

STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI – 86
(For candidates admitted from the academic year 2023 – 2024)

M.Sc. DEGREE EXAMINATION, APRIL 2024
BRANCH - BIOINFORMATICS
SECOND SEMESTER

COURSE : MAJOR CORE
PAPER : PYTHON AND R PROGRAMMING
SUBJECT CODE : 23BI/PC/PR24
TIME : 3 HOURS **MAX. MARKS: 100**

Q. No.	SECTION A (10 x 1=10 marks)	CO	KL
	All questions to be answered (Objective type)		
1	x <- c("boy", "girl", 100,"school") typeof(x) a) character b) string c) numeric d) char	CO1	K1
2	a<-1:4 will print ____ a) 0,1,2,3 b) 0,1,2,3,4 c) 1,2,3 d) 1,2,3,4	CO2	K2
3	True or False: To print the first 5 rows head() function is used.	CO1	K1
4	True or False: In R, lists can contain elements of same type only.	CO2	K2
5	What does the function ungap do?	CO1	K1
6	Give the output : rep(2:4, times=4)	CO2	K2
7	What is the output in Python? dnaSuite = 'cctttacttcgcctccgcgcctgcattccgttctctggcctcg' print(dna[0:6])	CO1	K1
8	In Python, Lists are in one way different from arrays. What is that?	CO2	K2
9	y <- c(1L,2L,3L,4L) typeof(y)	CO1	K1
10	Write a Python code to of 35, 56 find average and 8.	CO2	K2
Q. No.	SECTION B (10 x 2= 20 marks)	CO	KL
	Answers in about 50 words		
11	What is the purpose of using Jupyter notebooks in Python programming?	CO3	K3
12	What is Object-Oriented Programming (OOP) in Python?	CO4	K4
13	List two major uses of the Biopython package.	CO3	K3
14	Name two Python libraries used for data visualization.	CO4	K4
15	How does Matplotlib help in data visualization?	CO3	K3
16	What is ggplot in R and why is it used?	CO4	K4
17	Explain the significance of dataframes in both Python and R.	CO3	K3
18	What role does scikit-learn play in Python programming?	CO4	K4
19	What is the function of BioML(R) in biomedical data science?	CO3	K3
20	Explain the term 'protein-protein interaction graphs' and their importance.	CO4	K4

Q. No.	SECTION C (4 x 10= 40 marks)	CO	KL
	Answer in about 600 words - Internal choice		
21	a) Discuss the importance of installing Python and Jupyter notebooks for beginners in bioinformatics programming. OR b) Describe the process of parsing DNA and protein information from FASTA files using Python.	CO4	K4
22	a) Outline the steps for extracting and annotating gene information using the Entrez module in Biopython. OR b) Explain the biomedical datascience in R with tidyverse and shiny.	CO5	K5
23	a) Explain how Python's Pandas library is utilized in handling genomic data and why it is preferred. OR b) Discuss the advantages of using matplotlib and scikit-learn for data visualization in genomics research.	CO4	K4
24	a) Compare and contrast the use of vectors, matrices, and dataframes in R for managing genomic data. OR b) Explain how Bioconductor contributes to genomic data analysis and mention two specific packages that are widely used.	CO5	K5
Q. No.	SECTION D (2x 15=30 marks)	CO	KL
	Answer any TWO questions in about 1200 words		
25	Discuss the importance of programming in biology and how Python and R facilitate analysis of biological data with relevant examples.	CO5	K6
26	Summarize the steps and significance of data normalization, discretization, and sampling in R using examples.	CO5	K6
27	Describe using Biopython and Bioconductor for analyzing biological data from data acquisition to visualization.	CO5	K6
28	Compare Python and R for bioinformatics, focusing on their strengths, weaknesses, and preferred usage scenarios.	CO5	K6
