



# Exploring Individual Differences in Children's Knowledge Organization

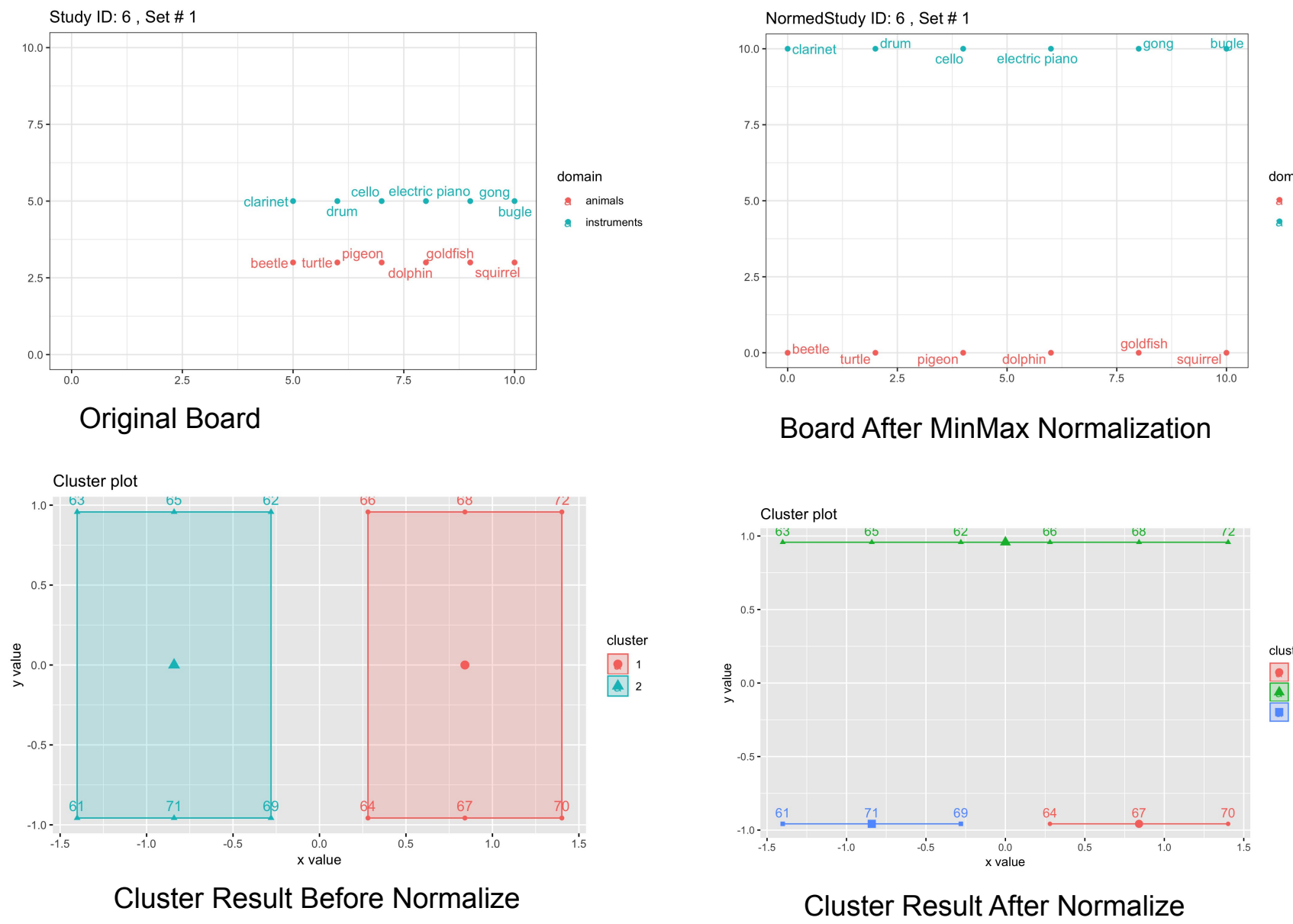
By Yizhi Zhang, Taeuk Kim, Jaeyeon Lee  
 Project Advisor: Zach Branson, Catarina Vales

## Background

In this project, we are interested in capturing individual differences in children's knowledge organization. The Department of Psychology conducted an experiment in which 62 children were asked to place 12 objects on a 10x10 board where similar objects are placed closer together. The goal of this project is to capture individual differences in children's arrangements.

