

## Exercises for Data Visualisation for Analysis in Scholarly Research

Exercise 1: Exploring network visualisations.....	1
Exercise 2: Comparing N-gram tools .....	2
Exercise 3: Trying entity recognition .....	3
Exercise 4: Exploring scholarly data visualisations.....	3
Exercise 7: Applying data visualisation to your own work .....	7
<i>Explore more visualisations</i> .....	7
<i>Sketch ideas for visualisations</i> .....	7
<i>Try visualising data in different tools</i> .....	7

### Exercise 1: Exploring network visualisations

Time: approx 5 minutes.

Goal: hands-on experience with an interactive visualisation; practice critical analysis.

1. In your browser, go to <http://bit.ly/11qqXuj>
2. Scroll down the page to the network graph.
3. Take a few minutes to explore the visualisation: try holding the cursor over items, clicking, dragging, etc.
4. Discuss with your neighbour:
  - Is what the visualisation does clear?
  - Did using it feel intuitive?
  - Does interacting with the network graph give you more or less information than the other representations of the data further down on the same page?
  - Does the visualisation open up new questions about the data?

If you have time, try the opening of the Book of Genesis at

[http://texttexture.com/index.php?text\\_id=603](http://texttexture.com/index.php?text_id=603)

Other examples include 'Kindred Britain' <http://kindred.stanford.edu/> (click on People or Connections) or Les Misérables

<http://hci.stanford.edu/jheer/files/zoo/ex/networks/force.html>

Tools for network visualisations include Gephi <http://gephi.org/>

## Exercise 2: Comparing N-gram tools

Time: approx 5 minutes.

Goal: explore the impact of different tools and datasets on visualised results

1. Think of two words or phrases you'd like to compare over time (e.g. Burma, Burmah; different countries or languages of publication).
2. Open two browser windows
3. In one, go to <http://books.google.com/ngrams>
4. In the other, go to <http://benschmidt.org/OL/>
5. Enter your words or phrases in each and compare the results
6. Discuss with your neighbour: what differences did you find, and what might have caused them?

Google Ngram tips: <http://books.google.com/ngrams/info>

Bookworm Open Library tips: click the 'cog' icon next to the 'i' to change the time period or click the underlined words next to the search term to change which books are searched (e.g. subject, language, country, gender of author). You can also compare the same word or phrase in different corpus.

Tip: if you're more interested in newspapers, try the Library of Congress's Chronicling America collection at <http://arxiv.culturomics.org/ChronAm/> or Australia and New Zealand newspapers at <http://dhistory.org/querypic/create/>

NB: in both tools, copyright affects the availability of 20th century sources.

Transcription errors may also affect results, particularly for older books (e.g. the 'long s' as f)

### Exercise 3: Trying entity recognition

Time: approx 5 minutes.

Goal: explore methods for extracting data from text

1. In your browser, go to <http://nlp.stanford.edu:8080/ner/>
2. Find a short paragraph of text (e.g. from a news site or digitised text) to paste into the box
3. How many of the things (concepts, people, places, events, references to time or dates, etc) you recognise did it pick up? Is any of the other information presented useful? Did it label anything incorrectly? What if you change classifiers?

If you're curious about other text processing, try

<http://nlp.stanford.edu:8080/corenlp/process>

An alternative demo site is TextRazor <http://www.textrazor.com/demo>

### Exercise 4: Exploring scholarly data visualisations

Time: approx 10-15 minutes.

Goal: hands-on experience; practice critical analysis.

Pair up with your neighbour to explore and discuss one of the visualisations overleaf, using these questions to get you started:

1. In your browser, go to one of the sites below
2. Take a few minutes to explore the visualisation
3. Discuss with your neighbour:
  - What do you think is being presented here?
  - Can you easily see where to start and how to use it?
  - What stories or trends can you start to see?
  - Does it work better at one scale over another?
  - Do you find it more effective at aggregate or detail level?
  - Does it present an argument or provide a space to develop and explore one?
  - If it was designed to present an argument or investigate a particular question, what do you think that was?
  - What have you learned from visualisation that you might not have learned from looking at the data or reading text about it?
4. Report back to the group: summarise the sites purpose, visualisation formats and data types in a sentence, then share the most interesting parts of your discussion

