nih\_table.Rmd

Peter Higgins

7/20/2018

## Read in data and clean up

ibd <- read\_csv("IBDDatabank2015\_DATA\_2018-06-29\_0901.csv")

## Parsed with column specification:  
## cols(  
## subject\_id = col\_character(),  
## redcap\_event\_name = col\_character(),  
## sex = col\_integer(),  
## marital\_status = col\_integer(),  
## race\_\_\_0 = col\_integer(),  
## race\_\_\_1 = col\_integer(),  
## race\_\_\_2 = col\_integer(),  
## race\_\_\_3 = col\_integer(),  
## race\_\_\_4 = col\_integer(),  
## race\_\_\_5 = col\_integer(),  
## race\_\_\_999 = col\_integer(),  
## ethnicity = col\_integer()  
## )

ibd %>%   
 filter(redcap\_event\_name == "baseline\_visit\_arm\_1") %>%   
 select(sex, race\_\_\_0:ethnicity) %>%   
 mutate(race = case\_when(  
 .$race\_\_\_0 == 1 ~ "White",  
 .$race\_\_\_1 == 1 ~ "Black or African-American",  
 .$race\_\_\_2 == 1 ~ "Asian",  
 .$race\_\_\_3 == 1 ~ "Native Hawaiian or Other Pacific Islander",  
 .$race\_\_\_4 == 1 ~ "American Indian or Alaska Native",  
 .$race\_\_\_5 == 1 ~ "More Than One Race",  
 .$race\_\_\_999 == 1 ~ "Unknown or Not Reported",  
 TRUE ~ "Unknown or Not Reported")) %>%   
 select(sex, race, ethnicity) %>%   
 mutate(ethnic\_cat = case\_when(  
 .$ethnicity == 1 ~ "Hispanic or Latino",  
 .$ethnicity == 0 ~ "Not Hispanic or Latino",  
 TRUE ~ "Unknown or Not Reported Ethnicity")) %>%   
 select(sex, race, ethnic\_cat) %>%   
 mutate(sex2 = case\_when(  
 .$sex == 1 ~ "Male",  
 .$sex == 0 ~ "Female",  
 TRUE ~ "Female")) %>%   
 select(sex2, race, ethnic\_cat) ->  
ibd  
  
ibd\_table <- ibd %>%   
 tabyl(race, sex2, ethnic\_cat) %>%   
 reduce(left\_join, by = "race")  
  
names(ibd\_table) <-(c("race", "female.Hispanic",   
 "male.Hispanic",   
 "female.Not", "male.Not",  
 "female.Unknown", "male.Unknown"))  
  
ibd\_table2 <- gather(ibd\_table, key= sex.eth, value = count, -race) %>%   
 separate(sex.eth, into = c('sex', 'ethnicity')) %>%   
 filter(count != 0)

Now create an empty table with zeroes in all possible 63 cells.

l <- list(race =   
 c("White",  
 "Black or African-American",  
 "Asian",  
 "Native Hawaiian or Other Pacific Islander",  
 "American Indian or Alaska Native",  
 "More Than One Race",  
 "Unknown or Not Reported"),  
 sex = c("male",  
 "female",  
 "Unknown or Not Reported Sex",  
 "male",  
 "female",  
 "male",  
 "female"),  
 ethnicity = c("Hispanic",  
 "Not",  
 "Unknown",  
 "Hispanic",  
 "Not",  
 "Hispanic",  
 "Not"),  
 count = rep(0,7))  
empty\_table <- as\_tibble(l) %>%   
 tidyr::complete(race,   
 nesting(sex), nesting(ethnicity),fill=list(count = 0))

## Now anti-join

# now anti-join ibd\_table2 with empty\_table to get missing rows  
# with no counts in our sample  
complement <- anti\_join(empty\_table, ibd\_table2, by = c('race', 'sex', "ethnicity")) %>%   
 print(n=Inf)

## # A tibble: 46 x 4  
## race sex ethnicity count  
## <chr> <chr> <chr> <dbl>  
## 1 American Indian or Alaska Native female Hispanic 0  
## 2 American Indian or Alaska Native female Not 0  
## 3 American Indian or Alaska Native female Unknown 0  
## 4 American Indian or Alaska Native male Hispanic 0  
## 5 American Indian or Alaska Native male Not 0  
## 6 American Indian or Alaska Native male Unknown 0  
## 7 American Indian or Alaska Native Unknown or N… Hispanic 0  
## 8 American Indian or Alaska Native Unknown or N… Not 0  
## 9 American Indian or Alaska Native Unknown or N… Unknown 0  
## 10 Asian female Hispanic 0  
## 11 Asian female Unknown 0  
## 12 Asian male Unknown 0  
## 13 Asian Unknown or N… Hispanic 0  
## 14 Asian Unknown or N… Not 0  
## 15 Asian Unknown or N… Unknown 0  
## 16 Black or African-American female Hispanic 0  
## 17 Black or African-American female Unknown 0  
## 18 Black or African-American male Hispanic 0  
## 19 Black or African-American male Unknown 0  
## 20 Black or African-American Unknown or N… Hispanic 0  
## 21 Black or African-American Unknown or N… Not 0  
## 22 Black or African-American Unknown or N… Unknown 0  
## 23 More Than One Race female Unknown 0  
## 24 More Than One Race male Hispanic 0  
## 25 More Than One Race male Unknown 0  
## 26 More Than One Race Unknown or N… Hispanic 0  
## 27 More Than One Race Unknown or N… Not 0  
## 28 More Than One Race Unknown or N… Unknown 0  
## 29 Native Hawaiian or Other Pacific Islander female Hispanic 0  
## 30 Native Hawaiian or Other Pacific Islander female Not 0  
## 31 Native Hawaiian or Other Pacific Islander female Unknown 0  
## 32 Native Hawaiian or Other Pacific Islander male Hispanic 0  
## 33 Native Hawaiian or Other Pacific Islander male Not 0  
## 34 Native Hawaiian or Other Pacific Islander male Unknown 0  
## 35 Native Hawaiian or Other Pacific Islander Unknown or N… Hispanic 0  
## 36 Native Hawaiian or Other Pacific Islander Unknown or N… Not 0  
## 37 Native Hawaiian or Other Pacific Islander Unknown or N… Unknown 0  
## 38 Unknown or Not Reported female Hispanic 0  
## 39 Unknown or Not Reported female Not 0  
## 40 Unknown or Not Reported male Hispanic 0  
## 41 Unknown or Not Reported Unknown or N… Hispanic 0  
## 42 Unknown or Not Reported Unknown or N… Not 0  
## 43 Unknown or Not Reported Unknown or N… Unknown 0  
## 44 White Unknown or N… Hispanic 0  
## 45 White Unknown or N… Not 0  
## 46 White Unknown or N… Unknown 0

