2 - Crowd-sourced online data

Diego Alburez-Gutierrez
MPIDR
European Doctoral School of Demography 2019-20

31/03/2020

MAX-PLANCK-INSTITUT   MAX PLANCK INSTITUTE
FÜR DEMOGRAFISCHE   FOR DEMOGRAPHIC
FORSCHUNG   RESEARCH

Agenda

Crowd-sourced data

Online genealogies

Familinx data

Kaplanis, J., et al. (2018). Quantitative analysis of population-scale family trees with millions of relatives. Science 360(6385):171–175.

Exploring Famlinx

```r
library(data.table)
library(tidyverse)
library(knitr)

# Read sample Familinx data using data.table
fread("../../Assignment/Data/sweden_genealogy.csv") %>%
  # To get a sense of the structure of the data:
  select(profileid, father, mother, birth_year, death_year)
  slice(1:4) %>%
  kable()
```

profileid	father	mother	birth_year	death_year
136	NA	NA	NA	NA
264	57536639	69836161	1875	NA
708	83768131	48140261	NA	NA
722	NA	NA	1694	1767

Assignment

- Questions about instructions
- Issues with Familinx data

Suggested homework activity

- ▶ Try out the `Exercise 1` from the final assignement