

Introduction to Data Visualization

BAN140 - Section NBB /NCC

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



Week 1

Week Topics



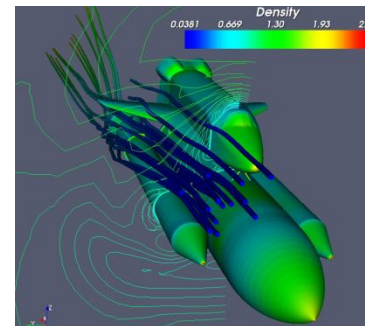
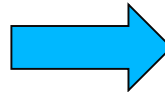
Current Week

- 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**

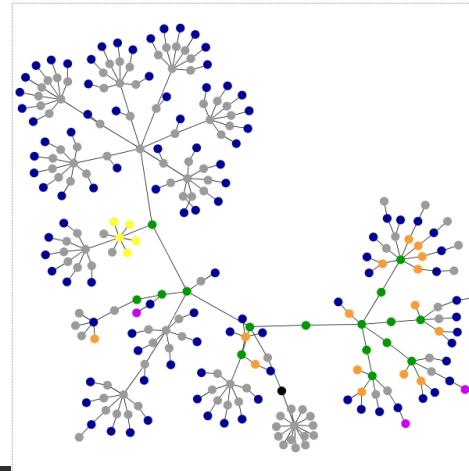
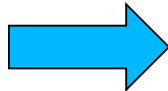
```
0265540 132204 133732 032051 037334 024721 015013 052226 001662
0265560 025537 064663 054606 043244 074076 124153 135216 126614
0265700 144210 056426 044700 042650 165230 137037 003655 006254
0265720 134453 124527 176005 027034 107614 170774 079702 067274
0265740 072451 007735 147620 061064 157435 113057 155356 114603
0265760 107204 102316 172451 046040 120223 001774 030477 046673
0265800 171317 116055 151117 134446 167210 041405 147127 050505
0266020 004137 046472 124015 134360 175550 053517 044639 021135
0266040 070176 047705 110754 175477 105392 076515 177364 056293
0266060 041023 074017 127113 003214 037026 037640 066171 123424
0266100 067701 037406 140000 155341 072410 100032 125455 056646
0266120 006716 071402 055672 132571 105645 170073 050376 072117
0266140 024451 007424 114200 077733 024434 012546 172404 102345
0266160 040223 050170 055164 164634 047154 126525 112514 032315
0266200 116041 176055 042766 025015 176314 017234 110060 014515
0266220 117156 030746 154234 125001 151144 163706 136237 164376
0266240 137055 062276 161755 115466 005322 132567 073216 002655
0266260 171466 126161 117155 065763 016177 014460 112765 055527
0266300 003767 175367 104754 036436 172172 150750 043643 145410
0266320 072074 000007 040627 070652 173011 002151 125132 140234
0266340 060115 014356 015164 067027 120206 070242 033065 131334
0266360 170601 170106 040437 127277 124446 136631 041462 116321
0266400 020243 005002 004146 121574 124651 006634 071331 102070
0266420 157504 160307 166330 074251 024520 114433 167273 030635
0266440 133614 106171 144160 101652 007365 026416 160716 100413
0266460 026630 007210 000630 121224 076033 140764 000737 003276
0266500 114060 042647 104475 110537 066716 104754 075447 112254
0266520 030374 144251 077734 015157 002513 173526 035531 150003
0266540 146207 015135 024446 130101 072457 040764 165513 156412
0266560 166410 067251 156160 106406 136770 030516 064740 022032
0266580 142165 123707 175121 071310 076357 032313 031136 015232
0266620 075074 016744 044055 102230 110063 033350 052765 172463
```



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/life/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**.

