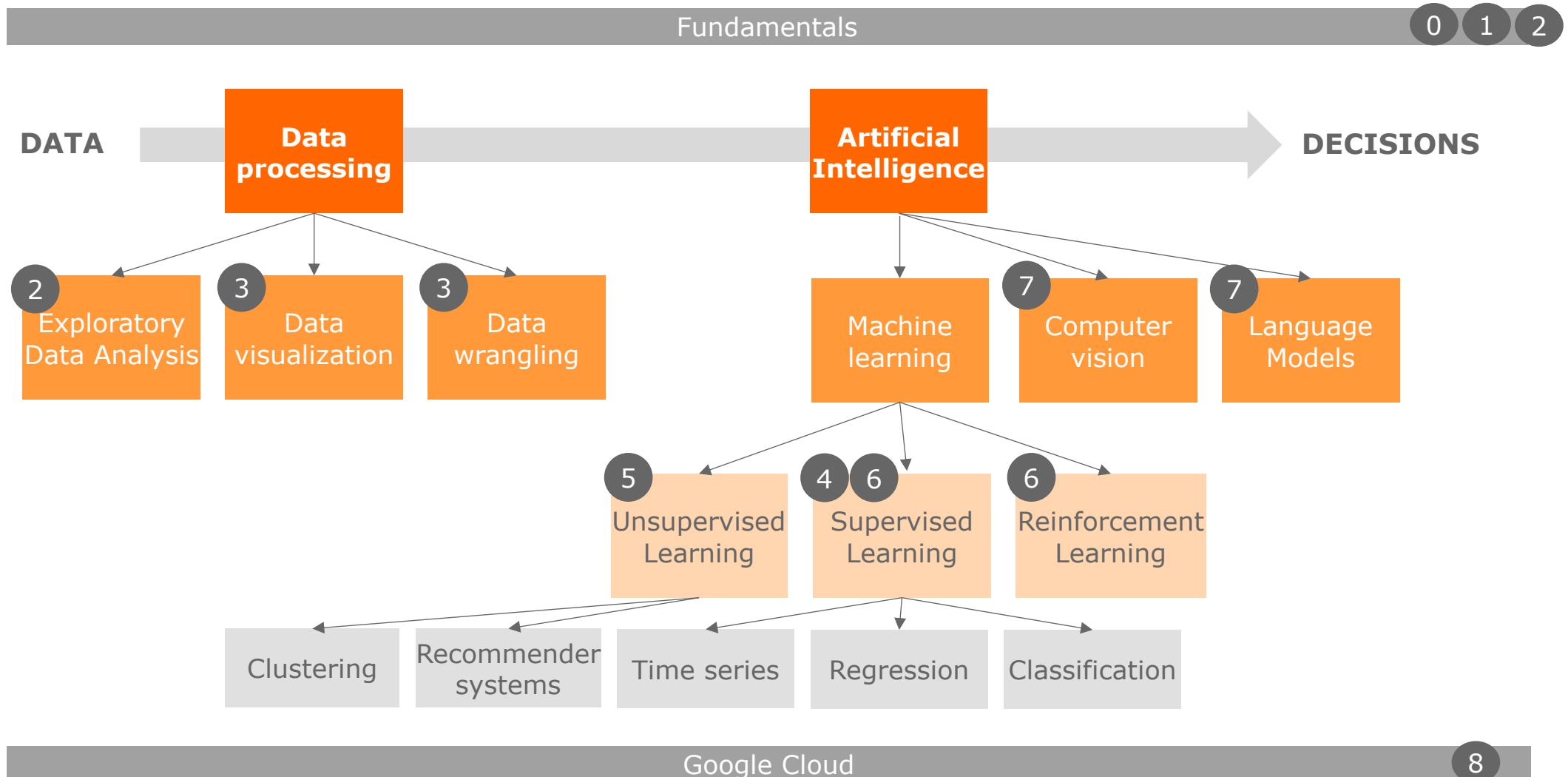
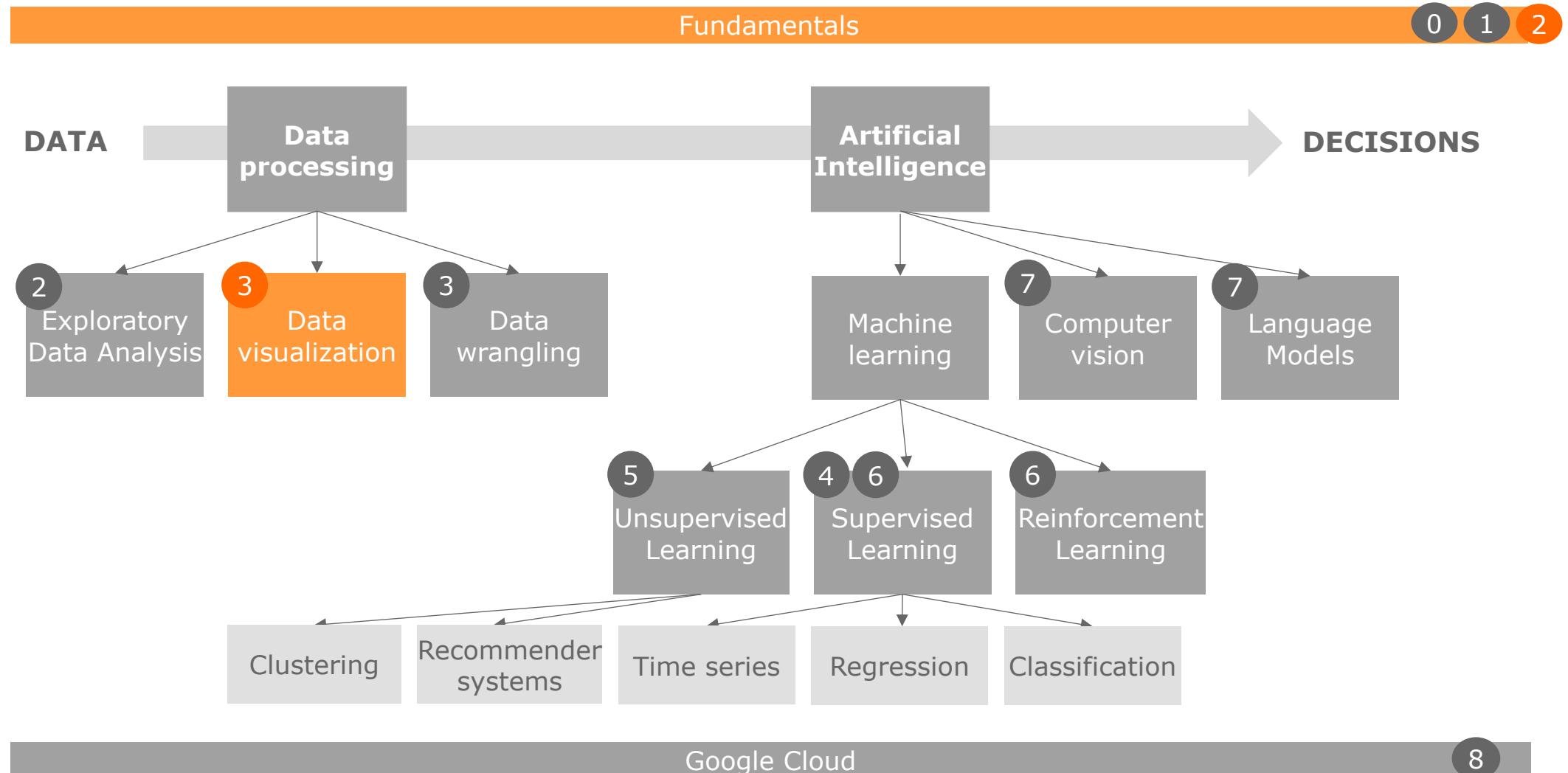


Data Visualization

Mário Amorim Lopes/Fábio Neves Moreira





Data Wrangling and Data Visualization

Data Engineering

- Different types of data (unstructured, semi-structured and structured data)
- Data importing
- Data cleaning
- Data transformation (date parsing, character encodings, etc.)

