

## Contents

Overview .....	2
Chapter 1: The Importance of Data Analytics .....	4
Come to LOGICAL conclusion .....	5
Avoid Avoidable Mistakes .....	5
Use Your Powers for Good .....	6
Safeguard Yourself with Data Analytics.....	7
Data Analytics versus Opinions on Social Media (Anecdotes) .....	7
A Rigorous Data Analysis of the Fitness Application .....	9
Data Analytics Triumphs Over Personal Testimonies .....	10
Chapter 2: Relationships, Types, Graphs.....	13
Quantitative versus Qualitative Insights .....	14
Continuity and Discreteness in Data .....	15
Visualization of Interactions: Scatterplots.....	17
Sequential Data Visualization: Time Series Plots .....	18
Quantitative Counts: Discrete Data.....	19
Qualitative Insights: Categorical, Binary, and Ordinal.....	20
Categorical Insights .....	21

## Overview

I am passionate about data analytics and sharing my insights through teaching, writing and speaking. My background includes a faculty position at the largest and best college in the United States and over two decades of experience at a variety of data analytics and data science companies to include Wells Fargo, Bank of America (3 times), IBM, and many others. During my tenure, I honed my skills in presenting data analytics and data science in a way that is easily understandable. My book is designed to help you grasp the essential concepts, practices, and knowledge in the data sciences, enabling you to analyze your data with confidence.

Throughout my career, I've received countless requests from budding data analysts and data scientists seeking guidance in data analytics. This book is the culmination of my years of experience and thousands of hours of dedicated work, distilled into a comprehensive, practical guide for mastering the data sciences.

My hope is that this book allows you to view data analytics as I do—a tool to unlock the hidden insights in your data. The journey to discovery is often challenging, with data analytics puzzles along the way. Yet, these challenges are both thrilling and rewarding to overcome. It's often the enigmatic puzzles in your data that spur you to play detective, leading to unexpected and exciting findings!

I hope you'll see this book as a journey into the world of data analytics. We begin by delving into the practical essentials of data analytics and data science, then broaden our scope to understand the challenges and processes involved in using data analytics to enhance human knowledge. This is truly an adventure in data analytics!

In this book, you'll explore the various aspects of data analytics. We start with an introduction to the field and its significance. Next, we cover essential skills and knowledge about different data types and methods for summarizing data, both graphically and numerically. A key focus will be on uncovering relationships between variables. We then advance to understanding populations and estimating their characteristics using sample data.

Further, we examine how data analytics contributes to the scientific method, aiding in expanding human understanding. This involves designing experiments that distinguish causation from correlation, and addressing common challenges in finding insights in your data. The book concludes with criteria for evaluating the quality of data and research designs, essential for assessing your work or that of others.

As you read, you'll be equipped to advance to more complex studies in data analytics, such as data wrangling, data mining, regression analysis, and experimental design. This knowledge lays a solid foundation for careers in data analytics, data science and machine learning.

You'll notice a minimal use of equations in this book. The emphasis is on conceptual understanding and coding practice, leaving calculations to statistical software. Expect more graphs than equations, although for those interested, equations can be found in standard textbooks.

Throughout, I use Google Colab as an example, but the focus is on basic graphing and analysis, applicable across all common data analytics software, not limited to any specific tool.

To enhance practical learning, the book includes links to download datasets from my website, allowing you to learn by doing and apply the concepts discussed.

## Chapter 1: The Importance of Data Analytics

Data analytics has become a cornerstone in our everyday experiences. Consider the recommendations from nutrition experts for certain diets or supplements. News channels often highlight market research data, such as consumer trends or brand popularity, with precise confidence intervals. In the realm of education, student performance is assessed not just through traditional grades but also through data-driven techniques like adaptive learning platforms, offering tailored insights into student progress.

In the wider world, data analytics is everywhere: from analyzing streaming service viewing habits to conducting detailed customer satisfaction surveys for products and services. Political strategies are often shaped by in-depth electoral data analysis. Insurance companies rely on sophisticated algorithms to set premiums.

The sports industry, too, has embraced analytics in game strategies and player performance evaluations. Businesses and organizations use data analytics for informed decision-making, optimizing operations, and enhancing profitability.

Data analytics is indispensable for propelling scientific, economic, and social advancement. It's the backbone of research, validating the significance of results and ensuring their validity for publication. This field, rich in its applications and impact, raises the question: what exactly encompasses the domain of data analytics?

The realm of data analytics is the art and science of extracting insights from information. This expertise equips you with the right techniques for gathering data, applying accurate analyses, and presenting findings effectively. Data analytics is foundational to scientific discovery, informed decision-making, and accurate forecasting. It enables a deeper comprehension of various subjects. Surprisingly, this field is not just about numerical outcomes. It encompasses a broad spectrum of practices, decisions, and methodologies for data collection and analysis, ensuring the generation of valid results and credible conclusions.

Let's explore two key reasons why mastering data analytics is vital in today's society. First, data analysts serve as navigators, helping us learn from data and steer clear of pitfalls that could lead to misleading conclusions. Second, in an era where decisions and opinions are increasingly data-driven, it's imperative to critically evaluate the analyses presented by others.

In my perspective, data analytics is an exhilarating field, embodying the joy of discovery, learning, and challenging preconceptions. It is a catalyst for generating new knowledge, progressively expanding the boundaries of what is known.

### Come to LOGICAL conclusion

Data analytics is more than just mere numbers and facts. It's not just about statistics like 'a large percentage of professionals recommend a certain product'. Rather, it's a comprehensive toolkit that enables reliable learning from data. Data analytics helps distinguish between sound and questionable conclusions, a crucial skill in an era with more data than ever and various misinterpretations, often influenced by biases.

Data analysts provide essential guidance in producing trustworthy analyses and forecasts. They assist researchers in avoiding a multitude of analytical pitfalls. When data analysis methods are applied correctly, they tend to yield precise results. In fact, these analyses take into account uncertainty and potential errors in findings. Data analysts ensure that every aspect of an insight adheres to the right methods for yielding dependable outcomes. These methods include:

- Gathering reliable data.
- Analyzing data correctly.
- Drawing logical conclusions.

