Winter semester 2019-2020, Mondays 9:30-14:00

# Research Methods and Ethics

# HOW TO LIE WITH STATISTICS?

**Oslo Metropolitan University**

2019.09.23, Pedro G. Lind

# What is statistics?

*"It is the practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample."*

Oxford English Dictionary

**And it deals with a fundamental concept...**

**The concept of PROBABILITY**

# What is probability?

*"The quality or state of being probable; the extent to which something is likely to happen or be the case."*

Oxford English Dictionary

*"The quality or state of being probable; the extent to which something is likely to happen or be the case."*

Oxford English Dictionary

# Why is probability so important?

**Because the universe is uncertain...**

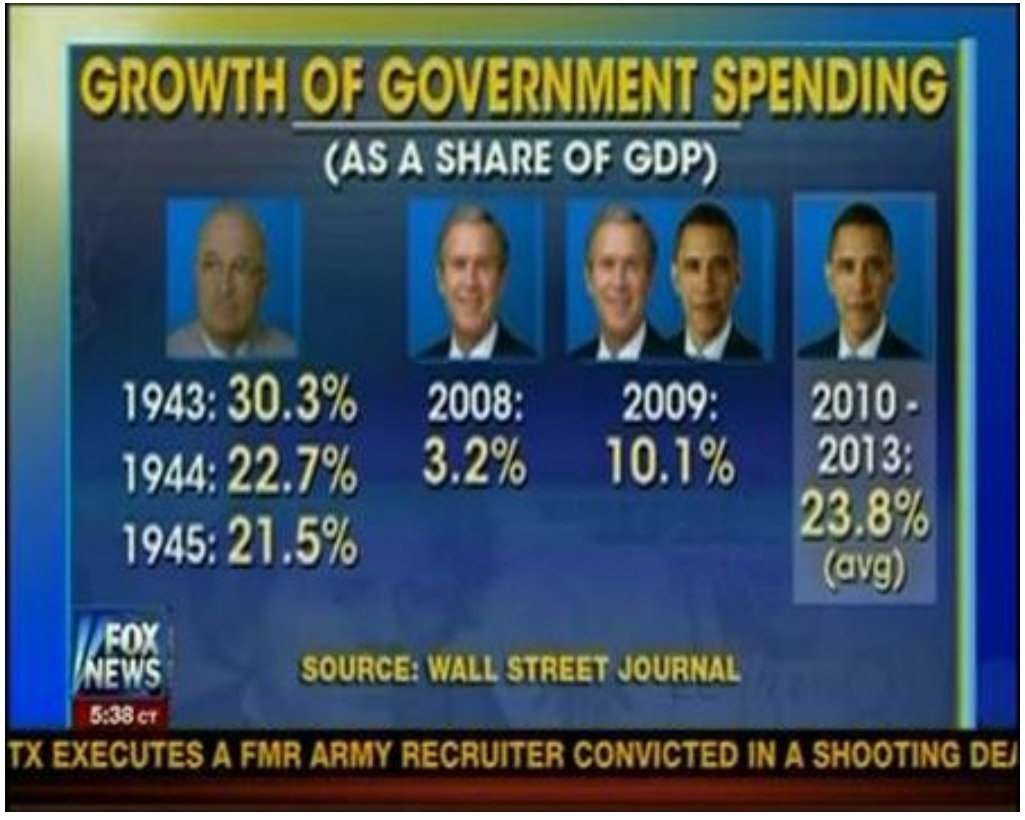
**... and probability is a weapon for surviving in it!**

- Important to know the weather tomorrow.
- Important to invest in the stock market.
- Important to evaluate risk in security systems.
- Important to handle big data in clouds.
- ... ... ... ...

# What do we do with probability and statistics?

- Probability is what we want to calculate for making decisions: How probable it is that our decision will not lead to a catastrophe?
- PROBLEM: we cannot measure probabilities directly.
- SOLUTION: We do statistics with data!
- **TAKE-AWAY:** From statistical analysis of real data we can estimate probabilities and make decisions!

