MSE-213
Probability and statistics for materials science

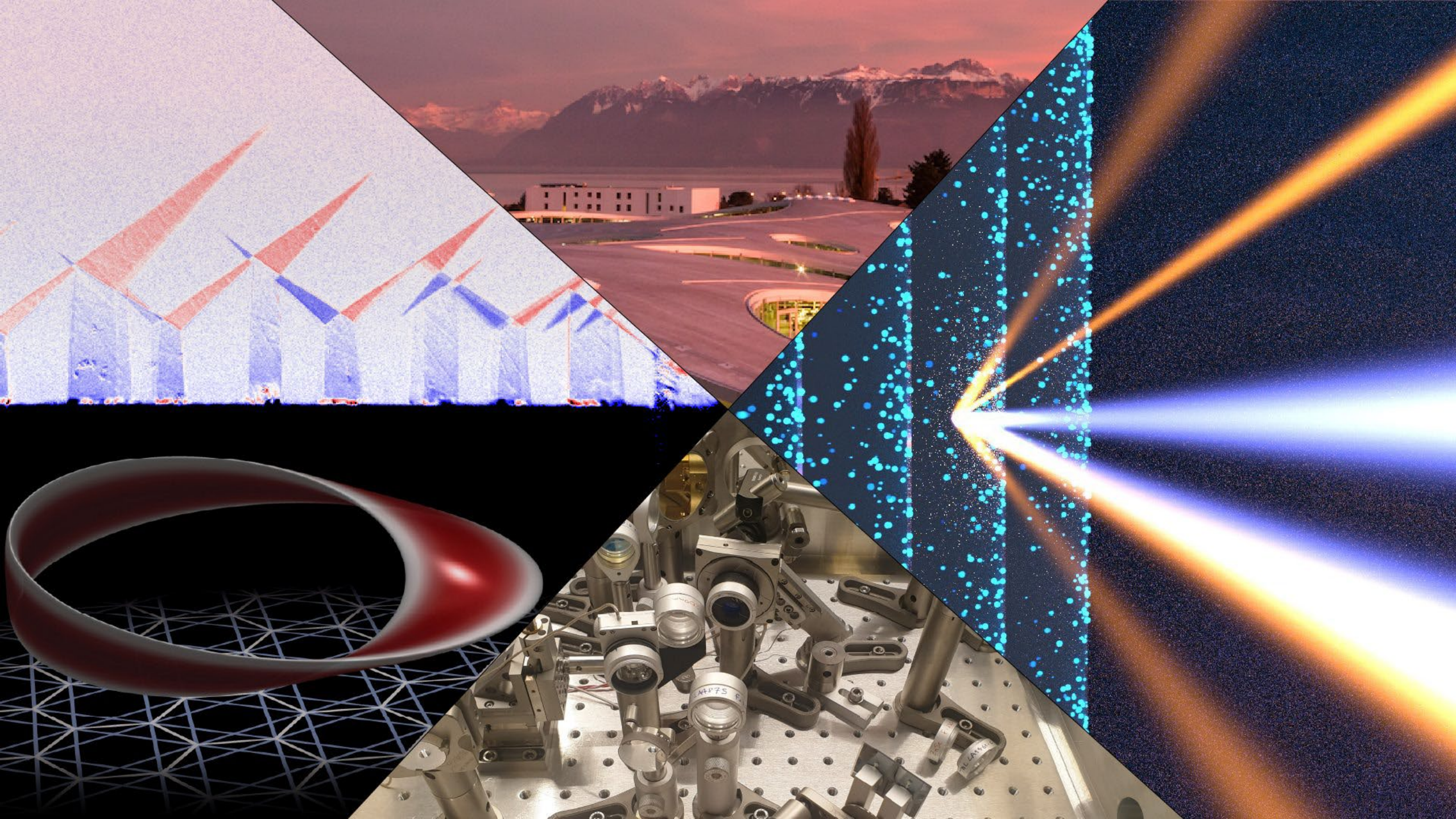
Gregor Jotzu

The Team

Arslan Mazitov
Laboratory of Computational Science and Modelling

Malo Herve
Dynamic Quantum Materials Laboratory

“Dynamic quantum materials laboratory”