### Avoid Avoidable Mistakes

Data analytics, much like its precursor statistics, begins well before the analysis phase. The journey to producing insightful findings from data is long and multifaceted. It encompasses the design of the study, the selection and measurement of variables, the development of sampling techniques and determining sample size, data cleaning, and choosing the right analytical methodology, among other factors. The overall integrity of the results relies on this entire process. Any weakness in this chain can lead to unreliable outcomes. Below is a glimpse into potential pitfalls and analytical errors that can impact a study:

**Biased Data Collection:** An improperly selected dataset can skew the conclusions right from the outset. For instance, in human subject research, if the subjects differ significantly from the general population, it can affect the results. Biased data collection in a retail analysis setting might occur if customer feedback is only gathered from high-end store locations, potentially leading to skewed insights that do not represent the broader customer base.

**Overgeneralization:** Conclusions drawn from one group may not be applicable to another. Data analytic inferences are inherently limited, and understanding these limitations is crucial. An example of overgeneralization in data analytics could be a tech company using social media trends from teenagers to predict technology preferences for all age groups, potentially leading to misinformed product development strategies.

**Causality Challenges:** Determining causality – when one factor causes a change in another – requires stringent standards. While many may quickly infer causality from correlation, a robust experimental

design with random assignment is essential to confidently attribute causality. An example of the causality challenge in data analytics is when a study finds a correlation between increased social media use and lower happiness levels, but without rigorous experimentation, it's unclear if social media use causes unhappiness or if less happy people are more drawn to social media.

**Inappropriate Analysis:** Missteps like analyzing a complex study with a simplistic approach, using an insufficient set of variables, or choosing the wrong statistical measure can lead to inaccurate conclusions. An example of inappropriate analysis in data analytics could be a financial firm using a simple linear regression to predict stock prices without considering market volatility, leading to unreliable investment strategies.

**Assumption Violations:** Most data analytic methods come with underlying assumptions about the sample, variables, data, and the model. While some assumptions can be relaxed under certain conditions (often thanks to principles like the central limit theorem), violating key assumptions can mislead the study's results. An example of assumption violations in data analytics occurs when an e-commerce company assumes normal distribution in customer purchase patterns for their predictive models, but the actual buying behavior is highly skewed due to a few big spenders, leading to inaccurate sales forecasts.

**Data Mining Risks:** Even with a meticulous approach, extensive exploration of a dataset can lead to false positives. This issue arises particularly when numerous tests are conducted, resulting in statistically significant findings by chance. Diligent data analysts keep track of the number of tests and contextualize their findings accordingly. An example of data mining risks in data analytics is when a researcher analyzing a large healthcare dataset performs multiple tests to find links between various lifestyle factors and a rare disease, potentially identifying coincidental associations that are not truly indicative of causal relationships.

To arrive at reliable conclusions, numerous aspects must align correctly. Unfortunately, there are many ways to err in data analysis, leading to misleading results. Data analysts guide others through this complex terrain. We'll delve into all these aspects and more, providing a comprehensive understanding of how to navigate these challenges.

### Use Your Powers for Good

Data analytics is leveraged across virtually all fields to interpret the immense volume of data available. Even if data analytics isn't your primary area of expertise, it can significantly boost your impact in your chosen field. A practical understanding of data analytic methods is crucial not only for generating new insights in your domain but also for comprehending the work of others.

On the flip side, as a data analyst, your skills are in high demand across diverse sectors: academic institutions, research facilities, government bodies, industry, and more. Moreover, careers in data analytics are often well-compensated. Reflecting on my favorite quote:

"The greatest joy in data analytics lies in the privilege to explore and discover insights in the diverse fields of others." - Dr. Ernesto Lee

My interests span a wide range, and knowledge in data analytics provides the tools to understand them all.

### Safeguard Yourself with Data Analytics

You might be familiar with the saying about lies and statistics, popularized by Mark Twain. But is it accurate? Certainly, unscrupulous analysts can misuse methodologies to draw false conclusions. However, data can easily be manipulated to create intentionally misleading analyses. But how can you identify such manipulations? Without a background in data analytics, these deceptions can be challenging to spot. Knowledge in data analytics is the key to defending yourself against such manipulation and to responding to misinformation with intelligence.

In today's world, we're inundated with more data and analyses intended to sway our opinions than ever before. Are you equipped to handle it?

### Data Analytics versus Opinions on Social Media (Anecdotes)

Understanding the strength of data analytics can be achieved by contrasting it with anecdotal evidence (junk on social media), a fundamentally different approach. Let's explore their differences.

Stuff on social media (anecdotal evidence) consists of individual stories. It appears in various forms, from personal endorsements to verbal recommendations. It's typically a personal account attesting to the validity or effectiveness of a claim, focusing on singular experiences, emotionally influenced, and often presented by non-experts in the subject.

Consider these instances of anecdotal evidence:

"Amazing! I switched to this new fitness app and saw immediate improvements in my health!"

"A friend has been eating fast food for years without health issues. Surely, the health risks are overblown!"

"This new skincare product gave me a radiant complexion. It's definitely effective!"

We often rely on friends for suggestions on various topics. However, the issue with anecdotal evidence is its potential to mislead, even when individual stories are true. Why is that?

The following table contrasts data analytics with anecdotal evidence:

Data Analytics	Anecdotal Evidence
Employs large, representative data samples, making findings more broadly applicable.	Uses small, biased samples, limiting generalizability.
Measurements are precise and taken in controlled settings.	Relies on spontaneous, subjective descriptions.
Considers and adjusts for other relevant variables.	Often overlooks other significant factors.
Establishes causality through stringent criteria.	Assumes causality without substantial evidence.

