Agenda

- categories & count

- tables

Resources

- Slides

- Data & Scripts

- RStudio Cloud

- Online Course

- Blog Post

Import Data

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

```
## # A tibble: 6 x 19
##   device    os      browser user_type    channel    gender frequency recency page
##   <fct>     <fct>   <fct>   <fct>        <fct>      <fct>       <dbl>   <dbl>
## 1 Desktop   Windo~  Chrome  New Visit~   Organic~   female          1       0
## 2 Mobile    iOS     Safari  Returning~   Organic~   <NA>            3       1
## 3 Desktop   Chrom~  Chrome  New Visit~   Direct     <NA>            1       0
## 4 Desktop   Macin~  Chrome  Returning~   Organic~   <NA>            2       0
## 5 Desktop   Macin~  Chrome  Returning~   Referral   <NA>            5       8
## 6 Mobile    Andro~  Chrome  New Visit~   Organic~   <NA>            1       0
## # ... with 10 more variables: hour_of_day <chr>, age <dbl>, duration <dbl>
## #   landing_page <fct>, exit_page <fct>, country <fct>, quantity <dbl>,
## #   revenue <dbl>, purchase_flag <lgl>, user_rating <dbl>
```

Number of Categories

```
nlevels(data$device)
```

```
## [1] 3
```

Categories

```
levels(data$device)
```

```
## [1] "Desktop" "Mobile"  "Tablet"
```

Unique Values

```
fct_unique(data$device)
```

```
## [1] Desktop Mobile  Tablet
## Levels: Desktop Mobile Tablet
```

```
fct_unique(data$device)
```

```
## [1] Desktop Mobile  Tablet
## Levels: Desktop Mobile Tablet
```

Tabulate

```
table(data$device)
```

```
## 
## Desktop  Mobile  Tablet 
##  177282   63482    3634 
```

Tabulate

```
fct_count(data$device)
```

```
## # A tibble: 3 x 2
##   f              n
##   <fct>      <int>
## 1 Desktop 177282
## 2 Mobile   63482
## 3 Tablet    3634
```

Tabulate

```
summary(data$device)
```

```
## Desktop   Mobile   Tablet
##  177282    63482     3634
```

Table

```
tab <- table(data$device)
tab
```

```
##
## Desktop  Mobile  Tablet
##  177282   63482    3634
```

Proportions

```
prop.table(tab)
```

```
##
##    Desktop     Mobile     Tablet
## 0.72538237 0.25974844 0.01486919
```

```
proportions(tab)
```

```
##
##    Desktop     Mobile     Tablet
## 0.72538237 0.25974844 0.01486919
```

Percentage

```
proportions(tab) * 100
```

```
##
##    Desktop     Mobile     Tablet
## 72.538237 25.974844  1.486919
```

```
round(proportions(tab) * 100, 2)
```

```
##
## Desktop  Mobile  Tablet
##   72.54   25.97    1.49
```

Cross Table

```
tab2 <- table(data$gender, data$device)
tab2
```

```
##
##            Desktop Mobile Tablet
##    female    32803   7268    494
##    male      46418  14503    696
##    <NA>      98061  41711   2444
```

Order of Variables

```
table(data$device, data$gender)
```

```
##
##            female   male  <NA>
##   Desktop  32803  46418 98061
##   Mobile    7268  14503 41711
##   Tablet     494    696  2444
```

Proportion

```
proportions(tab2)
```

```
##
##            Desktop       Mobile      Tablet
##   female 0.134219593 0.029738378 0.002021293
##   male   0.189927904 0.059341729 0.002847814
##   <NA>   0.401234871 0.170668336 0.010000082
```

```
proportions(tab2) * 100
```

```
##
##            Desktop     Mobile     Tablet
##   female 13.4219593  2.9738378  0.2021293
##   male   18.9927904  5.9341729  0.2847814
##   <NA>   40.1234871 17.0668336  1.0000082
```

Marginal Frequencies

```
margin.table(tab2, 1) # sum of rows
```

```
##
## female    male    <NA>
##  40565   61617  142216
```

```
margin.table(tab2, 2) # sum of columns
```

```
##
## Desktop  Mobile  Tablet
##  177282   63482    3634
```

Sum of all Cells

```
margin.table(tab2)
```

```
## [1] 244398
```

Group Labels

```
dimnames(tab2)
```

```
## [[1]]
## [1] "female" "male"    NA
##
## [[2]]
## [1] "Desktop" "Mobile"  "Tablet"
```

```
names(tab2)
```

```
## NULL
```

```
names(dimnames(tab2))
```