## Data Pre-Processing:

The data for this project were provided by Catarina Vales of CMU's Department of Psychology. Our datasets consist of 62 subjects placing 5 different sets of 12 objects on a board. The x,y coordinates of each object on a 10x10 board and a type of object is prespecified by the experimenter. In dataset1, children worked with animal and instrument objects which relation type is prespecified as different domain or same domain. In dataset 2, children worked with bird, mammal, insect objects. In our method, to better understand how children are clustering objects, we used Average silhouette method of K-mean clustering.



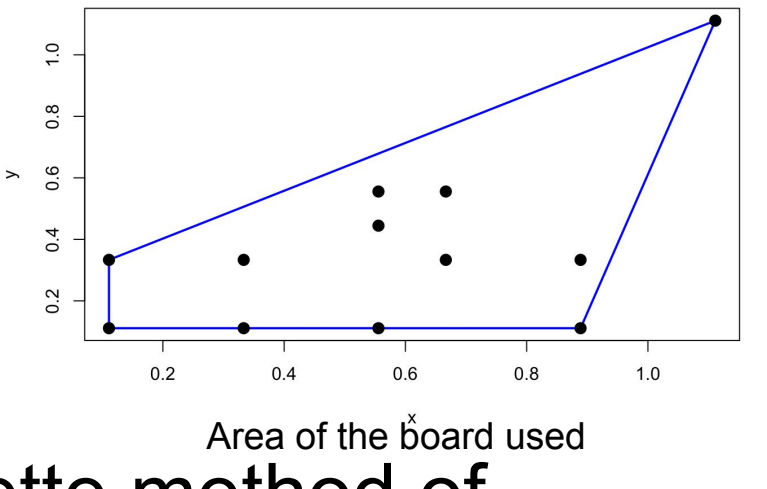
Before applying clustering methods, we normalized data points on the board using min-max method to rescale data points. This is to avoid from clusters tending to be separated along the axis with greater variance.

