

2000年合衆国センサスデータとその抽出ツールの開発 およびMANDARAでの地図化

谷 謙二 (埼玉大学)

I はじめに

日本では、統計データおよび地図データがインターネットで公開されることにより、GISを用いた地域統計の地図化が容易になりつつある。こうした傾向は海外でも同様と考えられるが、データフォーマット・統計制度等の情報が国内に比べて入手しにくいことなどにより、海外諸国の詳細な統計データの使用は当該地域に詳しい一部の専門家に限られている。しかし統計データのフォーマットを理解し、地図データを入手すれば、海外のデータであっても日本にいながらにして地図化し、分析を行うことができるようになる。

本研究では、こうした試みの一つとして、詳細な統計データファイルと地図データを無償で提供しており、また需要も多いと考えられる、アメリカ合衆国の2000年センサスから、ユーザーの必要なデータを抽出するツールを開発し、筆者が開発しているGISのMANDARAを使って地図化する方法を紹介する。

II 2000年合衆国センサスとCTPP2000

1. 2000年合衆国センサスの概要

合衆国センサスは1790年に開始され、10年ごとに実施されている。2000年合衆国センサスでは、short formと呼ばれる全数調査と、long formと呼ばれる1/6 サンプルの調査票を併用して行われている。short formでは、世帯に関する項目とし

て居住者数、住宅所有の2項目、個人に関する項目として氏名、電話番号、性別、年齢と生年月日、ヒスパニックかどうか、人種の6項目を尋ねているだけで、かなりシンプルである。図1にshort formの調査票を示した。

一方でlong formでは日本の国勢調査に比べて非常に詳細な調査が行われている。表1はlong formの調査票の項目を示したものであるが、世帯単位では住宅所有に加え、建築時期や部屋数、住宅設備、公共料金支出なども含んでいる。これらの項目は、日本の住宅・土地統計調査におおむね該当している。個人に関しては、人種、婚姻状態、学歴、使用言語、障害の有無、兵役、就業地、通勤手段、通勤時間、就業状況、産業、職業、収入など詳細な項目となっている。特に収入に関しては給与、年金などその内訳の記載も要求されており、日本の国勢調査と比較すると記入者にとっては負担が大きい¹⁾。

2. Web GISを使ってのセンサスデータの地図化

2000年合衆国センサスの内容は、センサス局ホームページ(<http://www.census.gov/>)を通じて公開されている。その中でWeb GIS機能を持つ "American FactFinder²⁾" を用いれば、調査項目や空間集計単位を指定してブラウザ上に主題図を描くことができる。American FactFinderでは、2000年センサスだけでなく、それ以降の調査についても地図化することができる。

Web GISとしての機能は、拡大・縮小、道路やストリートのオーバーレイ、階級区分の方法や色

1) このような詳細な項目を含むlong formであるが、収入や障害の有無などセンシティブな項目も含んでいるため、short formに比べて回答率が低く、センサスの精度を下げる要因となっており、2010年センサスからはshort formだけで実施することになっている (森, 2007)。

2) American FactFinderについては平井(2006)でも解説されている。



This is the official form for all the people at this address. It is quick and easy, and your answers are protected by law. Complete the Census and help your community get what it needs — today and in the future!

Start Here

Please use a black or blue pen.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2000?

Number of people

INCLUDE in this number:

- foster children, roomers, or housemates
- people staying here on April 1, 2000 who have no other permanent place to stay
- people living here most of the time while working, even if they have another place to live

DO NOT INCLUDE in this number:

- college students living away while attending college
- people in a correctional facility, nursing home, or mental hospital on April 1, 2000
- Armed Forces personnel living somewhere else
- people who live or stay at another place most of the time

2. Is this house, apartment, or mobile home — Mark ONE box.

- Owned by you or someone in this household with a mortgage or loan?
- Owned by you or someone in this household free and clear (without a mortgage or loan)?
- Rented for cash rent?
- Occupied without payment of cash rent?

3. Please answer the following questions for each person living in this house, apartment, or mobile home. Start with the name of one of the people living here who owns, is buying, or rents this house, apartment, or mobile home. If there is no such person, start with any adult living or staying here. We will refer to this person as Person 1.

What is this person's name? Print name below.

Last Name

First Name MI

4. What is Person 1's telephone number? We may call this person if we don't understand an answer.

Area Code + Number - -

5. What is Person 1's sex? Mark ONE box.

Male Female

6. What is Person 1's age and what is Person 1's date of birth?

Age on April 1, 2000

Print numbers in boxes.

Month Day Year of birth

→ NOTE: Please answer BOTH Questions 7 and 8.

7. Is Person 1 Spanish/Hispanic/Latino? Mark the "No" box if **not** Spanish/Hispanic/Latino.

- No, not Spanish/Hispanic/Latino Yes, Puerto Rican
- Yes, Mexican, Mexican Am., Chicano Yes, Cuban
- Yes, other Spanish/Hispanic/Latino — Print group. ↗

8. What is Person 1's race? Mark one or more races to indicate what this person considers himself/herself to be.

- White
- Black, African Am., or Negro
- American Indian or Alaska Native — Print name of enrolled or principal tribe. ↗

- Asian Indian Japanese Native Hawaiian
- Chinese Korean Guamanian or Chamorro
- Filipino Vietnamese Samoan
- Other Asian — Print race. ↗ Other Pacific Islander — Print race. ↗

- Some other race — Print race. ↗

→ If more people live here, continue with Person 2.

