UNIVERSITY OF CENTRAL FLORIDA
University Analysis and Planning Support

UNIVERSITY OF CENTRAL FLORIDA
University Analysis and Planning Support

Using Cluster Analysis to Identify University Benchmarking Peers

Fourth Annual Business Intelligence Symposium

Shuxin Li

Sandra Archer

Dr. Robert L. Armacost

April 11, 2008

Overview

- Motivation for identifying benchmarking peers
- Process used in 2001
- Current status
- Details of our initial “study” phase
- What’s next

Background and Motivation

- 2002-07 strategic plan development in 2001
  - UCF vision established: "The University of Central Florida will be the *nation's leading metropolitan research university* recognized for its intellectual, cultural, technological, and professional contributions and renowned for its outstanding programs and partnerships."
- Led to the basic questions:
  - What is a Metropolitan Research University (MRU)?
  - To which universities do we compare UCF?
  - How do we determine these peers?

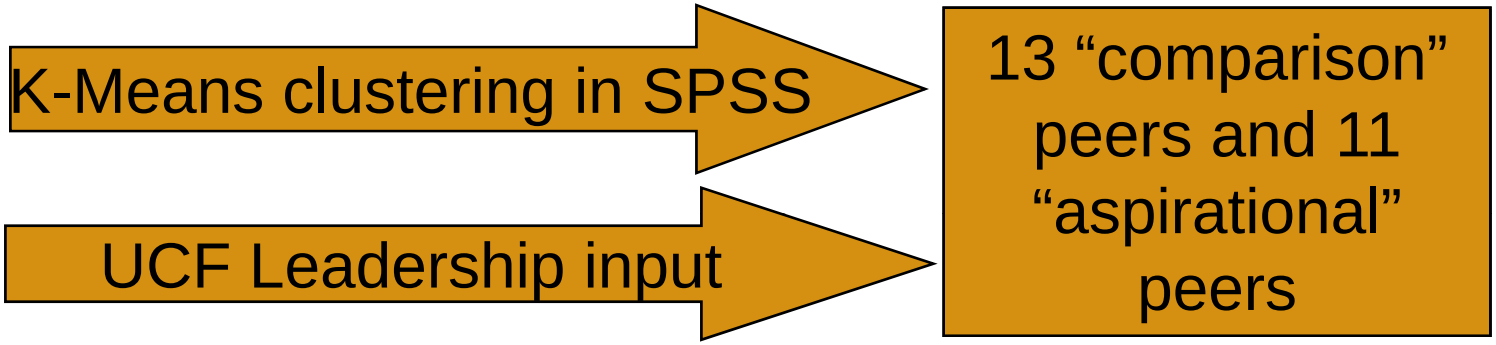
Institutions Initially Considered in 2001

- Prior years' peers (10)
- State University System (SUS) institutions (10)
- Collegiate Results Instruments (CRI) peer institutions (8)
- US News & World Report 3rd and 4th Tier (32)
- Members of the Coalition of Urban and Metropolitan Universities (14)
- Institutions of interest to UCF leadership

Variables Considered in 2001

- MRU Variables
  - Characteristics of the surrounding metropolitan region
  - Financial characteristics, including research expenditures
- Student characteristics
- Faculty characteristics

Method



- UCF 2002-2007 strategic plan
  - Peer list was held static over the course of the plan
  - Provided comparisons to help measure progress toward the plan

It is now 2008…

- UCF 2002-07 Strategic Planning timeframe has ended
- Time to repeat the exercise
  - We do not know if UCF's vision will change (MRU?)
- Much has changed since 2001:
  - UCF enrollment growth: 36K → 48.7K
  - Research funding: $64.9M → $121.4M
  - New medical college
- Is the 2001 peer list still relevant?
- We now have SAS available to us

The Process of Updating Our Peer List

	Step 1	Step 2	Step 3	Step 4
Clustering method	Same	New	Final Model	Final Model
Data values	2001 Data	2001 Data	Update	2006 Data
Variables to consider	MRU	MRU	MRU	New
Objective	Duplicate the 2001 process and explain results	Determine appropriate clustering methods	Update peer list using 2006 data	Determine new benchmarking peers

