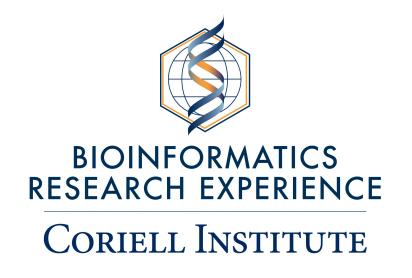


## BIOINFORMATICS RESEARCH EXPERIENCE

CORIELL INSTITUTE

### Introduction to Coriell

Introduction to the Bioinformatics
Research Experience



### Introductions

Two Truths and a Lie? Or personal bingo??

R and the Tidyverse



#### R Programming Language

- Open-source scripting language developed for statistical analysis
- Was originally known as S
  - S was developed in 1975
  - S was reimplemented as R in 1993
- Heavily used in bioinformatics, with our own archive Bioconductor, for biology-related packages



#### The Tidyverse

"The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures."

- Keep everything simple
  - Use existing data structures instead of custom, aka, use tidy data
  - Functions should do one thing well
- Glue the simple things together; simple things put together are more powerful than one complex thing
- Design for humans







# Rmarkdown and Rstudio

#### What is markdown (and Rmarkdown)?

- Markdown is just text, with a few optional symbols that allows a markdown interpreter to make it look good. Goal is to have something that still is human readable even without the interpreter.
- Rmarkdown is markdown for R
  - All the features of markdown, with extras
  - Intention is to make documenting data analysis easy
  - Execute code in Rmarkdown files (cannot do this in markdown)
  - Can also knit Rmarkdown files into other files
    - html, pdf, or Microsoft word reports
    - Can make websites and slides with Rmarkdown

#### Why use markdown and Rmarkdown?

Bioinformatics is mostly on a Linux machine using the command line terminal. The rest of the universe uses Macs or PCs.

#### This causes a bunch of problems

- operating systems don't talk to each other easily
- files are in proprietary formats (no Microsoft anything on Linux)
- files are not readable in plain text (and terminal needs them to be!)

#### Markdown solves all the problems

- Simple
- readable by every machine in both GUI and terminal AND humans
- allows some simple formatting to increase human readability

# Rstudio and Rmarkdown DEMO