

Column Modification - “Country”

Charles J. Gomez, Harshvardhan Singh

2023-06-21

The country column in our dataset contains a more detailed information about the authors’ affiliations, including first, middle, and last authors. However, this brings a level of complexity and potential confusion for machine learning models. Therefore, we propose a method to simplify the country column by retaining only the first and last author affiliations, ensuring a more straightforward representation of authorship.

```
library(stringr)
```

```
data = read.csv("INPUT_SQL_Text_Data_Astronomy_and_Astrophysics.csv")
data_CountryColumn <- data$country
```

#We removed all instances of middle authors' country codes. This step involved replacing the pattern "\\w+\\+middle" with an empty string.

```
data$country <- gsub("\\w+\\+middle", "", data$country)
data$country <- trimws(data$country)
print(head(data$country, 50))
```

```
## [1] "US+last US+first US+last US+first"
## [2] "US+first"
## [3] "US+first"
## [4] "IN+first US+last"
## [5] "US+last      US+first      US+last      US+first"
## [6] "US+first"
## [7] "US+first RS+last"
## [8] "US+last DE+first"
## [9] "CH+first CH+first"
## [10] "US+first US+first US+last US+last"
## [11] "GB+first GB+first US+last GB+first US+last GB+first"
## [12] "CN+first CN+last CN+first CN+last"
## [13] "US+first"
## [14] "US+first US+last"
## [15] "US+last US+last US+first US+first US+first US+last US+last US+first"
## [16] "FR+last US+first"
## [17] "IT+first IT+first IT+first IT+first"
## [18] "US+first US+first US+last US+last"
## [19] "RU+last RU+first GB+first"
## [20] "US+first US+last US+first US+last"
## [21] "US+last US+first CA+first US+first US+last CA+first"
## [22] "US+last US+first US+first US+last"
## [23] "US+first US+first"
## [24] "US+first US+last"
## [25] "US+last GB+first      US+last      GB+first"
## [26] "GB+first NL+last GB+first NL+last"
## [27] "AU+last AU+last AU+first AU+first"
```

```
## [28] "US+last GB+first US+last GB+first"
## [29] "US+first US+last US+last US+first"
## [30] "GB+last GB+first US+first GB+first US+first GB+last"
## [31] "GB+first GB+last IT+first GB+last GB+first IT+first"
## [32] "GB+last GB+last NL+first NL+first"
## [33] "US+first"
## [34] "AU+first AU+first AU+last AU+last"
## [35] "GB+last"
## [36] "DK+last US+first US+first DK+last"
## [37] "CA+first SE+last"
## [38] "US+first US+last US+first US+last"
## [39] "US+last US+last US+first US+first"
## [40] "FR+last US+first"
## [41] "AU+first AU+first AU+first AU+first"
## [42] "US+last US+first"
## [43] "CL+first US+last"
## [44] "BR+first BR+last BR+first BR+last"
## [45] "US+first US+first"
## [46] "AU+last US+first AU+last US+first"
## [47] "IT+first US+first US+last US+last US+first IT+first"
## [48] "US+last"
## [49] "US+first US+last"
## [50] "SE+first SE+last"
```

