Basic Statistics for Biomedical Research

Alex Sanchez, Anna Sanjuan, Miriam Mota,
Esther Camacho, Mireia Ferrer, Santiago Perez-Hoyos

Statistics and Bioinformatics Unit. Vall d'Hebron Institut de Recerca

Readme

- License: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License http://creativecommons.org/licenses/by-nc-sa/4.0/
- You are free to:
  - **Share** : copy and redistribute the material
  - **Adapt** : rebuild and transform the material
- Under the following conditions:
  - **Attribution** : You must give appropriate credit, provide a link to the license, and indicate if changes were made.
  - **NonCommercial** : You may not use this work for commercial purposes.
  - **Share Alike** : If you remix, transform, or build upon this work, you must distribute your contributions under the same license to this one.

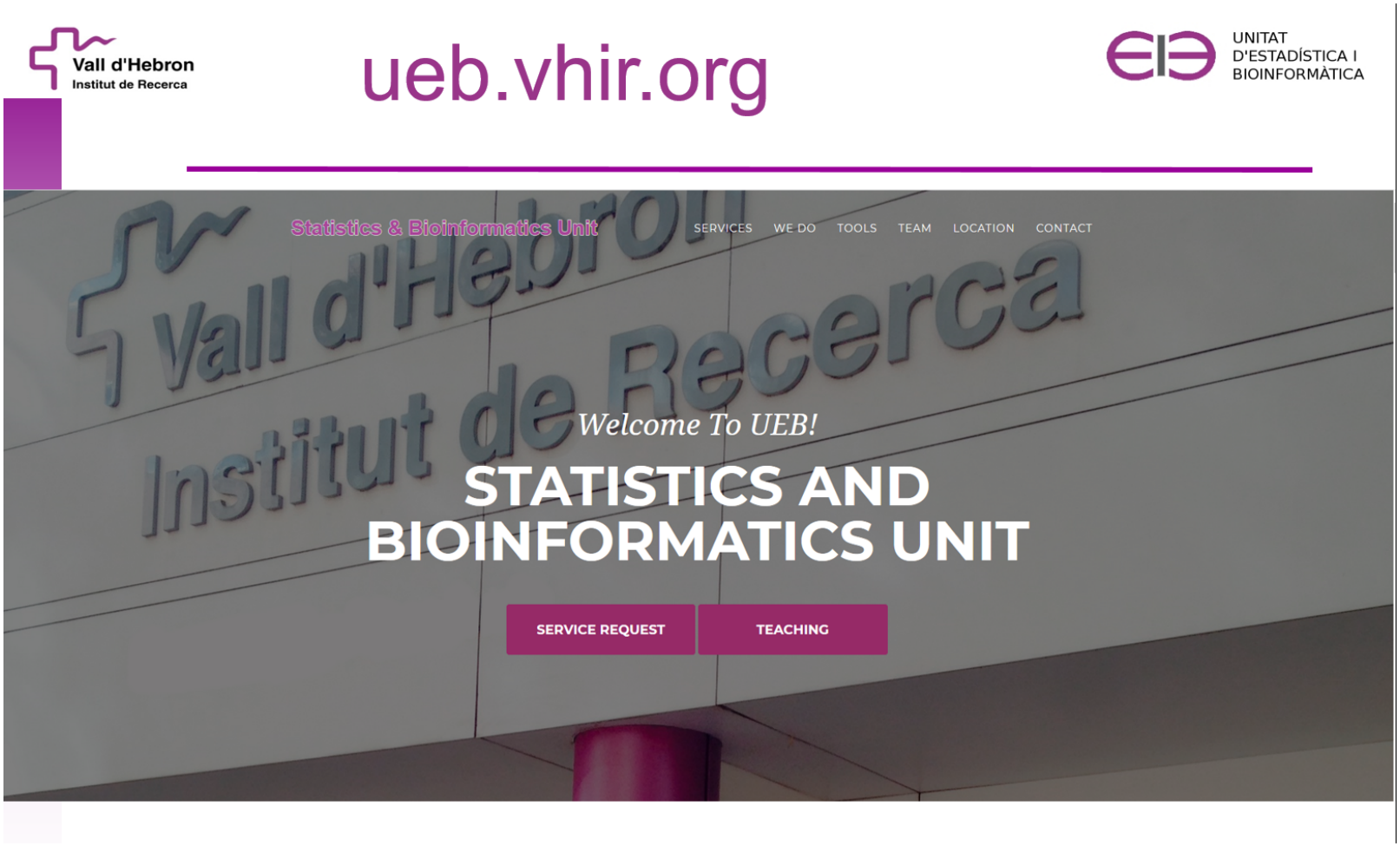
Section 1

Introduction

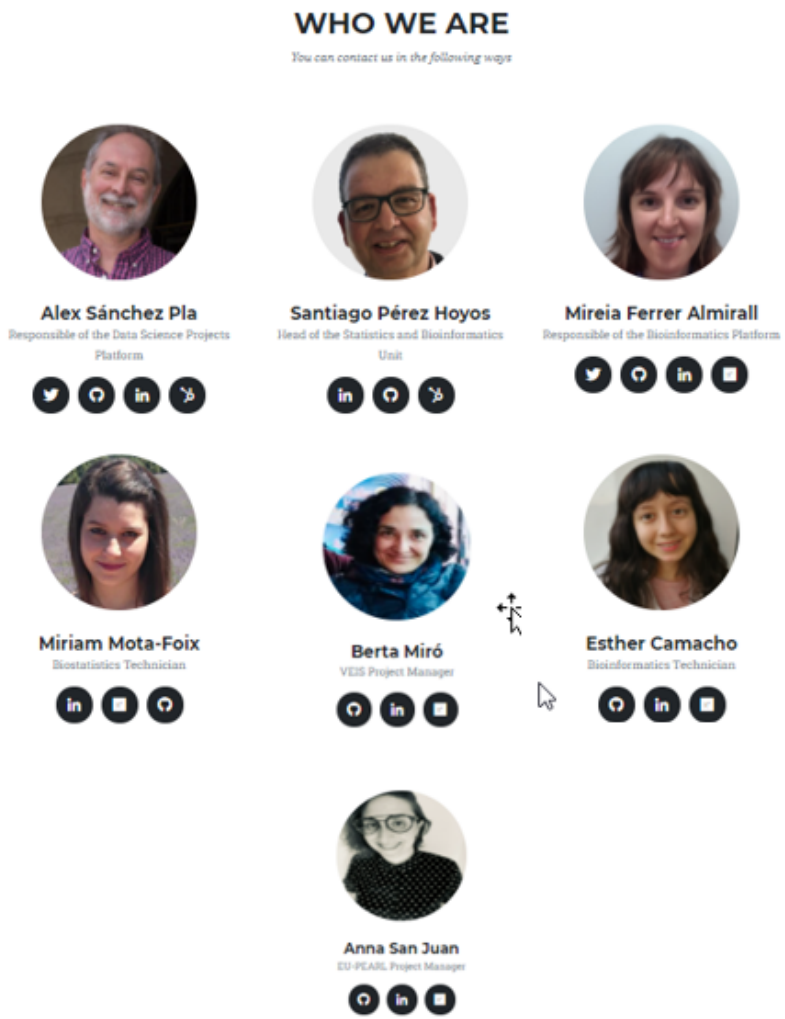
Outline

- Introduction
  - Who are we ("we"=teachers & students)
  - Why are we here (Why learn R?)
- How will we proceed: Methodology
- HW Data Science approach to using R
- References & Resources

