Geburtsdataset exploration

Zehui Bai

Stand: 2021-11-23 09:33 Uhr

Table of Contents

# Data Exploration

## Explore the missing data

## Load the sas7bat datasets  
require(haven)   
## scan("clipboard",what="string")  
geburt1 <- read\_sas("C:\\Users\\zbai\\Documents\\GitHub\\R-Projects\\SAS\\geburt1.sas7bdat")  
knitr::kable(sapply(geburt1,function(x)sum(is.na(x))))

|  |  |
| --- | --- |
|  | x |
| PPoint | 0 |
| OW | 0 |
| STALA | 0 |
| sex | 0 |
| altk | 0 |
| alt | 0 |
| AUSL | 0 |
| migrant | 80 |
| e001 | 233 |
| e004 | 220 |
| e005 | 518 |
| e005a | 519 |
| e005z1 | 4119 |
| e005z2 | 4122 |
| e005z3 | 4131 |
| e006 | 1009 |
| e006a | 1009 |
| e006z1 | 5032 |
| e006z2 | 5035 |
| e006z3 | 5038 |
| e007m | 311 |
| e007mc | 15679 |
| e007ma | 311 |
| e007v | 892 |
| e007vc | 15724 |
| e007va | 892 |
| e008m | 338 |
| e008mc | 14790 |
| e008v | 749 |
| e008vc | 14765 |
| e009m | 366 |
| e009mz | 14726 |
| e009v | 972 |
| e009vz | 14818 |
| e010m1 | 276 |
| e010m2 | 276 |
| e010m2c | 14752 |
| e011m | 15081 |
| e011v | 15239 |
| e012mz | 343 |
| e012vz | 1768 |
| e013amz | 552 |
| e013avz | 2386 |
| e013bmz | 639 |
| e013bvz | 2584 |
| Mbmi | 673 |
| Vbmi | 2630 |
| E014z | 1902 |
| e0151 | 2098 |
| e0152 | 2552 |
| e0152c1 | 15649 |
| e0152c2 | 17518 |
| e0152c3 | 17633 |
| e0153 | 248 |
| e0154 | 251 |
| e0155 | 260 |
| e0156 | 251 |
| e0157 | 277 |
| e016z | 1546 |
| e016 | 8573 |
| e017az | 453 |
| e017bz | 477 |
| e018 | 472 |
| e018m1 | 13654 |
| e018m2 | 13654 |
| e018m3 | 13654 |
| e018m4 | 13654 |
| e018m5 | 13654 |
| e018m6 | 13654 |
| e018m6z | 16076 |
| e019 | 227 |
| e029z | 551 |
| e057z | 675 |
| e089m | 345 |
| e089v | 1049 |
| e090m | 822 |
| e090v | 1431 |
| e091m | 422 |
| e091v | 1169 |
| e092m | 1003 |
| e092v | 1574 |
| e093 | 1088 |
| groe | 86 |
| gewi | 99 |
| kopf | 209 |
| arml | 237 |
| tail | 67 |
| huef | 79 |
| whr | 81 |
| WHtR | 72 |
| ebog | 184 |
| tricps | 356 |
| rueck | 395 |

Rohdatensatz <- read\_sas("C:\\Users\\zbai\\Documents\\GitHub\\R-Projects\\SAS\\Rohdatensatz.sas7bdat")  
library(tidyverse)  
Rohdatensatz <- na\_if(Rohdatensatz, '')   
Rohdatensatz <- na\_if(Rohdatensatz, 'N/A')   
  
knitr::kable(sapply(Rohdatensatz,function(x)sum(is.na(x))))

