YOUR NAME / GROUP MEMBER NAMES:

AMST 30161
Lab #2: Structured Data Lab Notebook

Link to lab procedure

How to Work Through This Lab

There are four main sections to this lab:

- Background, University Archives, and Sample Datasets
- Exploratory data analysis/visualization (EDA)
- Mapping
- Network Analysis

We'll spend two Thursday class periods leading up to the midterm break working on this lab (10/7, 10/14). Folks can (but are not required) to continue working over the break.

Our endpoint for this lab will be documenting your work going through the three sections of the lab. There will not be a "now it's your turn" section at the end of this lab.

Folks who want to continue working with these tools/methods for your own research questions/topics will have an opportunity to do that in our third and last lab.

Since there's not going to be a collaborative "now it's your turn" section at the end of the lab, folks will work in groups in class time, but you can also choose to turn in an individual lab notebook.

Collaborate to the degree that it is useful and helpful to work through the material.

I'm asking folks to submit a lab status update before you leave for break (by the end of Week #8) and submit the lab by the end of break (or sometime early in Week #10 once we're back).

Lab Notebook Components:

- Reflections/observations from the background, University Archives, and sample datasets sections of the lab
- Reflections/observations from exploratory data analysis/visualization section of the lab
- Reflections/observations from the mapping section of the lab
- Reflections/observations from the networks section of the lab
- Any data files you worked with (that aren't in the sample datasets)

- If you worked with Python/RStudio: Scripts or notebooks for any original code you generated
- Individual reflection
  - At least 300 words (or 3-4 minutes audio/video), addressing the following questions:
    - What you learned about working with structured data through this lab
    - What you learned about data analysis/visualization through this lab
    - How you're thinking about data analysis/visualization as they relate to ND football history and primary sources differently after this lab
    - Where you would go next with these tools/methods (other research questions/topics you're interested in exploring, other data sources or types of data you would want to be able to work with, etc)
    - Other comments/questions/observations

If you're working in Google Drive, you can just submit a link to Google Drive.

Notebooks can be individual or collaborative.

Link to lab procedure (GitHub)

Background

Link to this section of the lab

To prepare for this lab, we read Miriam Posner's 2015 keynote "Humanities Data: A Necessary Contradiction" and explored the companion "Early African American Film: Reconstructing the History of Silent Race Films, 1909-1930" digital project.

Discussion/Reflection Questions

- What stood out to you from this reading about how "data" can be a useful concept for research questions that might fall broadly under the umbrella of the humanities (or other qualitative disciplines/fields)?
- What stood out to you from this reading about how "data" can be a challenging or less than useful concept for research questions that might fall broadly under the umbrella of the humanities (or other qualitative disciplines/fields)?
- How does the way this reading talks about "data" relate to your own prior experience with different kinds of research/analysis methods?
- In what ways do you think the way Posner talks and thinks about data would be a useful starting point for engaging with archival materials related to Notre Dame football?
  - To put this another way, what kinds of questions or topics would thinking of primary sources as data allow us to explore or engage with?
- Other questions/observations from this reading

Thinking about the companion digital project that goes along with this reading...

- What can we tell about how the project is thinking about and using data in the way Posner frames it in the reading?
- How do we see the project (and reading) talking about, contextualizing, or framing the way it uses data within the scope of a particular set of research questions/topics?
  - To put this another way, what role does data play in framing the project's scope, goals, questions, etc?
- What types of research questions/topics does the project take up using structured data?
- Other questions/comments/observations from looking at the project

University Archive Resources

[Link to this section of the lab](#)

We've previously looked at some of the digital collections in the Notre Dame University Archives.

In the previous lab, we focused on using these source materials as unstructured text for different types of analysis facilitated by digital/computational tools.

In some cases, that analysis did lead to structured data (or data in a spreadsheet-like format). Or the back-end data driving the visualizations and analysis the different tools presented was made possible through analyzing and processing unstructured text to produce or generate structured data.

