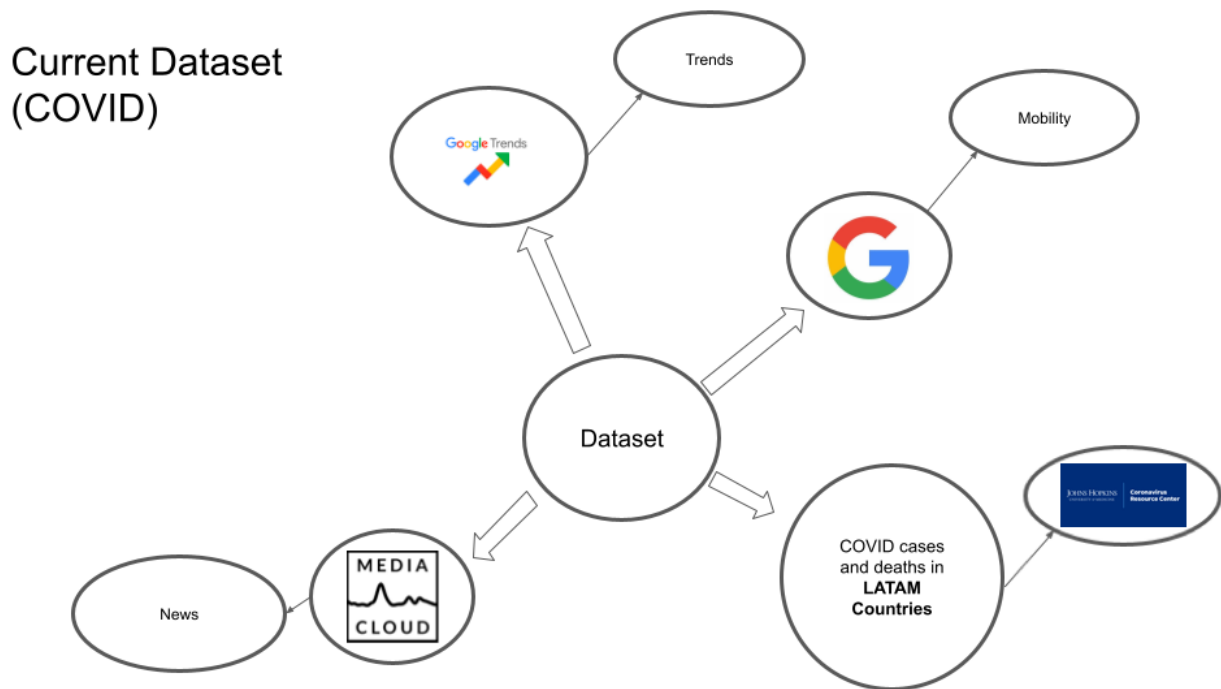


## Infodemic and misinformation in the spread of COVID-19 in LATAM.



The dataset is composed by 3 different sources.

1. COVID-19 data in LATAM from Our World in Data
2. Internet sources which are:
  - 2.1. Media Cloud
  - 2.2. Google Trends

### 1. COVID-19 data in LATAM

The LATAM COVID data is made up of 7 different datasets:

- **covid\_latam(raw).csv:** Contains COVID data for 32 LATAM countries. The dataset contains the columns:
  - 'iso\_code': ISO country code in 3 characters.
  - 'continent': Continent where the data comes from.
  - 'location': Country where the data comes from.
  - 'date': Date on which the data was recorded.
  - 'total\_cases': Total covid cases (cumulative) or probable cases.
  - 'new\_cases': New covid cases presented in the country and established date.

- 'New\_cases\_smoothed': New confirmed cases approximating the current week.
- 'Total\_deaths': Total deaths attributed to covid-19.
- 'New\_deaths': new deaths attributed to covid-19
- 'New\_deaths\_smoothed': new covid-19 deaths approximating the current week.
- 'Total\_cases\_per\_million': Total confirmed cases of COVID-19 per 1,000,000 people.
- 'New\_cases\_per\_million': New confirmed cases of COVID-19 per 1,000,000 people.
- 'New\_cases\_smoothed\_per\_million': New confirmed cases of COVID-19 per 1,000,000 people approximating the current week.
- 'Total\_deaths\_per\_million': Total deaths attributed to COVID-19 per 1,000,000 people
- 'New\_deaths\_per\_million': New deaths attributed to COVID-19 per 1,000,000 people.
- 'New\_deaths\_smoothed\_per\_million': New deaths attributed to COVID-19 approximating the current week per 1,000,000 people.
- 'Reproduction\_rate': Real-time estimate of the effective reproduction rate (R) of COVID-19, that is, this measure is how fast the virus is spreading in a population at a given time.
- 'Icu\_patients': Number of patients with COVID-19 in intensive care units (ICU) on a given day.
- 'Icu\_patients\_per\_million': Number of COVID-19 patients in intensive care units (ICU) on a given day per 1,000,000 people
- 'Hosp\_patients': Number of COVID-19 patients in the hospital on a given day.
- 'hosp\_patients\_per\_million': Number of COVID-19 patients in the hospital on a given day per 1,000,000 people.
- 'Weekly\_icu\_admissions': Number of COVID-19 patients recently admitted to intensive care units (ICUs) in a given week.
- 'Weekly\_icu\_admissions\_per\_million': Number of COVID-19 patients recently admitted to intensive care units (ICUs) in a given week per 1,000,000 people.
- 'Weekly\_hosp\_admissions': Number of newly admitted COVID-19 patients in hospitals in a given week.
- 'weekly\_hosp\_admissions\_per\_million': Number of newly admitted COVID-19 patients in hospitals in a given week per 1,000,000 people
- 'total\_tests': Total tests for COVID-19
- 'New\_tests': New tests for COVID-19 (only calculated for consecutive days).
- 'Total\_tests\_per\_thousand': Total tests for COVID-19 per 1,000 people.
- 'New\_tests\_per\_thousand': New tests for COVID-19 per 1,000 people.
- 'new\_tests\_smoothed': New tests for COVID-19 approximating the current week. For countries that do not report testing data on a daily basis, we assume that testing changed equally on a daily basis and are averaged over a 7-day rolling window.
- 'New\_tests\_smoothed\_per\_thousand': New tests for COVID-19 approximating the current week per 1,000 people
- 'positive\_rate': The proportion of COVID-19 tests that are positive, given as a 7-day moving average (this is the inverse of tests\_per\_case).

- 'tests\_per\_case': Tests performed per new confirmed case of COVID-19, given as a 7-day rolling average (this is the inverse positive\_rate)
- 'tests\_units': Units used by the location to report its test data.
- 'Total\_vaccinations': Total number of COVID-19 vaccination doses administered.
- 'people\_vaccinated': Total number of people who received at least one dose of vaccine.
- 'People\_fully\_vaccinated': Total number of people who received all the doses prescribed by the initial vaccination protocol.
- 'Total\_boosters': Total number of COVID-19 vaccination booster doses administered.
- 'new\_vaccinations': New COVID-19 vaccination doses administered (only calculated for consecutive days).
- 'New\_vaccinations\_smoothed': New COVID-19 vaccination doses administered, given as a 7-day rolling average.
- 'total\_vaccinations\_per\_hundred': Total number of COVID-19 vaccination doses administered per 100 people in the total population.
- 'People\_vaccinated\_per\_hundred': Total number of people who received at least one dose of vaccine per 100 people in the total population.
- 'people\_fully\_vaccinated\_per\_hundred': Total number of people who received all the doses prescribed by the initial vaccination protocol per 100 people in the total population.
- 'Total\_boosters\_per\_hundred': Total number of booster doses of vaccination against COVID-19 administered per 100 people in the total population.
- 'New\_vaccinations\_smoothed\_per\_million': New doses of COVID-19 vaccination administered (7 days smoothed) per 1,000,000 people in the total population.
- 'New\_people\_vaccinated\_smoothed': Daily number of people receiving their first dose of vaccine (7 days smoothed).
- 'new\_people\_vaccinated\_smoothed\_per\_hundred': Daily number of people receiving their first dose of vaccine (smoothed 7 days) per 100 people in the total population.
- 'Stringency\_index': Government Response Stringency Index: Composite measure based on 9 response indicators including school closures, workplace closures and travel bans, rescaled to a value from 0 to 100 (100 = response plus strict)
- 'Population': Population (last available values).
- 'population\_density': Number of people divided by land area, measured in square kilometers, most recent year available
- 'median\_age': Median age of the population, UN projection for 2020
- 'Aged\_65\_older': Proportion of the population aged 65 or older, most recent year available
- 'aged\_70\_older': Percentage of the population aged 70 or older in 2015
- 'gdp\_per\_capita': Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available.
- 'Extreme\_poverty': Proportion of the population living in extreme poverty, most recent year available since 2010.
- 'cardiovasc\_death\_rate': Mortality rate due to cardiovascular disease in 2017 (annual number of deaths per 100,000 people).

