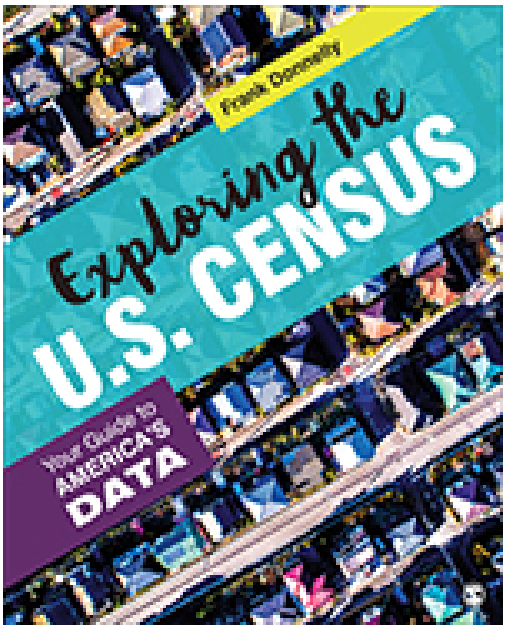

Sage Research Methods

Exploring the U.S. Census

For the most optimal reading experience we recommend using our website.

https://methods.sagepub.com/book/mono/exploring-the-us-census/toc

Author: Frank Donnelly

Pub. Date: 2023

Product: Sage Research Methods

DOI: https://doi.org/10.4135/9781071909645

Methods: Census data, Populations, Official statistics

Keywords: censuses, population, Bureau of the Census, estimates, counties, census tracts, race

Disciplines: Sociology, Political Science and International Relations, Geography

Access Date: September 19, 2025

Publisher: SAGE Publications, Inc.

City: Thousand Oaks

Online ISBN: 9781071909645

Front Matter

Chapters

Back Matter

Copyright

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London, EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Acquisitions Editor: Leah Fargotstein
Editorial Assistant: Claire Laminen
Content Development Editor: Chelsea Neve
Production Editor: Rebecca Lee
Copy Editor: QuADS Prepress Pvt. Ltd.
Typesetter: Integra
Proofreader: Ellen Brink
Indexer: Will Ragsdale
Cover Designer: Candice Harman
Marketing Manager: Shari Countryman

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Names: Donnelly, Frank (Francis P.), author.
Title: Exploring the U.S. census : your guide to America's data / Frank Donnelly, Baruch College.
Description: Los Angeles : SAGE, [2020] | Includes bibliographical references and index.
Identifiers: LCCN 2019029185 | ISBN 978-1-5443-5542-9 (paperback ; alk. paper) | ISBN 978-1-5443-5543-6 (epub) | ISBN 978-1-5443-5544-3 (epub) | ISBN 978-1-5443-5545-0 (adobe pdf)
Subjects: LCSH: United States. Census. | United States—Census. | United States—Population—Statistical methods. | United States—Population—Statistics.
Classification: LCC HA181 .D66 2019 | DDC 317.3—dc23
LC record available at https://lccn.loc.gov/2019029185

This book is printed on acid-free paper.

19 20 21 22 23 10 9 8 7 6 5 4 3 2 1

FOR INFORMATION:

SAGE Publications, Inc.

2455 Teller Road

Thousand Oaks, California 91320

E-mail: order@sagepub.com

SAGE Publications Ltd.

1 Oliver’s Yard

55 City Road

London, EC1Y 1SP

United Kingdom

SAGE Publications India Pvt. Ltd.

B 1/I 1 Mohan Cooperative Industrial Area

Mathura Road, New Delhi 110 044

India

SAGE Publications Asia-Pacific Pte. Ltd.

18 Cross Street #10-10/11/12

China Square Central

Singapore 048423

Printed in the United States of America

This book is printed on acid-free paper.

Library of Congress Cataloging-in-Publication Data

Names: Donnelly, Frank (Francis P.), author.

Title: Exploring the U.S. census : your guide to America's data / Frank Donnelly, Baruch College.

Description: Los Angeles : SAGE, [2020] | Includes bibliographical references and index.

Identifiers: LCCN 2019029185 | ISBN 978–1-5443–5542-9 (paperback ; alk. paper) | ISBN 978–1-5443–5543-6 (epub) | ISBN 978–1-5443–5544-3 (epub) | ISBN 978–1-5443–5545-0 (adobe pdf)

Subjects: LCSH: United States. Census. | United States—Census. | United States—Population—Statistical methods. | United States—Population—Statistics.

Classification: LCC HA181 .D66 2019 | DDC 317.3—dc23

LC record available at https://lccn.loc.gov/2019029185

19 20 21 22 23 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Leah Fargotstein

Editorial Assistant: Claire Laminen

Content Development Editor: Chelsea Neve

Production Editor: Rebecca Lee

Copy Editor: QuADS Prepress Pvt. Ltd.

Typesetter: Integra

Proofreader: Ellen Brink

Indexer: Will Ragsdale

Cover Designer: Candice Harman

Marketing Manager: Shari Countryman

Praise For Exploring the U.S. Census: Your Guide to America's Data

The most comprehensive source book on the subject that will be invaluable for anyone doing research using census data.

—A. Victor Ferreros, Florida State University

This book is interesting and easy to read, and it compiles information that I often have a hard time gathering before covering the census in my course. I feel much more prepared just having read it. I teach geography, and students make extensive use of census data but don't have a good understanding of where it comes from and its possibilities and limitations. While reading the book, I was considering teaching our census data class that has been on the shelf for a while. I was inspired by the text to offer it!

—Anne M. Mahar, Arcadia University

Exploring the U.S. Census provides an excellent overview of many parts of the U.S. Census Bureau's work. It is a timely reference for anyone interested in using their current and historic data.

—David S. Lamb, University of South Florida

This book integrates material that is difficult for new users of census data to find, as it is spread across dozens of technical manuals for different years and different data products. The attentive reader will skip years of novice traps and false starts. There is no comparable product on the market today.

—Dr. Stephanie Deitrick, Arizona State University

This is a much-needed text that will help students, researchers, and practitioners understand and properly use census data.

—Lee Hachadoorian, Temple University

This text is a welcome, comprehensive introduction to working with and understanding census data.

—Hugh Bartling, DePaul University

This is the most comprehensive guidebook to the use of the census. The author's experience is invaluable.

—Elizabeth C. Delmelle, University of North Carolina at Charlotte

Preface

The census, meaning the array of demographic and socioeconomic datasets produced by the U.S. Census Bureau, is an incredibly valuable and freely available resource for students, researchers, policy makers, and practitioners in many fields. In spite of this, in my 12 years of working as an academic librarian, I've discovered that this resource is either unknown or impenetrable to many. The sheer number of datasets, the unfamiliar terminology, difficult search interfaces, or conversely too many websites and tools to choose from, and the diffused nature of resources for learning about the census are formidable. Once I help students overcome these barriers, a new set typically emerges: Now that I have this data, what do I do with it? How do I study, manipulate, process, visualize, and integrate it into my work?

This book was designed to gather and organize all the essential information about the census in one place and to cover the basic geographic and subject-based concepts and the most important datasets in conjunction with tools and exercises that illustrate how to work with the data. I wrote it to appeal to the broadest possible audience, so that it doesn't focus on details that are relevant to only one or two specific fields. It's a hybrid of an academic text and a computer technology guidebook. Like the techie guidebooks, this book assumes that you learn best by doing, with practical step-by-step examples that all users will need to perform when working with the census. Like the academic texts, this book provides you with the necessary background and context needed for understanding, interpreting, and applying the statistics.

Given this broad audience, I chose to employ free and open source tools: spreadsheets, relational databases, GIS, and freely accessible websites to guarantee that anyone who reads this book will have ready access to both the data and applications. If you're a techie novice, you'll learn a lot about census data while simultaneously learning these tools. If you're already a tech expert and a seasoned analyst, but your knowledge of the census is limited, you'll learn a lot about these datasets and how to interpret and work with them properly. This book does not focus on statistical analysis, as the audience for this topic is narrower and there are already countless books devoted to learning it.

This book is divided into three parts. Part I is designed to provide essential context and background information that will apply to all the datasets that we cover: the role and function of census data, the primary interface used for obtaining data, and geographic and subject terminology used for organizing it. This part also provides initial exposure to several datasets that are explored in earnest in the second part. In Part II, each chapter covers a specific dataset in detail with exercises geared to common data-processing tasks. Part III covers advanced topics that will be valuable to many researchers.

The first six chapters should be read in order. If you skip Chapters 3 and 4 that discuss how the data is organized geographically and by subject, the subsequent chapters will make little sense. If you skip Chapter 2, you won't learn the basics of navigating and downloading data from the Census Bureau's website. Since downloading the data every time would be tedious, exercise data for this book can be downloaded in one package from the publisher's website. Chapters 5 and 6 cover the decennial census and American Community Survey, respectively, and should be covered in that order as the latter is based on the former. Chapters 4 through 6 also introduce the basics of working with spreadsheets and SQL, and in subsequent chapters, knowledge of these tools is presumed. The remaining chapters from 7 to 12 can be covered in any order based on the reader's individual research interests. These include additional population and business datasets (Chapters 7 and 8), integrating census data into research (Chapter 9), mapping and GIS (Chapter 10), creating derivatives for analysis (Chapter 11), and historical research and microdata (Chapter 12).

As I received questions from students and faculty throughout the most recent semester, I kept wishing that this book was finished so I could refer them back to specific chapters or sections that would answer their questions—and so I could remember the details too! My hope is that this book will initially be your key to unlocking the census, and then, it will become your indispensable reference once you have.

Acknowledgments

While writing is a solitary endeavor, the process of creating a book certainly is not, and I'd like to thank a number of people who helped make it happen. Thanks to Laura Horne-Popp, Stephen Francoeur, and Jason Haremza, who graciously agreed to review the proposal. Special thanks to Jean Yaremchuk, Jessie Braden, and Hannah Gurman, who reviewed it and met with me to strategize; Laura, Jessie, Steve Romalewski, and Joel Alvarez, who agreed to serve as references; my graduate students, past and present: Ryan Lennon, Abigail Andrews, and Chris Kim, who reviewed the exercises and took them for a test run; Anastasia Clark for her Python advice; and Janine Billadello for keeping the lab up and running while I was off writing.

Having a yearlong sabbatical is truly a privilege (thanks to the faculty of Baruch College who approved my fellowship leave), but after the first solitary month sitting in your one-bedroom New York City apartment, it becomes slightly dreary. So I'm grateful for many people who provided other places for releasing my creativity: the staff at the New York Public Library for the Performing Arts at Lincoln Center, Jane Huerta and Pastor John Flack for their coworking space at Our Savior's Atonement Lutheran Church in Washington Heights, and Jean-Paul Gheleyns and Jacqueline Degallaix for loaning us their place in Etel, Brittany, for one lovely month (where the one-bedroom apartments have views of the sea and a constant supply of fresh baguettes nearby).

Many thanks to the staff at SAGE, especially to my encouraging and patient editor Leah Fargotstein, and to all the reviewers of this book:

Hugh Bartling, *DePaul University*

Richard F. Bieker, *Delaware State University*

Courtney Broscious, *Eastern Connecticut State University*

Perry L. Carter, *Texas Tech University*

Dr. Stephanie Deitrick, *Arizona State University*

Elizabeth C. Delmelle, *University of North Carolina at Charlotte*

Rustu Deryol, *University of South Florida Sarasota-Manatee*

A. Victor Ferreros, *Florida State University*

Lee Hachadoorian, *Temple University*

David S. Lamb, *University of South Florida*

Anne M. Mahar, *Arcadia University*

Thanks to my mother and father for their lifelong support. Last but not least, thanks to my wife, Julie Poncelet, for simply everything.

About the Author

Frank Donnelly is a geospatial information professional whose practice blends the service-based and organizational skills of an academic librarian with the subject knowledge and analytical methods of a researcher. He has served as the Geospatial Data Librarian at Baruch College, CUNY, in midtown Manhattan since 2007, where he helps members of his university navigate geospatial and census data sources. He holds the rank of associate professor in the library and manages a GIS lab, where he and his graduate students provide research consultations, teach workshops, process and create data, and maintain a repository of GIS data. He was an early proponent of free and open source GIS software in academia and has introduced hundreds of people to GIS through his workshops and tutorial manual. He has written several academic articles, technical papers, and reports that utilize census data to study socioeconomic and demographic trends and provide information to researchers and policy makers. Prior to becoming a librarian, he worked as a planner and data analyst in the government and nonprofit sectors. He holds master's degrees in library and information science from the University of Washington and in geography from the University of Toronto. You can follow him at his blog **At These Coordinates: https://atcoordinates.info/**.

