Module 4
Manipulate Factors

Rsquared Academy

www.rsquaredacademy.com

Agenda

→ change value of levels

→ add /remove levels

→ combine / collapse levels

→ change order of levels

Resources

→ Slides

→ Code & Data

→ RStudio Cloud

→ Online Course

→ Blog Post

Function	Description
`fct_count()`	Return count/frequency of levels/categories
`fct_unique()`	Return unique level/category labels of a factor
`fct_match()`	Test for presence of levels in a factor

Sort

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

sort = TRUE

Channel	Count
Organic Search	139668
Direct	39853
Referral	35615
Social	8031
Affiliates	7388
(Other)	6073
Paid Search	4395
Display	3375

NOW, IT'S YOUR TURN

Function	Description
`fct_collapse()`	Collapse factor levels
`fct_recode()`	Change factor levels
`fct_lump_min()`	Lump factor levels with count lesser than specified value
`fct_lump_n()`	Lump all levels except the top n levels
`fct_lump_prop()`	Lump factor levels with count lesser than specified proportion
`fct_lump_lowfreq()`	Lump together least frequent levels
`fct_other()`	Replace levels with **Other** level
`fct_anon()`	Anonymize factor levels

Collapse

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Search = c(Paid, Organic)

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Search	144063
Referral	35615
Social	8031

Recode

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Search = "Paid Search"

Search = "Organic Search"

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Search	144063
Referral	35615
Social	8031

Lump

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Lump < 5000 →

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Organic Search	139668
Referral	35615
Social	8031
Other	7770

Lump

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Retain top 3 Channels →

Channel	Count
Direct	39853
Organic Search	139668
Referral	35615
Other	29262

Lump

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Lump < 2% Traffic →

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Organic Search	139668
Referral	35615
Social	8031
Other	7770

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Retain specific Channels →

Channel	Count
Direct	39853
Organic Search	139668
Referral	35615
Other	29262

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Drop specific channels →

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Organic Search	139668
Referral	35615
Social	8031
Other	7770

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Anonymize →

Channel	Count
ch_1	7388
ch_2	139668
ch_3	8031
ch_4	4395
ch_5	35615
ch_6	3375
ch_7	39853
ch_8	6073

NOW, IT'S YOUR TURN

Function	Description
`fct_expand()`	Add additional levels to a factor
`fct_drop()`	Drop unused factor levels
`fct_explicit_na()`	Make missing values explicit

Expand

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Add new channel →

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031
Blog	0

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031
Blog	0

Drop Blog Channel →

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Gender	Count
female	40565
male	61617
<NA>	142216

<NA>

(Missing)

Gender	Count
female	40565
male	61617
(Missing)	142216

NOW, IT'S YOUR TURN

Function	Description
fct_relevel()	Reorder factor levels
fct_shift()	Shift factor levels
fct_infreq()	Reorder factor levels by frequency
fct_rev()	Reverse order of factor levels
fct_inorder()	Reorder factor levels by first appearance
fct_shuffle()	Randomly shuffle factor levels

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

fct_relevel(channel, "Organic Search")

Channel	Count
Organic Search	139668
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Paid Search	4395
Referral	35615
Social	8031

Relevel

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Referral is the third level →

Channel	Count
(Other)	6073
Affiliates	7388
Referral	35615
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Social	8031

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

Display is the last level →

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031
Display	3375

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

fct_infreq(channel) →

Channel	Count
Organic Search	139668
Direct	39853
Referral	35615
Social	8031
Affiliates	7388
(Other)	6073
Paid Search	4395
Display	3375

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

fct_inorder(channel) →

Channel	Count
Organic Search	139668
Direct	39853
Referral	35615
Affiliates	7388
(Other)	6073
Social	8031
Display	3375
Paid Search	4395

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

fct_rev(channel)

Channel	Count
Social	8031
Referral	35615
Paid Search	4395
Organic Search	139668
Display	3375
Direct	39853
Affiliates	7388
(Other)	6073

Channel	Count
(Other)	6073
Affiliates	7388
Direct	39853
Display	3375
Organic Search	139668
Paid Search	4395
Referral	35615
Social	8031

fct_shuffle(channel)

Channel	Count
Direct	39853
(Other)	6073
Organic Search	139668
Social	8031
Affiliates	7388
Referral	35615
Display	3375
Paid Search	4395

NOW, IT'S YOUR TURN

➢ https://forcats.tidyverse.org/

➢ https://r4ds.had.co.nz/factors.html

➢ https://recipes.tidymodels.org/reference/discretize.html

➢ https://ggplot2.tidyverse.org/

➢ https://haleyjeppson.github.io/ggmosaic/

➢ https://rpkgs.datanovia.com/ggpubr/reference/ggdonutchart.html

→ Website

→ Free Online R Courses

→ R Packages

→ Shiny Apps

→ Blog

→ GitHub

→ YouTube

→ Twitter

→ Linkedin

Thank You

For more information please visit our website
www.rsquaredacademy.com