air forum

ORLANDO, FL
May 20-23, 2025
Pre-conference: May 19-20

Art or Science? Development of a Peer Group in IR Using Cluster Analysis

Sanjeewa Karunaratne, *MPA, MSR, LLM*

John D. Nugent, *Ph.D.*

Connecticut College, New London, CT

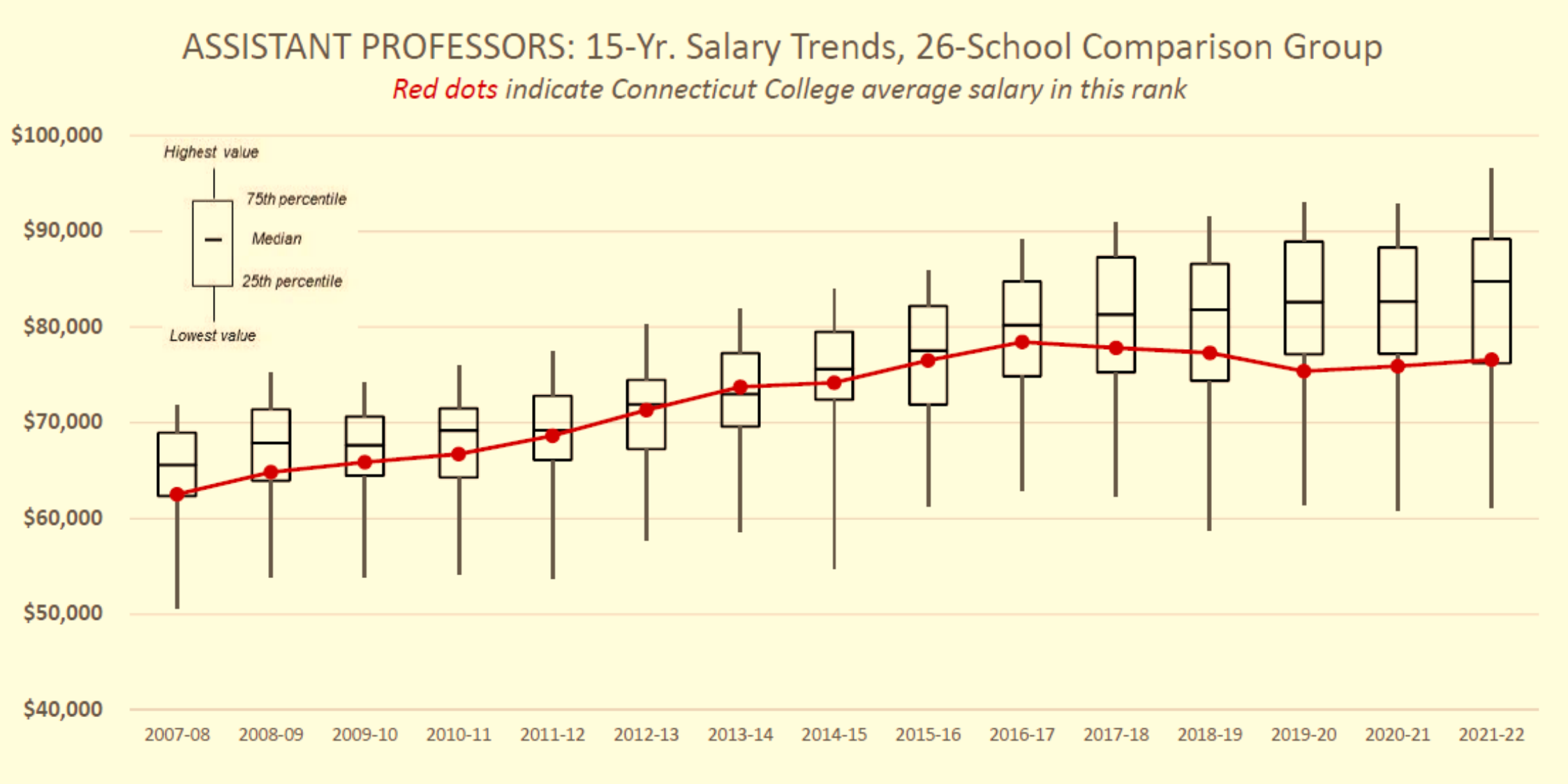
A Rewarding Day For Institutional Research

…Please note that as we move forward, the new peer group will be used for all comparison purposes….

