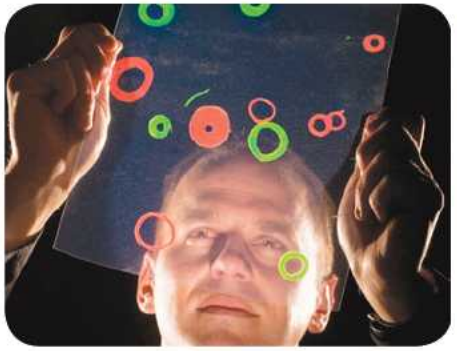
SAND2012-6546P

*Exceptional service in the national interest*

Sandia National Laboratories



# Development and Exploration of Cultural Archetypes Using JMP

Presented by Thor D. Osborn

September 12, 2012

U.S. DEPARTMENT OF ENERGY | NNSA

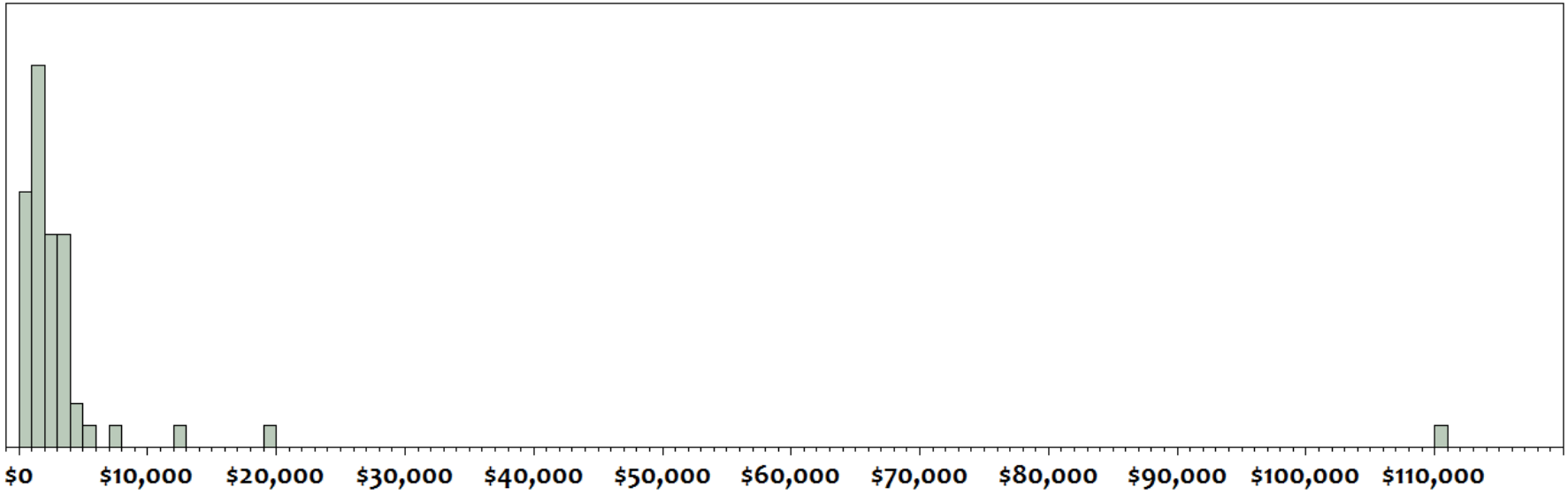
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

# Overview

- Societies are often said to have cultural similarities and differences
  - Can cultural archetypes having distinct tendencies be systematically identified?
  - How can archetype membership be determined?
    - Without inserting substantial personal bias?
    - Without a Ph.D. in anthropology?
  - JMP enables exploration
    - Fluid extraction of results from one platform as inputs to the next
    - Graphical design of equations for data pre-processing (similar to Mathcad)
    - Interactive graphical display of results for ease of presentation
- Onward – exploring inward-flowing greenfield FDI

# The Outcome Data

- Foreign Direct Investment (FDI) is important to economic development
- Inward-flowing “greenfield” FDI (IGFDI) is especially important to less developed economies
- There are vast differences between winners and losers – why?



