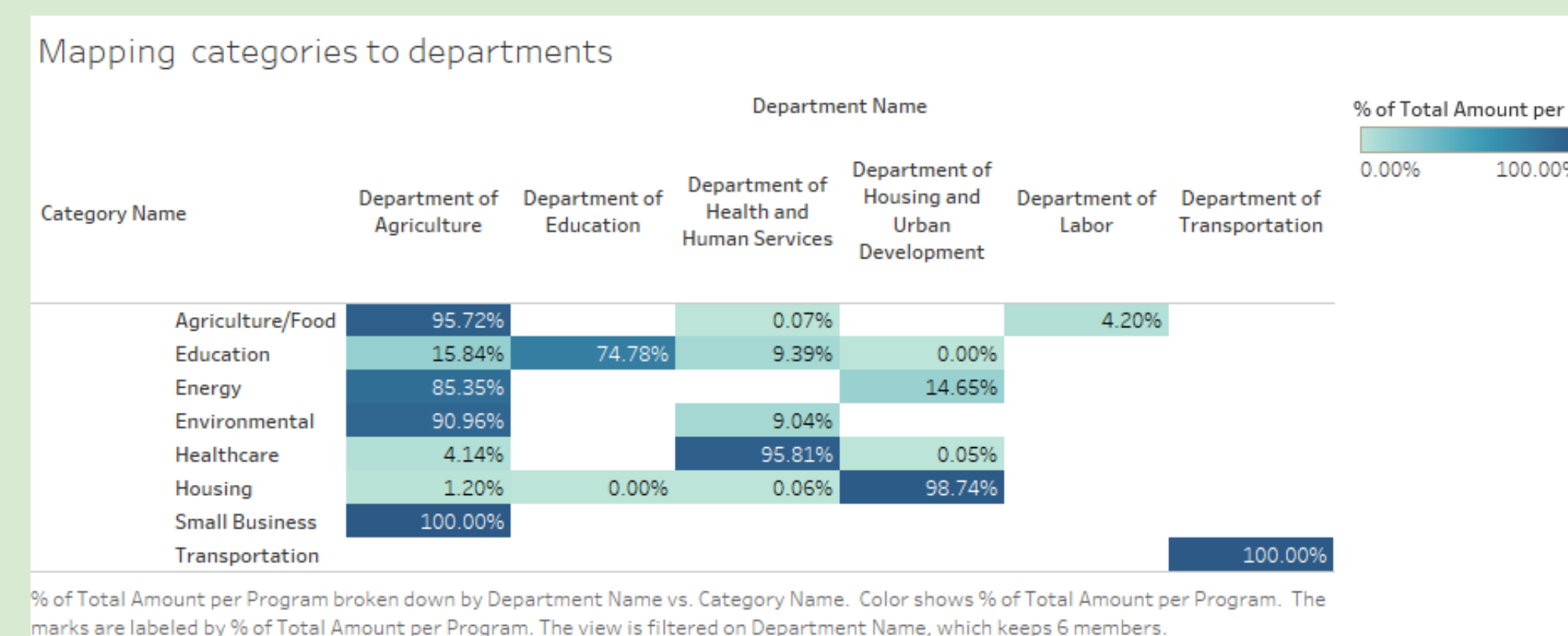




DATA PREPARATION

- The data set federalfunds.txt contains 2,666,153 rows, with each row describing one dollar amount allocation event (grant) from one CFDA program in one zip code in one year.
 - The data spans years 2005-2009 with similar number of records in each year
 - The data set shows program funds allocation across 9,335 unique CFDA codes.
 - The data set shows program funds allocation across 31,517 unique zip codes
- The data set federalfundsagencykey.xlsx contains a sample of 1,306 CFDA programs. The official list of CFDA programs can be found at the CFDA.gov website which shows 2,310 unique CFDA codes. We named this dataset CFDAPrograms.
 - Based on the official list of 2,310 CFDA programs, in the dataset CFDAPrograms, there are 136 Agencies managing between 102 and 1 CFDA programs
 - Our analysis of agency descriptions in the dataset CFDAPrograms and has revealed that there are 45 clearly identifiable departments overseeing between 1 and 28 agencies, and between 1 and 514 CFDA programs. Departments are identifiable by the first two digits of the CFDA program
 - We have developed an algorithm that makes departments identifiable, through a department code, for most individual CFDA programs based on the first two digits of the CFDA program
 - After this analysis, 34 CFDA programs for the official CFDA list remained unmatched with a department, accounting for less than 2% of the programs listed in the CFDAPrograms dataset.
- We needed to determine how many unique CFDA codes can be matched to an identifiable federal department.
 - Using our department matching algorithm, we were able to match 7,963 CFDA codes with one of the 45 identified departments, which amounts to 85% of all CFDA codes appearing in the data set federalfunds.txt.
 - The remaining 15% of unmatched CFDA codes amount to .3% of all grant spending.
 - Finally we attached the department code to the departments that are most identifiable.

