Yosuf I.Q. Barzinji
Image Color Analysis

Cultural Data Science

Date: January 04 5:00 PM
Topic - The Greyscale Epidemic

- Temporal Degradation of Color -

Investigating the Hypothetical Disappearance of Colorfulness from Clothing Over Time: Implications for Aesthetics and Culture

Yosuf Ismael Qasim Barzinji (202209093@post.au.dk)

Claude Monet, Twilight, Venice, 1908 (Artizon Museum). Grayscale effect added by Yosuf

School of Communication and Culture, Aarhus University,
Jens Chr. Schous Vej, 8000 Aarhus, Denmark
Supervisor: Joshua Skewes

Yosuf I.Q. Barzinji
Image Color Analysis

Cultural Data Science

Date: January 04 5:00 PM
Topic - The Greyscale Epidemic

Abstract

Exploring the colorfulness of everyday clothes across time to investigate the hypothetical decline of vibrant hues in the world of aesthetics proved to be quite challenging in terms of finding accurate data. Using color image extraction methods to quantify colorfulness, an insignificant decline appeared, regardless of the diverse methods of calculating colorfulness. While this decline could potentially be explained by insufficient data, the agreement across multiple models on a negative correlation may suggest a deeper insight into the true direction of the regression. However no null hypothesis has been rejected.

Yosuf I.Q. Barzinji Cultural Data Science Date: January 04 5:00 PM
Image Color Analysis Topic - The Greyscale Epidemic

Introduction

Modern-day society presumes that the world is turning monochrome or, otherwise, "grey", a black-and-white infestation potentially linked to the societal interest in minimalism *(Gonzalez, 2023)*.

This problem specifically pertains to day-to-day items, whether in fashion, the kitchen, or company aesthetics, with articles often presenting these claims without real statistical foundations. *(Bang, 2023).* A few attempts to quantify color trends over time have seen some success, such as the Science Museum Group Collection project. This initiative involved analyzing and plotting the colors of a large number of historical items from 1800 to 2020, showing a wave of greyness and reduced variation. *(Sleeman, 2021)*

