USC Department Of Translational Genomics
Keck Medicine of USC

TRGN 599: Applied Data Science and Bioinformatics

UNIT I. Introduction and Basic Data Science

Week 3 – Assignment 3

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 3

- Instructions: Please answer the following questions, each one worth 20 points (Total: 100 points).
- Load the file “trgn599.clinical.tsv” in your R-Studio.
- In order to answer the following questions you should create an R markdown file as explained during the last class.
- Question 1
  - Which of the following R codes generate a histogram of weight using the TCGA clinical data?
    - A) hist(clinical_data$weight)
    - B) hist(clinical_data$height)
    - C) hist(clinical_data$age)
    - D) plot(clinical_data$weight)
- Question 2
  - Which of the following R codes generate a histogram of height using the TCGA clinical data?
    - A) hist(clinical_data$weight)
    - B) hist(clinical_data$height)
    - C) hist(clinical_data$age)
    - D) plot(clinical_data$weight)

Assignment 3

- Question 3
  - Which of the following R codes generate a box plot of gender against race using the TCGA clinical data?
    - A) hist(clinical_data$weight)
    - B) boxplot(clinical_data$height)
    - C) boxplot(clinical_data$gender)
    - D) plot(clinical_data$weight)

- Question 4
  - Which of the following R codes generate a histogram of the variable cholesterol from the cholesterol related dataset?
    - A) hist(clinical_data$weight)
    - B) hist(clinical_data$height)
    - C) boxplot(clinical_data$gender)
    - D) hist(colesterol_Dataset$CHOL)

Assignment 3

- Question 5
  - Which of the following R codes generate a histogram of the variable weight from the cholesterol related dataset.
    - A) hist(clinical_data$WEIGHT)
    - B) boxplot(clinical_data$height)
    - C) hist(colesterol_Dataset$WEIGHT)
    - D) plot(clinical_data$WEIGHT)