The many misspellings of fuchsia

Nathan Moroney, Hewlett-Packard Laboratories, USA

Abstract

For nearly a decade, the World Wide Web has been used to collect unconstrained colour terms from thousands of volunteers. The resulting database consists of a red, green and blue triplet and a corresponding colour term. Focused analysis of the colour term fuchsia provides an informative exploration into the nature of colour terms. The colour term fuchsia is visualised as an image where each experimental participant is a pixel. This image can also be presented in frequency-sorted form. Finally, the convergence properties are investigated using the grouped median as a function of the number of subjects. In spite of the inability of a majority of English speakers to spell fuchsia, this colour term exhibits an approximate perceptual convergence with roughly 50 subjects.

Introduction

Fuchsia is both a genus flowering plant (Bartlett, 2002) and a colour term. The plant has a flower ranging from pinkish to purplish. In this chapter the colour term fuchsia is considered as one term, which occurs repeatedly in a database of unconstrained colour terms. These database entries consist of a red, green and blue triplet for a coloured patch displayed on the World Wide Web that elicited the colour term fuchsia. In this way, the colour term is anchored to a given device value. Furthermore, this anchoring has a context or a pragmatic intent, in this case a web-based colour naming experiment.

The experiment comprised seven randomly generated red, green and blue values that were shown on a web page with a white background. The red, green and blue values were selected from the six by six by six uniform sampling of the "web safe" palette, which has been in common use in the last years of the twentieth century. Volunteers were then instructed to provide "the best colour names" for each of the coloured patches. Over 4,000 volunteers have participated in this experiment. In this chapter only the results for the fuchsia colour term will be considered. Further specific details about this experiment have been described elsewhere (Moroney, 2003; Beretta and Moroney, 2008),

Lexical analysis of fuchsia

In this context, fuchsia is a distribution of red, green and blue triplets that have the colour term fuchsia associated with them. The first challenge of this interpretation of fuchsia is that only ten percent of the volunteers could actually spell the word. The many misspellings of fuchsia are shown in a pie chart in

figure 1. To generate this figure, candidate colour terms were selected by searching for the substrings fuc and fus and then manually reviewing the frequency statistics of these terms. The top three misspellings of fuchsia constitute two-thirds of this distribution, but there are also a large number of less frequent misspellings. While the rate of misspelling is potentially amusing, the nature of the distribution of the misspellings is intriguing. The presence of both a small number of predominant terms and a large number of increasingly rare terms are properties seen in processes exhibiting a heavy-tail distribution.

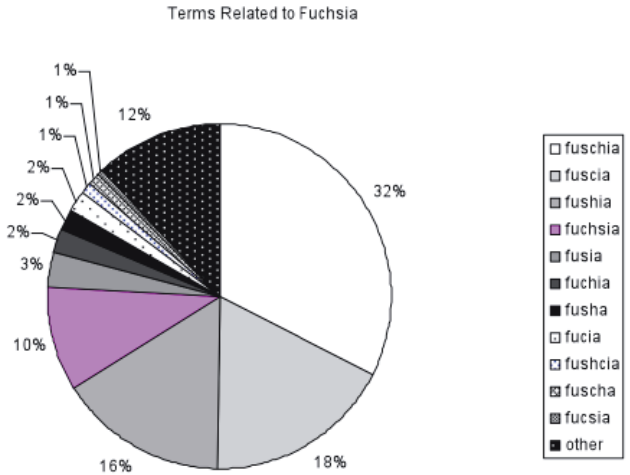

Fig. 1 A pie chart showing the many misspellings of fuchsia. Three misspellings account for 66% of the responses. Only 10% of the participants provided the correct spelling.

Colour terms as images

Given that the misspellings of fuchsia have been identified, it is then possible to return to the original question at the beginning of the introduction: what is the colour fuchsia? As an initial answer a colour visualisation of the corresponding red, green and blue values for the corrected spelling entries in the database including the term fuchsia are shown in figure 2. This figure represents each of the individual entries in the database as a fifty by ten coloured pixels. This block image of pixels is qualitatively informative on a number of levels. First is the distribution of colours for individual pixels, including, in the extreme, a few pixels that to the author appear greenish. Second is that this colour term as an image is still quite effective at communicating the overall perceptual colour corresponding to the term. Finally is the implicit deviance dithering of this image. This experiment included a large number of subjects, displays and colour proficiency and this image is a direct visualisation of this deviance.

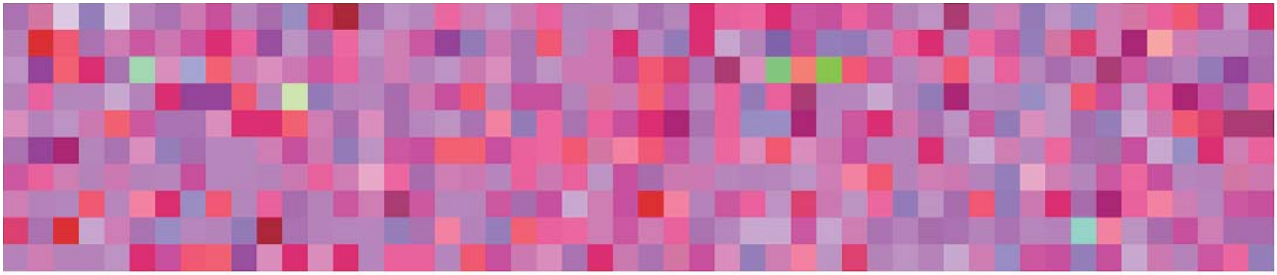

Fig. 2 The colour term fuchsia visualised as an image where individual pixels correspond to individual experimental subjects.