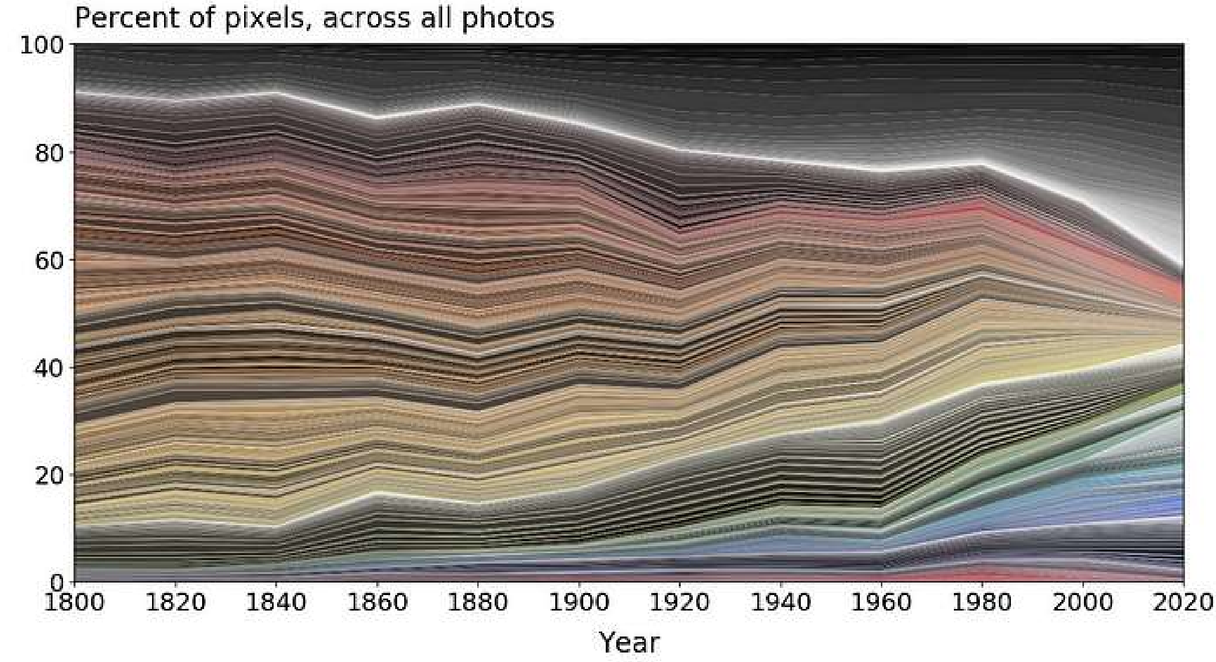

Figure 1 - Science Museum Group Collection *(Sleeman, 2021)*

While the causes and effects of this cultural shift are open to multifaceted debates, whether it is a change in the cultural western norms or a product of global shifting trends, its significance remains unclear. Among the many aspects impacted by this "greyscale epidemic," this paper focuses specifically on the color choices in daily fashion and clothing available through modern web shopping.

Methods

After evaluating multiple datasets, the Fashion Product Images Dataset was selected for its extensive clothing data (40,000+ rows) and metadata, such as categorical clothing identifiers and date, which are both essential. (*Fashion Product Images Dataset*, 2019)

The methodology aimed to quantify image colors into a measurable "colorfulness" metric that could be compared over time. While various established methods exist, such as dominant color detection and pixel color aggregation, most tend to be overly rigid, potentially introducing bias in the color analysis. Instead, the chosen method was RGB color averaging, which involves calculating the RGB values of each pixel and averaging them to represent the image. The process of quantifying colorfulness can be distilled into two key steps: *Image scaling* and *RGB conversion.*

To reduce noise, the image center was cropped for consistent color extraction. (Appendix A - Detailed Explanation)

