Data visualization

European Data Incubator, Bilbao

Mikel Madina & Miren Berasategi

9th November 2018

Deusto

1. Data visualization as artefact

> artefact (*US* **artifact**) **noun**

> [1] An object made by a human being, typically one of cultural or historical interest.
> *'gold and silver artefacts'*

> [2] Something observed in a scientific investigation or experiment that is not naturally present but occurs as a result of the preparative or investigative procedure.
> *'the curvature of the surface is an artefact of the wide-angle view'*

The Oxford Dictionary of English

The atomic level

Number of observations → marks

$$\frac{\text{data points}}{\text{observations}} < \text{pixels}$$

In a picture of 1000 x 1000 pixels, the maximum of observations to fit is 10^6 (1 million). How big is that?

The atomic level

Some strategies to overcome these constraints:

1. Filter
2. Split data into multiple charts
3. Augmenting visualizations
4. Densify

The atomic level

1. Filter observations TREND

- By design, communicating a selection of data
- By allowing the user to filter according to their interests
  - Innovative filtering INNOVATION (i.e. Smart brushing)

Smart Brushing of Parallel Coordinates, IEEE VIS 2018 talk, 25 Oct, Berlin

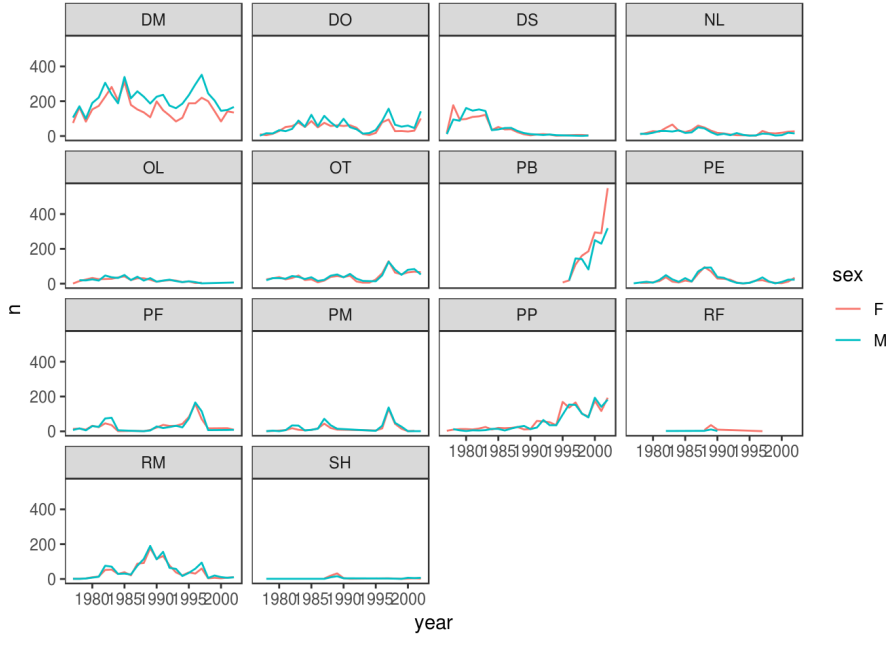
The atomic level

2. Split data into multiple charts TREND

Facets, trellis, small multiples

The atomic level

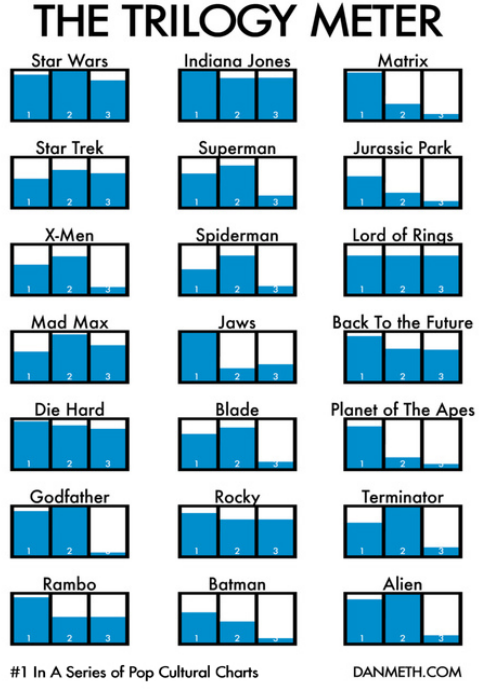
2. Split data into multiple charts TREND



Example facet visualization of the observation of animals by species and sex (Michonneau & Fournier 2018)

