Shining a light on NOIR - Rethinking Scales of Measurement

Chris Brunsdon

National Centre for Geocomputation
Maynooth University

contact: `christopher.brunsdon@nuim.ie`

Sep 2017 (`ECTQG@York`)

Steven's Original Scales:

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible statistics
NOMINAL	Determination of equality	Permutation group: $x' = f(x)$, f permutes x values	Number of cases, mode
ORDINAL	Determination of greater or less	Isotonic group: $x' = f(x)$ f any monotonic increasing function	Median, percentiles
INTERVAL	Determination of equality of intervals or differences	General linear group: $x' = ax + b$	Mean, standard deviation, rank order correlation
RATIO	Determination of equality of ratios	Similarity group: $x' = ax$	Coefficient of variation

Properties accumulate as columns are descended N-O-I-R. Restrictive on C3, expansive on C2,C4.

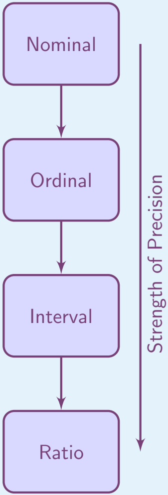
See Stevens, Stanley Smith. 1946. "On the Theory of Scales of Measurement." Science 103 (677–680)

A ‘permissible operator’ view

- $x \star y = z \iff f(x) \star f(y) = f(z)$

Scale	Permissible Operators ($\star$)
NOMINAL	$=, \neq$
ORDINAL	$>, <, \geq, \leq$
INTERVAL	$+, -$
RATIO	$\times, \div$

A Possible 'Nested' Arrangement ?



- Bad things come in threes
- ... if you count them in threes
- Scales of measurement are nested
- ... if you only look at the nesting scales
- Is the list universal?
- If not, what is missing?
- or is there anything else that slots in?

- ORDINAL $\Rightarrow$ (GRADED,RANK)
- Graded membership e.g. High, Medium, Low
- Rank - position in a race etc.
- In one respect the same - ie $\geq$ etc valid
- But also unique for each observation - no ties (mostly!)
- Rank-based statistics now meaningful for comparisons

Aside: Why are we doing this anyway?

- Driving force arguably comes from *Measurement Theory*
- An aspect of scientific thought eg Krantz, Luce, Suppes, and Tversky: Foundations of measurement vols I-III
- Chosen scale of measurement influence what kinds of analysis are meaningful
- Steven's uses the term 'permissible'
- Main idea is that results of analyses should be invariant if data is transformed by a permissible function f
- If $\mathcal{X}$ is the data then $\mathcal{A}(\mathcal{X}) = \mathcal{A}(f(\mathcal{X}))$
  - or possibly f$(\mathcal{A}(\mathcal{X})) = \mathcal{A}(f(\mathcal{X}))$

eg Temperature Data:

Year	Peak Daily Temperature - Week 1 of July (°C)						
2016	20.5	24.5	26.0	22.0	20.1	18.2	19.1
2017	22.4	20.1	18.7	19.2	19.1	20.3	22.7

Units	t-test $H_0 : \mu_a = \mu_b$	Mean (2016)	Mean (2017)	Δ% means
°C	$p = 0.38$	21.5 °C	20.4 °C	-5.3%
°F	$p = 0.38$	70.7 °F	68.6 °C	-2.9%

NB. 21.5 °C = 70.7 °F and 20.4 °C = 68.6 °F

This Gives Rise to a New Scale

- Example - Running shoes
- Is there a difference in the rate I run for shoes A and B?

Shoes	Units	Measurements									
A	Pace min/km	6.45	6.44	5.80	5.93	6.08	6.37	6.64	6.30	6.61	6.06
	Speed km/hr	9.30	9.32	10.34	10.12	9.87	9.42	9.04	9.53	9.08	9.90
B	Pace min/km	6.47	6.42	6.45	6.47	6.37	6.68	7.00	6.73	6.17	6.36
	Speed km/hr	9.28	9.35	9.30	9.28	9.42	8.98	8.58	8.92	9.72	9.43

- Should I choose pace or speed to test ?
- Both are measures of ‘rapidity’
- No obvious reason to favour one over the other

Looking at the test(s)

t-tests using both variables

Variable	*t*-statistic	D.F.	*p*-value
Speed	2.122	18	0.0480
Pace	-2.100	18	0.0501

Wilcoxon signed rank tests using both variables - replaces values by rank - 'demoting' the precision of information.

