Lab 02 Parallel Coordinates with HiPlot

COMP7507 Visualization and Visual Analytics

June 21, 2024

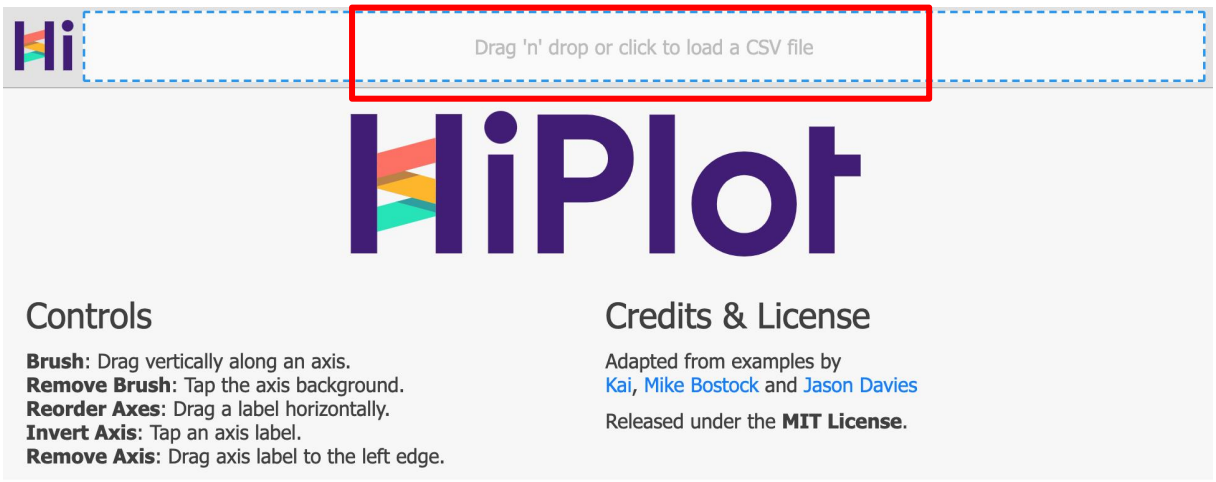
Brief introduction to HiPlot

HiPlot is a lightweight interactive visualization tool developed by the Facebook research team. The main aim of this tool is to help AI researchers discover correlations and patterns in high-dimensional data using parallel plots and other graphical ways to represent information. You may want to check their github link here: https://github.com/facebookresearch/hiplot

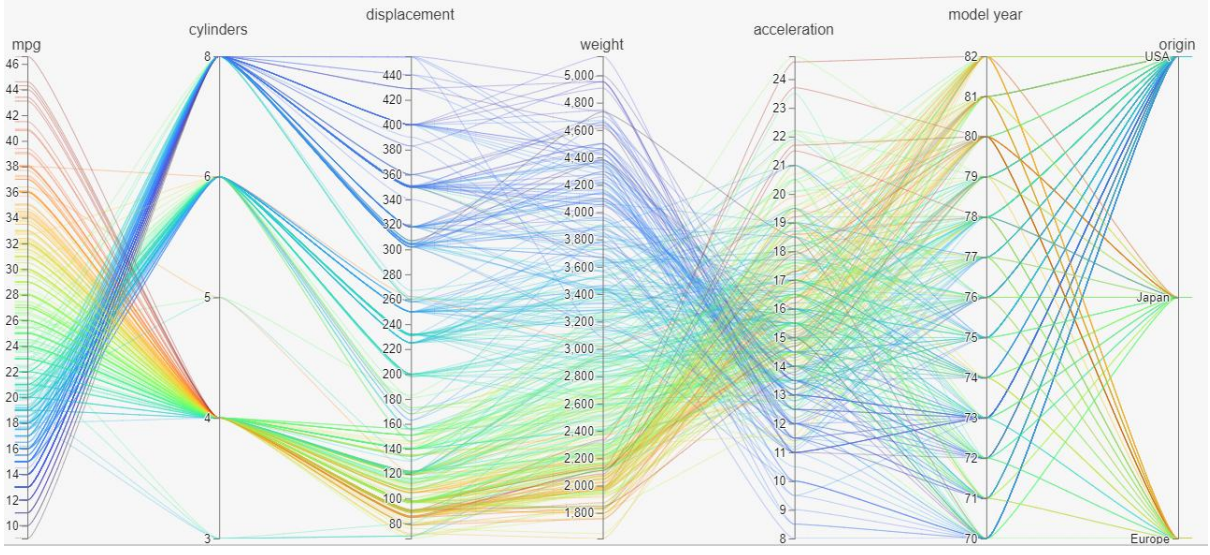
In this tutorial, you will play around with the HiPlot Web version to know how many features it can perform, and get familiar with parallel graphs at the same time.

