Data Management, Programming and Graphics with the R language

Alex Sanchez, Miriam Mota, Ricardo Gonzalo and Mireia Ferrer

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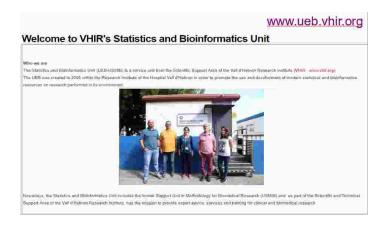
Readme

- 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.

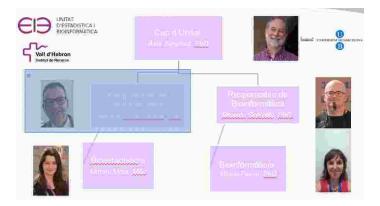Alex Sanchez, Miriam Mota, Ricardo Gonzalo and Mireia Ferre

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
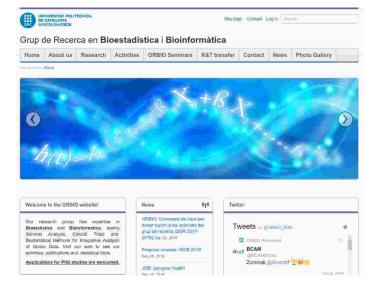- References & Resources

Who are we (1): The Statistics and Bioinformatics Unit
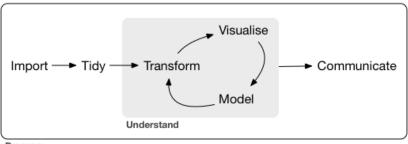
Who are we (2): Teachers

Who are we (3): The GRBio Research group

Why learn R

- Most people in most jobs have to *manage* information in their every day work.
- "Managing" may mean different things such as:
    - *retrieving*
    - *manipulating*
    - *visualizing*
    - *analyzing*
    - *reporting*
- R is a powerful tool that can be used to facilitate, improve or automate tasks such as those described above.

Hadley Wickam's approach to learning and applying Data Sciendce

Program

Your turn

- Provide examples of informations you may wish to manage
- Describe briefly
    - what this information is about
    - how it is stored
    - what you may wish to do with it
        - Transformations
        - Computations
        - Reports

How we will work

- Mastering R requires as many other disciplines

  - Time
  - Study, and
  - Practice.

- Our lectures will have the following structure (all but the first)

  - 1st part: Discuss the work you have done during the week
  - 2nd part: We introduce a few new ideas
  - 3rd part: Practice exercises and start working on the case study suggested/your data.

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