Variable	*W*-statistic	*p*-value
Speed	75	0.0630
Pace	25	0.0630

Is there a way to carry out a consistent test **without** loss of information and power?

The log interval scale

- Also proposed by Stevens
- Essentially $\log(x)$ is interval scale, not x
- Group structure is $f(x) = ax^b$, f is a permissible transform
- Note that pace $= 60 \times$ speed^{-1} so fits this structure
- So log(pace) and log(speed) are interval data, and t-test is permissible

t-tests using both variables **logged**

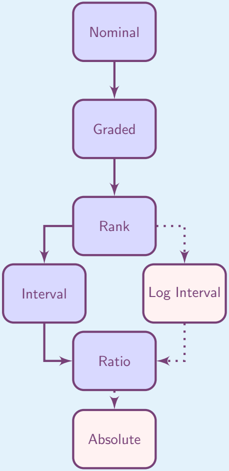
Variable	t-statistic	D.F.	p-value
Speed	2.112	18	0.0489
Pace	-2.112	18	0.0489

- Introducing this level of measurement leads to a better approach
- Note that it implies initial measurements only meaningful for $x > 0$

This brings focus to constrained measurements

- For log interval measurements we have $x > 0$
- Other constrained measurement levels exist:
  - probabilities $p \in [0, 1]$ (constrained in both directions)
  - counts n must be non-negative integers - $n \in \mathbb{Z}^+$
- Here the only permissible transform is $f(x) = x$ - the identity function
- This is the **absolute** scale

Augmenting NOIR

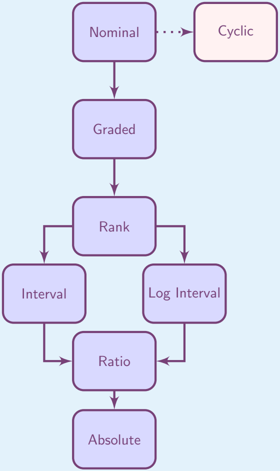


- The 'hierarchical' structure is gone
- A further thought for analysis - output statistic may be a different level of measurement than the data.
- So *p*-values (absolute) must be equal under permissible input transforms
- But means are measured at the same level as the input data, so can be equivalent under permissible interval or ratio transforms.

Measurement Scales for Statistics or Tests

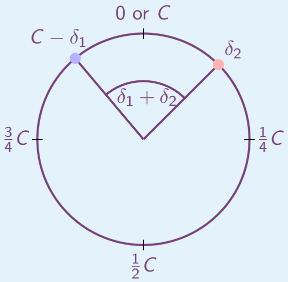
Statistic	Level of Measurement
Mean	Ratio or Interval
Quantiles	Rank, Graded, Interval, Log Interval, Ratio, Absolute
Standard deviation	Ratio ?
p-value	Absolute
Posterior Probabilities	Absolute

The Cyclic Measurement Scale



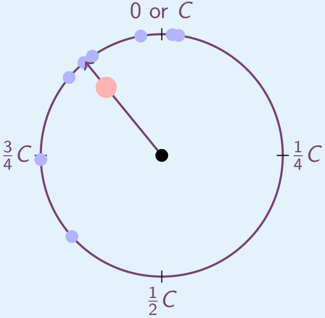
- Angles, Times of Day, Times of Year
- Difference between eg 359° and 357° same as between 359° and 1°
- Have a well-formed notion of $=, \neq, +, -, \times, \div$, but not $>, <, \geq, \leq$
- So in terms of NOIR they have some characteristics of Interval and Ratio data but not those of Ordinal

Cyclic Measurement Scales - Defining ‘Difference’

