

Media analysis workshop: a gateway for data journalism practice

Dr Xavier Ramon

Universitat Pompeu Fabra (Barcelona, Spain)

EJTA Teachers' Conference 2021



Universitat
Pompeu Fabra
Barcelona



Dataharvest
The European Investigative
Journalism Conference

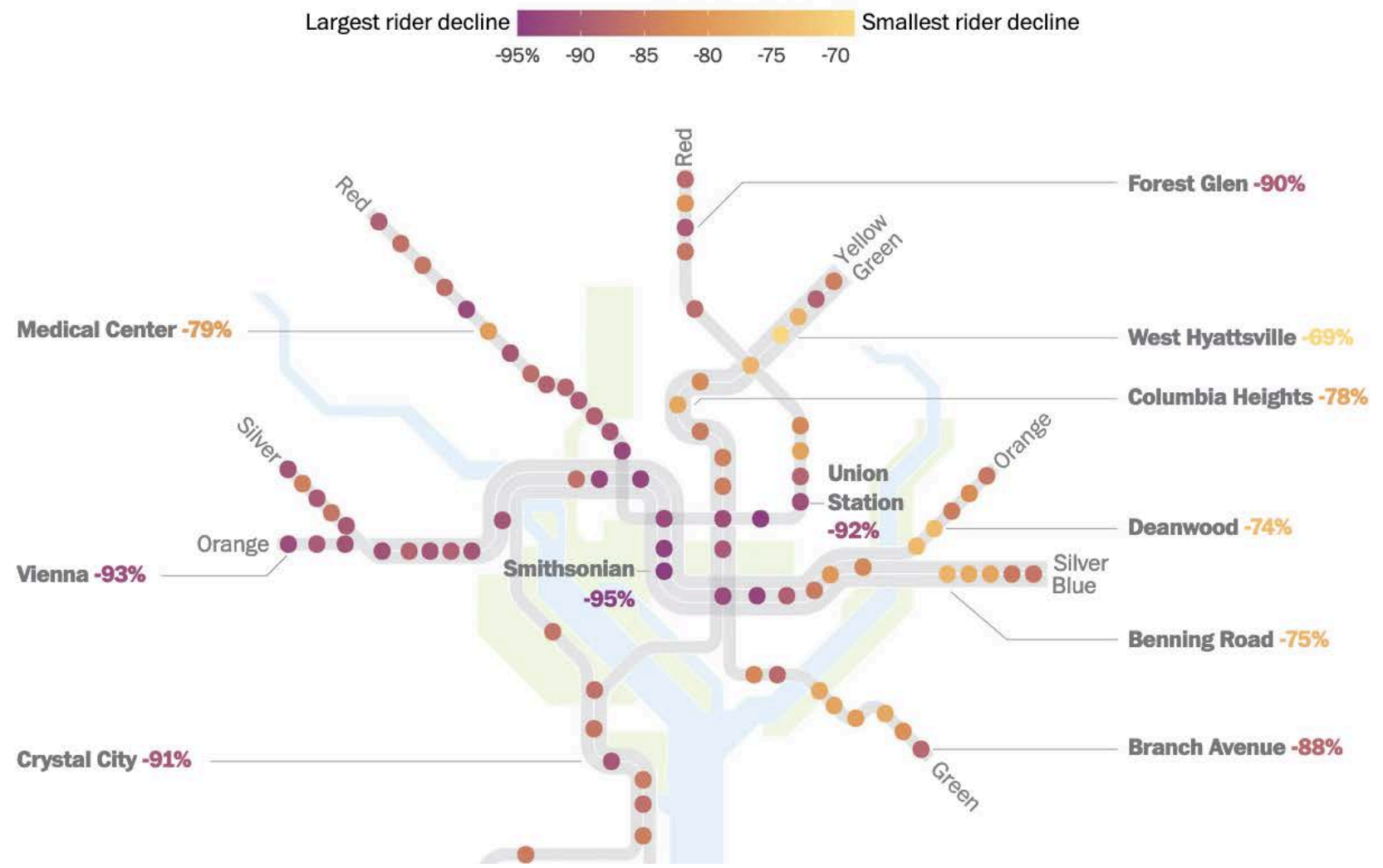
Introduction

- In a challenging and ever-evolving landscape, “journalism schools have an obligation to provide the **latest skills and techniques** to make students competitive for the news market” (Weiss & Retis-Rivas, 2018, p. 2).
- Bearing in mind the growing importance of numerical treatment and representation (Lewis, 2021), **data journalism training should be expanded** across undergraduate and graduate programmes.