### **University of Richmond, "Visualizing Emancipation"**

<http://www.americanpast.org/emancipation/>

Further information: <http://dirt.terrypbrock.com/2012/04/visualizing-emancipation-examining-its-process-through-digital-tools/>

### **Stanford "Mapping the Republic of Letters"**

<http://www.stanford.edu/group/toolingup/rplviz/rplviz.swf>

Further information: <http://openglam.org/2012/03/21/mapping-the-republic-of-letters/>, <http://danbri.org/words/2010/11/22/603>

### **GAPVis**

<http://gap.alexandriaarchive.org/gapvis/index.html#index>

Further information: <http://googleancientplaces.wordpress.com/>

### **Digital Harlem :: Everyday Life 1915-1930**

<http://digitalharlem.org/>

Further information: <http://digitalharlemblog.wordpress.com/>  
<http://writinghistory.trincoll.edu/evidence/robertson-2012-spring/>

### **Orbis**

<http://orbis.stanford.edu/>

Further information: <http://hestia.open.ac.uk/updating-orbis/>

### **Digital Public Library of America's timeline, map, bookshelf**

<http://dp.la/>

Further information: <http://dp.la/info/> and <http://dp.la/info/news/blog/>

### **Lost Change**

<http://tracemedia.co.uk/lostchange/>

Further information: <http://blog.britishmuseum.org/2014/02/19/lost-change-mapping-coins-from-the-portable-antiquities-scheme/>

### **The State of the Union in Context**

<http://benschmidt.org/poli/2015-SOTU>

## Exercise 5: create a chart using Google Fusion Tables

These exercises are based on data provided by Tate. I have cleaned and created simplified versions of the data; a version of the original data is in the handouts folder.

Time: approx 5 minutes. Goal: gain skills handling data and practice a specific tool

NB: Google updates these sites frequently. If your screen options don't match the instructions, please ask for help. An alternate version using Excel is available.

### Load the data

1. Find the file Tate\_artists\_percountry.xlsx in the zipped folder supplied earlier
2. Go to <https://drive.google.com/> and log into Google (if you aren't already)
3. Go to <http://bit.ly/Xw0zNJ> (or <https://www.google.com/fusiontables/data?dsrcid=implicit>) to access Fusion Tables from your account.
4. You should see a screen 'Import new table' with the option called 'From this computer' highlighted below.
5. Click 'Choose file' and select Tate\_artists\_percountry\_Englishnames.xlsx. Click 'Next'.
6. Click 'Next' on the next screen, then click 'Finish' on the following screen.

If you want to fill in the options on the Import screen you can update them as follows: untick 'Allow export'. For 'Attribute data to' put 'Tate' and for Attribution page link put <http://www.tate.org.uk/about/our-work/digital/collection-data>.

### Review the loaded data

7. The screen should load in 'Row' view that looks something like a spreadsheet with two columns.

### Make a pie chart

8. At the end of the row of menu options, there should be a red box with a plus sign in it. Click that, then select 'Add chart'.
9. Scroll down the left-hand side to find the Pie Chart option. Click the Pie Chart image.
10. On the 'Configure pie chart screen', check that the Category is set to CountryOfBirth and Value is set to Number of artists.
11. Change 'Maximum slices' to 80.
12. Click 'Done' (over on the right-hand side).
13. You should have a pie chart of your data!

If you finish early, try: changing other options on the 'Configure pie chart' screen; making a line or bar chart with the uploaded data or trying the map. Which formats best suit the data?

An Excel version of this exercise is available at

<http://www.openobjects.org.uk/2015/03/creating-simple-graphs-with-excel-pivot-tables-and-tates-artist-data/>

## Exercise 6: Geocoding data and creating a map using Google Fusion Tables

Time: approx 10 minutes

Google Fusion Tables can geocode data directly from the table, but it sometimes needs some help. Fusion will have recognised some columns as containing location data, but it will not know much about those locations. The best way to see how it copes with a dataset is to geocode it and look over the resulting map to check that records have ended up in the right place. The slides handout includes step-by-step screenshots.

