



# Composite quantitative analysis of international healthcare accreditation survey data

Never Stand Still

Faculty of Medicine

Australian Institute of Health Innovation

Max Moldovan, Charles Shaw and Jeffrey Braithwaite



# Healthcare accreditation:

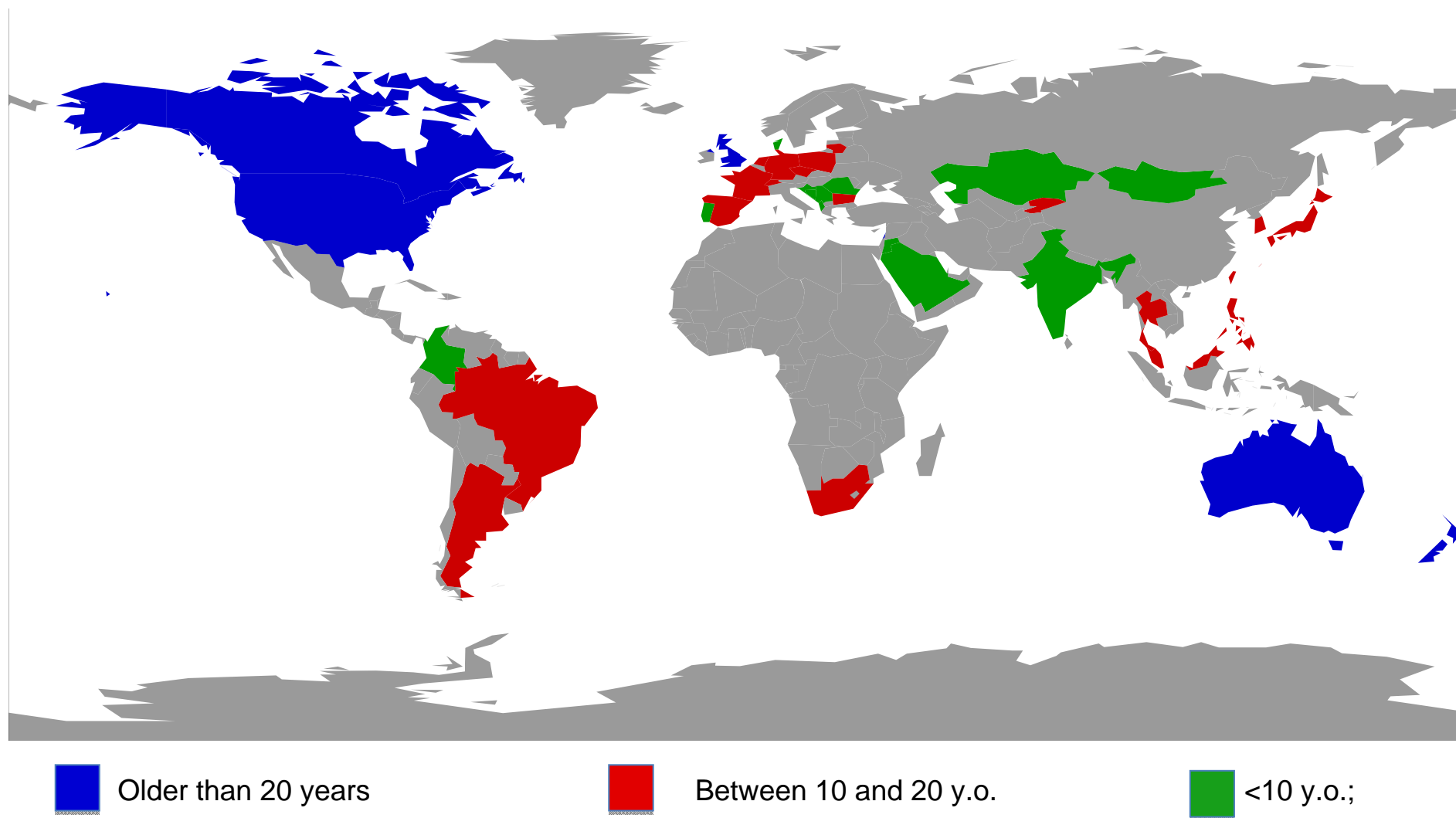
a process that a health care institution, provider, or program undergoes to demonstrate compliance with standards developed by an official agency.

