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**

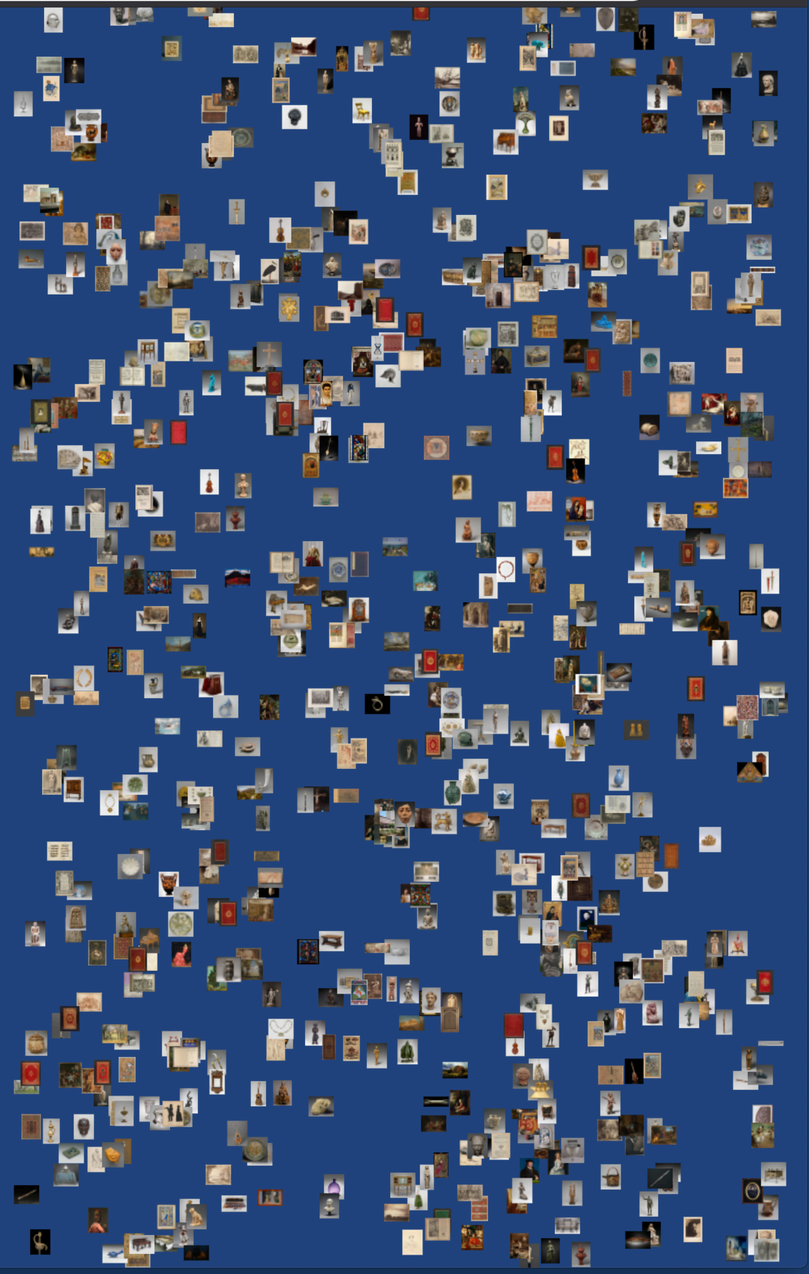
Workshop on Visualization for AI Explainability
October 22, 2018

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].

