

البيانات الثانوية

كيفن سانت مارتين و ماريانا بافلوفسكايا

الفصل 11

ترجمة بتصرف

أ.د. مضر خليل عمر

مقدمة

البيانات الثانوية هي بيانات لا يُنتجها الباحثون بأنفسهم ، بل يستخدمونها في أبحاثهم. مقارنةً بالبيانات الأولية التي تُنتج خلال العمل الميداني (والتي تتضمن ، على سبيل المثال ، قياس جودة المياه أو إجراء مقابلات مع المستجيبين) ، فإن البيانات الثانوية تُنتج بالفعل من قبل جهة أخرى . تشمل جهات توفير البيانات الثانوية الوكالات الحكومية والشركات الخاصة ، أو مصادر مثل الدراسات العلمية المنشورة ، أو الأرشيفات ، أو المجموعات . يشير مصطلح البيانات الثانوية عادةً إلى قواعد بيانات كبيرة نسبياً لا يستطيع الباحثون الأفراد جمعها بأنفسهم ، مثل بيانات التعداد السكاني ، أو أرشيفات الصحف ، أو قوائم جرد الموارد ، أو صور الأقمار الصناعية . على الرغم من تسميتها بالبيانات الثانوية ، إلا أن هذه البيانات تُثري قدرًا كبيرًا من العمل الأكاديمي ، وتُشكل محورًا أساسيًا لتخصصات فرعية كاملة في العلوم الاجتماعية والبيئية .

علاوة على ذلك ، من المرجح أن تزداد أهمية البيانات الثانوية في البحث وتطوير السياسات مع مرور الوقت . ويرجع ذلك إلى أن تقنيات المعلومات قد سهّلت انتشارًا واسعًا لمجموعة واسعة من المعلومات الرقمية البيئية والاجتماعية والاقتصادية ، بالإضافة إلى أساليب تحليلها . وبتوافرها على نطاق واسع وقبولها كبيانات شرعية ، أصبحت البيانات الثانوية تؤثر بشكل كبير على نوع المعرفة التي ننتجها وكيفية إنتاجها . ويتطلب انتشار البيانات الثانوية ، وخاصةً في شمال الكرة الأرضية ، إجراء تقييم دقيق لإمكاناتها وقيودها قبل دمجها في أي مشروع بحثي أو استخدامها للإجابة عن أسئلة بحثية محددة .

يتناول هذا الفصل بعض القضايا المتعلقة باستخدام الجغرافيين للبيانات الثانوية . ونشير إلى التنوع الكبير للبيانات الثانوية ومصادرها المتعددة ، ونناقش مزاياها وقيودها المهمة . كما نتناول بعض القضايا ذات الأهمية الخاصة للجغرافيين ؛ وهي المغالطة البيئية ومشكلة الوحدة المساحية القابلة للتعديل (MAUP) من حيث صلتها بالبيانات الثانوية . أخيرًا ، نُبرز ضرورة التعامل بإبداع ونقد مع البيانات الثانوية من خلال التركيز على مناهج تحليلية غير تقليدية تستخدم أساليب بحثية مختلطة . وفي هذا السياق ، نعتمد على أمثلة من أعمالنا وأعمال طلابنا في الجغرافيا الحضرية وإدارة الموارد .

أنواع ومصادر متعددة للبيانات الثانوية

تشمل البيانات الثانوية أنواعًا مختلفة من المعلومات حول العمليات الطبيعية والبشرية ، والتي تجمعها جهات حكومية ومنظمات غير حكومية وشركات مختلفة . ومن أمثلة هذه البيانات : بيانات تعداد السكان ، وإحصاءات الصحة ، ونتائج التحصيل الدراسي ، وبيانات مراقبة الطقس ، والصور الملتقطة بالاستشعار عن بُعد ، وقياسات درجة حرارة سطح المحيط ، وحسابات وفرة المخزون السمكي ، وكميات المواد الخطرة المُطلقة في البيئة ، ونتائج استطلاعات الرأي العام وغيرها من المسوحات السكانية أو التجارية ، بالإضافة إلى البيانات التي غالبًا ما تُعرض على شكل خرائط ، مثل أنماط التصويت ، واستخدام الأراضي ، والارتفاع

في الولايات المتحدة ، تُجمع وتُوزَّع الكثير من البيانات الثانوية من قِبل منظمات حكومية ، مثل مكتب الإحصاء ، ووكالة حماية البيئة (EPA) ، والمعهد الوطني للصحة (NIH) ، والإدارة الوطنية للمحيطات والغلاف الجوي (NOAA) ، وهيئة المسح الجيولوجي الأمريكية (USGS) ، على سبيل المثال لا الحصر . بالإضافة إلى ذلك ، تجمع وتبيع العديد من الوكالات الخاصة كميات كبيرة من البيانات . وتشمل هذه الوكالات شركات الاستشارات العقارية والبيئية ، وشركات التأمين والتمويل ، وشركات التسويق ، وغيرها . وأخيراً ، تُعيد عدد من الوكالات الخاصة معالجة البيانات التي تجمعها الحكومة ، وغالباً ما تُنجز الكثير من العمل المطلوب قبل تحليل هذه البيانات بفعالية ، أو تعمل كموزعين لمنتجات بيانات قد لا ترغب الحكومة في إنتاجها بكميات كبيرة .

تشمل البيانات الثانوية بوضوح معلومات متفرقة تنشأ من مجموعة واسعة من المواقع . وبالتالي ، يجب افتراض أن هذه البيانات ستختلف اختلافاً كبيراً من حيث شكلها ونوعها ، وتغطيتها المكانية أو الزمنية ، والفئات أو التصنيفات التي تُنظَّم من خلالها . في كثير من الحالات ، تُحدِّد هذه المؤهلات مدى فائدة مجموعة بيانات مُعيَّنة لمشروع بحثي مُعَيَّن . بالإضافة إلى ذلك ، لكل طريقة جمع ، تقنية التسجيل والتجميع ، وتُدمج مجموعة البيانات الناتجة في السياق التاريخي والاجتماعي للوكالة أو المؤسسة التي طورتها . على سبيل المثال ، تُجمع بيانات الهيئة الوطنية لمصايد الأسماك البحرية (NMFS) وتُسجل كوسيلة لتقييم وفرة مخزون الأسماك كميّاً .

يتجلى هذا التركيز بوضوح في تفويض الهيئة التاريخي بإدارة موارد مصايد الأسماك بحيث يمكن الحصول على أقصى إنتاج ، بدلاً من الحفاظ على مجتمعات الصيد ، على سبيل المثال . وبالتالي ، تُعنى مجموعة البيانات الأساسية لهذه الهيئة في المقام الأول بكمية الأنواع المختلفة من الأسماك في البحر . يتم أخذ عينات بحرية من مجموعات الأسماك باستخدام شبكة مكانية بدقة مناسبة لأخذ العينات الإحصائية ، ولكنها غير دقيقة للغاية للدراسات على مستوى المجتمع . المعلومات المُجمَّعة كمية ونوعية (أي أنها تتكون من قياسات رقمية ، بالإضافة إلى أنها تعكس الاختلافات في النوع) ، ولكن نظراً لتخزين المعلومات في قاعدة بيانات ، فإن هذه الأخيرة تقتصر على أوصاف سلاسل قصيرة بدلاً من النص المفصل الذي قد يُنتج من العمل الميداني (الموجود في ملاحظات العمل الميداني) .

كذلك ، في حين أن التغطية الزمنية لمجموعات بيانات الخدمة الوطنية لمصائد الأسماك البحرية الأساسية مثيرة للإعجاب (عدة عقود) ، فإن الكثير من البيانات التي قد تكون مفيدة لعلماء الاجتماع (مثل حجم طاقم سفن الصيد) لم تُجمع إلا منذ عام 1994 . وأخيراً ، يتم تنظيم بيانات الخدمة الوطنية لمصائد الأسماك البحرية ، التي قد تساعد في التحليل الاجتماعي والاقتصادي ، حسب فئة سفن الصيد بدلاً من المشاركين الفرديين في قطاع الصيد . هذا يجعل التحليل الاجتماعي والاقتصادي على مستوى مصائد الأسماك (فيما يتعلق ، على سبيل المثال ، بالقضايا المتعلقة بالتوظيف ، ووصف الوظيفة ، والأجور والمزايا، ومدة العمل) مستحيلاً تقريباً .

في حين أن الخدمة الوطنية لمصايد الأسماك البحرية (NMFS) مكلفة بجمع البيانات المتعلقة بمصايد الأسماك في الولايات المتحدة ، فمن الواضح أن البيانات التي يتم جمعها محدودة الفائدة لعلماء الاجتماع المهتمين بنطاق المجتمع ، أو مسائل التوظيف ، أو التغير الاجتماعي والاقتصادي بمرور الوقت . وبالتالي ، تتنوع البيانات الثانوية بشكل كبير ، وتُنتجها مجموعة واسعة من المنظمات ، وتعكس التاريخ الفريد لتلك المنظمات . ومع ذلك ، هناك العديد من القضايا المشتركة بين مجموعات البيانات . وينطبق هذا بشكل خاص

على تخزين المعلومات بشكل متزايد داخل قواعد البيانات الرقمية التي تشترك في مبادئ التنظيم ، وطرق الاستعلام ، وأشكال إعداد التقارير.

من الورق إلى قواعد البيانات الرقمية

قبل بضعة عقود فقط ، كانت البيانات الثانوية موجودة على الورق فقط ؛ كانت جميع التحويلات والحسابات تُجرى يدوياً أو باستخدام آلة حاسبة . وكان الورق هو الوسيط الذي تُخزَّن عليه البيانات ويُستخدم لتوفير نتائج أي استعلام أو عملية تحليلية أو تفسير . واليوم ، تُنشأ وتُخزَّن وتُحلَّل وتُورَّع الكثير من البيانات الثانوية رقمياً ، وخاصةً في مجتمعات ما بعد الصناعة . وقد حُلَّت جداول البيانات الرقمية وقواعد البيانات العلائقية محل الجداول المطبوعة . وأثار ذلك عميقة . فعلى سبيل المثال ، ازداد حجم البيانات المنشأة والمُخزَّنة بشكل كبير ، وأصبح الوصول إلى مجموعات البيانات أسرع بكثير ، ويمكن ربط مجموعات البيانات الموجودة في مواقع مختلفة عن بُعد لتعمل كقاعدة بيانات واحدة عبر الإنترنت ، كما يُمكن استيراد قواعد البيانات الضخمة جدًّا وتصورها وتحليلها بسهولة باستخدام حزم برامج متنوعة ، بما في ذلك برامج التحليل الإحصائي ونظم المعلومات الجغرافية (ينظر الفصل 22).

