

Statistical & Data Sciences Writing Plan

EXECUTIVE SUMMARY

One goal of the Statistical & Data Science (SDS) program is that we want our students to be leaders in statistics and data science. This leadership includes excellence in writing. This document lays out our program's writing plan, including our writing goals for students and the proposed criteria for use in our curriculum to achieve these goals.



SDS was particularly interested in the WEC Program due in large part to our project-heavy curriculum and the writing products that we ask our students to create. Many of our courses culminate in a poster or a paper that reports on a statistical procedure or task undertaken by our students. Despite the emphasis of writing in our course projects, the faculty of SDS felt limited in our ability to effectively teach and evaluate writing.

Given the ubiquity of Rmarkdown (Rmd) files both in and beyond SDS's curriculum, we have decided to center our writing plan around Rmd files, using them as the vehicle both for procedural writing (like problem sets) and for summative writing (like project write-ups and posters). Using this consistent file type allows SDS faculty to coordinate the introduction of components of Rmd files across the curriculum. Additionally, our students will be able to develop their writing skills by focusing on the content, instead of being distracted by the minutiae of the different file types necessary to create high-quality documents, reports, presentations, and dashboards.

Our writing plan also makes thoughtful use of skills and resources both within SDS and at Smith more broadly. In our implementation plan, we are going to work closely with the Jacobson Center to develop our writing pedagogy as faculty and to develop writing-based training sessions for student peer-tutors and graders. Within the department, we will create rubrics for course projects and assignments that align with the writing criteria in this plan.

SECTION #1: WHAT CHARACTERIZES WRITING IN SDS FIELDS (ACADEMIC, PROFESSIONAL, ALL LEVELS)?

1. Contextual; depends on:
 - a. Purpose
 - b. Audience (about tone based on modalities/venue).
2. Translational and framed: depicting a specific contextual problem. Writing in SDS connects abstract ideas to people, places and things. (Goes to heart of difference between SDS and MTH/CSC)
3. Grounding in real-world problem solving with explicit connections between the problem/solution to the data analysis and collection methods. Considerations include: How is the work relevant in the real world? What are the peculiarities of the data? How is the data collection then impacting your analysis which is then impacting your conclusions?
4. Visual: dependent on quick readings of images. Toolbox of ways to express ideas, and you use the correct tool for the correct job (text, table, histogram, etc.). Not only multimodal, so making appropriate choices (including visual vs. text) is critical.

SECTION #2: WHAT WRITING ABILITIES SHOULD SDS MAJORS BE ABLE TO DEMONSTRATE BY THE TIME THEY GRADUATE?

Our writing goals are structured from the most consequential writing decisions to the more nuanced ones. We begin with the goals concerning selection of writing form and engaging in an iterative writing process. The third goal focuses on using precision and clarity of language. The next two goals concern structure with the fourth goal focused on the overall structure of the written piece while the fifth centers on the introduction. Impactful figures and tables are discussed in the sixth writing goal. Coding is an important kind of writing in statistical and data sciences; the final writing goal focuses on the nuances of coding.

By the time they graduate, Smith SDS majors should:

1. Be specific & competent in terms of types of writing — sci report, journalistic article, etc. — including:
 - When notebooks (like Rmarkdown) are appropriate as a final product (as opposed to a poster, paper, or presentation)
 - Level of conversation (blog post vs. peer-reviewed publication)
 - Adjust tone for the appropriate audience
2. Engage in a process (writing or coding), including
 - Brainstorming
 - Outlining
 - Initial drafting
 - Peer review
 - Editing
 - Revising
3. Write with precision and clear language, regardless of audience and venue. This includes being able to write about ambiguity and uncertainty.
 - Reason well about the complexity of what they did through reflection and exploration about their process without writing in an overly complex manner
 - Use precise, correct, and accurate language
 - Use present tense and understand the difference between active voice vs. passive (including using the former instead of the latter)
4. Prioritize the important parts of the process and/or project to communicate, including
 - Being critical about what is useful or necessary to communicate, instead of — for example — showing off the method that they think is fancy, cool, or most “advanced” regardless of the relevance of that tool
 - Constructing a sequence and flow of ideas that build on each other
 - Constructing narratives that tell a well-motivated story instead of reporting a chronolog listing every step that the student(s) did and every thought the student(s) had in the order that they had them
 - Using formatting intentionally and mindfully to organize ideas (and not just as bells and whistles)
5. Clearly communicating the research question and how analysis will support the question in the introduction and abstract (if applicable). The introduction (and abstract) should preview the following:
 - A persuasive argument and/or list of processes and analysis strategies
 - Results that are supported by analysis/evidence
6. Create impactful figures and tables, including

