**Data Appendix**
**Dataset: celebrity_image_data.csv**

The data file contains eight variables, with each row containing information about a different celebrity. The unit of observation is the celebrity.

**Variables:**

### *celebrity_name*

- Definition: Name of celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: celebrity_name
- 2500(0)
- Steps for processing the variable for cleaned dataset: No changes were made

### *picture_number*

- Definition: Image number of a specific celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: picture_number
- 2500(0)
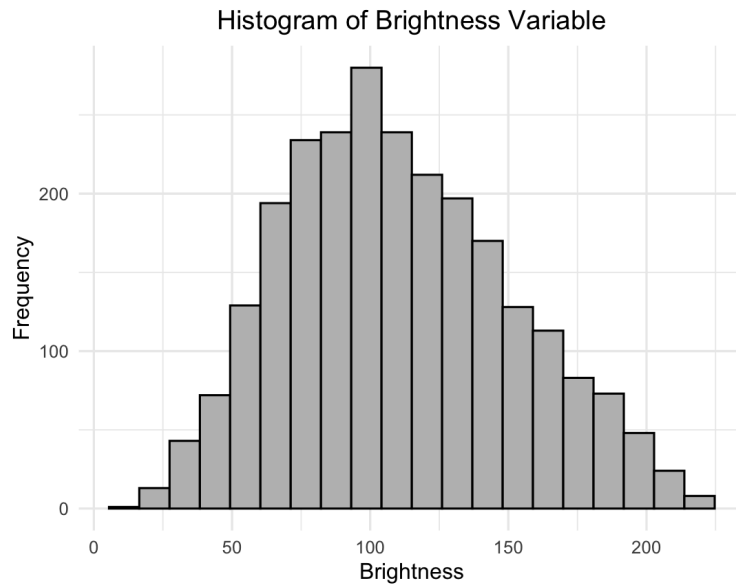- Steps for processing the variable for cleaned dataset: No changes were made

### *file_path*

- Definition: String of characters that indicates the location of the image in the GitHub repository
- Units of Measurement: N/A
- Exact Name of Variable in original data file: file_path
- 2500(0)
- Steps for processing the variable for cleaned dataset: No changes were made

### *brightness*

- Definition: Brightness of an image, measured as a number above 0
- Units of Measurement: N/A
- Exact Name of Variable in original data file: brightness
- 2500(0)
- Steps for processing the variable for cleaned dataset: No changes were made

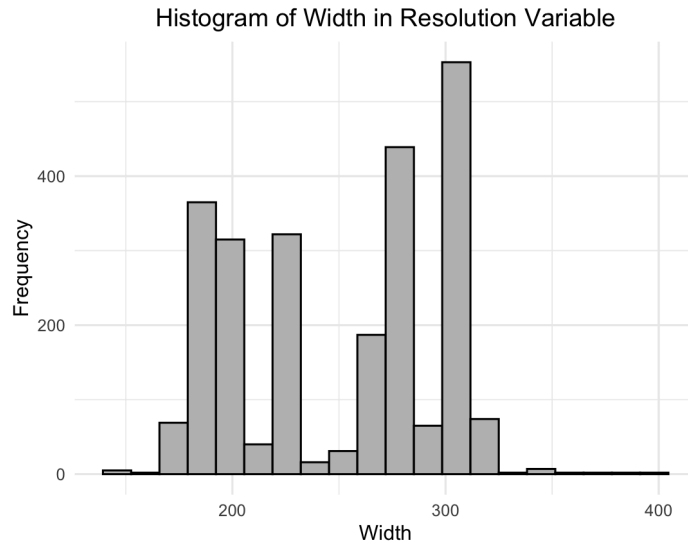| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
| --- | --- | --- | --- | --- | --- | --- |
| 15.85 | 79.54 | 106.53 | 110.30 | 138.40 | 224.07 | 40.92457 |

