

# mini-review

Yingjie

2023-03-05

## Data import

### data cleaning

```
expand_col_to_long <- function(data, target_col = "Mental health indicators") {  
  data_long <- data %>%  
    dplyr::rename("col_split" = target_col) %>%  
    cSplit(  
      indt = .,  
      splitCols = c("col_split"),  
      sep = ";|",  
      drop = F, # drop the original col or not  
      direction = "long", # this is better than "wide"  
      stripWhite = T  
    ) %>% # clean white space  
    dplyr::mutate(across(where(is.factor), as.character)) %>%  
    dplyr::mutate(col_split = trimws(col_split)) %>%  
    dplyr::mutate(col_split = ifelse(  
      nchar(col_split) > 10,  
      str_to_sentence(col_split),  
      col_split  
    )) %>%  
    ## capitalizes first word but not subsequent words  
    dplyr::mutate(col_split = Hmisc::capitalize(col_split)) %>%  
    as.data.frame() %>%  
    group_by(col_split) %>%  
    dplyr::summarise_at(c("n"), sum, na.rm = T) %>%  
    as.data.frame()  
  
    ## change back to the original column name  
    names(data_long)[names(data_long) == "col_split"] <- target_col  
  
  return(data_long)  
}  
  
plot_freq <- function(data, var = "Mental health indicators") {  
  p <- ggplot(  
    data = data,
```

```

aes(
  x = reorder(eval(parse(text = var)), n),
  y = n,
  # fill = n
)) +
geom_col() +
theme_bw() +
# theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
coord_flip() +
# scale_fill_distiller(name = '', palette = "Blues", guide = "colorbar", direction = 1) +
xlab("")

return(p)
}

```

```

region <- dt %>%
  dplyr::select(1, Country) %>%
  dplyr::filter(!is.na(Country)) %>%
  dplyr::filter(!Country %in% c('m', 'NA', 'Global', 'Europe', 'all over the world')) %>%
  dplyr::mutate(
    Country = case_when(
      str_detect(string = Country, pattern = 'Scotland|UK') ~ 'United Kingdom',
      str_detect(string = Country, pattern = 'Indonasia') ~ 'Indonesia',
      str_detect(string = Country, pattern = 'Malasia') ~ 'Malaysia',
      str_detect(string = Country, pattern = 'Danmark') ~ 'Denmark',
      str_detect(string = Country, pattern = 'Beigium') ~ 'Belgium',
      str_detect(string = Country, pattern = 'Brasil') ~ 'Brazil',
      str_detect(string = Country, pattern = 'Chili') ~ 'Chile',
      T ~ Country)
    )

library(SDGdetector)
packageVersion('SDGdetector')

```

```
## [1] '2.7.2'
```

```

library(dplyr)
library(stringr)
library(ggplot2)
library(cowplot)

```

```
## Warning: package 'cowplot' was built under R version 3.6.3
```

```

codelist.supp <- data.frame(
  country.name.en = c('USA'),
  iso3c = c('USA')
)
codelist <- SDGdetector::codelist_panel %>%
  dplyr::distinct(country.name.en, iso3c) %>%
  rbind(., codelist.supp) %>%
  dplyr::mutate(

