

EL DATA ANALYTICS COMO HERRAMIENTA EN LA GESTIÓN DE CRISIS

Sergio Gil
TECNALIA

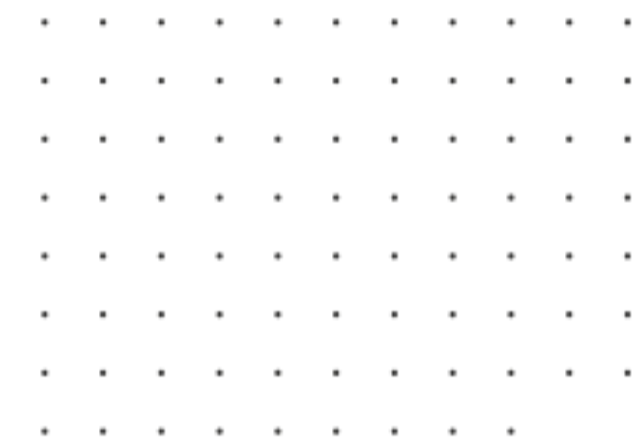
ÍNDICE

—	Contexto	04
—	Introducción	05
—	El papel del Data Analytics	06
—	Proyectos en los que ha participado TECNALIA	06
—	Consideraciones finales	07





CONTEXTO



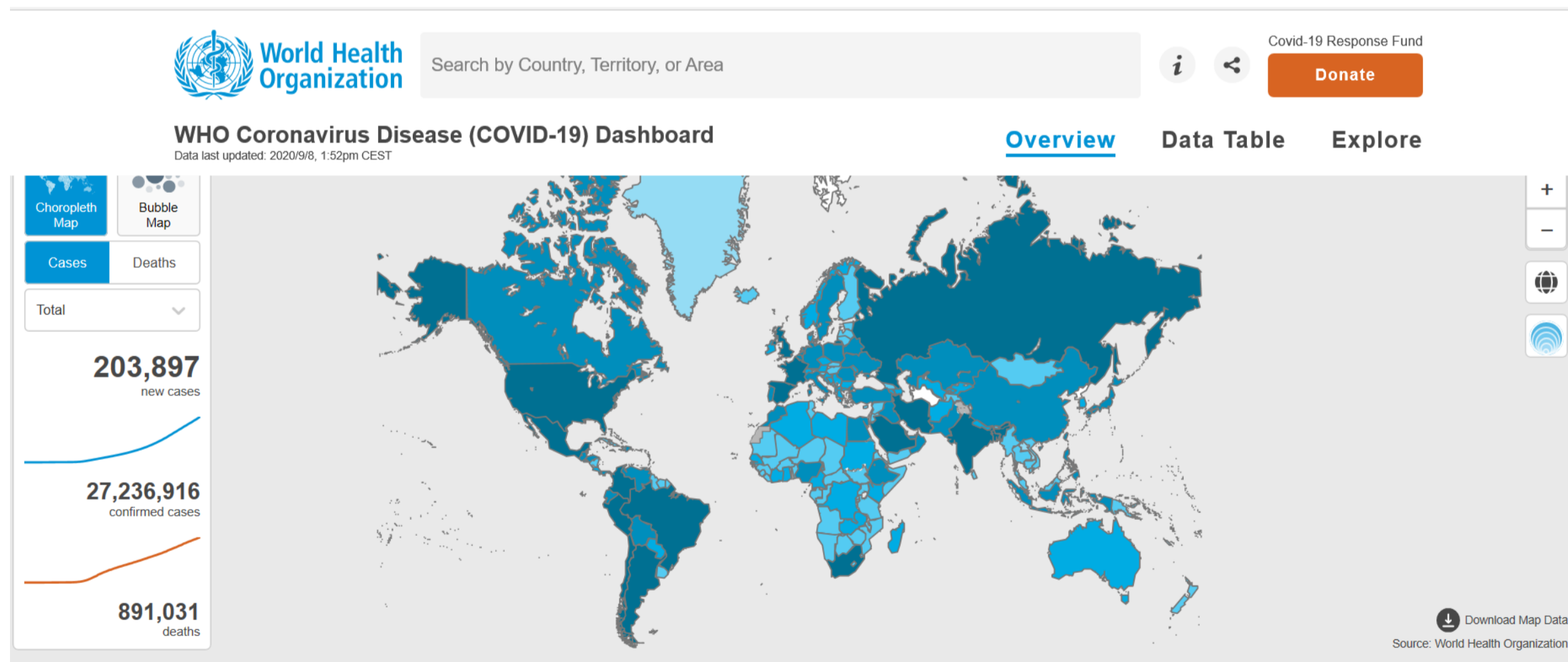
Origen: Wuhan (China)

Nombre: SARS-Cov-2, Severe Acute Respiratory Syndrome Coronavirus 2 (Covid19)

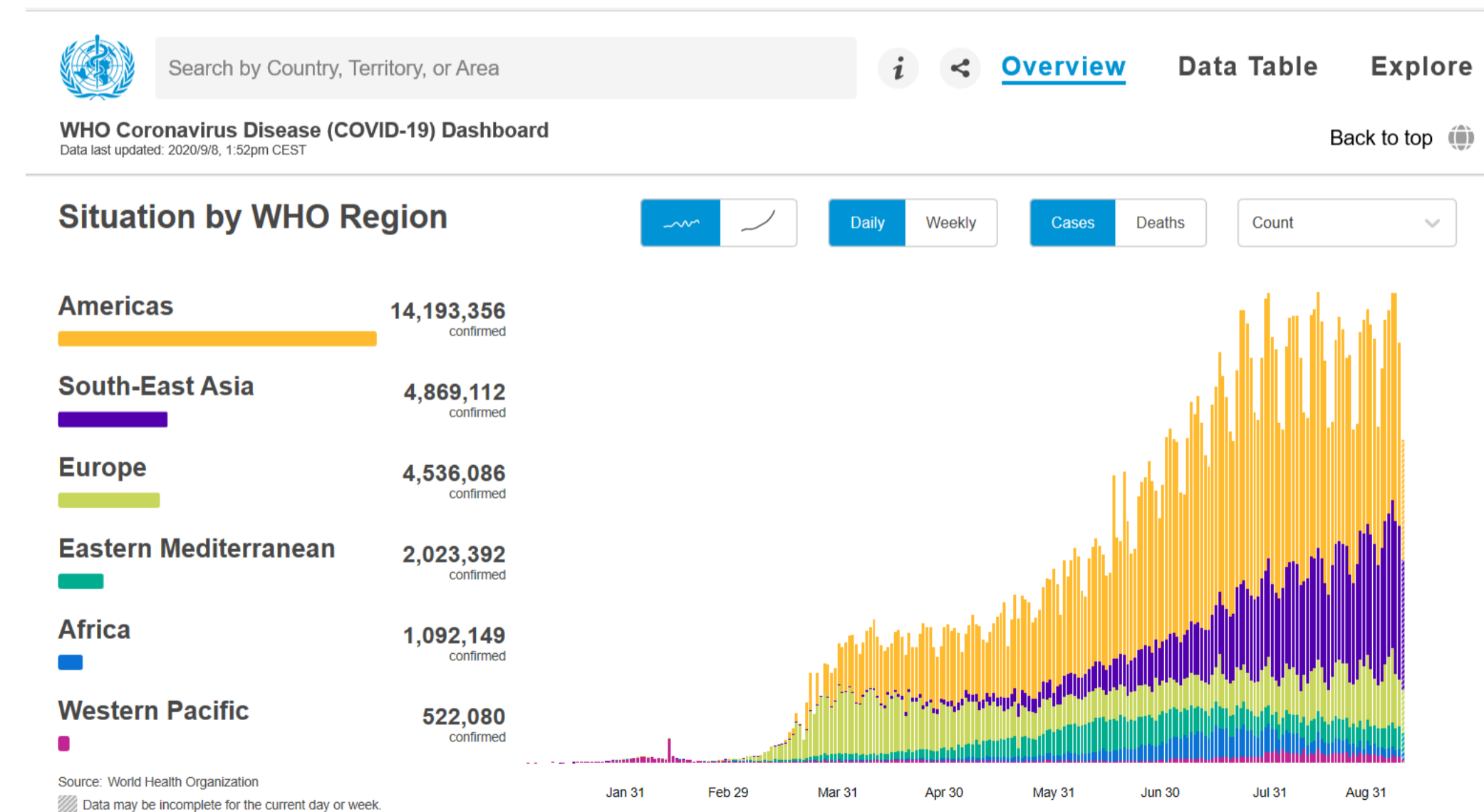
Fechas importantes:

- 30 de Enero de 2020, la OMS declara emergencia sanitaria de preocupación internacional.
- 11 de Marzo de 2020 la OMS lo reconoce como pandemia. Presente en 114 países y más de 4000 muertos.
- 4 de Abril se llega al millón de afectados a nivel mundial.
- 8 de Septiembre de 2020 más de 27 millones de infectados a nivel mundial.

Situación actual:

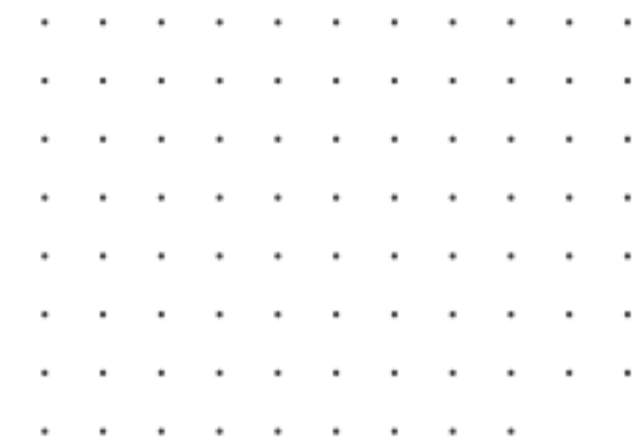


Globally, as of 1:52pm CEST, 8 September 2020, there have been 27.236.916 confirmed cases of COVID-19, including 891.031 deaths, reported to WHO.





INTRODUCCIÓN



Según la OMS la gestión de la pandemia se divide en 3 grandes fases:

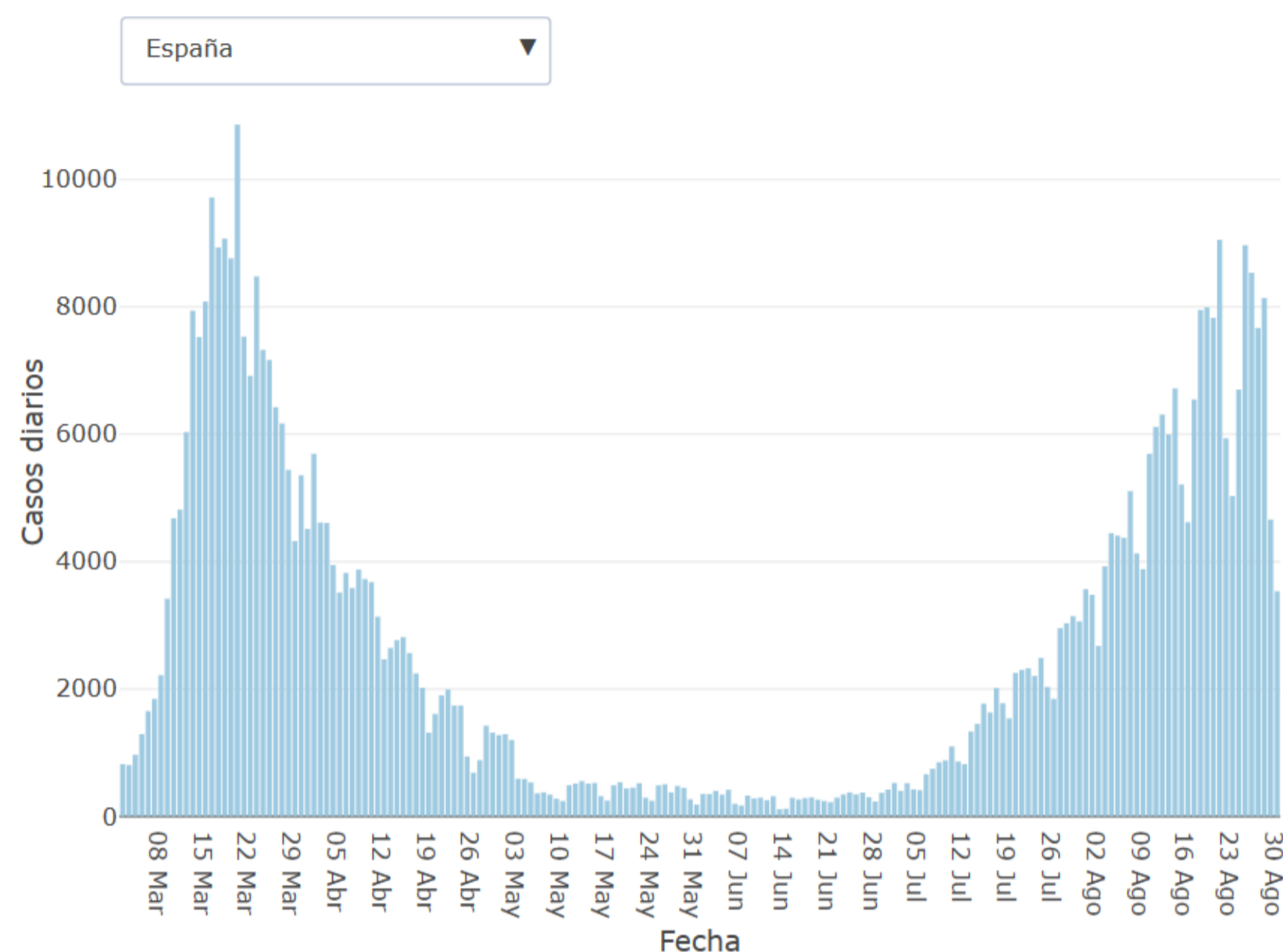
1. **Propagación.** Control del origen, bloqueo de la transmisión, prevenir otros brotes.
2. **Afección (intensidad).** Objetivo control de la intensidad de la pandemia (pico) y aliviar sistemas sanitarios.
3. **Monitorización y control.** Objetivo reducir los casos, controlar completamente la epidemia y lograr un equilibrio entre la prevención, el control de la pandemia y el desarrollo económico y social. Implementación de políticas basadas en evidencias científicas.

