

Dollars and Diplomas: Project Report

1. Project Data

For this project, three different datasets were accessed, each from a separate source:

- [High school graduation rates by state \(2012-2013\) - Governing Data](#)
- [Public school system finances \(2012\) - Census Bureau](#)
- [Average starting teacher salaries by state \(2012-2013\) - National Education Association](#)

The data from the Census Bureau was available as a downloadable Excel file. However, the data from Governing Data and the National Education Association were in the form of HTML tables.

GOVERNING

THE STATES AND LOCALITIES

FINANCE | HEALTH | INFRASTRUCTURE | MANAGEMENT | ELECTIONS | POLITICS | PUBLIC SAFETY | URBAN | EDUCATION



State graduation rates for the three most recent school years are shown below. Please note that data is unavailable for a few states ([see notes below](#)).



State High School Graduation Rates for All Students



State	2012-13 Graduation Rate	2011-12 Graduation Rate	2010-11 Graduation Rate
Kentucky	86.1%	-	-
Louisiana	73.5%	72%	71%
Maine	86.4%	85%	84%
Maryland	85.0%	84%	83%
Massachusetts	85.0%	85%	83%
Michigan	77.0%	76%	74%
Minnesota	79.8%	78%	77%

SOURCE: U.S. Department of Education, EDData/Consolidated State Performance Reports

State High School Graduation Rates for Low Income Students

State	2012-13 Graduation	2011-12 Graduation	2010-11 Graduation
-------	--------------------	--------------------	--------------------

High School Graduation Rates by State. (n.d.). Retrieved November 28, 2016, from <http://www.governing.com/gov-data/high-school-graduation-rates-by-state.html>

The screenshot shows the NEA website's page for '2012-2013 Average Starting Teacher Salaries by State'. The page includes a navigation bar with 'ISSUES AND ACTION', 'TOOLS AND IDEAS', and 'GRANTS'. A sidebar on the left lists various topics like 'ESSA Implementation Begins', 'College Affordability and Debt Services', and 'Human & Civil Rights'. The main content area features a title, social sharing buttons (Like 1.2K, Share, Tweet, Print, 360), a 'HIGHLIGHTS' section stating the national average starting teacher salary is \$36,141, and a table of average starting salaries by state.

State	Avg. Starting Salary
Alabama	\$36,198
Alaska	\$44,166
Arkansas	\$32,691
Arizona	\$31,874
California*	\$41,259
Colorado	\$32,126
Connecticut	\$42,924
District of Columbia*	\$51,539

2012-2013 average starting teacher salaries by state. (2013, December).

Retrieved November 29, 2016, from <http://www.nea.org/home/2012-2013-average-starting-teacher-salary.html>

In order to work with all datasets as CSV files, I used an [online conversion tool](#), which was simple and effective. The graduation rate data was broken down into three separate CSVs—all students; students from low-income households; and students receiving special education services. Therefore, there were five CSV files in total to work with.

Three identically-formatted CSVs for graduation rates:

```
mroy@holden ~/assign/project $ head gradrates.csv ]
State,2012-13 Graduation Rate,2011-12 Graduation Rate,2010-11 Graduation Rate
Alaska,71.8%,70%,68%
Alabama,80.0%,75%,72%
Arkansas,84.9%,84%,81%
Arizona,75.1%,76%,78%
California,80.4%,79%,76%
Colorado,76.9%,75%,74%
Connecticut,85.5%,85%,83%
District of Columbia,62.3%,59%,59%
Delaware,80.4%,80%,78%
```

t

CSV for average starting teacher salary:

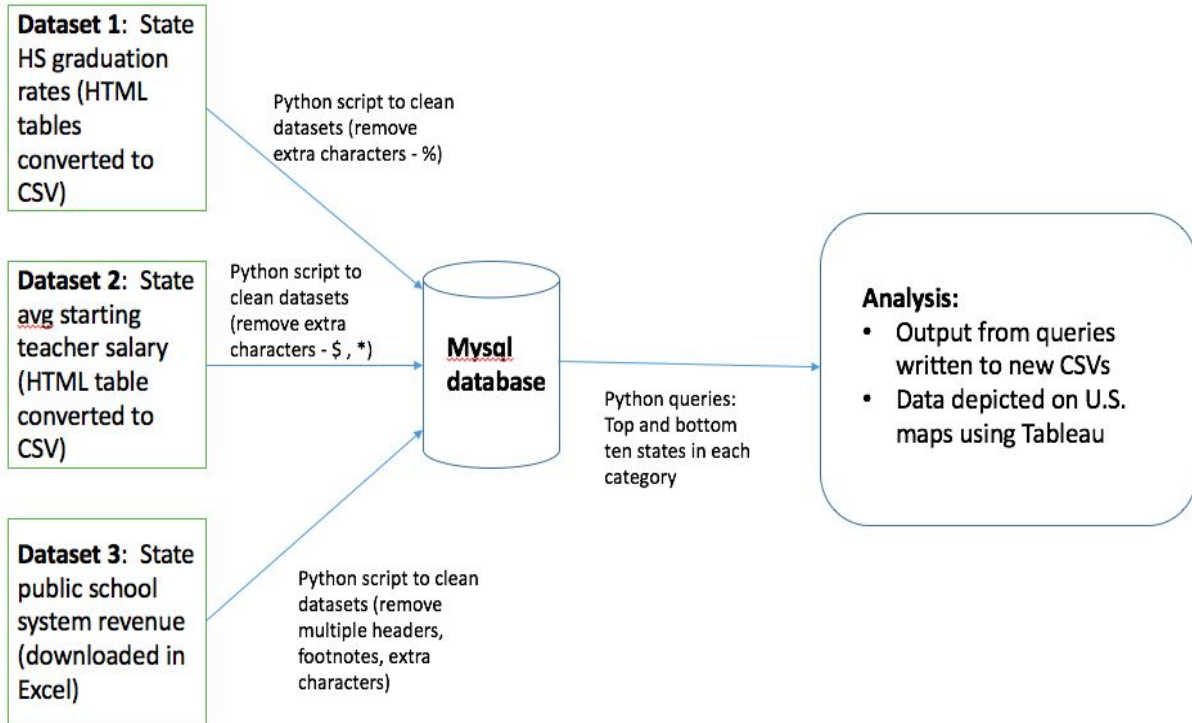
```
mroy@holden ~/assign/project $ head teachersalary.csv ]
State,Avg. Starting Salary
Alabama,"$36,198"
Alaska,"$44,166"
Arkansas,"$32,691"
Arizona,"$31,874"
California*,"$41,259"
Colorado,"$32,126"
Connecticut,"$42,924"
District of Columbia*,"$51,539"
Delaware,"$39,338"
```

CSV for public school system revenue:

```
mroy@holden ~/assign/project $ head schoolrevenue.csv ]
head: cannot open 'schoolrevenue.csv' for reading: No such file or directory
mroy@holden ~/assign/project $ head staterevenue.csv ]
Table 1.,Summary of Public Elementary-Secondary School System Finances by State:
Fiscal Year 2012 ,,,,,,,,,,
"(Thousand dollars. Detail may not add to total because of rounding. For meaning
of abbreviations and symbols, see footnotes.) ",,,,,,,,,
,,,,,,,,,
,,,,,,,,,
Geographic area,,Elementary-secondary revenue 1 ,,,Elementary-secondary expendi
ture 1 ,,,Debt,,
,,,,,,,,,outstanding,,
,,,From federal,From state,From local,,Current,Capital,,at end of,Cash and,
,,Total,sources,sources,sources,Total,spending,outlay,Other 2 ,fiscal year,secur
ities,
,,,,,,,,,
United States.....,
,"594,525,532","59,532,214","270,431,959","264,561,359","593,464,483","523,559,8
01","50,202,327","19,702,355","402,158,950","175,836,175",
mroy@holden ~/assign/project $ ]
```

2. Project Workflow

2.1. Overview



2.2. Database Design

I opted for simplicity in my database design and created one table, as that was all that was really needed for my purposes.

