

GROUP MEMBER NAMES:

YOUR NAME:

AMST 30161
Lab #1: Textual Analysis Lab Notebook

Required Components

- Reflections/observations from Thursday 9/9 in-class work (Voyant Tools and Distant Reader)
- Reflections/observations from Thursday 9/16 in-class work (AntConc and Python)
- Reflections/observations from your own textual analysis experiment
- Any PDF or TXT files you worked with
- Any Python scripts/Jupyter Notebooks used in the lab

If you're working in or using Google Drive, you can download the project folder as a compressed zip folder and submit to Canvas. You can also include a link to the Google Drive folder as part of the lab notebook.

[Link to lab procedure](#) (GitHub)

Thursday 9/9 (Voyant Tools and Distant Reader)

Preliminary discussion questions:

- What stood out to you from the resources we engaged with on computational text analysis?
- Questions you have about some of the research methods and approaches mentioned?
- Things you think would be interesting to explore related to Notre Dame Football using some of these methods?
- Other questions/observations

University Archive Resources

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

Head to the University Archives' "[Digital Collections](#)" web page.

Browse the resources and materials listed, thinking about...

- Time period the resource covers

- Type of material/publication (student publication versus University publication, newspaper versus yearbook versus magazine, etc)

Focus on a specific text resource and select a PDF to download. For serialized publications, this may involve clicking through multiple pages to get to specific issues.

- If you're not sure where to get started or what resource to focus on, [the Scholastic](#) is a student publication from the period we've been discussing, and their annual football review has [its own digital collection](#).

Looking at a specific document...

- 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)?
- What topics are covered or addressed in this resource?
- What kinds of research questions or topics could you analyze/explore using this resource?

Voyant Tools

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

Preliminary reflection questions for Voyant:

- What do you immediately notice after uploading a text to Voyant?
- What stands out, or catches your attention?
- What types of information are contained in this page?
- What do you have questions about, or what is confusing and not immediately clear?
- Based on your initial scan, what function do you think these various components serve?

Reflection questions to consider as you explore the different tools and panels in Voyant:

- What differences do you notice across the tools/panels- how are they similar, how are they different?
- What words or features stood out in each representation of the textual analysis?
- How did you understand the textual analysis differently based on how it was presented?
- What questions do you have about how this tool calculated these results (or what can we tell about how the tool is calculating results/generating visualizations)?
- What questions do you have about the textual data after exploring this tool?

Final Voyant reflection questions:

- What did you find engaging or interesting about Voyant Tools?
- What types of historical research questions could a researcher have about a text or collection of texts?
- How effectively might Voyant be able to address those questions?
- What challenges, frustrations, or limitations did you encounter while using Voyant?

- What remaining questions do you have about the source you focused on?
- How would you move forward with Voyant as a tool for historical textual analysis?

Data Basic

Word Counter Reflection Questions

- What do you notice about the text through the results displayed in WordCounter?
- How does WordCounter's results shape your understanding of the text?
- What additional questions do you have about the text?
- Other comments/questions/observations

Same Diff Reflection Questions

- What do you notice about the individual text and the combined texts through the results displayed in SameDiff?
- How do the SameDiff results shape your understanding of the individual text and combined text?
- What additional questions do you have about these resources? Where would you go next using this tool?
- Other comments/questions/observations

Distant Reader

Reflection questions:

- What did you find engaging or interesting about Distant Reader?
- What challenges, frustrations, or limitations did you encounter while using Distant Reader?
- What words or features stood out in each representation of the textual analysis?
- How did you understand the textual analysis differently based on how it was presented?
- What questions do you have about how this tool calculated these results (or what can we tell about how the tool is calculating results/generating visualizations)?
- What remaining questions do you have about the source you focused on?
- How would you move forward with Distant Reader as a tool for historical textual analysis?
 - To put that another way, what types of historical research questions could a researcher have about a text or collection of texts? How effectively might Distant Reader be able to address those questions?
- Other questions/observations

AntConc

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

Searching Keywords in Context:

- How did your search results change?
- What happens if you continue to customize or edit the Kwic Sort options?
- How do you understand a key word differently based on how you tell the program to calculate context?

Visualizing Keywords in Context:

- How useful do you find these preliminary visualizations?
- How do they compare to the types of visualizations generated on Voyant?
- How do these visualizations impact your understanding of the text?
- What questions do you have based on these visualizations?

Collocates and Word Lists:

- What terms are statistically likely to appear in proximity to your search term?
- What happens when you change the Window Span (number of words to the right and left of your search term AntConc will include in the analysis)?

