Beuth University of Applied Sciences Berlin

Project 1 for Data Visualization ST 2019

Color Measurements

Authors: Florian Becker

Rupali Sharma

Kerstin Wagner

Date: 26/05/19

Table of Content

[I. Abstract 2](#_Toc9793350)

[II. Introduction 3](#_Toc9793351)

[III. Materials and Methods 4](#_Toc9793352)

[A. Algorithms 4](#_Toc9793353)

[B. Datasets 5](#_Toc9793354)

[C. R Packages 5](#_Toc9793355)

[D. Procedure 6](#_Toc9793356)

[IV. Results 6](#_Toc9793357)

[A. Statistics 6](#_Toc9793358)

[B. Hypothesis A 7](#_Toc9793359)

[C. Hypothesis B 9](#_Toc9793360)

[V. Discussion 11](#_Toc9793361)

[A. Hypothesis A 11](#_Toc9793362)

[B. Hypothesis B 11](#_Toc9793363)

[VI. Conclusion 12](#_Toc9793364)

[VII. References 12](#_Toc9793365)

# Abstract

… will be written at the end …

# Introduction

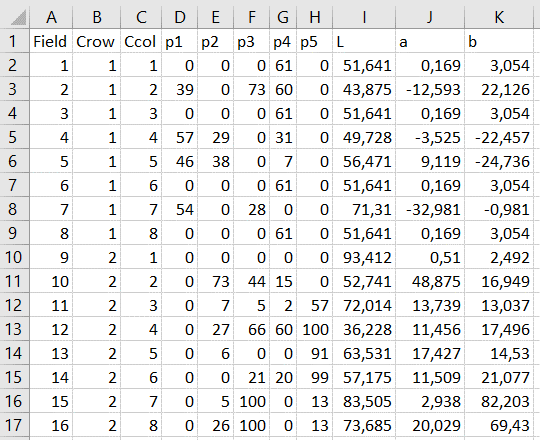
Our first project in Data Visualization is about evaluating printed colors cards in comparison with a master color card. The cards are from Douglas for determining skin color.

We have two datasets:

* MasterColorCard.csv (MCC) contains intended colors for color card production
* LabMeasurements-Color-Card.csv (LMCC) contains the measurements for 13 sheets with each has 42 color cards printed on in 7 rows and 6 columns

Figure 1 and Figure 2 show the structure of the datasets for a better understanding of the data:

* The MCC gives us the colors in two color spaces: CMYK including a special skin color for the print and calculated CIELAB for comparing the measurements with the LMCC.
* The LMCC only uses the CIELAB color definition.



CIELAB Definition

of one color

CMYKS Definition

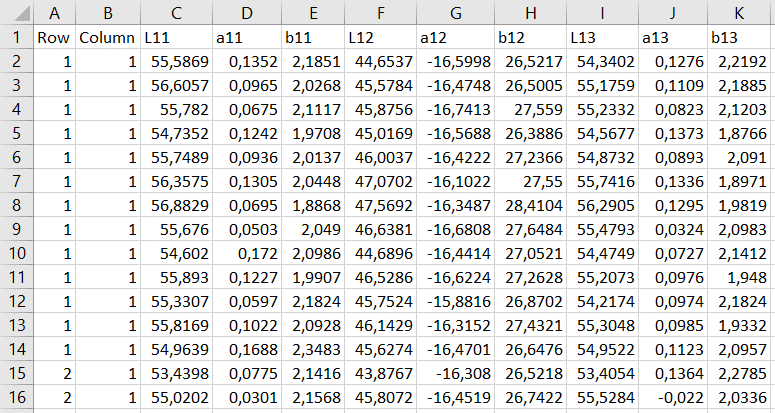
of one color

Position on the

mastercard



Figure 1: Structure of MasterColorCard.csv



First row and first column

(first card)

of all the 13 sheets

CIELAB Definition

of the 1st color

CIELAB Definition

of the 2nd color

Figure 2: Structure of LabMeasurements-Color-Card.csv

The first aspect, that we wanted to cover in the project, was the similarity between colors, cards and sheet with the MCC as given baseline.