Introduction

Introducing the Census

When most Americans think of the census, they think of the form they fill out every 10 years that's used to count the nation's population. Or they think of the Census Bureau, which is the agency that conducts the census. This book will demonstrate that the census is more than just a basic 10-year counting exercise; it is actually a rich collection of public datasets that are updated on an ongoing basis. The census plays an essential role within the government, as it is used for reapportioning seats in Congress between the states and redrawing legislative districts. It is also used to distribute hundreds of billions of dollars in federal funds to states and local governments to support education, transportation infrastructure, and social services.

Beyond simply counting the number of people and houses, the census includes data on educational attainment, commute time to work, home values, internet access, components of population change (births, deaths, migration), types and numbers of businesses, and much more. The data is summarized by age, sex, race, income brackets, occupations, and industries for large geographic areas like states, metropolitan areas, and counties as well as small areas like ZIP Codes and census tracts that can be used to study neighborhoods. The Census Bureau uses consistent methods and definitions across datasets and, to the extent possible, across time periods.

This high-quality, geographically detailed data is in the public domain and, thus, is freely available for anyone to use for any purpose. Census data is used by researchers, policy makers, urban and regional planners, academics, activists, journalists, entrepreneurs, data scientists, nonprofits, businesses, and federal, state, and local governments to study and understand society, identify markets, solve problems, provide services, and create profiles, maps, charts, reports, and web applications.

There is a learning curve in working with census data. There are many datasets to choose from, there is unfamiliar terminology, and there is a dizzying array of websites of varying complexity where the data can be accessed and downloaded. Some of the datasets require a degree of technical knowledge to interpret and use them properly, and you will need some technical skills to effectively process and summarize the data in order to use it to meet your needs. This book is your guide for exploring, understanding, and mastering the essentials of working with census summary data and will do the following:

- Demonstrate how to access and obtain census data from several sources, in varying quantities.
- Introduce the terminology and concepts that are used across all census datasets as they relate to subjects and geography.
- Cover each of the primary census datasets in detail: their purpose, how they're constructed, and what data is available.
- Demonstrate basic techniques for processing and summarizing data using spreadsheets.
- Demonstrate how to create derived data and how to work with essential census formulas.
- Show you how census data can be applied in different use cases.
- Illustrate advanced techniques like storing and querying data in a relational database and mapping data with geographic information systems (GIS).

This book is an introductory guide aimed at researchers and practitioners in many fields. It was designed as a hybrid between an academic text and a technology-oriented guidebook based on the fundamental principle that you learn best by doing, and you will work through examples with relevant tools rather than having concepts explained to you without a meaningful application. If you are new to the census and know absolutely nothing about the datasets except that you want to use them for your research, and your technical skills are limited to copying and pasting values in spreadsheets, this is the perfect book for you. At the beginning, we will use basic tools like spreadsheets and web applications and will gradually introduce more advanced techniques. You won't be able to fully learn tools like relational databases or GIS here, but I will illustrate the basics of each and how you can apply them to the census to plant the seeds of what's possible in your mind and prepare you to go off and learn more.

On the other hand, if you are a census novice but possess good technical skills, this book is also a good fit for you. You will learn all the census fundamentals, so you can work with the data confidently and correctly. You will be able to move more quickly through the technical basics and will appreciate the advanced techniques introduced in later chapters. You will be able to readily draw from the book's examples and apply them.

Now that we covered what's in the book, let's discuss what this book is *not* about:

- *Working with individual, case-level data.* This book focuses on census summary data, which has been aggregated by population groups and geographic areas. The individual responses to the census forms are not publicly available and are kept confidential for 72 years. While the census does release some sample data of individual responses that we will cover in Chapter 12, it is not the primary focus of this text.

- *Genealogy.* This book is not going to show you how to research your family history. We will focus largely on summary data that is contemporary. While Chapter 12 covers historical census data, the focus is on working with data that is summarized by places and groups and not on retrieving historic records of individuals.
- *Learning statistics.* This book is not a primer for statistical analysis, so if you want to learn how to do linear regression, this is not the book for you. This book covers plain data analysis: obtaining, processing, summarizing, creating derivatives, and understanding what the data represents, so you can use it to support research or create new information.
- *Summarizing everything on the Census Bureau's website.* The Census Bureau publishes a ton of datasets, applications, and reports. Instead of covering them all, we focus on essential datasets that are widely used by most researchers and on a selection of the most salient and useful tools.

Roadmap of Topics

This book is divided into three sections. The chapters in Part I (Census Fundamentals) cover the essential information that applies to all the datasets included in the book. Chapter 1 provides a summary of what the census is with examples of how you can use it for research, and places the census within the context of U.S. society and the world of data. In Chapter 2, we will take a crash course in using the new data.census.gov, which is the primary portal for accessing all of the Census Bureau's datasets. By 2020, it will fully supplant the veritable American Factfinder, which had served as the previous portal for 20 years. In navigating data.census.gov, you will gain exposure to the different datasets we will explore throughout the book. One of the best aspects of working with the various census datasets is that they all employ the same concepts and terminology. In Chapters 3 and 4, we explore the two principal means by which census data is summarized: geography and population categories that are known as universes. We will navigate the census geographic hierarchy, which goes from the nation down to individual census blocks, and will explore the fundamental categories of race, ethnicity, households and families, and the labor force.

In Part II (The Primary Datasets), we will tackle each of the major census datasets in turn. Each chapter includes a discussion of the purpose, collection methods, variables, and use cases for each dataset and addresses how each one fits into the larger census canon, while highlighting when you should use one relative to another dataset. In these chapters, we'll also explore a number of different sources for accessing census data in varying quantities. Chapter 5 focuses on the decennial census, which is the foundation of all

census datasets. We will discuss how the 2010 census was a break from the past and what the 2020 census will look like. Chapter 6 covers the American Community Survey (ACS), which is the chief source for annual, geographically detailed data on the socioeconomic characteristics of the nation. A detailed exploration of the ACS methodology will give you a solid understanding for how to use and interpret the estimates. Chapter 7 introduces the Population Estimates Program, which is a compact and easy to use dataset for studying annual population growth. Chapter 8 pivots away from population and housing to businesses with the introduction of the County Business Patterns and the Economic Census and concepts such as the North American Industrial Classification System that are essential for studying businesses. Once each dataset has been covered, Chapter 9 illustrates how you can integrate census data into your writing and research and provides useful background information on geographic considerations and historical trends that influence research findings.

Each of the chapters in Parts I and II include hands-on exercises that use online tools, spreadsheets, and relational databases to explore, process, and analyze data. We will begin using spreadsheets in Chapter 2 and will introduce relational databases and SQL in Chapter 5. Part III (Advanced Topics) is devoted to more advanced topics: types of analyses we haven't covered in the first two sections, datasets that are important but not as commonly used as the primary ones, and advanced tools like GIS that allow you to geographically visualize and analyze data.

Chapter 10 introduces the different geographic products that the Census Bureau produces that can be used with GIS. You will get a crash course in GIS, so that you can use these products and make your own maps of census data. Chapter 11 focuses on creating derivative data, everything from simple aggregates of geographies to population distribution measures. This chapter consists of a series of vignettes that demonstrate different measures and techniques that can be implemented in spreadsheets, databases, and GIS.

Chapter 12 concludes with two advanced topics. The first is analyzing historic census summary data and coping with changes in subjects and geographic boundaries over time. The second is working with census microdata, which are samples of individual responses to census questions. This chapter will introduce the Current Population Survey and will use tools from the Minnesota Population Center for accessing both historical census data (the NHGIS) and microdata (IPUMS).

Data and Software

We will be doing exercises throughout the book, and you will learn how to navigate the Census Bureau's website and a few other resources, such as the Missouri Census Data Center and the Minnesota Population Center, to access data. Since it would be tedious to have to download each and every data table we need, most of them are bundled together and can be downloaded from the publisher's website at **study.sagepub.com/census**. The data is organized in folders by chapter and exercise number. Each of the data tables and sources are listed prior to beginning each exercise, so if you prefer to download them from the original source you could. Some exercises are not included in this book but exist as supplementary material on the website. They consist of more specialized material that will be of interest to some readers and to others who would like more practice working with the data.

You will need to download some software in order to complete the exercises and learn the concepts. Since the census is a free and open dataset, we are going to use free and open source software to work with it. The packages listed below are freely available, are easy to install, and will run on any operating system: MS Windows, Mac, or Linux. They are well-established products that are stable and widely used. Links for downloading the software are below each title; we will start using Calc in Chapter 2 and SQLite® in Chapter 5 but won't get to QGIS until Chapter 10. Several of the exercises will use web-based resources and tools, and you are free to use your favorite web browser with the caveat that the rendering of websites and the mechanics of downloading files may vary between browsers. Mozilla Firefox was used for the examples and exercises in this book.

Spreadsheet: LibreOffice® Calc 6

https://www.libreoffice.org/

The spreadsheet is still the Swiss Army Knife of the data world, and given its ubiquity and familiarity, it will be our first tool of choice. LibreOffice® is a free and open source office suite that includes all the standard office tools: a word processor, a slide presentation program, a spreadsheet, and so on. Originally a proprietary package, it was relaunched as the open source OpenOffice suite and then subsequently branched into LibreOffice in 2010, so that development could continue via an independent open source team. In this book, we will be using LibreOffice version 6.

The LibreOffice spreadsheet is called Calc, and it looks and behaves just like any other spreadsheet package. If you are familiar with MS Excel or Mac Numbers, you can easily pick up Calc. Most of the formulas in Calc are identical to formulas used in Excel, and in Calc, you can open and save files in either the modern (.xlsx) or old (.xls) Excel format in addition to the open office format (.ods). Compared with Excel, Calc has much better support for working with CSV and text-delimited files, although its graphing functions are not as slick.

If you are a die-hard Excel user, you can still follow most of the examples in this book with Excel; the primary differences will be with the interfaces for accessing various tools and the charting functionality. I will mention some of the significant differences between the two packages as we move through the exercises.

Database: SQLite

http://sqlitebrowser.org/

While a spreadsheet offers flexibility and ease of use, it's less effective when you need to tie related but separate tables of data together, or when you need to group or summarize data. Relational databases give you the power to easily organize, summarize, and relate data because they offer structure and integrity over flexibility, and they don't mix formatting and presentation of data with organization and storage.

In this book, you will get a brief introduction to databases and the structured query language (SQL) using SQLite. While you may have never heard of SQLite, you probably use it everyday. It is a small, public domain database that is embedded in just about everything for storing information, such as your web browser (for storing bookmarks), phone (for storing contacts), and millions of websites. While it's commonly used as an embedded database within software, it can also be employed as a simple desktop database where each database is stored in a single file.

The SQLite developers provide the core program as well as a basic command-line interface for interacting with the database. Since the project is in the public domain, many independent developers have built graphical interfaces that make the database easier to use. In this book, we will use the DB Browser for SQLite. Of all the possible options, it has a clean and simple design and can easily be downloaded and installed on any operating system.

Geographic Information Systems: QGIS 3

https://www.qgis.org/en/site/

GIS are used to organize, process, and study spatial data, which is data that is tied to a specific location or place. Census data is inherently geographic, as each record represents a place and each column contains attributes that describe that place. We can use GIS to tie census data together, do geographic analysis, and make maps. GIS software works with special kinds of data files that incorporate the geometry and location of features so that they can be displayed graphically. The Census Bureau produces boundary files for every single geography that it tabulates data for. It also publishes features that it uses for creating its boundaries, such as roads, railroads, water, and other landmarks. If we have census boundaries for areas like counties, and we have census data for those counties in a table, we can join the table to the boundary file based on the common geographic code that they share and voila, that data is now visual and we can map it!

QGIS has emerged as one of the most popular free and open source GIS packages. It has a large number of developers and contributors, and updates are rolled out on a regularly scheduled basis. It has solid tools for both visualization and mapping as well as analysis and is highly extensible. Compared with other packages, there is a lot more user documentation devoted to it, and it's now commonly deployed in academia, the geospatial tech sector, nonprofits, and increasingly in government. In this book, we will use QGIS version 3.4.

If you have never used GIS, this book will introduce enough of the basics, as well as the common stumbling blocks, so that you can see the power of using it to work with census data. But when it is time to work on your own projects, you will need to learn more on your own. I'll offer suggestions on resources that you can turn to.

Conventions

Before we get started, some notes about terminology. When we talk about "the census," we could be referring to many things: the decennial census, the Census Bureau, or the census in the general sense referring to the collection of all census datasets and products. To keep things straight, I will refer to the 10-year count as the decennial census or will reference the specifc decennial census such as the 2020 census. I will refer to the agency as the Census Bureau, or simply as the Bureau, and will use the term *census* to generically reference

all datasets and products.

When referring to “the census,” the word is seldom capitalized. According to the Government Printing Office, you capitalize “census” only when you are referring to the agency itself (either in full as the Census Bureau or in part when referring to the agency as “the Census” or “the Bureau”) or when referring to an official title of a product (U.S. Census 2020, the Twenty-Third Decennial Census, 2012 Census of Agriculture). You would not capitalize “census” if you are referring to a product generally (2020 census, the decennial census, the census of agriculture). Throughout this book, I am going to avoid official titles and will stick with general references.

