

Introduction to Data Visualization

BAN140 - Section NBB /NCC

Mufleh Al-Shatnawi, Ph.D., P.Eng.,



Week 1

Week Topics



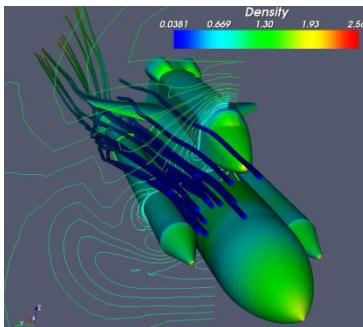
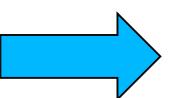
Current Week

- Introduction
- Communicating Visually

What is data visualization?

- **Data visualization** is the process of acquiring, interpreting and comparing data in order to clearly communicate complex ideas, thereby facilitating the identification and analysis of meaningful **patterns**.
- Computer-based visualization systems provide visual representations of datasets designed to help **people carry out tasks more effectively**

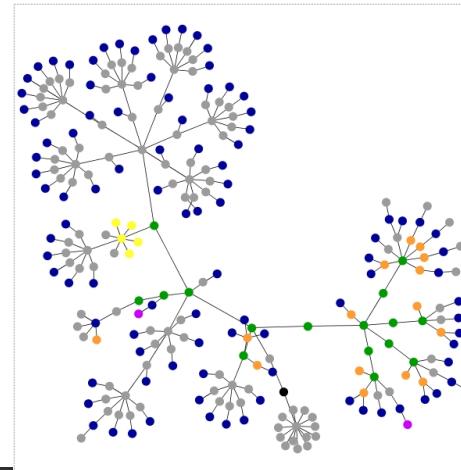
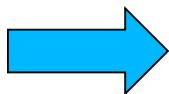
0265540 132304 132732 032051 037334 024721 015013 052325 001662
0265550 026557 026559 054506 043244 071095 132123 135235 122144
02655700 144210 054206 0446700 042650 165230 1707037 006355 006254
02655720 134453 124327 1765005 027093 107623 1707024 073702 067224
02655730 134453 124327 1765005 027093 107623 1707024 073702 067224
02655760 107204 102316 177451 0465040 020223 001724 030477 0466573
02656000 171317 116055 155117 134446 167232 041405 147127 050505
02656020 004137 046472 124015 134480 178550 053517 044635 021135
02656040 041023 074017 127113 003214 037026 037640 065171 123424
02656100 067701 037400 140000 165341 072410 100032 125455 056646
02656120 026411 005174 116200 077733 024434 012546 112544 102345
02656140 040223 050170 055161 164633 047154 126525 112544 102345
02656160 040223 050170 055161 164633 047154 126525 112544 102345
02656200 016041 176055 042768 025015 176531 017234 110060 014515
02656220 016041 176055 042768 025015 176531 017234 110060 014515
02656240 137055 062276 161755 115466 005322 182567 073236 002655
02656260 171466 125161 117155 065763 016127 014460 112765 055527
02656280 072057 100307 040627 070652 173013 121251 125126 140110
02656300 072057 100307 040627 070652 173013 121251 125126 140110
02656320 060115 014356 015164 067027 120200 070242 033065 131394
02656340 170601 171010 040437 127227 124446 136631 041462 116321
02656360 157504 163037 1655330 074251 024520 114433 167227 030635
02656420 157504 163037 1655330 074251 024520 114433 167227 030635
02656440 133834 106171 144160 010652 007385 1607316 100413
02656460 026553 0265720 0106530 121235 002893 140154 041237 0083276
02656480 026553 0265720 0106530 121235 002893 140154 041237 0083276
02656500 030374 144251 077734 015157 002513 173526 038531 150003
02656520 146207 015135 024446 130101 072457 040764 165513 156412
02656540 146207 015135 024446 130101 072457 040764 165513 156412
02656560 142166 123707 175121 071170 076357 037233 031136 015232
02656580 075074 016744 044055 102230 110063 038350 052765 124623



More about What is data visualization?

- Data visualization is a *graphic representation* that expresses the significance of data. It reveals insights and patterns that are **not immediately visible** in the raw data. It is an art through which information, numbers, and measurements can be made more understandable. According to (Friedman [2008](#)):

```
http://www.bu.edu/students/life/  
http://www.bu.edu/students/life/housing/  
http://www.bu.edu/students/life/dining-vending/  
http://www.bu.edu/students/life/phone/  
http://www.bu.edu/students/life/safety/  
http://www.bu.edu/students/life/transportation/  
http://www.bu.edu/students/life/activities/  
http://www.bu.edu/students/bu-global-orientation-registration/  
http://www.bu.edu/students/academics/  
http://www.bu.edu/students/academics/link/  
http://www.bu.edu/students/academics/admissions/  
http://www.bu.edu/students/academics/registration/  
http://www.bu.edu/students/academics/advising/  
http://www.bu.edu/students/academics/grades/  
http://www.bu.edu/students/academics/services/  
http://www.bu.edu/students/academics/support/  
http://www.bu.edu/students/health/  
http://www.bu.edu/students/health/services/  
http://www.bu.edu/students/health/counseling/  
http://www.bu.edu/students/health/facilities/  
http://www.bu.edu/students/health/clubsports/  
http://www.bu.edu/students/health/varsitysports/
```



Visualization offers a method for **seeing the unseen**.

Dictionary Definition

- Definition (www.oed.com)
- The **action** or **fact** of visualizing; the power or process of forming a mental picture or vision of something not actually present to the sight; a picture thus formed.