"study" phase in progress

"Study" Phase

- Review 2001 process and explain differences
- Explore clustering methods

Initial Institutions - 63

- ARIZONA STATE UNIVERSITY-MAIN CAMPUS
- BOSTON UNIVERSITY
- CLEVELAND STATE UNIVERSITY
- SAN DIEGO STATE UNIVERSITY
- IOWA STATE UNIVERSITY
- OHIO STATE UNIVERSITY-MAIN CAMPUS
- FLORIDA ATLANTIC UNIVERSITY-BOCA RATON
- FLORIDA INTERNATIONAL UNIVERSITY
- FLORIDA STATE UNIVERSITY
- INDIANA UNIVERSITY-BLOOMINGTON
- UNIVERSITY OF FLORIDA
- UNIVERSITY OF ILLINOIS AT CHICAGO
- UNIVERSITY OF VIRGINIA-MAIN CAMPUS
- UNIVERSITY OF WASHINGTON-SEATTLE CAMPUS
- UNIVERSITY OF WISCONSIN-MADISON
- WESTERN MICHIGAN UNIVERSITY
- WRIGHT STATE UNIVERSITY-MAIN CAMPUS
- etc…….

Final Variables Used in 2001

- MRU Variables
  - Expenditures per student
  - NSF founded research and development expenditures
  - Grant a medical degree (Y/N)
  - Value of endowment assets-market
  - Number of programs awarding Bachelor's, Master's & Doctoral degrees
- Student characteristics
  - Baccalaureate degrees awarded / undergraduate enrollment ratio
  - Graduate degrees awarded / graduate enrollment ratio
  - Full time / part time student ratio
  - Undergraduate / graduate student ratio
- Faculty characteristics
  - Number of full time faculty
  - Publications per faculty ratio

Methodology – Cluster Analysis

- Cluster analysis = exploratory data analysis tool for solving segmentation problems
  - Objective: classify objects (in our case universities) into clusters
  - Homogeneity ***within*** a cluster and heterogeneity ***between*** clusters is statistically significant
- Two main clustering methods in SAS
  - PROC FASTCLUS
  - PROC CLUSTER

PROC FASTCLUS

- "Distance-based disjoint" clustering method
- Steps:
  - Arbitrarily choose k universities as the "seeds"
  - Assign all other universities to their closest "seed"
  - Update the cluster mean and redefine the center
- Advantages:
  - Fast - only requires two or three passes over the data set
  - Can handle large datasets
- Disadvantages:
  - The number of clusters must be specified (potential for user bias)
  - Results may vary depending on initial random seeds
- Equivalent to SPSS k-means method (used in 2001)

