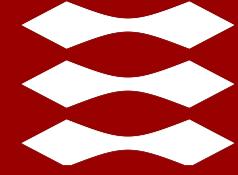


# Computational Data Analysis - Case 2

DTU

Group 16 - Anastasia Maftei (s226765), Beatriz Braga De Carvalho (s233576),

Raquel Moleiro Marques (s243636), Sree Keerthi Desu (s243933)



## Introduction

This project investigates whether self-perceived stress levels during cognitively demanding tasks aligns with physiological responses. Using data from the puzzle-based competition, we aim to answer: **Do self-reported emotional states reflect physiological signals during a puzzle-solving task?**

To do this, we apply clustering methods on specific features to identify participant profiles that reflect similar stress levels. These clusters correspond to the “**felt**” stress levels. These categories are then compared to the subjective self-reported variable *Frustrated* that represents “**perceived**” stress levels.

The task is now simply resolved to determining whether the felt and perceived stress levels match per individual.

## Dataset

- **EmoPairCompete:** 26 participants, 4 rounds each
- **Phases:** Rest → Task (Tangram puzzle) → Recovery

### Data collected:

- Physiological: HR, skin temp, EDA (phasic & tonic) via Empatica E4
- Self-reports: Emotion, frustration, difficulty, activation (per phase)

→ Our analysis **focuses on the extracted features provided**, not the raw signals.

## Feature Selection

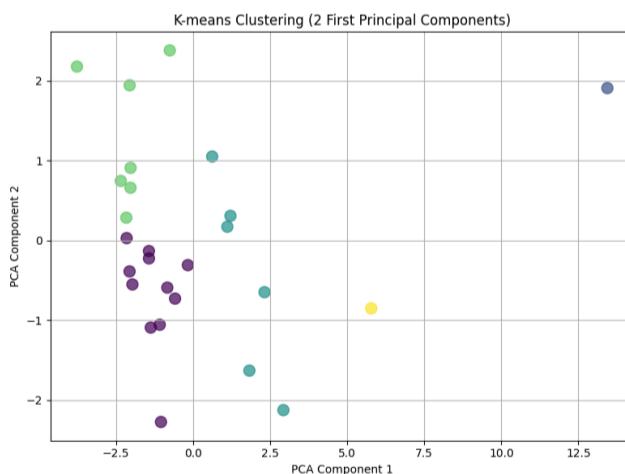
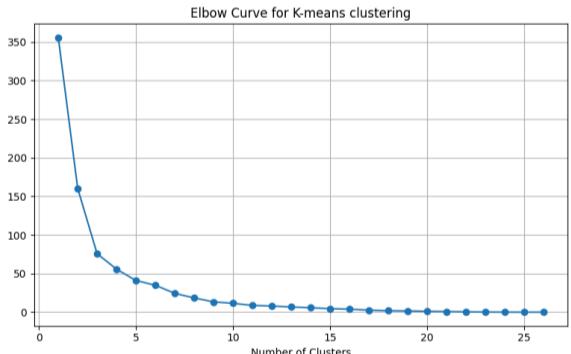
The EDA phasic component represents the event-related activity. Focusing only on **Phase 2** related data, we chose the EDA phasic features to analyse and extract the “**felt**” stress categories by using clustering methods. The resulting clusters/categories from performing the unsupervised analysis then are used to compare with the “Frustrated” variable from the questionnaire which represents an individual’s “**perceived**” stress.

## Clustering Methods

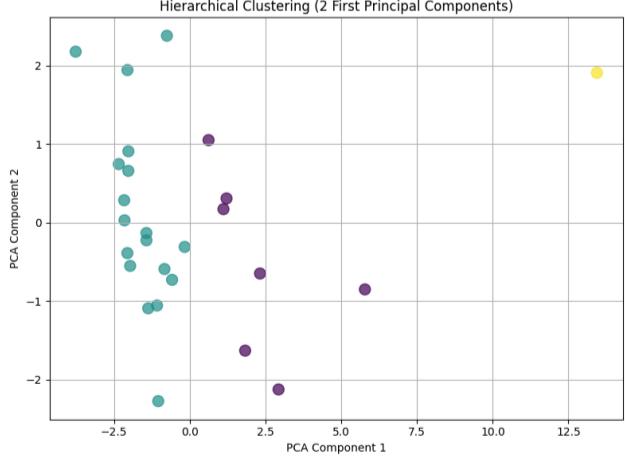
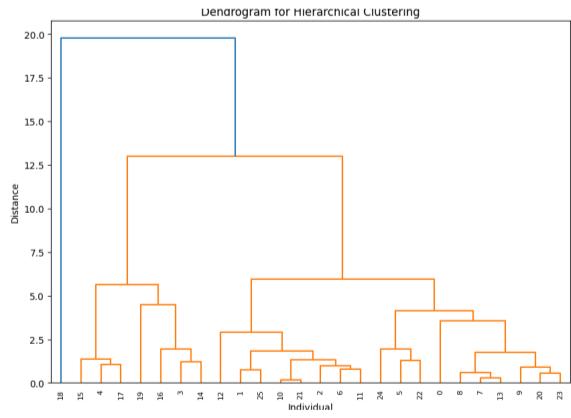
Before clustering, the data was **standardized** and **PCA** was applied to reduce dimensionality, retaining 3 components that explained over 90% of the variance, to ensure consistency and improve model performance.

