

# Let's Begin!

Cell Discovery Network

# Topics Covered in the next 45 minutes

## **Intro to R**

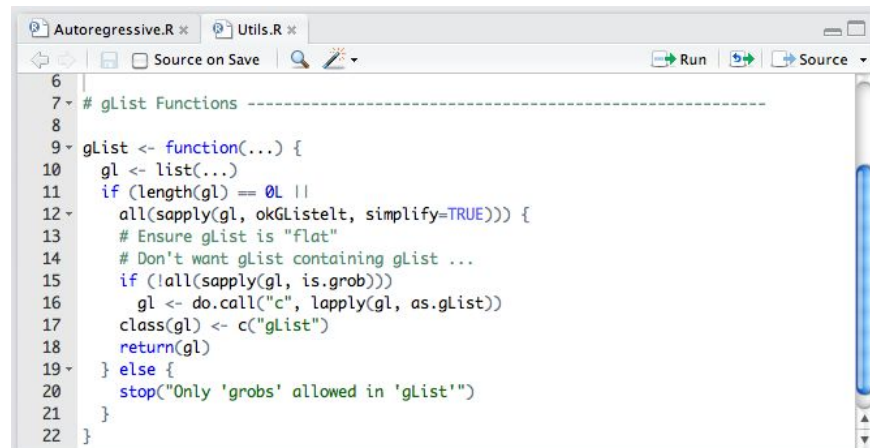
1. R Basics
2. GitHub (Downloading code)
3. File Types
4. How to use RStudio

## **R and Data Analysis**

1. Data frame basics
2. Seurat R package
3. Seurat object

# R Basics

- R is a programming language for **statistical analysis and visualization**.
- R is not a general programming language like Python or Java but its functions and use are domain-specific (statistical computing).
- Popular tool for data analysts in all fields.
- Users code through editors like **RStudio** or **VSCode**.
- R is known for its multitude of extension packages:
  - Popular packages in bioinformatics include: *Seurat*, *Bioconductor*, *tidyverse*.

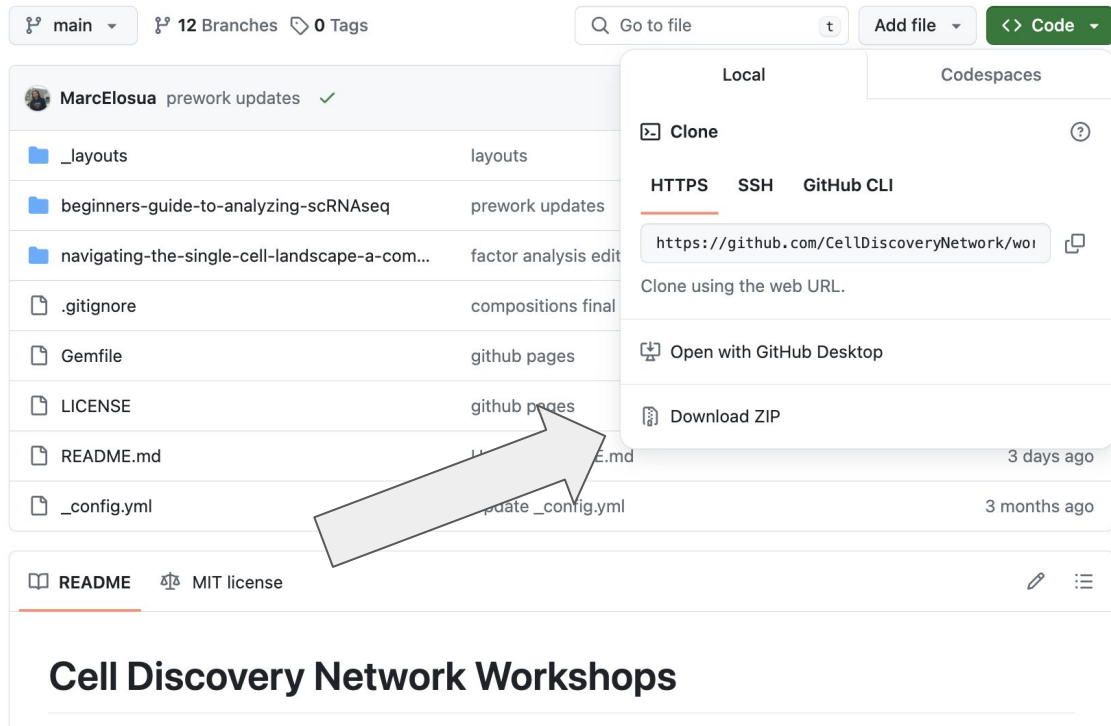


```
Autoregressive.R x  Utils.R x
Source on Save  Run  Source
6
7 # gList Functions -----
8
9 gList <- function(...) {
10   gl <- list(...)
11   if (length(gl) == 0L ||
12       all(sapply(gl, okGListelt, simplify=TRUE))) {
13     # Ensure gList is "flat"
14     # Don't want gList containing gList ...
15     if (!all(sapply(gl, is.grob)))
16       gl <- do.call("c", lapply(gl, as.gList))
17     class(gl) <- c("gList")
18     return(gl)
19   } else {
20     stop("Only 'grobs' allowed in 'gList'")
21   }
22 }
```

# GitHub

Download workshop code from GitHub by:

1. Download the entire repository zip file from GitHub.
2. Unzip the files.
3. Move them to the correct location on your device.