AntConc Reflection Questions:

- What do you notice is similar about Voyant Tools and AntConc as digital tools for textual analysis? What are the differences?
- Which did you prefer working with (and how/why)?
- How was your understanding of the text impacted by the analysis we did in AntConc? What questions do you still have?
- What would be some of your next steps for conducting textual analysis related to Notre Dame football, using Voyant or AntConc?
- What types of historical research questions can you see textual analysis being useful to answer or respond to (especially related to the conversations we've been having about college football/ND football)?

Python

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

Reflection Questions:

- What did you find engaging or interesting about the different aspects of these Python workflows?
- What challenges, frustrations, or limitations did you encounter while using and/or modifying these Python workflows?
- What words or features stood out in each component of the sample text analysis workflow?
- How did you understand the textual analysis differently based on how it was presented?

- What questions do you have about how the sample text analysis workflow and related Python libraries generated the analysis presented (or what can we tell about how the tool is calculating results/generating visualizations)?
- What remaining questions do you have about the source you focused on?
- How would you move forward with specific components of the sample Python text analysis workflow and related libraries as a tool for historical textual analysis?
 - To put that another way, what types of historical research questions could a researcher have about a text or collection of texts? How effectively might the different Python libraries and analysis methods highlighted be able to address those questions?
- Other questions/observations

Your Turn!

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

By this point in the lab we've covered a number of different tools for (computational) text analysis, including some that run in a web browser, some that run on your local computer, and some that involve working within a programming language.

Now it's your turn to come up with a small-scale question, topic, etc. related to the period of ND football history we've been looking at (or some of the broader themes we've been talking about) that you could explore using some of the primary sources and analysis/research methods covered in this lab.

Template/outline for the report:

- Background or context for the research question or topic you're exploring
 - At least 500 words (or 5 minutes audio/video) on how the question you're asking or topic you're exploring is grounded in the texts we've been reading and conversations we've been having about ND football history and cultural significance.
 - To put it another way, this is a "background" section.
- Sources and research/analysis methods you'll be using
 - At least 200 words (or 2 minutes audio/video) on how the specific sources you're engaging with and research/analysis methods you're using animate or connect to the question(s) you're asking or the topic(s) you're exploring.
 - To put it another way, this is a "methods" or "methodology" section
- What you found through the exploration/analysis
 - At least 500 words (or 5 minutes audio/video) on what you found through analyzing this source material using specific research/analysis methods. Folks are encouraged to include visualizations/screenshots/figures/etc in this section.
 - To put it another way, this is a "findings" section
- Conclusions and next steps

- At least 250 words (or 2-3 minutes audio/video) on how your findings relate to the broader context or background you outlined at the start of the report.
- How can the work you did in this lab inform how we think about or understand ND football's history and cultural significance? How does it connect to the readings and conversations we've been having in class?
- Where would you go next with this research? What questions emerged through doing this work? What questions were you not able to address?
- To put it another way, this is a hybrid "discussion" and "conclusion" section
- Any PDF or TXT files you created or worked with (that aren't already in the Lab 1 Google Drive)
- Any Python scripts/Jupyter Notebooks (or RStudio/RMarkdown files) used in the lab
- Individual reflections
 - At least 200 words (or 2-3 minutes audio/video) from each member of the group, addressing the following questions:
 - Your contribution to the lab report
 - Challenges you faced in this lab and how you solved them
 - What you learned about text analysis through this lab
 - How you're thinking about text analysis (and ND football history/primary sources) differently after this lab
 - Other comments/questions/observations

Collaborating Well

Collaboration is also your friend here- you don't have to come up with all the ideas or do all the work on your own. For example, folks with Python skills can focus on assembling a corpus of primary source materials as plain-text formats. Folks who are more comfortable with statistical analysis and/or data analysis/visualization can focus on those aspects of the project. Folks who are more comfortable with close reading approaches for analyzing primary sources can focus on the components of the lab that ask you to think about background/context/broader significance.

In short, leverage the group's strengths. And divide/conquer.

Folks are welcome to meet with me (individually or as a group) to ask questions and think about next steps for how to approach this lab.

Where to Start: Articulating a Research Question

Our end-point for this class is an open-ended final project that takes up course topics/themes/etc. This lab is not a final project.

We'll have in-class work time on Thursday 9/23 to continue working on this lab, and folks should anticipate spending 5-7 hours (total) outside class time working on this lab.

Given the limited scope of this assignment, I encourage folks to drill down and narrow their focus for what you want to analyze or explore.

- This could be a particular player, team, or coach.
- Or you could focus on a particular season or period of time.
- It could be a specific publication or type of source material.

The final project is a space where you take up more advanced computational methods or engage in larger-scale, multi-faceted research questions- hanging out in the world of exploratory data analysis is absolutely fine for this lab.

For now, think about the questions you might be interested in asking or the topics you might be interested in exploring- we've already done some brainstorming in this direction with our in-class work.

Building a Corpus

Then, thinking about the primary sources or materials that would be useful for exploring this question/topic.