State
name
grad_rate
grad_rate_lowinc
grad_rate_spd
teacher_salary
school_revenue

As it is a single table, there is no relational vocabulary.

State

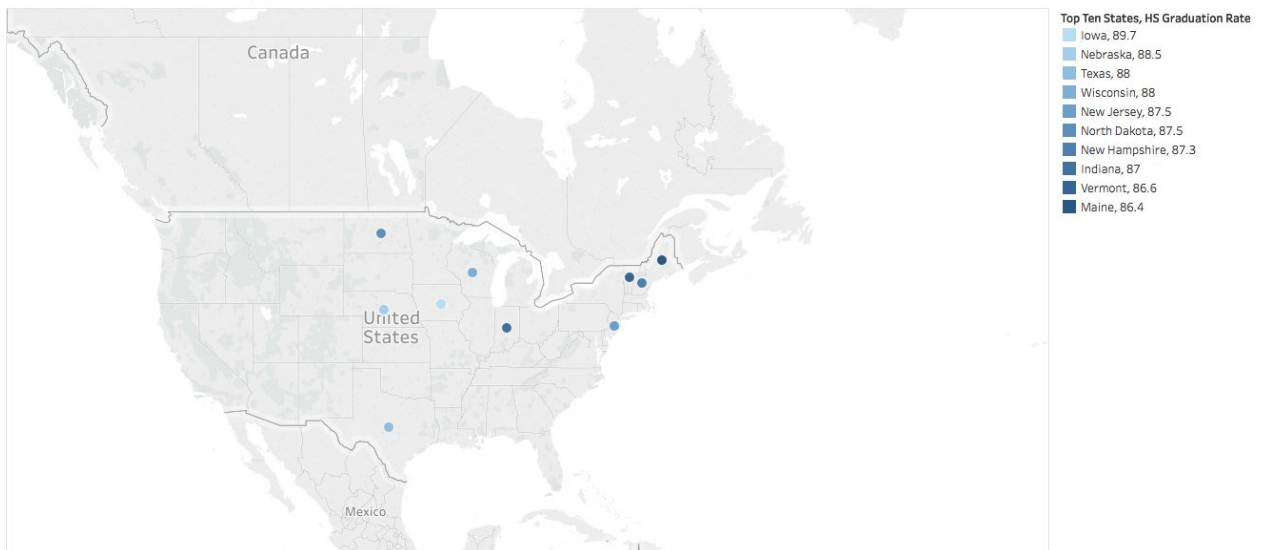
<u>id</u>	<u>name</u>	<u>grad_rate</u>	<u>grad_rate_lowinc</u>	<u>grad_rate_sped</u>	<u>teacher_salary</u>	<u>school_revenue</u>
-----------	-------------	------------------	-------------------------	-----------------------	-----------------------	-----------------------

2.3. Analysis

The data analysis aimed to highlight any patterns, trends, or correlations between high school graduation rates, teacher salaries, and school system revenue. Obviously each of these elements is complex and multi-layered, with webs of interrelated, influential factors. Examining this data side by side is in no way an assertion of causation, or even a vague suggestion of relationship. With that understanding, patterns are still interesting nonetheless, particularly regional ones in this case. As such, I wanted the data to be presented geographically, which I was able to do using Tableau.

