

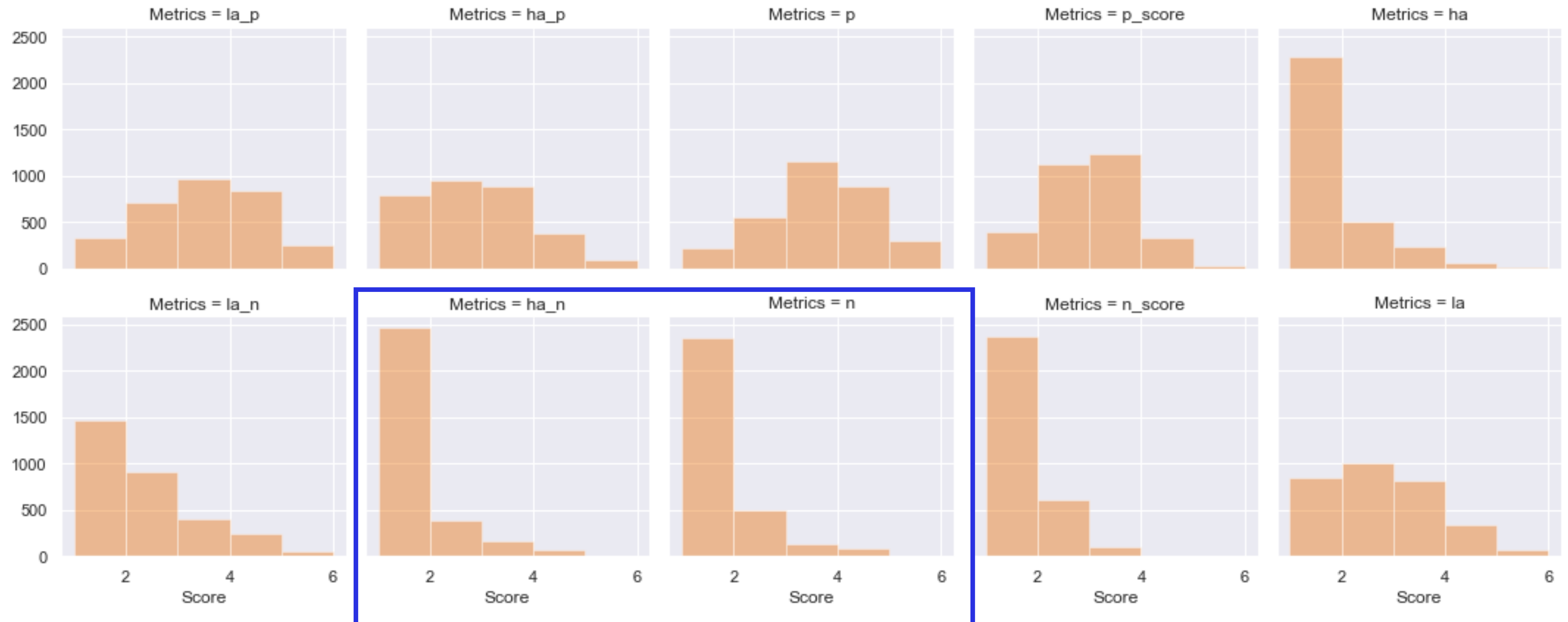
Exploratory Data Analysis

For Emotional States Prediction Projects

Sicong Zhao @ Duke University
sicong.zhao@duke.edu

1. EDA Insights - Distribution of Labels

