

# Clean NHANES linked NDI mortality data

*Martin Skarzynski*

*2018-04-19*

```
## read in data processed using sas and clean w/ dplyr ####  
library(here)
```

```
## here() starts at /Users/marskar/gdrive/nhanes  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
## The following objects are masked from 'package:stats':  
##  
##     filter, lag  
## The following objects are masked from 'package:base':  
##  
##     intersect, setdiff, setequal, union
```

```
library(readr)  
read_csv(here("dat/mort.csv")) %>%  
  select(-starts_with("MORTSRCE")) %>%  
  filter(!is.na(UCOD_LEADING) & !is.na(PERMTH_INT)) %>%  
  mutate_at(.vars = vars(-starts_with("PERMTH_"),  
                        -SEQN),  
            .funs = funs(as.factor)) %>%  
  write_rds(here("dat/1-clean-mort.rds"))
```

```
## Parsed with column specification:  
## cols(  
##   SEQN = col_integer(),  
##   ELIGSTAT = col_character(),  
##   MORTSTAT = col_character(),  
##   CAUSEAVL = col_character(),  
##   UCOD_LEADING = col_character(),  
##   DIABETES = col_character(),  
##   HYPERTEN = col_character(),  
##   PERMTH_INT = col_integer(),  
##   PERMTH_EXM = col_integer(),  
##   MORTSRCE_NDI = col_character(),  
##   MORTSRCE_CMS = col_character(),  
##   MORTSRCE_SSA = col_character(),  
##   MORTSRCE_DC = col_character(),  
##   MORTSRCE_DCL = col_character()  
## )
```

```
#names(mort)  
#length(mort)  
#glimpse(mort)
```

```
#any(is.na(mort$PERMTH_INT))
```

```
#all(is.na((select(mort, contains('WGT')))))  
#all(is.na((select(mort, contains('MORTSRCE_DCL')))))  
#unique(select(mort, -contains('MORTSRCE')))  
  
#library(purrr, help)  
#map(mort, nlevels)  
#nlevels(mort)
```