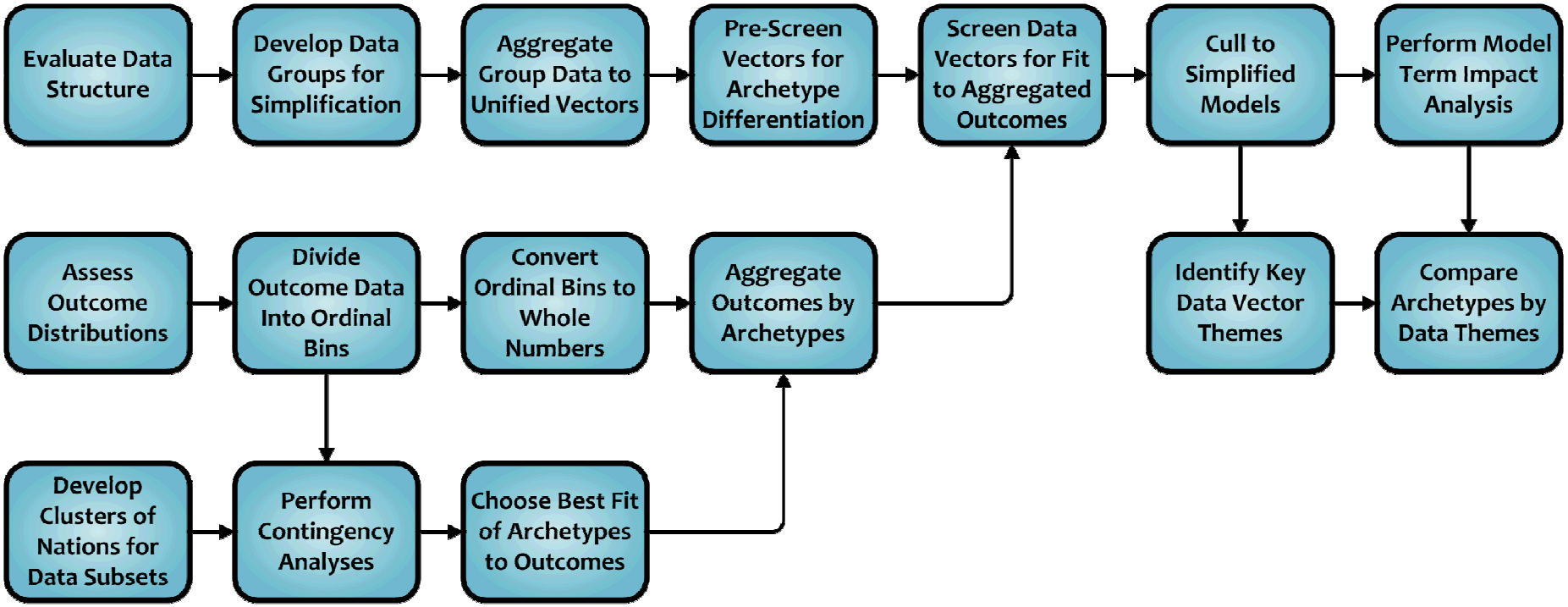
# The Comparison Data

- All of the data used in this analysis are from published materials – most can be readily downloaded from online sources
- GLOBE study of leadership traits and expectations around the world – 57 societies used in this analysis
- Societal traits from Culturally Endorsed Implicit Leadership Theory
- Socioeconomic data and characterizations from a broad range of sources
  - World Bank
  - World Economic Forum
  - Heritage Foundation
  - Transparency International
  - Freedom House
  - Vision for Humanity
  - International Food Policy Research Institute
  - United Nations

**The data chosen to represent the societies necessarily impacts the contextually-based societal archetype structure:**

**different data → different results**

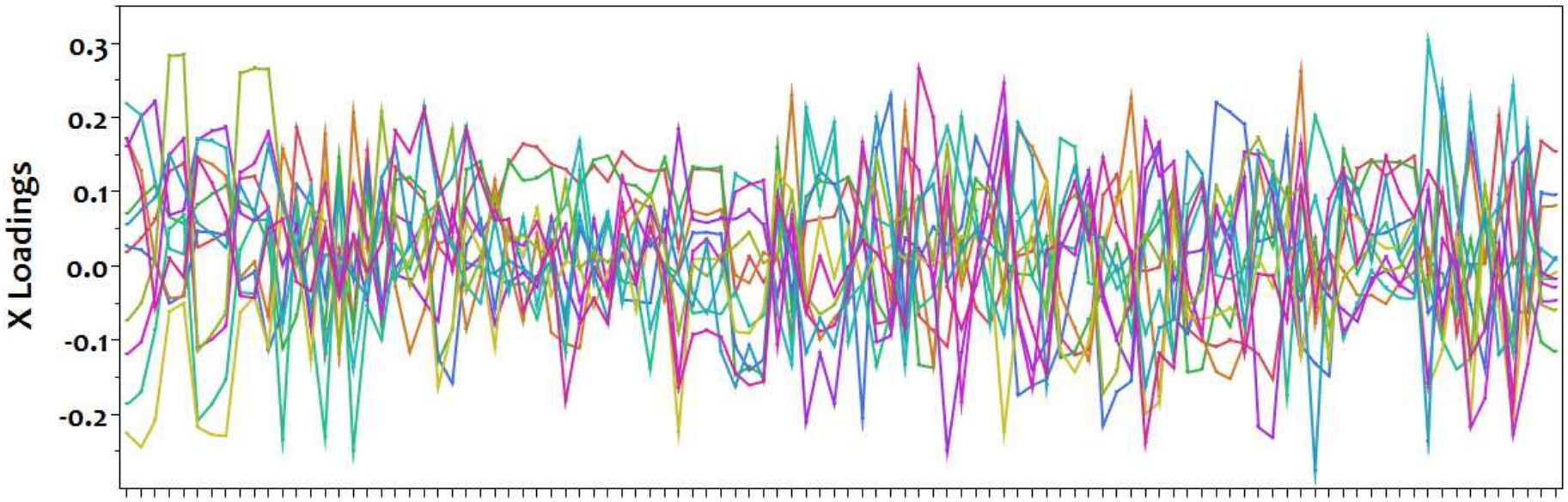
# The Model Development Process




**Data consolidation and pattern-matching for intuitive comprehension and hypothesis generation – *not* hypothesis testing**

