

A METHOD OF DATA MINING TO VISUALIZE MULTINOMIAL LOGISTIC REGRESSION MODELS

Kelsey Gonzalez
School of Sociology
University of Arizona

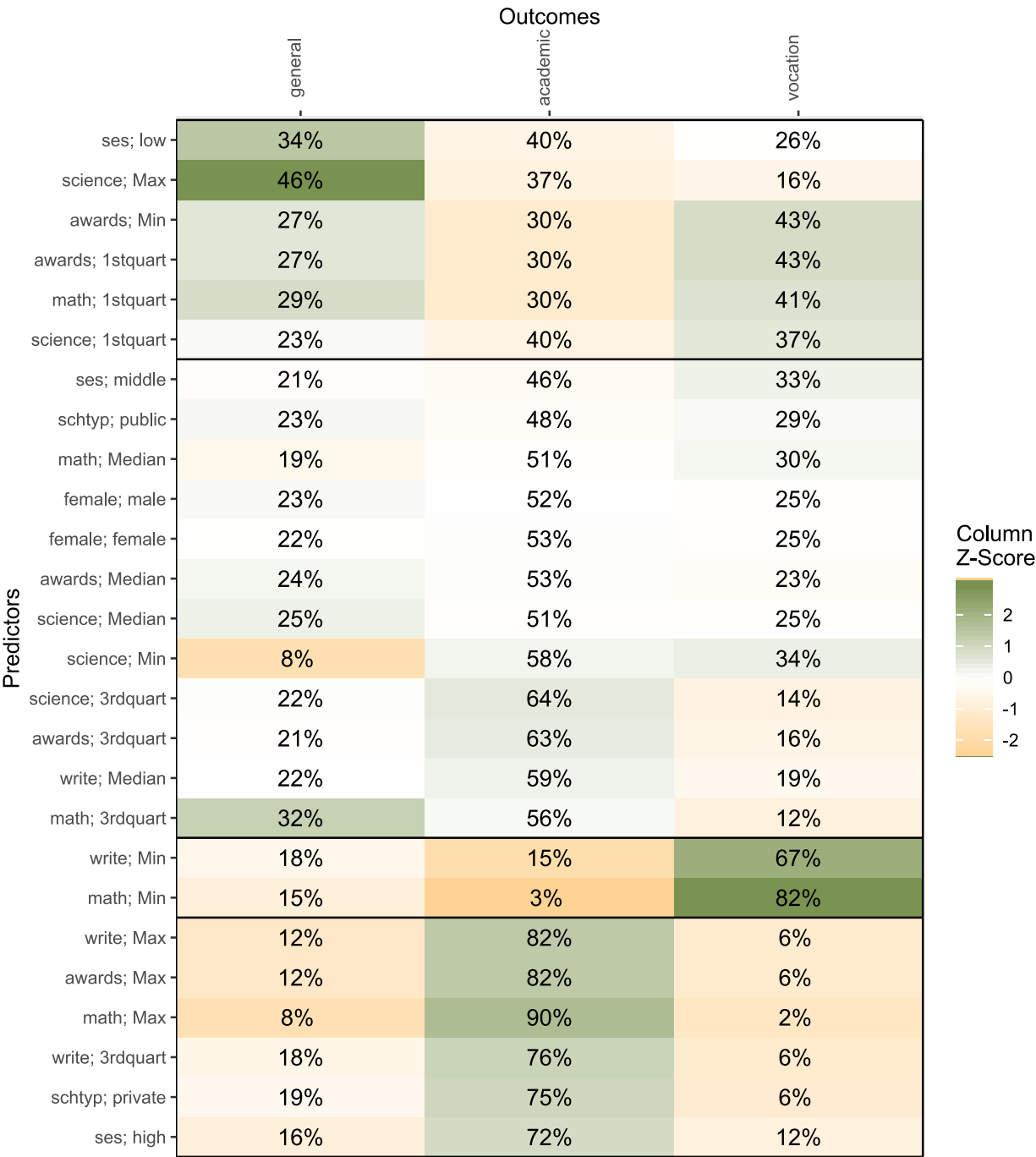
December 06, 2019

THE PROBLEM

Multinomial Logistic Regression Model predicting which type of high school is chosen

Predictors	Odds Ratios of Academic v General	Odds Ratios of Vocational v General
(Intercept)	0.01 *	79.84
ses [middle]	1.41	3.22 *
ses [high]	2.91	1.97
write	1.07	0.97
math	1.13 ***	0.97
science	0.92 **	0.96
female [female]	0.78	1.24
schtyp [private]	1.64	0.26
awards	0.89	1.08
Observations	200	
R ² Nagelkerke	0.398	
* p<0.05 ** p<0.01 *** p<0.001		

Heatmap of Predicted Probabilities



MY SOLUTION

THE PROCESS

- Mine average predicted probabilities
- Cluster & sort predicted probabilities
- Create heatmap of scaled predicted probabilities
- Overlay heatmap with original predicted probabilities