|  |  |
| --- | --- |
|  | x |
| \_Antwort\_ID | 0 |
| Zufallsgeneratorstartwert | 0 |
| Welches\_Geschlecht\_haben\_Sie\_ | 11 |
| Bitte\_geben\_Sie\_Ihr\_Alter\_an | 26 |
| Ist\_Ihr\_aktueller\_Erstwohnsitz\_i | 26 |
| Wie\_oft\_besuchen\_Sie\_Ihre\_Frauen | 45 |
| Warum\_besuchen\_Sie\_selten\_eine\_F | 286 |
| Seit\_Beginn\_der\_Corona\_Pandemie | 50 |
| VAR12 | 93 |
| VAR13 | 93 |
| VAR14 | 93 |
| VAR15 | 93 |
| VAR16 | 93 |
| Haben\_Sie\_eine\_HPV\_Impfung\_erhal | 48 |
| Bestehen\_oder\_bestanden\_bei\_Ihne | 45 |
| VAR19 | 45 |
| VAR20 | 45 |
| VAR21 | 46 |
| VAR22 | 161 |
| VAR23 | 320 |
| VAR24 | 81 |
| VAR25 | 81 |
| VAR26 | 81 |
| VAR27 | 81 |
| VAR28 | 81 |
| Woraus\_besteht\_das\_Angebot\_der\_F | 81 |
| VAR30 | 81 |
| VAR31 | 81 |
| VAR32 | 81 |
| VAR33 | 81 |
| VAR34 | 81 |
| VAR35 | 81 |
| VAR36 | 81 |
| VAR37 | 81 |
| Was\_wird\_allgemein\_als\_Vorteil\_d | 81 |
| VAR39 | 81 |
| VAR40 | 81 |
| VAR41 | 81 |
| VAR42 | 81 |
| VAR43 | 81 |
| VAR44 | 81 |
| VAR45 | 81 |
| VAR46 | 101 |
| VAR47 | 101 |
| VAR48 | 101 |
| VAR49 | 101 |
| VAR50 | 101 |
| VAR51 | 101 |
| VAR52 | 101 |
| VAR53 | 101 |
| VAR54 | 101 |
| VAR55 | 101 |
| VAR56 | 101 |
| VAR57 | 101 |
| Was\_sind\_typische\_erste\_Symptome | 101 |
| VAR59 | 101 |
| VAR60 | 101 |
| VAR61 | 101 |
| VAR62 | 101 |
| VAR63 | 101 |
| VAR64 | 101 |
| VAR65 | 101 |
| VAR66 | 101 |
| VAR67 | 101 |
| VAR68 | 313 |
| VAR69 | 101 |
| Wie\_ist\_Ihr\_Krankenversicherungs | 109 |
| VAR71 | 109 |
| Bitte\_geben\_Sie\_Ihr\_eigenes\_mona | 119 |
| VAR73 | 328 |
| Bitte\_ordnen\_Sie\_Ihr\_Selbstwertg | 114 |
| Rauchen\_Sie\_Zigaretten\_*Wenn\_ja* | 111 |
| VAR76 | 325 |
| In\_welchem\_Land\_ist\_Ihr\_Vater\_ge | 110 |
| VAR78 | 277 |
| In\_welchem\_Land\_ist\_Ihre\_Mutter | 110 |
| VAR80 | 279 |
| Seit\_wann\_leben\_Sie\_in\_Deutschla | 110 |
| VAR82 | 329 |
| Ist\_Deutsch\_Ihre\_Muttersprache\_ | 110 |
| VAR84 | 275 |
| VAR85 | 109 |
| VAR86 | 116 |
| Bitte\_geben\_Sie\_Ihren\_Familienst | 112 |
| VAR88 | 331 |
| Wie\_viele\_Geburten\_hatten\_Sie\_ | 110 |
| Mit\_wem\_wohnen\_Sie\_zusammen\_in\_e | 110 |
| VAR91 | 333 |
| In\_welchem\_der\_folgenden\_Stadt\_\_ | 111 |
| In\_welchem\_Bundesland\_leben\_Sie\_ | 112 |

str(Rohdatensatz)

