COMP6234 - Data visualisation

Project Report for “Childhood Obesity in Scotland” Data Story

John **Philip** Wilson

Student: 29767482

Jpw1n17@soton.ac.uk

*This report describes the motives the titled data story and the rationale behind choice of graphics to support it.*

# Data Story Overview

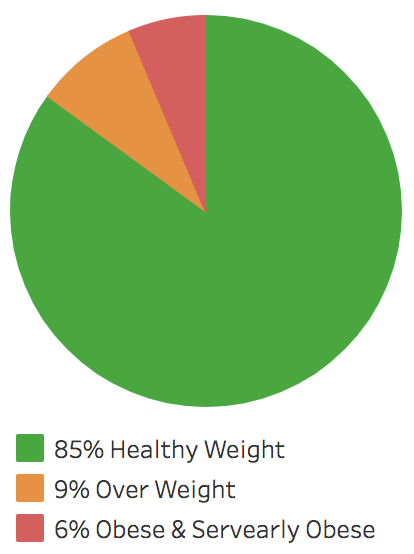
When setting out on creating this data story I intended to explore the following areas found in the data:

1. General prevalence of obesity in Primary 1 children today
2. The national trend over the last decade that the data covers
3. Overview of the regions the data was collected from and when all areas where covered. Highlight the worst and least affected areas.
4. Examine if there is any gender bias.
5. Show how economic background affects a child’s chances of being obese and how the economic down turn has impacted this.

The data story is intended to be in the style of a public service poster to raise awareness of childhood obesity in Scotland. For this reason, the general styling is clean and simple with each graphic focusing on a single point.

# Graphic Details

## General Preverlance

The intention of this graphic is to set the overall context for how prevalent childhood obesity is in primary 1 children. The source data segments the population into 5 different weight categories: -

* Under Weight: 0.4%
* Healthy Weight: 85%
* Overweight: 9%
* Obese: 3.1%
* Severely Obese: 2.4%

I chose a simple pie chart as a means of quickly communicating the relative proportions of each weight category following the Gestalt Theory of Emergency.

I felt displaying all 5 categories would over complicate the graphic and the underweight category would not be visible, so I look for ways to reduce and simplify.

As the main focus of the story is regarding Obesity as a health issue I choose to remove underweight since when rounding to whole numbers, it would be zero anyway. Health issues relating to being overweight become more significant when you cross the obesity threshold so I chose to combine Obese and Severely Obese into one group (OSO).

This left me with three categories ranging from health (good), overweight (concern) and obese or worse (bad). For this reason, the obvious metaphor of traffic lights felt appropriate when choosing the colour assignment.

5% of males suffer from Deuteranopia which affects their ability to distinguish between red and green. To alleviate this issue the layout of the pie chart uses consistent positioning to aid the reading of the legend. The first item in the list is positioned in the pie chart as the first segment starting at 12 o’clock and rotating clockwise, then followed by the second and third segment/list item.

## Trend Over time

I deliberated long and hard over whether to include this graphic as it is a boring flat line. When starting out on this analysis I was expecting to find either an upward or downward trend in the levels of obesity however the data showed that it had stabilised at between 5% and 6% over the last decade. As this contradicted my expectations I felt that this would be interesting to others and that it was necessary to draw attention to this stabilisation with a graphic. The line graph is a simple clear but boring way of showing rends over time. However As in this case the point was to draw attention to the lack of change I felt the metaphor of “flat lining” was a strong way to communicate this.

I chose the colour red for the line so as to remain consistent with the colouring of the OSO category in the previous graphic.

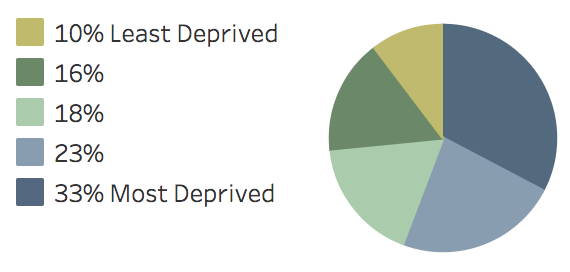
I also added a 95% confidence interval shown as a light grey region behind the line. This did not work very well as the region is to small to be visible. It also interferes with the tooltip highlighting preventing the display of the main lines value. On reflection I would choose to remove this as it does not add any really value and confidence intervals would need some explain to the general public which would be beyond the scope of a short poster of this nature.

## Regional Distribution

I had two key aims with this heat map:

1. Subtlety convey the lack of data collection in the early years of the survey before 2011.
2. That once the survey was established that there were little variations across the regions. The largest variations could be seen in the less populated areas where a small variation has a big impact.

For this reason, I created a custom colour map so that I could colour areas which had no sample data as white in contrast with the main key which ranged from amber (low values) to red (high values). This means there is clear switching of colour when a region transitions from no data to data. I chose the amber to red colour range linking the colour read with areas of most concern.

The regional data I had was captured and the local authority level which by default is not supported in Tableau. To enable this followed these guidelines.

To aid the scrolling I setup a Page which automatically would cycle through the Year filter. Unfortunately, this was not supported by Tableau Public so I had to resort to the manual slide filter.

