Data Types and Structures

Different objects in R

Download the section 3 .Rmd handout to `STAT240/lecture/03-data-types`.

Material in this section is covered by Chapters 2, 3 and 4 on the notes website.

Variables

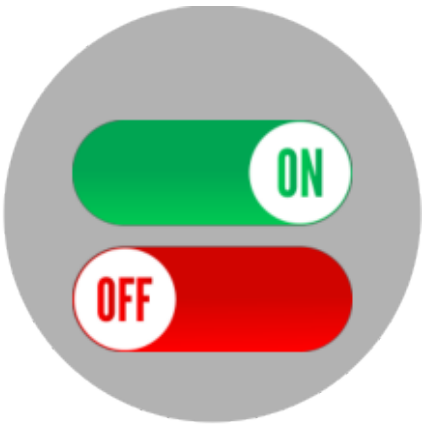
- Created with <- or =
- Capitalization matters!

Use an underscore _ in long variable names.

Basic variable types:

Numeric

Character



Logical

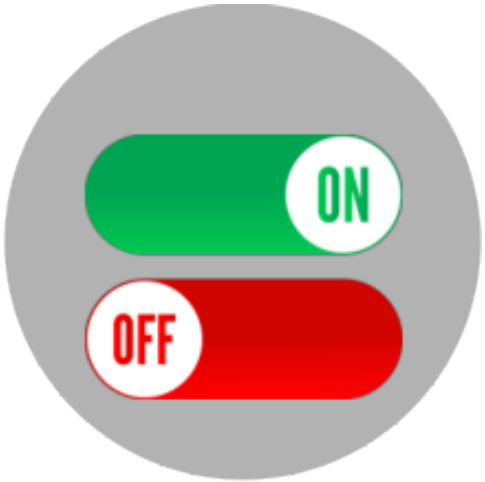
Found with `class()`.

Numeric

- Integer or decimal numbers
- Used with operators `+`, `-`, `*`, etc.

Character

- Interpreted as letters, no special meaning
- Indicated by quotation marks `"like this"`



Logical

- Can be `TRUE` or `FALSE`
- The result of a logic statement

Which type should be used for these variables?

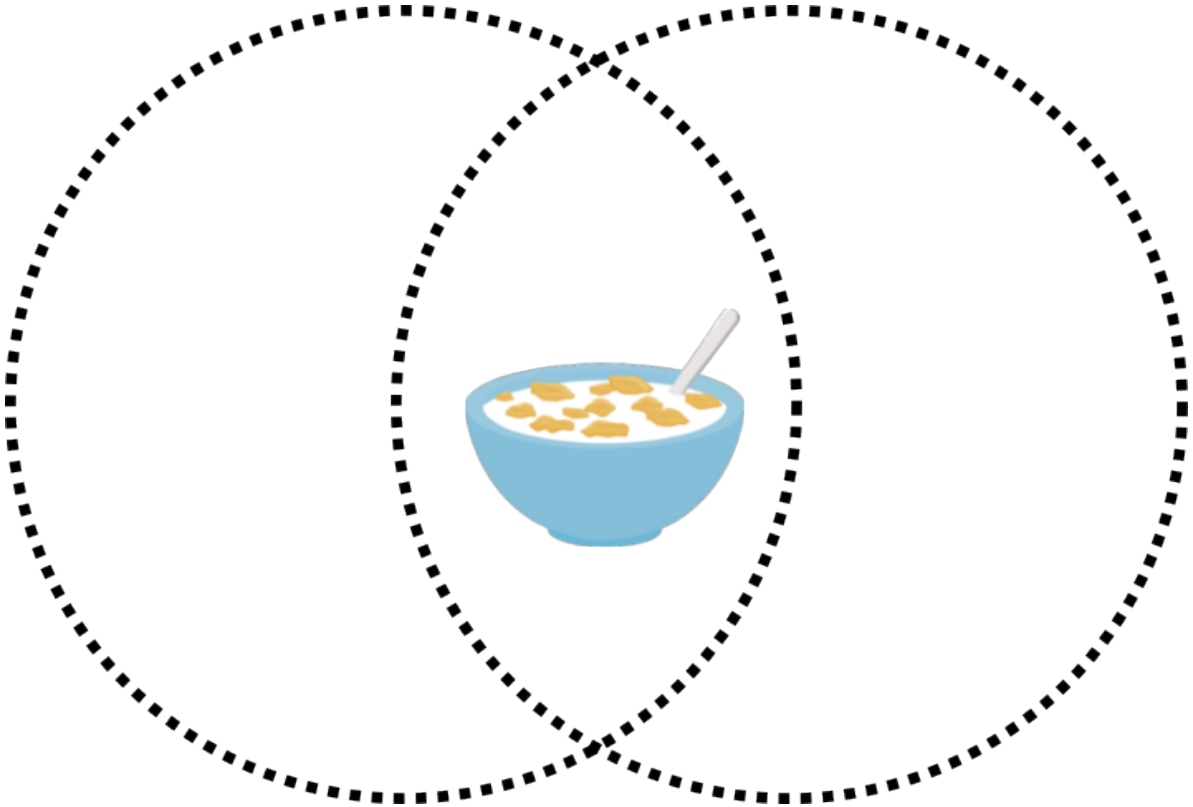
- Name
- Social security number
- Height
- Highest level of education
- If someone voted in last election
- Whether someone is married

We've seen operators for numerics (like +, −).

