EDU Stats: Descriptive & Comparative: Assignment #2

**Once you have identified a dataset, explain how you accessed it. What links did you click? Was there any registration required? Did you download directly or was there an online system you navigated?**

Initially, I googled open infertility data sets. The results provided a link to the fertility clinics registered with the CDC. I selected the data set regarding clinic success rates in Texas. The data were downloaded into a spreadsheet for use.

**In R: Import the data**

* + Provide a list of variables in the dataset
    - Copy and paste this list in your word document

names(cdc2)

[1] "TotNumberCycles" "IVF\_Rate" “Unstimulated\_Rate" "Diag\_OvulatoryRate" "Diag\_MaleRate"

**What is the structure of the data? Which variables are character and which are numeric? Copy and paste this list in your word document**

> str("TotNumberCycles")

chr "TotNumberCycles"

> str("IVF\_Rate")

chr "IVF\_Rate"

> str("TotUsingFrozenEggs")

chr " TotUsingFrozenEggs "

> str("Diag\_OvulatoryRate")

chr "Diag\_OvulatoryRate"

> str("Diag\_MaleRate")

chr "Diag\_MaleRate"

*Changes had to be made in excel to convert any spaces or commas into “numeric” under “text-to-column” function.*

**Terms Explained**

TotNumberCycles Total cycles excluding approved research

IVF\_Rate ART types & procedures % of cycles using IVF

TotUsingFrozenEggs Cycles using frsh emb from frzn nondnr eggs

Diag\_OvulatoryRate Patient diagnosis % of cycles with ovulatory dysfunction

Diag\_MaleRate Patient diagnosis % of cycles with male factor

**Describe the data through Central Tendencies:**

**Median, Mean, & Mode of 5 variables:**

TotNumberCycles**:** Median – 118.0, Mean -220.0, Mode – 118 and 130

IVF\_Rate: Median – 1, Mean – 1, Mode – 1

TotUsingFrozenEggs Median – 01.000, Mean – 1.907, Mode – 0

Diag\_OvulatoryRate Median – 0.1600, Mean – 0.2109, Mode – 0.11

Diag\_MaleRate Median – 0.3400, Mean – 0.3614, Mode – 0.31

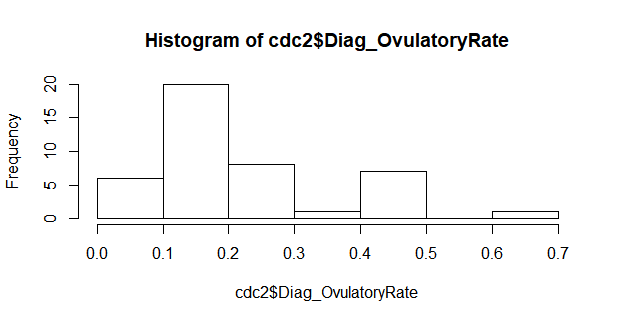
**Variance, Range, and Standard Deviation of Diag\_OvulatoryRate**

Variance – 0.02140864

Standard Deviation – 0.1463169

Range – 0.00 to 0.63

**Histogram**



**Box-Plot**

