

Visualizing 360 Data Points in a Single Display

Stephen Few

This paper explores ways to visualize a dataset that Jorge Camoes posted on the Perceptual Edge Discussion Forum. Jorge's initial visualization examples were expanded and supplemented by subsequent designs from several others. My intention in this paper is to examine these solutions and to propose one more of my own design.

Let's begin with the dataset itself. In his initial post, Jorge stated the following:

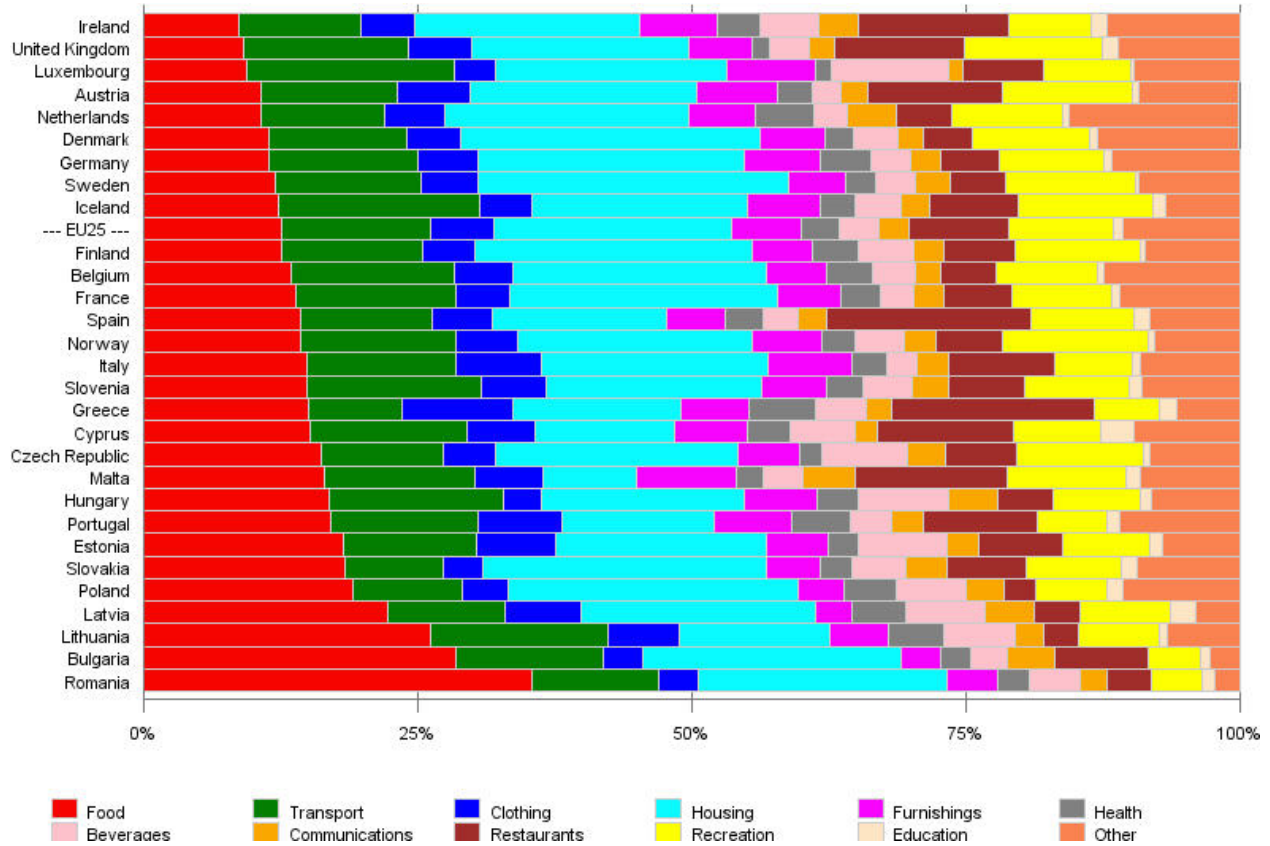
The content doesn't really matter, of course. What I am interested in is how we can plot on a single image all those 360 data points in a way that can help us to understand the data with a minimum cognitive processing.

Contrary to this statement, I believe that the content always matters when you are trying to design a data visualization. The nature of the content is one of the primary factors that determines the appropriate visualization, and also limits the types of questions that we can ask about the data.

This dataset measures the average household expenditure divided into 12 categories (including "other") for 29 countries, which I assume are all members of the European Union (EU). Each category of expenditure is expressed as a percentage of the country's total expenditures. The dataset also includes one more row labeled EU25, which appears to represent the average of each category of expenditures across all of the EU countries. In other words, the expenditure values for each country represent a part-to-whole relationship. We don't know the expenditure amounts (for example, in Euros); only the percentages of total expenditures for each country. This means that we cannot compare total expenditures per country, but we can compare how each country's expenditures are distributed across the 12 categories.

By its very nature, the dataset limits us to examining a limited number of quantitative relationships: part-to-whole (category expenditures to total expenditures), ranking (within individual expenditure categories), deviation (country expenditures compared to EU25 average expenditures), and multivariate (similarities and differences of the countries across the entire set of expenditure categories). The best visualization design can only be determined in light of which relationship that you wish to examine.

If you wish to examine the part-to-whole relationship for each country or to compare the part-to-whole relationships of different countries, the stacked bar design seems to work best. Here is grasshopper's design:



The stacked bars are much easier to compare than the pie charts.