# The dark side of the force...




# ...7 tips to better lie with statistics...

**Sources: "Lessons from how to lie with statistics"**

available at `https://towardsdatascience.com/`

Tip 1 to better lie with statistics

# Use big names to support your claims

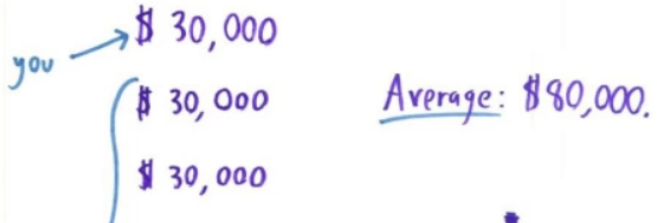
... don't let a fact prove an argument wrong!


Tip 2 to better lie with statistics

# Use the mean as an authority!

HOW TO LIE WITH STATISTICS

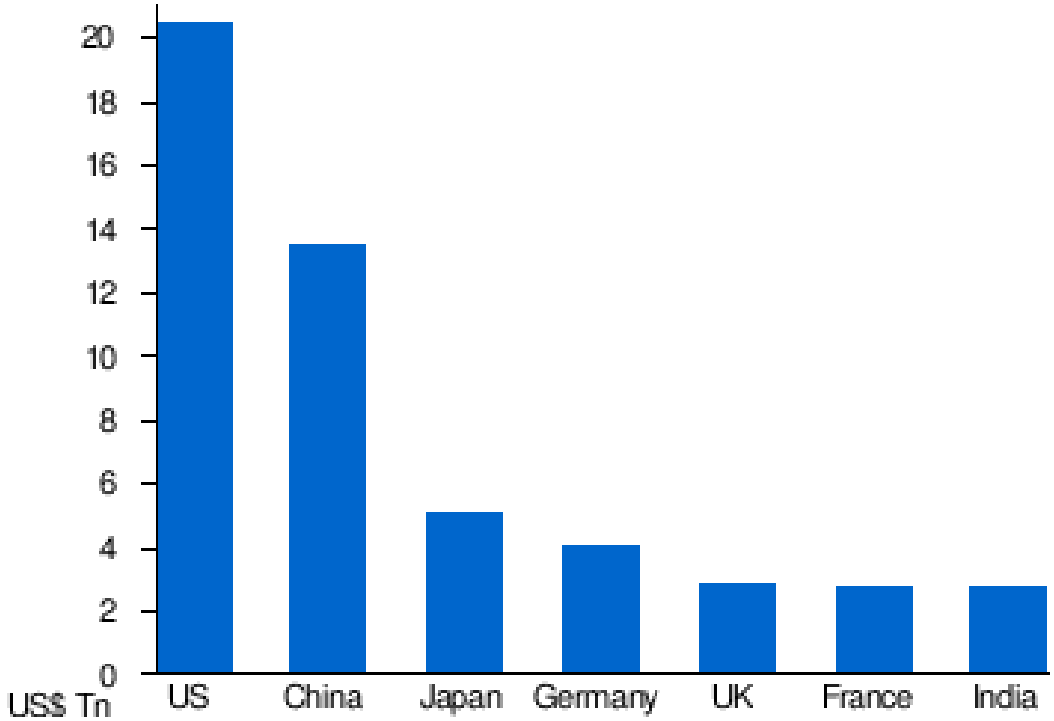

HOW TO LIE WITH STATISTICS



Tip 3 to better lie with statistics

## Use numbers without comparing them with a reference!

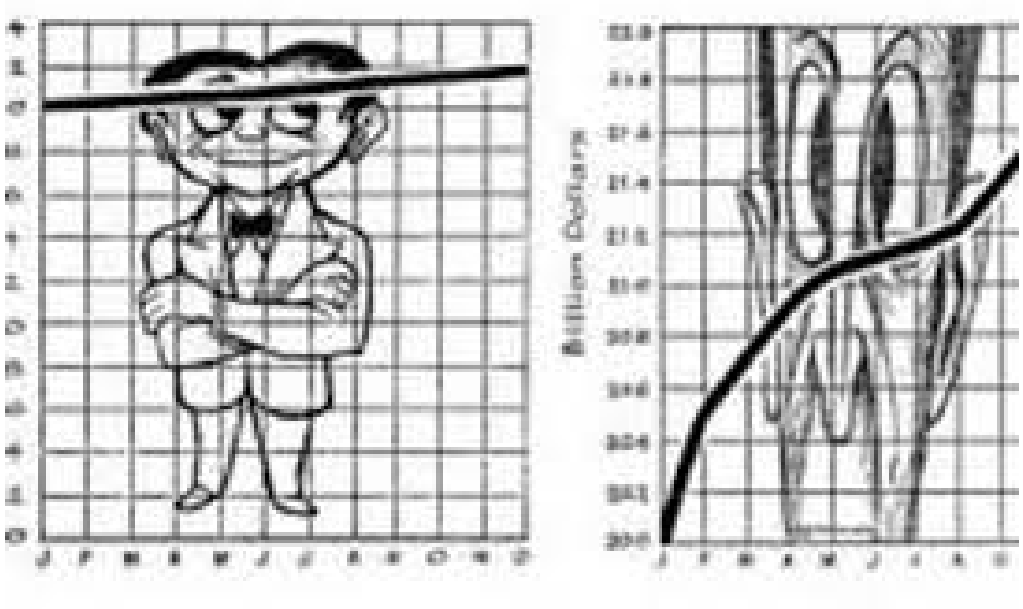
Estimated GDP for Norway in 2019: $\sim$ 0.427 US$Trillion