The atomic level

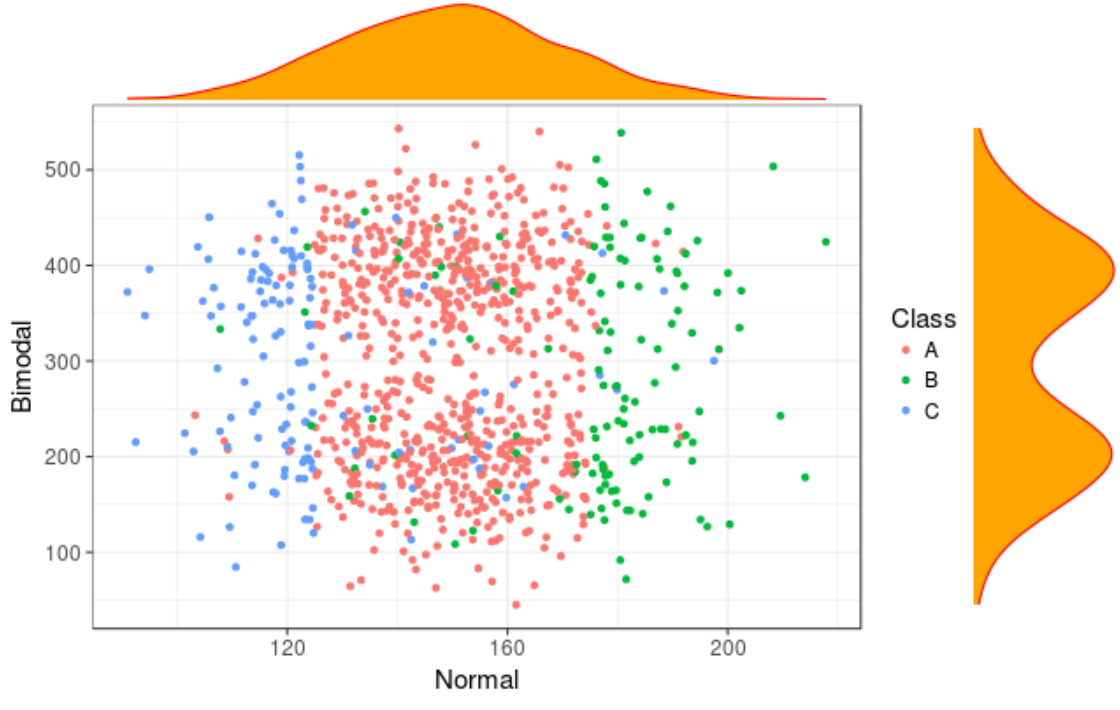
2. Split data into multiple charts TREND



Example of small multiples: The Trilogy Meter (Meth 2009)

The atomic level

3. Augmenting visualizations TREND

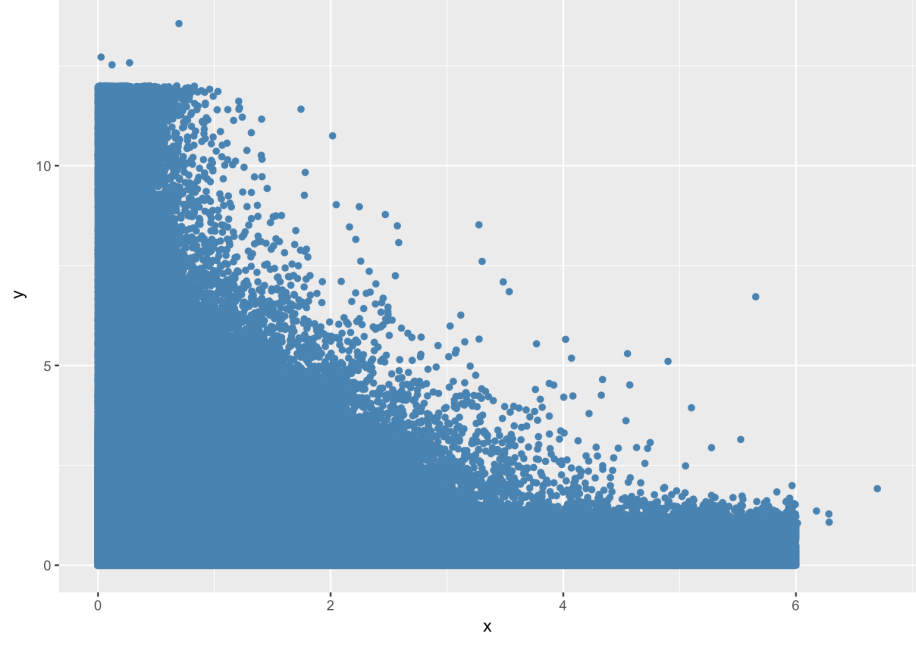


ggExtra, adding marginal histograms to ggplot2 (by Dean Attali)

The atomic level

4. Densify

- Escaping overplotting in scatterplots TREND

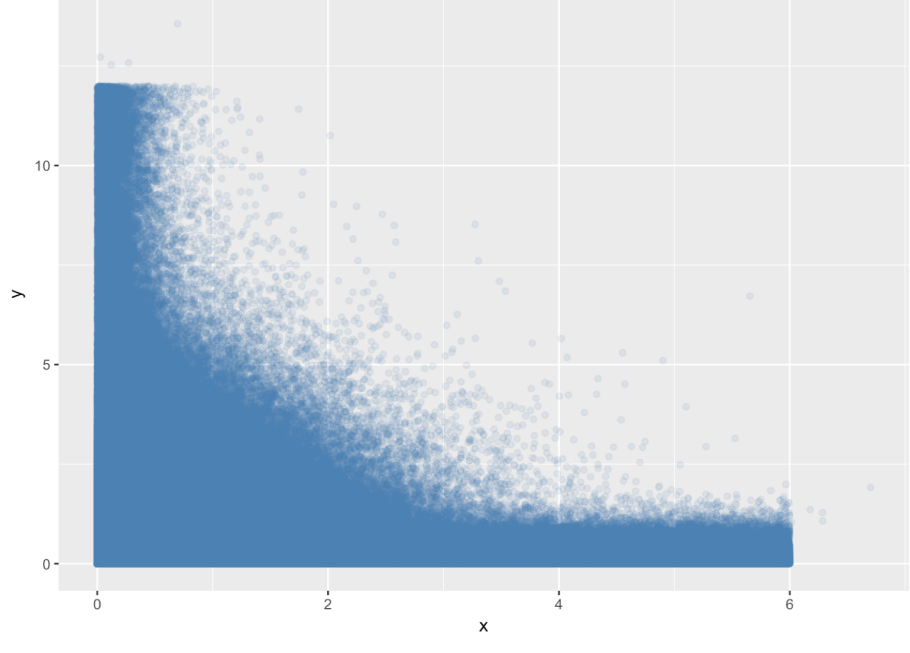


Clearly overplotted scatterplot (Kovalyshyn 2017)

