**DATS 6101 - Introduction to Data Science**

**Spring 2017**

**Course and Contact Information**

Section: 11

Time: Mondays 7:10 – 9:40 pm

Location: Duques Hall 361

**Instructor**

Name: Brian Wright

Campus Address: Staughton Hall 304

Phone: 703-244-5573

Email: [bwright6@email.gwu.edu](mailto:bwright6@email.gwu.edu)

Office hours: Wednesday

**Course Description**

Called the “[Sexiest Job Title of the 21st Century](https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century)” by the Harvard Business Review, Data Science and analytics are a booming industry. But what is a Data Scientist, what do they do, and how do you become one? These questions and more will be discussed and answered in this introductory course. This class covers the basic ideas and techniques of data science, including its definition and the context in data-driven computation and practical applications.

**Learning Outcomes**

As a result of completing this course, students will be able to

1. Understand the fundamental principles of data science and the variety and types of data scientists
2. Understand the research process and lifecycle associated with answering data driven problems
3. Initial understanding of statistical models and how they are applied to data problems
4. Gain experience with the using R to do analyses

**Course Prerequisite**

You are expected to have a basic understanding of R coding, either through online resources or previous experiences.

**Tentative Schedule**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class Date** | **Reading (Before Class)** | **Topic** | **Activity** | **Assignment (Due following Week)** |
| **Jan 23rd** | Doing Data Science: Preface/Chapter 1  Art of Data Science: Sections 1 and 2  Blog Posts:   * [Thinking Like a Data Scientist](https://www.kaggle.com/forums/f/15/kaggle-forum/t/2947/thinking-like-a-data-scientist) * [The Only Skill you Should be Concerned With](http://www.datasciencecentral.com/profiles/blogs/the-only-skill-you-should-be-concerned-with) | **What is a Data Scientist?**  **Data Science Lifecycle** | [R Installed](https://cran.rstudio.com/)  [R Studio Installed](https://www.rstudio.com/products/rstudio/download2/)  Discussion/Ice Breaker | Find a Data Science Job  [Datacamp:](https://www.datacamp.com/)  Intro to R: 1&2 |
| **Jan 30th** | Art of Data Science Chapters 3  The Art of R Programming: Intro and Getting Started  Analyzing the Analyzers  Critical Questions for Big Data | **Research Process/Asking a Question/R Overview (Tidyverse/Dplyr)/ Exploratory Data Analysis I** | Discussion/Set Project Groups |  |
| **February 6th** | Art of Data Chapter 3  Doing Data Science: Chapter 2 | **Intro to Descriptive Statistics (EDA) II** | Lab: Descriptive Statistics/Guest Speaker | Assignment |
| **February 13th** | Open Intro to Stats 3.1, 3.2 and Chpt 4  Introduction to Stats Learning 2.1 and 2.2 | **Hypothesis Testing and Statistical Inference** | In Class Exercise | Assignment |
| **February 20th** | Intro to Stats Learning Chapter 3  Open Intro to Stats Chapter 7 | **Intro to Model Building: Prediction Regression I** | In Class Exercise | Assignment |
| **February 27th** | Open Intro to Stats Chapter 8 | **Intro to Model Building: Prediction Regression II (Outliers)** | Guest Speaker/Quiz |  |
| **March 6th** | **Project I Presentations** | | | |
| **March 13th** | **No Classes Spring Break** | | | |
| **March 20th** | Intro to Stats Learning 4.3  Open Intro to Stats 8.4 | **Model Building: Classification I: Logistic Regression/GLM** | In Class Exercise |  |
| **March 27th** | Intro to Stats Learning 39-42 | **Classification II: K-nearest neighbor** |  | Assignment |
| **April 3rd** | Intro to Stats Learning 10.1, 10.2 | **Feature Selection: Principal Component Analysis** |  |  |
| **April 10th** | Doing Data Science Chapter 9, 14 | **Big Data Engineering/ Communicating Data Model Results(Manipulate/ggplot)** | Guest Speaker  Quiz | Assignment |
| **April 17th** | Doing Data Chapter 18 | **Data Ethics** | Discussion |  |
| **April 24th** |  | **Course Review** | Guest Speaker |  |
| **May 1st** | **Final Projects Presentations** | | | |

**Textbooks and Course Resources**

No one book fully captures the variety of skills and knowledge being given in this course. Consequently, I have pre-selected chapters from a variety of sources. These will all be provided prior or during the course.

Please download R and Rstudio located here: <https://www.rstudio.com/products/rstudio/download/>

Begin to explore Data Science blogs, listservs, and journals. A few are listed below to help get you started:

* <http://www.datasciencecentral.com/>
* <https://www.analyticsvidhya.com/>
* <http://www.datatau.com/>
* <https://www.facebook.com/data?_rdr=p>
* <http://freakonomics.com/>
* <http://harvarddatascience.com/>
* <http://www.informationisbeautiful.net/blog/>
* <http://blog.kaggle.com/>
* <https://www.r-bloggers.com/>
* <http://blog.yhat.com/>
* <https://analytics.googleblog.com/>
* <http://datasciencemasters.org/>

Start networking now:

* Follow the following on Twitter for Predictive Analytics: [@DataScienceCtrl](https://twitter.com/DataScienceCtrl)**,** [@analyticbridge](http://twitter.com/analyticbridge), [@mgualtieri](http://twitter.com/mgualtieri), [@doug\_laney](http://twitter.com/doug_laney), [@Hypatia\_LeslieA](http://twitter.com/Hypatia_LeslieA), [@hyounpark](http://twitter.com/hyounpark), and [@anilbatra](http://twitter.com/anilbatra)
* Follow the following on Twitter for Big Data and Data Sciences: [Vincent Granvill](https://twitter.com/analyticbridge), [Alistair Croll](https://twitter.com/acroll), [Alex Popescu](https://twitter.com/al3xandru), [@rethinkdb](https://twitter.com/rethinkdb), [Amy Heineike](https://twitter.com/aheineike), [Anthony Goldbloom](https://twitter.com/antgoldbloom), [Ben Lorica](https://twitter.com/bigdata), [@oreillymedia](https://twitter.com/OReillyMedia)., [Bill Hewitt](https://twitter.com/Kalido_CEO), [Carla Gentry CSPO](https://twitter.com/data_nerd), [David Smith](https://twitter.com/revodavid), [David Feinleib](https://twitter.com/vcdave), [Derrick Harris](https://twitter.com/derrickharris), [DJ Patil](https://twitter.com/dpatil), [Doug Laney](https://twitter.com/doug_laney) - [Edd Dumbill](https://twitter.com/edd), [Eric Kavanagh](https://twitter.com/eric_kavanagh), [Fern Halper](https://twitter.com/fhalper), [Gil Press](https://twitter.com/GilPress), [Hilary Mason](https://twitter.com/hmason), [Jake Porway](https://twitter.com/jakeporway), [James Gingerich](https://twitter.com/jamesvgingerich), [James Kobielus](https://twitter.com/jameskobielus), [Jeff Hammerbacher](https://twitter.com/hackingdata), [Jeff Kelly](https://twitter.com/jeffreyfkelly), [Jim Harris](https://twitter.com/@ocdqblog), [Justin Lovell](https://twitter.com/JustinLovell), [Kevin Weil](https://twitter.com/kevinweil), [Krish Krishnan](https://twitter.com/datagenius), [Manish Bhatt](https://twitter.com/imbigdata), [Merv Adrian](https://twitter.com/merv), [Michael Driscoll](https://twitter.com/medriscoll), [Monica Rogati](https://twitter.com/mrogati), [Neil Raden](https://twitter.com/NeilRaden), [Paul Philp](https://twitter.com/pphilp), [Peter Skomoroch](https://twitter.com/peteskomoroch), [Philip (Flip) Kromer](https://twitter.com/mrflip), [Philip Russom](https://twitter.com/prussom), [Paul Zikopoulos](https://twitter.com/BigData_paulz), [Russell Jurney](https://twitter.com/rjurney), [Sid Probstein](https://twitter.com/sidprobstein), [Stewart Townsend](https://twitter.com/stewarttownsend), [Todd Lipcon](https://twitter.com/tlipcon), [Troy Sadkowsky](https://twitter.com/tsadkowsky), [William McKnight](https://twitter.com/williammcknight), [Yves Mulkers](https://twitter.com/YvesMulkers)

**Independent or Out-of-Class Learning**

Average minimum amount of out-of-class or independent learning expected per week is 5 hours, for our 2.5 hour of classroom meeting. With the knowledge of basic statistics that a student must possess to take this class, the terms “average” and “minimum” should require no further explanation.

**Learning Assignments**

**In-Class Exercise:** These assessments will be used to check learning and give feedback on areas for improvement. Reading prior to class, class attendance, and participation in activities are essential for success on this part of the course.

**Homework Assignments:** Details on requirements will be given during class periods. Most assignments will be due the next class period and can be submitted via blackboard. We will work to provide feedback at the next class session.

**Class Projects:** Two group projects will be assigned over the semester to give students practice on applying data science principles and methods to various problems. Students will also build teamwork, communication, and technical skills.

**Quizzes:** Two quizzes will be given to ensure students are gaining knowledge and doing out of class readings

**Grading**

Your final grade will be determined by:

* Assignments (25%)
* Quizzes (25%)
* Project I (25%)
* Project II (25%)

**UNIVERSITY POLICIES**

**Observance of Religious Holidays**

In accordance with University policy, students should notify faculty during the first week of the semester of their intention to be absent from class on their day(s) of religious observance.   
For details and policy, see: [students.gwu.edu/accommodations-religious-holidays](http://students.gwu.edu/accommodations-religious-holidays).

**Academic Integrity Code**

Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information.   
For details and complete code, see: [studentconduct.gwu.edu/code-academic-integrity](http://studentconduct.gwu.edu/code-academic-integrity)

**Safety and Security**

In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.

**Support for Students Outside the Classroom**

*DISABILITY SUPPORT SERVICES (DSS)*

Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 in the Rome Hall, Suite 102, to establish eligibility and to coordinate reasonable accommodations.   
For additional information see: [disabilitysupport.gwu.edu/](http://disabilitysupport.gwu.edu/)

*MENTAL HEALTH SERVICES* ***202-994-5300***

The University's Mental Health Services offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include: crisis and emergency mental health consultations confidential assessment, counseling services (individual and small group), and referrals.   
For additional information see: [counselingcenter.gwu.edu/](http://counselingcenter.gwu.edu/)