

Reproducible Document Stack

Giuliano Maciocchi
Head of Product @ eLife

About eLife

@eLifeInnovation

innovation@elifesciences.org



MAX-PLANCK-GESELLSCHAFT

*Knut and Alice
Wallenberg
Foundation*

eLife is a non-profit organisation inspired by research funders and led by scientists



eLIFE

Helping scientists accelerate discovery by
operating a platform for research communication
that encourages and recognises the most
responsible behaviours in science

What do we mean by “responsible behaviours”?

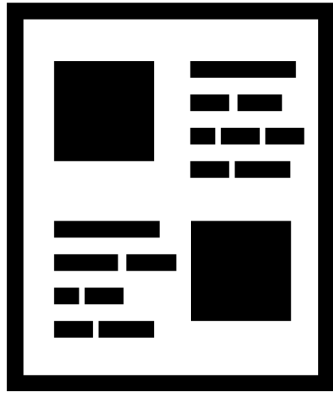
- Sharing of data, tools, and resources
- Objective and comprehensive reporting
- Cooperation and collaboration
- Constructive feedback and encouragement

eLife Innovation Initiative

We invest in open source technologies, tools and processes that improve the way cutting-edge research is discovered, shared, consumed and evaluated

Publishing reproducible research documents

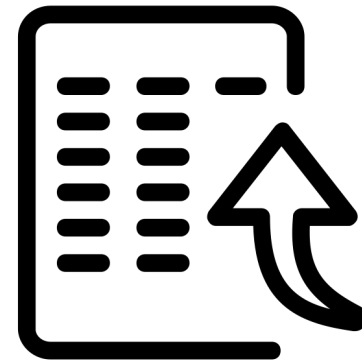
Research publishing today



Created by Dmitry Podluzny
from Noun Project

[CC-BY 3.0](#)

+



Created by IconfactoryTeam
from Noun Project

[CC-BY 3.0](#)



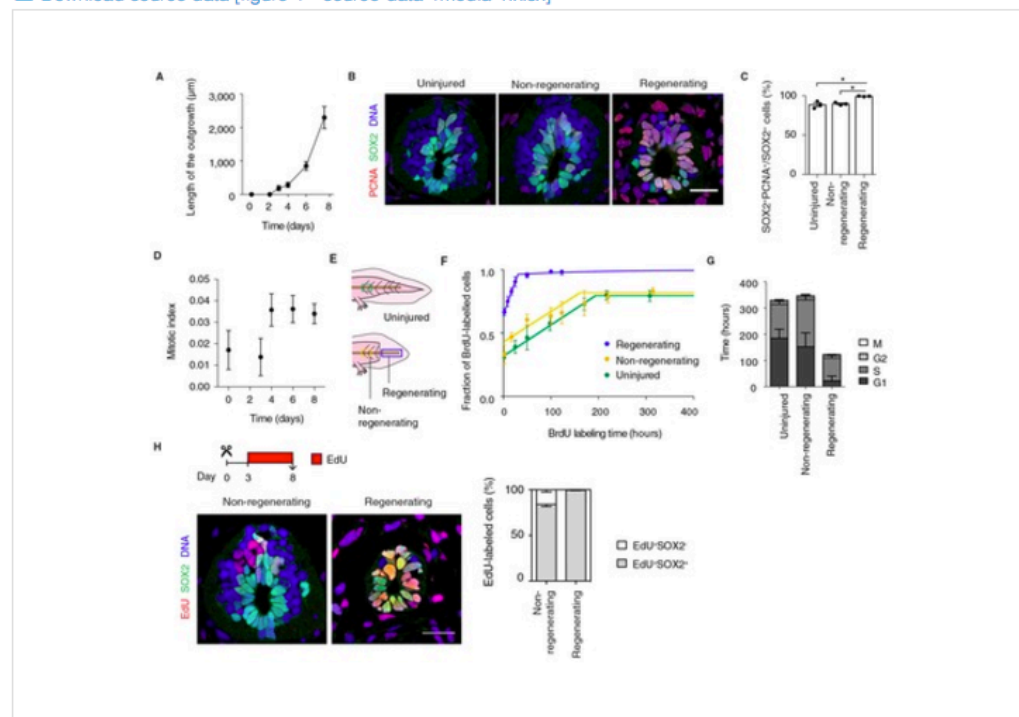
Complex datasets, and sometimes the code used to analyse them, are ultimately reduced to little more than PNG images embedded in an article.

