

List of Tables and Illustrations

- Figure 1.1 The recent wave of economic globalization has outstripped the prior one
- Figure 1.2 Global cross-border capital flows, 1990–2016E (E as estimate)
- Figure 1.3 Related terms evidenced in the number of searches receives in Google Trends between 2018 and 2019
- Figure 1.4 Various forms of capitalism globally
- Figure 1.5 The distribution of the US Bureau of Labor Statistics 2010 occupational employment (a total of 702) over the probability of being replaced by AI, along with the share in low, medium and high probability categories
- Figure 2.1 World inequality with its share of income accruing at top 1%
- Figure 2.2 Using a line (a) and radar (b) chart to show “intangibles” references in scientific journals
- Figure 2.3 Chernoff fishes for depicting a firm’s investment in moving to cloud services
- Figure 2.4 A mind map extracted from the US News & World Report’s best hospital ranking methodology
- Figure 2.5 Network structures discovered by wiring pattern mining
- Figure 3.1 An illustrator of revised WeWork’s prospectus “SUMMARY” based on Maslow’s hierarchy of needs 1.
- Figure 3.2 Capital classification described in economic literature
- Figure 3.3 Monetary Bearings: From the gold standard to unanchored monetary era, and to bitcoin
- Figure 3.4 The US stock market’s four meltdowns within 10 days during the COVID-19 epidemic
- Figure 3.5 What is human capital
- Figure 3.6 When spiral timelines combined with a geographical map
- Figure 3.7 Some available candidate data sources (LinkedIn not included)
- Figure 3.8 Corpus description of analyzing social clout capital
- Figure 3.9 Two venture-capital groups’ networks with/without topic classification

- Figure 4.1 Data nature and data science
- Figure 4.2 Human brain and artificial neural networks (mapping human brain)
- Figure 4.3 A kind of data integration architecture of big data brain
- Figure 4.4 A river diagram along the timeline
- Figure 4.5 Virtual time and actual time
- Figure 5.1 Complicated terms related to data rights
- Figure 5.2 Signposts (in pros and cons) to the automotive future
- Figure 5.3 A comparison of 2015 and 2025 on EU prototypical OEM under four scenarios given by Deloitte
- Figure 5.4 The timeline of a post on Raddi R/nosleep titled “My dead girlfriend keeps messaging me ...”
- Figure 6.1 An example of Stitch Fix’s data and personalized service
- Figure 6.2 A tongue image with a comparison of its parameters before and after annotating
- Figure 6.3 The wordles of all words relating to the titles of tables, charts, and boxes in the World Trade Report from 2003 to 2017
- Figure 6.4 A stream graph of indicating theme evolution relating to the titles of tables, charts, and boxes in the World Trade Report from 2003 to 2017
- Figure 6.5 A correlation of different years’ reports with Top/Bottom 10 similar sentence pairs that come from the titles of tables, charts, and boxes in the World Trade Report from 2003 to 2017
- Figure 6.6 A visual description of the Allergy Screener with the data relating to tongue diagnosis
- Figure 6.7 Capital in five countries from 1700 to 2010
- Figure 7.1 Important job skills for data scientists
- Figure 9.1 “Yu-Gong” (i.e., 《禹贡》) is the results map of an earlier census for landscapes in ancient China
- Figure 9.2 The simplified experimental evolution of Escherichia coli
- Figure 12.1 The difference between self-governing closeness and openness of data
- Figure 12.2 A diagram of interpreting self-governing openness of data: DataBox
- Figure 13.1 A possible path to data as currency within the financial sector
- Figure 13.2 Dimensions of data governance within a single enterprise
- Figure 13.3 Three types of chart to visualize year-over-year (YoY) vs. convergence or divergence bands
- Table 1.1 Types of general purpose technologies
- Table 1.2 Globalization: Then (twentieth century) vs. now (twenty-first century)
- Table 1.3 The six possible outcomes in the game involving bluffing
- Table 2.1 Sequential income data
- Table 2.2 Common Types of Capital
- Table 2.3 An example of education investment and economic returns
- Table 3.1 An illustration of Day-by-Day data

Table 3.2 Market reactions during virus emergencies: S&P 500 performance over the course of each outbreak

Table 3.3 Post topic classification among two groups

Table 4.1 Using a non-temporal database for the records from a fabricated short biography

Table 4.2 A comparison of using the non-/bi-temporal database for a fabricated short biography