Naturally, when you work with government information, you are going to encounter a lot of acronyms. Here are the most common ones that we will use throughout:

ACS: American Community Survey (a census dataset)

BLS: The Bureau of Labor Statistics (a federal statistical agency)

CBP: County Business Patterns (a census dataset)

CPS: Current Population Survey (a census dataset)

GIS: geographic information systems (a type of software)

MCDC: Missouri Census Data Center (a data provider)

MOE: margin of error (statistical concept used with ACS data)

NAICS: North American Industrial Classification System (a classification system)

NHGIS: National Historic Geographic Information System (a census data repository)

OMB: Office of Management and Budget (a federal agency)

PEP: Population Estimates Program (a census dataset)

PUMA: Public Use Microdata Area (a census geography)

USDA: U.S. Department of Agriculture (a federal agency)

ZBP: ZIP Code Business Patterns (a census dataset)

ZCTA: ZIP Code Tabulation Area (a census geography)

Supplementary Digital Content: Find datasets and supplemental exercises at the companion website at http://study.sagepub.com/census.

Placing the Census in Context

This chapter provides an introduction to the census in the broadest sense: as a series of datasets, a statistical agency, and a social and political concept. We begin with a summary of the fundamental datasets that are covered in this book and explore how you can use census data in your research with some examples. In doing so, we will touch on concepts that we will cover throughout the text. While this book is primarily a practitioner's guide to working with census data, this chapter provides essential background information so you can better understand and appreciate the importance and value of the census. We will discuss the roles the census plays within American society and how census data fits within the context of the ever-expanding universe of data that includes the federal statistical system, the open data movement, and big data.

What is Census Data?

We can think of the U.S. Census as a collection of datasets about population, housing units, and businesses that is created by the Census Bureau, which is part of the U.S. Department of Commerce. Census data is collected at regular intervals using methodologies such as total counts, sample surveys, and administrative records. After it is collected or generated, census data is summarized to represent counts or estimates of groups of people for different geographic areas. Census geographies, categories, and terminologies are relatively consistent across the different census datasets, and we will explore them in Chapters 3 and 4. A comparative summary of the datasets covered in this book is provided in Table 1.1.

TABLE 1.1 COMPARISON OF CENSUS DATASETS COVERED IN THIS BOOK

Dataset	Method	Frequency	Subjects	Geographies	Variables
Decennial Census	100% Count	10 years	Population, housing	Many	Several
American Community Survey	Sample survey 3.5 million addresses	Annual	Population, housing	Many	Many
Population Estimates Program	Administrative records	Annual	Population	Several	Few

Current Population Survey	Sample survey 60k households	Monthly	Population	Few	Many
Business Patterns	Administrative records	Annual	Businesses	Several	Few
Economic Census	100% count and sample survey	5 years	Businesses	Several	Several

When most Americans think of the census, they think of the 10-year or decennial census that is used to gather basic data about the total population. The decennial census is an actual count of people and housing units, and it serves as the baseline for measuring and generating other census datasets. Demographers refer to data that is collected from total counts as enumerations, or simply as populations. The American Community Survey (ACS) and the Current Population Survey (CPS) are ongoing sample surveys of the population that collect detailed demographic and socioeconomic characteristics. Sample surveys collect information from just a small subset of the population, either randomly or from targeted groups, which is used to estimate what the total population is. The size of the sample is carefully determined, so that the sample data can be used to estimate the total population for a given geographic area with a reasonable level of precision. The ACS is a large survey that is published annually for large and small geographic areas, while the CPS is a smaller survey that is published monthly and is summarized for the nation as a whole or for the states. The Population Estimates Program (PEP) is produced from administrative records and other census datasets to create annual estimates for areas like states, counties, and municipalities. The Census Bureau produces data for businesses via the Economic Census, which is a 5-year count of most types of businesses and a sample of other types, and the County and ZIP Code Business Patterns (ZBP), which is created from administrative records on an annual basis.

Who is counted in the census? It varies based on the dataset, and we will cover the specific details about the different methodologies that are used and the variables that are collected in Part II of this book. For now, the short answer is “everyone.” The decennial census counts all people residing in the United States on census day: citizens and permanent residents; documented and undocumented immigrants; people living in households; people living in institutionalized settings like college dormitories, military bases, prisons, and hospitals; and the homeless.

The ACS and CPS are primarily sample surveys of residential addresses, so they do not capture the full spectrum of the population that the decennial census captures. The ACS does sample people living in group

quarters (institutionalized settings), but the sample is small enough that it is able to publish coarse estimates only for large areas like states. The PEP is derived from the decennial census and administrative records that include birth and death certificates, so in theory it captures everyone. The business datasets capture most businesses, with some exceptions.

Census data is captured from households, institutions, and businesses through paper and online forms and, when necessary, through on-site visits and canvassing. One of the reasons that the Census Bureau is able to produce reasonably accurate and geographically detailed counts and estimates of the population is that it is a government agency that is backed by law. People are required to fill out and return their census forms. The Census Bureau sends out a series of reminders to nonrespondents, and if a household still does not respond, the Bureau sends an actual enumerator out to interview them for the decennial census and follows up with a sample of nonrespondents in person or on the phone for the ACS. In contrast, private polling agencies would never be able to accomplish a count or survey at the same scale due to the cost of conducting it and their inability to compel people to respond.

Applications of Census Data

What can you use census data for? At the simplest level, you may want to look up information for your town, city, or state to get some basic facts to support a story you are writing or research you are doing. The Census Bureau publishes profiles that contain a broad swath of data for one place. With the Bureau's new data discovery platform, **data.census.gov** (Figure 1.1), a simple place name search will provide you with quick facts, charts, and maps. We will explore this platform in Chapter 2.

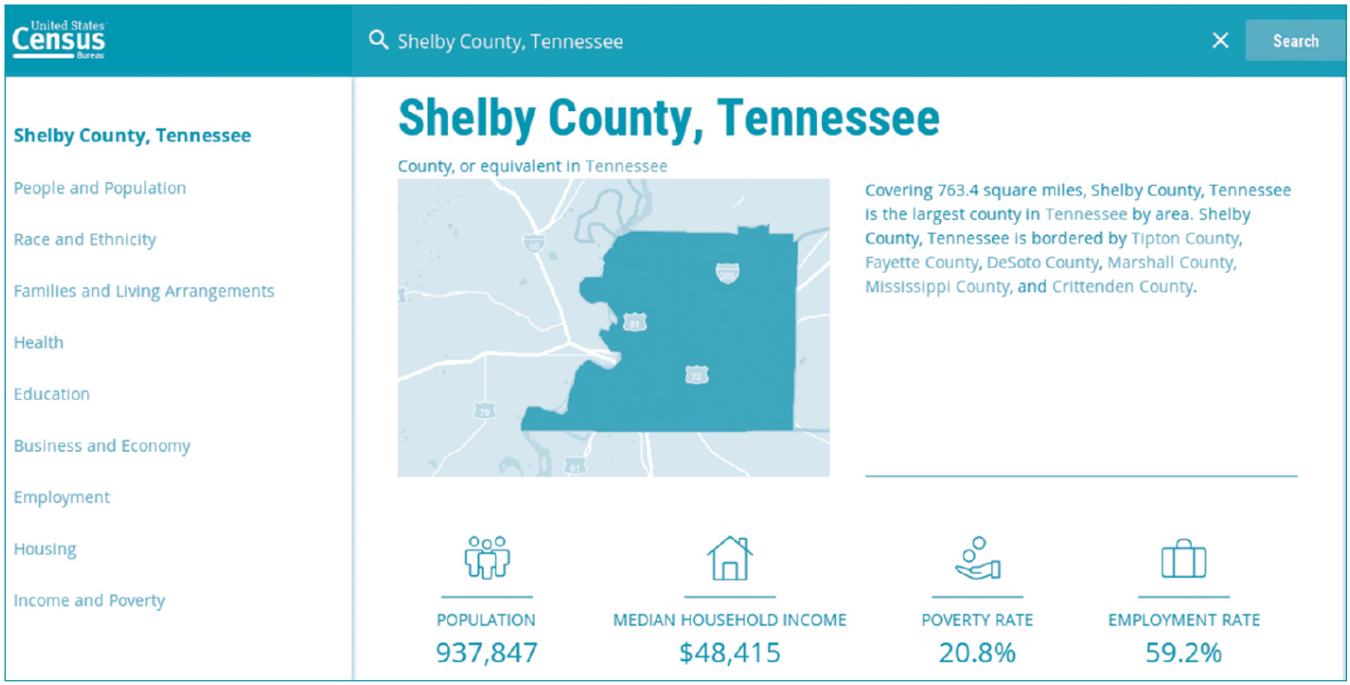

FIGURE 1.1 GEOGRAPHIC PROFILES USING DATA.CENSUS.GOV

Alternatively, you might want to compare one variable for many places in order to see which cities are growing fastest or which areas have the highest income or most unemployment. The Census Bureau publishes comparison tables that you can search through, modify, and download. Or maybe you need to gather many census variables for many places for a research project where you are creating new data, maybe even with data from other sources. The Census Bureau allows you to download data in bulk or to access it via a computer program or script using an API (application programming interface).

Or perhaps, you need to visualize census data. You can do this using a number of online tools, or you can download the data and visualize it on your desktop using spreadsheets or geographic information systems (GIS). In this book, we will demonstrate several of these methods to create charts and maps like the examples in Figure 1.2 that depict commuting data for the Northeast Corridor and New York City from the ACS.

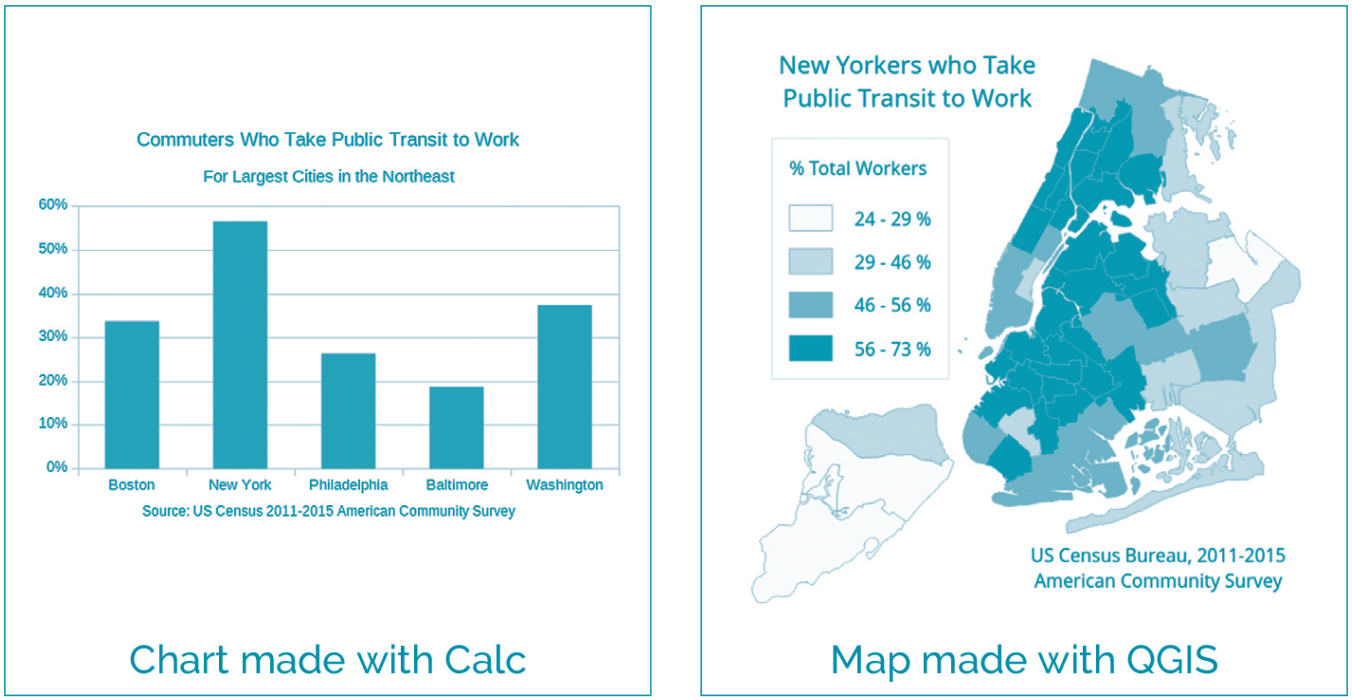

FIGURE 1.2 VISUALIZING CENSUS DATA

As an academic librarian who specializes in geographic datasets, I have helped hundreds of people find, process, and interpret census data for their projects to support arguments in their research and to create new information. This is the kind of data analysis that professor Gary Klass describes in his book *Just Plain Data Analysis* and that we will cover in this text. Klass makes a distinction between "plain data analysis" as processing, presenting, and evaluating statistics to support social and political research as opposed to statistical analysis, which focuses on the testing of hypotheses (Klass, 2012). Here are a few examples that illustrate the kinds of research you can do with census data and what you will learn from reading this book.

