Topic ideas

STA 210 - Project

Bayes' Harem - Kat Cottrell, David Goh, Ethan Song, Christina Wang

```
# load packages
library(tidyverse)
```

Project idea 1 - Abortion attitudes

Introduction and data

- State the source of the data set.

The dataset was put together by user svmiller on github. The user obtained the data from six waves of the World Values Survey (1982-2011). Link: https://github.com/svmiller/wvs-usa-abortion-attitudes

- Describe when and how it was originally collected (by the original data curator, not necessarily how you found the data)

The data were collected as part of the World Values Survey, which is administered every few years and collects information about people's values and beliefs worldwide. The survey aims to get a nationally representative sample of a minimum of 1200 for most countries, and the data are collected via face-to-face interviews at the respondents' homes. The data included in this set specifically include responses from 6 waves of the survey (administered over the period 1982-2011). The responses included in this set are from people in the United States, and it examines their attitudes towards abortion.

- Describe the observations and the general characteristics being measured in the data

The dataset includes 10,387 observations across all of the survey waves. The main variable that is assessed in this survey is abortion attitudes on a scale of 1-10 (1 being abortion is never justified, 10 being abortion is always justified). The survey also measures some demographic characteristics of the respondents (gender, education, age, etc.) and their attitudes on other

issues (political ideology, their respect for authority, how important religion is to their lives, etc.)

Research question

- Describe a research question you're interested in answering using this data.

How is an individual's attitude towards abortion influenced by their demographic characteristics (such as age, gender, and education level) and personal attitudes towards other issues (such as political ideology, importance of religion to their lives, and their respect for authority)?

Glimpse of data

- Use the `glimpse` function to provide an overview of the dataset

```
abortion_attitudes_data <- read_csv("data-1/wvs-usa-abortion-attitudes-data.csv")
```

```
Rows: 10387 Columns: 16
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (16): wvsccode, wave, year, aj, age, collegeed, female, unemployed, ideo...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(abortion_attitudes_data)
```

```
Rows: 10,387
Columns: 16
$ wvsccode         <dbl> 840, 840, 840, 840, 840, 840, 840, 840, 840, 840, 840~
$ wave             <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ year             <dbl> 1982, 1982, 1982, 1982, 1982, 1982, 1982, 1982, 1982,~
$ aj               <dbl> 5, 5, NA, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 5, 6, 3, 2, 5~
$ age              <dbl> 40, 43, 18, 18, 22, 21, 37, 45, 30, 72, 22, 47, 56, 5~
$ collegeed        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
$ female           <dbl> 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0,~
$ unemployed       <dbl> 0, 0, 0, 0, NA, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0~
$ ideology         <dbl> 8, NA, 10, NA, NA, 4, 8, 5, NA, 2, NA, 7, 10, 10, 10,~
$ satisfinancial   <dbl> 5, 3, 2, 6, 5, 8, 9, 10, 8, 6, 1, 8, 3, 8, NA, 2, 5, ~
$ postma4          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

```
$ cai              <dbl> 0, -2, 0, -1, 0, 1, -1, -1, 0, -1, 0, -2, 0, -1, -2, ~
$ trustmostpeople  <dbl> 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0,~
$ godimportant     <dbl> 10, 10, 8, 10, 5, 9, 10, 10, 7, 10, 5, 10, 10, 10, 10~
$ respectauthority <dbl> 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ nationalpride    <dbl> 1, NA, 1, 1, 1, 0, 0, NA, NA, 1, NA, 0, 1, 1, 1, 1, 1~
```

Project idea 2 - Nutrition

Introduction and data

- State the source of the data set.

This data comes from the National Health and Nutrition Examination Survey (NHANES), conducted in 2005-2006 by the Inter-university Consortium for Political and Social Research (ICPSR). It was retrieved from an ICPSR host page on the University of Michigan website. Possible variables used: DS218: Questionnaire: Drug Use, DS234: Questionnaire: Prescription Medications, DS229: Questionnaire: Osteoporosis. Link: https://www.icpsr.umich.edu/web/ICPSR/studies/25504/variables

- Describe when and how it was originally collected (by the original data curator, not necessarily how you found the data)

The data were collected from approximately 5,000 interviewees who were first interviewed by trained administrators in the home and then later completed health examinations in mobile examination centers. The target population is average American non-institutionalized civilians. Low-income persons, adolescents 12-19 years of age, persons 60 years of age and older, African Americans, and Mexican Americans were oversampled in this data collection.

- Describe the observations and the general characteristics being measured in the data

The NHANES 2005-2006 measured 7,449 variables pertaining to the health and nutritional status of Americans. However, our research question only assesses injury and drug use variables.

Research question

- Describe a research question you're interested in answering using this data.

Does there exist any positive correlation between injuries and drug use?

Glimpse of data

- Use the `glimpse` function to provide an overview of the dataset

