# CAI
# PROJECT
# – Analysis–

Project Duration: 2020.02.01 ~ 2020.04.27

Project Team Members: Team Leader Yang Hee-seung, Team Members Park Jae-jung, Kim Eun-ju, Jo Ye-seul

Analysis Responsibility: Jo Ye-seul, Park Jae-jung

# Contents

# Report Objective

### Society Emphasizing Individuality

Modern society values individuality, striving to enhance strengths, address weaknesses, and discover the most suitable beauty standards. In line with this societal trend, the concept of personal color has garnered public attention. Personal color refers to colors that harmonize best with one's hair, eye, and skin tones. Utilizing their personal color, individuals aim to present their best appearance and create their desired image.

### Limitations of Personal Color Diagnosis

However, current methods for evaluating personal color lack clear criteria for selecting and diagnosing individual characteristics, relying primarily on qualitative sensory evaluations.

### Overcoming these Limitations

To overcome these limitations, research is conducted on "the construction of a quantitative evaluation model for the classification of personal color skin tone types." This aims to identify quantitative criteria for classifying personal color types and develop suitable algorithms accordingly. Subsequently, through web implementation, the goal is to popularize personal color diagnosis among the public.

# Data Approach

| Analysis Topic and Hypothesis Setting | Data Collection | Data Cleaning and Processing | Data Analysis | Report Writing |
|---|---|---|---|---|
| Exploration of Papers on Color Theory and Personal Color<br><br>After reviewing the papers, Selecting the primary area for color assessment and Choosing the diagnostic colors. | Execution of a Survey on the Rationality of Personal Color Theory and the Selection of Assessment Areas and Diagnostic Colors<br><br>Collection of External Data | Data Cleaning and Processing for Validation of Collected Data | Analysis of Collected Data<br><br>Application of Multifaceted Analysis Techniques | Revision of Personal Color Classification Criteria Based on Analysis Results<br><br>Compilation of Data Findings |
| Desk Research | Survey Data | Data Preprocessing with Python | Statistics Result | Report Design |
| Set Personal Color Standards | Naver Searching Data | | Data Analysis & Visualization with R | Output Image Design |
| Make Survey | | | | Data Examination |

# Make Survey

**Paper Findings:**

Luminance: Differences in luminance exist pre and post makeup application across all areas.
Redness: No significant difference in redness across the entire face.
Yellowness: Differences in color values exist between the forehead and chin but not on the cheeks.

Conclusion drawn from the analysis suggests that women do not correct the yellowness of their cheeks compared to their foreheads and chins. This leads to the conclusion that the skin tone of the cheeks represents their overall skin tone, and preserving the cheek's skin tone is the most natural way to express individuality.

**Hypothesis:**

Is the cheek the representative skin tone of the face?

**Verification Method:**

Adjust the white balance of sample photos.
Extract the average color values for the forehead, cheeks, eyes, chin, and the entire face.
Conduct a survey to determine the facial area perceived as the representative skin tone when people recognize others' faces.
Present the average colors of the cheeks, chin, eyes, forehead, and the entire face for each facial sample.
Validate the results based on survey responses.

**Example:**

"Compared to the following picture, which color do you feel is most similar to the facial color?"

① ② ③ ④ ⑤
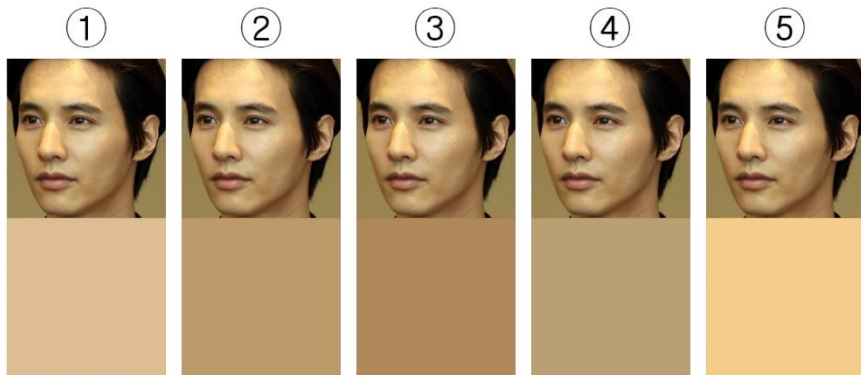
(From left to right: Cheek, Chin, Face, Eyes, Forehead)

# Make Survey

**Paper Findings:**

Personal color aims to make individuals appear brighter based on color theory.
Brightness enhancement in color theory is achieved through color contrast, luminance contrast, and saturation contrast.

