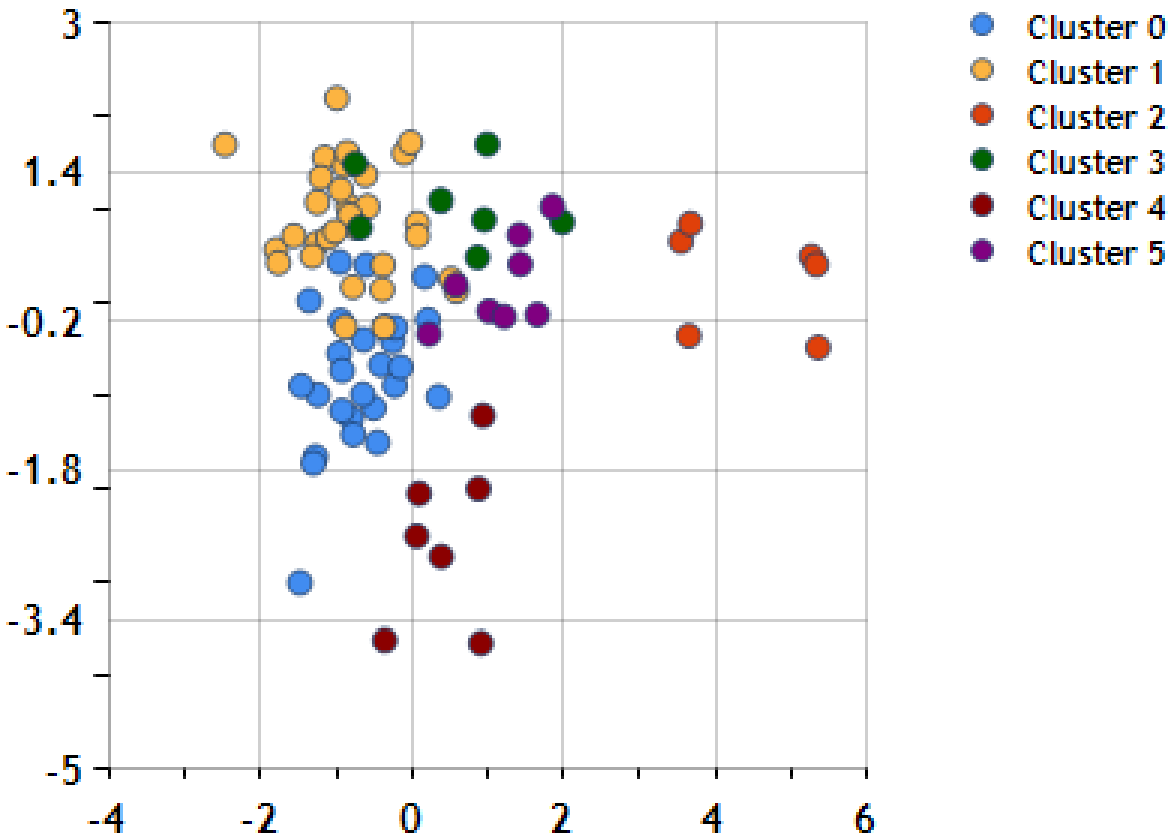


Cluster Analysis

Sean Bambrick
Matt Makowski

COM 631 Spring 2011



Note: This graph is not based on our data. We just thought it was cute.

Internal Variables:

C5: I like jokes that involve wordplay--puns, riddles, etc.

C6: I like sight gags.

C8: I enjoy Slapstick Comedy.

C20: I like sick humor.

C21: I enjoy humor that criticizes society.

C22: I like sarcasm.

C23: I like humor that is naughty.

C41: I like dark comedy.

C48: I enjoy irony.

C50: I like gross-out humor.

C66: I like "inside" jokes

External Variables:

G1: Are you male or female?

G2: What is your age?

G4: Which of the following categories best describes your political philosophy?

G9: What is your annual household income?

B1: Barack Obama is doing a good job as president.

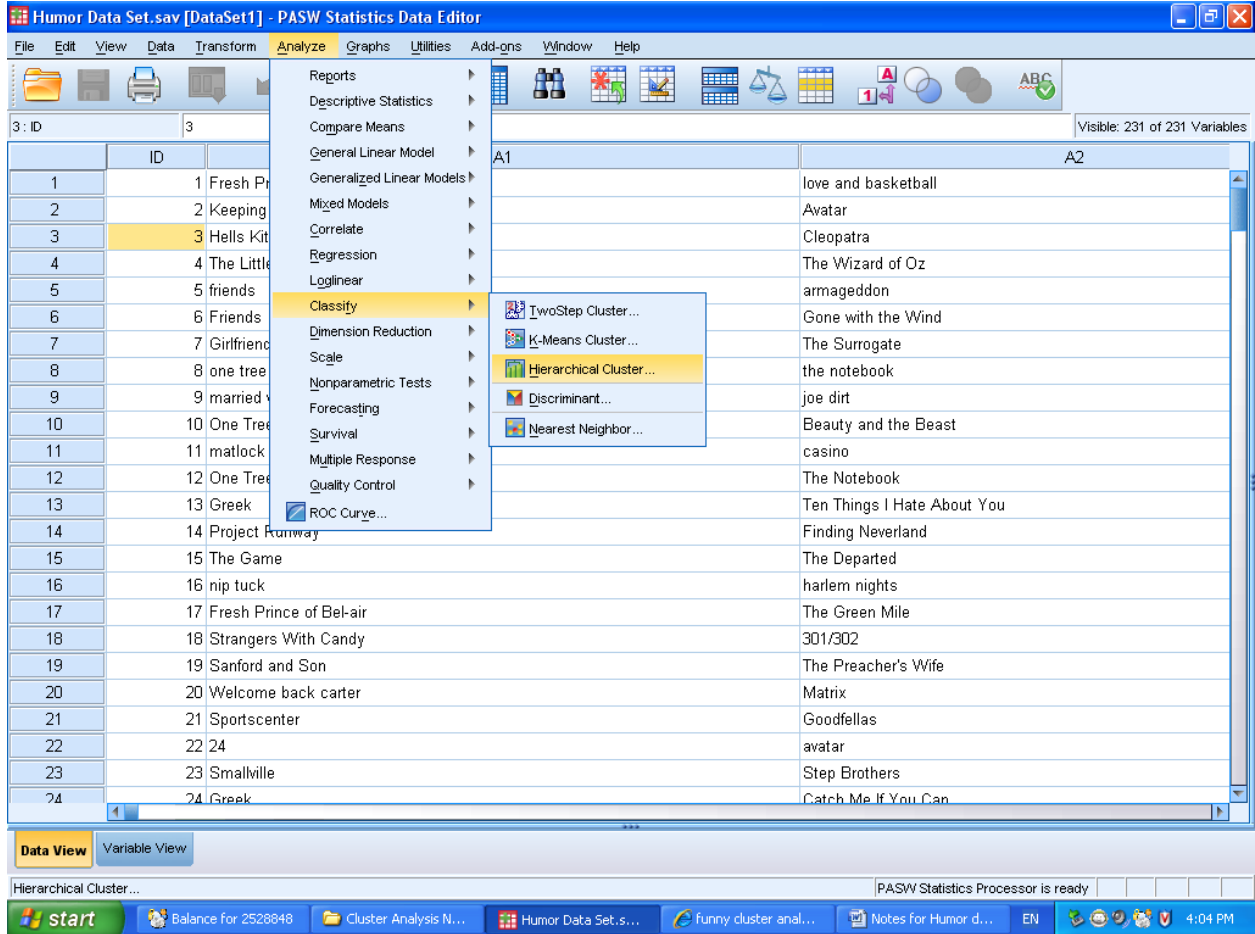
B5: I like finding unexpected linkages on the Internet.