- Each semester, I help journalism students with neighborhood reporting projects in New York City. The Census Bureau does not collect data for "neighborhoods" like Midtown, Chelsea, or Harlem, as these are areas that are defined locally. I assist them with translating the Census Bureau's geographies like census tracts or Public Use Microdata Areas into what we consider to be neighborhoods, and we walk through a number of different online sources for census data where they can get profiles. We will cover sources for data in Chapter 2, geography in Chapter 3, creating aggregates for neighborhoods in Chapter 6, and integrating census data into writing in Chapter 9.
- I worked with two journalism professors to combine ACS data and presidential election results in order to identify counties that had low median income, high unemployment, and high poverty compared with the U.S. average and that had switched to voting for the Republicans in 2016 after having voted

for the Democrats in the previous two presidential elections. I gathered and loaded the variables into a database, so that we could select counties that met the criteria. The professors combined the results of this analysis with other data to select a county that would serve as a field trip destination for an investigative reporting class. We will cover population groups in Chapter 4 and the ACS in Chapter 6 and will introduce databases in Chapter 5.

- A PhD student was doing research on heat waves, heat-related death, and poverty. She was working with county-level data from many sources from the 1970s to the early 2010s. We not only dove into the historical census files to obtain population and poverty data but also discovered another important variable she could use: From 1960 to 1980, the census asked people whether they had air conditioning in their homes. She was able to use this data for older decades and then created estimates for recent decades using data from the Department of Energy. We will explore historical census data in Chapter 12.
- Our lab advised the New York City Comptroller's Office in creating a series of statistical profiles on the economy of each of the city's Community Districts (New York City Comptroller's Office, 2017). It took data from the ZBP and summed it to ZIP Code Tabulation Areas so the data could be related and assigned to the districts. They collated the ZBP, ACS, and decennial census data into a concise and attractive report and web-based interface. We will cover census geography in Chapter 3, business datasets in Chapter 8, and creating derivatives and relating different geographies in Chapter 11.
- As part of a workshop I teach on GIS, I incorporate an example where we use demographic data from the census, TIGER geographic boundary files, and other geographic data such as the location of subway stations and coffee shops to identify possible locations for opening a new neighborhood coffee shop. We will cover GIS and the Bureau's geographic products in Chapter 10.

Role of the Census in American Society

In this section, we will consider how census data and the Census Bureau fit within the context of American society. In doing so, we will also touch on various aspects of the Census Bureau's history. For a fuller historical treatment, Margo Anderson's *The American Census: A Social History* (2015) is a definitive account, and the history portion of the Census Bureau's website at https://www.census.gov/history/ is quite comprehensive.

The census has played a vital role in American democracy since the country's founding. The United States was the first country to institute a population census for the purpose of assigning representatives to a de-

mocratically elected legislature (Emigh, Riley, & Ahmed, 2016a). Article I, Section 2, of the U.S. Constitution provides the original, legal basis for the census:

> Representatives and direct Taxes shall be apportioned among the several states which may be included in this Union, according to their respective Numbers. … The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct.

The first census was taken as stipulated in 1790 and has been conducted every 10 years since. The decennial census is used to reapportion seats in Congress based on the differential growth in population between the states, and the data is subsequently used to redraw legislative districts in states that either gained or lost seats. Other provisions in the Constitution provide justification for a federal statistical system. Article I, Section 9, requires that federal appropriations have legal authorization and that the government provides regular statements of its accounts, and Article II, Section 3, stipulates that the president must give Congress an annual update on the state of the union. Statistics were seen as one method for fulfilling these obligations and measuring the nation's progress (Anderson, 2010).

Outside the Constitution, a large body of federal law requires that specific census variables be collected. The statutory uses of each variable that will be collected in the 2020 census and the future iterations of the ACS are published in a report that the Census Bureau submitted to Congress (U.S. Census Bureau, 2017h). For example, the Civil Rights and Voting Rights Acts require data on age, sex, race, employment, and disabilities to evaluate whether civil rights are being protected. Census data is used for allocating hundreds of billions of dollars of funding for federal assistance programs to states and local governments, such as Medicaid, Highway Planning and Construction, Title I Education grants, Temporary Aid for Needy Families, Community Development Block Grants, and more (U.S. Government Accountability Office, 2009). In fiscal year 2016, approximately 320 federal programs used census data to allocate more than $880 billion in federal funds, primarily to state and local governments but also to individuals through direct assistance programs (Reamer, 2017, 2018).

The original decennial census was conducted by U.S. Marshalls, who fanned out across the country on horseback and counted people based on instructions from Congress. As the country grew in size and population and demands for census data increased, the mechanisms for collecting, tabulating, and presenting data grew in complexity and sophistication to meet the demands. Beginning in 1850, a temporary Census Office was established prior to each census to direct operations and tabulate the results (Anderson, 2015, pp. 41–58),

and by 1880, this office, staffed with professional statisticians, took control over all census operations (Anderson, 2015, pp. 89–101). In 1902, Congress established the Census Bureau as a permanent office under the Department of Commerce and Labor that remained in operation year-round, and it became one of the chief statistical agencies within the expanding federal government. The number of questions on the census grew from the mid-19th to early 20th century at the instigation of stakeholder groups that included professional statistical societies and business interests. From the mid- to late 20th century, the needs of the federal government for allocating funding and directing policy became the driving force behind the addition and standardization of questions.

Placed squarely in the middle of America's political mechanisms, the census is a strongly debated and contested issue. In their two-volume, comparative, historical study of censuses in the United Kingdom, the United States, and Italy, Emigh et al. (2016a, 2016b) conclude that this intense interaction around the U.S. Census ensures that it remains a vibrant social institution, whereas in other countries, population counting is seen as either a bureaucratic or symbolic exercise since it is disconnected from political outcomes. In the United States, there are fierce debates and lawsuits over how the census is conducted, what questions are asked, how categories for race and ethnicity are defined, and whether the census is accurate or not. Undercounting specific areas or population groups can result in the loss of federal aid and political power for these places or groups.

There are two peculiarities of American government vis-à-vis the census that help ensure that it remains relevant to society at large. First, since it is stipulated as part of the Constitution, it is reasonably assured that the census will be conducted every 10 years in a manner that's relatively consistent. Many other countries have abandoned doing actual counts of the population in favor of using annual sample surveys, estimates based on administrative data, or population registers (Baffour, King, & Valente, 2013). Following years of controversy and lawsuits against the Census Bureau concerning population undercounts in the 1980 and 1990 census, the Supreme Court ruled that the decennial census must be an actual count: It cannot be based on a sample and the count cannot be adjusted using statistical means in any way (Anderson, 2015, pp. 228–247). The Census Bureau can create and adopt new innovations like the ACS, but it must continue to do an actual 10-year count for apportionment purposes.

Second, according to U.S. copyright law (Title 17, Section 105), all works produced by the U.S. government, with few exceptions, are automatically released into the public domain and can be used by anyone for any purpose. This makes the census widely available and accessible, and throughout history, public, private, aca-

demic, and nonprofit agencies have employed census data for their own purposes. Stakeholder and interest groups throughout society lobby for changes in the census to meet their needs and also lobby on behalf of the Census Bureau to keep programs funded. In contrast, many other countries copyright their census data and limit what is available. With few exceptions, the United States is rather unique for having a large and established ecosystem of census data users.

Given the accessibility and quality of the census, it is employed for many purposes throughout American society. The Economics and Statistics Administration published a study on the value of the ACS with the subtitle: *Smart Government, Competitive Businesses, and Informed Citizens* (2015) that contains detailed statistics and vignettes on how census data is used. Examples include Kroger supermarkets creating sales projections and siting new stores and academic researchers and governments in Hawaii creating disaster management plans to cope with volcanic eruptions and lava flows.

State and local governments have always been heavy users of census data, because they can use it to study their own communities and create appropriate policies for urban and regional planning, social assistance, public services, and economic development. Within academia, the census serves as a foundational dataset within the social sciences. Compared with many other datasets, the census is geographically detailed, rich in its breadth of variables, relatively consistent, longitudinal, relatively accurate, and well documented. In academic research, the census is used to provide context and a frame of reference for describing places and population groups, can be used as criteria for selecting areas of study, can help define sampling strategies, and serves as the basis for new and derived estimates (Dickason, 2012). For example, census data is particularly important in the field of public health, where it is used as the basis for studying different populations in relation to risk and exposure to public health threats (Wilson et al., 2017).

In the private sector, there are two types of business that routinely use census data: (1) those who use it to make business decisions and (2) information brokers who use it as a commodity. Census data is used extensively within the fields of marketing and real estate. Marketers use it to identify populations and areas that would be good targets for their products and services, while real estate agents use census data to supplement their own information in order to understand housing markets and characterize neighborhoods. Information brokers gather data from many different sources, aggregate it, and use it to produce intelligence that they can sell to third parties, while others create web-based products that can be used for doing research.

Census data plays a key role within American representative democracy for apportioning political power and the resources of the U.S. government. Over time, it has become a vital piece of the nation’s infrastructure that

is similar to other public goods and services, in that it provides a piece of the foundation on which the country's society and economy rest through the basic yet essential information it provides. Based on the definition of public goods as described by political philosopher Angela Kallhoff in her book *Why Democracy Needs Public Goods* (2011), census data fits the definition as it is nonrival (each individual can use it without affecting someone else's use) and nonexcludable (it is free for all to share). Census data helps generate a public sphere by providing reliable information that creates mutual awareness of others in our society, and it serves as a focal point for debate over issues of common concern.

Criticism of the Census

The census is not without flaws or critics. In this section, we'll summarize some of the philosophical and political objections to the census (we will cover issues related to methodology such as undercounting and sample sizes as we discuss each dataset in later chapters). The first and certainly the most earnest concern is the confidentiality and privacy of an individual's responses to the census questionnaires. Throughout the 20th century, federal law stipulated that the Census Bureau would not publish or share records of individual responses to census questionnaires (U.S. Census Bureau, 2009). The current law, established in the 1950s and amended in the 1970s, prohibits the disclosure of individual-level census information for 72 years from the date the census was conducted. Despite these laws, there is a general and growing suspicion of government surveillance (fanned by controversies such as the National Security Agency's PRISM system) and government data-gathering programs from all sides of the political spectrum. Census confidentiality statues were rescinded during the First and Second World Wars under emergency security measures (Anderson, 2015; Aratani, 2018), so there are concerns that it could happen again given some future emergency.

While there is reason for concern, it is important to consider the environment in which the Census Bureau operates. The federal government is composed of hundreds of agencies that operate according to their own missions, needs, and interests and that compete for resources. What's in the best interest of one agency may not be in the best interest of another. The Census Bureau's goal is to create the most accurate population statistics that it possibly can. To achieve this, it must establish a high level of trust with the American people and ensure that each individual's responses will be held in confidence according to the law. Therefore, it is not in the Census Bureau's best interest to share information with other government agencies as it will erode the public's trust and jeopardize the accuracy of the statistics, if people refuse to respond out of fear for their

privacy.

In Margo Anderson's (2015) account of the Census Bureau's history, she describes how the Bureau fought to maintain its independence within the federal statistical system so that it could fulfill its mission of generating accurate statistics. In particular, the Bureau successfully resisted every attempt to tie its statistical-gathering activities to other branches of the government that specialized in regulatory enforcement, so that it could reassure individuals and businesses that their data would be used only for the purpose of generating summary statistics.

The Bureau continues with this struggle today. In late 2017 and early 2018, the Justice Department lobbied the Census Bureau to include a question on citizenship on the 2020 census form, which they deemed necessary for upholding the Voting Rights Act and fighting voter fraud (Baumgaertner, 2018; H. L. Wang, 2018h). Given the bitter partisan debates over immigration and the uncertainty and fear among immigrant groups (both legal and undocumented) about their status, the Census Bureau and its supporters (including the two previous secretaries of Commerce under the Obama and Bush administrations; Pritzker & Gutierrez, 2018) strenuously objected to this suggestion. All residents, regardless of their status, are counted in the decennial census. Given deepening suspicion of the government's motives, it's likely that many would refuse to participate and thus would jeopardize the accuracy of the count and all the programs that depend on it. In June 2019 the Supreme Court ruled against the addition of a citizenship question to the 2020 census.

