Data Visualization

A picture is worth a thousand words!

A Second Exercise

What is Data Visualization? Please define it in your own words!

Let's Look at Some Data

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Let's Look at Some Data

With Statistical Analysis

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89
9	7.5	9	7.5	9	7.5	9	7.5
11	4.1	11	4.1	11	4.1	11	4.1
0.816		0.816		0.816		0.816	
$y = 0.5x + 3$		$y = 0.5x + 3$		$y = 0.5x + 3$		$y = 0.5x + 3$	

Mean of x

Variance of x

Correlation between x & y

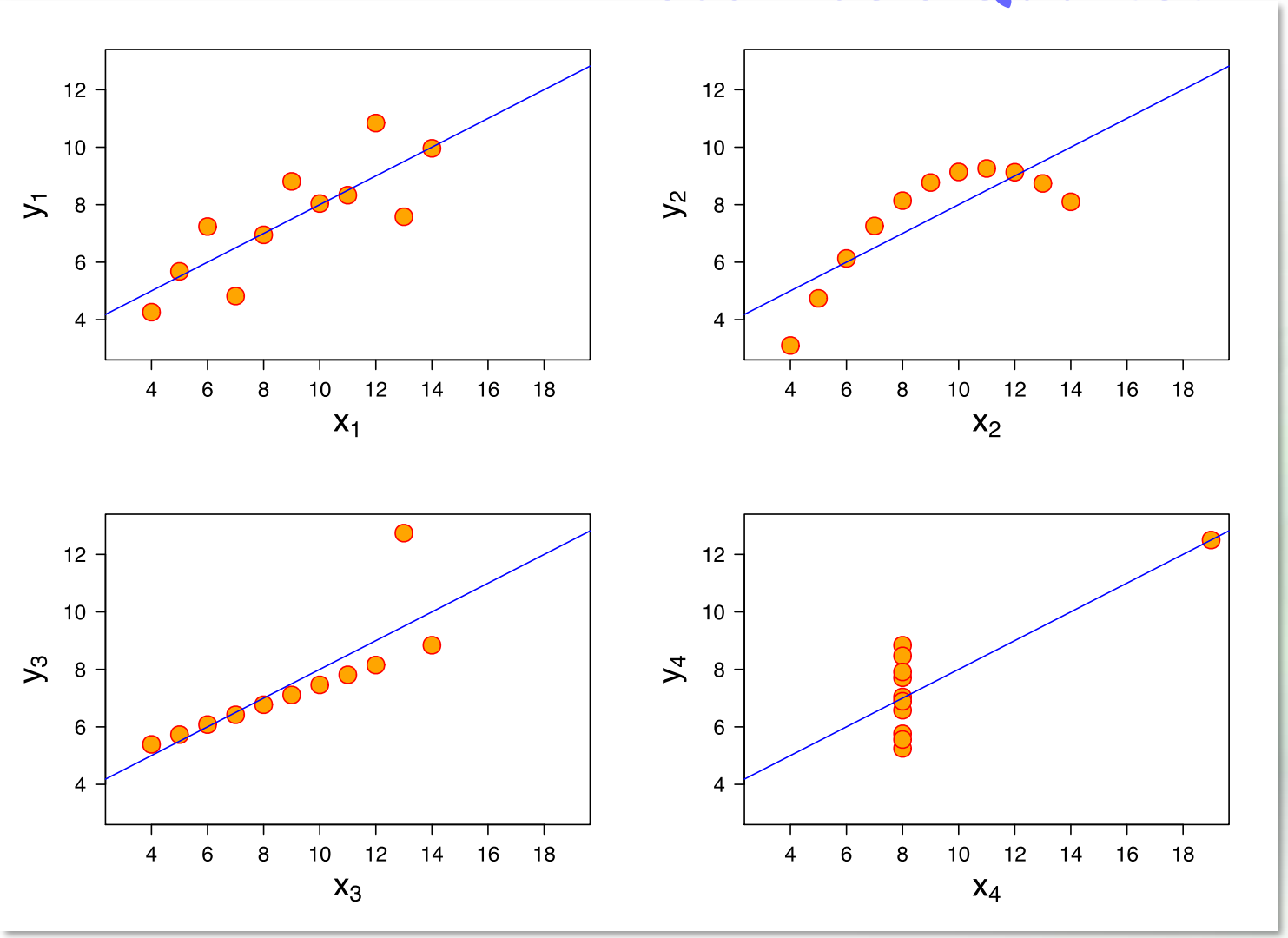
Mean of y

Variance of y

Linear Regression Line

Let's Look at the Data Visually

Anscombe's Quartet



What is Data Visualization?

- *Visualize*: To *form* a mental vision, image, or picture of (something not visible or present to the sight, or of an abstraction), to *make visible to the mind or imagination*
- *Visualization*: *Use of computer graphics* to create visual images which aid in the *understanding of complex*, often *massive representations of data*
- *Data Visualization*: The *process of discovering implicit but useful knowledge* from large data sets using visualization techniques, providing an accessible way to *see and understand trends, outliers, and patterns in data*

What is Data Visualization?