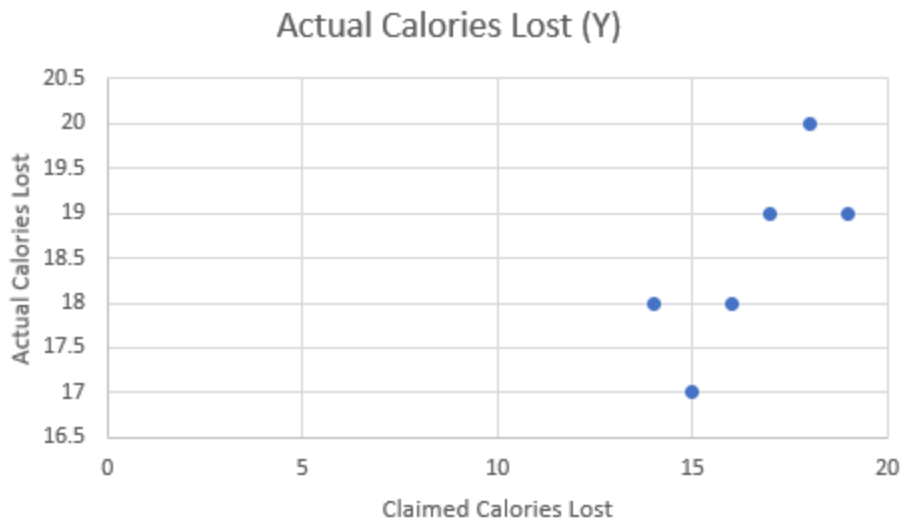
A brief examination of this table reveals the unreliability of anecdotal evidence. This problem is exacerbated by psychological tendencies. People are naturally drawn to and remember emotionally charged, extraordinary personal stories. In contrast, data analytics, although methodically collected and analyzed, lacks this emotional component.

Additionally, humans are predisposed to perceive causality in sequential events. In other words, humans often see patterns where there are none. Anecdotal evidence often selectively highlights successful cases, ignoring the numerous instances of failure.

For instance, when someone emotionally and enthusiastically shares their positive experience with a new fitness app, we might remember and believe in the app's effectiveness. However, we rarely hear from those who did not see any benefit, nor do we know the other lifestyle changes the individual might have made. This selective storytelling distorts perception, underscoring why data analytics, with its comprehensive and objective approach, is more reliable for drawing conclusions from data.

These biases inherently skew the interpretation of personal testimony towards atypical results and unwarranted causal inferences.





Now, let's visually demonstrate these issues and discuss how data analytics and scientific rigor address them.

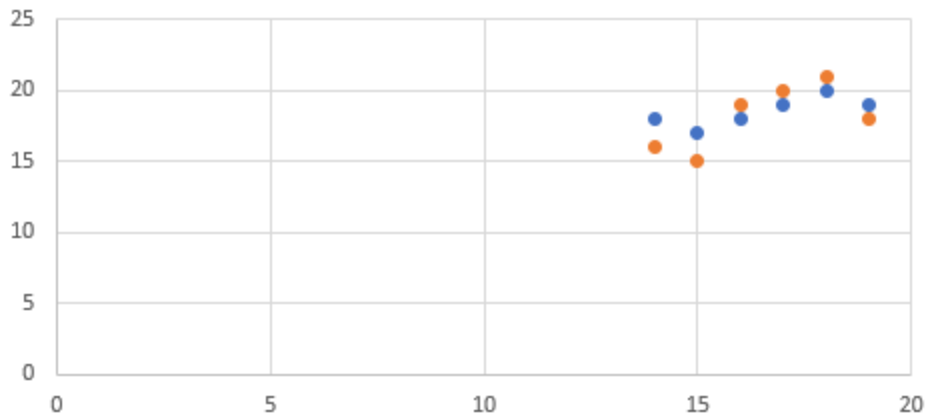
The chart depicts outcomes from individual reports of people using a speculative fitness application. Imagine this visual as part of a promotional segment on a streaming platform. For argument's sake, we'll accept that these individuals are completely truthful. The app appears successful, doesn't it? Significant health improvements are shown! When you see the users and listen to their heartfelt experiences of better health, it's tempting to credit the app for these improvements.

However, this chart doesn't tell us everything we need to know. Recall, individual reports often rely on small, non-representative samples that don't translate to the wider population. These users might have been selected for their remarkable success stories, or they may have volunteered their experiences. In either case, the data shown is from those with the most striking and emotionally resonant outcomes. As the disclaimer often states, their experiences are not typical!

Regrettably, we are naturally inclined to believe such compelling narratives. We give undue emphasis to sensational, personal accounts.

### A Rigorous Data Analysis of the Fitness Application

Let's envision a rigorous data analysis carried out through a comprehensive experiment with a larger, randomly selected sample that mirrors the general population. This approach involves both a group using the app (the treatment group) and a group not using it (the control group) for a fair comparison. We must look past individual success stories to grasp the broader reality that a methodical data analysis can unveil.



In a hypothetical graph, imagine blue dots indicating users of the fitness app and red dots for non-users. The outcome here might not seem as remarkable as the individual success stories. Why is this? The combination of random sampling and a control group offers an objective overview, providing a wider context. You'll notice that while some app users might show the significant health improvements advertised, many others only show marginal changes. These individuals typically don't share their less dramatic experiences. Moreover, the non-users, represented by the red dots, might exhibit similar health trends to the app users. Altogether, using the app doesn't seem to result in notably better health outcomes than not using it.

Since this hypothetical data analysis employs a random sample reflective of the wider population, as opposed to a self-chosen subset, we can more confidently generalize these results to a broader context.

Individual accounts on social media, the entertainment news industry, business reports, aren't inherently untrue. Rather, they fail to depict average outcomes, consider other influencing factors, or involve a comparison group. In fact, the individuals from the personal testimonies could also be part of the larger data set. Suppose they participated in the rigorous study before featuring in promotional material. Their stories remain factual.

Nevertheless, once they opt into a marketing campaign without a comparative framework, their narratives can mislead regarding the actual effectiveness of the app.

A fundamental takeaway highlighted in this book is the significance of how you gather, clean and analyze your data. The methodology behind data collection fundamentally influences the insights you can derive from it.

Data Analytics Triumphs Over Personal Testimonies

In data analytics, there are two fundamental approaches to assess whether a fitness app leads to better health: observational analyses and randomized controlled experiments (RCEs).