図 1 2000年合衆国センサスにおけるshort form調査票

出典 : Summary File 1 Technical Documentation

表1 2000年合衆国センサス long formでの調査項目

番号	居住者一覧	
1	住宅の居住者数	
2	1の人々の氏名	
番号	Person 1	
1	Person1の氏名	28 職業
2	電話番号	・仕事の種類
3	性別	・最も重要な活動
4	2000年4月1日現在の年齢と生年月日	29 雇用主
5	スパニッシュ/ヒスパニック/ラティーノか?	去年の仕事
6	人種	30 1999年に何週間仕事をしたか
7	婚姻状態	・1週間あたりの就業時間
8	2000年2月1日の通学状況	1999年の収入額
9	通学先の学年	・給与
10	最終学歴	・自営収入
11	祖先	31 30 利息、地代
11	家庭で英語以外を話すか?	・社会保障
11	その言語は何か?	・補助的社会保障
11	英語をうまく話すか?	・公的補助、福祉
12	出生地	・年金
13	合衆国の国民か?	・その他
14	合衆国での居住開始年	32 1999年の全収入
15	5年前の居住地	
16	以下の継続的な状態にあるか?	33-53は世帯について回答
16	・盲目、難聴	33 住宅所有
16	・歩行困難	34 住宅の建て方
17	肉体的、精神的状況により、過去6ヶ月以上次の活動は困難か?	35 建築時期
17	・学習、記憶、集中	36 入居時期
17	・着替え、入浴、室内歩行	37 部屋数
17	・買い物や通院	38 寝室数
17	・就業	39 水道、水洗トイレ、バスルームの完全な配管設備の有無
18	15歳未満か? その場合33へ	40 シンク・レンジ・冷蔵庫の完全なキッチン設備の有無
18	18歳未満の孫との同居	41 電話の有無
19	・この祖父母は孫の基本的ニーズに責任を負っているか?	42 暖房燃料の種類
19	・それはどのくらいの期間か	43 家族が利用する自動車台数
20	兵役に就いたことがあるか	44 One Family HouseまたはMobile Homeのみ回答
20	・その時期	・ビジネス(店舗や理容院)や診療所の有無
20	・全体での兵役期間	・住宅の広さ
21	先週の就業。就業していない場合25へ	・この不動産からの農業生産額
21	就業地	住宅施設の公共料金
22	・住所	45 電気代
22	・市、町、郵便局名	・ガス代
22	・この就業地はその市・町の内部か	・上下水代
22	・カウンティ名	・灯油代
22	・州または国名	46 住宅に賃料を払っている場合に回答
22	・ZIPコード	・毎月の賃料
23	通勤手段	・賃料には食事代を含むか
23	・自動車の同乗者数	47-53はこの住宅を所有している場合に回答
24	職場への出発時間	47 この資産への抵当の有無
24	・片道通勤時間	・毎月の抵当払い額
25-26	25-26は先週仕事をしていなかった者のみ	・ここでは不動産を含むか
25	・レイオフ中か?	・火災、危険、洪水保険料を含むか
25	・一時的に仕事を離れていたか?	48 第二抵当を持つか
25	・6ヶ月以内に職場に復帰すると告知されているか?	・その毎月の支払い額
25	・過去4週間で職を探したか?	49 昨年の不動産税額
25	・先週もし仕事のオファーがあったり呼び戻されていたら仕事を始めていたか?	50 この資産に対する火災、危険、洪水保険の年間支払額
26	最後に仕事をした時期	51 この資産の価格(この資産を売りに出すとしたらいくらと考えるか)
27	産業・雇用	52 コンドミニアムの場合の毎月のcondominium fee
27	・会社名	Mobile Homeの場合のみ回答
27	・ビジネスの名称	53 このMobile Homeは分割ローンか
27	・産業分類	・このMobile Houseのための分割ローンや不動産税、地代、登録料などの全体の費用

