Exploring Data Visualization with Flourish

UIC Spring 2020 DH Workshop Series
Kristen Mapes and Kate Topham
Agenda

1. Overview: What is visualization and why do we do it?
2. Working with data
   a. Compare two datasets
   b. Asking questions of our data
3. Introduction to Flourish
4. Visualizing with Flourish
5. Breakout activity: Design Lab
Overview - What’s Visualization?
Definitions

use of abstract, non-representational pictures to show numbers


the use of computer-supported, interactive, visual representations of abstract data to amplify cognition


@kmapesy @tophkat CC-BY
Why do we visualize?

- **Exploratory**
  - What’s here?
  - What questions can I ask?

- **Analytical**
  - How does this answer my research question?

- **Communicative**
  - Making your argument
  - How can I explain this data to others?
Structured data vs unstructured data

Structured:
- Spreadsheets
- Encoded text

Unstructured:
- Raw text
- Images
- Audio recordings
Data Types

nominal

ordinal

quantitative

relational
Data-ink ratio = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}

= \text{proportion of a graphic’s ink devoted to the non-redundant display of data-information}

= 1.0 - \text{proportion of a graphic that can be erased without loss of data-information.}

PROPORTION OF NEGROES IN THE TOTAL POPULATION OF THE UNITED STATES.

RAPPORT DES NÈGRES À LA POPULATION TOTALE DES ÉTATS UNIS.

DONE BY ATLANTA UNIVERSITY.
Our forecast for every district

The chance of each candidate winning, with all 435 House districts shown at the same size
Our forecast for every district

The chance of each candidate winning, with all 435 House districts shown at the same size
“Marks” - Means of Representing
“Marks” - Means of Representing

What is the best way to represent your data?

How easy will it be to understand your visualization?

From Mackinlay
“Marks” - Ways of Representing

- **Quantitative**
  - Position
  - Length
  - Angle
  - Slope
  - Area
  - Volume
  - Density
  - Color Saturation
  - Color Hue

- **Ordinal**
  - Position
  - Density
  - Color Saturation
  - Color Hue
  - Texture
  - Connection
  - Containment

- **Nominal**
  - Position
  - Color Hue
  - Texture
  - Connection
  - Containment
  - Density
  - Color Saturation
  - Shape

From Mackinlay
Our forecast for every district

The chance of each candidate winning, with all 435 House districts shown at the same size.
Our forecast for every district

The chance of each candidate winning, with all 435 House districts shown at the same size

Position
Color Hue
Texture
Connection
Containment
Density
Color Saturation
Shape
Length
Angle
Slope
Area
Volume

@kmapesy @tophkat CC-BY
Tools we’re not using:

- Tableau [tableau.com](https://tableau.com)
- Graph Commons [graphcommons.com](https://graphcommons.com)
- Onodo [onodo.org](https://onodo.org)
- Vistorian [vistorian.net](https://vistorian.net)
- Gephi [gephi.org](https://gephi.org)

Programming Heavy:

- D3 [d3js.org](https://d3js.org)
- Vega Lite [vega.github.io/vega-lite](https://vega.github.io/vega-lite)
Working with data
Survey of Scottish Witchcraft Database, 1563-1736

- From University of Edinburgh, Julian Goodare, Lauren Martin, Joyce Miller, and Louise Yeoman, in January 2003
- Nearly 4,000 people accused
- Project website; Database website

**Scottish History**

The Survey of Scottish Witchcraft

1563-1736

By Julian Goodare, Lauren Martin, Joyce Miller and Louise Yeoman
January 2003

- Accessing the Database
  - About the Project
    - History of the Project
    - Acknowledgements
    - Authorship and Copyright
    - Disclaimer
    - How to Cite Us
    - Sources and Bibliography
    - The Illustration
    - Contact Information
  - Introduction to Scottish Witchcraft
  - Further Reading on Scottish Witchcraft
  - Links

Welcome to the Survey of Scottish Witchcraft. This is an electronic resource for the history of witchcraft supporting web pages.