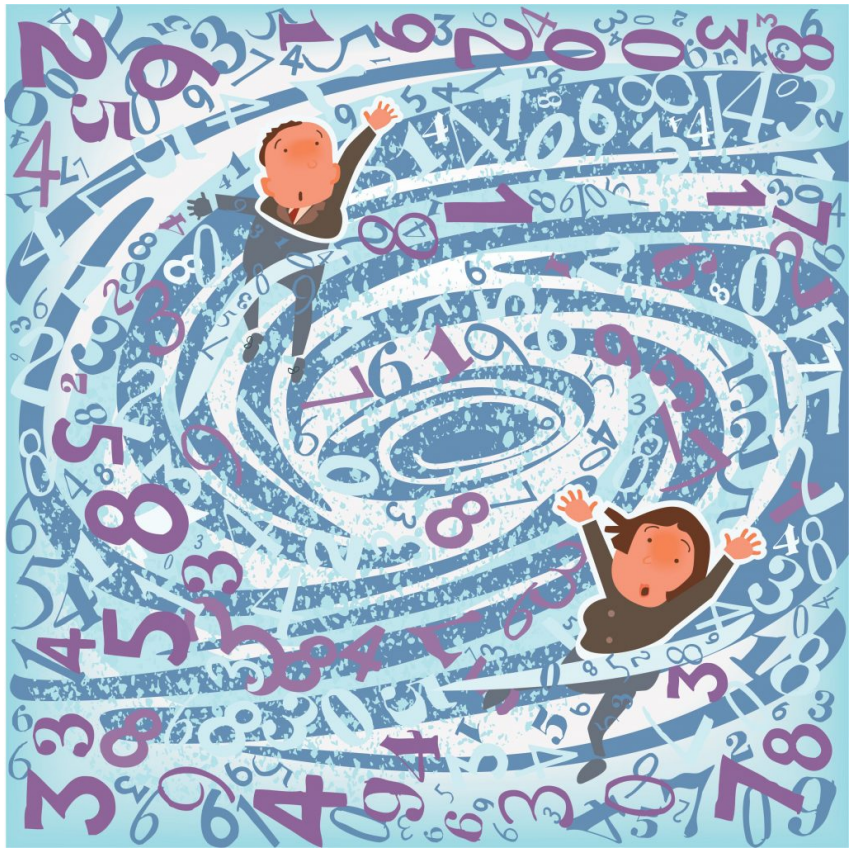
Who are we (1): The Statistics and Bioinformatics Unit



Who are we (2): Teachers



Why this course (1)



"We are drowning in information but starving for knowledge"

Why this course (2)

- **(Biomedical) research**, as well as many other human activities (social networks, sports, COVID . . .) generate **huge quantities** of -often complex- **data.**
  - Although sometimes we will also have small datasets
- We believe that data leads to information that leads to knowledge, but we need to be able to extract one from the other.
- This can be attempted in many ways, artificial intelligence, machine learning, data science or something which is common to all of them: **Plain Statistics**!!!

What are our goals (1)

- The main goal of this course is to introduce a variety of statistical methods and tools, which is good enough to:
  - Help you analyze your own data when it makes sense
  - Suggests you when the analysis is complex enough to contact an expert statistician (such us those in the UEB)
  - Help you to distiguish one from the other
- A secondary, but not least important objective: Show how to do it using R

Course contents

- This is a Standard course on Statistics using R as analysis tool
- Sessió 1. Introducció a R i RStudio i Tidyverse.
- Sessió 2. Estadística descriptiva I: Resums numèrics, taules i gràfics. Reproduïbilitat amb R

*Sessió 3. Estadística descriptiva II: Gràfics i taules bivariants. Visualització de dades amb ggplot2.

- Sessió 4. Més manegament de dades amb R. Automatització de tasques amb scripts..
- **Sessió 5. Sessió presencial de pràctiques en R i resolució de dubtes**
- Sessió 6. Introducció a la inferència. Intervals de Confiança.
- Sessió 7. Proves d'hipòtesis I: Conceptes bàsics
- Sessió 8. Proves d'hipòtesis II: Variables quantitatives.
- Sessió 9. Proves d'hipòtesis III: Taules de contingència, Khi^2, Tests diagnòstics: Sensibilitat, especificitat i corbes ROC.
- **Sessió 10. Sessió presencial de resolució d'exercici**

Why doing statistics with R

- R has become a "de facto" standard for statistical analysis
- R is free to use!!
- Practically all existing statistical methods available
- Powerful graphics that can be used interactively to explore data
- Possibility of scripting analysis $\longrightarrow$ **Reproducibility**
- Enormous and collaborative Community

