

Peck - Final Capstone Project

Jennifer Peck

2023-01-07

Introduction

For my final project in the Data Science certification course, I decided to use a data set that was pertaining to student loans. Not only is this pertinent to me as a student myself, but it is also the industry I have worked in for the past 20 years. On Kaggle, I was able to find a data set that is a scrape of COD (Common Origination and Disbursement) for the 2020-2021 academic school year. COD is a website used by the Department of Education, student loan servicers, schools, and students to track loan information.

This data set is focused to school level items. There is information about what type of school each is, where it is located, the number of recipients of loans, as well as the number of loans and dollar amounts for both originations and disbursements. For this project, I am going to focus on originations as that is the picture of the total loan debt taken out. I am going to view and present the data by modeling different methods of query and graphing, and determine if I locate any trends. This trending data may be on type of school, state location, or loan type. I will present the results in multiple formats, utilizing both tables and graphs. Let's take a look!

Methods and Analysis

Set-Up

To get started, I need to download the data set into R, and get it into a format that will make it easier to use going forward. I also want to make sure I have all the various library packages installed and loaded. The file on Kaggle is called "The schools that create the most student debt", which I am renaming to simply "debt".

I want to take a quick look at what we are starting with. First, I determine the class type of my file is a data frame, and I look at an example of the structure. I see that there are 42765 objects and 13 variables.

```
## [1] "data.frame"
```

```
## 'data.frame':   42765 obs. of  13 variables:
## $ index          : int  0 1 2 3 4 5 6 7 8 9 ...
## $ OPE.ID         : chr  "001051" "001052" "001009" "001057" ...
## $ School         : chr  "UNIVERSITY OF ALABAMA" "UNIVERSITY OF ALABAMA AT BIRMINGHAM" "AUBURN" ...
## $ City           : chr  "TUSCALOOSA" "BIRMINGHAM" "AUBURN" "MOBILE" ...
## $ State          : chr  "AL" "AL" "AL" "AL" ...
## $ Zip.Code       : num  35487 35294 36849 36688 36082 ...
## $ School.Type    : chr  "Public" "Public" "Public" "Public" ...
## $ Loan.Type      : chr  "Subsidized" "Subsidized" "Subsidized" "Subsidized" ...
## $ Recipients     : num  7594 4272 4504 3438 4334 ...
## $ X.of.Loans.Originated : num  7622 4352 4504 3487 4355 ...
```

```
## $ X..of.Loans.Originated.1: num 33040862 18590061 19474542 15061071 18467508 ...
## $ X..of.Disbursements : num 7622 4352 4535 3487 4355 ...
## $ X..of.Disbursements.1 : num 16774054 9058344 10142490 7557853 7395839 ...
```

I decided to rename all of the columns as some of them have special characters that I don't want to deal with.

```
#The original csv has special characters in some of the column names so I want
#to rename the columns for ease going forward
names(debt)[1] <- 'index'
names(debt)[2] <- 'opeid'
names(debt)[3] <- 'school'
names(debt)[4] <- 'city'
names(debt)[5] <- 'state'
names(debt)[6] <- 'zip'
names(debt)[7] <- 'school_type'
names(debt)[8] <- 'loan_type'
names(debt)[9] <- 'num_recipients'
names(debt)[10] <- 'num_loans_originated'
names(debt)[11] <- 'dollars_originated'
names(debt)[12] <- 'num_disbursements'
names(debt)[13] <- 'dollars_disbursed'

#check to make sure the columns are renamed
colnames(debt)
```

```
## [1] "index" "opeid" "school"
## [4] "city" "state" "zip"
## [7] "school_type" "loan_type" "num_recipients"
## [10] "num_loans_originated" "dollars_originated" "num_disbursements"
## [13] "dollars_disbursed"
```

I want to view a tibble of this data frame to see if it is getting into shape. It's starting to look pretty good!

```
## # A tibble: 42,765 x 13
## index opeid school city state zip schoo~1 loan_~2 num_r~3 num_l~4 dolla~5
## <int> <chr> <chr> <chr> <chr> <dbl> <chr> <chr> <dbl> <dbl> <dbl>
## 1 0 001051 UNIVE~ TUSC~ AL 35487 Public Subsid~ 7594 7622 3.30e7
## 2 1 001052 UNIVE~ BIRM~ AL 35294 Public Subsid~ 4272 4352 1.86e7
## 3 2 001009 AUBUR~ AUBU~ AL 36849 Public Subsid~ 4504 4504 1.95e7
## 4 3 001057 UNIVE~ MOBI~ AL 36688 Public Subsid~ 3438 3487 1.51e7
## 5 4 001047 TROY ~ TROY AL 36082 Public Subsid~ 4334 4355 1.85e7
## 6 5 001036 SAMFO~ BIRM~ AL 35229 Privat~ Subsid~ 638 638 2.73e6
## 7 6 001024 UNIVE~ LIVI~ AL 35470 Public Subsid~ 1085 1098 4.55e6
## 8 7 042267 ALABA~ DOTH~ AL 36303 Privat~ Subsid~ 0 0 0
## 9 8 001020 JACKS~ JACK~ AL 36265 Public Subsid~ 2894 2895 1.21e7
## 10 9 001050 TUSKE~ TUSK~ AL 36088 Privat~ Subsid~ 1104 1107 4.96e6
## # ... with 42,755 more rows, 2 more variables: num_disbursements <dbl>,
## # dollars_disbursed <dbl>, and abbreviated variable names 1: school_type,
## # 2: loan_type, 3: num_recipients, 4: num_loans_originated,
## # 5: dollars_originated
```

Now I want to start getting an idea of some of my populations. Based on this summary, I have 4078 schools, 7 loan types, and 8 school types.