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

Row & Column Labels

```
names(dimnames(tab2)) <- c("Gender", "Device")
tab2
```

```
##         Device
## Gender   Desktop Mobile Tablet
##    female   32803   7268    494
##    male     46418  14503    696
##    <NA>     98061  41711   2444
```

Display Group, Row & Column Labels

```
dimnames(tab2)
```

```
## $Gender
## [1] "female" "male"   NA
##
## $Device
## [1] "Desktop" "Mobile"  "Tablet"
```

Margin Totals

```
addmargins(tab2)
```

```
##         Device
## Gender   Desktop Mobile Tablet     Sum
##   female   32803   7268    494   40565
##   male     46418  14503    696   61617
##   <NA>     98061  41711   2444  142216
##   Sum     177282  63482   3634  244398
```

Row & Column Totals

```
rowSums(tab2)
```

```
## female    male    <NA>
##  40565   61617  142216
```

```
colSums(tab2)
```

```
## Desktop   Mobile   Tablet
##  177282    63482     3634
```

xtabs

```
tabx <- xtabs(~gender+device, data = data)
tabx
```

```
##          device
## gender    Desktop Mobile Tablet
##    female   32803   7268    494
##    male     46418  14503    696
##    <NA>     98061  41711   2444
```

Proportions

```
proportions(tabx)
```

```
##         device
## gender     Desktop      Mobile      Tablet
##   female 0.134219593 0.029738378 0.002021293
##   male   0.189927904 0.059341729 0.002847814
##   <NA>   0.401234871 0.170668336 0.010000082
```

Marginal Frequencies

```
margin.table(tabx, 1)
```

```
## gender
## female   male    <NA>
##  40565  61617 142216
```

```
margin.table(tabx, 2)
```

```
## device
## Desktop  Mobile  Tablet
##  177282   63482    3634
```

Margin Totals

```
addmargins(tabx)
```

```
##          device
## gender    Desktop  Mobile Tablet      Sum
##    female   32803    7268    494    40565
##    male     46418   14503    696    61617
##    <NA>     98061   41711   2444   142216
##    Sum     177282   63482   3634   244398
```

Multi Dimensional Table

```
tab3 <- xtabs(~gender+device+channel, data = data)
tab3
```

```
## , , channel = (Other)
##
##         device
## gender    Desktop Mobile Tablet
##   female      786    258      0
##   male       1063    507     19
##   <NA>       2173   1186     81
##
## , , channel = Affiliates
##
##         device
## gender    Desktop Mobile Tablet
##   female     1314     60      0
##   male       1714    169      0
##   <NA>       3518    548     65
##
## , , channel = Direct
##
##         device
```

Flat Tables

```
ftable(tabx)
```

```
##         device Desktop Mobile Tablet
## gender
## female          32803   7268    494
## male            46418  14503    696
## NA              98061  41711   2444
```

Flat Tables

```
ftable(tab2)
```

```
##         Device Desktop Mobile Tablet
## Gender
## female          32803   7268    494
## male            46418  14503    696
## NA              98061  41711   2444
```

Flat Tables

```
ftable(tab3)
```

```
##                   channel (Other) Affiliates Direct Display Organic Search Pa
## gender device
## female Desktop                786       1314   4785     123          17109
##        Mobile                 258         60    977     753           4480
##        Tablet                   0          0     59     104            282
## male   Desktop               1063       1714   7010     210          25016
##        Mobile                 507        169   2381     491           9563
##        Tablet                  19          0     95      73            448
## NA     Desktop               2173       3518  15824     554          54071
##        Mobile                1186        548   8292     911          27223
##        Tablet                  81         65    430     156           1476
```

Key Functions

- `nlevels()` for number of levels/categories

- Levels/ Categories

  - `levels()`

  - `unique()`

  - `fct_unique()`

- Count / Frequency

  - `table()`

  - `summary()`

  - `fct_count()`

Key Functions

- Two Way / Multidimensional Tables

    - `xtabs()`

    - `ftable()`

    - `gmodels::CrossTable()`

    - `descriptr::ds_cross_table()`

Key Functions - Table

- `proportions()`

- `margin.table()`

- `dimnames()`

- `addmargins()`

- `rowSums()`/`colSums()`

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