*[Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health, Seventh Edition. © 2003 by Saunders, an imprint of Elsevier, Inc. All rights reserved.]*

# Study and data set description

- ✓ The first international healthcare accreditation survey was undertaken in 2000, commissioned by the World Health Organisation [WHO Review 2003, WHO/EIP/OSD/2003.1]
- ✓ The current data set has been collected over 2009-2011
- ✓ 44 accreditation agencies, located in 38 countries
- ✓ 10 broad categories of attributes (e.g. “Policy and governance” and “Funding of the organization”)
- ✓ 173 questions leading to categorical, date/numeric or free text responses

# Geographical and age distribution of accreditation agencies\*



\* If more than one program per country, the oldest program is displayed

# Methods of analysis

- ✓ Stratified categorical analysis
- ✓ Supervised learning via penalised regression
- ✓ Unsupervised learning via hierarchical clustering

# Stratified categorical analysis

- ✓ Organisations are stratified into two or more groups based on observed factors
- ✓ Responses arranged into contingency tables
- ✓ Fisher's exact conditional testing is applied
- ✓ The results are assessed based on the plots of sorted P-values

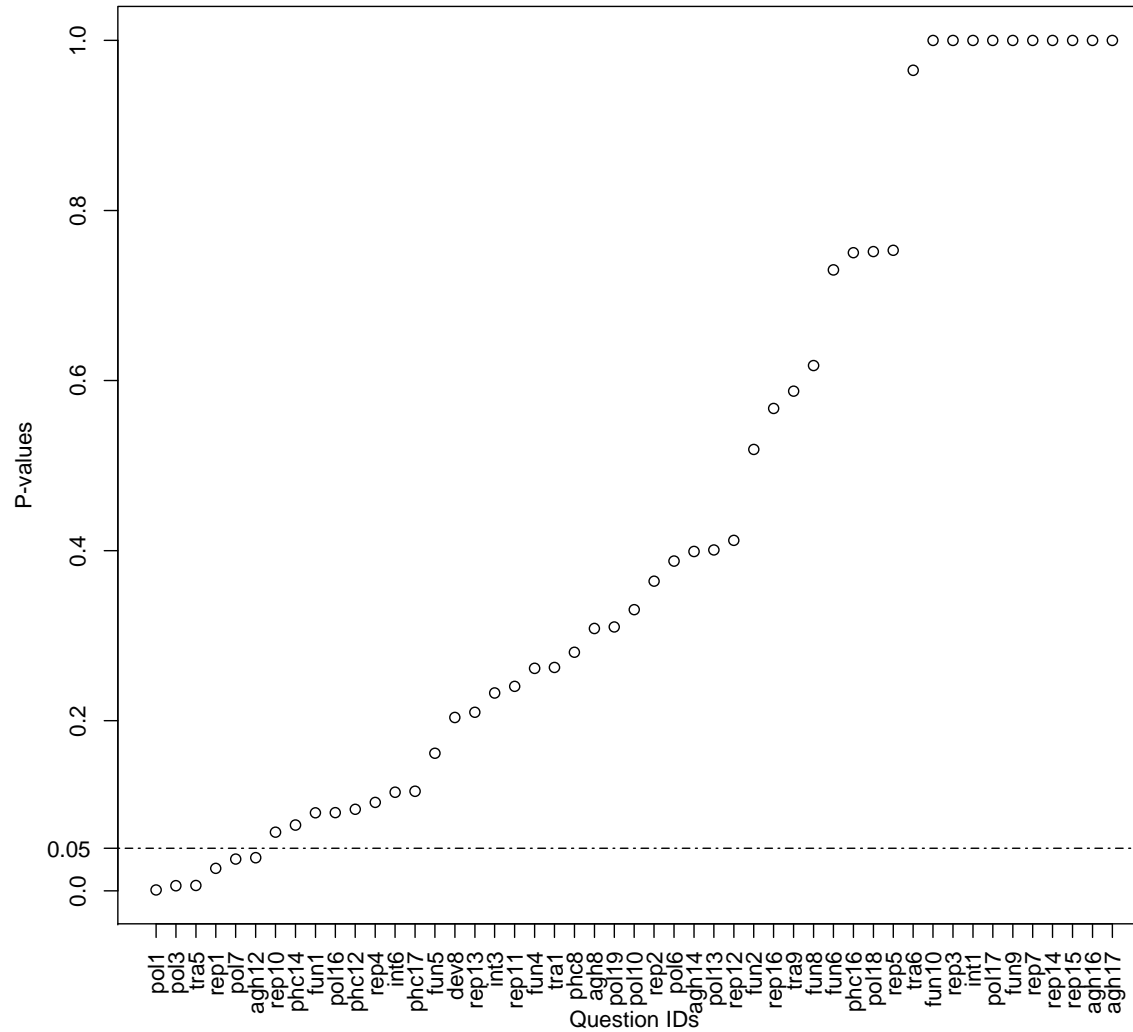
# Stratified categorical analysis

## Stratification factors:

1. Purchasing Power Parity adjusted GDP per capita
2. Government relatedness
3. Regional factor
4. Age of organisation
5. Country of origin population

# Stratified categorical analysis

GDP PPP factor (US\$20,000 split value)