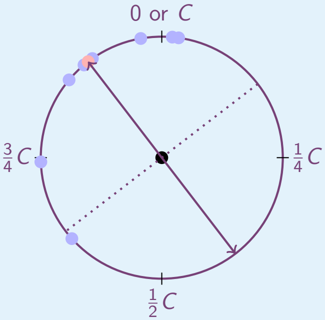


- Difference is not exactly the same for cyclic data
- Mean and circular variance also defined differently, but permissible.
- Quantiles not well defined - occasionally mean also undefined
- Median defined but not in terms of order - also sometimes undefined
- ... the latter if locations on the circle have centre of gravity at the centre of the circle.
- Also statistical tests exist eg for comparing two samples.

Circular Mean and Standard Deviation



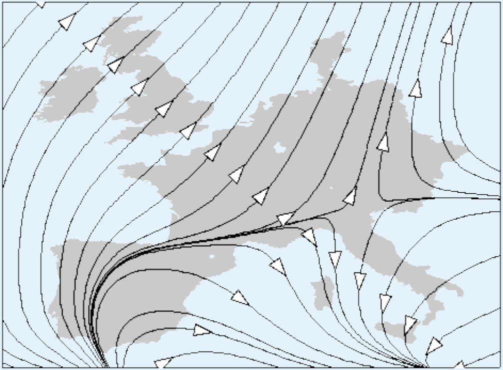
- Circular Mean:
- $\tilde{x} = \tan_2^{-1}\left(\sum_i \sin(x_i), \sum_i \cos(x_i)\right)$
- Circular SD:
- $\nu = \sqrt{-\ln\left(\left(\frac{1}{n}\sum_i \sin(x_i)\right)^2 + \left(\frac{1}{n}\sum_i \cos(x_i)\right)^2\right)}$



- Circular Median:
- If working in radians:
- $\text{argmin}_{\psi} \left\{ \frac{1}{n} \sum_{j=1}^{n} (\pi - |\pi - |\theta_j - \psi||) \right\}$
- ψ is any angle for which half of the data points lie in $[\psi, \psi + \pi)$ and the majority of points are nearer to ψ than $\psi + \pi$
- ψ may not be unique...

Example - Adding spatial weighting to circular means

- Moving Window Mean Directions ⇒ Streamlines
- NOAA Wind direction data



Proposed Alternative Lists of Levels

Tukey and Mosteller

1. Names
2. Grades (e.g. freshmen, sophomores etc.)
3. Counted fractions bound by 0 and 1
4. Counts (non-negative integers)
5. Amounts (non-negative real numbers)
6. Balances (any real number)

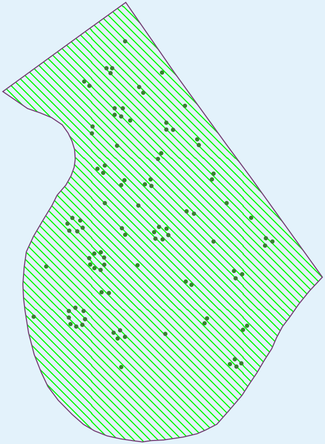
Chrisman

1. Nominal
2. Graded membership
3. Ordinal
4. Interval
5. Log-Interval
6. Extensive Ratio
7. Cyclical Ratio
8. Derived Ratio
9. Counts
10. Absolute

Some Further Extensions

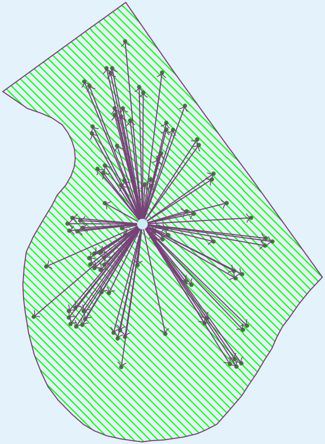
- Increased Dimension
  - Initially to 2D
  - Similar to direction, no $\geq$,$>$,$\leq$,$<$
  - Obviously important for geographers!
- Constraints
  - eg Values must be in positive (in $\mathbb{R}^+$), or in $[0,1]$ or an integer (in $\mathbb{Z}^+$)
  - Already there in Mosteller and Tukey or Chrisman implicitly
  - Look into this in a multidimensional context
- Partially ordered sets