資料 : Summary File 4 Technical Documentation

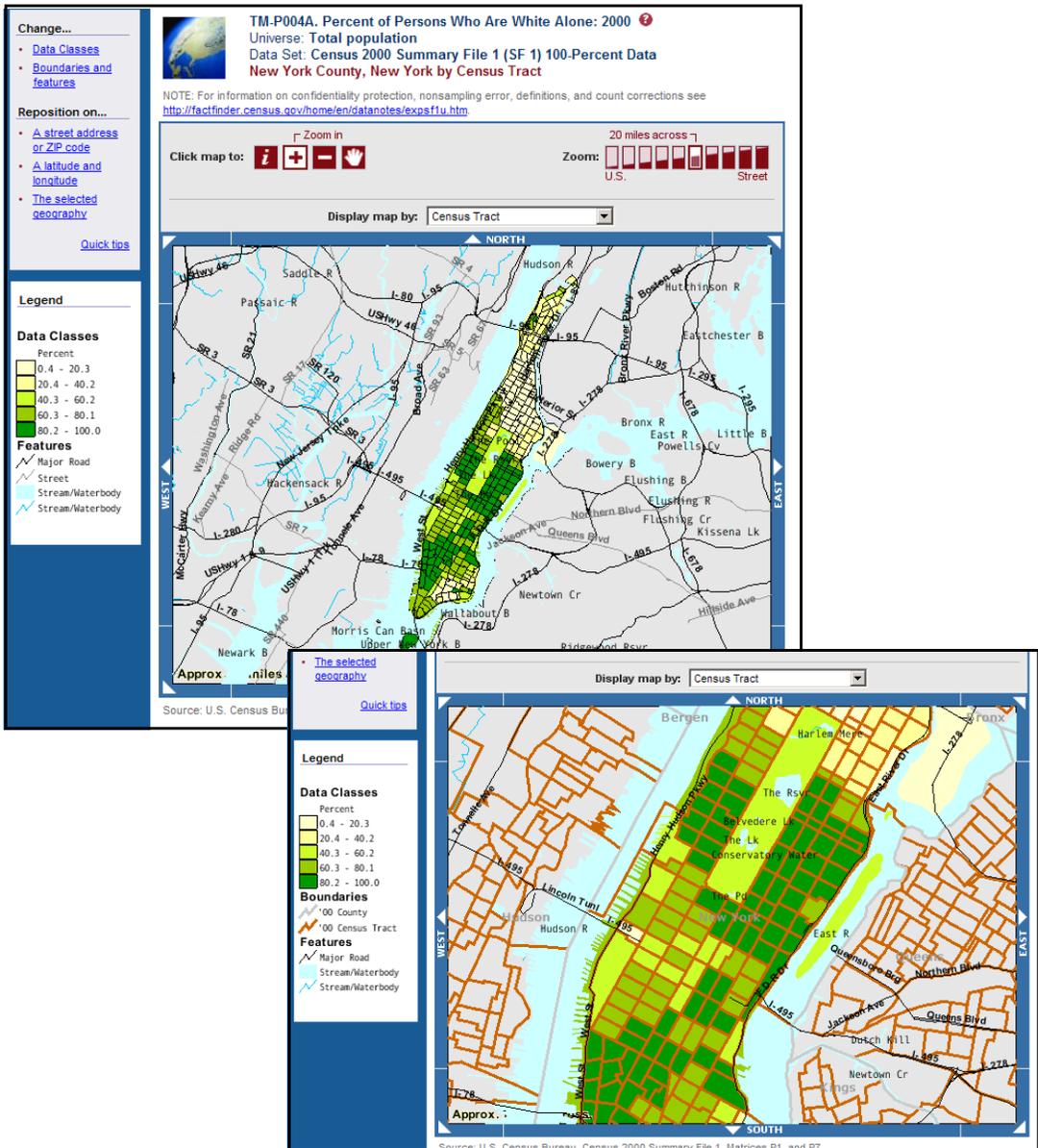


図2 American FactFinderを用いて描いたニューヨーク市・ニューヨークカウティにおける白人居住者割合

の変更などが含まれている。図2は、この機能を利用してニューヨーク市のニューヨークカウティ内のセンサストラクトごとの白人割合を示したものである。このようなWeb GISは、不特定多数の人々に対して空間的情報を安価に伝達するため

には有効な方法である。しかしながら、描画した主題図を加工することが困難であったり、独自に加工したデータを地図化できない、階級区分図しか描けない、操作性が十分でない、表示までに時間がかかる、といった点で、専門家が利用するに

は問題点もある。

3. 公開されているデータファイル

日本の国勢調査のデータは、総務省統計局ホームページにおいてExcel形式のファイルとして公開されており、それは都道府県あるいは市区町村単位である。電子媒体のExcel形式で提供されているにもかかわらず、紙媒体と同様のフォーマットでとなっているため、コンピュータで集計するには不便なものとなっている。Excelでの加工が行われていないデータや、市町村より詳細な町丁・字等別集計やメッシュデータについては、統計情報研究開発センターより販売されているもの、地図データも含め、かなり高額である。

一方2000年合衆国センサスのデータは、American FactFinderで閲覧できるだけでなく、データファイルとしても無償でダウンロード可能となっている。そこには、さまざまな空間集計単位のデータが含まれており、さらにそれぞれの空間集計単位に対応した境界線データもホームページで公開されている。こうしたデータファイルを直接操作することで、任意の集計単位で広範囲のデータを抽出することが可能となり、またExcel上でのデータの加工や、GISでの地図化も容易になる。

データファイルはSummary Filesと呼ばれる次の4つに分けられる。

Summary File 1: short formに関係する286の表が含まれている。

Summary File 2: 249の人口集団ごとにshort formに関係する47の表が含まれている。

Summary File 3: long formに関係する813の表が含まれている。

Summary File 4: 336の人口集団ごとに、long formに関係する323の表が含まれている。

Summary Fileとは、一つのファイルではなく、1～4それぞれに国全体と州ごとのファイルが存在する。カウンティ単位の空間スケールであれば、国全体のファイルであるFinal NationalまたはAdvanced Nationalに含まれており、それよりも詳細なセンサストラクト等の集計単位は州ごとのフ

ァイルに含まれている。空間集計単位の詳細については次節で説明する。

また、国全体、州単位のSummary Fileであっても、データの量が膨大であるので、ダウンロードするファイルはさらに細分化されている。たとえばそれぞれの州のSummary File 1のファイルは39個に分割されている。しかし必要なデータがどのファイルに含まれているかを知ることはかなり手間のかかる作業である。

また、一つのファイルには一つの州内のあらゆる集計単位のデータが含まれている。さらにその集計単位は州ごと・Summary Fileごとに論理番号で管理され、論理番号に対応する実体は別にGeographical Header(GeoHeader)ファイルで管理されている。

また、Summary File 2と4は、それぞれSummary File 1と3に対応した、人種や民族ごとに分割された詳細なデータであり、ファイルの数も膨大である。

このような複雑なデータファイルから、必要なデータを必要な集計単位で抽出することは、表計算ソフトの機能だけでは困難で、データベースソフトを用いる必要がある。しかしより簡便な抽出ツールを開発すれば、多くの人がデータを利用することが可能になるだろう。

4. 合衆国センサスの空間集計単位

センサスデータはカウンティ、センサストラクトなど、多様な空間単位で集計されており、Summary File 内では全ての集計単位ごとに数値が表章されている。また集計単位に対応した境界線データもシェープファイルまたはExport形式ファイルとして、ホームページから無償でダウンロードできる（ただし最小単位のBlockレベルの境界線データは公開されていない）。