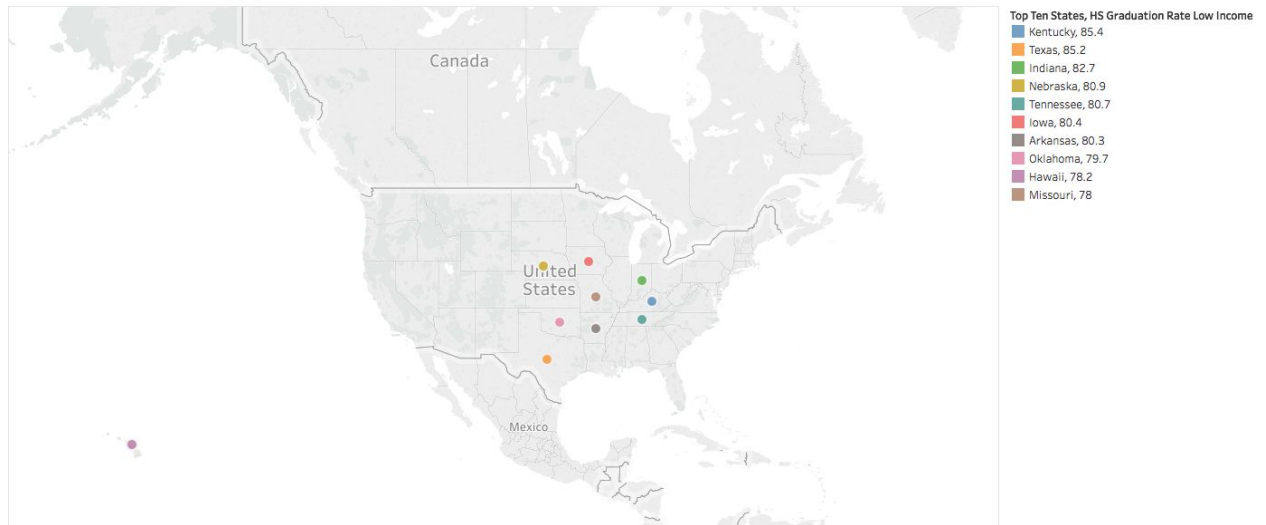
Each of my queries was written to a new CSV, which was uploaded to Tableau and depicted on a United States map. Through this visualization, what emerged was more patterning, or regional clustering, within some of the datasets, as opposed to between them. Not surprisingly, there appears to be no correlation between graduation rate, teacher salary, and school system revenue.

Highest Graduation Rates (%) - All Students



Nothing in the Deep South or west of the Rockies.

Highest Graduation Rates (%) - Low Income Students



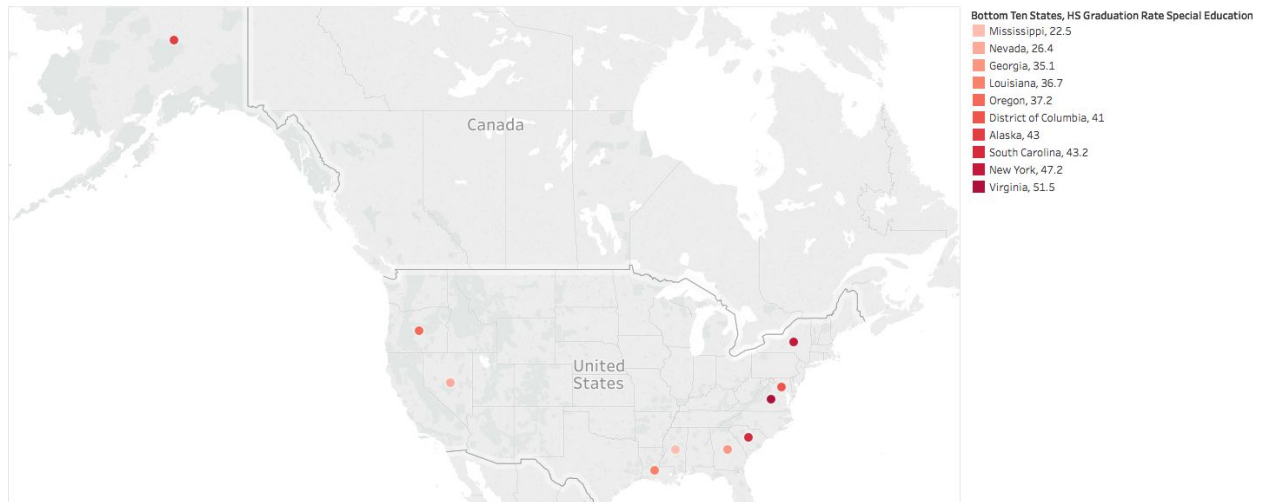
Obvious regional clustering, with Hawaii as an outlier.

