Agenda

- change value of levels
- add or remove levels
- change order of levels

Resources

- Slides
- Data & Scripts
- RStudio Cloud
- Online Course
- Blog Post

Import Data

```
data <- readRDS('data/analytics.rds')
data
```

```
##       device            os                 browser         user_type
## 1    Desktop       Windows                  Chrome       New Visitor
## 2     Mobile           iOS                  Safari Returning Visitor
## 3    Desktop     Chrome OS                  Chrome       New Visitor
## 4    Desktop     Macintosh                  Chrome Returning Visitor
## 5    Desktop     Macintosh                  Chrome Returning Visitor
## 6     Mobile       Android                  Chrome       New Visitor
## 7    Desktop       Windows                  Chrome       New Visitor
## 8    Desktop     Macintosh                  Chrome Returning Visitor
## 9    Desktop       Windows                  Chrome       New Visitor
## 10   Desktop     Chrome OS                  Chrome       New Visitor
## 11   Desktop         Linux                  Chrome       New Visitor
## 12    Tablet       Android                  Chrome Returning Visitor
## 13   Desktop       Windows                  Chrome       New Visitor
## 14   Desktop       Windows                  Chrome       New Visitor
## 15    Mobile       Android                  Chrome       New Visitor
## 16   Desktop       Windows                  Chrome       New Visitor
## 17   Desktop     Macintosh                  Chrome Returning Visitor
## 18   Desktop       Windows                  Chrome       New Visitor
```

Channel Column

```
channel <- data$channel
head(channel)
```

```
## [1] Organic Search Organic Search Direct         Organic Search Referral
## [6] Organic Search
## 8 Levels: (Other) Affiliates Direct Display Organic Search ... Social
```

Tabulate

```
fct_count(channel)
```

```
## # A tibble: 8 x 2
##   f                   n
##   <fct>           <int>
## 1 (Other)          6073
## 2 Affiliates       7388
## 3 Direct          39853
## 4 Display          3375
## 5 Organic Search 139668
## 6 Paid Search      4395
## 7 Referral        35615
## 8 Social           8031
```

Tabulate & 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

Tabulate & Sort

```
fct_count(channel,
          sort = TRUE)
```

```
## # A tibble: 8 x 2
##   f                   n
##   <fct>           <int>
## 1 Organic Search 139668
## 2 Direct          39853
## 3 Referral        35615
## 4 Social           8031
## 5 Affiliates       7388
## 6 (Other)          6073
## 7 Paid Search      4395
## 8 Display          3375
```

Proportion

```
fct_count(channel,
          prop = TRUE)
```

```
## # A tibble: 8 x 3
##   f                   n      p
##   <fct>           <int>  <dbl>
## 1 (Other)          6073 0.0248
## 2 Affiliates       7388 0.0302
## 3 Direct          39853 0.163
## 4 Display          3375 0.0138
## 5 Organic Search 139668 0.571
## 6 Paid Search      4395 0.0180
## 7 Referral        35615 0.146
## 8 Social           8031 0.0329
```

Validity of Levels

```
channel %>%
  fct_match("Social") %>%
  table()
```

```
## .
##  FALSE   TRUE
## 236367   8031
```

Combine Levels

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

Combine Levels

```
# levels to be combined
channel %>%
  fct_collapse(Search = c("Paid Search", "Organic Search")) %>%
  fct_count()
```

```
## # A tibble: 7 x 2
##   f                n
##   <fct>        <int>
## 1 (Other)       6073
## 2 Affiliates    7388
## 3 Direct       39853
## 4 Display       3375
## 5 Search      144063
## 6 Referral     35615
## 7 Social        8031
```

Combine Levels

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

Combine Levels

```
channel %>%
  fct_recode(
    Search = "Paid Search",
    Search = "Organic Search") %>%
  fct_count()
```

```
## # A tibble: 7 x 2
##   f                n
##   <fct>        <int>
## 1 (Other)       6073
## 2 Affiliates    7388
## 3 Direct       39853
## 4 Display       3375
## 5 Search      144063
## 6 Referral     35615
## 7 Social        8031
```