- Suggested Categories for the 1306 CFDA numbers are not necessarily the best fit for the analysis
 - One CFDA program code does not map uniquely to a category code does not map uniquely to.
 - Therefore we choose to use Department based classifications for our analysis rather than category classification. For most departments there is a predominant category but not for all (Department of Agriculture, for example).



- Preparing the final table for analysis:
 - Initially, we reduced the size of federalfunds.txt by focusing on only the columns that we needed (AUDIT YEAR, STATE, CFDA, AMOUNT) and grouped by AUDIT YEAR, STATE, CFDA to create table StateLevelAnalysis1 (SA1).
 - In order to get the department code to join to the we created a table, SA2, that joined the tabled with department code with SA1 by the CFDA numbers.
 - Finally we joined SA2 with the given state information giving us our table ready for analysis.

EDUCATION DESCRIPTIVE STATISTICS

- For each state and for each year, 2005-2009, we analyzed Per Capita Income and Spending on Education per Capita

2005

- In 2005 the highest per capita income was CT with \$50,226 (per capita spending was \$608.63)
- The Lowest was MS with \$29,675 (per capita spending was \$866.29).
- The average was \$34,667.96.
- In 2005 the highest educational funding per capita was SD that had \$5,060 (per capita income was \$33,772)
- The lowest was NV with \$186 (per capita income was \$38,637).
- The average was \$916.20.

2006

- In 2006 the highest per capita income was CT with \$54,191 (per capita spending was \$533.75) and the lowest was MS with \$27,711 (per capita spending was 1071.41).
- The average was \$36,821.08.
- In 2006 the highest educational funding per capita was SD that had \$5288.18 (per capita income was \$35,203), the lowest was NV with 248.08 (per capita income was \$38,637).
- The average was \$956.90.

2007

- In 2007 the highest per capita income was CT with \$54,191 (per capita spending was \$459.59)
- The lowest was MS with 29,237 (per capita spending was \$1,010.16).
- The average was \$38,567.68.
- In 2007 the highest educational funding per capita was SD that had \$5,375.97 (per capita income was \$38,729), the lowest was NV with \$188.63 (per capita income was \$40,137).
- The average was \$956.90.

