Data Analysis for Decision-Making

Lecture Notes

September 17, 2024

Table of contents

List of figures

List of tables

Preface

About the notes

> 💡 A PDF version of these notes is available here.
>
> Please note that while the PDF contains the same content, it has not been optimized for PDF format. Therefore, some parts may not appear as intended.

- These notes aims to support my lecture at the HS Fresenius but are incomplete and no substitute for taking actively part in class.
- I appreciate you reading it, and I appreciate any comments.
- This is work in progress so please check for updates regularly.
- For making an appointment, you can use the online tool that you find on my private homepage: https://hubchev.github.io/
- These notes are published under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

About the author

Figure 1: Prof. Dr. Stephan Huber

I am a Professor of *International Economics and Data Science* at HS Fresenius, holding a Diploma in Economics from the University of Regensburg and a Doctoral Degree (summa cum laude) from the University of Trier. I completed postgraduate studies at the Interdisciplinary Graduate Center of Excellence at the Institute for Labor Law and Industrial Relations in the European Union (IAAEU) in Trier. Prior to my current position, I worked as a research assistant to Prof. Dr. Dr. h.c. Joachim Möller at the University of Regensburg, a post-doc at the Leibniz Institute for East and Southeast European Studies (IOS) in Regensburg, and a freelancer at Charles University in Prague.

Throughout my career, I have also worked as a lecturer at various institutions, including the TU Munich, the University of Regensburg, Saarland University, and the Universities of Applied

Sciences in Frankfurt and Augsburg. Additionally, I have had the opportunity to teach abroad for the University of Cordoba in Spain, the University of Perugia in Italy, and the Petra Christian University in Surabaya, Indonesia. My published work can be found in international journals such as the Canadian Journal of Economics and the Stata Journal. For more information on my work, please visit my private homepage at hubchev.github.io.

> **ⓘ Contact:**
>
> Prof. Dr. Stephan Huber
> Hochschule Fresenius für Wirtschaft & Medien GmbH
> Im MediaPark 4c
> 50670 Cologne
> Office: 4e OG-3
> Telefon: +49 221 973199-523
> Mail: stephan.huber@hs-fresenius.de
> Private homepage: www.hubchev.github.io
> Github: https://github.com/hubchev

I was always fascinated by data and statistics. For example, in 1992 I could name all soccer players in Germany's first division including how many goals they scored. Later, in 2003 I joined the introductory statistics course of Daniel Rösch. I learned among others that probabilities often play a role when analyzing data. I continued my data science journey with Harry Haupt's *Introductory Econometrics* course, where I studied the infamous Jeffrey M. Wooldridge [2002] textbook. It got me hooked and so I took all the courses Rolf Tschernig offered at his chair of Econometrics, where I became a tutor at the University of Regensburg and a research assistant of Joachim Möller. Despite everything we did had to do with how to make sense out of data, we never actually used the term *data science* which is also absent in the more 850 pages long textbook by Wooldridge [2002]. The book also remains silent about *machine learning* or *artificial intelligence*. These terms became popular only after I graduated. The *Harvard Business Review* article by Davenport and Patil [2012] who claimed that data scientist is "The Sexiest Job of the 21st Century" may have boosted the popularity.

The term "data scientist" has become remarkably popular, and many people are eager to adopt this title. Although I am a professor of *data science*, my professional identity is more like that of an applied, empirically-oriented international economist. My hesitation to adopt the title "data scientist" also stems from the deep respect I have developed through my interactions with econometricians and statisticians. Considering their in-depth expertise, I feel like a passionate amateur.

Ultimately, I poke around in data to find something interesting. Much like my ten-year-old younger self who analyzed soccer statistics to gain a deeper understanding of the sport. The only thing that has changed since then is that I know more promising methods and can efficiently use tools for data processing and data analysis.

About the Course

The course totals 125 hours over one semester at the Master's level, granting 5 ECTS points. It consists of 3 weekly contact hours (42 hours in total) and 83 hours of private study.

