**Alcohol Related Death Analysis**

**Introduction**

This project utilises the following resources:   
  
[Population Estimates for UK, England and Wales, Scotland and Northern Ireland] - <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland>

[Statistics on Alcohol, England] - <https://data.gov.uk/dataset/statistics_on_alcohol_england>

The following project will conduct an in-depth analysis of the Statistics on Alcohol, England provided by data.gov.uk. The main aim of the project is to unveil meaningful information about Alcohol related death across England from 2001 to 2014, making use of the available information regarding gender and specific cause of death as recorded at the time.

This project is a non-profit educational piece being conducted independently by A Humble Devonian.

**Additional Resources**

To bring context to the raw numbers provided in the Statistics on Alcohol, England, additional resources were pulled from the Population Estimates for the UK, England and Wales, Scotland and Northern Ireland. This provides the Population Estimates for England from 2001-2014 which should bring greater clarity to the meaning of the data provided in the main resource.

**Raw Numbers**

The first piece of analysis will take place using the raw numbers. A preliminary glance shows that the number of deaths across males is almost double of that across females (see R Files – 2014RawNumPieChart.R and RawNumBarChart.R)

The raw data however does not take into account the size of the population. Whilst the raw data seems to indicate a gradual increase in the number of alcohol related crimes, when measured against the total population of England we can see the true picture.

**Population Weighted Analysis**

The population of England has not been static between 2001 and 2014 so the raw figures themselves are a limited value in displaying a trend. To bring some meaning to the date the raw numbers must be weighted in terms the population increase. To that end, 2001 will be used as a baseline to calculate the population increase for each year from the base value. The weighted value of the raw numbers will be calculated by:

Weight = 1 / ( pop[year] / pop[2001] )  
Weighted Value = Raw Value \* Weight

This creates a baseline of Weighted Values that can then be compared in real terms. However, for 2001 – 2010 we do not have values for the breakdown of males and females, and so we can only calculate the weighted value as a population as a whole. None the less, this still provides some interesting results.

**Determining Significance**

Observing the line of best fit, we can visibly see there is a slight trend upward, even after the weighted values are taken into account (see R Files – WeightedScatterPlot.R).

However, it may not be a statistically significant trend. Further analysis is required. Fortunately for us, R has its summary function which provides us with everything we need to analyse our Linear Model. Examining the summary for our Linear Model, we can see that according to the p-values there is a low significance correlation between the year and the number of alcohol related deaths (at a 5% significance rate).

From this we can conclude that yes, there does appear to be a unsual increase in alcohol related deaths across the population, even after taking a population increase into account. Obtaining values for 2015 and 2016 when they become available will be interesting to see if they fit into the trend. We can make predictions of what values these years are likely to take, using the Linear Model we have generated.