If you wish to examine ranking relationships, bar graphs work best. Nixnut's version (shown on the next page) looks a lot like a "table lens" visualization. It is easy to not only rank the countries based on any one category of expenditures, but to also see if there is a significant correlation between that category and any other category. If the descending magnitude pattern formed by the bars in the housing expenditure column could also be seen in another column, this would indicate a positive correlation between them, and if another column displayed an ascending pattern, this would indicate a negative correlation. Nixnut's example appears to show that there is no significant correlation between housing expenditures and any other category. If we found two expenditure categories that appeared to be correlated, we could examine this more precisely using a scatterplot.

Country	Housing	Food	Transport	Recreation	Restaurants	Furnishings	Clothing	Beverages	Health	Communication	Education	Other
Sweden	23.3	12.0	13.2	11.8	5.1	5.2	5.3	3.7	2.7	3.1	0.3	9.3
Denmark	27.3	11.4	12.5	10.6	4.5	5.9	5.0	4.1	2.6	2.3	0.8	12.8
Poland	26.4	19.1	9.9	6.5	2.9	4.3	4.2	6.3	4.7	3.5	1.5	10.7
Slovakia	25.8	18.3	9.0	8.7	7.2	5.0	3.6	5.0	2.8	3.7	1.4	9.4
Finland	25.3	12.5	12.9	11.4	6.5	5.5	4.8	5.0	4.2	2.8	0.4	8.7
France	24.5	13.8	14.7	9.2	6.1	5.8	4.8	3.1	3.5	2.8	0.7	11.1
Germany	24.3	11.4	13.6	9.5	5.3	7.0	5.4	3.6	4.6	2.8	0.7	12.0
Bulgaria	23.6	28.5	13.4	4.8	8.6	3.5	3.6	3.3	2.8	4.3	0.8	2.7
Belgium	23.1	13.4	14.9	9.3	5.1	5.4	5.4	3.8	4.3	2.3	0.5	12.4
Romania	22.7	35.3	11.6	4.6	4.1	4.8	3.6	4.6	3.0	2.4	1.1	2.3
Netherlands	22.2	10.6	11.4	10.1	5.1	6.2	5.4	3.0	5.3	4.5	0.5	15.5
Czech Republic	22.1	16.1	11.2	11.6	6.5	5.6	4.8	7.8	2.0	3.5	0.5	8.2
EU25 *	21.7	12.5	13.6	9.6	9.0	6.3	5.8	3.6	3.5	2.8	0.9	10.7
Norway	21.5	14.3	14.1	13.3	6.1	6.3	5.6	4.5	3.0	2.9	0.5	7.9
Latvia	21.4	22.2	10.7	8.3	4.1	3.4	6.9	7.2	4.9	4.5	2.3	4.0
Luxembourg	21.1	9.3	19.0	7.9	7.4	8.1	3.7	10.6	1.5	1.4	0.3	9.8
Austria	20.7	10.6	12.5	11.8	12.3	7.3	6.6	2.6	3.2	2.5	0.6	9.1
Italy	20.7	14.8	13.6	7.1	9.7	7.7	7.8	2.7	3.2	2.8	0.8	9.2
Ireland	20.5	8.7	11.1	7.5	13.8	7.1	4.9	5.4	3.8	3.6	1.4	12.5
Iceland	19.7	12.2	18.4	12.3	8.0	6.7	4.7	4.2	3.2	2.6	1.2	6.8
United Kingdom	19.7	9.0	15.1	12.6	11.9	5.8	5.8	3.7	1.6	2.2	1.4	11.1
Slovenia	19.6	14.9	15.9	9.6	6.9	6.0	5.9	4.5	3.3	3.3	1.1	9.1
Estonia	19.2	18.2	12.1	7.9	7.7	5.6	7.3	8.0	2.8	2.9	1.1	7.3
Hungary	18.4	16.9	15.8	7.9	5.0	6.7	3.6	8.3	3.7	4.5	1.1	8.0
Spain	18.0	14.2	12.0	9.3	18.6	5.3	5.5	3.1	3.5	2.7	1.5	8.3
Greece	15.4	15.0	8.5	6.0	18.4	6.2	10.1	4.6	6.0	2.4	1.8	5.9
Portugal	13.9	17.0	13.5	6.4	10.3	7.1	7.6	3.8	5.3	2.9	1.2	10.9
Lithuania	13.8	26.1	16.2	7.4	3.1	5.3	6.5	6.8	5.0	2.6	0.7	6.8
Cyprus	12.7	15.2	14.2	8.0	12.4	6.7	6.2	6.0	3.9	2.0	3.0	9.8
Malta	8.5	16.5	13.7	10.9	13.9	9.1	6.2	3.6	2.4	4.8	1.3	9.1

* EU25 represents the mean average of the 25 member states of the European Union

If you wish to examine the deviation of country expenditures from the EU average, once again bars would do the job nicely. In this case, however, you could display each expenditure as a positive or negative deviation from the EU average, with bars extending to the right or left of the zero baseline in the middle.

What if you want to examine similarities and differences between countries, as well as significant exceptions, across all categories of expenses at once? To do so would qualify as multivariate analysis (each category of expenses equals one variable). The bars charts in the form of a table lens display gives us a glimpse of the countries multivariate profiles, but it is much better designed to support comparisons within a given column than comparisons across the columns in individual rows.

A heatmap (values encoded as color) structured as a matrix is one way to examine multivariate profiles, as seen in Jorge's Excel 2007 conditional formatting example.

