La visualització de dades com a eina fonamental en un món complex

Jornada Florence Nightingale
Facultat de Matemàtiques i Estadística
Universitat Politècnica de Catalunya

May 12th, 2021   **By Gerard Giménez i Adsuar**

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Gerard Giménez i Adsuar

Work experience:

- **Research Assistant (GRECS, UdG)** (Des 20' - **Currently**)
- Data Scientist for Health Department at Generalitat (Jul 20' - Des 20')
- Financial analyst at a private equity firm in Barcelona (May 19' - May 20')
- Research Assistant in economics at Pompeu Fabra University

Education:

- **MESIO student at FME UPC-UB**
- BBA graduate at UB (2018) & Biotech studies at UB (2011-2014).

- I'm here to briefly explain my experience as a data journalist focusing on COVID's impact in Catalonia

- This self-learning project has brought about many professional surprises. I believe this only displays that data science & statistics are in great demand

- On May 9th, 2020, I had 41 followers (my friends, mainly). Today I've built a community of over 26.000 people

- **Adding value is KEY!**

Before sharing my experience… what's COVID?

SARS-CoV-2 virus produces the so-called COVID19 disease. Its name is #2 since it's the second coronavirus to produce a Severe Acute Respiratory Syndrome (SARS).

The first one appeared during 2003 causing a few thousand cases, with a reported IFR about 10%. That's around 10 times more deadly than COVID19.

Why has this one become a pandemic? Almost half of the infected patients are asymptomatic and contagious. This has proved to be a **HUGE** problem. **Seemingly fine people were spreading the virus unknowingly**, and its spread was hugely asymmetrical: **some didn't infect at all, whereas others infected 20 other individuals in a matter of a hours**

Some worldwide data

More than **150 million COVID19 cases** have been confirmed to date. Real incidence is **strictly higher** due to limited testing.

With more than **3.200.000 confirmed deaths**. This is also an underestimate.

Real incidence and deaths will only be available after the proper **excess deaths estimates for the period**. It will prove to be an arduous task, to say the least, since non democratic countries are highly reluctant to share data revealing the whole extent of the COVID19 crisis.

Let's talk about my little story...

Can we get data of the pandemic at a **local level**, almost real-time?

Yes, thanks to [Dades Obertes](#) de Catalunya! This is the raw data that we can get:

TipusCas ↓	ComarcaC...	ComarcaD...	MunicipiC...	MunicipiD...	SexeCodi	SexeDescr...	TipusCasD...	NumCasos
01/07/2020	07	Bages	08047	Cardona	1	Dona	Positiu per ELISA	1
01/07/2020	35	Solsonès	25207	Solsona	0	Home	Sospitós	1
01/07/2020	24	Osona	08183	Roda de Ter	1	Dona	Sospitós	2
01/07/2020	33	Segrià	25120	Lleida	1	Dona	Sospitós	10
01/07/2020	41	Vallès Oriental	08096	Granollers	0	Home	Positiu per ELISA	1
01/07/2020	07	Bages	08113	Manresa	1	Dona	Positiu PCR	1
01/07/2020	35	Solsonès	25166	Pinell de Solso...	0	Home	Sospitós	1
01/07/2020	19	Garrotxa	17185	Sant Joan Les F...	1	Dona	Sospitós	5
01/07/2020	21	Maresme	08219	Vilassar de Mar	1	Dona	Sospitós	2
01/07/2020	03	Alt Penedès	08091	Gelida	0	Home	Sospitós	1
01/07/2020	21	Maresme	08035	Calella	1	Dona	Sospitós	1

At first, I just wanted to see the evolution of weekly cases for my village… but ended up doing it for all the municipalities

It was more informative to analyse weekly cases rather than daily cases, specially for the ones with small populations.