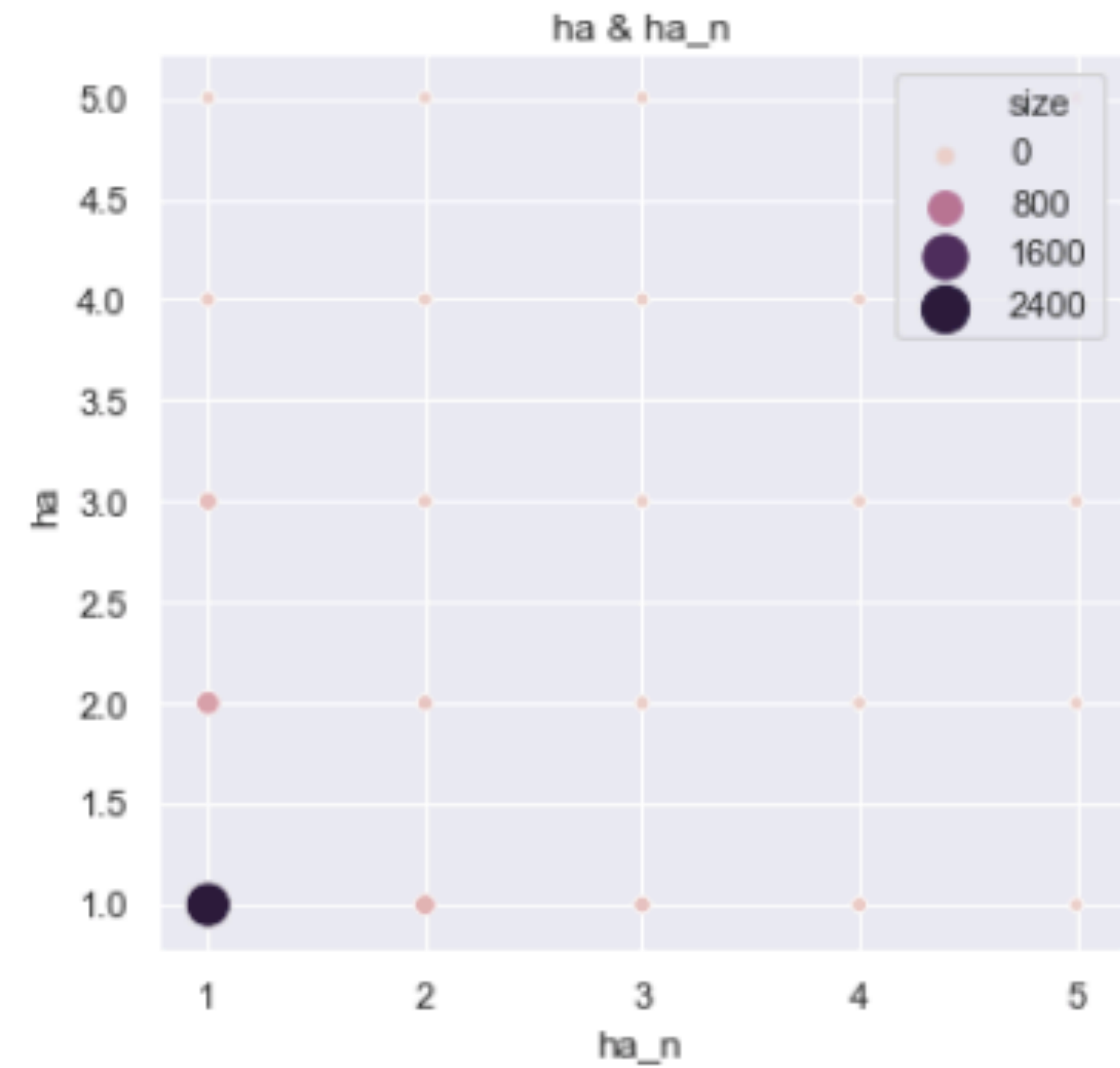
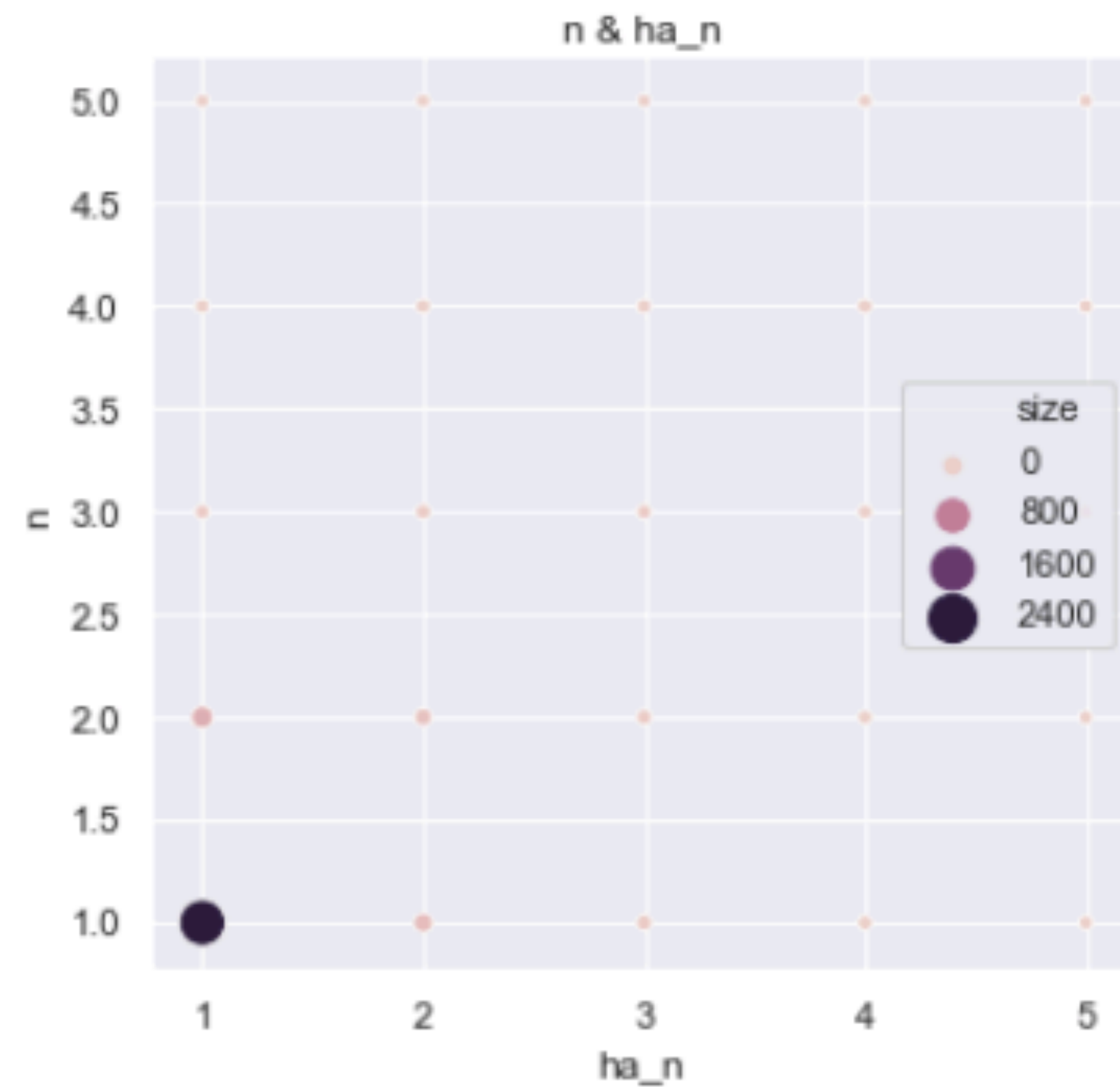
The `ha_n` and `n` metrics are very right skewed. And the scatter plots of the pairs of labels (`n` & `ha_n`, `ha` & `ha_n`) indicate high concentration.