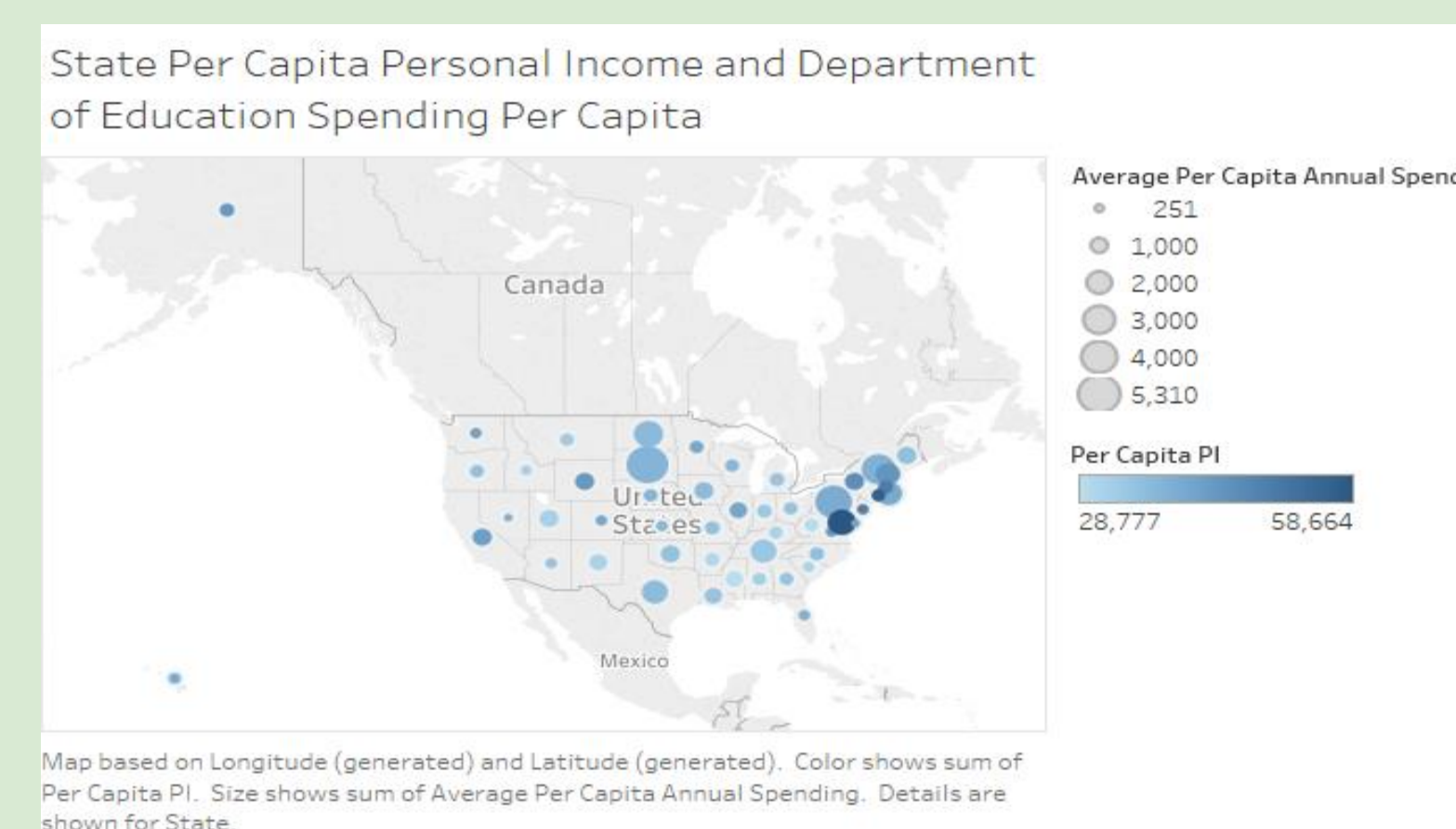
