

Venezuela-Colombia Relations through the News

Maria, Emanuel and Nibia

Stats 456

Sat 14th, 2024

A Topic Modeling Analysis

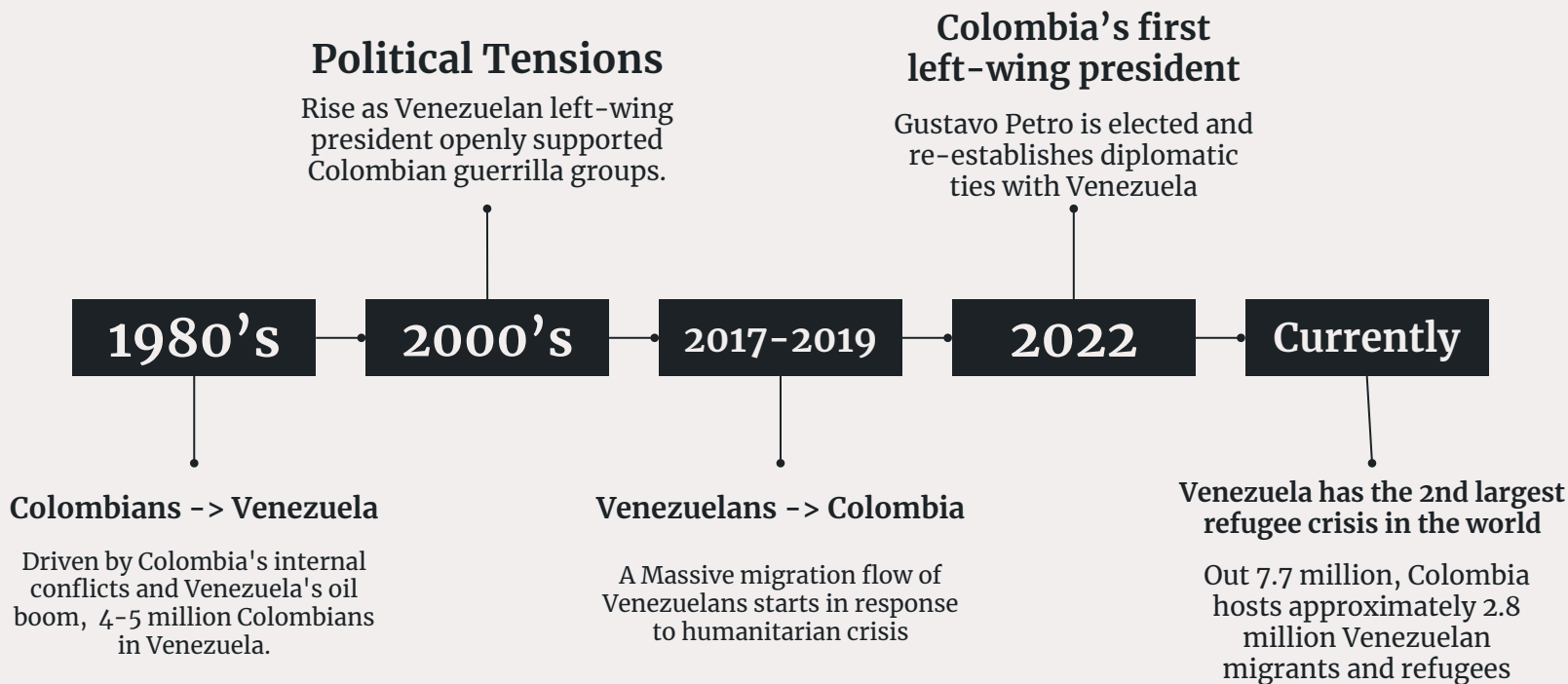


Shared history between Venezuela and Colombia

Sat 14th, 2024

Motivation

Edition: 001



Research question

What are the main themes discussed in Colombian newspapers about Venezuela–Colombia relations and how do they change in response to the current migration trend?

