

Namita Saxena

Thursday, May 8, 2014

Course: 30211.(002) Introduction to Data Analysis

Introduction:

Pupils ordinarily spend from 6 to 8 years in elementary schools, which may be preceded by 1 or 2 years in nursery school and kindergarten. The elementary school program is followed by a 4- to 6-year program in secondary school. Although elementary school runs from kindergarten through either grade 5 or grade 6, I chose to study til 6th grade. Why I choose this topic, the reason behind this topic is my toddler son who will start going elementary school soon.

***Questions that interest me* :**

- * Analyzing past trends of enrollment in public elementary schools, is there any change in trend over the past few years?
- * Is pattern of enrollment differ in rural and urban area?
- * What's the impact of teacher-pupil ratio on enrollment?
- * Distribution of elementary students enrollment with states?

2. Sources of Data:

National Center for Education Statistics (NCES) [<http://nces.ed.gov/>]
(<http://nces.ed.gov/>)

USA state population [<http://www.demographia.com/db-statedens.htm/>]

3. About the data:

I analyzed data-sets of 5 public school obtained from the given source from year 2007-2012. Each dataset has its own documentation and over 250 attributes. I cleaned them individually (as their column, attribute names and formatting were different from each other). The datasets provide essential background information on the education environment in the states including students enrolled in different grades, school type, total number of teachers.

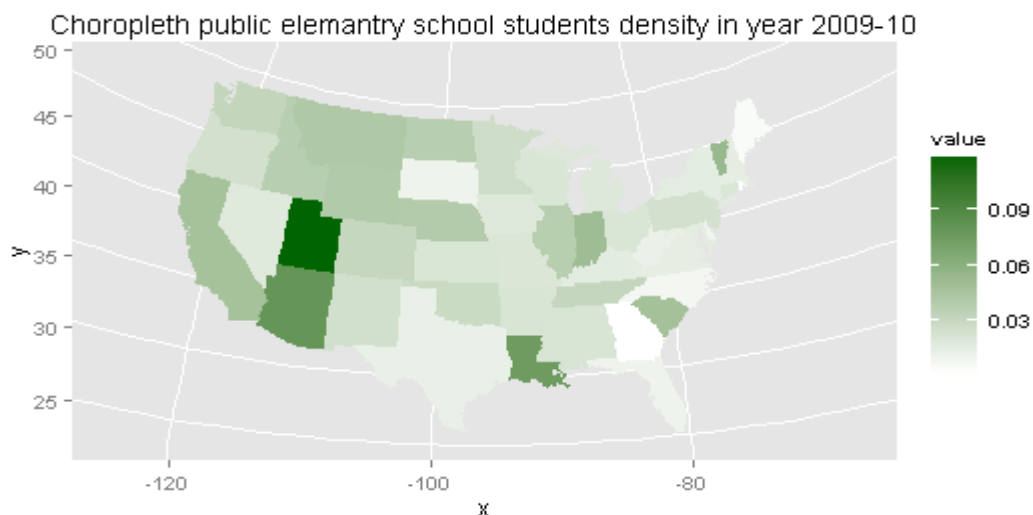
4. Data manipulation:

The dataset were in text format and I loaded it in R. After that I checked them for NA (missing values) and duplicated values. I also put it in right format for analysis after

checking their type, removing unnecessary rows and columns. The columns of total enrolled students in each grade has missing values (0,-1 for numeric values: missing data,-2 for numeric values: not applicable). After cleaning them I take subset of each dataset for KG to Grade6 (elementary school dataset). I merge them in one by state to analyze it further. The variable of interest are total number of teachers, total number of students, number of schools, number of elementary students, number of schools in urban and rural areas, etc.

5. Analysis:

Following graph shows the distribution of elementary school children in year 2009-2010. It is clear from the choropleth that Utah has highest



student density in 2009-2010.