More Definitions

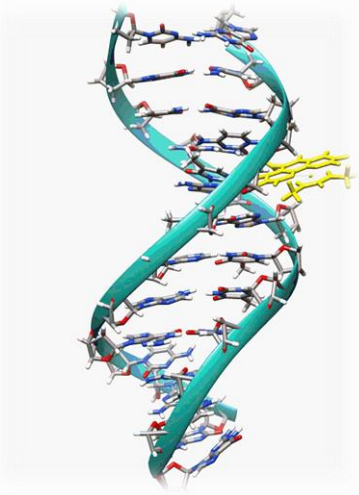
- “Transformation of the symbolic into the geometric” (McCormick et al. 1987)
- “... finding the artificial memory that best supports our natural means of perception” (Bertin 1967)
- “The use of computer-generated, interactive, visual representations of data to amplify cognition” (Card, Mackinlay, & Shneiderman 1999)

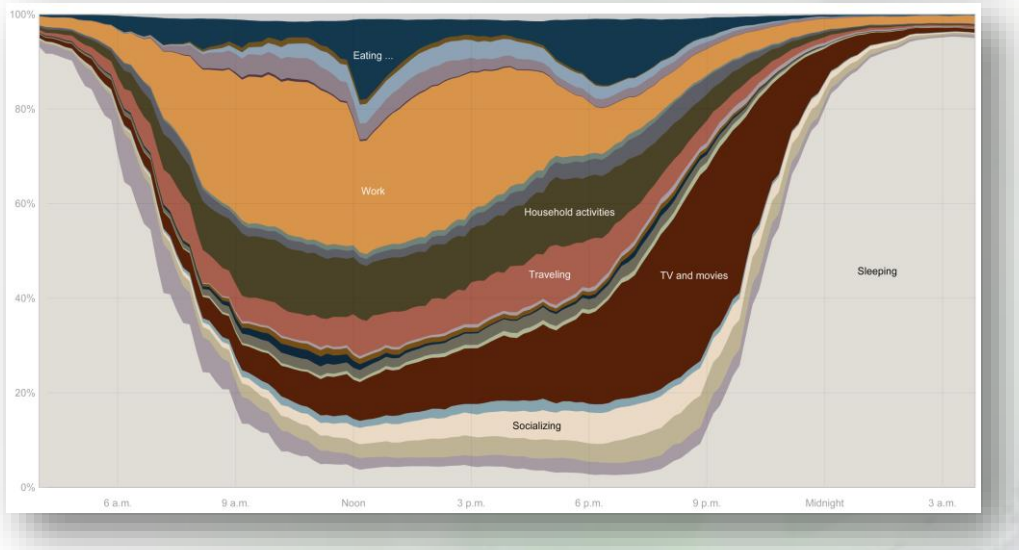
Two Types of Visualization

Exploration vs Explanation

Data Exploration Visualization

Figuring out what is true



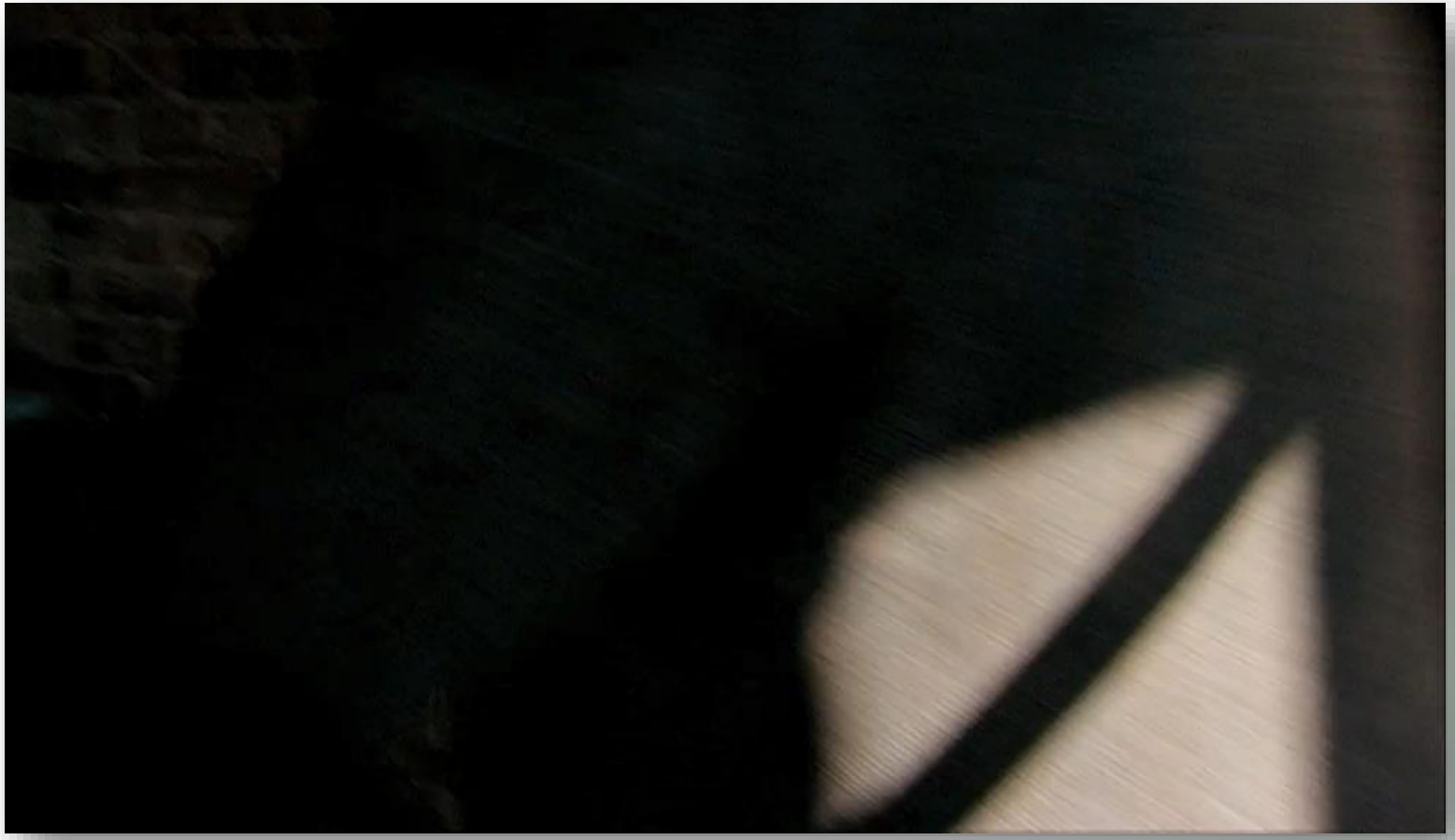


Data Presentation Visualization

Convincing other people it is true

