Reference

- Ward, M., Grinstein, G. G., Keim, D. **Interactive data visualization foundations, techniques, and applications**. Natick, Mass., A K Peters, 2010.
- Binary distance : http://people.revoledu.com/kardi/tutorial/Similarity/BinaryVariables.html

Visual Representation of Data

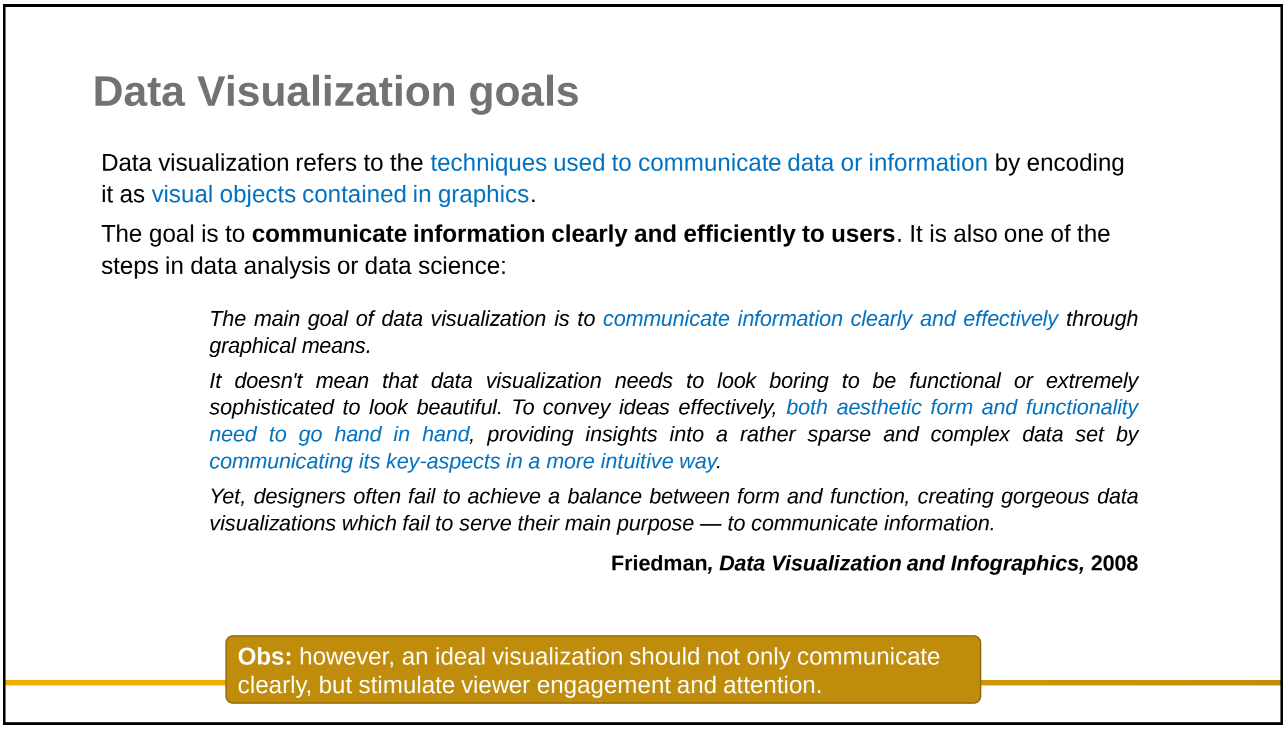
Álvaro Figueira, PhD.

MSc in Data Science - Data Visualization

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

FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO

U.PORTO



Data Visualization goals

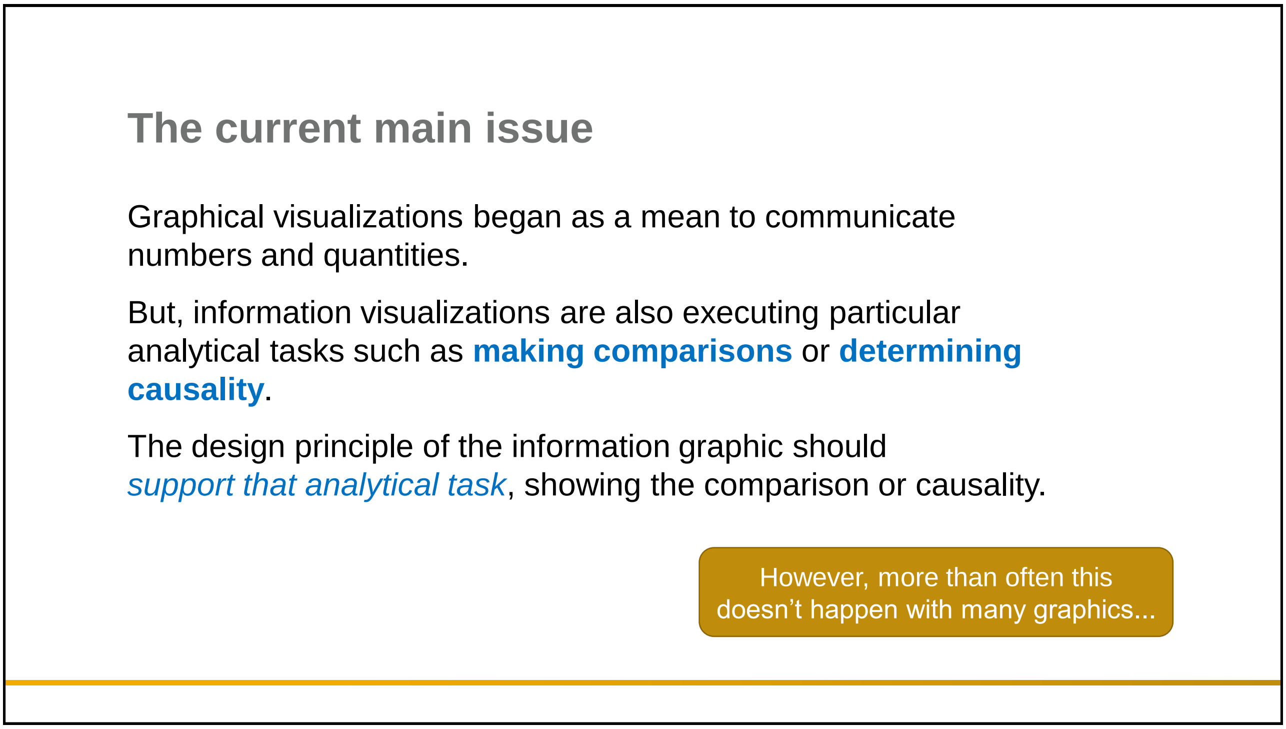
Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects contained in graphics.

The goal is to **communicate information clearly and efficiently to users**. It is also one of the steps in data analysis or data science:

> *The main goal of data visualization is to communicate information clearly and effectively through graphical means.*
>
> *It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way.*
>
> *Yet, designers often fail to achieve a balance between form and function, creating gorgeous data visualizations which fail to serve their main purpose — to communicate information.*
>
> **Friedman, *Data Visualization and Infographics*, 2008**

Obs: however, an ideal visualization should not only communicate clearly, but stimulate viewer engagement and attention.

143



The current main issue

Graphical visualizations began as a mean to communicate numbers and quantities.

But, information visualizations are also executing particular analytical tasks such as **making comparisons** or **determining causality**.

