Government Expenditure Data Exploration & Analysis Using Python

By Pedro Mena[*], Leslie Kerby[±], Derick Nielson[‡], Katherine Wilsdon[+], Paul Gilbreath[§], Connie Hill[◊], Konner Casanova[♦] & Kyle Massey[**]

The goal of improving cost efficiencies is a constant endeavor of all organizations. This is especially true for governments, where public perception often has the ability to affect budget allocations. The data used in this analysis consisted of publically available state expenditures from 2018 and 2019 for the state of Idaho. The dataset contains the record of over 2 million state expenditures across all state agencies. The data analysis was performed using Python and the Pandas library. Visualizations were created using the Matplotlib package. The data exploration showed that Idaho's Departments of Health and Welfare, Education and Transportation spent the most in this time period. The analysis also determined which Summary Objects, Sub-Object and Vendors experienced the greatest changes between the two years. Comparisons were also done using publicly available data on reported budget allocations by the states of Arkansas, California, Texas and Montana to see how spending differs between Idaho and these states based on percentage and per capita. Finally, suggestions for improvement in the areas of health care and employee transportation were given. These include methods of improving competition in health care, reducing travel through expanded teleconferencing and providing incentives to employees for reduced travel cost.

Keywords: *data science, budget analysis, python, pandas, government spending*

Introduction

Improving cost efficiencies is a goal that all organizations constantly seek to achieve. This is true for both private business as well as the public sector. In the United States, many of the individual states have laws that limit debt or require a balanced budget. This is different from the federal government which has the ability to carry a deficit, though there is a self-imposed debt ceiling. Due to these constraints, there is much incentive for state governments to analyze their respective budgets to identify areas for improvement and areas of waste. Still, this task can be very difficult, as even smaller states typically have hundreds of

[*]Graduate Student/Graduate Research Assistant, Idaho State University, USA.
[±]Professor, Idaho State University, USA.
[‡]Undergraduate Student, Idaho State University, USA.
[+]Undergraduate Student, Idaho State University, USA.
[§]Graduate Student, Idaho State University, USA.
[◊]Graduate Student, Idaho State University, USA.
[♦]Graduate Student, Idaho State University, USA.
[**]Graduate Student, Idaho State University, USA.

employees and potentially millions of different expenditures each year. The scope of these expenditures is also large; states typically have expenditures in infrastructure, education, public health, public worker's travel, etc. Analyzing this data can be extremely difficult and inefficient without the proper tools. The recent focus on the development and expansion of data science tools, such as the Pandas package in Python, presents a great opportunity for state governments to improve their fiscal spending. These tools can allow state governments to easily, quickly and efficiently analyze expenditures at a low cost. A comparison between the state of Idaho's spending and that of four different states will also be done. These states were chosen for the comparison due to each of the state's demographics and the availability of data from these states. Finally, this paper will analyze budget expenditures and identify areas of possible improvement for the state of Idaho, using data analytic and data mining techniques in Python.

Literature Review

The use of Big Data in business and finance has been expanding more and more in recent times. The reason for this is a belief that the more in depth analysis can result in better decision making and improve performance of the business and deliver a high return on investment. In a study done in 2011 by McKinsey Global Institute (MGI), it was estimated that the use of Big Data and analytics could save health providers over 300 million in annual value (Manyika et al. 2011). The study also estimated that Big Data could improve productivity growth by 0.5 percent in the European Union. Benefits cited by MGI in the use of Big Data include: improved automation, increased transparency and a quicker, more effective approach to identify vulnerability.

Big Data has also been used in many government applications. According to a publication in Communications of the ACM, The United States federal government began using Big Data in 2009 to help improve transparency. The focus of this was on transportation, health care and education related data (Gang-Hoon and Ji-Hyong 2014). Other government agencies that have taken advantage of big data include Singapore's Risk Assessment and Horizon Scanning, Japan's Ministry of Education and South Korea's Ministry of Public Administration.

State of Idaho Background

The state of Idaho is located in the northwest area of the United States. According to the United States Census Bureau, Idaho is ranked 13th in total area, approximately 83,569 square miles, but only has a population slightly below 1.8 million as of July 2019 (United States Census Bureau 2019a). This yields population density around 20 people per square mile, ranked 39th out of 50. The most populous area of the state is in the capital city, Boise and the surrounding towns. The state is divided into 44 counties, most of which are rural. The key industries of the state include energy, technology, agriculture, outdoor recreation

and tourism (Commerce Idaho 2020). Figure 1 shows Idaho's geographical location and county in the United States.

Figure 1. *State of Idaho*

In the last decade, Idaho has experienced major population growth. In 2010, the census reported the population of the state to be near 1.56 million, with the population in 2019 being reported around 1.8 million. The state has experienced an approximate 15% population increase in the last 10 years. This ranks very high within the United States, with the census bureau reporting that Idaho experienced the largest growth in the country in 2017 at 2.2% (United States Census Bureau 2017). Due to this large increase in population, it is likely the state will need to adjust its spending to ensure that public services are adequately funded and supported, and that infrastructure is maintained to accommodate the needs of the increasing population. This increases the need to use modern data analytic tools to reduce waste and monitor spending.

The State of Idaho employed 25,423 people in 2019 (Buxton 2020). This includes 13,070 employees listed as classified. Classified employees are subject to the state's evaluation process for promotion, merit and dismissal. The remaining employees are listed as non-classified. The average age of each state employee is 45 years, slightly above the national average of 42. The state government consists of 90 different agencies. These include areas such as health and welfare, defense, education, executive operations, etc. It should be noted that each department is divided differently. One example of this is the Industrial Commission. This agency is divided into three districts, while the Department of Agriculture is divided into four districts. Larger state agencies are typically organizationally structured

differently by region. According to the State Controller's office, most of the boundaries for these regional districts are county driven. It should be noted that not all state agencies have regions or districts through the state. For example, an agency may only have one office in Boise.

Methodology

State Controller's Office and Data

The data used for this analysis was obtained from the state of Idaho's Controller's office. This office is responsible for paying and managing all of the state's expenditures. This includes payroll for state employees, providing public access to state budget information and maintaining the state's accounting infrastructure. The Controller's office is required to collect data on every transaction that uses funds from the state. The head of this office is an elected office by the people of Idaho. Each term is four years in length and there are no term limits for this office. The current state controller for Idaho is Brandon D. Woolf who has been in the position since 2012.

The data from the State Controller Office was originally downloaded and saved as two CSV files. The two CSV files consist of all state expenditures from early July of 2017, beginning of the 2018 fiscal year, through June of 2019, end of the 2019 fiscal year. There were approximately 13.6 million transactions lines for expenditures during this time period between the two fiscal years. The data size of each CSV was approximately 2 gigabytes. Each expenditure in these datasets consists of 21 features including the data of the expenditure, the area of government where it occurred, the amount of the expenditure and the vendor who received the payment. The full list of features is given in Table 1. All data used in the project is publically available as part of the Transparent Idaho Program. As such, there are no issues with this dataset containing data that is confidential or otherwise covered by privacy laws.

Table 1. *Features from Expenditure Dataset*

Unnamed: 0	Fiscal Year	GL Account Code	GL Account Name	Agency Name
Agency Code	Amount	Effective Date	Process Date	Area of Government
Fund Name	Fund Code	Fund Category	Fund Detail	Fund Category
Object Name	Summary Object Name	Sub Object Code	Sub Object Name	Earn Ben Group
Vendor Name				

Pandas

In order to perform the analysis of the expenditure data it was necessary to make use of Python and the data science packages available. The majority of the analysis was done using the Pandas package. Pandas is a free open source package

designed to aid users in data analysis. Pandas is available standard or through download on most Python distributions, such as Anaconda. The package was developed in 2008 in response to the need for better data manipulation tools in Python. The project receives funding from many different entities, including the University of Paris Saclay Center for Data Science and Two Sigma. Python 3 is strongly recommended for use with Pandas, as all Python 2 support has been discontinued as of January 2020. Pandas only requires the use of the NumPy package in order to operate properly. The package is designed to work on Windows, Mac and Linux environments (McKinney 2011).

