visualizing health and healthcare data

Creating clear and compelling visualizations to "see how you're doing"

katherine **rowell**
lindsay **betzendahl**
cambria **brown**

WILEY

Table of Contents

List of Illustrations

Visualizing Health and Healthcare Data

Creating Clear and Compelling Visualizations to “See How You’re Doing”

Katherine Rowell
Lindsay Betzendahl
Cambria Brown

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Names: Rowell, Katherine L., author. | Betzendahl, Lindsay, author. | Brown, Cambria, author.

Title: Visualizing health and healthcare data : creating clear and compelling visualizations to “see how you’re doing” / by Katherine Rowell, Lindsay Betzendahl, Cambria Brown.

Description: First edition. | Hoboken, New Jersey : Wiley, [2020] | Includes bibliographical references and index.

Identifiers: LCCN 2020032174 (print) | LCCN 2020032175 (ebook) | ISBN 9781119680888 (paperback) | ISBN 9781119682325 (adobe pdf) | ISBN 9781119680864 (epub)

Subjects: LCSH: Medical statistics—Data processing. | Information visualization.

Classification: LCC RA409.5 .R69 2020 (print) | LCC RA409.5 (ebook) | DDC 610.2/1—dc23

LC record available at https://lccn.loc.gov/2020032174

LC ebook record available at https://lccn.loc.gov/2020032175

Cover Design and artwork: Wiley

Preface

We understand the challenges you face when trying to communicate health and healthcare data. We wrote this book to share our deep experience and expertise in creating clear, compelling, and actionable displays of health and healthcare data that empower people to “see how they're doing.”

As a group, we've spent the last 30 years working on this book—29 conducting the research, and one writing it. Our research can be categorized into three areas: formal education, work experience, and self-education.

We've each had the privilege of a formal education delivered by recognized health and healthcare experts and thought leaders. This education is foundational to our ability to think critically about the various aspects of what health is, how health and healthcare systems operate, and the structures and policies that influence it all.

We've played a role in the development, capture, and analysis of just about every type of health and healthcare data imaginable—from administrative claims data to risk-adjusted clinical outcomes data to complex public health survey data. We've deciphered the inner workings of innumerable transactional systems and untangled databases created by evil geniuses.

Each of us can describe the moment we were bitten by the data visualization bug and set on a path of self-education. We sought out thought leaders in the field—some we know only through their books and blogs; others we have developed lifelong professional and personal friendships with. We learned and honed our viz skills through deep practice—stretching ourselves outside our comfort zone,

stopping to reflect on successes and errors, making adjustments, and continuing the process over time. As a result, we have proudly joined the ranks of recognized data visualization experts, and we work with leading health and healthcare organizations throughout the world. And now, we have written this book to share our experience and expertise with you.

This book is organized into three sections. Section I is focused on understanding different types of data to be displayed, and requirement-gathering methods. All too often teams want to jump directly to creating displays, but this is a mistake to be avoided. Before you can determine how to display data and information, you must understand what will be displayed and for whom.

Section II provides an overview of the research that informs the best practices of table and graph design. Included in this chapter are the four fundamental shapes you can use alone or in combination to create clear and compelling displays of data. This section also includes detailed examples of the most common mistakes people make in selecting a chart type, with explanations about why they don't work, and examples of what works better and why.

Section III of the book defines and describes the characteristics of dashboards, reports, multidimensional exploratory displays (MEDS™) and infographics. It includes examples of each type of display and tips and tricks to designing and building them.

