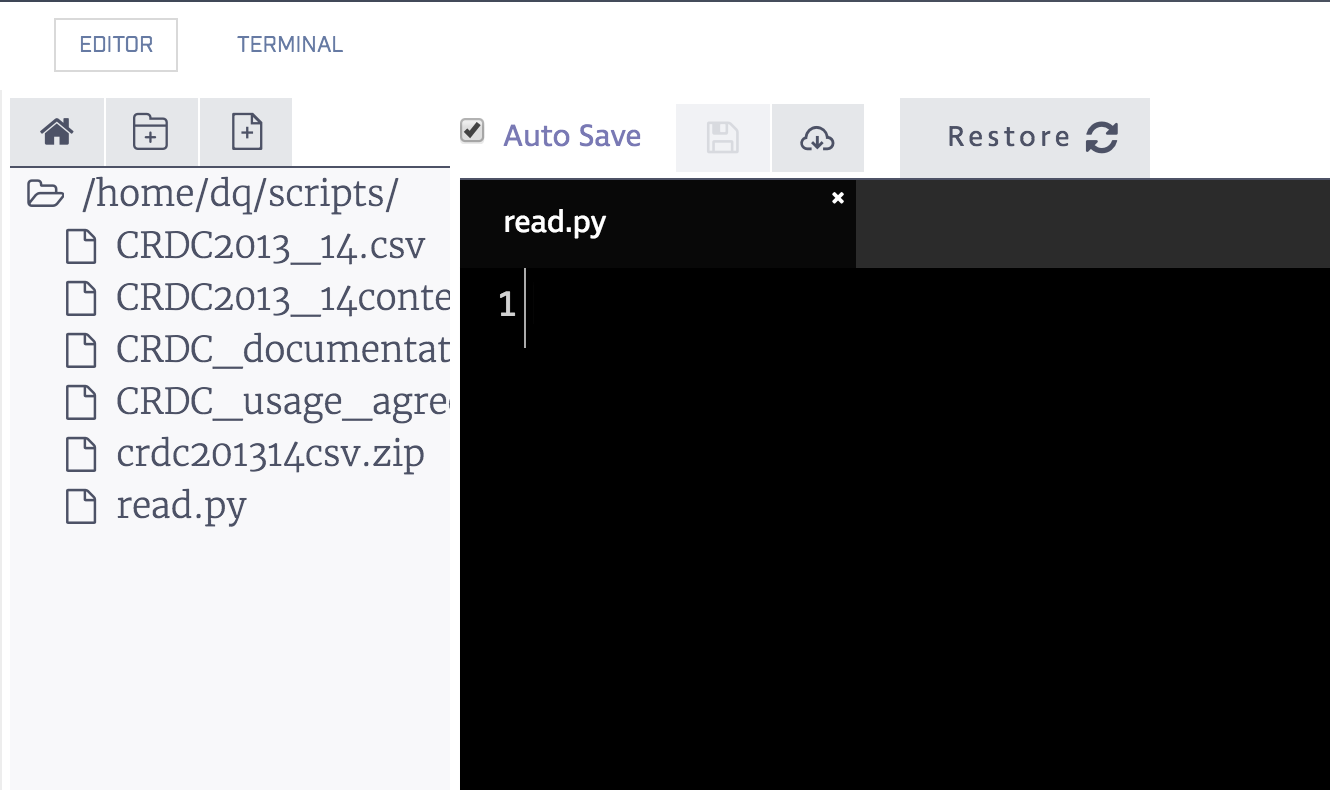
Almost all of the data you work with as a data scientist will come from a remote source, such as another website on the Internet. File downloads sometimes come in analysis-ready formats like CSV. At other times, the data will be in an archive format like TAR or ZIP. These formats compress files to reduce overall size, which makes them faster to download. Archive formats can also bundle multiple data files into a single archive file.

In this guided project, we'll be working with an education data set in ZIP format. We'll learn how to extract the files inside the ZIP download and then work with them.

You may have noticed that the interface for this guided project is a bit different than the one you've seen before. You'll notice that this interface has two views:

* **Editor**: you can create, edit, download, and restore files
* **Terminal**: you can run shell commands, like the ones we've explored in this course



Introduction to the dataset

The data set we'll work with is called the Civil Rights Data Collection. It contains information on educational achievement and opportunities in the U.S., broken down by race and school. For example, it records the racial composition of the students enrolled in advanced classes at each school. Each row represents a school, while each column records an indicator of academic achievement.

For the purposes of this exercise, we'll be using a subset of the data that only contains 1out of every 100 rows in the original data set. If you'd like to work with the original version, you can download it from [data.gov](https://catalog.data.gov/dataset/civil-rights-data-collection-2013-14).

Before we can load and analyze the data, we'll need to extract the files from the archive file, crdc201314csv.zip. We can run the [unzip](https://linux.die.net/man/1/unzip) command on an archive file to extract the files within it. Here's an example:



unzip test.zip

This command will extract all of the files from the archive test.zip into the current directory. Once we've extracted the files inside an archive file, it's good practice to delete the original archive to save space.

Instructions

* Switch to the terminal tab.
* List the contents of the current directory with the ls command, and take note of the archive file crdc201314csv.zip.
* Extract the files in crdc201314csv.zip using the unzip command.
* List the contents of the current directory, and make sure there are 4 new data files.
* Delete crdc201314csv.zip.