, similar to a **c-section**.

The controversy

The morality and legality of abortion is a highly important topic in **applied ethics** and is also discussed by **legal scholars** and **religious leaders**. Important facts about abortion are also reported by **sociologists** and **historians**.

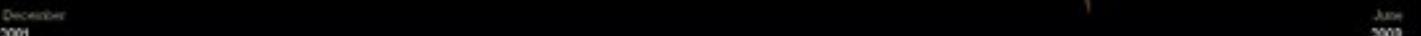
Abortion has been common in most societies, although it has often been opposed by some institutionalized religions and governments. **Controversy politics** in the **United States** and elsewhere saw abortion become commonly accepted by the mid-20th century. Additionally, abortion is accepted in **China**, **India** and other populous countries. The **Catholic Church** remains opposed to the procedure, however, and in other countries, notably the **United States** and the (predominantly Catholic) **Republic of Ireland**, the controversy is extremely active, to the extent that even the respective positions are subject to heated debate. While those on both sides of the issue are generally peaceful, if heated, in their defense of their position, the debate is sometimes characterized by violence. Though true of both sides, this is more marked on the side of those who oppose abortion, because of what they see as the gravity and urgency of their views.

The central question

The central question is the abortion debate: a clash of presumed or claimed rights. On one hand, is a fetus (sometimes called the "unborn") a person with **pro-life**/anti-abortion advocates; a human being with a right to life, and if so, at what point in pregnancy does the fetus become human? On the other hand, is a fetus part of a woman's body?

Wikipedia History Flow (IBM)

December
2001



June
2009

INSIGHTS INTO DATA

Sometimes the data is
unclear / even misleading

ANScombe QUATRET (73)

Set I		Set II		Set III		Set IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

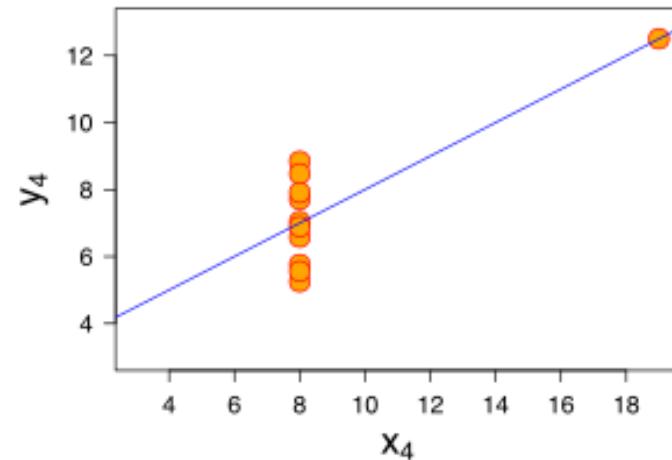
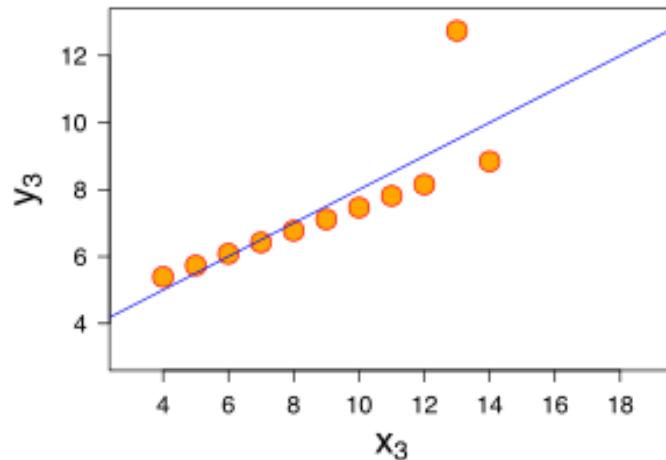
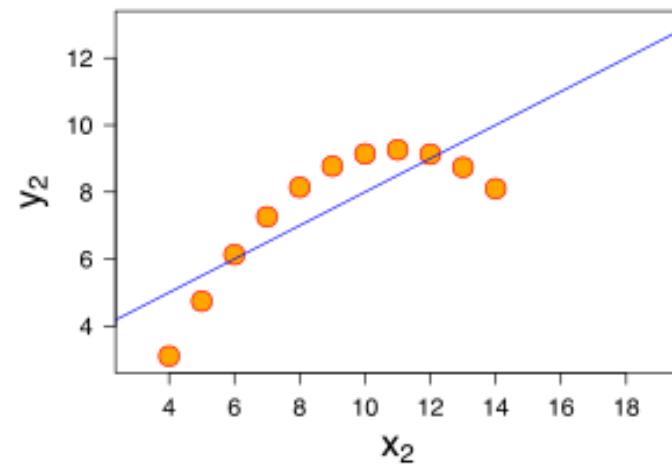
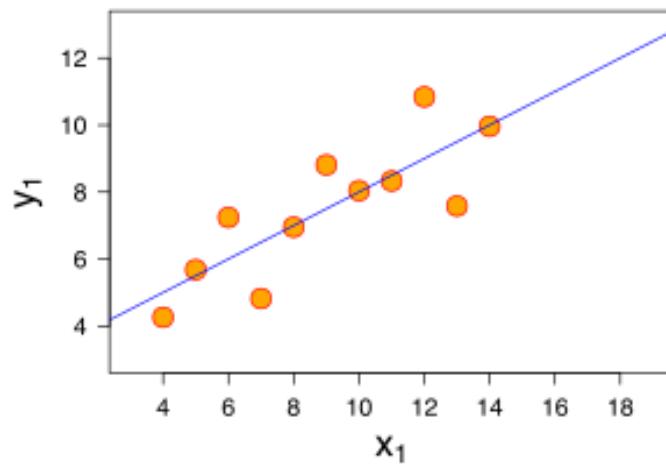
ANSCOMBE QUATRET (73)

Set I		Set II		Set III		Set IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Similar Statistical properties of all sets