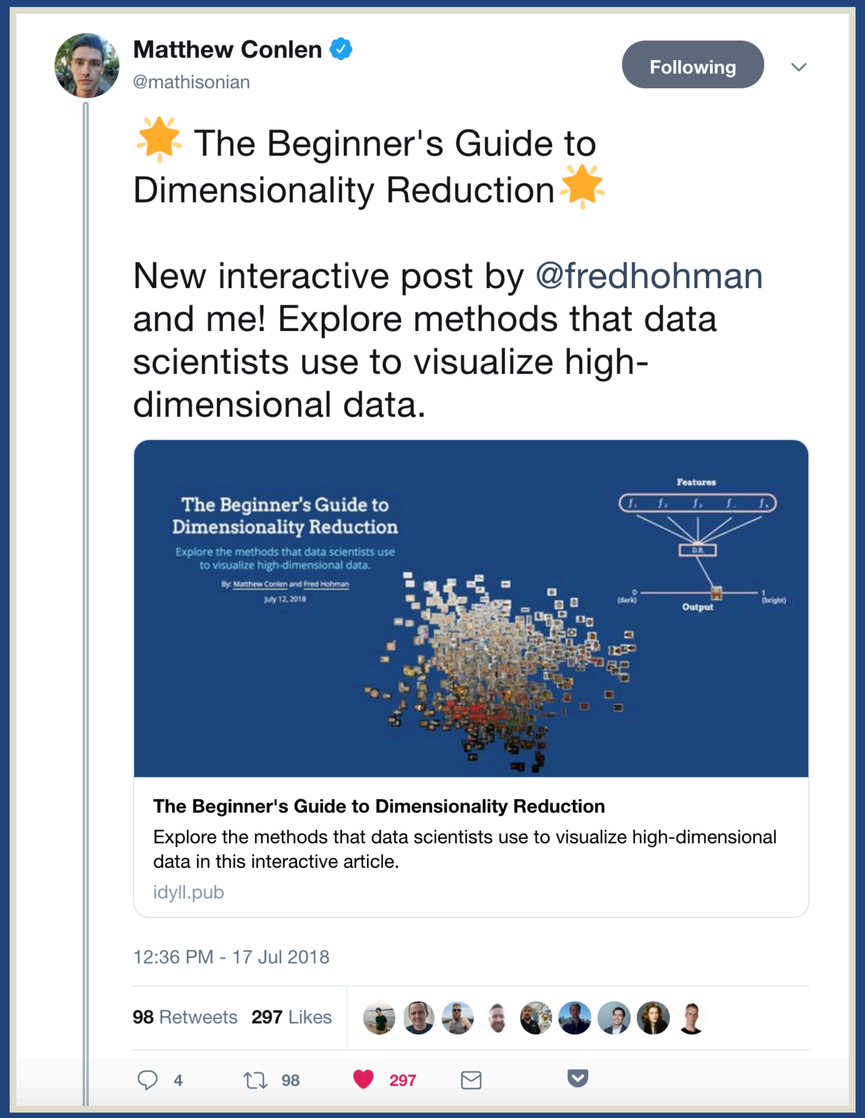
The results of a dimensionality reduction algorithm can be visualized to reveal patterns and clusters of similar or dissimilar data. Even though the data is displayed in only two or three dimensions, structures present in higher dimensions are maintained, at least roughly[7].

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.

