Data Structures in R: Data Frames part 1

Stat 133 with Gaston Sanchez

Lists reminder

single data type *multiple data types*

dimensions

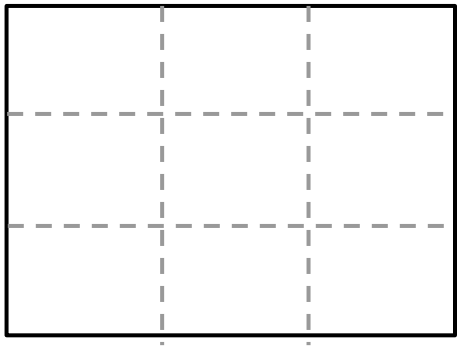
1D

Vector

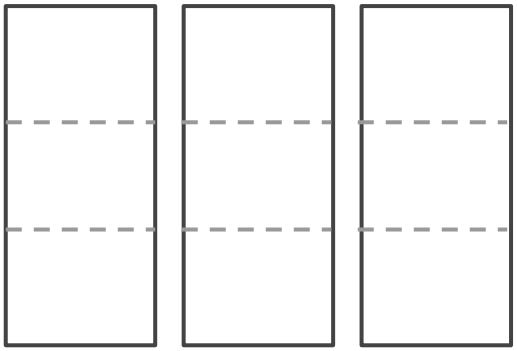
List

2D

Matrix



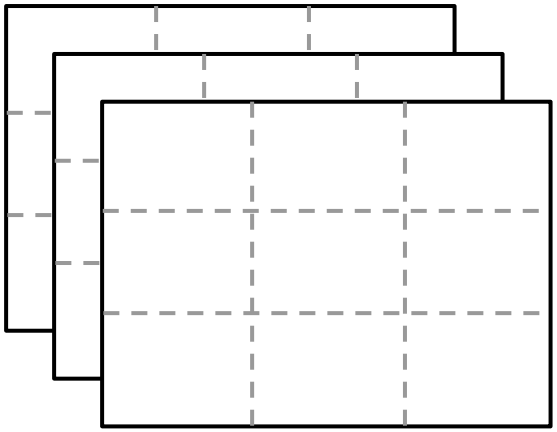
Data Frame



non-atomic structures

nD

Array



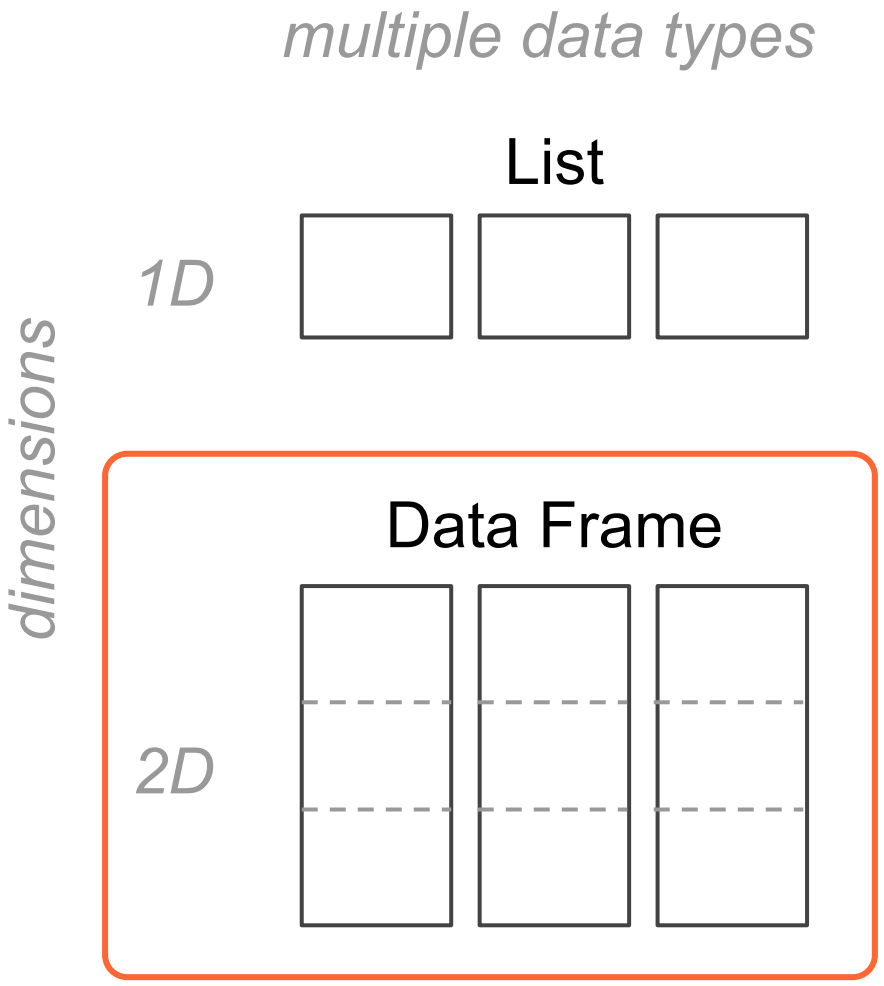
R lists

A list is the most general data structure in R

Lists can contain any other type of data structure

Lists can even contain other lists

Data Frames


multiple data types
List
1D
dimensions
Data Frame
2D

R data frames

A data.frame is the primary data structure that R provides for handling tabular data sets

Creating a data frame

```
# data frame

df <- data.frame(
  name = c('Anakin', 'Padme', 'Luke', 'Leia'),
  gender = c('male', 'female', 'male', 'female'),
  height = c(1.88, 1.65, 1.72, 1.50),
  weight = c(84, 45, 77, 49)
)
```

R data frames

R data frames are special kinds of lists

Stored in R as a list of vectors (or factors)

Columns are typically atomic structures

But since a data frame is a list, you can mix different types of columns

Data frames are NOT matrices but they behave a lot like matrices

There's a bunch of functions to inspect a data.frame object

Function	Description
`str()`	structure
`head()`	First rows
`tail()`	Last rows
`summary()`	Descriptive statistics
`dim()`	Dimensions (# rows, # columns)
`nrow()`	Number of rows
`ncol()`	Number of columns
`names()`	Column names
`colnames()`	Column names
`rownames()`	Row names
`dimnames()`	List with row and column names

Functions to inspect a data frame

```
# display structure
str(airquality)

# display structure but showing
# few elements
str(airquality, vec.len = 1)
```

Functions to inspect a data frame

```
# first n rows
head(airquality, n = 5)

# last n rows
tail(airquality, n = 5)
```

Functions to inspect a data frame

```
# column summaries
summary(airquality)

# memory size
object.size(airquality)

# attributes
attributes(airquality)
```

Functions to inspect a data frame

```
# data frame dimensions
dim(airquality)

# number of rows
nrow(airquality)

# number of columns
ncol(airquality)
```

Functions to inspect a data frame

```
# row names
rownames(airquality)

# column names
colnames(airquality)

# column names
names(airquality)
```

Functions to inspect a data frame

```
# object class ('data.frame')
class(airquality)

# check if object is data.frame
is.data.frame(airquality)

# data.frame is also a list
is.list(airquality)
```

More about data frames in next slides