**COURSE:** CE 599-002 Data Science in Transportation (Spring 2018)

**TIME and PLACE:** Tuesday and Thursday at 12:30 pm - 1:45 pm
OHR C053 (Basement of Raymond Building)

**INSTRUCTOR:** Dr. Greg Erhardt, 261 OHR, 323-4856, greg.erhardt@uky.edu

**OFFICE HOURS:** Monday 9-10 am or by appointment.

**PREREQUISITES:** Introductory course in computer programming, such as CS 115, CS 221 or EGR 102.
Introductory course in statistics, such as STA 381.

**NOTE:** Most of the communication for this course will be performed through email; check your email **at least once a day**!

**COURSE CONTENT AND OBJECTIVES:**

This course is designed around the Data Science Venn Diagram, as shown in **Figure 1**. It takes applications from the transportation realm, and introduces the practical skills needed to pursue data science both in the workplace and as a research student.

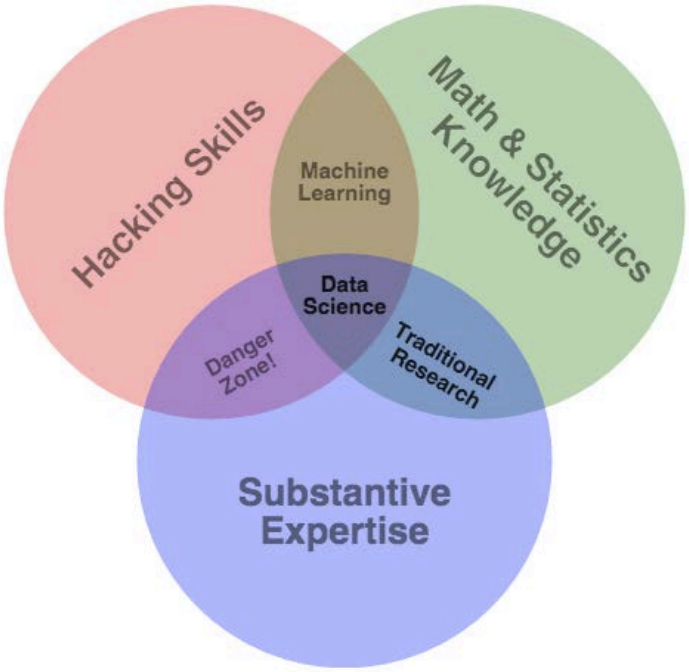


***Figure 1 Data Science Venn Diagram[1]***

Main topics to be covered include:

- Fundamentals of programming and data wrangling in Python
- Data visualization
- Applied statistical modelling and interpretation
- Identifying and perpetuating intellectually honest analysis.

1 Drew Conway, "The Data Science Venn Diagram," *Drew Conway*, accessed October 26, 2016, http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram.

## COURSE COMPONENTS, REQUIREMENTS, AND GRADING:

### Github

Github will be used for communication and posting course content outside of lectures, and for assignment submissions. Please be sure you are checking emails and messages sent through Github.

### Readings

Two textbooks are required for this class:

Downey, Allen B. *Think Python: How to Think Like a Computer Scientist*. 2 edition. Sebastopol, CA: O'Reilly Media, 2015. ISBN: 978-1-4919-3936-9

Note: This is available under a Creative Commons license, and can be downloaded as a PDF for free from here: http://greenteapress.com/wp/think-python-2e/

McKinney, Wes. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 1 edition. Beijing: O'Reilly Media, 2012. ISBN: 978-1-4493-1979-3

Note: There is a PDF version and e-book that are available, and are suitable.

An additional book is required reading and will serve as the a foundation for the mid-term report, as described below. Other sources will be identified in class and will be used as references as needed.

### Class Participation and Exercises

Class attendance and participation is required, except in the cases described in the attendance policy below. Classes will require the active completion of specific programming exercises in Python. In some cases, these exercises will be completed in class, and in other cases they will be assigned as homework. In all cases, students will be required to successfully complete the exercise prior to the next class.

Questions and comments from students are welcome at any time during class. We are here to help and want to see you succeed.

### Mid-Term Report

Students will write a mid-term report in which they will recommend mechanisms to identify and perpetuate intellectually honest analysis. Reports should start from a recognition of the issues in transportation forecasting. A book, and supplementary readings will be assigned as a starting point for these reports. Further details of the assignment will be provided during the course.

### Final Project

A semester project is part of the course requirements. Students will be required to apply the skills they have learned to a problem of interest in transportation. The project should draw from the three areas of data science.