```

```

    country.name.en = case_when(
      str_detect(string = iso3c, pattern = 'HKG') ~ 'Hong Kong',
      T ~ country.name.en)
  ) %>%
  dplyr::distinct(country.name.en, iso3c)

shp <- SDGdetector::shp

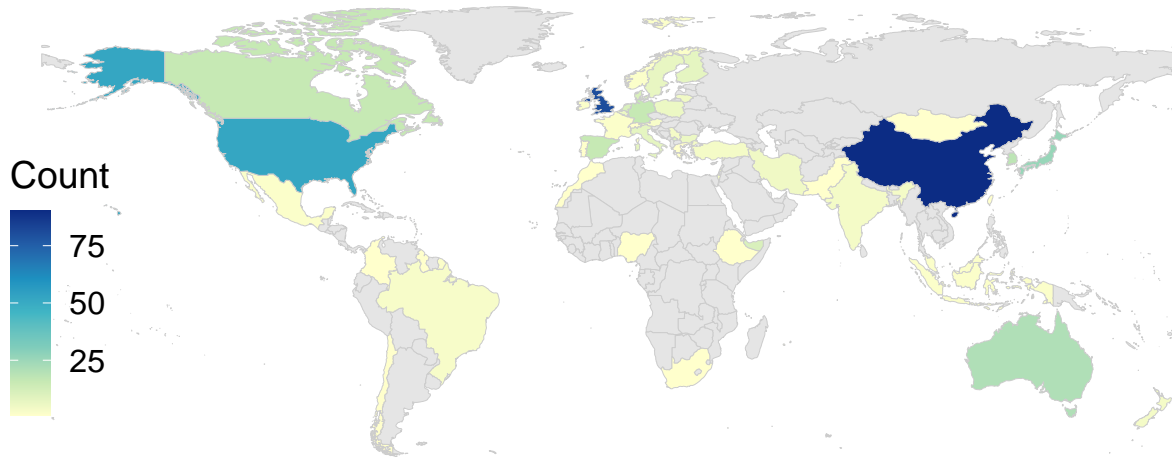
dt_match <- region %>%
  merge(x = .,
        y = codelist,
        by.x = 'Country',
        by.y = 'country.name.en', all.x = T) %>%
  arrange(!is.na(iso3c))

region_count <- dt_match %>%
  group_by(iso3c) %>%
  tally()

dt_sf <- region_count %>%
  merge(x= shp, y = ., by.x = 'iso_a3', by.y = 'iso3c', all.x = T)

dt_sf %>%
  dplyr::filter(name != 'Antarctica') %>%
  ggplot(.) +
  geom_sf(aes(fill = n), size = 0.1, color = 'gray80') +
  # geom_sf_text(aes(label = iso_a3), colour = "gray", size = 1) +
  scale_fill_distiller(name= 'Count', palette = 'YlGnBu', direction = 1, na.value = "gray90") +
  # theme_bw() +
  theme_nothing() +
  theme(legend.position = c(0.06, 0.4))

```



```
fname <- paste0(dir.fig, 'mini-review_paper_count_', 'map.png'); fname
```

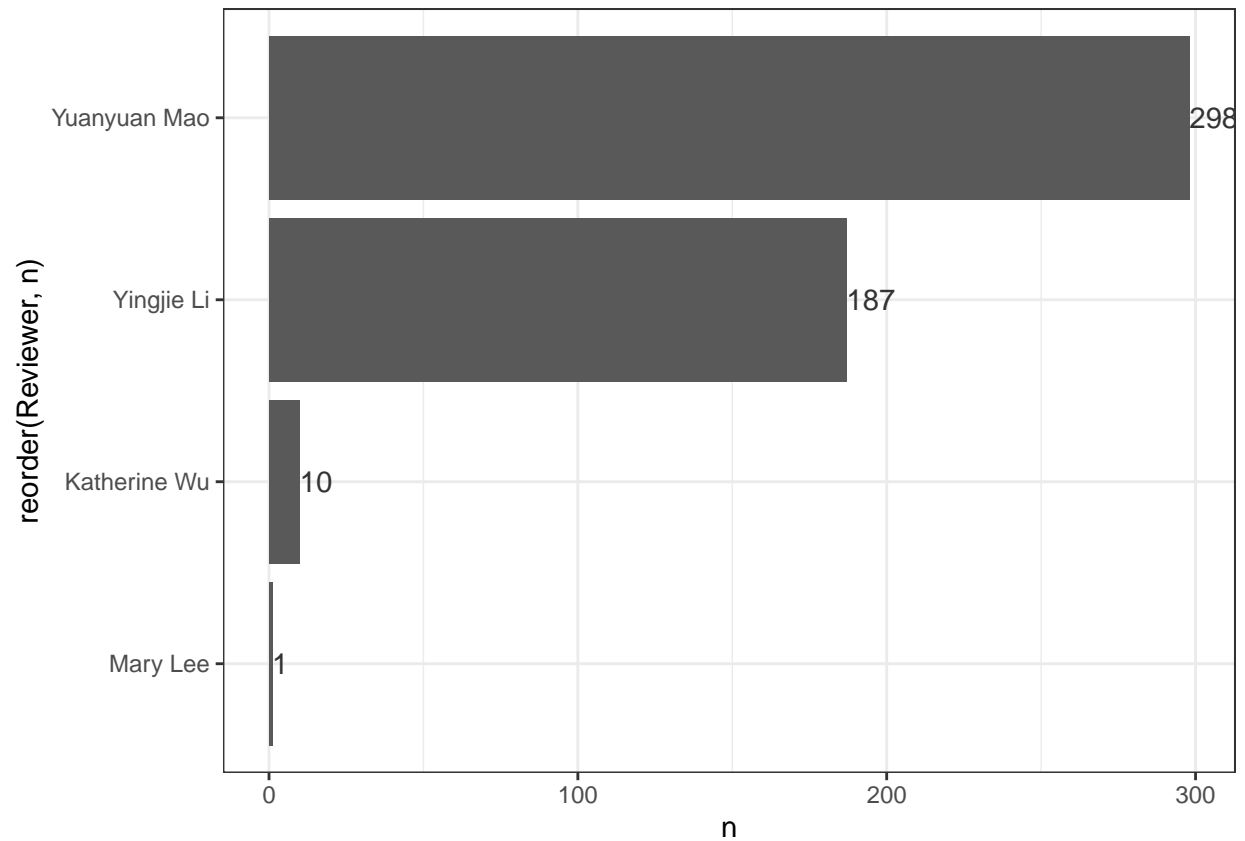
```
## [1] "./figures/mini-review_paper_count_map.png"
```

```
ggsave(filename = fname, plot = last_plot(), width = 6.4*2, height = 3.2*2, units = 'in', dpi = 300)
```