### Histogram of Brightness Variable



### *resolution*

- Definition: The resolution, or quality, of an image listed in the format of "width x height"
- Units of Measurement: Pixels
- Exact Name of Variable in original data file: resolution
- 2500(0)
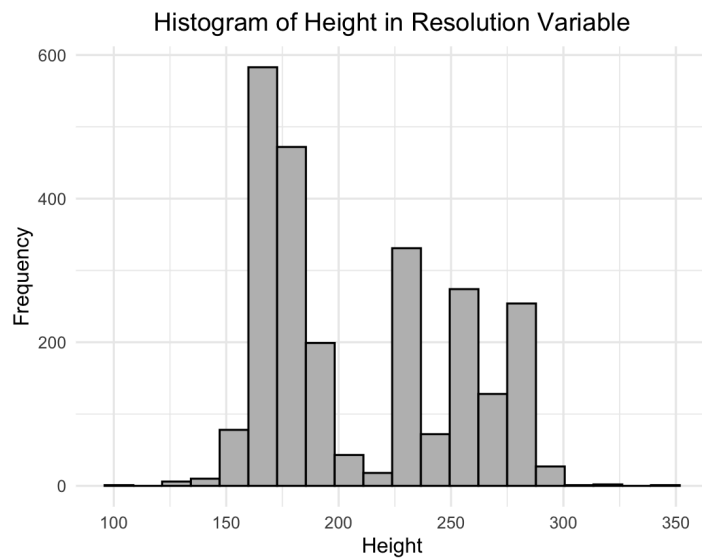- Steps for processing the variable for cleaned dataset: No changes were made

WIDTH

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|
| 140 | 201 | 259 | 248 | 299 | 392 | 47.15581 |

Histogram of Width in Resolution Variable

HEIGHT

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|
| 100.0 | 168.0 | 194.0 | 210.2 | 251.0 | 343.0 | 41.63988 |


Histogram of Height in Resolution Variable

*colorfulness*

- Definition: The level of color contained within an image, quantified as a number above 0
- Units of Measurement: N/A
- Exact Name of Variable in original data file: colorfulness
- 2500(0)

- Steps for processing the variable for cleaned dataset: No changes were made

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---------|-----------------|--------|------|-----------------|---------|--------------------|
| 0.00 | 29.11 | 39.11 | 42.50 | 51.77 | 155.30 | 22.52963 |
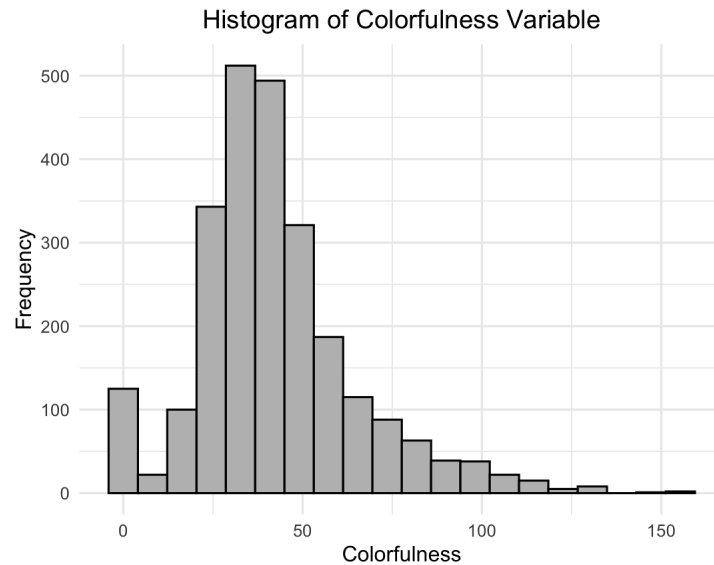
Histogram of Colorfulness Variable



### *gender*

- Definition: Gender of the celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: gender
- 2500(0)
- Steps for processing the variable for cleaned dataset: No changes were made

### *race*

- Definition: Race of the celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: race
- 2500(0)
- Steps for processing the variable for cleaned dataset: No changes were made

**Dataset: all_celebs_image_data.csv**

The data file contains eight variables, with each row containing information about celebrity images (including augmented images). The unit of observation is the celebrity image.