Check these data sets

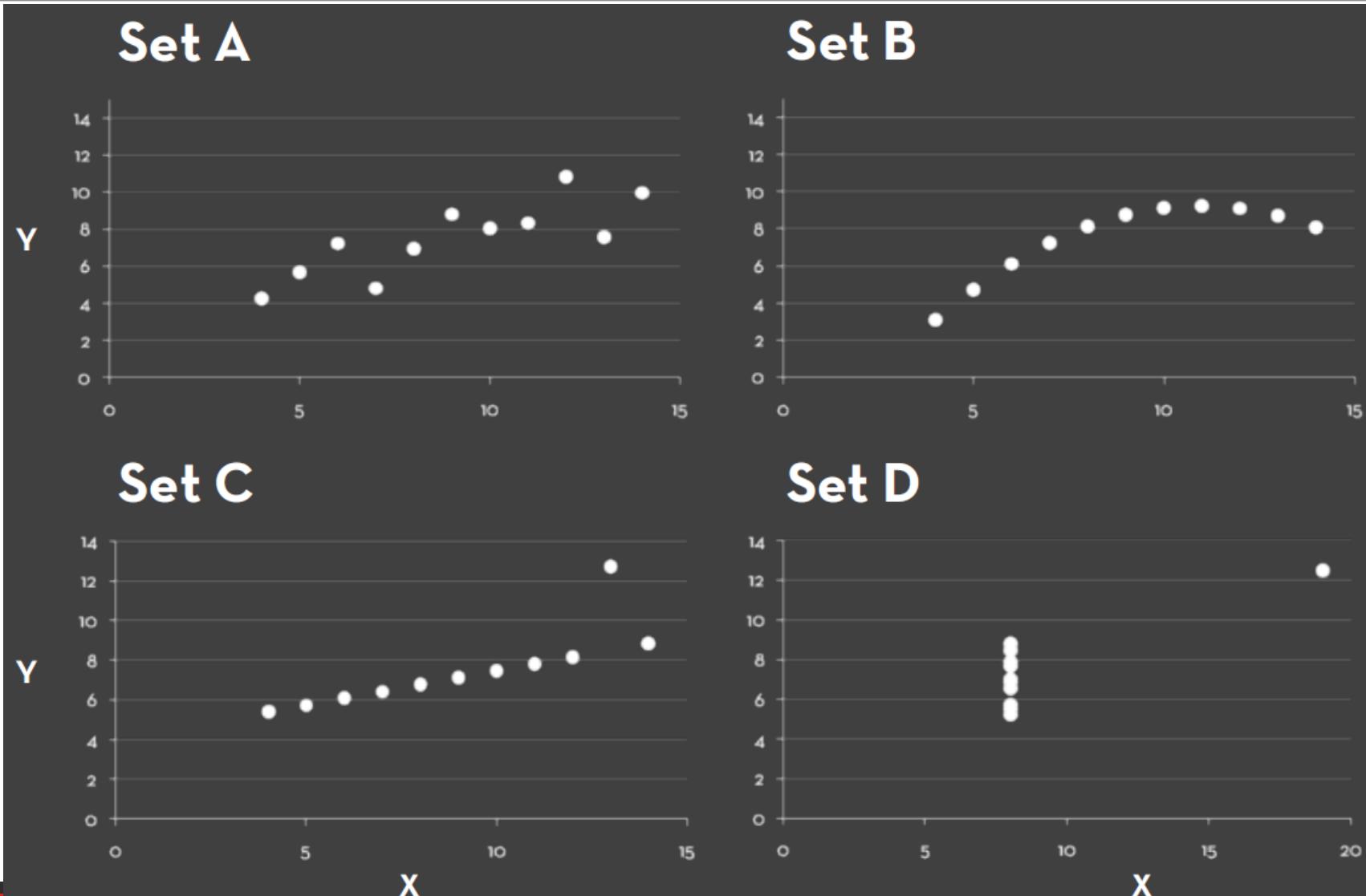
Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	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

Summary Statistics
 $\mu_X = 9.0$ $\sigma_X = 3.317$
 $\mu_Y = 7.5$ $\sigma_Y = 2.03$

Linear Regression
 $Y = 3 + 0.5 X$
 $R^2 = 0.67$

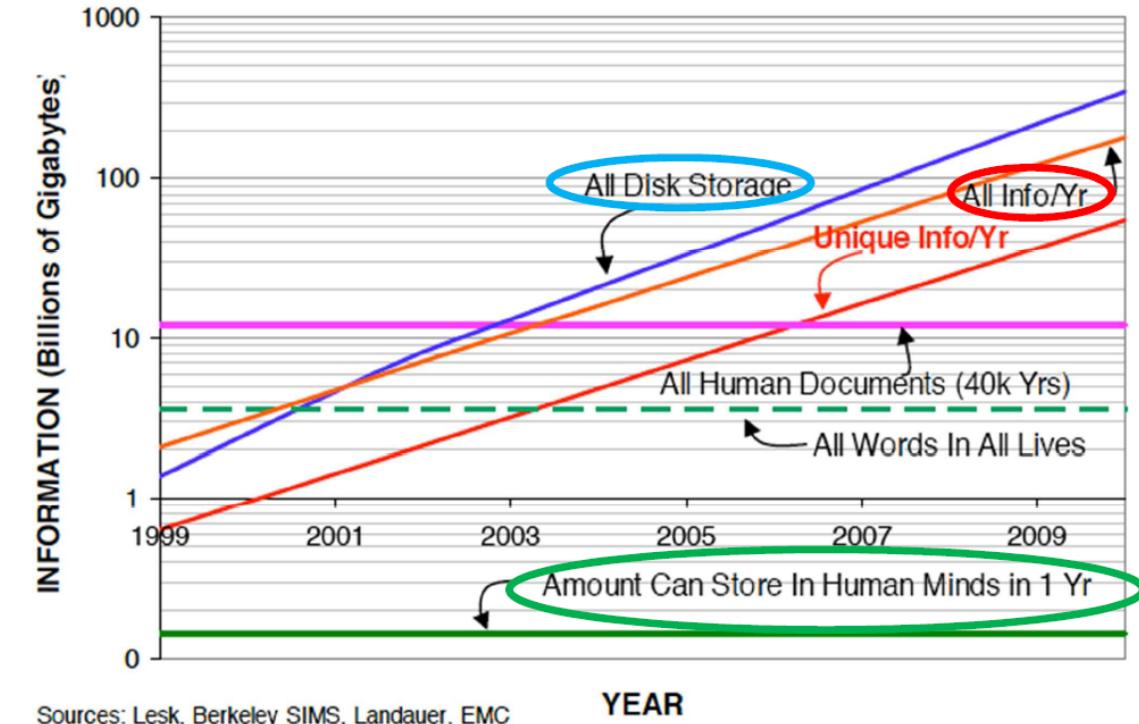
[Anscombe 73]

What Now?



