
Installation

```
# The following 2 lines causes a compilation issue when knitting, uncomment if you run yourself
# install.packages('tidyverse')
# install.packages('dplyr')
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(haven)
library(dplyr)
```

Pre-process dataset

```
# Treatment Variable: H4WP3 (if the biological mother has ever spent time in jail or prison)
# Outcome Variable: H4ED1 (respondent's high school graduation status)

# CHANGE THIS TO WHERE YOUR WAVE 4 DATASET IS. THE WAVE 4 FOLDER SHOULD BE NAMED 'Wave4'
wave4_dir <- "./Wave4/"

concat <- function(s1, s2) {
  return (paste(s1, s2, sep=""))
}

# w4inhome_dvn.sav is the main datatable that contains our treatment and outcome variable
inhome_df <- read_sav(concat(wave4_dir, "w4inhome_dvn.sav"))
birth_df <- read_sav(concat(wave4_dir, "w4birth_dvn.sav"))
segment_df <- read_sav(concat(wave4_dir, "w4segment_dvn.sav"))

# join against these table because they could have useful information
# keep rows with unique AIDs only
birth_df <- birth_df %>% distinct(AID, .keep_all = TRUE)
segment_df <- segment_df %>% distinct(AID, .keep_all = TRUE)

combined_df <- inhome_df %>%
  left_join(birth_df, by = "AID") %>%
  left_join(segment_df, by = "AID")
```

```

relevant_df = combined_df %>%
  filter(H4WP3 <= 1) %>% # Filter out refused or don't know for H4WP3
  filter(H4ED1 <= 4) %>% # Filter out don't know for H4ED1
  mutate(H4WP4 = case_when(H4WP4 >= 97 ~ NA, .default = H4WP4),
         H4WP5 = case_when(H4WP5 >= 94 ~ NA, .default = H4WP5),
         H4WP24 = case_when(H4WP24 == 7 ~ NA, .default = H4WP24),
         H4EC1 = case_when(H4EC1 >= 96 ~ NA, .default = H4EC1),
         H4ED7 = case_when(H4ED7 >= 5 ~ NA, .default = H4ED7),
         H4EC7 = case_when(H4EC7 >= 96 ~ NA, .default = H4EC7),
         H4ED1 = case_when(H4ED1 %in% c(1, 2) ~ 1, # Graduated high school
                           H4ED1 %in% c(3,4) ~ 0), # Did not graduate high school, This is dichotomizing
         H4EO7 = case_when(H4EO7 > 4 ~ NA, .default = H4EO7)) %>% #Change Don't Know/Missing/Refuse val
  select(H4WP3, H4WP4, H4WP5, H4WP24, H4OD4, H4EC1, H4EC7, H4LM28, H4ED7, H4EO7, BIO_SEX4, H4ED1)

relevant_df <- relevant_df %>%
  rename(Mother.Incarcerated = H4WP3,
         Mother.Incarcerated.Times = H4WP4,
         Mother.Incarcerated.Respondent.Age = H4WP5,
         Mother.Relationship.Respondent = H4WP24,
         US.Citizen = H4OD4,
         Household.Income = H4EC1,
         Household.Assets = H4EC7,
         Family.Support.Education = H4ED7,
         Family.Responsibilities = H4LM28,
         Neighborhood.Safety = H4EO7,
         Sex = BIO_SEX4,
         Highschool.Graduated = H4ED1)

# Print colnames, look up colnames on
# https://addhealth.cpc.unc.edu/documentation/codebook-explorer/#/
# To see if they are relevant
colnames(combined_df)

