INFO 590

Data Visualization

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

May 11, 2025

Office hours & TAs: see Canvas
Homepage: https://yyahn.com/dviz-course/

COURSE DESCRIPTION

Data visualization plays a pivotal role in understanding information, from news articles to cutting-edge scientific research, and is employed across diverse settings, from home offices to the world's largest corporations. As an integral component of data analysis, data visualization has become a crucial skill for all knowledge workers.

This introductory course delves into the core concepts of statistical data analysis and visualization. You will explore the foundations of data visualization, covering topics such as perception, integrity, design principles, statistical methods, data classifications, and various visualization techniques. Through hands-on exercises utilizing the Python stack, students will develop practical skills in data processing and visualization.

Relationships with E483/E583 Information Visualization (IVMOOC)   This course differs from E583/Z637 in that it places greater emphasis on fundamental statistical visualizations and conducting exploratory data analysis through coding, utilizing the Python data science and visualization stack. As a result, it may be a more suitable choice for students aiming for careers in research, development, engineering, and data analysis, or for those who will directly handle and analyze complex datasets.

COURSE OBJECTIVES

Upon completing the course, you are expected to acquire the ability to produce data visualizations that are effective, accurate, and ethical. You will also be able to critically evaluate data visualizations and effectively communicate your findings to others. This ability will build on your understanding of the fundamental principles of data visualization, including human perception, design, data types,

and visualization techniques. Your proficiency will be showcased through a course project through which you will not only create data visualizations but also document the process of creating effective, accurate, and ethical data visualizations.

COMMUNICATION

We will use Canvas and Slack for communication. **Canvas is for official communications** as well as for anything that contains personal and sensitive information. Additionally, you can use **Slack** for day-to-day information sharing, Q&As, team discussions, and other casual conversations.

Announcements, Q&As, and other official communication will be sent via Canvas. When joining the course Slack, feel free to avoid using your full name (e.g., use "John D." instead of "John Doe") to protect your privacy. Also never post your personal information or sensitive data (e.g., grades) to the course Slack.

The address of the course slack is: https://iu-dviz-course.slack.com, and visit https://join.slack.com/t/iu-dviz-course/signup to signup. You can create an account by using your IU email address. Please email the instructor if you want to use other email addresses or you have any issues.

Please be aware that the instructors may respond much more quickly to your messages on Slack than they do to emails or Canvas messages (too many emails!). If your message contains sensitive information, you can send it via email or Canvas, while sending a message on Slack to notify the instructor that you have sent an email or Canvas message.

Whenever you have something to say about the course or have a suggestion for improving the course, please share your thoughts! We will be extremely grateful if you point out issues in the quizzes, discussions, grading, and so on. You can simply send a message on slack, or anonymously share your opinion:

https://forms.gle/MzzNSV6Y8deJWGC77

PREREQUISITES

Because producing visualizations using Python data & visualization stack is an integral part of the course, it is required to have a good understanding and working knowledge of programming (esp. Python), as well as working knowledge of using open-source libraries. It is also recommended to have a basic understanding of mathematics, statistics, and the Web (HTML, CSS, Javascript, and JSON).

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

EXPECTATIONS AND REQUIREMENTS

The primary assessment will be through the assignments, the final exam, and the final course project. The topic of the final project will be of your (team's) choice, but I encourage everyone to consult with the instructors. You are required to submit a final paper that contains not only the *results* but

also detailed explanation of the visualization *process* to demonstrate your knowledge on visualization principles and techniques, as well as your ability to apply them to create visualizations.

(Online) You are expected to complete all course modules and assignments. The modules have many discussions that are graded as complete/incomplete. You are not expected to provide "correct" answers to those discussions. They are there to help you think about the topic. So, don't worry if you don't know the answer. You don't have to search the web or ask AI for the answer! Just share your initial thoughts and ideas. You are also expected to engage in discussions on Canvas and Slack.

GRADING

I sincerely hope that you focus on your learning and not on the grades! See https://www.youtube.com/watch?v=u6XAPnuFjJc

The grade may be curved at the end of the course. Because the gradebook often has ungraded items, *the grade that you can see on Canvas may not be a faithful reflection of your projected grade!*.

There will be extra credits based on your strong engagement in the course, in terms of sharing useful resources & interesting visualization-related articles, participating in discussions, and helping other students.