# One Alternative - PLS

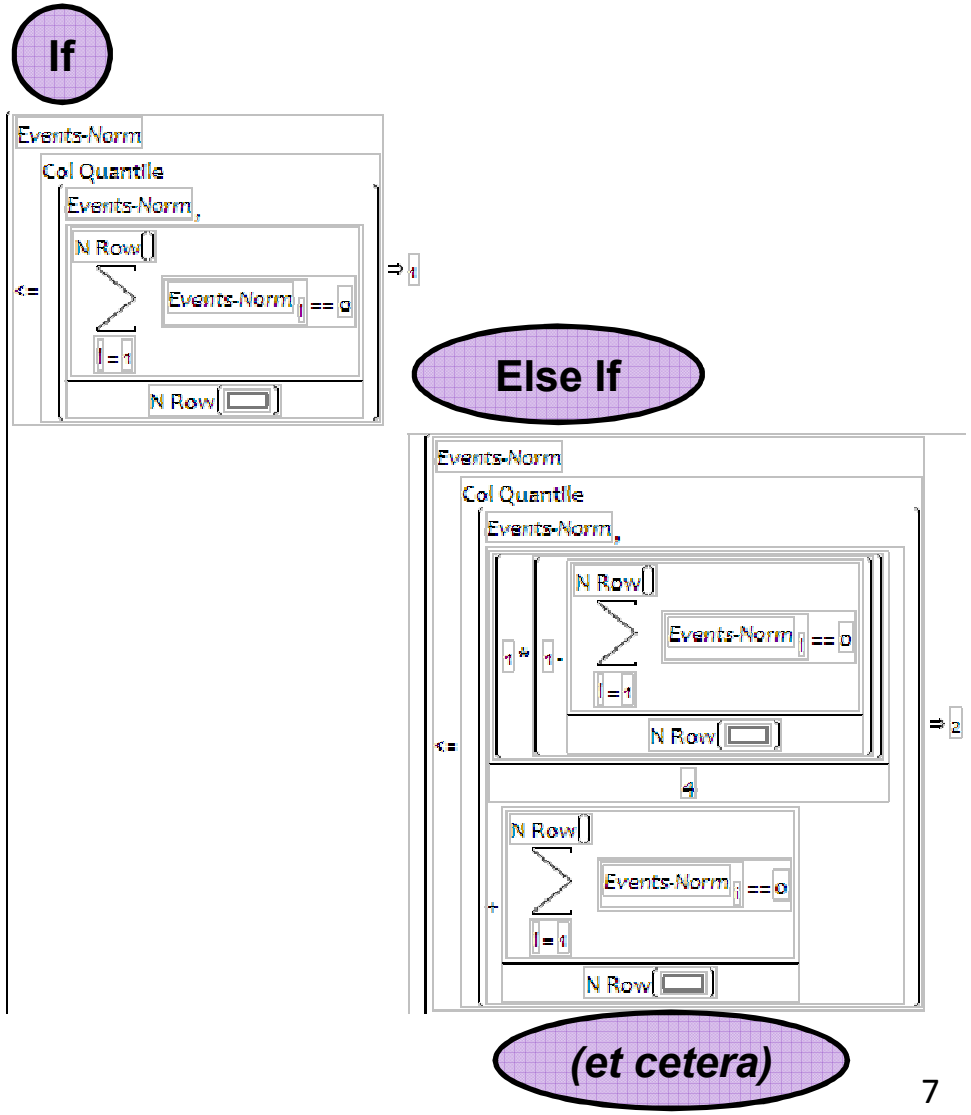
- Loadings plot
  - 12 latent variables, each is composite of original 90 inputs (not shown)
  - Model explains 99.3% of variation in outputs



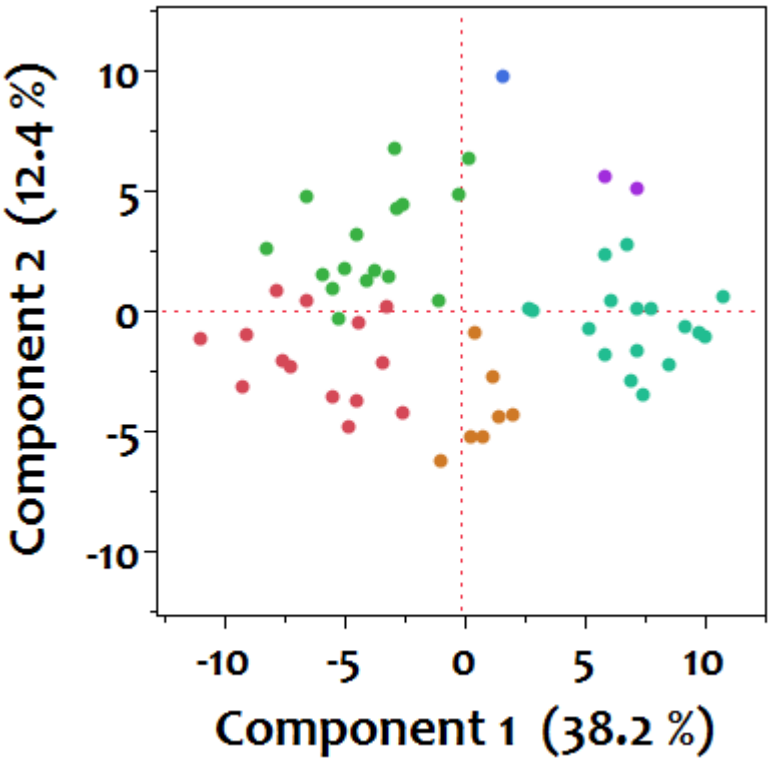

**Mathematically, PLS can "explain" nearly all of the variation – intuitively, somewhat less**

# Convert Outcome Data to Ordinals

- Generate new column and use Formula Editor
- Note traditional graphic mathematical notation
- Regular quantiles generated for IGFDI data, but JMP enables alternatives
  - The formula at right creates modified “quintiles” that place all zeroes in the first level instead of arbitrarily assigning equal values to separate levels



