

Clustering Works from the Cleveland Museum of Art

Sidney Miller, Hannah Li, Harshal Dahake, Aditya Tapshalkar, Kewal Khatiwala, William Braga

Summary

In this project, we sought to create a new visual tool to help view related pieces of art provided by the Cleveland Museum of Art (CMA). To do this, we can look at traits of each piece such as the title, creation date, culture, type, and technique

Our result was a visual depicting clusters of related art as determined by our model.

Clustering Interest Groups

The CMA artwork contains over 35000 pieces of art. While these can be conventionally ordered by a single criterion (time period, culture, etc.), our dataset contains many different criteria and subtleties.

For example, our data includes over 1800 years of art containing over 6000 art style techniques!

To best address this issue and match artwork to potential similar interest groups, we can use K-means unsupervised machine learning to cluster pieces of art into groups.

CMA Dataset

The Cleveland Museum of Art provided us with API data on their artwork. After cleaning, we found our dataset includes the following:

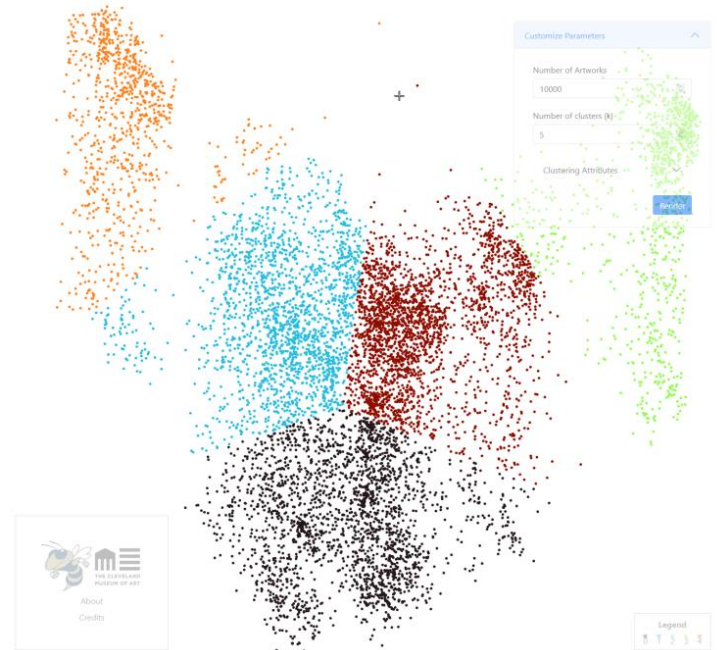
Earliest Date:	200 AD
Latest Date:	2000 AD
Number of Artworks:	35,035
Number of Artists:	4,128
Cultures Included:	5,381
Techniques:	6,647

The data collected included information on both the artwork and the artists themselves.

Artwork data included artwork titles, creation dates, culture/movement, painting types and techniques.

Artist data included names, roles, nationalities, and birth/death years.

Example of Visual



How does it work?

Our visual looks at some number of randomly selected pieces of art and groups it based on several criteria. The closer two art pieces are together, the more likely they are to be related

Under the hood, the clusters and 2D node coordinates are calculated in a three-stage process: encoding, feature reduction, and clustering.

In other words, we convert many text entries into a numerical format, convert these values to 2 dimensions (via PCA) and create clusters based on these results

Experiments and Results

Our visualization offers a whole new lens of analysis and comparison on CMA's collection. A user has the freedom to set the parameters for how they want the artwork to be clustered, explore and zoom in, and click on nodes to reveal an expanded description along with a picture.

While some other platforms may use a backend relational model, none of them depict these relationships on so many art pieces like ours does