تصنيف الجار الاقرب في الجغرافيا

ترجمة بتصرف وتعليق
أ.د. مضر خليل عمر

ملاحظة ،

هذه ترجمة بتصرف لكتيب صادر عن معهد الجغرافيين البريطانيين IBG ضمن سلسلة CATMOG برقم (6) عنوانه : التصنيف Classification ، لمؤلفه R . J. Johnston ، 1976 . قد يقول البعض بان برامج الحاسبة الاحصائية (SPSS و Minitab مثلا) و تقنيات نظم المعلومات الجغرافية GIS والتحليل المكاني Spatial Analysis تغنينا عن العمل اليدوي . هذا صحيح ، ولكن يبقى جوهري التعرف على طبيعة التقنية واسلوب عملها ليقوم الباحث بتفسر النتائج بشكل سليم ، فالباحث عليه اختيار البرامج و التقنيات المناسبة وتوجيهها بما يخدم هدف الدراسة لا ان تقوده هي و توصله الى موقع يجهل كيف وصل اليه وتضعه في موقف حرج .

مقدمة

يهدف أي نشاط علمي إلى الفهم ؛ فبمجرد أن نتمكن من تفسير أسباب وقوع أحداث معينة (ومتى وأين) *((وكيف))* ، قد نتمكن من التحكم فيها ، بما يُسهم في بناء عالم أفضل . وهكذا ، **يتضمن العلم تطوير منهجيات تنبؤية تُنتج تعميمات ونظريات وقوانين** . *((يشكل البحث العلمي الرصين لبنة مضافة في عملية بناء معرفة علمية توصل الى تعميم منطقي و استدلال منهجي بقصد بناء نظرية و صياغة قوانين .))* لا تنطبق هذه التعميمات على حالات أو أحداث محددة ، بل على فئات من الظواهر؛ فنحن نضع نظريات وقوانين ليس لتيار واحد قبل العصر الجليدي ، بل لجميع هذه التيارات . بعبارة أخرى ، نُعرّف فئةً نُطلق على أعضائها اسم التيارات قبل العصر الجليدي ؛ ونُسلّم بأن لكل تيار من هذه التيارات خصائصه الخاصة - كموقعه مثلاً - ولكننا نُجادل بأن له خصائص مشتركة كافية مع جميع التيارات قبل العصر الجليدي الأخرى ، مما يُتيح لنا دراستها كمجموعة .

إن تبني هذه الفلسفة العلمية يُثير مباشرةً مسألة تعريف الفئة : ما هي العناصر التي يُمكننا تجميعها معًا لتشكيل فئة ظاهرة مُحددة ؟ قد يُحدد لنا نموذج منطقي ، أو استنتاجي ، لموضوعنا فئاتنا ، كما هو الحال مع مجموعة القواعد التي وضعها علماء شكل الأرض لترتيب المجاري المائية : يُقدم جريج (1965) مقدمة عامة لهذا الأسلوب من البحث العلمي . ولكن ماذا لو لم تكن لدينا مجموعة قواعد محددة مسبقًا ، واضطررنا إلى العمل بطريقة استقرائية ، وتحديد الفئات بين الظواهر و/أو الأشياء التي لاحظناها ؟ بالنسبة للمجاري المائية ، على سبيل المثال ، قد نرغب في دراستها ليس فقط وفقًا لترتيبها في نظام ترتيب ستراهلر، ولكن أيضًا من خلال سمات تصريفها ، وملامح ثالوجها ، وما إلى ذلك . بالنسبة لهذه المتغيرات الأخيرة ، ربما لا توجد لدينا فئات محددة مسبقًا - وإذا كانت لدينا ، فمن المحتمل أنها تستند إلى خطوط حدودية عشوائية . ولذلك ، يجب أن نجد وسيلة لتجميع المجاري المائية معًا وفقًا للمتغيرات التي نقرر استخدامها . هذه هي عملية التصنيف ، التي يتم من خلالها تخصيص الكائنات الفردية ، بأكبر قدر ممكن من الموضوعية ، إلى فئات منفصلة .

تقوم فلسفة العلم على قبول مبدأ التصنيف ، القائل بوجود مجموعات من الظواهر / الأشياء المتشابهة يمكن عدها وحدة واحدة بغرض التوصل إلى تعميمات صحيحة حول جوانب سلوكها . ومن الواضح أن هناك العديد من التصنيفات المختلفة الممكنة ، وذلك تبعًا لهدف الدراسة المُجراة : فالتيارات ، على سبيل المثال ، تُشكل فئة واحدة ، ولكن يمكن تقسيمها بناءً على معايير متنوعة ، مثل الترتيب والعمق والتدرج ، وما إلى ذلك . والغرض من إجراءات التصنيف هو توفير تجميع مناسب للنشاط العلمي المُجرى : **يجب أولًا تحديد معايير التصنيف** ، ثم تصنيف الأشياء التي تُقاس على أساسها إلى فئات .

ويتعلق الجزء الأول من هذا الإجراء بالمجال الموضوعي ؛ أما الجزء الثاني - اختيار عملية التصنيف وتنفيذها - فهو محور هذه الدراسة .

للجغرافيا تاريخ طويل في مجال التصنيف ، ولعل أفضل الأمثلة عليه تأتي من علم المناخ . *((وقد البعض الجغرافيا علم التصانيف ، لكونها لا تركز على الحالات المفردة ، فمنهج دراسة المجموعة معروف في الجغرافيا بفرعيها البشرية و الطبيعية))* . قام العديد من الباحثين بتعريف المناطق المناخية . وكانت بعض جهودهم بمثابة تمهيدات لأعمال تنبؤية ، على سبيل المثال ، حول العلاقات بين المناخ وإنتاجية المحاصيل . بينما بُذلت جهود أخرى لأغراض تعليمية فقط ، كوسيلة لتعميم خريطة المناخ المعقدة على أرض الواقع ، التي تتوفر لها كميات هائلة من المواد .

لا أحد منا قادر على استيعاب الكم الهائل من المواد المتاحة في السجلات المناخية اليومية ، أو الكم الهائل من مجلدات التعداد العام للسكان ، أو الصور الجوية والفضائية التي لا تعد ولا تحصى والتي تملأ مكتباتنا بمعدلات متزايدة . وقد نكتفي بأخذ عينات من مجموعات المعلومات ذات الصلة ، أو قد نحاول تقليل "الحمل الزائد للمعلومات" ، تمامًا كما نختزل ، على سبيل المثال ، الأرقام المتعلقة بنسبة القوى العاملة في كل مقاطعة إنجليزية إلى بضع فئات مئوية ، مما يتيح إنشاء خريطة كوروبليث بسيطة تشير إلى السمات البارزة للنمط الوطني .

لذلك ، هناك سببان للتصنيف . **الأول** يتضمن عامل مقياس ، يختزل عددًا كبيرًا من الأفراد إلى عدد صغير من المجموعات ، لتسهيل الوصف والتوضيح . في مجال الجغرافيا ، تُعد الأدبيات الواسعة حول التصنيف الوظيفي للمدن (سميث ، 1966) مثالًا على هذا العمل . يتضمن **الثاني** تعريف الظواهر - وهي فئات تُشتق منها عبارات عامة ، كما هو الحال في العمل على الأماكن المركزية ، *((نظرية كرستالر))* حيث تُعرّف فئات المدن كمقدمة لدراسة مواضيع مثل تباعد المدن وسلوك المستهلك . أيًا كان الغرضان المناسبان ، فإن الطرق واحدة ، لأن المبادئ متطابقة : **دمج أفراد متفرقين مع أفراد متشابهين لإنتاج مجموعات** . تتوفر العديد من مجموعات القواعد البديلة لتطبيق هذه الطرق . يعتمد الاختيار بينها على عوامل عدة ، مثل طبيعة قياس الأشياء قيد الدراسة وأهداف التصنيف . ومع ذلك ، يوجد نوعان رئيسيان (تصنيف إجراءات ، التصنيف) ، ويُطلق عليهما الطرق التجميعية والتقسيمية - وسيتم تقديمهما في الأقسام الاتية .

الطرق التجميعية

تبدأ هذه الطرق بعدد كبير من الأفراد ، ثم تمضي قدمًا ، باستخدام مجموعة من القواعد المحددة مسبقًا ، لتوزيعهم على مجموعات وفقًا لمستويات التشابه طبقا لمعايير مختارة . ربما يكون الإجراء الأمثل هو أن يُنتج الباحث كل تجميع ممكن ، ثم يُقرر أيهما الأنسب لغرضه المُحدد . لكن هذا يكاد يكون مُستحيلًا في أغلب الأحيان ، نظرًا لكثرة الأفراد المعنيين . على سبيل المثال ، إذا كان من المقرر توزيع أربعة أفراد على إحدى مجموعتين ، فهناك سبع تركيبات مُمكنة ، كما مبين في ادناه :

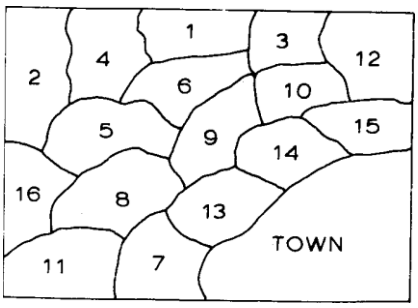
Combination	Group 1	Group 2
A	1,2	3,4
B	1,3	2,4
C	1,4	2,3
D	1	2,3,4
E	2	1,3,4
F	3	1,2,4
G	4	1,2,3

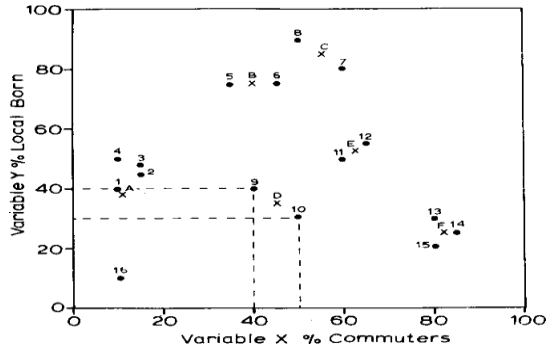
إذا كان هناك خمسة أفراد ، فهناك 15 احتمالاً (احسبها) ، وهكذا . وبما أن التصنيف ينطوي على الأرجح على أعداد كبيرة من الأفراد ، فإن البحث في مجموعة الاحتمالات الكاملة للوصول إلى الخيار الأمثل أمرٌ مستحيلٌ بلا شك (حتى باستخدام جهاز كمبيوتر عالي السرعة) . لذلك ، اضطر الباحثون المهتمون بالتصنيف ، في مجموعة واسعة من التخصصات ، إلى تطوير أساليب (غالبًا ما تُسمى أساليب التصنيف العددي) ، تُنتج "أفضل" تصنيف لمجموعة بياناتهم الخاصة . ونتناول هنا بعض هذه الأساليب - الأبسط عمومًا .

تحليل الارتباط الأولي

لدينا بيانات حول متغيرين لست عشرة قرية تقع إلى الشمال الغربي لمدينة كبيرة (الشكل 1) ، والمتغيران هما % للقوى العاملة في القرية التي تنتقل إلى المدينة للعمل (المتغير X) ، و% لسكان القرية الذين ولدوا داخل المنطقة المحلية (المتغير Y) . يوضح الرسم البياني التشتتي ثنائي الأبعاد (الشكل 2) قيم كل قرية على هذين المتغيرين ؛ *(اسقاط قيم X بدلالة قيم Y)* مهمتنا هي تصنيف القرى إلى مجموعات تحتوي على أعضاء متشابهين في موقعهم على المحورين . المهمة الأولى هي إنتاج مقياس لدرجة التشابه بين كل زوج من القرى . يتم ذلك باستخدام نظرية فيثاغورس المعروفة من الهندسة الإقليدية ، والتي تنص على أن مربع المسافة على وتر المثلث القائم الزاوية يساوي مجموع مربعي المسافتين على الضلعين الآخرين (الشكل 3) . وبالتالي ، وبشكل عام :

شكل (1)





مجموعة افتراضية لستة عشر قرية

$$D_{ij} = \sqrt{(i_x - j_x)^2 + (i_y - j_y)^2} \quad \text{.....(1)}$$

where i_x, j_x = the values for i and j on variable x

i_y, j_y = the values for i and j on variable y

and D_{ij} = the distance between i and j

Figure 2 shows how we apply equation (1) to finding the distance between places 9(i) and 10(j). The term $(i_x - j_x)^2$ is $(40 - 50)^2$ or 100. Similarly $(i_y - j_y)^2$ is $(40 - 30)^2$ or 100, so $D_{ij} = \sqrt{100 + 100}$, which is 14.1.

PYTHAGORAS' THEOREM

$$AC^2 = AB^2 + BC^2$$
$$= 3^2 + 5^2$$
$$= 9 + 25 = 34$$
$$\therefore AC = \sqrt{34} = 5{\cdot}831$$

شكل (3)

نظرية فيثاغورس

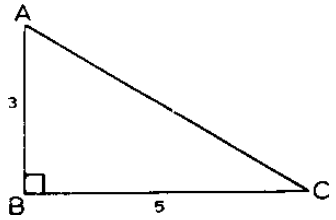


Table 1 Distances between places on Figure 2.