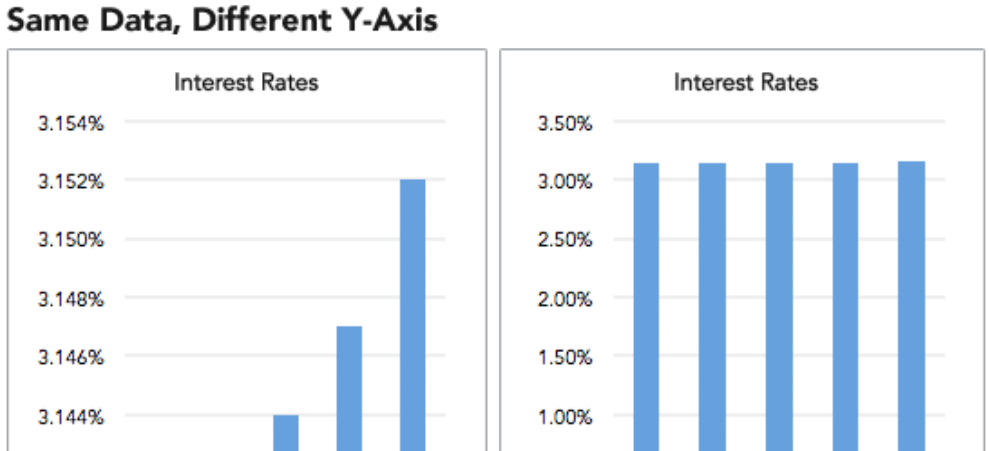



## (Percentual variation would be the correct approach...)

Tip 4 to better lie with statistics

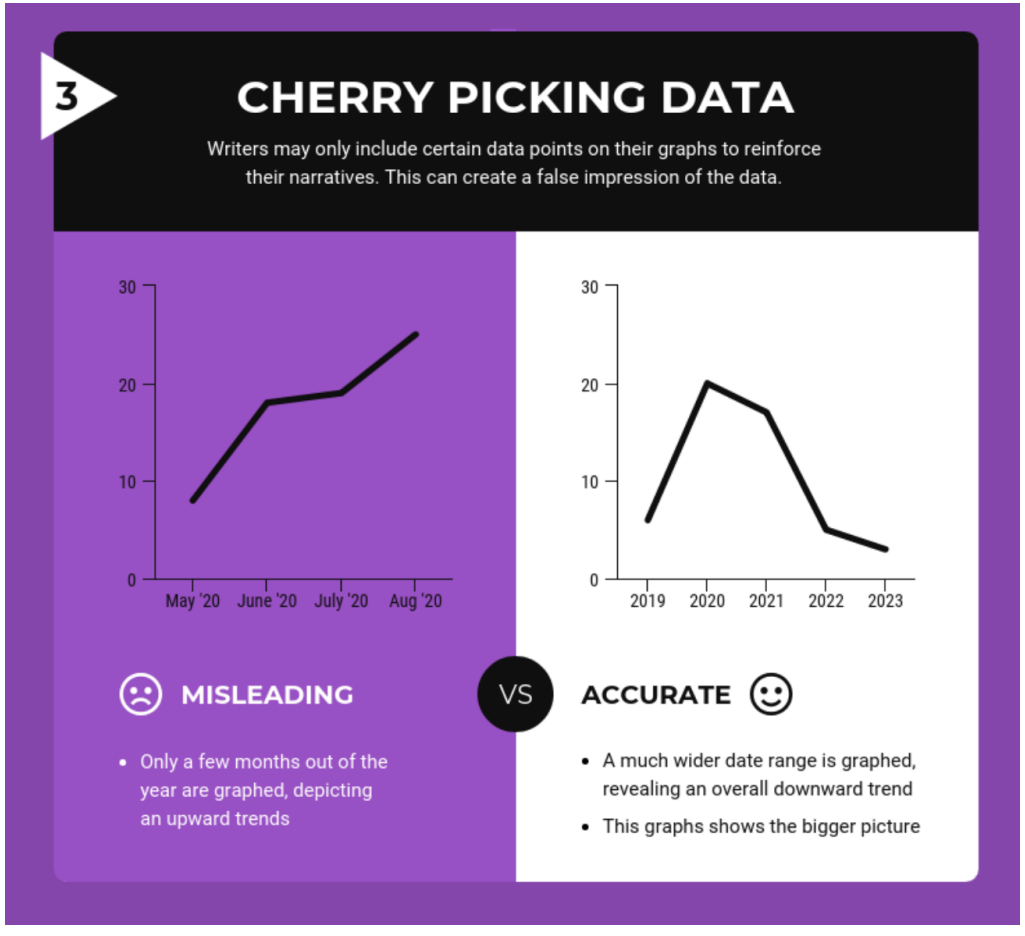
# Show your data according the “needs”...