The atomic level

4. Densify

- Escaping overplotting in scatterplots TREND

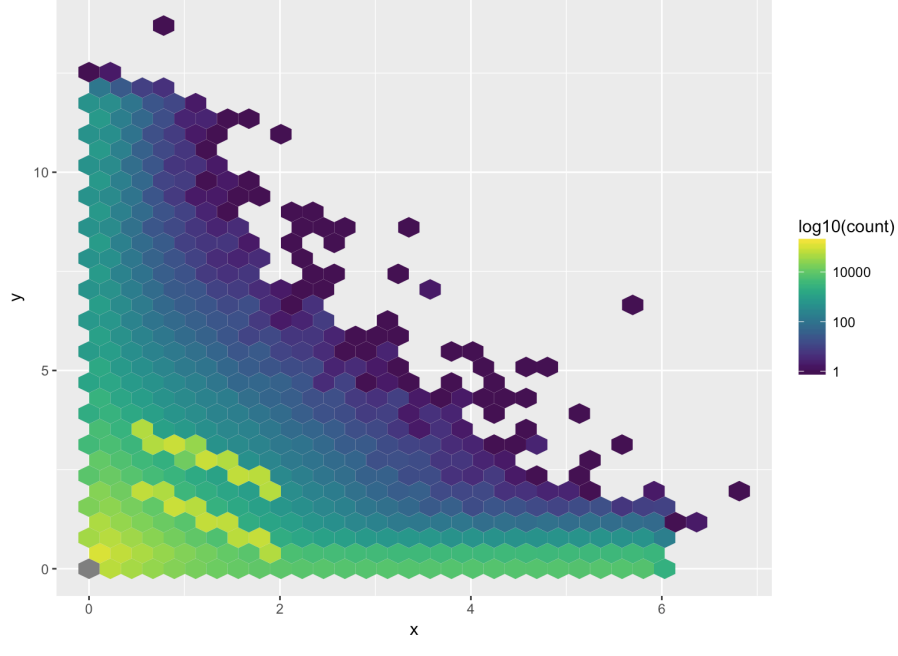


Using transparency is not helpful (Kovalyshyn 2017)

The atomic level

4. Densify

- Escaping overplotting in scatterplots TREND



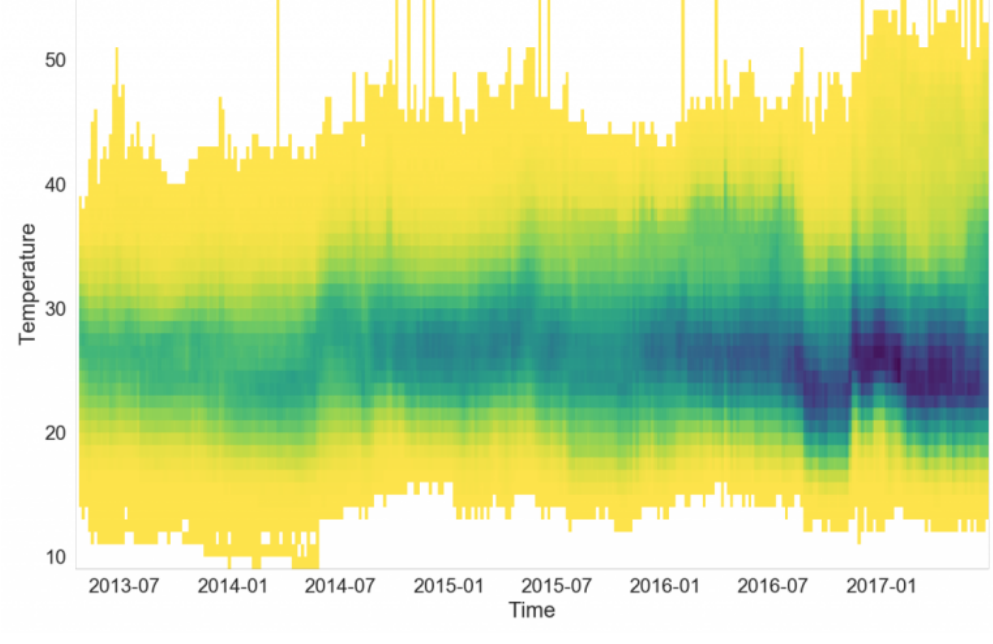
Now we can **see** density (Kovalyshyn 2017)

The atomic level

4. Densify

- Timelines INNOVATION

step-by-step



The atomic level

Adding the time factor:

- Static visualizations with real data (at the time of loading)
- Real-time visualizations, static and auto-refreshed
- Streaming data visualizations showing the flow of data

Require an additional effort for operational intelligence, where immediate decision making could be a requirement.

Source: Aragues 2018

The atomic level

How to communicate *nothingness*? (Kirk 2014)

Andy Kirk - The Design of Nothing: Null, Zero, Blank

The atomic level

How to communicate *nothingness*?

- **Null** Absence of measurement
- **Zero** Absence of amount/magnitude
- **Blank** Try to use *nothing* to represent *something*

Andy Kirk - The Design of Nothing: Null, Zero, Blank

The atomic level

How to communicate *nothingness*?