We explored three clustering approaches to analyze physiological data collected during the task phase:

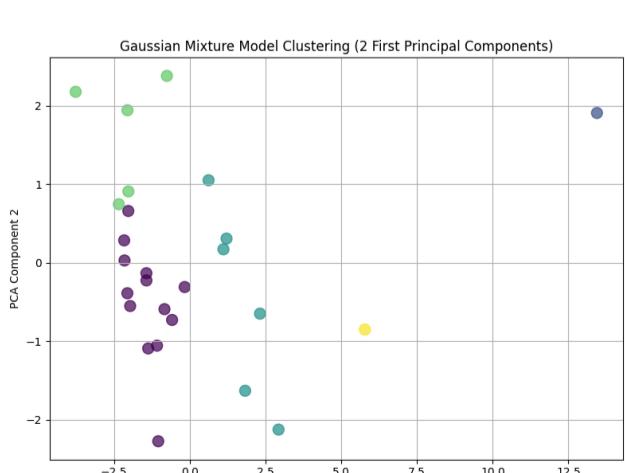
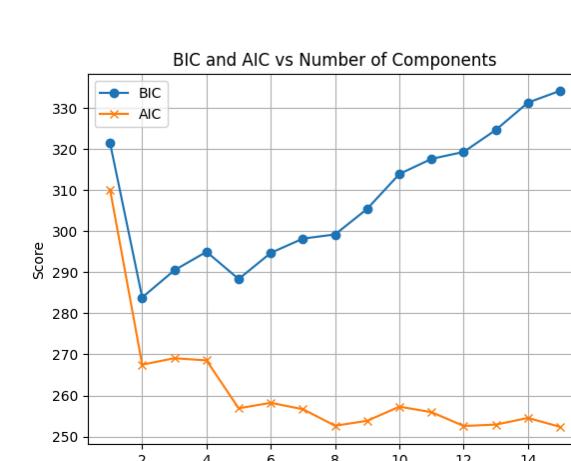
### K-means Clustering (K=5)



### Hierarchical Clustering (K=3)



### Gaussian Mixture Modelling (K=5)



## Model Selection

In order to choose the best model from the 3 clustering models, we use the **Davies-Bouldin score** which computes the average similarity between clusters. The lower the score, the better, as it implies that the clusters are more far apart and distinct.

Model	No. of clusters	Davies-Bouldin score
K-means clustering	5	0.838
Hierarchical clustering	3	0.690
Gaussian Mixture Model	5	0.832



## Analysis

We then found that the most important feature within the first principal component is the **EDA\_TD\_P\_AUC** feature. Using the mean of this feature within each cluster individually, we assigned each cluster a categorical mapping of stress level as follows,

Cluster	EDA_TD_P_AUC_Task	Felt stress
1	1408.764	Medium
2	131.291	Low
3	7328.021	High

The **Frustrated** variable from the questionnaire is then grouped by individual for phase 2 and the mean is calculated. Certain ranges of these values are then mapped to the different stress categories.

“Frustrated” values from questionnaire	Perceived stress
[0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2]	Low
[2.25, 2.5, 2.75, 3, 3.25, 3.5, 3.75, 4, 4.25]	Medium
[4.5, 4.75, 5, 5.25, 5.5, 5.75, 6]	High

These perceived and felt stress categories are then matched to answer our research question.

## Results

Individual	Frustrated	Perceived Stress	Felt Stress
1	2.0	Low	Low
2	2.0	Low	Low
3	0.5	Low	Low
4	0.25	Low	Medium
5	3.0	Medium	Medium
6	1.0	Low	Low
7	3.75	Medium	Low
8	4.5	High	Low
9	2.0	Low	Low
10	4.0	Medium	Low
11	3.5	Medium	Low
12	4.75	High	Low
13	4.75	High	Low
14	3.75	Medium	Low
15	5.25	High	Medium
16	2.25	Medium	Medium
17	3.0	Medium	Medium
18	4.5	High	Medium
19	1.25	Low	High
20	5.5	High	Medium
21	2.5	Medium	Low
22	2.0	Low	Low
23	3.75	Medium	Low
24	4.25	Medium	Low
25	4.75	High	Low
26	4.5	High	Low

Therefore with the above clusters, we ended up with the below:

→ **Reported SAME stress as they actually felt: 9**  
**Reported MORE stress than they actually felt: 15**  
**Reported LESS stress than they actually felt: 2**

## Conclusion

Our analysis reveals a disparity between the physiological and reported stress levels by the individuals. Not many are self-aware of the intensity of their stress, and **many reported that they felt more stressed** than the physiological features conveyed. This suggests that emotional interpretation of stress may be influenced by psychological or situational factors rather than purely physical responses.