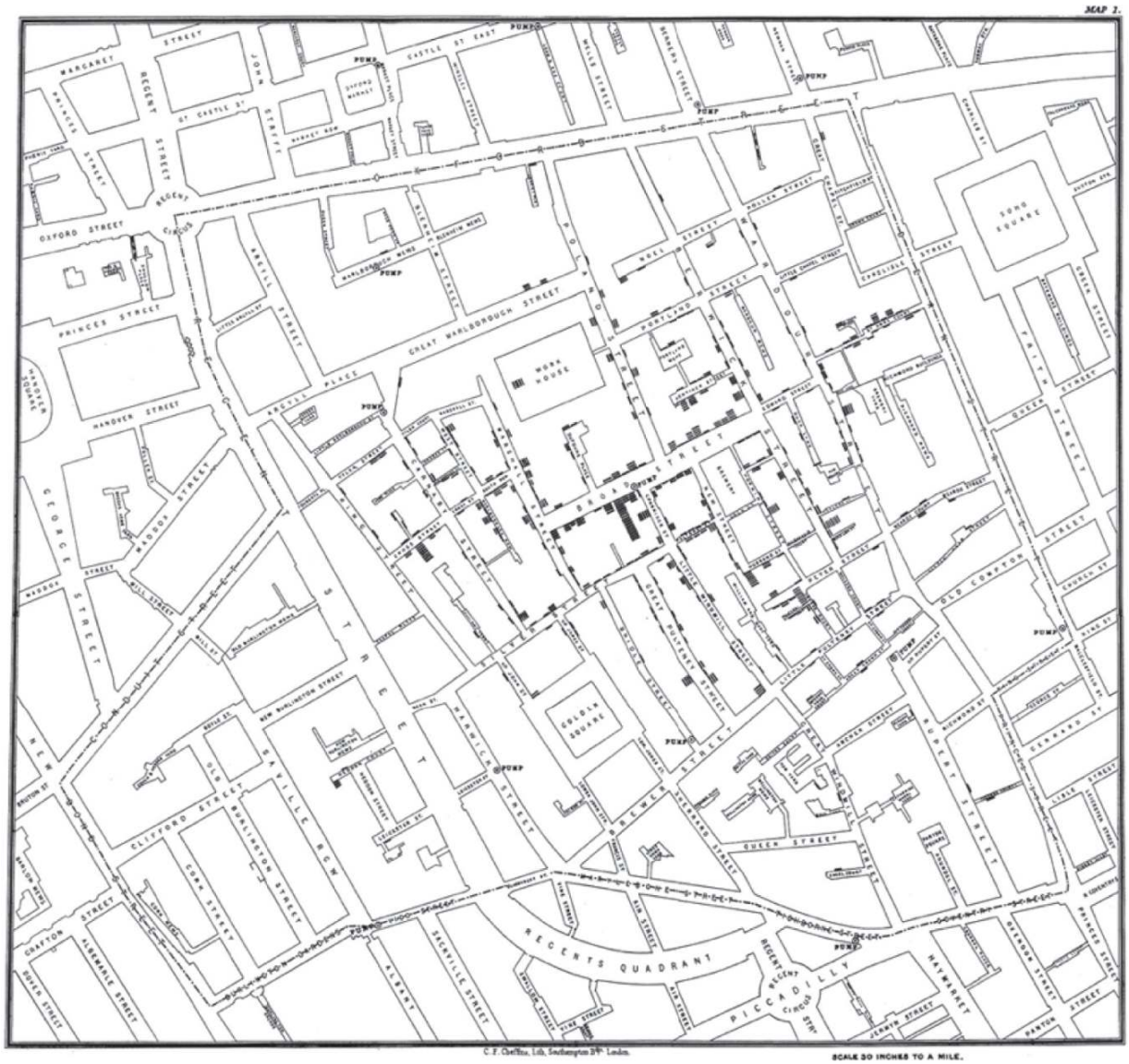
We ardently believe that making the message and the story in the data clear will improve health and healthcare systems. We hope this book will help you to join in our efforts.

SECTION ONE
Establishing a Framework and Process

> “If you can't describe what you're doing as a process, you don't know what you're doing.”
>
> —W. Edwards Deming, American engineer, statistician, and leading quality management thinker