HOUSEHOLD EXPENDITURE IN EUROPE (%)
2005

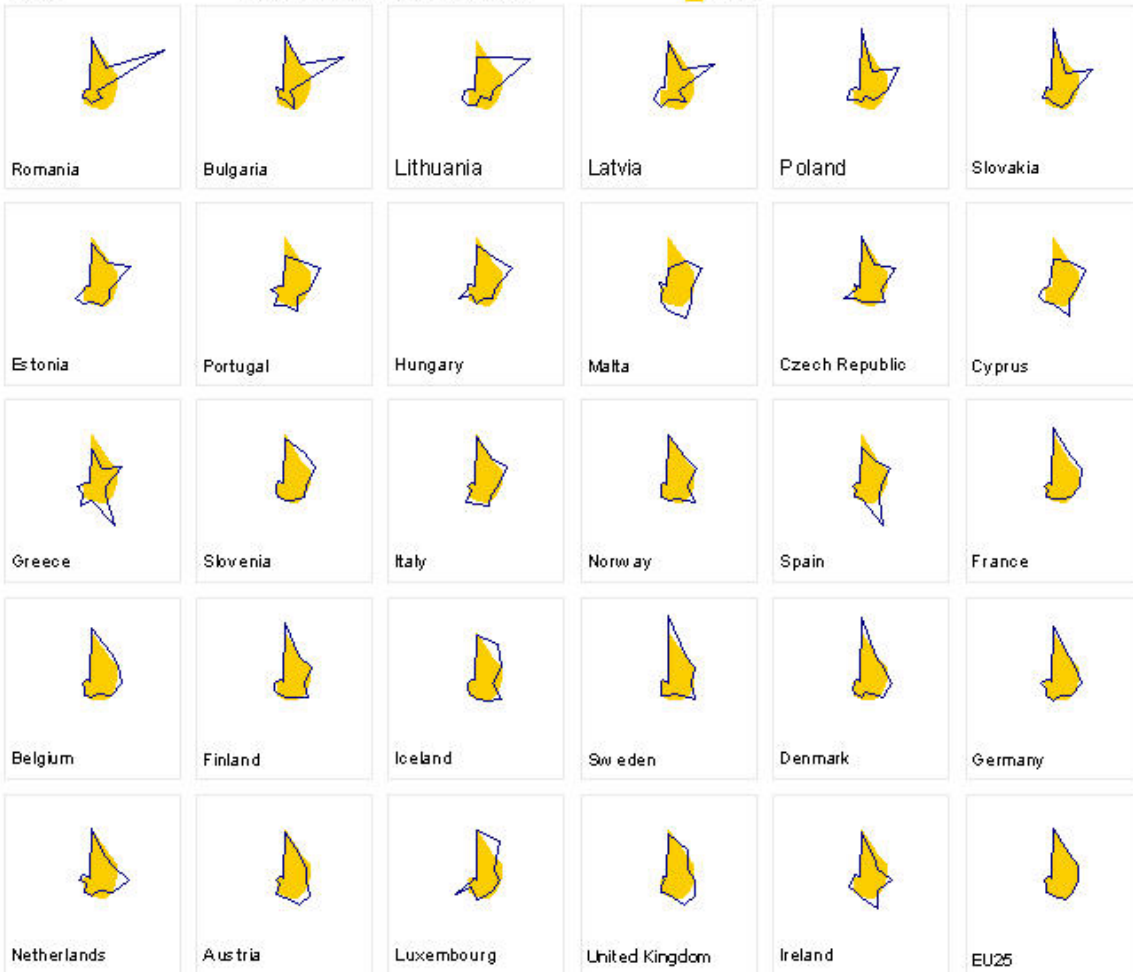
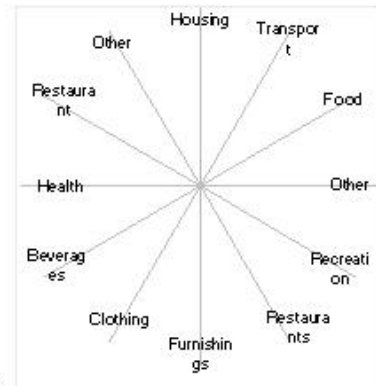
	Housing	Transport	Food	Recreation	Restaurants	Furnishings	Clothing	Beverages	Health	Communications	Education	Other
EU25	21,7	13,6	12,5	9,6	9,0	6,3	5,8	3,6	3,5	2,8	0,9	10,7
Sweden	28,3	13,2	12,0	11,8	5,1	5,2	5,3	3,7	2,7	3,1	0,3	9,3
Denmark	27,3	12,5	11,4	10,6	4,5	5,9	5,0	4,1	2,6	2,3	0,8	12,8
Poland	26,4	9,9	19,1	6,5	2,9	4,3	4,2	6,3	4,7	3,5	1,5	10,7
Slovakia	25,8	9,0	18,3	8,7	7,2	5,0	3,6	5,0	2,8	3,7	1,4	9,4
Finland	25,3	12,9	12,5	11,4	6,5	5,5	4,8	5,0	4,2	2,8	0,4	8,7
France	24,5	14,7	13,8	9,2	6,1	5,8	4,8	3,1	3,5	2,8	0,7	11,1
Germany	24,3	13,6	11,4	9,5	5,3	7,0	5,4	3,6	4,6	2,8	0,7	12,0
Bulgaria	23,6	13,4	28,5	4,8	8,6	3,5	3,6	3,3	2,8	4,3	0,8	2,7
Belgium	23,1	14,9	13,4	9,3	5,1	5,4	5,4	3,8	4,3	2,3	0,5	12,4
Romania	22,7	11,6	35,3	4,6	4,1	4,6	3,6	4,6	3,0	2,4	1,1	2,3
Netherlands	22,2	11,4	10,6	10,1	5,1	6,2	5,4	3,0	5,3	4,5	0,5	15,5
Czech Republic	22,1	11,2	16,1	11,6	6,5	5,6	4,8	7,8	2,0	3,5	0,5	8,2
Norway	21,5	14,1	14,3	13,3	6,1	6,3	5,6	4,5	3,0	2,9	0,5	7,9
Latvia	21,4	10,7	22,2	8,3	4,1	3,4	6,9	7,2	4,9	4,5	2,3	4,0
Luxembourg	21,1	19,0	9,3	7,9	7,4	8,1	3,7	10,6	1,5	1,4	0,3	9,8
Italy	20,7	13,6	14,8	7,1	9,7	7,7	7,8	2,7	3,2	2,8	0,8	9,2
Austria	20,7	12,5	10,6	11,8	12,3	7,3	6,6	2,6	3,2	2,5	0,6	9,1
Ireland	20,5	11,1	8,7	7,5	13,8	7,1	4,9	5,4	3,8	3,6	1,4	12,5
Iceland	19,7	18,4	12,2	12,3	8,0	6,7	4,7	4,2	3,2	2,6	1,2	6,8
United Kingdom	19,7	15,1	9,0	12,6	11,9	5,8	5,8	3,7	1,6	2,2	1,4	11,1
Slovenia	19,6	15,9	14,9	9,6	6,9	6,0	5,9	4,5	3,3	3,3	1,1	9,1
Estonia	19,2	12,1	18,2	7,9	7,7	5,6	7,3	8,0	2,8	2,9	1,1	7,3
Hungary	18,4	15,8	16,9	7,9	5,0	6,7	3,6	8,3	3,7	4,5	1,1	8,0
Spain	16,0	12,0	14,2	9,3	18,6	5,3	5,5	3,1	3,5	2,7	1,5	8,3
Greece	15,4	8,5	15,0	6,0	18,4	6,2	10,1	4,6	6,0	2,4	1,6	5,9
Portugal	13,9	13,5	17,0	6,4	10,3	7,1	7,6	3,8	5,3	2,9	1,2	10,9
Lithuania	13,8	16,2	26,1	7,4	3,1	5,3	6,5	6,6	5,0	2,6	0,7	6,8
Cyprus	12,7	14,2	15,2	8,0	12,4	6,7	6,2	6,0	3,9	2,0	3,0	9,8
Malta	8,5	13,7	16,5	10,9	13,9	9,1	6,2	3,6	2,4	4,8	1,3	9,1