Also, we needed to filter for active cases, rather than Ag positive cases (which show past infection):

```
#Import the csv
mun22m<- read.csv("https://analisi.transparenciacatalunya.cat/api/views/jj6z-iyrp/rows.csv?accessType=DOWNLOAD&sorting=true")
head(mun22m)
mun22m_2 <- mutate(mun22m, data = as.Date(TipusCasData, "%d/%m/%y"))
head(mun22m_2)


positius_municipi <- mun22m_2 %>%
  filter(TipusCasDescripcio %in% c("Positiu PCR", "Positiu per Test Ràpid")) %>%
  filter(!ComarcaDescripcio %in% "") %>%
  group_by(ComarcaCodi, data) %>%
  summarize(total_positius = sum(NumCasos))
```

After a bit of data wrangling I finally got:

Municipi	Fa 4 setm.	Fa 3 setm.	Penúltima setm.	Última setm.	Dif. última setm.	% canvi	Últ. 3 dies
Lleida	89	51	52	149	↑ 97	↑ 186%	24
Barcelona	235	181	148	142	↓ -6	↓ -4%	52
Alcarràs	9	3	17	29	↑ 12	↑ 71%	2
L'Hospitalet de Llobregat	42	50	33	27	↓ -6	↓ -18%	3
Guissona	24	12	23	16	↓ -7	↓ -30%	0
Manresa	15	23	12	14	↑ 2	↑ 17%	6
Badalona	13	12	9	12	↑ 3	↑ 33%	1
Terrassa	11	4	2	12	↑ 10	↑ 500%	3
Aitona	0	0	1	11	↑ 10	↑ 1000%	2
Santa Coloma de Gramenet	9	8	10	11	↑ 1	↑ 10%	1

Making it accessible to everyone by means of [a shiny app](#):

Positius COVID19, per tots els municipis de Catalunya

Act. 02/07/2020

Show 10 ∨ entries   Search: [ ]

	Municipi	Fa 6 setm.	Fa 5 setm.	Fa 4 setm.	Fa 3 setm.	Penúltima setm.	Última setm.	Dif. última setm.	% canvi	Últ. 3 dies	Total positius	% població positiva	Habitants
1	Lleida	116	94	85	47	68	182	114	167.6	65	1578	1.14	138956
2	Barcelona	574	477	217	183	145	142	-3	-2.1	235	19248	1.18	1636762
3	Alcarràs	4	8	9	2	23	37	14	60.9	8	122	1.28	9514
4	L'Hospitalet de Llobregat	57	55	46	45	34	31	-3	-8.8	16	3142		
5	Aitona	11	0	0	0	1	17	16	1600	5	84	3.24	2591
6	Guissona	24	17	23	9	22	16	-6	-27.3	3	232	3.25	7136
7	Manresa	18	20	17	21	12	14	2	16.7	11	1008	1.3	77714
8	Terrassa	27	18	6	4	3	14	11	366.7	0	1903	0.86	220556
9	Badalona	28	41	11	11	10	12	2	20	4	2127	0.96	220440
10	Martorell	8	10	2	7	2	10	8	400	1	351	1.25	28189

Showing 1 to 10 of 723 entries   Previous 1 2 3 4 5 … 73 Next

Els positius de les últimes setmanes inclouen: Positius PCR, Positius per Test Ràpid. Els positius totals, també inclouen els Positius per ELISA i Epidemiològic

Elaborat per Gerard Giménez Adsuar, amb les Dades Obertes de Catalunya. Els errors són la meva responsabilitat.

Per més informació, Twitter: @gmnzgerard

Web's *feedback* has been extremely positive

For several months, this little shiny app has received more than **a thousand daily visits**

I have also come to realise that more often than not, specially in the "fake news" era, information tends to calm down people!

But… how's the epidemic at the comarca level?

- I wanted to get in just one picture the dynamics of each comarca, with just a glimpse.

