EEB R Bootcamp
Intro, basic stats, and plotting

Noa Pinter-Wollman

nmpinter@ucla.edu

Welcome!

- **Goals**: Introduce graduate students to the R environment and provide a foundation for skills in scientific programming.

- Course webpage with materials: http://michaelalfaro.github.io/eeb201/

- Syllabus:

https://docs.google.com/document/d/17qyGIk6CO1dmmKhIePUt6plxqmlwkvyhe6LoNopvjUQ/edit

Who R we?

Instructors

Noa Pinter-Wollman

Mike Alfaro

Kirk Lohmueller

Jamie Lloyd-Smith

Graduate students

Kenji Hayashi

Alayna Mead

Rachel Turba

Rachel Potter

Rosa McGuire

Lizzie Blackmore

Dylan Morris

Amandine Gamble (postdoc)

Learning objectives

By the end of the course students will:

- Be comfortable executing basic commands in the R environment.
- Be able to load packages in R and make use of their added functionality.
- Be able to read in data files, manipulate data, and perform simple analyses in R.
- Be capable of plotting curves, scatter-plots, histograms, and other graphic outputs in R.
- Be able to write their own computer programs to simulate population models in discrete or continuous time.

What is R?

- "R is a free software environment for statistical computing and graphics."
- A free, open-source, community-built software package.
- Powerful, flexible computational tool with a broad array of add-on 'packages' that perform advanced analyses.
- A high-level programming language for scientific computing.
- A powerful platform for statistics.
- A major research tool in ecology and other fields.

What is R not?

- A commercial package

- A drop-down menu environment

- Excel

- JMP, SPSS, SAS, etc.

- Matlab

Why R?

- Free
- Widely used in ecology and evolution
- Helpful, active, and large user community
- Lots of resources (formal packages and informal community)
- Versatile
- Flexible
- Data manipulation capabilities
- Reproducible
- Scalable
- Transferable programming skills

R Bootcamp schedule

- Tuesday September 14th
  - 9-10:30 intro to the bootcamp and Basic stats and plotting – Pinter-Wollman
  - 10:30-12 Flow control, scripts, functions –Alfaro
  - 12-1:00 Lunch
  - 1:00-3:00PM Work on assignments

- Friday September 17th
  - 9-10:30 Advanced plotting - Lohmueller
  - 10:30-12 Working with data and population dynamics - Lloyd Smith
  - 12:30-1:00 Lunch
  - 1:00-3:00PM Work on assignments

What will you be learning?

- Command line, scripts, and functions
- Getting help in R
- Packages/libraries
- Reading data
- Variables
- Subsetting and indexing specific data
- Control flow (for, if, while…)
- Functions
- Plotting
- Basic modelling

R resources

- https://cran.r-project.org/
- R studio (https://www.rstudio.com/)
- R for biologists (https://cran.r-project.org/doc/contrib/Martinez-RforBiologistv1.1.pdf)
- Springer series 'Use R!' is available free through UCLA library
- R for Data Science http://r4ds.had.co.nz/
- Quick R: http://www.statmethods.net/
- R reference card: ftp://cran.r-project.org/pub/R/doc/contrib/Short-refcard.pdf
- R colors: http://research.stowers-institute.org/efg/R/Color/Chart/ColorChart.pdf
- Other grad students (Hacky hour)
- Stats consulting on campus
- Collaboratory and IDRE workshops
- Google!

What is the plan for the next 1.5 hours?

- Running statistical tests
- Plotting

Basic statistics

- T-tests
- ANOVA
- Correlation
- Other linear models
- More advances statistics – EEB202C

Some advice about script layout

1. Setup (e.g., clean workspace, load packages)

    ```
    # set up
    rm(list=ls()) # clean workspace
    graphics.off() #close all figures

    #load packages
    library('igraph')
    ```

2. Load data and prepare it for analysis (e.g., subset, assign data type...)

3. Perform analysis

4. Display results (plot, save output etc...)

Dataset to work on today

library(palmerpenguins)

There are two data sets in the package, we will use 'penguins' which is the simplified data frame of the raw data (which is also available for those into penguins...)

Bill length

Bill depth

Flipper length and Body mass

CHINSTRAP!