PROC CLUSTER

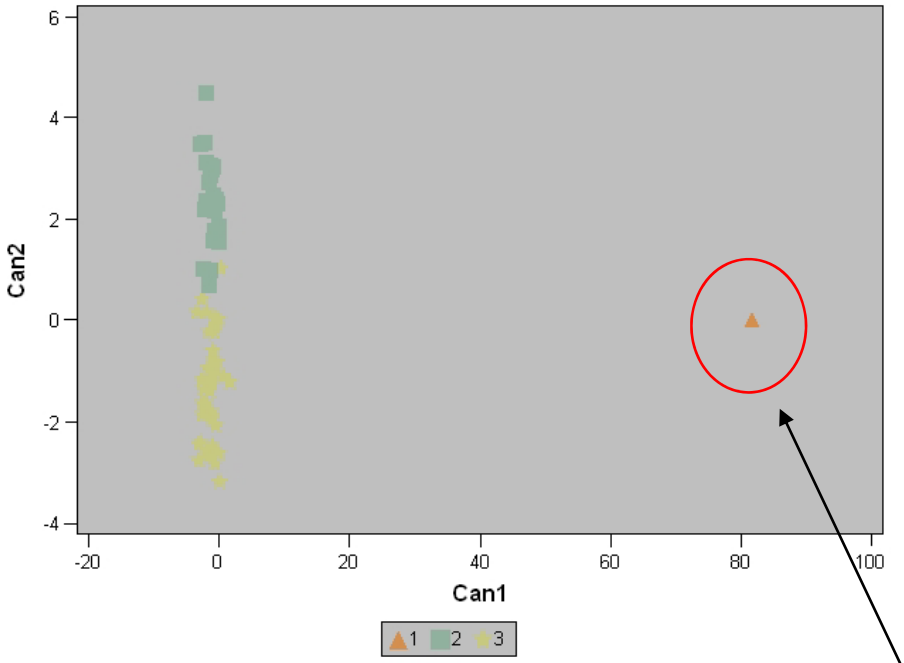
- "Hierarchical" clustering method
- Steps:
  - Each observation begins in a cluster by itself
  - The two closest clusters are merged to form a new cluster
  - Merging continues until only one cluster is left
- Advantages:
  - Can be visualized in a tree structure
  - Details of the clustering process is provided
- Disadvantages:
  - Relatively slower
  - Not suitable for larger datasets
  - Imputation of missing values required

Process Overview

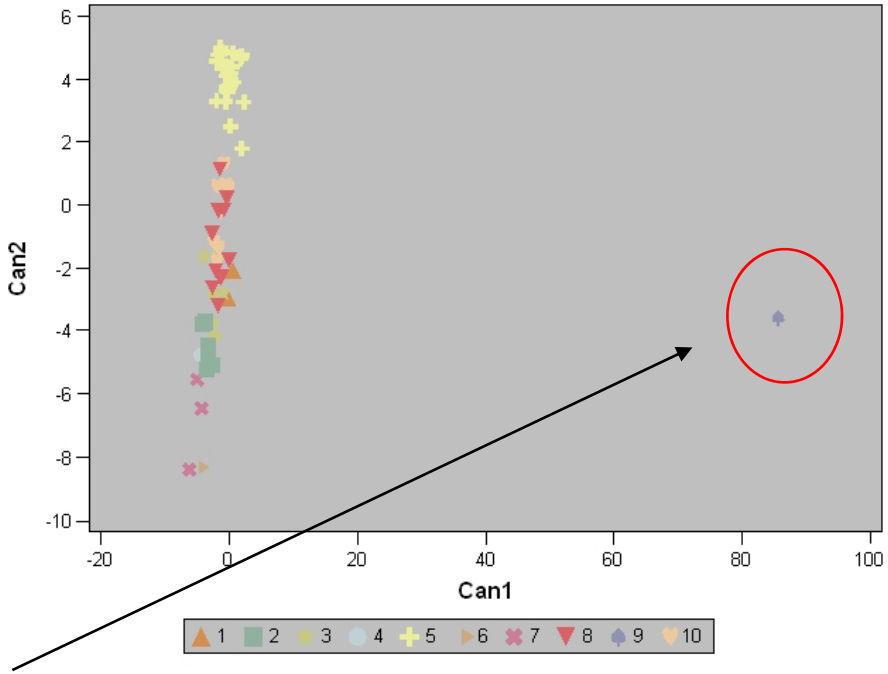
- Standardized variables – PROC STANDARD
- Ran PROC FASTCLUS
  - Number of clusters from 2 to 10
  - Reduced initial set of 63 institutions to 25 institutions
- Ran PROC CLUSTER
  - Used PROC TREE to display hierarchical clusters
- Compared results