Discussion/Reflection Questions

- In what ways have you seen or engaged with that workflow (unstructured text to structured data) in our first lab?
- How is that similar or different from the ways we see structured data show up in the Posner reading and companion digital project?

Now we're going to think about how we can engage with primary source material with structured data as a starting point for other kinds of analysis and exploration that rely on modes of data analysis/visualization that work with structured data.

Let's take a look at a few different types of primary sources that can lend themselves to working as structured data.

Sample primary sources:

- 1922 Student Directory
- 1923 Football Season Review

Discussion Questions:

- What can we tell about who created this resource (specific authors as well as publisher)?
- What kinds of information are included in this resource (images, text, etc)?
  - Where do we see opportunities for this information to work as structured data?
- What topics are covered or addressed in this resource?
- What kinds of research questions or topics could you analyze/explore using this resource?
  - Specifically, if we were able to engage with the information in this resource as structured data, what kinds of analysis/visualization or questions/topics could we explore.
- Other comments/questions/observations that surface as you think about structured data in relation to this primary source.

Sample Datasets

Link to this section of the lab

This lab includes four sample datasets.

- Directories takes a PDF of digitized historical student directories and uses a combination of OCR, OpenRefine, and regular expressions to transform the PDF to a CSV file
- Football Rosters uses Python to scrape Sports Reference web pages for ND football to get a data table
- Football Schedules uses Python to scrape Sports Reference web pages for ND football schedule and game results to get data table
- Rockne Coaching Tree transforms information about Rockne's coaching tree into a data table that could be used for network analysis

I encourage folks to skim the information provided about the background/data processing workflow for all the sample datasets.

For the lab notebook, you can engage with the discussion/reflection questions for 1-2 of the sample datasets.

Directories

Link to this section of the lab

Discussion/Reflection Questions:

- Where would you get started transforming this primary source into structured data?
- What would that data structure look like?
  - Specifically, what columns would you have, what data would go in each column, etc.

- What types of analysis/visualization or exploration would you be able to do with this resource as structured data?
- Other comments/questions about this source as structured data

Football Rosters

Link to this section of the lab

Discussion/Reflection Questions:

- What parts of information on this page would you want to work with as structured data?
- What types of analysis/visualization or exploration would you be able to do with this resource as structured data?
  - What questions or topics would you be able to explore using data from this resource?
- Other comments/questions about this source as structured data

Football Schedules

Link to this section of the lab

Discussion/Reflection Questions:

- What parts of information on this page would you want to work with as structured data?
- What types of analysis/visualization or exploration would you be able to do with this resource as structured data?
  - What questions or topics would you be able to explore using data from this resource?
- Other comments/questions about this source as structured data

Rockne Coaching Tree

Link to this section of the lab

Discussion/Reflection Questions:

- What parts of information on this page would you want to work with as structured data?
- What types of analysis/visualization or exploration would you be able to do with this resource as structured data?
  - What questions or topics would you be able to explore using data from this resource?
- Other comments/questions about this source as structured data

Exploratory Data Analysis/Visualization

Link to this section of the lab

Section table of contents

I'm asking everyone to spend some time with WTFcsv and Tableau.

PICK ONE:

- Folks with programming backgrounds can spend time working in RStudio/Python.
- Folks who don't have a programming background can do more extensive work with Microsoft Excel and Google Sheets.

The discussion/reflection questions at the end of the lab ask you to think holistically across the different programs/tools you focused on.

- There will be some questions that relate to specific programs/tools. You are not required to respond to or engage with all of the questions.
- Definitely cover the questions specific to WTFcsv and Tableau, but pick and choose from the others based on what programs/tools you worked with.

