STAT GU4241/GR5241 Summer 2020

HW will be submitted online in Courseworks as a .pdf file before the due date/time. Late HW is not accepted. All submitted HW (except code) needs to be hand-written (not typed) unless you have a documented disability, verified through disability services, that prevents this; or unless otherwise stated in the HW.
Show all work to get full credit.

TA Office Hours &Location:
LongZhao, lz2639@columbia.edu

You can find the codes and datasets that you will need for some of the HW problems on one of the resource pages for the book:
http://faculty.marshall.usc.edu/gareth-james/ISL/

If you use code/reference from these or other sites/resources make sure to cite them appropriately in your HW separately for each problem.
Make also sure to include a print-out copy of the code you are using in the HW separately for each problem.

*If you are working on the HW together with your classmates, you can submit one HW for a group of up to 3 students as long as the HW is accompanied by contribution statements clearly stating each student's contribution to the HW, in this case the students submitting the HW as a group will receive the same HW grade. Keep in mind that for the exams and the course each student is responsible from the full assigned set of HW problems.

HW1 (out: 06/01, Due: 06/06 before 9pm):

1) (20 points, will count towards each student's Project Grade)

Find 4 websites/sources of data on Coronavirus/Covid-19 Pandemic or on its implications. For each website/source of data, in at most a total of one page:
i) summarize the datasets available
ii) describe how/if at all the website/data source is in line with your interest for the project
iii) critique briefly the validity of the dataset (ex. data collection, possible errors, missing data, etc.)
iv) what would you be interested in learning from the datasets using statistical machine learning techniques? If possible, give one application for each of the following: a) a classification problem, b) regression analysis, c) cluster analysis.

2) (100 points) Solve the following exercises from the course textbook. Selected problems will be graded.
Ch 2: 1-3, 5, 6, 8, 10.
Ch 3: 3, 4, 14, 15.