01 - Course Overview

ml4econ, HUJI 2020

Itamar Caspi

March 15, 2020 (updated: 2020-03-22)

10-Year challenge

2010: ML = Maximum Likelihood
2020: ML = Machine Learning

An aside: about the structure of these slides

- The course's slide decks are created using the xaringan (/ʃæ.ˈriŋ.gæn/) R package and Rmarkdown.
- Some slides include hidden comments. To view them, press **p** on your keyboard

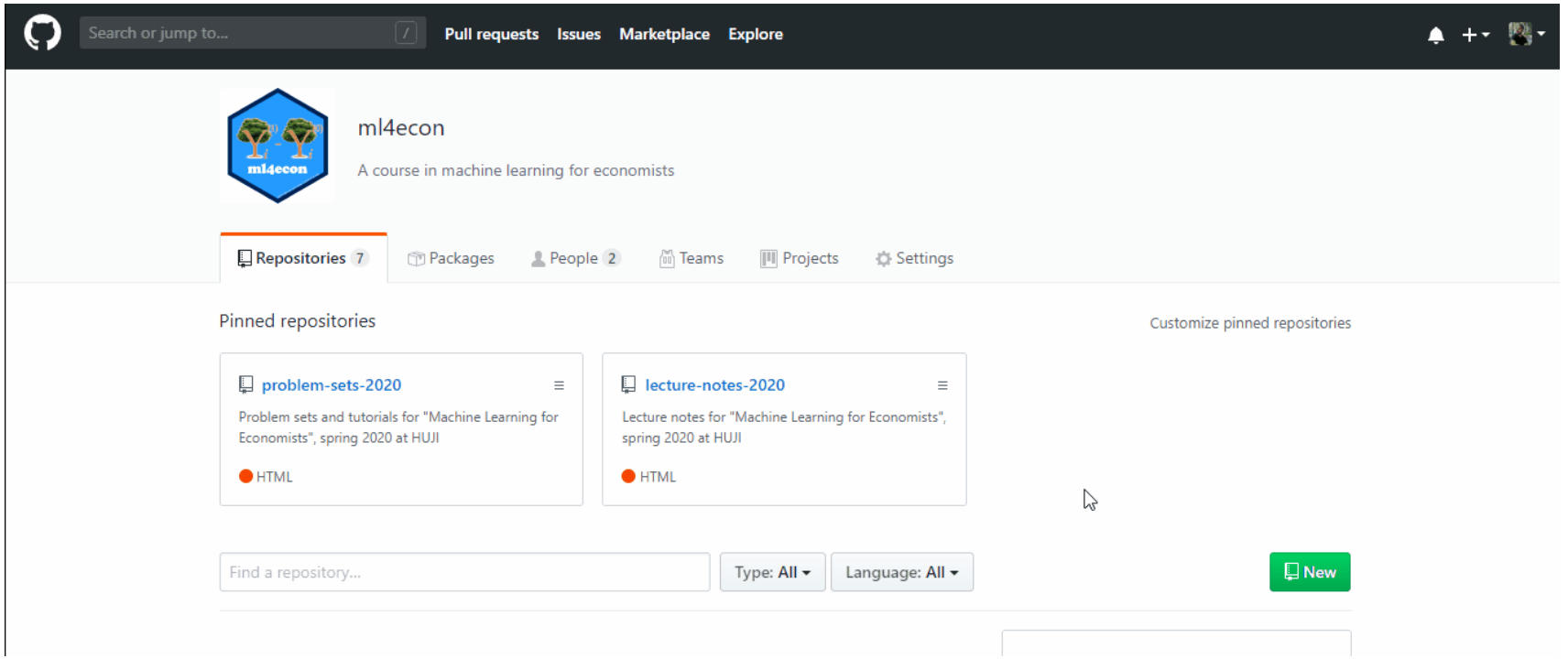
Outline

1. Logistics
2. About the Course
3. To Do List

Logistics

ml4econ GitHub repository

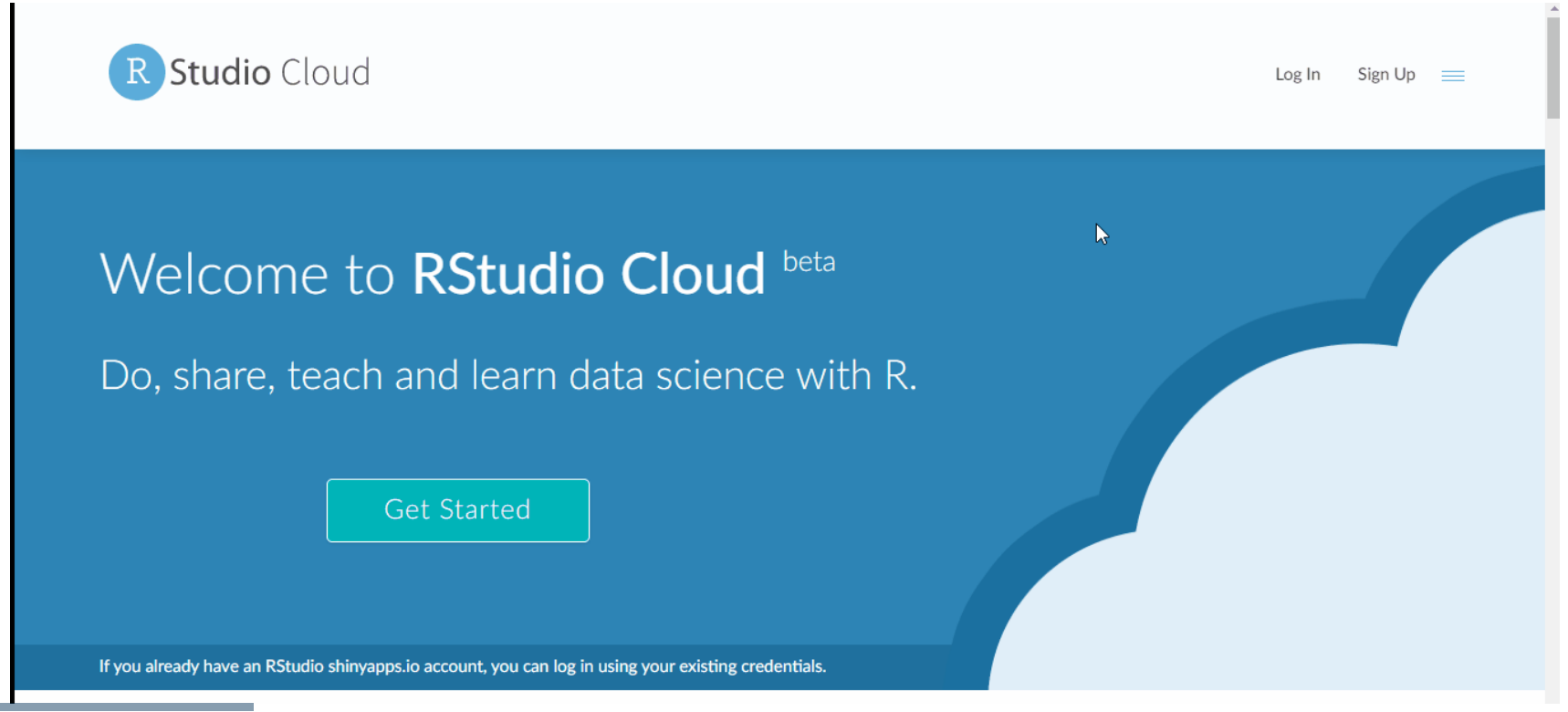
The class's GitHub repository: https://github.com/ml4econ



ml4econ RStudio Cloud workspace

RStudio Cloud is a hosted version of RStudio in the cloud that will make it easy for R and RStudio novices to learn data science and machine learning using R.

You can access our ml4econ workspace here.



People

- **Itamar Caspi**
  - Head of Monetary Analysis Unit, Research Department, Bank of Israel.
  - email: caspi.itamar@gmail.com
  - homepage: https://itamarcaspi.rbind.io/
- **Dor Goldenberg**
  - Asistant economist, Monetary Analysis Unit, Research Department, Bank of Israel; MA student, HUJI.
  - email: dorgoldenberg@gmail.com
- Meeting hours: after class, on demand.

Feedback

This is the second time we run this course $\Rightarrow$ your continuous feedback is important!

Please feel free to contact us by

- email
- in person
- or open an issue in our discussion forum

About the Course

Prerequisites

- Advanced course in econometrics.
- Some experience with R (or another programming language) are a plus.

This course is

About

How and when to apply ML methods in economics

- estimate treatment effects.
- prediction policy.
- work with new types of data (e.g., text).

To do that we will need to understand

- what is ML?
- how it relates to stuff you already know?
- how it differs?