**Hypothesis:**

Is the personal color perceived by individuals intended to make the skin appear brighter, or is it intended to complement the skin tone?

**Verification Method:**

For the base color (assumed to be skin tone), adjust the complementary colors according to color theory, selecting a total of 6 colors.
Adjust the hue, luminance, and saturation values to present a total of 5 colors.
Conduct a survey to determine which color makes the base color appear the brightest or most harmonious, ensuring that the populations for the brightest and most harmonious colors do not overlap.

| Survey A | Survey B |
|---|---|
| (Color selection that makes the skin appear the brightest) | (Color selection that complements the skin tone the most) |
| Q: When considering the color in the center of each palette as the skin tone, **Which color appears the brightest?** | Q: When considering the color in the center of each palette as the skin tone, **Which color appears the most harmonious?** |

# Make Survey

Rationality Verification of Personal Color Diagnostic Colors

**Paper Findings:**

Personal color aims to make individuals appear brighter based on color theory. Enhancing brightness in color theory involves color contrast, luminance contrast, and saturation contrast.

**Hypothesis:**

According to contrast theory, colors that contrast with the base color (skin tone) are perceived as most suitable for personal color.

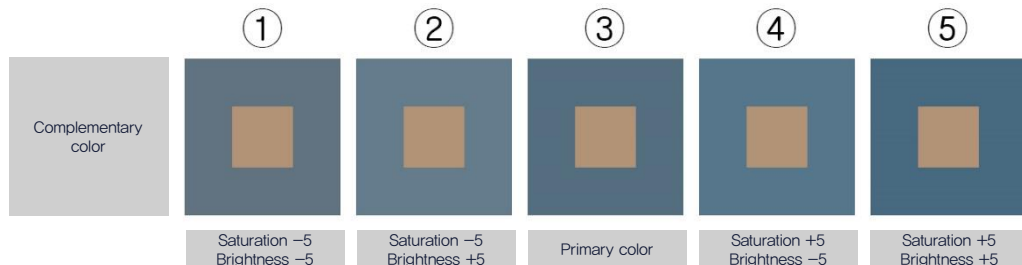**Verification Method:**

Using the base color as a sample, extract complementary colors, triadic colors, and split-complementary colors (total of 6). Fix the color values of each contrast and adjust the luminance/saturation to compare the colors.
Survey the most suitable color for personal color among each contrast and adjusted color.

**Example:**

When considering the color in the center of each palette as the skin tone, which color appears the brightest (most harmonious)?

| | ① | ② | ③ | ④ | ⑤ |
|---|---|---|---|---|---|
| Complementary color | | | | | |
| | Saturation −5 Brightness −5 | Saturation −5 Brightness +5 | Primary color | Saturation +5 Brightness −5 | Saturation +5 Brightness +5 |

# Make Survey

Rationality Verification of Personal Color Diagnostic Colors

**Paper Findings:**

Personal color aims to make individuals appear brighter based on color theory. Enhancing brightness in color theory involves color contrast, luminance contrast, and saturation contrast.

**Hypothesis:**

According to contrast theory, colors that contrast with the base color (skin tone) are perceived as most suitable for personal color.

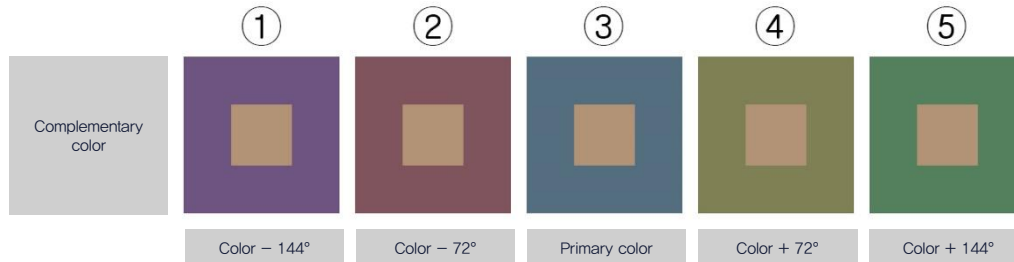**Verification Method:**

Using the base color as a sample, extract complementary colors, triadic colors, and split-complementary colors (total of 6). Fix the color values of each contrast and adjust the luminance/saturation to compare the colors.
Survey the most suitable color for personal color among each contrast and adjusted color.

