Code together: From messy data to insightful visualizations

R-Ladies Frankfurt Meetup #4

11th of July 2019

Artificial HIV-dataset

- Source: Github https://github.com/NFilmann/RLadiesFRA
  - Datasets basedata.csv, labdata.csv

Use `readr::delim` for import

```
library(tidyverse)
labdata <-
read_delim("https://raw.githubusercontent.com/NFilmann/RLadies
FRA/master/labdata.csv", delim=";")
```

Which factors are associated with therapeutic success?

- Therapeutic success in HIV-positive individuals, defined as
  - Primary goal is virologic response:, i.e. reduction of the viral load to an undetectable level below <20 copies per ml) by 24 weeks after start of treatment
  - CD4 cell counts: Key measure of immune status; they should rise 50 to 100 cells per ml in the first year of therapy (a CD4 count < 200 is defined as AIDS).

Source:
https://en.wikipedia.org/wiki/Management_of_HIV/AIDS#Response_to_therapy

basedata

```
> glimpse(basedata)

Observations: 1,299

Variables: 15

$ PatientID        <dbl> 1, 2, 3,
4, 5, 6, 7, 8, 9, 10, 11,...

$ DateOfBirth      <chr>
"17.02.1944", "02.06.1966", ...

$ Start_therapy    <chr>
"09.09.2015", "24.07.2014",
"13.06.2...

$ Gender           <chr> "male",
"male", "male", "male", ...

$ StudyCenter      <dbl> 1, 1, 1,
1, 1, 1, 1, 1, 1, 1...

$ DateOfDiagnosis <chr>
"01.01.1985", "01.01.1989", ...
```

continued

```
$ DateOfDeath      <chr> NA, NA,
NA, NA, NA, NA, NA, ...

$ PreMedication    <chr> "N", "N",
"N", "N", "Y", "N",...

$ HBVpos           <chr> NA, NA,
NA, "Y", NA, NA, "...

$ HCVpos           <chr> NA, NA,
NA, NA, NA, NA, NA, NA,...

$ MedID1           <dbl> 1, 1, 2,
4, 2, 1, 2, 2, 4, 2, 2...

$ MedID2           <dbl> 3, 4, 1,
7, 1, 2, 1, 1, 2, 1, 1...
```

(MedID3, MedID4, MedID5 accordingly)

labdata

```
> glimpse(labdata)
Observations: 7,794
Variables: 7
$ PatientID       <dbl> 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3,
3, 4, 4, 4, 4...
$ Time_weeks      <dbl> 0, 4, 8, 12, 16, 24, 0, 4, 8, 12, 16, 24, 0, 4, 8,
12, 16, 24, 0...
$ Test            <chr> "CD4", "CD4", "CD4", "CD4", "CD4", "CD4", "CD4",
"CD4", "CD4", "...
$ Value           <dbl> 354, 595, 427, 699, 606, 660, 64, 102, 152, 112,
141, 172, 146, ...
$ Test_1          <chr> "HIVPCR", "HIVPCR", "HIVPCR", "HIVPCR", "HIVPCR",
"HIVPCR", "HIV...
$ TErgNumOperator <chr> NA, NA, NA, NA, NA, NA, NA, NA, "<", NA, "<", "<",
NA, NA, NA, "...
$ TErgNum         <dbl> 500, 49, 97, 31, 49, 23, 2840000, 3670, 20, 23, 20,
20, 3136, 38...
```

Now it‘s your turn!