For more background and context on the various University Archive publications and collections, users are encouraged to consult Thomas E. Blantz's [*The University of Notre Dame: A History*](#) (University of Notre Dame Press, 2020).

- [Link to online access via Hesburgh Libraries \(ND users\)](#)

The text analysis tools and resources covered in this lab will work best (and in many cases require) plain-text files as the starting point for analysis.

We can extract plain-text from scanned documents using a process called "optical character recognition" or OCR.

For folks with Python proficiency, the Python portions of the lab include a sample OCR workflow that converts digitized PDFs to image files, then takes the image files as binary data and extracts text. The notebooks provided in that section of the lab also include a number of resources for refining or further developing/customizing an OCR workflow.

Acrobat Adobe Pro

But, depending on the number of documents you're working with as well as the quality of OCR output needed for your project, Adobe Acrobat Pro will likely get you where you need to go.

- NOTE: This is a different program than the Adobe Acrobat Reader DC free program you may have on your computer.

Notre Dame students have access to the Adobe Pro software title through the Adobe Creative Cloud Desktop Application, available free through OIT.

- NOTE: The full Adobe Creative Cloud suite is available on any [OIT lab computer](#) and can be accessed through the OIT general purpose [Virtual Computer Lab](#)

To get Adobe Pro on your own computer:

- Download the program: [OIT, Adobe Creative Cloud Desktop Application](#)
- Request a license: OIT, "[Request a license for Adobe Creative Cloud](#)," *ND Service Now*

Once you have Adobe Pro on your own computer, you can use the expanded features to run an OCR workflow on a single PDF or multiple PDFs.

Adobe Pro also allows you to export the contents of a single PDF or multiple PDFs to plain-text (.txt) files.

To run OCR within Adobe Pro:

- University of Illinois Scholarly Commons, "[Introduction to OCR and Searchable PDFs: Adobe Acrobat Pro](#)" *University of Illinois Library*
- Rowan University Library Digital Scholarship Center, "[Acrobat Pro: Optical Character Recognition for Research, Learning, and Accessibility](#)" *YouTube video* (20 August 2020)
- Pixascene, "[Perform an OCR on a PDF document using Adobe Acrobat Pro](#)" *YouTube video* (15 June 2020)
- Tulane University Libraries, "[Digitize Your Sources: OCR and Adobe Acrobat Pro](#)" *Tulane University Library Guide*

To export a single PDF as TXT from Adobe Pro (using "Export"):

- Adobe Acrobat User Guide, "[Convert or export PDFs to other file formats](#)" *Adobe* (26 August 2021)

To export multiple PDFs as TXT files from Adobe Pro (using "Action Wizard"):

1. View -> Tools -> Action Wizard -> Create New Action
 2. Choose 'Save & Export' -> Save -> add to right-hand pane
 3. At right-hand pane -> choose folder and click Specify Settings to change export format to TXT
- Source: StackOverflow, "[How to convert batch pdf files to text using Adobe Acrobat Pro?](#)" *StackOverflow* (2015)
 - For more on Action Wizard: Adobe Acrobat User Guide, "[Action Wizard \(Acrobat Pro\)](#)" *Adobe* (2 June 2020)

