

R-Basics

Session 3: Statistical Calculations and Graphics with R

Dr. Ulrike Göbel

2024-05-22

Where We Left Off and Where We Go

- We have made a tour of R's capabilities as a programming language, but **so far we didn't explore applications to real tasks**.
- R has developed out of the S language, which was originally a dedicated **statistical programming** language. It has inherited a rich assortment of statistical tests. I am not an expert in this area, but I will comment on some examples given [here](#).
- These examples also use another strength of R: **graphical representation**. R's **base package** of built-in functions includes well developed plotting capabilities, however these are increasingly superseded by the [ggplot2 package](#), which is part of the [tidyverse package ecosystem](#). We will explore how the same dataset can be represented using base plots and the ggplot -- where ggplot is much richer, but has also a much steeper learning curve.
- Finally I will introduce the [Bioconductor package ecosystem](#) of R packages. Its DESeq2 package is the de-facto standard for bulk RNAseq analysis in R. For single-cell analysis, the [Seurat package\(/ecosystem\)](#) may be more popular than [Bioconductor's OSCA sub-system](#) for this task.

Statistical Analysis with R and the Tidyverse

[Here](#) is a complete little statistical analysis workflow, which we will follow directly on their website.

Their analyses revolve around the question:

Do managers loose more time with multi-tasking than senior individual contributors (IC)?

When you copy R code from the website, be aware of what you are doing and **see whether you can recognize syntax patterns!**

Statistical Analysis with R and the Tidyverse

Before we start, you need to install the **vivainsights** and probably the **tidyverse** packages:

```
install.packages(c("vivainsights","tidyverse"),  
                 repos = "http://cran.us.r-project.org")  
  
library(vivainsights,quietly=TRUE)  
library(tidyverse,quietly=TRUE)
```

Load and examine the table data to be explored:

```
data("pq_data",package = "vivainsights") # download the data  
View(pq_data) # have a look -- it is a HUGE table!
```

The Bioconductor Package Ecosystem

Here again we will directly explore the [Bioconductor website](#).