

Assignment 04- Parallel Coordinates

Initial design thoughts

- Views
 1. Vertical axes
 - Not a line, but a thin long rectangle
 - Color coded (assign to colors to min & max and smooth color transition from min to max) (Not sure if it's a good idea to color axis as it might interfere with line colors)
 2. Label
 - All attributes shown
 - With min & max values for non-string attributes (i.e., int / float)
 3. Ticks
 - Int/Float attributes – Show 5-6 ticks at uniform intervals. Might result in unpleasant floating point tick values. Consider using nearest unit/tenth of tick value.
 - Unique string attribute – Attributes with unique string for each item.
 - If the item count \leq threshold (set at 60), show string values as ticks.
 - Else, ticks are hidden by default and only shown on mouse hover.
 - Group string attributes – Attributes whose values repeat among a set of values for all items. Will identify first such attribute of the table as a grouped column.
 - If groups count \leq threshold (set as 60), show string group values as ticks.
 - Else, ticks are shown by default and only shown on mouse hover.
 4. Line color
 - Is data grouped for any attribute?
 - Yes, choose a continuous color map as a transition between two base colors for first and last values of the first attribute.
 - Vary saturation or luminescence or one of RGB colors.
 - No, choose a categorical color map for groups. For each item, use the group color for line.
 5. Progressive rendering
 - If item count $>$ threshold, use progressive rendering for smooth transition
 6. Bundling
 - If item count $>$ threshold, use bundling to reduce clutter.
 - If grouped, can bundle group items
 - Or can bundle based on similarity/proximity
 7. Aggregation
 - Use boxplots to aggregate and show median, lower & upper quartile and lower & upper whisker for each attribute

- Interactions
 1. Filtering – should be able to brush multiple attributes axes and data to be filtered accordingly
 2. Reordering axes – should be able reorder INT/FLOAT axes
 3. Invert axes – should be able invert INT/FLOAT axes
 4. Magnifying lens – should show a magnifying lens for string attributes values when mouse hovers on axis (for the ones for which we couldn't show ticks).
 5. Show/Hide box plot on axes.

Project 1

Handling inconsistent data

I **did not** modify the data to handle missing attribute values for some items. Instead, while parsing, if I encounter a missing attribute value, I add a 0 for that attribute and continue forward.

Handling different data types

For each attribute, I store the type in the meta data that defines if the attribute is of type INT or FLOAT or STRING. In case of STRING type, I also save a subtype that stores UNIQUE or GROUP.

Attributes

While parsing,

- For a STRING and GROUP attribute, I identify and save all possible group values
- For a INT/FLOAT attribute, I calculate and save min, max, median, lower & upper quartile and lower & upper whisker values.

Interactions

1. Filtering – User can filter one/more axes by (brushing) clicking close to the axis and dragging the mouse up/down. The darkened region is brushed. I chose this mechanism as this is very intuitive and easy to grasp. We had witnessed this in the scatter plot too. To assist the user, as the mouse moves up/down, the brush region changes and filters lines correspondingly.
2. Reordering the axes – User can click on an attribute label, drag it and drop it on another label to reorder axes. This is also very natural as we are used to drag & drop objects to reorder even in the physical world. Once the dragged attribute lands in its new position, the other attributes adjust accordingly to accommodate the change. And the dataset, i.e., lines change immediately to reflect the change.
3. Inverting the axes – User can click on an attribute label to just invert the axis. Once the axis is inverted, the min-max labels, ticks and lines passing through that axis invert accordingly. I could have made inversion action upon click of min-max labels. But I chose to do it on attribute label because it's a transformation on the axis which is better represented by (name) label than by min-max labels.
4. Coloring – I also provided three different configurations for line coloring.
 - a. Single – A single color is best for visualizing a non-grouped dataset. Useful when item count is relatively less.

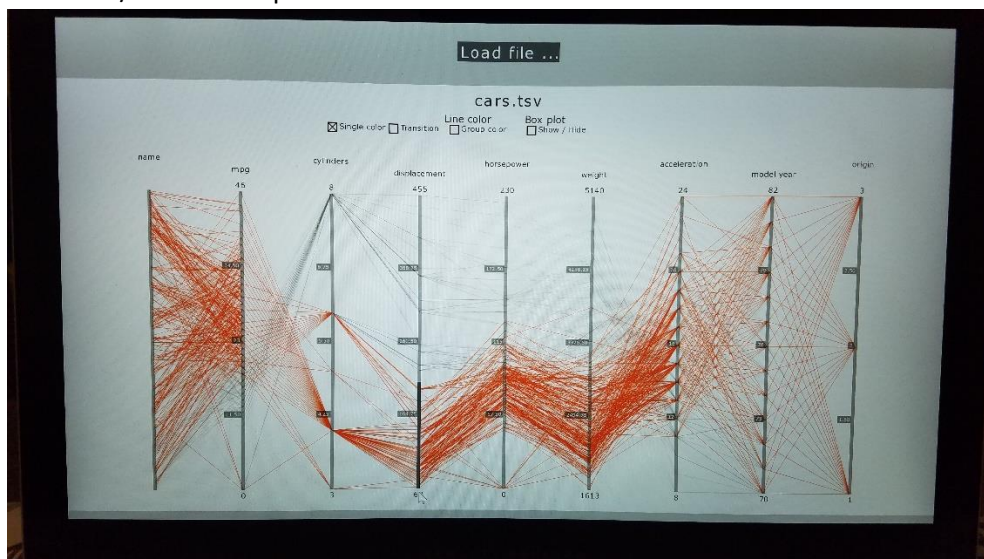
- b. Transition – Line colors have a smooth transition from one base color to another. It is useful when visualizing a huge, non-grouped dataset. Can identify items based on bands of color and is relatively easier to trace lines than a single color.
 - c. Grouped – Line colors are decided on the group to which they belong. Groups follow categorical color map based on HSB. It is extremely useful when visualizing grouped dataset as it is easier draw inferences and trace lines to groups because it is not always possible to trace lines to actual items because of the abundance of lines. Also, it is extremely good at highlighting hidden relationships between groups and attributes. (Will explain with an example in next section)
- 5. Box plot – Box plots show summarized information on the attributes and can be quite useful. I've intentionally kept box plot hidden by default as user needs a little time to get used to visualization and showing box plot at first can be overwhelming. Once used to the viz, user can click 'show' box plot and get an idea of variance in attribute values.
- 6. Magnifying glass – When there are many items, it's impossible to show all their string values at once effectively. It would only clutter the UI. Therefore, I decided to keep them hidden by default. And show them, a few at a time in a magnifying glass when user hovers over the axis. If string values don't fit in the small hovering window, complete text can be viewed at the top.

Some of the inferences from datasets

1. I used parallel coordinates with box plots to understand a variance in a few attributes of cameras. I observed that storage included in cameras has very little variance, meaning, most competitors have similar storage options which is on the lower end. Also, I observed that Normal focus range has more variance whereas Macro focus range has less variance.

Usage guidelines

1. Click on the "Load file ..." button to load dataset.
2. Once loaded, you'll see the parallel coordinates being drawn. By default, it uses the single line color and box plot is not shown. Use the checkboxes at the top to change to color configuration or enable/disable box plot.



3. To invert the axis, click on the axis label.
4. To reorder the axes, click and drag axis label 'A' and drop it (note: exactly) on the axis label 'B' where you want 'A' to be inserted.
5. Box plots ex:

