

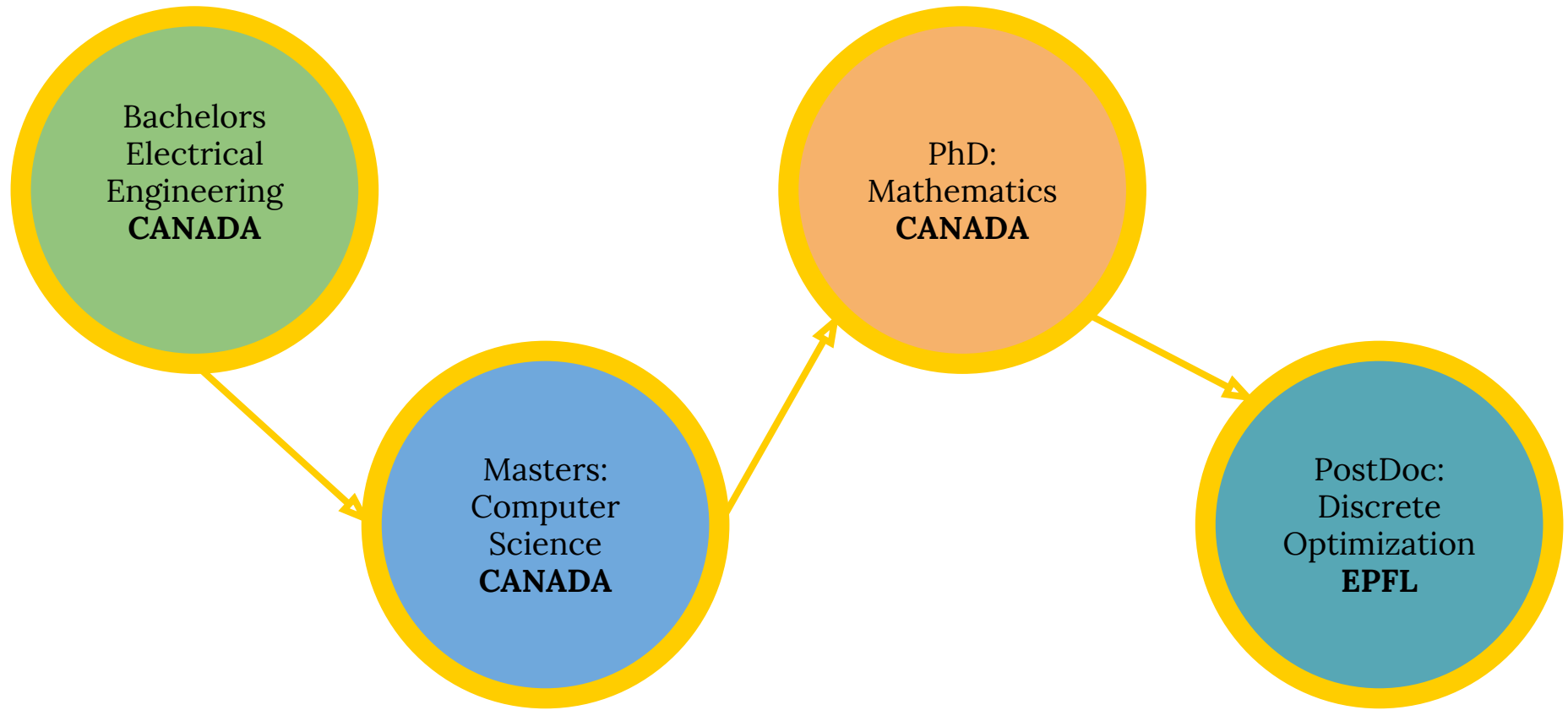


# Data analytics with Python



**Hands-On Data Workshops || EPFL Extension School || May 26-27, 2018**

# About me





# Today's agenda

---

## 1. Data Preparation

- ◉ Read and load our data
- ◉ Summary statistics
- ◉ Selecting, indexing, transforming
- ◉ Basic plotting



# Today's agenda

---

## 2. Data wrangling

- Handling missing values
- Transformations
- Outliers
- Correlations



# Today's agenda

## 3. Machine Learning

- ◉ Supervised learning:
  - Regression
  - Classification
- ◉ Unsupervised learning
  - Clustering



# Today's agenda

---

## 3. Visualizations

- Scatter plots
- Histograms
- Distributions
- Correlation matrix
- Model evaluations

## ***The toolset***



“

A vertical grey line extends from the bottom of the yellow circle to the bottom edge of the slide.