#We removed the "+first" and "+last" strings from the remaining country codes using another regular exp

```
data$country <- gsub("\\+last|\\+first", "", data$country)
# Remove any extra spaces resulting from the removal
data$country <- sapply(strsplit(data$country, "\\s+"), function(x) paste(x[x != ""], collapse = " "))
print(head(data$country, 50))
```

```
## [1] "US US US US" "US"
## [3] "US" "IN US"
## [5] "US US US US" "US"
## [7] "US RS" "US DE"
## [9] "CH CH" "US US US US"
## [11] "GB GB US GB US GB" "CN CN CN CN"
## [13] "US" "US US"
## [15] "US US US US US US US US" "FR US"
## [17] "IT IT IT IT" "US US US US"
## [19] "RU RU GB" "US US US US"
## [21] "US US CA US US CA" "US US US US"
## [23] "US US" "US US"
## [25] "US GB US GB" "GB NL GB NL"
## [27] "AU AU AU AU" "US GB US GB"
## [29] "US US US US" "GB GB US GB US GB"
## [31] "GB GB IT GB GB IT" "GB GB NL NL"
## [33] "US" "AU AU AU AU"
## [35] "GB" "DK US US DK"
## [37] "CA SE" "US US US US"
## [39] "US US US US" "FR US"
## [41] "AU AU AU AU" "US US"
## [43] "CL US" "BR BR BR BR"
## [45] "US US" "AU US AU US"
```

```
## [47] "IT US US US US IT"      "US"
## [49] "US US"                  "SE SE"

# calculate the percentage count of each country code in a vector
#calculate_percentage <- function(vec) {
# counts <- table(vec)
# percentages <- prop.table(counts) * 100
# formatted <- paste0(round(percentages, 1), "%", names(percentages))
# paste(formatted, collapse = " ")
#}

# Apply the calculate_percentage function to each row in the 'country' column
#data$country <- sapply(strsplit(data$country, "\\s+"), calculate_percentage)

# Print the modified 'country' column (first 20)
#print(head(data$country, 50))

# Calculate percentage count of each country code in a vector
calculate_percentage <- function(vec) {
  counts <- table(vec)
  percentages <- prop.table(counts) * 100
  return(percentages)
}

# Apply the calculate_percentage function to each row in the 'country' column
percentage_counts <- lapply(strsplit(data$country, "\\s+"), calculate_percentage)

# Get all unique countries
all_countries <- unique(unlist(lapply(percentages_counts, names)))

# Initialize a new data frame to hold the percentages
percentage_df <- data.frame(matrix(ncol = length(all_countries), nrow = length(percentages_counts)))
names(percentages_df) <- all_countries

# Fill the data frame with percentages
for (i in seq_along(percentages_counts)) {
  country_names <- names(percentages_counts[[i]])
  country_percentages <- percentages_counts[[i]]
  percentage_df[i, country_names] <- country_percentages
}

# Replace NA values with 0
percentage_df[is.na(percentages_df)] <- 0

# Combine the original data with the new percentage-encoded country columns
data <- cbind(data, percentage_df)

# Summary of dataframe
summary(data)
```

```
##   concept_id      work_id      publication_year      title
## Length:63999      Length:63999      Min.      :1824      Length:63999
```

## Class :character	Class :character	1st Qu.:2001	Class :character
## Mode :character	Mode :character	Median :2005	Mode :character
##		Mean :2005	
##		3rd Qu.:2010	
##		Max. :2022	
## paperabstract	country	year_concept	US
## Length:63999	Length:63999	Length:63999	Min. : 0.00
## Class :character	Class :character	Class :character	1st Qu.: 0.00
## Mode :character	Mode :character	Mode :character	Median : 50.00
##			Mean : 49.79
##			3rd Qu.:100.00
##			Max. :100.00
## IN	RS	DE	CH
## Min. : 0.000	Min. : 0.0000	Min. : 0.000	Min. : 0.0000
## 1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.0000
## Median : 0.000	Median : 0.0000	Median : 0.000	Median : 0.0000
## Mean : 1.397	Mean : 0.5605	Mean : 5.642	Mean : 0.8845
## 3rd Qu.: 0.000	3rd Qu.: 0.0000	3rd Qu.: 0.000	3rd Qu.: 0.0000
## Max. :100.000	Max. :100.0000	Max. :100.000	Max. :100.0000
## GB	CN	FR	IT
## Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.000
## 1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000
## Median : 0.000	Median : 0.000	Median : 0.000	Median : 0.000
## Mean : 9.964	Mean : 2.149	Mean : 1.947	Mean : 3.778
## 3rd Qu.: 0.000	3rd Qu.: 0.000	3rd Qu.: 0.000	3rd Qu.: 0.000
## Max. :100.000	Max. :100.000	Max. :100.000	Max. :100.000
## RU	CA	NL	AU
## Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.000
## 1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000
## Median : 0.000	Median : 0.000	Median : 0.000	Median : 0.000
## Mean : 1.822	Mean : 2.628	Mean : 1.975	Mean : 2.777
## 3rd Qu.: 0.000	3rd Qu.: 0.000	3rd Qu.: 0.000	3rd Qu.: 0.000
## Max. :100.000	Max. :100.000	Max. :100.000	Max. :100.000
## DK	SE	CL	BR
## Min. : 0.0000	Min. : 0.000	Min. : 0.0000	Min. : 0.0000
## 1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.0000
## Median : 0.0000	Median : 0.000	Median : 0.0000	Median : 0.0000
## Mean : 0.2942	Mean : 0.342	Mean : 0.4229	Mean : 0.7424
## 3rd Qu.: 0.0000	3rd Qu.: 0.000	3rd Qu.: 0.0000	3rd Qu.: 0.0000
## Max. :100.0000	Max. :100.000	Max. :100.0000	Max. :100.0000
## BE	JP	VE	MX
## Min. : 0.0000	Min. : 0.000	Min. : 0.00000	Min. : 0.0000
## 1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.00000	1st Qu.: 0.0000
## Median : 0.0000	Median : 0.000	Median : 0.00000	Median : 0.0000
## Mean : 0.4449	Mean : 2.551	Mean : 0.00521	Mean : 0.7465
## 3rd Qu.: 0.0000	3rd Qu.: 0.000	3rd Qu.: 0.00000	3rd Qu.: 0.0000
## Max. :100.0000	Max. :100.000	Max. :50.00000	Max. :100.0000
## PL	ES	IE	FI
## Min. : 0.000	Min. : 0.000	Min. : 0.0000	Min. : 0.0000
## 1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.0000
## Median : 0.000	Median : 0.000	Median : 0.0000	Median : 0.0000
## Mean : 0.573	Mean : 1.749	Mean : 0.2064	Mean : 0.2596
## 3rd Qu.: 0.000	3rd Qu.: 0.000	3rd Qu.: 0.0000	3rd Qu.: 0.0000
## Max. :100.000	Max. :100.000	Max. :100.0000	Max. :100.0000

