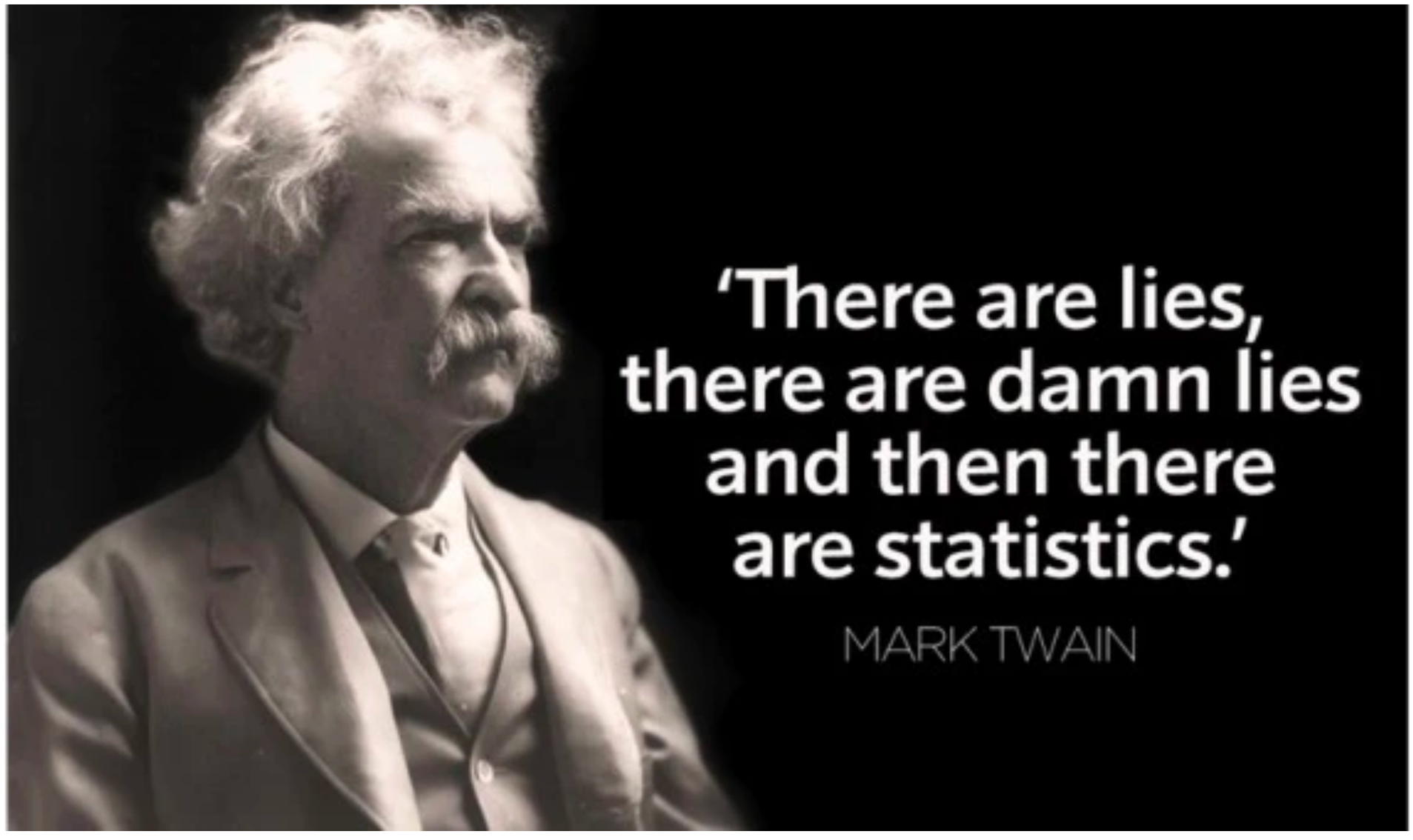
Data Science Ethics

Dr. Melissa Innerst, Ph.D. Fall 2021

Introduction

One of the most common quotes with “statistics” in it is…



Introduction

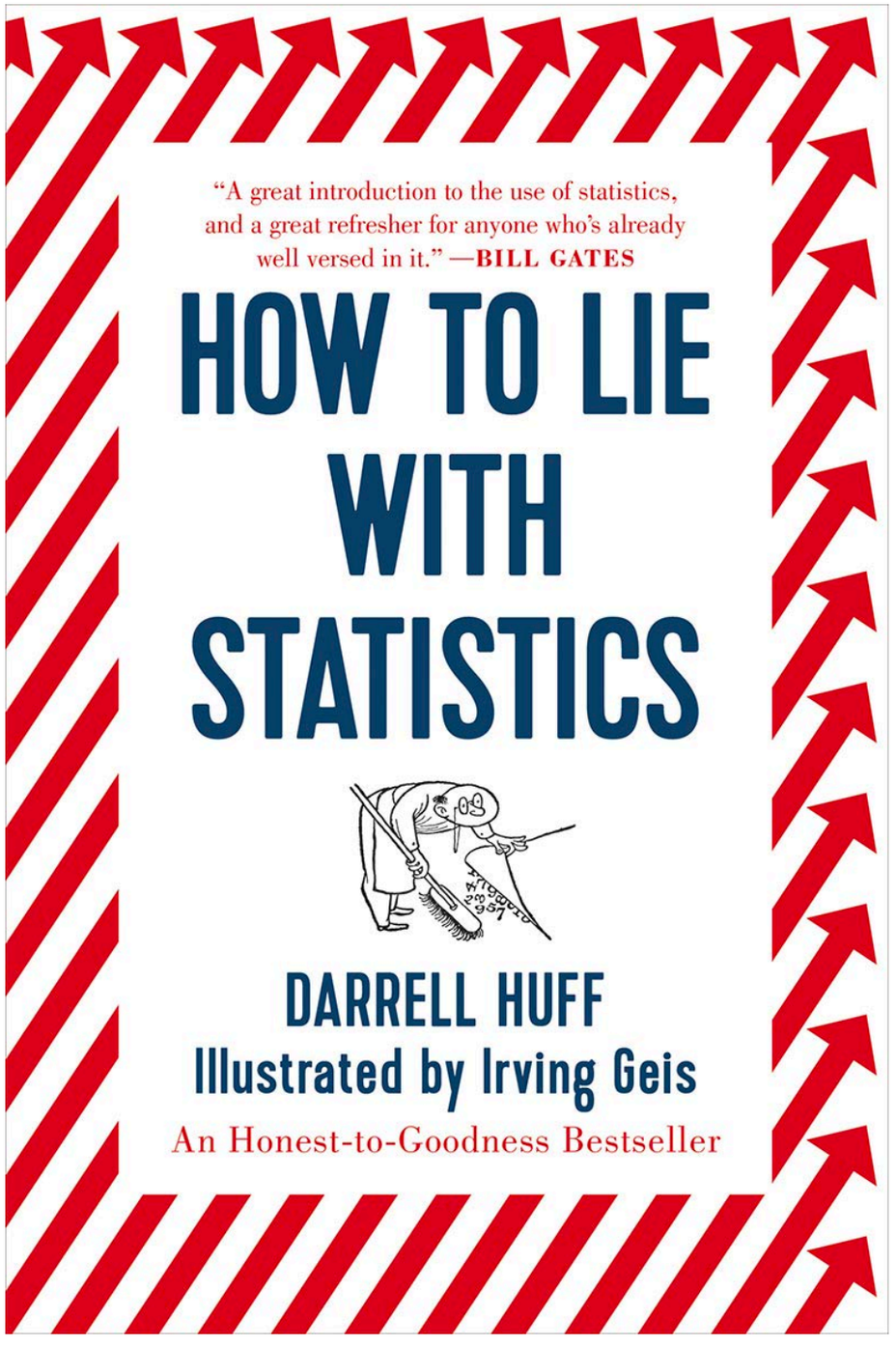
The best-selling book with “statistics” in the title is *How to Lie with Statistics* by Darrell Huff

Shows graphical ploys used to fool people even with accurate data

Statistics has received a bad reputation (and wrongly so!)

The book isn’t suggesting the readers do these things, it is warning readers be on the lookout for them

The tactics shown in this book are employed by politicians, journalists, and business people *not* by statisticians



Introduction

As a data scientist, in much of your work, you will be relying on the trust and confidence that your clients place in you

The term **professional ethics** describes the special responsibility not to take unfair advantage of that trust

That involves more than just being thoughtful and using common sense!

