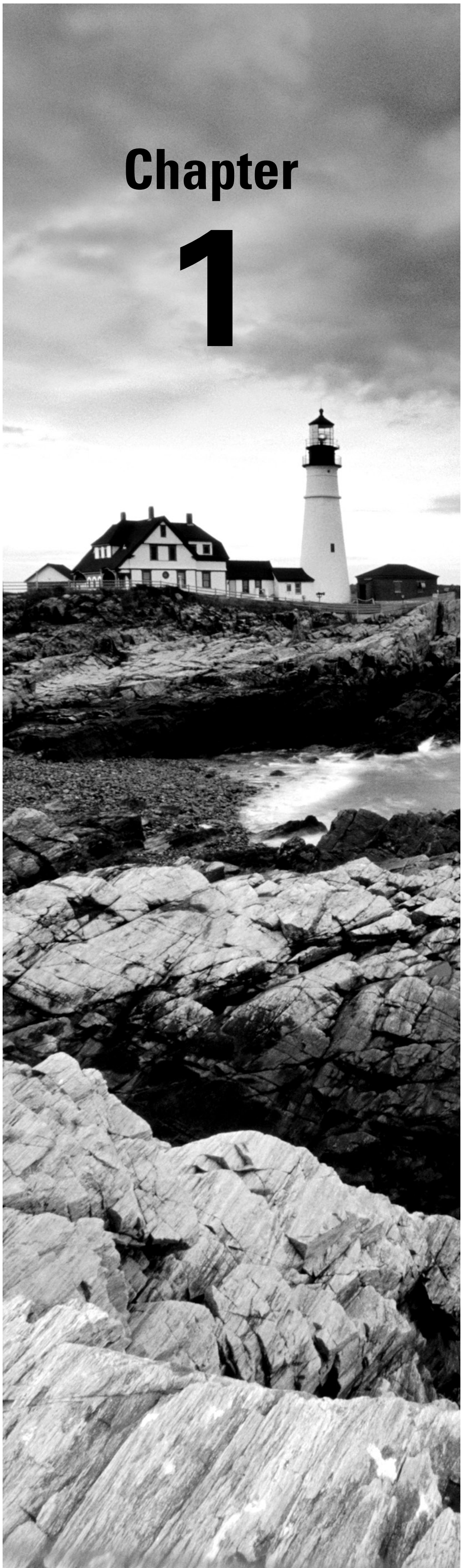


Chapter 1

Today's Data Analyst

THE COMPTIA DATA+ EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE:

✔ **Domain 1.0: Data Concepts and Environments**

- 1.5. Identify artificial intelligence (AI) concepts

✔ **Domain 3.0: Data Analysis**

- 3.2. Given a scenario, select the appropriate statistical method or function

Analytics is at the heart of modern business. Virtually every organization collects large quantities of data about its customers, products, employees, and service offerings. Managers naturally seek to analyze that data and harness the information it contains to improve the efficiency, effectiveness, and profitability of their work.

Data analysts are the professionals who possess the skills and knowledge required to perform this vital work. They understand how the organization can acquire, clean, and transform data to meet the organization's needs. They are able to take that collected information and analyze it using the techniques of statistics and machine learning. They may then create powerful visualizations that display this data to business leaders, managers, and other stakeholders.

Welcome to the World of Analytics

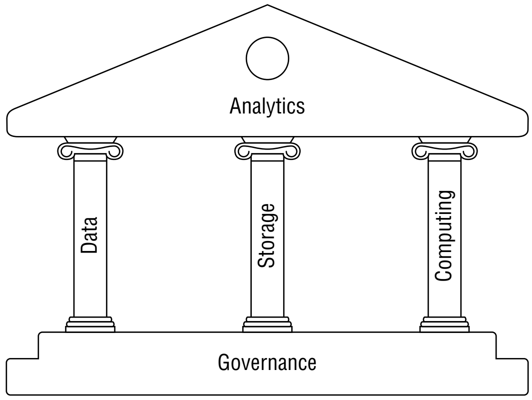
We are fortunate to live in the Golden Age of Analytics. Businesses around the world recognize the vital nature of data to their work and are investing heavily in analytics programs designed to give them a competitive advantage. Organizations have been collecting this data for years, and many of the statistical tools and techniques used in analytics work date back decades. But if that's the case, why are we just now in the early years of this Golden Age? Figure 1.1 shows the three major pillars that have come together at this moment to allow analytics programs to thrive: data, storage, and computing power.

Data

The amount of data the modern world generates on a daily basis is staggering. From the organized tables of spreadsheets to the storage of photos, video, and audio recordings, modern businesses create an almost overwhelming avalanche of data that is ripe for use in analytics programs.

