Tidyverse lecture - Solutions

datalab

2022-05-20

Exercise A

1. Make a new R script
2. Load the climate data which you can find in the Exercises folder

```
library(tidyverse)
library(readxl)
climate <- read_excel("climate.xlsx")
climate
```

```
## # A tibble: 60 x 7
##    station  year month    af  rain   sun device
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
##  1 armagh   2016     1     5 132.   44.5 Campbell Stokes
##  2 armagh   2016     2    10  62.6  71.3 Campbell Stokes
##  3 armagh   2016     3     4  43.8 117.  Campbell Stokes
##  4 armagh   2016     4     5  54   140.  Campbell Stokes
##  5 armagh   2016     5     0  41.4 210.  Campbell Stokes
##  6 armagh   2016     6     0  75.1 114.  Campbell Stokes
##  7 armagh   2016     7     0  80.6 113.  Campbell Stokes
##  8 armagh   2016     8     0  52.5 135.  Campbell Stokes
##  9 armagh   2016     9     0  65.4  91.1 Campbell Stokes
## 10 armagh   2016    10     0  37.1  89.8 Campbell Stokes
## # ... with 50 more rows
```

3. Is your data a dataframe or a tibble?

```
class(climate)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

It's a tibble.

4. Have a look at:

4.1 How many observations are there?

```
dim(climate)
```

```
## [1] 60  7
```

4.2 How many data columns are there and what are their types?

```
summary(climate)
```

```
##    station              year          month            af
##  Length:60          Min.   :2016   Min.   : 1.00   Min.   : 0.000
##  Class :character   1st Qu.:2016   1st Qu.: 3.75   1st Qu.: 0.000
##  Mode  :character   Median :2016   Median : 6.50   Median : 0.000
```

```
##                        Mean   :2016   Mean   : 6.50   Mean   : 1.683
##                        3rd Qu.:2016   3rd Qu.: 9.25   3rd Qu.: 3.000
##                        Max.   :2016   Max.   :12.00   Max.   :10.000
##       rain            sun           device
##  Min.   :  3.4   Min.   : 11.5   Length:60
##  1st Qu.: 45.9   1st Qu.: 74.6   Class :character
##  Median : 64.0   Median :109.4   Mode  :character
##  Mean   : 75.8   Mean   :113.8
##  3rd Qu.: 90.2   3rd Qu.:155.3
##  Max.   :221.6   Max.   :227.6
```

4.3 How many different stations are there?

```
distinct(climate,station)
```

```
## # A tibble: 5 x 1
##   station
##   <chr>
## 1 armagh
## 2 camborne
## 3 lerwick
## 4 oxford
## 5 sheffield
```

4.4 How many observations per station?

```
climate %>% count(station)
```

```
## # A tibble: 5 x 2
##   station       n
##   <chr>     <int>
## 1 armagh       12
## 2 camborne     12
## 3 lerwick      12
## 4 oxford       12
## 5 sheffield    12
```

Exercise B: 5-10 mins

On the climate dataset, select:

1. all rows from the station in Oxford

```
climate %>%
  filter(station == 'oxford')
```

```
## # A tibble: 12 x 7
##    station  year month    af  rain   sun device
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
##  1 oxford   2016     1     5  83.9  59.1 Campbell Stokes
##  2 oxford   2016     2     6  47.6 113.  Campbell Stokes
##  3 oxford   2016     3     4  74.2 124.  Campbell Stokes
##  4 oxford   2016     4     1  53.1 164.  Campbell Stokes
##  5 oxford   2016     5     0  86.1 203.  Campbell Stokes
##  6 oxford   2016     6     0  95.7 100.  Campbell Stokes
##  7 oxford   2016     7     0   3.4 228.  Campbell Stokes
##  8 oxford   2016     8     0  41.2 204.  Campbell Stokes
##  9 oxford   2016     9     0  44.6 113.  Campbell Stokes
```

```
## 10 oxford    2016    10     0  26.5 112.  Campbell Stokes
## 11 oxford    2016    11     3  76.1  88.3 Campbell Stokes
## 12 oxford    2016    12    10  25.8  62.3 Campbell Stokes
```

2. all rows from the station in Oxford when there were at least 100 hours of sunlight

```
climate %>%
  filter(station == 'oxford' & sun >= 100)
```

```
## # A tibble: 9 x 7
##    station  year month    af   rain    sun device
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
## 1 oxford    2016     2     6  47.6  113. Campbell Stokes
## 2 oxford    2016     3     4  74.2  124. Campbell Stokes
## 3 oxford    2016     4     1  53.1  164. Campbell Stokes
## 4 oxford    2016     5     0  86.1  203. Campbell Stokes
## 5 oxford    2016     6     0  95.7  100. Campbell Stokes
## 6 oxford    2016     7     0   3.4  228. Campbell Stokes
## 7 oxford    2016     8     0  41.2  204. Campbell Stokes
## 8 oxford    2016     9     0  44.6  113. Campbell Stokes
## 9 oxford    2016    10     0  26.5  112. Campbell Stokes
```

3. all rows from the stations in Oxford and Camborne when there were at least 100 hours of sunlight