```r
drugs <- read.table(file ='data-2/25504-0218-Data.tsv', sep = '\t', header = TRUE)
osteo <- read.table(file ='data-2/25504-0229-Data.tsv', sep = '\t', header = TRUE)
rx    <- read.table(file ='data-2/25504-0234-Data.tsv', sep = '\t', header = TRUE)
health <- inner_join(drugs, osteo, by=c('SEQN'='SEQN'))
health <- inner_join(health, rx, by=c('SEQN'='SEQN'))

glimpse(health)
```

```
Rows: 5,699
Columns: 286
$ SEQN       <int> 31131, 31131, 31144, 31151, 31151, 31151, 31151, 31151, 311~
$ DUAISC     <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 3, 3, 1,~
$ DUQ200     <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 2, NA, NA, NA,~
$ DUQ210     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 15, 15, 15, 15,~
$ DUQ220Q    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 15, 15, 15, 15,~
$ DUQ220U    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 4, 4, 4, 4, NA,~
$ DUQ230     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ240     <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 2, NA, NA, NA,~
$ DUQ250     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 1, 1, 1, 1, NA,~
$ DUQ260     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 16, 16, 16, 16,~
$ DUQ270Q    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 7, 7, 7, 7, NA,~
$ DUQ270U    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 4, 4, 4, 4, NA,~
$ DUQ272     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 5, 5, 5, 5, NA,~
$ DUQ280     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ290     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, 2, 2, 2, NA,~
$ DUQ300     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ310Q    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ310U    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ320     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ330     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 1, 1, 1, 1, NA,~
$ DUQ340     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 15, 15, 15, 15,~
$ DUQ350Q    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 8, 8, 8, 8, NA,~
$ DUQ350U    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 4, 4, 4, 4, NA,~
$ DUQ352     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 3, 3, 3, 3, NA,~
$ DUQ360     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ370     <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, NA, NA, NA,~
$ DUQ380A    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ380B    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ380C    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ380D    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ380E    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ390     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```
$ DUQ400Q    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ400U    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ410     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ420     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DUQ430     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 1, 1, 1, 1, NA,~
$ SDDSRVYR.x <int> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,~
$ RIDSTATR.x <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ RIDEXMON.x <int> 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2,~
$ RIAGENDR.x <int> 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2,~
$ RIDAGEYR.x <int> 44, 44, 21, 59, 59, 59, 59, 59, 59, 59, 27, 44, 44, 44, 44,~
$ RIDAGEMN.x <int> 535, 535, 255, 711, 711, 711, 711, 711, 711, 711, 329, 528,~
$ RIDAGEEX.x <int> 536, 536, 256, 711, 711, 711, 711, 711, 711, 711, 330, 528,~
$ RIDRETH1.x <int> 4, 4, 2, 4, 4, 4, 4, 4, 4, 4, 1, 5, 5, 5, 5, 5, 4, 4, 4, 3,~
$ DMQMILIT.x <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ DMDBORN.x  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 3, 1, 1, 1, 1,~
$ DMDCITZN.x <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1,~
$ DMDYRSUS.x <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 6, NA, NA, NA, NA, ~
$ DMDEDUC3.x <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DMDEDUC2.x <int> 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 5, 4, 4, 4, 2,~
$ DMDSCHOL.x <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DMDMARTL.x <int> 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5, 5, 1, 4, 4, 4, 1,~
$ DMDHHSIZ.x <int> 4, 4, 6, 2, 2, 2, 2, 2, 2, 2, 5, 2, 2, 2, 2, 4, 7, 7, 7, 5,~
$ DMDFMSIZ.x <int> 4, 4, 6, 2, 2, 2, 2, 2, 2, 2, 5, 2, 2, 2, 2, 4, 7, 7, 7, 5,~
$ INDHHINC.x <int> 11, 11, 3, 7, 7, 7, 7, 7, 7, 7, 7, 3, 3, 3, 3, 10, 7, 7, 7,~
$ INDFMINC.x <int> 11, 11, 3, 7, 7, 7, 7, 7, 7, 7, 7, 3, 3, 3, 3, 10, 7, 7, 7,~
$ INDFMPIR.x <dbl> 4.65, 4.65, 0.46, 3.03, 3.03, 3.03, 3.03, 3.03, 3.03, 3.03,~
$ RIDEXPRG.x <int> 2, 2, NA, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, NA, 2, 2, 2, ~
$ DMDHRGND.x <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 1,~
$ DMDHRAGE.x <int> 36, 36, 21, 60, 60, 60, 60, 60, 60, 60, 27, 44, 44, 44, 44,~
$ DMDHRBRN.x <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 3, 1, 1, 1, 1,~
$ DMDHREDU.x <int> 5, 5, 3, 4, 4, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 5, 4, 4, 4, 2,~
$ DMDHRMAR.x <int> 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5, 5, 1, 4, 4, 4, 1,~
$ DMDHSEDU.x <int> 4, 4, NA, 3, 3, 3, 3, 3, 3, 3, 1, NA, NA, NA, NA, 5, NA, NA~
$ SIALANG.x  <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ SIAPROXY.x <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ SIAINTRP.x <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ FIALANG.x  <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ FIAPROXY.x <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ FIAINTRP.x <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ MIALANG.x  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, NA, NA, NA,~
$ MIAPROXY.x <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, NA, NA, NA,~
$ MIAINTRP.x <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, NA, NA, NA,~
$ AIALANG.x  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, NA, NA, NA,~
```