How we will work

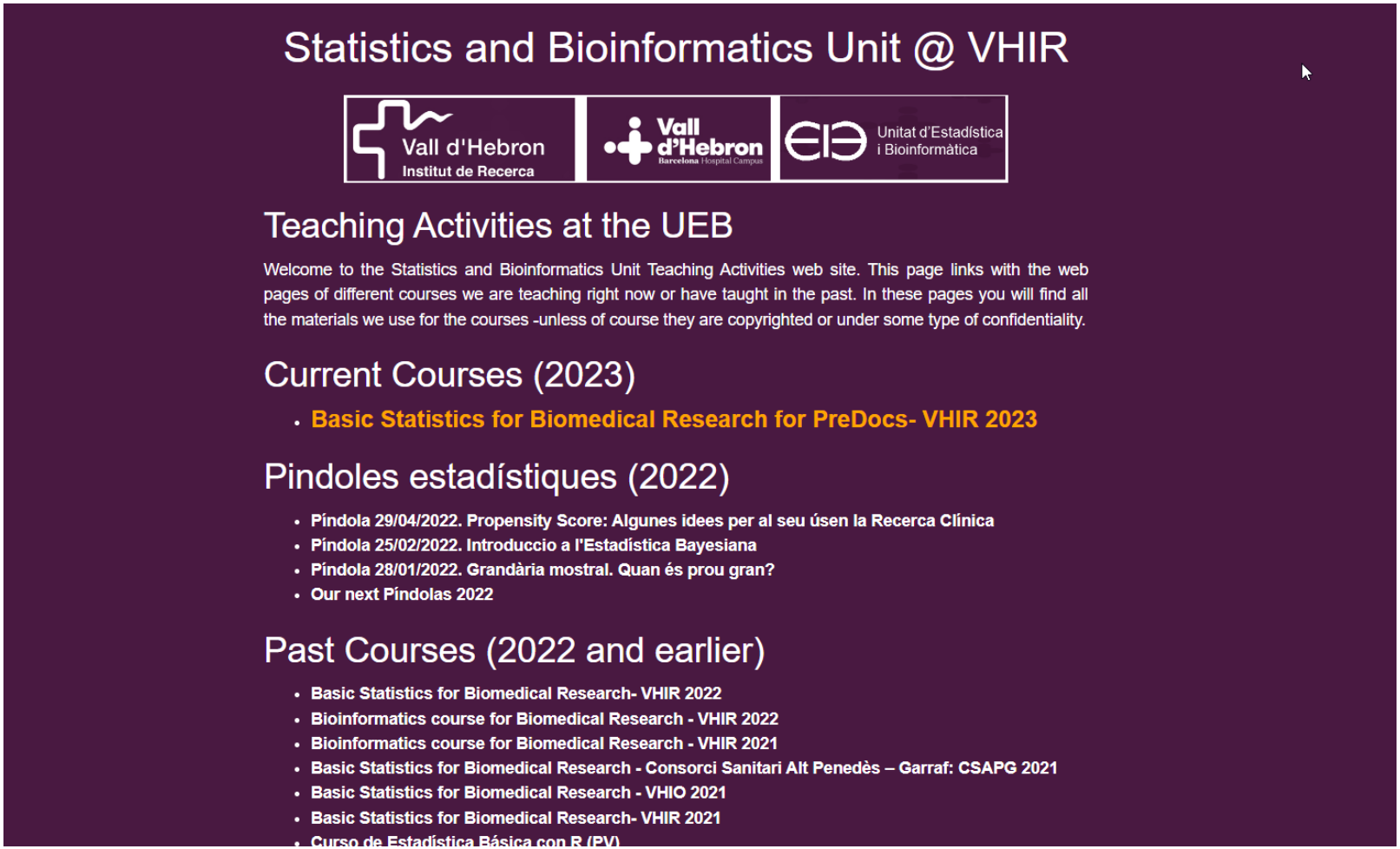
1. Our session will have the following structure (all but the first)
   - 1st part: We introduce a few new theoretical ideas
   - 2nd part: We wil show how to do it in R
     3rd part: Practice exercises and start working on the case study suggested or if you want in your data.
2. Mastering R requires as many other disciplines
   - Time
   - Study, and
   - Practice.

So take it easy. You can go back to R once the course finish

Class atendance and participation (class atendance should be at least 80%)