غالبًا ما تُنظَّم البيانات الثانوية الرقمية في قواعد بيانات ، مُنظمة كجدول واحد أو عدة جداول يُمكن ربطها منطقياً وفقاً لسمات مُشتركة (تُسمى قاعدة بيانات علائقية) . في هذه الجداول ، تُمثَّل الصفوف حالات فردية ، مثل : محطات ، أو قطع أراضي ، أو مناطق تعداد سكاني ، وتمثَّل الأعمدة (أو الحقول) خصائصها أو متغيراتها الكمية أو النوعية . وبينما توجد بيانات ثانوية كثيرة غير منظمة في قواعد بيانات علائقية ، فمن الواضح أن هناك اتجاهًا نحو هذا الاتجاه حتى بالنسبة للبيانات التي لا ترتبط عادةً بشكل جدولي أو حتى بتخزين رقمي . على سبيل المثال ، تُنظم المقالات الصحفية الآن في الغالب بشكل رقمي وتُفهرس كحالات داخل قاعدة بيانات .

ويمكن لعمليات البحث والاستعلام نفسها ، التي يمكنها ، على سبيل المثال ، إرجاع جميع مواقع عينات البحر التي رُصد فيها (أو لم يُرصد) عدد معين من صغار سمك القد من قبل عالم من الخدمة الوطنية لمصايد الأسماك البحرية ، أن تُرجع ، في ظل قاعدة بيانات مختلفة تمامًا ، جميع المقالات الصحفية المنشورة في السنوات الخمس الماضية والتي تذكر أزمة مصايد سمك القد وفقدان سبل العيش المحلية . حتى أرشيفات المعلومات المرئية، مثل الصور والخرائط ، تُنظم عبر جداول علائقية وتُحكمها نفس المبادئ والمنطق . بشكل عام ، ازداد حجم المعلومات الرقمية بشكل كبير خلال العقدين الماضيين ، وسيستمر في الازدياد في المستقبل . وعلى الرغم من تسميتها بالثانوية ، إلا أن هذه الأنواع من البيانات أصبحت المصدر "الأساسي" للعديد من مشاريع البحث . ولذلك ، من المهم فهم مزاياها وقبورها الأساسية ، بالإضافة إلى السياسات المحيطة بإنتاج البيانات الثانوية وتوزيعها واستخدامها.

مزايا البيانات الثانوية

من بين المزايا الواضحة للبيانات الثانوية ، والتي سنتناولها بإيجاز ، نطاقها وحجمها ، وجودتها المهنية وإمكانية الوصول إليها ، وارتباطها بالمراجع المكانية . توفر سمات البيانات الثانوية هذه فرصًا لأشكال معينة من التحليل لم تكن لتوجد لولاها . ومع ذلك ، وكما أوضح العديد من مؤرخي العلوم ، فإن نوع البيانات التي يتم جمعها ، ونطاقها وشكلها ، وفئاتها ، ومخططات تصنيفها ، ستعزز مصالح البعض وليس الجميع . على سبيل المثال ، قد نشير إلى كيفية توافق مجموعات بيانات إدارة مصايد الأسماك البحرية الوطنية مع مصالح

الشركات وقطاع صيد الأسماك واسع النطاق ؛ في الواقع ، إن العلاقة الوثيقة بين القوة الاقتصادية وجمع البيانات برعاية الدولة ليست نادرة (أو غير موثقة) . ومع ذلك ، فإن الفحص الدقيق لأي مجموعة بيانات يمكن أن يكشف عن قدرتها على القيام بعمل غير مقصود أو غير متصور ، ويمكن لعلماء الاجتماع إعادة تفسير المتغيرات داخل مجموعات بيانات الخدمة الوطنية لمصايد الأسماك البحرية بطرق جديدة . على سبيل المثال ، يتم جمع عدد أفراد الطاقم على متن سفينة كمؤشر على ضغط الصيد ، وهو متغير مهم في التقييمات البيولوجية لمخزون الأسماك ، ولكن يمكن إعادة تفسير هذا المتغير نفسه واستخدامه كمؤشر على العمالة وتغيرها بمرور الوقت ، على الرغم من دفته داخل جدول معني بمخزون الأسماك بدلاً من التحليل الاجتماعي والاقتصادي .

إن مسألة من يمكن أن تكون البيانات الثانوية ميزة له ، سواء كان مالكو السفن من الشركات أو أفراد الطاقم أو المنظمات العمالية ، لم يتم حلها أبدًا ؛ الميزة هي أن البيانات الثانوية متاحة لمن يرغب في قضاء الوقت للتعرف عليها ، ومن ثم الاستفادة من نطاقها وشرعيتها وسهولة الوصول إليها.

النطاق

معظم البيانات الثانوية ، نظرًا لتغطيتها المكانية الواسعة وكمية المعلومات التي تُجمع ، لا بديل لها . الباحثون الأفراد أو حتى فرق البحث لا يمكن إنتاج مجموعات بيانات مماثلة في الحجم أو النطاق . على سبيل المثال ، تُغطي إحصاءات السكان الحكومية الأراضي الوطنية ومجموعات سكانية بأكملها . وتنتج مئات المتغيرات للوحدات المكانية المفصلة ، وذلك بمعدل مرة كل عشر سنوات . كما تُصنف قواعد بيانات العقارات مخزون المساكن بتفصيل كبير ، وفي بعض البلدان ، تُدمج قواعد بيانات العقارات المحلية عبر الإنترنت بحيث يمكن الاستعلام عن مئات الآلاف من العقارات . ولا تُجمع جرد الموارد ، مثل مصايد الأسماك ، على المستوى الوطني فحسب ، بل تُدمج أيضًا في النظم الدولية لجمع البيانات والإبلاغ عنها ، مثل قواعد بيانات مصايد الأسماك التابعة لمنظمة الأغذية والزراعة (الفاو) التابعة للأمم المتحدة ، مما يُتيح إجراء تحليلات بيئية عالمية . بالإضافة إلى ذلك ، تحتوي بعض مجموعات البيانات الثانوية على بيانات يكاد يكون من المستحيل تقنيًا الحصول عليها دون استثمار حكومي كبير ، بما في ذلك منتجات الاستشعار عن بُعد (ينظر الفصل 10).

غالبًا ما يبدأ جمع البيانات بحالات فردية ومناطق صغيرة ؛ ثم تُجمع البيانات لتشمل حالات متعددة ووحدات مكانية أكبر . في الماضي ، عندما كانت البيانات تُخزن على الورق ، كانت التجميعات ثابتة . على سبيل المثال ، كانت بيانات التعداد المُخططة تُجمع إلى نوع واحد من الوحدات المكانية ، وهو منطقة التعداد ، بحيث لا توجد فعليًا إلا بهذا النطاق (بالإضافة إلى البيانات الخام) . مع قواعد البيانات الرقمية ، لم تعد مستويات التجميع ثابتة ، وفي معظم الحالات ، يجب على مُزوّد البيانات أو الباحث تحديد مستوى التجميع المناسب والمطلوب في ضوء المشروع الجاري ؛ وبالتالي ، يُمكن تحديد كتل ومجموعات التعداد بالإضافة إلى مناطق التعداد.

علاوةً على ذلك ، إذا تم استلام المقياس المكاني لمجموعة بيانات ثانوية على مستوى ثابت أو على شكل خريطة ثابتة ، مثل بيانات التعداد على مستوى الولاية (عندما يركز مشروع ما على المجتمعات المحلية) ، فقد تنتج الجهة التي أنشأت البيانات أيضًا البيانات على مقاييس أخرى ، بدقة مكانية أدق ، مثل المقاطعة أو الرمز البريدي أو منطقة التعداد . في الواقع ، غالبًا ما توجد البيانات في شكل يمكن إخراجها بمقاييس متنوعة تختلف عن مقياس منتجات البيانات القياسية .

المصادقية

عادةً ما تكون المعلومات الواردة في مجموعات البيانات الثانوية منظمة بشكل متسق ، مما يجعلها مناسبة للعديد من أنواع التحليل الكمي أو الإحصائي ، وهو غالبًا ما يكون السبب الرئيسي لجمع هذه البيانات . بالإضافة إلى ذلك ، عادةً ما يُنشئ البيانات الثانوية متخصصون مُدرَّبون تدريبًا خاصًا ، يُجرون اختبارًا مُسبقًا للأسئلة ويتحققون من الفئات لإنتاج معلومات قياسية وقابلة للمقارنة ، عبر الزمان والمكان ، والتي يُمكن استخدامها لفحص الاتجاهات أو مقارنة المعلومات عبر وحدات جغرافية مُماثلة ، مثل المقاطعات أو المحافظات . كما يُتيح الشكل الموحد للبيانات الثانوية للباحثين تصميم مشاريع جمع بيانات تُضيف إلى مجموعات البيانات الثانوية الحالية أو يُمكن مقارنتها بها.

ومن المهم أن تُضفي الأنظمة المهنية لجمع البيانات وتجميعها وتخزينها واسترجاعها ، والتي تُشكل البيانات الثانوية ، شرعية مُعترف بها على نطاق واسع ، وتعمل على تمكين البيانات الثانوية ، وجعلها مُقنعة بلاغياً ، والسماح لها بالإقناع بطرق لا تستطيع مجموعات البيانات الأخرى تحقيقها . على سبيل المثال ، تُشتق العديد من مجموعات البيانات من معلومات مُشكوك فيها يتم الإبلاغ عنها ذاتيًا من قِبل الشركات أو الأفراد أو مُستخدمي الموارد . ومع ذلك ، بمجرد تجميعها في شكل مُتسق ومُنظم ، تُصبح هذه المعلومات ، على الرغم من أصولها ، أساسًا للتحليلات العلمية الرسمية . في مصايد الأسماك ، على سبيل المثال ، تُعدّ دفاتر السجلات شكلاً من أشكال الإبلاغ الذاتي ، حيث يُبلغ قباطنة السفن عن صيد الأسماك ، والتخلص منها ، وموقع الرحلة ، ومتغيرات أخرى إلى دائرة مصايد الأسماك البحرية الوطنية .

وبينما تُستهزأ قصص كل صياد على حدة كونها قصصاً أو مبالغاً ، فإن إدخالات سجلاتهم تُضفي عليها مصداقية من خلال الأنظمة التقنية التي تُدمج فيها . وبالمثل ، نلاحظ أن قدرًا كبيرًا من البيانات المكانية الرقمية (طبقات الخرائط) المتوفرة حاليًا في قواعد البيانات الثانوية قد حُولت إلى صيغة رقمية من خرائط ورقية قد تكون قديمة ، أو مُدخلة من نقاط تحكم متفرقة ، أو مُرمزة جغرافيًا بشكل غير صحيح (مثل موقع السفارة الصينية في بلغراد) . ومع ذلك ، فإن هذه الطبقات ، بمجرد أن تُصبح رقمية ، تبدو دقيقة وتُضفي ثقة على التحليلات التي تُجرى.