Parallel Coordinates using HiPlot

1. Download "**CarData.txt**" from Moodle.
2. Open HiPlot's Web Version: https://facebookresearch.github.io/hiplot/_static/hiplot_upload.html
3. Upload the data file to HiPlot Web.



4. Now you are shown with the auto-generated parallel coordinations plot with car information.



5. Scroll down the webpage, there is a table showing the original data.

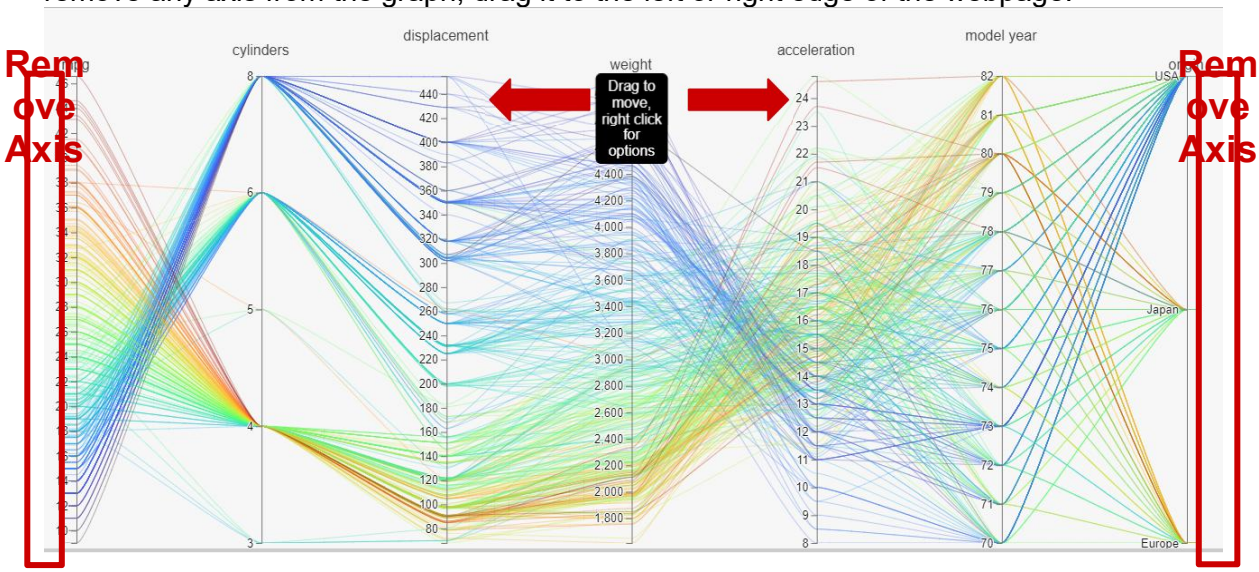
entries

	uid	from_uid	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin
■	0	null	18	8	307	130	3504	12	70	USA
■	1	null	15	8	350	165	3693	11.5	70	USA
■	10	null	15	8	383	170	3563	10	70	USA
■	100	null	18	6	250	88	3021	16.5	73	USA
■	101	null	23	6	198	95	2904	16	73	USA
■	102	null	26	4	97	46	1950	21	73	Japan
■	103	null	11	8	400	150	4997	14	73	USA
■	104	null	12	8	400	167	4906	12.5	73	USA
■	105	null	13	8	360	170	4654	13	73	USA
■	106	null	12	8	350	180	4499	12.5	73	USA