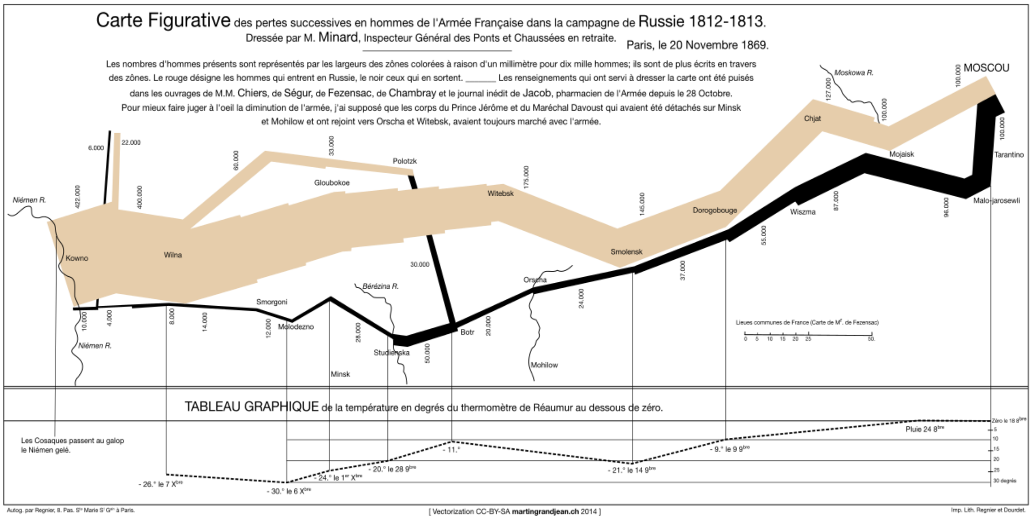
The design principle of the information graphic should *support that analytical task*, showing the comparison or causality.

However, more than often this doesn't happen with many graphics...

144

Graphical displays should

- **Show the data**
- Induce the viewer to think about the **substance** rather than about the methodology, the graphic design, or even the technology of graphic production
- Present many numbers in a **small space**
- **Avoid distorting** what the data has to “say”
- Make **large data sets coherent**
- Encourage the eye to **compare different pieces** of data
- Reveal the data at **several levels of detail**, from a broad overview to the fine structure
- Serve a **clear purpose**: description, exploration, tabulation or decoration
- Be closely **integrated with the statistical and verbal** descriptions of a data set.



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.”]

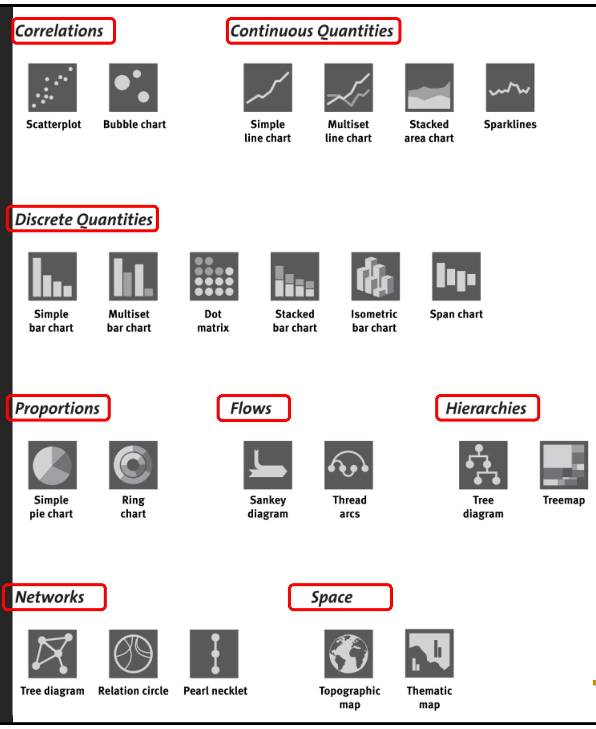
Quantitative Information Types

- Simple Quantities
- Time-series
- Rankings
- Part-to-whole
- Deviation
- Frequency distribution
- Correlation
- Nominal comparison
- Geographic or geospatial

147

Taxonomies for visual representation of data quantities

Christian Behrens, 2008



148

Some current techniques

Abstract structures

- Proportions: pie chart and ring chart
- Correlations: scatterplot, bubble chart
- Discrete quantities:
- bar chart, dot matrix, stacked bar charts
- Continuous quantities:
- line chart, stacked chart, sparklines
- Multidimensional: parallel coordinates

Hierarchical structures: trees

- Node-link layout (cartesian and polar)
- Treemaps: rectangular, circular and Voronoi
- Sunburst.

Relational structures: networks

- Node-link diagrams
- Layouts: matrix, linear, force directed, Sankey, circular, polar, geographical

Temporal structures

- Timelines: linear and polar
- Flows (Sankey diagrams)

Spatial structures: maps

- Dot distribution maps
- Isometric maps: isolines and heatmaps
- Choropleth maps
- Magnification and fish-eye views
- Cartograms: Dorling’s, area-value, isochronic

Temporal structures

- Animated maps
- Representation of trajectories
- Temporal flows

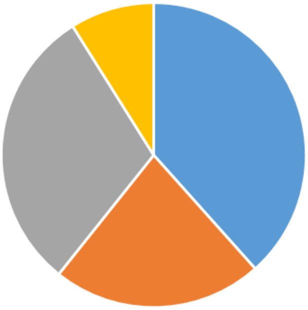
Textual structures

- Word clouds
- Textual trees

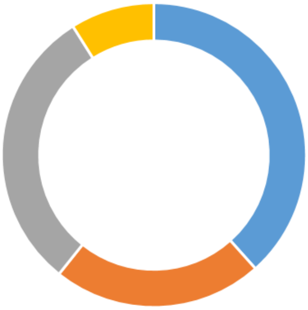
149

Proportions

Pie Chart



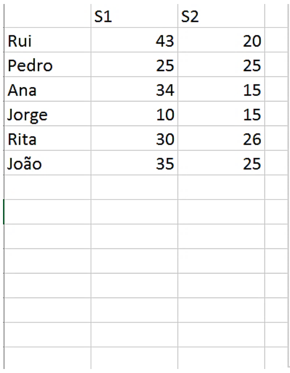
Ring Chart



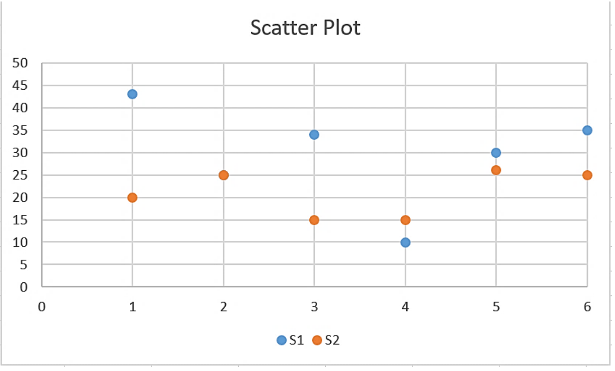
When to avoid?