- Making well-reasoned choices concerning the kind of plot used for a figure; similarly making well-reasoned choices when including variables and results in a table
 - Seeking aesthetic balance in terms of colors selected, the ink to information ratio, etc. (colors, ink/info ratio, etc.)
 - Using visual elements that accurately reflect data representation (e.g., using a diverging palette for variables that have a natural midpoint)
 - Ensuring visualizations and tables are both legible and provide meaningful context (including labeling axes, using legible font sizes, and writing meaningful captions)
7. Code with documentation and comments that are correctly indented, use a consistent style, and are human-readable. Their code should:
- Be the result of an iterative process as outlined in Writing Goal 2 above
 - Provide minimally reproducible examples (e.g. using the repress package) when trying to recreate errors
 - Engage in technical writing when appropriate, including:
 - i. Writing a README.md for GitHub repositories
 - ii. Filing issues on GitHub

SECTION #3: INTEGRATION OF WRITING INTO THE DEPARTMENT'S CURRICULUM

At present, writing exists within the SDS curriculum in the following forms:

- Problem sets
- Lab assignments and activities
- Project write-ups
- Project posters
- Capstone papers

Many SDS assignments and materials are prepared in the Rmarkdown notebook format. The Rmarkdown (or Rmd) format allows for prose and code to coexist in the same document. This coexistence allows for narrative code and for computational narratives; that is, a statistician can share their methods both in the prose that is typical for scholarly articles and with code snippets of their methods being implemented. This is an implementation of literate programming (Knuth 1983).

Like a journal article or a scientific poster, there is an inherent structure to Rmd files supported by section headings, text blocks, code blocks, figures, and tables. Given the wide adoption of Rmd files in both educational and industrial settings, writing Rmd files that make thoughtful use of these elements is critical to the development of statistical and data science students. It also enables students to share their work with colleagues or prospective employers in a readable, reproducible, and portable format. This is our motivation for centering our writing plan around Rmd files.

By centering our writing enriched curriculum around Rmd files, we give our students a consistent vehicle for written assignments during their whole major experience. This consistency allows for us — as faculty — to develop a coherent curriculum that builds on previous courses work in Rmd. As such, the writing curriculum in SDS will demonstrate iteratively revisiting and improving writing skills in a procedural sense as well as in terms of our writing goals from Section 2.

SECTION #4: ASSESSMENT OF STUDENT WRITING INCLUDING THE CRITERIA

In our discussions, SDS was most excited about having explicit criteria for each writing goal with example rubrics (to be developed as part of our implementation plans). We list our criteria below for each writing goal. These can be adapted into rubrics or into assignment prompts:

Writing Goal 1 - Be specific & competent in terms of types of writing

Demonstrates understanding of the venue through the selection of delivery medium.

Possible choices include:

- Blog post
- Journal-styled rendered Rmarkdown (rticles, papaja packages)
- Poster (e.g., with posterdown)
- Newspaper article

Demonstrates understanding of the audience and level of abstraction, including:

- Balancing tone between casual and technical
 - For example, using scientific tone for academic research reports while being more casual when writing for a lay audience (and not using too much jargon)
- Using types of visualization that are appropriate for audience (e.g., not assuming too much mathematical or information design knowledge for general audience)
 - Simple visualizations (e.g., bar charts) might be better than too wordy a piece.
- Paying attention to their assumptions about level of mathematical fluency/literacy
- Using a level of abstraction appropriate for audience
 - Starting with personal stories and experience are important for general audiences. For example: many NYTimes articles follow this flow: start by telling a story about an individual, zoom out to larger/high-level societal discussion, but then close by refocusing on the individual.
 - Jumping in at a high level appropriate for academic research papers

Create writing that is comparable to other writings for a similar audience and venue, including using citations and references that are appropriate to the venue.