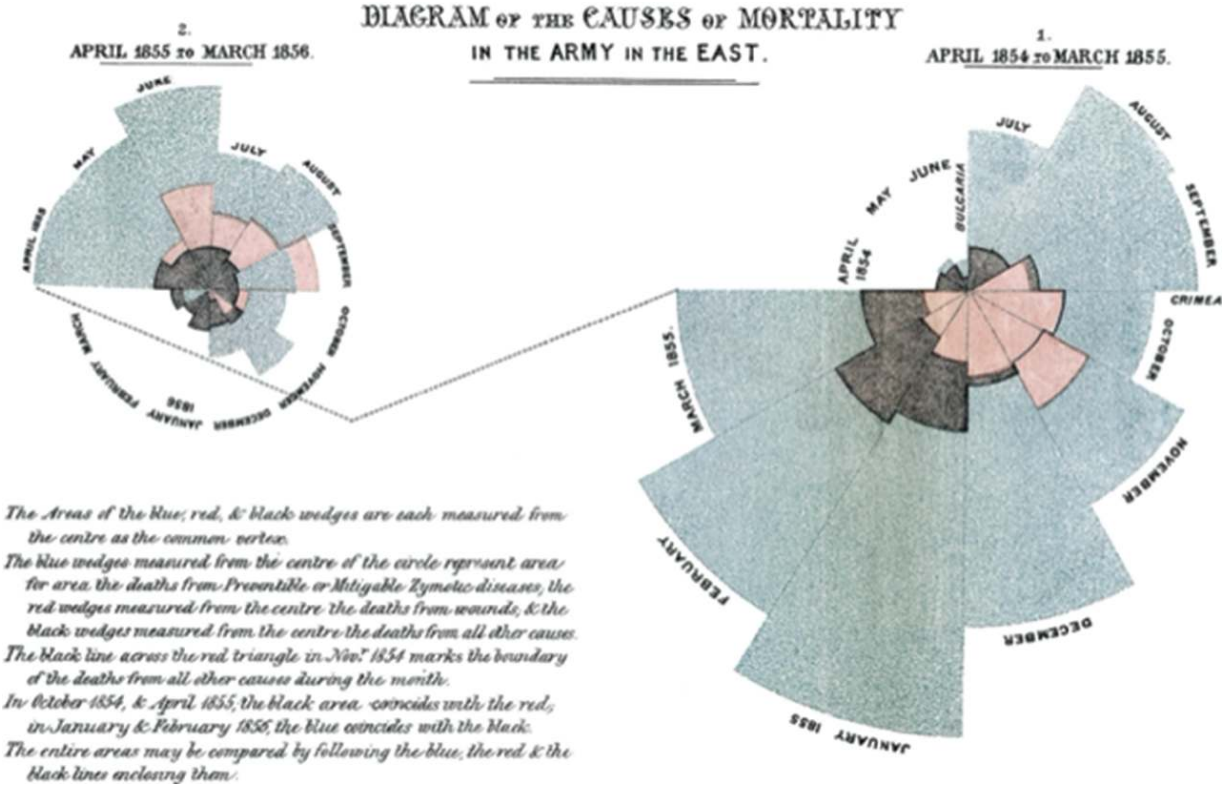
Chapter One
Health and Healthcare Data Visualizations of Historical Importance

Even before modern-day visualization research validated the direct and powerful relationship between the way information is presented and the way we see and understand it, pioneering healthcare statisticians and caregivers like John Snow (1813–1858) and Florence Nightingale (1820–1910) understood that visual display could be a highly effective method for grasping and communicating the messages buried in data. No one who has ever taken an epidemiology course can forget Dr. John Snow's classic work, *On the Mode of Communication of Cholera*. By mapping the London street addresses of residents who had become sick (and in many cases died) and their distance from City water pumps, Snow could visually and effectively communicate the relationship between a single pathogen-tainted water source and the homes of people who contracted the disease. Most people who had fallen ill, it turned out, lived near the Broad Street pump. Snow persuaded the town council to remove the pump's handle, and the outbreak abated.



In 1868, British nurse Florence Nightingale—distressed by the alarmingly high mortality rates in the Crimean War—began to compile statistics on causes of death. Her analysis revealed that of the 900,000 soldiers who died during the war—more than half of 1,650,000 combatants from all countries involved—most had succumbed to preventable diseases arising from unsanitary conditions in the hospitals where they were treated, and not as a direct result of battlefield wounds. Nightingale recognized the buried message: better hygiene could have saved—and could still save—thousands of lives.