Our data

Sat 14th, 2024

Newspapers

Edition: 001

El Tiempo

Most read
Historically right-leaning

El Espectador

Oldest
Left-leaning



Nexis Uni

2017

2020

2023

Data Processing

Sat 14th, 2024

Steps

Edition: 001

Nexus Uni.

Download ~5000 newspaper
articles

Text mining.

Preparing text

TF-IDF.

Filtering only words with a
TF-IDF over 0.1

.PDF tools

Create a dataframe with the
relevant information

.Tokenization

Split the text into individual
words based on spaces

What our data looked like

Sat 14th, 2024

Edition: 001

ID	date	title	content	source_name	language	length	nchar	processed_text	year
0001	2023-09-17	"Venezuela se considera víctima del conflicto..."	El delegado diplomático de Colombia ante ese país...	El Tiempo	SPANISH	1833 words	11478	delegado diplomático colombia país milton rengifo ase...	2023
0002	2023-11-19	"Venezuela va a ser un motor de la economía colombian..."	El presidente Gustavo Petro habló con su homólogo sob...	El Tiempo	SPANISH	627 words	3992	Presidente gustavo petro habló homólogo migracion...	2023

Topic Modelling

Sat 14th, 2024

Edition: 001

LDA (Latent Dirichlet Allocation)

LDA assumes that:

- Each document is generated from a mix of topics.
- Each topic is a distribution over words.

Using Bayesian inference, LDA works backward from the observed words to uncover:

- The most likely topics for each document.
- The most likely words for each topic.

LDA (Latent Dirichlet Allocation)

Sat 14th, 2024

Example

Edition: 001

“X”