The database contains all people known to have been accused of witchcraft in early modern Scotland—accused, how they were tried, what their fate was, and on a wide range of themes relating to social and instance, you can find all known cases involving neighbourhood quarrels, or demonic possession, or feuds showing how witchcraft cases were distributed; this is important because prosecutions tended to

There is also supporting material. An Introduction to Scottish witchcraft explains some of the findings and is also important; the database won't tell you everything on its own. However, it will tell you some things you should help you think about the history of witchcraft and what it means to us today.
Survey of Scottish Witchcraft Database, 1563-1736

- Subset of data from Miriam Posner’s Tableau Mapping tutorial
- Includes 3220 records:
  - Name
  - Gender
  - Latitude & Longitude
  - County
  - Date
Places of Residence for Accused Witches (total named accused witches: 3141)

https://witches.is.ed.ac.uk/
“Dollars and Cents” and the Bechdel Test

- Assembled by fivethirtyeight for “The Dollar-And-Cents Case Against Hollywood’s Exclusion of Women” (2014)
- Original dataset
- Bechdel test:
  - Two named women characters talk to each other about anything but a man
“Dollars and Cents” and the Bechdel Test

- 1795 films, dating from 1970 through 2013
  - Title
  - Does it pass the test?
  - Date made
  - Genre
  - Budget
  - Gross income (domestic & international)

h/t to Etan Adar, @eytanadar for the expanded dataset!
Activity!

Break into groups of 3 - each group chooses one of the two datasets

In 5 minutes, brainstorm 10 questions that you could ask of or explore with the dataset

Report/share out

Download the data here: http://bit.ly/UICVizData

@kmapesy @tophkat CC-BY
Kristen’s Questions

1. What role did gender play in accusations of witches in early modern Scotland?
2. Were witches accused in certain areas of Scotland more than others?
3. Was there a rural/urban difference in accusations?
4. Were more witches accused closer to England or further away from English influence?
5. Where accusations concentrated in certain moments of time or were they spread out more evenly over this time period?
6. Did accusations tend to take place during specific times of the year?
7. During summer? Or winter?
8. During holidays?
Kate’s Questions

1. How has the percentage of movies that pass changed over time?
2. **Does the rate of passing vary by genre? Which genres do worse or better?**
3. Are films that pass the bechdel test given higher or lower budgets than those that fail?
4. Are films directed by women more likely to pass the test?
5. **Do movies that pass the test make more or less money than those that fail?**
6. Does the gross income vary by genre?
7. How does the US stack up against other countries [good one to show bias]
8. Are movies that pass the test rated higher or lower than those that fail?
9. How do the highest-grossing directors do in the bechdel test?
10. How has the gross differential between male- and female-identified directors changed over time?
Intro to Flourish
What is flourish? Why flourish?

- Free!
- Web-based
- Flexible
- Interactive
- Embed on external sites

NB: Do not use data that is private or sensitive! Free Flourish makes all data public.
You Have to Aggregate Your Data

Be prepared to iterate over your data!

If you can do what you need in excel or google sheets, do it! You don’t need a fancy thing!
Create a Flourish account here: [app.flourish.studio/login]
Flourish walkthrough
Flourish walkthrough
Design Lab!
Your own research question

Take two minutes to come up with a research question that you could answer using structured data
What data might you find (or do you need) in doing this research?

Take three minutes to come up with a list of data types that would help you undertake this research.

Don’t worry about whether or not the actually data exists; imagine you can get whatever you want.
Draw 2 visualizations that answer your question

Take 5 minutes.
What marks will you use?
What kind of interaction would be helpful?

@kmapesy @tophkat CC-BY
What viz types might be suitable for your question?
Thanks!
dh.msu.domains/uic/visualization

Kristen Mapes
kmapes@msu.edu | @kmapesy

Kate Topham
tophamka@msu.edu | @tophkat