```
$ WTINT2YR.x <dbl> 26457.708, 26457.708, 46374.162, 32632.520, 32632.520, 3263~
$ WTMEC2YR.x <dbl> 26770.585, 26770.585, 49416.756, 32058.654, 32058.654, 3205~
$ SDMVPSU.x  <int> 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1,~
$ SDMVSTRA.x <int> 48, 48, 45, 53, 53, 53, 53, 53, 53, 53, 57, 54, 54, 54, 54,~
$ OSQ010A    <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ010B    <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ010C    <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ020A    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ020B    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ020C    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030AA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040AA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050AA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030AB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040AB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050AB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030AC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040AC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050AC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030BA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040BA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050BA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030BB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040BB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050BB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030BC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040BC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050BC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030BD   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040BD   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050BD   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030BE   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040BE   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050BE   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030BF   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040BF   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050BF   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030BG   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040BG   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050BG   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050CA   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```
$ OSD030CB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050CB   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050CC   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CD   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CD   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050CD   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CE   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CE   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050CE   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CF   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CF   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD050CF   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CG   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CG   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CH   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CH   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CI   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CI   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD030CJ   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ040CJ   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ080     <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ090A    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100A    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110A    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120A    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090B    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100B    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110B    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120B    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090C    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100C    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110C    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120C    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090D    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100D    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110D    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120D    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090E    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100E    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110E    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```
$ OSQ120E     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090F     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100F     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110F     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120F     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090G     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100G     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110G     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120G     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090H     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100H     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110H     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120H     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090I     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100I     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110I     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120I     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ090J     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ100J     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSD110J     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ120J     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ060      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ070      <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ130      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ140Q     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ140U     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ150      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ160A     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ160B     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ170      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ180      <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ190      <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ200      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ OSQ210      <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ OSQ220      <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ SDDSRVYR.y <int> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,~
$ RIDSTATR.y <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ RIDEXMON.y <int> 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2,~
$ RIAGENDR.y <int> 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2,~
$ RIDAGEYR.y <int> 44, 44, 21, 59, 59, 59, 59, 59, 59, 59, 27, 44, 44, 44, 44,~
$ RIDAGEMN.y <int> 535, 535, 255, 711, 711, 711, 711, 711, 711, 711, 329, 528,~
$ RIDAGEEX.y <int> 536, 536, 256, 711, 711, 711, 711, 711, 711, 711, 330, 528,~
$ RIDRETH1.y <int> 4, 4, 2, 4, 4, 4, 4, 4, 4, 4, 1, 5, 5, 5, 5, 5, 4, 4, 4, 3,~
```

