

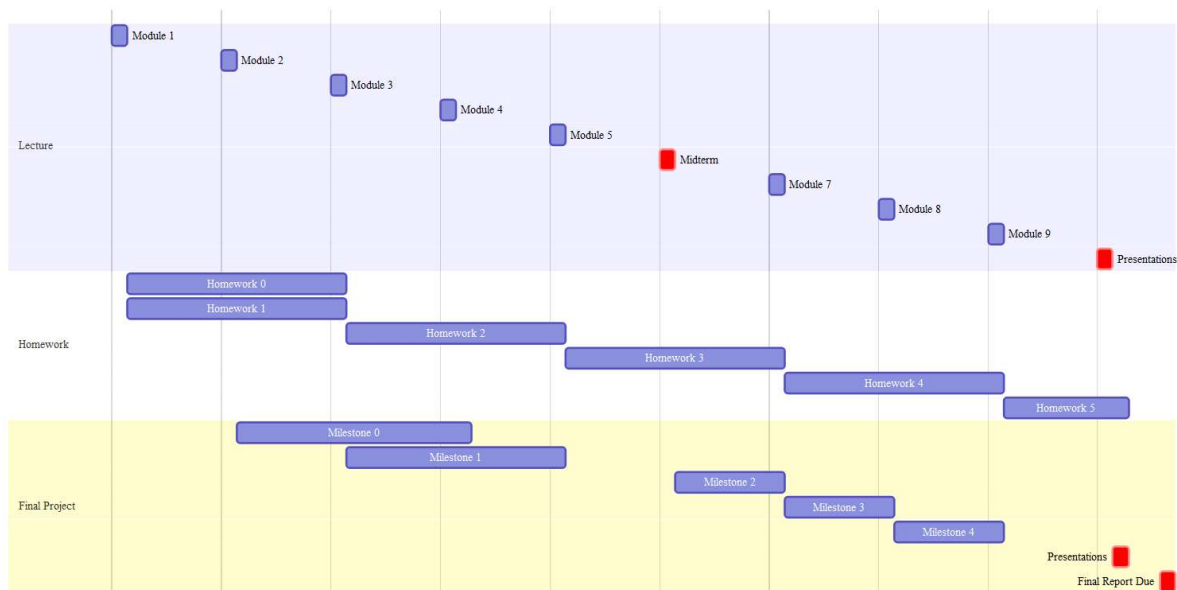
# DSC 465

## Course Project

### 1. Purpose

The goal of this project is to visualize a large or complex dataset with several different rich visualization techniques and get practice creating high-quality explanatory visualizations that communicate a story from your data. Take the opportunity to use visualization for data analysis through the full process, from finding and exploring data to expressing your findings.

### 2. Schedule and Milestones



The group project has five milestones (M0-4) and culminates in a presentation and a written report. The first two milestones are about establishing groups and happen in the first half of the class. After you submit proposed groups in Milestone 1 at the end of Week 5, I double check and approve them so that by the time you take the midterm, you are ready to start the project. There are three checkins during weeks 7, 8 and 9. We use week 10 for presentations and feedback. The final report is due at the time when the final exam would be if we had one (we do not). Note that grading is based almost exclusively on the report. See the Grading section for details.

For **Milestone 0**, everyone will present their project group preferences in the discussion forums. There is a template in the *Milestones 0 and 1* document to make sure everyone includes sufficient information. Posting is required. Some people will post about data or topics of interest, and groups often form around those ideas.

Use some time to form groups via conversations in the forums. **Milestone 1** requires all groups that have formed to submit a brief form indicating their members. This is due week 5 so I can review the groups and make sure everyone is covered. I will release official groups by the midterm so you can begin working together immediately afterward. I expect everyone will act professionally and courteously while we deal with any changes to group assignments that may be required.

**Milestones 2-5** are check-ins about group progress. The first will be about chosen data and directions, and by the last one I will expect a plan for how to finish the project. Each week I will review the materials and give feedback.

All groups will give a **Presentation** the last week of class. When the course is taught in-class, this means members must be there to present. If the course is online-in-class-hybrid, at least one person must be there to present. For fully-online offerings of this course, the presentations will be by video and commenting on other groups' presentations is required in lieu of feedback provided in class. Whether in class or online, I will also give feedback. The main purpose of these presentations is to gather feedback you can use to tune your final project report.

On the day of the final exam (which we do not have), the **Final Reports** are due. The final grading is based on the report, but the milestones on the way will help make sure everyone is on track.