إمكانية الوصول

ومن المهم أن أكبر مجموعات البيانات وأكثرها شمولاً ، مثل بيانات التعداد السكاني ، غالبًا ما تُنتجها الجهات الحكومية ، وتكون متاحة للجمهور بتكلفة منخفضة أو مجاناً . وهذا يجعلها في متناول الأكاديميين ، وكذلك المحللين العاملين في المنظمات غير الحكومية والمنظمات الشعبية ، الذين يمكنهم تحليل هذه البيانات وفقاً لاحتياجاتهم أو قضاياهم السياسية (الوود، ٢٠٠٦) . وبشكل عام ، يُسهّم هذا التوسع في ديمقراطية التقنيات والمعلومات الرقمية في تمكين مجموعة متنوعة من الجهات الفاعلة الاجتماعية خارج نطاق الدولة والشركات . كما يُسهّل تزايد إمكانية الوصول إلى البيانات الثانوية استخدامها كخطوة أولى استكشافية في مشاريع البحث التي تُركز بعد ذلك على جمع البيانات الأولية .

ويمكن استخدام البيانات الثانوية ، المتاحة على نطاق واسع وبأسعار معقولة وسهلة الاستخدام ، لاستهداف عمليات جمع البيانات الأولية المكلفة والمستهلكة للوقت بكفاءة أكبر . من بين أمور أخرى ، تُستخدم البيانات غالبًا لتحديد الأماكن و/أو السكان لإجراء دراسات نوعية أو كمية أكثر تعمقاً . في أحد المشاريع ، على سبيل المثال ، استخدمنا بيانات التعداد السكاني لتحديد الأحياء داخل مدينة نيويورك التي تضم أعدادًا كبيرة من الناطقين بالإسبانية والروسية . بالإضافة إلى ذلك ، فُحصت معلومات على مستوى البلدية (متوفرة على موقع إدارة التعليم في نيويورك) لتقدير عدد الطلاب المهاجرين الملحقين بالمدارس العامة في تلك

المناطق نفسها . وجمع هذه البيانات ، أمكن تحديد الأحياء التي يقيم فيها مهاجرون جدد مع أطفال صغار. ثم أصبحت هذه الفئات السكانية هدفًا لمشروع بحثي رئيسي قائم على المقابلات ، ركز على الممارسات الاقتصادية المتعددة للأسر المهاجرة.

ومع ذلك ، فبينما تبدو البيانات الثانوية منتشرة في كل مكان من منظور الشمال العالمي ، إلا أن هناك حدودًا لإمكانية الوصول إليها . وعلى وجه الخصوص ، توجد فجوة كبيرة في القدرات النسبية للدول الغنية والفقيرة على الوصول إلى المعلومات الرقمية وإنتاجها واستخدامها والتحكم فيها . وبما أن هذه الفجوة تعكس اختلافات في القوة الاقتصادية والسياسية ، فإن المجتمعات المتقدمة ما بعد الصناعية تتمتع بمزايا واضحة . ومع ذلك ، فإن دول الجنوب العالمي تدرك بشكل متزايد ضرورة تضيق الفجوة الرقمية ، ومع ازدياد تكلفة التقنيات الرقمية وسهولة استخدامها ، تُطلق حكوماتها مشاريعها الخاصة لجمع البيانات . كما تملأ الشركات الدولية بنوك بياناتها الرقمية بمعلومات حول أسواق الموارد والعمالة والاستهلاك الجديدة في الجنوب العالمي. يبدو أن التغطية الرقمية للعالم ، للأفضل أو للأسوأ ، تتوسع بسرعة ، وتوفر مصادر معلومات أكثر فأكثر للبحث. كجغرافيين ، نلاحظ أيضًا أن تزايد إمكانية الوصول إلى البيانات الثانوية الرقمية يرتبط ارتباطًا وثيقًا بنمو التقنيات الجيومكانية ، بحيث يعني الوصول إلى البيانات الثانوية بشكل متزايد الوصول إلى البيانات ذات المراجع الجغرافية . يُجمع جزء كبير من البيانات في مجموعات البيانات الثانوية إما حسب وحدات مكانية ، مثل مناطق التعداد أو الدوائر الانتخابية ، أو يتضمن معلومات موقعية أخرى مثل عنوان الشارع أو الإحداثيات الجغرافية . وبالتالي ، يمكن تصور هذه البيانات واستكشافها وتحليلها باستخدام نظم المعلومات الجغرافية (GIS) . للتعامل مع البيانات الثانوية العديد من المزايا ؛ بما في ذلك مزايا النطاق والحجم ، والمصادقية الواسعة ، وسهولة الوصول إليها المتزايدة باستمرار ، مما يجعلها مصدرًا لا يُضاهى للمعلومات الاجتماعية والبيئية للجغرافي . ومع ذلك ، لا يمكننا الاعتماد على البيانات الثانوية دون تمحيص . من المهم تذكر أنه ينبغي تقييم مزايا البيانات الثانوية مقارنةً بقيودها التي قد تكون شديدة في بعض الأحيان.

القيود

على الرغم من المزايا العديدة للبيانات الثانوية ، إلا أن استخدامها قد يُضيق ، ومن المفارقات ، فُرض البحث ويُقل من جودة النتائج . في هذا القسم ، سنناقش قيود البيانات الثانوية التي قد تُعيق مشروعك دون دراسة نقدية . سنناقش كيف أن البيانات الثانوية لا تُنشأ صراحةً لمشروعك المُحدد ، وكيف يُمكن أن تُصبح مجموعات البيانات غير مُتسقة داخليًا بمرور الوقت أو عبر المكان ، وكيف يُمكن أن يستند ما يبدو تغطيةً كاملةً إلى أخذ العينات ، وكيف يُمكن ألا تُمثل هذه البيانات المجتمع الذي تعتقد أنها تُمثله ، وكيف يجب موازنة دقتها مع قضايا الخصوصية والأخطاء وعدم الدقة المكانية.

لأي غرض أنشئت البيانات؟

يعني استخدام البيانات الثانوية أننا نستخدم البيانات التي أنشأها شخص آخر ولأغراضه الخاصة للإجابة عن أسئلة بحثنا المُحددة . حتى مجموعات البيانات الكبيرة مُتعددة الاستخدامات تُنظَّم وفقًا لغرض أصلي . على سبيل المثال، تُجمع بيانات التعداد السكاني لأغراض التصويت أو الضرائب، بينما تُستخدم بيانات الخدمة الوطنية لمصايد الأسماك البحرية للتقييم البيولوجي لمخزون الأسماك . ويؤثر التصميم الأولي ، المُدمج في البيانات ، على أسئلة بحثنا وأساليبه ونتائجه ، ويُحدّ منها . على سبيل المثال ، سيكون من الصعب للغاية دراسة بعض جوانب تغير المناخ العالمي التي لم تُدمج بالفعل في مجموعات البيانات العالمية

الموجودة مسبقًا . وتُجمّع هذه الأخيرة العديد من المتغيرات التي تهّم الباحثين في تغير المناخ العالمي ، ولكن ليس جميعها بالتأكيد . وبالمثل ، يجب على عالم الاجتماع الذي يبحث في الفقر أن يعتمد على تعريف مُحدد للدخل مُدمج في فئات تعداد مُحددة . أي أن التعداد عادةً ما يُبلغ عن الدخل النقدي الرسمي للأسرة ، ولكن من غير المُرجّح أن يشمل أنواعًا أخرى من النشاط الاقتصادي ، مثل الإنتاج غير الرسمي و/أو غير مدفوع الأجر للسلع والخدمات التي قد تكون مُهمة للتكيف مع الفقر.

العمل المنزلي، والعمل غير الرسمي مقابل المال، ورعاية الأطفال في المنزل، والتبادلات بين الأسر وداخل المجتمع ، لا تقل أهمية عن الأجور الرسمية في إعادة الإنتاج الاجتماعي ، إلا أنها غائبة عن بيانات التعداد السكاني . وكما هو الحال في استخدام علماء الاجتماع لقواعد بيانات الخدمات الوطنية للأمن الغذائي والمائي ، يمكن إعادة تفسير الفئات المصممة لغرض معين بشكل إبداعي لأغراض أخرى . ومع ذلك ، فإن إعادة التفسير هذه محدودة بتاريخ وسياق الوكالة أو المنظمة ، وهو ما ينعكس بدوره في قواعد البيانات التي تُنشئها . من الواضح أن مجموعات البيانات الثانوية تُنشأ وتُحفظ لأغراض محددة ، وبالتالي ، قد لا تكون قابلة للاستخدام إلا إذا تمكن الباحثون من إعادة تفسير البيانات الموجودة بشكل إبداعي ، أو كما هو الحال في كثير من الحالات ، تعديل أسئلة بحثهم الأصلية لتناسب مع البيانات .

إن استخدام البيانات الثانوية لديه القدرة على الحد من الإمكانات التحليلية ، بحيث قد تظل أسئلة البحث الأصلية ، في النهاية ، دون إجابة . تتغير ممارسات جمع البيانات ، لا تبقى ممارسات جمع البيانات واسعة النطاق ثابتة ، ولا يتحكم الباحثون الذين يستخدمون بيانات ثانوية في هذه التغيرات . حتى في مجموعة بيانات موحدة ومتسقة مثل تعداد الولايات المتحدة ، قد تتغير الفئات التحليلية (المتغيرات) أو حدود الوحدات المكانية ، مثل مناطق التعداد ، من عقد إلى آخر . بالإضافة إلى ذلك ، غالبًا ما تُضاف متغيرات جديدة وتُفصل الوحدات المكانية الحالية . وبالتالي ، يصبح إجراء المقارنات الطولية (الزمنية) أمرًا صعبًا ، كما هو موضح بالإشارة إلى مشكلة الوحدة المساحية القابلة للتعديل (MAUP) أدناه (الملحق 11.1) ، وأحيانًا يكون الأمر ببساطة مستحيلًا.

ممكّن ، على سبيل المثال ، بعد انهيار الاتحاد السوفيتي ، قلّصت الإدارة الجديدة في موسكو عدد المقاطعات الرئيسية من 33 مقاطعة عام 1992 إلى 9 مقاطعات فقط عام 1993. وقد تم ذلك بهدف تعديل هياكل السلطة في الحقبة السوفيتية جذريًا والانفصال عنها . ونتيجةً لذلك ، لم يعد من الممكن مقارنة الوضع الاجتماعي والاقتصادي في موسكو مباشرةً قبل وبعد الانتقال إلى الرأسمالية ، لأن جميع البيانات الاجتماعية والاقتصادية كانت مرتبطة بأحد نظامين مكانيين لا يمكن مقارنتهما . وللأسف ، فُقدت فرص بحثية بالغة الأهمية والمثيرة للاهتمام بسبب التغيير في جمع البيانات وتنظيمها (بافلوفسكايا 2002).

تغطية كاملة أم عينة مُقسّمة؟

على الرغم من حجمها ونطاقها ، فإن القليل من قواعد البيانات تُغطي بالكامل السكان الذين تدّعي تمثيلهم . في أغلب الأحيان ، تُقدّر المتغيرات من عينات مختارة ، وبالتالي ، قد تكون عُرضة لأخطاء العينة والتحيزات . بصفته تعدادًا سكانيًا ، يُقدّم مكتب الإحصاء الأمريكي تغطيةً شاملةً كل عشر سنوات . ومع ذلك ، تُكمل أسرة واحدة فقط من كل ست أسر "النموذج المطول" الذي يطلب بعضًا من أهم البيانات الاجتماعية والاقتصادية . وبالمثل ، تُقدّم ما يُسمى بالبيانات الجزئية من التعداد (عينة البيانات الجزئية للاستخدام العام أو PUMS) معلومات مُفصلة عن الوحدات السكنية والسكان فيها (على عكس المناطق الجغرافية مثل كتل التعداد أو مجموعات الكتل أو المناطق) . وبينما تُتيح هذه البيانات جدولة المعلومات بطرق لا تُتيحها مجموعة

بيانات التعداد العادية ، فإن النتائج لا تكون صحيحة إلا إذا كانت العينة كافيةً وتتضمن عددًا صحيحًا إحصائيًا من الحالات (ينظر الفصل 6) .