In an observational analysis, analysts gather and scrutinize all relevant data within a representative sample to construct a model that highlights each variable's influence. For each individual, you would evaluate variables such as heart rate variability, activity levels, nutritional intake, overall wellness, etc., and the use of the fitness app. Once you consider the impact of all these variables, you can then determine if there's a correlation between the app usage and health improvements. Personal testimonies, in contrast, offer none of this vital, broader context. This is classic data analytics.

Randomized controlled experiments (RCEs) are the premier standard for they enable you to infer causal links about the intervention's impact. We aim to ascertain whether the fitness app directly contributes to better health. RCEs randomly allocate individuals to either the experimental group using the app or the control group without it. This stratagem aims to balance the groups at the outset, making any differences observed by the experiment's end likely due to the app's effect. This is advanced analytics and is often left to the realm of researchers.

Don't fret; we'll delve into the intricacies of experimental design in far greater detail later. For the interim, grasp that utilizing data to pinpoint causal connections (the app improving health) and the ability to extrapolate findings beyond the initial cohort requires using assorted data collection methods and experimental frameworks.

Relying on personal testimonies for decisions isn't always detrimental. If you're seeking a music streaming service recommendation, the stakes are relatively low, especially if you share musical preferences with your advisor. Yet, when it comes to significant choices regarding finance, healthcare, and personal fitness, steer clear of decisions based solely on personal accounts. Instead, consider scientific data and professional analyses, even though they may lack the allure of emotionally charged narratives from seemingly relatable individuals.

Should you find yourself swayed by personal testimonies, remind yourself of their unrepresentative nature!

Imagine scrolling through your social media feed and coming across a post from an acquaintance, Alex. Alex claims, "Just one month using MindBoost meditation app and I've never felt more focused and productive in my life!"

This statement is compelling—Alex shares a before-and-after snapshot of a cluttered vs. an organized desk, attributing this transformation to the app. The post is filled with positive comments and has been shared widely. It seems everyone is convinced of MindBoost's effectiveness based on Alex's experience.

However, reflecting on the principles outlined so far in this chapter, we should approach Alex's claim with a healthy dose of skepticism. While it's a powerful personal testimony, it lacks the broader context necessary to draw a reliable conclusion. There's no mention of other lifestyle changes Alex might have made during this time, nor is there any data on how representative Alex's experience is among all users of the app.

Alex's story is precisely the type of personal testimony that can influence perceptions on social media, yet it doesn't provide the objective analysis necessary to validate the claim. Without a structured approach like the one offered by data analytics, including randomized controlled experiments or observational analyses, we cannot be certain if MindBoost is genuinely effective or if Alex's experience is an outlier.

As data analysts and scientists, we must look beyond individual narratives and seek out patterns in larger, more diverse datasets to form our conclusions. Only then can we determine if there is a true cause-and-effect relationship, as opposed to a coincidental correlation amplified by social media's echo chamber.

## Chapter 2: Relationships, Types, Graphs

In the field of data analytics, data is the cornerstone. Data provides the crucial insights that analysts and scientists gather to understand patterns, test theories, and make informed decisions. Data analytics, after all, is the discipline of gleaning knowledge from data. However, data come in various forms, each capturing different types of information. The nature of the data critically determines the insights you can draw from it and, just as importantly, the insights you cannot. Thus, recognizing the diverse types of data is essential.

The term "data" might conjure specific notions, seeming almost abstracted from tangible reality. Throughout this book, I encourage you to view data as pieces of evidence collected during your investigation. Data are the building blocks used to construct answers to pressing questions, such as:

- Can a new app increase productivity?
- Does regular meditation lead to better mental health?
- Can a software update enhance smartphone battery life?

When examining such questions, you might record a ton of attributes. In a study involving app users, for instance, you might note quantitative metrics like hours of usage and qualitative categories like user satisfaction levels. You might include discrete counts, like the number of tasks completed, or binary data indicating whether the app was opened each day or not.

For certain attributes, there might be multiple recording methods. Take, for example, monitoring stress levels: you could quantify this through a stress score or categorize individuals simply as 'stressed' or 'not stressed'.

This chapter will guide you through the various types of variables, the insights they afford, and how to visually represent them using intuitive examples. We'll begin with singular variables and progress to exploring how graphs can elucidate the interplay between pairs of different data types. At this juncture, we'll use graphs solely for gaining preliminary insights. Later, we'll commence with computing numerical summaries.

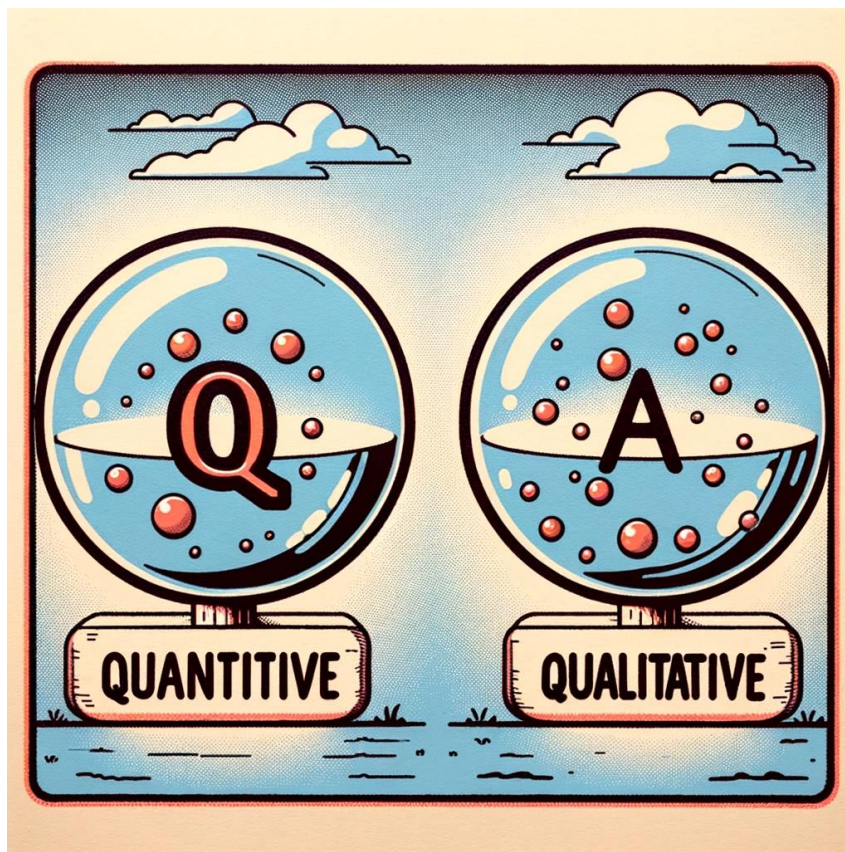
A relationship between two variables suggests that knowing the value of one can help you better predict the value of the other. Different professions and analysts may adopt different taxonomies for data categorization. For each data type, I'll offer synonyms to encompass these classifications. Also, I'll present sample data to demonstrate how they might be formatted in your analytical software.

