

Portfolio Instructions

Dr. Jessica Hauschild-Eastman, Maj Justin Merrick

Introduction

All Math 378 students must submit a portfolio at mid-term and the end of the semester. The portfolio assesses your understanding and performance in the course. You will assess your grade in the course and provide a collection of materials that demonstrate and reflect on your growth regarding the course learning goals. There is not a final exam for this course; instead, the portfolio will be your final project. You will create the portfolio using the template provided in Posit Cloud.

Your final product must be a rendered PDF, submitted on Gradescope. Excluding the documentation and optional additional materials, it should not exceed 20 pages. Do not submit a Quarto (.qmd) file or save an HTML file as a PDF. Do not submit a document where TEMPLATE is included in the title, YOUR NAME is given as your name, or placeholder text has not been removed - adjust all components accordingly. Make sure to provide a complete, professional product.

Because there are no points assigned to assessments in this course and because there is some subjectivity involved in all methods of assigning grades, you will have the opportunity to give your instructor input as to what you think your Prog and Final grades should be through your portfolio presentation. The goal is to arrive at grades that accurately reflect your performance, development, and learning.

At prog and at the end of the semester, you will meet in-person with your instructor(s) to present your portfolio. Following each portfolio conference, you will receive a letter grade with feedback from your instructor. At prog, there will be no portfolio revisions. At the end of the semester, you will have one opportunity to contest your grade and provide additional evidence in a final revision of the portfolio (this is completely optional), due during finals week. Your course grade will be decided by your instructor, based on the quality of your work and the effort exhibited and explained in your portfolio. Your instructor is the final authority on your letter grade at prog and final. Prog grades will be whole letter grades, but plus and minus grades may be used for final grades in the course.

Due Dates

The due dates for each submission of the portfolio are specified below:

- Prog – Due T17 (Tuesday 24 February), prior to Portfolio Conferences M18-T19
- Final – Due T38 (Thursday 04 May), prior to Portfolio Conferences M39-T40
 - Optional Final Portfolio Revisions due Wednesday 08 May at 2359

Portfolio Contents

Your portfolio must contain the following:

1. A listing of your letter grades on each assessment included in the course grade so far. This will include your completion of the AI questionnaires and all video quizzes, homework quizzes, DataCamp, and GRs. Grades will be entered into the tables provided in the template.
2. A reflection on your preparation and performance on the GRs and optional GR Refinements.
3. A Novel Application.
4. Discussions that include examples demonstrating the degree to which you have met the learning goals of Math 378. The learning goals for this course are to:
 - Develop an applied knowledge of traditional and computationally intensive statistical models.
 - Advance, use, and improve computer programming skills.
 - Continue to develop the habits of mind to be an independent learner.

You will include examples from in-class work and assessments that demonstrate your level of accomplishment **and your ability to learn and improve** in these areas. You can refer to other sources or any and all conversation threads in your selected generative AI platform (e.g., ChatGPT) and provide relevant excerpts for your portfolio.

5. An assessment of your learning and effort in the form of an overall letter grade based on the provided grading criteria for the course.
6. An overall reflection of your learning and effort in the course.
7. Documentation statement.

8. **OPTIONAL** - Additional material/evidence supporting your claims of learning in the course. These might include, but are not limited to, a Statement of Accomplishment or other certificate of completion for additional online learning courses, selected excerpts from LLM transcripts to demonstrate your learning in a particular area, or additional code, models, or figures for the Novel Application.

Each of the above components will have its own section (and writing prompt, if applicable) in the provided template.

Portfolio contents - GR Reflection

After reviewing your GRs and the solutions, you will have the opportunity to complete an optional GR Refinement, where you will work with an LLM of your choice to correct mistakes and refine your understanding of the course material. You will submit the optional GR Refinement to Gradescope as a check of your progress toward the course learning goals.

Regardless of submitting optional GR Refinements, take some time to reflect on your performance and preparation for the three GRs in the course. Consider the following:

- What you learned
- What study techniques were successful
- How you utilized course materials (videos, textbook, video and homework quizzes, assigned conceptual and applied exercises, your peers, and/or LLMs) to prepare for the GR (e.g., did you work with your table group on video and homework quizzes, did you utilize an LLM for help coding or to create personalized practice problems)
- Your experience using LLMs to correct mistakes and refine your understanding of course material in the GR Refinements (required for the final end-of-semester portfolio only)
- Areas where you can improve

Specifically, reflect on how your preparation and performance on the GRs and your effort and learning on the GR Refinements aligned with the course learning outcomes. Write a 2-3 paragraph summary of your reflection, addressing each of these points.