2D Point Data



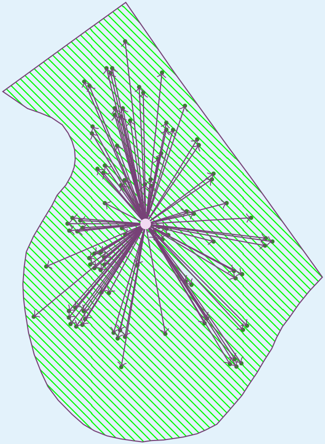
- Eg. locations of people sitting in Gordon square (near UCL)
- 2D measurements are 'integral' - eg easting on its own means little
- group structure is set of Euclidean transforms - combinations of:
  - Scaling
  - Rotation
  - Translation

2D Mean (Mean Centre)



- $\bar{\mathbf{x}} = \text{argmin}_{\mathbf{x}} \left[\sum_{i=1}^{n} (\mathbf{x} - \mathbf{x}_i)^2 \right]$
- So $\bar{\mathbf{x}}$ minimises squared distance to each of the data points
- Associated measure of spread:
- $D_s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\bar{\mathbf{x}} - \mathbf{x}_i)^2}$
- Standard distance - root mean squared distance from $\bar{\mathbf{x}}$ to data points.
- Both consistent under Euclidean transform

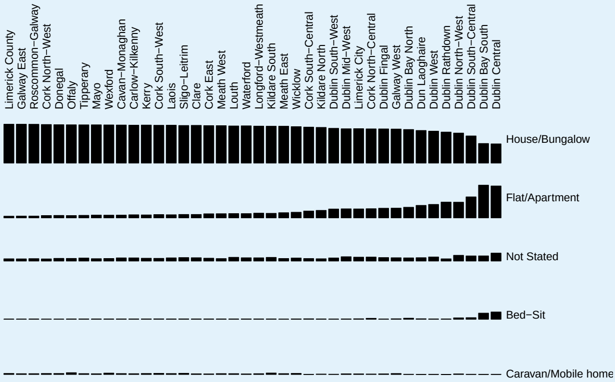
2D Median(Median Centre)



- $\tilde{\mathbf{x}} = \text{argmin}_{\mathbf{x}} \left[\sum_{i=1}^{n} |\mathbf{x} - \mathbf{x}_i|\right]$
- So $\tilde{\mathbf{x}}$ minimises summed absolute distance to each of the data points
- Associated measure of spread:
- $D_m = \text{median}\left(|\tilde{\mathbf{x}} - \mathbf{x}_i|\right)$
- Median distance - median distance from $\tilde{\mathbf{x}}$ to data points.
- Both also consistent under Euclidean transform

Thoughts on Medians

- They can be defined even for scales of measurement without $\geq, >, \leq, <$ operators
- ... on the basis of distance
- This also implies measures of spread
- ... based on this distance
- Generally (ie it needs proving!) if level of measurement may be ordered it corresponds to 50th percentile
- But it doesn't need this to be defined!



Household Type by Dáil Constituency There are a set of proportions for each constituency - they add up to one. That is, sums down columns below all add up to one - and all values must be greater than or equal to zero.