Some members of Congress have suggested that since the census is used for apportioning seats in Congress, either people who can legally vote or only U.S. citizens should be counted. This would be contrary to the intentions of the Founding Fathers and the 220-year history of the census, which has always counted every single person as it was deemed to be the simplest and fairest method for conducting the count. Children cannot vote, but there are approximately 74 million children in the United States and they depend on basic government services like schools. Legal permanent residents cannot vote and are not citizens, but they pay taxes and contribute to society. Politicians are elected to represent all members of their districts, and the Supreme Court agrees. The Court reconsidered the practice of counting every person as opposed to counting eligible voters during the drafting of the Fourteenth Amendment after the Civil War and decided to uphold the population count as the simplest and fairest approach. Since then, the Court has upheld this opinion on several occasions, most recently in 2015, when they ruled that states may count all residents, regardless of whether they are eligible to vote, when drawing legislative districts (Liptak, 2016).

Beyond the issue of confidentiality is a simpler issue of personal privacy that can be summarized as follows:

"Why is the government asking me all these questions? It's none of their business!" In this view, the Constitution says that there must be a 10-year count and says nothing about asking other questions or running additional surveys like the ACS. Therefore, some believe that most of the questions are unconstitutional and people have a right to refuse to answer them. However, as discussed in the previous section, there are several sections of the Constitution that provide a basis for establishing a federal statistical system. There are also federal laws and court decisions that require the government to collect statistics in order to fulfill many obligations. The Census Bureau cannot ask questions simply because they might be novel or interesting; every single question is asked because it has some basis in the law.

The Census Bureau explicitly ties each question to the law that requires it and presents this information to Congress (U.S. Census Bureau, 2017h).

In terms of privacy, the image of the government as a 1984 Big Brother that's gathering information about every citizen through coercion seems less plausible given life in the early 21st century. Every day, millions of Americans freely share information about themselves (knowingly and unknowingly) on social media and the internet that is infinitely more personal and potentially compromising than anything they share on a census form. This information is held by technology companies, credit agencies, and data brokers, many of whom sell it to third parties. By and large, these companies are completely unaccountable, and we cannot even know or request what data has been collected about us (Kitchin, 2014). Concerns about the Census Bureau's collection of basic demographic information seems minor in comparison.

The economics of the census is another issue that's frequently raised by fiscal conservatives. The census has been criticized as a waste of tax payer dollars, and it has been suggested that the private sector could do a better job. In reality, the federal statistical agencies' share of federal budget resources represented about 0.04% of gross domestic product in 2016 (Executive Office of the President, 2017), a trifling amount compared with the budgets for defense, Social Security, and Medicare. The private sector cannot compel people to fill out census forms and could not possibly conduct a count or survey of the same scope and detail. Businesses rely on census data the same way they rely on other public goods, such as roads, mass transit, and schools, as fundamental pieces of infrastructure that the economy is built on. When census programs are threatened by budget cuts, business leaders and trade groups are among the first to lobby against them. For example, when the American Community Survey was threatened with cuts in 2012, the Target Corporation collaborated with the Census Bureau to produce a YouTube video that showcased how Target uses census data to tailor its stores and products to different markets (U.S. Census Bureau, 2012b). Many fiscal critics fail

to measure the cost of the census against the benefits that it provides (Wilson et al., 2017).

There are good reasons to scrutinize the census. It is important to debate the census questions and categories to ensure that they reflect the changing nature, interests, and needs of our society. It is necessary to highlight issues with methodology that could result in unforeseen consequences regarding the accuracy of statistics. Given the creeping amount of surveillance in our society and the growing number of data breaches, confidentiality must be of utmost concern. But like every other political or public policy issue, it is important to study the underlying arguments to determine whether they are rooted in facts or opinions, either informed or uninformed. The ability to have reliable information for the purpose of checking facts is one of the reasons why we create census data to begin with.

The Census within the Data Universe

Where does the census fit into our data-saturated world? In this section, we'll situate census data within this context. The census datasets exist as part of a larger federal system of data collecting and publishing activities. The census can also be considered as part of the growing open data movement with some caveats, while it is largely distinct and separate from what most people think of as big data.

The Federal Statistical System

The census is part of the U.S. Federal Statistical System, whose mission is to provide evidence-building functions, which the government describes as "the collection, compilation, processing, analysis, and dissemination of data to create general purpose, policy- and program-specific, and research-oriented statistics and datasets. They also include program evaluation, performance measurement, and public health surveillance" (Executive Office of the President, 2017). The Census Bureau is one of the 13 principal statistical agencies whose primary mission is the production and analysis of statistics, and it receives the largest share of the statistical program's budget ($1.4 billion out of $7.2 billion in 2016).

Given the Census Bureau's size and the depth and breadth of its knowledge for creating statistics, it supports many other federal and state agencies in gathering and creating data. It has a long history of innovation in this field. In the late 19th century, it pioneered the use of mechanical punch card technology for tabulating data (Figure 1.3 shows women reading entries from 1940 census enumerator forms to create punch cards, which

would be fed into machines to tabulate results). The Census Bureau was in the forefront of developing statistical sampling methods in the 1930s, which were envisioned as efficient ways for collecting timely data on an ongoing basis (Anderson, 2015, pp. 176–179). Sample survey methods and the Bureau's early adoption of digital computer technology in the mid-20th century allowed for the expansion and growth of data collection and tabulation. In the late 20th century, the Census Bureau helped spread the adoption of GIS in the United States through the creation and distribution of TIGER, a database of geographic boundary files. They were also one of the earliest agencies to publish data on the internet.

FIGURE 1.3 CARD PUNCH OPERATORS CREATING POPULATION CARDS FOR THE 1940 CENSUS

Source: National Archives https://catalog.archives.gov/id/7741405

Many of the statistical agencies specialize in providing data that is specific to their departments and missions. The Bureau of Transportation Statistics focuses on transportation-related data while the National Center for Education Statistics focuses on education-specific data. The Census Bureau is unique in that its datasets appeal to a broad range of fields and interests; there are questions on commuting, education, the labor force, disabilities, and housing. It is also distinct in its ability to provide detailed data for small geographies that is uniform and comparative; almost all of the other datasets published by the government are not tabulated below the state or county level (Hartnett, Sevetson, & Forte, 2016).

Datasets can be classified into three categories based on how they are created: (1) statistical, (2) administrative, and (3) derived. Most of the Census Bureau's datasets are statistical datasets, while the majority of the other agencies produce data from administrative sources. Statistical datasets are created for the specific purpose of having data to answer specific questions. Statistical datasets can be generated from a total count, like the decennial census, or from sample surveys, like the ACS. In contrast, administrative datasets are created as a by-product of some process. For example, the primary function of the (IRS) is to collect taxes to raise revenue for the federal government, and it uses forms like the 1040 to gather this information. The purpose is to collect taxes, not to produce data. As a by-product, the IRS creates datasets that are used to measure migration between states and counties, based on whether a person's address changed from one year to the next. Derived datasets are data created from other data. The Census Population Estimates Program is a derived dataset that uses data from the decennial census, the ACS, the IRS, the Medicare Enrollment Program, and the National Center for Health Statistics to estimate the annual population for states, counties, and metropolitan areas.

Open Data

Over the past decade, there has been increasing interest around the concept of open data. In the most basic sense, data is considered open if it's free to use, reuse, and redistribute with minimal requirements. The open

data movement seeks to build collaboration and participation around free datasets that can be used to study and improve public services and to spur economic growth (Goldtstein & Dyson, 2013). Open data should meet a number of technical requirements that are intended to ensure that it is as accessible as possible. Data should be complete, primary (not summarized), timely, well-documented, and machine readable in nonproprietary data formats, so it can be easily processed without restrictions or expensive tools (Kitchin, 2014).

In many ways, the census could be considered as the original open dataset, as it has been well-documented, widely distributed, and publicly available since its inception in 1790. It falls within the public domain and can be used by anyone for any purpose. It is highly accessible as it can be discovered via several different web interfaces. The data is stored in machine-readable formats (CSV, text-delimited, spreadsheets, database tables, XML), which are formats that have a suitable organization and structure that allows data to be directly retrieved and manipulated. Given the number of datasets and their size and complexity, search engines cannot always crawl and index them directly, but in many instances, users can get persistent URLs to tables so that data can be linked to and cited. It is well-structured and indexable; every record represents a piece of American geography, and that geography is assigned a unique ID code (called a GEOID) that is relatively consistent within and across datasets and years. Census data stored in separate tables can be related and tied together using these identifiers.

One of the challenges for both description and accessibility is the sheer size and complexity of the datasets, which makes them confusing for new and even seasoned users to understand and navigate. The Census Bureau invests a lot of effort in providing tools to cater to many users, and the process for creating datasets is transparent. Each of the Bureau's data discovery tools includes links to glossaries with definitions and terminology, and each of the individual statistical program websites (the decennial census, the ACS, the Economic Census, etc.) includes detailed and frequently updated information that describes the methodology used for collecting and processing the data.

The census is not "complete" in the sense that it's not a primary or secondary dataset. Primary data is data that's collected by an individual or organization for their own use, secondary data is primary data that's distributed for use by outside researchers for their own purposes, and tertiary data is derived data: data that's been aggregated and summarized from the primary set. The Census Bureau's primary data, records of individual responses to census questionnaires, is subject to confidentiality regulations that protect individual's privacy. An individual's responses to the census are kept confidential for 72 years before they can be released, and

until then, the data cannot be shared with anyone, including other branches of the government. The Census Bureau provides samples of individual responses with personal identifying information removed in public use microdata files, but not the complete datasets. Most of the data is summarized by population groups and geographic areas, some of which are quite small in size.

Whether the census is timely is a matter of opinion. It is more timely than it used to be, as the detailed socioeconomic characteristics of the population are provided annually as part of the ACS, rather than just every 10 years with the decennial census. The Population Estimates Program and the County Business Patterns are published on an annual basis, and the Current Population Survey is published monthly. But in the big data world where data is provided in real time, the census is not considered timely. It is published at set intervals, and there is a time lag from the time the data is collected to the time it is processed and released.

Given their transparency, accessibility, documentation, structure, and geographic detail, the census datasets do serve as foundational layers in the open data universe. From the open data perspective, they cannot be considered primary or timely, and on these points, we can contrast the census with “big” datasets.

Big Data

In the colloquial sense, the census is quite large, but in the technical sense, it is not big data. Big data is captured in real time and has a granular level of detail, representing a specific person or event at a specific geographic location. Cameras and sensors that constantly monitor the environment are capturing big data, as are websites monitoring clicks, social media sites registering every comment and post, and online forms like 311 requests that are capturing individual complaints. In his book *The Data Revolution*, Rob Kitchin (2014) contrasts big data with small data, and he characterizes the latter as traditional datasets that are produced in a tightly controlled manner with limited scope, size, and time frame. Census data is a prime example of a small dataset: The size and scope are limited to a specific number of questions: if the data is sample based, it is limited to a certain number of respondents, and the time frame ranges from 1 to 10 years. Because of confidentiality reasons, the data is often coarse, summarized by groups and by places. The design is also inflexible; once a count or survey begins, the methods cannot be changed without compromising the dataset and generating great expense.

In contrast, big data seeks to be exhaustive and finely detailed, and is flexible and scalable in production.

Kitchin describes big data as being high in velocity (the speed in which it is produced) and volume (the sheer amount that is produced). The allure of big data is the notion that we can simply collect as much information as possible and analyze it in the hope of uncovering trends and making connections and predictions that were previously impossible to conceive without access to modern resources like machine learning and infinite disk space. So why bother with small data like the census when we have big data?

Big data captures what's easy to capture and whatever is openly expressed. It is often represented at face value by technology enthusiasts, even though the data is often not designed to answer specific research questions it's being applied to and is often dirty or unprocessed. While limited in volume and velocity, small data has a long development history with established practices and a design that seeks to answer specific research questions. Kitchin uses the analogy of gold mining; small data studies look for gold in narrow seams while big data studies attempt to extract nuggets from large-scale open pit mining. The principal difference is in investing resources to collect data to answer specific, targeted questions versus searching through tons of big data and hoping it tells us something (Kitchin, 2014). Ultimately, big data has the same limitations as small data; it is merely a representation of reality that is influenced and biased by the context in which it's created. Despite the hype generated around big data, it is not the objective, exhaustive, perfect, and sole solution for answering all of the world's questions.

Let's look at two recent examples that illustrate the limitations of big data versus the value of census data. In 2017, Facebook was touting its strengths to advertisers and investors by saying that its social media platform had the ability to reach 41 million adults in the United States between the ages of 18 and 24, and 60 million adults between the ages of 25 and 34. This sounded pretty impressive, until an analyst at an investment firm checked Facebook's numbers against the latest census data and found that there are only 31 million adults in the United States aged 18 to 24 and 35 million aged 25 to 34 (Hem, 2017; Swant, 2017). When presented with these discrepancies, Facebook responded that their data was designed to estimate how many people in a given area are eligible to see an advertisement that a business might run. Their estimates were derived based on Facebook user behavior, user demographics, and location data from devices and were not designed to match population or census estimates. They concluded by saying that they are always working to improve their estimates.