Python 3



Jupyter Notebook



Numpy



Scikit-learn



Matplotlib



Pandas



Seaborn



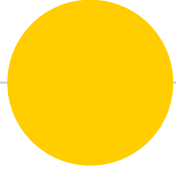


*Show me the data*



“

A thin vertical grey line extends downwards from the bottom of the yellow circle.



## The athlete's **social network**

- Users upload activities from tracking devices
- In depth analytics for premium users
- Strava Metro: data service for traffic and commute patterns



# Use Strava's API to download entire activity history

	A	B	C	D	E	F	G	H	I	J
	my_strava_activities									
1		average_speed	average_heartrate	average_watts	distance	elapsed_time	total_elevation_gain	type	commute	start_date_local
2	0	5.465	129.9	131.7	51690	3:23:50	1414	Ride	False	2018-05-21T09:59:24
3	1	6.948	131.3	127.8	133126	6:17:54	2280	Ride	False	2018-05-20T08:46:44
4	2	6.853	135.4	136.4	108418	4:39:40	1950	Ride	False	2018-05-19T08:12:35
5	3	6.756		135.3	59019.1	2:38:43	937	Ride	False	2018-05-17T17:50:55
6	4	7.454		117.7	23830.9	0:55:01	122	Ride	False	2018-05-17T07:15:06
7	5	6.259	129.8	129	23773.2	1:06:00	504	Ride	False	2018-05-16T18:09:43
8	6	2.864			10847.6	1:08:12	347.3	Run	False	2018-05-15T19:49:01
9	7	2.928			4995.2	0:31:15	50	Run	False	2018-05-14T19:38:03
10	8	7.289	132.8	121.9	129491	5:55:04	1833	Ride	False	2018-05-12T09:13:34

