

“Are men looking for similar dating partner?”

An Analysis of Speed Dating Dataset

Presented By Aarushi Mishra (# 16)

Data Description

- The Speed Dating – Male dataset consist of 3413 observations and 59 variables, in total
- For analysis, I used the data corresponding to 22 males who signed-up for the event
- In addition to many other details, event authorities asked each participant to mention the qualities they are looking for in the opposite sex
- Each participant was asked to distribute points out of 100 among the desired qualities

Total rows: 3413 Total columns: 59

	iid	id	gender
1	11	1	1
2	11	1	1
3	11	1	1
4	11	1	1
5	11	1	1
6	11	1	1
7	11	1	1
8	11	1	1
9	11	1	1
10	11	1	1
11	12	2	1
12	12	2	1
13	12	2	1
14	12	2	1
15	12	2	1
16	12	2	1
17	12	2	1
18	12	2	1
19	12	2	1
20	12	2	1

...

•
•
•

Data Description

- These qualities include:

➤ Attractiveness **attr1_1**

➤ Sincerity **sinc1_1**


➤ Intelligence **intel1_1**

➤ Fun **fun1_1**

➤ Ambitious **amb1_1**

➤ Shared Interests **shar1_1**

After sorting the data suitably, the final data looks like this

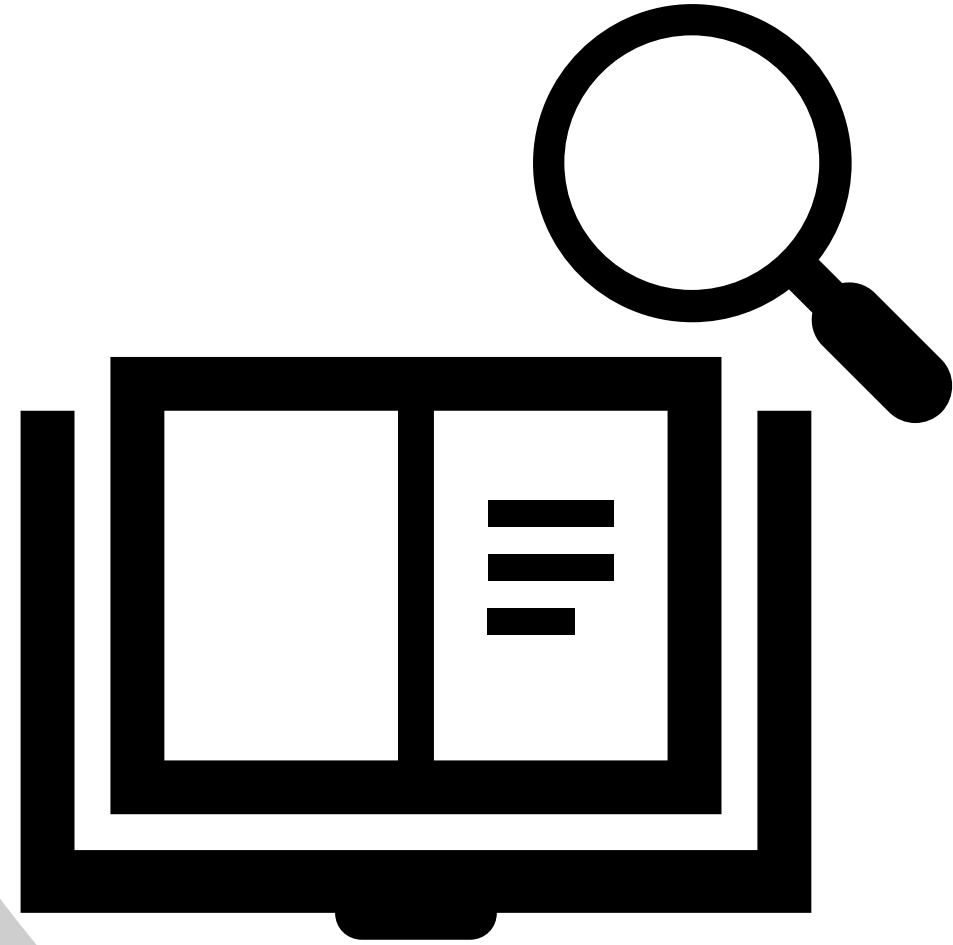