There are three lessons we can draw from this example. First, big data suffers from the same limitations as small data. The Facebook data was never cleaned or processed to estimate the actual population; it was sim-

ply taken at face value and accepted as is. Their statistics can be inflated because people misrepresent their age, have multiple or fake accounts, and because the location services capture people who are in a given area but don't necessarily live there, such as tourists. The census suffers from these same issues; people can be undercounted or overcounted, and there are challenges determining what a person's residence is (people with vacation homes, military personnel overseas, the homeless). The Census Bureau has an advantage relative to Facebook regarding data accuracy, as a person has more to gain by having multiple accounts or lying about their age on a social media platform versus a government form that most people fill out once every 10 years. More important, the Census Bureau has experience with addressing these issues and has methodologies for coping with missing and possibly false information.

This leads to the second lesson. Because the census is in the public domain, the process is transparent. We can go on the census website and freely access the data and all the documentation that's associated with collecting, processing, and disseminating it. People who disagree with the data can lobby the Census Bureau and Congress to try to force changes. While this might be difficult to achieve, you still have the right to do it and have the information at your disposal to create meaningful arguments. In sharp contrast, the Facebook data is in a black box. We can only guess and make assumptions about how it is created. While tech companies constantly push and pull us to freely give them data about ourselves, they resist any attempt to share the data they collect about us with us, or even disclose what they do with it. Many of the big datasets, especially the data generated from social media, suffer from this lack of transparency.

Last, this example illustrates one of the important use cases of census data: The census can serve as a baseline that we can check other datasets against. It can be used for fact checking, as the analyst used it in checking Facebook's claims, and for benchmarking, calibrating, and adjusting population estimates that are generated from other sources, which is what Facebook failed to do.

Meanwhile, *The Washington Post* reported in 2017 that scientists were now able to estimate what the demographic characteristics of different neighborhoods were based on the cars that are parked in the neighborhood (Ingraham, 2017). The researchers collected Google Street View images from 200 U.S. cities, created a schema that correlated the makes and models of thousands of cars with cars in the images, and used this data to build a model that predicts the race, income, and voting characteristics of the population in small census areas. The researchers compared their findings with the ACS data and found a high correlation between their estimates and the actual census data. They suggested that their method could be used to provide more

data that is just as accurate as the ACS but could do so in a timely fashion at a fraction of the cost (Gebru et al., 2017).

There are a number of lessons that can be drawn here. Like many experiments that take place in Silicon Valley, the results are novel and interesting, but the exercise takes place in a moral vacuum. Instead of asking a person to identify themselves on a government census form with information that describes how the data will be used, a private company takes pictures of a neighborhood and a third party uses this information to estimate who lives there. Is it just or fair to estimate what a neighborhood's population is like by photographing the cars parked on its streets? The ACS is used to allocate federal funds for everything from transportation projects, to programs for schools, to assistance for needy families. Would it be ethical to use the car-based data to allocate this money, instead of the ACS? Or what if another third party uses this data instead of the ACS to make decisions on whom to give a home loan? While the researchers never explicitly claim that their method should be used to replace the ACS, they implicitly point in this direction as they emphasize how expensive and untimely the ACS data is.

As part of its mission, one of the Census Bureau's goals is to ensure that everyone in the United States is counted as part of the decennial census and that a representative sample of the entire population is included in all their sample-based products. Furthermore, the categories that are used for tabulating the data must represent the entire U.S. population, and the census data itself must be accessible to everyone as a public good.

This mission cannot be fulfilled by the private interests in any one of the examples just described. Researchers in the car study state that they were unable to reliably estimate the presence of children or people employed as farmers, as children don't drive cars and the study omitted rural areas. In the Facebook example, even though a large percentage of the population uses Facebook, there are groups of people that tend to use social media less than others. In the United States, about 7 in 10 Americans used social media in 2018, and people who were older, lower income, or living in rural areas used it less than people who were younger, higher income, and urban (Pew Research Center, 2018). In essence, the big datasets that seek to be exhaustive are not truly exhaustive, but suffer from selection bias based on their context and, in some cases, by conscious decisions made by the people who shape how the data is created.

This does not mean that big data should be dismissed entirely, but it should not be considered as a holy grail. "Small" datasets like the census continue to play a valuable role as high-quality open datasets that are

designed to answer targeted questions regarding the demographic and socioeconomic characteristics of the United States. As a public good, the census is transparent, accessible, representative of the entire population, and accountable to the public in ways that private or proprietary datasets cannot be.

Conclusion and Next Steps

This chapter was designed to give you an overview of the census, so you can understand its legal justification, see the various roles it plays in U.S. society, and place it within the context of a broader data universe. The rest of this book is devoted to teaching you the practical concerns of understanding, finding, retrieving, processing, analyzing, and interpreting census data. As we address these concerns and learn about the different geographies, subjects, and datasets, we will touch on some of the contextual and ethical issues that we covered in this chapter. While many of our concerns will seem practical (How do I represent these racial categories? How can I combine these areas to study a neighborhood? What threshold should I use for establishing some criteria?), decisions made in creating and using data will always have social, political, economic, or ethical consequences.

In the next chapter, we'll get moving right away: We'll go directly to the main source for census data and start exploring the different datasets and tables, and then we'll step back in Chapters 3 and 4 to understand how this data is summarized and organized geographically and categorically.

- censuses
- population
- surveying
- Bureau of the Census

- Census data
- Survey research
- Populations

https://doi.org/10.4135/9781071909645

References

Abrams, R., & Gebeloff, R. (2018, May 3). A fast-food problem: Where have all the teenagers gone? *The New York Times*. Retrieved from https://nyti.ms/2FCKvXL.

Albrecht, J. (2007). *Key concepts and techniques in GIS*. Thousand Oaks, CA: SAGE.

Anderson, M. J. (2010). The census and the federal statistical system: Historical perspectives. *The ANNALS of the American Academy of Political and Social Science*, *631*, 152–162.

Anderson, M. J. (2015). *The American Census: A social history* (2nd ed.). New Haven, CT: Yale University Press.

Aoyama, Y., Murphy, J. T., & Hanson, S. (2011). *Key conecpts in economic geography*. Thousand Oaks, CA: SAGE.

Apparicio, P., Abdelmajid, M., Riva, M., & Shearmur, R. (2008). Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics*, *7*(7), 1–14.

Aratani, L. (2018, April 3). Secret use of census info helped send Japanese Americans to internment camps in WWII. *The Washington Post*. Retrieved from https://wapo.st/2IqLZWT?tid=ss_tw&utm_term=.36da085f7e30.

Arias, E., Heron, M., & Xu, J. (2017, August). *United States life tables, 2014 (Vol. 66; National Vital Statistics Reports No. 4)*. Washington, DC: National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/products/nvsr.htm.

Baffour, B., King, T., & Valente, P. (2013). The modern census: Evolution, examples, and evaluation. *International Statistical Review*, *81*(3), 407–425.

Bauder, J. (2014). *The reference guide to data sources*. Chicago, IL: American Library Association.

Baumgaertner, E. (2018, March 26). Despite concerns, census will ask respondents if they are US citizens. *The New York Times*. Retrieved from https://nyti.ms/2pN40b5.

Bazuin, J. T., & Fraser, J. C. (2013). How the ACS gets it wrong: The story of the American Community Survey and a small, inner city neighborhood. *Applied Geography*, *45*, 292–302.

Beaulac, J., Kristjansson, E., & Cummins, S. (2009). A systematic review of food deserts, 1966–2007. *Preventing Chronic Disease*, *6*(3), Article 105. Retrieved from https://www.cdc.gov/pcd/issues/2009/jul/08_0163.htm.

Bellu, L. G., & Liberati, P. (2006, December). *Inequality analysis: The GINI coefficient (EASYPol Analytical Tools, Module 040)*. Rome, Italy: Food and Agriculture Organization of the United Nations. Retrieved from http://www.fao.org/3/a-am352e.pdf.

Bishaw, A., & Benson, C. (2017, September). *Poverty: 2015 and 2016 (American Community Survey Briefs: ACSBR/16–01)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/content/census/en/library/publications/2017/acs/acsbr16-01.html.

Boburg, S. (2017, May 31). How Jared Kushner built a luxury skyscraper using loans meant for job-starved areas. *The Washington Post*, p. A01. Retrieved from http://wapo.st/2qGLDSz.

Boscoe, F. P., Henry, K. A., & Zdeb, M. S. (2012). A nationwide comparison of driving distance versus straight-line distance to hospitals. *The Professional Geographer*, *64*(2), 188–196.

Brennan Center for Justice. (2017, April 17). *Who draws the maps? Legislative and Congressional redistricting (Tech. Rep.)*. New York, NY: Author. Retrieved from http://www.brennancenter.org/analysis/who-draws-maps-states-redrawing-congressional-and-state-district-lines.

Burnett, K. D. (2011, November). *Congressional apportionment (2010 Census Briefs: C2010BR-08)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/publications/2011/dec/c2010br-08.html.

California State Data Center. (2016, April). *Re-calculating medians and their margin of errors for aggregated ACS data (January 2011 Network News, revised 2016)*. Sacramento: California State Data Center, Demographic Research Unit. Retrieved from http://www.dof.ca.gov/Forecasting/Demographics/Census_Data_Center_Network/.

Chapman, J. (2017, September). *Migration in the Washington region: Trends between 2000 and 2015 and characteristics of recent migrants (Tech. Rep.)*. Arlington, VA: The Stephen S. Fuller Institute, Schar School of Policy and Government, George Mason University. Retrieved from http://sfullerinstitute.gmu.edu/research/reports/migration-washington-region/.

Charreire, H., Casey, R., Salze, P., Simon, C., Chaix, B., Banos, A.,… Oppert, J.-M. (2009). Measuring the food environment using geographical information systems: A methodological review. *Public Health Nutrition*, *13*(11), 1773–1785.

Connerly, W., & Gonzalez, M. (2018, January 3). It's time the Census Bureau stops dividing America. *The Washington Post*. Retrieved from http://wapo.st/2CCT8Fa?tid=ss_tw&utm_term=.35a9257598b2.

Cook, T. M., Norwood, J. L., & Cork, D. L. (2011). *Change and the 2020 Census: Not whether but how*. Washington, DC: National Academies Press.

Coontz, S. (2010). Why American families need the census. *The ANNALS of the American Academy of Political and Social Science*, *631*, 141–149.

Cope, M., & Elwood, S. (2009). *Qualitative GIS: A mixed methods approach*. Thousand Oaks, CA: SAGE.

Cunningham, J. (2015, September). Measuring wage inequality within and across US metropolitan areas, 2003–13. *Monthly Labor Review*. Retrieved from https://doi.org/10.21916/mlr.2015.35.

D'Antonio, P., & Whelan, J. C. (2009). Counting nurses: The power of historical census data. *Journal of Clinical Nursing*, *18*(19), 2717–2724.

Desmond, M. (2016). *Evicted: Poverty and profit in the American city*. New York, NY: Broadway Books.

Dickason, D. G. (2012). Census of population; "Panning for gold." In J. P. Stoltman (Ed.), *21st century geography: A reference handbook* (pp. 267–276). Thousand Oaks: SAGE.

Donnelly, F., Clark, A., & Billadello, J. (2018, February). *New Yorkers on the move: Recent migration trends for the city and metro area* (WCIB Occasional Papers Series No. 15). New York, NY: Weissman Center for International Business, Zicklin School of Business, Baruch College, City University of New York. Retrieved from https://zicklin.baruch.cuny.edu/faculty-research/centers-institutes/weissman-center-international-business/occasional-paper-series/.

Economics and Statistics Administration. (2015, April). *The value of the American Community Survey: Smart government, competitive businesses, and informed citizens (Tech. Rep.)*. Washington, DC: U.S. Department of Commerce. Retrieved from https://www.commerce.gov/news/reports/2015/05/value-american-community-survey-smart-government-competitive-businesses-and.

Eichel, L., & Schmitt, M. (2016, July). *A portrait of Philadelphia's migration: Who is coming to the city—and who is leaving (Philadelphia Research Initiative Issue Brief)*. Philadelphia, PA: Pew Charitable Trust. Retrieved from http://pew.org/29OhwCT.

Emigh, R. J., Riley, D., & Ahmed, P. (2016a). *Antecedents of censuses from medieval to nation states: How societies and states count*. New York, NY: Palgrave Macmillan.

Emigh, R. J., Riley, D., & Ahmed, P. (2016b). *Changes in censuses from imperialist to welfare states: How societies and states count*. New York, NY: Palgrave Macmillan.

Evergreen, S. D. (2017). *Effective data visualization: The right chart for the right data*. Thousand Oaks, CA: SAGE.

Executive Office of the President. (2017, January). *Statistical programs of the United States government fiscal year 2017 (Tech. Rep.)*. Washington, DC: Office of Management and Budget. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2017.pdf.

Fox, L. (2017, September). *The supplemental poverty measure: 2016 (Current Population Reports: P60-261 (RV))*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/publications/2017/demo/p60-259.html.

