

The inner_join verb

JOINING DATA WITH DPLYR



Chris Cardillo

Data Scientist at DataCamp

LEGO dataset



The sets table

sets

```
# A tibble: 4,977 x 4
  set_num  name                                year theme_id
  <chr>    <chr>                                <dbl>   <dbl>
1 700.3-1  Medium Gift Set (ABB)                   1949     365
2 700.1.1-1 Single 2 x 4 Brick (ABB)                 1950     371
3 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB) 1950     371
4 700.1-2   Extra-Large Gift Set (Mursten)          1953     366
5 700.F-1   Automatic Binding Bricks - Small Brick Set (Lego Mursten) 1953     371
6 700.24-1  Individual 2 x 12 Bricks                  1954     371
7 700.C.1-1 Individual 1 x 6 x 4 Panorama Window (with glass) 1954     371
8 700.C.4-1 Individual 1 x 4 x 3 Window (with glass) 1954     371
9 700.H-1   Individual 4 x 4 Corner Bricks            1954     371
10 1200-1   LEGO Town Plan Board, Large Plastic      1955     372
# ... with 4,967 more rows
```

Linking two tables

sets

```
# A tibble: 4,977 x 4
  set_num  name
  <chr>    <chr>
1 700.3-1  Medium Gift Set (ABB)
2 700.1.1-1 Single 2 x 4 Brick (ABB)
3 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB)
4 700.1-2   Extra-Large Gift Set (Mursten)
5 700.F-1   Automatic Binding Bricks - Small Brick Set
6 700.24-1  Individual 2 x 12 Bricks
7 700.C.1-1 Individual 1 x 6 x 4 Panorama Window (with g
8 700.C.4-1 Individual 1 x 4 x 3 Window (with glass)
9 700.H-1   Individual 4 x 4 Corner Bricks
10 1200-1    LEGO Town Plan Board, Large Plastic
# ... with 4,967 more rows
```

themes

```
# A tibble: 665 x 3
      id name      parent_id
  <dbl> <chr>    <dbl>
1      1 Technic      NA
2      2 Arctic Technic    1
3      3 Competition    1
4      4 Expert Builder    1
5      5 Model      1
6      6 Airport      5
7      7 Construction    5
8      8 Farm      5
9      9 Fire      5
10     10 Harbor      5
# ... with 655 more rows
```

Inner join

```
sets %>%  
  inner_join(themes, by = c("theme_id" = "id"))
```

```
# A tibble: 4,977 x 6  
  set_num name.x year theme_id name.y parent_id  
  <chr>    <chr> <dbl>    <dbl> <chr>    <dbl>  
1 700.3-1 Medium Gift Set (ABB) 1949      365 System      NA  
2 700.1.1-1 Single 2 x 4 Brick (ABB) 1950      371 Supplemental 365  
3 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB) 1950      371 Supplemental 365  
4 700.1-2 Extra-Large Gift Set (Mursten) 1953      366 Basic Set 365  
5 700.F-1 Automatic Binding Bricks - Small Brick Set (Lego Mursten) 1953      371 Supplemental 365  
6 700.24-1 Individual 2 x 12 Bricks 1954      371 Supplemental 365  
7 700.C.1-1 Individual 1 x 6 x 4 Panorama Window (with glass) 1954      371 Supplemental 365  
8 700.C.4-1 Individual 1 x 4 x 3 Window (with glass) 1954      371 Supplemental 365  
9 700.H-1 Individual 4 x 4 Corner Bricks 1954      371 Supplemental 365  
10 1200-1 LEGO Town Plan Board, Large Plastic 1955      372 Town Plan 365  
# ... with 4,967 more rows
```

Customizing your join

```
sets %>%  
  inner_join(themes, by = c("theme_id" = "id"), suffix = c("_set", "_theme"))
```

```
# A tibble: 4,977 x 6  
  set_num name_set year theme_id name_theme parent_id  
  <chr>    <chr>   <dbl>   <dbl> <chr>      <dbl>  
1 700.3-1 Medium Gift Set (ABB) 1949     365 System      NA  
2 700.1.1-1 Single 2 x 4 Brick (ABB) 1950     371 Supplemental 365  
3 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB) 1950     371 Supplemental 365  
4 700.1-2 Extra-Large Gift Set (Mursten) 1953     366 Basic Set 365  
5 700.F-1 Automatic Binding Bricks - Small Brick Set (Lego Mursten) 1953     371 Supplemental 365  
6 700.24-1 Individual 2 x 12 Bricks 1954     371 Supplemental 365  
7 700.C.1-1 Individual 1 x 6 x 4 Panorama Window (with glass) 1954     371 Supplemental 365  
8 700.C.4-1 Individual 1 x 4 x 3 Window (with glass) 1954     371 Supplemental 365  
9 700.H-1 Individual 4 x 4 Corner Bricks 1954     371 Supplemental 365  
10 1200-1 LEGO Town Plan Board, Large Plastic 1955     372 Town Plan 365  
# ... with 4,967 more rows
```

