Quizzes due Apr 29, 2021 23:01 +03

Assignment Prep

Load the **GSE5859Subset** data:

```
library(GSE5859Subset)
data(GSE5859Subset)
```

Question 1

2/2 points (graded)

Inspect the dimensions of the `geneExpression` matrix.

A: How many samples are in the dataset?

8793 ✔ **Answer:** 8793

8793

B: How many features are in the dataset?

24 ✔ **Answer:** 24

24

Explanation

Part A:

```
dim(geneExpression)[1]
```

```
## [1] 8793
```

Part B:

```
dim(geneExpression)[2]
```

```
## [1] 24
```

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

Question 2

2/2 points (graded)

Inspect the `sampleInfo` data frame.

A: How many samples are from the ethnicity "ASN"?

23 **Answer:** 23

23

B: Which sample is from the ethnicity "CEU"?

15 **Answer:** 15

15

Explanation

Part A:

```
sum(sampleInfo$ethnicity == "ASN")
```

```
## [1] 23
```

Part B:

```
which(sampleInfo$ethnicity == "CEU")
```

```
## [1] 15
```

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

Question 3

2/2 points (graded)

Inspect the `sampleInfo` data frame.

A: What is the distance between samples 3 and 7?

36.05656

Answer: 36.05656

36.05656

B: What is the distance between samples 4 and 14?

46.72515

Answer: 46.72515

46.72515

Explanation

Part A:

```
sqrt(crossprod(geneExpression[,3]-geneExpression[,7]))
```

```
##          [,1]
## [1,] 36.05656
```

Part B:

```
sqrt(crossprod(geneExpression[,4]-geneExpression[,14]))
```

```
##          [,1]
## [1,] 46.72515
```

Submit

You have used 1 of 5 attempts

Answers are displayed within the problem

Question 4

1/1 point (graded)

This code finds the mean distance between the first sample (column 1) and all other samples:

```
column = 1
x = 1:ncol(geneExpression)
dists = sapply(x, function(x){
    test = geneExpression[,x]
    target = geneExpression[,column]
    sqrt(crossprod(target-test))
})
mean(dists)
```

```
## [1] 34.2757
```

Add an extra `sapply()` loop to this code to check the mean distance between each sample (column) and all other samples.

Which sample (column) has the largest mean distance from other samples?
(Report the column number)

15

✔ **Answer:** 15

15

Explanation

```
columns = 1:ncol(geneExpression)
mean_dists = sapply(columns, function(column){
    dists = sapply(x, function(x){
        test = geneExpression[,x]
        target = geneExpression[,column]
        sqrt(crossprod(target-test))
    })
mean(dists)
})
which.max(mean_dists)
```

```
## [1] 15
```

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

Question 5

1/1 point (graded)

Use `dist()` to calculate the distance between all pairs of samples.

What is the maximum distance between any two samples?

56.9792 ✔ **Answer:** 56.97692

56.9792

Explanation

```
d = dist(t(geneExpression))
max(d)
```

```
## [1] 56.97692
```

Submit You have used 2 of 5 attempts

Answers are displayed within the problem

Question 6

2/2 points (graded)

A: What is the distance between features "1007_s_at" and "201371_s_at"?

12.42643 ✔ **Answer:** 12.42643

12.42643

B: What is the distance between features "202138_x_at" and "202152_x_at"?

2.3093

✔ **Answer:** 2.3093

2.3093

Explanation

Part A:

```
sqrt(crossprod(geneExpression["1007_s_at",]-geneExpression[
```

```
##           [,1]
## [1,] 12.42643
```

Part B:

```
sqrt(crossprod(geneExpression["202138_x_at",]-geneExpressio
```

```
##         [,1]
## [1,] 2.3093
```

Submit

You have used 1 of 5 attempts

ℹ Answers are displayed within the problem

Question 7

1/1 point (graded)

Use `dist()` to calculate the distance between all pairs of features.

What is the maximum distance between any two features?

52.04894

✔ **Answer:** 52.04894

52.04894

Explanation

```
d2 = dist(geneExpression)
max(d2)
```

```
## [1] 52.04894
```

Submit

You have used 1 of 5 attempts

Answers are displayed within the problem