

DATA ANALYTICS FOR LIBERAL ARTS STUDENTS

George Mount
Stringfest Analytics
george@stringfestanalytics.com

I imagine this as a two-semester sequence for upperclassmen. The first course, “Intro to Data Analytics” is more of a general survey of data analytics with an emphasis on applications in the workforce. The second, “Data Analytics for the Liberal Arts” is geared toward students interested in graduate school in the humanities or similar. However, these topics would be of use to others as natural language processing has become popular in industry.

1. *Intro to Data Analytics*

This is a full-semester, 3-credit hour course. I consider this as a sort of “data analytics bootcamp” conducted as a typical semester-long college course.

Learning outcomes:

- Student can use multiple data analytics technologies in building a data analysis project
- Student can create visualizations and other data-backed assets using information design principles
- Student can conduct compelling presentations based on data analysis

Topics:

- What is data analytics? Why is it so popular? *Should* it be so popular? What can liberal arts students bring to the table? (This can be a bit contrarian and liberal arts triumphalist.)
- Basic spreadsheet modeling: be able to build a break-even pricing model or a staffing model. Almost anyone in business will have to build a budget at some point and it will probably be in Excel.
- Databases: Understand the basic architecture that powers modern business and how it’s changed in the last 15 years. Be able to write basic commands in structured query



language (SQL) to read from databases. Understand the difference between relational and non-relational databases, structured and unstructured data.

- Information design: Much of this is based on cognitive science. Learn when to use which chart versus tables, etc. How to design effective slide decks, visualizations and dashboards and use these assets to present sound data-backed presentations and recommendations. Pairs nicely with the liberal arts emphasis on sound communication.
- Coding: Perform basic task automation and data analysis with R.

Assessments:

- Build a break-even model for an ecommerce store. Present findings to management.
- Build an interactive dashboard investigating sales performance over time
- Perform basic data exploration and analysis of a relational database
- Conduct reproducible data analysis exercise and presentation using a series of data analytics tools

2. Data Analytics for the Liberal Arts

This is a full-semester, 3-credit hour course. The text could be [Text Analysis with R for Student of Literature](#) with some supplemental material. It may be wise to make “Intro to Data Analytics” a pre-req along with a basic stats course.

This course will be conducted entirely in R. The “target demo” would be students interested in grad school for the humanities or related, but the appeal could easily extend as natural language processing is becoming quite popular in industry.

Learning outcomes:

- Student can conduct exploratory data analysis of unstructured data sets
- Student can build reproducible data analysis projects in line with the [CRISP-DM](#) or similar methodology
- Student can create compelling visualizations to aid in their findings



Topics:

- Advantages and disadvantages of data analytics in the liberal arts
- Basic exploratory data analysis: describing, summarizing, visualizing one and multiple variables
- Unstructured data collection: APIs, web scraping
- What makes data “tidy” and why it matters
- Structured vs unstructured data
- Word frequency analysis
- Document similarity
- Document clustering & classification
- Topic modeling
- Network analysis

Assessments:

- Collect data from an API or the Web, explore and prepare it for quantitative analysis
- Visualize word frequencies and occurrences across the corpus of one or more authors
- Build a model to classify documents based on predicted original author
- Conduct network analysis on historical or literary figures

