****

**STAT 436: Statistical Data Visualization**

**3 Credits**

**Course Designations and Attributes**

Natural Science

LAS Credit

Intermediate

**Course Description**

Techniques for visualization within data science workflows. Topics include data preparation; exploratory data analysis; spatial, tabular, and graph structured data; dimensionality reduction; model visualization and interpretability; interactive queries and navigation.

**Requisites**

(Stat 240 or 303), graduate/professional standing, member of Statistics Visiting International Scholars program

**Meeting Time and Location**

Tuesday/Thursday 2:30-3:45pm, 594 Van Hise Hall

**Instructional Modality**

In-person

**Specify How Credit Hours are Met by the Course**  
The credit standard for this course is met by an expectation of a total of 135 hours of student engagement

with the courses learning activities (45 hours per credit), which include regularly scheduled: readings,

recorded lectures, in-class exercises, a midterm exam, portfolio assignments, and a group project, as described in this syllabus.

**Regular and Substantive Student-Instructor Interaction**

Participation in regularly scheduled lectures each week will include the opportunity for direct interaction between students and the instructor. The instructor will also frequently interact and post announcements in Canvas and email students about academic aspects of the class.

Instructors & Teaching Assistants

**Instructor**

Kris Sankaran ([ksankaran@wisc.edu](mailto:ksankaran@wisc.edu))

~~Office Hours will be set by~~ [~~this poll~~](https://go.wisc.edu/p5k1sm)~~.~~

Tuesdays 3:45 – 5:00pm

Medical Science Center 7225C, and

[Zoom](https://uwmadison.zoom.us/j/8622164885) [ID: 862 216 4885, passcode: 298858]

**Teaching Assistants**

Margaret Turner ([mlturner3@wisc.edu](mailto:mlturner3@wisc.edu))

Fridays 2 – 3pm

Medical Sciences Center 1274

Course Learning Outcomes

1. Develop a vocabulary of visual encodings that support exploration of geospatial, temporal, tree-structured, and network data, and demonstrate facility implementing them using packages in the R programming language.
2. Design dynamic queries that support interactive visualization of heterogeneous data, and demonstrate facility implementing them using the shiny package in the R programming language.
3. Design effective visualizations to summarize the results of dimensionality reduction and clustering algorithms.
4. Use visual artifacts derived from complex statistical and machine learning models to discuss the patterns they learn and mistakes they make.
5. Evaluate strengths and weaknesses of visualizations in relation to the tasks they support.

**A picture containing blur

Description automatically generated**

**Grading**

In-Class Exercises: 15%

Midterm: 20%

Homework: 35%

Group Project: 25%

Grades will be assigned according to the percentage scale, A = 92-100, AB = 88-91.9, B = 82-87.9, BC = 78-81.9, C = 70-77.9, D = 60-69.9, F = 0-59.9 (92% of points => A); and according to the percentile scale, A = 75, AB = 65, B = 45, BC = 30, C = 10, D = 5, F = 0 (performing better than 75% of the class => A). Your grade will be the higher of these two grades.

Course Website, Learning Management System & Digital Instructional Tools

<https://canvas.wisc.edu/courses/346654>

<https://piazza.com/class/lc5ndbnznzy6q2>

<https://www.gradescope.com/courses/498858> (sign-up code: WV68WB)

Required Textbook, Software & Other Course Materials

* There are no required textbooks. All readings are linked in the table below.
* [The R Project for Statistical Computing](https://www.r-project.org/)
* [ggplot2 documentation](https://ggplot2.tidyverse.org/reference/)
* [shiny documentation](https://shiny.rstudio.com/)

Campus provides students with [technology guidelines and recommendations](https://it.wisc.edu/learn/guides/learning-online-technology-tips-tools/) for instruction. Students should consult these resources prior to the start of the semester.

Homework & Other Assignments

* In-Class Exercises
  + This class follows a partially flipped format. Practice problems will be introduced during class and time will be set aside for you to complete them then. Submissions must be made on Gradescope before the start of the next course session.
  + In-class exercises will be graded for completeness only. They are designed more for practice than evaluation.
  + We will discuss exercises in class, but your submission must reflect your original work. A rule of thumb is that you may speak freely with your peers or instructors but shouldn’t write anything verbatim from discussion.
* Homework Assignments
  + Four homework assignments will be released throughout the course. There will be two types of homework: Problems sets from the course exercise sheet and more open-ended portfolio assignments. The portfolio assignments are small-scale visualization projects that ask you to practice course concepts by identifying a dataset of interest, discussing relevant visual queries, preparing a visualization, and revising it based on peer feedback. Homework will be due **February 20**, **March 13**, **April 3**, and **May 1**.
  + You will provide constructive criticism for your peers’ portfolio submissions within one week of the original submission. Your peer review will be graded, and this score will be included in the homework component of the course grade.
* Both in-class exercises and homework must be submitted on Gradescope. For every 24 hours late that a submission is made, it will be penalized 5%, for up to 4 days, after which no submissions will be accepted. The only exception for late acceptance will be in documented medical or family emergencies.
* For help on problems after initial attempts, students are encouraged to ask the instructor or TA during office hours, so that all students can benefit equally from discussion.

