Data Cleaning

Tejasvini Mavuleti

2022-07-28

library(lubridate) library(dplyr) library(reprex) library(ggplot2)

clean\_data <- function() {  
 patientdata = "C:/Users/mavul/OneDrive/Desktop/Health care data.csv"  
 if (file.exists(patientdata)) {  
 #dataset <- read.csv(file=patientdata, stringsAsFactors = FALSE, header = T)  
 dataset <- read.csv(file=patientdata, header = T)  
 }  
 return(dataset)  
}

age <- function(dob, age.day = today(), units = "years", floor = TRUE) {  
 calc.age = interval(dob, age.day) / duration(num = 1, units = units)  
 if (floor) return(as.integer(floor(calc.age)))  
   
 return(calc.age)  
}  
  
age\_group <- function(ag) {  
 ifelse(ag<30,20, ifelse(ag<40, 30, ifelse(ag<50,40, ifelse(ag<60, 50, ifelse(ag<70, 60,70)))))  
}

e\_europe <- c('Ukraine','Russia','Poland','Czech Republic','Hungary')  
w\_europe <- c('Austria','Belgium','France','Germany','Italy','Netherlands','Portugal','Spain','Switzerland')  
n\_europe <- c('Sweden', 'Finland', 'Denmark')  
c\_europe <- c('England','Scotland','Ireland')  
  
ethnic\_group <- function(countryname) {  
 ifelse((countryname %in% e\_europe), 1,  
 ifelse((countryname %in% w\_europe) ,2,  
 ifelse((countryname %in% n\_europe), 3,  
 ifelse((countryname %in% c\_europe), 4,  
 countryname))))  
 ifelse((countryname %in% e\_europe), 'e\_europe',  
 ifelse((countryname %in% w\_europe) ,'w\_europe',  
 ifelse((countryname %in% n\_europe), 'n\_europe',  
 ifelse((countryname %in% c\_europe), 'c\_europe',  
 countryname))))  
}

binary\_value <- function(value, compare\_to) {  
 ifelse(value==compare\_to,1,0)  
}

# Cleaning up data set  
patients <- clean\_data()  
str(patients)

## 'data.frame': 2000 obs. of 14 variables:  
## $ id : chr "Amelia Nixon" "Clara Hicks" "Mason Brown" "Michael Rice" ...  
## $ gender : chr "female" "female" "male" "male" ...  
## $ dob : chr "1944-03-09" "1966-07-02" "1981-05-31" "1945-02-13" ...  
## $ zipcode : int 89136 94105 89127 44101 89136 94105 60612 43221 89127 43210 ...  
## $ employment\_status : chr "retired" "employed" "employed" "retired" ...  
## $ education : chr "bachelors" "phd/md" "masters" "bachelors" ...  
## $ marital\_status : chr "married" "married" "married" "married" ...  
## $ children : int 1 4 2 2 3 2 0 2 2 7 ...  
## $ ancestry : chr "Portugal" "Sweden" "Germany" "Denmark" ...  
## $ avg\_commute : num 13.4 15.2 23.6 19.6 36.5 ...  
## $ daily\_internet\_use: num 2.53 6.77 3.63 5 7.75 3.34 6.75 3.01 4.12 3.15 ...  
## $ available\_vehicles: int 2 2 1 3 1 0 2 3 1 1 ...  
## $ military\_service : chr "no" "no" "no" "no" ...  
## $ disease : chr "hypertension" "endometriosis" "prostate cancer" "multiple sclerosis" ...

# Detect and delete emoty cells from education and correct spelling mistakes  
patients$education <- ifelse(patients$education == 'highscool', as.character('highschool'), as.character(patients$education))  
patients$education <- ifelse(as.factor(patients$education) == 'phD/MD', as.character('phd/md'), as.character(patients$education))  
patients$education <- as.factor(patients$education)

# Match ethnic groups with ancestry  
patients$ancestry <- as.factor(ethnic\_group(patients$ancestry))  
patients[ which(patients$employment\_status=='employed' & patients$age > 65), ]

## [1] id gender dob zipcode   
## [5] employment\_status education marital\_status children   
## [9] ancestry avg\_commute daily\_internet\_use available\_vehicles  
## [13] military\_service disease   
## <0 rows> (or 0-length row.names)

patients[ which(patients$employment\_status=='retired' & patients$age > 65), ]

## [1] id gender dob zipcode   
## [5] employment\_status education marital\_status children   
## [9] ancestry avg\_commute daily\_internet\_use available\_vehicles  
## [13] military\_service disease   
## <0 rows> (or 0-length row.names)

patients[ which(patients$employment\_status=='student' & patients$gender == 'male' & patients$disease == 'breast cancer'), ]

## [1] id gender dob zipcode   
## [5] employment\_status education marital\_status children   
## [9] ancestry avg\_commute daily\_internet\_use available\_vehicles  
## [13] military\_service disease   
## <0 rows> (or 0-length row.names)

patients[ which(patients$employment\_status=='student'), ]

## id gender dob zipcode employment\_status education  
## 410 Aubree Barron female 1990-03-09 43221 student phd/md  
## 459 Mason Lewis male 1988-01-15 90015 student phd/md  
## 472 Jeremiah Martin male 1966-01-12 10026 student masters  
## 555 Eva Evans female 1988-02-01 89136 student phd/md  
## 635 Sophia Lewis female 1987-08-10 89136 student phd/md  
## 650 Andrew White male 1990-11-22 60612 student phd/md  
## 707 Alyssa Meyer female 1991-10-14 60612 student phd/md  
## 736 Carson Brown male 1988-07-22 44101 student phd/md  
## 912 Emily Peters female 1962-07-30 94105 student masters  
## 952 Grace Patrick female 1977-01-07 89127 student masters  
## 974 Olivia Haney female 1954-06-23 10001 student bachelors  
## 1003 Emmanuel Brandt male 1990-12-02 94110 student phd/md  
## 1267 Annabelle Carrillo female 1998-07-06 89136 student bachelors  
## 1292 Zoe Roth female 1990-06-22 94105 student phd/md  
## 1332 Carter Bell male 1989-02-08 94105 student phd/md  
## 1364 Alexandra Young female 1987-01-30 89136 student phd/md  
## 1512 Angela Frazier female 1952-06-03 10026 student masters  
## 1531 Isabella Jensen female 1998-05-21 89136 student bachelors  
## 1555 Adalynn Waters female 1954-08-04 10001 student phd/md  
## 1654 Brooklyn Walker female 1987-11-12 94105 student phd/md  
## 1907 Colton Meyers male 1978-04-11 10026 student bachelors  
## marital\_status children ancestry avg\_commute daily\_internet\_use  
## 410 single 0 w\_europe 28.55 3.26  
## 459 single 0 n\_europe 17.45 6.46  
## 472 single 1 w\_europe 24.34 5.54  
## 555 single 0 n\_europe 15.90 4.56  
## 635 married 0 w\_europe 20.83 4.62  
## 650 single 0 e\_europe 38.69 5.73  
## 707 single 0 n\_europe 12.09 2.79  
## 736 single 0 e\_europe 28.10 6.12  
## 912 married 1 e\_europe 37.33 4.33  
## 952 married 1 c\_europe 48.53 4.64  
## 974 single 1 w\_europe 35.02 6.81  
## 1003 single 0 w\_europe 28.90 5.89  
## 1267 single 0 w\_europe 37.87 5.11  
## 1292 single 0 e\_europe 9.45 5.37  
## 1332 single 0 e\_europe 23.96 7.19  
## 1364 married 0 w\_europe 21.35 4.73  
## 1512 single 5 e\_europe 35.08 6.30  
## 1531 single 0 w\_europe 37.58 7.17  
## 1555 married 3 e\_europe 40.68 4.30  
## 1654 single 0 e\_europe 37.09 4.51  
## 1907 single 4 w\_europe 19.78 5.97  
## available\_vehicles military\_service disease  
## 410 1 no hypertension  
## 459 2 no kidney disease  
## 472 1 no gastritis  
## 555 2 no HIV/AIDS  
## 635 4 no schizophrenia  
## 650 0 no schizophrenia  
## 707 1 no diabetes  
## 736 0 no gastritis  
## 912 2 no Alzheimer's disease  
## 952 1 no hypertension  
## 974 3 no multiple sclerosis  
## 1003 1 no multiple sclerosis  
## 1267 1 no diabetes  
## 1292 1 no schizophrenia  
## 1332 0 no HIV/AIDS  
## 1364 0 no gastritis  
## 1512 1 no diabetes  
## 1531 1 no multiple sclerosis  
## 1555 2 no breast cancer  
## 1654 2 no schizophrenia  
## 1907 2 no hypertension

patients[ which(patients$employment\_status=='unemployed'), ]