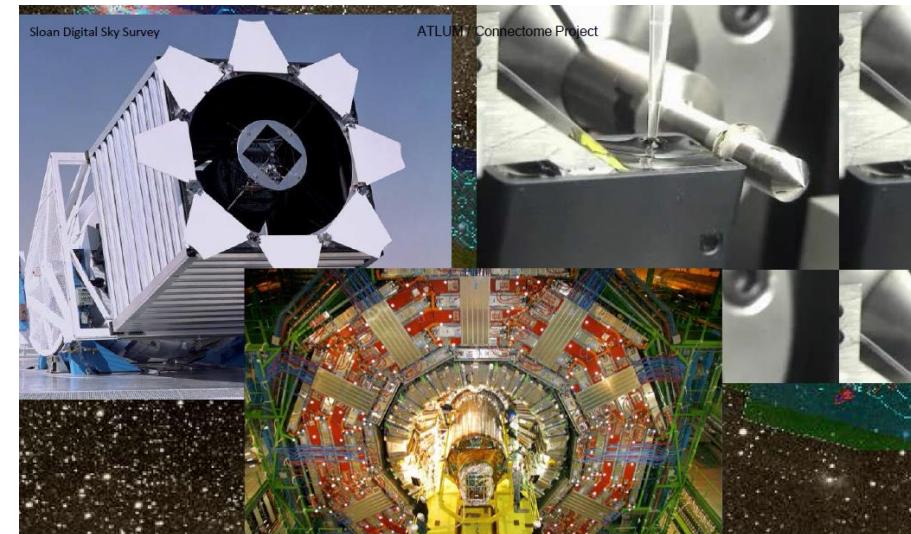
Why Data Visualization is important?

- Thanks to the rise of **social media**, the **popularity** of mobile devices, and service digitalization, data is available on any human activity that utilizes technology.
- The world produces **2.5 quintillion** bytes of data every day, and 90% of all data has been created in the last two years.



Why Data Visualization is important?

- Information Explosion
- Instrument Data Explosion
- “The Industrial Revolution of Data”

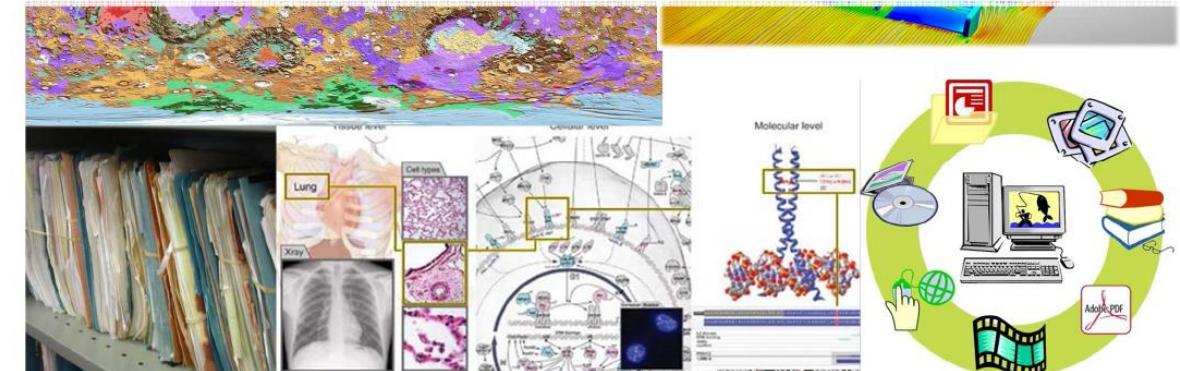


Why Data Visualization is important?

- The generated information is hugely valuable and makes it possible to analyze trends and patterns, and to use big data to **draw connections between events**. Thus, **data visualization can be an effective mechanism for presenting the end user with understandable information in real time.**