Abstract

This module provides essential skills for transforming data into actionable business insights. Upon completion of the module, students will be able to summarize the importance of analyzing business data and engage in discussions about different workflows for leveraging data analytics for new business trends. The module systematically develops skills to create plans for data collection and management. Students learn to recognize the challenges and opportunities of different quantitative empirical strategies. Decision principles, frameworks and tools, including decision trees and payoff tables, are covered in depth, focusing on the central role of decision support systems (DSS). Students will gain an in-depth understanding of the different roles in business analysis and a comprehensive overview of the entire data analysis workflow in a business context.

Learning outcomes/competences

After a successful completion of the module, the students are able to:

- summarize the significance of business data analysis for decision-making and demonstrate the ability to justify and articulate diverse workflows to convert data into actionable information,
- explain the role of data analytics to emerging trends in business and how organizations can use data analytics and decision support systems to solve problems,
- identify and contrast the competencies required to solve business problems with data and be able to assign the various tasks of a data scientific workflow to professionals with the appropriate profile,
- set up a plan for collecting, managing, analyzing, and applying data, reflectively applying quantitative methods,
- distinguish and discuss empirical research strategies to identify causal mechanisms, causes, and effects,
- identify the need of decision support systems and assess the possibilities of data-driven methods to improve decision-making of humans and organizations.

Module content

Introduction

- Significance of business data analysis for decision-making.
- Emerging trends: Evolution of computers and data processing, digitalization, artificial intelligence, machine learning, deep learning, big data, internet of things, cloud computing, and blockchain, industry 4.0, and remote working.
- The role of business analytics and intelligence in converting data into actionable information for decision-makers.
- Overview of various job roles for business analytics: data engineer, data analyst, machine learning engineer, business intelligence analyst, database administrator, data product manager, market research analyst, fraud analyst, …
- Analytics applications in strategic insights.
- Types of analytics: Descriptive, predictive, and prescriptive analytics.
- Workflows and Data science life cycles: OSEMN, DSLC, CRISP-DM, Kanban, TDSP, …

Data literacy competencies

- Types of data: Cross-section, panel, time-series, georeferenced, …
- Types of variables: Continuous, count, ordinal, categorial, …
- Conceptual framework: Knowledge and understanding of data and applications of data.
- Data collection: Identify, collect, and assess data.
- Data management: Organize, clean, convert, curate, and preserve data.
- Data evaluation: Plan, conduct, evaluate, and assess data analyses.
- Data application: Share, reflect, and evaluate results of analyses, comparing them with other findings while considering ethical issues and scientific standards.

Quantitative research design

- Techniques for measuring socio-economic and business phenomena.
- Strategies for identifying causes of effects and effects of causes.
- The fundamental problem of causal inference.
- Techniques to establish causation: matching, natural experiments, field experiments, and laboratory experiments.

Introduction to decision-making

- Overview of the principles guiding rational decision-making.
- Frameworks for structuring and representing the decision-making process.
- Utilization of decision trees, payoff tables, Lagrange multipliers, and expected utility theory.
- Exploration of bounded rationality in human decision-making behaviors.
- The necessity, concept, and evolution of

Assessment Methods and Criteria

Students complete this module with an **academic presentation**. The presentation takes place during the lecture period; the exact date is set by the lecturer. The presentation last for 10-15 minutes per student. In addition, a handout (3-5 pages per student) should be produced outlining the key features of the project and the literature on which these decisions are based (project outline). The handout should be submitted to the lecturer by the date of the presentation at the latest.

Group work is permitted. The maximum group size is 5 students. In case of group work, it must be possible to clearly define and assess each student's individual performance on the basis of specified sections, page numbers, or other objective criteria.

The presentation contributes 65% to the module grade, the handout contributes 35%. A passing grade in this module is achieved when the overall grade is greater than or equal to 4.0.

Literature

There are tons of books around that are both insightful and entertaining and support the lecture. In Figure 2, I present a short list of books I recommend: Bergstrom and West [2021], Chivers and Chivers [2021], Dougherty and Ilyankou [2021], Ellenberg [2015], Harford [2020], Huff [1954], Huntington-Klein [2022], and Jones [2020].

