Latent Gender Score

Preprocess Data

Charité - Universitätsmedizin Berlin, Institut für Public Health  
Hans-Aloys Wischmann, Mascha Kern

Revised: March 26, 2025  
Executed: March 27, 2025

## SEM: Domain Knowledge Model

## Configuration: Variables and Mappings

Configuration: Data variables

| Variable | Name | Values | Mappings |
| --- | --- | --- | --- |
| hid | hid |  |  |
| sex\_binary | sex | 1 (male) 2 (female) -1 (no answer) -2 (does not apply) -3 (answer improbable) | 1 > 0 2 > 1 <0 > NA |
| birth\_year | gebjahr | 1909 … 1972 … 1984 … | <0 > NA |
| immi\_year | immiyear | 1950 … | <0 > NA |
| immigration\_history | migback | 1 (no mig back) 2 (direct mig back) 3 (indirect mig back) -1 (no answer) | 1 > 0 3 > 1 <0 > NA |
| refugee\_exp | arefback | 1 (no evidence) 2 (with evidence of direct ref exp) 3 (with evidence of indirect ref exp) -1 (no answer) | 1 > 0 3 > 1 <0 > NA |
| east\_german\_residence | sampreg | 1 (West-G) 2 (East-G) -1 (no answer) -2 (does not apply) | 1 > 0 2 > 1 <0 > NA |
| sex\_or | sexor | 0 (probably hetero) 1 (probably bi or homo) 2 (insufficient info) | 2 > NA |
| partner | partner | 0 (no partner) 1 (spouse or registered partner) 2 (partner) 3 (probably spouse or registered pertner) 4 (probably partner) 5 (not clear) -1 (no answer) -2 (does not apply) | 2 > 1 3 > 1 4 > 1 5 > NA <0 > NA |
| sex\_entry\_change | pla0048 | 1 (yes) 2 (no) -1 (no answer) -5 (not included) -8 (question this year not part of survey) | 1 > 0 2 > 1 <0 > NA |
| felt\_discriminated\_gender\_12mo | plh0415i03 | 1 (yes) 2 (no) | 2 > 0 <0 > NA |
| felt\_discriminated\_ethnic\_12mo | plh0415i01 | 1 (yes) 2 (no) | 2 > 0 <0 > NA |
| using\_period\_of\_care | plb0020\_h | 1 (yes) 2 (no) | 2 > 0 <0 > NA |
| current\_mat\_parent\_leave | plb0019\_v2 | 1 (yes maternity) 2 (yes parental) 3 (no) | 3 > 0 2 > 1 <0 > NA |
| risk\_taking\_scale | plh0204\_v2 | 0-10 | <0 > NA |
| risk\_taking\_driving\_scale | plh0197 | 0-10 | <0 > NA |
| risk\_taking\_health\_scale | plh0201 | 0-10 | <0 > NA |
| highest\_educational\_degree | pgisced11 | 0 (in school) 1 (primary education) 2 (lower secondary eduation) 3 (upper secondary education) 4 (post-secondary non-tertiary education) 5 (short-cycle tertiary) 6 (Bachelors or equivalent) 7 (Masters or equivalent) 8 (Doctoral or equivalent) | <0 > NA |
| current\_monthly\_gross\_labor\_income | pglabgro | Euros gross | <0 > NA |
| daily\_hours\_childcare\_weekdays | pli0044\_h | 0-80 | <0 > NA |
| daily\_hours\_housework\_weekdays | pli0043\_h | 0-84 | <0 > NA |
| employment\_status\_imp | pgemplst | 1 (full-time) 2 (regular part-time) 3 (vocational training) 4 (irregular part-time) 5 (not employed) 6 (sheltered workshop) 7 (short-time work) | 5 > 0 1 > 13 3 > 1 6 > 1 4 > 2 7 > 2 13 > 3 <0 > NA |
| work\_time | pgtatzeit | 0-80 | <0 > NA |
| leadership\_position | plb0067 | 1 (yes) 2 (no) -1 (no answer) -5 (not included) -8 (question this year not part of survey) | 2 > 0 <0 > NA |
| frequency\_homeoffice | plb0096\_v1 | 1 (daily) 2 (several times per week) 3 (once every 2 to 4 weeks) 4 (rarely) -1 (no answer) -2 (does not apply) -5 (not included) -8 (not part of syear) | -2 > 0 1 > 2 3 > 1 4 > 1 <0 > NA |
| working\_overtime | plb0193\_v1 | 1 (yes) 2 (no) 3 (does not apply self-employed) -1 (no answer) -2 (does not apply) | -2 > 0 2 > 0 3 > 2 <0 > NA |
| worried\_health | plh0035 | 1 (big worries) 2 (some worries) 3 (no worries) -1 (no answer) -2 ( does not apply) | <0 > NA |
| worried\_pension | plh0335 | 1 (big worries) 2 (some worries) 3 (no worries) -1 (no answer) -2 ( does not apply) | <0 > NA |
| career\_sacrifices | plb0115 | 1 (Strongly disagree) 2 (Rather disagree) 3 (Rather agree) 4 (Fully agree) -1 (No answer) -2 (Does not apply) | <0 > NA |
| current\_life\_satisfaction\_scale | plh0182 | 0-10 | <0 > NA |
| political\_interest | plh0007 | 1 (Very strongly) 2 (strongly) 3 (not so strongly) 4 (not at all) -1 (no answer) -2 (does not apply) | 4 > 0 1 > 4 3 > 1 4 > 3 <0 > NA |
| current\_health | ple0008 | 1 (very good) 2 (good) 3 (satisfactory) 4 (less good) 5 (bad) -1 (no answer) | <0 > NA |
| self\_esteem | plh0206i11 | 1-7 | <0 > NA |
| health\_insurance\_status | ple0099\_v5 | 1 (Pflichtmitglied) 2 (freiwillig) 3 (mitversichters Familienmitglied) 4 (versichert als Rentner) -1 (no answer) -2 (does not apply) -3 (implausible) -4 (multiple response) -5 (not included) -8 (not in syear) | <0 > NA |
| alcohol\_consumption | ple0177 | 1 (daily) 2 (4 to 6 days) 3 (2 to 3 days) 4 (2 to 4 days) 5 (once a month or less) 6 (never) | 6 > 0 5 > 0 1 > 2 4 > 1 3 > 1 <0 > NA |
| current\_smoke | ple0081\_v2 | 1 (yes) 2 (no) -1 (no answer) | 2 > 0 <0 > NA |
| ever\_smoke | ple0080\_v3 | 1 (yes) 2 (no) -1 (no answer) | 2 > 0 <0 > NA |
| num\_physician\_visits | ple0072 | 0-99 | <0 > NA |
| diabetes | ple0012 | 1 (Yes) -1 (no answer) -2 (does not apply) | -2 > 0 -1 > 0 <0 > NA |
| hypertension | ple0018 | 1 (Yes) -1 (no answer) -2 (does not apply) | -2 > 0 -1 > 0 <0 > NA |
| migraine | ple0017 | 1 (Yes) -1 (no answer) -2 (does not apply) | -2 > 0 -1 > 0 <0 > NA |
| stroke | ple0016 | 1 (Yes) -1 (no answer) -2 (does not apply) | -2 > 0 -1 > 0 <0 > NA |
| no\_illness | ple0024 | 1 (Yes) -1 (no answer) -2 (does not apply) | -2 > 0 -1 > 0 <0 > NA |
| weight\_factor | phrf |  |  |

