rr

Iman

2025-05-24

```r
data(starwars)
str(starwars)
```

```
## tibble [87 x 14] (S3: tbl_df/tbl/data.frame)
##  $ name      : chr [1:87] "Luke Skywalker" "C-3PO" "R2-D2" "Darth Vader" ...
##  $ height    : int [1:87] 172 167 96 202 150 178 165 97 183 182 ...
##  $ mass      : num [1:87] 77 75 32 136 49 120 75 32 84 77 ...
##  $ hair_color: chr [1:87] "blond" NA NA "none" ...
##  $ skin_color: chr [1:87] "fair" "gold" "white, blue" "white" ...
##  $ eye_color : chr [1:87] "blue" "yellow" "red" "yellow" ...
##  $ birth_year: num [1:87] 19 112 33 41.9 19 52 47 NA 24 57 ...
##  $ sex       : chr [1:87] "male" "none" "none" "male" ...
##  $ gender    : chr [1:87] "masculine" "masculine" "masculine" "masculine" ...
##  $ homeworld : chr [1:87] "Tatooine" "Tatooine" "Naboo" "Tatooine" ...
##  $ species   : chr [1:87] "Human" "Droid" "Droid" "Human" ...
##  $ films     :List of 87
##   ..$ : chr [1:5] "A New Hope" "The Empire Strikes Back" "Return of the Jedi" "Revenge of the Sith"
##   ..$ : chr [1:6] "A New Hope" "The Empire Strikes Back" "Return of the Jedi" "The Phantom Menace" .
##   ..$ : chr [1:7] "A New Hope" "The Empire Strikes Back" "Return of the Jedi" "The Phantom Menace" .
##   ..$ : chr [1:4] "A New Hope" "The Empire Strikes Back" "Return of the Jedi" "Revenge of the Sith"
##   ..$ : chr [1:5] "A New Hope" "The Empire Strikes Back" "Return of the Jedi" "Revenge of the Sith"
##   ..$ : chr [1:3] "A New Hope" "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr [1:3] "A New Hope" "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr "A New Hope"
##   ..$ : chr "A New Hope"
##   ..$ : chr [1:6] "A New Hope" "The Empire Strikes Back" "Return of the Jedi" "The Phantom Menace" .
##   ..$ : chr [1:3] "The Phantom Menace" "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr [1:2] "A New Hope" "Revenge of the Sith"
##   ..$ : chr [1:5] "A New Hope" "The Empire Strikes Back" "Return of the Jedi" "Revenge of the Sith"
##   ..$ : chr [1:4] "A New Hope" "The Empire Strikes Back" "Return of the Jedi" "The Force Awakens"
##   ..$ : chr "A New Hope"
##   ..$ : chr [1:3] "A New Hope" "Return of the Jedi" "The Phantom Menace"
##   ..$ : chr [1:3] "A New Hope" "The Empire Strikes Back" "Return of the Jedi"
##   ..$ : chr "A New Hope"
##   ..$ : chr [1:5] "The Empire Strikes Back" "Return of the Jedi" "The Phantom Menace" "Attack of the
##   ..$ : chr [1:5] "The Empire Strikes Back" "Return of the Jedi" "The Phantom Menace" "Attack of the
##   ..$ : chr [1:3] "The Empire Strikes Back" "Return of the Jedi" "Attack of the Clones"
##   ..$ : chr "The Empire Strikes Back"
##   ..$ : chr "The Empire Strikes Back"
##   ..$ : chr [1:2] "The Empire Strikes Back" "Return of the Jedi"
##   ..$ : chr "The Empire Strikes Back"
##   ..$ : chr [1:2] "Return of the Jedi" "The Force Awakens"
##   ..$ : chr "Return of the Jedi"
```

```
##     ..$ : chr "Return of the Jedi"
##     ..$ : chr "Return of the Jedi"
##     ..$ : chr "Return of the Jedi"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr [1:3] "The Phantom Menace" "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr [1:3] "The Phantom Menace" "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr [1:2] "The Phantom Menace" "Attack of the Clones"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr [1:2] "The Phantom Menace" "Attack of the Clones"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr [1:2] "The Phantom Menace" "Attack of the Clones"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr "Return of the Jedi"
##     ..$ : chr [1:3] "The Phantom Menace" "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr [1:3] "The Phantom Menace" "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr [1:3] "The Phantom Menace" "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr [1:3] "The Phantom Menace" "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr [1:2] "The Phantom Menace" "Revenge of the Sith"
##     ..$ : chr [1:2] "The Phantom Menace" "Revenge of the Sith"
##     ..$ : chr [1:2] "The Phantom Menace" "Revenge of the Sith"
##     ..$ : chr "The Phantom Menace"
##     ..$ : chr [1:3] "The Phantom Menace" "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr [1:2] "The Phantom Menace" "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr "Attack of the Clones"
##     ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##     ..$ : chr "Revenge of the Sith"
##     ..$ : chr "Revenge of the Sith"
##     ..$ : chr [1:2] "A New Hope" "Revenge of the Sith"
##     ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
```

```
##   ..$ : chr "Revenge of the Sith"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr "The Force Awakens"
##  $ vehicles  :List of 87
##   ..$ : chr [1:2] "Snowspeeder" "Imperial Speeder Bike"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Imperial Speeder Bike"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Tribubble bongo"
##   ..$ : chr [1:2] "Zephyr-G swoop bike" "XJ-6 airspeeder"
##   ..$ : chr(0)
##   ..$ : chr "AT-ST"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Snowspeeder"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Tribubble bongo"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Sith speeder"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
```