## Quantitative versus Qualitative Insights

The primary division in the types of insights pulled from data is between quantitative and qualitative. Is the attribute one that can be precisely quantified with numbers, or is it more about the quality that isn't numerically based?

**Quantitative:** This insight is captured numerically and reflects a precise measurement or count. Examples include the number of users on an app, the speed of a website's response time, and sales figures. These are often termed numerical insights in data analytics.

**Qualitative:** This insight describes characteristics without using numbers. Observations are categorized into a finite set of classifications. This type of insight can encompass aspects that are challenging to measure and might contain elements of subjectivity. User experience, software interface design, and customer service quality fall under qualitative insights.



Within these overarching categories, there are further subdivisions.

## Continuity and Discreteness in Data

When the insights you're extracting are numeric, you're delving into quantitative insights. This encompasses two classifications:

### *Continuity in Data*

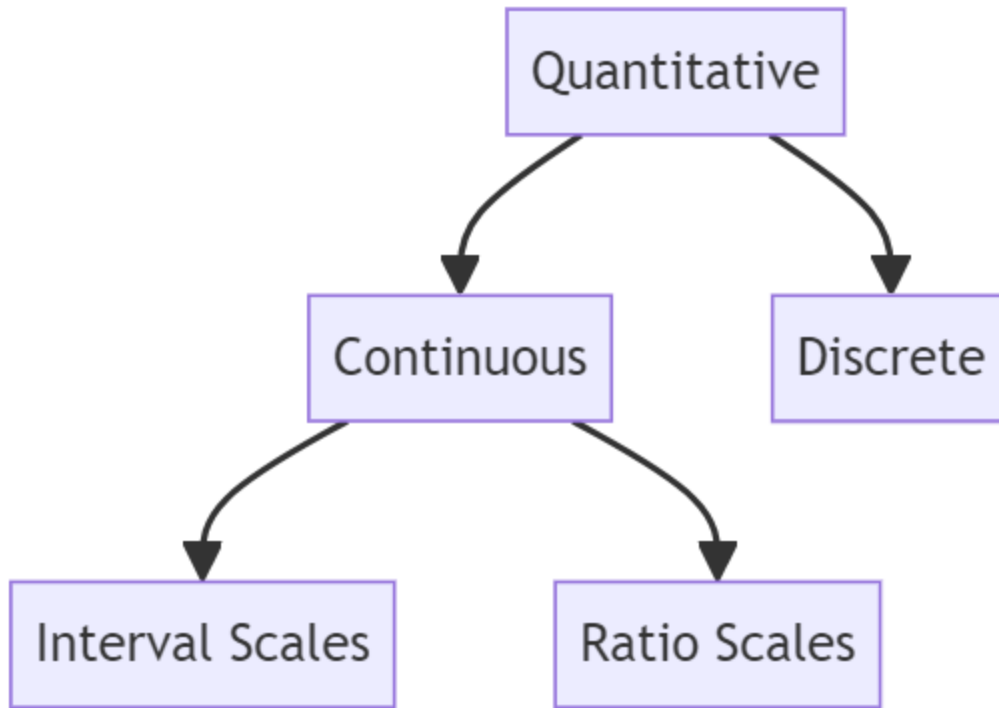
Continuous insights can assume any value along a continuum, including fractions and decimals, representing an infinite spectrum of possibilities between any two points (real numbers). Every nuance in these values is meaningful. Such insights are often gauged on a scale. For instance, the time spent on a webpage, the amount of data used by an app, or the response time of a server represent continuous data.

Data analysts distinguish continuous insights further based on the scales of measurement:

**Interval Scales:** With interval scales, the difference between values is significant. For instance, the time difference between 2:00 PM and 4:00 PM is the same as between 9:00 AM and 11:00 AM. However, such scales lack a true zero that signifies the absence of the attribute. For instance, while the time of day has a zero hour, it doesn't indicate the absence of time altogether.

Given this absence of a true zero, comparing magnitudes through ratios doesn't hold on interval scales. 20 degrees on a server's temperature scale isn't twice as hot as 10 degrees. You can operate with addition and subtraction on an interval scale, but multiplication and division don't apply.

**Ratio Scales:** With ratio scales, not only are the intervals significant, but these scales also include a zero measurement that denotes the absence of the attribute. For instance, zero bytes used signifies no data usage. Thus, measurement ratios are meaningful on these scales. 60 GB of data is six times more than 10 GB. Operations like addition, subtraction, multiplication, and division are all applicable on a ratio scale.



When examining continuous insights, you can analyze attributes such as the average, median, distribution, range, and variability. For example, the average data usage per smartphone in the U.S. might be 8 GB per month for adult users. The subsequent chapter will delve into these summary insights.

