The folder you unzipped contains two files:

* CRDC2013\_14.csv: the actual subset of the data we'll explore
* CRDC2013\_14content.csv: an explanatory file that describes each of the columns in CRDC2013\_14.csv

The explanatory file, also known as a [data dictionary](https://en.wikipedia.org/wiki/Data_dictionary), was included because the dataset contains more than 2000 columns. Let's start by reading in the data dictionary into pandas and exploring it more.

Here's the workflow we'll use:

* use the editor to edit read.py
* save the file
* run read.py using the terminal

Recall that in order to run a script from the terminal, we need to add the following scaffolding to the code.



if \_\_name\_\_ == "\_\_main\_\_":

   ## Add code here

We'll have to run our Python script using a virtual environment that has a pandas installation on it. The virtual environment we need to activate is in the /dataquest/system/env/python3 folder. As you may recall from a previous mission, we'll need to run source /dataquest/system/env/python3/bin/activate to activate the virtual environment.

Instructions

* Using the terminal:
  + Activate the python3 virtual environment.
  + Run pip freeze to verify that pandas is installed and available.
* Using the editor:
  + Edit read.py so that it will run from the command line.
    - You'll notice that the tab for the current file you're editing changed from read.pyto \*.
  + Add code to read.py to:
    - Add the scaffolding necessary for the script to run from the terminal.
    - Import pandas.
    - Read CRDC2013\_14content.csv into a pandas dataframe named contents.
      * Print the first few rows of contents.
  + Save read.py.
    - Wait until the \* has disappeared in the tab and you can see read.py.
* Using the terminal:
  + Run read.py from the terminal, and verify that it worked properly.
* Continue iterating on the code to explore the contents dataframe and find any column names that interest you.