Table 1: Summary

distinct_schools	distinct_loantype	distinct_schooltype
4078	7	8

Let's check out the loan types:

Table 2: Loan Types and #Rows

loan_type	count
Total	21370
Grad Plus	4274
Parent Plus	4274
Subsidized	4274
Unsubsidized - Graduate	4274
Unsubsidized - Undergraduate	4274
	25

This brings up a few interesting points. I see there are 25 rows that have no loan type, the count of the labeled loan types are all equal, and there is a huge number under 'total'. I will deal with the blanks in a bit and I assume the number of each loan type is even because there is a line item for each type, whether the school reported any info there or not. First though, I am curious about what the 'totals' are. I am going to narrow down to looking at only the info for one school to try and figure it out.

```
debt %>% filter(school == 'UNIVERSITY OF ALABAMA')
```

```
##      index opeid      school      city state  zip school_type
## 1         0 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 2      4274 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 3      8548 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 4     12822 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 5     17096 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 6     21370 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 7     25644 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 8     29918 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 9     34192 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
## 10    38466 001051 UNIVERSITY OF ALABAMA TUSCALOOSA    AL 35487    Public
##              loan_type num_recipients num_loans_originated
## 1              Subsidized              7594              7622
## 2 Unsubsidized - Undergraduate              9192              9474
## 3      Unsubsidized - Graduate              1795              1808
## 4              Parent Plus              2774              2828
## 5              Grad Plus              248              252
## 6              Total              21603              21984
## 7              Total              35612              36346
## 8              Total              62032              63218
## 9              Total              122269              124628
## 10             Total              241764              246428
```

```
##      dollars_originated num_disbursements dollars_disbursed
## 1          33040862           7622          16774054
## 2          36307779           9474          18396030
## 3          29936391           1808          15127471
## 4          80141304           2828          41174095
## 5           3995419            252           2094711
## 6          183421755          21984          93566361
## 7          333802648          36346          170358668
## 8          631297517          63218          322321306
## 9          1232658643          124628          629515141
## 10         2385175982          246428          1217856187
```

It appears that one of the total lines is the sum of the values of the other loan types added together. I cannot determine what the other 4 total lines are showing. To have it not skew my results going forward, I am going to throw out all of the rows in the set with a loan type of 'total'. I am also going to throw out the 25 loans with no loan type. I will call this new result 'clean_debt'.

```
clean_debt <- debt %>% filter(loan_type != 'Total')
clean_debt <- clean_debt %>% filter(loan_type != "")

#make sure that I no longer show either totals or blanks
clean_debt %>% group_by(loan_type) %>% summarize(count = n())
```

```
## # A tibble: 5 x 2
##   loan_type          count
##   <chr>          <int>
## 1 Grad Plus          4274
## 2 Parent Plus        4274
## 3 Subsidized          4274
## 4 Unsubsidized - Graduate 4274
## 5 Unsubsidized - Undergraduate 4274
```

```
#check to see how far row count went down
nrow(clean_debt)
```

```
## [1] 21370
```

When I ran the first tibble of this data, it jumped out at me that one of the lines showing had 0's reporting for a row - indicating that they reported their school's information to COD, but didn't happen to have any loans of a particular type. Even though 0's won't add to my results, I am going to toss those lines as well, just to make our data set as clean as possible.

Finally, in terms of set up, I discovered later in my research that there are schools listed in the data that are not located in America. Because I intend to look at trending by state, I want to remove these lines as well. An example of this is shown with this school in Ireland.

```
##   index opeid          school city state zip
## 1  2069 007705 ROYAL COLLEGE OF SURGEONS IN IRELAND    NA
## 2  6343 007705 ROYAL COLLEGE OF SURGEONS IN IRELAND    NA
## 3 10617 007705 ROYAL COLLEGE OF SURGEONS IN IRELAND    NA
## 4 14891 007705 ROYAL COLLEGE OF SURGEONS IN IRELAND    NA
## 5 19165 007705 ROYAL COLLEGE OF SURGEONS IN IRELAND    NA
```

```
##      school_type      loan_type num_recipients
## 1 Foreign-Private      Subsidized           5
## 2 Foreign-Private Unsubsidized - Undergraduate           5
## 3 Foreign-Private      Unsubsidized - Graduate          79
## 4 Foreign-Private      Parent Plus           5
## 5 Foreign-Private      Grad Plus           72
##  num_loans_originated dollars_originated num_disbursements dollars_disbursed
## 1                5          20500           5          10250
## 2                5          10000           5           5000
## 3               79         1619500          79         820000
## 4                7          363000           7         181500
## 5               73         4876008          75        2456679
```

```
#how many rows left
nrow(clean_debt)
```

