



Trabajo Práctico Final

Covid: causas y consecuencias a nivel mundial

Jonás Levy Alfie, Sofía Massobrio,
Sebastián Taboh y Juan Tollo

Trabajo
exploratorio:
incidencia Covid-19
por países

Data set principal: Covid-vaccinations

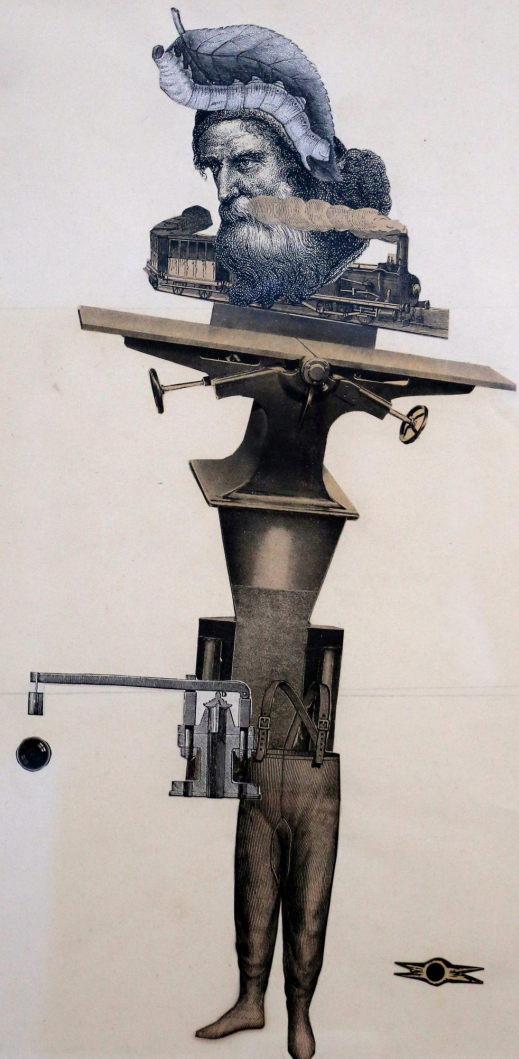
Información sociodemográfica y sanitaria desde el inicio de la pandemia:

- Por país
- Por día
- Variables que nos interesa estudiar y factores que puedan explicarlos

Nos interesa comparar las *situaciones* respecto del Covid en distintos países, representadas mediante las siguientes variables

- Casos por millón
- Muertes por millón
- Testeos por mil
- Vacunación por cien

Lógica del trabajo exploratorio:
Una pregunta, una respuesta,
y una pregunta en base a esa respuesta.



Desarrollo

Pregunta: ¿los países que fueron colonia de un mismo imperio tuvieron comportamiento similar?



Agrupamiento por colonizadores


Reducción de dimensionalidad
con PCA

- Agregamos un segundo dataset
- Países de los que eran colonias
- Trabajamos sobre las 4 variables mencionadas
- Redujimos de 4 variables relevadas a 2 dimensiones