Most common themes

```
sets %>%  
  inner_join(themes, by = c("theme_id" = "id"), suffix = c("_set", "_theme")) %>%  
  count(name_theme, sort = TRUE)
```

```
# A tibble: 419 x 2  
  name_theme      n  
  <chr>         <int>  
1 Supplemental    180  
2 Basic Set       171  
3 Technic        144  
4 Friends        133  
5 Gear           122  
6 City           120  
7 Town           117  
8 Ninjago         95  
9 Service Packs   94  
10 Star Wars      94  
# ... with 409 more rows
```


Other LEGO tables

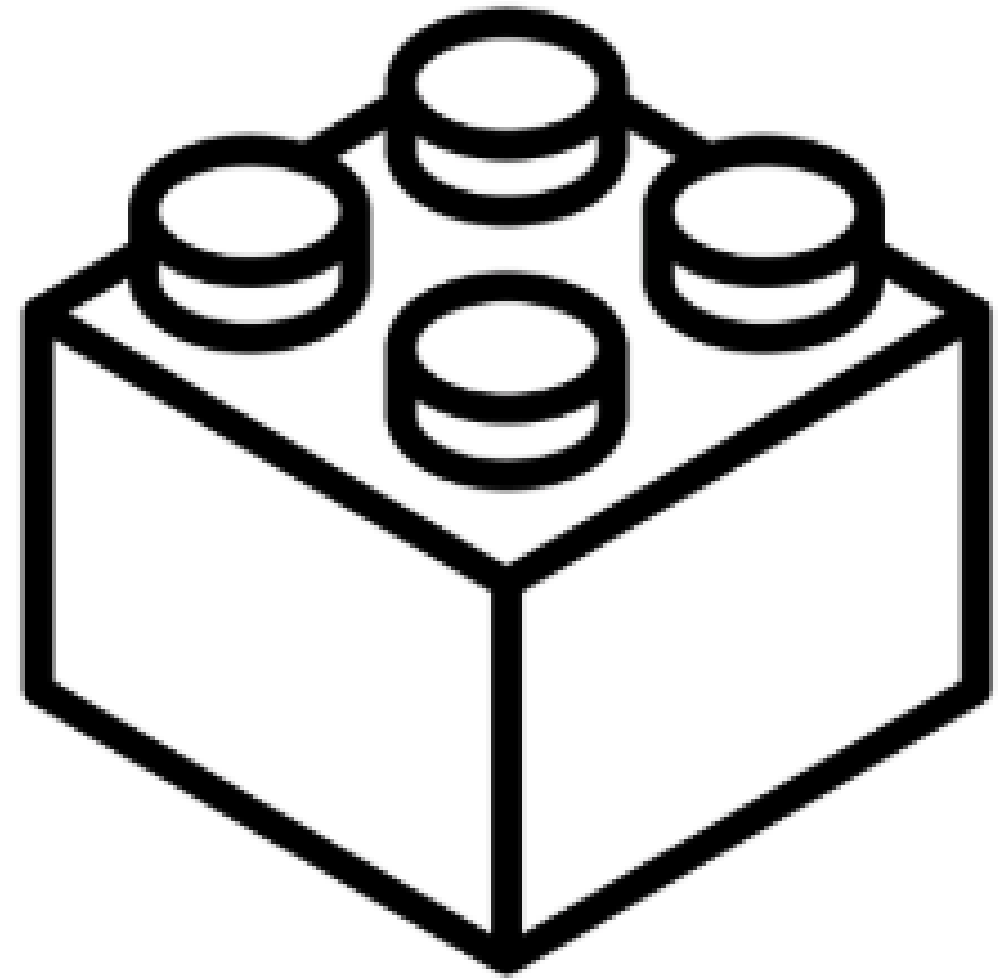
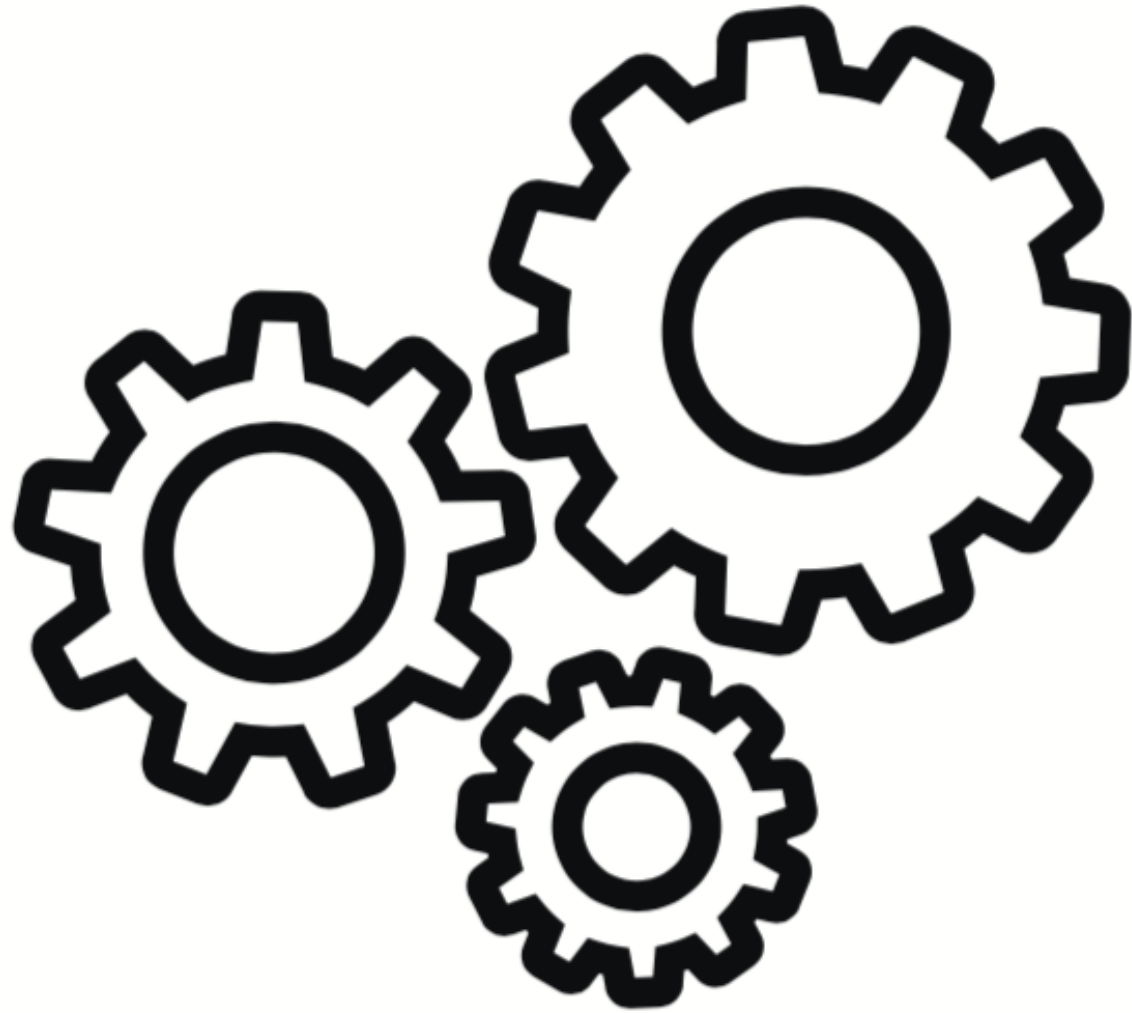
parts

```
# A tibble: 17,501 x 3
  part_num name
  <chr>    <chr>
1 0901    Baseplate 16 x 30 with Set 080 Yellow House
2 0902    Baseplate 16 x 24 with Set 080 Small White
3 0903    Baseplate 16 x 24 with Set 080 Red House P
4 0904    Baseplate 16 x 24 with Set 080 Large White
5 1       Homemaker Bookcase 2 x 4 x 4
6 10016414 Sticker Sheet #1 for 41055-1
7 10026stk01 Sticker for Set 10026 - (44942/4184185)
8 10039    Pullback Motor 8 x 4 x 2/3
9 10048    Minifig Hair Tousled
10 10049    Minifig Shield Broad with Spiked Bottom and
# ... with 17,491 more rows
```

