INFO I590 (Summer 2019)

Data Visualization

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

April 7, 2019

Assistant Instructors: TBD

COURSE DESCRIPTION

From news to cutting-edge scientific papers, from a home office to the largest companies in the world, data visualization is a critical method for revealing patterns in data and telling stories. Because data visualization is indispensible in understanding data, and because more and more decisions are driven by data, data visualization has become 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 aim to equip you with practical data visualization skills and they will be an integral part of the course.

Relationships with E583/Z637 Information Visualization (IVMOOC): Compared with E583/Z637, this course is geared more towards producing 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, engineering, and data analysis, and those who will directly process and analyze complex datasets.

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 explanatory visualizations. You are also expected to be able to evaluate and improve the effectiveness of data visualizations based on the principles of human perception, design, types of data, and visualization techniques.

1

COMMUNICATION

Announcements and all communication will be through Slack:

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

The first thing you need to do when you enroll this course is to join the course slack. You can create an account by using one of the following IU email addresses: indiana.edu, umail.iu.edu, iu.edu, and iupui.edu. Visit https://join.slack.com/t/iu-dviz-course/signup to signup.

Slack allows both private (group) communications and public fora. The number-one recommendation from the previous students for the new students was to learn to use Slack and get used to it as soon as possible. Email and Canvas will not be very effective due to the large email volume that the instructor receives usually. Please expect the following response time:

· **Email or Canvas**: the instructor will respond within *one week* (likely within several days).

· **Slack**: the instructor will respond within *one day* (likely within several hours).

PREREQUISITES

Because producing visualizations using Python data & visualization stack is an integral part of the course, it is required to have 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 basic understanding of mathematics, statistics, and 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 final assessment will be the final course project. The choice of your project topic can be guided by the instructors but you have freedom to choose your topic. 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.

You are expected to complete all course modules (quizzes and discussions) and assignments. You are also expected to engage in discussion on Canvas and Slack.

GRADING

The grade may be shifted at the end of the course because of new assignments, changes in the course from the previous offerings, etc. Therefore, the grade that you can see on Canvas may not be the grade you receive finally.

There will be small extra credits for active engagement on Slack, in terms of sharing useful resources & interesting visualization related articles, participating in discussions, and helping other students (e.g. answering questions)

- · Attendance, Quiz, and Participation: 30%

- · Assignments: 30%

- · Final project: 40%

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 (book preview available online at https://serialmentor.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 on my wiki (http://yyahnwiki.appspot.com/Information_visualization).

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

Python and data analysis

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://work.thaslwanter.at/Stats/html/: 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. http://www.learnpython.org: A web-based interactive tutorial

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

POLICIES

1. *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`.

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. *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.*

4. *You have the responsibility of backing up all your data and code.* Always back up your code and data. You should at least use Box, Dropbox, or Google Drive at the minimum. 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.

COURSE SCHEDULE AND READINGS

The schedule may change due to unexpected circumstances. See also IU Official Calendar.

Week 1 (5/7-): Why visualization? | Overview of Python visualization libraries

- · 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`

Week 2 (5/13-): History and integrity | Plotting basics with the Snow's map

- · 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`).

Week 3 (5/20-): Perception | Perception experiments

- · 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`).

Week 4 (5/27-): Design | Colors, Vector, and Bitmaps

- · 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`).

Week 5 (6/3-): Data Types and 1-D data | Tidy data and 1D 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`).

Week 6 (6/10-): Histogram and Boxplot | Histogram and CDF

 · C.O. Wilke, Fundamentals of Data Visualization Ch. 6–7 (`https://serialmentor.com/dataviz/overlapping-points.html`).

Week 7 (6/17-): Estimation | Estimation

 · C.O. Wilke, Fundamentals of Data Visualization Ch. 8–9 (`https://serialmentor.com/dataviz/overlapping-points.html`).

Week 8 (6/24-): Logscale and Beyond 1-D | Logscale and CCDF

 · Khan Academy: Logarithmic scale with Vi Hart (`https://www.khanacademy.org/math/algebra2/exponential-and-logarithmic-functions/logarithmic-scale`).

Week 9 (7/1-): High-dimensional data | 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`.

Week 10 (7/8-): Maps | Maps

 · Vsauce, What does earth look like? `https://youtu.be/2lR7s1Y6Zig`
 · Vox, Why all world maps are wrong `https://youtu.be/kIID5FDi2JQ`

Week 11 (7/15-): Text and Networks | 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/`.

Week 12 (7/22-): Final project week