

R in Grenoble

DATA CHALLENGES

Magali Richard & Florent Chuffart

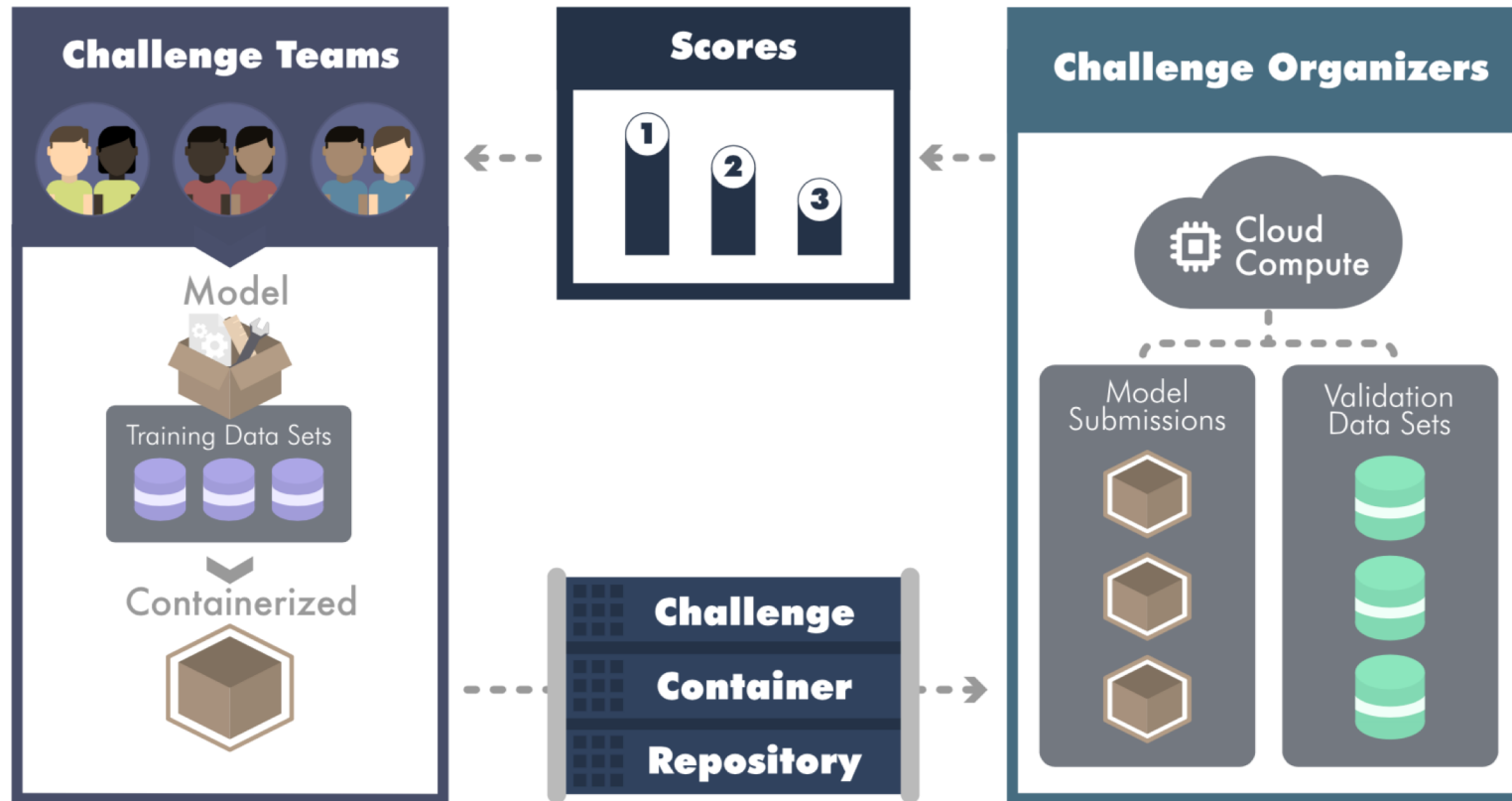
Introduction

Data challenges in class

Data challenges for scientists

Tutorial

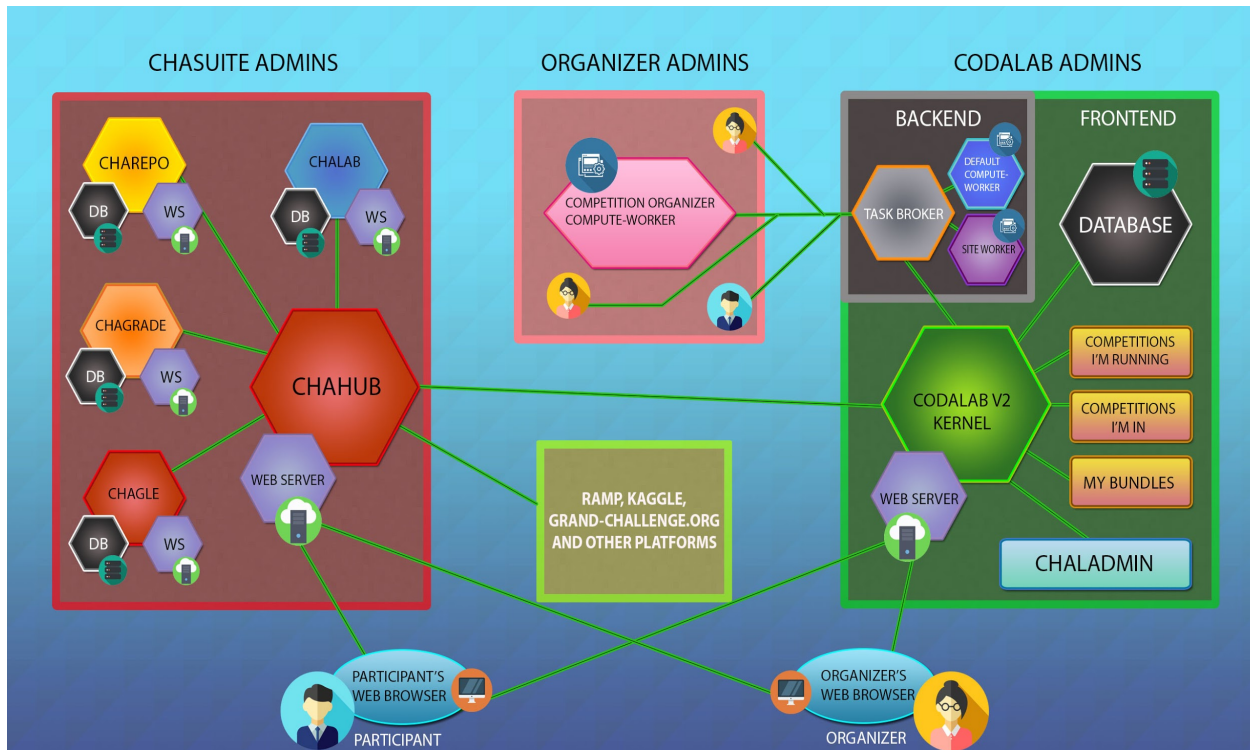
What is a data challenge?



Ellrott et al. Genome Biology (2019) 20:195

The challenge platform

CodaLab

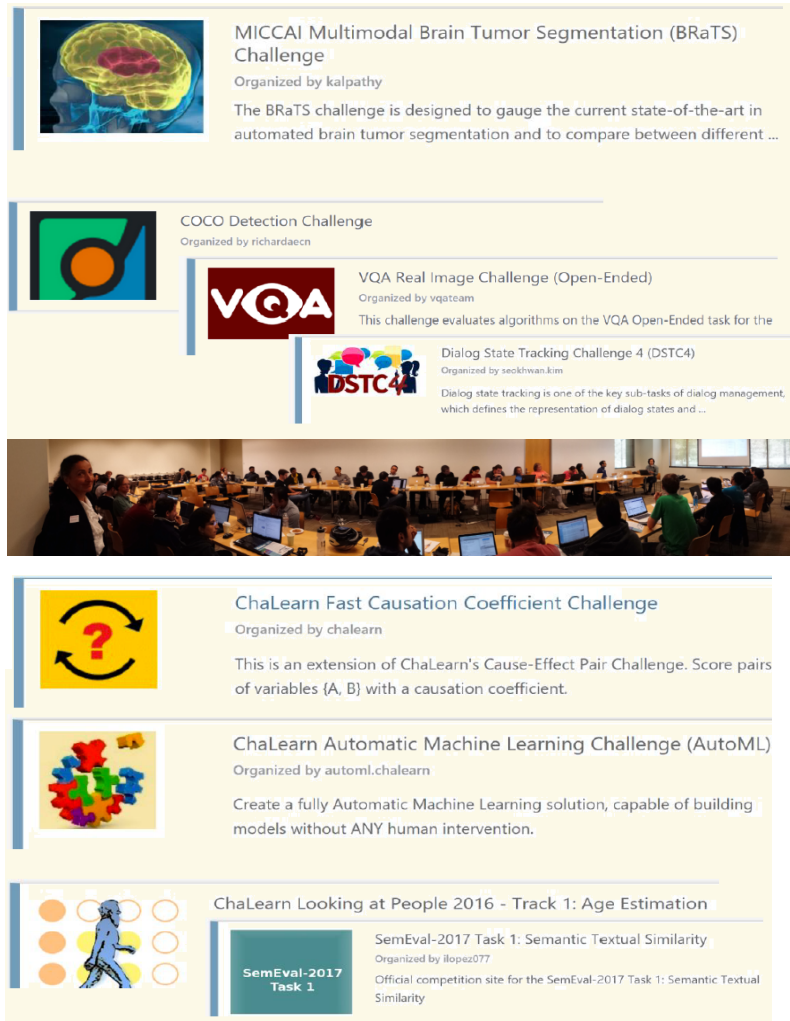


- Enables participants to submit their codes
- Automatically rank the participants



Alexis Arnaud

History



MICCAI Multimodal Brain Tumor Segmentation (BRaTS) Challenge
Organized by kalpathy
The BRaTS challenge is designed to gauge the current state-of-the-art in automated brain tumor segmentation and to compare between different ...

COCO Detection Challenge
Organized by richardaecn

VQA Real Image Challenge (Open-Ended)
Organized by vqateam
This challenge evaluates algorithms on the VQA Open-Ended task for the

Dialog State Tracking Challenge 4 (DSTC4)
Organized by seokhwan.kim
Dialog state tracking is one of the key sub-tasks of dialog management, which defines the representation of dialog states and ...

ChaLearn Fast Causation Coefficient Challenge
Organized by chalearn
This is an extension of ChaLearn's Cause-Effect Pair Challenge. Score pairs of variables (A, B) with a causation coefficient.

