R: installation, variables and data structure

Most examples are from

https://www.youtube.com/watch?v=_V8eKsto3Ug

&

R help documents

&

https://www.benjaminbell.co.uk/2018/03/quick-guide-to-annotating-plots-in-r.html

Why R?

- The most widely used one for Data Science, 50% more than Python
- Tools are free
- Many packages are developed

Download and installation

- https://www.r-project.org/
- Rstudio

Useful packages

- CRAN, Crantastic, or Github.com/trending/r
- 21938 available packages in CRAN
- Type help("INSTALL") or help("install.packages") in R for information on how to install packages from CRAN
- For example, in R type install.packages(“package name”) to install the specific package with the specified name.

Topics

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
Databases	Databases with R
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
ExtremeValue	Extreme Value Analysis
Finance	Empirical Finance
FunctionalData	Functional Data Analysis
Genetics	Statistical Genetics
GraphicalModels	Graphical Models
HighPerformanceComputing	High-Performance and Parallel Computing with R
Hydrology	Hydrological Data and Modeling
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
MetaAnalysis	Meta-Analysis
MissingData	Missing Data
ModelDeployment	Model Deployment with R
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
NumericalMathematics	Numerical Mathematics

CRAN package task view

Https://mirror.las.iastate.edu/CRAN/

An example of package installation

- install.packages("pacman") #install
- library(pacman) #load
- pacman::p_load(dplyr,psych,ggplot2, rio) #rio for input/ouptu from files

Variable types

- Type: numeric (integer, single, double), logical, character, complex, (raw)
- Structure: vector, matrix/array, data frame, list
- Vector, 1-d, the same type of data
- Matrix, 2-d, the same type of data, the same length, use the index number to work with the columns and rows.
- Array, similar to matrix, >=3 dimensions
- Data frame, 2-d vectors of multiple types, the same length, analogue to spreadsheet
- List, 1-d, ordered collection of any type of elements including lists

Variable names

- Variable names can be a combination of letters, digits, period, and underscore.
- Underscore and digits cannot be the first one while others can in a variable name.
- Period must be followed by letter.
- Case sensitive

Variable type examples

- n1<-15
- typeof(n1)
- X<-TRUE, x<-F
- Y<-”this is a string or character”
- typeof(Y)
- v1<-c(1,2,3,4,5,6)
- is.vector(v1)
- M1<-matrix(v1, nrow=2, byrow=T)
- A1<-array(c(1:24), c(4,3,2))

list

- O1<-c(1,2,3)
- O2<-c(T, F,T, T, F)
- O3<-c(2,3,4,5)
- List1<-list(O1,O2, O3)
- List1
- List2<-c(O1,O2,O3, List1)
- List2

Convert data types 1

- vN<-c(1,2,3)
- vC<-c(“a”,”b”,”c”)
- vL<-c(T,F, T)
- Df1<-cbind(vN,vC,vL)
- Df1
- Df2<-as.data.frame(Df1)
- as.numeric
- as.character

Convert data types 2

- x<-1:3
- y<-1:9
- df1<-cbind (x,y)
- typeof(df1)
- typeof(df1$x)
- df1<-cbind.data.frame(x,y)
- typeof(df1$x)
- str(df1)
- typeof(df1$x)

In-class question #3

Open an account at galaxy, (1) download a ChIP-seq dataset, (2) run trimmomatic to clean the downloaded dataset, (3). Map the clean paired reads to the reference genome with Bowtie2. Submit a picture showing your log history of the above three steps at galaxy through the webcourses. Due 6pm on January 30.