# IN2P3 School of Statistics

Yann Coadou
*for the SOS organising committee*

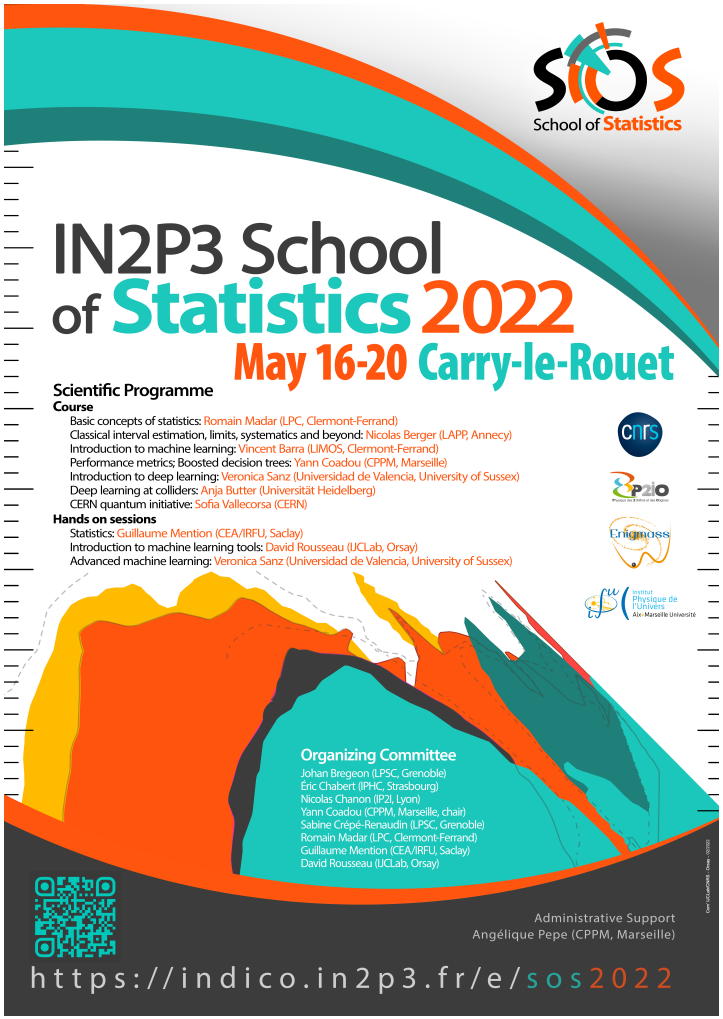
CPPM Marseille

IN2P3/IRFU Machine Learning workshop
APC, 27 September 2022



# School presentation

- **Origin:**
  - strong demand from IN2P3 community for statistical analysis training
  - usually not much developed during initial training
  - nowadays key ingredient of scientific results production
- **Goal:** *describe, explain and manipulate statistical concepts and tools necessary to perform statistical analysis of data in particle physics, astroparticle physics and cosmology*
  - proper usage of tools
  - real impact in day-to-day work
  - competitiveness and visibility in our big collaborations
- **Evolution:** Over the years, increased part related to machine learning and hands-on sessions
- **Target audience:** PhD students, postdocs, staff (junior and senior) interested in basic concepts and new tools