```
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr "Flitknot speeder"
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr "Koro-2 Exodrive airspeeder"
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr "Tsmeu-6 personal wheel bike"
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##  $ starships :List of 87
##    ..$ : chr [1:2] "X-wing" "Imperial shuttle"
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr "TIE Advanced x1"
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr "X-wing"
##    ..$ : chr [1:5] "Jedi starfighter" "Trade Federation cruiser" "Naboo star skiff" "Jedi Interceptor"
##    ..$ : chr [1:3] "Naboo fighter" "Trade Federation cruiser" "Jedi Interceptor"
##    ..$ : chr(0)
##    ..$ : chr [1:2] "Millennium Falcon" "Imperial shuttle"
```

```
##   ..$ : chr [1:2] "Millennium Falcon" "Imperial shuttle"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "X-wing"
##   ..$ : chr "X-wing"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Slave 1"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Millennium Falcon"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "A-wing"
##   ..$ : chr(0)
##   ..$ : chr "Millennium Falcon"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr [1:3] "Naboo fighter" "H-type Nubian yacht" "Naboo star skiff"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Naboo Royal Starship"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Scimitar"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Jedi starfighter"
##   ..$ : chr(0)
##   ..$ : chr "Naboo fighter"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
```

```
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Belbullab-22 starfighter"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "X-wing"
##   ..$ : chr(0)
##   ..$ : chr(0)
```

```r
library(dplyr)
```

```r
#select columns
select(starwars, name, eye_color)
```

```
## # A tibble: 87 x 2
##    name              eye_color
##    <chr>             <chr>
##  1 Luke Skywalker    blue
##  2 C-3PO             yellow
##  3 R2-D2             red
##  4 Darth Vader       yellow
##  5 Leia Organa       brown
##  6 Owen Lars         blue
##  7 Beru Whitesun Lars blue
##  8 R5-D4             red
##  9 Biggs Darklighter brown
## 10 Obi-Wan Kenobi    blue-gray
## # i 77 more rows
```

```r
#select rows
filter(starwars, mass > 70)
```

```
## # A tibble: 39 x 14
##    name      height  mass hair_color skin_color eye_color birth_year sex    gender
##    <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr>  <chr>
##  1 Luke Sk~     172    77 blond      fair       blue              19 male   mascu~
##  2 C-3PO        167    75 <NA>       gold       yellow           112 none   mascu~
##  3 Darth V~     202   136 none       white      yellow          41.9 male   mascu~
##  4 Owen La~     178   120 brown, gr~ light      blue              52 male   mascu~
##  5 Beru Wh~     165    75 brown      light      blue              47 fema~  femin~
##  6 Biggs D~     183    84 black      light      brown             24 male   mascu~
##  7 Obi-Wan~     182    77 auburn, w~ fair       blue-gray         57 male   mascu~
##  8 Anakin ~     188    84 blond      fair       blue            41.9 male   mascu~
##  9 Chewbac~     228   112 brown      unknown    blue             200 male   mascu~
```

```
## 10 Han Solo     180    80 brown      fair       brown            29    male  mascu~
## # i 29 more rows
## # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

```r
#sort rows
arrange(starwars, height)
```

```
## # A tibble: 87 x 14
##    name       height  mass hair_color skin_color eye_color birth_year sex   gender
##    <chr>       <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
##  1 Yoda           66    17 white      green      brown            896 male  mascu~
##  2 Ratts T~       79    15 none       grey, blue unknown           NA male  mascu~
##  3 Wicket ~       88    20 brown      brown      brown              8 male  mascu~
##  4 Dud Bolt       94    45 none       blue, grey yellow            NA male  mascu~
##  5 R2-D2          96    32 <NA>       white, bl~ red               33 none  mascu~
##  6 R4-P17         96    NA none       silver, r~ red, blue         NA none  femin~
##  7 R5-D4          97    32 <NA>       white, red red               NA none  mascu~
##  8 Sebulba       112    40 none       grey, red  orange            NA male  mascu~
##  9 Gasgano       122    NA none       white, bl~ black             NA male  mascu~
## 10 Watto         137    NA black      blue, grey yellow            NA male  mascu~
## # i 77 more rows
## # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

```r
starwars %>% group_by(hair_color) %>% summarise(count = n())
```

```