Data Visualization

- Visualizing data in Python using Seaborn
- Line charts
- Bar charts and heat maps
- Scatter plots
- Histograms and density plots

EDA in Python

- Building a data processing pipeline (import, transform, visualize)
- Descriptive statistics of a dataset
- Visual outlier detection and correction

Adv. Topics in Dataviz

- Examples of advanced data visualization
- Univariate visualization
- Multivariate visualization
- Whole dataset visualisations

#Name of the Module

Session 1

09h30	Introduction
10h00	Line charts
10h20	Bar charts
10h40	Heatmaps
11h00	Break
11h15	Scatter plots
11h45	Histograms
12h15	Density plots
12h45	Wrap-up

«Where is the wisdom we have lost in knowledge?

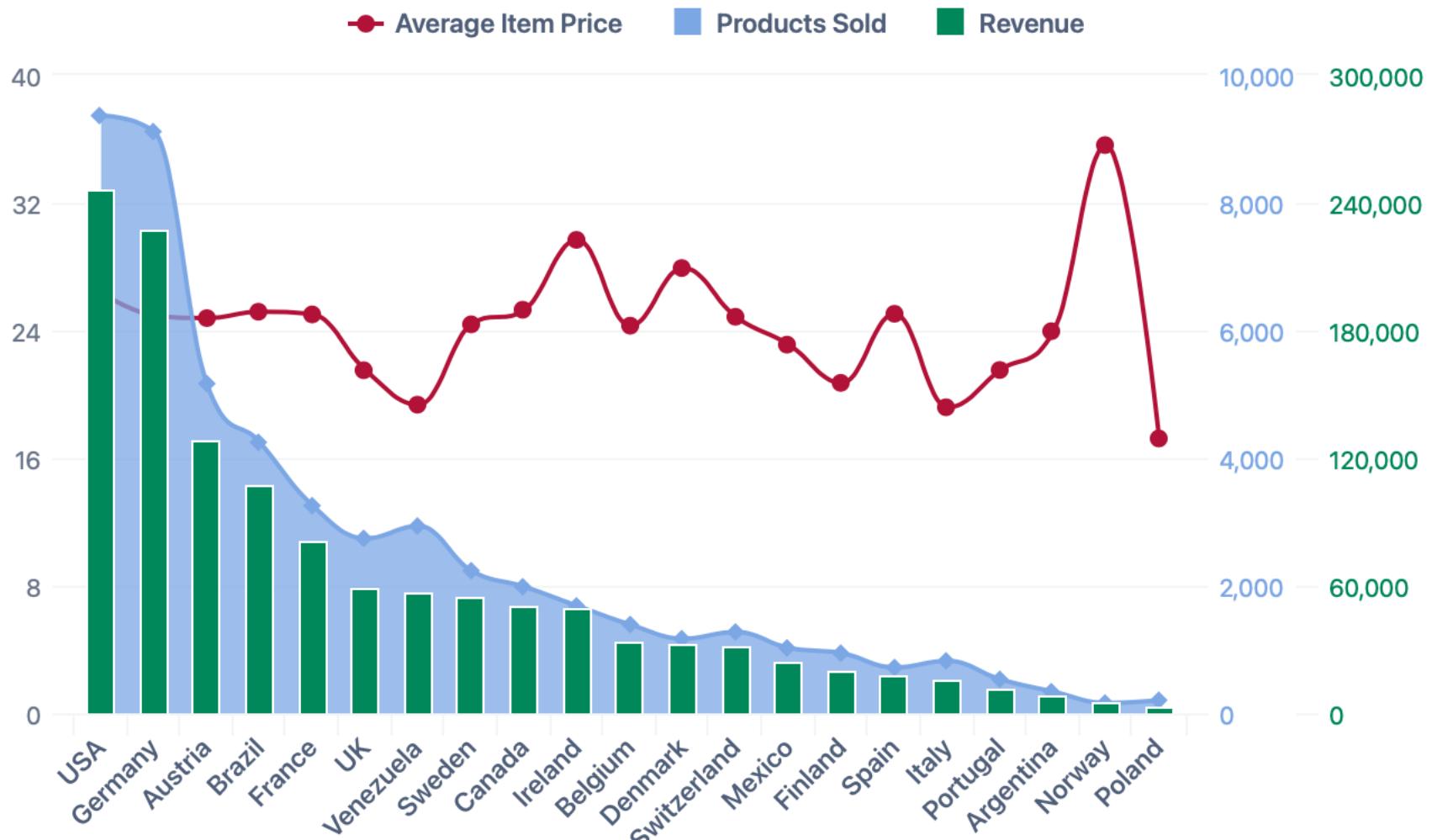
Where is the knowledge we have lost in information?»

— T. S. Eliot, the Roch (1934)

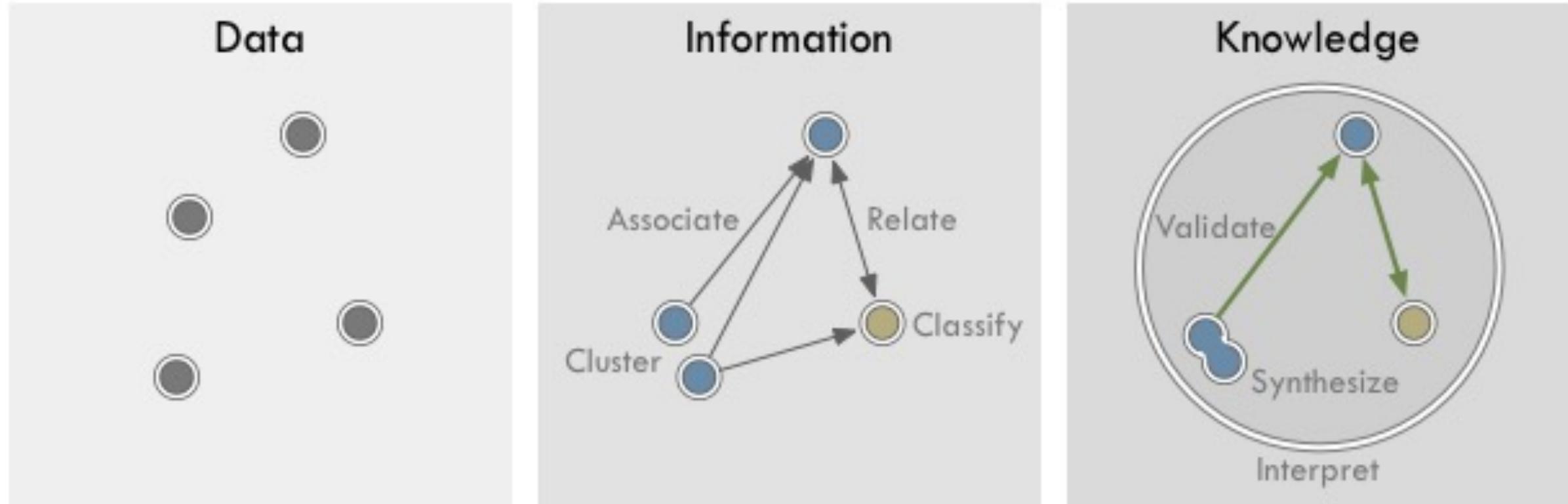
Plenty of data in this table – what about information and knowledge?

