Using Analytics to Minimize Student Course Withdrawals

Wednesday May 27, 2015

9:45 AM - 10:30 AM

Meeting Room 201 (Theater)

Greg V. Michalski, Ph.D., PMP®

Florida State College at Jacksonville

Session Description (Session Id: 1095)

Excessive course withdrawals are costly to both the student and the institution. While most institutions have systems to track and report basic descriptive information (e.g., counts and percentages), less attention is typically paid to a student's precise (and often complex) reason(s) for withdrawal. Building upon the results of **prior empirical work**, this session provides and demonstrates the use of both qualitative and quantitative analytics to process large volumes of raw, unstructured (open) text as extracted from a student withdrawals (text) database. The session focuses on how these text data can be **qualitatively structured** and then transformed numerically for subsequent **quantitative analyses** using appropriate multivariate procedures.

Agenda

- Introductions
  - Audience Quick Poll
    - Student Course Withdrawal Research/Projects, Text Mining, Analytics
  - Presenter Introduction
- Situating the Problem: Significance, Scope, Context
- Text Analytics Overview (consider “Mining Text Data for Useful Information” session 1125 tomorrow, for more information)
- Text Analytics Application: Summary and Extension of Prior Results
- Overview of Process and Methods
- Quantitative Analysis Results
  - Hierarchical Agglomerative Cluster Analysis (HCA)
  - Principal Components Analysis (PCA)
  - Multiple Correspondence Analysis (MCORA)
- So What? Practical Application: The REASON (assessment)
- Questions and Conclusion

Presenter

Greg V. Michalski
Florida State College at Jacksonville
Institutional Research, Analytics, Educational Evaluation
Verified email at fscj.edu

- Director of Student Analytics and Research, Florida State College at Jacksonville
  - Karen Stearns, Sr. Research Analyst and SQL database extraordinaire
  - Steve Kruszewski,Assistant Research Analyst
- Ph.D. Educational Measurement and Evaluation, University of Ottawa (Ontario, Canada)
- Current research interest and publication background in Institutional Research, Analytics, Evaluation
- Professional Background in Public, Private, and Non-Profit Businesses and Organizations
  - Bell Northern Research (BNR), Nortel Networks, Convergys, Winn-Dixie Information Technology, Zenith Data Systems, GVM Solutions, ICATT Consulting
- Post-Secondary Teaching (mostly adjunct)
  - Lake Michigan College, Jordan College, Illinois Technical College, University of Phoenix, Virginia Polytechnic Institute and State University (VA Tech., Blacksburg), Jacksonville University
- Education Measurement/Evaluation, Psychometrics, and Testing
  - ACT, Inc. (American College Testing, Iowa City)
- Phone: 904-632-3017 and E-mail: gmichals@fscj.edu
- LinkedIn

Section 1:

Situating the Problem: Significance, Scope, Context

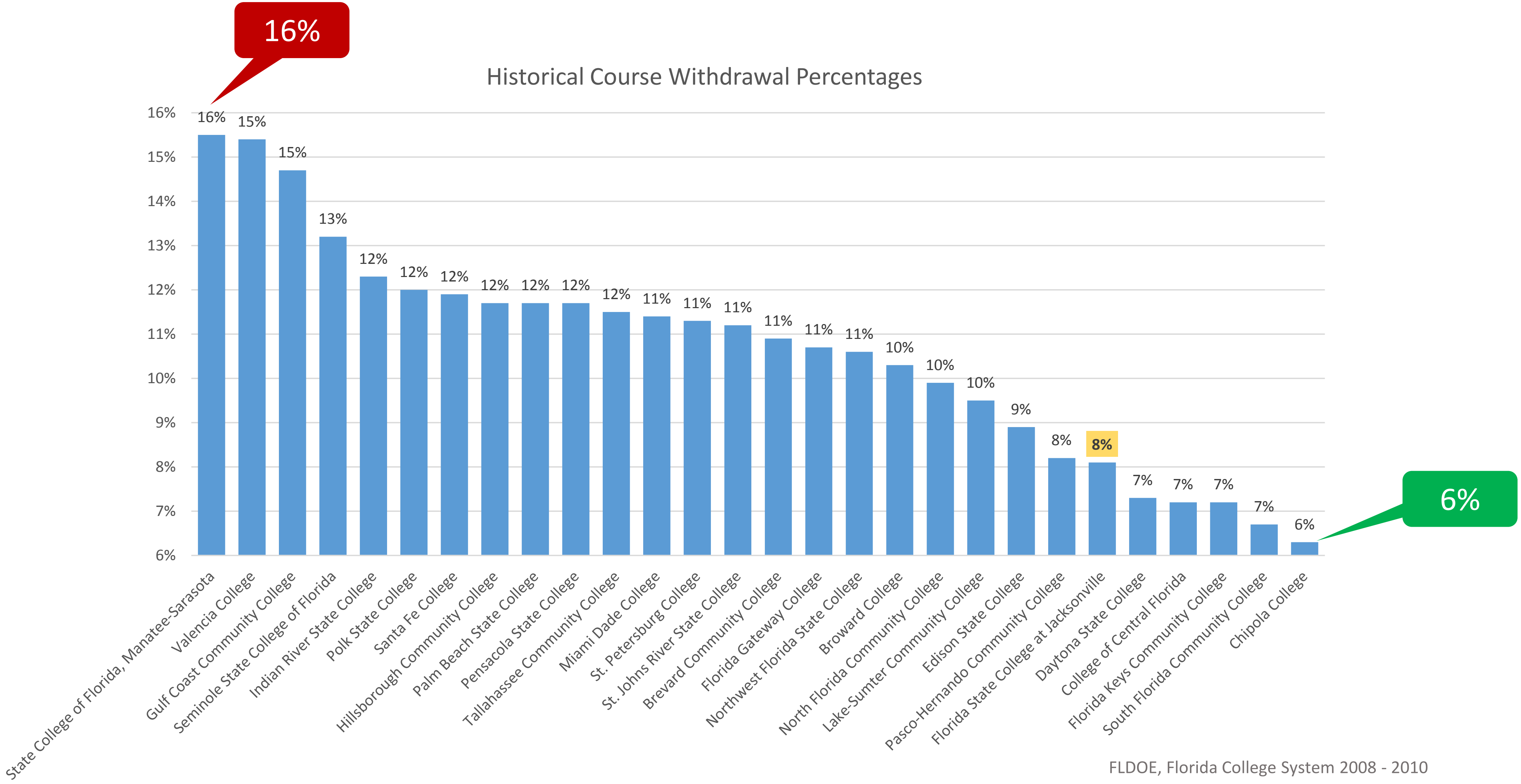
Problem Significance

- When students enroll in, but fail to complete a course, it costs the student and the state money, reduces available classroom space, and increases the amount of time for the student to complete their degree. Clearly, many withdrawals are necessary for personal and academic reasons, but when withdrawals become excessive they pose a significant burden on the student, the college, and the state.

(Florida Department of Education, March, 2011, p. 1)

- Student retention is one of the most widely studied areas in higher education.
  - In addition to the extensive body of research literature that now spans more than four decades, there are books and edited volumes, a journal, and a variety of conferences dedicated solely to student retention (Tinto, 2006)

Context: Course Withdrawal Rates System Wide



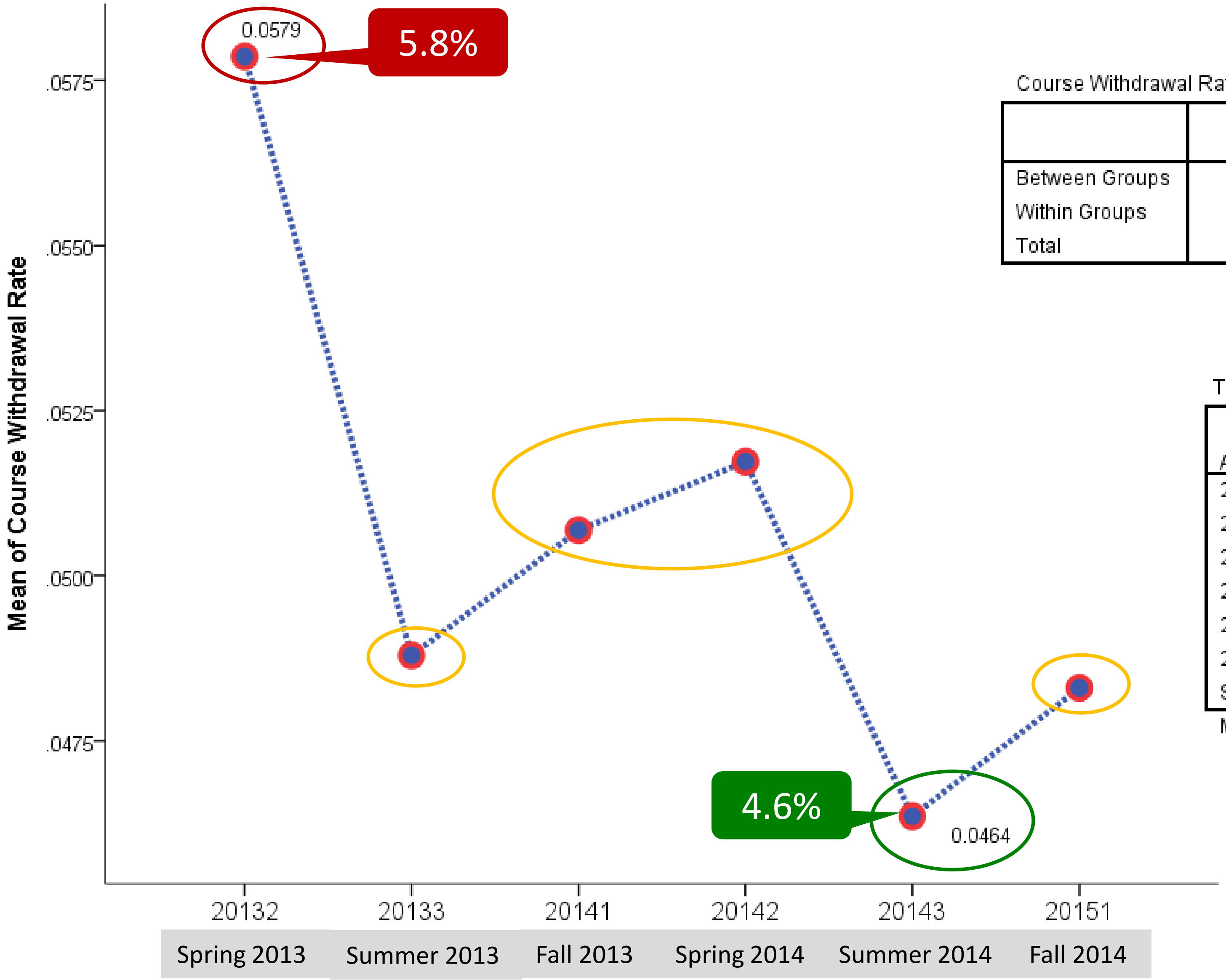
FLDOE, Florida College System 2008 - 2010

Retention and Course Withdrawal (Research Nuggets)

- At least two general classes of reasons involving largely academic, related to areas such as grades, instructors, and course and non-academic, related to areas such as family, illness, and military service (Dunwoody & Frank, 1995; Wiley, 2009)
- Most frequently cited reasons for student course withdrawal (1) job conflict, (2) inadequate preparation for the course, (3) dislike of the class, (4) assignments too heavy, (5) indefinite motivation, (6) illness, and (7) dislike of the instruction (Friedlander, 1981)
- Others cite (1) instructor, (2) class, (3) grade/grading system, (4) course load, (5) time-schedule conflict with other activities, and (6) personal/health/family (Lunnenborg, 1974)

Prior Six-Term FSCJ Course Withdrawal Rates (Overall)

n = 347,302



ANOVA

Course Withdrawal Rate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.746	5	.949	19.524	.000
Within Groups	16885.657	347296	.049		
Total	16890.403	347301			

Course Withdrawal Rate

Tukey HSD[a,b]

