Lab 3

Univariate Statistics with R

Below is a set of suggested answers for questions 1 to 9 at the end of Lab 3. Note, this is just one way of doing the tasks in the question set. If you have used different code, provided the end results look similar to those presented below, this is fine. So, here we go;

Read in data:

```
data = read.csv("Univar_Lab_3_Data.csv", header = T, sep = ",")
head(data)
```

```
##            V1         V2 V3 V4 V5
## 1 -0.30237540  1.9806825  0  0  8
## 2  1.53332331  1.6365009  0  1  4
## 3  3.25256819  1.2381232  0  0  5
## 4  0.42935839 -0.5642602  0  0  1
## 5  0.41289491  0.9782996  0  0  3
## 6  0.02135718 -0.1012774  0  0  1
```

Adding an ID variable:

```
ID = paste("ID", 1:200, sep = "")
data = as.data.frame(cbind(ID, data))
head(data)
```

```
##    ID          V1         V2 V3 V4 V5
## 1 ID1 -0.30237540  1.9806825  0  0  8
## 2 ID2  1.53332331  1.6365009  0  1  4
## 3 ID3  3.25256819  1.2381232  0  0  5
## 4 ID4  0.42935839 -0.5642602  0  0  1
## 5 ID5  0.41289491  0.9782996  0  0  3
## 6 ID6  0.02135718 -0.1012774  0  0  1
```

Name the variables:

```
colnames(data)[2:6] <- c("Daily Energy (vig)", "Daily enery (light)",
                         "Sex", "BDNF", "Attitude")
names(data)
```

```
## [1] "ID"                  "Daily Energy (vig)"  "Daily enery (light)"
## [4] "Sex"                 "BDNF"                "Attitude"
```

Add labels to sex and BDNF-alpha

```
data$Sex <- as.factor(data$Sex)
levels(data$Sex) = c("Female", "Male")
data$BDNF <- as.factor(data$BDNF)
levels(data$BDNF) = c("Not Present", "Present")
head(data)
```

```
##    ID Daily Energy (vig) Daily enery (light)    Sex        BDNF Attitude
## 1 ID1        -0.30237540           1.9806825 Female Not Present        8
## 2 ID2         1.53332331           1.6365009 Female     Present        4
## 3 ID3         3.25256819           1.2381232 Female Not Present        5
## 4 ID4         0.42935839          -0.5642602 Female Not Present        1
## 5 ID5         0.41289491           0.9782996 Female Not Present        3
```

```
## 6 ID6         0.02135718          -0.1012774 Female Not Present        1
```

Descriptives:

```
library(psych)
descript = describe(data)
descript_out = descript[ , c(2, 3)]
```

The table above has some values for central tendency that are not optimal for some variables. So we want to tidy this. **Note**: this is really just an R-skills questions, we would not want to present a table like this in a paper.

```
# Empty the cells for the inappropriate mean estimates
descript_out[1, 2] = NA
descript_out[4, 2] = NA
descript_out[5, 2] = NA
descript_out
```

```
##                        n mean
## ID*                  200   NA
## Daily Energy (vig)   200 0.36
## Daily enery (light) 200 0.14
## Sex*                 200   NA
## BDNF*                200   NA
## Attitude             200 4.45
```

We would likely just want to present the frequencies for sex and BDNF so we could just save the output from `table()`:

```
sex_freq = table(data$Sex)
BDNF_freq = table(data$BDNF)
sex_freq
```

```
##
## Female   Male
##    100    100
```

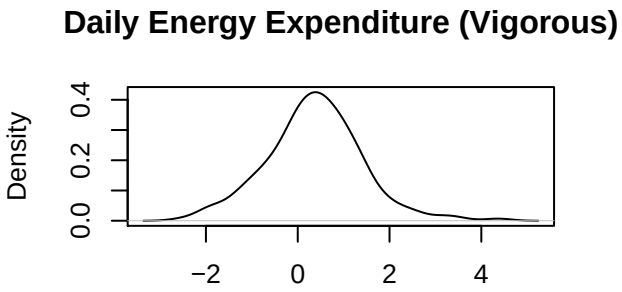
```
BDNF_freq
```

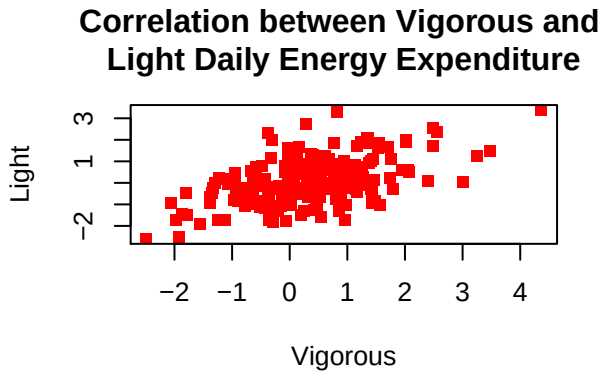
```
##
## Not Present     Present
##          95         105
```