PROC FASTCLUS: 63 Institutions

Number of clusters=3



Number of clusters=10



Outlier

PROC FASTCLUS

Initial list: 63
Clusters = 10

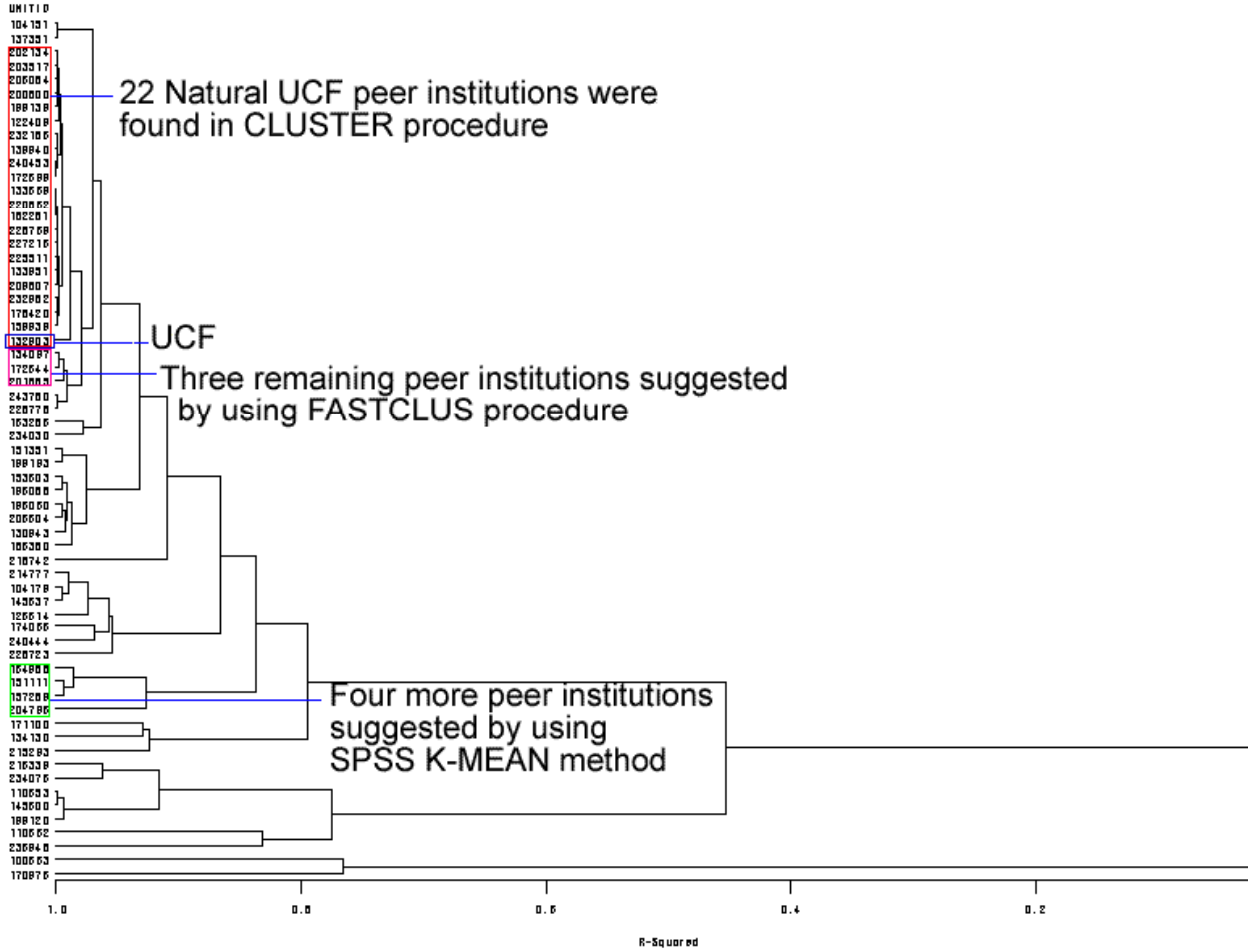
- 25 universities or institutions are recommended for further analysis:

ARIZONA STATE UNIVERSITY-MAIN CAMPUS
CLEVELAND STATE UNIVERSITY
FLORIDA ATLANTIC UNIVERSITY-BOCA RATON
FLORIDA INTERNATIONAL UNIVERSITY
GEORGE MASON UNIVERSITY
GEORGIA STATE UNIVERSITY
KENT STATE UNIVERSITY-MAIN CAMPUS
OLD DOMINION UNIVERSITY
PORTLAND STATE UNIVERSITY
SAN DIEGO STATE UNIVERSITY
SUNY AT ALBANY
THE UNIVERSITY OF TEXAS AT ARLINGTON
UNIVERSITY OF AKRON MAIN CAMPUS
UNIVERSITY OF CENTRAL FLORIDA
UNIVERSITY OF DELAWARE
UNIVERSITY OF HOUSTON-UNIVERSITY PARK
UNIVERSITY OF MEMPHIS
UNIVERSITY OF MISSOURI-ST LOUIS
UNIVERSITY OF NEVADA-LAS VEGAS
UNIVERSITY OF NEW ORLEANS
UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE
UNIVERSITY OF NORTH TEXAS
UNIVERSITY OF TOLEDO
UNIVERSITY OF WISCONSIN-MILWAUKEE
WESTERN MICHIGAN UNIVERSITY

PROC CLUSTER

Initial list: 63
Clusters = 10

- Only 22 universities or institutions are recommended for further analysis:



Compare Results

Initial list: 63
Clusters = 10

SPSS K-means results (29)

ARIZONA STATE UNIVERSITY-MAIN CAMPUS
CLEVELAND STATE UNIVERSITY
FLORIDA ATLANTIC UNIVERSITY-BOCA RATON
FLORIDA INTERNATIONAL UNIVERSITY
GEORGE MASON UNIVERSITY
GEORGIA STATE UNIVERSITY
INDIANA UNIVERSITY-PURDUE UNIVERSITY-INDIAN
KENT STATE UNIVERSITY-MAIN CAMPUS
OLD DOMINION UNIVERSITY
PORTLAND STATE UNIVERSITY
SAN DIEGO STATE UNIVERSITY
SUNY AT ALBANY
THE UNIVERSITY OF TEXAS AT ARLINGTON
UNIVERSITY OF AKRON MAIN CAMPUS
UNIVERSITY OF CENTRAL FLORIDA
UNIVERSITY OF HOUSTON-UNIVERSITY PARK
UNIVERSITY OF LOUISVILLE
UNIVERSITY OF MEMPHIS
UNIVERSITY OF MISSOURI-ST LOUIS
UNIVERSITY OF NEVADA-LAS VEGAS
UNIVERSITY OF NEW ORLEANS
UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE
UNIVERSITY OF NORTH TEXAS
UNIVERSITY OF SOUTH FLORIDA
UNIVERSITY OF TOLEDO
UNIVERSITY OF WISCONSIN-MILWAUKEE
VIRGINIA COMMONWEALTH UNIVERSITY
WESTERN MICHIGAN UNIVERSITY
WRIGHT STATE UNIVERSITY-MAIN CAMPUS

PROC FASTCLUS results (25)