Figure 2: Books for data literacy

In addition, I highly recommend the two books mentioned in Figure 3. Vaughan [2020] teaches important skills for working effectively with data. For a more technical approach that focuses on statistics, Spiegelhalter [2019] is an excellent choice. Both books are easy to read even without advanced math skills. We will use these books in the course and some of your presentations will refer to them as well.

Figure 3: Books for skills and statistics
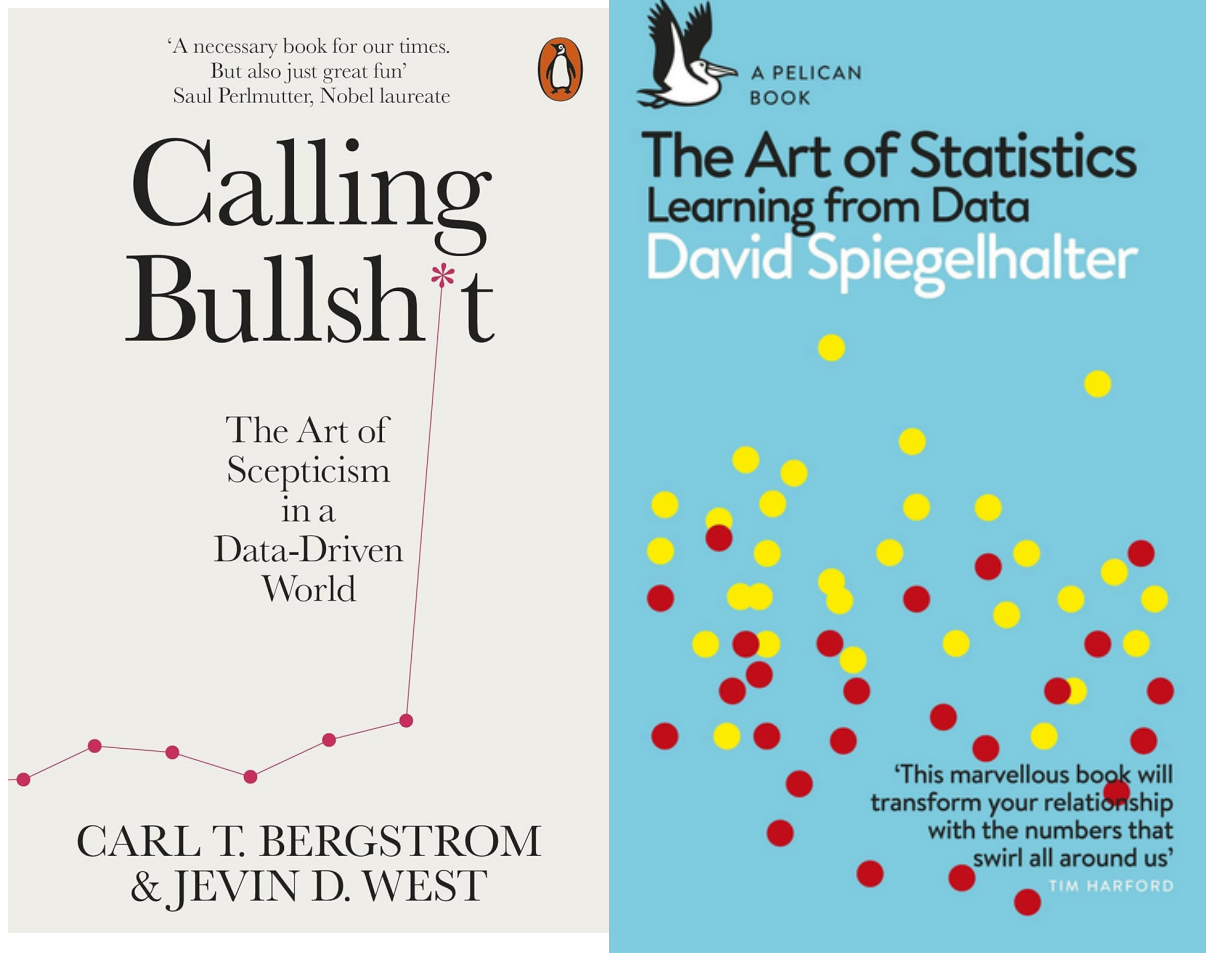
1 Introduction

