Final project for ESE5023 (Fall 2020)

1. Goals

The final project of ESE5023 is designed to be a capstone of the course. It is intended to be an original data analysis using real data (preferred from your research group), treating the problem in-depth, and presenting your results in a formal manner.

The project contains two parts – a presentation and a report. There will be two presentation sessions: one is on Dec. 24, the other one is on Dec. 31, 2020. The report is due by Jan. 10, 2021.

The topic cannot be part of the student's research in the past. The topic and scope should be approved by the instructor. If you have problems in finding suitable topics, please contact the instructor before Dec. 01. A science (or data-science) question should be clearly articulated, and then addressed using tools covered in this class.

2. Requirements

a. Excellence in data wrangling
b. Excellence in data visualization
c. In addition to the above, at least 3 components from the list below should be applied:
   ✓ data cleaning and/or outlier removal, possibly including filling or interpolation
   ✓ comparing means
   ✓ linear modeling
   ✓ variable selection
   ✓ forecasting
   ✓ time-series analysis
   ✓ spatial data analysis
   ✓ other methods to model or statistically assess the data
   ✓ Linux usage
   ✓ FORTRAN usage

3. Report and presentation

a. The primary product is a written report with a clear style and format.
b. A full bibliography and citations to data sources.
c. The actual data sets that were used in the work (even if the data set is downloadable, the data set itself is required). Formats may be NetCDF, flat ascii, csv, or R-objects as .RData.
d. The code and scripts used in the analysis. The instructor and TAs should be able to reproduce your results by running the code and scripts.
e. A 10-minute presentation to the class, to be followed by a 3-minute discussion with the class.