Portfolio contents - Novel Application

To earn above a D in this course, you are required to demonstrate your independent learning by learning something new, beyond the provided course materials, and apply it to a data set. A student earning a C in Math 378 has explored a novel topic beyond the provided course materials, though they may not have fully applied this knowledge to a new data set or context. A student earning a B in Math 378 will explore a novel topic beyond the provided course materials, and they have made a solid attempt to apply this knowledge to a new data set or

context with some errors. A student earning an A in Math 378 will thoroughly explore a novel application, learning something new beyond the provided course materials and successfully apply it to a new data set or in a new way.

Examples

You are encouraged to use outside resources for the Novel Application. Some ideas for the Novel Application include, but are not limited to:

- Read a chapter in the ISLP textbook that is not covered this semester in Math 378 (Chapters 7, 9, 11, or 13), watch the associated YouTube videos from the authors, and apply that technique to a new data set. Some possible data sets will be provided by your instructors.
- Complete a DataCamp skills track (usually 3-8 related courses) on a topic new to you and apply it. For example, you might take a skills track on natural language processing and then apply those skills to a new data set like the Federalist papers or a data set of social media posts. As another example, you might take a skills track on building APIs and then build your own API. Skills tracks are not limited to the Python language. You may also consider online learning through another platform such as Coursera, Udemy, or Digital University. Example skills tracks in DataCamp include, but are not limited to:
 - Time Series in Python or Time Series in R
 - Natural Language Processing in Python
 - Machine Learning in Production in Python
 - Building APIs in Python
 - Big Data with PySpark or Big Data in R
 - Reinforcement Learning in Python
 - Interactive Data Visualization in R or Shiny Fundamentals in R
 - SQL Fundamentals
 - Julia Fundamentals
 - AI Fundamentals
 - AWS Cloud Practitioner or Microsoft Azure Fundamentals
 - Developing AI Applications or Developing Large Language Models
 - Github Foundations
- Complete courses or a specialization on DeepLearning.AI and apply those skills.

- Learn about Natural Language Processing, either through DataCamp or other sources, and apply that technique to a new data set.
- Read one or a few scholarly journal articles (number depending on length and complexity) on a topic new to you and apply that technique to a new data set.
- Learn a new topic (through reading journal articles, completing online courses, or reading a chapter in the course textbook) and write a lesson for Math 378 students. Your final product should be in the form of Notes to Instructors, an outline of content and activities for a 50-minute class period. If you're interested in this Novel Application, make sure to discuss with your instructor.
- Build an interactive web application to deploy models for this course (e.g., classification or tree-based techniques applied to a data set).
- Work with the OpenAI API to fine tune your own GPT or create a RAG.
- Other topics to learn about might include attention and transformers (related to generative AI), reinforcement learning, backpropagation and/or the gradient descent algorithm for neural networks, techniques such as the genetic algorithm or particle swarm, or anything else you want to learn about.

Data Sets

We have provided a few data set options for the Novel Application in the **data** folder.

- The [Bike Sharing data set](#) contains the hourly and daily count of rental bikes between years 2011 and 2012 in the Capital bikeshare system, along with weather and seasonal information: `Bikeshare.csv`
- The [Penguins data set](#) contains data for 344 penguins. There are three different species of penguins, collected from three islands in the Palmer Archipelago, Antarctica. There are 17 variables: `penguins_raw.csv`
- The Fuel data set contains data on 1,971 sorties flown around early October 2015. There are 46 variables: `Fuel_Data.xlsx`

Additionally, you can find data sets on the following websites:

- [UCI Machine Learning Repository](#)
- [538 Data](#)
- [Kaggle](#)
- [Data is Plural Newsletters](#)

- The Department of the Air Force E-Publishing ([EPUBS](#)): contains publications and forms from Air Force, Space Force, and Air National Guard. You may be able to scrape the PDF documents for a natural language processing application.
- Anywhere else you find data!

You are welcome to use additional data sets available in the **data** folder or in the **ISLP** package, but they should not have been previously used for your chosen topic. For example, if you are reading the chapter on Support Vector Machines, you should not use the **Heart** or **Khan** data sets because those are utilized in the chapter and lab.

It is your responsibility to choose a data set that is appropriate for your chosen application. If you have questions or need guidance, please see one of the Math 378 instructors.

Deliverable

You will propose your Novel Application topic and application to your instructor at the prog Portfolio Conferences and receive feedback on whether the topic and amount of work are appropriate for this course. If applicable, you will also share your chosen data set with the instructors at prog to ensure that it is appropriate for your chosen application.