60%

Topic 1

{government, election,
policy, law}



40%

Topic 2

{team, goal, match, win}



20%

Topic 3

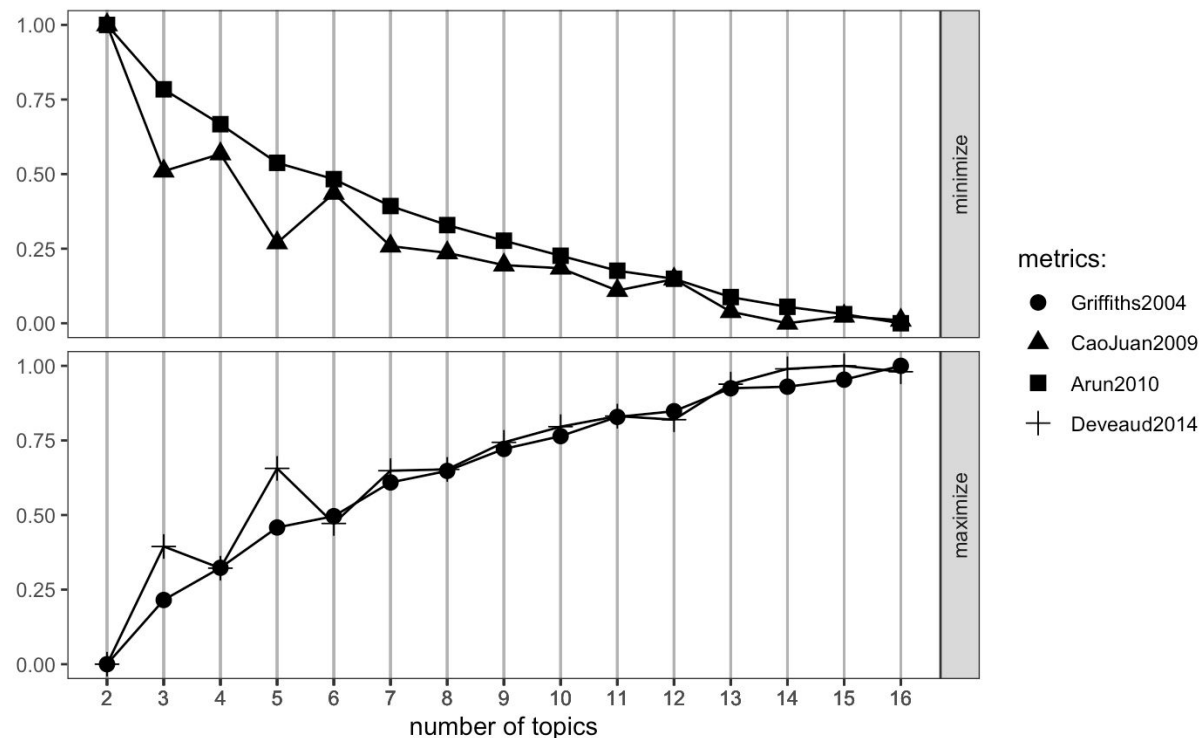
{economy, stocks, market,
trade}



Determining optimal number of topics

Sat 14th, 2024

Edition: 001



- **Griffiths2004:**
Aims to identify coherence.
- **Deveaud2014:**
Reflects interpretability.
- **CaoJuan2009:**
Measures distinctiveness.
- **Arun2010:**
Analyzes topic separation.

Assigning Topic Names

Sat 14th, 2024

Top Words

Edition: 001

Topic 1

- años (years)
- colombia (colombia)
- vida (life)
- ser (to be)
- tiempo (time)
- solo (only)
- cómo (how)
- hace (does)
- día (day)
- así (so)
- persona (person)
- gente (people)
- vez (time)
- historia (history)
- país (country)
- hacer (to do)
- mundo (world)
- bogotá (Bogotá)
- cada (each)
- después (after)

Topic 2

- presidente (president)
- petro (Petro)
- venezuela (venezuela)
- gobierno (government)
- maduro (Maduro)
- colombia (Colombia)
- dijo (said)
- oposición (opposition)
- nicolás (Nicolás)
- país (country)
- elecciones (elections)
- nacional (national)
- gustavo (Gustavo)
- política (politics)
- guaidó (Guaidó)
- campaña (campaign)
- benedetti (Benedetti)
- electoral (electoral)
- político (political)
- tiempo (time)

Topic 3

- venezuela (Venezuela)
- eln (ELN)
- colombia (Colombia)
- paz (peace)
- fiscalía (prosecutor's office)
- gobierno (government)
- según (according to)
- grupo (group)
- policía (police)
- justicia (justice)
- farc (FARC)
- autoridades (authorities)
- persona (person)
- general (general)
- país (country)
- caso (case)
- unidos (United States)
- corte (court)
- dijo (said)
- fiscal (prosecutor)

Topic 4

- colombia (Colombia)
- brasil (Brazil)
- argentina (Argentina)
- américa (America)
- selección (selection)
- venezuela (Venezuela)
- equipo (team)
- tiempo (time)
- chile (Chile)
- grupo (group)
- partido (party)
- mundial (world)
- ecuador (Ecuador)
- uruguay (Uruguay)
- Perú (Peru)
- gda (GDA)
- fútbol (soccer)
- diarios (newspapers)
- copyright (copyright)
- reservados (reserved)

Topic 5

- colombia (Colombia)
- persona (person)
- país (country)
- migrantes (migrants)
- venezuela (Venezuela)
- venezolanos (Venezuelans)
- ciento (percent)
- salud (health)
- según (according to)
- frontera (border)
- migración (migration)
- países (countries)
- millones (millions)
- año (year)
- coronavirus (coronavirus)
- población (population)
- casos (cases)
- unidos (United States)
- situación (situation)
- merlano (Merlano)

Topic 6

- colombia (Colombia)
- países (countries)
- país (country)
- venezuela (Venezuela)
- unidos (United States)
- gobierno (government)
- américa (America)
- política (politics)
- región (region)
- presidente (president)
- puede (can)
- internacional (international)
- ser (to be)
- mundo (world)
- trump (Trump)
- años (years)
- parte (part)
- biden (Biden)
- latina (Latin)
- derechos (rights)