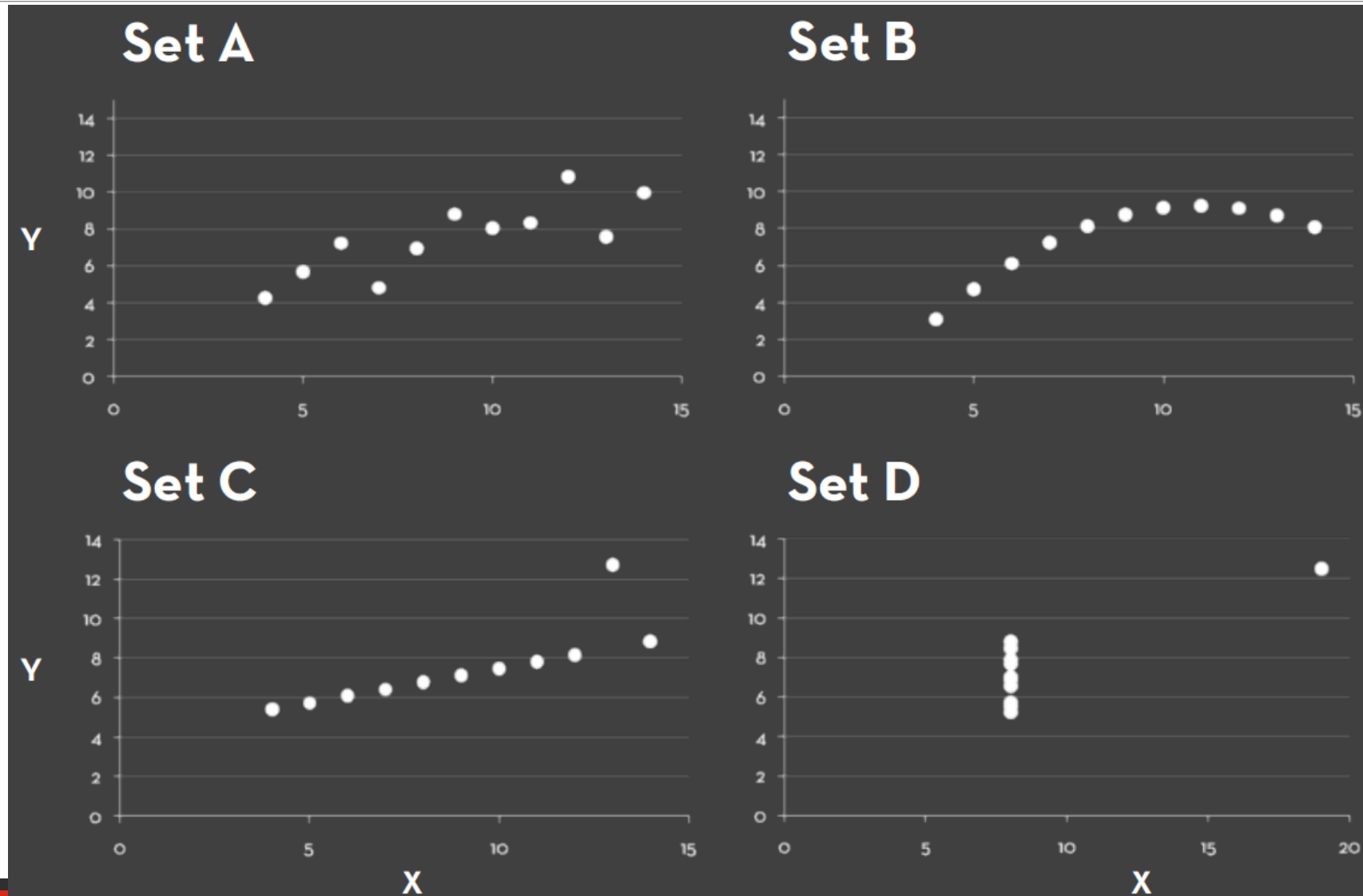
Dictionary Definition

- 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		Linear Regression					
$\mu_X = 9.0 \quad \sigma_X = 3.317$		$Y = 3 + 0.5 X$					
$\mu_Y = 7.5 \quad \sigma_Y = 2.03$		$R^2 = 0.67$					
				[Anscombe 73]			

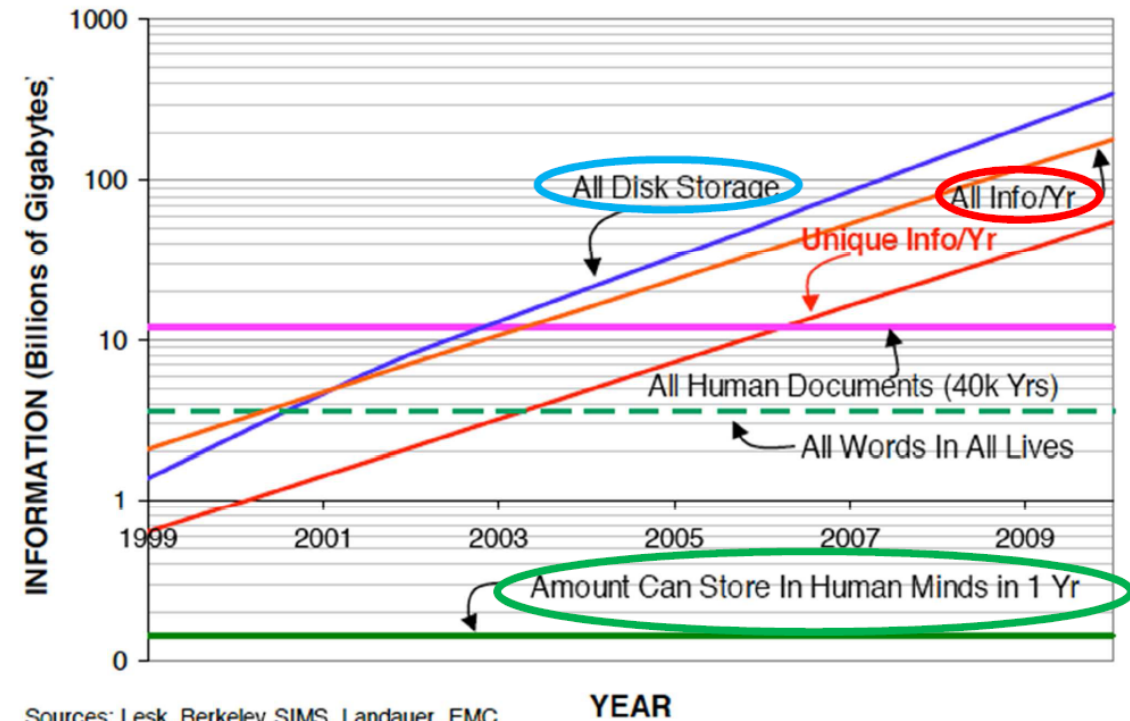
What Now?