(source: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.PDF>)

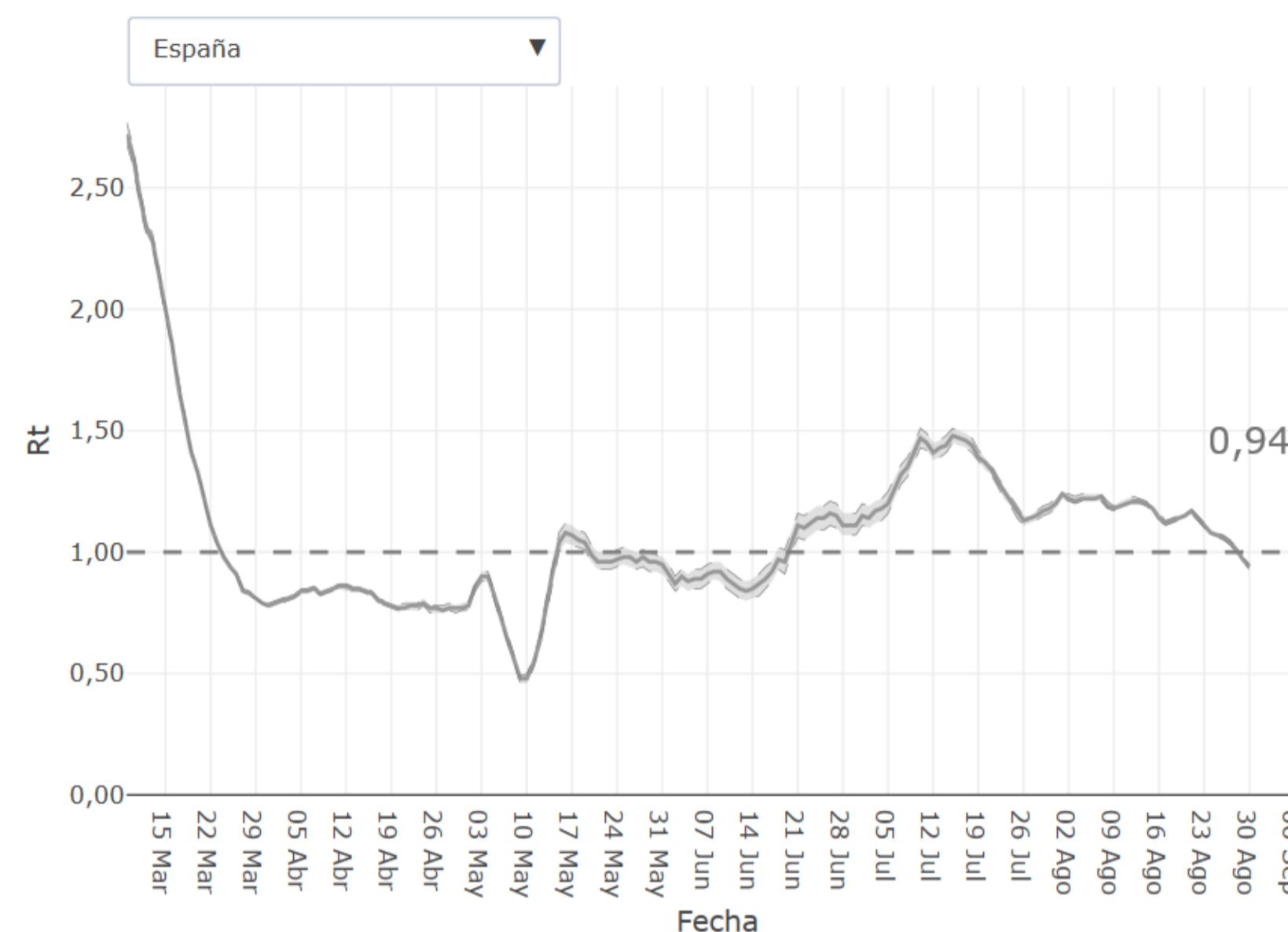
Algunas fechas importantes en España:

- Primer caso detectado 31 de Enero de 2019.
- Se declara el estado de alarma 15 de Marzo de 2020. (1512 casos).
- Entrada en **fase 0** (en la mayoría de las regiones) el 4 de Mayo de 2020. (218011 casos).
- Entrada en **fase 1** (en la mayoría de las regiones) el 11 de Mayo de 2020. (227436 casos).
- Entrada en **fase 2** (en la mayoría de las regiones) el 25 de Mayo de 2020. (235400 casos).
- Entrada en **fase 3** (en la mayoría de las regiones) el 8 de Junio de 2020. (241717 casos).
- Entrada en la **“nueva normalidad”** (en la mayoría de las regiones) el 22 de Junio de 2020. (246504 casos).
- ...A fecha de 10/09/2020 se ha acumulado un total de 543000 casos de personas infectadas.

Curva epidémica



Número reproductivo básico instantáneo (Rt)



Fuente: <https://cnecovid.isciii.es/covid19/#ccaa>
<https://cnecovid.isciii.es/covid19/#documentación-y-datos>

La tasa de crecimiento y decrecimiento con el tiempo de duplicación de los casos y el tiempo en reducirse a la mitad. Se utilizan ventanas de tiempo de siete a nueve días. Para ello se realiza un análisis de **regresión lineal con el logaritmo del número de casos diario de COVID-19 como variable dependiente y el tiempo de declaración del caso como independiente. El paquete utilizado es: R: incidence** (Thibaut Jombart, Zhian N. Kamvar, Rich FitzJohn, Jun Cai, Sangeeta Bhatia, Jakob Schumacher and Juliet R.C. Pulliam (2020). incidence: Compute, Handle, Plot and Model Incidence of Dated Events. R package version 1.7.1. URL <https://doi.org/10.5281/zenodo.2584018>) Además, estimamos el **número de reproducción básico instantáneo (Rt. Estimate Time Varying Reproduction Numbers from Epidemic Curves. R package version 2.2-1. <https://CRAN.R-project.org/package=EpiEstim>) modelo de Poisson, usando un marco bayesiano para la construcción de los intervalos de confianza** Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. Am J Epidemiol 2013;178(9):1505-1512).



EL PAPEL DEL DATA ANALYTICS



Data Analytics

Predicción y alertas

BlueDot published the [first scientific paper on COVID-19](#), accurately predicting its global spread using proprietary models.

Inteligencia Artificial

Tracing y anticipación

Every year the US Centers for Disease Control and Prevention holds a competition to see who can accurately forecast the flu. a group at Carnegie Mellon University The lab gathers recent and historical data from the CDC and other partner organizations, including flu-related Google searches, Twitter activity, and web traffic on the CDC, medical sites, and Wikipedia. Those data streams are then fed into [machine-learning algorithms](#) to [make predictions in real time](#). Oxford University provides forecasts of the spread of the virus based on the [GLEAMviz](#) epidemiological model. [Metabiota](#), a San Francisco-based company, offers an [Epidemic Tracker](#) and a near-term forecasting model of disease spread, which they use to make [predictions](#)

Machine Learning

Visual Analytics

rank the top dashboards to be those of [UpCode](#), [NextStrain](#), the John's Hopkins' [JHU CSSE](#), [Thebaselab](#), the [BBC](#), the [New York Times](#), and [HealthMap](#). Other notable dashboards include [Microsoft Bing's AI tracker](#).

Diagnostico y pronóstico

to a [recent review](#) of AI applications against COVID-19 by researchers working with UN Global Pulse, studies have shown that AI can be as accurate as humans, can save radiologists' time, and perform a diagnosis faster and cheaper than with standard tests for COVID-19.

called [COVID-Net](#) has been developed to diagnose COVID-19 in chest x-rays using data from patients with various lung conditions, including COVID-19

[Dutch University of Delft](#) who released an AI model for diagnosing COVID-19 from X-rays

Researchers at China's Huazhong University of Science and Technology have used ML to develop a [prognostic prediction](#) algorithm to predict the likelihood of someone surviving the infection.

Tratamiento y cura

Google's [DeepMind](#) has predicted the structure of the proteins of the virus — information that could be useful in developing new drugs

Researchers from South Korean and the USA have [published](#) results from using ML to identify an existing drug

Chinese firm [Baidu](#) is one of the producers of such infrared cameras that uses computer vision to scan crowds.