Tip 5 to better lie with statistics

# Select data that best serve your purposes

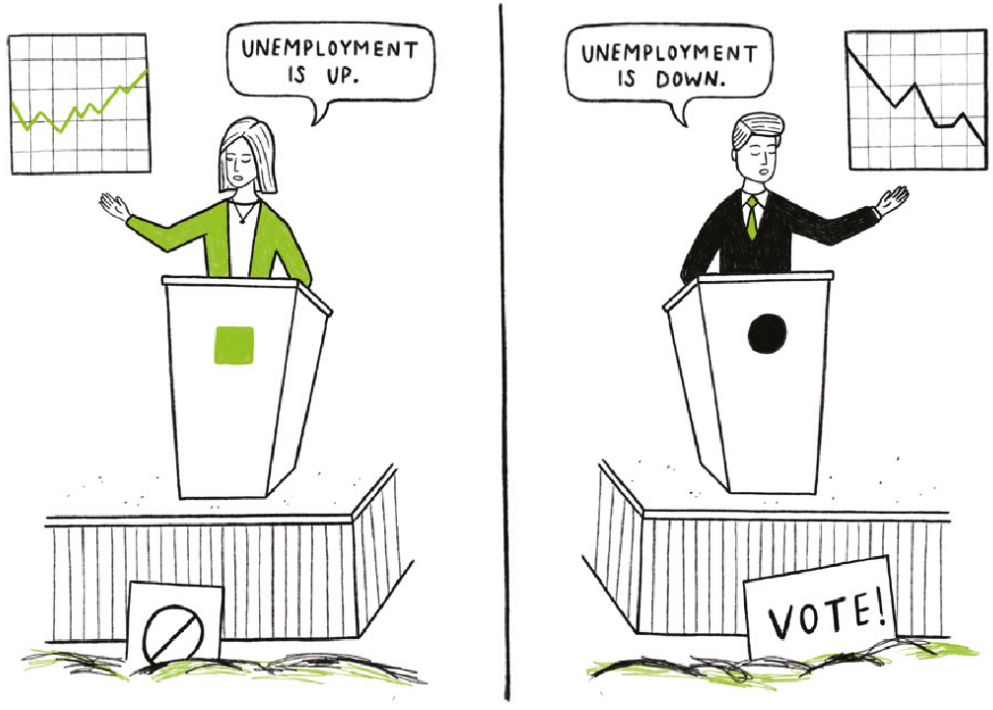


Tip 5 to better lie with statistics

# Select the data that best serve your purposes

**Use** *special* **samples**

... and confirm them with your own (always biased) experience...

