CE 599

Data Science for Transportation

Spring 2021
3 Credits
Time: Tuesday and Thursday at 12:30 pm - 1:45 pm
Location: OHR C053 (Basement of Raymond Building)
Zoom: https://uky.zoom.us/j/2204743066

Instructor Information

Instructor: Professor Greg Erhardt
Office Building & Room Number: OHR 261
Email: greg.erhardt@uky.edu
Mobile Phone: (859) 699-1761
Office Hours: Monday 10-12 am, after class, or by appointment

Either in person or on Zoom: https://uky.zoom.us/j/2204743066

Course Description

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.

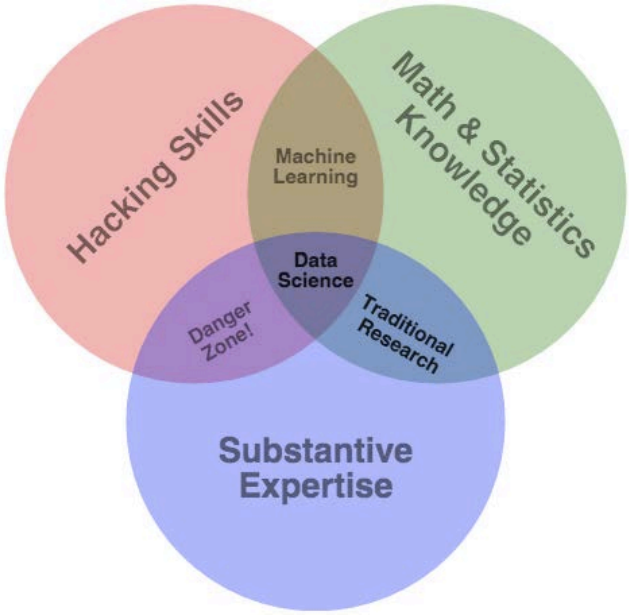

Figure 1 Data Science Venn Diagram[1]

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.

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.

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

Format

This will be an in-person COVID-adaptive course. We will follow the risk levels defined by the COVID-19 incidence rates for Fayette County, as reported on the Kentucky COVID-19 dashboard. The expected participation levels are:

- Green or Yellow (<=10 per 100k): Fully in-person.
- Orange (>10 to 25 per 100k): Instructor will provide in-person lectures, and students will have the choice to participate in-person or online.
- Red (25+ per 100k): Lockdown – online until infection rates come down.

These levels and details may be adjusted based on health considerations and teaching needs. Expect a re-evaluation 2 weeks in and at the mid-term. Students will be notified of any changes and are expected to participate accordingly. Masks and social distancing are required for all levels.

At all levels, this remains an interactive class. Students are required to bring a laptop computer to class and install the required software. If you need access to a computer, please contact the instructor. You should have your camera turned on during class, anticipate asking questions and being called upon to answer questions, and anticipate being asked to share your screen to show your solutions to coding problems.

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

Course Materials

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.

One book will be required reading during this class:

> Bergston, Carl T., West, Jevin D., Calling Bull****: The Art of Skepticism in a Data-Driven World. Random House, 2020. ISBN: 978-0525509189

Two reference books are recommended for this class. We will not follow these closely, but many students find them helpful:

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.

Additional readings and assignments may be required during the course.

Course Components

Class Participation and Exercises

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.

Mid-Term Report

Students will write a mid-term report in which they will recommend mechanisms to identify and perpetuate intellectually honest analysis. 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 Transport Findings guidelines. A formal project proposal and at least one interim product 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	60%	90 - 100 = A
Mid-Term Report	10%	80 - 89 = B
Final Project & Presentation	30%	70 - 79 = C
	100%	60 - 69 = D
		<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.

Tentative Course Schedule

The expected course schedule is below. All components are subject to change.

Date	Topic
Jan 26	Intro to Jupyter and Python
Jan 28	Python Data Types, Version Control with Git
Feb 2	Programming Logic
Feb 4	Working with Data Files, Data Cleaning
Feb 11	Pandas Basics, Part 1
Feb 16	Pandas Basics, Part 2
Feb 18	Introduction to Open Data APIs
Feb 23	Resources: Open Data, Open-Source Software and StackExchange
Feb 25	Data Wrangling with Pandas
Mar 2	Interactive Mapping
Mar 4	Spatial Analysis
Mar 9	Group, Split, Apply, Combine
Mar 11	Data Visualization, Part 1
Mar 16	Data Visualization, Part 2
Mar 18	Regression Models, Part 1
Mar 23	Regression Models, Part 2
Mar 25	Correlation, Causality and Co-linearity
Mar 30	**Project Proposals**
Apr 1	Beyond Notebooks: Software Design and Classes, Part 1
Apr 6	Beyond Notebooks: Software Design and Classes, Part 2
Apr 8	Version Control with Multiple Users
Apr 15	Simulation Models, Part 1
Apr 20	Simulation Models, Part 2
Apr 22	Networks, Part 1
Apr 27	Networks, Part 2
Apr 29	Optimization Models, Part 1
May 4	Optimization Models, Part 2
May 6	Ethics in Data Science
May 11	**Final Presentations**
May 13	**Final Project Report Due**

Technology Information

Minimum technical requirements for UK courses and suggested hardware, software, and internet connections are available at ITS Student Hardware & Software Guidelines.

Students should have a computer with a webcam, microphone, and stable internet connection suitable for participation in an online class. Students should sign up for their own Zoom account for use in interacting with small groups and homework partners. All UK students receive a free Zoom Pro account. For collaborative problem solving, I recommend that you use your phone as a document camera, or that you use a drawing program such as the Zoom whiteboard or Google Jamboard in conjunction with a tablet.

For account help, contact UK's Information Technology Customer Services online, by email, or by phone at 859-218-HELP (4357). For general technical support, contact Engineering Computing Services via their request form.

Attendance

Students should attend all classes and participate in all field work associated with assignments. Excused absences are allowed in accordance with *Senate Rules 5.2.5.2.3.3*, described below. Students who do not feel well, who have a fever, who are under a quarantine order, or who may have been exposed to COVID-19 should self-isolate and avoid coming to class. They should notify the instructor as soon as possible. When practical, they should participate remotely and when not practical, a suitable make-up assignment will be arranged.

Behavior

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, acts of hate, or discrimination on the basis of race, sex, religion, national origin, age, disability status or sexual orientation will not be tolerated. Masks and social distancing is required for any in-person interaction. Appendix A lists detailed class and university policies.

Resources

Appendix B lists relevant resources, including mental health resources. The Writing Center may be of value in this class.

Please ask if you need help academically, professionally, personally, or in any way. Your professors, advisors, teaching assistants, many of your peers are here to help. We many not have

the answers, but we really do care about your well-being, and will do our best to find those who might.

Acknowledgments

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