Fuentes de Datos:

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov>

<https://datascience.nih.gov/covid-19-open-access-resources>

<https://www.tableau.com/covid-19-coronavirus-data-resources>

<https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data>

<https://data.europa.eu/euodp/en/data/dataset/covid-19-clinical-trials-in-the-european-economic-area-eea>

<https://data.europa.eu/euodp/en/data/dataset/bd2f71e7-0551-4f57-8e82-fcfca8c1a462>

<https://azure.microsoft.com/es-es/services/open-datasets/catalog/covid-19-open-research/>

<https://www.kaggle.com/search?q=covid+19>

<https://dj2taa9i652rf.cloudfront.net/> (AWS data lake)



PROYECTOS EN LOS QUE HA PARTICIPADO TECNALIA



EPIDIG

- App, vía bluetooth, que rastrea las exposiciones de las personas trabajadoras en los centros de trabajo, de forma totalmente anónima y confidencial.
- Permite detectar, de forma temprana, los trabadores y trabajadoras que han estado en contacto con las personas que han reportado una infección.

RAPID (Micro + Macro)

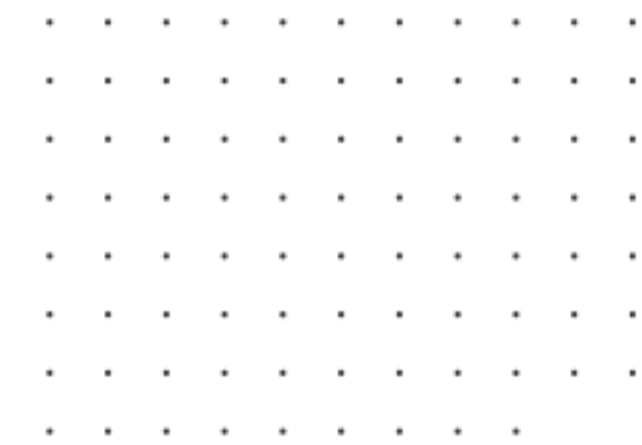
- App, vía códigos QR, que ayuda a la toma de decisiones en la gestión del COVID19, a partir de datos sobre los movimientos de las personas trabajadoras en el entorno micro (centro de trabajo) o macro (zonas de trabajo).
- También permite incorporar la funcionalidad adicional de rastreo de contactos por bluetooth **con identificación.**

QR Blockchain

- Proporciona certificados digitales sobre los resultados de las pruebas analíticas con tecnología blockchain.



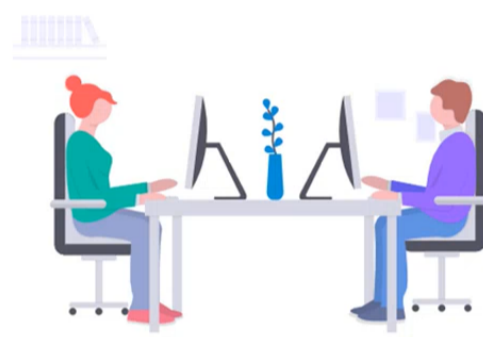
CERTIDIG: QR BLOCKCHAIN



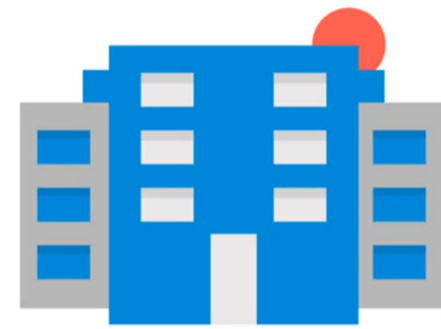
Certificado QR Blockchain test (CERTIDIG)

CERTIDIG

Bienvenido, selecciona tu perfil



Personal riesgos laborales

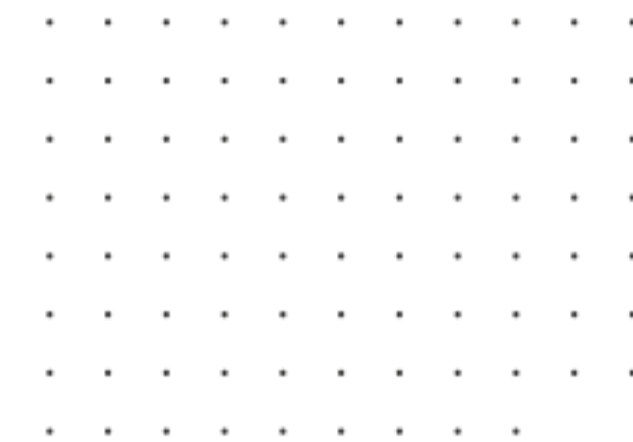


Acceso resultado test

- Vinculado a la realización de test, emisión de certificados digitales con tecnología, permite visualizar los resultados de las pruebas del Covid19 (PCR, Elisa o Test Rápidos), además de garantizar la fiabilidad y trazabilidad de la información, mediante tecnología BlockChain.
- Blockchain permite el intercambio de datos a través de operaciones codificadas y cifradas, garantizando la seguridad de los datos, inmutabilidad y fiabilidad



— EPIDIG (TRACING)



EPIDIG App tracing



Modo Usuario



Modo Gestor (Dpto. Riesgos laborales)

Panel de control 1

- Generador de los COVIDCodes que permiten activar los contactos tras un positivo.

Panel de control 2

- Dashboard con las estadísticas de códigos generados, usados, evolución de usuarios/as y aplicaciones conectadas, etc.

Implementa el protocolo DP3-T sin necesidad de las APIs GAEN (Google Apple Exposure Notification). **En producción desde finales de mayo.**



—
RAPID MACRO



RAPID Macro (BRTA)

Consiste en:

- Web scraping.
- Fuentes de datos heterogéneas.
- Integración flujos de movilidad.
- Integración datos epidemiológicos.
- Integración datos demográficos.
- Integración datos económicos. Para la interpretación de los datos económicos el BRTA ha dispuesto de la colaboración del Instituto Vasco de la Competitividad ORKESTRA
- Fusión de datos.
- Integración de datos hasta la unidad mínima de gestión (espacial) de zona sanitaria.
- Técnicas de Visual y Data Analytics.

Proporciona:

- Dashboard con la evolución de la pandemia en las diferentes zonas de salud de Euskadi.
- Dashboard geográfico con la información fusionada e integrada por zona de salud.
- Dashboard pandémico-económico por zona de salud.