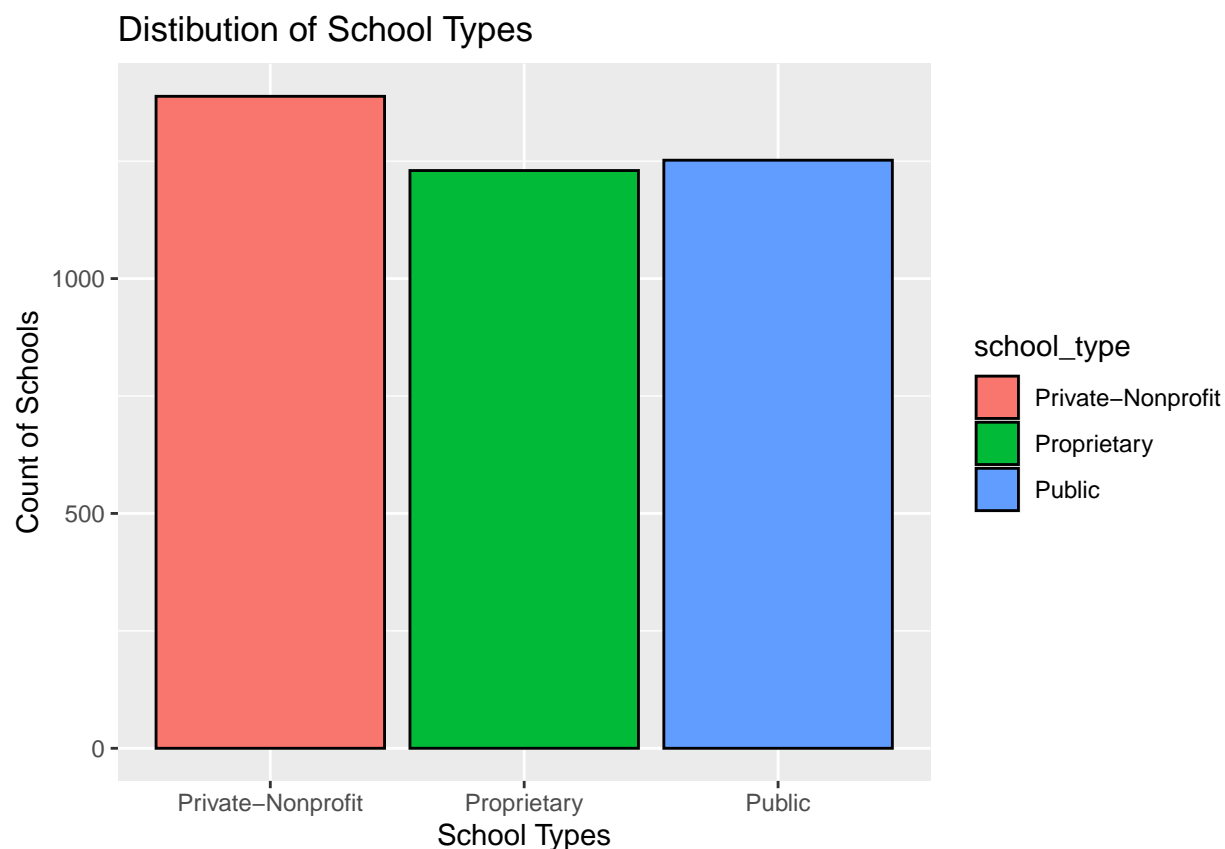
```
## [1] 13283
```

This gets the cleaned up data down to 13283 lines. The data is now in good shape and cleaned up of items that will give skewed results.

Data Modeling

Ok, I am ready to start looking for trends! I will split this section by looking for trends from four angles - school type trends, loan type trends, state trends, and school trends.

School Type The first thing I want to check out is how many schools there are for each school type. I will view this via bar graph.



The graph above shows 3 school types: Private-Nonprofit, Proprietary, and Public. Private-Nonprofit has the most schools reporting. If you recall, before we cleaned up the data, we originally had 8 school types showing. The other 5 that are no longer here were indicating either foreign schools, blanks, or invalid information (info in wrong column).

Next, I am going to see how many loan recipients there are at each school type. The results show that there are close to double the amount of recipients at public schools vs private schools, even though there are more private schools on the list, as we saw above.

Table 3: Recipients at School Types

school_type	number_recipients
Public	4204834
Private-Nonprofit	2284681
Proprietary	485428

For the last portion of my review by loan type, I want to look at the amount of dollars originated, by each. Here we see that the most money is in the realm of public schools, but it's a lot closer to the private school amount than the number of recipients would lead us to believe.