##	KR	TW	AT	IR
##	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000
##	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000
##	Median : 0.0000	Median : 0.0000	Median : 0.0000	Median : 0.0000
##	Mean : 0.4737	Mean : 0.2367	Mean : 0.2039	Mean : 0.1189
##	3rd Qu.: 0.0000	3rd Qu.: 0.0000	3rd Qu.: 0.0000	3rd Qu.: 0.0000
##	Max. :100.0000	Max. :100.0000	Max. :100.0000	Max. :100.0000
##	HU	IL	LV	UA
##	Min. : 0.0000	Min. : 0.0000	Min. :0.00e+00	Min. : 0.0000
##	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.:0.00e+00	1st Qu.: 0.0000
##	Median : 0.0000	Median : 0.0000	Median :0.00e+00	Median : 0.0000
##	Mean : 0.2299	Mean : 0.9436	Mean :4.69e-03	Mean : 0.2302
##	3rd Qu.: 0.0000	3rd Qu.: 0.0000	3rd Qu.:0.00e+00	3rd Qu.: 0.0000
##	Max. :100.0000	Max. :100.0000	Max. :1.00e+02	Max. :100.0000
##	GE	GR	AR	NG
##	Min. : 0.00000	Min. : 0.0000	Min. : 0.0000	Min. : 0.00000
##	1st Qu.: 0.00000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.00000
##	Median : 0.00000	Median : 0.0000	Median : 0.0000	Median : 0.00000
##	Mean : 0.03542	Mean : 0.2227	Mean : 0.3265	Mean : 0.04271
##	3rd Qu.: 0.00000	3rd Qu.: 0.0000	3rd Qu.: 0.0000	3rd Qu.: 0.00000
##	Max. :100.00000	Max. :100.0000	Max. :100.0000	Max. :100.00000
##	HR	TR	AM	CZ
##	Min. : 0.00000	Min. : 0.0000	Min. : 0.00000	Min. : 0.0000
##	1st Qu.: 0.00000	1st Qu.: 0.0000	1st Qu.: 0.00000	1st Qu.: 0.0000
##	Median : 0.00000	Median : 0.0000	Median : 0.00000	Median : 0.0000
##	Mean : 0.02995	Mean : 0.2077	Mean : 0.09844	Mean : 0.1352
##	3rd Qu.: 0.00000	3rd Qu.: 0.0000	3rd Qu.: 0.00000	3rd Qu.: 0.0000
##	Max. :100.00000	Max. :100.0000	Max. :100.00000	Max. :100.0000
##	KZ	KH	PT	CO
##	Min. : 0.00000	Min. : 0.00000	Min. : 0.0000	Min. : 0.00000
##	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.: 0.0000	1st Qu.: 0.00000
##	Median : 0.00000	Median : 0.00000	Median : 0.0000	Median : 0.00000
##	Mean : 0.01602	Mean : 0.02526	Mean : 0.1365	Mean : 0.02904
##	3rd Qu.: 0.00000	3rd Qu.: 0.00000	3rd Qu.: 0.0000	3rd Qu.: 0.00000
##	Max. :100.00000	Max. :100.00000	Max. :100.0000	Max. :100.00000
##	MZ	BW	ZA	SK
##	Min. : 0.00000	Min. : 0.00000	Min. : 0.0000	Min. : 0.00000
##	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.: 0.0000	1st Qu.: 0.00000
##	Median : 0.00000	Median : 0.00000	Median : 0.0000	Median : 0.00000
##	Mean : 0.02552	Mean : 0.00156	Mean : 0.2787	Mean : 0.08971
##	3rd Qu.: 0.00000	3rd Qu.: 0.00000	3rd Qu.: 0.0000	3rd Qu.: 0.00000
##	Max. :100.00000	Max. :50.00000	Max. :100.0000	Max. :100.00000
##	PK	IS	NO	NZ
##	Min. : 0.00000	Min. : 0.00000	Min. : 0.00000	Min. : 0.00000
##	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.: 0.00000
##	Median : 0.00000	Median : 0.00000	Median : 0.00000	Median : 0.00000
##	Mean : 0.05391	Mean : 0.00612	Mean : 0.08399	Mean : 0.08985
##	3rd Qu.: 0.00000	3rd Qu.: 0.00000	3rd Qu.: 0.00000	3rd Qu.: 0.00000
##	Max. :100.00000	Max. :66.66667	Max. :100.00000	Max. :100.00000
##	BG	VN	MY	GH
##	Min. : 0.00000	Min. : 0.0000	Min. : 0.00000	Min. : 0.00000
##	1st Qu.: 0.00000	1st Qu.: 0.0000	1st Qu.: 0.00000	1st Qu.: 0.00000
##	Median : 0.00000	Median : 0.0000	Median : 0.00000	Median : 0.00000
##	Mean : 0.07552	Mean : 0.0013	Mean : 0.01042	Mean : 0.00365

