Visualização de Dados

Representações visuais de dados

U.PORTO FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO [dcc] DEPARTAMENTO DE CIÊNCIA DE COMPUTADORES FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO

1

Topic 1

Milestones in Data Visualization

2

Some Milestones in Information Visualization

- **Raw data and tabular presentation** have been present since the **2nd century**, when it was used in Egypt to **organize astronomical information, and to aid navigation**.
- The representation of quantitative data in **2D graphs** did not arise until the **17th century**, when Rene Descartes invented the method, not originally for presenting data but by linking Euclidean geometry with algebra by inventing the Cartesian coordinate system.
- It was by the **late 18th and early 19th centuries** that arrived the ***golden age of graphics*** mainly through the advances of William Playfair by inventing or improving many of the graphs that we use today, such as the bar graph or the pie chart.
- The dawn of the **21st** century witnessed an unprecedented surge in the realm of visualization information, largely attributable to the **advancements in computing power**. This rapid progress has not only transformed the way data is handled, but also introduced a whole new epoch – the era of ***visual analytics***.

Examples

Visualization in scientific discovery (past)

Dr. John Snow

Discovering the cause of the London cholera out brake, 1853-54 (Wikipedia)

The **Minard diagram** shows the losses suffered by Napoleon's army in the 1812–1813 period. Six variables are plotted: the size of the army, its location on a two-dimensional surface, time, direction of movement, and temperature. The line width illustrates a comparison (size of the army at points in time) while the temperature axis suggests a cause of the change in army size. This multivariate display on a two-dimensional surface tells a story that can be grasped immediately while identifying the source data to build credibility.

[Tufte wrote in 1983 that: "It may well be the best statistical graphic ever drawn."]

Visualization in scientific discovery (present)

(e.g., surgery training)

- Temporal bone surgery
- Movement of the drill is controlled with a force feedback device

https://www.voxel-man.com/simulators/tempo/

https://www.youtube.com/watch?v=CUOm6fxCJqI

Example of fluid mechanics visualization

https://www.youtube.com/watch?v=3NsKpHftx7c

Graphical displays should

- **Show the data**
- **Induce the viewer to think about the substance** and not the ~~methodology~~ used, the ~~graphic design~~ employed, or even the ~~technology~~ for the graphic production
- Present many numbers in a **small space**
- **Avoid distorting** what the data has to "say". Do not introduce biases or misinterpretations
- **Make large data sets coherent**
- **Encourage the eye to compare different pieces** of data: identify pattern, trends, differences
- Reveal the data at **several levels of detail**, from a broad overview to the in-depth exploration of fine structures
- **Serve a clear purpose**: description, or exploration, or verification
- Be closely **integrated with the statistical and verbal descriptions** of a data set.

9

Topic 2

Taxonomies for Data Visualization

10

Taxonomies for visual representation of data

Christian Behrens, **2008**

11

Designed by Adedamola Ladipo

12

Topic 3

Examples of common visual representations of data

13

Proportions

Pie Chart

Ring Chart

When to avoid this type of representation?

14

Correlations: Scatter plot

What does the x-axis represent?

15

Correlations: Scatter and bubble chart

At least, how many properties must have each record in this chart?

16

Visualization explorer for every major film 2008-2016

17

Box Plot

Can you explain a box plot?

18

Waterfall graphs

A waterfall chart helps in understanding the **cumulative effect on the initial value** which is increased or decreased by a series of intermediate values leading to a final value

19

Continuous quantities and Sparklines

Tufte's (Tufte, 2006) own Sparklines are **data-intense, design-simple, word-sized graphics**. Sparklines, can **display the temporal evolution of variables**, its most recent value, its name and the out-of-the-norm values, **everything in a highly condensed graphic**.

Obs: Sparklines can run in a text layout.

	Region	magnitude_trend	count	avg(Magnitude)
1	Fox Islands, Aleutian Islands, Alaska		14	3.271429
2	Island of Hawaii, Hawaii		14	3.035714
3	Puerto Rico region		14	3.035714
4	Southern Alaska	3	10	2.880000
5	Andreanof Islands, Aleutian Islands, Alaska		8	2.712500
6	Central California		8	2.925000
7	Baja California, Mexico		7	2.957143
8	Virgin Islands region		7	3.185714
9	Kodiak Island region, Alaska		6	2.733333
10	Central Alaska		5	2.920000

Note: the aspect ratio of line chart or a sparkline are crucial for good reading. The visual average of the hill-slopes within the line should be ideally 45 degrees.

20

21

Multidimensions and Parallel Coordinates

Parallel coordinates can be arranged in order to show certain correlations among attributes for the same type of object

22

Hierarchical structures: Trees

23

Relational structures with linear layout

24

Relational structures: networks

What do they represente?
- Nodes?
- Edges?
- Size of nodes?
- Thickness of edges?

Human disease network ©

Álvaro Figueira • VD • 2023 • 1ª ed.

