

# Árbol Filogenético de variantes del virus SARS-CoV-2, Mapeo Representativos por País

## Equipo GATTACA:

Ricardo Bernabé Nicolás

Raymundo Méndez García

Fernando Sánchez Olivares

Ivanna Saucedo Dorantes

## Introducción

El SARS-CoV-2 es un virus perteneciente a la familia de los coronavirus, y presenta una estructura muy similar al SARS-CoV al igual que al MERS-CoV (Síndrome respiratorio del Oriente Medio). Este se caracteriza por ser un virus de RNA monocatenario y de cadena positiva. Al igual que con los otros representantes de esta familia se estima que su origen es quiróptero, ya que comparte del 89-96% de identidad de nucleótidos con el coronavirus del murciélago (Clerkin, *et al.*, 2020; Shang, *et al.*, 2020). Por lo anterior, la enfermedad provocada por este agente se considera de tipo zoonosis. De manera general, esta zoonosis se manifiesta con problemas respiratorios intermedios a severos que pueden causar la muerte de las personas infectadas.

Una particularidad desde el punto de vista genómico de este virus es que su genoma es el más grande conocido para un virus de RNA hasta la fecha. Entre sus características más importantes se encuentra que el extremo 3' del genoma codifica para 4 proteínas importantes: proteínas de la espiga (S), envoltura (E), membrana (M) y nucleocápside (N), además de proteínas accesorias que son específicas del género y pueden ayudar a SARS-CoV2 a evadir el sistema inmunológico o incluso a aumentar su potencial de virulencia (Li, *et al.* 2020).

Varias de las proteínas anteriormente mencionadas son esenciales para el ciclo de vida del virus. Tal como es el caso de S, la proteína causante de la infección de las células humanas al ser capaz de unirse al receptor de la angiotensina-convertisina 2 humana (ACE-2). ACE-2 es expresada principalmente en los pulmones y el corazón, así como el epitelio intestinal, el endotelio vascular y en los riñones, lo cual le confiere a este virus una capacidad de infección multiorgánica dificultando de esta forma el diagnóstico de infección por SARS-CoV2 causante de COVID-19 (Clerkin *et al.*, 2020).

Hay muchas formas en que se clasifican los virus del SARS-CoV-2. Cada tipo de clasificación puede ser adecuada en función del contexto en el que se comunica el SARS-CoV-2. A menudo se habla del SARS-CoV-2 en el contexto de los linajes (y sublinajes). El sistema de clasificación de linajes más utilizado es el Pango. Nextclade también puede utilizarse en este contexto. En un contexto más amplio, la Organización Mundial de la Salud (OMS) puede clasificar los linajes o grupos de linajes relacionados mediante el uso de letras griegas (como ómicron). Estos métodos de clasificación permiten que los científicos puedan comunicar las similitudes y diferencias entre los virus SARS-CoV-2 (CDC, 2023).

Para el presente trabajo se utilizaron las variantes omicron, alpha, gamma, delta, epsilon, eta, iota, kappa, zeta, mu, para realizar un árbol filogenético que muestre las diferencias filogenéticas entre cada una de las variantes del SARS-CoV2, también se realizó un mapeo representativo por país tomando en cuenta diferentes variables que se han presentado en el mundo debido al COVID-19.