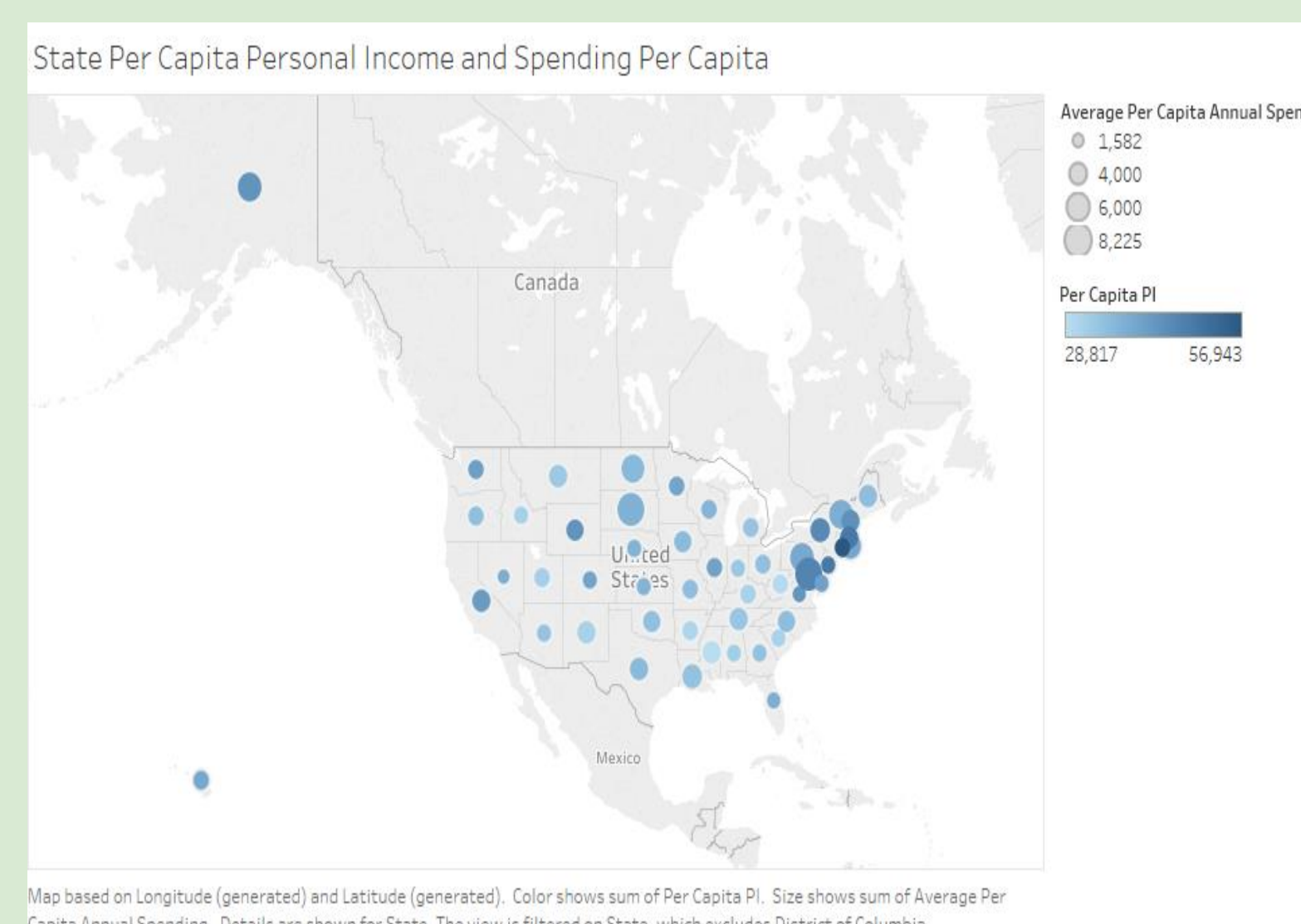
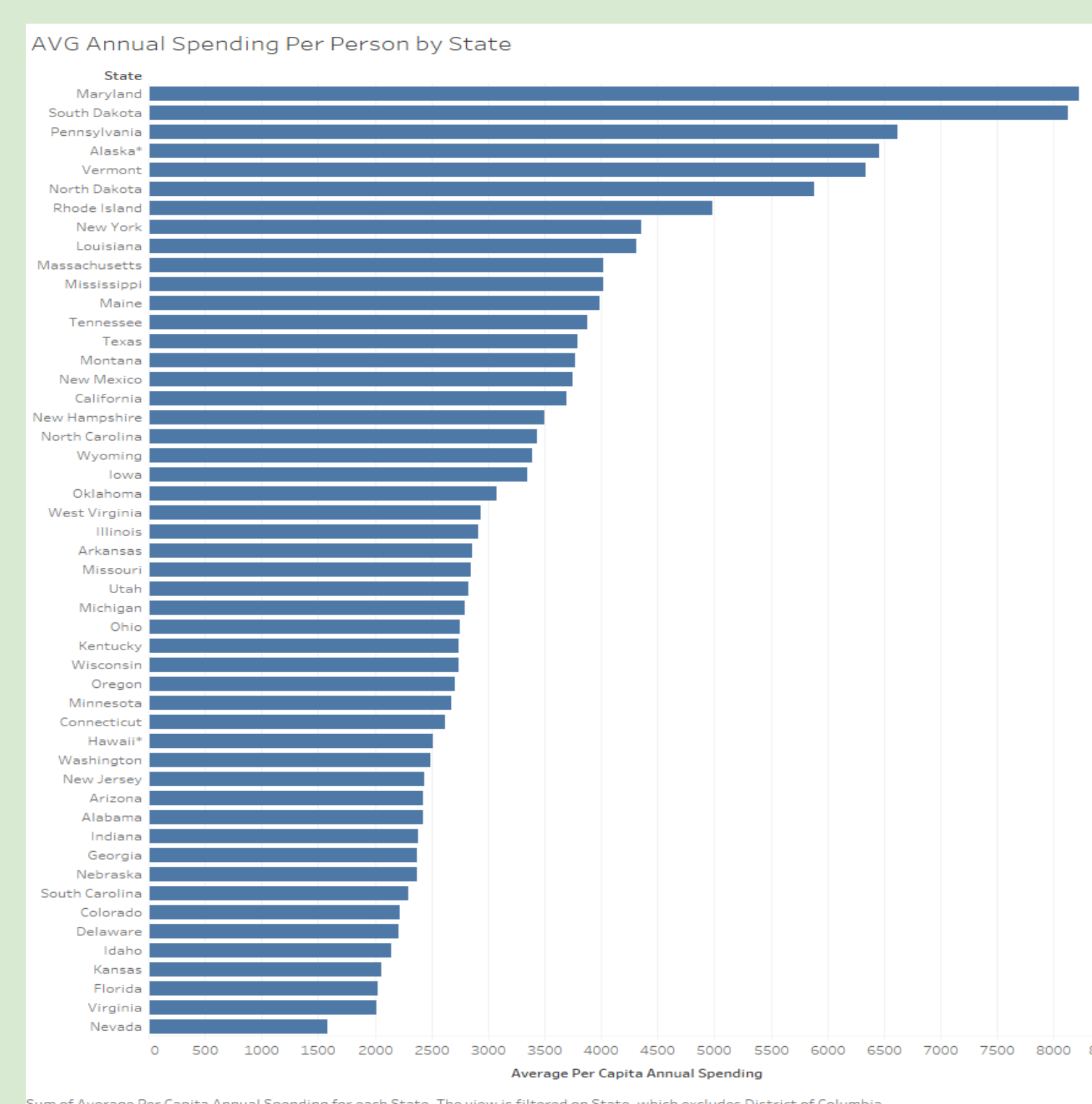