Segment	Country	Product	Discount Band	Units Sold	Manufacturing	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Number	Month Name	Year
Government	Canada	Carretera	None	1618,5	\$ 3,00	\$ 20,00	\$ 32 370,00	\$ -	\$ 32 370,00	\$ 16 185,00	\$ 16 185,00	01/01/2014	1	January	2014
Government	Germany	Carretera	None	1321	\$ 3,00	\$ 20,00	\$ 26 420,00	\$ -	\$ 26 420,00	\$ 13 210,00	\$ 13 210,00	01/01/2014	1	January	2014
Midmarket	France	Carretera	None	2178	\$ 3,00	\$ 15,00	\$ 32 670,00	\$ -	\$ 32 670,00	\$ 21 780,00	\$ 10 890,00	01/06/2014	6	June	2014
Midmarket	Germany	Carretera	None	888	\$ 3,00	\$ 15,00	\$ 13 320,00	\$ -	\$ 13 320,00	\$ 8 880,00	\$ 4 440,00	01/06/2014	6	June	2014
Midmarket	Mexico	Carretera	None	2470	\$ 3,00	\$ 15,00	\$ 37 050,00	\$ -	\$ 37 050,00	\$ 24 700,00	\$ 12 350,00	01/06/2014	6	June	2014
Government	Germany	Carretera	None	1513	\$ 3,00	\$ 350,00	\$ 529 550,00	\$ -	\$ 529 550,00	\$ 393 380,00	\$ 136 170,00	01/12/2014	12	December	2014
Midmarket	Germany	Montana	None	921	\$ 5,00	\$ 15,00	\$ 13 815,00	\$ -	\$ 13 815,00	\$ 9 210,00	\$ 4 605,00	01/03/2014	3	March	2014
Channel Partners	Canada	Montana	None	2518	\$ 5,00	\$ 12,00	\$ 30 216,00	\$ -	\$ 30 216,00	\$ 7 554,00	\$ 22 662,00	01/06/2014	6	June	2014
Government	France	Montana	None	1899	\$ 5,00	\$ 20,00	\$ 37 980,00	\$ -	\$ 37 980,00	\$ 18 990,00	\$ 18 990,00	01/06/2014	6	June	2014
Channel Partners	Germany	Montana	None	1545	\$ 5,00	\$ 12,00	\$ 18 540,00	\$ -	\$ 18 540,00	\$ 4 635,00	\$ 13 905,00	01/06/2014	6	June	2014
Midmarket	Mexico	Montana	None	2470	\$ 5,00	\$ 15,00	\$ 37 050,00	\$ -	\$ 37 050,00	\$ 24 700,00	\$ 12 350,00	01/06/2014	6	June	2014
Enterprise	Canada	Montana	None	2665,5	\$ 5,00	\$ 125,00	\$ 333 187,50	\$ -	\$ 333 187,50	\$ 319 860,00	\$ 13 327,50	01/07/2014	7	July	2014
Small Business	Mexico	Montana	None	958	\$ 5,00	\$ 300,00	\$ 287 400,00	\$ -	\$ 287 400,00	\$ 239 500,00	\$ 47 900,00	01/08/2014	8	August	2014
Government	Germany	Montana	None	2146	\$ 5,00	\$ 7,00	\$ 15 022,00	\$ -	\$ 15 022,00	\$ 10 730,00	\$ 4 292,00	01/09/2014	9	September	2014
Enterprise	Canada	Montana	None	345	\$ 5,00	\$ 125,00	\$ 43 125,00	\$ -	\$ 43 125,00	\$ 41 400,00	\$ 1 725,00	01/10/2013	10	October	2013
Midmarket	United States of America	Montana	None	615	\$ 5,00	\$ 15,00	\$ 9 225,00	\$ -	\$ 9 225,00	\$ 6 150,00	\$ 3 075,00	01/12/2014	12	December	2014
Government	Canada	Paseo	None	292	\$ 10,00	\$ 20,00	\$ 5 840,00	\$ -	\$ 5 840,00	\$ 2 920,00	\$ 2 920,00	01/02/2014	2	February	2014
Midmarket	Mexico	Paseo	None	974	\$ 10,00	\$ 15,00	\$ 14 610,00	\$ -	\$ 14 610,00	\$ 9 740,00	\$ 4 870,00	01/02/2014	2	February	2014
Channel Partners	Canada	Paseo	None	2518	\$ 10,00	\$ 12,00	\$ 30 216,00	\$ -	\$ 30 216,00	\$ 7 554,00	\$ 22 662,00	01/06/2014	6	June	2014
Government	Germany	Paseo	None	1006	\$ 10,00	\$ 350,00	\$ 352 100,00	\$ -	\$ 352 100,00	\$ 261 560,00	\$ 90 540,00	01/06/2014	6	June	2014
Channel Partners	Germany	Paseo	None	367	\$ 10,00	\$ 12,00	\$ 4 404,00	\$ -	\$ 4 404,00	\$ 1 101,00	\$ 3 303,00	01/07/2014	7	July	2014
Government	Mexico	Paseo	None	883	\$ 10,00	\$ 7,00	\$ 6 181,00	\$ -	\$ 6 181,00	\$ 4 415,00	\$ 1 766,00	01/08/2014	8	August	2014
Midmarket	France	Paseo	None	549	\$ 10,00	\$ 15,00	\$ 8 235,00	\$ -	\$ 8 235,00	\$ 5 490,00	\$ 2 745,00	01/09/2013	9	September	2013
Small Business	Mexico	Paseo	None	788	\$ 10,00	\$ 300,00	\$ 236 400,00	\$ -	\$ 236 400,00	\$ 197 000,00	\$ 39 400,00	01/09/2013	9	September	2013
Midmarket	Mexico	Paseo	None	2472	\$ 10,00	\$ 15,00	\$ 37 080,00	\$ -	\$ 37 080,00	\$ 24 720,00	\$ 12 360,00	01/09/2014	9	September	2014
Government	United States of America	Paseo	None	1143	\$ 10,00	\$ 7,00	\$ 8 001,00	\$ -	\$ 8 001,00	\$ 5 715,00	\$ 2 286,00	01/10/2014	10	October	2014
Government	Canada	Paseo	None	1725	\$ 10,00	\$ 350,00	\$ 603 750,00	\$ -	\$ 603 750,00	\$ 448 500,00	\$ 155 250,00	01/11/2013	11	November	2013
