

Syllabus

Contents

Course title and instructor

Purpose and contents of the class

Prerequisites

Textbook

Topics covered

Computing requirements

Course site

Assignments and grading

Late assignment policy

Extra assignment for graduate course credit

Office hours

Email policy

Special accommodations

Academic dishonesty

Sharing of Course Materials is Prohibited

Class Recordings

Course title and instructor

Title: SDS 375/395 Data Visualization in R

Semester: Spring 2023

Unique: 57635 and 57745, TTH 3:30pm–5:00pm, UTC 4.110

Instructor: Claus O. Wilke

Email: wilke@austin.utexas.edu

Office Hours: Mon. 9am - 10am (open Zoom), Thurs. 10am - 11am (open Zoom), or by appointment

GitHub: [clauswilke](https://github.com/clauswilke)

Teaching Assistant: Alexis Hill

Email: alexis.hill@utexas.edu

Office Hours: Monday 2pm - 3pm (open Zoom), Tuesday 10am - 11am (open Zoom), or by appointment

appointment

GitHub: [alexismhill3](#)

Purpose and contents of the class

In this class, students will learn how to visualize data sets and how to reason about and communicate with data visualizations. A substantial component of this class will be dedicated to learning how to program in R. In addition, students will learn how to compile analyses and visualizations into reports, how to make the reports reproducible, and how to post reports on a website or blog.

Prerequisites

The class requires no prior knowledge of programming. However, students are expected to have successfully completed an introductory statistics class taught with R, such as SDS 320E, and they are expected to have some basic familiarity with the statistical language R.

This class is not a biology class, and no knowledge of biology is expected. Thus, the class is suitable for anybody interested in data visualization, including mathematicians, engineers, physicists, etc.

Textbook

This class draws heavily from materials presented in the following book:

- Claus O. Wilke. Fundamentals of Data Visualization. O'Reilly Media, 2019.

Additionally, we will also make use of the following books:

- Hadley Wickham, Danielle Navarro, and Thomas Lin Pedersen. ggplot2: Elegant Graphics for Data Analysis, 3rd ed. Springer, to appear.
- Kieran Healy. Data Visualization: A Practical Introduction. Princeton University Press, 2018.

All these books are freely available online and you do not need to purchase a physical copy of either book to succeed in this class.

Topics covered

Class	Topic	Coding concepts covered
	Introduction	reproducible

Class	Topic	Studio setup online, R Markdown Coding concepts covered
1.	Introduction, reproducible workflows	
2.	Aesthetic mappings	ggplot2 quickstart
3.	Telling a story	
4.	Visualizing amounts	<code>geom_col()</code> , <code>geom_point()</code> , position adjustments
5.	Coordinate systems and axes	coords and position scales
6.	Visualizing distributions 1	<code>stats</code> , <code>geom_density()</code> , <code>geom_histogram()</code>
7.	Visualizing distributions 2	violin plots, sina plots, ridgeline plots
8.	Color scales	color and fill scales
9.	Data wrangling 1	<code>mutate()</code> , <code>filter()</code> , <code>arrange()</code>
10.	Data wrangling 2	<code>group_by()</code> , <code>summarize()</code> , <code>count()</code>
11.	Visualizing proportions	bar charts, pie charts
12.	Getting to know your data 1: Data providence	
13.	Getting to know your data 2: Data quality and relevance	handling missing data, <code>is.na()</code> , <code>case_when()</code>
14.	Getting things into the right order	<code>fct_reorder()</code> , <code>fct_lump()</code>
15.	Figure design	ggplot themes
16.	Color spaces, color vision deficiency	colorspace package
17.	Functions and functional programming	<code>map()</code> , <code>nest()</code> , purrr package
18.	Visualizing trends	<code>geom_smooth()</code>
19.	Working with models	<code>lm</code> , <code>cor.test</code> , broom package
20.	Visualizing uncertainty	frequency framing, error bars, ggdist package
21.	Dimension reduction 1	PCA
22.	Dimension reduction 2	kernel PCA, t-SNE, UMAP
23.	Clustering 1	k-means clustering
24.	Clustering 2	hierarchical clustering
25.	Visualizing geospatial data	<code>geom_sf()</code> , <code>coord_sf()</code>
26.	Redundant coding, text annotations	ggrepel package
27.	Interactive plots	ggiraph package

Class	Topic	Coding concepts covered
28.	Overplotting	filtering, 2D histograms, contour plots
29.	Compound figures	patchwork package

Computing requirements

Programming needs to be learned by doing, and a significant portion of the in-class time will be dedicated to working through simple problems. All programming exercises will be available through a web-based system, so the only system requirement for student computers is a modern web browser.

Course site

All materials and assignments will be posted on the course webpage at:
<https://wilkelab.org/SDS375>

Assignment deadlines are shown on the schedule at:
<https://wilkelab.org/SDS375/schedule.html>

Assignments will be submitted and grades will be posted on Canvas at:
<https://utexas.instructure.com>

R compute sessions are available at: <https://edupod.cns.utexas.edu>
 Note that edupods will be unavailable due to maintenance approximately two hours per month, usually on a Thursday afternoon between 4pm and 6pm. Specific maintenance times are published in advance here: <https://wikis.utexas.edu/display/RCTFusers>

An online discussions forum will be available at the following private GitHub repo:
https://github.com/wilkelab/SDS375_spring2023
 You will be given access to this repo during the first week of class.

Assignments and grading

The graded components of this class will be homeworks, projects, and online participation. Each week either one homework or one project is due. Homeworks will be relatively short visualization problems to be solved by the student, usually involving some small amount of programming to achieve a specified goal. Projects are larger and more involved data analysis problems that involve both programming and writing. Students will have at least one week to complete each homework and two weeks to complete each project. The submission deadlines for homeworks and projects will be Mondays at 11am.