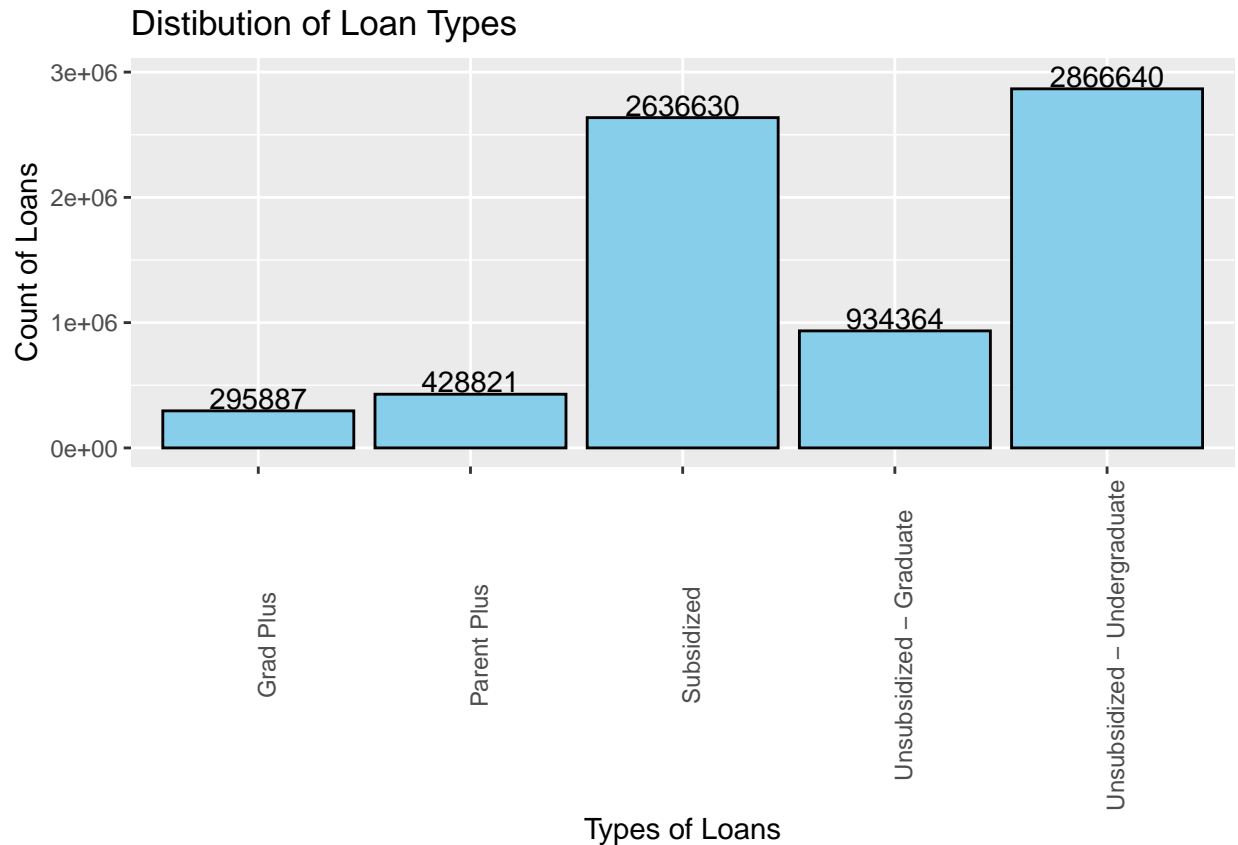
Table 4: Total Orig by School Type

school_type	dollars_orig
Public	28494531594

school_type	dollars_orig
Private-Nonprofit	22219073998
Proprietary	3139216228

Loan Type The second angle I want to look into is trends by loan type.

In the graph below, we see that the amounts of Subsidized and Unsubsidized - Undergraduate loans are fairly close. This is what I would expect to see as it is common for students to get both a subsidized and unsubsidized loan at the same time.



I want to see how many loan recipients there are by each loan type. If we compare this to the graph above, the counts are very close but there are more loans by type than there are recipients. This makes sense as some students get more than one loan. What we can't tell here is distinct recipient. If one student got both a subsidized and unsubsidized loan, they will show up as a recipient for each column.

Table 5: Recipients and Loan Types

loan_type	number_recipients
Unsubsidized - Undergraduate	2769772
Subsidized	2605973
Unsubsidized - Graduate	899011
Parent Plus	420037
Grad Plus	280150

Next, let's check out the amount of money originated by each loan type. Unsubsidized - Graduate is the highest dollar amount. This is likely due to the cost of Grad school being higher than Undergrad.

Table 6: Orig by Loan Type

loan_type	dollars_orig
Unsubsidized - Graduate	17264551867
Unsubsidized - Undergraduate	11576673233
Subsidized	10658152362
Parent Plus	7463173728
Grad Plus	6890270630

For the final look at loan types, I want to determine what the average dollar amount originated by recipient is. The results show that on average, recipients that take out Grad PLUS loans have the highest average loan amount. Subsidized is the lowest.

Table 7: Average Orig by Person and Loan Type

loan_type	avgorig
Grad Plus	24594.934
Unsubsidized - Graduate	19203.938
Parent Plus	17767.896
Unsubsidized - Undergraduate	4179.648
Subsidized	4089.894

State Information The third angle that I would like to investigate is at the state level. I am curious if there are trends of how many loans or amount of money that is originated by state.

First up, I'd like to find out which 10 states have the most schools reporting in this data set. California has over 1300 schools, while the next closest state is New York with 859.

Table 8: Top 10 Schools per State

state	n
CA	1303
NY	859
TX	832
PA	766
FL	590
OH	561
IL	556
MO	369
MI	359
MA	349

I queried how many schools, per school type, are being reported in each state. This is a full look, so the table is long.