- The average x value is 9 for each dataset
- The average y value is 7.50 for each dataset
- The variance for x is 11 and the variance for y is 4.12
- The correlation between x and y is 0.816 for each dataset
- A linear regression (line of best fit) for each dataset follows the equation $y = 0.5x + 3$

ANScombe QUATRET (73)



THE REASON OF VISUALIZATION

The ability to take data—to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it—that's going to be a hugely important skill in the next decades, . . .because now we really do have **essentially free and ubiquitous data**. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Hal Varian, Google's Chief Economist

The McKinsey Quarterly, Jan 2009

SOME MORE TERMINOLOGIES

- Data Visualization
- Scientific Visualization
- Information Visualization
- Statistical Graphics
- Visual Analytics
- Information Dashboards
- Infographics

TERMINOLOGY

- Differences between terms are often fuzzy
 - Information visualization versus infographics
- Differences between terms are sometimes highly contested
 - Information visualization versus statistical graphics
- Differences often come down to two aspects
 - Type of data being visualized
 - Why data is being visualized

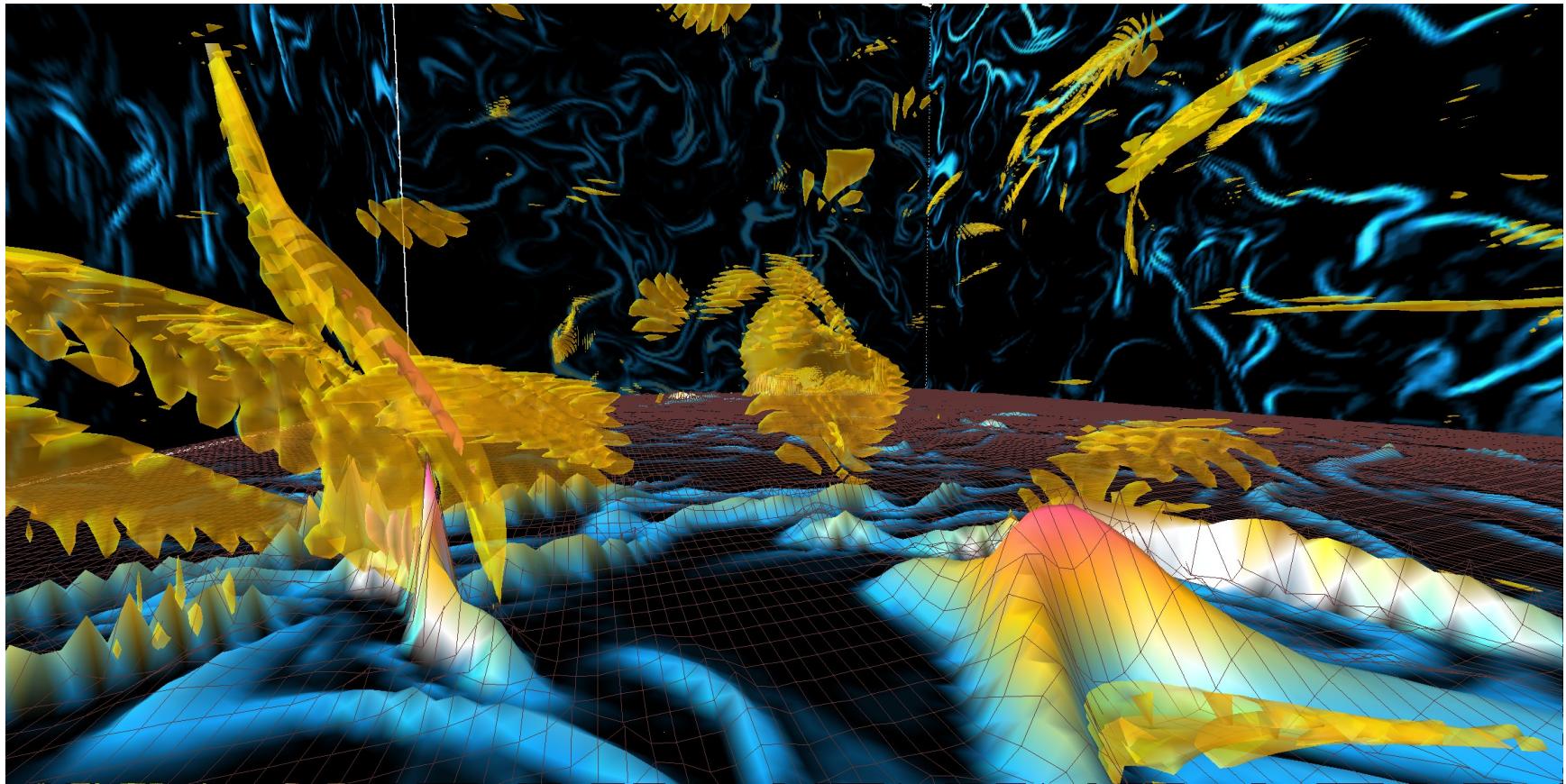
SCIENTIFIC VISUALIZATION

Type of Data

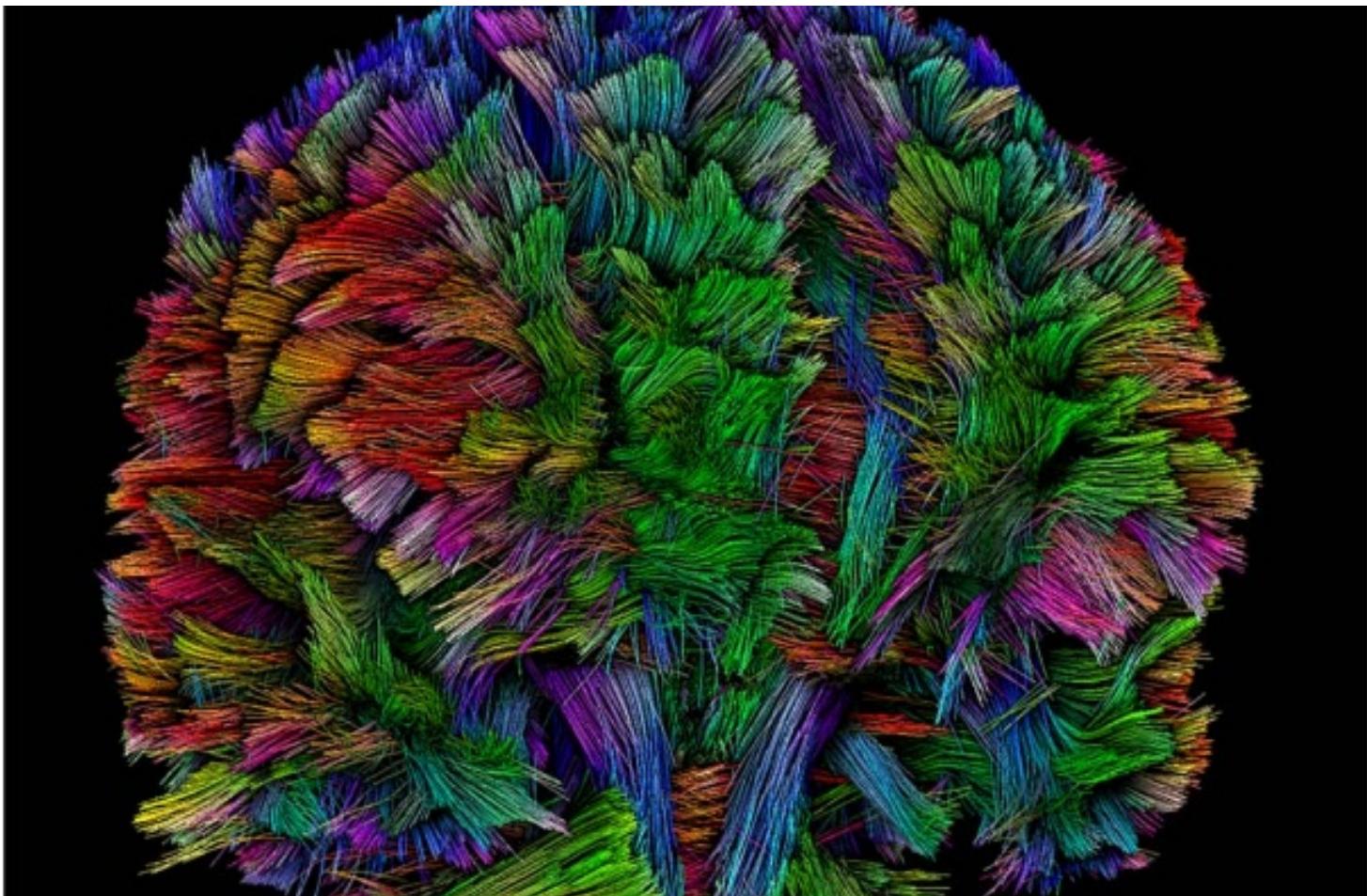
- Scientific data (objects exist in 1D, 2D, or 3D space)
- Often scalar or vector fields from computer simulations

Primary Purpose

- Aims to convey **scientific** data accurately
- Aims to reveal underlying **structure** in data
- Aims to encourage **exploration** of data (interactivity)