في حالة مصائد الأسماك ، تُنظَّم عينات الأسماك البحرية التي تُجرىها الخدمة الوطنية لمصائد الأسماك البحرية (NMFS) بدقة مكانية مناسبة لإجراء جرد إقليمي وآليات تنظيمية إقليمية واسعة النطاق قد تضع ، على سبيل المثال ، قيودًا على حجم الصيد . ومع ذلك ، فإن الدقة التي تُجرى بها الخدمة الوطنية لمصائد الأسماك البحرية (NMFS) عيناتها تجعل تقييم مخزونات الأسماك المحلية وموائلها أمرًا صعبًا في أحسن الأحوال . ونتيجةً لذلك ، يجد صيادو الأسماك الصغار الذين تُمارس ممارسات صيدهم محليًا صعوبةً في فهم تصريحات الخدمة الوطنية لمصائد الأسماك البحرية (NMFS) بشأن صحة / نفوق المخزون ، ويجدون أن اللوائح الإقليمية اللاحقة لا تتوافق مع تجاربهم أو احتياجاتهم المحلية.

"الصمت" في البيانات الثانوية

مجموعات البيانات الثانوية هي في الأساس تمثيلات جزئية . فهي لا تحتوي إلا على معلومات حول ظواهر مختارة أو جوانبها ، وبالتالي ، تغفل دائمًا معلومات حول ظواهر أخرى أو جوانبها . والنتيجة هي إسكاتٌ وتهميشٌ فعليٌّ للعمليات أو الأشخاص أو الأماكن غير المُمَثَّلَة . على سبيل المثال ، لا تُوصف سوى أنواع مُعَيَّنَة من الظواهر "الرسمية" من خلال البيانات الاجتماعية والاقتصادية التي تجمعها وكالات الدولة بانتظام . ومع ذلك ، فإن هذه الظواهر ليست سوى غيضٍ من فيض ، إذ لا تُوثَّق الممارسات الاقتصادية أو الاجتماعية غير الرسمية الأخرى ، وتظل غير مرئية ضمن مجموعات البيانات التي تُقرأها الدولة (جيبسون - غراهام 2006).

وبالمثل ، في حين أن العمليات البيئية والتغيرات التي تطرأ عليها بمرور الوقت تتأثر بوضوح بالأنشطة البشرية ، فإن بعضها فقط يُؤخذ في الحسبان في قواعد البيانات الرسمية المتعلقة بالإدارة البيئية ، مما يُعيق تقديرنا لمحركات التغير البيئي . على سبيل المثال ، في حين تُراقب أنشطة الصيد التجاري بعناية مقارنةً بإدارة مصائد الأسماك ، فإن الصيد الترفيهي وصيد الكفاف لا يُراقب ، على الرغم من الاعتقاد بأن لهذه الأنشطة تأثيرًا كبيرًا على موارد مصائد الأسماك المُحددة . تُحتسب الأنشطة الرسمية وتُقاس ، وبالتالي ، يُمكن إدراجها ضمن قاعدة بيانات ثانوية . من الواضح أن توثيق التوظيف الرسمي ، وخدمات الرعاية الصحية الرسمية ، ورعاية الأطفال الرسمية ، والاستهلاك الرسمي ، وجمع الموارد الرسمي ، وما شابه ذلك ، أمرٌ بالغ الأهمية ؛ إلا أنها لا تستطيع استيعاب مجمل التجربة الإنسانية .

لا توجد معلومات متسقة بشأن الأسر غير الرسمية ، والشبكات ، واستخدام الموارد لمعيشة الكفاف ، أو اقتصادات المجتمع . ما هو أو من هو غير مُدرج ، أو ما يُعرف بـ "الصمت" في مجموعات البيانات الثانوية ، سيؤثر بوضوح على قدرتنا على بناء تفسيرات باستخدام البيانات الثانوية ، وسيحد منها . علاوة على ذلك ، قد لا تُمثل مجموعات البيانات التي تستهدف ظاهرة رسمية مُعَيَّنَة هذه الظاهرة تمثيلًا كاملاً . على سبيل المثال ، يُجري مكتب الإحصاء الأمريكي تعدادًا اقتصاديًا للشركات الأمريكية المُصنفة إلى قطاعات ، مثل العقارات أو الخدمات المهنية ، ومُجمَّعة في وحدات جغرافية ؛ الولايات والمقاطعات والرموز البريدية . تنتج هذه المجموعة الضخمة من البيانات إمكانية البحث في الاقتصادات الإقليمية والمحلية ، ولكنها تقتصر على المؤسسات التي لديها عمال بأجر ، وبالتالي تستبعد العديد من الشركات الصغيرة وجميع العاملين لحسابهم الخاص والعديد من الشركات العائلية . وبالتالي ، لا توفر هذه المجموعة سوى رؤية محدودة حول الاقتصادات والخدمات المحلية.

ما هي المتغيرات التي تقيس فعليًا؟

في حين تبدو البيانات الكمية، خاصةً إذا وزعتها الجهات الحكومية والمهنية، موضوعية ودقيقة وواضحة، فمن المهم فهم كيفية بناء هذه المتغيرات وما الذي قد تقيسه أو لا تقيسه تحديدًا. وأبرز مثال على ذلك هو مفهوم "العرق" المستخدم في تعداد الولايات المتحدة. فقبل تعداد عام 2000، لم يكن بإمكان الناس سوى اختيار فئة عرقية واحدة؛ مما منعهم من تحديد هويتهم بأكثر من عرق واحد. وقد أخفت الإحصاءات الناتجة عن العرق التنوع العرقي للعديد من الأفراد، وبسطت التركيبة العرقية لسكان الولايات المتحدة بشكل مفرط. بالإضافة إلى ذلك، كان للعدد المحدود من الفئات العرقية التي استخدمها التعداد قوة تأديبية، إذ أجبرت الناس على تعريف أنفسهم والآخرين بهذه المصطلحات - وهي عملية قوية عملت لقرون على بناء والحفاظ على الطبقات والتسلسلات الهرمية الأخرى القائمة على فئات عرقية معينة. قد يُسبب تحليل صحة الأفراد مقارنةً بمتغيرات اجتماعية واقتصادية أو صحية عامة أخرى مشاكل أيضًا. غالبًا ما تُبلغ الحالة الصحية ذاتيًا، وبالتالي، فهي مقياس ذاتي للغاية يعتمد على كيفية فهم المستجيبين لمعنى كونهم يتمتعون بصحة جيدة أو سيئة. ومع ذلك، يُستخدم هذا المقياس بالتزامن مع متغيرات أخرى أقل ذاتية نظرًا لعدم الإبلاغ عنها ذاتيًا.

الفئات

يُعد غموض فئات البيانات، بالإضافة إلى غموض المتغيرات، اعتبارًا مهمًا آخر. يُمثل البحث الذي أجرته إحدى طالبات الدراسات العليا لدينا حول الاقتصادات المتنوعة للمجتمعات العربية الأمريكية في شمال شرق الولايات المتحدة مثالًا على هذه المشكلة. لتحديد هذه المجتمعات، استخدمت بيانات التعداد السكاني، وهو أحد المصادر الوحيدة لهذه المعلومات. ومع ذلك، لا يُعرّف التعداد السكاني العرب الأمريكيين مباشرةً باستخدام متغير واحد. تُصنّف الفئات العرقية سكان الشرق الأوسط ضمن فئة "البيض". ومع ذلك، يُمكن تمييز السكان العرب، بطرق أخرى باستخدام بيانات التعداد السكاني. على سبيل المثال، يُمكن تعريف العرب بأنهم أولئك الذين يتحدثون اللغة العربية، أو أولئك القادمون من بلد ذي أغلبية عربية (أصل قومي)، أو أولئك الذين يُعلنون أنهم من أصول عربية. تُقدم هذه التعريفات ثلاث طرق متداخلة، وإن كانت غير متطابقة، لإحصاء الأمريكيين العرب.

يُرجّح أن يتحدث المهاجرون الجدد نسبيًا اللغة العربية، ويشمل الأصل القومي مجموعات غير عربية، ويُعدّ الأصل فئة غامضة في حد ذاته. باستخدام هذه التعريفات، أنتج الطالب خرائط تُظهر ثلاثة توزيعات مختلفة، وإن كانت متداخلة، للسكان الأمريكيين العرب، مما يؤكد على أن حتى مجموعات البيانات الشاملة، مثل تعداد الولايات المتحدة، لا تُقدم أحيانًا سوى تمثيل جزئي لإجمالي السكان.

الخصوصية

يتطلب التناقض بين الحاجة إلى بيانات مفصلة والحاجة إلى حماية خصوصية الأفراد أحيانًا من الباحثين اتخاذ قرارات مهمة بشأن كيفية إجراء أبحاثهم أو عدم إجرائها. على سبيل المثال، تحتوي بعض مجموعات البيانات على معلومات حساسة جُمعت على مستوى الأفراد أو الأسر. من ناحية، قد تكون هذه التفاصيل ضرورية للتحليل. من ناحية أخرى، قد يؤدي استخدامها في البحث إلى الكشف عن معلومات شخصية للفرد. لتجنب هذا الانتهاك للخصوصية، عادةً ما تُجمع البيانات في وحدات مكانية كبيرة نسبيًا،

مما يؤدي بالضرورة إلى فقدان المعلومات والدقة . ينطبق هذا على العديد من البيانات المتعلقة بالصحة ، وكذلك على عينات البيانات الدقيقة للاستخدام العام (PUMS) من تعداد سكان الولايات المتحدة . حتى في حال توفر بيانات عالية الدقة واستخدامها مقبول ، فإنها قد تُشكل مشكلة . على سبيل المثال ، ينبغي توخي الحذر عند رسم خرائط للمعلومات التفصيلية حول الدخل ، ومستويات التعليم ، والعرق ، ومعدلات الجريمة ، وما إلى ذلك ، إذ قد يؤدي ذلك إلى وصم أشخاص وأماكن معينة ، مع ما يترتب على ذلك من آثار على رفاههم الاقتصادي والاجتماعي . في هذه الحالة ، قد يقرر الباحثون رسم خرائط لنتائج التحليلية على مقاييس أصغر ، أي باستخدام وحدات مكانية أكبر من مقياس التحليل الفعلي (كروملي وماكلافرتي 2002).