Why Data Visualization is important?

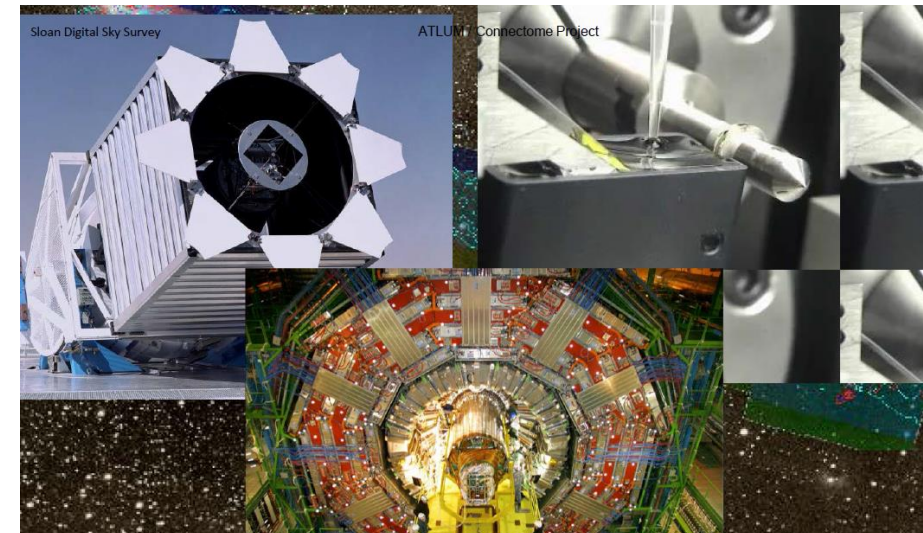
➤ Thanks to the rise of **social media**, the **popularity** of mobile devices, and service digitalization, data is available on any human activity that utilizes technology.

➤ The world produces **2.5 quintillion** bytes of data every day, and 90% of all data has been created in the last two years.



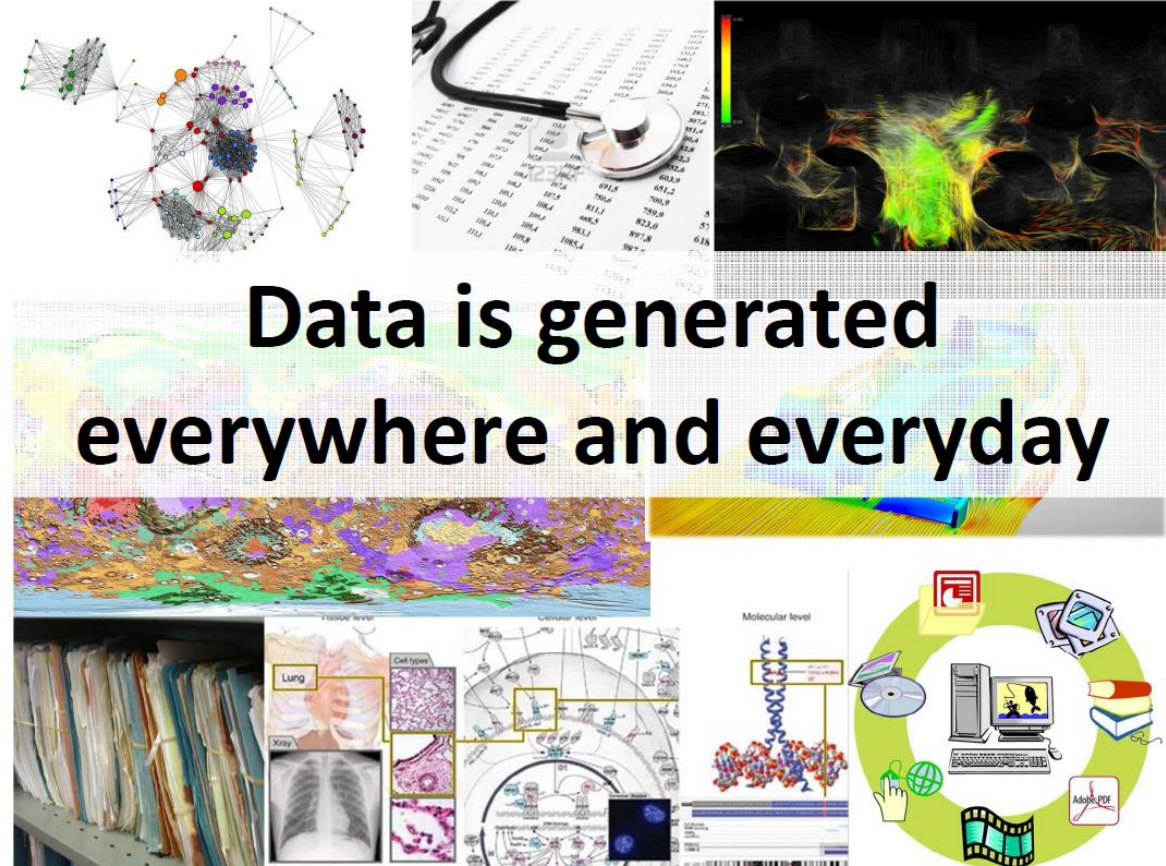
Why Data Visualization is important?