## Objetivo

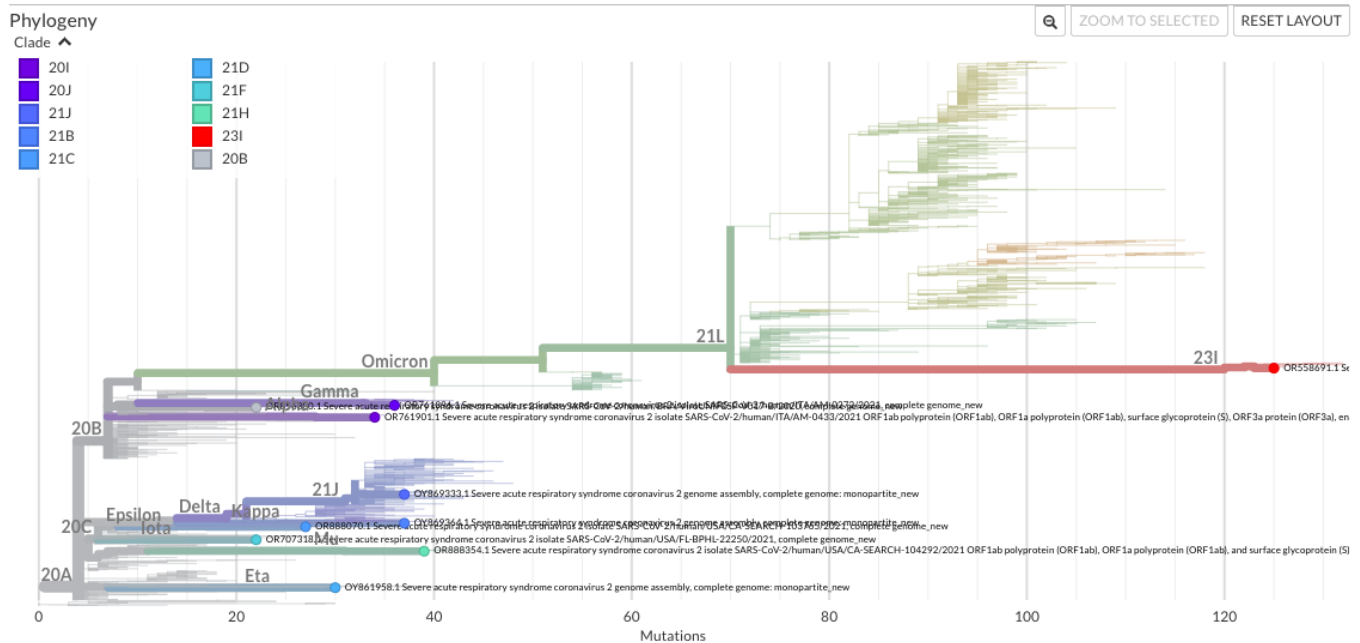
Mostrar un árbol filogenético del SARS-CoV2 y hacer un mapeo representativo de número de casos y número de muertes en todo el mundo y también trataremos de comprobar la siguiente hipótesis:

**“El gde per capita se relaciona directamente con el número de muertes de cada país.”**

## Análisis filogenético

Para el análisis filogenético de las variantes de SARS-CoV2 se seleccionaron 10 variantes reconocidas por la OMS omicron (23I), alpha (20I), gamma (20J), delta (21J), epsilon (21C), eta (21D), iota (21F), kappa (21B), zeta (20B), mu (21H) y a partir de las secuencias que podemos encontrar en el genbank y corresponden a personas de diferentes partes del mundo.

Se obtuvo el siguiente árbol filogenético:



Podemos observar 4 grandes clados, entre los cuales está la variante eta, en otro clado se observa la variante mu siendo esta la que más mutaciones presenta dentro del clado, iota, epsilon, kappa y delta, podríamos decir que estas variantes se encuentran más emparentadas, en otro clado encontramos a las variantes alpha, gamma, y zeta, estas 3 han sido clasificadas por la OMS como variantes “de preocupación”, y por último encontramos que la variante que presenta una diferencia de mutaciones significativas es la omicron más lejana.

## Mapeo de variantes por país

Al inicio del proyecto queríamos hacer una relación entre muertes por covid/ variante. pero resulta que obtener ese tipo de información resulta complicada de obtener debido a la naturaleza del evento que se estaba analizando.