B9: Marriages between same-sex couples should be recognized by the law as valid, with the same rights as traditional marriages.

B21. On this scale, how would you rate the Cleveland area?

B22. How would you rate the neighborhood you live in?

Analyze >> Classify >> Hierarchical Cluster



Select your Internal Variables for analysis (C5, C6, C8, C20, C21, C22, C23, C41, C48, C50, C66)

The screenshot shows the PASW Statistics Data Editor interface. A dialog box titled "Hierarchical Cluster Analysis" is open, allowing the user to select variables for analysis. The "Variables(s):" list contains the following variables: C23, C41, C48, C50, and C66. The "Label Cases by:" field is empty. The "Display" section has "Statistics" and "Plots" checked. The background data table has columns ID, A1, and A2.

ID	A1	A2
1	1 Fresh Prince of Belair	love and basketball
2	2 Keeping up with the Kardashians	Avatar
3	3 Hells Kitchen	Cleopatra
4	4 The Little Rascals	Oz
5	5 friends	Wind
6	6 Friends	Beast
7	7 Girlfriends	ate About You
8	8 one tree hill	and
9	9 married with children	
10	10 One Tree Hill	
11	11 matlock	
12	12 One Tree Hill	
13	13 Greek	
14	14 Project Runway	
15	15 The Game	
16	16 nip tuck	
17	17 Fresh Prince of Bel-air	
18	18 Strangers With Candy	
19	19 Sanford and Son	The Preacher's Wife
20	20 Welcome back carter	Matrix
21	21 Sportscenter	Goodfellas
22	22 24	avatar
23	23 Smallville	Step Brothers
24	24 Greek	Catch Me If You Can

Click “Statistics” box and select “Range of Solutions” and indicate your range (3 to 6 is common).

Also, make sure that “Agglomeration Schedule” is checked.

The screenshot shows the PASW Statistics Data Editor interface with the Hierarchical Cluster Analysis dialog box open. The background data table is as follows:

ID	Case Name	Variable
1	Fresh Prince of Belair	love and basket
2	Keeping up with the Kardashians	Avatar
3	Hells Kitchen	Cleopatra
4	The Little Rascals	
5	friends	
6	Friends	
7	Girlfriends	
8	one tree hill	
9	married with children	
10	One Tree Hill	
11	matlock	
12	One Tree Hill	
13	Greek	
14	Project Runway	
15	The Game	
16	nip tuck	
17	Fresh Prince of Bel-air	
18	Strangers With Candy	
19	Sanford and Son	The Preacher's Wife
20	Welcome back carter	Matrix
21	Sportscenter	Goodfellas
22	22 24	avatar
23	Smallville	Step Brothers
24	Greek	Catch Me If You Can

The Hierarchical Cluster Analysis dialog box has the following settings:

- Agglomeration schedule
- Proximity matrix
- Cluster Membership:
 - None
 - Single solution
 - Range of solutions
 - Minimum number of clusters: 3
 - Maximum number of clusters: 6
- Label Cases by:
 - Cluster
 - Cases
 - Variables
- Display:
 - Statistics
 - Plots

The 'Statistics...' button is highlighted with a blue box.

Select "Plots" tab -- note that you must select either the "dendrogram" box or something under "icicle."

(Note, you don't need either for five stats, but you are forced by SPSS!)

The screenshot shows the SPSS Statistics Data Editor interface with the 'Humor Data Set.sav' file open. The main window displays a list of 24 cases with their IDs and names. Two dialog boxes are open over the data:

- Hierarchical Cluster Analysis Dialog:**
 - Variables(s):** C23, C41, C48, C50, C66
 - Label Cases by:** (empty)
 - Cluster:** Cases Variables
 - Display:** Statistics Plots
- Hierarchical Cluster Plots Dialog:**
 - Dendrogram:**
 - Icicle:**
 - All clusters
 - Specified range of clusters
 - Start cluster: 1
 - Stop cluster: (empty)
 - By: 1
 - None
 - Orientation:**
 - Vertical
 - Horizontal

The taskbar at the bottom shows the Windows Start menu, system tray, and taskbar with several open applications including 'Balance for 2528848', 'Cluster Analysis N...', 'Humor Data Set.s...', 'funny cluster anal...', and 'Notes for Humor d...'. The system clock shows 4:20 PM.

Click “Method” select “Ward’s Method” from the drop-down menu and click “Continue”

The screenshot shows the SPSS Hierarchical Cluster Analysis: Method dialog box. The 'Cluster Method' is set to 'Between-groups linkage'. Under 'Measure', 'Interval' is selected, and 'Ward's method' is chosen from the list. The 'Transform Values' section has 'Standardize' set to 'None' and 'By case' selected. The 'Transform Measure' section has 'Rescale to 0-1 range' checked. The 'Display' section has 'Statistics' and 'Plots' checked. The background shows a data editor window with a list of TV shows and their IDs.