Channel Partners	United States of America	Paseo	None	912	\$ 10,00	\$ 12,00	\$ 10 944,00	\$ -	\$ 10 944,00	\$ 2 736,00	\$ 8 208,00	01/11/2013	11	November	2013
Midmarket	Canada	Paseo	None	2152	\$ 10,00	\$ 15,00	\$ 32 280,00	\$ -	\$ 32 280,00	\$ 21 520,00	\$ 10 760,00	01/12/2013	12	December	2013
Government	Canada	Paseo	None	1817	\$ 10,00	\$ 20,00	\$ 36 340,00	\$ -	\$ 36 340,00	\$ 18 170,00	\$ 18 170,00	01/12/2014	12	December	2014
Government	Germany	Paseo	None	1513	\$ 10,00	\$ 350,00	\$ 529 550,00	\$ -	\$ 529 550,00	\$ 393 380,00	\$ 136 170,00	01/12/2014	12	December	2014
Government	Mexico	Velo	None	1493	\$ 120,00	\$ 7,00	\$ 10 451,00	\$ -	\$ 10 451,00	\$ 7 465,00	\$ 2 986,00	01/01/2014	1	January	2014
Enterprise	France	Velo	None	1804	\$ 120,00	\$ 125,00	\$ 225 500,00	\$ -	\$ 225 500,00	\$ 216 480,00	\$ 9 020,00	01/02/2014	2	February	2014
Channel Partners	Germany	Velo	None	2161	\$ 120,00	\$ 12,00	\$ 25 932,00	\$ -	\$ 25 932,00	\$ 6 483,00	\$ 19 449,00	01/03/2014	3	March	2014
Government	Germany	Velo	None	1006	\$ 120,00	\$ 350,00	\$ 352 100,00	\$ -	\$ 352 100,00	\$ 261 560,00	\$ 90 540,00	01/06/2014	6	June	2014
Channel Partners	Germany	Velo	None	1545	\$ 120,00	\$ 12,00	\$ 18 540,00	\$ -	\$ 18 540,00	\$ 4 635,00	\$ 13 905,00	01/06/2014	6	June	2014
Enterprise	United States of America	Velo	None	2821	\$ 120,00	\$ 125,00	\$ 352 625,00	\$ -	\$ 352 625,00	\$ 338 520,00	\$ 14 105,00	01/08/2014	8	August	2014
Enterprise	Canada	Velo	None	345	\$ 120,00	\$ 125,00	\$ 43 125,00	\$ -	\$ 43 125,00	\$ 41 400,00	\$ 1 725,00	01/10/2013	10	October	2013
Small Business	Canada	VTT	None	2001	\$ 250,00	\$ 300,00	\$ 600 300,00	\$ -	\$ 600 300,00	\$ 500 250,00	\$ 100 050,00	01/02/2014	2	February	2014
Channel Partners	Germany	VTT	None	2838	\$ 250,00	\$ 12,00	\$ 34 056,00	\$ -	\$ 34 056,00	\$ 8 514,00	\$ 25 542,00	01/04/2014	4	April	2014
Midmarket	France	VTT	None	2178	\$ 250,00	\$ 15,00	\$ 32 670,00	\$ -	\$ 32 670,00	\$ 21 780,00	\$ 10 890,00	01/06/2014	6	June	2014
Midmarket	Germany	VTT	None	888	\$ 250,00	\$ 15,00	\$ 13 320,00	\$ -	\$ 13 320,00	\$ 8 880,00	\$ 4 440,00	01/06/2014	6	June	2014
Government	France	VTT	None	1527	\$ 250,00	\$ 350,00	\$ 534 450,00	\$ -	\$ 534 450,00	\$ 397 020,00	\$ 137 430,00	01/09/2013	9	September	2013
Small Business	France	VTT	None	2151	\$ 250,00	\$ 300,00	\$ 645 300,00	\$ -	\$ 645 300,00	\$ 537 750,00	\$ 107 550,00	01/09/2014	9	September	2014
Government	Canada	VTT	None	1817	\$ 250,00	\$ 20,00	\$ 36 340,00	\$ -	\$ 36 340,00	\$ 18 170,00	\$ 18 170,00	01/12/2014	12	December	2014
Government	France	Amarilla	None	2750	\$ 260,00	\$ 350,00	\$ 962 500,00	\$ -	\$ 962 500,00	\$ 715 000,00	\$ 247 500,00	01/02/2014	2	February	2014
Channel Partners	United States of America	Amarilla	None	1953	\$ 260,00	\$ 12,00	\$ 23 436,00	\$ -	\$ 23 436,00	\$ 5 859,00	\$ 17 577,00	01/04/2014	4	April	2014
Enterprise	Germany	Amarilla	None	4219,5	\$ 260,00	\$ 125,00	\$ 527 437,50	\$ -	\$ 527 437,50	\$ 506 340,00	\$ 21 097,50	01/04/2014	4	April	2014
Government	France	Amarilla	None	1899	\$ 260,00	\$ 20,00	\$ 37 980,00	\$ -	\$ 37 980,00	\$ 18 990,00	\$ 18 990,00	01/06/2014	6	June	2014
Government	Germany	Amarilla	None	1686	\$ 260,00	\$ 7,00	\$ 11 802,00	\$ -	\$ 11 802,00	\$ 8 430,00	\$ 3 372,00	01/07/2014	7	July	2014
Channel Partners	United States of America	Amarilla	None	2141	\$ 260,00	\$ 12,00	\$ 25 692,00	\$ -	\$ 25 692,00	\$ 6 423,00	\$ 19 269,00	01/08/2014	8	August	2014
Government	United States of America	Amarilla	None	1143	\$ 260,00	\$ 7,00	\$ 8 001,00	\$ -	\$ 8 001,00	\$ 5 715,00	\$ 2 286,00	01/10/2014	10	October	2014
Midmarket	United States of America	Amarilla	None	615	\$ 260,00	\$ 15,00	\$ 9 225,00	\$ -	\$ 9 225,00	\$ 6 150,00	\$ 3 075,00	01/12/2014	12	December	2014
Government	France	Paseo	Low	3945	\$ 10,00	\$ 7,00	\$ 27 615,00	\$ 276,15	\$ 27 338,85	\$ 19 725,00	\$ 7 613,85	01/01/2014	1	January	2014
Midmarket	France	Paseo	Low	2296	\$ 10,00	\$ 15,00	\$ 34 440,00	\$ 344,40	\$ 34 095,60	\$ 22 960,00	\$ 11 135,60	01/02/2014	2	February	2014