<http://vis.lbl.gov/Events/SC04/Incite3/index.html>



LineAO - Improved Three-Dimensional Line Rendering

<http://www.informatik.uni-leipzig.de/~ebaum/Publications/eichelbaum2012a/>

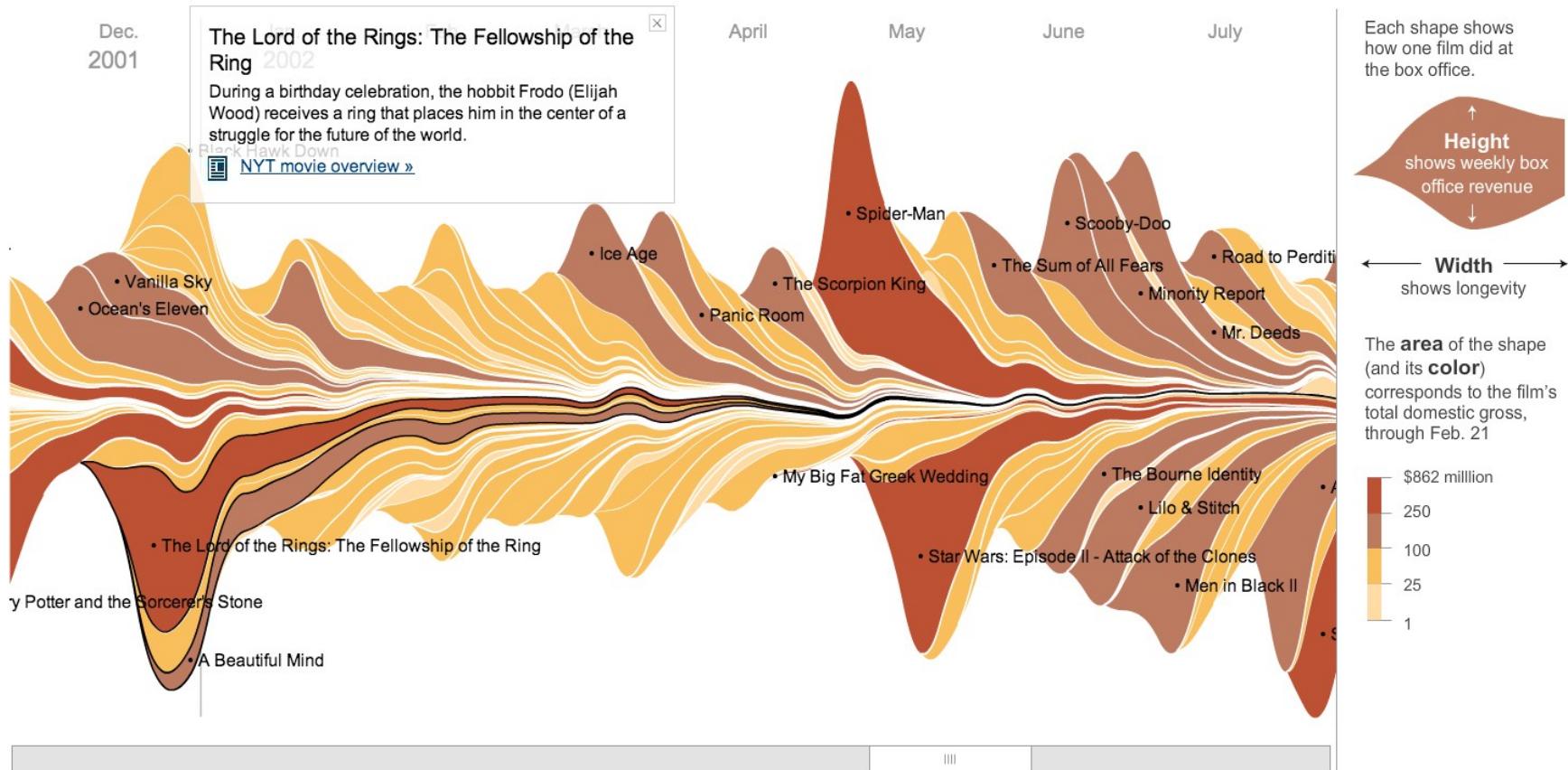
INFORMATION VISUALIZATION

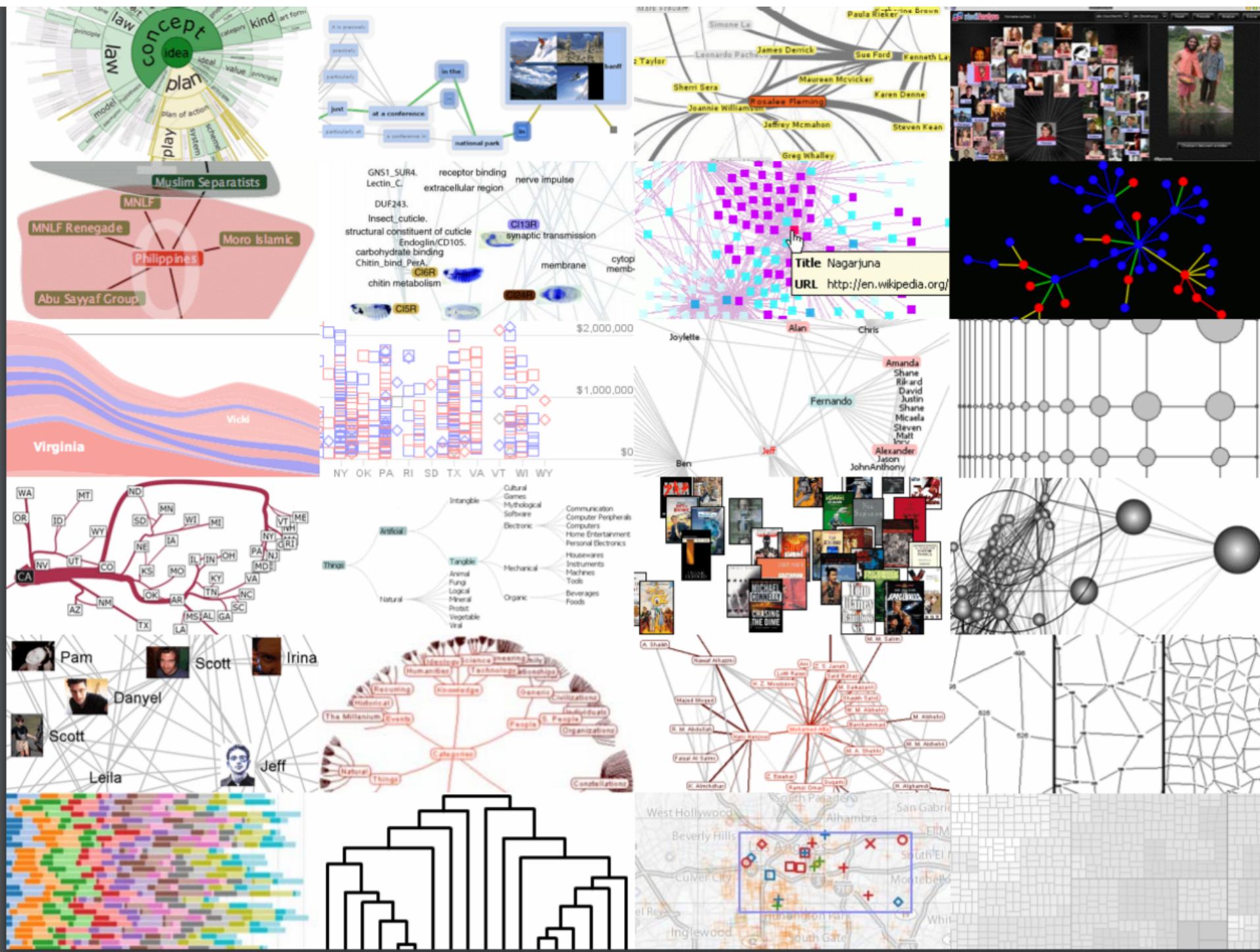
Type of Data

- Abstract data (has no inherent physical form)
- May be numerical, categorical, temporal, geospatial, or text data