The Power of Data Visualization

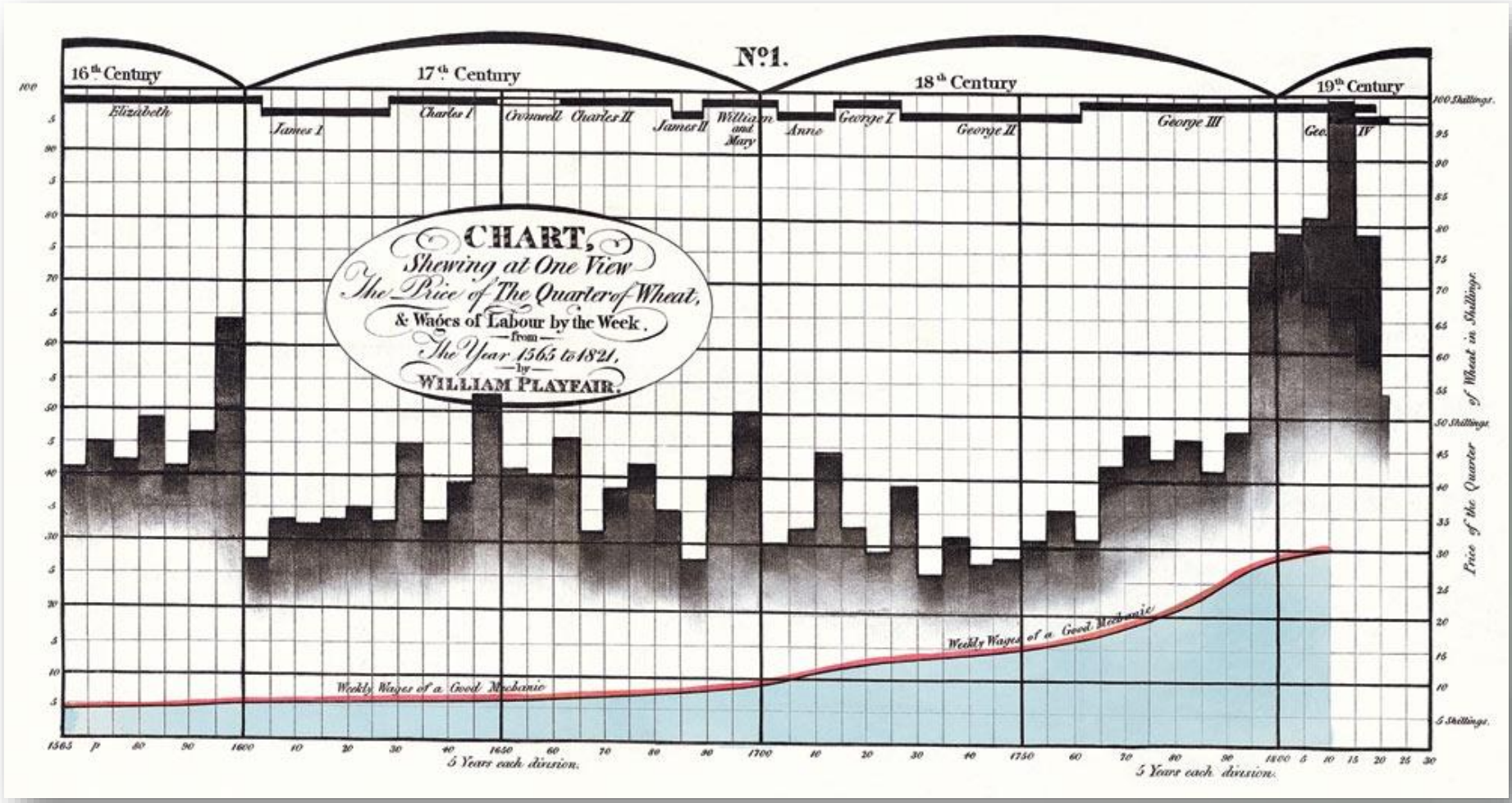
200 Countries, 200 Years



Hans Rosling

William Playfair's Early Chart

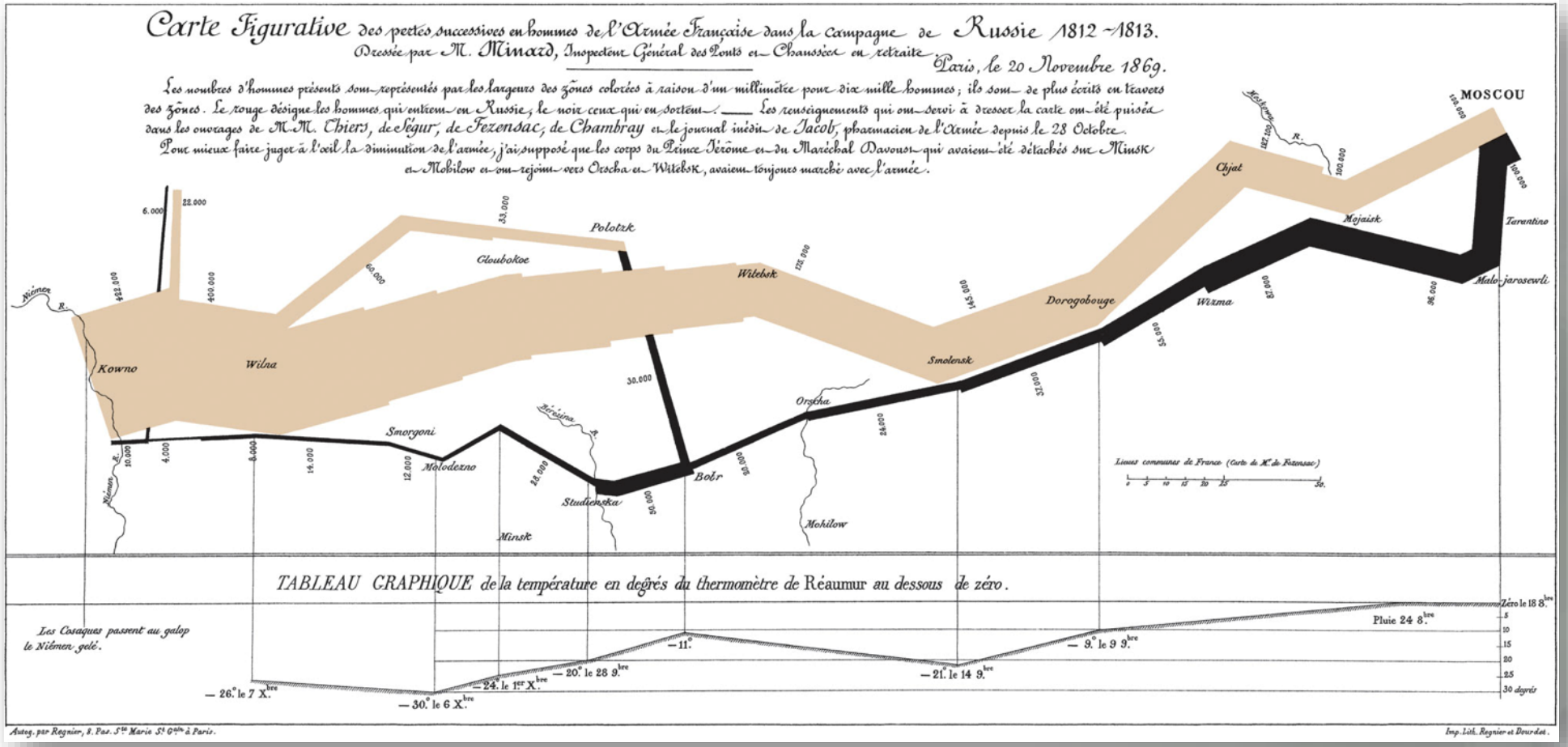
Founder of Graphical Method



(Playfair, 1786)

Visualization of Napoleon's Army

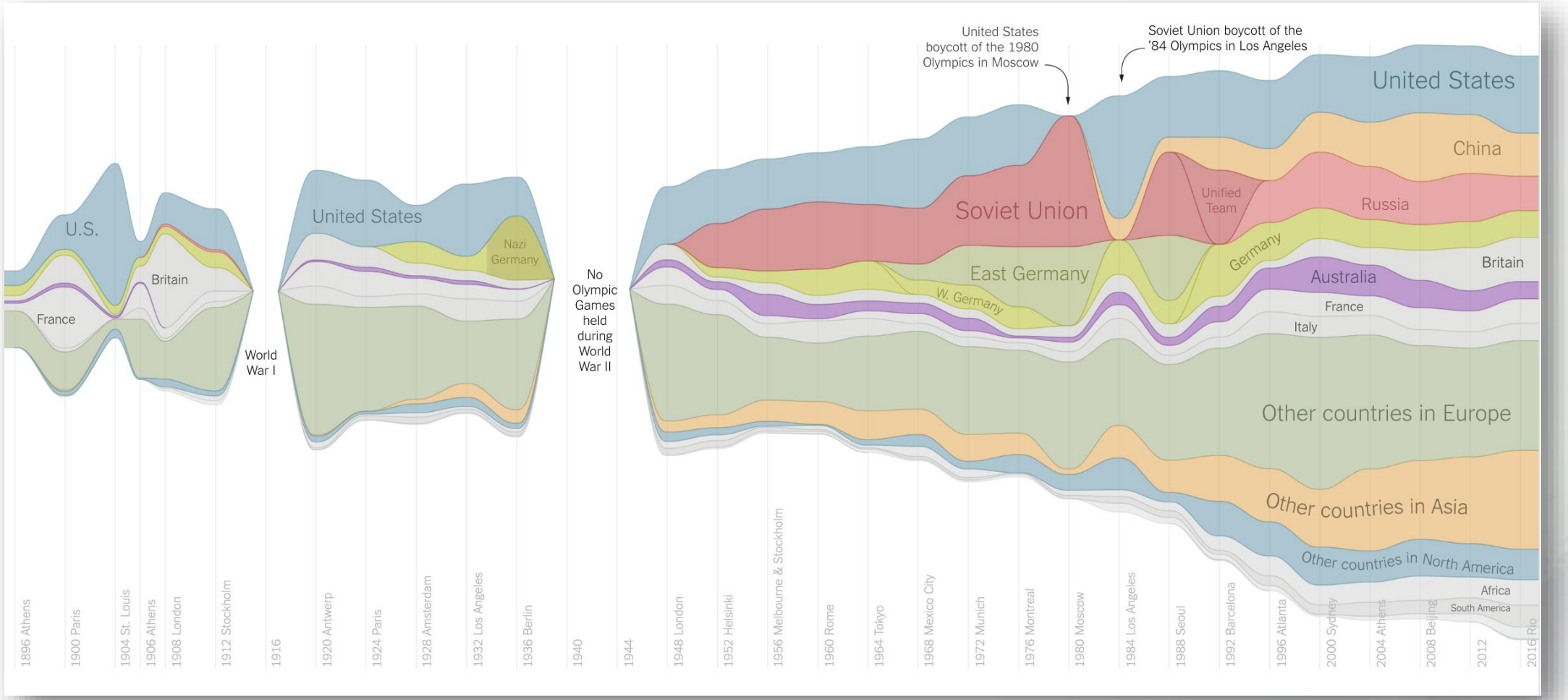
(Charles Joseph Minard 1812)