```
climate %>%
  filter(station %in% c('oxford','camborne') & sun >= 100)
```

```
## # A tibble: 17 x 7
##     station   year month    af   rain    sun device
##     <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
##  1 camborne 2016     3     0  88.4  140. Kipp Zonen
##  2 camborne 2016     4     0  81.4  184. Kipp Zonen
##  3 camborne 2016     5     0  45.6  206. Kipp Zonen
##  4 camborne 2016     6     0  65.8  132. Kipp Zonen
##  5 camborne 2016     7     0  23.2  161. Kipp Zonen
##  6 camborne 2016     8     0  57.4  171. Kipp Zonen
##  7 camborne 2016     9     0 154.   103. Kipp Zonen
##  8 camborne 2016    10     0  53.2  125. Kipp Zonen
##  9 oxford   2016     2     6  47.6  113. Campbell Stokes
## 10 oxford   2016     3     4  74.2  124. Campbell Stokes
## 11 oxford   2016     4     1  53.1  164. Campbell Stokes
## 12 oxford   2016     5     0  86.1  203. Campbell Stokes
## 13 oxford   2016     6     0  95.7  100. Campbell Stokes
## 14 oxford   2016     7     0   3.4  228. Campbell Stokes
## 15 oxford   2016     8     0  41.2  204. Campbell Stokes
## 16 oxford   2016     9     0  44.6  113. Campbell Stokes
## 17 oxford   2016    10     0  26.5  112. Campbell Stokes
```

4. a subset that only contains the station, year and rain columns

```
sub <- climate %>%
  select(station, year, rain)
sub
```

```
## # A tibble: 60 x 3
##    station  year  rain
##    <chr>   <dbl> <dbl>
##  1 armagh   2016 132.
```

```
##  2 armagh   2016  62.6
##  3 armagh   2016  43.8
##  4 armagh   2016  54
##  5 armagh   2016  41.4
##  6 armagh   2016  75.1
##  7 armagh   2016  80.6
##  8 armagh   2016  52.5
##  9 armagh   2016  65.4
## 10 armagh   2016  37.1
## # ... with 50 more rows
```

Exercise C

1. To the climate dataset, add:

1.1 A column that states the amount of hours with no sunshine for each month. A month has on average 730 hours, you can use the same amount of hours of all of them.

```
climate <- climate %>%
  mutate(no_sun = 730 - sun)
```

1.2. A column the says whether the weather this month was good. We consider good a month with at least 100 hours of sunlight and less than 100 mm of rain. Otherwise the weather was bad.

```
climate <- climate %>%
  mutate(good_weather = ifelse(sun > 100 & rain < 100, "Yes", "No"))
climate
```

```
## # A tibble: 60 x 9
##    station  year month    af  rain   sun device          no_sun good_weather
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <chr>            <dbl> <chr>
##  1 armagh   2016     1     5 132.   44.5 Campbell Stokes   686. No
##  2 armagh   2016     2    10  62.6  71.3 Campbell Stokes   659. No
##  3 armagh   2016     3     4  43.8 117.  Campbell Stokes   613. Yes
##  4 armagh   2016     4     5  54   140.  Campbell Stokes   590. Yes
##  5 armagh   2016     5     0  41.4 210.  Campbell Stokes   520. Yes
##  6 armagh   2016     6     0  75.1 114.  Campbell Stokes   616. Yes
##  7 armagh   2016     7     0  80.6 113.  Campbell Stokes   617. Yes
##  8 armagh   2016     8     0  52.5 135.  Campbell Stokes   595. Yes
##  9 armagh   2016     9     0  65.4  91.1 Campbell Stokes   639. No
## 10 armagh   2016    10     0  37.1  89.8 Campbell Stokes   640. No
## # ... with 50 more rows
```

2. Count the number of:

2.1 Months, i.e. lines, per station that did not have any days with air frost (so two conditions)

```
climate %>%
  count(station, af == 0)
```

```
## # A tibble: 9 x 3
##    station  `af == 0`     n
##    <chr>    <lgl>     <int>
## # 1 armagh   FALSE         6
## # 2 armagh   TRUE          6
## # 3 camborne TRUE         12
## # 4 lerwick  FALSE         5
## # 5 lerwick  TRUE          7
```

```
## 6 oxford    FALSE        6
## 7 oxford    TRUE         6
## 8 sheffield FALSE        6
## 9 sheffield TRUE         6
```

2.2 Months with good weather per station (use the column you made in 1.2). What's the place with the best weather in England?

```
climate %>%
  filter(good_weather == 'Yes') %>%
  count(station) %>%
  arrange(n)
```

```
## # A tibble: 5 x 2
##    station       n
##    <chr>     <int>
## 1 lerwick       4
## 2 sheffield     5
## 3 armagh        6
## 4 camborne      7
## 5 oxford        9
```

Oxford has the most months with good weather (9).