**Example:**

When considering the color in the center of each palette as the skin tone, which color appears the brightest (most harmonious)?



| | ① | ② | ③ | ④ | ⑤ |
|---|---|---|---|---|---|
| Complementary color | Color − 144° | Color − 72° | Primary color | Color + 72° | Color + 144° |

# Make Survey

Survey for Comparing Personal Color Perception

## Goals

Investigate whether there are differences in color perception based on age and gender.

Determine if there are differences in perception of personal color based on level of interest in fashion.

Explore whether differences in perception of personal color exist based on individuals' understanding and preference of the concept.

Determine the purpose of using personal color through surveys to assess future utilization.

## Verification Methods

- Add age and gender items to the survey questions.
- Investigate whether there are differences in color perception for each survey item.

- Add a check item for level of interest in fashion.
- Assume that those with a high interest in fashion will have a higher interest in color and better color distinction abilities.

- Survey whether the perceived personal color concept aligns with making the skin appear brighter or harmonizing with the skin tone.
- Assess whether differences in perception of the concept lead to differences in color perception.

- Add an item for the purpose of using personal color (styling, makeup, etc.).
- Decide whether to associate the survey results with the data from the "Musinsa" project conducted during the mid-term project phase and determine how to utilize personal color measurements in the future.
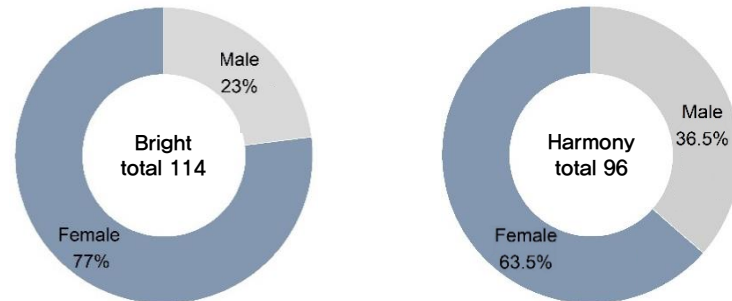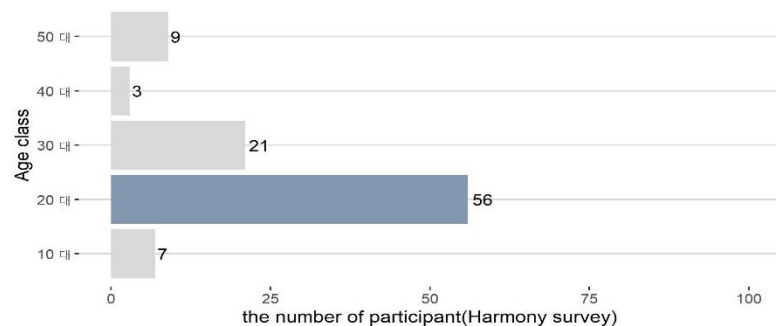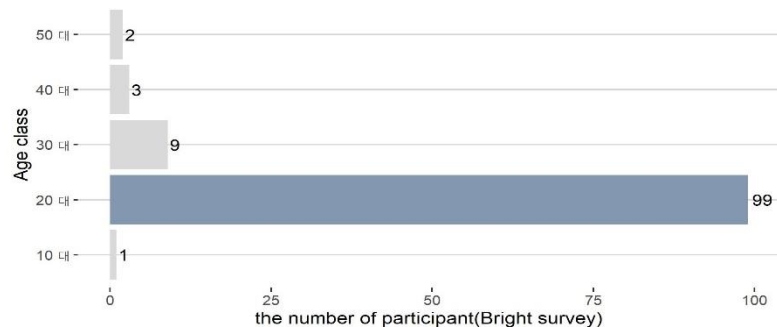
# Survey Result

Survey participation results

## Number of survey participants
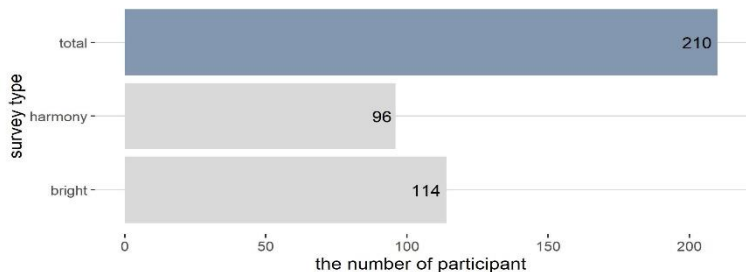


## Ratio of men and women in survey



Bright
total 114

Male
23%

Female
77%

Harmony
total 96

Male
36.5%

Female
63.5%

## Number of people in the survey by age

# Data Analysis

Logistic regression analysis is a method used to understand the impact of independent variables on a dependent variable and to produce a model that predicts the value of the dependent variable corresponding to certain values of the independent variables.

