

MATH218 Final Project
Report Due: Monday, May 18th at 11:59 pm
Please submit an electronic copy of your project to:
alyford@middlebury.edu

The goal of this project is to use the techniques learned in this class to tell a story of your choosing. My hope is that you identify a project that both proves you've learned the material in this class *and* is of genuine interest to you. I have provided one such outline about what a project might look like, but I encourage you to propose an alternative project if there is something specific you would like to work on. You should submit your final report as a knitted R Markdown document in HTML form (unless some other form makes more sense for your project *and* you've cleared it with me).

Potential project outline:

- Identify a data set of interest
- Conduct basic exploratory data analysis describing the variables of interest
- Determine metrics for optimization (e.g., Does it make sense to use Accuracy? Kappa? MSE? Precision? Recall?)
- Compare and contrast several of the techniques learned in class for supervised learning (e.g., kNN, linear regression, linear discriminant analysis, decision trees, random forests, splines, etc.)
- Utilize unsupervised learning techniques where appropriate (e.g., k-means clustering, hierarchical clustering, PCA)
- Draw conclusions about the efficacy of each technique and how the best technique can help answer your research question(s) of interest

There is no required length for your project write-up. As always, you should reach out to me if you have any questions about your project. Generally speaking, if you test a unique set of research questions *correctly* using techniques learned in this class, you will do well on this project. Specifically, your grade will be a function of: How well do you communicate your findings? How thoroughly have you convinced yourself (and me) that you have identified the best statistical learning approach(es) to answer your research questions? How have you justified the metric(s) used to determine what constitutes the 'best' approach?

Good luck!