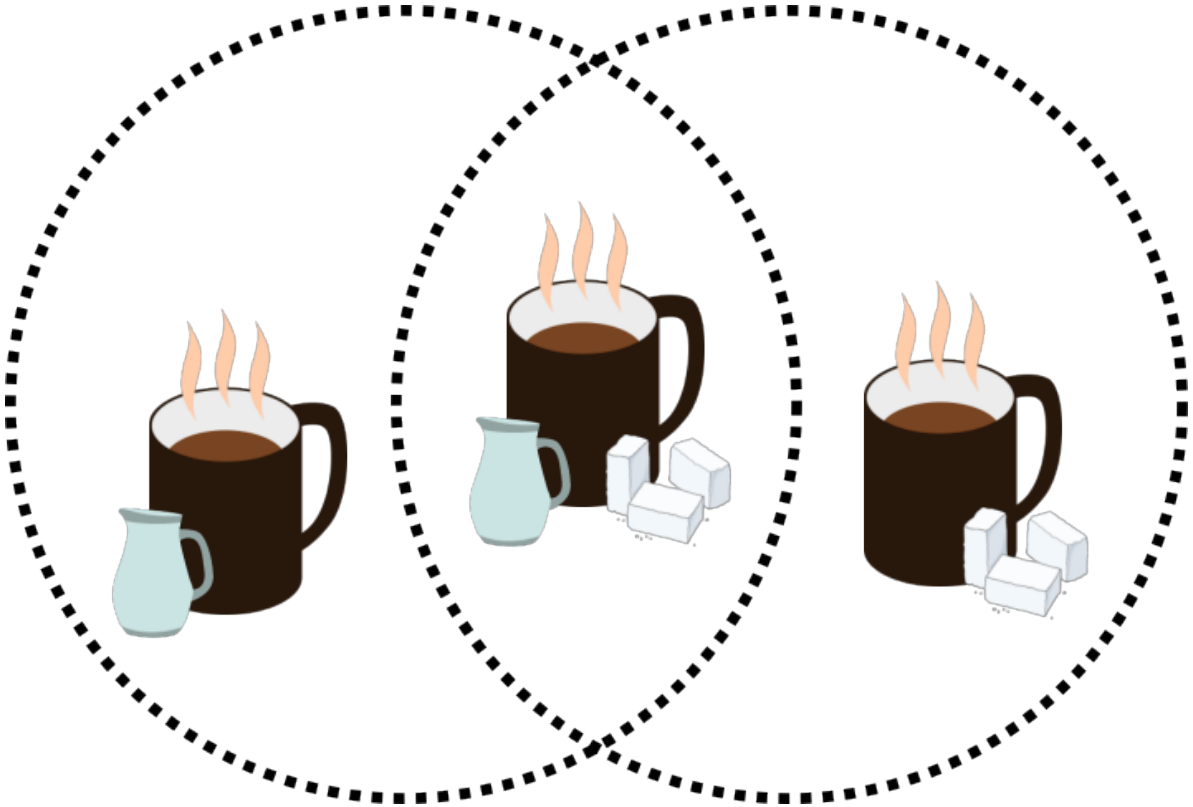
There are also **logical operators**.

- `<`, `>`, `<=`, `>=` to compare numbers
- `==` and `!=` to check equality

- `&` for "and", `|` for "or"



