

36-350: Data Mining

Handout 11

October 1, 2003

Comparing cars via parallel coordinates

Parallel-coordinate plot—Each object is represented not by a point but by a **profile**: a line connecting its values along several attribute axes. The axes are parallel, with a dot indicating the attribute value for each object. Points on neighboring axes corresponding to the same object are connected by a line.

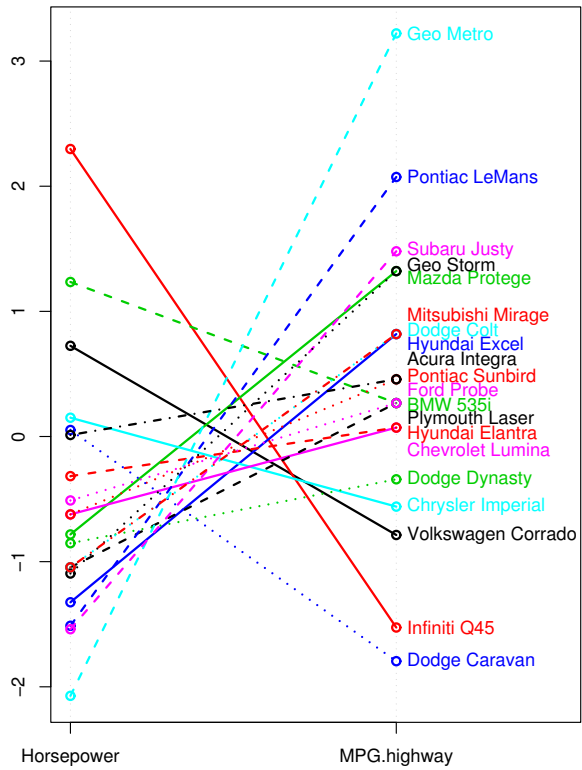
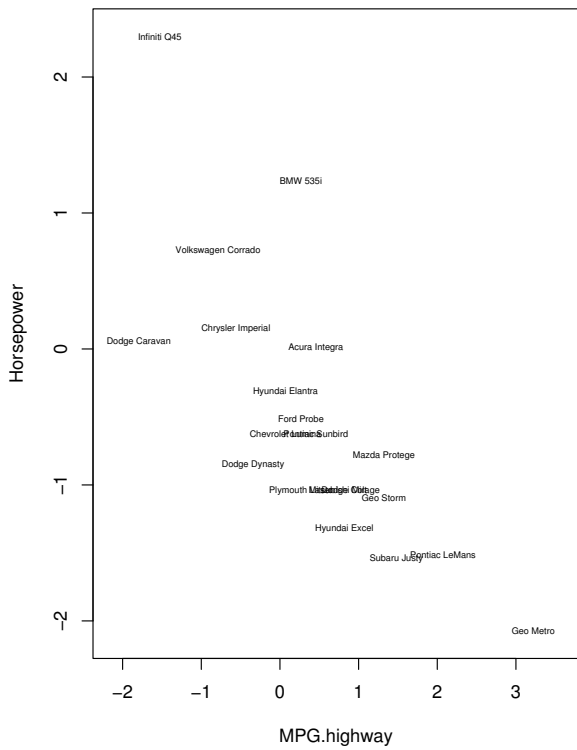
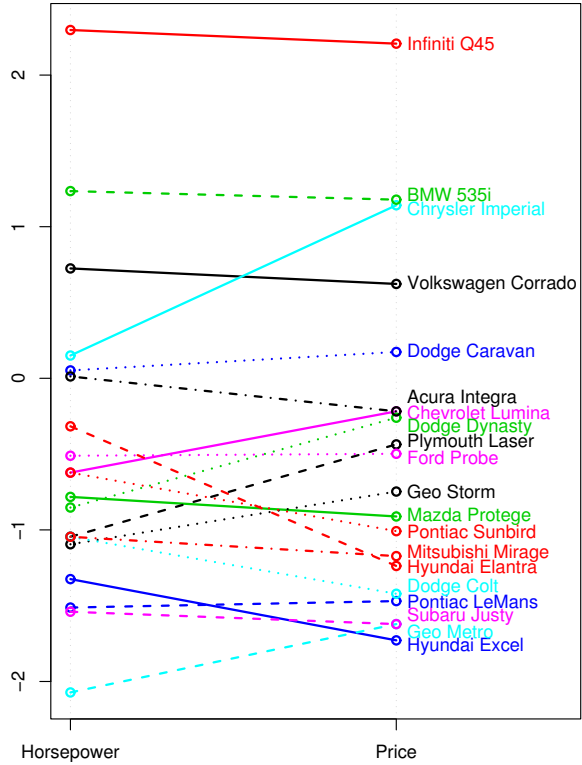
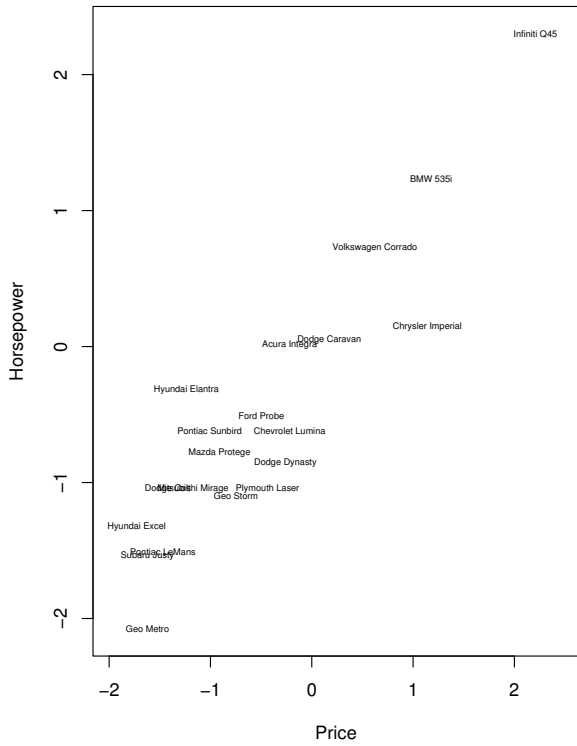
Duality:

- Points in a scatterplot become lines in a parallel-coordinate plot.
- A scatterplot can show many individuals (1000s) in a few dimensions (2). This makes it good for finding outliers and clusters, and judging the correlation between attributes.
- A parallel-coordinate plot can show many dimensions (10) for a few individuals (10). This makes it good for finding unusual attributes, grouping attributes together, and judging the similarity between objects on all attributes.

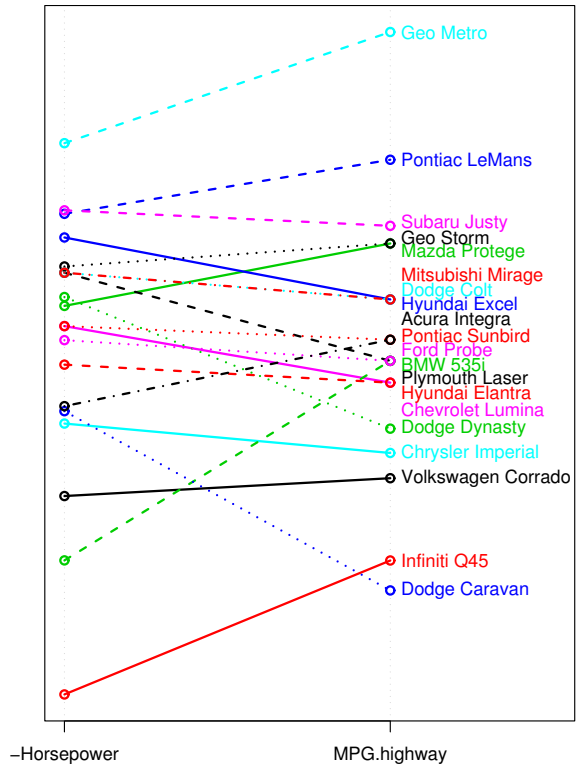
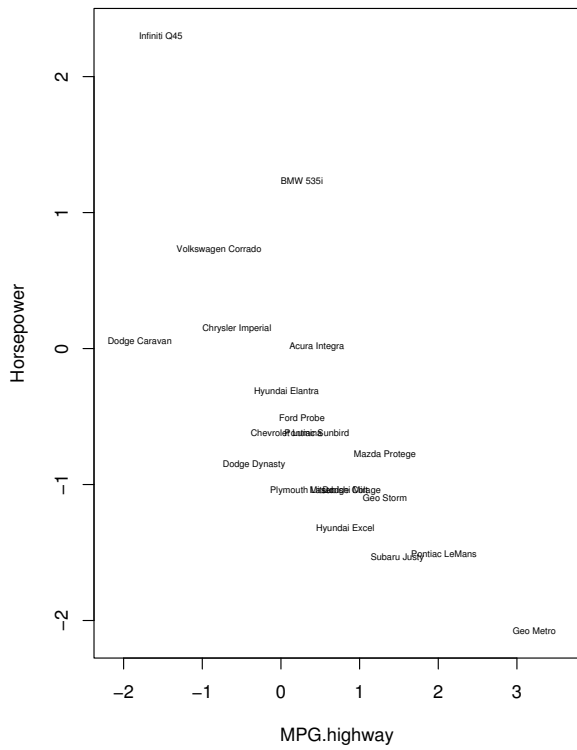
Scaling and sorting makes parallel-coordinates different from a simple line chart. To make the plot readable, the attributes need to be put onto a comparable scale, and sorted so that similar attributes are together. As in projection, this leads to a search for the ‘optimal’ parameters of the plot.