The materials

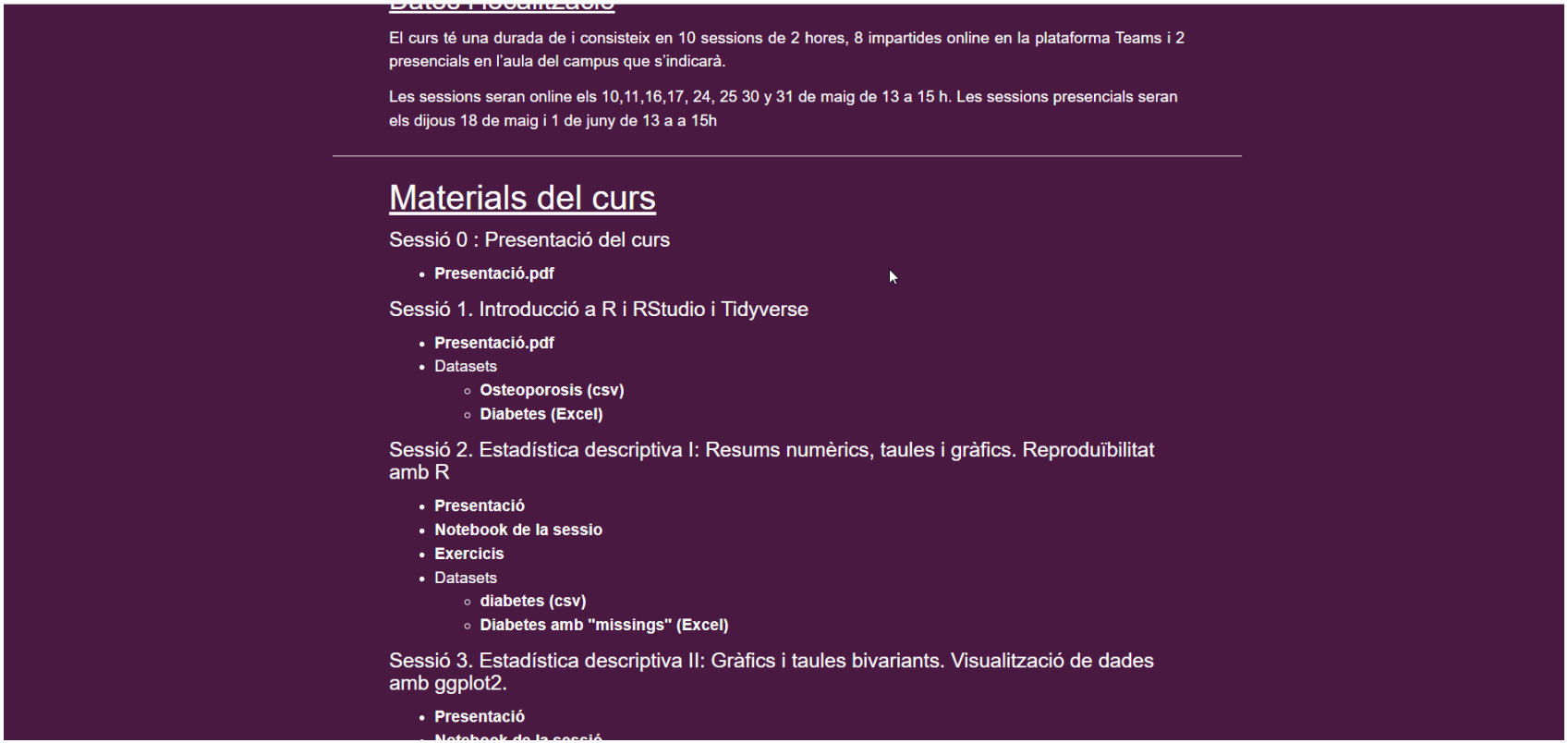
You can find the presentations, scripts , exercises in the teaching section of our web ueb.vhir.org



- Resources and references

There is a huge variety of resources to learn R, books, tutorials, free online courses, etc.

- This course is based on the book Data Science for R.
- Other interesting books + Using R and RStudio for Data Management, Statistical Analysis, and Graphics, 2nd edition
- Online courses + Coursera's Data Science Specialization
- A list of R tutorials and courses



El curs té una durada de i consisteix en 10 sessions de 2 hores, 8 impartides online en la plataforma Teams i 2 presencials en l'aula del campus que s'indicarà.

Les sessions seran online els 10,11,16,17, 24, 25 30 y 31 de maig de 13 a 15 h. Les sessions presencials seran els dijous 18 de maig i 1 de juny de 13 a a 15h

---

Materials del curs

Sessió 0 : Presentació del curs

- **Presentació.pdf**

Sessió 1. Introducció a R i RStudio i Tidyverse

- **Presentació.pdf**
- Datasets
  - **Osteoporosis (csv)**
  - **Diabetes (Excel)**

Sessió 2. Estadística descriptiva I: Resums numèrics, taules i gràfics. Reproduïbilitat amb R

- **Presentació**
- **Notebook de la sessio**
- **Exercicis**
- Datasets
  - **diabetes (csv)**
  - **Diabetes amb "missings" (Excel)**

Sessió 3. Estadística descriptiva II: Gràfics i taules bivariants. Visualització de dades amb ggplot2.

- **Presentació**