Mathematical methods for neuroscience

Introduction

Course overview

Mathematical methods for neuroscience BIO-612

Dr. Lida Kanari

Email: lida.kanari@epfl.ch

EPFL/BBP, Campus Biotech, Geneva

Where: EPFL, AAC006

When: Tuesdays 13:15 - 15:00 (exercises 15:00 - 16:00)

We will have a few experts to give invited talks towards the end of the semester.

Format of the course

- Each week: theoretical questions + computational implementations

  Due to deliver: next week

- Project: computational project

  Due to deliver: ~2 weeks before the end of the course

- Presentations of projects

  Last course

Details

- Attend the lectures each week: Tuesdays 13:15 - 15:00
- Attend the exercise session (work in groups is usually more efficient)
- Hand-in exercises by next Tuesday (deadline 12:00, before the next course) exercises should be handed **per person** and replies should be independent even if you worked with a group for part or all exercises (please in that case indicate the members of your group)

Materials / resources

- Execution of tutorial python notebooks happens online under: https://noto.epfl.ch

(if you really want, you can download the tutorials and necessary software locally but at your own risk)

- Moodle:
  - Slides will be up on the day of the lecture
  - Exercises will be up on the day of the lecture
  - Relevant references will also be included (papers, book chapters etc)

Project (teams of ~2-3 ppl)

- Theoretical part
- Computational part
- Report / Presentation

Projects

- Groups and projects will be selected (28/03)
- Project ideas will be discussed later in the course

Course grade

- ❖ 30% exercises assignments (~10 exercise assignments over 14 weeks, the best 9 will contribute to the grade)
- ❖ 50% project report (~2 weeks before the end of the course)
- ❖ 20% project presentation (last week of the course)

Outline and Objectives of the course

How to “mathematatize” a problem:

- Translate your problem to mathematical terms
- Set up equations
- Use mathematical tools to a) parametrize b) solve your equations
- Implement solutions computationally

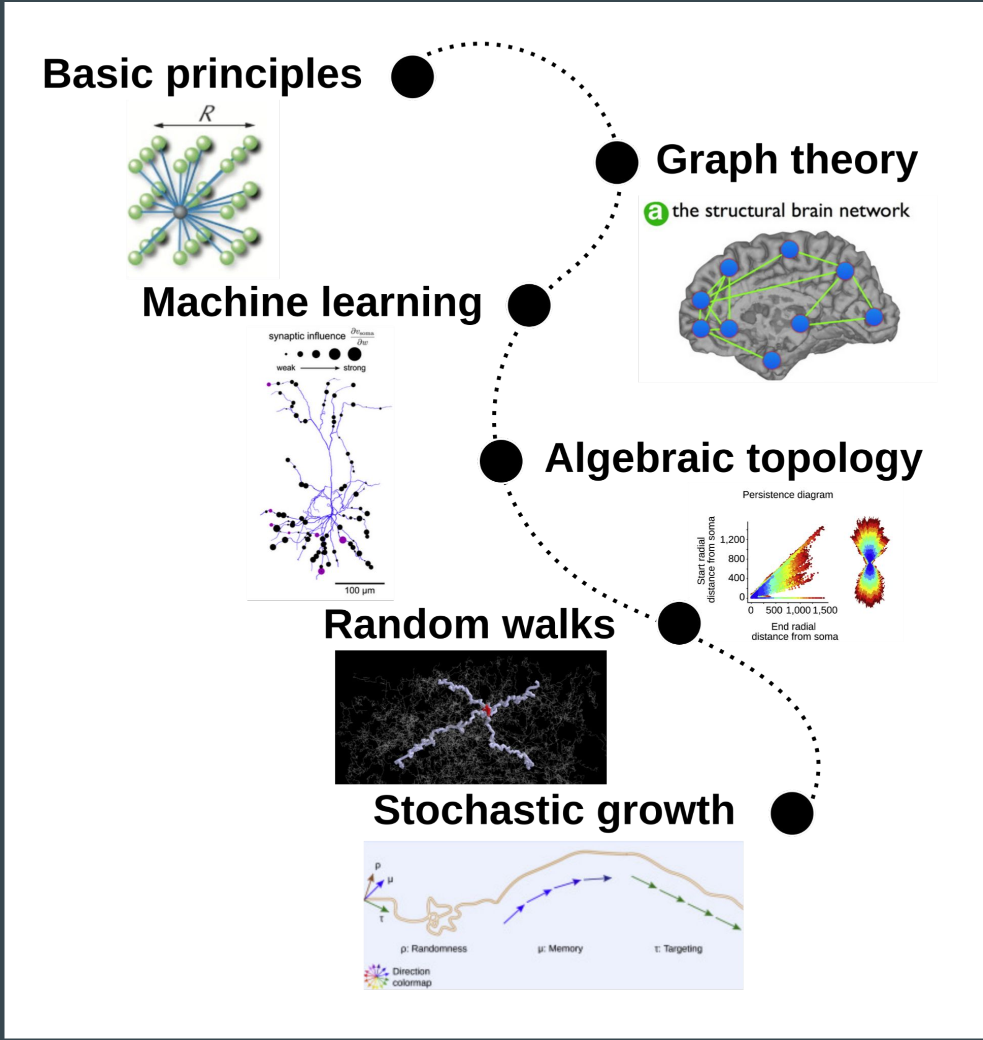
Important points!

- Select appropriate methods
- Learn the language - terminology
- Understand the limitations

Outline and Objectives of the course

- Statistical methods
- Machine learning
- Classification - Discrimination tasks
- Algebraic topology
- Random walks
- Stochastic growth

Outline


Basic principles
R
Graph theory
a the structural brain network
Machine learning
synaptic influence $\frac{\partial v_{soma}}{\partial w}$
weak
strong
100 µm
Algebraic topology
Persistence diagram
Start radial distance from soma
1,200
800
400
0
0
500
1,000
1,500
End radial distance from soma
Random walks
Stochastic growth
ρ
μ
τ
ρ: Randomness
μ: Memory
τ: Targeting
Direction colormap

Questionnaire

Now we will have a small questionnaire (not graded) about your background and where we should focus

Go to:

https://freeonlinesurveys.com/s/yYkJwZbx

Deconstructing a problem into fundamental components

Deconstructing a problem

Q: How does the brain give rise to behavior and cognition?

Deconstructing a problem

Q: How does the brain give rise to behavior and cognition?

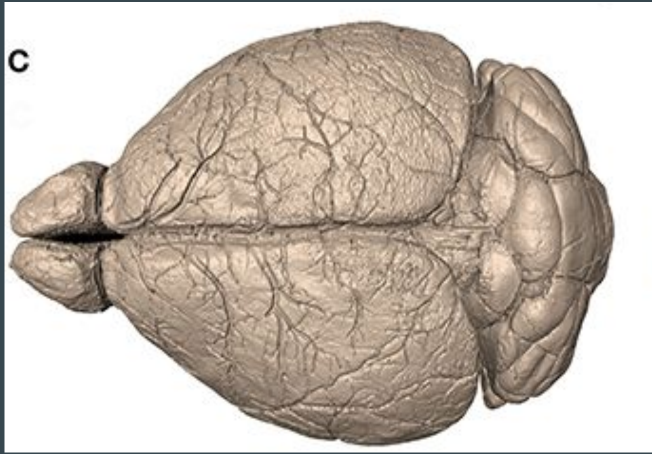
Deconstructing a problem

Q: How does the brain give rise to behavior and cognition?

What is the brain? Structural and functional properties.
What is behavior? Phenotype of actions.
What is cognition? How actions are decided?
How are the previous three related?

Deconstructing a problem

Let's examine each one of the fundamental components: *brain*



Xiong et al. 2017

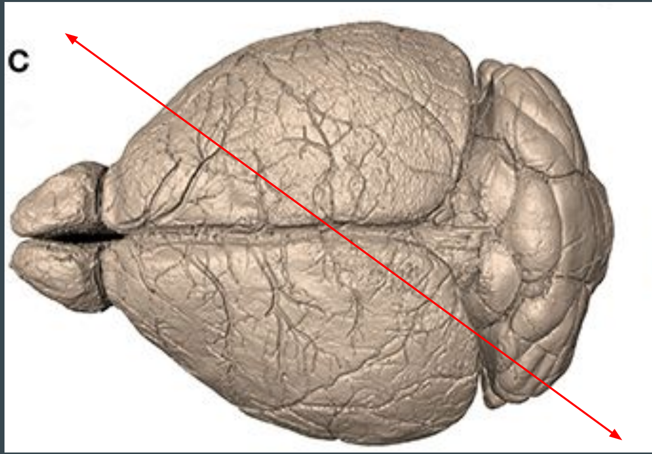
Deconstructing a problem

What is the brain?

How big is the brain?

How can we measure its dimensions?

Geometry



Xiong et al. 2017

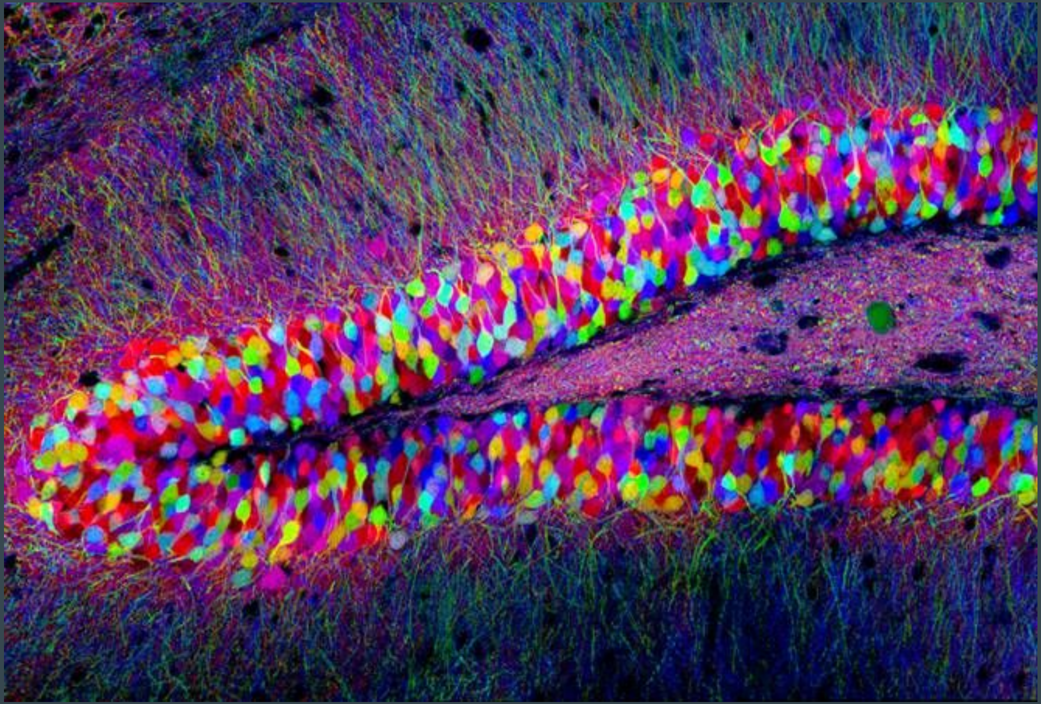
Deconstructing a problem

What is the brain? If we zoom in ... we see *neurons*

How many neurons we see?

How are they distributed?

Statistics



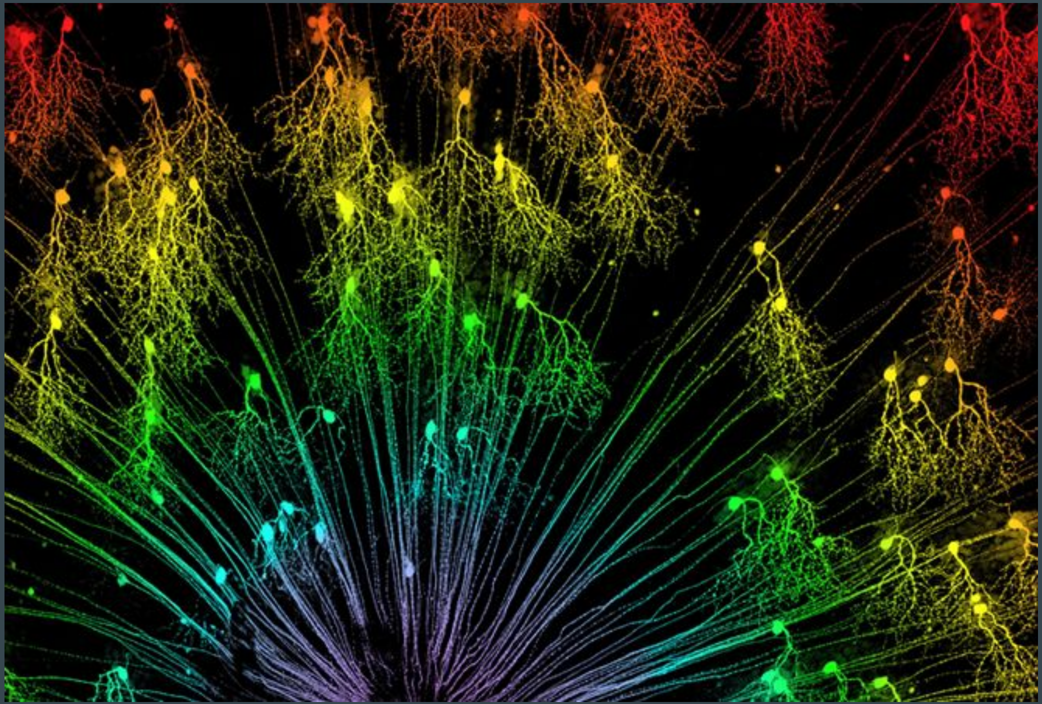
Brainbow,
Josh R. Sanes

Deconstructing a problem

What is the brain? If we zoom in further... we see *neuronal shapes*

How big is each neuron?