1.1 What is the value that I can add in this course?

As a professor, my goal is to educate business graduates so that they can add value to their future organizations. Unfortunately, neither you nor I know what situations you will end up in in business where you can take advantage of using data to make good decisions. There are countless ways to contribute by understanding "data analysis for decision-making". Therefore, I will try to provide knowledge that is as general as possible. This way, I increase the likelihood that at least some parts of the lecture will be important for your future.

Since access to the internet is virtually constant and instantaneous these days, it's hard to add value by memorizing something. In seconds, we get answers to almost any sort of question using tools like Google and ChatGPT. Therefore, I don't see much added value to your education by teaching you facts and schemes to solve problems. Instead, I try to build up your skills and your data literacy that will help you ask the right questions, search for the right information, and use and analyze data in a meaningful way.

Perhaps you know the saying

> *"A man is well educated when he knows where to find what he doesn't know."*[1]

I would like to add something here. A man is well educated when he knows…

- … what he needs to know.
- … where to find what he doesn't know.
- … how to search efficiently.
- … how to verify the information being found.
- … how to transform information into insights.
- … how communicate the insight.

In the spirit of this lecture, my goal is to help you become an educated person who understands how data can support good decision-making.

1.2 What is data analysis for decision-making

For successful communication, intersubjectivity is essential. Once we agree on a clear definition of the words used in the title of this lecture, we can identify the things that we need to know, where to find them efficiently, how to evaluate and use the knowledge.

Data refers to all types of recordable information – facts about something or someone. Data analysis involves the thorough examination of this information with a specific goal in mind. Decision-making is defined as the process of reaching a decision. The word decision comes from the Latin verb decidere, which has various meanings, including

---

[1]Loosely translated and based on the well-known saying "Gebildet ist, wer weiß, wo er findet, was er nicht weiß." which is often attributed to George Simmel (1858 - 1918).

- make explicit,
- put an end to,
- bring to conclusion,
- settle/decide/agree (on),
- die,
- end up,
- fail,
- fall in ruin,
- fall/drop/hang/flow down/off/over,
- sink/drop,
- cut/notch/carve to delineate,
- detach,
- cut off/out/down,
- fell.

Let's agree on the following working definition that is symbolized in **?@fig-decisionm**:

> Fitzgerald [2002, p. 8]: "A decision is the point at which a choice is made between alternative—and usually competing—options. As such, it may be seen as a stepping-off point—the moment at which a commitment is made to one course of action to the exclusion of others."

Exercise 1.1. Pass the course

To pass this course, you will be required to give a 10-minute presentation and produce a 3-5 page handout. Your performance will be measured by the quality of your work in relation to the time available. The value of the presentation lies not only in the content, but also in the impact it has on your audience – your classmates and your professor. Your goal is to add some value with your work.

Discuss:

- What do you need to know to do well in this course?
- Where can you find the information you need to be successful?
- What tools will help you find, store and use information efficiently?
- How can you verify the accuracy of the information you find?
- How can you analyze information to gain valuable insights?
- How can you communicate these insights effectively?

Exercise 1.2. Information is omnipresent

Facts are abundant and easily accessible, so education should not just focus on providing information. Instead, skills should be developed to deal effectively with the wealth of facts available. This includes mastering statistical methods and avoiding common pitfalls when working with data.

To find the appropriate statistical method and interpret the data correctly, you should have some knowledge about the type of data and variables that you look at. Use ChatGPT to inform yourself about the following points that are defined in the module description:

Data literacy competencies

- Types of data: Cross-section, panel, time-series, georeferenced, …
- Types of variables: Continuous, count, ordinal, categorial, …

Make a short presentation about it. At best, include definitions and examples to the different types of information.

Moreover, what tools would you pick to make the presentation? Discuss the pros and cons of the options available.

