

Replicate NZ OA Study

Libraries

```
library(tidyverse)
```

```
## Warning: package 'tibble' was built under R version 4.0.2
```

Load Supplement

```
nz_oa <- readxl::read_xlsx("peerj-51535-NZ_university_articles_2017_White_et_al_PeerJ.xlsx")
```

Print ten first rows

```
nz_oa
```

```
## # A tibble: 12,930 x 37
##   DOI Evidence Licence `OA Status` Title Authors `Author count`
##   <chr> <chr>    <chr>    <chr>    <chr> <chr>          <dbl>
## 1 10.1~ open (v~ <NA>    bronze    Effe~ "Jayne~          16
## 2 10.1~ oa repo~ <NA>    green     Effe~ "Scrag~           9
## 3 10.1~ <NA>    <NA>    closed    Vita~ "Scrag~           2
## 4 10.1~ open (v~ <NA>    bronze    Asso~ "Mckin~          15
## 5 10.1~ open (v~ <NA>    bronze    Pate~ "Under~           7
## 6 10.1~ <NA>    <NA>    closed    Asse~ "Under~           3
## 7 10.1~ <NA>    <NA>    closed    Effe~ "Merry~           3
## 8 10.1~ <NA>    <NA>    closed    Time~ "Sammo~           2
## 9 10.1~ open (v~ <NA>    bronze    Earl~ "Gon\\~           8
## 10 10.1~ open (v~ <NA>    bronze    Pneu~ "Walte~           4
## # ... with 12,920 more rows, and 30 more variables: `Author count`>20` <chr>,
## #   Journal <chr>, Year <dbl>, Publisher <chr>, `Is OA` <lgl>, Genre <chr>, `OA
## #   Journal` <lgl>, Version <chr>, `Host of best version` <chr>, `Green version
## #   available` <chr>, Repositories <chr>, `Number of repositories` <dbl>, `In
## #   DOAJ` <lgl>, ISSNs <chr>, `Archive accepted manuscript` <chr>, `Archive
## #   published version` <chr>, `Sherpa/Romeo colour` <chr>, `Sherpa/Romeo
## #   Licence` <chr>, `Crossref citations` <dbl>, `Free text url` <chr>, `APC
## #   charged in DOAJ` <chr>, `DOAJ Currency` <chr>, `DOAJ APC` <dbl>, `Flourish
## #   APC (USD)` <dbl>, `Publisher Currency` <chr>, `Publisher APC` <dbl>, `USD
## #   APC` <dbl>, Funders <chr>, Subjects <chr>, `NZ Reprint author` <chr>
```

In total, the supplement contains 12930 articles with DOI.

Selection criteria

From the manuscript (248-255):

DOIs for 2017 journal articles were gathered from each university, then amalgamated into a single file of more than 12,600 journal articles. If there was a local corresponding author at any

university for a given article then it was designated as having a New Zealand corresponding author. During the course of the project we found that a small percentage of articles with large numbers of authors and large numbers of citations skewed the data so articles with more than 20 authors were excluded on the grounds that they had a tenuous connection to the New Zealand University that had submitted the DOI. This reduced the sample size to 12,016. These were fed into The Program.

Accordingly, the data is subsetting using the following criteria:

- has a DOI
- publication year 2017
- journal articles
- author count ≤ 20

```
nz_oa_df <- nz_oa %>%
  filter(!is.na(DOI), Year == 2017, Genre == "journal-article", `Author count` <= 20) %>%
  distinct(DOI, .keep_all = TRUE)
# print first ten rows
nz_oa_df

## # A tibble: 11,613 x 37
##   DOI      Evidence Licence `OA Status` Title Authors `Author count`
##   <chr> <chr>      <chr>   <chr>      <chr> <chr>          <dbl>
## 1 10.1~ open (v~ <NA>   bronze    Effe~ "Jayne~          16
## 2 10.1~ oa repo~ <NA>   green     Effe~ "Scrag~           9
## 3 10.1~ <NA>      <NA>   closed    Vita~ "Scrag~           2
## 4 10.1~ open (v~ <NA>   bronze    Asso~ "Mckin~          15
## 5 10.1~ open (v~ <NA>   bronze    Pate~ "Under~           7
## 6 10.1~ <NA>      <NA>   closed    Asse~ "Under~           3
## 7 10.1~ <NA>      <NA>   closed    Effe~ "Merry~           3
## 8 10.1~ <NA>      <NA>   closed    Time~ "Sammo~           2
## 9 10.1~ open (v~ <NA>   bronze    Elec~ "Dodd,~           4
## 10 10.1~ open (v~ <NA>   bronze    Oral~ "Farqu~           5
## # ... with 11,603 more rows, and 30 more variables: `Author count`>20` <chr>,
## #   Journal <chr>, Year <dbl>, Publisher <chr>, `Is OA` <lgl>, Genre <chr>, `OA
## #   Journal` <lgl>, Version <chr>, `Host of best version` <chr>, `Green version
## #   available` <chr>, Repositories <chr>, `Number of repositories` <dbl>, `In
## #   DOAJ` <lgl>, ISSNns <chr>, `Archive accepted manuscript` <chr>, `Archive
## #   published version` <chr>, `Sherpa/Romeo colour` <chr>, `Sherpa/Romeo
## #   Licence` <chr>, `Crossref citations` <dbl>, `Free text url` <chr>, `APC
## #   charged in DOAJ` <chr>, `DOAJ Currency` <chr>, `DOAJ APC` <dbl>, `Flourish
## #   APC (USD)` <dbl>, `Publisher Currency` <chr>, `Publisher APC` <dbl>, `USD
## #   APC` <dbl>, Funders <chr>, Subjects <chr>, `NZ Reprint author` <chr>
```