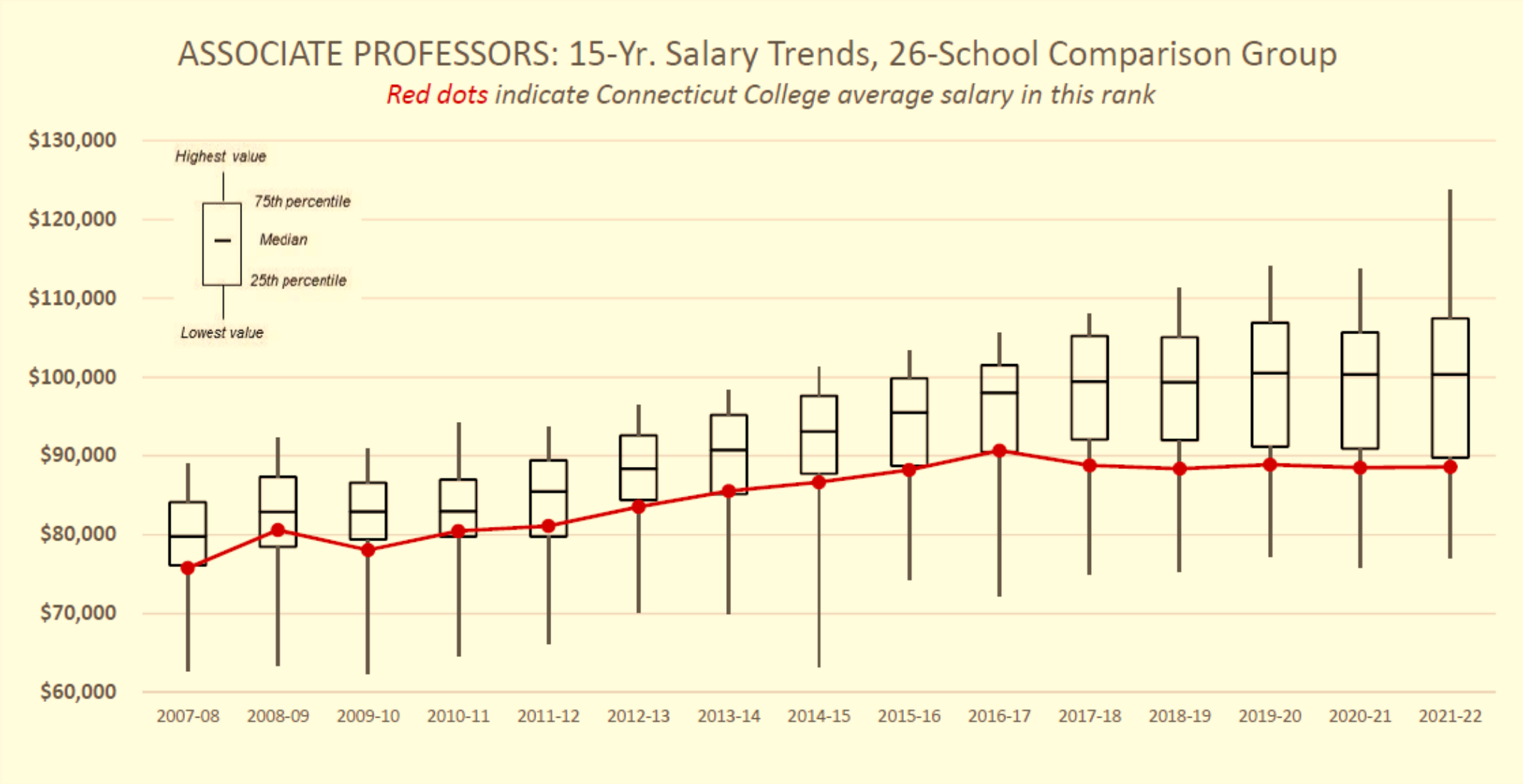
Sincerely,

Andrea E. Chapdelaine
President

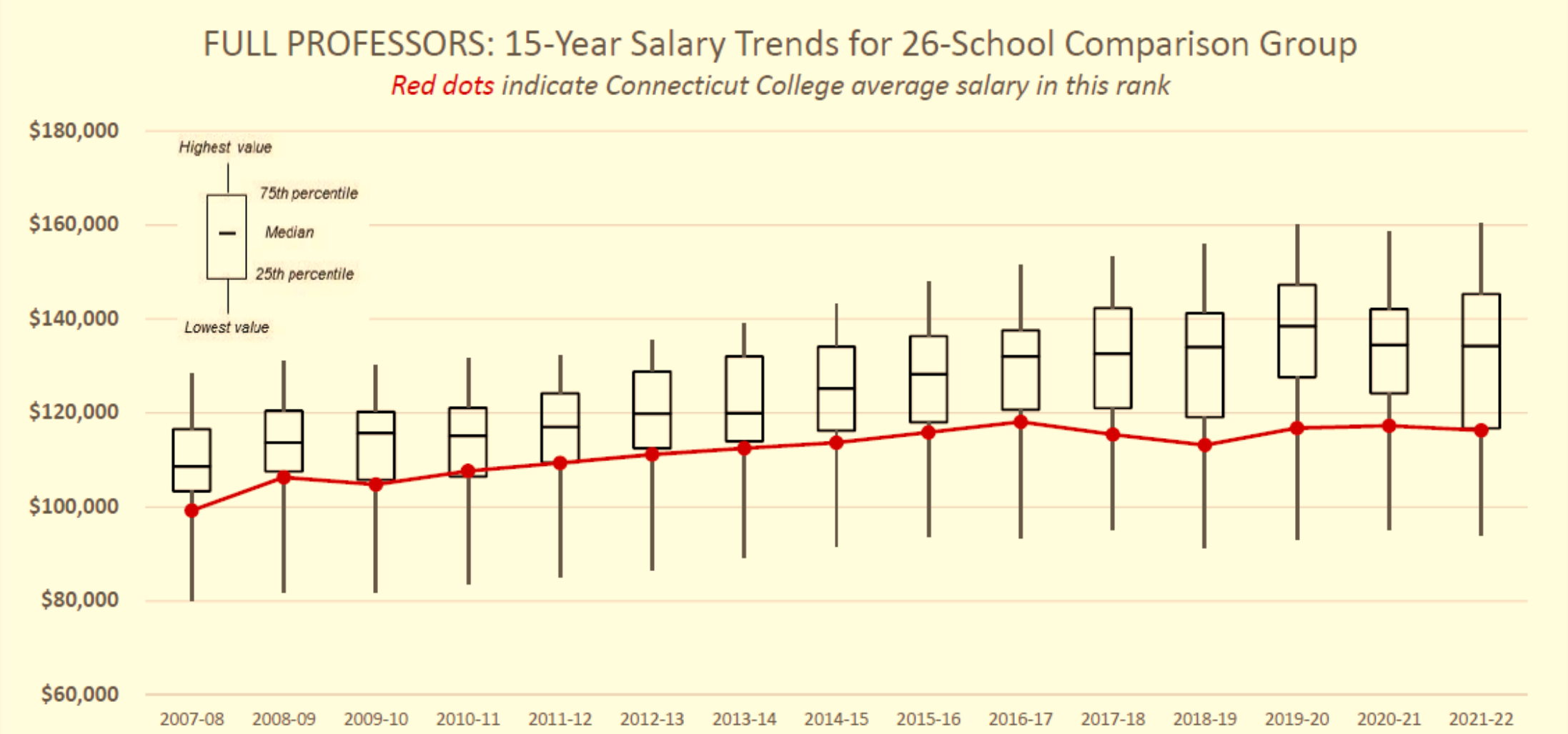
How Did We Get Here? Assistant



How Did We Get Here? Associate



How Did We Get Here? Professor



Possible Uses of a Peer Group

- Clarify an institution's market position in higher education
- Allow comparison of activities and outcomes to similar schools
- Set benchmarks for management decisions (e.g., faculty salaries, division budgets, perhaps, at the midpoint of the peer group)
- Use for peer group comparisons by the accreditor. (A peer institution is one that your institution considers comparable in terms of mission, scale, student population, or outcomes at this moment in time (NECHE))

Development of a Peer Group

- Part art, part science, part politics
- The IR Office had taken the approach of keeping a list of peer groups for different comparison purposes i.e., Human Resource, Faculty salaries
- Some schools in these groups look less “peer-y”, to Connecticut College in terms their resources (Smith, Bowdoin, Colby, Middlebury)
- There was demand for a single, unified, and more realistic peer group that can be used for many purposes including faculty salaries
- We decided to use cluster analysis to develop this new peer group

Questions…

- Who has used cluster analysis before?

- Any other practical uses of cluster analysis?

Our Approach

I. Case reduction

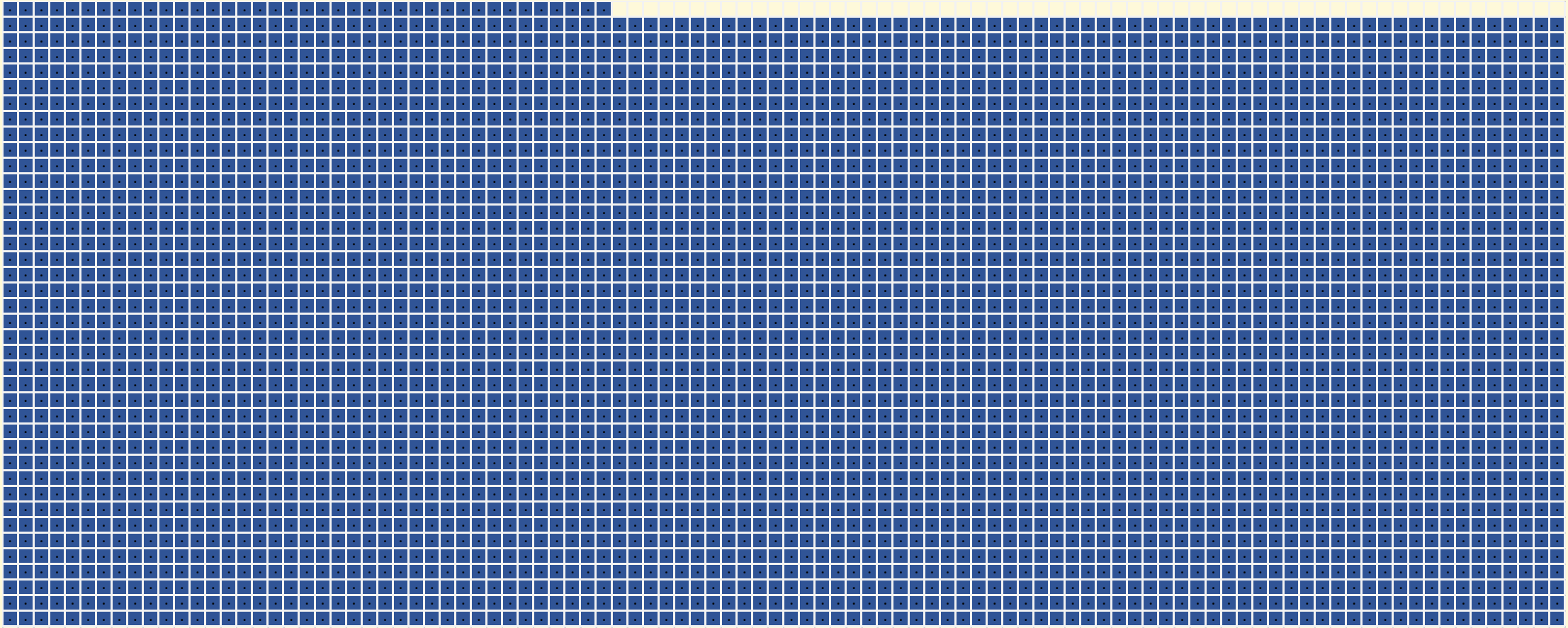
II. Variable reduction

III. Cluster analysis

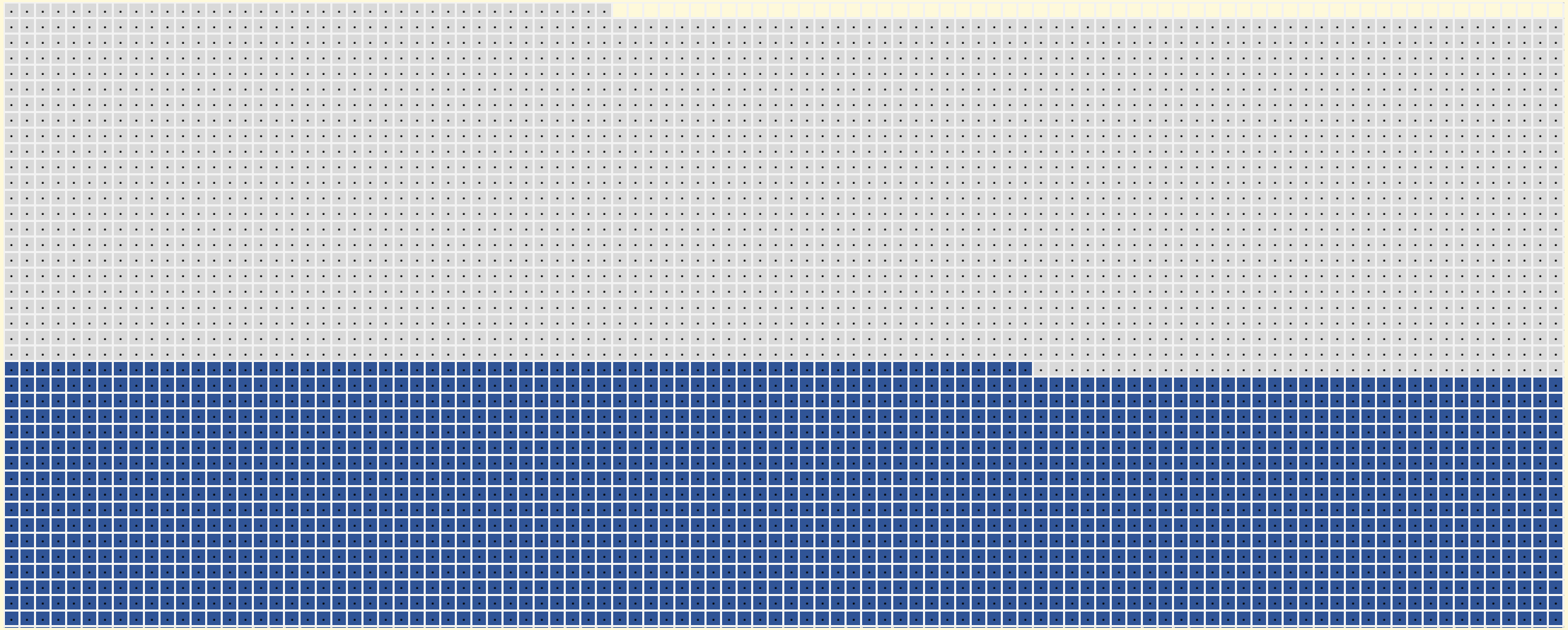
I. Case Reduction

- Cluster analysis can be susceptible to how the researcher go about doing it
- It can be swayed by outliers, number of variables and their correlations
- We started with all 3,939 schools in Carnegie Classification system and winnowed down to 105 on five factors:
  - **Control** (public, private non-profit, private for-profit)
  - **Basic categorization** (associates, baccalaureate, masters, doctoral, special focus)
  - **Size and setting** (2 yr./4 yr.; small/medium/large; how residential is it)
  - **Undergraduate profile** (inclusive/selective; full-time vs. part-time)
  - **Enrollment profile** (exclusively/very high/high/majority undergrad)
- The key is the reduction of unexplained variances