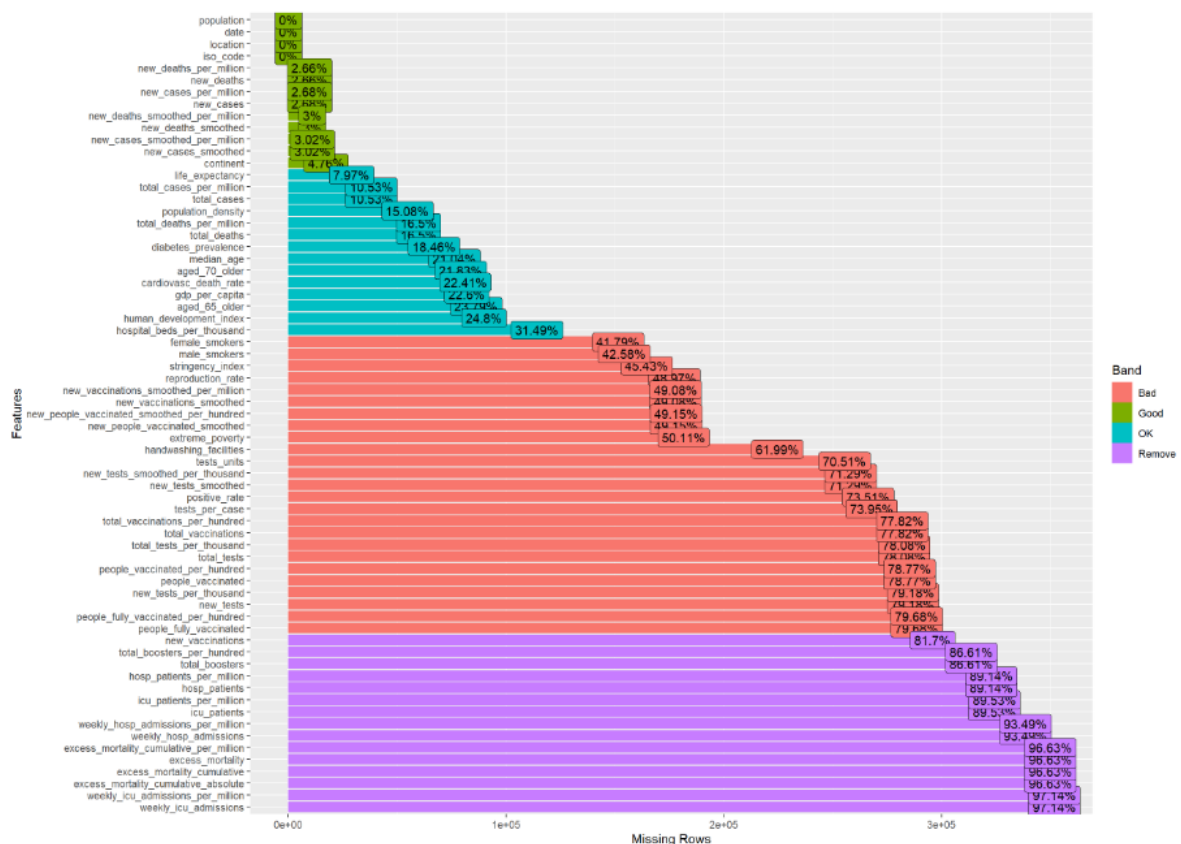
El dataset que fué utilizado fué el [siguiente](#)

Es un dataset que contiene la información general de cada país, actualizado diariamente, usamos la información recopilada hasta el 6 de diciembre del 2023.

Entre toda las variables que el dataset muestra, nos vamos a enfocar en las siguientes:

- Casos totales (Acomulados en tiempo)
- Muertes totales (Acomulados en tiempo)
- Población total
- Edad media
- Camas por mil habitantes
- Muertes por millón de habitantes

## Missing Data Profile



Observemos que la mayoría de las columnas del dataset vienen vacías, por lo que no les vamos a dar importancia para este análisis.

## Análisis Exploratorio

Podemos notar que las variables con alta correlación, por ejemplo:

Población / Casos totales

Casos totales/Muertes totales

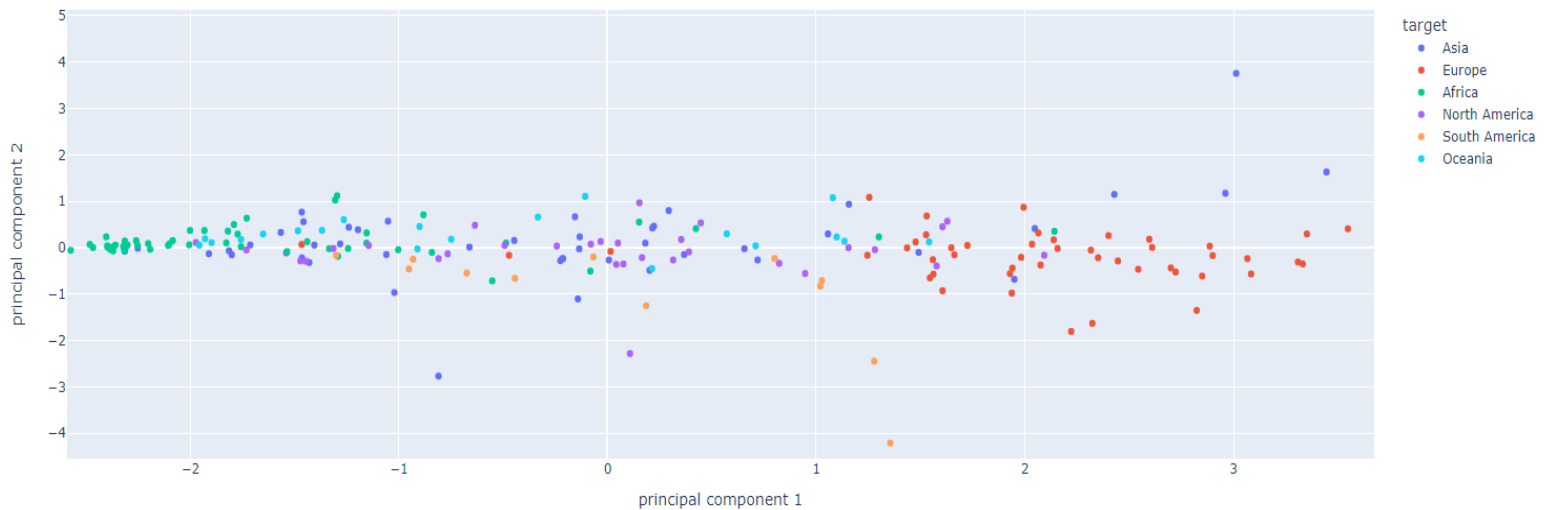
Muertes totales / Población

casos por millón/ media de edad

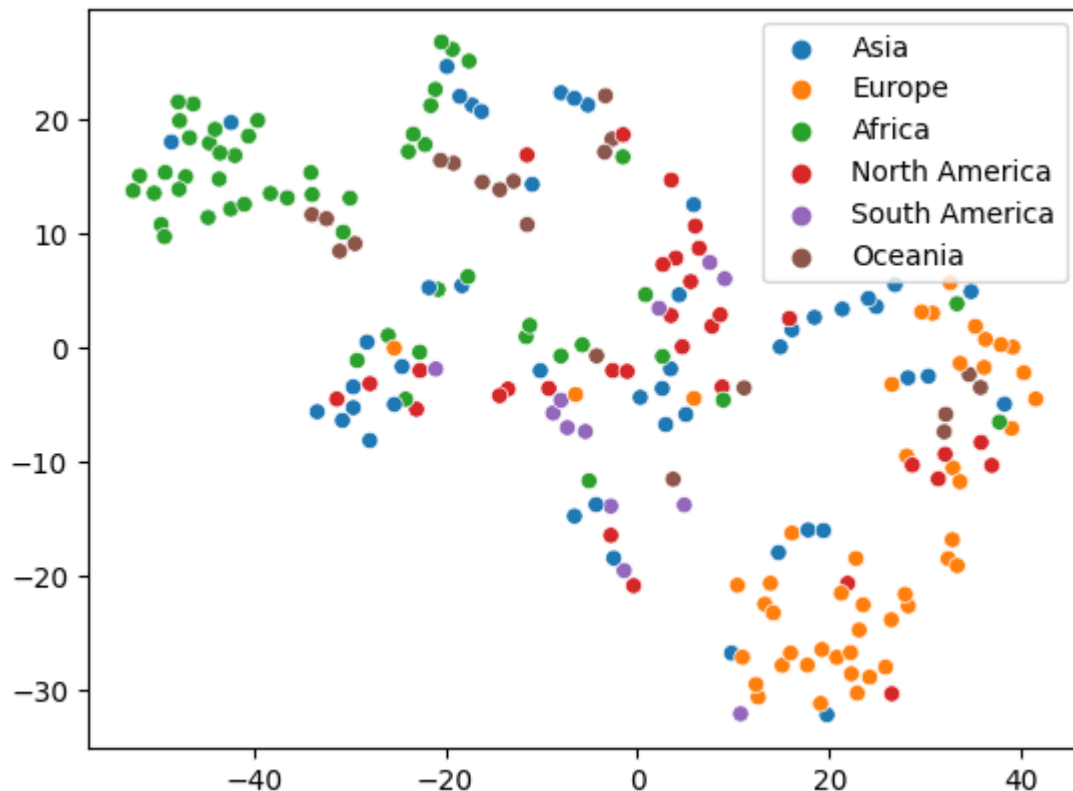
casos por millón/ gdp

	total_cases	total_deaths	total_cases_per_million	total_deaths_per_million	population_density	median_age	gdp_per_capita	hospital_beds_per_thousand	population
total_cases	1.00	0.76	0.12	0.14	-0.02	0.29	0.19	0.21	0.70
total_deaths	0.76	1.00	0.03	0.28	-0.04	0.19	0.13	0.06	0.46
total_cases_per_million	0.12	0.03	1.00	0.47	0.17	0.76	0.65	0.47	-0.12
total_deaths_per_million	0.14	0.28	0.47	1.00	-0.00	0.68	0.28	0.39	-0.08
population_density	-0.02	-0.04	0.17	-0.00	1.00	0.14	0.28	0.29	-0.02
median_age	0.29	0.19	0.76	0.68	0.14	1.00	0.65	0.61	0.03
gdp_per_capita	0.19	0.13	0.65	0.28	0.28	0.65	1.00	0.29	-0.05
ospital_beds_per_thousand	0.21	0.06	0.47	0.39	0.29	0.61	0.29	1.00	-0.04
population	0.70	0.46	-0.12	-0.08	-0.02	0.03	-0.05	-0.04	1.00

Para poder visualizar nuestros datos, optamos por utilizar PCA clasificando nuestros datos por continente.

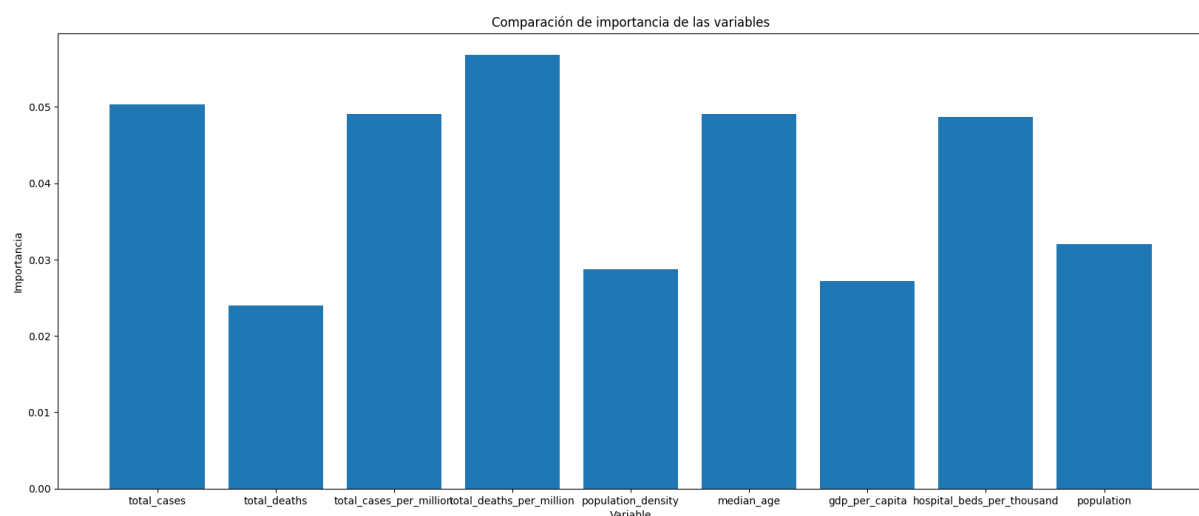


No tenemos una visualización clara de nuestro conjunto de datos, por lo vamos a utilizar T-SNE como herramienta de visualización.



En esta imagen podemos observar que países de África están relacionados entre sí y que en Europa algunos países están un poco relacionados, pero no lo suficiente para formar un cluster claro. por lo tanto suponemos que los datos tienen poca relación entre sí dado su continente.