- 'diabetes\_prevalence': Prevalence of diabetes (% of the population aged 20 to 79 years) in 2017.
- 'Female\_smokers': Proportion of women who smoke, most recent year available.
- 'male\_smokers': Percentage of men who smoke, most recent year available.
- 'Handwashing\_facilities': Proportion of population with basic handwashing facilities in facilities, most recent year available.
- 'Hospital\_beds\_per\_thousand': Hospital beds per 1,000 people, most recent year available since 2010.
- 'Life\_expectancy': Life expectancy at birth in 2019.
- 'Human\_development\_index': A composite index that measures average achievement in three basic dimensions of human development: a long and healthy life, knowledge, and a decent standard of living.
- 'excess\_mortality\_cumulative\_absolute': Cumulative difference between the reported number of deaths since January 1, 2020, and the projected number of deaths for the same period based on previous years.
- 'Excess\_mortality\_cumulative': Percentage difference between the cumulative number of deaths since January 1, 2020 and the projected cumulative deaths for the same period based on previous years.
- 'excess\_mortality': Percentage difference between the reported number of weekly or monthly deaths in 2020-2021 and the projected number of deaths for the same period based on previous years.
- 'Excess\_mortality\_cumulative\_per\_million': Cumulative difference between the reported number of deaths since January 1, 2020, and the projected number of deaths for the same period, based on previous years, per million people.

**Note:** For more details regarding the calculation of the excess mortality for the different countries, please refer to: <https://elifesciences.org/articles/69336>

- **daily\_covid\_cases\_latam.csv:** Daily COVID cases per million inhabitants for 32 LATAM countries.
  - The first 'location' column contains the name of the country where the cases were filed.
  - The following columns contain the number of cases per million inhabitants presented on that day in that country.
- **daily\_covid\_deaths\_latam.csv:** Daily confirmed COVID deaths per million inhabitants for 32 LATAM countries.
  - The first column 'location' contains the name of the country where the deaths occurred.
  - The following columns contain the number of confirmed deaths per million inhabitants presented on that day in that country.
- **daily\_covid\_vaccinations\_latam.csv:** Daily confirmed COVID vaccinations per hundred inhabitants for 32 LATAM countries (in a cumulative way).
  - The first column 'location' contains the name of the country where the vaccinations were applied.

- The following columns contain the number of confirmed vaccinations per hundred inhabitants applied till that day in that country.
- **Weekly\_covid\_cases\_latam.csv:** Weekly COVID cases using the epidemiological week per million inhabitants for 32 LATAM countries.
  - The first 'location' column contains the name of the country where the cases were filed.
  - The following columns contain the sum of cases per million inhabitants presented in that country during that epidemiological week. The format of the epidemiological week is:
    - First 4 digits corresponding to the year, followed by 2 digits corresponding to the epidemiological week.
- **Weekly\_covid\_deaths\_latam.csv:** Weekly confirmed COVID deaths using the epidemiological week per million inhabitants for 32 LATAM countries.
  - The first column 'location' contains the name of the country where the deaths occurred.
  - The following columns contain the sum of the deaths per million inhabitants presented in that country during that epidemiological week. The format of the epidemiological week is:
    - First 4 digits corresponding to the year, followed by 2 digits corresponding to the epidemiological week.
- **weekly\_covid\_vaccinations\_latam.csv:** Weekly confirmed COVID vaccinations applied using the epidemiological week per hundred inhabitants for 32 LATAM countries.
  - The first column 'location' contains the name of the country where the vaccinations were applied.
  - The following columns contain the sum of the total vaccinations per hundred inhabitants presented in that country till that epidemiological week. The format of the epidemiological week is:
    - First 4 digits corresponding to the year, followed by 2 digits corresponding to the epidemiological week.