MSE-213
Probability and statistics for materials science

MSE-213
Probability and statistics for materials science

and data analysis

MSE-213
Probability and statistics for materials science

and data analysis

and “pragmatic” Python

MSE-213
Probability and statistics for materials science

and data analysis

and “pragmatic” Python

and experiment design

MSE-213
Probability and statistics for materials science

and data analysis

and “pragmatic” Python

and experiment design

and how to be a responsible citizen with a scientific-technical background

MSE-213
Probability and statistics for materials science

and data analysis

and “pragmatic” Python

and experiment design

and how to be a responsible citizen with a scientific-technical background

and a bit of philosophy of science

Probability & statistics is everywhere

everywhere

8h ago 19.36 CEST

Polls finds Harris victorious in debate, building lead over Trump nationally

Reuters is out with one of the first polls since **Kamala Harris** and **Donald Trump** debated on Tuesday evening, and found voters view the vice-president as the winner in the debate, where Trump came off as the less sharp candidate.

Harris is also building her lead among registered voters nationally to 47% over Trump's 42%, a slight jump from previous weeks, according to the survey, which was conducted with Ipsos. Here's more about it:

> *The two-day poll showed Harris with a five percentage point lead among registered voters, just above the four-point advantage she had over Trump in an Aug 21-28 Reuters/Ipsos poll.*
>
> *Among voters who said they had heard at least something about Tuesday's debate, 53% said Harris won and 24% said Trump won. Some 52% of respondents said that Trump stumbled and didn't appear sharp, while 21% said that of Harris.*
>
> *Harris, 59, put Trump, 78, on the defensive with a stream of attacks on his fitness for office and his myriad legal woes.*
>
> *The poll surveyed 1,690 US adults nationwide, including 1,405 registered voters. It had a margin of error of around three percentage points for registered voters.*

Share

Updated at 20.49 CEST

Suède: le cancer de la peau recule chez les moins de 50 ans

Depuis 2015, l'incidence de la maladie chez les jeunes adultes a baissé de 5% par an dans le pays, révèle une étude publiée lundi. Une première en Europe.

Probability & statistics is everywhere

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Office fédéral de la statistique

Probability & statistics is everywhere

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Office fédéral de la statistique

Probability & statistics is everywhere

and doing it wrong can be deadly...

Expert told court multiple cot deaths 'one in 73m'

Roy Meadow gave evidence in four prominent trials of mothers accused of killing their babies, resulting in three convictions, reports **Jane Kirby**

The Telegraph News Sport Business Money Opinion Ukraine US election Travel Health

Wrongly-jailed Sally Clark died from drink

Probability & statistics is everywhere

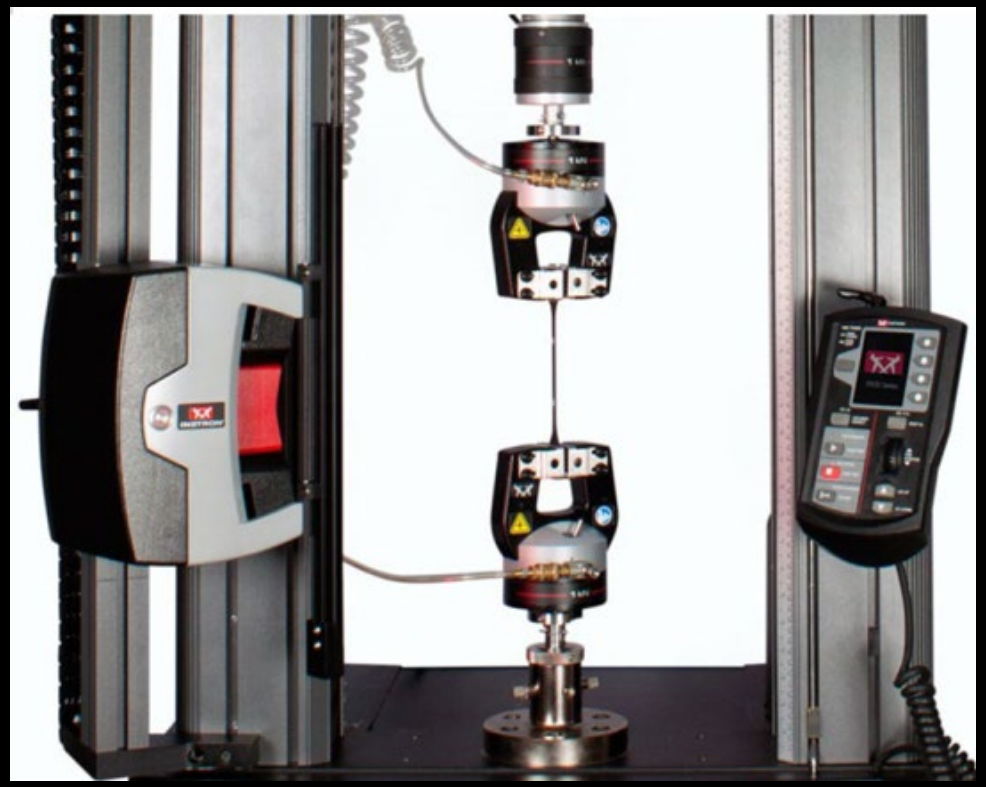
and doing it wrong can be deadly...