ChaLearn Automatic Machine Learning Challenge (AutoML)
Organized by automl.chalearn
Create a fully Automatic Machine Learning solution, capable of building models without ANY human intervention.

ChaLearn Looking at People 2016 - Track 1: Age Estimation

SemEval-2017 Task 1: Semantic Textual Similarity
Organized by ilopez077
Official competition site for the SemEval-2017 Task 1: Semantic Textual Similarity

2013: Medical data. Result submission.

COMPETITION BUNDLES

2014: Computer vision, speech, NLP, IR.
MSCOCO: 361 participants.

2015: **CODE SUBMISSION**

AutoML: 687 participants
Hackathons. Coopetitions.

2016: ChaLab Wizard.

USE IN EDUCATION

2017: 480 challenges, 10000 users.

SCALABILITY, REUSABILITY, DYNAMIC COMPETITIONS

See.4c: EU prize, 2 million Euros

Introduction

Data challenges in class

Data challenges for scientists

Tutorial

A novel pedagogic approach

As practical work

Sex Prediction Challenge (sexpred)

Organized by Alexis_Arnaud

The challenge provides the data.frame d of tumoral tissues described by genes expression values, histological and clinical attributes. The ...

Nov 15, 2019-Dec 30, 2019

37 participants

As homework

Histology Prediction Challenge (histpred)

Organized by Alexis_Arnaud

The challenge provides the data.frame d of tumoral tissues described by genes expression values, histological and clinical attributes. The ...

Nov 21, 2019-Dec 30, 2019

37 participants

As final evaluation



Virulence Prediction Challenge (virpred)

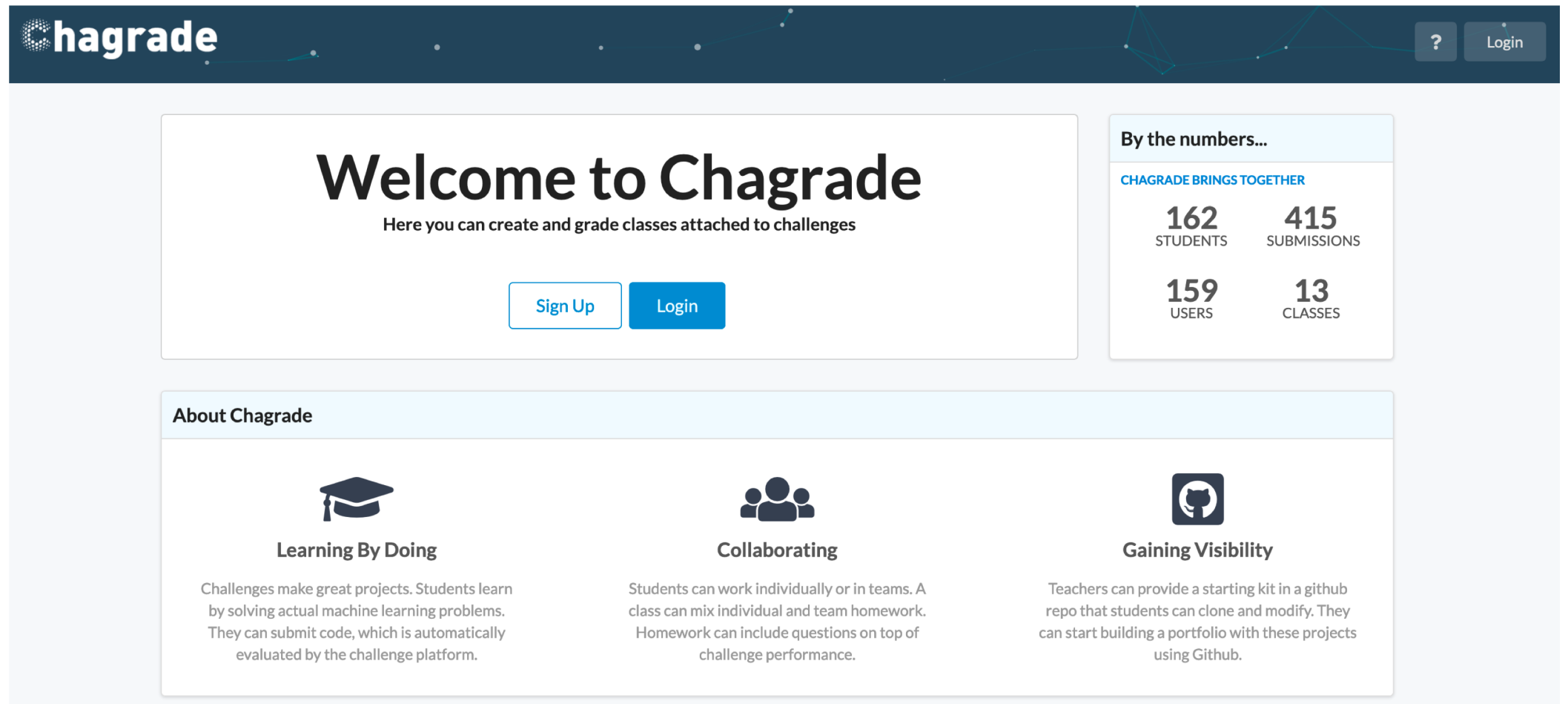
Organized by FlorentC

The challenge provides the data.frame d of tumoral tissues described by genes expression values, histological and clinical attributes. The ...