## Condition 1
## Logistic regression analysis is suitable when the dependent variable is measured as a categorical qualitative variable on a nominal scale.
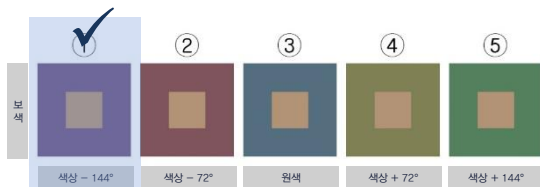


In this case, the dependent variable is the selection of correct or incorrect answers regarding personal color. Since the selection of correct and incorrect answers constitutes a categorical variable, logistic regression analysis is appropriate.

## Condition 2
## There should be no multicollinearity among the independent variables.

| | |
|---|---|
| sex | 1.50449421996316 |
| definition | 1.00000003184469 |
| need | 1.06355126809827 |
| purpose_clothes | 1.36940374676163 |
| purpose_cosmetic | 1.70145776269596 |
| interest | 1.06502533039289 |
| age_class | 1.09500636334636 |

Using the square root of the variance inflation factor (VIF), no value exceeds 2. Therefore, it is determined that there are no issues with multicollinearity among the variables.

## Condition 3
## The values of the dependent variable should be statistically independent of each other.

```
                OLS Regression Results
==========================================================
Omnibus:              20.947   Durbin-Watson:          1.943
Prob(Omnibus):         0.000   Jarque-Bera (JB):      27.207
Skew:                  1.291   Prob(JB):            1.24e-06
Kurtosis:              3.368   Cond. No.            2.01e+18
==========================================================
```

Using sm.OLS(), the Durbin–Watson statistic is not significantly different from 2. Therefore, it is concluded that the values of the dependent variable are independent.

## Condition 4
## The values of the dependent variable corresponding to the independent variables should follow a normal distribution, and the variances of all normal distributions should be equal.

```
Shapiro-Wilk normality test          Bartlett test of homogeneity of variances

data: resid(out0)                    data: comp by sex
W = 0.39922, p-value < 2.2e-16       Bartlett's K-squared = 28.49, df = 1, p-value = 9.42e-08
```

The p–value from the Shapiro–Wilk test is less than 0.05, indicating that the data does not follow a normal distribution. Additionally, the p–value from the Bartlett test is less than 0.05, indicating that the variances are not equal.

# Data Analysis

Based on the previous results, since our data does not follow a normal distribution and does not have equal variances, logistic regression analysis cannot be used. Consequently, attempts to identify significant variables for building a logistic regression model were unsuccessful.
The goal was to use logistic regression analysis to identify groups with high survey accuracy and plan algorithms accordingly. However, this is not feasible.

```
Call:
glm(formula = comp ~ sex + definition + need + purpose_clothes +
    purpose_cosmetic + interest + age_class, family = binomial,
    data = total)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.91974  -0.33018  -0.21494  -0.00009  2.65833

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)        -2.277e+01  3.526e+03  -0.006   0.9948
sexMale             3.238e+00  1.804e+00   1.795   0.0726 .
definitionharmony   1.742e+01  3.526e+03   0.005   0.9961
needharmony        -7.306e-01  1.298e+00  -0.563   0.5734
purpose_clothes     3.781e-04  1.802e+00   0.000   0.9998
purpose_cosmetic    2.192e+00  1.835e+00   1.194   0.2325
interest            1.271e-01  4.378e-01   0.290   0.7715
age_class          -1.256e-02  5.705e-02  -0.220   0.8258
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 33.255  on 95  degrees of freedom
Residual deviance: 26.849  on 88  degrees of freedom
AIC: 42.849

Number of Fisher Scoring iterations: 19
```

- glm() : Function used to identify statistically significant variables in logistic regression analysis.

- Call : Presentation of the constructed model.

