

Investigation of COVID-19 Problem

This assignment is an optional assignment that can be undertaken and graded as a replacement for the final exam. You may instead choose to take the final exam online, proctored by your instructor.

For this final exam replacement assignment, you will investigate an aspect of COVID-19 that intrigues you. **This assignment may be completed individually or in teams of 2-4 students.**

By the due date for this assignment, **each team member will upload:**

- their own R or (Rmd + html) scripts (50 pts).

Also, by the appropriate due dates **one designated team member will upload:**

- a team proposal for your investigation (20 pts)
- a team report of your investigation (50 pts)
- slides for your team's 10-minute presentation (20 pts)
- an account of which team member performed what tasks (10 pts).

The **team proposal (due April 18th)** will be uploaded by the designated team member. You are encouraged to discuss your proposed investigation with your instructor during office hours and during the limited time set aside for discussion on Wed April 15th. Your proposal should include a ~2 paragraph description of the problem that you will investigate and its relevance, a list of data sources you plan to use, and a list of team members.

Any aspect of COVID-19 that intrigues you is fine to propose. One example is the association of statewide Shelter-In-Place orders with flattening of the curve of new COVID-19 cases. Another is associations of factors (Ex PPE, age, hospital capacity, shelter-in-place orders) with health worker infection/mortality. Another is the prediction of new COVID-19 deaths in specific region(s) 2 or 3 weeks into the future using a type of model that interests you (ex: regression, SVM, tree-based models, ANN). For predictions, **bonus points for correctly predicting the future.**

For your team scripts, you may divide EDA, modeling, etc. tasks among members. Please include:

1. Perform any needed data cleaning/preparation.
2. Display aggregate information on your data, perhaps stratified by age, gender, etc..
3. Create visualizations of univariate distributions of key variables.
4. Create visualizations of bivariate or trivariate relationships that tell the story of your data (ex: scatter plot of regional mortality vs time-to shelter-in-place order, aligned according to date of 20th case, colored by a measure of wealth.
5. Optional: Fit model(s) of your choosing (classification/regression). Be sure to evaluate your models using appropriate visualizations (ex: Residuals vs Fitted plots, ROC curves) and statistics (ex: sensitivity/recall, specificity, precision, confusion matrix, AUC).

For your final report and presentation:

6. Prepare a docx or (Rmd + html) final report of your findings: Let your visualization tell much of the story of the data. Be sure to interpret your EDA/visualizations and also evaluate and compare your models. Draw inferences from your investigations, and summarize what your investigation reveals along with its relevance.
7. Prepare 10-15 slides for a 10-minute presentation of your findings. After class feedback on your presentation, there will be time to incorporate any helpful suggestions into your written report.