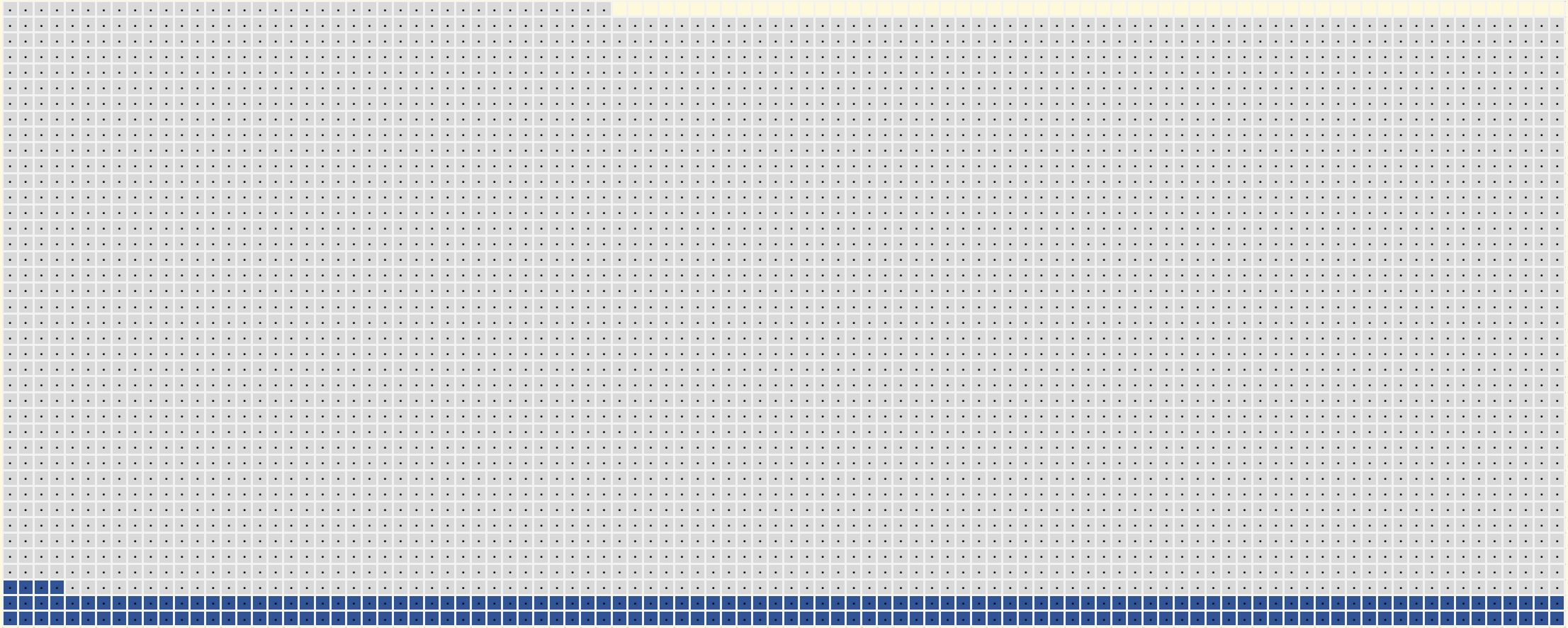
Starting With 3,939 U.S. Institutions…



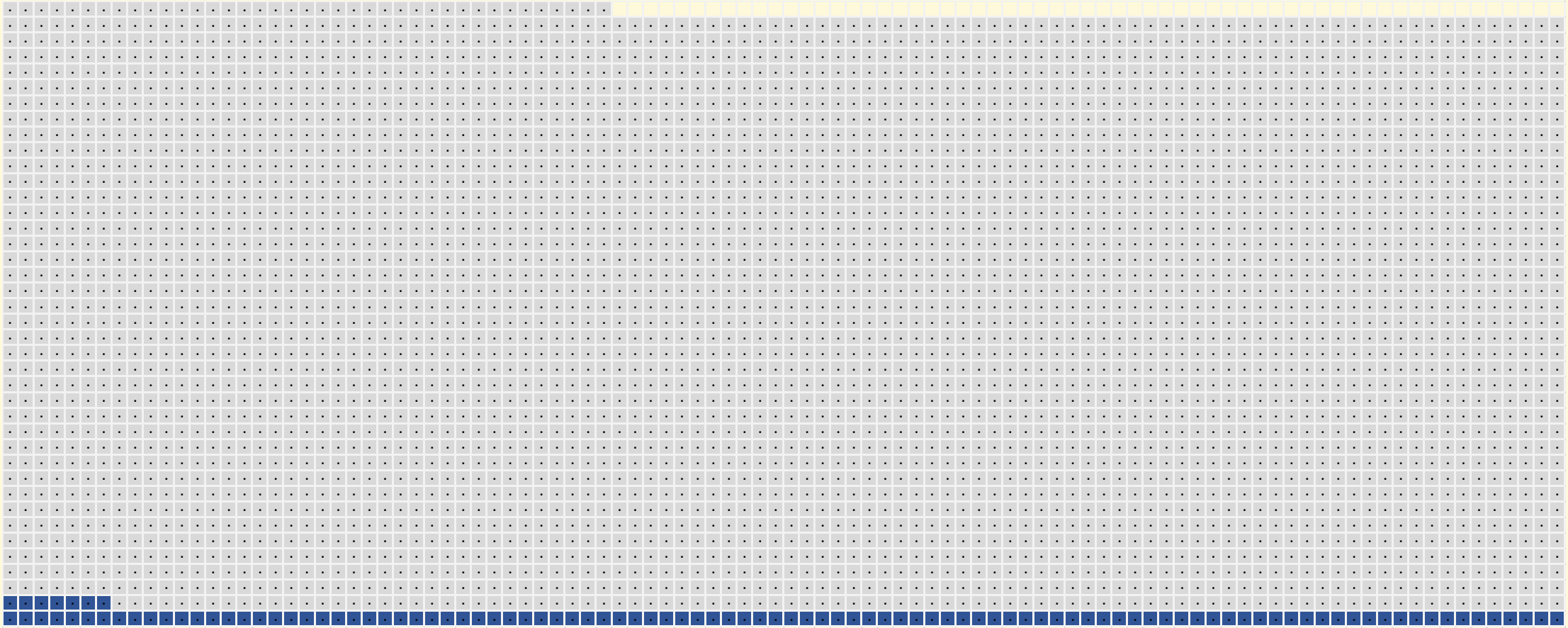
Eliminated All Publics And Private For-profits (-2,273)



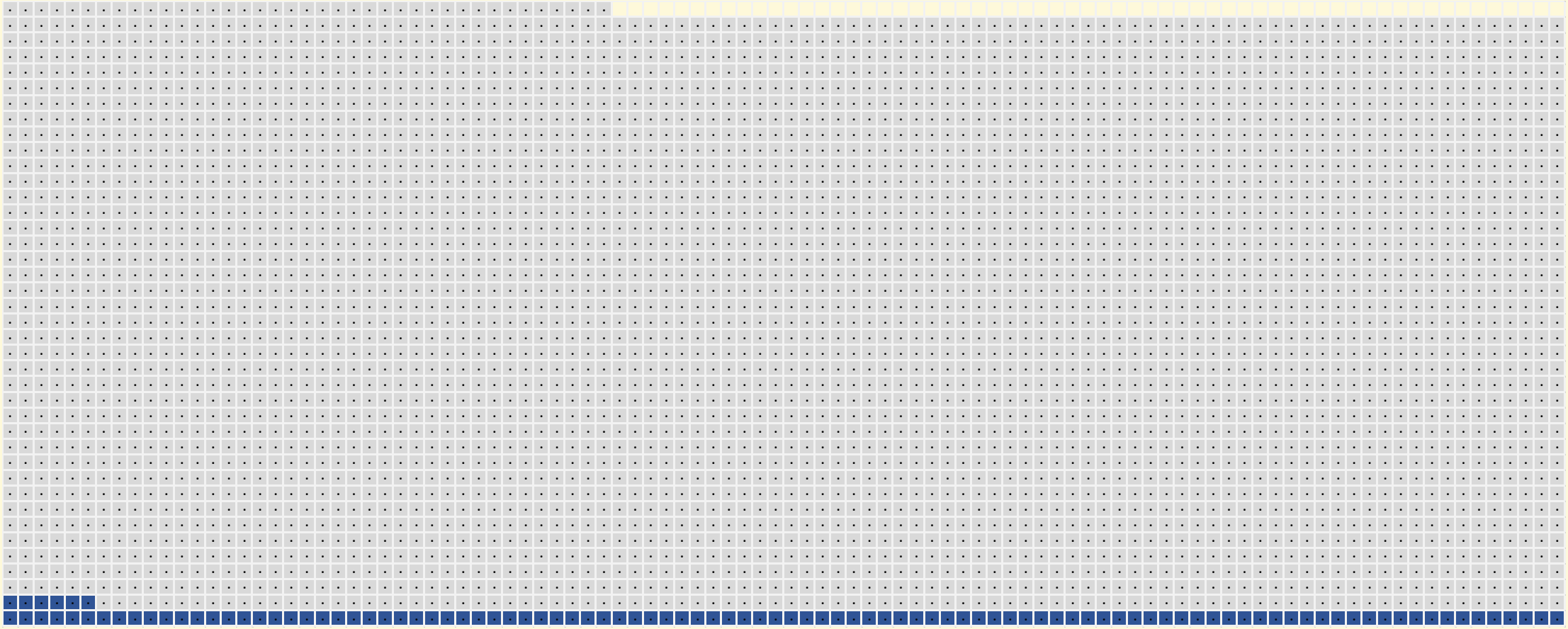
Eliminated All Non-bachelors Arts And Sciences (-1,462)



Kept Only Selective or More Selective And Fewer Transfer-ins (-26)



Kept Only Exclusively / Very High / High Undergrad (-1)