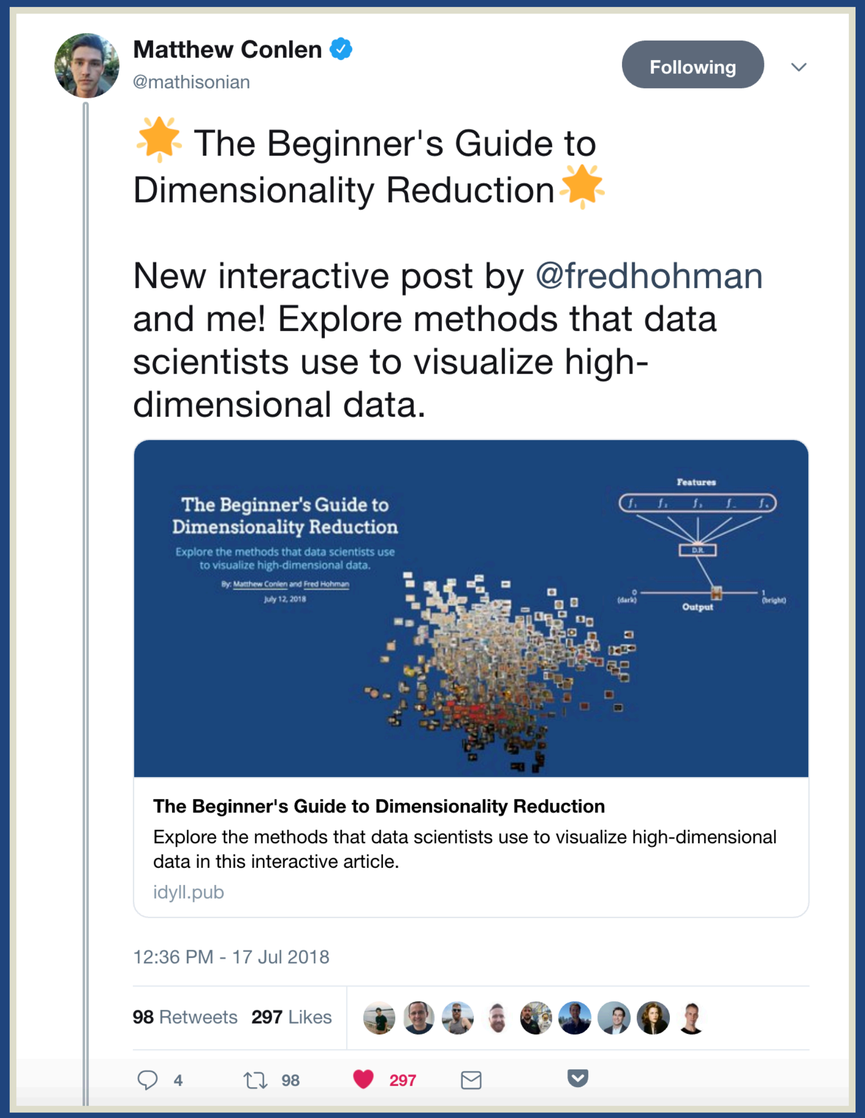
This guide will teach you how to think about these embeddings, and provide a comparison of some of the most popular dimensionality reduction algorithms used today.