```

```

## [1] "AID" "IMONTH4" "IDAY4" "IYEAR4" "BIO_SEX4" "VERSION4"
## [7] "BREAK_Q" "PRYEAR4" "PRETEST4" "PRISON4" "H4OD1M" "H4OD1Y"
## [13] "H4OD2A" "H4OD2B" "H4OD2C" "H4OD2D" "H4OD2E" "H4OD2F"
## [19] "H4OD3" "H4OD4" "H4OD5" "H4OD6Y" "H4OD7" "H4WP1"
## [25] "H4WP2Y" "H4WP3" "H4WP4" "H4WP5" "H4WP6" "H4WP7"
## [31] "H4WP8Y" "H4WP9" "H4WP10" "H4WP11" "H4WP12" "H4WP13"
## [37] "H4WP14" "H4WP15Y" "H4WP16" "H4WP17" "H4WP18" "H4WP19"
## [43] "H4WP20" "H4WP21" "H4WP22" "H4WP23" "H4WP24" "H4WP25"
## [49] "H4WP26" "H4WP27" "H4WP28" "H4WP29Y" "H4WP30" "H4WP31"
## [55] "H4WP32" "H4WP33" "H4WP34" "H4WP35" "H4WP36" "H4WP37"
## [61] "H4WP38" "H4WP39" "H4WP40" "H4WS1" "H4WS2" "H4WS4"
## [67] "H4WS5" "H4GH1" "H4GH2" "H4GH3M" "H4GH3D" "H4GH3Y"
## [73] "H4GH4A" "H4GH4B" "H4GH5F" "H4GH5I" "H4GH6" "H4GH7"
## [79] "H4GH8" "H4GH9" "H4GH10" "H4GH11H" "H4GH11M" "H4GH11T"
## [85] "H4GH12" "H4GH13H" "H4GH13M" "H4GH13T" "H4HS1" "H4HS2A"
## [91] "H4HS2B" "H4HS2C" "H4HS2D" "H4HS3" "H4HS4" "H4HS5"
## [97] "H4HS6" "H4HS7" "H4HS8" "H4HS9" "H4ID1" "H4ID2"
## [103] "H4ID3" "H4ID4" "H4ID5A" "H4ID5B" "H4ID5C" "H4ID5D"
## [109] "H4ID5E" "H4ID5F" "H4ID5G" "H4ID5H" "H4ID5I" "H4ID5J"
## [115] "H4ID5K" "H4ID5L" "H4ID5N" "H4ID7" "H4ID8" "H4ID9A"

