Assignment 1 (part 1-1)

- Purpose:
  - The purpose of this assignment is mainly to ensure that you each have a development environment that will **compile**, **test** and **run** Scala programs. If you successfully completed Assignment0, you should be all set. If not, go back and follow the appropriate instructions.

- How to do part 1:
  - Clone the class repo: https://github.com/rchillyard/CSYE7200
  - Using a command line "shell" (the *Terminal* app on a Mac, or if you're on Windows, you can create a shell from within *IDEA* or *Eclipse*):
    - `cd assignment-helloworld`
    - `sbt test run`
    - you will be asked which main program you wish to run: choose the *HelloWorld* option.

Assignment 1 (part 1-2)

- Alternatively…

  - you can create a project using IntelliJ or Eclipse and everything will (or *should*) be taken care of for you (just use the "Run" menu).

  - IntelliJ IDEA doesn't do a perfect job of importing projects (neither does Eclipse). But you should create a new project and choose "Import from Version Control." Provide the URL of the class repository and let IDEA do the rest. When prompted, you will need to provide a JDK (I recommend you use 1.8). It will probably require you to add a Scala SDK also (use any 2.12.x).

  - Once built, you should see a small green arrow by the classes that have main programs. Ideally, you just click…

  - If that doesn't work, try opening a Terminal shell; cd to the Hello World module, and type "sbt run." After doing that, IDEA should behave as described above. For some reason, it sometimes needs help getting started.

Assignment 1 (part 2-1)

- Go to kaggle.com
  - If you haven't already done so, register, log in and agree to their data usage terms (go to datasets).
  - Download this file:
    - https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset (you will need your cell phone to accept and enter a code)
  - Move it to some place that's easy to remember on your machine, for example: ~/kaggle/imdb-5000-movie-dataset.csv
  - If you have any difficulty with this, you can simple use the file provided.

Assignment 1 (part 2-2)

- Run the Ingest program (providing the argument "*~/kaggle/imdb-5000-movie-dataset.csv*" - or wherever you placed the file you downloaded from Kaggle)

  - You should have no problems with this so far — you should see a list of 5000 movies, each with its properties.

  - If you don't provide an argument to the Run configuration, then the program will use the built-in version of the dataset (provided in the class repository). It should be the same as the one you downloaded.

Assignment 1 (part 2-3)

- Now, edit the *Ingest.scala* file so that you only get to see the movies which originate in New Zealand
  - You should replace the line which prints the properties with these two lines:

    ```scala
    val kiwiMovies = for (m <- ingester(source); if (m.properties(20)=="New Zealand")) yield m
    println(kiwiMovies.size)
    ```

  - Let me know how many you get in your submission.