Expert told court multiple cot deaths 'one in 73m'

Roy Meadow gave evidence in four prominent trials of mothers accused of killing their babies, resulting in three convictions, reports **Jane Kirby**

What this course is for

- Thermodynamics, quantum, statistical mechanics
- Your research projects



What this course is for

- Thermodynamics, quantum, statistical mechanics
- Your research projects

- Your future work in
  - ... industrial/academic research,
  - ... government (climate, demographic change, migration...)
  - ... machine learning/AI/big data
  - ... starting your own company

What this course is for

- Thermodynamics, quantum, statistical mechanics
- Your research projects

- Your future work in
  - ... industrial/academic research,
  - ... government (climate, demographic change, migration...)
  - ... machine learning/AI/big data
  - ... starting your own company

+ "pragmatic" Python for all of the above
and to automate sending your e-mails ;)

Organizational matters

- Feedback, please!

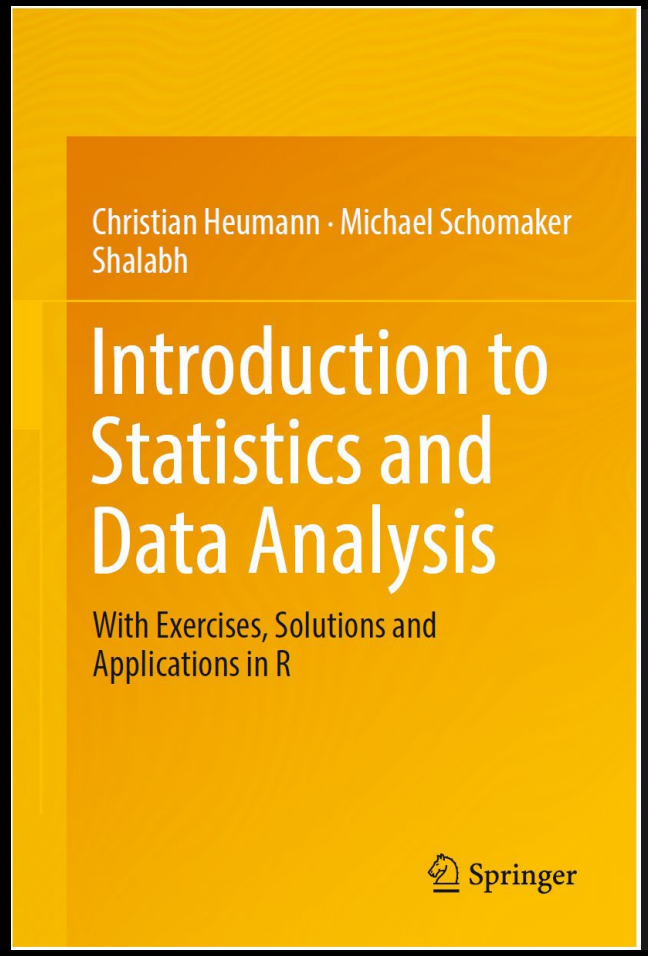
  > gregor.jotzu@epfl.ch
  use anonymousemail.me
  if you prefer