- Following **Robert Koch Institut's guidance,** I realised that the key indicator was weekly cases per 100.000 inhabitants

- Since the objective of all this was to prepare for the MESIO, this proved to be a nice programming challenge!
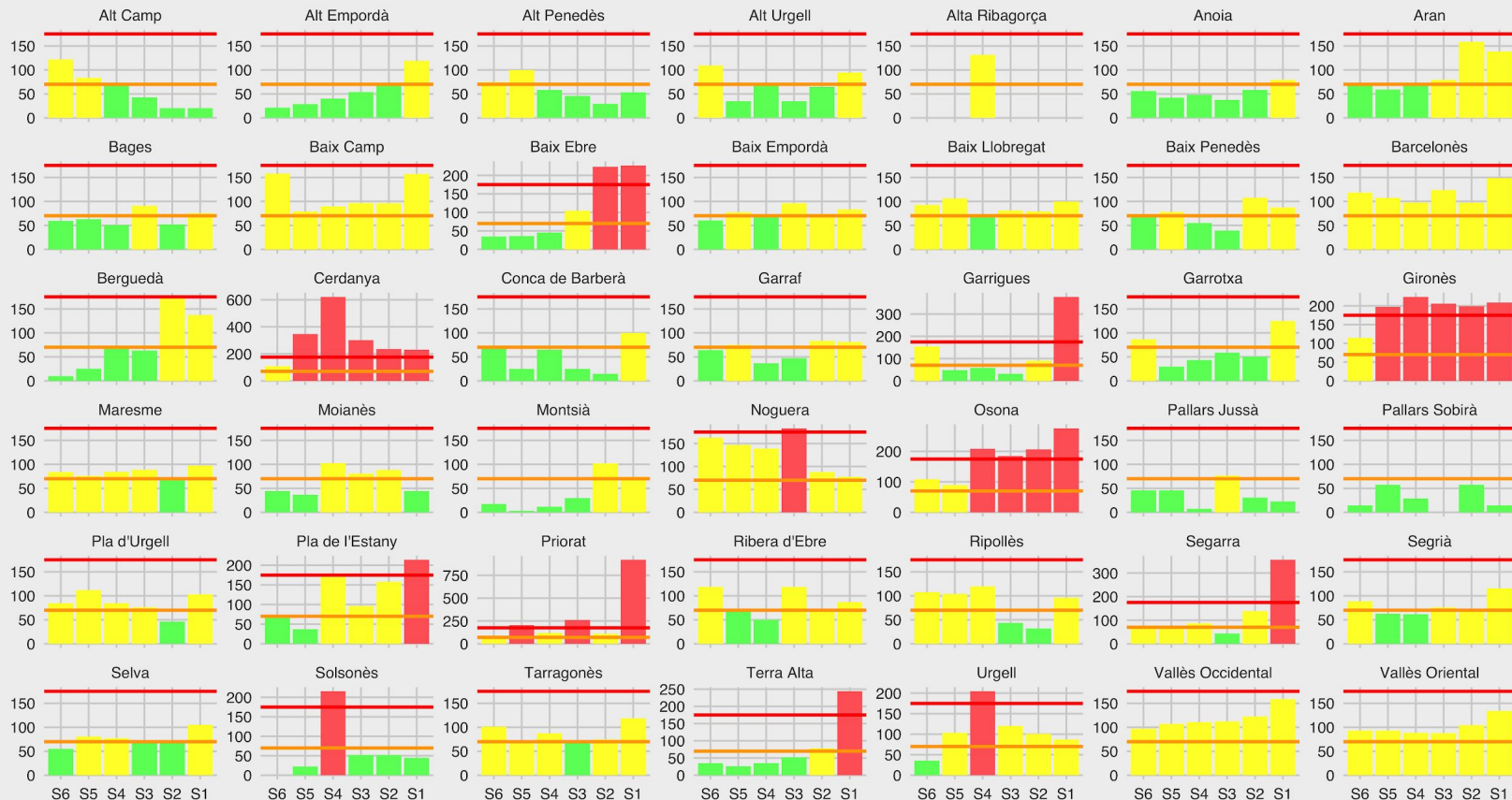
This was the result:

CATALUNYA: Nous PCR positius COVID19 per cada 100.000 habitants, últimes 6 setmanes fins al 07/10/2020

Segons la guia de Harvard, a partir de 175 casos / 100.000 hab. totes les escoles haurien de tancar (vermell).
Entre 175 i 70, només poden obrir Infantil i Primària (groc).
A menys de 70 (verd) pot estar tot obert, amb mesures.