2 From anecdote to insight

Anecdotes are great. They are true stories. They are often intriguing, thrilling, and often easy to understand. Oftenl, it is hard to deny the them.

Anecdotes are great. They are true stories—often intriguing, relatable, and easy to understand. They provide vivid examples that make abstract ideas more concrete and memorable. Whether it's a personal experience or a captivating story about a successful business leader, anecdotes resonate because they tap into our natural affinity for storytelling. Their simplicity and emotional impact can make them powerful teaching tools.

And importantly, anecdotes are hard to contradict. Take, for example, the argument that smoking can't be that harmful because your 88-year-old uncle has smoked his entire life and he is still in good health. It's a tough claim to refute, as it's a real-life example. However, the problem lies in extrapolating a single, isolated case to draw broader conclusions, which can be misleading.

However, while anecdotes can be persuasive, their strength is also their weakness. They represent isolated instances, and while it's hard to deny the truth of an individual story, the danger lies in overgeneralizing from it. Anecdotes lack the rigorous analysis and breadth of evidence necessary to draw reliable conclusions. They don't account for the full complexity of most situations, especially in business, where decisions are influenced by many interconnected factors.

In business, relying too heavily on anecdotes can lead to misguided conclusions. For example, a company might base its strategy on the success story of a famous entrepreneur without considering the countless failed ventures that didn't make the headlines. This is known as survivorship bias, where the successes are visible, but the failures are hidden.

The challenge, then, is to take anecdotes and go beyond them. Instead of drawing direct conclusions, use them as starting points for deeper investigation. They can provide valuable hypotheses but need to be supported by data, rigorous analysis, and an understanding of the underlying principles at play. Anecdotes can inspire curiosity and point us in interesting directions, but they should be tested against a larger body of evidence to ensure that the insights we draw are reliable and applicable in a broader context.

> **Exercise 2.1.** Survivorship bias
> Read "How Successful Leaders Think" by Roger Martin [2007] and the chapter "Identification" of "Quantitative Methods" by Huber [2024].
> Here is a summary of Martin [2007] taken from the Harvard Business Review Store:
>
> > In search of lessons to apply in our own careers, we often try to emulate what effective leaders do. Roger Martin says this focus is misplaced, because moves that work in one context may make little sense in another. A more productive, though more difficult, approach is to look at how such leaders think. After extensive interviews with more than 50 of them, the author discovered that most are integrative thinkers–that is, they can hold in their heads two opposing ideas at once and then come up with a new idea that contains elements

of each but is superior to both. Martin argues that this process of considera-
tion and synthesis (rather than superior strategy or faultless execution) is the
hallmark of exceptional businesses and the people who run them. To support
his point, he examines how integrative thinkers approach the four stages of
decision making to craft superior solutions. First, when determining which
features of a problem are salient, they go beyond those that are obviously
relevant. Second, they consider multidirectional and nonlinear relationships,
not just linear ones. Third, they see the whole problem and how the parts fit
together. Fourth, they creatively resolve the tensions between opposing ideas
and generate new alternatives. According to the author, integrative thinking
is an ability everyone can hone. He points to several examples of business
leaders who have done so, such as Bob Young, co-founder and former CEO
of Red Hat, the dominant distributor of Linux open-source software. Young
recognized from the beginning that he didn't have to choose between the two
prevailing software business models. Inspired by both, he forged an innovative
third way, creating a service offering for corporate customers that placed Red
Hat on a path to tremendous success.

a) Discuss the concepts introduced by Martin [2007] critically:

- Does he provide evidence for his ideas to work?
- Is there a proof that his suggestions can yield success?
- Is there some evidence about whether his ideas are superior to alternative causes of action?
- What can we learn from the article?
- Does his argumentation fulfill highest academic standards?
- What is his identification strategy with respect to the *causes of effects* and the *effects of causes*?
- Martin [2007, p. 81] speculates:

  "At some point, integrative thinking will no longer be just a tacit skill (culti-
  vated knowingly or not) in the heads of a select few."

b) If teachers in business schools would have followed his ideas of integrative thinkers
being more successful, almost 20 years later, this should be the dominant way to
think as a business leader. Is that the case? And if so, can you still gain some
competitive advantage by thinking that way?