Table 6.1 Top and Bottom 10 Sentence pairs in the World Trade Reports

List of Case Study Examples

- Case 1.1 The game theory of “beauty and coin”.
- Case 3.1 An example of using EBITDA to dig the true value of companies.
- Case 3.2 Some paragraphs from WeWork’s prospectus “SUMMARY”.
- Case 3.3 Some paragraphs related to capital classification from Haskel and Westlake’s *Capitalism without Capital: The Rise of the Invisible Economy*.
- Case 3.4 Some paragraphs related to capital classification from Piketty’s *Capital in the 21 Century*.
- Case 3.5 A comparison of translating Bourdieu’s exposition of “capitalization and capital conversion”.
- Case 3.6 A comparison of describing data assets and intangibles.
- Case 3.7 An online post to support “bitcoin is not backed by nothing” on Unchained-Capital.com.
- Case 4.1 Some paragraphs related to digital memories from Mayer-Schönberger’s *Delete*.
- Case 4.2 A fabricated short biography.
- Case 5.1 A scenario of authorizing consumer data privacy.
- Case 5.2 Common types of automatically delivering information flows.
- Case 5.3 A theoretical case of data capital migration.
- Case 6.1 An example of instructing how to “Fair Use”.
- Case 7.1 Greenleaf dairy farmer left with a lagoon of dumped milk.
- Case 7.2 Two commonly used tax digital errand-runners.
- Case 11.1 A comparison of income tax and tax on capital.
- Case 12.1 Google’s Privacy Policy across two decades and 30 versions.

List of Explaining Data Sources

Data Source 2.1 *Literature Data* Literature data is any single part of a scientific paper (e.g., titles, abstracts, backgrounds, methods, results, conclusions, or the whole article) collected by searching of several specific databases (e.g., DBLP: Computer Science Bibliography can be accessed at <https://dblp.uni-trier.de>, MEDLINE/PubMed, ScienceDirect), library catalog services (e.g., Harvard's HOLLIS), or online facilities (e.g., Google Scholar) during a given period of time. Quantitative approaches to archives and texts developed by digital humanists have offered one such expansion. These approaches often treat literature as a text mine. Each paper included in the analysis can be classified as either a paper describing article type (e.g., a description, or a review of the literature) or other information: year of publication, journal title, specialty area, the field of study, and characteristics given by authors to data reuse.

Data Source 3.1 *Market Data* Market data is price and trade-related data for a financial instrument (e.g., equities, fixed-income products, derivatives, and currencies) reported by a trading venue such as a stock exchange.

Data Source 3.2 *Day-by-Day Data* Day-by-Day data is recorded in the computer systems according to human routine rhythm, such as stock trading on each closing price in a week as $\{C_1, C_2, \dots, C_7\}$. Table II.4 shows the day-by-day behavior data, namely a collection of behaviors which include each person's daily behavior. Trade-related data for a financial instrument reported by a trading venue such as a stock exchange. Market data allows traders and investors to know the latest price and see historical trends for instruments such as equities, fixed-income products, derivatives, and currencies.

Data Source 3.3 *Social Media Data* Social media data refers to all of the raw insights and information collected from individuals (e.g. prospects and customers) social media activity. Social media data tracks how individuals engage with a specific content or channels like Facebook, Twitter, Instagram, and LinkedIn.

- Data Source 4.1 *Event-Related Potential Data* An event-related potential (ERP) is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event. More formally, it is any stereotyped electrophysiological response to a stimulus. ERPs are measured by means of electroencephalography (EEG). ERP data's format depends on the manufacturers of EEG, such as Brain Products .eeg files and Neuroscan .cnt files.
- Data Source 4.2 *Functional Magnetic Resonance Imaging Data* Functional magnetic resonance imaging (fMRI) measures brain activity by detecting changes associated with blood flow. This technique relies on the fact that cerebral blood flow and neuronal activation are coupled. When an area of the brain is in use, blood flow to that region also increases. The primary form of fMRI uses the blood-oxygen-level dependent (BOLD) contrast, discovered by Seiji Ogawa in 1990. DICOM as a format for neuroimaging with fMRI.
- Data Source 4.3 *Eye Tracking Data* Eye tracking is the process of measuring either the point of gaze (where one is looking) or the motion of an eye relative to the head. The format of data generated by different eye trackers, the devices for measuring eye positions and eye movement, is various but mostly can be exported as the text.
- Data Source 4.4 *Positron Emission Tomography Data* A brain positron emission tomography (PET) scan is typically an outpatient procedure of imaging test that allows doctors to see how the brain is functioning. The scan captures images of the activity of the brain after radioactive "tracers" have been absorbed into the bloodstream. These tracers are "attached" to compounds like glucose (sugar). Glucose is the principal fuel of the brain. Active areas of the brain will be utilizing glucose at a higher rate than inactive areas. PETs' data format depends on their manufacturers.

