**STAT 40001/ STAT 50001 Statistical Computing Fall 2022**

**Lab-4**

**1)** Go to <http://jse.amstat.org/jse_data_archive.htm>

a) Import the ***babyboom.dat.txt*** data which has Time of Birth, Sex, and Birth Weight of Babies and number of minutes after midnight.

b) How many observations are recorded?

c) Print first 5 observations

d) Print last 5 observations

**2)** For 1997-2002, from police-reported car crashes in which there is a harmful event (people or property), and from which at least one vehicle was towed (Data are restricted to front-seat occupants, include only a subset of the variables recorded, and are restricted in other ways also), are provided as Item “nassCDS” under Rich source of data in Useful link.

a) Import the data in R using appropriate R code

b) Print the first 5 observations.

**3)** Consider a data set where the columns are separated by $

Col1$Col2$Col3

1$2$3

4$5$6

7$8$9

a$b$c

Save the data in a local drive and import it in R by removing $ sign.

**4)** *bweight* data provided in the Brightspace are Stata data (.dta) sets describing the babyweight along with several variables. Import the data in R and identify the dimension of the data and extract the variables included in the datasets.

**5)** URL below contains data related to the number of publications produced by Ph.D. biochemists.

<http://www.stata-press.com/data/lf2/couart2.dta>

1. Import the data in R. (Note that this is Stata data)
2. List the variables included in the data
3. State the dimension of the data