

# Student Loan Defaults

Bo Suzow

July 30, 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 (<https://www.politifact.com/truth-o-meter/statements/2015/aug/14/jeb-bush/jeb-bush-student-loan-debt-has-doubled-under-obama/>). As the debt ever increases, what is the trend in its default rates? Are they going up or held at steady rates?

The U.S. Department of Education (USDOE) releases official cohort default rates for 3 cohorts (<https://www2.ed.gov/offices/OSFAP/defaultmanagement/cdr.html#table>) (2012, 2013 and 2014) in September 2017. This report analyzes the default data to identify any patterns and/or trend nationwide and by state.

## Data Load

The dataset is publicly available here (<https://www2.ed.gov/offices/OSFAP/defaultmanagement/peps300.xlsx>). The details on the data dictionary (explanation of each column/field) is found here (<https://www2.ed.gov/offices/OSFAP/defaultmanagement/instructions.html>). While USDOE released 5 Default Reports, this analysis focuses on Report 300 which reports "FY 2014, FY 2013, and FY 2012 official cohort default rates published for schools participating in the Title IV student financial assistance programs."

```
defaults=read_excel("peps300.xlsx")
defaults = tbl_df(defaults)
#defaults
```

The dataset features 4712 schools participating in the Title IV programs. 29 variables/columns are provided in the dataset. Let's rename them for easiness of the processing ensued.

```
names(defaults)=c("id","name","addr","city","state","state_name","zip","zipext","prog_len","school_type",
                  "yr1","num1","denum1","drate1","prate1",
                  "eth_code","program","congress_dist","region",
                  "yr2","num2","denum2","drate2","prate2",
                  "yr3","num3","denum3","drate3","prate3")
```

Are there any columns with missing values?

```
names(defaults)[colSums(is.na(defaults))>0]
```

```
## [1] "addr"      "zip"      "zipext"   "congress_dist"
```

No missing value strategy is needed as these columns will not be used in the analysis.

## School Counts by Categorical Variables/Columns

There are a few categorical (aka factor) variables in the dataset – State (state), Program Length (prog\_len), School Type (school\_type), Program, Ethnic Code (eth\_code). Let's convert them to factor variables (categorical variables are called factor variables in R) with proper labels (descriptions of the codes).

```

#-----
# State
#-----

defaults <- defaults %>% mutate(state = factor(state))

#-----
# Program Length
#-----

prog_len_list = c(0:12)
prog_len_descs = c("Short Term (300-599 hours)",
  "Graduate/Professional (>=300 hours)",
  "Non-Degree (600-899 hours)",
  "Non-Degree 1 Year (900 - 1799 hours)",
  "Non-Degree 2 Year (1800-2699 hours)",
  "Associate's Degree", "Bachelor's Degree",
  "First Professional Degree",
  "Master's Degree or Doctor's Degree",
  "Professional Certification",
  "Undergraduate (Previous Degree Required",
  "Non-Degree 3 Plus Years (>= 2700 hours)",
  "Two-Year Transfer")
defaults <- defaults %>% mutate(prog_len = factor(prog_len, levels=prog_len_list,
  labels=prog_len_descs))

#-----
# School Type
#-----

school_type_list = c(1:3, 5:7)
school_type_descs = c("Public",
  "Private, Nonprofit",
  "Proprietary",
  "Foreign public",
  "Foreign private",
  "Foreign For-Profit")
defaults <- defaults %>% mutate(school_type = factor(school_type, levels=school_type_list,
  labels=school_type_descs))

#-----
# Region
#-----

region_list = c(1:11)
region_descs = c("01", "02", "03", "04", "05", "06", "07", "08", "09", "10", "11")

defaults <- defaults %>% mutate(region = factor(region, levels=region_list,
  labels=region_descs))

sum_state <- defaults %>% group_by(state) %>% summarize(n())
sum_prog_len <- defaults %>% group_by(prog_len) %>% summarize(n())
sum_school_type <- defaults %>% group_by(school_type) %>% summarize(n())
sum_program <- defaults %>% group_by(program) %>% summarize(n())
sum_eth_code <- defaults %>% group_by(eth_code) %>% summarize(n())

```

## Data Wrangling

Data Exploration

Data Plots