## id gender dob zipcode employment\_status  
## 6 Sofia Wise female 1956-10-16 94105 unemployed  
## 13 Eli Wilson male 1949-02-09 94105 unemployed  
## 23 Jackson Alexander male 1961-03-30 94110 unemployed  
## 52 Vivian Daniels female 1962-09-09 94105 unemployed  
## 74 William Roberson male 1979-11-27 90008 unemployed  
## 76 Mason Pena male 1951-05-25 94105 unemployed  
## 80 Owen Brown male 1949-04-28 60612 unemployed  
## 82 Grayson Rodgers male 1949-11-27 10001 unemployed  
## 90 Chloe Estrada female 1949-09-30 10001 unemployed  
## 98 Bailey Morgan female 1962-06-01 90015 unemployed  
## 101 Mila Lopez female 1952-09-10 89127 unemployed  
## 107 Connor Aguirre male 1956-11-06 89136 unemployed  
## 108 Justin Swanson male 1954-06-11 94105 unemployed  
## 110 Sophia Villarreal female 1951-12-24 10026 unemployed  
## 114 Emma Cooper female 1965-05-26 43210 unemployed  
## 116 Ryan Solis male 1962-11-25 44101 unemployed  
## 117 Ryan Smith male 1969-06-03 89136 unemployed  
## 127 Maya Brown female 1965-08-21 60605 unemployed  
## 131 Samantha Harris female 1956-06-24 43210 unemployed  
## 144 Alexis Carpenter female 1957-10-24 94105 unemployed  
## 146 Audrey Buck female 1982-11-23 60605 unemployed  
## 156 Christopher Sullivan male 1950-06-08 44101 unemployed  
## 157 Aaron Mcdaniel male 1948-09-12 10001 unemployed  
## 159 Clara Ortiz female 1957-06-30 10001 unemployed  
## 173 Anthony Frank male 1982-10-15 90015 unemployed  
## 175 Brooklyn Moore female 1958-04-29 94110 unemployed  
## 176 Dominic Vazquez male 1953-08-20 10001 unemployed  
## 177 Aiden Thomas male 1950-12-11 89127 unemployed  
## 201 Jack Gray male 1964-11-27 10001 unemployed  
## 214 Zoe Walton female 1979-05-14 10026 unemployed  
## 233 Payton Perez female 1953-01-31 94105 unemployed  
## 237 Harper Jackson female 1949-09-15 94110 unemployed  
## 240 Makenzie Newton female 1960-01-02 89127 unemployed  
## 261 Aubree Navarro female 1969-04-18 60605 unemployed  
## 281 Micah Ingram male 1954-08-07 89127 unemployed  
## 283 Jeremiah Banks male 1957-06-05 43210 unemployed  
## 284 Eleanor Stewart female 1951-11-06 89136 unemployed  
## 289 Zoe Jackson female 1957-12-17 43221 unemployed  
## 295 Bentley Wilson male 1967-08-23 90008 unemployed  
## 325 Hudson Pennington male 1959-02-06 60605 unemployed  
## 326 Joshua Mason male 1961-04-02 94105 unemployed  
## 335 Amelia Gardner female 1947-06-28 60605 unemployed  
## 338 Elena Jones female 1979-07-08 43210 unemployed  
## 341 Samuel Parker male 1949-03-17 10001 unemployed  
## 349 Wyatt Hanna male 1955-04-23 94110 unemployed  
## 356 Savannah Glass female 1954-12-31 90008 unemployed  
## 369 Colin Harding male 1947-03-03 90008 unemployed  
## 373 Jaxon Walter male 1948-06-25 10001 unemployed  
## 379 Payton Neal female 1964-09-05 43210 unemployed  
## 384 Hunter Medina male 1970-09-23 94105 unemployed  
## 404 Ariel Mason female 1954-06-20 94105 unemployed  
## 409 Evan Morrow male 1958-12-13 89127 unemployed  
## 413 Hannah Owens female 1954-12-31 60605 unemployed  
## 430 Jackson Perez male 1967-01-25 10001 unemployed  
## 435 Samantha Thompson female 1969-05-27 94110 unemployed  
## 438 Brooklyn Mcclain female 1954-05-11 60605 unemployed  
## 453 Ethan Michael male 1969-10-02 94105 unemployed  
## 463 Nolan English male 1983-08-06 43221 unemployed  
## 470 Daniel Sanchez male 1951-04-26 89127 unemployed  
## 475 Ivan Beard male 1955-05-08 94105 unemployed  
## 478 Abigail Ryan female 1957-11-17 89136 unemployed  
## 482 Brooklyn Roman female 1966-12-08 90008 unemployed  
## 483 Daisy Fields female 1951-06-19 60605 unemployed  
## 495 Caleb Kennedy male 1948-03-03 60605 unemployed  
## 502 Amelia Vazquez female 1947-01-15 89127 unemployed  
## 503 Natalia Cox female 1953-08-16 43221 unemployed  
## 504 Nevaeh Mccoy female 1955-03-24 60605 unemployed  
## 544 Parker Mendoza male 1960-08-01 90008 unemployed  
## 556 James Baxter male 1956-05-28 10026 unemployed  
## 595 Alyssa Arnold female 1951-07-31 89127 unemployed  
## 597 Easton Dickerson male 1947-08-08 43221 unemployed  
## 598 Jacob Young male 1949-03-29 94105 unemployed  
## 602 Austin Holmes male 1978-11-16 10001 unemployed  
## 604 Leah Shaffer female 1953-03-09 60605 unemployed  
## 608 Delilah Smith female 1961-07-09 90008 unemployed  
## 615 Lydia Richardson female 1971-04-04 10026 unemployed  
## 618 Aubrey Macias female 1954-01-25 89127 unemployed  
## 636 Hunter Mann male 1981-12-20 43221 unemployed  
## 640 Kevin Gutierrez male 1972-06-14 94110 unemployed  
## 643 Reese Johnson female 1967-01-06 60605 unemployed  
## 660 Parker Brown male 1973-09-19 10001 unemployed  
## 669 Olivia Phillips female 1961-08-30 43210 unemployed  
## 691 Ethan Huber male 1968-06-21 60605 unemployed  
## 699 Alice Coleman female 1967-11-05 43210 unemployed  
## 704 Kendall Roberts female 1964-08-29 43210 unemployed  
## 709 Dylan Ayala male 1948-11-19 94105 unemployed  
## 713 Morgan Mills female 1971-10-11 44101 unemployed  
## 720 Melody Rhodes female 1957-12-13 43221 unemployed  
## 732 Carter Copeland male 1957-10-17 60605 unemployed  
## 735 Jameson Wells male 1969-12-11 89136 unemployed  
## 742 Ariana Miller female 1947-10-31 44101 unemployed  
## 763 Margaret Hunt female 1953-01-24 43221 unemployed  
## 779 Samuel Richards male 1956-07-20 43221 unemployed  
## 780 Oscar Lewis male 1959-09-22 94105 unemployed  
## 786 Ariana Cunningham female 1951-06-25 90015 unemployed  
## 811 Ryan Cox male 1948-10-05 10026 unemployed  
## 835 Lucas Jones male 1977-08-03 90015 unemployed  
## 848 Grace Wong female 1952-07-21 43221 unemployed  
## 893 Katelyn Dickerson female 1959-07-03 60612 unemployed  
## 896 Emily Townsend female 1970-05-12 60612 unemployed  
## 902 Jacob Smith male 1962-07-03 94110 unemployed  
## 915 Aaron Dawson male 1970-03-18 43221 unemployed  
## 916 William Lynch male 1956-06-07 94110 unemployed  
## 937 Margaret Reed female 1962-09-19 89127 unemployed  
## 940 Daisy James female 1975-09-18 89136 unemployed  
## 944 Avery Abbott female 1976-05-26 89127 unemployed  
## 962 Kimberly Nielsen female 1948-04-22 89127 unemployed  
## 964 Connor Salazar male 1947-06-16 10001 unemployed  
## 966 Hazel Lewis female 1959-12-12 94105 unemployed  
## 976 Matthew Davila male 1962-05-07 94105 unemployed  
## 977 Aaliyah Blair female 1961-12-14 89136 unemployed  
## 999 Aaron Watts male 1954-12-08 94110 unemployed  
## 1004 Anna Silva female 1958-10-13 60612 unemployed  
## 1010 Joel Jones male 1949-07-07 10026 unemployed  
## 1011 Elizabeth Holland female 1954-11-20 94110 unemployed  
## 1024 Julian Long male 1960-02-06 89127 unemployed  
## 1028 Avery Shaffer female 1966-03-22 60605 unemployed  
## 1029 Jessica Hall female 1978-05-18 89127 unemployed  
## 1032 Brayden Roberts male 1967-08-22 94110 unemployed  
## 1037 Caleb Vargas male 1955-11-23 60605 unemployed  
## 1040 Joshua Meyers male 1964-09-05 43210 unemployed  
## 1044 Jayden Foster male 1963-04-09 44101 unemployed  
## 1050 Allison Lopez female 1960-03-23 89136 unemployed  
## 1057 Sarah Potter female 1962-09-07 90008 unemployed  
## 1060 Elijah Pearson male 1955-06-06 89127 unemployed  
## 1066 Harper King female 1959-09-28 60612 unemployed  
## 1072 Mia Hart female 1952-07-29 43221 unemployed  
## 1073 Brady Durham male 1951-12-07 90015 unemployed  
## 1079 Colin Wise male 1948-03-22 60605 unemployed  
## 1081 Joel Williams male 1962-10-08 90015 unemployed  
## 1095 Connor Roach male 1953-12-03 43221 unemployed  
## 1096 Lillian Salinas female 1947-10-11 43210 unemployed  
## 1112 Aliyah Jones female 1954-09-21 44101 unemployed  
## 1114 Daniel Ellis male 1954-11-03 94105 unemployed  
## 1124 Nolan Brewer male 1975-01-30 90008 unemployed  
## 1130 Avery Lowery male 1978-09-20 60612 unemployed  
## 1132 Caleb Mcconnell male 1974-12-11 10001 unemployed  
## 1134 Riley Watson female 1951-11-20 60612 unemployed  
## 1135 Emmanuel King male 1954-05-16 90015 unemployed  
## 1137 Ava Gardner female 1947-09-15 94110 unemployed  
## 1165 Abigail Santiago female 1950-04-25 90008 unemployed  
## 1183 Riley Bailey female 1951-02-24 60605 unemployed  
## 1191 Gavin Wheeler male 1962-07-18 94110 unemployed  
## 1196 Aliyah Vega female 1961-06-28 89127 unemployed  
## 1197 Mia Palmer female 1987-09-07 43221 unemployed  
## 1204 Leah Miller female 1971-03-04 60605 unemployed  
## 1206 Adrian Montgomery male 1960-08-07 43221 unemployed  
## 1212 Penelope Cain female 1959-12-01 43221 unemployed  
## 1220 Richard Moore male 1958-09-17 89136 unemployed  
## 1224 Annabelle Kelly female 1960-09-10 60605 unemployed  
## 1225 Michael Smith male 1976-10-03 10001 unemployed  
## 1230 Izabella Peterson female 1974-12-28 90008 unemployed  
## 1231 Mia Moody female 1948-09-30 10001 unemployed  
## 1244 Matthew