Writing Goal 2 - Engage in a process (writing or coding)

With the six phases of an iterative process identified in this goal, we have potential criteria for each one:

- Brainstorming
 - a. An outcome of a good brainstorming session includes multiple ideas. For example, students could be asked to identify two “safe” ideas and one stretch/reach/beyond possibility idea. Alternatively, they could be asked for the lamest idea, the one they’ll probably do, the one they like the most, the craziest thing they can think of
 - b. In each case, students should articulate why each idea falls under the given classification.
- Outlining
 - a. A good outline includes evidence of multiple ideas, a scaffolded argument and may include ideas that didn’t work
 - b. In the case of coding, there may be pseudo-code and/or commented out code that didn’t work.
- Initial Drafting
 - a. A good initial draft is a complete document with an introduction, a complete argument borne out in the main text, and a conclusion

- b. Evidence of creativity and/or evidence of passion
- c. *Writing* - Describe relevant procedures (and removed the “chronology” aspects of the writing)
- d. *Code* - Only working relevant code included
- e. There should also be evidence of proofreading, including:
 - That the draft is readable
 - Free of basic typos (not perfect polish, but all errors that would be found by using spell check have been resolved)
 - The tools available to students to check the “easy” things (spell check, grammar check on google docs) have been used
 - Students should have read the document from start to finish with “fresh eyes” before submitting
 - In the case of code, the file compiles and produces no syntax errors
- Peer Review - There are two roles in a peer review: the author and the reviewer. We will denote these criteria separately:
 - a. Reviewer
 - Comments are empathetic “Comment as if it were you” Golden Rule
 - Offering specific advice and comments, as well as questions
 - Noting areas for improvement as well as highlighting strong parts of the work
 - b. Author
 - Full draft provided (see above criteria for the Initial Drafting)
 - If desired, questions or areas of concern noted to assist reviewer’s focus during the review
- Editing
 - a. We note that “editing” depends on the draft number and state. Earlier drafts will be inherently less polished than later ones. However for all drafts, students will
 - Complete line editing their own writing, including:
 - Checking for typos, grammar, consistency of tense
 - Checking for missing figure and table references
 - Checking for missing citations (i.e. the [?] in LaTeX)
 - Polish is commensurate with draft number/state, with increasingly fewer typos and grammatical errors
 - Ensuring that code is readable and/or meets a style guideline (e.g., run a linter, ensuring lines of code as much as possible don’t exceed 80 characters)
- Revision
 - a. Noting that a revision is a larger process than editing, students are expected to have improved and/or altered the structure of the work beyond simple line edits.
 - b. Students should visibly address at least 50% of the comments (including addressing comments with a response.)
 - Could include revision of figures, tables, etc
 - c. Creativity/passion for subject is honed (not sanitized), supported further with more evidence, citations, and/or process

Writing Goal 3 - Write with precision and clear language, regardless of audience and venue. This includes being able to write about ambiguity and uncertainty.

Writes with clarity, precision, and appropriate tone and tenor, including:

- Using active voice throughout
- Using present tense throughout. (Though future tense is appropriate for a section labeled “future work”)
- Detailing assumptions made of their data and by their methods

- Explaining their steps with a sequence using easy-to-follow logic
- Crisply defining terms and methods using precise language
- Detailing issues of complexity in an enumerative fashion (instead of simply stating “it’s complicated” or “it’s complex”)

Writing Goal 4 - Prioritize the important parts of the process and/or project to communicate

Creates a coherent structure for the work by prioritizing the important elements of the process and/or project. The work should:

- Announce an organizational logic so that readers know where they are going and why
- Directly communicate a scientific narrative using an overt logical structure: Moves from problem, to procedure, data, conclusions, and back to target problem
- State how the selected analysis methods facilitate answering the research question
- Describe the analysis method at a level appropriate for the primary audience
- Limit description of data cleaning tasks that support the narrative at hand
- Have justifications for methodology that are limited to the scope of problem/research question.

Writing Goal 5 - Clearly communicating the research question and how analysis will support the question in the introduction and abstract (if applicable).

For this writing goal, we have two options for criteria: one focused on an abstract and the second on an introduction.