Note that homework assignments 4 and 5 (**HW4** and **HW5**) are mostly about the project. They involve turning in parts of your individual contributions to the project. These assignments encourage you to try some of the visualizations we cover towards the end of the quarter, whether or not they end up getting used in the project.

### 3. Groups

The groups should be 5 people on average. Groups may vary in size, but that is the target. I want people to form groups on their own based on discussions (see Milestones 0 and 1), but I reserve the right to edit groups if necessary to make sure everyone has a suitable group.

It can be difficult to keep groups working smoothly together. The skills practiced in doing this will be useful to your careers. I expect you all to handle this professionally, courteously and respectfully. Deviations from this will adversely affect your grade in two ways: your participation score may be reduced, and your teammates may report their dissatisfaction on the peer evaluations.

I have a few suggestions that may be helpful. First, recognize that when you only communicate through text, you may miss your teammate's tone. Tone can be very important and if you misunderstand the meaning, frustration can compound as you all have a different understanding of what is happening. Be direct and clear. Talk in person or by video.

Second, notice that in PD 0, when you mention your group preferences for the project, one piece of information is when you are available for meetings. Consider this when forming groups because if anyone falls out of communication, there may be issues.

Finally, communication is key. Agreeing as you form your group on a communication protocol. For example, this could be email or any number of newer tools for managing work like Slack. Make sure everyone agrees to turn on alerts so no one is left behind.

If things go badly and get out of control, talk to me immediately. Note the policy at the bottom of this document about removing a group member is grave. Don't let it get that far.

## 4. Expectations

The aim of your deep analysis of your data is to draw out and display clearly a story from your data. The first part of the process is an exploratory analysis, using visualizations to explore variables and basic relationships. Your group should explore both relational (distribution, correlation, scatter, etc.) techniques as well higher dimensional methods (geospatial visualization, 2D or 3D plots, graph visualization, etc.). Think about what is showing in these graphs and look for patterns and stories that emerge. Then produce more graphs as you unearth the story and gather detail. The visualizations you use for this exploration do not need to be highly polished. Simply make sure you're not distorting data so you're getting a correct idea of the story.

Once you have the idea of your story, you transition to building explanatory visualizations to show it. These will be more sophisticated, detailed visualizations. The goal is deep plots that go beyond simple graphs to bring out interesting relationships in the data. This is a good time to break out some of the techniques we will discuss in the latter half of the course. You are welcome to use some of those for exploration, but I expect you will need them for explanation. These explanatory visualizations, highly-polished, will be the core visualizations of your final report.

Note that interactivity can help a lot with both exploratory and explanatory visualization, so please jump in and use it if you would like. In the report, include some screenshots and a description of how the interaction works.

The **number of visualizations** expected for the final report depends on the size of the group: at least 4 for groups of 5, but generally 5 unless the 4 include exceptional detail. If a group winds up with fewer than 4 people, fewer visualizations will be allowed. For one of the core, final visualizations, I will allow a highly polished and organized set of basic exploratory graphs that reveal interesting features of the underlying variables, but this should be more than just the collection of initial graphs thrown together. The set must fit together on a page.

## 5. Presentation

The presentation should focus on the data story and should be presented for a non-technical audience. Remember, the visualizations should be explanatory rather than exploratory, so when you present them, you should be using them to tell your data story. The final slides of your presentation can contain a show-and-tell of your best exploratory/technical images if you like, but note that time will probably be short.

## 6. Final Report

The final project report is almost the entire grade. This section provides details on expectations by giving an outline for the two major components plus a list of thought questions to help you figure out what to write. First, here are some overall important points:

- The report is submitted as a group as one PDF file. Additional files are allowed, e.g. for code or other artifacts as needed.
- In an appendix, each student must submit a personal writeup of their contribution and a reflection on what they have learned. This must be in the same document as the report.
- Clarity and organization are important – I cannot grade what I cannot find. Plus, it is essential to be able to communicate your analysis to others.
- Make explicit what visualizations should be graded. Include the visualization and its analysis alongside each other.
- You must explicitly quote any result from another source that you will be using in your analysis.
- Any developmental scripts/data analysis or materials that helped you build your visualizations should be included in the appendix.

### Structure/Outline

There are two major components of the report – group and individual. They must be included in the same document. That is, in one document include the group report plus all the individual reports.