Lump Levels

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 Levels

```
channel %>%
  fct_lump_min(min = 5000) %>%
  fct_count()
```

```
## # A tibble: 7 x 2
##   f                  n
##   <fct>          <int>
## 1 (Other)         6073
## 2 Affiliates      7388
## 3 Direct         39853
## 4 Organic Search 139668
## 5 Referral       35615
## 6 Social          8031
## 7 Other           7770
```

Lump Levels

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 Levels

```
channel %>%
  fct_lump_n(n = 3) %>%
  fct_count()
```

```
## # A tibble: 4 x 2
##   f                   n
##   <fct>           <int>
## 1 Direct          39853
## 2 Organic Search 139668
## 3 Referral        35615
## 4 Other           29262
```

Lump Levels

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

Lump Levels

```
channel %>%
  fct_lump_prop(prop = 0.02) %>%
  fct_count()
```

```
## # A tibble: 7 x 2
##   f                   n
##   <fct>           <int>
## 1 (Other)          6073
## 2 Affiliates       7388
## 3 Direct          39853
## 4 Organic Search 139668
## 5 Referral        35615
## 6 Social           8031
## 7 Other            7770
```

Retain Levels

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

Retain Levels

```
# channels to be retained
retain <- c("Organic Search", "Direct", "Referral")

channel %>%
  fct_other(keep = retain) %>%
  fct_count()
```

```
## # A tibble: 4 x 2
##   f                 n
##   <fct>         <int>
## 1 Direct        39853
## 2 Organic Search 139668
## 3 Referral      35615
## 4 Other         29262
```

Drop Levels

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

Drop Levels

```
# channels to be dropped
dropped <- c("Display", "Paid Search")

channel %>%
  fct_other(drop = dropped) %>%
  fct_count()
```

```
## # A tibble: 7 x 2
##   f                 n
##   <fct>         <int>
## 1 (Other)        6073
## 2 Affiliates     7388
## 3 Direct        39853
## 4 Organic Search 139668
## 5 Referral      35615
## 6 Social         8031
## 7 Other          7770
```

Retain Levels

```
# collapse
channel %>%
  fct_collapse(Other = c("(Other)",
                         "Affiliates",
                         "Display",
                         "Paid Search",
                         "Social")) %>%
  fct_count()
```

```
## # A tibble: 4 x 2
##   f                    n
##   <fct>            <int>
## 1 Other            29262
## 2 Direct           39853
## 3 Organic Search 139668
## 4 Referral         35615
```

Retain Levels

```
# recode
channel %>%
  fct_recode(Other = "(Other)",
             Other = "Affiliates",
             Other = "Display",
             Other = "Paid Search",
             Other = "Social") %>%
  fct_count()
```

```
## # A tibble: 4 x 2
##   f                    n
##   <fct>            <int>
## 1 Other            29262
## 2 Direct           39853
## 3 Organic Search 139668
## 4 Referral         35615
```

Anonymize

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

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

Anonymize

```
channel %>%
  fct_anon(prefix = "ch_") %>%
  fct_count()
```

```
## # A tibble: 8 x 2
##   f           n
##   <fct>   <int>
## 1 ch_1    39853
## 2 ch_2   139668
## 3 ch_3    35615
## 4 ch_4     4395
## 5 ch_5     7388
## 6 ch_6     8031
## 7 ch_7     3375
## 8 ch_8     6073
```

Add new Level

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

Add new Level

```
channel %>%
  fct_expand("Blog") %>%
  levels()
```

```
## [1] "(Other)"        "Affiliates"     "Direct"         "Display"
## [5] "Organic Search" "Paid Search"    "Referral"       "Social"
## [9] "Blog"
```

Drop unused Levels

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

Drop unused Levels

```
channel %>%
  fct_expand("Blog") %>%
  fct_drop() %>%
  levels()
```

```
## [1] "(Other)"        "Affiliates"    "Direct"        "Display"
## [5] "Organic Search" "Paid Search"   "Referral"      "Social"
```

