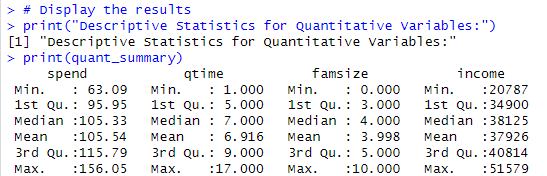
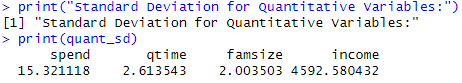
**Introduction:**

In this assignment, we investigate a dataset related to the supermarket shopping habits of UK families. The objective is to obtain meaningful insights and patterns from the data employing statistical methods and visualizations.

**Descriptive Statistics**

Quantitative variables





Displayed above is the table of descriptive statistics for Quantitative Variables which we will summarise below.

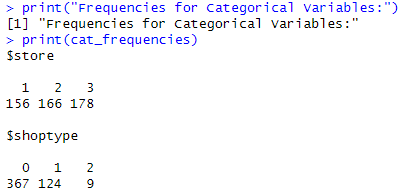
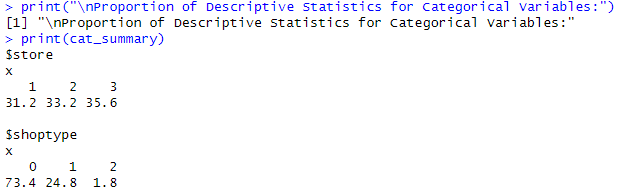
**Spend** - The amount of money people spend on their weekly supermarket shopping varies. On average, people spend around £105.54, but some spend as little as £63 while others spend up to £156. The standard deviation for spending is like a measure of how spread out or varied people's spending habits are. In this case, it is around £15.32.

**Queue Time (qtime)** - The time spent waiting in line at the checkout or collection point also varies. On average, people wait about 7 minutes with some spending only 1 minute in line and others waiting up to 17 minutes. For queue time, the standard deviation is about 2.61 minutes, this tells us how much individual waiting times varies around the average.

**Family Size (famsize)** - Families come in different sizes. On average, a family has about 5 people, but some individuals shop alone and others belong to larger families with up to 11 people. The standard deviation for family size is about 2 people, this means that individual family sizes can differ from the average.

**Income** - On average, it is around £38,125. Some households earn as little as £20,787, while others bring in as much as £51,579. For household income, the standard deviation is about £4592.58, this indicates how much individual household incomes vary around the average income.

In summary, these numbers give us a sense of what is typical for people's spending, waiting times, family sizes and incomes when shopping. The "average" values help us understand what most people experience and the range (from minimum to maximum) shows how much shopping habits can differ from person to person.

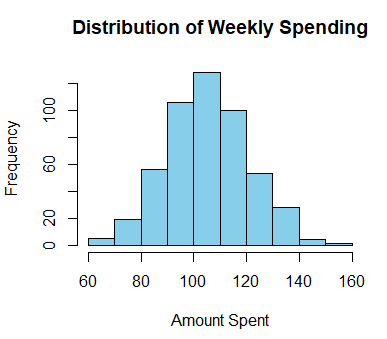
Categorical Variables

Displayed above is the table of descriptive statistics for Categorical Variables which we will summarise below.

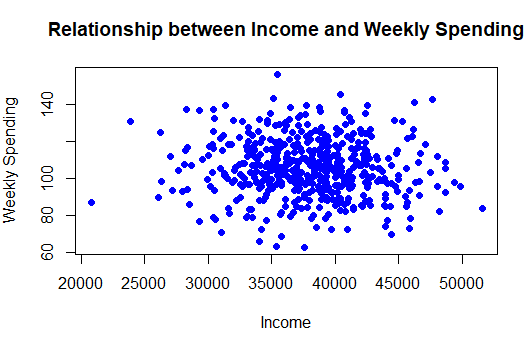
**Store** - For frequencies, there are 156 urban stores, 166 rural stores and 178 city stores in the dataset. The proportions of stores shows that urban stores represent approximately 31.2% of all stores, rural stores make up about 33.2% and city stores are the most prevalent at approximately 35.6%. This information helps us understand the popularity of different types of stores in the dataset.

**Shoptype** - For frequencies, there are 367 instances of Type 0, 124 instances of Type 1 and 9 instances of Type 2. In terms of proportions, Type 0 represents approximately 73.4% of all instances, Type 1 makes up about 24.8%, and Type 2 accounts for around 1.8%. This information provides insights into the distribution of different shopping types in the dataset.

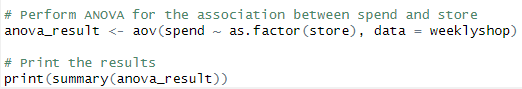
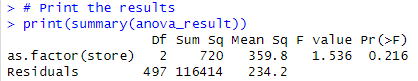
In summary, frequencies show us the raw count in each category and proportions provide a percentage-based view helping us understand the distribution across different stores and shopping types.

**Visualisation of a single variable**

The histogram on the left represents the distribution of the "spend" variable. We can see that most families spend between £90 and £120 with over 100 families spending around £100 per week. There are some outliers such as some families spending as less as £60 and as much as £160.

**A Visualisation that shows the relationship between two variables**

On the left is a scatter plot between Household income and Weekly Supermarket spending. When we look at the dots on the graph, it does not seem that there is any clear pattern as the points are scattered all around which means the correlation is close to 0. In simple terms, it seems that how much money a family makes does not really predict how much they spend on groceries.

**Examining the association between spend and store**

The hypothesis being tested is whether there are significant differences in the mean amount spent on a weekly supermarket shop across different types of stores. To investigate this, we employed a one-way Analysis of Variance (ANOVA).

ANOVA is suitable for comparing means across multiple groups, in our case the different store types. The null hypothesis assumes that there are no significant differences in spending among the various store types while the alternative hypothesis states that at least one store type has a different mean spending. The key point to interpret is the p-value associated with the "store" variable (categorical variable with three levels).

The p-value (0.216) is greater than the typical significance level of 0.05. Since the p-value is greater than 0.05, we fail to reject the null hypothesis. The null hypothesis in this context would be that there are no significant differences in spending among the different store types. In other words, we cannot conclude that the mean spending is significantly different across the three store types.

**Investigating association using correlation between famsize and income**

  
We have chosen to investigate association between famsize and income because there could be a relationship where the family size increases, the household income increases. The correlation coefficient we obtained between "famsize" and "income" is approximately 0.0057. Since the correlation coefficient is close to zero, it suggests a very weak linear relationship between family size and household income. To summarise, this small positive correlation does not imply a strong or meaningful association between the two variables.

**Conclusion**

In this exploration of the dataset on UK family’s supermarket shopping habits, we have gained valuable insights into both quantitative and categorical variables. The descriptive statistics, visualizations and statistical tests provided an understanding of spending patterns, family sizes and incomes offering a glimpse into the diversity of shopping behaviours.