Data representation (matrix)

R1: A list with 10 zeros represents a subject, if the subject has been to a location type, then change zero at corresponding location in the list to 1;

R2: A list with 10 zeros represents a subject, if the subject has been to a location type, then change zero at corresponding location in the list to duration the subject stay at that location type;

R3: Frequency of a subject been to a location type

Output

I cluster subjects by the three data representation format above, then make plot for each cluster based on age and gender. It should be convenient to explore the relationship between cluster and some attributes. But the results are not as what I imaged. As you can see in the plot, it is hard to identify cluster, the reason may be there are so many points at one coordinate point meantime.

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithms | Input requirement | Data representation | Output |
| K-Means  (k=3) | matrix | R1 | Figure 1 |
| R2 |  |
| R3 |  |
| SpectralClustering  (cluster = 2) | matrix | R1 | 536 subjects belong to a cluster, the rest is in another cluster |
| R2 | All subjects belong to one cluster |
| R3 | 6 subjects belong to a cluster, the rest is in another cluster |

Note: green points in figures represent first of the clusters and red points represent another cluster, yellow points represent last cluster. And x-axis is gender, y-axis is age.

The problem is I have no idea how to visualize the results.

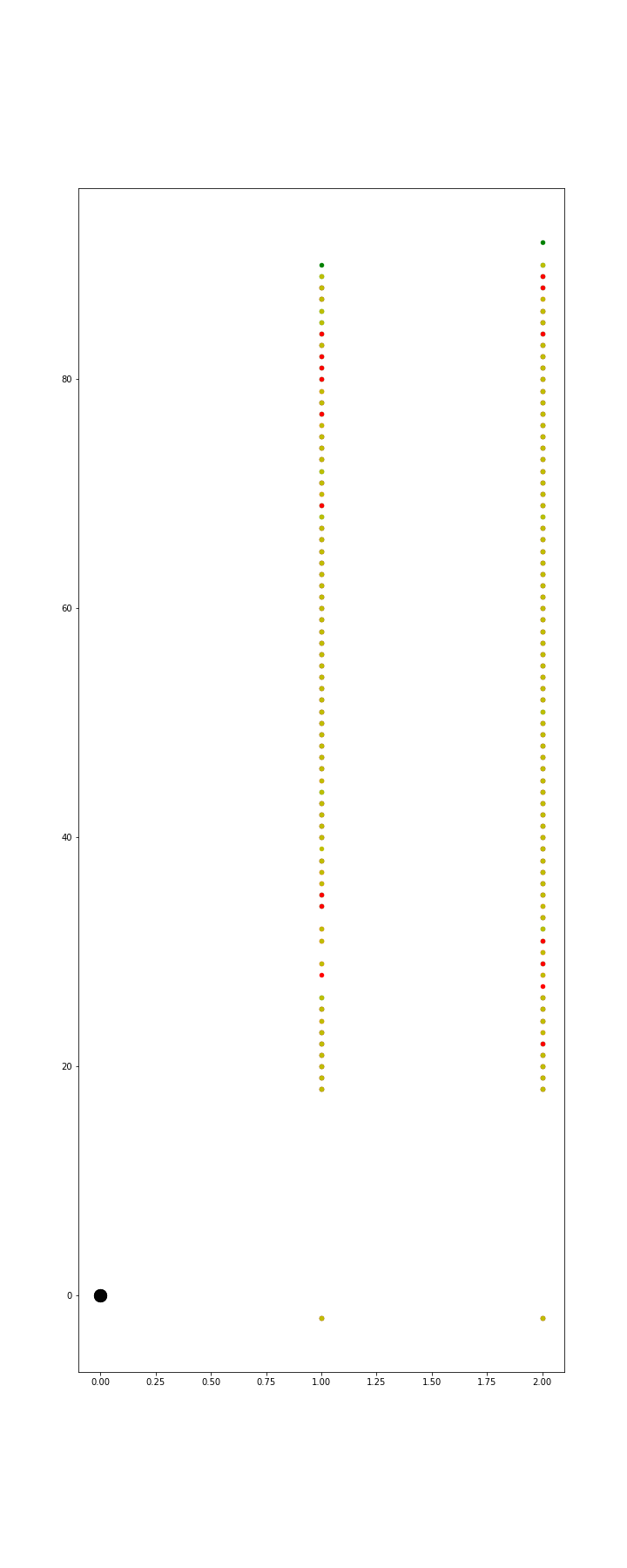


Figure 1