Conda environments

Cedric Arisdakessian

2022-08-10

What is a computing environment?

- The operating system: Example: Windows, Linux, MacOS, etc.
- The system libraries (binaries). Examples: Lapack, libpng, libcrypto, and many others
- The tool's library (e.g. ggplot2, phyloseq)
- The variables in your R session
- The existing bash variables when a program is launched (e.g. RSTUDIO_WHICH_R)

Why do we need to control the environment?

After all, if my program works, who cares?

- Some packages can have conflicting requirements. Let's say your current installed packages require R>=4.0. A new package you want to install requires R==3.6. There are multiple possible scenarios:
  - Maybe R will prevent you to install one of the package
  - Maybe R will still install the package and accept that one of the packages won't have all the requirements met. You run the risk of having some impredictible bugs in your code.
- Reproducibility to get consistent results and facilitate troubleshooting
- Portability: when we share code, we see directly what packages are required. So it's easier to setup on a new machine.
- Easy to install and delete new tools (conda, docker) because popular packages and tools are already documented

Existing tools

Virtual environments:

- anaconda/miniconda/mamba
- virtualenv

Virtualization:

- Virtual machines (e.g. VirtualBox)
- Containerization: docker/singularity/podman/kubernetes. . .

Miniconda

Go to: https://docs.conda.io/en/latest/miniconda.html Download miniconda3 for your operating system and run the installer.