If the differences in values from low to high were spread across a broader range of color intensity, ranging from a lighter orange to darker orange than is seen in this example, it would be a little easier to spot multivariate similarities and differences, and much easier to spot the extremes. If a diverging color scale were used instead of this sequential scale, using one color to show values that are above the EU average and another to show values below the average, and a shade of gray for values close to the average, multivariate profiles could be seen every more clearly. Also, for purposes of multivariate analysis, rather than assigning the range of color intensity to match represent the lowest and highest values in the entire dataset, the entire range of color from lightest to darkest could have been assigned independently to each category of expenditures, ranging from the lowest to the highest values in that category only, which would have expressed values more clearly and made multivariate profiles more comparable.

Jorge's star glyph display also provides a means to make multivariate comparisons. I am not a fan of glyphs, however, compared to other means of display. As far as I can tell, probably the only multivariate pattern in this dataset that the star glyph could display better than any other is one of symmetry between all the variables. In other words, if you wished to find those countries whose expenditures were distributed most uniformly across the 12 categories, a symmetrically shaped star glyph would represent this fairly well.

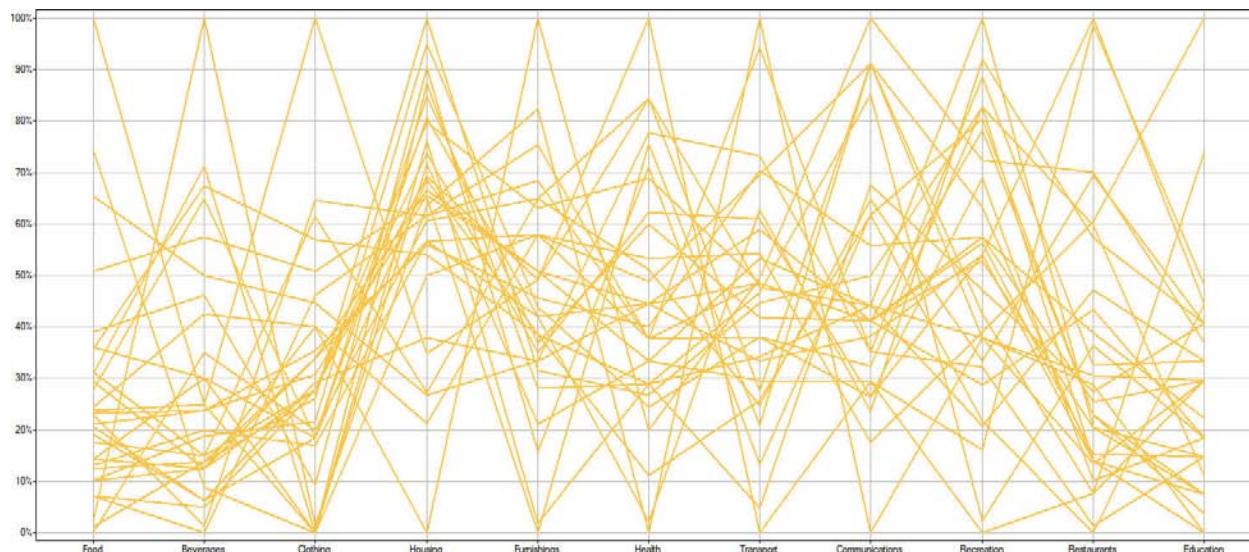
HOUSEHOLD EXPENDITURE IN EUROPEAN COUNTRIES, 2005