Not about

- Cutting-edge ML techniques (e.g., deep learning)
- Computational aspects (e.g., gradient descent)
- Data wrangling (a.k.a. "feature engineering")
- Distributed file systems (e.g., Hadoop, Spark)

Tentative schedule

Week	Topic
1	Course Overview & Reproducibility
2	Basic ML Concepts
3	Regression and Regularization
4	Classification
5	Non-parametrics
6	Unsupervised Learning
7	Text analysis
8	Causal Inference
9	Lasso and Average Treatment Effects
10	Trees and Heterogeneous Treatment Effects
11	Prediction Policy Problems
12	The Economics of AI

> **NOTE**: This schedule can (and probably will) go through changes!

Readings on ML for economists

> All materials and lecture notes will be available on the class website.

Please read the following excellent surveys:

- The impact of machine learning on economics Athey (2018)
  In *The Economics of Artificial Intelligence: An Agenda*.
  University of Chicago Press.

- Machine learning: an applied econometric approach Mullainathan and Spiess (2017)
  Journal of Economic Perspectives, 31(2), 87-106.

Readings on ML

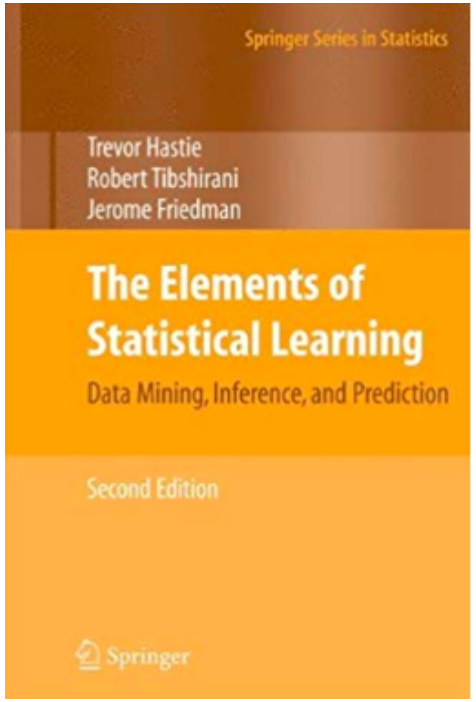
> All materials and lecture notes will be available on the course repo.

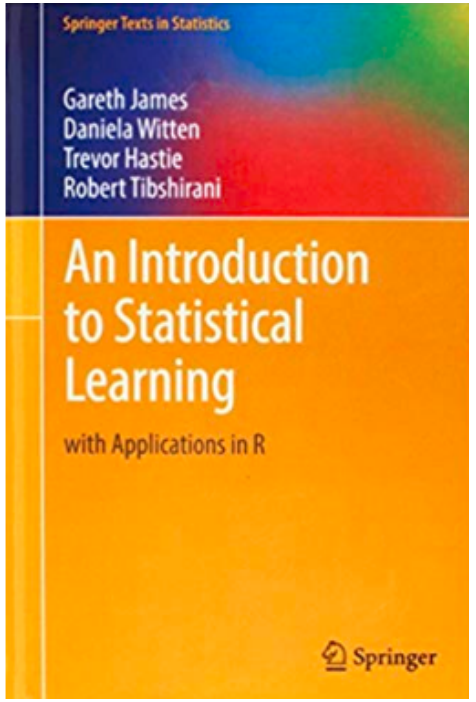
There are **no** required textbooks.

A couple of suggestions:

- An Introduction to Statistical Learning with Applications in R (ISLR)
  James, Hastie, Witten, et al. (2013)
  PDF available online

- The Elements of Statistical Learning (ELS)
  Hastie, Tibshirani, and Friedman (2009)
  PDF available online





