

INDIANA UNIVERSITY DATA SCIENCE SPRING 2016 SEMESTER ADVISING NOTES

David Wild - updated January 2016

GENERAL NOTES

General information about the program can be found on the website at <http://datascience.soic.indiana.edu>. All general questions, and questions about matters relating to student status and financial matters, should be directed to:

IU Data Science Graduate Office, 611 N Park, Bloomington, IN 47401. Email datasci@indiana.edu phone 812.856.5953
The office is managed by **Kayla Scroggins**, Data Science Graduate Services & Records Manager

ACADEMIC ADVISING

Questions regarding advising should be sent to the dsadvise@indiana.edu email address. This will get to the two Data Science faculty advisors, Professor David Wild and Professor Ying Ding.

Dr David Wild, Director, Data Science Academic Programs and Associate Professor, is the primary advisor for residential students. You can book an advising session in semester time with him on Tuesdays or Wednesday mornings at <https://djwild.youcanbook.me/>. Please use these sessions sparingly. Residential students should meet in Informatics West 207 (901 E 10th St); if you would like an online meeting, check the box in the appointment and a link will be provided for the call.

Dr Ying Ding, Associate Director, Data Science Online Programs and Associate Professor, is the primary advisor for online students. She is available in her office hours for physical or online meetings Monday and Friday 9-3:30PM open office (drop by anytime). Her physical address is the Main Library LI02. For online meetings she can be contacted on Skype (skype id: ying_ding). Otherwise, send me email to schedule an appointment.

PATHS AND TRACKS

There is understandably some confusion about *Paths* and *Tracks*. A path is not a formal part of your degree, but a guide to help you choose courses that are right for you. Courses are classed into *Decision Maker* or *Technical* paths. Decision maker courses are focused on the skills needed by data science decision makers such as utilization of data science techniques, social factors, and domain-specific applications. Technical courses are focused on the technical side of data science, often requiring strong programming skills. You do not have to “declare” a path for your degree, and being on one path doesn’t mean you can’t take courses from another path. In the course listings below, we tag the classes with paths (note that these tags are currently provisional, and some may change). A track is more formal specialization, and the track determines which courses you can

take. The default is the *general* track, in which you can pick whichever courses you wish. Currently we have just one specialization track, *computational & analytic*, the requirements of which are listed at the end of this document. We may add more tracks in the future.

WHICH COURSE SECTIONS DO I SIGN UP FOR?

Many data science courses have multiple sections associated with them. These are generally for different classes of students, and you should take care to sign up for the correct section.

Online students. Students registered for the online programs (certificate, MS) should sign up for the section with "Above class for students not in residence on the Bloomington campus". It should also say "Above class taught online". If you see more than one onlinesection, go for the one that says "For data science students only"

Residential students taking online class. For most online classes, residential students who are taking an online class should sign up for the section that says "Above class taught online" *but not* "Above class for students not in residence on the Bloomington campus". You should also be automatically signed up for a residential discussion section. For a few classes, this separate section may not exist, in which case you should sign up for the regular online section (but note that students on visas can normally only take one such "truly online" class per semester)

Residential students taking a residential class. Should sign up for normal section without the additional qualifiers. If you see more than one section, go for the one that says "For data science students only".

ONLINE COURSES - upcoming semester classes highlighted in green

| Course | Next Class | Instructor | Path | Notes |
|---|-------------|-------------------|----------------------------|--|
| INFO I571 Introduction to Cheminformatics | TBD | Wild | Decision Maker | Website: http://i571.wikispaces.com |
| INFO I590 Data Science in Drug Discovery, Health and Translational Medicine | Spring 2016 | Wild | Decision Maker | Website: http://dsdht.wikispaces.com Prerequisites: Ability to perform basic statistical tasks in R; conceptual understanding of machine learning. |
| INFO I590 Management, Access, and Use of Big and Complex Data | Fall 2016 | Plale | Decision Maker & Technical | Website http://datamanagementcourse.soic.indiana.edu/ |
| INFO I590 Big Data Applications and Analytics | Fall 2016 | Fox | Decision Maker | Some programming experience required, Python preferred. Website https://bigdatacourse.appspot.com/preview |
| INFO I590 Perspectives in Data Science | Spring 2016 | Stirling | Decision Maker | This course will introduce multiple perspectives of the application of data science through recorded interviews with leaders in Silicon Valley companies, and map these to the practical skillsets of the data scientist |
| INFO I590 Big Data Open Source Software and Projects | Spring 2016 | Fox / Abdul-Wahid | Technical | Familiarity in Scripting Languages such as Linux Shell and especially Python. Knowledge of Java helpful but not required. |
| INFO I526 Applied Machine Learning | Fall 2016 | Natarajan | Decision Maker & Technical | Rescheduled from Spring 2016. Requires programming ability in any one of the following programming languages – C / C++/ Java/Python/ Matlab. Basic algebra and probability knowledge. |
| CSCI B649 Cloud Computing for | Spring 2016 | Qiu / Abdul-Wahid | Technical | This is a programming intensive course. It has |

| | | | | |
|--|-------------|---------------------|----------------------------|--|
| Data Intensive Sciences | | | | similar requirements to the CS graduate level residential version. Students are expected to have weekly (or biweekly) programming homework. General programming experience with Windows or Linux using Java (2-3 years) and scripts is required. A background in parallel and cluster computing is a plus, although not necessary. |
| CSCI B649 High Performance Computing | Fall 2016 | Sterling | Technical | Intermediate C/C++ experience Familiarity with Linux/Unix command-line utilities |
| ILS Z636 Data Semantics | Spring 2016 | Ding | Technical | Basic knowledge of HTML and XML is necessary. Basic knowledge of Java can be helpful. |
| ILS Z637 Information Visualization | Spring 2016 | Börner / Ginda | Decision Maker & Technical | |
| ILS Z604 Social and Organizational Informatics of Big Data | TBD | Rosenbaum & Fichman | Decision Maker | |