図3は集計単位の階層関係を示したものであり、線で結ばれている集計単位間には階層関係が存在する。最も中心的な集計単位は、州からカウンティ、センサストラクト、ブロックグループ、ブロックと連なるもので、下位階層の集計単位の境界線は、上位階層の集計単位の内部に必ず収まる。それぞれの集計単位には、FIPS(Federal Information Processing Standards)コードが割り当

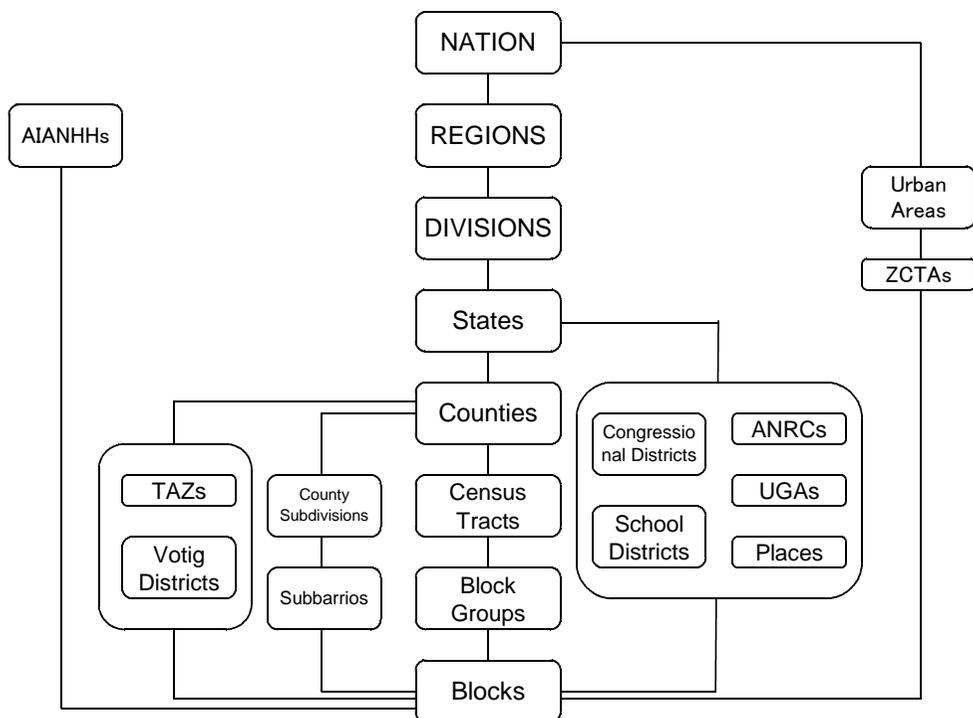


図3 2000年合衆国センサスにおける空間集計単位の階層関係

AIANHHs: American Indian Area/Alaska Native Area/Hawaian Home Land

ANRC: Alaska Native Regional Corporation

SLD: State Legislative District (upper & lower chambers)

TAZ: Traffic analysis Zone

UGA: Oregon Urban Area

ZCTA: ZIP Code Tabulation Area

資料 : Summary File 1 Technical Documentation

てられている³⁾。

ニューヨーク市などの「市・町・村」は図3の "Places"に含まれており、Placeには "Incorporated Place", "Consolidated City", "Census Designated Place" の3種類が存在する。Incorporated Placeは自治体化されたCity, Town, Villageなどであり、その境界線はカウンティやセンサストラクトの境界線とは必ずしも一致しない。Census Designated

Placeは、名称で区別されるが自治体化されていない (unincorporated) 人口・商業集積地である。Consolidated Cityは自治体とカウンティが一体化した地方政府である。

また図3には示されていないが、都市圏 (MSA), 都市圏中心都市といった集計単位も存在する。アメリカ合衆国の都市圏の設定方法は、2003年からCore Based Statistical Area (CBSA) とい

3) コード番号は、州は2桁、カウンティは3桁、センサストラクトは4桁または6桁の数字となっている。

う定義に変更されており、これは2000年センサスの人口・通勤データを基準に設定されている⁴⁾。しかしこの新しい定義による都市圏が使われるのは2003年の告知以降で(Office of Management and Budget, 2000)、2000年センサスのSummary File内で使われている都市圏は従来からの都市圏定義によるものである。

5. CTPP2000

2000年センサスデータのうち、通勤時間、通勤手段、中心都市への通勤者数などの通勤に関するデータはSummary File 3に含まれているが、地域間のODデータについてはSummary Fileには含まれていない。カウンティ間の通勤ODデータについてはセンサス局ホームページ内でExcelファイルとしてダウンロードできるようになっているが、より詳細なデータはセンサス局からは得られない。

センサストラクト間などの詳細なODデータや、就業地ベースのデータは、センサス局ではなく、合衆国運輸省において公開されている。そのデータはCTPP(Census Transportation Planning Package) 2000と呼ばれ、運輸統計局のホームページ(<http://www.bts.gov/>)からダウンロードできる。このデータは州ごとに集計単位を指定してCSVファイルとしてダウンロードできるので、ダウンロード後の操作はSummary Fileと比較すると容易である。しかし地域間のODデータに関しては、指定した州にかかわる全ての発着地の組み合わせが出力されるので、必要な流動パターンのみを抽出するためのツールがあるとより便利である。

通勤データは1/6サンプルのlong form調査によって得られたものなので、集計単位を小さくするほど誤差が大きくなる。たとえば、センサストラクト単位でダウンロードして得た数値をカウンティ単位で集計し、合計した場合と、最初からカウンティ単位でダウンロードした数値とは食い違っている。long formのデータを利用する際にはこの点にも注意を払う必要がある。