الأخطاء والدقة

وأخيرًا ، تحتوي جميع مجموعات البيانات الثانوية على أخطاء . حتى المسوحات التي تُجرى بشكل احترافي ، بما في ذلك استطلاعات الرأي العام ، قد تنطوي على مشاكل في أخذ العينات غير معروفة ، وقد تُشوّه تمثيل السكان المعنيين . على سبيل المثال ، يُقلّل مكتب الإحصاء الأمريكي باستمرار من عدد ملايين المهاجرين غير الشرعيين بشكل رئيسي ، بالإضافة إلى أولئك الموجودين في عناوين غير مدرجة في قاعدة بيانات التعداد . قد تُشوّه بعض هذه الأخطاء بشكل منهجي السكان الذين تُمثّلهم ، مما يُسهم في اتخاذ قرارات سياسية غير دقيقة . على سبيل المثال ، يؤدي نقص عدد المهاجرين الذين يُلبّون الطلب على العمالة الرخيصة في التعداد السكاني إلى إضعاف هذه الفئات العاملة سياسيًا . بالإضافة إلى ذلك ، عندما يشمل نقص العدد الأسر التي لديها أطفال ، فقد يُقلّل من تقدير الطلب على المدارس والخدمات الأخرى.

وهناك أيضًا أخطاء عشوائية قد لا تُشوّه المتوسطات الإجمالية ، ولكنها تُقلّل من جودة البيانات وقدرة الباحث على العمل بدقة أكبر . على وجه الخصوص ، تُعد أخطاء البيانات الثانوية في مرحلة إدخال البيانات ، بما في ذلك الأخطاء المطبعية في معلومات السمات أو الأخطاء في الطبقة المكانية ، شائعة جدًا . على سبيل المثال ، قد تؤدي التناقضات في عناوين الشوارع إلى حذف العديد من السجلات (تصل في بعض الحالات إلى 40%) التي لا يمكن مطابقتها مع قاعدة بيانات العناوين عبر عملية تُعرف باسم الترميز الجغرافي . وبالمثل ، فإن بيانات سجل رحلات الصيد الفردية التي تجمعها الخدمة الوطنية لمصائد الأسماك البحرية (NMFS) مليئة بالأخطاء . في حين أن العديد من الأخطاء ناتجة عن سوء إدخال البيانات (حيث استعانت الخدمة الوطنية لمصائد الأسماك البحرية بشركات استخدمت عمال السجون لإدخال البيانات من نماذج كانت الإدخالات فيها مكتوبة بخط اليد من قبل صيادي الأسماك في البحر) ، فإن أخطاء أخرى ناتجة عن تقديم صيادي الأسماك تقارير خاطئة متعمدة .

بالإضافة إلى ذلك ، قد تحدث أخطاء عشوائية بسبب أعطال تكنولوجية مثل مشاكل معايرة الأجهزة التي تقلل من جودة صور الأقمار الصناعية . تُعد أخطاء الموقع مهمة بشكل خاص في البيانات ذات المرجع الجغرافي . يمكن أن تؤدي هذه البيانات إلى استنتاجات خاطئة بشأن التداخل المكاني للظواهر المعنية . على سبيل المثال ، قد تُعرّف الأماكن خطأً بظاهرة اجتماعية سلبية ، كما نوقش بالإشارة إلى الخصوصية ، أو قد يكون تعرضها للمخاطر الصناعية ، كما هو مُقاس في قاعدة بيانات جرد انبعاثات المواد السامة (TRI) ، غير مُمثل تمثيلاً كافياً (سكوت وآخرون، 1997). ومن المشاكل المهمة الأخرى المغالطة البيئية ومشكلة الوحدة المساحية القابلة للتعديل (ينظر الملحق 11.1).

التعامل مع البيانات الثانوية

في حين أنه يجب دراسة مزايا وعيوب البيانات الثانوية بعناية ، إلا أنها تُسهم بوضوح في توسيع نطاق وقوة الأشكال القياسية للتحليلات ، مثل الاستعلام (طرح أسئلة على قاعدة البيانات واسترجاع البيانات التي تُجيب عن هذه الأسئلة) أو التحليل الإحصائي . تُناقش هذه الأشكال القياسية لتحليل قواعد البيانات بالتفصيل في مجموعة متنوعة من النصوص التمهيدية ، ولن نستعرضها هنا (ينظر شيخار وتشاولا 2003). بل سنناقش بإيجاز ثلاث استراتيجيات لاستخدام البيانات الثانوية بشكل إبداعي ، بهدف اقتراح إمكانية استخدام البيانات الثانوية بطرق تُكمل التحليلات الإبداعية والنقدية في الجغرافيا . تشمل الأمثلة الثلاثة التي نقدمها : تحويل البيانات الثانوية وتعديلها لتتوافق بشكل أفضل مع أسئلة البحث الأصلية ، وتصميم مقاييس ومؤشرات جديدة ، واستخدام البيانات الثانوية في نهج "مختلط" يجمع بين التحليل المكاني الكمي ومعلومات المقابلات النوعية . بالإضافة إلى ذلك ، سندرس الفرص التي توفرها مجالات استخراج البيانات والتصور الجغرافي الناشئة .

إعادة تصميم البيانات بما يتناسب مع احتياجاتنا البحثية.

كما ذكرنا سابقاً ، يمكن للفئات والمتغيرات المُضمنة في مجموعات البيانات الثانوية أن تؤثر على استراتيجيات البحث ونتائجه وتُشكلها . لتجنب ذلك ، نحتاج إلى فحص البيانات بدقة ، وإذا لزم الأمر ، تحديثها ومراجعتها و/أو دمجها مع جمع البيانات الأولية . يُعد بحثنا حول الوصول إلى المساحات المفتوحة في مدينة نيويورك مثالاً جيداً . لتحليل بيانات التعداد الاجتماعي والاقتصادي ، حصل الطالب من إدارة الحدائق والترفيه على قاعدة بيانات تُشير إلى مواقع المساحات المفتوحة في المدينة . تضمنت فئات المساحات المفتوحة في هذه القاعدة "المرافق المتاحة للجمهور ذات الأهمية الإقليمية" ولكنها استبعدت المساحات التي تخدم بشكل رئيسي مجتمعات فردية مثل "الملاعب ، وملاعب كرة السلة وكرة اليد ، والحدائق المجتمعية" . ومع ذلك ، تلعب هذه الأخيرة دوراً بالغ الأهمية في ممارسات الترفيه اليومية لسكان نيويورك . وبدون أخذها في الحسبان ، سيكون تحليل الوصول إلى المساحات المفتوحة غير مكتمل .

كان تحديث قاعدة البيانات جزءاً ضرورياً من البحث ، ولكنه يتطلب وقتاً وجهداً كبيرين . حصل الطالب على مجموعات البيانات الإضافية من عدة وكالات عامة ومنظمات غير حكومية ودمجها مع قاعدة البيانات الأصلية ، وتوضح الخريطة الناتجة وجود اختلافات ملحوظة في حسابات الوصول إلى المساحات المفتوحة بين قاعدة البيانات الأصلية وقاعدة البيانات المُحدثة (ينظر الشكل 11.1). في العديد من المناطق ، يتجاوز الفرق 0.5 فدان . يُمثل هذا تناقضاً كبيراً بالنظر إلى معايير المدينة لتعريف المناطق التي تعاني من نقص حاد في الخدمات (1.5 فدان لكل 1000 نسمة أو أقل). وقد ثبت أن تحديث قاعدة البيانات كان حاسماً أيضاً للحصول على إحدى النتائج الرئيسية للدراسة : ألا وهي أن الوصول إلى المساحات المفتوحة يختلف بشكل مختلف فيما يتعلق بالدخل ووضع الأقليات اعتماداً على حجم المساحات المفتوحة المتاحة على مسافة قريبة سيراً على الأقدام .

صمم مقاييسك التحليلية الخاصة.

عند العمل مع البيانات الثانوية ، يُساعد التفكير الإبداعي خلال المرحلة التحليلية على تجنب الاختصار على المقاييس الإحصائية القياسية . في الواقع ، يُمكن التلاعب بمجموعات البيانات الثانوية لإنتاج متغيرات ومقاييس جديدة تؤدي إلى نتائج مُلهمة . ومرة أخرى ، يُقدم بحثنا حول الوصول إلى المساحات المفتوحة في

نيويورك مثلاً جيداً . هنا ، صُمم مقياس جديد للتغلب على مشكلة الوحدة المساحية القابلة للتعديل ، والتي نتجت عن قياس الأنماط البشرية اليومية ، مثل استخدام الحدائق الحضرية ، استناداً إلى كيانات مكانية سياسية عشوائية مثل منطقة مجلس المجتمع (CBD) . تقليدياً، يُقاس الوصول بنسبة بسيطة على مستوى مناطق مجلس المجتمع ، ويُعبر عنه إما كنسبة مئوية من المساحة المفتوحة أو بمساحة فدان من المساحة المفتوحة لكل 1000 ساكن في كل وحدة.

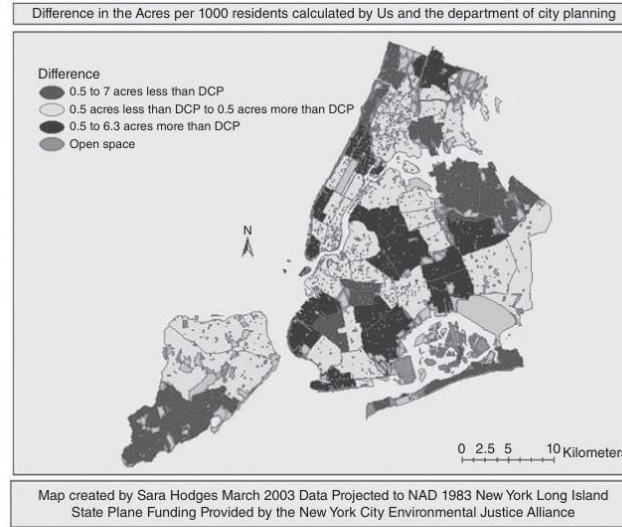


Figure 11.1 Differences in the amount of open space in New York City (acres per 1,000 residents); calculated by researchers and the Department of City Planning
Source: Sara Hodges, 2004. MA thesis "Open space in New York City: A GIS-based analysis of equity of distribution and access" Hunter College, New York. Reprinted with permission. See http://www.geo.hunter.cuny.edu/~mpavlov/Articles/Ch11_Figures_H1_res.pdf for full color maps.

ومع ذلك ، تُعد مناطق مجلس المجتمع كيانات كبيرة نسبياً ، وللترفيه اليومي ، لا يستخدم سكان نيويورك سوى المساحات المفتوحة التي تقع على مسافة قريبة سيراً على الأقدام . بالإضافة إلى ذلك ، يرتبط هذا المقياس القياسي للوصول إلى المساحات المفتوحة ارتباطاً واضحاً بحدود إدارية عشوائية (منطقة الأعمال المركزية) ، على الرغم من أن الناس يتجاهلون هذه الحدود ويذهبون ببساطة إلى أقرب حديقة أو ملعب . أردنا قياس الوصول إلى المساحات المفتوحة مع مراعاة كيفية استخدام الناس للمساحات المفتوحة في حياتهم اليومية . أشارت الدراسات إلى أن الأطفال يستخدمون الحدائق في نطاق ربع ميل ، بينما يمشي البالغون لمسافة تصل إلى نصف ميل .