Nota: actualitzat el 10/10/2020 amb les Dades Obertes de Catalunya. Els últims 3 dies no inclouen tots els positius, per aquest motiu, no els considerem
S1 és l'última setmana, del 07/10/2020 al 01/10/2020 , les altres, són les 5 setmanes precedents
Per Gerard Giménez Adsuar, Twitter: @gmnzgerard
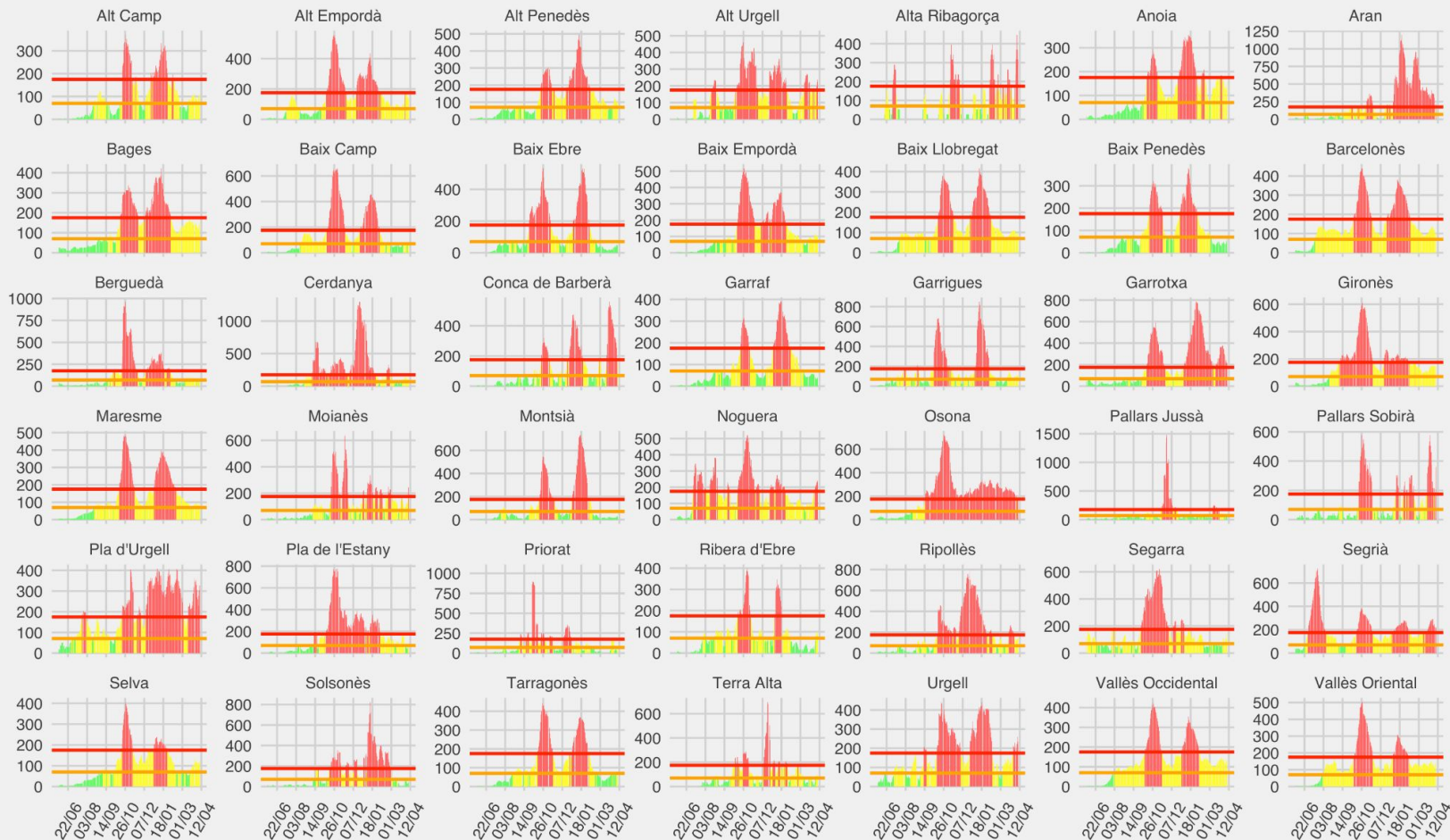
Maximizing the information in just one image

