Data used:

* Artsy API Artist Database: <https://developers.artsy.net/>
* Documentation: <https://developers.artsy.net/v2/>
* Main CSV used for analysis: **artists\_2.csv**
* Original data dump from API: **ArtistsALL.json**

Notes:

* Cleaned nationality values (document in zip file outling my steps and original nationalities/changes) in main csv
* Cleaned gender (mostly capitalization/spelling errors) in main csv
* *'artwork\_names'* and *'similar\_artists\_names'* information downloaded into artists\_2.csv as best as I could from 'Downloading artworks troubelshooting.ipynb' and 'Downloading similar artists troubleshooting.ipynb' BOTH in zip file
* Decided to keep url within main csv to keep original data sources in one place
* *Final\_genes.txt* = list of all genes UNRELATED to geographic location or time (ex. 1910s, 1950s abstract, etc.)

Process:

1. Download data
   1. Ipynb with code:
      1. **Artsy API Call\_1**
2. Make CSV to use for analyses
   1. Ipynbs with code:
      1. **Overview of making Artists\_2.csv**
      2. **Downloading artworks into Df**
      3. **Downloading similar artists**
3. Descriptive statistics on data and distributions of various tags/data columns
   1. Ipynbs with code:
      1. **Descriptive Statistics and Distributions**
      2. **Gene Distribution**
      3. **Lifespan Distributions**
4. Find correlations and make some sort of network graph
   1. Ipynbs with code:
      1. **Gene-Artist correlations and Similarities Graph** (only examined possible correlations/networks between genes and artists and their shared genes, used cosine similarity)