- Coefficients의 Pr
  : p–value of the constructed model.
  : Variables with p–values less than 0.05 are statistically significant.
  : The variable with the lowest p–value indicates the strongest predictive power.

- Interpretation of Results (Displayed in the left figure)
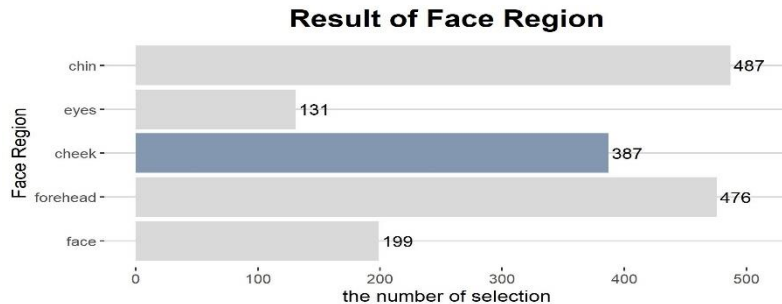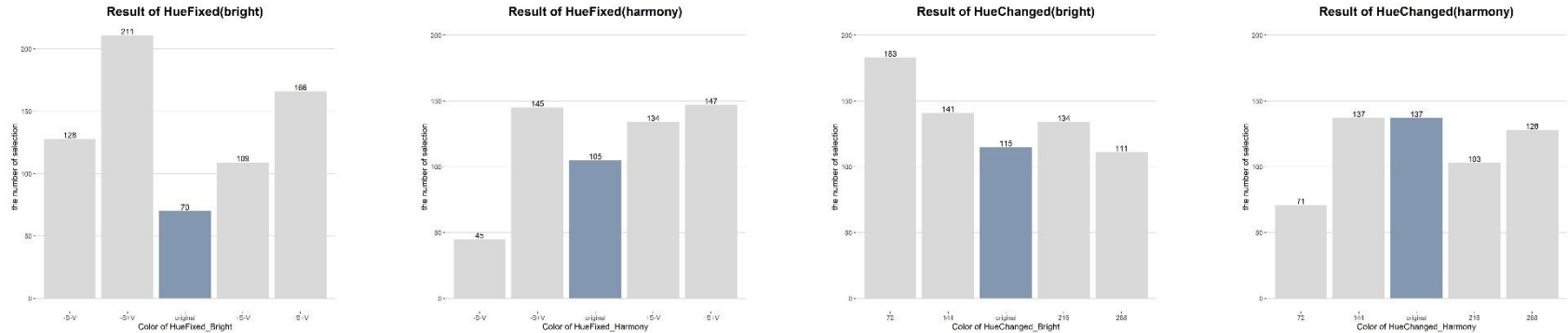
### Survey B (Harmony)

Among the fixed color items, the complementary color sample

All variables have p–values greater than 0.05
–〉None of the variables can be considered statistically significant.

https://github.com/slmteruto/CAI/cys/CAI_Logistic_useR.ipynb

# Data Analysis

— Furthermore, except for the item in the survey questionnaire that asks about harmony with facial color, where the color change item is excluded, the selection rate for correct answers is low in all other items.

— According to color theory, facial color is expected to appear brightest when it is accompanied by complementary colors, triadic colors, and split-complementary colors. However, colors that have been adjusted in terms of hue, saturation, and brightness from the primary colors receive more selections than the primary colors themselves.

**Result of HueFixed(bright)**

the number of selection — Color of HueFixed_Bright

211, 168, 128, 109, 70

**Result of HueFixed(harmony)**

the number of selection — Color of HueFixed_Harmony

145, 147, 134, 105, 45

**Result of HueChanged(bright)**

the number of selection — Color of HueChanged_Bright

183, 141, 134, 115, 111

**Result of HueChanged(harmony)**

the number of selection — Color of HueChanged_Harmony

137, 137, 126, 103, 71

**Result of Face Region**

Face Region — the number of selection

| Face Region | the number of selection |
| --- | --- |
| chin | 487 |
| eyes | 131 |
| cheek | 387 |
| forehead | 476 |
| face | 199 |

— The cheek is considered as the facial area representing the overall facial color as set in the paper. However, the survey results show that the number of selections for the cheek is low.

— The reason for the low selection of items related to the color selection area and diagnostic color is believed to be primarily due to significant differences in people's perceptions.

— As a result of these findings, it is concluded that it is difficult to believe in the accuracy of the paper and theory.

