# UNIVERSITY OF MARY WASHINGTON -- NEW COURSE PROPOSAL

Electronically submit this completed form with attachments in one file to the Chair of the College Curriculum Committee.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Arts and Sciences** | | **Business** | X | **Education** | | |
| Proposal Submitted By: Chris Garcia (with help of Ron Zacharski) | | | Date Prepared: 28 November 2016 | | | | |
| Course Title: | Introduction to Data Science | | | | | | |
| Department/discipline and course number*: | DSCI 101 (to be cross-listed as DATA 101) | | | | | | |
| Prerequisites: None | | | | | | | |
| **This course number must be approved by the Office of the Registrar before the proposal is submitted.* | | | | | | | |
| Number of credits: | 3 | Will this course meet for at least 700 contact minutes for each credit hour proposed? ***If no, provide a credit hour justification.*** | **YES** X | | **NO** | | |
| Will this be a ***new***, *repeatable* "special topics" course? (Do you want students to be able to take this new course more than once if the topic changes?) | | | **NO** X | | **YES** | | |
| | | | | | | | |
| Date of first offering of this ***new*** course: **FALL SEMESTER, year** | | | FALL 2017 | | | | |
| Proposed frequency of offering of the course: | | | every semester | | | | |
| List the faculty who will likely teach the course: | | | Chris Garcia, CPSC faculty | | | | |
| **Are ANY new resources** required? | **NO X** | | **YES** | | *Document in attached impact statement* | | |
| | | | | | | | |
| **This new course will be (check all that apply):** | | | | | | | |
| Required in the major | | Required in the minor | | General Elective | | | |
| Elective in the major | | Elective in the minor | | General Education** | X | | |
| ***AFTER the new course is approved, a separate proposal must be sent to the General Education Committee.* | | | | | | | |

**Catalog Description** (suggested length – less than 50 words)**:**

A gentle, hands-on introduction to the field of Data Science and its applications. Covers a wide range of topics to provide an overview of the use of data in different fields. Provides hands-on practice with basic tools and methods of data analysis. Prepares students to use data in their field of study and in their work and to effectively communicate quantitative findings. Cross-listed as DATA 101.

| **COURSE HISTORY:** | Was this course taught previously as a topics or experimental course? | **YES** | | **NO X** | |
|---|---|---|---|---|---|
| **Course Number and Title of Previous Course** | | **Semester Offered** | | **Enrollment** | |
| | | | | | |
| | | | | | |
| | | | | | |

| | |
|---|---|
| | **CHECK HERE** if the proposed course is to be ***equated*** with the earlier topics or experimental offerings. If equated, students who took the earlier "topics" course will only be able to take the new course as a repeat (C- grade or lower). |
| ***NOTE:*** If the proposed course has not been previously offered as a topics or experimental course, **explain in the attached rationale statement** why the course should be adopted even though it has not been tried out. | |

**REQUIRED ATTACHMENTS:**

1. **Rationale Statement** – Why is this course needed? What purposes will it serve?

2. **Credit Hour Justification** (if required) – explain how this course will comply with the UMW Credit Hours Policy (D.5.1)
3. **Impact Statement** – Provide details about the Library, space, staffing, budget, and technology impacts created by adding this new course. Include supporting statements from the Library, IT Department, etc. ***Any change that impacts another Department must have a written statement (such as a copy of an email) from the Chair(s) agreeing to the change.***
4. **Sample Syllabus**

**Department Chair Approval:**_____ **Date:**_____

**CCC Chair Approval:**_____ **Date:**_____

**UCC Chair Approval:**_____ **Date:**_____

**Rationale Statement**

(Why is this course needed? What purposes will it serve?)

The Virginia Governor's Office summarizes the need this way:

> For the past several decades, the application of data science and analytics in scientific fields, including physics and biology, has unlocked new knowledge from large data sets. However, the recent explosion in the amount of data available across a wide range of applications, and the development of new tools to access and analyze data, have created exciting new opportunities for intellectual exploration in other traditional liberal arts fields, including economics, sociology and psychology. As such, data science and analytics has the potential to engage an expanded pool of liberal arts faculty and students in innovative, cross-disciplinary education and research that could enhance and enrich their academic areas, and offers students avenues to develop new types of quantitative skills as they pursue their degrees.[1]

The Governor's Office is strongly encouraging the development of data science skills in both STEM and non-STEM graduates in the commonwealth's liberal arts colleges.

Our objective for this course is to provide an introduction to data analytics to students in a breadth of majors and this objective is aligned with the Commonwealth's goal of creating data-enabled graduates regardless of discipline. The skills learned in this proposed course will be directly applicable to students both in their coursework and on the job. Because we believe that integrating data science into traditional liberal arts programs will expand the occupational paths that are available to our students, our goal is to help as many students as possible develop basic data analytics skills. Our desire is to develop 'T-shaped' graduates who possess both a deep understanding of their discipline and the analytical skills necessary to work with people from other disciplines (this is a learning outcome promoted by the Association of American Colleges and Universities' LEAP program).[2]

This course will serve as the introductory course for the Data Science minor, and as a prerequisite for DATA 219, the second course in the minor's skill-building sequence. DATA 101 will contain two main

[1] background statement of the Commonwealth Data Analytics and Humanities Summit October 17-18, 2016

[2] Lumina Foundation. 2014. The Degree Qualification Profile 2.0. Indianapolis, IN. https://www.luminafoundation.org/files/resources/dqp.pdf

topics: (1) an overview of the Data Science field, so that students can build a cognitive framework for how the various activities in the discipline fit together, and (2) practical experience with fundamental Data Science skills, including Python programming and working with various common data formats. Students will exit the course with a grounding in what Data Science can accomplish and what minoring in Data Science would entail. We are also interested in getting the course QR designated and some sections honors designated.