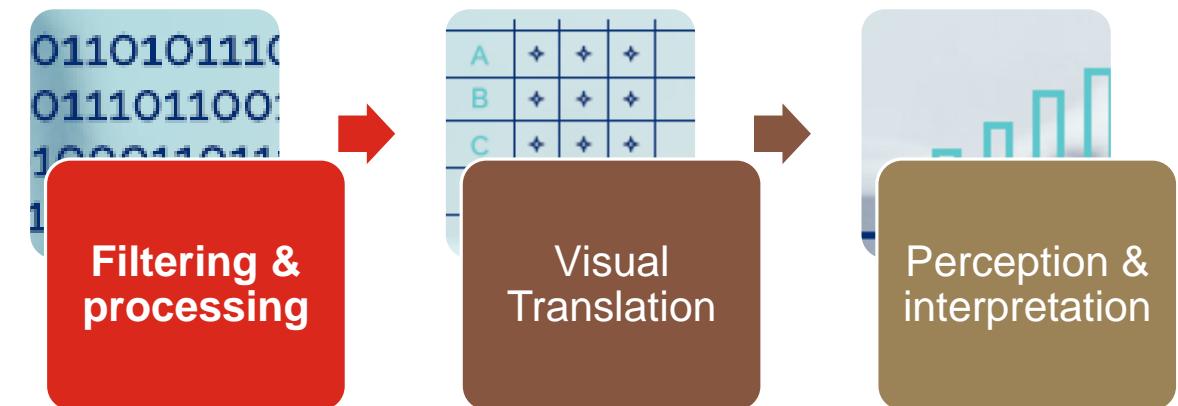


Data is generated everywhere and everyday

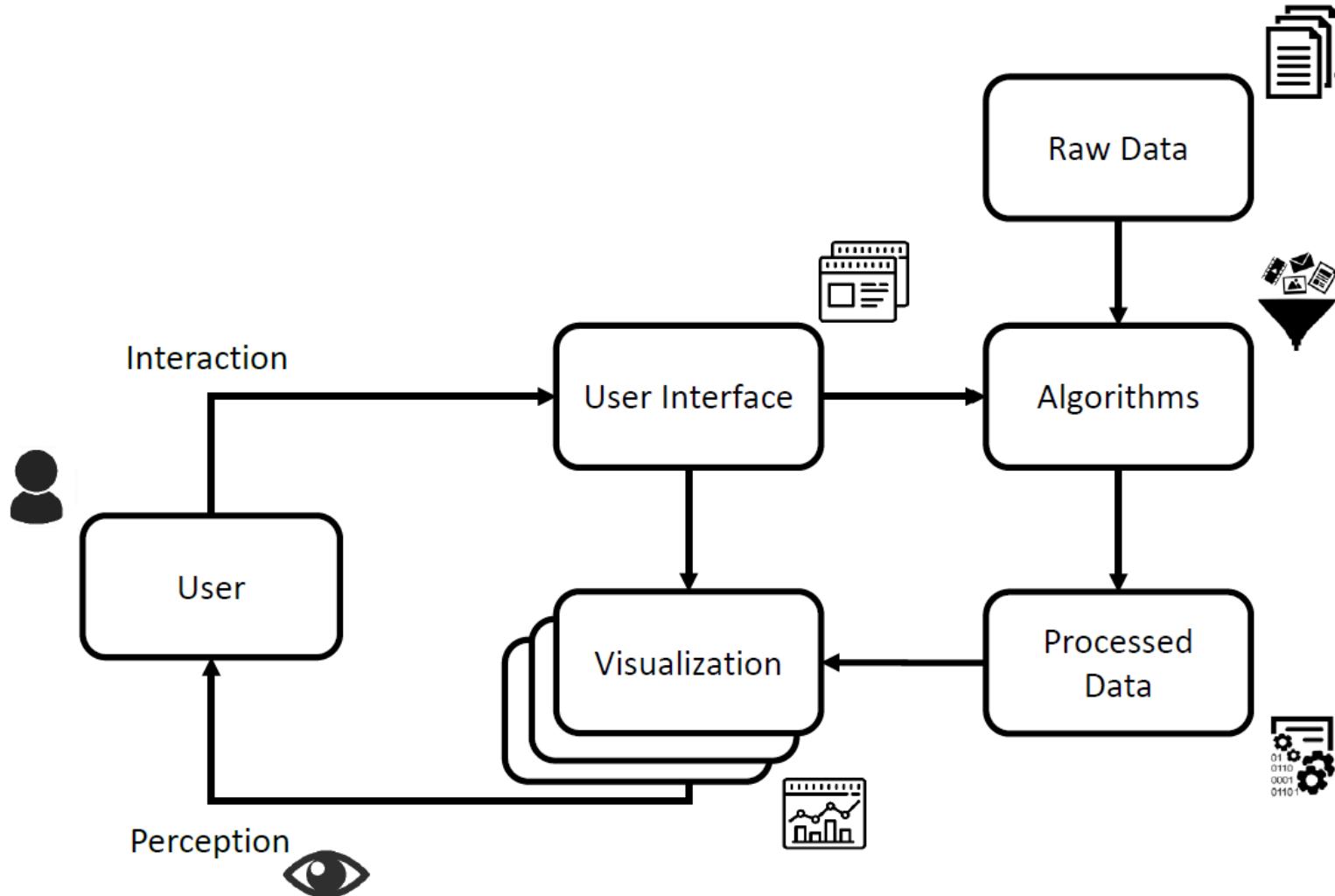


Top level: the Data Visualization Process

- **Filtering & processing:** Refining and cleaning data to convert it into information through analysis, interpretation, contextualization, comparison, and research.
- **Translation & visual representation:** Shaping the visual representation by defining graphic resources, language, context, and the tone of the representation, all of which are adapted for the recipient.
- **Perception & interpretation:** Finally, the visualization becomes effective when it has a perceptive impact on the construction of knowledge.



System Overview

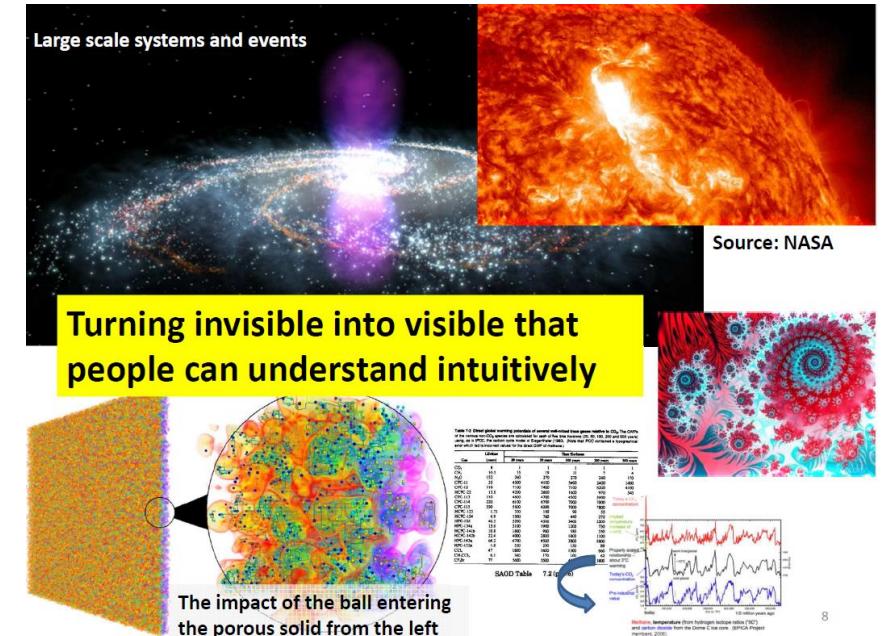


Why Data Visualization is needed?

