

Volleyball Stat Help

Ethan Trout

2023-04-12

Reading the Data in

```
RACED <- read_csv("D:/DeskingTop/MWCCS Public Data Sets (PDS)/WIHS PDS/CSV/race.csv",
                  show_col_types = FALSE)
HIVD <- read_csv("D:/DeskingTop/MWCCS Public Data Sets (PDS)/WIHS PDS/CSV/hivstat.csv",
                 show_col_types = FALSE)

f20D <- read_csv("D:/DeskingTop/MWCCS Public Data Sets (PDS)/WIHS PDS/CSV/f20.csv",
                 show_col_types = FALSE)
f26D <- read_csv("D:/DeskingTop/MWCCS Public Data Sets (PDS)/WIHS PDS/CSV/f26.csv",
                 show_col_types = FALSE)
f25D <- read_csv("D:/DeskingTop/MWCCS Public Data Sets (PDS)/WIHS PDS/CSV/f25c.csv",
                 show_col_types = FALSE)
f24D <- read_csv("D:/DeskingTop/MWCCS Public Data Sets (PDS)/WIHS PDS/CSV/f24beh.csv",
                 show_col_types = FALSE)
```

Finding Which ID's are in NC, GA, FL, and are African American.

From race.csv

```
## All possible values that matched NC GA and FL are put into vectors
Florida <- c("Florida","FLORIDA","fl","Fl","FL")
NC <- c("nc","NC","North Carolina","NORTH CAROLINA")
Georgia <- c("ga","GA","georgia","Georgia","GEORGIA","Georgia (GA)","Florida",
            "FLORIDA","fl","Fl","FL")
FNCG <- c("ga","GA","georgia","Georgia","GEORGIA","Georgia (GA)","Florida",
          "FLORIDA","fl","Fl","FL","nc","NC","North Carolina","NORTH CAROLINA")

## Subsetting the race data
RACE =
  RACED %>% mutate(STATE = if_else(STEBER %in% Florida,"FLORIDA",
    if_else(STEBER %in% NC,"NORTH CAROLINA",
      if_else(STEBER %in% Georgia,"GEORGIA","NA")))) %>%
  filter(IDTEER > 2012)%>%
  filter(AFAMER == 1) %>%
  mutate(CHECK = if_else(STEBER %in% FNCG,"1","0")) %>%
  filter(CHECK == 1) %>%
  select(STATE,STEBER,everything(),CHECK)

## Creating a Checking Index
CHECK = RACE %>% select(STATE,CASEID,AFAMER,IDTEER)
```

Finding which IDS are HIV Positive and merging it with IDs from previous

From hivstat.csv

```
#Subsetting HIV Data
IDS <- CHECK$CASEID
HIV = HIVD %>%
  filter(STATUS == 1) %>%
  select(CASEID,STATUS) %>%
  rename("HIVSTATUS" = STATUS) %>%
  filter(CASEID %in% IDS)
#Creating List of all IDS that are African American, HIV positive, and in correct states
IDS <- HIV$CASEID
```

Social Support Data Appendix B SOCIAL SUPPORT

```
## Creating the Social Support Data
SOCSUPP = f25D %>%
  select(CASEID,VISIT,IDTEEC,LCMPEC,LFTOEC,ISOLEC,SLSTEC,SINFEC,SADVEC,SCNFEC,
        SAWTEC,SSHREC,STRNEC,SUNDEC,SHLPEC,STAKEC,SPRPEC,SSHPEC,SCAREC,
        SPLCEC,SMNYEC) %>%
  filter(IDTEEC > 2012) %>%
  filter(CASEID %in% IDS)
```

Depression Data Appendix A Depression Scale and creating Depression Variables

```
## Creating the Depression Scale Data
DPRSS = f26D %>%
  select(CASEID,VISIT,IDTEPS,BTHRPS,NOAPPS,WBLUPS,GOODPS,CONCPS,DEPRPS,EFFTPS,FUTRPS,
        FAILPS,FEARPS,SLEPPS,WHYPYS,TALKPS,LONEPS,UNFRPS,
        ENJYPS,CRY_PS,SAD_PS,DISLPS,MOTVPS) %>%
  filter(IDTEPS > 2012)

## Changing the weight of needed variables
DPRSS =
  DPRSS %>%
  mutate(WEIGHTGOODPS =
    if_else(GOODPS == 1,"4",
      if_else(GOODPS == 2, "3",
        if_else(GOODPS == 3, "2",
          if_else(GOODPS == 4,"1",
            if_else(GOODPS < 0,"-1","NA")))))) %>%
  mutate(WEIGHTENJYPS =
    if_else(ENJYPS == 1,"4",
      if_else(ENJYPS == 2, "3",
        if_else(ENJYPS == 3, "2",
          if_else(ENJYPS == 4,"1",
            if_else(ENJYPS < 0,"-1","NA")))))) %>%
  mutate(WEIGHTFUTRPS =
    if_else(FUTRPS == 1,"4",
      if_else(FUTRPS == 2, "3",
        if_else(FUTRPS == 3, "2",
```

```

                                if_else(FUTRPS == 4, "1",
                                          if_else(FUTRPS < 0, "-1", "NA")))))) %>%
mutate(WEIGHTWHPYPS =
  if_else(WHPYPS == 1, "4",
    if_else(WHPYPS == 2, "3",
      if_else(WHPYPS == 3, "2",
        if_else(WHPYPS == 4, "1",
          if_else(WHPYPS < 0, "-1", "NA")))))) %>%

select(CASEID, VISIT, IDTEPS, WEIGHTGOODPS,
  WEIGHTENJYPS, WEIGHTWHPYPS, WEIGHTFUTRPS, everything()) %>%
select(-GOODPS, -ENJYPS, -WHPYPS, -FUTRPS)

## Creating Variables Specified in the Depression Questionnaire
DPRSS = DPRSS %>%
mutate(SCORETEMP = as.numeric(WEIGHTGOODPS) + as.numeric(WEIGHTENJYPS) +
  as.numeric(BTHRPS) + as.numeric(NOAPPS) + as.numeric(WBLUPS) +
  as.numeric(CONCPS) + as.numeric(DEPRPS) + as.numeric(EFFTPS) +
  as.numeric(WEIGHTFUTRPS) + as.numeric(FAILPS) + as.numeric(FEARPS) +
  as.numeric(SLEPPS) + as.numeric(WEIGHTWHPYPS) + as.numeric(TALKPS) +
  as.numeric(LONEPS) + as.numeric(UNFRPS) + as.numeric(CRY_PS) +
  as.numeric(SAD_PS) + as.numeric(DISLPS) + as.numeric(MOTVPS)) %>%
mutate(SCORE = SCORETEMP - 20) %>%
select(everything(), -SCORETEMP) %>%
mutate(DEPRESSED = if_else(SCORE > 15, "1", "0")) %>%
select(CASEID, VISIT, DEPRESSED, SCORE, everything()) %>%
filter(CASEID %in% IDS)