Let's try to quantify the amount of data that exists in the world. We'll begin with an estimate made by Google's then-CEO Eric Schmidt in 2010. At a technology conference, Schmidt estimated that the sum total of all of the stored knowledge created by the world

FIGURE 1.1 Analytics is made possible by modern data, storage, and computing capabilities.



at that point in time was approximately 5 exabytes. To give that a little perspective, the file containing the text of this chapter is around 100 kilobytes. So, Schmidt's estimate is that the world in 2010 had total knowledge that is about the size of 50,000,000,000,000 (that's 50 trillion!) copies of this book chapter. That's a staggering number, but it's only the beginning of our journey.

Now fast-forward just two years to 2012. In that year, researchers estimated that the total amount of stored data in the world had grown to 1,000 exabytes (or one zettabyte). Remember, Schmidt's estimate of 5 exabytes was made only two years earlier. In just two years, the total amount of stored data in the world grew by a factor of 200! But we're still not finished!

In the year 2020, IDC estimates that the world created 59 zettabytes (or 59,000 exabytes) of new information. Compare that to Schmidt's estimate of the world having a total of 5 exabytes of stored information in 2010. If you do the math, you'll discover that this means that on any given day in the modern era, the world generates an amount of brand-new data that is approximately 32 times the sum total of all information created from the dawn of civilization until 2010! Now, *that* is a staggering amount of data!

From an analytics perspective, this trove of data is a gold mine of untapped potential.

Storage

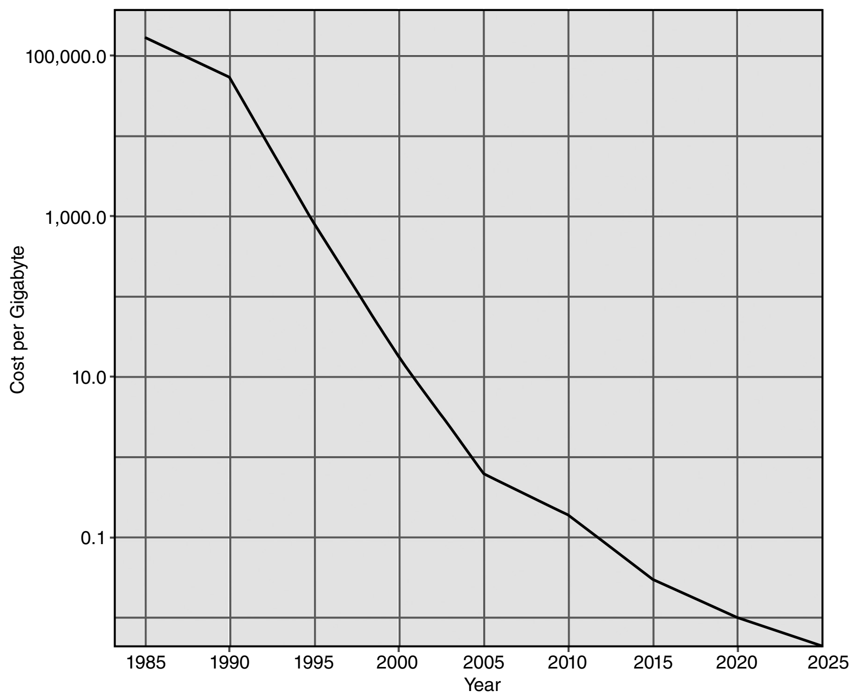
The second key trend driving the growth of analytics programs is the increased availability of storage at rapidly decreasing costs. Table 1.1 shows the cost of storing a gigabyte of data in different years using magnetic hard drives.

Figure 1.2 shows the same data plotted as a line graph on a logarithmic scale. This visualization clearly demonstrates the fact that storage costs have plummeted to the point

TABLE 1.1 Gigabyte Storage Costs over Time

Year	Cost per GB
1985	$169,900
1990	$53,940
1995	$799
2000	$17.50
2005	$0.62
2010	$0.19
2015	$0.03
2020	$0.01
2025	$0.01

FIGURE 1.2 Storage costs have decreased over time.



where storage is almost free and businesses can afford to retain data for analysis in ways that they never have before.

Computing Power

In 1975, Gordon Moore, one of the co-founders of Intel Corporation, made a prediction that computing technology would continue to advance so quickly that manufacturers would be able to double the number of components placed on an integrated circuit every two years. Remarkably, that prediction has stood the test of time and remains accurate today.

Commonly referred to as *Moore's Law*, this prediction is often loosely interpreted to mean that we will double the amount of computing power on a single device every two years. That trend has benefited many different technology-enabled fields, among them the world of analytics.

In the early days of analytics, computing power was costly and difficult to come by. Organizations with advanced analytics needs purchased massive supercomputers to analyze their data, but those supercomputers were scarce resources. Analysts fortunate enough to work in an organization that possessed a supercomputer had to justify their requests for small slices of time when they could use the powerful machines.

Today, the effects of Moore's Law have democratized computing. Most employees in an organization now have enough computing power sitting on their desks to perform a wide variety of analytic tasks. If they require more powerful computing resources, cloud services allow them to rent massive banks of computers at very low cost. Even better, those resources are charged at hourly rates, and analysts pay only for the computing time that they actually use.