# Supervised learning via penalised regression

- ✓ R package *penalized*: Penalised regressions via  $L_1$  and  $L_2$  penalties [Goeman (2010): Biometrical Journal 52 (1), 70–84]
- ✓  $L_1$ : **LASSO**;  $L_2$ : **ridge**;  $L_1$  and  $L_2$ : **elastic nets**
- ✓ Only  $L_1$  penalty (LASSO) has been used
- ✓ Supervisory (response) variable is an ordinal four-level categorical variable defined by field experts (similar to phenotypes defined by clinical experts)
- ✓ Explanatory variables are categorical responses coded as factors
- ✓ Logistic link function has been used

# Supervised learning via penalised regression

**LASSO** [Tibshirani (1996): J. Royal. Statist. Soc. B 58 (1): 267–288] :

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

# Supervised learning via penalised regression

## Vitality categorisation (“phenotype”):

- 0 – on “life support” and unlikely to survive
- 1 – hardly functional and deteriorating
- 2 – functional, with signs of improvement
- 3 – alive and well

# Supervised learning via penalised regression

The set of shifted (“phenotype”) response variables:

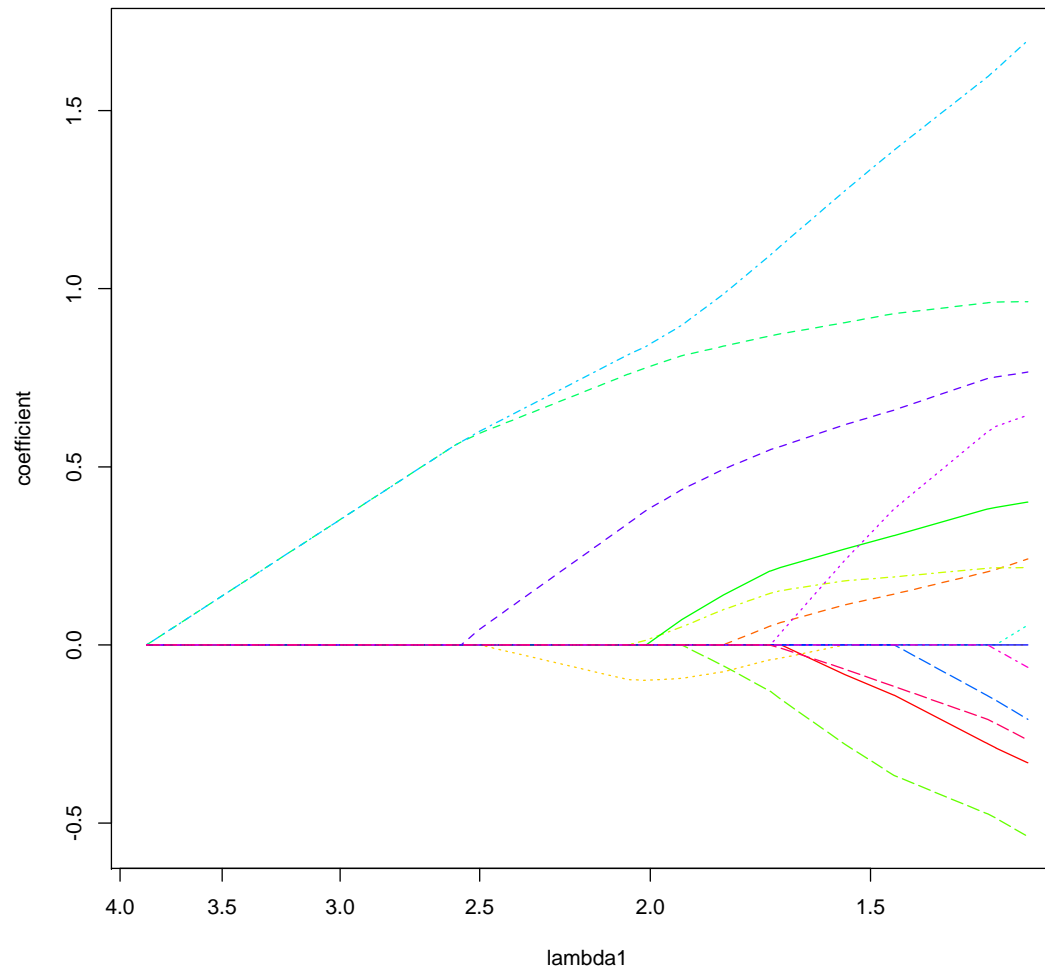
$y_{i,1}$  – refined “deprived”: **0 vs. 1-2-3**