Figure 2.1: Distribution of bullet holes in returned aircraft
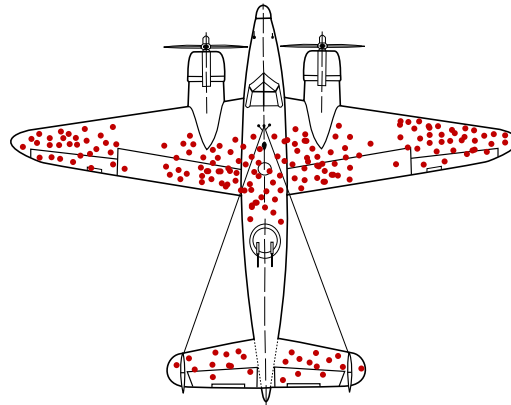
Source: Martin Grandjean (vector), McGeddon (picture), Cameron Moll (concept), CC BY-SA 4.0, Link

c) Figure 2.1 visualizes the distribution of bullet holes in aircraft that returned from combat in World War II. Imagine you are an aircraft engineer. What does this picture teach you?

d) Inform yourself about the concept of survivorship bias explained in Wikipedia [2024].

e) In Martin [2007], the author provides an example of a successful company to support his management ideas. Discuss whether this article relates to survivorship bias.

Drawing insights from anecdotes is challenging, especially in business, for several reasons:

1. **Limited sample size**: Anecdotes are usually individual cases that do not reflect the full extent of a situation. In business, decisions often require data from large, diverse populations to ensure reliability. Relying on a single story or experience can lead to conclusions that are not universally valid.

2. **Bias and subjectivity**: Anecdotes are often influenced by personal perspectives, emotions or particular circumstances. Moreover, anecdotes often highlight success stories while ignoring failures. This is an example for the so-called *Survivorship Bias*.

3. **Lack of context and the inability to generalize**: Anecdotes often lack the broader context necessary to understand the underlying factors of a situation. Business problems tend to be complex and influenced by numerous variables such as market trends, consumer behavior and external economic conditions. Many of these variables change significantly over time. Without this context, an anecdote can oversimplify the problem and lead to incorrect decisions. Anecdotes are usually specific to a particular time, place or set of circumstances. They may not apply to different markets, industries or economic environments, which limits their usefulness for general decision-making. For example, learning only from the tremendous success of figures like Steve Jobs while ignoring the countless people who failed is like learning how to live a long life by talking to a single 90-year-old person. If that person happens to be obese and a heavy smoker, it doesn't mean those behaviors contributed to their longevity.

4. **Lack of data rigor**: Anecdotes lack the rigor and precision of data-driven analysis where the empirical model that allows to identify causality and to measure the effect of causes is formally described.

Conclusion

To make informed business decisions, it is critical to base insights on systematic data analysis rather than anecdotal evidence, as anecdotes are too narrow, subjective and unreliable to guide complex business strategies.

Exercise 2.2. Systematic analysis as an alternative to anecdotal analysis

- What defines a systematic analysis?
- When can we say that we have 'found evidence'?
- When can we claim to have identified a causal effect?
- When can we trust the size of an effect that we have measured?

3 Epistemic errors

Many things can go wrong when analysing data. Various issues can lead to misleading interpretations, whether due to data not measuring what it is intended to, human misinterpretation, or an overreliance on probabilistic reasoning. The following examples illustrate these pitfalls and are drawn from the insightful book by Jones [2020], which I highly recommend for further reading.

Epistemic errors, that are mistakes related to knowledge, understanding, or the acquisition of information, arise from cognitive biases, misunderstandings, and incorrect assumptions about the nature of data and the reality it represents. Recognizing and addressing these errors is crucial for accurate data analysis and effective decision-making.

