Data lab 1

Ragnhild Thorseth Grevskott

Starting 12:15

If you have had trouble downloading R, please let me know ☺

Me

Ragnhild Thorseth Grevskott (Ragnhild.grevskott@usn.no,)

- Ecologist: behavioural ecology & botany

- NOT statistician, got some basic training
  (less than Jan with statistics but more with Rstudio)

- I can answer *some* of your questions!
  - We can also ask Jan or others at the institute

The data labs

- Repetition of lectures
- Get familiar with R software
- Learn how to do statistics in R(studio)
- Some hand calculations of statistics?

The lab manuals: Goals

1. Introduction to statistics on the computer and graphics

Goals

- Get started on the computers, learning how to start with R

- Collect a data set on ourselves for future use

- Make graphs, such as histograms, bar charts, box plots, scatter plots, and dot plots.

- Learn to graph data as the first step in data analysis

The lab manuals: Summary

Quick summary from text (see Chapter 2 in Whitlock and Schluter)

- Computers make data analysis faster and easier. However, it is still the human's job to choose the right procedures.

- Variables are either numerical (measured as numbers) or categorical (describing which category an individual belongs to).

- Graphing data is an essential step in data analysis and presentation. The human mind receives information much better visually than verbally or mathematically.

- The distribution of categorical variables can be presented in a bar chart. The distribution of numerical variables can be presented in a histogram, a box plot, or cumulative frequency plot.

- The relationship between two numerical variables can be shown in a *scatter plot*. The relationship between two categorical variables can be shown in a *grouped bar chart* or a *mosaic plot*. The association between a numerical variable and a categorical variable can be shown with *multiple histograms, grouped cumulative frequency plots*, or *multiple box plots*.

- A good graph should be honest and easy to interpret, with as much information as needed to interpret the graph readily available. At the same time, the graph should be uncluttered and clear.

The lab manuals: Learning the tools

- We go through this together:
  - The functions you will need to work with today's topic

Learning the Tools

Feel free to skip to section 2 if you have used R before.

1. Downloading the right software

a. Download *R* from http://www.r-project.org/. Click on *download R* in the top left corner and choose your preferred mirror (e.g., from the University of Bergen, https://cran.uib.no/). Download the correct version for your computer.

b. Download *RStudio* from https://www.rstudio.com/products/rstudio/. You want to choose the free *RStudio Desktop Open Source Edition*. Download the correct version for your computer.
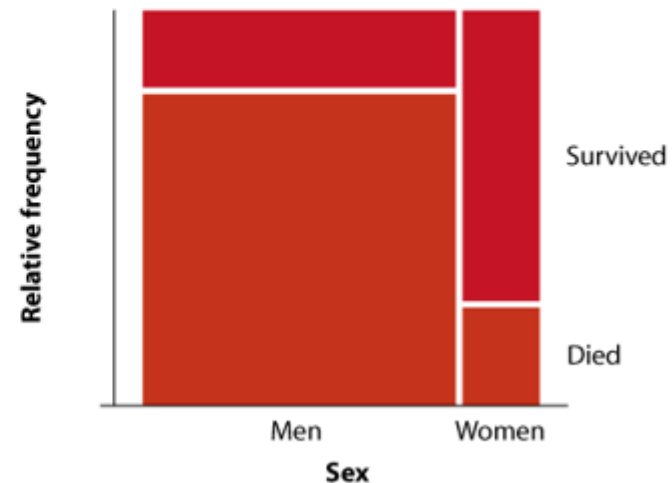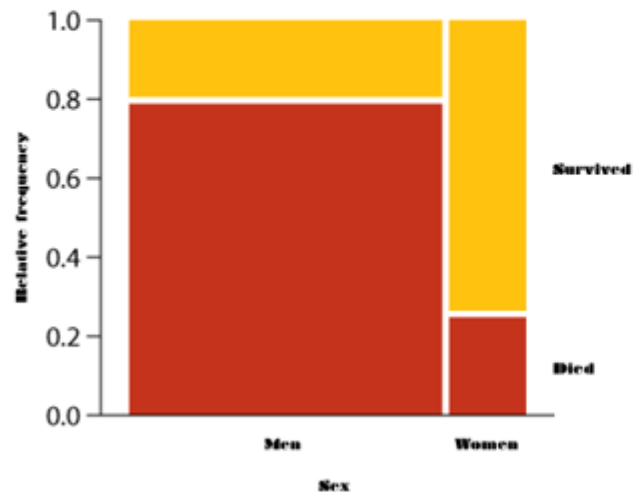
The lab manuals: Questions

- This lab: seven questions

Questions

1. For each of the following pairs of graphs, identify features that are better on one version that the other.

a. Survivorship as a function of sex for passengers of the *RMS Titanic*

Today

- We go through the whole «learning the tools» part together
- We will have some time for helping people getting along (but of course not endless

R

- Hope you managed to download R?

- R is a really good (and widely used!) tool for doing data cleaning, statistics, visualising etc.
  - BUT: For many of you: new software?
    - need to learn new 'language' etc.

Excel

- Forgot to mention that you will also need Excel
  - Just assumed you already have it

- Download from USN server (for free!)

Student data set

- https://forms.gle/rr7wWDoUxbviWz6s8
- Fill out the form

- I will upload the dataset to canvas
  - Then you can use it in R