Data clean-up -- Child #61 not recorded

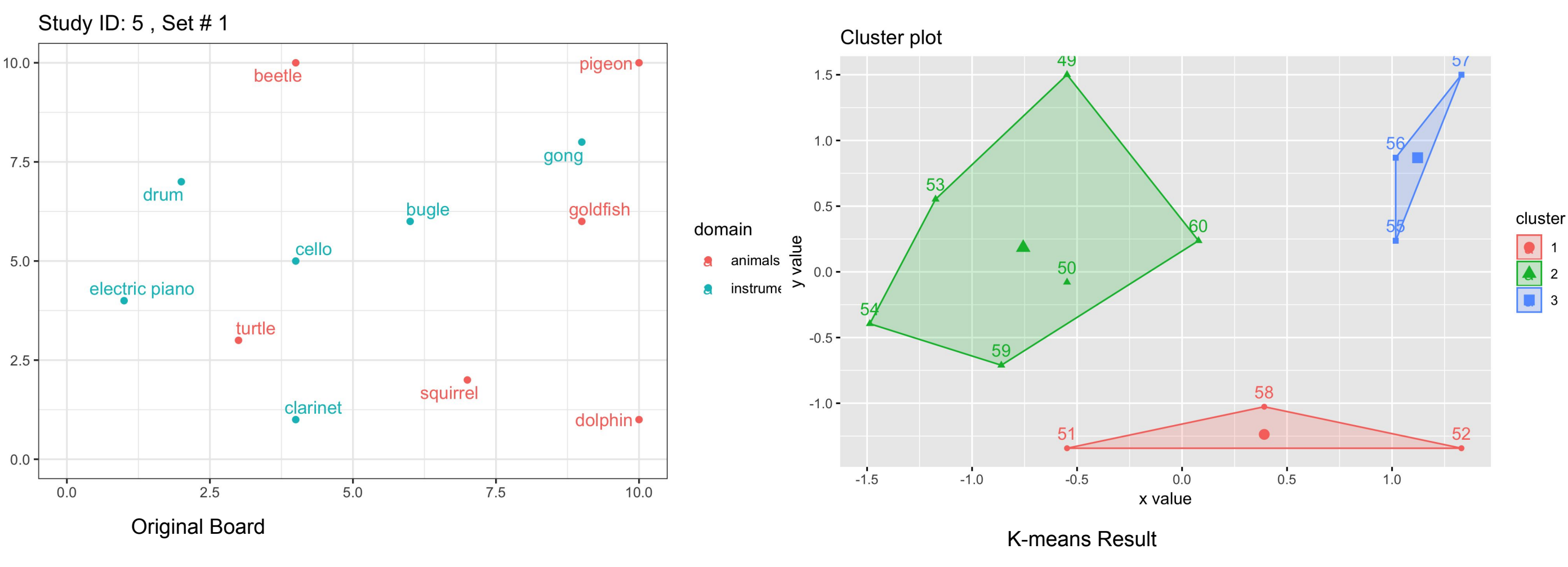
## Methods:

To understand how children are clustering pictures similarly or differently on boards, we use:

- the area of the board used by each child is captured by calculating the area of a convex polygon that encompass all picture points on the board.
- the number of clusters using the Average silhouette method of K-means Clustering
- the accuracy of clustering
- within-cluster sum of squared distances
- between-cluster sum of squared distances
- The ratio of within and between-cluster sum of squares distances



- Two figures below are children 5 board and the result of K-mean clustering of children 5 respectively.