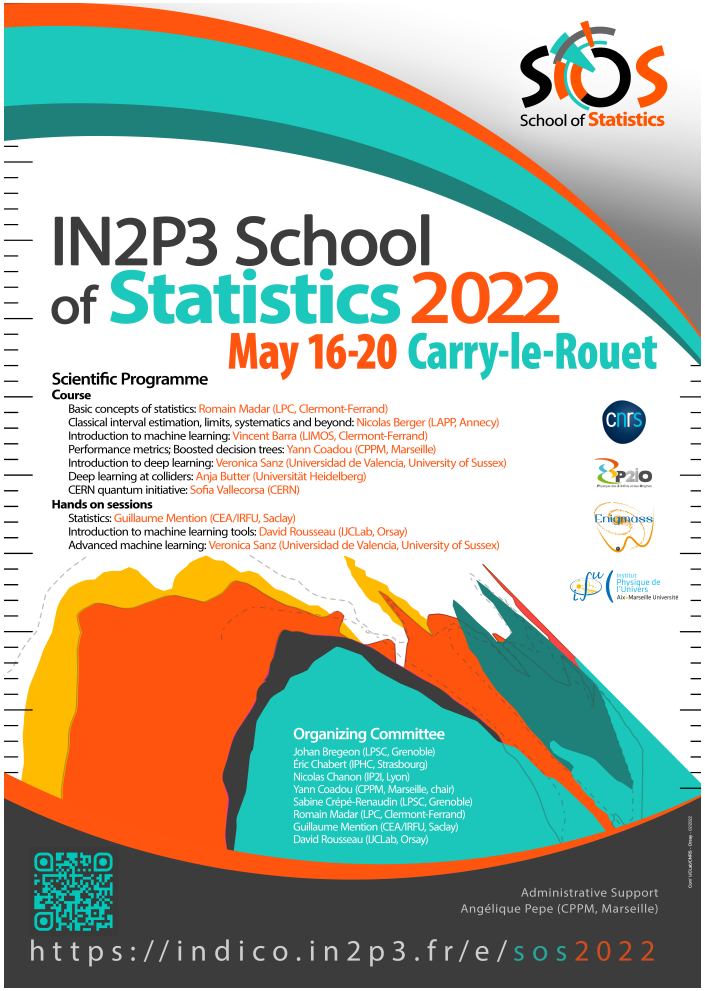

- 8th edition of successful biennial series
  ... after Strasbourg (2008), Autrans (2010–2016), La Londe-les-Maures (2018), Zoom (2021)
- Back to in-person mode
- 16–20 June 2022 in Carry-le-Rouet (13)
- Attendance: PhD students, postdocs, engineers
- In English (many foreign students in our labs)
- Indico: ▸ https://indico.in2p3.fr/e/SOS2022
- General web site: ▸ http://sos.in2p3.fr
- Financial support:

cnrs Enigmass

Institut Physique de l'Univers Aix*Marseille Université

P2IO Physics of 2 Infinities and Origins

# People behind the scene




▸ https://indico.in2p3.fr/e/SOS2022

▸ https://sos.in2p3.fr

## Organising committee

- Johan Bregeon (LPSC, Grenoble)
- Éric Chabert (IPHC, Strasbourg)
- Nicolas Chanon (IP2I, Lyon)
- Yann Coadou (CPPM, Marseille) –chair–
- Sabine Crépé-Renaudin (LPSC, Grenoble)
- Romain Madar (LPC, Clermont-Ferrand)
- Guillaume Mention (CEA/IRFU, Saclay)
- David Rousseau (IJClab, Orsay)

## Local support at CPPM

- Angélique Pèpe, Véronique Roux

# School programme

- **Three main components:**
  - fundamental concepts (statistics, limits, systematics)
  - machine learning: concepts, BDT, deep learning
  - hands-on sessions (statistics, ML)
- **Time for discussion:**
  - good to be back in-person

| | Mon 16 | Tue 17 | Wed 18 | Thu 19 | Fri 20 |
|---|---|---|---|---|---|
| 8:30-9:00 | | | | | Hands-on: advanced machine learning |
| 9:00-10:30 | Arrival | Machine learning | Intervals & limits I | Intervals & limits II | |
| 10:30-11:00 | | Break | | | |
| 11:00-12:30 | | Boosted decision trees; Metrics | Hands-on: basics machine learning | Deep learning I | CERN quantum technology initiative |
| 12:30-14:00 | Lunch | | | | |
| 14:00-15:30 | Basic concepts I | Hands-on: basic statistics | Free | Deep learning II | Departure |
| 15:30-16:00 | Break | | | Break | |
| 16:00-17:30 | Basic concepts II | | | Deep learning at colliders | |