Exams, Quizzes, Papers & Other Major Graded Work

* One midterm exam will take place in class on **March 23**. If you have exceptional circumstances which require a different exam date, please reach out to the teaching team as soon as possible.
* There will be a group final project. Students will work in a team of 3 - 5 people. Guiding project milestones will be due on **February 6**, **February 13**, **March 27**, and **May 7**.
* There is no final exam.
* No late submissions will be accepted for the group projects.
* Each student will rate their project team members. If a student has two or more peer-ratings below 75%, they will be penalized 25%. If below 50%, the penalty is increased to 50%.

|  |  |  |
| --- | --- | --- |
| **Week** | **Topic** | **Reading** |
| **1** | **Marks and Channels** | 1. [Fundamentals of Data Visualization: Chapters 2](https://clauswilke.com/dataviz/aesthetic-mapping.html) 2. [Introduction to Data Science: Chapter 7](https://rafalab.github.io/dsbook/ggplot2.html) 3. Optional: [Data Humanism](http://giorgialupi.com/data-humanism-my-manifesto-for-a-new-data-wold) |
| **2** | **Data Wrangling** | 1. [Fundamentals of Data Visualization: Chapter 5](https://clauswilke.com/dataviz/directory-of-visualizations.html) 2. [R for Data Science: Chapter 12](https://r4ds.had.co.nz/tidy-data.html) |
| **3** | **Faceting and Layout** | 1. [Fundamentals of Data Visualization: Chapter 21](https://clauswilke.com/dataviz/multi-panel-figures.html) 2. [Introduction to Data Science: Chapter 9.1 - 9.3, 9.7](https://rafalab.github.io/dsbook/gapminder.html) |
| **4** | **Interaction** | 1. [Visualization Analysis and Design: Chapter 12.1 - 12.3, 13.1 - 13.3](https://search.library.wisc.edu/catalog/9911196629502121) 2. [Mastering Shiny: Chapters 3, 4](https://mastering-shiny.org/basic-reactivity.html) 3. Optional:  [Up and Down the Ladder of Abstraction](http://worrydream.com/LadderOfAbstraction/) |
| **5** | **Interaction** | 1. [Mastering Shiny: Chapter 7](https://mastering-shiny.org/action-graphics.html?q=time%20ser#action-graphics) 2. [Crosstalk tutorial](https://emilyriederer.github.io/demo-crosstalk/tutorial/tutorial-rmd.html) |
| **6** | **Time Series** | 1. [Forecasting Principles and Practice: Chapter 2](https://otexts.com/fpp3/graphics.html) |
| **7** | **Geographic Data** | 1. [Geographic Data in R: Chapter 2](https://geocompr.robinlovelace.net/spatial-class.html) |
| **8** | **Network Data** | 1. [Visualization Analysis and Design: Chapter 9](https://search.library.wisc.edu/catalog/9911196629502121) 2. [Modern Statistics for Modern Biology: Chapter 10.1 - 10.2](http://web.stanford.edu/class/bios221/book/Chap-Graphs.html) |
| **9** | **Clustering** | 1. [Intro to Data Science: Chapter 34](https://rafalab.github.io/dsbook/clustering.html) 2. [Superheat Vignette (2, 3 & 6)](https://rlbarter.github.io/superheat/basic-usage.html) 3. [Cluster Analysis of Genomic Data](http://users.umiacs.umd.edu/~hcorrada/CMSC702/readings/Solutions_ch13.pdf) |
| **10** | **Dimensionality Reduction I** | 1. [Beginner’s Guide to Dimensionality Reduction](https://idyll.pub/post/visxai-dimensionality-reduction-1dbad0a67a092b007c526a45/) 2. [PCA and UMAP with tidymodels and #TidyTuesday cocktail recipes](https://juliasilge.com/blog/cocktail-recipes-umap/) 3. [Understanding UMAP](https://pair-code.github.io/understanding-umap/) |
| **11** | **Dimensionality Reduction II** | 1. [Text mining with R: Chapter 6](https://www.tidytextmining.com/topicmodeling.html) 2. [Visualizing the structure of RNA-seq expression data using grade of membership models](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006599) |
| **12** | **Model Building** | 1. [Partial-dependence Profiles](http://ema.drwhy.ai/partialDependenceProfiles.html) 2. [Visualization in Bayesian workflow](https://arxiv.org/abs/1709.01449) |
| **13** | **Visualizing Deep Learning Models** | 1. [Four Experiments in Handwriting with a Neural Network](https://distill.pub/2016/handwriting/) 2. [Visualizing what convnets learn](https://jjallaire.github.io/deep-learning-with-r-notebooks/notebooks/5.4-visualizing-what-convnets-learn.nb.html) 3. Optional: [Automation Makes Us Dumb](https://www-proquest-com.ezproxy.library.wisc.edu/docview/1626773318/E4C6F4FD50AF4437PQ/1?accountid=465) |
| **14** | **Conclusion** | 1. [655 Frustrations of Doing Data Visualization](https://www.datasketch.es/project/655-frustrations-doing-data-visualization) 2. [Tukey, Design Thinking, and Better Questions](https://simplystatistics.org/2019/04/17/tukey-design-thinking-and-better-questions/) 3. Optional: [A Brief History of Data Visualization](https://datavis.ca/papers/hbook.pdf) |