Lowest Graduation Rates (%) - All Students



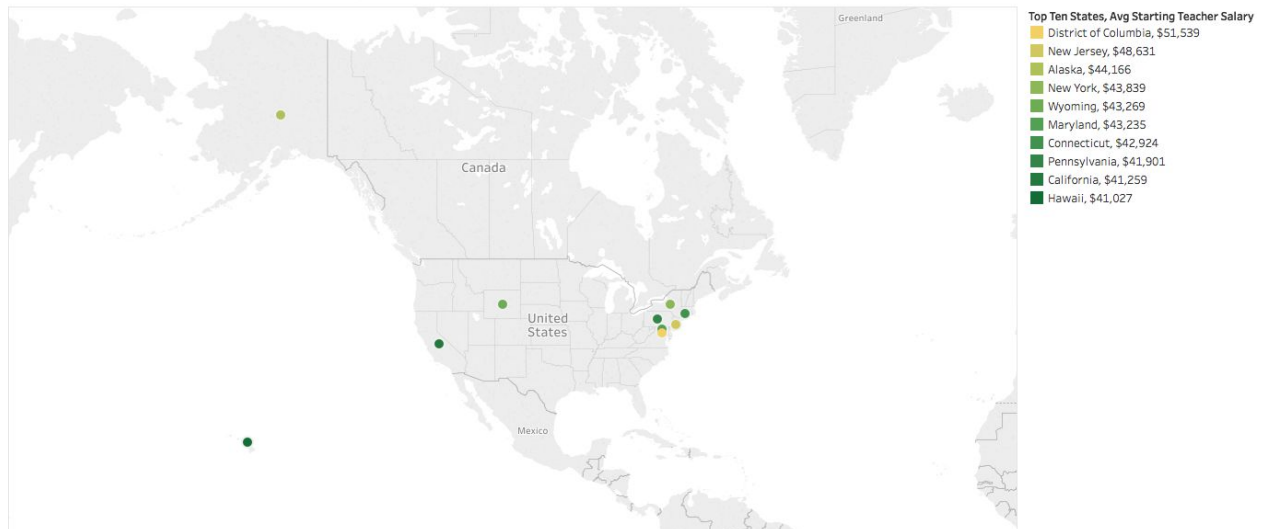
The South, the West, and Alaska.

Lowest Graduation Rates (%) - Special Education Students



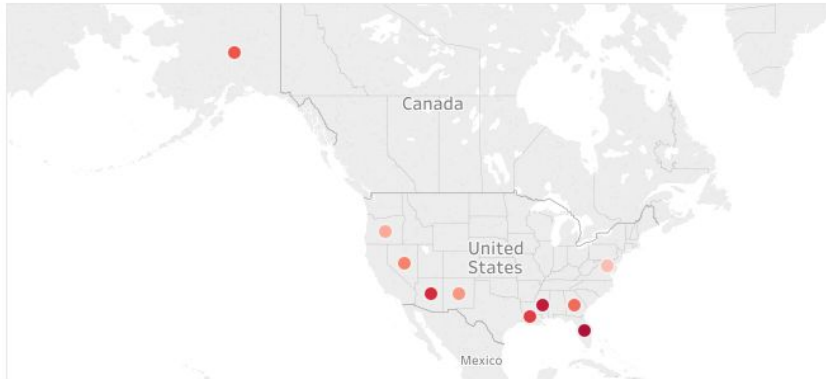
Nothing in the middle.

Highest Average Starting Teacher Salary



Alaska, Hawaii, California, and New York make sense in this category. Wyoming is somewhat unexpected. Also a clear Northeast trend.

Lowest Graduation Rates (%) - All Students



Bottom Ten States, HS ..