Jan 09, 2020-Jan 11, 2020

35 participants

Chagrade: a dedicated tool in codalab



The screenshot shows the Chagrade website homepage. At the top left is the Chagrade logo. On the top right, there are links for help (a question mark icon) and login. The main content area is divided into three sections: a large welcome message with sign-up and login buttons, a statistics section titled 'By the numbers...' showing 162 students, 415 submissions, 159 users, and 13 classes, and an 'About Chagrade' section with three columns: 'Learning By Doing' (with a graduation cap icon), 'Collaborating' (with a group of people icon), and 'Gaining Visibility' (with a GitHub icon).

Chagrade ? Login

Welcome to Chagrade

Here you can create and grade classes attached to challenges


[Sign Up](#) [Login](#)

By the numbers...

CHAGRADE BRINGS TOGETHER


162 STUDENTS	415 SUBMISSIONS
159 USERS	13 CLASSES

About Chagrade




Learning By Doing

Challenges make great projects. Students learn by solving actual machine learning problems. They can submit code, which is automatically evaluated by the challenge platform.



Collaborating

Students can work individually or in teams. A class can mix individual and team homework. Homework can include questions on top of challenge performance.



Gaining Visibility

Teachers can provide a starting kit in a github repo that students can clone and modify. They can start building a portfolio with these projects using Github.

Introduction

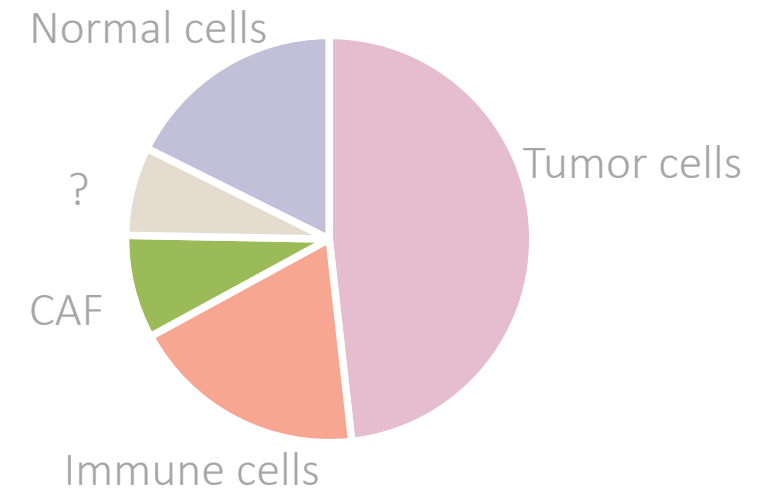
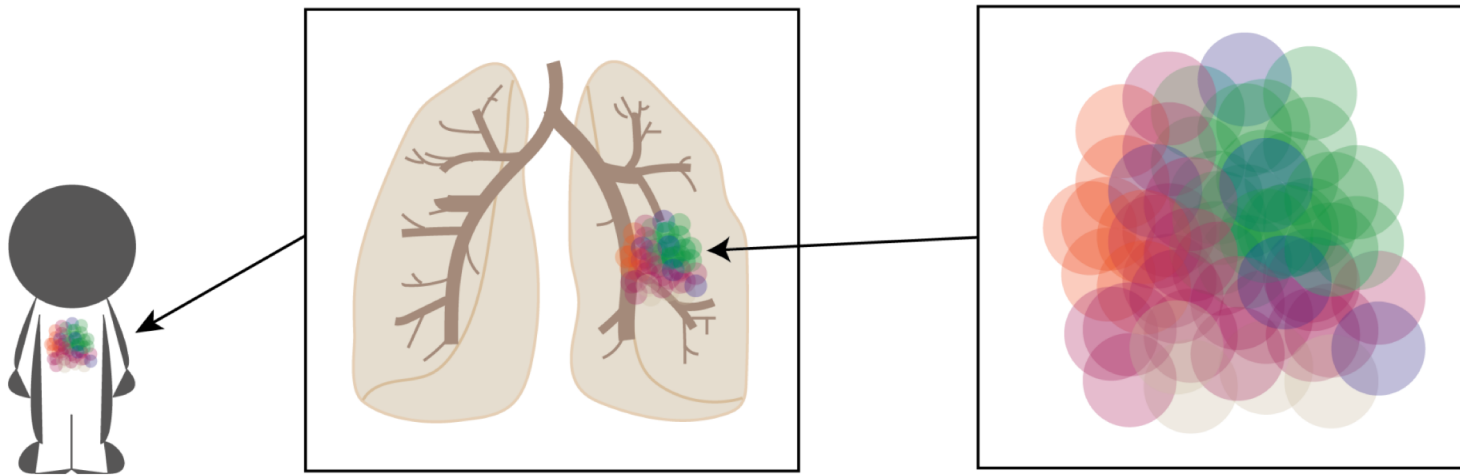
Data challenges in class