Pratt male 1966-10-12 43210 unemployed  
## 1248 Addison Walker female 1963-05-09 43210 unemployed  
## 1259 Skylar Lewis female 1971-11-08 10001 unemployed  
## 1266 Laila Hubbard female 1949-08-29 89127 unemployed  
## 1270 Eric Lloyd male 1957-02-21 10001 unemployed  
## 1273 Caleb Crane male 1980-01-03 60612 unemployed  
## 1276 Kayla Romero female 1955-05-29 10001 unemployed  
## 1293 Joshua Martin male 1952-07-29 94105 unemployed  
## 1296 Lincoln Howard male 1951-03-27 43210 unemployed  
## 1300 William Ali male 1969-04-13 90015 unemployed  
## 1306 Anna Castro female 1971-09-01 10001 unemployed  
## 1322 Mia Curtis female 1965-06-23 43221 unemployed  
## 1333 Jaxson Parker male 1949-09-15 44101 unemployed  
## 1349 Evelyn Baker female 1962-06-26 94105 unemployed  
## 1370 Isabella Cisneros female 1953-09-26 90008 unemployed  
## 1371 Genesis Riley female 1967-03-10 44101 unemployed  
## 1372 Abigail Rodriguez female 1968-02-03 94110 unemployed  
## 1374 Lyla Rodriguez female 1955-11-10 60605 unemployed  
## 1377 Autumn Guzman female 1949-08-26 90008 unemployed  
## 1387 Robert Dickson male 1961-12-31 10001 unemployed  
## 1402 Gabriella Nichols female 1948-07-03 89127 unemployed  
## 1409 Logan Hall male 1960-04-25 94105 unemployed  
## 1411 Wyatt Santiago male 1966-07-22 89127 unemployed  
## 1412 Santiago Goodman male 1973-08-18 60605 unemployed  
## 1414 Sofia Miller female 1955-05-19 10001 unemployed  
## 1422 Jasmine Lara female 1981-01-15 90015 unemployed  
## 1429 Kayden Young male 1953-03-07 10001 unemployed  
## 1435 Emma Thompson female 1983-12-08 89136 unemployed  
## 1496 Eleanor Willis female 1947-07-03 43210 unemployed  
## 1497 Lincoln Carroll male 1989-03-25 94110 unemployed  
## 1500 Arianna Moore female 1955-04-03 89127 unemployed  
## 1502 Isabelle Conrad female 1952-03-26 60605 unemployed  
## 1504 Mila Crawford female 1953-04-18 94110 unemployed  
## 1505 Gabriel Gentry male 1965-01-25 43221 unemployed  
## 1518 Faith Rivera female 1965-12-01 94105 unemployed  
## 1552 Michael Brown male 1967-06-17 90015 unemployed  
## 1563 Hannah Brady female 1947-05-02 10001 unemployed  
## 1564 Kaleb Mack male 1965-04-28 94105 unemployed  
## 1565 Bailey Morris female 1952-05-12 90008 unemployed  
## 1569 Faith Li female 1976-07-11 10001 unemployed  
## 1573 Dylan Diaz male 1952-12-20 90015 unemployed  
## 1575 Olivia Marks female 1977-10-20 44101 unemployed  
## 1581 Molly Wilson female 1962-10-25 94110 unemployed  
## 1585 John Wiggins male 1953-10-17 89136 unemployed  
## 1590 Julian Villarreal male 1980-12-15 94105 unemployed  
## 1597 Trinity Torres female 1976-08-04 94110 unemployed  
## 1598 Mia Wilson female 1971-08-15 43210 unemployed  
## 1600 Anthony Cordova male 1973-06-18 44101 unemployed  
## 1604 Parker Garcia male 1957-06-12 43221 unemployed  
## 1606 Lucas Gregory male 1949-05-11 44101 unemployed  
## 1611 Lillian Brock female 1963-07-07 60612 unemployed  
## 1615 Michael Mercado male 1955-12-19 89127 unemployed  
## 1626 Juliana Johnson female 1956-03-11 60612 unemployed  
## 1635 Isabella Perkins female 1958-10-15 94105 unemployed  
## 1638 Kingston Baker male 1953-01-06 10026 unemployed  
## 1646 Rylee Cooper female 1965-04-30 89136 unemployed  
## 1650 Nora Lloyd female 1954-04-02 60605 unemployed  
## 1660 Naomi Gallagher female 1950-11-10 43210 unemployed  
## 1669 Emily Johnson female 1958-09-08 94105 unemployed  
## 1677 Kaylee Johnson female 1952-04-04 10001 unemployed  
## 1683 Londyn Wood female 1972-03-18 90008 unemployed  
## 1700 Henry Francis male 1947-10-07 94110 unemployed  
## 1710 Ariana Kim female 1952-09-24 90008 unemployed  
## 1722 Jeremiah Powell male 1982-07-03 89127 unemployed  
## 1739 Victor Stewart male 1971-11-19 94110 unemployed  
## 1742 Colin Kelly male 1950-10-31 94110 unemployed  
## 1750 Jayce Cole male 1955-08-22 90008 unemployed  
## 1756 Jack Smith male 1947-08-20 60612 unemployed  
## 1760 Logan Ryan male 1957-07-27 94110 unemployed  
## 1768 Carson Allen male 1950-07-02 44101 unemployed  
## 1776 Elise Ramirez female 1958-06-03 60605 unemployed  
## 1787 Bailey King female 1967-11-23 89127 unemployed  
## 1791 Angel Conley male 1955-02-03 60605 unemployed  
## 1795 Maya Smith female 1973-03-08 43221 unemployed  
## 1797 Ariana Rogers female 1975-08-31 10026 unemployed  
## 1806 Kayla Carroll female 1986-04-01 43210 unemployed  
## 1810 Connor Reyes male 1959-03-15 10001 unemployed  
## 1815 Miguel Woodard male 1953-12-17 60605 unemployed  
## 1823 Ariel Lewis female 1950-06-19 43210 unemployed  
## 1824 David Dennis male 1958-01-12 43221 unemployed  
## 1841 Naomi Schultz female 1956-04-07 94105 unemployed  
## 1859 Logan Moore male 1957-07-21 90008 unemployed  
## 1862 Joshua Davis male 1957-01-10 43221 unemployed  
## 1869 Elijah Watson male 1949-04-20 90015 unemployed  
## 1873 Noah Wise male 1964-12-03 10001 unemployed  
## 1887 William Cameron male 1974-09-13 10026 unemployed  
## 1898 Ella Smith female 1948-08-18 89127 unemployed  
## 1911 Timothy Munoz male 1956-09-16 94110 unemployed  
## 1913 Madison Owen female 1966-07-07 60612 unemployed  
## 1916 Asher Martin male 1970-09-10 89127 unemployed  
## 1918 Riley Christensen male 1950-10-01 43221 unemployed  
## 1921 Brandon Williams male 1958-11-03 90008 unemployed  
## 1929 Sophia Ortiz female 1955-09-02 94105 unemployed  
## 1932 Aaliyah Gutierrez female 1983-02-21 10001 unemployed  
## 1935 Micah Nelson male 1958-07-27 43210 unemployed  
## 1938 Alexander Mcdowell male 1959-06-04 94105 unemployed  
## 1940 Mateo Malone male 1962-04-17 90008 unemployed  
## 1970 Austin Carey male 1949-05-23 60612 unemployed  
## 1972 Sophia Hopkins female 1959-01-30 94105 unemployed  
## 1993 Zoey Green female 1972-12-20 60612 unemployed  
## 1994 Mia Acosta female 1950-04-15 94105 unemployed  
## 1997 Natalia Dominguez female 1963-05-10 60612 unemployed  
## education marital\_status children ancestry avg\_commute daily\_internet\_use  
## 6 highschool married 2 w\_europe 48.68 3.34  
## 13 highschool married 3 c\_europe 49.61 2.52  
## 23 highschool married 2 w\_europe 43.83 6.91  
## 52 highschool married 2 n\_europe 37.11 4.93  
## 74 highschool single 0 e\_europe 41.11 6.36  
## 76 highschool married 4 w\_europe 33.71 5.55  
## 80 highschool single 0 w\_europe 25.76 3.31  
## 82 highschool married 2 w\_europe 36.19 3.33  
## 90 highschool married 0 w\_europe 29.26 6.38  
## 98 bachelors married 3 w\_europe 24.17 4.51  
## 101 bachelors married 2 w\_europe 17.83 6.35  
## 107 highschool single 7 c\_europe 38.73 5.13  
## 108 highschool single 3 w\_europe 41.08 3.57  
## 110 highschool married 2 w\_europe 23.18 5.02  
## 114 highschool single 6 w\_europe 40.85 5.91  
## 116 highschool single 0 w\_europe 38.08 3.10  
## 117 highschool married 2 n\_europe 13.94 5.36  
## 127 highschool married 2 c\_europe 30.09 4.23  
## 131 highschool married 3 c\_europe 38.67 4.53  
## 144 highschool single 3 w\_europe 30.36 6.01  
## 146 bachelors single 2 c\_europe 39.31 6.79  
## 156 highschool married 2 n\_europe 47.52 5.40  
## 157 highschool single 1 n\_europe 41.47 6.60  
## 159 highschool single 3 w\_europe 43.59 5.38  
## 173 bachelors married 2 w\_europe 27.29 3.95  
## 175 highschool married 1 w\_europe 31.17 6.99  
## 176 bachelors single 2 w\_europe 34.37 7.00  
## 177 highschool married 1 e\_europe 37.40 3.54  
## 201 bachelors married 0 n\_europe 37.05 4.85  
## 214 highschool single 2 e\_europe 35.89 4.18  
## 233 bachelors single 6 w\_europe 38.90 5.39  
## 237 highschool married 1 n\_europe 42.83 4.13  
## 240 bachelors single 3 e\_europe 25.82 5.96  
## 261 highschool single 2 c\_europe 31.52 5.10  
## 281 bachelors married 3 c\_europe 33.81 5.45  
## 283 highschool married 1 c\_europe 30.80 5.18  
## 284 bachelors single 3 n\_europe 37.07 6.59  
## 289 highschool married 3 n\_europe 44.80 7.13  
## 295 highschool single 4 e\_europe 23.50 1.71  
## 325 highschool married 1 w\_europe -0.74 6.61  
## 326 bachelors married 2 e\_europe 14.80 3.48  
## 335 highschool married 3 e\_europe 41.32 5.80  
## 338 highschool married 2 n\_europe 39.38 5.32  
## 341 highschool single 2 w\_europe 36.43 6.39  
## 349 highschool single 0 w\_europe 17.02 7.87  
## 356 highschool single 3 e\_europe 15.75 5.99  
## 369 highschool married 7 w\_europe 25.86 4.58  
## 373 highschool single 3 w\_europe 30.84 1.80  
## 379 bachelors married 1 e\_europe 15.16 3.90  
## 384 highschool married 5 w\_europe 39.91 5.06  
## 404 bachelors married 3 e\_europe 25.67 2.08  
## 409 highschool single 2 n\_europe 15.77 4.88  
## 413 highschool married 1 w\_europe 28.45 7.79  
## 430 highschool single 5 e\_europe 36.59 5.41  
## 435 highschool single 4 w\_europe 35.31 1.39  
## 438 highschool single 2 w\_europe 23.31 5.85  
## 453 highschool