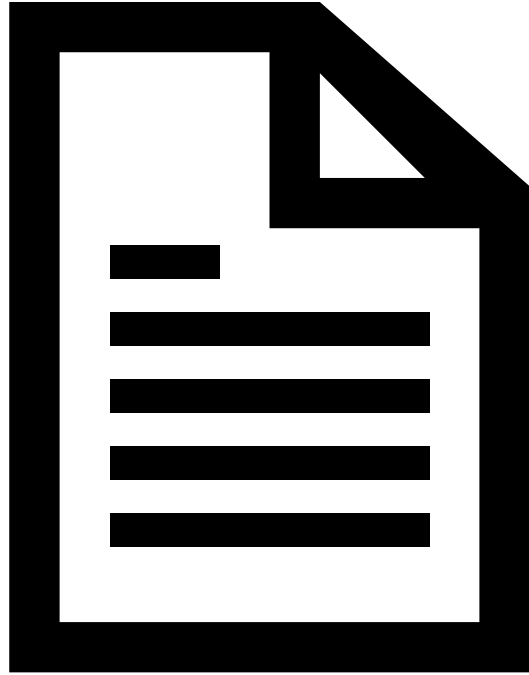
Obs	id	gender	attr1_1	sinc1_1	intel1_1	fun1_1	amb1_1	shar1_1	total_score
1	1	1	35	20	20	20	0	5	100.00
2	2	1	60	0	0	40	0	0	100.00
3	3	1	19	18	19	18	14	12	100.00
4	4	1	30	5	15	40	5	5	100.00
5	5	1	30	10	20	10	10	20	100.00
6	6	1	50	0	30	10	0	10	100.00
7	7	1	35	15	25	10	5	10	100.00
8	8	1	33.33	11.11	11.11	11.11	11.11	22.22	99.99
9	9	1	50	0	25	10	0	15	100.00
10	10	1	100	0	0	0	0	0	100.00
11	11	1	35	10	20	15	10	10	100.00
12	12	1	25	10	35	15	0	15	100.00
13	13	1	15	20	15	30	5	15	100.00
14	14	1	40	15	20	10	5	10	100.00
15	15	1	40	20	20	20	0	0	100.00
16	16	1	20	25	25	10	10	10	100.00
17	17	1	20	20	15	15	5	25	100.00
18	18	1	20	20	20	20	10	10	100.00
19	19	1	20	20	15	15	12	18	100.00
20	20	1	20	18	20	17	10	15	100.00
21	21	1	23	18	21	18	10	10	100.00
22	22	1	70	0	15	15	0	0	100.00

Objective of Analysis

- Based on the points distribution for these 6 qualities, perform clustering of observations
- Infer if there is any pattern in the way the observations are clustered
- Answer the question of analysis:

“Are men looking for similar dating partners?”