There will be ten homeworks and three projects. Both homeworks and projects need to be submitted electronically on Canvas. Homeworks are worth 10 points and projects are worth

100 points. The lowest-scoring homework will be dropped, so that a maximum of 90 points can be obtained from the homeworks.

Online participation assignments are small tasks meant to promote discussion and engagement with the class materials. They are worth between 5 and 10 points, for a maximum of 60.

Assignment type	Number	Points per assignment	Total points
Homework	9 (+1)	10	90
Project	3	100	300
Participation	9	5-10	60

Thus, in summary, each project contributes 22% to the final grade, the totality of all homeworks contributes another 20% to the final grade, and participation contributes 13%.

There are no traditional exams in this class and there is no final.

The participation assignments are meant to foster discussion about class materials on GitHub. Several of them can be performed more than once and will receive the indicated points each time they are performed successfully, up to the maximum number of times the specific activity is rewarded or the point maximum of 60, whichever is reached earlier. The following activities will be available:

Participation assignment	Points	Times rewarded
Submit GitHub username	5	1x
Make a comment in the sandbox	5	1x
Open an issue	10	3x
Provide a constructive comment	5	6x
Close an issue that was answered	5	3x

The class will use +/- grading, and the exact grade boundaries will be determined at the end of the semester. However, the following minimum grades will be guaranteed:

Points achieved	Minimum guaranteed grade
405 (90%)	A-
360 (80%)	B-
315 (70%)	C-
225 (50%)	D-

Late assignment policy

Homeworks that are submitted past the posted deadline will not be graded and will receive 0

points.

Project submissions will have a 2-day grace period. Projects submitted during the grace period will have 25 points deducted from the obtained grade. After the grace period, students who have not submitted their project will receive 0 points.

Participation assignments will be open for several weeks each and will be spaced approximately evenly throughout the semester, so that students have to demonstrate consistent participation to achieve full points.

Extra assignment for graduate course credit

Graduate students who are taking this class for graduate course credit will have to complete one additional assignment. The assignment will be to write a brief report (4-5 pages, no more than 3 figures) applying concepts from this class to a dataset of the student's choice. This assignment will be graded pass/fail, and a failing grade on this assignment will result in a 10% penalty on the total points obtained in the class. Students who receive a failing grade can submit a revised assignment for regrading. The last day by which a revised assignment can be submitted is the last day of class in the semester.

Office hours

Both the graduate TAs and myself will be available at posted times or by appointment. Office hours will be over Zoom. The most effective way to request an appointment for office hours outside of posted times is to suggest several times that work for you. I would suggest to write an email such as the following:

Dear Dr. Wilke,

I would like to request a meeting with you outside of regular office hours this week. I am available Thurs. between 1pm and 2:30pm or Fri. before 11am or after 4pm.

Thanks a lot,

John Doe

Note that we will not usually make appointments before 9am or after 5pm.

Email policy

When emailing about this course, please put "SDS375" into the subject line. Emails to the instructor or TA should be restricted to organizational issues, such as requests for

appointments, questions about course organization, etc. For all other issues, post an issue on GitHub, ask a question during open Zoom, or make an appointment for a one-on-one session.

Specifically, we will not discuss technical issues related to assignments over email. Technical issues are questions concerning how to approach a particular problem, whether a particular solution is correct, or how to use the statistical software R. These questions should be posted as issues on GitHub. Also, we will not discuss grading-related matters over email. If you have a concern about grading, schedule a one-on-one Zoom meeting.

Special accommodations

In-person attendance is not required. While the class is set up as an in-person class, there is no strict requirement for attendance. No grade component depends on any in-person activity, and every effort will be made to provide students with online access to course materials.

Students with disabilities. Students with disabilities may request appropriate accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471-6259, <https://diversity.utexas.edu/disability/>

Religious holy days. Students who must miss a class or an assignment to observe a religious holy day will be given an opportunity to complete the missed work within a reasonable time after the absence. According to UT Austin policy, such students must notify me of the pending absence at least fourteen days prior to the date of observance of a religious holy day.

Academic dishonesty

This course is built upon the idea that student interaction is important and a powerful way to learn. We encourage you to communicate with other students, in particular through the [discussion forums on GitHub](#). However, there are times when you need to demonstrate your own ability to work and solve problems. In particular, your homeworks and projects are independent work, unless explicitly stated otherwise. You are allowed to confer with fellow students about general approaches to solve the problems in the assignments, but you have to do the assignments on your own and describe your work in your own words. Students who violate these expectations can expect to receive a failing grade on the assignment and will be reported to Student Judicial Services. These types of violations are reported to professional schools, should you ever decide to apply one day. Don't do it—it's not worth the consequences.

Sharing of Course Materials is Prohibited

Any materials in this class that are not posted publicly may not be shared online or with anyone outside of the class unless you have my explicit, written permission. This includes but is not limited to lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional problem sets. Unauthorized sharing of materials promotes cheating. It is a violation of the University's Student Honor Code and an act of academic dishonesty. We are well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure in the course.

Any materials posted on the public class website (<https://wilkelab.org/SDS375/>) are considered public and can be shared under the Creative Commons Attribution CC BY 4.0 license.

Class Recordings

Any class recordings provided are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

Reuse

Text and figures are licensed under Creative Commons Attribution CC BY 4.0. Any computer code (R, HTML, CSS, etc.) in slides and worksheets, including in slide and worksheet sources, is also licensed under MIT. Note that figures in slides may be pulled in from external sources and may be licensed under different terms. For such images, image credits are available in the slide notes, accessible via pressing the letter 'p'.

Corrections

If you see mistakes or want to suggest changes, please [create an issue](#) on the source repository.