Who it Reached

Who it Reached



Who it Reached



Released online

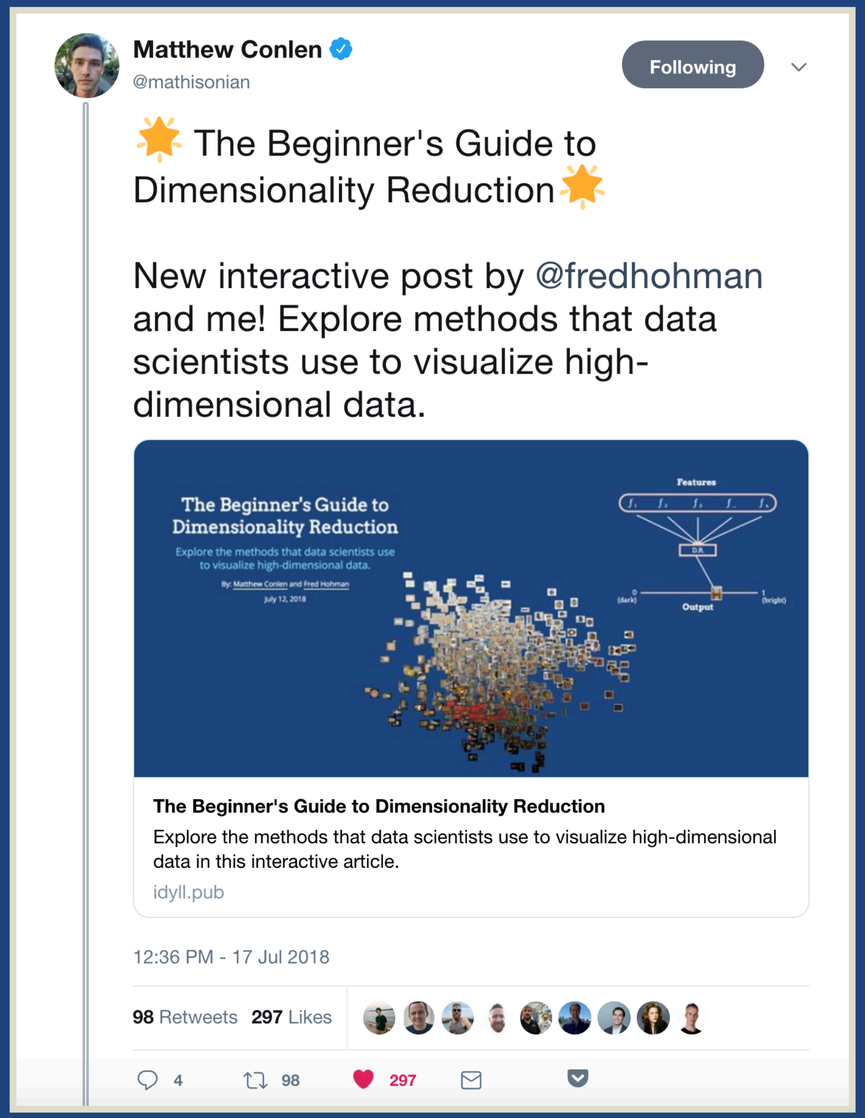
VISxAI community

Twitter

Reddit (r/datascience)