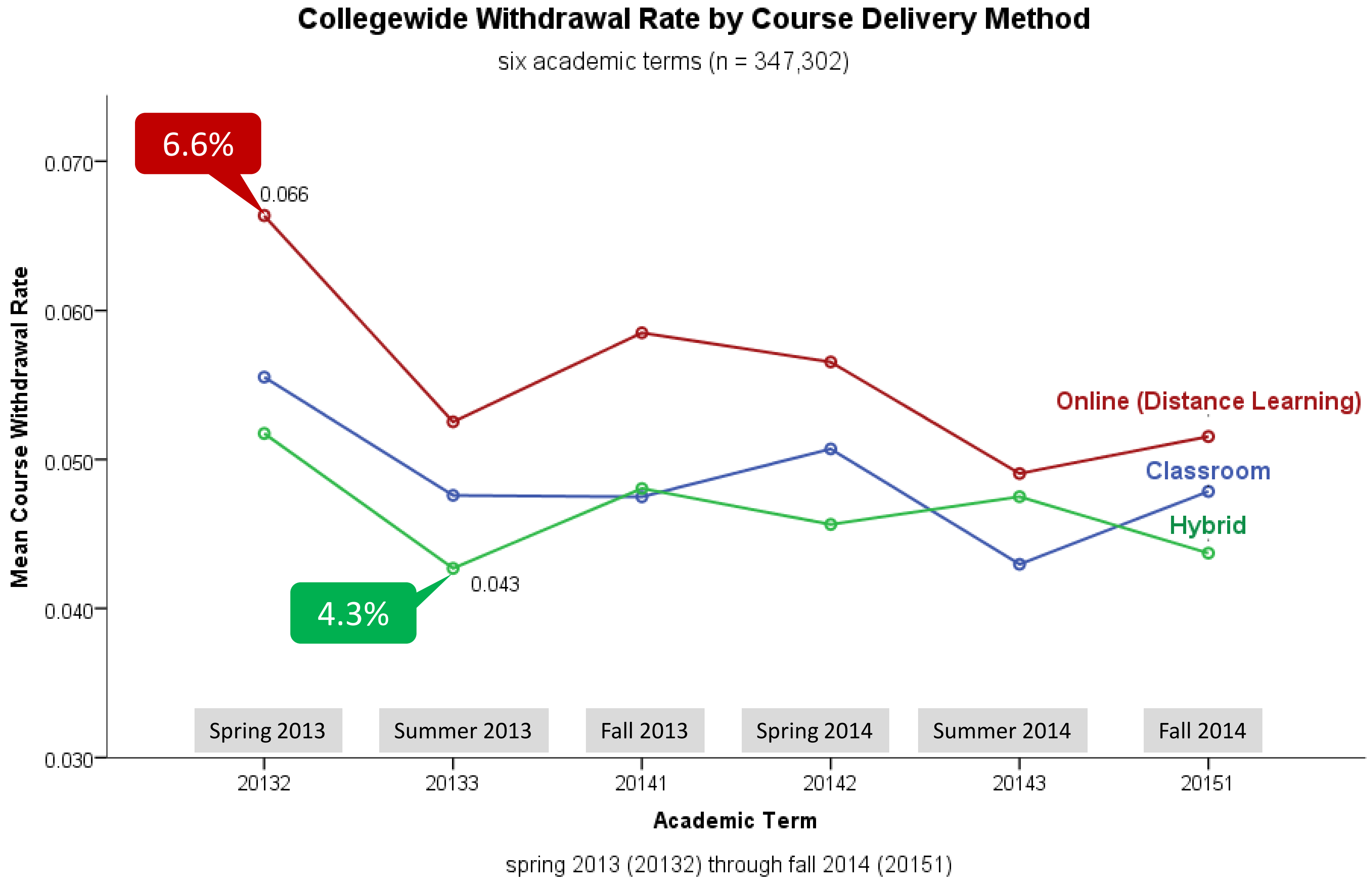
Academic Term	N	Subset for alpha = 0.05		
		1	2	3
20143	34344	.04635		
20151	66751	.04830	.04830	
20133	36462	.04879	.04879	
20141	71124		.05069	
20142	68091		.05172	
20132	70530			.05786
Sig.		.473	.119	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 52425.293.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Prior Six-Term FSCJ Course Withdrawal Rates (by Delivery Method)



Section 2

Text Analytics Overview

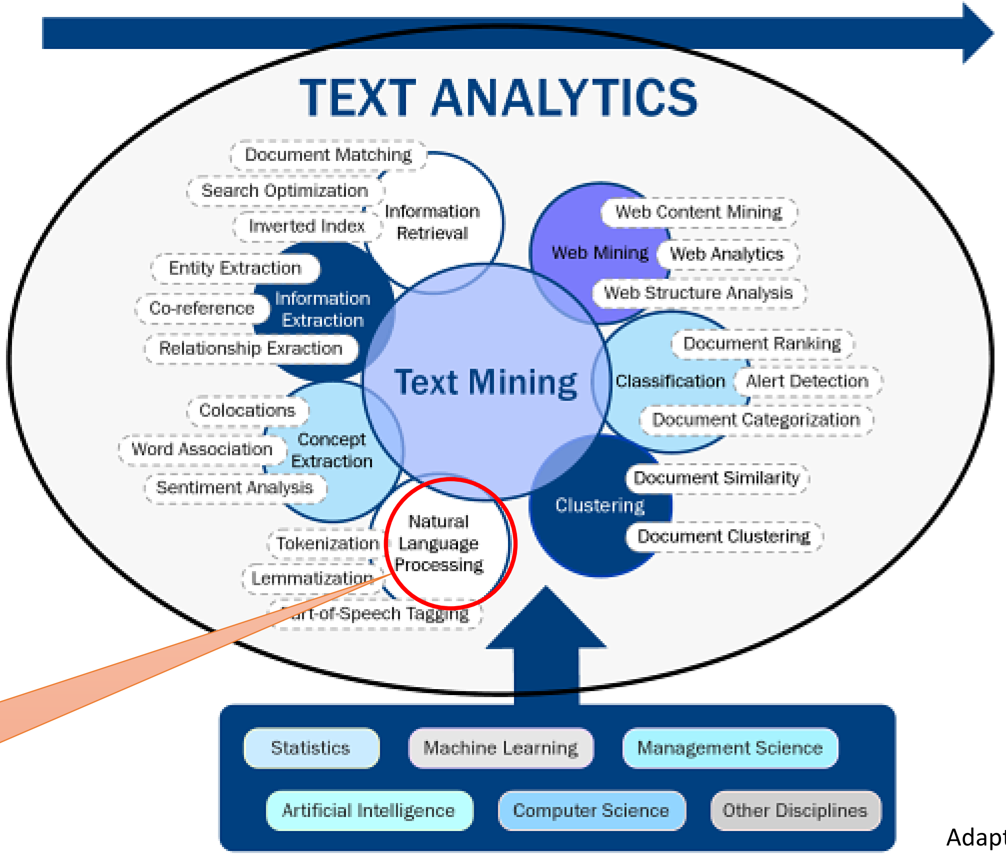
Text Analytics and Text Mining

- Generally considered part of the broader field of data mining, text analytics is a relatively new and still evolving field. Some current definitions include
  - Text mining and text analytics are broad umbrella terms describing a range of technologies for analyzing and processing semi structured and unstructured text data. (Miner, et al, 2012)
  - Text mining involves the discovery of useful and previously unknown "gems" of information from textual document repositories based upon patterns extracted from natural language (Zhang & Segall, 2010)
  - Text mining is the study and practice of extracting information from text using the principles of computational linguistics (Singh and Raghuvanshi, 2012)
  - Process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules (Nisbet, Elder, Elder, & Miner, 2009)
  - A form of qualitative analysis, [involving] the extraction of useful information from text (such as open-ended responses) so that the key ideas or concepts contained within this text can be grouped into an appropriate number of categories (SPSS, Inc., 2009, p. 5)
  - Text mining is the practical application of many techniques of analytical processing in text analytics. (Miner, et al, 2012)

Knowledge and Discipline Areas

Text Mining

- Draws from many techniques in the broader field of text analytics
- The practical application of many techniques of analytical processing



Natural Language Processing

Text Analytics

- Statistics
- Machine Learning
- Management Science
- Artificial Intelligence
- Computer Science
- Other disciplines

Adapted from Miner et all (2012)

Analytical modeling is an iterative process, just like sculpture. When we are satisfied that we have the best model (among alternatives), we can use the model (or deploy it) to make decisions… (p. xxxv)

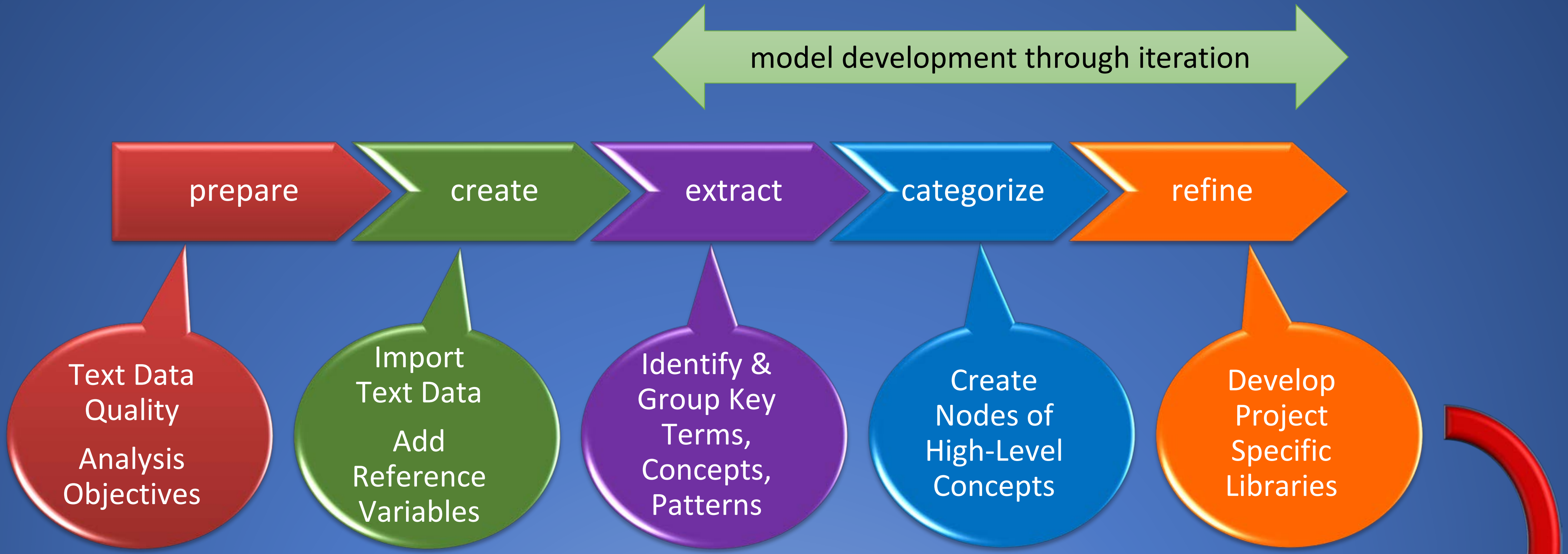
Miner, G. [et al] (2012). Practical text mining and statistical analysis for non-structured text data applications. Waltham, MA: Academic Press.

Natural Language Processing (NLP) (computational linguistics)

- Entity recognition—Find proper nouns and identify the type of noun (person, place, thing , location, product name, brand name)
- Information extraction—Identify specific information about an entity
- Relationship extraction—Follow the relationship between entities such as suppliers and distributors
- Morphological segmentation—Identify word forms (morphemes) such as "competitor" vs. "competitors " vs. "competing"
- Parsing—Understand and analyze the structure of a sentence
- Discourse analysis—Follow a discussion from one sentence to another to identify clarifications, elaborations, and add on information
- Disambiguation—Differentiate similar terms and phrases; parse irregular usages such as slang, sarcasm, exaggeration, and understatement
- Sentiment analysis—Determine the emotional content of a text

(Markham, Kowolenko, & Michaelis, 2015)

Process Overview



Export node contents (categories), ID, and reference (e.g., demographic) variables for further analysis and graphing. Output can be a set of multiple-response variables for quantitative analysis, for example, based on inter-node correlations, or cluster analysis of records coded into multiple categories.

Section 3

Text Analytics Application:
Summary and Extension of Prior Results

Prior Empirical Work

Michalski, G. V. (2014). In their own words: A text analytics investigation of college course attrition. *Community college Journal of Research and Practice, 38*(9), 811-826.