Data challenges for scientists

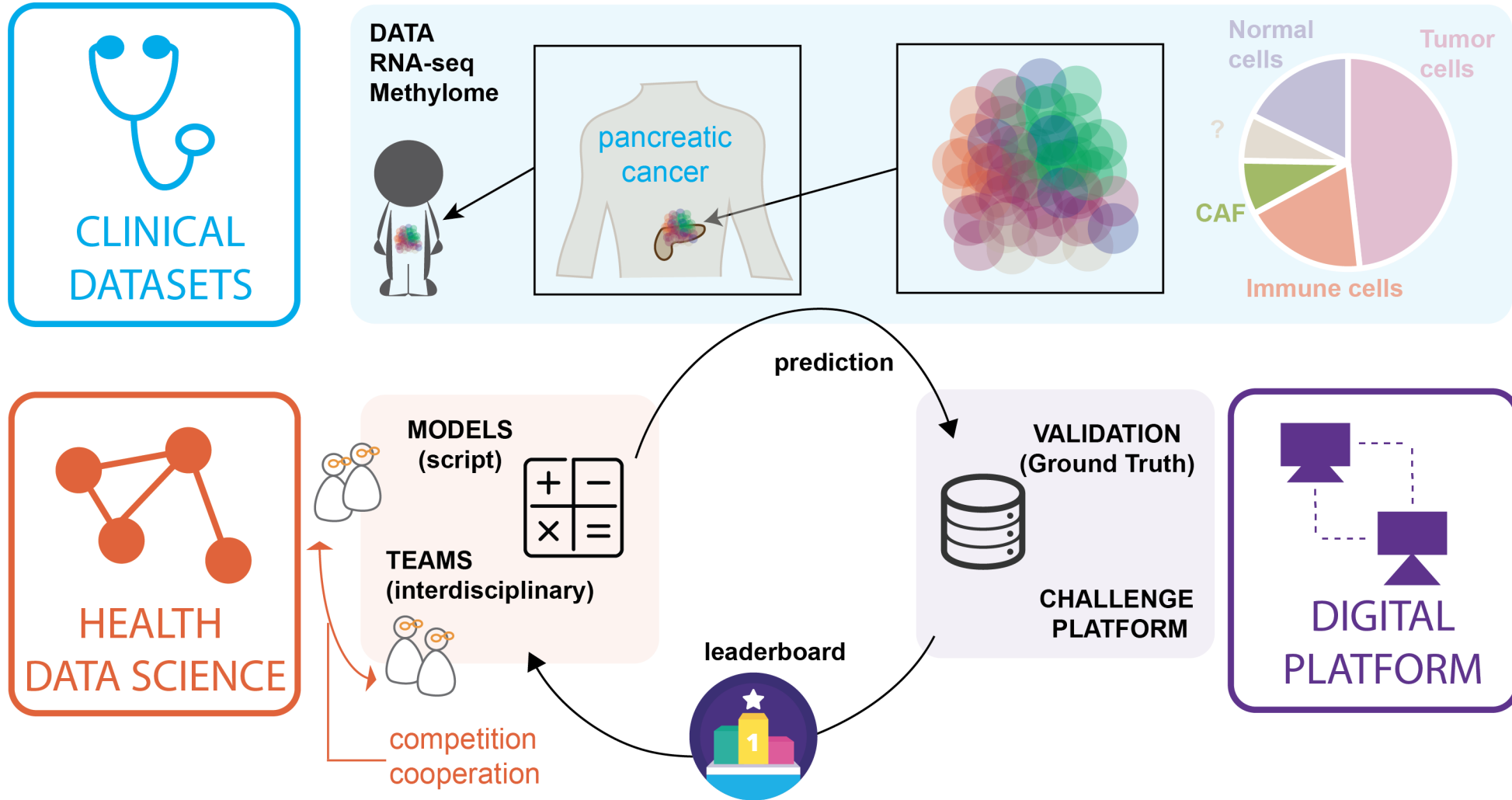
Tutorial

Goal of our data challenge

- **Quantification of tumor heterogeneity**



A challenge for scientists



2 editions of the data challenge

1st edition (2018)



- Methylation Data
- One cancer type
- Cell lines

2nd edition (2019)



- Methylation and transcriptomic Data
- Several cancer types
- Primary tumors / cell lines

When and where?

- **When:** 25-29 of November 2019
- **Where:** Aussois (CAES CNRS), French Alps



Health Data Challenge
(2nd edition)
2019

November 25 to 29
Centre Paul Langevin
Aussois, French Alps



Deconvolution methods
to quantify tumor heterogeneity
in cancer research

**Registration deadline extension
September 30**

Invited speakers

Michael Scherer
Zitax - Frank Inst. für Informatik,
Saarbrücken, Germany
Francisco Avila Cobos
Ghent University Ghent, Belgium
Jerome Cros
JAF 1, Paris, France

Organizers

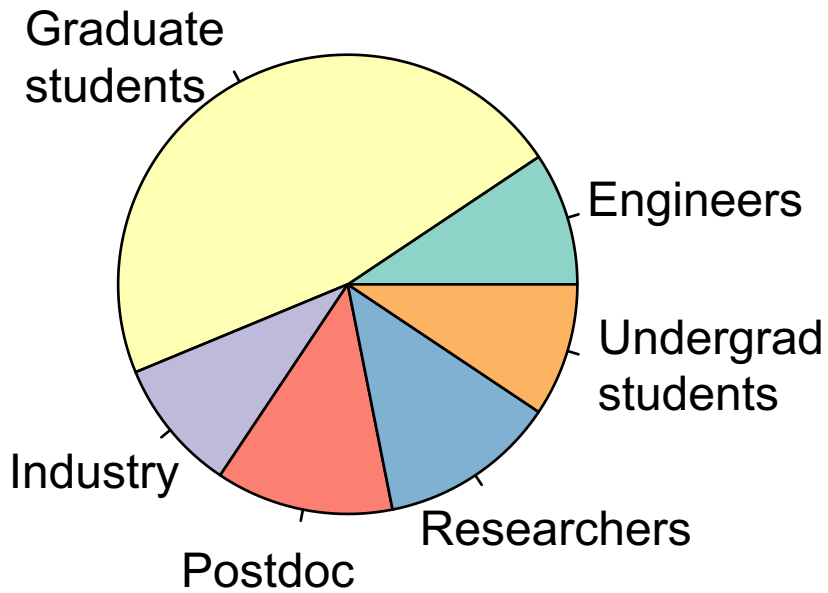
Yuna Blum & Florent Petitprez
Programme CIT, MCC, Paris, France
Magali Richard & Alexis Arnaud
UNRS, INRA, Grenoble, France