ID	3	A1	A2
1	1	Fresh Prince of Belair	
2	2	Keeping up with the Kardashians	
3	3	Hells Kitchen	
4	4	The Little Rascals	
5	5	friends	
6	6	Friends	
7	7	Girrfriends	
8	8	one tree hill	
9	9	married with children	
10	10	One Tree Hill	
11	11	matlock	
12	12	One Tree Hill	
13	13	Greek	
14	14	Project Runway	
15	15	The Game	
16	16	nip tuck	
17	17	Fresh Prince of Belair	
18	18	Strangers With Candy	
19	19	Sanford and Son	The Preacher's Wife
20	20	Welcome back carter	Matrix
21	21	Sportscenter	Goodfellas
22	22	24	avatar
23	23	Smallville	Step Brothers
24	24	Greek	Catch Me If You Can

“Save” and select range of solutions, “3” for minimum number of clusters and “6” for maximum number of clusters.

Click “Continue”

Click “Ok”

The screenshot shows the PASW Statistics Data Editor interface. The main window displays a data table with columns ID, A1, and A2. The data includes rows for various TV shows and movies, such as 'Fresh Prince of Belair', 'Keeping up with the Kardashians', 'Hells Kitchen', 'The Little Rascals', 'friends', 'Friends', 'Girrfriends', 'one tree hill', 'married with children', 'One Tree Hill', 'matlock', 'One Tree Hill', 'Greek', 'Project Runway', 'The Game', 'nip tuck', 'Fresh Prince of Belair', 'Strangers With Candor', 'Sanford and Son', 'Welcome back carter', 'Sportscenter', '24', 'Smallville', and 'Greek'. The A1 column contains show titles, and the A2 column contains related terms or genres.

Overlaid on the data table is the 'Hierarchical Cluster Analysis' dialog box. The dialog has the following settings:

- Cluster Membership: None, Single solution, Range of solutions
- Number of clusters: []
- Minimum number of clusters: 3
- Maximum number of clusters: 6

The 'Continue' button is highlighted in blue. Other buttons visible include 'OK', 'Paste', 'Reset', 'Cancel', 'Help', 'Statistics...', 'Plots...', 'Method...', and 'Save...'. The status bar at the bottom indicates 'PASW Statistics Processor is ready' and the system clock shows '4:22 PM'.

Now we examine the cluster groupings...
 Analyze > Descriptive Statistics > Frequencies

The screenshot shows the PASW Statistics Data Editor interface. The main window displays a data table with 24 rows and two columns: 'ID' and 'A2'. The 'A2' column contains movie titles. The 'Analyze' menu is open, and the 'Frequencies...' option is selected. The 'Data View' button is visible at the bottom of the window.

ID	A2
1	love and basketball
2	Avatar
3	Cleopatra
4	The Wizard of Oz
5	armageddon
6	Gone with the Wind
7	The Surrogate
8	the notebook
9	joe dirt
10	Beauty and the Beast
11	casino
12	The Notebook
13	Ten Things I Hate About You
14	Finding Neverland
15	The Departed
16	harlem nights
17	The Green Mile
18	301/302
19	The Preacher's Wife
20	Matrix
21	Goodfellas
22	avatar
23	Step Brothers
24	Catch Me If You Can

Add the cluster variables “Ward Method-- 6 Cluster [CLU6_1]” “Ward Method-- 5 Cluster [CLU5_1]” “Ward Method-- 4 Cluster [CLU4_1]” “Ward Method-- 3 Cluster [CLU3_1]” These new variables will be at the bottom of the list from the box on the left side.

Next, Click “ok” tab to run the Frequencies Statistics

The screenshot shows the PASW Statistics Data Editor interface. The main window displays a data table with columns ID, A1, A2, and a fourth column. The data table contains 30 rows of data. A dialog box titled "Frequencies" is open, showing a list of variables on the left and a list of selected variables on the right. The selected variables are Ward Method--6 Clusters [CLU6_1], Ward Method--5 Clusters [CLU5_1], Ward Method--4 Clusters [CLU4_1], and Ward Method--3 Clusters [CLU3_1]. The dialog box also has buttons for Statistics..., Charts..., and Format..., and a checkbox for "Display frequency tables" which is checked. The status bar at the bottom indicates "PASW Statistics Processor is ready" and the system clock shows 5:42 PM.

ID	A1	A2	
1	Fresh Prince of Belair	love and basketball	the hangover
2	Keeping up with the Kardashians	Avatar	Step Brothers
3	Hells Kitchen		
4	The Little Rascals		
5	friends		
6	6 Friends		
7	7 Girlfriends		
8	8 one tree hill		
9	9 married with children		
10	10 One Tree Hill		
11	11 matlock		
12	12 One Tree Hill		
13	13 Greek		
14	14 Project Runway		
15	15 The Game		
16	16 nip tuck	harlem nights	friday
17	17 Fresh Prince of Bel-air	The Green Mile	The Hang over
18	18 Strangers With Candy	301/302	The Hangover
19	19 Sanford and Son	The Preacher's Wife	Meet the Fu'kers
20	20 Welcome back carter	Matrix	Goonies
21	21 Sportscenter	Goodfellas	Knocked Up
22	22 24	avatar	the hangover
23	23 Smallville	Step Brothers	Step Brothers
24	24 Greek	Catch Me If You Can	The Hangover
25	25 Good Times	The Last Dragon	Harlem Nights
26	26 Keeping up with the Kardashians	Avatar	Step Brothers
27	27 Keeping Up With The Kardashians	Avatar	Step Brothers
28	28 The Sopranos	The Thing	Pee Wee's Big Adventure
29	29 Wonderfalls	Kill Bill	Zoolander
30	30 Friends	Cleopatra	Dodgeball or Animal Crackers

Run Means (with ANOVA tests) to compare means among the clusters (there are 6 selected in this case).

Click “Analyze” >> “Compare Means” >> “Means”

List Both Internal and External Variables Combined:

Internal Variables: C5, C6, C8, C20, C21, C22, C23, C41, C48, C50, C66

External Variables: G1, G2, G4, G9, B1, B5, B9, B21, B22

Under the “Independent List” select “Ward Method-- 6 Cluster [CLU_6]” (Note, only select this one, not 3-6 like we did in the Frequencies Table.)

The screenshot shows the PASW Statistics Viewer interface. The main window displays the 'Frequencies' output for 'Ward Method--6 Clusters'. A 'Means' dialog box is open, showing the 'Independent List' with 'Ward Method--6 Clusters' selected. Below the dialog, the 'Ward Method--6 Clusters' table is visible, showing the distribution of cases across six clusters.

Statistics
Ward Method--6 Clusters

N	Valid	248
	Missing	40
Mean		3,04

Ward Method--6 Clusters

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	69	24.0	27.8	27.8
2	22	7.6	8.9	36.7
3	48	16.7	19.4	56.0
4	66	22.9	26.6	82.7
5	24	8.3	9.7	92.3
6	19	6.6	7.7	100.0
Total	248	86.1	100.0	
Missing System	40	13.9		
Total	288	100.0		

Click “Options” tab and Check “ANOVA table and eta” box at the bottom left.