- **Null** Absence of measurement
- **Zero** Absence of amount/magnitude
- **Blank** Try to use *nothing* to represent *something*
- The design should be **invisible**

Andy Kirk - The Design of Nothing: Null, Zero, Blank

The atomic level

Projections, statistical models, uncertainty

Unpublished papers:

Alex Kale, Francis Nguyen, Matthew Kay, Jessica Hullman (2019), "Hypothetical Outcome Plots Help Untrained Observers Judge Trends in Ambiguous Data", *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2019

Number of variables TREND

- Reduce dimensionality (statistically): PCA, factors, clusters
- Restructure dataset (from wide to tall) (heatmaps)

Generating new idioms INNOVATION

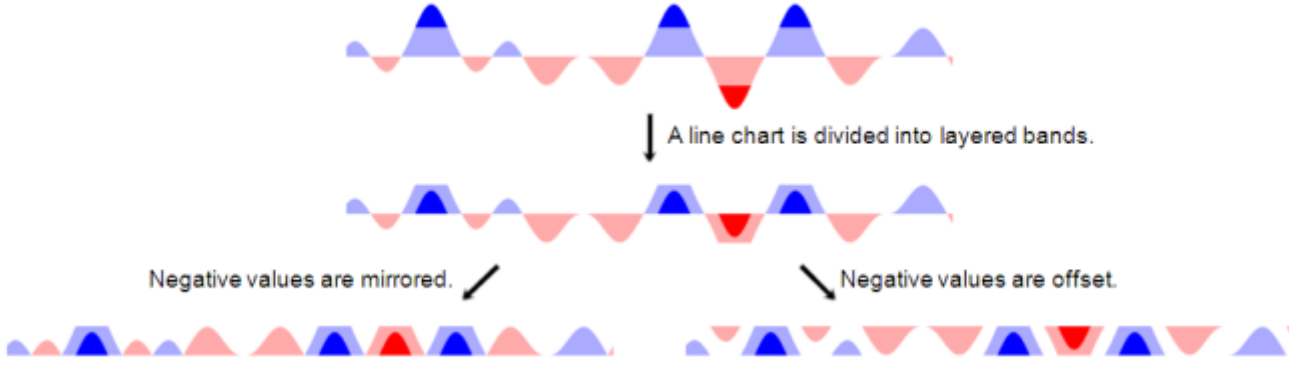
A word of caution:

- will need to be custom coded
- readers will require training
- correct interpretation may be more time demanding

Xenographics: Weird but (sometimes) useful charts

Generating new idioms

INNOVATION

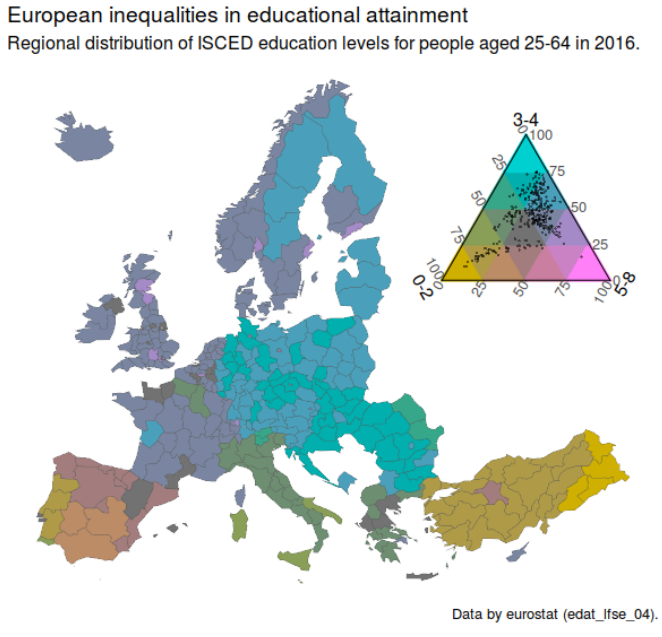


Horizon charts

Source: Jeffrey Heer, Nicholas Kong, Maneesh Agrawala (2009), "Sizing the Horizon: The Effects of Chart Size and Layering on the Graphical Perception of Time Series Visualizations". *ACM Human Factors in Computing Systems (CHI)*, pp. 1303 - 1312