```
$ DMQMILIT.y <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ DMDBORN.y  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 3, 1, 1, 1, 1,~
$ DMDCITZN.y <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1,~
$ DMDYRSUS.y <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 6, NA, NA, NA, NA, ~
$ DMDEDUC3.y <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DMDEDUC2.y <int> 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 5, 4, 4, 4, 2,~
$ DMDSCHOL.y <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DMDMARTL.y <int> 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5, 5, 1, 4, 4, 4, 1,~
$ DMDHHSIZ.y <int> 4, 4, 6, 2, 2, 2, 2, 2, 2, 2, 5, 2, 2, 2, 2, 4, 7, 7, 7, 5,~
$ DMDFMSIZ.y <int> 4, 4, 6, 2, 2, 2, 2, 2, 2, 2, 5, 2, 2, 2, 2, 4, 7, 7, 7, 5,~
$ INDHHINC.y <int> 11, 11, 3, 7, 7, 7, 7, 7, 7, 7, 7, 3, 3, 3, 3, 10, 7, 7, 7,~
$ INDFMINC.y <int> 11, 11, 3, 7, 7, 7, 7, 7, 7, 7, 7, 3, 3, 3, 3, 10, 7, 7, 7,~
$ INDFMPIR.y <dbl> 4.65, 4.65, 0.46, 3.03, 3.03, 3.03, 3.03, 3.03, 3.03, 3.03,~
$ RIDEXPRG.y <int> 2, 2, NA, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, NA, 2, 2, 2, ~
$ DMDHRGND.y <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 1,~
$ DMDHRAGE.y <int> 36, 36, 21, 60, 60, 60, 60, 60, 60, 60, 27, 44, 44, 44, 44,~
$ DMDHRBRN.y <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 3, 1, 1, 1, 1,~
$ DMDHREDU.y <int> 5, 5, 3, 4, 4, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 5, 4, 4, 4, 2,~
$ DMDHRMAR.y <int> 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5, 5, 1, 4, 4, 4, 1,~
$ DMDHSEDU.y <int> 4, 4, NA, 3, 3, 3, 3, 3, 3, 3, 1, NA, NA, NA, NA, 5, NA, NA~
$ SIALANG.y  <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ SIAPROXY.y <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ SIAINTRP.y <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ FIALANG.y  <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ FIAPROXY.y <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ FIAINTRP.y <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ MIALANG.y  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, NA, NA, NA,~
$ MIAPROXY.y <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, NA, NA, NA,~
$ MIAINTRP.y <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, NA, NA, NA,~
$ AIALANG.y  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, NA, NA, NA,~
$ WTINT2YR.y <dbl> 26457.708, 26457.708, 46374.162, 32632.520, 32632.520, 3263~
$ WTMEC2YR.y <dbl> 26770.585, 26770.585, 49416.756, 32058.654, 32058.654, 3205~
$ SDMVPSU.y  <int> 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1,~
$ SDMVSTRA.y <int> 48, 48, 45, 53, 53, 53, 53, 53, 53, 53, 57, 54, 54, 54, 54,~
$ RXDUSE     <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1, 1, 1,~
$ RXDDRUG    <chr> "CARVEDILOL", "RAMIPRIL", " ", "ACETAMINOPHEN; DICHLORALPHE~
$ RXDDRGID   <chr> "d03847", "d00728", " ", "d03459", "d00023", "d04785", "d00~
$ RXQSEEN    <int> 1, 1, NA, 1, 1, 1, 1, 1, 1, 1, NA, 1, 1, 1, 1, NA, 1, 1, 1,~
$ RXDDAYS    <int> 730, 730, NA, 30, 1095, 1095, 730, 1460, 730, 61, NA, 365, ~
$ RXDCOUNT   <int> 2, 2, NA, 7, 7, 7, 7, 7, 7, 7, NA, 4, 4, 4, 4, NA, 3, 3, 3,~
$ SDDSRVYR   <int> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,~
$ RIDSTATR   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ RIDEXMON   <int> 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2,~
```