También usamos el clasificador ExtraTrees para ver cuáles de nuestras variables eran las más importantes.



Podemos observar que, el gdp\_per\_capita rankea bajo en importancia y que cosas como las camas por mil habitantes, la media de edad y la población son más relevantes.

## Resultados:

En esta sección mostraremos los conteos resumidos del 6 de diciembre del 2023, Los clasificaremos por continente y los ordenaremos por GDP (PIB)

### ÁFRICA

continent	location	total_cases	total_deaths	total_deaths_per_million	gdp_per_capita	hospital_beds_per_thousand	population
Africa	Seychelles	51058	172	1605.451	26382.287	3.6	107135
Africa	Equatorial Guinea	17130	183	109.259	22604.873	2.1	1674916
Africa	Mauritius	311465	1055	811.864	20292.745	3.4	1299478
Africa	Libya	507269	6437	944.902	17881.509	3.7	6812344
Africa	Gabon	49046	307	128.506	16562.413	6.3	2388997
Africa	Botswana	330397	2800	1064.517	15807.374	1.8	2630300
Africa	Algeria	271997	6881	153.241	13913.839	1.9	44903228
Africa	South Africa	4072575	102595	1712.946	12294.876	2.32	59893884
Africa	Tunisia	1153361	29423	2381.25	10849.297	2.3	12356116
Africa	Egypt	516023	24830	223.714	10550.206	1.6	110990096

Podemos notar que el GDP no tiene relación directa con el número de camas pero pareciera que sí con el número de muertes totales y con el número de casos totales.