SAS Procedures Used

- PROC IMPORT
- PROC PRINT
- PROC CLUSTER
- PROC TREE
- PROC SORT
- PROC TABULATE
- PROC MEANS

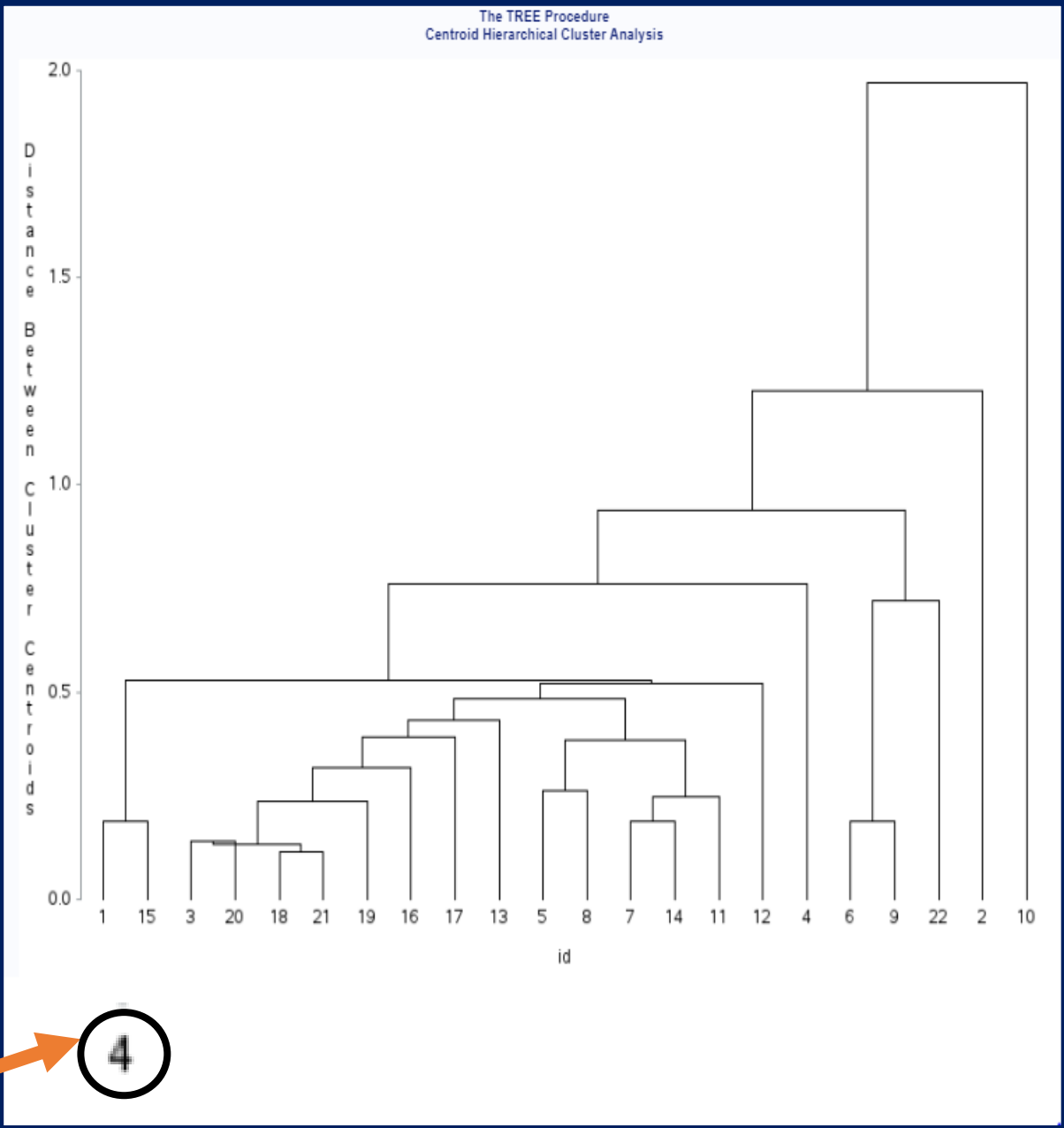
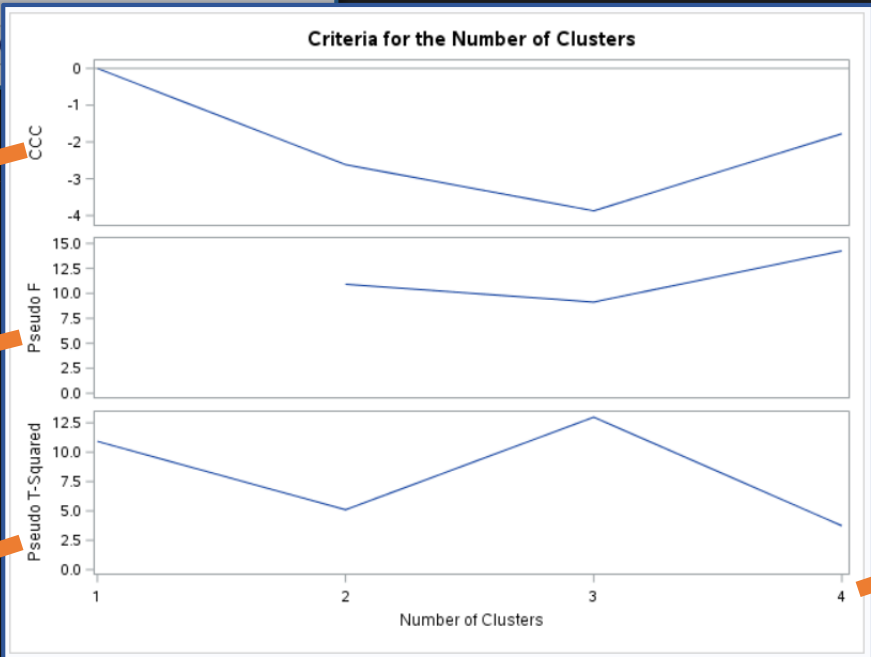
SAS Output