The following graph has

`summary(linearmodel)`

Call:

```
lm(formula = Total_elemantry_enstudents10 ~ ratio3, data = p_t_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-1955.03 -127.48 -25.33 102.93 1762.50

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	36.392	4.261	8.541	<2e-16 ***
ratio3	23.572	0.275	85.705	<2e-16 ***

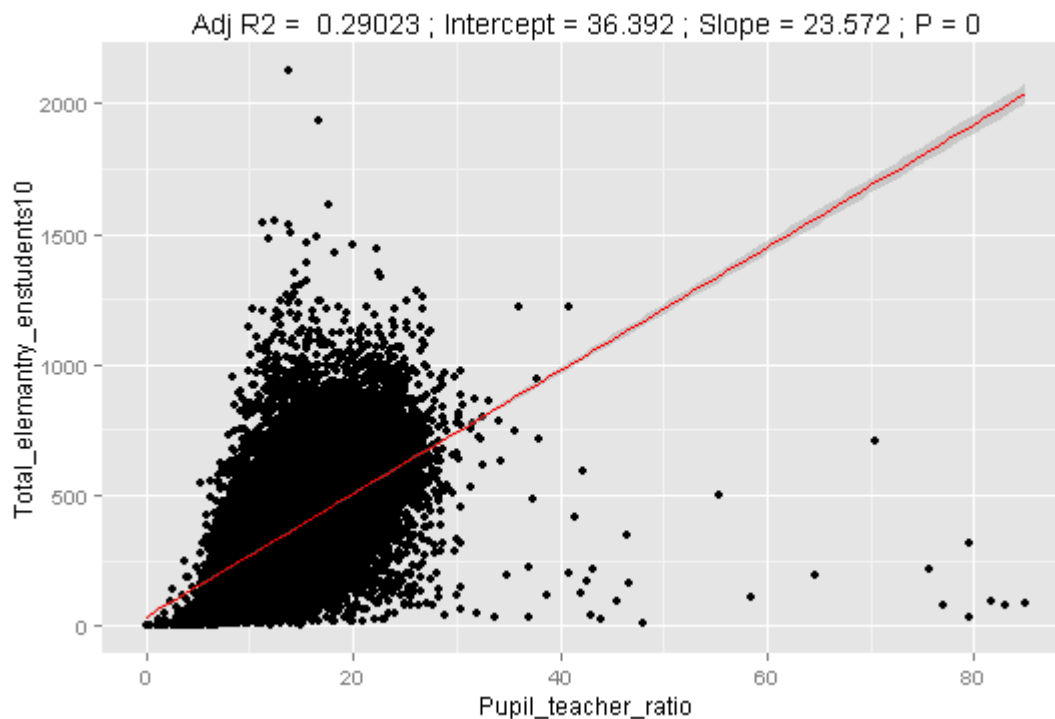
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 192.8 on 17960 degrees of freedom

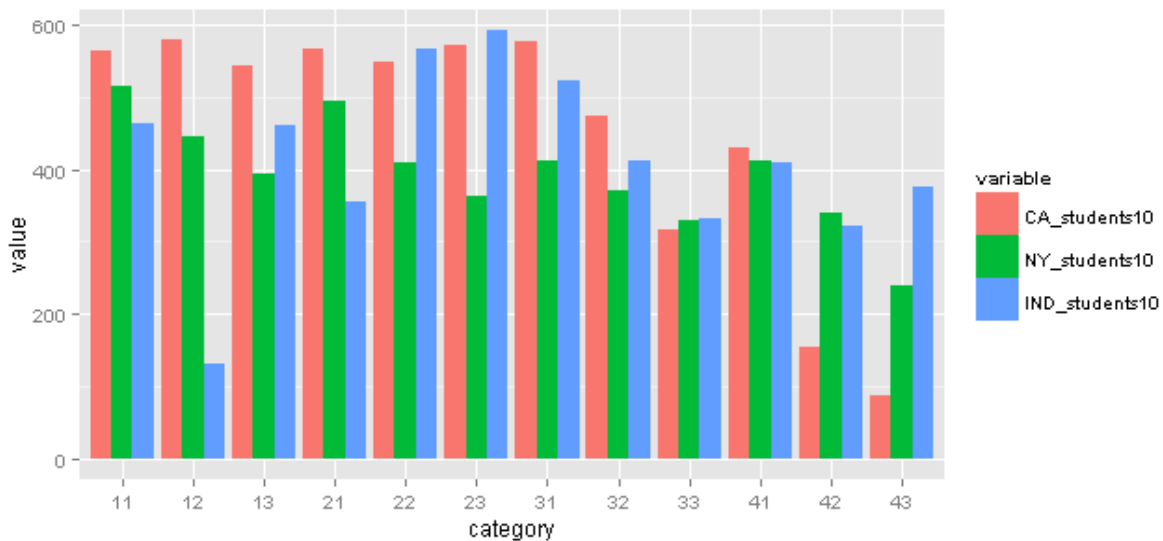
Multiple R-squared: 0.2903, Adjusted R-squared: 0.2902

F-statistic: 7345 on 1 and 17960 DF, p-value: < 2.2e-16

The p value is very small that shows linear model is significant for pupil teacher ratio. I have plotted the distribution for all states.



The following graph shows the comparision between California, Indiana and New York elemantry school Pupils. The description of number is as follows:



Locale codes are divided into four main locale types (city, suburb, town, and rural) and each of the four locale types has three subtypes (large, midsize, and small for city and suburb locale types and fringe, distant, and remote for town and rural locale types).

11 = City, Large with population of 250,000 or more.

12 = City, Midsize with population less than 250,000 and greater than or equal to 100,000.

13 = City, Small with population less than 100,000.

21 = Suburb, Large with population of 250,000 or more.

22 = Suburb, Midsize with population less than 250,000 and greater than or equal to 100,000.

23 = Suburb, Small with population less than 100,000.

31 = Town, Fringe

32 = Town, Distant

33 = Town, Remote

41 = Rural, Fringe

42 = Rural, Distant

43 = Rural, Remote

6. Conclusion:

A high proportion of 5- to 17-year-olds indicates a relatively large requirement for education expenditures

and services. Such a population profile poses a different set of public service requirements, such as a need for more hospitals or may be more libraries and park. State and local governments frequently use demographic data to help develop strategies for providing public services. The proportion of public school enrollment in elementary schools is a useful indicator of the potential for growth in secondary enrollment. Since elementary education is normally less expensive to provide than secondary education, it can also serve as one of the measures of current and future relative demand for more expensive secondary education services.

Acknowledgment:

I would like to thank my instructor Ram for his valuable suggestions especially for graphs.