More resources

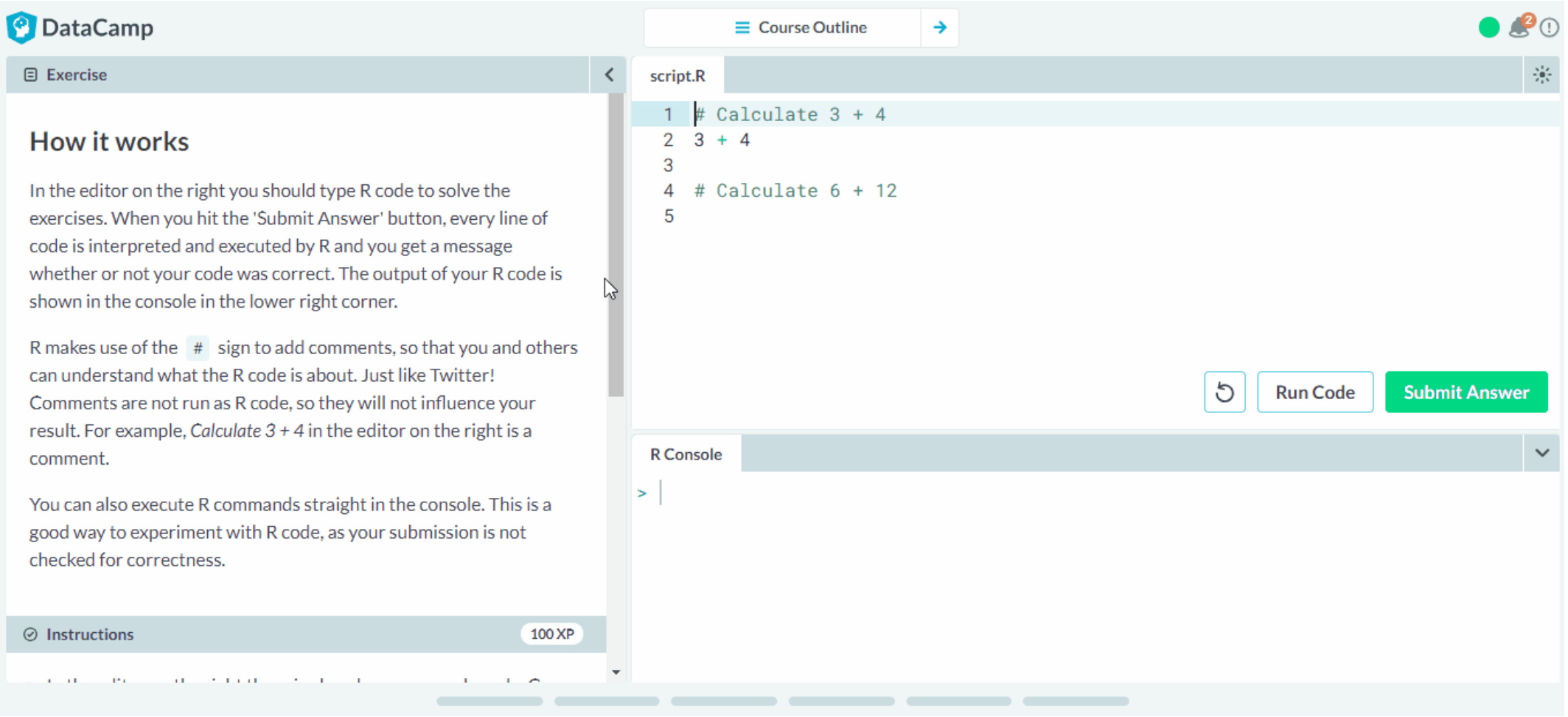
Can be found at our GitHub repo:

https://github.com/ml4econ/lecture-notes-2020/blob/master/resources.md

Programming

- Two of the most popular open-source programming languages for data science:
  - R
  - Python
- This course: R.
- Why R? See presentation notes and the FAQ section of our class website.
- We do encourage you to try out Python. However, we will only be able to provide limited support for Python users.

DataCamp in the classroom



Grading

Assignments:

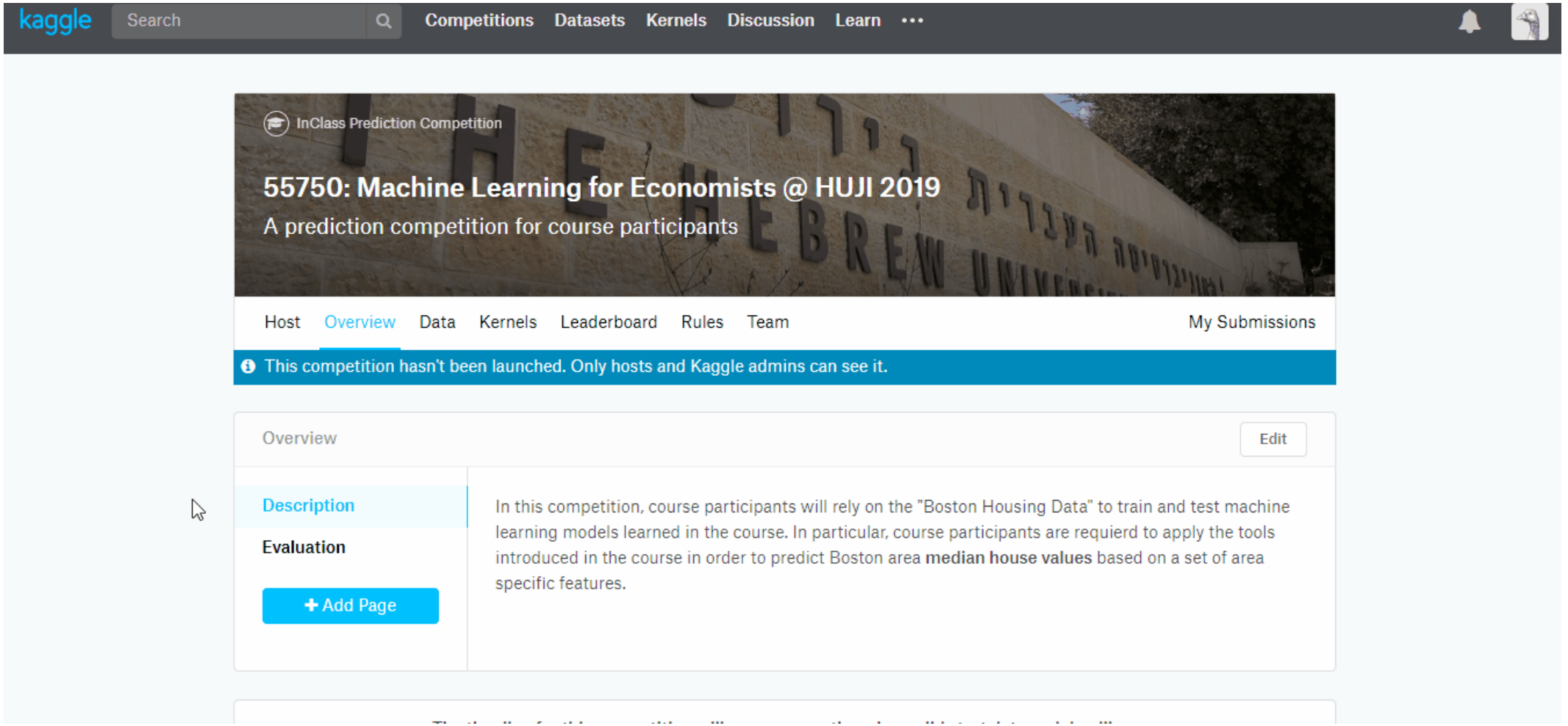
- DataCamp Classroom: we will provide you access to specific courses that will teach you essential R programming skills.
- $\sim$ 4 Problem sets.

Projects:

- Kaggle prediction competition: predict.
- Conduct a replication study based on one of the datasets included in the experimentdatar package.

GRADING: Assingments **20%**, project **40%**, final exam **40%**.

Kaggle


kaggle
Search
Competitions
Datasets
Kernels
Discussion
Learn
InClass Prediction Competition
55750: Machine Learning for Economists @ HUJI 2019
A prediction competition for course participants
Host
Overview
Data
Kernels
Leaderboard
Rules
Team
My Submissions
This competition hasn't been launched. Only hosts and Kaggle admins can see it.
Overview
Edit
Description
Evaluation
+ Add Page
In this competition, course participants will rely on the "Boston Housing Data" to train and test machine learning models learned in the course. In particular, course participants are requierd to apply the tools introduced in the course in order to predict Boston area median house values based on a set of area specific features.

experimentdatar

We will also make use of he `experimentdatar` data package that contains publicly available datasets that were used in Susan Athey and Guido Imbens' course "Machine Learning and Econometrics" (AEA continuing Education, 2018).

- You can install the **development** version from GitHub

```
# install.packages("devtools")
devtools::install_github("itamarcaspi/experimentdatar")
```

- **EXAMPLE:** Load the `experimentdatar` package and the `social` dataset:

```
library(experimentdatar)
data(social)
```

- Tips:
  1. Runnig `?social` privides variable definitions.
  2. Running `dataDetails("social")` will open a link to the paper associated with `social`.

To Do List

Homework*

- ☑ Download and install Git.
- ☑ Download and install R and RStudio.
- ☑ Create an account on GitHub
- ☑ Create an account on Rstudio Cloud and ask Dor/Itamar to invite you to our workspace.
- ☑ Create an account on DataCamp and ask Dor/Itamar to invite you to DataCamp Classroom.

[*] Please consult the Guides section in our course's website.

```
slides %>% end()
```

Source code

References

[1] S. Athey. "The impact of machine learning on economics". In: *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 2018.

[2] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, 2009 .פבר. ISBN: 9780387848570.

[3] G. James, T. Hastie, D. Witten, et al. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer London, Limited, 2013. ISBN: 9781461471370.

[4] S. Mullainathan and J. Spiess. "Machine learning: an applied econometric approach". In: *Journal of Economic Perspectives* 31.2 (2017), pp. 87-106.