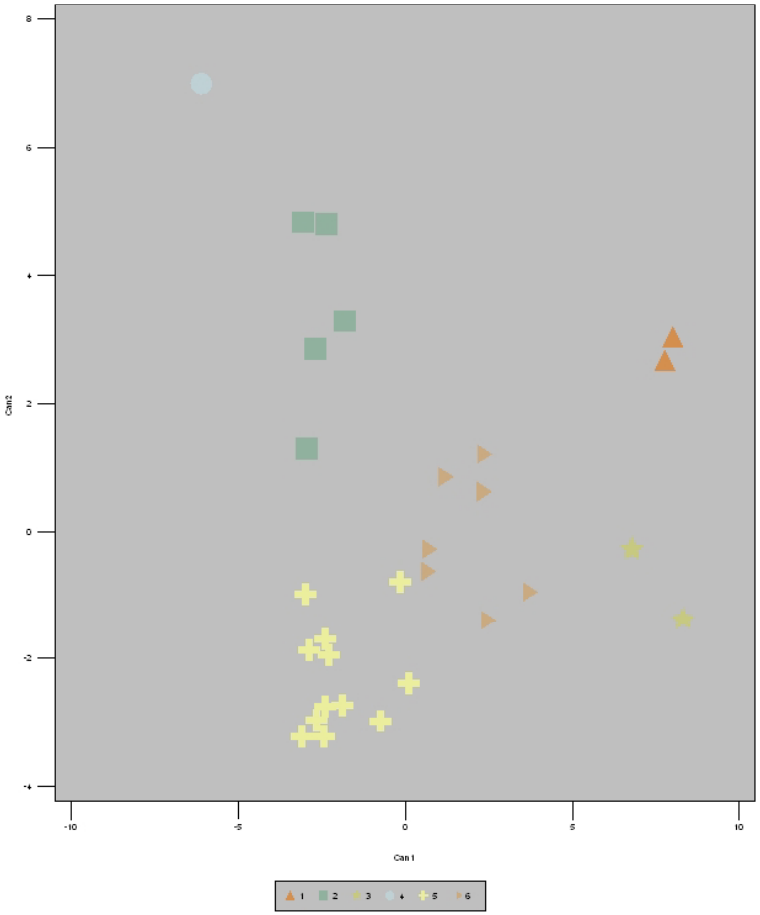
ARIZONA STATE UNIVERSITY-MAIN CAMPUS
CLEVELAND STATE UNIVERSITY
FLORIDA ATLANTIC UNIVERSITY-BOCA RATON
FLORIDA INTERNATIONAL UNIVERSITY
GEORGE MASON UNIVERSITY
GEORGIA STATE UNIVERSITY
KENT STATE UNIVERSITY-MAIN CAMPUS
OLD DOMINION UNIVERSITY
PORTLAND STATE UNIVERSITY
SAN DIEGO STATE UNIVERSITY
SUNY AT ALBANY
THE UNIVERSITY OF TEXAS AT ARLINGTON
UNIVERSITY OF AKRON MAIN CAMPUS
UNIVERSITY OF CENTRAL FLORIDA
UNIVERSITY OF HOUSTON-UNIVERSITY PARK
UNIVERSITY OF MEMPHIS
UNIVERSITY OF MISSOURI-ST LOUIS
UNIVERSITY OF NEVADA-LAS VEGAS
UNIVERSITY OF NEW ORLEANS
UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE
UNIVERSITY OF NORTH TEXAS
UNIVERSITY OF SOUTH FLORIDA
UNIVERSITY OF TOLEDO
UNIVERSITY OF WISCONSIN-MILWAUKEE
WESTERN MICHIGAN UNIVERSITY

PROC CLUSTER results (22)