$y_{i,2}$  – mid split: **0-1 vs. 2-3**

$y_{i,3}$  – refined “winners”: **0-1-2 vs. 3**

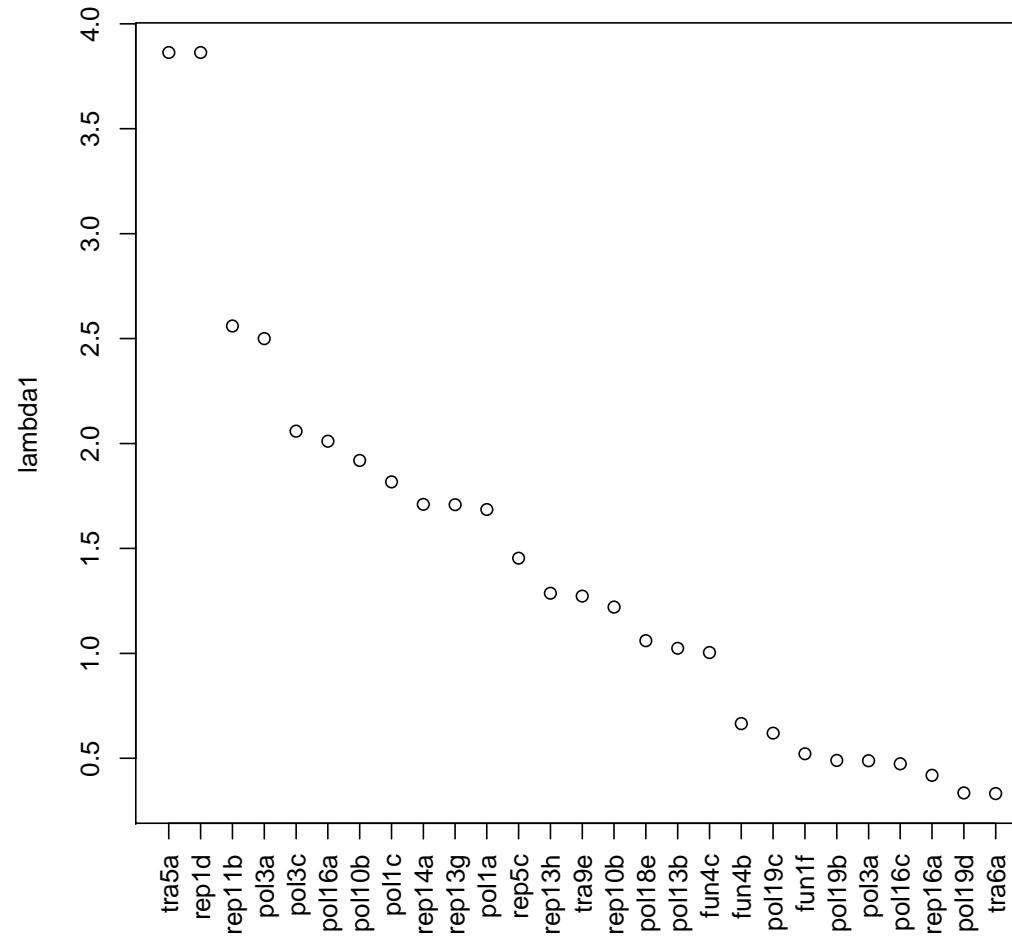
# Supervised learning via penalised regression

LASSO path



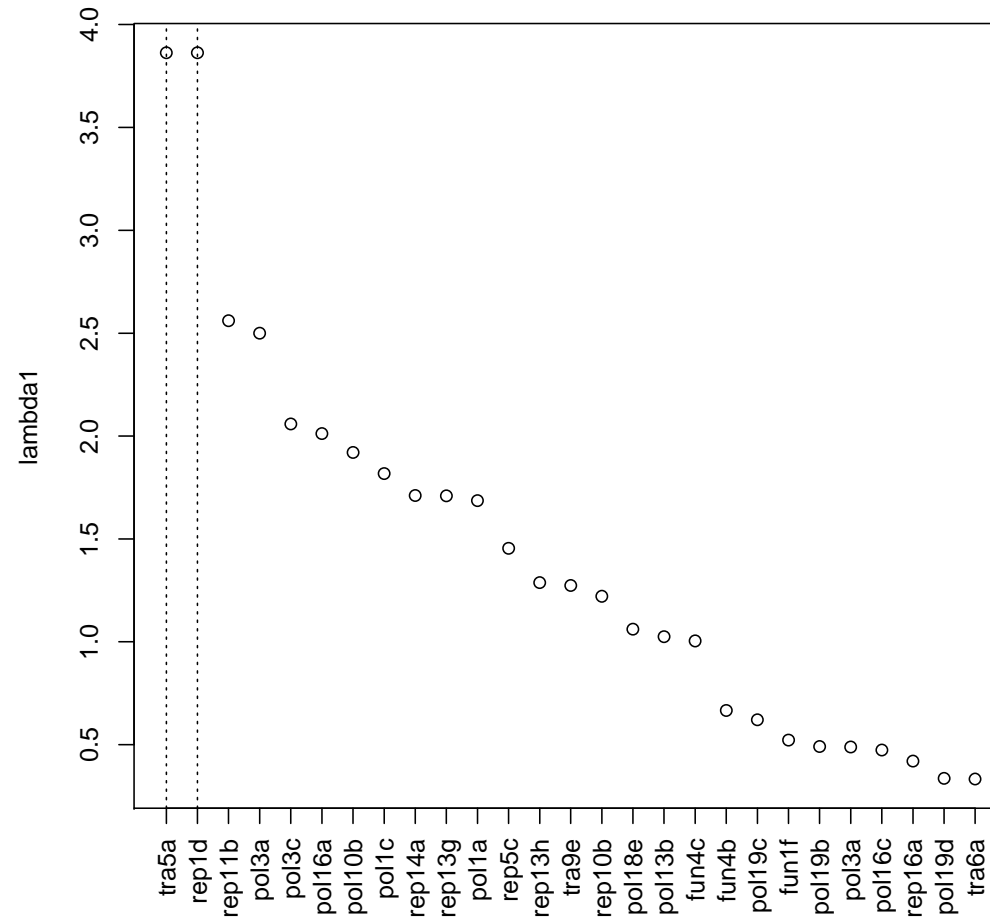
# Supervised learning via penalised regression