- Lecture, Lecture, Last exercise, Python & new exercise

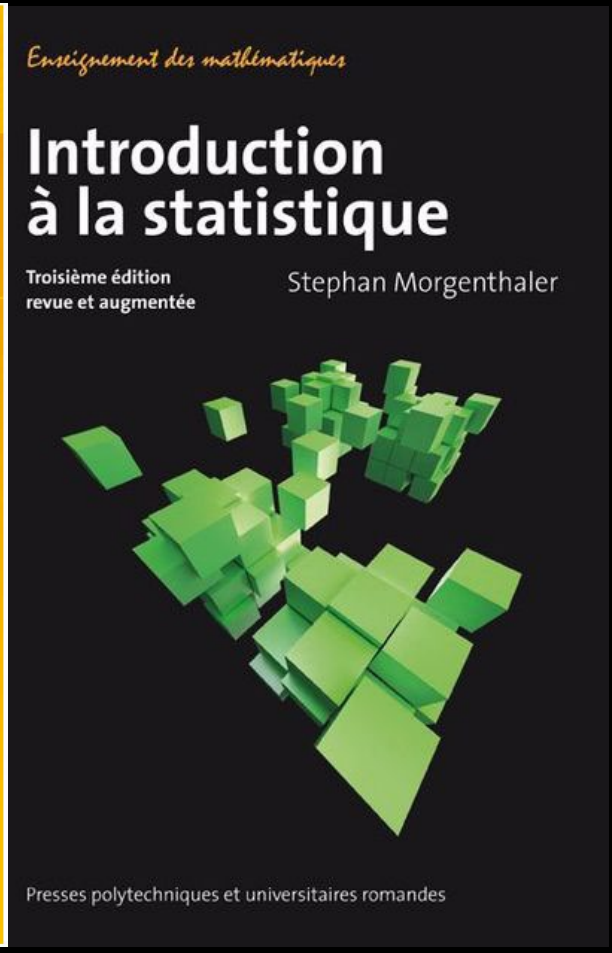
- Bring your computers please, leave your phones in your pocket/bag