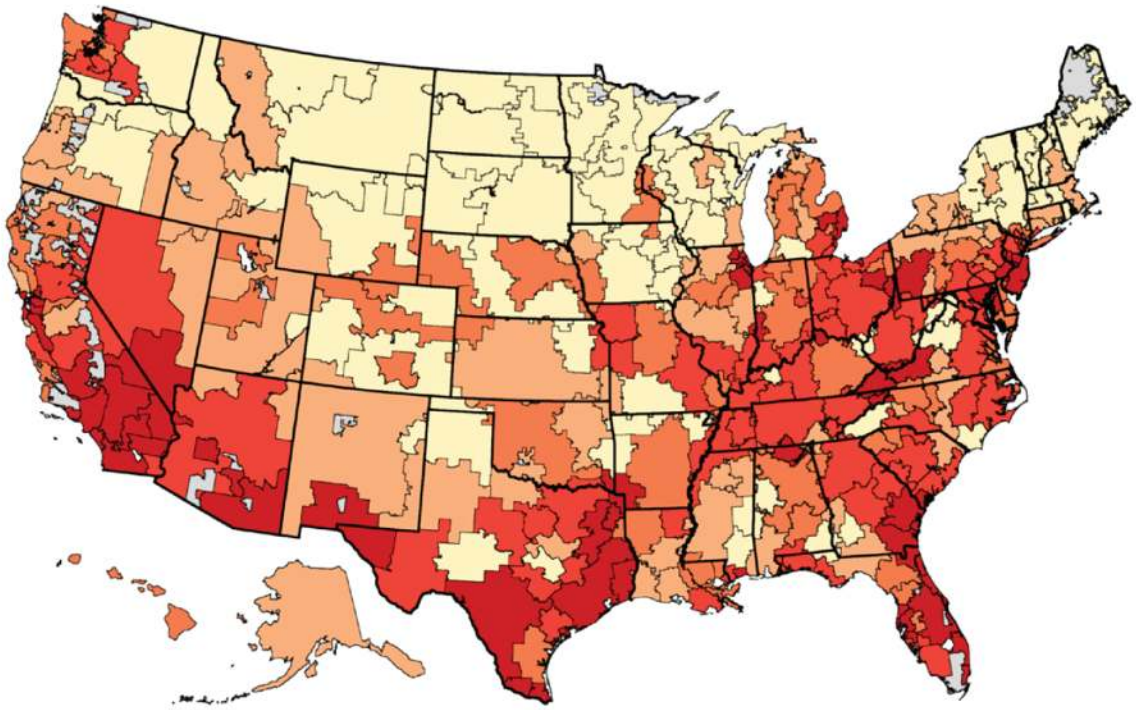
As impressive as her statistics were, Nightingale worried that the tables she presented to Queen Victoria would seem tedious, even incomprehensible, and feared that members of the British Parliament were unlikely to be swayed by numbers lying flat on a page. So Nightingale devised ingenious ways of presenting the information in charts.



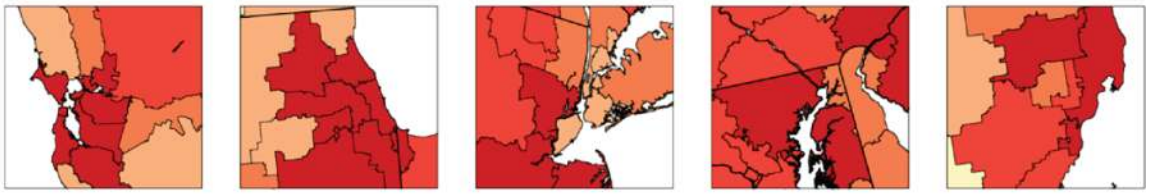
In the now-famous “Diagram of the Causes of Mortality in the Army in the East,” each month is represented as a twelfth of a circle. The months with more deaths are shown with longer wedges so that the area of each wedge represents the total number of deaths. Preventable deaths are blue, deaths due to wounds are red, and deaths from all other causes are black. Over the months after March 1855, when members of the Sanitary Commission began repairing, cleaning, and otherwise improving field hospital conditions, the blue wedges shrank dramatically. Showing incredible insight into the power of displaying the data in this way, Nightingale said her graph was designed “to

affect thro' the Eyes what we fail to convey to the public through their word-proof ears.”

More recent efforts by healthcare researchers like those led by Dr. Jack Wennberg at the Dartmouth Atlas Project have documented glaring—and, for the most part, inexplicable—variations in how medical resources are apportioned and delivered in the United States. The project builds on Medicare data to provide comprehensive information and analysis about national, regional, and local markets, as well as individual hospitals and their affiliated physicians.



Map 4. Percent of cancer patients admitted to intensive care during the last month of life (deaths occurring 2003-07)



Source: The Dartmouth Institute: *The Dartmouth Atlas of Healthcare*

Consider the map reproduced from the Dartmouth Atlas Report: “Quality of End-of-Life Cancer Care for Medicare Beneficiaries.” It displays the percent of cancer patients admitted to intensive care during the last month of life compared by hospital referral regions. About 24% of cancer patients nationwide were admitted to intensive care at least once during that last month. However, the percent thus admitted varied more than sevenfold across those regions (dark red versus light yellow areas on the map). This map leads the viewer to ask, “Why are these rates so dramatically different across the country?,” and perhaps to add an even more significant question: “What should the rate be?”

