

Strava analysis - deel 2

Over de data

Strava is een social platform waar je ritgegevens, hardloopegegevens etc in kan bijhouden. in deze opgave maken we gebruik van een beperkte dataset uit dit platform # Setup van document Inlezen libraries

Inlezen data

Vragen

In dit notebook gaan we vooral met tijdgebaseerde gegevens werken. Bij het inlezen heeft R al zelf bepaald dat de kolom `start_date` het datatype `POSIXct` (datum/tijd) is. we hebben echter nog een tijdsgebaseerde kolom: `moving_time` de functie `hms::as_hms` kan van een getal een hour/minute/second maken:

1. Maak een nieuwe kolom aan met de tijdsduur in uren/minuten/seconden

```
## # A tibble: 1,361 x 7
##       X1 distance moving_time start_date      gear_id type  timespan
##   <dbl>    <dbl>      <dbl> <dtm>      <chr>    <chr> <time>
## 1     1    313156      4115 2021-03-03 15:08:31 b6734226 Ride 01:08:35
## 2     2    315482      4084 2021-03-01 15:06:21 b6734226 Ride 01:08:04
## 3     3    530608      6911 2021-02-28 12:13:45 b6734226 Ride 01:55:11
## 4     4    542308      7166 2021-02-26 08:18:43 b6734226 Ride 01:59:26
## 5     5    412496      5426 2021-02-24 14:20:02 b6734226 Ride 01:30:26
## 6     6    433972      5531 2021-02-23 13:37:52 b6734226 Ride 01:32:11
## 7     7     40579      2779 2021-02-22 14:37:46 <NA>      Walk 00:46:19
## 8     8    285415      6343 2021-02-21 08:14:13 b5957471 Ride 01:45:43
## 9     9    532266      9111 2021-02-20 12:06:03 b5957471 Ride 02:31:51
## 10    10    289046      5391 2021-02-19 12:14:34 b5957471 Ride 01:29:51
## # ... with 1,351 more rows
```

2. Duur activiteiten per per kwartaal

Per activiteit will we het aantal per kwartaal weten. Daarvoor heb je de `quarter` functie nodig icm een aantal bekende functies. alleen activiteiten die meer dan 3x zijn gedaan hoeven meegenomen te worden & sorteer op type - Bepaal per activiteit in welk kwartaal deze plaatsvond - groepeer op 2 items - tel iets op & formatteer weer als hms - filter - sorteer Wat kun je concluderen?

```
## 'summarise()' has grouped output by 'type'. You can override using the '.groups' argument.
```

```
## # A tibble: 24 x 3
## # Groups:   type [9]
##   type      kwartaal totaal
```

```
##      <chr>          <int> <time>
## 1 Ride              3 789:07:06
## 2 Ride              2 728:09:25
## 3 Ride              4 324:06:51
## 4 Ride              1 301:25:18
## 5 VirtualRide       4  67:46:37
## 6 VirtualRide       1  55:55:22
## 7 Walk              4  26:47:12
## 8 IceSkate          4  23:25:00
## 9 IceSkate          1  22:59:57
## 10 Hike             3  22:15:32
## # ... with 14 more rows
```

3. Gebruik van gear per maand in kwartaal 4

We hebben nu een functie nodig om de maand te extraheren en gaan per & per gear maand tellen - nieuwe kolommen toevoegen - filteren - lege rijen weghalen - groeperen - tellen Wat concludeer je?

```
## # A tibble: 24 x 3
## # Groups:   gear_id, maand [24]
##   gear_id maand     n
##   <chr>    <dbl> <int>
## 1 b2992308    10     32
## 2 b2992308    11     27
## 3 b2992308    12     23
## 4 b1289756    10     22
## 5 b6734226    10     16
## 6 b1289756    11     14
## 7 b6734226    12     14
## 8 b1289756    12     13
## 9 b5957471    11     13
## 10 b5957471    12     12
## # ... with 14 more rows
```

4. Wat zijn de 2 meest recent aangeschafte fietsen en hoe lang gaan ze mee?

- tijd tussen activiteit en **vandaag** bepalen mbv `difftime`
- minimaal verschil per `gear_id`

```
## # A tibble: 11 x 3
## # Groups:   gear_id [11]
##   gear_id start_date      verschil
##   <chr>    <dtm>          <drtn>
## 1 b6734226 2020-02-14 11:47:45 -410.5085 days
## 2 b5957471 2019-06-12 18:06:07 -657.2458 days
## 3 b3995755 2017-07-07 12:00:01 -1362.5000 days
## 4 b3079109 2016-07-25 08:18:15 -1709.6540 days
## 5 b2992308 2015-10-16 17:41:47 -1992.2627 days
## 6 b2442923 2015-10-04 08:26:42 -2004.6481 days
## 7 b2266483 2015-08-20 09:18:46 -2049.6120 days
## 8 b1350013 2014-06-08 09:48:50 -2487.5911 days
## 9 b1289756 2014-05-10 16:24:29 -2516.3163 days
```

10 b716318 2013-04-30 10:09:58 -2891.5764 days
11 <NA> 2013-02-17 11:45:13 -2963.5103 days