	Place															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	7.1	9.4	10.0	43.0	49.5	64.0	64.0	30.0	41.2	51.0	57.0	70.0	76.5	72.8	30.0
2	7.1	0	3.0	7.1	36.1	42.4	57.0	57.0	25.5	38.1	45.3	51.0	66.7	72.8	69.6	35.4
3	9.4	3.0	0	5.4	33.6	40.4	55.2	54.7	26.3	39.4	45.1	50.5	67.5	73.7	70.8	38.3
4	10.0	7.1	5.4	0	35.4	43.0	58.3	56.6	31.6	44.7	50.0	55.2	72.8	79.1	76.2	40.0
5	43.0	36.1	33.6	35.4	0	10.0	25.5	21.2	35.4	47.4	35.4	36.1	63.6	70.7	71.1	69.6
6	49.5	42.4	40.4	43.0	10.0	0	15.8	15.8	35.4	45.3	29.2	28.3	57.0	64.0	65.2	73.8
7	64.0	57.0	55.2	58.3	25.5	15.8	0	14.1	44.7	51.0	30.0	25.5	53.9	62.1	63.3	86.0
8	64.0	57.0	54.7	56.6	21.2	15.8	14.1	0	51.0	60.0	41.2	38.1	67.1	73.8	76.2	89.4
9	30.0	25.5	26.3	31.6	35.4	35.4	44.7	51.0	0	14.1	22.4	29.2	41.2	46.4	44.7	42.4
10	41.2	38.1	39.4	44.7	47.4	45.3	51.0	60.0	14.1	0	22.4	29.2	30.0	35.4	31.6	44.7
11	51.0	45.3	45.1	51.0	35.4	20.2	30.0	41.2	22.4	22.4	0	7.1	28.3	35.4	36.1	64.0
12	57.0	51.0	50.5	57.0	36.1	28.3	25.5	38.1	29.1	29.2	7.1	0	29.2	36.1	38.1	71.1
13	70.7	66.7	67.5	70.7	63.6	57.0	53.9	67.1	41.2	30.0	28.3	29.2	0	7.1	10.0	72.8
14	76.5	72.8	73.7	76.5	70.7	64.0	62.1	73.8	46.4	35.4	35.4	36.1	7.1	0	7.1	76.5
15	72.8	69.6	70.8	72.8	71.1	65.2	68.3	76.2	44.7	31.6	36.1	38.1	10.0	7.1	0	70.7
16	30.0	35.4	38.3	40.0	69.6	73.8	86.0	89.4	42.4	44.7	64.0	71.1	72.8	76.5	70.7	0

يمثل كلٌّ من i وj بدورهما ، ست عشرة قرية ، فنحصل على مصفوفة 16 × 16 للمسافات بين القرى ، والموضحة في الجدول 1 . القيم على القطر الرئيسي لهذه المصفوفة (من أعلى اليسار إلى أسفل اليمين) تساوي صفرًا ، مما يشير إلى أن القرية i تُشبه القرية j تمامًا ، عندما i = j؛ والقيم على يمين

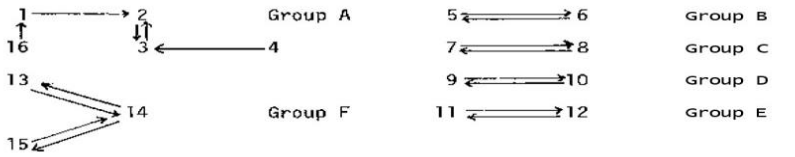
القطر الرئيسي هي صورة طبق الأصل لتلك الموجودة على يساره ، مما يعني أن لدينا مصفوفة مربعة متماثلة . يمكننا الآن الانتقال إلى التصنيف ، باستخدام معيار تجميع كل مكان مع المكان الأكثر تشابهًا معه . (سيكون هذا هو أقرب جار له في مخطط التشتت في الشكل 2). يُحدَّد أقرب جار من مصفوفة المسافة بإيجاد الصف الذي يحمل أصغر قيمة فيه لكل عمود - باستثناء القيم الصفرية ، على طول القطر الرئيسي ، والتي تشير إلى المسافة إلى الذات . وبالتالي ، بالنسبة للعمود 1 في الجدول 1، أصغر قيمة هي 7.1 في الصف 2؛ وأقرب جار لـ I هو 2 . المجموعة الإجمالية لأقرب الجيران هي : يمكن إظهار المجموعات من خلال رسم تخطيطي ، حيث يربط السهم كل مكان بأقرب جار له ، على النحو الاتي : تم تصنيف القرى الستة عشر إلى ست مجموعات ؛ والتشابه بين أزواج القرى هو المسافة بينها على الإحداثيات الديكارتية للرسم البياني (الشكل 2) ، ومعيار التصنيف يخصصها لمجموعات مع أقرب جيرانها ، أي أولئك الذين تشبههم أكثر.

Column	1	2	3	4	5	6	7	8
Smallest row value is in row	2	3	2	3	6	5	8	7

Column	9	10	11	12	13	14	15	16
Smallest row value is in row	10	9	12	11	14	13 15	14	1

(Note that place 14 has two equidistant nearest neighbours)

The groups can be shown by a diagram, in which an arrow links each place to its nearest neighbour, as follows:



E78.3 average = 78.3/4 = 19.6

Pairs of Places	Distance
5 - 7	25.5
5 - 8	21.2
6 - 7	15.8
6 - 8	15.8

هذه المسافات غير القطرية هي في الواقع المسافات بين مراكز مجموعات المجموعات ، وهي البؤر المركزية للمجموعات (كما هو موضح بـ Xs في الشكل 2) . في المجموعة ثنائية النقاط ، يقع مركز الثقل في منتصف المسافة بين العناصر؛ أما في المجموعات التي تضم أكثر من عنصرين ، فيقع مركز الثقل عند النقطة التي يبلغ عندها مجموع المسافات إلى جميع العناصر أصغر قيمة ممكنة . (سنتجاهل حاليًا القيم على القطر الرئيسي في الجدول 2) . وبالتالي، يمكننا الاستمرار في إنتاج مصفوفة المسافة في الجدول 3 ، وتجميع المجموعات بحيث تصبح القرى الست عشرة الآن جزءًا من مجموعة كبيرة واحدة . تُظهر القيم على طول القطر الرئيسي (في المربعات) متوسط المسافات داخل المجموعة.

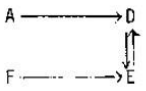
Table 2 Inter-group distances from Table l a

	Group					
	A	B	C	D	E	F
(members)	(1-2-3-4-16)	(5-6)	(7-8)	(9-10)	(11-12)	(13-14-15)
A	18.6	46.7	64.2	36.4	54.0	72.6
B	46.7	10.0	19.6	40.9	32.2	65.3
C	64.2	19.6	[illegible]	51.7	33.7	66.1
D	36.4	40.9	51.7	14.1	25.8	38.2
E	54.3	32.2	33.7	25.8	7.1	33.9
F	72.0	65.3	66.1	38.2	33.9	[illegible]

تتم عملية التجميع من الجدول 2 بالطريقة نفسها تمامًا مثل التجميع الأولي من الجدول 1. أولاً نجد أصغر قيمة صف في كل عمود ، على النحو الاتي :

Column	A	B	C	D	E	F
Smallest row value in row	D	C	B	E	D	D

From this we derive our grouping:



Group H

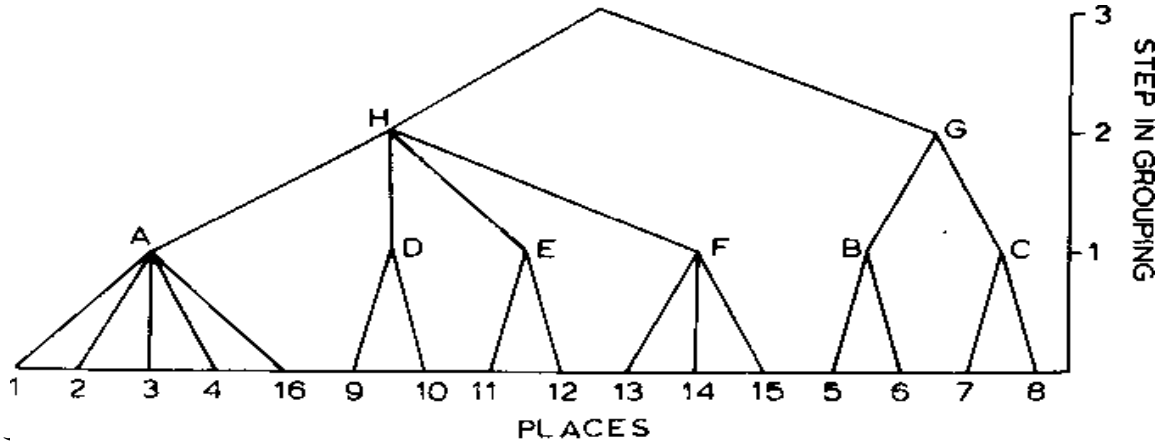
Group G

Table 3 Inter-group distances from Table 2 a

	Group	
	G (B-C)	H (A-D-E-F)
G (B-C)	17.1	52.7
H (A-D-E-F)	52.7	37.2

انتقلنا بعد ذلك من ست عشرة قرية إلى ست مجموعات ، ثم إلى مجموعتين ، حتى أصبحت كل قرية في المجموعة نفسها . يمكن تمثيل هذه العملية بشجرة ارتباط (الشكل 4) ، حيث تُشير إلى الارتباطات في كل خطوة . (لاحظ أن القرى مرتبة بترتيب محدد على طول المحور الأفقي لتجنب تقاطع الخطوط) . تنشأ مشكلتان من هذه المجموعة من الإجراءات . الأولى هي : **متى نتوقف** ؟ من الواضح أن الاستمرار حتى تصبح جميع القرى في مجموعة واحدة لا جدوى منه ، فمتى يجب إيقاف الإجراء ؟ ربما قررنا مسبقًا أننا نريد ست مجموعات ، لذا سنتوقف بعد الخطوة الأولى .

Fig. 4 Linkage tree for Elementary Linkage Analysis of Villages in Figure 2



متوسط المسافة داخل المجموعة ، الموضح في الأقطار الرئيسية لمصفوفات الجدولين 2 و 3. هذه هي متوسطات المسافات بين أعضاء المجموعة المحددة ، لذا تُحسب قيمة 17.1 للمجموعة Q في الجدول 3 على النحو الاتي :

Pairs of Places	5-6	5-7	5-8	6-7	6-8	7-8
Distance	10.0	25.5	21.2	15.8	15.8	14.1

average distance = 102.4/6 = 17.1

قد نقرر أن المجموعات المستخدمة في الجدول 3 كبيرة جدًا ، بمعنى أن متوسط المسافة داخلها كبير جدًا ، وبالتالي فإن الأعضاء ليسوا متشابهين جدًا . بشكل عام ، المسافات داخل المجموعة في الجدول 2 أصغر، لذا قد نقرر الالتزام بحل المجموعات الست . **المشكلة الثانية** هي أن الإجراء لا يتيح لنا الكثير من الخيارات . يمكن أن يكون لدينا ست عشرة قرية فريدة ، أو ست مجموعات من القرى ، أو مجموعتان من القرى ، أو جميع القرى الست عشرة في مجموعة واحدة . ماذا لو أردنا أربع مجموعات ؟ علاوة على ذلك ، فإن بعض المجموعات في الجدول 2 لها متوسط مسافات داخل المجموعة منخفض ، مما يشير إلى مجموعات متماسكة جدًا من القرى المتشابهة ، بينما لدى مجموعات أخرى مسافات أكبر بكثير . سيكون من الأفضل أن نتمكن من اختيار النوع الأول ، والاستغناء عن الأنواع الأخرى .

الاجراءات الهرمية

نظرًا للمشكلات التي اثيرت آنفًا ، عادةً ما تُعتمد الطريقة الموضحة أعلاه بصيغة مُعدّلة ، حيث تسير عملية التجميع ببطء أكبر . تعمل هذه الطريقة وفقًا للقواعد نفسها ، باستثناء أنه في كل خطوة ، يتم تجميع الزوج ذي المسافة الأقصر فقط ، بدلاً من تجميع جميع الأزواج دفعةً واحدة . يُتيح هذا مرونة أكبر في تحديد نقطة توقف العملية ، أو تحديد التجميع الذي سيتم اختياره بعد اكتمال العملية بأكملها ورسم شجرة الارتباط . يبدأ هذا الإجراء الهرمي بالجدول 1. بالنسبة لكل عمود من المصفوفة ، يتم ملاحظة الصف الذي يحتوي على أصغر إدخال (باستثناء القطر الرئيسي ، كما هو معتاد) ، على النحو الاتي :

Column	1	2	3	4	5	6	7	8
Row with smallest entry	2	3	2	3	6	5	8	7
value	7.1	3.0	3.0	5.4	10.0	10.0	14.1	14.1

Column	9	10	11	12	13	14	15	16
row with smallest entry	10	9	12	11	14	13/15	14	1
value	14.1	14.1	7.1	7.1	7.1	7.1	7.1	30.0

ثم تُعزل أصغر هذه القيم وتُستخدم لتحديد الزوج الذي سيتم تجميعه . المسافة بين القريتين ٢ و٣ هي ٣,٠ وحدات ، وتُشكل هذه القيم المجموعة (أ) في الخطوة الأولى من الإجراء الهرمي . يجب الآن إنشاء مصفوفة مسافات جديدة ، تُستبدل فيها القريتان ٢ و٣ المنفصلتان بالمجموعة (أ) (الجدول ٤) . هذه مصفوفة أبعادها ١٥ × ١٥، حيث تُمثل المسافات في العمود (أ) والصف أ متوسطات جميع أعضاء المجموعة أ إلى القرية المعنية (أي المسافة من الأخيرة إلى مركز ثقل المجموعة أ) ؛ وتُحسب هذه القيم بالطريقة السابقة نفسها تمامًا ، بحيث تكون المسافة من القرية ١ إلى المجموعة أ هي متوسط المسافات بين ١ و٢ و١ و٣ (٧,١ و٩,٤، مما يجعل الناتج ٨,٣) . لاحظ أن لدينا الآن متوسط مسافة داخل المجموعة في القطر الرئيسي عند تقاطع الصف (أ) والعمود (أ) في الجدول 4 .

