

PAC3 – Creació de visualització interactiva –

NOTA: El present document i tots els altres arxius que conformen el treball que es presenta, es poden trobar a https://github.com/marcosfvlaboa/Catalunya_COVID-19_DataViz

- 1- **Seleccionar algun dels conjunts de dades següents o algun altre de complexitat i característiques similars.**
 - <https://www.ecdc.europa.eu/en/covid-19/data>
 - <https://analisi.transparenciacatalunya.cat/es/browse?q=covid&sortBy=relevance>
 - <https://analisi.transparenciacatalunya.cat/Salut/Incidencia-de-la-COVID-19-a-Catalunya/623z-r97q/data>
 - https://datos.gob.es/en/catalogo?q=covid&theme_id=salud&sort=metadata_created+desc
 - <https://github.com/montera34/escovid19data>

S'ha optat per utilitzar el conjunt de dades de la incidència de la COVID-19 a Catalunya¹. Sobre aquest conjunt de dades cal remarcar els següents punts:

- Dades diàries de la incidència de la COVID-19 a Catalunya elaborades per l'IDESCAT a partir de dades del Departament de Salut i del Ministeri de Sanitat. La sèrie s'inicia el 24 de febrer de 2020 i s'actualitza cada dia fins al dia 22 de setembre de 2020.
- Es proporcionen les xifres diàries de nous casos confirmats, defuncions i altes hospitalàries, així com els totals acumulats.
- Les dades poden presentar oscil·lacions o incoherències derivades dels processos de depuració.
- Les dades del període del 15 a 23 d'abril de 2020 (que provenien del Departament de Salut) s'han revisat, a conseqüència de la substitució de la informació del Departament de Salut per la del Ministeri de Sanitat.
- El 16 d'abril disminueix la xifra total de defuncions i altes hospitalàries, per una revisió de dades del Ministeri de Sanitat.

¹ IDESCAT - a partir de les dades del Ministeri de Sanitat i del Departament de Salut de la Generalitat de Catalunya (2 de març del 2021). Incidència de la Covid-19 a Catalunya. <https://analisi.transparenciacatalunya.cat/Salut/Incidencia-de-la-COVID-19-a-Catalunya/623z-r97q/data>

- El Ministeri de Sanitat informa que en l'augment del total de casos confirmats els dies 10, 11, 15 i 16 de maig s'han considerat, a més dels casos nous, alguns casos antics.
- El 14 i el 18 de maig disminueix la xifra total de casos confirmats, per una revisió de dades del Ministeri de Sanitat.
- Les dades de defuncions fins al 23 de maig de 2020 fan referència a defuncions diàries i a partir del 24 de maig de 2020 fan referència a defuncions dels darrers 7 dies.²

Es crea un Jupyter Notebook anomenat "*transformation.ipynb*" per explorar, seleccionar i transformar les dades. El dataset resultant es desa a l'arxiu "*COVID-19_Catalunya_setmanal.csv*".

En l'exploració bàsica del data set original es comprova que:

- consta de 212 registres dividits en 7 variables. La primera d'elles és la data diària en el que es van registrar. La resta de variables són les dades recollides sobre la COVID-19.
- existeixen valors nuls que modifiquem a valors zero ja que es computaran com que no hi ha hagut cap ocurrència de la variable en qüestió en un dia concret. Per exemple, el fet de hi hagi un valor nul d'una "Alta diària", s'entén com que aquell dia no va haver-hi altes relacionades amb el COVID-19. Aquest fet té la seva lògica en el transcurs de la pandèmia.
- el valor màxim, per exemple, de la variable "Total de casos confirmats" és de 133751. El valor mínim és 1. La mitjana "Nous casos diaris confirmats" és de 297,86.

L'objectiu de la transformació és la d'aconseguir un dataset amb les dades recollides per setmanes i no per dies. D'aquesta manera reduïm la complexitat de la interacció en la visualització. Per fer-ho caldrà:

- Ordenar per "Data" de la observació més antiga a la més nova.
- Sumar les dades diàries per setmanes (cada 7 dies)
- Descartar l'últim registre ja que es tracta de la suma de només dos dies. Notar que el dataset original comença amb la data dilluns 24/02/2020 però finalitza un dimarts 22/09/2020.
- Reanomenar el títol de les columnes.

