CMSC 636 Data Visualization

Anudeep Nallamothu

October 17, 2016

1 Introduction

This is a document about design enhancements on Parallel coordinates(PC) visualization constructed upon tumor data set. Three problems are stated for Parallel coordinates on this particular dataset [1]. Among these three problems, our group tried to address second problem, which is about cluttering. Following section discusses why cluttering happened and our proposed solution for it.

Group Names: Anudeep Nallamothu, Kishan Pawar, Harsh Vardhan Vashista
Group Number: 4
Group Idea: Hierarchical Edge Bundling

2 Design

2.1 Dataset

Dataset consists of 36 columns(which we consider as dimensions). These dimensions are scaled to one axes each. There are 108 rows in total. 9 therapies which are sub categorized under 2 organ types.

2.2 Task

- One of our task is to find insights from data like maximum value, minimum value, median, etc.
- Effective representation of outliers.
- Finding out relationships between different dimensions.

2.3 Problem

As stated above we are addressing cluttering problem in the PC visualization. 36 * 108 values lead to too many mappings between axes leading to overlapping among them. This will create the problem of overlapping lines. This can be solved to some extent by creating

2.4 Mapping and Encoding choices

2.5 Design Choices

Hierarchical Edge Bundling[1]
In this design, 360 degrees is divided by n which number of dimensions to represent the complete data. Primary reason to choose this plot is to reduce cluttering and improve adjacency relationship representation(which later prove to be ineffective).

Box Whisker Plot
In this design, each Box whisker represents a tumor with specific therapy sample. Primary reason to choose this plot is to reduce cluttering and improve information insights from the data. Each box summarizes different kinds of metrics about the data(like median, 75 and 25 percent quartile).

2.6 Good and Bad

I will discuss good and bad points in Hierarchical bundling as well as Box whisker plot.

Hierarchical Bundling

- Advantage with this approach is clear insight about adjacency relationship.
- Disadvantage is (a) more cluttering. (b) Less interactivity like in parallel co-ordinates.

Box Whisker

- Advantage with Box whisker is (a) data can be summarize using median line and quartile data. (b) outliers can be clearly represented using circles which can not created unneccesary cluttering. (c) Comparison of dataset dimensions can be easy
- Disadvantage is (a) Relationship among different dimensions cannot be depicted(although the visualization in the assignment have relationships among them....in general box whisker plot this is not possible).

2.7 Issues

- To be specific, our idea(hierarchical bundling) is very bad idea. Although it represents the data completely, it failed in addressing basic problem of reducing cluttering problem. Instead of that it increased cluttering problem by grouping all the relation lines inside circle.
- Representing 32 dimensions is a very congested. If the number of dimensions increase more than this number, it will be more cluttered and congested.
- In basic Box whisker plots, representing relationships is not possible.

2.8 Future Fixes

- Using hierarchical bundling is a bad idea. Reasons are explained above.
- Box whisker plot works better than for this dataset. To represent relations, lines are drawn between set of 12 box whisker plots to other.

References

[1] Danny Holten *Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data.*

[2] Figure Link
`https://seeingcomplexity.wordpress.com/2011/02/05/hierarchical-edge-bundles/`

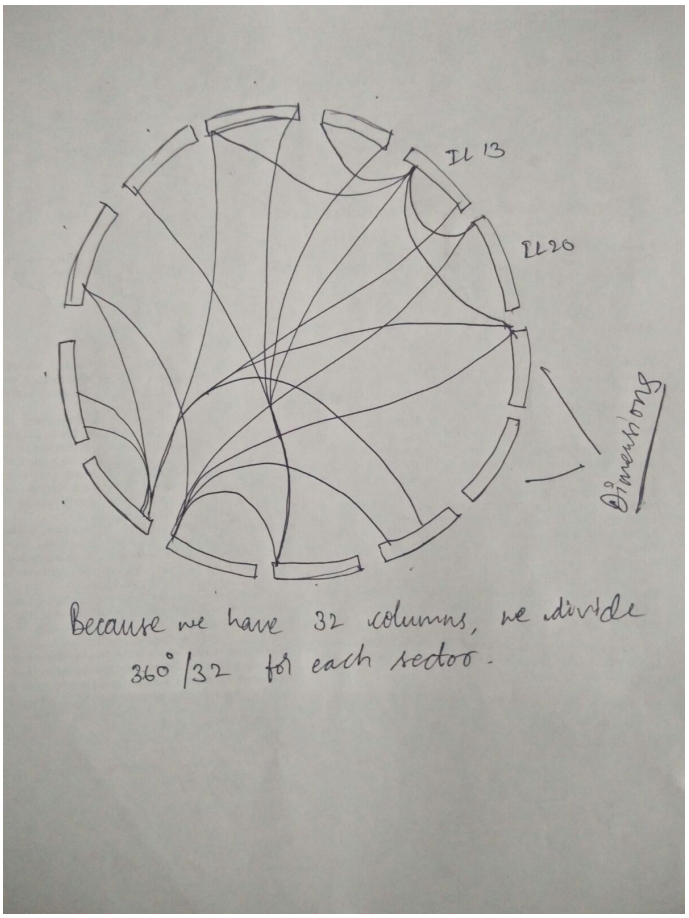

Figure 1: Sketch of Hierarchical Bundling

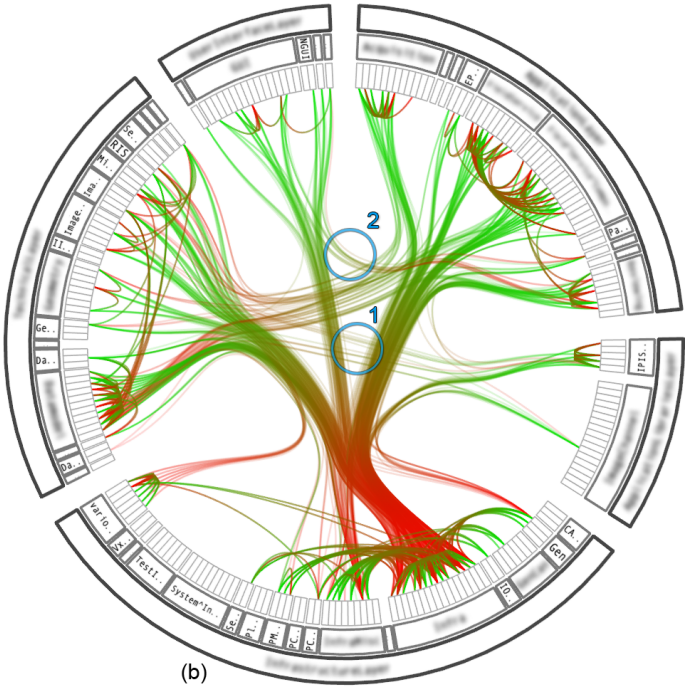

Figure 2: Sample image of Hierarchical Bundling[2]