بعد حساب مصفوفة المسافة ، نبحث فيها عن أقصر مسافة ، والتي ستشير إلى زوج القرى المراد تجميعها في الخطوة التالية . المسافة هي 6.2 ، وتُجمع القرية 4 مع المجموعة (أ) . مرة أخرى ، تُحسب مصفوفة مسافة جديدة ، هذه المرة 14 × 14 (لم تُذكر هنا لاعتبارات المساحة) ، وننتقل إلى الخطوة التالية . تستمر العملية ، تجميعًا تلو الآخر، حتى تصبح جميع القرى في مجموعة واحدة . تُلخص معاملاتها كما يلي :

Step	Places Grouped	Distance	Forming Group	Average Intra-Group Distance
1	2-3	3.0	A	3.0
2	A-4	6.2	B	5.2
3	11-12	7.1	C	7.1
	13-14-15	7.1	D	8.1
4	B-1	8.8	E	7.0
5	5-6	10.0	F	10.0
6	7-8	14.1	G	14.1
	9-10	14.1	H	14.1
7	F-G	19.6	J	17.1
8	C-H	25.8	K	20.7
9	D-K	36.1	L	27.7
10	E-16	37.9	M	18.6
11	L-J	50.8	N	38.3
12	N-M	68.7	O	30.1

شجرة الارتباط لهذا التجميع هي الشكل 5 ، حيث تُشكل المسافة المتضمنة في التجميع المحور الرأسي . ومن هنا ، نحصل على انطباع واضح عن المسافات القصيرة المتضمنة في المجموعات القليلة الأولى (حتى J، على سبيل المثال) ، والمجموعات الأكبر بعد ذلك . إذا لم يكن لدينا سبب مُسبق لقبول عدد مُعين من المجموعات ، فسيساعدنا هذا الرسم البياني في تحديد مكان توقف التجميع . ومن الفروقات المهمة التي يجب ملاحظتها بين هذا التجميع :

Table 4 Distances between places after first step in hierarchical grouping

	Place														
	1	4	5	6	7	8	9	10	11	12	13	14	15	16	A
1	0	10.0	43.0	49.5	64.0	64.0	30.0	41.2	51.0	57.0	70.7	76.5	72.8	30.0	8.3
4	10.0	0	35.4	43.0	58.3	56.6	31.6	44.7	50.0	55.2	72.8	79.1	76.2	40.0	6.2
5	43.0	35.4	0	10.0	25.5	21.2	35.4	47.4	35.4	36.1	63.6	70.0	71.1	69.6	34.8
6	49.5	43.0	10.0	0	15.8	15.8	35.4	45.3	29.2	28.3	57.0	64.0	65.2	73.8	41.1
7	64.0	58.3	25.5	15.8	0	14.1	44.7	51.0	30.0	25.5	53.9	62.1	63.3	86.0	56.1
8	64.0	56.6	21.2	15.8	14.1	0	51.0	60.0	41.2	38.1	67.1	73.8	76.2	89.4	55.6
9	30.0	31.6	35.4	35.4	44.7	51.0	0	14.1	22.4	29.2	41.2	46.4	44.7	42.4	25.9
10	41.2	44.7	47.4	45.3	51.0	60.0	14.1	0	22.4	29.2	30.0	35.4	31.6	44.7	38.8
11	51.0	50.0	35.4	29.2	30.0	41.2	22.4	22.4	0	7.1	28.3	35.4	36.1	64.0	45.2
12	57.0	55.2	36.1	28.3	25.5	38.1	29.2	29.2	7.1	0	29.2	36.1	38.1	71.1	50.8
13	70.7	72.8	63.6	57.0	53.9	67.1	41.2	30.0	28.3	29.2	0	7.1	10.0	72.8	67.1
14	76.5	79.1	70.7	64.0	62.1	73.8	46.4	35.4	35.4	36.1	7.1	0	7.1	76.5	73.3
15	72.8	76.2	71.1	65.2	63.3	76.2	44.7	31.6	36.1	38.1	10.0	7.1	0	70.7	70.2
16	30.0	40.0	69.6	73.8	86.0	89.4	42.4	44.7	64.0	71.1	72.8	76.5	70.7	0	32.7
A	8.3	6.2	34.8	41.4	56.1	55.8	25.9	38.8	45.2	50.8	67.1	73.3	70.2	32.7	3.0

وما ينتج عن تحليل الارتباط الأولي لمجموعة القرى نفسها هو أن القرية ١٦ لم تُجمع مع القرى ١ و٢ و٣ و٤ حتى الخطوة ١٠؛ ففي التصنيف السابق ، جُمعت القرى الخمس في الخطوة الأولى ، على الرغم من المسافة الكبيرة التي تفصل القرية ١٦ عن بقية المجموعة . أما طريقتنا الحالية (التجميع الهرمي مع استبدال مركز الثقل) فتترك القيم المتطرفة ، مثل القرية ١٦، كمجموعات منفصلة ؛ حيث يُخصص تحليل الارتباط الأولي فردًا على الفور للمجموعة التي تضم أقرب جار له ، بغض النظر عن المسافة .

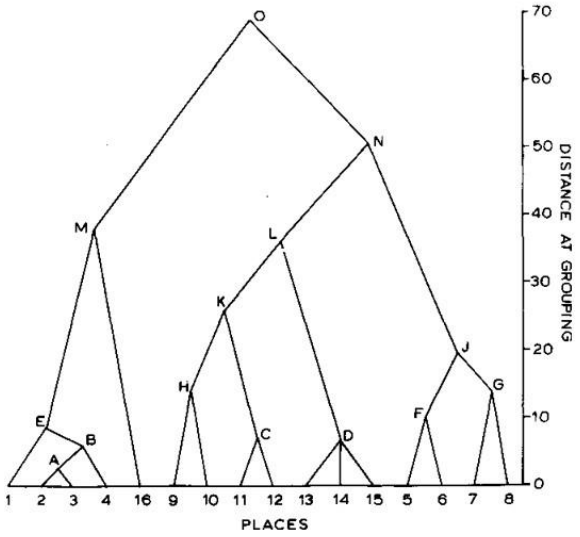


Fig. 5 Linkage tree for Hierarchical Clustering with Centroid Replacement of Villages in Figure 2.

أين يجب أن نوقف عملية التجميع ؟ أي خطوة من الإجراء توفر "أفضل" مجموعة من المجموعات ؟ للأسف ، لا يوجد تعريف بسيط لـ "الأفضل" ، لذلك قد نضطر إلى اتخاذ قرار عشوائي بناءً على فحص شجرة الارتباط . بدلاً من ذلك ، يمكننا القول إنه لن يتم قبول أي مجموعة إذا كانت المسافة المتضمنة في إنتاجها أكبر من المسافة المحددة مسبقًا . في حالتنا ، قد تكون هذه 20.0 وحدة مسافة ، وفي هذه الحالة سنتوقف عند الخطوة 7 ، مع ست مجموعات ناتجة .(E-16-H-C-D-J) أو يمكننا القول إنه لن يتم قبول أي مجموعة إذا تجاوز متوسط المسافة داخل المجموعة رقمًا معينًا . (لاحظ أن المسافات داخل المجموعة لا تكبر، كما هو الحال مع مسافات التجميع ، في كل خطوة لاحقة من عمليتنا ؛ والسبب في ذلك هو أن إضافة فرد واحد إلى مجموعة سيكون لها على الأرجح تأثير أقل على حجم متوسط المسافة داخل المجموعة من إضافة مجموعة إلى أخرى) .

لذا، إذا اعتمدنا مسافة قصوى قدرها 20.0 في مثالنا ، فسنقبل جميع المجموعات الناتجة حتى الخطوة 7، بالإضافة إلى تلك من الخطوة 10 . ولكن كيف يمكن تحديد هذه المسافات العشوائية ، ما لم تكن لدينا أسباب مسبقة وجيهة للاختيار؟ إذا رسمنا جميع المسافات في الجدول 1 (الشكل 6) ، فهناك بعض الإيحاءات بوجود انقطاع حول المسافة 20.0 ، لكن توزيع التكرار التراكمي لا يشير إلى أي انقطاع من هذا القبيل . في هذا ، كما هو الحال في العديد من الجوانب الأخرى لإجراء التصنيف ، يكون الاختيار العشوائي ضروريًا ، و"الموضوعية" الكاملة مستحيلة (جونستون، 1968) . تم اقتراح مجموعة متنوعة من الطرق الأخرى . على سبيل المثال ، يُمكن رسم متوسط المسافة داخل المجموعة - لجميع المجموعات - بيانيًا مقابل متوسط المسافة بين المجموعات في جميع خطوات الإجراء . ولكن في النهاية ، لا بد من اتخاذ قرار عشوائي .

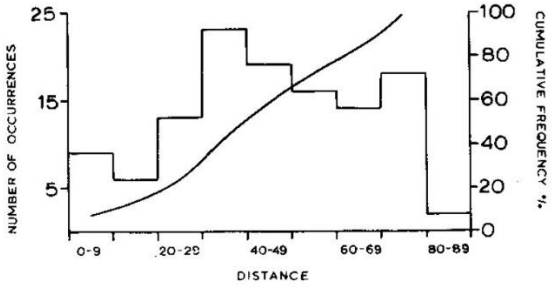


Fig. 6 Frequency Histogram and Cumulative Frequency Distribution of Distances in Table 1.

تعريف المجموعات البديلة

كانت القاعدة المعتمدة في إجرائنا الهرمي هي أنه في كل خطوة ، تُستبدل الأماكن المُجمّعة معًا بمركز ثقلها ، وتُظهر مصفوفة المسافة الجديدة المسافات بين مراكز ثقل المجموعة . (المجموعات ذات العضو الواحد ، بالطبع ، يكون مركز ثقلها هو الفرد) . ولكن ، كما أوضح العديد من الكُتّاب (مثل لانس وويليامز، 1967) ، يُمكن اقتراح قواعد أخرى ، والدفاع عنها منطقيًا بناءً على تعريفات مُحددة للمجموعات . وتشمل هذه :

(1) طريقة أبعد مسافة ، حيث تُعبّر المسافة بين مجموعتين عن أقصى مسافة بين عضو في إحداهما وعضو في الأخرى . وهذا من شأنه أن يُبرز اختلافات المجموعات ، وفي ظل قاعدة إيقاف عشوائية لمسافة مُعينة ، يعني بلا شك مجموعات أكثر مما يعنيه تطبيق استبدال مركز ثقل المجموعة على البيانات نفسها .

(2) طريقة أقرب مسافة ، والتي تعتمد تعريفًا "متحررًا" للمسافة بين المجموعات على أنها أقصر مسافة بين أي زوج من الأعضاء .

(3) طريقة متوسط المجموعة ، حيث عند وجود مجموعات متعددة الأعضاء ، يؤدي دمج مجموعتين إلى استبدالهما بمركز ثقل مركزي لمركزيهما ، بدلاً من مركز ثقل جميع أعضاء المجموعة . وبالتالي ، إذا كانت المسافة من المجموعة X التي تضم المكانين i و j إلى المجموعة Y هي 10.0 ، ومن المجموعة Z التي تضم الأماكن k ، 1، m، n إلى المجموعة Y هي 15.0 ، فإن المسافة من المجموعة الجديدة X-Z إلى Y ستكون 12.5 (متوسط 10.0 و15.0) ؛ ولا يُعتد هنا بكبر عدد أعضاء Z ، حيث تُعد المجموعة فردًا واحدًا .

(4) طريقة المسافة الكلية (المعروفة أحيانًا باسم تحليل الترتيب النموذجي ، ينظر جونستون، 1968)، حيث يُخصص فرد لمجموعة فقط إذا كان أقرب إلى جميع أعضاء المجموعة منه إلى أي فرد آخر . من الواضح أن هذا معيار صارم للغاية ، وهو تعريف "محافظ" .

طريقة وارد

يُعد الاختيار بين هذه المتغيرات من الإجراء الهرمي - ولم تُدرج هنا سوى عينة صغيرة من الاحتمالات - أمرًا صعبًا للغاية ما لم يكن لديك تعريف واضح للمجموعة . وحتى الآن ، لم نتناول إلا الطرق التي تعتمد على المسافات بين المجموعات وداخلها . هناك طريقة أخرى ، تحظى بشعبية متزايدة بين الجغرافيين ، طورها وارد (1963) . تستند طريقة وارد إلى حجة مفادها **أن هدف التصنيف ينبغي أن يكون إنتاج مجموعات تكون فيها المسافات بين أعضائها الفردية ونقاط مركزها المركزية عند الحد الأدنى** . بمعنى آخر، يجب تقليل تباين المسافة إلى الحد الأدنى . هذا التباين (يسميه وارد مجموع مربعات الخطأ المعياري- ESS) هو:

$$\left(\sum_{i=1}^{n} (D_{ix})^2\right)/n = ESS \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

حيث ان D تمثل المسافة بين المكان i و مركز المجموعة x ، وبافتراض ان i عضو المجموعة (X) ، و ان n يمثل عدد اعضاء المجموعة X وتكتب هذه المعادلة بالصيغة الاتية :

$$\sum_{i=1}^{n} i_x^2 - ((\Sigma i_x)^2/n) = ESS$$

where i_x is the value of place i on variable x
Summation is over all variables.