```

## [121]	"H4ID9B"	"H4ID9C"	"H4ID9D"	"H4ID9E"	"H4ID9F"	"H4ID9G"
## [127]	"H4ID10A"	"H4ID10B"	"H4ID10C"	"H4ID10D"	"H4ID10E"	"H4ID10F"
## [133]	"H4ID10G"	"H4ID10H"	"H4ID11"	"H4ID12"	"H4ID13"	"H4ID14"
## [139]	"H4ID15"	"H4ID16"	"H4ID17"	"H4ID18"	"H4ID19"	"H4ID20"
## [145]	"H4ID21"	"H4ID22"	"H4ID23"	"H4ID24"	"H4ID25"	"H4ID26"
## [151]	"H4SP1H"	"H4SP1M"	"H4SP1T"	"H4SP2H"	"H4SP2M"	"H4SP2T"
## [157]	"H4SP3H"	"H4SP3M"	"H4SP3T"	"H4SP4H"	"H4SP4M"	"H4SP4T"
## [163]	"H4SP5"	"H4SP6"	"H4SP7"	"H4HR1"	"H4HR2"	"H4HR3"
## [169]	"H4HR4"	"H4HR5A"	"H4HR6A"	"H4HR7A"	"H4HR8A"	"H4HR9A"
## [175]	"H4HR5B"	"H4HR6B"	"H4HR7B"	"H4HR8B"	"H4HR9B"	"H4HR5C"
## [181]	"H4HR6C"	"H4HR7C"	"H4HR8C"	"H4HR9C"	"H4HR5D"	"H4HR6D"
## [187]	"H4HR7D"	"H4HR8D"	"H4HR9D"	"H4HR5E"	"H4HR6E"	"H4HR7E"
## [193]	"H4HR8E"	"H4HR9E"	"H4HR5F"	"H4HR6F"	"H4HR7F"	"H4HR8F"
## [199]	"H4HR9F"	"H4HR5G"	"H4HR6G"	"H4HR7G"	"H4HR8G"	"H4HR9G"
## [205]	"H4HR5H"	"H4HR6H"	"H4HR7H"	"H4HR8H"	"H4HR9H"	"H4HR5I"
## [211]	"H4HR6I"	"H4HR7I"	"H4HR8I"	"H4HR9I"	"H4HR5J"	"H4HR6J"
## [217]	"H4HR7J"	"H4HR8J"	"H4HR9J"	"H4HR5K"	"H4HR6K"	"H4HR7K"
## [223]	"H4HR8K"	"H4HR9K"	"H4HR5L"	"H4HR6L"	"H4HR7L"	"H4HR8L"
## [229]	"H4HR9L"	"H4HR5M"	"H4HR6M"	"H4HR7M"	"H4HR8M"	"H4HR9M"
## [235]	"H4HR5N"	"H4HR6N"	"H4HR7N"	"H4HR8N"	"H4HR9N"	"H4HR5O"
## [241]	"H4HR6O"	"H4HR7O"	"H4HR8O"	"H4HR9O"	"H4HR5P"	"H4HR6P"
## [247]	"H4HR7P"	"H4HR8P"	"H4HR9P"	"H4HR10"	"H4HR11YA"	"H4HR11YB"
## [253]	"H4HR11YC"	"H4HR11YD"	"H4HR11YE"	"H4HR11YF"	"H4HR11YG"	"H4HR11YH"
## [259]	"H4HR11YI"	"H4HR11YJ"	"H4HR11YK"	"H4HR11YL"	"H4HR11YM"	"H4HR11YN"
## [265]	"H4HR11YO"	"H4ED1"	"H4ED2"	"H4ED3A"	"H4ED3B"	"H4ED3C"
## [271]	"H4ED3D"	"H4ED3E"	"H4ED3F"	"H4ED3G"	"H4ED3H"	"H4ED4A"
## [277]	"H4ED4B"	"H4ED4C"	"H4ED4D"	"H4ED4E"	"H4ED4F"	"H4ED4G"
## [283]	"H4ED4H"	"H4ED5A"	"H4ED5B"	"H4ED5C"	"H4ED5D"	"H4ED5E"
## [289]	"H4ED5F"	"H4ED5G"	"H4ED5H"	"H4ED6"	"H4ED7"	"H4ED8"
## [295]	"H4ED9"	"H4MI1"	"H4MI2"	"H4MI3"	"H4MI4A"	"H4MI4B"
## [301]	"H4MI4C"	"H4MI4D"	"H4MI4E"	"H4MI5"	"H4MI6A"	"H4MI6B"
## [307]	"H4MI6C"	"H4MI7"	"H4MI8Y"	"H4MI9Y"	"H4MI10"	"H4MI11Y"
## [313]	"H4MI11M"	"H4MI12Y"	"H4MI12M"	"H4MI13"	"H4MI14"	"H4MI15"
## [319]	"H4MI16A"	"H4MI16B"	"H4MI16C"	"H4MI16D"	"H4MI17"	"H4LM1"
## [325]	"H4LM2"	"H4LM3"	"H4LM4"	"H4LM5"	"H4LM6"	"H4LM7"
## [331]	"H4LM8"	"H4LM9M"	"H4LM9Y"	"H4LM10"	"H4LM11"	"H4LM12"
## [337]	"H4LM13"	"H4LM14"	"H4LM15Y"	"H4LM16Y"	"H4LM17"	"H4LM18"
## [343]	"H4LM19"	"H4LM20"	"H4LM21A"	"H4LM21B"	"H4LM21C"	"H4LM22"
## [349]	"H4LM23"	"H4LM24"	"H4LM25"	"H4LM26"	"H4LM27"	"H4LM28"
## [355]	"H4LM29"	"H4LM30"	"H4EC1"	"H4EC2"	"H4EC3"	"H4EC4"
## [361]	"H4EC5"	"H4EC6"	"H4EC7"	"H4EC8"	"H4EC9"	"H4EC10"
## [367]	"H4EC11"	"H4EC12"	"H4EC13"	"H4EC14"	"H4EC15"	"H4EC16"
## [373]	"H4EC17"	"H4EC18"	"H4EC19"	"H4RE1"	"H4RE2"	"H4RE3"
## [379]	"H4RE4"	"H4RE5"	"H4RE6"	"H4RE7"	"H4RE8"	"H4RE9"
## [385]	"H4RE10"	"H4RE11"	"H4MH1"	"C4WD90_1"	"C4WD90_2"	"C4WD90_3"
## [391]	"H4MH2"	"H4MH3"	"H4MH4"	"H4MH5"	"H4MH6"	"H4MH7"
## [397]	"H4MH8"	"H4MH9"	"H4MH10"	"C4WD60_1"	"C4WD60_2"	"C4WD60_3"
## [403]	"H4MH11A"	"H4MH11B"	"H4MH12A"	"H4MH12B"	"H4MH13A"	"H4MH13B"
## [409]	"H4MH14A"	"H4MH14B"	"H4MH15A"	"H4MH15B"	"H4MH16A"	"H4MH16B"
## [415]	"H4MH17A"	"H4MH17B"	"C4NUMSCR"	"H4MH18"	"H4MH19"	"H4MH20"
## [421]	"H4MH21"	"H4MH22"	"H4MH23"	"H4MH24"	"H4MH25"	"H4MH26"
## [427]	"H4MH27"	"H4MH28"	"H4MH29"	"H4SE1"	"H4SE2"	"H4SE3"
## [433]	"H4SE4"	"H4SE5"	"H4SE6"	"H4SE7"	"H4SE8"	"H4SE9"
## [439]	"H4SE10"	"H4SE11"	"H4SE12"	"H4SE13"	"H4SE14"	"H4SE15"

##	[445]	"H4SE16"	"H4SE17"	"H4SE18"	"H4SE19"	"H4SE20"	"H4SE21"
##	[451]	"H4SE22"	"H4SE23"	"H4SE24"	"H4SE25"	"H4SE26A"	"H4SE26B"
##	[457]	"H4SE26C"	"H4SE26D"	"H4SE26E"	"H4SE26F"	"H4SE26G"	"H4SE26H"
##	[463]	"H4SE26I"	"H4SE26J"	"H4SE26K"	"H4SE26L"	"H4SE26M"	"H4SE26N"
##	[469]	"H4SE26O"	"H4SE26P"	"H4SE26Q"	"H4SE26R"	"H4SE26S"	"H4SE26T"
##	[475]	"H4SE26V"	"H4SE27"	"H4SE28"	"H4SE29"	"H4SE30"	"H4SE31"
##	[481]	"H4SE32"	"H4SE33"	"H4SE34"	"H4SE35"	"H4SE36A"	"H4SE36B"
##	[487]	"H4SE36C"	"H4SE36D"	"H4SE36E"	"H4SE36F"	"H4SE36G"	"H4SE36H"
##	[493]	"H4SE36I"	"H4SE36J"	"H4SE36K"	"H4SE36L"	"H4SE36N"	"H4SE36O"
##	[499]	"H4SE37A"	"H4SE37B"	"H4SE37C"	"H4SE37D"	"H4SE37E"	"H4SE37F"
##	[505]	"H4SE37G"	"H4SE37H"	"H4SE37I"	"H4SE37J"	"H4SE37K"	"H4SE37L"
##	[511]	"H4SE37N"	"H4SE37O"	"H4TR1"	"H4TR2"	"H4TR3"	"H4TR4"
##	[517]	"H4TR5"	"H4TR6"	"H4TR7"	"H4TR8"	"H4TR9"	"H4TR10"
##	[523]	"H4TR11"	"H4TR12"	"PTNR_ID.x"	"H4RD1"	"H4RD2D"	"H4RD2M"
##	[529]	"H4RD2Y"	"H4RD3"	"H4RD4"	"H4RD5"	"H4RD6"	"H4RD7A"
##	[535]	"H4RD7B"	"H4RD7C"	"H4RD7D"	"H4RD7E"	"H4RD7F"	"H4RD7G"
##	[541]	"H4RD8"	"H4RD9"	"H4RD10"	"H4RD11"	"H4RD12"	"H4RD13"
##	[547]	"H4RD14"	"H4RD14T"	"H4RD15"	"H4RD16"	"H4RD17"	"H4RD18"
##	[553]	"H4RD19"	"H4RD20"	"H4RD21"	"H4RD22"	"H4RD23"	"H4RD24"
##	[559]	"H4RD25"	"H4KK14"	"H4KK15A"	"H4KK15B"	"H4KK15C"	"H4KK15D"
##	[565]	"H4DS1"	"H4DS2"	"H4DS3"	"H4DS4"	"H4DS5"	"H4DS6"
##	[571]	"H4DS7"	"H4DS8"	"H4DS9"	"H4DS10"	"H4DS11"	"H4DS12"
##	[577]	"H4DS13"	"H4DS14"	"H4DS15"	"H4DS16"	"H4DS17"	"H4DS18"
##	[583]	"H4DS19"	"H4DS20"	"H4CJ1"	"H4CJ2"	"H4CJ3"	"H4CJ4"
##	[589]	"H4CJ5"	"H4CJ6"	"H4CJ7A"	"H4CJ7B"	"H4CJ7C"	"H4CJ7D"
##	[595]	"H4CJ7E"	"H4CJ7F"	"H4CJ7G"	"H4CJ7H"	"H4CJ7I"	"H4CJ7J"
##	[601]	"H4CJ7K"	"H4CJ8"	"H4CJ9A"	"H4CJ9B"	"H4CJ9C"	"H4CJ9D"
##	[607]	"H4CJ9E"	"H4CJ9F"	"H4CJ9G"	"H4CJ9H"	"H4CJ9I"	"H4CJ9J"
##	[613]	"H4CJ9K"	"H4CJ10"	"H4CJ11"	"H4CJ12"	"H4CJ13A"	"H4CJ13B"
##	[619]	"H4CJ13C"	"H4CJ13D"	"H4CJ13E"	"H4CJ13F"	"H4CJ13G"	"H4CJ13H"
##	[625]	"H4CJ13I"	"H4CJ13J"	"H4CJ13K"	"H4CJ14"	"H4CJ15A"	"H4CJ15B"
##	[631]	"H4CJ15C"	"H4CJ15D"	"H4CJ15E"	"H4CJ15F"	"H4CJ15G"	"H4CJ15H"
##	[637]	"H4CJ15I"	"H4CJ15J"	"H4CJ15K"	"H4CJ16"	"H4CJ17"	"H4CJ18"
##	[643]	"H4CJ19"	