### ASIA

continent	location	total_cases	total_deaths	total_deaths_per_million	gdp_per_capita	hospital_beds_per_thousand	population
Asia	Qatar	514524	690	256.017	116935.6	1.2	2695131
Asia	Singapore	2739037	1933	342.912	85535.383	2.4	5637022
Asia	Brunei	315733	164	365.254	71809.251	2.7	449002
Asia	United Arab Emirates	1067030	2349	248.805	67293.483	1.2	9441138
Asia	Kuwait	666538	2570	602.031	65530.537	2	4268886
Asia	Saudi Arabia	841469	9646	264.936	49045.411	2.7	36408824
Asia	Bahrain	696614	1536	1043.31	43290.705	2	1472237
Asia	Japan	33803572	74694	602.606	39002.223	13.05	123951696
Asia	Oman	399449	4628	1011.297	37960.709	1.6	4576300
Asia	South Korea	34571873	35934	693.495	35938.374	12.27	51815808
Asia	Israel	4841772	12707	1344.798	33132.32	2.99	9449000
Asia	Malaysia	5140563	37201	1096.139	26808.164	1.9	33938216
Asia	Turkey	17004677	101419	1188.394	25129.341	2.81	85341248
Asia	Kazakhstan	1502857	19072	983.194	24055.588	6.7	19397998
Asia	Iran	7624699	146724	1656.952	19082.62	1.5	88550568
Asia	Turkmenistan				16389.023	7.4	6430777
Asia	Thailand	4760122	34495	481.122	16277.671	2.1	71697024
Asia	Azerbaijan	834135	10367	1000.861	15847.419	4.7	10358078
Asia	Iraq	2465545	25375	570.274	15663.986	1.4	44496124
Asia	China	99320723	121877	85.474	15308.712	4.34	1425887360

Podemos notar que el GDP no tiene relación directa con el número de camas, pero pareciera que sí con el número de muertes totales y con el número de casos totales.

## EUROPA

continent	location	total_cases	total_deaths	total_deaths_per_million	gdp_per_capita	hospital_beds_per_thousand	population
Europe	Luxembourg	386030	1000	1544.161	94277.965	4.51	647601
Europe	Ireland	1725026	9366	1864.583	67335.293	2.96	5023108
Europe	Norway	1489459	5732	1054.777	64800.057	3.6	5434324
Europe	Switzerland	4425596	14086	1611.584	57410.166	4.53	8740471
Europe	San Marino	25292	126	3739.982	56861.47	3.8	33690
Europe	Netherlands	8620051	22986	1308.698	48472.545	3.32	17564020
Europe	Sweden	2728931	25332	2401.286	46949.283	2.22	10549349
Europe	Denmark	3417017	8824	1500.104	46682.515	2.5	5882259
Europe	Iceland	209328	186	498.789	46482.958	2.91	372903
Europe	Austria	6081287	22534	2520.69	45436.686	7.37	8939617
Europe	Germany	38437756	174979	2098.829	45229.245	8	83369840
Europe	Belgium	4829921	34339	2946.056	42658.576	5.64	11655923
Europe	Finland	1499712	10864	1960.747	40585.721	3.28	5540745
Europe	United Kingdom	24812582	232112	3438.241	39753.244	2.54	67508936
Europe	France	38997490	167985	2599.316	38605.671	5.98	67813000
Europe	Malta	120705	885	1659.5	36513.323	4.485	533293
Europe	Italy	26318717	192909	3267.569	35220.084	3.18	59037472
Europe	Spain	13980340	121852	2562.143	34272.36	2.97	47558632

Podemos notar que el GDP no tiene relación directa con el número de camas, pero pareciera que sí con el número de muertes totales y con el número de casos totales.