List of Definitions

- Definition 2.1 *Production* Production is a process of combining various material inputs and immaterial inputs in order to make something for consumption (output).
- Definition 2.2 *Producer* A producer is any organization (e.g., corporations, industry bodies, government agencies) who plan, coordinates, and performs production.
- Definition 2.3 *Economic Resource* An economic resource is the (input) factors of production used in producing goods or providing services.
- Definition 2.4 *Investment* Investment is what happens when a producer either acquires a fixed asset or spends resources (money, effort, raw materials) to improve it.
- Definition 2.5 *Asset* An asset is an economic resource that is expected to provide a benefit over a period of time.
- Definition 2.6 *Fixed Asset* A fixed asset is an asset that results from using up resources in the process of its production.
- Definition 2.7 *Rate of Profit* The profit rate is the relative profitability of a whole investment project, but sometimes expresses the share of profits that may be relevant with income from capital.
- Definition 2.8 *Interest Rate* An interest rate is the amount of interest due per period, as a proportion of the amount lent, deposited, or borrowed (called the principal sum), which varies widely depending on the identity of the lender or borrower.
- Definition 2.9 *Rate of Return on Capital* The rate of return on capital measures the yield on capital over the course of a year regardless of its legal form (e.g., profits, interest, rents, dividends, royalties, capital gains, etc.), expressed as a percentage of the value of capital invested.
- Definition 2.10 *Confidence* Given an association rule that is an implication expression of the form $X \rightarrow Y$, where X and Y are disjoint itemsets, i.e., $X \cap Y = \emptyset$. Confidence determines how frequently items in Y appear in patterns that contain X .

Definition 2.11 *Mind Map* A mind map is a hierarchical diagram to be an image-like aid. It is often created around a single concept, drawn as an image in the center of a blank page, to which associated representations of keywords, ideas, tasks or other items in a radial way. Mind maps are considered to be a type of spider diagram.

Definition 3.1 *EBITDA* EBITDA is a metric that excludes expenses associated with debt by adding back interest expense and taxes to earnings.

Definition 3.2 *Intellectual Capital* Intellectual capital is the intangible value of a business, covering its people (human capital), the value relating to its relationships (relational capital), and everything that is left when the employees go home (structural capital), of which intellectual property is but one component. It is the sum of everything everybody in a company knows that gives it a competitive edge.

Definition 3.3 *Meaning Capital* Meaning capital involves several forms of capital (e.g., spiritual, psychological, and experiential capital), which more directly connect to an individual or a collective sense of purpose and what emotions are.

Definition 3.4 *Cross-Border Capital Flows* Cross-border capital flows, at the global level, can be calculated as the sum of capital inflows or the sum of capital outflows, excluding financial derivatives but including flows related to international banking transactions.

Definition 3.5 *Data Asset* A Data asset is the data asset as data exceeding a certain scale that is owned and controlled by a specific agent, collected from the agent's past transactions involved in information processes, and capable of bringing future economic benefits to the agent.

Definition 3.6 *Data Production* Data Production is a process of combining various analyzable data inputs in order to make something for consumption (output). Something for consumption (output) here means data product and the others.

Definition 3.7 *Data Producer* A data producer is any agent (e.g., corporations, industry bodies, government specifies, individuals) who has the ability to pursue adding value to raw data or other primary data products.

Definition 3.8 *Data Capital* Data capital, for a specific agent, can be calculated as the sum of add-valued data asset inflows or outflows.

Definition 3.9 *Digital Currency* The digital currency is the blanket term used to describe all digital or electronic, and intangible money, which includes both virtual currency and cryptocurrency. It can be regulated or unregulated and is used online.

Definition 3.10 *Virtual Currency* Virtual currencies are a type of digital currency, typically controlled by its creators and used and accepted among the members of specific virtual community.

Definition 3.11 *Cryptocurrency* "Crypto" in cryptocurrency refers to the fact that many encryption algorithms and cryptographic techniques are used to

ensure security across the network. This level of security also makes cryptocurrencies hard to counterfeit.

Definition 3.12 *Spatio-Temporal Visualization* The spatio-temporal visualization incorporates the legacy from the conventional cartography to track the kind of changes occurring over time, including: existential changes (e.g., appearance and disappearance), changes of spatial properties (e.g., location, sharp or/and size, orientation, altitude, height, gradient, and volume), and changes of thematic properties expressed through values of attributes (e.g., qualitative changes and changes of ordinal or numeric characteristics: increase and decrease). Of which, timeline is a graphical way of displaying a list of events in chronological order. Some timelines display as a straight (e.g., horizontal or vertical) line, while others can be a curve (e.g., a spire).