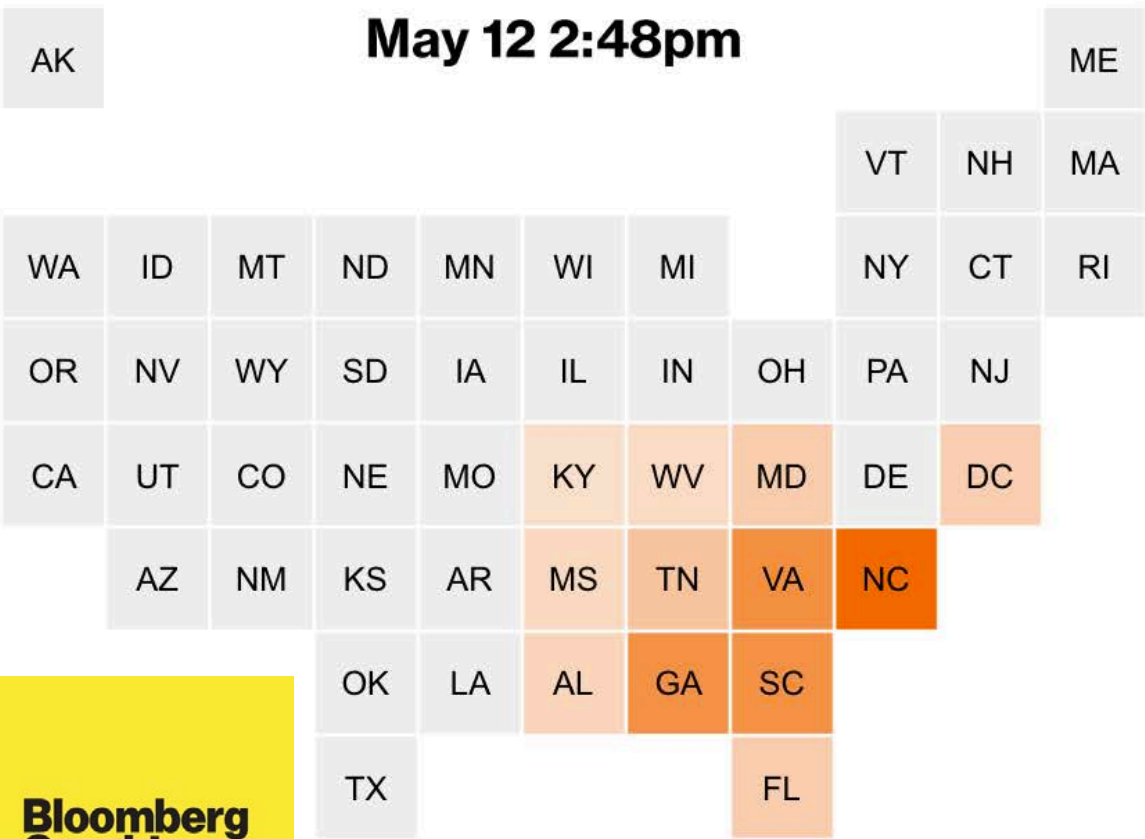
Decline in ridership on Metrorail

Change in median daily ridership, Jan. 1 to March 5, 2020 vs. 2021



Gasoline Station Outages

The percentage of gas stations without fuel in each state through May 12

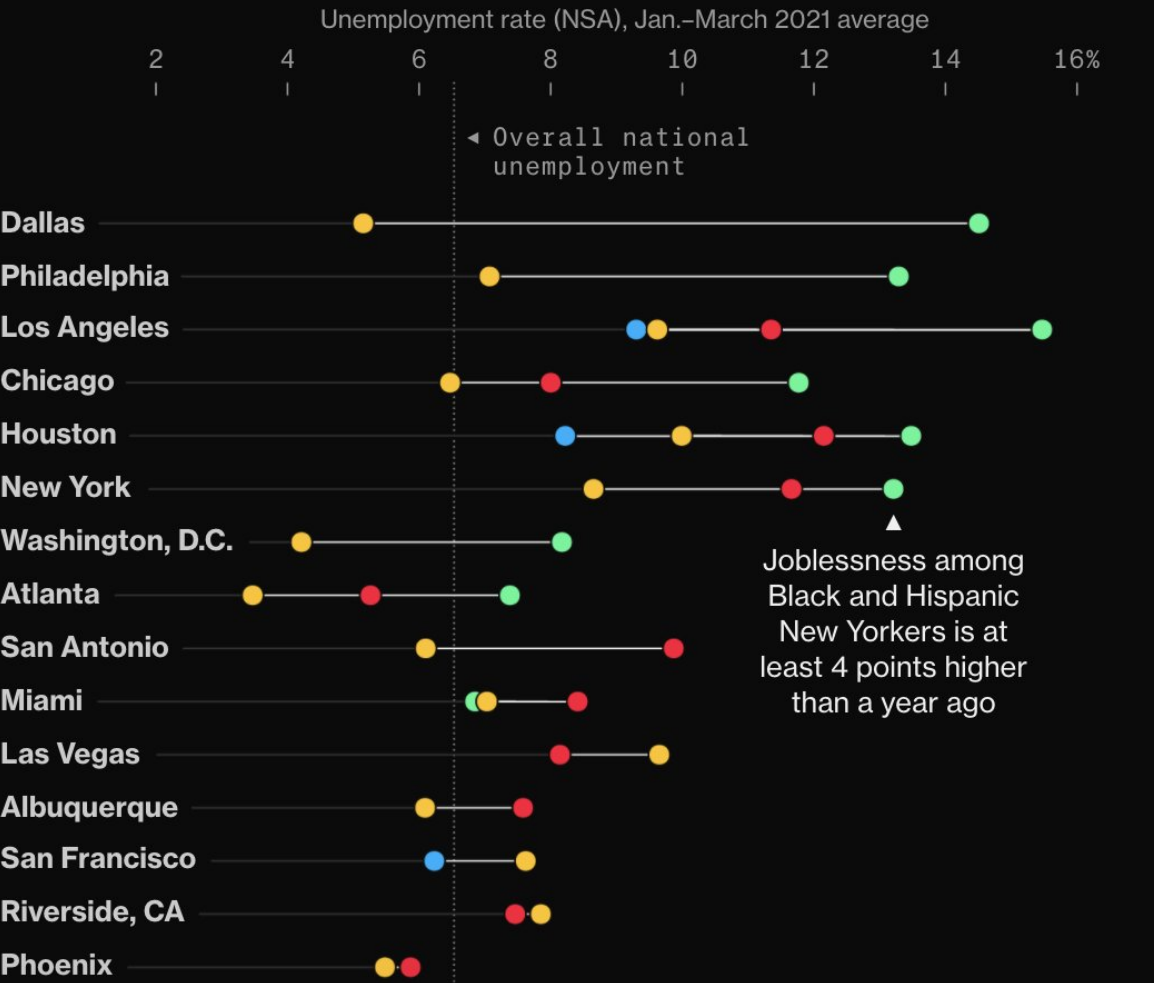


Bloomberg Graphics

Double-Digit Divide

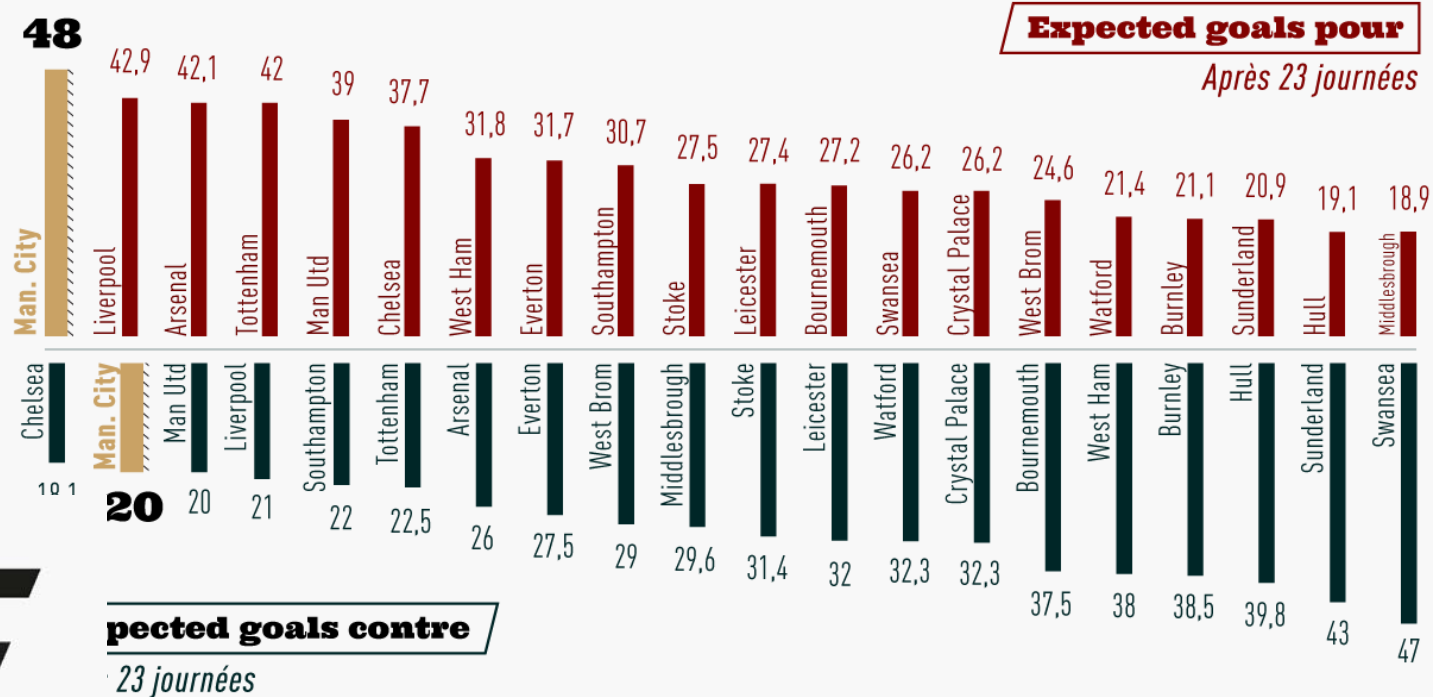
Minority communities face substantial unemployment gaps, though only wider disparities tend to be meaningful in this analysis because of sample-size limitations of local data

Asian Black Hispanic White



DES OCCASIONS À LA PELLE

Depuis le début de la saison 2016-2017, Guardiola met en avant le manque de réussite et d'efficacité offensive et défensive de son équipe. Les « expected goals » renforcent cette idée : City créé beaucoup d'occasions et en concède peu.