Frey, W. H. (2018a). *Diversity explosion: How new racial demographics are remaking America* (2nd ed.). Washington, DC: Brookings Institution Press.

Frey, W. H. (2018b, March). *US population disperses to suburbs, exurbs, rural areas, and "middle of the country" metros (The Avenue)*. Washington, DC: Brookings Institution Press. Retrieved from https://brook.gs/2l747ot.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(50), 13108–13113. Retrieved from http://www.pnas.org/content/114/50/13108.

Glassman, B. (2016, July). *Income inequality metrics and economic well-being in US metropolitan statistical areas (SEHSD Working Paper 2016–19)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/working-papers/2016/demo/SEHSD-WP2016-19.html.

Glenn, E. H. (2011). acs.R: An R package for neighborhood-level data from the U.S. census. *Computers in Urban Planning and Urban Management Conference*, 1–7.

Goldtstein, B., & Dyson, L. (2013). *Beyond transparency: Open data and the future of civic innovation*. San Francisco, CA: Code for America.

Graser, A. (2016). *Learning QGIS* (3rd ed.). Birmingham, UK: Packt.

Graser, A., & Peterson, G. N. (2016). *QGIS map design*. Chugiak, AK: Locate Press.

Gries, P., Campbell, J., & Montojo, J. (2013). *Practical programming: An introduction to computer science using Python 3* (2nd ed.). Dallas, TX: Pragmatic Bookshelf.

Grubesic, T. H. (2008). ZIP codes and spatial analysis: Problems and prospects. *Socio-Economic Planning Sciences*, *42*, 129–149.

Guzman, G. G. (2017, September). *Household income: 2016 (American Community Survey Briefs: ACSBR/16–02)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/publications/2017/acs/acsbr16-02.html.

Hartnett, C. J., Sevetson, A. L., & Forte, E. J. (2016). *Fundamentals of government information* (2nd ed.). Chicago, IL: Neal-Schuman.

Harvey, F. (2015). *A primer of GIS: Fundametal geographic and cartographic concepts* (2nd ed.). New York, NY: Guilford Press.

Hayslett, M., & Kellam, L. (2010). The American Community Survey: Benefits and challenges. *IASSIST Quarterly*, Winter/Spring, 31–39.

Hayter, C. T. (2015). Adding a little ZIP to the mail: The development and growth of the ZIP code. *Documents to the People*, *33*(3), 43–51.

Heimel, S. (2014, October). *American Community Survey fiscal year 2014 content review: Median county-level coefficients of variation of key estimates from ACS data (Final Report)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/programs-surveys/acs/operations-and-administration/2014-content-review/methods-and-results.html.

Hem, A. (2017, September 7). Facebook claims it can reach more young people than exist in UK, US and other countries. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/sep/07/facebook-claims-it-can-reach-more-people-than-actually-exist-in-uk-us-and-other-countries.

Herman, E. (2008). The American Community Survey: An introduction to the basics. *Government Information Quarterly*, *25*(2008),504–519.

Hewko, J., Smoyer-Tomic, K. E., & Hodgson, M. J. (2002). Measuring neighbourhood spatial accessibility to urban amenities: Does aggregation error matter? *Environment and Planning A*, *34*, 1185–1206.

Hobbs, F., & Stoops, N. (2002, November). *Demographic trends in the 20th century (Census 2000 Special Reports: CENSR-4).* Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/prod/2002pubs/censr-4.pdf.

Humes, K., & Hogan, H. (2009). Measurement of race and ethnicity in a changing, multicultural America. *Race and Social Problems*, *1*, 111–131.

Humes, K., Jones, N. A., & Ramirez, R. R. (2011, March). *Overview of race and Hispanic origin: 2010 (2010 Census Briefs: C2010BR-02)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/publications/2011/dec/c2010br-02.html.

Iceland, J. (2014). *A portrait of America: The demographic perspective*. Oakland: University of California Press.

Iceland, J., Weinberg, D. H., & Steinmetz, E. (2002, August). *Racial and ethnic residential segregation in the United States: 1980–2000 (Census 2000 Special Reports: CENSR-3)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/hhes/www/housing/housing_patterns/pdftoc.html.

Ingraham, C. (2017, November 30). Scientists can now figure out detailed, accurate neighborhood demographics using Google Street View photos. *The Washington Post*. Retrieved from http://wapo.st/2AnuP9L.

International Association for Social Science Information Services & Technology. (2012). *Quick guide to data citations (Special Interest Group on Data Citation)*. Retrieved from https://www.icpsr.umich.edu/files/ICPSR/enewsletters/iassist.html.

Jensen, J., & Kletzer, L. G. (2006). Tradable srvices: Understanding the scope and impact of services offshoring. In S. M. Collins & L. Brainard (Eds.), *Offshoring white-collar work* (pp. 75–116). Washington, DC: Brookings Institution Press.

Johnson, K. M. (2019, March). *Hispanic population of child-bearing age grows, but births diminish (Data Snapshot)*. Durham: University of New Hampshire, Carsey School of Public Policy. Retrieved from https://carsey.unh.edu/publication/snapshot/hispanic-pop.

Johnson, T. P., & Tourangeau, R. (2018, January 23). Yes, the census should be tracking race and ethnicity. *The Washington Post*. Retrieved from http://wapo.st/2n4GDau?tid=ss_tw&utm_term=.df0b40c25424.

Kallhoff, A. (2011). *Why democracy needs public goods*. New York, NY: Lexington Books.

Keith, T. (2018, March 27). Fact check: Has citizenship been a standard census question? *National Public Radio*. Retrieved from https://n.pr/2GdAVzL.

Kendall, B., & Adamy, J. (2019, February 15). Supreme court agrees to decide legality of census citizenship question. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/supreme-court-agrees-to-

decide-legality-of-census-citizenship-question-11550255991.

Kitchin, R. (2014). *The data revolution*. Thousand Oaks, CA: SAGE.

Klass, G. M. (2012). *Just plain data analysis* (2nd ed.). Lanham, MD: Rowman & Littlefield.

Klein, H. S. (2012). *A population history of the United States* (2nd ed.). New York, NY: Cambridge University Press.

Kochhar, R., & Cilluffo, A. (2018, July). *Income inequality in the U.S. is rising most rapidly among Asians* (Tech. Rep.). Washington, DC: Pew Research Center. Retrieved from https://pewrsr.ch/2NxhwcY.

Kreibich, J. A. (2010). *Using SQLite*. Sebastopol, CA: O'Reilly Media.

Krygier, J., & Wood, D. (2016). *Making maps: A visual guide to map design with GIS* (3rd ed.). New York, NY: Guilford Press.

Kugler, T. A., & Fitch, C. A. (2018). Interoperable and accessible census and survey data from IPUMS. *Scientific Data*, *5*(180007). Retrieved from https://www.nature.com/articles/sdata20187.

Liptak, A. (2016, April 4). Supreme Court rejects challenge on "one person one vote". *The New York Times*. Retrieved from https://nyti.ms/2ooDh5b.

Livingston, G. (2019, May 22). *Is U.S. fertility at an alltime low? Two of three measures point to yes* (Fact Tank). Washington, DC: Pew Research Center. Retrieved from http://pewrsr.ch/2DhW0U2.

Logan, J. R., Stults, B. J., & Xu, Z. (2016). Validating population estimates for harmonized census tract data, 2000–2010. *Annals of the American Association of Geographers*, *106*(5), 1013–1029.

Logan, J. R., Xu, Z., & Stults, B. J. (2014). Intepolating U.S. decennial census tract data from as early as 1970 to 2010: A longitudinal tract database. *The Professional Geographer*, *66*(3), 412–421.

Mackun, P., & Wilson, S. (2011, March). *Population distribution and change: 2000 to 2010 (2010 Census Briefs: C2010BR-01)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/publications/2011/dec/c2010br-01.html.

Maier, M. H., & Imazeki, J. (2013). *The data game: Controversies in social science statistics* (4th ed.). Armonk, NY: M.E. Sharpe.

Martin, J. A., Hamilton, B. E., Osterman, M. J., Driscoll, A. K., & Drake, P. (2018, January). *Births: Final data for 2016 (Vol. 67; National Vital Statistics Reports No. 1)*. Washington, DC: National Center for Health

Statistics. Retrieved from https://www.cdc.gov/nchs/products/nvsr.htm.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, *67*(2), 281–315.

Matthews, S. A., McCarthy, J. D., & Rafail, P. S. (2011). Using ZIP code business patterns data to measure alcohol outlet density. *Addictive Behaviors*, *36*, 777–780.

Menke, K., Smith, R., Jr., Pirelli, L., & Van Hoesen, J. (2016). *Mastering QGIS* (2nd ed.). Birmingham, UK: Packt.

Micarelli, W. F. (1998). Evolution of the United States economic censuses: The nineteenth and twentieth centuries. *Government Information Quarterly*, *15*(3), 335–377.

Monmonier, M. (2018). *How to lie with maps* (3rd ed.). Chicago, IL: University of Chicago Press.

MSU Land Policy Institute. (2014, October). *Northwest Michigan seasonal population analysis* (Networks Northwest). East Lansing: The Land Policy Institute at Michigan State University. Retrieved from http://www.canr.msu.edu/resources/northwest_michigan_seasonal_population_analysis.

Mule, T. (2012, May). *Census coverage measurement estimation report: Summary of estimates of coverage for persons in the United States (DSSD 2010 Census Coverage Measurement Memorandum Series 2010-G-01)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/coverage_measurement/post-enumeration_surveys/2010_results.html.

Muro, M., Rothwell, J., Andes, S., Fikri, K., & Kulkarni, S. (2015, February). *America's advanced industries: What they are, where they are, and why they matter (Metropolitan Policy Program)*. Washington, DC: Brookings Institution. Retrieved from http://brook.gs/2bwynvV.

Murray, C., & Schuetz, J. (2018, June). *Housing in the US is too expensive, too cheap, and just right: It depends on where you live (Metropolitan Policy Program)*. Washington, DC: Brookings Institution. Retrieved from https://brook.gs/2Mb3GM1.

National Research Council. (2007). *Using the American Community Survey: Benefits and challenges*. Washington, DC: National Academies Press.

Nesse, K., & Rahe, M. L. (2015). Conflicts in the use of the ACS by federal agencies between statutory requirements and survey methodology. *Population Research and Policy Review*, *34*, 461–480.

New York City Comptroller's Office. (2017, April). *New York City neighborhood economic profiles (2017*

ed.) New York, NY: Author. Retrieved from https://comptroller.nyc.gov/reports/new-york-city-neighborhood-economic-profiles/.

Nicotera, N. (2007). Measuring neighborhood: A conundrum for human services researchers and practitioners. *American Journal of Community Psychology*, *40*, 26–51.

O' hUallachain, B., & Leslie, T. F. (2013). Spatial pattern and order in Sunbelt retailing: Phoenix in the twenty-first century. *The Professional Geographer*, *65*(3), 396–420.

Office of the Federal Register. (1997, October). *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity (Federal Register, Vol. 62, Iss. 210; 62 FR 58782)*. Washington, DC: Author. Retrieved from https://www.gpo.gov/fdsys/granule/FR-1997–10-30/97–28653.

Office of the Federal Register. (2010, June). *2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas (Federal Register, Vol. 75, Iss. 123; 75 FR 37245)*. Washington, DC: Author. Retrieved from https://www.gpo.gov/fdsys/granule/FR-2010–06-28/2010–15605.

Pew Research Center. (2015, September 3). *The whys and hows of generations research (Tech. Rep.)*. Washington, DC: Author. Retrieved from http://www.people-press.org/2015/09/03/the-whys-and-hows-of-generations-research/.

Pew Research Center. (2018, February 5). *Social media fact sheet (Tech. Rep.)*. Washington, DC: Author. Retrieved from http://www.pewinternet.org/fact-sheet/social-media/.

Pew Research Center. (2018, March). *Wide gender gap, growing educational divide in voter's party identification (Survey Rep.)* Washington, DC: Author. Retrieved from http://www.people-press.org/2018/03/20/wide-gender-gap-growing-educational-divide-in-voters-party-identification/.

Prewitt, K. (2013). *What is your race? The census and our flawed efforts to count Americans*. Princeton, NJ: Princeton University Press.

Pritzker, P., & Gutierrez, C. (2018, April 4). U.S. Census is not about citizenship. *Bloomberg*. Retrieved from https://www.bloomberg.com/view/articles/2018–04-04/u-s-census-is-not-about-citizenship.

Ratcliffe, M., Burd, C., Holder, K., & Fields, A. (2016, December). *Defining rural at the U.S. Census Bureau (American Community Survey and Geography Brief: ACSGEO-1)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/publications/2016/acs/acsgeo-1.html.