Minard's chart shows six types of information: geography, time, temperature, the course and direction of the army's movement, and the number of troops remaining.

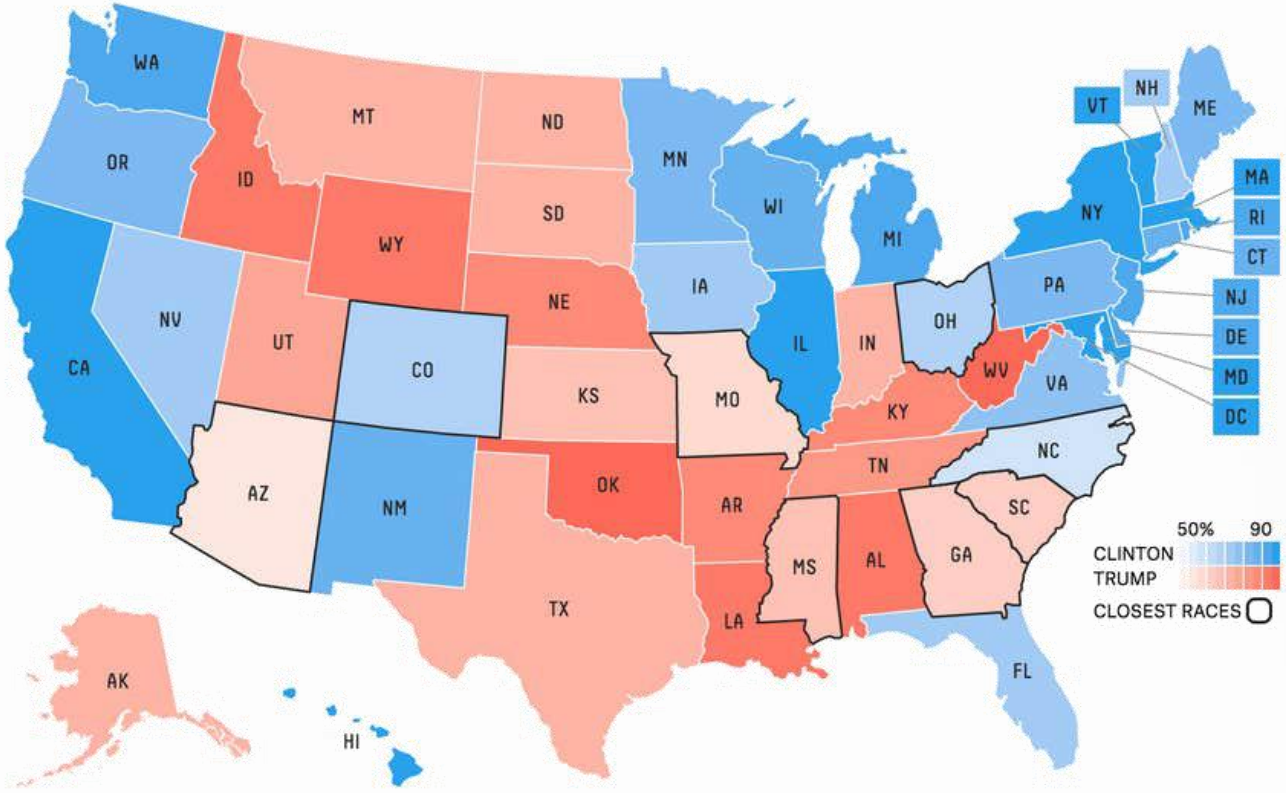
Olympic medals

The New York Times, 2016

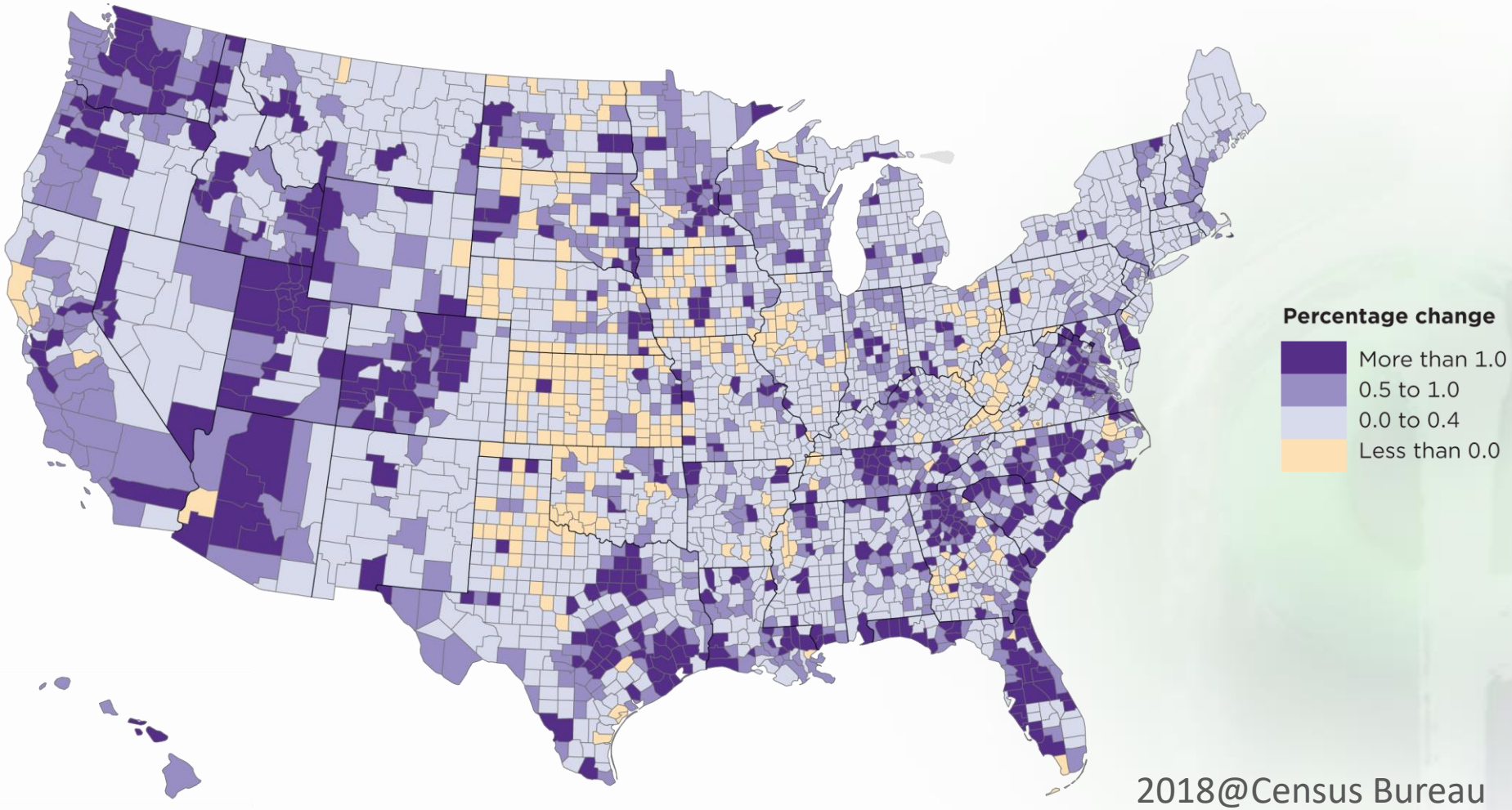


Who Will Win the President?