If I had more time I would have liked to aid dynamic highlighting of regions with the highest and lowest levels of OSO for each year filtered.

Finally, I could have chosen a simple bar graph for this graphic however the heat map also adds implicate information about which regions are rural and urban due to their geographical location

## Gender Bias

The data also provided segmentation with regards to gender. This showed that there tends to be more boys obese than girls however this difference has narrowed from a gap of 0.7% in 2006 to 0.5% in 2015. With this gap being so relatively small I did not feel it was worth mentioning in the story. When you include the overweight category the gap halves going from a gap of 1% in 2006 to 0.4% in 2015. Again, as this is so relatively small it did not feel worth mentioning so this aspect was excluded from the data story.

## Econmonic Influence

The economic background of a child is by far the strongest influence over there likelihood of being obese with the poorest having the highest perseverance. It is no surprise then that the economic down turn of 2008 has had the most impact on the bottom 20% of children.

I had the following points I wanted to communicate:

1. The most deprived children have seen an increase in obesity over the last decade
2. The least deprived have seen a reduction in obesity over the last decade
3. In present day a poor child is 3 times more likely to be obese than that of a child from a wealth background.

For this reason, I chose to show the trend over time with a coloured area graph and the present-day proportions with a simple pie chart. Although the present-day proportions is interpretable from the graph It is difficult to see this straight away whereas this is instantly visible in the pie chart.

I carefully choose a natural colour map that would not communicate any negative context to either end of the spectrum. I also ensured that I used the same colouring for both the graph and pie chart to reinforce that both graphics referred to the same segmentation of the data.

I had experimented with showing all five groups in a line graph however this resulted in a confusing graphic and detracted from the key point which was the divergence of the rich and poor. I wanted to draw attention to this divergence of these two groups. This is visible from the graph it’s self however dose not emphasis it very well and it is difficult to interpret the extent. It’s for this reason I felt the need to add additional annotation to the graph.

I used Gestalt Theory Invariances to draw attention to this annotation. I used a strong contrasting colour with the only inverted colouring on the page. To draw attention further I increase the font size of the “Increased by 12%” and similarly decreased the font size of “Decreased by 19%” text.

To reinforce the relative proportion of the pie chart I also added the percentage ratios to the pie chart legend. Having already introduce the segmentation grouping in the story text and previous graphic, I chose a simpler naming for each of the five segments on the legend.

# Making the Argument

At my heart I’m a strong socialist, I don’t believe in complete equality and feel effort should be rewarded (with some limits) however I do believe in helping those less well off than yourself and giving them the support, they need to develop their own success.

I’m also someone who has suffered from obesity all my adult life and for me this is a hot topic.

So when I discovered in the data how significant an impact economic background had on a child’s chance of being obese I know that this was the key message I wanted to leave the reader with. To this end I wanted to demon straight how economic background was the only key factor which influenced had a significant influence over childhood obesity.

There for I attempted to use “Toulmin’s model of argument” to support this message. To do this I tried to address the four key areas as follows:

* Grounds: I link to the source of the data showing that these are official government statistics from which this data story is derived. I also make several references to other news articles to emphasis the severity of the obesity issue being discussed. The opening of the story has the aim of leaving the read in no doubt that this is a worth topic to be discussed.
* Backing: This is provided in the form of the area graph, showing the divergence in the data of children from different ends of the economic spectrum. This is then reinforced by showing this divergence starts around 2008 after the economic crash.
* Warrants: Having established the validity of the data I now need to provide a connection from the data to the final claim of the data story. I do this by providing a link to the [SIMD quintiles](http://www.gov.scot/resource/0043/00439496.pdf) to support how this type of segmentation links to a child’s economic back ground. The final warrant is made in the form of a pie chart derived from this segmentation making it clear the relative ratios of each economic background.

The final claim in the last paragraph is then well supported by the Grounds, Backing and Warrant above.

# General Improvement

One of the distracting elements of each graphic is the tableau toolbar at the bottom. I had originally managed to remove this as I was using a licensed account. However, once I w=switched to a Tableau Public account the toolbar returned. I followed many online guides detailing different ways by with it could be removed before I finagling found the article which made it clear that this was not possible on the public version. I had concern over weather graphics published on my licensed account would be visible to others so I decided to stick with the public version.

The custom colour map in the geographic heat map could be better. The majority of the data ranges from 3% to 8% however in order to support the missing data and the smaller regions with a larger swing (Orkney at 11.6%) the colour map is spread out over a wider range. If I had more time I would construct a no linear colour map with more of a variation between 3% and 8%.

The addition dynamic linking between text and heat map would also support this between allowing the heat map to switch to the referenced year and highlight the referenced region.

At the moment the html design has targeted a desktop layout. This day and age the buzzword for web content is “mobile first” as mobile has overtaken desktop for the device people use the most. Although the content is viewable on my mobile the layout would benefit from a responsive CSS design to support both desktop and mobile browsers with dedicated layouts targeting their typical aspect ratios.

# Conclusion

The data story is intended as a public health message with a call to political action and there for needed to be accessible to a wider audience.

To this end the data story is deliberately short so as not to lose the read before the final key point is made.

The clean styling and simple graphics used support the message being conveyed well and guide the reader to a solid conclusion.