Task 1

Cheat Sheet of Best Practices for Reproducible Data Analysis

* Having a programming language with multiple capabilities such as data compression, integrated language,
* Having a good programming language that is widely accessible and easily used where collaborators can work together and users can share code
* Having an open source site for methods
* Built-in data integrity verification
* Unit tests to add code quality and functionality
* Have good self-documenting code
* Having an external remote so that you can have a backup of your work and to sync your remote/work to other computers and collaborate with others
* Having a continuous integration (CI) system that is able to perform data analysis/statistics on the data as you are working

Reinhart-Rogoff Affair

Reinhart and Rogoff were two Harvard economists who wrote a 2010 paper entitled, “Growth in a Time of Debt”, which claimed that the average yearly growth in countries with periods of gross national debt that are >90% of GDP around 1945-2009 was -0.1%. This number was calculated in an Excel spreadsheet, but Reinhart and Rogoff erringly excluded five rows from their calculations causing a gross miscalculation in their results. Open science is a way for any type of person to learn more about what others are doing or to help their own research by comparing methods and data analysis with other scientists. I think that it is necessary to pursue the ideal of open science because it allows us to pursue the idyllic image of unlimited knowledge. The “Reinhart-Rogoff Affair” may have never occurred if their information was not public knowledge. The methods of a study cannot be reproducible if they are not open to the public.