##	3rd Qu.: 0.00000	3rd Qu.: 0.0000	3rd Qu.: 0.00000	3rd Qu.: 0.00000
##	Max. :100.00000	Max. :50.0000	Max. :100.00000	Max. :50.00000
##	SI	GT	EE	EG
##	Min. : 0.00000	Min. :0.0e+00	Min. : 0.00000	Min. : 0.00000
##	1st Qu.: 0.00000	1st Qu.:0.0e+00	1st Qu.: 0.00000	1st Qu.: 0.00000
##	Median : 0.00000	Median :0.0e+00	Median : 0.00000	Median : 0.00000
##	Mean : 0.02076	Mean :7.8e-04	Mean : 0.01563	Mean : 0.01797
##	3rd Qu.: 0.00000	3rd Qu.:0.0e+00	3rd Qu.: 0.00000	3rd Qu.: 0.00000
##	Max. :100.00000	Max. :5.0e+01	Max. :100.00000	Max. :100.00000
##	RW	UG	UY	JM
##	Min. : 0.00000	Min. : 0.00000	Min. :0.00e+00	Min. :0.00e+00
##	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.:0.00e+00	1st Qu.:0.00e+00
##	Median : 0.00000	Median : 0.00000	Median :0.00e+00	Median :0.00e+00
##	Mean : 0.00193	Mean : 0.00052	Mean :2.08e-03	Mean :5.47e-03
##	3rd Qu.: 0.00000	3rd Qu.: 0.00000	3rd Qu.:0.00e+00	3rd Qu.:0.00e+00
##	Max. :50.00000	Max. :33.33333	Max. :1.00e+02	Max. :1.00e+02
##	SA	RO	UZ	
##	Min. : 0.00000	Min. : 0.00000	Min. : 0.00000	
##	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.: 0.00000	
##	Median : 0.00000	Median : 0.00000	Median : 0.00000	
##	Mean : 0.01979	Mean : 0.01979	Mean : 0.02349	
##	3rd Qu.: 0.00000	3rd Qu.: 0.00000	3rd Qu.: 0.00000	
##	Max. :100.00000	Max. :100.00000	Max. :100.00000	
##	TH	NP	TN	LK
##	Min. : 0.00000	Min. :0.00e+00	Min. : 0.0000	Min. :0.00e+00
##	1st Qu.: 0.00000	1st Qu.:0.00e+00	1st Qu.: 0.0000	1st Qu.:0.00e+00
##	Median : 0.00000	Median :0.00e+00	Median : 0.0000	Median :0.00e+00
##	Mean : 0.01406	Mean :6.25e-03	Mean : 0.0013	Mean :5.73e-03
##	3rd Qu.: 0.00000	3rd Qu.:0.00e+00	3rd Qu.: 0.0000	3rd Qu.:0.00e+00
##	Max. :100.00000	Max. :1.00e+02	Max. :50.0000	Max. :1.00e+02
##	PE	LT	CY	ID
##	Min. : 0.00000	Min. : 0.00000	Min. :0.00e+00	Min. :0.00e+00
##	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.:0.00e+00	1st Qu.:0.00e+00
##	Median : 0.00000	Median : 0.00000	Median :0.00e+00	Median :0.00e+00
##	Mean : 0.00495	Mean : 0.01172	Mean :8.91e-03	Mean :9.11e-03
##	3rd Qu.: 0.00000	3rd Qu.: 0.00000	3rd Qu.:0.00e+00	3rd Qu.:0.00e+00
##	Max. :66.66667	Max. :100.00000	Max. :1.00e+02	Max. :1.00e+02
##	EC	LU	CR	TJ
##	Min. : 0.00000	Min. :0.0e+00	Min. :0.00e+00	Min. :0.00e+00
##	1st Qu.: 0.00000	1st Qu.:0.0e+00	1st Qu.:0.00e+00	1st Qu.:0.00e+00
##	Median : 0.00000	Median :0.0e+00	Median :0.00e+00	Median :0.00e+00
##	Mean : 0.00182	Mean :7.8e-04	Mean :7.03e-03	Mean :3.13e-03
##	3rd Qu.: 0.00000	3rd Qu.:0.0e+00	3rd