The raw results shown in figure 2 can be refined by taking the data and re-arranging the corresponding red, green and blue values. In this way additional trends in the data can be visualised. Figure 3 shows the result of sorting and rearranging the colour pixels from the more random colour pixels in figure 2. In this way the more frequent colours are shown on the left and less frequent colours are shown on the right. This image provides a visualisation of both the clustering and the variation of the data. Qualitatively, the number of genuinely disruptive participants or patches of the opposite hue appears to be in the order of 1% of the data. This result shows that coloured patches that might be described as pinkish, purplish and reddish predominate. This data can form the basis of statistical and quantitative analysis of fuchsia.

Fig. 3 The colour term fuchsia visualised as a frequency-sorted image where individual pixels correspond to individual experimental subjects. The regions of colour are arranged from more to less frequent from left to right.

Convergence of fuchsia

The previous section provides a visual definition of the colour term fuchsia that includes colour term as image and the image pixels re-arranged according to a frequency sorting. However, given this data is it possible to estimate how quickly the colour term converges? If we have perceptual anchors in the form of red, green and blue anchor values, how small of a difference can be achieved by analysis of a given number of participants?

Figure 4 shows three sub-plots of the red, green and blue data corresponding to fuchsia. These figures are histograms for the given colour channel. The x-axis is the digital count and the y-axis is the proportion or frequency of colours with that digital count. The results for red are shown in the upper left, for green in the upper right and for blue in the lower left. In all cases the distributions are relatively smooth and bounded by one end the scale or other. Qualitatively these distributions have differing shapes; the red is steeper while the blue is less steep. This indicates that a relatively narrower range of red values correspond to fuchsia while a wider range of blue values occurs. These distributions are also not well modelled using a normal distribution and therefore the arithmetic mean is not an appropriate measure of the central tendency of these distributions.

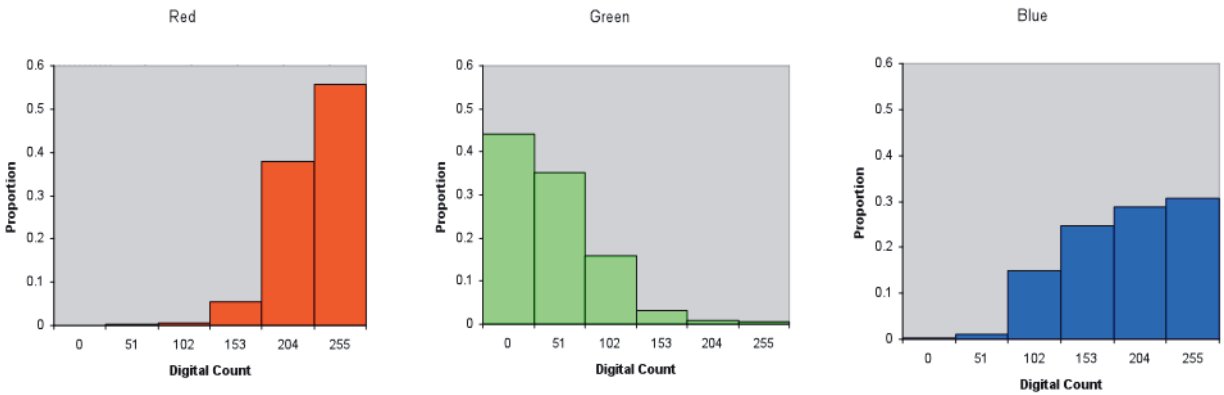

Figure 4. Distributions of red, green and blue digital counts for the colour term fuchsia.

For this analysis the grouped median (Black, 2009) was used as a measure of the centre of the distribution. The grouped median was computed:

$$GroupedMedan = x_1 + \frac{width \cdot \left(\frac{n}{2} - f_1\right)}{\left(f_2 - f_1\right)} \quad (1)$$

where x1 is the real lower limit of the median interval, width w is the size of the interval, n is the population size, f1 is the cumulative frequency count of the interval containing the median, and f2 is the cumulative frequency count for the interval following the one containing the median. The use of the grouped median as opposed to the ungrouped median results in a measure of the central tendency of

the distribution that is not one of the original, highly quantised colour values. This value can then be plotted versus the number data points used to compute the grouped median. This is shown in figure 5 and is a visualisation of the convergence of the red, green and blue channel values for the colour term fuchsia.