## NORTEAMÉRICA

continent	location	total_cases	total_deaths	total_deaths_per_million	gdp_per_capita	hospital_beds_per_thousand	population
North America	United States	103436829	1144877	3384.308	54225.446	2.77	338289856
North America	Bermuda	18860	165	2569.813	50669.315		64207
North America	Cayman Islands	31472	37	538.401	49903.029		68722
North America	Canada	4731777	53255	1384.89	44017.591	2.5	38454328
North America	Sint Maarten (Dutch part)	11051	92	2081.825	36327.232		44192
North America	Aruba	44224	292	2742.84	35973.781		106459
North America	Puerto Rico	1252713	5938	1825.722	35044.67		3252412
North America	Trinidad and Tobago	191496	4390	2867.326	28763.071	3	1531043
North America	Bahamas	38084	844	2058.592	27717.847	2.9	409989
North America	Saint Kitts and Nevis	6607	46	964.745	24654.385	2.3	47681
North America	Panama	1041696	8651	1962.309	22267.037	2.3	4408582
North America	Antigua and Barbuda	9106	146	1556.968	21490.943	3.8	93772
North America	Mexico	7702387	334917	2626.715	17336.469	1.38	127504120
North America	Barbados	110132	593	2105.48	16978.068	5.8	281646
North America	Costa Rica	1230653	9368	1808.202	15524.995	1.13	5180836
North America	Dominican Republic	661180	4384	390.424	14600.861	1.6	11228821
North America	Grenada	19693	238	1897.034	13593.877	3.7	125459
North America	Saint Lucia	30209	410	2279.399	12951.839	1.3	179872

Podemos notar que el GDP no tiene relación directa con el número de camas, pero pareciera que sí con el número de muertes totales y con el número de casos totales, de hecho este caso es raro, pues las estadísticas indican que a mayor GDP, mayor el número de muertes totales y casos totales.

## OCEANÍA

continent	location	total_cases	total_deaths	total_deaths_per_million	gdp_per_capita	hospital_beds_per_thousand	population
Oceania	Australia	11678853	23766	907.882	44648.71	3.84	26177410
Oceania	New Zealand	2422392	3549	684.436	36085.843	2.61	5185289
Oceania	Palau	6249	9	497.678	13240.405	4.8	18084
Oceania	Nauru	5393	1	78.796	12895.635	5	12691
Oceania	Fiji	69047	885	951.849	8702.975	2.3	929769
Oceania	Samoa	16780	31	139.395	6021.557		222390
Oceania	Tonga	16888	12	112.289	5425.621	2.6	106867
Oceania	Papua New Guinea	46864	670	66.058	3823.194		10142625
Oceania	Marshall Islands	16138	17	408.723	3819.202	2.7	41593
Oceania	Tuvalu	2943	1	88.222	3575.104		11335
Oceania	Micronesia (country)	26547	65	569.287	3299.464		114178
Oceania	Vanuatu	12019	14	42.847	2921.909		326744
Oceania	Solomon Islands	25954	199	274.759	2205.923	1.4	724272
Oceania	Kiribati	5085	24	182.875	1981.132	1.9	131237
Oceania	Cook Islands	7174	2	117.426			17032
Oceania	French Polynesia	78700	649	2118.893			306292
Oceania	Guam	52187	417	2427.481			171783
Oceania	New Caledonia	80064	314	1082.912			289959
Oceania	Niue	895					1952
Oceania	Northern Mariana Islands	14334	41	827.046			49574
Oceania	Pitcairn	4					47
Oceania	Tokelau	80					1893
Oceania	Wallis and Futuna	3550	8	689.893			11596

Podemos notar que el GDP no tiene relación directa con el número de camas, pero pareciera que sí con el número de muertes totales y con el número de casos totales, de hecho este caso es raro, pues las estadísticas indican que a mayor GDP, mayor el número de casos totales, sin embargo el número de muertes totales pareciera no seguir un patrón.

## AMÉRICA DEL SUR

continent	location	total_cases	total_deaths	total_deaths_per_million	gdp_per_capita	hospital_beds_per_thousand	population
South America	Chile	5319739	61996	3162.458	22767.037	2.11	19603736
South America	Uruguay	1040596	7656	2236.768	20551.409	2.8	3422796
South America	Argentina	10072905	130678	2871.392	18933.907	5	45510324
South America	Venezuela	552695	5856	206.913	16745.022	0.8	28301700
South America	Brazil	37519960	702116	3260.901	14103.452	2.2	215313504
South America	Suriname	82506	1404	2271.676	13767.119	3.1	618046
South America	Colombia	6383948	142727	2751.415	13254.949	1.71	51874028
South America	Peru	4520102	221564	6507.098	12236.706	1.6	34049588
South America	Ecuador	1065377	36036	2001.889	10581.936	1.5	18001002
South America	Paraguay	736118	19933	2939.647	8827.01	1.3	6780745
South America	Guyana	73673	1299	1606.228	7435.047	1.6	808727
South America	Bolivia	1209626	22407	1833.016	6885.829	1.1	12224114
South America	Falkland Islands	1923					3801
South America	French Guiana	98041	413	1356.019			304568

Podemos notar que el GDP no tiene relación directa con el número de camas, pero pareciera que sí con el número de muertes totales y con el número de casos totales, de hecho este caso es raro, pues las estadísticas indican que a mayor GDP, mayor el número de muertes totales y casos totales.