MORE INFORMATION & REGISTRATION
Website: bit.ly/2W7x7Y7
Contact: data.institutuniv.grenoble.alpes.fr
[@jrgrenobledata](https://twitter.com/jrgrenobledata) #hdc2019

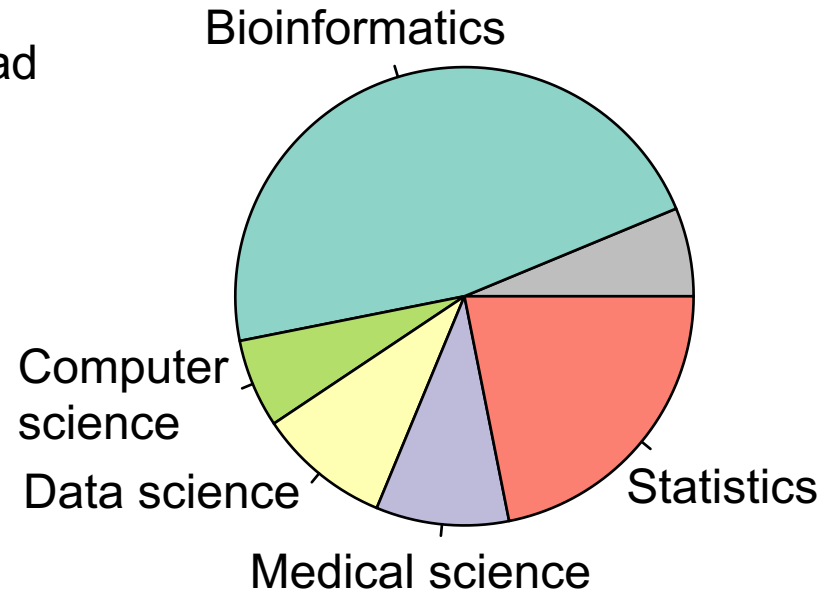




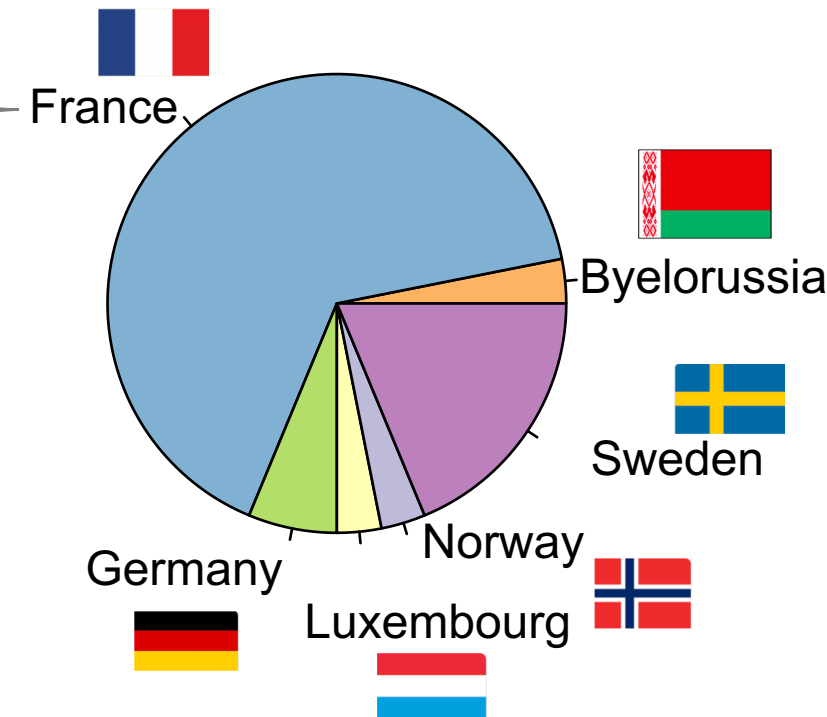
About the participants (n=34)



10 teams of 3-4 people



Institut Curie
UGA Grenoble
CRC Cordeliers
INSERM
CEA
Innate Pharma
Verteego
...