# Data Analysis

Reasons why the data analysis using survey results was not properly conducted

## Internal Factors

**1. Insufficient learning and understanding of the analysis method before proceeding with survey writing**

It is considered that the number of options (5) provided during the survey on diagnostic colors was too high.
If the survey writing had been conducted with a clear direction of 'prediction analysis based on correct and incorrect answer rates', only two options would have been provided.
Then, people could have made easier and faster choices, and the analysis would have proceeded in the desired direction.

**2. Setting more independent variables**

To increase the survey participation rate, only the minimum elements were included in the survey, thinking that the number of questions should be minimal.
However, it is judged that if there had been more diverse independent variables, multidimensional analysis would have been possible based on the results of logistic regression analysis.

## External Factors

**1. Need for more survey participants**
Currently, there are about 200 survey participants.
There were limitations in only asking acquaintances of the team members.

**2. Distribution of survey participants**
It is judged that more objective survey results could have been obtained if the distribution of participants' age and gender had been uniform.
Alternatively, if a specific group had been the focus, more standardized survey results could have been obtained.

# Data Analysis

Therefore, it is decided to select a group deemed to have the most interest in personal color using different data sources. The algorithms will be based on the answers chosen by this group.
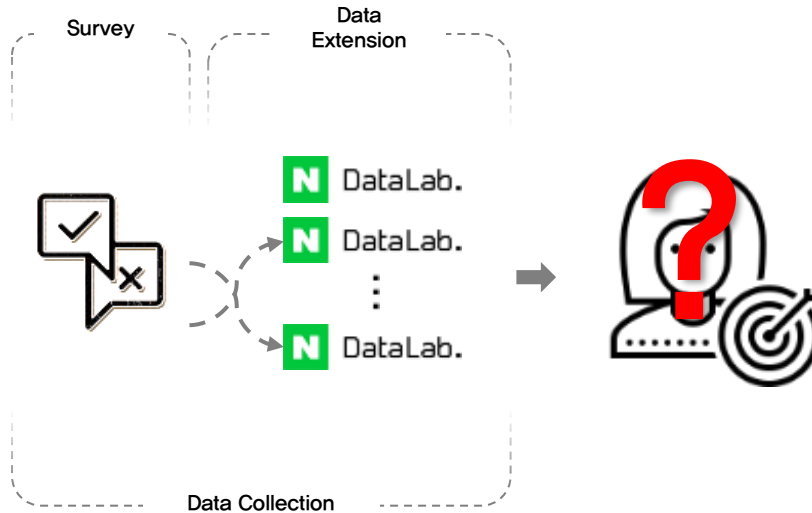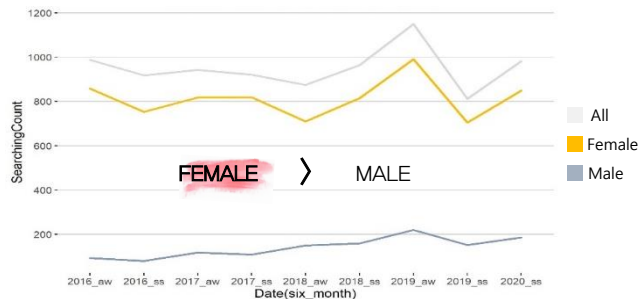


## Using Naver Data Lab's search trends

✓ Data that indexes the search volume trends for specific keywords.
✓ High market share of Naver in South Korea.
✓ Search trends can be segmented by age and gender.
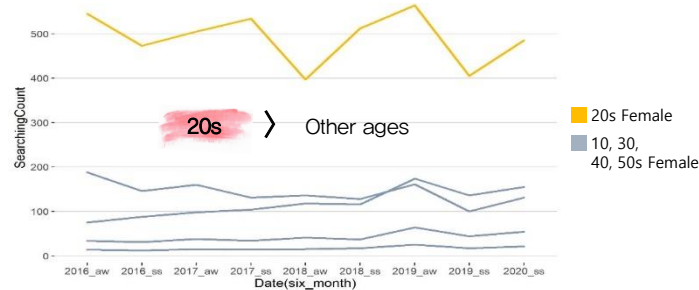✓ Data can be obtained in formats other than graphs.

