Steps to get notebooks/data into Databricks

1. Set up a free Databricks account for the community edition from https://accounts.cloud.databricks.com/registration.html#signup/community
2. Download entire github repo as a zip file from https://github.com/Lewkow/LAK_2017_Workshop
3. Unzip the repo on your local machine
4. Login to Databricks
5. **Import Notebooks**
    a. Go to Workspace -> Users -> <your username> and click arrow next to your name, select Import
    b. Import the notebooks from the unzipped repo, they are now in your Databricks workspace
6. **Create a Cluster**
    a. Click on Clusters icon on the left panel in Databricks
    b. Pick a name for the cluster and choose **Spark 2.1 (Auto-updating, Scala 2.10)** for the Spark version (This shouldn't matter, but it is good to be consistent)
    c. This will trigger a cluster to be created for you on Amazon Web Services (AWS)
7. **Import Data as Tables (https://docs.databricks.com/user-guide/importing-data.html)**
    a. Click on Tables icon on the left panel in Databricks
    b. Click Create Table
    c. Datasource -> File and drag and drop datafiles from LAK_2017_Workshop/data directory
    d. Copy names of files stored in Databricks somewhere (example: /FileStore/tables/gjfs9x661489346020543/train.csv)
    e. Click Preview Table, name titanic data "titanic"
    f. Click Create Table
    g. The data files are now in Databricks and are accessible from notebooks in Databricks