

TRGN 527: Applied Data Science and Bioinformatics

UNIT I. Introduction and Basic Data Science

Week 1 – Assignment 1

Enrique I. Velazquez Villarreal, M.D., Ph.D., M.P.H., M.S. | Assistant Professor

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Leader of the USC Bioinformatics Core – *USC CaRE2 Health Equity Center*

David W. Craig, Ph.D. | Professor and Vice Chair

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Co-Director, Institute of Translational Genomics

Assignment 1

- Instructions: Please answer the following questions, each one worth 20 points (Total: 200 points). Please first generate an R Markdown answering the questions and then go to Blackboard/Week 1/ Assignment 1 to answer these questions accordingly. If you have any questions please don't hesitate to ask to your instructor.
- For the following five questions please load the file “trgn599.rna.tsv” in your R-Studio.
- Question 1
 - What type of data contains this file:
 - A) Gene expression data
 - B) Clinical data
 - C) Surgical data
 - D) Demographic data
- Question 2
 - How many rows (without including the header) has this file:
 - A) 24621 rows
 - B) 24620 rows
 - C) 24630 rows
 - D) 24640 rows

Assignment 1

- Using the the file “trgn599.rna.tsv” from your R-Studio.
- Question 3
 - How many columns (without including the first column) has this file:
 - A) 466 columns
 - B) 470 columns
 - C) 465 columns
 - D) 485 columns
- Question 4
 - The number of rows (without including the informational header) in this file refers to:
 - A) Number of genes
 - B) Number of samples
 - C) Number of clinical variables
 - D) Number of variants

Assignment 1

- Using the the file “trgn599.rna.tsv” from your R-Studio.
- Question 5
 - The number of columns (without including the first column) in this file refers to:
 - A) Number of samples
 - B) Number of genes
 - C) Number of clinical variables
 - D) Number of variants

Assignment 1

- For the following five questions please load the file “trgn599.clinical.tsv” from your R-Studio.
- Question 6
 - What type of data contains this file:
 - A) Gene expression data
 - B) Clinical data
 - C) Financial data
 - D) Econometric data
- Question 7
 - How many rows (without including the header) has this file:
 - A) 485
 - B) 465
 - C) 470
 - D) 464

Assignment 1

- Use the file “trgn599.clinical.tsv” from your R-Studio.
- Question 8
 - How many columns has this file:
 - A) 22 columns
 - B) 24 columns
 - C) 23 columns
 - D) 25 columns
- Question 9
 - The number of rows (without including the informational header) in this file refers to:
 - A) Number of samples
 - B) Number of mutations
 - C) Number of translocations
 - D) Number of variables

Assignment 1

- Use the file “trgn599.clinical.tsv” from your R-Studio.
- Question 10
 - The number of columns in this file refers to:
 - A) Number of genes
 - B) Number of mutations
 - C) Number of clinical-related variables
 - D) Number of translocations