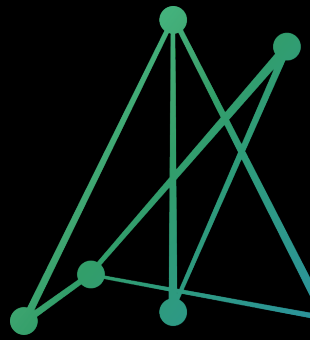


DATA SCIENCE
& ANALYTICS

University of Missouri

Data Visualization

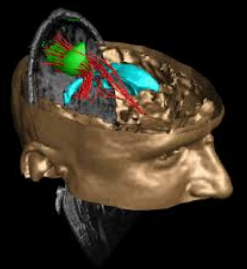
Introduction

DATA SCIENCE
& ANALYTICS
University of Missouri

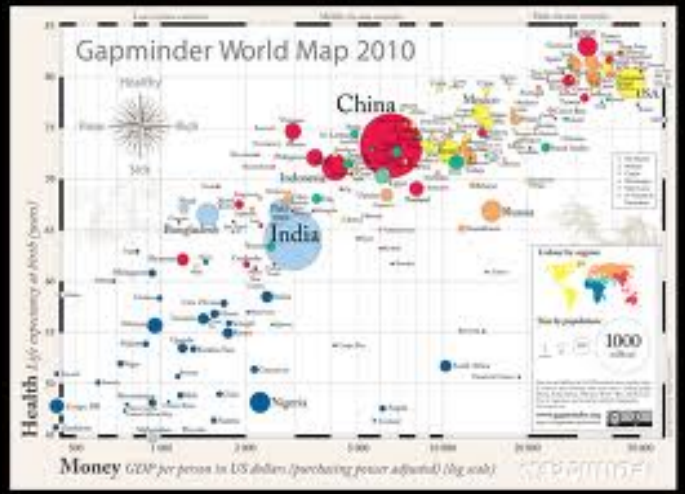
Course Objectives

This course covers the fundamentals of current visualization concepts and technologies. Students will develop an understanding of the principles of data visualization. You will learn about appropriate selection, modeling and evaluation of data visualizations, and using visualization tools and methods for exploratory data analysis.

What is Data Visualization?



VISUAL COMPLEXITY - MAPPING FACEBOOK FRIENDSHIP, PAUL BUTLER (2010)

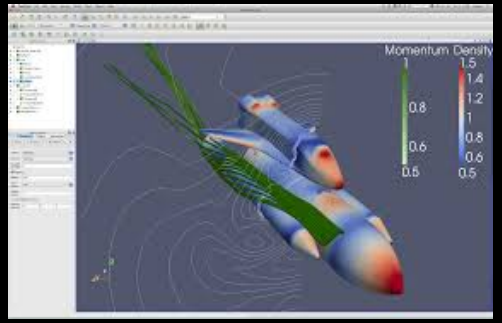


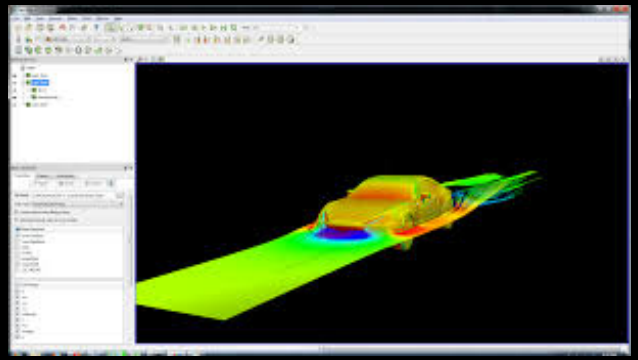
- Visual representation of data to communicate, explore, discover, analyze, gain insight about phenomena.
- Effective and efficient representation to communicate ideas, convey a message, convince people.