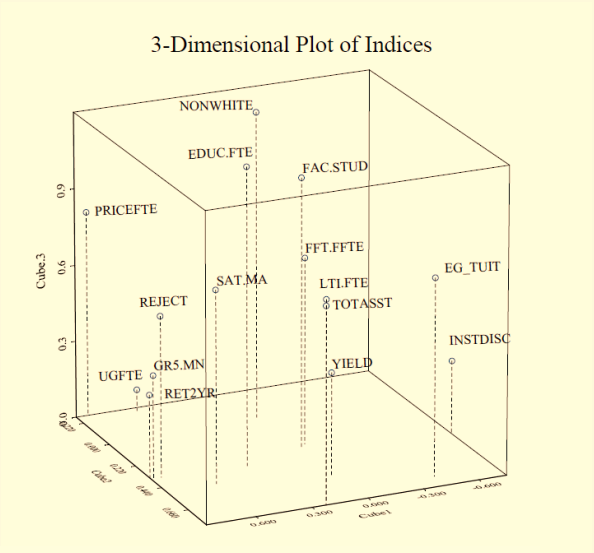


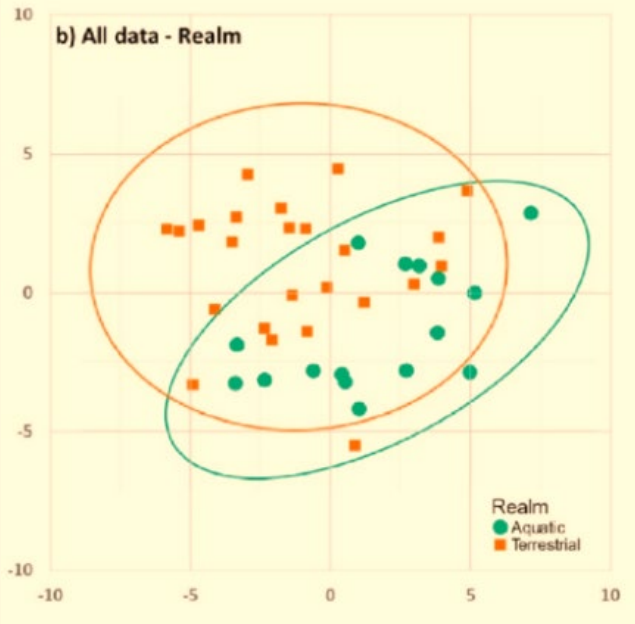
Variables We Looked At

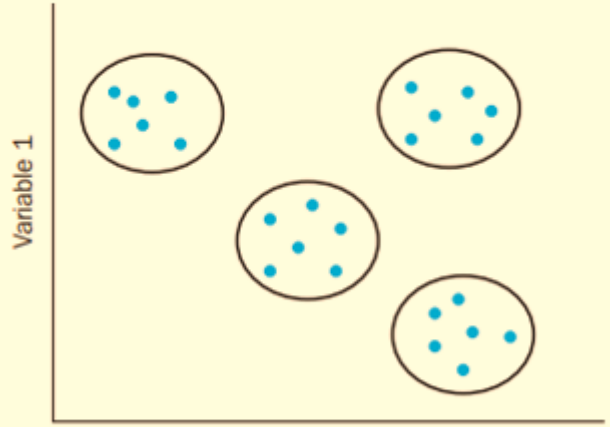
- Institutional characteristics: Carnegie Classification of Institutions of Higher Education
- Cost of living variation: Regional Price Parity by Metropolitan Statistical Area (or closest one to the campus)
- Endowment; endowment per student: FY2023 endowment data from NACUBO
- Student body size: Avg. of 3 years thru 2023 from National Center for Education Statistics (IPEDS)
- Faculty and staff size (FT, PT, total) (IPEDS)
- Student-faculty ratio; student-staff ratio: Avg. of 3 years thru 2023 (IPEDS)
- Admit rate and yield rate: Avg. of 3 years thru 2023 (IPEDS)
- Social mobility: Washington Monthly subrank
- Operating budget; operating budget per student: Avg. of 3 years thru 2023 (IPEDS)
- First-to-second-year retention rate: Avg. of 3 years thru 2023 (IPEDS)
- 6-year graduation rate: Avg. of 3 years thru 2023 (IPEDS)
- 2022-23 AAUP faculty salary data

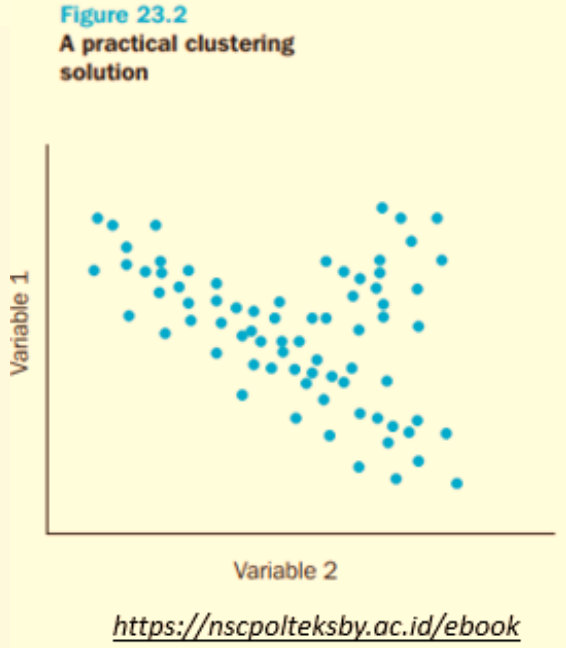
II. Variable Reduction

- Correlation matrix
- Factor analysis
- Regression
- Conceptual foundation





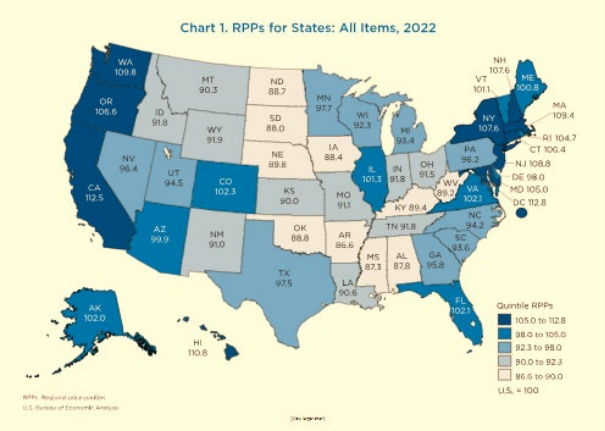




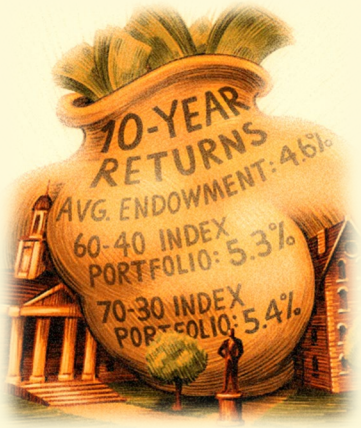
https://nscpolteksby.ac.id/ebook

Variable Reduction

- Regional price parity (RPP)



- Market value of endowment ($)



- Spending per student ($)

Normalization

- 3-year averages (to smooth out large variances)
- Z-scores (to streamline differences in scales)
- Missing values
- Outliers (boxplot and histogram)
- 79 schools left in the analysis after clean up
- Case and variable reduction complete

