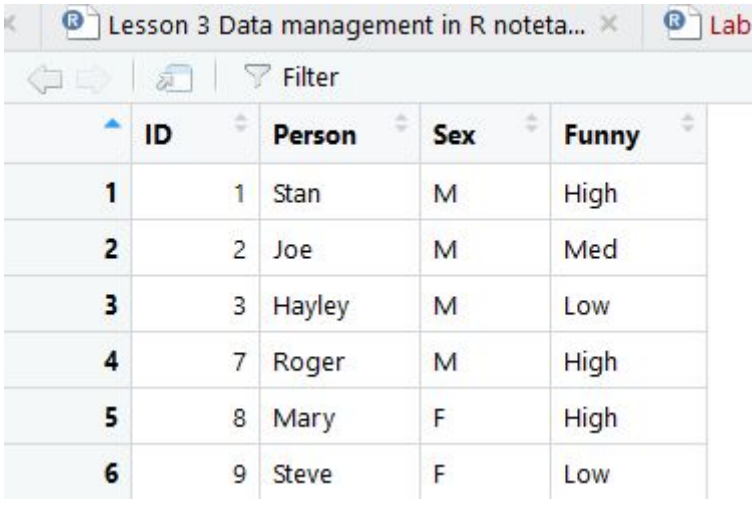
```
# STAT521_LAB_3_Data_Management              Rosana Lin Ho #

# Question 1 #

df1 = data.frame(ID = c(1,2,3,7,8,9),
                 Person
=(c('Stan','Joe','Hayley','Roger','Mary','Steve')),
                 Sex = c('M','M','M','M','F','F'),
                 Funny =c('High','Med','Low','High','High','Low'))
View(df1)
```




| | ID | Person | Sex | Funny |
|---|---|---|---|---|
| 1 | 1 | Stan | M | High |
| 2 | 2 | Joe | M | Med |
| 3 | 3 | Hayley | M | Low |
| 4 | 7 | Roger | M | High |
| 5 | 8 | Mary | F | High |
| 6 | 9 | Steve | F | Low |

```
str(df1)
'data.frame':   6 obs. of  4 variables:
 $ ID    : num  1 2 3 7 8 9
 $ Person: Factor w/ 6 levels "Hayley","Joe",..: 5 2 1 4 3 6
 $ Sex   : Factor w/ 2 levels "F","M": 2 2 2 2 1 1
 $ Funny : Factor w/ 3 levels "High","Low","Med": 1 3 2 1 1 2


# Question 2 #

df2_Age= data.frame(ID = c(1,2,3,7,8,9),
                    Age = c(41,60,21,1600,41,15))
View(df2_Age)
```

| | ID | Age |
|---|---|---|
| 1 | 1 | 41 |
| 2 | 2 | 60 |
| 3 | 3 | 21 |
| 4 | 7 | 1600 |
| 5 | 8 | 41 |
| 6 | 9 | 15 |

```
#Merging data frame #
df2 = merge(df1,df2_Age, by="ID",all=TRUE)

View(df2)
```

| | ID | Person | Sex | Funny | Age |
|---|---|---|---|---|---|
| 1 | 1 | Stan | M | High | 41 |
| 2 | 2 | Joe | M | Med | 60 |
| 3 | 3 | Hayley | M | Low | 21 |
| 4 | 7 | Roger | M | High | 1600 |
| 5 | 8 | Mary | F | High | 41 |
| 6 | 9 | Steve | F | Low | 15 |

```
#Print last 4 records

tail(df2,4)
```

```
  ID Person Sex Funny  Age
3  3 Hayley   M   Low   21
4  7  Roger   M  High 1600
5  8   Mary   F  High   41
6  9  Steve   F   Low   15
```

```
# Question 3 #

# a #
df3 = data.frame(ID = c(4,5,6),
                 Person = I(c('Peter', 'Lois', 'Meg')),
                 Sex = c('M','F','F'),
                 Funny = c('High','High','Low'),
                 Age = c(42,40,17))
View(df3)
```