part_categories

```
# A tibble: 64 x 2
      id name
  <dbl> <chr>
1      1 Baseplates
2      3 Bricks Sloped
3      4 Duplo, Quatro and Primo
4      5 Bricks Special
5      6 Bricks Wedged
6      7 Containers
7      8 Technic Bricks
8      9 Plates Special
9     11 Bricks
10    12 Technic Connectors
# ... with 54 more rows
```


Part



Joining with a one-to-many relationship

JOINING DATA WITH DPLYR



Chris Cardillo
Data Scientist at DataCamp

Joining sets and themes

```
sets %>%  
  inner_join(themes, by = c("theme_id" = "id"), suffix = c("_set", "_theme"))
```

```
# A tibble: 4,977 x 6  
  set_num name_set year theme_id name_theme parent_id  
  <chr>    <chr>   <dbl>   <dbl> <chr>      <dbl>  
1 700.3-1 Medium Gift Set (ABB) 1949     365 System      NA  
2 700.1.1-1 Single 2 x 4 Brick (ABB) 1950     371 Supplemental 365  
3 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB) 1950     371 Supplemental 365  
4 700.1-2 Extra-Large Gift Set (Mursten) 1953     366 Basic Set 365  
5 700.F-1 Automatic Binding Bricks - Small Brick Set (Lego Mursten) 1953     371 Supplemental 365  
6 700.24-1 Individual 2 x 12 Bricks 1954     371 Supplemental 365  
7 700.C.1-1 Individual 1 x 6 x 4 Panorama Window (with glass) 1954     371 Supplemental 365  
8 700.C.4-1 Individual 1 x 4 x 3 Window (with glass) 1954     371 Supplemental 365  
9 700.H-1 Individual 4 x 4 Corner Bricks 1954     371 Supplemental 365  
10 1200-1 LEGO Town Plan Board, Large Plastic 1955     372 Town Plan 365  
# ... with 4,967 more rows
```

The inventories table

inventories

```
# A tibble: 15,174 x 3
  id version set_num
  <dbl>   <dbl> <chr>
1     1       1 7922-1
2     3       1 3931-1
3     4       1 6942-1
4    15       1 5158-1
5    16       1 903-1
6    17       1 850950-1
7    19       1 4444-1
8    21       1 3474-1
9    22       1 30277-1
10   25       1 71012-11
# ... with 15,164 more rows
```

Joining sets and inventories

```
sets %>%  
  inner_join(inventories, by = "set_num")
```

```
# A tibble: 5,056 x 6  
  set_num      name      year theme_id    id version  
  <chr>      <chr>      <dbl>   <dbl> <dbl>   <dbl>  
1 700.3-1    Medium Gift Set (ABB)  1949     365 24197     1  
2 700.3-1    Medium Gift Set (ABB)  1949     365 24214     2  
3 700.3-1    Medium Gift Set (ABB)  1949     365 24215     3  
4 700.1.1-1 Single 2 x 4 Brick (ABB)  1950     371 11831     1  
5 700.1.1-1 Single 2 x 4 Brick (ABB)  1950     371 24230     2  
6 700.1.1-1 Single 2 x 4 Brick (ABB)  1950     371 24231     3  
7 700.1.1-1 Single 2 x 4 Brick (ABB)  1950     371 24232     4  
8 700.1.1-1 Single 2 x 4 Brick (ABB)  1950     371 24233     5  
9 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB)  1950     371    537     1  
10 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB)  1950     371 24240     2  
# ... with 5,046 more rows
```

