Computational Environment Management

Recording Your Software Environment

- Software packages have dependencies
- Software packages have versions
- Need to:
  - Make software runnable (i.e. install it)
  - Describe software that is installed
- Ideally, do these things in a way that others can execute to replicate

Suggested Tools

- **miniconda** - install non-python packages
- **pip** - install python packages
- **R package** install script
- **bash scripts** - specify and record packages
- **git** - track scripts and files
- **github/bitbucket** - make git repo public

miniconda

- https://conda.io/miniconda.html
- Install miniconda3 (don't use miniconda2!)
- `conda create -n name_of_env [<pkg>...]`
- `conda activate name_of_env`
- `conda search -c <channel> <pattern>`
- `conda install -c <channel> <pkg>`
- e.g.

  `conda install -c bioconda star=`
- `conda list --export > conda_packages.txt`

conda demonstration

pip

- Command to install python packages
- Packages listed on pypi.python.org
- More reliable than installing through conda
- `pip install <package>`
- e.g. `pip install pandas`
- Put packages in `requirements.txt`

pip demonstration

R install

- R notorious for install/version conflict issues
- Can install R using conda:
  - `conda install -c conda-forge r-base=3.6.2`
  - Bioconductor packages:

    ```
    conda install -c bioconda
    bioconductor-deseq2=1.26.0
    ```
- **Making your R code reproducible is a major pain! Do your best, it's worth it.**

R package install

- When you finally have R installed…
- Write an R script that contains install calls
  - install.package() calls
  - Bioconductor packages, e.g.

```
source("http://bioconductor.org/biocLite.R")
biocLite("DESeq2")
```

- E.g. `install_r_packages.R`

R install script demonstration

Master install script

- E.g. `install_packages.sh`
- bash script that contains software install command(s):
  - conda install commands
  - pip requirements.txt file
  - R script containing package install
  - conda list output to file

Master install script demonstration

Record your work with git

- `git` tracks changes to files repositories
- We have been recording software environment in files!
- Repos can be cloned - every clone has the complete change history
- Every change to a file is reversible
- Very few commands to learn, huge benefits!

git demonstration

github / bitbucket

- Public web platforms for sharing code
- Can be made private
- Easy to share and collaborate
- Code backup mechanism
- git isn’t efficient at storing big (binary) files
- **Only code! Not for data!**

bitbucket demonstration

Summary & Recommendations

- Use:
  - `conda install` for non-python packages
  - pip for python packages with requirements.txt
  - R script for R packages
  - `conda list --export` to record pkgs to `conda_packages.txt`
- Create bash script that calls all of these at once, e.g. install_packages.sh
- Check them all into a git repo
- Push to github/bitbucket
- Profit