One of the most powerful aspects about Pandas is the ability to read and convert datasets into a structured form known as a DataFrame. A Pandas DataFrame is a 2-D Python list that stores values in a tabular form. As is the case with most Python packages, Pandas is considered high level and as a result, there are trade-offs in efficiency for improved ease of use. In addition to the ability to store values, Pandas has many tools for data analytics as well. This includes the ability to group data by user input, such as values present in the dataset or the data type of the values in the set. Also, Pandas can perform basic statistical analysis, such as standard deviation, mean, median, etc. Pandas is also designed with the ability to add, remove and combine groups of data between different datasets (McKinney 2019).

Data Exploration

In data science, one of the most crucial parts of any project is to explore the data that will be used for analysis. Figure 2 shows a portion of the header for the 2019 fiscal year DataFrame. It should be noted that none of the methods used in this project change the contents of the original CSV files.

Figure 2. *Sample of Header from Fiscal Year 2019 Dataset*

Eight of the 22 features were found to be numerical. The remaining 14 columns consisted of string data. Figures 3 and 4 show the descriptive statistics for both fiscal year datasets.

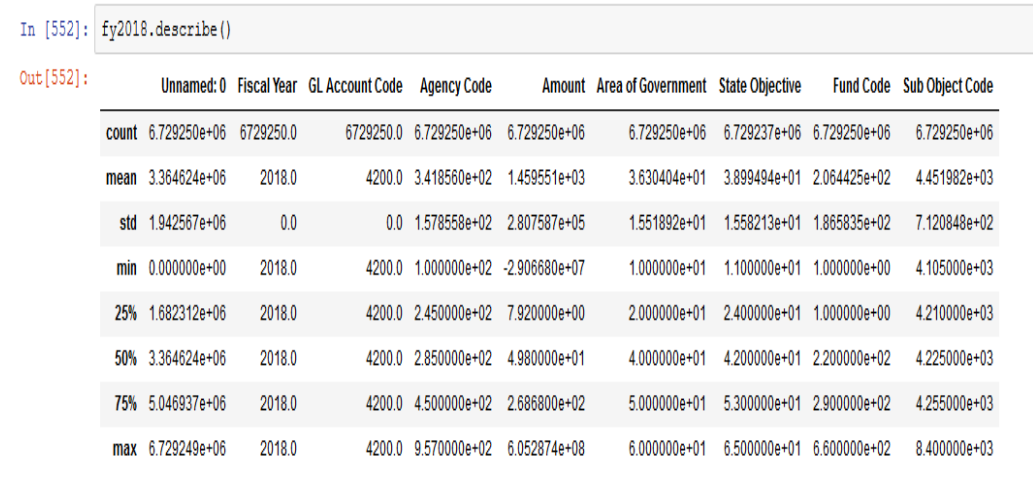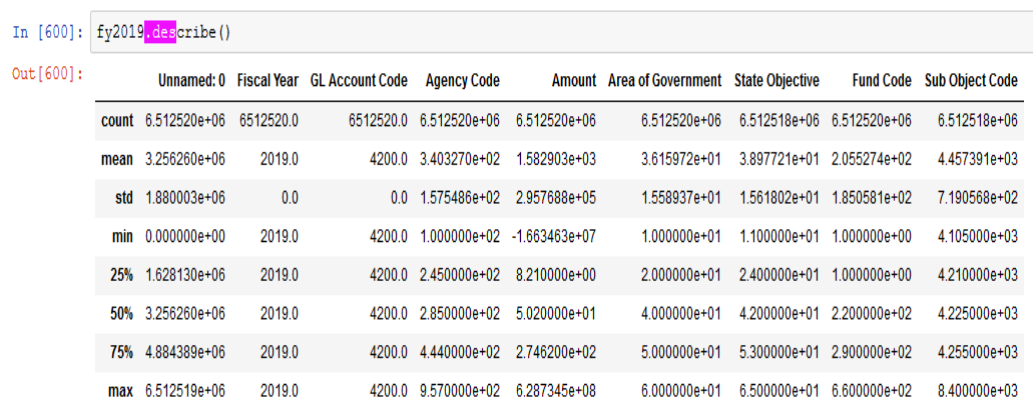**Figure 3.** *Descriptive Statistics from Fiscal Year 2018 Dataset*

In [552]: `fy2018.describe()`

Out[552]:

	Unnamed: 0	Fiscal Year	GL Account Code	Agency Code	Amount	Area of Government	State Objective	Fund Code	Sub Object Code
count	6.729250e+06	6729250.0	6729250.0	6.729250e+06	6.729250e+06	6.729250e+06	6.729237e+06	6.729250e+06	6.729250e+06
mean	3.364624e+06	2018.0	4200.0	3.418560e+02	1.459551e+03	3.630404e+01	3.899494e+01	2.064425e+02	4.451982e+03
std	1.942567e+06	0.0	0.0	1.578558e+02	2.807587e+05	1.551892e+01	1.558213e+01	1.865835e+02	7.120848e+02
min	0.000000e+00	2018.0	4200.0	1.000000e+00	-2.906680e+07	1.000000e+01	1.100000e+01	1.000000e+00	4.105000e+03
25%	1.682312e+06	2018.0	4200.0	2.450000e+02	7.920000e+00	2.000000e+01	2.400000e+01	1.000000e+00	4.210000e+03
50%	3.364624e+06	2018.0	4200.0	2.850000e+02	4.980000e+01	4.000000e+01	4.200000e+01	2.200000e+02	4.225000e+03
75%	5.046937e+06	2018.0	4200.0	4.500000e+02	2.686800e+02	5.000000e+01	5.300000e+01	2.900000e+02	4.255000e+03
max	6.729249e+06	2018.0	4200.0	9.570000e+02	6.052874e+08	6.000000e+01	6.500000e+01	6.600000e+02	8.400000e+03

Figure 4. *Descriptive Statistics for Fiscal Year 2019 Dataset*

In [600]: `fy2019.describe()`

Out[600]:

	Unnamed: 0	Fiscal Year	GL Account Code	Agency Code	Amount	Area of Government	State Objective	Fund Code	Sub Object Code
count	6.512520e+06	6512520.0	6512520.0	6.512520e+06	6.512520e+06	6.512520e+06	6.512518e+06	6.512520e+06	6.512518e+06
mean	3.256260e+06	2019.0	4200.0	3.403270e+02	1.582903e+03	3.615972e+01	3.897721e+01	2.055274e+02	4.457391e+03
std	1.880003e+06	0.0	0.0	1.575486e+02	2.957688e+05	1.558937e+01	1.561802e+01	1.850581e+02	7.190568e+02
min	0.000000e+00	2019.0	4200.0	1.000000e+00	-1.663463e+07	1.000000e+01	1.100000e+01	1.000000e+00	4.105000e+03
25%	1.628130e+06	2019.0	4200.0	2.450000e+02	8.210000e+00	2.000000e+01	2.400000e+01	1.000000e+00	4.210000e+03
50%	3.256260e+06	2019.0	4200.0	2.850000e+02	5.020000e+01	4.000000e+01	4.200000e+01	2.200000e+02	4.225000e+03
75%	4.884389e+06	2019.0	4200.0	4.440000e+02	2.746200e+02	5.000000e+01	5.300000e+01	2.900000e+02	4.255000e+03
max	6.512519e+06	2019.0	4200.0	9.570000e+02	6.287345e+08	6.000000e+01	6.500000e+01	6.600000e+02	8.400000e+03

The next step in the data exploration was to identify if any of the columns had consistent data entries. Three features were consistent throughout the set: Fiscal Year, GL Account Code and GL Account Name. Using the information from the Amount columns, it was found that Idaho's expenditures totaled over 9.9 billion USD in 2018 and 10.3 billion USD in 2019. It is important to note that many of these expenditures, while billed from the state, may come from sources other than tax payer funds. One example is retirement expenditures, where funds come from the state retirement program rather than the state budget.

It was now possible to begin exploring where the expenditures occurred. Idaho uses five different categories to classify an individual expenditure. The first is the state agency where the expenditure occurred. The next level is Object Name which is used to distinguish personnel costs versus capital outlays. Then Summary Object is used to determine what kind of expenditure is being billed for such as, employee benefits, computer services, supplies, etc. The next level is Sub-Object, the type of the expenditure. Some examples are employees, retirement/sick leave and worker's compensation. Finally, the last level is Vendor, which includes the individual employees or organization billing the state. One example would be

from an expenditure in July 2018. In this case the agency that created the expenditure was State Board of Accountancy. The summary object listed was administrator supplies and the sub-object was listed as office supplies. Finally, the vendor that billed the state was Treasure Valley Coffee INC. Table 2 shows this categorization as it appears in the dataset.