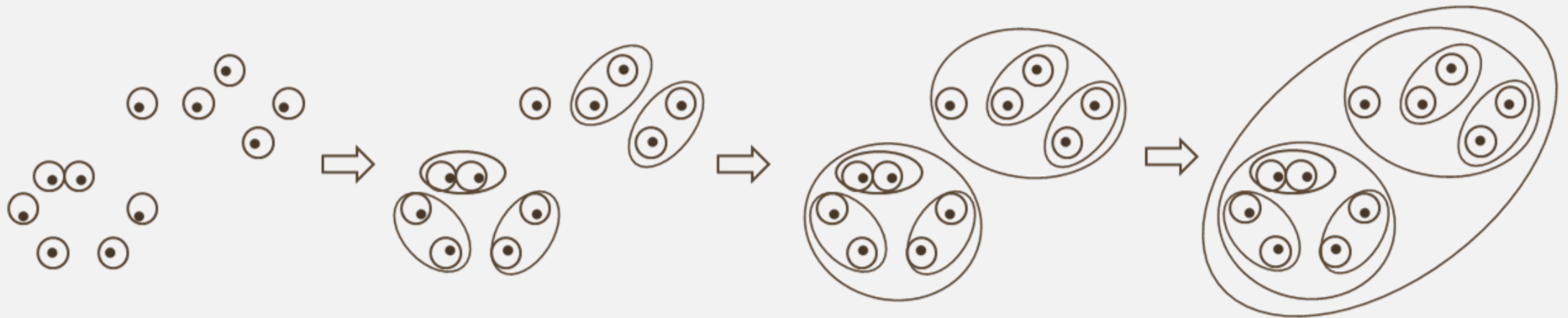
RUN MODEL

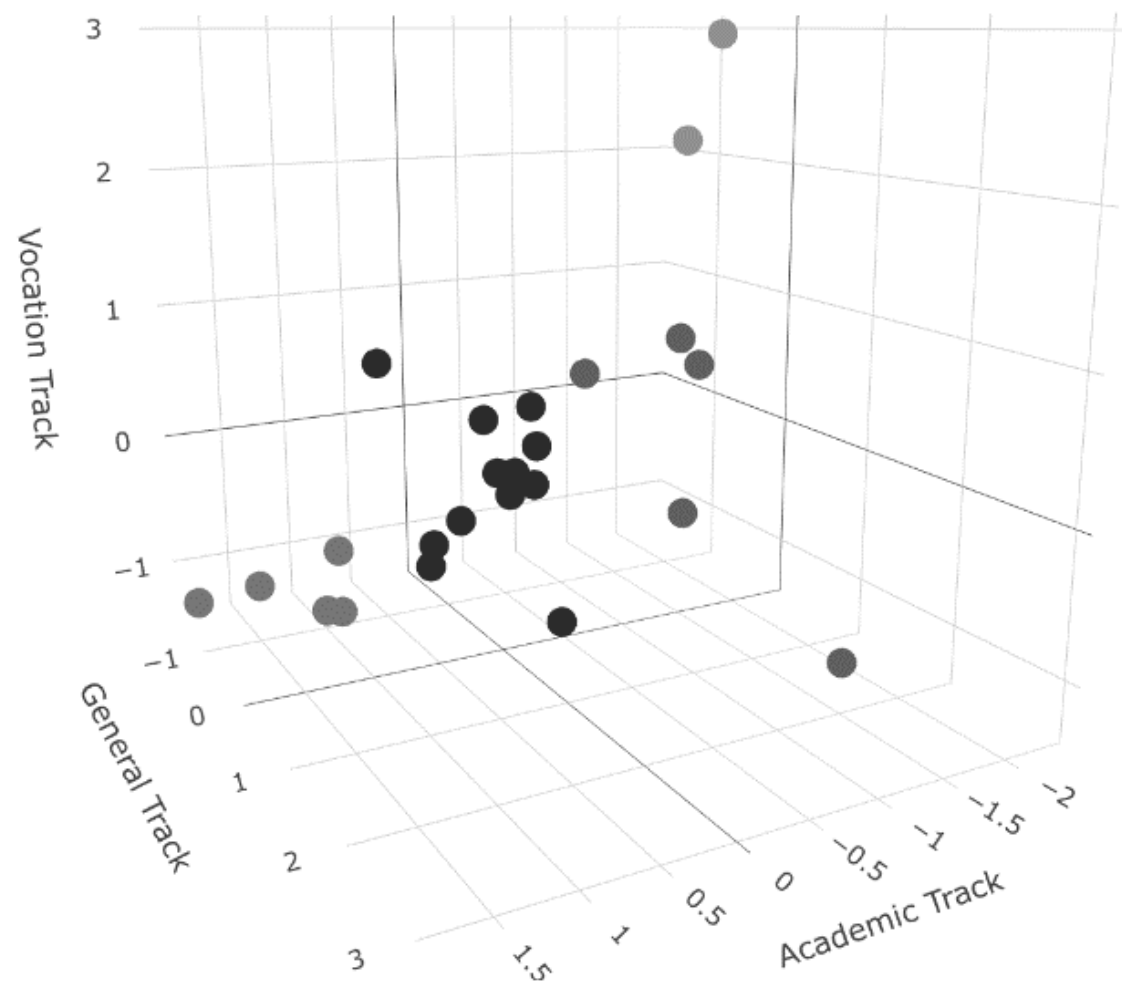
```
# run multinomial logit model  
model = multinom(prog ~ ses + write + math + science + female + schtyp + awards, data = data)
```