The screenshot shows the PASW Statistics Viewer interface. The main window displays the 'Frequencies' output for a dataset. Two dialog boxes are open: 'Means' and 'Means: Options'. The 'Means: Options' dialog is the primary focus, showing a list of statistical measures to be calculated. The 'Anova table and eta' checkbox is checked, indicating that ANOVA results and eta-squared values will be included in the output. The 'Test for linearity' checkbox is unchecked. The 'Means' dialog in the background shows the selection of dependent and independent variables for the analysis.

	Total	248	86.1	100.0
Missing System	40	13.9		
Total	288	100.0		

Click “Continue”

Click “OK” to compare means.

```

CLUSTER  C5 C6 C8 C20 C21 C22 C23 C41 C48 C50 C66
/METHOD WARD
/MEASURE=SEUCLID
/PRINT SCHEDULE CLUSTER(3,6)
/PLOT VICICLE
/SAVE CLUSTER(3,6).
    
```

Notice that the Cluster procedure does not generate much printed output that's of use to you, unless you are interested in specific data cases. The real valuable output comes with using the Means procedure to compare cluster means.

Cluster

[DataSet1] E:\Comm 631\Humor Data Set\Humor Data Set.sav

Case Processing Summary^{a,b}

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
248	86.1	40	13.9	288	100.0

a. Squared Euclidean Distance used

b. Ward Linkage

Ward Linkage

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	153	198	.000	0	0	16
2	98	124	3.000	0	0	57
3	67	287	6.500	0	0	13
4	114	261	10.500	0	0	124
5	120	229	14.500	0	0	113
6	136	185	18.500	0	0	22
7	76	155	22.500	0	0	79
8	105	125	26.500	0	0	177
9	21	273	31.000	0	0	37
10	109	260	35.500	0	0	67
11	163	255	40.000	0	0	119
12	128	225	44.500	0	0	36
13	67	202	49.000	3	0	78
14	80	142	53.500	0	0	154
15	54	89	58.000	0	0	24
16	75	153	62.667	0	1	20
17	236	282	68.167	0	0	62

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
18	122	269	73.667	0	0	99
19	85	226	79.167	0	0	114
20	75	137	84.750	16	0	43
21	199	277	90.750	0	0	158
22	136	274	96.750	6	0	73
23	106	213	102.750	0	0	51
24	54	246	108.917	15	0	107
25	116	134	115.417	0	0	71
26	4	132	121.917	0	0	140
27	237	270	128.917	0	0	96
28	173	254	135.917	0	0	29
29	173	181	142.917	28	0	79
30	126	174	149.917	0	0	86
31	38	117	156.917	0	0	113
32	13	212	164.417	0	0	73
33	110	167	171.917	0	0	46
34	179	271	179.917	0	0	151
35	83	182	187.917	0	0	65
36	8	128	196.083	0	12	56
37	21	118	204.917	9	0	118
38	196	245	213.917	0	0	76
39	139	204	222.917	0	0	75
40	24	151	231.917	0	0	111
41	96	138	240.917	0	0	72
42	50	130	249.917	0	0	78
43	75	100	259.267	20	0	77
44	29	232	268.767	0	0	95
45	86	127	278.267	0	0	69
46	110	141	288.100	33	0	91
47	221	265	298.100	0	0	70
48	143	206	308.100	0	0	76
49	81	161	318.100	0	0	97
50	71	119	328.100	0	0	123
51	68	106	338.100	0	23	83
52	78	102	348.100	0	0	208

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
228	12	20	7945.596	225	216	236
229	8	29	8125.920	207	186	238
230	34	48	8311.777	198	214	233
231	79	101	8506.416	210	221	240
232	3	39	8715.892	205	222	243
233	34	35	8936.890	230	199	244
234	22	25	9183.368	223	193	237
235	7	13	9431.606	212	218	238
236	12	42	9683.927	228	213	241
237	11	22	9948.727	217	234	243
238	7	8	10243.427	235	229	244
239	2	15	10563.521	224	226	242
240	30	79	10885.426	203	231	242
241	4	12	11346.200	227	236	245
242	2	30	11884.990	239	240	245
243	3	11	12493.119	232	237	246
244	7	34	13224.938	238	233	247
245	2	4	14362.007	242	241	246
246	2	3	16222.706	245	243	247
247	2	7	20719.073	246	244	0