I wanted to show:

1. Last 6 weeks' evolution (upward, stable or downward trend)
2. Immediate incidence level recognition using the color coding (red, yellow and green)

This was the result of several iterations. Interpreting plots can be trivial for us… however for the untrained it can be challenging. **My goal was to bridge that gap while maintaining its rigour.** Data visualization matters **A LOT**.

Looking at the last 10 months, we can readily see the history of the several waves from this pandemic...

CATALUNYA COVID19: Positius setmanals PCR/TA per cada 100.000 habitants, últims 10 mesos

Nota: actualitzat el 12/04/2021 amb les Dades Obertes de Catalunya. Els últims 3 dies no inclouen tots els positius, per aquest motiu, no els considerem
Per Gerard Giménez Adsuar, Twitter: @gmnzgerard

From sharing plots on Twitter, to working for the Health Department at Generalitat

They were looking for data scientists as the need for data visualizations was growing. I had a case of imposter syndrome since my knowledge was very limited.

However, that was enough. On my first day, I had ready a daily report to be sent to hospitals and city halls summarizing the situation from the vast datasets available (much more info than the open data, obviously!).

Moving forward, there was the challenge to automatically identify outbreaks...

My small script for outbreak identification:

By filtering and matching across cases and contacts, I could automate outbreak detection and increase it by 30%

Large outbreaks weren't an issue - they were already detected - but small ones (3-4-5 people) went easily undetected… causing serious issues.

Brots detectats automàticament, dia: 2020-09-16

```
## [1] "Tenim  0  brot(s) amb almenys 7 casos."

## [1] "Tenim  0  brot(s) amb almenys 6 casos."

## [1] "Tenim  0  brot(s) amb almenys 5 casos."

## [1] "Tenim  2  brot(s) amb almenys 4 casos."

## [1] "Tenim  6  brot(s) amb almenys 3 casos."

## [1] "Tenim  24  possibles brot(s) amb almenys 2 casos."
```