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



Why Data Visualization is important?

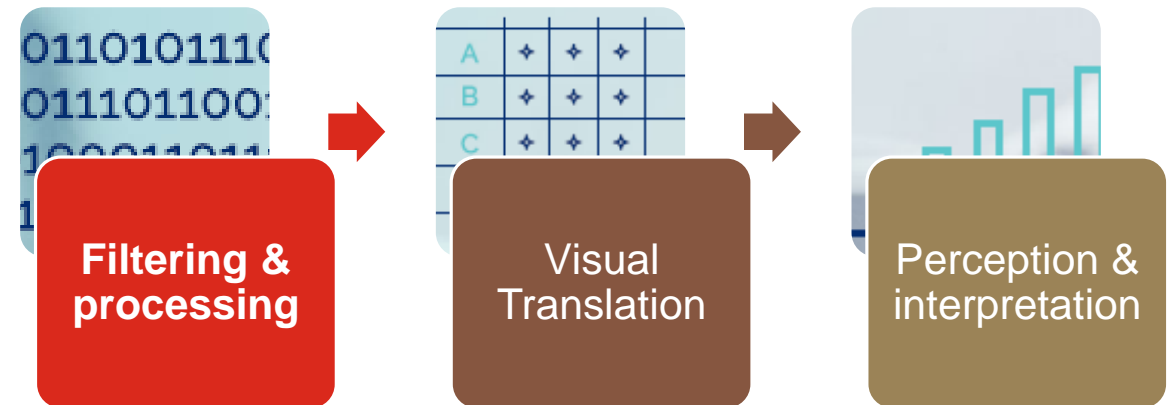
- The generated information is hugely valuable and makes it possible to analyze trends and patterns, and to use big data to **draw connections between events**. Thus, **data visualization** can be an effective mechanism for presenting the end user with understandable information in real time.



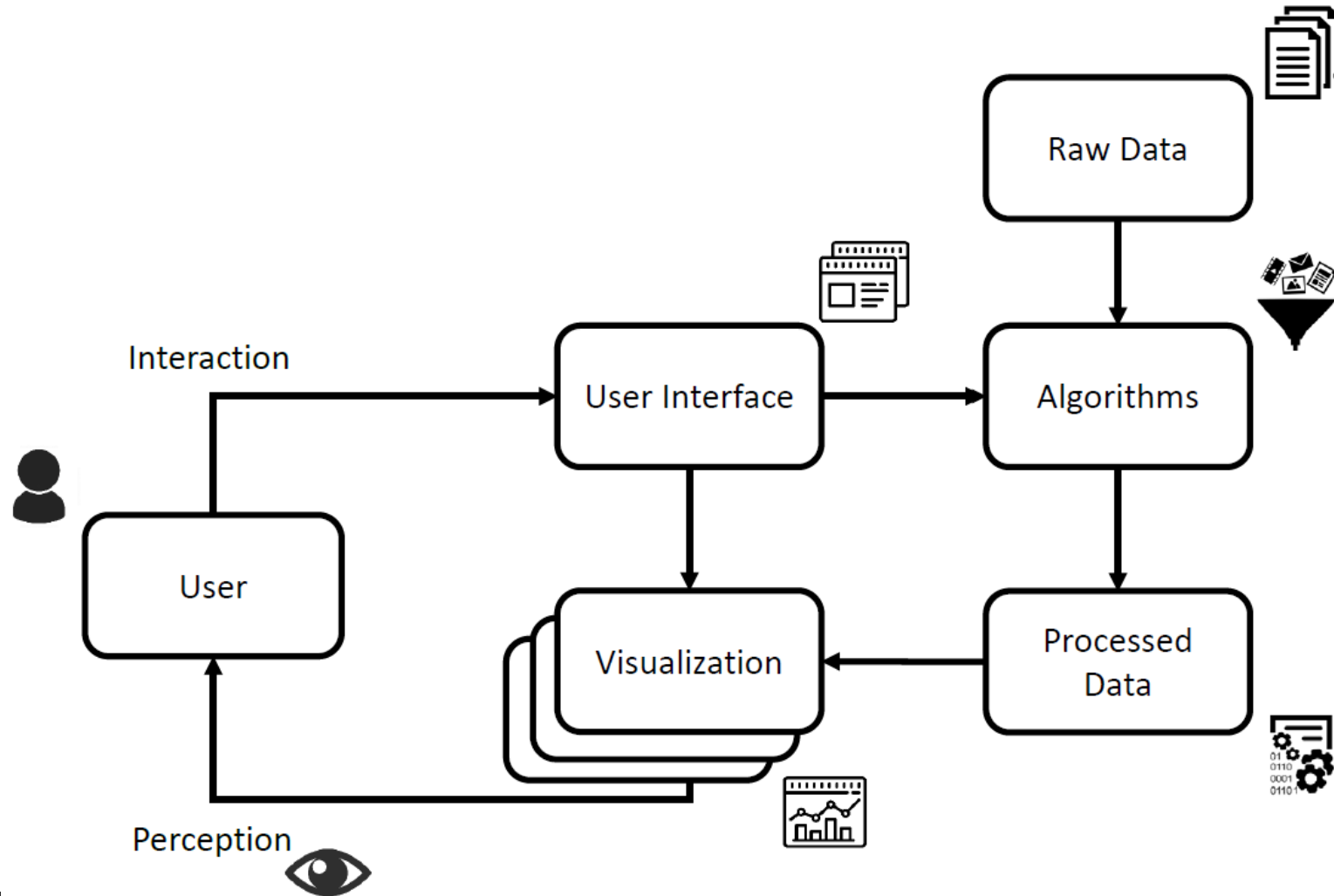
Data is generated everywhere and everyday