150

Correlations: Scatter plot

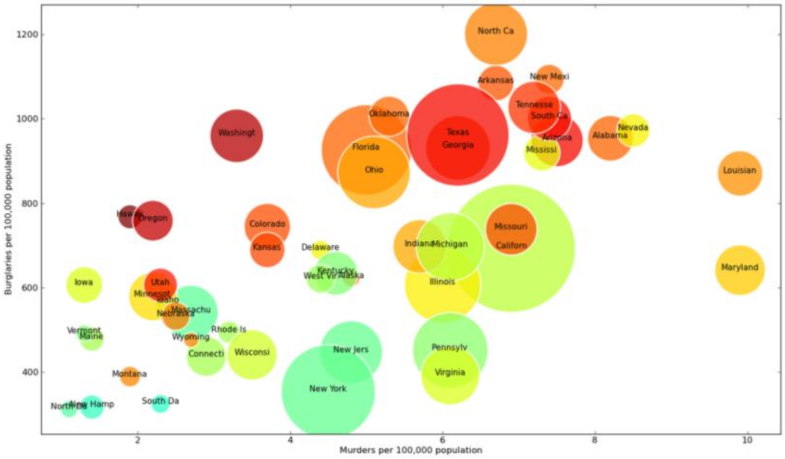


	S1	S2
Rui	43	20
Pedro	25	25
Ana	34	15
Jorge	10	15
Rita	30	26
João	35	25



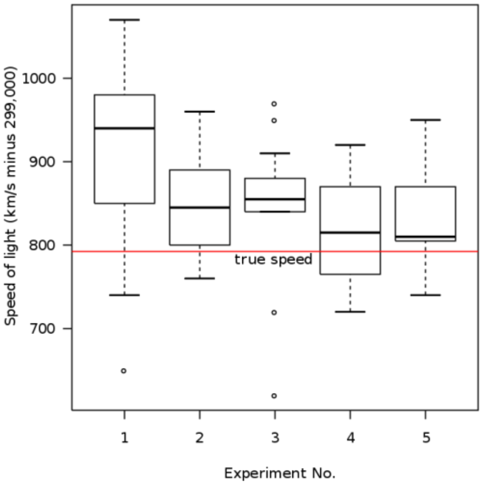
What does the x-axis represent?

Correlations: Bubble chart



How many properties must have each record in this chart?

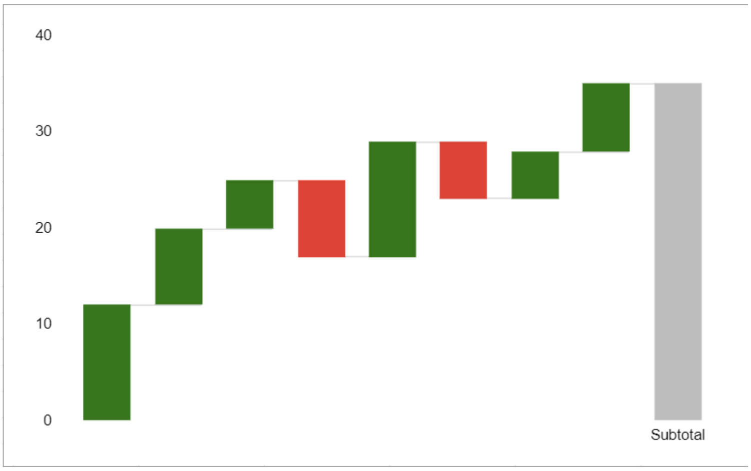
Box Plot



Can you explain a box plot?

153

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

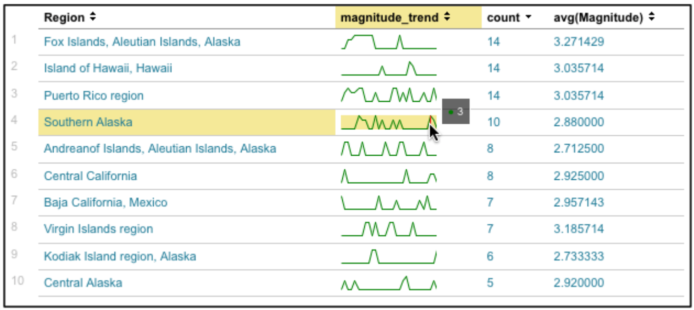
154

154

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**.

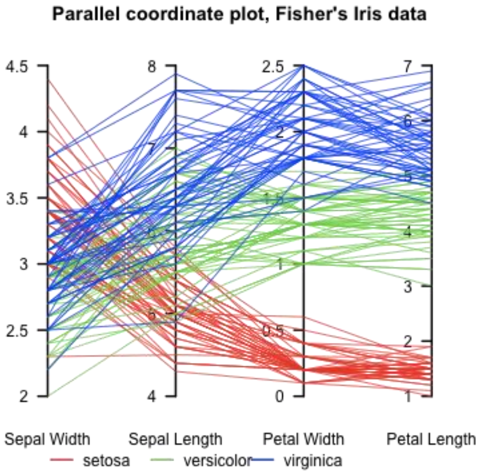
Obs: Sparklines can run in a text layout.



	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		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

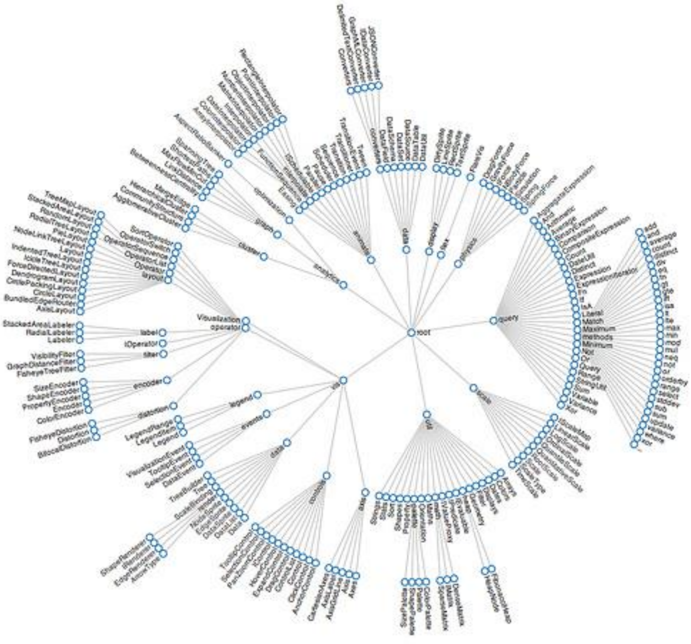
Note: the aspect ratio of line chart or a sparkline are crucial for good reading. The visual average of the hill-slopes within the line should be ideally 45 degrees.

Multidimensions and Parallel Coordinates



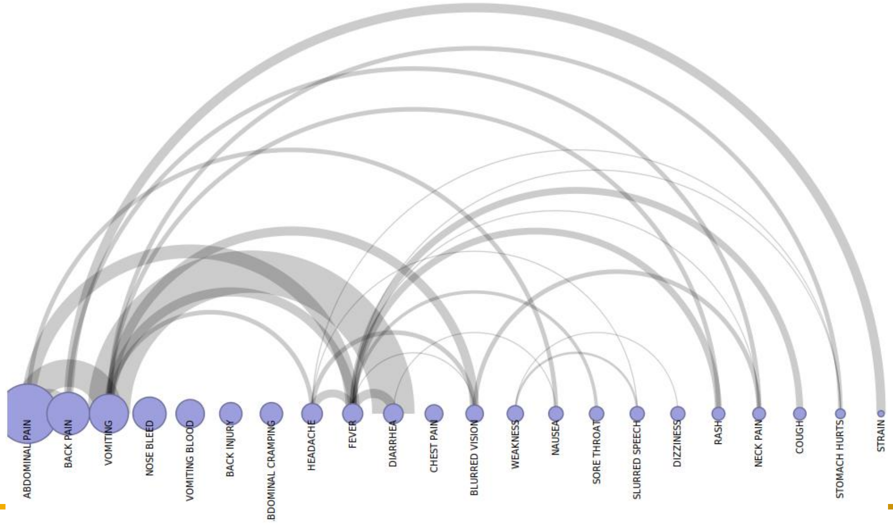
Parallel coordinates can be arranged in order to show certain correlations among attributes for the same type of object