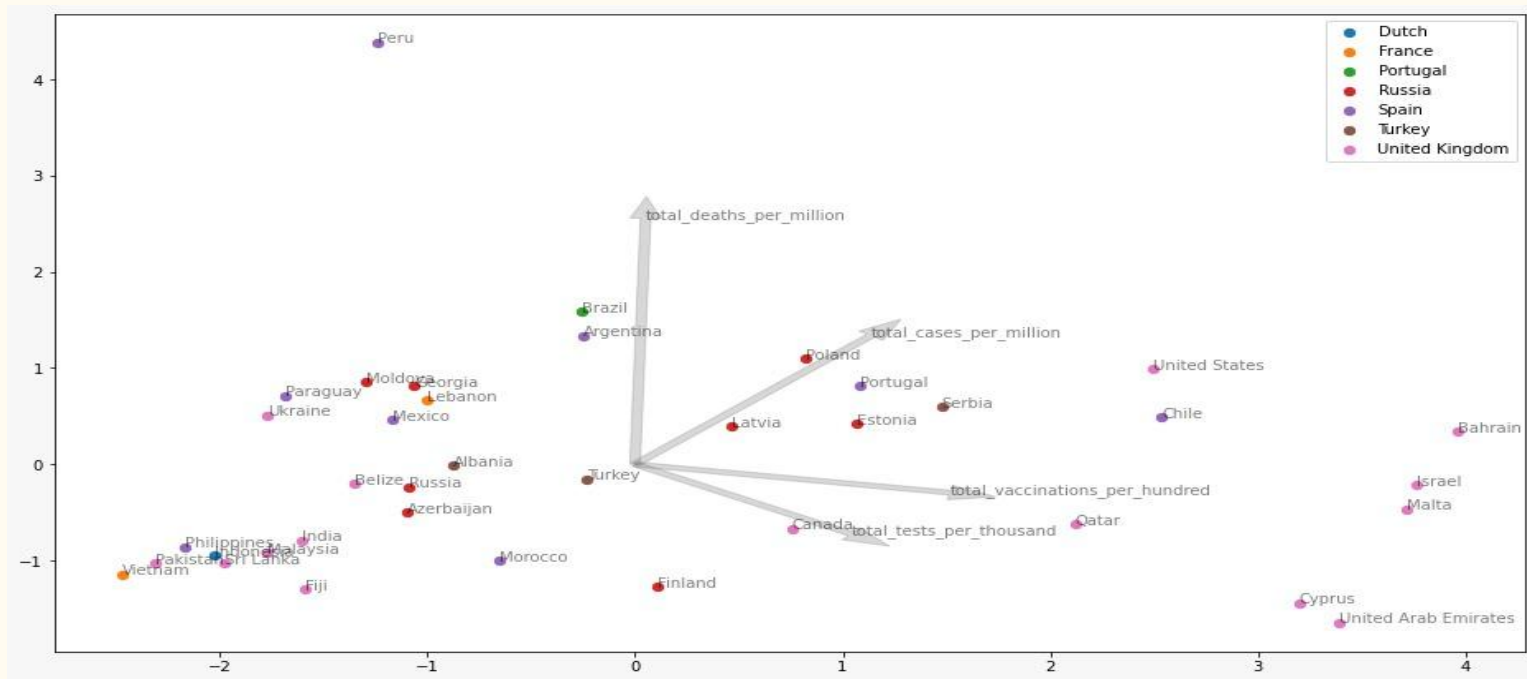
© Dataset

Independence days

Independence days around the world

 Rachael Tatman • updated 4 years ago (Version 2)

No vemos un agrupamiento según Imperio...



pero ¿habrá agrupamiento por regiones para nuestros targets?

Pregunta: ¿el Covid-19 afectó de manera diferente entre regiones a nivel mundial?