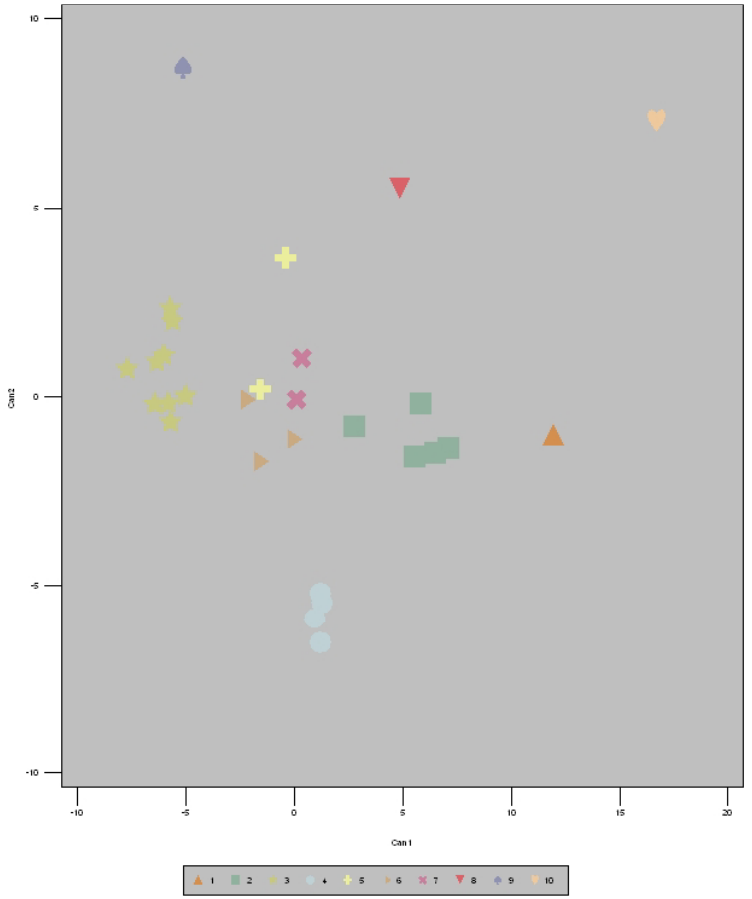
CLEVELAND STATE UNIVERSITY
FLORIDA ATLANTIC UNIVERSITY-BOCA RATON
FLORIDA INTERNATIONAL UNIVERSITY
GEORGE MASON UNIVERSITY
GEORGIA STATE UNIVERSITY
KENT STATE UNIVERSITY-MAIN CAMPUS
OLD DOMINION UNIVERSITY
PORTLAND STATE UNIVERSITY
SAN DIEGO STATE UNIVERSITY
THE UNIVERSITY OF TEXAS AT ARLINGTON
UNIVERSITY OF AKRON MAIN CAMPUS
UNIVERSITY OF CENTRAL FLORIDA
UNIVERSITY OF HOUSTON-UNIVERSITY PARK
UNIVERSITY OF MEMPHIS
UNIVERSITY OF MISSOURI-ST LOUIS
UNIVERSITY OF NEVADA-LAS VEGAS
UNIVERSITY OF NEW ORLEANS
UNIVERSITY OF NORTH CAROLINA AT CHARLOTT
UNIVERSITY OF NORTH TEXAS
UNIVERSITY OF TOLEDO
UNIVERSITY OF WISCONSIN-MILWAUKEE
WESTERN MICHIGAN UNIVERSITY