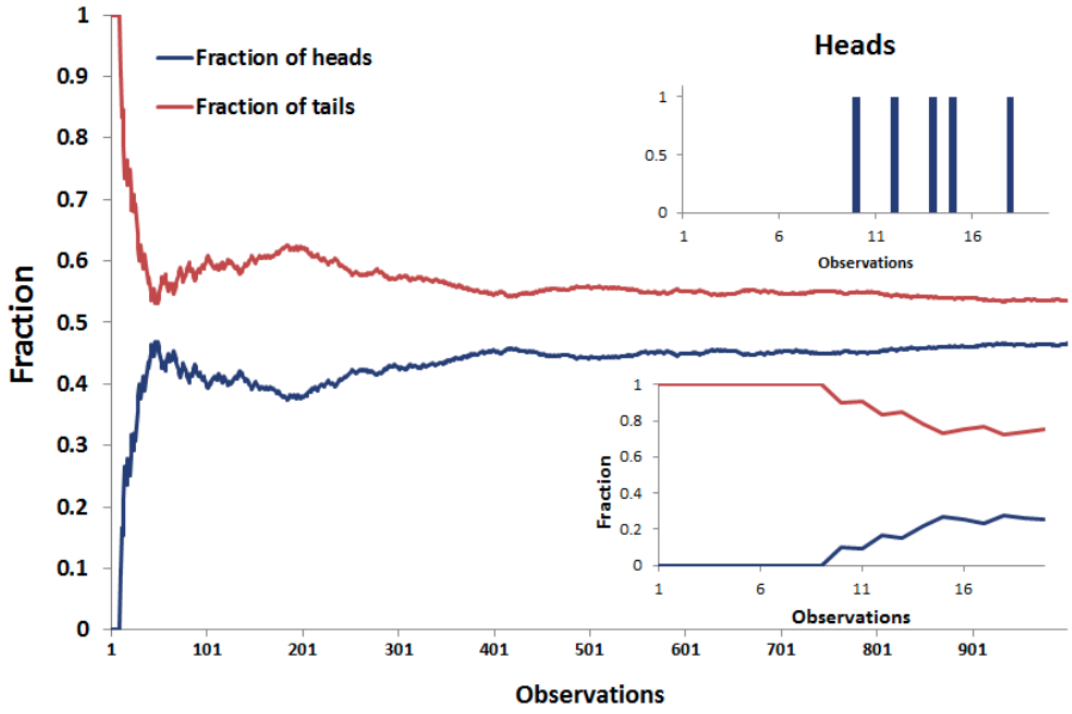


Tip 5 to better lie with statistics

# Select the data that best serve your purposes

**Use *small* samples**

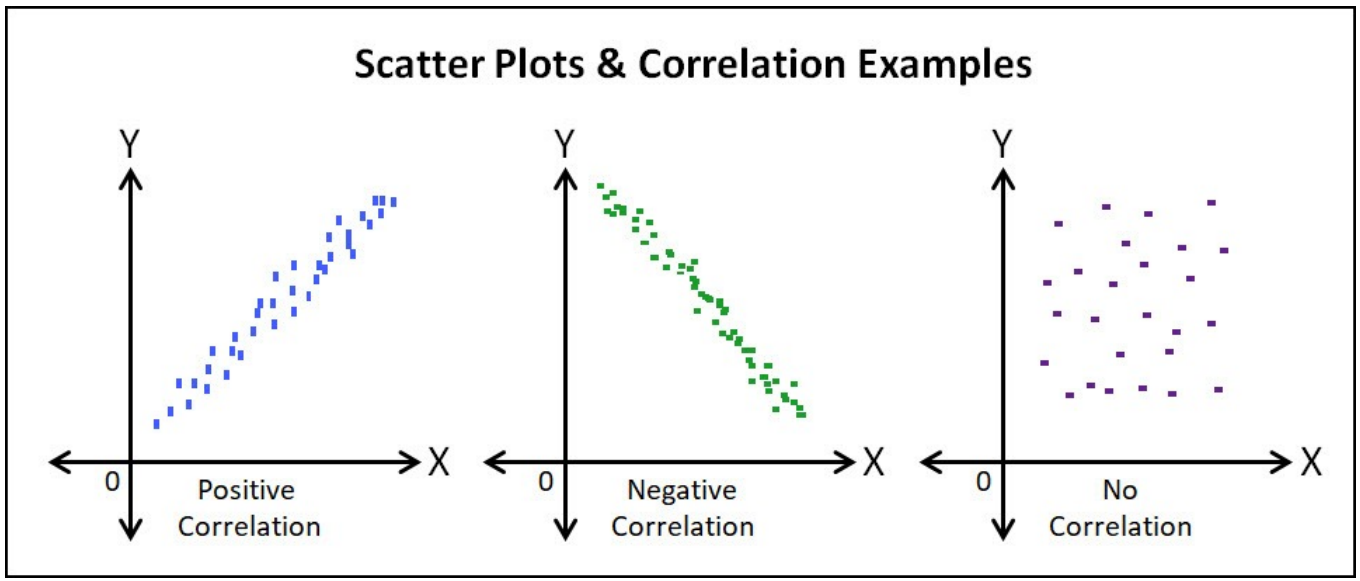
Errors will be so large that any outcome is possible!



Tip 6 to better lie with statistics

# Use CORRELATION as a synonym of CAUSATION

- Assume *X* and *Y* are correlated.
- **What does that mean?**
  It *only* means that they increase together or decrease together (positive), or one goes up as the other does down (negative).