Geospatial displays of data like this one make the variation in end-of-life care jump off the page in a way that it never would if the data were buried in a table or report narrative. Such geospatial maps and accompanying reports, coupled with the underlying research, have helped policymakers, the media, healthcare analysts, and others improve their understanding of the efficiency and effectiveness of our healthcare system. As with the map created by John Snow, the visualizations built for the Dartmouth project make the story easy to see and understand and have formed the foundation for many of the nation's ongoing efforts to improve American health and healthcare systems.

Throughout history the power of data visualizations to help us see and consider the stories buried in our health and healthcare data has resulted in profound insights and often changes and improvement in our health and healthcare. And now with the advent of technology that allows us to amass data and quickly explore it, we are on the cusp of revolutionary insights and improvements. It is an exciting time indeed.

Chapter Two
Stop Hunting Unicorns and Start Building Teams and Know The Data

People often ask us what it takes to create clear and compelling displays of health and healthcare data that people love to use, and which raise awareness and move people to take action. The answer is paradoxically simple. It requires strong teams of experts in the highly complex disciplines of health and healthcare, statistics, data, technology, accessibility design, data visualization, and user experience testing—teams who appreciate the unique skills, experiences, and expertise each person brings to a project and the ability to work collaboratively.

Because here is the secret you already know. The idea that any lone person will ever have every single bit of expert knowledge and skill in health and healthcare, technical applications, and data visualization and design required to deliver beautiful and compelling dashboards, reports, and infographics is sheer lunacy. That's why organizations have to stop hunting unicorns and commit to building diverse teams who bring the expertise required and the ability to communicate and collaborate.

With that in mind, here is a summary of tips for building great data-analytics, reporting, and data visualization teams.

Search for Characteristics and Core Competencies

It is imperative to understand what characteristics and core competencies are required to complete the work.

Here's where to begin:

Curiosity. When teams are curious, they question, probe, and inquire. Curiosity is a crucial impetus for uncovering interesting and relevant stories in our health and healthcare data. Above all else you need a team of really, genuinely inexhaustibly curious people!

Health and Healthcare Subject-Matter Expertise. Team members with front-line, boots-on-the-ground, clinical, quality, operational, policy, financial, research, and public health experience and expertise are essential for identifying the questions of interest and the decisions or needs of the stakeholders for and to whom data is being analyzed and communicated.

Data Analysis and Reporting. Without exception, at least one member of your team must have math, statistics, and data-analysis skills. Experience with data modeling is a plus if you can find it; at a minimum, some familiarity with the concept of modeling is beneficial. The ability to use data analysis, reporting, and display tools and applications is also highly desirable, but another more technically trained IT team member may be able to bring this ability to the table if necessary.

Technical: IT and Database Expertise. Often, groups will confuse this skill area with data-analysis and reporting competence. Data and database architecture and administration require an entirely different set of skills from those needed for data analysis, so it's important not to conflate the two. You'll need team members who know how to extract, load, and transform (ETL) and architect data for analysts to use. And while you may sometimes find candidates who have both skill

sets, don't assume that the presence of one means a lock on the other.

Data Visualization, Accessibility, and User Experience Testing. Knowledge of best practices and awareness of current research is required to create clear, useful, and compelling dashboards, reports, and infographics. But as you will learn in this book, these skills are not intuitive; they are based on research into human vision and cognition and must be learned and honed over time. And although it's not necessary for every team member to become an expert in this field, each should have some awareness of it to avoid working at cross-purposes with team members employing those best practices. (That is, everyone should know better than to ask for 3-D red, yellow, and green pie charts and other do's and don'ts you will learn in this book!)

We do wish that data-analysis and reporting unicorns were real! Life would be so much simpler. But they aren't and never will be, so we let go of that fantasy long ago and have found tremendous success in training and building great teams with the different skills that are required. We encourage you to do the same.

Get to Know the Data

The term *data governance* has become pervasive in the health and healthcare data environment. However, at present, any efforts at creating full-fledged data governance standards and associated documentation are at best nascent and more often nonexistent. Therefore, to perform correct analysis and create useful displays of data, analysts and data visualizers must establish discipline and

process to learn and understand—and get to know—health and healthcare data.