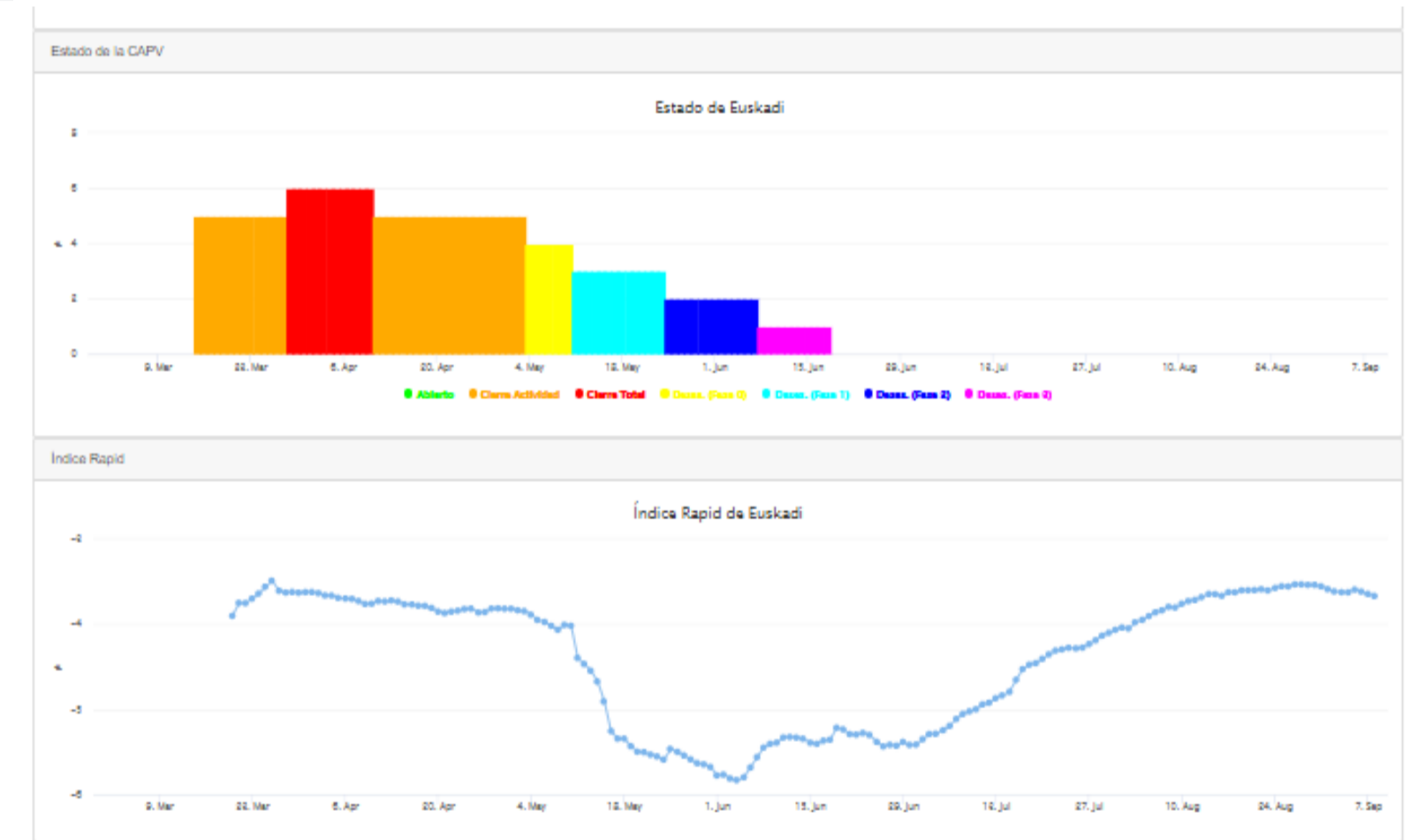
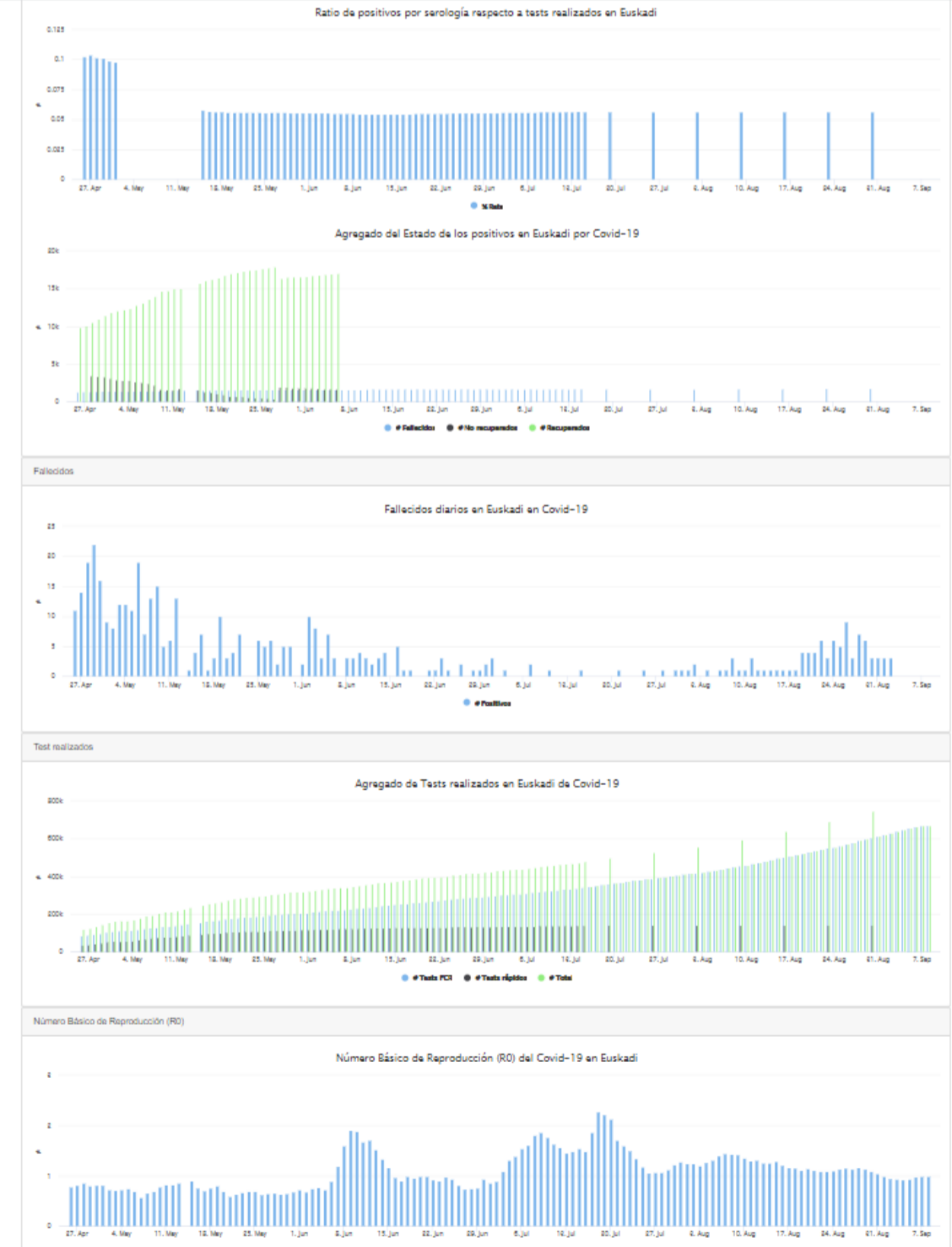
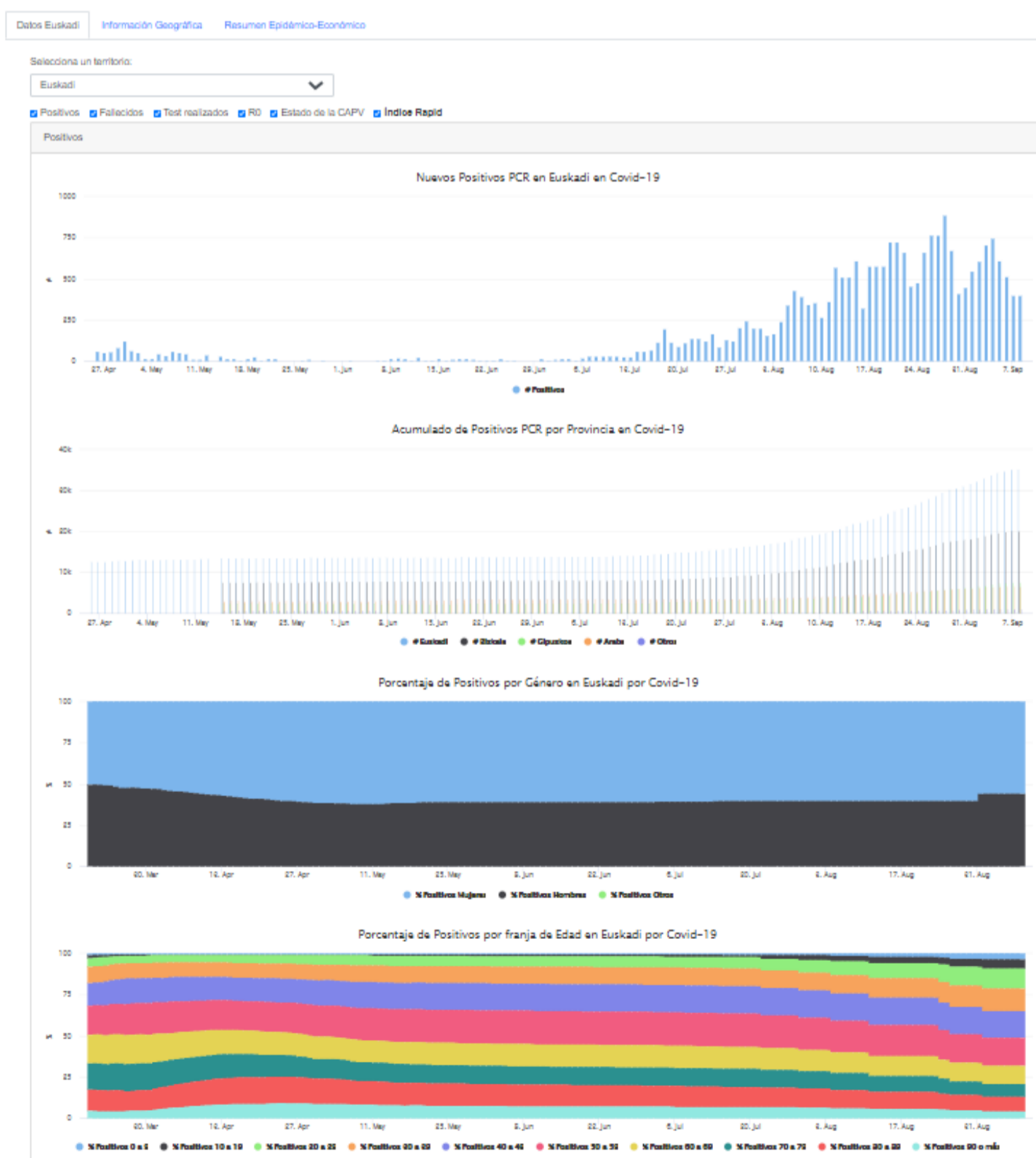
Proyectos en los que ha participado TECNALIA