Configuration: Match variables

| Variable | Name | File |
| --- | --- | --- |
| syear | syear | ../Data/ppathl.csv |
| pid | pid | ../Data/ppathl.csv |

## Load: Data

## Apply: Mappings

Study Years with Required Info

| syear | current\_smoke | ever\_smoke | stroke | migraine | hypertension | diabetes | total |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2002 | 0 | 23,834 | 0 | 0 | 0 | 0 | 23,834 |
| 2004 | 21,967 | 0 | 0 | 0 | 0 | 0 | 21,967 |
| 2006 | 22,332 | 0 | 0 | 0 | 0 | 0 | 22,332 |
| 2008 | 19,658 | 0 | 0 | 0 | 0 | 0 | 19,658 |
| 2009 | 0 | 0 | 20,792 | 20,792 | 20,792 | 20,792 | 83,168 |
| 2010 | 18,896 | 0 | 0 | 0 | 0 | 0 | 18,896 |
| 2011 | 0 | 0 | 21,069 | 21,069 | 21,069 | 21,069 | 84,276 |
| 2012 | 0 | 18,191 | 0 | 0 | 0 | 0 | 18,191 |
| 2013 | 0 | 0 | 19,149 | 19,149 | 19,149 | 19,149 | 76,596 |
| 2014 | 27,414 | 0 | 0 | 0 | 0 | 0 | 27,414 |
| 2015 | 0 | 0 | 25,394 | 25,394 | 25,394 | 25,394 | 101,576 |
| 2016 | 24,485 | 0 | 0 | 0 | 0 | 0 | 24,485 |
| 2017 | 0 | 0 | 26,914 | 26,914 | 26,914 | 26,914 | 107,656 |
| 2018 | 25,845 | 0 | 0 | 0 | 0 | 0 | 25,845 |
| 2019 | 0 | 0 | 26,048 | 26,048 | 26,048 | 26,048 | 104,192 |
| 2020 | 29,152 | 0 | 0 | 0 | 0 | 0 | 29,152 |
| 2021 | 2,164 | 0 | 20,142 | 20,142 | 20,142 | 20,142 | 82,732 |
| 2022 | 26,650 | 0 | 0 | 0 | 0 | 0 | 26,650 |

[1] "Removing study years before: 2002"

## Consistency: Dx vs. no\_illness

Records with inconsistent Dx, Individual Dx prevails

| pid | syear | no\_illness | hypertension | diabetes | migraine | stroke |
| --- | --- | --- | --- | --- | --- | --- |
| 301701 | 2021 | 1 | 1 | 0 | 0 | 0 |
| 366902 | 2021 | 1 | 1 | 0 | 0 | 0 |
| 2001701 | 2021 | 1 | 1 | 0 | 0 | 0 |
| 2893902 | 2009 | 1 | 1 | 0 | 0 | 0 |
| 5213002 | 2021 | 1 | 0 | 1 | 0 | 0 |
| 8004202 | 2021 | 1 | 1 | 0 | 0 | 0 |
| 21608001 | 2021 | 1 | 1 | 1 | 1 | 1 |
| 27342503 | 2021 | 1 | 0 | 0 | 0 | 1 |

## Incidence: Migraine and Stroke

## Risk factors: Sequence of Dx and Smoking

Sequence of Dx

| name | value |
| --- | --- |
| pid | 162,323 |
| smoke\_before\_stroke | 32,428 |
| smoke\_before\_migraine | 32,006 |
| diabetes\_before\_stroke | 4,497 |
| diabetes\_before\_migraine | 4,409 |
| hypertension\_before\_stroke | 14,883 |
| hypertension\_before\_migraine | 14,264 |

