STAT 2430 Data Visualization

This is the syllabus for the Data Visualization (STAT 2430) course offered in Winter 2024 at Dalhousie University.

- Instructor: Andrew Irwin, a.irwin@dal.ca
- Meeting times: Tuesday and Thursday 10:05-11:25 am, Henry Hicks (HH) 212
- Office hours: Tuesday and Thursday 2:30-3:30 pm in Chase 225, or by appointment.
- Math & Stats Learning Centre: Mohammad Amirian, Chase 119
    - Tuesday 11:30a-12:30p
    - Wednesday 1:30-2:30p
    - Thursday 1:30-2:30p

Course Description

Data visualization is the art and science of turning data into readable graphics. We will explore how to design and create data visualizations. This exploration will include both the principles and techniques of data visualization. Students will learn the value of visualization, specific techniques, and how to design and improve a data visualization.

Prerequisites

At least one MATH or STAT course at or above the 1000 level. No experience with R or computer programming is required, but a desire to learn about both is essential.

Course Materials

Course notes:

- Andrew J Irwin (2022) Data visualization.

Textbooks:

- Kieran Healy (2018) Data Visualization: A Practical Introduction. Princeton University Press.
- Claus O. Wilke (2019) Fundamentals of data visualization. O'Reilly.
- Garret Grolemund & Hadley Wickham (2017) R for data science. O'Reilly.

The textbooks are available in printed form from bookstores and online. The online versions are free to use. You are not required to buy the books on paper. All three are excellent books with very distinct goals. The books by Healy and Wilke focus on the design and interpretation of data visualizations. The book by Grolemund and

Wickham describes the use of R, particularly the set of packages known as the "tidyverse" for data visualization and the broader tasks of data science.

Software

We will use the statistical software R and RStudio and version control software git. No prior experience with R, RStudio or git is assumed. We will take class time to learn the software.

- To download and install R go to r-project.org and click on the link to download R
- To download and install Rstudio, go to Rstudio.com and click on the link to download Rstudio
- To download and install git:
  - on Windows, go to git-scm.org and click on the link to download a version for Windows
  - on Macintosh, use the Terminal app and type `xcode-select --install` to download and install git.

These tools can be installed on Linux computers as well. Contact me if you have trouble. If you have a Chromebook you can use all these tools through the cloud service rstudio.cloud. Everyone should learn to use the cloud service as a backup in case of problems with setting up the software on their own computer.

R and Rstudio are available on Dalhousie computer labs, but the git version control software must be installed following the instructions for Windows computers above. Since all your user files are erased from lab computers when you log out, this process must be repeated on each login. For this reason, I don't recommend using Dalhousie computer labs for this course, unless you are using rstudio.cloud.

You should install this software on your computer during the first week of the course. If you are unable to install the software, get in touch with me for assistance.

Course Structure

Each week will be structured around the following components:

- Reading course notes and suggested material from the textbooks.
- In class meetings which will be a mixture of lectures, demonstrations, and skills practice. Some class time will be be dedicated to computer demonstrations and exercises; these times will be identified by a file describing the demonstration (marked Demo in the detailed outline). You are encouraged to bring your computer to these classes.
- Opportunities to develop and demonstrate your knowledge in course work to be evaluated (tasks, assignments, and a term project).
- Scheduled office hours. Come to office hours if you have questions about the classes, readings, or work to be submitted for grading. If you can't participate in office hours, but you have a question, please get in touch with me and arrange another time to meet.

Follow the detailed outline for each week's plan to keep you on track.

Evaluation

Your goal in this course is to learn the principles and skills of data visualization. Most people learn best using a combination of collaborative and independent work. The general guideline is that work you submit should be your

own—your ideas, your thoughts, your choices, your code, your typing. You are encouraged to discuss your work with the instructor and students. When you help another student, be careful you don't help them so much that you inhibit their learning.

