**Exploratory analysis in R**

**First look at the data**

* X360 is the most frequently used platform in the data set, then PS3 and then PC.
* 2600 and GEN is the least frequently used platform to launch games.
* The relationship between Global, EU and NA sales is roughly linear, as expected.
* EU, NA and Global sales were all skewed towards the right before removal of outliers.
* There were five outlying observations.
* The distribution of Global Sales is still skewed towards the right, but less extreme.
* After removing the outliers, the data points do not seem as tightly clustered around the smoothing curve (ex: Global sales versus NA sales). The relationship between the variables is still linear. There are large amounts of deviation.
  + This could be due to the fact that global sales is made up of EU, NA and ‘Other’ undefined sales.
  + That means that NA sales could make up a smaller/larger proportion of global sales when global sales are high.
* Before removing outliers, NA\_sales and EU\_sales had a positive upward trEND. There were many data points that were far from the trend line indicating a large variance. This also implies that the variables are not very strongly correlated.
* After removing outliers, the trend line has a parabolic shape. The points on the graph do not have a clear linear relationship, implying that NA\_sales and EU\_sales are not strongly correlated.
  + It would be interesting to observe which products differ significantly in sales.
* The is a high frequency of products with the same number of sales in North American and Europe.
* The difference between the sales in the two countries has a median slightly set towards the left of the data values but there is a small tail towards the left.
  + There are more positive differences than negative. Turtle games sold more products in North America than in Europe.

**Impact per product**

* Global sales per prod:
  + Slightly skewed towards the right.
* NA sales per product:
  + Slightly skewed towards the right.
  + Many outliers.
* EU sales:
  + Also skewed towards the right with many outliers.
* Product with the highest sales:
  + Globally: 515
  + EU: 515
  + NA: 948
* Product with the lowest sales:
  + Globally: 518
  + EU: 51510
  + NA: 4491
* Aside from a few outliers, EU sales make up less than fifty percent of that product’s global sales.
* 75% of NA sales make up greater than 40% of that product’s global sales.
  + North America on spends more with Turtle Games than the EU.
* Of the top ten products for Europe, North America and Global sales:
  + 515, 876, 948, 979, 1945 appear in the top ten for all sales.
  + Identified top ‘money-makers’ for Turtle games.
* Of the bottom ten product:
  + None appeared in global, European and north American sales as well.
  + At most, a bottom ranking product appeared in two out the three graphs.
  + For business, this could mean that further research is required to determine why certain products are unpopular.