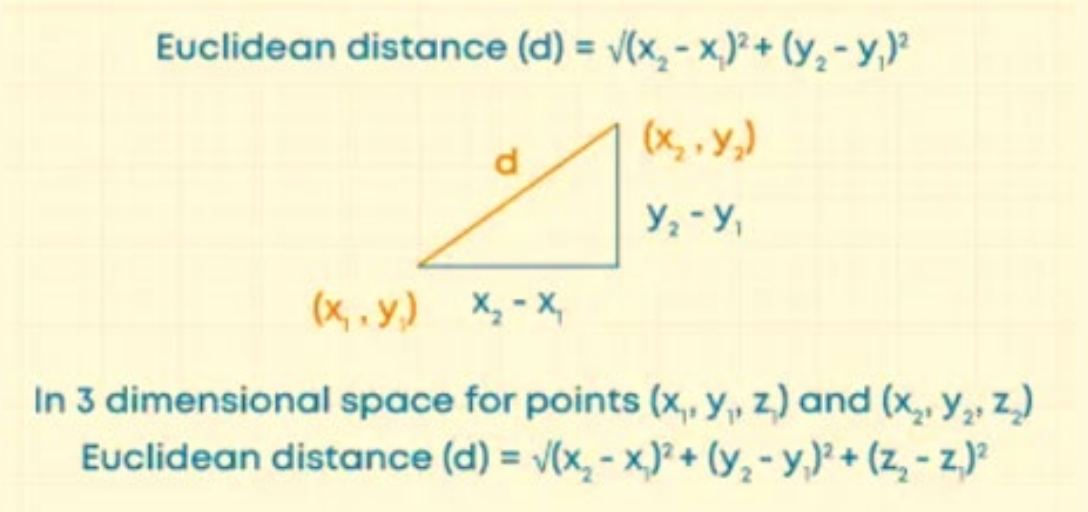
III. Cluster Analysis

I. Type of Cluster Analysis to Use
Hierarchical vs. K-means

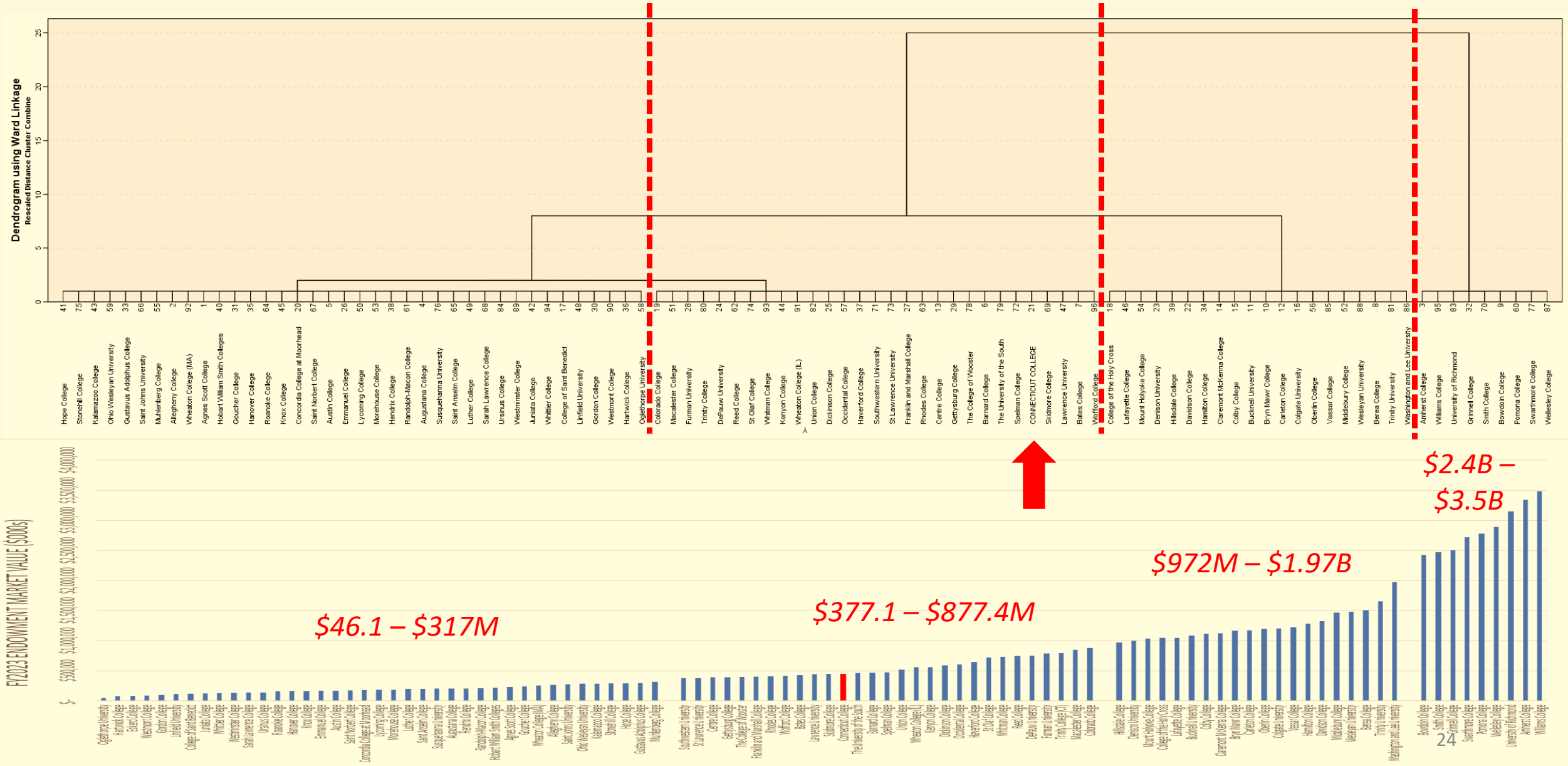
II. Linkage Method
Average linkage vs. Ward's method

III. How many clusters? (Priori)

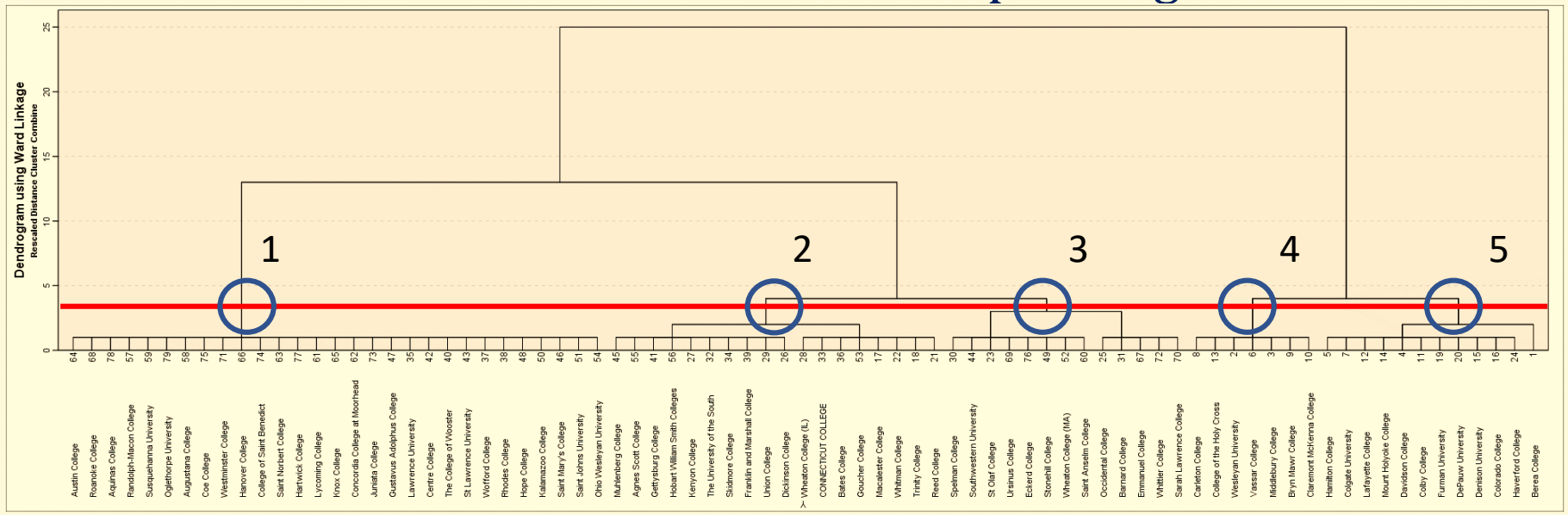
IV. How does hierarchical clustering work?



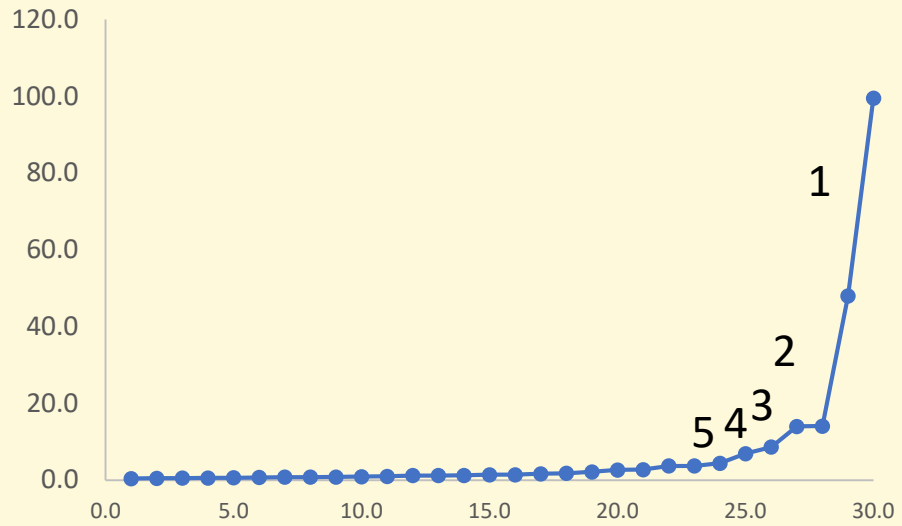
Hierarchical Clustering Dendrogram Using SPSS Explained: Endowment



Hierarchical Clustering Dendrograms With Financials in SPSS: RPP, Endowment And Spending Per Student



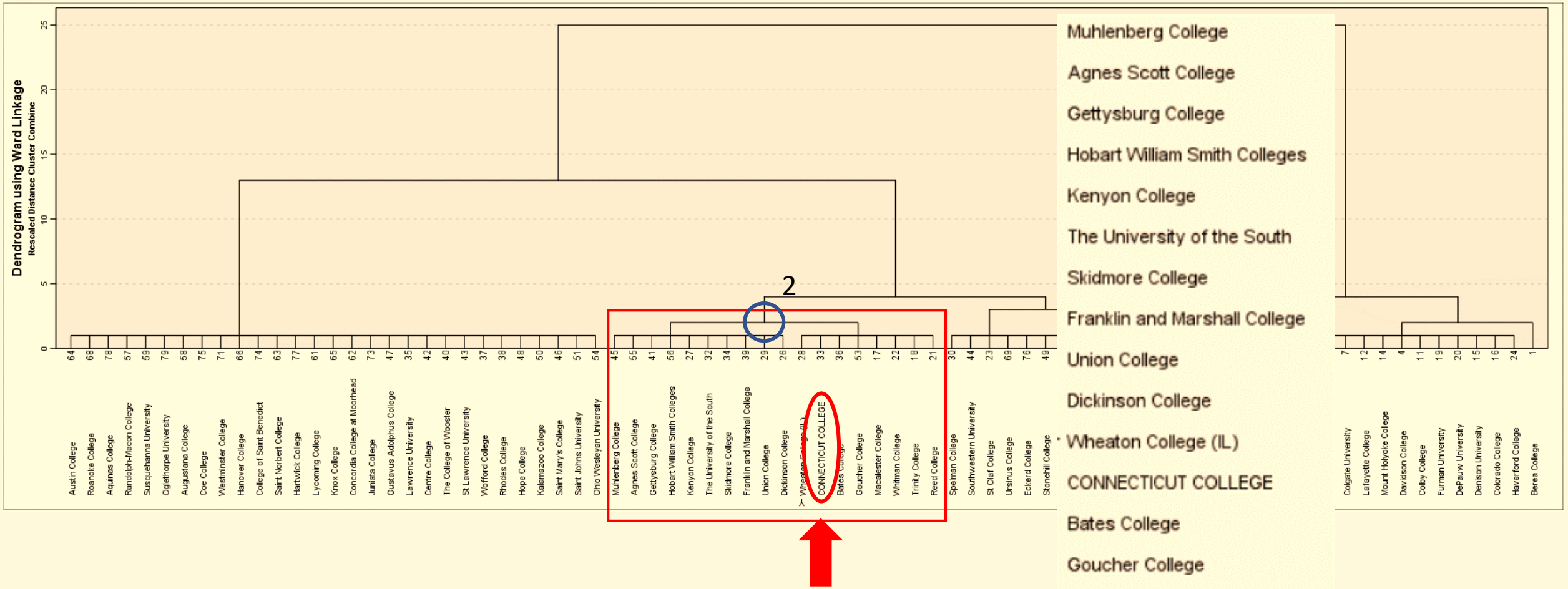
Agglomeration Schedule



Agglomeration Schedule

Stage	Cluster Combined Cluster 1	Cluster 2	Coefficients	Stage Cluster First Appears Cluster 1	Cluster 2	Next Stage	Diff
70	2	3	34.728	56	69	77	4
71	8	9	38.427	63	58	75	4
72	7	13	42.856	0	67	75	4
73	1	6	49.780	62	66	76	7
74	5	24	58.396	68	65	76	9
75	7	8	72.399	72	71	78	14
76	1	5	86.498	73	74	77	14
77	1	2	134.517	76	70	78	48
78	1	7	234.000	77	75	0	99