- Slides, Exercises, Python scripts will be on moodle

1 marks text on the blackboard



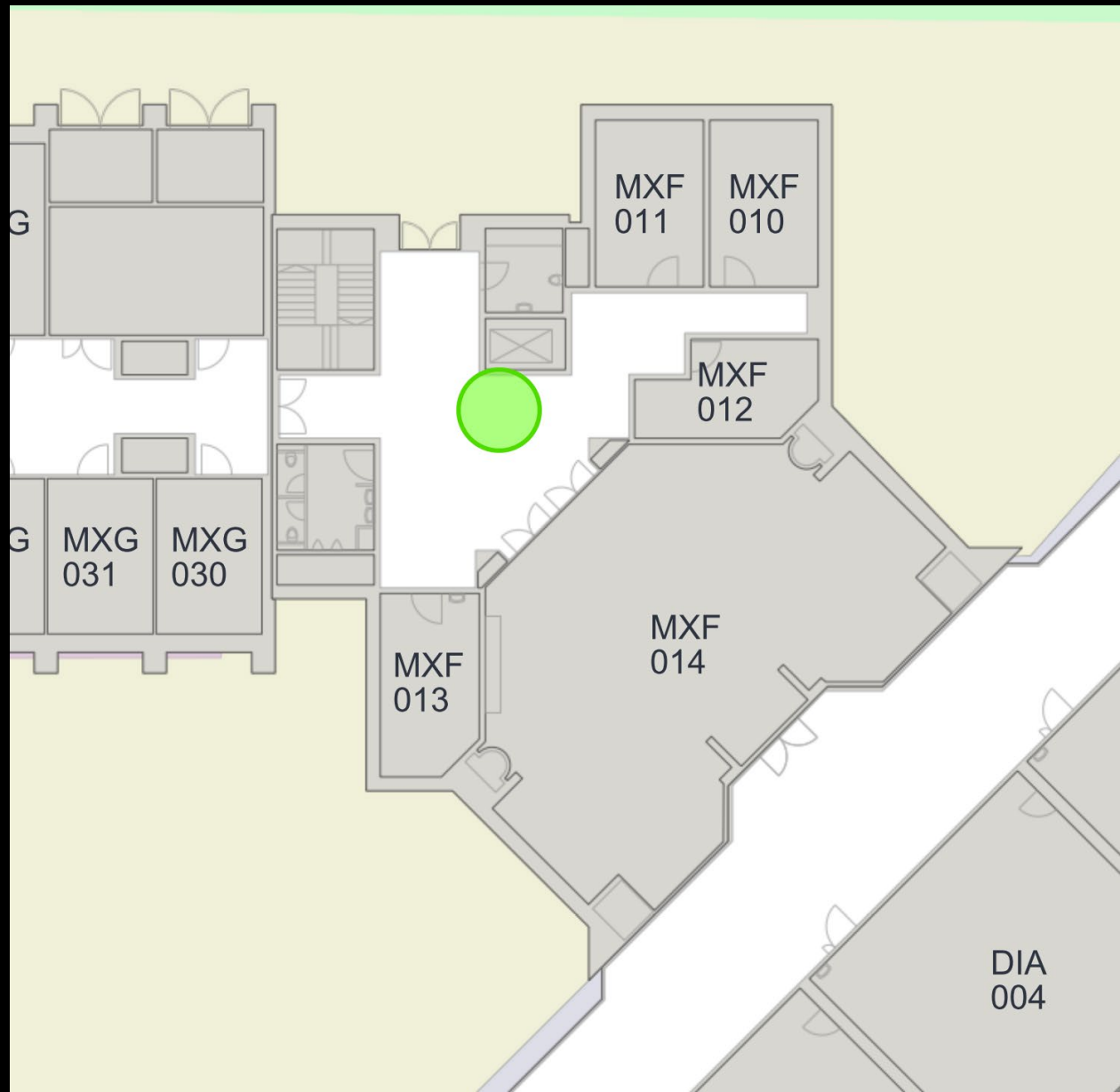
Heumann & Schomaker



Morgenthaler

Organizational matters

- Exercises will be on moodle
- You can hand them in by Wednesday 13:00 in
  - Moodle
  - mse213@groupes.epfl.ch
  - My mailbox, Station 12, MXF 094
- Solutions will be posted online afterwards



THE EXAM

- Written (on paper)
- There will be “pragmatic” questions on Python
- You can bring notes (1 A4 page, two-sided).
- You **should** bring a non-programmable non-graphic calculator

Real data can have real problems

Real data can have real problems

Liam Thorp
@LiamThorpECHO

So I'm not getting a vaccine next week - was feeling weird about why I'd been selected ahead of others so rang GP to check. Turns out they had my height as 6.2cm rather than 6 ft 2, giving me a BMI of 28,000 😂

9:43 AM · Feb 17, 2021

Real data can have real problems

Liam Thorp
@LiamThorpECHO

So I'm not getting a vaccine next week - was feeling weird about why I'd been selected ahead of others so rang GP to check. Turns out they had my height as 6.2cm rather than 6 ft 2, giving me a BMI of 28,000 😂

9:43 AM · Feb 17, 2021

Statistics is different

- Often no “single” answer – think critically!