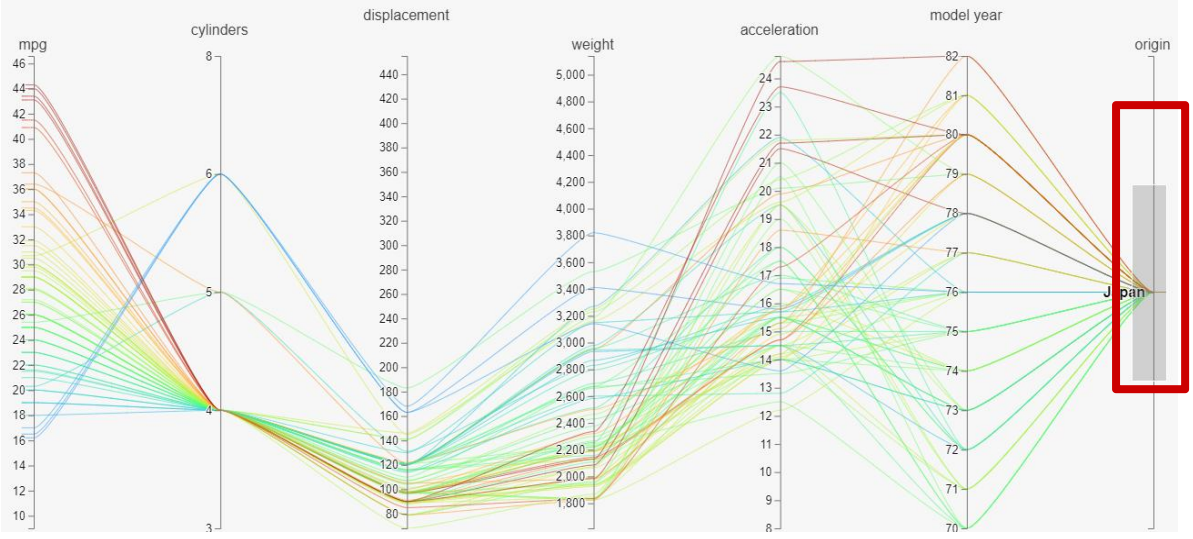
Showing 1 to 10 of 398 entries

Previous 1 2 3 4 5 ... 40 Next

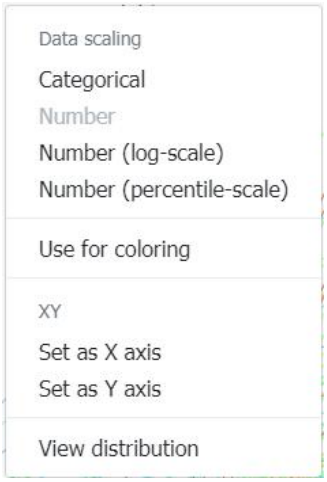
6. To play with the graph, there are some basic operations that you need to know:
   a. **Invert Axis**: Tap an axis label to invert the order.
   b. **Reorder Axis**: Drag any label horizontally to re-order the graph. The ordering of parallel graphs is important to find patterns and relationships between attributes. If you want to remove any axis from the graph, drag it to the left or right edge of the webpage.



   c. **Brush**: By dragging vertically along an axis, you are able to select some data within the brush's area. To clean the existing brush, tap the axis background.