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

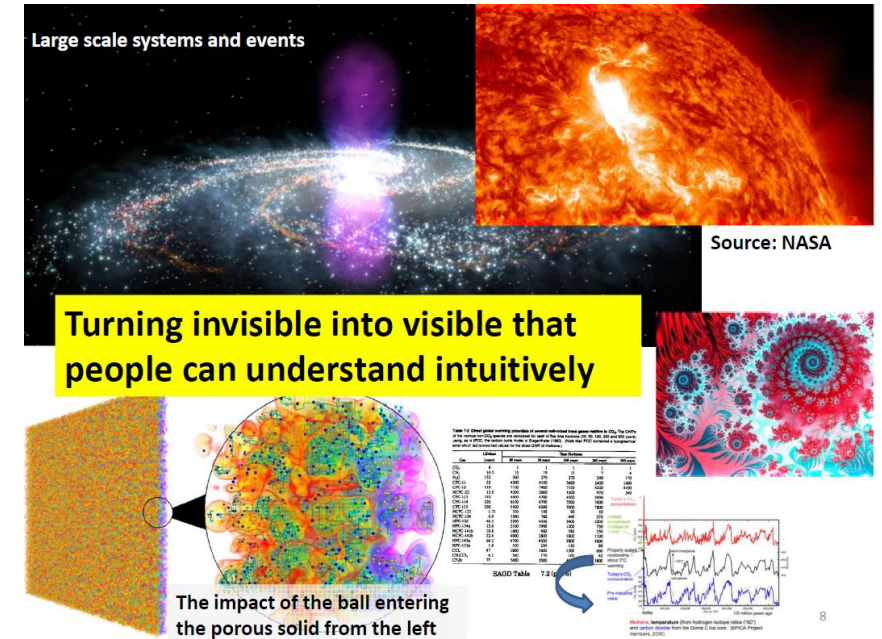


Why Data Visualization is needed?

