
# Department of Decision Sciences

**Course Title**: Programming for Analytics (DNSC 6211)
**Spring 2018**

## Course description

This is a foundation course on programming skills for data analytics. Programming is important for those who are serious about a career in business analytics. This course emphasizes and focuses on concepts, techniques and tools that will prepare you to apply data science techniques in the business domain. To that end, this course will help you develop skills primarily in *programming* as applied to *analytical techniques*. After taking this course, you will have a working knowledge of the analytics workflow (starting at the idea- stage and ending with creating basic, and hopefully compelling, interactive and automated scripts and programs). This class is designed so that you will 'learn by doing' individually as well as in groups, enabling you to practice the techniques with inputs from the instructor. We will be using R and Python and the recommended computing environment is Unix.

**Instructor**: Shivraj Kanungo
**Room**: Funger 415E
**Phone**: (202) 994-3735
**Email**: kanungo@gwu.edu

## Learning Objectives

Students who successfully complete the course requirements will be able to:

1. Write scripts and complete applications in Python and R
2. Recognize different sources of data and learn how to access those data
3. Visualize data using Python and R
4. Design develop and deliver reproducible data products in R and Python
5. Use Python and R for analytics workflow for descriptive, predictive and prescriptive analytics

## Course delivery

Each class session will include a lecture component and an instructor-led hands-on or implementation exercise. Students will work in teams and individually. Inside the class we will adopt the pair programming approach. Students are also permitted to work on their own in class.

## Course material

All course material will be provided; optional textbooks are recommended
The following book are recommended. They are available free online.

- For R I suggest that you refer to The Art of R Programming and R for Data Science.
- Learn Python the Hard Way is an excellent resource for Python. There is a nice Youtube supplement for this book.

## Software

Python, R and other open source software

## Grading framework

| Evaluation component | Weight | Due dates |
|---|---|---|
| Assignments | 10 | Weekly |
| Quizzes | 10 | Weekly |
| Individual Project (in R) | 25 | 3/11 |
| Group Project (Python) | 35 | 5/2 |
| Final exam (hands-on) | 20 | 5/9 |

## Grading scale

| Grade | Range | Grade | Range |
|---|---|---|---|
| A | 96-100 | C+ | 71-75 |
| A- | 91-95 | C | 66-70 |
| B+ | 86-90 | Fail | < 66 |
| B | 81-85 | | |
| B- | 76-80 | | |

# Weekly course coverage

| Week | Language | Topic | Asgn | Quiz |
|---|---|---|---|---|
| 01<br>0/17 | R | Introduction<br>□ Review the Syllabus<br>□ Lecture: Overview of R and its relevance to Business Analytics<br>□ Introduction: Reproducible data products in R<br>a. Review of basic R concepts (plots, functions, loops, conditions)<br>b. In-class exercises | Y | |
| 02<br>1/24 | R | Exploratory and rudimentary analysis<br>□ Review the assignment<br>□ Lecture: Advanced loops; vectorizing, subsetting<br>□ Exploratory/Descriptive analysis, Handling missing data, Titanic dataset<br>a. In-class exercises, simulation; base graphics | Y | Y |
| 03<br>1/31 | R | Data management and logistic regression<br>□ Review the assignment<br>□ Lecture: Data management<br>□ data analysis; ggplot<br>a. In-class exercises | Y | Y |
| 04<br>2/7 | R | Model-and data-driven analysis<br>□ Review the assignment<br>□ Lecture: Data driven analysis and workflow<br>□ Matrix approaches<br>a. In-class exercises (regression + logistic) | Y | Y |
| 05<br>2/14 | R | Model-and data-driven analysis<br>□ Review the assignment<br>□ Lecture: Decision Trees<br>□ Naïve Bayes<br>a. In-class exercises (decision trees, Naïve Bayes) | Y | Y |
| 06<br>2/21 | R | Interactive applications with Shiny and R<br>□ Review the assignment<br>□ Lecture: Shiny architecture<br>□ Matrix approaches<br>a. In-class exercises (regression + logistic) | Y | Y |
| 07<br>2/28 | Python | Interactive applications with Shiny and R<br>□ Review the assignment<br>□ Lecture: Overview of Python and its relevance to Business Analytics<br>□ Review of basic Python concepts (plots, functions, loops, conditions)<br>a. In-class exercises | Y | Y |
| 08<br>3/7 | Python | Pandas and data handling<br>□ Review the assignment<br>□ Lecture: Overview of Python and its relevance to Business Analytics<br>□ Review of basic Python concepts (plots, functions, loops, conditions)<br>a. In-class exercises (lists, list comprehensions and dictionaries) | Y | Y |
| 09<br>3/21 | Python | Regression workflow + logistic regression (+ classification) workflow<br>□ Review the assignment<br>□ Lecture: Data driven workflows for prediction<br>□Review of model evaluation<br>a. In-class exercises | Y | Y |
| 10<br>3/28 | Python | Web scraping<br>□ Review the assignment<br>□ Lecture: Document structure, tags<br>□ JSON and tags<br>a. In-class exercises | Y | Y |

| | | | | |
|---|---|---|---|---|
| 11<br><br>4/4 | Python | Clustering<br>☐ Review the assignment<br>☐ Lecture: Hierarchical and k-means<br>☐ Group project time<br>a. In-class exercises | Y | Y |
| 12<br><br>4/11 | Python | APIs and credentials; basic text processing; sentiment analysis<br>☐ Review the assignment<br>☐ Lecture: Authentication and credentials<br>☐ Parsing a tweet, and handling text data<br>a. In-class exercises | Y | Y |
| 13<br><br>4/18 | Python | Python objects and Prescriptive models<br>☐ Review the assignment<br>☐ Lecture: Linear optimization basics<br>☐ DEA example<br>a. In-class exercises | Y | Y |
| 14<br><br>4/25 | | Final presentations | | |

## Disability Support Services

Any student who may need an accommodation based on the impact of a disability should contact the Office of Disability Support Services (DSS) to inquire about the documentation necessary to establish eligibility, and to coordinate a plan of reasonable and appropriate accommodations. DSS is located in Rome Hall, Suite 102. For additional information, please call DSS at 202-994-8250, or consult www.disabilitysupport.gwu.edu.
For more information see https://disabilitysupport.gwu.edu/teaching-support

## Useful Resources

- Basic Linux commands
- The Comprehensive R Archive Network
- The Python official site

**Data sources**

- https://www.opensciencedatacloud.org/publicdata/
- http://www.kdnuggets.com/datasets/index.html
- http://datascience.berkeley.edu/open-data-sets/
- http://www.datasciencecentral.com/profiles/blogs/big-data-sets-available-for-free
- http://www.datasciencecentral.com/profiles/blogs/great-github-list-of-public-data-sets
- http://www.datascienceweekly.org/data-science-resources/data-science-datasets

**Competitions**

- https://www.kaggle.com/
- http://www.kdnuggets.com/competitions/
- http://www.drivendata.org/competitions/