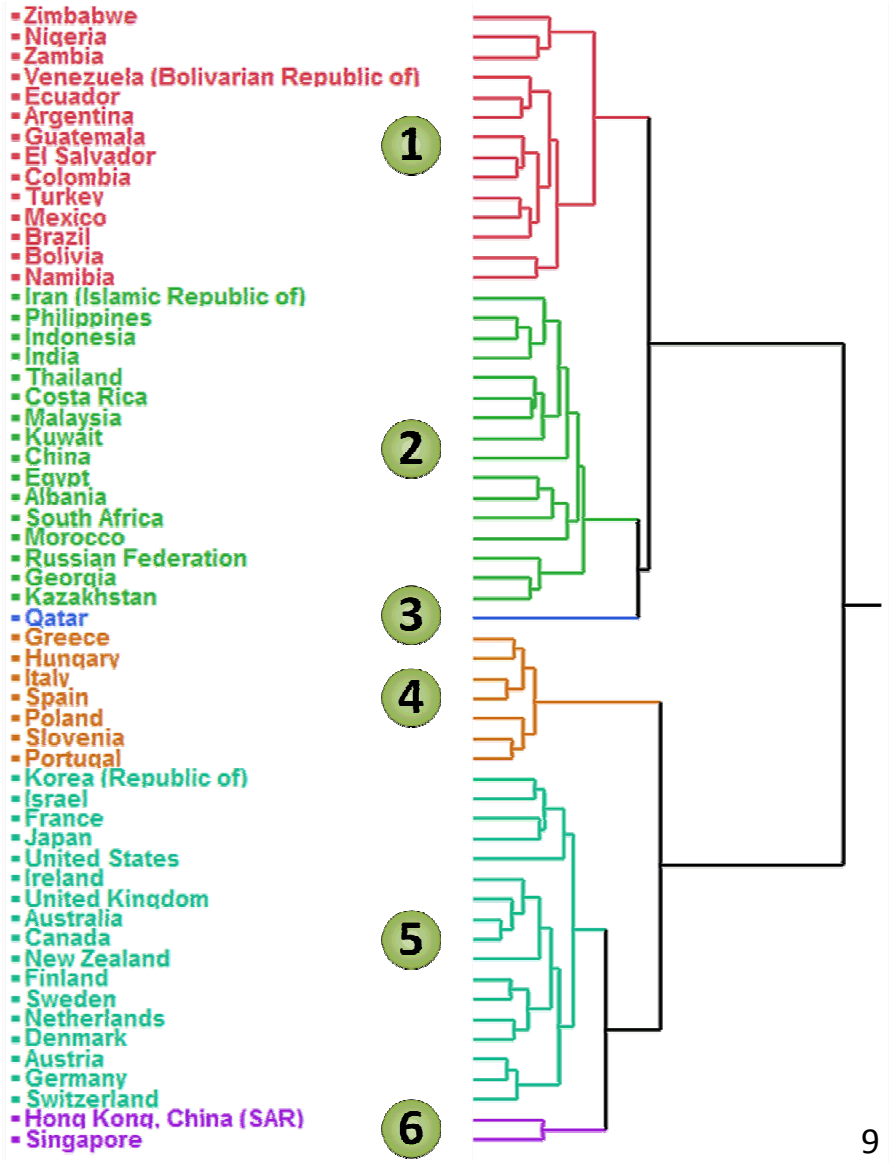
# Develop Clusters of Societies – 1/2




- Objective is to cluster societies by co-similarity
- Use principal components analysis to find first principal component of data subset
- Analyze → Multivariate Methods → Principal Components
  - Select all input data columns, hit "OK"
  - Save first principal component to data table

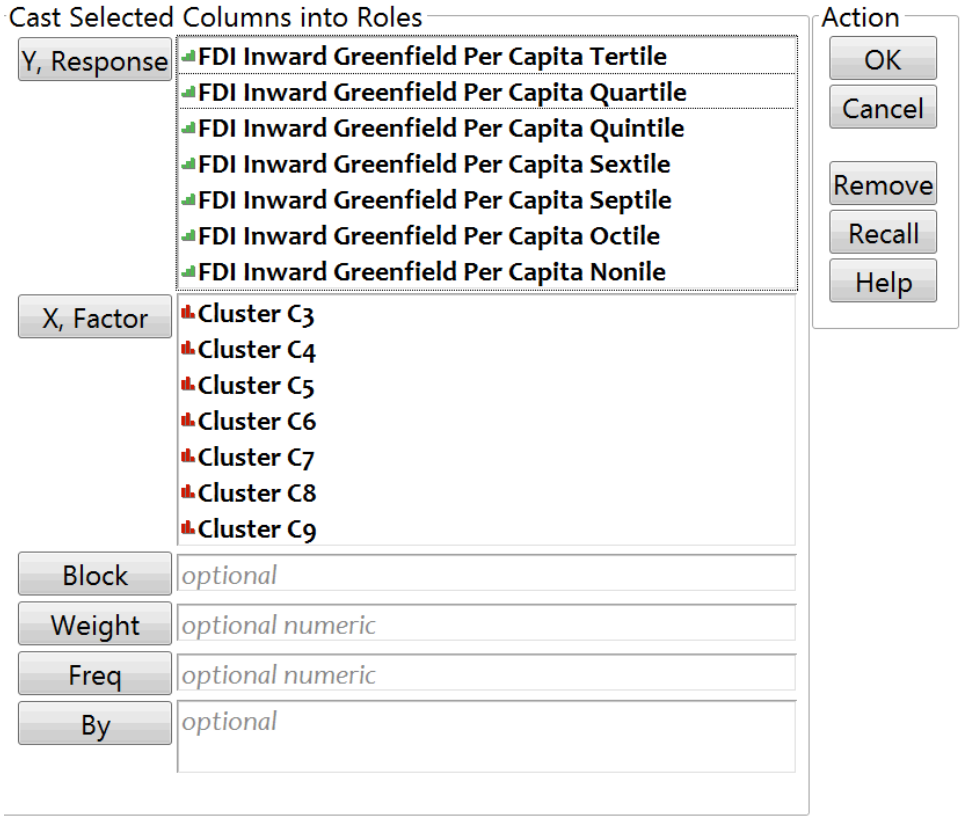
# Develop Clusters of Societies – 2/2

- Analyze → Multivariate Methods → Cluster
  - Select all data in subset for Ys
  - Select Country for Label
  - Select Prin 1 for Order
  - Hit "OK"
  - Red triangle menu
  - Set number of clusters, N
  - Save clusters to data table
  - Repeat for all N
- Post-processing
  - Image extracted using SnagIt
  - Identifiers superimposed using MS Visio
  - Cluster labeling: future enhancement for JMP?