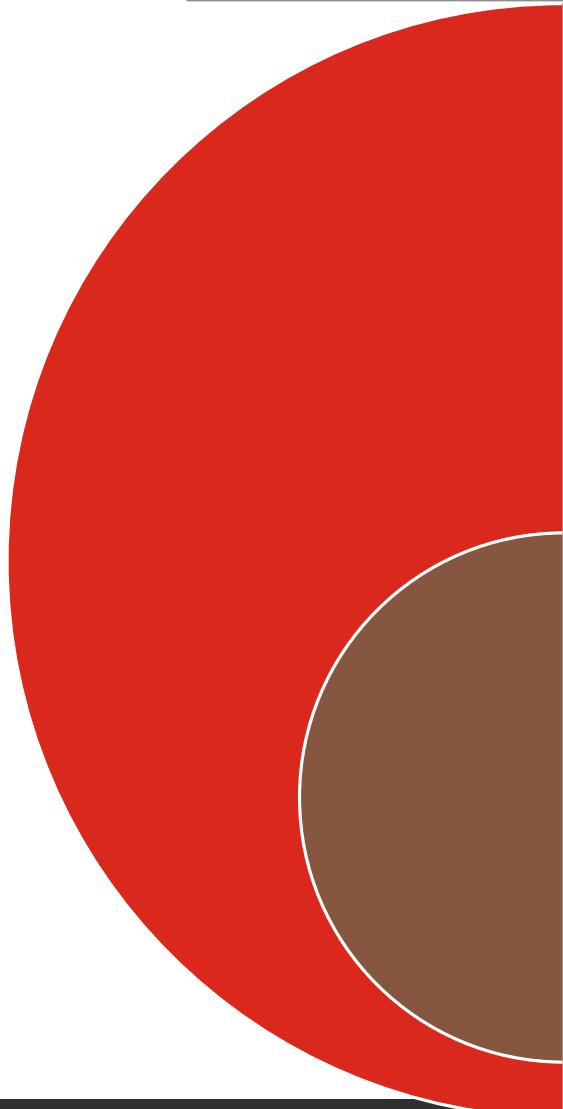
- **Intuitive**: People can easily understand graphs
- **Fast**: Our brains are great at identifying patterns, spotting trends and outliers in a tangible data visualization graph.
- **Flexible**: Allow users to represent different relationship in any given data.
- **Insightful**: Allows users to gain more **in-depth knowledge**.

Data Visualization used for

- Support reasoning about information (analysis)
 - Finding relationships
 - Discover structure
 - Quantifying values and influences
- Inform and influence others (communication)
 - Capture attention, engage
 - Tell a story visually
 - Focus on certain aspects, and omit others



More about Data Visualization usage



Can lead a user to

- Detect patterns
- Detect trends
- Detect correlations in data

Can then prompt a user to

- Draw inferences
- Anticipate potential trajectories and outcomes
- Ask new questions of the data that wouldn't have otherwise been considered

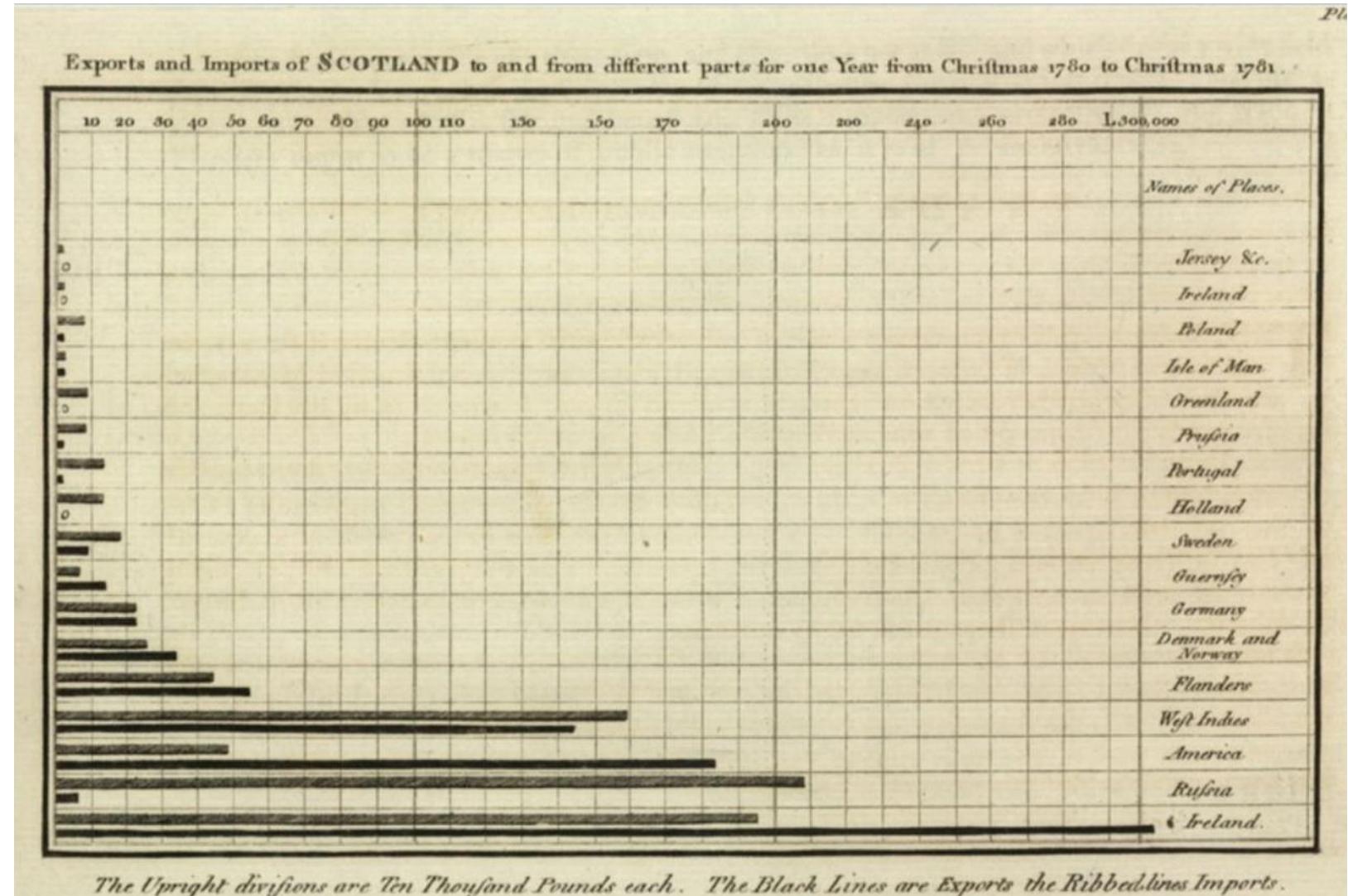
In class Discussion

- Some visualizations allow the user to filter out undesirable properties in the dataset. This produces a refined list of results to better help with decision making or understanding (think facets on a database).
- So let's apply these ideas to an example:
- <https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/>
- <https://informationisbeautiful.net/visualizations/diversity-in-tech/>
- <https://informationisbeautiful.net/visualizations/top-500-passwords-visualized/>