« Expected goals » c'est quoi?

Cet outil statistique en constante évolution tente d'estimer le nombre de buts qu'une équipe « aurait dû » marquer. Partant du principe que le nombre de tirs par match ne suffit pas à mesurer la qualité des occasions, on cherche à attribuer à chaque tir une « cote de réussite » qui dépendra de plusieurs paramètres (distance au but, surface du corps utilisée, type de passe reçue, etc.) basés sur des milliers de tirs tentés par le passé. En gros, deux tirs à cinq mètres du but « valent mieux » que quatre frappes décochées des 30 mètres. Précisons que les expected goals - dont l'une des limites est d'exclure les actions qui n'aboutissent pas à un tir - ne constituent pas en eux-mêmes une analyse et ne peuvent se substituer à l'observation des matches.

Introduction

- Students are **attracted to the breadth and clarity** of data journalism produced by leading organizations such as *The Washington Post*, *Bloomberg*, *The New York Times*, *The Economist* or *L'Équipe*.
 - However, many of them are still **afraid of the complexity** that data analysis and visualization entails.
- Therefore, as journalism educators we must **design strategies** to challenge students' preconceptions and counteract their lack of data literacy (Bradshaw, 2018).
- At Universitat Pompeu Fabra, we **capitalize on students' interest in media analysis** to introduce them to data analysis and visualization practices.

Introduction

Undergraduate course ‘Research Techniques Applied to Communication’ (21382)

- As part of this course, students engage in a **two-week media analysis workshop** that equips them with the foundations to work with data. In groups, students develop four small projects that involve manual coding and automatic data collection.



Newspaper analysis

Twitter analysis

YouTube analysis

TV news analysis

AI. Newspaper analysis



Description of the project

- To examine how US press covered the Boston marathon explosions the day after the events (16 April, 2013). Students manually code a sample of **180 newspaper frontpages**.
- The classroom is divided in groups, following the nine regional divisions established by the US Circulation Bureau.
 - New England; Mid-Atlantic; East North Central; West North Central; South Atlantic; East South Central; West South Central; Mountain; and Pacific.
 - Newspapers in each region are selected bearing in mind strategic criteria such as their circulation, history and impact.

AI. Newspaper analysis

Description of the project

- Each group is responsible for a regional division and has to **code a sub-sample** of 20 frontpages. Students receive different **materials**:
 1. PDF files of the frontpages
 2. Codebook (18 categories)
 3. Excel database
- After coding the materials, all teams work collaboratively to obtain results and visualize data.



Database.xlsx

Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista

Calibri (Cuer... 12 A- A- Ajustar texto General Formato condicional Dar formato como tabla

Normal 2 Normal Bueno Incorrecto Neutral Cálculo Celda de com... Celda vinculada Entrada Notas Insertar Eliminar Formato Autosuma Rellenar Ordenar y filtrar

	A	B	C	D	E	F	G	H	I	J	K	L
1	A. Dades contextuais de l'objecte d'estudi				B. Agenda		C. L'ús de la fotografia					
2	A1. Data	A2. Regió	A3. Nom del mitjà	A4. Estat	B1. Extensió	B2. Obertura	C1. Fotografies	C2. Número fotografies	C3. Sang	C4. Menors	C5. Font foto principal	C6. Infograf
3	16/4/13	5	The Washington Post	DC	2	1	1	2	2	2	1	
4	16/4/13	2	The Times-Tribune	PA	2	1	1	4	2	2	1	
5	16/4/13	1	Portland Press Herald	ME	3	1	1	2	2	2	3	
6	16/4/13	9	Mail Tribune	OR	1	1	1	1	2	2	4	
7	16/4/13	9	San Jose Mercury News	CA	1	1	1	3	1	2	7	
8	16/4/13	3	Detroit Free Press	MI	2	1	1	2	1	2	1	
9	16/4/13	5	The News Observer	NC	2	1	1	2	1	2	1	
10	16/4/13	1	Cape Cod Times	MA	3							
11	16/4/13	3	Evansville courier & press	IN	3							
12	16/4/13	8	The Arizona Republic	AZ	3							
13	16/4/13	2	The Buffalo News	NY	3							
14	16/4/13	1	Connecticut Post	CT	3							
15	16/4/13	5	The News Journal	DE	3							
16	16/4/13	9	Los Angeles Times	AK	2							
17	16/4/13	4	Star Tribune	MN	2							
18	16/4/13	8	Arizona Daily Star	AZ	2							
19	16/4/13	2	The Press of Atlantic City	NJ	2							
20	16/4/13	5	Richmond Times-Dispatch	VA	3							
21	16/4/13	1	Concord Monitor	NH	2							
22	16/4/13	2	The Philadelphia Inquirer	PA	2							
23	16/4/13	5	The Herald-Mail	MD	2							