Community College Journal of Research and Practice, 38: 811–826, 2014

ISSN: 1066-8926 print/1521-0413 online
DOI: 10.1080/10668926.2012.720865

Routledge
Taylor & Francis Group

In Their Own Words: A Text Analytics Investigation of College Course Attrition

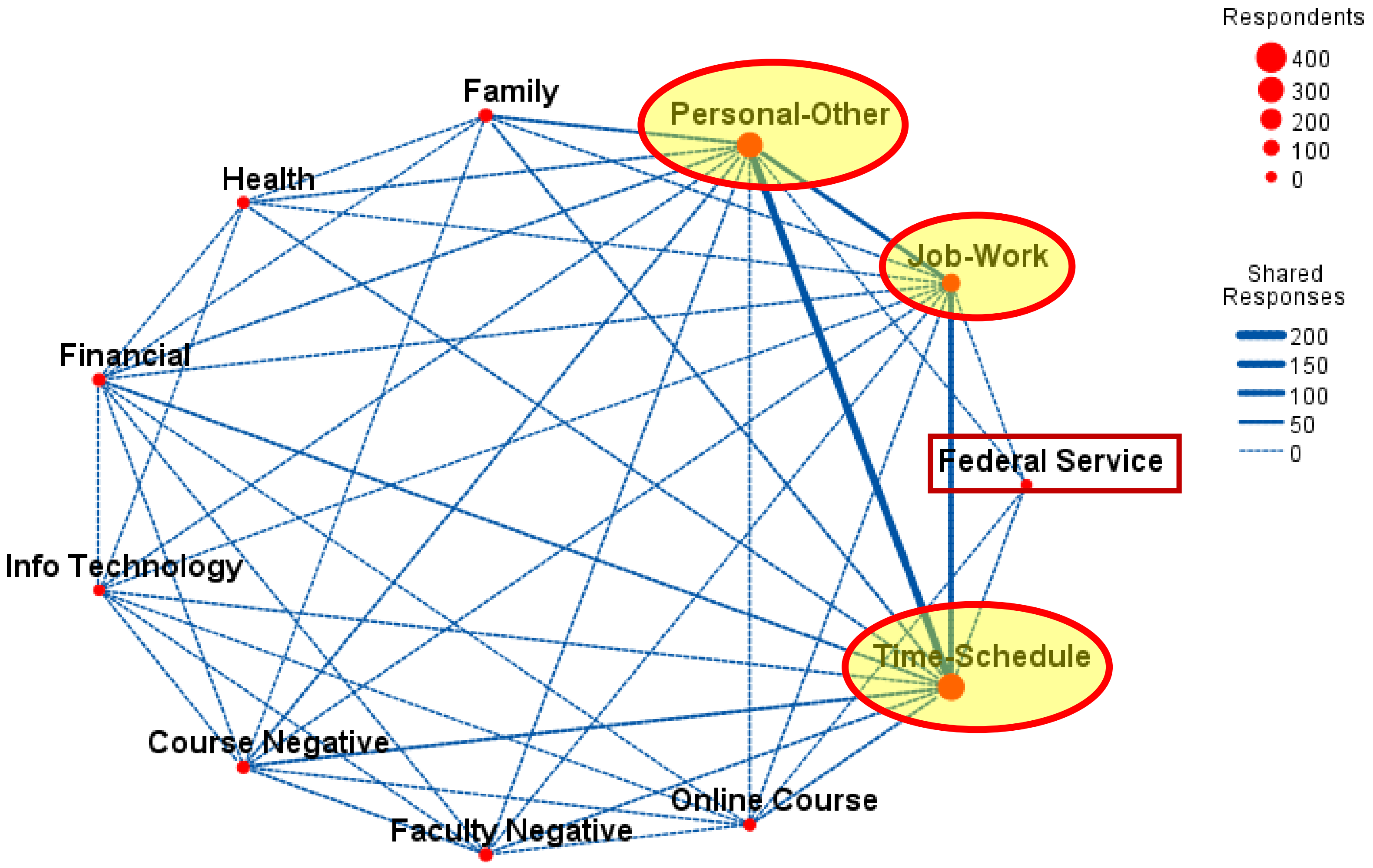
Greg V. Michalski

Office of Student Analytics and Research, Florida State College at Jacksonville, Jacksonville, Florida, USA

Natural language processing (NLP)

- Natural Language Processing text analysis of verbatim text
- 1,295 student comments describing reason(s) for course withdrawal
  - 616 comments used to develop preliminary model
    - categorized 96.1% of all records in eleven categories (nodes)
  - 679 comments used to test model
    - categorized 98.7% of spring term records
  - Referenced prior empirical work in college course withdrawal to label final model nodes
    - academic rationales involve course and faculty; academic schedule; course delivery mode changes
    - non-academic rationales involve personal issues especially job/work, family, financial, health
- Final text model results were exported and further analyzed using Hierarchical Cluster Analysis (HCA), Principal Component Analysis (PCA), and Multiple Correspondence Analysis (MCORA)

Model of Course Withdrawal Explanations



Michalski, 2014, p.7

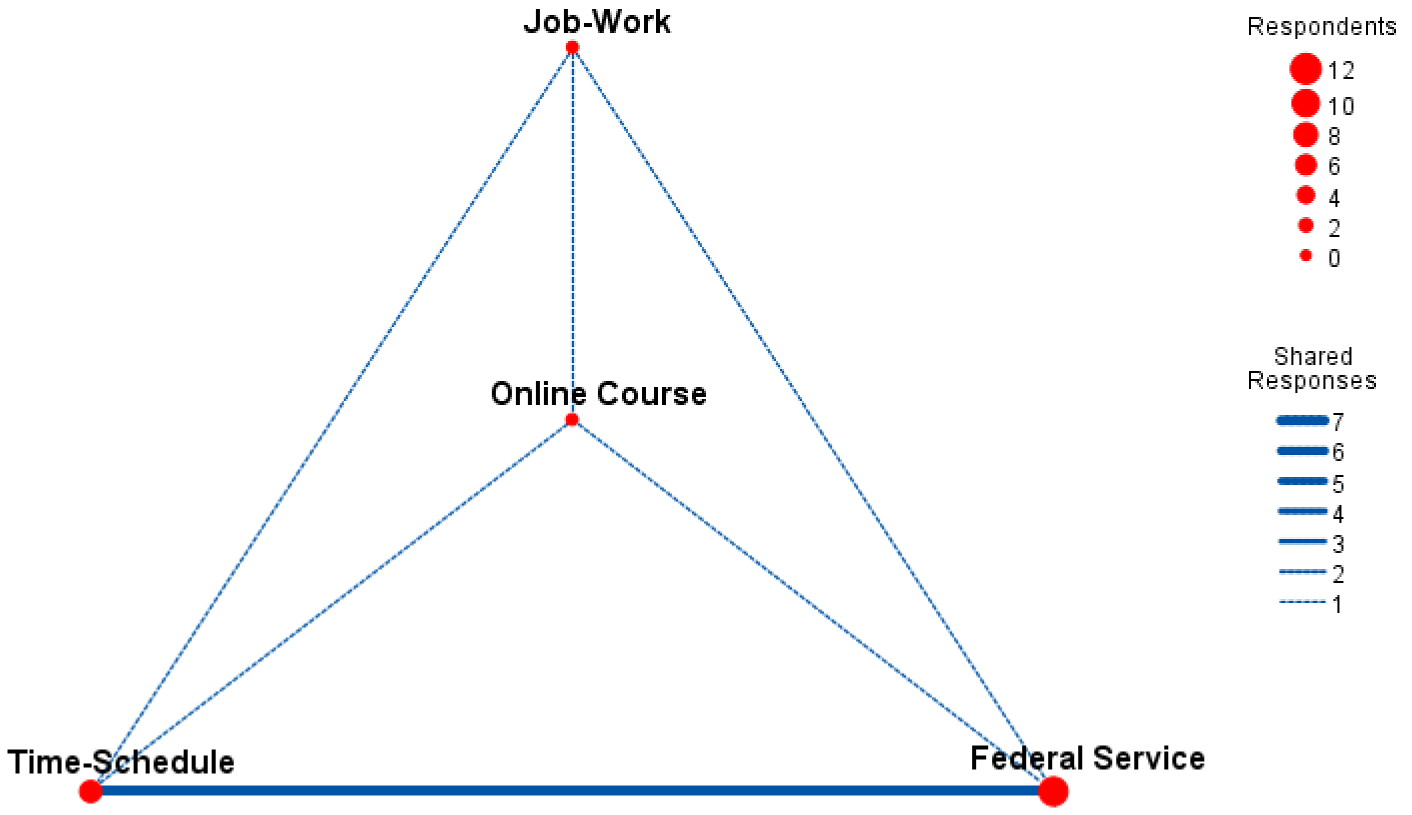
FIGURE 1
Category web model showing eleven labeled nodes. The model uses relative circle diameter to represent the number of student withdrawal comments coded into each node, and relative line thickness to represent the number of comments shared between nodes. The phase one model categorized 96.1% of student course withdrawal explanations from the fall 2010 academic term.

Model Coding Example: Federal Service Node

Complex Coding Example

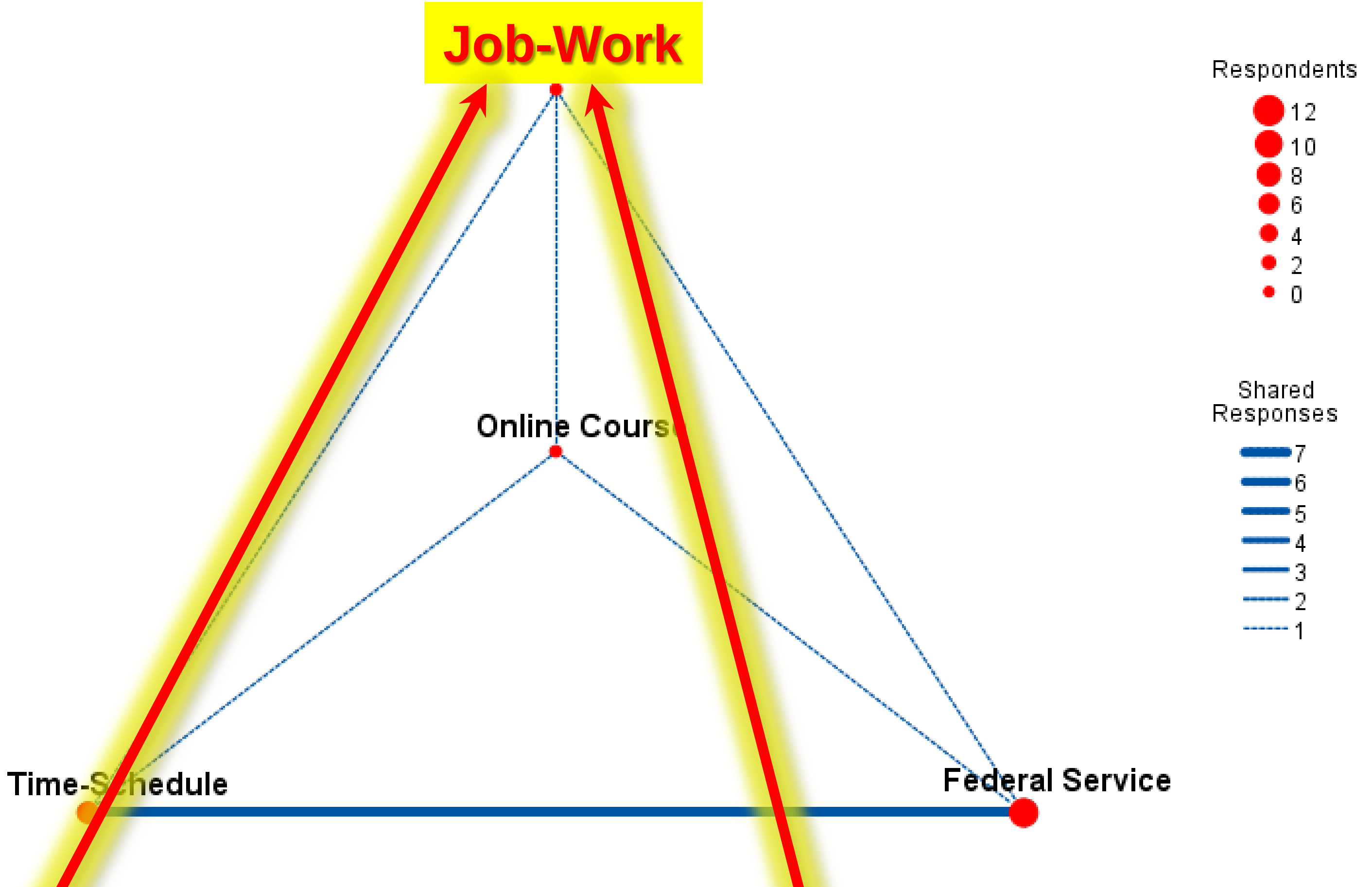
- How did the text miner code this student withdrawal comment into multiple nodes?

I have to withdraw from this class. I am a contractor for the Department of Homeland Security. I have to travel to Guantanamo Bay, Cuba every month for work and do not have time at this point to have an on campus dedicated class. My other two classes are online. If that is an option for this class I would like to do it online as well.