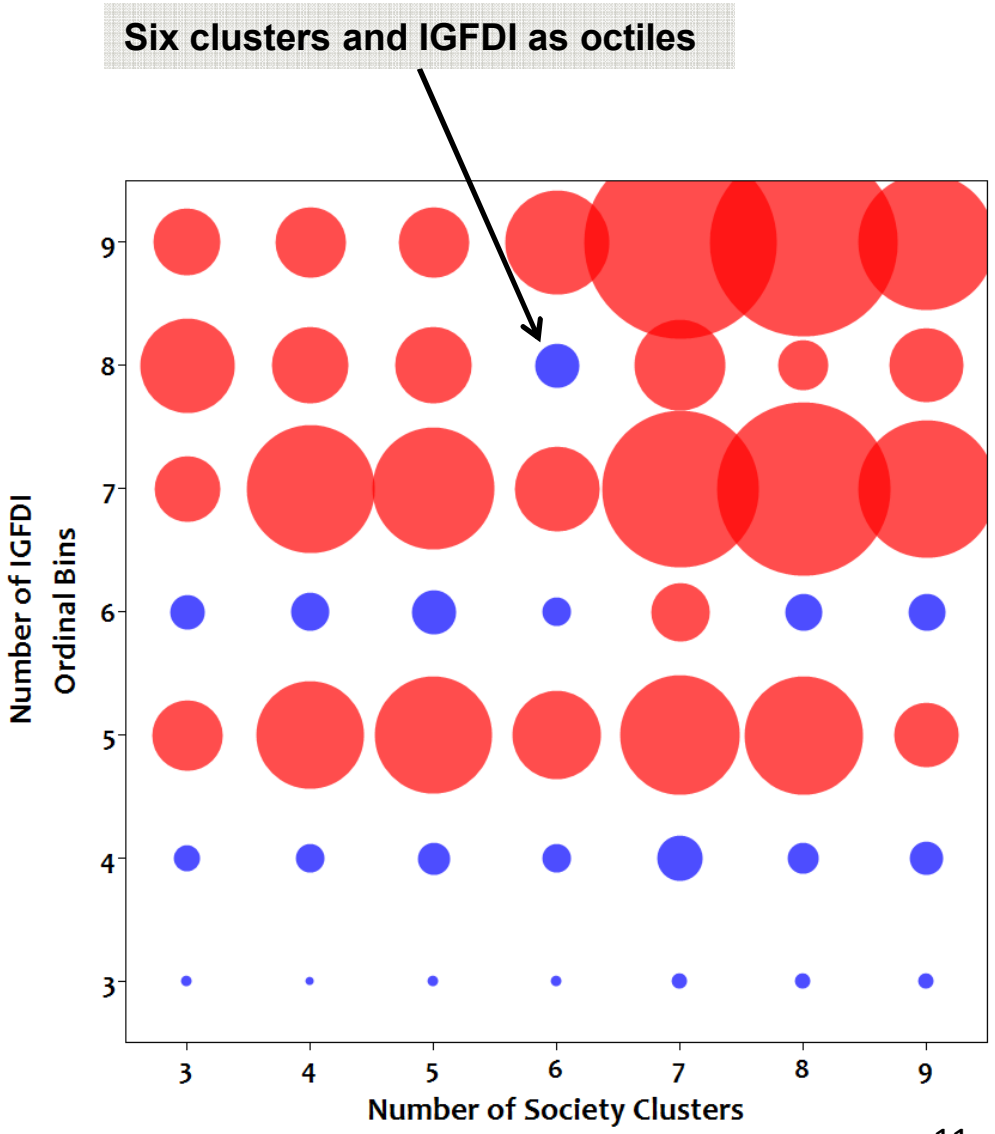
# Perform Contingency Analysis

- Analyze → Fit Y by X
  - Select cluster columns for X
  - Select normalized quantile columns for Y
  - JMP automatically engages the Contingency "personality" of the Fit Y by X platform because X is categorical and Y is ordinal
  - Hit "OK"
  - Statistical test outcomes appear below mosaic plot
  - Save output as text file for easy searching



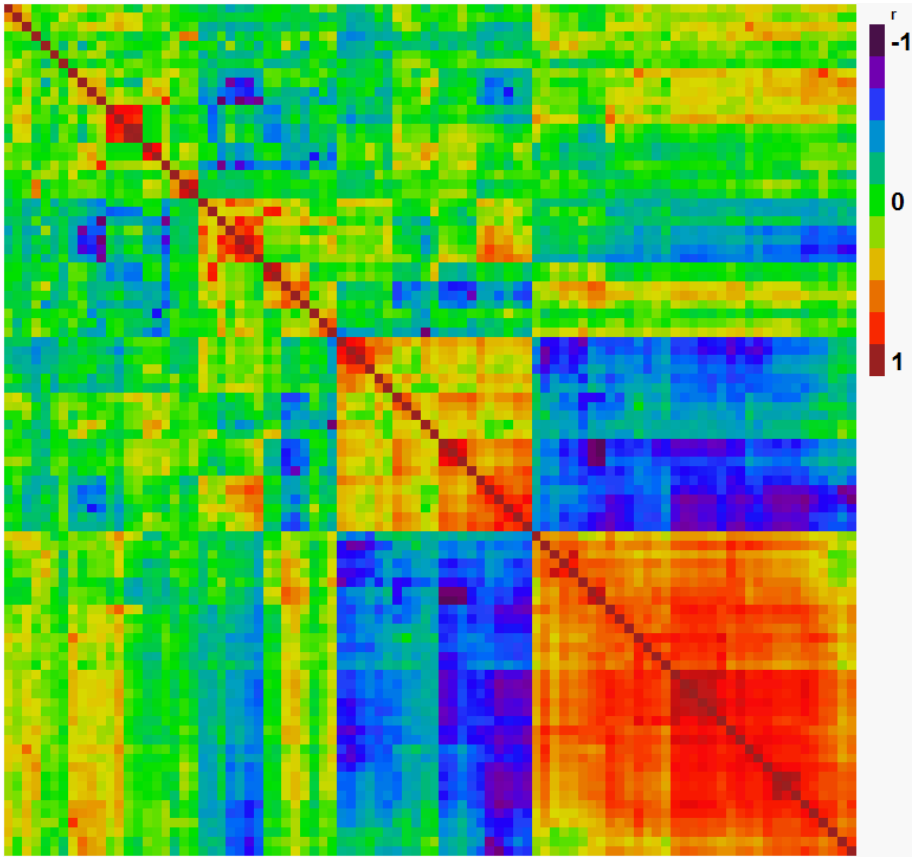
# Identify Best Match

- Use table of contingency analysis results
  - Pearson Chi-Square Probabilities
  - Use Formula Editor to make test column – check limit and whether Chi-Square suspect
- Graph → Bubble Plot
  - Ordinal bins as Y
  - Cluster size for X
  - Pearson Chi-Square probability for Data
  - Color by test column
- Use judgment to find most suitable combination