# Data Analysis

## Setting Algorithm Criteria

### Search frequency for personal color by gender



FEMALE 〉 MALE

The search volume for females is overwhelmingly higher than for males

### Search frequency for personal color by age among females



20s 〉 Other ages

The search volume for women in their 20s is overwhelmingly high, indicating a significant interest in personal color. Therefore, this group is designated as the representative group

### Search frequency for the definition of personal color among women in their 20s



Bright ≒ Harmony
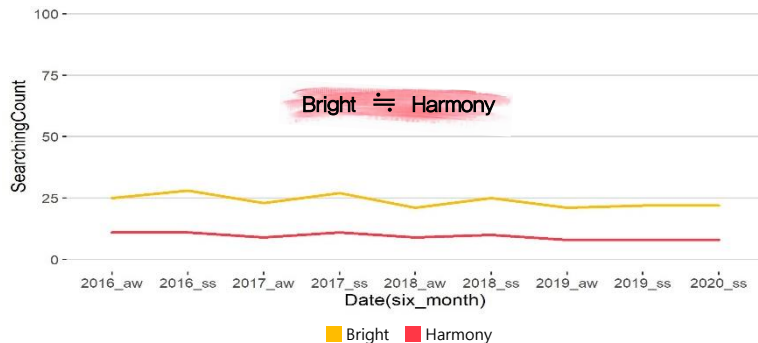
The search frequency for the definition of personal color among the designated representative group of women in their 20s shows little variation.

Thus, on the website, it is decided to present the colors that:
- Both show harmony with facial color and make the facial color appear brighter.
- Make the face appear brighter will reflect adjustments in brightness and saturation.
- Show harmony with facial color will reflect adjustments in color values.

〈 Search Method 〉
Period: January 2016 – February 2020 (averaged every 6 months)

Keywords:
Personal Color: personal color
Bright: bright, light, brightly, lightly
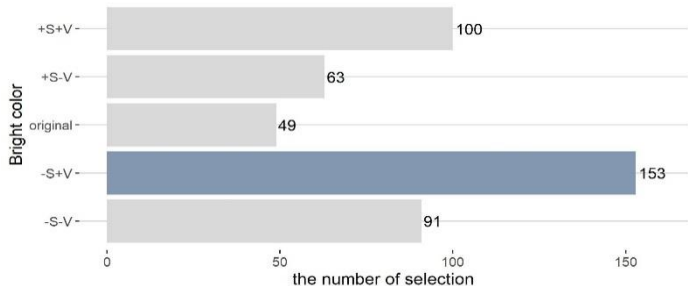Harmony: harmonious, balanced, well-matched, in harmony
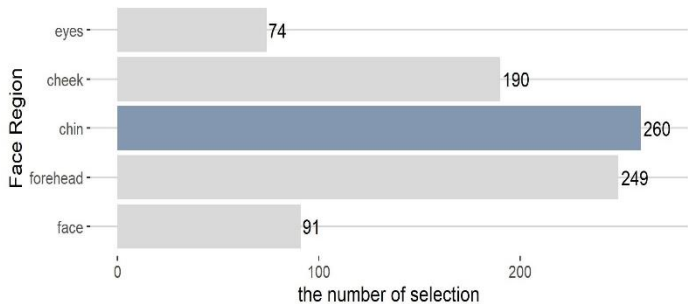
# Data Analysis

## Algorithm Implementation Criteria

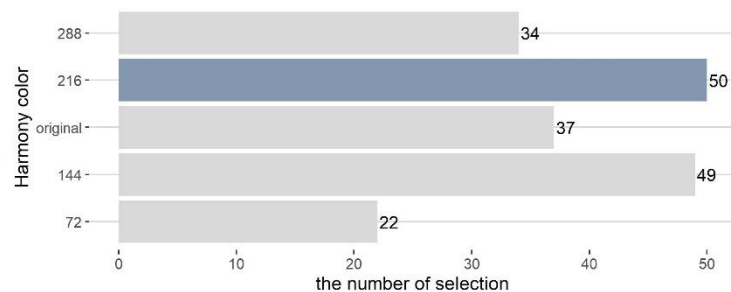### Representative color values that make the face appear brighter



According to women in their 20s, the color values that make the face appear the brightest are achieved by decreasing the saturation by 5 and increasing the brightness by 5 for each complementary color's primary color.

### Representative color values that show harmony with the facial color



According to women in their 20s, the color values that show the most harmony with the facial color are achieved by increasing the hue by 72 degrees for each complementary color's primary color.

