

# Statistical Methods and Data Analysis I

## Lecture 1: What We Will Deal With — Introduction

Including Administrivia

Oleg Goldshmidt

`oleg.goldshmidt@post.idc.ac.il`

Arison School of Business  
Interdisciplinary Center (IDC)  
Herzliya, Israel

March 03, 2019



- Assist the students in developing the necessary **intuition and habits** to formulate useful hypotheses, design experiments to test them while avoiding common pitfalls, choose the most appropriate and effective techniques in each case.
- Provide students with **methodology, techniques, and tools** to collect and analyse data, develop hypotheses and models, make rational data-driven decisions, present ideas effectively and convincingly.
- Teach the students develop a *habit* of *methodically* applying *intuition*, *tools*, and *techniques* to **critical evaluation** of information, proposals, and conclusions presented by others.

- Mathematics: some, not very much.
  - Some linear algebra
  - Some intuition about mathematical analysis (limits, derivatives, integrals, etc.)
  - Some common sense regarding probability.
- Basic programming skills.
  - Basic familiarity with the R programming language and some of its useful packages and libraries.

You have studied some of that stuff for sure.

Reasonable effort will be made to bridge gaps where necessary.

# Course Components

**Lectures.** The primary activity of the course will consist of lectures incorporating individual and group practical exercises.

**Mini-Seminar.** Close to the end of the semester some lectures will be in the form of a mini-seminar. Students will be assigned, well in advance, academic papers or real data sets to review. The work will be done in pairs, the participants will present their analysis (for instance, one supporting and the other criticizing the study's conclusions, or presenting conflicting points of view regarding the data) to the rest of the class.

**Homework.** At least 2, possibly more, homework assignments during the semester. Homework submission is not a requirement for taking the final test or passing the course. However, homework assignments not submitted on time will receive 0 grade that will be taken into account in the final weighted grade. Any delay in submission of homework may be allowed under special circumstances only (e.g., reserve military duty, serious illness, etc.), supported by official documentation, and strictly in accordance with IDC policies.

**Final test.** There will be a written test scheduled on two dates according to the IDC policies. Any special dispensations beyond the normal dates may be granted only through IDC administration and subject to agreement of the Lecturer. If granted, the format of the special test may be different (e.g., an oral examination).

**Final grade** = a weighted average of the following:

Final test: **60%**

Mini-seminar: **20%**

Homework: **20%**

# Definition of Statistics



# Definition of Statistics

## Wikipedia:

A branch of mathematics dealing with collection, analysis, interpretation, presentation, and organization of data.

# Definition of Statistics

## Wikipedia:

A branch of mathematics dealing with collection, analysis, interpretation, presentation, and organization of data.

## Merriam-Webster Dictionary:

A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.

# Definition of Statistics

## Wikipedia:

A branch of mathematics dealing with collection, analysis, interpretation, presentation, and organization of data.

## Merriam-Webster Dictionary:

A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.

## Oxford English Dictionary:

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.



**State**-istics:

collection of population and economic information vital to the **state**.

## State-istics:

collection of population and economic information vital to the **state**.

## 1770:

“science dealing with data about the condition of a state or community” — from German *Statistik*, popularized and perhaps coined by the German political scientist Gottfried Aschenwall (1719–1772).

## State-istics:

collection of population and economic information vital to the **state**.

## 1770:

“science dealing with data about the condition of a state or community” — from German *Statistik*, popularized and perhaps coined by the German political scientist Gottfried Aschenwall (1719–1772).

## Today:

widely applied not only to matters of state, but in business, finance, technology, and all the natural and social sciences.

# What Statistics Are Good For?

- Figuring out what is going on and compiling an **objective representation of the current situation**.
- Discovering and analyzing **past trends**.
- **Modeling relationships** between various factors and parameters.
- Making **rational decisions** relevant to the future.
- **Presenting** convincing reports, ideas, proposals, etc., supported by data and models.
- Avoiding being a sucker and **developing a nose** for bullshit.
  - This is an essential part of being **objective** and **rational**.
  - Understanding how it works will help you become more **convincing**.