#### *Visualization of Distributions: Histograms*

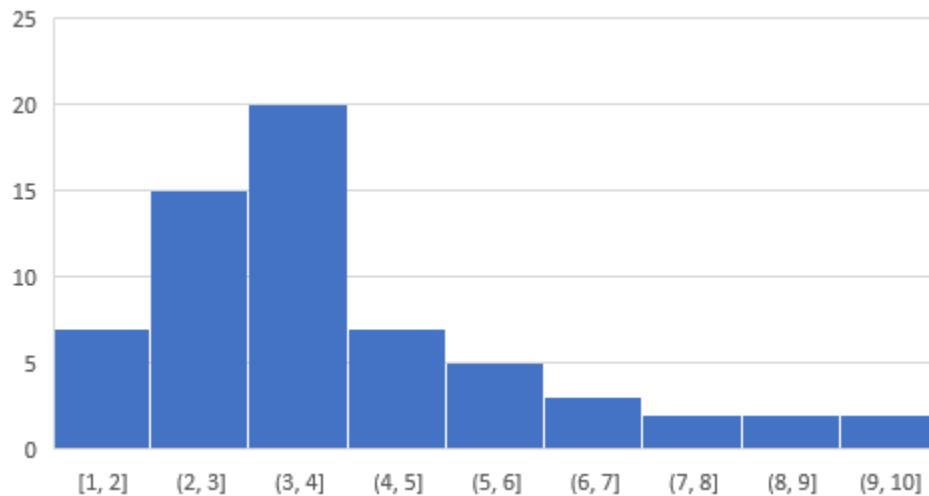
Histograms are superb for visualizing continuous insights because they illustrate the distribution of values. Grasping the distribution is key to understanding which values are common or rare, among other traits.

Each histogram bar covers a range of values for the continuous insight on the horizontal axis, known as bins. The bar's height indicates the count or proportion of observations within each bin.

Consider a histogram that portrays the distribution of daily screen time for teenagers. It reveals the frequency and spread of screen time values. Through this, one can discern if the distribution is symmetrical or skewed, the breadth of screen time values, and the most prevalent data points. The nuances of histograms will be explored further within this chapter.



The example histogram might show that most teenagers' daily screen time clusters around 3-4 hours. You might not find instances below 30 minutes, and the distribution could be skewed towards the higher end, with some teenagers reaching up to 10 hours of screen time.



Notice how continuous insights allow the examination of a broad array of characteristics, including central tendency, variability, distribution shape, and outliers. This will be contrasted with qualitative insights later, where the focus shifts from numerical analysis to categorical interpretation.

#### Visualization of Interactions: Scatterplots

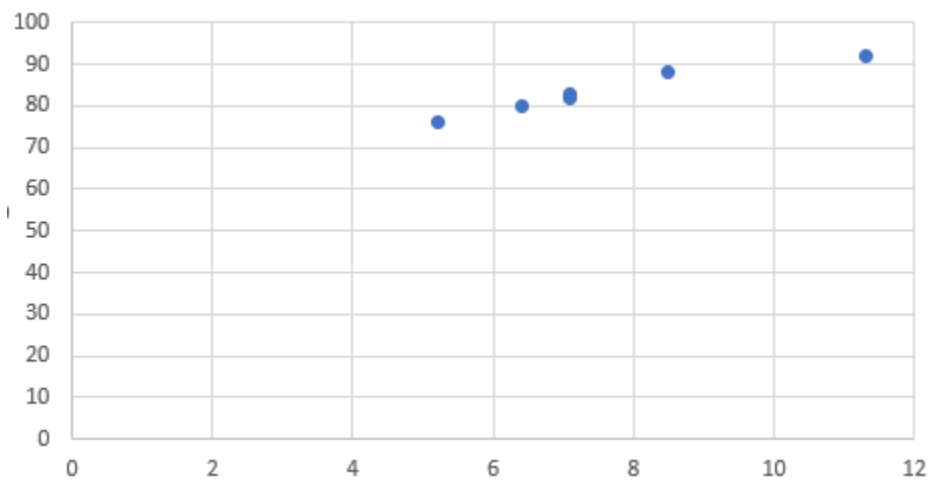
With two quantitative insights, you can depict their interplay using a scatterplot. Scatterplots excel at illustrating the relationship between two quantitative variables. On the plot, each point has an X and Y coordinate representing a paired value, corresponding to a single entity's recorded data.

For instance, consider a scatterplot of data points representing the hours spent on a learning app (X) and the resulting test scores (Y) for individual students:

Hours (X)	Score (Y)
5.2	76
8.5	88
11.3	92
6.4	80
7.1	83
7.1	82

- X-axis: The horizontal axis typically bears the predictor variable, which could elucidate variations in the other variable.

- Y-axis: On the vertical axis, you'll often find the response variable, whose changes are explained by the predictor variable.



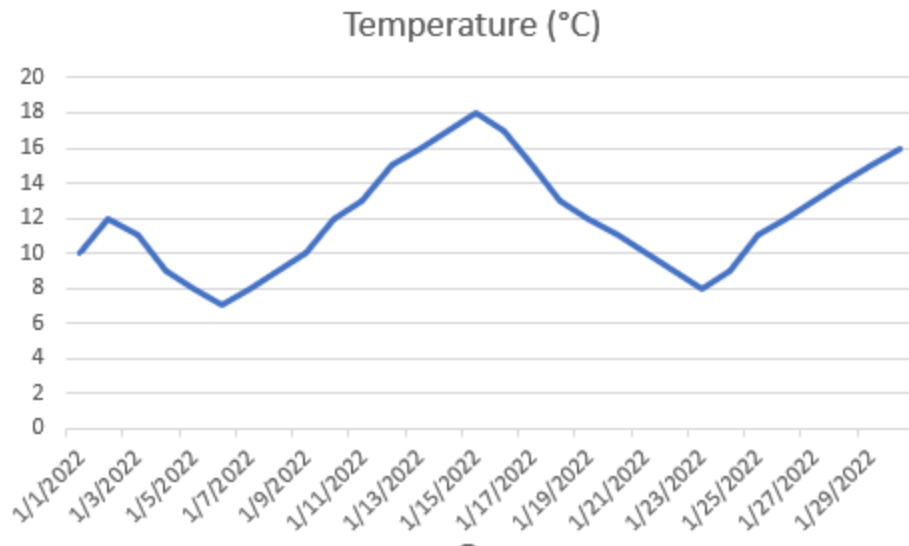
When there's no explicit predictor-response relationship between the variables, their placement on the axes is less critical.