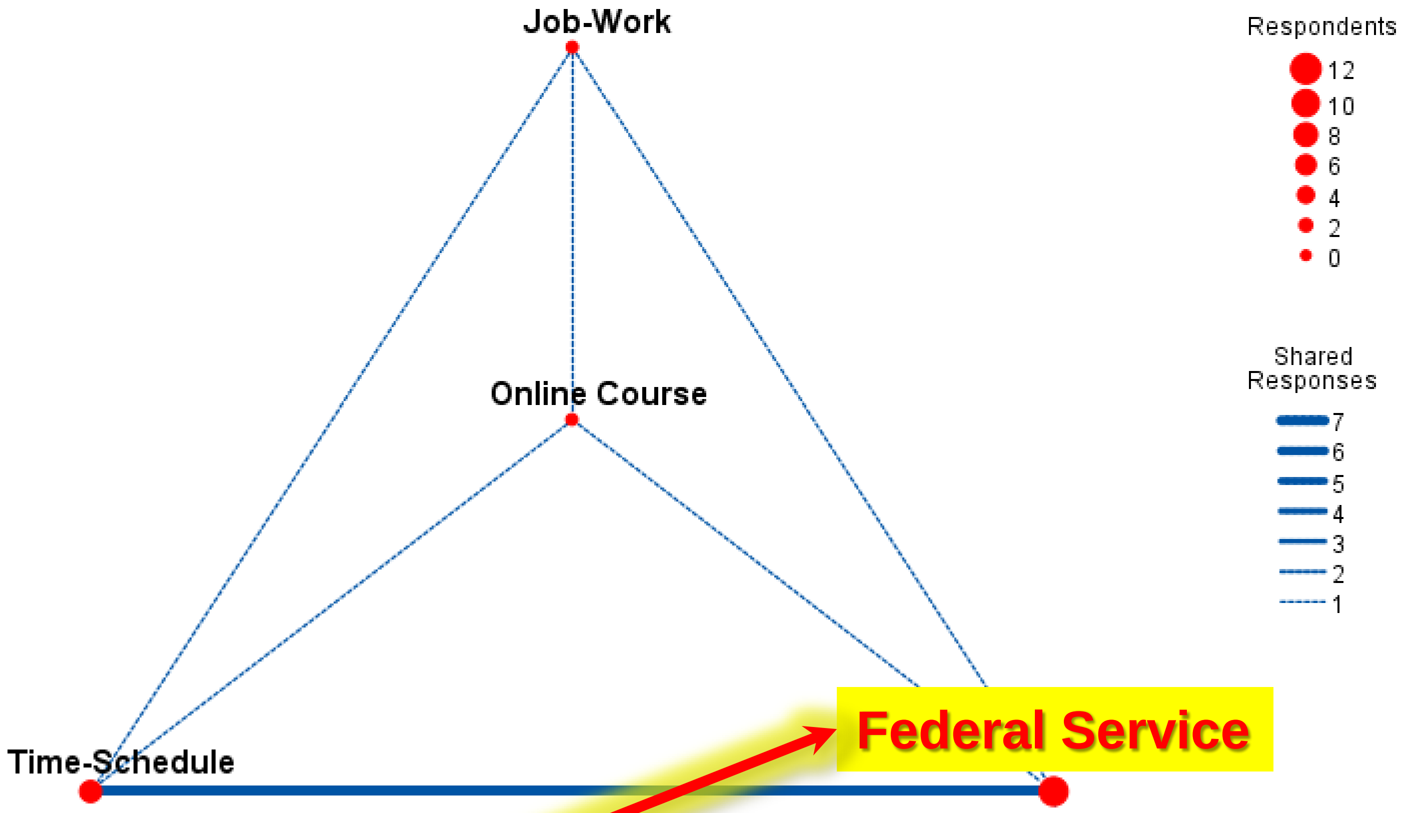
Federal Service Coding Example: Single Withdrawal Comment Coded into Four Nodes

I have to withdraw from this class. I am a contractor for the Department of Homeland Security. I have to travel to Guantanamo Bay, Cuba every month for work and do not have time at this point to have an on campus dedicated class. My other two classes are online. If that is an option for this class I would like to do it online as well.



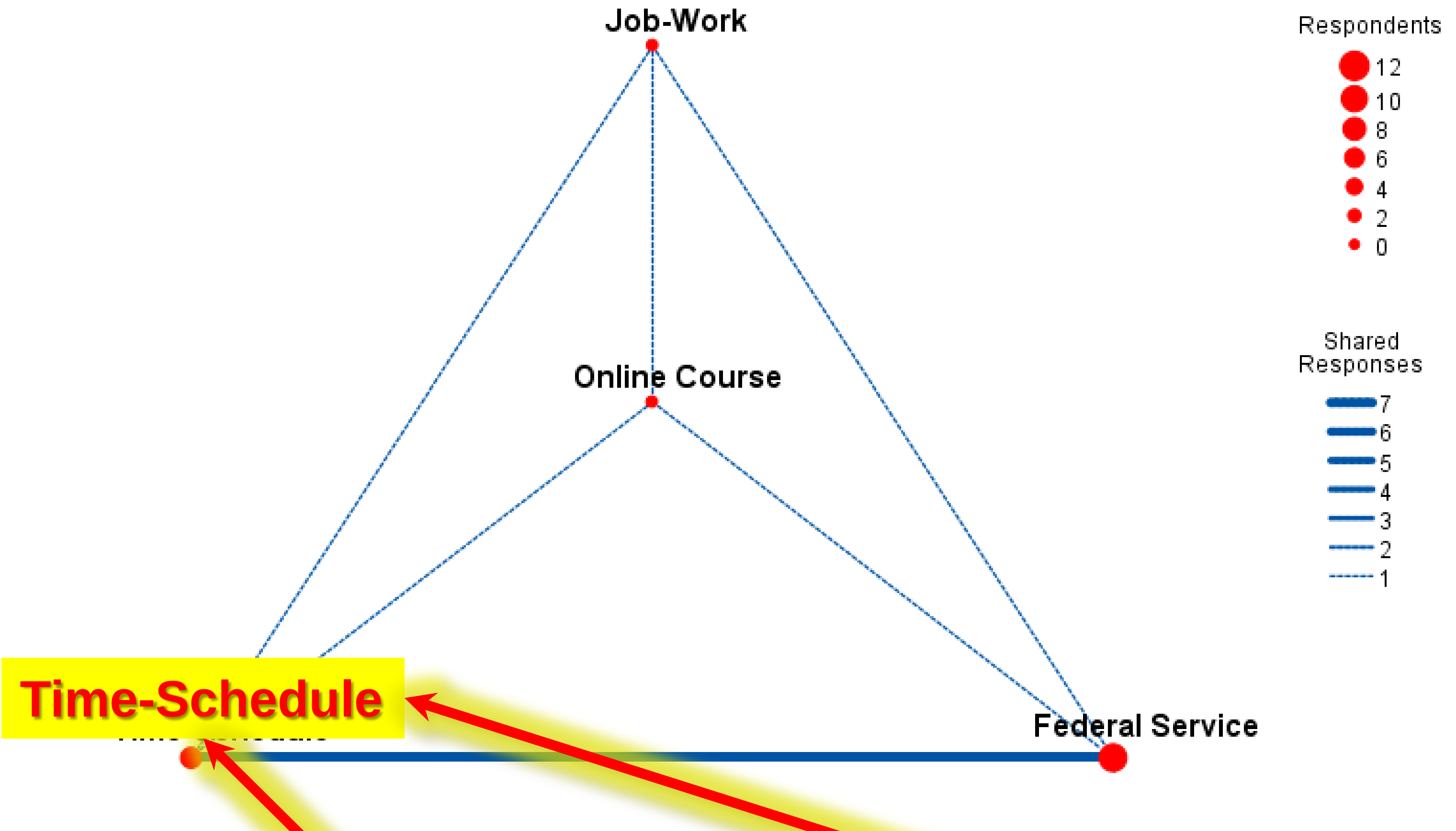
I have to withdraw from this class. **I am a contractor** for the Department of Homeland Security. I have to travel to Guantanamo Bay, Cuba every month for **work** and do not have time at this point to have an on campus dedicated class. My other two classes are online. If that is an option for this class I would like to do it online as well.

Federal Service Coding Example:
1 of 4
Job-Work



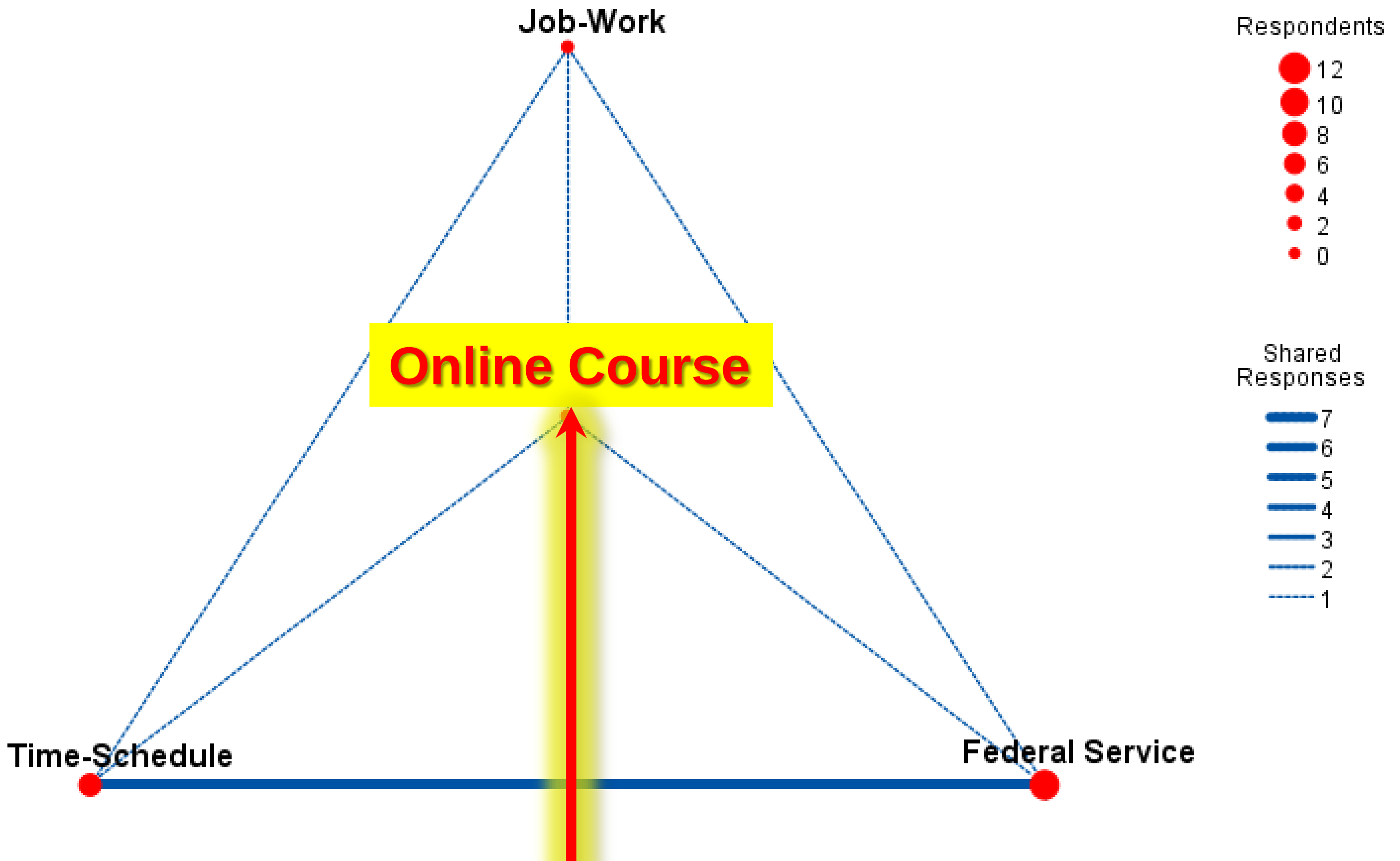
I have to withdraw from this class. I am a contractor for the **Department of Homeland Security**. I have to travel to Guantanamo Bay, Cuba every month for work and do not have time at this point to have an on campus dedicated class. My other two classes are online. If that is an option for this class I would like to do it online as well.

Federal Service Coding Example:
2 of 4
Federal Service



I have to withdraw from this class. I am a contractor for the Department of Homeland Security. I have to travel to Guantanamo Bay, Cuba **every month** for work and **do not have time** at this point to have an on campus dedicated class. My other two classes are online. If that is an option for this class I would like to do it online as well.