Geometry



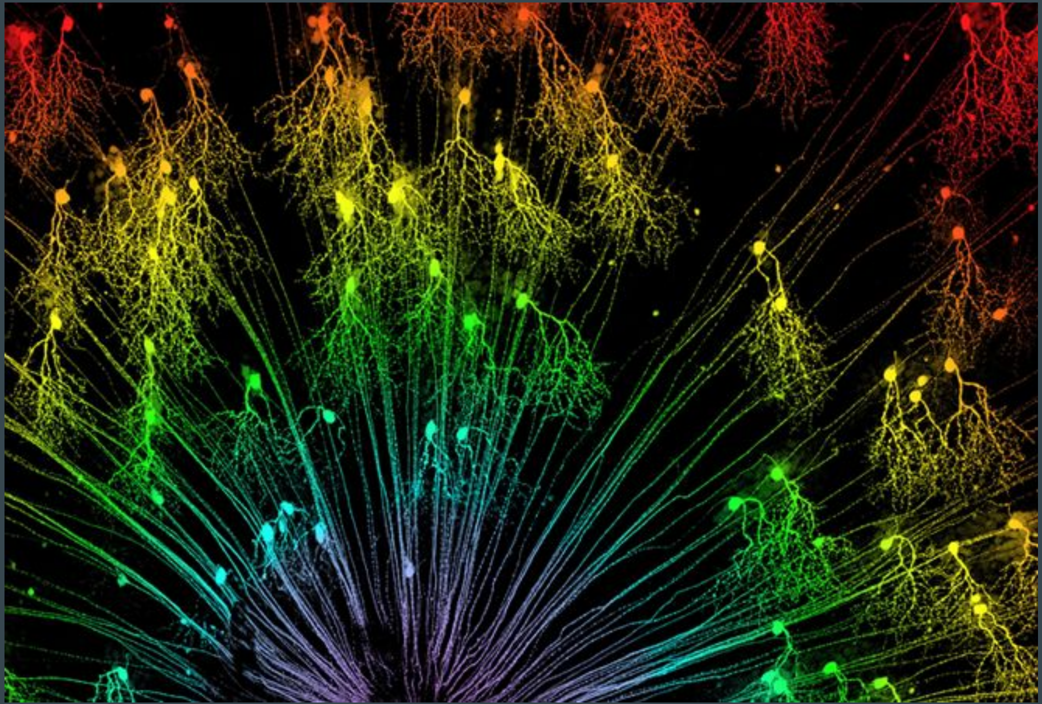
Brainbow,
Josh R. Sanes

Deconstructing a problem

What is the brain? If we zoom in further... we see neuronal shapes

How many branches does each neuron have?

Statistics



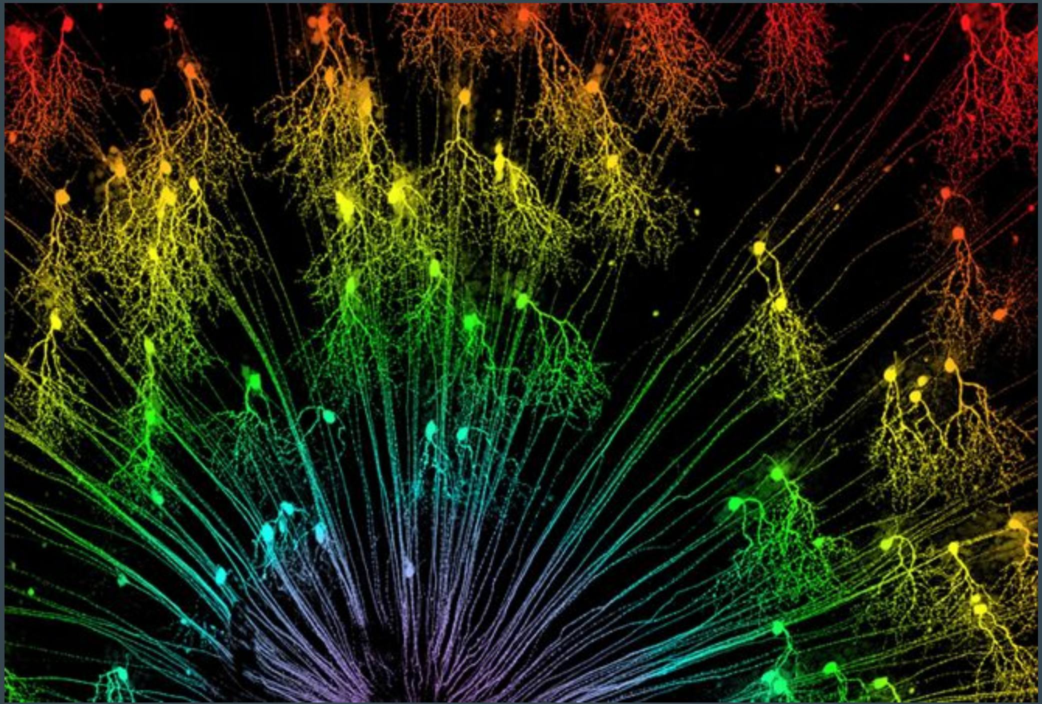
Brainbow,
Josh R. Sanes

Deconstructing a problem

What is the brain? If we zoom in further... we see neuronal shapes

How can we separate the neurons in classes?

Machine learning



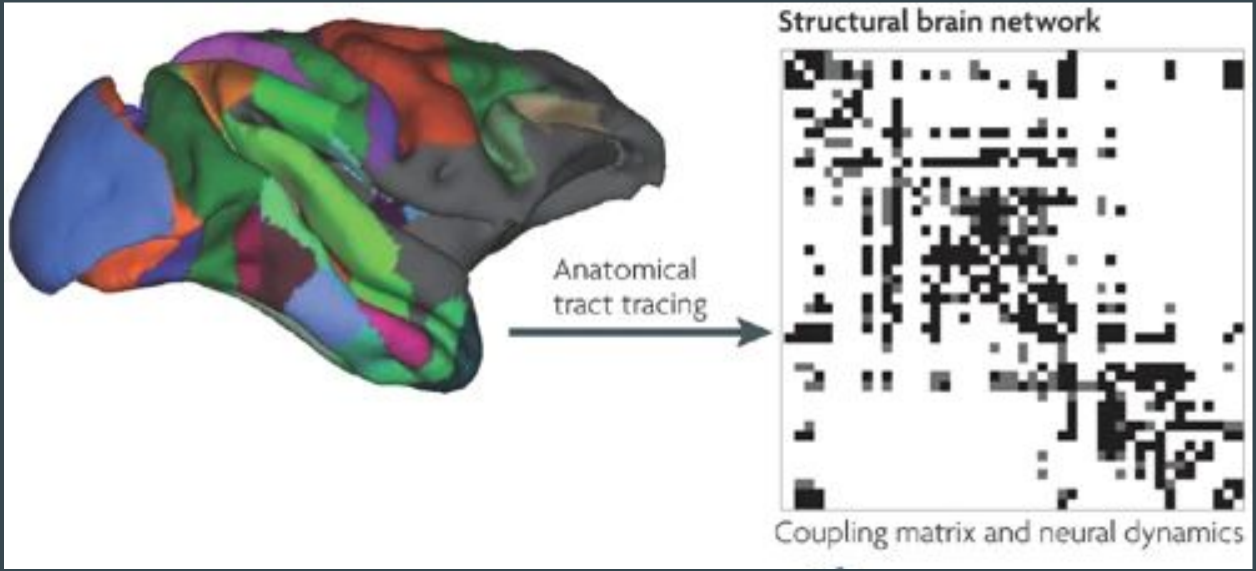
Brainbow,
Josh R. Sanes

Deconstructing a problem

What is the brain? Now we take a step back.... to the *network* level

The connections between brain regions and between single neurons define the brain connectome

Graph theory



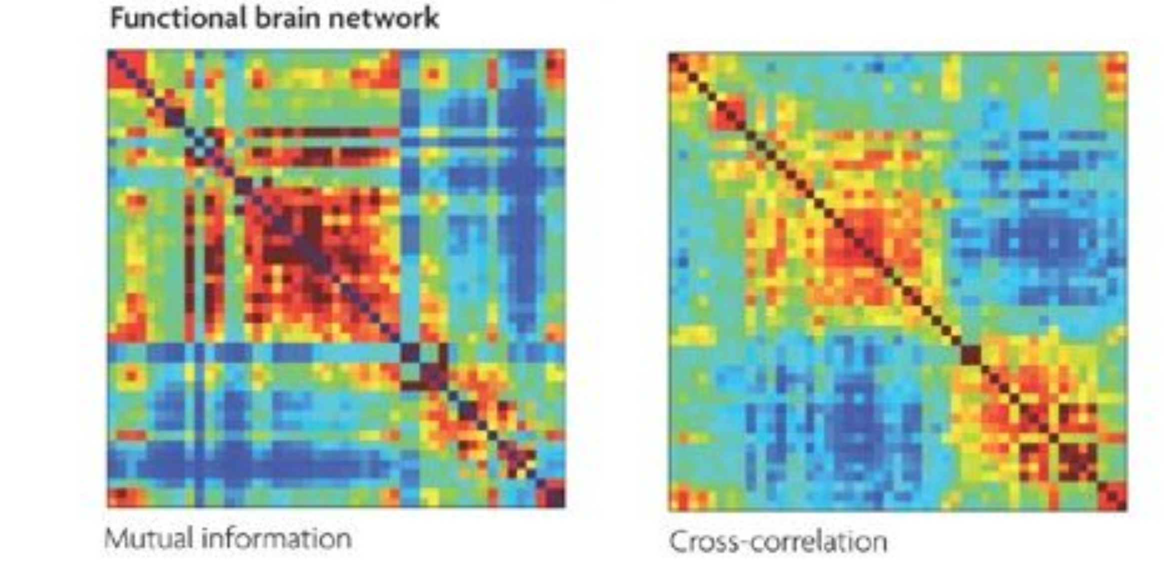
Bullmore & Sporns 2009

Deconstructing a problem

What is the brain? Now we take a step back... to the network level

From this network we can study how the information flows between nodes in the network

Graph theory



Bullmore & Sporns 2009

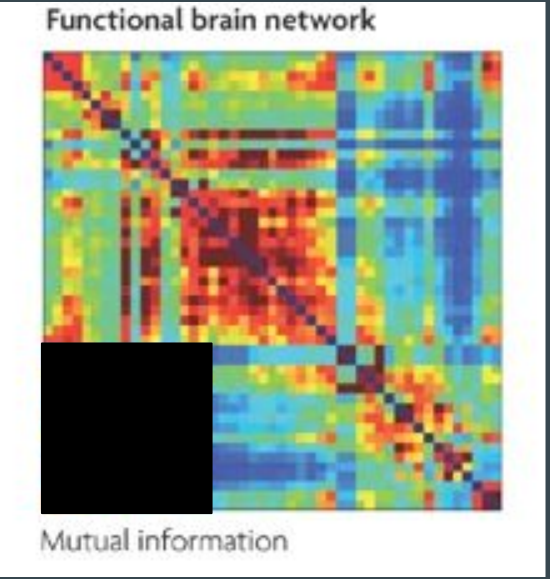
Deconstructing a problem

What is the brain? What happens if something doesn’t work?

But... what happens if the network is not working properly?

For example if one subregion does not receive signals, and the neurons remain silent?

How algebraic topology can be useful?



Modified from
Bullmore & Sporns 2009

Deconstructing a problem

What is the brain? What happens if something doesn't work?

Deconstructing a problem

What is the brain? What happens if something doesn't work?