Agenda



Talks

#1

#2

#2

#2

Guidelines

#1

Poster session

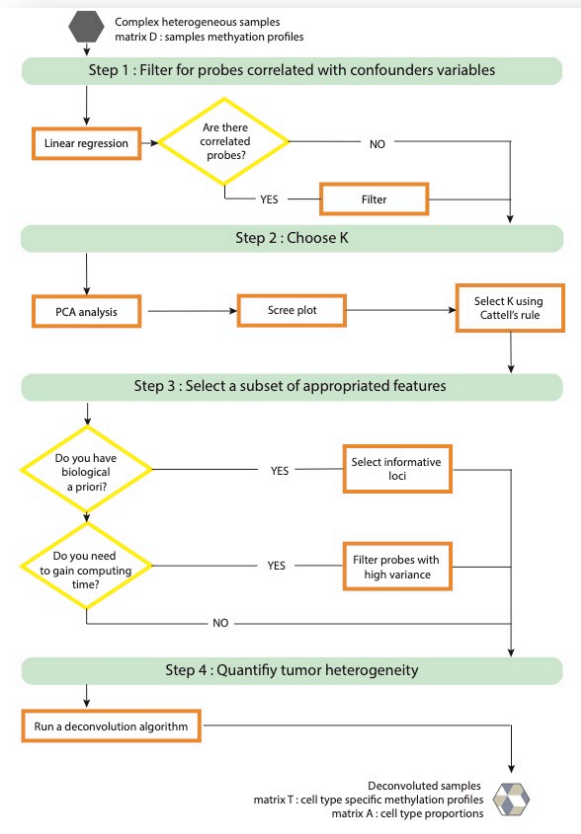
Social event

#2

Feed Back
Brain Storming

Valorization of the first edition

➤ Guidelines



➤ Article

CSH Cold Spring Harbor Laboratory
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

HOME | Search

In review in BCM bioinformatics

New Results Comment on this paper

Guidelines for cell-type heterogeneity quantification based on a comparative analysis of reference-free DNA methylation deconvolution software

Clementine Decamps, Florian Privé, Raphael Bacher, Daniel Jost, Arthur Wagué, HADACA consortium, Eugene Andres Houseman, Eugene Lurie, Pavlo Lutsik, Aleksandar Milosavljevic, Michael Scherer, Michael G.B. Blum, Magali Richard

doi: <https://doi.org/10.1101/698050>

➤ R package *medepir*

<https://rdrr.io/github/bcm-uga/medepir/man/medepir-package.html>



M Richard, C Decamps, F Privé, M Blum

➤ Blog posts

M

Towards Data Science
Sharing concepts, ideas, and codes

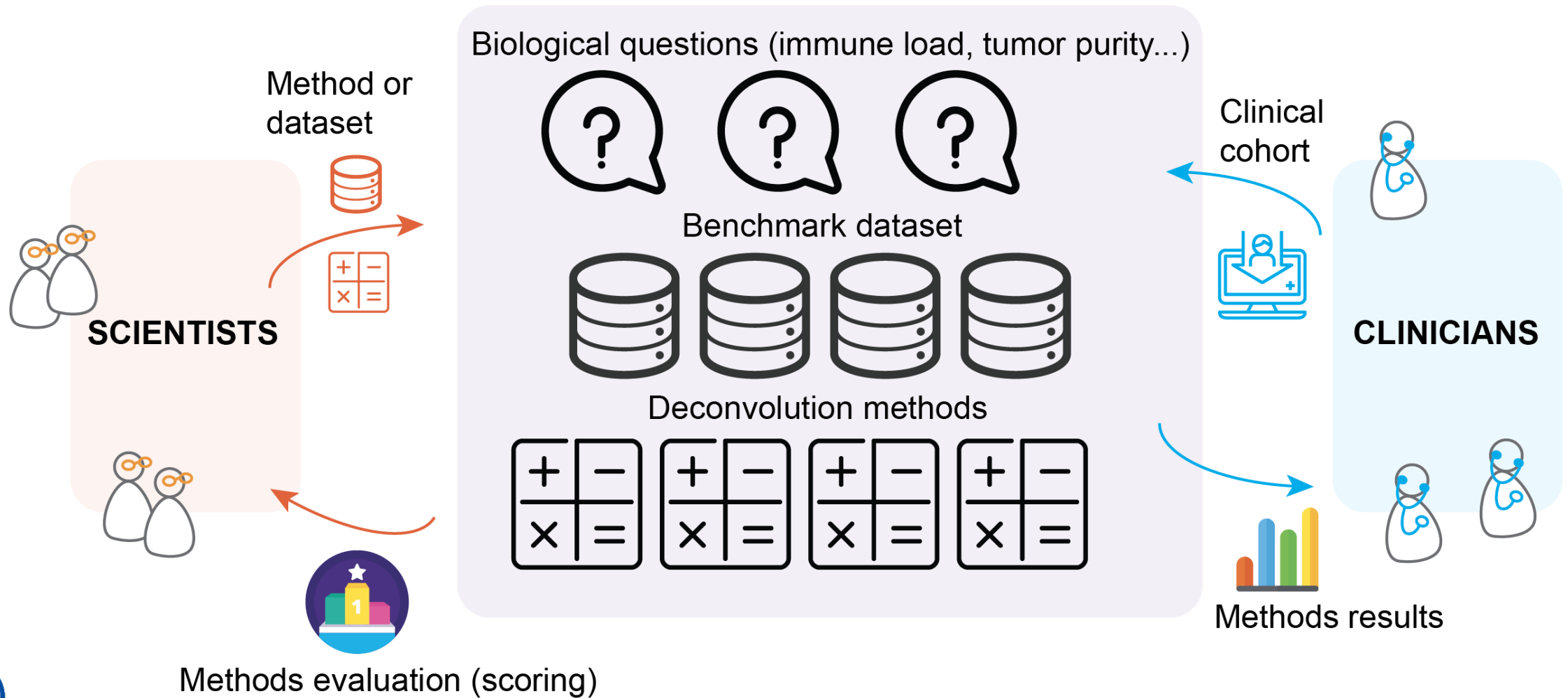
DATA SCIENCE MACHINE LEARNING PROGRAMMING VISUALIZATION

Health data challenges organization: feedback, comments and recommendations.

