

Introduction to R

Workshop-ggplot
Lokesh Mano • 28-Apr-2021

exclude: true count: false

name:

content

class:

spaced

##

Contents

* Course

and

webpage

*

Overview

of R *

Data

formats *

Data

frames *

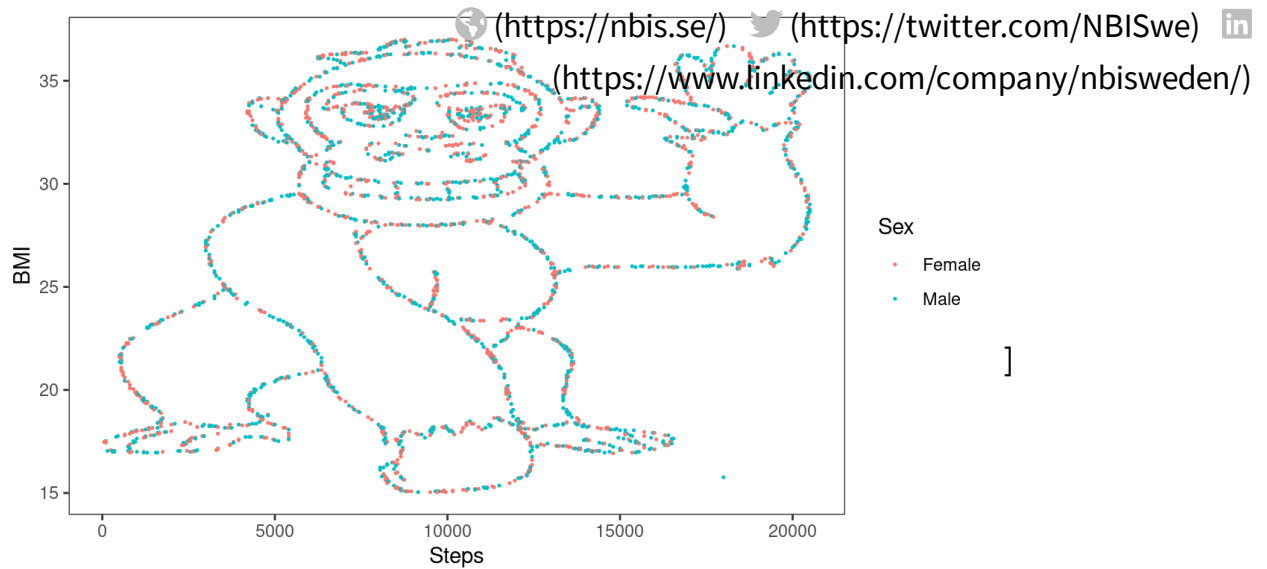
Important

functions

* Tips

name: demo

0.1 Quick checkups



If you don't recognize the correlation you see in the figure above, I would highly recommend you to read the following paper ;)

Yanai, I., Lercher, M. A hypothesis is a liability. *Genome Biol* 21, 231 (2020).

—

- Masks and social-distancing
- Coffe and tea breaks in Ecology
- Webpage structure
- Plots from drop-down
- Times mentioned in schedule are **super** arbitrary

name: r-intro

R

* Derived
 from a
 statistical
 programming
 language
 called **S** *
 You can
 write your
 own
 functions *
 Powerful and
 flexible. *
 Available for
 all platforms

- GUI with **Rstudio**

—

- **RMarkdown:** Embedding codes and results together



(<https://nbis.se/>)



(<https://twitter.com/NBISwe>)



(<https://www.linkedin.com/company/nbisweden/>)

.pull-center[



]

name:

data

Data

Formats

- Wide format

	Sample_1	Sample_2	Sample_3	Sample_4
ENSG00000000003	321	303	204	492
ENSG00000000005	0	0	0	0
ENSG00000000419	696	660	472	951
ENSG00000000457	59	54	44	109
ENSG00000000460	399	405	236	445
ENSG00000000938	0	0	0	0

—

- familiarity
- conveniency
- you see more data

name:

data-2

Data

Formats

- Long

format

Sample_ID	Gene	count
Sample_1	ENSG00000000003	321
Sample_1	ENSG00000000005	0

Sample_ID	Gene	count
Sample_1	ENSG000000000419	696
Sample_1	ENSG000000000457	59
Sample_1	ENSG000000000460	399
Sample_1	ENSG000000000938	0