The second aspect was to get a better understanding of different calculation and visualization of the similarity: what works good and is easily accessible and what is hardly readable and interpretable.

We calculate the distance by ∆E and the similarity by cosine similarity. For more details about the algorithms we have choosen, see “IV.B Algorithms”, p. 4).

The calculated results are visualized in histograms, boxplots, density plots, scatterplots and violin plots (see “V. Results”, p. 5)

Our hypothesis we wanted to proof or reject:

.

# Materials and Methods

## Algorithms

### Delta E 2000

The CIELAB color definition allows to calculate perceptual differences. “∆E is a metric for understanding how the human eye perceives color difference” [1]. Because Delta E 2000 is the current state-of-the-art, we used that algorithm for a first calculation of distances.

For the result are different interpretations possible. Figure 3 shows two possible: a more wide-ranged overview on the left [1], and one focussed on the printig quality on the right (errors between ΔE 2 to 4 are considered accurate implementation) [2].

Lots of relevant factors, for example the sample size or the experience of the observer, but also the time somebody need to recognize a difference to be considered [2].

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Delta E | Perception | | <= 1.0 | Not perceptible by human eyes. | | 1 - 2 | Perceptible through close observation. | | 2 - 10 | Perceptible at a glance. | | 11 - 49 | Colors are more similar than opposite | | 100 | Colors are exact opposite | | |  |  | | --- | --- | | Color Difference | Impact | | < 0.2 | Not visible | | 0.2 - 1.0 | Very low | | 1 - 3 | Low | | 3 - 6 | Medium | | > 6 | High | |

Figure 3: Two Color Difference Ratings

### Cosine similarity

The cosine similarity calculates the angle between two vectors, the result ranges from −1 (exactly opposite) to 1 (exactly the same).

## Datasets

We have decided to shrink the data we are working with and focus on the skin colors in the middle of the card and to leave the border color patches out, because they don’t seem to be relevant for determining a skin color.

The four grey patches in the middle of the mastercard represent the hole for the customer’s skin – that’s why they are also not relevant.

Figure 3 shows the remaining 32 of 64 colors that we have used for the comparions.



Figure 3: All Colours vs. Tested Colors

## R Packages

We use some additional packages for the project:

|  |  |  |
| --- | --- | --- |
| Package | Used function | Reason |
| data.table | fread | Fast import of the csv files with automatic controls detection, e.g. separator [2] |
| colorscience | deltaE2000 | Calculates the distance of two colors in CIELAB [3] |
| tcR | cosine.similarity | Calculates the similarity of two vectors [4] |

## Procedure

The procedure for distances and similarities is the same:

* Create the necessary data.frames with calculated distances / similarities.
* Plot and save the different charts as png in the folder “Images”.
* Use the same colors for the sheets in the charts.
* Use the colors of MCC for plotting color patch relevant charts.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |

Figure 4: Color Legend for Sheets

# Results

For describing the deviation we have chosen minimum, maximum, mean, median and number of outliers.

Section A shows these statistics for different data groups, which can be found partially in the charts.

## Statistics

### For all sheets by colors

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Distance | | | | | Similarity | | | | |
|  | Min | Max | Mean | Median | Outl. | Min | Max | Mean | Median | Out. |
| All | 0.047 | 26.759 | 2.476 | 1.383845 | 549 | 1.092 | 2.638 | 2.009 | 2.036915 | 0 |

Table 1: For all sheets by colors

### For all sheets by card

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Distance | | | | | Similarity | | | | |
|  | Min | Max | Mean | Median | Outl. | Min | Max | Mean | Median | Out. |
| All | 2.207 | 2.970 | 2.476 | 2.482172 | 3 | 1.989 | 2.027 | 2.009 | 2.009369 | 6 |

Table 2: For all sheets by card

### For each sheet

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Distance | | | | | Similarity | | | | |
| Sh. | Min | Max | Mean | Median | Outl. | Min | Max | Mean | Median | Out. |
| 1 |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |

Table 3: For each sheets

### For each colors

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Distance | | | | | Similarity | | | | |
| Col. | Min | Max | Mean | Median | Outl. | Min | Max | Mean | Median | Out. |
| 1 | 0.256 | 3.11 | 1.271 | 0.6215896 | 1 | 2.576 | 2.625 | 2.599 | 2.599173 | 0 |
| 2 | 0.216 | 1.998 | 0.66 | 3.565043 | 9 | 1.719 | 1.766 | 1.744 | 1.743961 | 1 |
| 3 | 2.622 | 4.773 | 3.567 | 0.6839234 | 5 | 2.485 | 2.525 | 2.503 | 2.501948 | 0 |
| 4 | 0.24 | 2.27 | 0.748 | 1.512918 | 11 | 1.957 | 2.018 | 1.991 | 1.992313 | 3 |
| 5 | 0.719 | 2.661 | 1.481 | 1.055471 | 1 | 2.086 | 2.133 | 2.113 | 2.114288 | 2 |
| 6 | 0.393 | 1.837 | 1.046 | 25.31076 | 3 | 2.056 | 2.081 | 2.066 | 2.065985 | 4 |
| 7 | 23.855 | 26.759 | 25.294 | 1.234916 | 6 | 1.371 | 1.398 | 1.386 | 1.386039 | 7 |
| 8 | 0.579 | 3.186 | 1.223 | 4.505533 | 4 | 1.667 | 1.722 | 1.692 | 1.691421 | 5 |
| 9 | 3.538 | 5.583 | 4.503 | 0.8432311 | 9 | 2.58 | 2.638 | 2.61 | 2.608857 | 0 |
| 10 | 0.098 | 2.226 | 0.825 | 1.229698 | 2 | 1.799 | 1.847 | 1.826 | 1.826865 | 3 |
| 11 | 0.562 | 2.237 | 1.238 | 1.811967 | 4 | 2.156 | 2.22 | 2.193 | 2.193381 | 3 |
| 12 | 1.044 | 3.01 | 1.849 | 2.562715 | 3 | 2.199 | 2.267 | 2.233 | 2.234101 | 3 |
| 13 | 1.437 | 4.658 | 2.63 | 3.031098 | 6 | 2.462 | 2.511 | 2.488 | 2.488132 | 0 |
| 14 | 2.31 | 4.2 | 3.047 | 0.52415 | 10 | 2.42 | 2.461 | 2.441 | 2.440897 | 0 |
| 15 | 0.047 | 2.145 | 0.591 | 0.6040576 | 8 | 1.838 | 1.896 | 1.868 | 1.867933 | 0 |
| 16 | 0.069 | 1.535 | 0.643 | 1.037628 | 1 | 1.646 | 1.688 | 1.669 | 1.669991 | 2 |
| 17 | 0.528 | 2.006 | 1.077 | 1.936186 | 5 | 1.582 | 1.627 | 1.607 | 1.608216 | 15 |
| 18 | 0.229 | 5.185 | 2.018 | 2.018438 | 0 | 1.092 | 1.127 | 1.108 | 1.107621 | 2 |
| 19 | 0.285 | 4.976 | 2.067 | 2.332186 | 0 | 1.092 | 1.124 | 1.107 | 1.107537 | 0 |
| 20 | 1.517 | 4.587 | 2.416 | 0.8583082 | 4 | 2.375 | 2.43 | 2.402 | 2.402575 | 0 |
| 21 | 0.208 | 1.978 | 0.887 | 1.208014 | 2 | 1.794 | 1.843 | 1.821 | 1.821359 | 4 |
| 22 | 0.842 | 1.866 | 1.226 | 0.8678517 | 2 | 1.686 | 1.729 | 1.706 | 1.706272 | 1 |
| 23 | 0.412 | 1.823 | 0.927 | 1.545201 | 2 | 1.681 | 1.723 | 1.705 | 1.70529 | 2 |
| 24 | 0.825 | 3.21 | 1.59 | 2.303437 | 7 | 2.146 | 2.197 | 2.176 | 2.176179 | 2 |
| 25 | 1.187 | 4.615 | 2.427 | 1.014791 | 1 | 2.44 | 2.492 | 2.468 | 2.468607 | 0 |
| 26 | 0.596 | 2.905 | 1.094 | 0.7054563 | 29 | 1.878 | 1.954 | 1.92 | 1.920033 | 1 |
| 27 | 0.237 | 1.805 | 0.784 | 1.269947 | 12 | 1.932 | 1.982 | 1.957 | 1.957615 | 0 |
| 28 | 0.418 | 2.689 | 1.31 | 0.968093 | 11 | 2.112 | 2.178 | 2.147 | 2.14874 | 4 |
| 29 | 0.435 | 1.948 | 1.032 | 2.911953 | 0 | 1.671 | 1.711 | 1.692 | 1.692088 | 0 |
| 30 | 2.09 | 4.155 | 2.935 | 3.11869 | 7 | 2.414 | 2.455 | 2.436 | 2.436329 | 2 |
| 31 | 2.217 | 4.396 | 3.115 | 3.715516 | 3 | 2.319 | 2.371 | 2.348 | 2.349743 | 0 |
| 32 | 3.131 | 4.387 | 3.729 | 0.6215896 | 3 | 2.238 | 2.279 | 2.26 | 2.260672 | 6 |