Generating new idioms

INNOVATION

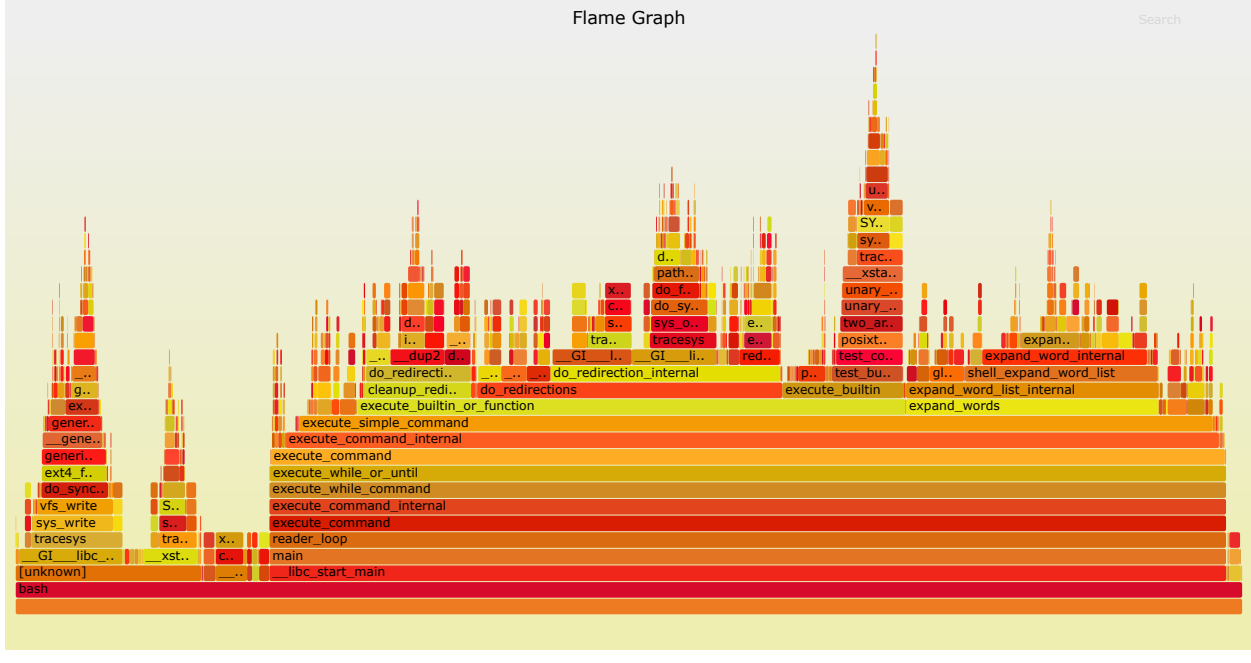


Choropleth maps with `tricolore`

Source: Jonas Schöley (2018), “Choropleth maps with tricolore”

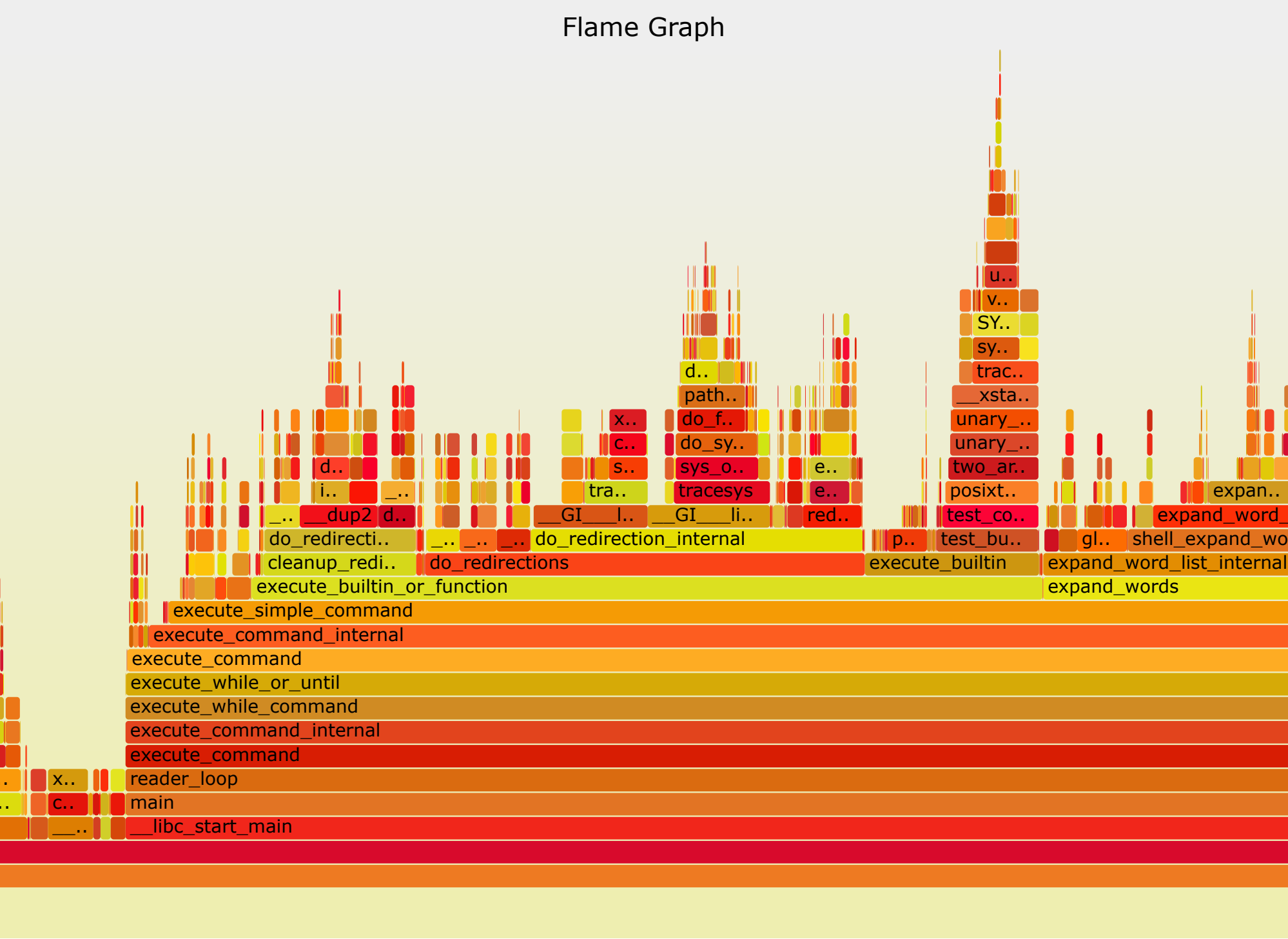
Generating new idioms

INNOVATION



Flame graphs

Source: Brendan Gregg (2016), “The Flame Graph”. *ACM Queue* Vol. 14, No. 2


Flame Graph
u..
v..
SY..
sy..
trac..
__xsta..
unary_..
unary_..
two_ar..
posixt..
test_co..
test_bu..
d..
path..
do_f..
do_sy..
sys_o..
tracesys
x..
c..
s..
tra..
e..
e..
red..
p..
d..
i..
_..
_..
__dup2
d..
__GI___l..
__GI___li..
do_redirecti..
_..
_..
_..
do_redirection_internal
cleanup_redi..
do_redirections
execute_builtin
execute_builtin_or_function
expand_words
expan..
gl..
expand_word_
shell_expand_wo
expand_word_list_internal
execute_simple_command
execute_command_internal
execute_command
execute_while_or_until
execute_while_command
execute_command_internal
execute_command
x..
reader_loop
c..
main
___..
__libc_start_main