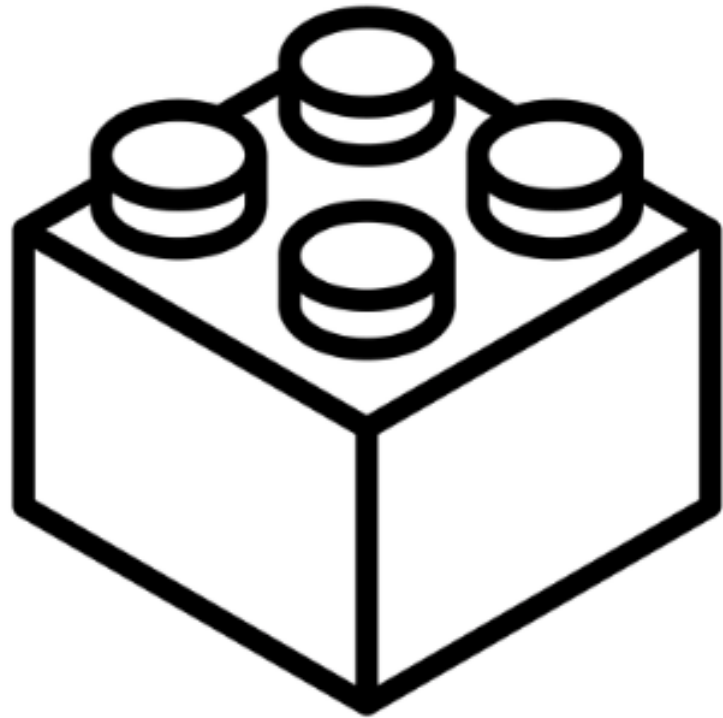
Filtering the joined table

```
sets %>%  
  inner_join(inventories, by = "set_num") %>%  
  filter(version == 1)
```

```
# A tibble: 4,976 x 6  
  set_num    name                                year theme_id    id version  
  <chr>      <chr>                                <dbl>   <dbl> <dbl>   <dbl>  
1 700.3-1    Medium Gift Set (ABB)                   1949     365 24197     1  
2 700.1.1-1  Single 2 x 4 Brick (ABB)                 1950     371 11831     1  
3 700.B.2-1  Single 1 x 2 x 3 Window without Glass (ABB) 1950     371   537     1  
4 700.1-2    Extra-Large Gift Set (Mursten)           1953     366 12985     1  
5 700.F-1    Automatic Binding Bricks - Small Brick Set (Lego Mursten) 1953     371 11265     1  
6 700.24-1   Individual 2 x 12 Bricks                  1954     371  7645     1  
7 700.C.1-1  Individual 1 x 6 x 4 Panorama Window (with glass) 1954     371  3896     1  
8 700.C.4-1  Individual 1 x 4 x 3 Window (with glass)  1954     371  3663     1  
9 700.H-1    Individual 4 x 4 Corner Bricks            1954     371 15503     1  
10 1200-1     LEGO Town Plan Board, Large Plastic       1955     372 10761     1  
# ... with 4,966 more rows
```

Parts and pieces

part



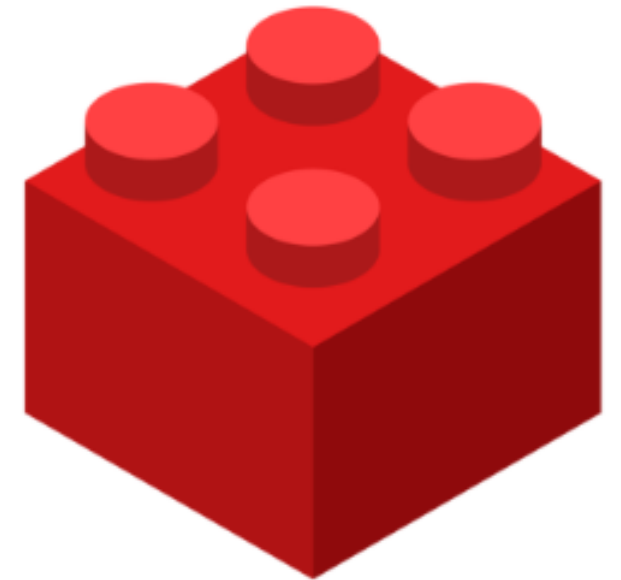
+

color



=

piece



The inventory parts

inventory_parts

```
# A tibble: 258,958 x 4
  inventory_id part_num          color_id quantity
      <dbl>   <chr>          <dbl>     <dbl>
1           21 3009              7         50
2           25 21019c00pat004pr1033  15          1
3           25 24629pr0002          78          1
4           25 24634pr0001           5          1
5           25 24782pr0001           5          1
6           25 88646              0          1
7           25 973pr3314c01          5          1
8           26 14226c11              0          3
9           26 2340px2             15          1
10          26 2340px3             15          1
# ... with 258,948 more rows
```