Important information (knowledge?) hidden in that Excel sheet



From data to knowledge



Some examples of information and knowledge better portrayed through charts

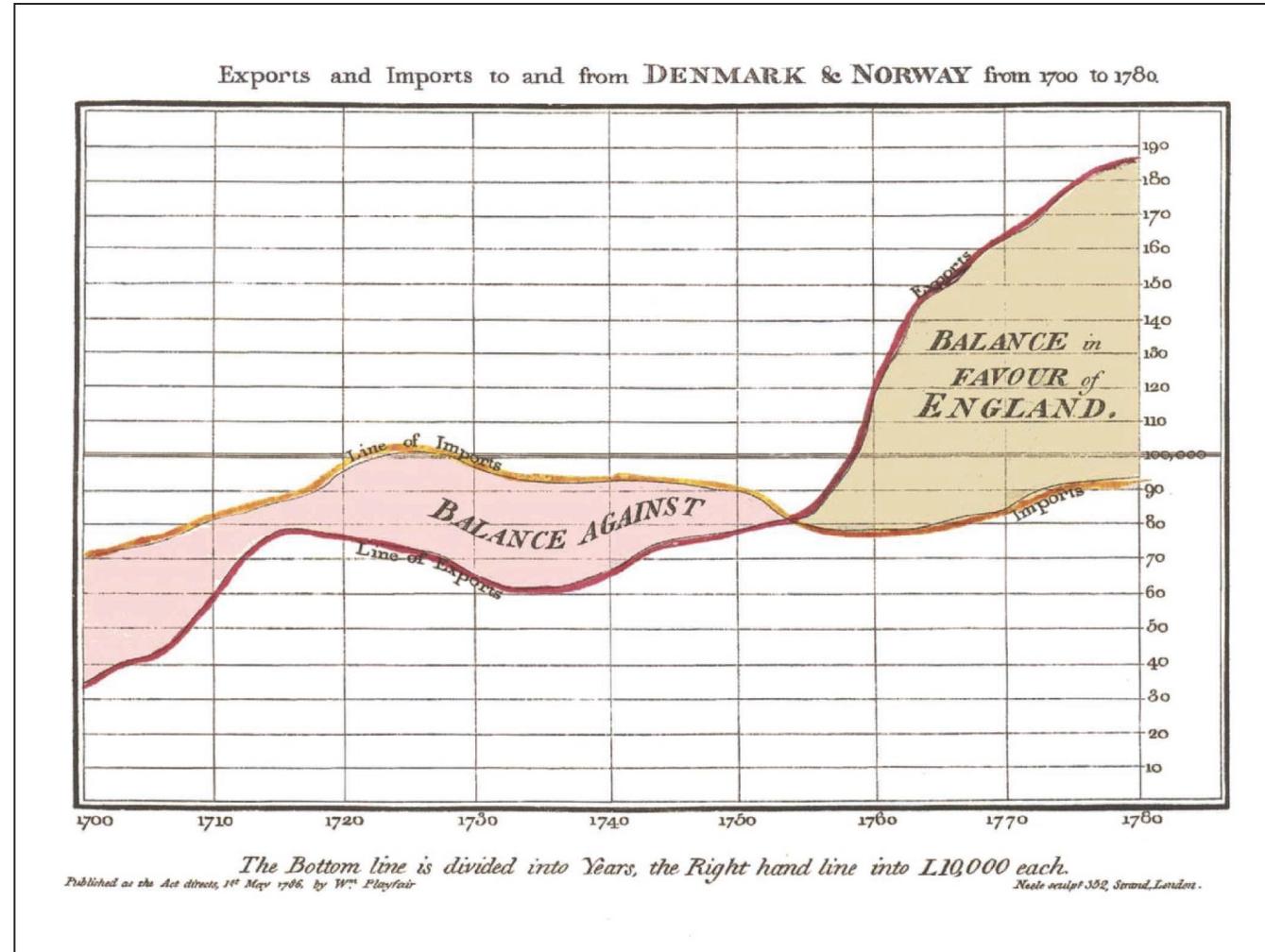


FIGURE 2.19 The First Time Series Line Chart Created by William Playfair in 1801.

How did it go for Napoleonic in its way to Moscow?

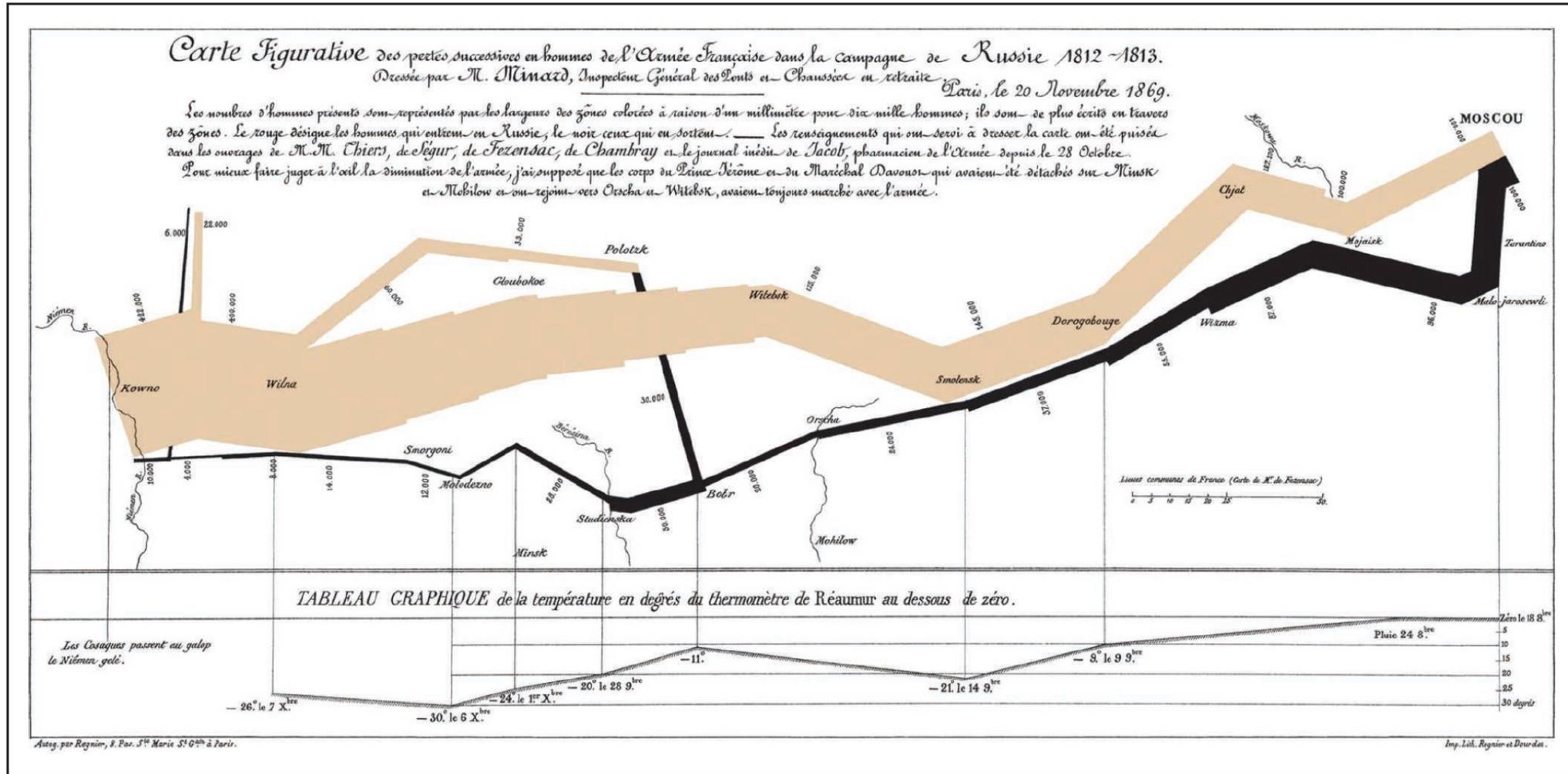
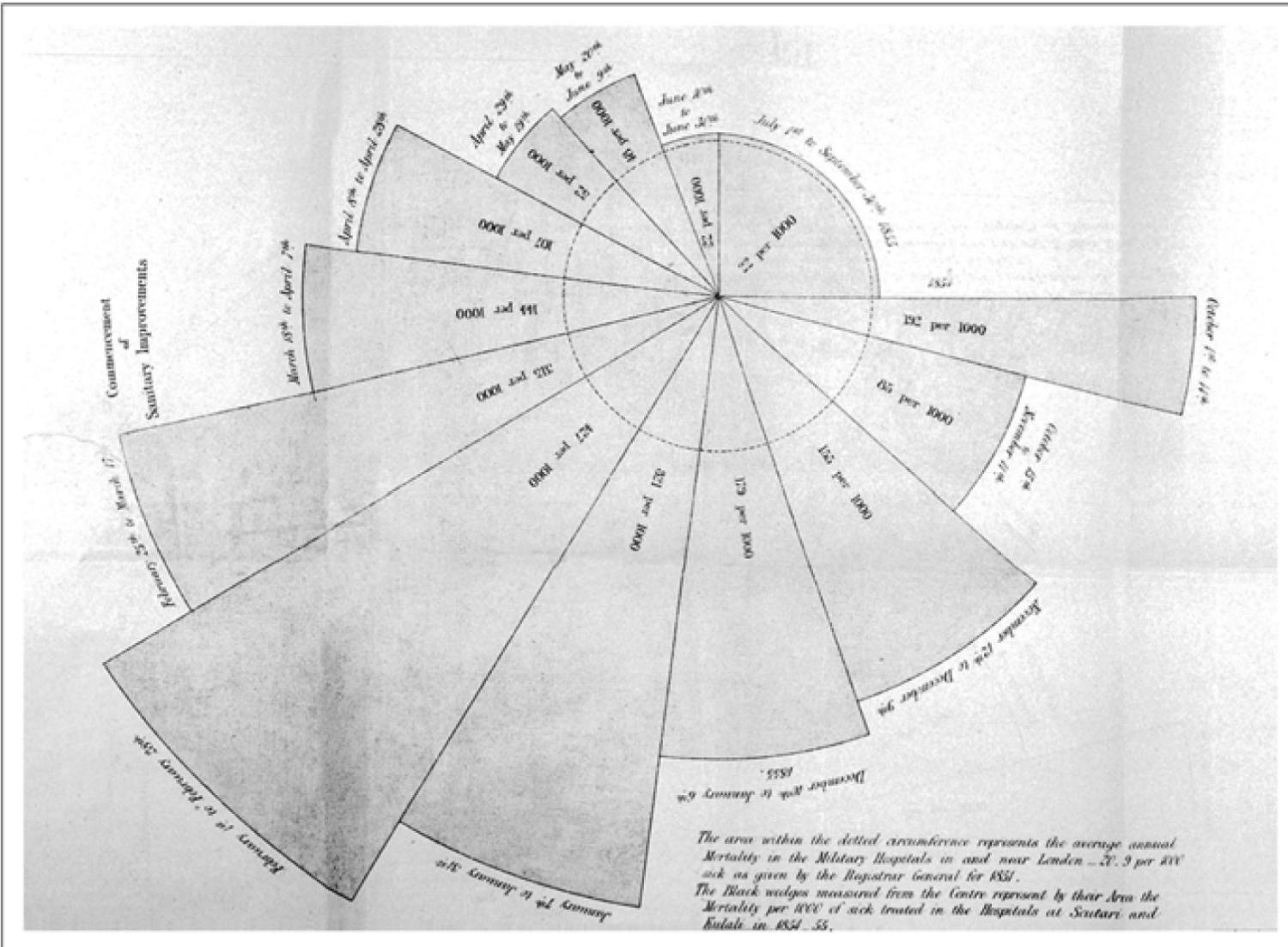