Table 2. *Expenditure Categorization Example*

Agency Name	Object Name	Summary Object	Sub-Object	Vendor
Accountancy, State Board of	Operating Expenses	Administrative Supplies	Office Supplies	Treasure Valley Coffee Inc.

Starting at the agency level it was found that the Department of Health and Welfare had the largest number of expenditures in both years with over 110,000 expenditures in both fiscal years. The Idaho Department of Corrections was second both years, with over 55,000 expenditures and the Idaho Department of Transportation was third with over 45,000 expenditures in each of the years. Figure 5 shows the top agencies for fiscal year 2018. In terms of dollar amounts, Health and Welfare spent the most, at over 2.5 billion USD in 2018 and nearly 3 billion USD in 2019. Department of Education was second at approximately 2 billion USD in both fiscal years. Department of Transportation was third, at approximately 1 billion USD in both years.

At the Summary Object level it was found that awards and claims and miscellaneous payments as agent made up the largest dollar amounts at this level. Both objects' total expenditures added up to over 2.5 billion USD in both fiscal years. Wages and employee benefits followed, with wages totaling close to 1 billion each year and benefits close to 500 million. It should be noted, that there were significantly more expenditures for employee benefits than wages. This is not unexpected, as there are many types of benefits that are billed each cycle. These include: dental coverage, unemployment insurance, retirement, medical, etc. Figure 6 shows the amounts spent by summary object for the 2018 fiscal year. At the Sub-Object level, employees and insurance made up the largest amount of expenditures across the two fiscal years by a large margin. Both had approximately 1 million USD in expenditures in both years. This is not surprising as medical benefits made up a large number of Summary Objects. Due to the large number of this category a graph could not be provided.

At the vendor level, it was discovered that the state paid out to 72,324 individual vendors during the 2018 fiscal year, and 76,371 vendors during the 2019 fiscal year. 1.5 billion USD of this was spent across 8,800 expenditures on vendors that are redacted. There are a number of reasons why a vendor may be redacted. This includes medical privacy issues relating to the federal Health Insurance Portability and Accountability Act (HIPPA), security concerns for the individual(s) or a vendor may be under the age of 18. Still, this could be an area of concern, as there are fewer accountability safeguards for redacted expenditures. It should be noted that the State Controller's office claims to have internal measures to allow internal auditors to evaluate these expenditures without violating privacy

laws. Other vendors of note include: St. Luke's Regional Medical Center, the largest hospital network in the state and United Behavioral Health. Both vendors had expenditures over 100 million in both fiscal years.

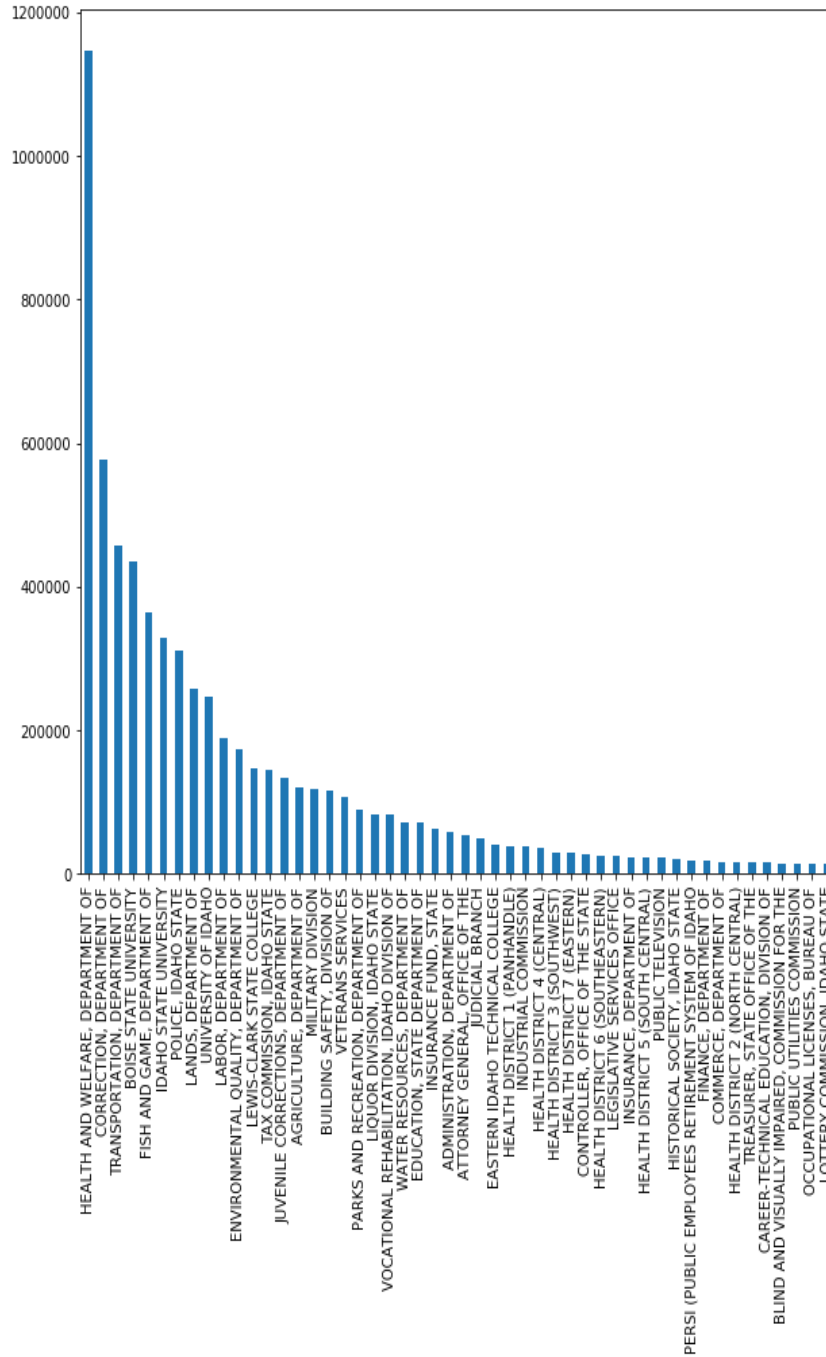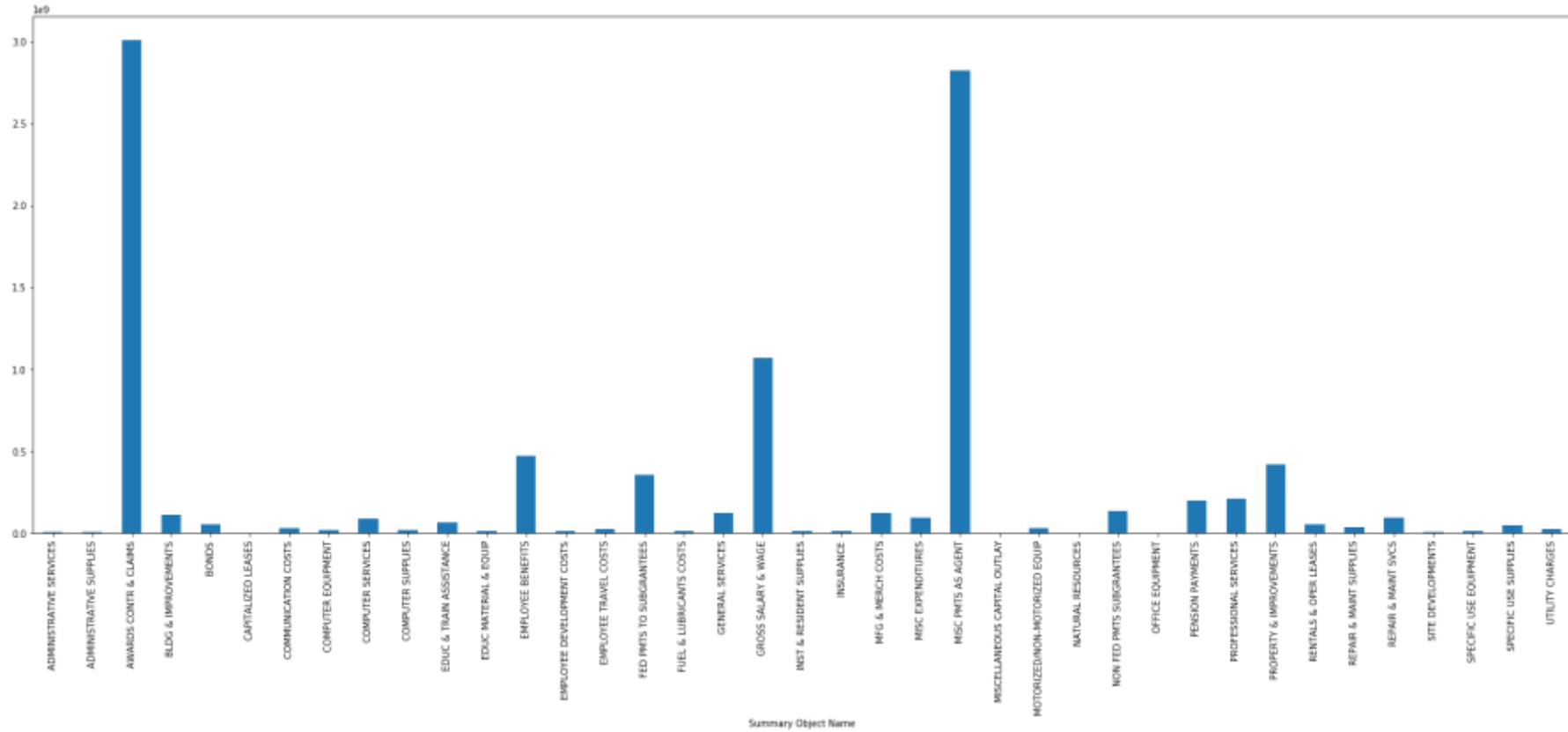**Figure 5.** *Fiscal Year 2018 Expenditure Counts*