Qu.:0.00e+00	3rd Qu.:0.00e+00
##	Max. :50.00000	Max. :5.0e+01	Max. :1.00e+02	Max. :1.00e+02
##	AE	LB	SG	BO
##	Min. :0.00e+00	Min. :0.0e+00	Min. :0.00e+00	Min. : 0.00000
##	1st Qu.:0.00e+00	1st Qu.:0.0e+00	1st Qu.:0.00e+00	1st Qu.: 0.00000
##	Median :0.00e+00	Median :0.0e+00	Median :0.00e+00	Median : 0.00000
##	Mean :8.33e-03	Mean :2.6e-03	Mean :9.11e-03	Mean : 0.00156
##	3rd Qu.:0.00e+00	3rd Qu.:0.0e+00	3rd Qu.:0.00e+00	3rd Qu.: 0.00000
##	Max. :1.00e+02	Max. :1.0e+02	Max. :1.00e+02	Max. :50.00000
##	JO	AZ	IQ	OM
##	Min. :0.00e+00	Min. :0.00e+00	Min. :0.00e+00	Min. : 0.00000
##	1st Qu.:0.00e+00	1st Qu.:0.00e+00	1st Qu.:0.00e+00	1st Qu.: 0.00000

```

## Median :0.00e+00 Median :0.00e+00 Median :0.00e+00 Median : 0.00000
## Mean :6.25e-03 Mean :4.38e-03 Mean :7.81e-03 Mean : 0.00195
## 3rd Qu.:0.00e+00 3rd Qu.:0.00e+00 3rd Qu.:0.00e+00 3rd Qu.: 0.00000
## Max. :1.00e+02 Max. :1.00e+02 Max. :1.00e+02 Max. :50.00000
## KW PR TZ BD
## Min. :0.0e+00 Min. :0.00e+00 Min. : 0.00000 Min. :0.00e+00
## 1st Qu.:0.0e+00 1st Qu.:0.00e+00 1st Qu.: 0.00000 1st Qu.:0.00e+00
## Median :0.0e+00 Median :0.00e+00 Median : 0.00000 Median :0.00e+00
## Mean :7.8e-04 Mean :7.81e-03 Mean : 0.00104 Mean :1.56e-03
## 3rd Qu.:0.0e+00 3rd Qu.:0.00e+00 3rd Qu.: 0.00000 3rd Qu.:0.00e+00
## Max. :5.0e+01 Max. :1.00e+02 Max. :33.33333 Max. :1.00e+02
## MU MT CU MK
## Min. :0.00e+00 Min. :0.00e+00 Min. :0.0e+00 Min. :0.00e+00
## 1st Qu.:0.00e+00 1st Qu.:0.00e+00 1st Qu.:0.0e+00 1st Qu.:0.00e+00
## Median :0.00e+00 Median :0.00e+00 Median :0.0e+00 Median :0.00e+00
## Mean :1.56e-03 Mean :1.56e-03 Mean :7.8e-04 Mean :1.56e-03
## 3rd Qu.:0.00e+00 3rd Qu.:0.00e+00 3rd Qu.:0.0e+00 3rd Qu.:0.00e+00
## Max. :1.00e+02 Max. :1.00e+02 Max. :5.0e+01 Max. :1.00e+02
## VI PS ZW NA
## Min. :0.00e+00 Min. :0.00e+00 Min. :0.00e+00 Min. : 0.00000
## 1st Qu.:0.00e+00 1st Qu.:0.00e+00 1st Qu.:0.00e+00 1st Qu.: 0.00000
## Median :0.00e+00 Median :0.00e+00 Median :0.00e+00 Median : 0.00000
## Mean :1.56e-03 Mean :1.56e-03 Mean :1.56e-03 Mean : 0.00156
## 3rd Qu.:0.00e+00 3rd Qu.:0.00e+00 3rd Qu.:0.00e+00 3rd Qu.: 0.00000
## Max. :1.00e+02 Max. :1.00e+02 Max. :1.00e+02 Max. :50.00000
## QA HN SD
## Min. : 0.00000 Min. :0.00e+00 Min. :0.0e+00
## 1st Qu.: 0.00000 1st Qu.:0.00e+00 1st Qu.:0.0e+00
## Median : 0.00000 Median :0.00e+00 Median :0.0e+00
## Mean : 0.00156 Mean :3.13e-03 Mean :7.8e-04
## 3rd Qu.: 0.00000 3rd Qu.:0.00e+00 3rd Qu.:0.0e+00
## Max. :50.00000 Max. :1.00e+02 Max. :5.0e+01