1. EDA Insights - Distribution of Labels

Distribution wise, `ha_n` and `n` metrics are very right skewed.

- About 80% subjects reported `ha_n==1`
- About 76% subjects reported `n==1`
- About 68% subjects reported `ha_n==1 & n==1`



Based on the skewness we have detected, after discussion with our PI, Gregory Samanez-Larkin, we decide to generate labels in following approach:

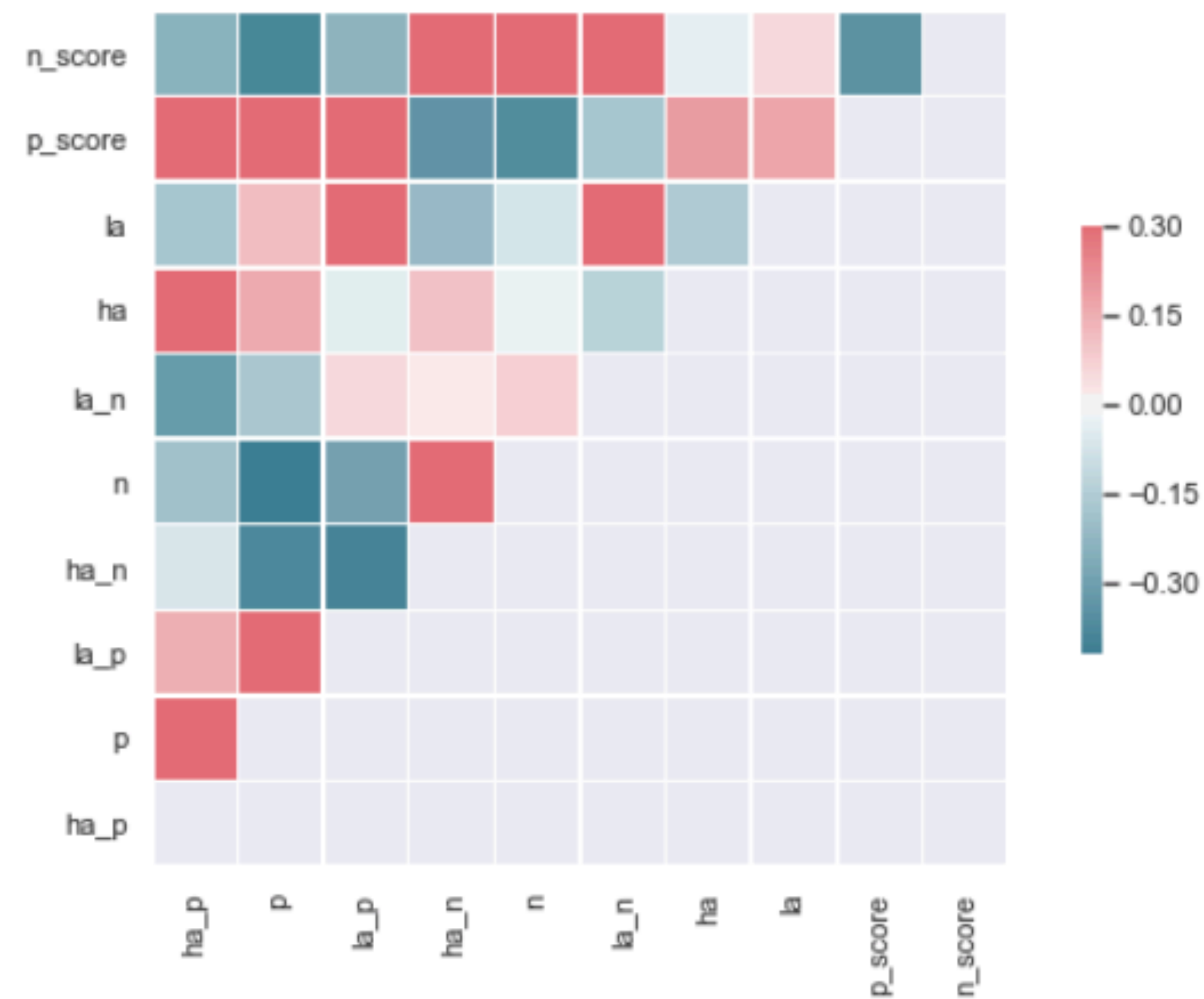
if $\text{max}(\text{la}_n, n, \text{ha}_n) \geq \text{max}(\text{la}_p, p, \text{ha}_p)$, then $\text{label} == 1$, which represents the subject is unhappy.