يوضح الشكل 7 مثالاً على هذه الطريقة . لدينا عشر دوائر انتخابية مُرتبة حسب نسبة ناخبيها الذين صوتوا لحزب العمال ، ونريد تجميع الدوائر . أي زوج سيُعطي أقل قيمة ؟ من الواضح أننا نحتاج إلى مراعاة الأزواج المتجاورة فقط ، فعلى سبيل المثال ، لن يتم تجميع الدائرة 1 مع الدوائر 3 قبل تجميعها مع الدوائر 2 . بالنسبة للدائرتين 1 و 2 ، يُحسب ESS كما يلي :

$$ESS_{12} = (1_x^2 + 2_x^2) - ((1_x + 2_x)^2/n)$$

where 1_x = the value for constituency 1 on variable x, etc.

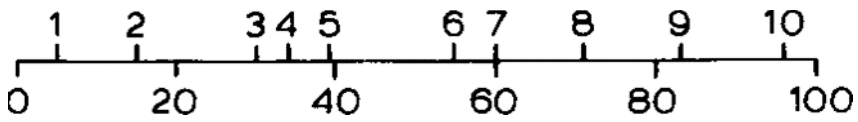
$$= (5^2 + 15^2) - (5 + 15)^2/n$$

$$= 250 - 400/2 = 50$$

(لاحظ أن هذه القيمة - 50 - هي مجموع المسافات التربيعية من المكانين 1 و2 إلى مركز ثقل المجموعة ، والتي سيكون لها قيمة 10 على متغير نسبة التصويت لحزب العمال) وبتتبع الازواج الممكنة ، قد نحصل على :

Pair	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
ESS	50	112.5	8	12.5	128	12.5	60.5	288	388

الحد الأدنى لـ ESS هو 8 ، ونقوم بتجميع الدوائر الانتخابية 3 و4 . في الخطوة التالية ، نستبدل 3 و4 بالمجموعة A ويجب علينا حساب ESS للمجموعات المحتملة الجديدة :



النسبة المئوية للمصوتين لحزب العمال في الدوائر الانتخابية

$$ESS \quad A-2 \text{ (i.e. 2-3-4)} = (15^2 + 30^2 + 34^2) - (79^2)/3 = 200.7$$
$$ESS \quad A-5 \text{ (i.e. 3-4-5)} = (30^2 + 34^2 + 39^2) - (103^2)/3 = 40.7$$

الحد الأدنى لقيمة ESS هو ١٢,٥، مما يؤدي إلى تجميع الدوائر الانتخابية ٦ و٧ (قيمة ١٢,٥ للدوائر ٤ و٥ أصبحت الآن زائدة عن الحاجة ، لأن الدائرة ٤ جُمعت مع الدائرة ٣) . الإجراء الكامل لتجميع دوائرنا الانتخابية العشر هو الاتي ، الشكل 8 هو شجرة الارتباط لهذه العملية . بما أن التجميع يعتمد على متوسط مربعات المسافات ، فإن التباين الكبير داخل المجموعة في الخطوات اللاحقة يتضخم (وهذه ميزة مقارنة بالطرق القائمة على المسافات نفسها) ، وفي هذه الحالة على الأقل ، يكون قرار إيقاف التجميع واضحًا تمامًا .

Step	Group	ESS	Forming Group	New ESS for possible groups			
1	3-4	8	A	A-2	200.7	A-5	40.7
2	6-7	12.5	B	B-5	240.7	8-8	134.0
3	A-5	40.7	C	C-2	321.0	B-C	696.2
4	1-2	50.0	D	C-D	801.2		
5	B-8	134.0	E	E-C	1322.8	E-9	464.8
6	9-10	338.0	F	E-F	1126.0		
7	C-D	801.2	G	G-E	3457.9		
8	E-F	1126.0	H	G-H	7783.6		
9	G-H	7783.6	J				

<u>بمجرد أن تتشكل المجموعة</u>... أيًا كانت الإجراءات الهرمية التي نتبعها ، تظهر دائمًا مشكلة واحدة . تُجمّع الإجراءات المجموعات دائمًا ، دون الاستفسار عما إذا كان من الأفضل ، في خطوات معينة ، استيفاء معايير التصنيف إذا تم تفكيك إحدى المجموعات الموجودة . على سبيل المثال ، ربما قُسّمنا أوروبا ، لتنظيم دوري كرة القدم الكبرى ، إلى ثلاث مجموعات - الغربية والوسطى والشرقية . في وقت لاحق ، قد نقرر أننا نحتاج إلى دوريين اثنين فقط . ربما يكون أفضل إجراء للتخصيص هو تخصيص الجزء الغربي من الدوري المركزي للدوري الغربي ، والباقي للدوري الشرقي . ولكن إذا اقتصرنا على التعامل مع الرابطة المركزية ككل ، فسيكون توازن الرابطتين الجديدتين غير متكافئ . يوضح الشكل 9 مثالاً على هذه المشكلة في إجراءات التجميع .

لدينا ست محطات تقع على طول خط سكة حديد ، ونريد تجميعها في "مناطق صيانة" لتقليل حركة المعدات بين المحطات . باستخدام التجميع الهرمي مع استبدال مركز الثقل ، سيتم التجميع على النحو المبين في ادناه . وإذا قررنا تقسيم المجموعات إلى ثلاث مجموعات (أ، ب، ج) ، فستحتوي كل مجموعة على محطتين . ولكن ماذا لو قررنا تقسيمها إلى مجموعتين فقط - ربما في برنامج ترشيد لاحق ؟ ستكون المجموعة الأولى المجموعة ب ، بمحطتين ، والأخرى المجموعة د ، بأربع محطات ، أبعدها مسافة 81 كيلومترًا . بديهيًا ، نعتقد أن الحل الأمثل للمجموعتين هو تخصيص المحطة 5 للمجموعة ب والمحطة 6 للمجموعة أ ، مما سيعطي متوسط مسافات داخل المحطة 28.3 و21.7 كيلومترًا ، بدلًا من 20 و33 كيلومترًا الناتجة عن عملية التجميع .

Step	Places Grouped	Distance	Forming Group	Average Intra-Group Distance
1	1-2	10.0	A	10.0
2	3-4	20.0	B	20.0
3	5-6	26.0	C	26.0
4	A-C	40.5	D	33.0

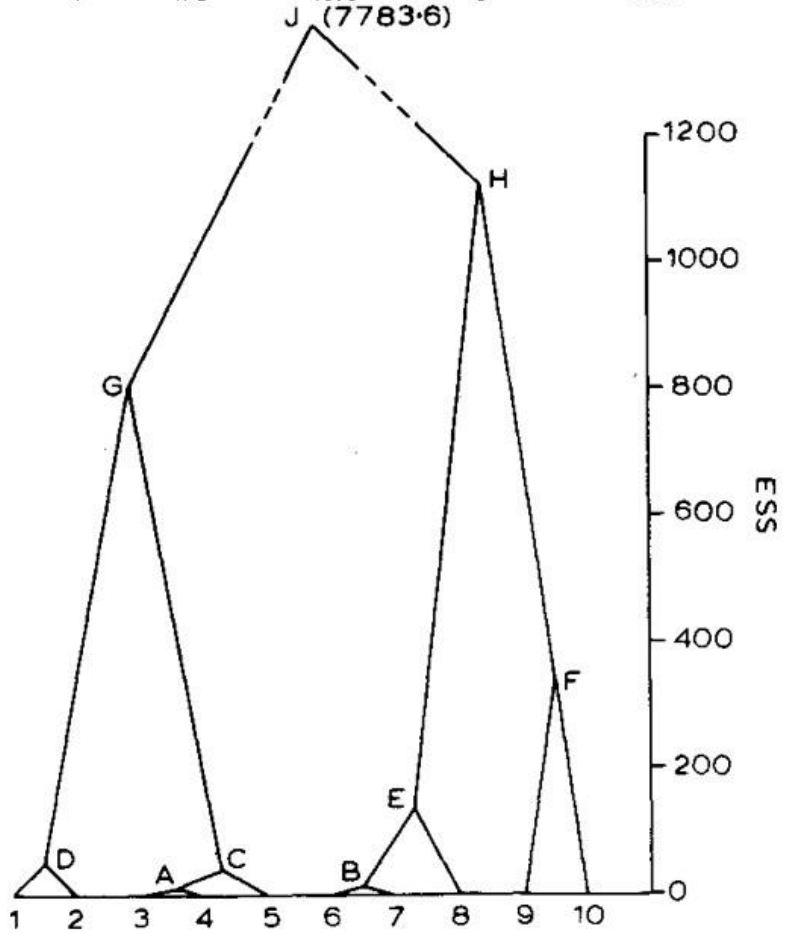


Fig. 8 Linkage tree for application of Ward's Method to Constituencies in Figure 7

تنشأ هذه المشكلة لأننا بعد الخطوة الثالثة نُجمّع مجموعات ، وليس محطات فردية . يمكننا تجاوز هذه المشكلة بالتحقق من كون التجميع مثاليًا ، سواءً في كل خطوة من خطوات التجميع ، أو - بشكل أكثر واقعية - عندما نصل إلى العدد المطلوب من المجموعات . يتضمن ذلك تحديد مراكز المجموعات ، ومعرفة ما إذا كانت كل مجموعة أقرب إلى مركز مجموعتها . بالنسبة لمحطاتنا الست ، ينتج عن ذلك المسافات المبينة لاحقا . في حل المجموعات الثلاث ، تكون كل محطة أقرب إلى مركز ثقل المجموعة التي خُصصت لها من مركز ثقل أي مجموعة أخرى . (نكتشف ذلك بملاحظة ما إذا كانت المسافة المسطرة هي الأقصر في الصف ذي الصلة) . أما في حل المجموعتين ، تكون المحطة 5 أقرب إلى مركز ثقل المجموعة ب منها إلى مركز ثقل مجموعتها د . يبدو أن تصنيفها خاطئ ، لذا نُعيد تخصيصها للمجموعة ب (التي تُسمى الآن المجموعة BI) ، ونُعيد حساب المسافات . لا توجد الآن أية تصنيفات خاطئة ، لذا يُعد الحل الأمثل ، وننتهي بمجموعتين ، إحداهما تحتوي على المحطات ١ و٢ و ٦، والأخرى المحطات ٣ و٤ و٥ . (تجدر الإشارة إلى أنه لو استخدمنا أقرب مسافة ، بدلًا من طريقة استبدال مركز الكتلة ، لحصلنا على حل أمثل لمجموعتين ، ولكن ليس لثلاث مجموعات : أترك لكم شرح ذلك).

Distance to Centroid of Group	Three-Group Solution			Two-Group Solution	
	A	B	C	B	D
Station					
1	5	91	45.5	91	25.25
2	5	81	35.5	81	15.25
3	75	10	35.5	10	55.75
4	95	10	55.5	10	75.75
5	52.5	32.5	13	32.5	33.25
6	27.5	57.5	13	58.5	7.25

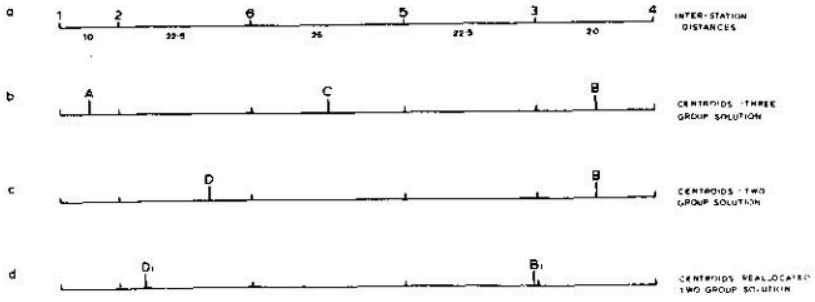


Fig. 9 The Optimal Grouping Problem: A Hypothetical Railway System

21

Distance to Centroid of Group	1 B	D
Station 1	80.2	14.2
2	70.2	4.2
3	0.8	66.8
4	20.8	86.8
5	21.7	44.3
6	47.7	18.3

إلى n من الابعاد

جميع الأمثلة المستخدمة حتى الآن تناولت تصنيف "مساحات" أحادية البعد (محطات على طول خط سكة حديد ؛ دوائر انتخابية وفقًا لتصويت حزب العمال) أو ثنائية الأبعاد (قرى تتميز بمتغيرين) . ويرجع ذلك إلى إمكانية إعادة إنتاج "خرائط" مواقع الأماكن ومراكز كتلة المجموعات بدقة في مخططاتنا ، مما يُتيح فهم آلية عمل إجراءات التجميع . لكن جميع الطرق التي ناقشناها تنطبق بالتساوي على التصنيفات ذات الأبعاد n ، حيث يُصنَّف 3 n وأفرادنا بناءً على عدد كبير من المتغيرات المختلفة . في حالة أحادية

البعد، تكون المسافات جمعًا ، بحيث عندما تكون المسافة AB > المسافة AC، فإن AC AB + BC . في الحالة ثنائية الأبعاد ، نطبق صيغة فيثاغورس للمعادلة (1) ، ويمكن توسيعها لتشمل n بُعدًا ، كما يلي:

$$D_{ij} = \sqrt{(i_1 - j_1)^2 + (i_2 + j_2)^2 .... + (i_n - j_n)^2} ..... (4)$$

where i_n, j_n = the values of i and j on the nth variable.