25

Relational structures: Sankey diagrams

Álvaro Figueira • VD • 2023 • 1ª ed.

26

27

Relational structures: Sankey diagrams

Figure 9. Thematic evolution based on authors' keywords

28

Time Series as Stacked Graphs (History of Olympics)

29

Filled regions

30

Hierarchies: Treemaps

31

Hierarchies: circular treemaps

32

Heatmaps

Value (%)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	-9.8%	84.5%	-79.5%	1.8%	23.4%	54.7%	48.2%	75.4%	-61.4%	-76.7%	-35.8%	
2010	-66.5%	41.2%	38.9%	-41.1%	20.5%	55.7%	-97.9%	77.8%	80.3%	-88.7%	-66.0%	10.6%
	87.6%	-12.5%	86.7%	-21.6%	56.8%	2.0%	57.5%	10.1%	-86.2%	8.6%	66.4%	82.2%
	95.8%	42.3%	64.8%	0.1%	50.8%	-70.0%	-6.7%	32.9%	47.0%	-16.9%	40.7%	-13.3%
	-35.0%	-8.2%	36.9%	38.4%	-3.2%	-76.1%	61.2%	9.8%	10.4%	5.2%	93.2%	92.8%
	-49.2%	24.5%	-46.3%	54.1%	-18.9%	-38.9%	3.1%	-59.4%	15.2%	67.5%	-73.7%	20.8%
2005	52.1%	-87.2%	76.2%	-26.4%	-42.3%	65.3%	-9.0%	88.4%	60.7%	34.8%	-59.2%	37.0%
	89.1%	-17.7%	5.6%	81.4%	11.1%	81.7%	-90.3%	-58.4%	42.7%	-69.4%	-0.3%	18.8%
	-61.5%	42.3%	-70.0%	2.9%	17.2%	85.1%	-20.1%	-43.7%	52.5%	70.1%	-21.0%	55.4%
	55.0%	4.7%	-68.1%	76.2%	-29.1%	33.8%	-95.7%	41.9%	69.8%	5.4%	-16.2%	43.7%
	8.0%	27.5%	17.3%	39.0%	26.1%	-14.5%	-56.4%	66.5%	66.5%	67.2%	-17.6%	63.1%
2000	43.4%	53.5%	-34.6%	49.9%	-56.9%	-49.9%	41.2%	77.7%	35.5%	-50.9%	-10.5%	72.6%
	-34.9%	-19.1%	-73.3%	44.7%	0.2%	70.5%	85.0%	-54.2%	-40.0%	13.4%	27.3%	-20.6%
	83.8%	-12.9%	-48.2%	65.4%	60.8%	50.0%	-31.0%	-39.7%	-5.2%	59.3%	46.7%	87.0%
	-33.0%	-37.6%	-86.2%	10.7%	-1.4%	91.2%	-49.8%	-31.2%	-91.5%	-38.6%	-49.8%	18.5%
	37.2%	43.2%	-45.4%	-32.2%	45.3%	25.2%	90.9%	-69.2%	42.3%	-40.8%	96.7%	10.9%
1995	-33.2%	-47.2%	98.1%	1.8%	-26.5%	72.6%	-21.8%	90.5%	20.6%	76.0%	50.6%	-42.0%

What's the meaning of the colors?

Heatmaps

Heatmaps with trees

Álvaro Figueira • VD • 2023 • 1ª ed.

35

Textual structures: Word Clouds

What can you infer from this visual represenation?

Álvaro Figueira • VD • 2023 • 1ª ed.

36

Textual Structures

37

Algorithmic Flow Charts

38

Recently, we have seen incredible, big data, real-time, dashboards

taken from bbc.com

39

taken from nbsnews.com

40

Presidential race Senate | House | Governors

	PARTY	VOTES	PCT.
Joe Biden ✓	**Democrat**	**3,701,835**	**55.7%**
Donald Trump*	Republican	2,852,042	42.9%
Jo Jorgensen	Libertarian	47,938	0.7%
Howie Hawkins	Green	23,349	0.4%
Brock Pierce	Independent	17,720	0.3%
Write-ins	Write-in	4,134	0.1%

84% of expected votes counted

*Incumbent

	PARTY	VOTES	PCT.
Donald Trump* ✓	**Republican**	**5,866,019**	**52.1%**
Joe Biden	Democrat	5,218,943	46.4%
Jo Jorgensen	Libertarian	125,291	1.1%
Howie Hawkins	Green	33,153	0.3%
Write-ins	Write-in	12,265	0.1%

99% of expected votes counted

*Incumbent

	PARTY	VOTES	PCT.
Joe Biden ✓	**Democrat**	**3,385,459**	**49.8%**
Donald Trump*	Republican	3,332,215	49.0%
Jo Jorgensen	Libertarian	78,025	1.1%
Write-ins	Write-in	8,384	0.1%

99% of expected votes counted

*Incumbent

taken from bbc.com