Tasks are opportunities to test your understanding of the key concepts from the lessons and demonstrate you have developed basic proficiency with essential computing skills. Most lessons will have a task for you to complete. After you work on the task with R, go to Brightspace to answer some questions about the task. Tasks are primarily designed to help you learn and you are encouraged to seek assistance from classmates, but work you submit must be your own.

Assignments combine skills from multiple lessons into a meaningful data visualization activity. These are opportunities to apply the content of lessons in thoughtful and sometimes creative ways. Assignments are assessments of your skills and should be done independently.

The **Term Project** is an opportunity to combine many of the skills learned in the course. You will explore, analyze, and present an analysis of a data set of your choice. Your proposal will select a data set and describe some of your planned analyses. The oral presentation will be a 5-10 minute overview of your data and key visualizations. The report will be a polished reproducible document demonstrating many of the ideas of the course using your selected data set. The term project is designed to be done collaboratively with a partner. The work on the project should be the work of members of the group and not outside collaborators or internet sources.

- Tasks (30%, due weekly, equally weighted)
- Assignments (30%, roughly 1 every two weeks, equally weighted)
- Term project, divided into three components
    - Proposal, due Thursday 7 March, 5%
    - Report, due Thursday 4 April, 15%

- Final exam, with two components
    - Oral presentation, due one day before the final exam, 10%
    - Peer evaluation of presentations, 10%

Tasks are due by the end of the day on Monday. Assignments are due on Friday, approximately one week after they are assigned. If Friday is a holiday (Feb 2, March 29) the assignment will be due the following Monday.

Copying work from any other sources is not allowed and will be considered an academic offense for this course. You are encouraged to *learn* from many different sources. If you make use of material outside of course materials on assignments or the project report, include references to the material and a description of what you used in a "Sources used" section at the end of your work.

Late policy. Tasks will not be accepted late, but I will drop your two lowest grades. Late assignments will be penalized 10% per day. If an accommodation is requested and granted due to illness or other unforeseeable circumstance, the value of that work will be redistributed to other tasks or assignments. *If you anticipate not being able to submit any component of the term project (proposal, oral presentation, or report) on time, please contact me by email.*

Grades will be reported on Brightspace.

Letter grades

Your numerical grade will be converted to a letter grade for the course using the Dalhousie Common Grade Scale.

First your numerical grade will be rounded up to the nearest integer, then it will be converted to a letter using this table.

A+	A	A-	B+	B	B-	C+	C	C-	D	F
90-100	85-89	80-84	77-79	73-76	70-72	65-69	60-64	55-59	50-54	0-49

Course Policies

Data retention policy

During the course you will create work on a third-party website (github.com). These files will be shared only with the instructor, TAs, and student team members. All files will be kept for 6 months after the end of the course. Repositories will be deleted after that time. If you wish to keep files after that date, either clone the repositories to your computer or fork them to your own GitHub account separate from the course.

Important dates

- Tuesday 9 January: Term begins
- February 19-23: Study break
- Thursday 4 April: Last Tuesday/Thursday class

If you know you will be unable to meet certain deadlines in advance, for *bona fide* reasons such as scheduled medical appointments or religious observances, please contact me as soon as practical.

I will not require you to provide a medical note or to use the Student Declaration of Absence form.

For due dates for evaluations, see the detailed course outline and evaluation schedule.

Territorial Acknowledgement

The Dalhousie University Senate acknowledges that we are in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq People and pays respect to the Indigenous knowledges held by the Mi'kmaq People, and to the wisdom of their Elders past and present. The Mi'kmaq People signed Peace and Friendship Treaties with the Crown, and section 35 of the Constitution Act, 1982 recognizes and affirms Aboriginal and Treaty rights.

The Dalhousie University Senate also acknowledges the histories, contributions, and legacies of African Nova Scotians, who have been here for over 400 years. For more information see, e.g., the Black Cultural Centre for Nova Scotia

University Policies and Statements

This course is governed by the academic rules and regulations set forth in the University Calendar and by Senate.

Please see the supplement to the syllabus on Brightspace for important statements, policies, and resources provided by the Faculty of Science and the University.