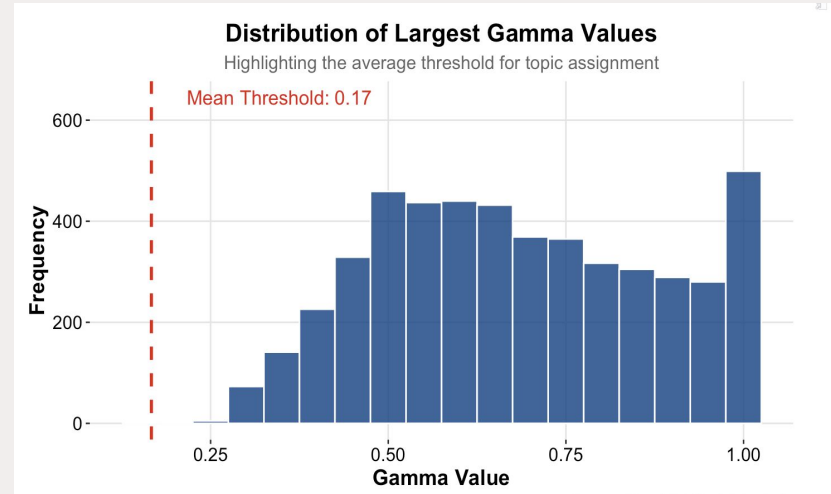
LDA Interactive Visualization

Classifying articles

Edition: 001

We categorized each newspaper article based on topic-document affiliations by Gamma Values

- **Gamma Value:**
 - Represents the degree of association between a document and each topic.
- **Assignment Criteria:**
 - Each document was assigned to the topic with the **largest gamma value**.
 - Documents with a **largest gamma value below the average gamma threshold** were omitted to maintain quality and relevance in topic assignments.



Assigning Topic Names

Sat 14th, 2024

Context

Edition: 001

Article titles		
Topic 1	Topic 2	Topic 3
<p>Mercado de fichajes, en vivo: movimientos de última hora en plena primera fecha (Soccer Transfer market, live: last-minute movements <u>during</u> the first matchday)</p>	<p>Portugal responde a Maduro y dice que fue Venezuela quien atacó a su crucero (Portugal responds to Maduro and says it was Venezuela that attacked its cruise ship)</p>	<p>Petro invitará a más sectores de la oposición venezolana: esto es lo que sabe (Petro will invite more sectors of the Venezuelan opposition: here's what we know)</p>
<p>Mercado de fichajes, en vivo: movimientos de última hora antes de la primera fecha (Soccer Transfer market, live: last-minute movements <u>before</u> the first matchday)</p>	<p>Álex Saab, señalado por corrupción en Pdvsa: este sería el monto desfalco (Álex Saab, accused of corruption in Pdvsa: this would be the embezzlement amount)</p>	<p>El balance de EE. UU. sobre la Cumbre de Venezuela, organizada por Gustavo Petro (The U.S. assessment of the Venezuela Summit organized by Gustavo Petro)</p>

Assigning Topic Names

Sat 14th, 2024

Context

Edition: 001

Article titles

Topic 4	Topic 5	Topic 6
Migrantes venezolanos no tienen protección ante el coronavirus en Norte de Santander <i>(Venezuelan migrants lack protection against coronavirus in Norte de Santander)</i>	'Es fundamental que se esté recuperando la certidumbre con empresarios': mincomercio <i>('It is essential to restore certainty with business owners': Ministry of Commerce)</i>	Fiscalía apela fallo que negó preclusión a favor de Álvaro Uribe <i>(Prosecutor appeals ruling that denied preclusion in favor of Álvaro Uribe)</i>
Accidentes viales, otro flagelo que enfrentan los migrantes, además del Darién <i>(Traffic accidents: another scourge faced by migrants, in addition to the Darién)</i>	Nuevo estudio revela el balance de las reservas de carbono en la Amazonia <i>(New study reveals the balance of carbon reserves in the Amazon)</i>	El balance de EE. UU. sobre la Cumbre de Venezuela, organizada por Gustavo Petro <i>(The U.S. assessment of the Venezuela Summit organized by Gustavo Petro)</i>