For this section of the portfolio, you will provide an extended abstract and reflection of your work. This should be approximately 4-6 pages. If you have any questions about what to include or not, make sure to discuss with your instructor. Your extended abstract and reflection on your work must contain the following:

1. A discussion of your learning and the resources you used to complete the Novel Application, including how you utilized these resources to learn and improve your skills. Did you take a DataCamp skills track or work through a chapter in the textbook? Detail that here. You should also discuss whether and how you used resources such as Git tutorials, your instructor, your classmates, Google, or an LLM. Additionally, you should make sure to mention all sources used in the Documentation section of the portfolio. If using generative AI such as an LLM, you must address the five part documentation statement required for use of an LLM.
2. A detailed summary of your application. You should describe what you did to apply your learning to a new data set or in a new way. Did you perform an analysis of a new data set? Did you create a lesson for Math 378? Did you create an API or interactive web application? Detail that here. You must also include a final output from your application. Examples might include a few figures demonstrating a new modeling technique, code or model output for a new modeling technique, or even screen shots of an interactive web application demonstrating its use.

3. Discussions of the degree to which you have met the learning goals of Math 378 in regards to the Novel Application. A paragraph on each outcome should be sufficient. Include specific examples from your learning sources that demonstrate your level of accomplishment and your ability to learn and improve.
4. Answers to the following questions:
 - What skills, knowledge, or experience did you gain from working on this Novel Application?
 - What was the most interesting part of the Novel Application?
 - What were some of the challenges you faced working on the Novel Application? How did you persist through difficulties?
 - How might you improve or further your learning in this area in the near future (six months to a year)?
5. Additional Materials. You will not provide all code/output in the portfolio itself. However, if you use extensive code for your Novel Application, you should submit a separate document with all code/output included. Note that your code does not need to be production level. This is simply a way for your instructors to see how you approached the Novel Application and what you accomplished. This document may also include DataCamp certificates or LLM transcripts that support your claims of learning in the portfolio.

Final Reflection

Each one of us brings a lot to this community and we each have accomplished a lot this semester in this course. The goal of this section of the portfolio is to help you reflect on what you have accomplished and how you have grown as a statistician and person. Make sure to fully address each part of each prompt.

If you feel that your responses to the prompts overlap each other or other reflections (such as that for the Novel Application), it is okay to refer to your previous reflection rather than repeat your thoughts. This section of the portfolio is only required at the end of the semester.

Prompt 1

Thinking about the course learning goals, answer the following questions:

- What are your strengths and weaknesses, specifically related to the course learning goals?
- What was the most interesting or valuable part of the course?

- What were some of the challenges you faced in this course? How did you persist through difficulties?
- What other resources did you use throughout the course? How did you utilize these resources to learn and improve your skills? Discuss whether and how you used resources such as your instructor, your classmates, Google, or LLMs. If you used generative AI, we strongly recommend including transcript links.
- What is one way that you could improve in each of the three course learning goals - in your statistical thinking, in your computer programming skills, and in your abilities as an independent learner?

Give specific examples to describe your effort and learning in the course. Make sure to discuss what you have learned, how you have improved, and where you can go next.

Prompt 2

During the first class meeting, you wrote about the strengths you carried with you as you entered the room that day. On Lesson 35, you had an opportunity to consider the strengths you carry now, a full semester (and really, year) later. Revisit your notes from both in-class free-writes (contained in the First Day of Class and Chapter 12 projects on Posit Cloud), If applicable, feel free to revisit your strengths assignments from the Z-course. Then discuss the following questions:

- Have your strengths changed since the first day of class (either from the Z-course or Math 378)?
- Have old strengths been reshaped, and have you developed new strengths?
- Have the ways you think, talk, and write about your strengths changed?

Prompt 3

For the first strengths assignment, you reflected on a previous goal and the progress you had made towards it. You also wrote about one particular thing that you wanted to improve upon during this course. Look back at that first writing assignment and think about either goal. Then address the following questions:

- Did you accomplish your goal? Make sure to state what the goal was.
- Did your progress towards achieving that goal change you, your study habits, or your world view?
- What is *one* goal you have for the next six months, in personal, academic, or other areas of your life?
- What are *two* things that you will do to help yourself accomplish that goal?

Authorized Resources and Documentation

Authorized resources for this portfolio include anyone and anything, including generative AI at the Level 4: Use of GenAI to co-create and/or revise work per the Dean's Guidance for Academically-related GenAI at USAFA. However, the portfolio is INDIVIDUAL EFFORT. You may consult with fellow Math 378 statisticians and LLMs when developing the portfolio, but ALL explanations and reflections must reflect your own understanding/learning and you must provide documentation of all resources used.

Documentation statements should be detailed and specific, including the name of cadets who assisted you and how they helped. Any outside resources used for the Novel Application (e.g., websites, journal articles, coding assistance, etc.) must be documented and proper citations provided. If using an LLM such as ChatGPT, you must detail how the resource was used.