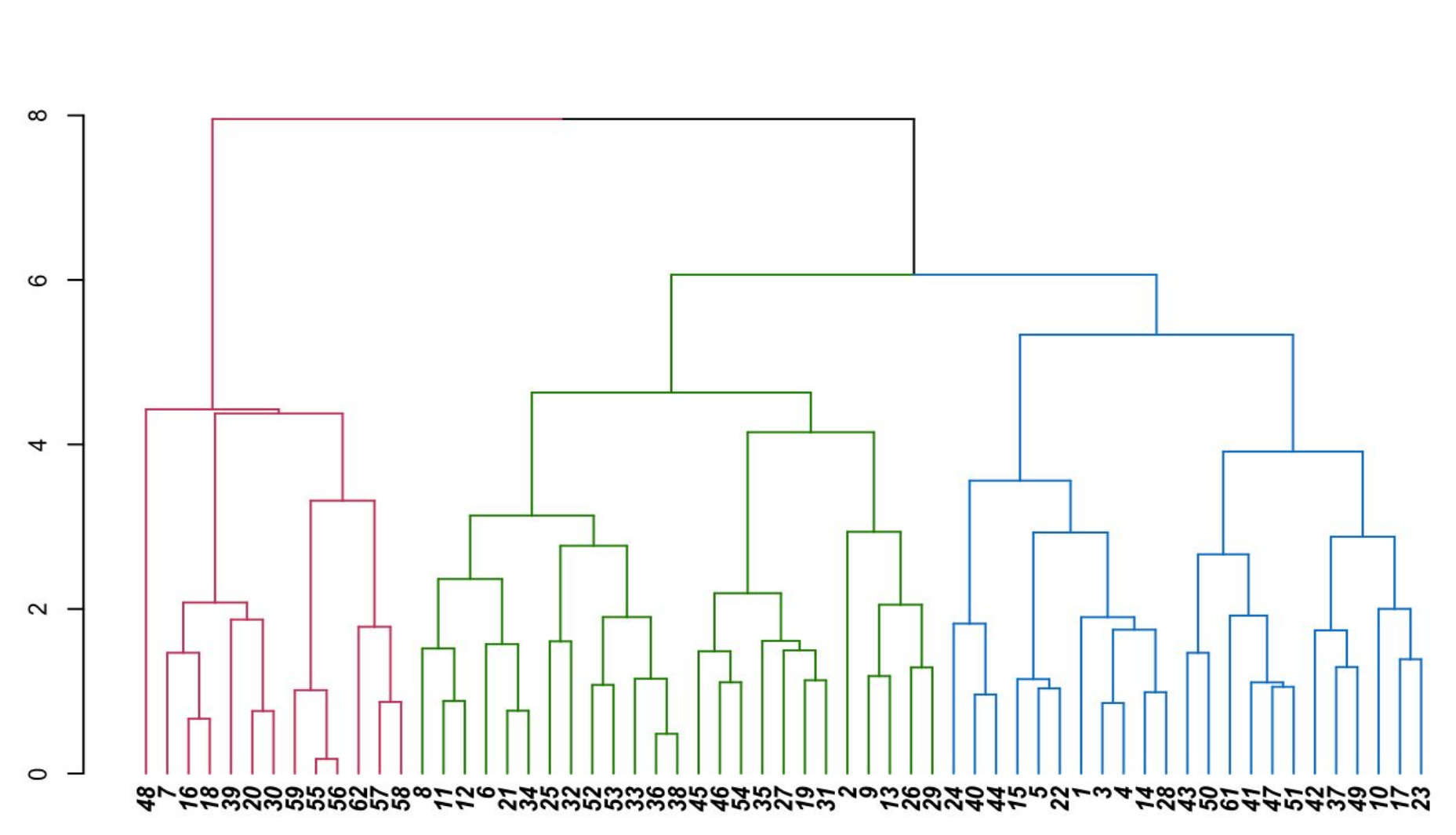


- The table below is the calculation result of 6 statistics for children 5.

# Clusters	withiness	betweeness	BSS/TSS	Area of Board	Accuracy
3	0.3	1.85	0.67	69	0.63

- We now have data frames with size 61 \* 6 because data for child 60 is not recorded. We then carry out the hierarchical clustering to compare clusters amongst children.
- For each kid and each board, we summarized the type of clustering done using 6 statistics (see above, refer to your table).
- Within each set, which kids are clustering similarly, and which are clustering differently?
- To answer this question, we can use hierarchical clustering on this 62x6 matrix to give us the groups of kids that are clustering most similarly to each other.