Statistics is different

- Often no “single” answer – think critically!
- Random, but surprisingly robust

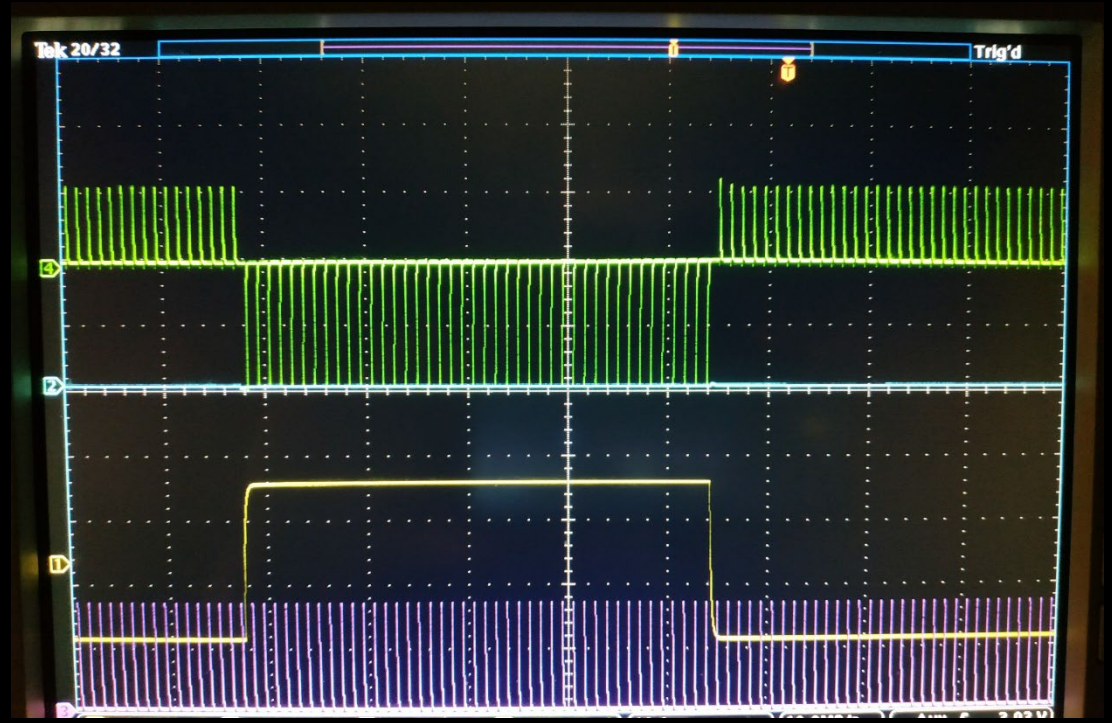
Statistics is different

- Often no “single” answer – think critically!
- Random, but surprisingly robust
- Notation is NOT consistent