## # A tibble: 12 x 2
##    hair_color    count
##    <chr>         <int>
##  1 auburn            1
##  2 auburn, grey      1
##  3 auburn, white     1
##  4 black            13
##  5 blond             3
##  6 blonde            1
##  7 brown            18
##  8 brown, grey       1
##  9 grey              1
## 10 none             38
## 11 white             4
## 12 <NA>              5
```

```r
starwars %>% group_by(skin_color) %>% summarise(count = n())
```

```
## # A tibble: 31 x 2
##   skin_color        count
##   <chr>             <int>
## 1 blue                  2
## 2 blue, grey            2
## 3 brown                 4
## 4 brown mottle          1
## 5 brown, white          1
## 6 dark                  6
## 7 fair                 17
## 8 fair, green, yellow   1
```

```
##  9 gold                      1
## 10 green                     6
## # i 21 more rows
```

```
#group data and display count
starwars %>% group_by(sex) %>% summarise(count = n())
```

```
## # A tibble: 5 x 2
##   sex            count
##   <chr>          <int>
## 1 female            16
## 2 hermaphroditic     1
## 3 male              60
## 4 none               6
## 5 <NA>               4
```

```
#add new columns with changes
mutate(starwars, mass2 = mass * 2)
```

```
## # A tibble: 87 x 15
##     name       height  mass hair_color skin_color eye_color birth_year sex    gender
##     <chr>       <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
##  1 Luke Sk~      172    77 blond      fair       blue              19  male  mascu~
##  2 C-3PO         167    75 <NA>       gold       yellow           112  none  mascu~
##  3 R2-D2          96    32 <NA>       white, bl~ red               33  none  mascu~
##  4 Darth V~      202   136 none       white      yellow          41.9  male  mascu~
##  5 Leia Or~      150    49 brown      light      brown             19  fema~ femin~
##  6 Owen La~      178   120 brown, gr~ light      blue              52  male  mascu~
##  7 Beru Wh~      165    75 brown      light      blue              47  fema~ femin~
##  8 R5-D4          97    32 <NA>       white, red red               NA  none  mascu~
##  9 Biggs D~      183    84 black      light      brown             24  male  mascu~
## 10 Obi-Wan~      182    77 auburn, w~ fair       blue-gray         57  male  mascu~
## # i 77 more rows
## # i 6 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>, mass2 <dbl>
```

```
summarise(starwars, avg = mean(height, na.rm = TRUE))
```

```
## # A tibble: 1 x 1
##     avg
##   <dbl>
## 1  175.
```

```
# calculat mean and na.rm = True -> to ignore null values
summarise(starwars, avg = mean(mass, na.rm = TRUE))
```

```
## # A tibble: 1 x 1
##     avg
##   <dbl>
## 1  97.3
```

```
#sum of null values in each column
colSums(is.na(starwars))
```

```
##       name     height       mass hair_color skin_color  eye_color birth_year
##          0          6         28          5          0          0         44
##        sex     gender  homeworld    species      films   vehicles   starships
##          4          4         10          4          0          0          0
```

```r
#replace all null values in birth_year with 0
starwars$birth_year[is.na(starwars$birth_year)] <- 0

#fill null values with height mean in height column
starwars$height[is.na(starwars$height)] <- mean(starwars$height, na.rm = TRUE)

#fill null values with mass median in mass column
starwars$mass[is.na(starwars$mass)] <- median(starwars$mass, na.rm = TRUE)

colSums(is.na(starwars))
```

```
##       name      height        mass hair_color skin_color  eye_color birth_year
##          0           0           0          5          0          0          0
##        sex      gender   homeworld    species      films   vehicles  starships
##          4           4          10          4          0          0          0
```

```r
starwars <- starwars %>% replace_na(list(sex = "unknown", gender = "unknown"))

get_mode <- function(x) {
  ux <- na.omit(unique(x))
  ux[which.max(tabulate(match(x, ux)))]
}

starwars$species[is.na(starwars$species)] <- get_mode(starwars$species)

starwars$homeworld[is.na(starwars$homeworld)] <- get_mode(starwars$homeworld)

colSums(is.na(starwars))
```

```
##       name      height        mass hair_color skin_color  eye_color birth_year
##          0           0           0          5          0          0          0
##        sex      gender   homeworld    species      films   vehicles  starships
##          0           0           0          0          0          0          0
```

```r
#if hair_color is null replace it with unknown using if
starwars <- starwars %>%
  mutate(hair_color = if_else(is.na(hair_color), "unknown", hair_color))

colSums(is.na(starwars))
```

```
##       name      height        mass hair_color skin_color  eye_color birth_year
##          0           0           0          0          0          0          0
##        sex      gender   homeworld    species      films   vehicles  starships
##          0           0           0          0          0          0          0
```