Teaching & Learning Data Transparency Statement

*The privacy and security of faculty, staff and students’ personal information is a top priority for UW-Madison. The university carefully evaluates and vets all campus-supported digital tools used to support teaching and learning, to help support success through*[learning analytics](https://teachlearn.provost.wisc.edu/learning-analytics/)*, and to enable proctoring capabilities. View the university’s full*[teaching and learning data transparency statement](https://teachlearn.provost.wisc.edu/teaching-and-learning-data-transparency-statement/)*.*

Privacy of Student Records & the Use of Audio Recorded Lectures Statement

*View* [more information about *FERPA*](https://registrar.wisc.edu/ferpa-facstaff/)*.*

Lecture materials and recordings for this course are protected intellectual property at UW-Madison. Students in this course may use the materials and recordings for their personal use related to participation in this class. Students may also take notes solely for their personal use. If a lecture is not already recorded, you are not authorized to record my lectures without my permission unless you are considered by the university to be a qualified student with a disability requiring accommodation. [Regent Policy Document 4-1] Students may not copy or have lecture materials and recordings outside of class, including posting on internet sites or selling to commercial entities. Students are also prohibited from providing or selling their personal notes to anyone else or being paid for taking notes by any person or commercial firm without the instructor’s express written permission. Unauthorized use of these copyrighted lecture materials and recordings constitutes copyright infringement and may be addressed under the university’s policies, UWS Chapters 14 and 17, governing student academic and non-academic misconduct.

Course Evaluations

Students will be provided with an opportunity to evaluate this course and your learning experience. Student participation is an integral component of this course, and your confidential feedback is important to me. I strongly encourage you to participate in the course evaluation.

UW-Madison uses a digital course evaluation survey tool called [AEFIS](https://kb.wisc.edu/luwmad/page.php?id=81069). For this course, you will receive an official email two weeks prior to the end of the semester, notifying you that your course evaluation is available. In the email you will receive a link to log into the course evaluation with your NetID. Evaluations are anonymous. Your participation is an integral component of this course, and your feedback is important to me. I strongly encourage you to participate in the course evaluation.

Students’ Rules, [Rights & Responsibilities](https://guide.wisc.edu/undergraduate/#rulesrightsandresponsibilitiestext)

Diversity & Inclusion Statement

[Diversity](https://diversity.wisc.edu/) is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals.The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background – people who as students, faculty, and staff serve Wisconsin and the world.

Academic Integrity Statement

By virtue of enrollment, each student agrees to uphold the high academic standards of the University of Wisconsin-Madison; academic misconduct is behavior that negatively impacts the integrity of the institution. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these previously listed acts are examples of misconduct which may result in disciplinary action. Examples of disciplinary action include, but is not limited to, failure on the assignment/course, written reprimand, disciplinary probation, suspension, or expulsion.

Accommodations for Students with Disabilities Statement

The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA. (See: [McBurney Disability Resource Center](https://mcburney.wisc.edu/))

[Academic Calendar & Religious Observances](https://secfac.wisc.edu/academic-calendar/)

*You can use the link above to provide your students with information about the current and future academic calendars, along with the university’s religious observance policy. As the start-date for the fall 2021 semester coincides with Rosh Hashanah, it is particularly important to reach out to your students and share your plans to provide flexibility for the first day of class.*