LASSO path: Entrance distances



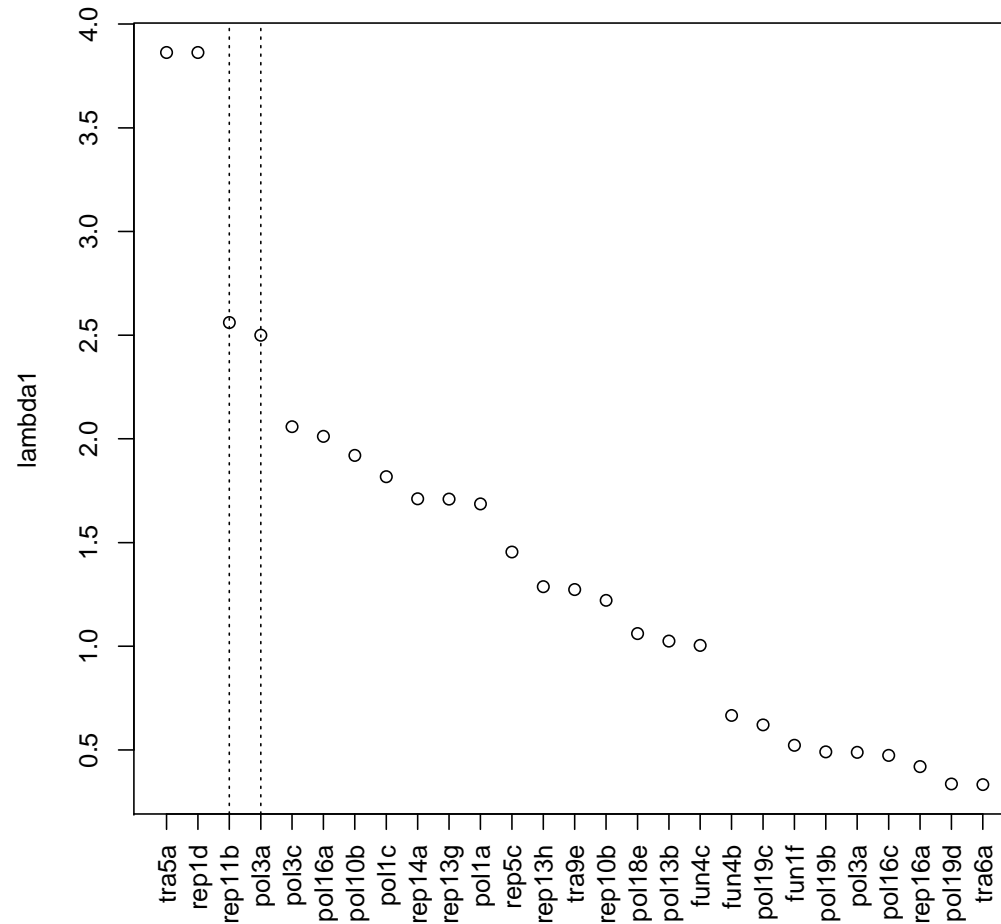
# Supervised learning via penalised regression

