INTEG 375

Data Visualization

Monday and Wednesday, 10:00 - 1:30 in EV1-132
w2016-integ375.slack.com

John McLevey, PhD
Knowledge Integration
Sociology & Legal Studies
University of Waterloo

DESCRIPTION

Visualizing data is central to scientific and social scientific research, and is increasingly prominent in science communication and journalism. This course provides an introduction to the history, principles, and techniques of data visualization and information design, with an emphasis on data used by social scientists, information scientists, and policy analysts. Students will learn how to create effective visualizations for a variety of different audiences and purposes, including making sense of new datasets and communicating evidence to others.

I assume that students in this class will have a minimal background in research methods (i.e. one undergraduate research methods class), but little to no programming experience. While some programming experience is advantageous, you will acquire the necessary skills in this course.

LEARNING OUTCOMES

There are three primary learning objectives in this course.

1. Students will learn principles and best practices for visualizing data.
2. Students will learn how to design and create their own visualizations using open source tools.
3. Students learn how to critically assess visualizations produced by others.

DELIVERABLES AND EVALUATION

Assignment	Deadline	Value
4 visualization challenges (collaborative)	In Class, but TBA	20%
Article / chapter review (800 words, collaborative)	Due the day before class	10%
Visualization review (1,000 words, collaborative)	March 21	20%
"Portfolio" presentation (collaborative)	March 23-30	10%
Visualization "portfolio" (collaborative)	April 8	35%
Engagement / Participation	Ongoing	5%

4 Visualization Challenges (Collaborative)

There will be 4 small visualization challenges over the course of the semester. I announce the challenge in class, and will collect your final work *before* the next scheduled class meeting. Students will work in groups of 3.

Article / Chapter Review (Collaborative)

You will review 1 article or chapter from the syllabus *in depth*. Each review is due **a full day before** the class where we will discuss the readings you wrote about. If your review is late, you will not receive credit under any circumstances. Reviews **will not** exceed 800 words. I expect them to be thoughtful, clear, and carefully edited. They should identify core points from the readings, offer any thoughts or reactions, and raise possible questions. These reviews will be shared with the class on #slack, where others can read and leave constructive and thoughtful comments. You are not required to comment on the reviews, but doing so is a good way to contribute to class discussions and improve your participation grade. Disrespectful comments will not be tolerated. They will be as damaging to your participation grade as hostile behaviour in the classroom.

Visualization Review (Collaborative)

You will do a "visualization review" that includes (1) at least one graph that has been published in a peer reviewed **academic** article in any of the sciences or social sciences, and (2) a graph intended for **broader publics**. You can find these graphs for broader publics almost anywhere, but I expect you to find and discuss graphs that lots of people would likely see – i.e. published in major newspapers and magazines, in election coverage, in ad campaigns, etc. Your review should include a critical discussion of the graphs, with a **heavy** emphasis on theories and principles from class meetings and readings. Criticisms should always be accompanied by suggestions about what could have been done better. In fact, you may want to actually *show* how it could be done better by including your own graph. Your review will not exceed 1,000 words.

Visualization Portfolio & Portfolio Presentation (Collaborative)

As we progress through the course, you will gradually build up a "portfolio" consisting of (1) a variety of graphs with both professional audiences and broader publics in mind, (2) the R code used to produce your graphs, and (3) an explanation of why you made the choices you did. At the end of the semester, you will present your portfolio to the class. Note that the portfolio and the presentation

make up 45% of your final grade. It is *not* something that you should leave until the last minute. You may revise your final portfolio based on class discussion following your presentation.

Participation

There will be "hands on" sessions almost every week. Your participation grade will be based on how engaged you are in those sessions, and how much you contribute to class discussions about the assigned readings.