"H4CJ20"	"H4CJ21"	"H4CJ22Y"	"H4CJ22M"	"H4CJ23Y"
##	[649]	"H4CJ23M"	"H4CJ24Y"	"H4CJ24M"	"H4CJ25Y"	"H4CJ25M"	"H4T01"
##	[655]	"H4T02"	"H4T03"	"H4T04"	"H4T05"	"H4T06"	"H4T07"
##	[661]	"H4T08"	"H4T09"	"H4T010"	"H4T011"	"H4T012"	"H4T013"
##	[667]	"H4T014"	"H4T015"	"H4T016"	"H4T017"	"H4T018"	"H4T019"
##	[673]	"H4T020"	"H4T021"	"H4T022"	"H4T023"	"H4T024"	"H4T025"
##	[679]	"H4T026"	"H4T027"	"H4T028"	"H4T029"	"H4T030"	"H4T031"
##	[685]	"H4T032H"	"H4T032M"	"H4T032T"	"H4T033"	"H4T034"	"H4T035"
##	[691]	"H4T036"	"H4T037"	"H4T038"	"H4T039"	"H4T040"	"H4T041"
##	[697]	"H4T042H"	"H4T042M"	"H4T042T"	"H4T043"	"H4T044"	"H4T045"
##	[703]	"H4T046"	"H4T047"	"H4T048"	"H4T049"	"H4T050"	"H4T051"
##	[709]	"H4T052"	"H4T053"	"H4T054"	"H4T055"	"H4T056"	"H4T057"
##	[715]	"H4T058"	"H4T059"	"H4T060"	"H4T061"	"H4T062"	"H4T063"
##	[721]	"H4T064A"	"H4T064B"	"H4T064C"	"H4T064D"	"H4T065A"	"H4T065B"
##	[727]	"H4T065C"	"H4T065D"	"H4T065E"	"H4T066"	"H4T067"	"H4T068"
##	[733]	"H4T069"	"H4T070"	"H4T071"	"H4T072"	"H4T073H"	"H4T073M"
##	[739]	"H4T073T"	"H4T074"	"H4T075"	"H4T076"	"H4T077"	"H4T078"
##	[745]	"H4T079"	"H4T080"	"H4T081"	"H4T082"	"H4T083"	"H4T084"
##	[751]	"H4T085"	"H4T086"	"H4T087"	"H4T088"	"H4T089"	"H4T090"
##	[757]	"H4T091"	"H4T092"	"H4T093"	"H4T094"	"H4T095"	"H4T096"
##	[763]	"H4T097"	"H4T098"	"H4T099"	"H4T100"	"H4T101H"	"H4T101M"

```
## [769] "H4T0101T" "H4T0102" "H4T0103" "H4T0104" "H4T0105" "H4T0106"
## [775] "H4T0107" "H4T0108" "H4T0109" "H4T0110" "H4T0111" "H4T0112"
## [781] "H4T0113" "H4T0114" "H4T0115" "H4T0116" "H4T0117" "H4T0118"
## [787] "H4T0119" "H4T0120" "H4MA1" "H4MA2" "H4MA3" "H4MA4"
## [793] "H4MA5" "H4MA6" "H4DA1" "H4DA2" "H4DA3" "H4DA4"
## [799] "H4DA5" "H4DA6" "H4DA7" "H4DA8" "H4DA9" "H4DA10"
## [805] "H4DA11" "H4DA12H" "H4DA12M" "H4DA12T" "H4DA13" "H4DA14"
## [811] "H4DA15A" "H4DA15B" "H4DA15C" "H4DA15D" "H4DA15E" "H4DA16"
## [817] "H4DA17" "H4DA18" "H4DA19" "H4DA20" "H4DA21" "H4DA22"
## [823] "H4DA23" "H4DA24" "H4DA25" "H4DA26" "H4DA27" "H4DA28"
## [829] "H4PE1" "H4PE2" "H4PE3" "H4PE4" "H4PE5" "H4PE6"
## [835] "H4PE7" "H4PE8" "H4PE9" "H4PE10" "H4PE11" "H4PE12"
## [841] "H4PE13" "H4PE14" "H4PE15" "H4PE16" "H4PE17" "H4PE18"
## [847] "H4PE19" "H4PE20" "H4PE21" "H4PE22" "H4PE23" "H4PE24"
## [853] "H4PE25" "H4PE26" "H4PE27" "H4PE28" "H4PE29" "H4PE30"
## [859] "H4PE31" "H4PE32" "H4PE33" "H4PE34" "H4PE35" "H4PE36"
## [865] "H4PE37" "H4PE38" "H4PE39" "H4PE40" "H4PE41" "H4CMONTH"
## [871] "H4CDAY" "H4CYEAR" "H4ARM" "H4CUFF" "H4CUFFLG" "H4SBP"
## [877] "H4DBP" "H4BPCLS" "H4BPFLG" "H4PR" "H4PP" "H4MAP"
## [883] "H4HGT" "H4WGT" "H4WTLBS" "H4BMI" "H4BMICLS" "H4WAIST"
## [889] "H4IR1" "H4IR2" "H4IR3" "H4IR4" "H4IR5" "H4IR6"
## [895] "H4IR7" "H4IR8" "H4IR8A" "H4IR9" "H4IR9A" "H4IR10"
## [901] "H4IR11" "H4IR12" "H4IR13" "H4IR14" "H4E01" "H4E02"
## [907] "H4E03" "H4E04" "H4E05A" "H4E05B" "H4E05C" "H4E05D"
## [913] "H4E05E" "H4E05F" "H4E05G" "H4E05H" "H4E05I" "H4E05J"
## [919] "H4E06" "H4E07" "PTNR_ID.y" "PRGN0" "LBN0" "H4LB1"
## [925] "H4LB2Y" "H4LB3" "H4LB4" "H4LB5" "H4LB6P" "H4LB60"
## [931] "H4LB7" "H4LB8" "H4LB9W" "H4LB9D" "H4LB10" "H4LB11Y"
## [937] "PTNR_ID" "H4TR25" "H4TR26" "H4TR27Y" "H4TR28Y" "TSDURATN"
## [943] "H4TR29"
```

```
# Print preview for dataset
relevant_df
```

```
## # A tibble: 5,064 x 12
##   Mother.Incarcerated Mother.Incarcerated.Times Mother.Incarcerated.Responden~1
##   <dbl+lbl>          <dbl+lbl>          <dbl+lbl>
## 1 0 [(0) No]        NA                      NA
## 2 0 [(0) No]        NA                      NA
## 3 0 [(0) No]        NA                      NA
## 4 0 [(0) No]        NA                      NA
## 5 0 [(0) No]        NA                      NA
## 6 0 [(0) No]        NA                      NA
## 7 0 [(0) No]        NA                      NA
## 8 0 [(0) No]        NA                      NA
## 9 0 [(0) No]        NA                      NA
## 10 0 [(0) No]       NA                      NA
## # i 5,054 more rows
## # i abbreviated name: 1: Mother.Incarcerated.Respondent.Age
## # i 9 more variables: Mother.Relationship.Respondent <dbl+lbl>,
## #   US.Citizen <dbl+lbl>, Household.Income <dbl+lbl>,
## #   Household.Assets <dbl+lbl>, Family.Responsibilities <dbl+lbl>,
## #   Family.Support.Education <dbl+lbl>, Neighborhood.Safety <dbl+lbl>,
## #   Sex <dbl+lbl>, Highschool.Graduated <dbl>
```

Overview of Data Pre-processing

Treatment and Outcome Variables

Our treatment variable is H4WP3 (if the biological mother has ever spent time in jail or prison). Our outcome variable is H4ED1 (respondent's high school graduation status).

$$A_i = \begin{cases} 0 & \text{if biological mother never spent time in prison} \\ 1 & \text{if biological mother ever spent time in prison} \end{cases}$$
$$Y_i = \begin{cases} 0 & \text{if did not received a high school diploma or equivalent} \\ 1 & \text{if received a high school diploma or equivalent} \end{cases}$$

Tables

We use the Wave 4 dataset from the ADD Health longitudinal study. Our treatment and outcome variables are found in the 'in-home' dataset within W4. Within this dataset, each respondent is classified by an AID. Since there are data present in order tables that we may need to adjust for, such as sex, partner relationships, birth information, etc., we join the 'in-home' table with the 'birth_dvn' and 'w4segment_dvn' tables also found in Wave 4 on AID to incorporate additional relevant information into our analysis.

Data Adjustment

For our treatment variable, we drop rows where our treatment variable is not well-defined (respondent entered don't know or did not respond).

For our outcome variable, we grouped 'finished high school with a diploma' and 'earned a high school equivalency degree' as one and 'earned a certificate of attendance or certificate of completion' and 'did not receive a high school diploma, equivalency degree (GED), or other certificate' as 0. Furthermore, we drop rows whose graduation status is 'don't know.'

For confounding and other relevant variables, we changed the encoded numerical value for 'don't know', 'refused', or other similar missing data to NA values.

Description of Columns

Visit <https://addhealth.cpc.unc.edu/documentation/codebook-explorer/> for a more detailed description of the columns.

Mother.Incarcerated (H4WP3) [Treatment Variable]: (Has/did) your biological mother ever (spent/spend) time in jail or prison?

Mother.Incarcerated.Times (H4WP4): How many times (has/did) your biological mother (spent/spend) time in jail or prison?

Mother.Incarcerated.Respondent.Age (H4WP5): How old were you when your biological mother went to jail or prison (the first time)?

Mother.Relationship.Respondent (H4WP24): How close do you feel to your [mother figure]?

US.Citizen (H4OD4): Were you born a U.S. citizen?

Household.Income (H4EC1): Thinking about your income and the income of everyone who lives in your household and contributes to the household budget, what was the total household income before taxes and deductions in {2006/2007/2008}? Include all sources of income, including non-legal sources.

Household.Assets (H4EC7): What is your best estimate of the total value of your assets and the assets of everyone who lives in your household and contributes to the household budget? Include all assets, such as bank accounts, retirement plans and stocks. Do not include equity in your home.

Family.Support.Education (H4ED7): In the past 12 months, have any relatives, including your parents or in-laws, helped you out by paying some of your educational expenses, such as tuition or books?

Family.Responsibilities (H4LM28): Indicate how much you would agree or disagree with this statement: Family responsibilities have interfered with my ability to work.

Neighborhood.Safety (H4EO7): How safe did you feel when you were in the sample member's/respondent's neighborhood?

Sex (BIO_SEX4): Respondent's Gender

Highschool.Graduated (H4ED1) [Outcome Variable]: What is your high school graduation status?