2016@Independent



Housing Unit Percent Change



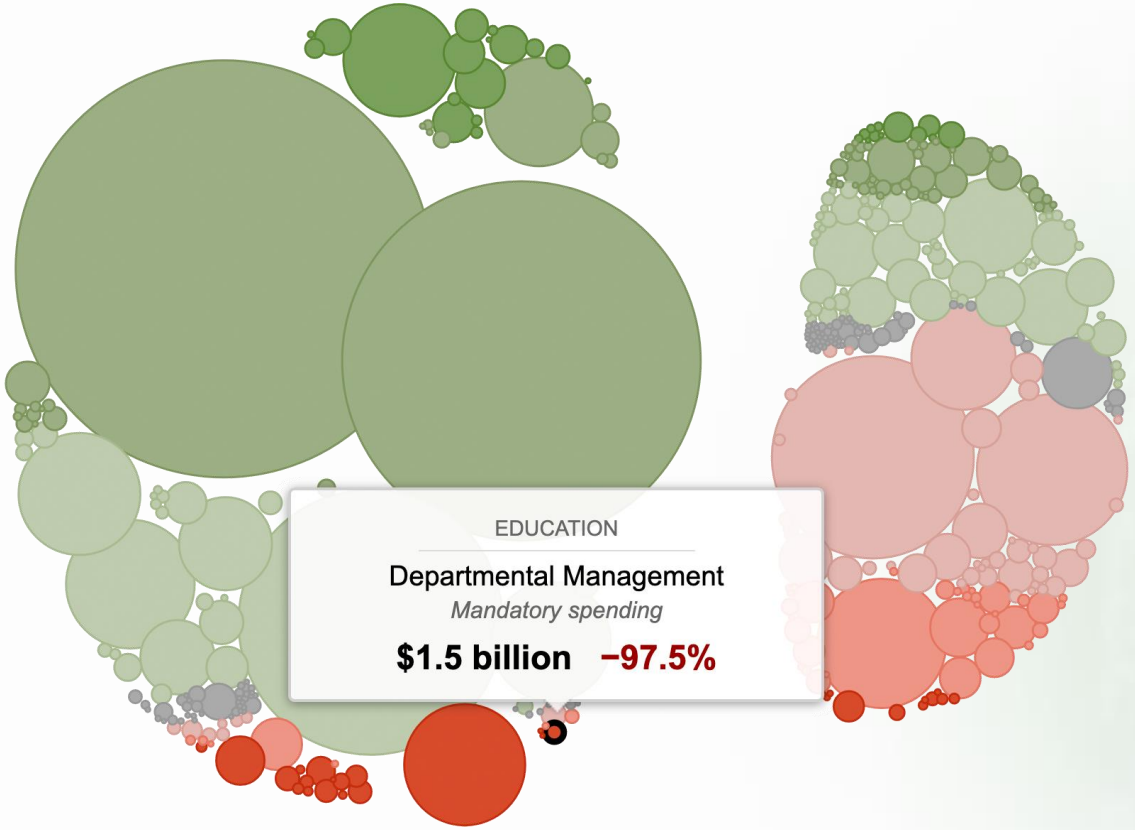
2018@Census Bureau

Obama's 2013 Budget Proposal

Mandatory

$2.5 trillion

About 70 percent of budgetary spending is controlled by existing laws, including entitlements like Medicare, Medicaid and Social Security.



Discretionary

$1.1 trillion

Only about 30 percent of the budget is controlled by the annual budget process. Last August, the White House and Congress agreed to a cap on this spending.

2012@The New York Times

Visualization Tools

Which software should you use to build data graphics?

Visualization Tools

Categories

Basic Productivity Applications

Google Chart

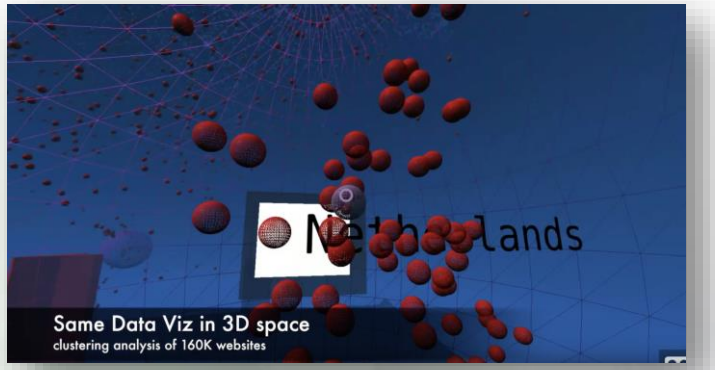
Common productivity tools that are *good enough* for *most* visualization tasks with *basic chart types* as well as some sophisticated displays

Visualization Tools

Categories

Basic Productivity Applications

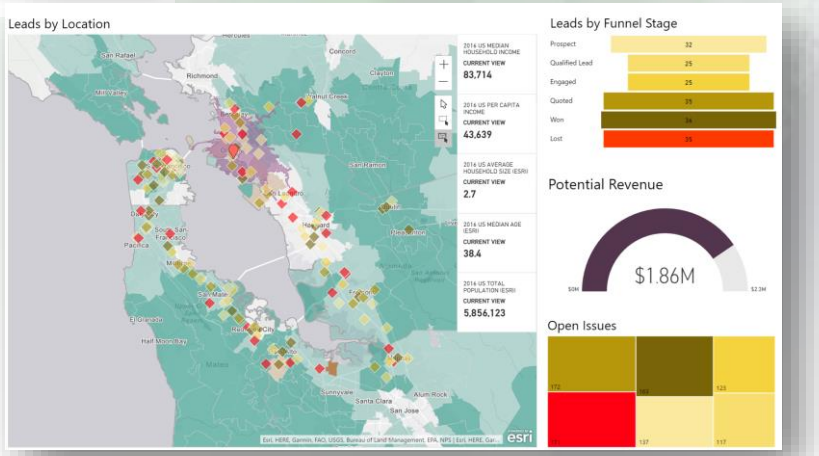
These applications focus on usability through *a drag and drop interface*. They are designed for everyone from *novices* to *expert* visualization designers and analysts.



QUANTUMVIZ

esri

+ a b | e a u



Visualization Software

Visualization Tools

Categories

Basic Productivity Applications

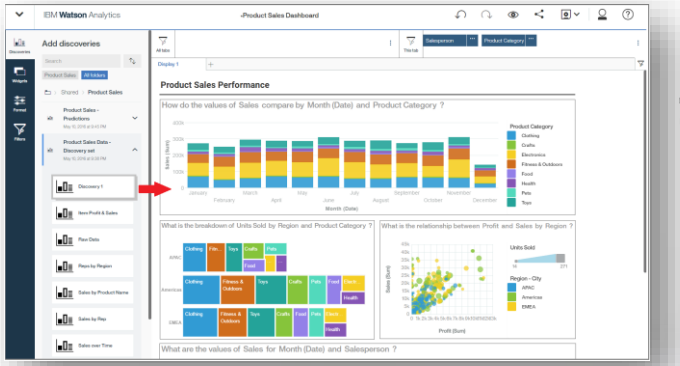
These applications focus on usability through *a drag and drop interface*. They are designed for everyone from *novices* to *expert* visualization designers and analysts.

