# PROJECT MILESTONE 1: STUDY GOALS AND LITERATURE REVIEW

**Instructions**

1. Have one member of your team submit a write-up on Canvas. This write-up should be in a PDF, Word, or other well-formatted text document.
2. Write-ups must be no longer than 1200 words.

# Learning Outcomes

By the end of this milestone, you will be able to,

1. Discuss the goals of a data analysis problem within its larger context.
2. Search for and explain literature on data visualization relevant to a problem domain.
3. Identify key themes from across a broad collection of visualization concepts.

# Overall Description

You and your teammates have a choice between two types of projects. In both, you will choose a dataset or datasets and questions to study, and develop your interface or report over the course of the semester.

* + Exploratory Interface: Many experts in data visualization build interfaces that help a specific audience accurately answer questions about their data. For example, newspapers often have stories built from interactive visualizations, and scientific research labs often use dashboards to check and learn from experiments. In this type of project, you will design an interface to help your readers reach their analysis goals. You will map out relevant visual tasks and then design a suitable interface.
  + Critical Report: A real-world statistical investigation often culminates in a report that answers their audience’s questions. The report must present a logical series of visualizations that highlight the most salient aspects of the data, anticipate potential critiques, and presents multiple angles to give readers confidence in the answers that have been shared.

You may switch modalities later if your group decides it is appropriate.

When considering which data to use within the project, think carefully about whether it helps your audience answer its essential questions. Do not simply choose a dataset because it is easy to access or because it has many rows or variables. To evaluate a dataset and its collection methodology, consider whether the data is truly representative of the relevant population, and whether the data is comprised of objective, precise measurements. No dataset is perfect, and part of your later analysis will be to check for biases and reliability. Nonetheless, the more you find the data meaningful, the easier it will be to study it in depth.

You will develop your project over the course of the semester. Please choose a topic in an area that you are genuinely interested! Consider this project as an opportunity to develop your team’s critical and creative data science skills. If you are involved in a research project on campus, you may use them as a source of data for either type of project. Nonetheless, don’t confuse “seriousness” with “interest.” We have seen great projects about personal hobbies, like music, sports, art, and astrology, and some about business and healthcare topics were too distant from any teammates’ expertise to allow for in-depth study.

Professional examples belonging to each project type are given at the end of these instructions. We do not expect you to write a report at that professional level in a one-semester introduction to data visualization — view these instead as ideal reports within each project modality.

# Milestone Description

For this initial step, you will choose a project type, focus on a problem of interest, and survey visualization literature relevant to the task. We expect that by the time you submit your milestone, you have become broadly familiar with prior work in the area and to have read several sources in close detail. In your writing, do not simply list facts — explain key themes and develop your own commentary (e.g., how are proposed methods related to what we have learned in class, what are the main existing approaches, which visual idioms recur most frequently, what concerns do you have about existing work…). Be generous with including diagrams or screenshots to illustrate your points.

When you structure your report, make sure the following points are clearly visible,

1. Motivation and goals: You should identify at least one real dataset and explain several tasks/questions you would like your visualizations to support/answer. Describe your intended audience (e.g. how much prior knowledge do they have about the field?) and the most important questions you want to help them answer.
2. Literature review: Prepare an overview of visualization research literature and / or packages related to your problem. Provide a discussion of the main challenges, approaches, themes, and packages that may be relevant to your project.

# Recommended Roles

There are many sub-tasks and responsibilities that go into each group project milestone; your group is responsible for dividing them appropriately. However, we encourage you to use these recommended roles as a starting point for that conversation. **You are welcome to edit these roles as the group agrees. However, the group maintains collective responsibility for ensuring everyone knows their responsibilities, as well as agreeing on a project topic, modality, and dataset.**

* **Coordinator:** Set up group meetings and documents, take detailed notes on group meetings and other communications, document progress and clearly communicate responsibilities and schedules, aid data reviewer and literature reviewer in collecting suggested datasets and literature sources, responsible for final formatting/grammar/citation/fact-checking/overall cohesion review of write-up and submission to Canvas.
* **Data Reviewer:**Collect candidate datasets alongside Coordinator. Critically engage with the group's chosen project purpose as it relates to multiple candidate datasets, taking notes on important variables, biases, and features of datasets. Lead group discussion on which dataset to use, aid Writer in describing dataset.
* **Literature Reviewer:** Collect sources for literature review alongside Coordinator. Critically engage with the group's chosen project purpose as it relates to multiple literature sources, taking notes on main challenges, approaches, themes, and packages relevant to the visualization problem at hand. Lead group discussion on which literature sources to include, aid Writer in describing them.
* **Writer:**Writer is primarily responsible for creating and formatting the write-up. This will involve talking with the other members as appropriate to ensure it effectively communicates the group's cohesive, high-level project approach and purpose, as well as how the selected dataset and literature sources relate to that central purpose, and critically engaging with the chosen dataset and literature sources.

# Rubric

*Motivation and goals* (9 points): The report demonstrates that its authors have immersed themselves in the problem context and make a compelling case for their visualization goals.

*Literature review* (9 points): The report draws from a variety of complementary resources and connects concepts across them. Commentary demonstrates a deep familiarity with prior work.

*Clarity and style* (7 points): The writing is compact, well-structured, and free from technical errors. Figures are annotated and citations are formatted consistently.

# Professional Examples

# (These are NOT meant to be an indication of the quality you have to provide, but use these as inspiration!)

# Interactive Dashboards

1. [Poemage: Visualizing the Sonic Topology of a Poem](https://miriah.github.io/publications/poemage.pdf)
2. [Voltaire’s Correspondence Network](http://republicofletters.stanford.edu/publications/voltaire/letters/)
3. [Metagenomics with The Banfield Lab](https://stamen.com/work/banfield/)
4. [Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time](https://doi.org/10.1109/TVCG.2012.291)
5. [Ocupado: Visualizing Location-Based Counts Over Time Across Buildings](https://doi.org/10.1111/cgf.13968)
6. [EcoLens: Integration and interactive visualization of ecological datasets](https://doi.org/10.1016/j.ecoinf.2007.03.005)
7. [MatrixQCvis: shiny-based interactive data quality exploration for omics data](https://doi.org/10.1093/bioinformatics/btab748)
8. [Visinity: Visual Spatial Neighborhood Analysis for Multiplexed Tissue Imaging Data](https://doi.org/10.1109/TVCG.2022.3209378)

Reports

1. [The State of the World’s Children 2024 (Section 2)](https://www.unicef.org/media/165156/file/SOWC-2024-full-report-EN.pdf)
2. [Milky Way Census Shows Stars Take Varied Paths](https://www.scientificamerican.com/article/milky-way-census-shows-stars-take-varied-paths/)
3. [NASA’s World of Change](https://earthobservatory.nasa.gov/world-of-change)
4. [Americans’ Social Media Use](https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/)
5. Most articles from the [Upshot](https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.nytimes.com/international/section/upshot&ved=2ahUKEwjooKfOt_OKAxXmQkEAHdpeBUkQFnoECBgQAQ&usg=AOvVaw1Hb4rCAXIZHBphv8gc_RjV)

2