Primary Purpose

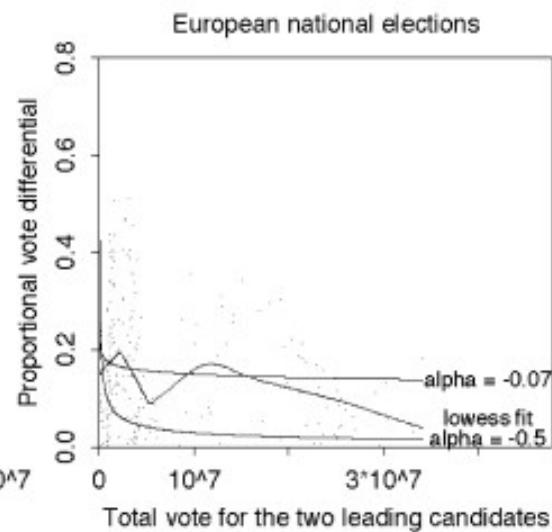
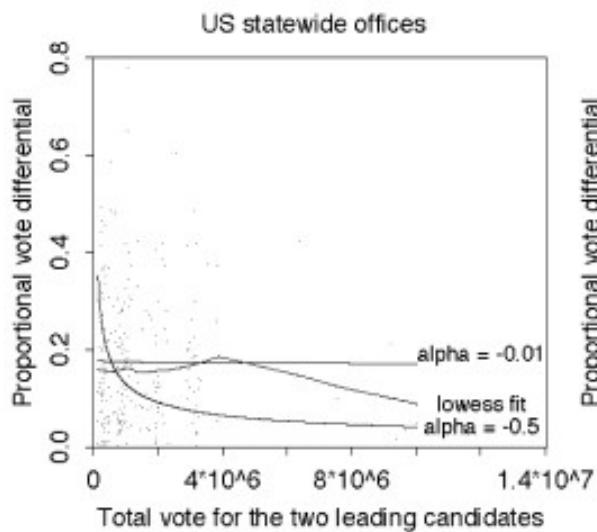
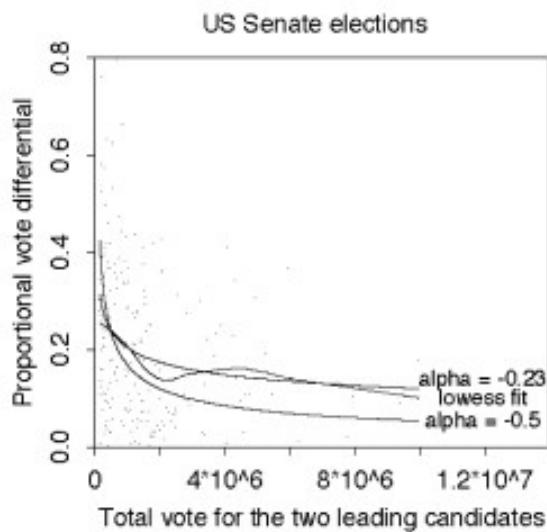
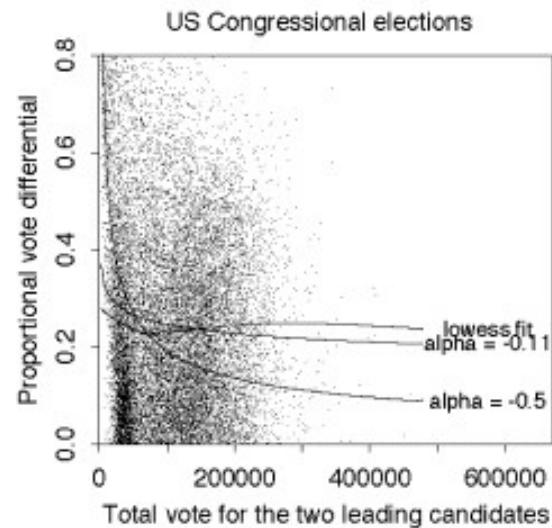
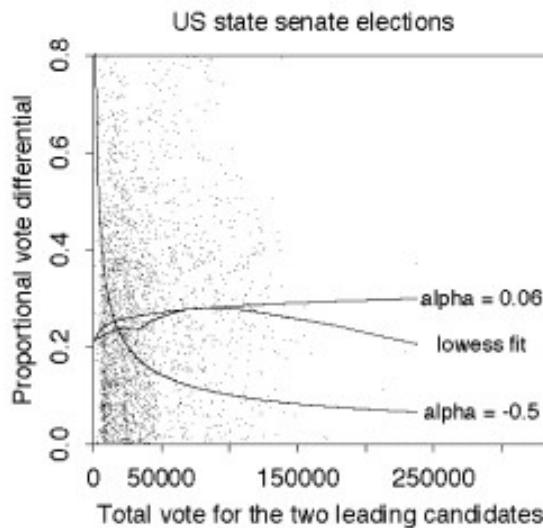
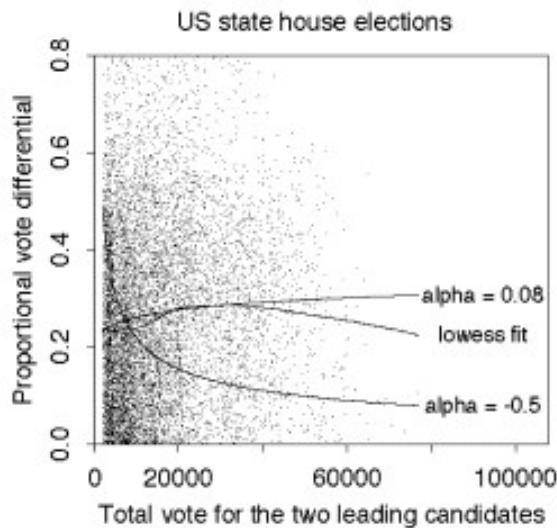
- Aims to convey **abstract** data accurately
- Aims to reveal underlying **structure** in data
- Aims to encourage **exploration** of data (interactivity)
- Aims to display data **aesthetically**