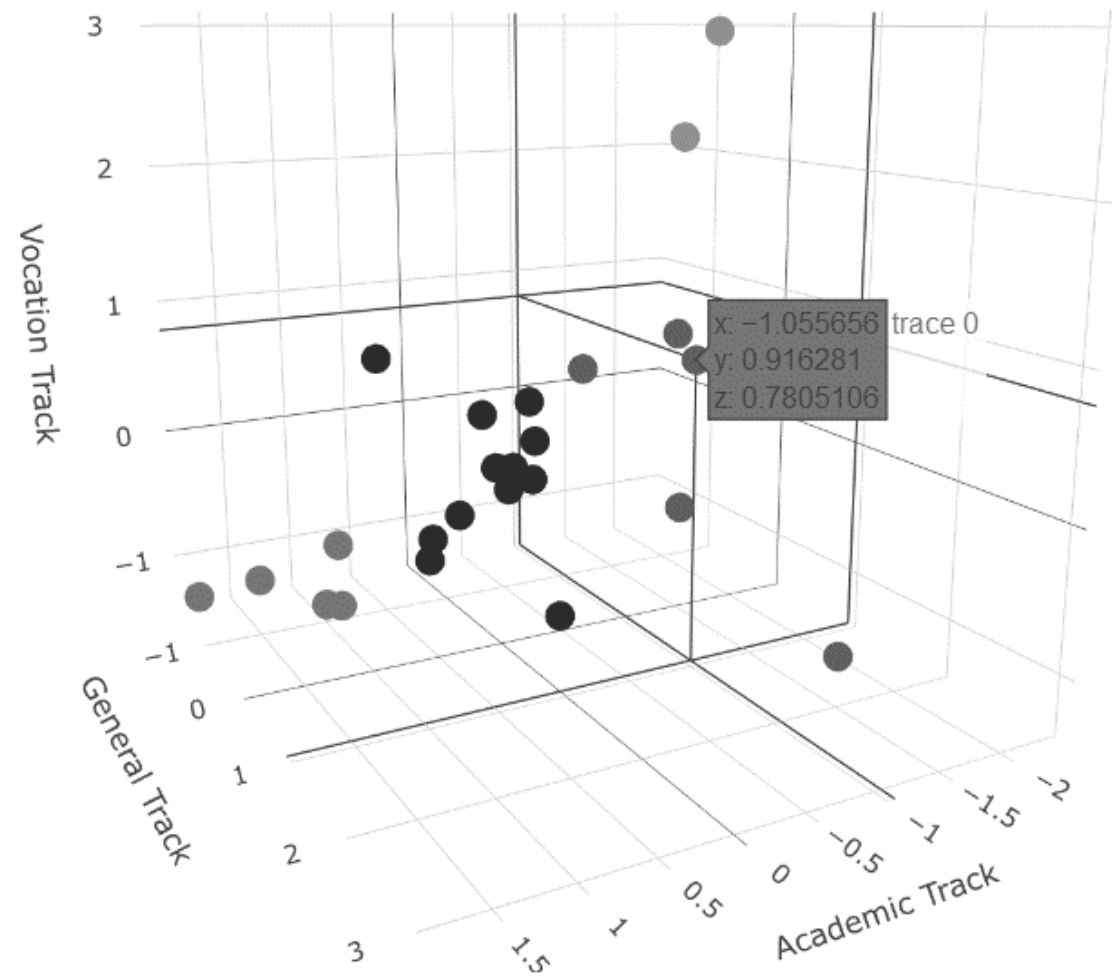
EXTRACT PREDICTED PROBABILITIES FOR ALL OUTCOMES

Variable	General	Academic	Vocation
awards; 3rdquart	21%	63%	16%
awards; 1stquart	27%	30%	43%
awards; Max	12%	82%	6%
awards; Median	24%	53%	23%
awards; Min	27%	30%	43%
schtyp; private	19%	75%	6%
schtyp; public	23%	48%	29%
female; female	22%	53%	25%
female; male	23%	52%	25%
science; 3rdquart	22%	64%	14%
science; 1stquart	23%	40%	37%
science; Max	46%	37%	16%
science; Median	25%	51%	25%
science; Min	8%	58%	34%
math; 3rdquart	32%	56%	12%
math; 1stquart	29%	30%	41%
math; Max	8%	90%	2%
math; Median	19%	51%	30%
math; Min	15%	3%	82%
write; 3rdquart	18%	76%	6%
write; Max	12%	82%	6%
write; Median	22%	59%	19%
write; Min	18%	15%	67%
ses; high	16%	72%	12%
ses; middle	21%	46%	33%
ses; low	34%	40%	26%