Permissible Transforms

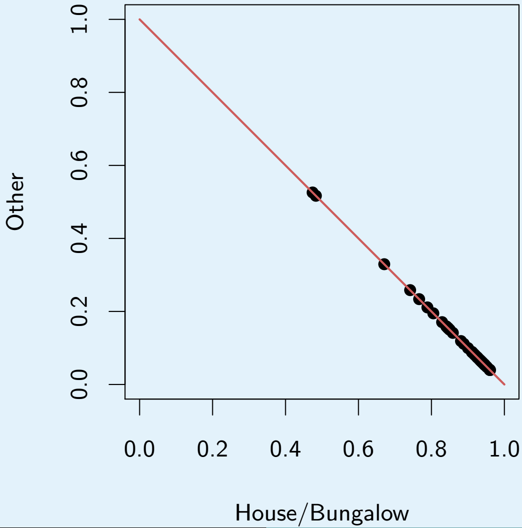
- Arguably none
  - Measurements are $\mathbf{p} = (p_1, p_2, \cdots, p_m)$ such that $p_1, p_2, \cdots, p_m \geq 0$ and $\sum_{j=1}^{m} p_j = 1$
  - Any translation, rotation, scaling etc. would violate these
- In this sense, there are similar to a multidimensional absolute level
- Note that the dimension of $\mathbf{p}$ is $m - 1$.
- Multidimensional mean and median as for 2D data still make sense
- If $p_1, p_2, \cdots, p_m$ meet constraints, so do their mean and median centres
- Weighted versions usually more useful
  - $\bar{\mathbf{x}} = \text{argmin}_{\mathbf{x}} \left[\sum_{i=1}^{n} w_i \left(\mathbf{x} - \mathbf{x}_i \right)^2 \right]$
  - $\tilde{\mathbf{x}} = \text{argmin}_{\mathbf{x}} \left[\sum_{i=1}^{n} w_i \left| \mathbf{x} - \mathbf{x}_i \right| \right]$

Means, Medians etc.

Statistic	Weighted	House/ Bungalow	Flat/ Apartment	Bed-Sit	Caravan/ Mobile	Not Stated
Median	N	0.829	0.142	0.007	0.003	0.019
	Y	0.827	0.144	0.008	0.003	0.019
Mean	N	0.871	0.106	0.003	0.003	0.017
	Y	0.870	0.107	0.003	0.003	0.017

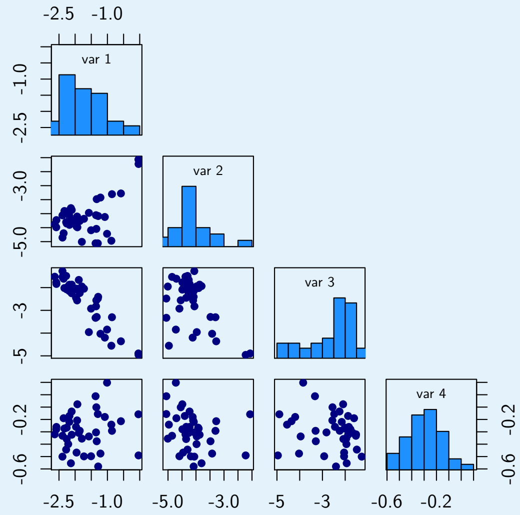
Inhomogeneity of Distance?

More ‘potential’ for large distances further away from the constraints – is a transformation onto $m - 1$ dimensional *unconstrained* space useful?



Proposed Approach (Aitchison)

- Isometric Log Ratios
- Firstly transform to $\left(\log(\frac{p_1}{g_{\mathbf{p}}}), \log(\frac{p_1}{g_{\mathbf{p}}}), \cdots, \log(\frac{p_m}{g_{\mathbf{p}}})\right)$
- Then express in as coordinmates with an $m - 1$ dimensional orthogonal basis
- Euclidean distances in this space correspond to an alternative measure of distance for $(p_1, p_2, \cdots, p_m)$ proposed by Aitchison
- An inverse transform exists
- Can compute mean and median on a distance basis in transformed space
- Then transform back to composition space


var 1
var 2
var 3
var 4
-2.5
-1.0
-5.0
-3.0
-5
-3
-0.6
-0.2

The recomputed summary statistics

Statistic	Weighted	House/ Bungalow	Flat/ Apartment	Bed-Sit	Caravan/ Mobile	Not Stated
Median	N	0.884	0.094	0.002	0.002	0.018
	Y	0.882	0.095	0.002	0.002	0.018
Mean	N	0.900	0.079	0.002	0.002	0.017
	Y	0.898	0.081	0.002	0.002	0.017

Discussion

- The *ilr* transform rather like log
- Transformed Data is a multidimensional measure scale
- Permitted transforms - Euclidean - rotation, translation
    - Similar to earlier 2D