Database.xlsx

Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista

Calibri (Cuer... 12 A- A- Ajustar texto General Formato condicional Dar formato como tabla

Normal 2 Normal Bueno Incorrecto Neutral Cálculo Celda de com... Celda vinculada Entrada Hipervinculo Insertar Eliminar Formato Autosuma Rellenar Ordenar y filtrar

	A	B	C	D	E	F	G	H	I	J
14										
15	C1. Does the frontpage include photographs from the Boston marathon explosions?									
16										
17		N	%							
18	Yes	179	99,44							
19	No	1	0,56							
20	Total	180	100,00							
21	Source: Authors' analysis									
22										
23	C2. Number of photographs dedicated to the Boston marathon explosions									
24										
25		N	%							
26	One photograph	51	28,33							
27	Two photographs	97	53,89							
28	Three photographs	20	11,11							
29	Four photographs	9	5,00							
30	More than four photographs	3	1,67							
31	Total	180	100,00							
32	Source: Authors' analysis									
33										

Step I

Students produce tables with descriptive statistics

Step 2

Explore tools integrated in Google's Journalism Studio to produce graphics

A collection of tools to empower journalists to do their work more efficiently, creatively, and securely



<https://journaliststudio.google.com>

Extension dedicated to the Boston marathon explosions on frontpage

■ Less than half page ■ From half page to one page ■ One page ■ Double page

From half page to one page



One page



Less than half page

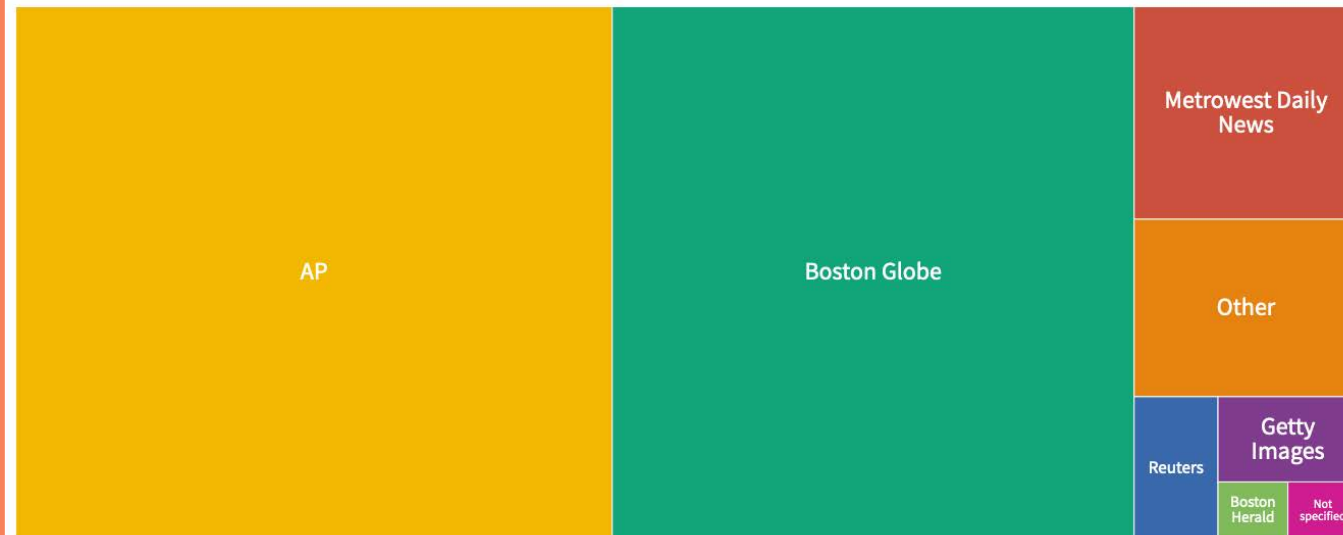


Double page



Source: Auth

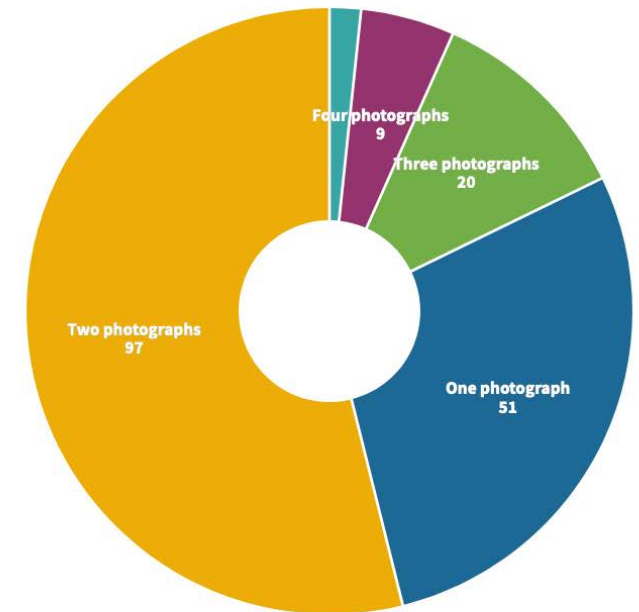
Source of the main photograph



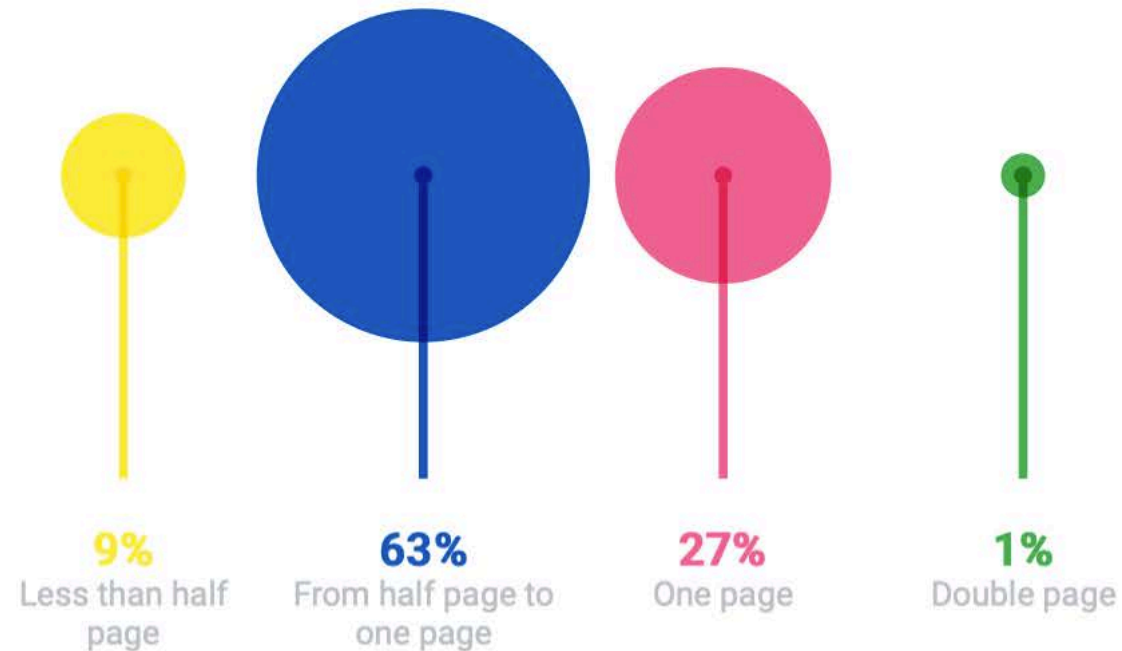
Authors' analysis

Number of photographs dedicated to the Boston marathon explosions

■ One photograph ■ Two photographs ■ Three photographs ■ Four photographs ■ More than four photographs



Extension on frontpage



Extension on frontpage

VALUE TYPE



123

PREVIEW



SAVE GIF



TERM ONE	Less than half page	16	X
TERM TWO	From half page to one page	114	X
TERM THREE	One page	48	X
TERM FOUR	Double page	2	X

+ ADD TERM

A2. Twitter analysis

Description of the project

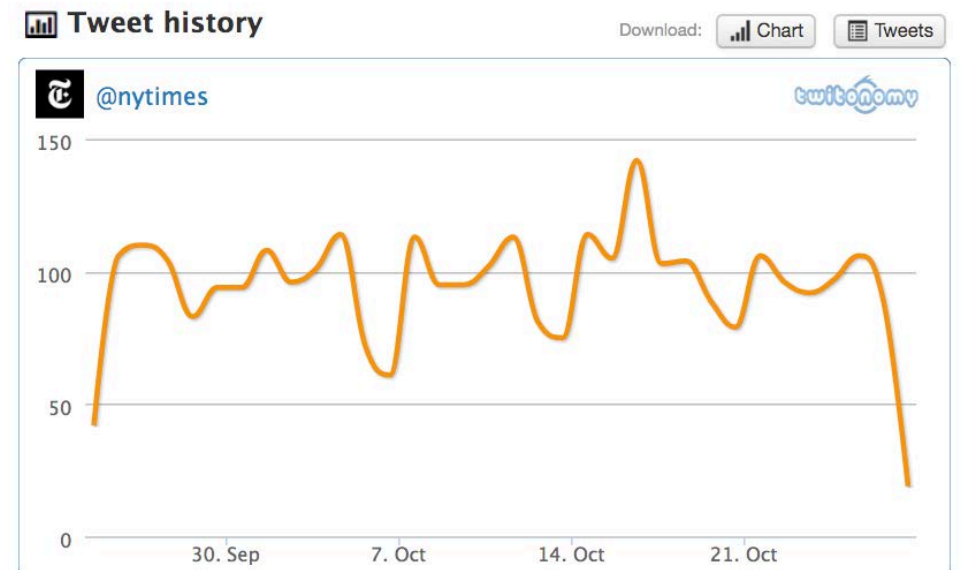
- By employing Twitonomy (<http://twitonomy.com>), a specialized tool created by Digitonomy, students **scrap data from different Twitter accounts** to compare the volume and pace of publication, the use of links and hashtags and the content that generates more engagement in those accounts.