1. **Group technical report:** not exceeding 7 pages, with an additional page for the display of each visualization, with details of your project. This section is intended for a visualization literate audience and must be written in a clear organized fashion. Here is the recommended structure for the document (use this unless deviation has a clear reason):
  - a. Introduction – explain the data you are using and introduce the story you will be telling. Include the variables you are working with.
  - b. Exploratory Analysis – show a sample of your exploratory visualizations and explain how you did your initial analysis and how it lead you to your story.
  - c. Visualizations – for each visualization, show the image and explain what type it is and how all the variables are mapped. Discuss how you refined it through your drafting process to bring out patterns and relationships. Then explain how it fits into your analysis/story.
  - d. Analysis and Discussion – wrap-up briefly with a review of what you discovered in your data and how you showed it.
  - e. Add appendices with your code and results of formative, exploratory data analysis.
2. **Individual report:** Each team member will submit a 1-2 page summary of their work on the project to be included as the first appendices of the main document. This should detail what role they played as a part of the team and what specific visualizations they performed/contributed to. In addition, include a short reflection summary of what you learned about data visualization in the project.

### What to Report

Here are some guiding questions to help you think about your process and what to write about:

1. A review of your data including what variables are involved and what commonalities/differences they have. What are the distributions/correlations/categorical levels involved? Are there any important variables that stand out as key variables that will be important for drawing conclusions?
2. What visualization techniques you applied and why they were appropriate to the data? You should discuss the kinds of data they explore and what benefits they have for visualizing the data relative to other methods.
3. How the methods you chose create rich and deep displays of the data that go beyond simple graphs.
4. The design criteria that went into your visualizations. You should discuss how color palette choice, scale choices, plot symbol/weight/size and how you arrived at the overall organization of your visualizations.
5. Whether coding/scripting was used to build the visualization. This is not required for every visualization, but your team is expected to explore all available paths towards building visualizations, and credit will be given for coding effort.
6. If you have included interactivity (not required, but helps move a visualization technique into a deeper/richer domain), you should discuss the type of interactivity and what its purpose is for either explaining or exploring the data.
7. How the techniques relate to and complement each other. How does each display a different side of the data and how does each visualization help understand aspects in other visualizations that you've done.
8. What conclusions can be drawn about the data from your visualizations and how might they help people analyze the data further. Are there any statistical techniques that might be suggested in the visualization that a data analyst might try?
9. If you had more time, where might you have liked to have developed your visualizations further?

## 7. Grading

The grading for the project will be based on a rubric. It is included below. The grading guidelines are helpful to consider when planning your work and report. There are two top-level categories: group and individual. The group portion is divided into: **structure** of the report, explanation of the **data**, **exploratory** investigation, **visualizations** and **conclusions**. The **milestones** component includes 15 points for completing the progress check-ins satisfactorily.

Each visualization is evaluated on the choice of visualization for the data/audience/message (did you make the right thing?), and the implementation of the specific visualization (did you make it well?).

The individual component is divided into: **structure** of your discussion, summary of your **contribution**, and personal **takeaway**, **evaluations** and **partner's evals**. The last two parts are about peer evaluations. I will provide a template for a peer evaluation and each of you must fill it out about each other student in the group. These are submitted to an individual submission box (separate from project). The **evaluation** row of the rubric indicates that you filled out and submitted evaluations, and the **partners' evals** row is how your group reflected on you (full credit unless there were concerns about your contribution).

*Note that the structure of the rubric mimics what should be the structure of your documents. For further guidance see the outline in the *Final Report* section.*

Some extra credit will be given to any group that researches and correctly applies a visualization technique not specifically covered in class. Your efforts in this direction must be clearly documented in your report. The amount of extra credit will depend on how different the technique is.

### Group Component

	<b>Excellent</b>	<b>Good/Fair</b>	<b>Poor</b>
<b>Layout and clarity</b>	Report is clear and neatly organized, with appropriate use of headings, tables and figures to enhance readability. Report is well written with appropriate use of grammar and visualization terminology. (10pts)	Report is mostly clear with parts that are not well organized under sections. Use of appropriate visualization terminology is limited. (5pts)	Report is not clear. The layout is cluttered and not organized in sections. Major editing and revision are required. Errors in spellings, capitalization, punctuation and grammar distract readers. (1pts)
<b>Data explanation</b>	Explanation of the data is clear and targeted towards a non-expert audience. (10pts)	Explanation of the data is reasonably clear but may contain some assumptions about audience or may be missing some important detail. (5 pts)	Explanation of the data is missing or confusing. Non-expert audience would be confused by jargon. (1 pts)
<b>Initial Exploratory Visualizations</b>	The core variables were well explored for their distributions and relationships. Time and location variables were explored. (5 pts)	Some crucial variables were missing from exploration. Some incorrect conclusions were drawn. (3 pts)	Variables were missing and conclusions were drawn incorrectly. (1 pts)
<b>Visualization Approach (per vis.)</b>	The approach is appropriate for the data, variables, and intended message. (5 pts)	The approach is not completed suited to the data or variables, or the interpretation is incorrect. (3 pts)	The approach is inappropriate for the data or variables or has seriously incorrect or misleading interpretation. (1 pts)