LASSO path: Entrance distances



# Supervised learning via penalised regression

LASSO path: Entrance distances



# Supervised learning via penalised regression

“Winners” refined	50-50	“Losers” refined
tra5a	pol3a	pol3a
rep1d	pol13a	pol10b
rep11b	fun5b	pol13a
pol3a	dev8a	rep11b
pol3c	tra9b	pol7a
pol16a	rep11b	fun1a
pol10b	pol1a	int6b
pol1c	rep14b	tra1c
rep14a	tra5a	fun10a
rep13g	pol16a	pol7d
pol1a	fun4d	pol19a
rep5c	pol16d	pol16d
rep13h	fun1a	rep3a
tra9e	pol1c	rep15a
rep10b	fun8g	tra1b
pol18e	pol10b	fun8g
pol13b	pol6k	fun4d

# Supervised learning via penalised regression

“Winners” refined	50-50	“Losers” refined
tra5a	<b>pol3a</b>	<b>pol3a</b>
rep1d	pol13a	<b>pol10b</b>
<b>rep11b</b>	fun5b	pol13a
<b>pol3a</b>	dev8a	<b>rep11b</b>
pol3c	tra9b	pol7a
pol16a	<b>rep11b</b>	fun1a
pol10b	pol1a	int6b
pol1c	rep14b	tra1c
rep14a	tra5a	fun10a
rep13g	pol16a	pol7d
pol1a	fun4d	pol19a
rep5c	pol16d	pol16d
rep13h	fun1a	rep3a
tra9e	pol1c	rep15a
<b>rep10b</b>	fun8g	tra1b
pol18e	<b>pol10b</b>	fun8g
pol13b	pol6k	fun4d

# Supervised learning via penalised regression

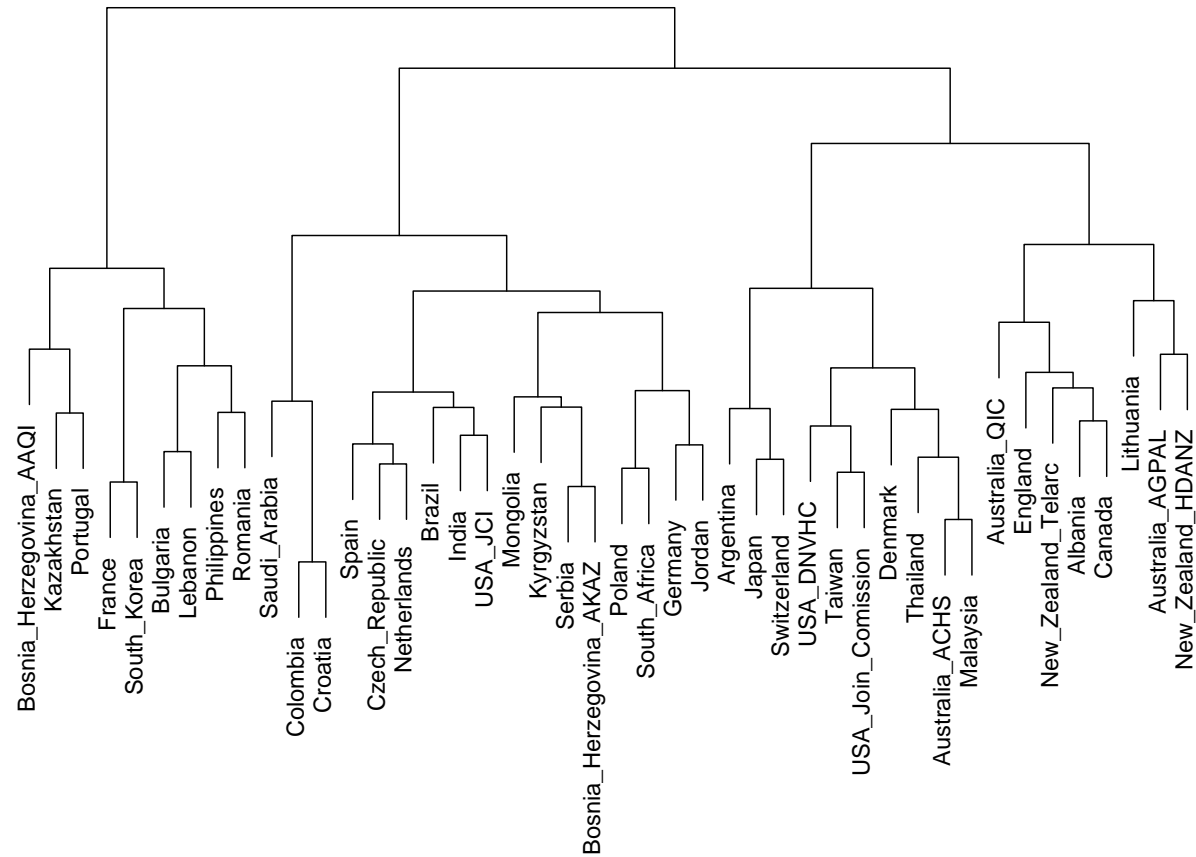
“Winners” refined	50-50	“Losers” refined
<b>tra5a</b>	pol3a	pol3a
rep1d	pol13a	pol10b
rep11b	fun5b	pol13a
pol3a	dev8a	rep11b
pol3c	tra9b	<b>pol7a</b>
pol16a	rep11b	fun1a
pol10b	pol1a	int6b
pol1c	rep14b	tra1c
rep14a	tra5a	fun10a
rep13g	pol16a	pol7d
pol1a	fun4d	pol19a
rep5c	pol16d	pol16d
rep13h	fun1a	rep3a
tra9e	pol1c	rep15a
rep10b	fun8g	tra1b
pol18e	pol10b	fun8g
pol13b	pol6k	fun4d