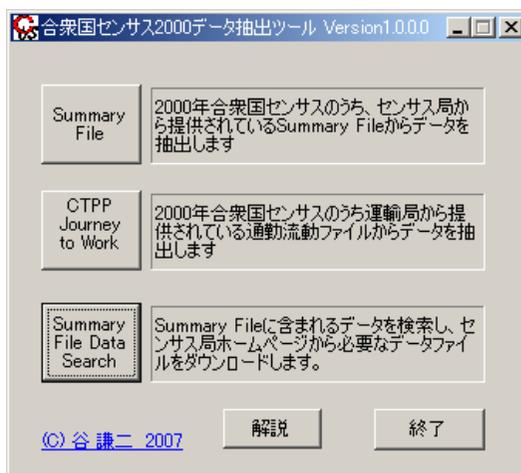


図4 データ抽出ツールの起動画面

III データ抽出ツールの開発

上記のようなSummary FileとCTPPのODデータから必要なデータを抽出するツールを開発した。開発言語はVisual Basic 2005で、ツールを使用するためには、NET Framework 2.0がインストールされたWindowsパソコンが必要である。ツールおよび詳細なマニュアルは谷(2007)を参考にするか、筆者のホームページ(<http://www5f.biglobe.ne.jp/~ktani/>)からダウンロードできる。

抽出ツールの機能は以下の3つである。①必要なSummary Fileを検索し、ダウンロードする、②ダウンロードしたSummary Fileから必要なデータを必要な集計単位で抽出し、Excelファイルとして出力する、③ダウンロードしたCTPP2000のODデータから、発着地を限定して必要なデータを抽出する。

図4はツールを起動した画面であり、3つのボタンがそれぞれ上記3つの機能に対応している。図5は機能①のSummary Fileをダウンロードする画面である。図5の画面では、必要なデータをまずキーワードで検索し、データの含まれるSummary Fileとそのファイル番号を確定する。次