### Representative facial area



According to women in their 20s, the facial area that has a color most similar to the overall facial color is the chin

〈 Algorithm Criteria Setting 〉
Based on the choices of women in their 20s from the survey

- ✓ Representative facial area: Chin

- ✓ Color values that make the face appear brighter: Decrease the saturation by 5 and increase the brightness by 5 for each complementary color's primary color

- ✓ Color values that show harmony with the facial color: Increase the hue by 72 degrees for each complementary color's primary color.
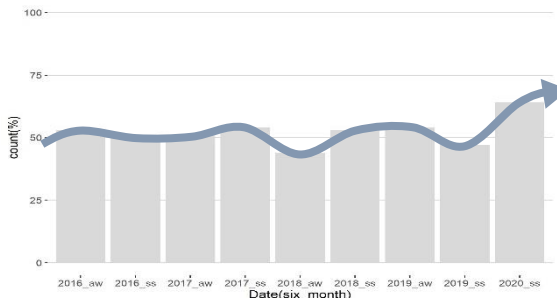
# Data Analysis

## Personal Color Purpose

### Personal Color Purpose for Clothing
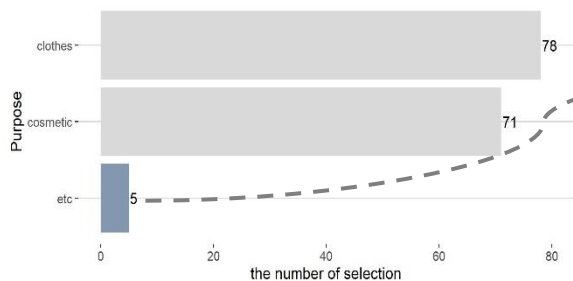
**Results of Personal Color Purpose Survey**
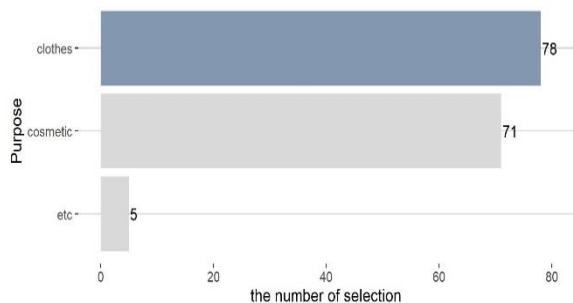


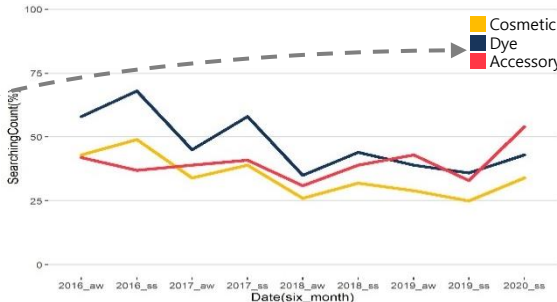**Search Results for Personal Color Purpose for Clothing**



- ✓ According to the survey, the most common purpose for using personal color is matching clothing.
- ✓ Using Naver Data Lab to analyze search results for the personal color purpose for clothing shows an upward trend.
- ✓ It was concluded that matching with "Musinsa clothing," used in the midterm project, would be beneficial.

### Results of Personal Color Purpose Survey

**Results of Personal Color Purpose Survey**



**Search Results for Personal Color Purpose**



- ✓ When surveying the purpose of personal color, many respondents indicated they wanted to use it for accessories and hair dyeing.
- ✓ Using Naver Data Lab to analyze search results for each purpose, accessories show an upward trend, while makeup and hair dyeing show a downward trend.
- ✓ Therefore, it is anticipated that it would be beneficial to introduce a recommendation system on the website in the order of accessories, makeup, and hair dyeing.

# " Personal Color Data Analysis Results "

✓ Algorithm Criteria Set Using Surveys and Naver Data Lab:

Representative Group: Women in their 20s

Representative Facial Area: Jaw

Color Value to Make the Face Look Brighter: Decrease saturation by 5 and increase brightness by 5 for each complementary color's original hue.

Color Value to Show Harmony with the Face: Increase hue by 72 degrees for each complementary color's original hue.

✓ Personal Color Usage Purposes

Survey results show high responses for using personal color in clothing matching.

Naver Data Lab also shows a consistent increase in searches linking personal color with clothing.

It is deemed beneficial to include a fashion item recommendation system on the website.

Cosmetics, hair dyeing, and accessories are selected as additional recommendation systems for future implementation.

Thank you