

# Data Classes

Introduction to R for Public Health Researchers

## Data Types:

- One dimensional types ('vectors'):
  - Character: strings or individual characters, quoted
  - Numeric: any real number(s)
  - Integer: any integer(s)/whole numbers
  - Factor: categorical/qualitative variables
  - Logical: variables composed of TRUE or FALSE
  - Date/POSIXct: represents calendar dates and times

# Seq

- `seq(from, to, by = )` can create sequences

```
seq(from = 1, to = 5)
```

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

```
seq(from = 1, to = 5, by = 0.1)
```

```
## [1] 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9
## [20] 2.9 3.0 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4.0 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8
## [39] 4.8 4.9 5.0
```

# Logical

`logical` is a type that only has two possible elements: `TRUE` and `FALSE`

```
x = c(TRUE, FALSE, TRUE, TRUE, FALSE)
class(x)
```

```
## [1] "logical"
```

```
z = c("TRUE", "FALSE", "TRUE", "FALSE")
class(z)
```

```
## [1] "character"
```

```
as.logical(z)
```

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

## as. and is. functions

- `as.numeric`, `as.character`, `as.logical`, `as.integer` - "coerces"/changes a vector into that data type - MAY RESULT in NA
- `is.numeric`, `is.character`, `is.logical`, `is.integer` gives a single TRUE or FALSE if that vector is that class

```
is.logical(c(TRUE, FALSE))
```

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

```
is.numeric(c(TRUE, FALSE))
```

```
## [1] FALSE
```

```
as.numeric(c(TRUE, FALSE))
```

```
## [1] 1 0
```

```
as.numeric(c("5", "0", "$0 "))
```

```
## Warning: NAs introduced by coercion
```

```
## [1] 5 0 NA
```

```
as.character(c(TRUE, FALSE))
```

```
## [1] "TRUE" "FALSE"
```

# Factors

A factor is a special character vector where the elements have pre-defined groups or 'levels'. You can think of these as qualitative or categorical variables:

```
x = factor(c("boy", "girl", "girl", "boy", "girl"))
```

```
x
```

```
## [1] boy  girl girl boy  girl  
## Levels: boy girl
```

```
class(x)
```

```
## [1] "factor"
```

Note that levels are, by default, in alphanumerical order.

# Factors

- don't use `as.factor`, use `factor`, even when re-creating a factor
- don't use the `relevel` function. Use the `levels` function to grab the levels if you need.
- The `fct_relevel` function in `forcats` (in `tidyverse`) is fine to use.
- Check out the `forcats` functions `fct_inorder`, `fct_infreq`, `fct_lump`

# Dates

- Use the `lubridate` package - period.
- Change dates using `ymd`, `dmy`, or `mdy` or other combinations.
  - `lubridate` cannot guess this - you also don't want it to
  - If some are `ymd` and others are `dmy`, you need to clean
  - `as_date` also is a good function to try
- Make datetimes using `ymd_hms`, `ymd_hm`, or `ymd_h`
  - `as_datetime` also is a good function to try



Lab

Lab

Website

Website

Website