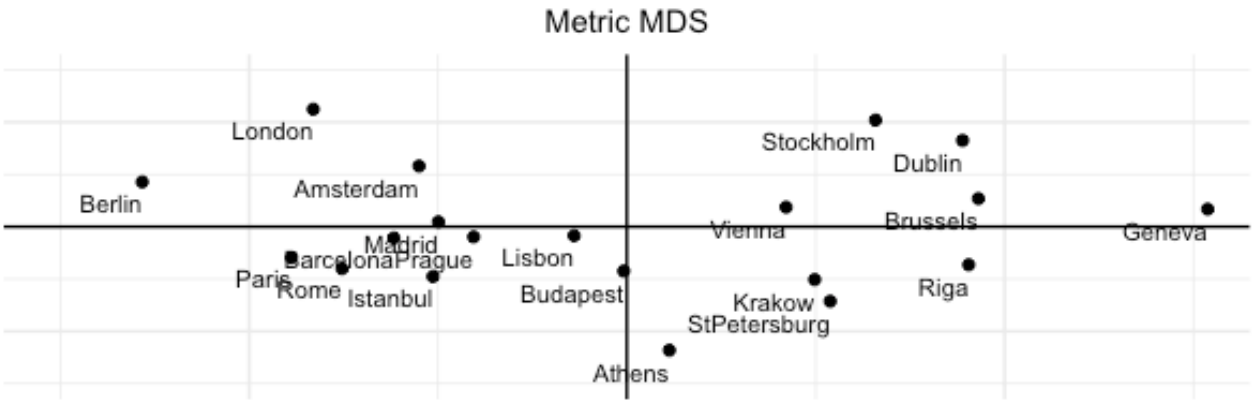
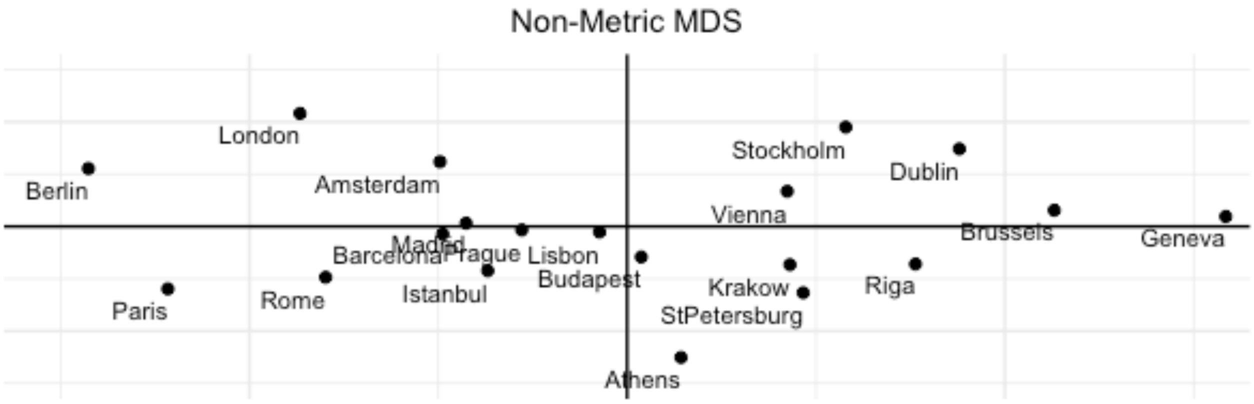
Multidimensional Scaling

In the first part multidimensional scaling (MDS) was applied, which is a visual representation of the distances or dissimilarities between objects in a cartesian space based on the distances between each of them. Objects are more similar to each other, the closer they appear in the cartesian space.

MDS can be divided into two subtypes, metric and non-metric MDS, which differ in the way the dissimilarities are transformed into distances. The metric MDS creates a linear relationship, i.e. if one doubles the dissimilarity one also doubles the distance in the cartesian space, whereas the non-metric MDS uses an ordinal scale, it is therefore a representation of the similarity of rankings instead of the actual distances.

MDS is always being applied to a matrix, in this case a triangular matrix of euclidian distances between the cities based on their average attribute scores. Both metric and non-metric MDS were computed and displayed in a 2-dimensional cartesian space using ggplot2.