WTFcsv Discussion and Reflection Questions

Link to this section of the lab
- What do you notice about the results for this dataset (or these datasets) displayed in WTFcsv?
- How does the tool's results shape your understanding of the data?
- What additional questions do you have about this data (or these datasets)? Where would you go next with exploring this dataset using some of the analysis tools/approaches highlighted in the tool?
- Other comments/questions/observations

Excel Discussion/Reflection Questions

Link to this section of the lab
- Why would it be helpful to be able to sort/filter/etc within a dataset?
- As you explore the sort/search/filter functionality, what questions emerge about the data?

- How does working in Excel shape your understanding of the data?
- What types of visualizations (or aggregations) were you able to generate in Excel using PivotCharts/PivotTables?
- How could those visualizations shape or impact your understanding of the data?
- What additional questions do you have about this data (or these datasets)? Where would you go next with exploring this dataset using some of the analysis or visualization tools/approaches?
- Other comments/questions/observations

Tableau Discussion/Reflection Questions

Link to this section of the lab
- How does working in Tableau shape your understanding of the data?
- What types of visualizations (or aggregations) were you able to generate in Tableau?
- How could those visualizations shape or impact your understanding of the data?
- What additional questions do you have about this data (or these datasets)? Where would you go next with exploring this dataset using some of the analysis or visualization tools/approaches?
- Other comments/questions/observations

Python and RStudio Discussion/Reflection Questions

Link to Python section of the lab
Link to RStudio section of the lab
- How does working in Python/RStudio shape your understanding of the data?
- What types of visualizations (or aggregations) were you able to generate in Python/RStudio?
- How could those visualizations shape or impact your understanding of the data?
- What additional questions do you have about this data (or these datasets)? Where would you go next with exploring this dataset using some of the analysis or visualization tools/approaches?
- Other comments/questions/observations

EDA General Discussion/Reflection Questions

- Thinking holistically across the tool(s) you interacted with, what questions were you interested in asking or exploring the sample datasets using the available analysis/visualization tools?
- How did using the programs shape your understanding of the data?
- Where would you go next with these tools/approaches?
  - What questions or topics would you want to explore using the sample datasets?

○ Or, what other types of information/data related to ND football would you want to work with?
● Other comments/questions/observations

Mapping

Link to this section of the lab

I'm asking everyone to spend some time with ArcGIS Online.

● Folks who don't have a background with data/mapping might want to start with Google MyMaps before you jump into ArcGIS Online.

PICK ONE:

● Folks with programming backgrounds can spend time working in RStudio/Python.
● Folks interested in doing more in-depth work with mapping can explore Carto (web-based program that does not involve any coding).

The discussion/reflection questions at the end of the lab ask you to think holistically across the different programs/tools you focused on.

● There will be some questions that relate to specific programs/tools. You are not required to respond to or engage with all of the questions.
● Definitely cover the general questions and any specific to ArcGIS Online, but pick and choose from the others based on what programs/tools you worked with.

Google MyMaps Discussion/Reflection Questions

Link to this section of the lab
● How does working in Google MyMaps shape your understanding of the data?
● What types of visualizations (or aggregations) were you able to generate?
● How could those visualizations shape or impact your understanding of the data?
● What additional questions do you have about this data (or these datasets)? Where would you go next with exploring this dataset using some of the analysis or visualization tools/approaches?
● Other comments/questions/observations

Sample Google MyMap
● Public map
● MyMap project

ArcGIS Online Discussion/Reflection Questions

Link to this section of the lab
Link to sample ArcGIS Online map

- How does working in ArcGIS Online shape your understanding of the data?
- What types of visualizations (or aggregations) were you able to generate?
- How could those visualizations shape or impact your understanding of the data?
- What additional questions do you have about this data (or these datasets)? Where would you go next with exploring this dataset using some of the analysis or visualization tools/approaches?
- Other comments/questions/observations

Carto Discussion/Reflection Questions

Link to this section of the lab
Link to sample Carto map

- How does working in Carto shape your understanding of the data?
- What types of visualizations (or aggregations) were you able to generate?
- How could those visualizations shape or impact your understanding of the data?
- What additional questions do you have about this data (or these datasets)? Where would you go next with exploring this dataset using some of the analysis or visualization tools/approaches?
- Other comments/questions/observations