Directional Data Revisited

- Also interpretable as 2D data?
- $\mathbf{x} = (x, y)$
- Constraint is $x^2 + y^2 = 1$
- Also a log connection:
- Complex representation as $e^{i\theta}$
- log of this is $i\theta$
  - this can work as interval scale
  - although inverse transform is cyclic
  - $e^{i\theta} = e^{i\theta + 2k\pi}$ if $k \in \mathbb{Z}$

Partially Ordered Sets - The Basics

- A partially ordered set has some pairs of members which $\preceq$ holds – but not **all** pairs

Properties of $\preceq$ and friends

1. $a \preceq a$ (Reflexivity)
2. If $a \preceq b$ and $b \preceq a$ then $a = b$ (Antisymmetry)
3. If $a \preceq b$ and $b \preceq c$ then $a \preceq c$ (Transitivity)
4. $a \prec b$ implies $a \preceq b$ but $a \neq b$
5. $b \succ a$ means the same as $a \prec b$

Take-home $\prec$ etc. work like comparison operators like $<$ etc. but **only** on **some** pairs of objects...

An Example Data Set

Data Description

Indicator	Name	Description
s_1	Income	Per capita income (1974)
s_2	Illiteracy	Illiteracy (1970 percept of popn.)
s_3	LifeExp	Life expectancy in years (1969-71)
s_4	Murder	Murder and non-negligent manslaughter rate per 100,000 popn. (1976)
s_5	HSGrad	Percent high-school graduates (1970)

Table: US Well-being variables by State

US states data as a poset

Definition of $\preceq$ etc here:

For US states a and b, $a \preceq b$ if and only if $s_{1a} \leq s_{1b}$ and $s_{2a} \geq s_{2b}$ and $s_{3a} \leq s_{3b}$ and $s_{4a} \geq s_{4b}$ and $s_{5a} \leq s_{5b}$

- $a \preceq b$ implies state b is 'doing better' that state a on all indicators.
- States no longer fully rankable, but some still precede others
- Only requires consensus on sign of variables, not on weighting

Visualising the US poset:

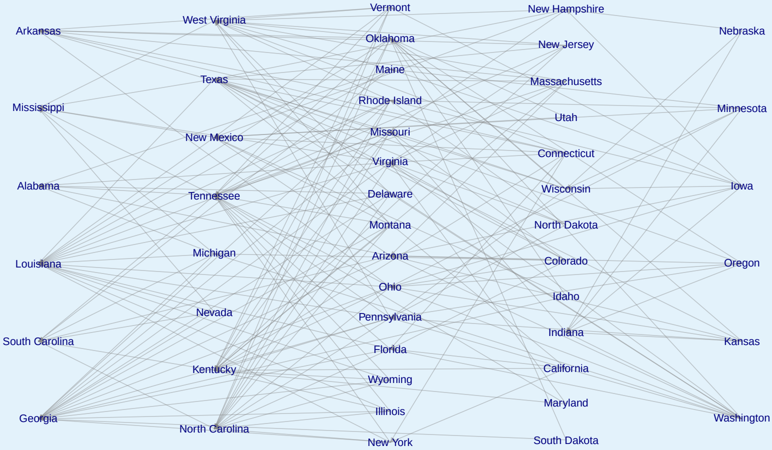


Figure: Hasse Diagram (Peeled Minimal Elements)

Some terminology

- A *chain* $\mathcal{C} \subseteq \mathcal{P}$ is a set such that all a, b in $\mathcal{C}$ are comparable. Note that a chain is therefore an ordered set. A chain is *maximal* if no other chain $\mathcal{C}'$ exists such that $\mathcal{C} \subset \mathcal{C}'$.
- The *depth* of a poset $\{\mathcal{P}, \preceq\}$ is the length of its longest chain.
- An *antichain* $\mathcal{A} \subseteq \mathcal{P}$ is a set such that no distinct a, b in $\mathcal{A}$ are comparable. An antichain is *maximal* if no other antichain $\mathcal{A}'$ exists such that $\mathcal{A} \subset \mathcal{A}'$.
- An element $a \in \mathcal{P}$ is a *maximal element* if there is no element $b \in \mathcal{P}$ such that $a \preceq b$. The *maximal element set* is the set of all such elements. Similar for *minimal*—

Geographical Hasse Diagram is Revealing. . .

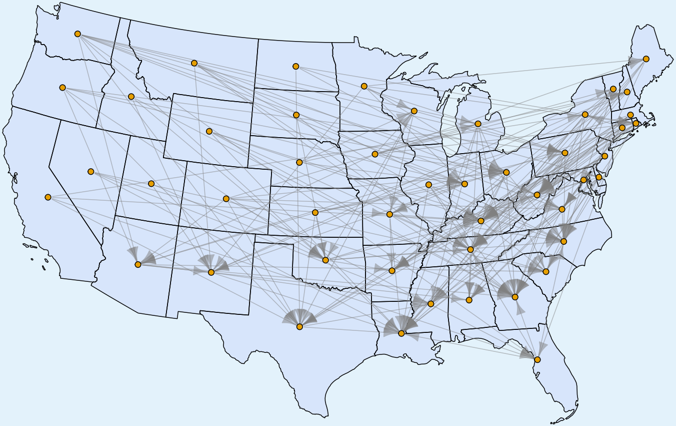


Figure: Hasse Diagram (Based on Geographical Location)

"In general states in the north west tend to enjoy a better state of well being (at least on the basis of this index) . . ."

Some chains of well-being

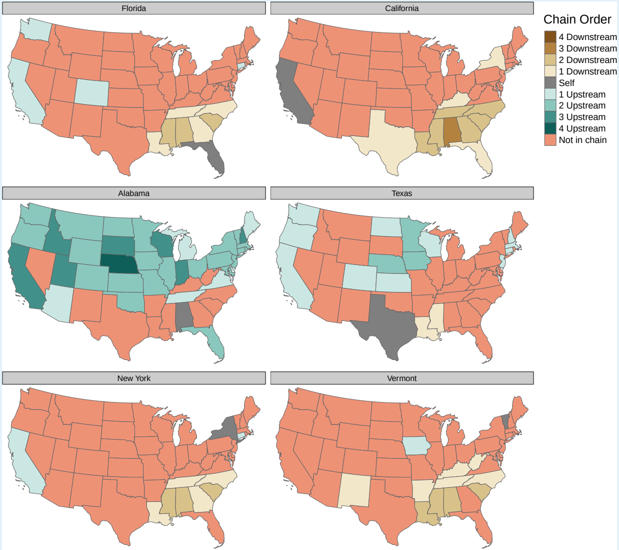


Figure: State-focused Relationship Maps

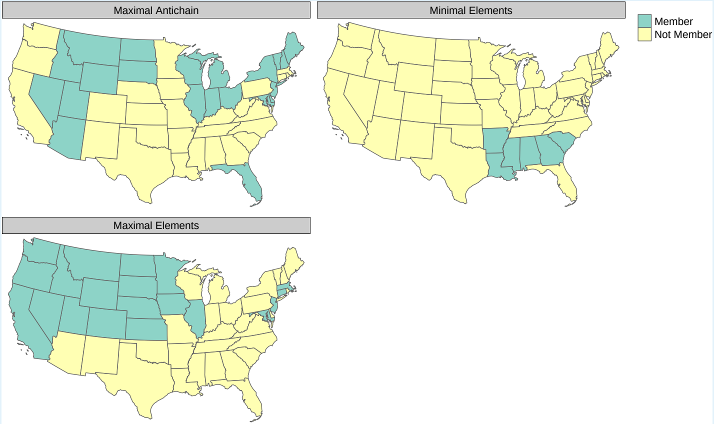


Figure: Significant Set Maps

Do these sets cluster?

... It looks like they do, at least here

	Minimal Elements	Maximal Elements	Maximal Antichain
Join Count statistic	5.043	4.076	2.817
p-value	0.000	0.000	0.002

Tobler revisited?