Food	Food and non-alcoholic beverages
Beverages	Alcoholic beverages, tobacco and narcotics
Clothing	Clothing and footwear
Housing	Housing, water, electricity, gas and other fuels
Furnishing	Furnishings, household equipment and routine maintenance
Health	Health
Transport	Transport
Communication	Communications
Recreation	Recreation and culture
Education	Education
Restaurant	Restaurants and hotels
Other	Miscellaneous goods and services



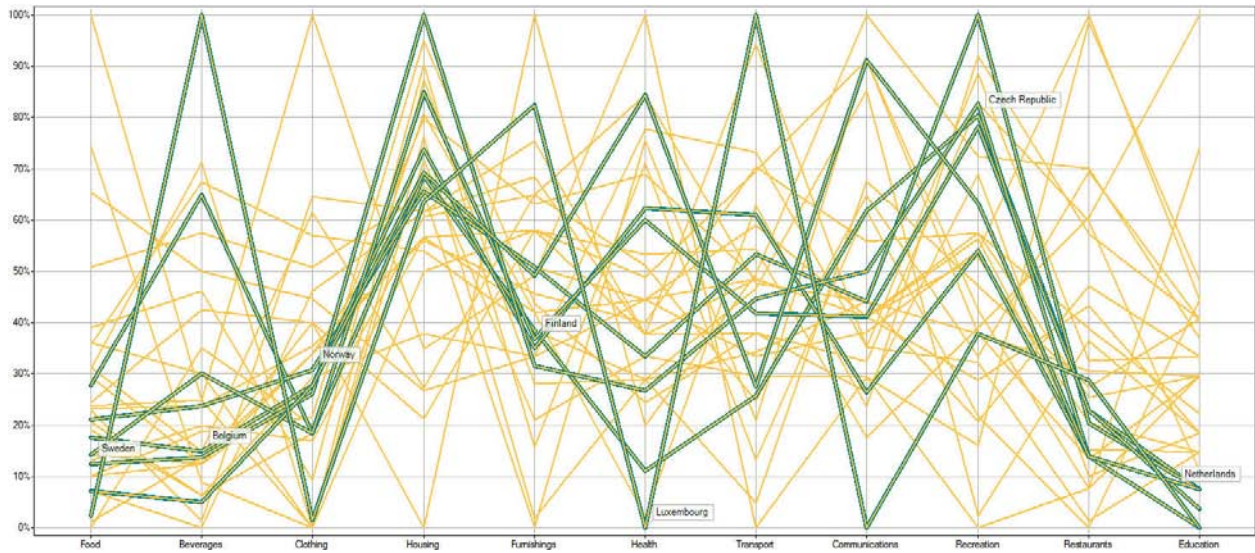
The best visualization for multivariate analysis, however, does not appear in the solutions that have been proposed so far: parallel coordinates. Let's look at what we could discover by using parallel coordinates.

Here is the full dataset displayed as parallel coordinates, which the exception of the "other" category, which I have excluded because I don't think it adds much of value to the multivariate display.

