SIESMIC AP Analysis Workflow

1. **Data Processing (Institution Specific -** see [here](https://docs.google.com/spreadsheets/d/1SzU4PcIEUsAGnKKyAcugHO2O2aZW29sf9a_cC-FAElk/edit#gid=1679989021) for shared SEISMIC variable names)
   1. Load pkgs
   2. Load full dataset
   3. Student level variables
      1. Rename std level variables
      2. Generate/recode std level variables
      3. Create std level variable dataframe
   4. Course level variables
      1. Rename crs level variables
      2. Generate/recode crs level variables
      3. Create crs level variable dataframe
      4. By course 1 and 2:
         1. Create dataframe of only first time taking 1st and 2nd course
   5. AP Level variables
      1. By AP subject
         1. Rename AP level variables
         2. Generate/recode AP level variables
         3. Create dataframe
   6. Create stacked dataset
      1. By course subject:
         1. Join dataframes for first time taking Course1, Course2, and AP
      2. Stack each subject by course
         1. Include new variable: “discipline” as flag for each subject
            1. BIO
            2. CHEM
            3. PHYS
      3. Should end up with something that looks like [this](https://drive.google.com/open?id=1Sj5kaFNGUkBhRoOH3cIPm-97UEBZmcFkbKGjzBbKWc0).
2. **Data Analysis (Same Across Institutions)**
   1. Load pkgs
   2. Load cleaned dataset
      1. Filter for student level inclusion/exclusion criteria
         1. Include:
            1. First time freshmen
            2. Took course 2
         2. Exclude:
            1. Transfer students
            2. International students
            3. Honors
   3. Create dataframes for analysis
      1. For each course:
         1. Took course 2 dataframe
            1. Filter by “discipline” flag (i.e. BIO, CHEM, PHYSICS)
            2. Filter by apyear (i.e. 2013 for BIO, etc)
         2. Took AP dataframe
            1. Filter by “discipline” flag (i.e. BIO, CHEM, PHYSICS)
            2. Filter by aptaker = 1
         3. Eligible to Skip dataframe
            1. Filter by “discipline” flag (i.e. BIO, CHEM, PHYSICS)
            2. Filter by apskipper = 1
         4. Eligible to Skip by each cutoff score
            1. For each cutoff score:

Filter by “discipline” flag (i.e. BIO, CHEM, PHYSICS)

Filter by apscore = (cutoff score, i.e., 3, 4, 5)

* 1. Analyses (see also [here](https://docs.google.com/spreadsheets/d/1rN8W_iz1mr7lEzBGfdTZHa45wKOSLiSF8VEpChCPsmE/edit#gid=129222174))
     1. RQ1
     2. For each course:
        1. RQ1a
           1. Sample: Took Course 2 dataframe
           2. DV: aptaker
           3. IV: factor(firstgen) + factor(lowincomflag) + factor(female) + factor(urm) + scale(hsgpa) + scale(mathsr) + scale(englsr)
           4. COV: factor(cohort)
        2. RQ1b
           1. Sample: Took AP dataframe
           2. DV: apscore
           3. IV: factor(firstgen) + factor(lowincomflag) + factor(female) + factor(urm) + scale(hsgpa) + scale(mathsr) + scale(englsr)
           4. COV: factor(crs\_term)

Etc….