FIGURE 2.20 Decimation of Napoleon's Army during the 1812 Russian Campaign.

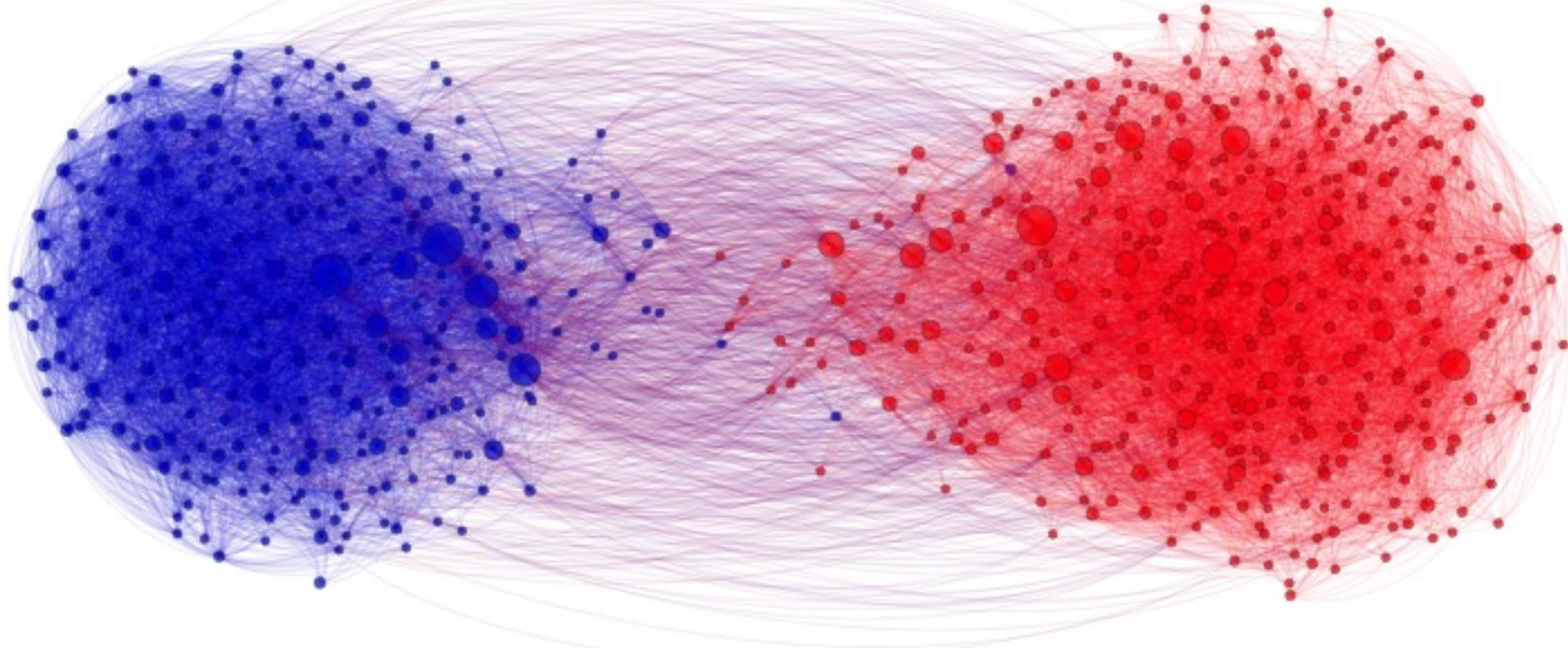
Florence Nightingale was the first to introduce analytics in a hospital!