General Topic Names

Sat 14th, 2024

Topic naming

Edition: 001

Final
assigned
topic names

01 Soccer

04 Migration

02 Venezuela
Politics

05 Environmental
Degradation

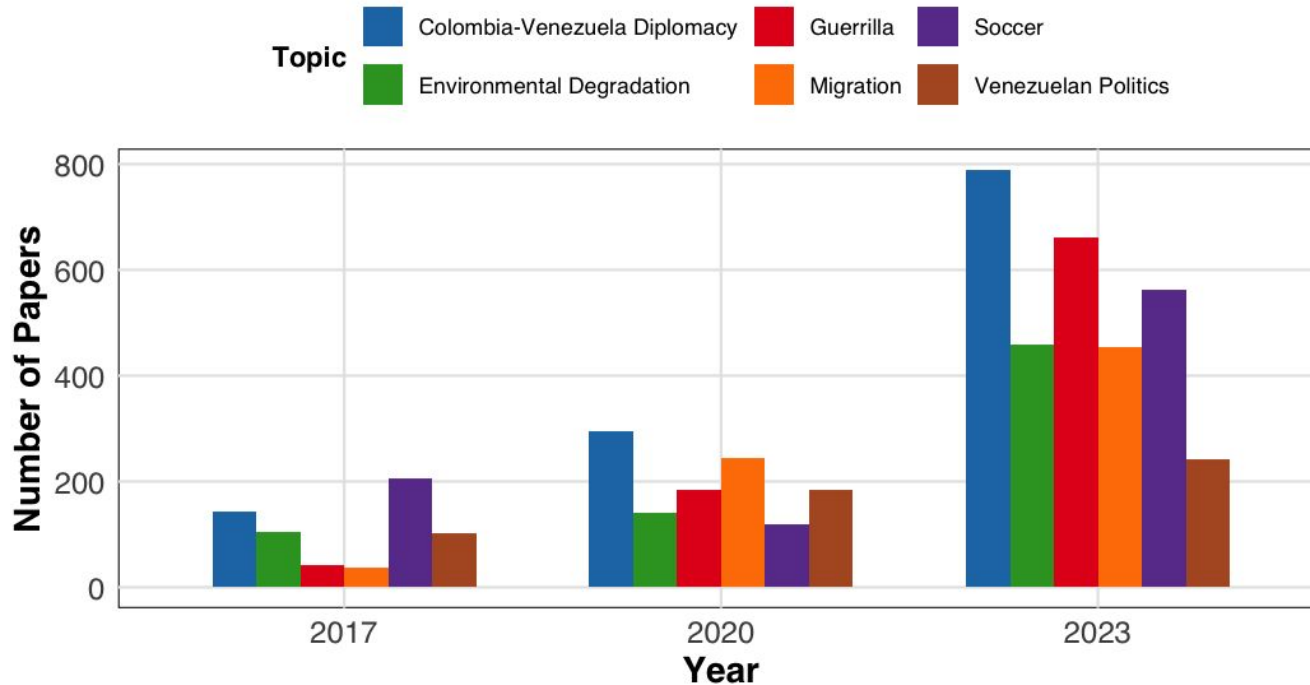
03 Colombia-
Venezuela
Diplomacy

06 Guerrilla

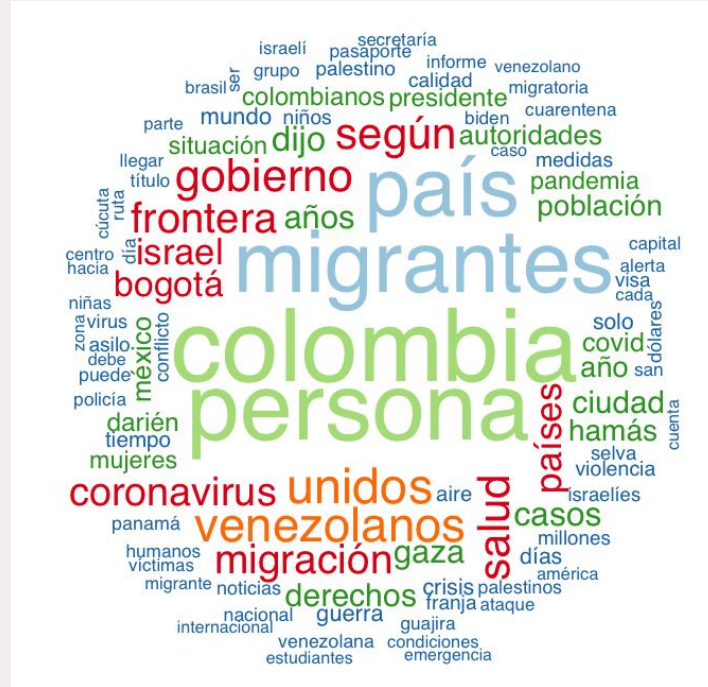
Key Themes

Topic Distribution per Year (Count of Papers)

Number of papers assigned to each topic per year



Zooming into topic 5: Migration



Zooming into the “Migration” topic

Words

- colombia (Colombia)
- persona (person)
- país (country)
- migrantes (migrants)
- venezuela (Venezuela)
- venezolanos (Venezuelans)
- ciento (percent)
- salud (health)
- según (according to)
- frontera (border)
- migración (migration)
- países (countries)
- millones (millions)
- año (year)
- coronavirus (coronavirus)
- población (population)
- casos (cases)
- unidos (United States)
- situación (situation)
- merlano (Merlano)

Titles

Migrantes venezolanos no tienen protección ante el coronavirus en Norte de Santander

(Venezuelan migrants lack protection against coronavirus in Norte de Santander)

Así está la situación en la frontera de EE. UU. tras una semana del fin del Título 42

(This is the situation on the U.S. border a week after the end of Title 42)

Accidentes viales, otro flagelo que enfrentan los migrantes, además del Darién

(Traffic accidents: another scourge faced by migrants, in addition to the Darién)

Amnistía Internacional pidió a Colombia reconocer como refugiados a venezolanos