- **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

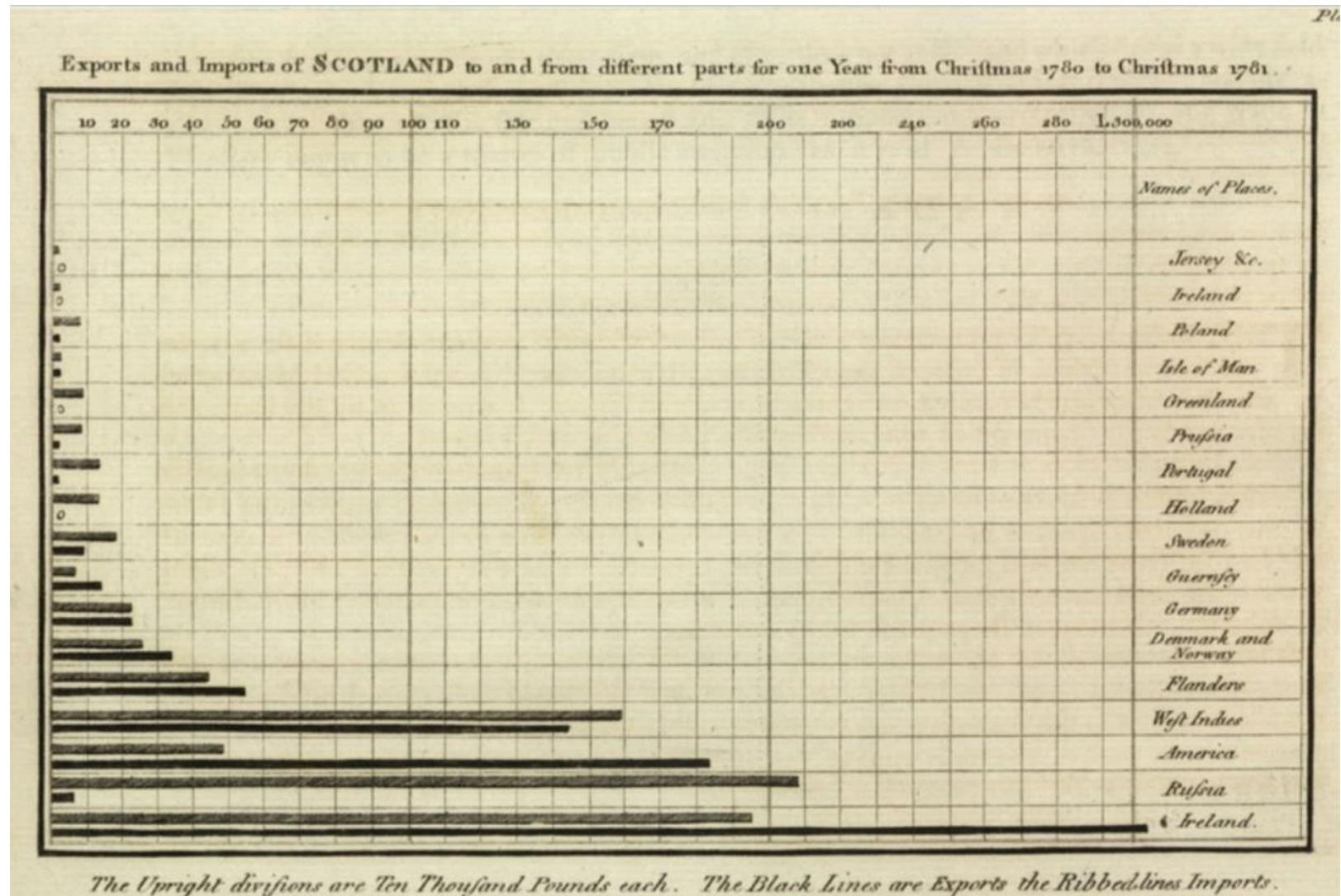
In class Discussion

- 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/>

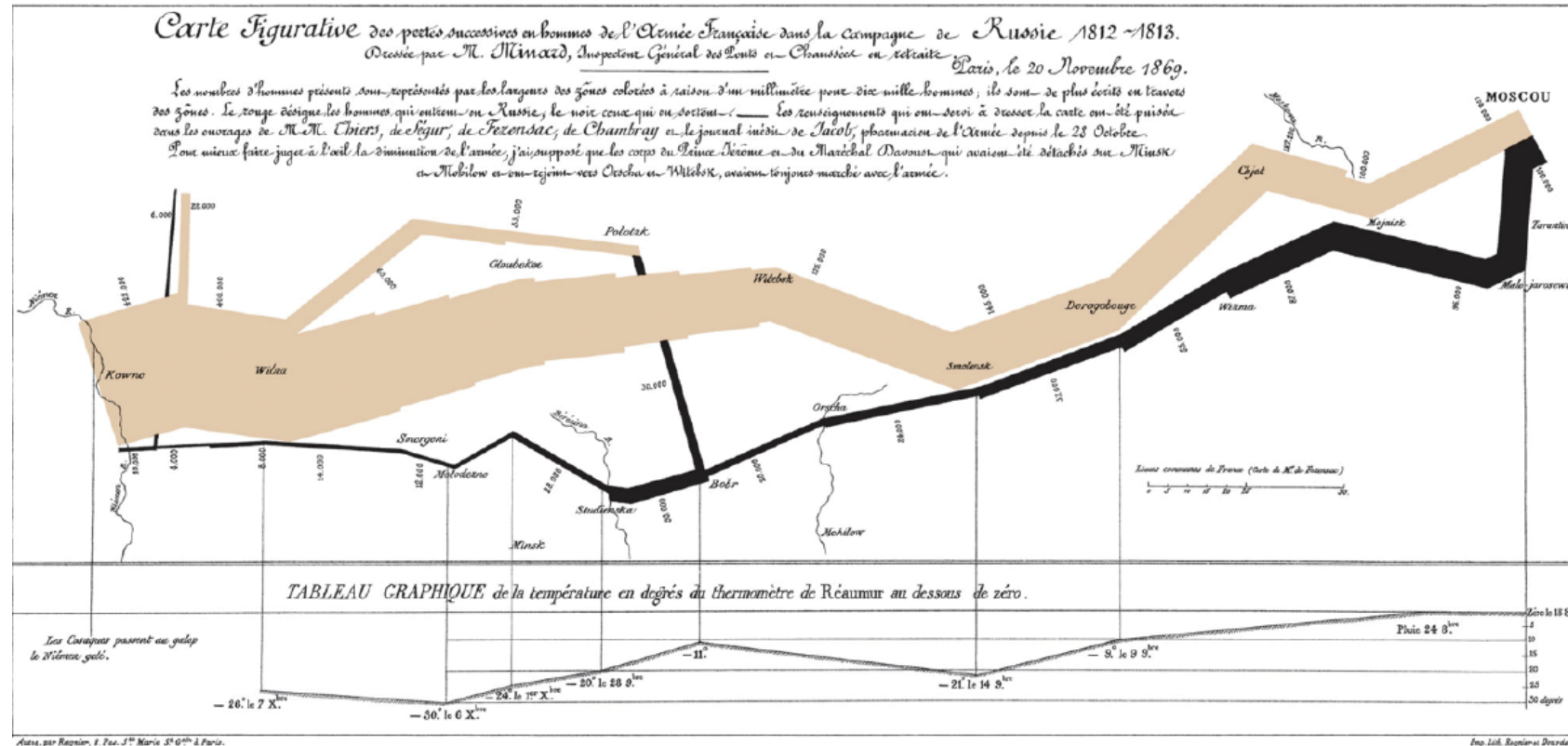
History of Data Visualization