Hierarchical structures: Trees



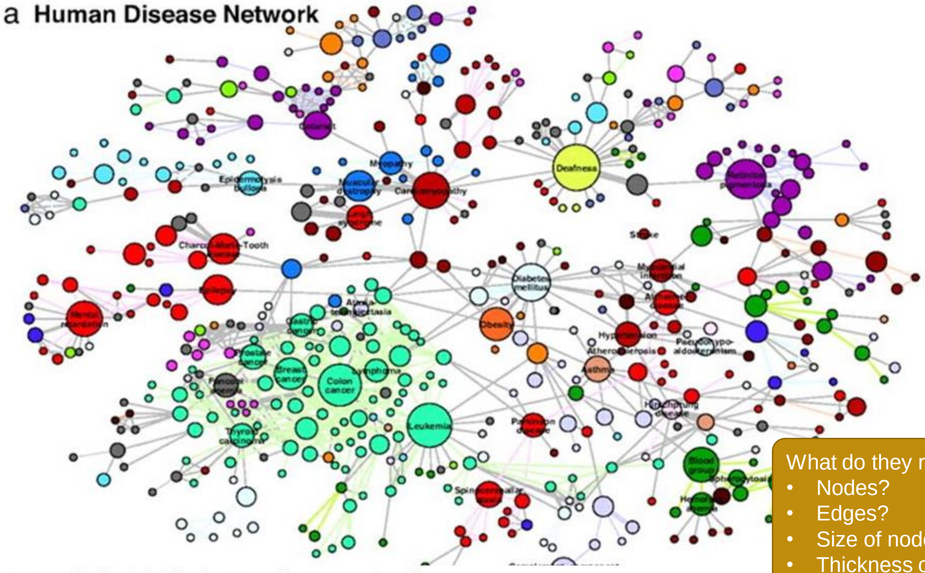
157

Relational structures with linear layout



158

Relational structures: networks



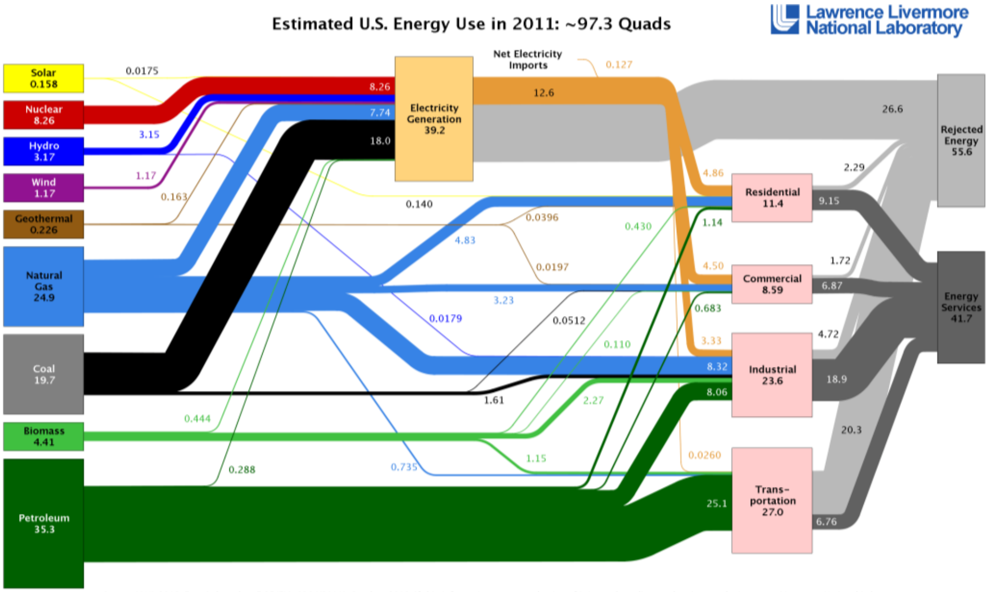
source: Goh et al. The human disease network

What do they represent:

- Nodes?
- Edges?
- Size of nodes?
- Thickness of edges?

159

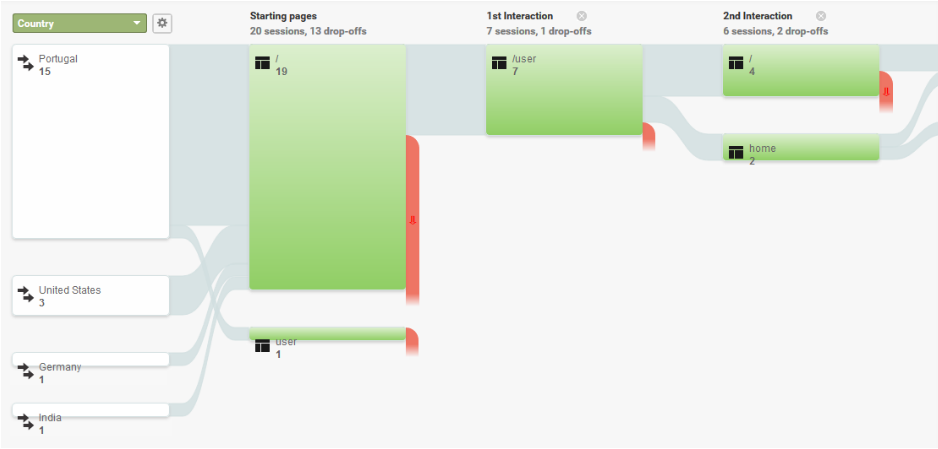
Relational structures: Sankey diagrams



Source: LLNL 2012. Data is based on DOE/EIA-0384(2011), October, 2012. If this information or a reproduction of it is used, credit must be given to the Lawrence Livermore National Laboratory and the Department of Energy, under whose auspices the work was performed. Distributed electricity represents only retail electricity sales and does not include self-generation. EIA reports flows for non-thermal resources (i.e., hydro, wind and solar) in BTU-equivalent values by assuming a typical fossil fuel plant "heat rate." The efficiency of electricity production is calculated as the total retail electricity delivered divided by the primary energy input into electricity generation. End use efficiency is estimated as 80% for the residential, commercial and industrial sectors, and as 25% for the transportation sector. Totals may not equal sum of components due to independent rounding. LLNL-MI-410527