# People in front of the screen

## Lectures

- Romain Madar (LPC, Clermont-Ferrand): Basic concepts
- Vincent Barra (LIMOS, Clermont-Ferrand): Machine learning
- Yann Coadou (CPPM, Marseille): Boosted decision trees; metrics
- Nicolas Berger (LAPP, Annecy): Intervals & limits
- Veronica Sanz (Universities of Valencia and Sussex): Deep learning
- Anja Butter (ITP Universität Heidelberg): Deep learning at colliders
- Sofia Vallecorsa (CERN): CERN quantum technology initiative

## Hands-on sessions

- Guillaume Mention (CEA/IRFU, Saclay): basics of statistics
- David Rousseau (IJClab, Orsay): Introduction to ML tools
- Veronica Sanz (Valencia and Sussex): advanced machine learning

# People watching

# Machine learning @ SOS 2022

- Complete overview of founding principles
- Metrics to evaluate ML algorithms in HEP
- Boosted decision trees
- Deep learning
  - basics of neural networks: empirical risk, gradient descent, backpropagation
  - DNN / CNN / RNN / LSTM
  - autoencoders
  - generative models
  - clustering algorithms
- Deep learning at colliders: top tagging, event generation, detector simulation, unfolding, uncertainty estimation
- CERN QTI: focus on quantum machine learning
- Hands-on

Advice: watch video recording of lectures

▸ https://indico.in2p3.fr/e/SOS2022

# Hands-on sessions

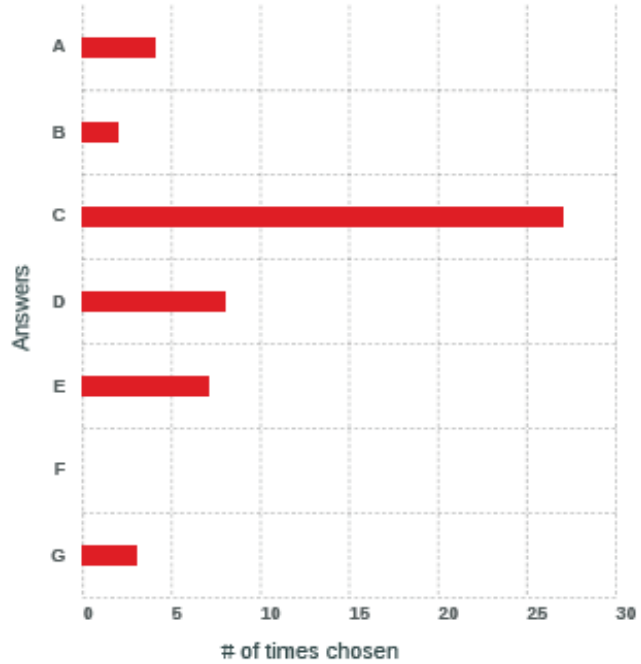
- All done with Jupyter notebooks, all available in indico
- Can run locally (install anaconda) or on Google Colab
- **Statistics**
  - basic python packages (NumPy, SciPy, MatPlotLib)
  - probabilities
  - sampling
  - statistical inference
- **Introduction to machine learning**
  - dataset handling, features
  - BDT with XGBoost, LightGBM, scikit-learn GBDT
  - hyperparameter optimisation
  - figures of merit
  - variable selection (feature vs. permutation importance)
- **Advanced machine learning**
  - exploratory data analysis
  - deep networks
  - unsupervised learning with clustering: PCA, k-means, t-SNE, UMAP

# SOS 2022: know your audience

- 46 participants (40 survey answers)
  - 39 PhD students (85%)
  - 2 postdocs (4%)
  - 5 engineers (11%)
- No Zoom link during school (but video recording)
  - favour in-person
  - not a showstopper for staff anymore?

