QUEEN'S UNIVERSITY BELFAST
SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Data Analysis & Visualisation

CSC3062

BEng (CS & SE), MEng (CS & SE), BIT & CIT

Dr Reza Rafiee

Semester 1 – 2019/2020

This is R

R Programming; Dataframes

- The frequently used **R-Objects**:
  - Vectors
  - Lists
  - Matrices
  - Arrays
  - Factors
  - Dataframes

R Programming; Dataframes

How to create a dataframe?

Filter

	name	score	attempts	qualify
1	Anastasia	12.5	1	yes
2	Dima	9.0	*NA*	no
3	Katherine	16.5	2	yes
4	James	12.0	*NA*	no
5	Emily	9.0	2	no
6	Michael	20.0	*NA*	yes
7	Matthew	14.5	1	yes
8	Laura	13.5	*NA*	no
9	Kevin	8.0	2	no
10	Jonas	19.0	1	yes

Showing 1 to 10 of 10 entries, 4 total columns

Dataframe is a tabular collection of data in which columns can have different modes

R Programming; Dataframes

How to create a dataframe?

```
exam_data <- data.frame( name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily',
'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'), score = c(12.5, 9, 16.5, 12, 9, 20,
14.5, 13.5, 8, 19), attempts = c(1, NA, 2, NA, 2, NA, 1, NA, 2, 1), qualify = c('yes',
'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes') )

Veiw(exam_data)
```

How to access to the elements of a dataframe?
What is the total number of missing?

Filter

	name	score	attempts	qualify
1	Anastasia	12.5	1	yes
2	Dima	9.0	*NA*	no
3	Katherine	16.5	2	yes
4	James	12.0	*NA*	no
5	Emily	9.0	2	no
6	Michael	20.0	*NA*	yes
7	Matthew	14.5	1	yes
8	Laura	13.5	*NA*	no
9	Kevin	8.0	2	no
10	Jonas	19.0	1	yes