Multiple Linked Views (MLV) TREND

INNOVATION

Making Data Visual

Ricardo Langner, Ulrike Kister, Raimund Dachselt (2019). "Multiple Coordinated Views at Large Displays for Multiple Users: Empirical Findings on User Behavior, Movements, and Distances", *IEEE Transactions on Visualization and Computer Graphics* 25(1) (InfoVis 2018, Berlin) 10.1109/TVCG.2018.2865235

[VIS'18] Multiple Coordinated Views at Large Displays for Multiple Users

Beyond 2 dimensions

INNOVATION

FiberClay: Sculpting Three Dimensional Trajectories to Reveal Structural Insights

DXR Overview

Beyond 2 dimensions INNOVATION

- Virtual Reality
- Augmented reality

References:

Ronell Sicat, Jiabao Li. *DXR: A Toolkit for Building Immersive Data Visualizations*

Christophe Hurter, Nathalie Henry Riche, Steven M. Drucker, Maxime Cordeil, Richard Alligier, Romain Vuillemot (2018), "FiberClay: Sculpting Three Dimensional Trajectories to Reveal Structural Insights", IEEE Transactions on Visualization and Computer Graphics_ 25(1) (InfoVis 2018, Berlin)

Senses and control interfaces INNOVATION

(How is this *visualization*?)

- Sound
- *smell necklace*
  https://twitter.com/NElmqvist/status/1055370456528773

2. Data visualization as a tool for communication

2. Data visualization as a tool for communication

> *The modern approach to data visualization is focused on quickly making data visualization.*

(Meeks 2018)

2. Data visualization as a tool for communication

Focus on speed affects:

- how data visualization products are designed
- what tools are used to create them
- the role of the creator in relation to the product
- how engagement with readers in envisioned

2. Data visualization as a tool for communication

> *Ultimately, data visualization is not a technical problem, it's a design problem and, more than that, a communication problem .*

(Meeks 2018)

Let's look at what charts **say**, **mean**, and **do**.

What charts say

Explicitly

Charts *do* "show me the data" (but remember that it's more that they **tell** the data than actually show it).

Means chosing the right specific chart to use in order to display and query the data.

How to improve: Expose data cleanly and clearly. Accuracy *vs.* precision.

What charts say

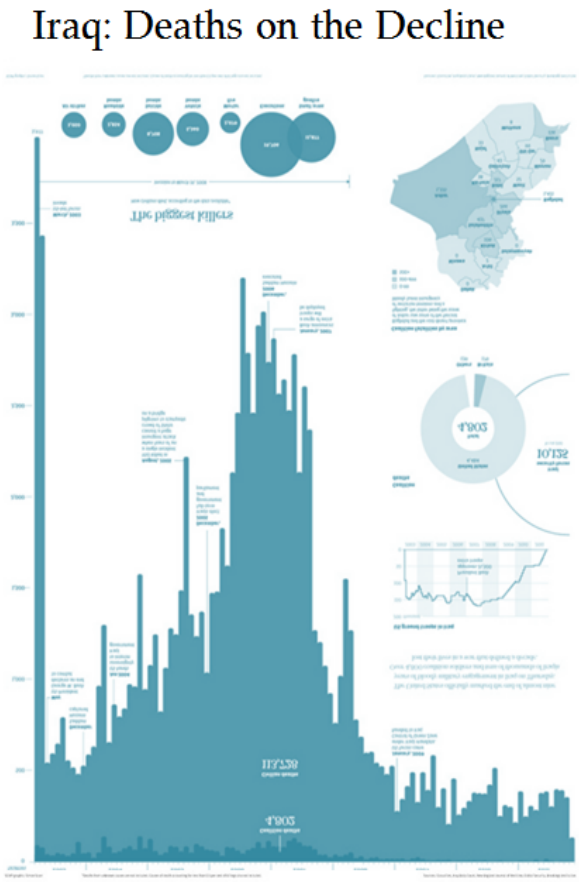
Implicitly

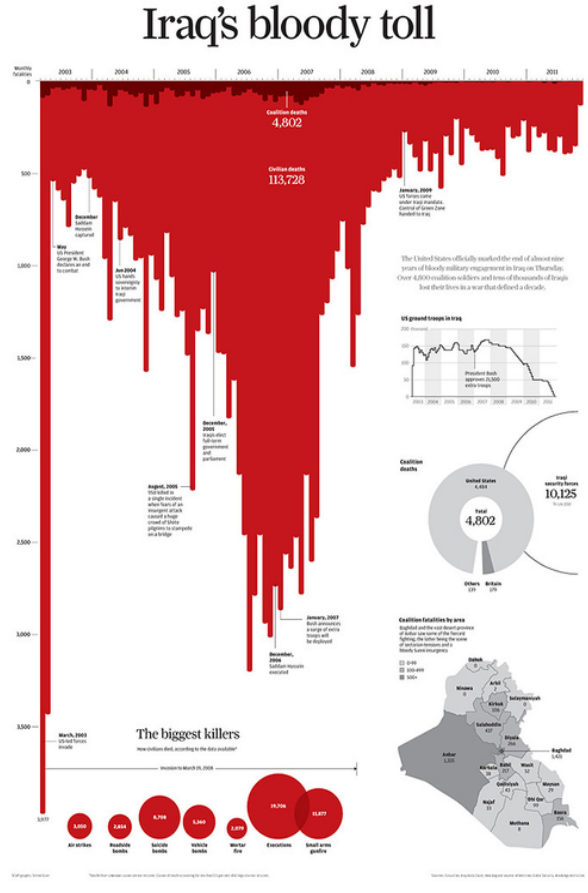
No chart is an unbiased view of the data, as data visualization is a manufactured artefact.