# Developing a Nose for Bullshit



# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.

# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.
- Neither do journalists.

# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.
- Neither do journalists.
- Neither do pundits on the Internet, from blogs to TED talks.

# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.
- Neither do journalists.
- Neither do pundits on the Internet, from blogs to TED talks.
- To say nothing about advertisers, aided by Google, Facebook, and just about everyone else.

# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.
- Neither do journalists.
- Neither do pundits on the Internet, from blogs to TED talks.
- To say nothing about advertisers, aided by Google, Facebook, and just about everyone else.
- Businesses bullshit us to sell us stuff.

# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.
- Neither do journalists.
- Neither do pundits on the Internet, from blogs to TED talks.
- To say nothing about advertisers, aided by Google, Facebook, and just about everyone else.
- Businesses bullshit us to sell us stuff.
- Startup culture elevates it to the level of high art (and VCs are complicit).

# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.
- Neither do journalists.
- Neither do pundits on the Internet, from blogs to TED talks.
- To say nothing about advertisers, aided by Google, Facebook, and just about everyone else.
- Businesses bullshit us to sell us stuff.
- Startup culture elevates it to the level of high art (and VCs are complicit).
- Administration seems to consist solely of it.

# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.
- Neither do journalists.
- Neither do pundits on the Internet, from blogs to TED talks.
- To say nothing about advertisers, aided by Google, Facebook, and just about everyone else.
- Businesses bullshit us to sell us stuff.
- Startup culture elevates it to the level of high art (and VCs are complicit).
- Administration seems to consist solely of it.
- Science and education are not immune.
  - “Cargo cult science”

# Developing a Nose for Bullshit

- Politicians do not feel constrained by facts.
- Neither do journalists.
- Neither do pundits on the Internet, from blogs to TED talks.
- To say nothing about advertisers, aided by Google, Facebook, and just about everyone else.
- Businesses bullshit us to sell us stuff.
- Startup culture elevates it to the level of high art (and VCs are complicit).
- Administration seems to consist solely of it.
- Science and education are not immune.
  - “Cargo cult science”
- *And yet... The world seems to go round and good things are happening... Not everything smells badly...*

# Old School Bullshit

## *Old school business*

Our collective mission is to functionalize customer-driven enterprize solutions for leveraging underutilized portfolio transparencies.

## *Old school / New Age*

We exist as transmissions. To embark on the myth is to become one with it.

## *Computer-generated*

My goal, in essence, is to juxtapose musical semitones. Working binarily means that my focus is always percussively-based, and never orchestral.

## *Somewhat better than the current politicians are capable of*

Just as our forefathers before us, we look to the unending horizons of our great nation with minds fixed and hearts aflame to rekindle the dampened sparks of our collective destiny.

*Adapted from <https://callingbullshit.org>*



# Bullshit in the Age of Data Science

## Quantitative business

Adjusted for international exchange rates, our top-performing global fund beat the market in 7 of the past 9 years.

## Statistics-driven medicine

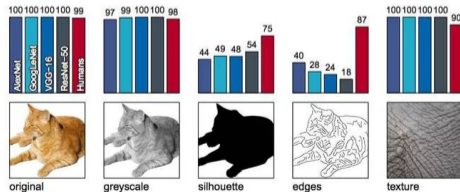
While short of statistical significance after Bonferroni correction ( $p = 0.13$ ), the results underscore a clinically important effect size (relative odds of survival at 5 years = 1.2) that challenges the current therapeutic paradigm.

## Is it real?

34% of behaviorally challenged second-graders admit to having sniffed magic markers at least once in the past year.