## Now bind with complement

# now combine complement with ibd\_table2  
full\_table <- bind\_rows(ibd\_table2, complement)

## Set up to Spread

# set up to spread  
ibd\_table <- full\_table %>%   
 unite(col = "eth.sex", c('ethnicity', "sex"), sep=".") %>%   
 spread(key = eth.sex, value = count)

## Now add margin totals and arrange, write xls

# convert race col to rownames to make numbers into a matrix  
m <- as.matrix(ibd\_table[ ,-1])  
rownames(m) <- ibd\_table$race  
ibd\_table2 <- addmargins(m, FUN=c(Total=sum), quiet = T)  
ibd\_table <- rownames\_to\_column(as.data.frame(ibd\_table2), "Racial Categories")  
  
#arrange without total  
ibd\_table3 <- ibd\_table[1:7,] %>% arrange(`Racial Categories`)  
  
ibd\_table <- as.data.frame(rbind(ibd\_table3, ibd\_table[8,]))  
  
write\_xlsx(ibd\_table, here("ibd\_table.xlsx"))

## Now make flextable

#make flextable  
myft <- ibd\_table %>%   
 regulartable() %>%   
 theme\_booktabs() %>%   
 set\_formatter\_type(fmt\_double = "%0.0f") %>%   
 add\_header(`Racial Categories` = "Racial Categories",  
 Hispanic.female = "Hispanic or Latino",  
 Hispanic.male = "Hispanic or Latino",  
 `Hispanic.Unknown or Not Reported Sex` = "Hispanic or Latino",  
 Not.female = "Not Hispanic or Latino",  
 Not.male = "Not Hispanic or Latino",  
 `Not.Unknown or Not Reported Sex` = "Not Hispanic or Latino",  
 Unknown.female = "Unknown/Not Reported",  
 Unknown.male = "Unknown/Not Reported",  
 `Unknown.Unknown or Not Reported Sex` = "Unknown/Not Reported",  
 Total = "Total") %>%   
 add\_header(`Racial Categories` = "Racial Categories",  
 Hispanic.female = "Ethnic Categories Divided by Sex",  
 Hispanic.male = "Ethnic Categories Divided by Sex",  
 `Hispanic.Unknown or Not Reported Sex` = "Ethnic Categories Divided by Sex",  
 Not.female = "Ethnic Categories Divided by Sex",  
 Not.male = "Ethnic Categories Divided by Sex",  
 `Not.Unknown or Not Reported Sex` = "Ethnic Categories Divided by Sex",  
 Unknown.female = "Ethnic Categories Divided by Sex",  
 Unknown.male = "Ethnic Categories Divided by Sex",  
 `Unknown.Unknown or Not Reported Sex` = "Ethnic Categories Divided by Sex",  
 Total = "Total") %>%   
 set\_header\_labels(`Racial Categories` = "Racial Categories",  
 Hispanic.female = "Female",  
 Hispanic.male = "Male",  
 `Hispanic.Unknown or Not Reported Sex` = "Unknown",  
 Not.female = "Female",  
 Not.male = "Male",  
 `Not.Unknown or Not Reported Sex` = "Unknown",  
 Unknown.female = "Female",  
 Unknown.male = "Male",  
 `Unknown.Unknown or Not Reported Sex` = "Unknown",  
 Total = "Total") %>%   
 font(fontname = "Arial") %>%   
 fontsize(size = 10) %>%   
 fontsize(size=12, part="header") %>%   
 merge\_h(part= "header") %>%   
 merge\_v(part= "header") %>%   
 align(align = "center", part = "all") %>%   
 width(j = 1, width = 2.0)  
  
myft

| Racial Categories | Ethnic Categories Divided by Sex | | | | | | | | | Total |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hispanic or Latino | | | Not Hispanic or Latino | | | Unknown/Not Reported | | |
| Female | Male | Unknown | Female | Male | Unknown | Female | Male | Unknown |
| American Indian or Alaska Native | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Asian | 0 | 1 | 0 | 3 | 5 | 0 | 0 | 0 | 0 | 9 |
| Black or African-American | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 2 |
| More Than One Race | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 3 |
| Native Hawaiian or Other Pacific Islander | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Unknown or Not Reported | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 1 | 0 | 4 |
| White | 5 | 1 | 0 | 43 | 46 | 0 | 1 | 3 | 0 | 99 |
| Total | 6 | 2 | 0 | 48 | 54 | 0 | 3 | 4 | 0 | 117 |