GENTOO!

ADÉLIE!
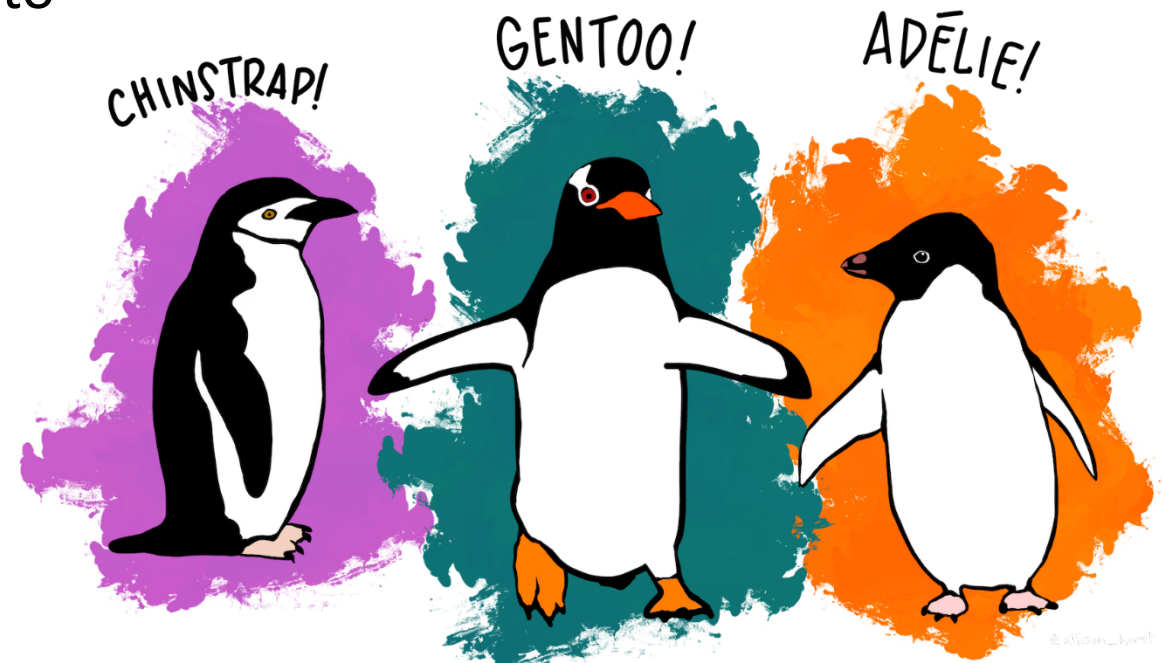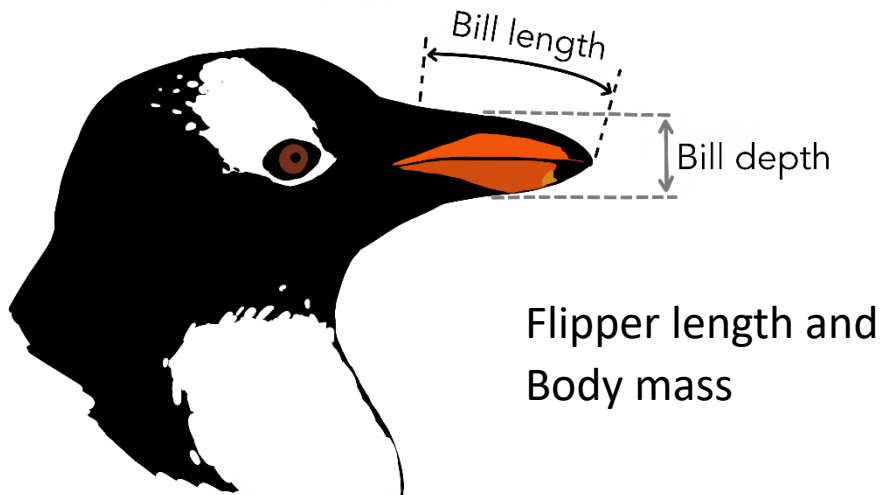
Advanced plotting

- Color palettes
- Heatmaps
- The R graph gallery:

http://www.r-graph-gallery.com/all-graphs/

- ggplot2