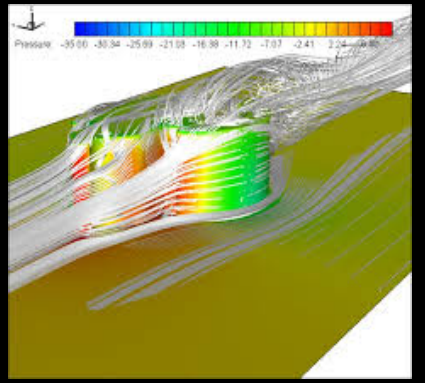
Types of Data Visualization

- Information Visualization (InfoVis)
  - No inherent structure: relational (facebook connections, geographical data, stock market, measurements, etc.)
- Scientific Visualization (SciVis)
  - Structural data: engineering simulations, medical imaging, seismic data, storm forecasting, fluid dynamics, etc.
  - Visual Analytics: Use visualization to explore and gain insight toward understanding of some phenomenon.

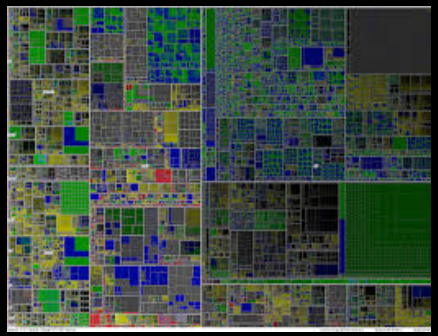
Scientific Visualization

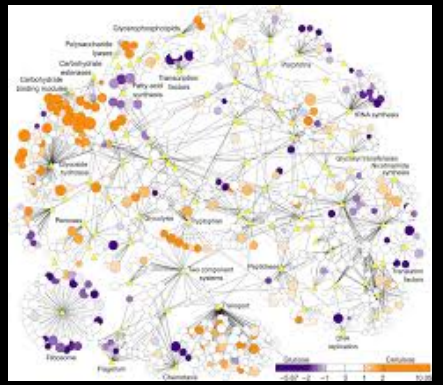


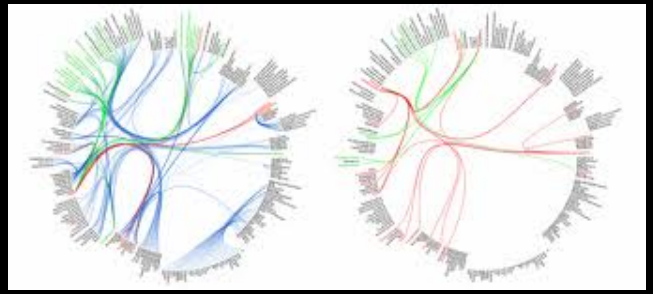


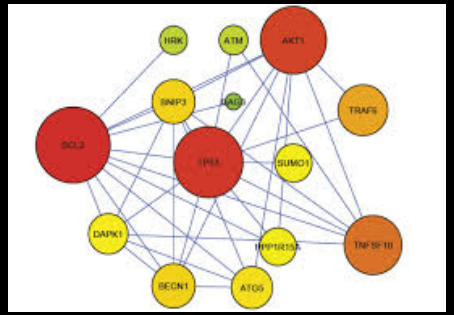


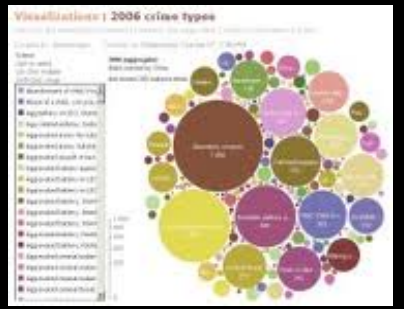
Information Visualization











Why Visualize Data?

- How many V’s do you see ?

DFVHDYJDWYSEPSBCWNQWZNCXETRBX
QECMRTHJPCVORCGMXNXZEZFKYJHVCT
XECRFVPTOJNBKVCMXNRXWVMYBMACQ
RTRPFEOFGVMCNSZXNCEHOCYJHOBVCM

Why Visualize Data?