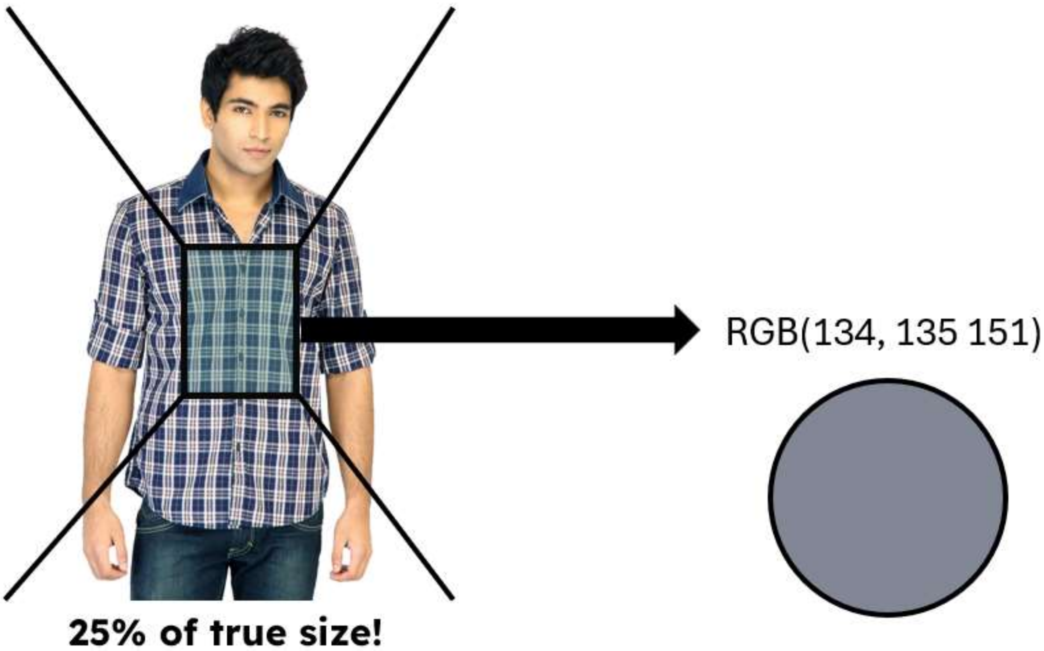

Figure 2 - Scaling Visualized; Image licensed from the Kaggle dataset

Secondly, once an RGB code is obtained, numerous methods of converting it into a "colorfulness" unit are available. For the sake of maintaining a concise yet exploratory analysis, two methods were considered:

$$CF = Colorfulness$$

$$RGB = RED,\ GREEN,\ BLUE$$

$$SB = Saturation,\ Brightness$$

Yosuf I.Q. Barzinji Cultural Data Science Date: January 04 5:00 PM
Image Color Analysis Topic - The Greyscale Epidemic

1. Hasler's Equation: Proposed by Hasler and Süsstrunk (2003), this equation is designed to evaluate the perceptual impact of color diversity within an image.

$$CF_{Hasler} = \sqrt{(R - G)^2 (G - B)^2 (B - R)^2}$$

2. Saturation-Brightness (SB) Calculation: A simpler approach that combines the saturation and brightness values of the image to derive its colorfulness.

$$CF_{SB} = S * B$$

Both methods provide unique perspectives on quantifying colorfulness. Hasler's equation offers a more nuanced analysis of color variation, capturing the perceptual impact of diverse colors, while the Saturation-Brightness (SB) calculation highlights the broader difference between grayscale and the corresponding color. Together, these contrasting approaches allow for a speculative yet informative comparison in the analysis.

Four Python scripts were developed for data merging, filtering, color extraction and color analysis, with an R script for statistical evaluation. (Appendix B - Detailed Description)

While the dataset includes a "baseColor" column for categorizing clothing by color, it lacks the capacity to capture variations within similar shades, such as differences among "red" tones(e.g. Maroon, Crimson, etc). Nonetheless, these labels proved useful during modeling and graphing, adding an extra explorative parameter to the study. To address their limitations, both Hasler's and SB colorfulness measures were compared with values derived from the "baseColor" column.

Yosuf I.Q. Barzinji
Image Color Analysis
Cultural Data Science
Date: January 04 5:00 PM
Topic - The Greyscale Epidemic

Results

Variable Distributions and Linear Modelling

Variable Distribution

An analysis of gender distribution within the dataset revealed that the majority of clothing items were designed for men and women. Specifically, 49.87% of the sample consisted of men's clothing, while women's clothing accounted for 41.9%. Although gender is not the primary focus of this analysis, assessing the distribution was necessary to ensure that the dataset does not exhibit any disproportionate representation that could introduce bias or random variance in subsequent analyses.