Figure 6. *Expenditure Amount for Individual Summary Objects*

The next objective of the data exploration was to determine where the largest changes had occurred between the two fiscal years. This process was done using all four categories for identifying expenditures. The first category examined was the agency category. Agencies as a whole spent 487 million USD more during the 2019 fiscal year than in 2018. On average, each agency's expenditure spending increased by 5.2 million USD or 3.8%. It was found that in terms of dollars, the Departments of Health and Welfare, Transportation, and Education had the largest increase. Health and Welfare was the largest with an increase close to 160 million USD. In terms of percentage, the Office of the Secretary of State, Department of Lands and Military Division experienced the largest increases. The Office of Secretary of State was the largest with an increase of over 54%, likely due to elections. Tables 3 and 4 summarize the largest agency increases by both percent and dollar amount. It should be noted that this process can be used to determine where the largest decreases occurred. This can be useful in identifying areas that are underfunded or have made significant efficiency improvements. Table 5 summarizes the largest agency decreases in terms of dollars.

Table 3. *Top 5 Greatest Year over Year Changes (by Dollars)*

Agency Name	Difference($)	Percent (%)
Department of Health And Welfare	1.595946e+08	5.601805
State Department of Education	9.317064e+07	4.541356
Department of Transportation	6.970577e+07	7.508004
Military Division	4.131726e+07	41.668744
Department of Lands	3.586630e+07	41.965609

Table 4. *Top 5 Greatest Year over Year Changes by Percent*

Agency Name	Difference ($)	Percent (%)
Idaho Office of the Secretary of State	2.254961e+06	54.023836
Department of Lands	3.586630e+07	41.965609
Military Division	4.131726e+07	41.668744
Department of Environmental Quality	1.605245e+07	32.710614
Correctional Industries	2.884625e+06	30.489316

Table 5. *Top 5 Smallest/Negative Year over Year Changes (in Dollars)*

Agency Name	Difference ($)	Percent (%)
Catastrophic Health Care	-2.429092e+06	-12.062386
Persi (Public Employees Retirement System of Idaho)	-2.771477e+06	-1.353824
Idaho State Historical Society	-2.916141e+06	-28.482509
Department of Labor	-8.692220e+06	-6.019738

The analysis of the other three categories showed similar results. The average increase for each Summary Object in 2019 was 12.8 million USD or an increase of 4.04% from 2018. In terms of Summary Objects, the largest increases in dollars were Awards & Claims, Property & Improvements, and Buildings & Improvements. Awards & Claims, was the largest, increasing by 141 million USD between the two fiscal years. In terms of percentage,

Buildings & Improvements, Computer Supplies and Specific Use Equipment experienced the largest increase. Tables 6, 7 and 8 summarize the top five Summary Object increases.

Table 6. *Top 5 Greatest Year over Year Changes (in Dollars)*

Summary Object Name	Difference ($)	Percent (%)
Awards Contr & Claims	1.410918e+08	4.693434
Property & Improvements	1.119937e+08	26.599756
Bldg & Improvements	6.569845e+07	59.706208
Misc Pmts as Agent	6.420825e+07	2.272240
Gross Salary & Wage	3.813386e+07	3.568694

Table 7. *Top 5 Greatest Year over Year Changes by Percent*

Summary Object Name	Difference ($)	Percent (%)
Bldg & Improvements	6.569845e+07	59.706208
Computer Supplies	4.841004e+06	28.262855
Specific Use Equipment	2.869434e+06	27.173937
Property & Improvements	1.119937e+08	26.599756
Capitalized Leases	2.014129e+05	14.745854

Table 8. *Top 5 Smallest/Negative Year over Year Changes (in Dollars)*

Summary Object Name	Difference ($)	Percent (%)
Educ & Train Assistance	-1.245134e+06	-1.896331
Pension Payments	-2.491714e+06	-1.265241
Motorized/Non-Motorized Equip	-2.607726e+06	-8.510727
Computer Equipment	-3.270671e+06	-16.140637
Employee Benefits	-2.209452e+07	-4.692094

The Sub-Object category experienced an average increase of 1.17 million dollars or a 208% increase between the 2018 and 2019 fiscal years, across all Sub-Objects. Many Sub-Objects experienced major increases in expenditure spending percentage-wise, which skews the arithmetic average. The machine and equipment Sub-Object had the largest percentage increase during this time, at over 45,000%. Other significant increases include the construction in progress and Land Non-1099 Reportable Sub-Objects, which also had increases of 13,129% and 7,929%, respectively. These are most likely due to the starting of major projects, as these amounts were relatively small in the previous year. In dollar amounts, the largest increase occurred with the Medical Assistance Vendors with an increase of 147 million dollars. Other significant increases include School Appointments, with an increase of 95.7 million USD, and Land Non-1099 Reportable, with an increase of 58 million USD. Tables 9, 10 and 11 summarize the Sub-Object increases.

Table 9. *Top 5 Greatest Year over Year Changes by USD Amount*

Sub Object Name	Difference ($)	Percent (%)
Medical Assistance-Vendors	1.472141e+08	6.484279
School Apportionment	9.579945e+07	5.386734
Land - Non 1099misc Reportable	5.804148e+07	7929.909276
Infrastructure - 1099m Reportable	4.951701e+07	12.079128
Employees	3.682643e+07	3.696804

Table 10. *Top 5 Greatest Year over Year Changes by Percent*

Sub Object Name	Difference ($)	Percent (%)
Machinery & Equipment	185692.00	45512.745098
Construction In Progress-Land-Building-Equipment	433548.08	13129.624537
Land - Non 1099misc Reportable	58041479.33	7929.909276
Educational Equipment-Furniture	39697.18	4411.680114
Non-State Employee-1099misc Box 3	37900.00	3661.835749

Table 11. *Top 5 Smallest/Negative Year over Year Changes*

Sub Object Name	Difference ($)	Percent (%)
School District-Federal Subgrant	-5.559486e+06	-2.554015
Environmental or Ecological Testing	-9.763956e+06	-90.027430
Premiums	-2.324098e+07	-7.160197
Group Insurance Health & Accident	-3.009131e+07	-12.228138
Refunds	-6.661183e+07	-14.849007

The vendor category experienced an average increase of 8,030 dollars per vendor in fiscal year 2019. This is an average increase of 6.6% with a median of 0.96%. The largest increase was Knife River Corporation Mountain West with an increase of over 79 million. It needs to be noted that Knife River Mountain West changed names from Knife River Northwest during this time period. As a result, there was a corresponding decrease of nearly the same amount from Knife River Northwest. The actual change was an increase of 137,000 dollars. In order to perform effective analysis of many vendors it will be necessary to learn the context of many of the vendors. Other significant vendor increases were from the State Treasures Office and The First American Title Company. Tables 12, 13 and 14 summarize the changes in vendor expenditures between the 2 fiscal years.

