

# Package ‘package10883408’

January 10, 2024

**Type** Package

**Title** Article Data Analysis and Visualization with Package

**Version** 0.1.0

**Author** 10883408

**Maintainer** 10883408 <10883408@plymouth.ac.uk>

**Description** The package\_10883408 made for text analysis. It is perfect for researchers because it provides essential features for analysing word frequencies in datasets. It provides below features.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Imports** dplyr,

tm,

ggplot2,

tidytext,

scales,

tidyr,

forcats,

lubridate,

gridExtra,

tidyverse

**RoxygenNote** 7.2.3

**Depends** R (>= 2.10)

**Suggests** knitr,

rmarkdown

**VignetteBuilder** knitr

## R topics documented:

analyze_articles . . . . .	2
calculate_and_plot_tfidf . . . . .	2
plot_zipfs_law . . . . .	3
word_frequency_plot . . . . .	3

<b>Index</b>	<b>5</b>
--------------	----------

---

analyze_articles	<i>Analyze Articles from a File</i>
------------------	-------------------------------------

---

### Description

This function performs an analysis of articles contained in a given file. It includes the distribution of articles over time and the count of articles per journal. It generates visualizations for these analyses.

### Usage

```
analyze_articles(article_data, published, journal)
```

### Arguments

filepath	The path to the file containing article data.
----------	---

### Value

A grid of plots showing the distribution of articles over time and the count of articles per journal.

### Examples

```
# Example usage:
analyze_articles(article_data, published, journal)
```

---

calculate_and_plot_tfidf	<i>Calculate and Plot TF-IDF</i>
--------------------------	----------------------------------

---

### Description

This function calculates the Term Frequency-Inverse Document Frequency (TF-IDF) for words in a collection of articles, grouped by journal. It then plots the top TF-IDF words for each journal, providing a visualization of the most distinctive words used in different journals' articles about the Ukraine war in 2022.

### Usage

```
calculate_and_plot_tfidf(article_data, articles, journal)
```

### Arguments

article_data	A dataframe containing the articles and their metadata.
articles	The column name in 'article_data' that contains the article text.
journal	The column name in 'article_data' that contains the journal names.

### Value

A ggplot object representing the top TF-IDF words for each journal.

**Examples**

```
# Example usage:
calculate_and_plot_tfidf(article_data, "articles", "journal")
```

---

plot_zipfs_law	<i>Plot Zipf's Law for Article Data</i>
----------------	---

---

**Description**

This function applies Zipf's law to a given dataset of articles. It tokenizes the words in the articles, performs a count and a TF-IDF (Term Frequency-Inverse Document Frequency) calculation, then uses linear regression to analyze the relationship between the ranks of words and their frequencies. The result is a log-log plot illustrating Zipf's law, with different colors for each journal.

**Usage**

```
plot_zipfs_law(article_data, articles, journal)
```

**Arguments**

article_data	A dataframe containing the articles and their metadata.
articles	The column name in 'article_data' that contains the article text.
journal	The column name in 'article_data' that contains the journal names.

**Value**

A list containing the processed article data, regression analysis summary, and the ggplot object.

**Examples**

```
# Example usage:
plot_zipfs_law(article_data, "articles", "journal")
```

---

word_frequency_plot	<i>Word Frequency Plot</i>
---------------------	----------------------------

---

**Description**

Creates a plot of word frequencies over time from a set of articles.

**Usage**

```
word_frequency_plot(article_data, articles)
```

**Arguments**

article_data	A dataframe containing the articles and their metadata.
articles	The column name in 'article_data' that contains the article text.

**Details**

This function takes a dataframe containing articles and their metadata and produces a plot showing the frequency of specific words over time. It is particularly useful for analyzing trends in article content.

**Value**

A ggplot object representing the word frequency plot.

**Examples**

```
# Example usage:  
word_frequency_plot(article_data, specific_words)
```

# Index

`analyze_articles`, [2](#)

`calculate_and_plot_tfidf`, [2](#)

`plot_zipfs_law`, [3](#)

`word_frequency_plot`, [3](#)