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**

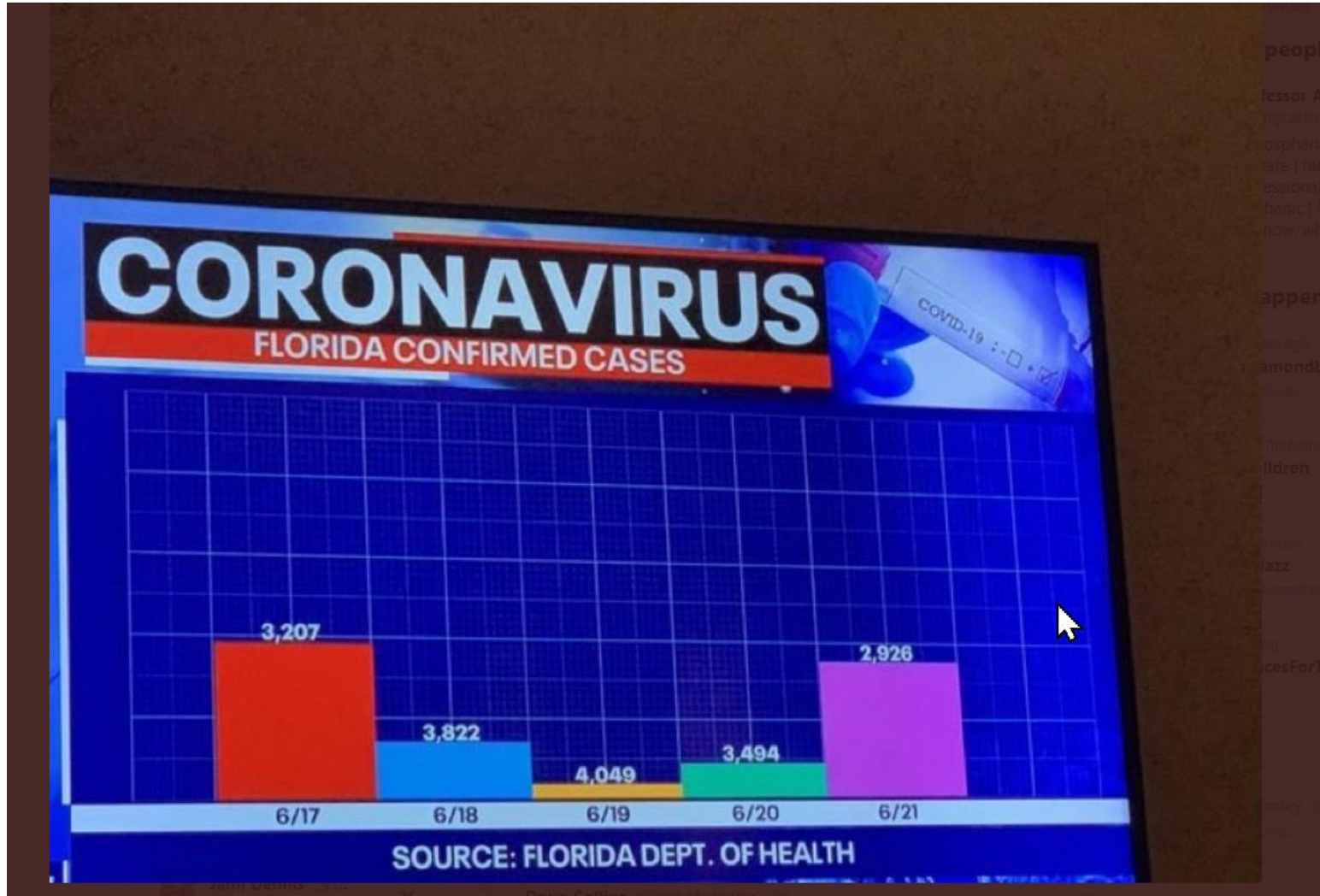
In Class Examples - Modern

- 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/>

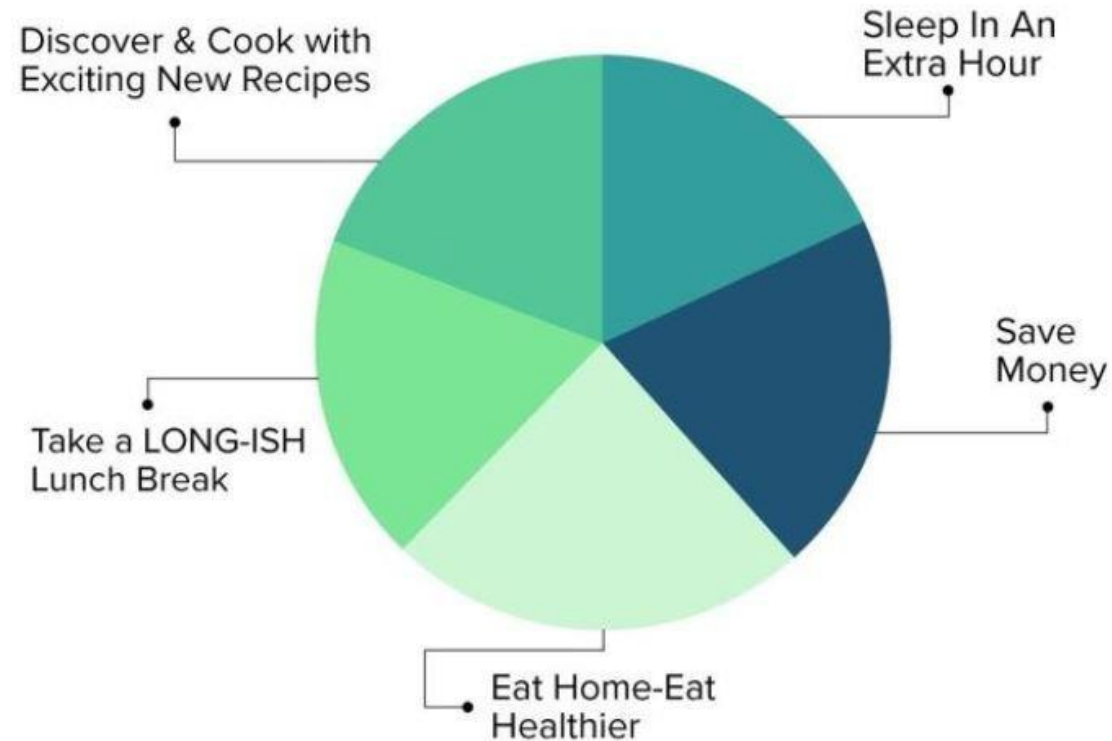
Example 2