Herramienta para la ayuda a la toma de decisiones en la reactivación económica en tiempos de pandemia, con una **esfera epidemiológica**, otra **social** y otra **económica**.

Dashboard con la evolución de la pandemia en las diferentes zonas de salud de Euskadi.

Información que se puede visualizar:

- Positivos: nuevos positivos detectados mediante pruebas PCR por día.
- Fallecidos diarios.
- Porcentaje de Positivos por Género: Mujer (azul) /Hombre (negro)
- Porcentaje de Positivos por franja de edad. Se consideran las franjas de edad de 10 en 10 años.
- Ratio de positivos por test rápidos, o serológicos, respecto al número de test realizados.
- Estado de los positivos.
- Número total de test realizados (tanto PCR, como serológicos).
- R0: número básico de reproducción. el número promedio de casos nuevos que genera un caso dado a lo largo de un período infeccioso.

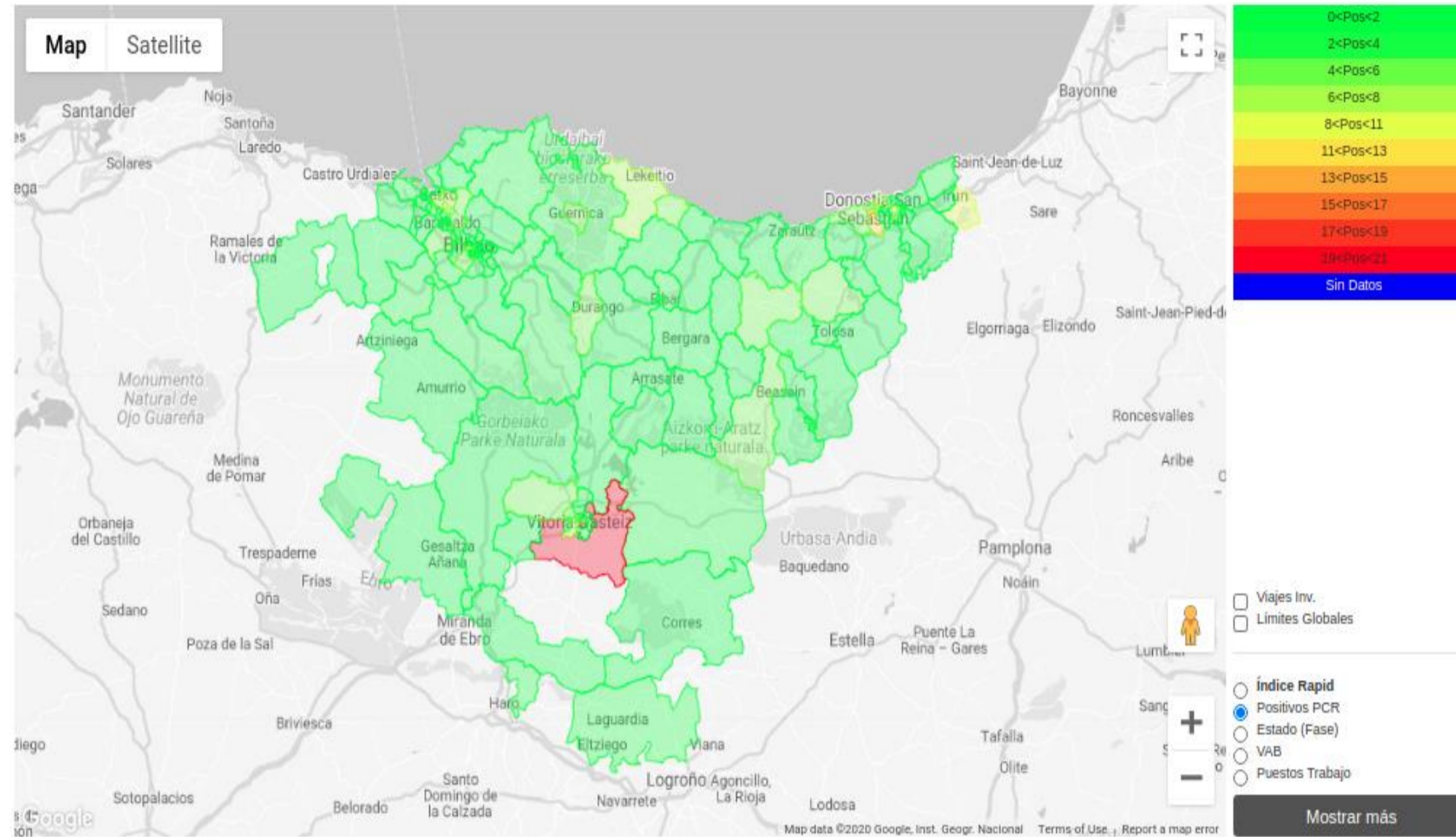


Proyectos en los que ha participado TECNALIA

Dashboard geográfico con la información fusionada e integrada por zona de salud.