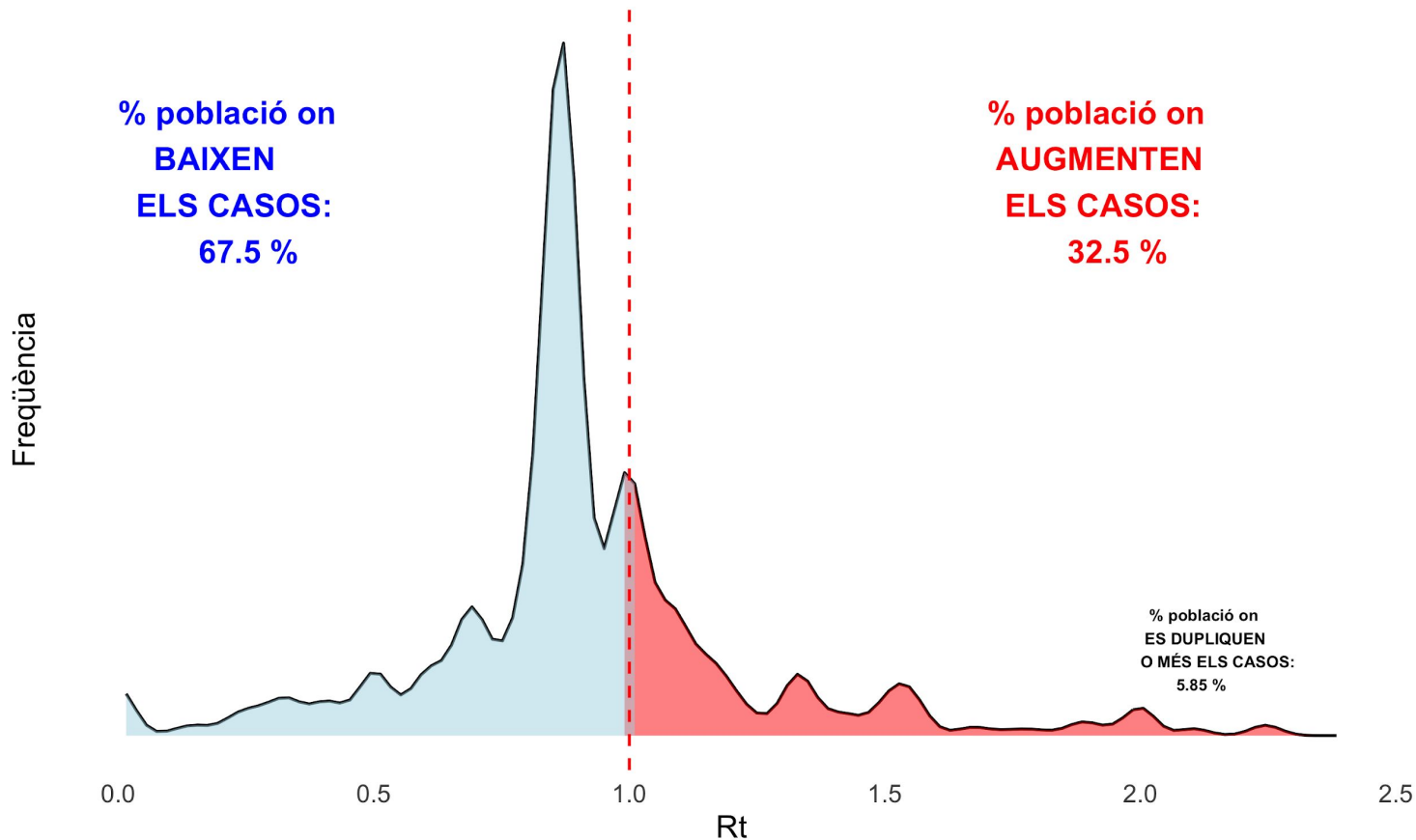Aquesta és la taula resum dels brots detectats:

Mida dels brots, nº de casos	Nombre de brots
2	24
3	6
4	2
5	0
6	0
7	0

Some of my other data visualizations...

