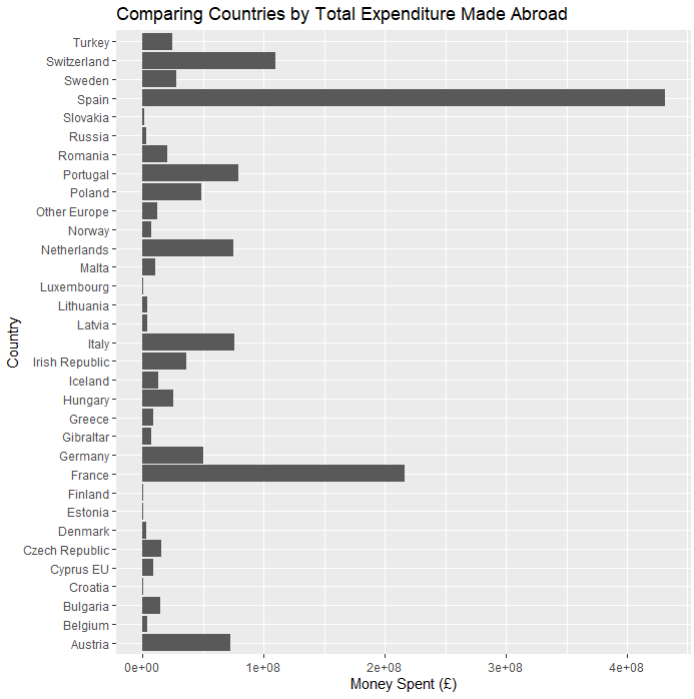
Math6147 - R Coursework Report

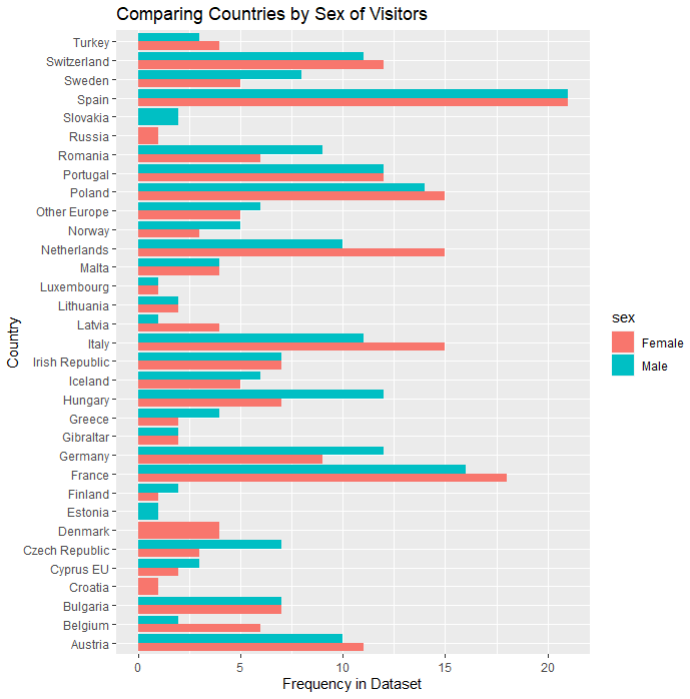
Student ID: 2940 4045

October 2021

Characteristics of Travel Habits between Countries

To begin, we may note that the mode of transport is always "air" across every record in the data-set. This may suggest either that air-transport is significantly more popular than all other forms of transport to the point of being the overwhelming majority, or that other forms of transport were simply not surveyed when data collection was performed for this data-set.
We may also compare total spend made by visits and sex of visitors country-by-country:





We see from the above that significantly more is spent during visits to Sweden than all other location, with France second highest though also significantly higher than all other countries.
We may also observe that, despite mainly even splits by sex in each country, only male visitors visited Slovakia and Estonia. Additionally, only female visitors visited Russia, Denmark and Croatia. This

suggests either that these particular are popular depending explicitly on the sex of the visitors, or that this sample contains anomalous ratios of males-to-females for these countries during the 3-month sampling period.