***You work in (check all relevant boxes)***

Answered: 40
**A. nuclear physics**: 4 (7.84%)
**B. theory**: 2 (3.92%)
**C. particle physics**: 27 (52.94%)
**D. astro / cosmo**: 8 (15.69%)
**E. neutrino**: 7 (13.73%)
**F. private company**: 0 (0.00%)
**G. other**: 3 (5.88%)




***if you chose "other", please specify:***

Answered: 2
- DAQ Engineer
- Nuclear imaging

# SOS 2022: courses balance

***How did you find the balance between basic concepts and advanced methods?***

Answered: 40
**A. I would have preferred to spend more time on basics**: 12 (30.00%)
**B. I would have preferred to spend more time on advanced methods**: 11 (27.50%)
**C. The balance was good**: 17 (42.50%)

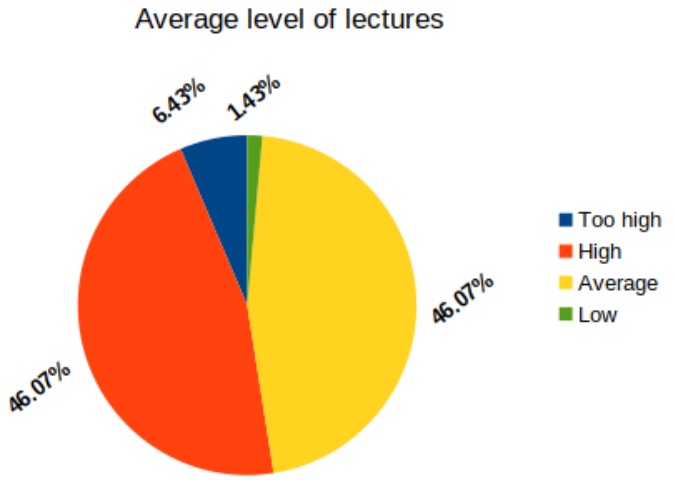

***How did you find the balance between lectures and hands-on sessions?***

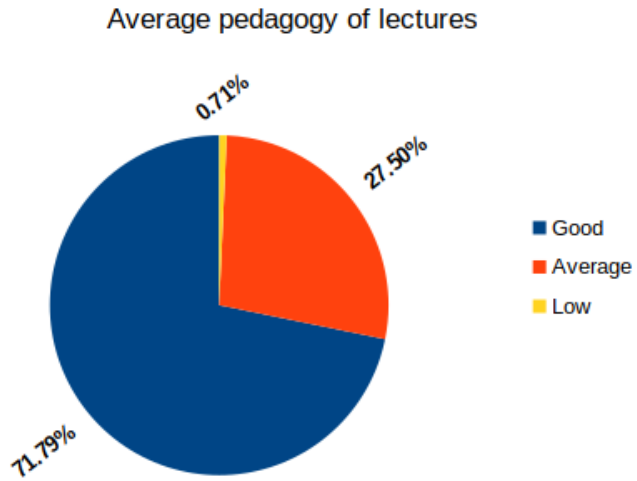
Answered: 40
**A. I would have preferred to spend more time on lectures**: 5 (12.50%)
**B. I would have preferred to spend more time on hands-on sessions**: 13 (32.50%)
**C. The balance was good**: 22 (55.00%)


- Right balance between topics
- Hard to have more hands-on, possibly some code examples within lectures (done in limits lecture)