لتجنب تأثير حدود منطقة الأعمال المركزية ، حوّل الطالب خريطة المساحات المفتوحة إلى تنسيق نقطي بحجم خلية يبلغ 40 قدماً . يتطابق هذا الدقة تقريباً مع حجم قطعة أرض ضريبية ، وبالتالي يمكن أن يأخذ في الحسبان حتى أصغر المساحات المفتوحة ، مثل الحدائق المجتمعية . بالإضافة إلى ذلك ، يقارب حجم الخلية هذا أيضاً حجم مسكن معين يمكن من خلاله قياس الوصول إلى المساحات المفتوحة . ثم حسبنا لكل بكسل (ممثّل للمباني السكنية) مجموعاً بالأقدنة من المساحات المفتوحة ضمن مسافة سير على الأقدام لنصف قطر ربع ميل ونصف ميل . بدلاً من قيمة واحدة لكل منطقة كبيرة ، أنشأنا سطحاً يعكس اختلافات أدق بكثير في إمكانية الوصول إلى المساحات المفتوحة (ينظر الشكلين 11.2 و 11.3) . لا تُظهر هذه الخرائط إمكانية الوصول إلى المساحات المفتوحة بمصطلحات جديدة باستخدام مقاييس جديدة فحسب ، بل تُظهر أيضاً وجود اختلافات كبيرة (ولكنها مُتجاهلة) في مقدار المساحات المفتوحة المتاحة للأطفال والبالغين في نيويورك.

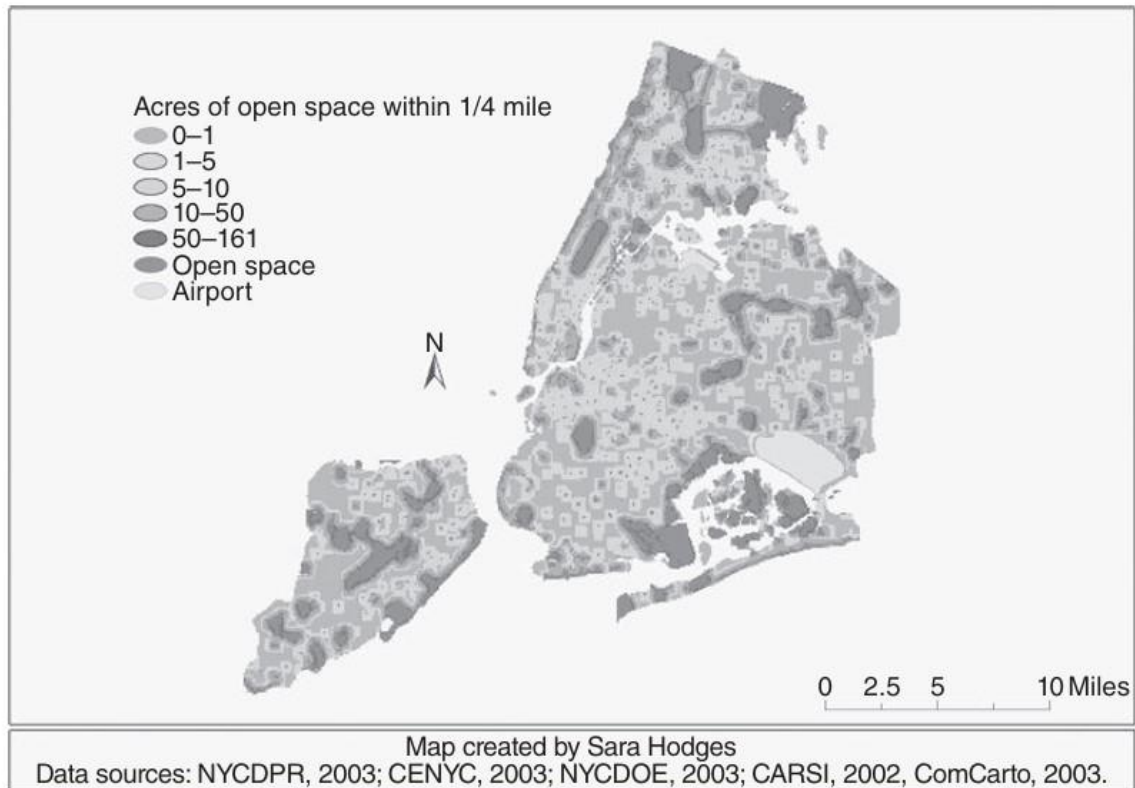


Figure 11.2 Children's access to open space in New York City

Source: Sara Hodges, 2004. MA thesis "Open space in New York City: A GIS-based analysis of equity of distribution and access" Hunter College, New York. Reprinted with permission. See http://www.geo.hunter.cuny.edu/~mpavlov/Articles/Ch11_Figures_H1_res.pdf for full color maps.

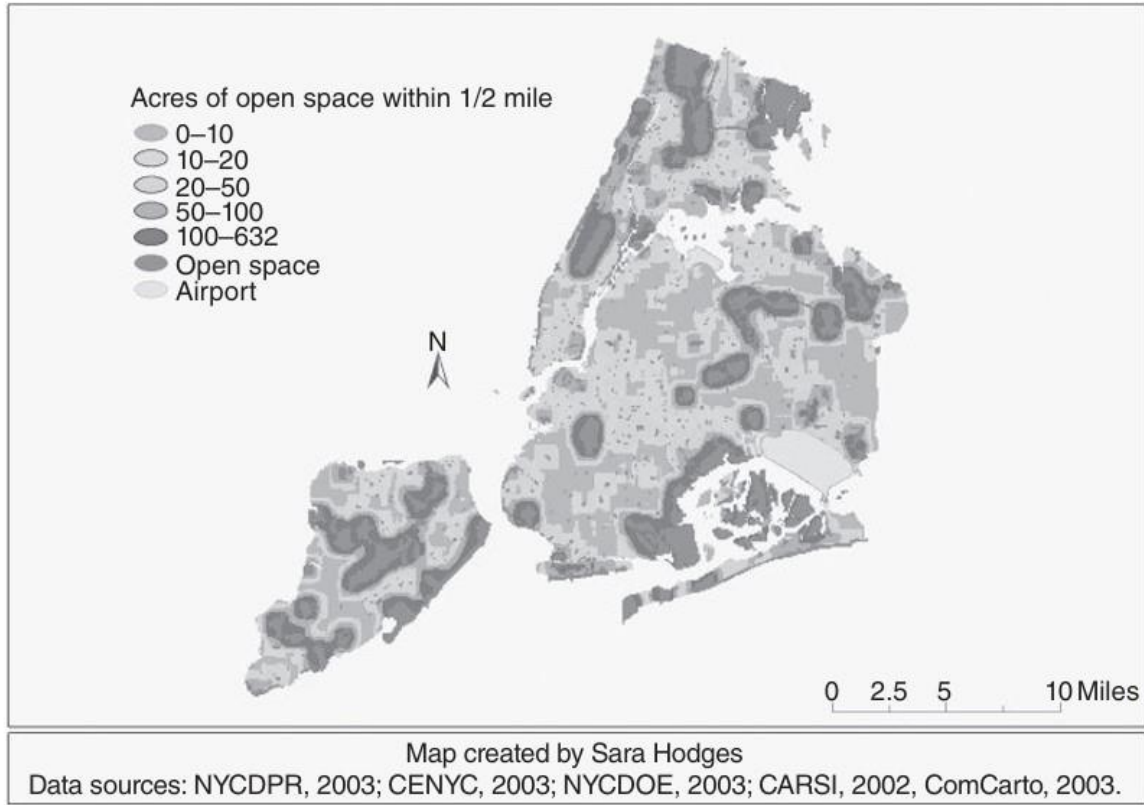


Figure 11.3 Adults' access to open space in New York City

Source: Sara Hodges, 2004. MA thesis "Open space in New York City: A GIS-based analysis of equity of distribution and access" Hunter College, New York. Reprinted with permission. See http://www.geo.hunter.cuny.edu/~mpavlov/Articles/Ch11_Figures_H1_res.pdf for full color maps.

الطرق المختلطة: رسم خرائط للمشهد الاجتماعي لمجتمعات صيد الأسماك

في ما سبق ، تُحوّل البيانات من قاعدة بيانات ثانوية وتُعاد صياغتها إلى متغير جديد عبر مساحة جديدة ، وهي مساحة الوصول إلى المساحات المفتوحة . في حالة مصايد الأسماك ، أنتجنا مقياساً جديداً موزعاً ضمن مساحة جديدة (سانت مارتن وهول أربور، 2008) . يُمثل علم مصايد الأسماك وإدارتها بشكل متكرر وجود مجتمعات الصيادين وصيادي الأسماك في المحيط كجهد صيد إجمالي مُعبّر عنه بكميات الأسماك المصطادة . ورغم فائدته في تقدير مخزون الأسماك المتبقي على مستوى المنطقة أو العائدات المستقبلية ، فإن تجميع الجهد في متغير واحد يُزيل الاختلافات المحلية واعتماد مجتمعات معينة على موارد معينة . وللتعبير عن وجود مجتمعات صيد معينة واعتمادها على مناطق صيد معينة ، فُحصت مجموعات البيانات البيولوجية التابعة للخدمة الوطنية لمصايد الأسماك البحرية (NMFS) لإيجاد طريقة لرسم خريطة للمشهد الاجتماعي لمجتمعات الصيد.

وباستخدام بيانات سجل الصيد ، رُبطت المعلومات الموقعية لكل رحلة صيد بميناء السفينة ، مما كشف عن مكان السفن . من مجتمعات صيد محددة ، دُمجت بيانات عدد أفراد الطاقم وطول الرحلة لتحديد

وقت العمل . لكل رحلة ، ضُرب عدد أفراد الطاقم على متن السفينة بطول الرحلة لإنشاء متغير جديد ، وهو "أيام صيد الأسماك" الذي يمكن ربطه بمواقع أو مناطق صيد محددة . استُخدم هذا المتغير المكاني الجديد لإنشاء خرائط فردية ومركبة لمجموعة متنوعة من المجتمعات تُظهر المناطق التي تعتمد عليها (ينظر الشكل 11.4) .