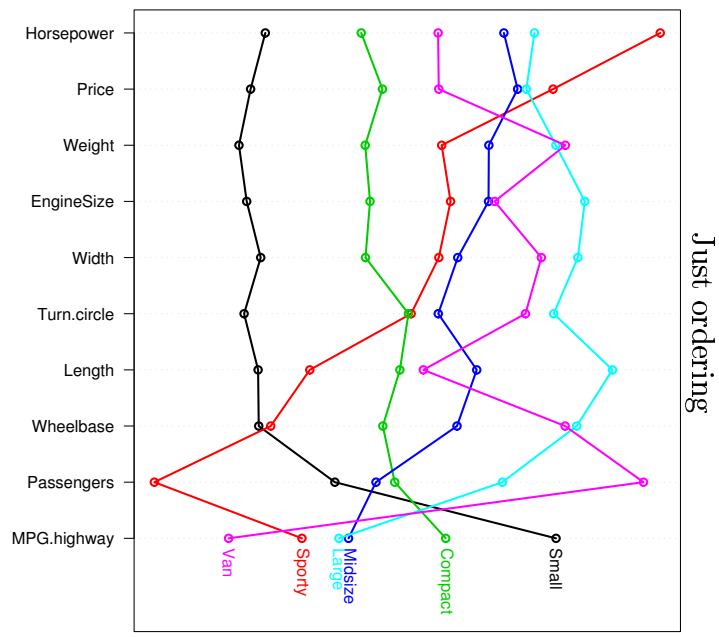
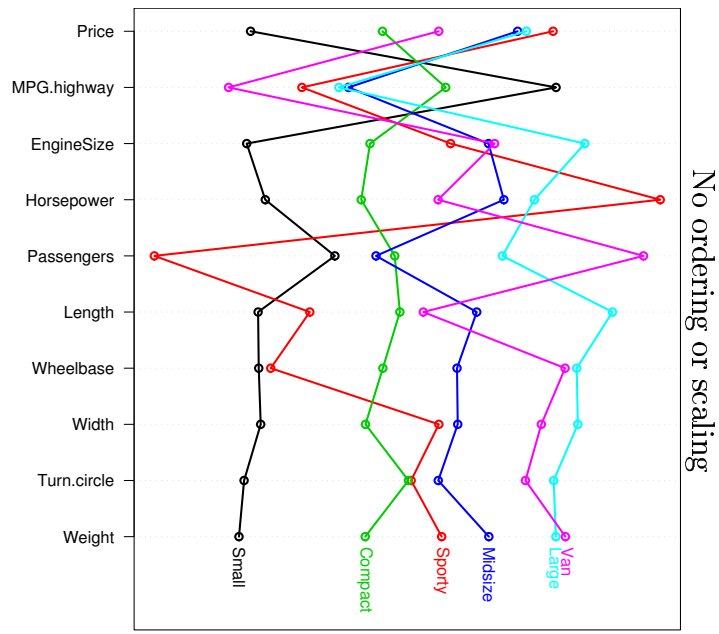
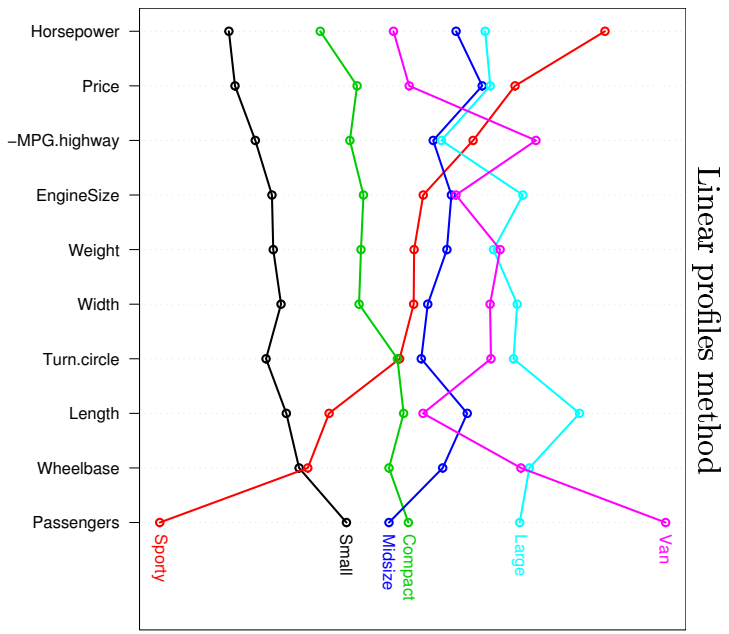
Linear profiles method—Attributes are scaled and sorted to make the profiles turn out as straight as possible (minimize wiggle). Sometimes this involves reversing (negating) an axis. The axes can also be unevenly spaced, to make the profiles more linear. The optimal scaling turns out to be the inverse of the PCA projection weights (attributes with high weight are scaled down, and vice versa).