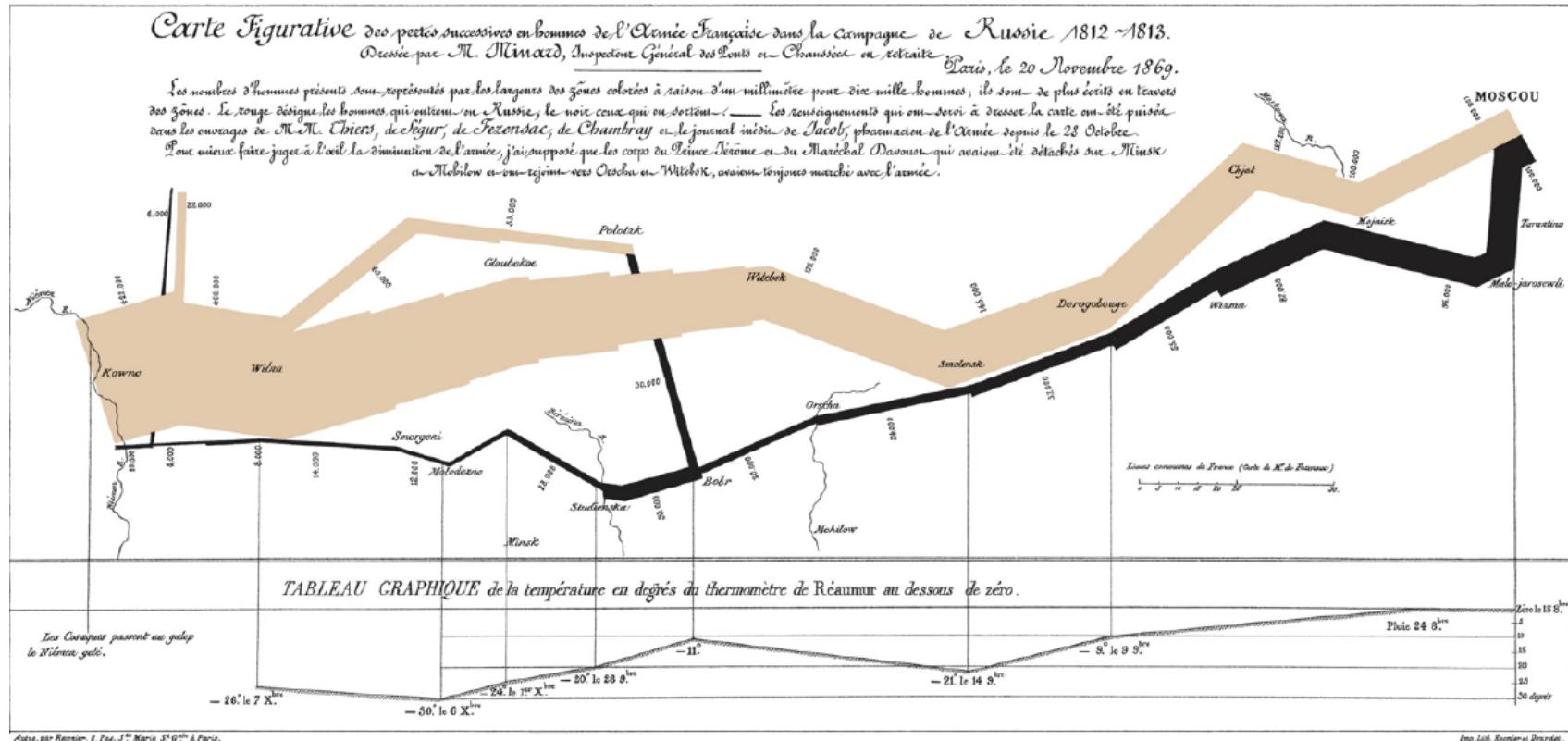
History of Data Visualization

William Playfair (1821) – line, bar charts, etc.

Chart of the Exports
and Imports of Scotland
to and from different
parts for one year from
Christmas 1780 to
Christmas 1781.



Charles Joseph Minard 1869: Napoleon's March



According to Tufte: “It may well be the best statistical graphic ever drawn.”
5 variables: **Army Size, location, dates, direction, temperature**

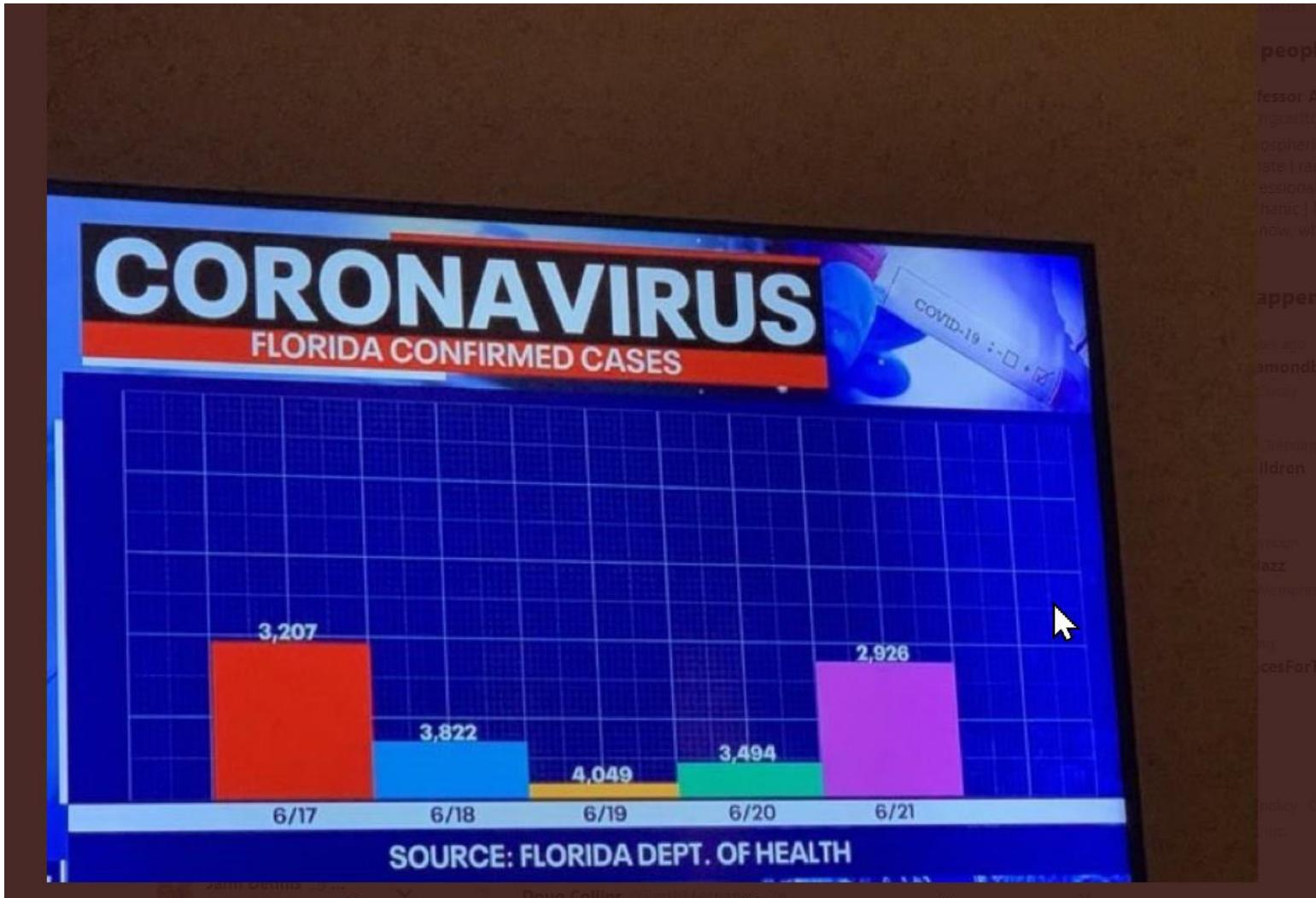
In Class Examples - Modern

- The famous Gapminder Video, Hans Rosling:
200 Countries, 200 Years, 4 Minutes
 - <https://youtu.be/jbkSRLYSojo>
- NY Times [words](#), [words](#), [numbers](#)
- [50 examples](#) (from June 2009, somewhat dated)
- [D3 Gallery](#)
- [Netflix Queues](#)
- [Unemployment Visualization](#)