Stroke: Unadjusted (Incident = 786 of 54714)

| Variable | Contrast | RR | Signif |
| --- | --- | --- | --- |
| hypertension\_before\_stroke | (1) / (0) | 2.726 [2.320, 3.132] | \* |
| diabetes\_before\_stroke | (1) / (0) | 1.989 [1.640, 2.337] | \* |
| smoke\_before\_stroke | (1) / (0) | 1.883 [1.615, 2.150] | \* |
| sex\_binary | (1) / (0) | 0.798 [0.685, 0.910] | \* |

Migraine: Unadjusted (Incident = 2146 of 52087)

| Variable | Contrast | RR | Signif |
| --- | --- | --- | --- |
| sex\_binary | (1) / (0) | 3.173 [2.865, 3.480] | \* |
| smoke\_before\_migraine | (1) / (0) | 1.466 [1.345, 1.588] | \* |
| hypertension\_before\_migraine | (1) / (0) | 0.834 [0.748, 0.919] | \* |
| diabetes\_before\_migraine | (1) / (0) | 0.794 [0.651, 0.936] | \* |

## Compute: Age, Age at Immigration, Children in Household

## Filter: Adults with Reported Sex

## Valid Data: Most Recent with Outcome (Yes/No for Migraine Incidence and Stroke Incidence)

Adults: Prevalence and Incidence of Dx by Sex

| sex binary | N prev | migraine prevalence firstDx | stroke prevalence firstDx | N inci mig | migraine incidence | N inci str | stroke incidence |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 61,192 | 792 | 445 | 26,167 | 532 | 26,514 | 450 |
| 1 | 60,472 | 2,618 | 339 | 25,914 | 1,614 | 28,193 | 336 |

[1] "Participants (Valid data elements)"

[1] "All: 162323 <- 19142644"

[1] "Adults: 121664 <- 16697401"

[1] "Recent: 121664 <- 3317870"

[1] "Analysis: 51370 <- 2286130"

[1] "Weight!=0: 48819"

[1] "Incidence over study years (excluding prevalent Migraine or prevalent Stroke), N=51370: Migraine (2116), Stroke (742)"