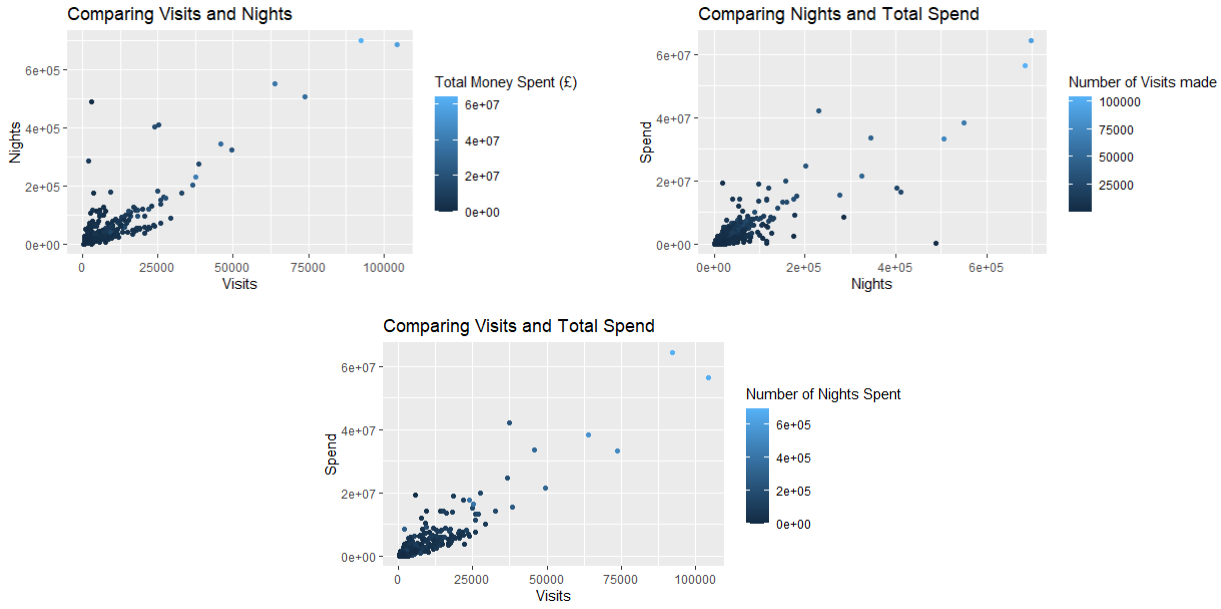
Potential Relationships between Visits, Spends and Nights

We may also test for correlations directly between each of the three variables in turn by using Pearson's product-moment correlation:

- Visits and Nights: 0.8382415
- Nights and Spend: 0.9100663
- Nights and Visits: 0.8445919

The above all demonstrate a strong positive between each of the three examined variables. From this, we may conclude that more popular tourist destinations to make a visit to will generally also have visitors staying for more nights and spending more in doing so.

However, we may also compare the three variables directly in turn via scatter-plots to examine their distributions, with the third variable in each plot being given by a heat-map across the scattered points.



The above make it very clear that it is still more common within the data itself for the number of visits to be low, spends to be low and for the number of nights stayed to be low. This agrees with our positive correlation, as our three variables all decrease together when clustering in the above plots. The positive correlations all appear again in the heat-map, with darker points clustering near the axis and lighter points always near the top-right of the axis.

The Effect of Age and Sex on Visits, Spends and Nights

To examine the effect of age and sex on these three variables, we produce three column plots with age along the x-axis and sex represented with colours.