These three trends—the massive volume of data generated by our businesses on a daily basis, the availability of inexpensive storage to retain that data, and the cloud's promise of virtually infinite computing power—come together to create fertile ground for data analytics.

Careers in Analytics

As businesses try to keep up with these trends, hiring managers find themselves struggling to identify, recruit, and retain talented analytics professionals. This presents a supply-and-demand situation that is problematic for businesses but excellent news for job candidates seeking to break into the field.

In a 2024 survey of executive leaders, Gartner found that 58 percent of CEOs and 62 percent of CFOs believe that artificial intelligence will be the trend with the most significant impact on their industries over the next three years. This will inevitably lead to increased demand for hiring analytics professionals, a fact that was confirmed by the World

TABLE 1.2 Highest-Growth Occupations

Rank	Occupation
1	AI and machine learning (ML) specialists
2	Sustainability specialists
3	Business intelligence analysts
4	Information security analysts
5	FinTech engineers
6	Data analysts and scientists
7	Robotics engineers
8	Big Data specialists
9	Agricultural equipment operators
10	Digital transformation specialists

Economic Forum in their 2023 Future of Jobs Report. That study listed occupations with the highest rate of new job creation. The results, shown in Table 1.2, found that AI and machine learning specialists are the most in-demand of any career field, with three other analytics-related fields in the top ten.

The future is bright. There's no reason to anticipate a reduction in this demand any time soon. It's the right time to enter the exciting field of data analytics!

The Analytics Process

Analysts working with data move through a series of different steps as they seek to gain business value from their organization's data. Figure 1.3 illustrates the process they move through as they acquire new data, clean and manipulate that data,

FIGURE 1.3 The analytics process.

analyze it, create visualizations, and then report and communicate their results to business leaders.

Data Acquisition

Analysts work with a wide variety of data, using data sources generated by the business itself or obtained from external sources. For example, data analysts might look at their own organization's sales data (an internal source) and augment it with census data (an external source) as they try to identify new potential markets for their firm's products and services.

In Chapter 3, "Understanding Data," you'll learn about the different data types, data structures, and file formats that analysts might encounter as they carry out data acquisition tasks. In Chapter 4, "Databases and Data Acquisition," you'll learn about the techniques used to collect this data and integrate it with existing systems as well as the use of relational databases to store, maintain, and query those datasets.

Cleaning and Manipulation

In an ideal world, we'd acquire data from internal and external sources and then simply pull it directly into our analysis. Unfortunately, the world of data is far from ideal, and you'll quickly discover (if you haven't already!) that datasets often contain errors, are missing crucial values, or come in a format that simply makes analysis difficult. Analysts spend a large portion of their time cleaning and manipulating data to get it ready for transformation. In fact, many analytics professionals estimate that cleaning and manipulation work consumes 80 percent of the time spent on most analytics projects!

In Chapter 5, "Data Quality," you'll learn more about the cleaning and manipulation work performed by data analysts. You'll discover the common reasons for cleansing and profiling datasets and different data manipulation methods. You'll also learn about the importance of data quality control and techniques you can use to improve the quality of your data.

Analysis

Once you have clean datasets in hand, you're ready to begin analyzing your data. This work typically begins with a process known as *exploratory data analysis* (EDA), which uses simple statistical techniques to summarize a dataset and draw high-level conclusions. EDA creates hypotheses that analysts may further explore using the techniques of machine learning and artificial intelligence.

In Chapter 6, "Data Analysis and Statistics," you'll learn about the tools and techniques of data analysis. You'll learn how to use descriptive statistics to perform EDA. You'll also learn about the processes used to continue analyses, including clarifying business questions, identifying data sources, and applying analytic techniques.

Visualization

The old adage "a picture is worth a thousand words" is as true in the world of analytics as it is in other aspects of life. The human mind excels at processing visual information and isn't so good at handling large quantities of numeric data.

You've already seen this at play once in this chapter. Table 1.1 presented a set of data points on the cost of storage. Looking at that table, you could tell that the cost of storage decreased over time, but you probably had to do a little thinking to reach that conclusion. A quick look at the same data visualized in Figure 1.2 likely led you to the same conclusion without all the mental gymnastics.

The storage dataset was fairly simple, however. Figure 1.4 shows you an excerpt from a 51-row dataset containing the average college tuition in each state. Can you quickly get a sense of the trend from state to state by looking at that data? It's probably not so easy for you.

