

DATA ANALYSIS FOR MOLECULAR BIOLOGY AND BIOCHEMISTRY



#Bioinformatics

#Genomics

#DryLab

#Techniques

Introductory data analysis focusing on molecular biology data sets and examples and including basic programming skills using Python and basic statistics skills using R. Prerequisite: MATH 12 or equivalent is recommended. Students with credit for MBB 243 may not take this course for further credit. CMPT 120 will be accepted in lieu of MBB 110.

TOPICS

- Flavours of data in molecular biology and biochemistry
- Genomic data
- Fundamentals of R and Python
- Regular expressions and patterns
- Quantitative DNA/RNA sequence analysis
- Exploratory data analysis in the Tidyverse
- Generic visualization methods using ggplot2
- Advanced visualizations for molecular biology

INSTRUCTORS

Sophie Sneddon



Bioinformatician, video game enthusiast, and cat person

Ryan Morin



Just a regular everyday normal guy



JULIA EVANS @bork

sort & uniq

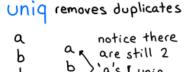
16

sort sorts
its input
\$ sort names.txt

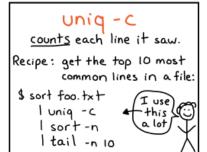
the default sort is alphabetical.

	Sor	t -n	
	numer	ric sort	
'sort'	order	'sort -n'	or

sort -h: human sort			
'sort -n' order 'sort -h' order			
156	¦ 45 K		
11 30 M	30M U		
11 30 M 11 45 K	156 5		
2006	2006		
useful example: du-sh * sort-h			



b a are still 2
b > b 'a's! uniq
a c only uniquifies
c adjacent
matching lines



Group Data

dplyr::group_by(iris, Species)

Group data into rows with the same value of Species.

dplyr::ungroup(iris)

Remove grouping information from data frame.

iris %>% group_by(Species) %>% summarise(...)
Compute separate summary row for each group.