Federal Service Coding Example: 3 of 4 **Time-Schedule**



I have to withdraw from this class. I am a contractor for the Department of Homeland Security. I have to travel to Guantanamo Bay, Cuba every month for work and do not have time at this point to have an on campus dedicated class. **My other two classes are online**. If that is an option for this class **I would like to do it online as well**.

Federal Service Coding Example:
4 of 4
Online Course

Section 4

Overview of Process and Methods

Text Extraction, Preparation, Analysis: Practical Process Overview

Raw Text Data
- Open ended or semi-structured text (e.g., from database, survey, or other ERP system(s)

Query/Extract
- Define parameters (e.g., date range, question ID, etc.) extract / export text to Excel

Text Analysis File Preparation
- Add unique record key/ID
- Create/name reference variables or fields
- Minor error corrections (e.g., spell check, standardize acronyms, etc.

IBM/SPSS Text Analytics Import
- Cleaned text data with unique record key identifiers imported at new project

Text Mining and Analysis
- Extract terms, concepts, types
- Build categories
- Iterate, and finalize
- **Export** for quantitative analysis (SPSS)

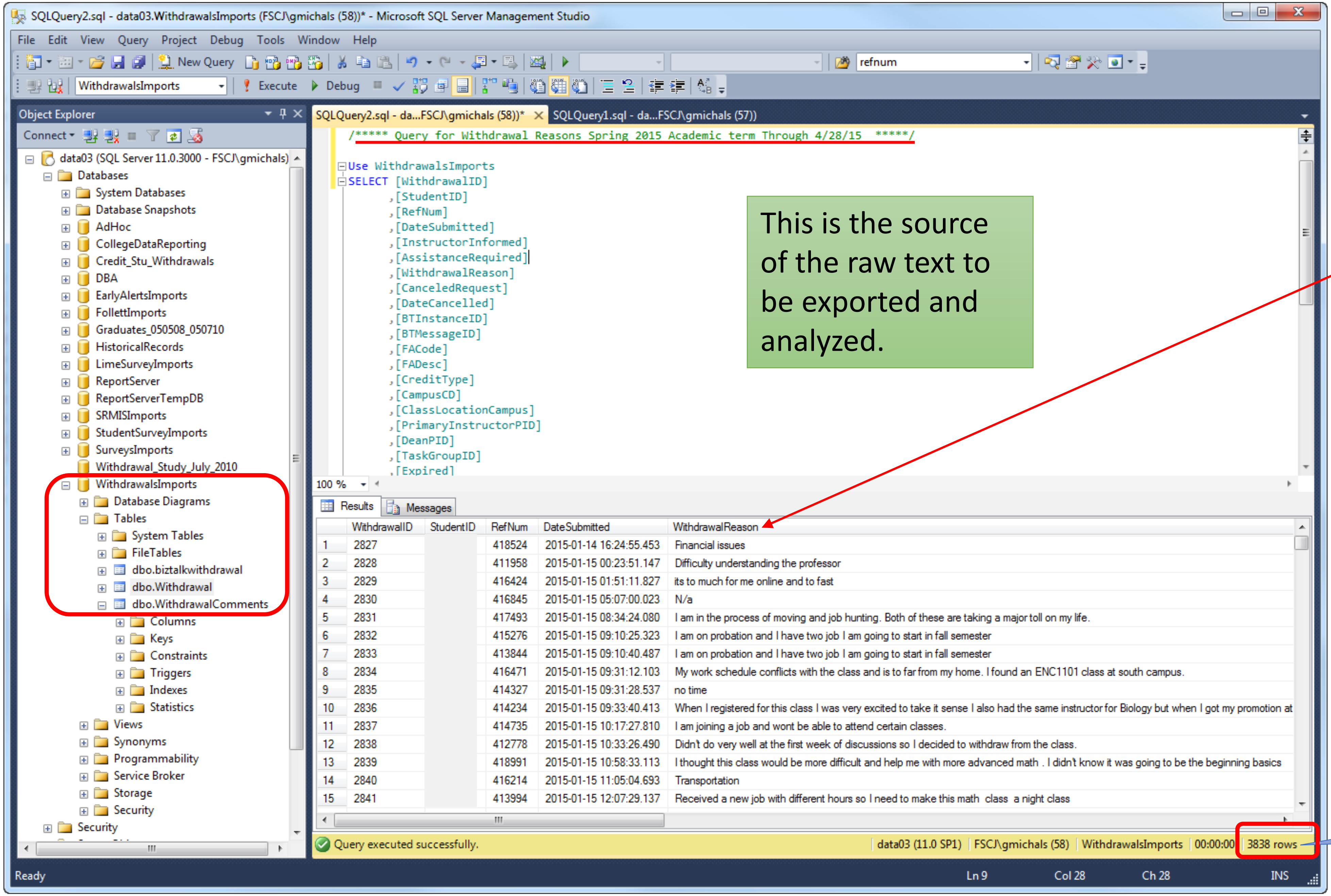
Microsoft® SQL Server 2012

Excel

IBM® SPSS® Text Analytics for Surveys
Version 4.0.1

IBM® SPSS® Modeler
Version 16.0

Data Source: Withdrawals Database



This is the source of the raw text to be exported and analyzed.

Spring 2015 Term

SQL Query to extract course withdrawal comments (n = 3,838)

'WithdrawalReason' field contains written explanation for course withdrawal as provided by the student at the time of withdrawal.

3,838 rows

Variable View (first 20 cases to illustrate coding)

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values
1	WD_ID_Combine	Numeric	3	0	WithdrawalID_Combined	None
2	Term	Numeric	8	0	Academic Term	{1, fall 2010}...
3	time_sched	Numeric	4	0	Time-Schedule	{0, false}...
4	personal_other	Numeric	4	0	Personal-Other	{0, false}...
5	job_work	Numeric	4	0	Job-Work	{0, false}...
6	family	Numeric	4	0	Family	{0, false}...
7	course_neg	Numeric	4	0	Course Negative	{0, false}...
8	faculty_negative	Numeric	4	0	Faculty Negative	{0, false}...
9	financial	Numeric	4	0	Financial	{0, false}...
10	online_course	Numeric	4	0	Online Course	{0, false}...
11	health	Numeric	4	0	Health	{0, false}...
12	info_tech	Numeric	4	0	InfoTechology	{0, false}...
13	fed_service	Numeric	4	0	Federal Service	{0, false}...

Data View (first 20 cases to illustrate coding)

*WD_Text_Fall-Spring_ALL01.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	WD_ID_Combine	Term	time_sched	personal_other	job_work	family	course_neg	faculty_negative	financial	online_course	health	info_tech	fed_service
1	8394	1	1	1	0	0	0	0	0	0	0	0	0
2	8395	1	1	0	0	0	0	0	0	0	0	0	0
3	8396	1	1	0	0	0	0	0	0	0	0	0	0
4	8397	1	0	0	1	0	0	0	0	0	0	0	0
5	8398	1	1	0	0	0	0	0	0	0	0	0	0
6	8399	1	1	0	0	0	0	0	0	0	0	0	0
7	8400	1	1	1	0	0	0	0	1	0	0	1	0
8	8401	1	0	0	0	0	0	0	0	0	1	0	0
9	8402	1	1	0	0	0	0	0	0	0	0	0	0
10	8403	1	0	0	1	0	0	0	0	0	0	0	0
11	8404	1	1	1	0	0	0	0	1	0	0	0	0
12	8405	1	1	0	0	0	0	0	0	0	0	0	0
13	8406	1	1	0	0	0	0	0	0	0	0	0	0
14	8407	1	1	1	0	0	0	0	0	0	0	0	0
15	8408	1	1	0	0	0	0	0	0	0	0	0	0
16	8409	1	0	0	0	0	0	0	0	0	0	1	0
17	8410	1	1	0	0	0	0	0	0	0	0	0	0
18	8411	1	0	0	0	0	0	0	0	0	0	0	0
19	8412	1	0	1	0	0	0	0	0	0	0	0	0
20	8413	1	0	0	1	0	0	0	0	0	0	0	0

Text Model Export for Quantitative Analysis

- Text model data exported for quantitative analysis
  - Objective
    - Analyze text model coding, identify significant quantitative relationships, confirm coding of records, especially those categorized into multiple nodes
- Statistical Procedures Include
  - Hierarchical Agglomerative Cluster Analysis (HCA)*
  - Principal Components Analysis (PCA)
  - Multiple Correspondence Analysis (MCORA)*

Schmidt (2010) used CATPAC application for text analysis (Quantification of transcripts from depth interviews, open ended responses and focus groups). Discussed subsequent analysis using **Hierarchical Cluster Analysis (Ward's Method), Multiple Correspondence Analysis (Optimal Scaling),** Rule-based web categorization (word link frequency), Suggested the application of other techniques including Bayesian heuristics, CHAID, and Latent Class Analysis.

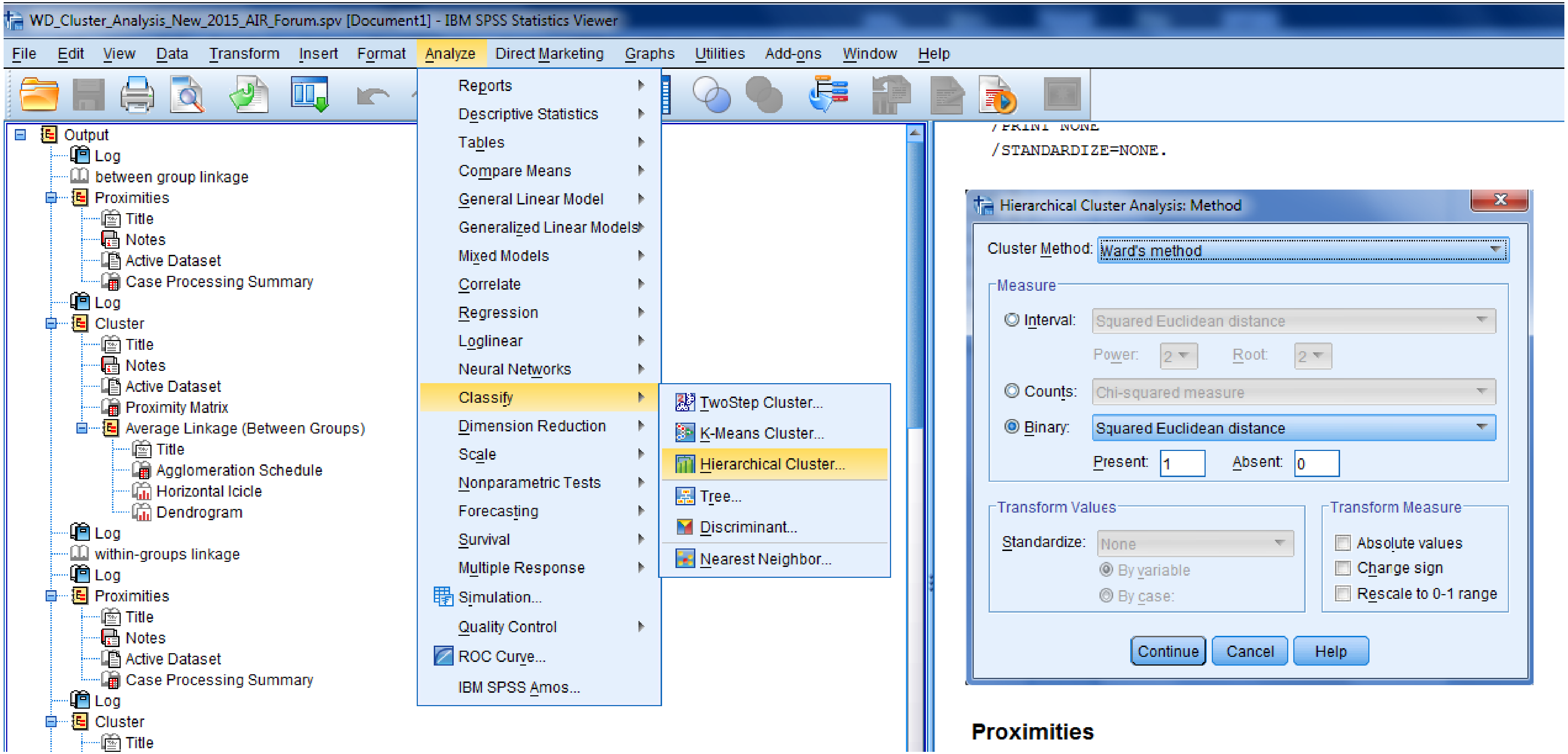
Section 5

Quantitative Analysis 1: Hierarchical Cluster Analysis (HCA)