married 4 e\_europe 31.77 3.70  
## 463 highschool married 3 w\_europe 26.17 5.17  
## 470 highschool single 0 c\_europe 29.50 6.88  
## 475 highschool single 1 w\_europe 29.76 6.35  
## 478 highschool married 0 w\_europe 33.81 4.73  
## 482 highschool single 0 w\_europe 19.61 4.43  
## 483 highschool married 2 w\_europe 43.98 6.41  
## 495 highschool married 1 w\_europe 36.61 5.56  
## 502 highschool single 0 w\_europe 29.54 5.16  
## 503 highschool married 2 c\_europe 44.23 8.08  
## 504 bachelors married 4 c\_europe 39.81 5.83  
## 544 highschool single 3 e\_europe 24.47 6.18  
## 556 highschool single 3 w\_europe 22.53 3.05  
## 595 bachelors single 2 n\_europe 54.10 2.40  
## 597 highschool married 1 w\_europe 27.90 8.82  
## 598 highschool married 3 n\_europe 28.52 5.98  
## 602 highschool single 4 n\_europe 16.49 7.20  
## 604 highschool married 2 e\_europe 29.45 5.48  
## 608 bachelors single 3 e\_europe 43.79 4.47  
## 615 bachelors married 1 w\_europe 17.08 3.85  
## 618 highschool single 1 c\_europe 52.58 5.99  
## 636 highschool single 2 w\_europe 29.05 6.20  
## 640 highschool single 2 w\_europe 44.94 5.12  
## 643 highschool single 1 w\_europe 28.71 4.63  
## 660 highschool single 2 c\_europe 45.95 3.24  
## 669 bachelors married 2 c\_europe 21.47 4.98  
## 691 highschool married 3 n\_europe 26.89 6.72  
## 699 highschool single 2 w\_europe 44.19 5.90  
## 704 bachelors single 1 n\_europe 22.05 3.62  
## 709 bachelors single 5 n\_europe 31.25 3.68  
## 713 bachelors single 2 w\_europe 22.69 6.07  
## 720 highschool married 0 w\_europe 21.19 3.63  
## 732 highschool married 3 w\_europe 23.65 4.87  
## 735 highschool married 2 w\_europe 18.86 2.76  
## 742 highschool single 1 n\_europe 11.31 5.94  
## 763 highschool married 2 n\_europe 32.07 4.81  
## 779 bachelors single 3 w\_europe 36.39 6.25  
## 780 highschool single 5 w\_europe 21.29 2.50  
## 786 highschool single 2 e\_europe 35.41 4.36  
## 811 highschool single 6 w\_europe 38.70 4.32  
## 835 highschool married 3 w\_europe 10.44 8.55  
## 848 highschool married 2 e\_europe 49.02 2.26  
## 893 bachelors single 1 e\_europe 34.53 6.43  
## 896 highschool married 2 w\_europe 34.32 5.91  
## 902 highschool single 3 w\_europe 31.70 7.26  
## 915 highschool single 2 w\_europe 31.98 4.76  
## 916 highschool single 2 w\_europe 32.04 4.15  
## 937 highschool single 6 w\_europe 31.53 2.97  
## 940 bachelors married 2 c\_europe 28.98 5.66  
## 944 highschool single 1 e\_europe 35.82 4.15  
## 962 bachelors single 6 c\_europe 16.58 3.32  
## 964 highschool married 1 w\_europe 25.54 5.74  
## 966 highschool single 1 c\_europe 41.56 5.54  
## 976 highschool married 2 e\_europe 41.23 5.95  
## 977 highschool single 2 e\_europe 34.59 4.51  
## 999 highschool single 2 w\_europe 40.94 3.72  
## 1004 highschool married 1 c\_europe 31.83 3.64  
## 1010 bachelors married 1 w\_europe 33.30 4.57  
## 1011 bachelors single 5 n\_europe 22.34 5.26  
## 1024 highschool single 2 w\_europe 41.91 4.27  
## 1028 bachelors single 0 w\_europe 24.44 6.02  
## 1029 bachelors single 0 e\_europe 27.05 7.06  
## 1032 highschool married 2 w\_europe 15.18 7.06  
## 1037 highschool single 4 n\_europe 26.37 7.11  
## 1040 highschool single 1 w\_europe 31.34 6.59  
## 1044 highschool married 0 w\_europe 47.52 7.99  
## 1050 highschool married 2 n\_europe 28.84 5.47  
## 1057 highschool single 1 n\_europe 38.12 7.02  
## 1060 bachelors single 2 c\_europe 36.14 7.15  
## 1066 highschool single 0 e\_europe 23.77 5.40  
## 1072 highschool married 2 w\_europe 40.73 6.31  
## 1073 highschool married 1 e\_europe 17.68 5.24  
## 1079 highschool single 1 w\_europe 24.05 3.93  
## 1081 highschool single 2 e\_europe 38.75 5.12  
## 1095 bachelors single 1 e\_europe 20.68 5.03  
## 1096 bachelors single 2 w\_europe 46.04 5.70  
## 1112 highschool married 3 w\_europe 30.21 3.72  
## 1114 highschool single 1 c\_europe 35.01 4.56  
## 1124 highschool single 1 w\_europe 28.37 2.76  
## 1130 highschool single 3 n\_europe 32.99 6.18  
## 1132 highschool single 3 w\_europe 28.68 3.79  
## 1134 bachelors single 1 e\_europe 39.21 7.88  
## 1135 highschool married 1 e\_europe 23.16 5.43  
## 1137 highschool single 6 w\_europe 33.67 4.51  
## 1165 highschool single 2 c\_europe 30.77 3.93  
## 1183 highschool married 0 w\_europe 24.16 8.09  
## 1191 highschool married 1 e\_europe 27.29 4.03  
## 1196 highschool single 1 w\_europe 32.60 3.83  
## 1197 highschool married 0 w\_europe 34.22 5.92  
## 1204 highschool married 2 w\_europe 27.69 6.58  
## 1206 highschool married 2 e\_europe 27.02 4.02  
## 1212 highschool single 2 n\_europe 31.15 4.24  
## 1220 highschool married 1 w\_europe 32.53 6.26  
## 1224 bachelors single 2 w\_europe 38.40 5.78  
## 1225 highschool married 2 w\_europe 23.81 3.22  
## 1230 highschool single 0 w\_europe 21.68 4.58  
## 1231 highschool married 2 w\_europe 28.49 4.92  
## 1244 highschool married 1 w\_europe 2.75 4.74  
## 1248 bachelors married 1 w\_europe 57.37 4.19  
## 1259 highschool married 2 w\_europe 32.74 4.46  
## 1266 highschool married 4 n\_europe 24.98 6.13  
## 1270 highschool single 3 w\_europe 27.68 3.05  
## 1273 highschool married 2 w\_europe 29.94 5.04  
## 1276 highschool single 2 n\_europe 31.00 3.70  
## 1293 bachelors single 3 n\_europe 16.60 7.28  
## 1296 bachelors single 0 w\_europe 11.05 4.92  
## 1300 bachelors married 0 w\_europe 31.77 2.40  
## 1306 highschool single 2 n\_europe 32.45 4.08  
## 1322 bachelors single 1 e\_europe 35.00 2.63  
## 1333 highschool single 4 c\_europe 29.53 2.66  
## 1349 highschool married 1 w\_europe 31.12 3.77  
## 1370 highschool married 1 e\_europe 34.48 3.16  
## 1371 highschool single 3 n\_europe 31.71 4.23  
## 1372 highschool single 3 n\_europe 40.64 5.24  
## 1374 highschool married 3 c\_europe 24.80 5.18  
## 1377 highschool married 1 c\_europe 20.71 4.20  
## 1387 highschool single 2 w\_europe 50.81 4.71  
## 1402 bachelors single 2 n\_europe 30.07 6.83  
## 1409 highschool single 3 n\_europe 55.86 7.59  
## 1411 bachelors married 2 e\_europe 26.20 5.60  
## 1412 highschool married 6 e\_europe 40.64 6.19  
## 1414 highschool single 3 n\_europe 44.99 5.26  
## 1422 bachelors single 2 n\_europe 39.13 5.80  
## 1429 bachelors married 2 n\_europe 37.49 5.28  
## 1435 bachelors single 6 e\_europe 24.07 6.22  
## 1496 bachelors single 4 e\_europe 24.71 1.57  
## 1497 highschool married 0 w\_europe 43.80 6.73  
## 1500 highschool single 6 w\_europe 5.96 6.16  
## 1502 bachelors married 2 w\_europe 35.15 7.74  
## 1504 highschool married 3 n\_europe 37.25 4.10  
## 1505 highschool married 3 n\_europe 18.82 4.01  
## 1518 bachelors married 1 c\_europe 38.80 3.06  
## 1552 bachelors single 2 w\_europe 30.84 4.22  
## 1563 highschool single 4 e\_europe 27.07 6.53  
## 1564 highschool married 4 n\_europe 18.54 5.72  
## 1565 highschool single 4 w\_europe 32.57 5.74  
## 1569 bachelors single 2 e\_europe 20.43 3.53  
## 1573 highschool single 1 n\_europe 32.53 5.02  
## 1575 bachelors married 1 w\_europe 32.95 4.59  
## 1581 bachelors single 2 w\_europe 15.83 8.04  
## 1585 highschool married 3 w\_europe 36.66 6.70  
## 1590 bachelors married 2 w\_europe 19.95 5.98  
## 1597 highschool married 3 w\_europe 27.83 2.14  
## 1598 highschool married 6 n\_europe 17.46 2.74  
## 1600 highschool single 1 w\_europe 30.68 4.35  
## 1604 highschool married 1 w\_europe 18.62 3.58  
## 1606 bachelors single 5 e\_europe 26.69 1.94  
## 1611 highschool single 5 e\_europe 23.72 4.18  
## 1615 highschool single 1 w\_europe 18.38 3.23  
## 1626 highschool married 3 e\_europe 31.46 3.27  
## 1635 highschool single 2 c\_europe 38.26 6.61  
## 1638 highschool married 0 e\_europe 33.59 5.41  
## 1646 highschool married 0 w\_europe 26.30 5.68  
## 1650 highschool single 4 w\_europe 42.74 2.39  
## 1660 bachelors single 2 n\_europe 5.94 7.03  
## 1669 highschool single 1 w\_europe 35.89 5.08  
## 1677 highschool single 2 e\_europe 32.11 5.69  
## 1683 bachelors single 1 w\_europe 30.37 4.77  
## 1700 highschool single 1 w\_europe 42.75 4.58  
## 1710 bachelors single 2 w\_europe 21.90 3.15  
## 1722 highschool married 2 e\_europe 22.61 5.05  
## 1739 highschool married 2 w\_europe 28.03 3.48  
## 1742 highschool single 1 w\_europe 29.16 6.16  
## 1750 highschool single 2 w\_europe 32.47 4.46  
## 1756 highschool married 1 n\_europe 35.49 1.87  
## 1760 highschool married 1 w\_europe 34.43 2.77  
## 1768 highschool single 7 w\_europe 36.16 7.54  
## 1776 highschool married 4 w\_europe 35.66 4.91  
## 1787 bachelors married 2 e\_europe 17.46 3.20  
## 1791 bachelors single 2 e\_europe 8.05 2.69  
## 1795 highschool married 3 w\_europe 33.33 4.60  
## 1797 highschool single 3 n\_europe 26.36 5.34  
## 1806 highschool single 7 e\_europe 28.52 5.47  
## 1810 highschool