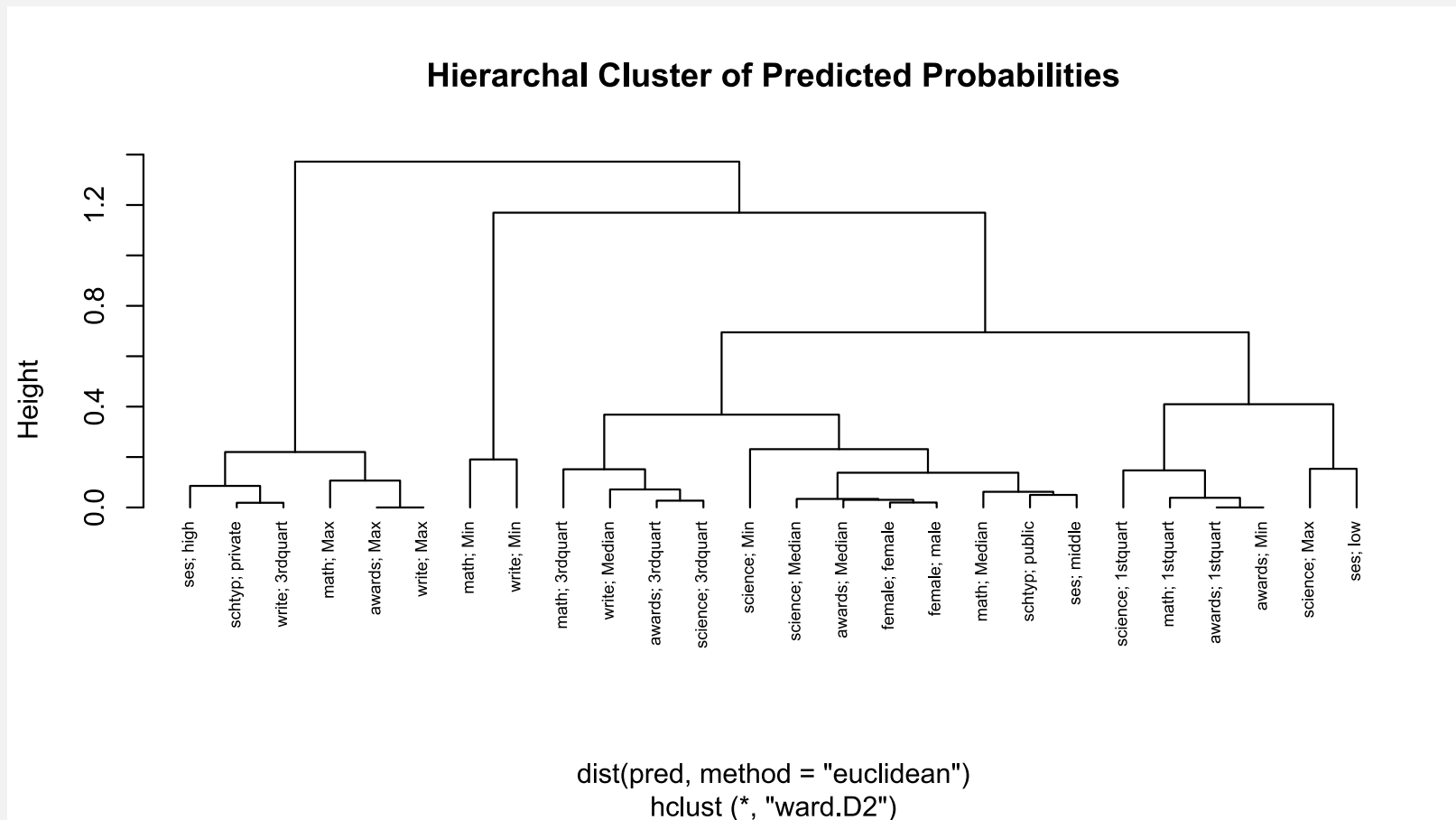
AGGLOMERATIVE HIERARCHICAL CLUSTERING



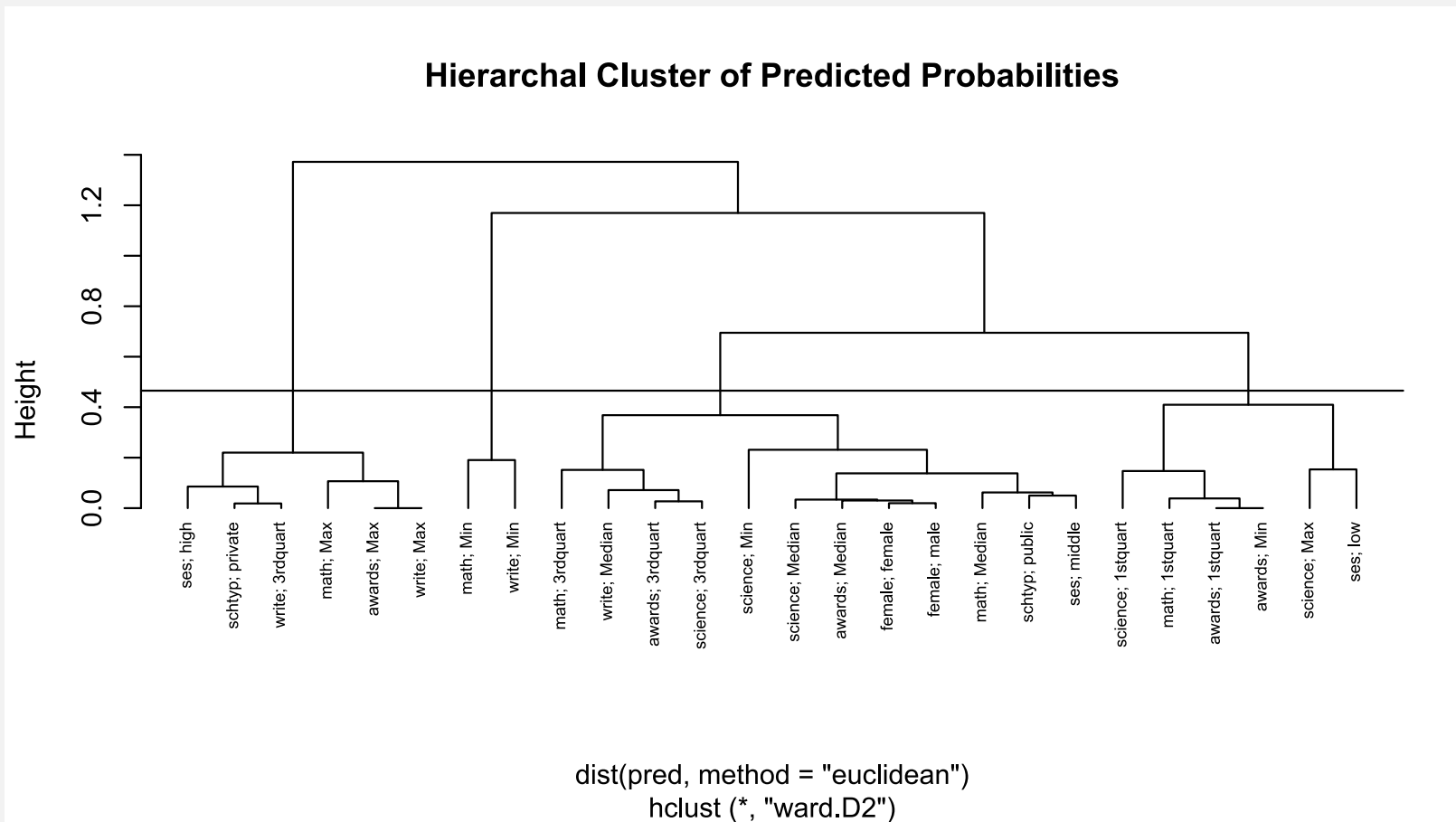




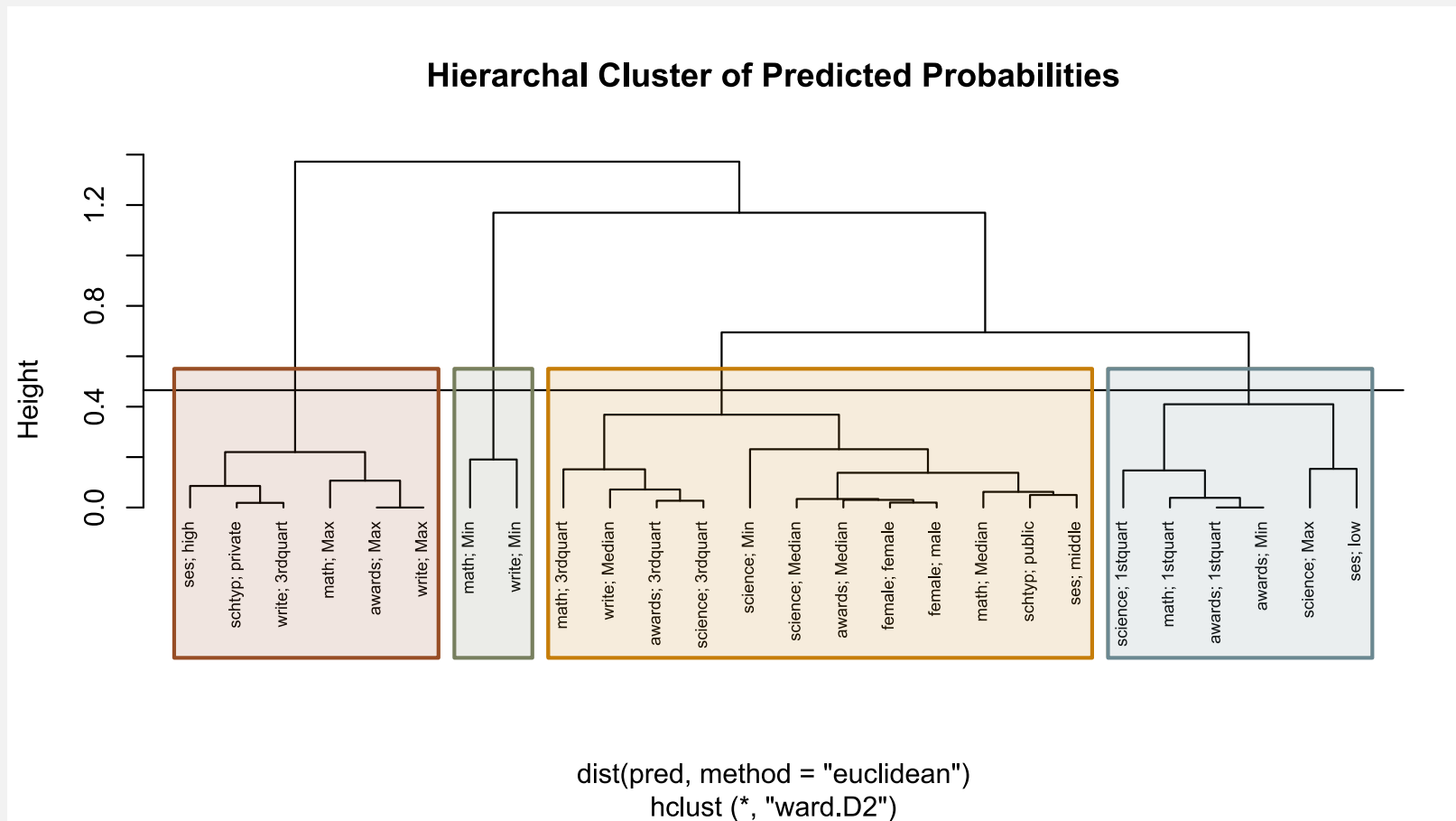
SORT USING HIERARCHICAL CLUSTERING ON DISTANCES

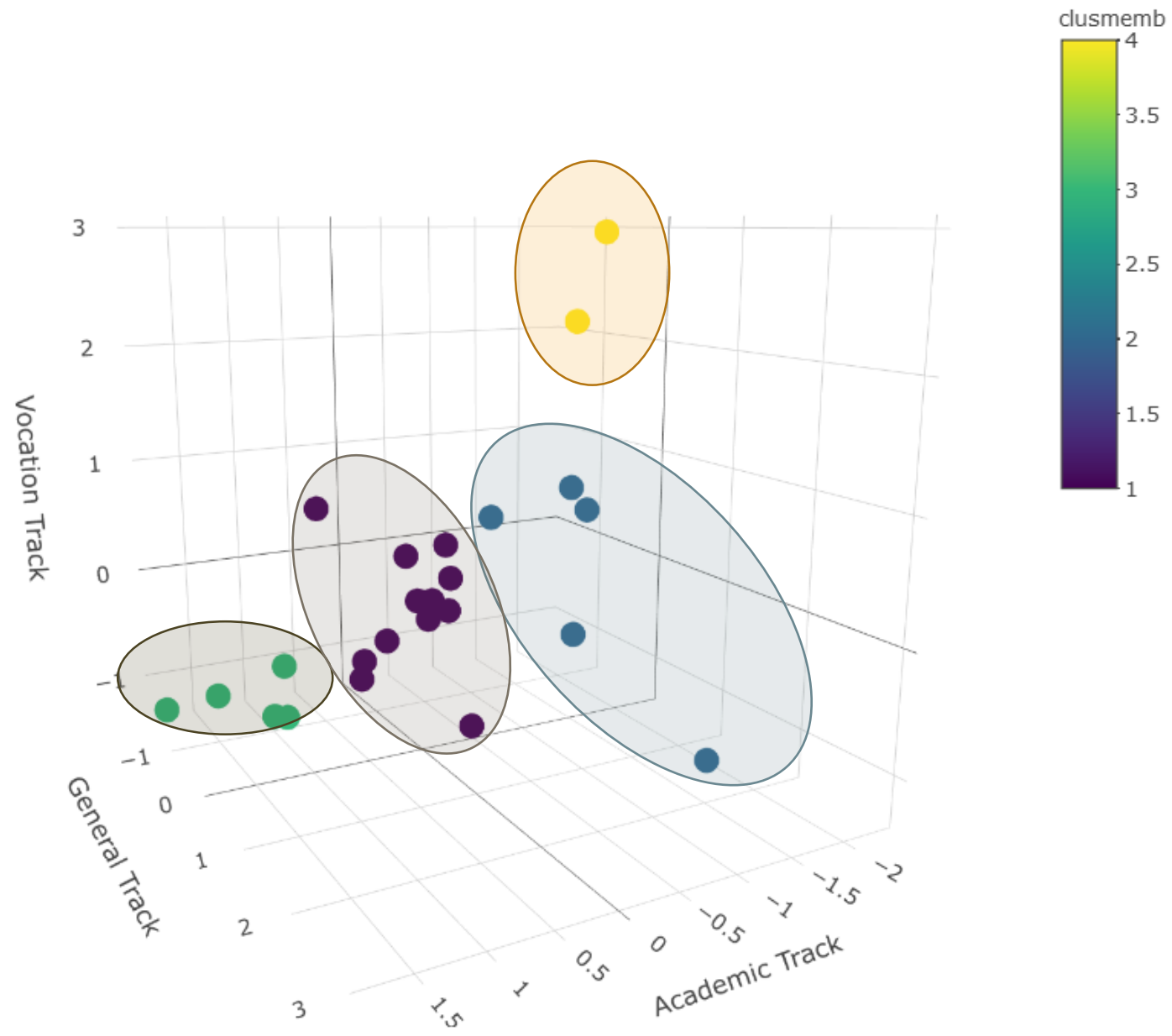


CUT THE TREE AT 75TH PERCENTILE OF EUCLIDEAN DISTANCE MATRIX



CREATING 3 CLUSTERS





RETURN TO PREDICTED PROBABILITIES

Variable	General	Academic	Vocation
awards; 3rdquart	21%	63%	16%
awards; 1stquart	27%	30%	43%
awards; Max	12%	82%	6%
awards; Median	24%	53%	23%
awards; Min	27%	30%	43%
schtyp; private	19%	75%	6%
schtyp; public	23%	48%	29%
female; female	22%	53%	25%
female; male	23%	52%	25%
science; 3rdquart	22%	64%	14%
science; 1stquart	23%	40%	37%
science; Max	46%	37%	16%
science; Median	25%	51%	25%
science; Min	8%	58%	34%
math; 3rdquart	32%	56%	12%
math; 1stquart	29%	30%	41%
math; Max	8%	90%	2%
math; Median	19%	51%	30%
math; Min	15%	3%	82%
write; 3rdquart	18%	76%	6%
write; Max	12%	82%	6%
write; Median	22%	59%	19%
write; Min	18%	15%	67%
ses; high	16%	72%	12%
ses; middle	21%	46%	33%
ses; low	34%	40%	26%