Showing 1 to 10 of 10 entries, 4 total columns

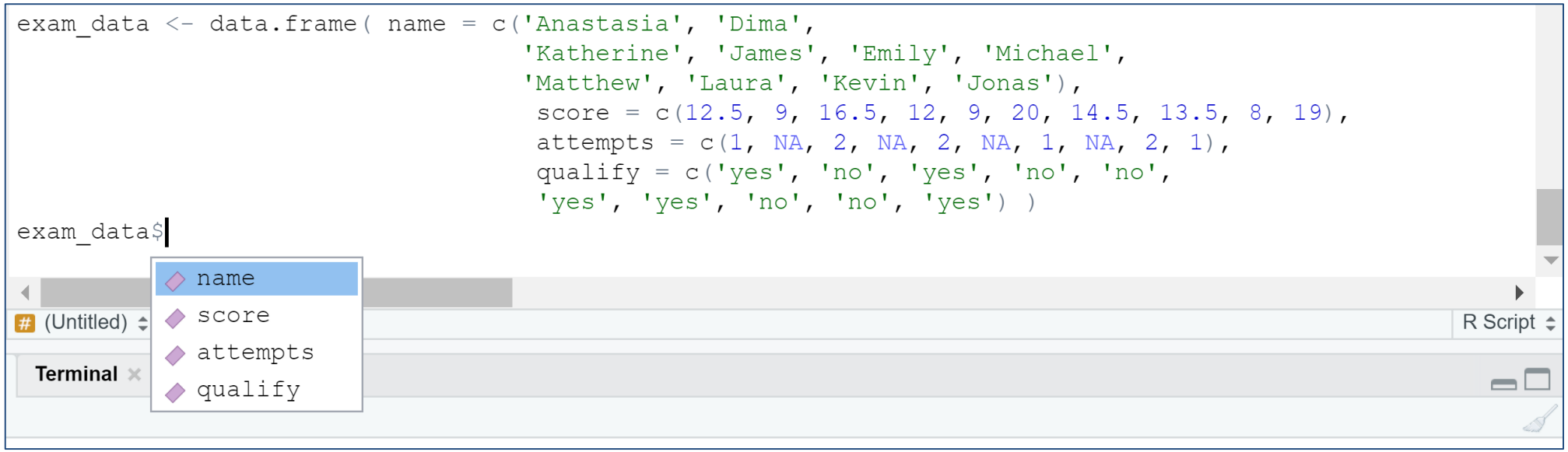
R Programming; Dataframes

How to create a dataframe?

```
exam_data <- data.frame( name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily',
'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'), score = c(12.5, 9, 16.5, 12, 9, 20,
14.5, 13.5, 8, 19), attempts = c(1, NA, 2, NA, 2, NA, 1, NA, 2, 1), qualify = c('yes',
'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes') )

Veiw(exam_data)
```

How to access to the elements of a dataframe?



R Programming; Dataframes

What is the total number of missing?

Write one line of code to create the following result in the console?

```
[1] "The number of NA values in attempts column is 4"
```

Filter

	name	score	attempts	qualify
1	Anastasia	12.5	1	yes
2	Dima	9.0	*NA*	no
3	Katherine	16.5	2	yes
4	James	12.0	*NA*	no
5	Emily	9.0	2	no
6	Michael	20.0	*NA*	yes
7	Matthew	14.5	1	yes
8	Laura	13.5	*NA*	no
9	Kevin	8.0	2	no
10	Jonas	19.0	1	yes

Showing 1 to 10 of 10 entries, 4 total columns

R Programming; Dataframes

How to access to a column of a dataframe?

```
print(paste("The number of NA values in attempts column is",
            sum(is.na(exam_data$attempts)), sep = " "))
```

Filter

	name	score	attempts	qualify
1	Anastasia	12.5	1	yes
2	Dima	9.0	NA	no
3	Katherine	16.5	2	yes
4	James	12.0	NA	no
5	Emily	9.0	2	no
6	Michael	20.0	NA	yes
7	Matthew	14.5	1	yes
8	Laura	13.5	NA	no
9	Kevin	8.0	2	no
10	Jonas	19.0	1	yes

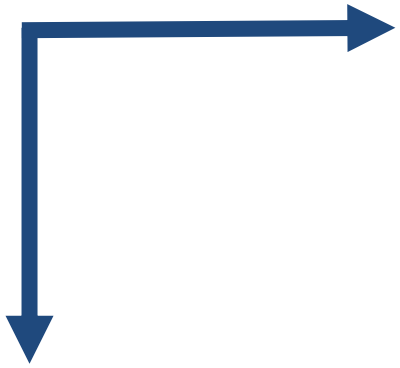
Showing 1 to 10 of 10 entries, 4 total columns