Presentation in the KI Seminar Series

We are booked for the March 11 spot in the Knowledge Integration seminar series. I will speak about some of the goals of the course, and you will all present some of the graphs that you have produced over the course of the semester. As with your portfolios, I expect you to describe the larger motivation behind your graphs (e.g. what questions are you asking, why are you asking them), and perhaps explain some of the design decisions you made in creating the graphs. The KI seminar series is held on Fridays. If you are not able to make it (e.g. because you work on Fridays), please let me know as soon as possible.

READINGS

There are two required books for this course. They are available for purchase at the UW bookstore or online.

> Stephen Few (2012) *Show Me the Numbers*
> Winston Change (2013) *R Graphics Cookbook*

SOFTWARE

I am teaching this course using R, which is an open source language / environment for statistical computing and data visualization. If you have experience programming in other language, you will likely notice that R has some quirks. (We will talk about them in class.) One of the many reasons why we are using R instead of some other language is because it has *fantastic* libraries for data visualization, and for doing reproducible work. We will do all our our R coding in RStudio. In addition to R and RStudio, you should install LaTeX, Haskell, and Pandoc on your laptops.

We will have an "installation party" in class, but **please try and install as much as you can first**. All software is open source. You do not have to pay for anything.

SUBMITTING WORK & LATE POLICY

You will submit all work electronically in the general channel on #slack **and** on LEARN. **You may not submit hard copies or Microsoft Word documents under any circumstances**. Please submit PDF files instead. I will deduct **5 points** a day for every day, or part of a day, that your work is late, including weekends. I will not make exceptions without a medical note.

COMMUNICATION

We will be using the collaboration tool slack for all class communication. Of course you are free to email me, but I tend to respond to slack messages from students faster than I respond to emails. Sign up with and sign into slack by going to slack.com. There are slack apps for Mac OS X, iOS, and Android. If you are a Linux or Windows user, there is a very good web app.

Feedback

I will solicit brief, informal, and confidential course evaluations three or four times throughout the semester. These will only take a few minutes of your time. The purpose is to make sure that we are moving at a comfortable pace, that you feel you understand the material, and that my teaching style is meeting your needs. I will use this ongoing feedback to make adjustments as the course progresses. Although you are not obligated to do so, please fill out the evaluations so that I can make this the best learning experience for you, and the best teaching experience for me.

ON CAMPUS RESOURCES

The Writing Centre

Although I will be giving you feedback on your work throughout the term, I encourage you to make appointments with people at the writing centre. Their services are available to all UW students.

Access Ability Services

The AccessAbility Office, located in Needles Hall, Room 1132, collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations to lessen the impact of your disability, please register with the AccessAbility Office at the beginning of each academic term.

Mental Health

The University of Waterloo, the Faculty of Environment, and our Departments consider students' well-being to be extremely important. We recognize that throughout the term students may face health challenges – physical and / or emotional. Please note that help is available. Mental health is a serious issue for everyone and can affect your ability to do your best work. Counselling Services is an inclusive, non-judgmental, and confidential space for anyone to seek support. They offer confidential counselling for a variety of areas including anxiety, stress management, depression, grief, substance use, sexuality, relationship issues, and much more.

UNIVERSITY POLICIES

Academic Integrity

In order to maintain a culture of academic integrity, members of the University of Waterloo community are expected to promote honesty, trust, fairness, respect and responsibility.

We will all uphold academic integrity policies at University of Waterloo, which include but are not limited to promoting academic freedom and a community free from discrimination and harassment. You can educate yourself on these policies – and the disciplinary processes in place to deal with violations – on the Office of Academic Integrity website.

A student is expected to know what constitutes academic integrity, to avoid committing academic offense, and to take responsibility for his/her actions. A student who is unsure whether an action constitutes an offense, or who needs help in learning how to avoid offenses (e.g., plagiarism, cheating) or about 'rules' for group work / collaboration should seek guidance from the course professor, academic advisor, or the Undergraduate Associate Dean. For information on categories of offences and types of penalties, students should refer to Policy 71, Student Discipline. For typical penalties, check Guidelines for Assessment of Penalties.

