Introduction to R Programming

RealAnalytics

Representing Data using Data Structures

An Overview: Vectors, Matrics and Arrays, Lists, Data Frames, Factors and Tables Why is this important?

- Code errors are many times due data represented as an incorrect data structure
- As a more advanced R user, need to know when creating packages and deploying models etc.

Stories make everything easier to understand. Let's go with Harry Potter

Vectors

Homogenous - logical - character - integer - numeric

```
is_wizard_as_logical_var <- c(TRUE, TRUE, FALSE)
is.logical(is_wizard_as_logical_var)
```

```
## [1] TRUE
```

```
creatures_as_character_var <- c("Skrewt", "Hippogriff")
is.character(creatures_as_character_var)
```

```
## [1] TRUE
```

```
owls_as_integer_var <- c(1L, 6L, 10L)
is.integer(owls_as_integer_var)
```

```
## [1] TRUE
```

```
house_points_as_numeric_var <- c(1000, 2000, 3000)
is.numeric(house_points_as_numeric_var)
```

```
## [1] TRUE
```

Challenge: Represent your own variables as the different types of vectors

Matrics and Arrays

Homogenous

```
a <- matrix(1:6, ncol = 3, nrow = 2)
a
```

```
##      [,1] [,2] [,3]
## [1,]    1    3    5
## [2,]    2    4    6
```

```
b <- array(1:12, c(2, 3, 2))
b
```

```
## , , 1
##
##      [,1] [,2] [,3]
## [1,]    1    3    5
## [2,]    2    4    6
##
## , , 2
##
##      [,1] [,2] [,3]
## [1,]    7    9   11
## [2,]    8   10   12
```

Lists

Heterogenous

```
hermione_subject_scores <- list(c("herbology","muggle studies"),c(95, 99))
hermione_subject_scores
```

```
## [[1]]
## [1] "herbology"      "muggle studies"
##
## [[2]]
## [1] 95 99
```

```
is.list(hermione_subject_scores)
```

```
## [1] TRUE
```

Data Frames

Heterogenous

```
df <- data.frame(x = c(1,2,3,4), houses = c("gryffindor","slytherin", "ravenclaw","
hufflepuff"))
str(df)
```

```
## 'data.frame':    4 obs. of  2 variables:
##  $ x      : num  1 2 3 4
##  $ houses: Factor w/ 4 levels "gryffindor","hufflepuff",..: 1 4 3 2
```