As to be expected the results are fairly similar for both types of MDS. Geographical proximity seems to affect the placing on the cartesian space to a certain degree, possibly because of cultural similarities. One can see that Madrid, Lisbon and Barcelona are close, just like Brussels and Geneva. Furthermore Budapest, Krakow, St. Petersburg and Riga as well as Berlin, Amsterdam and London appear in the same quadrant, which suggests a high degree of similarity. In the following section property fitting will shed some light on which attributes may be associated with the positions of the individual cities.

Property Fitting

Property fitting is a method of analyzing the proximities based on dimension to get an idea of which attributes may have affected the coordinates of each city. In our case it makes it possible to plot the attributes and preferences into the cartesian space that resulted from the MDS.

First it is necessary relate the position in the cartesian space to the individual properties. A linear regression with the attributes as the dependent variables, and the coordinates of the cities as the independent variables is being used to achieve this goal. The regression summary for the attribute „romantic“ is displayed below.

Response romantic :

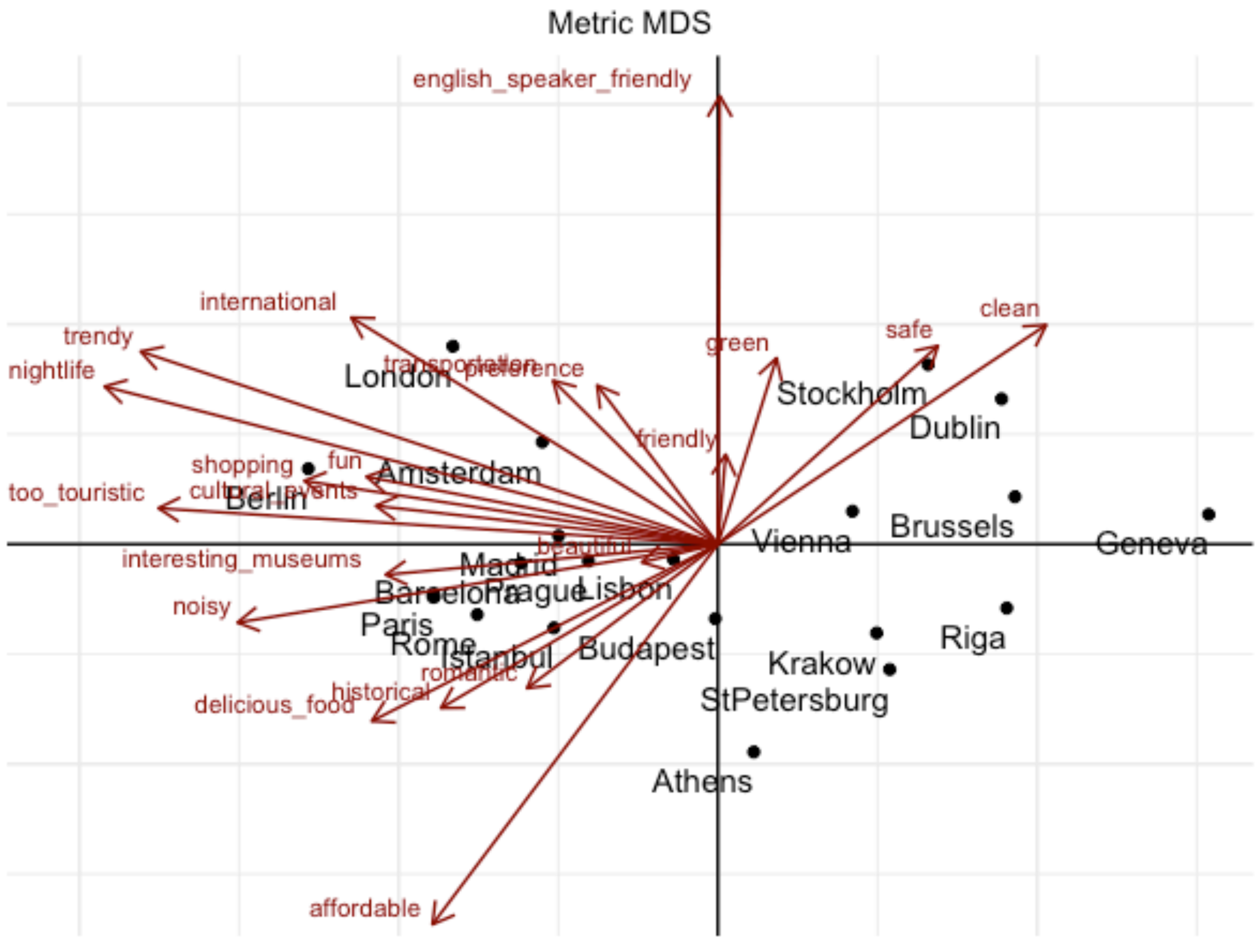
Coefficients:

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
x	*-0.1196*	*0.5596*	*-0.214*	*0.833*
y	*-0.1639*	*0.7239*	*-0.226*	*0.823*

Multiple R-squared: 0.005356, Adjusted R-squared: -0.1052
F-statistic: 0.04846 on 2 and 18 DF, p-value: 0.9528

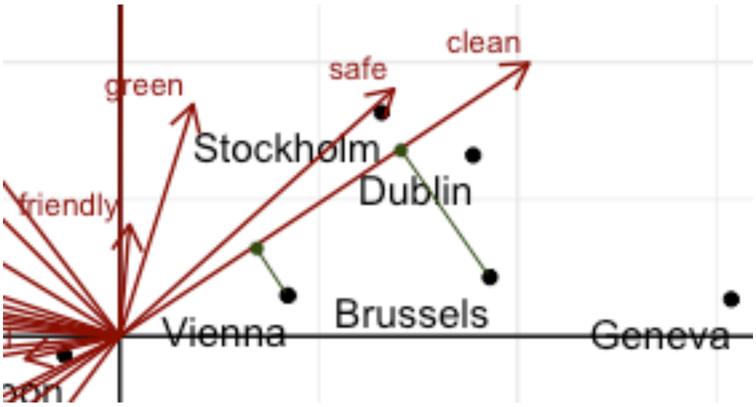
In general, the impact of the cities coordinates on the attributes is statistically insignificant. The standard error (ranging from 0.52 to 0.97) is relatively high compared to the estimates, the t-statistic values (ranging from -0.64 to 0.66) are fairly close to zero and relatively small compared to the standard error, the R-squared (ranging from 0.001 to 0.027) is very low and the p-values (ranging from 0.51 to 0.97) are very high.

However the purpose of the regression is not to get a particularly good fit, which would be unlikely given the nature of the data, but rather to give a rough idea of the relation of the attributes to the coordinates. The graph below shows the results of the regression plotted over the results of the metric MDS using ggplot.



To interpret the visualization one needs to project the location of the city onto the individual attribute by drawing a perpendicular line to the arrow. The distance from the city to the arrow is meaningless and irrelevant for the interpretation. The only relevant thing is the point in which the line meets the arrow.

For example, in the given graph, Brussels would be considered more clean than Vienna, because the point of intersection is closer to the tip of the arrow, even though it is further away from the arrow itself and the direction it is pointing towards. It is also worth pointing out, that the arrows do extend past the origin, so Geneva is less trendy than Vienna.



The results of the property fitting are in line with what one might expect. Paris and Rome are related to the attributes „romantic“ and „historical“. London and Dublin are outstandingly „english speaker friendly“. Stockholm, Brussels and Geneva score high on being „clean“ and „safe“. Berlin, Amsterdam and London are both „fun“, „trendy“ and „international“, but also „too touristic“.

Lastly the cities in the second quadrant appear to be particularly correlated with the preferences for a good city trip of the queried respondents.

Key Findings and Usefulness of the Analysis

First of all the results of the MDS visualize the similarities of the cities in the dataset. This can be used in order to build a recommender system for city trips, as it may be likely that customers will want to visit a city which is similar to the one they have visited previously (or not). It could also prove useful to offer city trips in bundles of similar cities, as this might increase the probability of them being bought by a single customer, who’s preferences are aligned with the properties of the cities.

The property fitting allows to judge which attributes of the cities are causing their proximity in the cartesian space and also how the participants of the questionnaire view the cities. Additionally it gives an insight into which aspects of a city are responsible for its popularity. Those insights may be used to focus on these very aspects when advertising a trip to a specific city or choose the most reasonable target audience for certain advertisements.

For a city trip provider it can be of great value to know which cities are particularly affordable. He may want to keep his business risks as low as possible or attract younger customers with very inexpensive vacation trips.

Finally, the property fitting gives an insight into which cities are most in line with the customers preferences. This can be used to focus a marketing campaign on the less popular cities or calibrate economical decisions. A small city trip provider who is scaling his business may be interested in focussing only on the most popular of cities.