² Detalls extrets de l'apartat "Sobre" en la propia página de les dades

2- Utilitzar alguna eina per crear una proposta de disseny inicial o Mockup que respongui una pregunta bàsica sobre el conjunt de dades.

A la figura 1 es presenta una proposta de disseny inicial de la futura visualització que es vol crear. La intenció és la de poder respondre a la pregunta bàsica “Cóm ha evolucionat el COVID-19 a Catalunya?”.

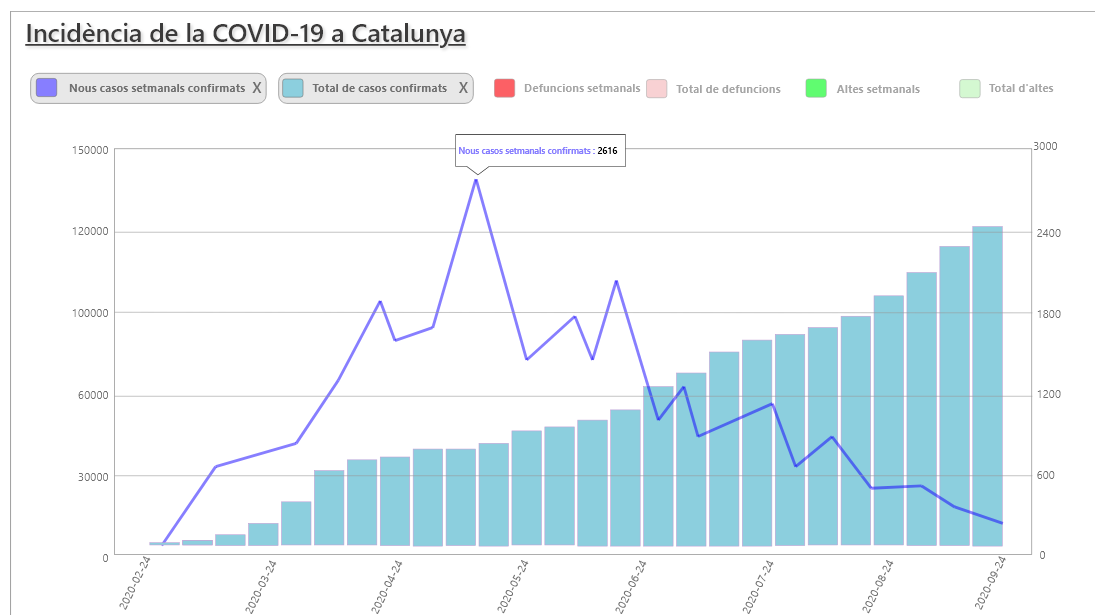


Figura 1. Mockup per la visualització de les dades del COVID-19 a Catalunya

L'esbós pretén mostrar l'evolució de la pandèmia segons les diferents variables de les que disposem. Per fer-ho:

- s'utilitza un gràfic en funció del temps per dates.
- les dades setmanals es presenten de manera lineal, mentre que les dades totals es presenten amb gràfics de barres.

Per la realització del Mockup s'opta per fer servir Adobe Xd. L'arxiu original es pot trobar al repositori (veure NOTA a l'inic d'aquest document).

3- Determinar quins elements interactius poden afegir-se a la proposta inicial per millorar les seves capacitats com a eina d'anàlisi visual.

Amb la figura 1 es poden intuir els elements interactius a afegir:

- Una de les propostes del Mockup es tracta de poder veure els valors en passar el ratolí per sobre d'un punt concret. A l'esbós es pot apreciar, amb una caixa amb fletxa, el valor en aquell punt per a la variable "Nous casos setmanals confirmats".
- Una altra proposta que es desprèn de l'esbós és la selecció de les dades. Sota el títol es pot veure una llegenda amb totes les variables de les dades. La intenció és que es pugui visualitzar totes les dades o bé una selecció de una o més variables. En la figura es veu com les dades seleccionades són "Nous casos setmanals confirmats" i "Total de casos confirmats". I són precisament aquestes dues les que es visualitzen.