Tableau is one of the leading data visualization software packages

- Various *data sources* and *files*, from *basic files*, *databases*, to *live* data sources
- User-friendly *drag* and *drop* interface for visualizing
- Strong data *preprocessing* capabilities
- The workbooks can be *published* to the web and *shared* securely across an organization

Visualization Software

Visualization Tools
Categories



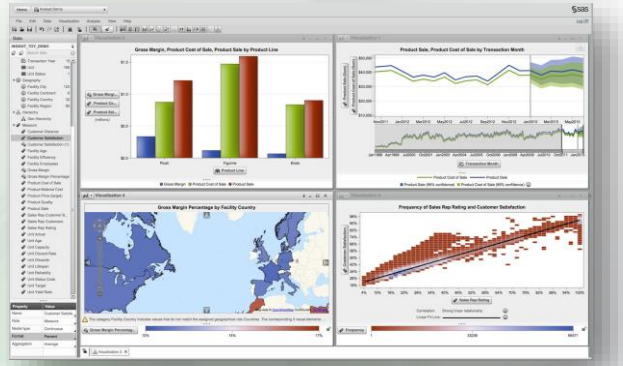
Basic Productivity Applications

Research describes *business intelligence* tools are the *next wave* of *advanced visualization software*. They provide the ability to show *dynamic content*, *visual querying*, *multiple dimensional-linked visualizations*, *animated visualizations*, *personalization*, and *alerts* based on changing data

Business Intelligence Tools

IBMWatson

sas



Microsoft Power BI

Visualization Software

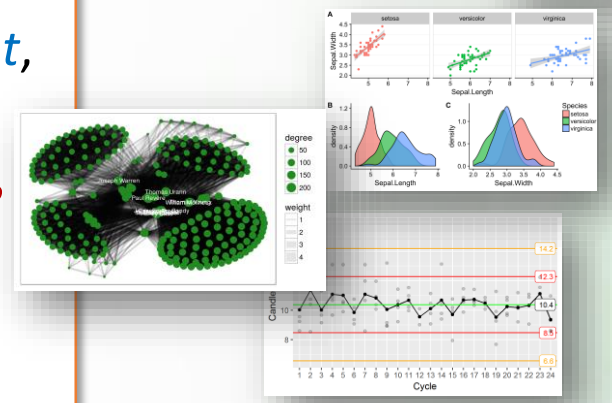
Visualization Tools

Categories

Basic Productivity Applications

Business Intelligence

For *developers*, *analysts*, and *designers* who want to visualize data in their *own programming environment*, there are several contenders. Most programming languages have *data graphic packages*. *Python* and *R* have a sophisticated set of libraries or packages for data visualization. In addition, there are numerous *JavaScript* libraries for web-based data graphics.



Visualization Software

Programming Packages

Selecting Tools for Visualizing Data

A Criteria

When *evaluating* a new data visualization tool, consider the following:

Output: Can you *publish* visualizations to the *web*, create high quality *print graphics*, and *embed* them in other applications?

Interoperability: How easily can you connect to *other data sources*? Does the software allow you to import diverse file types?

Sharing: Can others *view* and *edit* your visualization and analysis?

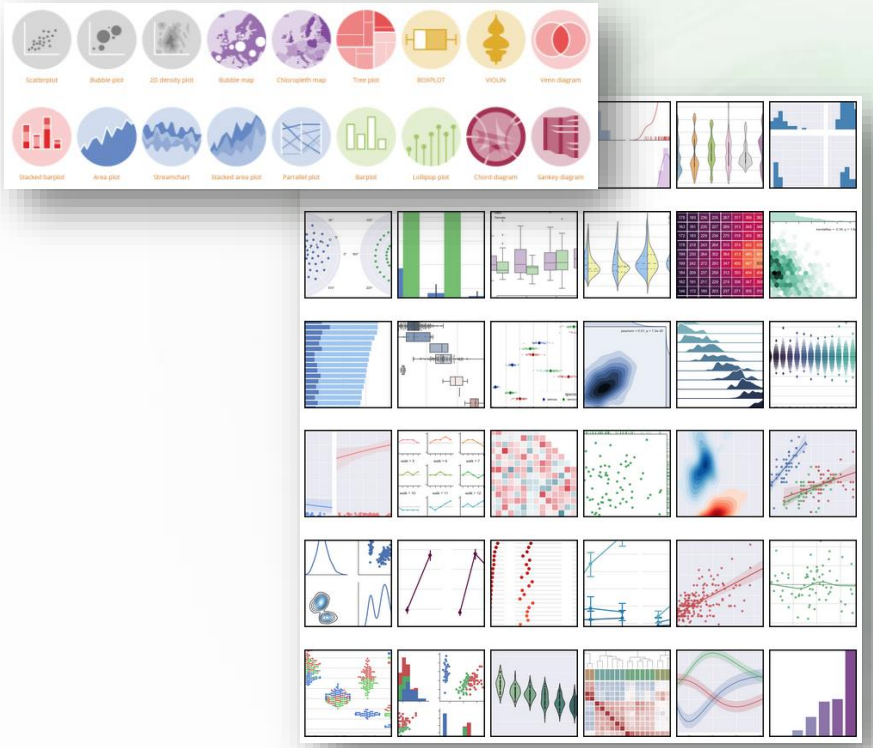
Display Types: What *types of visualizations* do you need to build? *Maps*, *networks*, and *text-based visualizations* are not available in every tool.

Data Exploration: Do you need a tool to *explore your data* and *present it visually*? Features such as visual querying are *not standard* in every tool.

Simplicity: Do you want to create charts and graphs *quickly*?

Persistence: Do you think you'll need to *revise* the visualizations you created?

In This Course



References

- Michael Sandberg (2013), *DataViz History: Charles Minard's Flow Map of Napoleon's Russian Campaign of 1812*.
- Costigan-Eaves, P., & Macdonald-Ross, M (1990), *William Playfair (1759-1823)*. Statistical Science, 318-326.

Acknowledgements

Some of the materials are adapted from:

- Neha Vaidya, 2019
- Teemu Roos, 2018