Figure 1—source data 1

Gene expression changes in the regenerating axolotl spinal cord compared to the uninjured axolotl spinal cord.

DOI: <http://dx.doi.org/10.7554/eLife.10230.004>

[Download source data \[figure-1—source-data-1media-1.xlsx\]](#)

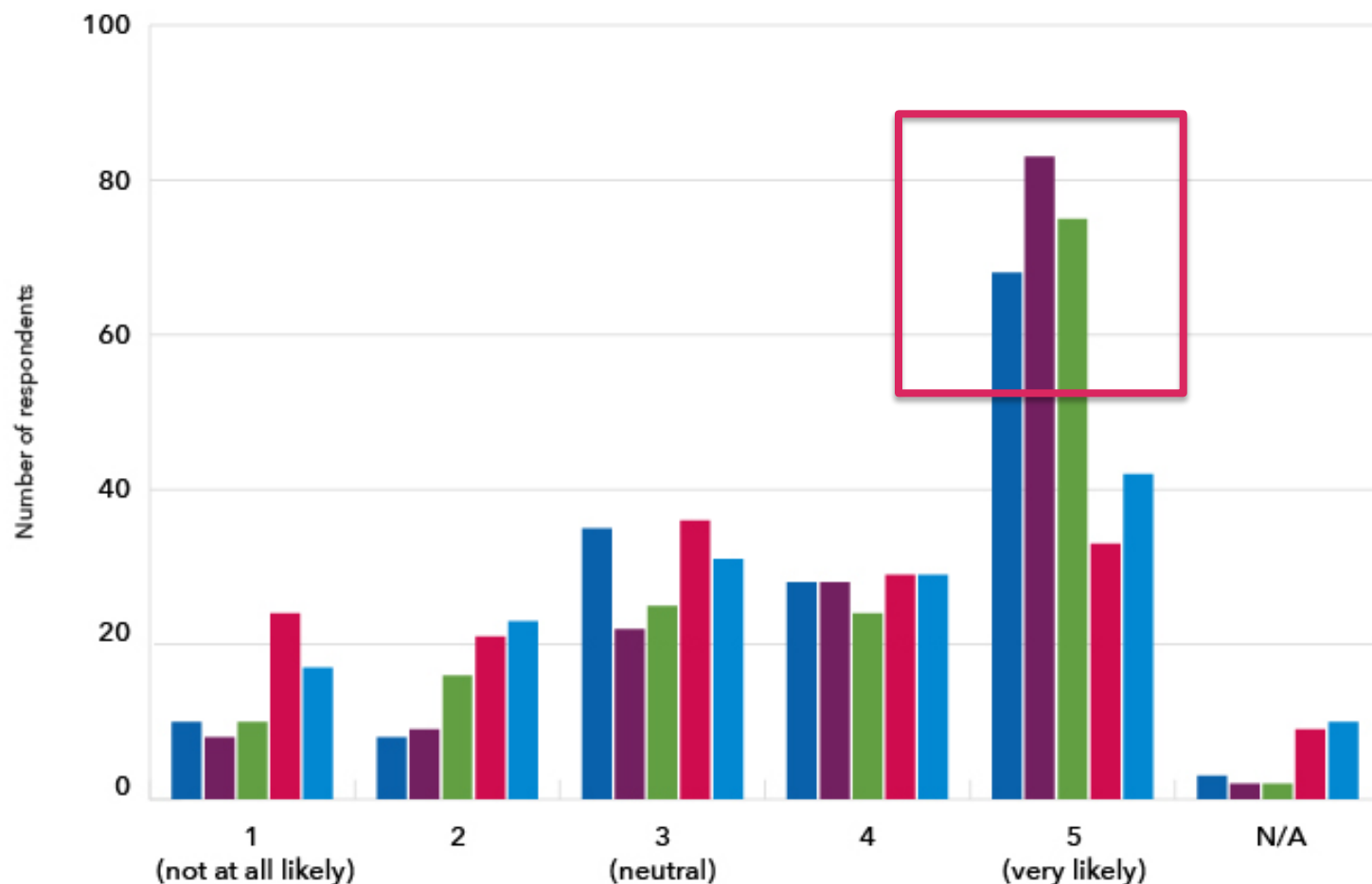


populations as in Figure 2E. Error bars, mean \pm s.d. of $n \geq 5$ tails/time-point. (G) Cell cycle parameters of uninjured, non-regenerating, and regenerating neural stem cell populations, calculated from data in D and F. Error bars, mean \pm 1 σ confidence intervals. Data can be found in [Supplementary file 1](#) and <http://nbviewer.jupyter.org/gist/fabianrost84/3cc58a27b5688f4e2eba>. (H) EdU was injected daily from day 3 to day 8 of regeneration, when tails were collected for analysis. Representative images of non-regenerating and regenerating regions of the spinal cord. SOX2 labels neural stem cells, EdU labels cells that underwent DNA synthesis and their progeny, and Hoechst labels DNA. Quantification of the percentage of EdU+ cells that remain as neural stem cells (SOX2+) or differentiate (SOX2-) over total EdU+ cells after the six-day chase. Error bars, mean \pm s.e.m. of $n=4$ tails/region. Scale bar, 50 μ m. A supplementary IPython notebook (Pérez and Granger, 2007) containing all the raw data and the code used for the estimations of mitotic index, cell cycle length and cell cycle parameters is available at <http://nbviewer.jupyter.org/gist/fabianrost84/3cc58a27b5688f4e2eba>.

DOI: <http://dx.doi.org/10.7554/eLife.10230.005>

Survey of life scientists, June 2017 – Likelihood of using a feature if provided (n=152, all respondents)

<https://elifesciences.org/inside-elife/e832444e/innovation-understanding-the-demand-for-reproducible-research-articles>