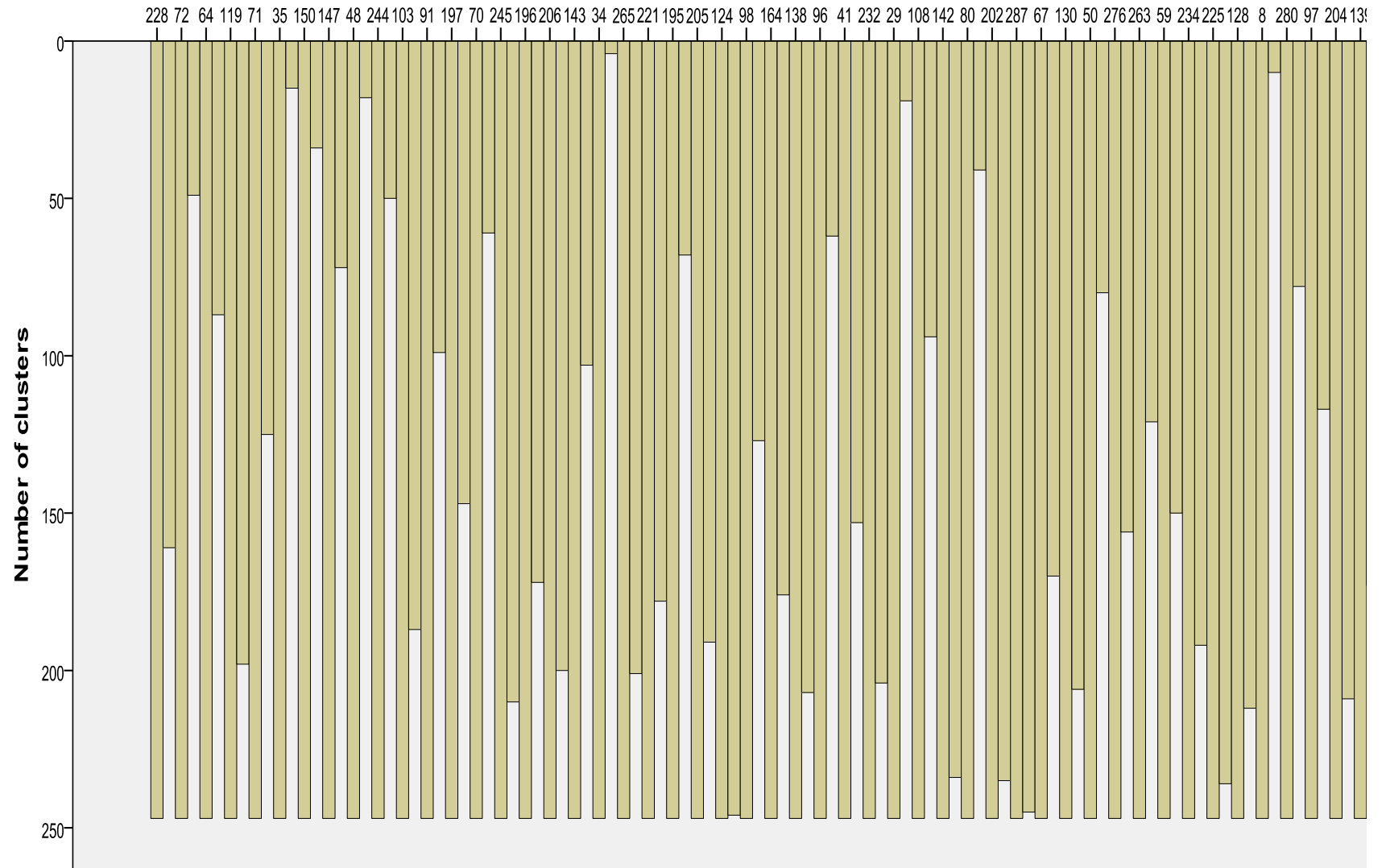
Cluster Membership

Case	6 Clusters	5 Clusters	4 Clusters	3 Clusters
2	1	1	1	1
3	2	2	2	2
4	3	3	3	1
6	3	3	3	1
7	4	4	4	3
8	4	4	4	3
9	4	4	4	3
11	5	2	2	2
12	3	3	3	1
13	4	4	4	3
14	1	1	1	1
15	1	1	1	1
16	1	1	1	1
17	1	1	1	1
18	1	1	1	1
20	3	3	3	1
21	1	1	1	1
22	5	2	2	2
23	2	2	2	2
24	4	4	4	3
25	5	2	2	2
27	1	1	1	1
29	4	4	4	3
30	1	1	1	1
31	1	1	1	1
32	3	3	3	1

Cluster Membership

Case	6 Clusters	5 Clusters	4 Clusters	3 Clusters
34	6	5	4	3
35	6	5	4	3
36	5	2	2	2
38	1	1	1	1
39	2	2	2	2
41	4	4	4	3
42	3	3	3	1
44	5	2	2	2
45	4	4	4	3
46	1	1	1	1
48	6	5	4	3
49	3	3	3	1
50	4	4	4	3
51	2	2	2	2
52	3	3	3	1
53	5	2	2	2
54	1	1	1	1
56	1	1	1	1
57	1	1	1	1
58	5	2	2	2
59	4	4	4	3
60	3	3	3	1
61	3	3	3	1
64	6	5	4	3
65	1	1	1	1
66	3	3	3	1
67	4	4	4	3
68	4	4	4	3
69	4	4	4	3
70	6	5	4	3
71	6	5	4	3
72	6	5	4	3
73	1	1	1	1
74	1	1	1	1
75	1	1	1	1
76	1	1	1	1
77	4	4	4	3
78	2	2	2	2
79	1	1	1	1
80	4	4	4	3
81	5	2	2	2
82	2	2	2	2
83	4	4	4	3
85	4	4	4	3
86	3	3	3	1
87	5	2	2	2
88	1	1	1	1
89	1	1	1	1
90	5	2	2	2
91	6	5	4	3
92	4	4	4	3
93	4	4	4	3

Cases



Note: This Icicle Plot has been cut down for to save room. This is an example and you do not need to have this in your work for the five stats assignment.

GET

```
FILE='D:\Comm 631\Humor Data Set\Humor Data Set.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
FREQUENCIES VARIABLES=CLU6_1 CLU5_1 CLU4_1 CLU3_1
/ORDER=ANALYSIS.
```

Frequencies

[DataSet1] D:\Comm 631\Humor Data Set\Humor Data Set.sav

Statistics

		Ward Method--6 Clusters	Ward Method--5 Clusters	Ward Method--4 Clusters	Ward Method--3 Clusters
N	Valid	248	248	248	248
	Missing	40	40	40	40

Frequency Table

Ward Method--6 Clusters

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	69	24.0	27.8	27.8
	2	22	7.6	8.9	36.7
	3	48	16.7	19.4	56.0
	4	66	22.9	26.6	82.7
	5	24	8.3	9.7	92.3
	6	19	6.6	7.7	100.0
Total		248	86.1	100.0	
Missing	System	40	13.9		
Total		288	100.0		

Ward Method--5 Clusters

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	69	24.0	27.8	27.8
	2	46	16.0	18.5	46.4
	3	48	16.7	19.4	65.7
	4	66	22.9	26.6	92.3
	5	19	6.6	7.7	100.0
Total		248	86.1	100.0	
Missing	System	40	13.9		
Total		288	100.0		

Ward Method--4 Clusters

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	69	24.0	27.8	27.8
	2	46	16.0	18.5	46.4
	3	48	16.7	19.4	65.7
	4	85	29.5	34.3	100.0
	Total	248	86.1	100.0	
Missing	System	40	13.9		
Total		288	100.0		

Ward Method--3 Clusters

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	117	40.6	47.2	47.2
	2	46	16.0	18.5	65.7
	3	85	29.5	34.3	100.0
	Total	248	86.1	100.0	
Missing	System	40	13.9		
Total		288	100.0		

```

MEANS TABLES=C5 C6 C8 C20 C21 C22 C23 C41 C48 C50 C66 G1 G2 G4 G9 B1 B5 B9 B21
B22 BY CLU6_1
/CELLS MEAN COUNT STDDEV
/STATISTICS ANOVA.