### Load the data

1. Go to <http://bit.ly/Xw0zNJ> (or <https://www.google.com/fusiontables/data?dsrclid=implicit>)
2. Upload Tate\_artists\_gender\_yearofbirth\_placeofbirth\_subset.xlsx

### Tell Fusion Tables about your data

3. Hover your cursor over the 'placeOfBirth' column until an arrowhead appears.
4. Click it to open the menu and select 'Change'.
5. A 'Change column' page will open. On this page, change the 'Type' value from 'Text' to 'Location'.
6. Click 'Save' to go back to the spreadsheet view.

This should trigger Google's geocoding process and it will start to find latitude and longitudes for the places listed.

### Create a map from your geocoded data

1. Click the '+' at the end of the menu row (starts with the File option and has a tab per data view)
2. When the menu opens, select 'Add map'.
3. The 'Geocode' view will probably open automatically. If it doesn't, change 'Location' to placeOfBirth on the 'Configure map' screen.
4. Geocoding takes a few minutes. And then...
5. Congratulations! You've created a map!

If you have extra time, you could try

- changing options on the map e.g. change icons from 'fixed' to 'bucket' and deciding on the best sizes for each 'bucket' of data
- visualising your own data
- load other Tate data from the 'Extra data for extended exercises' folder.

## Exercise 7: Applying data visualisation to your own work

Choose an option that suits your interests and skills:

- explore more visualisations
- sketch possible visualisations based on your own data or research questions
- try more tools for creating visualisations with your own or supplied data

### Explore more visualisations

The more visualisations you review, the better your ability to describe your own goals with visualisation and critically analyse existing visualisations.

There are links to visualisation blogs and other specialist sites on the Resources post at <http://bit.ly/UJwgEz> (i.e. <http://www.miaridge.com/resources-for-data-visualisation-for-analysis-in-scholarly-research/>) and on <https://www.tumblr.com/blog/scholarlyvision>

For each visualisation, you might like to consider: what sources have they used? Do they explain how they've prepared them? What effect have their choices of visualisation formats and tools had? What data or queries are prioritised, and which are more difficult or impossible?

If you have a particular type of data, process, format or audience in mind, ask for suggestions for sites to review.

### Sketch ideas for visualisations

Grab some markers and sketch out possible visualisations. Sketching helps you work out what story you want to tell so that you can choose the best format for your data and goals.

Ann K Emery's [Dataviz Design Process: 7 Steps for Beginners](#) includes some examples of sketches.

You may wish to review the 'planning' slides to help define what it is that you want to learn or express about the data. <http://extremepresentation.com/design/7-charts/> can help with choosing a structure. Once you are happy with your sketched ideas, you can work out how to clean or summarise your data and decide which tools to use.

### Try visualising data in different tools

These sites can be used with your own or public data:

- Google Fusion Tables
- Viewshare <http://viewshare.org>
- Voyant Tools, a 'reading and analysis environment for digital texts' <http://voyeurtools.org>
- Plot.ly, graph and share your data' <https://plot.ly>
- Raw, 'the missing link between spreadsheets and vector graphics' <http://raw.densitydesign.org>
- Palladio <http://palladio.designhumanities.org>
- Tableau Public, download from <http://tableausoftware.com>
- Dipity <http://www.dipity.com> for visual timelines without programming

Some visualisation sites provide example datasets that can be used in any application. Datasets designed for visualisations include those listed at <http://www.tableau.com/public/community/sample-data-sets> and <http://www.tableau.com/public/community/data-catalogue> A range of cultural and historical data is listed at [http://museum-api.pbworks.com/w/page/21933420/Museum APIs](http://museum-api.pbworks.com/w/page/21933420/Museum%20APIs)

You can also try visualising the data from the British Library Pin-a-tale project, available in Google Docs at <http://bit.ly/WT1Ai5>

Try a visualisation and evaluate the results. Is more cleaning or transformation needed? You may need to iterate with different versions of a dataset after cleaning or enhancing it, or re-organising it to meet the requirements of different tools.