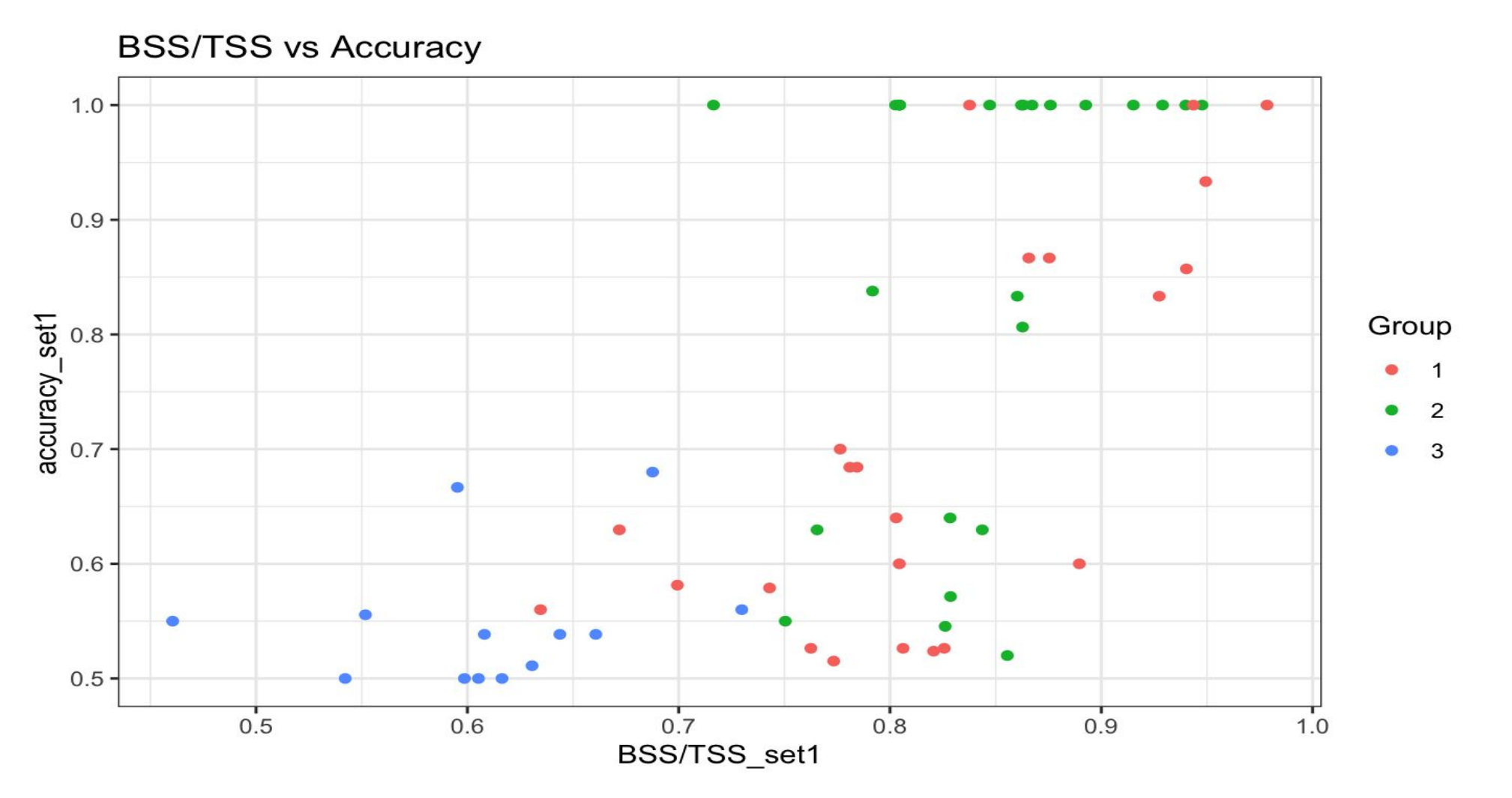
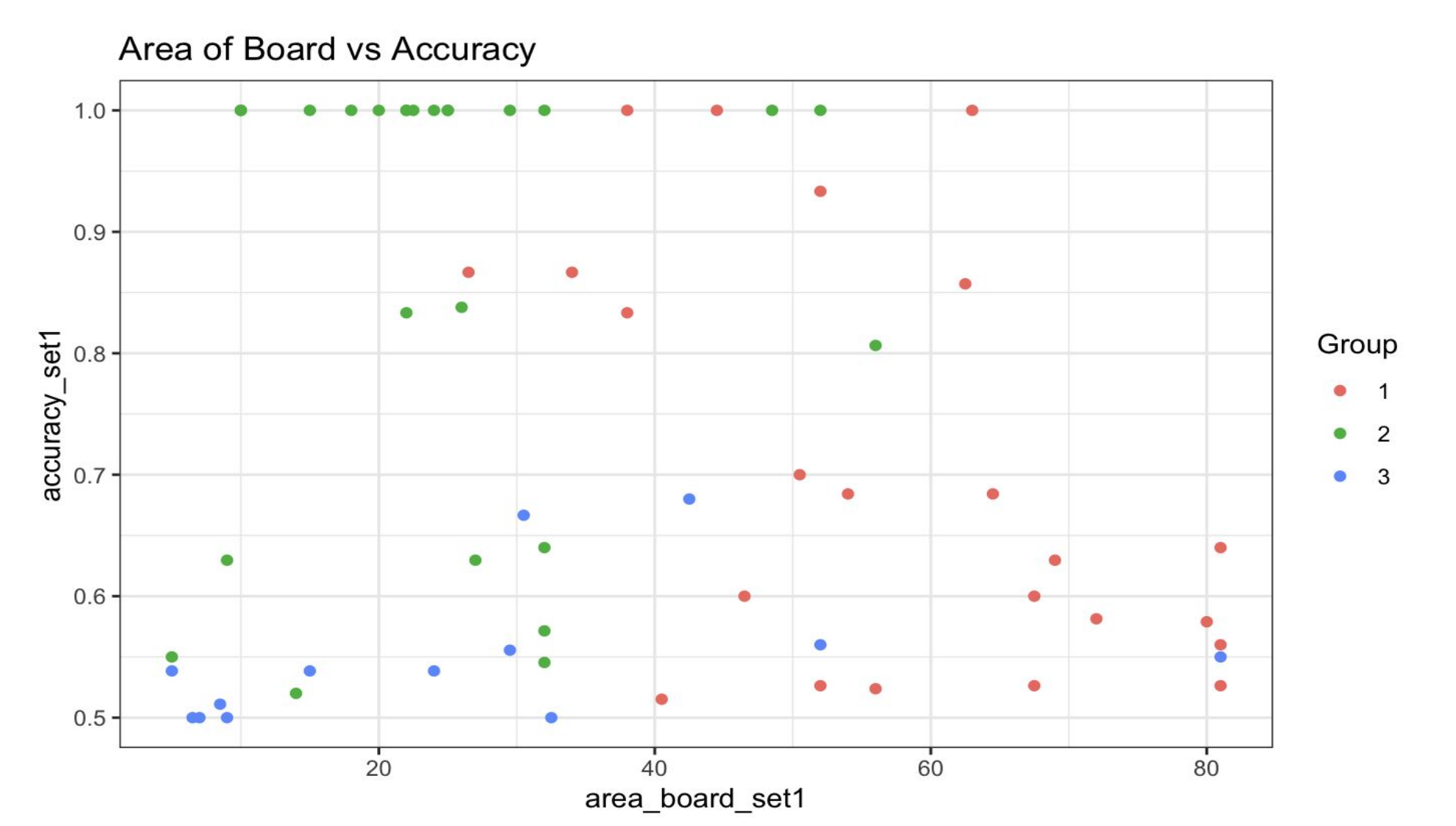
## Analysis and Results:



Number of Children in Each Group	
Group 1	23
Group 2	25
Group 3	13

- Hierarchical Clustering Dendrogram of Children Using Set 1

- Hierarchical clustering gives us groups of kids that are deemed to be clustering similarly to each other, but in what ways they are clustering similarly? We can plot the data in terms of the original variables, colored by group membership.



- Children in group 2 hierarchical clustering showed the highest accuracy rate of clustering different items.
- Children who used the least area of board showed lowest accuracy rate.
- Children with the highest accuracy maximized the distance between different groups but minimized the difference within each group.

## Conclusions:

- Using k-means and hierarchical clustering, we found children that tend to cluster similarly to each other,
- Children in group2 of hierarchical clustering demonstrated highest knowledge in discriminating items in different category with highest group age average
- Children ID (10, 15), (14, 28), (40, 41), (8, 9, 11, 12), (13, 19), (21, 27), (35, 36), (25, 34) remained in same group throughout the different sets of items. This suggest that analyzing the similarities and differences of these groups of participants could further suggest in what ways they are similar.