**Impact Statement**
(Provide details about the Library, space, budget, and technology impacts created by adding this new course. Include supporting statements from the Library, IT Department, etc. as needed.)

The college has the equipment and facilities currently on hand needed to teach this course. The course does not require additional staff time from COB. The library has the necessary material to support the course.

## Sample Syllabus – DATA 101 Introduction to Data Science

Welcome to DATA 101, the introductory course in the Data Science minor! In this course, you will get a broad overview of one of the newest and most exciting fields today: the science of working with data of all shapes and sizes. You'll learn the fundamentals of the ways data is commonly structured and how it should be interpreted; what kinds of domains it can represent; how it can be visualized, processed, and analyzed; how to use basic Data Science tools (including the Python programming language) to work with data; and some fundamental techniques for extracting knowledge from data, including hypothesis testing, clustering, and classification.

By the end of this course, you should have a very good idea of what the discipline of Data Science is all about, and what its strengths and limitations are. You should know enough to determine whether you want to continue in future courses in the minor, and you should know where the content areas of those future courses fit in to the big picture.

Lectures: MWF 10-10:50 HCC327
General Structure:
Friday lectures are designed to provide an overview of the different areas of data science and the use of data science in a wide range of disciplines. For example, we will hear how data is used in biology, international relations, and linguistics.

The Monday and Wednesday classes are guided labs (short lecture followed by hands-on work on interactive worksheets) designed to have you develop basic skills in analyzing data and effectively communicating your results.

Topics

- Overview of data science (2 weeks)
  - Overview
  - Examples of how "Big Data" is used in a number of disciplines
- Data analysis tools (2.5 weeks)
  - Python, Anaconda, Jupyter Notebooks
  - Getting started with Python
  - Gentle introduction to Pandas/Numpy
- Describing data (2.5 weeks)
  - Determining data types
  - Understanding numerical data (mean, median, standard deviation)
  - Normalization
  - Basic probability
  - Reading data from csv file
  - Using descriptive statistics to develop and support an argument
- Visualizing data (2 weeks)
  - The "grammar of graphics" and graphical perception (position < length < angle <direction < area < volume < saturation < hue)
  - Types of charts
  - (Im)proper use of scales
  - Visualization tools: ggplot, Bokeh
  - Using visualization to develop and support an argument
- Hypothesis testing (2 weeks)
  - Welch's t-test
  - Using hypothesis testing to develop and support an argument
- Clustering (2 weeks)
  - "unsupervised learning"
  - a broad qualitative understanding of your data
  - clustering algorithms (k-means, hierarchical)
  - dimensionality reduction techniques (Principal Component Analysis - PCA)
  - understanding significant differences in your data through clustering and PCA
- Classification (2 weeks)
  - "supervised learning"
  - classification algorithms: k-nearest neighbors

Grading

- Topic unit labs – 20%: These are typically done in class (or at least started in class). Each unit consists of a short lecture followed by working on an interactive worksheet. Points are based on completion of a worksheet. The worksheets are designed to be completed in two hours.

- Rudiments – 30%: There are 8 rudiments, each of which should roughly take you one hour to complete. These typically involve using what you learned in a lab on a different data set.
- Projects – 25%: Five projects will involve both analyzing data and communicating the results to support an argument. For example, one project involves making a single presentation slide that makes an argument and supports it with data.
- In-class challenges – 5%: Throughout the semester the instructor will present very short in-class "challenges."
- Final exam – 20%: Open-notes, open-book, Dec 11th, 8am

## Late work policy

Work must be completed on time to receive full credit. If you miss a deadline by a week or less, 10% of the points will be deducted. If work is over 1 week late 20% of the points will be deducted.

## Accommodations for students with special needs

Any student with a documented disability may receive a special accommodation to complete any requirements of this course. If you are have a disability or believe you have one you may wish to self-identify. You may do so by providing documentation to the Office of Disability Services located in Room 203 of George Washington Hall (Phone: Voice 540-654-1266, Fax: 540-654-1163). Appropriate accommodations may then be provided for you. If you have a condition that may affect your ability to exit the premises in an emergency or that may cause an emergency during class, you are encouraged to discuss this in confidence with me and/or anyone at the Office of Disability Services. This office can also answer any questions you have about the Americans with Disabilities Act (ADA).

## Academic Integrity

I assume you are an ethical student and a person with integrity. I expect that you will follow the university honor code and the Computer Science Department's Honor Policy (see http://cs.umw.edu/mediawiki/index.php/CPSC_Department_Honor_Code_Policy). Please use common sense and ask yourself what would a person with integrity do?

## Class participation

I expect students to attend classes regularly. Since you will be spending the majority of class time working on projects, if you miss a class you will miss the opportunity to earn points. That said, attendance is not taken and no points will be awarded based directly on attendance. If you are going to miss a class, please be courteous and inform me and your teammates.

## General Education Student Learning Outcomes

- Students will demonstrate an ability to interpret quantitative/symbolic information..
- Students will have the ability to convert relevant information into various mathematical/analytical forms.

- Students will be able to apply analytical techniques or rules to solve problems in a variety of contexts.
- Students will gain an appreciation for how analytical techniques or rules are used to address real-world problems across multiple disciplines.