Conducting the Matching:

```
library(MatchIt)
# treatment formula based off of what the sufficient adjustment set is (those are the variables we incl
# NOT THIS: treatment_formula<-Mother.Incarcerated+Household.Income+Sex+Household.Assets+Family.Support
treatment_formula2<-Mother.Incarcerated~Household.Income+Neighborhood.Safety

#Perform Matching
clean_relevant_df<-relevant_df %>% select(-Mother.Incarcerated.Times, -Mother.Incarcerated.Respondent.A
matching<-matchit(treatment_formula2,
                  method="nearest",
                  distance="glm",
                  estimand="ATT",
                  data=clean_relevant_df)
matching
```

```
## A matchit object
## - method: 1:1 nearest neighbor matching without replacement
## - distance: Propensity score
## - estimated with logistic regression
## - number of obs.: 3812 (original), 246 (matched)
## - target estimand: ATT
## - covariates: Household.Income, Neighborhood.Safety
```

```
summary(matching)
```

```
##
## Call:
## matchit(formula = treatment_formula2, data = clean_relevant_df,
##          method = "nearest", distance = "glm", estimand = "ATT")
##
## Summary of Balance for All Data:
```

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio
## distance	0.0432	0.0319	0.4687	1.4032
## Household.Income	7.0163	8.1095	-0.3712	1.2851
## Neighborhood.Safety	1.7073	1.3670	0.4446	1.5397

```
##                      eCDF Mean eCDF Max
## distance              0.1162  0.2910
## Household.Income      0.0911  0.1999
## Neighborhood.Safety   0.0851  0.2327
##
## Summary of Balance for Matched Data:
##                      Means Treated Means Control Std. Mean Diff. Var. Ratio
## distance              0.0432      0.0432          0          1
## Household.Income      7.0163      7.0163          0          1
## Neighborhood.Safety   1.7073      1.7073          0          1
##                      eCDF Mean eCDF Max Std. Pair Dist.
## distance              0          0          0
## Household.Income      0          0          0
## Neighborhood.Safety   0          0          0
##
## Sample Sizes:
##           Control Treated
## All           3689     123
## Matched        123     123
## Unmatched     3566        0
## Discarded        0        0
```

```
summary(matching)$nn
```

```
##           Control Treated
## All (ESS)      3689     123
## All            3689     123
## Matched (ESS)   123     123
## Matched        123     123
## Unmatched     3566        0
## Discarded        0        0
```

method="exact" was too redundant and didn't work well with data and had terrible covariate balance be

```
table(clean_relevant_df$Mother.Incarcerated, matching$weights)
```

```
##
##           0      1
## 0 3566   123
## 1      0   123
```

```
fit<-lm(Highschool.Graduated~Mother.Incarcerated+Household.Income+Neighborhood.Safety,
        data=clean_relevant_df,
        weights=matching$weights)
fit
```

```
##
## Call:
## lm(formula = Highschool.Graduated ~ Mother.Incarcerated + Household.Income +
##     Neighborhood.Safety, data = clean_relevant_df, weights = matching$weights)
##
## Coefficients:
```



```
##           (Intercept)  Mother.Incarcerated  Household.Income
##           0.76486      -0.06504            0.01559
## Neighborhood.Safety
##           0.02604
```

```
summary(fit)
```

```
##
## Call:
## lm(formula = Highschool.Graduated ~ Mother.Incarcerated + Household.Income +
##     Neighborhood.Safety, data = clean_relevant_df, weights = matching$weights)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0040  0.0000  0.0000  0.0000  0.2586
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.764855   0.075176  10.174  <2e-16 ***
## Mother.Incarcerated -0.065041  0.040148  -1.620  0.1065
## Household.Income    0.015590  0.006884   2.265  0.0244 *
## Neighborhood.Safety 0.026041  0.026488   0.983  0.3265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3148 on 242 degrees of freedom
## Multiple R-squared:  0.03319,    Adjusted R-squared:  0.02121
## F-statistic: 2.769 on 3 and 242 DF,  p-value: 0.04233
```

#Estimate Causal Effect-coefficient on treatment in the linear regression we fit on the matched data

```
fit$coefficients["Mother.Incarcerated"]
```

```
## Mother.Incarcerated
##           -0.06504065
```