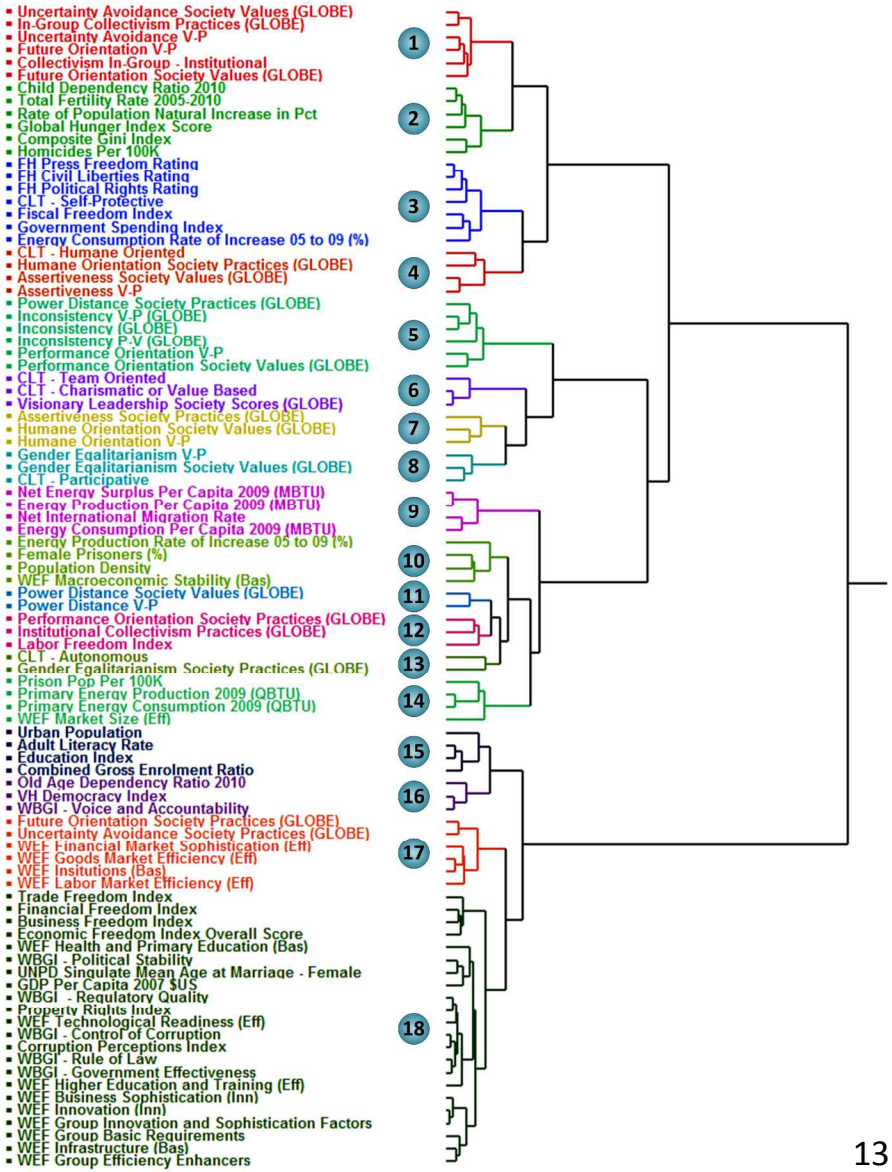
# Examine Data Structure

- Use full data table
- Analysis → Multivariate Methods → Multivariate
  - All input data for Ys
  - Hit “OK”
  - Red triangle menu: Color Maps → Cluster the Correlations
  - Visual indication that groupings of similar data exist in data set
  - Right-click over correlation table to save as a new table



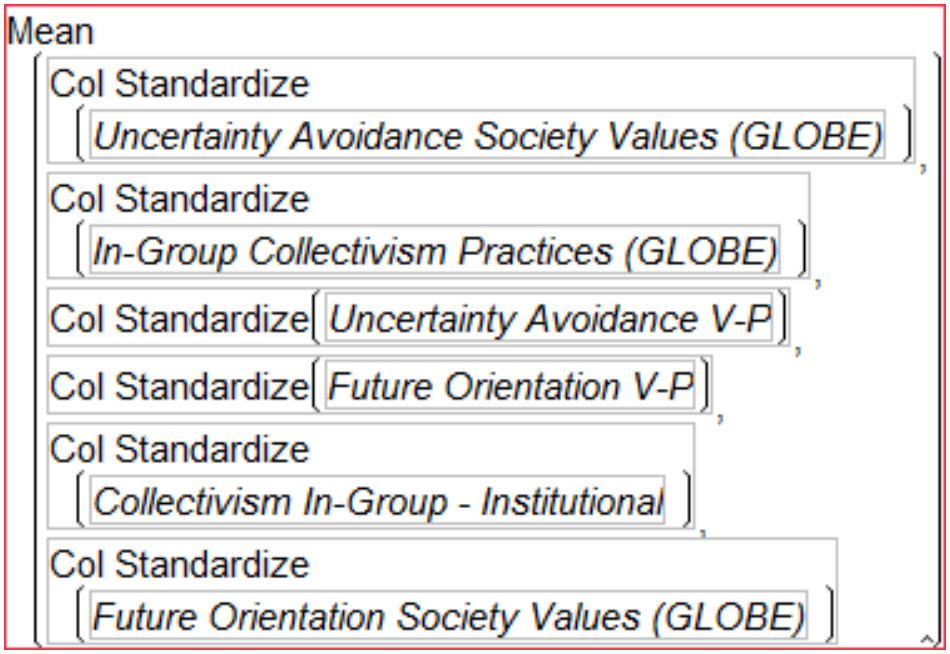
# Simplify Using Data Groups

- Use table generated from correlation matrix
- Perform PCA to obtain first principal component
- Cluster as before using Ward's method and Prin1
- Choose N=18 as a compromise between simplification and granularity
- Save Clusters to data table