يمثل كل حد بين قوسين المسافة بين المكانين (i وj) على المتغير المحدد ؛ في المعادلة (4) نحصل على مجموع مربع المسافة لجميع المتغيرات ، ثم نأخذ الجذر التربيعي لهذه القيمة لنحصل على طول الوتر في الفضاء ذي الأبعادn .

قد تنشأ مشاكل في تطبيق المعادلة (4) بسبب اختلافات المقياس بين المتغيرات . في مثال قريتنا ، لوحظ تقريبًا النطاق الكامل للقيم من 0 إلى 100% لكل متغير . مع ذلك ، لو كان نطاق أحد المتغيرات من 0 إلى 100% ، ونطاق الآخر من 40 إلى 60% فقط ، فمن الواضح أن المتغير الأول سيساهم بشكل أكبر في المسافة بين أزواج الأماكن مقارنةً بالمتغير الثاني ، لأن نطاق قيمه كان أكبر بخمس مرات . إذا تم اختيار عدة متغيرات لتصنيف ما ، فمن المفترض أن يكون لكل منها وزن متساوٍ في تحديد المجموعات (مع ذلك ، راجع جونستون، 1965، 1970 لمناقشة هذا الأمر) . لضمان ذلك ، عادةً ما تُعبَّر قيم المتغيرات الأصلية بصيغة الانحراف Z المعياري ، حيث :

$$Z_i = \frac{(x_i - \bar{x})}{S_x} ..... (5)$$

where x_i = value of variable x for ith observation

$\bar{x}$ = the mean of all values of x

S_x = the standard deviation for x

Z_i = the Z score for observation i

بعد هذا التحويل ، يكون لكل متغير متوسط يساوي 0 وانحراف معياري يساوي 1.0، ويكون له وزن جميع المتغيرات الأخرى في التصنيف نفسه .

تتعلق إحدى المشكلات المحتملة الأخرى المتعلقة بالتصنيف متعدد الأبعاد بالعلاقات المتبادلة بين المتغيرات . يفترض حساب المسافات باستخدام صيغة فيثاغورس أن جميع المتغيرات متعامدة ، أي غير مترابطة : إذا لم تكن كذلك ، تكون المسافات متحيزة . للتغلب على هذه المشكلة ، من الممكن توحيد المسافات ، ولكن الإجراء الأكثر شيوعًا هو استبدال مجموعة المتغيرات الأصلية بمجموعة جديدة ، من خلال استخدام تحليل المكونات الرئيسية . (تم توضيح تحليل المكونات الرئيسية في دراسة أخرى من هذه السلسلة - بقلم س. دولتري) . ما يفعله بشكل أساسي هو أخذ مصفوفة من الملاحظات على مجموعة من المتغيرات واستبدال الأخيرة بمجموعة جديدة من المتغيرات – عادةً ، وإن لم يكن بالضرورة ، عددًا أصغر - والتي تكون متعامدة .

بالإضافة إلى ذلك ، فإن درجات الملاحظات على هذه المتغيرات الجديدة - المكونات - لها متوسط 0 وانحراف معياري 1.0 ، بحيث يمكن حساب مسافات فيثاغورس على الفور . ومن المزايا الأخرى لتحليل المكونات الرئيسية أنه يزيل التكرار في مصفوفة البيانات الأصلية . قد يكون متغيران تكرارًا لبعضهما البعض - متوسط الدخل ومتوسط سنوات الدراسة كمؤشرات للحالة الاجتماعية والاقتصادية ،

على سبيل المثال - وسيعطي مفهومهما المشترك وزنًا غير مستحق إذا استخدم التجميع كليهما . يحل تحليل المكونات الرئيسية محلهما بمتغير واحد . بعد حساب مصفوفة المسافة ، تُجرى عملية التجميع كما في السابق ، باستخدام أيٍّ من الإجراءات يُعد مناسبًا للغرض المُحدد . يمكن إنشاء شجرة ارتباط لعرض عملية التجميع ، ولكن للأسف ، لا يُمكن تمثيل مواقع الملاحظات في فضاء يتجاوز الأبعاد الثلاثة بيانيًا . يُمكن إجراء اختبار أمثلية التجميع ؛ عند التعامل مع أكثر من بُعد واحد ، يُستخدم أسلوب التعدد .

تحليل تمييزي خطي متعدد الأبعاد؛ لا يمكن توضيح تفاصيل هذه الطريقة هنا (ينظر كينج ، ١٩٦٩) ، ولكن أهدافها ونتائجها هي نفسها التي نوقشت أعلاه في مسألة محطة السكة الحديد . لا تعتمد جميع التصنيفات على مصفوفات المسافة ؛ تتوفر مجموعة واسعة من مقاييس التشابه والاختلاف بين الملاحظات (إيفريت، ١٩٧٤؛ سبنس وتايلور، ١٩٧٠) . أحد هذه المقاييس شائع الاستخدام هو معامل الارتباط (بيرسون، سبيرمان، أو كيندال) ، والذي يُظهر درجة التوافق بين متغيرين على مجموعة من الملاحظات ، أو بين ملاحظتين على مجموعة من المتغيرات . يمكن تطبيق إجراءات التجميع الهرمي على مصفوفات الارتباط ، باستثناء أن التجميع يكون في المكانين اللذين لهما أكبر قيمة في المصفوفة (أي الأكثر تشابهًا) بدلاً من الأصغر . ومع ذلك ، هناك بعض المشاكل في حساب متوسط الارتباطات بين المجموعات .

مثال: المناطق الاجتماعية في ملبورن

أنتج الجغرافيون الحضريون وعلماء الاجتماع أدبياتٍ غزيرة حول المناطق الاجتماعية داخل المدن (جونستون، ١٩٧١) . وكان من أهداف هذه الأعمال توفير خريطة للمناطق الاجتماعية (أنواع الأحياء السكنية) ، مُعرّفة وفقًا لخصائص السكان والأسر والمساكن . قدّم البيان النظري الأولي نموذجًا استنتاجيًا يتضمن مخططًا للتصنيف ، إلا أن أحدث الأعمال اعتمدت على الإجراءات الاستقرائية ، باستخدام أساليب مثل تلك الموضحة هنا لاقتراح أنواع المناطق الاجتماعية (بيري وريس، ١٩٦٩).

ومن الأمثلة على هذه الأعمال دراسةٌ لمدينة ملبورن ، أستراليا أجراها جونز (١٩٦٩) . حيث اختصر ٢٤ متغيرًا أصليًا إلى ثلاثة مكونات متعامدة على بعضها البعض ، ثم شرع في تصنيف ٦١١ منطقة إحصاء في المنطقة الحضرية ، باستخدام إجراء التجميع الهرمي - القائم على مصفوفة فيثاغورس للمسافات - مع استبدال مركز الثقل . تم عزل عشرين مجموعة ، تراوحت أحجامها بين ثلاث مجموعات تضم كل منها منطقة واحدة فقط ، وأخرى تضم 183 منطقة . وفي عملية التجميع ، استُبدلت تسع عشرة مجموعة بثلاث مجموعات كبيرة - كما هو موضح في الشكل 10. (لاحظ استخدام المسافات المربعة في التجميع). ثم أُنشئ جدول يوضح متوسط ترتيب كل مجموعة من المجموعات العشرين (الرتب من 1- منخفضة إلى 611-مرتفعة) . ومن (الجدول 5)، يمكن تحديد المجموعات الرئيسية الثلاث بشكل منفصل على النحو الاتي :

1- ضواحي الطبقة المتوسطة : عالية من حيث الوضع الاجتماعي والاقتصادي ، ومنخفضة من حيث الوضع العرقي .

2- ضواحي الطبقة المتوسطة : عالية من حيث الوضع العائلي ، ومنخفضة من حيث الوضع الاجتماعي والاقتصادي ، ومنخفضة من حيث الوضع العرقي .

3- ضواحي الدخل المنخفض ذات النسب الكبيرة من المهاجرين غير البريطانيين : منخفضة من حيث الوضع العائلي ، ومنخفضة من حيث الوضع العرقي ، ومنخفضة من حيث الوضع الاجتماعي والاقتصادي.

4- أحياء المهاجرين في وسط المدينة : منخفضة من حيث الوضع العائلي ، ومنخفضة من حيث الوضع العرقي .

5- أحياء المهاجرين في وسط المدينة : منخفضة من حيث الوضع العائلي ، ومنخفضة من حيث الوضع العرقي .

6- أحياء المهاجرين في وسط المدينة : منخفضة من حيث الوضع العرقي .

لا تمثل كل مجموعة فرعية المجموعة الأكبر التي تم دمجها فيها بشكل حقيقي ؛ ففي العديد من الحالات ، تنحرف المجموعة الفرعية عن متوسط المجموعة في أحد المتغيرات الثلاثة ، ويمكن عادةً تحديد ذلك بوضوح في شجرة الارتباط .

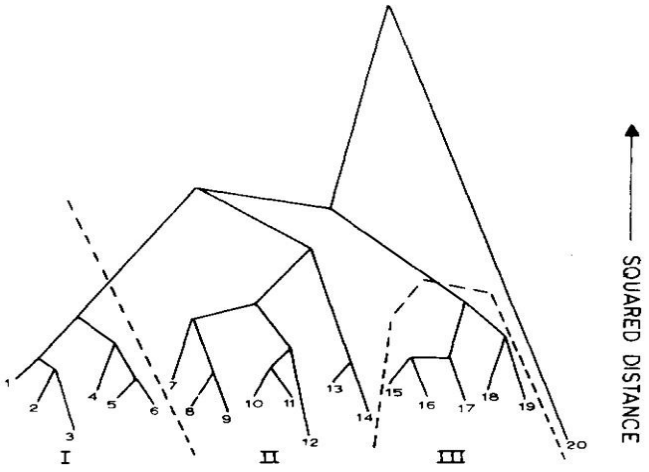


Fig. 10 Linkage tree for the Final Stages of a Classification of Melbourne's Social Areas (after Jones, 1969, p. 98).

Table 5 Mean ranks for Melbourne social areas

(N = 611)

Group	No. of Members	Mean Rank for Socio-Economic Status	Family Status	Ethnic Status
I	287	421	383	155
1	183	375	450	158
2	14	567	534	61
3	1	598	435	1
4	81	519	225	176
5	4	244	209	119
6	3	301	330	15
II	83	104	502	390
7	8	37	602	320
8	41	152	494	407
9	3	226	601	342
10	14	30	463	545
11	5	8	483	607
12	1	36	606	563
13	6	128	518	18
14	5	37	437	149
III	240	236	145	456
15	7	493	7	370
16	47	477	77	402
17	27	265	32	503
18	94	101	144	537
19	66	215	254	362
20	1	57	2	606.

Source: Jones (1969, p.99)

المجموعات والاقاليم

ركزنا في نقاشنا حتى الآن على المشكلات العامة وعمليات التصنيف ، ولم نتطرق إلا إلى جزء صغير من مجال التصنيف العددي (سوكال وسنيث، ١٩٦٣) . في بعض مجالات الدراسة ، يواجه الجغرافيون مشكلةً ، وإن لم تكن مقتصرة على الجغرافيا ، إلا أنها شائعة في تخصصنا ؛ ألا وهي مسألة **التشابه بين المجموعات والمناطق** (جريج، ١٩٦٥، ١٩٦٧) .

تقع المنطقة في صميم الكثير من العمل الجغرافي . عادةً ما يتم التعرف على نوعين من المناطق : (1) المناطق الرسمية ، التي تضم أماكن ذات خصائص متشابهة ؛ و(2) المناطق الوظيفية ، أو العقدية ، التي تضم أماكن ذات أنماط ارتباط مماثلة بأماكن أخرى . يمكن تقسيم كل نوع من هذه الأنواع إلى: (1) أنواع إقليمية ، و(2) مناطق . الفرق بين النوعين الأخيرين هو أنه بينما يشمل النوع الإقليمي أماكن متشابهة في خصائص محددة مسبقًا (المظاهر الطبيعية ، والهياكل السكانية ، وما إلى ذلك) ، فإن المنطقة تنطوي أيضًا على قيد التجاور - يجب أن تشمل المنطقة وحدات متجاورة مكانيًا . وبالتالي ، لدينا مخطط تصنيف 2 × 2 ، كما هو موضح أدناه مع مثال لكل خلية :

	Regional Type	Contiguous Region
Formal	Urban Ghettos (e.g. Morrill, 1965)	Political constituencies (e.g. Taylor, 1973)
Functional	World trade groups (e.g. Russett, 1967)	School catchment zones (e.g. Shepherd and Jenkins, 1972)

لا يتطلب إنتاج الأنواع الإقليمية أي إجراءات خاصة ؛ فأي طريقة تصنيف قابلة للتطبيق على مجموعة البيانات المحددة ستكون ذات صلة . ومع ذلك ، قد تكون هناك حاجة إلى طرق خاصة للمناطق المتجاورة . من بين عدد من الطرق المقترحة (ينظر، على سبيل المثال، تايلور، 1969) ، يمكن تحديد نهجين أساسيين . يقترح النهج الأول (Czyz 1970، Johnston؛1968)، اعتماد إجراءات التصنيف المعتادة ، وأنه عند تحديد الأنواع الإقليمية ، يجب إجراء اختبارات لمعرفة ما إذا كانت تُشكل أيضًا مناطق متجاورة . أما البديل ، وهو الأكثر استخدامًا ، فيُدخل قيدًا على عملية التجميع .