single 4 w\_europe 41.34 6.10  
## 1815 highschool married 3 c\_europe 33.66 4.54  
## 1823 highschool single 3 w\_europe 36.44 5.40  
## 1824 bachelors married 3 c\_europe 39.81 5.44  
## 1841 highschool married 4 c\_europe 39.33 3.66  
## 1859 bachelors single 2 n\_europe 22.17 5.03  
## 1862 highschool single 2 w\_europe 53.31 5.38  
## 1869 highschool married 0 e\_europe 33.63 5.58  
## 1873 highschool single 1 w\_europe 33.00 5.99  
## 1887 highschool married 1 n\_europe 55.12 3.61  
## 1898 highschool married 7 c\_europe 33.58 6.65  
## 1911 highschool married 2 w\_europe 27.96 7.03  
## 1913 bachelors married 2 e\_europe 20.57 5.44  
## 1916 highschool single 2 w\_europe 7.53 2.15  
## 1918 highschool married 4 n\_europe 27.43 6.54  
## 1921 bachelors married 4 e\_europe 49.93 5.44  
## 1929 highschool single 3 e\_europe 18.14 5.98  
## 1932 highschool single 4 w\_europe 25.85 6.59  
## 1935 bachelors single 4 e\_europe 41.63 3.81  
## 1938 highschool married 2 e\_europe 17.23 5.27  
## 1940 bachelors single 2 w\_europe 24.48 4.70  
## 1970 highschool married 3 e\_europe 47.71 2.72  
## 1972 highschool married 0 e\_europe 34.71 4.98  
## 1993 highschool single 4 e\_europe 28.50 7.08  
## 1994 bachelors married 4 w\_europe 24.46 6.69  
## 1997 highschool married 4 n\_europe 21.09 5.92  
## available\_vehicles military\_service disease  
## 6 0 no Alzheimer's disease  
## 13 0 no hypertension  
## 23 0 no HIV/AIDS  
## 52 0 no HIV/AIDS  
## 74 0 yes kidney disease  
## 76 0 no Alzheimer's disease  
## 80 0 no schizophrenia  
## 82 0 yes multiple sclerosis  
## 90 0 no Alzheimer's disease  
## 98 0 no heart disease  
## 101 0 no Alzheimer's disease  
## 107 0 no HIV/AIDS  
## 108 0 no kidney disease  
## 110 0 no kidney disease  
## 114 0 no kidney disease  
## 116 0 no prostate cancer  
## 117 0 no Alzheimer's disease  
## 127 0 no hypertension  
## 131 0 no hypertension  
## 144 0 no Alzheimer's disease  
## 146 0 no multiple sclerosis  
## 156 0 no diabetes  
## 157 0 no multiple sclerosis  
## 159 0 no kidney disease  
## 173 0 yes Alzheimer's disease  
## 175 0 no hypertension  
## 176 0 no prostate cancer  
## 177 0 no kidney disease  
## 201 0 no gastritis  
## 214 0 no endometriosis  
## 233 0 no endometriosis  
## 237 0 no Alzheimer's disease  
## 240 0 no heart disease  
## 261 0 no kidney disease  
## 281 0 no prostate cancer  
## 283 0 no heart disease  
## 284 0 no Alzheimer's disease  
## 289 0 yes hypertension  
## 295 0 no prostate cancer  
## 325 0 yes Alzheimer's disease  
## 326 0 no schizophrenia  
## 335 0 no breast cancer  
## 338 0 no hypertension  
## 341 0 no heart disease  
## 349 0 yes Alzheimer's disease  
## 356 0 no skin cancer  
## 369 0 yes hypertension  
## 373 0 no diabetes  
## 379 0 no hypertension  
## 384 0 yes hypertension  
## 404 0 no endometriosis  
## 409 0 no prostate cancer  
## 413 0 no hypertension  
## 430 0 no kidney disease  
## 435 0 no HIV/AIDS  
## 438 0 no kidney disease  
## 453 0 yes skin cancer  
## 463 0 no HIV/AIDS  
## 470 0 no prostate cancer  
## 475 0 yes HIV/AIDS  
## 478 0 no breast cancer  
## 482 0 no Alzheimer's disease  
## 483 0 no breast cancer  
## 495 0 yes prostate cancer  
## 502 0 no hypertension  
## 503 0 no skin cancer  
## 504 0 yes Alzheimer's disease  
## 544 0 no skin cancer  
## 556 0 no diabetes  
## 595 0 no breast cancer  
## 597 0 no prostate cancer  
## 598 0 no diabetes  
## 602 0 no hypertension  
## 604 0 no breast cancer  
## 608 0 no heart disease  
## 615 0 no Alzheimer's disease  
## 618 0 no breast cancer  
## 636 0 yes kidney disease  
## 640 0 no Alzheimer's disease  
## 643 0 no hypertension  
## 660 0 no gastritis  
## 669 0 no kidney disease  
## 691 0 no prostate cancer  
## 699 0 no skin cancer  
## 704 0 no diabetes  
## 709 0 yes Alzheimer's disease  
## 713 0 no hypertension  
## 720 0 no Alzheimer's disease  
## 732 0 no prostate cancer  
## 735 0 no hypertension  
## 742 0 no multiple sclerosis  
## 763 0 no multiple sclerosis  
## 779 0 no gastritis  
## 780 0 no schizophrenia  
## 786 0 no kidney disease  
## 811 0 no skin cancer  
## 835 0 no HIV/AIDS  
## 848 0 no Alzheimer's disease  
## 893 0 no Alzheimer's disease  
## 896 0 no multiple sclerosis  
## 902 0 no hypertension  
## 915 0 no skin cancer  
## 916 0 yes Alzheimer's disease  
## 937 0 no kidney disease  
## 940 0 no hypertension  
## 944 0 no skin cancer  
## 962 0 no multiple sclerosis  
## 964 0 no Alzheimer's disease  
## 966 0 no hypertension  
## 976 0 no hypertension  
## 977 0 no diabetes  
## 999 0 no heart disease  
## 1004 0 no hypertension  
## 1010 0 no hypertension  
## 1011 0 no Alzheimer's disease  
## 1024 0 no Alzheimer's disease  
## 1028 0 no kidney disease  
## 1029 0 no skin cancer  
## 1032 0 yes prostate cancer  
## 1037 0 no multiple sclerosis  
## 1040 0 no hypertension  
## 1044 0 no multiple sclerosis  
## 1050 0 no Alzheimer's disease  
## 1057 0 no endometriosis  
## 1060 0 yes HIV/AIDS  
## 1066 0 no hypertension  
## 1072 0 no endometriosis  
## 1073 0 no prostate cancer  
## 1079 0 no prostate cancer  
## 1081 0 no diabetes  
## 1095 0 no hypertension  
## 1096 0 no Alzheimer's disease  
## 1112 0 no breast cancer  
## 1114 0 no multiple sclerosis  
## 1124 0 no hypertension  
## 1130 0 yes Alzheimer's disease  
## 1132 0 no prostate cancer  
## 1134 0 no Alzheimer's disease  
## 1135 0 no schizophrenia  
## 1137 0 no hypertension  
## 1165 0 no skin cancer  
## 1183 0 no skin cancer  
## 1191 0 no prostate cancer  
## 1196 0 no breast cancer  
## 1197 0 no gastritis  
## 1204 0 no skin cancer  
## 1206 0 no gastritis  
## 1212 0 no kidney disease  
## 1220 0 no Alzheimer's disease  
## 1224 0 no breast cancer  
## 1225 0 no Alzheimer's disease  
## 1230 0 no gastritis  
## 1231 0 no schizophrenia  
## 1244 0 no hypertension  
## 1248 0 no hypertension  
## 1259 0 no breast cancer  
## 1266 0 no schizophrenia  
## 1270 0 no Alzheimer's disease  
## 1273 0 no Alzheimer's disease  
## 1276 0 no Alzheimer's disease  
## 1293 0 no gastritis  
## 1296 0 no diabetes  
## 1300 0 no hypertension  
## 1306 0 no endometriosis  
## 1322 0 no heart disease  
## 1333 0 no Alzheimer's disease  
## 1349 0 no gastritis  
## 1370 0 no HIV/AIDS  
## 1371 0 no heart disease  
## 1372 0 no breast cancer  
## 1374 0 no breast cancer  
## 1377 0 no kidney disease  
## 1387 0 yes gastritis  
## 1402 0 no hypertension  
## 1409 0 no heart disease  
## 1411 0 yes kidney disease  
## 1412 0 yes prostate cancer  
## 1414 0 no endometriosis  
## 1422 0 no Alzheimer's disease  
## 1429 0 no prostate cancer  
## 1435 0 no heart disease  
## 1496 0 no hypertension  
## 1497 0 no prostate cancer  
## 1500 0 no gastritis  
## 1502 0 no endometriosis  
## 1504 0 no HIV/AIDS  
## 1505 0 no skin cancer  
## 1518 0 no multiple sclerosis  
## 1552 0 no skin cancer  
## 1563 0 no Alzheimer's disease  
## 1564 0 no hypertension  
## 1565 0 no skin cancer  
## 1569 0 no heart disease  
## 1573 0 no gastritis  
## 1575 0 no kidney disease  
## 1581 0 no hypertension  
## 1585 0 yes prostate cancer  
## 1590 0 no prostate cancer  
## 1597 0 no Alzheimer's disease  
## 1598 0 no breast cancer  
## 1600 0 no gastritis  
## 1604 0 no prostate cancer  
## 1606 0 yes kidney disease  
## 1611 0 no breast cancer  
## 1615 0 no skin cancer  
## 1626 0 no hypertension  
## 1635 0 no hypertension  
## 1638 0 no schizophrenia  
## 1646 0 no Alzheimer's disease  
## 1650 0 no kidney disease  
## 1660 0 no Alzheimer's disease  
## 1669 0 no breast cancer  
## 1677 0 no heart disease  
## 1683 0 no kidney disease  
## 1700 0 no prostate cancer  
## 1710 0 no skin cancer  
## 1722 0 no kidney disease  
## 1739 0 no kidney disease  
## 1742 0 yes diabetes  
## 1750 0 no kidney disease  
## 1756 0 no diabetes  
## 1760 0 no kidney disease  
## 1768 0 no skin cancer  
## 1776 0 no breast cancer  
## 1787 0 no multiple sclerosis  
## 1791 0 no HIV/AIDS  
## 1795 0 no Alzheimer's disease  
## 1797 0 no diabetes  
## 1806 0 no HIV/AIDS  
## 1810 0 no prostate cancer  
## 1815 0 yes schizophrenia  
## 1823 0 no diabetes  
## 1824 0 no HIV/AIDS  
## 1841 0 no kidney disease  
## 1859 0 no heart disease  
## 1862 0 no kidney disease  
## 1869 0 no hypertension  
## 1873 0 no hypertension  
## 1887 0 no Alzheimer's disease  
## 1898 0 no endometriosis  
## 1911 0 no prostate cancer  
## 1913 0 no Alzheimer's disease  
## 1916 0 no skin cancer  
## 1918 0 no diabetes  
## 1921 0 no prostate cancer  
## 1929 0 no breast cancer  
## 1932 0 no HIV/AIDS  
## 1935 0 no kidney disease  
## 1938 0 no prostate cancer  
## 1940 0 no skin cancer  
## 1970 0 no Alzheimer's disease  
## 1972 0 no breast cancer  
## 1993 0 no diabetes  
## 1994 0 no kidney disease  
## 1997 0 no gastritis

