Objectives

Demonstrate the ability to perform a sufficiently difficult computational task using the Python programming language and the students’ knowledge of data structures and algorithms

Work collaboratively on a technically sophisticated project

Develop technical presentation skills
Task: Form a Team and Target

Working in teams, students must identify a topic of mutual interest

Examples:

- Demonstration of a machine learning algorithm
- Analysis of data
- Performing facial recognition on input images
- Working with SQL databases

Instructor shall assign orphans to a team
Task: Find the Data

For the selected topic, students must identify an appropriate data set.

Examples:
- UCI's Iris Machine Learning data set
- National Archives Fixed Format data
- Images of rabbits and kale found on the Internet
- An SQL database collected from Kaggle
- An English language text corpus in the NLTK
Data analysis tasks require the students explore the data and identify *several* interesting, unanswered questions lurking within

Examples using the COLED-V expenditure data:

• Several massive ammunition expenditure outliers appear in the data set. What’s the deal? When were they expended, under what conditions? Are these real?

• What was the 155mm M449 Ammunition type, how frequently was it used? Which units expended the most under combat and training situations?
Task: . . . Or the Goal

Machine learning, computer vision, and algorithm teams approach this by defining how they intend to use the data to accomplish the major objective.

Examples:

Using public domain images of street scenes and the OpenCV module, we shall identify the locations of faces in the image.

With the NLTK, we shall perform part of speech tagging on the collected works of Mark Twain (in public domain).
Constraints

Roughly five team members

Must work with an external data set or a collection of files on the file system

- CSV or JSON file collected from an Internet source
- Machine learning data set from UCI or other source
- Fixed format files collected from the National Archives
- Images stored in a local folder
- SQL Database
Delivery: Demonstration

Groups shall elect one or more team members to give a five to ten minute presentation and project demonstration.

Discuss:

The goal

Problems/successes during implementation

Final analysis

Teams *may* perform the demonstration on their personal computers:

- Large data sets
- Custom module installations
- Reduced Stress

The classroom PC and MAC computers remain accessible, but lack most useful modules.

Due 2019-MAY-09
Each team shall prepare a single document detailing:

- Team member names
- Data source files selected
- Any necessary modules (installed)
- Operational Instructions
- References
- Interesting observations

A hard-copy of this document is due at the start of class on 2019-MAY-09