# File Types: R vs Rmd vs Qmd

R File - Standard R script file that execute line-by-line or sequentially. Best for running within pipelines where dynamic visual output is not necessary. Often scripts are used to store functions and load them to your environment.

Rmd File - R Markdown files are dynamic documents that allow for text, code, and a graphics to be visualized in one space.

Qmd File - Quarto Markdown files are a newer and more capable extension of R markdown that support multiple programming languages.

# File Types: csv, tsv, RDS, h5

csv - Comma-Separated Values

- Used to store count matrices or metadata.

tsv - Tab-Separated Values

- Used for metadata when text data contains commas.

RDS - R Data Serialization

- Store single R objects in binary format such as Seurat or SingleCellExperiment objects.

h5 - Hierarchical Data Format version 5

- Stores large and complex data in an efficient way.
  - h5ad - Scanpy specific way of storing AnnData objects.

*Next: Navigate to **prework** directory.*

# How to Use RStudio

Let's Tour RStudio:

- File ⇒ New File or New Project
- Review Console & Terminal
- Environment & History
- Files, Plots, Packages, Help
  - *?function\_name* functionality

# Using R Projects

1. File ⇒ New R Project
  - a. An R project contains all of the package versions and environment variables of a specified workspace.
2. Existing Directory
3. Set present working directory to *beginners-workshop*

If you do not feel comfortable coding along with us throughout the day or if you hit a roadblock or burnout, find the pre-run code in HTML files in the repository we just downloaded.



# R and Data Analysis



# Data Frame Basics

*Walk through data-wrangle-viz.R*

# Seurat R package

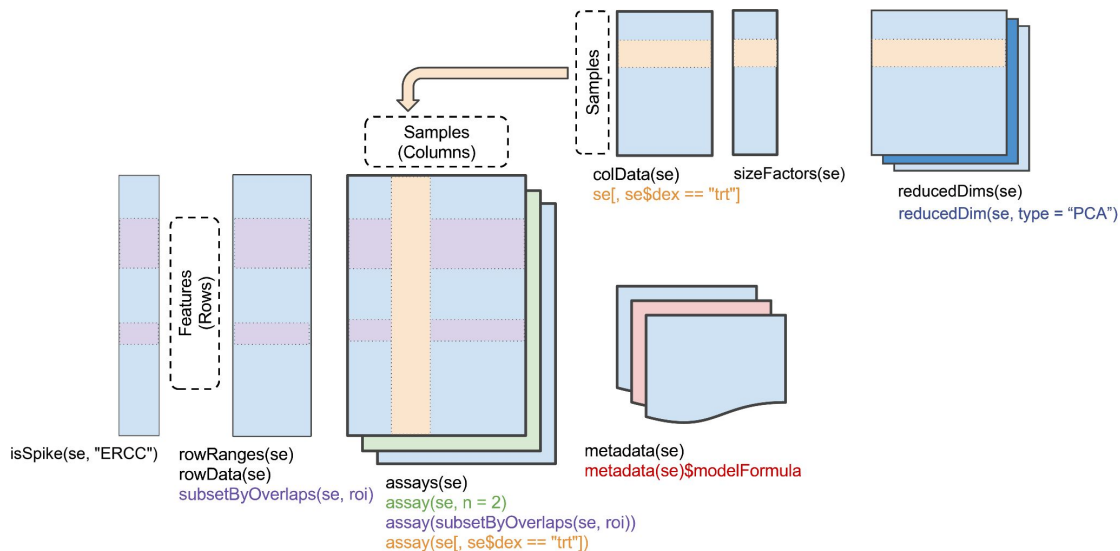
*Seurat is an R package designed for QC, analysis, and exploration of single-cell RNA-seq data.*

*Seurat aims to enable users to identify and interpret sources of heterogeneity from single-cell transcriptomic measurements, and to integrate diverse types of single-cell data.*

	Cell1	Cell2	...	CellN
Gene1	3	2	.	13
Gene2	2	3	.	1
Gene3	1	14	.	18
...	.	.	.	.
...	.	.	.	.
...	.	.	.	.
GeneM	25	0	.	0

# Seurat Object

	Cell1	Cell2	...	CellN
Gene1	3	2	.	13
Gene2	2	3	.	1
Gene3	1	14	.	18
...	.	.	.	.
...	.	.	.	.
...	.	.	.	.
GeneM	25	0	.	0



# Seurat Object

An object of class Seurat  
33234 features across 59572 samples within 1 assay  
Active assay: RNA (33234 features, 0 variable features)  
1 layer present: data  
1 dimensional reduction calculated: tsne

Name	Type	Value
se	S4 [33234 x 59572] (SeuratObject::Seurat)	S4 object of class Seurat
assays	list [1]	List of length 1
meta.data	list [59572 x 46] (S3: data.frame)	A data.frame with 59572 rows and 46 columns
active.assay	character [1]	'RNA'
active.ident	factor	Factor with 1 level: "local"
graphs	list [0]	List of length 0
neighbors	list [0]	List of length 0
reductions	list [1]	List of length 1
images	list [0]	List of length 0
project.name	character [1]	'local'
misc	list [6]	List of length 6
version	list [1] (S3: package_version, numeric_version)	List of length 1
commands	list [0]	List of length 0
tools	list [0]	List of length 0

The image features a dark blue background with a diagonal white line running from the bottom-left corner towards the top-right corner, creating a split-screen effect.

**Thank You!**