```
$ RIAGENDR   <int> 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2,~
$ RIDAGEYR   <int> 44, 44, 21, 59, 59, 59, 59, 59, 59, 59, 27, 44, 44, 44, 44,~
$ RIDAGEMN   <int> 535, 535, 255, 711, 711, 711, 711, 711, 711, 711, 329, 528,~
$ RIDAGEEX   <int> 536, 536, 256, 711, 711, 711, 711, 711, 711, 711, 330, 528,~
$ RIDRETH1   <int> 4, 4, 2, 4, 4, 4, 4, 4, 4, 4, 1, 5, 5, 5, 5, 5, 4, 4, 4, 3,~
$ DMQMILIT   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ DMDBORN    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 3, 1, 1, 1, 1,~
$ DMDCITZN   <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1,~
$ DMDYRSUS   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 6, NA, NA, NA, NA, ~
$ DMDEDUC3   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DMDEDUC2   <int> 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 5, 4, 4, 4, 2,~
$ DMDSCHOL   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ DMDMARTL   <int> 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5, 5, 1, 4, 4, 4, 1,~
$ DMDHHSIZ   <int> 4, 4, 6, 2, 2, 2, 2, 2, 2, 2, 5, 2, 2, 2, 2, 4, 7, 7, 7, 5,~
$ DMDFMSIZ   <int> 4, 4, 6, 2, 2, 2, 2, 2, 2, 2, 5, 2, 2, 2, 2, 4, 7, 7, 7, 5,~
$ INDHHINC   <int> 11, 11, 3, 7, 7, 7, 7, 7, 7, 7, 7, 3, 3, 3, 3, 10, 7, 7, 7,~
$ INDFMINC   <int> 11, 11, 3, 7, 7, 7, 7, 7, 7, 7, 7, 3, 3, 3, 3, 10, 7, 7, 7,~
$ INDFMPIR   <dbl> 4.65, 4.65, 0.46, 3.03, 3.03, 3.03, 3.03, 3.03, 3.03, 3.03,~
$ RIDEXPRG   <int> 2, 2, NA, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, NA, 2, 2, 2, ~
$ DMDHRGND   <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 1,~
$ DMDHRAGE   <int> 36, 36, 21, 60, 60, 60, 60, 60, 60, 60, 27, 44, 44, 44, 44,~
$ DMDHRBRN   <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 3, 1, 1, 1, 1,~
$ DMDHREDU   <int> 5, 5, 3, 4, 4, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 5, 4, 4, 4, 2,~
$ DMDHRMAR   <int> 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5, 5, 1, 4, 4, 4, 1,~
$ DMDHSEDU   <int> 4, 4, NA, 3, 3, 3, 3, 3, 3, 3, 1, NA, NA, NA, NA, 5, NA, NA~
$ SIALANG    <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ SIAPROXY   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ SIAINTRP   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ FIALANG    <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ FIAPROXY   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ FIAINTRP   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ MIALANG    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, NA, NA, NA,~
$ MIAPROXY   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, NA, NA, NA,~
$ MIAINTRP   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, NA, NA, NA,~
$ AIALANG    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, NA, NA, NA,~
$ WTINT2YR   <dbl> 26457.708, 26457.708, 46374.162, 32632.520, 32632.520, 3263~
$ WTMEC2YR   <dbl> 26770.585, 26770.585, 49416.756, 32058.654, 32058.654, 3205~
$ SDMVPSU    <int> 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1,~
$ SDMVSTRA   <int> 48, 48, 45, 53, 53, 53, 53, 53, 53, 53, 57, 54, 54, 54, 54,~
```

Project idea 3 - Heart disease

Introduction and data

- State the source of the data set.

This dataset was published in this form by Kamil Pytlak, and it is taken from the Centers of Disease Control and Prevention (CDC) in the USA. It was retrieved from Pytlak's post on Kaggle.com. Link: https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease/versions/2?resource=download

- Describe when and how it was originally collected (by the original data curator, not necessarily how you found the data)

The data was collected by the Behavioral Risk Factor Surveillance System (BRFSS) which conducts annual telephone surveys about the health status of U.S. residents in all 50 states, the District of Columbia and in the 3 US territories. This data set includes data from 2020, consisting of 401958 rows and 279 columns. Of note is that the classes are not balanced, and the author advises fixing the weights or undersampling.

- Describe the observations and the general characteristics being measured in the data

The key indicators for the risk of heart disease include smoking, diabetic status, obesity (high BMI), not getting enough physical activity or drinking too much alcohol. In addition to these, the data records other information such as history of stroke, mental health, age category, race, general health, sleep time, kidney disease and skin cancer.

Research question

- Describe a research question you're interested in answering using this data.

What can we conclude about a person's likelihood of heart disease from their (1) BMI, (2) sex, (3) age category, (4) sleeping hours, (5) race, and (6) whether they are physically active? Note that whether someone has heart disease is a binary variable. I have another research question prepared with the outcome variable set to a person's BMI, which is a continuous variable:

What can we conclude about a person's BMI from their (1) alcohol consumption, (2) sex, (3) age category, (4) sleeping hours, (5) race, and (6) whether they are physically active?

Glimpse of data

- Use the `glimpse` function to provide an overview of the dataset

