participant-demographics

Dan Spakowicz

1/7/2021

```
y <- readxl::read_xlsx(file.path(paths$box,
                                 "participants",
                                 "Be Well Study Schedule.xlsx"),
                       trim_ws = TRUE,
                       sheet = 2)
```

```
## New names:
## * `` -> ...24
## * `` -> ...25
## * `` -> ...26
## * `` -> ...27
## * `` -> ...28
## * ...
```

```
head(y)
```

```
## # A tibble: 6 x 30
##   `Subject ID` Gender `Smoking status`   Age `Race/ethnicity` `how did you he~
##   <chr>        <chr>  <chr>            <dbl> <chr>            <chr>
## 1 HONC60-01    male   former              66 white            LCSC
## 2 HONC60-02    male   Current             58 white            LCSC
## 3 HONC60-03    female former              60 white            LCSC
## 4 HONC60-04    female former              73 decline to say   LCSC
## 5 HONC60-05    female current             59 <NA>             LCSC
## 6 HONC60-06    female current             74 white            LCSC
## # ... with 24 more variables: `V1 S1, S2` <chr>, `V1 St1-9` <chr>, `V1
## #   U1-10` <chr>, `V1 Blood` <chr>, `V2 S1, S2` <chr>, `V2 St1-9` <chr>, `V2
## #   U1-10` <chr>, `V2 Blood` <chr>, `V3 S1, S2` <chr>, `V3 St1-9` <chr>, `V3
## #   U1-10` <chr>, `V3 Blood` <chr>, `V4 S1, S2` <chr>, `V4 St1-9` <chr>, `V4
## #   U1-10` <chr>, `V4 Blood` <chr>, Notes <chr>, ...24 <lgl>, ...25 <lgl>,
## #   ...26 <lgl>, ...27 <lgl>, ...28 <lgl>, ...29 <lgl>, ...30 <chr>
```

```
df <-
  y %>%
  mutate(recruitment.style = if_else(`how did you hear about us?` == "LCSC",
                                     true = "in person",
                                     false = "social media"),
         dropout = if_else(is.na(`V1 Blood`),
                           true = TRUE,
                           false = FALSE),
         across(c("Gender", "Smoking status", "Race/ethnicity"), tolower),
         `Race/ethnicity` = if_else(`Race/ethnicity` == "white",
                                    true = "caucasian",
                                    false = `Race/ethnicity`),
```

```
        `Race/ethnicity` = if_else(`Race/ethnicity` == "white caucasian",
                                   true = "caucasian",
                                   false = `Race/ethnicity`),
        `Race/ethnicity` = if_else(`Race/ethnicity` == "black",
                                   true = "african american",
                                   false = `Race/ethnicity`),
        `Race/ethnicity` = if_else(`Race/ethnicity` == "african",
                                   true = "african american",
                                   false = `Race/ethnicity`),
        across(c("Gender", "Smoking status", "Race/ethnicity", "dropout"),
               as.factor)
        )


head(df)
```

```
## # A tibble: 6 x 32
##   `Subject ID` Gender `Smoking status`   Age `Race/ethnicity` `how did you he~
##   <chr>        <fct>  <fct>            <dbl> <fct>            <chr>
## 1 HONC60-01    male   former              66 caucasian        LCSC
## 2 HONC60-02    male   current             58 caucasian        LCSC
## 3 HONC60-03    female former              60 caucasian        LCSC
## 4 HONC60-04    female former              73 decline to say   LCSC
## 5 HONC60-05    female current             59 <NA>             LCSC
## 6 HONC60-06    female current             74 caucasian        LCSC
## # ... with 26 more variables: `V1 S1, S2` <chr>, `V1 St1-9` <chr>, `V1
## #   U1-10` <chr>, `V1 Blood` <chr>, `V2 S1, S2` <chr>, `V2 St1-9` <chr>, `V2
## #   U1-10` <chr>, `V2 Blood` <chr>, `V3 S1, S2` <chr>, `V3 St1-9` <chr>, `V3
## #   U1-10` <chr>, `V3 Blood` <chr>, `V4 S1, S2` <chr>, `V4 St1-9` <chr>, `V4
## #   U1-10` <chr>, `V4 Blood` <chr>, Notes <chr>, ...24 <lgl>, ...25 <lgl>,
## #   ...26 <lgl>, ...27 <lgl>, ...28 <lgl>, ...29 <lgl>, ...30 <chr>,
## #   recruitment.style <chr>, dropout <fct>
```

```
table(df$`how did you hear about us?`)
```

```
##
##    Craigslist   Facebook ad            FB    FB- sister          LCSC
##             1             1            17             1            52
## ResearchMatch            RM
##             1            10
```

```
table(df$recruitment.style)
```

```
##
##    in person social media
##           52           31
```

```
table(df$recruitment.style, df$dropout)
```

```
##
##                FALSE TRUE
##   in person       15   37
##   social media    21   10
```

```
fisher.test(df$recruitment.style, df$dropout)
```

```
##
```

```
##  Fisher's Exact Test for Count Data
##
## data:  df$recruitment.style and df$dropout
## p-value = 0.0006754
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.06548566 0.55706900
## sample estimates:
## odds ratio
##  0.1974018
```

Create a table of the results stratified by `recruitment.style`.

```
str(df$Gender)
```

```
##  Factor w/ 2 levels "female","male": 2 2 1 1 1 1 2 1 2 1 ...
```

```
listVars <- c("Age", "Gender", "Smoking status", "Race/ethnicity", "dropout")
catVars <- c("Gender", "Smoking status", "Race/ethnicity", "dropout")
```

```
table1 <- CreateTableOne(vars = listVars,
                         data = df,
                         factorVars = catVars,
                         strata = "recruitment.style")
table1
```

```
##                               Stratified by recruitment.style
##                                in person     social media  p      test
##   n                               52            31
##   Age (mean (SD))              67.26 (5.77)  63.68 (6.77)   0.017
##   Gender = male (%)               18 (40.0)     13 (41.9)   1.000
##   Smoking status = former (%)     30 (68.2)     16 (51.6)   0.226
##   Race/ethnicity (%)                                        0.408
##      african american              4 (10.8)      0 ( 0.0)
##      caucasian                    25 (67.6)     11 (73.3)
##      decline to say                8 (21.6)      4 (26.7)
##   dropout = TRUE (%)              37 (71.2)     10 (32.3)   0.001
```

```
table1 <- print(table1, printToggle = FALSE)

# flextable

table2 <- table1 %>%

  as.data.frame() %>%

  rownames_to_column("Variables")%>%

  flextable() %>%

  autofit(add_w = 0.2) %>%

  align(align = "left") %>%

  align(align = "left", part = "header") %>%

  add_header_lines("Table 1: Demographics stratified by recruitment style.")
```

```
table2 <- print(table2, printToggle = FALSE)
```

```
## a flextable object.
## col_keys: `Variables`, `in person`, `social media`, `p`, `test`
## header has 2 row(s)
## body has 9 row(s)
## original dataset sample:
##                      Variables     in person  social media      p test
## 1                            n            52            31
## 2              Age (mean (SD))  67.26 (5.77)  63.68 (6.77)  0.017
## 3           Gender = male (%)    18 (40.0)     13 (41.9)   1.000
## 4 Smoking status = former (%)    30 (68.2)     16 (51.6)   0.226
## 5          Race/ethnicity (%)                               0.408
```

```
tab3Mat <- print(table2, quote = FALSE, noSpaces = TRUE, printToggle = FALSE)
```

```
## NULL
```

```
## Save to a CSV file
write.csv(tab3Mat, file = "../data/demographics-table.csv")
```