# DATS 2102: Data Visualization for Data Science

**Instructor**: Junjun Yin  
**Email**: [j.yin@gwu.edu](mailto:j.yin@gwu.edu)  
**Semester**: Fall 2025  
**Dates**: 08/25/25 – 12/08/25  
**Class Time**: Tuesday & Thursday, 11:10am – 12:25pm  
**Location**: Tompkins Hall (TOMP) 107, 725 23rd Street NW  
**Office Hours**: Wednesday, 11:00am – 12:30pm (or by appointment)  
**Office Location**: 2036 H St NW, Room 309

## Course Description

This course introduces students to the core principles and practices of data visualization within the context of data science. Students will learn how to collect, process, analyze, and communicate data-driven insights using effective and ethical visualization techniques. Emphasis will be placed on hands-on programming with Python’s visualization ecosystem (pandas, matplotlib, seaborn, plotly, altair, geopandas) and applying best practices for clarity, accuracy, and storytelling. The course will cover visualization theory, design principles, and practical skills, including geographic data mapping and visualizing results from machine learning models. By the end of the semester, students will be able to produce high-quality visualizations that effectively communicate data insights to diverse audiences.

## Course Prerequisites

DATS 1001 and STAT 1051/1053/1111/1127, or permission of the instructor.

## Learning Outcomes

As a result of completing this course, students will be able to: 1. Process and tidy real-world data using pandas. 2. Apply visual perception and design principles to create truthful, clear graphics. 3. Visualize univariate, bivariate, and multivariate patterns; compare groups effectively. 4. Map and analyze geographic data using geopandas, contextily, and folium/plotly. 5. Visualize relationships and communicate model context and uncertainty. 6. Build interactive, annotated visuals and simple data stories/dashboards. 7. Apply visualization to ML & NLP tasks (feature importance, confusion matrices/ROC, word clouds, BERTopic topic maps, embedding plots).

## Course Workload

This is a 3-credit course. Students are expected to engage in 2.5 hours of direct instruction and a minimum of 5 hours of independent learning each week, for a combined minimum total of 7.5 hours per week or 112.5 hours over the semester.

## Required Tools and Texts

**Tools**: Anaconda (or Python 3.10+), JupyterLab, VS Code, Sublime Text, PyCharm, Google Colab, or other tools that support Python programming and visualization.  
**Core Libraries**: pandas, numpy, matplotlib, seaborn, altair, plotly, geopandas, mapclassify, contextily, folium, scikit-learn, umap-learn, sentence-transformers, bertopic, wordcloud.

**Documentation & Guides**:  
- [Matplotlib](https://matplotlib.org/stable/contents.html)  
- [Seaborn](https://seaborn.pydata.org/)  
- [Plotly](https://plotly.com/python/)  
- [Altair](https://altair-viz.github.io/)  
- [GeoPandas](https://geopandas.org/)  
- [Pandas](https://pandas.pydata.org/docs/)

**Texts**: No required textbook. Recommended: *Fundamentals of Data Visualization* by Claus O. Wilke (available free online) and *Storytelling with Data* by Cole Nussbaumer Knaflic.

## Weekly Topics & Schedule

| Week | Topic | Description |
| --- | --- | --- |
| 1 | Getting Started | Python setup, Jupyter, pandas basics, first plot with matplotlib. |
| 2 | Language of Graphs | Encodings, tidy data, seaborn & altair grammar. |
| 3 | Distributions & Variation | Hist/KDE/violin/ECDF, binning & outliers. |
| 4 | Wrangling with pandas | select/filter/mutate/groupby/merge, reshape, dates. |
| 5 | Perception & Principles | Cleveland–McGill, preattentive features, clutter. |
| 6 | Comparisons | Bars/dots/small multiples, ordering & baselines, log scales. |
| 7 | Text, Labels, & Tables | Direct labeling, captions, tables. |
| 8 | Mapping I | Choropleths, CRS, spatial joins, geopandas, mapclassify, folium. |
| 9 | Color & Accessibility | Sequential/diverging/qualitative palettes, pitfalls. |
| 10 | Relationships & Modeling | Scatter/line, smoothing, statsmodels, model checks. |
| 11 | Uncertainty | Error bars, intervals, bootstrap visuals. |
| 12 | Visualization for ML/NLP | Feature importance, confusion/ROC-PR, word clouds, BERTopic, UMAP embeddings. |
| 13–14 | Final Project Workshops | Scoping, refinement, narrative. |

## Assignments & Grading

| Assignment | Weight |
| --- | --- |
| Weekly Notebooks & Exercises | 40% |
| Mid-Semester Visualization Project | 15% |
| Final Project | 25% |
| Participation & Peer Feedback | 10% |
| Quizzes (2 total) | 10% |

## Final Project

The final project will synthesize the skills learned throughout the course. Students will: - Propose a project idea by Week 9. - Develop a prototype by Week 13. - Submit the final project by December 8.