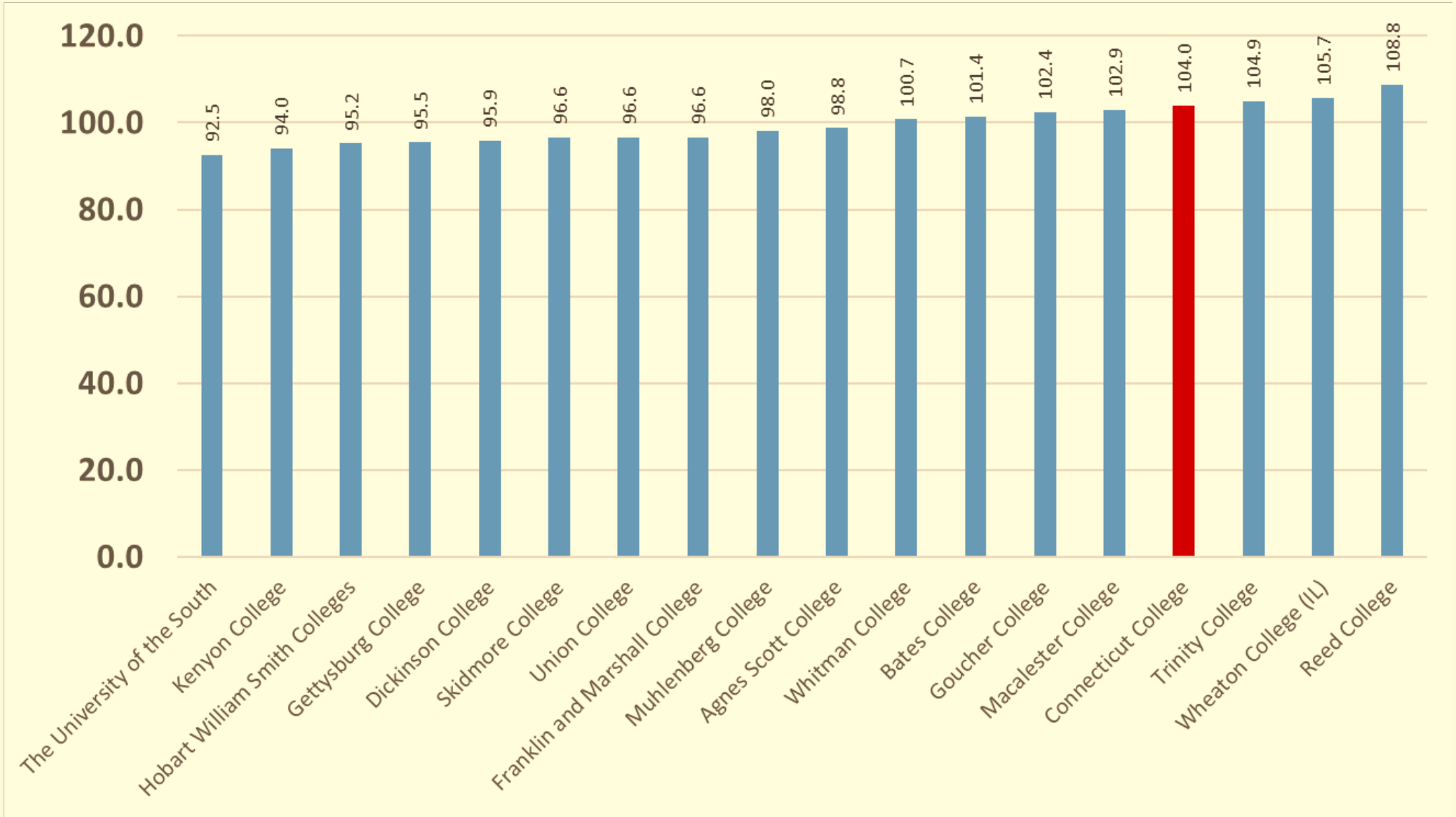
Extracting Cluster # 2: RPP, Endowment And Spending Per Student



17 schools
+ Conn

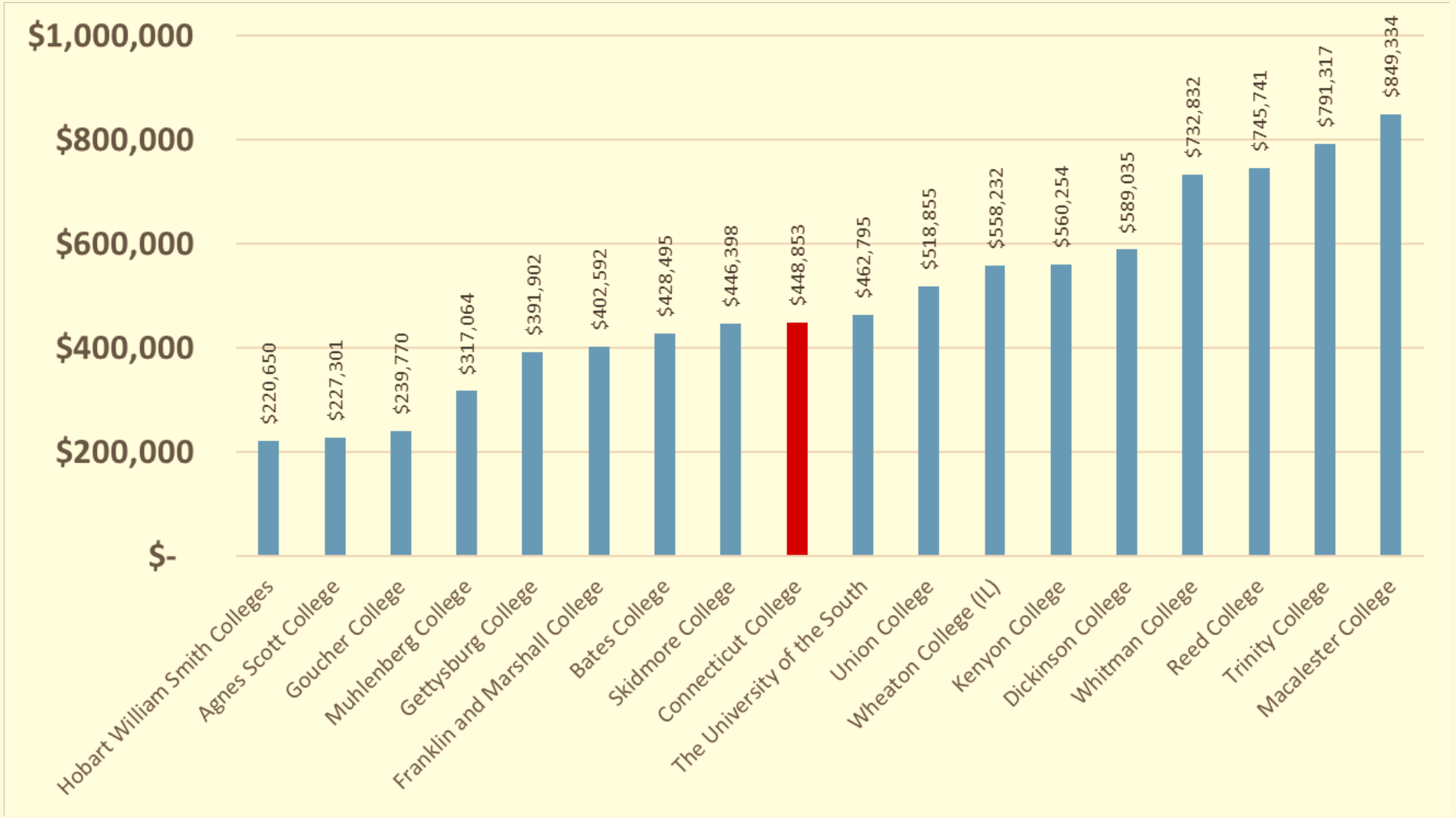
- Muhlenberg College
- Agnes Scott College
- Gettysburg College
- Hobart William Smith Colleges
- Kenyon College
- The University of the South
- Skidmore College
- Franklin and Marshall College
- Union College
- Dickinson College
- Wheaton College (IL)
- CONNECTICUT COLLEGE
- Bates College
- Goucher College
- Macalester College
- Whitman College
- Trinity College
- Reed College

Details of Each Matric: 2023 Cost of Living (RPP)