When researchers analyse data they should be aware of the fact that data can represent reality but do not necessarily do so. For example, a survey on customer satisfaction that only includes responses from a self-selected group of highly engaged customers may not accurately reflect the overall customer base, see the box Hite Report. To avoid this pitfall, it is essential to ensure that your data collection methods are representative and unbiased, and to validate your data against external benchmarks or additional data sources.

> Hite Report
>
> In 1976, when the *The Hite Report* (see Hite [1976] and Figure 3.1) was published it instantly became a best seller. Hite used an individualistic research method. Thousands of responses from anonymous questionnaires were used as a framework to develop a discourse on human responses to gender and sexuality. The following comic concludes the main results.
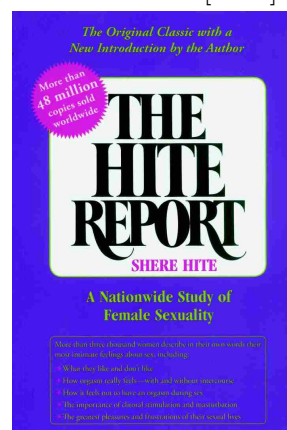>
> Figure 3.1: The Hite [1976] Report
>
>

Figure 3.2: Comic on the Hite Report

Source: Picture is taken from www.theparisreview.org.

The picture of womens' sexuality in Hite [1976] was probably a bit biased as the sample can hardly be considered to be a **random and unbiased** one:

- Less than 5% of all questionnaires which were sent out were filled out and returned (response bias).
- The questions were only sent out to women's organizations (an opportunity sample).

Thus, the results were based on a sample of women who were highly motivated to answer survey's questions, for whatever reason.

The Data-Reality Gap

The difference between the data we collect and the reality it is supposed to represent.

Another common epistemic error involves the influence of human biases during data collection and interpretation. Known as the all too human data error, this occurs when personal biases or inaccuracies affect the data. An example would be a researcher's personal bias influencing the design of a study or the interpretation of its results. To mitigate this, implement rigorous protocols for data collection and analysis, and consider using double-blind studies and peer reviews to minimize bias.

> **All Too Human Data**
>
> Errors introduced by human biases or inaccuracies during data collection and interpretation.

Inconsistent ratings can also lead to epistemic errors. This happens when there is variability in data collection methods, resulting in inconsistent or unreliable data. For example, different evaluators might rate the same product using different criteria or standards. To avoid this issue, standardize data collection processes and provide training for all individuals involved in data collection to ensure consistency.

> **Inconsistent Ratings**
>
> Variability in data collection methods that leads to inconsistent or unreliable data.

The black swan pitfall refers to the failure to account for rare, high-impact events that fall outside regular expectations. Financial models that did not predict the 2008 financial crisis due to the unexpected nature of the events that led to it are an example of this error. To prevent such pitfalls, consider a wide range of possible outcomes in your models and incorporate stress testing to understand the impact of rare events.

> **The Black Swan Pitfall**
>
> The failure to account for rare, high-impact events that fall outside the realm of regular expectations.

Falsifiability and the God pitfall involve the tendency to accept hypotheses that cannot be tested or disproven. This error might occur when assuming that a correlation between two variables implies causation without the ability to test alternative explanations. To avoid this, ensure that your hypotheses are testable and that you actively seek out potential falsifications. Use control groups and randomized experiments to validate causal relationships.

> **Falsifiability and the God Pitfall**
>
> The tendency to accept hypotheses that cannot be tested or disproven.

To avoid epistemic errors, critically assess your assumptions, methodologies, and interpretations. Engage in critical thinking by regularly questioning your assumptions and seeking alternative explanations for your findings. Employ methodological rigor by using standardized and validated methods for data collection and analysis. Engage with peers to review and critique your work, providing a fresh perspective and identifying potential biases. Finally, stay updated with the latest research and best practices in your field to avoid outdated or incorrect methodologies.

Understanding and addressing epistemic errors can significantly improve the reliability and accuracy of your data analyses, leading to better decision-making and more trustworthy insights.