In total, the dataset contains 11613 articles with DOI. These are 403 articles less than reported in the manuscript.

Table 3: Proportion of articles with at least one New Zealand author which were open

```
nz_oa_df %>%
  group_by(`Is OA`) %>%
  summarise(articles = n(),
            citations = mean(`Crossref citations`, na.rm = TRUE)) %>%
  mutate(prop = articles / sum(articles))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
## # A tibble: 2 x 4
##   `Is OA` articles citations prop
##   <lgl>      <int>      <dbl> <dbl>
## 1 FALSE      6803        4.62 0.586
## 2 TRUE       4810        5.91 0.414
```

Table 4: Proportion of articles with New Zealand corresponding author which were open

```
nz_oa_df %>%
  filter(`NZ Reprint author` == "Yes") %>%
  group_by(`Is OA`) %>%
  summarise(articles = n(),
            citations = mean(`Crossref citations`, na.rm = TRUE)) %>%
  mutate(prop = articles / sum(articles))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
## # A tibble: 2 x 4
##   `Is OA` articles citations prop
##   <lgl>      <int>      <dbl> <dbl>
## 1 FALSE      3350        3.77 0.657
## 2 TRUE       1751        4.98 0.343
```

Table 5: Articles by type of access

```
nz_oa_df %>%
  group_by(`OA Status`) %>%
  summarise(articles = n(),
            citations = mean(`Crossref citations`, na.rm = TRUE)) %>%
  mutate(prop = articles / sum(articles))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
## # A tibble: 6 x 4
##   `OA Status` articles citations prop
##   <chr>      <int>      <dbl> <dbl>
## 1 bronze      1056        5.29 0.0909
## 2 closed      6803        4.62 0.586
## 3 diamond      258        1.78 0.0222
## 4 gold       1694        5.17 0.146
## 5 green      1194        7.27 0.103
## 6 hybrid       608        8.09 0.0524
```

Table 6: Articles by type of access for New Zealand corresponding authors

```
nz_oa_df %>%
  filter(`NZ Reprint author` == "Yes") %>%
```

```
group_by(`OA Status`) %>%
summarise(articles = n(),
          citations = mean(`Crossref citations`, na.rm = TRUE)) %>%
mutate(prop = articles / sum(articles))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 6 x 4
##   `OA Status` articles citations    prop
##   <chr>         <int>     <dbl> <dbl>
## 1 bronze           409       5.08 0.0802
## 2 closed          3350       3.77 0.657
## 3 diamond           92       1.40 0.0180
## 4 gold             693       4.76 0.136
## 5 green            410       5.34 0.0804
## 6 hybrid           147       6.97 0.0288
```

Session info

```
sessionInfo()
```

```
## R version 4.0.0 (2020-04-24)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/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] forcats_0.5.0  stringr_1.4.0  dplyr_1.0.0    purrr_0.3.4
## [5] readr_1.3.1    tidyr_1.1.0    tibble_3.0.3   ggplot2_3.3.0
## [9] tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.1.0 xfun_0.14      haven_2.2.0    lattice_0.20-41
## [5] colorspace_1.4-1 vctrs_0.3.1    generics_0.0.2 htmltools_0.5.0
## [9] yaml_2.2.1       utf8_1.1.4     rlang_0.4.7    pillar_1.4.6
## [13] glue_1.4.1       withr_2.2.0    DBI_1.1.0      dbplyr_1.4.3
## [17] modelr_0.1.8     readxl_1.3.1   lifecycle_0.2.0 munsell_0.5.0
## [21] gtable_0.3.0     cellranger_1.1.0 rvest_0.3.5    evaluate_0.14
## [25] knitr_1.28       fansi_0.4.1    broom_0.5.6    Rcpp_1.0.4.6
## [29] scales_1.1.1     backports_1.1.7 jsonlite_1.7.0 fs_1.4.1
## [33] hms_0.5.3        digest_0.6.25  stringi_1.4.6  grid_4.0.0
## [37] cli_2.0.2        tools_4.0.0    magrittr_1.5    crayon_1.3.4
## [41] pkgconfig_2.0.3  ellipsis_0.3.1 xml2_1.3.2     reprex_0.3.0
## [45] lubridate_1.7.8  assertthat_0.2.1 rmarkdown_2.1  httr_1.4.1
```

```
## [49] rstudioapi_0.11  R6_2.4.1          nlme_3.1-147      compiler_4.0.0
```