Joining three or more tables

JOINING DATA WITH DPLYR



Chris Cardillo
Data Scientist at DataCamp

Joining sets and inventories

```
sets %>%  
  inner_join(inventories, by = "set_num")
```

```
# A tibble: 5,056 x 6  
  set_num    name                                year theme_id    id version  
  <chr>      <chr>                                <dbl>   <dbl> <dbl>   <dbl>  
1 700.3-1    Medium Gift Set (ABB)                   1949     365 24197     1  
2 700.3-1    Medium Gift Set (ABB)                   1949     365 24214     2  
3 700.3-1    Medium Gift Set (ABB)                   1949     365 24215     3  
4 700.1.1-1 Single 2 x 4 Brick (ABB)                 1950     371 11831     1  
5 700.1.1-1 Single 2 x 4 Brick (ABB)                 1950     371 24230     2  
6 700.1.1-1 Single 2 x 4 Brick (ABB)                 1950     371 24231     3  
7 700.1.1-1 Single 2 x 4 Brick (ABB)                 1950     371 24232     4  
8 700.1.1-1 Single 2 x 4 Brick (ABB)                 1950     371 24233     5  
9 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB) 1950     371   537     1  
10 700.B.2-1 Single 1 x 2 x 3 Window without Glass (ABB) 1950     371 24240     2  
# ... with 5,046 more rows
```

The themes table

themes

```
# A tibble: 665 x 3
  id name      parent_id
  <dbl> <chr>      <dbl>
1     1 1 Technic      NA
2     2 2 Arctic Technic    1
3     3 3 Competition    1
4     4 4 Expert Builder    1
5     5 5 Model          1
6     6 6 Airport         5
7     7 7 Construction    5
8     8 8 Farm            5
9     9 9 Fire            5
10    10 10 Harbor         5
# ... with 655 more rows
```

Adding another join

```
sets %>%  
  inner_join(inventories, by = "set_num") %>%  
  inner_join(themes, by = c("theme_id" = "id"))
```

```
# A tibble: 5,056 x 8  
  set_num name.x year theme_id id version name.y parent_id  
  <chr>   <chr>   <dbl>   <dbl> <dbl>   <dbl> <chr>      <dbl>  
1 700.3-1 Medium Gift Set (ABB) 1949 365 24197 1 System NA  
2 700.3-1 Medium Gift Set (ABB) 1949 365 24214 2 System NA  
3 700.3-1 Medium Gift Set (ABB) 1949 365 24215 3 System NA  
4 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 11831 1 Supplemen... 365  
5 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 24230 2 Supplemen... 365  
6 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 24231 3 Supplemen... 365  
7 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 24232 4 Supplemen... 365  
8 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 24233 5 Supplemen... 365  
9 700.B.2-1 Single 1 x 2 x 3 Window without ... 1950 371 537 1 Supplemen... 365  
10 700.B.2-1 Single 1 x 2 x 3 Window without ... 1950 371 24240 2 Supplemen... 365  
# ... with 5,046 more rows
```

Recall: suffix

```
sets %>%
  inner_join(inventories, by = "set_num") %>%
  inner_join(themes, by = c("theme_id" = "id"), suffix = c("_set", "_theme"))
```

```
# A tibble: 5,056 x 8
  set_num name_set year theme_id id version name_theme parent_id
  <chr>    <chr>   <dbl>   <dbl> <dbl>   <dbl> <chr>         <dbl>
1 700.3-1 Medium Gift Set (ABB) 1949 365 24197 1 System NA
2 700.3-1 Medium Gift Set (ABB) 1949 365 24214 2 System NA
3 700.3-1 Medium Gift Set (ABB) 1949 365 24215 3 System NA
4 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 11831 1 Supplement... 365
5 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 24230 2 Supplement... 365
6 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 24231 3 Supplement... 365
7 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 24232 4 Supplement... 365
8 700.1.1-1 Single 2 x 4 Brick (ABB) 1950 371 24233 5 Supplement... 365
9 700.B.2-1 Single 1 x 2 x 3 Window without... 1950 371 537 1 Supplement... 365
10 700.B.2-1 Single 1 x 2 x 3 Window without... 1950 371 24240 2 Supplement... 365
# ... with 5,046 more rows
```