On the scatterplot, each dot might symbolize the relationship between time spent on the learning app and the corresponding test score for a student. Such a scatterplot could reveal that as study hours increase, so do test scores. The strength of this relationship can be quantified using correlation, which we'll explore in chapter 3.

### Sequential Data Visualization: Time Series Plots

While scatterplots map the correlation between two quantitative variables, time series plots trace one quantitative variable over time. These plots show you how the variable evolves, allowing for the detection of trends and cyclical patterns. Measurements in time series plots are taken at uniform intervals, like hourly, daily, or monthly.

On these plots, time spans the X-axis, and the Y-axis charts the quantitative measurement.



A time series plot could, for example, illustrate the growing number of subscribers to a streaming service. Seasonal viewership patterns might also emerge from such data.

### Quantitative Counts: Discrete Data

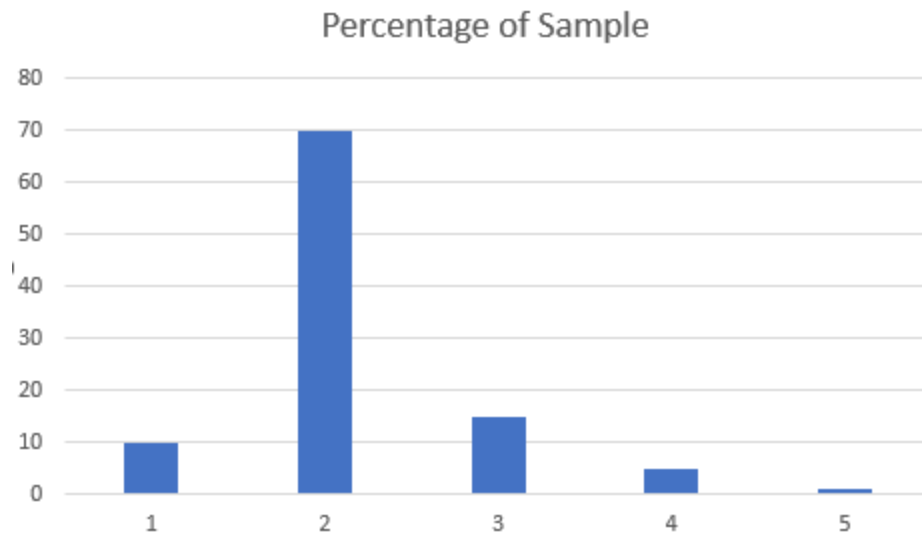
Discrete quantitative insights tally occurrences of a characteristic or event. These counts are whole numbers and don't break down into smaller parts. For instance, a company may report having 15 or 20 computers but not 15.3. The possible values are numerable and distinct. Other discrete data examples include the number of items sold, the number of visitors to a website, or the volume of customer service inquiries.

With discrete data, rates of occurrence or aggregated summaries like averages, totals, and measures of spread are analyzable. For example, an online retailer might report an average of 150 daily transactions.

### Bar Charts

Bar charts are a fundamental tool in data analytics for visualizing discrete insights. Each bar signifies a distinct category, with the height indicating its frequency within the dataset. Bar charts are useful for showcasing the prevalence of different categories.

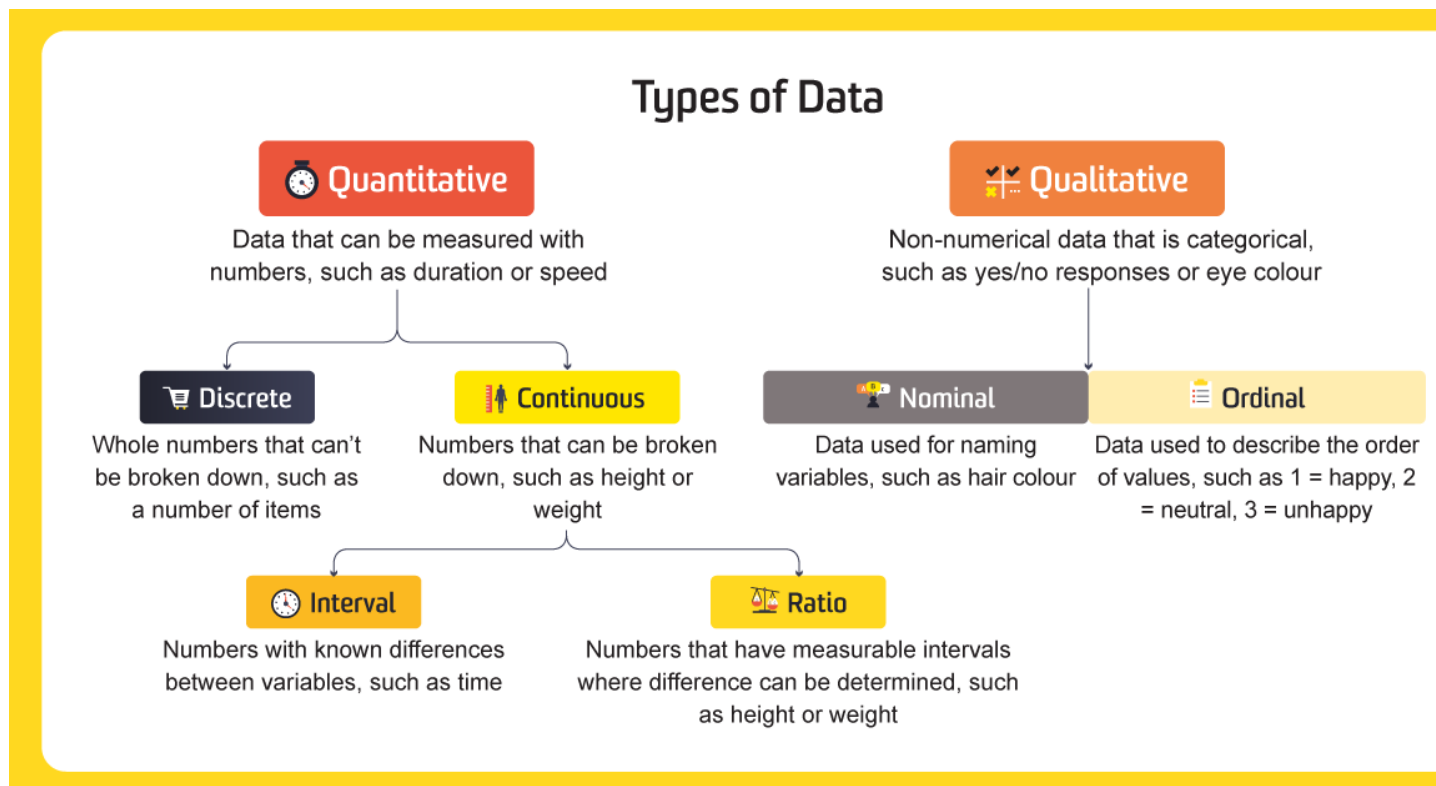
Although bar charts might resemble histograms, there's a clear distinction: bars in a histogram are adjacent, reflecting the continuous nature of the data they represent. In contrast, bars in a bar chart are separated, each corresponding to a singular, discrete category.