**Note:** For weekly documents, each column represents que epiweek in epiweek format (Year-Week). You can use the [epiweek library](#) of python to convert the columns to date format.

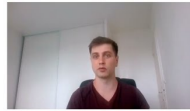
## 2. Internet sources:

### 2.1. Media Cloud

Media cloud dataset was extracted using Explorer tool from Media Cloud. Some examples of the use of the tool can be seen in the videos:

1. [Media Cloud - Topic Mapper](#):

# Media Cloud Topic Mapper



Adrien Carrel



2. [Media Cloud - Explorer](#):

## Media Cloud - Explorer

How it works:



Powerful query tool to search through diverse digital news media sources



Allows search by multiple languages (English, Spanish, Arabic, ...)



Includes temporal information which enables the observation of trends



Features various representations of how topics are covered (word-clouds, bi-grams, embeddings)



Provide additional information about stories to better understand the scope of your topic of interest



The datasets that we find in this section are:

- **covid-19\_attention\_latam.csv**: Attention to the word COVID-19 in LATAM countries.
  - The first column 'date' contains the date on which attention to the word is being analyzed.
  - For each country, we have the following features of the word “covid-19” in each country:
    - “count”: which represents the number of news media mentioning our topic.
    - “total\_count” which is the total number of news media published.

- and "ratio" which indicates the proportion of news media that contain the word we are interested in.
  - The column is presented in the format:
    - count\_NameOfCountry, total\_count\_NameOfCountry, ratio\_NameOfCountry
- **covid\_vaccination\_attention\_latam.csv**: Attention to the word COVID vaccination in LATAM countries.
  - The first column 'date' contains the date on which attention to the word is being analyzed.
  - For each country, we have the following features of the word “COVID vaccination” in each country:
    - “count”: which represents the number of news media mentioning our topic.
    - "total\_count" which is the total number of news media published.
    - and "ratio" which indicates the proportion of news media that contain the word we are interested in.
  - The column is presented in the format:
    - count\_NameOfCountry, total\_count\_NameOfCountry, ratio\_NameOfCountry

## 2.2. Google Trends

The Google Trends dataset was extracted using the Google Trends Online tool. Some examples of the use of the tool can be seen in the video:

### 2. [Google Trends tutorial](#):


## Google Trends

**How it works:**

1. Compares the number of times the keyword appeared in searches with the total number of searches in a specific region and time.
2. The results of the comparison are ranked between 0 and 100 representing the popularity of the topic in a certain time and location.

**Examples:**

1. Forecasting Influenza using google trends [1].
2. Early warning COVID-19 monitor [2].
3. COVID-19 lockdowns and well-being [3].



[1] - Lu P, Hou S, Baltusaitis K, Shah M, Leskovec J, Sosic R, Hawkins J, Brownstein J, Conidi G, Gunn J, Gray J, Zink A, Santillana M. Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. JMIIR Public Health Surveill. 2019;4(1):e4. DOI: 10.2196/publichealth.8950

[2] - Kogan NE, Clemente L, Lautaud P, Kaashoek J, Link NB, Nguyen AT, Lu FS, Huybers P, Resch B, Havas C, Petutschnig A, Davis J, Chinazzi M, Mustafa B, Hanage WP, Vespignani A, Santillana M. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. Sci Adv. 2021 Mar 5;7(10):eabd6989. doi: 10.1126/sciadv.abd6989. PMID: 33674304; PMCID: PMC7935356.

[3]- Brodeur A, Andrew EC, Sarah F, Nattavudh P. (2020). COVID-19, lockdowns and well-being: Evidence from Google Trends. Journal of Public Economics. 193. 10.1016/j.jpubeco.2020.104346.

The dataset that we find in this section is:

- **covid-19\_Symptoms\_latam.csv:** Google trends of some Covid-19 symptoms in LATAM countries.
  - The first column 'Week' contains the date on which symptom trends in LATAM are being analyzed.
  - Each country has Fever, Headache and Cough columns, which represent some of the symptoms of covid-19 in each country. The column is presented in the format:
    - Fever\_NameOfCountry, Headache\_NameOfCountry, Cough\_NameOf-Country