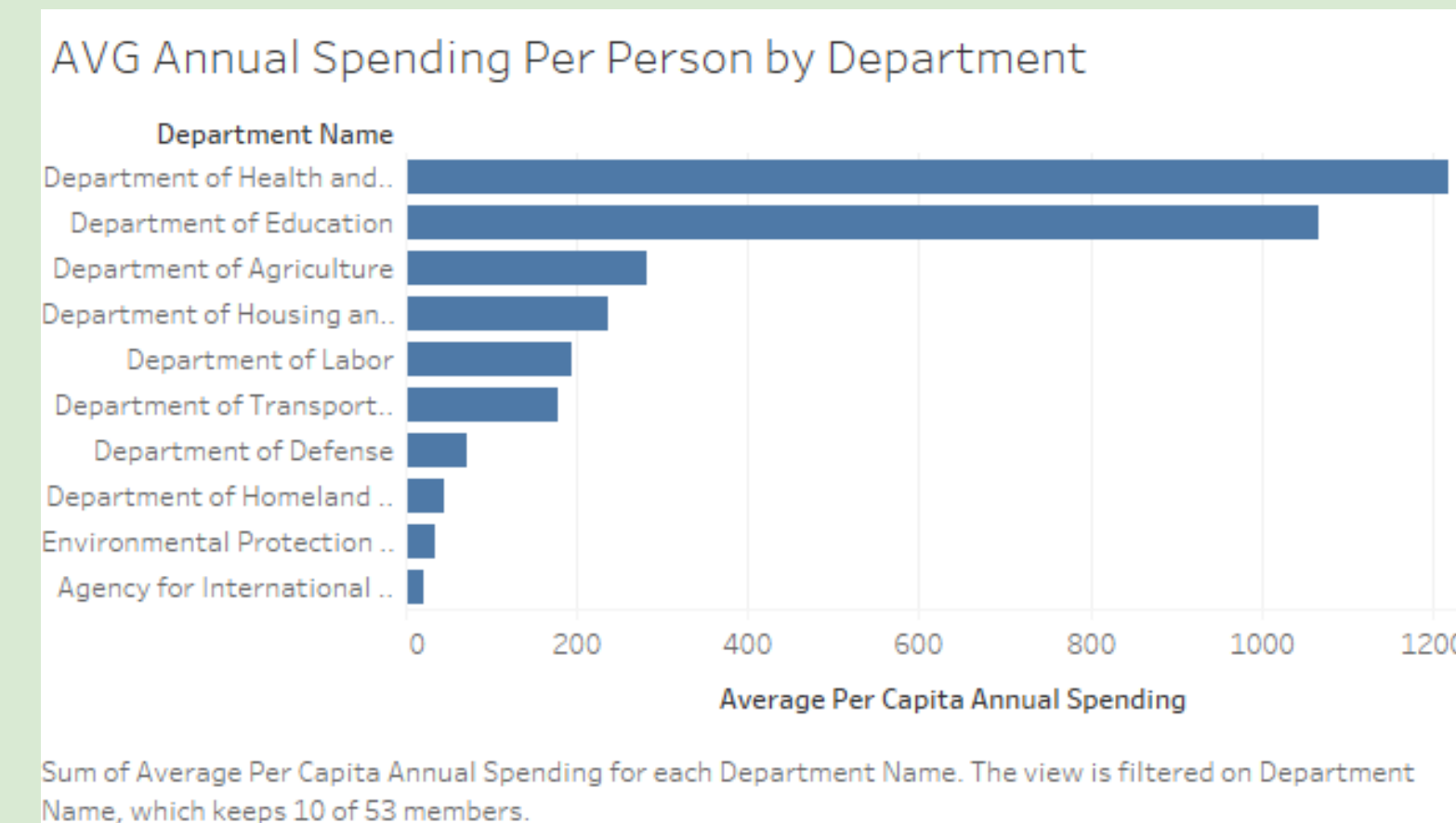
2008