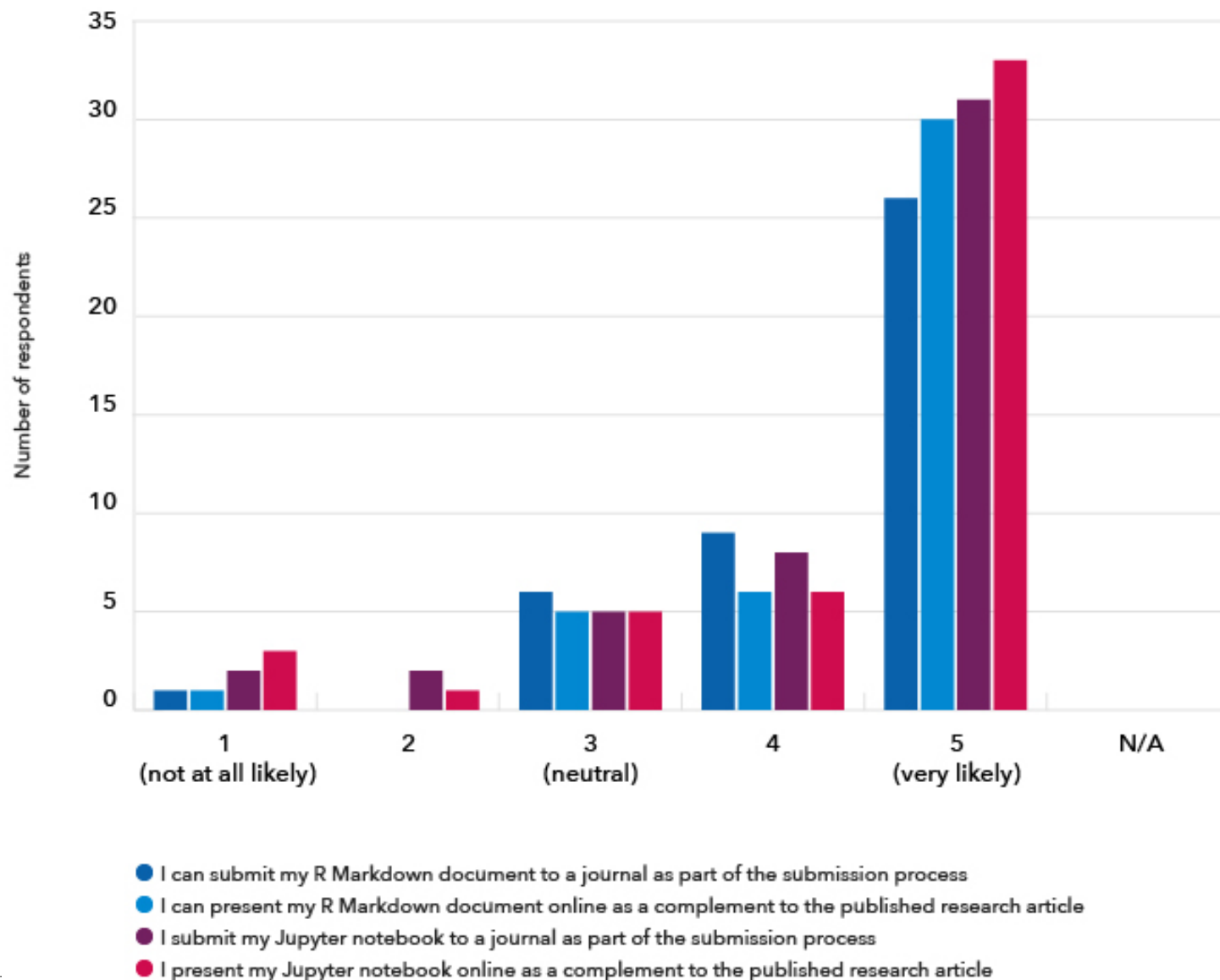


- I submit my research manuscript in a file format that preserves code and data alongside the text of a manuscript
- I view and download the data and code associated with a research article directly from the online publication
- I interact online with the data, code and/or plots in a published research article to gain additional insight
- I run the code directly from an online published research article to reproduce the author's results
- I modify the code directly from an online published research article and run my modified version to produce a new result



Survey of life scientists, June 2017 – Likelihood of using features specific to R Markdown (42 respondents) or Jupyter notebooks (48 respondents), were they to be made available

<https://elifesciences.org/inside-elif/e832444e/innovation-understanding-the-demand-for-reproducible-research-articles>



Opportunities

- Facilitating sharing **via the journal** means:
 - Persistence
 - Leverages incentive system
 - Streamline the process from initial discovery to replication
 - Engages community



Opportunities

- Doing it **open source** means:
 - Progress can be shared
 - Stimulates further innovation
 - Engages community

Our vision: Reproducible Documents

- Encapsulates usable code and data within the flow of a manuscript.
- Delivers progressive enhancement from static research article, to full data and code interaction
- Platform, tool, language agnostic
- Easy and accessible for everyone
- Encourage reuse of published research

Reproducible Document Stack project

- Authoring platform
- Reproducible Document Container
- Publisher tools for Reproducible documents



RDS = Manuscript + Data + Methods



RDS: Authoring tool

- Initial prototype based on stenci.la authoring tools
- User friendly, clean text editing
- Executable code blocks
- Support for R, Python
- Plot generation
- “Mini” formula language for excel-like graphing

Paragraph ▾ Insert ▾ View ▾

package). But for now, there is just one function for that: `filter`, which takes as input a data table and an expression describing the filtering criteria (for now the expression is a string, but in the future it will be a Mini lambda expression).



```
filter(data, 'sepal_length < 4.5 & petal_width >= 0.2')
```

species	sepal_length	sepal_width	petal_length	petal_width
setosa	4.4	2.9	1.4	0.2
setosa	4.4	3	1.3	0.2
setosa	4.4	3.2	1.3	0.2

