Group:
Harsh Vashistha (me)
Sushant Chaudhari
Sushant Athley

Chord diagrams:
Initially, we went ahead with Chord diagrams which is solution 4 in discussions during the class. Chord diagrams are radial plots which also uses a technique of hierarchical edge bundling.
The problems which we tried to address was problem 1 ('Parallel coordinates do not let one see the distribution easily') and Problem 2 ('clutter').

The main idea for going with Chord diagram was that it helps in identifying relationships easily among different attributes than horizontally laid out parallel coordinate plots. The argument for this was that, the attributes arrangement within these plots is radial and a direct link can be established between each one of them. Then there was the problem of clutter which was targeted by using edge bundling to reduces the number of independent links.

Some of the pros and cons of the technique are as follows:

Pros:
- Chord diagram can reduce clutter between generated by links between different attributes by using edge bundling techniques.
- Chord diagrams can also accommodate bar charts and line charts in the plots periphery for more detailed analysis.
- Chord diagrams are very effective to reduces clutter when there are large number of links or relationships between attributes. These plots work well with adjacency data wherein there might exist a link between two or more attributes but necessarily with all of them. This is evident in infographic published by The New York TImes on 'Close-Ups of Genome,Species by Species by Species' on Jan'2007.
- Chord diagrams can be used on most datasets which displays relationship between attributes

Cons:

- Although Chord diagrams work great with adjacency data, they don't work well when there exist a relationship between each and every attribute within a plot and with large number of attributes. This is mainly because even with bundling of edges, there remains large number of crisscrossing lines in the center of the plot which are not easy to distinguish.

Box and whisker:

Box and whisker plot is the general consensus of the class. This solution tries to visualize distributions rather than direct relationships. It allows for easy outlier detections, general distribution of data and helps in removing the unnecessary edges or clutter. Although the cost of removing those edges comes with abstracting information into a distribution but that can be overcome by overlaying box-whisker plot on top of the original parallel coordinate plot. Another problem with box and whiskers idea is that the horizontal span of the graph is very large. Therefore, it hides large parts of the graph which become visible only by scrolling to it. This makes it hard to follow relationships which are far apart in the plot.

Conclusion:

We went out with chord diagrams considering all the pros of the for the method but it turns out chord diagram is not a very good choice for the problem set because there are too many attributes to the data with each having a relationship with other. This makes the solution weigh in more on negative side than on positive creating a solution that introduces more clutter in the center of the plot if not on the periphery. On the other hand, box and whisker solution, helps with outlier detection as well as solves clutter problem to some extent but it also comes with its own problems of excessive horizontal span and difficulty with following a path.

Future Improvements:

We as a group directly went ahead looking for a visualization and tried to fit in current dataset and plot into our visualization. Doing this is inherently is inherently wrong because we also picked up the good and the bad of the original solution without improving on much. In future, we would try to create exact problems and tasks first. Explore the dataset, its features and drawbacks. Then from there move towards searching for a solution.