قيود التجاور

يمكن توضيح استخدام قيود التجاور بالعودة إلى قرانا الست عشرة . تُظهر الخريطة الأصلية (الشكل 1) موقع هذه القرى ، ومن خلالها يُمكننا استخلاص مصفوفة تجاور (الجدول 6) ، حيث يُشير الرقم 1 في الخلية إلى أن القرى المعنية (أرقام الصفوف والأعمدة) لها حدود مشتركة ، ويُشير الرقم 0 إلى أن القريتين غير متجاورتين . (لاحظ أنه قد نواجه مشاكل عند التقاء أربعة حدود ؛ فقد قررنا ، على سبيل المثال ، أن القرى 3 و 9، وكذلك القرى 6 و10، ليس لها حدود مشتركة).

ننتقل الآن إلى التجميع الهرمي باستخدام إجراء استبدال مركز الثقل كما في السابق ، ولكن مع فارق واحد ؛ قبل تجميع قريتين ، نتحقق في المصفوفة مما إذا كانتا متجاورتين ، وإذا لم تكونا كذلك ، فإننا لا نُجمّعهما وننتقل إلى أقصر مسافة تالية . وبالتالي ، فإن أقصر مسافة في الجدول 1 هي 3.0 ، بين القريتين 2 و3 ، ولكن يُظهر لنا الجدول 6 والشكل 1 أنهما ليس لهما حدودا مشتركة ، وبالتالي لن تُشكلا منطقة . أقصر قيمة تالية هي 5.4، للقرى 3 و4 ، ولكنهما أيضًا غير متجاورتين . في الواقع ، أقصر مسافة ننتقل بها إلى التجميع هي 7.1 ، حيث تُصنف القرى 2 و4 معًا ، وكذلك القرى 13 و14 و15 (ولكن ليس 11 و12).

بعد كل تجميع ، يجب إعادة حساب مصفوفتي المسافة والتجاور ، مع أنه يُمكننا كبديل حساب مصفوفة تُحسب فيها المسافات يتم إدخال الوحدات المتجاورة فقط ، وتُستبدل جميع المسافات الأخرى بقيمة غير منطقية (مثلاً ١٠٠٠,٠) . وبالطبع ، لا يمكن استخدام هذه القيمة الأخيرة في حساب مصفوفة المسافات الجديدة ؛ إذ يلزم الرجوع إلى القيمة الأصلية . بالنسبة لقرانا الست عشرة ، تم التجميع الكامل مع قيد التجاور على النحو الاتي :

Table 6 Contiguity Matrix for Figure 2[a]

	Place															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	-	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
2	0	-	1	1	0	0	0	0	0	0	0	0	0	0	0	1
3	1	0	-	0	0	1	0	0	0	1	0	1	0	0	0	0
4	1	1	0	-	1	1	0	0	0	0	0	0	0	0	0	0
5	0	1	0	1	-	1	0	1	1	0	0	0	0	0	0	1
6	1	0	1	1	1	-	0	0	1	0	0	0	0	0	0	0
7	0	0	0	0	0	0	-	1	0	0	1	0	1	0	0	0
8	0	0	0	0	1	0	1	-	1	0	1	0	1	0	0	1
9	0	0	0	0	1	1	0	1	-	1	0	0	1	1	0	0
10	0	0	1	0	0	0	0	0	1	-	0	1	0	1	1	0
11	0	0	0	0	0	0	1	1	0	0	-	0	0	0	0	1
12	0	0	1	0	0	0	0	0	0	1	0	-	0	0	1	0
13	0	0	0	0	0	0	1	1	1	0	0	0	-	1	0	0
14	0	0	0	0	0	0	0	0	1	1	0	0	1	-	1	0
15	0	0	0	0	0	0	0	0	0	1	0	1	0	1	-	0
16	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	-

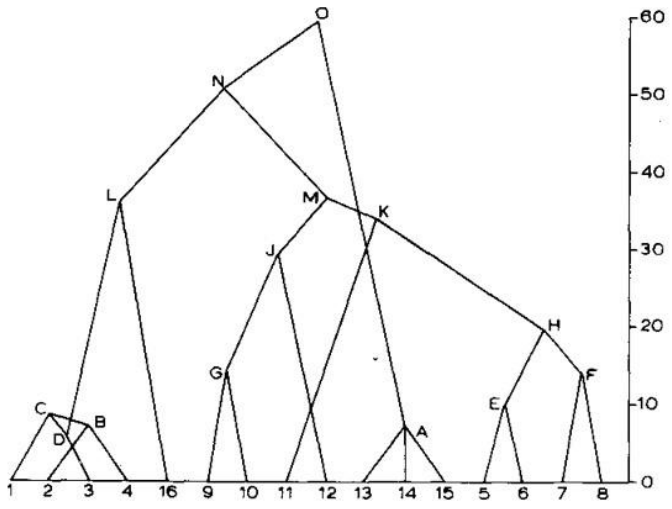
[a] 1 = two places are contiguous 0 = non-contiguous

شجرة الارتباط موضحة في الشكل 11. تشير المقارنة بين هذا وذاك في التجميع غير المقيد (الشكلان 5 و11؛ حيث وُضعت القرى في الموضع نفسه على المحورين الأفقيين لتسهيل المقارنة) إلى الاختلافات الرئيسية . وتشمل هذه الاختلافات أولًا ، القريتين 11 و12 ، واللتين ، نظرًا لبعدهما المكاني الكبير، لا تُجمعان معًا إلا في الخطوة الأخيرة ، بل اثنتين فقط في الحل المقيد ، مقارنةً بالخطوة الثالثة في الحل غير المقيد ؛ وثانيًا ، الهوية المنفصلة حتى الخطوة الأخيرة لمجموعة القرى الأقرب إلى المدينة (13، 14، 15) . لاحظ أيضًا النقطة المثيرة للاهتمام وهي أن المسافة في الخطوة الثالثة كانت أقصر من المسافة في الخطوتين الأخريين ، نتيجة التجميع السابق للقرى 1 و2 و4 ؛ الرقم 1 فقط مجاور للرقم 3، وهو الأقرب في فضاء المتغيرات (الشكل 2) إلى الرقمين 2 و4 .

Step	Places Grouped	Distance	Forming Group	Average Intra-Group Distance
1	(13-14-15	7.1	A	8.1
	(2-4	7.1	B	7.1
2	B-1	8.6	C	8.1
3	C-3	5.9	D	7.0
4	5-6	10.0	E	10.0
5	(7-8	14.1	F	14.1
	(9-10	14.1	G	14.1
6	E-F	19.6	H	17.1
7	G-12	29.2	J	24.2
8	H-11	33.9	K	23.8
9	D-16	35.9	L	18.6
10	J-K	36.7	M	
11	L-M	50.9	N	
12	A-N	59.2	O	

التجميع والتقسيم الإقليمي : مثال

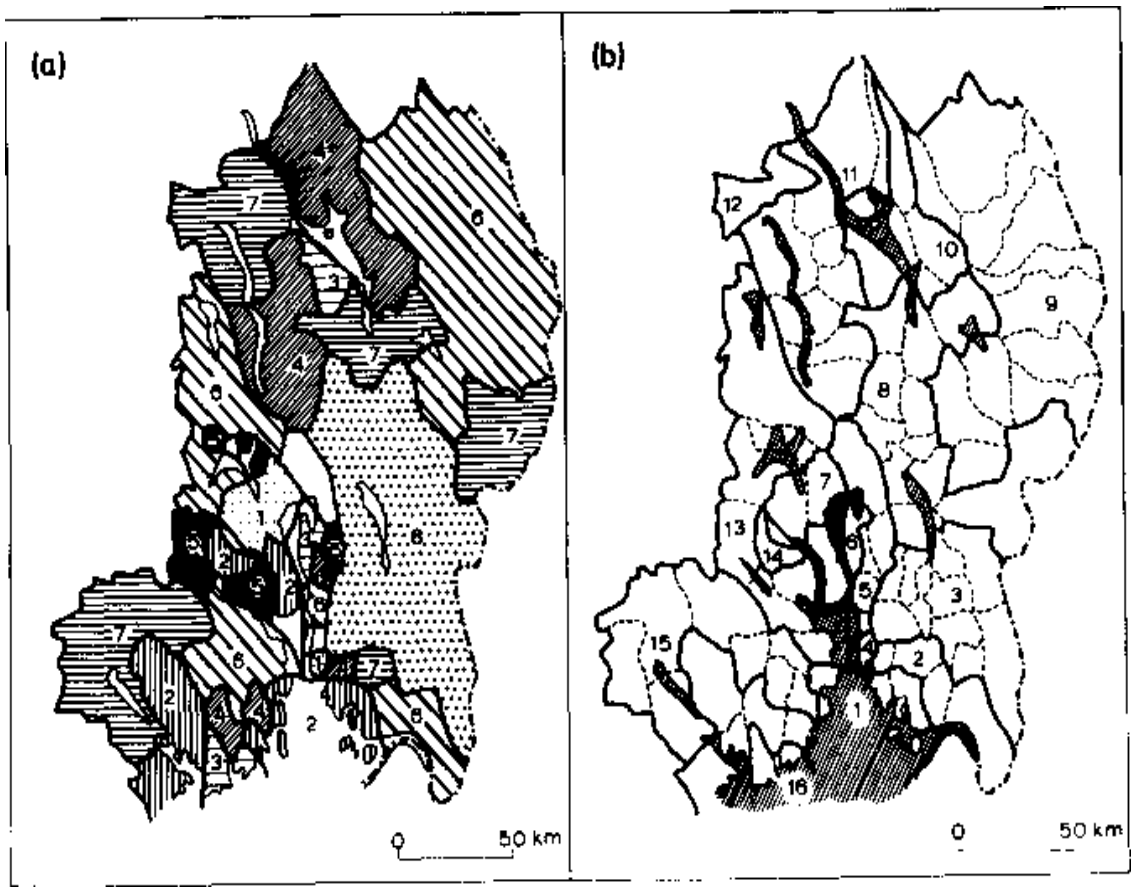
من الأمثلة على إجراءات التقسيم الإقليمي ، باستخدام طريقة وارد ، ورقة بحثية حديثة لبايفوجلين و نوردغارد (1974) حول أنواع الزراعة في شرق النرويج . أشارت بياناتهما إلى 86 وحدة مكانية (بلديات ومجموعات من البلديات) ، ولكل منها معلومات عن ستة متغيرات تتعلق بالنشاط الزراعي . تم اختصار المتغيرات الستة إلى أربعة ، متعامدة ، باستخدام تحليل المكونات الرئيسية PCA .



في تحليلهم لهذه البيانات ، أنتج المؤلفون نوعًا من تصنيف المناطق الزراعية وإقليمية لأنواع الزراعة . بالنسبة للأول ، لم يُفرض أي قيد على التجاور، وحددوا ثمانية أنواع مختلفة من المناطق الزراعية . ولأن معظم هذه الأنواع تضم مجموعات عدة من المناطق غير المتجاورة ، فقد تألفت الخريطة الناتجة من ثلاثين منطقة منفصلة (الشكل 12أ) . وفي إقليميتهم ، التي استخدمت الإجراءات نفسها بالإضافة إلى قيد التجاور، تم تحديد ست عشرة منطقة من الأنواع الزراعية (الشكل 12ب).

ستوضح مقارنة الخريطتين في الشكل 12 الاختلافات بين تحليل النوع الإقليمي (12أ) وتحليل المناطق المتجاورة (12ب) في تقسيمهما للمنطقة نفسها . من الواضح أن الباحث يجب أن يكون متأكدًا من نوع التصنيف الذي يرغب في إنتاجه . وكدليل عام ، من المحتمل أن يكون إجراء النوع الإقليمي أكثر "كفاءة" في جميع الدراسات تقريبًا ، ما لم تكن هناك حاجة إلى إقليمية متجاورة . هذا يُشار إلى النقطة في الشكل 13، حيث يُعبَّر عن مجموع مربعات المسافات داخل المجموعات كنسبة مئوية من مجموع مربعات جميع المسافات في المصفوفة الأصلية .

على يسار هذا الرسم البياني ، حيث تُحدَّد جميع الوحدات الـ 86 بشكل منفصل ، تكون النسبة المئوية للداخل / الإجمالي صفرًا . كلما تحرك المرء إلى اليمين ، مع عدد أقل من المجموعات ، تزداد النسبة المئوية ، ببطء في البداية ثم بسرعة . في كل خطوة ، تكون النسبة المئوية لتحليل النوع الإقليمي أقل من تلك الخاصة بتحليل المنطقة المتجاورة ، لأن المجموعات الأولى تجمع الوحدات الأكثر تشابهًا ، بينما لا تفعل الأخيرة ذلك إلا إذا كانت متجاورة أيضًا . في هذا ، كما هو الحال في العديد من الجوانب الأخرى لإجراء التصنيف ، يمكن أن يكون لاختيار الباحث تأثير كبير على النتيجة . المناطق المتجاورة هي حالة خاصة من الأنواع الإقليمية، ويجب على المحلل التأكد ، قبل أن يتابع ، من المنتج النهائي الذي يرغب فيه ، وما هي المنتجات الثانوية التي سينتج عنها .