Python and RStudio Discussion/Reflection Questions

Link to the Python section of the lab
Link to the RStudio section of the lab

- How does working in Python/RStudio shape your understanding of the data?
- What types of visualizations (or aggregations) were you able to generate?
- How could those visualizations shape or impact your understanding of the data?
- What additional questions do you have about this data (or these datasets)? Where would you go next with exploring this dataset using some of the analysis or visualization tools/approaches?
- Other comments/questions/observations

Mapping General Discussion/Reflection Questions

- Thinking holistically across the mapping tool(s) you interacted with, what questions were you interested in asking or exploring about the spatial components of the sample datasets?
- How did using spatial analysis/visualization tools shape your understanding of the data?
- Where would you go next with spatial analysis/visualization tools?
  - What questions or topics would you want to explore using the sample datasets?
  - Or, what other types of spatial information/data related to ND football would you want to work with?
- Other comments/questions/observations

Networks

Link to this section of the lab

Section table of contents

I'm asking everyone to spend some time with ConnectTheDots.

PICK ONE:

- Folks who want to more in-depth work with network analysis/statistics can spend time with Gephi (open-source software, no coding)
- Folks who want to focus more on network graphs/visualizations can spend time with Palladio (web-based program, no coding)
- Folks with programming backgrounds are welcome to explore some of the linked RStudio/Python resources and tutorials

The discussion/reflection questions at the end of the lab ask you to think holistically across the different programs/tools you focused on.

- There will be some questions that relate to specific programs/tools. You are not required to respond to or engage with all of the questions.
- Definitely cover the general questions and any specific to ConnectTheDots, but pick and choose from the others based on what programs/tools you worked with.

ConnectTheDots Discussion/Reflection Questions

Link to this section of the lab
Link to sample ConnectTheDots results
- What do you notice about the results for this dataset displayed in ConnectTheDots?
- How does the tool's analysis/visualization results shape your understanding of the data?
- What additional questions do you have about this data? Where would you go next with exploring this dataset using some of the analysis tools/approaches highlighted in the tool?
- Other comments/questions/observations

Palladio Discussion/Reflection Questions

Link to this section of the lab

- What do you notice about the results for this dataset displayed in Palladio?
- How does the tool's analysis/visualization results shape your understanding of the data?
- What additional questions do you have about this data? Where would you go next with exploring this dataset using some of the analysis tools/approaches highlighted in the tool?
- Other comments/questions/observations

Gephi Discussion/Reflection Questions

[Link to this section of the lab](#)

- What do you notice about the results for this dataset displayed in Gephi?
- How does the tool's analysis/visualization results shape your understanding of the data?
- What additional questions do you have about this data? Where would you go next with exploring this dataset using some of the analysis tools/approaches highlighted in the tool?
- Other comments/questions/observations

Networks General Discussion/Reflection Questions

- Thinking holistically across the network analysis/visualization tool(s) you interacted with, what questions were you interested in asking or exploring about the sample network data?
- How did using network analysis shape your understanding of the data (or the Rockne coaching tree more broadly)?
- Where would you go next with network analysis/visualization tools?
    - What questions or topics would you want to explore using the sample datasets?
    - Or, what other types of network information/data related to ND football would you want to work with?
- Other comments/questions/observations

Individual Reflection

[Link to this section of the lab](#)

At least 300 words (or 3-4 minutes audio/video), addressing the following questions:

- What you learned about working with structured data through this lab
- What you learned about data analysis/visualization through this lab
- How you're thinking about data analysis/visualization as they relate to ND football history and primary sources differently after this lab
- Where you would go next with these tools/methods (other research questions/topics you're interested in exploring, other data sources or types of data you would want to be able to work with, etc)
- Other comments/questions/observations