This bar chart indicates that Florida pre-teens with 2 electronic devices make up nearly 70% of the sample and are the most common.

### Qualitative Insights: Categorical, Binary, and Ordinal

In data analytics, when you document data that sorts your observations into specific categories, you're handling qualitative insights. There are three kinds of qualitative variables—categorical, binary, and ordinal. With these types of data, the focus is typically on the proportion each category occupies within the dataset. Hence, bar charts and pie charts are popular tools for visualizing qualitative variables, effectively representing the counts and relative proportions of each category.



As noted earlier with quantitative insights, observe that qualitative insights provide a different depth of understanding. This point is particularly underscored in discussions about binary variables. Wherever there is an option to record data as either quantitative or qualitative, it's advantageous to opt for the quantitative form, given the richer insights it offers.

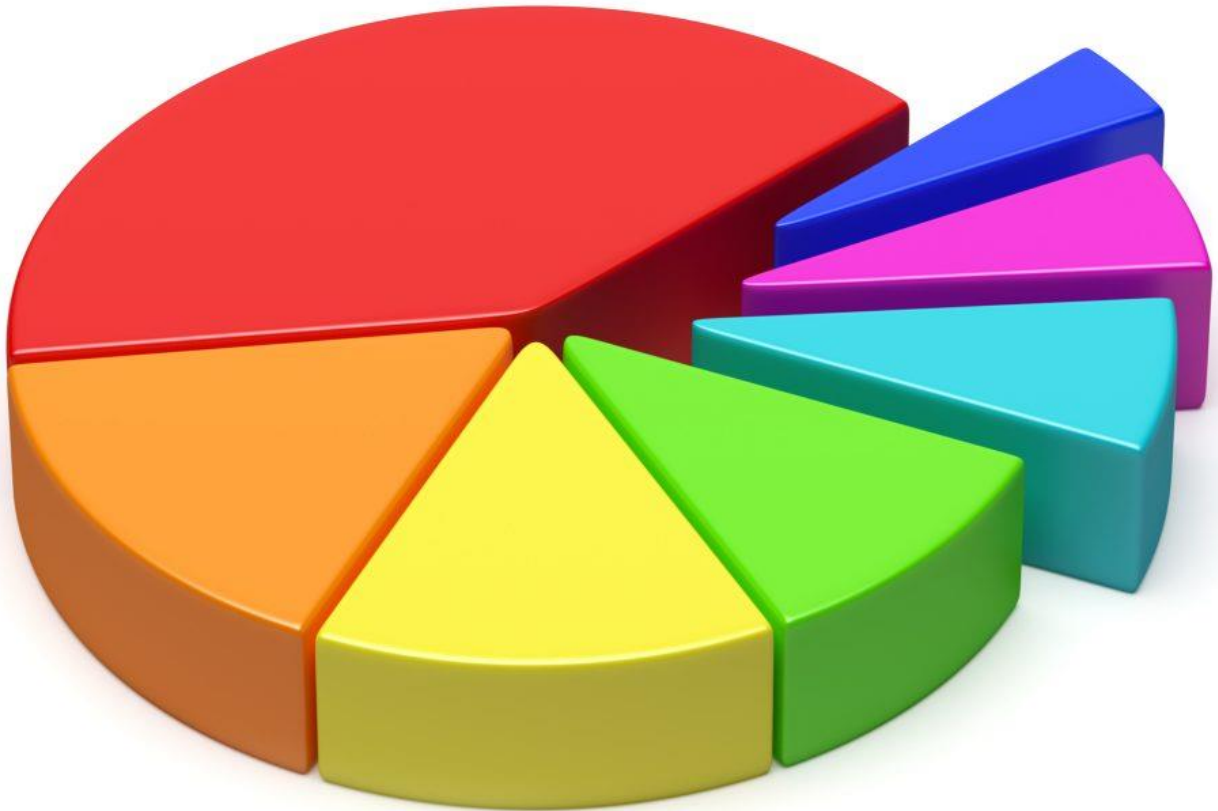
### Categorical Insights

Categorical insights consist of values that fall into a limited set of distinct groups based on a specific characteristic. In Data Science, categorical values are often encoded for algorithmic processing. For a categorical variable, you assign categories without any inherent ranking. These variables can delineate comparison groups within your data, such as different user segments in an app's user base. Categorical insights are also known as attribute or nominal variables.

For instance, smartphone brands can be a categorical variable with values like Brand A, Brand B, Brand C, and so on.

The categorical insights in the pie chart might be the outcome of a survey examining preferred smartphone brands in 2022.

Pie charts excel at emphasizing the fractional composition of the whole. They are suitable for categorical, binary, and ordinal insights that define segments in your dataset.



This chart might show that Brands A, B, and C collectively make up a significant majority of the market share, with the rest distributed among several smaller competitors.