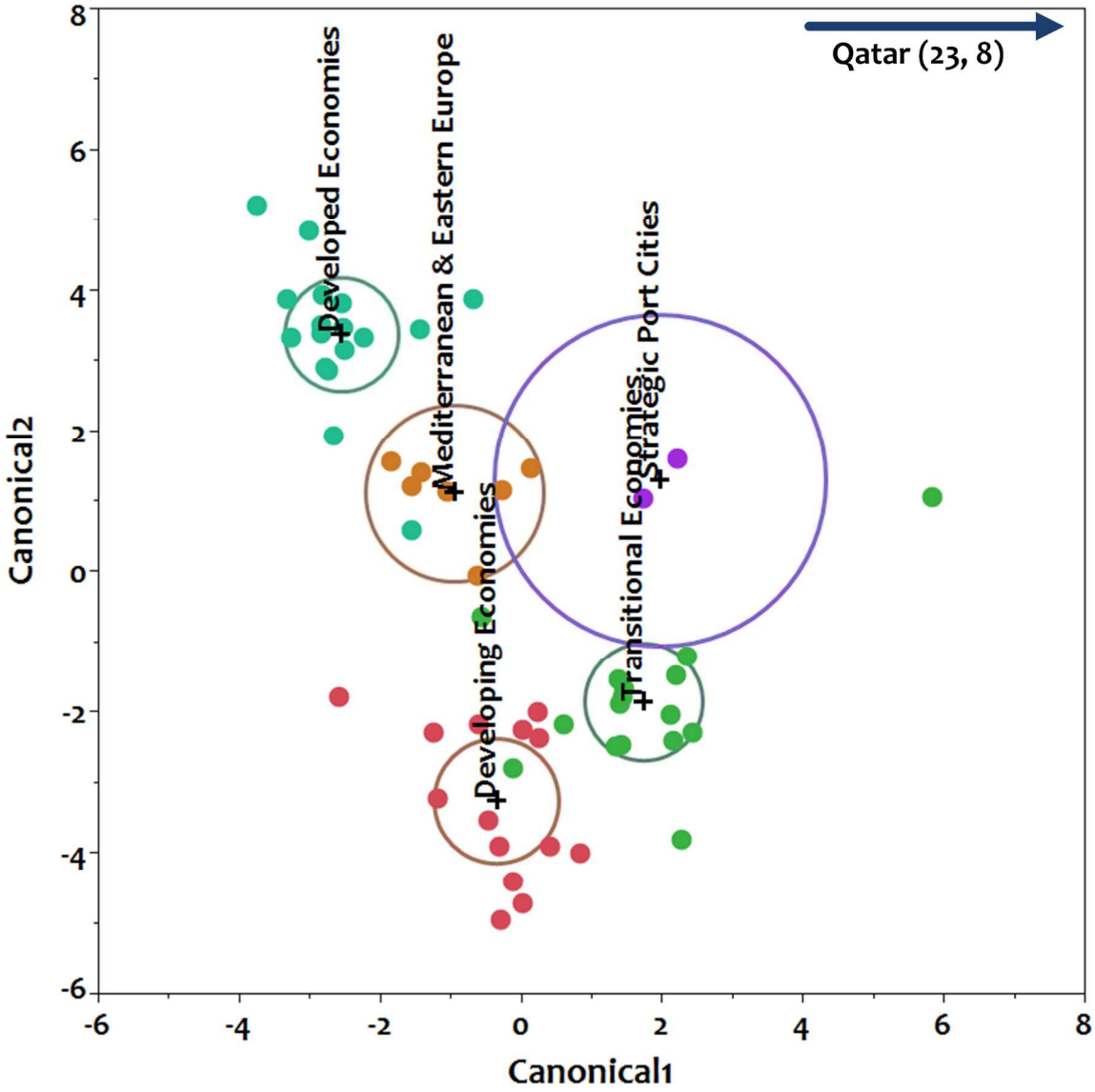


# Make Unified Group Vectors

- Add new columns to main data table
- Use Formula editor to generate means of column normalized cluster members
  - Example: Data Group 1



# Pre-Screen to Differentiators

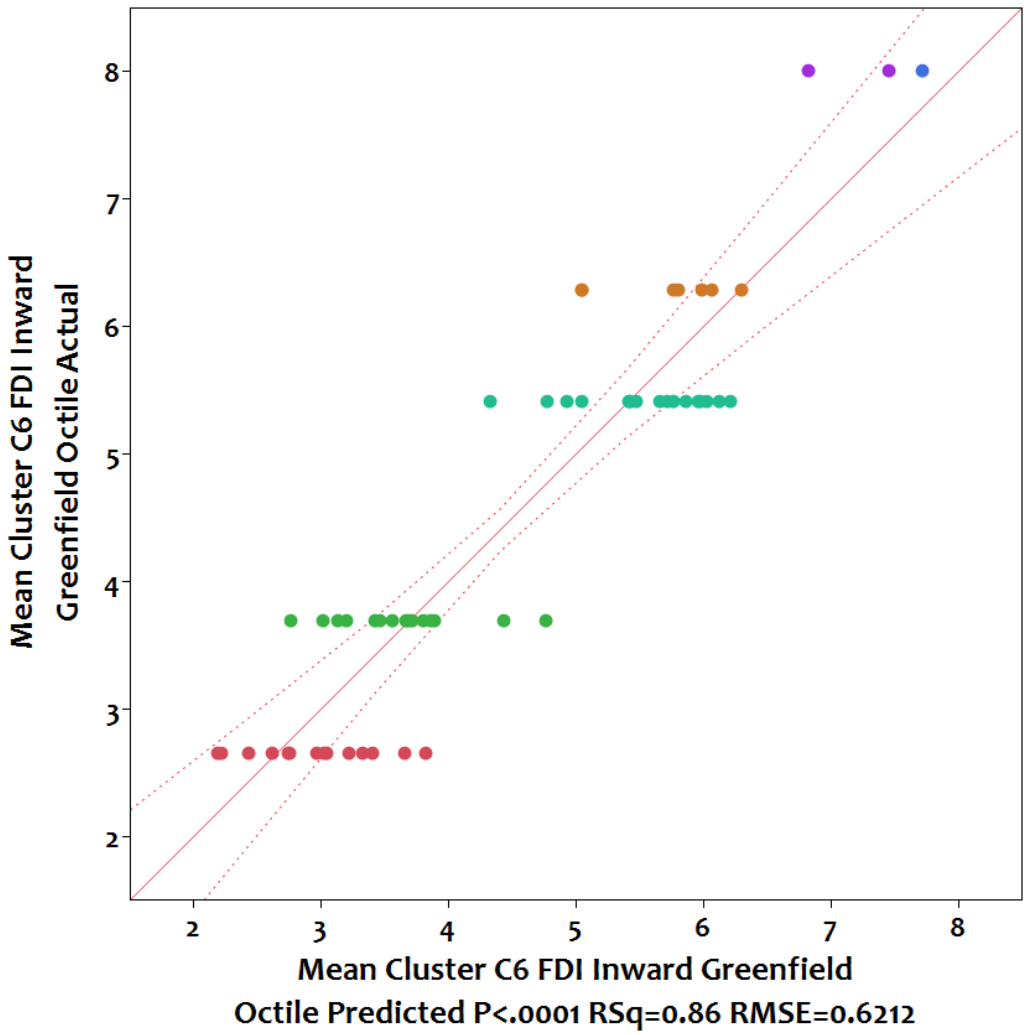



- Use Discriminant platform
- Stepwise variable selection
- Add all, remove one at a time until p=0.1

