

The Beginner's Guide to Dimensionality Reduction

Explore the methods that data scientists use to visualize high-dimensional data.

By: [Matthew Conlen](#) and [Fred Hohman](#)

Workshop on Visualization for AI Explainability
October 22, 2018

The Beginner's Guide to Dimensionality Reduction

Explore the methods that data scientists use to visualize high-dimensional data.

By: [Matthew Conlen](#) and [Fred Hohman](#)

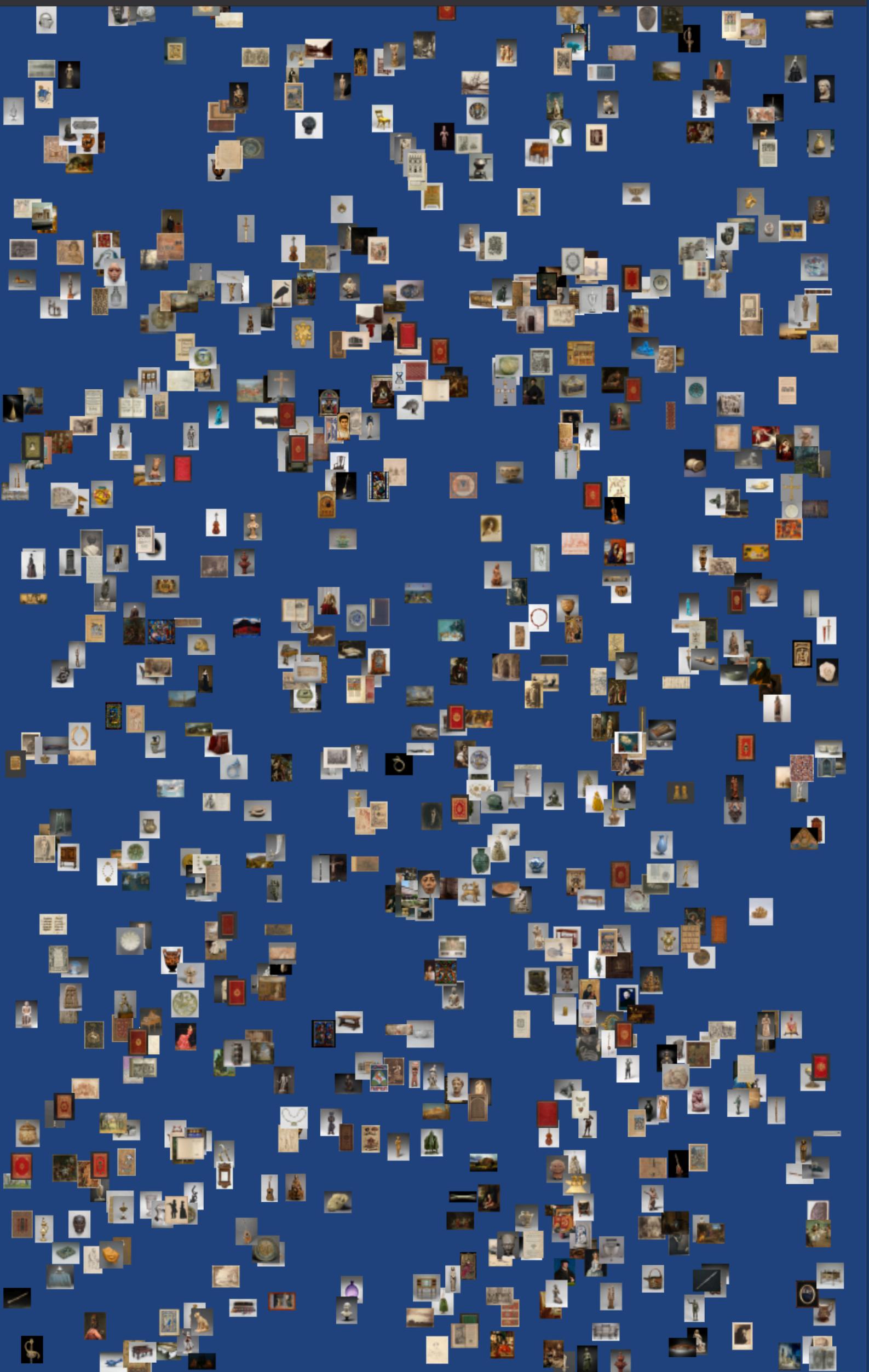
July 16, 2018

Dimensionality reduction is a powerful technique used by data scientists to look for hidden structure in data. The method is useful in a number of domains, for example document categorization, protein disorder prediction, and machine learning model debugging^[2].

The results of a dimensionality reduction algorithm can be visualized to reveal patterns and clusters of similar or dissimilar data. Even though the data is displayed in only two or three dimensions, structures present in higher dimensions are maintained, at least roughly^[7].

The technique is available in many applications, for example Google's [Embedding Projector](#)^[10] let's you view high-dimensional datasets embedded in two or three dimensions under a variety of different projections.

