

## A9: Creació de la visualització i lliurament del projecte

### Presentació

El títol del projecte i de la visualització és

**OCUPACIÓ DE LES AULES A LES ESCOLES CATALANES CURS 2020/2021**

La visualització s'ha fet amb Flourish Studio, i per això es pot consultar de manera pública al següent enllaç:

<https://public.flourish.studio/story/886027/>

Tant el codi en R com les fonts de dades descarregades, comentades a la primera part, com les dades preparades per a fer la visualització es poden trobar al repositori de Github:

[https://github.com/AlvaroDiazS/UOC\\_VisualizacioDades\\_Practica](https://github.com/AlvaroDiazS/UOC_VisualizacioDades_Practica)

En el Github es pot trobar els dos documents descarregats de les dades obertes de la Generalitat de Catalunya amb les dades sobre les places ofertades a les escoles catalanes i també les edats per any de la població de Catalunya.

També s'ha afegit un fitxer extra en format .json per tal de poder crear mapes. La font de dades és <https://github.com/aariste/GeoJSON-Mapas>

També s'adjunta la imatge de la portada de la visualització creada.

### Explicació

He elegit crear un seguit de visualitzacions, totes elles relacionades per a respondre les preguntes plantejades. Aquest no era el pla inicial. La primera idea era fer ús de Tableau per geolocalitzar les escoles, els municipis i comarques i a partir d'aquí poder anar responnent les preguntes amb gràfics diferents de manera interactiva. Per molt que ho he intentat, he sigut del tot incapaç de poder fer un ús entenedor del mapa en Tableau i he passat al pla B, fer un seguit de visualitzacions, gràfic de bombolles, gràfic de dònuts, gràfic de barres i també mapes per poder donar resposta a cada pregunta però a Flourish Studio.

A la primera part vaig plantejar una sèrie de preguntes que tractaré de respondre una a una si és que la visualització ho permet:

**1. Quin és el ràtio d'ocupació de les aules a Catalunya? Hi ha diferències segons el nivell de l'educació o la titularitat del centre?**

Amb la segona diapositiva de l'story creat a Flourish Studio respon clarament a aquestes preguntes, ja que es pot anar descobrint fent zoom in i zoom out per les diferents bombolles la pregunta. Es pot filtrar per nivell educatiu i la grandària de les bombolles es pot elegir, entre les places ofertades, assignacions i % de places ocupades.

En la tercera diapositiva s'arriba a l'últim nivell on en passar el ratolí surt el pop up amb les dades filtrades i també s'afegeixen altres indicadors.

**2. Quin és el percentatge de places oferides segon els tipus d'educació i tipus de centre?**

Aquesta pregunta també es resol amb les diapositives 2 i 3 jugant amb les bombolles. De fet s'ha reformulat la pregunta per saber el nombre de places ofertades, no el percentatge, ja que és més significativa.

Es pot comprovar per a la gran majoria de comarques que hi ha més oferta pública que privada i que els cursos que més places s'ofereixen són pel primer any al col·legi, amb 3 anys, i el primer any d'institut amb dotze anys.

**3. Quin percentatge d'assignacions són a 1a petició? Aquest valor canvia geogràficament?**

A la quarta diapositiva es pot respondre a aquesta pregunta. Es pot filtrar per comarca i es pot veure segons 3 indicadors, assignacions totals, assignacions en 1a petició i el % d'assignacions en 1a petició segons la titularitat del centre i el cicle educatiu.

Es pot tornar a comprovar que aquest percentatge és més elevat al 2n cicle d'educació infantil i a l'Educació Secundària Obligatoria, ja que hi ha més places ofertades i per tant és més possible que et donin la primera assignació. En altres cursos amb molt poques places i moltes sol·licituds el percentatge baixa, especialment en l'educació primària.

**4. Es necessiten més o menys aules segons el municipi?**

Aquesta pregunta no he sigut capaç de respondre-la mitjançant visualitzacions.