> *All data is transformed to be in a chart, and the inaction of not designing that transformation carries just as strong an implication as the action of transforming it.*

(Meeks 2018)

What charts say





Original infographic (right) by Simon Scarr and redesigned, more default representation of an histogram (left), redesigned by Andy Cotgreave (Meeks 2018)

What charts say

Implicitly

The implicit channel of a data visualization (the title and other framing elements) can be even more powerful than the explicit channel.

How to improve: Style should be intentional, purposeful and thematically appropriate, not the result of defaults or superficial decisions.

What charts say

Systematically

> *[…] all charts display data and all data is a proxy for the systems that created and measured that data.*

(Meeks 2018)

How to improve: Caution not to reveal an underlying system that is proprietary or confidential.

What charts say

Descriptively

- internally: axes, labels, annotations
- externally: surrounding text, figure descriptions, discussions

Unlike the implicit channel, the descriptive channel is active and purposeful (not subconscious).

How to improve: Consider annotations, labels, axis elements as part of the data visualization.

What charts say

> *By being more explicit in our own understanding of what charts say and how we can systematically describe what they say, we can grow more capable of using the channels available in that expression to our advantage.*

(Meeks 2018)

What does your chart say that you didn't intend?

What charts mean

Charts mean more than just what they say.

What charts mean

Intentionally

> *The mode and purpose of a chart should be well understood by the chart maker and immediately apparent to the chart reader.*

(Meeks 2018)

What charts mean

Historically

Charts are products of their time.

It is important to provide background about the data sources, to enable checking whether they are still based on relevant priorities, dimensions and metrics.

What charts mean

Historically

Obesity chart

What charts mean

Culturally

Charts should be adapted to the culture they will be consumed in (think user-centered design techniques).

What charts mean

Contextually

Enable removing and adjusting data visualization elements to reduce complexity, not based on screen size as in responsive data visualization, but on priority.

What charts mean

> *Meaning-making may sound too soft to the kind of technical professionals that make and read data visualization but communication without meaning is just noise.*

(Meeks 2018)

What charts do

The most important thing about a chart is its impact.

What charts do

Provide insights

Identify and emphasize the insights that the readers might expect

- simple line chart and highlighted features of the line chart

simple line chart and highlighted features of the line chart

What charts do

Cause change

As difficult to measure as it is important.

How have they impacted business decisions? How were they used in presentations? Where they modified (changed colours, cropped, annotated) somehow?

What charts do

Cause visual literacy

> *All data visualization was, at some point, complex data visualization, until an audience grew comfortable and literate enough to read it.*

(Meeks 2018)

(Netflix connected scatterplot chart?)

What charts do

Create new charts

- Imaginary genealogy for charts with scatterplots as a common ancestor (Meeks 2018)

Imaginary genealogy for charts with scatterplots as a common ancestor (Meeks 2018)

What charts do

> *All communication is evaluated based on content, but persuasive communication, which is all data visualization unless it is purely decorative, is rightly also evaluated based on effect.*

(Meeks 2018)

3. The artefact goes social

Data counseling

> *[…] brings domain expertise into the operationalization process to help inform decisions about good proxies as well as to uncover insights using the resulting visualizations.*

(Meyer & Fisher 2018)

Data counseling

Based on **interviews** (1) for

- gaining an understanding of the questions and data
- get feedback on proxies, **explorations** (2), and visualization **prototypes** (3)

Data counseling

1. Interviews

> *The role of the interviewer is to ask questions that will guide the stakeholders toward elucidating the information necessary for working through an operationalization process and designing visualizations.*

(Meyer & Fisher 2018)

Data counseling

1. Interviews

Identify stakeholders:

- analysts
- data producers
- gatekeepers
- decision makers
- connectors

Data counseling

1. Interviews

Require practice and experience.

Semistructured: be prepared, but also be open.

- start with open ended questions (problem, data, context)
- use the conversation to search out the more abstract question

Data counseling

1. Interviews

Use traditional conversation / interpersonal communication skills to prevent dead ends: keep them talking

- rephrase responses back to the stakeholder
- ask the same or similar questions in different ways
- explore a completely different conversational topic

Data counseling

1. Interviews

Contextual interviews

- take place in the stakeholder's work environment
- consist of demonstrations of the tools and data inspection methods currently in use

Data counseling

2. Data exploration

Look to the data as early as possible.

???

Data counseling

3. Rapid prototyping

> *[…] is a process of trying out many visualization ideas as quickly as possible and getting feedback from stakeholders on their efficacy.*

(Meyer & Fisher 2018)

Data counseling

3. Rapid prototyping

≠ *fast* data visualization

≈ agile/lean methodologies and user-centered design

Data counseling

3. Rapid prototyping

- Prototypes range from low-fidelity sketches to high-fidelity working models (Meyer & Fisher 2018)

Prototypes range from low-fidelity sketches to high-fidelity working models (Meyer & Fisher 2018)

Data counseling

3. Rapid prototyping

Prototypes are made to obtain feedback on them: get to the stakeholders early and often.