DataCamp.com

Who it Reached



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

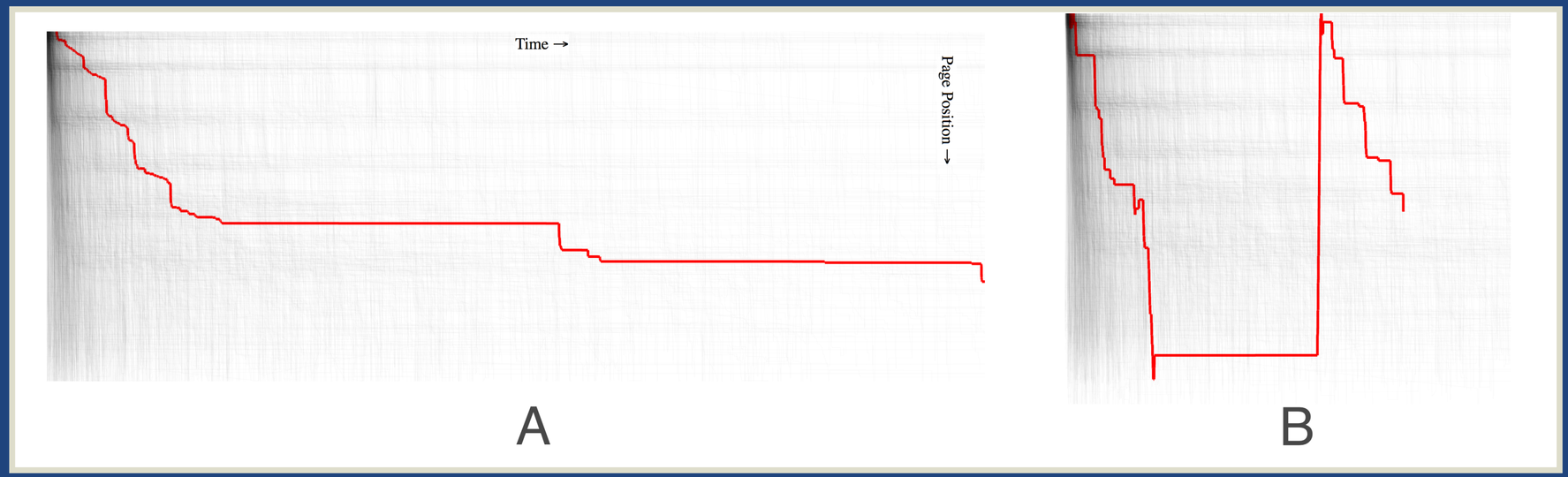
Overall

*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

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

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

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

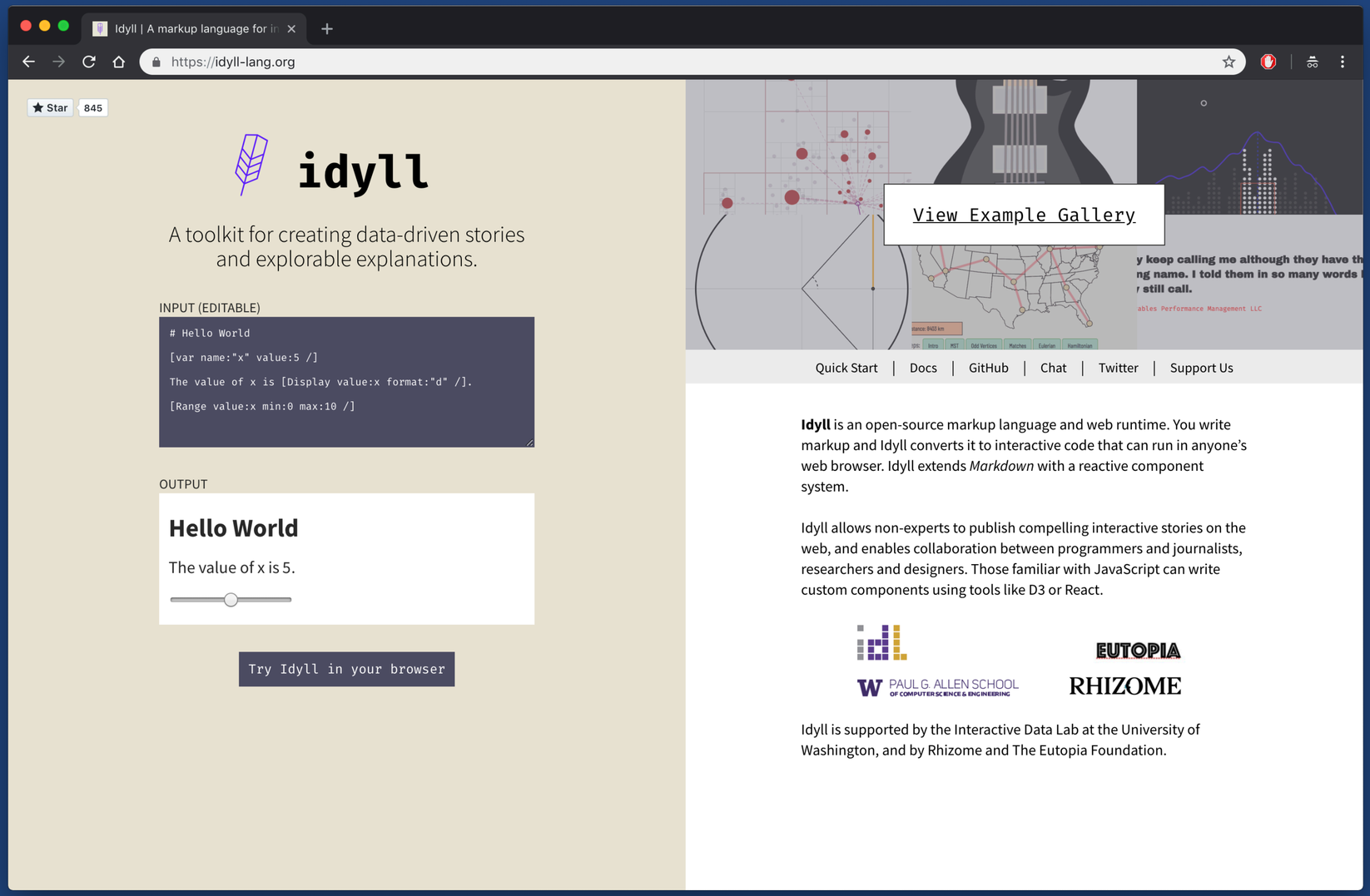
OP had a good intro to whet the appetite. If you want to learn more about what PCA is and how to implement it (including Python code), check this out.