Table 9: Schools, by Type, per State

state	school_type	n
AK	Private-Nonprofit	7
AK	Proprietary	2
AK	Public	15
AL	Private-Nonprofit	63
AL	Proprietary	36
AL	Public	99
AR	Private-Nonprofit	51
AR	Proprietary	49
AR	Public	92
AZ	Private-Nonprofit	16
AZ	Proprietary	106
AZ	Public	56
CA	Private-Nonprofit	438
CA	Proprietary	509
CA	Public	356
CO	Private-Nonprofit	33
CO	Proprietary	89
CO	Public	110
CT	Private-Nonprofit	67
CT	Proprietary	55
CT	Public	28
DC	Private-Nonprofit	42
DC	Proprietary	9
DC	Public	5
DE	Private-Nonprofit	14
DE	Proprietary	15
DE	Public	10
FL	Private-Nonprofit	175
FL	Proprietary	285
FL	Public	130
GA	Private-Nonprofit	124
GA	Proprietary	75
GA	Public	130
GU	Private-Nonprofit	2
HI	Private-Nonprofit	16
HI	Proprietary	11
HI	Public	31
IA	Private-Nonprofit	135
IA	Proprietary	42
IA	Public	63
ID	Private-Nonprofit	14
ID	Proprietary	35
ID	Public	29
IL	Private-Nonprofit	260
IL	Proprietary	160
IL	Public	136
IN	Private-Nonprofit	146
IN	Proprietary	70
IN	Public	71
KS	Private-Nonprofit	82

state	school_type	n
KS	Proprietary	28
KS	Public	98
KY	Private-Nonprofit	86
KY	Proprietary	39
KY	Public	76
LA	Private-Nonprofit	42
LA	Proprietary	89
LA	Public	109
MA	Private-Nonprofit	263
MA	Proprietary	40
MA	Public	46
MD	Private-Nonprofit	56
MD	Proprietary	49
MD	Public	93
ME	Private-Nonprofit	50
ME	Proprietary	14
ME	Public	41
MI	Private-Nonprofit	127
MI	Proprietary	99
MI	Public	133
MN	Private-Nonprofit	111
MN	Proprietary	34
MN	Public	134
MO	Private-Nonprofit	166
MO	Proprietary	76
MO	Public	127
MS	Private-Nonprofit	32
MS	Proprietary	19
MS	Public	57
MT	Private-Nonprofit	16
MT	Proprietary	11
MT	Public	41
NC	Private-Nonprofit	169
NC	Proprietary	46
NC	Public	116
ND	Private-Nonprofit	13
ND	Proprietary	9
ND	Public	40
NE	Private-Nonprofit	62
NE	Proprietary	22
NE	Public	47
NH	Private-Nonprofit	41
NH	Proprietary	18
NH	Public	22
NJ	Private-Nonprofit	73
NJ	Proprietary	96
NJ	Public	47
NM	Private-Nonprofit	11
NM	Proprietary	24
NM	Public	49
NV	Private-Nonprofit	11
NV	Proprietary	25

state	school_type	n
NV	Public	23
NY	Private-Nonprofit	443
NY	Proprietary	183
NY	Public	233
OH	Private-Nonprofit	260
OH	Proprietary	124
OH	Public	177
OK	Private-Nonprofit	44
OK	Proprietary	40
OK	Public	96
OR	Private-Nonprofit	71
OR	Proprietary	50
OR	Public	54
PA	Private-Nonprofit	446
PA	Proprietary	167
PA	Public	153
PR	Private-Nonprofit	99
PR	Proprietary	17
PR	Public	7
RI	Private-Nonprofit	42
RI	Proprietary	6
RI	Public	10
SC	Private-Nonprofit	91
SC	Proprietary	54
SC	Public	103
SD	Private-Nonprofit	24
SD	Proprietary	12
SD	Public	40
TN	Private-Nonprofit	145
TN	Proprietary	78
TN	Public	72
TX	Private-Nonprofit	221
TX	Proprietary	263
TX	Public	348
UT	Private-Nonprofit	24
UT	Proprietary	65
UT	Public	35
VA	Private-Nonprofit	139
VA	Proprietary	84
VA	Public	79
VI	Public	5
VT	Private-Nonprofit	39
VT	Proprietary	4
VT	Public	19
WA	Private-Nonprofit	86
WA	Proprietary	61
WA	Public	100
WI	Private-Nonprofit	104
WI	Proprietary	41
WI	Public	97
WV	Private-Nonprofit	37
WV	Proprietary	34