```
heart_disease_data <- read_csv("data-3/heart_2020_cleaned.csv")
```

```
Rows: 319795 Columns: 18
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (14): HeartDisease, Smoking, AlcoholDrinking, Stroke, DiffWalking, Sex, ...
dbl  (4): BMI, PhysicalHealth, MentalHealth, SleepTime

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(heart_disease_data)
```

```
Rows: 319,795
Columns: 18
$ HeartDisease     <chr> "No", "No", "No", "No", "No", "Yes", "No", "No", "No"~
$ BMI              <dbl> 16.60, 20.34, 26.58, 24.21, 23.71, 28.87, 21.63, 31.6~
$ Smoking          <chr> "Yes", "No", "Yes", "No", "No", "Yes", "No", "Yes", "~
$ AlcoholDrinking  <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No",~
$ Stroke           <chr> "No", "Yes", "No", "No", "No", "No", "No", "No", "No"~
$ PhysicalHealth   <dbl> 3, 0, 20, 0, 28, 6, 15, 5, 0, 0, 30, 0, 0, 7, 0, 1, 5~
$ MentalHealth     <dbl> 30, 0, 30, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 30, 0, 2,~
$ DiffWalking      <chr> "No", "No", "No", "No", "Yes", "Yes", "No", "Yes", "N~
$ Sex              <chr> "Female", "Female", "Male", "Female", "Female", "Fema~
$ AgeCategory      <chr> "55-59", "80 or older", "65-69", "75-79", "40-44", "7~
$ Race             <chr> "White", "White", "White", "White", "White", "Black",~
$ Diabetic         <chr> "Yes", "No", "Yes", "No", "No", "No", "No", "Yes", "N~
$ PhysicalActivity <chr> "Yes", "Yes", "Yes", "No", "Yes", "No", "Yes", "No", ~
$ GenHealth        <chr> "Very good", "Very good", "Fair", "Good", "Very good"~
$ SleepTime        <dbl> 5, 7, 8, 6, 8, 12, 4, 9, 5, 10, 15, 5, 8, 7, 5, 6, 10~
$ Asthma           <chr> "Yes", "No", "Yes", "No", "No", "No", "Yes", "Yes", "~
$ KidneyDisease    <chr> "No", "No", "No", "No", "No", "No", "No", "No", "Yes"~
$ SkinCancer       <chr> "Yes", "No", "No", "Yes", "No", "No", "Yes", "No", "N~
```