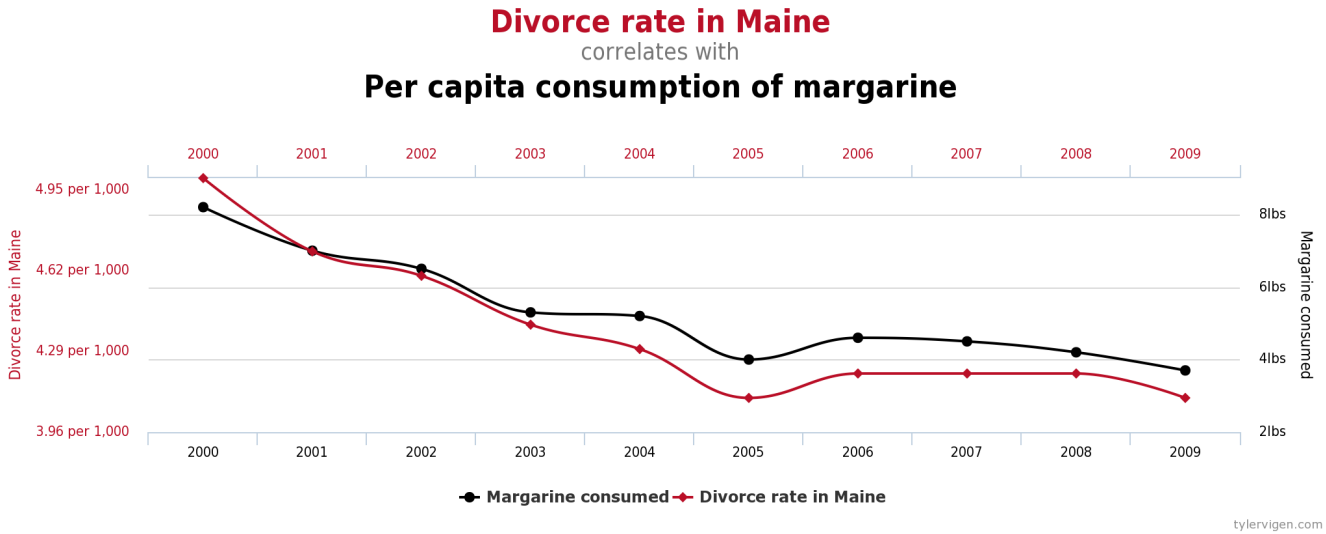


Tip 6 to better lie with statistics

# Use CORRELATION as a synonimous for CAUSATION

## Three different explanations are possible:

- One **causes** the other ($X$ =clouds and $Y$ =rain).
- Both are caused by one **third thing** ($X$ =sun glasses and $Y$ =ice-creams).
- Both things are completely **unrelated**.

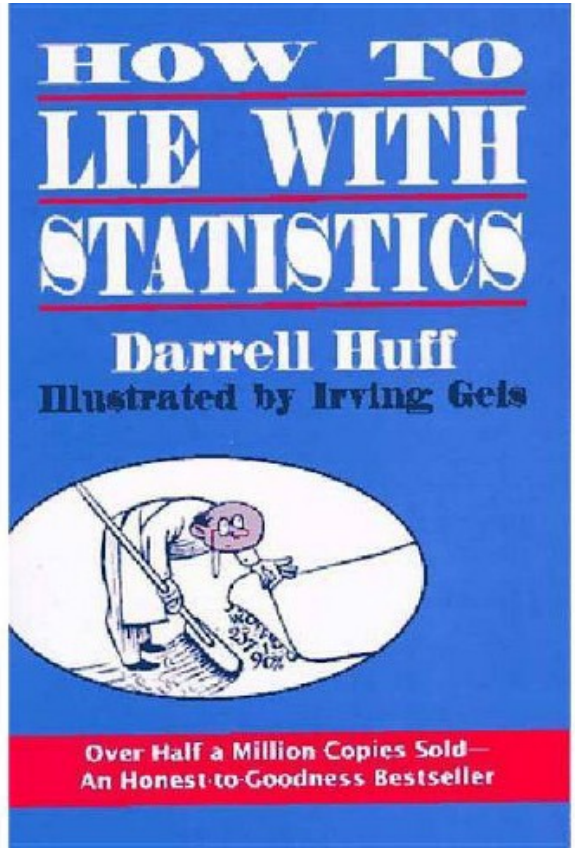


## Tip 7 to better lie with statistics

# Extrapolate as much as you can!

- **In statistics no relationship lasts forever!**
- Extrapolations are very dangerous in a serious statistical analysis.

# How to lie with statistics?



## Take-home messages part I

### If you want to lie to other:

- As future producers of tables and graphs, follow these tips for misleading your statistical results to your peers or costumers.
- But to succeed on that, **develop your "data literacy":** learn how to properly read, understand and explain data.

## Take-home messages part II

### If you do not want to be tricked by other:

- Do not forget: statistical models are possible interpretations of uncertain data, sometimes designed to inform, sometime to persuade. Try to **distinguish between them.**
- Be as **objective** as possible while analyzing your data.
- Be **critical** about the data-gathering process.
- Check the values and their ranges.
- Report the **confidence intervals** (details next week...)