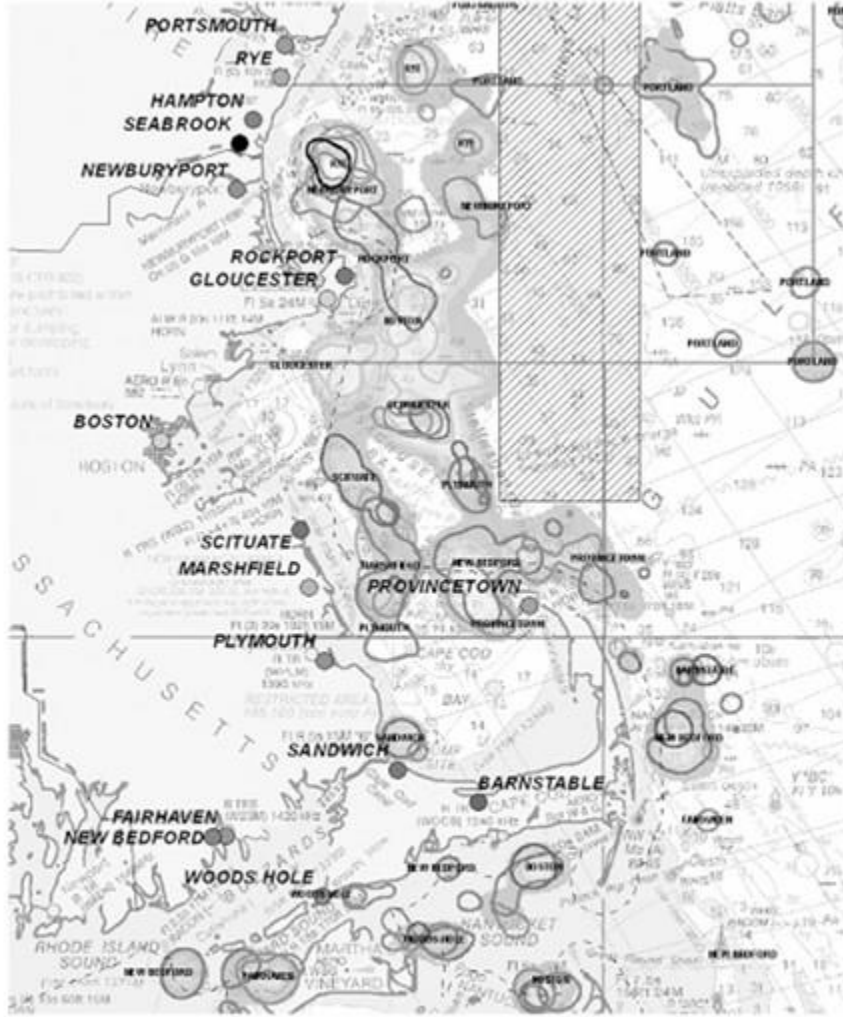


Figure 11.4 An extract from a map depicting the primary fishing grounds (based on labor time) of small trawl vessels from particular communities/ports in New England (color coded outlines in the original correspond to port markers, see <http://fisheries.rutgers.edu> for full color version). The outlines are superimposed upon a NOAA nautical chart. The map also contains a raster density surface based on the aggregate of all vessels.

وكجزء من مشروع بحثي تشاركي ، دُعي صيادو الأسماك من كل مجتمع لتصحيح أو تعديل الخرائط ، وقد أثبتت هذه الخرائط قيمتها في حث المجتمعات المحلية على إجراء تقييمات أكثر محلية لمخزون الأسماك وزيادة مساهمة المجتمع في إدارة المخزون . يُعدّ نهج "المنهج المختلط" المُستخدم في هذا العمل (التحليل الإحصائي وتحليل نظم المعلومات الجغرافية للبيانات الثانوية ، إلى جانب المقابلات التشاركية وورش العمل)

طريقة ناشئة وفعالة للاستفادة من مجموعات البيانات الثانوية ، وتحديد حدودها ، وتوظيف أساليب بديلة لمعالجة تلك الحدود ، والأهم من ذلك ، تعميم فوائد البيانات الثانوية على المجتمعات وعامة الناس.

فرص بحثية جديدة في عالم رقمي

تُتيح قواعد البيانات الثانوية نفسها ، بخصائصها وانتشارها الواسع ، أشكالاً وأساليب تحليل جديدة . بل إنها تُسهّل أشكالاً من إنتاج المعرفة تنفرد بها قواعد البيانات الثانوية . وعلى وجه الخصوص ، استُخدمت أساليب جديدة تتناول خصائص مُحددة لمجموعات البيانات الكبيرة في عدد من المجالات ، بما في ذلك الجغرافيا ونظم المعلومات الجغرافية . ظهرت هذه التقنيات لأول مرة في أبحاث التسويق التي تطلبت تقنيات جديدة لدمج وتحليل المعلومات المتزايدة ، وإن كانت غير مترابطة وغير منهجية ، حول المستهلكين وسلوكهم . و وعد "التنقيب" الإحصائي في قواعد البيانات هذه بالكشف عن أنماط غير معروفة بعد في سلوك المستهلك ، والتي يمكن الاستفادة منها لاحقاً لتحقيق أرباح الشركات .

المهم هو أن هذه التقنيات تعكس النهج التقليدي للبحث ؛ فبدلاً من اختبار الفرضيات ، تهدف خوارزميات التنقيب الجديدة في البيانات إلى الكشف عن أنماط لم تُفترض أو تُلاحظ بعد ، وهي أنماط تظهر بشكل فريد من قواعد البيانات الرقمية الضخمة . واليوم ، يُعد استكشاف البيانات الرقمية اتجاهاً بحثياً رائداً في الجغرافيا ونظم المعلومات الجغرافية . بالإضافة إلى المناهج الإحصائية التي تستخرج البيانات المكانية (شيخار وتشاولا 2003) ، تسمح مناهج التصور الجغرافي الجديدة بالمثل بالتعرف على الأنماط في البيانات الثانوية . يتضمن تحليل البيانات الاستكشافية المكانية عروض بيانات متقدمة تجمع بين الخرائط والرسوم البيانية والجدول ، مما يساعد الباحث على فحص البيانات بصرياً واكتشاف أنماط مكانية جديدة Longley et al. 2005 .

MacEachren 1995 .

الخلاصة

لقد ازدادت كمية وحجم مجموعات البيانات الرقمية العامة والتجارية ، وخاصة تلك التي تحتوي على معلومات مكانية ، بشكل ملحوظ ، وستستمر في الازدياد . تُعد البيانات الثانوية الآن وستظل مهمة للبحث الجغرافي كمصدر أساسي للمعلومات لعدد متزايد من التطبيقات كثيفة البيانات . ومن الواضح أن استخدام هذه البيانات يمنح الباحث مزايا مهمة من حيث تغطية البيانات وجودتها وتكاليفها ، بالإضافة إلى فرصة تحليل الظواهر التي يستحيل تحليلها لولا ذلك ، تحليل البيانات ، مثل توزيع السكان على المستوى الوطني . ومع ذلك ، ينبغي مراعاة القيود المهمة للبيانات الثانوية ، مثل خطر أسئلة البحث القائمة على البيانات ، والتمثيل غير الكامل للظواهر ، وغموض الفئات ، وقضايا الخصوصية .

بالإضافة إلى ذلك ، ينبغي أن يفهم الجغرافيون بوضوح السهولة المحتملة التي يمكن أن تؤدي بها البيانات الثانوية إلى مغالطة بيئية أو خطأ في تقدير القيمة (MAUP) . في حين أن مزايا وعيوب البيانات الثانوية تُعد معطيات مهمة لأي شكل من أشكال التحليل ، فإننا متحمسون لإمكانات التقنيات التحليلية الجديدة والمبتكرة التي تُسهّلها البيانات الثانوية . بالإضافة إلى ذلك ، يجب ألا تقتصر على الغرض الأصلي لأي مجموعة بيانات ؛ ولا ينبغي أن نتردد في معالجة البيانات وتحويلها لبناء متغيرات أو مقاييس أو خرائط جديدة ؛ ولا ينبغي أن نتردد في دمج تحليل البيانات الثانوية مع أساليب أخرى كما هو الحال في البحث بالطرق المختلطة . على الرغم من ارتباط قواعد البيانات الثانوية في كثير من الأحيان بتقنيات التحليل القياسية ، إلا أنه من المفيد عدها مناطق شاسعة يجب استكشافها وتصورها وفهمها باستخدام مناهج نقدية وإبداعية جديدة.

التمرين ١١,١

الاستعلام في قاعدة بيانات TRI الإلكترونية

في متصفح الإنترنت ، افتح موقع جرد انبعاثات المواد السامة (TRI) التابع لوكالة حماية البيئة <http://www.epa.gov/tri/>. يوفر هذا الموقع إمكانية الوصول إلى قاعدة بيانات عامة تتضمن معلومات حول انبعاثات المواد الكيميائية السامة وأنشطة إدارة النفايات الأخرى في بعض الصناعات في الولايات المتحدة . يوفر الموقع أدوات تتيح لك جدولاً (تلخيص) هذه البيانات حسب الوحدات الجغرافية (مثل الولايات) ، والصناعة ، ونوع المواد الكيميائية المنبعثة . كما يتيح لك إعداد تقرير عن حوادث TRI في أحياء محددة مُحددة برموزها البريدية . لمعرفة ما إذا كانت هناك أي انبعاثات في حيّك أو أي آخر في الولايات المتحدة ، أدخل الرمز البريدي المقابل . على سبيل المثال ، يكشف إدخال 80524 (فورت كولينز، كولورادو) أن مصنعاً في هذه المنطقة يُنتج مشروبات الشعير قد أطلق الأمونيا والمركبات العطرية متعددة الحلقات في الهواء عام 2004.

يتميز موقع TRI الإلكتروني بواجهة سهلة الاستخدام تتيح لك الاستعلام ، بطرق قياسية متنوعة ، عن قاعدة بيانات حكومية ضخمة تغطي العديد من الصناعات ، على نطاق وطني ، وعلى مدى سنوات عديدة . لإنشاء استعلامات متعددة السمات ، ينظر أسفل صفحة النتائج حيث يوجد رابط إلى الصفحة الرئيسية لـ TRI Explorer. باستخدام TRI Explorer ، يمكنك البحث على مدار عدة سنوات (1988-2004)، حسب نوع المادة الكيميائية المنبعثة ، والموقع الجغرافي ، ونوع الصناعة . لاستخدام هذه البيانات بطرق أكثر إبداعاً وجغرافياً (على سبيل المثال ، لاستكشاف العلاقة بين الانبعاثات والمناطق الفقيرة أو التي تقطنها الأقليات) ، يمكنك تنزيل نوع وموقع (إحداثيات x و y) انبعاثات المواد السامة واستيرادها إلى نظام المعلومات الجغرافية.

التمرين ١١,٢

رسم الخرائط باستخدام أداة American Fact Finder

من أبرز ميزات أداة American Fact Finder ، وهي قاعدة البيانات التفاعلية التابعة لمكتب تعداد الولايات المتحدة ، قدرتها على رسم خرائط للتعداد السكاني والبيانات ذات الصلة . في هذا التمرين ، سنرسم خريطة لمتغير تعداد واحد (متوسط العمر) بدقة مكانية مختلفة (الولايات والمقاطعات) ، وندرس تأثير المقياس على كيفية عرض البيانات وتفسيرها أي MAUP . افتح الصفحة الرئيسية لأداة American Fact Finder <http://factfinder.census.gov/> في متصفحك ، وانقر على رابط "التعداد العشري" الموجود في الشريط الأيسر أسفل "مجموعات البيانات" . عند فتح الصفحة ، تأكد من تحديد خيار "ملف ملخص تعداد عام ٢٠٠٠ SF ١" .