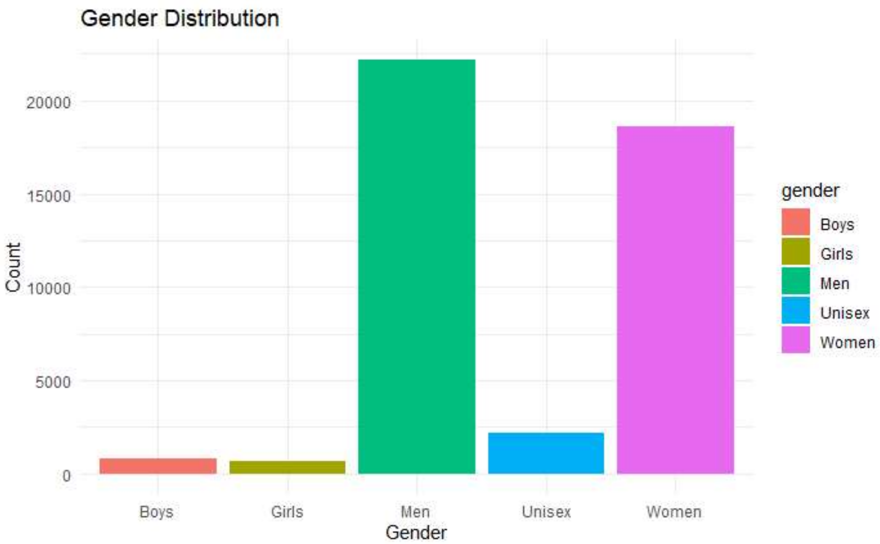

Figure 3: Gender Distribution

The year distribution was the most surprising variable, as the variation was not clearly apparent at first. Upon closer examination, it was observed that more than 91% of all data points lied notably at the specific periods 2011.50, 2011.75, and 2012.50. The decimals in these year values represent the seasons, with 0.25 indicating spring and 0.75 indicating fall. While there was still a notable number of data points from the period 2015 to 2016, the topwear distribution was predominantly concentrated in the 2011 to 2013 range.