4) 設定方法の変更については金本・徳岡(2002)に詳しい。

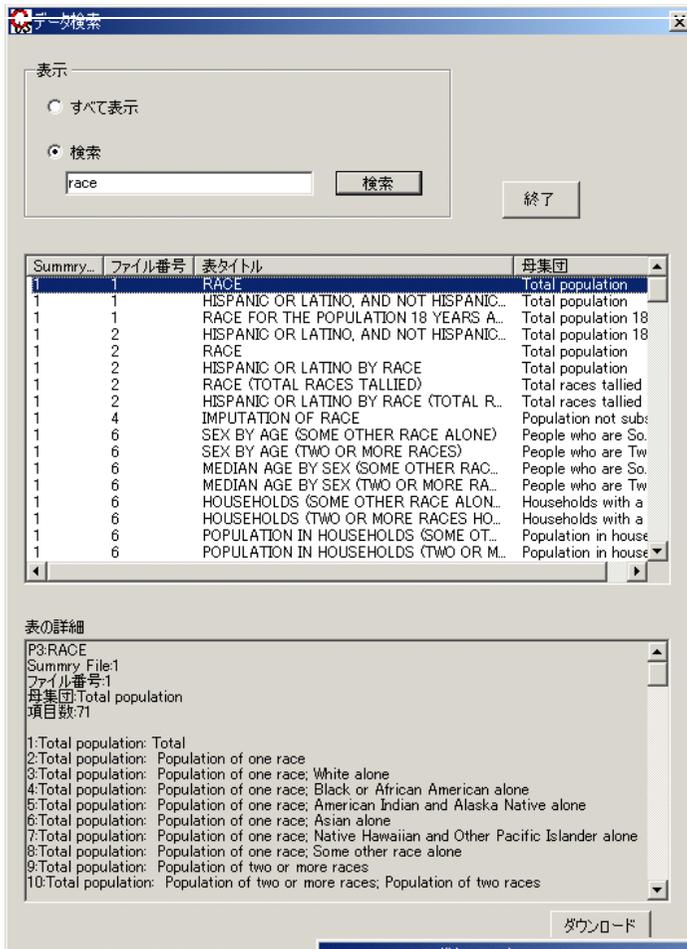


図5 Summary Fileのダウンロード画面

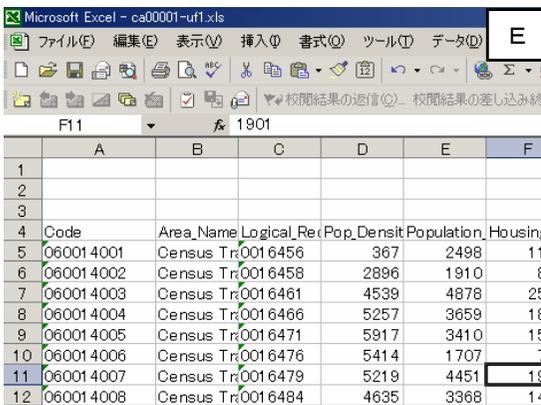
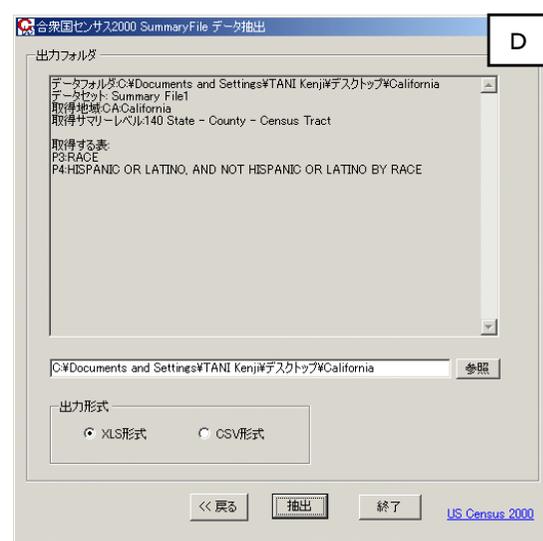
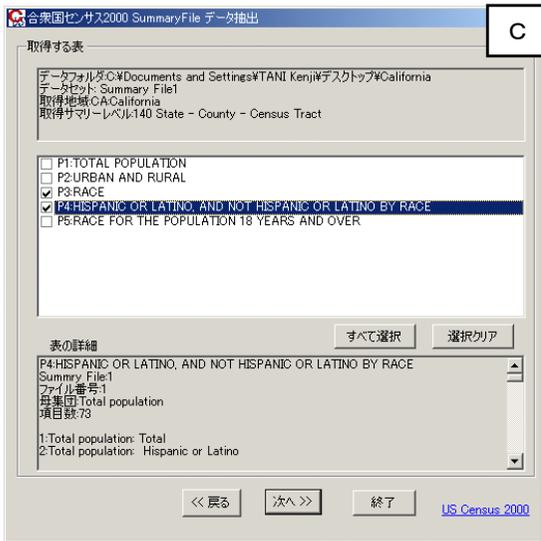
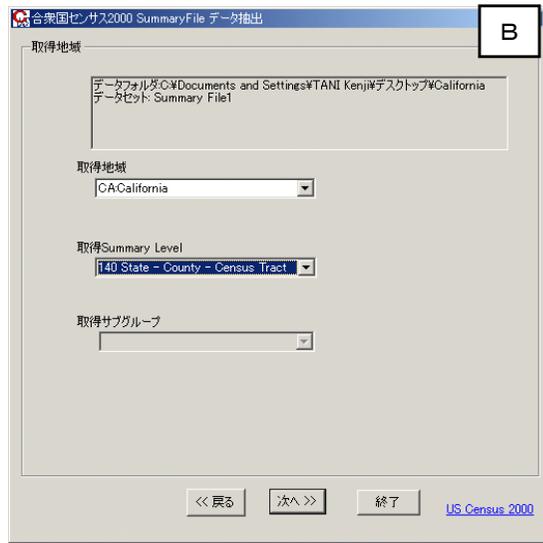


図 6 Summary Fileデータ抽出画面

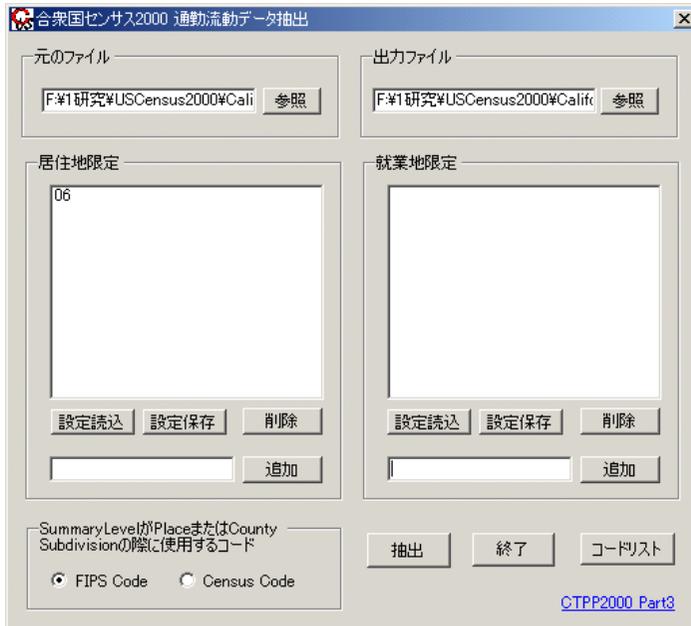


図 7 CTPP2000のODデータ抽出画面

の段階で、必要な州・国を選択し、ダウンロードする。Summary File 2と4の場合は、さらに下位人口集団も指定する。統計データの入っているSummary Fileと同時に、その地理的実体の情報が含まれるGeoHeaderファイルも必要となるので、ここで同時にダウンロードできる

図6は機能②の画面で、操作はウィザード形式で行われる。まずSummary Fileの入っているフォルダを指定し (A)、次に取得地域と空間集計単位であるSummary Levelを指定する (B)。さらにSummary File内のファイルから、抽出したいデータを指定する (C)。最後にデータの保存先フォルダと出力形式 (XLSまたはCSV) を選択し、実行する (D)。そうして抽出されたデータをExcelで示したものがEである。EのA列に含まれるものはコード番号で、たとえば集計単位がセンサストラクトの場合は、州-カウンティ-センサストラクトのコードが入る。人口、世帯数、面積はGeoHeaderファイルから抽出したもので、その後の列にSummary File内の指定したデータが入っている。

図7は機能③の画面である。この機能では、フ

ァイル中のすべての発着地間のODデータから、必要な発着地間のODデータを抽出する。たとえば、周辺地域からニューヨーク市への通勤者数をセンサストラクト単位で取得する場合には、まずニューヨーク市周辺の、ニューヨーク州、コネチカット州、ペンシルバニア州のODデータをセンサストラクト単位でダウンロードしておく。それぞれの州に関して、居住地に州のFIPSコードを指定し、就業地にニューヨーク市を構成する5つのカウンティのコードを指定する。センサストラクトのコードは上位の桁が州番号-カウンティ番号となっているので、ファイル内のセンサストラクトコードと、指定された条件の州・カウンティのコードを比較し、条件に合う発着地とその間の通勤者数が抽出される。ただし、ここで抽出されるのはあくまで通勤者数のみなので、就業者数に占めるニューヨーク市通勤者の割合を求めるような場合には、別途CTPP2000のデータをダウンロードし、Excelで加工してから、FIPSコードをキーとして、抽出したデータと一致させなければならない。詳細な手順はマニュアルに記載されている。