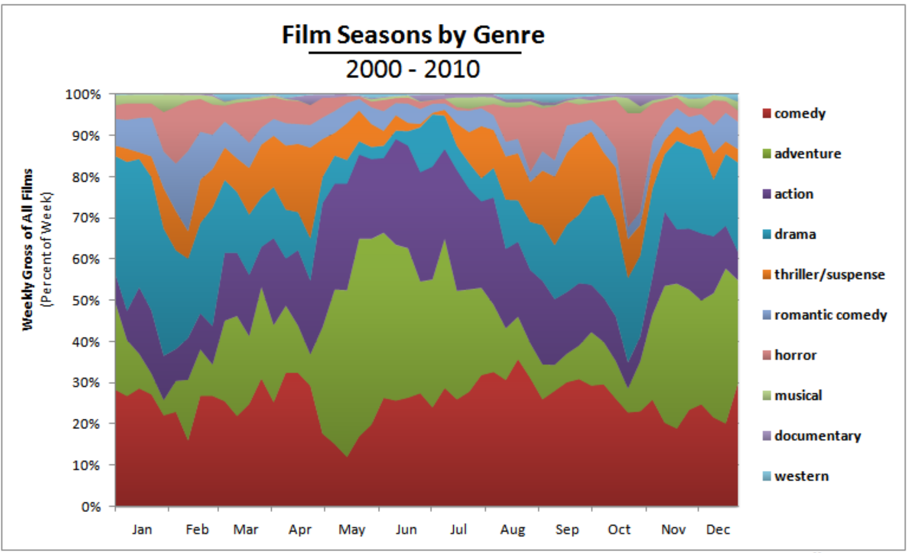
160

Relational structures: Sankey diagrams



161

Time series: stacked graphs

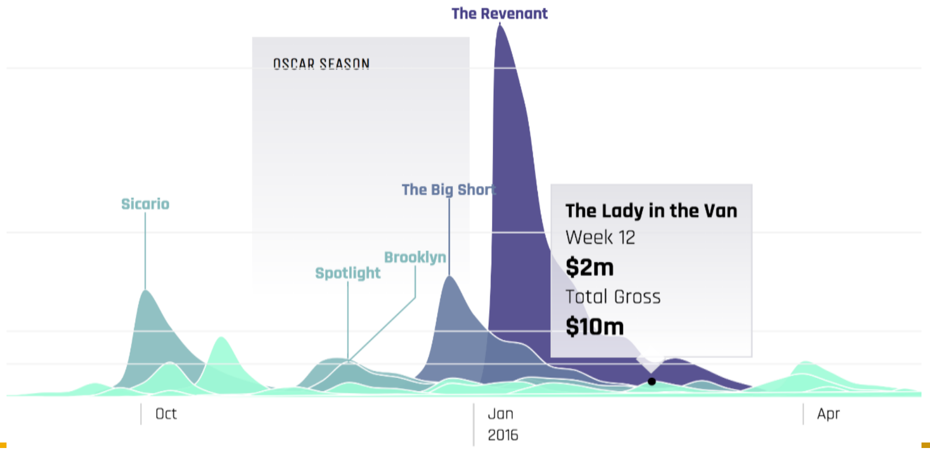


BoxOfficeQuant.com

162

Filled regions

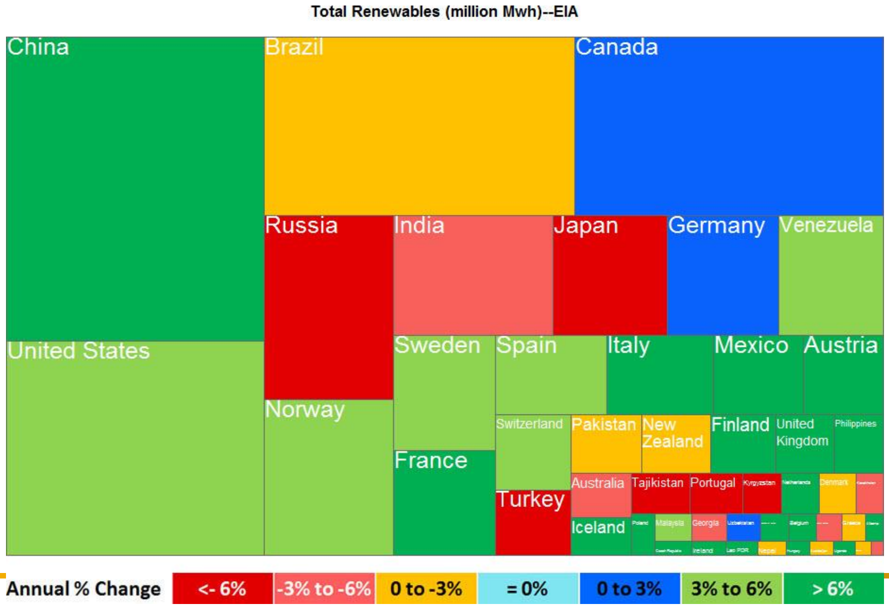
This is just a teaser for you all...



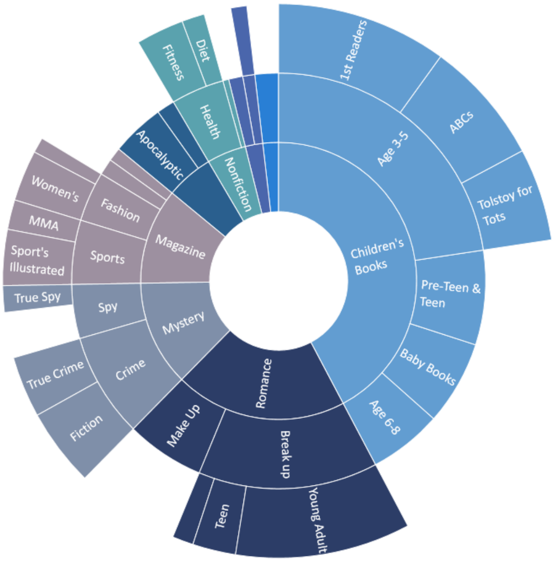
The Lady in the Van
Week 12
$2m
Total Gross
$10m

163

Hierarchies: Treemaps

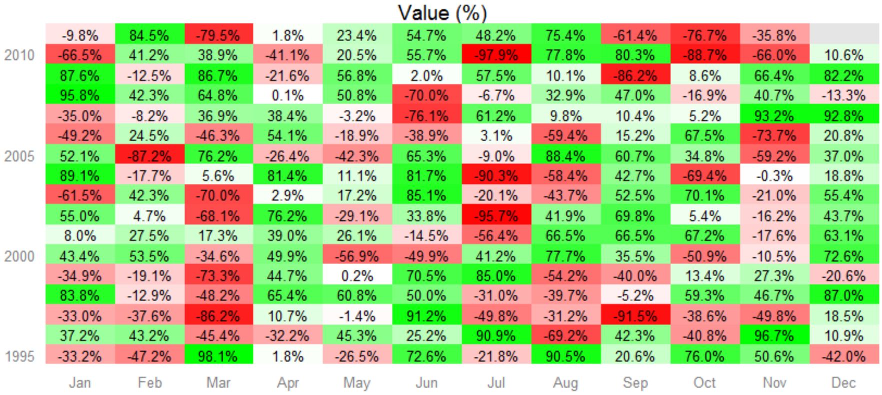


Hierarchies: circular treemaps



165

Heatmaps

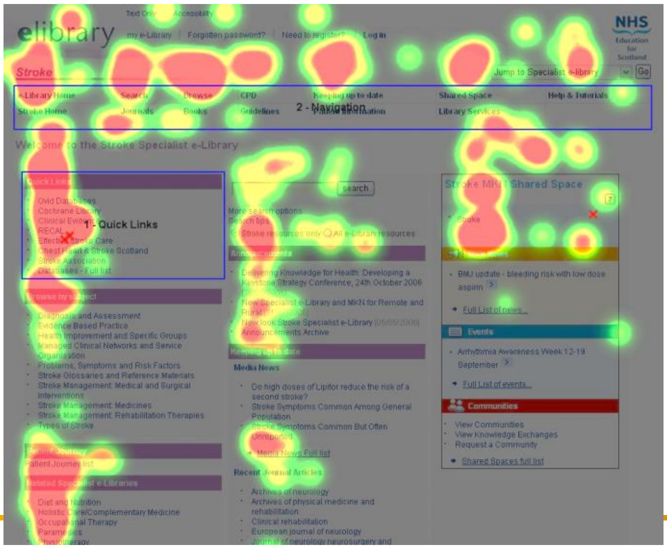

Value (%)

	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?