- In 2008 the highest per capita income was CT with 61,232 (per capita spending was 457.05)
- The lowest was MS with 30,563 (per capita spending was 963.04).
- The average was 40138.18.
- In 2008 the highest educational funding per capita was SD that had \$5,310.82 (per capita income was \$41,311)
- The lowest was MT with \$210.28 (per capita income was \$35,448).
- The average was \$1,112.20

2009

- In 2009 the highest per capita income was CT with \$60,428 (per capita spending was \$391.67)
- The lowest was MS with \$29,801 (per capita spending was \$983.22).
- The average was \$38,672.98.
- In 2009 the highest educational funding per capita was SD that had \$5,513.74 (per capita income was \$39,524) per capita and the lowest was HI with \$355.38 (per capita income was \$41,473). The average was \$1,254.43.

EXPLORATORY DATA ANALYSIS



CLUSTER ANALYSIS

- The data does not reveal an obvious casual relationship as to how states actually received funding, which required us to dig deeper.
- Cluster analysis was chosen to be the model of choice because it would allow us to see the variation between the states income and the variation between the state's educational funding.
- A cluster analysis was performed separately for each year, 2005-2009.
- The clusters were formed based upon the following dimensions: Per Capita Personal Income and Educational Spending Per Capita.



2005 shows that those states with the highest incomes can be grouped into having the least amount of funding towards education. Cluster 2 has above average income but has more educational funding, than does cluster 3 with below average per capita income.

These results are very similar to 2005, where cluster 2 is represents the states with the 4 highest per capita incomes. Cluster 1 now represents the states that have above average per capita incomes and more educational funding than the third cluster, which has below average incomes per capita.

These results are very similar to 2005 and 2006, where cluster 2 represents the states with the 4 highest per capita incomes. Cluster 1 now represents the states that have above average per capita incomes and more educational funding than the third cluster, which has below average incomes per capita. The only change between this and past years is that Maryland is now in cluster 1 rather than cluster 3.