The pdf for the "famous" Rt

Distribució de la Rt a Catalunya

**% població on
BAIXEN
ELS CASOS:
67.5 %**

**% població on
AUGMENTEN
ELS CASOS:
32.5 %**

**% població on
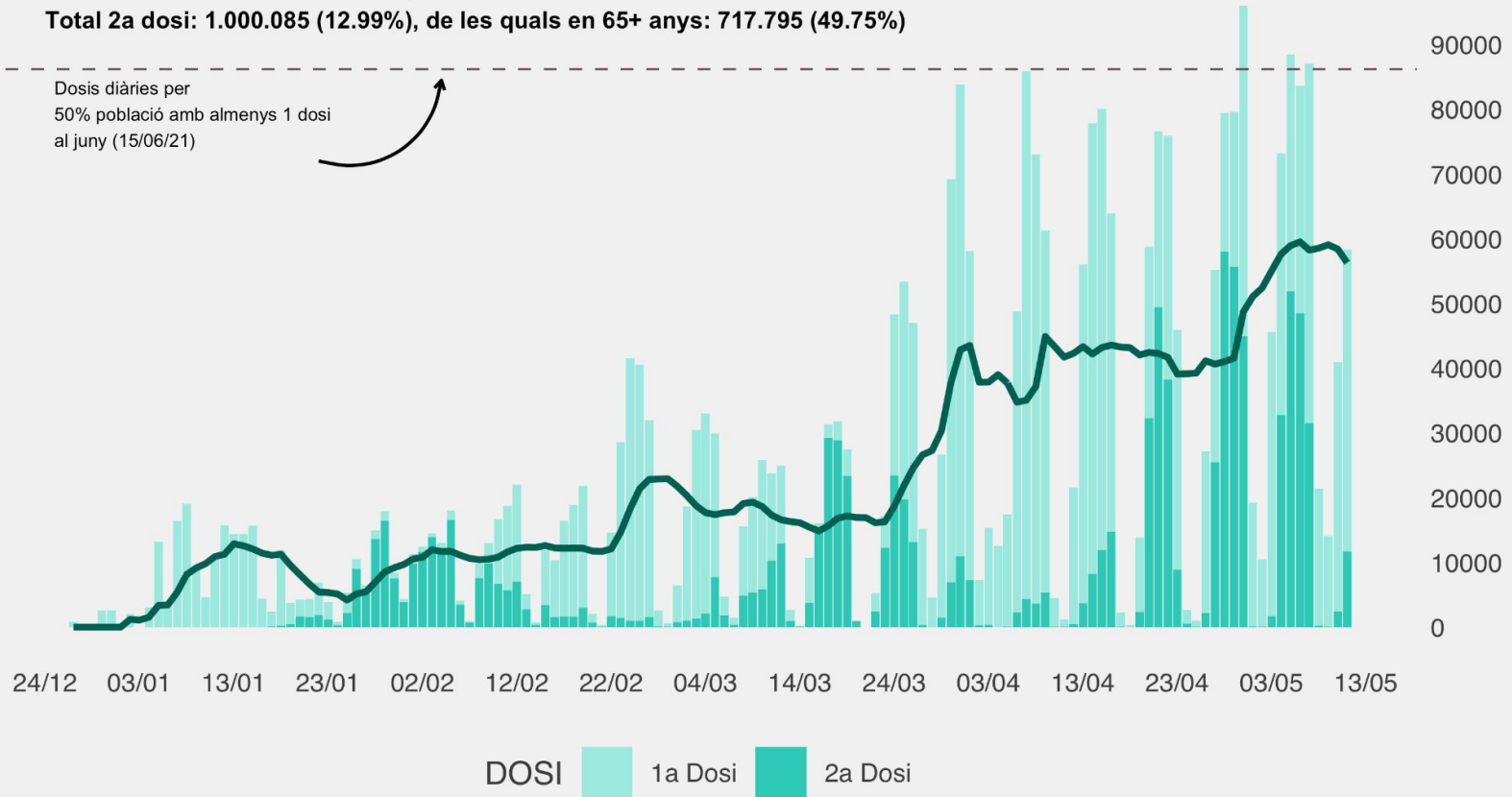ES DUPLIQUEN
O MÉS ELS CASOS:
5.85 %**

Freqüència

Rt

Vacunacions diàries de la COVID19 a Catalunya, act: 12/05/2021

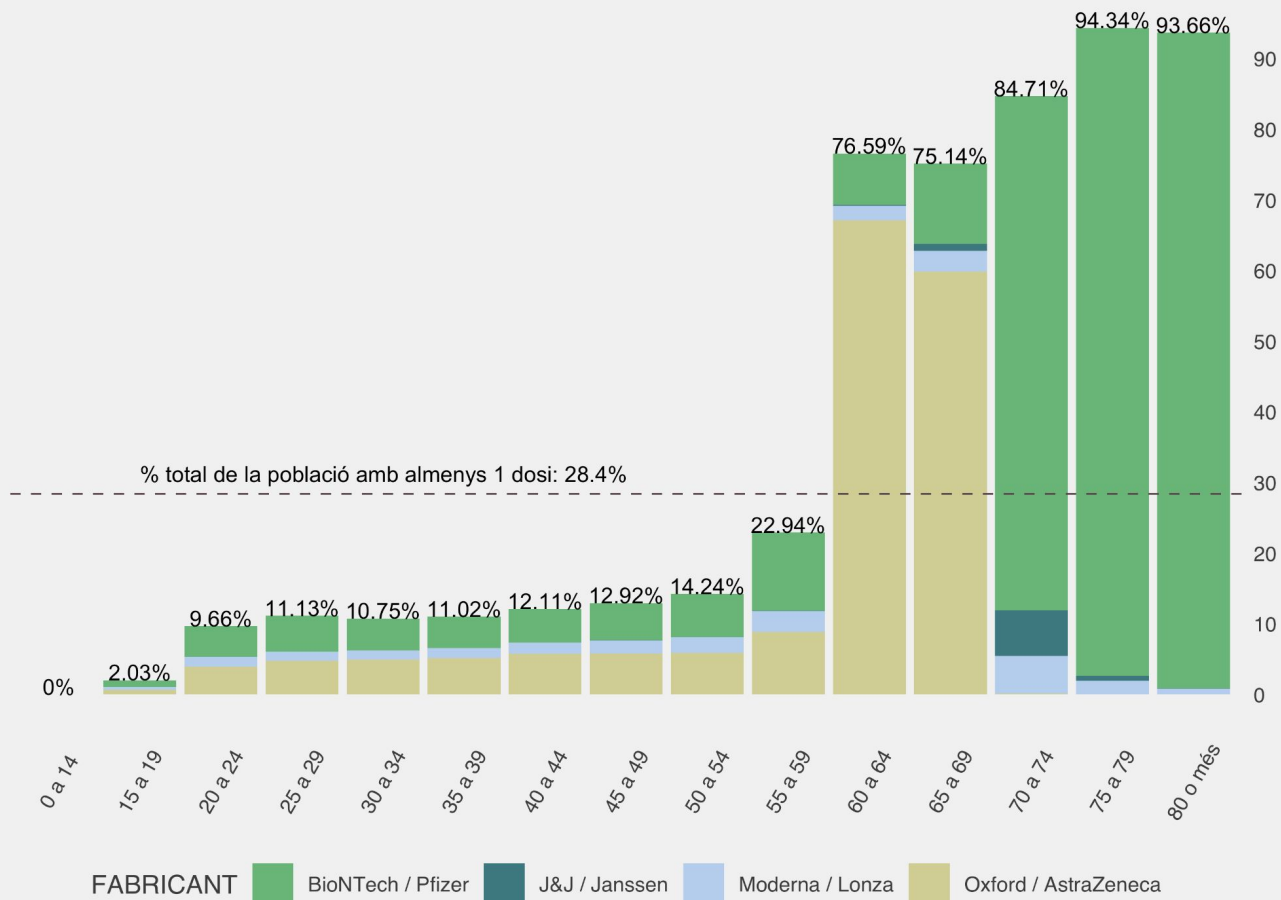**Total 1a dosi: 2.229.360 (28.95%), de les quals en 65+ anys: 1.285.121 (89.07%)**
Total 2a dosi: 1.000.085 (12.99%), de les quals en 65+ anys: 717.795 (49.75%)

Dosis diàries per
50% població amb almenys 1 dosi
al juny (15/06/21)

DOSI   1a Dosi   2a Dosi

Gràfica per @gmnzgerard. Font: dadescovid.cat

21

% de la població almenys 1 dosi a CAT, per edats

% total de la població amb almenys 1 dosi: 28.4%

0% 2.03% 9.66% 11.13% 10.75% 11.02% 12.11% 12.92% 14.24% 22.94% 76.59% 75.14% 84.71% 94.34% 93.66%

0 a 14 · 15 a 19 · 20 a 24 · 25 a 29 · 30 a 34 · 35 a 39 · 40 a 44 · 45 a 49 · 50 a 54 · 55 a 59 · 60 a 64 · 65 a 69 · 70 a 74 · 75 a 79 · 80 o més

FABRICANT BioNTech / Pfizer   J&J / Janssen   Moderna / Lonza   Oxford / AstraZeneca

Act. 12/05/2021. Gràfica per @gmnzgerard. Font: dadescovid.cat

Key takeaways

- **Open data** is key for a modern democracy, as people can **join efforts towards a common goal**.
- The demand for **data scientists** has been on the **rise** for more than a decade.
- **MESIO at FME UPC-UB is an excellent stepping stone** for acquiring such skills.
- In this **world aflush with data**, knowing the tools to analyse it is a **superpower**.
- **Always bear in mind which is your audience, and adapt the message accordingly**

Thank you very much for your attention!

I am more than happy to answer your questions