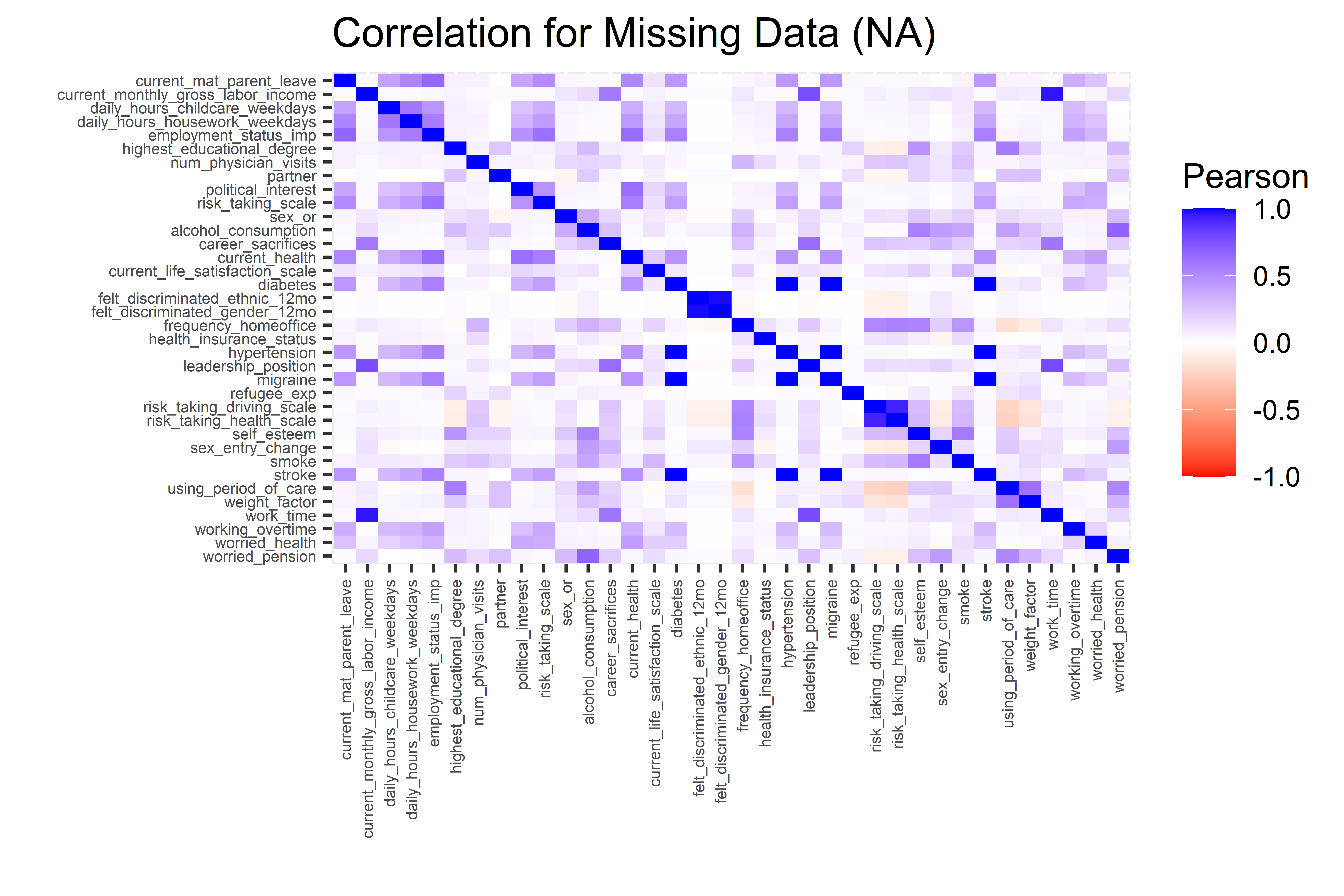
Valid Data Points

| SEM | Variable | N | Fraction [%] |
| --- | --- | --- | --- |
| Scaled \* 0.1 | age | 51,370 | 100.0 |
| As is | diabetes\_before\_migraine | 51,370 | 100.0 |
| As is | diabetes\_before\_stroke | 51,370 | 100.0 |
| As is | east\_german\_residence | 51,370 | 100.0 |
| As is | hypertension\_before\_migraine | 51,370 | 100.0 |
| As is | hypertension\_before\_stroke | 51,370 | 100.0 |
| As is | immigration\_history | 51,370 | 100.0 |
| As is | migraine\_incidence | 51,370 | 100.0 |
| As is | num\_children\_in\_household | 51,370 | 100.0 |
| As is | sex\_binary | 51,370 | 100.0 |
| As is | smoke\_before\_migraine | 51,370 | 100.0 |
| As is | smoke\_before\_stroke | 51,370 | 100.0 |
| As is | stroke\_incidence | 51,370 | 100.0 |
| As is | employment\_status\_imp | 51,354 | 100.0 |
| As is | current\_mat\_parent\_leave | 51,346 | 100.0 |
| As is | risk\_taking\_scale | 51,342 | 99.9 |
| As is | daily\_hours\_housework\_weekdays | 51,337 | 99.9 |
| As is | political\_interest | 51,323 | 99.9 |
| As is | daily\_hours\_childcare\_weekdays | 51,319 | 99.9 |
| As is | partner | 49,864 | 97.1 |
| As is | highest\_educational\_degree | 48,548 | 94.5 |
| As is | num\_physician\_visits | 46,727 | 91.0 |
| As is | sex\_or | 43,505 | 84.7 |
| As is | current\_monthly\_gross\_labor\_income | 39,495 | 76.9 |
| As is | gross\_hourly\_wage | 38,791 | 75.5 |
|  | age\_group | 51,370 | 100.0 |
|  | age\_group\_at\_immigration | 51,370 | 100.0 |
|  | current\_health | 51,344 | 99.9 |
|  | diabetes | 51,335 | 99.9 |
|  | hypertension | 51,335 | 99.9 |
|  | migraine | 51,335 | 99.9 |
|  | stroke | 51,335 | 99.9 |
|  | working\_overtime | 51,305 | 99.9 |
|  | worried\_health | 51,266 | 99.8 |
|  | refugee\_exp | 51,096 | 99.5 |
|  | current\_life\_satisfaction\_scale | 50,896 | 99.1 |
|  | weight\_factor | 48,819 | 95.0 |
|  | smoke | 46,840 | 91.2 |
|  | using\_period\_of\_care | 46,272 | 90.1 |
|  | health\_insurance\_status | 44,723 | 87.1 |
|  | self\_esteem | 41,785 | 81.3 |
|  | frequency\_homeoffice | 40,258 | 78.4 |
|  | work\_time | 38,791 | 75.5 |
|  | worried\_pension | 36,721 | 71.5 |
|  | leadership\_position | 33,959 | 66.1 |
|  | risk\_taking\_health\_scale | 32,598 | 63.5 |
|  | alcohol\_consumption | 31,306 | 60.9 |
|  | risk\_taking\_driving\_scale | 30,888 | 60.1 |
|  | career\_sacrifices | 27,036 | 52.6 |
|  | sex\_entry\_change | 16,438 | 32.0 |
|  | felt\_discriminated\_ethnic\_12mo | 871 | 1.7 |
|  | felt\_discriminated\_gender\_12mo | 868 | 1.7 |

Table 1: Population (Current Status)