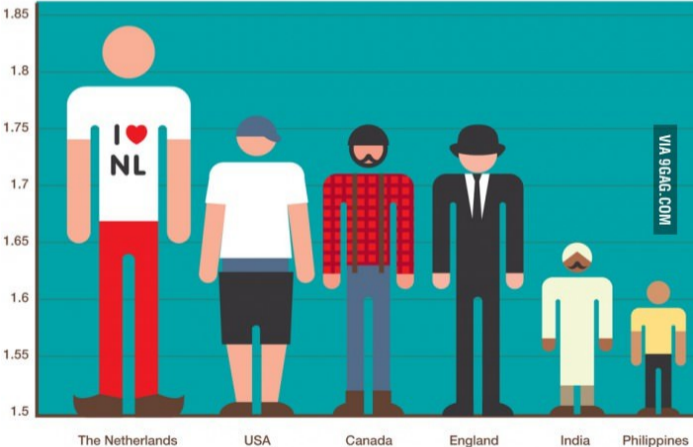
Adapted from <https://callingbullshit.org>

An example of an image being distorted in different ways and the accuracy of neural networks and humans in analyzing it. *Source: Geirhos et al., ICLR 2019.*



## LOOKING DOWN ON THE REST OF THE WORLD

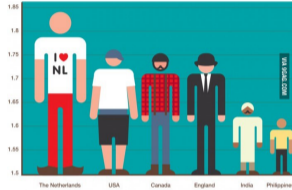
(Average male height in m)



# Beyond the Message: Deciphering Bullshit

## LOOKING DOWN ON THE REST OF THE WORLD

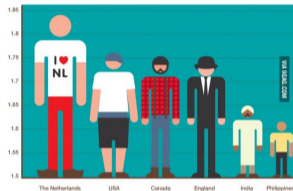
(Average male height in m)



# Beyond the Message: Deciphering Bullshit

## LOOKING DOWN ON THE REST OF THE WORLD

(Average male height in m)



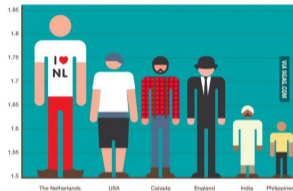
## Some key points:

- Look at the numbers! If there are none — ask why!

# Beyond the Message: Deciphering Bullshit

## LOOKING DOWN ON THE REST OF THE WORLD

(Average male height in m)



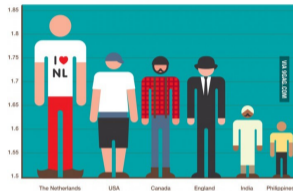
## Some key points:

- Look at the numbers! If there are none — ask why!
- Always strive to get qualitative statements into quantitative ones: the Dutch are the tallest in the world — but by how much? Are quantitative results hidden from view or misrepresented in any way?

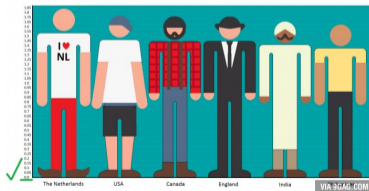
# Beyond the Message: Deciphering Bullshit

## LOOKING DOWN ON THE REST OF THE WORLD

(Average male height in m)



## ACTUAL MALE HEIGHTS TO SCALE



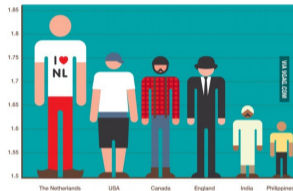
## Some key points:

- Look at the numbers! If there are none — ask why!
- Always strive to get qualitative statements into quantitative ones: the Dutch are the tallest in the world — but by how much? Are quantitative results hidden from view or misrepresented in any way?