**Out of 18 vectors, seven are non-differentiating with respect to cluster identification – these can be removed for final model determination**

# Model Fit to Vectors

- Only five input data vectors with cohesive themes
- Clear separation of societal clusters aka cultural archetypes
- Useful for qualitative understanding

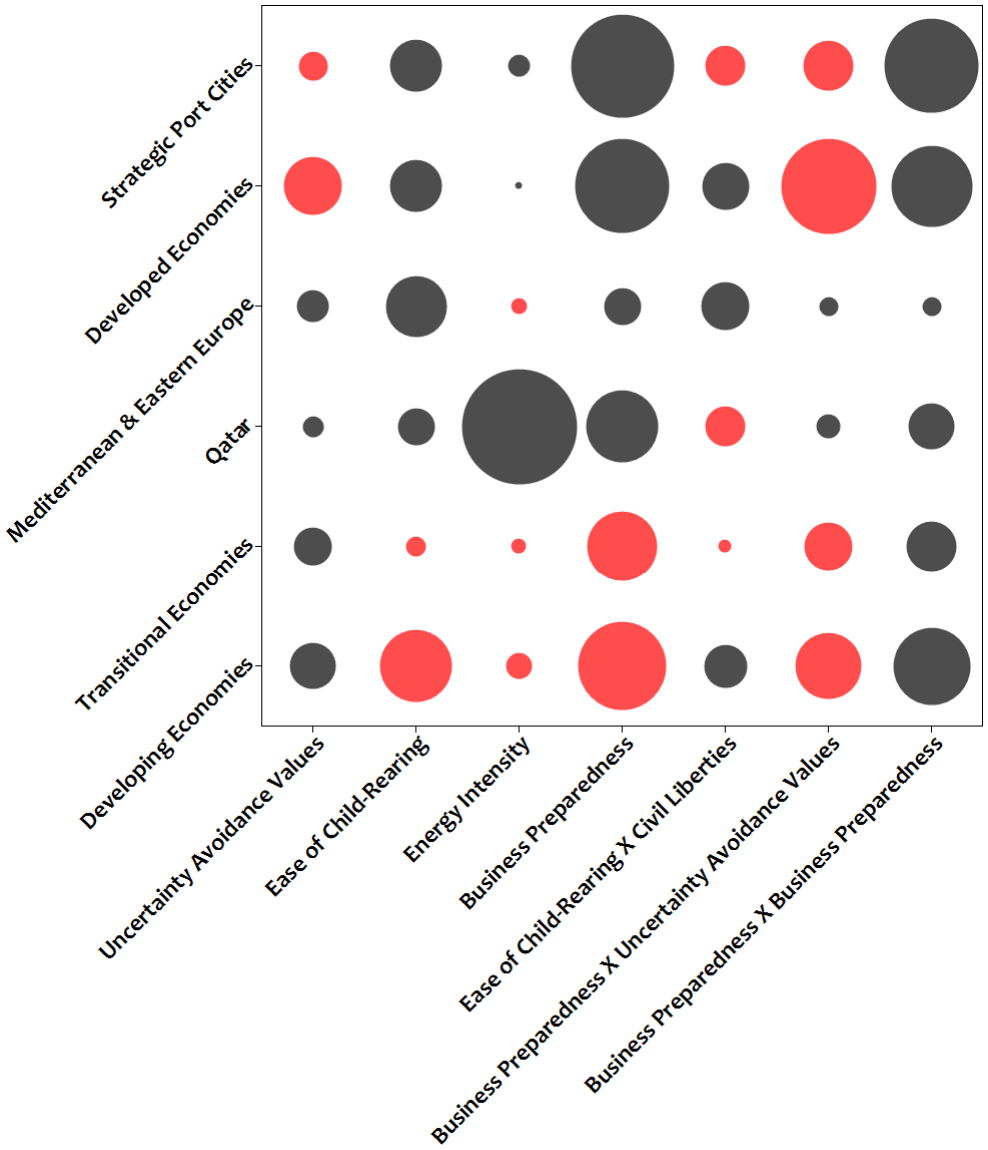


# Semi-Quantitative Model Drivers

Sandia National Laboratories

| Group | Group Theme | Representative Variable | Correlation |
|---|---|---|---|
| 1 | Uncertainty Avoidance Values | Uncertainty Avoidance V-P (Values – Practices) | 0.9376 |
| 2 | Ease of Child-Rearing | Child Dependency Ratio | 0.9691 |
| 3 | Civil Liberties | FH Civil Liberties Rating | 0.8850 |
| 9 | Energy Intensity | Energy Production Per Capita | 0.9818 |
| 18 | Business Preparedness | Government Effectiveness | 0.9836 |

# Characterize Societies by Impact




- Model terms multiplied by society cluster mean input values
- Illustrates societal differences and similarities
- Identifies factors with limited relevance – see Energy Intensity

# Conclusions

- Relatively stable cultural factors correlated with IGFDI
  - Inference: IGFDI trends unlikely to change rapidly regardless of immediate changes in economic marker data
- Method Has Limitations
  - Hierarchical clustering still dependent on data order
  - Influence of personal biases curtailed but not eliminated
  - Results are path-dependent and must be assessed for reasonableness
- Method Has Benefits
  - Useful cultural groupings can be developed from openly available data
  - Approach offers qualitative heuristic benefit for comprehension