patients[ which(patients$gender == 'female' & patients$disease == 'breast cancer'), ]

## id gender dob zipcode employment\_status education  
## 8 Addison Combs female 1963-01-17 43221 employed bachelors  
## 22 Ivy Freeman female 1960-06-25 89127 employed masters  
## 25 Lily Williams female 1977-05-05 10001 employed bachelors  
## 33 Eliana Stephenson female 1942-09-15 44101 retired highschool  
## 34 Valeria Vaughan female 1949-05-31 90015 retired phd/md  
## 55 Abigail Carey female 1958-05-27 43210 employed bachelors  
## 93 Isla Smith female 1958-01-07 43221 employed bachelors  
## 100 Mila Morgan female 1943-07-12 89136 retired phd/md  
## 124 Sophia Parsons female 1981-10-28 43210 employed bachelors  
## 126 Mckenzie Gamble female 1932-10-14 94110 retired bachelors  
## 140 Amelia Thompson female 1970-10-12 90008 employed bachelors  
## 154 Ava Roth female 1979-08-21 60612 employed highschool  
## 165 Natalie Chavez female 1968-03-09 44101 employed bachelors  
## 178 Jasmine Wolf female 1934-07-09 60605 retired bachelors  
## 193 Madison English female 1950-09-12 44101 retired bachelors  
## 197 Eliana Walker female 1945-05-13 10026 retired bachelors  
## 209 Isabella Davis female 1954-07-09 43221 retired bachelors  
## 218 Emma Jones female 1924-03-28 60605 retired bachelors  
## 219 Ariana Smith female 1955-05-21 60605 employed bachelors  
## 223 Summer Peterson female 1954-03-31 10001 employed highschool  
## 234 Isabella Nunez female 1948-04-05 44101 retired masters  
## 241 Katelyn Williams female 1935-08-22 10026 retired bachelors  
## 280 Ella Burton female 1931-07-01 89127 retired highschool  
## 294 Jocelyn Bradley female 1949-12-11 90008 retired masters  
## 296 Arianna Wolfe female 1957-03-10 10001 employed masters  
## 316 Mia Cowan female 1988-10-21 10001 employed bachelors  
## 317 Valentina Becker female 1958-06-16 10001 retired bachelors  
## 335 Amelia Gardner female 1947-06-28 60605 unemployed highschool  
## 376 Isabella Kirk female 1959-12-24 89127 employed bachelors  
## 383 Eliana Kelley female 1940-09-26 94105 retired bachelors  
## 387 Charlotte Johnson female 1957-09-15 10026 employed highschool  
## 390 Payton Shaw female 1938-01-03 90008 retired bachelors  
## 406 Paisley Wood female 1974-04-10 94105 employed bachelors  
## 425 Lucy Roberts female 1932-05-18 89127 retired bachelors  
## 429 Violet Walton female 1943-08-09 94110 retired bachelors  
## 448 Samantha Huang female 1949-06-07 44101 retired highschool  
## 456 Sophia Campbell female 1954-12-20 44101 employed highschool  
## 478 Abigail Ryan female 1957-11-17 89136 unemployed highschool  
## 483 Daisy Fields female 1951-06-19 60605 unemployed highschool  
## 520 Adalyn Atkinson female 1946-03-07 94105 retired bachelors  
## 525 Stephanie Boyd female 1930-01-15 94110 retired bachelors  
## 580 Angelina Hogan female 1979-10-19 90008 employed bachelors  
## 585 Sofia Jones female 1945-05-02 90008 retired bachelors  
## 595 Alyssa Arnold female 1951-07-31 89127 unemployed bachelors  
## 604 Leah Shaffer female 1953-03-09 60605 unemployed highschool  
## 616 Aurora Elliott female 1941-06-20 43210 retired bachelors  
## 618 Aubrey Macias female 1954-01-25 89127 unemployed highschool  
## 675 Kimberly Miller female 1933-02-09 90008 retired bachelors  
## 683 Jocelyn Taylor female 1983-07-03 94110 employed bachelors  
## 696 Eva Newman female 1950-12-22 89127 retired bachelors  
## 708 Caroline Bell female 1935-12-05 60612 retired bachelors  
## 712 Michelle Jimenez female 1964-07-12 90008 employed masters  
## 717 Harper Cole female 1974-06-04 94110 employed bachelors  
## 730 Brielle Mason female 1967-08-10 43221 employed highschool  
## 758 Penelope Mack female 1930-10-26 60605 retired highschool  
## 760 Lucy Jones female 1959-06-23 89136 employed masters  
## 781 Madison Hernandez female 1942-07-16 60612 retired bachelors  
## 785 Bella Sanchez female 1951-07-28 94110 employed phd/md  
## 794 Zoey Frazier female 1967-11-23 94110 employed bachelors  
## 796 Hannah Morales female 1976-12-10 89127 employed bachelors  
## 805 Ella Townsend female 1977-10-22 60612 employed masters  
## 807 Aubrey Parsons female 1965-10-16 10001 retired masters  
## 870 Lily Gomez female 1941-07-16 44101 retired bachelors  
## 880 Makayla Morrison female 1969-08-31 90015 employed bachelors  
## 886 Sofia Smith female 1933-03-19 90008 retired phd/md  
## 897 Nevaeh Werner female 1959-03-21 60612 employed bachelors  
## 904 Paige Williams female 1941-01-11 43210 retired highschool  
## 913 Leah Martin female 1960-06-07 94105 employed masters  
## 919 Rebecca Stewart female 1937-10-16 90008 retired bachelors  
## 930 Gianna Martin female 1939-06-14 94105 retired highschool  
## 946 Anna Hall female 1968-01-23 89127 employed phd/md  
## 950 Sarah Payne female 1931-11-17 10001 retired bachelors  
## 956 Isabella Bennett female 1929-08-23 89136 retired bachelors  
## 959 Harper Madden female 1937-07-06 60605 retired bachelors  
## 969 Abigail Brown female 1930-07-02 94110 retired bachelors  
## 970 Kylie Russell female 1933-06-14 43221 retired bachelors  
## 1013 Valeria Gibson female 1972-11-12 90008 employed bachelors  
## 1014 Alaina Hunt female 1963-03-10 94105 employed bachelors  
## 1022 Stella Deleon female 1959-01-31 60605 employed masters  
## 1069 Isabella Rogers female 1936-08-12 90008 retired masters  
## 1089 Sophie Bell female 1970-03-28 94105 employed bachelors  
## 1092 Faith Davis female 1935-02-06 10026 retired masters  
## 1112 Aliyah Jones female 1954-09-21 44101 unemployed highschool  
## 1120 Layla Rojas female 1973-07-08 89136 employed masters  
## 1128 Sophia Johnson female 1969-04-03 10001 employed masters  
## 1157 Serenity Sanders female 1947-12-05 90015 retired bachelors  
## 1159 Gabriella Johnson female 1935-04-20 90015 retired bachelors  
## 1179 Penelope Pitts female 1938-08-05 90008 retired bachelors  
## 1196 Aliyah Vega female 1961-06-28 89127 unemployed highschool  
## 1207 Savannah Harmon female 1957-05-30 89127 employed masters  
## 1210 Audrey Oconnell female 1975-10-26 90015 employed bachelors  
## 1224 Annabelle Kelly female 1960-09-10 60605 unemployed bachelors  
## 1249 Isabel Morris female 1939-03-06 89136 retired bachelors  
## 1252 Kylee Johnson female 1934-03-29 94110 retired masters  
## 1259 Skylar Lewis female 1971-11-08 10001 unemployed highschool  
## 1260 Elise Ross female 1963-01-14 60605 employed bachelors  
## 1265 Isabella Williams female 1964-04-11 10001 employed bachelors  
## 1320 Sophia Guerra female 1943-12-06 44101 retired highschool  
## 1330 Elizabeth Ferguson female 1964-02-05 89136 employed bachelors  
## 1331 Sophie Ward female 1951-07-25 10001 retired masters  
## 1341 Melanie Nguyen female 1934-12-06 10001 retired bachelors  
## 1361 Addison Jones female 1940-07-27 43210 retired bachelors  
## 1372 Abigail Rodriguez female 1968-02-03 94110 unemployed highschool  
## 1374 Lyla Rodriguez female 1955-11-10 60605 unemployed highschool  
## 1380 Avery Lambert female 1951-05-10 10026 retired bachelors  
## 1410 Eva Moore female 1955-03-22 89136 employed masters  
## 1458 Molly Perry female 1941-04-15 43210 retired bachelors  
## 1465 Alana Weaver female 1961-12-21 60612 employed highschool  
## 1467 Lauren Hayes female 1943-08-27 94105 retired bachelors  
## 1474 Lyla Rogers female 1947-07-10 94105 retired phd/md  
## 1479 Isabella Cox female 1962-11-06 10026 retired masters  
## 1524 Hailey King female 1962-12-24 10001 retired bachelors  
## 1541 Emma Richardson female 1938-11-25 94110 retired masters  
## 1555 Adalynn Waters female 1954-08-04 10001 student phd/md  
## 1568 Vivian Harris female 1967-04-16 94110 employed bachelors  
## 1598 Mia Wilson female 1971-08-15 43210 unemployed highschool  
## 1611 Lillian Brock female 1963-07-07 60612 unemployed highschool  
## 1618 Kaylee Bennett female 1938-11-19 43210 retired bachelors  
## 1649 Emery Manning female 1962-11-09 10026 employed masters  
## 1655 Lilly Davis female 1959-11-07 60605 employed bachelors  
## 1669 Emily Johnson female 1958-09-08 94105 unemployed highschool  
## 1671 Katherine Hunt female 1970-04-29 90015 employed bachelors  
## 1709 Elizabeth Sanchez female 1960-06-13 94110 employed bachelors  
## 1724 Brianna Brown female 1942-08-18 10001 retired bachelors  
## 1753 Harmony Torres female 1935-12-06 89136 retired masters  
## 1759 Kinsley Lewis female 1933-02-21 43221 retired bachelors  
## 1776 Elise Ramirez female 1958-06-03 60605 unemployed highschool  
## 1781 Sophia Fritz female 1959-07-06 89127 employed bachelors  
## 1796 Liliana Meyer female 1954-12-30 94105 employed bachelors  
## 1808 Josephine Galvan female 1955-12-05 43210 employed bachelors  
## 1830 Amelia Warren female 1982-09-03 60605 employed bachelors  
## 1845 Amy Carpenter female 1975-12-08 60612 employed bachelors  
## 1849 Catherine Mullins female 1936-11-01 43210 retired bachelors  
## 1864 Autumn Arnold female 1953-09-29 94110 retired bachelors  
## 1867 Autumn Stevenson female 1956-11-27 94105 retired masters  
## 1874 Payton Gonzales female 1952-01-29 90008 employed bachelors  
## 1879 Sophia Wilson female 1958-01-26 60605 employed bachelors  
## 1886 Gianna Gilbert female 1936-11-04 10026 retired phd/md  
## 1902 Avery Rice female 1983-07-28 94110 employed bachelors  
## 1929 Sophia Ortiz female 1955-09-02 94105 unemployed highschool  
## 1946 Liliana Neal female 1982-11-11 60612 employed masters  
## 1953 Layla Crane female 1926-01-27 60612 retired bachelors  
## 1956 Abigail Rodgers female 1969-11-26 89127 employed bachelors  
## 1972 Sophia Hopkins female 1959-01-30 94105 unemployed highschool  
## 1973 Ruby Rosario female 1961-06-15 43221 employed masters  
## marital\_status children ancestry avg\_commute daily\_internet\_use  
## 8 married 2 c\_europe 20.47 3.01  
## 22 married 2 c\_europe 43.37 5.65  
## 25 married 2 w\_europe 