INTRODUCTION TO R

Explore the Data Frame

Datasets

- Observations
- Variables
- Example: people
  - each person = observation
  - properties (name, age ...) = variables
- Matrix? Need different types
- List? Not very practical

name	age	child
Anne	28	FALSE
Pete	30	TRUE
Frank	21	TRUE
Julia	39	FALSE
Cath	35	TRUE

Data Frame

- Specifically for datasets
- Rows = observations (persons)
- Columns = variables (age, name, ...)
- Contain elements of different types
- Elements in same column: same type

name	age	child
Anne	28	FALSE
Pete	30	TRUE
Frank	21	TRUE
Julia	39	FALSE
Cath	35	TRUE

Create Data Frame

- Import from data source
- CSV file
- Relational Database (e.g. SQL)
- Software packages (Excel, SPSS ...)

Create Data Frame `data.frame()`

```
> name <- c("Anne", "Pete", "Frank", "Julia", "Cath")
> age <- c(28, 30, 21, 39, 35)
> child <- c(FALSE, TRUE, TRUE, FALSE, TRUE)

> df <- data.frame(name, age, child)
```

column names match variable names

```
> df
   name age child
1  Anne  28 FALSE
2  Pete  30  TRUE
3 Frank  21  TRUE
4 Julia  39 FALSE
5  Cath  35  TRUE
```

Name Data Frame

```
> names(df) <- c("Name", "Age", "Child")
> df
   Name Age Child
1  Anne  28 FALSE
2  Pete  30  TRUE
   ...
5  Cath  35  TRUE

> df <- data.frame(Name = name, Age = age, Child = child)
> df
   Name Age Child
1  Anne  28 FALSE
2  Pete  30  TRUE
   ...
5  Cath  35  TRUE
```

Data Frame Structure

Factor instead of character

```
> str(df)
'data.frame':5 obs. of  3 variables:
 $ Name : Factor w/ 5 levels "Anne","Cath",..: 1 5 3 4 2
 $ Age  : num  28 30 21 39 35
 $ Child: logi  FALSE TRUE TRUE FALSE TRUE

> data.frame(name[-1], age, child)
Error : arguments imply differing number of rows: 4, 5

> df <- data.frame(name, age, child,
                   stringsAsFactors = FALSE)
> str(df)
'data.frame':5 obs. of  3 variables:
 $ name : chr  "Anne" "Pete" "Frank" "Julia" ...
 $ age  : num  28 30 21 39 35
 $ child: logi  FALSE TRUE TRUE FALSE TRUE
```

INTRODUCTION TO R

Let’s practice!