

# Springboard Intro to DS - Capstone Project

Bo Suzow

July 31, 2018

## Introduction

Student loan debt (SLD) total in the US reached a staggering number in 1.5 trillion dollars, borrowed by 44 million people. It is 2.4 times larger than the total of credit card debt. The SLD total in 2008 was 640 billion dollars [which ballooned to 1.2 trillion by 2015](#). While the importance and value of higher education (HED) can never be marginalized, student loans accumulated for financing higher education are reported as societal issues such as [reasons for divorce](#), financial dependence on parents, and lack of home ownership observed in Gen-Y.

In order for an student to be able to pay off HED loans, the loan total should not exceed his/her annual income from gainful employment the HED would provide. For unfortunate some, this rule of thumb was violated and loan defaults resulted. As the SLD total ever increases, what is the trend in its default rates? Are they going up or held at steady rates?

Seeking the data related this concern, I found datasets that the U.S. Department of Education (USD OE) released in September 2017 – [official default rates for 3 cohorts](#) (2012, 2013 and 2014).

Using one of the datasets, known as [Report 300](#), I plan on answering the following questions:

- Is the default trend upward, downward or steady?
- Any strong predictors of higher default rates?
- What are default rate distributions by various categorical variables (program length, school type, or region/state)

## Deliverables

- Shiny app for data exploration
- Default rate prediction (contingent upon extra dataset procurement)
- Presentation for the Shiny app details, storytelling

## Extra

- Search for a dataset of HED institution profiles and integrate additional independent variables into the default dataset.