

# PPOL564 Data Science I: Foundations

## Study Suggestions for Final

*Fall 2019*

### Test Structure

The final will be in two parts. Part 1 will be composed of multiple choice and short answer questions. Part 1 will be given in hard copy and must be completed and turned in before moving onto Part 2. Part 2 will be a coding problem. A Jupyter Notebook and data will be posted on CANVAS under the “Finals” Assignment tab. Students will submit a completed notebook to CANVAS before the end of the exam. A total of 2 hours are allotted for the final exam with 15 possible points.

#### Part 1

- 5 Multiple Choice (5 points)
- 5 Short Answer (5 points)
- Questions will be similar to those on the assignments.
- See below for “General Topics” to study.

#### Part 2

- 1 programming question (5 points)
- Coding task will be similar to those on the assignments and coding discussions.
- Multiple components to the question, e.g. clean data, build algorithm, run algorithm, interpret.
- Part 2 is “closed book” with respect to the class materials.
- *You will not be asked to do anything that you have not done before.*

### General Topics

#### Reproducibility + Version Control

- What are important components for a reproducible project?
- How does Git work?
- How does Github work?
- Know how Jupyter Notebooks work.

#### Programming

- Different data types and structures in Python.
- Iteration, control sequences, comprehensions, and generators.
- Functions and scope.
- File management.
- Using `Numpy` arrays for data manipulation (and linear algebra operations)
- Using `Pandas` data frames for data manipulation and data exploration.

## Linear Algebra

- Vectors and vector operations.
- Linear independence, orthogonality, and projections.
- Matrix operations and inversions
- Linear regression
- Eigen decomposition

## Calculus

- Derivatives
- Optimization (uni-/multi-variate)
- Gradient Descent
  - Know how the algorithm work;
  - Be able to implement a gradient descent algorithm.

## Probability

- Conditional probability
- Bayes theorem
- Probability distributions
- Expectation and variance
- Naive Bayes Classifier
  - Know how the algorithm work;
  - Be able to implement a Naive Bayes Classifier.