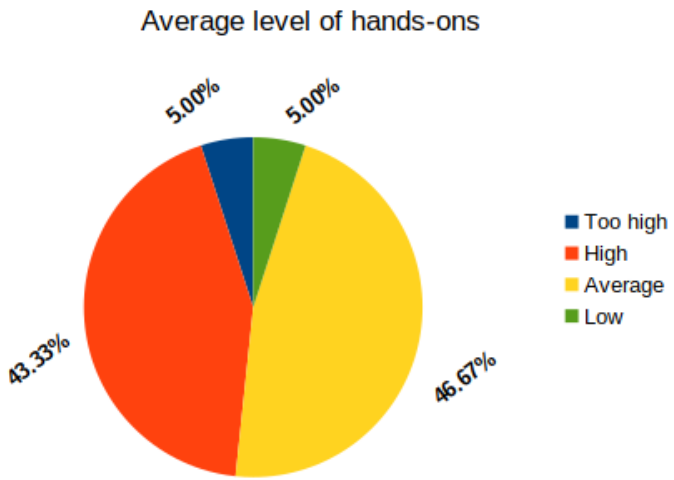
# SOS 2022: courses quality

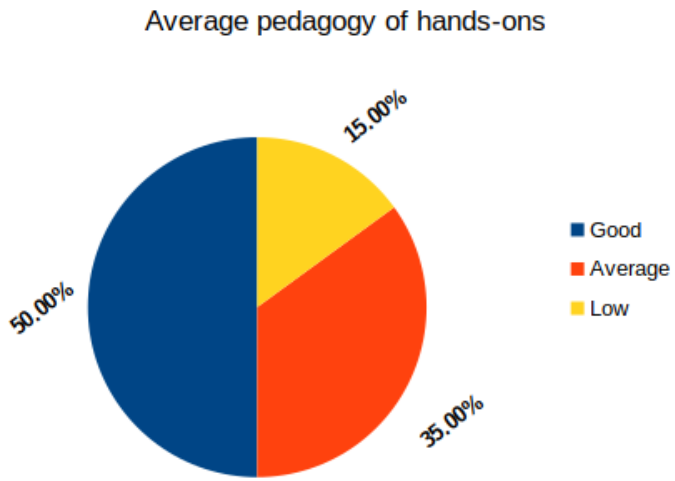
Lectures




Hands-on



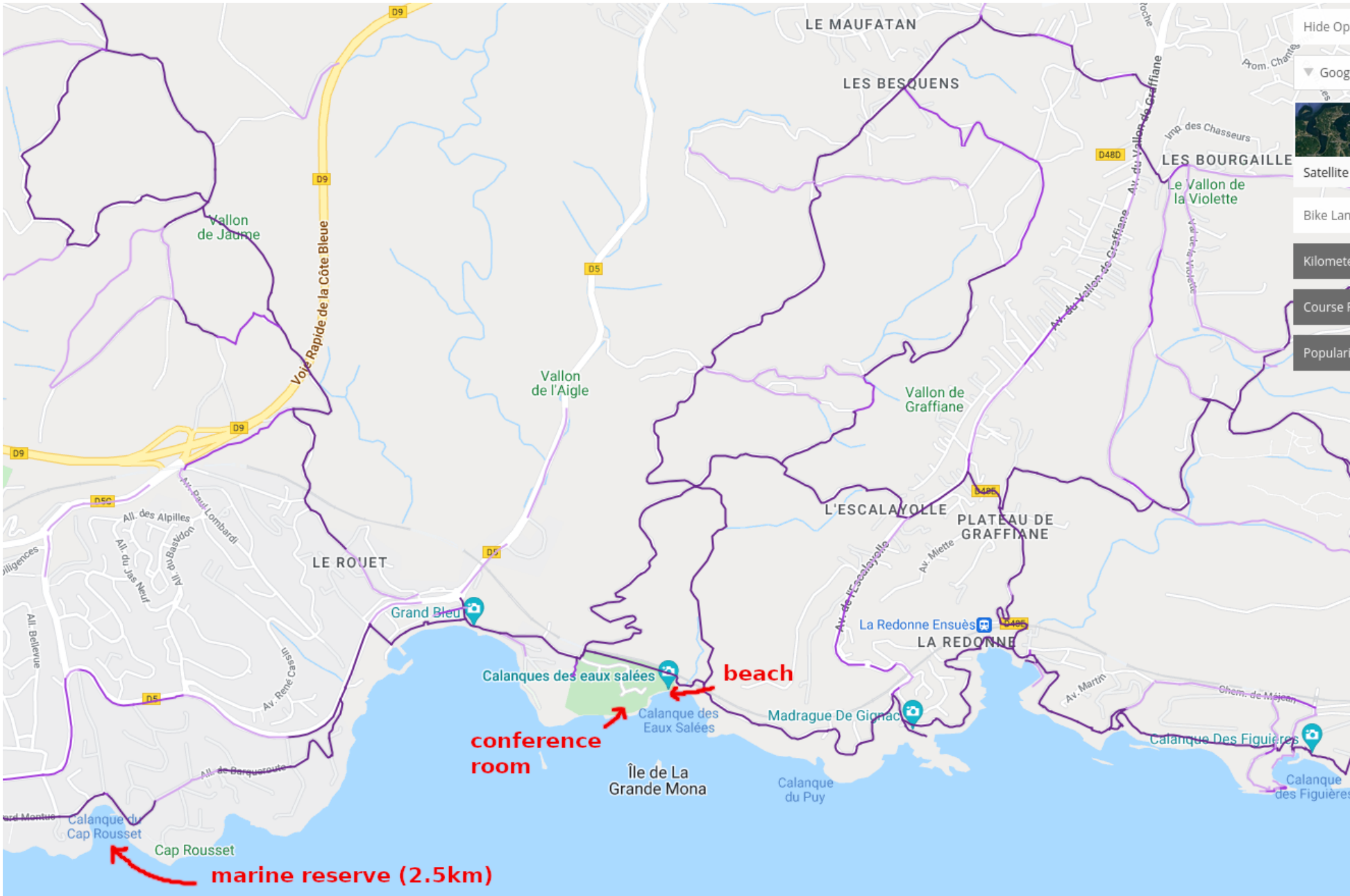




- Brainstorming about ways to improve hands-on experience