(Amnesty International called on Colombia to recognize Venezuelans as refugees)

La Guajira atraviesa otra ola migratoria, esta vez durante la pandemia

(La Guajira faces another migration wave, this time during the pandemic)

Assigning Topic Names

Sat 14th, 2024

Top Words

Edition: 001

Topic 1

- israel (Israel)
- gaza (Gaza)
- hamás (Hamas)
- mundo (World)
- guerra (War)
- presidente (President)
- persona (Person)
- franja (Strip)
- palestino (Palestinian)
- unidos (United)
- dijo (Said)
- según (According to)
- israelíes (Israelis)
- conflicto (Conflict)
- noticias (News)
- gobierno (Government)
- palestinos (Palestinians)
- ataque (Attack)
- país (Country)
- israelí (Israeli)

Topic 2

- venezolanos (Venezuelans)
- persona (Person)
- migrantes (Migrants)
- autoridades (Authorities)
- ciudad (City)
- país (Country)
- dijo (Said)
- según (According to)
- frontera (Border)
- años (Years)
- policía (Police)
- grupo (Group)
- gobierno (Government)
- colombia (Colombia)
- tiempo (Time)
- centro (Center)
- crisis (Crisis)
- venezolano (Venezuelan)
- zona (Zone)
- derechos (Rights)

Topic 3

- migrantes (Migrants)
- persona (Person)
- darién (Darien)
- migración (Migration)
- dijo (Said)
- selva (Jungle)
- colombia (Colombia)
- dólares (Dollars)
- panamá (Panama)
- unidos (United)
- día (Day)
- frontera (Border)
- años (Years)
- llegar (Arrive)
- días (Days)
- ruta (Route)
- solo (Only)
- año (Year)
- cada (Each)
- hacia (Toward)