يحتوي هذا الملف على بيانات تغطي جميع سكان الولايات المتحدة . انقر على "الخرائط المواضيعية" في يمين الشاشة . يمكنك الآن تحديد المقياس الجغرافي الذي ترغب في استخدامه لعرض البيانات . لعرض الولايات المتحدة الأمريكية بالكامل حسب الولاية ، حدد "الدولة" كنوع جغرافي و"الولايات المتحدة" كمنطقة جغرافية . انقر على "التالي" . في الشاشة التالية ، حدد الموضوع PO17 - TM متوسط العمر: 2000 (متغير محدد من ملف الملخص 1). سيؤدي هذا المتغير ، وهو متوسط العمر ، إلى تقسيم السكان إلى نصفين . بمعنى آخر ، نصف السكان أصغر سناً والنصف الآخر أكبر سناً من متوسط العمر - فكلما ارتفع

متوسط العمر، زاد عمر سكان تلك المنطقة . بلغ متوسط العمر في الولايات المتحدة الأمريكية عام 2000 35.3 عامًا .

انقر على زر "إظهار النتيجة" لتحميل خريطة متوسط العمر حسب الولاية . مفتاح الخريطة موجود على اليسار . يشير هذا إلى القيم المضمنة في كل فئة من الفئات الخمس المعروضة بألوان مختلفة . ووفقاً لهذه القيم ، يختلف متوسط العمر بشكل كبير حسب الولاية ، حيث تفصل عشر سنوات بين السكان الأصغر سناً (27.1 عاماً في ولاية يوتا) والسكان الأكبر سناً (38.9 عاماً في غرب فرجينيا و38.7 عاماً في فلوريدا) . بالإضافة إلى ولاية يوتا ، تتمتع ولايات تكساس وكاليفورنيا وأيداهو ولويسيانا وميسيسيبي وجورجيا وألاسكا بمتوسط عمر منخفض نسبياً (يظهر باللون الأصفر الفاتح) . يُفعل زر "i" وظيفة الاستعلام التي يمكنك استخدامها لمعرفة قيمة متوسط العمر لكل ولاية على حدة .

حدد الولايات التي تضم أكبر عدد من السكان سناً . لنر الآن ما إذا كان عرض نفس البيانات حسب المقاطعة يُغير توزيع السكان الأكبر سناً والأصغر سناً . في القائمة المنسدلة "عرض الخريطة حسب" أعلى الخريطة ، اختر "المقاطعة" بدلاً من "الولاية" كوحدة مكانية . عند تحميل الخريطة الجديدة ، انظر إلى مفتاح الخريطة ولاحظ تغير قيمتي متوسط العمر الأدنى والأقصى . على هذا المستوى المكاني ، يتراوح متوسط العمر بين 20 و58.6 عاماً لكل مقاطعة ، مما يُنتج فجوة تقارب 40 عاماً بدلاً من 10 سنوات كما في الخريطة السابقة . مع أن كلا الإحصائيتين حُسبتا من نفس البيانات ، إلا أن بيانات المقاطعة تحتفظ بتباين أكبر من المتوسط على مستوى الولاية . افحص الخريطة وحدد ما إذا كانت الولايات "الأصغر سناً" (مثل يوتا ، تكساس ، كاليفورنيا) تتميز بكثافة سكانية متساوية ؟ هل الولايات "الأقدم سناً" تتميز بكثافة سكانية متجانسة ؟ ما التفسيرات المحتملة لمتوسط العمر وتباينه داخل مختلف أنحاء البلاد ؟ ما هو الاستنتاج الخاطئ الذي قد نتوصل إليه من هذه البيانات والذي قد يشكل مغالطة بيئية واضحة؟

الملحق ١١،١:

المغالطة البيئية ومشكلة الوحدة المساحية القابلة للتعديل

في حين أن القيود المتأصلة في مجموعات البيانات الثانوية تُثير القلق ، فإن العلاقة بين مجموعات البيانات الثانوية وعدد من قضايا التحليل المكاني تُثير القلق أيضاً . على وجه الخصوص ، ندرس البيانات الثانوية وميلها إلى زيادة حدوث المغالطة البيئية ، بالإضافة إلى علاقتها بمشكلة الوحدة المساحية القابلة للتعديل (MAUP) ، وكلاهما يُمثلان مصدر قلق بالغ للجغرافيين.

المغالطة البيئية

تشير المغالطة البيئية إلى افتراض أن جميع الأفراد في مجموعة ما يشتركون في متوسط خصائص تلك المجموعة . في حالة البيانات المكانية ، يجب أن نحرص على عدم افتراض أن جميع الأشخاص المقيمين في منطقة جغرافية معينة ، مثل منطقة إحصائية أو منطقة مدرسية ، لديهم خصائص مطابقة لمتوسط المنطقة ككل . يوضح المثال التالي الذي أعده طالب دراسات عليا هذه المشكلة . كان الهدف من بحث هذا الطالب هو معرفة ما إذا كانت الاختلافات في سلوك إعادة التدوير لدى سكان نيويورك تُحدد باختلاف مواقفهم تجاه إعادة التدوير ومعرفتهم بها . أجاب أفراد من مناطق ذات مستويات منخفضة وعالية من إعادة التدوير عن أسئلة حول مواقفهم تجاهها . ومع ذلك ، لم يُقيّم سلوكهم في إعادة التدوير إلا باستخدام ما يُسمى "معدل التحويل" ، أو النسبة المئوية للنفايات المُحوّلة من التخلّص ، والمُقدّرة لكل منطقة في المدينة .

افترض أن المشاركين في الاستطلاع يُعيدون التدوير بشكل أقل أو أكثر بناءً على متوسط الإحصائيات لمنطقتهم بدلاً من سلوكهم الفعلي ، وهو ما يُمثل حالة من المغالطة البيئية . يتضمن تجنب ذلك سؤال المشاركين مباشرة عن سلوكهم في إعادة التدوير . تزيد قواعد البيانات الثانوية من احتمالية حدوث المغالطة البيئية نظراً لتوافر وفرة من البيانات تُشبه البيانات التي يحتاجها الباحث ، أي سلوك إعادة التدوير في هذه الحالة ، ويمكن الوصول إليها بسهولة عبر وحدات مكانية متعددة . من المشكلات التحليلية الشائعة الأخرى التي يواجهها الجغرافيون مشكلة وحدة المساحة القابلة للتعديل (MAUP) ، والتي تشير إلى تأثير الحدود السياسية على البيانات المكانية وتحليلها . وتُعد هذه الحدود ، على وجه الخصوص ، مفاهيم اجتماعية قد لا تكون لها علاقة تُذكر بالظاهرة قيد الدراسة .

ففي الولايات المتحدة ، على سبيل المثال ، قد يكون تأثير حدود الولايات ، وخاصةً حدود المقاطعات ، على انتشار الأمراض ، أو الفصل السكاني ، أو الهجرة ، محدوداً للغاية ، ومع ذلك ، تُجمع البيانات وتُحلل وتُرسَم خرائطها بشكل متكرر باستخدام هذه الحدود . بمعنى آخر ، غالباً ما نحدد أنماطاً في البيانات بناءً على حدود لا علاقة لها بالظاهرة قيد الدراسة . ومن المهم مراعاة جانبين آخرين من مشكلة وحدة المساحة القابلة للتعديل (Wong 2004) . أولاً ، تتغير حدود الوحدات التي تُجمع بياناتها بمرور الوقت ، مما يجعل من الصعب ، أو من المستحيل ، مقارنة مجموعات البيانات التي تصف نفس المنطقة ولكن في فترات زمنية مختلفة . تُمثل التغييرات الجذرية في الحدود الإدارية في موسكو (التي نوقشت سابقاً) ، والتي تُشكل أساس تنظيم البيانات الاجتماعية والاقتصادية ، حالةً متطرفةً من تباين استخدام البيانات (MAUP) .

ثانياً ، يُمكن أن يؤثر نطاق عرض البيانات وتحليلها على النتائج . على سبيل المثال ، قد يُسفر تحليل بيانات التعداد نفسها على مقاييس مختلفة (مستوى كتل التعداد ، أو مجموعات كتل التعداد ، أو مناطق التعداد) عن إحصاءات وأنماط مكانية مختلفة (ينظر التمرين 11.2) . يُعد الوعي بتأثير اختيار مقياس مكاني أو آخر أمراً بالغ الأهمية . في بعض الحالات ، يُعالج اختيار مقياس واحد للتحليل المشكلة المطروحة بدقة ، بينما في حالات أخرى ، يكون التحليل على مقاييس متعددة ضرورياً لالتقاط تلك العمليات التي تتجلى بشكل مختلف على مقاييس مختلفة . على سبيل المثال ، لا تظهر الأدلة الجغرافية على الروابط بين هياكل السيطرة السياسية والاقتصادية في الحقبة السوفيتية (الوزارات الاقتصادية ومقرات كومسومول) والتطور الرأسمالي اللاحق (الخدمات المصرفية الخاصة الجديدة والشركات المالية) إلا عند الدقة المكانية الأولية لعناوين الشوارع . فقط على هذا النطاق ، يُمكن رؤية تركيز الشركات الجديدة داخل مواقع ومكاتب هياكل السلطة في الحقبة السوفيتية نفسها ، أما عند الدقة الأقل ، فلا يظهر هذا التطابق المكاني .

تُظهر دراسة حول الوصول إلى المساحات المفتوحة في مدينة نيويورك ، أجراها طالب دراسات عليا ، ضرورة اتباع نهج متعدد المقاييس . حُسبت نسب المساحات المفتوحة التي تقيس الوصول إلى المساحات المفتوحة (ينظر أدناه لمزيد من التفاصيل) وارتباطاتها بالمتغيرات الاجتماعية والاقتصادية على ثلاثة مستويات : مستوى منطقة مجلس المدينة ، ومنطقة التعداد السكاني ، والحي (مقاسة كمساحة مفتوحة على مسافة قريبة سيراً على الأقدام) . في حين ارتبط عدد من المتغيرات الاجتماعية والاقتصادية ارتباطاً وثيقاً بالمساحات المفتوحة على نطاق منطقة الأعمال المركزية (إيجابياً في حالة متوسط دخل الأسرة وسلبياً في حالة نسبة الأشخاص الملونين) ، إلا أنه لم يكن من الممكن استخدام نفس المتغيرات للتنبؤ بإمكانية الوصول إلى المساحات المفتوحة على مقاييس مكانية أدق .

على هذه المقاييس ، كانت الارتباطات أكثر تعقيداً . على سبيل المثال ، على مستوى الحي ، كان لدى كل من الأحياء الغنية والفقيرة إمكانية الوصول إلى المساحات المفتوحة ، ولكن في الأحياء الغنية كانت

هناك مساحات مفتوحة كبيرة (حدائق حضرية) ، بينما لم يكن لدى الأحياء الفقيرة إمكانية الوصول إلا إلى مساحات مفتوحة صغيرة جدًا.