## Viz

### proregss

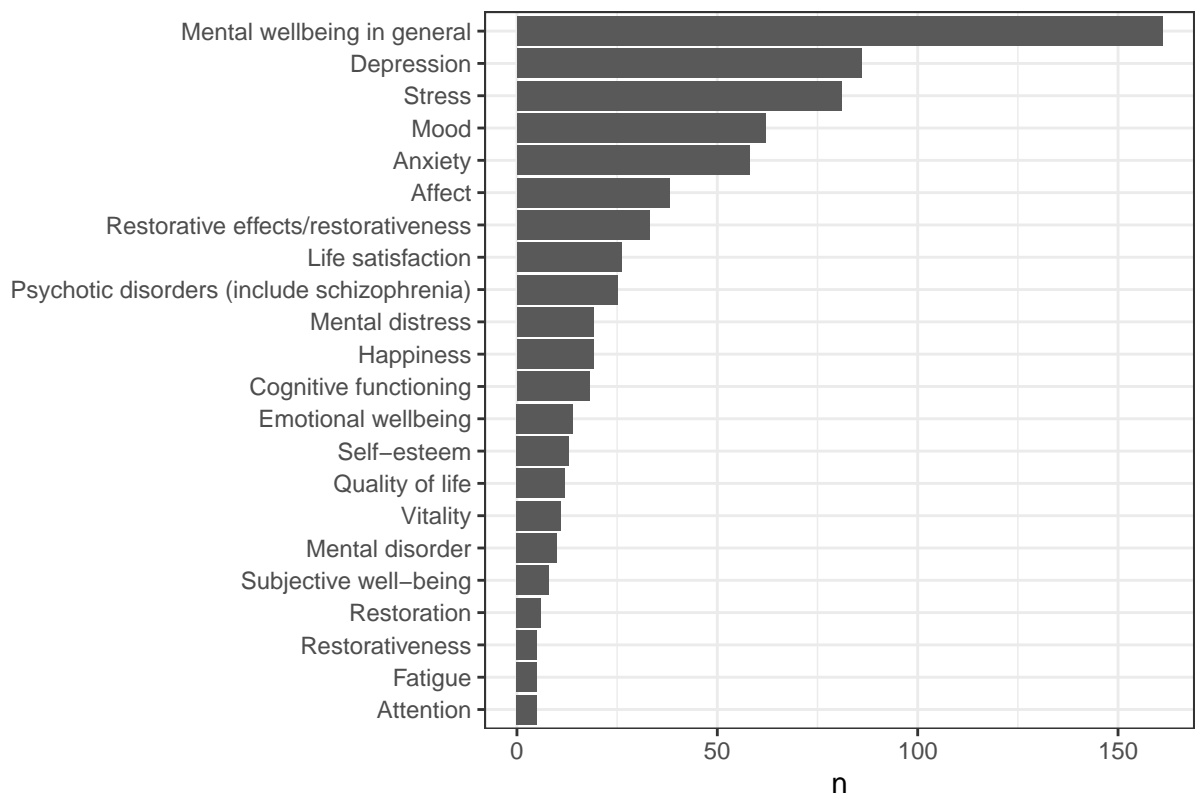
```
dt %>%
  group_by(Reviewer) %>%
  tally() %>%
  ggplot(aes(x = reorder(Reviewer, n), y = n)) +
  geom_col() +
  geom_text(aes(label = n), vjust = 0.5, hjust = 0, color = 'gray20') +
  coord_flip() +
  theme_bw()
```