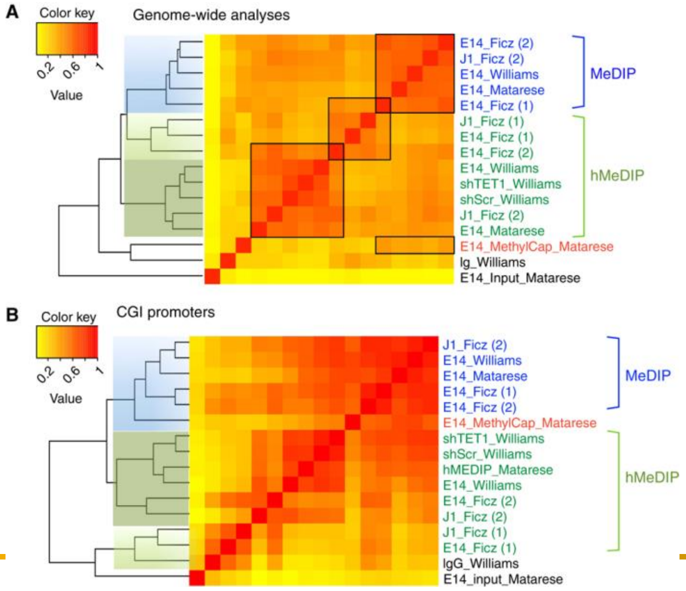
166

Heatmaps


elibrary
NHS Education for Scotland
Text Only
Accessibility
my e-Library | Forgotten password? | Need to register? | Log in
Stroke
Jump to Specialist e-library
Go
e-Library Home
Search
Browse
CPD
Keeping up to date
Shared Space
Help & Tutorials
Stroke Home
Journals
Books
Guidelines
2 - Navigation
Library Services
Welcome to the Stroke Specialist e-Library
Quick Links
1 - Quick Links
Ovid Databases
Cochrane Library
RECAL
Effective Stroke Care
Chest Heart & Stroke Scotland
Stroke Association
Databases - Full list
search
More search options
Search tips
Stroke resources only
All e-Library resources
Stroke MKN Shared Space
Browse by subject
Diagnosis and Assessment
Evidence Based Practice
Health Improvement and Specific Groups
Managed Clinical Networks and Service Organisation
Problems, Symptoms and Risk Factors
Stroke Glossaries and Reference Materials
Stroke Management: Medical and Surgical Interventions
Stroke Management: Medicines
Stroke Management: Rehabilitation Therapies
Types of Stroke
Delivering Knowledge for Health: Developing a Keystone Strategy Conference, 24th October 2006
New Specialist e-Library and MKN for Remote and Rural
New look Stroke Specialist e-Library
Announcements Archive
BMJ update - bleeding risk with low dose aspirin
Full List of news
Events
Arrhythmia Awareness Week 12-19 September
Full List of events
Communities
View Communities
View Knowledge Exchanges
Request a Community
Shared Spaces full list
Media News
Do high doses of Lipitor reduce the risk of a second stroke?
Stroke Symptoms Common Among General Population
Stroke Symptoms Common But Often Unreported
Media News Full list
Recent Journal Articles
Archives of neurology
Archives of physical medicine and rehabilitation
Clinical rehabilitation
European journal of neurology
Journal of neurology neurosurgery and
Patient Journey
Patient Journey list
Related Specialist e-Libraries
Diet and Nutrition
Holistic Care/Complementary Medicine
Occupational Therapy
Paramedics
Physiotherapy

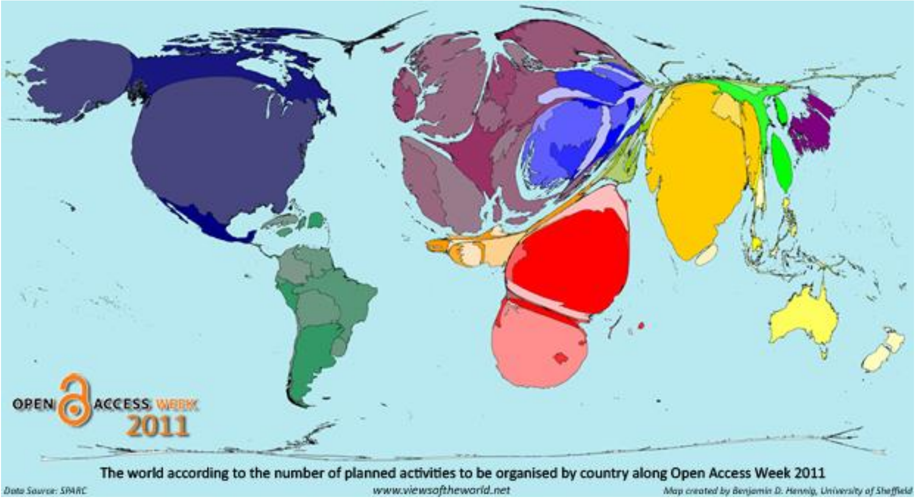
167

Heatmaps with trees



168

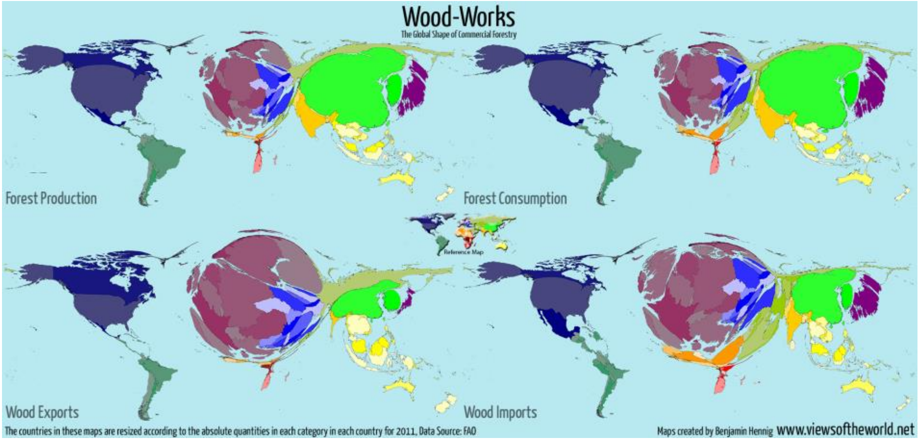
Maps and area cartograms



The world according to the number of planned activities to be organised by country along Open Access Week 2011

Data Source: SPARC www.viewsoftheworld.net Map created by Benjamin D. Hennig, University of Sheffield

Maps and area cartograms



The countries in these maps are resized according to the absolute quantities in each category in each country for 2011, Data Source: FAO