This approach gives 26.6% unhappy label, which is relatively balanced than our previous approach, which is:

if $\text{mean}(\text{la}_n, n, \text{ha}_n) \geq \text{mean}(\text{la}_p, p, \text{ha}_p)$, then $\text{label} == 1$

2. EDA Insights - Correlations between labels

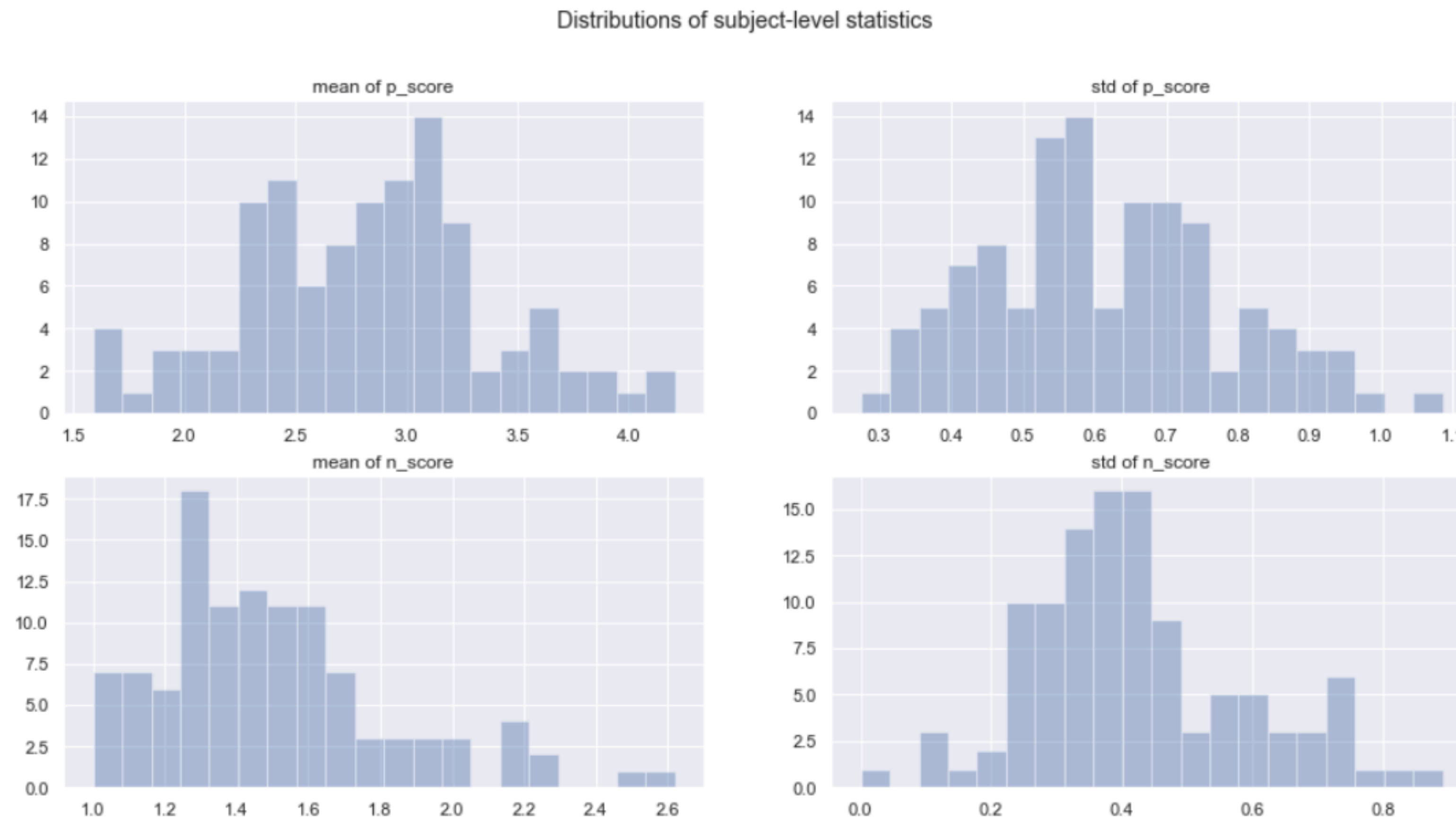
Correlations between labels make sense.
Positive measures and negative measures are positively correlated among each other respectively. And positive measures are negatively correlated with negative measures.