Definition 3.13 *Data credibility* Data credibility is the extent to which the good faith of a provider of data or source of data can be relied upon to ensure that the data really represents is what the data is supposed to represent, and that there is no intent to misrepresent what the data is supposed to represent.

Definition 4.1 *Atomic Data* Atomic data is an atomic state of information that cannot be broken down into smaller pieces or sources, such as unit sales (e.g., for a particular model of the washer), gross revenue (that cannot be calculated from anything even if being broken down by transaction; note that net revenue is based on multiple inputs and is not considered atomic data), and base salary (as a hard number that is not typically calculated from the others).

Definition 4.2 *Data Item* A data item describes an atomic data concerning a specific property at a certain time point. Values of a data item are not necessarily atomic. The complexity of a value (represented by symbols like numbers, texts, images, sounds, videos or others) depends on the complexity of the property and time component.

Definition 4.3 *Data Object* A data object is a unique location in a computer memory (i.e., a specific memory address) that contains the value of *something*. This *something* can be accessed by giving a unique name that is known as the identifier of the variable that contains the data object. Data objects are often composed of a collection of limited data items, which are identified by object (o), property (p) and time (t), while the value (v) is a function of o , p , and t , $v = F(o, p, t)$.

Definition 4.4 *Dataset* A dataset (or a data set) is a collection of data objects. Generally, data objects in a specific dataset are limited. Even though building a dataset that contains an infinite stream, there exist limited data objects at a certain (arbitrary) time point.

Definition 4.5 *Data* Data as a general concept refers to the fact that some existing information or knowledge is digitally coded in some form suitable for better usage or processing. The data can be represented as an atomic data, a data item, a data object, or a dataset arbitrarily. The size of the data is

measured in bytes (i.e., *B*). Think of *NULL* data is with an empty value, of which the size is zero *B*.

Definition 4.6 *Metadata* Metadata is the data that provides information about other data. There exist multiple distinct types of metadata, including descriptive metadata, structural metadata, administrative metadata, reference metadata, and statistical metadata.

Definition 4.7 *Data Tool* A data tool involves algorithms, data models, programs, and software, which can run in a computer system. Data tools belong to a specific data object designed for some purposes (e.g., an analytical purpose). The data tool itself is also data, which is usually used to process data and can also be processed by other data tools.

Definition 4.8 *Data Resource* As data organizations connected by logics and data areas assembled by data volumes reach a certain scale, the massive distinct datasets become a data resource.

Definition 4.9 *Data Nature* All data in cyberspace constitutes data nature, but the computer systems (or other IT physical facilities) as carriers do not belong to the part of data nature.

Definition 4.10 *Data Experiment* Data experiment is a scientific method to discovering how things (e.g., specific natural phenomena, or any rules in a designated dataset and/or data tool) in the universe work though the direct intervention and control of datasets and/or data tools. Three major types of data experiments include taking datasets, data tools, or simulated natural phenomena (e.g., natural disasters, urban or business operations) as the experimental object, respectively. Data experiment is characteristic that the experimental objects and tools of such a research approach are both composed of data: data verifies data through data.

Definition 4.11 *Repeatability* The measurement can be obtained with stated precision by the same team using the *same* measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials. For data experiments, this means that a researcher can reliably repeat her own computation.

Definition 4.12 *Replicability* The measurement can be obtained with stated precision by a different team using the *same* measurement procedure, the same measuring system, under the same operating conditions, in the same or a different location on multiple trials. For data experiments, this means that an independent group can obtain the same result using the author's own artifacts.

Definition 4.13 *Reproducibility* The measurement can be obtained with stated precision by a different team, a *different* measuring system, in a different location on multiple trials. For data experiments, this means that an independent group can obtain the same result using artifacts which they develop completely independently.