| | ID | Person | Sex | Funny | Age |
|---|---|---|---|---|---|
| 1 | 4 | Peter | M | High | 42 |
| 2 | 5 | Lois | F | High | 40 |
| 3 | 6 | Meg | F | Low | 17 |

```
# b #
df3_b = rbind(df2,df3) #concatenate (stack) datasets
df3_b[order(df3_b$ID),]  # order dataset by variable
```

```
  ID Person Sex Funny  Age
1  1   Stan   M  High   41
2  2    Joe   M   Med   60
3  3 Hayley   M   Low   21
7  4  Peter   M  High   42
8  5   Lois   F  High   40
9  6    Meg   F   Low   17
4  7  Roger   M  High 1600
5  8   Mary   F  High   41
6  9  Steve   F   Low   15
```

```
# c #
summary(df3_b)
```

```
       ID       Person           Sex    Funny        Age        
 Min.   :1   Length:9           F:4   High:5   Min.   :  15.0  
 1st Qu.:3   Class :AsIs        M:5   Low :3   1st Qu.:  21.0  
 Median :5   Mode  :character         Med :1   Median :  41.0  
 Mean   :5                                     Mean   : 208.6  
 3rd Qu.:7                                     3rd Qu.:  42.0  
 Max.   :9                                     Max.   :1600.0  
```

```
# d #
df3_b[df3_b$Sex =='F',]
```

```
  ID Person Sex Funny Age
5  8   Mary   F  High  41
6  9  Steve   F   Low  15
8  5   Lois   F  High  40
9  6    Meg   F   Low  17
```

```
# e #
df3_b[((df3_b$Sex =='F') & (df3_b$Funny != 'Low')),]
df3_b
```

```
  ID Person Sex Funny Age
5  8   Mary   F  High  41
8  5   Lois   F  High  40
```

```
# Question 4 #

# a #
df4 = data.frame(ID = c(1,2,4,5,7),
                 EnterStudy = I(c('July 12,2017', 'March 4,2016', 'April
1,2018','May 5,2017','November 11,2016')),
                 Group = c(0,1,0,2,1))
View( df4)
```

| | ID | EnterStudy | Group |
|---|---|---|---|
| 1 | 1 | July 12,2017 | 0 |
| 2 | 2 | March 4,2016 | 1 |
| 3 | 4 | April 1,2018 | 0 |
| 4 | 5 | May 5,2017 | 2 |
| 5 | 7 | November 11,2016 | 1 |

```
str(df4)
'data.frame':   5 obs. of  3 variables:
 $ ID        : num  1 2 4 5 7
 $ EnterStudy: 'AsIs' chr  "July 12,2017" "March 4,2016" "April 1,2018"
"May 5,2017" ...
 $ Group     : num  0 1 0 2 1

typeof(df4$EnterStudy)
[1] "character"

# Changing format
df4$EnterStudy = as.Date(df4$EnterStudy, format = "%B %e, %Y") #doesnt
work
df4$EnterStudy = strptime(df4$EnterStudy, "%B %e, %Y")

str(df4)
'data.frame':   5 obs. of  3 variables:
 $ ID        : num  1 2 4 5 7
 $ EnterStudy: POSIXlt, format: "2017-07-12" "2016-03-04" ...
 $ Group     : num  0 1 0 2 1

#Merging data frame #
df4_a = merge(df3_b,df4, by="ID",all=TRUE)
View(df4_a)
```

| | ID | Person | Sex | Funny | Age | EnterStudy | Group |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Stan | M | High | 41 | 2017-07-12 | 0 |
| 2 | 2 | Joe | M | Med | 60 | 2016-03-04 | 1 |
| 3 | 3 | Hayley | M | Low | 21 | *NA* | *NA* |
| 4 | 4 | Peter | M | High | 42 | 2018-04-01 | 0 |
| 5 | 5 | Lois | F | High | 40 | 2017-05-05 | 2 |
| 6 | 6 | Meg | F | Low | 17 | *NA* | *NA* |
| 7 | 7 | Roger | M | High | 1600 | 2016-11-11 | 1 |
| 8 | 8 | Mary | F | High | 41 | *NA* | *NA* |
| 9 | 9 | Steve | F | Low | 15 | *NA* | *NA* |

```
# b #
t_stan_peter = df4_a$EnterStudy[1] - df4_a$EnterStudy[4]
t_stan_peter
Time difference of -263 days

# c #
df4_a$Group= factor(df4_a$Group, labels = c("Control","Low intensity",
"High intensity"))
View(df4_a)
```

| | ID | Person | Sex | Funny | Age | EnterStudy | Group |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Stan | M | High | 41 | 2017-07-12 | Control |
| 2 | 2 | Joe | M | Med | 60 | 2016-03-04 | Low intensity |
| 3 | 3 | Hayley | M | Low | 21 | *NA* | *NA* |
| 4 | 4 | Peter | M | High | 42 | 2018-04-01 | Control |
| 5 | 5 | Lois | F | High | 40 | 2017-05-05 | High intensity |
| 6 | 6 | Meg | F | Low | 17 | *NA* | *NA* |
| 7 | 7 | Roger | M | High | 1600 | 2016-11-11 | Low intensity |
| 8 | 8 | Mary | F | High | 41 | *NA* | *NA* |
| 9 | 9 | Steve | F | Low | 15 | *NA* | *NA* |