- District of Columbia, 68.7
- Oregon, 68.7
- New Mexico, 70.3
- Nevada, 70.7
- Georgia, 71.7
- Alaska, 71.8
- Louisiana, 73.5
- Arizona, 75.1
- Mississippi, 75.5
- Florida, 75.6

Highest Average Starting Teacher Salary



Top Ten States, Avg Sta..

- District of Columbia, \$48,631
- New Jersey, \$48,631
- Alaska, \$44,166
- New York, \$43,839
- Wyoming, \$43,269
- Maryland, \$43,235
- Connecticut, \$42,900
- Pennsylvania, \$41,259
- California, \$41,259
- Hawaii, \$41,027

Our nation's capital interestingly tops both these lists.

Highest Graduation Rates (%) - All Students



Top Ten States, HS Grad..

Iowa, 89.7
Nebraska, 88.5
Texas, 88
Wisconsin, 88
New Jersey, 87.5
North Dakota, 87.5
New Hampshire, 87..
Indiana, 87
Vermont, 86.6
Maine, 86.4

Highest Public School System Revenue



Top Ten States, Total Pu..

California, \$66,568,...
New York, \$58,757,...
Texas, \$49,815,063
Illinois, \$29,153,117
New Jersey, \$26,61,...
Pennsylvania, \$26,...
Florida, \$24,131,880
Ohio, \$22,035,362
Michigan, \$17,671,...
Georgia, \$17,556,0..

Texas and New Jersey are the only two states to be on both of these lists, and they fall in the same spot on each, third and fifth respectively.

3. Resolving Challenges

As a complete and true beginner in this field, every step of this project has come with an associated set of challenges. As with any new learning endeavor, these skills and processes have required practice, trial and error, and on a more personal note, management of frustration level. To help navigate the hurdles, I took advantage of all available course materials, including screencasts, tutorials, example code, notes, and handouts, using them as consistent reference points and scaffolding in my skill development. I tried to use the internet for help when possible, though this can be rather difficult when learning any type of new language, as you lack a supportive context or framework for interpreting the search results. However, on a few occasions during this project I found Googling my error messages to be a useful tactic in troubleshooting.

One project step I found particularly challenging was working with the school system revenue CSV, as it had much more complicated formatting than the others, including many lines of headers and footnotes and extra characters that needed to be dealt with before I could upload it properly into my database. To address the headers, I initially thought I was going to have to use a bit of code and a spread function that I found relatively confusing and difficult to adapt to my exact needs. However, after seeking some additional help and clarifying what parts of the data I was focused on, a simpler solution was found, skipping lines with `next()`, and this made much more sense to me. To fix another issue in the same CSV, I was trying to figure out how to use `re.sub` to remove rows of dots from my state names, and I was struggling with the concept. So I took an alternate approach that was easier for me to grasp in the moment, and solved the issue another way, using `re.search`. Though after asking some additional questions, I was able to understand `re.sub`, and implement it successfully in another area.

This situation provided an important reminder that there is more than one right way to get where you need to go in programming. Because there are so many rules and details, I have a tendency to view the process as fixed and rigid, as opposed to something that can be wiggled and turned and examined from different perspectives. Though the solution I came up with here, and likely throughout the project, was arguably not pretty or precise or ideal, it worked, and I could wrap my brain around it.

4. Learning an Analysis Tool

I chose to use Tableau as my project analysis tool, though I had never worked with it before. My first step was watching several of the Getting Started tutorial videos available on their website. This was helpful in providing a general orientation to the interface and basic functions.

Though there is a large library of instructional material I could have accessed, my next strategy was essentially trial and error. I uploaded some CSVs and clicked around a lot all over the place until bit by bit, I made something happen. What I ultimately came up with is not precisely what I had initially intended, but that is very often the case the first time we set ourselves loose to explore a tool. And indeed, exploration was part of my goal, so in that regard, the process was a success.

Tableau is clearly a powerful tool with some advanced capabilities. I have only just started to skirt around the edges, but I was glad to have this opportunity to do so. Though there were some features I found frustrating or counterintuitive, I am still interested to learn more, continue my experimentation, and expand my skillset.