Table 12. *Top 5 Greatest Year over Year Changes*

Vendor Name	Difference ($)	Percent ($)
Knife River Corporation Mountain West	79270938.93	1289.430678
State Treasurers Office	51451052.14	3.695622
First American Title Company	43573135.43	3951.012729
Health & Welfare Department	35022656.15	78.447689
Blue Cross Of Idaho Care Plus, Inc.	31092732.75	77.237668

Table 13. *Top 5 Greatest Year over Year Changes by Percent*

Vendor Name	Difference ($)	Percent (%)
Nicholas And Company	58979.52	8.300630e+19
Home Depot	4747.30	6.681231e+19
Napa Parts Inc.	4083.55	5.747086e+19
Home Depot Store 1808	10295.49	4.829871e+19
Paypal Inc	12213.04	2.864721e+19

Table 14. *Top 5 Smallest/Negative Year over Year Changes*

Vendor Name	Difference ($)	Percent (%)
Veyo, Llc	-1.506158e+07	-98.122833
Staker And Parson Companies	-1.553464e+07	-38.048156
Diageo North America	-1.919531e+07	-99.997565
Knife River Corporation Northwest	-7.913310e+07	-99.924590
Redacted	-9.047389e+07	-6.184988

Exploring the expenditures using the four different categories can provide great insight into the expenditure data, especially as it changes from year to year. This analysis was able to show which categories had the largest increases and decreases both in terms of dollars and percentages, between both fiscal years. In terms of state agency and vendors, performing this analysis is useful in determining which groups are having the largest change in spending. This analysis showed that the Department of Health and Welfare was the agency with the largest dollar increase, while the Department of the Secretary of State had the largest percentage increase of all agencies. This allows for a quick determination if spending is trending according to expectations or if it is becoming out of control. Looking at the Summary Objects and Sub-Objects allows for analysis of what the expenditures are being spent on. Also, it can allow the state to make more informed decisions if an agency or summary object is adequately funded. In terms of Summary objects, it was found that Awards & Claims had the largest dollar increase, while Buildings and Improvements had the largest percentage increase. It is important to note that the decreases are just as important to analyze as the increases. Deeper analysis into the decreases can help identify methods that can help other groups improve spending efficiency. Also, it can be a useful tool for analyzing outside vendors to make better determinations of which vendors are performing the best.

Comparison to Other States

This section will compare Idaho's expenditures in different areas to that of several other states. The data will be compared against the states of Arkansas, Texas, Montana and California. Doing this comparison helps identify areas where Idaho's spending may be inefficient. Also, it may help identify areas of spending that are performing well in relation to other states. First, it is important to understand how Idaho total costs are distributed. According to the exploration of the expenditure data, 30.2% of Idaho's expenditure spending came from health and human services. 20.1% of expenditure spending was

from state government expenses. 9.8% was in transportation related expenses, 5.5% in higher education, 2.6% in corrections, 22.8% in kindergarten through 12th grade education (K-12), and the remaining expenditures are classified as other. Based on this data, the total per capita spending in 2019 for Idaho was approximately $6,129 per person. Figure 7 shows a visualization of this spending.

Figure 7. *Idaho Spending by Percent (Fiscal Year 2019)*
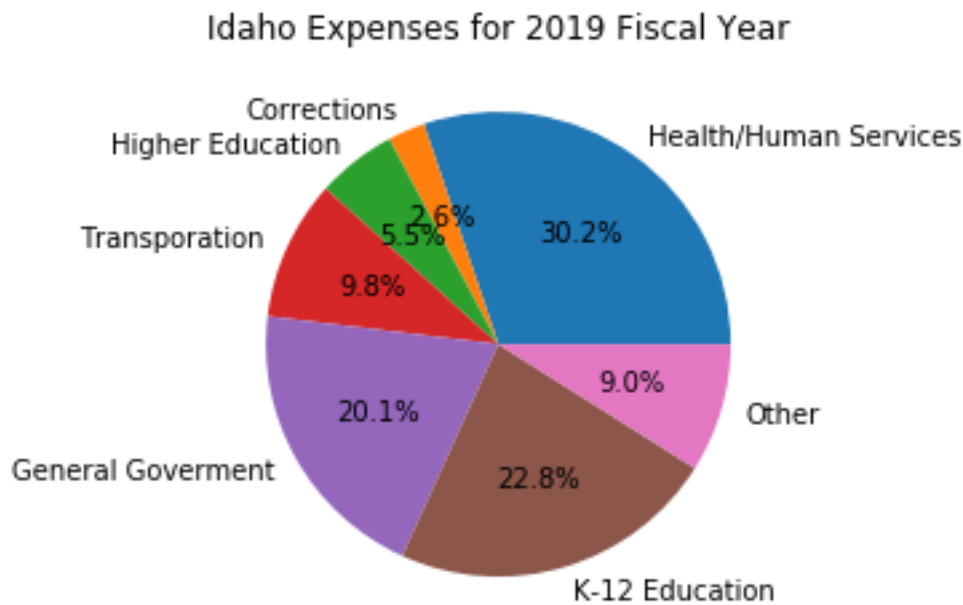
Arkansas

Arkansas is a smaller state in terms of land size, ranked 29th, and has a population of around 3 million. It was chosen for analysis in order to compare Idaho to a state that is very different in terms of size and population density. According to the state's appropriations for the 2020 fiscal year, the state has allocated 29.8% of its total allocations for healthcare (State of Arkansas 2020). This is close to Idaho's 30.2%. Areas of significant difference include state government expenditures, where Arkansas has allocated 14.7%, compared to Idaho's 20% in this area. Arkansas also has allocated more in the area of education. Arkansas allocated 33.8% in education. 14.2% for K-12 education and 19.6% for higher education. This is more than Idaho which allocated 28.3% for education, with 22.8% for K-12 and 5.5% for higher education. One possible reason for the larger higher education rates is that the universities in Arkansas have much higher endowments when compared to those in Idaho. For example, University of Arkansas has an endowment over 1 billion USD, while Boise State University has an endowment of 115 million (NACUBO 2019). Also, Idaho universities may rely more on federal funding due to the presence of the Idaho National Lab. Idaho and Arkansas have similar spending in the area of corrections, with Arkansas allocating 1.6% to Idaho's 2.6%. Also similar are the state's allocations in the area of transportation, with Arkansas

spending 10.3% of its allocations in this area and Idaho spending 9.8%. Figure 8 shows Arkansas' spending allocations in terms of percentage.