3. EDA Insights - By Subject Analysis

The subject level statistics indicates there is huge variance across subjects.Â

Below is the distribution of the individual mean and standard deviation of positive valence score and negative valence score. As we can see, they are widely spread when compared with the value range (from 1 to 5)



4. EDA Insights - Valence by Age Groups

As show in fig.1, the distribution of subjects' age. Two age groups (20s & 50s) constitute ~70% of all subjects. Fig.2, fig.3 are the distribution plots of p_score, n_score by age group. As we can see, the 50s group seem to be happier, and less unhappier than the 20s group. It might be because their emotional regulation skills are better.

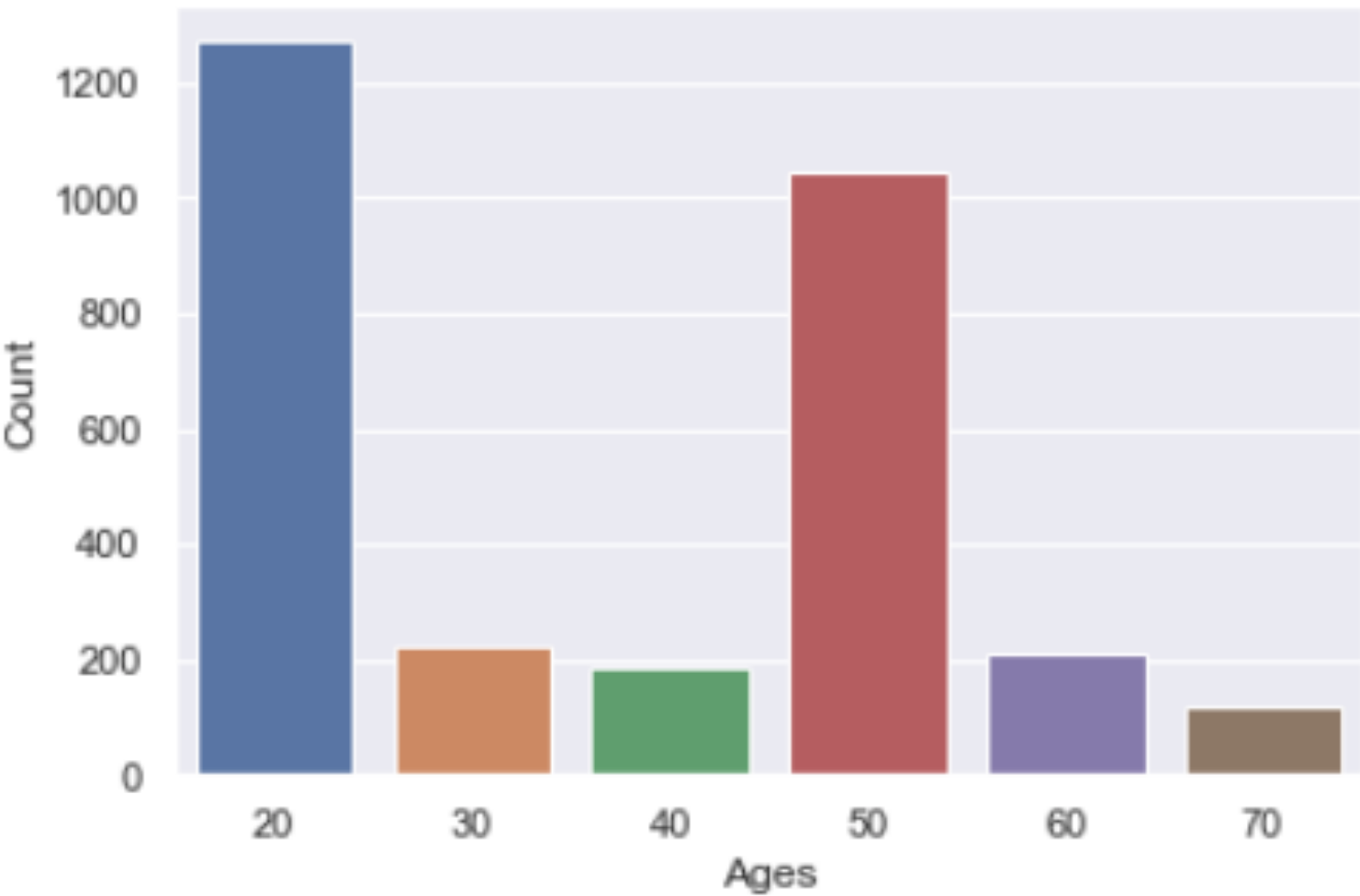


Fig1. Distribution of Participants' Age

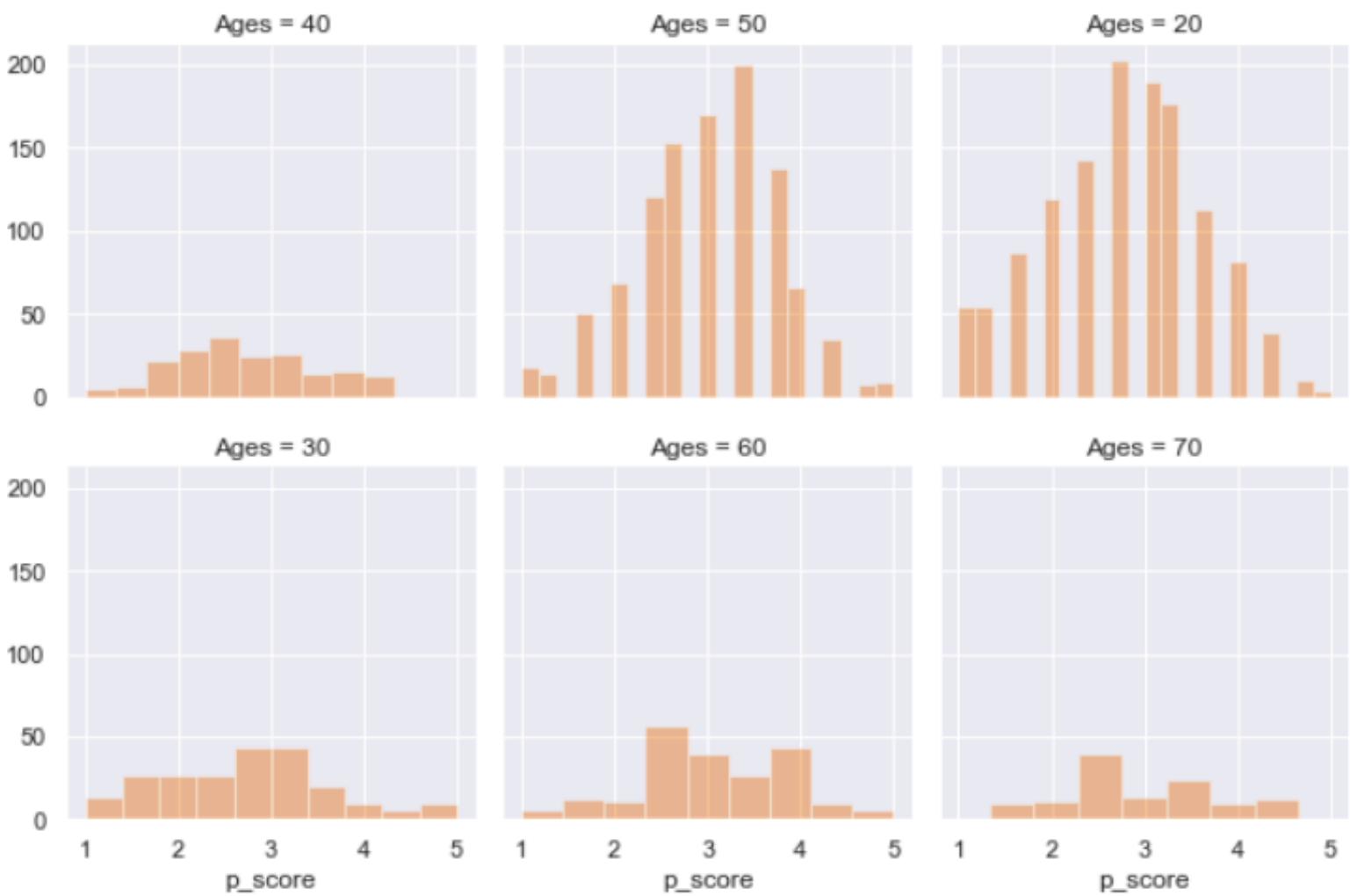


Fig2. Distribution p_score by Age Group

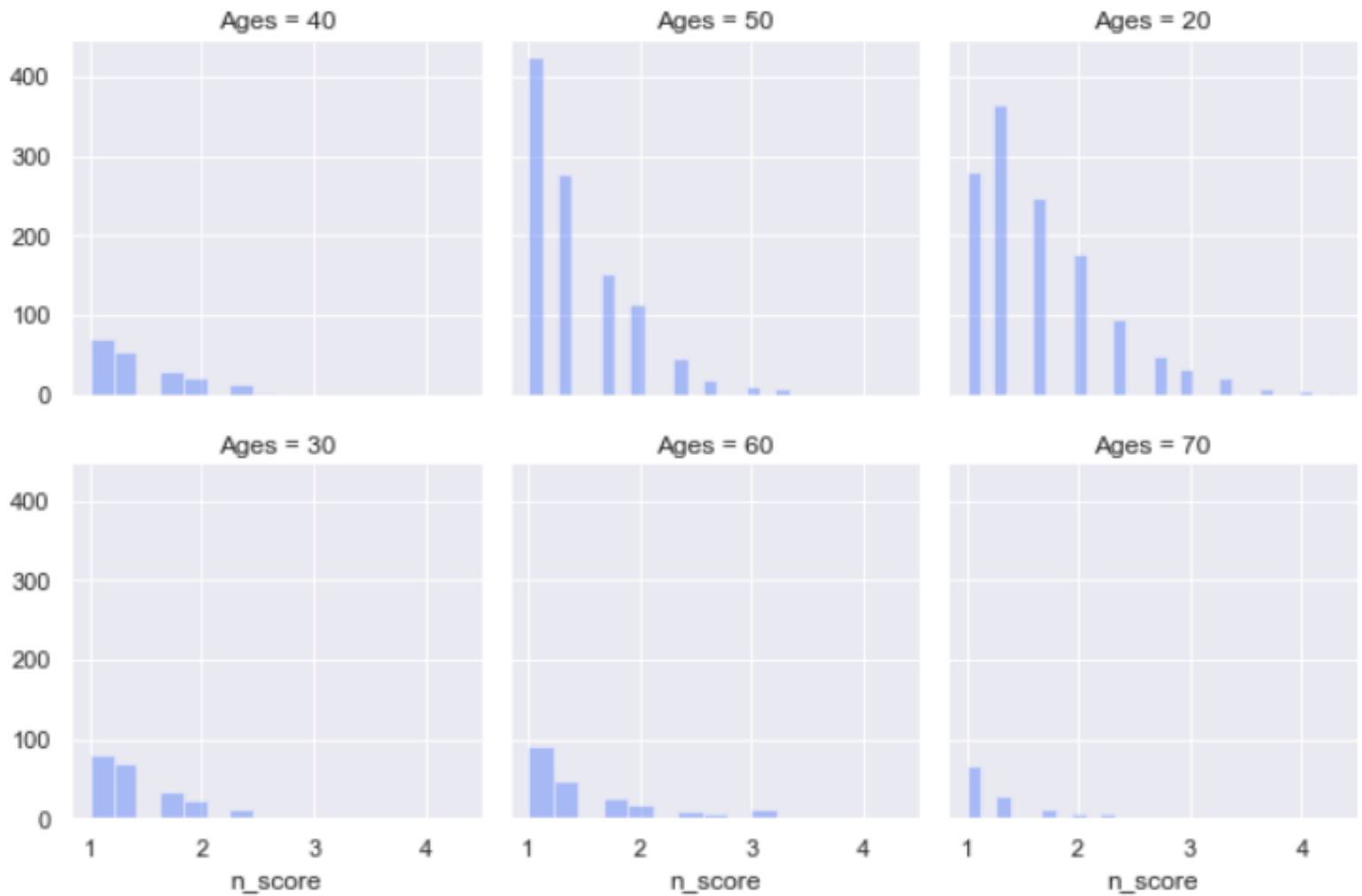
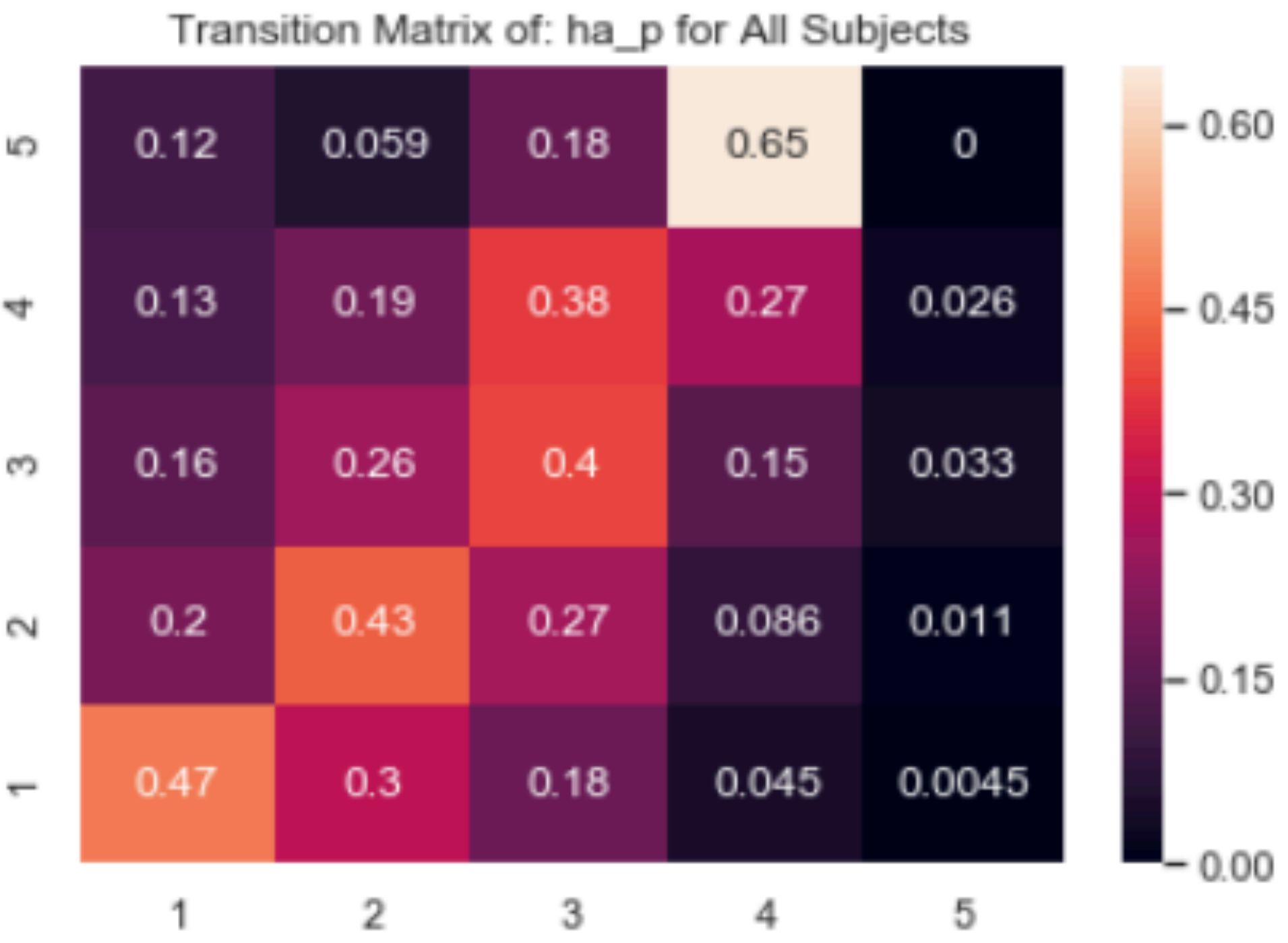
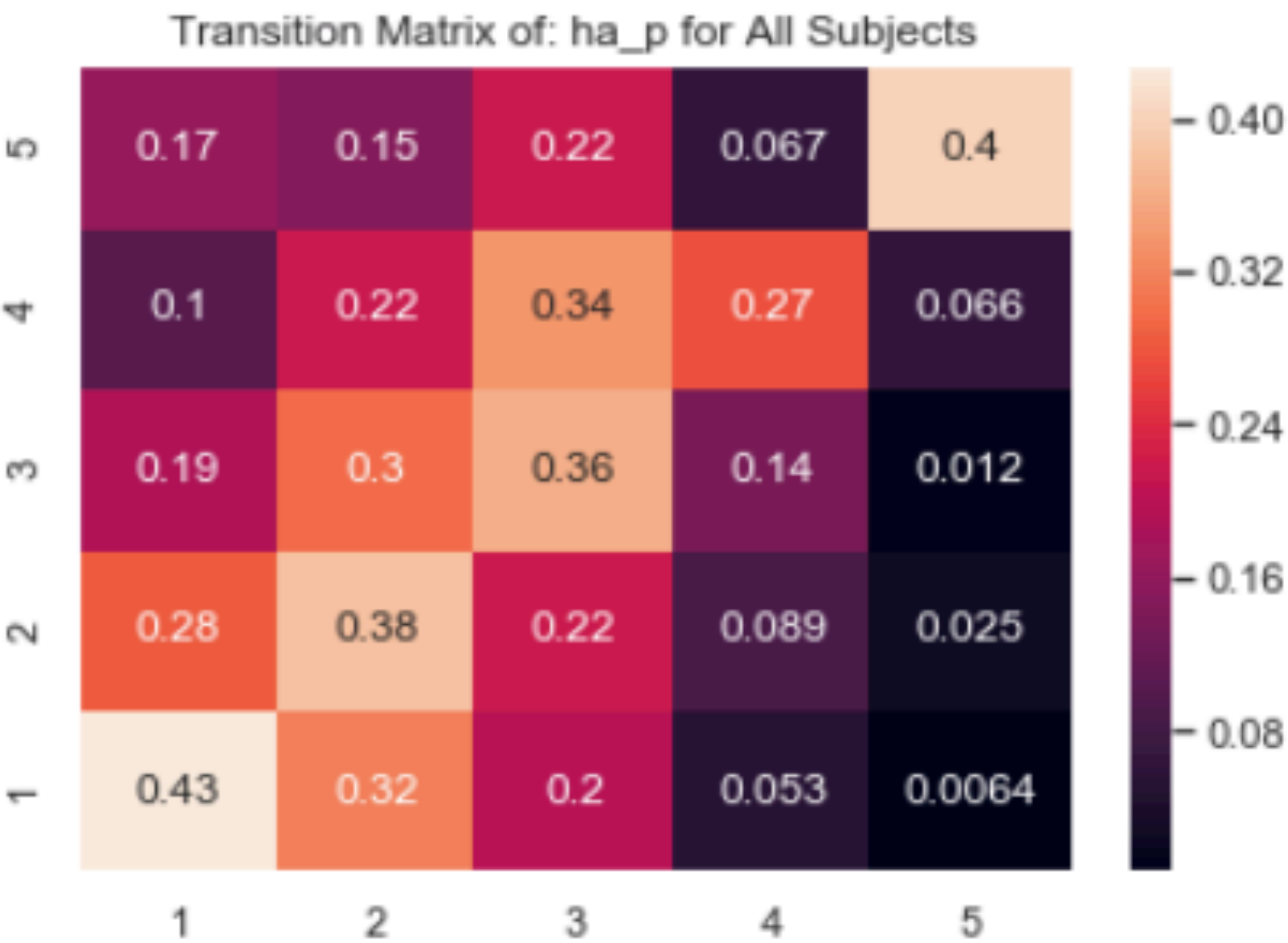


Fig3. Distribution n_score by Age Group

5. EDA Insights - Transition of Emotional States

This is the transition matrix of emotion during day time and before/after sleep for state 'ha_p' (high arousal positive).

As indicated below, sleep potentially helps maintain high score of 'ha_p'.



5. EDA Insights - Stationary Distribution of Emotional States