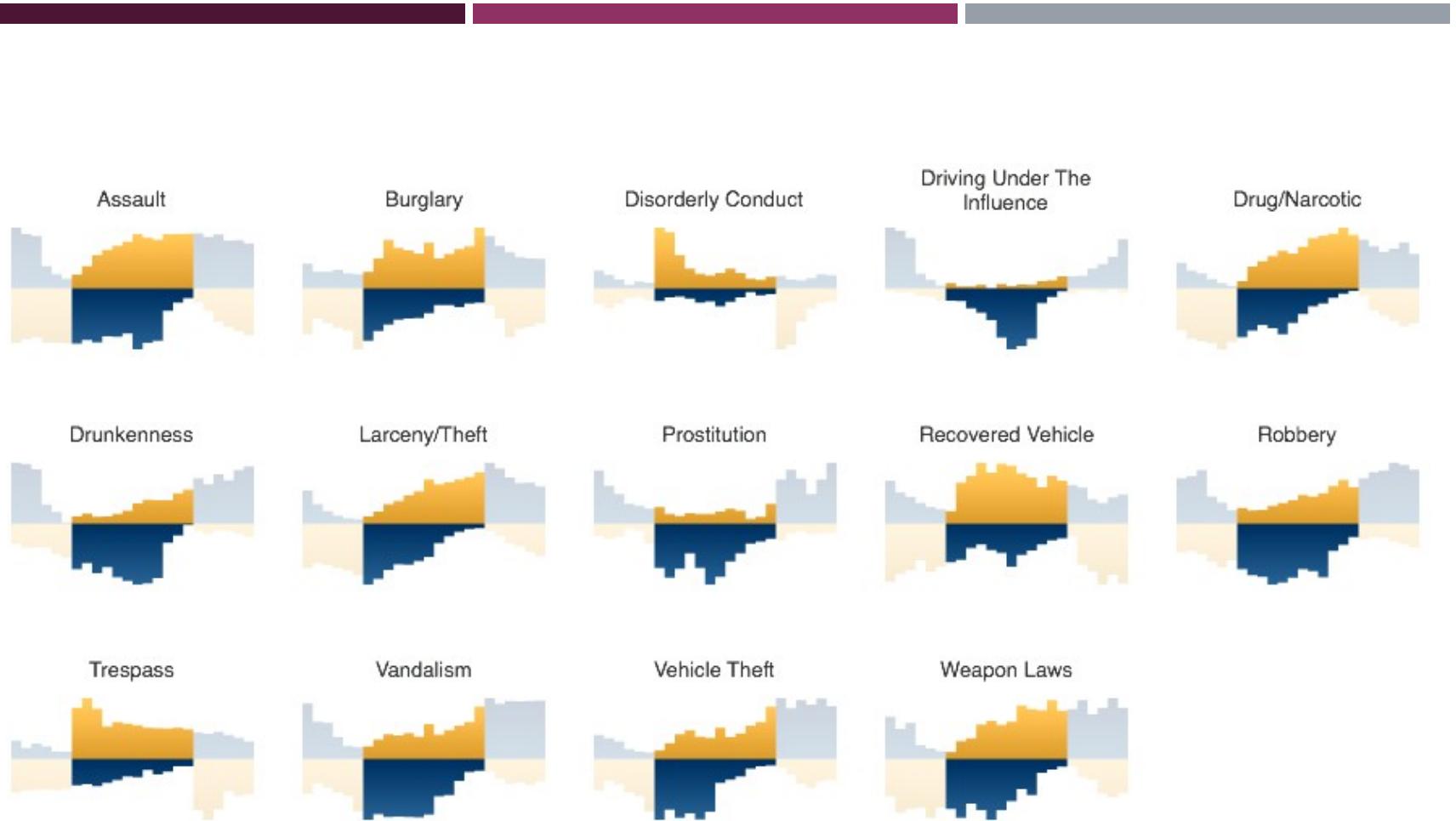




STATISTICAL GRAPHICS

- Type of Data
 - Abstract data
 - Mostly statistical, quantitative, or numerical data
- Primary Purpose
 - Aims to convey data accurately
 - Aims to convey underlying structure in data
- Not Emphasized
 - May not be aesthetically pleasing
 - May not encourage exploration or be interactive