Plotting data

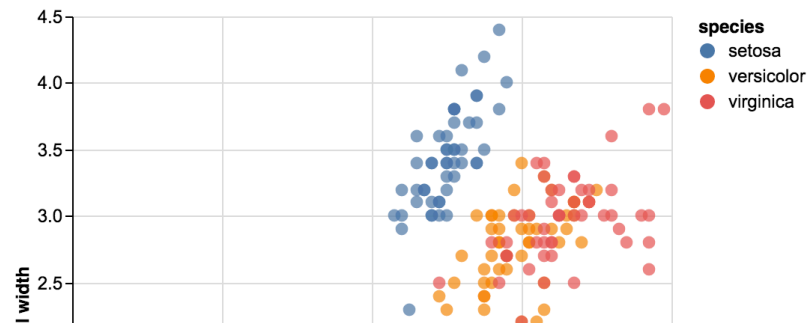
At present, a few functions are available for creating simple plots. The implementation of these is currently based on Vega-Lite, a high level grammar for visualization rendered using D3.js.

The core `plot` function takes a data table, a mark name e.g. points, bars, and the names of columns in the data to encode into various plotting channels e.g. x, y, size, color. Let's create a simple `scatterplot` of our data:



I **B** x_2 x^2 `</>`

```
points(data, 'sepal_length', 'sepal_width', color='species')
```



RDS: Reproducible Document Container

- Container for text, code, data and media assets
- Standard JATS XML based data formats
- Open format

RDS: Publisher tools

- Journal submission infrastructure integration
- Quick export to PDF for legacy systems
- Web-based publishing of fully reproducible documents with in-browser code interaction and execution

Reproducible documents in practice

Authoring tool supports 'dynamic figures'

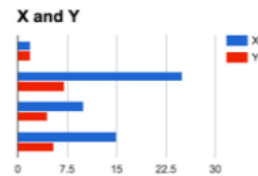
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam vel libero sit amet dolor mollis aliquet. Donec gravida massa at d

Etiam fermentum sodales luctus. Donec potenti. Sed facilis

purus bibendum. Sed

ras commodo ex et justo
attis tellus. Suspendisse
s in sem mollis mollis.

Select a plot to create a new figure



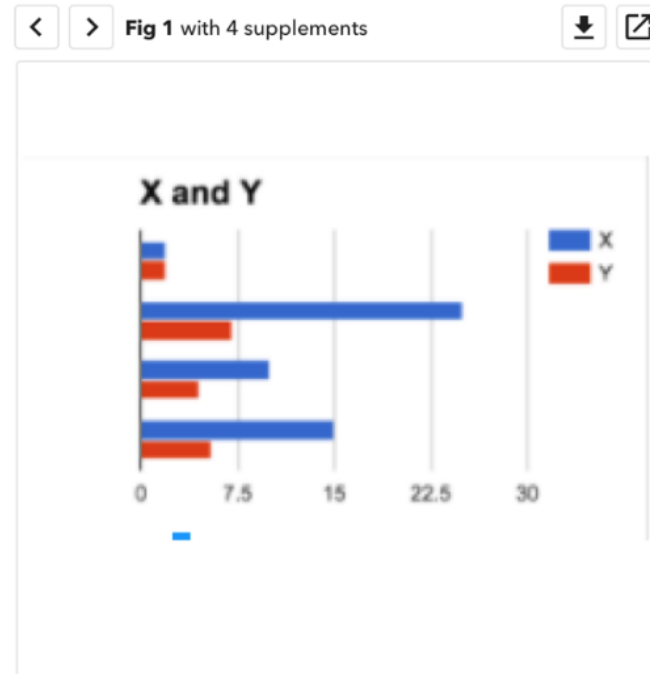
Dynamic figures are generated by code/data

X	Y
2	2
25	7.071067812
10	4.472135955
15	5.477225575

Dynamic figures preserved in published article

communicate with the secondary neurons. Important mutations that blocked this chemical signaling prevented worms from responding appropriately to the smell of:

Aging animals, including people, often have impaired olfactory function and can therefore find it difficult to identify and respond to odors. In a study, Leinwand et al. found that aged worms were no different from young worms in their response to odors. Experiments suggested that aging worms' responses to odors decline because the communication between the primary and secondary olfactory neurons may be impaired with age. Leinwand et al. strengthened this communication to reverse the effects of aging on the worms' sense of smell. Moreover, experiments also showed that an animal's performance on an odor task was correlated with its longevity, such that top performers also lived longer. A challenge for the future is to understand the precise changes that occur at early stages that lead to impair the sense of smell. Future studies could also investigate whether similar combinations of olfactory neurons are needed to elicit certain behavioral responses to odors in young and old animals.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sed malesuada ex.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sed malesuada ex. Suspendisse efficitur metus ac nisl tempus... [more](#)

Code/data still interactive in published figure

communicate with the secondary neurons. Important mutations that blocked this chemical signaling prevented worms from responding appropriately to the smell of :

Aging animals, including people, often have impaired can therefore find it difficult to identify and respond to. Leinwand et al. found that aged worms were no different. experiments suggested that aging worms' responses to decline because the communication between the primary secondary olfactory neurons may be impaired with age. Leinwand et al. strengthened this communication to reverse effects of aging on the worms' sense of smell. Moreover experiments also showed that an animal's performance on odor task was correlated with its longevity, such that top performers also lived longer. A challenge for the future is to understand the precise changes that occur at early stages to impair the sense of smell. Future studies could also test similar combinations of olfactory neurons are needed to elicit certain behavioral responses to odors in young and old

< > Fig 1 with 4 supplements

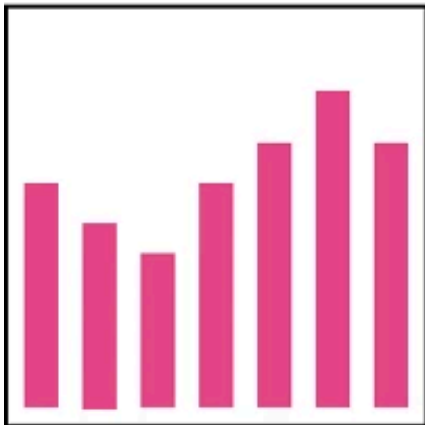
X	Y
2	2
25	7.071067812
10	4.472135955
15	5.477225575

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sed malesuada ex.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sed malesuada ex. Suspendisse efficitur metus ac nisl tempus... [more](#)

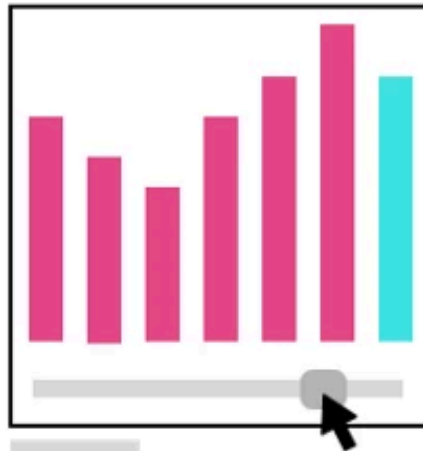
Progressive enhancement via multi-level output

1.



Plain .PNG
(e.g., mobile)

2.



Interactive

3.



Code / Data view
(edit, execute)

4.



Core principles

- Open: not trying to “win” a tools race
- Future-proof: extensible document exchange format
- Usable: should not add undue burden to time-strapped researchers
- Portable: minimise or eliminate dependencies for reuse



eLIFE



substance



Stencila



More info

- Announcement:
<https://elifesciences.org/for-the-press/e6038800/elif-supports-development-of-open-technology-stack-for-publishing-reproducible-manuscripts-online>
- About the project:
<https://elifesciences.org/labs/7dbeb390/reproducible-document-stack-supporting-the-next-generation-research-article>
- June 2017 survey results:
<https://elifesciences.org/inside-elif/e832444e/innovation-understanding-the-demand-for-reproducible-research-articles>



Thank you

Want to find out more? Join our webinar 25/01 @ 3pm
elifesci.org/reprodoc-update

g.maciocci@elifesciences.org