> **Exercise 3.1.**

Figure 3.3: Bananas in various stages of ripeness
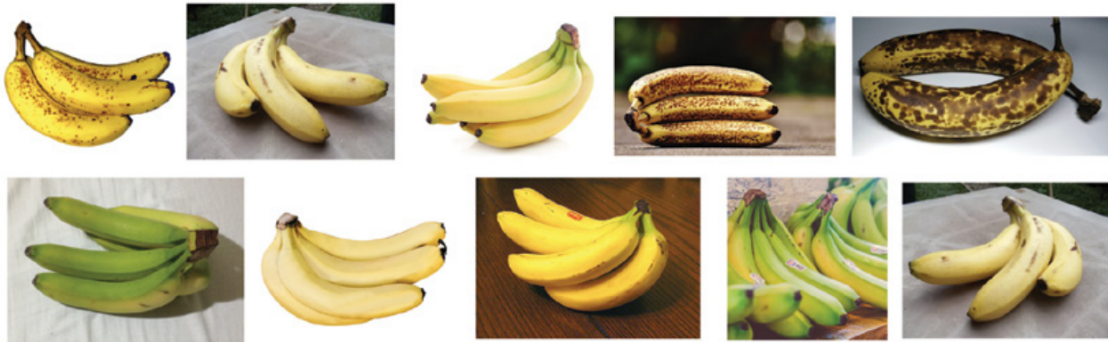
Source: Jones [2020, p. 33]

a) Rate the ripeness level of the bananas pictured by Figure 3.3. Compare your assessment to that of a colleague and discuss any differences in your ratings. What might account for the variance in perception of the bananas' ripeness between you and your colleague?

b) Specify how you rated the second and the last bananas on the ripeness scale?

c) Upon reevaluation, it appears that the second and the last bananas are identical in ripeness. How would you justify your initial decision now? This scenario underscores an important lesson for interpreting polls and surveys: it illustrates how subjective assessments can lead to variance in results. It highlights the necessity of ensuring clarity and consistency in the criteria used for evaluations to minimize subjective discrepancies.

References

Carl T Bergstrom and Jevin D West. *Calling Bullshit: The Art of Skepticism in a Data-Driven World.* Penguin Books, 2021.

Tom Chivers and David Chivers. *How to Read Numbers: A Guide to Statistics in the News (And Knowing When to Trust Them).* Weidenfeld & Nicolson, 2021.

Thomas H Davenport and DJ Patil. Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 90(5):70–76, 2012.

Jack Dougherty and Ilya Ilyankou. *Hands-On Data Visualization Interactive Storytelling from Spreadsheets to Code.* O'Reilly, 2021. URL https://handsondataviz.org/.

Jordan Ellenberg. *How Not To Be Wrong: The Power of Mathematical Thinking.* Penguin Books, 2015.

Stephen P. Fitzgerald. *Decision Making.* Capstone Publishing, 2002.

Tim Harford. *How to Make the World Add Up: Ten Rules for Thinking Differently about Numbers.* The Bridge Street Press, 2020.

Shere Hite. *The Hite Report. A Nationwide Study of Female Sexuality.* New York: Dell, 1976.

Stephan Huber. Quantitative methods, 2024. URL https://hubchev.github.io/qm/.

Darrell Huff. *How to Lie with Statistics.* WW Norton & company, 1954.

Nick Huntington-Klein. *The Effect: An Introduction to Research Design and Causality.* CRC Press, 2022. URL https://theeffectbook.net.

Ben Jones. *Avoiding Data Pitfalls: How to Steer Clear of Common Blunders When Working with Data and Presenting Analysis and Visualizations.* John Wiley & Sons, Hoboken, New Jersey, 2020.

Roger Martin. How successful leaders think. *Harvard Business Review*, 85(6):71–81, 2007. URL https://hbr.org/2007/06/how-successful-leaders-think.

David Spiegelhalter. *The Art of Statistics: Learning From Data.* Penguin UK, 2019.

Daniel Vaughan. *Analytical skills for AI and data science.* O'Reilly Media, Sebastopol, CA, June 2020.

Wikipedia. Survivorship bias, 2024. URL https://en.wikipedia.org/wiki/Survivorship_bias.

Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach.* South-Western, 2nd edition, 2002.