Clustering Print Collections

Transmissions from the Front

DAMON CROCKETT, LEAD SCIENTIST, LENS MEDIA LAB

February 2023

If the gory details of clustering a small collection of photographs is not your idea of a fun half hour, this guide is not for you. If you’d rather skip ahead to the part where I tell you which prints fall into which clusters—tough luck, since I won’t do that here. This is a series of lessons about clustering print collections against the backdrop of a large reference collection. Although my case study is highly particular, I believe it to have much wider significance in data-driven cultural heritage research.

# Reference Collections and Art Collections

A collection of photographic prints can be called, in the general case, an *art collection*—for lack of a better term—and I take the lessons described herein to apply to any art collection whatever, although I will continue to use the terms “print” and “print collection”, because I work with photographic prints.

Collections of photographic prints come to us in different sizes, but none thus far have been smaller than ~50, and none have been larger than ~1000. For a data analyst, this is *small* data, despite what it may seem to the museum lifer. For data this small, the best methods are whichever seem the best to you, the expert, and good news—your data is so small that you’ll be able to check. Data analysis at this scale cannot tolerate obvious error, and obvious errors are manageable.