Introduction Criteria

- For scoping the introduction, there are three parts:
 1. Setting the scene - Why should we care about this? (prior work and motivation)
Provide an explanation of who this is important to and why?
 2. What is your contribution to the conversation? This should be framed for a general audience without statistical jargon. (How should a person with no prior knowledge of this course (e.g., a parent, roommate, etc.) understand your work?)
 3. Explain how you are going to do part 2 (in one sentence)
- A good introduction:
 - Provides minimally necessary relevant background information to motivate the importance of the main question being asked, so that those without prior subject matter knowledge understand why the question is important to answer.
 - Identifies response variable (or outcome) and one or more of the most relevant explanatory variables, and specifies the scope (who or what is covered by the analysis?) and method of inquiry (what kind of model?).
 - Example - “Want to investigate relationship between X and Y among {Z}, and we want to investigate how this relationship is related to these constructs”
 - Previews the remainder of the paper in one (non-run on) sentence

Abstract Criteria

- A good abstract should:
 1. Set the scene: Explain and justify the importance of the work, and its constituency.
 2. Articulate the specific contributions of this work. This should be framed for a general audience without statistical jargon. A person with no prior knowledge of this course (a parent, roommate, etc.) should be able to understand this.
 3. Explicate methodology: Explain how the paper will articulate these contributions.

4. Provide the minimally necessary relevant background information to motivate the importance of the main question being asked, so those without prior subject matter knowledge understand why the question is important to answer.
5. Identify the response (or outcome) variable and one or more of the most relevant explanatory variables
6. Specify the scope (who or what is covered by the analysis?) and method of inquiry (what kind of model?)
 - Example - "Want to investigate relationship between Y and X, and we want to investigate how this relationship is related to these constructs"
7. Preview the remainder of the paper in one (non-run on) sentence

Writing Goal 6 - Create impactful figures and tables

For this writing goal, we have two options for criteria: one focused on visualizations and the second on tables

A good visualisation should:

- Include readable axis labels, units and legends
- Reveal the data
- Be attention grabbing, but uses a low ink density to data ratio
- Tell the same story of the task you are addressing in prose
- Is not reliant on the prose to understand the graphic
- Include a meaningful and self-contained caption with proper grammar
- Help the reader see themselves in the graphic (e.g., "That's my state")

A good table should:

- Use readable font in every cell
- Use vertical and horizontal lines to bound data and results
- Be sized to fit either in a column or on a single page
- Reveal meaningful information about the data, experiments, or results
- Enhance the story you are telling
- Not rely on the prose to explain the significance of the table's contents
- Include a meaningful and self-contained caption with proper grammar
- Include explicit units
- Avoid pseudoprecision / use appropriate digits of precision or scientific notation

Writing Goal 7 - Code with documentation and comments that are correctly indented, use a consistent style, and are human-readable.

Demonstrates code is the result of an iterative process, with a final product such that

- Code is readable, formatted, follows a style guide, linted, and contains only that which is necessary.
- Code expresses the ideas/logic/sequence used in analysis process

Code that is impenetrable to anyone but the author is clearly not the result of an iterative process

Can provide minimally reproducible examples when trying to recreate errors, such that an Instructor/TA can copy-and-paste code into a fresh R session and produce the exact same error. Examples of not meeting this goal include:

- Creating GitHub repos that are incomprehensible to outside reviewers

- Reporting errors in a manner that they are impossible to reproduce, or not fully specified.
- Using passive voice with no information to report errors (e.g., “it doesn’t work”, “my R hates me”)

Demonstrates engagement with technical writing when appropriate. Examples include:

- Creating a complete home page for GitHub repository that has a README file illustrating how to use project, and includes both brief of description of the goal and a simple example
- Writing GitHub issues in a manner so that they can be quickly understood and are accessible and inviting for collaboration from:
 - a. Your most important collaborator: Future you
 - b. Other current collaborators
 - c. Future collaborators

SECTION #5: SUMMARY OF IMPLEMENTATION PLANS AND REQUESTED SUPPORT

With our writing goals and criteria in place, the next step is incorporating both into our courses. To that end, we plan to focus SDS faculty efforts on creating rubrics and project prompts, including:

- Create rubrics with examples of student work (both positive and less successful ones)
- Course specific rubrics and project specific rubrics (for example)
- Starting project prompts (i.e., what do we give to students?)
- Potentially - Create a function in the SDS package with vector inputs to generate rubrics

Additionally, in an effort to be transparent with our efforts to incorporate writing into the SDS curriculum, we plan to develop a pamphlet on writing expectations for students. A version of this pamphlet would be on our department website.

