Supplement 1
Demopgraphics

Peter Kamerman

Last knitted: 27 April 2020

Contents

Note: sites RBP and RISI removed at data cleaning stage because n < 10.

Import and check data

```r
# Import
demo <- read_rds('data-cleaned/data-demographics.rds')

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

```
## [1] 595    7
```

```r
names(demo)
```

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

```r
glimpse(demo)
```

```
## Rows: 595
## Columns: 7
## $ ID                <chr> "RPB73", "RPB74", "RPB75", "RPB76", "RPB77", "RPB...
```

```
## $ 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", ...
```

Basic descriptive statistics

Full cohort

Descriptives

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

Table 1: Data summary

Name	Piped data
Number of rows	595
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: 479, Mal: 116
Employment_status	49	0.92	4	Une: 329, Ful: 131, Par: 52, Oth: 34
ART_currently	5	0.99	2	Yes: 459, No: 131
Education	37	0.94	3	Sec: 393, Pri: 100, Ter: 65

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	8	0.99	37.31	9.06	19	31	36.0	42.00	76
CD4_recent	99	0.83	337.17	400.32	1	155	261.5	437.75	7300

By site

Sample size

```r
demo %>%
    # Select ID
    select(ID) %>%
    # Clean-up ID column to keep only study site information
    mutate(ID = str_remove(ID,
                            pattern = '[0-9][0-9]?[0-9]?[0-9]?')) %>%
    # Group data
    group_by(ID) %>%
    summarise(Count = n()) %>%
    kable(caption = 'Sample size by study site')
```

Table 4: Sample size by study site

ID	Count
DD	60
NM	240
RESI	99
RPA	20
RPB	102
RPC	14
RPD	10
STIG	50

Descriptives