```

Sex Behavior Data Appendix C. D1 - D7 from f20

```

## Creating the Sex Behavior Data
SEXBHVLFT =
  f20D %>%
  select(CASEID, VISIT, IDTENR, MALSNR,
    SXAGNR, MINJNR, MHEMNR, MHIVNR, MMANNR, MSX5NR) %>%
  filter(IDTENR > 2012) %>%
  filter(CASEID %in% IDS)

```

Sex Behavior Data for last 6 months from f24

```

SEXBHV6MO =
  f24D %>%
  select(CASEID, VISIT, IDTEBH, MSX6BH, MS6NBH, VAGSBH, MVCDBH, PORLBH, MPCDBH, MPOSBH, RORLBH, MODDBH,
    ANALBH, MACDBH, PSSTBH, HPOSBH, PJALBH, YSSTBH) %>%
  filter(IDTEBH > 2012) %>%
  filter(CASEID %in% IDS)

```

ISSUES PRESENTED

Mismatched Case IDs

When trying to merge data sets together, there must be the same CASEID's from each data set. This becomes a problem as data sets during different years contain different patients. I.e. the social support data does not contain any of the patients that the other data sets include during 2013. Another problem arises in 2015 where the sexual behavior data set only includes 4 rows of patient data. 2014 has a sufficient number of patients in each data set, which is why it is included.

Multiple Visits in one year

Another problem that arose was when a single patient had multiple VISITs in a single year. To stop from a single CASEID taking up multiple lines, patients are limited to their latest visit from every year.

Files Included

The 2014.csv file contains the data from all four forms for 38 different patients.

Creating and outputting the file

```
# List of IDs that have data in all 4 different forms
ID2014 <- c(1130,1159,1409,1476,1508,1535,1646,1780,1834,1856,2083,2144,2515,
            2761,2894,3357,4110,4127,4635,4812,5514,5652,6130,6197,6577,6705,
            7254,7875,7903,8551,8749,8917,9056,9156,9376,9818,9900)
```

```
DPRSS2014 =
  DPRSS %>%
  filter(IDTEPS == 2014) %>%
  group_by(CASEID) %>% mutate(a = max(VISIT)) %>%
  filter(VISIT == a) %>%
  select(-a,-VISIT) %>%
  filter(CASEID %in% ID2014) %>%
  write_csv("D:/DeskingTop/VolleyballStats/DPRSS2014.csv")
```

```
SEXBHVLFT2014 =
  SEXBHVLFT %>%
  filter(IDTENR == 2014) %>%
  group_by(CASEID) %>% mutate(a = max(VISIT)) %>%
  filter(VISIT == a) %>%
  select(-a,-VISIT) %>%
  filter(CASEID %in% ID2014) %>%
  write_csv("D:/DeskingTop/VolleyballStats/SEXBHVLFT2014.csv")
```

```
SEXBHV6M02014 =
  SEXBHV6M0 %>%
  filter(IDTEBH == 2014) %>%
  group_by(CASEID) %>% mutate(a = max(VISIT)) %>%
  filter(VISIT == a) %>%
  select(-a,-VISIT) %>%
  filter(CASEID %in% ID2014) %>%
  write_csv("D:/DeskingTop/VolleyballStats/SEXBHV6M02014.csv")
```

```

SOCSUPP2014 =
  SOCSUPP %>%
  filter(IDTEEC == 2014) %>%
  group_by(CASEID) %>% mutate(a = max(VISIT)) %>%
  filter(VISIT == a) %>%
  select(-a,-VISIT) %>%
  filter(CASEID %in% ID2014) %>%
  write_csv("D:/DeskingTop/VolleyballStats/SOCSUPP2014.csv")

#Merging the data together
a <- merge(DPRSS2014,SEXBHVLFT2014,by="CASEID")
b <- merge(a,SEXBHV6M02014,by="CASEID")
c <- merge(b,SOCSUPP2014,by="CASEID")

# Increasing Readability
DATA2014 = c %>%
  rename("DPRSS" = IDTEPS, "SEXBHVLFT" = IDTENR, "SEXBHV6M0" = IDTEBH,
         "SOCSUPP" = IDTEEC) %>%
  select(CASEID,DPRSS,everything()) %>%
  write_csv("D:/DeskingTop/VolleyballStats/2014.csv")

```

Formatting

Columns B, Y, AG, AW are all dummy columns containing the year. The name of each of these columns is to indicate where the data from each form starts to help with readability.