Undergraduate students will select from one or more topics identified by the instructor. Their submission will be a software product.

Graduate students will select their own topic and prepare a final report. The format of the report will be that of a research paper, and will follow the Transportation Research Board (TRB) guidelines. High-quality papers will be recommended for submission to the TRB Annual Meeting. A formal project proposal and at least one interim report will be required to ensure sufficient progress.

Students will also present the results of their project orally during class time. Further details of both components will be provided during the course.

**Grading**

| | | |
|---|---|---|
| Attendance and Exercises | 50% | 90 - 100 = A |
| Mid-Term Report | 15% | 80 - 89 = B |
| Final Project | 30% | 70 - 79 = C |
| Final Presentation | 5% | 60 - 69 = D |
| | 100% | <60 = E |

**Appeals:**

After each assignment has been returned, you will have one week to review it and plead your case for a grade change. Beyond the one-week period no changes will be made.

**CODE OF CONDUCT**

Students are expected to live up to the principles of the University of Kentucky creed:

- I promise to strive for academic excellence and freedom by promoting an environment of creativity and discovery.
- I promise to pursue all endeavors with integrity and compete with honesty.
- I promise to embrace diversity and inclusion and to respect the dignity and humanity of others.
- I promise to contribute to my University and community through leadership and service.
- I promise to fulfill my commitments and remain accountable to others.

The student code of conduct, along with the policies of the university and the College of Engineering, puts these principles into practice. Details of the code of conduct can be found at http://www.uky.edu/StudentAffairs/Code/part1.html.

Any form of academic dishonesty will not be tolerated. Bullying, harassment, acts of hate, or discrimination on the basis of race, sex, religion, national origin, age, disability status or sexual orientation will not be tolerated.

**ATTENDANCE POLICY:**

Students are required to attend all classes and participate in all fieldwork associated with assignments. Excusable absences include: (1) illness of student or serious illness of an immediate family member; (2) the death of a member of the student's immediate family; (3) trips for members or student organizations sponsored by an academic unit, trips for university classes, and trips for participation in intercollegiate athletic events; and (4) major religious holidays. Students are responsible for notifying the instructor in writing (NO EMAILS) prior to an event in categories (3) and (4). The instructor has the right to request verifications for all cases. Such verification needs to be provided by the student within two weeks of the absence in question.

**ACKNOWLEDGEMENTS:**

Many of the exercises are based on those developed by Paul Waddell and Geoff Boeing for CP255: Urban Informatics and Visualization at the University of California at Berkeley. Those materials are available from:

https://github.com/waddell/urban-informatics-and-visualization

**Course Schedule (subject to revision):**

| Date | Topic |
|---|---|
| Jan 11 | Intro to Jupyter and Python |
| Jan 16 | Python Data Types, Version Control with Git |
| Jan 18 | Programming Logic |
| Jan 23 | Working with Data Files, Data Cleaning |
| Jan 25 | Pandas Basics, Part 1 |
| Jan 30 | Pandas Basics, Part 2 |
| Feb 1 | Introduction to Open Data APIs |
| Feb 6 | Resources: Open Data, Open-Source Software and StackExchange |
| Feb 8 | Data Wrangling with Pandas |
| Feb 13 | Interactive Mapping |
| Feb 15 | Spatial Analysis |
| Feb 20 | Group, Split, Apply, Combine |
| Feb 22 | Version Control with Multiple Users: Webpage Example |
| Feb 27 | Group Work: Class Webpage |
| Mar 1 | Data Visualization, Part 1 |
| Mar 6 | Data Visualization, Part 2 |
| Mar 8 | Project Proposals and Selection |
| Mar 13 | **Spring Break – No Class** |
| Mar 15 | **Spring Break – No Class** |
| Mar 20 | Beyond Notebooks: Software Design and Classes |
| Mar 22 | Agent-Based Modeling |
| Mar 27 | Network Models |
| Mar 29 | Agent-Based / Network Modeling Example |
| Apr 3 | Regression Modeling, Part 1<br>**Mid-Term Report Due** |
| Apr 5 | Regression Modeling, Part 2 |
| Apr 10 | Discrete Choice Modeling |
| Apr 12 | Time-Series Modeling |
| Apr 17 | Correlation, Causality and Co-linearity |
| Apr 19 | Detecting Bull**** |
| Apr 24 | Ethics in Data Science<br>Course Evaluations and Wrap-Up |
| Apr 26 | **Final Presentations** |
| May 4 | **Final Project Report Due** |