To implement and test this initial writing plan within the SDS curriculum, we will follow the below timeline:

- **Summer 2020** - Draft of writing plan submitted to SDS Chair and Jacobson Center
 - Capstone work reviewed by Kinnaird and outside reader (TBD)
 - Implementation ideas deployed (if possible)
- **Fall 2020** - Beta test of criteria and writing plan (This is totally optional)
 - Use the plan in your courses, if you want
 - Implementation ideas deployed (as possible)
- **January 2021** - Edits to plan based on Faculty feedback from beta test
- **Spring 2021** - Deployment of full writing plan across SDS curriculum
- **Summer 2022** - Review and edit writing plan based on Spring 2021

We believe that the following would best support the deployment of this writing plan:

- As the writing center on campus, the Jacobson Center would play a critical role in the deployment of our writing plan. We have identified a few training programs that would support our writing plan’s success:
 - Training for faculty on how to teach writing, and writing resources, including
 - Workshop on drafting the prompts to include student success
 - Individual course consultations (perhaps with our WEC liaison)

- Lunch discussions as a department with our WEC liaison centered around writing topics applicable across the SDS curriculum
- Mimicking the success of our course assigned Spinelli tutors, we would like to connect with Jacobson Center peer writing tutors for individual classes or projects
 - The SDS faculty has sent nominations to Sara Eddy about SDS students to be peer writing tutors
 - Sara has begun the process of identifying and hiring students for this role with the approval of the Jacobson Center.
- In collaboration with Sara Eddy, the current WEC liaison, we will create an online archive of writing support materials and pedagogy. This repository will include:
 - Project prompts
 - Examples from Dr. Flash of materials from statistics departments who have completed the WEC program and incorporated writing across their curriculum
- In collaboration with both the Jacobson and Spinelli Centers, we would like to create training about giving writing feedback on a variety of SDS assignments for the Spinelli center tutors as well as SDS graders
- There are three implementation strategies that could require additional funding:
 - We would like to test our criteria on prior student work. To do this, we would like to offer a stipend (or other payment) for someone to apply the draft criteria to student work
 - We would like to bring Prof. Annette Vee of the University of Pittsburgh to discuss her recent book and paper on coding literacy
 - As we develop new curriculum to implement a writing-enriched SDS curriculum, we plan on applying for curricular grants through Sherrerd Center

SECTION #6: PROCESS USED TO CREATE THIS WRITING PLAN

Crafting this plan followed a version of the Writing Enriched Curriculum (WEC) by Dr. Pamela Flash (University of Minnesota), involving all tenure ladder members of the SDS faculty. The SDS faculty engaged in the four in-person meetings as detailed by WEC, two with Dr. Flash and all four included our WEC liaison, either Naila Moreira or Sara Eddy, from the Jacobson Center at Smith.

Beyond the synchronous meetings, we generated several materials to act as a starting point for various discussions and sections of this document. Between the first and the second meeting, we created a spreadsheet to map the characteristics that we identified as critical to writing in SDS to the courses that we currently teach. We filled this in during our second WEC meeting, and it provided the starting point for our conversations about how to integrate writing into the curriculum. Between the second and third meeting, we generated a series of (aspirational) expectations that we have for our students' fluency with the Rmarkdown (Rmd) file format, as they enter and exit all courses. We refer to this as the "I/O document" (or the "in and out document"). Then Katherine Kinnaird, as SDS's WEC liaison, mapped the writing goals (Section 2) to the expectations in this I/O document. We used this document as the starting point for our discussion about writing criteria in the third meeting. Between the third and fourth meeting, we developed our first draft for two of our writing criteria. The process of creating these criteria became the starting point for planning our next steps implementing our writing plan, including drafting the remaining five criteria. After the fourth (and final required meeting by WEC), the SDS faculty have continued to generate criteria.

To create the proposed SDS Writing Plan, Kinnaird used all the materials generated by SDS in the WEC process and synthesized them into a coherent plan. She then submitted this plan to

the faculty for comment and approval. Kinnaird then made the requested edits and has submitted this plan to both the Chair of SDS and the Director of the Jacobson Center.