<https://viz.wtf/>

Example 3

Advantages of Working From Home

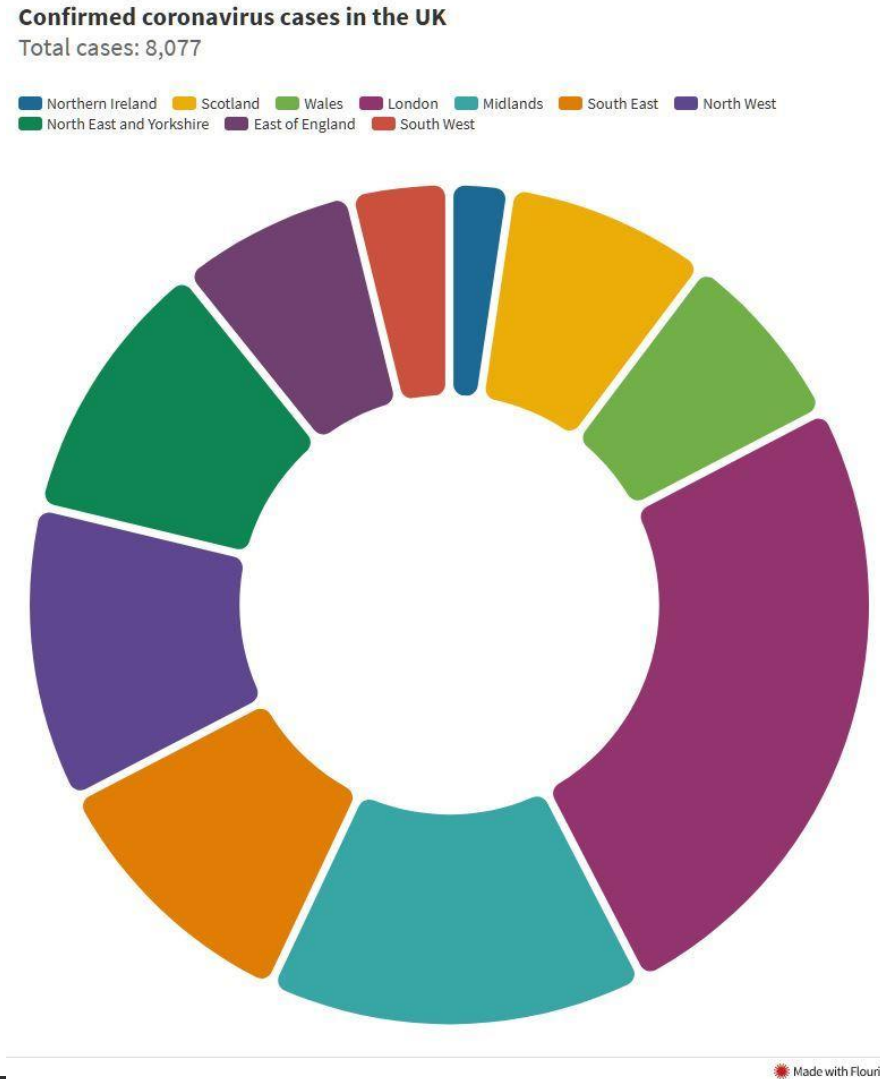


<https://viz.wtf/>

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/wefeelfine>

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