**Variables:**

### celebrity_name

- Definition: Name of celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: celebrity_name
- 10000(0)
- Steps for processing the variable for cleaned dataset: No changes were made

### picture_number

- Definition: Image number of a specific celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: picture_number
- 10000(0)
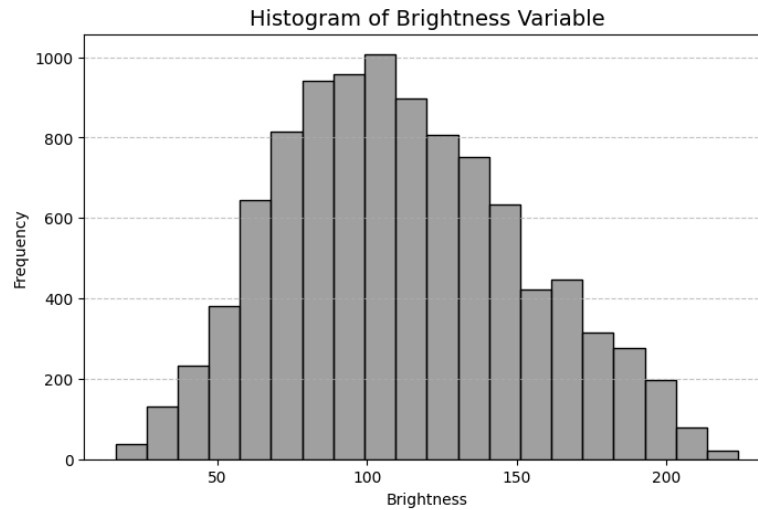- Steps for processing the variable for cleaned dataset: No changes were made

### file_path

- Definition: String of characters that indicates the location of the image in the GitHub repository
- Units of Measurement: N/A
- Exact Name of Variable in original data file: file_path
- 10000(0)
- Steps for processing the variable for cleaned dataset: No changes were made

### brightness

- Definition: Brightness of an image, measured as a number above 0
- Units of Measurement: N/A
- Exact Name of Variable in original data file: brightness
- 10000(0)
- Steps for processing the variable for cleaned dataset: No changes were made

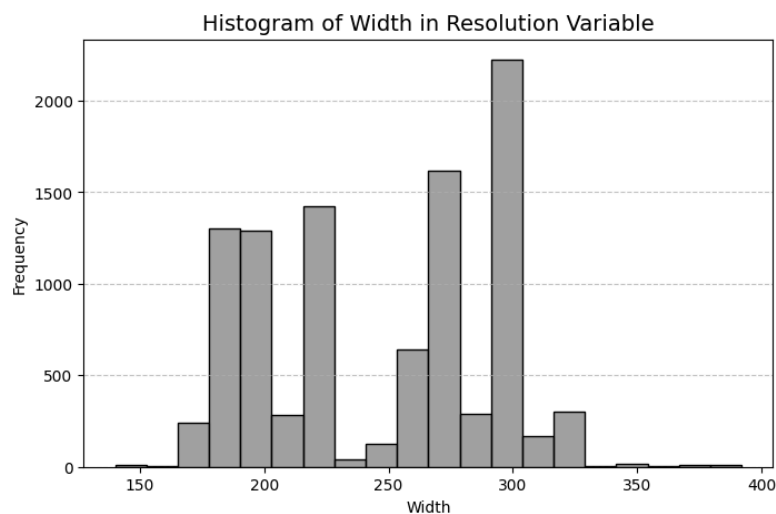| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---------|-----------------|--------|------|-----------------|---------|--------------------|
| 15.85 | 81.21 | 108.03 | 111.50 | 139.13 | 224.07 | 40.40 |