There are specific professional standards that should guide your actions

Introduction

The best known professional standards are those in the Hippocratic oath for physicians

The following three principles in the oath mirror standards for professionalism in data science:

> “I will not be ashamed to say ‘I know not,’ nor will I fail to call in my colleagues when the skills of another are needed for a patient’s recovery”

> “I will respect the privacy of my patients, for their problems are not disclosed to me that the world may know”

> “I will remember that I remain a member of a society, with special obligations to all my fellow human beings, those sound of mind and body as well as the infirm”

Introduction

In this section, we will discuss principles of professional ethics for data science

We will look at specific examples of legal obligations as well as guidelines issued by professional societies

Note: There is not a “data analyst’s oath,” there are only guidelines

These guidelines provide a description of the ethical expectations on which your clients can reasonably rely

As a consensus statement of professional ethics, the guidelines also establish standards of accountability

Principles to Guide Ethical Action

There are a few common sense starting places for behaving ethically in a given situation

Tell the truth

Don’t harm innocent people

Don’t steal

Professional ethics also require a neutral, unemotional, and informed assessment of the situation

Principles to Guide Ethical Action

As a professional, you possess skills that are not widely available

A fundamental notion of professional ethics is to avoid using those skills in a way that is misleading

Non-professionals are not always in a position to make an informed judgement about whether your methods and conclusions are appropriate

There is an implicit promise that, in every professional action you take, you will use appropriate methods and draw appropriate conclusions

Principles to Guide Ethical Action

It is necessary to believe that your methods and conclusions are appropriate, but it is not sufficient

It is easy to mislead yourself, particularly in the heat of the moment and the excitement of satisfying your client/research team

It isn’t a matter of absolutes. It is not always certain that a method is appropriate. There is almost always a risk that something is wrong

Deal with this by *drawing on* generally recognized professional standards. For example,

Use software systems that have been vetted by the community

Check that your data are what you believe them to be

Don’t use analytical methods that would not pass scrutiny by professional colleagues

Principles to Guide Ethical Action

Be open and honest!

Don’t overstate your confidence in results

If there is a part of the analysis that you aren’t confident in and would be uncomfortable with an expert scrutinizing, you should point that out yourself (don’t wait for someone to notice)

Point out to clients the substantial risks of error or unexpected outcome

Note: You do not have the time, money, or data needed to make every aspect of your work perfect

Use professional judgement and identify the most salient risks and ensure that your work is good enough even if it isn’t perfect

Principles to Guide Ethical Action

You have a professional responsibility to particular stakeholders

It is important that you consider/recognize all the various stakeholders to whom you have this responsibility

You could have a responsibility to…

- Your employer or client
- The general public
- The subjects in your study
- The individuals represented in your data
- Your research community or profession

You could have legal responsibilities as well if there are laws relevant to your work or you are called upon as an expert witness in a trial

Principles to Guide Ethical Action

Another concern is the potential for a conflict of interest

A conflict of interest is not itself unethical. We will all have such conflicts. It is natural.

We want do work that will advance us professionally, which instills a temptation to satisfy the expectations of our employers or colleagues or the marketplace

We have to be careful not to let our personal goals cloud or bias or otherwise shape our professional judgement

Professional Ethical Precepts

#1 - Do your work well by your own standards and by the standards of your profession

#2 - Recognize the parties to whom you have a special professional obligation

#3 - Report results and methods honestly and respect your responsibility to identify and report flaws/shortcomings in your work

Principles to Guide Ethical Action

Example: Legal ethics

Lawyers for an accused murderer found the bodies of two victims whose deaths were unknown to the authorities and to the victims' families

The responsibility to confidentiality for their client precluded the lawyers from following their hearts and reporting the discovery

The lawyers careers were destroyed by the public!

Yet, courts and legal scholars have confirmed that the lawyers were right to do what they did

Such drama is rare! In this section, we will consider a few, less dramatic scenarios

Your Turn

Suppose you are a consultant for a client who wants a proprietary model to predict commercial outcomes

After reviewing the literature, you find an existing multiple linear regression model that matched the scenario well and use available public data to fit the parameters of the model

The client's staff were pleased, but the CEO wanted a model that would give a competitive advantage to the company

The CEO asks you, the statistical consultant, whether the coefficients in the model can be "tweaked" to reflect the specific values of the company

You tell the CEO those values best match the data and to arbitrarily change them would be "playing God". The CEO says they *want* to play God.

How should you respond?

Ethics Codes and Frameworks

Several organizations have developed detailed statements on topics such as…

- Professionalism
- Integrity of data and methods
- Responsibilities to stakeholders
- Conflicts of interest
- Response to allegations of misconduct

Those organizations include…

- American Statistical Association (ASA)
- Association for Computing Machinery (ACM)
- Data Science Association
- United Nations Statistics Division