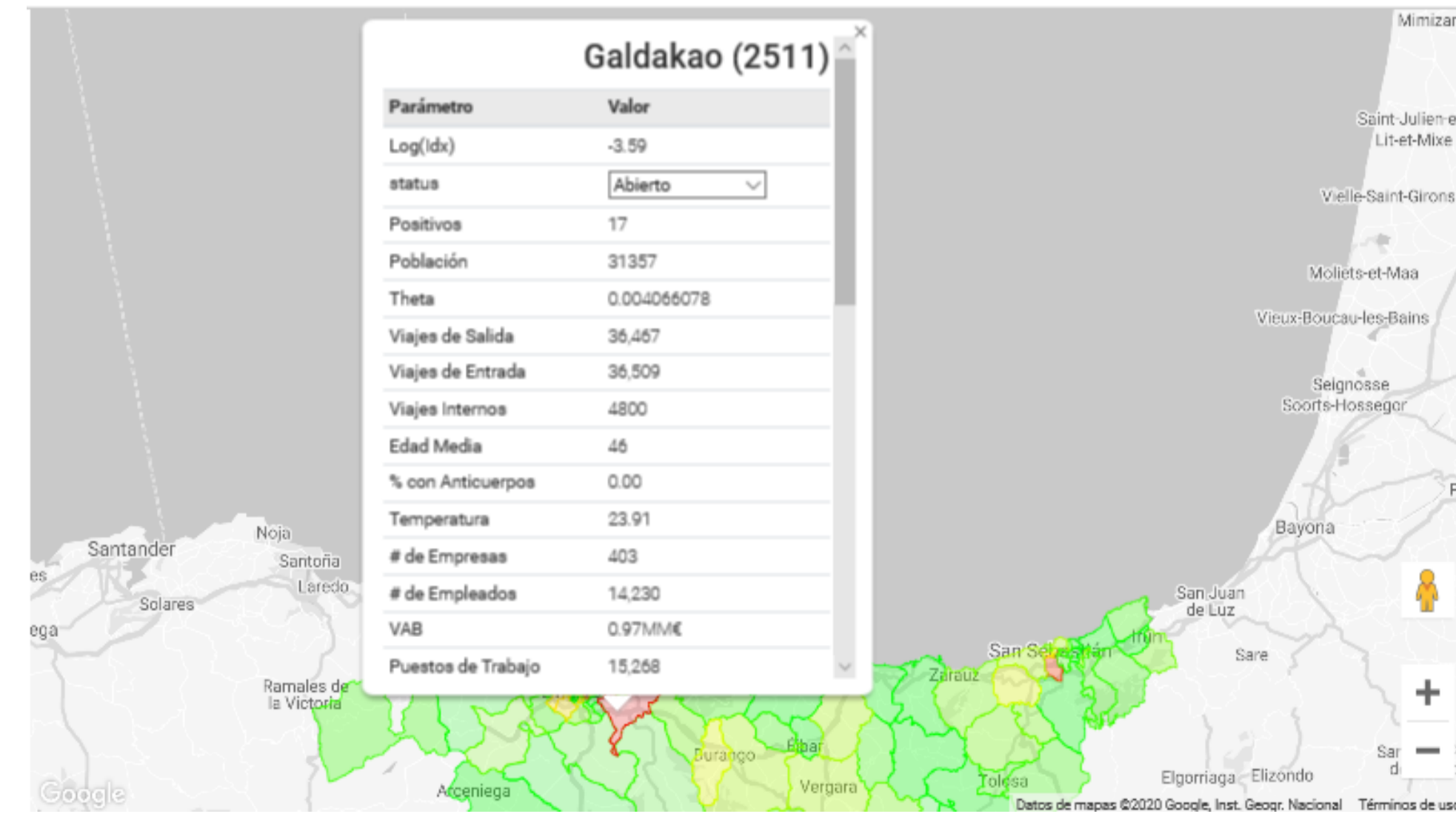
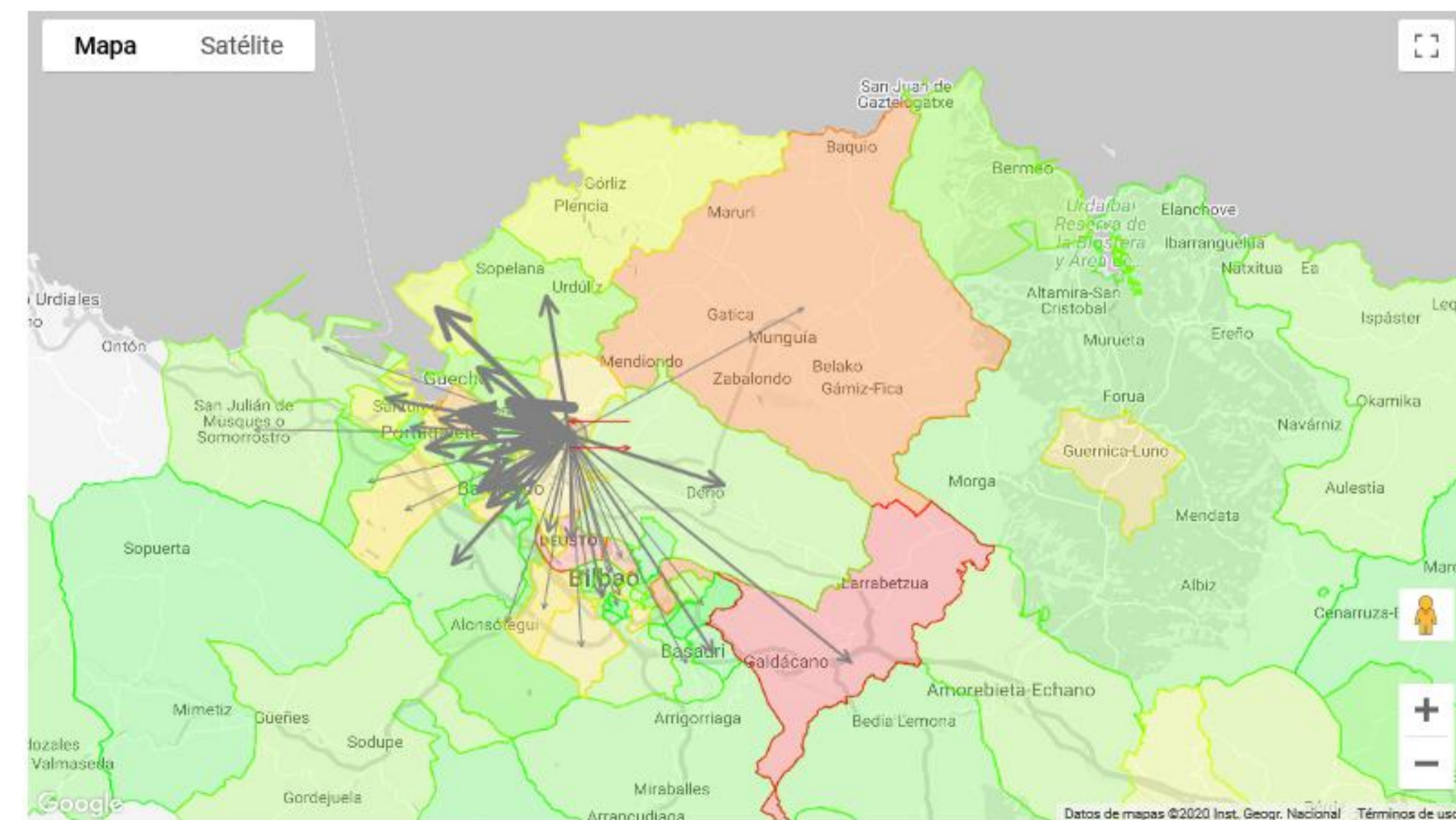
Fecha: 2020-9-3