`&` returns `TRUE` if **both** sides are `TRUE`

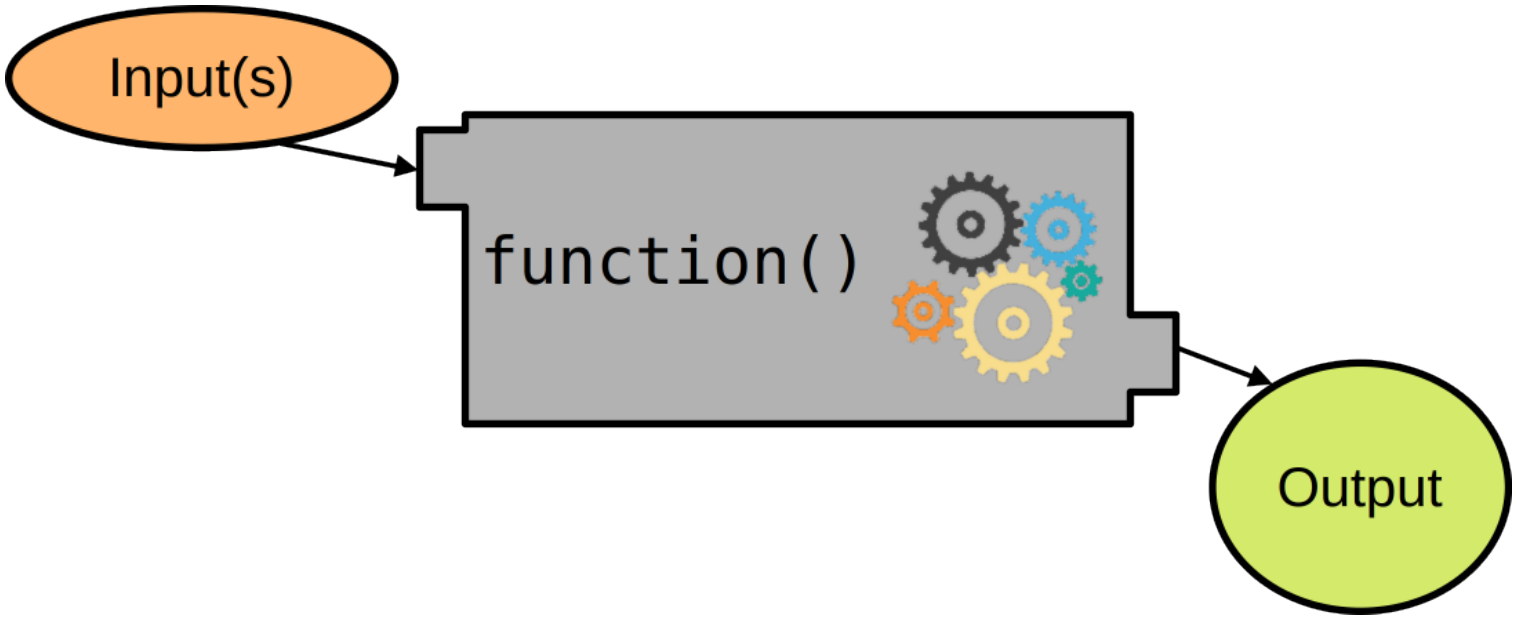


| returns `TRUE` if **either or both** sides are `TRUE`

`NA` refers to an empty or "missing" space.

- Different from 0 and different from `NaN`.
- Doesn't work with regular operators

Check missing values with `is.na`.



Functions are a set of instructions for R.

- `sqrt()`, `class()`

The function's arguments must be the correct type.

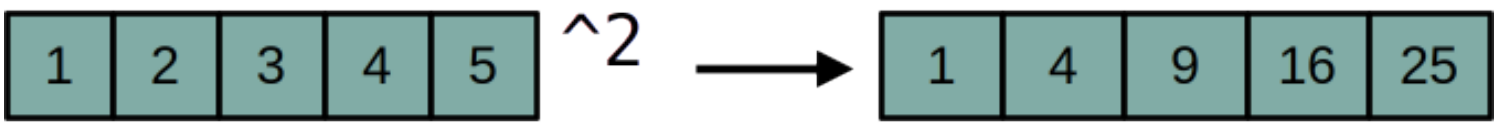
We can save the output to use later.

A **vector** is an ordered sequence of items

- All items must be the same type
- Created with `c()`
- Find an item by index with `[]`.

R does **elementwise** operations with vectors.

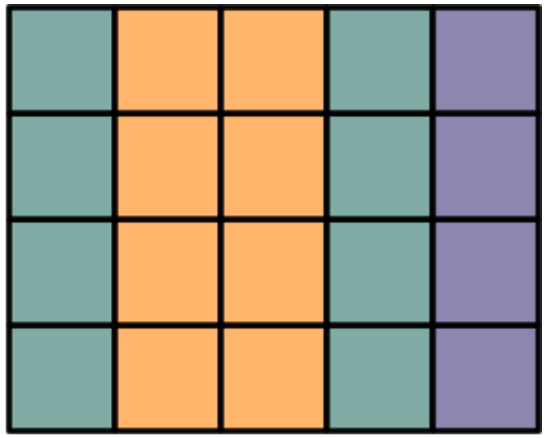
- Works on each element separately



Try stuff out!

- Find the sum of all the odd numbers between 1 and 99.
- Evaluate $[(\frac{1}{2}x - 8)^2 - 20]$ for all of the integers between 0 and 50. For what values of x is the function negative?

Vector



Dataframe

A **dataframe** stores multiple vectors.

- Row = item in sample, column = variable

Create a dataframe with `tibble`.

- Give any number of columns as input
- Columns should have the same length

Let's make a datafame of the alphabet.

Use `df[r, c]` to get a specific value.

- `df[r, ]` to get a specific row
- `df[ , c]` to get a specific column

Get a row by name with `$`.

Now we have the tools to work with real data!

- Run `read_csv` from `tidyverse` to get a dataset on recent volcano eruptions.

We can use `View(eruptions_recent)` and others to get a look at the data.

Test out the functions in the .Rmd.

What does each command do? Which ones do you prefer to use to explore the dataframe?