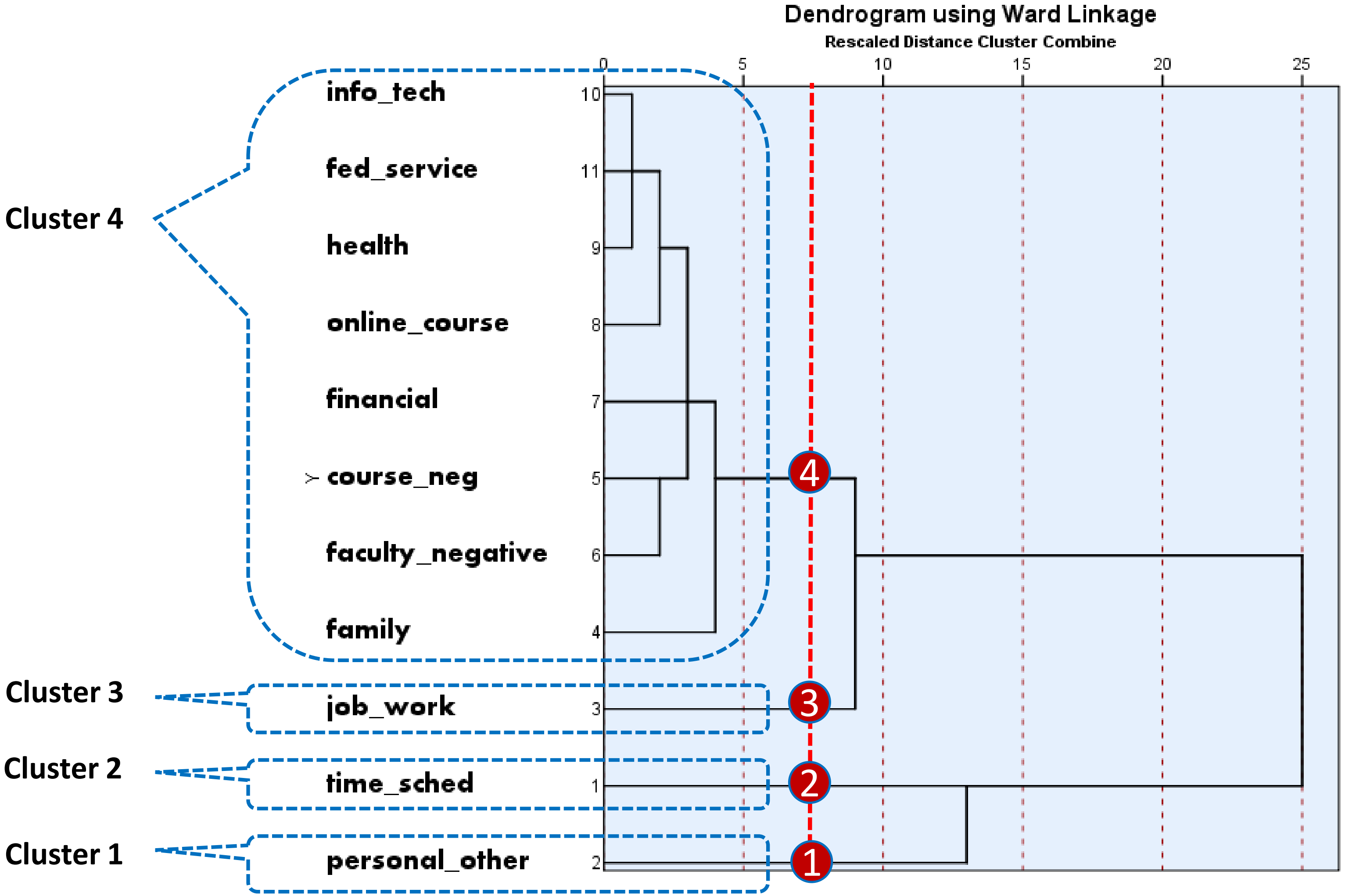
Hierarchical Cluster Analysis (HCA)

- Exploratory multivariate technique designed to reveal natural groupings (or clusters) within a data set that would otherwise not be apparent
  - The objects in hierarchical cluster analysis can be cases or variables, depending on whether you want to classify cases or examine relationships between the variables
  - Most useful for clustering a small number (less than a few hundred) objects
- Commonly used in the social sciences for classification
  - Useful to reveal natural groupings (or clusters) using a variety of methods
- The objective of HCA is to identify relatively homogeneous groups of variables (or cases) based on selected characteristics
- Used in present study to explore relationship among the eleven model categories identified in the text model

HCA Procedure Using SPSS

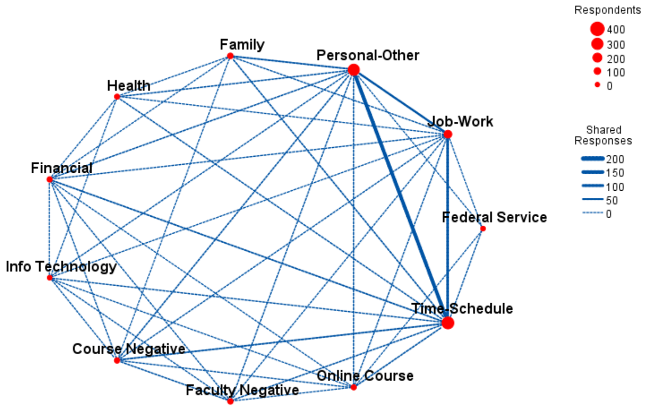


Hierarchical Cluster Analysis (HCA)



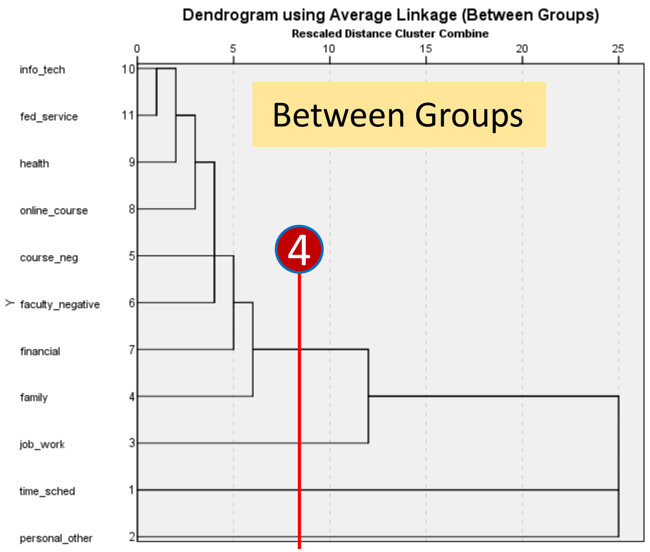
HCA Notes

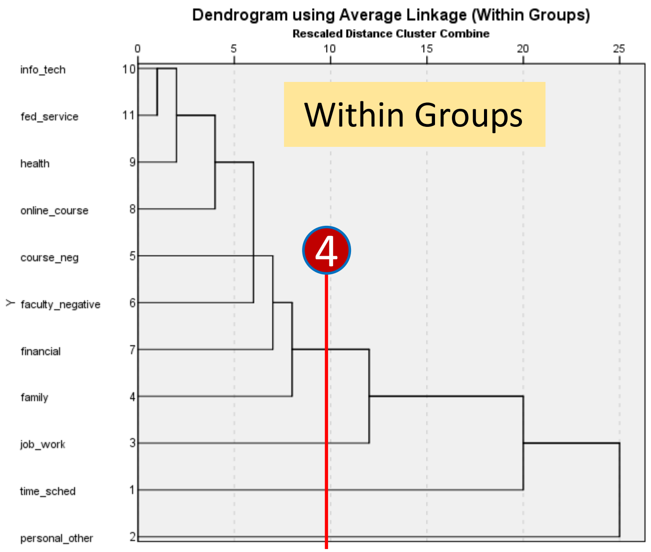
With Measure level set to Binary "Squared Euclidean Distance" the Cluster Methods between-groups, within groups, nearest neighbor, furthest neighbor, centroid linkage, median linkage, and Ward's linkage all provide similar good four-cluster solutions.

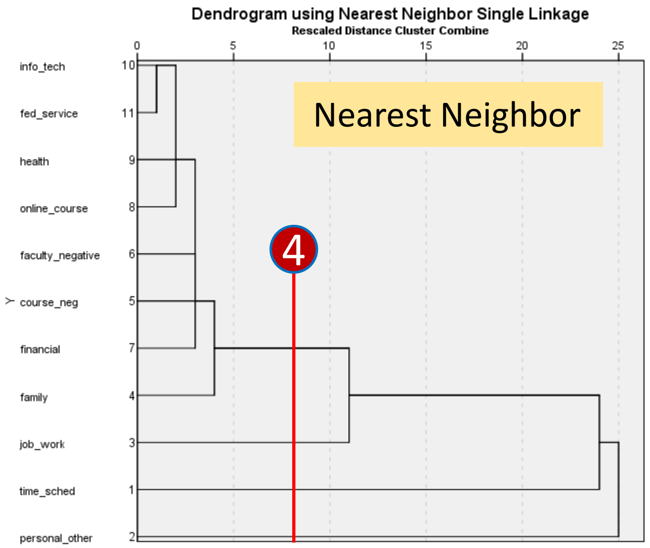


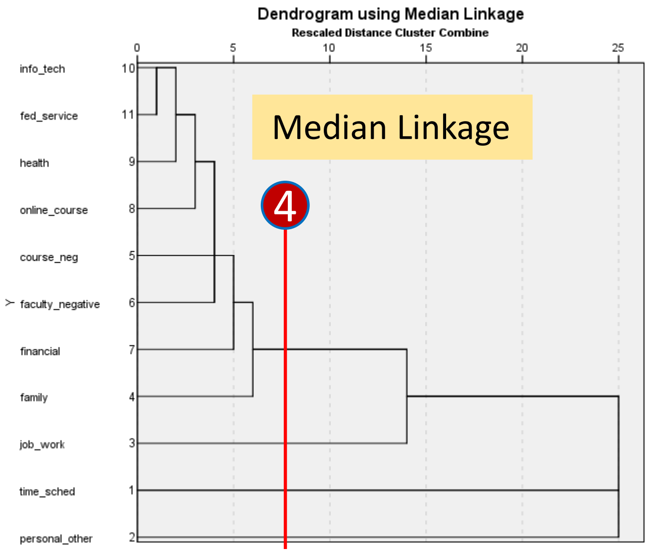
Alternative HCA Methods

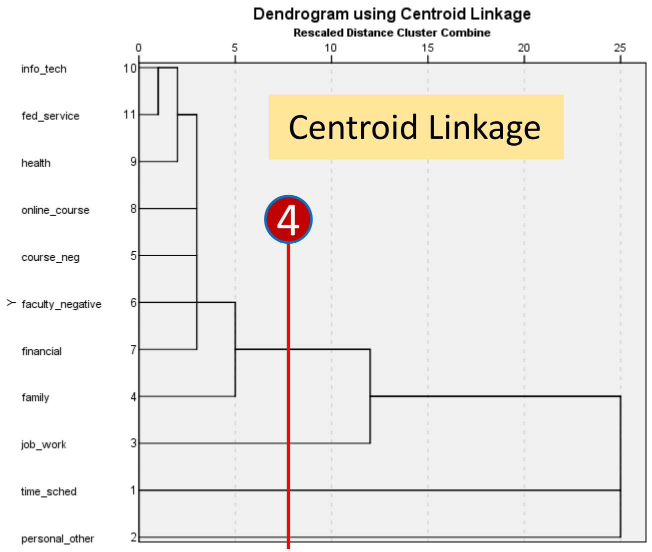
Measure = Binary "Squared Euclidean Distance"

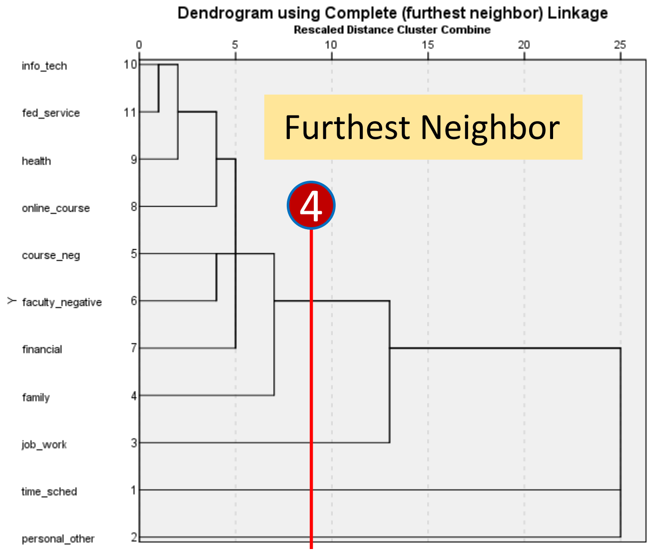












Section 6

Quantitative Analysis 2: Principal Component Analysis (PCA)

