

# Basic Statistics for Biomedical Research

Alex Sanchez, Anna Sanjuan, Augusto Sao and  
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

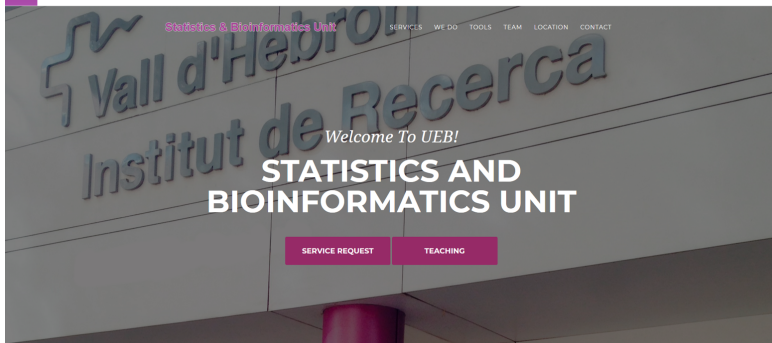
# Outline

- 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



ueb.vhir.org



# Who are we (2): Teachers

## WHO WE ARE

You can contact us in the following ways



**Alex Sánchez Pla**

Responsible of the Data Science Projects

Platform



**Santiago Pérez Hoyos**

Head of the Statistics and Bioinformatics

Unit



**Mireia Ferrer Almirall**

Responsible of the Bioinformatics Platform



**Miriam Mota-Foix**

Bioinformatics Technician



**Augusto Sao Aviles**

Statistics Technician



**Esther Camacho**

Bioinformatics Technician



**Anna San Juan**

EU-PRISM Project Manager



**Berta Miró**

VED Project Manager



**Angel Blanco**

ESPC-Life Project Manager



**Guillem Fernández**

Classical Data Specialist



# 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 as those in the UEB)
  - Help you to distinguish 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. Introducció a la inferència estadística. Intervals de confiança.
  - Sessió 6. Proves d'hipòtesis I: Conceptes bàsics
  - Sessió 7. Proves d'hipòtesis II: Variables quantitatives.
  - Sessió 8. Proves d'hipòtesis III: Variables qualitatives.
  - Sessió 9. Tests diagnòstics: Sensibilitat, especificitat i corbes ROC. Exercici de anàlisis de dades reals.
  - Sessió 10. I Exercici resum d'anàlisis de dades reals
- If there is anything else you would like to learn, let us know and we'll try not to let it out.

# 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 → **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 will 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 attendance and participation (class attendance should be at least 80%)

# The materials

You can find the presentations, scripts , exercises in the teaching section of our web [ueb.vhir.org](http://ueb.vhir.org)

## Statistics and Bioinformatics Unit @ VHIR



## Teaching Activities at the UEB

Welcome to the Statistics and Bioinformatics Unit Teaching Activities web site. This page links with the web pages of different courses we are teaching right now or have taught in the past. In these pages you will find all the materials we use for the courses -unless of course they are copyrighted or under some type of confidentiality.

## Current Courses (2021)

- **Basic Statistics for Biomedical Research - VHIO 2021**
- Basic Statistics for Biomedical Research- VHIR 2021
- Curso de Estadística Básica con R (PV)

## Past Courses (2020 and earlier)

- Basic Statistics for Biomedical Research - VHIO
- Curs de Bioinformàtica per a la Recerca Biomèdica
- Basic Statistics for Biomedical Research - VHIO
- Basic Statistics for Biomedical Research
- Sessió formativa Servei Otorrino
- Bioinformatics for clinical and biomedical research

## - Resources and references

- Course materials at:

[https://uebvhir.github.io/Course\\_Basic\\_Statistics\\_VHIR\\_2021.htm](https://uebvhir.github.io/Course_Basic_Statistics_VHIR_2021.htm)

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

# Have a good course !!!!



“Data don’t make any sense,  
we will have to resort to statistics.”



15/14