4- Implementar usant qualsevol eina la versió interactiva que permeti la manipulació de diferents elements de la visualització interactiva.

S'ha optat per realitzar la visualització interactiva amb l'eina Flourish Studio. Una eina tipus aplicació web que permet la creació, exportació i publicació de visualitzacions de dades.

El resultat es pot apreciar a la figura 2 i la visualització es pot visitar al següent enllaç:

- <https://public.flourish.studio/visualisation/6245825/>

La captura mostra el gràfic per a totes les variables.

Incidència de la COVID-19 a Catalunya

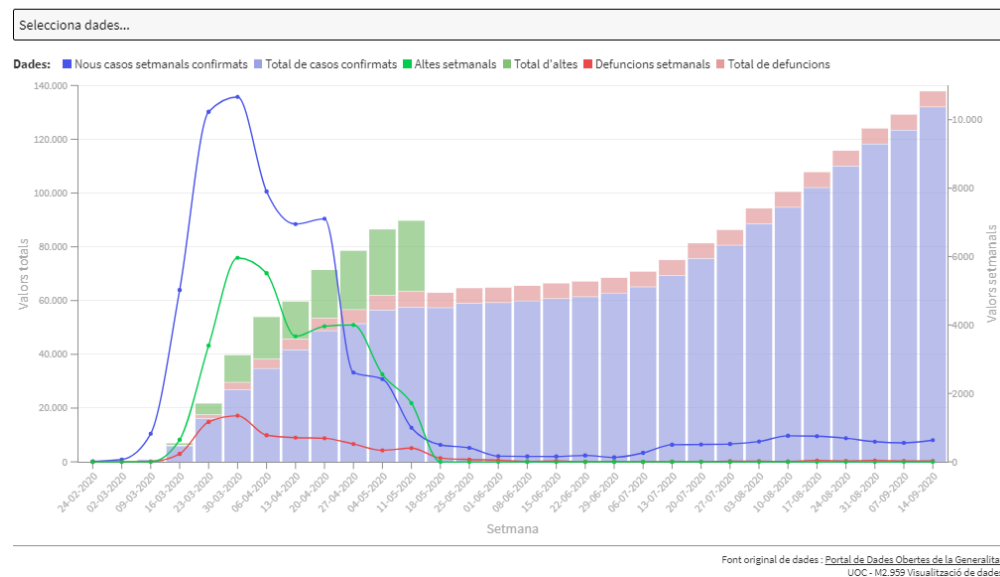


Figura 2. Visualització de les dades de la incidència del COVID-19 a Catalunya - General.

S'ha decidit fer servir colors semblants per a dades connexes. Per exemple, per a les defuncions setmanals i el total de defuncions s'han escollit dues tonalitats de vermell diferents. L'eix de les X representa la línia de temps. Als eixos de les Y s'indiquen els valors totals (esquerra) i setmanals (dreta). Com en la proposta inicial, les dades setmanals es representen amb línies i les dades totals amb barres. S'inclou una llegenda i una barra de selecció de dades.

Si es passa el ratolí per sobre del gràfic apareix el valor que té la variable en aquell punt. Si es fa clic es pot fixar. A la figura 3 es pot veure un exemple a on s'ha fixat l'element emergent al valor màxim de "Total d'altres". En aquest cas marca que es tracta de la setmana del 01/05/2020 amb un valor de 24454.