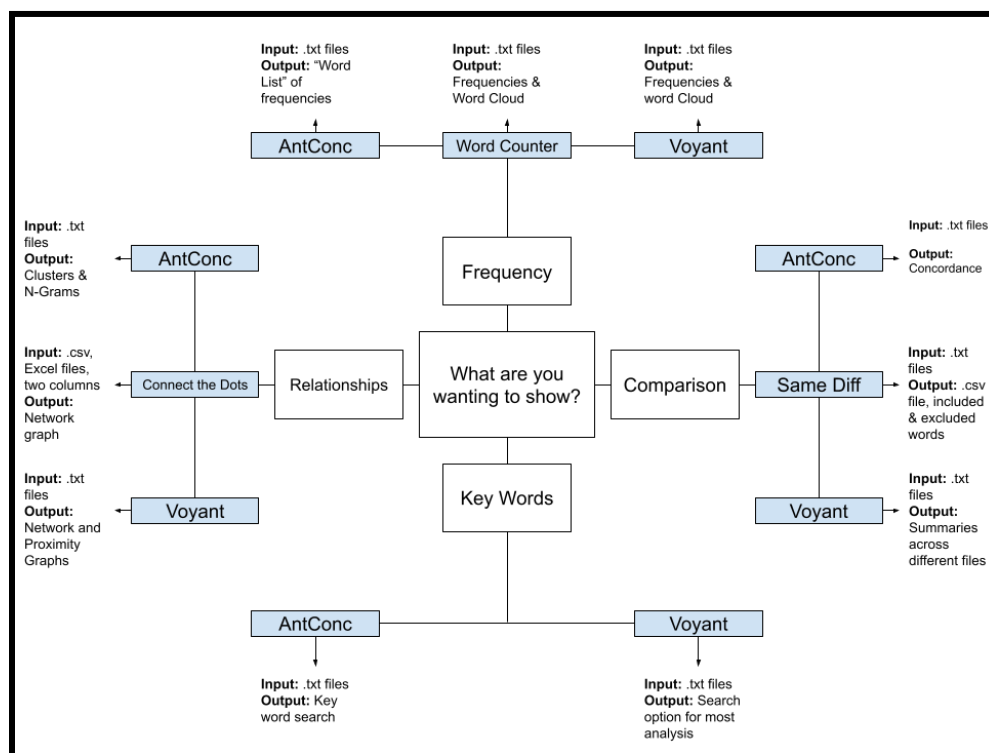
Address Your Research Question

Any of the tools and methods covered in this lab are fair game. Folks are also welcome to branch out and/or go more deeply into any of these areas, keeping in mind the limited assignment scope.

But the tools you choose and the analysis methods you land on should be driven by your topic/question and what you want to know about it.

For example, are we wanting to show relationships, similarity, difference. Or change over time. Or comparison.

The "Text Analysis Flow Guide" visual is a great place to start for thinking about your options.



Ferdio's [DataVizProject](#) is another useful resource, although it focuses more on general data analysis/visualization and not specifically text sources.

Some questions from [the "Distant Reader" documentation](#) that might get you started:

- "The aforementioned process is akin to the application of descriptive statistics and merely provides an overview of the results. For more in-depth analysis and understanding, you will want to do more than read frequencies. For example, you will want to identify documents containing interesting words or phrases (search). You will want to read in context where those words occur; employ a concordance. You may want to count & tabulate more specific things, like only the people in a set of given documents; refine the results. You may want to create a timeline or plot things on a map (geolocate). You might want to employ some machine learning techniques that can extract latent themes and plot those themes over time (topic model). You may want to characterize some or all of your carrel into a set of nodes and edges, and then visualize the resulting network graph."

The thought experiments in Jennifer Guiliano's "Toward a Praxis of Critical Digital Sport History" might also be a starting point.

- Jennifer Guiliano, "Toward a Praxis of Critical Digital Sport History" *Journal of Sport History* 44:2 (2017): 146-159. [Link to online access via Hesburgh Libraries.](#)

Lab Notebook Components

Required Components

- Reflections/observations from Thursday 9/9 in-class work (Voyant Tools and Distant Reader)
- Reflections/observations from Thursday 9/16 in-class work (AntConc and Python)
- Reflections/observations from your own textual analysis experiment
- Any PDF or TXT files you worked with
- Any Python scripts/Jupyter Notebooks used in the lab

If you're working in or using Google Drive, you can download the project folder as a compressed zip folder and submit to Canvas. You can also include a link to the Google Drive folder as part of the lab notebook.

Template/outline for the report:

- Background or context for the research question or topic you're exploring
 - At least 500 words (or 5 minutes audio/video) on how the question you're asking or topic you're exploring is grounded in the texts we've been reading and conversations we've been having about ND football history and cultural significance.
 - To put it another way, this is a "background" section.
- Sources and research/analysis methods you'll be using
 - At least 200 words (or 2 minutes audio/video) on how the specific sources you're engaging with and research/analysis methods you're using animate or connect to the question(s) you're asking or the topic(s) you're exploring.
 - To put it another way, this is a "methods" or "methodology" section
- What you found through the exploration/analysis
 - At least 500 words (or 5 minutes audio/video) on what you found through analyzing this source material using specific research/analysis methods. Folks are encouraged to include visualizations/screenshots/figures/etc in this section.
 - To put it another way, this is a "findings" section
- Conclusions and next steps
 - At least 250 words (or 2-3 minutes audio/video) on how your findings relate to the broader context or background you outlined at the start of the report.
 - How can the work you did in this lab inform how we think about or understand ND football's history and cultural significance? How does it connect to the readings and conversations we've been having in class?

- Where would you go next with this research? What questions emerged through doing this work? What questions were you not able to address?
 - To put it another way, this is a hybrid "discussion" and "conclusion" section
- Any PDF or TXT files you created or worked with (that aren't already in the Lab 1 Google Drive)
- Any Python scripts/Jupyter Notebooks (or RStudio/RMarkdown files) used in the lab
- Individual reflections
 - At least 200 words (or 2-3 minutes audio/video) from each member of the group, addressing the following questions:
 - Your contribution to the lab report
 - Challenges you faced in this lab and how you solved them
 - What you learned about text analysis through this lab
 - How you're thinking about text analysis (and ND football history/primary sources) differently after this lab
 - Other comments/questions/observations