## tibble[,90] [333 x 90] (S3: tbl\_df/tbl/data.frame)  
## $ \_Antwort\_ID : num [1:333] 2 3 4 5 6 7 8 9 10 11 ...  
## ..- attr(\*, "format.sas")= chr "BEST"  
## $ Zufallsgeneratorstartwert : num [1:333] 9.48e+08 5.91e+08 5.95e+08 7.96e+08 2.11e+09 ...  
## ..- attr(\*, "format.sas")= chr "BEST"  
## $ Welches\_Geschlecht\_haben\_Sie\_ : chr [1:333] "Weiblich" NA "Weiblich" "Weiblich" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Bitte\_geben\_Sie\_Ihr\_Alter\_an : chr [1:333] "20 - 34" NA NA "20 - 34" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Ist\_Ihr\_aktueller\_Erstwohnsitz\_i: chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Wie\_oft\_besuchen\_Sie\_Ihre\_Frauen: chr [1:333] "Jährlich" NA NA "Halbjährlich" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Warum\_besuchen\_Sie\_selten\_eine\_F: chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Seit\_Beginn\_der\_Corona\_Pandemie : chr [1:333] "...genauso regelmäßig wie vor der Pandemie besucht" NA NA "...genauso regelmäßig wie vor der Pandemie besucht" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR12 : chr [1:333] "Nein" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR13 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR14 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR15 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR16 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Haben\_Sie\_eine\_HPV\_Impfung\_erhal: chr [1:333] "Nein ich habe keine HPV-Impfung erhalten" NA NA "Ja, im Alter von 15 bis 17" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Bestehen\_oder\_bestanden\_bei\_Ihne: chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR19 : chr [1:333] "Nein" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR20 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR21 : chr [1:333] "Ja" NA NA "Weiß ich nicht" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR22 : chr [1:333] "Jährlich" NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR23 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR24 : chr [1:333] "Nein" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR25 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR26 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR27 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR28 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Woraus\_besteht\_das\_Angebot\_der\_F: chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR30 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR31 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR32 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR33 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR34 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR35 : chr [1:333] "Nein" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR36 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR37 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Was\_wird\_allgemein\_als\_Vorteil\_d: chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR39 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR40 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR41 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR42 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR43 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR44 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR45 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR46 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR47 : chr [1:333] "Ja" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR48 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR49 : chr [1:333] "Nein" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR50 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR51 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR52 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR53 : chr [1:333] "Ja" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR54 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR55 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR56 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR57 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Was\_sind\_typische\_erste\_Symptome: chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR59 : chr [1:333] "Ja" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR60 : chr [1:333] "Ja" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR61 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR62 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR63 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR64 : chr [1:333] "Nein" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR65 : chr [1:333] "Nein" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR66 : chr [1:333] "Nein" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR67 : chr [1:333] "Ja" NA NA "Nein" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR68 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR69 : chr [1:333] "Ja" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Wie\_ist\_Ihr\_Krankenversicherungs: chr [1:333] "Gesetzlich versichert" NA NA "Privat versichert" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR71 : chr [1:333] "(Fach-)Hochschulabschluss (Bachelor, Master, Magister, Diplom, Staatsexamen, Promotion)" NA NA "Abitur (Allgemeine Hochschulreife), Fachgebundene Hochschulreife oder Fachhochschulreife" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Bitte\_geben\_Sie\_Ihr\_eigenes\_mona: chr [1:333] "500 bis unter 1.000\200" NA NA "1.000 bis unter 1.500\200" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR73 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Bitte\_ordnen\_Sie\_Ihr\_Selbstwertg: num [1:333] 4 NA NA 4 3 NA NA 2 4 4 ...  
## ..- attr(\*, "format.sas")= chr "BEST"  
## $ Rauchen\_Sie\_Zigaretten\_\_Wenn\_ja\_: chr [1:333] "Ich habe noch nie geraucht" NA NA "Ich rauche derzeit gelegentlich (weniger als 1 Zigarette pro Woche)" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR76 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ In\_welchem\_Land\_ist\_Ihr\_Vater\_ge: chr [1:333] "Sonstiges" NA NA "Deutschland" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR78 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ In\_welchem\_Land\_ist\_Ihre\_Mutter : chr [1:333] "Sonstiges" NA NA "Deutschland" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR80 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Seit\_wann\_leben\_Sie\_in\_Deutschla: chr [1:333] "Weniger als 10 Jahre" NA NA "Seit Geburt" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR82 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Ist\_Deutsch\_Ihre\_Muttersprache\_ : chr [1:333] "Nein" NA NA "Ja" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR84 : chr [1:333] "Gut" NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR85 : chr [1:333] "Eine befristete Aufenthaltsgenehmigung" NA NA "Die Deutsche Staatsbürgerschaft" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR86 : chr [1:333] "Christlich" NA NA "Christlich" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Bitte\_geben\_Sie\_Ihren\_Familienst: chr [1:333] "Alleinstehend" NA NA "Alleinstehend" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR88 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Wie\_viele\_Geburten\_hatten\_Sie\_ : chr [1:333] "(Noch) keine Geburt" NA NA "(Noch) keine Geburt" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ Mit\_wem\_wohnen\_Sie\_zusammen\_in\_e: chr [1:333] "Ich wohne alleine" NA NA "Mit der Familie (Partner\*in/ Kindern/ Eltern/ Geschwister)" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ VAR91 : chr [1:333] NA NA NA NA ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ In\_welchem\_der\_folgenden\_Stadt\_\_: chr [1:333] "Großstadt (ab 100.000 Einwohner)" NA NA "Mittelstadt (20.000 bis 100.000 Einwohner)" ...  
## ..- attr(\*, "format.sas")= chr "$"  
## $ In\_welchem\_Bundesland\_leben\_Sie\_: chr [1:333] "Bremen" NA NA "Nordrhein-Westfalen" ...  
## ..- attr(\*, "format.sas")= chr "$"