VISUAL ANALYTICS

Type of Data

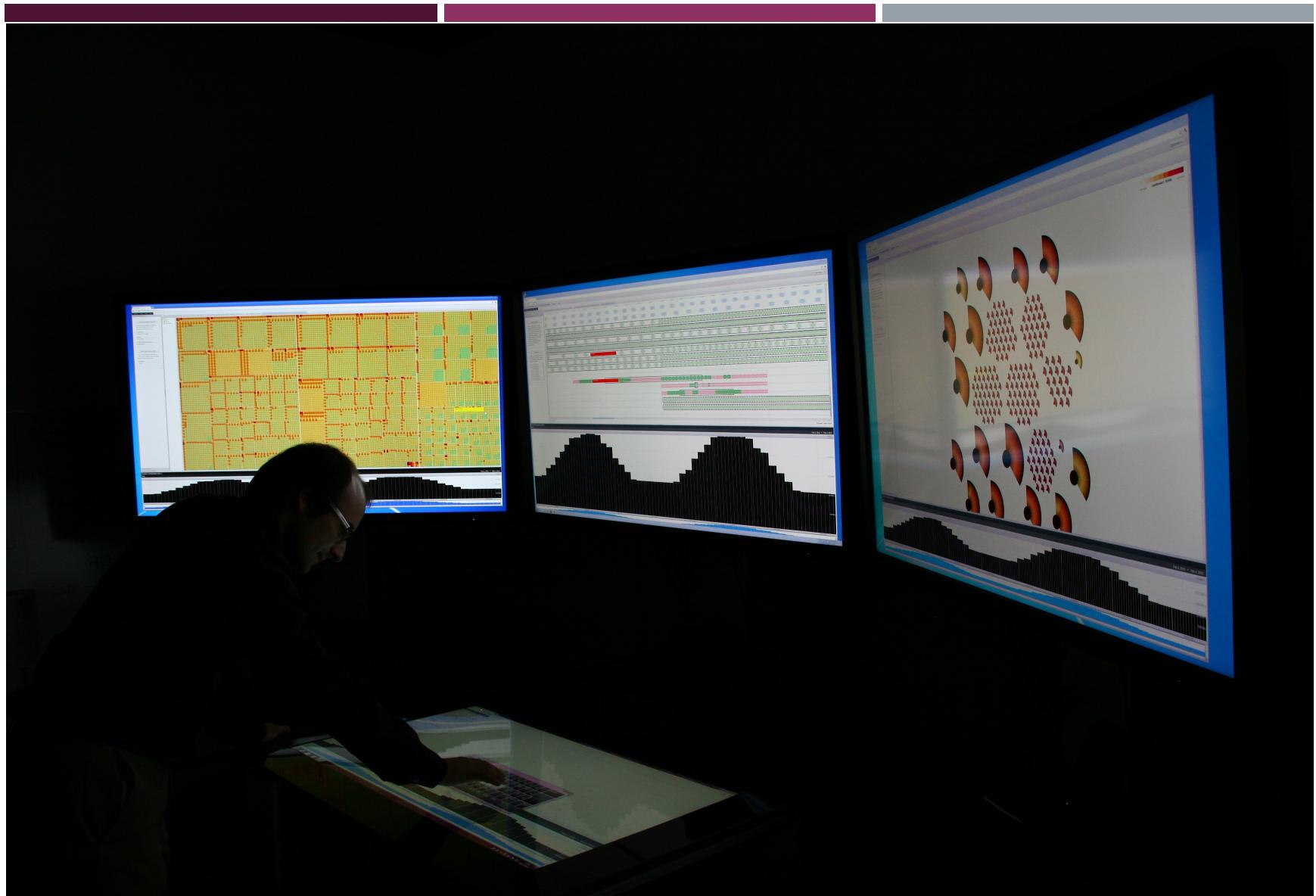
- Abstract data

Primary Purpose

- Aims to answer a specific question (goal-oriented)
- Aims to support analytical reasoning with interactive visual interfaces

Not Emphasized

- May not be aesthetically pleasing
- May not be constrained by a single display



<http://ff.cx/vast-challenge-2012/>

INFORMATION DASHBOARDS

Type of Data

- Abstract data
- Temporal data or time series

Primary Purpose

- Aims to convey large amount of information quickly
- Aims to convey outliers and trends at a glance