PROC FASTCLUS: 29 institutions

Number of clusters=6



Number of clusters=10



PROC FASTCLUS

Initial list: 29
Clusters = 10

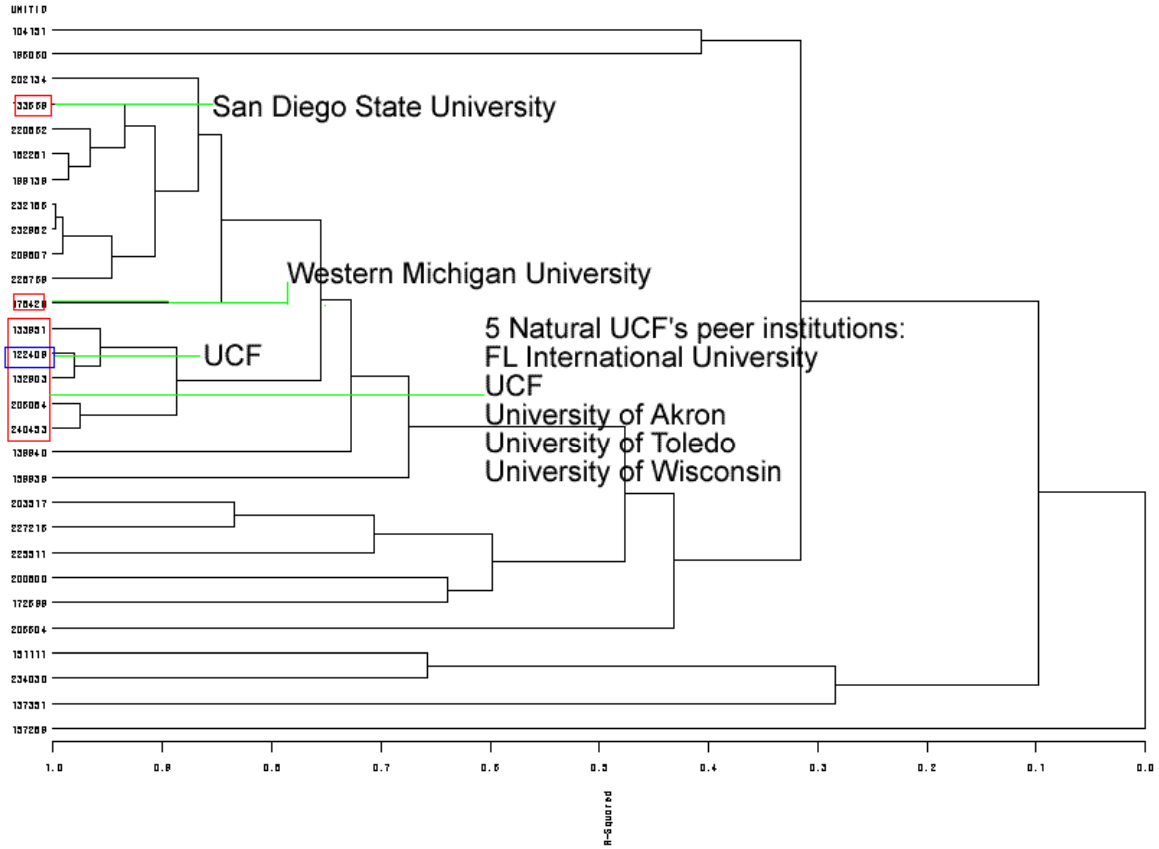
- 5 universities or institutions are found to be UCF's peers in 2001:

FLORIDA INTERNATIONAL UNIVERSITY
UNIVERSITY OF AKRON MAIN CAMPUS
UNIVERSITY OF CENTRAL FLORIDA
UNIVERSITY OF WISCONSIN-MILWAUKEE
UNIVERSITY OF TOLEDO

PROC CLUSTER

Initial list: 29
Clusters = 10

- 5 natural universities or institutions are found to be UCF's peers in 2001:



Compare Results

Initial list: 29
Clusters = 10

SPSS K-means results (7)	PROC FASTCLUS results (5)	PROC CLUSTER results (5)
FLORIDA INTERNATIONAL UNIVERSITY	FLORIDA INTERNATIONAL UNIVERSITY	FLORIDA INTERNATIONAL UNIVERSITY
UNIVERSITY OF AKRON MAIN CAMPUS	UNIVERSITY OF AKRON MAIN CAMPUS	UNIVERSITY OF AKRON MAIN CAMPUS
UNIVERSITY OF CENTRAL FLORIDA	UNIVERSITY OF CENTRAL FLORIDA	UNIVERSITY OF CENTRAL FLORIDA
UNIVERSITY OF WISCONSIN-MILWAUKEE	UNIVERSITY OF WISCONSIN-MILWAUKEE	UNIVERSITY OF WISCONSIN-MILWAUKEE
UNIVERSITY OF TOLEDO	UNIVERSITY OF TOLEDO	UNIVERSITY OF TOLEDO
SAN DIEGO STATE UNIVERSITY		
WESTERN MICHIGAN UNIVERSITY		

- 2001 data

Conclusion

- Duplicated the 2001 process
  - Some differences in the results
- Explored new clustering methods
  - Found that using more than one clustering method provides validation for the results
  - Gained an understanding of which process is most appropriate for that situation

What's Next?

- Finalize modeling methods
- Update data values for 2006
- Consider new variables
  - MRU variables?
  - Dependent on UCF's strategic direction

Questions

Sandra Archer, Director
archer@mail.ucf.edu

Shuxin Li, Graduate Research Assistant
shuxinli@mail.ucf.edu

Dr. Robert Armacost
armacost@mail.ucf.edu

UCF University Analysis and Planning Support

Presentation will be available online: http://uaps.ucf.edu

407-882-0285