Principal Component Analysis (PCA)

- Data reduction technique to identify a small number of factors that explain most of the variance in a data set
- Attempts to identify a smaller set of uncorrelated principal components that explain the pattern of correlations within a set of observed variables
- Used here to understand the structure and patterns of correlations in the model for records coded into multiple nodes
- Prepared for PCA by checking
  - Correlation matrix finding multiple significant correlations
  - Kaiser-Meyer-Olkin Measure of Sampling Adequacy*
  - Bartlett's Test of Sphericity*
- Determined number of components to extract using both Scree plots, and Parallel Analysis

*Kaiser-Meyer-Olkin Measure of Sampling Adequacy was found to be slightly less than 0.60 (0.466 for fall 2010), however, Bartlett's Test of Sphericity was highly significant (approximate chi-square = 388, $p < 0.0001$) supporting the marginal factorability of the correlation matrix

Internode Rank Correlations for Fall 2010

Category[a]		Time-Schedule	Personal-Other	Job-Work	Family	Course Negative	Faculty Negative	Financial	Online Course	Health	Info Technology	Federal Service
Time-Schedule	Spearman's rho (ρ)	1.000	-.070	.004	-.092*	.114**	-.119**	.063	.025	-.032	.076	.027
	Sig. (2-tailed)	.	.083	.917	.022	.005	.003	.121	.541	.427	.060	.507
Personal-Other	Spearman's rho (ρ)	-.070	1.000	-.071	.087*	-.038	-.175**	-.051	-.009	.020	.025	-.107**
	Sig. (2-tailed)	.083	.	.077	.030	.342	.000	.205	.829	.616	.532	.008
Job-Work	Spearman's rho (ρ)	.004	-.071	1.000	-.038	-.033	-.091*	-.018	-.018	-.055	.017	-.046
	Sig. (2-tailed)	.917	.077	.	.349	.416	.024	.658	.661	.171	.665	.251
Family	Spearman's rho (ρ)	-.092*	.087*	-.038	1.000	-.062	-.090*	-.040	-.075	.037	-.047	-.042
	Sig. (2-tailed)	.022	.030	.349	.	.122	.025	.323	.063	.365	.241	.300
Course Negative	Spearman's rho (ρ)	.114**	-.038	-.033	-.062	1.000	.301**	.075	.157**	-.062	.215**	-.037
	Sig. (2-tailed)	.005	.342	.416	.122	.	.000	.063	.000	.124	.000	.360
Faculty Negative	Spearman's rho (ρ)	-.119**	-.175**	-.091*	-.090*	.301**	1.000	.087*	.009	-.066	.199**	-.039
	Sig. (2-tailed)	.003	.000	.024	.025	.000	.	.031	.818	.103	.000	.331
Financial	Spearman's rho (ρ)	.063	-.051	-.018	-.040	.075	.087*	1.000	.017	-.032	.129**	-.037
	Sig. (2-tailed)	.121	.205	.658	.323	.063	.031	.	.665	.422	.001	.360
Online Course	Spearman's rho (ρ)	.025	-.009	-.018	-.075	.157**	.009	.017	1.000	-.055	.154**	.021
	Sig. (2-tailed)	.541	.829	.661	.063	.000	.818	.665	.	.175	.000	.601
Health	Spearman's rho (ρ)	-.032	.020	-.055	.037	-.062	-.066	-.032	-.055	1.000	.016	-.031
	Sig. (2-tailed)	.427	.616	.171	.365	.124	.103	.422	.175	.	.690	.450
Info Technology	Spearman's rho (ρ)	.076	.025	.017	-.047	.215**	.199**	.129**	.154**	.016	1.000	-.021
	Sig. (2-tailed)	.060	.532	.665	.241	.000	.000	.001	.000	.690	.	.610
Federal Service	Spearman's rho (ρ)	.027	-.107**	-.046	-.042	-.037	-.039	-.037	.021	-.031	-.021	1.000
	Sig. (2-tailed)	.507	.008	.251	.300	.360	.331	.360	.601	.450	.610	.

a. Academic Term = fall 2010 (n=616)

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Total Variance Explained by Principal Components Analysis, Both Terms Combined (n=1,295)

Parallel Analysis was used to determine appropriate number of components

Total Variance Explained

Component	Initial Eigenvalues				Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	Parallel Analysis[1]	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	**1.486**	**1.1487 *****	13.512	13.512	1.486	13.512	13.512	1.404	12.761	12.761
2	**1.212**	**1.1066 *****	11.022	24.535	1.212	11.022	24.535	1.206	10.962	23.723
3	**1.136**	**1.0743 *****	10.329	34.863	1.136	10.329	34.863	1.176	10.689	34.412
4	**1.078**	**1.0479 *****	9.798	44.661	1.078	9.798	44.661	1.127	10.248	44.661
5	1.020	1.0213	9.269	53.929						
6	1.002	0.9971	9.113	63.043						
7	.973	0.9745	8.844	71.886						
8	.893	0.9508	8.120	80.006						
9	.876	0.9234	7.966	87.973						
10	.742	0.8962	6.746	94.718						
11	.581	0.8593	5.282	100.000						

Extraction Method: Principal Component Analysis.

1. Randomly Generated Parallel Analysis Eigenvalues for 11 variables, n=1,295 subjects, 100 replications (Watkins, 2006)
*** indicates component should be retained

4 Principal Components explain 45% of Total Variance . Components 1 and 2 explain 24% of Total Variance.

Rotated Component Matrix from PCA of Both Terms Combined (n=1,295)

Rotated Component Matrix[a]

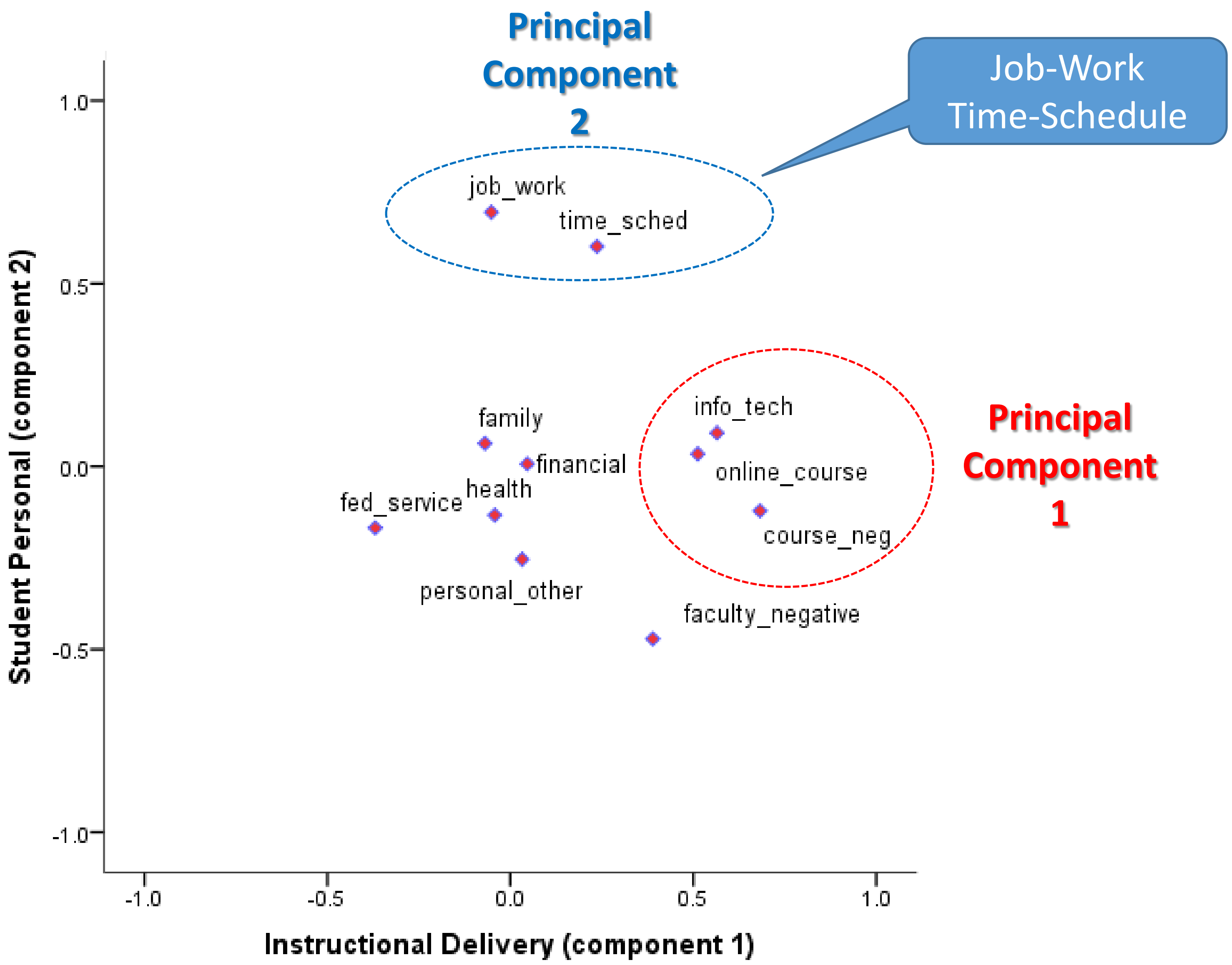
Node	Component			
	1	2	3	4
Course Negative (1)	.683	-.121	.172	-.121
Info Technology (1)	.565	.092	.024	.114
Online Course (1)	.512	.034	-.056	-.174
Job-Work (2)	-.052	.695	.141	.099
Time-Schedule (2)	.237	.602	.025	-.277
Faculty Negative (3)	.390	-.471	.459	.037
Federal Service (3)	-.369	-.168	.442	-.356
Financial (3)	.046	.007	.226	-.059
Personal-Other	.032	-.254	-.811	-.196
Family (4)	-.069	.063	-.081	.687
Health (4)	-.042	-.133	.031	.583

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Labeled Component View of Both Terms Combined (n=1,295)



Section 7

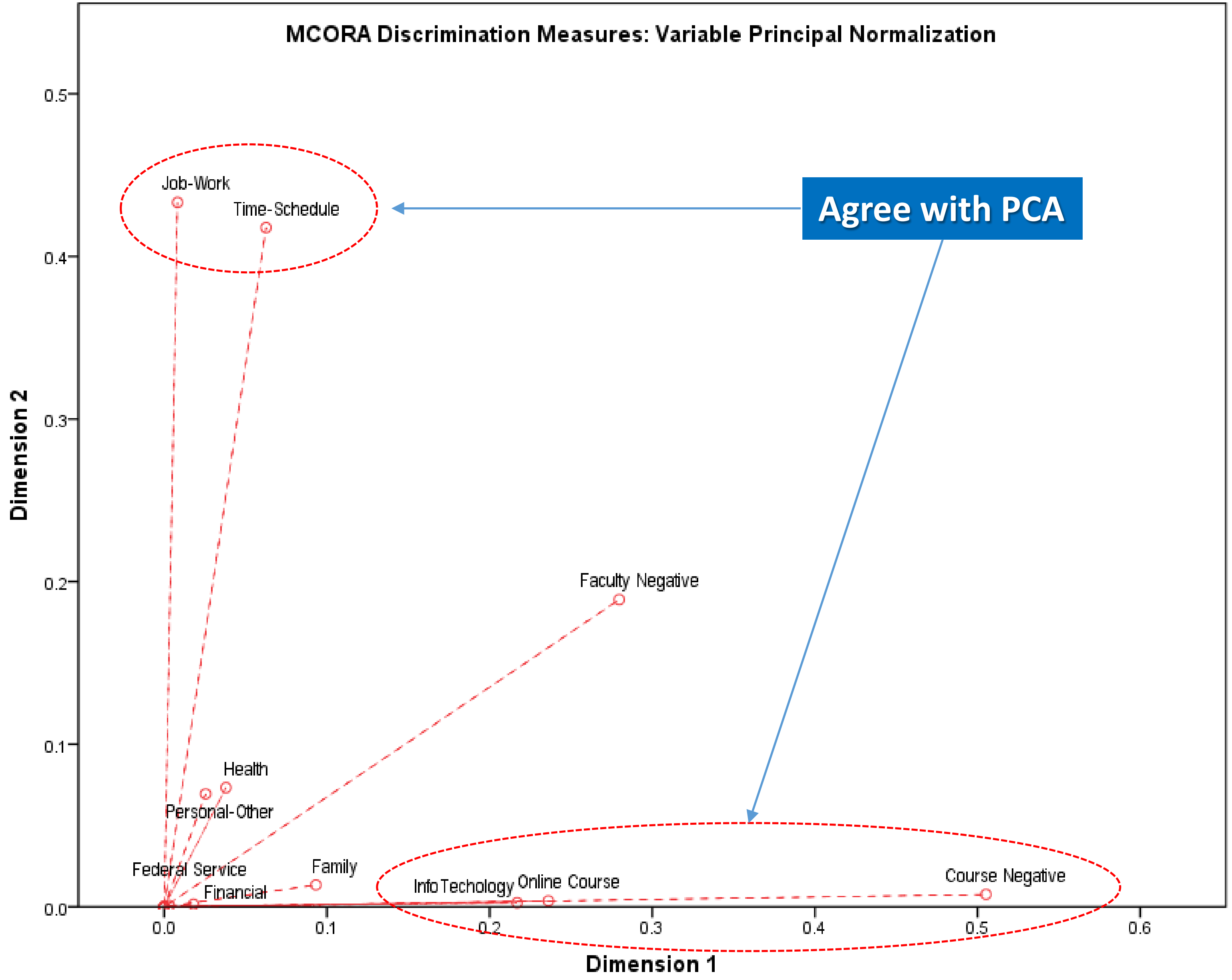
Quantitative Analysis 3: Multiple Correspondence Analysis (MCORA)

