

Choose one of the following projects.

1 Develop

I think that one of the reasons R is becoming a popular programming language outside of the statistical sciences is because of the extensive work being done to develop and distribute packages. While many of these packages provide statistical tools (for instance, the package ‘`rnn`’ (link) allows users to fit Recurrent Neural Networks) some of the most useful ones have more general applications. For instance the package `knitr` link facilitates document generation through R, allowing a person to generate not only the text of the document, but the statistical, mathematical, and computational work that goes along with it. Shiny (link) provides this for webpages, allowing you to create interactive tools that can be deployed on the web. The package ‘`magrittr`’ provides a wonderfully useful tool (the pipe `%>%`) that allows you to chain functions together in a way that makes code more transparent and natural.

Details You may (and I would encourage this) work in groups for this project but you need to work in an environment that allows me to track what your progress is and how people are contributing - that means GitHub. If you want the work to be private I can provide you with a private repo to work from. Your results should include a link to the package, a sufficient amount of documentation, and an approximately two page write-up detailing why you think the package will be helpful.

2 Document

A classic literature report. Read a paper, summarize the important methods, use what you’ve learned to do something new - an application to a new problem, an alteration to the method, a simulation to illustrate its strengths and weaknesses, etc. Possible topics that might be interesting to you (and me): Block Bootstraps, Deep Learning, The Kernel Trick and other topics associated with Support Vector Machines, recent developments in Machine Learning, Kernel Smoothing, Relevance Vector Machines. Your selected paper would need approval before Thanksgiving break. I am more than happy to discuss the paper with you or help you find one. **Details** Not a group project. Your results should include a clear and concise write up outlining the paper and providing your analysis (approximately 5-10)

3 Data Analysis

Kaggle (link) hosts competitions in which you and a team can compete with people around the world to build the best predictive model for real-world data. They also allow instructors to set up competitions among their students. I have gotten a lot of fun out of competitions like these and have a competition in mind. If you would like to participate, organize a team and let me know which competitions you are interested in. If enough people are interested in this option, we can make the competition in-class.

Details Teams should be around 3-4 members. You will have the opportunity to submit results as the competition proceeds. You and your teammates should organize using something that can document your contributions (i.e., GitHub/Dropbox). Teams should submit a 5 page document explaining their approach to the problem.