# 585 runs 364 rides

Whoa! That's a big number





## About the data

### Average speed

Average speed  
in meters per  
second

### Elapsed Time

The total  
elapsed time

### Distance

Total distance in  
meters

### Total elevation gain

Total elevation  
in meters

### Commute

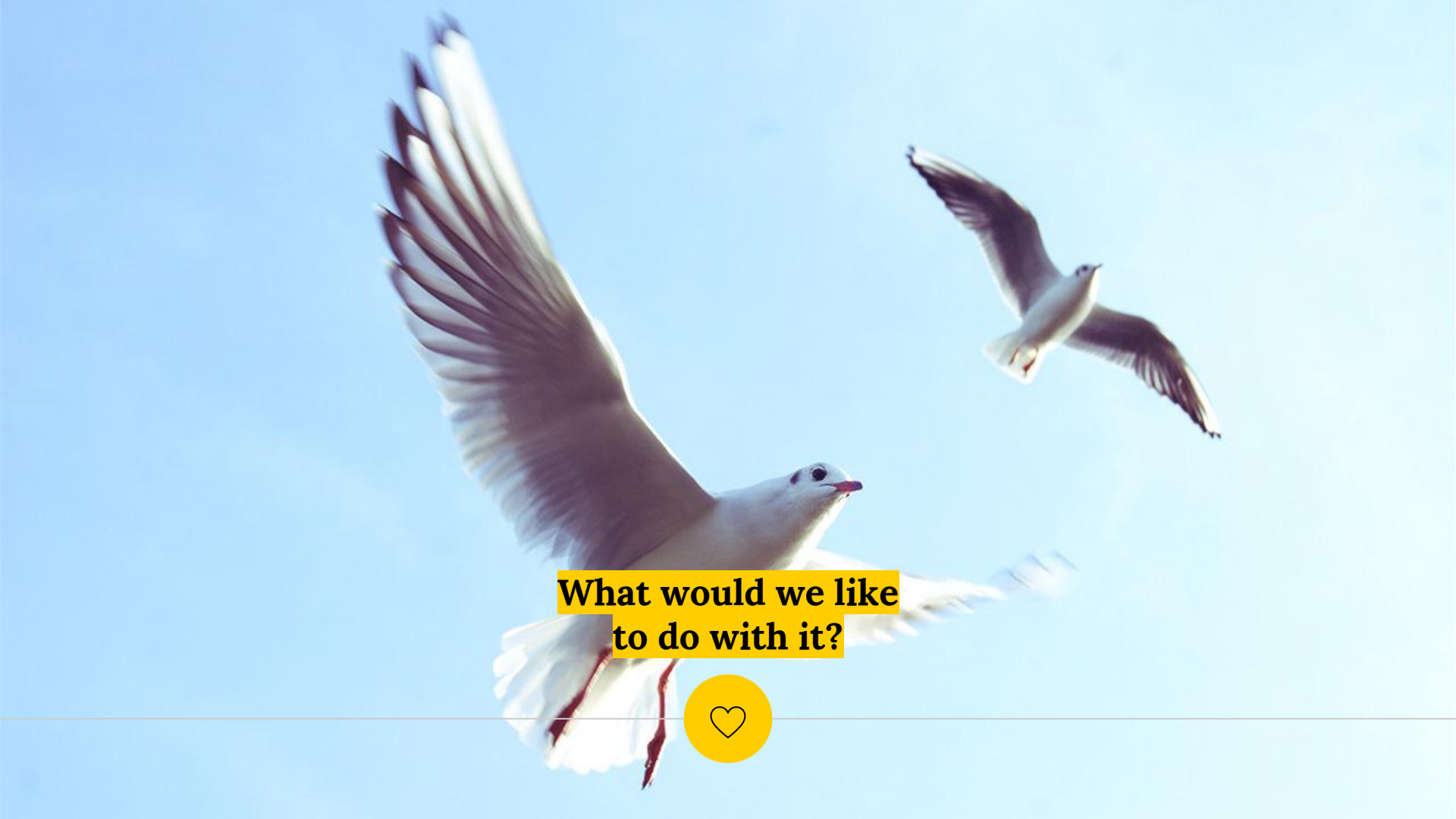
Whether the activity  
is a commute

### Average Heartrate

Beats per minute

### Average Watts

Cycling specific

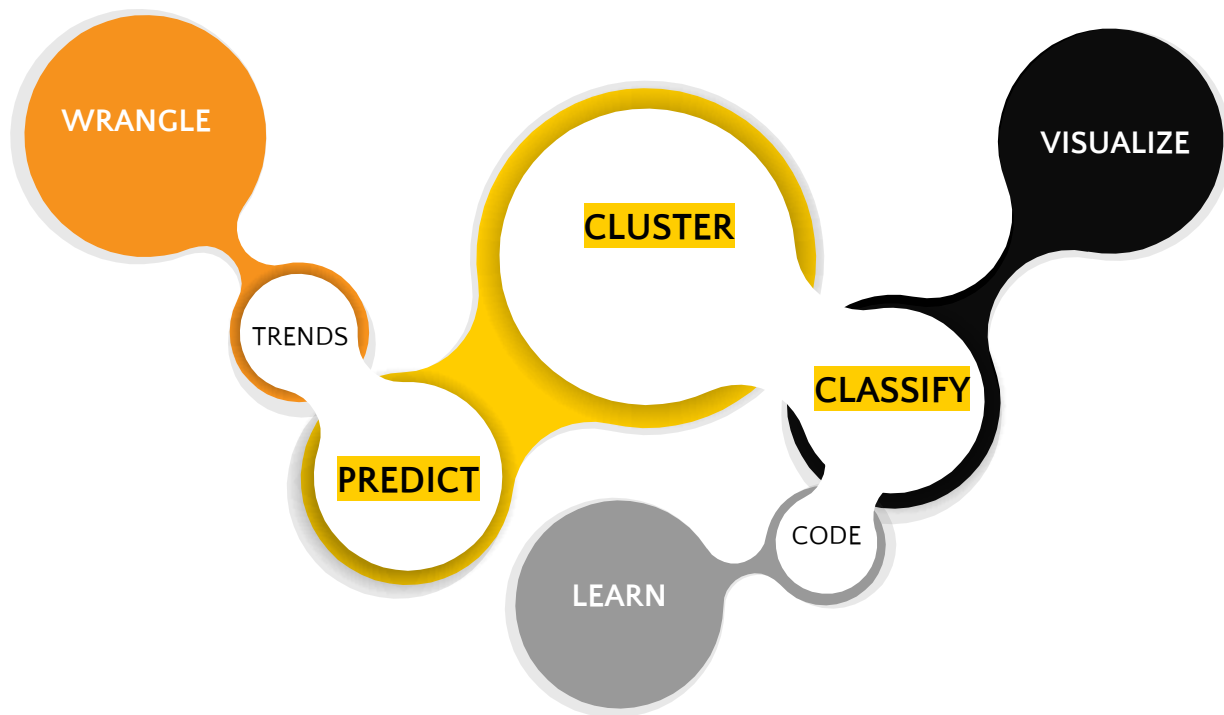
A photograph of two white doves in flight against a clear, bright blue sky. The dove in the foreground is shown from a low angle, with its wings fully extended, revealing the intricate structure of its feathers. Its head is turned slightly to the right. The second dove is positioned further back and higher up in the frame, also in flight. The overall mood is peaceful and hopeful.

**What would we like  
to do with it?**





## Outline of goals





**let's get started!**