R Programming; Comparison

n=1

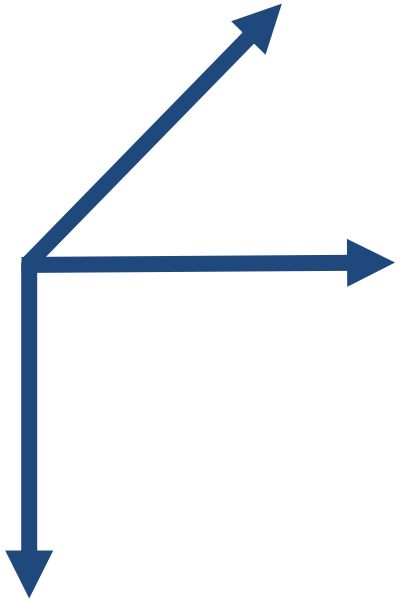
Vector
Factor
List

n=2



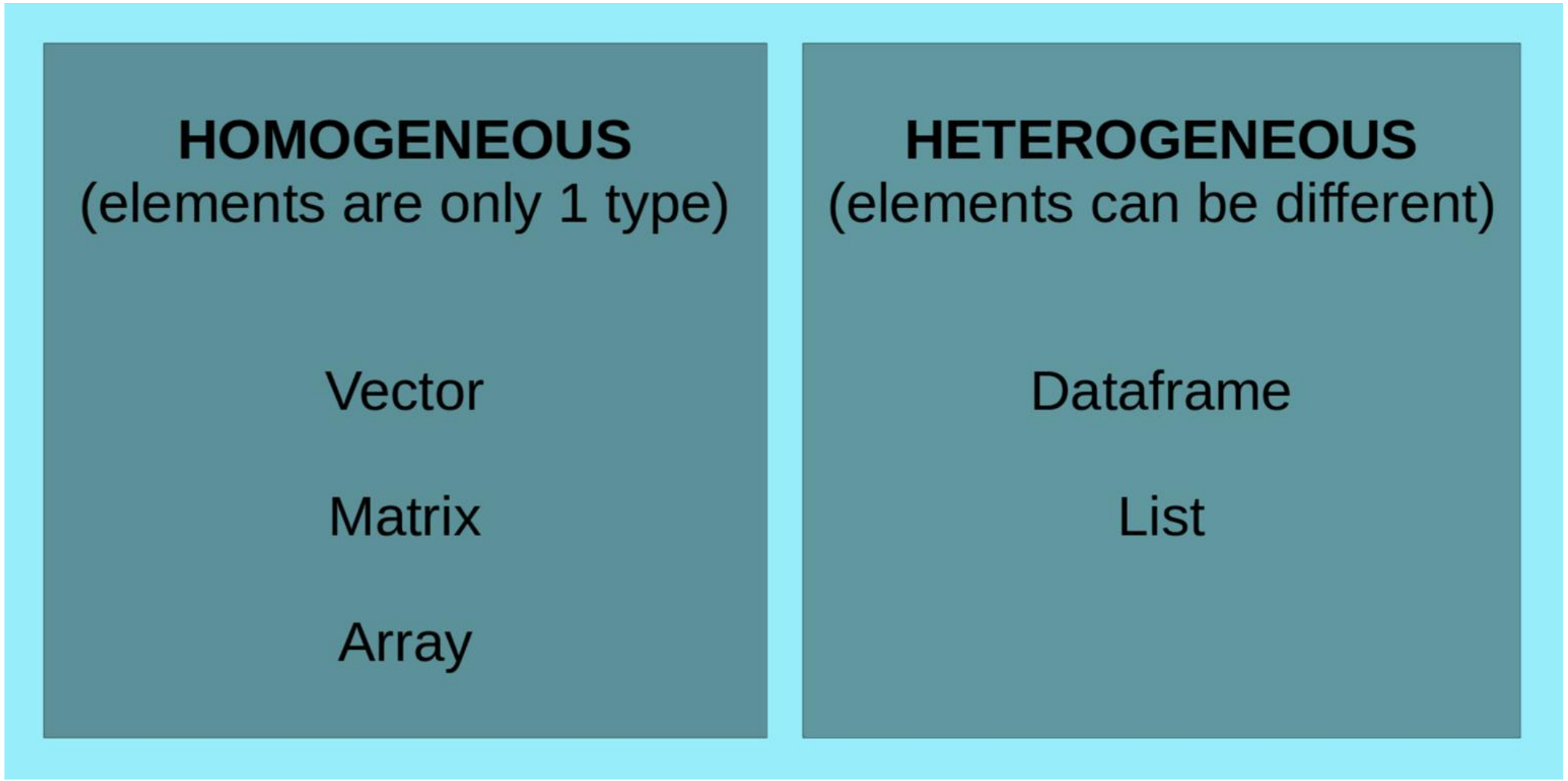
Matrix
Dataframe

n > 2



Array

R Programming; Comparison



R Programming; R-Objects

- The frequently used **R-Objects**:
  - Vectors
  - Lists
  - Matrices
  - Arrays
  - Factors
  - Dataframes

R Programming; R-Objects

Any Questions?