Christmas lights

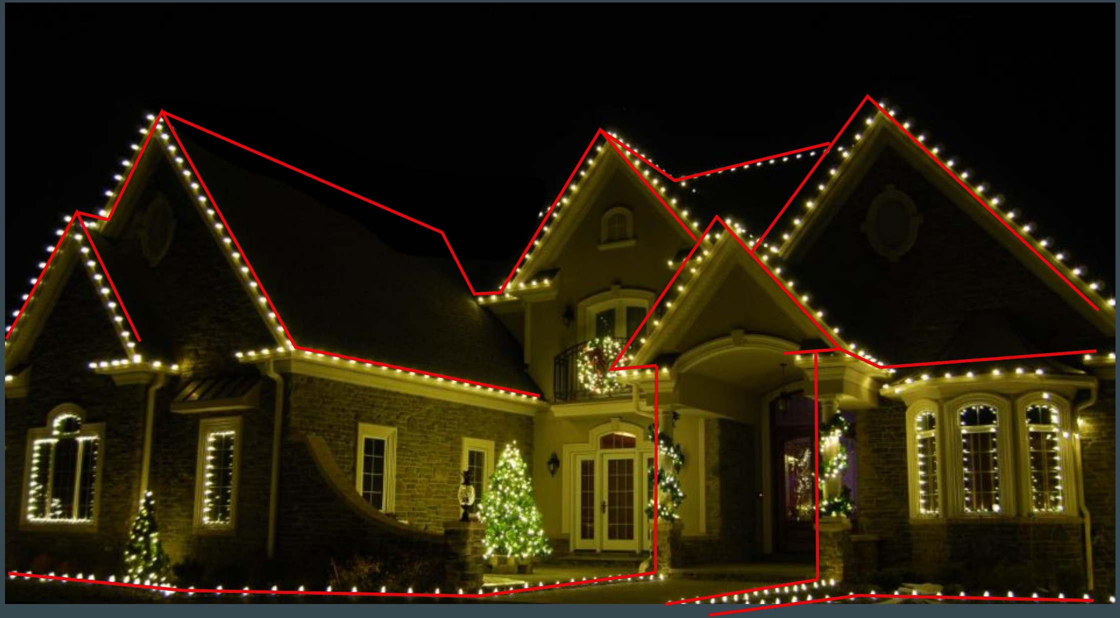
One part of the lights' line breaks down...

Deconstructing a problem

What is the brain? What happens if something doesn't work?

We need to see this as a graph, or a collection of points and lines

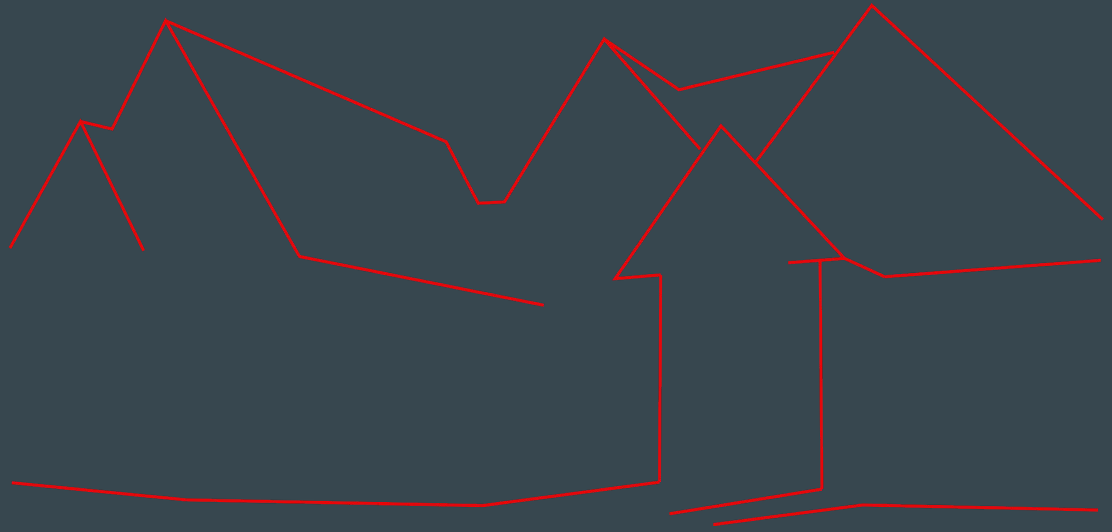
And now we can use algebraic topology



Deconstructing a problem

What is the brain? What happens if something doesn't work?

Then these lines need to be translated into a meaningful skeleton



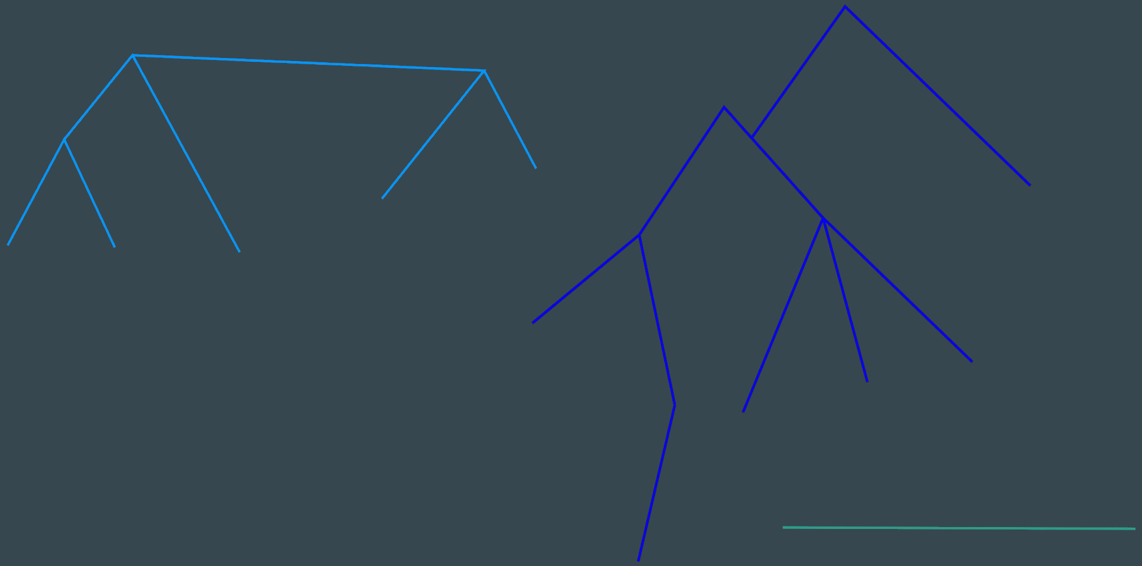
Deconstructing a problem

What is the brain? What happens if something doesn't work?

Then these lines need to be translated into a meaningful skeleton

That will allow us to detect:

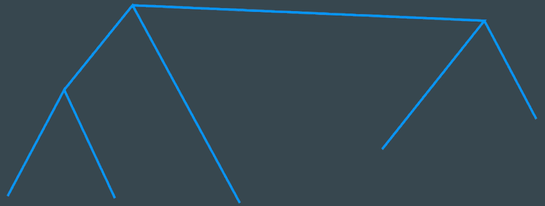
Continuous shapes: lines
Connected components: simplices
Gaps: holes

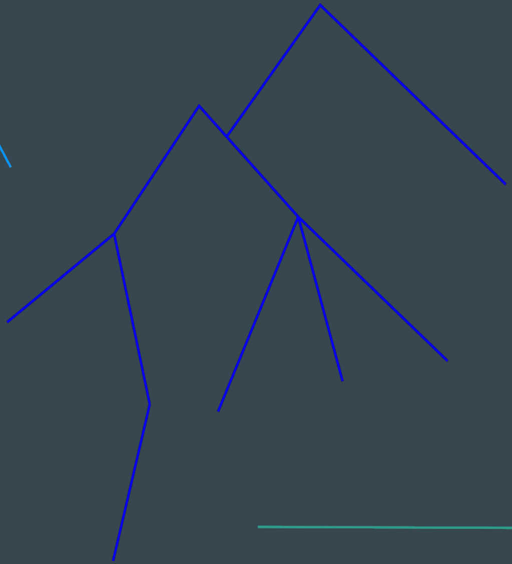


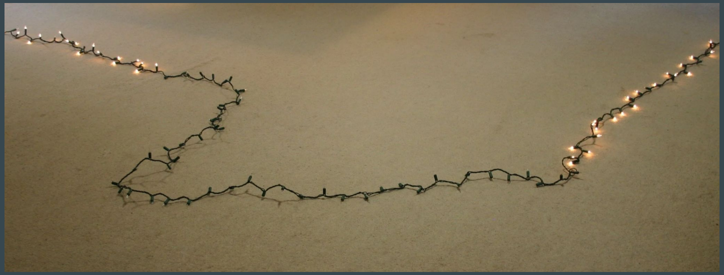
Deconstructing a problem

What is the brain? What happens if something doesn’t work?

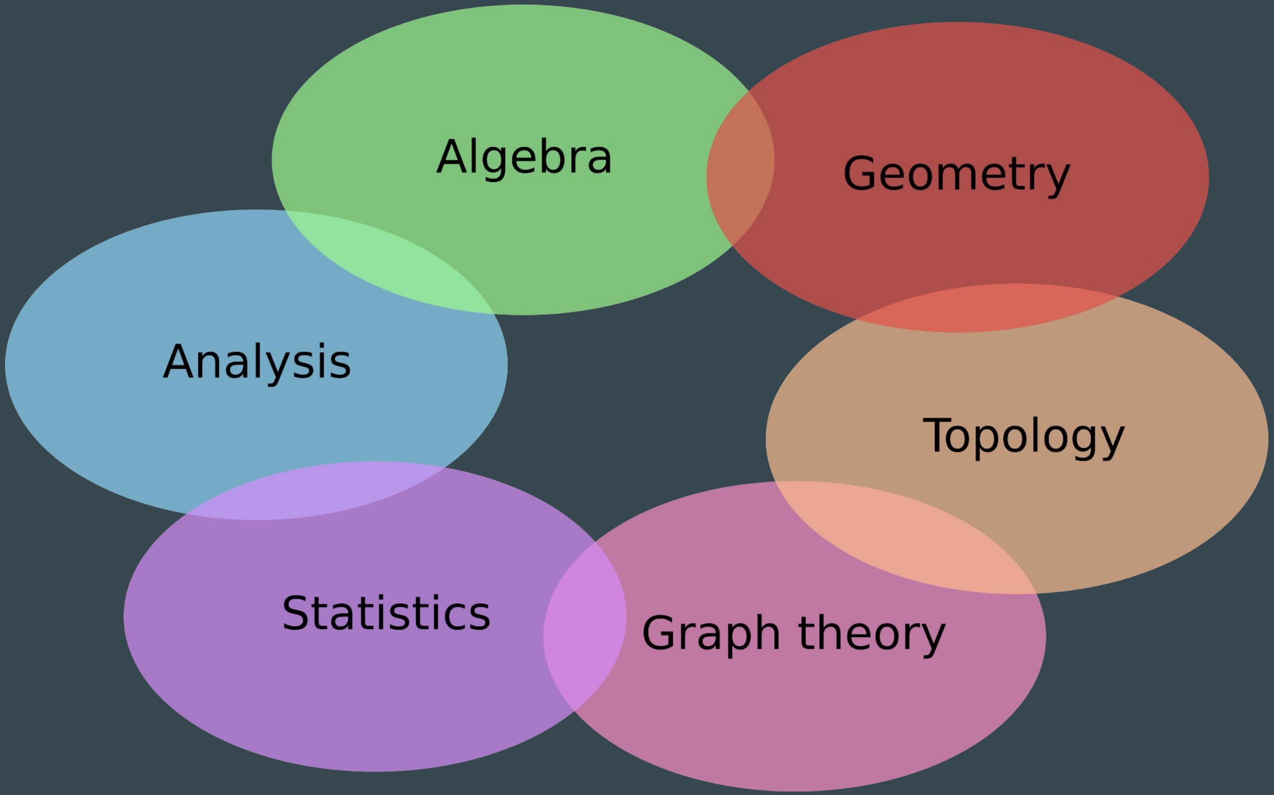
And identify where the system breaks, and what can be fixed in order to close the circuit







Mathematical tools



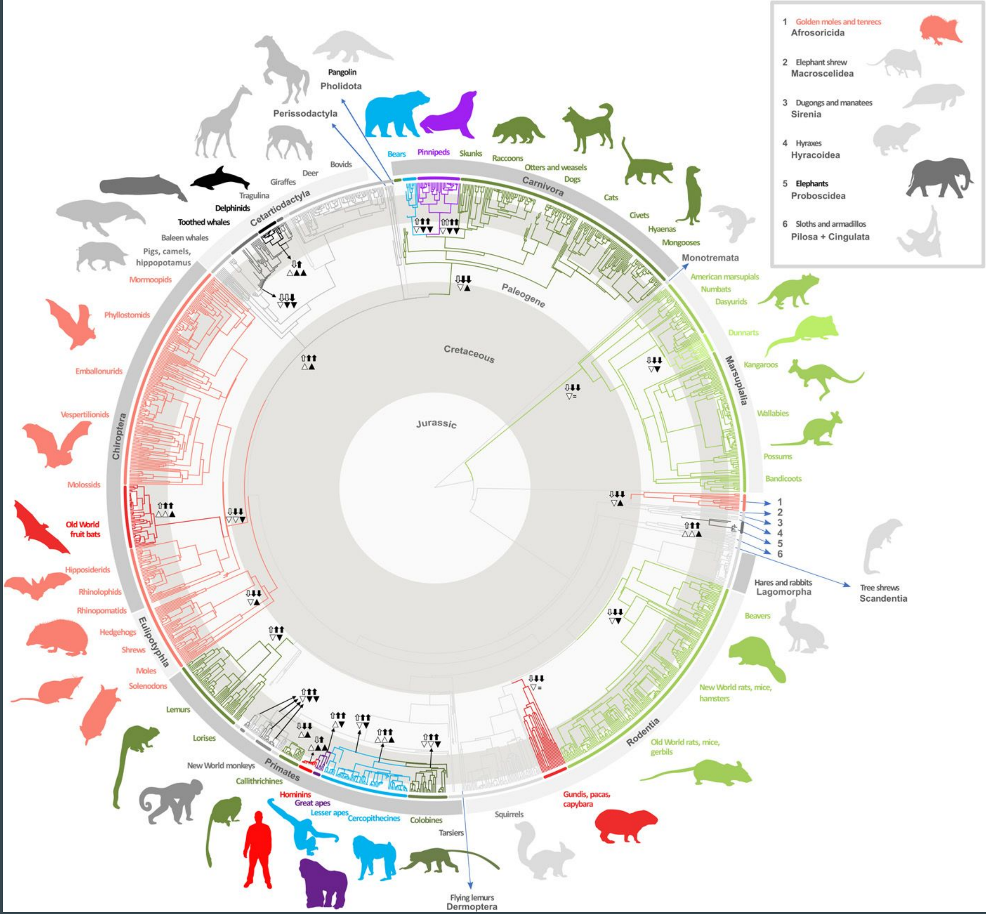
I. Geometry

What is geometry?

