Data Assignment 4

Yash Patel

01 April 2022

Question 1

Create three lists (x, y and z) from randomly generated numbers from a uniform distribution. (Hint – see "runif" and "round") in the cookbook.) Each list (x,y, and z) should contain 10 numbers

```
q1x <- round(runif(10, min = 0, max = 100))
q1y <- round(runif(10, min = 0, max = 100))
q1z <- round(runif(10, min = 0, max = 100))
```

Run correlations between x and y, x and z, and y and z (5 pts.)

```
cor(q1x, q1y)
```

```
## [1] -0.2263951
```

```
cor(q1x, q1z)
```

```
## [1] -0.3233139
```

```
cor(q1y, q1z)
```

```
## [1] -0.1655712
```

Create a data frame from the three lists. Cut and paste the output here (5 pts.)

```
df1 <- data.frame(q1x, q1y, q1z)
library(knitr)
kable(df1)
```

q1x	q1y	q1z
3	42	68
62	42	54
3	78	1

q1x	q1y	q1z
68	91	57
46	10	85
33	72	71
6	78	98
79	7	32
73	89	2
60	37	23

Find the means of the three columns of the data frame (hint: use "apply" command (5 pts.)

```
apply(df1, 2, mean)
```

```
##  q1x  q1y  q1z
## 43.3 54.6 49.1
```

Using the data frame you created, rerun the correlation between x and y. (Hint, you can select individual columns from a data frame. One way is to list the data frame and then select the specific column. For example, "df$x" refers to the x column in data frame "df". The "x" column from the data frame is specified by the dollar sign ($)) (5 pts.)

```
cor(df1$q1x, df1$q1y)
```

```
## [1] -0.2263951
```

```
cor(df1$q1x, df1$q1z)
```

```
## [1] -0.3233139
```

```
cor(df1$q1y, df1$q1z)
```

```
## [1] -0.1655712
```

Question 2

Download and install two R packages – "ppcor" and "psych". (Hint – use "install.packages(" package name ") - keep the quotes – to download a package. Use library(package name) – no quotes – to make the R package available to your current R session. (10 pts.)

```
install.packages("ppcor")
install.packages("psych")
```

```
library(ppcor)
```

```
## Warning: package 'ppcor' was built under R version 4.1.3
```

```
## Loading required package: MASS
```

```
library(psych)
```

Create a vector of data. Run the function from the "psych" package called "describe" on your vector (hint – if the psych package installed correctly, the command describe (vector) should produce an output that includes 13 different descriptive statistics. (5 pts.)

```
q2v <- c(1, 2, 3, 4, 5, 6, 7)
describe(q2v)
```

```
##    vars n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 7    4 2.16      4       4 2.97   1   7     6    0    -1.71 0.82
```

Question 3

(Since you created unique vectors, your plots and correlations will be based on those values. Note the partial correlations. Does the partial correlation change from the original correlation of X and Y? If so, how? What does this mean? You don't have to explain this in the assignment, but you should be able to interpret the results.)

Run the partial correlation on your data frame using

pcor (5 pts)

```
pcor(df1)
```

```
## $estimate
##            q1x        q1y        q1z
## q1x  1.0000000 -0.2999543 -0.3756002
## q1y -0.2999543  1.0000000 -0.2590455
## q1z -0.3756002 -0.2590455  1.0000000
##
## $p.value
##           q1x       q1y       q1z
## q1x 0.0000000 0.4329189 0.3191754
## q1y 0.4329189 0.0000000 0.5009034
## q1z 0.3191754 0.5009034 0.0000000
##
## $statistic
##            q1x        q1y        q1z
## q1x  0.0000000 -0.8319109 -1.0722531
## q1y -0.8319109  0.0000000 -0.7095918
## q1z -1.0722531 -0.7095918  0.0000000
##
## $n
## [1] 10
##
## $gp
## [1] 1
##
## $method
## [1] "pearson"
```

partial.r (5 pts.)