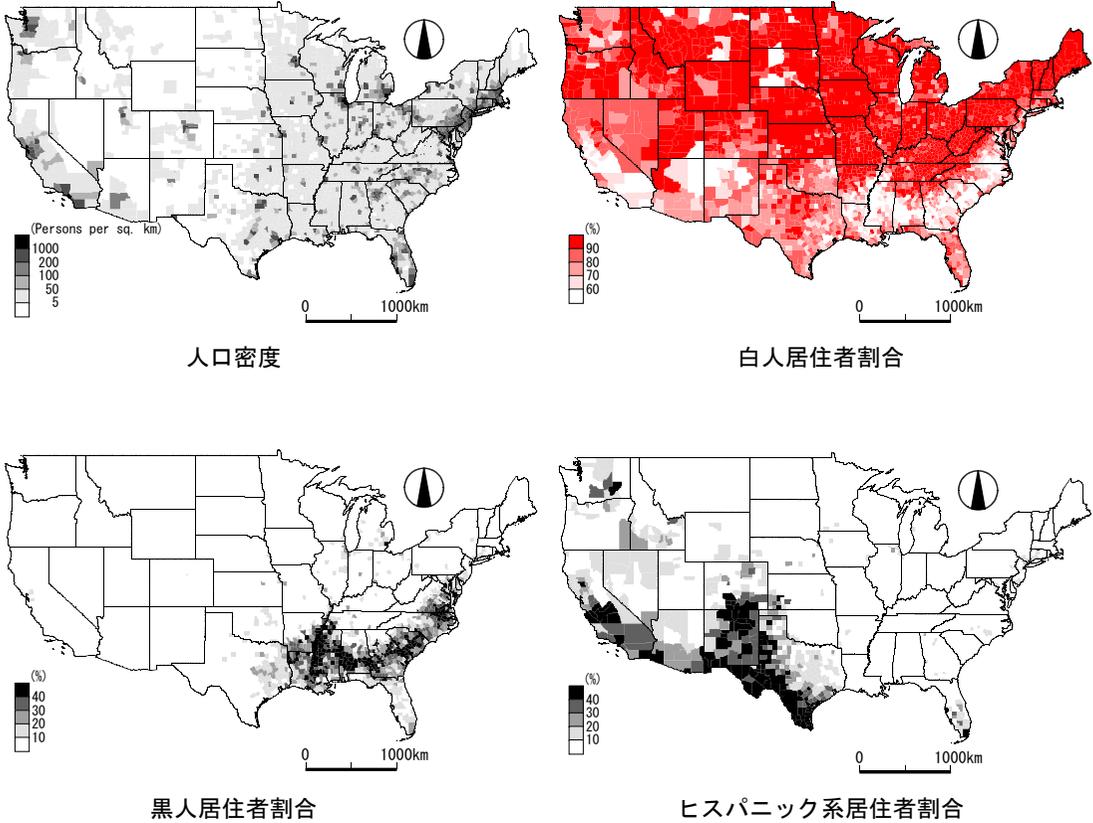


図8 カウンティ単位での地図化

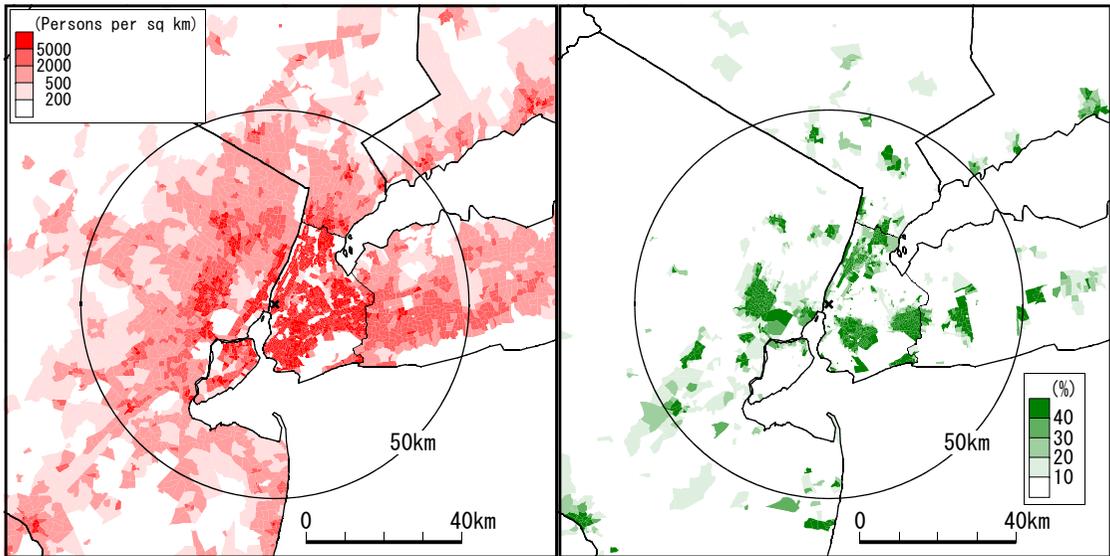
IV MANDARAでの地図化

統計データを抽出しただけでは地図化することはできないが、センサス局ホームページからはカウンティやセンサストラクトといった各種集計単位に対応した境界データをダウンロードすることができる。データはExport(.e00)形式とShape形式の両方が用意されている。どちらの形式のデータも、筆者の開発しているGISのMANDARAで取り込むことができる。ただし、Shape形式は位相構造化されていないデータなので、隣接する地区間の境界線は重複しており、あまり効率的でない。したがって位相構造を持つExport形式のファイルから読み込んだ方が、後の処理は容易である。