"not everything is comparable to everything else, but near things are less likely to be comparable than distant things."

Are Measurement Theory and Steven's Scales Helpful Anyway?

- Idea is not without its critics
- The original simple idea would be helpful if comprehensive
  - But it isn't !
  - Especially for geographers ...
  - Rather like 'i' before 'e' except after 'c'
  - "My neighbour is agreeing to reimburse the conciege with madiera and caffeine."
  - So many contradictions, hardly a structural rule...

Against Proscription

- Previous points were a critique of Steven particular categorisation
- ... but not of measurement theory *per se*
- Are there times when it malkes sense to use an analysis technique that isn't permissible?
- A lot of non-parametric statistical methods do this

Counterexamples

- Spearman's Rank Correlation Coefficient
  - equivalent to Pearson's coefficent applied to ranks
  - ... calculating means and variances of ranks - NOT PERMISSIBLE!
- Wilcoxon Rank Sums test
  - ADDING RANKS NOT PERMISSIBLE!

Tensions between Measurement Theory and Statistical Models

- Doesn't statistical modelling make this redundant?
- Choosing a log interval scale might imply *t*-tests on logged data
  - But so would a log-normal model
- Indeed although logs in the running example ensures an invariant *p*-value
  - ... it would be numerically incoorect of model assumption not true
- Also it is quite possible to derive the distribution of a sum of ranks
  - ... even though measurement theory says this is meaningless !

But in some ways **not** meaningless. . .

- Higher average ordinal score does imply more high ranking scores
- Its just that difference don't make sense -
  - $4.5 > 4.2$ but
  - 4.5 to 4.2 is not the same as 3.5 to 3.2 or 1.9 to 1.6 . . .
- Similar ID numbers may be thought of as nominal BUT
  - If allocated in sequence they may be a proxy for ordinal time
- Floor on an apartment block is ordinal, but could be ratio if all floors same height
  - It depends on context as well as measurement level

Some quotes. . .

- “Permission is not required in data analysis.”
- “If a mathematician gives or witholds ‘permission’ . . . , he (sic) may be accessory to helping the practioner escape the reality of defining the research problem.”
  - Guttman, 1977
- “Experience has shown that in a wide range of situations that the application of proscribed statistics to data can yield results that are scientifically meaningful, useful in making descisions, and valuable as a basis for further research.”
  - Velleman and Wilkinson, 1993

Perhaps if not axiomatic, still sometimes helpful ?

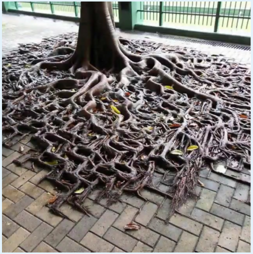
- Ultimately need to think about research questions of themselves
- Not the scale of measurement of data used to investigate them
- They can occasionally provide useful guidelines, though
  - Thus although means of Likert scales can be compared, they do not convey the full richness of interval or ratio means
- Possibly the idea of *casting* as in the C programming language is useful
- `x = (float) i` or `n = (int) y` convert data of one type to another
  - but sometimes with a loss of detail, or future flexibility

Occasionally the idea of Measurement Scale is Food For Thought

- Spearman's Rank Correlation Coefficient flawed in measurement scale terms but Kendall's τ coefficient isn't
- Kendall - suppose we have two variables for each case i; x_i and y_i. If we choose two cases at random, say j and k
- let $p = \Pr(x_i > x_j \text{ and } y_i > y_j)$
- then $\tau = 2 * p - 1$
- Only uses $>$, no means etc. - therefore fine for any ordered levels.
- Can make a local statistic out of it if a location $\mathbf{l}$ and radius r is associated with p and a further condition that observations i and j are with a distance r from $\mathbf{l}$.

Final Thoughts

- NOIR was a useful starting point but
- actually a lot more going on - suggested revised diagram below



- I haven't covered all possible measurement scales here
- Perhaps an axiomatic approach is unhelpful
- But viewed as one way to assess analysis it has some uses. . .
- But perhaps we need to move beyond NOIR as quantitative *and* theoretical geographers