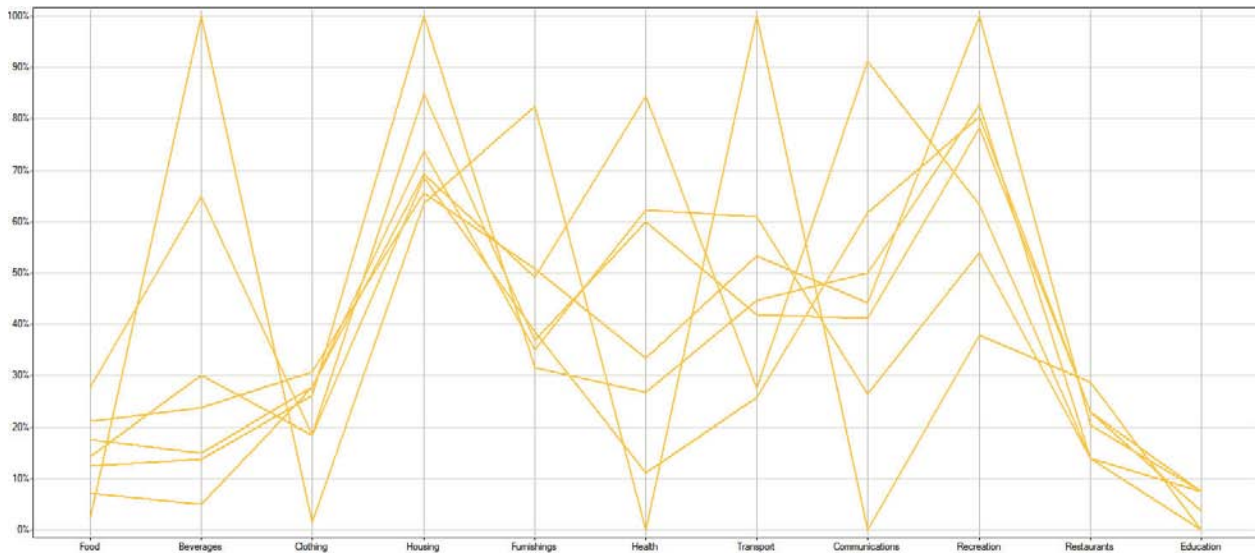


Each line is a country (with one for the EU25 average as well). Each vertical axis measures a difference category of expenditures, beginning with "food" on the left and ending with "education" on the right. With this view we certainly cannot see the multivariate profile of any single country, but we can spot exceptions and predominant patterns. For instance, the country with the highest beverage expenditures spends 43% more than the country with the next highest beverage expenditures (100% divided by 70%). We can also see that by far most countries spend more than twice as much on housing than the country with the lowest housing expenditures.

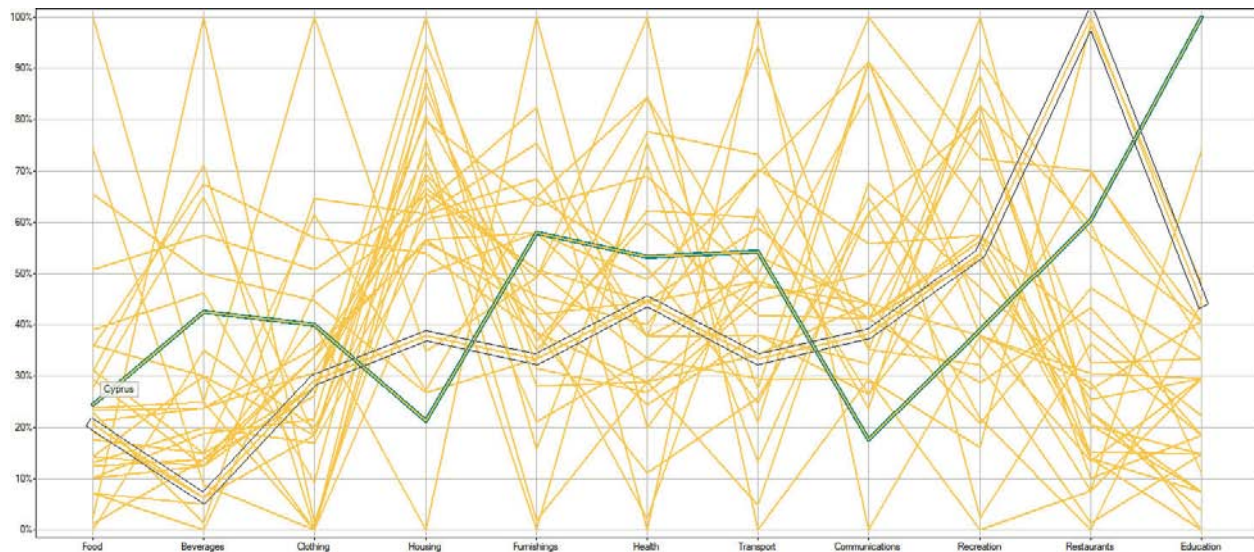
I intentionally arranged the expenditure categories with education on the right because I was curious about multivariate patterns that might be associated with especially low or high education expenditures. Placing education on the right makes it easy to focus on selected groups of countries based on their education expenditures. For instance, in the next example I've highlighted the seven countries with the lowest education expenditures.



By doing this, I can begin to look for significant patterns in the other expenditure categories. For instance, I can see that all seven of these countries have housing expenditures that fall within the top 40% of the range. If I want to eliminate the distraction of the other countries, I can easily filter out all but these seven, as shown below. Notice that those countries that spend the least on education also tend to spend little on restaurants.