Topic 4

- colombia (Colombia)
- venezolanos (Venezuelans)
- migrantes (Migrants)
- migración (Migration)
- país (Country)
- población (Population)
- persona (Person)
- frontera (Border)
- gobierno (Government)
- millones (Millions)
- situación (Situation)
- venezolana (Venezuelan)
- países (Countries)
- crisis (Crisis)
- migratoria (Migratory)
- colombianos (Colombians)
- cúcuta (Cucuta)
- migrante (Migrant)
- según (According to)
- internacional (International)

Topic 5

- coronavirus (Coronavirus)
- salud (Health)
- país (Country)
- casos (Cases)
- covid (Covid)
- pandemia (Pandemic)
- colombia (Colombia)
- persona (Person)
- gobierno (Government)
- medidas (Measures)
- virus (Virus)
- países (Countries)
- cuarentena (Quarantine)
- presidente (President)
- días (Days)
- brasil (Brazil)
- caso (Case)
- autoridades (Authorities)
- emergencia (Emergency)
- según (According to)

Topic 6

- bogotá (Bogotá)
- aire (Air)
- ciudad (City)
- calidad (Quality)
- colombia (Colombia)
- salud (Health)
- puede (Can)
- nacional (National)
- pasaporte (Passport)
- san (San)
- guajira (Guajira)
- capital (Capital)
- alerta (Alert)
- secretaría (Secretariat)
- país (Country)
- condiciones (Conditions)
- según (According to)
- debe (Must)
- cuenta (Account)
- estudiantes (Students)

Assigning Topic Names

Sat 14th, 2024

Top Words

Edition: 001

Topic 7

- persona (Person)
- mujeres (Women)
- derechos (Rights)
- colombia (Colombia)
- violencia (Violence)
- años (Years)
- niños (Children)
- país (Country)
- salud (Health)
- casos (Cases)
- situación (Situation)
- población (Population)
- humanos (Humans)
- informe (Report)
- niñas (Girls)
- ser (To be)
- víctimas (Victims)
- parte (Part)
- solo (Only)
- año (Year)

Topic 8

- unidos (United)
- colombia (Colombia)
- méxico (Mexico)
- migrantes (Migrants)
- persona (Person)
- países (Countries)
- país (Country)
- frontera (Border)
- colombianos
(Colombians)
- asilo (Asylum)
- visa (Visa)
- año (Year)
- gobierno (Government)
- título (Title)
- biden (Biden)
- migración (Migration)
- américa (America)
- según (According to)
- tiempo (Time)
- derechos (Rights)

Assigning Topic Names

Sat 14th, 2024

Context

Edition: 001

Article titles

Topic 2

Lo que se sabe del naufragio de inmigrantes venezolanos
(What is known about the Venezuelan migrant shipwreck)

Al menos siete muertos tras ser arrollados por conductor en Brownsville, Texas
(At least seven dead after being run over by a driver in Brownsville, Texas)

Topic 3

La industria migratoria del Darién, un lucrativo negocio
(The migration industry of the Darién, a lucrative business)

El pueblo de Colombia donde algunos se vuelven millonarios con la tragedia de migrantes
(The Colombian town where some become millionaires from the migrant tragedy)

Topic 7

El grupo de jóvenes que buscan abolir la explotación sexual en Medellín
(The group of young people seeking to abolish sexual exploitation in Medellín)

En el Concejo piden investigar explotación sexual de migrantes por el Tren de Aragua
(In the Council, they call for an investigation into sexual exploitation of migrants by the Tren de Aragua)

Topic Names within “Migrants”