Yosuf I.Q. Barzinji
Image Color Analysis

Cultural Data Science

Date: January 04 5:00 PM
Topic - The Greyscale Epidemic

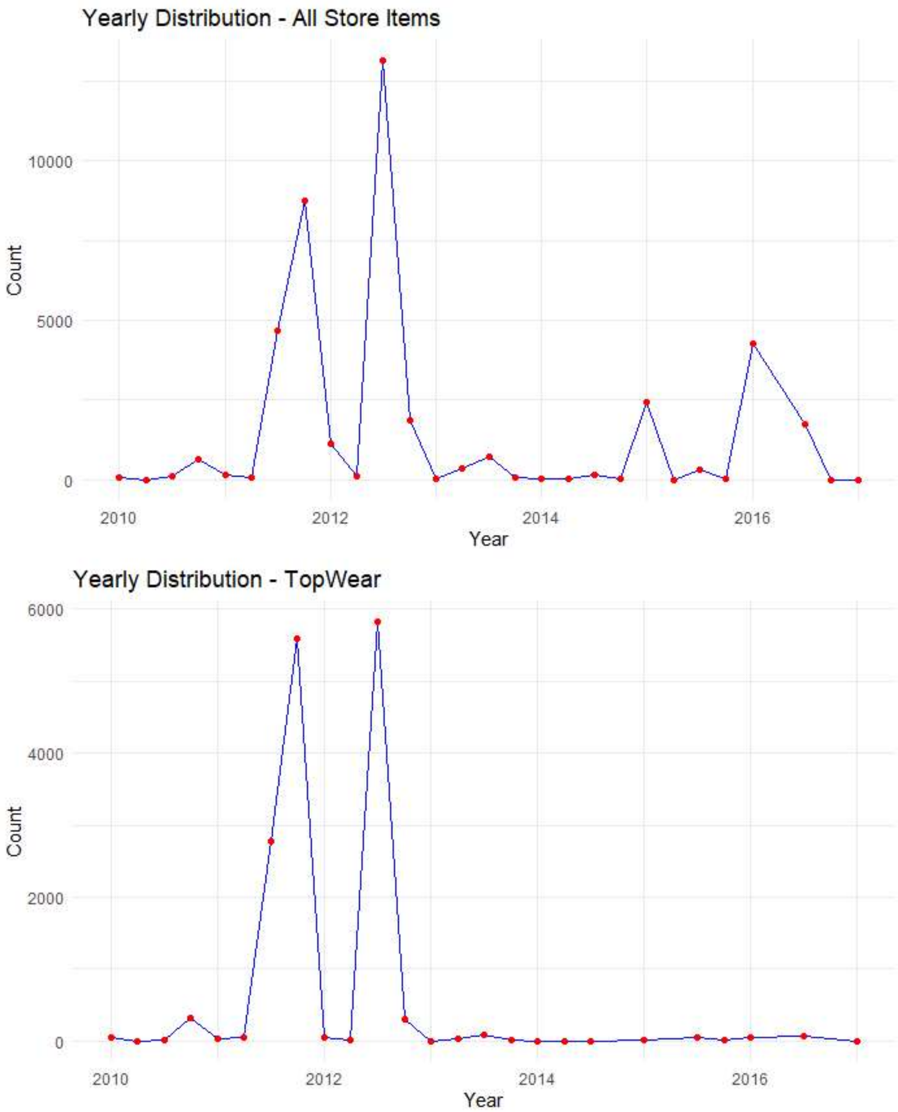

Figure 4: Year Distribution of Topwear Items

Linear Modelling

This section will present four graphs in total, each displaying the respective linear model coefficients in the top-left corner of the regression plot. The four graphs correspond to two equations for colorfulness and two sets of RGB color codes.

Yosuf I.Q. Barzinji Cultural Data Science Date: January 04 5:00 PM
Image Color Analysis Topic - The Greyscale Epidemic