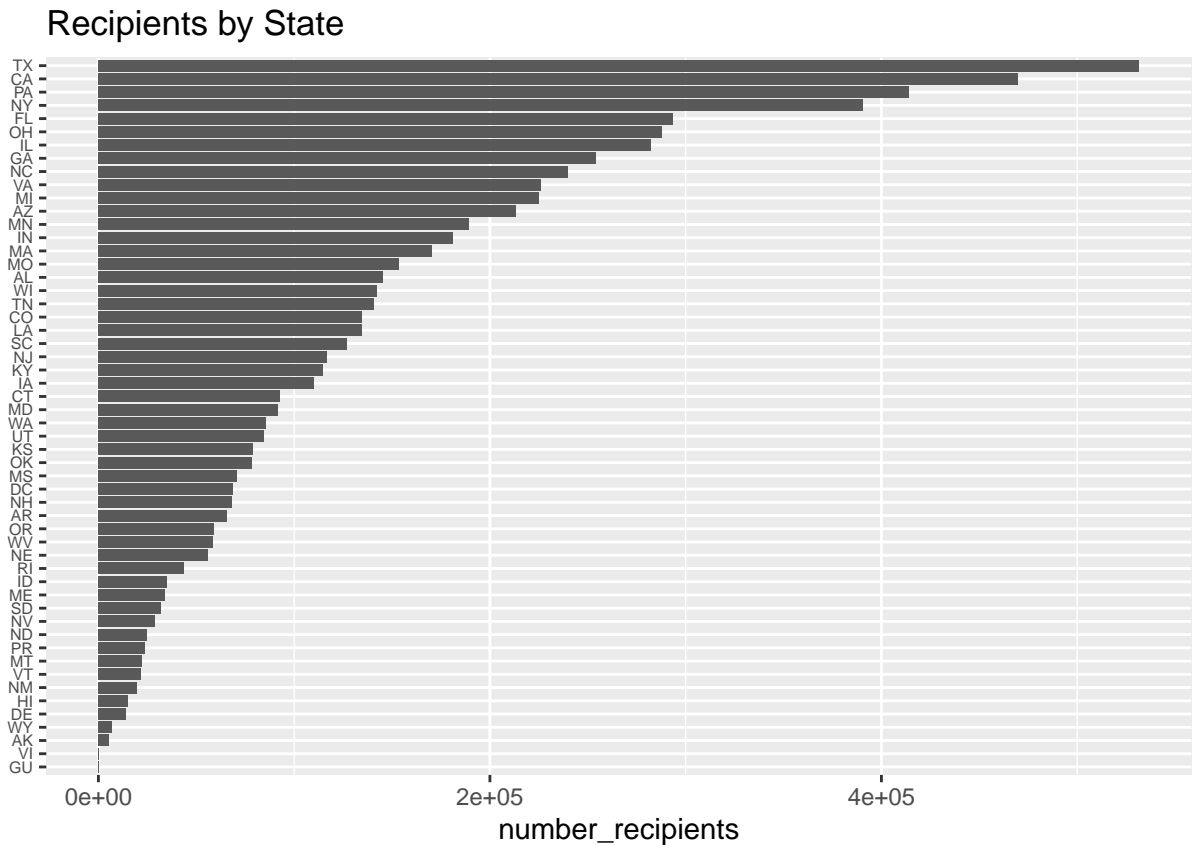
state	school_type	n
WV	Public	67
WY	Proprietary	5
WY	Public	25

Now I want to look at the 10 states that have the highest number of loan recipients Texas is the highest with almost 532 thousand.

Table 10: Top 10 Recipients by State

state	number_recipients
TX	531764
CA	469773
PA	414193
NY	390524
FL	293675
OH	287646
IL	282287
GA	254150
NC	239807
VA	225933

Here is a larger overview of recipients by state in a descending graph view.