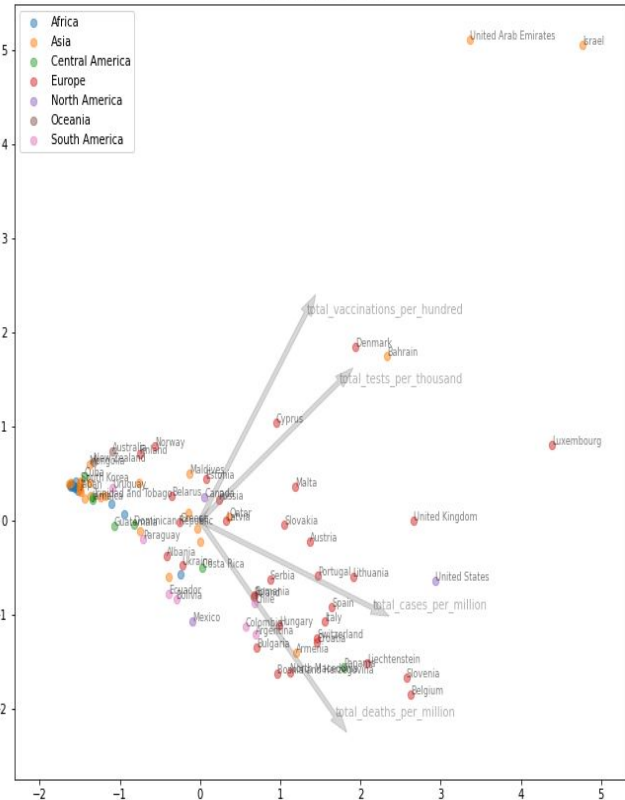
—

Agrupamiento por región

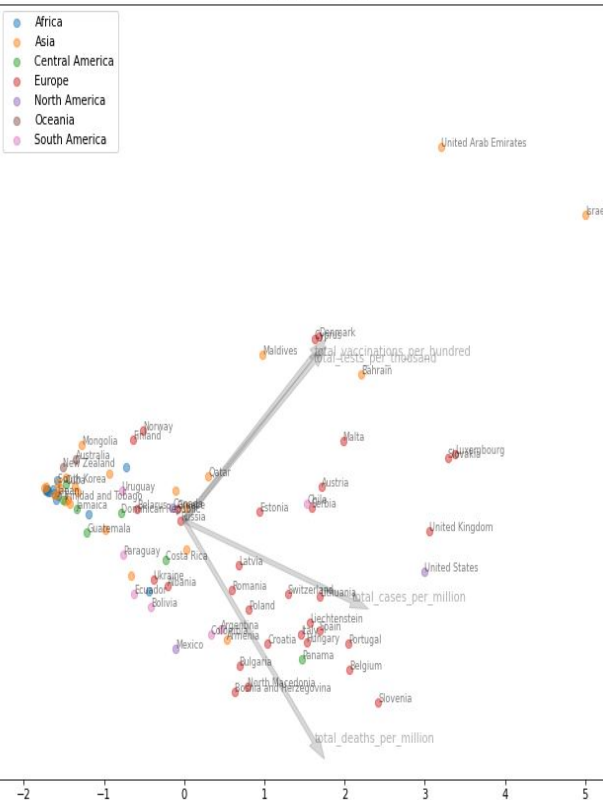
Reducción de dimensionalidad
con PCA

- Trabajamos sobre el dataset de Covid
- Agrupamos por fecha
- Trabajamos sobre las 4 variables mencionadas
- Observamos el comportamiento de los datos en 2 dimensiones

2021-01-01 (83.99%)



2021-03-01 (84.04%)



2021-06-01 (83.80%)

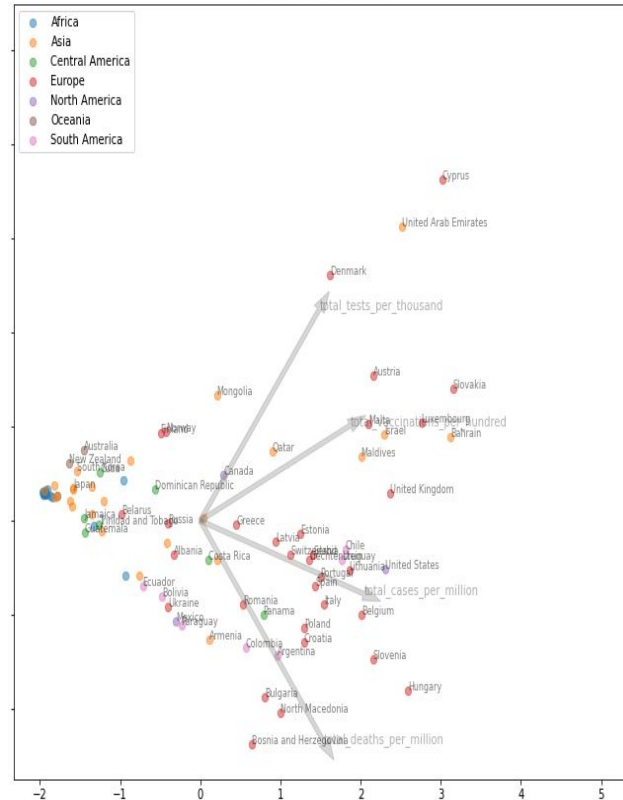
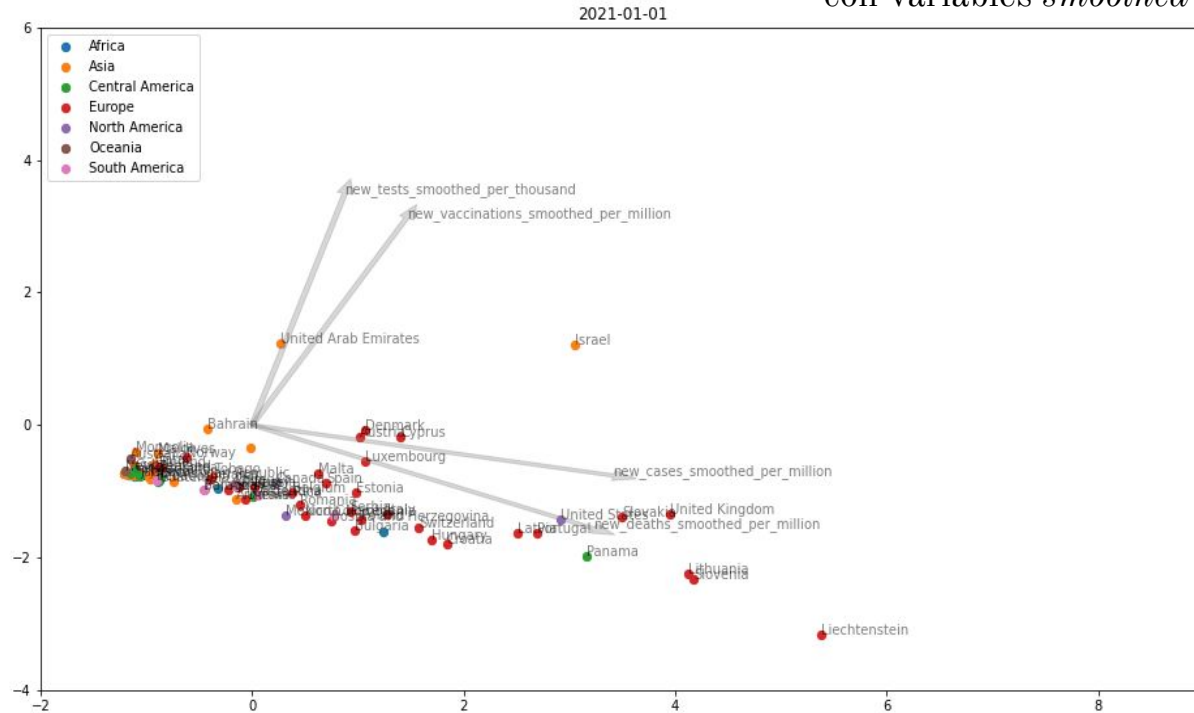


