Supplement 1

Demographics

Peter Kamerman

Last knitted: 03 March 2021

Contents

---

1 Import and check data

```
# Import
demo <- read_rds('data-cleaned/data-demographics.rds')%>%
    # Fix site names:
    ## DD = Site 1, NM = Site 2, RESI = Site 3, RP = Site 4, STIG = Site 5
    mutate(Site = case_when(
        Site == 'DD' ~ 'Site 1',
        Site == 'NM' ~ 'Site 2',
        Site == 'RESI' ~ 'Site 3',
        Site == 'RP' ~ 'Site 4',
        Site == 'STIG' ~ 'Site 5'
    ))

# Check
## Demographics
dim(demo)
```

```
## [1] 596   8
```

```
names(demo)
```

```
## [1] "ID"                "Site"              "Sex"
## [4] "Age"               "Employment_status" "CD4_recent"
## [7] "ART_currently"     "Education"
```

```r
glimpse(demo)
```

```
## Rows: 596
## Columns: 8
## $ ID                <chr> "RPB73", "RPB74", "RPB75", "RPB76", "RPB77", "RPB...
## $ Site              <chr> "Site 4", "Site 4", "Site 4", "Site 4", "Site 4",...
## $ Sex               <chr> "Female", "Female", "Female", "Female", "Female",...
## $ Age               <dbl> 36, 27, 39, 36, 31, 32, 28, 37, 31, 25, 31, 24, 3...
## $ Employment_status <chr> "Other", "Unemployed", "Other", "Unemployed", "Un...
## $ CD4_recent        <dbl> 391, 571, 591, 207, 126, 225, 543, 410, 74, 212, ...
## $ ART_currently     <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", ...
## $ Education         <chr> "Tertiary", "Secondary", "Secondary", "Primary", ...
```

2 Basic descriptive statistics

2.1 Full cohort

2.1.1 Descriptives

```r
demo %>%
    # Remove ID column
    select(-ID, -Site) %>%
    # Convert character to factor
    mutate_if(is.character, factor) %>%
    my_skim()
```

Table 1: Data summary

Name	Piped data
Number of rows	596
Number of columns	6
Column type frequency:	
factor	4
numeric	2
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	n_unique	top_counts
Sex	0	1.00	2	Fem: 481, Mal: 115
Employment_status	49	0.92	4	Une: 330, Ful: 131, Par: 52, Oth: 34
ART_currently	5	0.99	2	Yes: 460, No: 131
Education	37	0.94	3	Sec: 395, Pri: 99, Ter: 65

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	8	0.99	37.28	9.06	19	31	36	42	76
CD4_recent	99	0.83	320.71	238.92	1	155	261	432	1232

2.2 By study site

2.2.1 Sample size

```
demo %>%
    # Select ID
    select(Site) %>%
    # Group by study site
    group_by(Site) %>%
    summarise(Count = n()) %>%
    kable(caption = 'Sample size by study site')
```

Table 4: Sample size by study site

Site	Count
Site 1	60
Site 2	239
Site 3	99
Site 4	148
Site 5	50

2.2.2 Descriptives

```
demo %>%
    # Convert character to factor
    mutate_if(is.character, factor) %>%
    # Remove ID
    select(-ID) %>%
    # Group by site
    group_by(Site) %>%
    my_skim()
```

Table 5: Data summary

Name	Piped data
Number of rows	596
Number of columns	7
Column type frequency:	
factor	4
numeric	2
Group variables	Site