**5. Hi ha indrets de Catalunya on hi ha una evolució ascendent o descendent d'alumnes?**

A la cinquena diapositiva es pot comprovar aquesta evolució per comarca de quatre indicadors diferents. Per una banda hi ha les places totals, les places ofertades, les places assignades i el % de places assignades.

És l'única diapositiva que s'han fet servir les dades dels quatre anys disponibles.

## **6. Existeixen desigualtats de gènere?**

Aquesta pregunta tampoc s'ha pogut respondre. Ja vaig anticipar que seria molt complicat fer-ho i efectivament amb les dades disponibles, seria especular molt donant unes dades sobre gènere.

A banda de les diapositives comentades, s'ha afegit dos mapes, un per aules de tres anys i altre per aules de dotze anys (on hi ha més oferta de places) amb les dades geolocalitzades de cada escola del curs 2020/2021.

He pensat que era interessant afegir aquests dos mapes per dues raons. La primera perquè es pot navegar per la comarca o zona que es desitgi i poder extreure informació de cada centre. La segona és com ja he comentat, la intenció era integrar tota la visualització en un mapa, però he sigut incapaç de fer-ho i en passar al pla B de les diapositives, creia convenient fer ús d'algun mapa.

Penso que l'usuari de la visualització són les famílies que volen escolaritzar al seu fill o la seva filla amb 3 o 12 anys i vol tenir alguna informació extraestadística a més de la que puguin tenir a partir de les portes obertes.

Crec que per un perfil tècnic del Departament d'Educació es queda una mica simple i els indicadors que es descriuen es poden enfocar d'una altra manera.

## **Descripció tècnica del projecte**

Pel que fa a les dades emprades per a la visualització he fet front a diversos problemes que crec que he solucionat de manera correcta.

Per una banda quan vaig començar a geolocalitzar les escoles del curs 2020/2021 em vaig adonar que les coordenades de longituds i latituds estaven equivocades, ja que alguna d'elles si estaven ben situades per n'hi ha d'altres que no. Vaig veure que les del curs 2017/2018 si eren correctes. Una de les feines ha sigut corregir les coordenades de les escoles per poder fer la geolocalització.

Per una altra banda s'ha enriquit la base de dades principals de 2 maneres, una afegint l'edat dels estudiants i les dades del nombre de nens i nenes per edat d'aquell municipi.

Per fer aquesta fase es van corregir els codis dels municipis, ja que els de la província de Barcelona tenien 4 dígit i es va afegir un 0 al principi per tal de poder fer de manera correcta els joins entre base de dades.

També s'han calculat alguns camps o s'han recodificat alguna variable com si la classe era mixta o no.

En fer l'exportació de les dades a Flourish Studio, els accents no s'acceptava bé, i per això vaig haver d'eliminar tots i per això la visualització no en tenen. Els títols o textos extres sí que en tenen.

Per últim, he fet les agrupacions pertinents i guardats els .csv preparats per a carregar les dades.

He fet servir R en tota la preparació de les dades, des de la lectura de les fonts de dades originals, a les correccions necessàries i la preparació de les dades finals. He fet servir la llibreria dplyr, una de les més conegudes i usades per tal de fer aquestes feines comentades.

El fitxer .json de les comarques no l'he hagut de tocar amb cap moment, el vaig descarregar de l'enllaç anterior i el vaig carregar al Flourish Studio directament.

La part de la visualització s'ha fet Flourish Studio. S'han carregat les dades preparades, s'han definit quin tipus de gràfics i quines columnes es feien servir i el tuning dels paràmetres perquè la visualització quedés com volia.

Per últim, s'ha creat un repositori a Github amb els documents que s'han fet servir, tant els csv, com el json o script d'R.

La llicència és CC0-1.0 License, que permet l'ús comercial de les dades, la modificació, la distribució i l'ús privat. És a dir, pràcticament és com si no tingués llicència.

L'elecció d'aquests programaris ha sigut per una banda R, ja que estic acostumat a fer-ho servir i Flourish Studio com a alternativa i pla B al Tableau.