Histogram of Brightness Variable

## *resolution*

- Definition: The resolution, or quality, of an image listed in the format of "width x height"
- Units of Measurement: Pixels
- Exact Name of Variable in original data file: resolution
- 10000(0)
- Steps for processing the variable for cleaned dataset: No changes were made
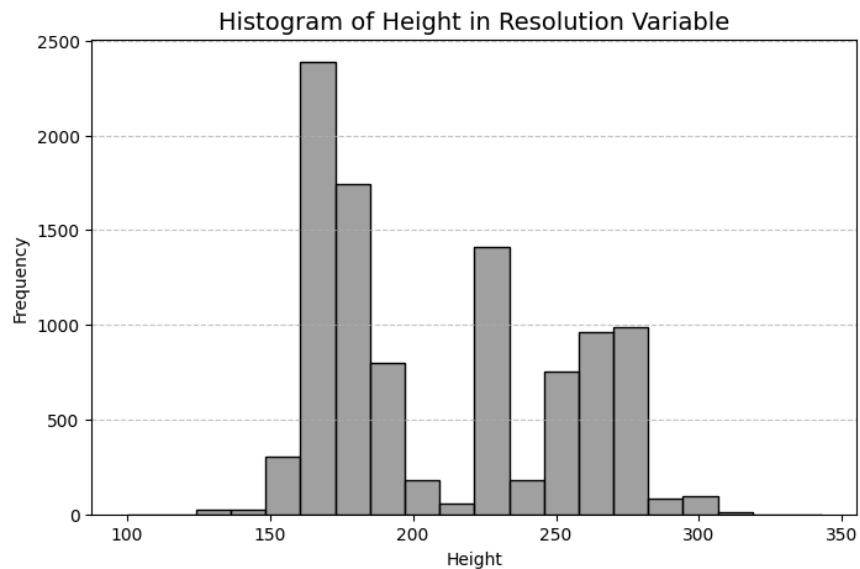
WIDTH

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---------|-----------------|--------|------|-----------------|---------|--------------------|
| 140 | 201 | 259 | 248 | 299 | 392 | 47.15581 |



Histogram of Width in Resolution Variable

HEIGHT

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|
| 100.0 | 168.0 | 194.0 | 210.2 | 251.0 | 343.0 | 41.63988 |



Histogram of Height in Resolution Variable

*colorfulness*

- Definition: The level of color contained within an image, quantified as a number above 0
- Units of Measurement: N/A
- Exact Name of Variable in original data file: colorfulness
- 10000(0)
- Steps for processing the variable for cleaned dataset: No changes were made

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|
| 0.00 | 29.60 | 39.01 | 42.40 | 51.04 | 155.57 | 22.04 |
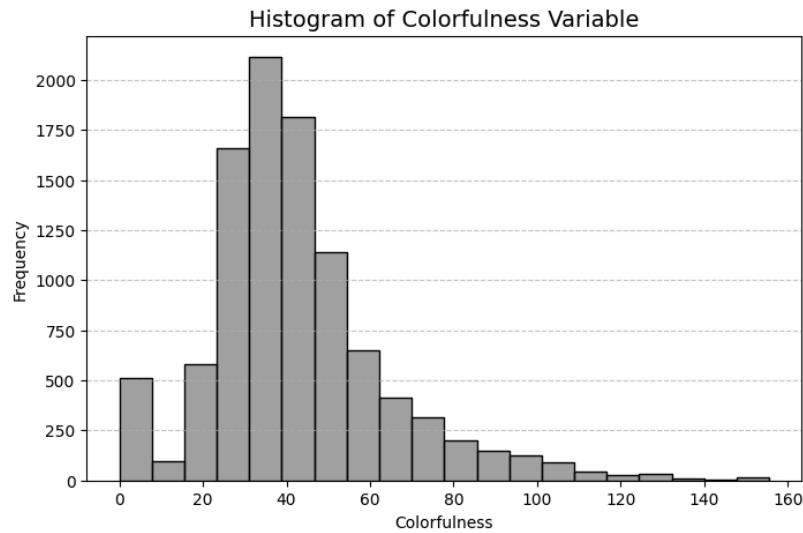
Histogram of Colorfulness Variable

## gender

- Definition: Gender of the celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: gender
- 10000(0)
- Steps for processing the variable for cleaned dataset: No changes were made

## race

- Definition: Race of the celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: race
- 10000(0)
- Steps for processing the variable for cleaned dataset: No changes were made

**Dataset: incorrect_data_guesses.csv**

The data file contains sixteen variables, with each row containing information about a different celebrity image which was incorrectly identified in epoch 100. The unit of observation is the celebrity image.