```

```

# structure of the dataframe
str(data)

```

```

## 'data.frame': 63999 obs. of 110 variables:
## $ concept_id : chr "https://openalex.org/C44870925" "https://openalex.org/C44870925" "https://
## $ work_id : chr "https://openalex.org/W1993867637" "https://openalex.org/W2022503540" "http
## $ publication_year: int 2004 1991 2003 2003 2002 1999 2003 1997 2007 2002 ...
## $ title : chr "KINEMATIC TREATMENT OF CORONAL MASS EJECTION EVOLUTION IN THE SOLAR WIND"
## $ paperabstract : chr "We present a kinematic study of the evolution of coronal mass ejections (
## $ country : chr "US US US US" "US" "US" "IN US" ...
## $ year_concept : chr "2004+https://openalex.org/C44870925" "1991+https://openalex.org/C44870925
## $ US : num 100 100 100 50 100 100 50 50 0 100 ...
## $ IN : num 0 0 0 50 0 0 0 0 0 0 ...
## $ RS : num 0 0 0 0 0 0 50 0 0 0 ...
## $ DE : num 0 0 0 0 0 0 0 50 0 0 ...
## $ CH : num 0 0 0 0 0 0 0 0 100 0 ...
## $ GB : num 0 0 0 0 0 0 0 0 0 0 ...
## $ CN : num 0 0 0 0 0 0 0 0 0 0 ...
## $ FR : num 0 0 0 0 0 0 0 0 0 0 ...
## $ IT : num 0 0 0 0 0 0 0 0 0 0 ...
## $ RU : num 0 0 0 0 0 0 0 0 0 0 ...
## $ CA : num 0 0 0 0 0 0 0 0 0 0 ...

```

##	\$ NL	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ AU	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ DK	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ SE	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ CL	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ BR	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ BE	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ JP	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ VE	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ MX	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ PL	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ ES	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ IE	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ FI	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ KR	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ TW	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ AT	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ IR	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ HU	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ IL	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ LV	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ UA	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ GE	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ GR	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ AR	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ NG	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ HR	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ TR	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ AM	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ CZ	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ KZ	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ KH	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ PT	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ CO	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ MZ	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ BW	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ ZA	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ SK	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ PK	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ IS	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ NO	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ NZ	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ BG	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ VN	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ MY	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ GH	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ SI	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ GT	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ EE	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ EG	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ RW	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ UG	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ UY	:	num	0	0	0	0	0	0	0	0	0	0	...
##	\$ JM	:	num	0	0	0	0	0	0	0	0	0	0	...


```

## $ SA      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ RO      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ UZ      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ TH      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ NP      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ TN      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ LK      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PE      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ LT      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ CY      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ ID      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ EC      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ LU      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ CR      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ TJ      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AE      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ LB      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ SG      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ BO      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ JO      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AZ      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ IQ      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ OM      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ KW      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PR      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ TZ      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ BD      : num 0 0 0 0 0 0 0 0 0 0 ...
## [list output truncated]

```