Geometry (from Ancient Greek measuring the earth) is studying the properties of space such as the distance, shape, size...

Why is geometry useful?

Geometry in neuroscience studies the shape and size of the brain, as well as its underlying structures

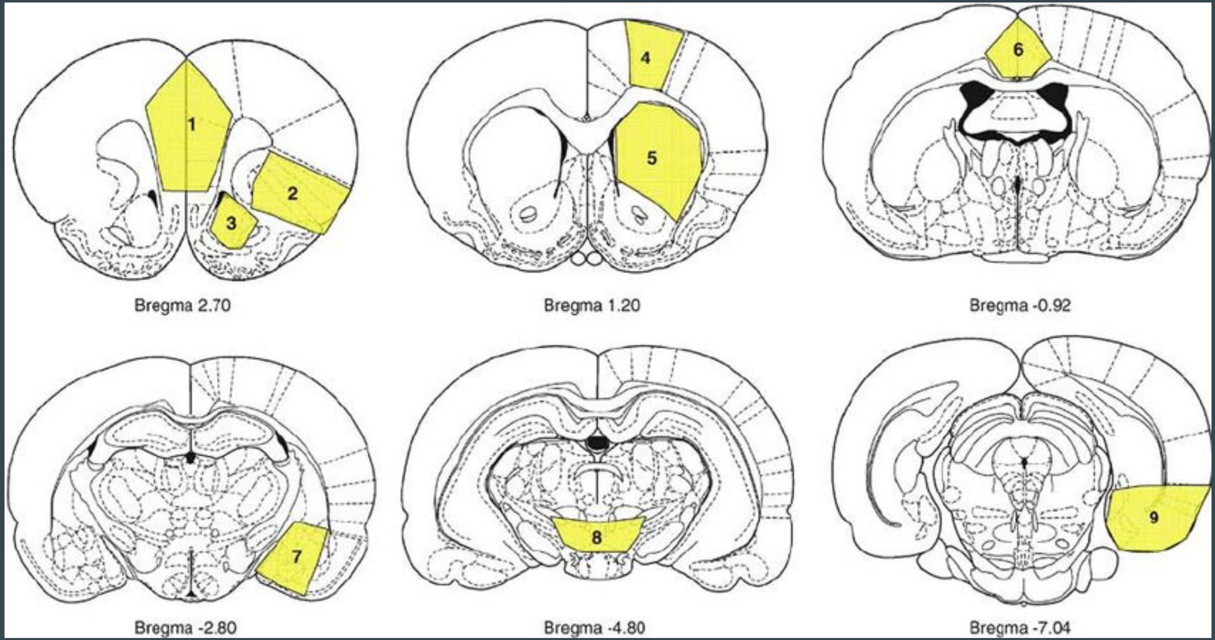


Allometric study of the brain in an effort to explain brain size

Smaers et al. 2021

Why is geometry useful?

Geometry in neuroscience studies the shape and size of the brain, as well as its underlying structures



Geometry is useful to study brain atlases, i.e. the decomposition in brain regions

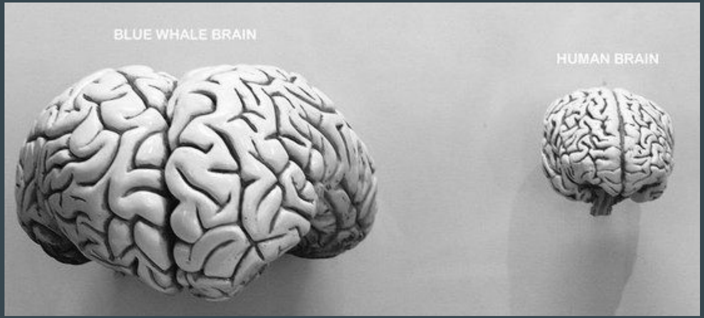
Coronal rat brain atlas diagrams from Paxinos and Watson (1998)

Why is geometry useful?

An interesting example of geometry in neuroscience is the attempt to understand what is special about the human brain.

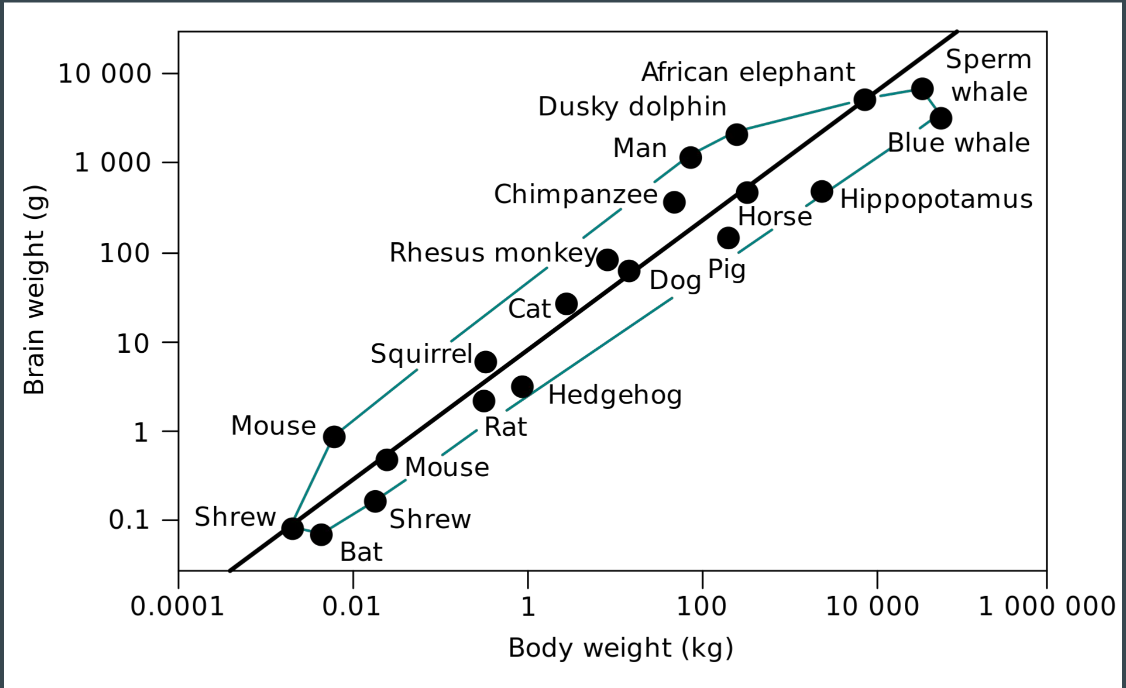
Trying to find what characteristic is essential for intelligence it was initially thought that brain size should be an important indicator

Indeed human brain size is "big" but compared to other animals such as whale or elephant it is by far not the largest animal brain.



Why is geometry useful?

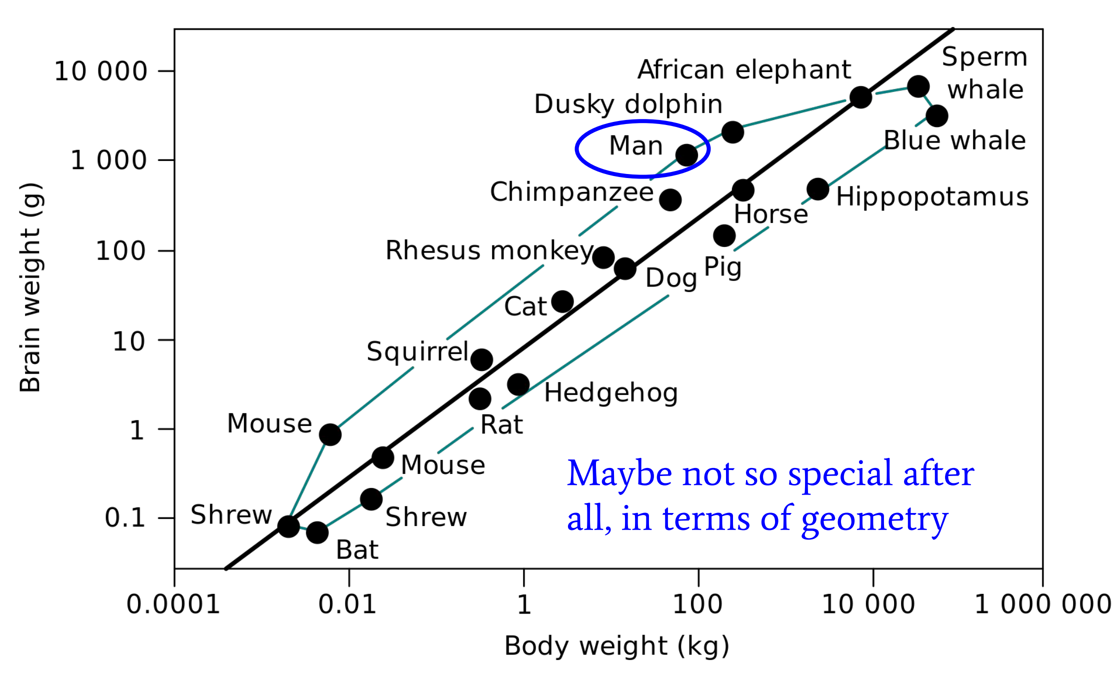
Maybe it's the relative size of the brain in comparison to body mass.



Von Haller's law, source: wikipedia

Why is geometry useful?

Maybe it's the relative size of the brain in comparison to body mass.

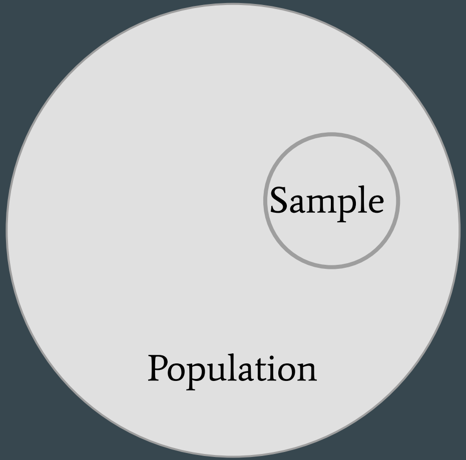


Von Haller's law, source: wikipedia

II. Statistics

What is statistics?

Statistics is the mathematical discipline that studies how collected data can be analyzed rigorously, to prove or disprove a hypothesis. This requires the extrapolation (inference) of conclusions on a larger population, while data are only available on a subset (observations) of this population. The “validity” of these conclusions is formally judged by certain statistical criteria. It is clear that this endeavor rests in large part on probability theory which is essential to all statistics.

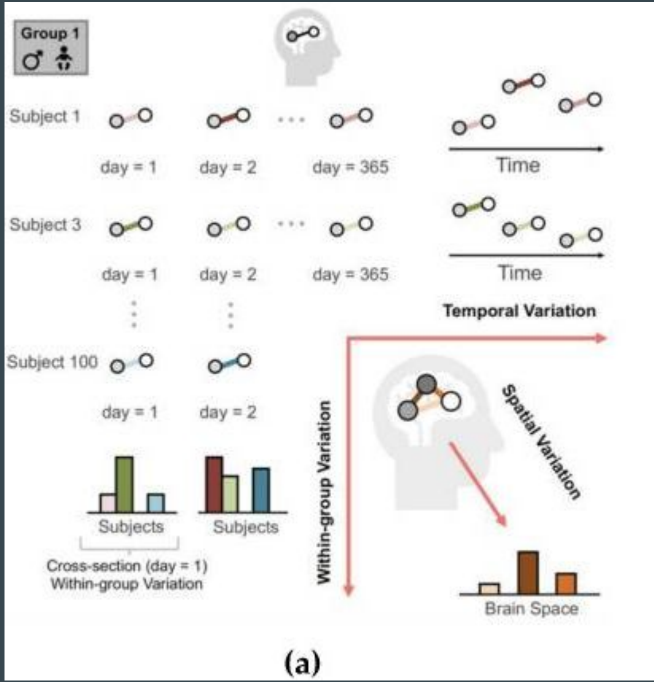


Why is statistics useful?

Statistics is used in every study either directly, or indirectly, so it’s essential to know how to correctly use statistical tools and which models are appropriate depending on the use case.

Why is statistics useful?