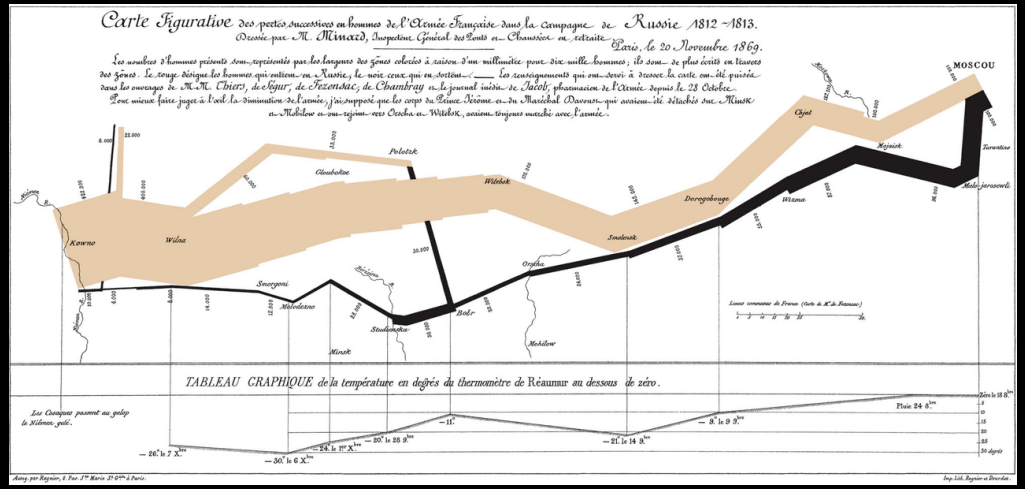
- How many V's do you see ?

DFVHDYJDWYSEPSBCWNQWZNCXETRBX
QECMRTHJPCVORCGMXNXZEZFKYJHVCT
XECRFVPTOJNBKVCMXNRXWVMYBMACQ
RTRPFEOFGVMCNSZXNCEHOCYJHOBVCM

Why Visualize Data?

- Analysis: Exploit human visual perception to explore relationships in data.
- Communication: Filter, mine, represent data in a manner suitable to convey an idea effectively.

