

---

Name: Krish Agarwal  
Registration Number: 21112016  
Class: 2BSc DS A  
Lab 5 (Asynchronous)

---

Q1) From the package “datasets” use the dataset called “longley” and assign to a variable “df”. Description: Predict number of people employed from economic variables Type: Regression Dimensions: 16 instances, 7 attributes

```
library(datasets)
data("longley", package = "datasets")
df <- longley
df
```

##	GNP.deflator	GNP	Unemployed	Armed.Forces	Population	Year	Employed
## 1947	83.0	234.289	235.6	159.0	107.608	1947	60.323
## 1948	88.5	259.426	232.5	145.6	108.632	1948	61.122
## 1949	88.2	258.054	368.2	161.6	109.773	1949	60.171
## 1950	89.5	284.599	335.1	165.0	110.929	1950	61.187
## 1951	96.2	328.975	209.9	309.9	112.075	1951	63.221
## 1952	98.1	346.999	193.2	359.4	113.270	1952	63.639
## 1953	99.0	365.385	187.0	354.7	115.094	1953	64.989
## 1954	100.0	363.112	357.8	335.0	116.219	1954	63.761
## 1955	101.2	397.469	290.4	304.8	117.388	1955	66.019
## 1956	104.6	419.180	282.2	285.7	118.734	1956	67.857
## 1957	108.4	442.769	293.6	279.8	120.445	1957	68.169
## 1958	110.8	444.546	468.1	263.7	121.950	1958	66.513
## 1959	112.6	482.704	381.3	255.2	123.366	1959	68.655
## 1960	114.2	502.601	393.1	251.4	125.368	1960	69.564
## 1961	115.7	518.173	480.6	257.2	127.852	1961	69.331
## 1962	116.9	554.894	400.7	282.7	130.081	1962	70.551

Q2) Find the structure of the dataset and write an insight that you get from the data.

```
str(df)
```

```
## 'data.frame':  16 obs. of  7 variables:
## $ GNP.deflator: num  83 88.5 88.2 89.5 96.2 ...
## $ GNP          : num  234 259 258 285 329 ...
## $ Unemployed   : num  236 232 368 335 210 ...
## $ Armed.Forces: num  159 146 162 165 310 ...
## $ Population   : num  108 109 110 111 112 ...
## $ Year         : int  1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 ...
## $ Employed     : num  60.3 61.1 60.2 61.2 63.2 ...
```

*#The dataset contains data ranging from year 1947 to 1962 showing total number of population amongst wh*

Q3) Find the descriptive statistics of the dataframe generated.

```
summary(df)
```

```
##      GNP.deflator      GNP      Unemployed      Armed.Forces
## Min.   : 83.00   Min.   :234.3   Min.   :187.0   Min.   :145.6
## 1st Qu.: 94.53   1st Qu.:317.9   1st Qu.:234.8   1st Qu.:229.8
## Median :100.60   Median :381.4   Median :314.4   Median :271.8
## Mean   :101.68   Mean   :387.7   Mean   :319.3   Mean   :260.7
## 3rd Qu.:111.25   3rd Qu.:454.1   3rd Qu.:384.2   3rd Qu.:306.1
## Max.   :116.90   Max.   :554.9   Max.   :480.6   Max.   :359.4
##      Population      Year      Employed
## Min.   :107.6   Min.   :1947   Min.   :60.17
## 1st Qu.:111.8   1st Qu.:1951   1st Qu.:62.71
## Median :116.8   Median :1954   Median :65.50
## Mean   :117.4   Mean   :1954   Mean   :65.32
## 3rd Qu.:122.3   3rd Qu.:1958   3rd Qu.:68.29
## Max.   :130.1   Max.   :1962   Max.   :70.55
```

Q4) Find the mean, min and max of each column using indexing (both column index and name).

```
cat("----- Mean ----- \n\n")
```

```
## ----- Mean -----
```

```
cat("Population:")
```

```
## Population:
```

```
mean(df$Population)
```

```
## [1] 117.424
```

```
cat("\n")
```

```
cat("Employed:")
```

```
## Employed:
```

```
mean(df$Employed)
```

```
## [1] 65.317
```

```
cat("\n")
```

```
cat("Unemployed:")
```

```
## Unemployed:
```

```
mean(df$Unemployed)
```

```
## [1] 319.3313
```

```
cat("\n")
```

```
cat("Armed Forces:")
```

```
## Armed Forces:
```

```
mean(df[,4])
```

```
## [1] 260.6687
```

```
cat("\n")
```

```
cat("GNP:")
```

```
## GNP:
```

```
mean(df[,2])
```

```
## [1] 387.6984
```

```
cat("\n")
```

```
cat("GNP Deflator:")
```

```
## GNP Deflator:
```

```
mean(df[,1])
```

```
## [1] 101.6813
```

```
cat("\n\n")
```

```
cat("----- Max ----- \n\n")
```

```
## ----- Max -----
```

```
cat("Population:")
```

```
## Population:
```

```
max(df$Population)
```

```
## [1] 130.081
```

```
cat("\n")
```

```
cat("Employed:")
```

```
## Employed:
```

```
max(df$Employed)
```

```
## [1] 70.551
```

```
cat("\n")
```

```
cat("Unemployed:")
```

```
## Unemployed:
```

```
max(df$Unemployed)
```

```
## [1] 480.6
```

```
cat("\n")
```

```
cat("Armed Forces:")
```

```
## Armed Forces:
```

```
max(df[,4])
```

```
## [1] 359.4
```

```
cat("\n")
```

```
cat("GNP:")
```

```
## GNP:
```

```
max(df[,2])
```

```
## [1] 554.894
```

```

cat("\n")

cat("GNP Deflator:")

## GNP Deflator:

max(df[,1])

## [1] 116.9

cat("\n\n")

cat("----- Min ----- \n\n")

## ----- Min -----

cat("Population:")

## Population:

min(df$Population)

## [1] 107.608

cat("\n")

cat("Employed:")

## Employed:

min(df$Employed)

## [1] 60.171

cat("\n")

cat("Unemployed:")

## Unemployed:

min(df$Unemployed)

## [1] 187

```

```
cat("\n")
```

```
cat("Armed Forces:")
```

```
## Armed Forces:
```

```
min(df[,4])
```

```
## [1] 145.6
```

```
cat("\n")
```

```
cat("GNP:")
```

```
## GNP:
```

```
min(df[,2])
```

```
## [1] 234.289
```

```
cat("\n")
```

```
cat("GNP Deflator:")
```

```
## GNP Deflator:
```

```
min(df[,1])
```

```
## [1] 83
```

Q5) Use a while loop to iterate through the Unemployed column to find the second largest value of the column and display the year of the same.

```
count <- 1
largest <- 0

while (count <= length(df$Unemployed))
{
  if(df$Unemployed[count] > largest){
    largest2 <- largest
    largest <- df$Unemployed[count] }

  count = count + 1

  #df$Year["largest2", 3]
  #for (i in df$Year){
    #if (i == 468.1){
      #print(df$Year[i, 3])
    #}
  #}
}

paste("Second Largest value:", largest2)
```

```
## [1] "Second Largest value: 468.1"
```

Q6) Write a while loop starting with  $x = 100$ . The loop should skip all the multiples of 7 till u get 100 values.

```
x <- 100
count <- 1

while (count<=100){
  if (x%%7!=0){
    print(x)
  }

  x <- x+1
  count = count +1
}
```

```
## [1] 100
## [1] 101
## [1] 102
## [1] 103
## [1] 104
## [1] 106
## [1] 107
## [1] 108
## [1] 109
## [1] 110
## [1] 111
## [1] 113
## [1] 114
## [1] 115
## [1] 116
## [1] 117
## [1] 118
## [1] 120
## [1] 121
## [1] 122
## [1] 123
## [1] 124
## [1] 125
## [1] 127
## [1] 128
## [1] 129
## [1] 130
## [1] 131
## [1] 132
## [1] 134
## [1] 135
## [1] 136
## [1] 137
## [1] 138
## [1] 139
## [1] 141
## [1] 142
```

```
## [1] 143
## [1] 144
## [1] 145
## [1] 146
## [1] 148
## [1] 149
## [1] 150
## [1] 151
## [1] 152
## [1] 153
## [1] 155
## [1] 156
## [1] 157
## [1] 158
## [1] 159
## [1] 160
## [1] 162
## [1] 163
## [1] 164
## [1] 165
## [1] 166
## [1] 167
## [1] 169
## [1] 170
## [1] 171
## [1] 172
## [1] 173
## [1] 174
## [1] 176
## [1] 177
## [1] 178
## [1] 179
## [1] 180
## [1] 181
## [1] 183
## [1] 184
## [1] 185
## [1] 186
## [1] 187
## [1] 188
## [1] 190
## [1] 191
## [1] 192
## [1] 193
## [1] 194
## [1] 195
## [1] 197
## [1] 198
## [1] 199
```

Q7) Write a for loop to iterate through the dataframe of longely and create a new column GNP.PC that has the values GNP per capita (or income per person).  $\text{GNP per capita} = \text{GNP} / \text{Population}$