Incidència de la COVID-19 a Catalunya

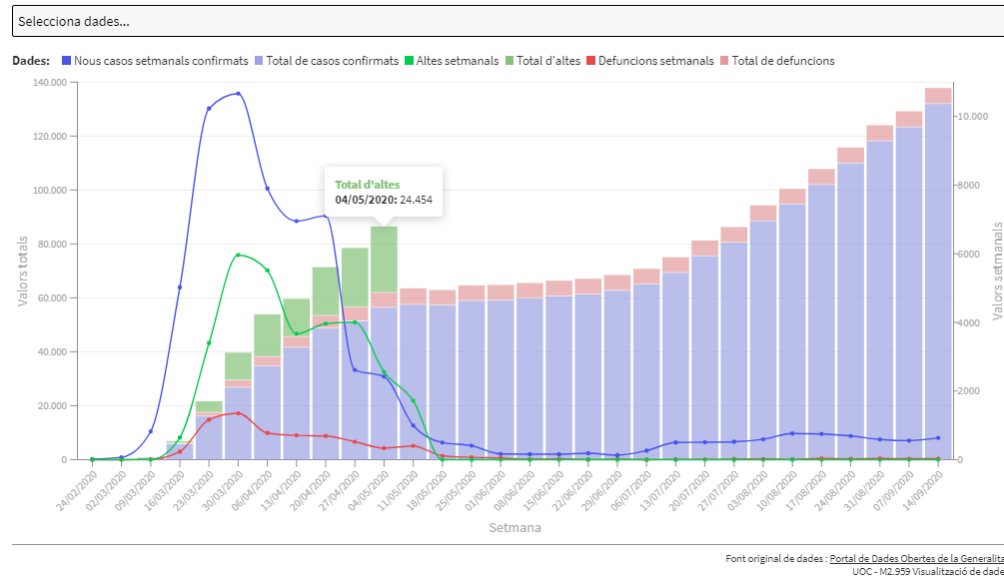


Figura 3. Detall de la interacció en passar el ratolí per un punt.

Si es seleccionen dades, utilitzant la barra de selecció es comprova com es filtren les dades per a la creació del nou gràfic. A la figura 4 es mostra un exemple a on s'ha inclòs les dades "Defuncions setmanals" i "Total defuncions". Com es pot veure, es presenten noves seccions en els eixos de les Y per adaptar les noves dades a l'espai, la llegenda es modifica per incloure només les seleccionades i la barra de selecció inclou les variables amb un botó per eliminar-les de la selecció. Però, es manté la coherència en la manera de presentar les dades gràficament, es a dir, les dades totals es mostren amb barres, les dades setmanals amb línies, es mantenen els colors i també la interacció anterior a on s'obre un element emergent en un punt seleccionat.

Incidència de la COVID-19 a Catalunya

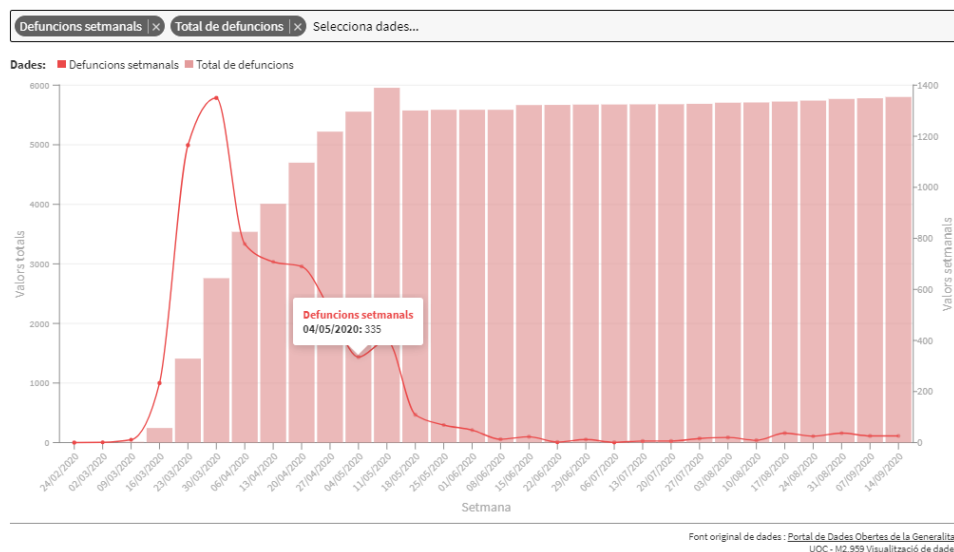


Figura 4. Detall de la interacció de la selecció de variables. En aquest cas corresponents a les defuncions.

5- Descriure preguntes amb llurs respostes que la visualització interactiva permet resoldre.

A banda de respondre a la pregunta general de “Cóm ha evolucionat el COVID-19 a Catalunya?”, es poden extreure preguntes més concretes.