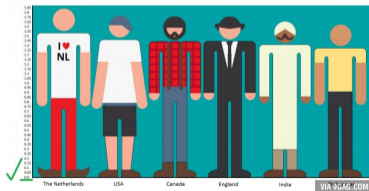
# Beyond the Message: Deciphering Bullshit

## LOOKING DOWN ON THE REST OF THE WORLD

(Average male height in m)



## ACTUAL MALE HEIGHTS TO SCALE



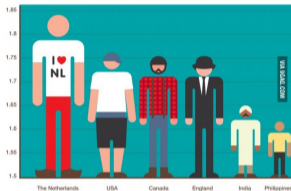
## Some key points:

- Look at the numbers! If there are none — ask why!
- Always strive to get qualitative statements into quantitative ones: the Dutch are the tallest in the world — but by how much? Are quantitative results hidden from view or misrepresented in any way?
- What exactly is being measured? Average height? Tallest persons? Shortest persons? Comparable age groups? Etc., etc.

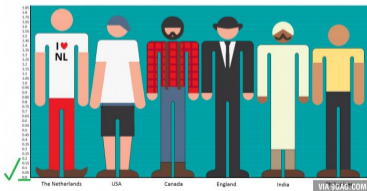
# Beyond the Message: Deciphering Bullshit

## LOOKING DOWN ON THE REST OF THE WORLD

(Average male height in m)



## ACTUAL MALE HEIGHTS TO SCALE



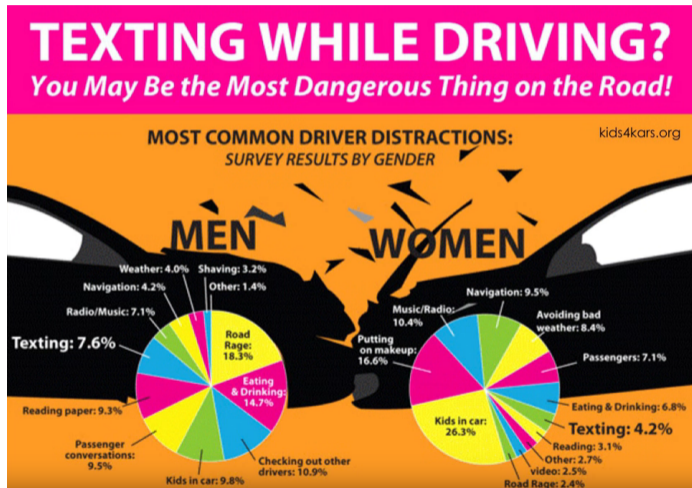
## Some key points:

- Look at the numbers! If there are none — ask why!
- Always strive to get qualitative statements into quantitative ones: the Dutch are the tallest in the world — but by how much? Are quantitative results hidden from view or misrepresented in any way?
- What exactly is being measured? Average height? Tallest persons? Shortest persons? Comparable age groups? Etc., etc.
- Consider applicability limits.

# Beyond the Message: More Bullshit Still



# Beyond the Headline and the Graphics

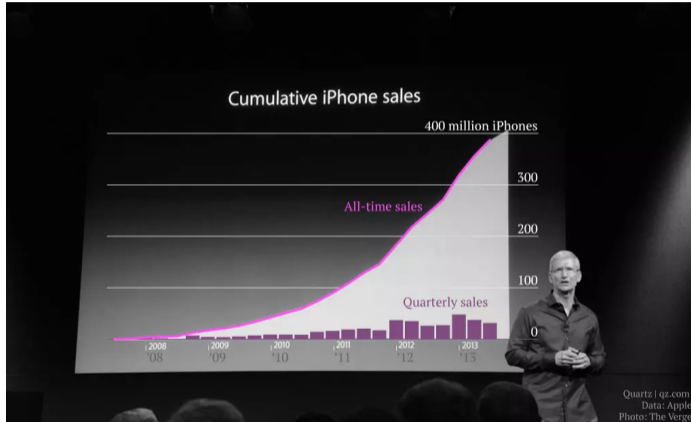


# The Art of the Presentation



*Apple's iPhone presentation, September 6, 2013*

# Lying — And Calling Bullshit — Effectively



Adapted from <https://qz.com/122921/the-chart-tim-cook-doesnt-want-you-to-see/>