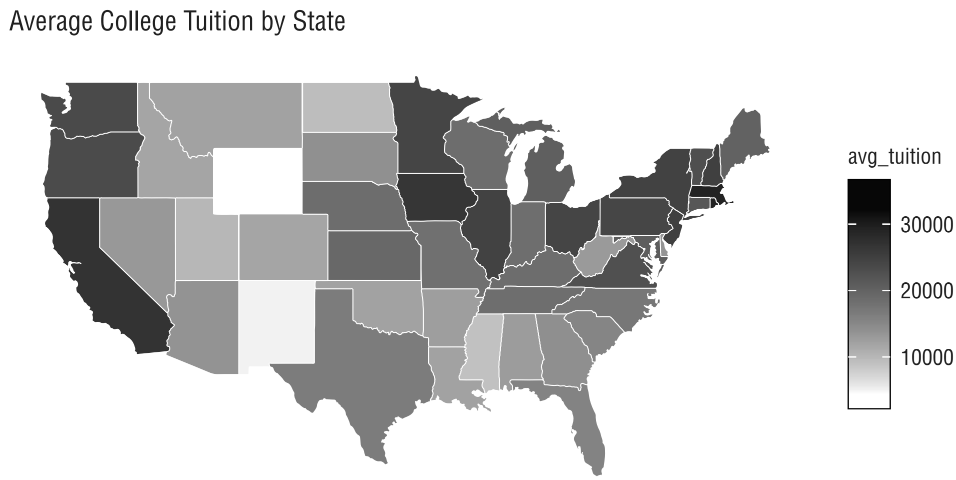
Figure 1.5 presents the same data in a map-based visualization. Darker shades represent higher tuition costs. We'll bet that you can draw conclusions from this visualization much more quickly than you can from the raw data!

In Chapter 7, "Data Visualization with Reports and Dashboards," you'll learn about different types of data visualizations, including line charts, histograms, infographics, and more. You'll discover how to select a visualization method appropriate for your needs and use it to tell the story of your data.

FIGURE 1.4 Table of college tuition data.

	state	avg_tuition	region
1	AK	19610.00	alaska
2	AL	13736.33	alabama
3	AR	13637.44	arkansas
4	AZ	14760.17	arizona
5	CA	27036.34	california
6	CO	12900.64	colorado
7	CT	22108.86	connecticut
8	DC	36009.50	district of columbia
9	DE	14592.67	delaware
10	FL	16639.59	florida
11	GA	15381.37	georgia
12	HI	10884.00	hawaii
13	IA	26735.68	iowa
14	ID	12851.29	idaho
15	IL	24911.74	illinois
16	IN	22677.10	indiana
17	KS	19973.68	kansas

FIGURE 1.5 Visualization of college tuition data.



Reporting and Communication

Although visualizations are very useful, they often can't stand alone. In most cases, you'll need to provide business leaders with multiple visualizations as well as supporting text to help communicate the story of your data. That's where *reports* and *dashboards* enter the picture. Reports provide the reader with textual analysis and supporting data in tabular and/or visualization form. They're an extremely common work product in the field of data analytics.

Reports, however, only present a point-in-time analysis. Business leaders often want to monitor business activities in real time using visualizations. Dashboards provide this real-time look at an organization's data using continuously updated visualizations.

Chapter 7 covers reporting and communication in more detail. You'll see examples of different types of reports and dashboards. You'll discover how to translate business requirements into appropriate reporting tools and how to design effective reports and dashboards.

The Analytics Process Is Iterative

While we describe the steps of the analytics process as a series of sequential actions, it's more accurate to think of them as a set of interrelated actions that may be revisited frequently while working with a dataset.

For example, an analyst reviewing a visualization may notice unusual data points that don't seem to belong in the dataset, causing them to return to the data cleaning stage and rerun their analysis with the newly cleaned dataset. Similarly, an analyst running an analysis might discover that their analysis would be enriched by adding another source of data, causing them to return to the data acquisition stage.

This process is meant to help you understand the different activities that take place during a data analysis effort and the approximate order in which they typically occur. You shouldn't view it as a rigid process, but rather as a rough guide.

Analytics Techniques

Analysts use a variety of techniques to draw conclusions from the data at their disposal. To help you understand the purpose of different types of analysis, we often group these techniques into categories based on the purpose of the analysis and/or the nature of the tool. Let's take a look at the major categories of analytics techniques.

Descriptive Analytics

Descriptive analytics uses statistics to describe your data. For example, if you perform descriptive analytics on your customer records, you might ask questions like, what proportion of your customers are female? And how many of them are repeat customers?

You can perform descriptive analytics using very basic analysis tools, including simple descriptive statistics and analytic tools. You'll learn more about the use of statistics in descriptive analytics in Chapter 6.

Inferential Analytics

Inferential analytics allows analysts to make generalizations about a population based on a sample of data. For instance, if you survey a subset of your customers to learn about their satisfaction, you can use inferential analytics to draw conclusions about the entire customer base. This type of analysis often involves hypothesis testing, confidence intervals, and regression models to determine relationships, test predictions, and assess whether observed patterns are statistically significant. By doing so, inferential analytics helps analysts make informed decisions about broader populations without needing to examine every single individual.

Predictive Analytics

Predictive analytics seek to use your existing data to predict future events. For example, if you have a dataset on how your customers respond to direct mail, you might use that dataset to build a model that predicts how individual customers will respond to a specific future mailing. That might help you tweak that mailing to improve the response rate by changing the day you send it, altering the content of the message, or even making seemingly minor changes like altering the font size or paper color.

