Instructions for Accessing InfoUSA data

* After you have signed the pledge of confidentiality, you will be given access to the remote volume that contains the InfoUSA data files.
* You will use the path below on Windows Explorer or Mac Finder. Your login will be win\<NetID> and NetID password on Windows, or <NetID> and NetID password on Mac (the domain, if it asks is [win.duke.edu](http://win.duke.edu)). You will need to use the VPN client if off campus.
	+ Windows: [\\oit-nas-fe11.oit.duke.edu\ssri-infousa](file://oit-nas-fe11.oit.duke.edu/ssri-infousa)
	+ Mac: smb://oit-nas-fe11.oit.duke.edu/ssri-infousa
* After connecting, you will see several folders. The two main folders are “documentation” and “derived set”. These contain documentation and the data files, respectively.
* The documentation folder includes a list of variables and their codings as well as a description of the InfoUSA ethnicity codes.
* The data files are divided into folders by year. Each year contains plain text files for each ZIP code. If you would like to merge years and/or ZIP codes, you must load and merge the files in the software of your choice. The .txt files are tab-delimited and can be opened in all statistical software programs.
	+ In R, the “read.delim()” function will import each individual text file.
	+ In Stata, the “import delimited” command will import each individual text file.
* Please note that, especially if accessing the data from off campus, the folders for each year may take a long time to load. The folders contain thousands of files and are quite large.