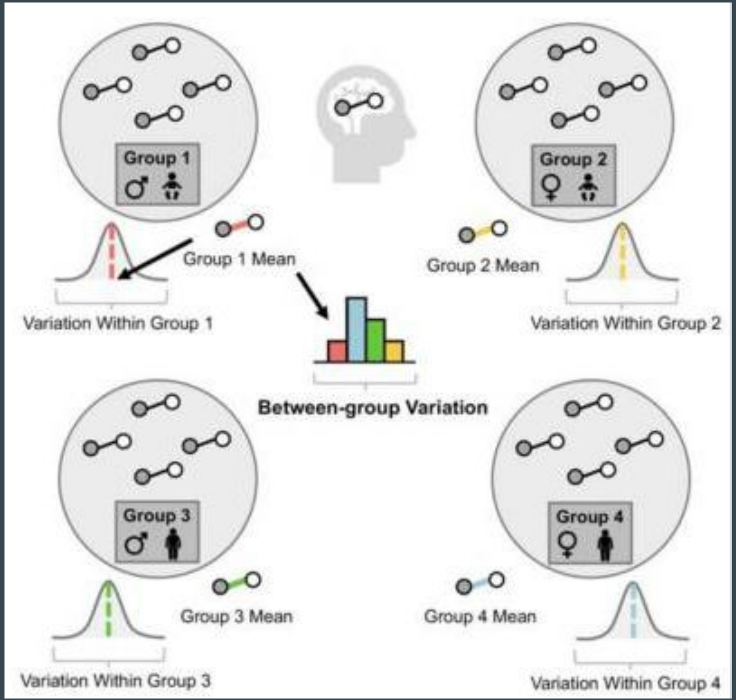
Examples of statistical operations:



Collection of scientific data,
Chen, 2019

Why is statistics useful?

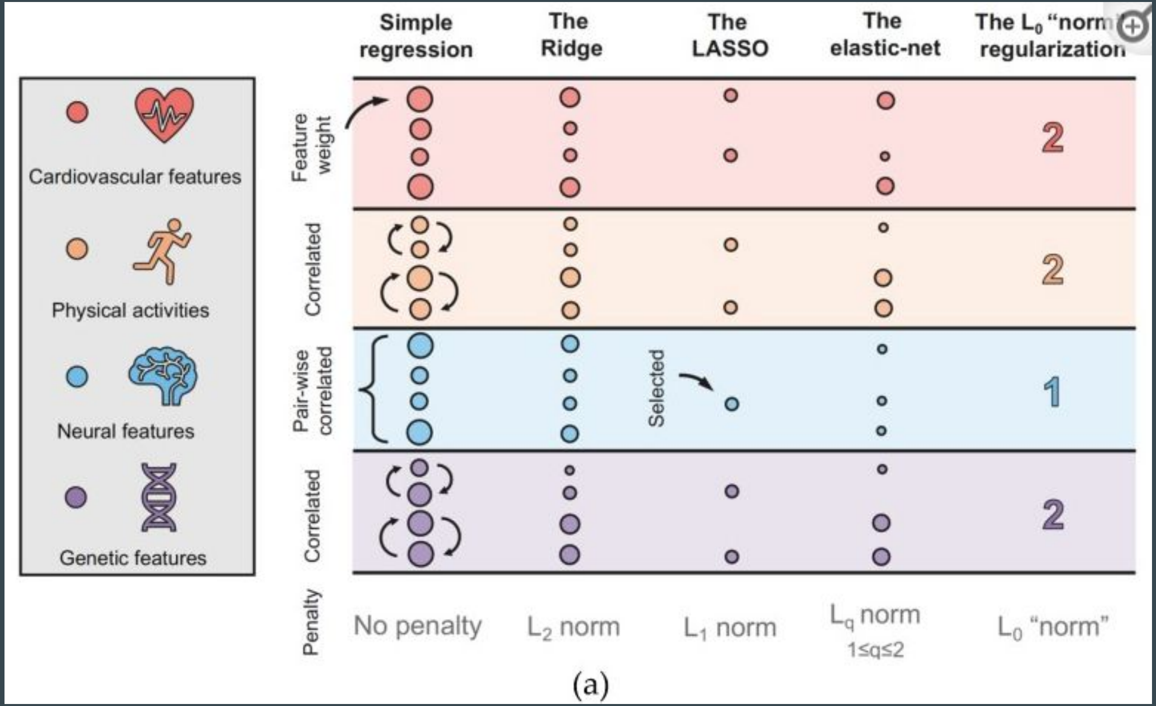
Examples of statistical operations:



Statistical analysis of data, Chen, 2019

Why is statistics useful?

Examples of more statistical operations:



Chen, 2019

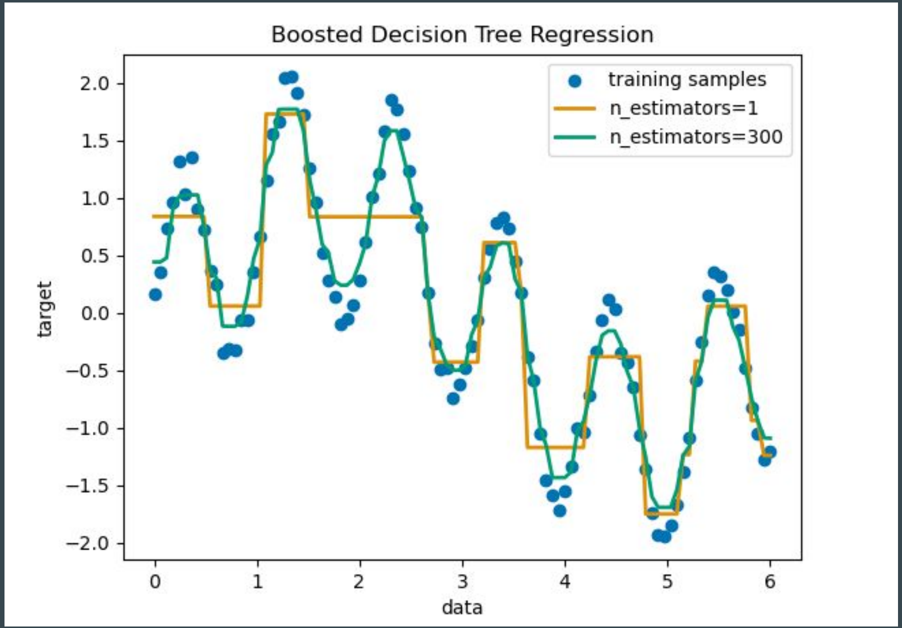
III. Machine Learning

What is machine learning?

Machine learning is more of a computational tool, rather than a mathematical method. However, underneath the computational implementation ML requires advanced mathematical computations. Better understanding some of the concepts underlying ML algorithms will be helpful to decide what tools are appropriate to use in each problem.

Why is machine learning useful?

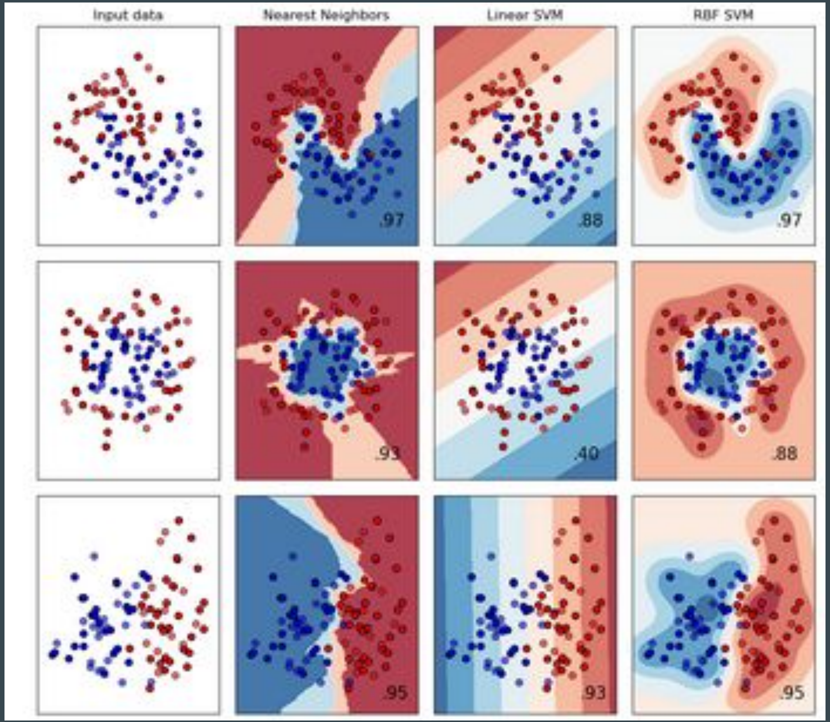
Regression: a way to approximate datasets and predict future points in a sequence:



Sklearn library

Why is machine learning useful?

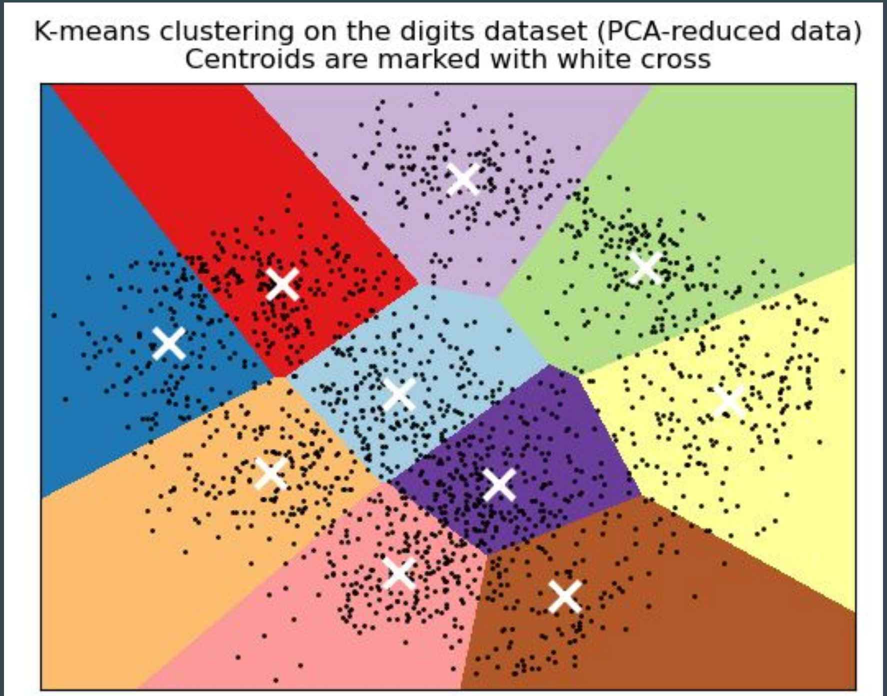
Classification: a way to classify data into known groups and understand differences:



Sklearn library

Why is machine learning useful?

Clustering: a way to cluster data into groups without prior knowlegde:



Sklearn library

IV. Algebraic Topology

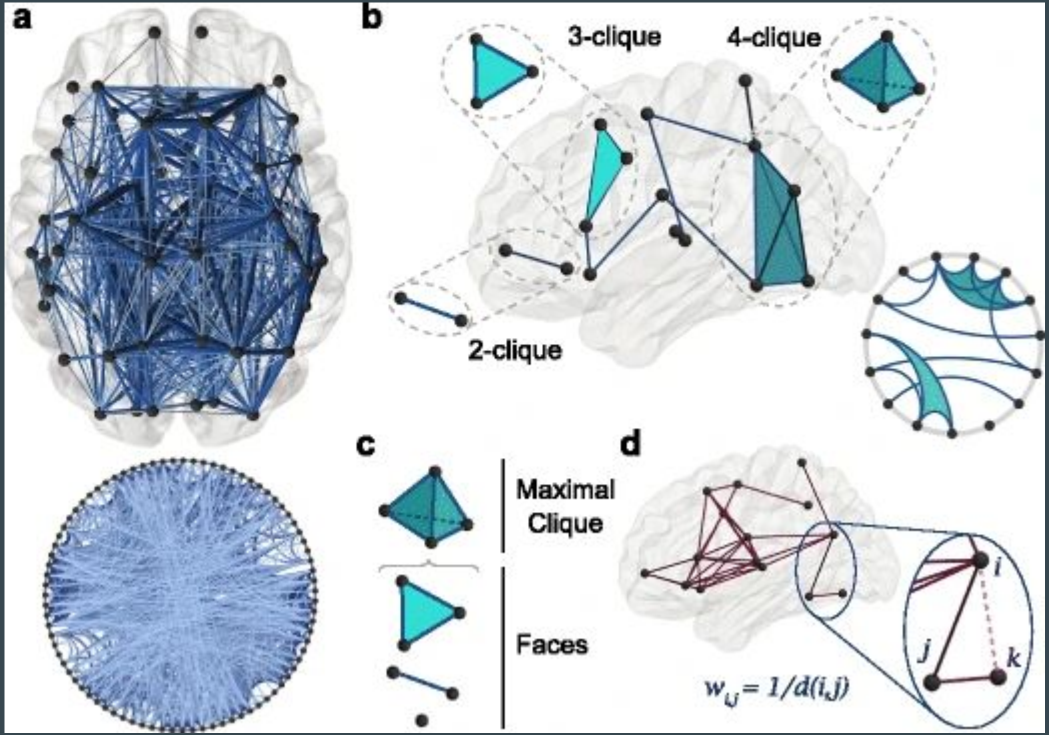
What is algebraic topology?