Not Emphasized

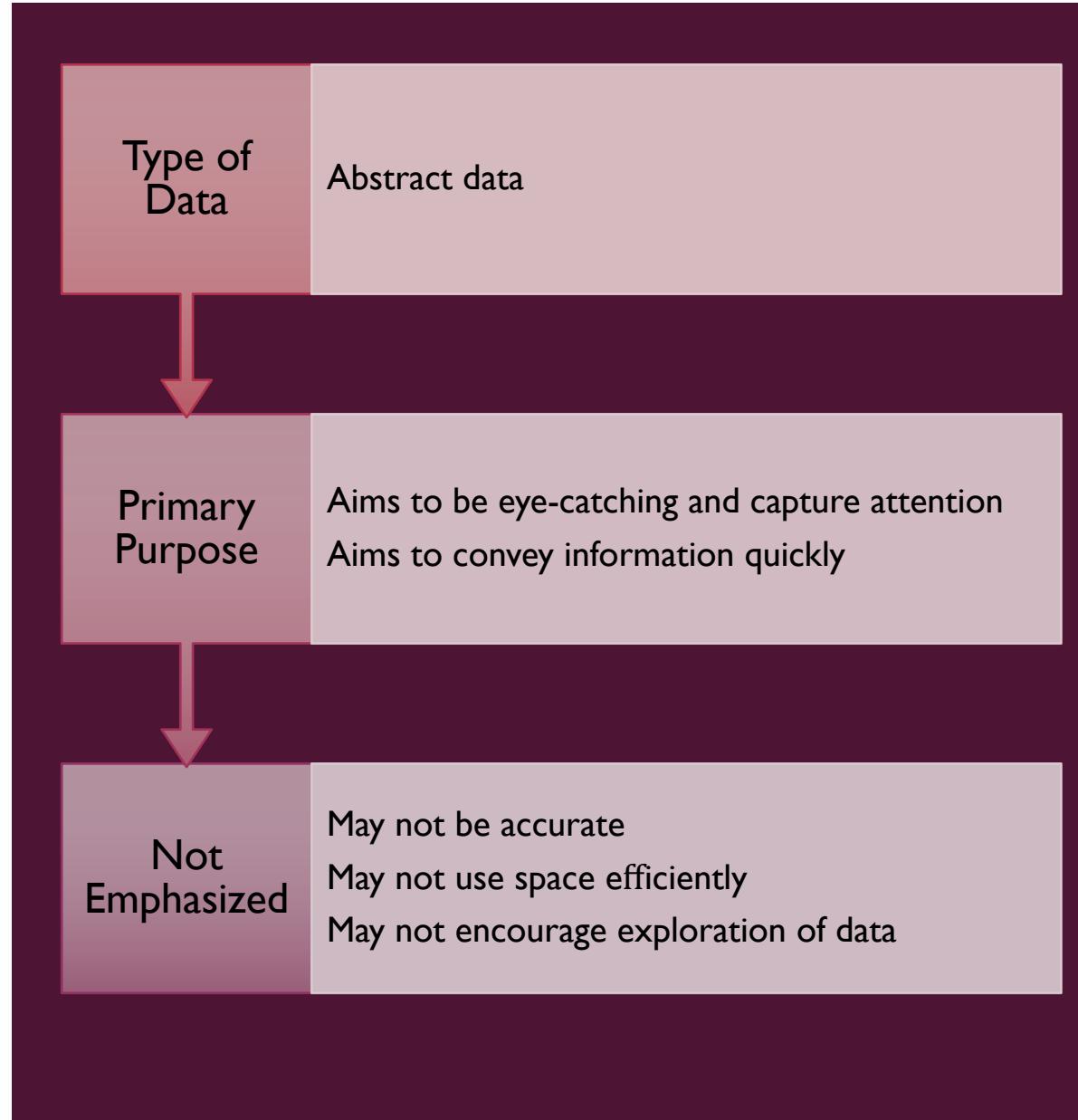
- May be extremely dense

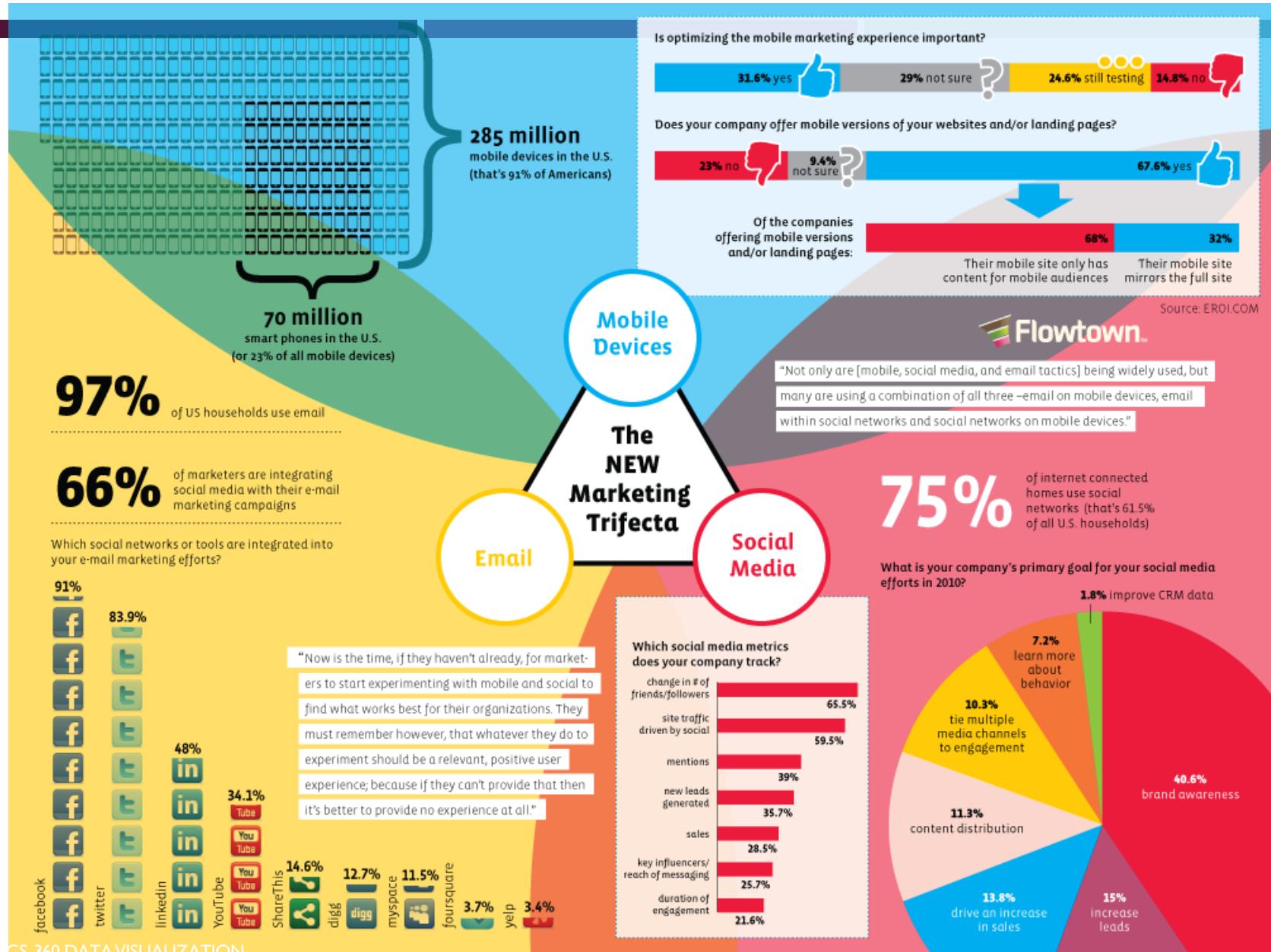
* No english language proficiency

[†] Special education

Note: Assessment and assignment scores are being expressed as the percentage of points that were earned out of the total points possible.

INFOGRAPHICS







INFORMATION VISUALIZATION VS INFOGRAPHICS

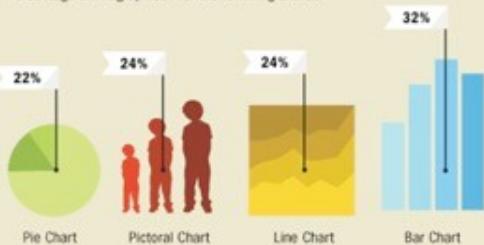
INFOGRAPHIC OF INFOGRAPHICS

Data visualization is a popular new way of sharing research. Here is a look at some of the visual devices, informational elements, and general trends found in the modern day infographic.

DESIGN

CHART STYLE

Percentage of infographics with the following charts:



FONT

Sans Serif Condensed Sans Serif
Serif



KEY INFO

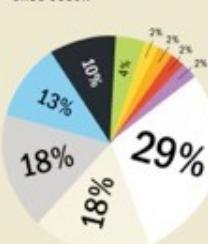
Percentage of infographics with key:



Average number of symbols per key: 5.1

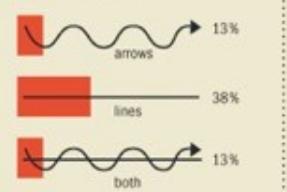


BASE COLOR

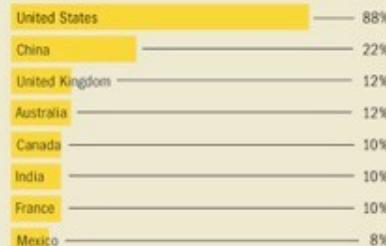


NAVIGATIONAL ICONOGRAPHY

Frequency of arrows & connecting lines in infographics:



COUNTRIES FEATURED

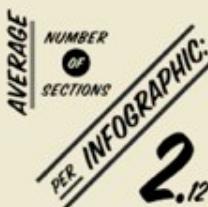


THEME

Relative popularity of different infographic themes:



SECTIONS



CREDITED SOURCES

Average number of sources per infographic: 2.29



TITLE

Average number of words per infographic title: 4.36

"RICHEST AND POOREST AMERICAN NEIGH



■ Questions ?