**Variables:**

### celebrity_name

- Definition: Name of celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: celebrity_name
- 287(0)
- Steps for processing the variable for cleaned dataset: No changes were made

### picture_number

- Definition: Image number of a specific celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: picture_number
- 287(0)
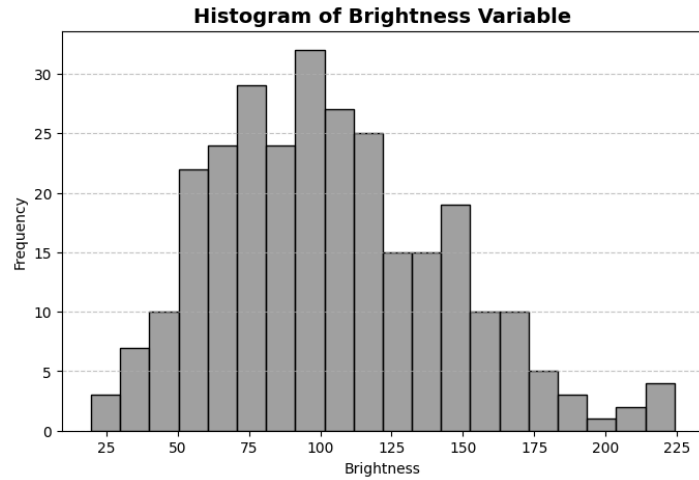- Steps for processing the variable for cleaned dataset: No changes were made

### file_path

- Definition: String of characters that indicates the location of the image in the GitHub repository
- Units of Measurement: N/A
- Exact Name of Variable in original data file: file_path
- 287(0)
- Steps for processing the variable for cleaned dataset: No changes were made

### brightness

- Definition: Brightness of an image, measured as a number above 0
- Units of Measurement: N/A
- Exact Name of Variable in original data file: brightness
- 287(0)
- Steps for processing the variable for cleaned dataset: No changes were made

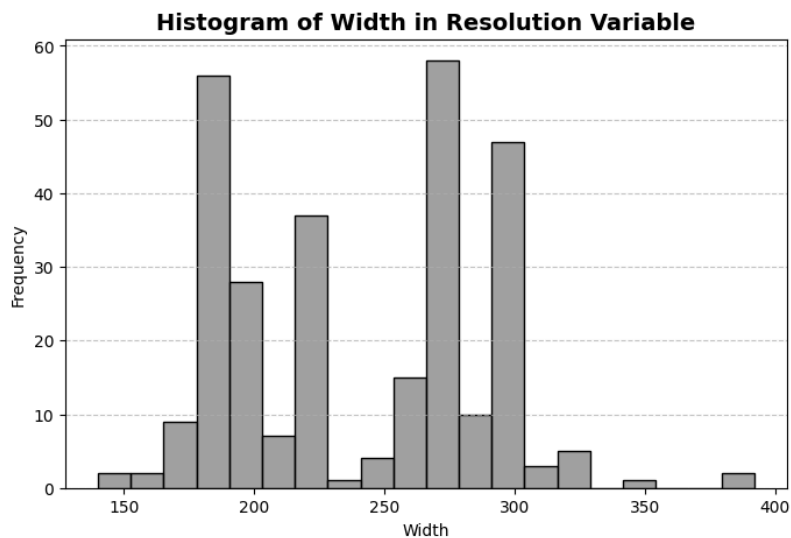| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|
| 19.71 | 73.28 | 99.11 | 103.45 | 129.78 | 224.07 | 40.88 |