Nombre	Posit.	Estado	Log(Idx)	Pob.	VAB (MM€)	Puestos Trabajo
Llanada Alavesa	3	Abierto	-2.97	12,975	0.36	5,207
Aiala	1	Abierto	-3.00	19,290	0.51	8,414
Norte-Legutio	0	Abierto	-3.41	4,566	0.52	8,168
Arrasate	2	Abierto	-3.34	23,445	0.81	14,205
Valles Alaveses II	0	Abierto	-3.49	2,993	0.17	2,479
Rioja Alavesa	3	Abierto	-3.00	11,509	0.68	9,630
Montaña Alavesa	0	Abierto	-3.47	2,654	0.09	1,759
Olaguibel	21	Abierto	-2.69	26,312	2.29	37,246
Norte-Zuia	2	Abierto	-2.98	6,598	0.15	2,739
Llodio	4	Abierto	-2.98	22,356	0.77	9,606
Valles Alaveses I	1	Abierto	-3.10	5,610	0.23	3,890
Abetxuko	1	Abierto	-2.76	3,557	0.03	563
Aranbizkarra I	6	Abierto	-2.02	13,641	0.22	3,628
Aranbizkarra II	2	Abierto	-2.20	13,850	0.27	4,339
Casco Viejo (Vitoria-Gasteiz)	6	Abierto	-1.95	17,176	0.51	8,252
Gazalvide-Txagorritxu	2	Abierto	-1.95	12,823	0.28	4,496
La Habana	5	Abierto	-1.91	10,002	0.13	2,164
Lakua-Arriaga	1	Abierto	-2.22	13,568	0.56	9,163

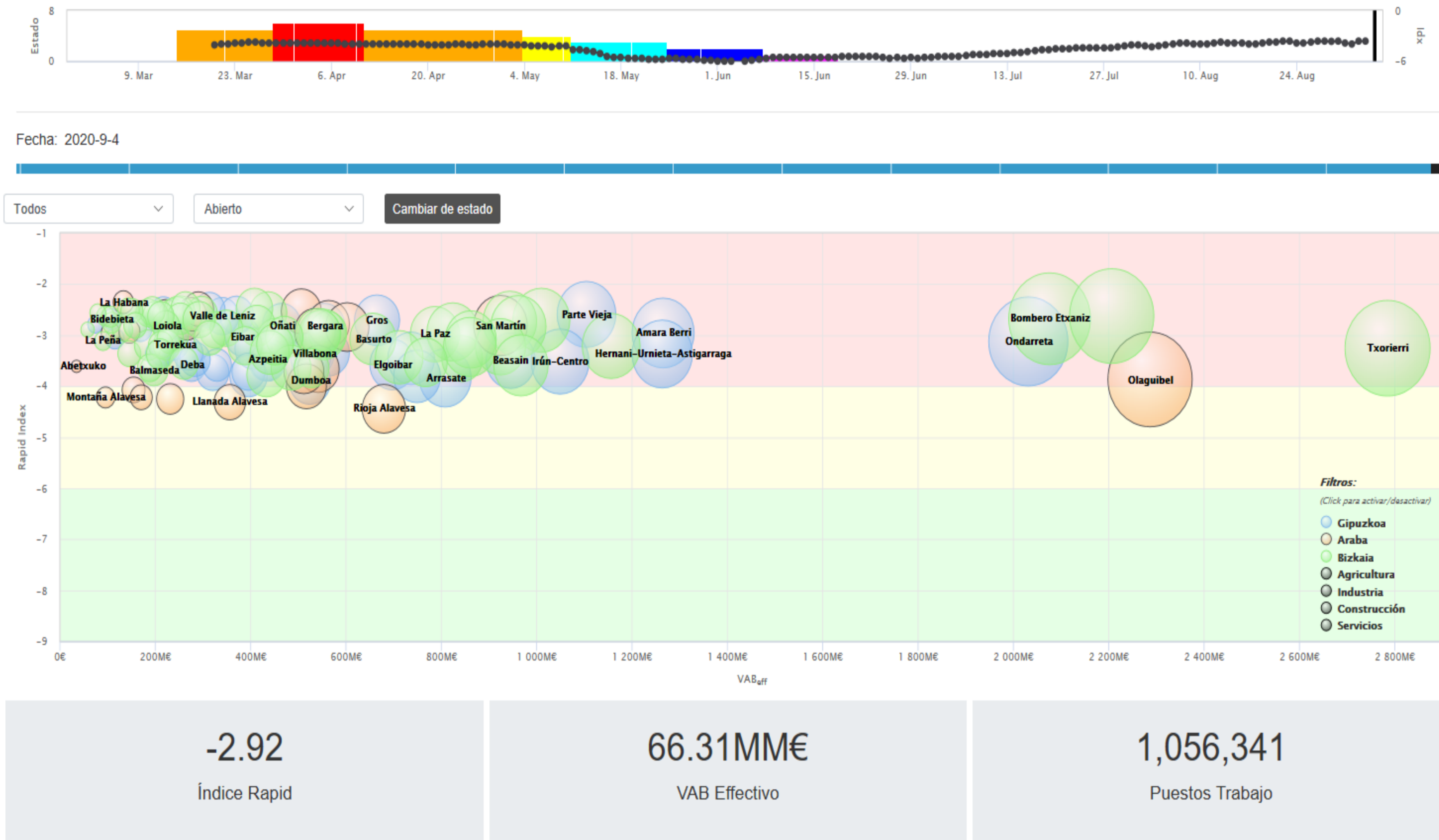


Información que se puede visualizar:

- Índice Rapid
- Número de Positivos PCR detectados (seleccionado por defecto)
- Estado de cada zona
- VAB, valor añadido bruto
- Puestos de trabajo
- Porcentaje de la población con Anticuerpos
- Población
- Edad Media
- Número de empresas Totales
- Establecimientos
- Empleados
- % Puestos de trabajo Esenciales



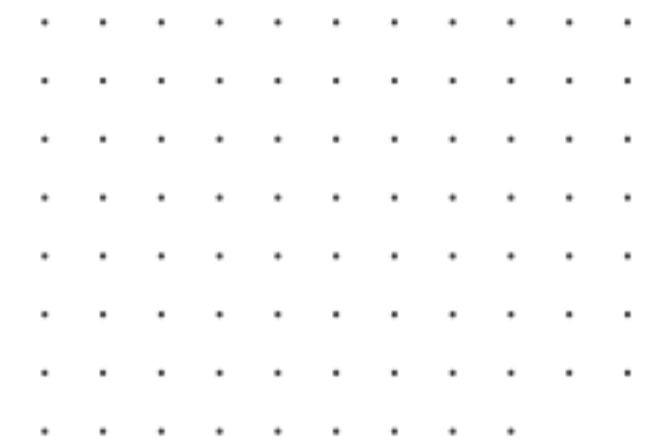
Dashboard pandémico-económico por zona de salud.



- Visualización en Pareto: Valor Añadido bruto vs Índice RAPID.
- El Índice RAPID es un índice que fusiona información pandemeológica, movilidad y demográfica para, por zona de salud, proporcionar un valor de riesgo de propagación de la pandemia.
- Permite una gestión específica por zona de salud.



CONSIDERACIONES FINALES



Algunas consideraciones finales:

- El **Data Analytics** en su concepto más amplio se ha demostrado que puede **ayudar en la gestión de la toma de decisiones** ante **situaciones de crisis**.
- El primer gran **reto** es el **acceso a los datos, tanto a las fuentes como a los expertos de dominio (responsables de los datos)**.
- El segundo gran reto es la **conformación de data sets homogéneos y normalizados**. Será fundamental temas de **gobernanza del dato**.
- El tercer gran reto es la conformación de **equipos multidisciplinarios** que permitan conformar soluciones basadas en Data Analytics conjugando conocimiento de dominio.
- El cuarto gran reto es dar **explicabilidad e interpretabilidad** a los resultados.
- El quinto gran reto es el debate sobre cuestiones **legales y éticas** en el manejo de los datos.
- El sexto gran reto concebir el **Data Analytics** como lo que es, una **herramienta y no una solución**. El Data Analytics nos permitirá digerir datos, extraer conocimiento, mejorar en la gestión, aprender para el futuro... pero no solucionará la pandemia.

GRACIAS

sergio.gil@tecnalia.com

blogs.tecnalia.com



www.tecnalia.com

tecnalia  Inspiring
Business