Names: Sean Carney / Cameron Carton

Student Numbers: 19349341 / 19720959

Dataset description: Cost of Living by City 2018 Dataset.

This dataset contains information pertaining to the cost of living in various cities around the world. The Indexes in the data show the relative Cost of Living(Plus Rent), Groceries Prices, Restaurant Prices, Rent, and the Local Purchasing Power of each city. The indexes are measured relative to New York City, USA which has an index of 100.00 in every category.

Initial questions: Which variables have the strongest correlation with cost of living?, Why do these variables correlate strongly with the cost of living?

Description: The scatterplot shows the relationship between the Cost of Living, Groceries, and Restaurant price in Cities. Each point represents a city. The colour and size of a point represents the Restaurant price, starting at black, leading into a small red point, then a larger orange point, and an even larger yellow point. Hovering over a point will reveal the associated cityâ€™s name.

Insight: The groceries index and restaurant price index were found to be the best indication of the cost of living index. The scatter plot clearly shows the general trend for the cost of living to increase with the increasing cost of groceries. The increasing size of the points as each of these increases as well as the colour shifting to yellow visualises the correlation of restaurant prices and cost of living.

There doesnâ€™t appear to be any extreme outliers in the dataset. All the points generally follow the same trend. Although not the most surprising outcome (higher cost of living indicates higher grocery and restaurant prices), what is interesting is that the graph does look as though it is beginning to curve, possible meaning the prices of groceries begin to taper past a certain point of cost of living.

The direct correlation between groceries and cost of living is obvious. The best reasoning we can offer is that around 20-30% of household income is spent on food for most countries, but some countries spend significantly less or more.

Design considerations: A scatter plot was chosen as it is the best way to show the correlation between the 3 quantitative datasets.

The groceries and restaurant price indexes were chosen as the best indicators after creating scatter plots of cost of living index against all available indexes. These scatter plots confirm that groceries and restaurant prices correlate the strongest with the cost of living.

Cost of Living Index was chosen instead of Cost of Living Plus Rent because it seems obvious that rent would impact the latter greatest.

The size and colour of the points was chosen to display the Restaurant Price Index because it is the most clear visual way to display all 3 variables. Using both size and colour is useful because it makes the general trend more apparent which is what we aimed to do, although it does cost the readability of the individual points.

The colour scheme we chose was better than many of the alternatives because it had a distinct gradient which makes understanding the plot easier.

Data filtering and transformation: No data filtering or transformation was needed as there were no significant outliers or missing data.