21.94 7.20  
## 33 married 2 w\_europe 50.57 5.14  
## 34 married 2 e\_europe 40.66 7.02  
## 55 married 2 n\_europe 23.30 4.91  
## 93 married 2 w\_europe 24.28 5.22  
## 100 married 3 e\_europe 29.97 7.66  
## 124 single 0 w\_europe 17.15 5.08  
## 126 married 2 w\_europe 39.66 8.07  
## 140 married 4 w\_europe 38.42 6.22  
## 154 married 7 e\_europe 33.60 5.14  
## 165 married 0 w\_europe 42.95 4.87  
## 178 married 2 c\_europe 24.68 4.43  
## 193 single 2 w\_europe 38.87 4.53  
## 197 married 1 e\_europe 11.42 5.68  
## 209 single 1 w\_europe 30.20 5.62  
## 218 single 3 e\_europe 36.27 4.93  
## 219 married 1 c\_europe 27.31 5.94  
## 223 married 2 e\_europe 37.64 4.53  
## 234 married 1 e\_europe 36.69 4.72  
## 241 married 1 w\_europe 19.74 7.53  
## 280 married 2 w\_europe 10.00 6.74  
## 294 married 3 w\_europe 38.27 5.02  
## 296 married 2 w\_europe 33.49 5.25  
## 316 single 0 w\_europe 37.54 7.55  
## 317 married 5 w\_europe 21.78 4.48  
## 335 married 3 e\_europe 41.32 5.80  
## 376 married 3 e\_europe 48.00 4.59  
## 383 married 2 w\_europe 22.95 5.05  
## 387 married 2 w\_europe 40.63 4.69  
## 390 married 5 w\_europe 40.37 5.39  
## 406 married 2 n\_europe 26.32 5.97  
## 425 married 1 e\_europe 39.72 5.70  
## 429 married 2 w\_europe 20.92 6.43  
## 448 married 4 n\_europe 44.27 6.75  
## 456 married 3 w\_europe 20.33 5.56  
## 478 married 0 w\_europe 33.81 4.73  
## 483 married 2 w\_europe 43.98 6.41  
## 520 single 3 c\_europe 22.72 4.53  
## 525 married 1 w\_europe 18.50 7.02  
## 580 single 1 c\_europe 47.22 4.32  
## 585 married 7 w\_europe 37.55 3.78  
## 595 single 2 n\_europe 54.10 2.40  
## 604 married 2 e\_europe 29.45 5.48  
## 616 married 4 e\_europe 23.54 4.91  
## 618 single 1 c\_europe 52.58 5.99  
## 675 single 1 w\_europe 16.53 4.98  
## 683 married 1 e\_europe 39.62 4.55  
## 696 married 2 n\_europe 15.35 5.73  
## 708 married 1 e\_europe 34.57 7.64  
## 712 married 1 w\_europe 39.09 2.23  
## 717 married 5 c\_europe 33.35 3.72  
## 730 married 2 w\_europe 43.94 4.04  
## 758 single 0 w\_europe 30.65 4.85  
## 760 married 2 n\_europe 36.41 1.79  
## 781 married 3 e\_europe 20.01 3.89  
## 785 married 1 c\_europe 28.82 4.12  
## 794 married 5 w\_europe 30.92 4.27  
## 796 married 3 w\_europe 23.71 4.92  
## 805 married 6 e\_europe 28.80 4.31  
## 807 married 2 e\_europe 31.00 6.05  
## 870 single 1 w\_europe 29.08 4.87  
## 880 married 3 w\_europe 16.51 4.48  
## 886 married 0 w\_europe 40.32 3.75  
## 897 single 4 e\_europe 15.04 3.11  
## 904 married 0 w\_europe 22.28 5.83  
## 913 married 5 c\_europe 37.69 5.15  
## 919 married 5 e\_europe 21.94 3.31  
## 930 married 2 w\_europe 39.97 3.52  
## 946 married 2 w\_europe 32.55 4.75  
## 950 married 3 w\_europe 27.59 5.77  
## 956 married 3 w\_europe 15.46 6.74  
## 959 married 0 n\_europe 41.51 4.19  
## 969 married 0 w\_europe 35.61 3.35  
## 970 married 2 w\_europe 40.59 2.30  
## 1013 married 1 c\_europe 38.89 4.44  
## 1014 married 1 w\_europe 25.52 7.57  
## 1022 single 3 e\_europe 18.80 6.02  
## 1069 married 5 w\_europe 39.75 6.43  
## 1089 single 2 e\_europe 31.16 4.52  
## 1092 married 1 n\_europe 23.92 3.62  
## 1112 married 3 w\_europe 30.21 3.72  
## 1120 married 3 e\_europe 40.24 4.25  
## 1128 married 2 w\_europe 25.94 5.91  
## 1157 single 1 e\_europe 28.10 6.72  
## 1159 married 2 n\_europe 48.73 3.23  
## 1179 married 1 e\_europe 25.16 3.98  
## 1196 single 1 w\_europe 32.60 3.83  
## 1207 married 5 w\_europe 18.21 3.17  
## 1210 married 2 w\_europe 15.23 4.41  
## 1224 single 2 w\_europe 38.40 5.78  
## 1249 married 1 n\_europe 11.91 4.52  
## 1252 married 5 w\_europe 50.85 2.35  
## 1259 married 2 w\_europe 32.74 4.46  
## 1260 married 3 e\_europe 28.16 3.23  
## 1265 married 4 w\_europe 20.73 7.04  
## 1320 married 1 e\_europe 50.96 6.08  
## 1330 married 2 c\_europe 39.05 5.84  
## 1331 married 2 c\_europe 14.57 5.43  
## 1341 married 7 w\_europe 16.49 4.84  
## 1361 married 1 n\_europe 44.60 7.04  
## 1372 single 3 n\_europe 40.64 5.24  
## 1374 married 3 c\_europe 24.80 5.18  
## 1380 single 2 w\_europe 40.66 3.95  
## 1410 married 7 w\_europe 47.49 5.60  
## 1458 married 3 w\_europe 22.02 5.87  
## 1465 married 1 e\_europe 42.84 6.47  
## 1467 married 1 c\_europe 14.35 6.47  
## 1474 single 7 w\_europe 29.47 5.14  
## 1479 single 1 e\_europe 14.19 6.33  
## 1524 married 2 w\_europe 16.77 4.68  
## 1541 married 2 e\_europe 16.75 6.38  
## 1555 married 3 e\_europe 40.68 4.30  
## 1568 married 3 w\_europe 46.61 2.44  
## 1598 married 6 n\_europe 17.46 2.74  
## 1611 single 5 e\_europe 23.72 4.18  
## 1618 married 2 w\_europe 16.49 6.44  
## 1649 married 2 e\_europe 49.25 4.52  
## 1655 married 0 w\_europe 27.75 6.81  
## 1669 single 1 w\_europe 35.89 5.08  
## 1671 married 1 e\_europe 35.67 3.99  
## 1709 married 1 w\_europe 44.07 4.52  
## 1724 married 4 w\_europe 25.92 6.02  
## 1753 married 3 w\_europe 5.71 3.56  
## 1759 married 4 e\_europe 30.99 3.73  
## 1776 married 4 w\_europe 35.66 4.91  
## 1781 married 3 w\_europe 19.03 8.64  
## 1796 married 1 e\_europe 32.82 4.18  
## 1808 single 1 n\_europe 41.97 5.62  
## 1830 single 0 w\_europe 32.59 6.70  
## 1845 married 3 e\_europe 32.98 6.31  
## 1849 single 2 w\_europe 22.03 4.54  
## 1864 single 0 c\_europe 45.76 4.93  
## 1867 married 1 e\_europe 34.07 6.18  
## 1874 married 1 e\_europe 33.17 3.71  
## 1879 married 1 w\_europe 37.37 3.48  
## 1886 married 3 w\_europe 27.00 5.74  
## 1902 married 4 w\_europe 35.05 4.01  
## 1929 single 3 e\_europe 18.14 5.98  
## 1946 married 5 w\_europe 33.81 4.95  
## 1953 married 1 w\_europe 32.30 5.01  
## 1956 married 1 w\_europe 32.13 3.81  
## 1972 married 0 e\_europe 34.71 4.98  
## 1973 single 2 w\_europe 34.87 3.50  
## available\_vehicles military\_service disease  
## 8 3 no breast cancer  
## 22 1 no breast cancer  
## 25 2 no breast cancer  
## 33 1 no breast cancer  
## 34 2 no breast cancer  
## 55 3 no breast cancer  
## 93 1 no breast cancer  
## 100 1 no breast cancer  
## 124 3 no breast cancer  
## 126 1 no breast cancer  
## 140 2 no breast cancer  
## 154 2 no breast cancer  
## 165 3 no breast cancer  
## 178 2 no breast cancer  
## 193 2 no breast cancer  
## 197 1 no breast cancer  
## 209 0 no breast cancer  
## 218 2 no breast cancer  
## 219 4 no breast cancer  
## 223 2 no breast cancer  
## 234 2 no breast cancer  
## 241 3 no breast cancer  
## 280 2 no breast cancer  
## 294 1 no breast cancer  
## 296 2 no breast cancer  
## 316 2 no breast cancer  
## 317 4 no breast cancer  
## 335 0 no breast cancer  
## 376 0 no breast cancer  
## 383 0 no breast cancer  
## 387 1 no breast cancer  
## 390 3 no breast cancer  
## 406 0 no breast cancer  
## 425 3 no breast cancer  
## 429 3 no breast cancer  
## 448 1 no breast cancer  
## 456 1 no breast cancer  
## 478 0 no breast cancer  
## 483 0 no breast cancer  
## 520 1 no breast cancer  
## 525 1 no breast cancer  
## 580 1 yes breast cancer  
## 585 3 no breast cancer  
## 595 0 no breast cancer  
## 604 0 no breast cancer  
## 616 2 no breast cancer  
## 618 0 no breast cancer  
## 675 3 no breast cancer  
## 683 2 no breast cancer  
## 696 2 no breast cancer  
## 708 2 no breast cancer  
## 712 4 no breast cancer  
## 717 1 no breast cancer  
## 730 1 no breast cancer  
## 758 2 no breast cancer  
## 760 2 no breast cancer  
## 781 3 no breast cancer  
## 785 3 no breast cancer  
## 794 3 no breast cancer  
## 796 1 no breast cancer  
## 805 3 no breast cancer  
## 807 1 no breast cancer  
## 870 2 no breast cancer  
## 880 1 no breast cancer  
## 886 1 no breast cancer  
## 897 1 no breast cancer  
## 904 2 no breast cancer  
## 913 2 no breast cancer  
## 919 2 no breast cancer  
## 930 3 no breast cancer  
## 946 2 no breast cancer  
## 950 2 no breast cancer  
## 956 1 no breast cancer  
## 959 2 no breast cancer  
## 969 4 no breast cancer  
## 970 0 no breast cancer  
## 1013 2 no breast cancer  
## 1014 3 no breast cancer  
## 1022 2 no breast cancer  
## 1069 2 no breast cancer  
## 1089 3 no breast cancer  
## 1092 2 no breast cancer  
## 1112 0 no breast cancer  
## 1120 4 no breast cancer  
## 1128 2 no breast cancer  
## 1157 3 no breast cancer  
## 1159 2 no breast cancer  
## 1179 3 no breast cancer  
## 1196 0 no breast cancer  
## 1207 2 no breast cancer  
## 1210 2 no breast cancer  
## 1224 0 no breast cancer  
## 1249 2 no breast cancer  
## 1252 2 no breast cancer  
## 1259 0 no breast cancer  
## 1260 3 no breast cancer  
## 1265 1 no breast cancer  
## 1320 2 no breast cancer  
## 1330 3 no breast cancer  
## 1331 2 no breast cancer  
## 1341 1 no breast cancer  
## 1361 2 no breast cancer  
## 1372 0 no breast cancer  
## 1374 0 no breast cancer  
## 1380 2 no breast cancer  
## 1410 3 no breast cancer  
## 1458 2 no breast cancer  
## 1465 2 no breast cancer  
## 1467 3 no breast cancer  
## 1474 2 no breast cancer  
## 1479 1 no breast cancer  
## 1524 2 no breast cancer  
## 1541 2 no breast cancer  
## 1555 2 no breast cancer  
## 1568 3 no breast cancer  
## 1598 0 no breast cancer  
## 1611 0 no breast cancer  
## 1618 0 no breast cancer  
## 1649 3 no breast cancer  
## 1655 3 no breast cancer  
## 1669 0 no breast cancer  
## 1671 2 no breast cancer  
## 1709 1 no breast cancer  
## 1724 2 no breast cancer  
## 1753 1 no breast cancer  
## 1759 4 no breast cancer  
## 1776 0 no breast cancer  
## 1781 3 no breast cancer  
## 1796 3 no breast cancer  
## 1808 2 no breast cancer  
## 1830 2 no breast cancer  
## 1845 2 no breast cancer  
## 1849 2 no breast cancer  
## 1864 1 no breast cancer  
## 1867 2 no breast cancer  
## 1874 2 no breast cancer  
## 1879 1 no breast cancer  
## 1886 2 no breast cancer  
## 1902 3 no breast cancer  
## 1929 0 no breast cancer  
## 1946 2 no breast cancer  
## 1953 1 no breast cancer  
## 1956 3 no breast cancer  
## 1972 0 no breast cancer  
## 1973 3 no breast cancer