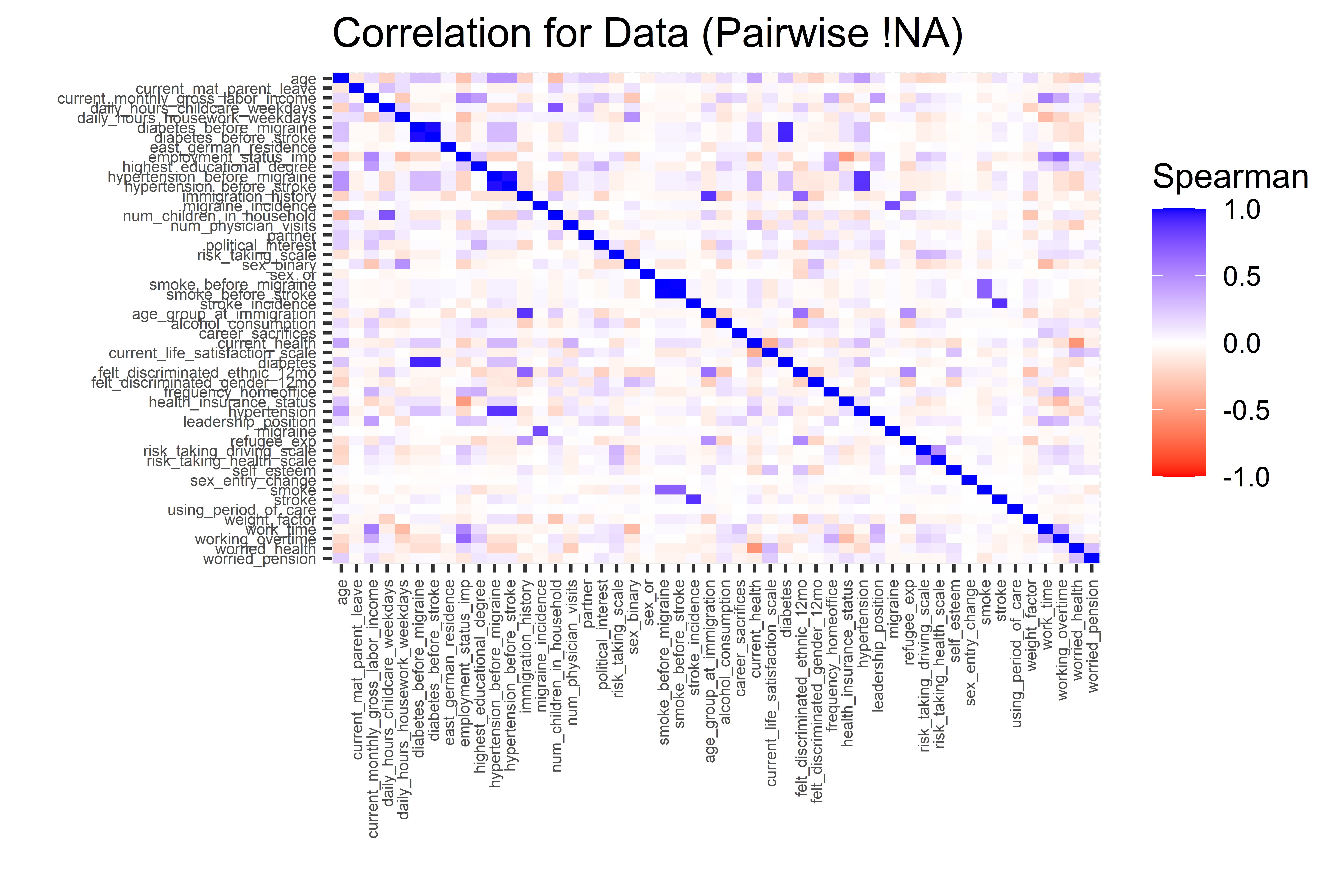
|  | Sex\_M (N=25741) | Sex\_F (N=25629) | Overall (N=51370) |
| --- | --- | --- | --- |
| **age** |  |  |  |
| Mean (SD) | 50.1 (18.4) | 50.3 (18.5) | 50.2 (18.5) |
| Median [Min, Max] | 50.0 [18.0, 106] | 50.0 [18.0, 104] | 50.0 [18.0, 106] |
| **age\_group** |  |  |  |
| [18,35) | 6359 (24.7%) | 5997 (23.4%) | 12356 (24.1%) |
| [35,50) | 6242 (24.2%) | 6796 (26.5%) | 13038 (25.4%) |
| [50,65) | 7066 (27.5%) | 6697 (26.1%) | 13763 (26.8%) |
| [65,Inf) | 6074 (23.6%) | 6139 (24.0%) | 12213 (23.8%) |
| **sex\_or** |  |  |  |
| Prob. Hetero | 21196 (82.3%) | 21186 (82.7%) | 42382 (82.5%) |
| Prob. Bi or Homo | 533 (2.1%) | 590 (2.3%) | 1123 (2.2%) |
| Missing | 4012 (15.6%) | 3853 (15.0%) | 7865 (15.3%) |
| **partner** |  |  |  |
| No | 10396 (40.4%) | 11139 (43.5%) | 21535 (41.9%) |
| Yes | 14591 (56.7%) | 13738 (53.6%) | 28329 (55.1%) |
| Missing | 754 (2.9%) | 752 (2.9%) | 1506 (2.9%) |
| **smoke** |  |  |  |
| No | 16063 (62.4%) | 18112 (70.7%) | 34175 (66.5%) |
| Yes | 7266 (28.2%) | 5399 (21.1%) | 12665 (24.7%) |
| Missing | 2412 (9.4%) | 2118 (8.3%) | 4530 (8.8%) |
| **hypertension** |  |  |  |
| No | 19631 (76.3%) | 20054 (78.2%) | 39685 (77.3%) |
| Yes | 6093 (23.7%) | 5557 (21.7%) | 11650 (22.7%) |
| Missing | 17 (0.1%) | 18 (0.1%) | 35 (0.1%) |
| **diabetes** |  |  |  |
| No | 23635 (91.8%) | 23938 (93.4%) | 47573 (92.6%) |
| Yes | 2089 (8.1%) | 1673 (6.5%) | 3762 (7.3%) |
| Missing | 17 (0.1%) | 18 (0.1%) | 35 (0.1%) |
| **migraine** |  |  |  |
| No | 25405 (98.7%) | 24615 (96.0%) | 50020 (97.4%) |
| Yes | 319 (1.2%) | 996 (3.9%) | 1315 (2.6%) |
| Missing | 17 (0.1%) | 18 (0.1%) | 35 (0.1%) |
| **stroke** |  |  |  |
| No | 25382 (98.6%) | 25373 (99.0%) | 50755 (98.8%) |
| Yes | 342 (1.3%) | 238 (0.9%) | 580 (1.1%) |
| Missing | 17 (0.1%) | 18 (0.1%) | 35 (0.1%) |
| **num\_children\_in\_household** |  |  |  |
| None | 18128 (70.4%) | 17304 (67.5%) | 35432 (69.0%) |
| Single | 3397 (13.2%) | 3877 (15.1%) | 7274 (14.2%) |
| Multiple | 4216 (16.4%) | 4448 (17.4%) | 8664 (16.9%) |
| **daily\_hours\_childcare\_weekdays** |  |  |  |
| Mean (SD) | 0.644 (1.75) | 1.85 (4.14) | 1.25 (3.23) |
| Median [Min, Max] | 0 [0, 24.0] | 0 [0, 24.0] | 0 [0, 24.0] |
| Missing | 28 (0.1%) | 23 (0.1%) | 51 (0.1%) |
| **daily\_hours\_housework\_weekdays** |  |  |  |
| Mean (SD) | 0.979 (0.970) | 2.16 (1.54) | 1.57 (1.42) |
| Median [Min, Max] | 1.00 [0, 20.0] | 2.00 [0, 24.0] | 1.00 [0, 24.0] |
| Missing | 22 (0.1%) | 11 (0.0%) | 33 (0.1%) |
| **employment\_status\_imp** |  |  |  |
| Not employed | 8704 (33.8%) | 10933 (42.7%) | 19637 (38.2%) |
| Training/Shelter | 835 (3.2%) | 659 (2.6%) | 1494 (2.9%) |
| Part-Time/Short-Time | 2206 (8.6%) | 7525 (29.4%) | 9731 (18.9%) |
| Full-Time | 13987 (54.3%) | 6505 (25.4%) | 20492 (39.9%) |
| Missing | 9 (0.0%) | 7 (0.0%) | 16 (0.0%) |
| **work\_time** |  |  |  |
| Mean (SD) | 39.8 (13.3) | 30.5 (13.7) | 35.4 (14.3) |
| Median [Min, Max] | 40.0 [0.400, 80.0] | 33.0 [0.500, 80.0] | 40.0 [0.400, 80.0] |
| Missing | 5354 (20.8%) | 7225 (28.2%) | 12579 (24.5%) |
| **current\_monthly\_gross\_labor\_income** |  |  |  |
| Mean (SD) | 3700 (19400) | 2090 (1940) | 2940 (14200) |
| Median [Min, Max] | 2800 [0, 2700000] | 1700 [0, 75600] | 2300 [0, 2700000] |
| Missing | 4968 (19.3%) | 6907 (27.0%) | 11875 (23.1%) |
| **east\_german\_residence** |  |  |  |
| No | 20570 (79.9%) | 20547 (80.2%) | 41117 (80.0%) |
| Yes | 5171 (20.1%) | 5082 (19.8%) | 10253 (20.0%) |
| **immigration\_history** |  |  |  |
| None | 18572 (72.1%) | 18650 (72.8%) | 37222 (72.5%) |
| Indirect | 1632 (6.3%) | 1599 (6.2%) | 3231 (6.3%) |
| Direct | 5537 (21.5%) | 5380 (21.0%) | 10917 (21.3%) |