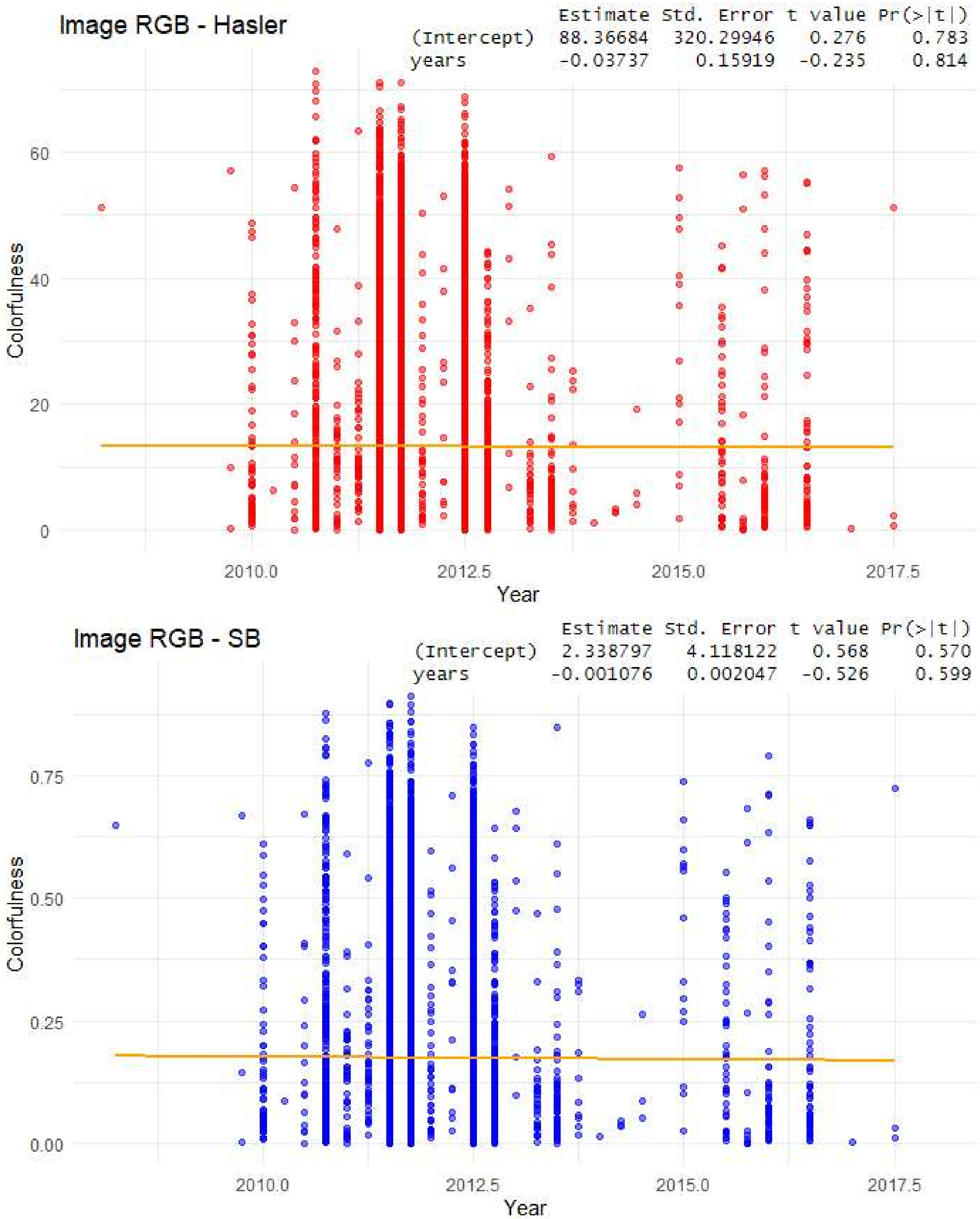

Figure 5 - Hasler and SB image RGB models and coefficients

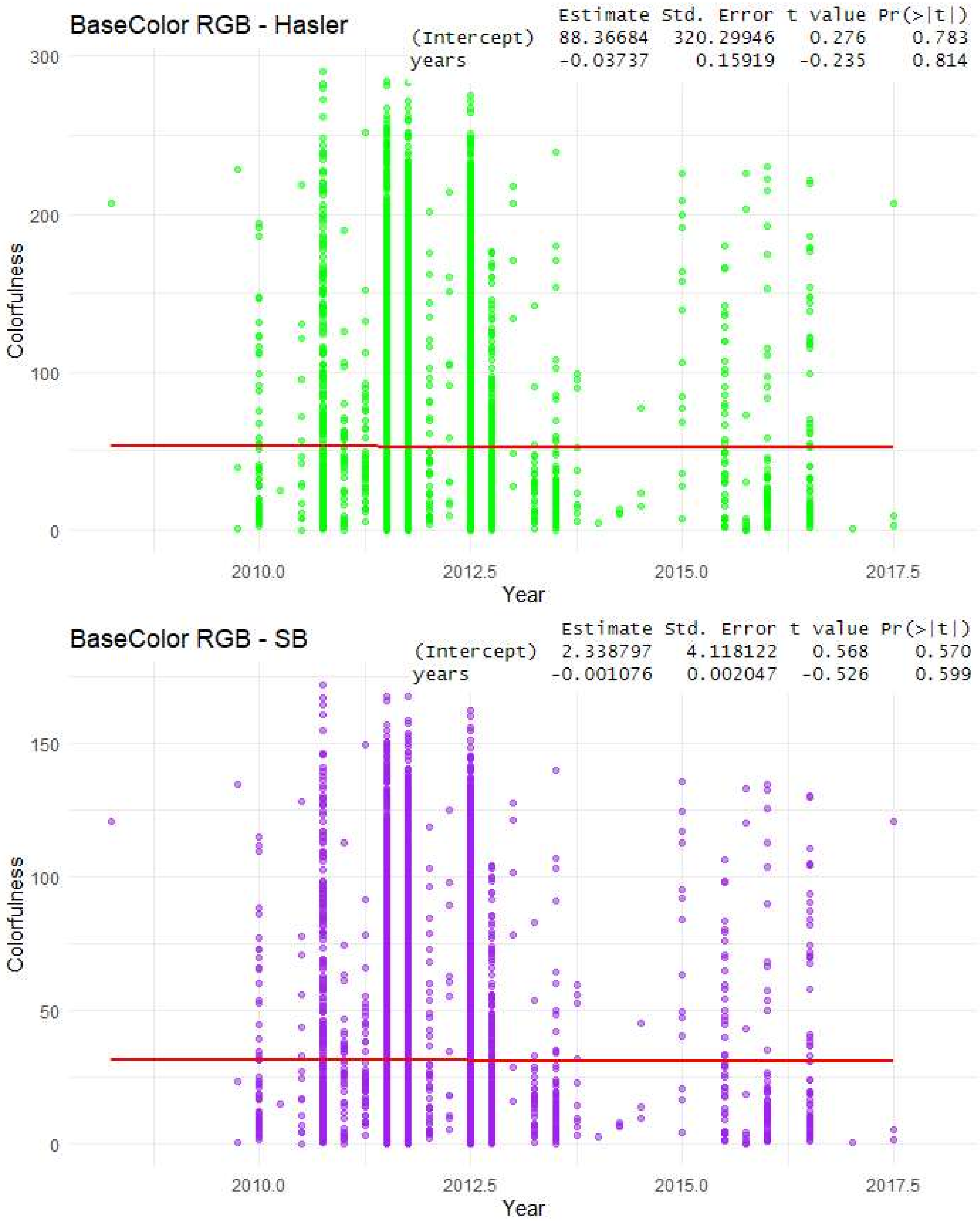

Figure 5 - Hasler and SB ColorBased RGB models and coefficients

Yosuf I.Q. Barzinji Cultural Data Science Date: January 04 5:00 PM
Image Color Analysis Topic - The Greyscale Epidemic

Discussion

Yearly Distribution

This skewed year distribution highlighted a serious issue with the dataset. The dominance of data from 2011 to 2013 meant that the dataset would produce poor results regardless of the analysis. Filtering years with fewer than 50 data points or removing a large number of data points from the years with the highest representation were considered as potential solutions. However, these methods would not substantially improve the dataset due to the lack of data in other years, revealing a fundamental error in data collection.
Furthermore, exploring other subcategories, such as "bottomwear," showed a similar problem, as most accessories also came from the same concentrated years. This issue was consistent across subcategories, emphasizing the challenges posed by the dataset's limitations.