**Figure 8.** *Arkansas Spending by Percent (2020 Fiscal Year)*
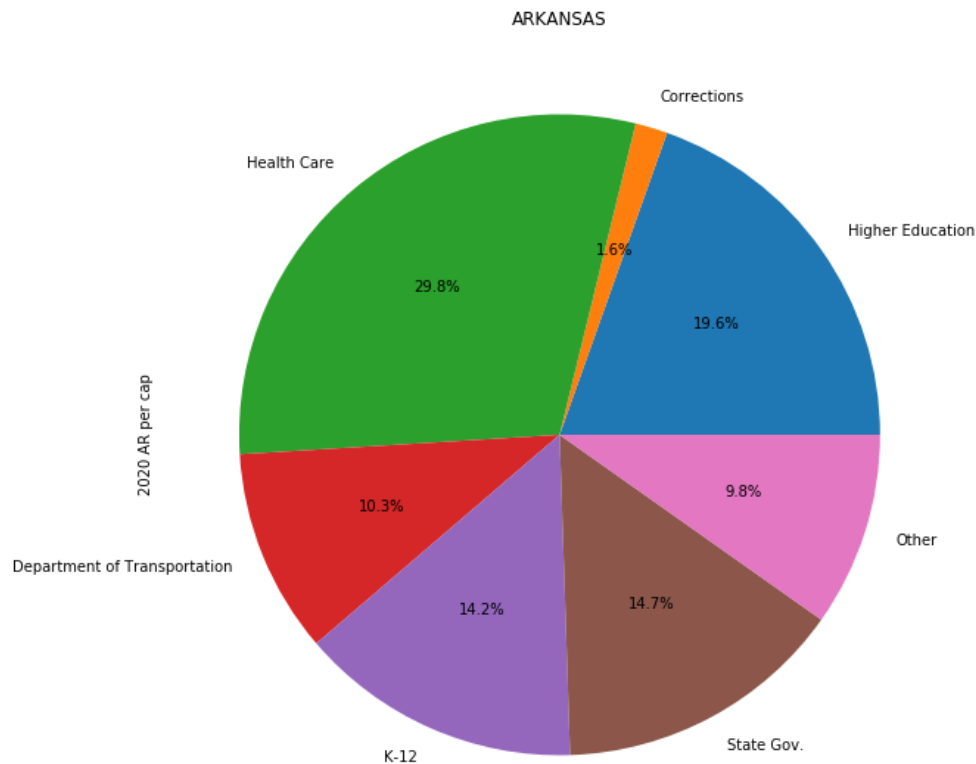
Using population estimates from the census bureau, it was also possible to determine estimates of Arkansas spending in terms of per capita and to compare these to Idaho's for another view on spending differences. Arkansas spent more in many areas than Idaho, on a per capita basis. These include education, K-12 and higher education combined, where Arkansas spent over $3,300 per person, compared to Idaho's $1,720. Transportation was another area with a noticeable difference, where Arkansas spent over $1,000 per person and Idaho spent $603. In terms of health care, Arkansas spent $2,900 per person, compared to Idaho at $1,865. Overall, Arkansas spent significantly more per capita than Idaho. Arkansas spent over $9,747 per person in total spending, while Idaho only spent $6,129. Figure 12 compares the per capita spending of Idaho and all the states analyzed.

California

California is the most populous state in the Unites States; its population is estimated to be over 39 million, and 3rd in terms of land size. California was chosen in order to compare Idaho's spending patterns to the state with the largest population in the country. According to the Governor's Budget Office in California, the state allocated more in K-12 for 2020 with 27.7% of total

expenditures allocated for K-12 education compared to 22.8% for Idaho. California was also higher in corrections, with 11.4% of spending going to this area. In higher education, California allocated 8.2% similar to Idaho's 5.5% (Newsom 2020). The two states also have allocated a similar amount in transportation with California allocating 8.4% compared to 9.8% for Idaho. California also allocated close to the same amount, percentage-wise, in 2018 in health and human services at 32%. State government allocations were significantly less at 5.3%. Figure 9 shows California's spending allocations for 2020.

Figure 9. *California Spending Distribution*

California Allocated Expenses for 2020
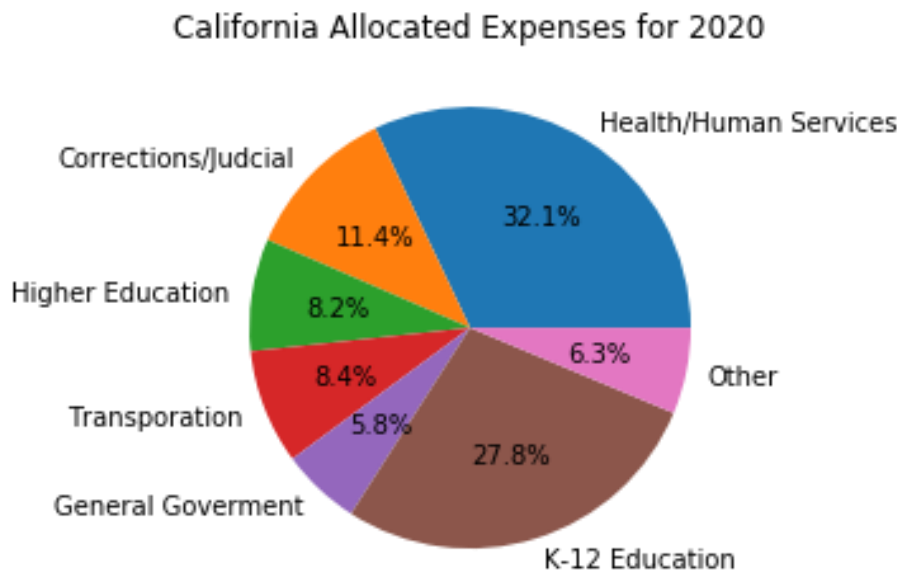
Data was also available to perform analysis on a per capita basis for expenditures in California. This is useful as Idaho and California are very different in terms of population and size. This analysis may help determine areas where either too much or too little may be allocated. In the area of transportation, California spent $473 USD per person, while Idaho spent $603 per person. In health care, California was similar, at $1,803 per person, while Idaho spent $1,865 per person. Finally in education, California spent approximately $2,018 per person, while Idaho spent approximately $1,720 per person. California's overall per capita allocation for 2020 was $5,625.

Texas

Texas is the second largest state in terms of size and it is also second in population, with approximately 29 million people living in the state. Texas was chosen for the same reason as California, to provide a comparison between Idaho and a large state. However, Texas is considered to be a very different state than California and is more similar to Idaho. According to data available from the Texas' Legislative Budget Board, the state's budget from 2018–2019 was 221 billion in HB-1 (Texas Legislative Budget Board 2018). It should be

noted that Texas divides up health care cost differently with spending for both acute health services and long term health services. These two services accounted for 29.3% of the budget, approximately 54 billion dollars. This is close percentage wise to Idaho's 30.2%. Texas's exact spending on corrections was found to be 2.7%, which compares close to Idaho's 2.6%. Higher education for Texas is close as well, with 7.2% allocated compared to Idaho's 5.5%. Also, transportation spending was similar, with Idaho spending, 9.3% of its allocations on transportation and Texas spending 12.1%. Idaho's state government expenses are higher in percentage than Texas at 20% compared to 2.6%, though Texas allocates its employee benefits separately. Texas spent a similar amount on K-12 education than Idaho, at 23.4%. Figure 10 shows how Texas allocated funds for 2018–2019 fiscal period.

Figure 10. *Texas Spending by Percent (2018–2019)*
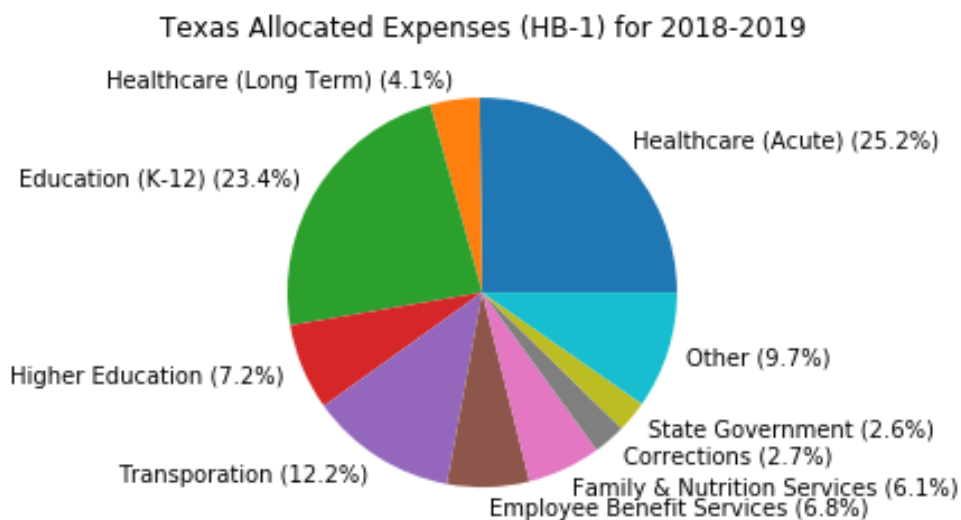
On a per capita basis, Texas spent near $7,727 per capita overall. $980 was spent per capita in transportation, compared to $379 in Idaho. In the area of health care, Texas spent $2,351 on a per capita basis, compared to $1,865 for Idaho. Finally, in the area of education, Texas spent $1,816 per person on K-12 education and $514 per person on higher education. This gives Texas a per capita cost of education total of $2,362 per person, compared to Idaho's $1,720.