Cholera outbreak – Dr John Snow saw the information and concluded: the problem was in the water pumps.



Polarized social networks: red (Republicans) and blue (Democrats)

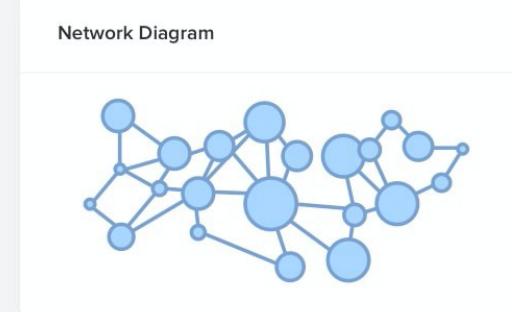
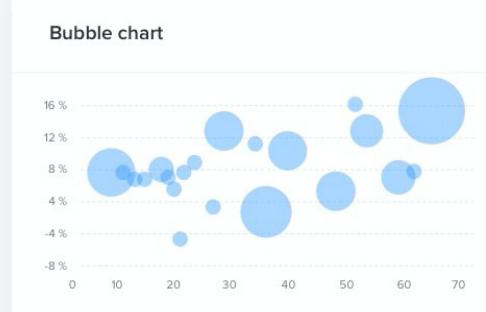
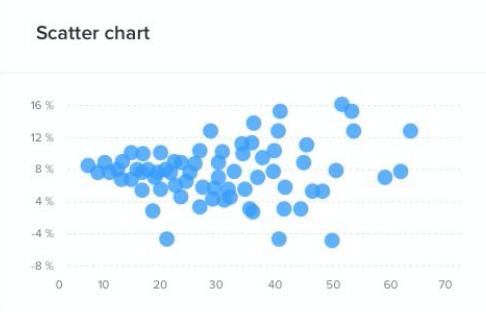


Some dataviz tools to get you started!

What if we want to understand if there's a **relationship** between variables? Use these charts.

See relationship

Static

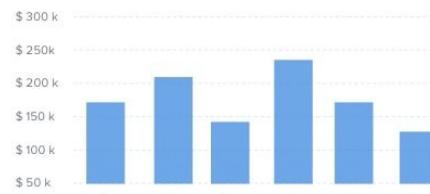


Some dataviz tools to get you started!

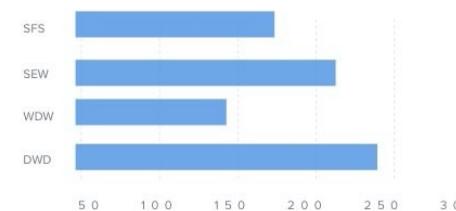
Use these charts if you want to **compare** static or dynamic (time-dependent) data

Static

Column chart

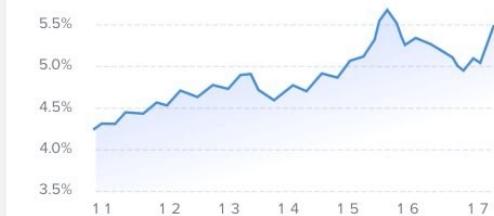


Bar chart



Over Time

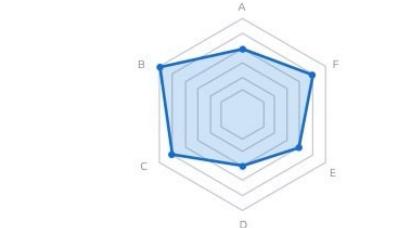
Line chart



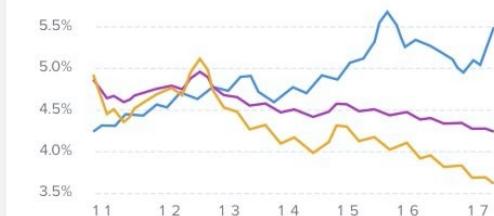
Column overlap chart



Circular area chart

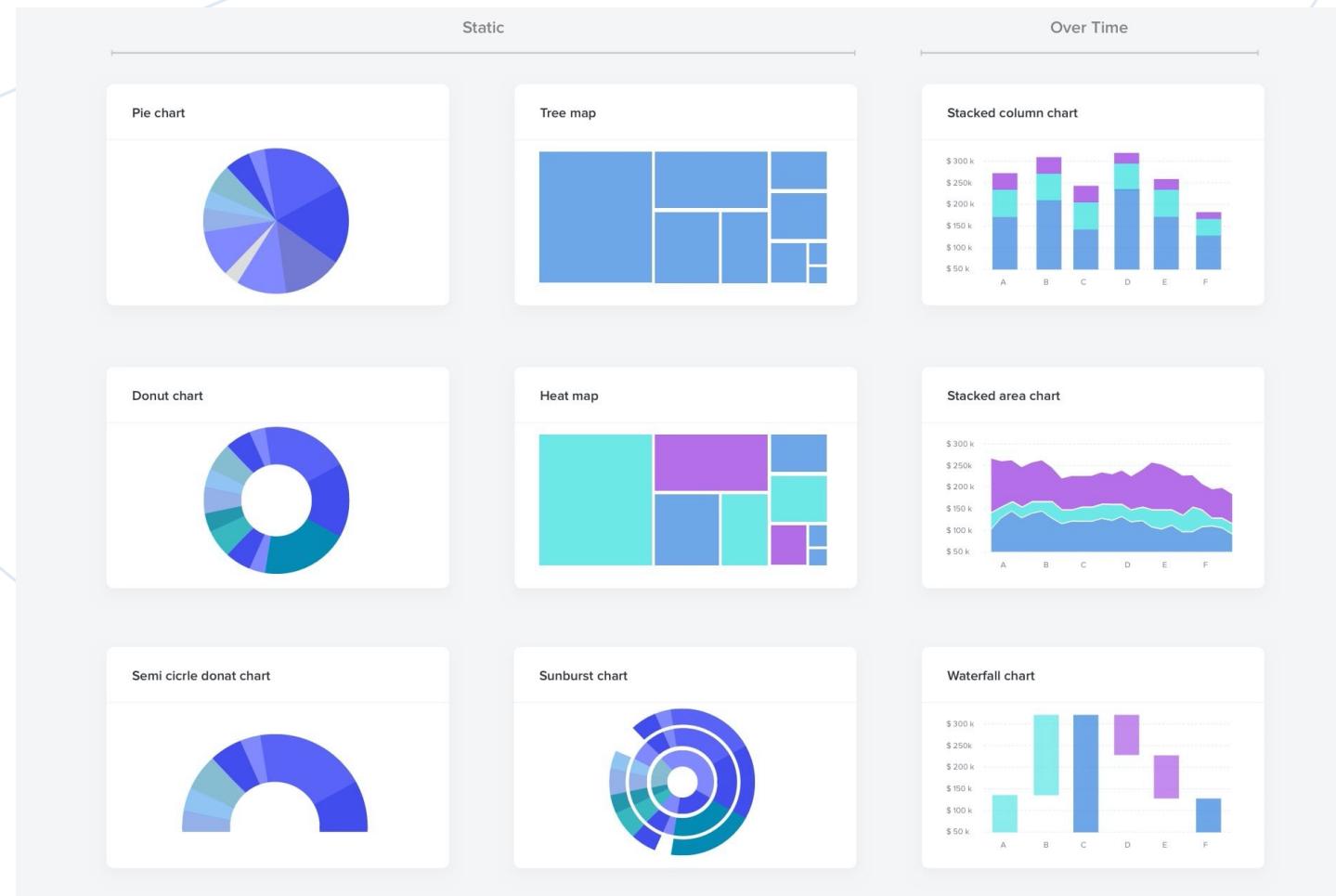


Line chart (several items)



Some dataviz tools to get you started!