Sample_ID	Sample_Name	Time	Replicate	Cell	Gene	count
Sample_1	t0_A	t0	A	A431	ENSG000000000003	321
Sample_1	t0_A	t0	A	A431	ENSG000000000005	0
Sample_1	t0_A	t0	A	A431	ENSG000000000419	696
Sample_1	t0_A	t0	A	A431	ENSG000000000457	59
Sample_1	t0_A	t0	A	A431	ENSG000000000460	399
Sample_1	t0_A	t0	A	A431	ENSG000000000938	0

name: data-3

0.2 Data Formats

- Long format

Sample_ID	Sample_Name	Time	Replicate	Cell	Gene	count
Sample_1	t0_A	t0	A	A431	ENSG000000000003	321
Sample_1	t0_A	t0	A	A431	ENSG000000000005	0
Sample_1	t0_A	t0	A	A431	ENSG000000000419	696
Sample_1	t0_A	t0	A	A431	ENSG000000000457	59
Sample_1	t0_A	t0	A	A431	ENSG000000000460	399
Sample_1	t0_A	t0	A	A431	ENSG000000000938	0

- easier to add data to the existing
- Most databases store and maintain in long-formats due to its efficiency
- R tools **like ggplot** require data in long format.



(<https://nbis.se/>)



(<https://twitter.com/NBISwe>)



(<https://www.linkedin.com/company/nbisweden/>)

name: data-
frame

Data
Frames

- Let us take
a quick look
into

`data.frame`
in R :

`.pull-center[`



`]`

* imported

files re

usually in

`data.frame`

* Structured

matrix with

`row.names`

and

`colnames` *

Probably

most used

`data.type`

in Biology!

name: func

0.3 Vectors

```
n <- c(2,3,4,2,1,2,4,5,10,11,8,9)
print(n)
```

```
## [1] 2 3 4 2 1 2 4 5 10 11 8 9
```

-

```
z <- n +3  
print(z)
```

 (<https://nbis.se/>)  (<https://twitter.com/NBISwe>)  (<https://www.linkedin.com/company/nbisweden/>)

```
## [1] 5 6 7 5 4 5 7 8 13 14 11 12
```

—

```
z <- n +3  
mean(z)
```

```
## [1] 8.083333
```

—

```
s <- c("I", "love", "Batman")  
print(s)
```

```
## [1] "I"      "love"    "Batman"
```

name:
vec-typ

Vector
types

* `int`

stands for
integers *

`dbl`

stands for
doubles or
real
numbers *

`chr`

stands for
character
vectors or
strings *

`dtm`

stands for
date and
time, *

`lgl`

stands for
logical
with just
TRUE or
FALSE *

`fctr`

stands for
factors
which R
uses to
state
categorical
variables.

* `date`

stands for
dates

You can
find what
kind of
vectors
you have
or
imported
by using
the
function

`class()`

 (<https://nbis.se/>)  (<https://twitter.com/NBISwe>) 
(<https://www.linkedin.com/company/nbisweden/>)

0.4 Important tips

- `?` and `??`
 - `?` help manual for a particular function
 - `??` searches the entire R library for the term
 - `vignette("ggplot2")`

—

- TAB completion
 - Probably most useful to avoid unnecessary error messages (and/or frustration)!

—

- Case sensitive

```
print(N)
```

```
## Error in print(N): object 'N' not found
```

```
print(n)
```

```
## [1] 2 3 4 2 1 2 4 5 10 11 8 9
```

name: end_slide class: end-slide, middle count: false

1 Thank you. Questions?

.end-text[

R version 4.0.5 (2021-03-31)




Platform: x86_64-pc-linux-gnu (64-bit)

OS: Ubuntu 20.04.2 LTS

Built on : 📅 28-Apr-2021 at ⌚ 02:14:59 PM

2021 • SciLifeLab (<https://www.scilifelab.se/>) • NBIS (<https://nbis.se/>)]

2021 • NBIS • SciLifeLab

 (<https://nbis.se/>)  (<https://twitter.com/NBISwe>) 
(<https://www.linkedin.com/company/nbisweden/>)