Gráfico que muestra el comportamiento de cada continente durante el avance del Covid.
Tomamos los datos acumulados en 3 momentos del año diferentes (enero, marzo y junio)

Gráfico de representación de avance del Covid-19 por quincena con variables *smoothed*



Observaciones

A lo largo del tiempo:

- se da una dispersión de los datos.
- cada región se mantiene agrupada.
- África y varios países de Asia no muestran grandes cambios.
- Europa manifiesta un cambio de tendencia de daño a prevención.

Surge la pregunta de qué factores intervinieron en el comportamiento de cada región.

Pregunta: ¿podemos explicar la situación de los países respecto al Covid-19 a partir de otras features relevantes?

—

Planteo como problema de clasificación

Clustering de países según las variables de Covid, y clasificación utilizando nuevas features

- Etiquetamos cada país en base a las 4 variables de Covid.
- Utilizamos KMeans con el método del codo, obtuvimos 5 clusters que conforman las etiquetas.
- Utilizamos como predictores:
 - indicadores de factores de riesgo en la población (diabetes, cardiovasc.)
 - indicadores de desarrollo (PBI, IDH)
 - datos de la población (distribución etaria, densidad poblacional)
 - inversión en salud (camas hosp.)
- Utilizamos Regresión Logística.
- Precision/Recall cercano a 0.5, faltaría información o explorar más los métodos.

Palabras finales

Comentarios y conclusiones

- Las variables de los datasets están incompletas y un futuro trabajo requiere reconstruir las variables faltantes.
- Los análisis temporales con PCA muestran gran dispersión entre los países y se observan patrones interesantes:
 - Agrupamiento por regiones
 - Cambio de tendencia a partir del inicio de la vacunación

Líneas posibles de trabajo futuro

- La prueba del clasificador fue acotada para la complejidad del problema. Se puede explorar la incorporación de nuevas variables relevantes y métodos de clasificación.
- En vez de observar el impacto de los países colonizadores, se puede analizar la influencia de grandes potencias.
- Relacionar avance del Covid con cantidad de celulares en el país cada 1000 habitantes. Ej. ¿Menos celulares implica más contagios por reuniones sociales?

Data set: países del mundo

- Datos interesantes como:
 - educación de la población.
 - nivel de trabajo en el sector de servicios (que tienen más interacción social).
- Inconvenientes:
 - datos de distintas fechas para los distintos países según cuándo se hubiera hecho el último censo en cada país.
 - datos viejos respecto a aquellos del dataset de Covid.



Fuente:

<https://www.kaggle.com/fernandol/countries-of-the-world/>