- Definition 4.14 *Raw Data* Raw data (sometimes called source data or atomic data) is data that has not been processed for use. Raw data that has undergone processing is sometimes referred to as cooked data.
- Definition 4.15 *Data Products* Data products (i.e., cooked data), such as models, intermediate results, and annotated training corpora, are the outcomes from data preparation, processing, and analysis (e.g., statistical analysis, data mining, and machine or deep learning). Data products also include visualizations and dashboards created by the artistic manual work of data scientists to assist in the interpretation of the analysis in an actionable way.
- Definition 4.16 *Stream Graph* A stream graph, as a variation of a stacked area graph, displays the changes in data (between the maximum and minimum) over time of different categories through the use of flowing, organic shapes that somewhat resemble a river-like stream.
- Definition 5.1 *Data Sovereign* Data sovereign is a data producer with the ability to defend data resources against any attack.
- Definition 5.2 *Data Migration* Data migration is a project by means of which data will be moved or copied from one environment to another, and removed or decommissioned in the source.
- Definition 5.3 *Data Capital Migration* Data capital migration involves multiple data sovereigns, which makes data capital flowing in one or more directions (and there may be numerous migrations taking place simultaneously). So, its representatives will be moved or copied from one environment to another, and removed or decommissioned in the source.
- Definition 6.1 *Feature Engineering* Feature engineering is the process of using domain knowledge to extract features from raw data via data technologies. These features can be used to improve the performance of algorithms.
- Definition 6.2 *Labeled Data* Labeled data is a group of samples that have been tagged with one or more labels. Labeling typically takes a set of unlabeled data and augments each piece of that unlabeled data with meaningful tags that are informative.
- Definition 6.3 *Wordle* A wordle is a visual depiction of the words contained in a piece of free text. The more frequent a particular word was within the source text, the bigger it's displayed in the wordle—font and color variation, as well as giving weight to particular words, which are positioned vertically or horizontally. It can help filter out both low-frequency and low-quality information to help a viewer understand content by just glancing it.
- Definition 6.4 *Barrier to Entry* Barriers to entry, an economic term, describe the existence of high startup costs or other obstacles that prevent new competitors from quickly entering an industry or area of business. Barriers to entry may be natural (high startup costs to drill a new oil well), created by governments (licensing fees or patents stand in the way), or by other companies (monopolists can buy or compete away startups).

Definition 12.1 *Data Jigsaw* Data jigsaw is a behavior that is to splice multiple data fragments or numerous times to restore the data to be integral.

Definition 12.2 *Algorithmic Bias* Algorithmic bias describes repeatable errors caused by algorithms and models that create unfair outcomes. Bias can emerge due to many factors, including but not limited to the design of the algorithm or the unintended or unanticipated use or decisions relating to the way data is collected, selected, or used to train the algorithm. Algorithmic bias can have impacts ranging from inadvertent privacy violations to reinforcing social prejudices of race, gender, sexuality, and ethnicity.

Index

A

ACM, 26

B

Bank for International Settlements (BIS), 5

B.C.E., 3, 4

Becket, 14, 15

Borgman, C.L., 18

C

Chandler, A.C., 19

Consumer price index (CPI), 5

D

Data capital, ix, x

Data economy, ix, 2

Data science, 2

Data speak for itself, x, 2

Data technology, 8

Disruptive innovation, 1

E

Engels, F., 16

European Union (EU), 9, 122

F

Foreign direct investment (FDI), 7

G

Gelman, A., 21

Generally Accepted Accounting Principles (GAAP), 50

General purpose technologies (GPT), 3, 4

Gross domestic product (GDP), 7

H

Harvard Business School (HBS), 23

I

Information technology (IT), 1, 10

Intellectual property (IP), 49

L

Laissez-faire, 11, 18

Leonelli, S., 18

M

Marx, K., x, 11, 12, 15, 16, 20

Massachusetts Institute of Technology (MIT), 21

Mayer-Schönberger, V., 24, 26, 29, 31, 83, 119, 120, 122, 123, 128, 132, 140, 145, 146, 179, 181, 219, 223, 230, 231, 234, 248, 257, 267, 292, 294, 303, 312, 327–331, 337, 339, 345, 351

Meng, Xiao-Li, 17, 129

Mintz, S.W., 14

N

Neal, L., 15

P

Piketty, T., 1, 19–21, 24, 27, 31, 35, 41–47, 66, 68, 77, 79, 91–94, 202, 203, 283, 292, 299, 301, 304, 305, 308–315, 321, 323–327, 329, 331, 333–335, 337, 339, 350, 356

R

Research and development (R&D), 47

S

Schumpeter, J.A., 12, 15
Science, Technology, Engineering, and Mathematics (STEM), 18
Small and medium-sized enterprises (SMEs), 6
Smith, A., 11, 15, 16, 18

T

Tirole, J.M., 20

U

United Kingdom (UK), 5
United States of America (USA), 5

W

Weber, M., 11, 18, 150, 204
World Top Incomes Database (WTID), 20
World Trade Organization (WTO), 9

Z

Zuboff, S., 23