| **age\_group\_at\_immigration** |  |  |  |
| [18,35) | 20540 (79.8%) | 20607 (80.4%) | 41147 (80.1%) |
| [35,50) | 381 (1.5%) | 370 (1.4%) | 751 (1.5%) |
| [50,65) | 599 (2.3%) | 538 (2.1%) | 1137 (2.2%) |
| [65,Inf) | 4221 (16.4%) | 4114 (16.1%) | 8335 (16.2%) |
| **refugee\_exp** |  |  |  |
| None | 23657 (91.9%) | 24213 (94.5%) | 47870 (93.2%) |
| Indirect | 76 (0.3%) | 67 (0.3%) | 143 (0.3%) |
| Direct | 1866 (7.2%) | 1217 (4.7%) | 3083 (6.0%) |
| Missing | 142 (0.6%) | 132 (0.5%) | 274 (0.5%) |
| **weight\_factor** |  |  |  |
| Mean (SD) | 3100 (3990) | 3030 (3790) | 3070 (3890) |
| Median [Min, Max] | 1740 [1.24, 54200] | 1710 [1.70, 50500] | 1720 [1.24, 54200] |
| Missing | 1250 (4.9%) | 1301 (5.1%) | 2551 (5.0%) |
| **sex\_entry\_change** |  |  |  |
| No | 7923 (30.8%) | 8286 (32.3%) | 16209 (31.6%) |
| Yes | 95 (0.4%) | 134 (0.5%) | 229 (0.4%) |
| Missing | 17723 (68.9%) | 17209 (67.1%) | 34932 (68.0%) |
| **felt\_discriminated\_gender\_12mo** |  |  |  |
| No | 354 (1.4%) | 372 (1.5%) | 726 (1.4%) |
| Yes | 13 (0.1%) | 129 (0.5%) | 142 (0.3%) |
| Missing | 25374 (98.6%) | 25128 (98.0%) | 50502 (98.3%) |
| **felt\_discriminated\_ethnic\_12mo** |  |  |  |
| No | 169 (0.7%) | 316 (1.2%) | 485 (0.9%) |
| Yes | 203 (0.8%) | 183 (0.7%) | 386 (0.8%) |
| Missing | 25369 (98.6%) | 25130 (98.1%) | 50499 (98.3%) |
| **using\_period\_of\_care** |  |  |  |
| No | 23110 (89.8%) | 22847 (89.1%) | 45957 (89.5%) |
| Yes | 94 (0.4%) | 221 (0.9%) | 315 (0.6%) |
| Missing | 2537 (9.9%) | 2561 (10.0%) | 5098 (9.9%) |
| **current\_mat\_parent\_leave** |  |  |  |
| No | 25627 (99.6%) | 24633 (96.1%) | 50260 (97.8%) |
| Yes | 97 (0.4%) | 989 (3.9%) | 1086 (2.1%) |
| Missing | 17 (0.1%) | 7 (0.0%) | 24 (0.0%) |
| **risk\_taking\_scale** |  |  |  |
| Mean (SD) | 5.23 (2.45) | 4.40 (2.46) | 4.82 (2.49) |
| Median [Min, Max] | 5.00 [0, 10.0] | 5.00 [0, 10.0] | 5.00 [0, 10.0] |
| Missing | 12 (0.0%) | 16 (0.1%) | 28 (0.1%) |
| **risk\_taking\_driving\_scale** |  |  |  |
| Mean (SD) | 3.83 (2.63) | 2.83 (2.56) | 3.32 (2.64) |
| Median [Min, Max] | 4.00 [0, 10.0] | 2.00 [0, 10.0] | 3.00 [0, 10.0] |
| Missing | 10546 (41.0%) | 9936 (38.8%) | 20482 (39.9%) |
| **risk\_taking\_health\_scale** |  |  |  |
| Mean (SD) | 3.41 (2.49) | 2.75 (2.40) | 3.07 (2.47) |
| Median [Min, Max] | 3.00 [0, 10.0] | 2.00 [0, 10.0] | 3.00 [0, 10.0] |
| Missing | 10051 (39.0%) | 8721 (34.0%) | 18772 (36.5%) |
| **leadership\_position** |  |  |  |
| No | 11079 (43.0%) | 12789 (49.9%) | 23868 (46.5%) |
| Yes | 6592 (25.6%) | 3499 (13.7%) | 10091 (19.6%) |
| Missing | 8070 (31.4%) | 9341 (36.4%) | 17411 (33.9%) |
| **frequency\_homeoffice** |  |  |  |
| Never | 15824 (61.5%) | 17715 (69.1%) | 33539 (65.3%) |
| Rarely | 1770 (6.9%) | 1101 (4.3%) | 2871 (5.6%) |
| Regularly | 2149 (8.3%) | 1699 (6.6%) | 3848 (7.5%) |
| Missing | 5998 (23.3%) | 5114 (20.0%) | 11112 (21.6%) |
| **working\_overtime** |  |  |  |
| No | 13258 (51.5%) | 16012 (62.5%) | 29270 (57.0%) |
| Yes | 9945 (38.6%) | 8387 (32.7%) | 18332 (35.7%) |
| Self-Employed | 2498 (9.7%) | 1205 (4.7%) | 3703 (7.2%) |
| Missing | 40 (0.2%) | 25 (0.1%) | 65 (0.1%) |
| **worried\_health** |  |  |  |
| Yes | 4425 (17.2%) | 5019 (19.6%) | 9444 (18.4%) |
| Somewhat | 12543 (48.7%) | 13029 (50.8%) | 25572 (49.8%) |
| No | 8716 (33.9%) | 7534 (29.4%) | 16250 (31.6%) |
| Missing | 57 (0.2%) | 47 (0.2%) | 104 (0.2%) |
| **worried\_pension** |  |  |  |
| Yes | 3770 (14.6%) | 4724 (18.4%) | 8494 (16.5%) |
| Somewhat | 7440 (28.9%) | 8074 (31.5%) | 15514 (30.2%) |
| No | 7022 (27.3%) | 5691 (22.2%) | 12713 (24.7%) |
| Missing | 7509 (29.2%) | 7140 (27.9%) | 14649 (28.5%) |
| **career\_sacrifices** |  |  |  |