# Unsupervised learning via hierarchical clustering

- ✓ R function *daisy* (package *cluster*) has been used
- ✓ 207 categorical attributes have been coded as asymmetric binomials
- ✓ Dissimilarity matrix has been computed based on Gower's dissimilarity coefficient [Gower (1971) *Biometrics* 27, 857–874 ]
- ✓ Dendrograms are based on Ward's hierarchical clustering method
- ✓ Weights are imposed as an inverse of attribute counts within each category

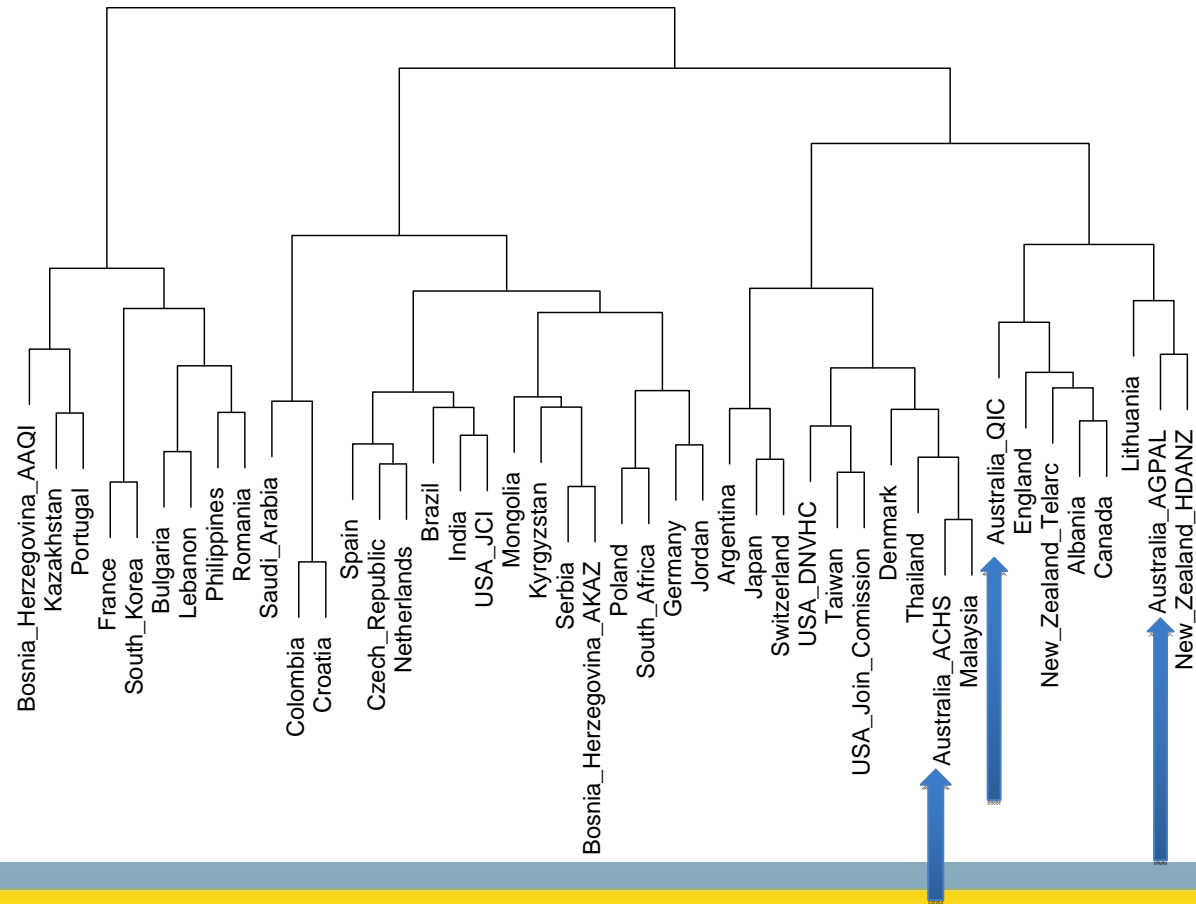
# Unsupervised learning via hierarchical clustering