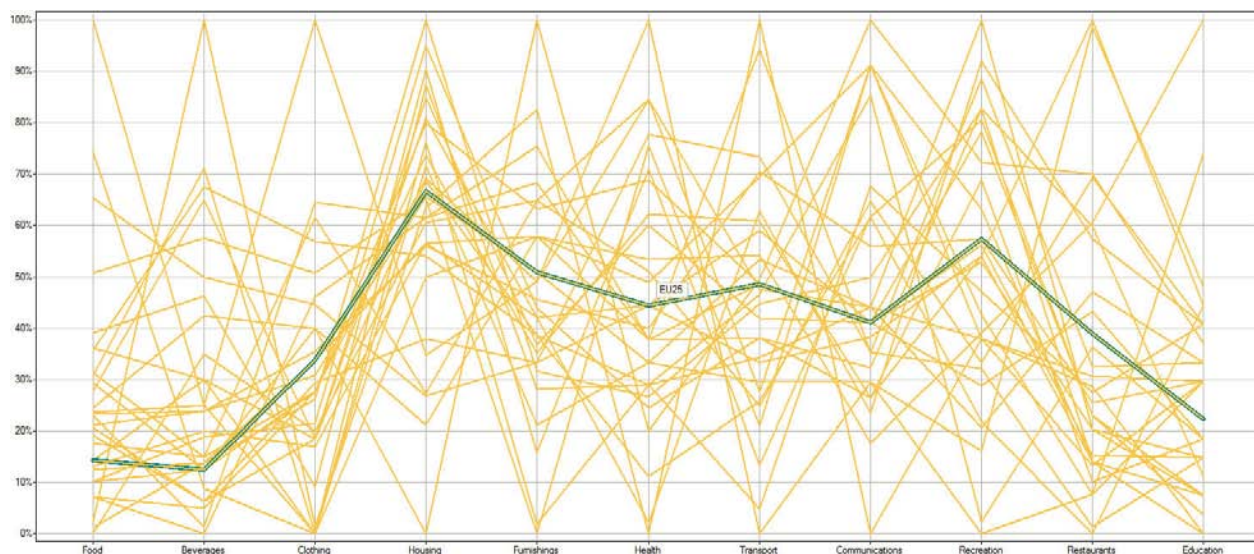


What about the country that spends the most on education? I wonder if it exhibits some interesting characteristics in any of the other categories. In the next example, I've highlighted this country (Cyprus, with the darkest line) and have also featured the EU25 average (the line that is outlined) to compare them.



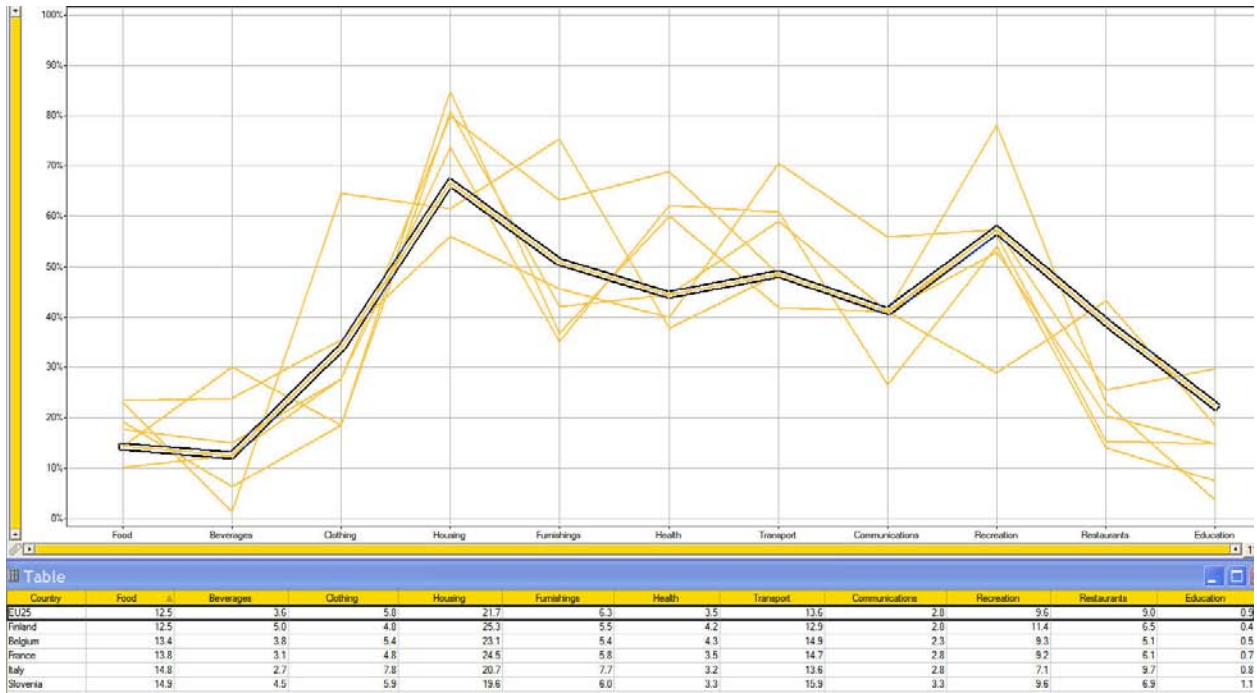
Cyprus appears to have a multivariate profile that is much different from the norm, although it comes very close to the norm in food expenditures.

What if I want to determine which countries have multivariate profiles that are most similar to the norm? In the following example, the entire dataset is shown with the EU25 average highlighted.