Maps created by Benjamin Hennig www.viewsoftheworld.net

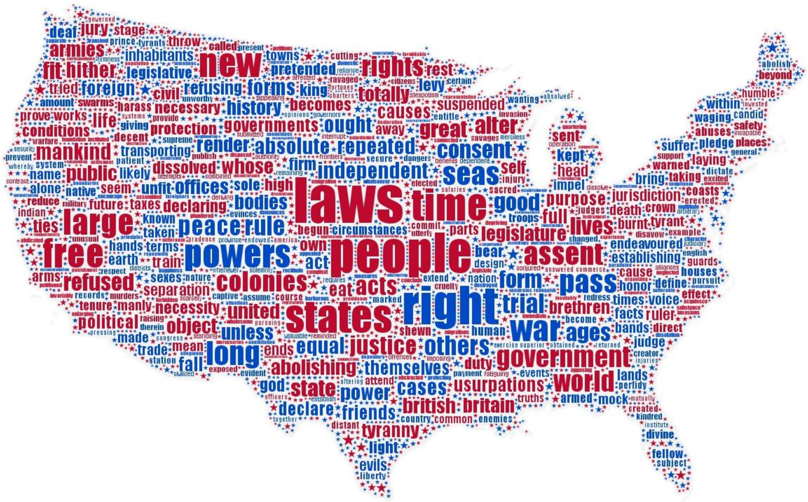
Textual structures: Word Clouds

ways
create
target
asked
tutorial
interesting
either
cloud
activities
piece
survey
video
word
projects
comfortable
reading
review
pairs
easy
get
largest
text
tool
dictogloss
great
passages
use
visual
future
food
poll
entered
corridor
identify
images
speeches
links
many
include
sentence
ambitions
often
recycle
containing
nice
seen
speculate
websites
sentences
teaching
coursebook
suggestions
edit
language
easier
news
key
improving
theme
wordle
vocabulary
works
emphasised
predict
starting
screen
link
passage
write
projector
hobbies
shown
ideas
etc
series
used
perhaps
obviously
posters
make
level
seems
surveys
already
frequently
display
type
comments
analysed
resulting
higher
best
displaying
displayed
colour
using
useful
change
varied
may
creating
recently
wish
done
appear
created
number
clouds
predictions
benefits
provide
computer
please
checked
results
synonyms
ask
place
favourite
bias
looks
part
words
fun
students
covered
writing
image
topic
ok
new
original
reconstruction
shows
anytime
groups
good
support
tell
save
class
texts
describe
answers
based

What can you infer from this?

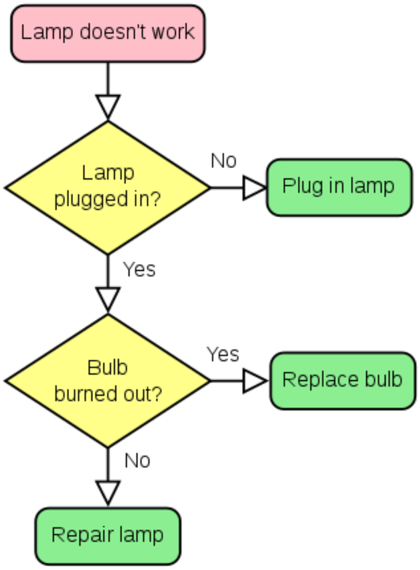
171

Textual Structures



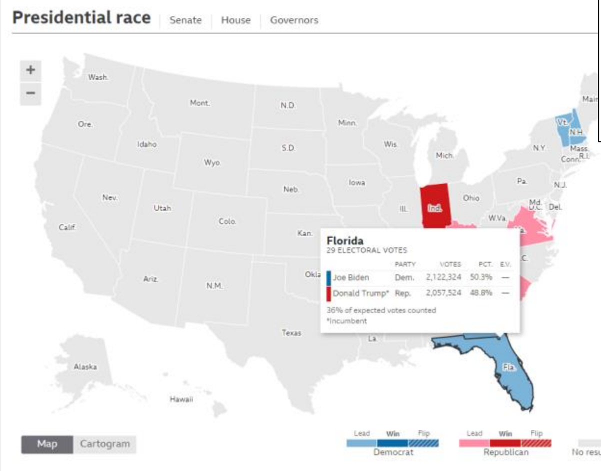
172

Algorithmic Flow Charts



173

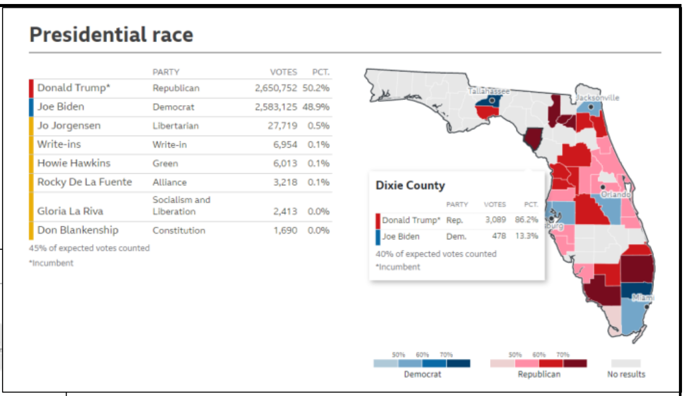
Recently, we have seen incredible, big data, real-time, dashboards



Florida
29 ELECTORAL VOTES

	PARTY	VOTES	PCT.	E.V.
Joe Biden	Dem.	2,122,324	50.3%	—
Donald Trump*	Rep.	2,057,524	48.8%	—

36% of expected votes counted
*Incumbent



Presidential race

	PARTY	VOTES	PCT.
Donald Trump*	Republican	2,650,752	50.2%
Joe Biden	Democrat	2,583,125	48.9%
Jo Jorgensen	Libertarian	27,719	0.5%
Write-ins	Write-in	6,954	0.1%
Howie Hawkins	Green	6,013	0.1%
Rocky De La Fuente	Alliance	3,218	0.1%
Gloria La Riva	Socialism and Liberation	2,413	0.0%
Don Blankenship	Constitution	1,690	0.0%

45% of expected votes counted
*Incumbent

Dixie County

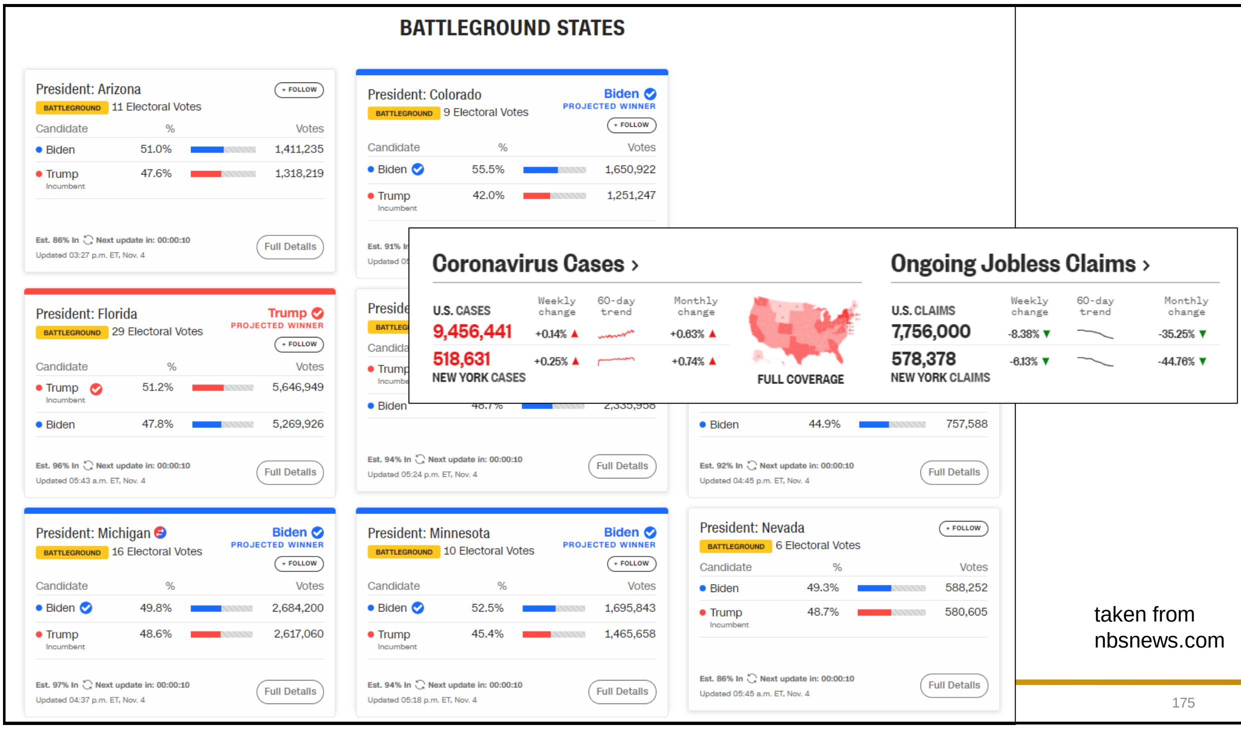
	PARTY	VOTES	PCT.
Donald Trump*	Rep.	3,089	86.2%
Joe Biden	Dem.	478	13.3%