This guide will teach you how to think about these embeddings, and provide a comparison of some of the most popular dimensionality reduction algorithms used today.



Who it Reached

Matthew Conlen

@mathisonian

Following

The Beginner's Guide to Dimensionality Reduction

New interactive post by @fredhohman and me! Explore methods that data scientists use to visualize high-dimensional data.

The Beginner's Guide to Dimensionality Reduction

Explore the methods that data scientists use to visualize high-dimensional data.

By Matthew Conlen and Fred Hohman

July 12, 2018

The Beginner's Guide to Dimensionality Reduction

Explore the methods that data scientists use to visualize high-dimensional data in this interactive article.

idyll.pub

12:36 PM - 17 Jul 2018

98 Retweets 297 Likes

4 98 297

Released online
VISxAI community
Twitter
Reddit (r/datascience)
DataCamp.com

Audience
7k visits from 4.5k “readers”
Practicing data scientists,
researchers, and students

How People Engaged

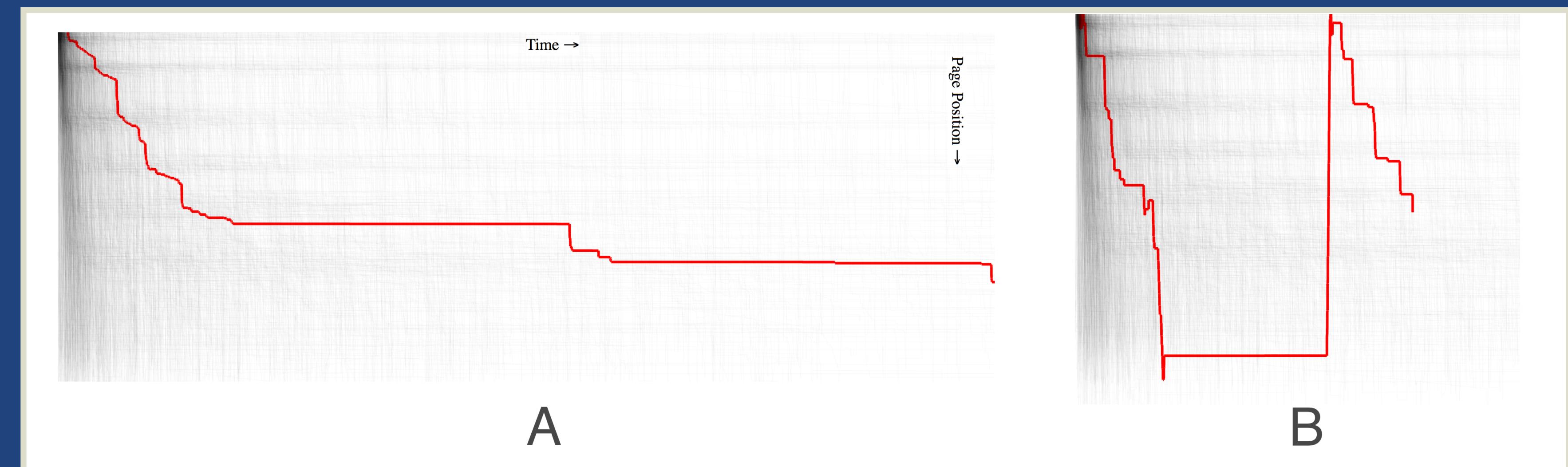
Overall

*76% finish the article

Long tail engagement distributions across most interactions

Most visitors skim, 6% read through multiple times

Scroll Patterns



**desktop, mobile coming soon*

Feedback



[arcane_neptune](#) 2 points · 3 months ago

Thank you for this. I'm trying to learn PCA for undergrad research.

[Reply](#) [Share](#) [Report](#) [Save](#)

[Omnislip](#) 14 points · 3 months ago

Just note that this is the very "dangerous" side to data science, where only the output of various methods are described, and not the underlying methods themselves!

[Reply](#) [Share](#) [Report](#) [Save](#)