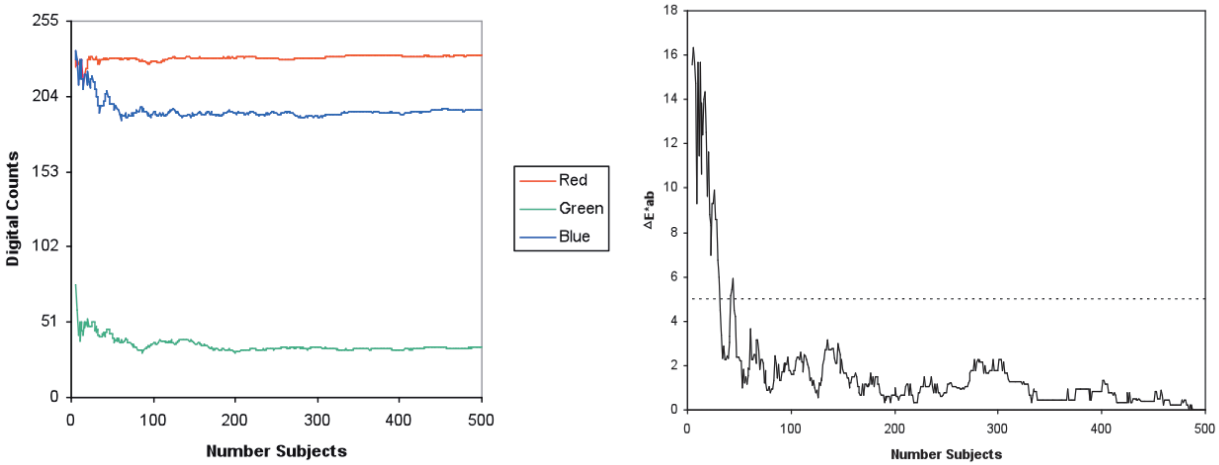

Fig. 5 (left) Convergence of the grouped median of the red, green and blue values for the term fuchsia.
Fig. 6 (right) Convergence of the grouped median of the CIELAB colour difference for the term fuchsia.

Another way of looking at the convergence of the term fuchsia is to plot the colour difference between the grouped median of the entire sample population versus the grouped mean of portions of the population. This then provides a measure of convergence that is not computed on a channel by channel basis. Figure 6 shows a plot of this convergence, where the x-axis is the number of subjects used to compute the grouped median, and the y-axis is the corresponding CIELAB colour difference between the sub-population under consideration and the entire population. For all computations the grouped median was still computed on a channel by channel basis. The resulting red, green and blue values were then assumed to be sRGB values (IEC, 1999) and converted to CIELAB values. This figure also shows a dotted line at the 5 DE*ab level. This value is of interest because it is an approximate threshold used in image colour difference evaluation, and is roughly the same as the variation seen in the real world, such as cereal boxes and lemons (Moroney, 2006).

The results shown in figure 6 show a rather rapid convergence of the colour coordinate corresponding to the colour term fuchsia. On the order of 50 subjects are enough to achieve a grouped median colour difference of around 5. This result was also confirmed through a repeated iterative pseudo-random shuffling of the database to compute a smoothed average curve. So while relatively few people may be able to spell fuchsia, the result of analysing the data provided by 50 people is quite consistent. In spite of a lack of lexical accuracy, the colour term fuchsia has a fairly robust corresponding perceptual anchor, as seen with the convergence of the colour coordinates and differences.

Conclusions

Only 10% of English speakers can correctly spell the colour term 6fuchsia. This colour term can be directly visualised by converting individual experimental data into coloured pixels in an image. This fuchsia-as-image can also be re-arranged into a frequency sorted image. These images show both the variation in the red, green and blue values that correspond to the colour term fuchsia and monotonic red, green and blue histograms. The grouped median was used with this heavily quantised data and the end result is that roughly 50 subjects are enough to get an estimate of fuchsia that has converged to within 5 DE*ab units. In spite of the lack of lexical accuracy, there is a relatively rapid convergence for the colour coordinates for the colour term fuchsia.

References

BARTLETT, G., (2002), Fuchsias: A Colour Guide, Wiltshire, UK, Crowood Press Ltd.
BERETTA, G., MORONEY, N. (2008), Cognitive aspects of color, HP Labs Technical Report, HPL-2008-109, 1-26.
BLACK, K., (2009), Business Statistics: Contemporary Decision Making, 6th Edition, New York: John Wiley and Sons. 71-72.
INTERNATIONAL ELECTROTECHNICAL COMMISSION, (1999), Part 2-1: Colour Management – Default RGB Colour Space – sRGB, First Edition, IEC 61966-2-1.
MORONEY, N. (2003). Unconstrained web-based color naming experiment in R. ESCHBACH and G. G. MARCU eds. Color Imaging VII: Processing, Hardcopy, and Applications, Bellingham, WA, SPIE 36-46.
MORONEY, N. (2006). Uncalibrated color in R. ESCHBACH and G. G. MARCU eds. Color Imaging XI: Processing, Hardcopy, and Applications, Bellingham, WA, SPIE 69-74.