CTS2000: Scripting for the Humanities

CALENDAR DESCRIPTION:
This course introduces students to core concepts in data creation, representation, and interpretation. Students will develop basic programming skills in Python as well as skills for ethically encoding and working with data.

Departments: Culture & Technology Studies, College of Arts

COURSE DESCRIPTION:
This course introduces students to core concepts in data creation, representation, and interpretation. In addition to thinking critically about the data as a form of representation, students will also analyse the ethical considerations of data and develop a number of models for the ethical contextualization of data. Students will learn to model data in software, understand the history of the concept of data, develop a data schema, understand differences between data and capta, and critique data as representation. Students will develop basic programming skills in Python programming languages. They will create data structures, use programming libraries to manipulate and work with data, and develop their own functions. The final project will combine data mining of social media or online archives while also reflecting on the construction of their dataset.

This course presumes no prior programming experience or knowledge of math. It is designed for programming novices.

LEARNING OUTCOMES
The successful completion of this course will enable the student to:
1.   Model information and forms of representation as data.
2.   Critically evaluate multiple forms of data.
3.   Discuss data structures and basic algorithms.
4.   Program in Python using datasets, data structures, libraries, and functions.
5.   Use Python to collect content, encode it as data, and analyze it.

REQUIRED TEXTS:
Students will require use of a computer that is less than five years old and is able to run Python. Students without access to a computer should contact the instructor to arrange borrowing a laptop through the College of Arts.

METHODS OF ASSESSMENT:

Participation 20%
Critical Data Study 15%
Data Schema Assignment 10%
Data History Analysis 10%
Python programming exercises 20%
Data Mining Project 25%

ASSIGNMENT DETAILS:

Participation (20%)

Students will take part in class discussions both during class lectures and labs as well as on the class Slack channel.

Critical Data Study (15%)
Students will choose a dataset from a selection, examine the dataset and answer the following questions:

- How is the dataset structured?
- What relevant information is missing?
- Why was the missing information excluded?
- How could the data be structured differently?
- What are three questions that you might ask of this data?

Data Schema Assignment (10%)
Students will be required to develop a data schema: a non-technical description of how a cultural object can be represented as a form of data.

Data History Analysis (10%)
Students will analyze a historical form of data and write a short analysis of the manner in which conceptions of race, gender, equity, and power are encoded in that data.

Python Programming Exercises (20%)
Students will be required to complete four (out of five) short Python programs that accomplish a range of tasks. Each exercise will introduce students to concepts in programming such as variables, functions, data structures and simple algorithms.

Data Mining Project (25%)
Students will develop a larger Python program that will acquire social media content, encode it as data, analyze data, and display their results in a combined visual and textual form.

READINGS:
Week 1-2 Histories of data, thinking critically about data
Johanna Drucker, "Humanities Approaches to Graphical Display."
Bruno Latour, "Visualization and Cognition: Drawing Things Together."
Exploring Datasets: Torn Apart

Week 3-4 Data and capta, race, indigeneity, and data
Niels Kerssens, "The Shifting Power of Metaphor in 1990s Data Mining."
ProPublica: "Machine Bias"

Week 5-6 Introduction to R and Python
Jennifer Giuliano and Carolyn Heitman, "Difficult Heritage and the Complexities of Indigenous Data"
Christen, "Does information really want to be free? Indigenous knowledge systems and the question of openness."
Introduction to Python
Introduction to R

Week 7-8 Big Data, working with large datasets
Understanding Twitter data structures
Sentiment analysis using R

Week 9-10 Data mining, linked open data
Andrea Zeffiro, "Towards a Queer Futurity of Data."
"Introduction to the Semantic Web"
Dominic Oldman et al, "Zen and the Art of Linked Data."
Using LOD: Wikidata, HuViz, Sparql queries

Week 11-12 Data analysis, displaying data visually