## Count by indicator

```
dt.bymhi.clean %>%
  dplyr::slice_max(order_by = n, n = 20) %>%
  ggplot(aes(
    x = reorder(Indicator, n),
    y = n,
    # fill = n
  )) +
  geom_col() +
  theme_bw() +
  # theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
  coord_flip() +
  # scale_fill_distiller(name = '', palette = "Blues", guide = "colorbar", direction = 1) +
  xlab("") +
  ggtitle("Frequency of Mental health indicators")
```

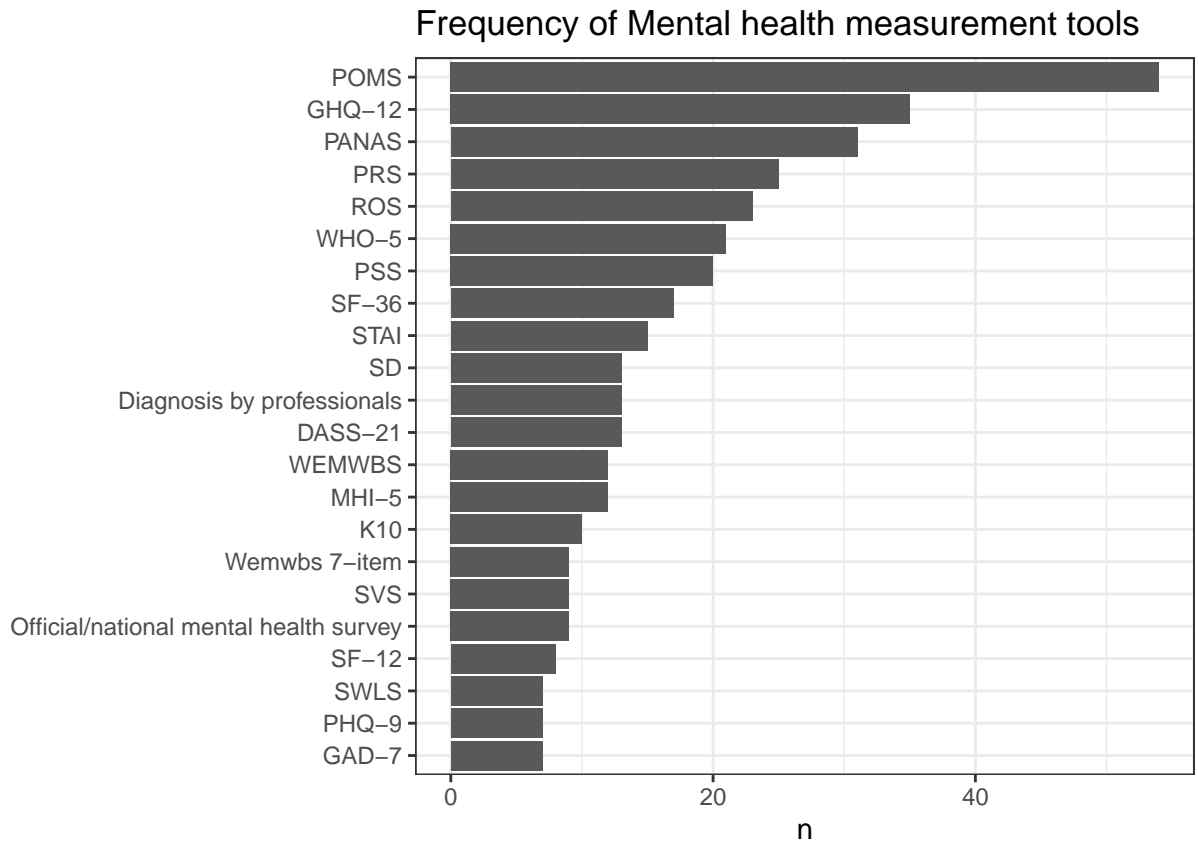
Frequency of Mental health indicators



```
fname <- paste0(dir.fig, "mini-review_mh_freq_", today, ".png")
ggsave(filename = fname, plot = last_plot(), width = 16 / 2, height = 9 / 2, units = "in", dpi = 300, bg = "white")
```

## Count by tool

```
dt.bytool.clean %>%
  dplyr::slice_max(order_by = n, n = 20) %>%
  plot_freq(data = ., var = 'Tool') +
  ggtitle("Frequency of Mental health measurement tools")
```



```
fname <- paste0(dir.fig, "mini-review_mh_tool_freq_", today, ".png")
ggsave(filename = fname, plot = last_plot(), width = 16 / 2, height = 9 / 2, units = "in", dpi = 300, bg = "white")
```