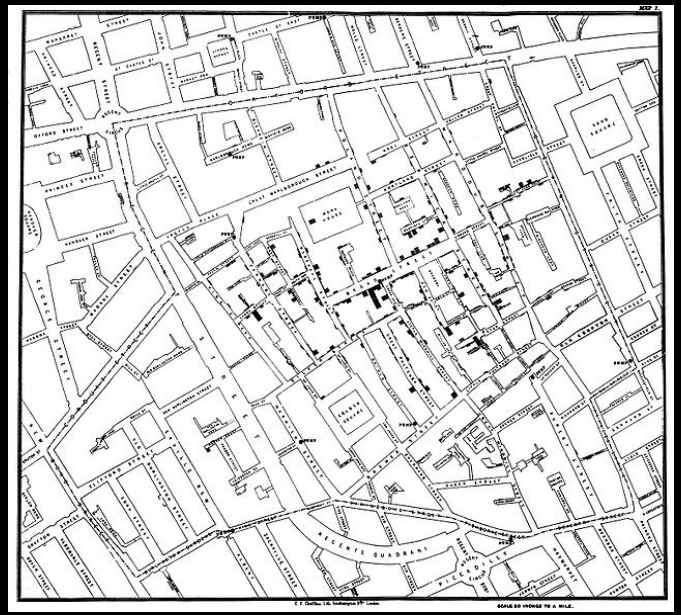
Impact of Visualization



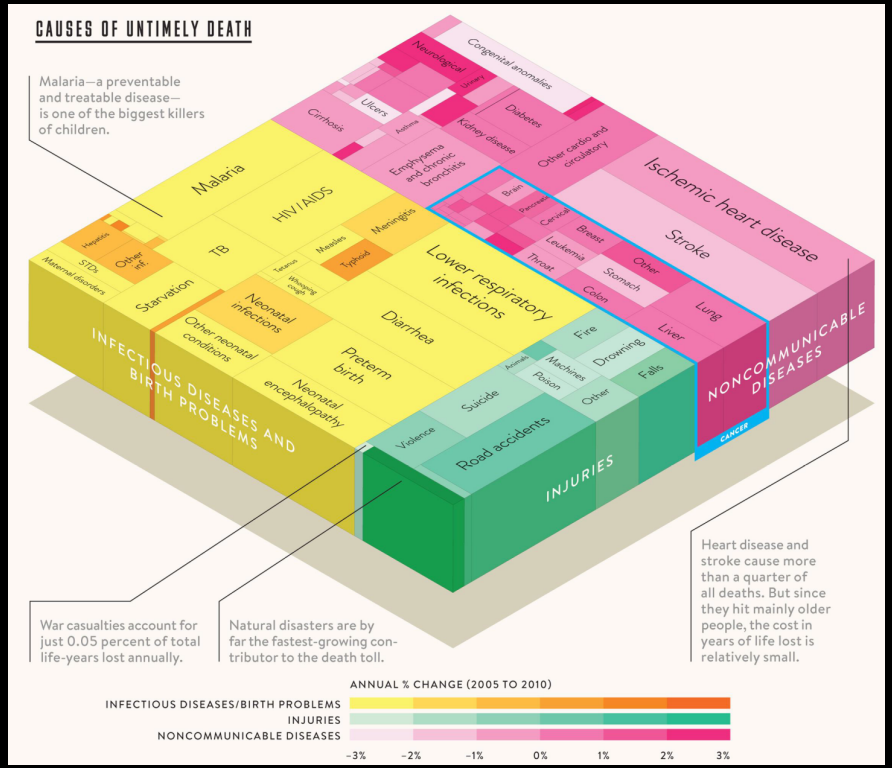
Minard's map of Napoleon's Russian Campaign of 1812. Represents six types of data: number of troops, distance, temperature, lat and long, direction of travel, location w.r.t. dates.

Impact of Visualization

John Snow's cholera map in the London Epidemic of 1854.



Impact of Visualization

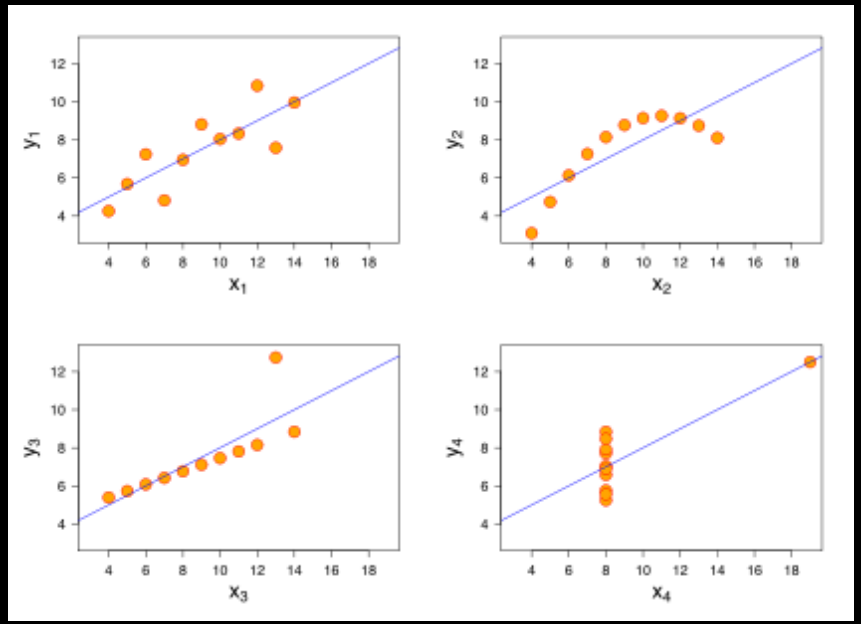


Why Visualize Data?

- Anscombe's Quartet

I		II		III		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

Why Visualize Data?



- Anscombe's Quartet

All four sets have identical simple summary statistics.

Property	Value
Mean of *x* in each case	9 (exact)
Sample variance of *x* in each case	11 (exact)
Mean of *y* in each case	7.50 (to 2 decimal places)
Sample variance of *y* in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between *x* and *y* in each case	0.816 (to 3 decimal places)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)

END OF SLIDES