```

Means

[DataSet1] E:\Comm 631\Humor Data Set\Humor Data Set.sav

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
C5. I like jokes that involve wordplay--puns, riddles, etc. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C6. I like sight gags. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C8. I enjoy slapstick. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C20. I like sick humor. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C21. I enjoy humor that criticizes society. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C22. I like sarcasm. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C23. I like humor that is naughty. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C41. I like dark comedy. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C48. I enjoy irony. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C50. I like gross-out humor. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
C66. I like "inside" jokes (jokes only certain people "get"). * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
G1. Are you male or female? * Ward Method--6 Clusters	216	75.0%	72	25.0%	288	100.0%
G2. What is your age? * Ward Method--6 Clusters	216	75.0%	72	25.0%	288	100.0%
G4. Which of the following categories best describes your political philosophy? * Ward Method--6 Clusters	189	65.6%	99	34.4%	288	100.0%

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
G9. What is your annual household income? * Ward Method--6 Clusters	216	75.0%	72	25.0%	288	100.0%
B1. Barack Obama is doing a good job as president. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
B5. I like finding unexpected linkages on the Internet. * Ward Method--6 Clusters	245	85.1%	43	14.9%	288	100.0%
B9. Marriages between same-sex couples should be recognized by the law as valid, with the same rights as traditional marriages. * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
B21. On this scale, how would you rate the Cleveland area? * Ward Method--6 Clusters	248	86.1%	40	13.9%	288	100.0%
B22. How would you rate the neighborhood you live in? * Ward Method--6 Clusters	246	85.4%	42	14.6%	288	100.0%

Report

Ward Method--6 Clusters		C5. I like jokes that involve wordplay--puns, riddles, etc.	C6. I like sight gags.	C8. I enjoy slapstick.	C20. I like sick humor.
1	Mean	6.17	5.51	5.04	4.84
	N	69	69	69	69
	Std. Deviation	2.307	1.763	2.212	1.812
2	Mean	4.50	2.77	2.68	1.86
	N	22	22	22	22
	Std. Deviation	2.464	1.901	2.338	1.726
3	Mean	7.21	6.29	4.88	1.52
	N	48	48	48	48
	Std. Deviation	2.535	2.518	3.036	1.924
4	Mean	8.03	7.30	6.82	7.68
	N	66	66	66	66
	Std. Deviation	1.655	1.655	2.398	1.915
5	Mean	4.33	3.50	2.08	1.71
	N	24	24	24	24
	Std. Deviation	2.777	2.449	2.430	2.312
6	Mean	3.00	4.05	4.11	8.89
	N	19	19	19	19
	Std. Deviation	2.427	2.223	3.035	1.286
Total	Mean	6.30	5.59	4.92	4.70
	N	248	248	248	248
	Std. Deviation	2.746	2.491	2.924	3.276

Report

Ward Method--6 Clusters		C21. I enjoy humor that criticizes society.	C22. I like sarcasm.	C23. I like humor that is naughty.	C41. I like dark comedy.
1	Mean	6.41	7.48	5.97	6.36
	N	69	69	69	69
	Std. Deviation	2.039	2.139	2.121	2.142
2	Mean	6.86	8.27	4.41	2.73
	N	22	22	22	22
	Std. Deviation	1.983	1.202	2.323	2.453
3	Mean	7.46	8.98	7.40	4.00
	N	48	48	48	48
	Std. Deviation	2.593	1.523	2.219	2.806
4	Mean	8.55	9.24	8.94	7.27
	N	66	66	66	66
	Std. Deviation	1.405	1.009	.892	2.324
5	Mean	1.17	4.17	3.63	2.29
	N	24	24	24	24
	Std. Deviation	1.523	3.239	2.039	2.216
6	Mean	6.42	9.16	8.74	7.00
	N	19	19	19	19
	Std. Deviation	2.854	1.119	1.558	2.925
Total	Mean	6.71	8.12	6.88	5.48
	N	248	248	248	248
	Std. Deviation	2.849	2.315	2.581	3.019

Report

Ward Method--6 Clusters		C48. I enjoy irony.	C50. I like gross-out humor.	C66. I like "inside" jokes (jokes only certain people "get").	G1. Are you male or female?
1	Mean	6.88	4.26	5.93	1.54
	N	69	69	69	54
	Std. Deviation	1.778	2.140	2.702	.503
2	Mean	5.64	1.55	5.95	1.50
	N	22	22	22	20
	Std. Deviation	2.460	1.738	2.627	.513
3	Mean	8.38	1.69	8.69	1.59
	N	48	48	48	44
	Std. Deviation	1.721	2.308	1.417	.497
4	Mean	8.29	7.62	8.08	1.48
	N	66	66	66	60
	Std. Deviation	1.699	1.944	1.809	.504
5	Mean	4.79	2.50	6.50	1.70
	N	24	24	24	23
	Std. Deviation	3.107	2.322	2.226	.470
6	Mean	8.26	7.47	7.58	1.60
	N	19	19	19	15
	Std. Deviation	1.821	1.806	2.479	.507
Total	Mean	7.34	4.49	7.22	1.55
	N	248	248	248	216
	Std. Deviation	2.304	3.204	2.456	.499

Report

Ward Method--6 Clusters		G2. What is your age?	G4. Which of the following categories best describes your political philosophy?	G9. What is your annual household income?	B1. Barack Obama is doing a good job as president.
1	Mean	22.61	3.43	2.72	5.22
	N	54	47	54	69
	Std. Deviation	7.487	1.098	1.559	2.344
2	Mean	21.85	3.44	2.60	5.82
	N	20	18	20	22
	Std. Deviation	5.905	1.097	1.465	2.108
3	Mean	22.34	3.78	3.09	6.15
	N	44	37	44	48
	Std. Deviation	5.918	.947	1.626	2.775
4	Mean	22.05	3.46	2.37	6.15
	N	60	52	60	66
	Std. Deviation	3.652	1.146	1.438	2.662
5	Mean	24.26	2.86	2.43	6.13
	N	23	21	23	24
	Std. Deviation	6.635	1.108	1.619	3.591
6	Mean	21.20	3.07	2.93	3.74
	N	15	14	15	19
	Std. Deviation	4.092	1.328	1.668	3.052
Total	Mean	22.41	3.42	2.67	5.67
	N	216	189	216	248
	Std. Deviation	5.805	1.120	1.552	2.752