Algebraic topology studies the topological properties of spaces and objects, using tools from algebra. One of the main objectives is to find algebraic invariants in order to classify topological spaces.

Topological Data Analysis (TDA) is a mathematical field that combines the techniques of topology and geometry with the analysis of large, complex data sets. It aims to study the shape and structure of data by treating it as a topological space and using algebraic and geometric methods to extract meaningful information and identify patterns and relationships within the data.

Why is algebraic topology useful?

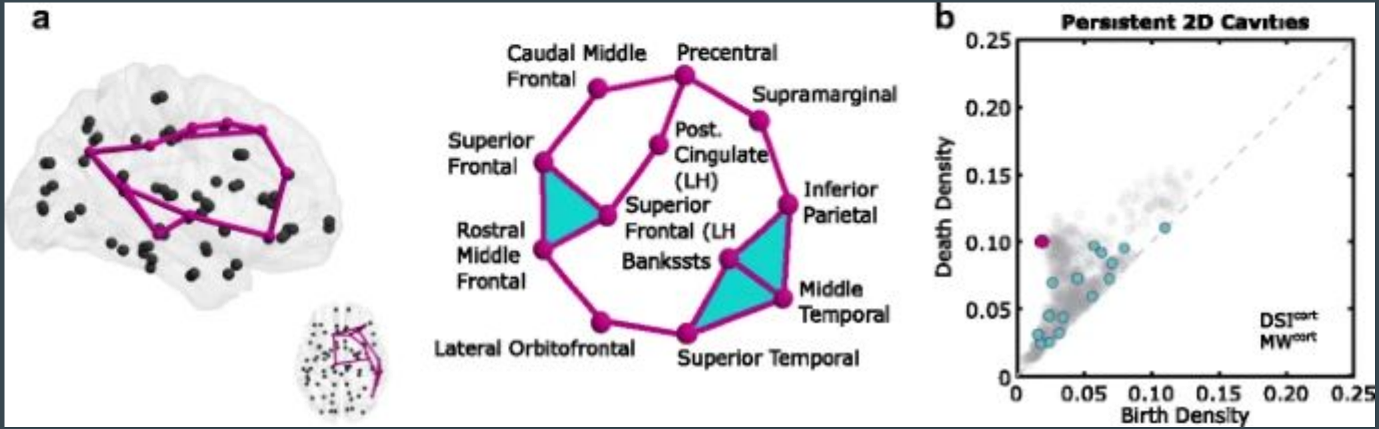
Seen as a network of points and connections, TDA is an excellent tool to study the brain connectivity.



Sizemore et al. 2017

Why is algebraic topology useful?

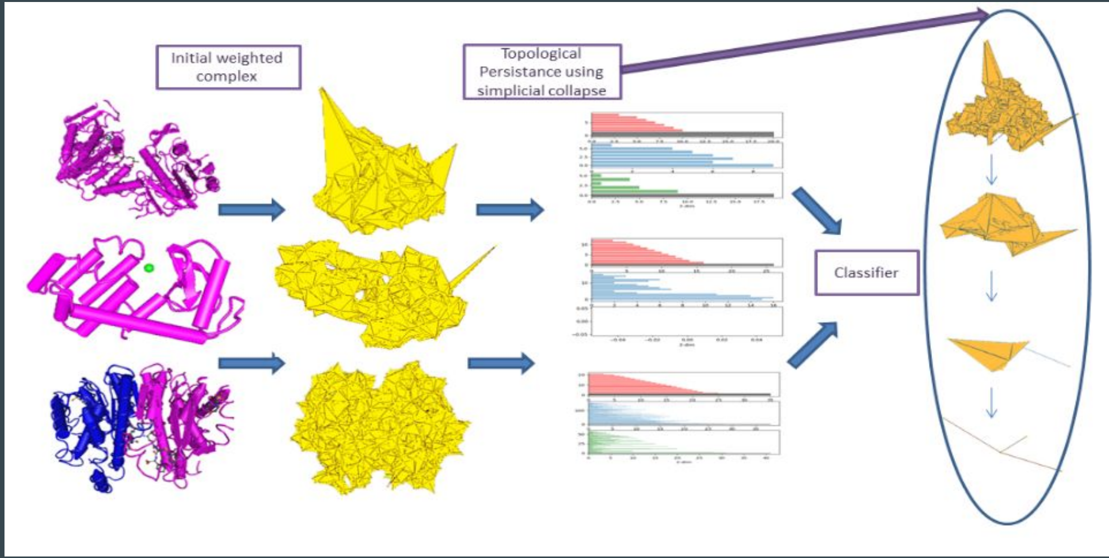
TDA is a powerful tool to simplify complex networks and reveal the underlying structure. TDA is useful to study brain networks, identify important nodes and discover abnormalities in brain connectivity.



Sizemore et al. 2017

Why is algebraic topology useful?

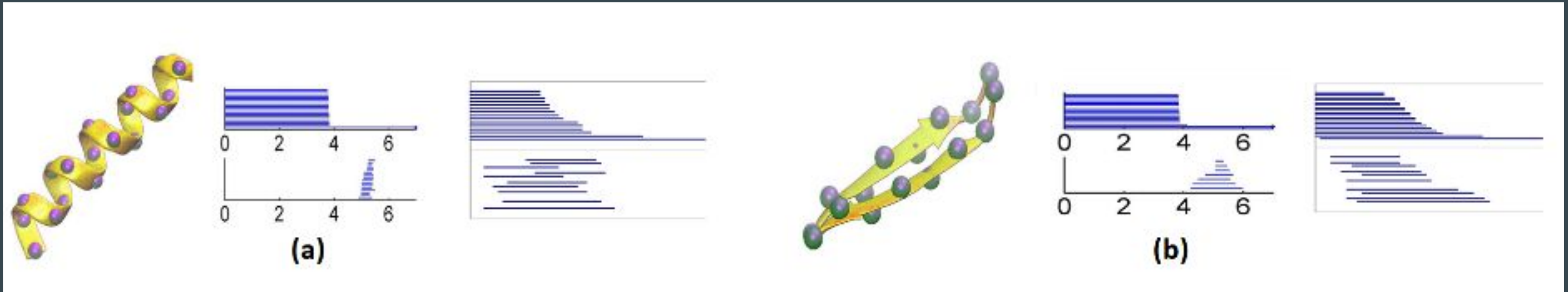
TDA is useful to extract topological signatures from protein structure and can be used for the automatic annotation and classification of proteins.



Day and Mandal 2012

Why is algebraic topology useful?

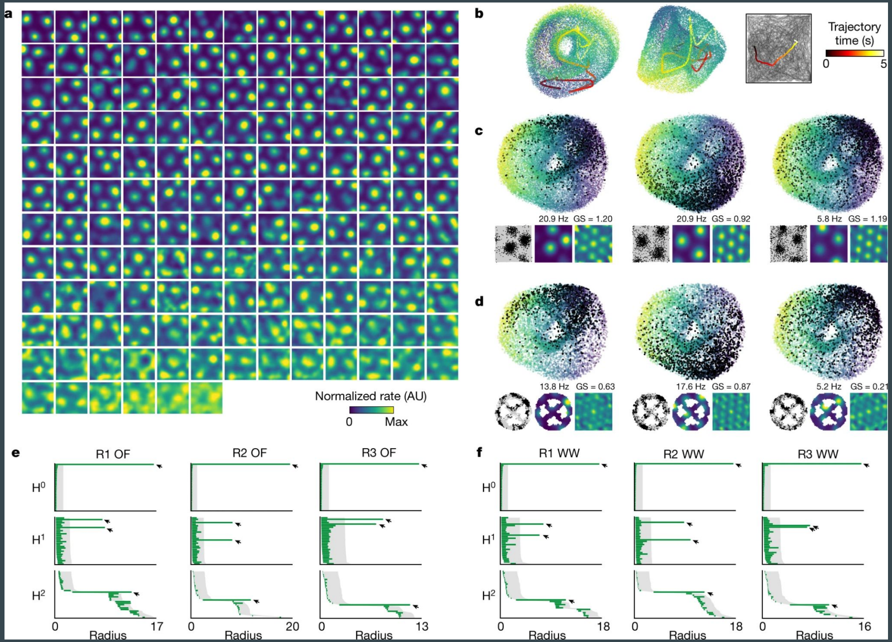
TDA is useful to extract topological signatures from protein structure and can be used for the automatic annotation and classification of proteins.



Day and Mandal 2012

Why is algebraic topology useful?

TDA is useful to study the topology of place cells in hippocampus, and identify subtypes based on the topological properties of firing during specific trajectories



Gardner et al. 2022

VI. Stochastic processes

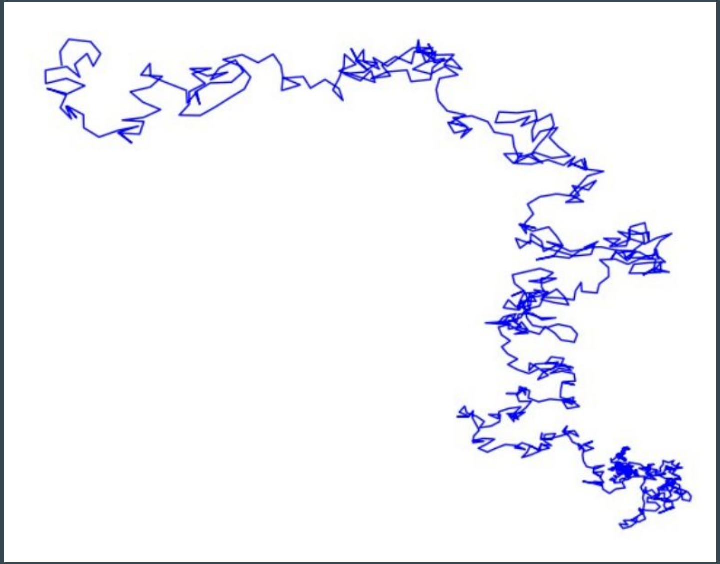
What is a stochastic process ?

Stochastic processes are mathematical tools to describe the evolution of a system over time, under conditions of uncertainty or randomness. They are used to model various types of random phenomena, including time series data, financial markets, communication networks and biological systems.

In a stochastic process, the state of the system is described by a random variable that changes over time, and the future state of the system is determined probabilistically, based on the current state and the underlying probability distributions.

What is a stochastic process ?

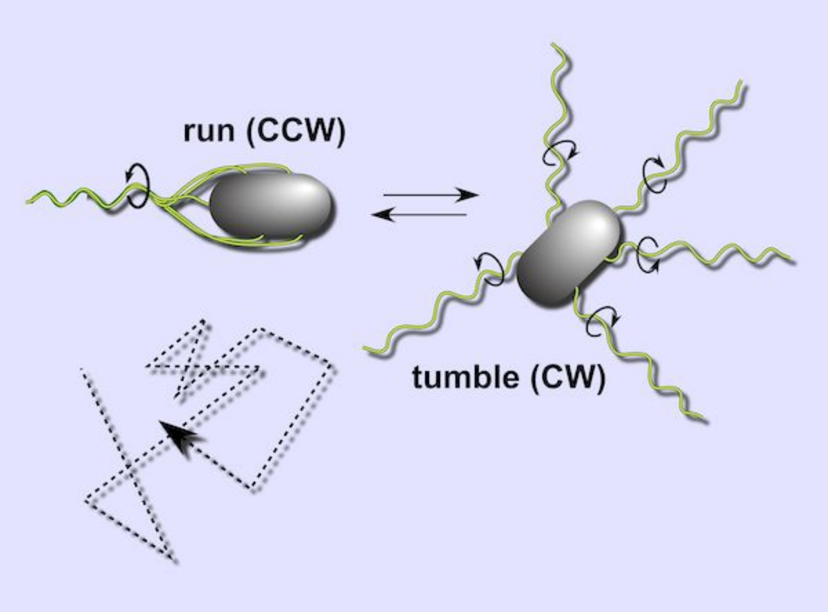
We will explore stochastic processes using the model of random walks. Starting from an initial position x0, we move on the lattice (up, down, left, right) with predefined probabilities (p_u, p_d, p_l, p_r)



Why are stochastic processes useful?

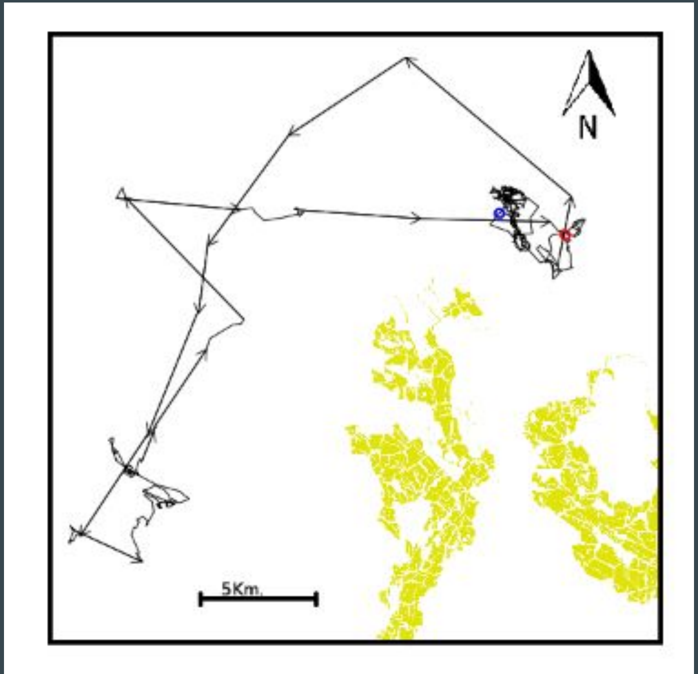
Random walks are used to model cell and animal movement:

Ecoli movement, Sandy Parkinson



Why are stochastic processes useful?