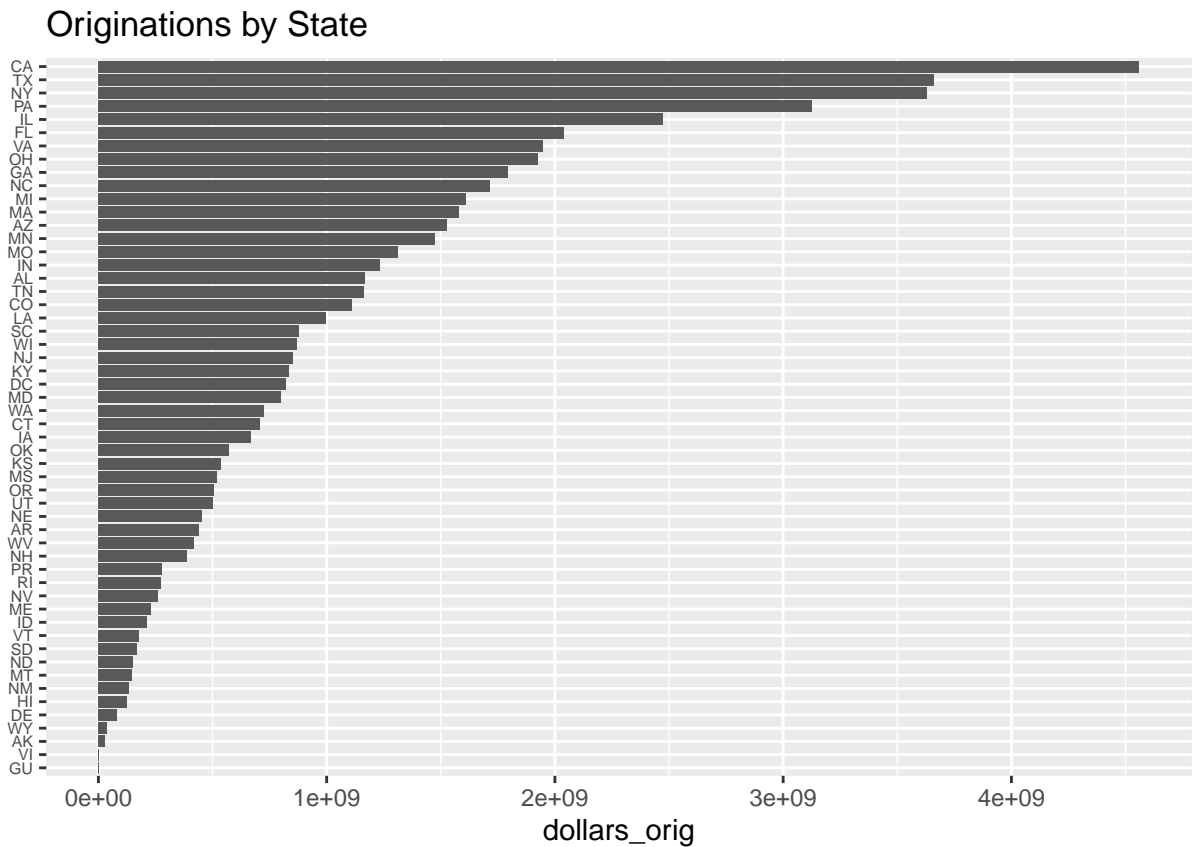


Next, I want to view the dollars originated by state. Here are the top 10 states with the highest originated amount. California lands in the top spot.

Table 11: Dollars Orig per State

state	dollars_orig
CA	4560513271
TX	3659070238
NY	3628169842
PA	3124201398
IL	2475197981
FL	2037438833
VA	1945669906
OH	1923909608
GA	1792658962
NC	1716090531

Similar to before, I want to look at a descending graph by state, this time with a full view of total originated amounts for all states.

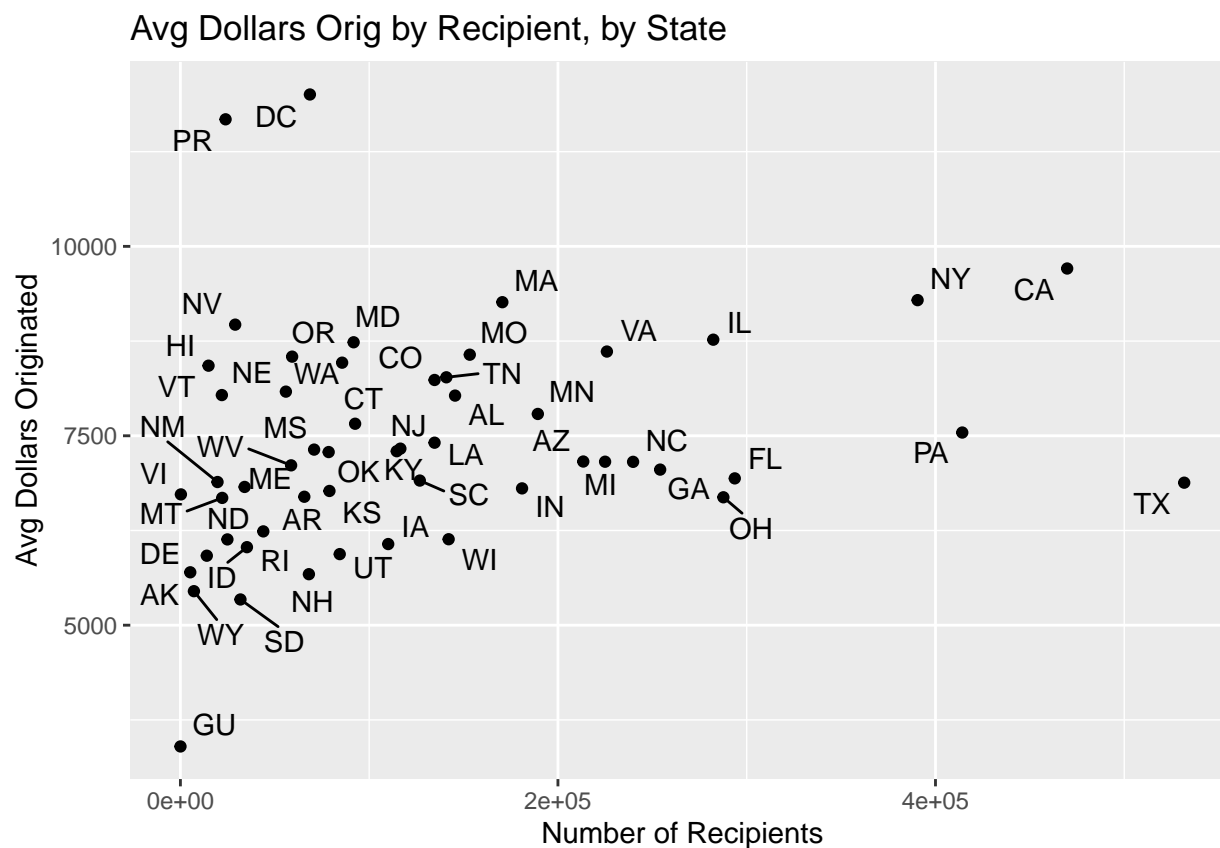


I want step away from totals for a bit and look at the average amount originated by recipient, by state. Interestingly, Washington DC and Puerto Rico borrow the highest amount per person.

Table 12: Average Orig, by Recipient, by State

state	avgorig
DC	12004.630
PR	11676.997
CA	9707.908
NY	9290.517
MA	9263.365
NV	8967.005
IL	8768.374
MD	8734.330
VA	8611.712
MO	8570.357

I decided to graph out this data in a different way. This makes it a little easier to see that PR and DC are outliers at the high end of dollars per person, Guam is an outlier on the low end of the same thing. TX, CA, PA, and NY are outliers in terms of the amount of people who got loans, but their average dollar amount falls in line with most other states.