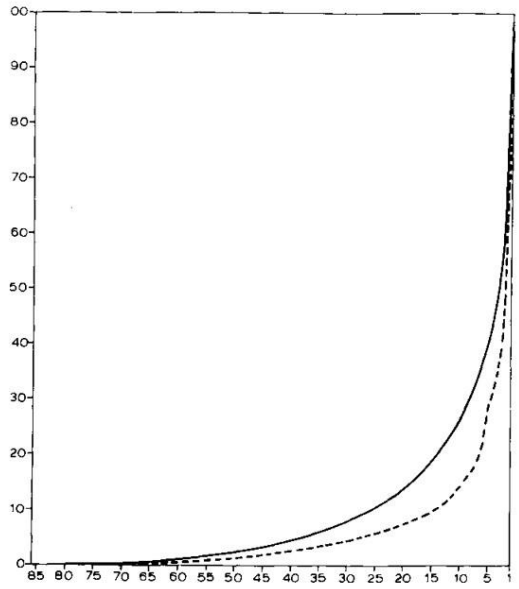


Fig. 13 The efficiency of the classifications shown in Figure 12. On the vertical axis, the within-group distances as a percentage of the total distances are shown; on the horizontal axis the number of groups is indicated. The solid graph refers to the grouping with contiguity constraint, the pecked line to the grouping without such a constraint (after Byfuglien and Nordgard, 1974).

طرق التقسيم

ركزت معظم المناقشات في هذه المقدمة الموجزة لإجراءات التصنيف في الجغرافيا على الطرق التجميعية ، نظرًا لكونها الأكثر استخدامًا . نبدأ في هذه الطرق بمعالجة كل ملاحظة على حدة كمجموعة منفصلة ، ثم ننتقل إلى تجميع المجموعات . تتوفر مجموعة أخرى من التقنيات ، والتي تعمل بطريقة معاكسة تمامًا : تبدأ بافتراض أن جميع الأفراد ينتمون إلى مجموعة واحدة ، ثم تقسم المجموعة إلى مجموعات فرعية . مرة أخرى ، تم ابتكار مجموعة متنوعة من التقنيات - مثل التحليل الارتباطي التكراري بين الأعمدة (ماكويتي وكلارك، 1968) - ولكننا سنتناول هنا طريقة واحدة فقط ، وهي تحليل الارتباط ، والتي استُخدمت على نطاق واسع لنوع معين من البيانات .

تحليل الارتباط

طور هذه الطريقة عالما نبات ، ويليامز ولامبرت (1959) ، للتعامل مع تصنيف المجتمعات النباتية . يُعد هذا التحليل مناسبًا بشكل خاص للبيانات الاسمية (المعروفة غالبًا بالبيانات الثنائية) ، التي تُسجل فقط وجود أو غياب نوع (متغير) في موقع ما (ملاحظة) ، بدلاً من أي شدة لوجود النوع . ويمكن ، بالطبع ، إعادة كتابة أي بيانات مُقاسة بمقاييس أخرى - ترتيبية أو فاصلة أو نسبية - في صيغة ثنائية للافادة من هذه الطريقة .

يعتمد تحليل الارتباط على الاختلافات بين الأنواع في توزيعها على المواقع ، ويستخدم إحصاء مربع كاي . يُقدم هنا مثال على عمله باستخدام البيانات الافتراضية للجدول 7 ، والتي تُظهر وجود/غياب أعشاش ثمانية أنواع مختلفة من الطيور (أ-ح) في عشرين عينة من غابات مختلفة . يُمثل الوجود بالرقم 1 في الخلية ذات الصلة ؛ ويُمثل الغياب بالرقم صفر. (لقد اتخذنا الوجود/الغياب خطًا فاصلًا ، ولكن كان بإمكاننا ، على سبيل المثال ، ترميز أي غابة تحتوي على أقل من خمسة أعشاش لكل 100 متر مربع بالصفر) . تتم عملية التجميع على النحو الاتي . أولًا ، يُحسب إحصاء مربع كاي (× ٢) لتوزيعات كل زوج من

الأنواع . يتضمن ذلك إعداد جدول ٢ × ٢ . تُدرج الخلية (أ) عدد المواقع (الغابات) التي يتواجد فيها كلا النوعين ، وتُسمى (ب) المواقع التي تحتوي على النوع (ص) دون النوع (س) ، وهكذا : N هو العدد الإجمالي للمواقع . ثم يُحسب مربع كاي ، باستخدام الصيغة الشائعة لتطبيقه على جدول 2 × 2 :

		Species x: present	Species x: absent	
Species y	present	a	b	a+b
	absent	c	d	c+d
		a+c	b+d	N

$$\chi^2 = \frac{(ad-bc)^2 N}{(a+b)(a+c)(b+d)(c+d)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Table 7 Distribution of bird species by woodlands

Woodland	Species A	B	C	D	E	F	G	H
1	1	0	1	0	1	0	1	0
2	1	1	1	0	0	.0	1	1
3	1	1	1	0	0	0	0	0
4	1	1	1	0	0	0	1	1
5	0	0	0	1	1	1	0	0
6	1	0	0	0	1	1	1	1
7	0	0	1	0	0	0	0	0
8	1	1	1	0	0	0	1	1
9	0	1	1	1	0	0	1	0
10	1	1	1	0	0	0	0	1
11	0	0	1	1	0	0	1	0
12	1	0	1	0	0	0	1	1
13	0	1	0	1	1	1	0	0
14	0	0	0	0	1	1	1	1
15	1	0	0	0	1	0	0	0
16	0	0	0	1	0	1	1	1
17	0	1	0	1	1	1	0	1
18	0	0	0	1	1	0	1	0
19	0	0	0	1	1	1	0	1
20	0	0	0	1	1	1	1	0

For species A and B in Table 7, the 2 x 2 contingency table and the calculation for χ^2 are:

		Species A: present	Species A: absent	
Species B	present	5	3	8
	absent	4	8	12
		9	11	20

$$\chi^2 = \frac{((5x8) - (4x3))^2 \times 20}{(5+3) \times (5+4) \times (3+8) \times (4+8)}$$

$$= ((40-12)^2 \times 20)/8 \times 9 \times 11 \times 12 = (28^2 \times 20)/9504$$

$$(784 \times 20)/9504 = 15680/9504 = 1.6$$

Table 8 Inter-Species chi-square statistics

	Species A	B	C	D	E	F	G	H
A	-	1.6	5.1	13.4	1.8	5.7	0.4	1.8
B	1.6	-	3.3	0.4	3.0	1.2	0.5	0.8
C	5.1	3.3	-	5.1	12.8	13.3	0.8	0
D	13.4	0.4	5.1	-	1.8	5.1	0.1	1.8
E	1.8	3.0	12.8	1.8	-	7.5	0.8	0.8
F	5.7	1.2	13.3	5.1	7.5	-	0.6	0.8
G	0.4	0.5	0.8	0.1	0.8	0.6	-	0.8
H	1.8	0.8	0	1.8	0.8	0.8	0.8	-
	29.8	10.8	40.4	27.7	28.5	34.2	4.0	6.8

The full 8 x 8 matrix of inter-species χ^2 values is presented in Table 8.

في المرحلة الثانية من الإجراء ، نقرر أي الأنواع سنستخدم لتقسيم الغابات إلى مجموعتين . النوع المختار هو الأكثر اختلافًا عن جميع الأنواع الأخرى ، وبالتالي يُفترض أنه أفضل مُميز بين أنواع الغابات . بما أن حجم الفرد X4 يرتبط ارتباطًا إيجابيًا بدرجة الاختلاف بين توزيعات نوعين ، فإن النوع الأكثر اختلافًا بشكل عام هو النوع الذي يمتلك أكبر مجموع لقيم x2 السبع . لذلك ، تُحسب مجاميع الأعمدة ، وتُعطى في الجدول 8. ومن هذه المجاميع ، يُلاحظ أن النوع C هو الأكثر اختلافًا ، وهذا يُشكل أساس التقسيم : تُوضع جميع الغابات التي تحتوي على النوع C في مجموعة ، وتلك التي لا تحتوي على النوع C في مجموعة أخرى ، مما يُعطي عضوية المجموعة الاتية :

Group	Members
I - C present	1,2,3,4,7,8,9,10,11,12
II - C absent	5,6,13,14,15,16,17,18,19,20

لدينا الآن مجموعتان ، ويُكرَّر الإجراء الكامل لكلٍّ منهما على حدة . أُعِدَّت جداول توزيع الأنواع السبعة المتبقية (الجدول 9) ؛ ولم يعد للنوع "ج" استخدامٌ في التصنيف ، إذ تشير المجاميع الناتجة إلى أنه ينبغي تقسيم المجموعة الأولى بناءً على وجود أو غياب النوع "أ" ، بينما يُمثِّل النوع "د" المُميِّز في المجموعة الثانية . يُعطي هذا أربع مجموعات ، تُكرَّر العملية في كلٍّ منها . يستمر التقسيم حتى تحتوي جميع المجموعات على عضوين أو أقل ، وبعد ذلك يصبح التصنيف الإضافي بلا معنى . أحيانًا ، تُقسَّم المجموعات الأكبر.

Table 9 Distribution of bird species by two woodland groups

Group I									Group II							
Woodland	Species							Woodland	Species							
	A	B	D	E	F	G	H		A	B	D	E	F	G	H
1	1	0	0	1	0	1	0	5	0	0	1	1	1	0	0
2	1	1	0	0	0	1	1	6	1	0	0	1	1	1	1
3	1	1	0	0	0	0	0	13	0	1	1	1	1	0	0
4	1	1	0	0	0	1	1	14	0	0	0	1	1	1	1
7	0	0	0	0	0	0	0	15	1	0	0	1	0	0	0
8	1	1	0	0	0	1	1	16	0	0	1	0	1	1	1
9	0	1	1	0	0	1	0	17	0	1	1	1	1	0	1
10	1	1	0	0	0	0	1	18	0	0	1	1	0	1	0
11	0	0	1	0	0	1	0	19	0	0	1	1	1	0	1
12	1	0	0	0	0	1	1	20	0	0	1	1	1	1	0

Table 10 Inter-species chi-square statistics

	Group I							Group II						
	Species							Species						
	A	B	D	E	F	G	H	A	B	D	E	F	G	H
A	-	1.3	6.0	0.5	0	0	4.3	-	0.6	6.0	0.3	1.4	0	0
B	1.3	-	0.1	1.6	0	0.1	1.6	0.6	-	1.1	0.3	0.6	2.5	0
D	6.0	0.1	-	0.3	0	1.1	2.5	6.0	1.1	-	0.5	0.5	0.5	0.5
E	0.5	1.6	0.3	-	0	0.5	1.1	0.3	0.3	0.5	-	0.3	1.1	1.1
F	0	0	0	0	-	0	0	1.4	0.6	0.5	0.3	-	0	2.5
G	0	0.1	1.1	0.5	0	-	0.5	0	2.5	0.5	1.1	0	-	0.4
H	4.3	1.6	2.5	1.1	0	0.5	-	0	0	0.5	1.1.	2.5	0.4	-
E	12.1	4.7	10.0	4.0	0	2.2	10.0	8.3	6.1	9.1	3.6	5.3	4.5	4.5

لا يمكن تقسيمها أكثر من ذلك ، لأن جميع أعضائها متشابهة تمامًا . في مثالنا ، تُشكل الغابات ٢ و٤ و٨ مجموعة من ٣ غابات في الخطوة الرابعة . لكل من هذه الغابات الثلاثة درجة حضور لثلاثة من الأنواع التي ما تزال قيد الدراسة (B وG وH) ودرجة غياب للنوعين الآخرين (D وF) ؛ وهي متشابهة من جميع النواحي ، والتقسيم الإضافي الوحيد هو إلى غابات فردية فريدة . (لن يكون تفردها قائمًا على معايير ثمانية أنواع من الطيور مستخدمة في التصنيف ، ولكن ربما معيار آخر (الموقع) . ويتم التجميع الكامل على النحو الاتي :

Step	Group	Divide on Species		Forms Groups of Woodlands	Largest X^2
1		C	I II	1,2,3,4,7,8,9,10,11,12 5,6,13,14,15,16,17,18,19,20	13.3
2	I II	A D	III V	1,2,3,4,8,10,12.IV 7,9,11 5,13,16,17,18,19,20.VI 6,14,15	6.0 6.0
3	III IV V VI	E D and G G F,G, and H	VII IX XI XIII	1 XIII 2,3,4,8,10,12 7 X 9,11 5,13,17,19 XII 16,18,20 15 XIV 6,14	2.9 3.0 2.1 3.0
4	VIII XI XII	G B and H E and H	XV XVII	2,4,8,12 XVI 3,10 four groups of one member each 16 XVIII 18,20	2.8 3.0
5	XV	B	XIX	2,4,8 XX 12	-
6	XIX			division not possible - three members perfectly alike	

كما هو الحال مع إجراءات التجميع ، نواجه السؤال الاتي : "**متى يجب إيقاف التجميع** ؟" الإجابة عشوائية عمومًا ، عندما تكون هناك درجة حرية واحدة فقط ؛ (غريغوري، ١٩٧٢) . اختيار مستوى ٠,٠٥ بدلًا من أي مستوى آخر يعكس العر ف . علاوة على ذلك ، لا يُجدي اختبار الدلالة الإحصائية نفعًا في حال عدم استخدام العينات . تُوفر القاعدة أداة تقييم مفيدة . مع ذلك ، الإجراء ، مما يُشير إلى أنه - وفقًا لمعيار الدلالة الإحصائية - يجب أن تنتهي العملية بعد الخطوة ٢. تشير شجرة الارتباط مع قيم X ٢ على المحور الرأسي إلى الإجراء (الشكل ١٤) . في هذا ، تشير الحروف الموجودة عند الانقطاعات في عملية الانقسام إلى النوع الذي تم إجراء التقسيم عليه ، ويُشار إلى عدد أعضاء كل مجموعة داخل الدائرة ، ويُظهر تحديد المجموعة بالأرقام الرومانية ، وتشير العلامات الإيجابية (الحضور) والسلبية (الغياب) إلى المجموعة التي تحتوي على النوع الذي حدث عليه الانقسام ، والتي لم يحدث عليه .