Authors: Elise Amblard, Yuna Blum, Jane Merlevede, Magali Richard

In preparation

From data challenges to benchmarks



Introduction

Data challenges in class

Data challenges for scientists

Tutorial

On line tutorial by Alexis Arnaud

Codalab >  Bundle_generation > Details




Bundle_generation 




Project ID: 5596



★ Star 0

🍴 Fork 0

Clone 

🔗 11 Commits  1 Branch  0 Tags  748 KB Files

 README.md

To create a Codalab competition, you have to generate a zip file with the required files inside and upload the archive on the Codalab platform. The *R* script `Bundle_generator_--cmd.Rmd` allows you to automatically create a valid bundle by using the *R* terminal to ask you some questions in order to personalize the competition.

Generate a Codalab bundle

In a *R* terminal, run the following command : `rmarkdown::render(input = "Bundle_generator_--cmd.Rmd", envir = new.env())`

You can also generate a toy bundle, with the parameters by default, with *Rscript* : `Rscript -e 'rmarkdown::render(input = "Bundle_generator_--cmd.Rmd", envir = new.env())'`

Test the Codalab bundle

In a *R* terminal, run the following command : `rmarkdown::render(input = "Bundle_test.Rmd", envir = new.env())`

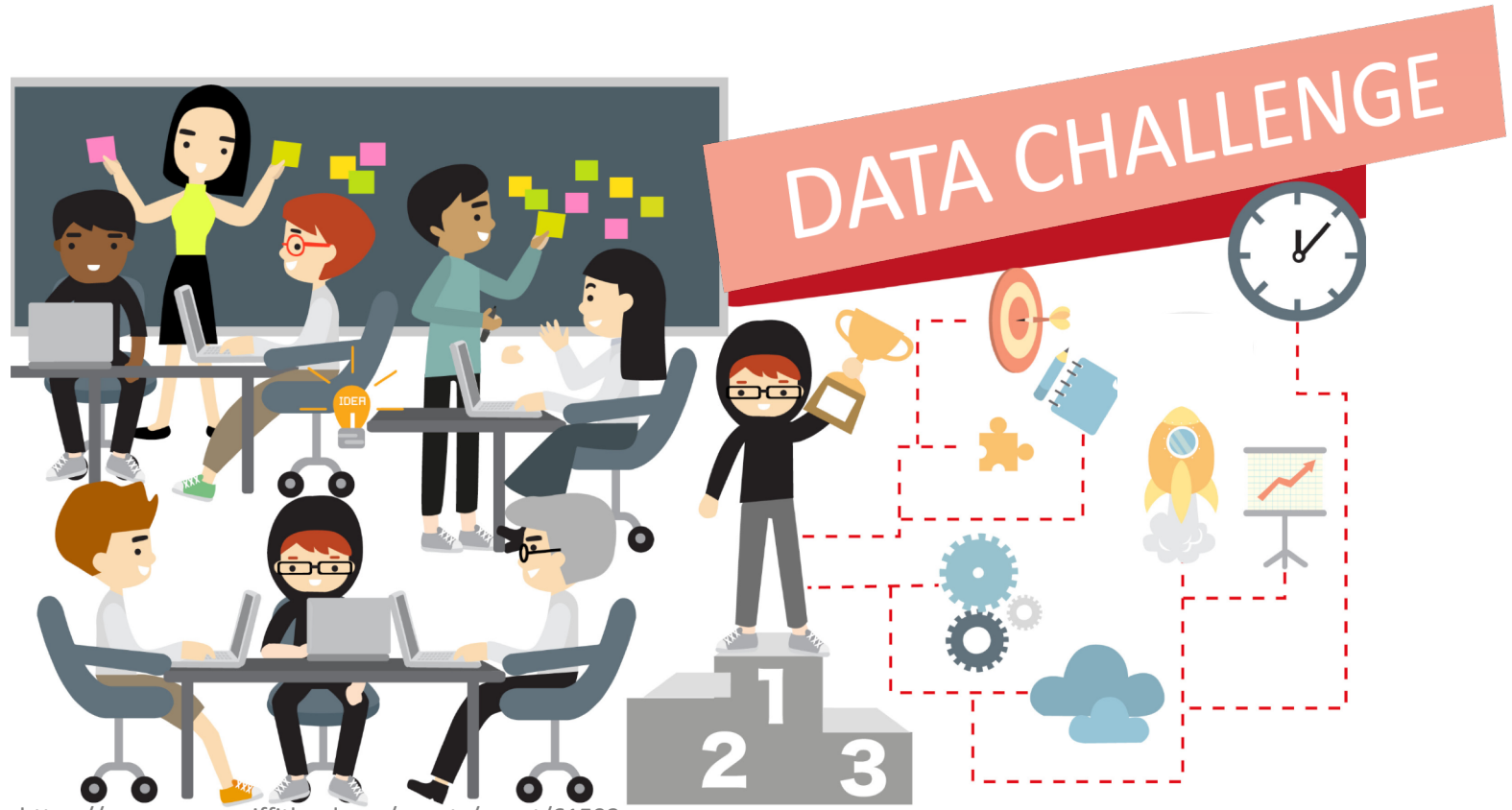
[WIP] Render a web site to previsualize the competition

It gives you the possibility to visualize your competition with a minimalist website which looks like the Codalab platform.

`rmarkdown::render_site(input = "Web_page_generator.Rmd", envir = new.env())`

It's up to you now!

Thank you for your attention !



<https://app.secure.griffith.edu.au/events/event/61593>