## Datos concluyentes

Al tratarse de un proyecto global, la recabación de datos, la metodología de recabación de datos, corresponde a la voluntad humana y política específica de la región que está siendo analizada; muchos de los datos necesarios para la obtención de un árbol filogenético por países, no son de fácil acceso o fueron recabados. Por otro lado se encontraron resultados que llamaron la atención, de acuerdo a la anormalidad que representan, como lo fue la última ola de severidad del covid en China del mes enero de 2023 con un repunte de casos positivos Covid, así como un repunte de muertes confirmadas a partir de dicha enfermedad, que contrastado con los datos analizados previamente, coinciden con los datos recopilados por la OMS.

### Caso China

Se confirmó la muerte de 60.000 personas en el periodo del 8 de diciembre al 12 de enero. El pico máximo de pacientes hospitalizados de covid en estado crítico alcanzó su máximo el 5 de enero y descendió a 44.3% hasta el 17 de enero. Una de las causas de preocupación del aumento de casos se debió al Año Nuevo Lunar que comienza el 20 de enero y termina el 28, donde la población regresa a sus ciudades de origen, generando grandes movilizaciones de personas.

#### China Situation

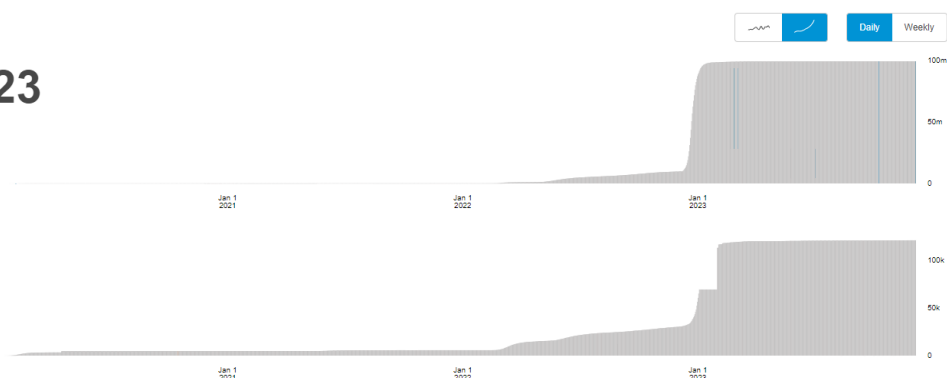
**99.320.723**

confirmed cases

**121.877**

deaths

Source: World Health Organization



<https://covid19.who.int/region/wpro/country/cn>

Así mismo, pudimos observar como el Gross Domestic Product de cada región no tuvo una relación directa con el número de casos totales confirmados, ni con el número de muertes. En otras palabras, la propagación del virus y sus variantes, no dependen de las condiciones del desarrollo económico del país, sino de los factores de propagación biológicos propios del virus e, incluso factores como la cantidad de camas, la población y la edad son más relevantes en este sentido.

## Referencias

-Clasificaciones y definiciones de las variantes del SARS-CoV-2. (2023, septiembre). Centro para el Control y la prevención de enfermedades.

-<https://espanol.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>

Clerkin, K. J., Fried, J. A., Raikhelkar, J., Sayer, G., Griffin, J. M., Masoumi, A., Jain, S. S.,

-Burkhoff, D., Kumaraiah, D., Rabbani, L., Schwartz, A., & Uriel, N. (2020). COVID-19 and Cardiovascular Disease. *Circulation*, 141(20), 1648–1655.

<https://doi.org/10.1161/CIRCULATIONAHA.120.046941>

-Li, Y.-D., Chi, W.-Y., Su, J.-H., Ferrall, L., Hung, C.-F., & Wu, T.-C. (2020). Coronavirus vaccine development: from SARS and MERS to COVID-19. *Journal of Biomedical Science*, 27(1). doi:10.1186/s12929-020-00695-2