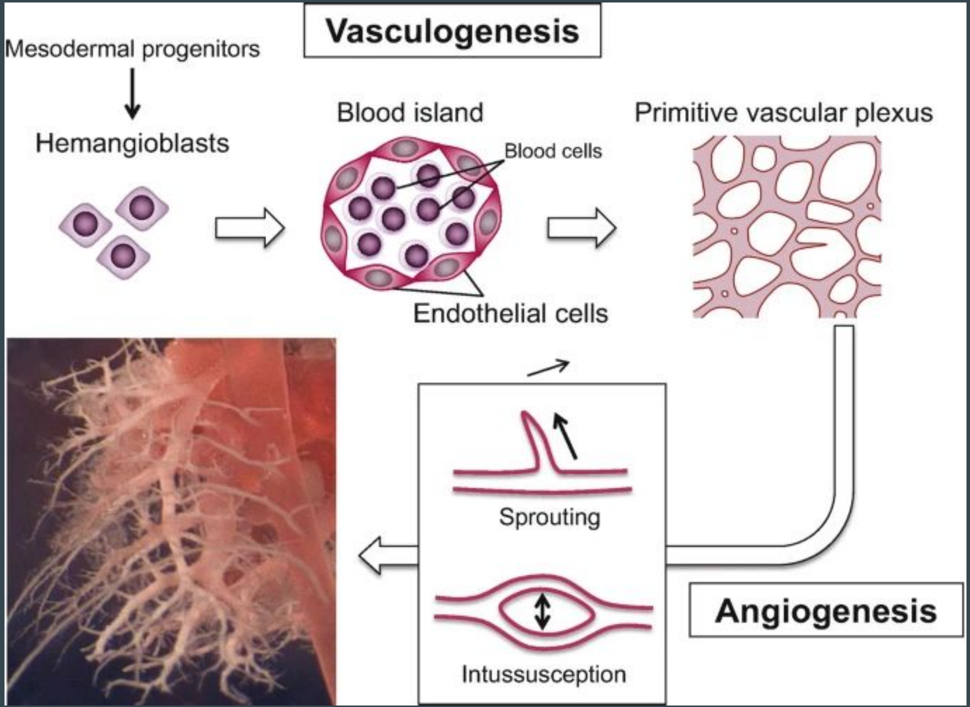
Random walks are used to model cell and animal movement:



Caribou movement, Nicosia et al. 2017

Why are stochastic processes useful?

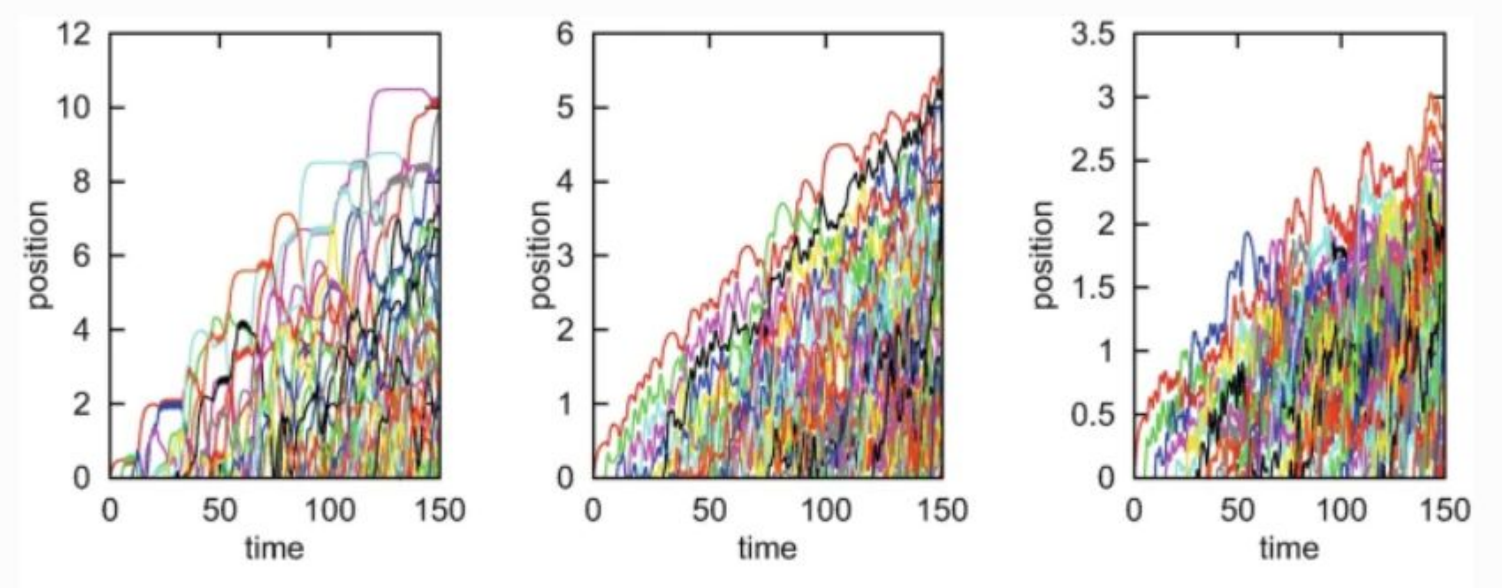
Another example of random walks modeling biological processes is angiogenesis



Kurihara et al. 2022

Why are stochastic processes useful?

Another example of random walks modeling biological processes is angiogenesis



Simulating trajectories of movement, Kurihara et al. 2022

Why are stochastic processes useful?

In neuroscience random walks were initially used to model spike activity of a neuron

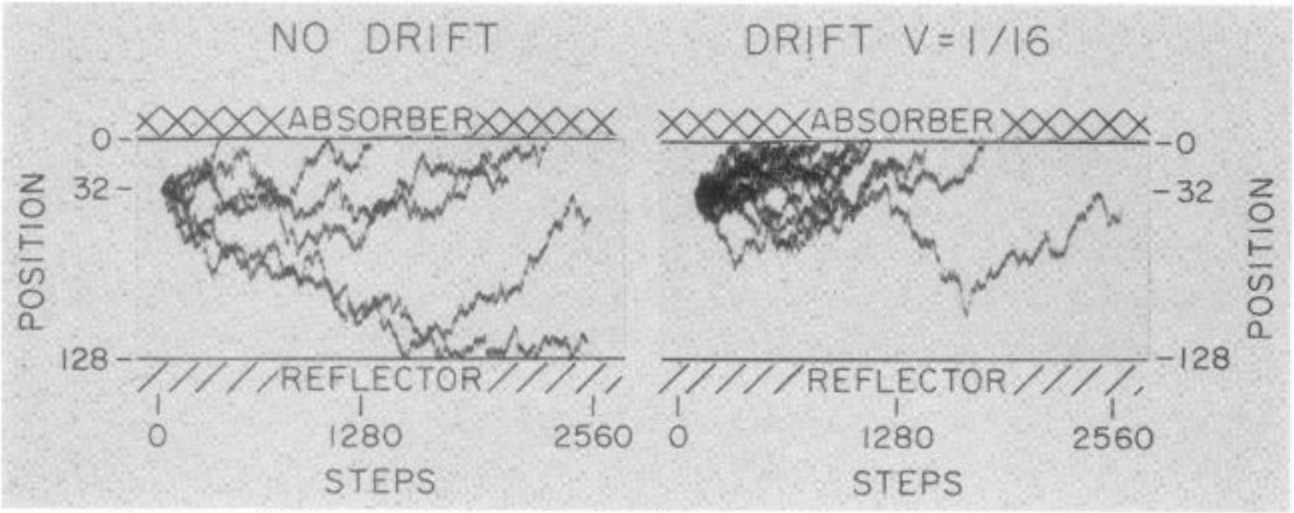
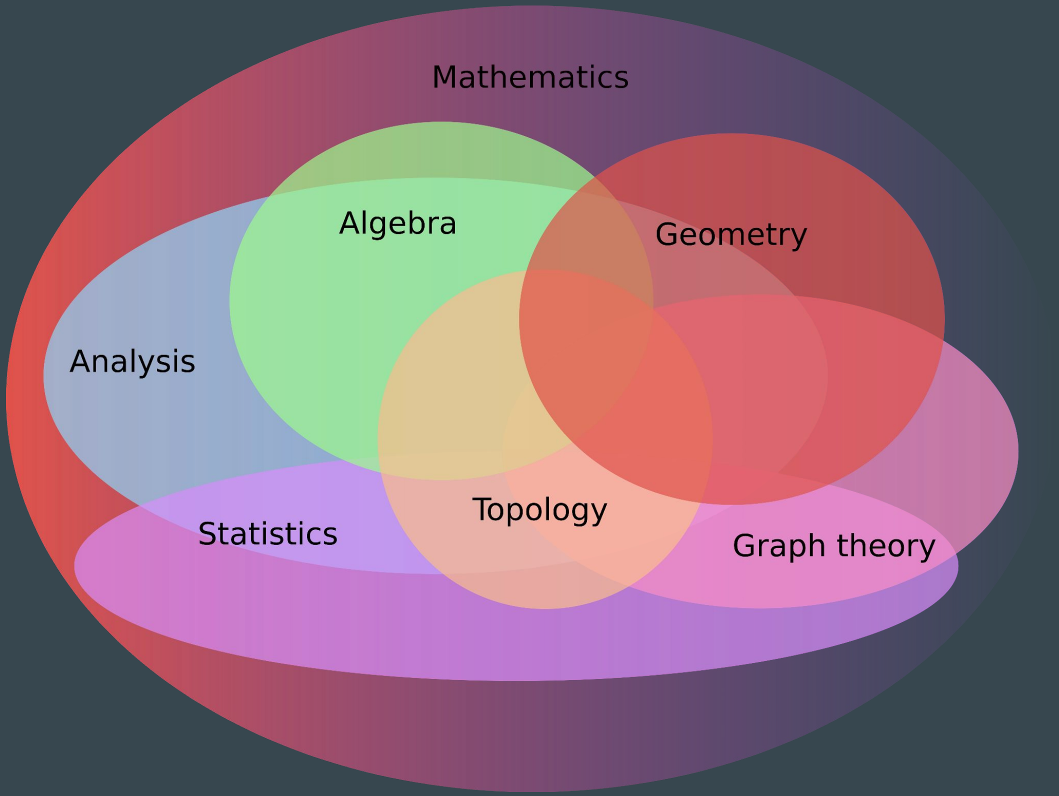


Figure 3 Typical random walks in one dimension in a computer simulation of the model.

Gerstein and Mandelbrot 1964

Mathematical tools



Data collection

Introduction to “mathematical thinking”

Examples: Data collection

Problem: I have collected a set of data...

	A	B	C	D	E
1	property|name	axon|mean_number_of_bifurcatio	axon|mean_number_of_leav	axon|mean_number_of_sectio	axon|mean_remote_bifurcation_ang
2	C050800E2_cor.h5	59	60	119	1.2029861559104
3	C120398A-P2.h5	25	26	51	1.31538621816681
4	C120398A-P3.h5	40	41	81	1.42040152246122
5	C271097A-P1.h5	37	38	75	1.17331024837382
6	C271097A-P2.h5	21	22	43	1.18553678311865
7	C271097A-P3.h5	34	35	69	1.30503361454563
8	rp110111_L5-2_idH.h5	61	62	123	1.39761281141949
9	rp120608_P_3_idC.h5	44	47	92	1.18236947545108
10	rp120608_P_3_idD.h5	69	70	139	1.08486835209442
11	sm100429a1-5_INT_idE.h	38	39	77	1.4799778645049
12	sm100429a1-5_INT_idG.h	28	29	57	1.73913560953665
13	vd101102b_INT_idA.h5	43	44	87	1.45461104222372
14	C060998B-P4.h5	10	11	21	1.47125481505973
15	C120398A-P1.h5	10	11	21	1.6064295953642
16	C140600C-P3.h5	12	13	25	1.11694986036925
17	C150897B-P2.h5	32	33	65	1.21858139632941
18	C200897C-P2.h5	19	20	39	1.70497037908832
19	C200897C-P4.h5	41	42	83	1.37332234508014
20	C220498B-P3_cor.h5	25	26	51	1.51601242237315
21	C231296A-P4B2.h5	16	17	33	1.5199786142752
22	C260199A-P2.h5	4	5	9	1.47696729610387
23	C300797C-P2.h5	17	18	35	1.52795803970122
24	C300797C-P4.h5	10	11	21	1.5769699580797
25	C310897A-P4.h5	69	70	139	1.24321043670234
26	C310897B-P3.h5	29	30	59	1.31227750969831
27	C310897B-P4.h5	8	9	17	1.25173531651878
28	Fluo9_left.h5	28	29	57	1.30945300169143

Examples: data collection

What I need in order to understand my data? We first need to ask the *right* questions

Establish a hypothesis: i.e.