Sat 14th, 2024

Topic naming

Edition: 001

Final
assigned
topic names

01 Gaza

02 Colombia-
Venezuela
Border Security

03 Darien
Migration
Route

04 Venezuela
Migration Crisis

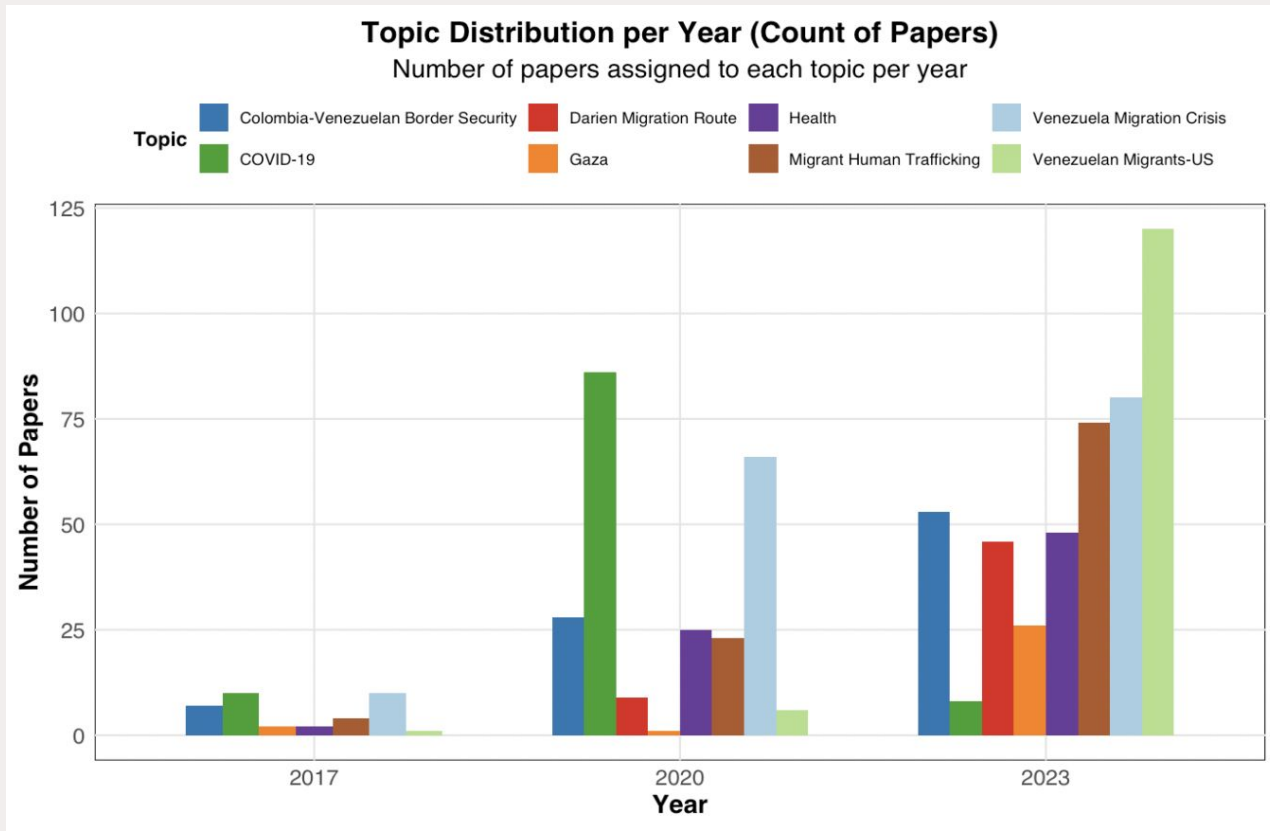
05 COVID-19

06 Health

07 Migrant
Human
Trafficking

08 Venezuela
Migrants-US

Key Migration Themes



Limitations

Sat 14th, 2024

Edition: 001

Lemmatizing

01

02

**Fixed Number of
Topics**

Misinterpretation

03

04

Nexis Uni

**Computationally
expensive**

05

06

**Misclassification
of Documents**

Thanks!

Maria, Emanuel and Nibia

Stats 456

Sat 14th, 2024