Variable type: factor

skim_variable	Site	n_missing	complete_rate	n_unique	top_counts
Sex	Site 1	0	1.00	2	Fem: 39, Mal: 21
Sex	Site 2	0	1.00	2	Fem: 184, Mal: 55
Sex	Site 3	0	1.00	2	Fem: 66, Mal: 33
Sex	Site 4	0	1.00	1	Fem: 148, Mal: 0
Sex	Site 5	0	1.00	2	Fem: 44, Mal: 6
Employment_status	Site 1	0	1.00	4	Une: 36, Ful: 17, Par: 6, Oth: 1
Employment_status	Site 2	49	0.79	3	Une: 126, Ful: 55, Par: 9, Oth: 0
Employment_status	Site 3	0	1.00	4	Une: 47, Ful: 35, Par: 16, Oth: 1
Employment_status	Site 4	0	1.00	4	Une: 99, Oth: 31, Par: 14, Ful: 4
Employment_status	Site 5	0	1.00	4	Une: 22, Ful: 20, Par: 7, Oth: 1
ART_currently	Site 1	0	1.00	2	Yes: 58, No: 2
ART_currently	Site 2	3	0.99	2	Yes: 130, No: 106
ART_currently	Site 3	1	0.99	2	Yes: 95, No: 3
ART_currently	Site 4	0	1.00	2	Yes: 128, No: 20
ART_currently	Site 5	1	0.98	1	Yes: 49, No: 0
Education	Site 1	2	0.97	3	Sec: 40, Pri: 9, Ter: 9
Education	Site 2	27	0.89	3	Sec: 160, Pri: 44, Ter: 8
Education	Site 3	4	0.96	3	Sec: 61, Pri: 19, Ter: 15
Education	Site 4	2	0.99	3	Sec: 108, Ter: 20, Pri: 18
Education	Site 5	2	0.96	3	Sec: 26, Ter: 13, Pri: 9

Variable type: numeric

skim_variable	Site	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	Site 1	0	1.00	38.63	5.87	27	34.00	38.5	42.00	58
Age	Site 2	6	0.97	36.73	8.54	20	31.00	36.0	41.00	76
Age	Site 3	0	1.00	43.40	9.32	19	35.00	42.0	49.00	66
Age	Site 4	2	0.99	30.82	4.72	20	27.00	31.0	34.00	40
Age	Site 5	0	1.00	44.90	9.67	25	37.25	44.5	52.50	65
CD4_recent	Site 1	18	0.70	428.60	255.93	16	256.00	374.5	577.25	1095
CD4_recent	Site 2	37	0.85	222.10	180.15	1	103.75	185.5	308.75	963
CD4_recent	Site 3	16	0.84	452.76	278.07	11	256.50	405.0	651.00	1232
CD4_recent	Site 4	4	0.97	322.19	205.43	26	174.00	270.0	415.00	1040
CD4_recent	Site 5	24	0.52	482.88	292.90	50	291.00	434.5	582.00	1138

3 Session information

```
sessionInfo()
```

```
## R version 3.6.3 (2020-02-29)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 10 (buster)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib/x86_64-linux-gnu/libopenblasp-r0.3.5.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=C
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.28      skimr_2.1.1     forcats_0.5.0   stringr_1.4.0
##  [5] dplyr_0.8.5     purrr_0.3.4     readr_1.3.1     tidyr_1.0.2
##  [9] tibble_3.0.1    ggplot2_3.3.0   tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.0.0 xfun_0.13        repr_1.1.0       haven_2.2.0
##  [5] lattice_0.20-38  colorspace_1.4-1 vctrs_0.2.4      generics_0.0.2
##  [9] htmltools_0.4.0  base64enc_0.1-3  yaml_2.2.1       utf8_1.1.4
## [13] rlang_0.4.5      pillar_1.4.3     glue_1.4.0       withr_2.2.0
## [17] DBI_1.1.0        dbplyr_1.4.3     modelr_0.1.6     readxl_1.3.1
## [21] lifecycle_0.2.0  munsell_0.5.0    gtable_0.3.0     cellranger_1.1.0
## [25] rvest_0.3.5      evaluate_0.14    fansi_0.4.1      highr_0.8
## [29] broom_0.5.6      Rcpp_1.0.4.6     scales_1.1.0     backports_1.1.6
## [33] jsonlite_1.6.1   fs_1.4.1         hms_0.5.3        digest_0.6.25
## [37] stringi_1.4.6    grid_3.6.3       cli_2.0.2        tools_3.6.3
## [41] magrittr_1.5     crayon_1.3.4     pkgconfig_2.0.3  ellipsis_0.3.0
## [45] xml2_1.3.2       reprex_0.3.0     lubridate_1.7.8  assertthat_0.2.1
## [49] rmarkdown_2.1    httr_1.4.1       rstudioapi_0.11  R6_2.4.1
## [53] nlme_3.1-144     compiler_3.6.3
```