Explanation of columns in "counts-raw.txt.gz":

- **doi** - digital object identifier
- **pubDate** - Date of publication, format: yyyy-mm-dd
- **journal** - Abbreviation of PLOS journal
- **title** - Title of article
- **articleType** - Article classification
- **authorsCount** - Number of authors
- **f1000Factor** - Score assigned by Faculty of 1000
- **backtweetsCount** - Number of tweets
- **deliciousCount** - Number of bookmarks in Delicious
- **pmid** - PubMed ID number
- **plosSubjectTags** - Descriptions of the subject areas of the article
- **plosSubSubjectTags** - More specific descriptions of the subject areas of the article
- **facebookShareCount** - Number of Facebook shares
- **facebookLikeCount** - Number of Facebook likes
- **facebookCommentCount** - Number of Facebook comments
- **facebookClickCount** - Number of Facebook clicks
- **mendeleyReadersCount** - Number of Mendeley users that bookmarked article
- **almBlogsCount** - Number of blog posts that link to article
- **pdfDownloadsCount** - Number of PDF downloads
- **xmlDownloadsCount** - Number of XML downloads
- **htmlDownloadsCount** - Number of page views
- **almCiteULikeCount** - Number of saves in CiteULike
- **almScopusCount** - Number of citations in Scopus
- **almPubMedCentralCount** - Number of citations in PubMed Central
- **almCrossRefCount** - Number of citations in CrossRef
- **plosCommentCount** - Number of comments on PLOS website
- **plosCommentResponsesCount** - Number of responses to comments on PLOS website
- **wikipediaCites** - Number of links to article
- **year** - Year of publication, format: yyyy
- **daysSincePublished** - Number of days since publication (out-dated)
- **wosCountThru2010** - Number of citations in Web of Science as of 2010
- **wosCountThru2011** - Number of citations in Web of Science as of 2011

dplyr

- **filter** - subset rows
- **select** - subset columns
- **arrange** - sort column(s)
- **mutate** - create new column(s)
- **group_by** - split data into groups based on values in column(s)
- **summarize** - reduce all rows (per group) to one summary row
- %>% - pipe output of one function to the next

ggplot2

- **aes** - Map columns of data frame to plot aesthetics

  - **x** - data on x-axis
  - **y** - data on y-axis
  - **col** - color of points and lines
  - **shape** - shape of points
  - **size** - size of points
  - **fill** - color of geometric shapes
- **geom_*** - The geometric objects to be plotted
  - **geom_point** - scatter plot
  - **geom_bar** - bar plot
  - **geom_histogram** - histogram
  - **geom_smooth** - loess curve
  - **geom_text** - use text labels instead of points
  - **geom_errorbar** - Add error bars
- **scale_x_log10, scale_y_log10** - Log transform an axis
- **scale_x_continous, scale_x_discrete** - Change breaks and labels on axis
- **scale_color_manual, scale_fill_manual, scale_color_brewer** - Change colors used for color or fill aesthetics
- **facet_grid, facet_wrap** - Create a plot per group
- **theme** - Change the appearance of the plot

Debugging

- **debug(function_name)** - Enter debugger whenever function is called
- **browser()** - Enter debugger
- **options(error = recover)** - Set this option to enter debugger whenever an error occurs

Defensive programming

- **stopifnot(cond1, cond2, . . .)** - Stop if any of the listed conditions are FALSE
- From package testit:
  - **assert("message", cond1, . . .)** - Stop and print message if any of the listed conditions are FALSE
  - **has_warning(expr)** - Return TRUE if expression creates a warning
  - **has_error(expr)** - Return TRUE if expression creates an error