| No | 3926 (15.3%) | 4212 (16.4%) | 8138 (15.8%) |
| Rather no | 4799 (18.6%) | 4171 (16.3%) | 8970 (17.5%) |
| Rather yes | 3827 (14.9%) | 3372 (13.2%) | 7199 (14.0%) |
| Yes | 1355 (5.3%) | 1374 (5.4%) | 2729 (5.3%) |
| Missing | 11834 (46.0%) | 12500 (48.8%) | 24334 (47.4%) |
| **current\_life\_satisfaction\_scale** |  |  |  |
| Mean (SD) | 7.30 (1.75) | 7.33 (1.77) | 7.32 (1.76) |
| Median [Min, Max] | 8.00 [0, 10.0] | 8.00 [0, 10.0] | 8.00 [0, 10.0] |
| Missing | 234 (0.9%) | 240 (0.9%) | 474 (0.9%) |
| **political\_interest** |  |  |  |
| Not at all | 3654 (14.2%) | 5245 (20.5%) | 8899 (17.3%) |
| Not strong | 9586 (37.2%) | 12461 (48.6%) | 22047 (42.9%) |
| Strong | 9090 (35.3%) | 6530 (25.5%) | 15620 (30.4%) |
| Very strong | 3382 (13.1%) | 1375 (5.4%) | 4757 (9.3%) |
| Missing | 29 (0.1%) | 18 (0.1%) | 47 (0.1%) |
| **current\_health** |  |  |  |
| Very good | 3432 (13.3%) | 2831 (11.0%) | 6263 (12.2%) |
| Good, Satisfactory | 10350 (40.2%) | 9830 (38.4%) | 20180 (39.3%) |
| Less good | 7639 (29.7%) | 8005 (31.2%) | 15644 (30.5%) |
| Bad | 3222 (12.5%) | 3805 (14.8%) | 7027 (13.7%) |
| Missing | 1098 (4.3%) | 1158 (4.5%) | 2256 (4.4%) |
| **self\_esteem** |  |  |  |
| Mean (SD) | 5.78 (1.20) | 5.56 (1.34) | 5.67 (1.28) |
| Median [Min, Max] | 6.00 [1.00, 7.00] | 6.00 [1.00, 7.00] | 6.00 [1.00, 7.00] |
| Missing | 4955 (19.2%) | 4630 (18.1%) | 9585 (18.7%) |
| **health\_insurance\_status** |  |  |  |
| Pflichtversicherung | 12519 (48.6%) | 12874 (50.2%) | 25393 (49.4%) |
| Freiwillige Versichung | 3010 (11.7%) | 1612 (6.3%) | 4622 (9.0%) |
| Familienversicherung | 1672 (6.5%) | 3936 (15.4%) | 5608 (10.9%) |
| Rentnerversicherung | 4421 (17.2%) | 4679 (18.3%) | 9100 (17.7%) |
| Missing | 4119 (16.0%) | 2528 (9.9%) | 6647 (12.9%) |
| **alcohol\_consumption** |  |  |  |
| Rarely | 5833 (22.7%) | 8598 (33.5%) | 14431 (28.1%) |
| Regularly | 6747 (26.2%) | 5876 (22.9%) | 12623 (24.6%) |
| (Almost) Daily | 2981 (11.6%) | 1271 (5.0%) | 4252 (8.3%) |
| Missing | 10180 (39.5%) | 9884 (38.6%) | 20064 (39.1%) |
| **num\_physician\_visits** |  |  |  |
| Mean (SD) | 2.93 (3.96) | 3.10 (3.69) | 3.02 (3.82) |
| Median [Min, Max] | 2.00 [0, 90.0] | 2.00 [0, 96.0] | 2.00 [0, 96.0] |
| Missing | 2915 (11.3%) | 1728 (6.7%) | 4643 (9.0%) |
| **highest\_educational\_degree** |  |  |  |
| Mean (SD) | 3.90 (1.75) | 3.68 (1.66) | 3.79 (1.71) |
| Median [Min, Max] | 3.00 [0, 8.00] | 3.00 [0, 8.00] | 3.00 [0, 8.00] |
| Missing | 1429 (5.6%) | 1393 (5.4%) | 2822 (5.5%) |
| **smoke\_before\_stroke** |  |  |  |
| No | 14063 (54.6%) | 16903 (66.0%) | 30966 (60.3%) |
| Yes | 11678 (45.4%) | 8726 (34.0%) | 20404 (39.7%) |
| **smoke\_before\_migraine** |  |  |  |
| No | 14076 (54.7%) | 16943 (66.1%) | 31019 (60.4%) |
| Yes | 11665 (45.3%) | 8686 (33.9%) | 20351 (39.6%) |
| **diabetes\_before\_stroke** |  |  |  |
| No | 23440 (91.1%) | 23691 (92.4%) | 47131 (91.7%) |
| Yes | 2301 (8.9%) | 1938 (7.6%) | 4239 (8.3%) |
| **diabetes\_before\_migraine** |  |  |  |
| No | 23430 (91.0%) | 23732 (92.6%) | 47162 (91.8%) |
| Yes | 2311 (9.0%) | 1897 (7.4%) | 4208 (8.2%) |
| **hypertension\_before\_stroke** |  |  |  |
| No | 18443 (71.6%) | 18996 (74.1%) | 37439 (72.9%) |
| Yes | 7298 (28.4%) | 6633 (25.9%) | 13931 (27.1%) |
| **hypertension\_before\_migraine** |  |  |  |
| No | 18449 (71.7%) | 19172 (74.8%) | 37621 (73.2%) |
| Yes | 7292 (28.3%) | 6457 (25.2%) | 13749 (26.8%) |
| **migraine\_incidence** |  |  |  |
| No | 25218 (98.0%) | 24036 (93.8%) | 49254 (95.9%) |
| Yes | 523 (2.0%) | 1593 (6.2%) | 2116 (4.1%) |
| **stroke\_incidence** |  |  |  |
| No | 25303 (98.3%) | 25325 (98.8%) | 50628 (98.6%) |
| Yes | 438 (1.7%) | 304 (1.2%) | 742 (1.4%) |