- · Attendance, Quiz, and Participation: 20%
- · Assignments: 20%
- · Exam: 30%
- · Final project: 30%

BOOKS AND KEY MATERIALS

There is no required textbook, but we will mainly use materials from the following books:

1. Fundamentals of Data Visualization by Claus O. Wilke (available online at https://clauswilke.com/dataviz/)
2. The Visual Display of Quantitative Information (2nd ed.) by E.R. Tufte: one of the foundational book on visualization. It contains a rich set of historical visualization, thoughtful discussion on visualization principles.

See also Visualization books and Data Visualization page on my wiki.

If you are still in the process of learning the basics of Python, the following books and websites may be helpful for you:

1. https://docs.python.org/3/: Python 3 Official Documentation
2. http://www.diveintopython3.net/index.html: Dive Into Python by Mark Pilgrim
3. http://www.learnpython.org: A web-based interactive tutorial
4. https://github.com/ipython-books/minibook-2nd-code: Learning IPython for Interactive Computing and Data Visualization by Cyrille Rossant: Introduction to IPython as well as lots of advanced analysis

If you are interested in web-based visualizations, you should check out the ObservableHQ and its tutorials for D3.js, Vega-Lite, and Observable Plot, all available at the following URL:

· https://observablehq.com/tutorials

FINAL PROJECT

See https://github.com/yy/dviz-course/wiki/Projects for the final project details, including the deliverables, types of projects, and some project ideas.

COURSE SCHEDULE

The schedule may change. See also IU Official Calendar.

Key dates

Mark your calendar and plan ahead! The summer course is fast-paced and end-heavy (both the exam and the final project are due in the last couple of weeks of the course). So, be sure to start the project early and plan ahead.

· Project proposal due: **06/06**
· Project due: **07/28**
· Exam: **07/28–08/01**

Schedule

Week	Date	Topic	Key dates
01	05/13–	Why visualization?	Visualization tools	
02	05/19–	History and Integrity	
03	05/26–	Perception	
04	06/02–	Design principles	Project proposal due: 6/6
05	06/09–	Data types and tidy data	
06	06/16–	Histogram and box plot	
07	06/23–	Estimation and log-scale	
08	06/30–	High-dimensional data	
09	07/07–	Maps	
10	07/14–	Text	Networks	
11	07/21–	Project week	Presentation and paper due: 7/28
12	07/28–	Final exam week	End of course: 8/1

Readings

Week 1: Why visualization?

· J. Heer *et al*. A Tour through the Visualization Zoo. https://queue.acm.org/detail.cfm?id=1805128
· J. VanderPlas, The Python Visualization Landscape. https://youtu.be/FytuB8nFHPQ

· Further readings: https://github.com/yy/dviz-course/blob/master/m01-intro/class.md

Week 2: History and integrity

· E.R. Tufte, The Visual Display of Quantitative Information, Ch. 1–2.
· C.O. Wilke, Fundamentals of Data Visualization Ch. 1 (https://serialmentor.com/dataviz/introduction.html).
· Further readings: https://github.com/yy/dviz-course/blob/master/m02-history/class.md and https://github.com/yy/dviz-course/blob/master/m03-integrity/class.md

Week 3: Perception

· C.G. Healey, Perception in Visualization, https://www.csc2.ncsu.edu/faculty/healey/PP/index.html
· B. Wong, Color Coding, Nature Methods (2010).
· B. Wong, Avoiding color, Nature Methods (2011).
· C.O. Wilke, Fundamentals of Data Visualization Ch. 4 Color scales (https://serialmentor.com/dataviz/color-basics.html).
· C.O. Wilke, Fundamentals of Data Visualization Ch. 15 Common pitfalls of color use (https://serialmentor.com/dataviz/color-pitfalls.html).
· Further readings: https://github.com/yy/dviz-course/blob/master/m04-perception/class.md

Week 4: Design

· B. Wong, Gestalt Principles I & II, Nature Methods (2010).
· E.R. Tufte, The Visual Display of Quantitative Information, Ch. 4.
· S. Bateman et al., Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts, CHI'10.
· C.O. Wilke, Fundamentals of Data Visualization Ch. 18–21 (https://serialmentor.com/dataviz/optimize-data-signal.html).
· Further readings: https://github.com/yy/dviz-course/blob/master/m05-design/class.md