Report

Ward Method--6 Clusters		B5. I like finding unexpected linkages on the Internet.	B9. Marriages between same-sex couples should be recognized by the law as valid, with the same rights as traditional marriages.	B21. On this scale, how would you rate the Cleveland area?	B22. How would you rate the neighborhood you live in?
1	Mean	5.29	6.62	4.87	6.24
	N	69	69	69	68
	Std. Deviation	2.921	3.671	2.057	2.045
2	Mean	4.45	5.77	4.95	6.45
	N	22	22	22	22
	Std. Deviation	2.874	3.878	2.278	1.945
3	Mean	5.57	6.83	4.17	6.27
	N	46	48	48	48
	Std. Deviation	3.324	3.622	2.215	2.229
4	Mean	6.69	7.79	5.23	7.03
	N	65	66	66	66
	Std. Deviation	2.952	3.126	2.096	1.977
5	Mean	4.38	4.42	4.00	5.65
	N	24	24	24	23
	Std. Deviation	3.187	3.798	1.719	2.424
6	Mean	6.05	6.47	4.53	5.95
	N	19	19	19	19
	Std. Deviation	3.153	4.033	1.982	1.747
Total	Mean	5.61	6.67	4.73	6.40
	N	245	248	248	246
	Std. Deviation	3.119	3.673	2.110	2.097

ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
C5. I like jokes that involve wordplay--puns, riddles, etc. * Ward Method--6 Clusters	Between Groups	609.317	5	121.863	23.544	.000
	Within Groups	1.253E3	242	5.176		
	Total	1.862E3	247			
C6. I like sight gags. * Ward Method--6 Clusters	Between Groups	542.135	5	108.427	26.507	.000
	Within Groups	989.913	242	4.091		
	Total	1.532E3	247			
C8. I enjoy slapstick. * Ward Method--6 Clusters	Between Groups	554.888	5	110.978	17.256	.000
	Within Groups	1.556E3	242	6.431		
	Total	2.111E3	247			
C20. I like sick humor. * Ward Method--6 Clusters	Between Groups	1.799E3	5	359.887	102.356	.000
	Within Groups	850.882	242	3.516		
	Total	2.650E3	247			
C21. I enjoy humor that criticizes society. * Ward Method--6 Clusters	Between Groups	995.200	5	199.040	47.716	.000
	Within Groups	1.009E3	242	4.171		
	Total	2.005E3	247			
C22. I like sarcasm. * Ward Method--6 Clusters	Between Groups	543.068	5	108.614	33.675	.000
	Within Groups	780.541	242	3.225		
	Total	1.324E3	247			
C23. I like humor that is naughty. * Ward Method--6 Clusters	Between Groups	803.803	5	160.761	46.215	.000
	Within Groups	841.806	242	3.479		
	Total	1.646E3	247			
C41. I like dark comedy. * Ward Method--6 Clusters	Between Groups	825.544	5	165.109	28.013	.000
	Within Groups	1.426E3	242	5.894		
	Total	2.252E3	247			
C48. I enjoy irony. * Ward Method--6 Clusters	Between Groups	360.962	5	72.192	18.379	.000
	Within Groups	950.586	242	3.928		
	Total	1.312E3	247			

ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
C50. I like gross-out humor. * Ward Method--6 Clusters	Between Groups	1.483E3	5	296.529	68.126	.000
	Within Groups	1.053E3	242	4.353		
	Total	2.536E3	247			
C66. I like "inside" jokes (jokes only certain people "get"). * Ward Method--6 Clusters	Between Groups	317.084	5	63.417	13.082	.000
	Within Groups	1.173E3	242	4.848		
	Total	1.490E3	247			
G1. Are you male or female? * Ward Method--6 Clusters	Between Groups	.925	5	.185	.739	.595
	Within Groups	52.515	210	.250		
	Total	53.440	215			
G2. What is your age? * Ward Method--6 Clusters	Between Groups	117.194	5	23.439	.690	.631
	Within Groups	7.129E3	210	33.947		
	Total	7.246E3	215			
G4. Which of the following categories best describes your political philosophy? * Ward Method--6 Clusters	Between Groups	13.352	5	2.670	2.195	.057
	Within Groups	222.627	183	1.217		
	Total	235.979	188			
G9. What is your annual household income? * Ward Method--6 Clusters	Between Groups	15.873	5	3.175	1.329	.253
	Within Groups	501.789	210	2.389		
	Total	517.662	215			
B1. Barack Obama is doing a good job as president. * Ward Method--6 Clusters	Between Groups	116.759	5	23.352	3.222	.008
	Within Groups	1.754E3	242	7.247		
	Total	1.871E3	247			
B5. I like finding unexpected linkages on the Internet. * Ward Method--6 Clusters	Between Groups	153.003	5	30.601	3.292	.007
	Within Groups	2.221E3	239	9.294		
	Total	2.374E3	244			
B9. Marriages between same-sex couples should be recognized by the law as valid, with the same rights as traditional marriages. * Ward Method--6 Clusters	Between Groups	224.211	5	44.842	3.491	.005
	Within Groups	3.108E3	242	12.844		
	Total	3.333E3	247			

ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
B21. On this scale, how would you rate the Cleveland area? * Ward Method--6 Clusters	Between Groups	47.580	5	9.516	2.190	.056
	Within Groups	1.052E3	242	4.346		
	Total	1.099E3	247			
B22. How would you rate the neighborhood you live in? * Ward Method--6 Clusters	Between Groups	45.686	5	9.137	2.126	.063
	Within Groups	1.031E3	240	4.297		
	Total	1.077E3	245			

Measures of Association

	Eta	Eta Squared
C5. I like jokes that involve wordplay--puns, riddles, etc. * Ward Method--6 Clusters	.572	.327
C6. I like sight gags. * Ward Method--6 Clusters	.595	.354
C8. I enjoy slapstick. * Ward Method--6 Clusters	.513	.263
C20. I like sick humor. * Ward Method--6 Clusters	.824	.679
C21. I enjoy humor that criticizes society. * Ward Method--6 Clusters	.705	.496
C22. I like sarcasm. * Ward Method--6 Clusters	.641	.410
C23. I like humor that is naughty. * Ward Method--6 Clusters	.699	.488
C41. I like dark comedy. * Ward Method--6 Clusters	.605	.367
C48. I enjoy irony. * Ward Method--6 Clusters	.525	.275
C50. I like gross-out humor. * Ward Method--6 Clusters	.765	.585
C66. I like "inside" jokes (jokes only certain people "get"). * Ward Method--6 Clusters	.461	.213
G1. Are you male or female? * Ward Method--6 Clusters	.132	.017
G2. What is your age? * Ward Method--6 Clusters	.127	.016