Multiple Correspondence Analysis (MCORA)

- Exploratory technique to analyze cross-classifications of two or more categorical variables in multi-way frequency table
- Aim is to transform a table of numbers into a plot of points in a small number of—usually two—dimensions
- Also called homogeneity analysis, a technique used to find optimal categorical quantifications by separating categories from each other as much as possible
- Objects in the same category are plotted close to each other and objects in different categories are plotted as far apart

*See Bartholomew et al., 2008

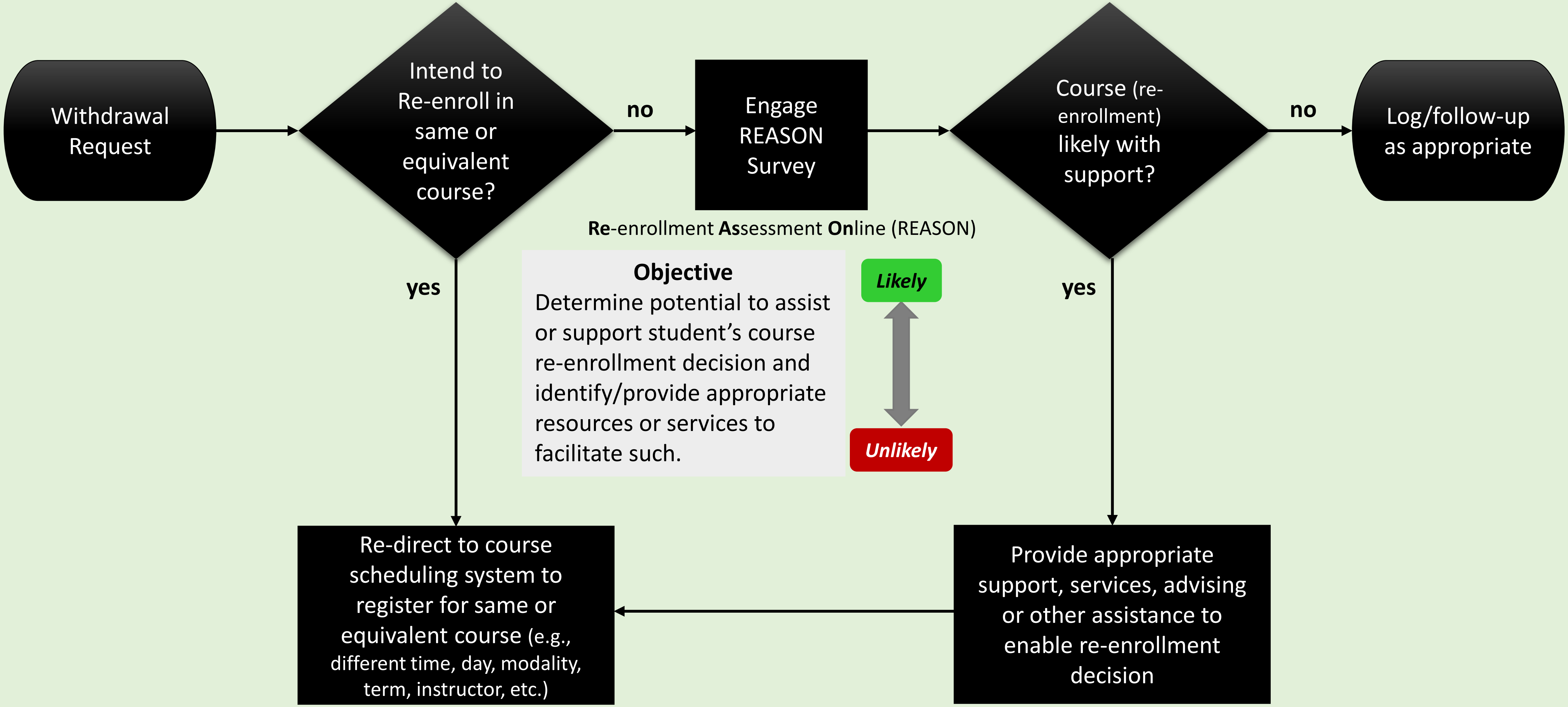
Multiple Correspondence Analysis Results



Section 8

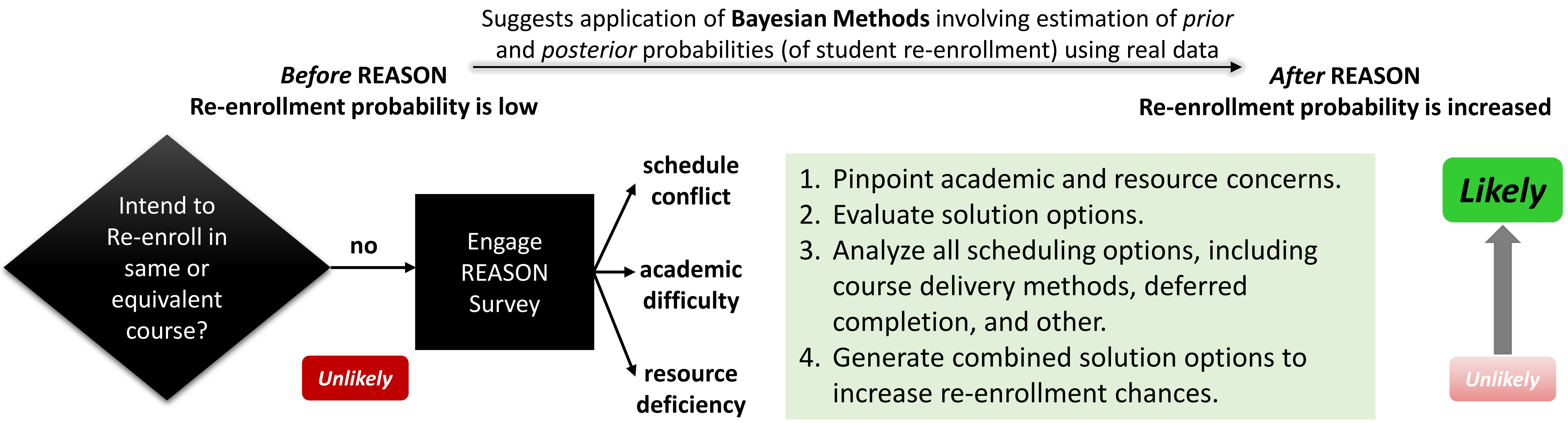
So What? Practical Application: The REASON (assessment)

Re-enrollment Assessment Online (REASON)



Re-enrollment Assessment Online (REASON)

- Purpose: given student's stated intention to not enroll in the same or equivalent course (i.e., course re-enrollment likelihood is low), use REASON to further assess support options, resources, or strategies that would *increase* course re-enrollment likelihood
  - STUDENT A withdrew from a given course and has indicated their intention to not re-enroll. REASON would gather further information from the student and use that information to offer support, options, and solutions to increase course re-enrollment probability.



Questions and Close

- Thank you for attending!
- Please feel free to contact me with follow-up questions and discussion or connect via LinkedIn

gmichals@fscj.edu

Selected References

Aggarwal, C. (2012). Mining text data. New York: Springer.

Angus, D., Rintel, S., & Wiles, J. (2013). Making sense of big text: a visual-first approach for analysing text data using Leximancer and Discursis. *International Journal of Social Research Methodology*, *16*(3), 261-267.doi:10.1080/13645579.2013.774186

Chiang, R. H. L., Goes, P., & Stohr, E. A. (2012). Business Intelligence and Analytics Education, and Program Development. *ACM Transactions on Management Information Systems, 3*(3), 1-13.doi:10.1145/2361256

Dunne, C., Shneiderman, B., Gove, R., Klavans, J., & Dorr, B. (2012). Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. *Journal of the American Society for Information Science and Technology, 63*(12), 2351-2369.

Fisher, I. E., Garnsey, M. R., Goel, S., & Tam, K. (2010). The Role of Text Analytics and Information Retrieval in the Accounting Domain. *Journal of Emerging Technologies in Accounting*, *7*(1), 1-24.doi:10.2308/jeta.2010.7.1.1

Kent, E. L. (2014). Text analytics - techniques, language and opportunity. *Business Information Review, 31*(1), 50-53.doi:10.1177/0266382114529837

Lim, E.-P., Chen, H., & Chen, G. (2013). Business Intelligence and Analytics. *ACM Transactions on Management Information Systems, 3*(4), 1-10.doi:10.1145/2407740

Lustigova, Z., & Brom, P. (2014). Educational Datamining in Virtual Learning Environments. *International Journal of Advanced Corporate Learning (iJAC)*, *7*(1), 39.doi:10.3991/ijac.v7i1.3557

Markham, S. K., Kowolenko, M., & Michaelis, T. L. (2015). Unstructured Text Analytics to Support New Product Development Decisions. *Research-Technology Management, 58*(2), 30-38.

Michalski, G. V. (2014). In their own words: A text analytics investigation of college course attrition. *Community college Journal of Research and Practice, 38*(9), 811-826. doi: 10.1080/10668926.2012.720865

Miner, G. [et al] (2012). Practical text mining and statistical analysis for non-structured text data applications. Waltham, MA: Academic Press.

Singh, P. D., & Raghuvanshi, J. (2012). Rising of Text Mining Technique: As Unforeseen-part of Data Mining. *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), 1*(3), 139-144.

Schmidt, M. (2010). Quantification of transcripts from depth interviews, open ended responses and focus groups: Challenges, accomplishments, new applications and perspectives for market research. *International Journal of Market Research, 52*(4), 483.doi:10.2501/S1470785309201417

Selected Presenter Works

Michalski, G. V. (2014). In Their Own Words: A Text Analytics Investigation of College Course Attrition. *Community College Journal of Research and Practice, 38*, 9, 811-826.

Michalski, G. V. (2014). *Student analytics: Data driven decision making to increase student success*. Workshop delivered at the 25th International Conference on College Teaching and Learning, Florida State College at Jacksonville. March 25, 2014. Jacksonville, FL (USA).

Michalski, G. V. (2013). *Using predictive analytics to understand and foster degree completion*. Paper presented at the 53rd International AIR Forum. May 21, 2013. Long Beach, CA (USA): Association for Institutional Research.

Michalski, G. V. (2012). *Using predictive analytics to support decision making and improve academic results*. National Alliance of Community and Technical Colleges Conference (presentation). July 13, 2012. Minneapolis, MN.

Michalski, G. V. (2011). *Complementing the Numbers: A Text Mining Analysis of College Course Withdrawals*. Paper presented at the 51st International AIR Forum. May 24, 2011. Toronto, Ontario (Canada): Association for Institutional Research.

Michalski, G. V., & Cousins, J. B. (2000). Differences in stakeholder perceptions about training evaluation: a concept mapping/pattern matching investigation. *Evaluation and Program Planning, 23*, 2, 211-230.

Michalski, G. V., & Cousins, J. B. (2001). Multiple Perspectives on Training Evaluation: Probing Stakeholder Perceptions in a Global Network Development Firm. *American Journal of Evaluation*, 22, 1, 37-53