Focus not on whether they like it or not, but rather on **what the visualization can and cannot do** (contextual interview where the stakeholder uses the visualization).

Responsive data visualization

> *Responsive web design, and responsive data visualization are not simply a way to make our content accessible on smaller screens. We need to build an ergonomic web that feels natural regardless of device type.*

(Hinderman 2018)

Responsive data visualization

Unknowns require adaptability.

- the context in which **the user** is trying to consume the visualization
- changes in **the data** that is being displayed

Responsive data visualization

Output side (the client)

Making things work in all screen types by redrawing charts to fit its container.

Match CSS breakpoints + add any new ones as the content requires: group data to fit (trade-off precision for reduced rendering complexity and performance).

Responsive data visualization

Input side (the data)

Adapting at breakpoints. No need to just redraw the exact same elements:

> *As long as the message being conveyed by the data is the same, and the point you're trying to prove is always present, you should prove it with as much firepower as you have available.*

(Hinderman 2018, p.361)

Responsive data visualization

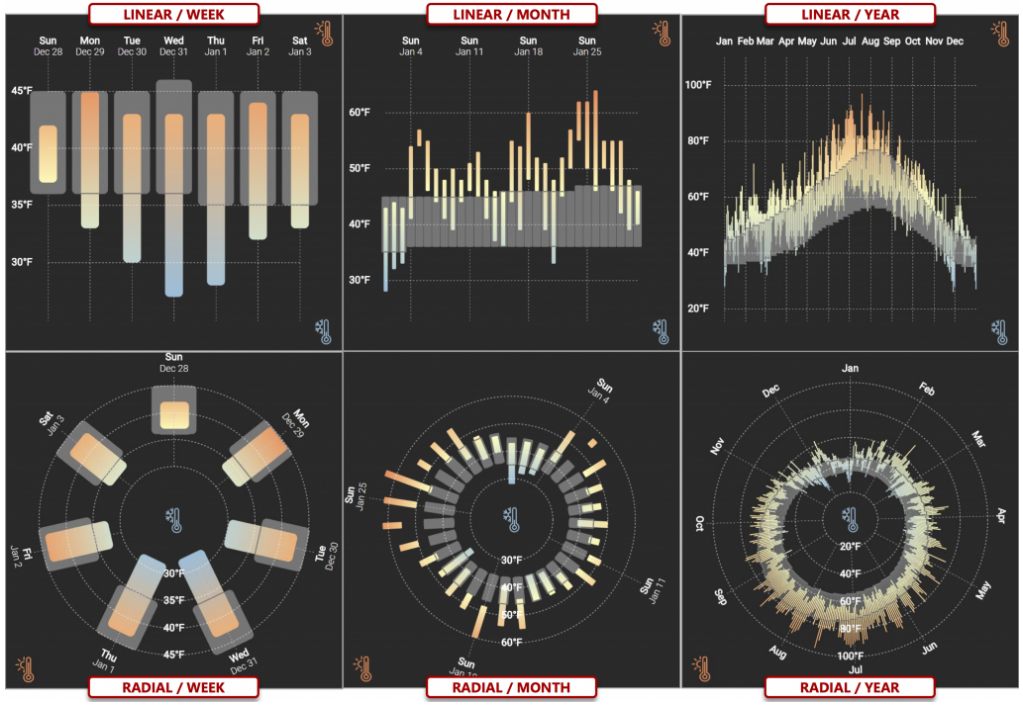
Input side (the data)

Adapting at interaction points.

> *[…] present a rational default but enable users to dig into more complex or specific layers of data when the device's capabilities limit the presentation of both at the same time.*

(Hinderman 2018, p.362)

Responsive data visualization



Linear and radial temperature range charts designed for mobile phone displays (Brehmer *et al.* 2019)

Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches

Responsive data visualization

The other “side” of adaptation is on the side of the user.

Data served

Responsive data visualization

Device dependant

Responsive data visualization

Alerts

Example: sports tracker on watch vs. phone vs. computer

Epilogue

Resources

http://oavis.steveharoz.com/

http://ieeevis.org/year/2018/welcome

References

Ihor Kovalyshyn (2017), *When Scatter Plot Doesn't Work*

Dominik Moritz and Danyel Fisher (2018), "Visualizing a Million Time Serieswith the Density Line Chart" arXiv:1808.06019v2 [cs.HC]

Andy Kirk 2014 Null zero blank

François Michonneau & Auriel Fournier (2018), "Data Carpentry: R for data analysis and visualization of Ecological Data." doi: 10.5281/zenodo.569338, http://datacarpentry.org/R-ecology-lesson/

Dan Meth (2009), "The Trilogy Meter" https://danmeth.myportfolio.com//post/77471620/my-trilogy-meter-1-in-a-series-of-pop-cultural

Anthony Aragues (2018), *Visualizing Streaming Data*. O'Reilly Media

Miriah Meyer & Danyel Fisher (2018), *Making Data Visual*. O'Reilly Media

Tanja Blascheck, Lonni Besançon, Anastasia Bezerianos, Bongshin Lee, Petra Isenberg (2019). "Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches". *IEEE Transactions on Visualization and Computer Graphics* 25(1)

Matthew Brehmer, Bongshin Lee, Petra Isenberg, Eun Kyoung Choe (2019). "Visualizing Ranges over Time on Mobile Phones: A Task-Based Crowdsourced Evaluation". *IEEE Transactions on Visualization and Computer Graphics* 25(1)

Thank you!

@DeustoKomunika @DeustoData

@mberasategi @neregauzak