Scatterplot vs. parallel-plot (no scaling)

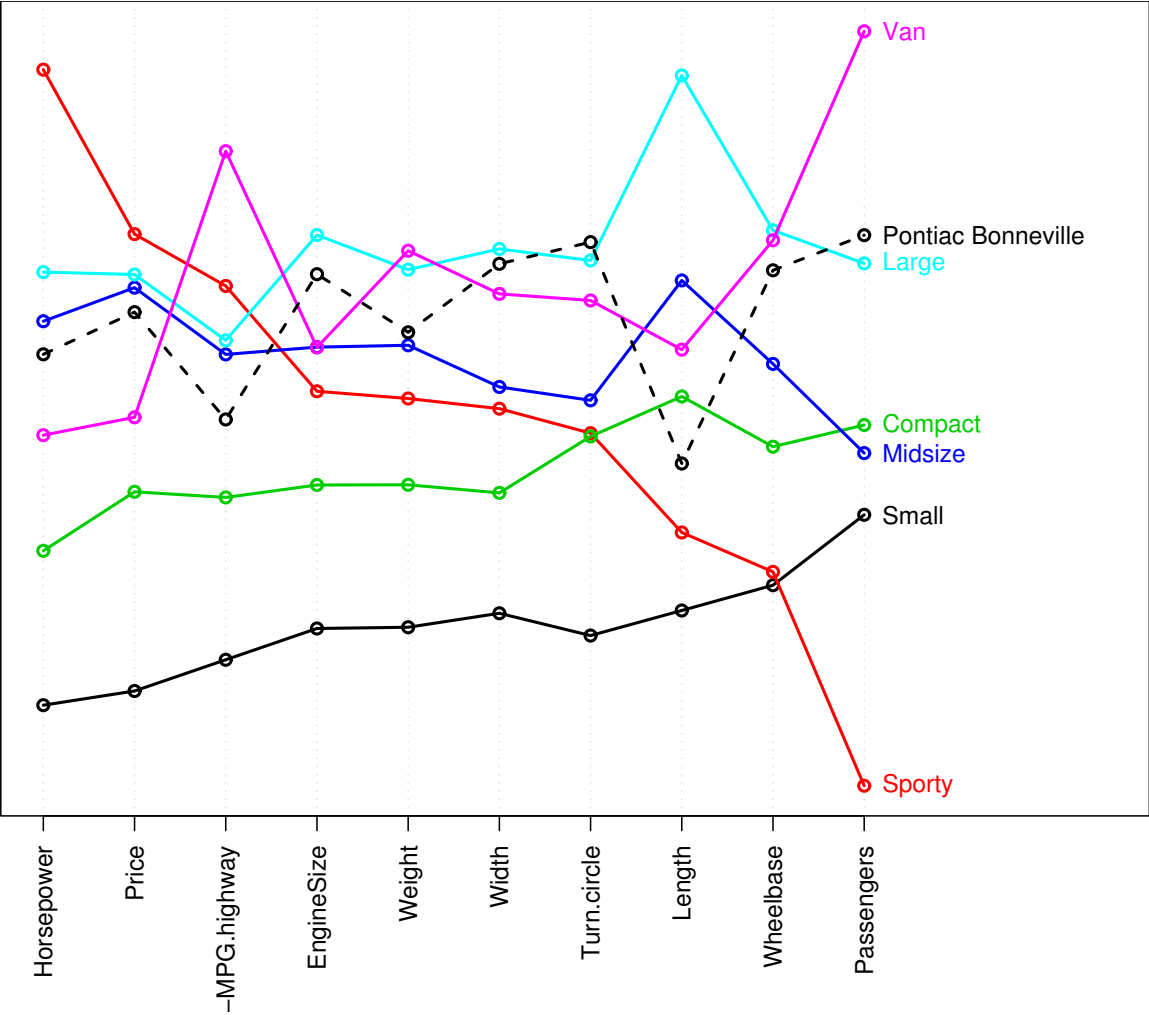


Scatterplot vs. parallel-plot (linear profiles method)





Is the Bonneville Large or Van?



Bonneville's profile resembles Large, but has very small Length (like a Van).

Is the Scoupe Sporty or Small?

