**Review of the National Student Loan Default**

Is there a correlation between the national student loan default rate (this study will use the 3 YR default rate; randomly selected) regionally and the percentage of Federal loans issued per region? The number of cohorts per region will also compare the average student loan debt upon separation.

**History**

The Department of Education issues The College Scorecard annually to help empower perspective students with tools to assist them in making informed decisions regarding their secondary education. One of the components in the scorecard is the financial status or stability of universities/colleges (AKA as cohorts) regarding federal student loans.

**Dataset Used**

The dataset used is: The College Scorecard (<https://collegescorecard.ed.gov/data/>. The report is an annual study gathering data from self-reporting cohorts within the United States and its territories. This study covers the 2017 – 2018 Academic Year. Extracted from the database: UNITID, INSTNM (changed to NAME), STABBR (changed to STATE), REGION, UGDS (changed to UNDERGRADS), PCTFLOAN, CDR3 (changed to RPYMT\_RATE\_YR3), DEBT\_MDN (changed to AVG\_LOAN\_DEBT).

**Glossary and Data Extracted**

**UNITID:** ID assigned to each Cohort

**NAME:** Name of the Cohort

**STATE:**  State cohort is located in

**REGION:** Defined from 1 – 9

**UNDERGRADS:** Total number of undergraduates per cohort

**CDR3:** Cohort Default Rate After 3 years

**PCTFLOAN:** Mean of federal loans issued (excluding Perkins and PLUS loans)

**RPYMT\_RATE\_YR3:** Loans entered into repayment 3 years after separation

**AVG\_LOAN\_DEBT:** Average student loan debt at separation

**PROCEDURE**:/**ANALYSIS**

* We identified the region with the largest cohorts to see if the size of the region would be a factor with the analysis.
* All data used is averaged (including NaN) for simplicity.
* Manipulation of data: some column names are changed, and some data was rounded to three decimal points.
* The analysis was performed in the order of three questions above.
* The results were plotted with Matplotlib