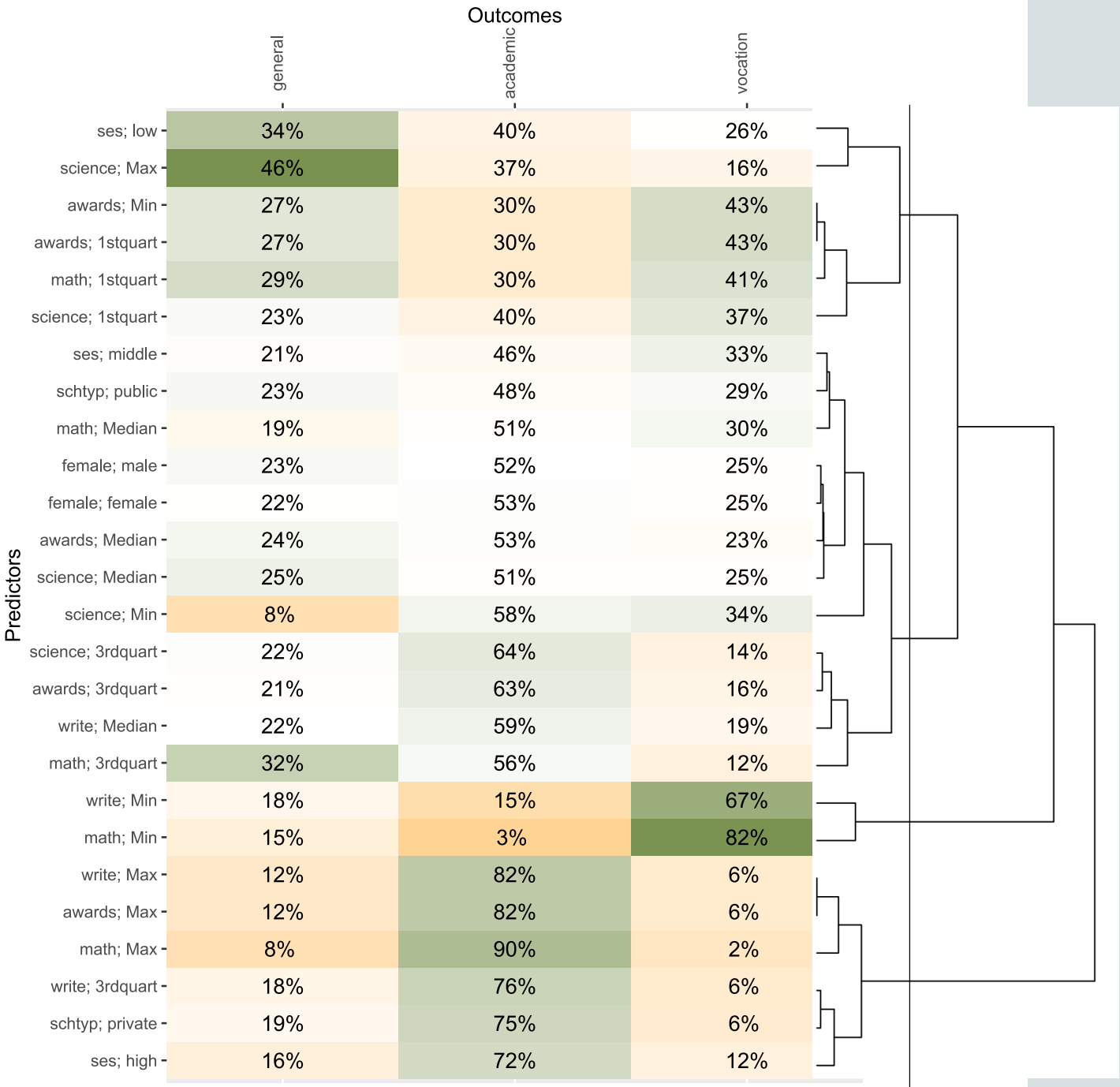
SORT PREDICTED PROBABILITIES IN HIERARCHICAL ORDER

Variable	General	Academic	Vocation
ses; high	16%	72%	12%
schtyp; private	19%	75%	6%
write; 3rdquart	18%	76%	6%
math; Max	8%	90%	2%
awards; Max	12%	82%	6%
write; Max	12%	82%	6%
math; Min	15%	3%	82%
write; Min	18%	15%	67%
math; 3rdquart	32%	56%	12%
write; Median	22%	59%	19%
awards; 3rdquart	21%	63%	16%
science; 3rdquart	22%	64%	14%
science; Min	8%	58%	34%
science; Median	25%	51%	25%
awards; Median	24%	53%	23%
female; female	22%	53%	25%
female; male	23%	52%	25%
math; Median	19%	51%	30%
schtyp; public	23%	48%	29%
ses; middle	21%	46%	33%
science; 1stquart	23%	40%	37%
math; 1stquart	29%	30%	41%
awards; 1stquart	27%	30%	43%
awards; Min	27%	30%	43%
science; Max	46%	37%	16%
ses; low	34%	40%	26%