All categorical attributes



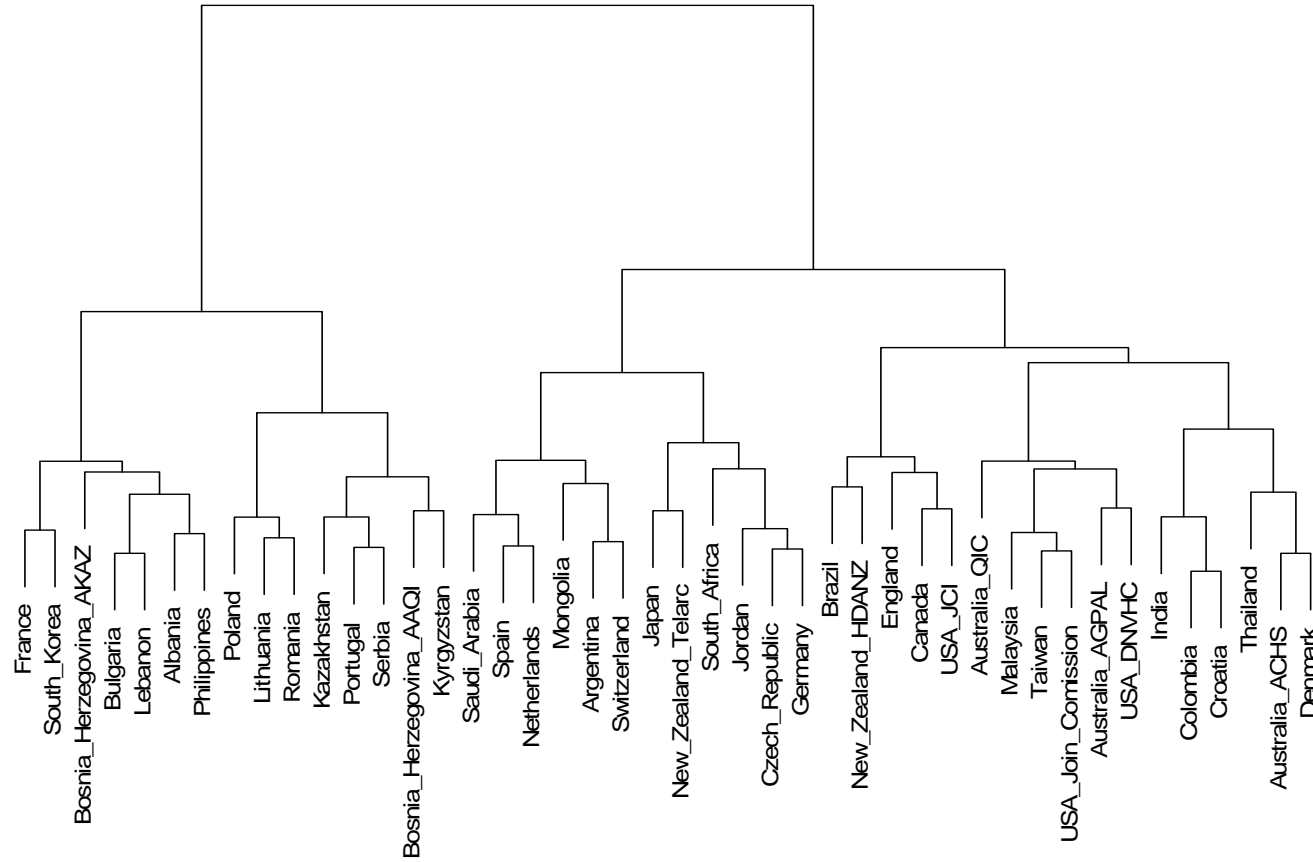
# Unsupervised learning via hierarchical clustering

All categorical attributes



# Unsupervised learning via hierarchical clustering

Policy and governance



# Example of results triangulation

- ✓ It has been identified that the legal status of organization is associated with the GDP PPP per capita factor (stratified analysis) => This association can be attributed to economic/social factors beyond the direct control
- ✓ Relatively wealthy Portugal fell into a “deprived” cluster in the Policy and Governance motivated dendrogram (cluster analysis)
- ✓ POL3a (What is the legal status of the accreditation organization?: government agency) has been identified as highly important attribute contributing to failure (LASSO)
- It can be recommended to change the legal status

# Challenges and extensions

- ✓ Missing variables
- ✓ Attributes interactions
- ✓ Nonlinear associations

# Challenges and extensions

- ✓ Missing variables
- ✓ Attributes interactions
- ✓ Nonlinear associations
  
- ✓ Risk factors analysis
- ✓ Predictive modeling
- ✓ Regulatory impact analysis

# Questions