Requirements: - Multiple well-designed visualizations with an accompanying narrative. - At least one map or ML/NLP visualization. - Accessibility considerations (color choice, labeling, alt text). - A reproducible Jupyter Notebook and any necessary datasets or data sources.

Projects will be graded on clarity, creativity, technical proficiency, and adherence to visualization best practices.

## University Policies

**Academic Integrity Code**  
Academic integrity is an essential part of the educational process, and all members of the GW community take these matters very seriously. As the instructor of record for this course, my role is to provide clear expectations and uphold them in all assessments. Violations of academic integrity occur when students fail to cite research sources properly, engage in unauthorized collaboration, falsify data, and otherwise violate the Code of Academic Integrity. If you have any questions about whether particular academic practices or resources are permitted, you should ask me for clarification. If you are reported for an academic integrity violation, you should contact Conflict Education and Student Accountability (CESA) to learn more about your rights and options. Consequences can range from failure of assignment to expulsion from the University and may include a transcript notation. More info: [students.gwu.edu/code-academic-integrity](https://students.gwu.edu/code-academic-integrity) or cesa@gwu.edu.

**University policy on observance of religious holidays**  
Students must notify faculty during the first week of the semester, or as early as possible, but no later than three weeks prior to the absence, of their intention to be absent for religious observance. See details at [provost.gwu.edu/policies-procedures-and-guidelines](https://provost.gwu.edu/policies-procedures-and-guidelines).

**Use of Electronic Course Materials and Class Recordings**  
Students are encouraged to use electronic course materials for private personal use in connection with their academic program of study. These materials should not be shared or used for non-course related purposes unless express permission is granted by the instructor.

## Academic Support

**Academic Commons**  
Academic Commons is the central location for academic support resources for GW students. To schedule a peer tutoring session for a variety of courses visit [go.gwu.edu/tutoring](https://go.gwu.edu/tutoring). Visit [academiccommons.gwu.edu](https://academiccommons.gwu.edu) for study skills tips, finding help with research, and connecting with other campus resources. For questions email academiccommons@gwu.edu.

**GW Writing Center**  
GW Writing Center cultivates confident writers in the University community by facilitating collaborative, critical, and inclusive conversations at all stages of the writing process. Working alongside peer mentors, writers develop strategies to write independently in academic and public settings. Appointments can be booked online at [gwu.mywconline.com](https://gwu.mywconline.com).

**Disability Support Services (DSS)**  
Any student who may need an accommodation based on the potential impact of a disability should contact Disability Support Services at [disabilitysupport.gwu.edu](https://disabilitysupport.gwu.edu) to establish eligibility and coordinate reasonable accommodations.

**Student Health Center**  
The Student Health Center (SHC) offers medical, counseling/psychological, and psychiatric services to GW students. More information about the SHC is available at [healthcenter.gwu.edu](https://healthcenter.gwu.edu). Students experiencing a medical or mental health emergency on campus should contact GW Emergency Services at 202-994-6111, or off campus at 911.

## GW Campus Emergency Information

**GW Emergency Services**: 202-994-6111  
For situation-specific instructions, refer to GW’s Emergency Procedures guide.

**GW Alert**  
GW Alert is an emergency notification system that sends alerts to the GW community. GW requests students, faculty, and staff maintain current contact information by logging on to [alert.gwu.edu](https://alert.gwu.edu). Alerts are sent via email, text, social media, and other means, including the Guardian app.

**Protective Actions**  
GW prescribes four protective actions that can be issued by university officials depending on the type of emergency. All GW community members are expected to follow directions according to the specified protective action: Shelter, Evacuate, Secure, and Lockdown. Learn more at [safety.gwu.edu/gw-standard-emergency-statuses](https://safety.gwu.edu/gw-standard-emergency-statuses).