- Do I want to approximate a distribution?
- Do I want to identify subtypes?
- Do I want to find correlations between measurements?
- Do I want to find averages for specific measurements?

The question will hint to which mathematical tools are appropriate to use!

Examples: data collection

Question: which mathematical tools can I use?

Statistics

???

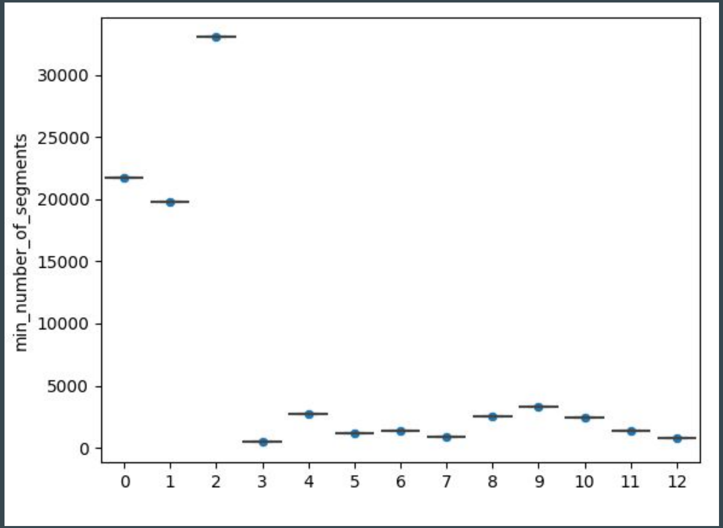
Machine learning

Graph theory

Algebraic topology

Data collection

Question: what do my data represent? First I need to visualize the dataset

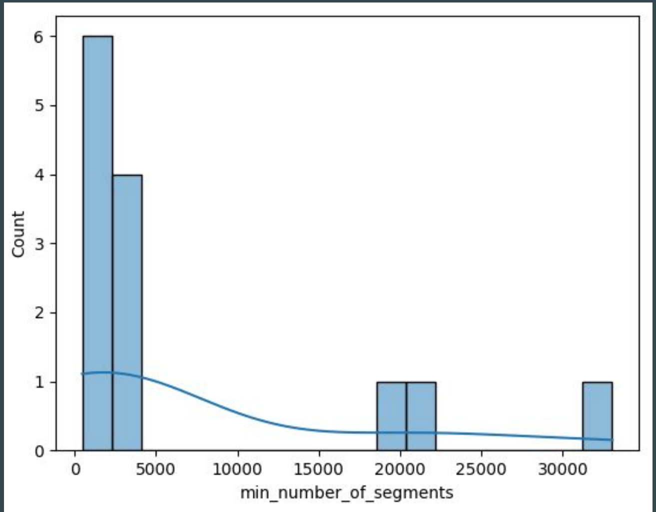


As a set of independent points

Data collection

Question: what do my data represent? First I need to visualize the dataset

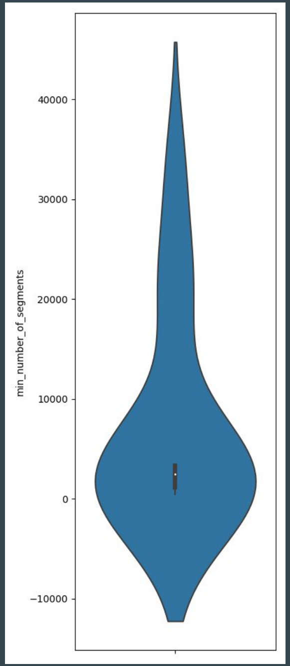
As a set of points from the same distribution



Data collection

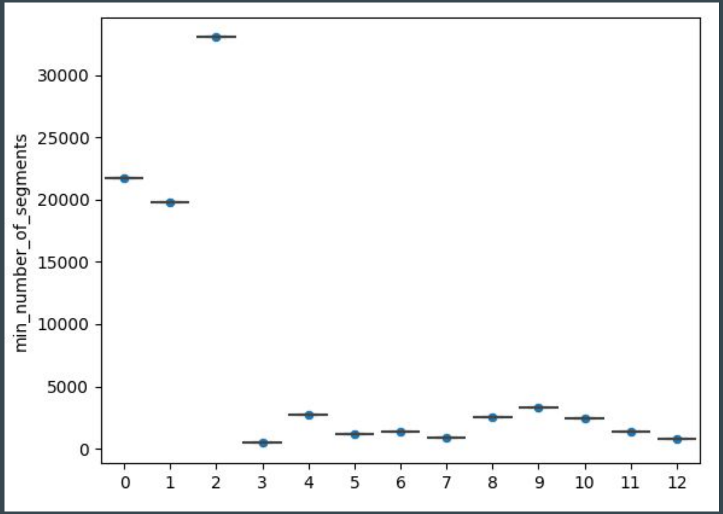
Question: what do my data represent? First I need to visualize the dataset

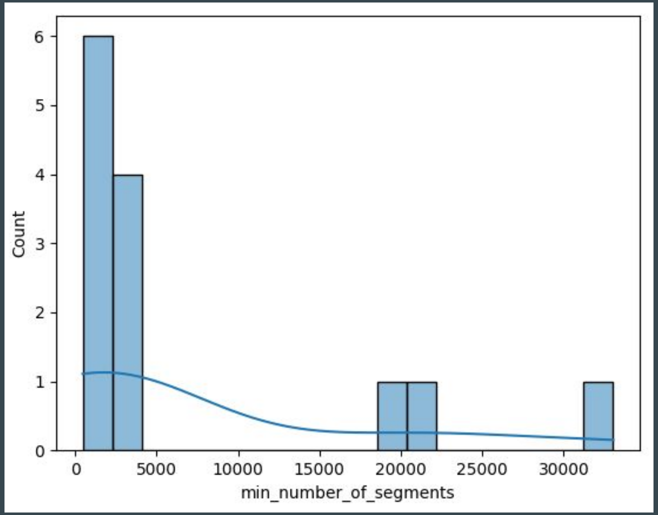
As a collection of points, in a violin plot

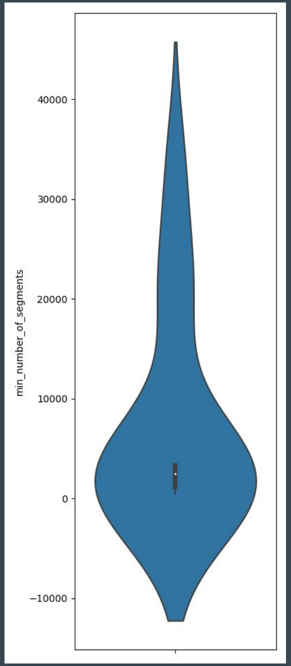


Data collection

As a single population of data points...

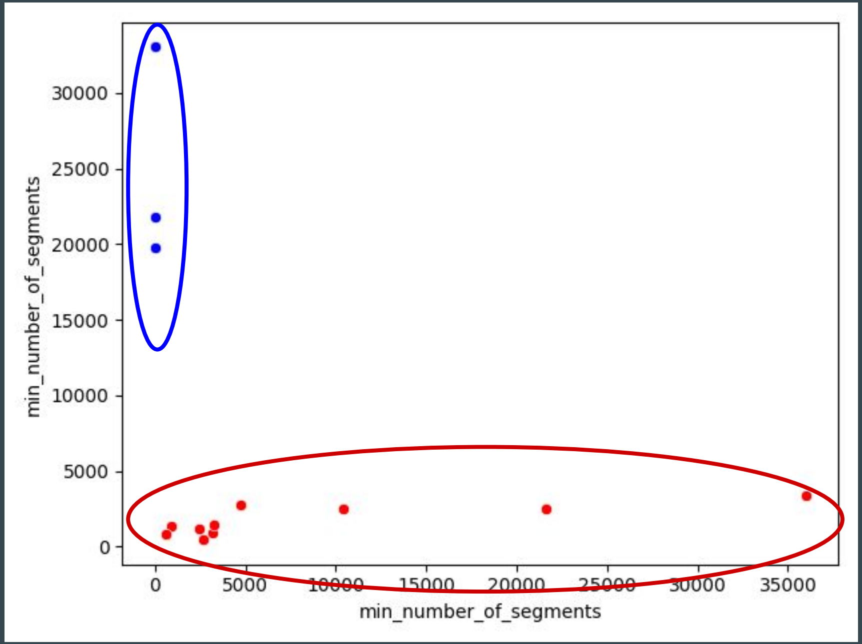






Data collection

Question: what do my data represent? First I need to visualize the dataset

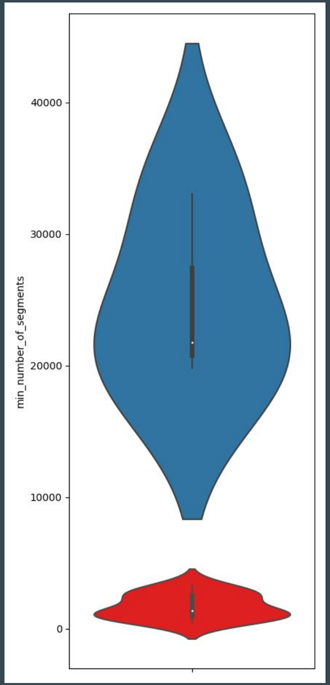


Or as two sets of points

Data collection

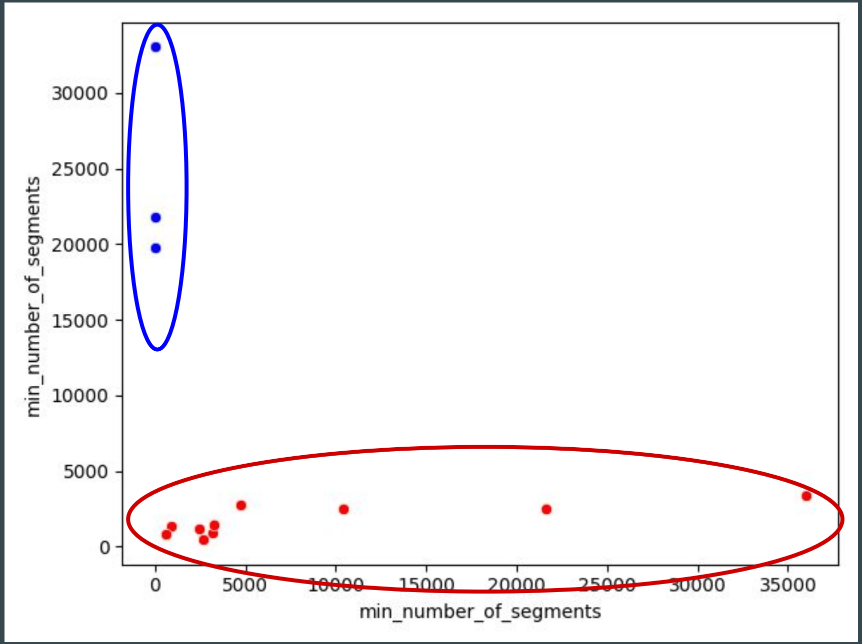
Question: what do my data represent? First I need to visualize the dataset

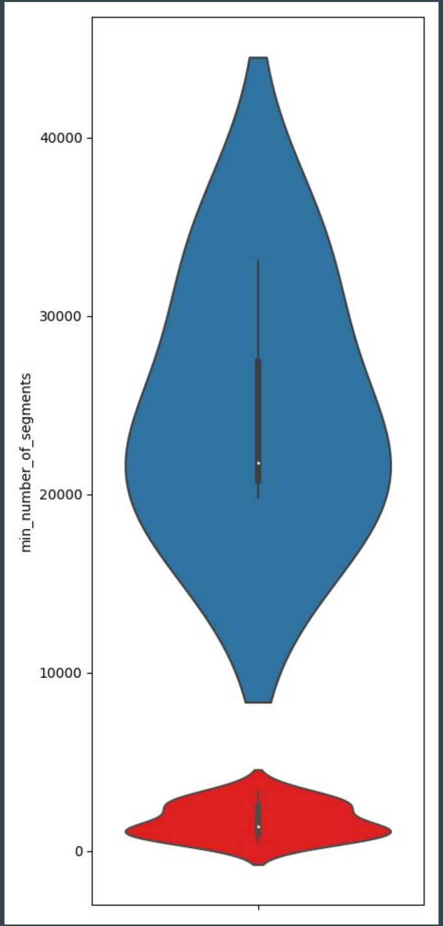
Two violin plots



Data collection

Or as two distinct sets of groups





What are the challenges in data collection?

Data compatibilities

- Data sources - are the collection methods compatible ?
- Metadata - are the data of the same kind? For example, animal age, species etc
- File format - is the format of the data compatible?
- Units - are the data in the same units?
- Normalization - do I need to normalize my dataset?

Depending on the task, each of the above might have a different answer.

What are the challenges in data collection?

I need to ask for my data:

- What is the *question* I am trying to solve? (same dataset can be used for different questions)

For example:

If I need a statistical analysis between two datasets to evaluate if my transformations introduced a bias: I might need to normalize both datasets

But... if I need to identify classes in my dataset, normalization might hide differences between classes

What are the challenges in data collection?

Data compatibilities

Question to be addressed

Methods that should be used

Questions?