Montana

Montana is the 4th largest state in the United States and has an estimated population of slightly over 1 million. Montana's demographics are similar to Idaho in terms of population. According to Montana's fiscal report for 2019 budget, HB-2 accounted for over 86% of state expenses. The amount of HB-2 was approximately 10.31 billion dollars. Of this amount, HB-2 allocated approximately 4.25 billion dollars, 42% of the total budget, to health and human services (Legislative Fiscal Division 2017). This likely includes the cost

of health benefits to state employees. This is more than Idaho at 30.2% of total allocations for the same period. Idaho spent more in government expenses with 20% of the allocated budget, compared to Montana's 6%. Montana allocated close to the same amount in education, with Montana allocating 26% of the budget, approximately 2.65 billion dollars, between both K-12 and higher education. In comparison, Idaho allocated 28.3% toward education. Montana also spent more in the areas of transportation, and corrections at 18% and 8% respectively. Idaho spent 9.8% and 2.6% respectively in these areas. Figure 11 shows a visualization of Montana's spending allocations. Overall, Montana spent significantly more than Idaho, with a total of $10,140 per person.

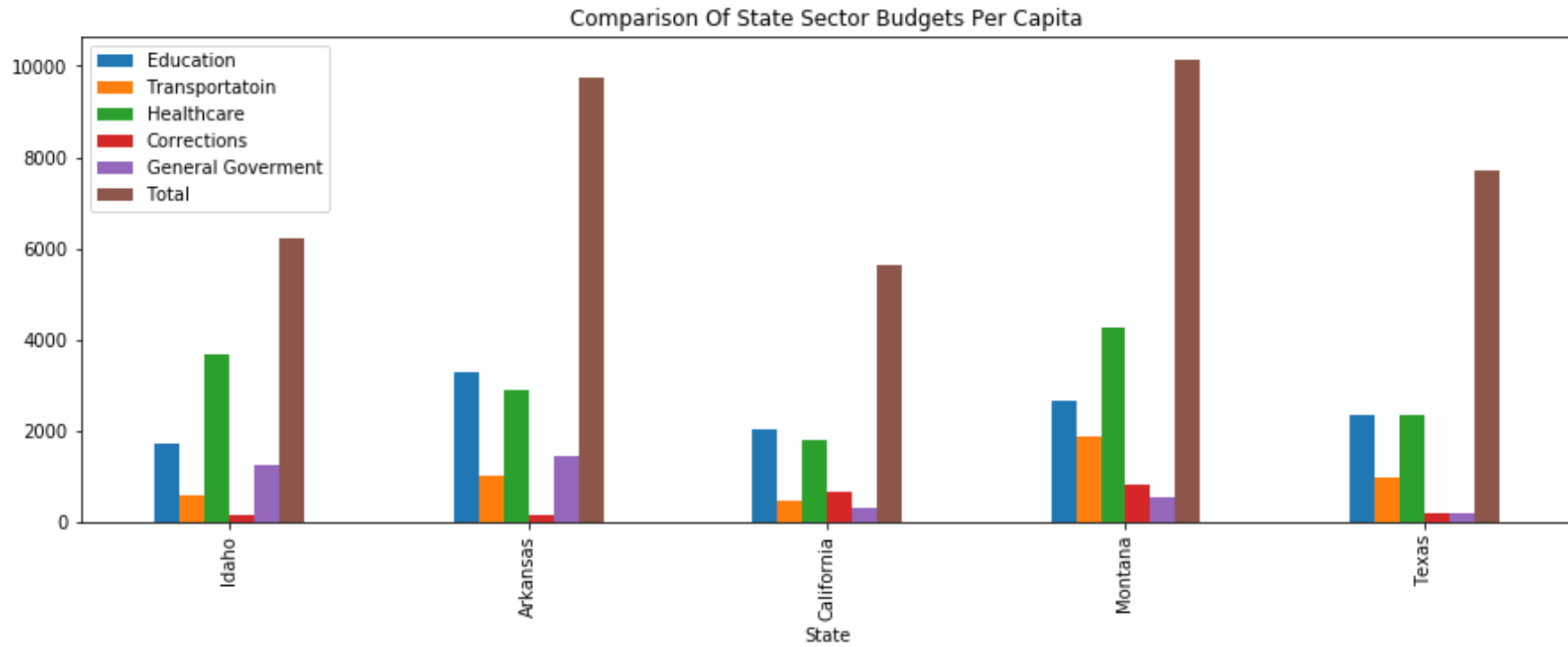**Figure 11.** *Montana Allocation Estimates for 2019*
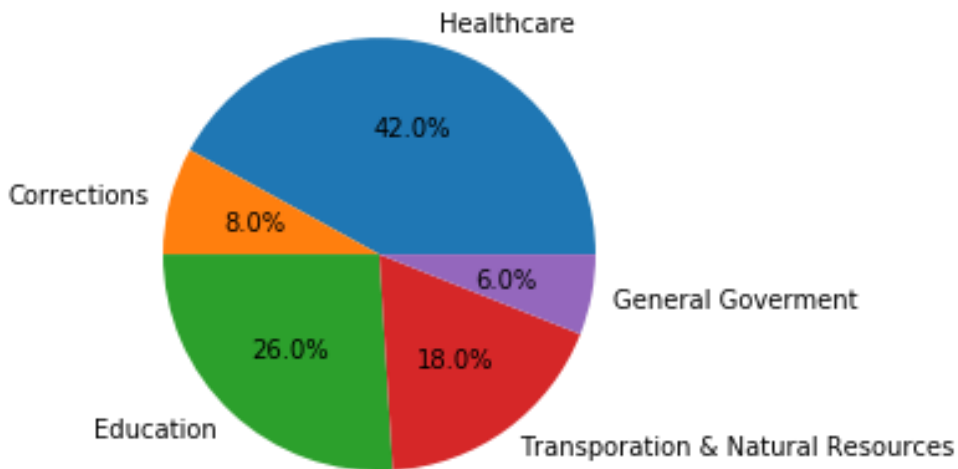
Figure 12. *Comparison of State's Expenditure Spending (per Capita)*

Areas of Improvement

The results of the data exploration show that there are many areas for possible improvements in reducing cost in Idaho. The analysis showed that there is a great deal of expense in both health care and transportation. As a result, these areas were further analyzed to see if suggestions could be offered to provide guidance on improvement. Health care was chosen as it is an important issue and employee transportation was chosen due to Idaho's large geographical area.

Health Care

Health care is a complex issue in the United States. In previous years, both state and federal governments have sought to reduce health care costs and allow more access. The results have been mixed, with more people having access to health care, but the costs of health care and health insurance have increased, in some cases dramatically. In Idaho for example, in 2019 Blue Cross of Idaho proposed a premium of $305 for a Bronze health plan, a 5% increase from the previous year (Dutton 2018). In 2020, the lowest priced Bronze plan from Blue Cross of Idaho was quoted at $374 per month according to the YourhealthIdaho. org. Figure 13 shows a screenshot from the Idaho health insurance website. The state government is facing similar issues, as a large percentage of its budget expenditures are from the areas of health care benefits provided to employees and health care services provided by the state. Again, the Department of Health and Welfare is Idaho's largest state agency and by far has the largest number of expenditures. According to the State Controller's Office, the reason for this is the growth and expanded use of Medicaid. While it is unlikely that the suggestions in this section will solve the issues surrounding health care, the savings could allow Idaho to continue providing health benefits to its employees and health services to its citizens without major cost increases in the short term.

Figure 13. *Section of a Quote from YourhealthIdaho.org*
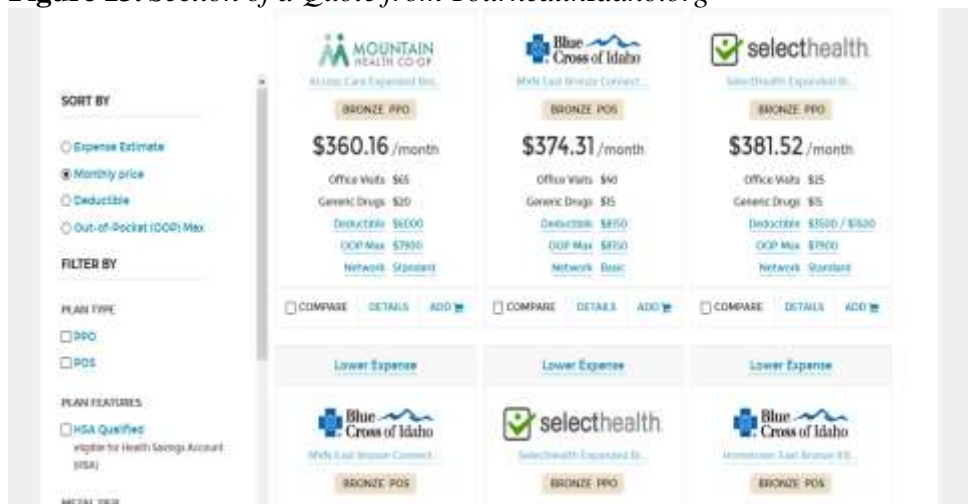
The first issue that should be addressed is the issue of redacted vendors. These vendors made up over 1.37 billion USD of the state's expenditures in the 2019 fiscal year. It is true that there are likely good reasons for redactions and the state does claim to have measures to help internal auditors in this area. Still, transparency is questionable, especially to external parties. A system to classify redactions by reason, for example redacted for security, could go a long way in helping external auditors identify waste and improve accountability.