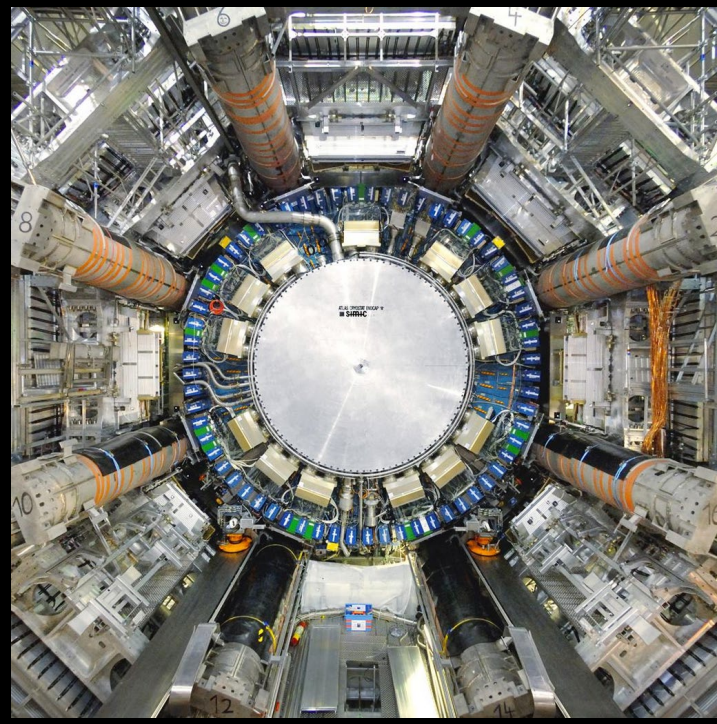
Statistics is different

- The best use depends on the situation

SML@EPFL



DQML@EPFL



ATLAS@CERN

Exploring/Inspecting Data

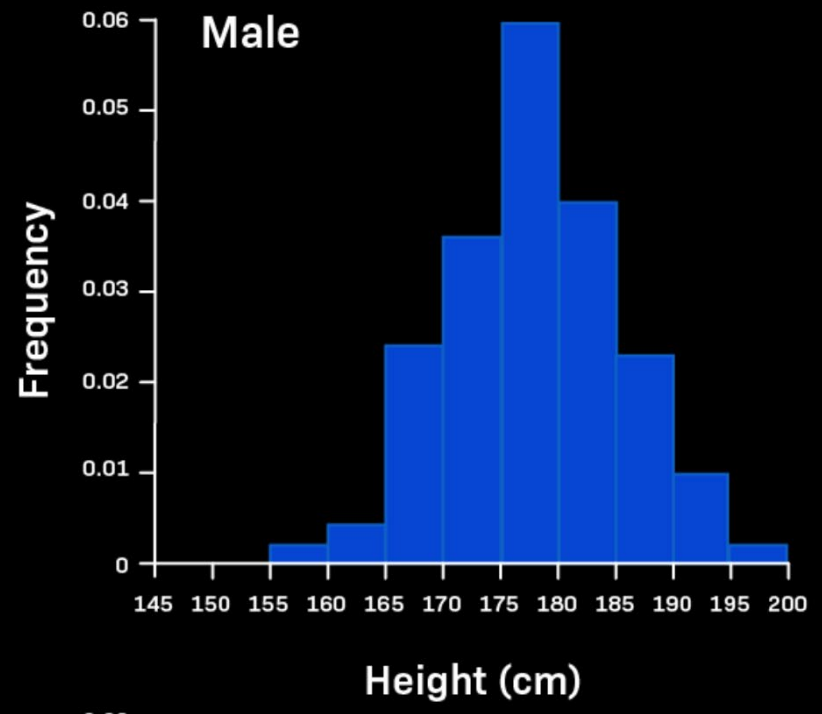
Types of data

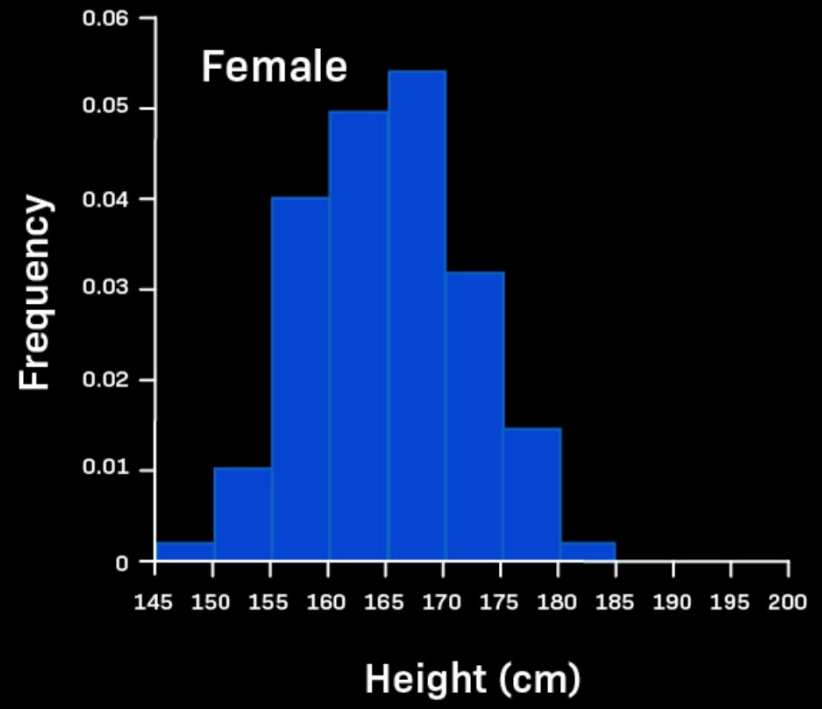
Absolute & Relative Frequency

Histogram

Binning Data

Binning Data





Reducing Data

Single-value characteristics: max, min, mean, mode, median

```
[50, 55, 45, 52, 21, 15, 55]
```