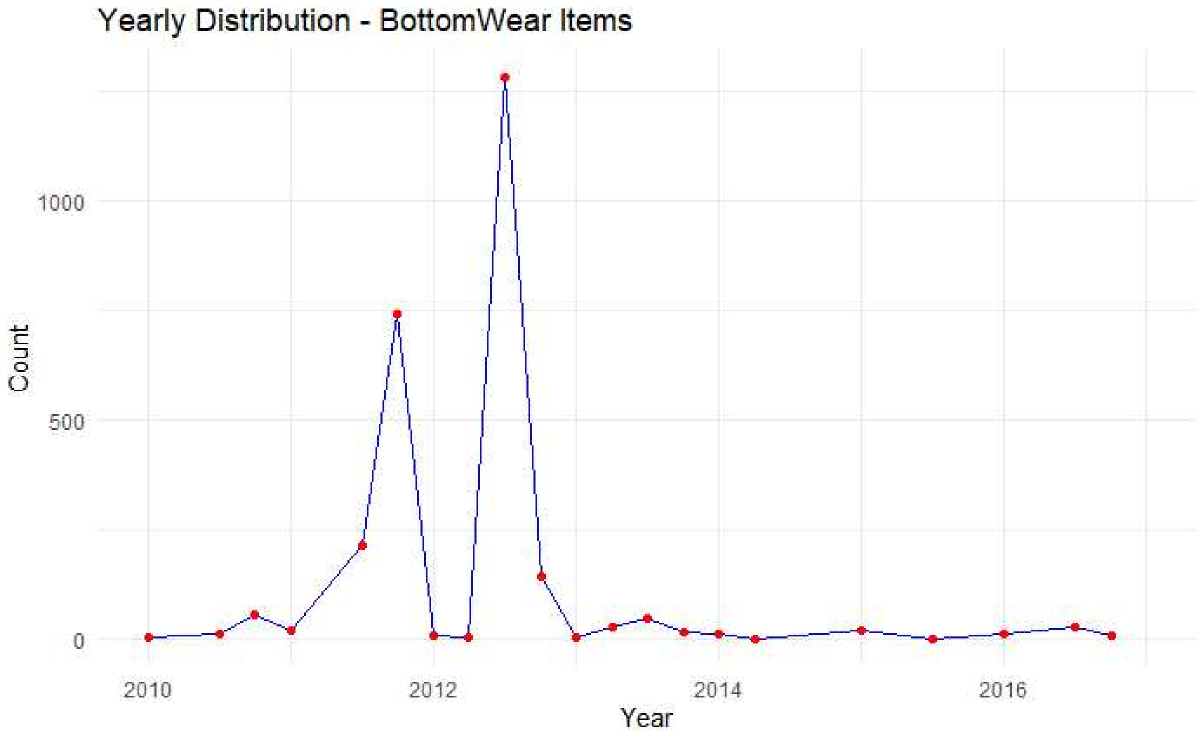

Figure 6 - Yearly Distribution of BottomWear Items

Significance Evaluation

None of the four models showed any significance at any level, likely due to the uneven yearly distribution of the data. Considering this, it is difficult to determine whether there was a clear error in the methodology, correlation tests and variations of t-testing will never accurately represent the data as it is fundamentally flawed. While all models showed a slight negative correlation, which aligns with the proposed hypothesis, this is likely a result of decentralized averaging, suggesting that outliers in highly dense regions are the primary cause.

For future projects, a full and detailed analysis of each categorical variable in the dataset will be necessary to produce results that can better highlight cultural shifts over time.

Appendix

Appendix A - Color Image Extraction

Given that each image contains considerable noise, such as models, backgrounds, or additional clothing items, a straightforward approach of zooming into the center of the image proved highly effective. This technique consistently delivered optimal results and was less resource intensive compared to alternative methods. Once a clip of the center of the image was taken, each pixel was evaluated with an RGB value and averaged to represent the image's overall color.

Appendix B - Script Descriptions

1. **Merging.py**
   This script merged two datasets, images.csv and styles.csv. The images.csv file included image ID and their respective image links, while styles.csv contained all relevant metadata, including year, season, subcategory, and gender.
2. **Filtered.py**
   To ensure consistency in the analysis, a single clothing type was selected from the database. Given the dataset's diverse entries, "TopWear" was chosen as it had the highest quantity of 15,402 data points.
3. **Color_Extraction.py**
   This script cropped each image to its center to minimize noise, such as models or extraneous elements, and extracted the RGB color code for the cropped region.
4. **Colorfulness.py**
   Using the extracted RGB values, this script calculated two new columns in the dataset, one for each method of determining colorfulness (Hasler's equation and SB calculation).
5. **Color_Analysis.rmd**
   Finally, an R script was used to perform statistical analysis on the processed data, enabling a comprehensive evaluation of colorfulness over time.

Yosuf I.Q. Barzinji Cultural Data Science Date: January 04 5:00 PM
Image Color Analysis Topic - The Greyscale Epidemic

References

Gonzalez, A. (2023, December 24). *Blaming minimalism: What happens when color disappears?* TheCollector. https://www.thecollector.com/minimalism-color-grey/

Sleeman, C. (2021, December 16). Colour & Shape: Using Computer Vision to Explore the Science Museum Group Collection. *Medium*. https://lab.sciencemuseum.org.uk/colour-shape-using-computer-vision-to-explore-the-science-museum-c4b4f1cbd72c

Bang, J. (2023, January 11). *Is color disappearing from the world?* The Trailblazer. https://phhstrailblazer.org/featured/in-depth/2023/01/11/is-color-disappearing-from-the-world/

Hasler, D., & Suesstrunk, S. E. (2003). Measuring colorfulness in natural images. *Proceedings of SPIE, the International Society for Optical Engineering/Proceedings of SPIE*. https://doi.org/10.1117/12.477378

Fashion Product Images Dataset. (2019, March 14). Kaggle. https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-dataset

Artizon Museum. (1908). *Collection Highlights: | Artizon Museum*. https://www.artizon.museum/en/collection/art/22492