School Information In the last section of data analysis, I am going to look at some trends by individual schools. If we look at the schools with the most money originated, Arizona State University landed on top. This is interesting because Arizona hasn't popped up as being at the top of any of our other queries.

Table 13: Top 10 Most Orig by School

school	dollars_orig
ARIZONA STATE UNIVERSITY	567643466
LIBERTY UNIVERSITY	446670128
NEW YORK UNIVERSITY	405470421
PENNSYLVANIA STATE UNIVERSITY (THE)	387508772
CAPELLA UNIVERSITY	332359466
RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY	308751556
GRAND CANYON UNIVERSITY	305945501
TEMPLE UNIVERSITY	263246236
UNIVERSITY OF SOUTHERN CALIFORNIA	242900907
NOVA SOUTHEASTERN UNIVERSITY-DAVIE	232295695

I also want to see which schools have the highest average amount originated by recipient. An interesting discovery here is that the top 10 are all either medical schools or schools best known for their medical programs.

Table 14: Schools with Highest Avg Orig by Recipient

school	avgorig
CALIFORNIA UNIVERSITY OF SCIENCE AND MEDICINE	44944.00
MIDWESTERN UNIVERSITY	41914.61
ALABAMA COLLEGE OF OSTEOPATHIC MEDICINE	40292.14
ROCKY VISTA UNIVERSITY	40076.30
HACKENSACK MERIDIAN SCHOOL OF MEDICINE	39051.28
A. T. STILL UNIVERSITY OF HEALTH SCIENCES	38659.43
MEHARRY MEDICAL COLLEGE	38510.64
WESTERN UNIVERSITY OF HEALTH SCIENCES	37708.31
WESTERN MICHIGAN UNIVERSITY HOMER STRYKER M.D. SCHOOL OF MEDICINE	37369.35
PACIFIC NORTHWEST UNIVERSITY OF HEALTH SCIENCES	37180.73

Results

Looking at this student loan debt by school information in multiple ways has shown a number of interesting results.

- Far more people obtain the assistance of student loans if they attend public schools. This may indicate a couple of things - lower income people are more likely to attend public school, or more people in general attend public school. We can't determine this without data regarding total number of attendees.
- Even though there was close to double the amount of recipients at public vs private schools, the dollar amount originated was fairly close. This tells us that the private school students need to borrow more money, which makes sense as private schools tend to have far more expensive tuition.
- Looking at the data by loan type doesn't help to draw very many conclusions, but there was one interesting result - graduate students need to borrow far more per person than anybody else. Unsubsidized-Graduate and Grad PLUS loans average out to \$20-25,000 per recipient, while Unsubsidized-Undergrad and Subsidized loans come out to roughly \$4100 per recipient.

- Trending by state showed that the most schools are in CA, followed by NY and TX, however the state with the most students getting a loan is TX, followed by CA. NY was 4th on that list. 7 of the top 10 states by number of recipients are also on the list of top 10 states with the most amount of schools. The top 10 states by recipients match the top 10 states by dollars originated. This likely tells us that these states aren't outliers, but simply at the top due to volume.
- When the state data was graphed by average originated dollars per recipient, it was easier to find some outliers. Puerto Rico and Washington DC have a higher proportion of money originated per person than the rest of the group. Guam was at the other end of the spectrum, with very few recipients and originated loans.
- Arizona State University was the school that originated the most amount of debt. This is interesting as Arizona as a state didn't come up anywhere else while we were looking at trends. It isn't in the top 10 for overall originations, number of recipients or anything else, yet the school on its own originated over a half billion dollars in one academic year. My assumption is if we had numbers of total school attendees, we would find that a higher proportion of students who attend ASU borrow money than other colleges.
- The final query regarding which schools have the highest average of dollars originated by recipient was interesting, though not entirely surprising. All of the top 10 schools were either medical schools, or colleges that were best known for their medical programs.

Conclusion

Based on this data set, if you are a student obtaining a graduate degree or a medical degree, and need to borrow money in order to do it, you are going to be taking on a large debt. Undergraduate degrees and public schools, are likely to cause less of a financial strain, though they may add up if you obtain multiple loans. Since this data was a snapshot of one academic year, we are unable to see the average amount borrowed by recipient over the length of their college degree. The academic year for this data was also 2020-2021 so COVID would play a part in decreasing how many people went to school and needed to obtain loans. I would expect that if we looked at this same data from 2018-2019, the numbers would be even larger.

References

The schools that create the most student debt Data set - Kaggle <https://www.kaggle.com/datasets/thedevastator/the-schools-that-create-the-most-student-debt?resource=download>

Introduction to Data Science: Data Analysis and Prediction Algorithms with R Rafael A. Irizarry <http://rafalab.dfci.harvard.edu/dsbook/>

R Markdown Cookbook Yihui Xie, Christophe Dervieux, Emily Riederer <https://bookdown.org/yihui/rmarkdown-cookbook/>

Github data visualization cheat sheet <https://github.com/rstudio/cheatsheets/blob/main/data-visualization.pdf>

Github R markdown cheat sheet <https://github.com/rstudio/cheatsheets/blob/main/rmarkdown.pdf>

My github for this project https://github.com/jenpeck/Capstone_Final_Project