Single-value characteristics: max, min, mean, mode, median

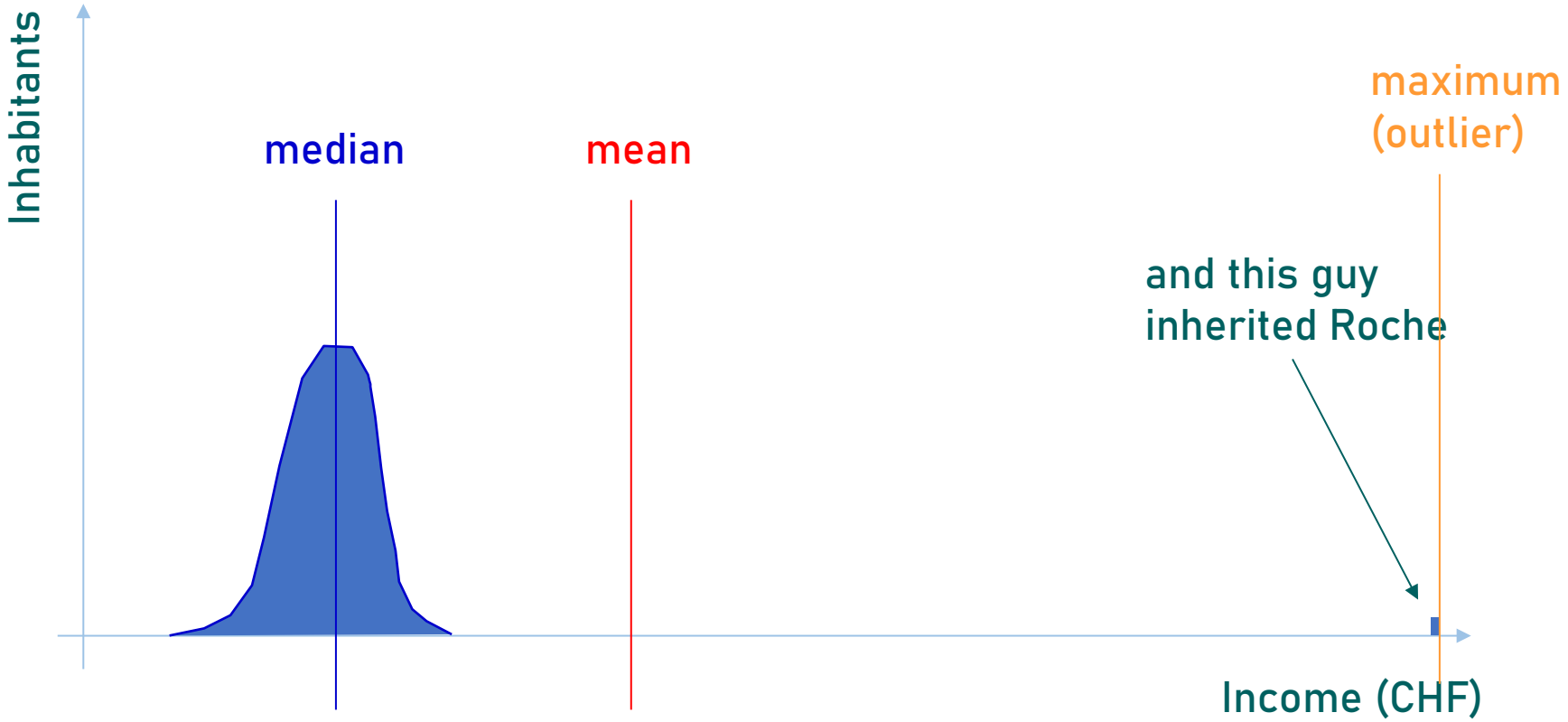
```
[15, 21, 45, 50, 52, 55, 55]
```

Limits of data reduction

Vaux-sur-Morges
169 Inhabitants

Average annual income
338.779CHF / person

Limits of data reduction



Quantiles – Percentiles, Quartiles

Quantiles – Percentiles, Quartiles

[15, 21, 45, 50, 52, 55, 55]

Quantiles – Different definitions

method : ***str, optional***

This parameter specifies the method to use for estimating the quantile. There are many different methods, some unique to NumPy. The recommended options, numbered as they appear in [1], are:

1. 'inverted_cdf'
2. 'averaged_inverted_cdf'
3. 'closest_observation'
4. 'interpolated_inverted_cdf'
5. 'hazen'
6. 'weibull'
7. 'linear' (default)
8. 'median_unbiased'
9. 'normal_unbiased'

The first three methods are discontinuous. For backward compatibility with previous versions of NumPy, the following discontinuous variations of the default 'linear' (7.) option are available:

- 'lower'
- 'higher',
- 'midpoint'
- 'nearest'

See Notes for details.

Empirical Cumulative Distribution Function (ECDF)

[15, 21, 45, 50, 52, 55, 55]

The “Box Plot”

[15, 21, 45, 50, 52, 55, 57]

The “Box Plot”- with “outliers”

[-10, 21, 45, 50, 52, 55, 57]