Table 4: For each color

## Hypothesis A

.

Histogram: contains a combined distribution over all 13 sheets with the distribution of the mean distances (mean of distances over all colors) of each card.

Box, Violin, and Density: contain 13 colors for 13 sheets where each box/violin/density\_line represents the distribution of the mean distances (mean of distances over all colors) of each card on the sheet from the master card: 42 points (for 42 cards) for each sheet.

The similarity plots are much more varied than the distance ones.

The violin plots have similar readability in terms of box region and means to the boxplots, but the outliers are not visible because the violin includes them, whereas in the boxplots, the outliers are clearly visible, so between these two types of blots, boxplots are better for outlier detection, otherwise they work similarly.

### Histogram

|  |  |
| --- | --- |
| Distance | Similarity |
|  |  |

### Boxplot

|  |  |
| --- | --- |
| Distance | Similarity |
|  |  |

### Violin Plot

|  |  |
| --- | --- |
| Distance | Similarity |
|  |  |

### Density plot

|  |  |
| --- | --- |
| Distance | Similarity |
|  |  |

## Hypothesis B

Four colors have been selected (37, 57, 65, 77): four plots on one representing the plots for resp. colors

Histogram: each box represents the distribution of the distances of resp. color on each card on the sheet from the resp. color on the master card: 42 points (for 42 cards) for each sheet

Box, Violin, and Density: contain for each sheet the distance/similarity between those colors from each card makes for a point in the box/violin/density plot.

### Histogram

|  |  |
| --- | --- |
| Distance | Similarity |
|  |  |

### Boxplot

|  |  |
| --- | --- |
| Distance | Similarity |
|  |  |

### Violin Plot

|  |  |
| --- | --- |
| Distance | Similarity |
|  |  |

### Density plot

|  |  |
| --- | --- |
| Distance | Similarity |
|  |  |

# Discussion

## Hypothesis A

.

## Hypothesis B

# Conclusion

Some more before

To improve the results, it would be helpful to go more in the detail regarding the different meanings of Delta E 2000 and cosine similarity. A scatterplot “distance vs similarity” can visualize possible relationships between these statistics, e.g. for each color.

# References

|  |  |
| --- | --- |
| [1] | Z. Schuessler, “Delta E 101,” 11 11 2016. [Online]. Available: http://zschuessler.github.io/DeltaE/learn/. [Accessed 24 05 2019]. |
| [2] | M. Dowle, “Package 'data.table',” 07 04 2019. [Online]. Available: https://cran.r-project.org/web/packages/data.table/data.table.pdf. [Accessed 25 05 2019]. |
| [3] | J. Gama and G. Davis, “Package ‘colorscience’,” 25 07 2018. [Online]. Available: https://cran.r-project.org/web/packages/colorscience/colorscience.pdf. [Accessed 25 05 2019]. |
| [4] | V. Nazarov, “Package 'tcR',” 25 03 2019. [Online]. Available: https://cran.r-project.org/web/packages/tcR/tcR.pdf. [Accessed 25 05 2019]. |