Predictive analytics programs rely on the use of advanced statistical tools and specialized artificial intelligence, machine learning, and deep learning techniques.

Prescriptive Analytics

Prescriptive analytics seek to optimize behavior by simulating many scenarios. For example, if you want to determine the best way to allocate your marketing dollars, you might run different simulations of consumer response and then use algorithms to prescribe your behavior in that context. Similarly, you might use prescriptive analytics to optimize the performance of an automated manufacturing process.

Machine Learning, Artificial Intelligence, and Deep Learning

The work of analytics is intellectually and computationally demanding. Fortunately, you don't always have to do this work yourself; you can rely on automated techniques to help you unlock the hidden value in your data.

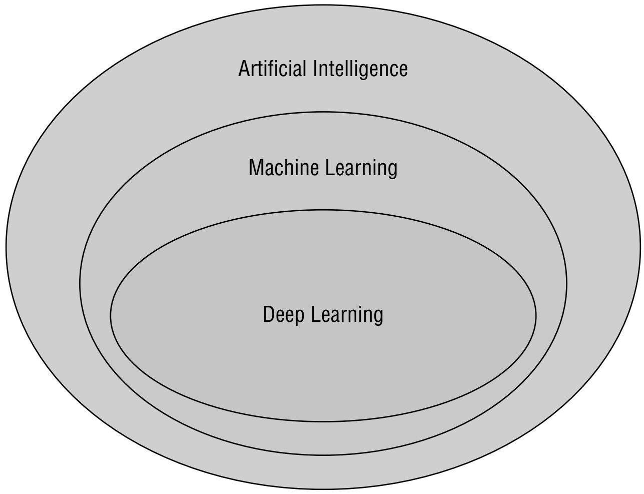
Machine learning uses algorithms to discover knowledge in your datasets that you can then apply to help you make informed decisions about the future. That's true regardless of the specific subject matter expertise where you're working, as machine learning has applications across a wide variety of fields. For example, here are some cases where machine learning commonly adds value:

- Segmenting customers and determining the marketing messages that will appeal to different customer groups
- Discovering anomalies in system and application logs that may be indicative of a cybersecurity incident
- Forecasting product sales based on market and environmental conditions
- Recommending the next movie that a customer might wish to watch based on their past activity and the preferences of similar customers
- Setting prices for hotel rooms far in advance based on forecasted demand

Of course, those are just a few examples. Machine learning can bring value to almost every field where discovering previously unknown knowledge is useful—and we challenge you to think of a field where knowledge doesn't offer an advantage!

As we move through the world, we hear the terms *artificial intelligence, machine learning*, and *deep learning* being used almost interchangeably to describe any sort of technique where computers are working with data. Now that you're entering the world of data, it's important to have a more precise understanding of these terms.

FIGURE 1.6 The relationship between artificial intelligence, machine learning, and deep learning.



Artificial intelligence (AI) includes any type of technique where you are attempting to get a computer system to imitate human behavior. As the name implies, you are trying to ask computer systems to artificially behave as if they were intelligent. Now, of course, it's not possible for a modern computer to function at the level of complex reasoning found in the human mind, but you can try to mimic some small portions of human behavior and judgment.

Machine learning (ML) is a subset of AI techniques. ML techniques attempt to apply statistics to data problems in an effort to discover new knowledge. Or, in other terms, ML techniques are AI techniques designed to learn.

Deep learning is a further subdivision of machine learning that uses quite complex techniques, known as neural networks, to discover knowledge in a particular way. It is a highly specialized subfield of machine learning that is most commonly used for image, video, and sound analysis.

Figure 1.6 shows the relationships between these fields.

Generative AI

Generative AI is a rapidly growing branch of artificial intelligence that focuses on creating new content, such as text, images, music, and even videos, rather than simply analyzing or interpreting existing data. If you've used tools like ChatGPT, Claude, or Midjourney, you've used generative AI.

At the heart of generative AI are *foundational models*, which are large, pre-trained deep learning models capable of tackling a wide range of tasks. These models are trained on massive datasets to learn general patterns and structures, allowing them to be fine-tuned for specific applications with minimal additional training.

A key category of foundational models is *large language models (LLMs)*, which specialize in processing and generating human language. LLMs are trained on extensive corpora of text to understand the nuances of language, such as grammar, context, and semantics. For example, models like GPT-4 (Generative Pre-trained Transformer 4) can perform tasks ranging from drafting emails and summarizing reports to generating creative writing or answering complex questions. LLMs use advanced deep learning techniques to produce text that often feels human-like, making them indispensable in fields such as customer support, content creation, and coding.

Generative AI builds upon the principles of *natural language processing (NLP)*, a field dedicated to enabling computers to understand and interact with human language. NLP techniques allow generative AI models to parse, interpret, and generate meaningful text. While traditional NLP tasks focused on narrower objectives—such as translating text or identifying sentiment—modern generative AI expands the scope to include creating entirely new and original content.