كما هو الحال مع جميع التقنيات الأخرى التي نوقشت هنا ، تم إدخال مجموعة من الاختلافات الطفيفة على موضوع تحليل الارتباط ، مع إمكانية تضمين قيود مختلفة ، وتُنتج الطريقة أنواعًا إقليمية فقط . تتمثل إحدى مزايا الطريقة في سهولة تطبيقها على مصفوفة البيانات في كلا البعدين . حلل تحليلنا الاختلافات بين المتغيرات (أنواع الطيور) وصنف الغابات ، ولكننا تمكنا من تصنيف أنواع الطيور بالقدرنفسه ؛ تُعرف الطريقة الأولى باسم تحليل الارتباط "العادي" والثانية باسم تحليل الارتباط "العكسي"

استُخدم تحليل الارتباط على نطاق واسع في العمل البيئي والجغرافي الحيوي ، حيث غالبًا ما تكون بياناته ثنائية (فرينكل وهاريسون، 1974) ، ولكنه ذو قيمة في مجموعة من المجالات الأخرى . على سبيل المثال ، استخدم كارو (1968) كلاً من النهج "العادي" و"العكسي" في دراسة للهيكل الوظيفي للمستوطنات في شرق أنجليا عام 1846. وتضمنت البيانات ترميز الوجود/الغياب لـ 61 دالة تجارية وخدمية لـ 76 مستوطنة . واقترح التحليل الطبيعي وجود إحدى عشرة مجموعة من المستوطنات عند مستوى 0.05، بينما أشار التحليل العكسي إلى اثنتي عشرة مجموعة من الوظائف المركزية . ويشير دمج هاتين المجموعتين في رسم تخطيطي (الشكل 15) ، حيث تمثل النقطة وجود دالة في تلك المستوطنة ، إلى نظام

مكان مركزي هرمي واضح يتراوح من مراكز السوق الإقليمية إلى القرى الصغيرة . وهذا في الواقع إعادة ترتيب لمصفوفة البيانات الأصلية ، في شكل يسلط الضوء على الترتيب بين خلاياها البالغ عددها 4636 خلية .

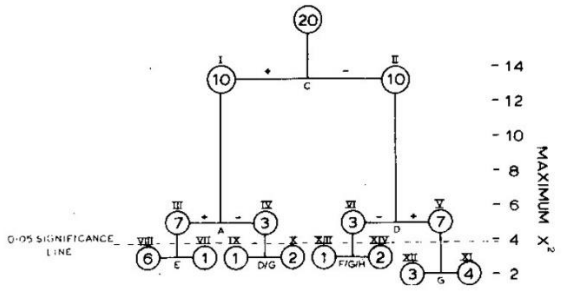
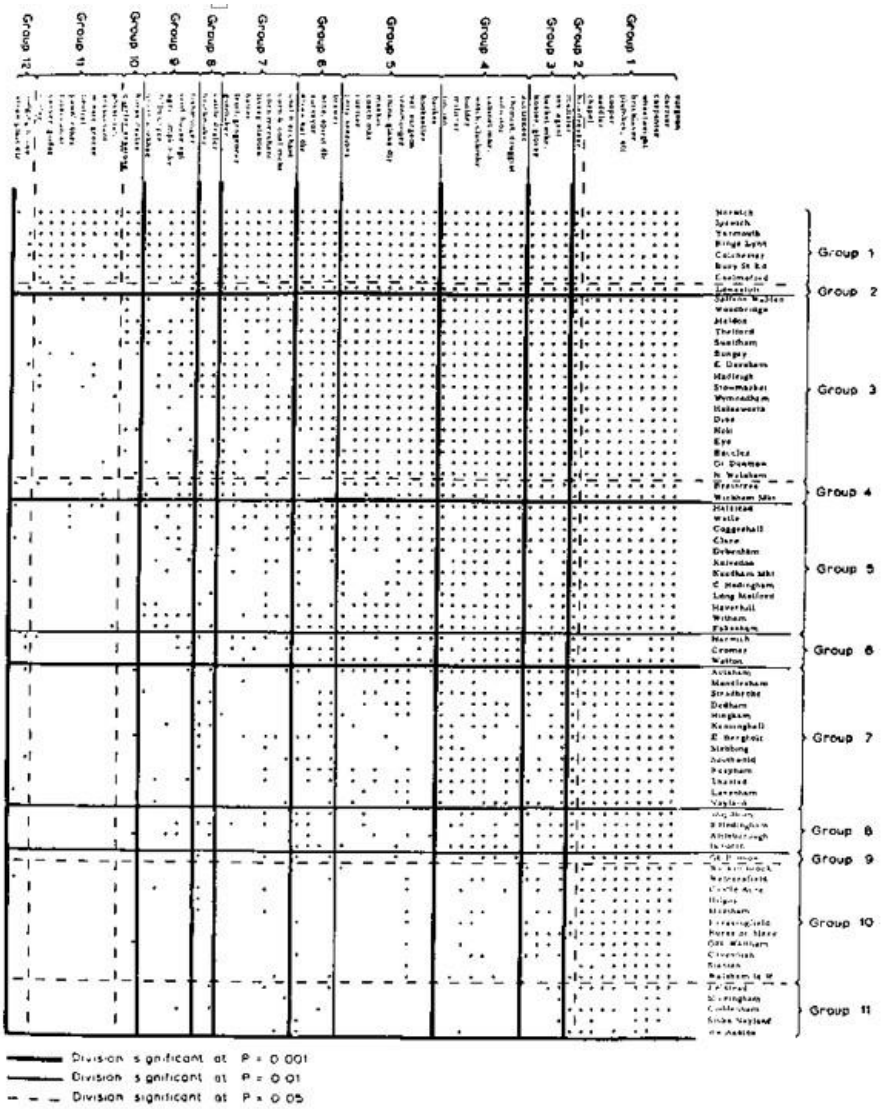


Fig. 14 Linkage tree for Association Analysis of the Woodlands in Table 7.



ASSOCIATION ANALYSIS: FINAL GROUPING BY TRADES AND SERVICES AND BY SETTLEMENTS, 1846

Fig. 15 Groups of Settlements and of Central Functions in East Anglia 1876, according to Association Analysis (after Carqe, 1968, p.264).

الاستنتاجات

يُعدّ التصنيف مجالًا واسعًا ، كما تُشير العديد من المراجعات الحديثة (إيفريت، 1974؛ كورماك، 1971؛ سيبسون وجاردين، 1971) . لم يُستخدم سوى جزء صغير نسبيًا من الطرق المتاحة في البحث الجغرافي ، ومع ذلك ، فإن هذا يعني تباينًا كبيرًا في الأدبيات من حيث الإجراءات الفعلية المُستخدمة . لم يُناقش هنا سوى بعض التقنيات الأبسط والأسهل تطبيقًا ، على الرغم من أن قواعد عملها تُشبه قواعد العديد من التقنيات الأخرى التي لم يسمح لنا المجال بدراستها . ربما يُمكن استخلاص ثلاثة استنتاجات رئيسية ذات صلة من الأمثلة التي ناقشناها .

أولها هو أنه على الرغم من أن الإجراءات التي بحثناها واضحة نسبيًا ، إلا أنه إذا كنا نتعامل مع مجموعة بيانات مهما كان حجمها - لا سيما من حيث عدد الملاحظات المراد تجميعها (كما في دراسة جونز) - فإن العملية سرعان ما تُصبح مُملة وتستغرق وقتًا طويلاً ، وبالتالي قد تكون عُرضة لخطأ بسيط . لحسن الحظ ، أصبحت برامج الحاسوب متاحة على نطاق واسع لتخفيف الملل ، ولكن هنا يكمن الاستنتاج الثاني.

ثانيا ، قد يؤدي الاختيار غير الدقيق للطريقة إلى نتائج قد تكون مضللة ، وأحد مخاطر توافر البرامج - للعديد من الإجراءات الإحصائية وليس التصنيف فقط - هو **أن الناس قد يختارون طريقة ما لمجرد توفر البرنامج بدلاً من أسباب منطقية** . عندما يقدم البرنامج عددًا كبيرًا من الخيارات (مثل إجراء CLUSTAN الشائع Wishart :، 1969) فقد يكون الاختيار أكثر عشوائية . ومن المشاكل الأخرى المتعلقة بإجراءات الحاسوب أن معظمها ، بسبب حجم الذاكرة المحدود ، لا يمكنه التعامل مع مجموعات البيانات الكبيرة ومصفوفات المسافات ، مما دفع العديد من الباحثين (مثل1969 Pocock وWishart) إلى البحث عن طرق ممكنة لتقليل الجهد المبذول .

ينبثق الاستنتاج **النهائي** من الاستنتاجين الأولين . على الرغم من إدراكنا لضرورة التصنيف ، إلا أن النقاش بين الجغرافيين حول مشاكل التعريفات المعنية كان محدودًا . ليس لدينا نظريات متطورة للمناطق والأنواع الإقليمية لمقارنتها بتلك التي طورها ، على سبيل المثال ، علماء النفس التربوي لأغراضهم الخاصة (ماكويتي، 1967) . بدأ جريج (1965، 1967) بعض المناقشات على أساس مخططه الدقيق لمنطق الإقليمية (ينظر بونج، 1966أ؛ جريج، 1966)، ولكن معظم الجغرافيين كانوا مهتمين أكثر بإيجاد تقنية والمضي قدمًا في التصنيف بدلاً من التأكد من ماهية التصنيف الذي يصنفونه . لذلك ، **يُعد التصنيف مرحلة حاسمة في التطور العلمي** . لم يتم حل جميع المشكلات التقنية (على الرغم من ادعاء بونج 1966ب) ، ولكن تتوفر مجموعة من الأساليب سهلة التطبيق للباحث . يتطلب الاختيار من بين هذه الأساليب بعض الاختيار، وبالتالي دراسة متأنية لغرض الدراسة ، وكذلك تقييم المخرجات .

التعليق

يعد التصنيف Classification الخطوة الاولى للانتقال من المعرفة العامة الى مستوى العلم في الشيئ ، فهو الاساس الذي ستبنى عليه النتائج اللاحقة . وطرائق التصنيف عديدة جدا ، ومتنوعة ، و متناقضة في النتائج احيانا . فالتصنيف قد يركز على التشابه في الخصائص Similarities ، او التباين فيها Variation . و قد تكون النتيجة متشابهة بدرجة كبيرة ، ولكن الهدف من وراء التصنيف يختلف كليا . ولان الخريطة تمثل العمود الفقري للدراسة الجغرافية ، و رسمها مبني على اساس التصنيف المكاني في الخصائص ، لذا فان معرفة جيدة بطرائق التصنيف و استخدامها بموضوعية يساهم عمليا في توضيح صورة الجغرافيا عند المختصين و غيرهم .

من خلال ما ورد في هذا الكتيب ، و سياقات عرض طريقة التصنيف على اساس التجاور المكاني و الموضوعي ، لتقنية محددة من طرائق التصنيف العديدة ، يتوجب على الجغرافي الحصيف ان يدرك ان ما يكتبه و ينشره من اوراق (بحثية او مقالات) لها اهميتها و انعكاسها على القارئ ، سواء اكان مختصا ام لا . وبما ان طرائق التحليل و التصنيف اضحت متيسرة جدا على الحاسوب ، و تجذب الانظار و تستهوي الباحثين الشباب ، فان مزالق سوء الاختيار والتطبيق ، و عدم فهم الطريقة بشكل سليم يؤديان حتما الى تشويه صورة الاختصاص ، و كذلك صورة الكاتب ايضا في الوقت نفسه . فما يكتب وينشر انما يعبر عن قدرة الكاتب على اختيار الموضوع و تقنياته المناسبة و تفسيره للنتائج بشكل سليم .

يضاف الى ذلك ، فان النهج التطبيقي في الجغرافيا خاصة و اتجاه العلم عامة لمعالجة مشاكل البيئة والمجتمع Problem Oriented Approach ، و توجه الجامعات لتطبيق وظيفتها الثالثة (خدمة المجتمع) ، و اعتماد الابحاث والدراسات الاكاديمية كمعلومات اساسية لصناعة القرارات الرسمية Decision making يستوجب بالضرورة ان تكون الاوراق البحثية رصينة بكل جوانبها ، من حيث : اختيار الموضوع و البيانات والمعلومات ، و اعتماد طرائق تحليل مناسبة توصل الى الهدف ، و النتائج ، والاستنتاجات ، و تفسيرها و عرضها نصيا و بصريا .