Twitter #analytics and much more...

- ✓ Get detailed and visual **analytics** on anyone's tweets, retweets, replies, mentions, hashtags...
- ✓ Browse, search, filter and get **insights** on the people you follow and those who follow you
- ✓ **Backup/export** tweets, retweets, mentions and reports to Excel & PDF in just one click
- ✓ Monitor your interactions with other Twitter users: **mentions**, retweets, favorites...
- ✓ Get and export **Search Analytics** on any keywords, #hashtags, URL or @users
- ✓ Get insights on and download any user's **retweeted & favorited tweets**
- ✓ Browse, search, filter, sort and batch add/remove people to your **lists**
- ✓ **Monitor** tweets from your favorite users, lists and keyword searches
- ✓ Get actionable insights on your followers with **Followers Report**
- ✓ Find out easily those you follow but **don't follow you back**
- ✓ **Download** your followers and following lists to Excel
- ✓ Get the list of the **followers** you don't follow back
- ✓ **Track clicks** on the links in your tweets
- ✓ Track your **follower growth** over time
- ✓ And much more...

Get started, try Twitonomy now!



A2. Twitter analysis

Description of the project

- Students produce **comparative tables** and have to think about the most suitable type of **chart** to represent some of the obtained data.

	A	B	C	D	E
21			TIME	The New Yorker	
22		Analysed tweets	1238	492	
23		Tweets per day	82,43	32,8	
24		Retweets	42	20	
25		% of tweets being retweets	3,39	4,07	
26		User mentions	150	79	
27		Mentions per tweet	0,12	0,16	
28		Replies	13	13	
29		% of tweets being replies	1,05	2,64	
30		Links	1166	448	
31		Links per tweet	0,94	0,91	
32		Hashtags	90	16	
33		Hashtags per tweet	0,07	0,03	
34		Tweets retweeted	1193	472	
35		% of tweets being retweeted	96,37	95,93	
36		Total number of retweets	112260	32537	



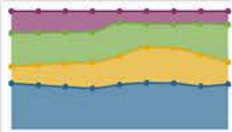
Choose a template

Show [Featured](#) Favourites Mine Company All by Flourish

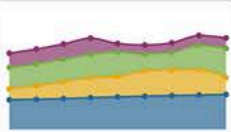
Line, bar and pie charts

Basic types of chart, single or in a grid

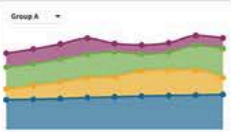
STARTING POINTS ?



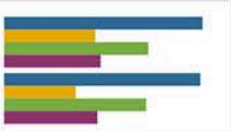
Area chart (proportional)



Area chart (stacked)



Area chart (with menu)



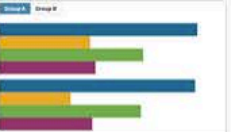
Bar chart



Bar chart (proportional)



Bar chart (stacked)



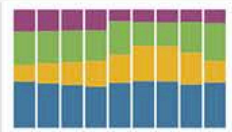
Bar chart (with menu)



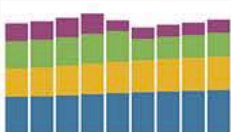
Column chart



Column chart (grouped)



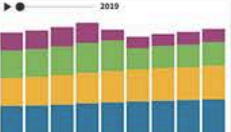
Column chart (proportional)



Column chart (stacked)



Column chart (with highlight)



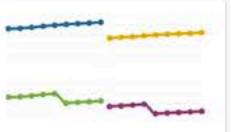
Column chart (with menu)



Donut chart



Grid of column charts



Grid of line charts



Grid of pie charts



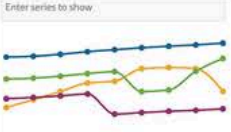
Line + Column



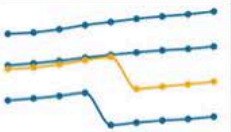
Line chart



Line chart (projected)



Line chart (searchable)



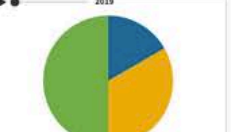
Line chart (with highlight)



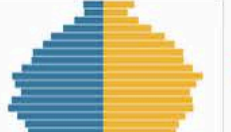
Multiseries grid



Pie chart



Pie chart with time slider



Population pyramid



Streamgraph

Projection map

Make data maps with region-shading and/or points. Adapt an existing example or upload your own geographic boundary file.

STARTING POINTS ?

<https://flourish.studio/>



Create stunning presentations, infographics, and more in just seconds

Make something that matters

Create your free account



<https://www.genial.ly/en>

Enrich your stories with charts, maps, and tables.