地図データはMANDARAに取り込んだ後、それぞれのオブジェクトに該当するコード番号をオブジェクト名として設定する。一方、データ抽出ツールを使って出力されるデータにも集計単位に応じたコードが付けられているので、これをキーとしてMANDARA上で両者を結合させる。

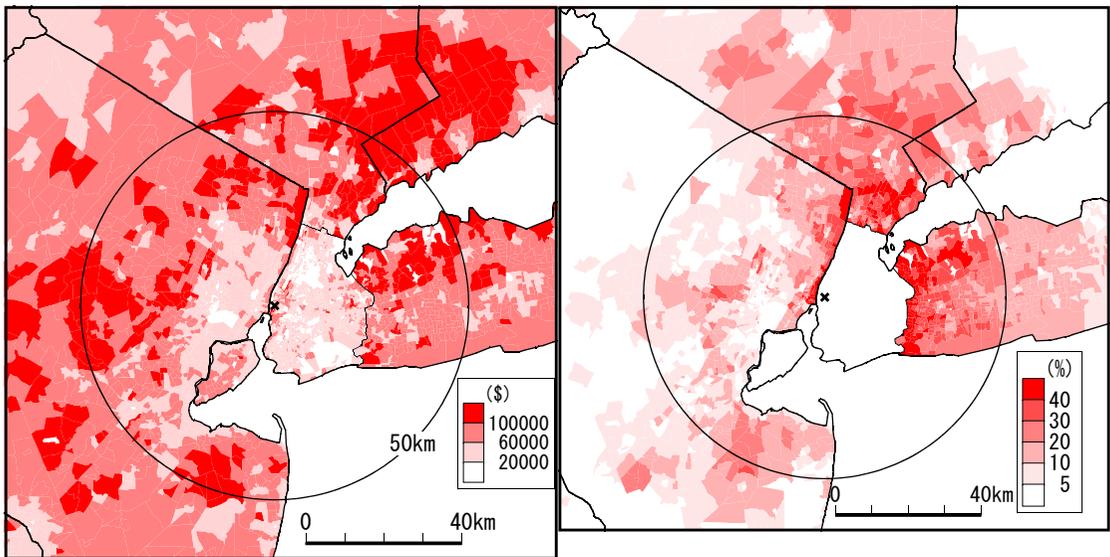
図8は、Summary File 1の人種データを、Final Nationalデータを使ってカウンティ単位で地図化したものである。カウンティの境界線データは、合衆国全体のものがセンサス局ホームページに用意されている。

図9はセンサストラクト単位で作成した地図であり、ニューヨーク市を中心としている。センサストラクトの境界線データは、州単位で用意されているので、複数の州にまたがる地図データを編



人口密度

黒人居住者割合



世帯収入（中位数）

ニューヨーク市への通勤率

図9 ニューヨーク市周辺のセンサストラクト単位での地図化

集して統計データと組み合わせるのは若干手間がかかる。黒人居住者割合から、ニューヨーク市内での顕著なセグレーションが存在し、また西側のニューアークにも黒人居住地区が見られる。Summary File 3に含まれる世帯収入では、中心部のマンハッタン島南部で高く、周囲のインナーシ

ティで低くなり、さらにニューヨーク市外の郊外になると高くなるという構造で、アメリカ合衆国都市の構造の一つの典型を示している。CTPP2000のODデータから取得したニューヨーク市への通勤率では、西側のペンシルバニア州からの通勤率は比較的lowく、東側のロングアイランドや、北側、

北東側のコネチカット州の通勤率が高いことがわかる。50km圏を超えても20%以上の地域が広がっており、ニューヨーク市が広大な通勤圏を持っていることが読み取れる。しかし人口密度で見ると、50km圏外の人口密度は500人/km²未満と低密度の地区となっている。ニューヨーク大都市圏の郊外では低密度地域に高所得層が居住し、ニューヨーク市へ通勤していることがわかる。なおセンサスの境界線データでは、ハドソン川やイーストリバー、ロングアイランドの入り江などもセンサストラクト内に含まれているので、別の河川や海岸線データとあわせて利用すると見栄えがよくなるであろう。

V おわりに

従来、詳細なデータは現地に行かなければ収集することが困難であった。そのためアメリカ合衆国の小地域のデータは、従来日本の地理学研究で使われることは少なかった。しかしながら、インターネットの普及とデータの無償公開の進展、さらにGISの発展により、日本にいながらにして合衆国の詳細なセンサスデータを地理学研究で活用できるようになった。しかしデータの構造は複雑で、簡単には理解できない。そこで本研究で開発

したデータ抽出ツールが役立つであろう。

今後は、合衆国以外の国についても同様なデータが公開されているかを調査し、その地図化の可能性について検討してみたい。

本研究の概要は2007年度人文地理学会大会において発表した。

文 献

- 金本良嗣・徳岡一幸 2002. 日本の都市圏設定基準. 応用地域学研究:7, 1-15.
- 谷 謙二 2007. 2000年合衆国センサスデータ抽出ツール利用マニュアル. 埼玉大学教育学部地理学研究報告:27, 50-79.
- 平井 誠 2006. American FactFinderを用いた統計データの利用. 統計:57-2, 16-21.
- 森 博美 2007. 合衆国における人口センサスの新展開—2010年センサス計画を中心に—. 日本統計研究所報:36, 29-48.
- Office of Management and Budget 2000. Standards for defining metropolitan and micropolitan statistical areas; notice. *Federal Register*:65-249, 82228-82238.

Development of a data extracting tool for US Census 2000 and mapping using MANDARA-GIS

Kenji TANI

Dept. Geography, Saitama Univ.