The CLUSTER Procedure Centroid Hierarchical Cluster Analysis

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	503.550184	393.958345	0.7085	0.7085
2	109.591839	59.040306	0.1542	0.8627
3	50.551533	14.500875	0.0711	0.9338
4	36.050658	25.074352	0.0507	0.9846
5	10.976305	10.976305	0.0154	1.0000
6	0.000001		0.0000	1.0000

Root-Mean-Square Total-Sample Standard Deviation 10.88363

Root-Mean-Square Distance Between



Pseudo T-Squared

Pseudo F

CCC

4

SAS Output

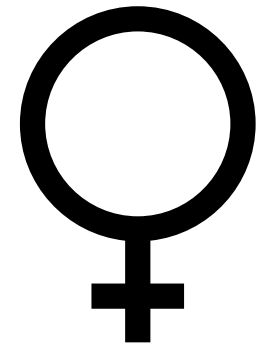
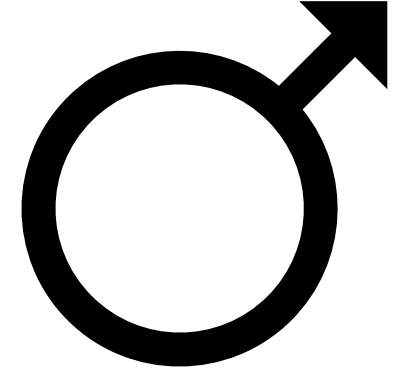
CLUSTER	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
1	17	attr1_1	17	27.0782353	8.1431138	15.0000000	40.0000000
		sinc1_1	17	16.1829412	5.2639289	5.0000000	25.0000000
		intel1_1	17	19.7711765	5.3335142	11.1100000	35.0000000
		fun1_1	17	17.3005882	7.7948544	10.0000000	40.0000000
		amb1_1	17	7.1829412	4.4110880	0	14.0000000
		shar1_1	17	12.4835294	6.4346415	0	25.0000000
2	3	attr1_1	3	56.6666667	11.5470054	50.0000000	70.0000000
		sinc1_1	3	0	0	0	0
		intel1_1	3	23.3333333	7.6376262	15.0000000	30.0000000
		fun1_1	3	11.6666667	2.8867513	10.0000000	15.0000000
		amb1_1	3	0	0	0	0
		shar1_1	3	8.3333333	7.6376262	0	15.0000000
3	1	attr1_1	1	60.0000000	.	60.0000000	60.0000000
		sinc1_1	1	0	.	0	0
		intel1_1	1	0	.	0	0
		fun1_1	1	40.0000000	.	40.0000000	40.0000000
		amb1_1	1	0	.	0	0
		shar1_1	1	0	.	0	0
4	1	attr1_1	1	100.0000000	.	100.0000000	100.0000000
		sinc1_1	1	0	.	0	0
		intel1_1	1	0	.	0	0
		fun1_1	1	0	.	0	0
		amb1_1	1	0	.	0	0
		shar1_1	1	0	.	0	0

- While most of the subjects are looking for a combination of attributes, few of them have rated few attributes to be absolutely more desirable over the others
- Cluster 1: men looking for combination of attributes but most desired one is attractiveness and least is ambitious
- Cluster 2: men do not desire for sincerity and sharing at all, attractiveness is highly rated!!
- Cluster 3 : 1 person who wants partner to be attractive and fun loving only
- Cluster 4: 1 person who desires for only good looks

Conclusion

- All the subjects can be broadly divided into 4 clusters / groups
- Out of 6 qualities, “attractiveness” is highest rated quality that men are looking for, followed by “intelligence”
- Majority of the subjects belong to one cluster

“Yes, most of the men are looking for similar dating partners!”





Thank You



Questions / Feedback?