Another area that could be addressed is the issue of health cost transparency. This is an issue that is being evaluated not only at the state level, but at the federal level as well. In Idaho, both St. Luke's and United Health were among the top five vendors in terms of dollar amounts to the state of Idaho. Each had total expenditures valued over 100 million USD for the 2018 and 2019 fiscal years. Increasing the transparency of health cost to the consumer could provide incentive for them to search for alternatives that might better suit their individual needs and increase competition in health industry and reduce cost which would help Idaho maintain its ability to offer health insurance to its employees and health services to its citizens.

Employee Transportation

The next area to be examined for possible improvements was travel/ transportation for public employees. Idaho is a vast state, ranked 13th in the nation in terms of land size. Idaho also has a significant number of people that live in rural areas that are a number of hours from the major population centers in the state. As a result, many state employees are required to travel a large number of hours in order to perform their duties in these rural areas. For example, employees for the Department of Health and Welfare may have to travel from the city of Idaho Falls to the town of Victor, a drive of over an hour each way. To further complicate matters, Idaho's climate is very unpredictable. It is common for areas of the state to receive significant snowfall into the months of March and April. This can make travel more difficult and dangerous for the employees. It is very likely that an employee may spend half of their work day or more serving a single client in a rural area, for a routine task. According to the 2010 Census, Teton County has a significant Hispanic population, 16.7% (United States Census Bureau 2019b). Due to this, there is an increased chance that an interpreter may be needed to assist with services, which would further increase the costs of providing services.

A possible solution to this issue is expanding the use of teleconferencing technology where possible. The use of this technology would significantly reduce the need for state employees to travel to these rural areas. This would save hours of travel time, as well as reduce costs associated with travel, including expenditures needed to pay for gasoline, car maintenance, insurance, etc. Also, the cost of language translation could be reduced, as interpreters could also participate in the teleconferences from other locations. This would largely eliminate the need for interpreters to travel to rural areas and bill less hours. According to the Bureau of Labor Statistics, the median hourly rate of a translator/interpreter is 24.95 USD per

hour (United States Bureau of Labor Statistics 2020). Another benefit of the reduced travel requirements would be the possible increase in competition in the bidding for interpreting contracts with the state, as there would be fewer constraints for the contractors. This would have the benefit of reducing cost and improving the quality of the contractors.

Teleconferences could also be used for other applications as well. As noted, Idaho is a large state and travel to the capital, Boise, can take 4–5 hours from other areas inside the state. Expanding the use of teleconferences for things such as meetings, department trainings, etc. would also help reduce cost, by allowing state workers to attend these functions from their typical place of work, rather than having to travel. This not only would reduce cost in terms of travel, but may also improve productivity, inner department communication and allow for more training opportunities to improve operations.

It should be noted that there are issues that would need to be addressed with the expansion of teleconference technology. First, it is likely that lower income clients would need to be provided with a tablet or other device to participate in teleconferences. Still, the cost of devices that are needed for teleconferences are generally low when compared to the cost needed for travel. A formal and effective training service may also need to be provided to both clients and employees to ensure that the service can be used as intended.

The next area of possible improvement for travel/transportation involves reducing the cost of both in and out of state travel with the use of contracts. The data exploration showed that 40%, around 8–9 million dollars, was spent on out-of-state travel during both the 2018 and 2019 fiscal years while the remaining portion was used on in state travel. The benefit of contracts is that it would provide guaranteed income to travel vendors, such as hotels and airlines, while allowing the state employees to enjoy a lower rate for travel. This should be very effective in-state, as the contracts would be easier to negotiate, but out-of-state contracts would make sense in other areas, such as air travel.

Contracts could help in reducing the cost of both in-state and out-of-state travel, but it ignores an underlying issue with the travel system. As the system is set up now, there are no incentives for employees to participate in reducing the cost of travel. Travel can be difficult on employees and when making arrangements such as airfare, employees generally will plan for the most convenient arrangement, rather than the most practical. Examples of this are flying instead of taking a shuttle bus, or choosing a more expensive flight instead of a lower cost one. If the state was to determine a target price for types of travel and then reward employees for meeting this price, it would increase the probability that an employee would trade some convenience for cost reduction, as they also would benefit. This would also increase and incentivize employee involvement in the cost reductions.

Conclusions

The analysis done in this study shows the ability that data science tools, such as the Pandas package in Python, have in analyzing large amounts of data easily and in a reasonable amount of time. In this case, Pandas was able to read, display and order over 6 million different expenditures from the state of Idaho's controller office. Using basic Pandas functions with this data, it was possible to explore and determine which state agencies had the most expenditures, as well as which agencies spent the most during the 2018 and 2019 fiscal years. The data exploration was also able to group data by Summary Object and Sub-Object to determine what the use and the amount of these expenditures was. For example, Idaho spent near 5% more on health than Texas, but over 12% less than Montana. Also, using Pandas it was possible to see which vendors had billed the state the most times over the two fiscal years. This analysis allowed for comparison with other states such as, California and Texas, to see how Idaho's expenditure allocations varied from other states, on both a percentage and per capita basis. This was done successfully and the areas where Idaho's spending was significantly different from others were identified. This included incentives for employees to find savings during travel, expanded use of teleconferencing and remote technology and improved health cost transparency. Finally, using the insight gained from this data exploration, it was possible to identify areas where improvement could and should be made, specifically; health care and employee transportation. It is hoped that results from this study will help encourage state and local governments to make greater use of data analytic tools in order to improve spending efficiency and increase transparency. The Controller's Office has expressed interest in expanding this analysis to answer more questions regarding expenditures as well as performing regression analysis in the future.

Acknowledgments

References

Buxton SE (2020) *Report to the governor FY 2021 change in employee compensation & benefits report*. Idaho: Idaho Division of Human Resources.

Dutton A (2018) *Idaho health insurers raised rates 27 percent this year.   For 2019? 8 percent*. Idaho Statesman.

Gang-Hoon K, Ji-Hyong C (2014) Big data applications in the government sector: a comparative analysis among leading countries. *Communications of the ACM* 57(3): 78–85.

Idaho Commerce (2020) *Key industries*. State of Idaho.

Legislative Fiscal Division (2017) *Legislative fiscal report 2019 Bienniumvolume 1– Statewide perspectives*. State of Montana.

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C et al. (2011) *Big data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute Key Industries.

McKinney W (2011) *Pandas: a foundational python library for data analysis and statistics*. PyHPC 2011.

McKinney W (2019) *Pandas: powerful python data analysis toolkit release 0.24.2*. Pydata.org.

NACUBO (2019) *U.S. and Canadian Institutions listed by fiscal year (FY) 2019 endowment market value and change in endowment market value from FY18 to FY19 (Feb 2019)*. NACUBO.

Newsom G (2020) *Governor's budget summary 2020-2021*. State of California.

State of Arkansas (2020) *Appropriation summary fiscal year 2020*. Arkansas: State of Arkansas.

Texas Legislative Budget Board (2018) *State budget by program*. Texas: Texas Legislative Budget Board.

United States Bureau of Labor Statistics (2020) *Interpreters and translators*. Occupational Outlook Handbook. U.S. Department of Labor.

United States Census Bureau (2017) *Idaho is nation's fastest-growing state, Census Bureau reports*. United State Census Bureau.

United States Census Bureau (2019a) *Quick facts, Idaho*. United States Census Bureau.

United States Census Bureau (2019b) *Quick facts, Teton County, Idaho*. United States Census Bureau.