Data Visualization

Yong-Yeol (YY) Ahn
yyahn@iu.edu

Summer 2018

Assistant Instructors
Silvia Karim (sikarim@iu.edu)
Abu Saleh Md Noman (amdnoman@indiana.edu)

COMMUNICATION

Announcements and other communication will be through Slack:

https://iu-dviz-course.slack.com

Therefore, you should join as soon as possible. You can join with your IU email adddress by visiting the following link.

https://join.slack.com/t/iu-dviz-course/signup

COURSE DESCRIPTION

From TV news to cutting-edge scientific papers, from a home office to the largest companies in the world, data visualization is extensively used to reveal patterns in data and to tell stories. More and more data is collected through devices and services, and more and more decisions are made through data. Data visualization is indispensible for data analysis, and thus is an essential skill for every knowledge worker. This course is an introduction to basic statistical data analysis and visualization. We will

learn fundamentals of data visualization in the context of perception, integrity, design, statistics, types of data, and visualization techniques. The hands-on exercises using the Python stack will be an integral part of the course.

Relationships with S637 Information Visualization (IVMOOC): Compared with S637, this course is geared more towards fundamental statistical visualizations and exploratory data analysis, using the Python data science and visualization stack. Therefore, this course may be more suitable for students who pursue their careers in research, developement, and data analysis.

COURSE OBJECTIVES

By the end of the course, you are expected to be able to understand, explain, and manipulate basic types of data, analyze them by applying basic exploratory visualization techniques, and create basic explanatory web-based visualizations. You will also be able to evaluate the effectiveness of data visualizations based on the principles of human perception, design, types of data, and visualization techniques.

PREREQUISITES

This course's residential version is open to advanced undergraduate students as well as graduate students; online version of this course is open to graduate students. Because creating visualizations using programming languages (Python) is an integral part of the course, it is required to have good understanding and working knowledge of programming.

For the undergraduate students, the prerequisites are:

· Programming foundations (I210 & I211, or equivalent courses).

It is strongly recommended to have basic understanding of mathematics, statistics, and Web. The following courses are recommended before taking this course:

· I123: Data Fluency

· I300: HCI/Interaction Design

· I308: Information Representation

· I360: Web Design

For self-assessment, visit the following link: http://bit.ly/dviz_self. Contact the instructor if you are uncertain about your background.

EXPECTATIONS AND REQUIREMENTS

(All sections) The final assessment will be mainly based on the exam and course project. The choice of project topic can be guided by the instructors but you will have freedom to choose projects topics. You are required to submit a final paper that contains detailed explanation of the visualization process and results as well as the visualization artifact itself (e.g. a visualization tool or a webpage) depending on the nature of your project.

(Residential course) You are expected to attend every class and engage in discussions. You are not allowed to use your phone or computer during the class unless explicitly asked to do so. You are expected to read reading materials *prior* to the class. At the beginning of most class meetings, there will be an *in-class quiz* based on the assigned readings and previous materials. You are expected to complete all weekly assignments.

(Online) You are expected to complete all course modules and assignments. You are also expected to actively engage in discussions on Slack.

BOOKS AND KEY MATERIALS

There is no required textbook, but the following books and websites are recommended.

Python and data analysis

1. Python 3 Official Documentation

2. Dive Into Python by Mark Pilgrim (available online)

3. An introduction to statistics (with Python) by Thomas Haslwanter (available online): this book uses Python to explain basic statistics. It also contains a succinct tutorial for Python and data visualization using Python.

4. Learnpython.org: A web-based interactive tutorial

5. Learning IPython for Interactive Computing and Data Visualization by Cyrille Rossant: Introduction to IPython as well as lots of advanced analysis

Visualization and Design

See Visualization books on my wiki.

POLICIES

1. *Disabilities.* Every attempt will be made to accommodate qualified students
   with disabilities (e.g. mental health, learning, chronic health, physical, hear-
   ing, vision, neurological, etc.). You must have established your eligibility
   for support services through Disability Services for Students. Note that ser-
   vices are confidential, may take time to put into place, and are not retroac-
   tive. Captions and alternate media for print materials may take three or more
   weeks to get produced. Please contact Disability Services for Students at
   http://disabilityservices.indiana.edu or 812-855-7578 as soon as possible if
   accommodations are needed. The office is located on the third floor, west tower,
   of the Wells Library (Room W302). Walk-ins are welcome 8 AM to 5 PM, Mon-
   day through Friday. You can also locate a variety of campus resources for stu-
   dents and visitors who need assistance at http://www.iu.edu/~ada/index.shtml.

2. *Sexual misconduct and Title IX.* As your instructor, one of my responsibilities
   is to create a positive learning environment for all students. Title IX and IU's
   Sexual Misconduct Policy prohibit sexual misconduct in any form, including
   sexual harassment, sexual assault, stalking, and dating and domestic violence.
   If you have experienced sexual misconduct, or know someone who has, the
   University can help. If you are seeking help and would like to speak to someone
   confidentially, you can make an appointment with:

   a) The Sexual Assault Crisis Services (SACS) at (812) 855-8900 (counseling
      services)

   b) Confidential Victim Advocates (CVA) at (812) 856-2469 (advocacy and
      advice services)

   c) IU Health Center at (812) 855-4011 (health and medical services)

   It is also important that you know that Title IX and University policy require
   me to share any information brought to my attention about potential sexual
   misconduct, with the campus Deputy Title IX Coordinator or IU's Title IX
   Coordinator. In that event, those individuals will work to ensure that appropriate
   measures are taken and resources are made available. Protecting student privacy
   is of utmost concern, and information will only be shared with those that need
   to know to ensure the University can respond and assist. I encourage you to
   visit *stopsexualviolence.iu.edu* to learn more.

3. *No electronics—laptops, tablets, and smartphones—may be used in class*, unless
   the usage is specifically requested by the instructors. It has been shown that

using laptops in class hurts learning, *even if* you are using it to take notes. If you must have electronics due to a special reason, please obtain permission beforehand.

4. *Be honest.* Your assignments and papers should be your own work. If you find useful resources for your assignments, share them and cite them. If your friends helped you, acknolwedge them. You should feel free to discuss both online and offline, but do not show your code directly. Any cases of academic misconduct (cheating, fabrication, plagiarism, etc) will be reported to the School and the Dean of Students, following the standard procedure. Cheating is not cool.

5. *You have the responsibility of backing up all your data and code*. Always use at least Box, Dropbox, or Google Drive. Ideally, learn version control systems and use `https://github.iu.edu` or `https://github.com`. Loss of data, code, or papers due to various reasons (e.g. malfunction of your laptop) is not an acceptable excuse for delayed or missing submission.

6. *Inform your excused absences prior to class*. Please contact the instructor until the previous day for an excused absence.

7. If you have any issues, don't hesistate to contact me or IU's Counseling and Psychological Services.

GRADING

· Attendance, Quiz, and Participation: 30%

· Assignments: 30%

· Final project: 40%

COURSE SCHEDULE

The schedule is subject to change.

Week 1 (5/7-): Why visualization?

Week 2 (5/14-): History and integrity

Week 3 (5/21-): Perception

Week 4 (5/28-): Design

Week 5 (6/4-): Data Types and 1-D data

Week 6 (6/11-): Histogram and Boxplot

Week 7 (6/18-): Estimation

Week 8 (6/25-): Logscale and Beyond 1-D

Week 9 (7/2-): High-dimensional data

Week 10 (7/9-): Maps

Week 11 (7/16-): Text and Networks

Week 12 (7/23-): Project hack & presentation