- You may NOT use an LLM to write any of the reflections or entire sections of your portfolio, or to complete the Novel Application for you. Use of AI in this way is equivalent to asking your friend to do your assignment for you, which is not permitted under the honor code (e.g., "Perform a linear regression analysis on the attached data set" or "Answer the following question: <INSERT HOMEWORK PROBLEM HERE>").
- You may use an LLM to edit and improve sections of your portfolio that you have written. Use of AI for editing your work is equivalent to asking your friend for editorial feedback on your work and should be documented as such (e.g., "Help me to make my explanation of logistic regression more clear and concise.").
- You may use an LLM to help you in the Novel Application. Example uses include brainstorming ideas based on your interests related to this course, helping you find outside resources related to your chosen topic, helping you find data sets, and providing coding assistance.
- If you have any questions about appropriate use of LLMs, please discuss with your instructor.

If using an LLM on any part of your portfolio, your documentation statement must include the following components:

1. The name of the GenAI tool
2. What you provided the LLM/an explanation of your prompts (a summary of your prompts is sufficient here)
3. What the LLM provided you/an explanation of the output (a summary of the outputs is sufficient here)
4. How you altered the LLM's output (e.g., how you iterated on prompts or how you edited the response before using it).

5. A full transcript of your interaction with LLMs, preferably in the form of a link. If a link is unavailable, you must provide a separate PDF of your full transcript(s).

Grade Descriptions

An “A” grade in Math 378

A student who has earned an A has developed a strong applied knowledge of traditional and computationally intensive statistical models. They thoroughly explore modeling techniques, identify appropriate models, and interpret and evaluate models and output. This student effectively utilizes generative AI to further their understanding of statistical concepts. They thoughtfully craft prompts, actively question responses, and iteratively improve solutions assisted by the AI. This student documents their use of generative models appropriately.

This student has greatly advanced, uses appropriately, and has substantially improved their computer programming skills, particularly in R and/or Python. They understand how to iterate code, debug errors, and how to find and utilize code examples from class, generative AI, and other online resources. They understand how to adjust the code to analyze their own data and can appropriately interpret the output. This student thoughtfully questions the examples to evaluate their appropriateness and iteratively improves solutions.

This student has fully embraced independent and lifelong learning. They take the initiative to explore statistical concepts in more depth. They actively participate in class and collaborate effectively with peers. This student continually improves their statistical knowledge and programming skills over the duration of the course, leveraging resources like their peers and instructor, online learning and generative AI. This student has thoroughly explored a novel application, learning something new beyond the provided course materials and successfully applying it to a new data set or in a new way.

A “B” grade in Math 378

A student who has earned a B has developed a solid applied knowledge of statistical models. They explore problems and can identify reasonable traditional and computational models. This student interprets output appropriately but may miss some nuances.

This student has improved their programming skills in R and/or Python. They write working code and understand core concepts but may have minor deficiencies with iteration, abstraction, or debugging. This student can find and implement existing code for statistical models, understand the output, and make minor adjustments as needed. They may occasionally struggle to identify the most appropriate code examples or fully grasp the underlying methodology. This student participates in class and works effectively with peers but relies more heavily on instructor guidance.

This student has made progress toward independent learning habits but not fully embraced them. They utilize generative AI to assist with their learning but occasionally requires guidance crafting effective prompts or questioning model responses. They document AI usage appropriately. This student shows development in their statistical and programming abilities over the semester, but growth may be inconsistent across topics. This student has explored

a novel topic beyond the provided course materials, and they have made a solid attempt to apply this knowledge to a new data set or context with some errors.

A “C” grade in Math 378

A student who has earned a C has developed basic applied statistical knowledge. They follow examples to build models but may struggle to explore problems independently or identify appropriate techniques. This student attempts to leverage generative AI to support their learning but struggles crafting prompts or iterating solutions with the model. Their documentation of AI usage is limited. This student shows marginal improvement in applied statistical knowledge, despite available resources.

Their code works but lacks understanding of how to iterate, debug errors, or abstract concepts. This student attempts to utilize existing code but requires guidance adjusting it to new data or interpreting the output. They have a basic grasp of core methods but limited ability to manipulate more complex examples.

This student participates in class inconsistently. They collaborate occasionally with peers but require significant instructor guidance. This student is beginning to take ownership of their learning but has not fully embraced lifelong habits. This student has explored a novel topic beyond the provided course materials, though they may not have fully applied this knowledge to a new data set or context.

A “D/F” grade in Math 378

A student who has earned a D/F has not adequately developed applied statistical knowledge or programming skills. They struggle to build models even from examples. Their code is inefficient or nonfunctional. This student cannot effectively find and implement existing code even with support. They do not understand how to adjust examples for new data or interpret model output. This student does not participate effectively in class activities.

This student demonstrates little to no growth in their statistical knowledge or programming skills throughout the duration of the course, neglecting to leverage available resources. This student does not effectively utilize generative AI as a learning tool, either by failing to craft appropriate prompts or neglecting to question and improve solutions. They likely do not document their use of the resource appropriately.

This student has not demonstrated the habits of an independent learner. They have not completed a novel application or shown any evidence of independent learning beyond the provided course materials