Robotic Process Automation

Robotic process automation (RPA) is a technology that uses software robots—or "bots"—to automate repetitive, rule-based tasks traditionally performed by humans. These bots are programmed to follow a predefined set of instructions, interacting with digital systems just like a person would, but with greater speed and accuracy. RPA excels at tasks such as data entry, form processing, and system integration, making it an indispensable tool for improving efficiency and reducing human error in business operations.

RPA can assist data analysts by automating time-consuming processes, enabling us to focus on more strategic and value-added work. For instance, RPA can streamline data collection from multiple sources, consolidate the information into a central database, and prepare it for analysis. This is particularly useful in organizations that rely on many legacy systems or deal with large volumes of data spread across different platforms. RPA can automate these tedious tasks and accelerate the analytics workflow.

One of the most impactful applications of RPA in analytics is automated reporting. Generating reports often requires pulling data from various systems, organizing it into the desired format, and applying specific business rules. RPA bots can easily perform these monotonous tasks, ensuring reports are delivered consistently and on time. For example, an RPA bot can automatically extract sales data from a customer relationship management (CRM) system, update a spreadsheet, apply calculations, and generate a report summarizing key performance metrics—all without any manual intervention. This not only saves time but also ensures that stakeholders have access to accurate, up-to-date information for decision-making.

Data Governance

In the beginning of this chapter, we discussed the three major forces that have come together to create the Golden Age of Analytics: data, storage, and computing. In Figure 1.1, we illustrated how those three forces support modern analytics programs. However, there is one element of that figure that we haven't yet discussed. Notice that there is a slab of stone that supports the three pillars of analytics. This slab represents the important role of *data governance* in analytics programs. Without strong governance, analytics programs can't function effectively.

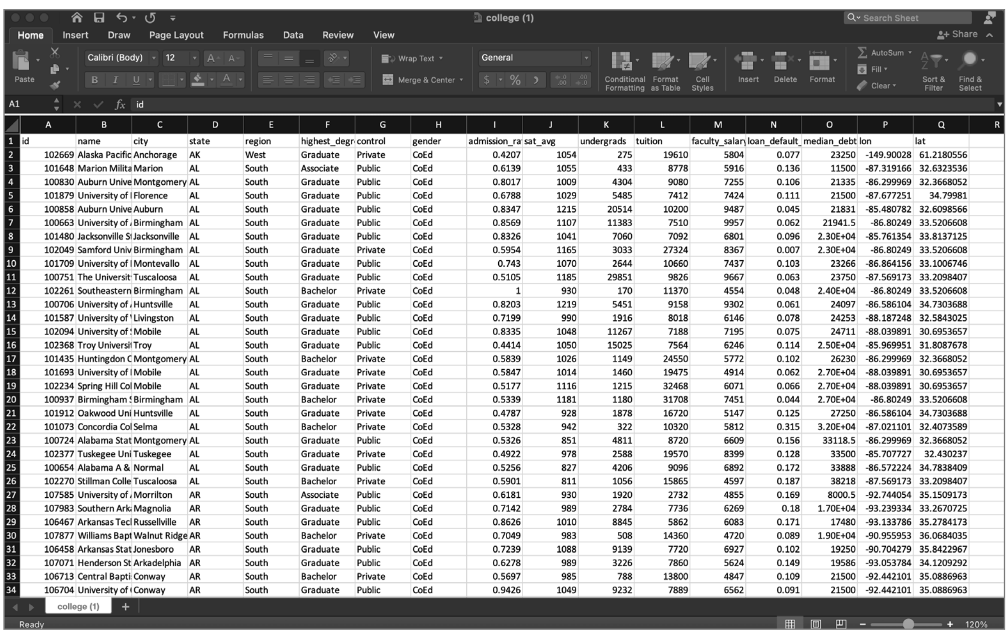
Data governance programs ensure that the organization has high-quality data and is able to effectively control that data. In Chapter 8, "Data Governance," you'll learn the major concepts of data governance and how organizations use *master data management (MDM)* programs to maintain and improve the quality of their data.

Analytics Tools

Software helps analysts work through each one of the phases of the analytics process. These tools automate much of the heavy lifting of data analysis, improving the analyst's ability to acquire, clean, manipulate, visualize, and analyze data. They also provide invaluable assistance in reporting and communicating results.