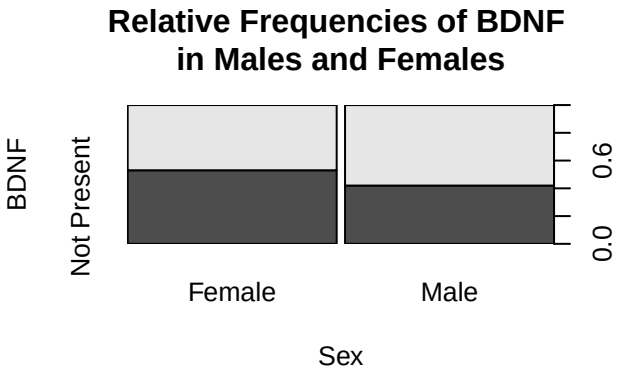
For the plots I provide a single code to produce the 2x2 single figure. If you want to check any individual plot, you can look at this aspect of the code.

```
par(mfrow = c(2, 2))

plot(as.factor(data$Attitude), main = "Attitude to Exercise",
     xlab = "Attitude Rating",
     ylab = "Frequency", col = "lightblue")

plot(density(data[ , 2]), main = "Daily Energy Expenditure (Vigorous)", xlab = "")

plot(data[ , 2], data[ , 3],
     main = "Correlation between Vigorous and \nLight Daily Energy Expenditure",
     xlab = "Vigorous", ylab = "Light", pch = 15, col = "red")

plot(data[ , 4], data[ , 5],
```

```
     main = "Relative Frequencies of BDNF \nin Males and Females",
     xlab = "Sex", ylab = "BDNF")
```





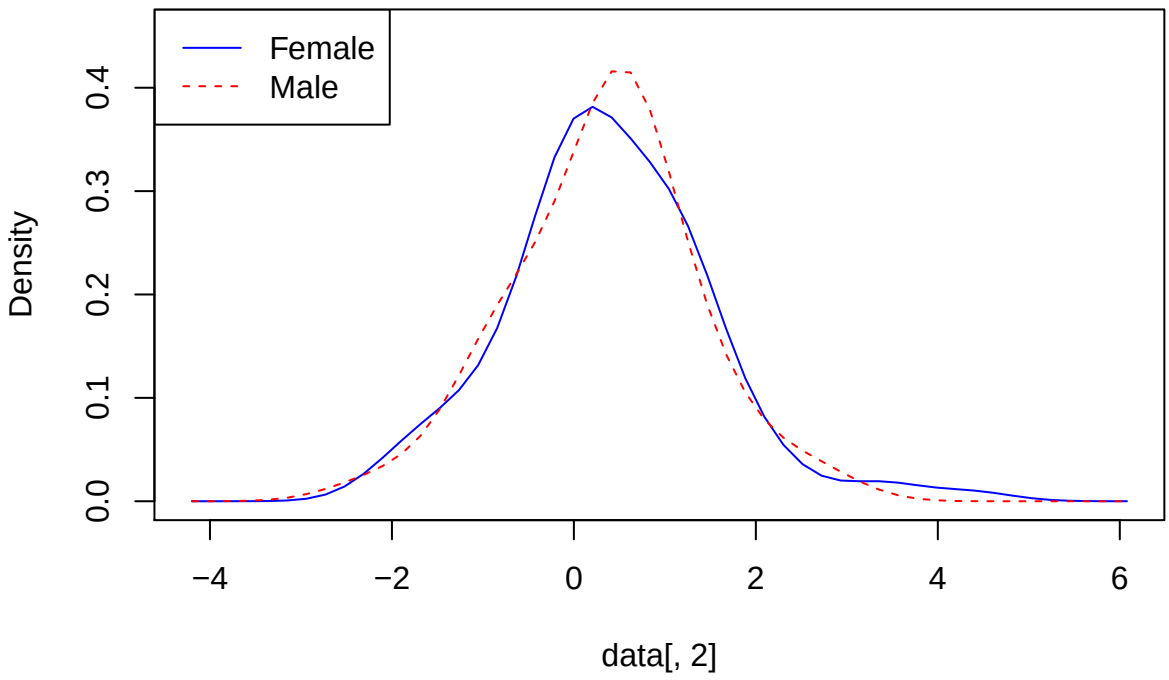


The final plot asked for was the comparison of densities. Code for this plot would look something like:

```
## Warning: package 'sm' was built under R version 3.4.2
```

```
## Package 'sm', version 2.2-5.4: type help(sm) for summary information
```

```
sm.density.compare(data[ , 2], data[ , 4], col = c("blue", "red"), lty = c(1, 2))
legend("topleft", levels(data[ , 4]), lty = c(1, 2), col = c("blue", "red"))
```



Any combination of these 5 plots in a single figure is appropriate.