Missing Values

Gender	Count
female	40565
male	61617
<NA>	142216

<NA>

(Missing)

Gender	Count
female	40565
male	61617
(Missing)	142216

Missing Values

```
data %>%
  use_series(gender) %>%
  fct_explicit_na() %>%
  fct_count()
```

```
## # A tibble: 3 x 2
##   f                n
##   <fct>        <int>
## 1 female       40565
## 2 male         61617
## 3 (Missing)   142216
```

Relevel

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

Organic Search is first level

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

Relevel

```
levels(channel)
```

```
## [1] "(Other)"        "Affiliates"     "Direct"         "Display"
## [5] "Organic Search" "Paid Search"    "Referral"       "Social"
```

```
channel %>%
  fct_relevel("Organic Search") %>%
  levels()
```

```
## [1] "Organic Search" "(Other)"        "Affiliates"     "Direct"
## [5] "Display"        "Paid Search"    "Referral"       "Social"
```

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

Relevel

```
levels(channel)
```

```
## [1] "(Other)"        "Affiliates"     "Direct"         "Display"
## [5] "Organic Search" "Paid Search"    "Referral"       "Social"
```

```
channel %>%
  fct_relevel("Referral", after = 2) %>%
  levels()
```

```
## [1] "(Other)"        "Affiliates"     "Referral"       "Direct"
## [5] "Display"        "Organic Search" "Paid Search"    "Social"
```

Relevel

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

Relevel

```
levels(channel)
```

```
## [1] "(Other)"        "Affiliates"     "Direct"         "Display"
## [5] "Organic Search" "Paid Search"    "Referral"       "Social"
```

```
channel %>%
  fct_relevel("Display", after = Inf) %>%
  levels()
```

```
## [1] "(Other)"        "Affiliates"     "Direct"         "Organic Search"
## [5] "Paid Search"    "Referral"       "Social"         "Display"
```

Reorder Levels

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

Reorder Levels

```
levels(channel)
```

```
## [1] "(Other)"        "Affiliates"     "Direct"         "Display"
## [5] "Organic Search" "Paid Search"    "Referral"       "Social"
```

```
# order levels by frequency
channel %>%
  fct_infreq() %>%
  levels()
```

```
## [1] "Organic Search" "Direct"         "Referral"       "Social"
## [5] "Affiliates"     "(Other)"        "Paid Search"    "Display"
```

Reorder Levels

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

Reorder Levels

```
levels(channel)
```

```
## [1] "(Other)"        "Affiliates"    "Direct"        "Display"
## [5] "Organic Search" "Paid Search"   "Referral"      "Social"
```

```
# order levels in order of appearance
channel %>%
  fct_inorder() %>%
  levels()
```

```
## [1] "Organic Search" "Direct"        "Referral"      "Affiliates"
## [5] "(Other)"        "Social"        "Display"       "Paid Search"
```

Reorder Levels

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

Reorder Levels

```
levels(channel)
```

```
## [1] "(Other)"        "Affiliates"     "Direct"         "Display"
## [5] "Organic Search" "Paid Search"    "Referral"       "Social"
```

```
# reverse order of levels
channel %>%
  fct_rev() %>%
  levels()
```

```
## [1] "Social"         "Referral"       "Paid Search"    "Organic Search"
## [5] "Display"        "Direct"         "Affiliates"     "(Other)"
```

Reorder Levels

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

Reorder Levels

```
levels(channel)
```

```
## [1] "(Other)"        "Affiliates"     "Direct"         "Display"
## [5] "Organic Search" "Paid Search"    "Referral"       "Social"
```

```
# randomly shuffle order of levels
channel %>%
  fct_shuffle() %>%
  levels()
```

```
## [1] "Direct"         "Affiliates"     "Paid Search"    "Social"
## [5] "Referral"       "Organic Search" "Display"        "(Other)"
```

References

- 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

Connect with Us

- Website
- Free Online R Courses
- R Packages
- Shiny Apps
- Blog
- eBook
- GitHub
- YouTube
- Twitter
- Linkedin