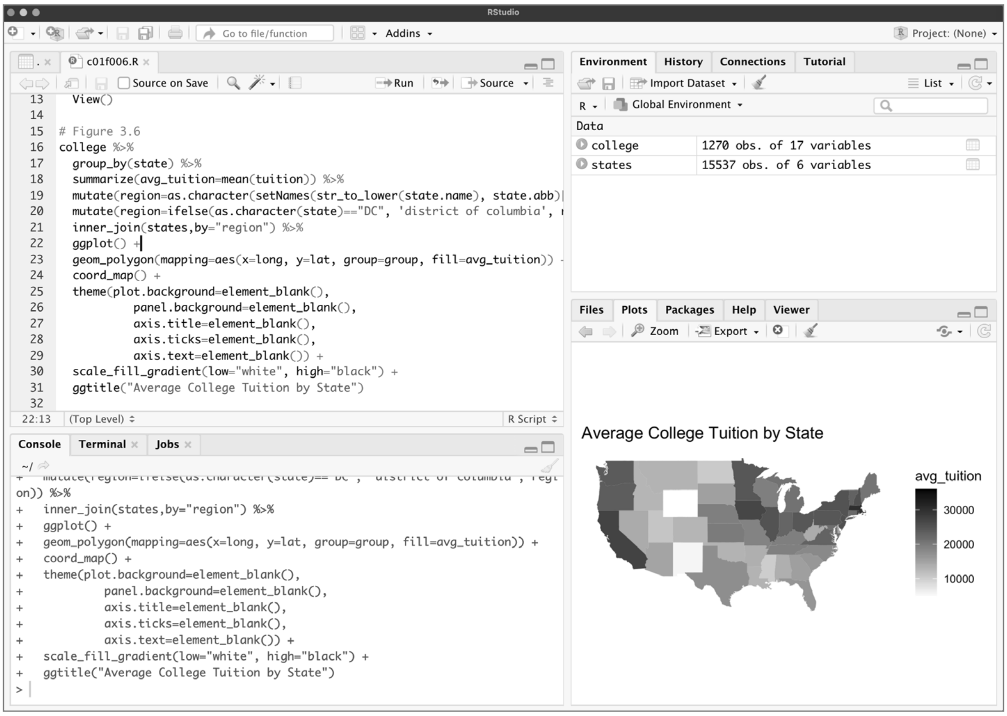
Some of these tools are well known to most computer users. For example, people are generally familiar with spreadsheet tools such as Microsoft Excel or Google Sheets. Figure 1.7 shows an example of the college dataset used to create Figures 1.4 and 1.5 loaded in Excel.

FIGURE 1.7 Data analysis in Microsoft Excel.



