July 23

I.   Project 3 - Specify and estimate a model using high resolution photos and population counts of Accra (Korle Gonno)
    A. Download the 4 compressed files from slack titled **Accra1.zip**, **Accra2.zip**, **Accra3.zip** and **Accra4.zip**. These files contain 10,000 equally sized, high resolution (~60cm), bird's eye images (orthophotos) of the Korle Gonno neighborhood in Accra. Also download the file titled **labels.csv** from the same location. This .csv file is also populated with 10,000 rows, with each row corresponding to the orthophoto with the same name. For example, row 1 in the .csv file corresponds to the image file 1.jpeg. Row 1000 corresponds to the image file 1000.jpeg and so forth. The .csv file has a second column that provides the counts of all persons estimated to have resided within that area at the time these orthophotos photos were taken (about Sept. 2005).
    B. Setup a file structure to house the orthophotos as well as the .csv file. It's a good idea to review some of the scripts you have produced from the tensorflow exercises as well as from the Maroney lectures (and those subsequent exercises) and consider their design as a model for your project file structure. For example you could create one folder to house training images and a second one to hold your test images.
    C. Import the orthophotos and labels into your python / tensorflow workspace. Creating training and test images as well as training and test labels.
    D. Specify a small toy model, just to see if you can train it using your arrays. First one I rand had only 12 observations. Don't worry too much about the size of your dataset at this point. Using the DNN model is also perfectly fine. After obtaining some model results, you could increase the source data size to 140 or 150.
    E. Next try to specify a more robust machine learning model that is appropriate for a computer vision problem such as the one presented by trying to identify the relationship between urban form and population. Think about the response variable and why it is important when further specifying arguments within the layers of your model. Continue to use the smaller model and see if setting batch size and steps as a promotion of input is effective at increasing processing speeds. If you have a GPU, try to employ it.
    F. Try to improve the predictive power of your model by increasing the training and test images & label dataset sizes to the maximum capacity your computer can handle. Modify layer specifications appropriately and seek to improve your model's predictive power. Consider validation measures to further assess your results.
    G. Await receipt of a randomly selected image from Accra and use it as input to predict the population of that area.
    H. Write a one to two page description regarding your work. Be sure to include plots and illustrations that describe your application of the selected machine learning method.
II.   For Monday: **Data** - identify at least 3 sources that can openly provide the data needed for input data to train your model. Identify the variables that you expect to use in the specification of your machine learning method. What feature will be seeking to predict as a means to solve your problem?
III. Daily responses from this week - due by Saturday 5PM. Project 3 due by **Sunday midnight**.