CONVERT TO Z-SCORES

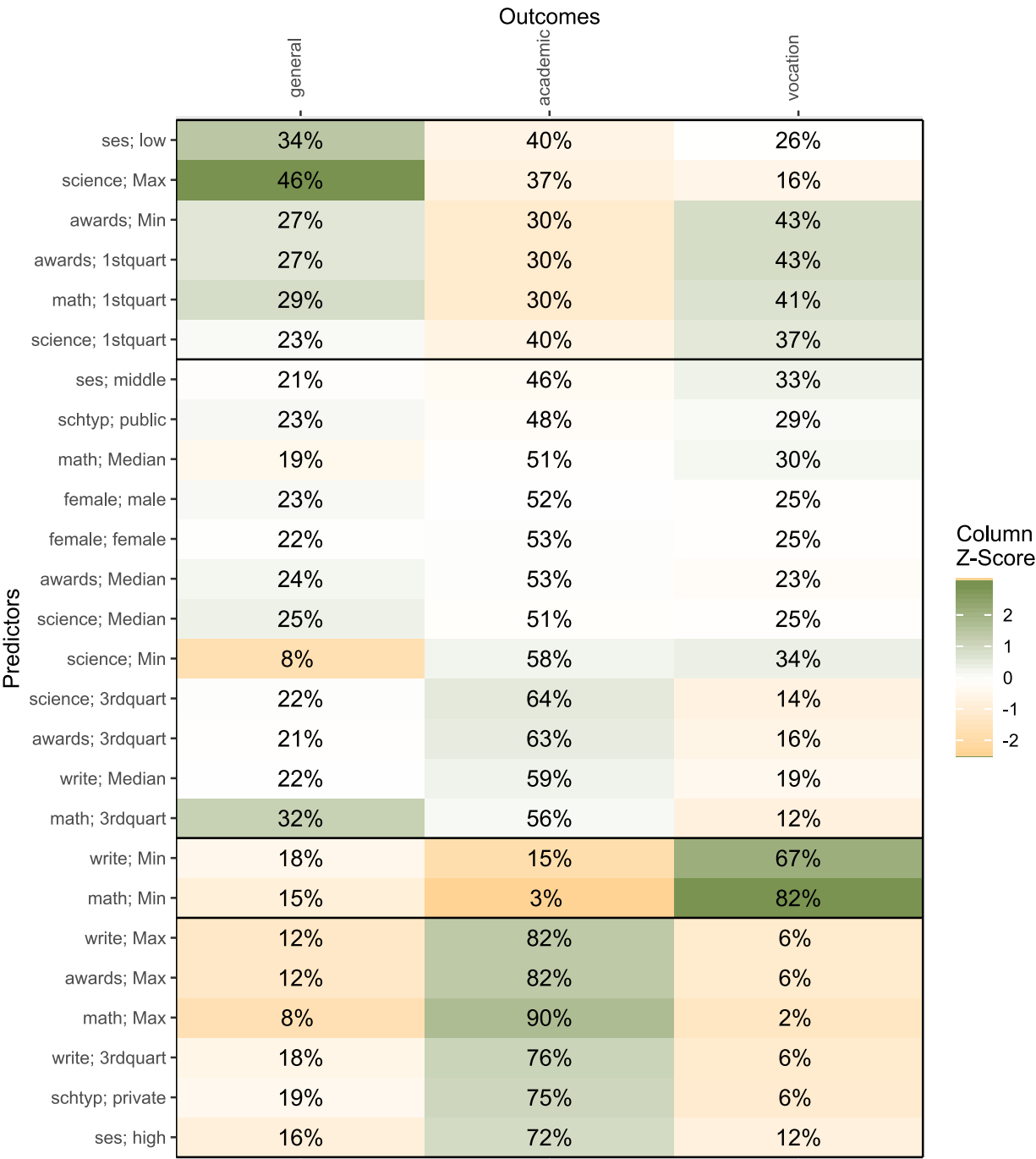
Variable	General	Academic	Vocation
ses; high	-0.77	0.96	-0.74
schtyp; private	-0.37	1.09	-1.05
write; 3rdquart	-0.50	1.16	-1.07
math; Max	-1.64	1.80	-1.29
awards; Max	-1.21	1.44	-1.08
write; Max	-1.21	1.44	-1.08
math; Min	-0.80	-2.34	2.96
write; Min	-0.41	-1.79	2.18
math; 3rdquart	1.23	0.17	-0.72
write; Median	0.02	0.33	-0.37
awards; 3rdquart	-0.05	0.51	-0.55
science; 3rdquart	0.05	0.57	-0.66
science; Min	-1.68	0.28	0.42
science; Median	0.37	-0.08	-0.07
awards; Median	0.25	0.05	-0.16
female; female	0.03	0.05	-0.06
female; male	0.16	-0.03	-0.04
math; Median	-0.39	-0.04	0.22
schtyp; public	0.18	-0.19	0.14
ses; middle	-0.09	-0.28	0.35
science; 1stquart	0.14	-0.60	0.61
math; 1stquart	0.92	-1.06	0.78
awards; 1stquart	0.65	-1.07	0.91
awards; Min	0.65	-1.07	0.91
science; Max	2.98	-0.70	-0.51
ses; low	1.50	-0.56	-0.02

Heatmap of Predicted Probabilities



HEAT MAP

Heatmap of Predicted Probabilities



HEAT MAP
+
CLUSTERS