Problem Set 0
6.S897/HST.956
Due: **Mon 2/11/19 at 11:59pm** through Stellar

The purpose of this exercise is for you to get access to sensitive healthcare datasets. This requires 3 steps:
1. CITI certification to ensure you are trained to handle health data.
2. Access to the MIMIC dataset.
3. Access to the IBM MarketScan dataset.

Once you complete all three of these steps (details below), please zip your CITI completion report and your IBM Data Use Agreement together and submit them via Stellar. Do so before the deadline.

[1] CITI certification
You will need to complete the CITI "Data or Specimens Only Research" course.

1. Log into the CITI site via the MIT portal (i.e. follow the link which is provided).
2. First register on the CITI program website, selecting "Massachusetts Institute of Technology (SSO)" as your organization affiliation (not "independent learner").
3. Follow the links to add a Massachusetts Institute of Technology Affiliates course. In the Human Subjects training category, select the "Data or Specimens Only Research" course.
4. Complete the course and save a copy of your completion report. The completion report lists all modules completed, with dates and scores.
5. You will submit this completion report twice: once to PhysioNet (for MIMIC access) and once to the course staff (via Stellar).

[2] MIMIC Access
Please complete the steps required to get access to MIMIC data outlined here. Please use your MIT email address username as your PhysioNetWorks username (otherwise, please email us your PhysioNetWorks username). One of the steps will require you to fill out a Data Use Agreement (DUA) where you will be asked for:

1. A reference name. Write David Sontag.
2. If asked for a phone number contact info, put 617-253-3476.
3. The general research area for which the data will be used: Write 6.S897/HST.956.

[3] IBM MarketScan Access
- Please download the IBM MarketScan Data Use Agreement from Stellar.
- Complete this agreement and upload it to Stellar.
- By signing this agreement, students are agreeing that they will not download any data from the IBM servers, nor will they attempt to re-identify any of the patients in the dataset.