40% of expected votes counted
*Incumbent

taken from bbc.com

174

174



BATTLEGROUND STATES

President: Arizona + FOLLOW
BATTLEGROUND 11 Electoral Votes

Candidate	%	Votes
● Biden	51.0%	1,411,235
● Trump Incumbent	47.6%	1,318,219

Est. 86% In Next update in: 00:00:10
Updated 03:27 p.m. ET, Nov. 4
Full Details

President: Colorado
Biden ✓ PROJECTED WINNER + FOLLOW
BATTLEGROUND 9 Electoral Votes

Candidate	%	Votes
● Biden ✓	55.5%	1,650,922
● Trump Incumbent	42.0%	1,251,247

Est. 91% In
Updated 05

President: Florida
Trump ✓ PROJECTED WINNER + FOLLOW
BATTLEGROUND 29 Electoral Votes

Candidate	%	Votes
● Trump ✓ Incumbent	51.2%	5,646,949
● Biden	47.8%	5,269,926

Est. 96% In Next update in: 00:00:10
Updated 05:43 a.m. ET, Nov. 4
Full Details

Preside
BATTLEG

Candida		
● Trump Incumbe		
● Biden	48.7%	2,335,958

Est. 94% In Next update in: 00:00:10
Updated 05:24 p.m. ET, Nov. 4
Full Details

● Biden	44.9%	757,588

Est. 92% In Next update in: 00:00:10
Updated 04:45 p.m. ET, Nov. 4
Full Details

President: Michigan
Biden ✓ PROJECTED WINNER + FOLLOW
BATTLEGROUND 16 Electoral Votes

Candidate	%	Votes
● Biden ✓	49.8%	2,684,200
● Trump Incumbent	48.6%	2,617,060

Est. 97% In Next update in: 00:00:10
Updated 04:37 p.m. ET, Nov. 4
Full Details

President: Minnesota
Biden ✓ PROJECTED WINNER + FOLLOW
BATTLEGROUND 10 Electoral Votes

Candidate	%	Votes
● Biden ✓	52.5%	1,695,843
● Trump Incumbent	45.4%	1,465,658

Est. 94% In Next update in: 00:00:10
Updated 05:18 p.m. ET, Nov. 4
Full Details

President: Nevada + FOLLOW
BATTLEGROUND 6 Electoral Votes

Candidate	%	Votes
● Biden	49.3%	588,252
● Trump Incumbent	48.7%	580,605

Est. 86% In Next update in: 00:00:10
Updated 05:45 a.m. ET, Nov. 4
Full Details

Coronavirus Cases ›

	Weekly change	60-day trend	Monthly change
U.S. CASES 9,456,441	+0.14% ▲		+0.63% ▲
518,631 NEW YORK CASES	+0.25% ▲		+0.74% ▲

FULL COVERAGE

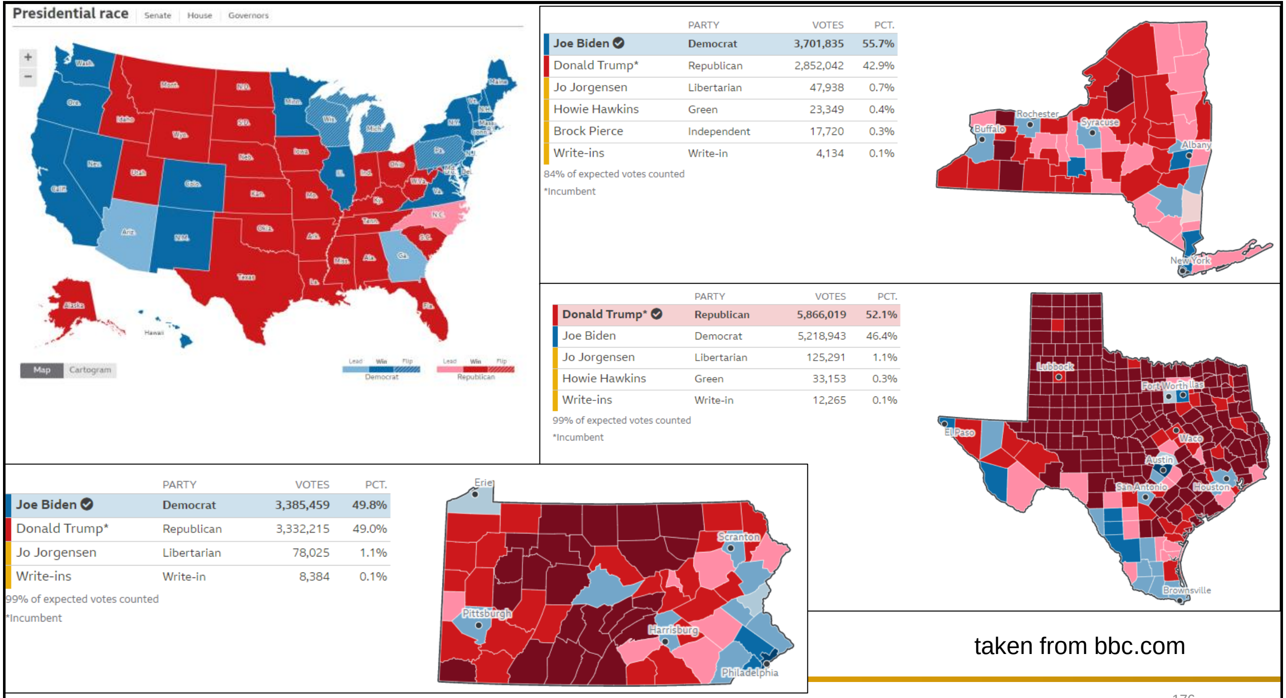
Ongoing Jobless Claims ›

	Weekly change	60-day trend	Monthly change
U.S. CLAIMS 7,756,000	-8.38% ▼		-35.25% ▼
578,378 NEW YORK CLAIMS	-6.13% ▼		-44.76% ▼

taken from nbsnews.com

175

175



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

176

176

https://www.nytimes.com/interactive/2020/11/03/us/elections/results-president.html

177