# split the diseases into columns with binary values  
patients$hypertension <- binary\_value(patients$disease,'hypertension')  
patients$endometriosis <- binary\_value(patients$disease,'endometriosis')  
patients$multiple\_sclerosis <- binary\_value(patients$disease,'multiple sclerosis')  
patients$schizophrenia <- binary\_value(patients$disease,'schizophrenia')  
patients$kidney\_disease <- binary\_value(patients$disease,'kidney disease')  
patients$gastritis <- binary\_value(patients$disease,'gastritis')  
patients$alzheimer <- binary\_value(patients$disease,'Alzheimer disease')  
patients$cancers <- binary\_value(patients$disease,'cancer')  
patients$diabetes <- binary\_value(patients$disease,'diabetes')  
patients$heart\_disease <- binary\_value(patients$disease,'heart disease')

patients$gender <- ifelse(patients$gender == 'male', 0, 1)  
patients$marital\_status <- ifelse(patients$marital\_status == 'single', 0, 1)  
patients$military\_service <- ifelse(patients$military\_service == 'no', 0, 1)  
  
str(patients)

## 'data.frame': 2000 obs. of 24 variables:  
## $ id : chr "Amelia Nixon" "Clara Hicks" "Mason Brown" "Michael Rice" ...  
## $ gender : num 1 1 0 0 1 1 1 1 1 0 ...  
## $ dob : chr "1944-03-09" "1966-07-02" "1981-05-31" "1945-02-13" ...  
## $ zipcode : int 89136 94105 89127 44101 89136 94105 60612 43221 89127 43210 ...  
## $ employment\_status : chr "retired" "employed" "employed" "retired" ...  
## $ education : Factor w/ 4 levels "bachelors","highschool",..: 1 4 3 1 3 2 4 1 3 2 ...  
## $ marital\_status : num 1 1 1 1 1 1 1 1 0 1 ...  
## $ children : int 1 4 2 2 3 2 0 2 2 7 ...  
## $ ancestry : Factor w/ 4 levels "c\_europe","e\_europe",..: 4 3 4 3 4 4 2 1 4 2 ...  
## $ avg\_commute : num 13.4 15.2 23.6 19.6 36.5 ...  
## $ daily\_internet\_use: num 2.53 6.77 3.63 5 7.75 3.34 6.75 3.01 4.12 3.15 ...  
## $ available\_vehicles: int 2 2 1 3 1 0 2 3 1 1 ...  
## $ military\_service : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ disease : chr "hypertension" "endometriosis" "prostate cancer" "multiple sclerosis" ...  
## $ hypertension : num 1 0 0 0 0 0 0 0 0 0 ...  
## $ endometriosis : num 0 1 0 0 0 0 0 0 0 0 ...  
## $ multiple\_sclerosis: num 0 0 0 1 0 0 0 0 0 0 ...  
## $ schizophrenia : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ kidney\_disease : num 0 0 0 0 0 0 1 0 0 0 ...  
## $ gastritis : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ alzheimer : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ cancers : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ diabetes : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ heart\_disease : num 0 0 0 0 0 0 0 0 0 0 ...

write.table(patients,file="C:/Users/mavul/OneDrive/Desktop/Health\_care\_data.csv", sep=",")

#Bar plotting diseases   
counts <- table(patients$disease)  
  
par(las=2) # make label text perpendicular to axis  
par(mar=c(5,8,4,2)) # increase y-axis margin.  
  
barplot(sort(counts, decreasing = TRUE), main="Disease Names",   
 xlab="Diseases Frequency",   
 col="red",  
 horiz=TRUE,  
 cex.names=0.8,  
 xlim = c(0, 350))