Classifications, Intent, Purpose, and Lineage

Powerful data analysis and visualizations require subject-matter expertise about health and healthcare, and knowledge and understanding about the data being captured and reported. Whether it is a survey instrument, electronic health record, billing system, or the classification systems used to organize them and the databases to store them, diligent inquiry and research must be performed to ensure that the available data are fully understood. This discovery includes but is not limited to learning:

- How the data are categorized (classified) and relate to each other.
- Whether the data are part of a widely recognized and accepted classification system such as ICD10 Codes or a custom, unique system created for one particular project.
- Why, for what intent and purpose, a set of data is collected in the first place, such as medical claim payment, clinical research, patient care, public health initiatives.
- Moreover, and as necessary, how the data may have changed over time, their lineage.
- What the data definitions are.

For example, the HEDIS measure set was created in 1991 (version 1) and originally stood for the Health Maintenance Organization (HMO) Employer Data and Information Set. In

1994 version 2 was renamed the Health Plan Employer Data and Information Set. In the original design, HEDIS measures provided consumers and regulators with information to reliably compare competing managed care health plans. The focus was on how they were paying out premium dollars to providers for services. HEDIS was a marketing tool for HMOs to demonstrate that they were providing the most benefit to subscribers as compared to all other HMO plans.

A fundamental change occurred in 1997 when the National Committee for Quality Assurance (NCQA) adopted the HEDIS measure set as a way to measure and report about Medicare beneficiaries' quality of care. That is, the type of services delivered to specific cohorts of patients, such as mammograms for women, were used as proxies for quality care. At the same time, the NCQA also announced that the HEDIS acronym would be changed to stand for Healthcare Effectiveness Data and Information Set. Commercial insurers quickly followed suit, using HEDIS compliance targets in contract and payment negotiations with providers.

In some ways, it may be logical to use the HEDIS measure set as a proxy for the quality of care provided. For example, if the evidence supports that women of a certain age would benefit from an annual mammogram, and the number of eligible women receiving them can be reliably measured, then it may be that providing a mammogram is a reasonable proxy for quality care. However, as anyone who works in the health and healthcare profession knows, HEDIS falls short as a comprehensive measure of quality care because it does not capture the various outcomes that are important to clinicians, patients, researchers, and payers.

In other words, WYSIWYG (what you see is what you get) does not always (or even usually) apply to health and healthcare data. Therefore, it is crucial to take time to research the structure and categories of classification systems, and the primary purpose of a dataset, and how it may or may not be able to be reliably used for analysis and reporting.

Deciding on a set of terms (terminology) that accurately represent a system of concepts, creating a vocabulary with definitions of those terms, and arranging and organizing related entities into a classification system or database would appear to be a fairly straightforward (if long and detailed) task. Well, as the old joke goes, if you want 12 different opinions, put six people in a room and go around twice. (If they are experts in the field, the total will be closer to 24 different opinions.) Reaching consensus on how to define and classify health and healthcare information is tough. Evidence changes; treatments, procedures, and patients metamorphose; stakeholders define terms differently based on ever-shifting goals and objectives.

All these transformations mean that teams must dedicate time to research and consider the underlying classification systems of data definitions, and the intent, purpose, and lineage of health and healthcare databases. Starting here will save much frustration and wasted time later on and increase confidence about the ability to create valuable dashboards, reports, infographics, and other displays of health and healthcare data.

In addition to understanding the intent, purpose, and lineage of health and healthcare data, a fundamental understanding of data types, scales/levels of measure, and data relationships is required to create meaningful displays of data.

There is a wealth of books, publications, and blog postings on this topic, written by experts in the field of statistics and data analysis. The objective is not to recreate that level of detail here. Instead, the intention is to define and describe the data concepts that must be understood to ensure they are displayed correctly.

Two Types of Data

Qualitative/Categorical Data

Qualitative/categorical data are most easily understood as nonnumerical data that may be observed but not measured or have mathematical functions performed on them, like a sum or average. Some examples may include a patient's eye color, sex, and perceptions about their health status or quality of care they received. Other examples may include measuring organizational change or physicians' implementation of evidence-based guidelines.

Sometimes these data are also described as *yes/no* data. For example, do people have food allergies, wear their seatbelt, smoke cigarettes, feel safe at home?

Although qualitative/categorical data can be coded with numerical values (e.g., 1 = male, 2 = female), those values do not have any mathematical meaning.

Quantitative/Numerical Data

Quantitative/numerical data measure the quantity or amount of something, and they are numerical, that is, they can be mathematically quantified and allow for the calculation of metrics such as sum, average. Quantitative/numerical data fall into one of two categories: continuous or discrete.