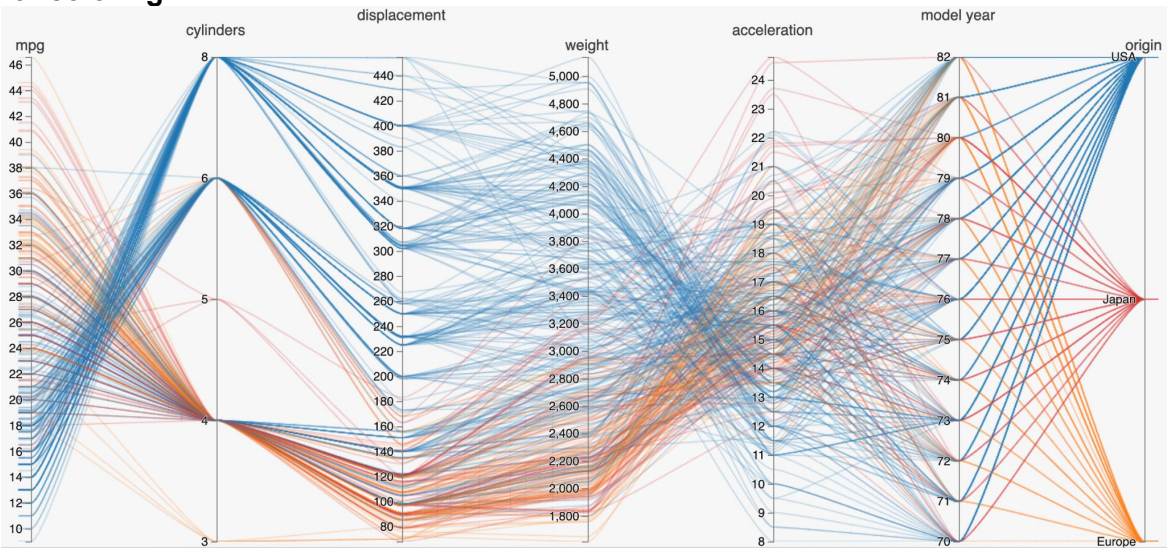


7. Right-click on the label of any axis, you will see a menu that enables you to perform many tasks:

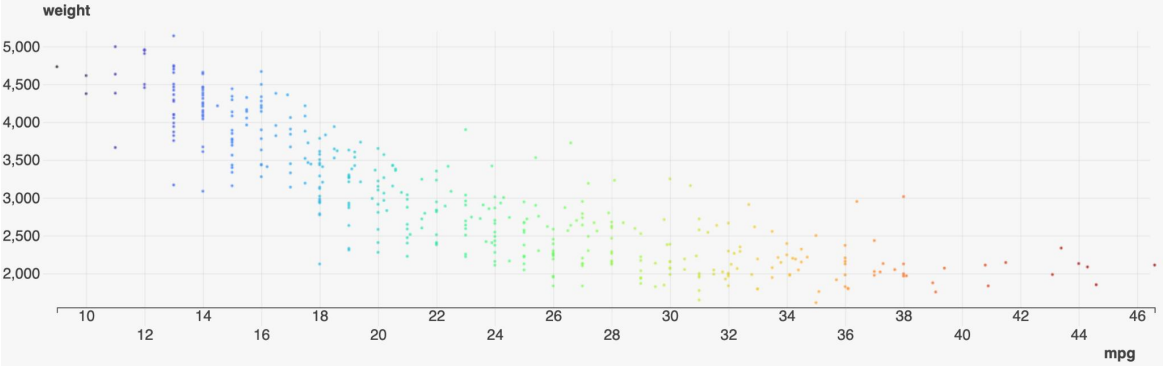


a. **Data scaling**: Switch your selection in the data scaling menu between Categorical, Number, Number(log-scale) and Number (percentile-scale), the corresponding axis will change its scale accordingly.

b. **Coloring and clustering**: By selecting "**Use for coloring**", you will make the color of each line in this parallel plot change according to its data on this axis. Apply this function on a categorical axis and you can do clustering. Let's try with the "**origin**" axis. As you can see from the last column, there are 3 origins: Europe, Japan, USA. To visualize the data in different origins with different colors, right-click the "**origin**" label and select "**Use for coloring**".

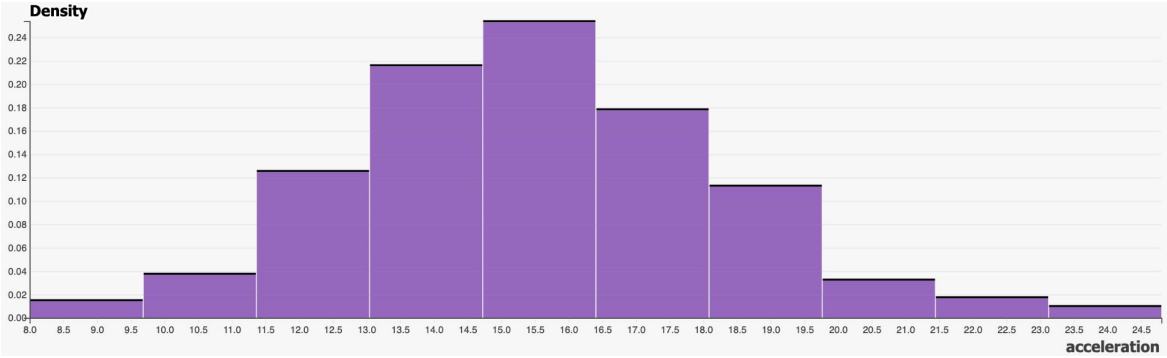


c. **Scatter plot**: The function of "**Set as X/Y axis**" is to create a scatter plot for two attributes. Scatter plots help us to clearly see the relationship between two variables. Right-click the "**mpg**" label and select "**Set as X axis**" to set "**mpg**" as the X axis of the scatter plot. Then right-click the "**weight**" label and select "**Set as Y axis**" to set "**weight**" as the Y axis of the scatter plot. Now there is a scatter plot generated below the parallel chart, which shows the inversely proportional relationship between "mpg" and "weight".



d. **Histogram**: The "**View Distribution**" function generates a histogram of the axis, allowing you to view its data distribution. Right-click on the "**acceleration**" label, select "**View**

Distribution". You will find the histogram below the scatter plot, which shows that the distribution of acceleration follows a normal distribution.



Some tasks for you

1) Create a parallel coordinates plot of the iris data set **("IrisData.txt" from Moodle)** and see if you can identify the 3 classes of iris, without seeing the "Species" variable.
2) Upload the screenshot of task 1) to Moodle before **Oct. 8, 2024**. An Example is shown as below: