Assignment: Random Sampling and Data Structures

YOUR_NAME_HERE

Problem 1

Let us assign treatment groups A or B randomly to individuals. (Done together in class)

```
# Individuals in the study
Name = c("Alice","Bob","Hector","Steve","Tony","Steph","Mary", "Charlie", "Diana", "Ethan")

# two things in data frame
# 1. Names
# 2. Assignments

n_names <- length(Name)
study <- data.frame(
  Names = Name,
  Assignments = sample(c("Treatment A","Treatment B"),n_names,replace=TRUE)
)
print(study)
```

```
     Names Assignments
1    Alice Treatment B
2      Bob Treatment A
3   Hector Treatment A
4    Steve Treatment A
5     Tony Treatment B
6    Steph Treatment B
7     Mary Treatment A
8  Charlie Treatment B
9    Diana Treatment A
10   Ethan Treatment B
```

Problem 2

Using the data frame you created, let's subset it in various way.

Write code to get only individuals from treatment group B. (Done together in class)

```
mask <- study$Assignments == "Treatment B"
study_Bgroup <- study[mask,]
study_Bgroup
```

```
     Names Assignments
1    Alice Treatment B
5     Tony Treatment B
6    Steph Treatment B
8  Charlie Treatment B
10   Ethan Treatment B
```

Write code to get the first 5 individuals of the group.

```
study_first5 <- study[1:5,]
study_first5
```

```
   Names Assignments
1  Alice Treatment B
2    Bob Treatment A
3 Hector Treatment A
4  Steve Treatment A
5   Tony Treatment B
```

Problem 3

Create a list that includes the following elements:

- The full data frame of individuals and treatment groups
- The list of names of individuals in treatment group A
- The list of names of individuals in treatment group B

Create these lists using subsetting and logicals as opposed to hard coding the names. (Done together in class)

```r
mask <- study$Assignments == "Treatment A"
study_Agroup <- study[mask,]
study_list <- list(
  full_study <- study,
  group_A <- study_Agroup,
  group_B <- study_Bgroup
)
study_list
```

```
[[1]]
     Names Assignments
1    Alice Treatment B
2      Bob Treatment A
3   Hector Treatment A
4    Steve Treatment A
5     Tony Treatment B
6    Steph Treatment B
7     Mary Treatment A
8  Charlie Treatment B
9    Diana Treatment A
10   Ethan Treatment B

[[2]]
   Names Assignments
2    Bob Treatment A
3 Hector Treatment A
4  Steve Treatment A
7   Mary Treatment A
9  Diana Treatment A

[[3]]
     Names Assignments
1    Alice Treatment B
5     Tony Treatment B
6    Steph Treatment B
8  Charlie Treatment B
10   Ethan Treatment B
```

Problem 4

We have a new clinical trial approaching. We have 300 individuals. Instead of names, we will only use ID numbers, which will range from 501 to 800. We know that the first 100 of these are male and the rest are female. We must randomly assign a treatment out 4 possible values: {A, B, C, D}. Create a data frame that includes an individuals ID number, whether they are male or female, and the treatment randomly assigned. **Print the first 10 rows of the resulting data frame.**

```
# make ID numbers
ID <- seq(501,800,1) # 501:800

# M/F
Gender <- c(rep("M",100),rep("F",200))

# Random Assignments
Assignments <- sample(c("A","B","C","D"),length(ID),replace=TRUE)

# Create Data Frame
another_study <- data.frame(
  ID,
  Gender,
  Assignments
)

another_study[1:10,]
```

```
    ID Gender Assignments
1  501      M           B
2  502      M           B
3  503      M           B
4  504      M           A
5  505      M           A
6  506      M           B
7  507      M           B
8  508      M           C
9  509      M           D
10 510      M           A
```

Problem 5

You receive some new instruction to have exactly 75 in each group. So instead of randomly assigning the group to each individual, randomly shuffle the ID numbers, and ensure there are 75 in each of the 4 treatment groups. **Print the first 10 rows of the resulting data frame.**