Reply Share Report Save

How We Made It

Idyll

https://idyll-lang.org/



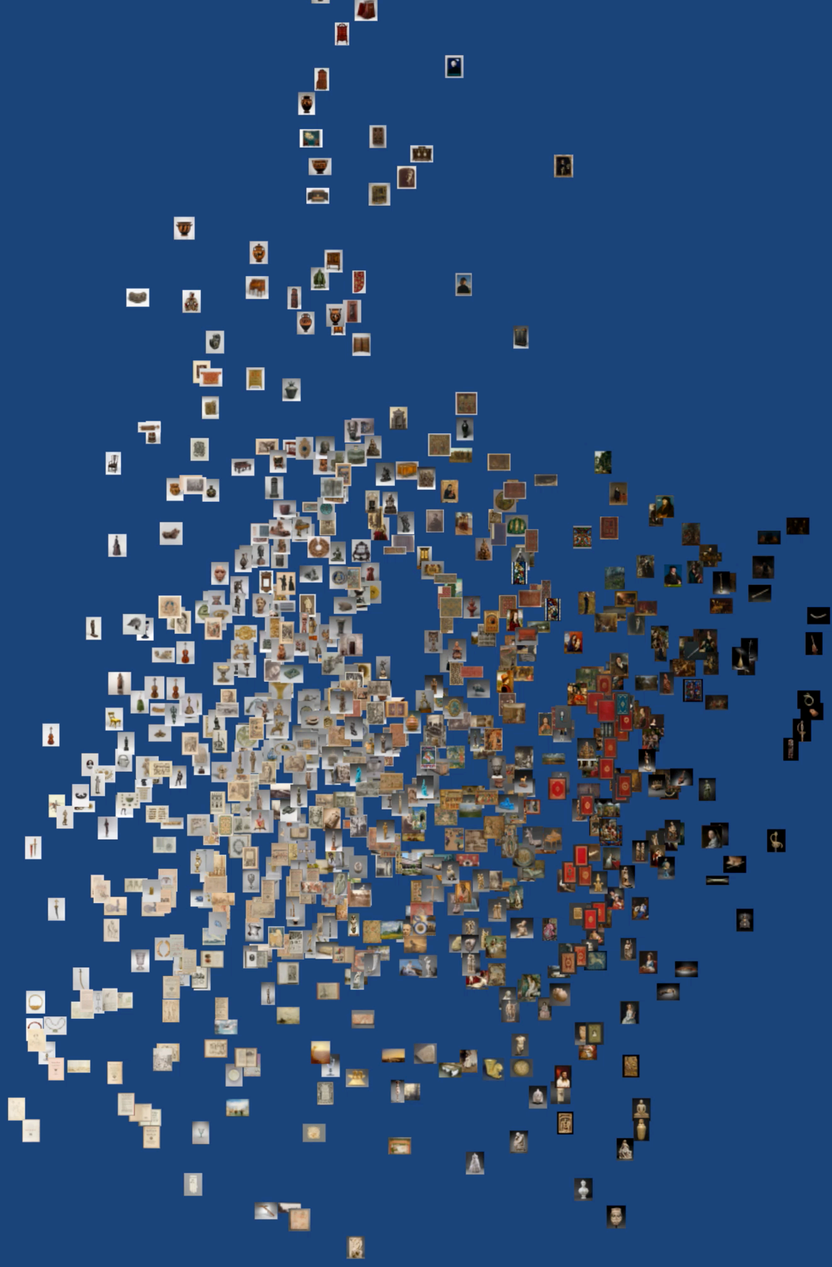
The Beginner's Guide to Dimensionality Reduction

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