Histogram of Brightness Variable

## *resolution*

- Definition: The resolution, or quality, of an image listed in the format of "width x height"
- Units of Measurement: Pixels
- Exact Name of Variable in original data file: resolution
- 287(0)
- Steps for processing the variable for cleaned dataset: No changes were made
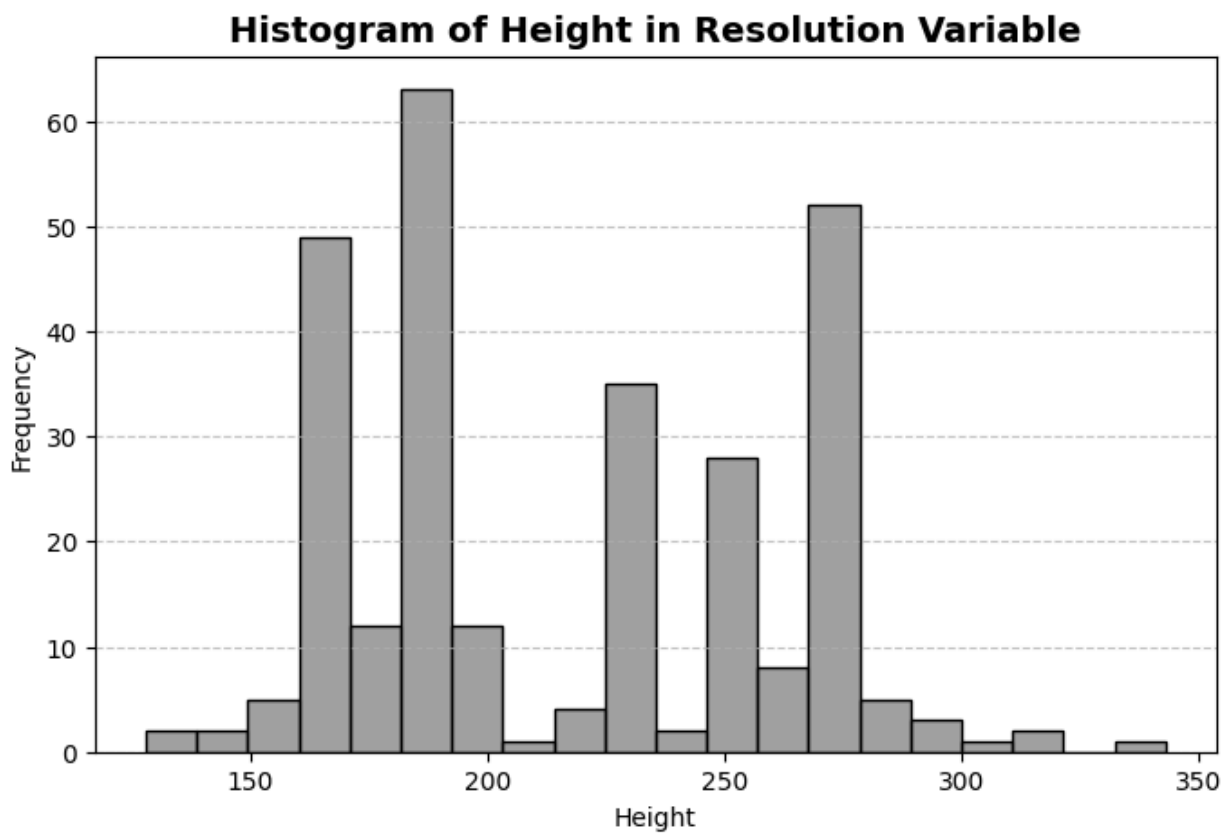
WIDTH

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---------|-----------------|--------|------|-----------------|---------|--------------------|
| 140 | 194 | 249 | 241 | 276 | 392 | 48.91 |



Histogram of Width in Resolution Variable

HEIGHT

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|
| 128 | 182 | 202 | 216.16 | 256.50 | 343.0 | 44.31 |

## Histogram of Height in Resolution Variable



*colorfulness*

- Definition: The level of color contained within an image, quantified as a number above 0
- Units of Measurement: N/A
- Exact Name of Variable in original data file: colorfulness
- 287(0)
- Steps for processing the variable for cleaned dataset: No changes were made

| Minimum | 25th Percentile | Median | Mean | 75th Percentile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|
| 0.00 | 26.71 | 36.87 | 42.67 | 56.29 | 117.07 | 25.54 |

## Histogram of Colorfulness Variable



### *gender*

- Definition: Gender of the celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: gender
- 2500(0)
- Steps for processing the variable for cleaned dataset: No changes were made

### *race*

- Definition: Race of the celebrity
- Units of Measurement: N/A
- Exact Name of Variable in original data file: race
- 2500(0)
- Steps for processing the variable for cleaned dataset: No changes were made