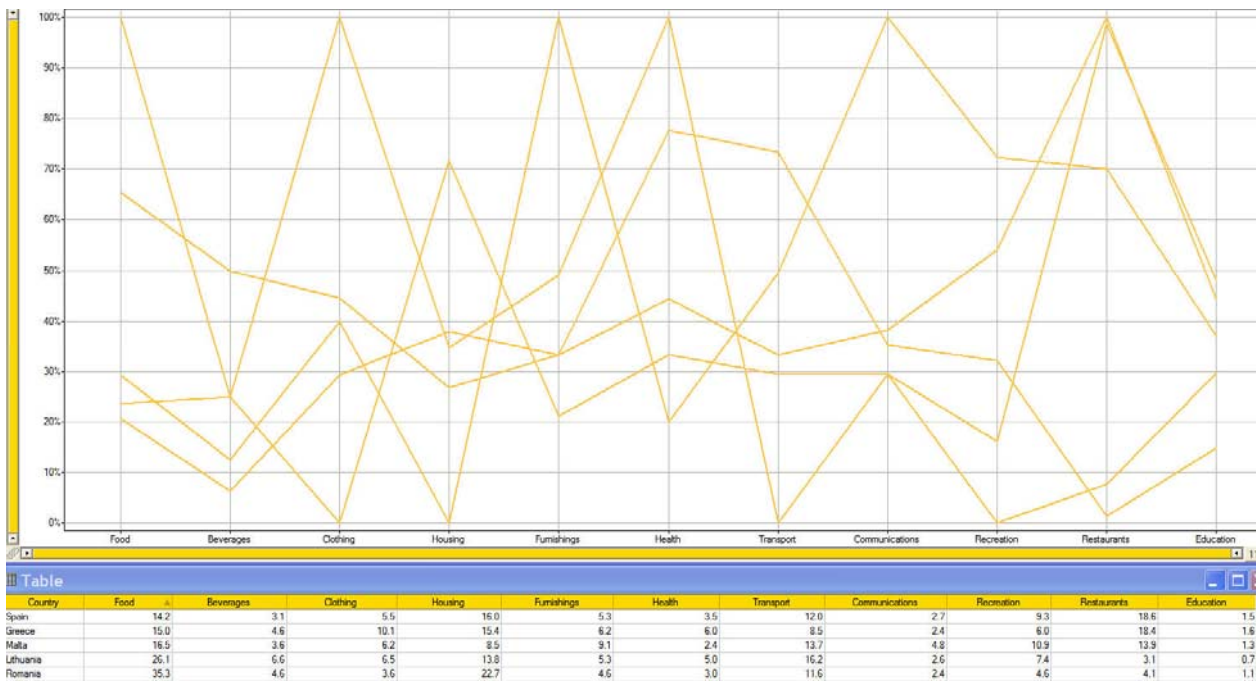


With this view, I can't see which countries are similar to the norm, but I can take advantage of profile matching to make this easy. I'm using the parallel coordinates that are available in Spotfire DecisionSite to create these examples. This product has profile matching built in, so all I need to do is ask it to rank all of the countries based on their multivariate similarity to the EU25 average. After doing this, I filtered all by the six countries that are most similar to the norm, which you can see in the next example. I've

included a table below the parallel coordinates to provide the names and precise values for each of these countries.

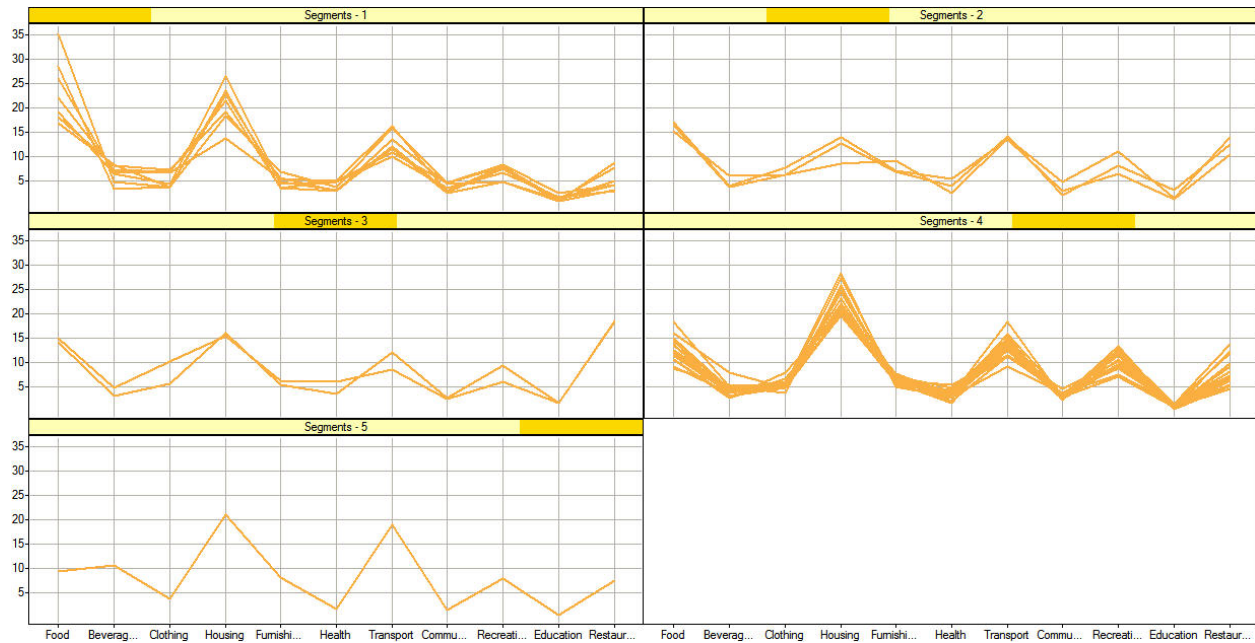


How about the least similar multivariate profiles to the norm? The next example displays the least similar countries.



I hope that you're recognizing that the visualization itself is limited if the software does not provide useful ways to interact with the data. Another example of a useful interaction

is to see if there are meaningful clusters of multivariate profiles—natural groups of countries that are alike, based on some measure of similarity, such as the highest correlations across all sets of variables. The next example shows the results of asking the software to search for such clusters, which I've shown as a series of small multiples to make it easier to compare and contrast them.



Parallel coordinates, although a little confusing and cluttered looking at first glance, provide a powerful means of performing multivariate analysis. They were invented by [Alfred Inselberg](#), who is currently a professor of mathematical sciences at the University of Tel Aviv. Although his work can get quite technical, it is brilliant. Parallel coordinates can bring a great many otherwise rare insights to the multivariate analysis of business data.