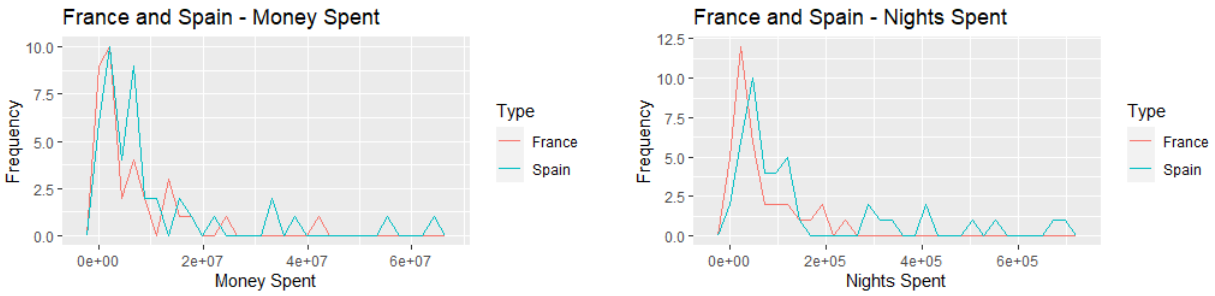
Broadly, we may first observe that number of visits, total spent and number of nights spent all increase with age.
Though, we may also see that visits, total spend and nights spent are all much higher for female visitors in the age ranges "45-54" and "55-64" than for male visitors in the same age ranges or for visitors of either sex in any other age category.
Looking specifically at visits outside those age ranges, we see that visits are generally longer for male visitors than female visitors, with this difference being more pronounced in the higher age ranges.
The same trend can be found within nights, excluding the same two age ranges.

Looking at spends specifically outside of the same two age ranges, we struggle to find a clear correlation.

From these observations, we can infer that total spend, number of visits and number of nights spend are all much higher for female visitors between ages 45 and 64. With a minor correlation between number of visits and number of nights spent with age, more-so for male visitors.

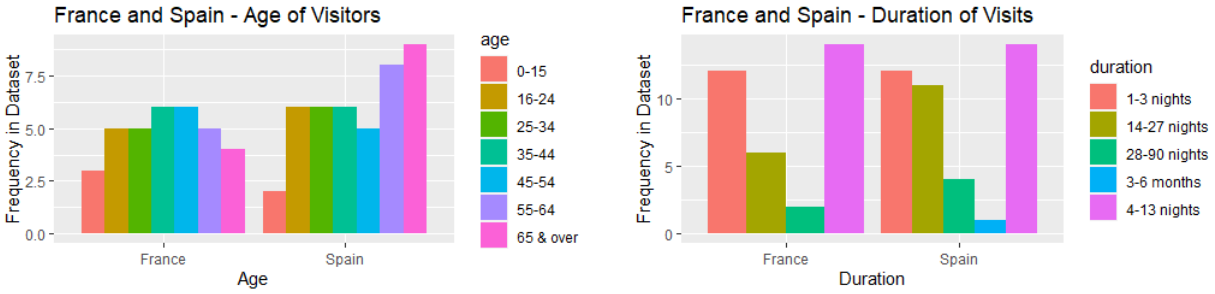
Comparing France and Spain

Comparing spends and nights between these two countries gives the following plots:



From the above, we see that France and Spain follow roughly the same distribution for nights spent and for money spent while abroad. Both have a significantly higher frequency of fewer visits made and smaller amounts of money spent during the visit. This may suggest that shorter visits are popular to both destinations, the reduced spend on such a visit would then come as a natural result of this.

Comparing the ages of visitors and the durations of visits gives the following:



From the above, we see that visitors aged "55-64" and aged "65 & over" are significantly more common visiting Spain than visiting France. Beyond this, the distribution of ages of visitors between countries is the same. This tells us that Spain is the more popular holiday destination for specifically older tourists.
We also see that, though stays of "1-3 nights" and "4-13 nights" are similarly popular between the countries, Spain is significantly more popular as a destination to stay for "14-26 nights". The remaining durations are slightly more popular for Spain than for France. This tells us that Spain is more popular for longer holidays.

In conclusion, we may claim that France and Spain both follow broadly very similar patterns for all the variables examined, except for age of visitors.