Use these charts if you want to **disaggregate** data (i.e., view partial sales by category, location, etc.)

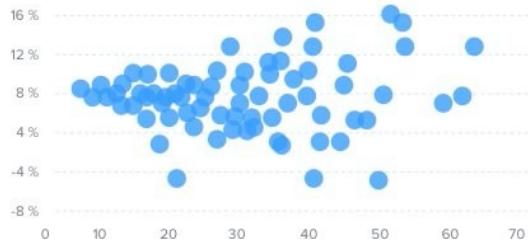


Some dataviz tools to get you started!

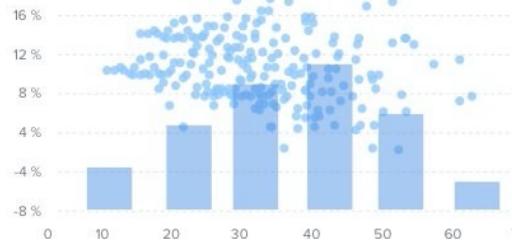
Use these charts if you want to see the **distribution** of your data.

Static

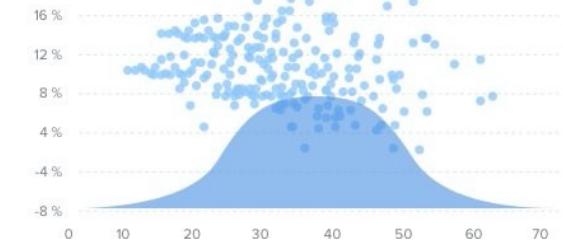
Scatter chart



Histogram chart



Bell curve



Cheat sheet

Relationship

Static



Scatter chart



Bubble chart



Network Diagram

Comparison

Static



Column chart



Bar chart



Column overlap chart



Circular area chart

Composition

Static



Pie/Donut chart



Tree map



Heat map



Sunburst chart

Distribution

Static



Scatter chart



Histogram chart



Bell curve

Over Time



Line chart

Over Time



Stacked column chart



Stacked area chart



Waterfall chart

What to avoid

Avoid these charts, as they are **difficult to interpret** and sometimes even **ambiguous**.

Don't

Gauges



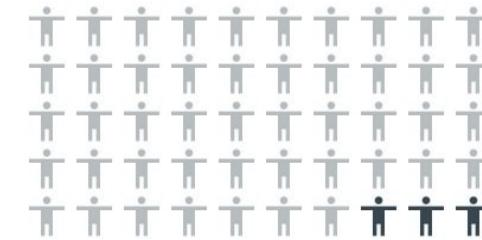
Don't

3D charts



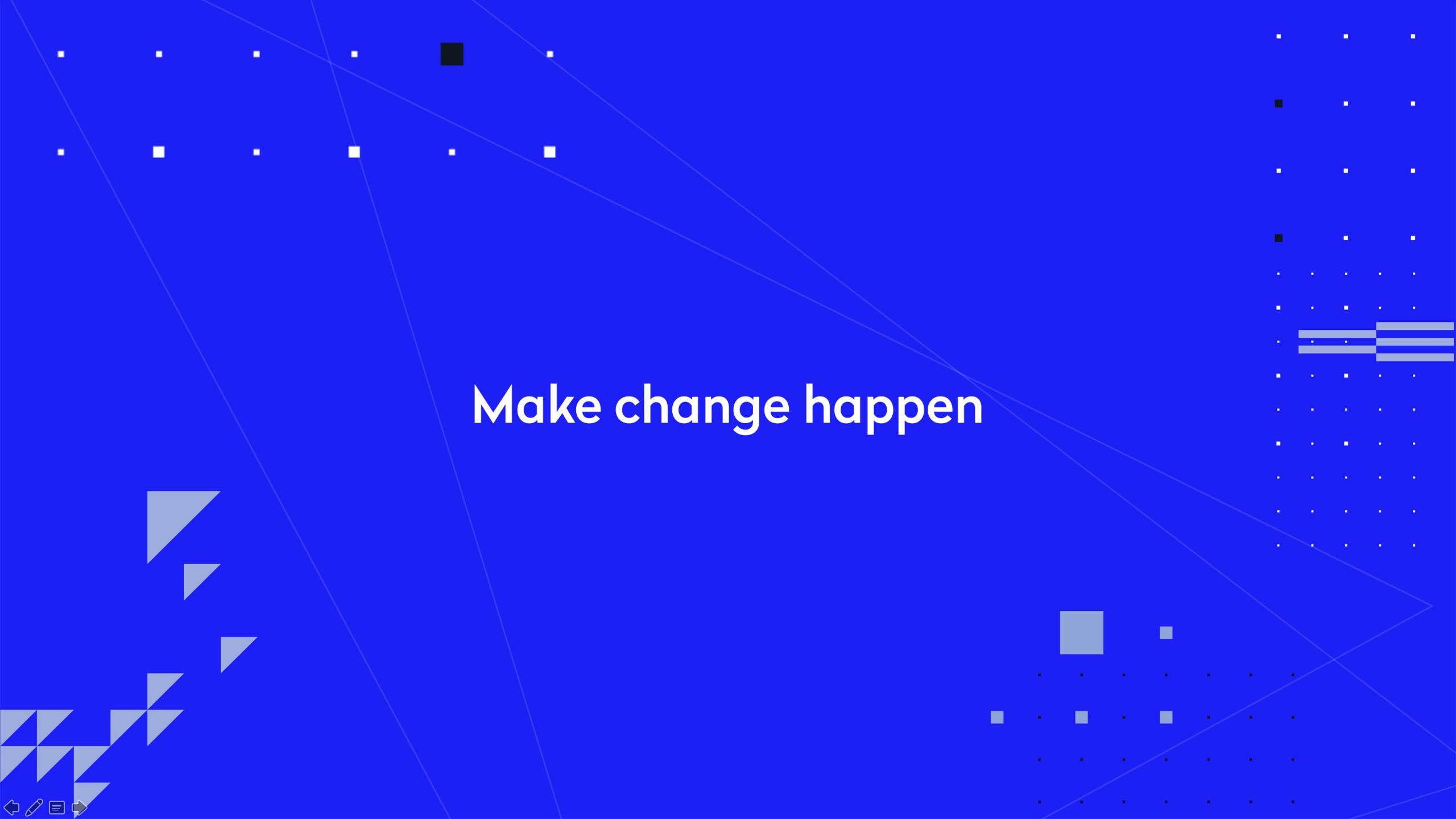
Don't

?????????



Wrap-up

- ✓ **Visualization can help uncover non-obvious relations** — looking at data in tabular form is a very limited (and laborious) way of gaining knowledge
- ✓ **You need to use the right chart with the right type of data** — anything goes does not apply; use the appropriate tool for the job
- ✓ **Building charts is like an art** — there are clear guidelines for what to do, but making the difference requires a personal touch



Make change happen