Cases for Bad or Ugly visualizations

Courtesy of WTFViz.net

Example 1



<https://viz.wtf/>

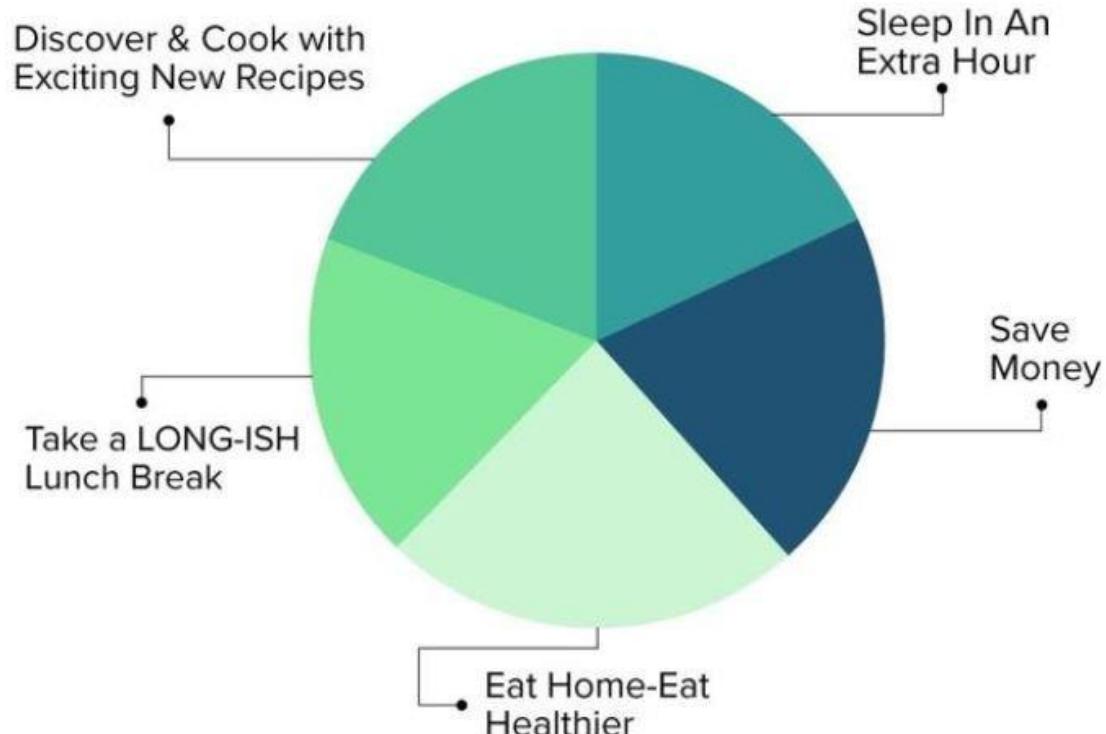
Example 2



<https://viz.wtf/>

Example 3

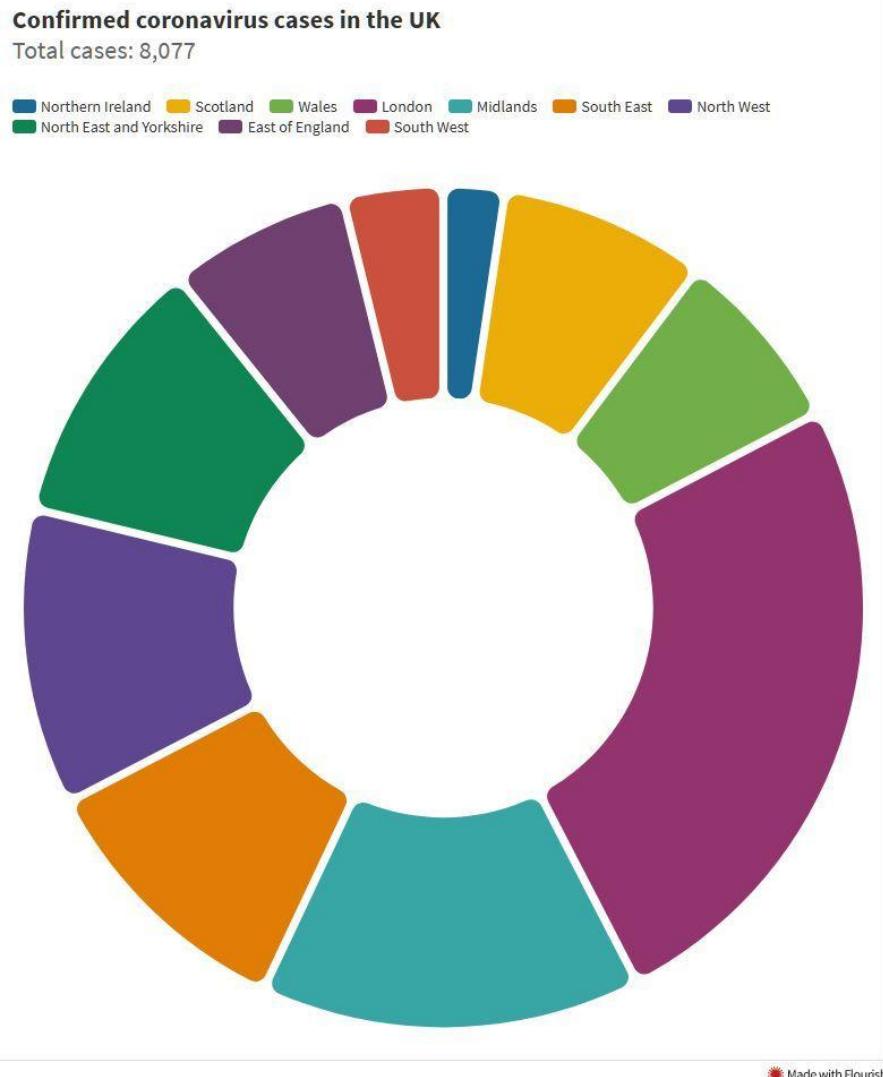
Advantages of Working From Home



<https://viz.wtf/>

Example 4

Select an appropriate view for your data



<https://viz.wtf/>

Thinking about Creating a Data Visualization

- Determine what you are ultimately trying to communicate
- Determine which visualization best achieves your aim in the clearest manner possible
- Check
 - [The Data Visualization Catalogue](#)

Further Considerations

- What data (or subset of data) is **relevant** for your aim?
- Who is your **audience**?
- How will you **encode** your data?
- How will you **structure** the visualization?
- How will you demonstrate the **relationship** between data?
- Determine the **scale** of your visualization.
- Which elements will the user **interact** with?
- Will a user **intuitively understand** how the visualization works and what it represents?

Data visualization Vs data narrative

Data Narrative

The data narrative include:

- Context
- Clarity
- Emotional connection
- Engagement

Storytelling with data

“I think people have begun to forget how powerful human stories are, exchanging their sense of empathy for a fetishistic fascination with data, networks, patterns, and total information ... Really, the data is just part of the story. The human stuff is the main stuff, and the data should enrich it.”

Jonathan Harris, Creator We Feel Fine
<http://number27.org/wefelfine>

Examples

ProPublica

Workers Comp Benefits: How Much is a Limb Worth?

<https://projects.propublica.org/graphics/workers-compensation-benefits-by-limb#>

Faces of Fracking

California's Getting Fracked

<http://www.facesoffracking.org/data-visualization/>

Visualization software

Software Tools for analyzing and data visualization

- [Tableau](#),
- [Microsoft Excel](#)
- [Google charts](#)
- [Many Eyes: IBM's Free Online Data Visualization Tool](#)
- R Language ([Data Visualization with R](#))
- Statistical Analysis System [SAS](#)
- Statistical Package for the Social Sciences [SPSS](#)
-

References and Resources

- [Knaflic] Cole Nussbaumer Knaflic, **Storytelling with Data: A Data Visualization Guide for Business Professionals**, Wiley, 2017
 - Available online through Seneca Libraries: https://senecacollege-primo.hosted.exlibrisgroup.com/permalink/f/t3376v/01SENC_ALMA5146374280003226
- [Ryan] Lindy Ryan, **Visual Data Storytelling with Tableau**, Pearson Addison-Wesley, 2018
 - Available online through Seneca Libraries: https://senecacollege-primo.hosted.exlibrisgroup.com/permalink/f/t3376v/01SENC_ALMA5167006190003226
- [Healy] Kieran Healy, **Data Visualization: A Practical Introduction**, Princeton University Press, 2018.
 - Available (hardcopy) at Seneca Libraries: https://senecacollege-primo.hosted.exlibrisgroup.com/permalink/f/t3376v/01SENC_ALMA2172469250003226
- **A Reader on Data Visualization:** https://mschermann.github.io/data_viz_reader/
- **Data visualization:** https://en.wikipedia.org/wiki/Data_visualization