# DSC 190: Statistical Data Analysis and Inference
## Fall 2024

### HW and Final Project Submission Format

## Objective

One of the most important goals of this course is to learn how to write a data analysis report. The HW and Final Project are to be submitted in a format similar to a data analysis report. The difference is that the report in this class will be more structured so that it can be more easily composed and graded.

## Structure

The overall structure of the HW and Final Project report should be as follows:

0. Header
1. Introduction
2. Analysis
3. Conclusion(s)/Discussion
4. Appendix/Appendices

Now let's consider the basic outline of the data analysis report in more detail:

0. **Header.** This includes important general information:
   - *Title*: Choose a succinct but specific title that reflects the goals of the analysis.
   - *Author contributions*: A brief description of the respective contribution of each of the team members.
   - *Use of GPT*: A brief description of how GPT was used to help write the report, if relevant.

1. **Introduction.** Good features for the Introduction include:
   - Brief summary of the study and data, as well as any relevant substantive context, background, or framing issues.
   - The "big questions" answered by your data analyses, and summaries of your conclusions about these questions. These questions should include: 1) the questions posed by the HW prompts; 2) other questions that you may propose.
   - Brief outline of remainder of the report.

2. **Basic analysis.** In this format, the analysis is organized by research questions. Devote a subsection for each question raised in the Introduction. These questions should be organized according to the HW prompts. Within each subsection, statistical method, analyses, and conclusion would be described (for each question). For example:

   **2.1 Data processing and summaries**
   Methods
   Analysis
   Conclusions

   **2.2 Comparison between males and females**
   Methods
   Analysis

Conclusions

**2.3 Effect of Age**

Methods

Analysis

Conclusions

**Etc. . .**

3. **Advanced analysis.** This section contains analysis that goes beyond the HW prompts. It will display your own interest and creativity. It may include:
   - An additional analysis question, e.g. estimating another parameter, considering the effect of another variable in the data, evaluating the validity of statistical assumptions not considered, etc.
   - Using a more advanced analysis method to answer one of the HW questions or a new question.

4. **Conclusion(s)/Discussion.** This section closes the report:
   - *Conclusion summary*: It should reprise the questions and goals of the analysis stated in the introduction. It should also summarize the findings and compare them to the original goals.
   - *Discussion*: If relevant, include additional observations or details gleaned from the analysis section. If relevant, discuss relevance to the science and other studies. **Discuss data limitations.** New questions, future work, etc., can also be raised here.

5. **Appendix/Appendices.** This section is not mandatory but it may be necessary depending on what you do. This is the place to put details and ancillary materials, that is, materials that you want to include but would disturb the reading flow if they were put in the main text. These might include such items as
   - Technical descriptions of (unusual) statistical procedures
   - Detailed tables or computer output
   - Figures and Tables that were not central to the arguments presented in the body of the report

6. **Computer code.** In a general data analysis report, computer code may be included in the Appendix. **In our course, code should be submitted as a separate file**. Make sure to document your code by including appropriate section headers, text sentences, comments and annotations, to make it easier for the reader to follow what you are doing.

## Formatting and length

A good data analysis report should present all the necessary information in a concise fashion. To exercise this and facilitate grading, please abide to the following constraints:

- Use 12-point font for the main text, with full space between lines.
- Length guidelines:
  - Header + Introduction: 1 page
  - Each question: 1 to 2 pages each, including tables and figures
  - Advanced analysis: 1 to 2 pages, including tables and figures
  - Summary/conclusions/discussion: 1 page
- The total length of the report should not exceed 12 pages (not including Appendix or code). Any additional material, if it is really necessary, should go in the Appendix.

## Specific guidelines for the Final Project

The structure of the final project will be more similar to a general data analysis report. The structure will be the same as the one described above except for the following:

- You come up with your own data analysis questions, both Basic and Advanced. Do not simply replicate the tasks given in the description of the datasets in Kaggle.
- Propose 4 questions for the Basic Analysis section. The analysis here can be carried using methods covered in class.
- Propose 1 question for the Advanced Analysis section. This analysis may be a deeper look at one of the questions in the Basic Analysis, possibly using more advanced methods not covered in class, or a new question that you consider harder than the ones you proposed in the Basic Analysis.

## Presentation style

Points will be given for good presentation style and abiding to the formatting constraints. As a guideline, a good data analysis report has several important features:

- It is organized in a way that makes it easy for different audiences to skim/fish through it to find the topics and the level of detail that are of interest to them.
- The writing is plain and unremarkable, so that the content of the analysis is what the reader remembers, not distracting quirks or emotions in the writing. Examples of distractions include:
  - Overly formal or flowery prose.
  - Overly casual or overly brief prose.
  - Grammatical and spelling errors.
  - Placing the data analysis in too broad or too narrow a context for the questions of interest to your primary audience.
  - Focusing on process rather than reporting procedures and outcomes.
  - Getting bogged down in technical details, rather than presenting what is necessary to properly understand your conclusions on substantive questions of interest to the primary audience.
- Tables are well organized, with well labeled columns and rows. Do not make the table too large so that they can be easily followed and the reader does not get lost.
- Figures are well composed, with well labeled axes and large enough fonts. If relevant, use colors and line types to distinguish between different results and include a legend. Do not make the figure too busy so that it can be easily understood.

## Use of GPT

The use of GPT and similar tools is permitted, as long as it is acknowledged specifically. The guidelines described above still apply. The authors of the report are still responsible for the entire written content. Be careful with potential errors that GPT or similar tools may produce.