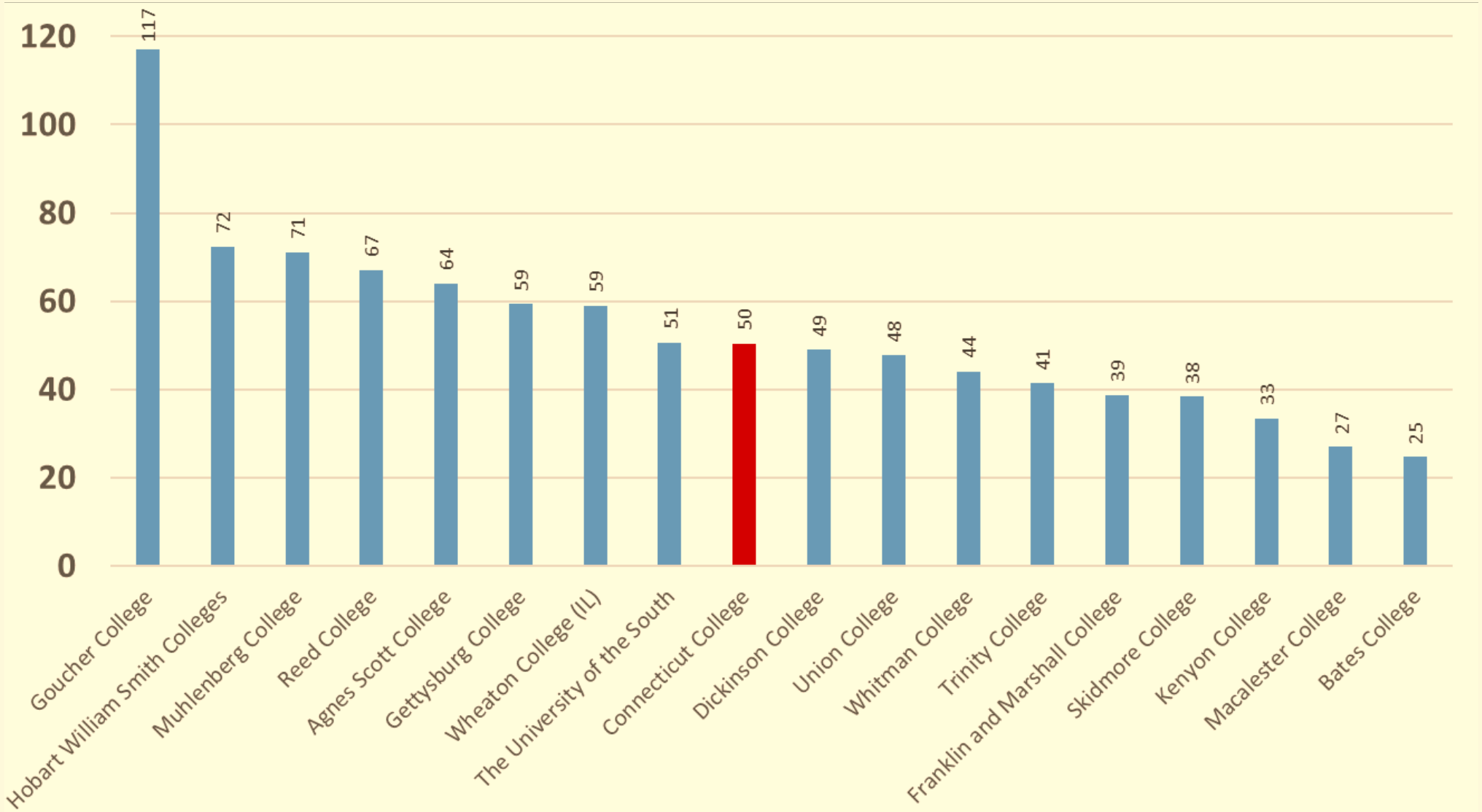


Details: Per-student Spending (3-yr. avg., thru 2023)

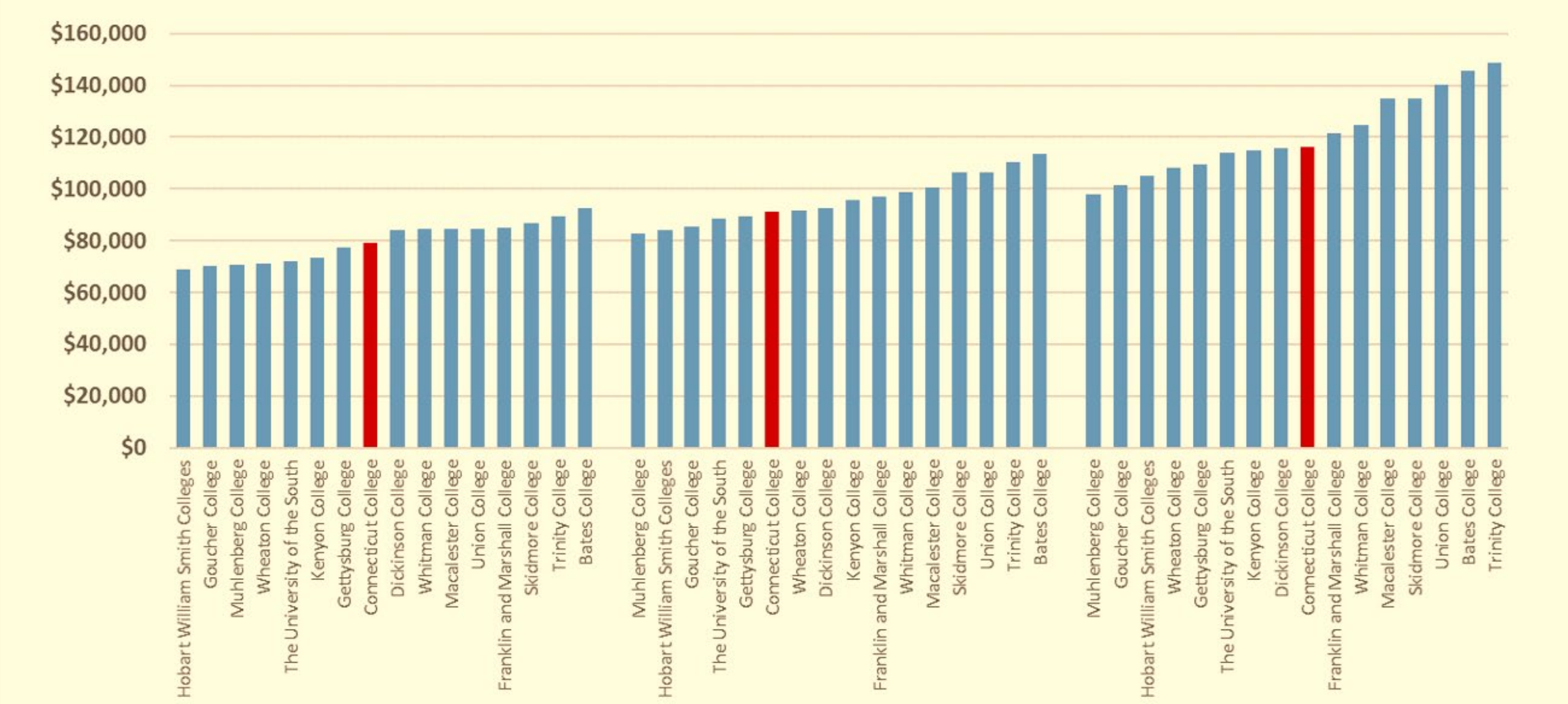
Details: 2023 Endowment ($000s)



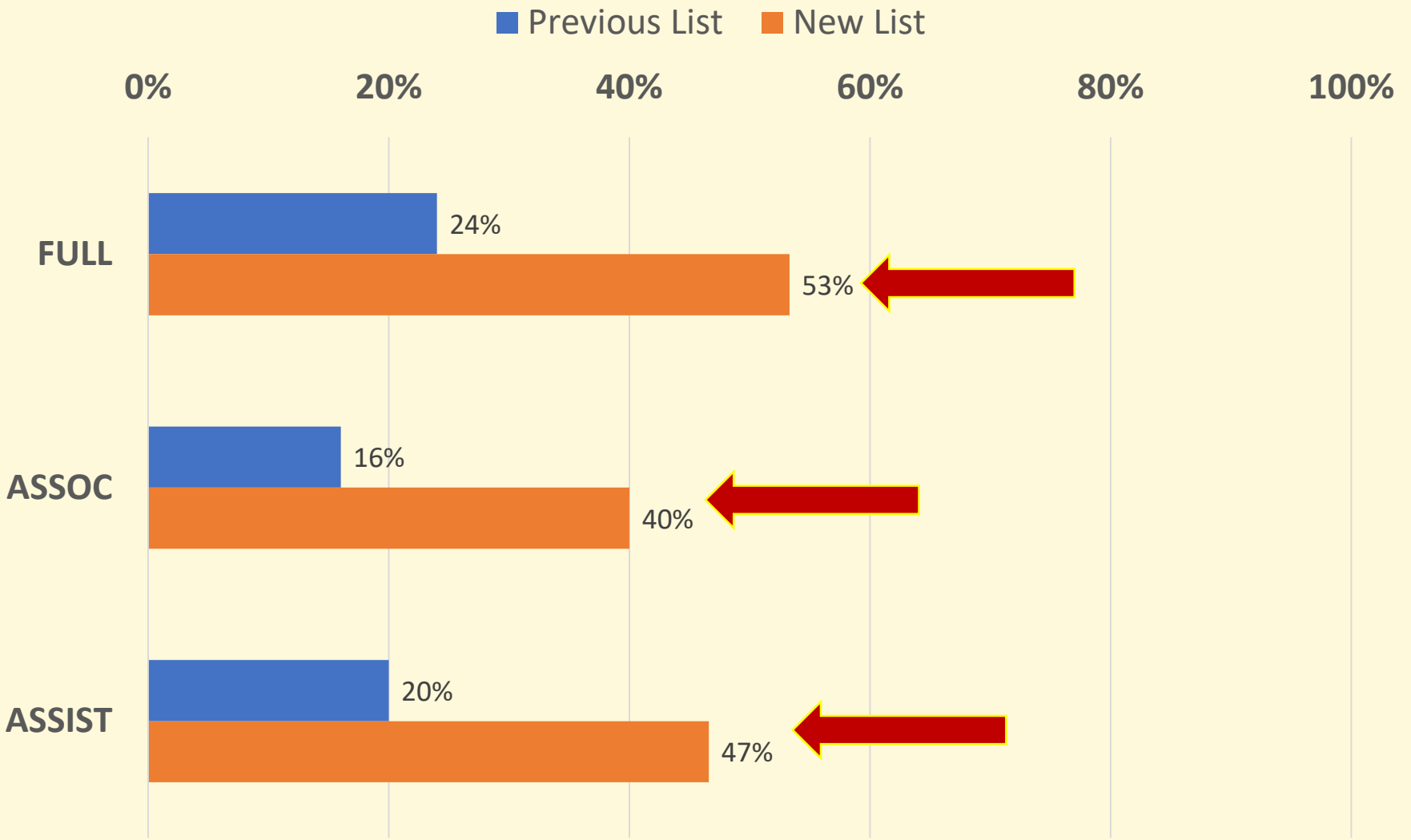
Details: U.S. News Ranking (3-yr. avg., thru 2023)