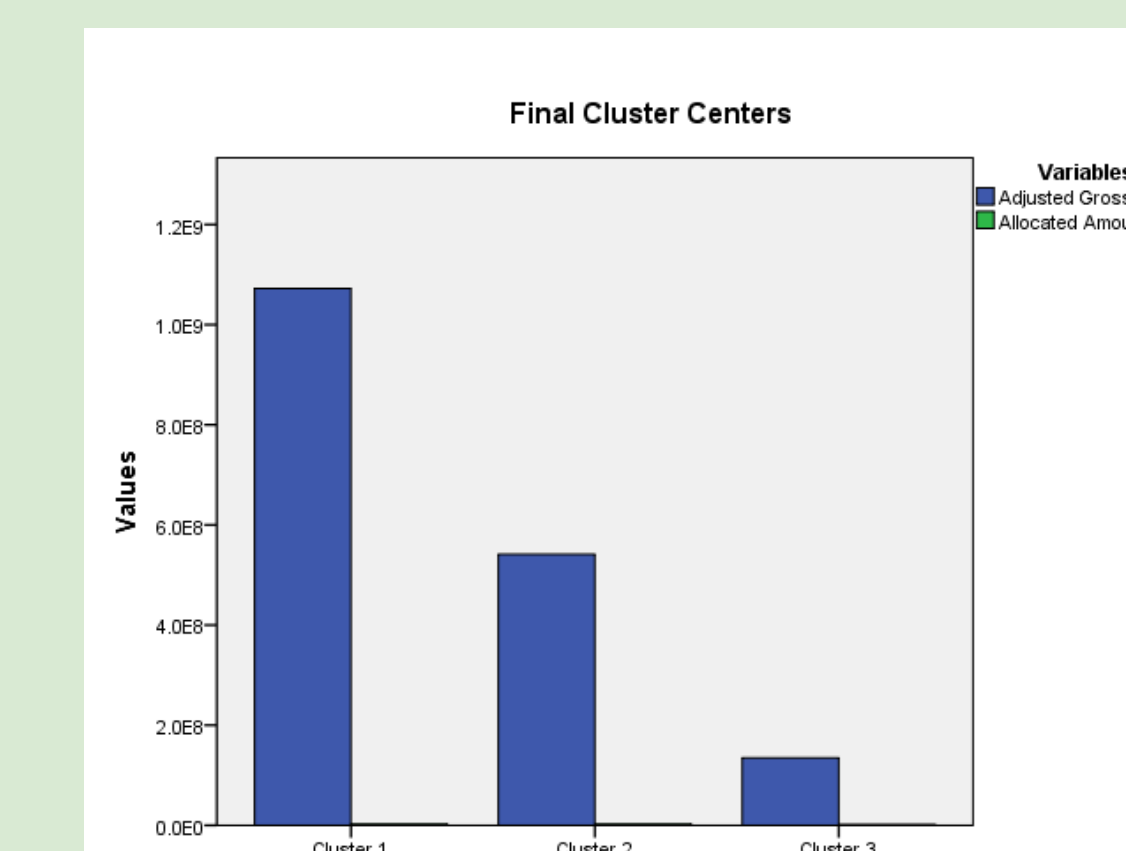
In 2008 the gap between the poorest states' education funding and the richest states' educational funding has decreased to within a couple of dollars with both groups still lagging behind the average representing above average income states. Four more states have joined the richest group, which is represented by cluster 1.

In a three cluster model the highest cluster only contains one state: CT, but the other two clusters show again this difference between states that have above average incomes vs. those that have below average incomes. If we look at only two clusters we lose this insight and the first cluster is that of states that have higher incomes and lower educational funding. The second is higher educational funding and lower income.

ZIP CODE LEVEL ANALYSIS

- Similar to the state level analysis we examined zip codes based upon income and funding allocated to education.
- We again ran a cluster analysis on the data from 2005 to see if the zip code data would follow the trend that appeared on the state level where higher than average incomes received more funding than did lower income states.
- Due to the variation in the incomes and funding per zip code outliers needed to be removed and therefore any zip code that had an extraordinarily high per capita Adjusted Gross Income (AGI) funds were removed, as were the zip codes with funding outside the normal range
- This left us with approximately 6,000 aggregated records of yearly spending data by the on CFDA programs offered by the Department of Education per zip code.

	Final Cluster Centers		
	Cluster 1	Cluster 2	Cluster 3
Adjusted Gross Income	1072476991	541294019	134610995
Allocated Amount	2339346	2320680	1587594



- This cluster shows the following trend:
- Those with the greatest AGI have the most funding allocated to their educational programs as opposed to those zip codes with the lesser AGIs.

RECOMMENDATIONS

- Increase educational funding for states/municipalities with below average Personal Income/AGI:
 - In our statistical analysis of the data we found that those states with higher than average incomes are receiving the largest amounts of educational funding. We recommend that this be changed to the reverse where those states that have lower than average incomes receive more educational funding.
 - The states that were continuously low in education only have about 5% of their funding allocated to education, where as states such as South Dakota, who had the highest per capita education spending had 12% on average.
 - In addition to this education is competing with health and human services for the largest amount of funding.
- Creating a better category system with machine learning algorithms.