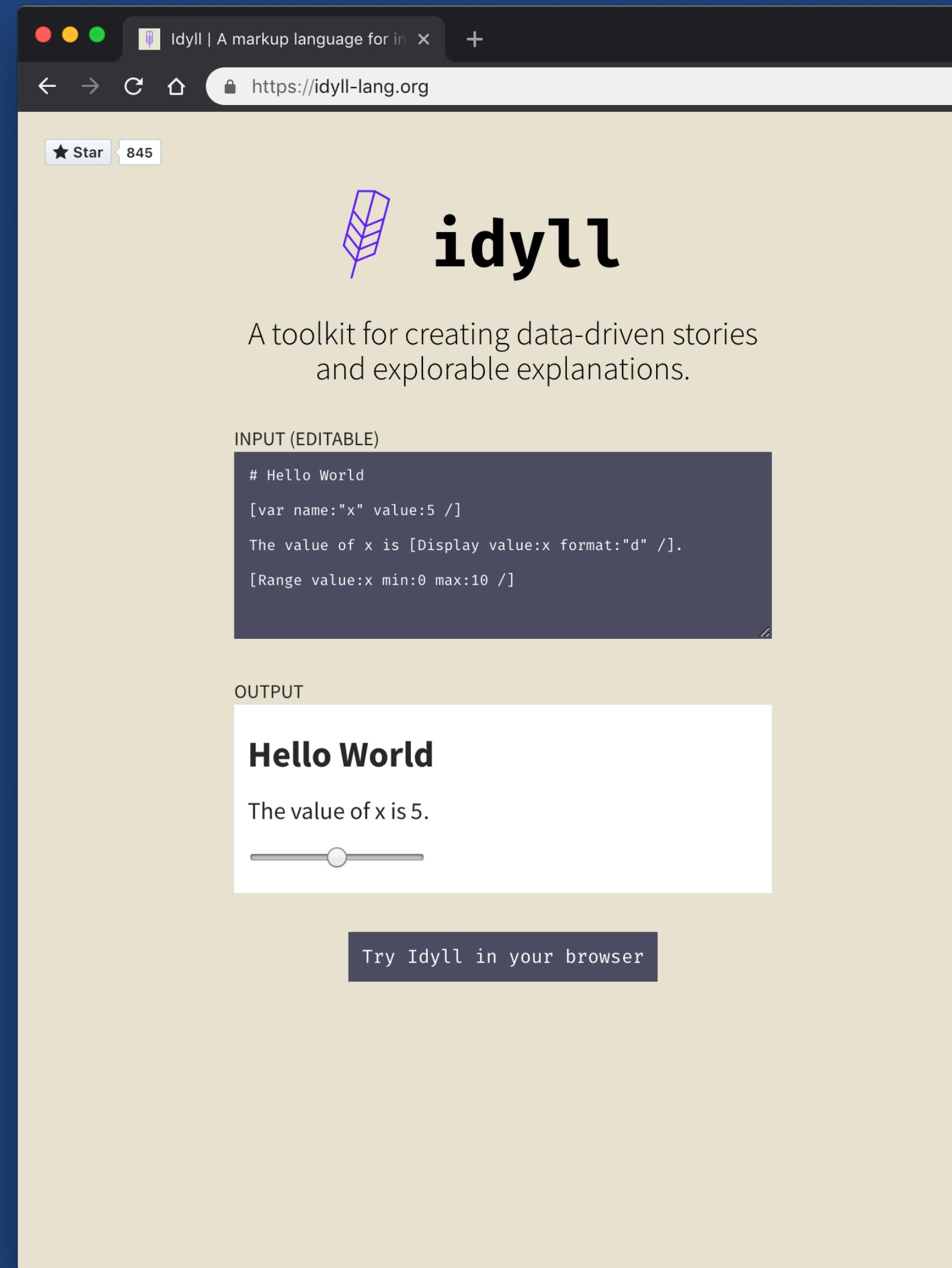
[funnythingaboutmybak](#) [MS | Data Scientist | Energy](#) 14 points · 3 months ago

OP had a good intro to whet the appetite. If you want to learn more about what PCA is and how to implement it (including Python code), [check this out](#).

[Reply](#) [Share](#) [Report](#) [Save](#)

How We Made It

Idyll
<https://idyll-lang.org/>



The screenshot shows the Idyll website at <https://idyll-lang.org>. On the left, the homepage features the Idyll logo and a brief description: "A toolkit for creating data-driven stories and exploratory explanations." Below this is an "INPUT (EDITABLE)" section containing the following code:

```
# Hello World
[var name:"x" value:5 /]
The value of x is [Display value:x format:"d" /].
[Range value:x min:0 max:10 /]
```

On the right, the "OUTPUT" section displays the result: "Hello World" and "The value of x is 5." with a slider control. A button at the bottom says "Try Idyll in your browser".

The right side of the page shows a complex, interactive visualization featuring a map of the United States with various data overlays, including a network graph and a grid pattern. A callout box points to a link labeled "View Example Gallery". Below the visualization, there is a summary text and navigation links.

Idyll is an open-source markup language and web runtime. You write markup and Idyll converts it to interactive code that can run in anyone's web browser. Idyll extends *Markdown* with a reactive component system.

Idyll allows non-experts to publish compelling interactive stories on the web, and enables collaboration between programmers and journalists, researchers and designers. Those familiar with JavaScript can write custom components using tools like D3 or React.

idl **EUTOPIA**
W PAUL G. ALLEN SCHOOL
RHIZOME

Idyll is supported by the Interactive Data Lab at the University of Washington, and by Rhizome and The Eutopia Foundation.

The Beginner's Guide to Dimensionality Reduction

Explore the methods that data scientists use
to visualize high-dimensional data.

<https://github.com/mathisonian/dimensionality-reduction>



Matthew Conlen
University of Washington
@mathisonian



Fred Hohman
Georgia Tech
@fredhohman

Acknowledgments

Jeffrey Heer, Polo Chau, Caleb Robinson, Nicky Case, and Hamish Todd