Reamer, A. (2017, August). *Counting for dollars 2020: The role of the decennial census in the geographic distribution of federal funds (Initial analysis: Report No. 1)*. Washington, DC: George Washington Institute of

Public Policy. Retrieved from https://gwipp.gwu.edu/counting-dollars-2020-initial-analysis.

Reamer, A. (2018, December). *Census-derived datasets used to distribute federal funds (Counting for dollars 2020: Report No. 4)*. Washington, DC: George Washington Institute of Public Policy. Retrieved from https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds.

Reed, S. B. (2014, April). One hundred years of price change: The Consumer Price Index and the American inflation experience. *Monthly Labor Review*. Retrieved from https://doi.org/10.21916/mlr.2014.14.

Rogerson, P. A., & Plane, D. A. (2013). The Hoover Index of population and demographic components of change: An article in memory of Andy Isserman. *International Regional Science Review*, *36*(1), 97–114.

Saenz, R., & Johnson, K. M. (2018, June). *White deaths exceed births in a majority of U.S. states (Census Brief)*. Madison: Applied Population Lab, Department of Community and Environmental Sociology, University of Wisconsin-Madison. Retrieved from https://apl.wisc.edu/data-briefs/natural-decrease-18.

Salvo, J. J., & Lobo, A. P. (2010). The federal statistical system: The local government perspective. *The ANNALS of the American Academy of Political and Social Science*, *631*, 75–88.

Salvo, J. J., Lobo, A. P., Willett, A. L., & Alvarez, J. A. (2007). *An evaulation of the quality and utility of ACS five year estimates for Bronx census tracts and neighborhoods (Tech. Rep.)*. New York, NY: Population Division, New York City Department of City Planning. Retrieved from http://www1.nyc.gov/assets/planning/download/pdf/data-maps/nyc-population/eval_quality_utility_acs_five_year_estimates_bx.pdf.

Schlossberg, M. (2003). GIS, the US Census and neighbourhood scale analysis. *Planning, Practice, and Research*, *18*(2–3), 213–217.

Schmidt, E. P. (2018, April). *Postsecondary enrollment before, during, and after the Great Recession (Curent Population Reports: P20-580)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/content/census/en/library/publications/2018/demo/p20-580.html.

Shaffer, J. (2016, April 20). *5 tips on designing colorblind-friendly visualizations (Tech. Rep.)*. Tableau Blog. Retrieved from https://www.tableau.com/about/blog/2016/4/examining-data-viz-rules-dont-use-red-green-together-53463.

Short, J. R., & Mussman, M. (2014). Population change in U.S. cities: Estimating and explaining the extent of decline and level of resurgence. *The Professional Geographer*, *66*(1), 112–123.

Smith, A. S., Holmberg, C., & Jones-Puthoff, M. (2012, September). *The emergency and transitional shelter population: 2010 (2010 Census Special Reports: C2010SR-02)*. Washington, DC: U.S. Census Bureau.

Retrieved from https://www.census.gov/library/publications/2012/dec/c2010sr-02.html.

Smith, S. K., Tayman, J., & Swanson, D. A. (2013). *A practitioners guide to state and local population projections*. New York, NY: Springer.

Sperling, J. (2012). The tyranny of census geography: Small-area data and neighborhood statistics. *Cityscape: A Journal of Policy Development and Research*, *14*(2), 219–223. Retrieved from https://www.huduser.gov/portal/periodicals/cityscpe/vol14num2/ch11.html.

Spielman, S. E., Folch, D., & Nagle, N. (2014). Patterns and causes of uncertainty in the American Community Survey. *Applied Geography*, *46*, 147–157.

Stevens, N. (2006, June). *Changing postal ZIP code boundaries (RL33488)*. Washington, DC: Congressional Research Service. Retrieved from https://fas.org/sgp/crs/misc/RL33488.pdf.

Storper, M., Kemeny, T., Makarem, N. P., & Osman, T. (2015). *The rise and fall of urban economies: Lessons from San Francisco and Los Angeles*. Stanford, CA: Stanford University Press.

Sun, M., & Wong, D. W. (2010). Incorporating data quality information in mapping American Community Survey data. *Cartography and Geographic Information Science*, *37*(4), 285–300.

Sunnucks, M. (2017, May 15). BBC: Forget Wall Street, Arizona leads US in finance, banking jobs. *Phoenix Business Journal*. Retrieved from https://www.bizjournals.com/phoenix/news/2017/05/15/bbc-forget-wall-street-arizona-leads-us-in-finance.html.

Swant, M. (2017, September 6). Facebook claims it reaches more people than the U.S. census data says exists. *Adweek*. Retrieved from http://www.adweek.com/digital/facebook-claims-it-reaches-more-people-than-the-u-s-census-data-says-exist/.

Sweigart, A. (2015). *Automate the boring stuff with Python*. San Francisco, CA: No Starch Press.

Talen, E. (2003). Neighborhoods as service providers: A methodology for evaluating pedestrian access. *Environment and Planning B: Planning and Design*, *30*(2), 181–200.

Teaford, J. (2006). *The metropolitan revolution: The rise of post-urban America*. New York, NY: Columbia University Press.

Terrazas, A. (2016, September). *The elusive truth on domestic migration: Comparing data sources on U.S. cross-county migration (Tech. Rep.)*. Retrieved from https://www.zillow.com/research/comparing-migration-data-sources-13252/.

Terry, R. L., & Fond, M. (2013). Experimental U.S. Census Bureau race and Hispanic origin survey questions: Reactions from Spanish speakers. *Hispanic Journal of Behavioral Sciences*, *35*(4), 524–541.

The Economist. (2018, December 22). Less than the sum of their parts: Why American cities are so weirdly shaped. *The Economist*. Retrieved from https://www.economist.com/united-states/2018/12/22/why-american-cities-are-so-weirdly-shaped.

U.S. Census Bureau. (1994, November). *Geographic areas reference manual*. Washington, DC: Author. Retrieved from https://www.census.gov/geo/reference/garm.html.

U.S. Census Bureau. (2009). *Events in the chronological development of privacy and confidentiality at the US Census Bureau (Tech. Rep.)*. Washington, DC: Author. Retrieved from https://www.census.gov/history/www/reference/privacy_confidentiality/privacy_and_confidentiality_2.html.

U.S. Census Bureau. (2010) *Table P1, total population (2010 Decennial Census) [Data file]* Washington, DC: Author. Retrieved from https://data.census.gov/.

U.S. Census Bureau. (2012a, September). *2010 census summary file 1: Technical documentation (SF1/10–4 (RV))*. Washington, DC: Author. Retrieved from https://www.census.gov/prod/cen2010/doc/sf1.pdf.

U.S. Census Bureau. (2012b, February 12). *Stats in action: Target uses American Community Survey (ACS) data (Tech. Rep.)*. Washington, DC: Author. Retrieved from https://www.census.gov/library/video/sia_target_uses.html.

U.S. Census Bureau. (2016a). *American Community Survey and Puerto Rico Community Survey 2016 subject definitions (Tech. Rep.)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.2016.html.

U.S. Census Bureau. (2016b, September 15). *Measuring same-sex couples, sexual orientation, and gender identity on Census Bureau and federal surveys (Webinar for the National LGBTQ Task Force)*. Washington, DC: Author. Retrieved from https://www.census.gov/library/working-papers/2016/demo/file-01.html.

U.S. Census Bureau. (2017a, February 28). *2015 National Content Test: Race and ethnicity analysis report (Tech. Rep.)*. Washington, DC: Author. Retrieved from https://www.census.gov/programssurveys/decennial-census/2020-census/planning-management/final-analysis/2015nct-race-ethnicity-analysis.html.

U.S. Census Bureau. (2017b, September). *ACS summary file technical documentation (2016 ACS 1-year and 2012–2016 ACS 5-year Data Releases)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/acs/technical-documentation/summary-file-documentation.2016.html.

U.S. Census Bureau. (2017c). *American Community Survey: Accuracy of the data (2016) (Tech. Rep.)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html.

U.S. Census Bureau. (2017d). *American Community Survey: Multiyear accuracy of the data (5-year 2012–2016) (Tech. Rep.)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html.

U.S. Census Bureau. (2017e, July). *Methodology for state and county total housing unit estimates (Population and housing unit estimates vintage 2017)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/popest/technical-documentation/methodology.html.

U.S. Census Bureau. (2017f, July). *Methodology for the subcounty total resident population estimates (Population and housing unit estimates vintage 2017)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/popest/technical-documentation/methodology.html.

U.S. Census Bureau. (2017g, July). *Methodology for the United States population estimates: Nation, states, counties, and Puerto Rico (Population and housing unit estimates vintage 2017)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/popest/technical-documentation/methodology.html.

U.S. Census Bureau. (2017h, March). *Subjects planned for the 2020 Census and the American Community Survey (Federal Legislative and Program Uses)*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/publications/2017/dec/planned-subjects-2020-acs.html.

U.S. Census Bureau. (2018a, March 2). *2020 Census operational plan executive summary (version 2.0)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/planning-docs/exe-sum.html.

U.S. Census Bureau. (2018b, March 29). *Questions planned for the 2020 Census and American Community Survey (Federal Legislative and Program Uses)*. Washington, DC: Author. Retrieved from https://census.gov/library/publications/2018/dec/planned-questions-2020-acs.html.

U.S. Census Bureau. (2018c, July). *Understanding and using American Community Survey data: What all data users need to know (Tech. Rep.)*. Washington, DC: Author. Retrieved from https://www.census.gov/programs-surveys/acs/guidance/handbooks/general.html.

U.S. Government Accountability Office. (2009, December). *Formula grants: Funding for the largest federal assistance programs is based on census-related data and other factors (Report to Congressional requesters*

No. GAO-10–263). Washington, DC: Author. Retrieved from https://www.gao.gov/products/GAO-10–263.

U.S. Government Accountability Office. (2016, September). *Community development block grants: Sources of data on community income are limited (Report to Congressional requesters; GAO-16–734)*. Washington, DC: Author. Retrieved from https://www.gao.gov/products/GAO-16–734.

Van Auken, P. M., Hammer, R. B., Voss, P. R., & Veroff, D. L. (2006). The American Community Survey in counties with seasonal populations. *Population Research and Policy Review*, *25*, 275–292.

Wang, F., & Luo, W. (2005). Assessing spatial and non-spatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. *Health & Place*, *11*, 131–146.

Wang, H. L. (2018a, April 3). 17 states, 7 cities sue to remove citizenship question from 2020 Census. *National Public Radio*. Retrieved from https://n.pr/2uJ4Ryj.

Wang, H. L. (2018b, February 8). 2020 Census to count deployed troops at home bases, prisoners at facilities. *National Public Radio*. Retrieved from https://n.pr/2FSsHZ5.

Wang, H. L. (2018c, January 26). 2020 Census to keep racial, ethnic categories used in 2010. *National Public Radio*. Retrieved from https://n.pr/2GhjoCY.

Wang, H. L. (2018d, March 30). 2020 Census will ask about same-sex relationships. *National Public Radio*. Retrieved from https://n.pr/2uy24Ie.

Wang, H. L. (2018e, March 13). 2020 Census will ask black people about their exact origins. *National Public Radio*. Retrieved from https://n.pr/2tJa8W1.

Wang, H. L. (2018f, February 1). 2020 Census will ask white people more about their ethnicities. *National Public Radio*. Retrieved from https://n.pr/2GAd1Ld.

Wang, H. L. (2018g, April 7). 2 More lawsuits join legal fight over 2020 Census citizenship question. *National Public Radio*. Retrieved from https://n.pr/2JU07cE.

Wang, H. L. (2018h, January 10). Adding citizenship question risks “bad count” for 2020 Census, experts warn. *National Public Radio*. Retrieved from https://n.pr/2EESQuO.

Wang, H. L. (2018i, January 29). No Middle Eastern or North African category on 2020 Census, Bureau says. *National Public Radio*. Retrieved from https://n.pr/2GqX9dT.

Wang, H. L. (2018j, March 29). U.S. census to leave sexual orientation, gender identity questions off new surveys. *National Public Radio*. Retrieved from https://n.pr/2njH5Rf.

Wang, H. L. (2019, February 15). Supreme court to decide whether 2020 Census will include citizenship question. *National Public Radio*. Retrieved from https://n.pr/2WIXy3s.

Wang, M., Kleit, R. G., Cover, J., & Fowler, C. S. (2012). Spatial variations in US poverty: Beyond metropolitan and non-metropolitan. *Urban Studies*, *49*(3), 563–585.

Warshawer, G. (2013, January 25). Bronx's Little Italy tries to add new spice. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/SB10001424127887323539804578261800609567478.

Wilson, R., Hasanali, S., Sheikh, M., Cramer, S., Weinberg, G., Firth, A.,… Soskolne, C. (2017). Challenges to the census: International trends and a need to consider public health benefits. *Public Health*, *151*, 87–97.

Wong, D. M. (2010). *The Wall Street Journal guide to information graphics*. New York, NY: W. W. Norton.