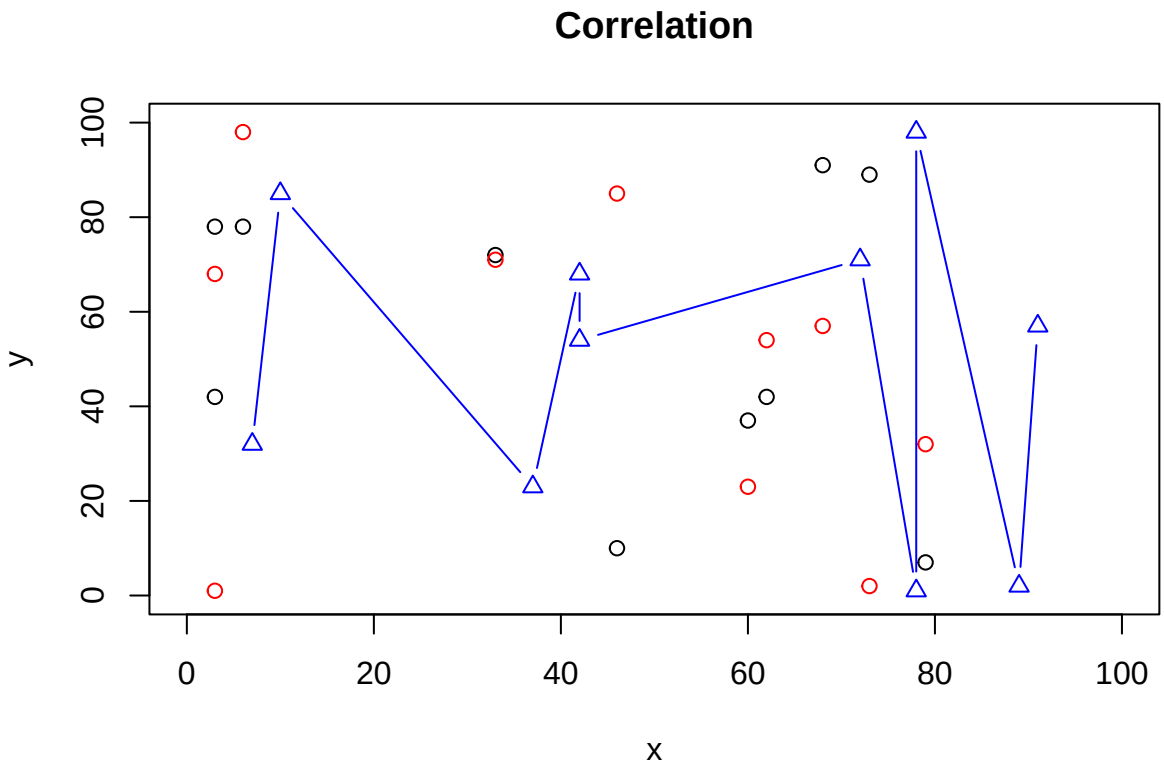
```
partial.r(df1)
```

```
##            q1x        q1y        q1z
## q1x  1.0000000 -0.2999543 -0.3756002
## q1y -0.2999543  1.0000000 -0.2590455
## q1z -0.3756002 -0.2590455  1.0000000
```

Question 4

Create a scatter plot of x and y. The title of the plot should be "Correlation". Make sure that the X and Y axes are labelled ("x" and "y", or another title is fine). (5 pts.)

```
plot(q1x, q1y,
     main = "Correlation",
     xlab = "x",
     ylab = "y",
     xlim = c(0, 100),
     ylim = c(0, 100)
     )
points(q1x, q1z, col = "red")
df4 <- df1[order(df1$q1y),] # Sorts the data frame so that lines on plot connect properly
points(df4$q1y, df4$q1z, col = "blue", pch = 2, type = "b")
```



Add another set of data to the (same) plot of x and z using points and a different color for the points. (5 pts)

See above.

Add another set of data to the (same) plot of y and z using points with a line connecting the points. (Also use a different color line and different symbol type (pch).) (5 pts.)

See above.