# Next edition

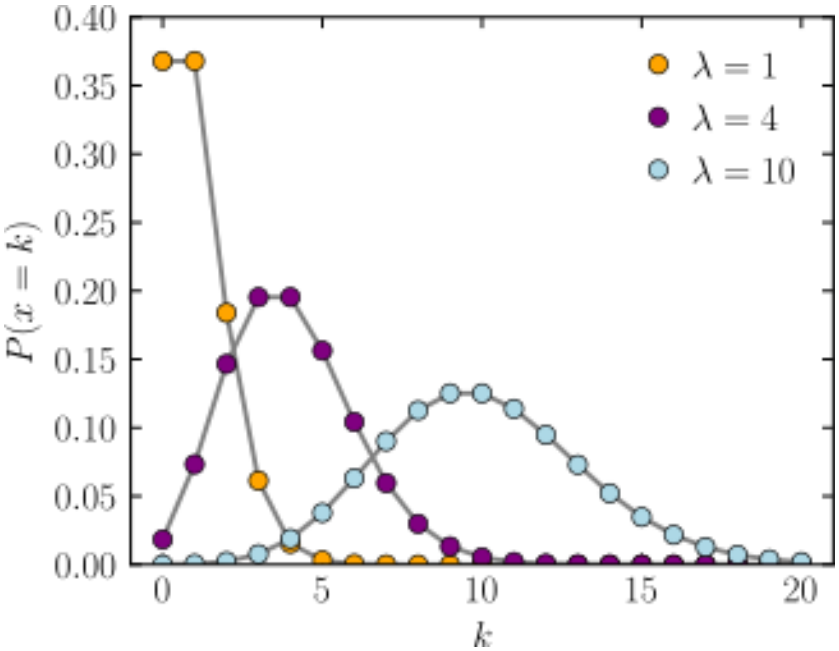
- Most likely in May 2024:
  - in person
  - one week in a relatively isolated place to foster discussions
  - possibly in Carry again
- Interested in joining the organisation? Come and talk to me!