- Quina ha estat la evolució de la pandèmia pel que fa a defuncions/casos/altes?

La figura 4 anterior mostra l'evolució pel que fa a defuncions. La figura 5 es l'exemple per a casos. Per a altes es faria de manera anàloga amb les dades corresponents.

Incidència de la COVID-19 a Catalunya

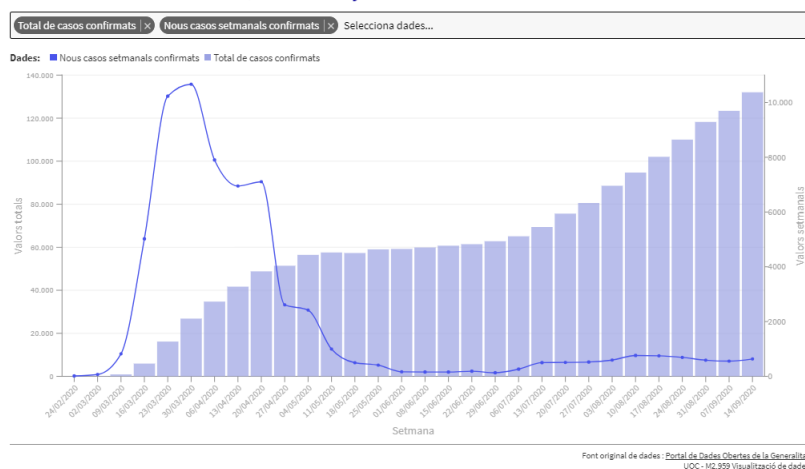


Figura 5. Detall de les variables corresponents als casos confirmats.

- Quina setmana hi va haver més defuncions?

La figura 6 mostra que va ser la setmana del 30/03/2020 amb un total de 1350 persones.

Incidència de la COVID-19 a Catalunya

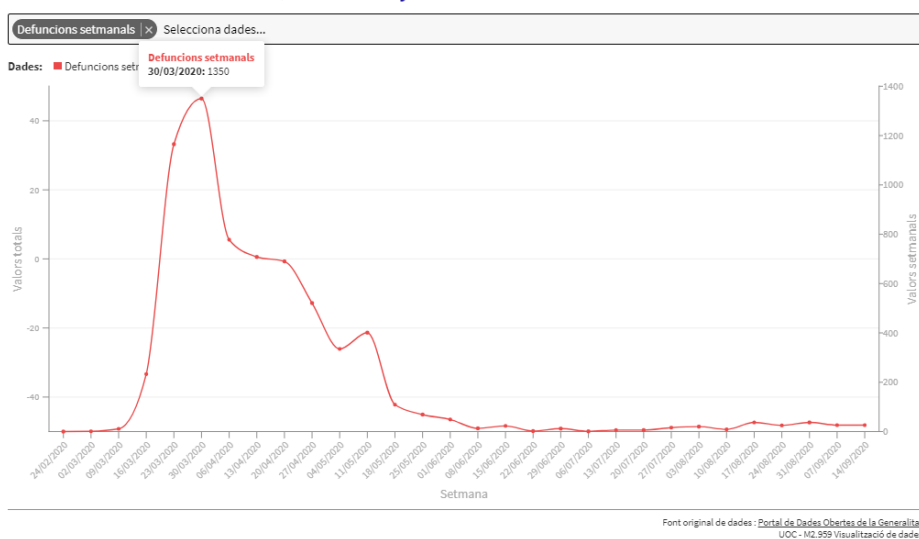


Figura 6. Detall de la variable “Defuncions setmanals”

- *Hi ha relació entre el nombre de casos i el de morts? I el d'altres?*

En la següent captura (figura 7) es pot apreciar certa relació entre aquestes tres variables. Per exemple, es pot extreure que tendeix a haver més altes i defuncions com més casos hi hagi.