## Missingness: Correlation Plot



TRUE

## Correlations: Spearman



TRUE

## Monthly Gross: Clip at Percentile 99.99

Current monthly gross labor incomes clipped at percentile 99.99 (100322 EUR)

| current\_monthly\_gross\_labor\_income |
| --- |
| 106,371 |
| 120,614 |
| 2,700,000 |
| 500,000 |

## Imputation

[1] "MICE imputations: m = 20"

MICE: Imputation methods applied

| Method | Variables |
| --- | --- |
| - | age, age\_group, age\_group\_at\_immigration, diabetes\_before\_migraine, diabetes\_before\_stroke, east\_german\_residence, hypertension\_before\_migraine, hypertension\_before\_stroke, immigration\_history, migraine\_incidence, num\_children\_in\_household, sex\_binary, smoke\_before\_migraine, smoke\_before\_stroke, stroke\_incidence, weight\_factor |
| logreg | current\_mat\_parent\_leave, diabetes, felt\_discriminated\_ethnic\_12mo, felt\_discriminated\_gender\_12mo, hypertension, leadership\_position, migraine, partner, sex\_entry\_change, sex\_or, smoke, stroke, using\_period\_of\_care |
| pmm | current\_life\_satisfaction\_scale, current\_monthly\_gross\_labor\_income, daily\_hours\_childcare\_weekdays, daily\_hours\_housework\_weekdays, highest\_educational\_degree, num\_physician\_visits, risk\_taking\_driving\_scale, risk\_taking\_health\_scale, risk\_taking\_scale, self\_esteem, work\_time |
| polr | alcohol\_consumption, career\_sacrifices, current\_health, employment\_status\_imp, frequency\_homeoffice, health\_insurance\_status, political\_interest, refugee\_exp, working\_overtime, worried\_health, worried\_pension |

## Save

Time difference of 1.50807 hours