Measures of Association

	Eta	Eta Squared
G4. Which of the following categories best describes your political philosophy? * Ward Method--6 Clusters	.238	.057
G9. What is your annual household income? * Ward Method--6 Clusters	.175	.031
B1. Barack Obama is doing a good job as president. * Ward Method--6 Clusters	.250	.062
B5. I like finding unexpected linkages on the Internet. * Ward Method--6 Clusters	.254	.064
B9. Marriages between same-sex couples should be recognized by the law as valid, with the same rights as traditional marriages. * Ward Method--6 Clusters	.259	.067
B21. On this scale, how would you rate the Cleveland area? * Ward Method--6 Clusters	.208	.043
B22. How would you rate the neighborhood you live in? * Ward Method--6 Clusters	.206	.042

CLUSTER ANALYSIS TABLE:

Cluster Name:	<i>British Humor</i>	Sarcastic Humor	???	High Humor	No Humor	Nasty Humor	Total	F	Sig
Variable:	1 (69)	2 (22)	3 (48)	4 (66)	5 (24)	6 (19)	248		
Internal Variables:							Grand Mean		
C5: Jokes that involve wordplay	6.17	4.50	7.21	8.03	4.33	3.00	6.30	23.544	<.001
C6: Sight gags	5.51	2.77	6.29	7.30	3.50	4.05	5.59	26.507	<.001
C8: Slapstick Comedy.	5.04	2.68	4.88	6.82	2.08	4.11	4.92	17.256	<.001
C20: Sick humor	4.84	1.86	1.52	7.68	1.71	8.89	4.70	102.356	<.001
C21: Humor that criticizes society	6.41	6.86	7.46	8.55	1.17	6.42	6.71	47.716	<.001
C22: Sarcasm.	7.48	8.27	8.98	9.24	4.17	9.16	8.12	33.675	<.001
C23: Naughty Humor	5.97	4.41	7.40	8.94	3.63	8.74	6.88	46.215	<.001
C41: Dark Comedy.	6.36	2.73	4.00	7.27	2.29	7.00	5.48	28.013	<.001
C48: Irony.	6.88	5.64	8.38	8.29	4.79	8.26	7.34	18.379	<.001
C50: Gross-out Humor	4.26	1.55	1.69	7.62	2.50	7.47	4.49	68.126	<.001
C66: "Inside" Jokes	5.93	5.95	8.69	8.08	6.50	7.58	7.22	13.082	<.001
External Variables:									
G1: Are you male or female? (represented in % of females)	1.54 (54%)	1.50 (50%)	1.59 (59%)	1.48 (48%)	1.70 (70%)	1.60 (60%)	1.55 (55%)	.739	.595
G2: age? (in years)	22.61	21.85	22.34	22.05	24.26	21.20	22.41	.690	.631
G4: political philosophy	3.43	3.44	3.78	3.46	2.86	3.07	3.42	2.195	.057
G9: Annual household income?	2.72	2.60	3.09	2.37	2.43	2.93	2.67	1.329	.253
B1: Barack Obama is doing a good job as president.	5.22	5.82	6.15	6.15	6.13	3.74	5.67	3.222	.008
B5: find unexpected linkages funny	5.29	4.45	5.57	6.69	4.38	6.05	5.61	3.292	.007
B9: Marriages between same-sex couples...	6.62	5.77	6.83	7.79	4.42	6.47	6.67	3.491	.005
B21. How would you rate the Cleveland area?	4.87	4.95	4.17	5.23	4.00	4.53	4.73	2.190	.056
B22. Rate you neighborhood?	6.24	6.45	6.27	7.03	5.65	5.95	6.40	2.126	.063

Bold: indicates a score above the mean.

Cluster Analysis Writeup

Eleven variables were selected from Neuendorf and Skalski's humor scale data set to be selected as the internal (independent, or cluster) variables. These items were all done with similar interval/ratio measures, allowing for the use of the unstandardized raw data. Variables are selected but it is the actual objects/cases (participants) who are clustered (in total, $n = 248$).

C5: I like jokes that involve wordplay--puns, riddles, etc.

C6: I like sight gags.

C8: I enjoy Slapstick Comedy.

C20: I like sick humor.

C21: I enjoy humor that criticizes society.

C22: I like sarcasm.

C23: I like humor that is naughty.

C41: I like dark comedy.

C48: I enjoy irony.

C50: I like gross-out humor.

C66: I like "inside" jokes

From this, a hierarchical cluster analysis was run (allowing for 3 to 6 clusters), from which 6 clusters were created from the data. This was done using Ward's method (agglomerative approach using squared Euclidean distances of objects), selected for being one of the most common as well as detailed approaches to cluster analysis. All of the variables selected resulted in a significance of $<.001$. Ward's analysis intends to produce clusters of objects/cases (truly meaning participants) that will all be near in size to each other, but this analysis results in three high-membership clusters (69, 48, and 66) and three low-membership clusters (19, 22, 24), although this generally is not an issue of concern. Dendrogram and icicle plots were also run, producing a visual breakdown of individual objects (respondents) that make up the clusters.

Finally, an ANOVA was run to test significance between the created clusters and the external variables of interest. ANOVA also produced the means, number of cases, and standard deviations for both the internal and external variables. By saving the clusters created, we were able to run this ANOVA many times over and see which interested variables (demographics and public opinion questions) proved to be significant those which did not. The final list of external variables tested for significance is as follows:

G1: Are you male or female?

G2: What is your age?

G4: Which of the following categories best describes your political philosophy?

G9: What is your annual household income?

B1: Barack Obama is doing a good job as president.

B5: I like finding unexpected linkages on the Internet.

B9: Marriages between same-sex couples should be recognized by the law as valid, with the same rights as traditional marriages.

B21. On this scale, how would you rate the Cleveland area?

B22. How would you rate the neighborhood you live in?

The survey this data set was based from was administered to college students, which is why most of the demographic questions proved to be non-significant (e.g., age, as each cluster has roughly the same mean age). Political philosophy, however, which one can look at as more of an opinion vs. demographic question, proved to be near significant ($p = <.057$) with a mean score of 2.195 (lower scores being conservative and higher being liberal). Public opinion questions all proved to be of significance, with the exception being of the community quality of life questions, B21 and B22, which did show near significance ($p = <.056$ & $p = <.063$, respectively).

The mean scores produced in the ANOVA for each cluster allow for an understanding of specific details for each cluster (versus the overall mean). Even more so, when compared specifically with how the clusters were created the internal humor variables, one can name each cluster accordingly, which we did. For instance, the fourth cluster of people was named “High Humor” because all of their responses are scored high, on average, across all of the humor variables. And, the cluster means for external variables may be examined to further describe the clusters. In doing so we can see that, for instance, the “No Humor” cluster has the highest average age (24.26) and that the “High Humor” cluster (besides just really enjoying funny things) thinks very highly of Cleveland and the neighborhood they live in.