

## Assignment 4: Visualizing Information

**Objective:** In this assignment, you'll visualize information from a data table exported by xlsx or csv files. You can also create your own data table. In this assignment, you'll work with multivariate data. The number of variables in the data table should be more than five. Each student should use their own unique data table for this assignment.

**Specification:** You'll visualize the data using Parallel Coordinates Plot as discussed in the class. You should be able to draw relationships among different variables. You should also be able to cluster the observations. The following example (Table 1) shows a data table (education) with the following variables: state, reading, math, writing, percent\_graduates\_sat, pupil\_staff\_ratio, and dropout\_rate.

**Table 1: Education Data Table:**

education[2:52,]							
	state	reading	math	writing	percent_graduates_sat	pupil_staff_ratio	dropout_rate
	Alabama	557	552	549	7	6.7	2.3
	Alaska	520	516	492	46	7.9	7.3
	Arizona	516	521	497	26	10.4	7.6
	Arkansas	572	572	556	5	6.8	4.6
	California	500	513	498	49	10.9	5.5
	Colorado	568	575	555	20	8.1	6.9
	Connecticut	509	513	512	83	6.6	2.1
	Delaware	495	498	484	71	7.9	5.5
	District of Columbia	466	451	461	79	6.3	7.1
	Florida	497	498	480	59	8.1	3.8
	Georgia	490	491	479	71	7.0	4.6
	Hawaii	479	502	469	58	8.3	5.4
	Idaho	541	540	520	18	10.0	2.6
	Illinois	588	604	583	6	9.9	4.0
	Indiana	496	507	480	63	7.5	2.7
	Iowa	610	615	588	3	6.8	2.3
	.	.	.	.	.	.	.
	.	.	.	.	.	.	.
	.	.	.	.	.	.	.

You need to use **Parallel Coordinates Plot** to find relationships among different variables for different observations. You need to use 'clustering' technique to cluster the observations to find a better relationship among different variables among different observations. Figure 2 shows the parallel coordinates plot for Education Data Table.

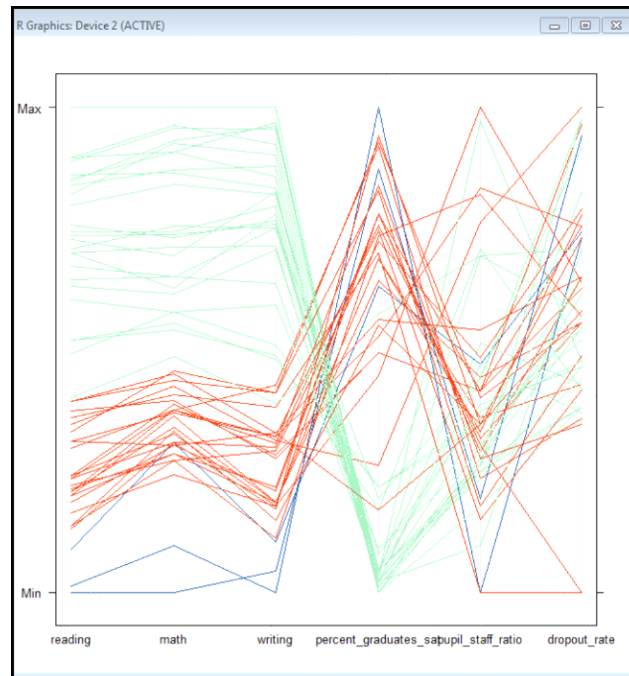


Figure 2: Parallel Coordinates Plot (with Clustering) for Education Data Table (Table 1).

Next, you need to plot the map of the states that helps you to further analyze a particular variable across the states. Figure 3 shows how **reading** ( a variable in Education data table) varies across the states. Use map to show the variations of at least three variables across the states.

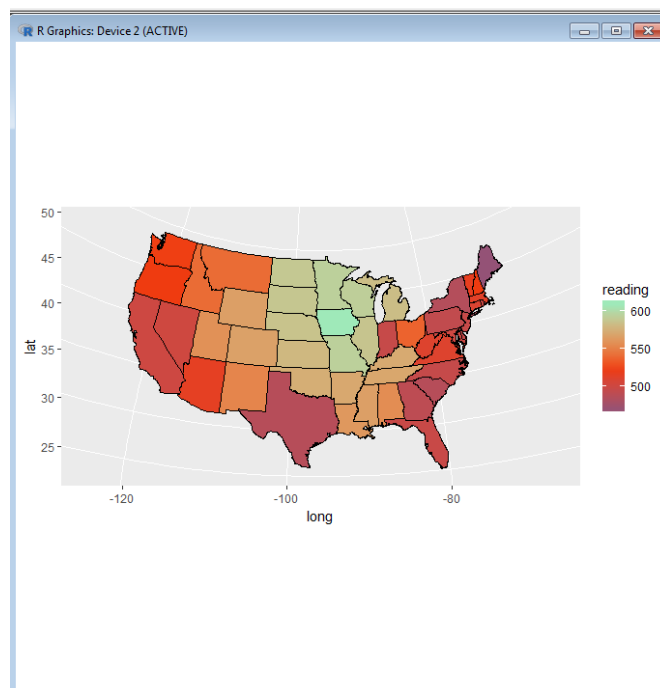


Figure 3: State map showing the variation of a variable across the states.

### **To do lists:**

1. Export data using excel sheet or csv files.
2. You may need to clean the data; adjust rows, columns or order the data table using a particular variable.
3. Use Parallel Coordinate Plot to plot the data for different states;
4. Use 'clustering technique to cluster the data
5. Use map to visualize the variation of a variable across the states; Use map to visualize the variations of at least three different variables across the states.

### **Submission:**

Each student should have his own unique data table. **No two submissions should be the same.**

The submission should consist of R codes that you will use for visualizing the data table. You also need to write how your work is able to answer different questions, i.e,

- How you will draw relationships among different variables in a particular observation,
- How you will establish relation among the variables for different observation.
- Does clustering help in visualizing information?

You need to write down about the clarity of your presentation.

So, your final submission will be a zip file consisting of the following:

1. Data Table ( csv file or excel sheet);
2. A pdf file named 'R.pdf' for all R codes used for the assignment;
3. A pdf file named Assignment 4.pdf that explains your work and describes different steps as found in **To do lists**.

Name the zipped file as follows: **LastName\_FirstLetter ofFirstName\_Assignment4.zip**.

Submission deadline is **Friday, December 1**.

This assignment carries 15% of the course evaluation.