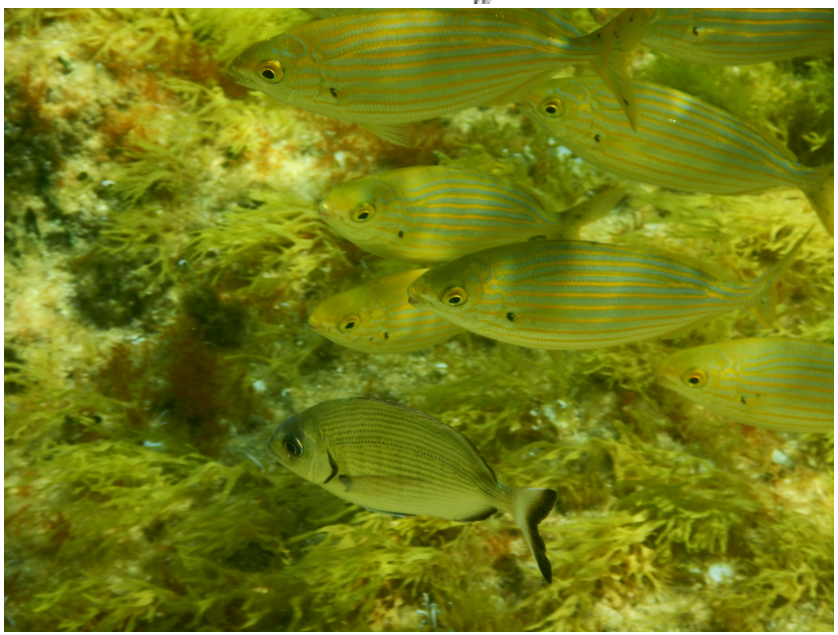
# Carry heatmap

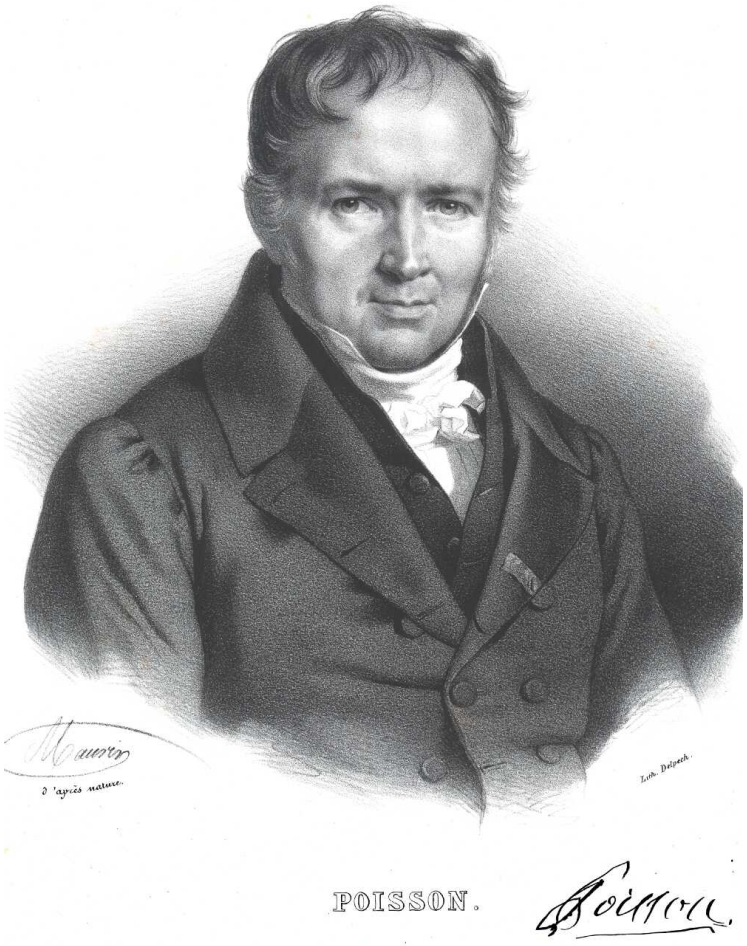


# Neurons making decisions. . .

# The meanings of “poisson” in Carry






```
[ ] import scipy.stats
    scipy.stats.poisson
```