<b>Visualization Execution (per vis.)</b>	Design criteria covered in the class are adhered to, and the data is not distorted or cluttered. (5 pts)	The execution of the visualization has minor flaws. Visualization may have some clutter or data distortion. (3 pts)	The visualization is cluttered or data is distorted significantly. Design criteria for the visualization are not adhered to. (1 pts)
<b>Milestones 2-4</b>	All milestones were completed satisfactorily, demonstrating progress by the full group. (15 pts)	Some milestones were not completed or did not include sufficient progress or information. (8 pts)	Milestones were mostly either missing or insufficient. (1 pts)
<b>Conclusions</b>	Results from the three approaches results are correctly interpreted and conclusions are clearly reported. Including practical significance (5 pts)	Interpretation of results contains one or two errors. (3 pts)	Interpretation of results contains three or more errors, or is missing. (1 pts)
<b>Code</b>	Code is included in appendix (5pts)	Code is missing (0pts)	
<b>Deadline</b>	Information received by due date (0 pts)	Information not received by due date (-10 pts)	
<b>Extra Credit for technique not covered in class</b>	Analysis includes use of visualization technique beyond those covered in class (+ 5 pts)	No additional technique covered (+0 pts)	
<b>Total</b>	<b>/ 55 + 10 * #vis pts</b>	+ 5 points for possible extra credit	

### Individual Component

	<b>Excellent</b>	<b>Good/Fair</b>	<b>Poor</b>
<b>Layout and Clarity</b>	Report is clear and neatly organized, with appropriate use of	Report is mostly clear with parts that are not well organized under	Report is missing or unclear. The layout is cluttered and not

	headings and tables to enhance readability. Report is well written with appropriate use of grammar and terminology. (5 pts)	sections. Small amount inappropriate usage of visualization terminology is limited. (3 pts)	organized in sections. Major editing and revision are required. Errors in spellings, capitalization, punctuation and grammar distract readers. (2-0 pts)
Summary of Work	Summary describes your contributions clearly and concisely. Report shows evidence of clear and equitable contribution to the group’s effort. Effort is described in language that shows understanding and proper usage of the visualization methods applied. (5pts)	Parts of summary are unclear, may include some limited misunderstandings or misapplication of techniques, or report reveals some lack of equitable contribution. (3 pts)	Summary of work is missing, shows lack of equitable contribution, or summary shows misunderstanding of visualization techniques. (2-0 pts)
Summary of Takeaways	Thoughtful discussion of what you learned from the course, connected to the project and your results. (5 pts)	Report of what you learned lacks some clarity or is unspecific concerning what you learned about data visualization from the project. (3 pts)	Summary of takeaways is missing or lacks a clear description of what you learned about data visualization from the project. (2-0 pts)
Evaluation	Submitted your evals (5 pts)	Not submitted (0 pts)	
Partner’s Evals	Depends on the evaluation by your peers (10-0 pts)		
Total	/ 30 points		

## 8. Group Member Non-Performance

The final project in this course is very broad in its scope allowing your group to focus on a wide range of dataset types for visualization, and on a wide range of techniques for visualizing the data. Group members are expected to participate fully and equitably in the group,

Usually, the peer evaluation component of the grade and documentation, including the meeting minutes, in addition to an overall desire for excellence, is sufficient motivation for individuals to



contribute a fair share to the team project. However, in extreme cases, individuals have been known to completely cease contributing to a team project. If this is the case, a team has the right to notify the instructor **unanimously (other than the individual being sanctioned)** that the individual is no longer contributing and the team no longer wants the individual on the team.

It is expected that a team will be able to show significant effort towards reconciling the issue prior to such an extreme action. Note also that this is not a decision to be made lightly, as expulsion from a team will result in **the loss of 40% of the of the final project grade**, i.e. the group portion of the grade, for the person expelled. Because this is such a serious decision, any team that makes this decision will also experience a deduction of **10% of the final project grade**.