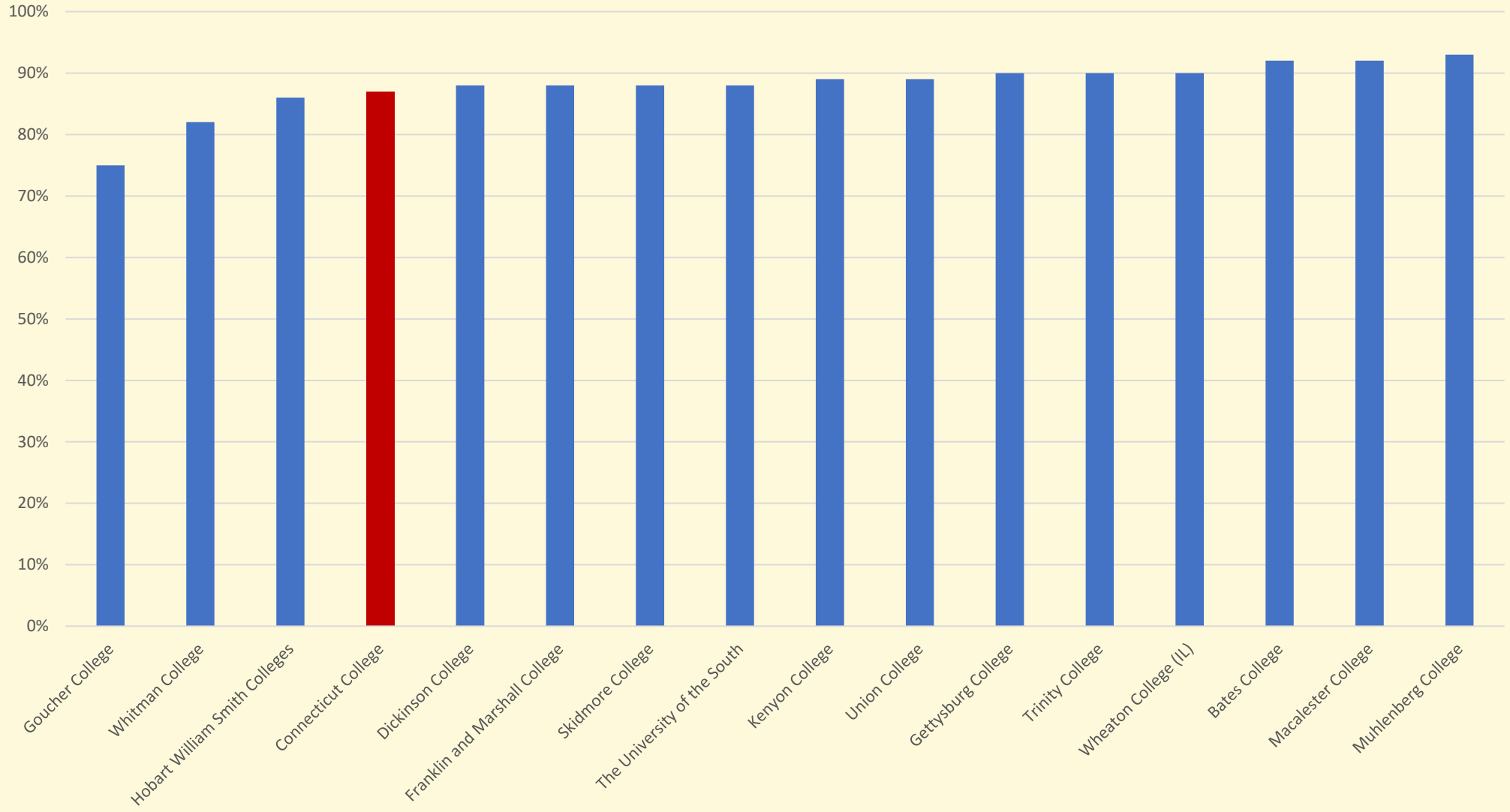
Details: 2023-24 Avg. Faculty Salaries



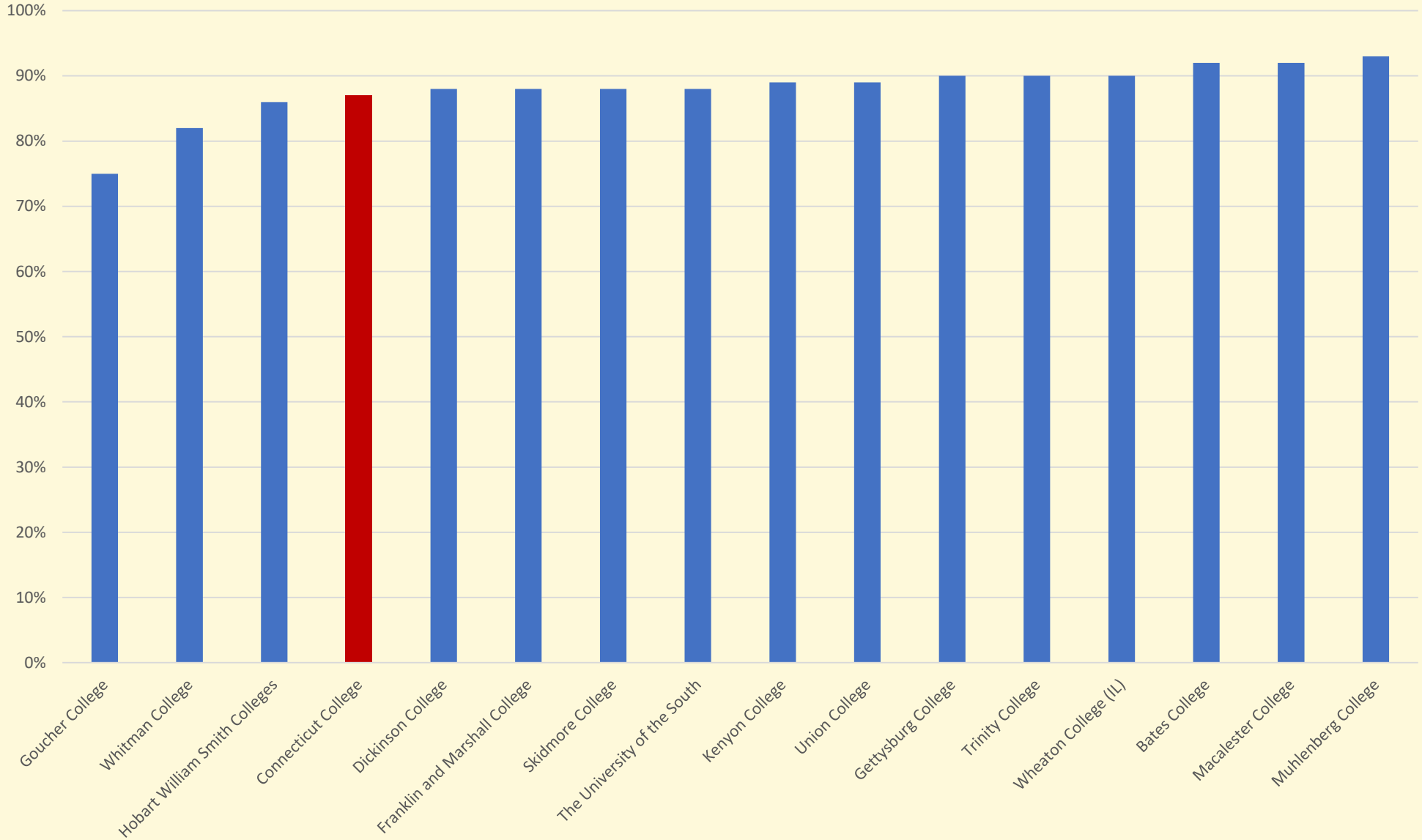
Conn's AY 2022-23 Salary Percentiles Between Previous and New Peer Group



Details: Retention rates (3-yr. avg., thru 2023)



Details: 6-yr Grad rates (3-yr. avg., thru 2023)



So….Takeaways

- The peer group update process is both political (art) and technical (science)
- Our College had not been revised its peer group since 2006. It should be reviewed regularly (recommendation: every five years)
- Connecticut College cluster analysis process was novel because of the case and variable reduction approach and Ward's method
- Cluster analysis is one technique that can be used to develop a peer group
- Cluster analysis can be deployed with widely available software (SPSS, R, or Stata) and using readily available datasets (IPEDS, NACUBO)
- Robust and transparent methodology for peer selection that can be explained to a larger audience and is defensible
- Once the peer group gets approved, then we can do other sorts of analysis for variables we didn't use for selecting the peer group

References

- Armacost, Robert L., et al. Selecting Peer Institutions Using Cluster Analysis. AIR Forum, 2003.
- Blashfield, Roger K. "Mixture Model Tests of Cluster Analysis: Accuracy of Recovery." Psychological Bulletin, vol. 83, no. 3, 1976, pp. 377–388.
- Boronico, Jess, and Shail Choksi. "Identifying Peer Institutions Using Cluster Analysis." American Journal of Business Education, vol. 5, no. 2, Apr. 2012, pp. 51–60.
- Ferreira, Luis, and David B. Hitchcock. "A Comparison of Hierarchical Methods for Clustering Functional Data." Communications in Statistics—Simulation and Computation, vol. 37, no. 3, 2008, pp. 545–568.
- Fuller, Andrea. "In Selecting Peers for Comparison's Sake, Colleges Look Upward." Chronicle of Higher Education, 10 Sept. 2012.
- Hands, David, and Brian S. Everitt. "A Multivariate Approach to the Analysis of Variance." The Statistician, vol. 36, no. 4, 1987, pp. 365–374.
- Hair, Joseph F., et al. Multivariate Data Analysis. 7th ed., Pearson, 2010.
- Kuiper, Samuel, and Lloyd Fisher. "A Monte Carlo Comparison of Six Clustering Procedures." Biometrics, vol. 31, no. 3, 1975, pp. 627–637.
- Luna, Andrew L. Selecting Peer Institutions Using Cluster Analysis. Austin Peay State University, 2018.
- McCormick, Alexander C. Hidden in Plain View. 10 May 2007.
- Milligan, Glenn W., and Laura M. Sokol. "A Two-Stage Clustering Algorithm with Robust Recovery Characteristics." Educational and Psychological Measurement, vol. 40, no. 3, 1980, pp. 755–759. https://doi.org/10.1177/001316448004000320.
- Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "FaceNet: A Unified Embedding for Face Recognition and Clustering." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823. https://arxiv.org/abs/1503.03832.
- Smith, Kent W. Multi-Dimensional Scaling Approach to Identifying Comparison Schools. Trinity College, Hartford, CT, 2000.
- Suhr, Diana. Factors vs. Clusters. SIR Consulting, 2018.
- University of Central Florida. A Multi-Phase Approach for Identifying University “Organizational Peers” Using Cluster Analysis.
- University of Rhode Island. Peer Group Executive Summary. 2004.

Questions?