```r
demo %>%
    # Clean-up ID column to keep only study site information
    mutate(ID = str_remove(ID,
                            pattern = '[0-9][0-9]?[0-9]?[0-9]?')) %>%
    # Convert character to factor
    mutate_if(is.character, factor) %>%
    # Group by site
    group_by(ID) %>%
    skim()
```

Table 5: Data summary

Name	Piped data
Number of rows	595
Number of columns	7
Column type frequency:	
factor	4
numeric	2
Group variables	ID

Variable type: factor

skim_variable	ID	n_missing	complete_rate	n_unique	top_counts
Sex	DD	0	1.00	2	Fem: 39, Mal: 21
Sex	NM	0	1.00	2	Fem: 185, Mal: 55
Sex	RESI	0	1.00	2	Fem: 65, Mal: 34
Sex	RPA	0	1.00	1	Fem: 20, Mal: 0
Sex	RPB	0	1.00	1	Fem: 102, Mal: 0
Sex	RPC	0	1.00	1	Fem: 14, Mal: 0
Sex	RPD	0	1.00	1	Fem: 10, Mal: 0
Sex	STIG	0	1.00	2	Fem: 44, Mal: 6
Employment_status	DD	0	1.00	4	Une: 36, Ful: 17, Par: 6, Oth: 1
Employment_status	NM	49	0.80	4	Une: 126, Ful: 55, Par: 9, Oth: 1
Employment_status	RESI	0	1.00	4	Une: 47, Ful: 35, Par: 16, Oth: 1
Employment_status	RPA	0	1.00	3	Une: 12, Oth: 4, Par: 4, Ful: 0
Employment_status	RPB	0	1.00	4	Une: 72, Oth: 20, Par: 7, Ful: 3
Employment_status	RPC	0	1.00	4	Une: 8, Oth: 4, Ful: 1, Par: 1
Employment_status	RPD	0	1.00	3	Une: 6, Oth: 2, Par: 2, Ful: 0
Employment_status	STIG	0	1.00	4	Une: 22, Ful: 20, Par: 7, Oth: 1
ART_currently	DD	0	1.00	2	Yes: 58, No: 2
ART_currently	NM	3	0.99	2	Yes: 131, No: 106
ART_currently	RESI	1	0.99	2	Yes: 95, No: 3
ART_currently	RPA	0	1.00	2	Yes: 18, No: 2
ART_currently	RPB	0	1.00	1	Yes: 102, No: 0
ART_currently	RPC	0	1.00	2	No: 12, Yes: 2
ART_currently	RPD	0	1.00	2	No: 6, Yes: 4
ART_currently	STIG	1	0.98	1	Yes: 49, No: 0
Education	DD	2	0.97	3	Sec: 40, Pri: 9, Ter: 9
Education	NM	27	0.89	3	Sec: 160, Pri: 45, Ter: 8
Education	RESI	4	0.96	3	Sec: 61, Pri: 19, Ter: 15
Education	RPA	0	1.00	3	Sec: 14, Pri: 5, Ter: 1
Education	RPB	2	0.98	3	Sec: 74, Ter: 15, Pri: 11
Education	RPC	0	1.00	3	Sec: 11, Ter: 2, Pri: 1
Education	RPD	0	1.00	3	Sec: 7, Ter: 2, Pri: 1
Education	STIG	2	0.96	3	Sec: 26, Ter: 13, Pri: 9

Variable type: numeric

skim_variable	ID	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	DD	0	1.00	38.63	5.87	27	34.00	38.5	42.00	58
Age	NM	6	0.98	36.79	8.56	20	31.00	36.0	41.00	76
Age	RESI	0	1.00	43.33	9.36	19	35.00	42.0	49.00	66
Age	RPA	0	1.00	30.35	3.70	22	28.75	31.0	33.00	36
Age	RPB	1	0.99	31.18	4.79	20	28.00	31.0	35.00	40
Age	RPC	1	0.93	29.77	4.62	23	27.00	29.0	32.00	40
Age	RPD	0	1.00	29.60	5.42	21	25.50	30.0	33.25	37
Age	STIG	0	1.00	44.90	9.67	25	37.25	44.5	52.50	65
CD4_recent	DD	18	0.70	428.60	255.93	16	256.00	374.5	577.25	1095
CD4_recent	NM	37	0.85	230.86	218.78	1	104.50	187.0	309.50	2000
CD4_recent	RESI	16	0.84	529.69	800.31	11	256.50	405.0	651.00	7300
CD4_recent	RPA	0	1.00	163.40	84.13	26	106.25	162.5	220.25	313
CD4_recent	RPB	3	0.97	385.16	214.73	50	224.50	327.0	501.50	1040

skim_variable	ID	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
CD4_recent	RPC	1	0.93	197.08	61.50	101	166.00	175.0	246.00	305
CD4_recent	RPD	0	1.00	189.20	54.84	133	152.50	173.0	202.75	295
CD4_recent	STIG	24	0.52	482.88	292.90	50	291.00	434.5	582.00	1138

Session information

```r
sessionInfo()
```

```
## R version 3.6.3 (2020-02-29)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Catalina 10.15.4
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.28        skimr_2.1         forcats_0.5.0      stringr_1.4.0
##  [5] dplyr_0.8.5       purrr_0.3.3       readr_1.3.1        tidyr_1.0.2
##  [9] tibble_3.0.0      ggplot2_3.3.0.9000 tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.0.0 xfun_0.12         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     withr_2.1.2       glue_1.3.2
## [17] DBI_1.1.0        dbplyr_1.4.2     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_0.5.5      Rcpp_1.0.4       backports_1.1.5   scales_1.1.0
## [33] jsonlite_1.6.1   fs_1.3.1         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_1.3.0       reprex_0.3.0     lubridate_1.7.4   assertthat_0.2.1
## [49] rmarkdown_2.1    httr_1.4.1       rstudioapi_0.11   R6_2.4.1
## [53] nlme_3.1-145     compiler_3.6.3
```