id	name	city	state	region	highest_degr	control	gender	admission_ra	sat_avg	undergrads	tuition	faculty_salary	loan_default_	median_debt	lon	lat
102669	Alaska Pacific	Anchorage	AK	West	Graduate	Private	CoEd	0.4207	1054	275	19610	5804	0.077	23250	-149.90028	61.2180556
101648	Marion Milita	Marion	AL	South	Associate	Public	CoEd	0.6139	1055	433	8778	5916	0.136	11500	-87.319166	32.6323536
100830	Auburn Unive	Montgomery	AL	South	Graduate	Public	CoEd	0.8017	1009	4304	9080	7255	0.106	21335	-86.299969	32.3668052
101879	University of I	Florence	AL	South	Graduate	Public	CoEd	0.6788	1029	5485	7412	7424	0.111	21500	-87.677251	34.79981
100858	Auburn Unive	Auburn	AL	South	Graduate	Public	CoEd	0.8347	1215	20514	10200	9487	0.045	21831	-85.480782	32.6098566
100663	University of /	Birmingham	AL	South	Graduate	Public	CoEd	0.8569	1107	11383	7510	9957	0.062	21941.5	-86.80249	33.5206608
101480	Jacksonville S	Jacksonville	AL	South	Graduate	Public	CoEd	0.8326	1041	7060	7092	6801	0.096	2.30E+04	-85.761354	33.8137125
102049	Samford Univ	Birmingham	AL	South	Graduate	Private	CoEd	0.5954	1165	3033	27324	8367	0.007	2.30E+04	-86.80249	33.5206608
101709	University of I	Montevallo	AL	South	Graduate	Public	CoEd	0.743	1070	2644	10660	7437	0.103	23266	-86.864156	33.1006746
100751	The Universit	Tuscaloosa	AL	South	Graduate	Public	CoEd	0.5105	1185	29851	9826	9667	0.063	23750	-87.569173	33.2098407
102261	Southeastern	Birmingham	AL	South	Bachelor	Private	CoEd	1	930	170	11370	4554	0.048	2.40E+04	-86.80249	33.5206608
100706	University of /	Huntsville	AL	South	Graduate	Public	CoEd	0.8203	1219	5451	9158	9302	0.061	24097	-86.586104	34.7303688
101587	University of '	Livingston	AL	South	Graduate	Public	CoEd	0.7199	990	1916	8018	6146	0.078	24253	-88.187248	32.5843025
102094	University of !	Mobile	AL	South	Graduate	Public	CoEd	0.8335	1048	11267	7188	7195	0.075	24711	-88.039891	30.6953657
102368	Troy Universit	Troy	AL	South	Graduate	Public	CoEd	0.4414	1050	15025	7564	6246	0.114	2.50E+04	-85.969951	31.8087678
101435	Huntingdon C	Montgomery	AL	South	Bachelor	Private	CoEd	0.5839	1026	1149	24550	5772	0.102	26230	-86.299969	32.3668052
101693	University of I	Mobile	AL	South	Graduate	Private	CoEd	0.5847	1014	1460	19475	4914	0.062	2.70E+04	-88.039891	30.6953657
102234	Spring Hill Col	Mobile	AL	South	Graduate	Private	CoEd	0.5177	1116	1215	32468	6071	0.066	2.70E+04	-88.039891	30.6953657
100937	Birmingham !	Birmingham	AL	South	Bachelor	Private	CoEd	0.5339	1181	1180	31708	7451	0.044	2.70E+04	-86.80249	33.5206608
101912	Oakwood Uni	Huntsville	AL	South	Graduate	Private	CoEd	0.4787	928	1878	16720	5147	0.125	27250	-86.586104	34.7303688
101073	Concordia Col	Selma	AL	South	Bachelor	Private	CoEd	0.5328	942	322	10320	5812	0.315	3.20E+04	-87.021101	32.4073589
100724	Alabama Stat	Montgomery	AL	South	Graduate	Public	CoEd	0.5326	851	4811	8720	6609	0.156	33118.5	-86.299969	32.3668052
102377	Tuskegee Uni	Tuskegee	AL	South	Graduate	Private	CoEd	0.4922	978	2588	19570	8399	0.128	33500	-85.707727	32.430237
100654	Alabama A &	Normal	AL	South	Graduate	Public	CoEd	0.5256	827	4206	9096	6892	0.172	33888	-86.572224	34.7838409
102270	Stillman Colle	Tuscaloosa	AL	South	Bachelor	Private	CoEd	0.5901	811	1056	15865	4597	0.187	38218	-87.569173	33.2098407
107585	University of /	Morrilton	AR	South	Associate	Public	CoEd	0.6181	930	1920	2732	4855	0.169	8000.5	-92.744054	35.1509173
107983	Southern Ark	Magnolia	AR	South	Graduate	Public	CoEd	0.7142	989	2784	7736	6269	0.18	1.70E+04	-93.239334	33.2670725
106467	Arkansas Tecl	Russellville	AR	South	Graduate	Public	CoEd	0.8626	1010	8845	5862	6083	0.171	17480	-93.133786	35.2784173
107877	Williams Bapt	Walnut Ridge	AR	South	Bachelor	Private	CoEd	0.7049	983	508	14360	4720	0.089	1.90E+04	-90.955953	36.0684035
106458	Arkansas Stat	Jonesboro	AR	South	Graduate	Public	CoEd	0.7239	1088	9139	7720	6927	0.102	19250	-90.704279	35.8422967
107071	Henderson St	Arkadelphia	AR	South	Graduate	Public	CoEd	0.6278	989	3226	7860	5624	0.149	19586	-93.053784	34.1209292
106713	Central Bapti	Conway	AR	South	Bachelor	Private	CoEd	0.5697	985	788	13800	4847	0.109	21500	-92.442101	35.0886963
106704	University of (	Conway	AR	South	Graduate	Public	CoEd	0.9426	1049	9232	7889	6562	0.091	21500	-92.442101	35.0886963

FIGURE 1.8 Data analysis in RStudio.



Other analytics tools require more advanced skills. For example, the R programming language is designed to provide analysts with direct access to their data, but it requires learning some basic coding skills. Figure 1.8 shows the RStudio integrated development environment with the code used to create Figure 1.5.

You'll likely work with several different tools in your work as a data analyst. Your choice of tools will depend on the work at hand, the standards used by your organization, and the software licenses available to you. We'll discuss many common analytics tools in Chapter 2, "Data Analytics Tools."

Exam Tip

The bad news is that the Data+ exam covers more than 20 different analytics tools that you'll need to understand to answer test questions. The good news is that you won't need deep knowledge of each of these tools. We'll explore everything that you need to know in Chapter 2.

Summary

Analytics programs allow businesses to access the untapped value locked within their data. Today, many organizations recognize the potential value of this work but are still in the early stages of developing their analytics programs. These programs, driven by the unprecedented availability of data, the rapidly decreasing cost of storage, and the maturation of cloud computing, promise to create significant opportunities for businesses and, in turn, for data professionals skilled in the tools and techniques of analytics.

As analysts develop analytic work products, they generally move through a series of stages. Their work begins with the acquisition of data from internal and external sources and continues with the cleaning and manipulation of that data. Once data is in a suitable form, data professionals apply analytic techniques to draw conclusions from their data, create visualizations to depict the story of their data, and develop reports and dashboards to effectively communicate the results of their work to business leaders.

Generative AI and robotic process automation (RPA) are two emerging technologies reshaping the analytics landscape. Generative AI leverages foundational models to create new content, enabling data professionals to draft reports, summarize insights, and enhance creative tasks with unprecedented efficiency. RPA complements these efforts by automating repetitive, rule-based processes, such as data collection and report generation, freeing analysts to focus on more strategic work. Together, these technologies empower businesses to accelerate analytics workflows, improve accuracy, and unlock new opportunities for innovation in data-driven decision-making.