Incidència de la COVID-19 a Catalunya

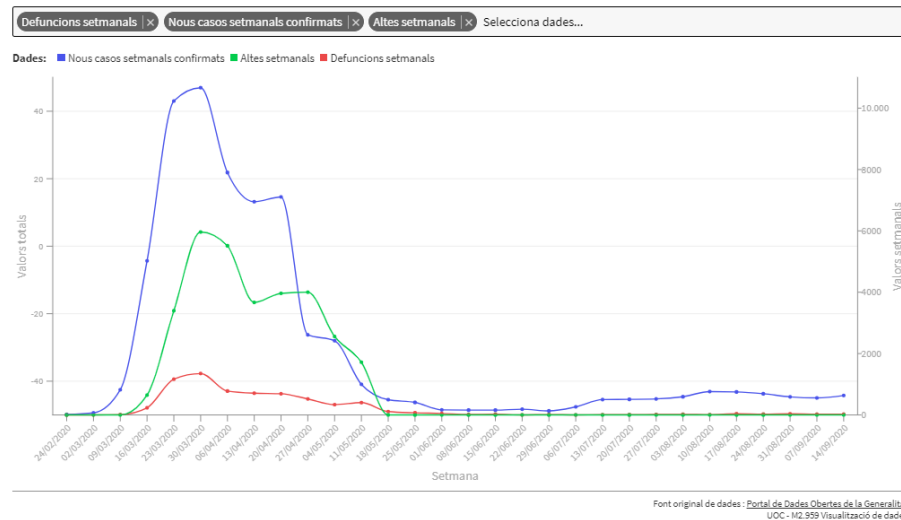


Figura 7. Detall de les variables setmanals.

També es poden descobrir curiositats com, per exemple, el que es mostra a la figura 8

Incidència de la COVID-19 a Catalunya

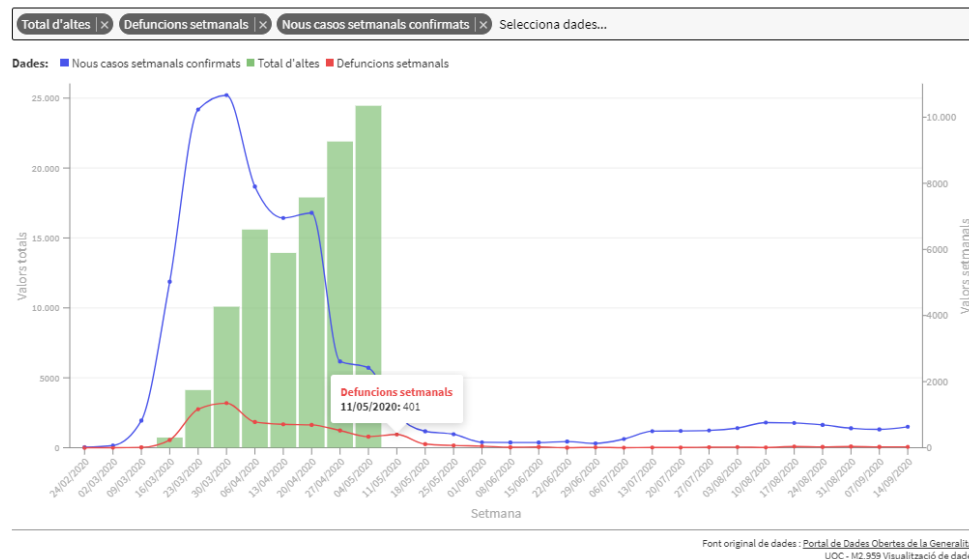


Figura 8. Detall de les variables "Total d'altres", "defuncions setmanals" i "Nous casos setmanals confirmats"

. La captura ens diu, per a la setmana de l'11/05/2020 que:

- *els casos confirmats van disminuir considerablement.* Podria tractar-se dels resultats del confinament però tal com apuntaven des de la pròpia web de descàrrega, el 14 i

el 18 de maig disminueix la xifra total de casos confirmats, per una revisió de dades del Ministeri de Sanitat.

- *el nombre d'altres va deixar d'augmentar*. Potser no se'n van enregistrar més al dataset.
- en canvi, *va haver un pic de defuncions (401) en una setmana*. Potser degut a la saturació de les UCIs o que van coincidir les defuncions d'un nombre elevat de casos confirmats de períodes anteriors.