Grievances and Appeals

A student who believes that a decision affecting some aspect of his / her university life has been unfair or unreasonable may have grounds for initiating a grievance. Read Policy 70: Student Petitions and Grievances, Section 4. When in doubt please contact your Undergraduate Advisor for details.

A decision made or penalty imposed under Policy 70 – Student Petitions and Grievances (other than a petition) or Policy 71 – (Student Discipline) may be appealed if there is a ground. A student who believes he/she has a ground for an appeal should refer to Policy 72 (Student Appeals).

Religious Observances

Student needs to inform the instructor at the beginning of term if special accommodation needs to be made for religious observances that are not otherwise accounted for in the scheduling of classes and deliverables.

DATASETS

In class, I will primarily use data from the International Social Survey Programme, including modules on attitudes, social networks, work, social inequality, and the environment. I may also occasionally use data from research projects that I am working on, or from the following sources:

1. Women in Parliament, 1945-2003: Cross-National Dataset
2. Aid Data: Open Data for International Development
3. Canadian Community Health Survey - Annual Component (CCHS)
4. British Household Panel Survey
5. Public use data from Statscan, including crime statistics
6. Government of Canada Open Data Portal
7. Any of the datasets available from the ICPSR repository or from Dataverse. There is a Dataverse repository for researchers affiliated with Ontario universities.

You are free to use any of these datasets in your own work.

USEFUL "CHEAT SHEETS" FOR R

Rstudio has useful cheatsheets for R Markdown, data visualization using ggplot2, and data wrangling with dplyr. You can get them at https://www.rstudio.com/resources/cheatsheets/.

SCHEDULE & READINGS

Date	Topic	Articles / Chapters
M, Jan 4	Introduction	Healy and Moody (2014)
W, Jan 6	Issues in data visualization	Few 1
M, Jan 11	Introduction to R and RStudio	Few 2
W, Jan 13	Differing roles of tables and graphs	Few 3 & 4
M, Jan 18	Visual perception	Few 5
W, Jan 20	Introduction to R and RMarkdown	Skim Healy (2014)
M, Jan 25	Data munging / tidy data (R library: dplyr)	Wickham (2014)
W, Jan 27	Selecting the right graphs	Few 6
M, Feb 1	General design principles for tables and graphs	Few 7 & 8
W, Feb 3	Graph design	Few 9 & 10
M, Feb 8	Distributions, relationships, and time series	Chang 3, 4, 5, & 6
W, Feb 10	Small multiples and generalized pairs plots	Few 11, Chang 11
—	—	—
M, Feb 22	Terrible graphs that you should never make	Few 12
W, Feb 24	Visualizing networks: igraph	Robins (2015) 8
M, Feb 29	Visualizing networks: igraph	Chang 12
W, March 2	Visualizing networks: igraph	Few 13
M, March 7	Visualizing data that is too BIG to fit on your machine	Few 14
W, March 9	Production: graphs for screens, slides, and paper	Chang 14
F, March 11	**Presentations in KI Seminar (see "deliverables")**	
M, March 14	In class project work	
W, March 16	In class project work	
M, March 21	In class project work	
W, March 23	Portfolio presentations	
M, March 28	Portfolio presentations	
W, March 30	Portfolio presentations and debriefing	

REFERENCES

Chang, Winston. 2013. *R Graphics Cookbook*. O'Reilly.

Few, Stephen. 2012. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press.

Healy, Kieran. 2014. *Plain Text, Papers, Pandoc*. http://kieranhealy.org/blog/archives/2014/01/23/plain-text/.

Healy, Kieran, and James Moody. 2014. "Data Visualization in Sociology." *Annual Review of Sociology* 40: 105–28.

Robins, Gary. 2015. *Doing Social Network Research: Network-Based Research Design for Social Scientists*. SAGE.

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software.*