Week 5: Data Types and 1-D data

· H. Wickham, Tidy Data, Journal of Statistical Software, https://vita.had.co.nz/papers/tidy-data.pdf
· C.O. Wilke, Fundamentals of Data Visualization Ch. 14 (https://serialmentor.com/dataviz/overlapping-points.html).
· Further readings: https://github.com/yy/dviz-course/blob/master/m06-data/class.md

Week 6: Histogram and Boxplot

· C.O. Wilke, Fundamentals of Data Visualization Ch. 6–7 (https://serialmentor.com/dataviz/overlapping-points.html).
· Further readings: https://github.com/yy/dviz-course/blob/master/m07-1D/class.md and https://github.com/yy/dviz-course/blob/master/m08-histogram/class.md

Week 7: Estimation and logscale

- · C.O. Wilke, Fundamentals of Data Visualization Ch. 8–9 (https://serialmentor.com/dataviz/overlapping-points.html).
- · Further readings: https://github.com/yy/dviz-course/blob/master/m09-estimation/class.md
- · Khan Academy: Logarithmic scale with Vi Hart (https://www.khanacademy.org/math/algebra2/exponential-and-logarithmic-functions/logarithmic-scale).

Week 8: High-dimensional data

- · C.O. Wilke, Fundamentals of Data Visualization Ch. 11 (https://serialmentor.com/dataviz/visualizing-associations.html).
- · 3Blue1Brown, Eigenvectors and eigenvalues https://www.youtube.com/watch?v=PFDu9oVAE-g.
- · Victor Powell, PCA http://setosa.io/ev/principal-component-analysis/.
- · L. van der Maaten & G. Hinton, Visualizing data using t-SNE, JMLR 2008 http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.
- · Further readings: https://github.com/yy/dviz-course/blob/master/m10-logscale/class.md and https://github.com/yy/dviz-course/blob/master/m11-highdim/class.md

Week 9: Maps

- · Vsauce, What does earth look like? https://youtu.be/2lR7s1Y6Zig
- · Vox, Why all world maps are wrong https://youtu.be/kIID5FDi2JQ
- · Further readings: https://github.com/yy/dviz-course/blob/master/m12-maps/class.md

Week 10: Text and Networks

- · J. Harris, Word clouds considered harmful, http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/.
- · The Observatory of Economic Complexity, https://atlas.media.mit.edu/en/profile/country/usa/.
- · Further readings: https://github.com/yy/dviz-course/blob/master/m13-text/class.md and https://github.com/yy/dviz-course/blob/master/m14-networks-and-interactive/class-network.md

Week 11: Project week

Week 12: Final exam week

POLICIES

1. *Be honest.* Don't be a cheater. 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, acknowledge them. You should feel free to discuss both online and offline (except for the exam), 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.*

2. *You have the responsibility of backing up all your data and code.* Always back up your code and data. You should at least use Google Drive or Dropbox at the minimum. You can also use cloud services like Google Colaboratory. Ideally, learn version control systems and use https://github.iu.edu or https://github.com. Loss of data, code, or papers (e.g. due to malfunction of your laptop) is not an acceptable excuse for delayed or missing submission.

3. *Disabilities.* Every attempt will be made to accommodate qualified students with disabilities (e.g. mental health, learning, chronic health, physical, hearing, vision, neurological, etc.). You must have established your eligibility for support services through Disability Services for Students. Note that services are confidential, may take time to put into place, and are not retroactive. 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, Monday through Friday. You can also locate a variety of campus resources for students and visitors who need assistance at http://www.iu.edu/~ada/index.shtml.

4. *Bias-based incidents.* Any act of discrimination or harassment based on race, ethnicity, religious affiliation, gender, gender identity, sexual orientation, or disability can be reported to biasincident@indiana.edu or to the Dean of Students Office at (812) 855-8188.

5. *Sexual misconduct and Title IX.* 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, you can use university resources:
   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. Visit *stopsexualviolence.iu.edu* to learn more.

6. If you have any mental health issues, don't hesitate to contact IU's Counseling and Psychological Services, which provides free counseling sessions. Also, please contact Disability Services for Students at http://disabilityservices.indiana.edu or 812-855-7578 as soon as possible if accommodations are needed. See "Disabilities" section for more information.