Start creating

It's free & no sign-up is required

apocalypse” of announced store closings meant that more jobs than normal left the economy during the month. This month, the disappointing March number was revised down from 98,000 to 79,000. But the April jobs report provides a bounce back in part because of warmer weather and fewer layoffs. The Labor Department reported gains in hospitality, mining, healthcare, and finance. Including the revisions for the February and March reports, an average of 174,000 jobs were added per month over the last three months.

Monthly Changes in U.S. Employment (Non-Farm), 2006-2017



3. An interest-rate hike in June now looks more likely.

Jeder Vierte hat keinen beruflichen Abschluss

Berufsqualifikations-Abschlüsse

ohne Migrationshintergrund mit Migrationshintergrund



Auch beim Berufsabschluss ist es ähnlich: 27 Prozent der Migrationshintergrund haben keinen Beruf gelernt und kein abgeschlossen, deutlich mehr als unter den Deutschen ohne Migrationgeschichte. Tendenz ist die Neuzuwanderer besser qualifiziert als die Migranten, die seit 20, 30 Jahren in Deutschland leben. "Deshalb sind bei Abitur und Hochschulabschlüssen die Anteile inzwischen vergleichbar mit der deutschen Bevölkerung." Mit den Flüchtlingen habe sich das Qualifikationsniveau der Neuzukommenden jedoch verschlechtert, sagt Brückner, dass es in den meisten Herkunftsländern ein duales Ausbildungssystem wie in Deutschland gebe.

Migranten und ihre Kinder verdienen weniger

Monatliches Nettoeinkommen in Euro



<https://www.datawrapper.de>

A3.YouTube analysis



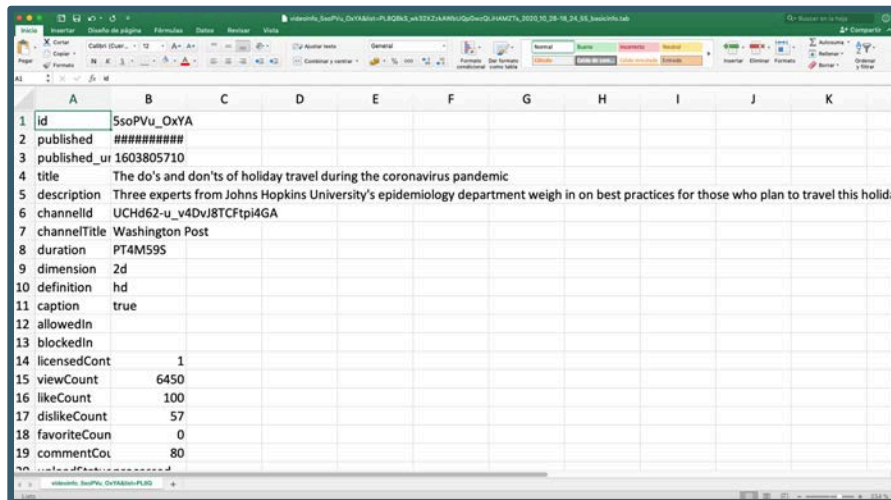
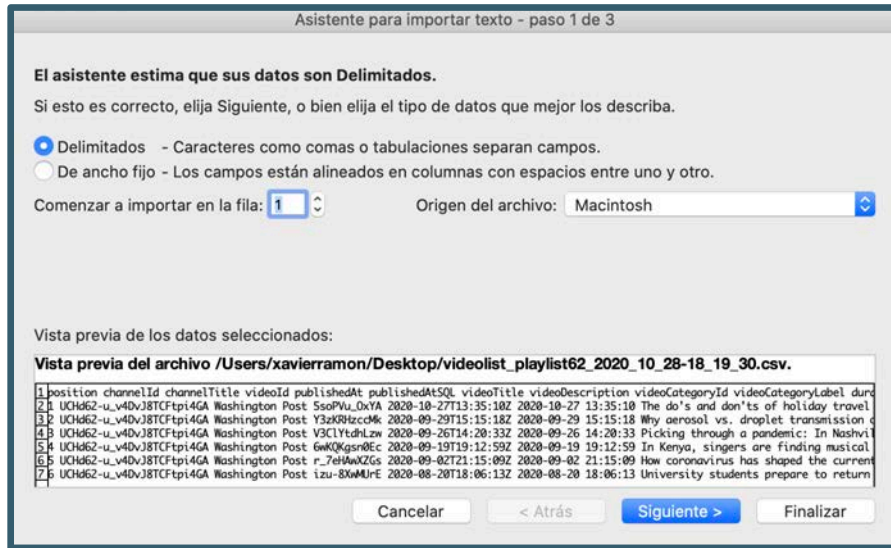
Description of the project

- By using YouTube DataTools, created by the Digital Methods Initiative from the University of Amsterdam (<https://tools.digitalmethods.net/netvizz/youtube/>) (Rieder, 2015), students **download data from channels and video lists** to scrutinize how particular media organizations use YouTube to disseminate their output.

The screenshot shows the "YouTube Data Tools" website. At the top, there's a dark header with the title "YouTube Data Tools" and navigation links: "blog", "software", "research", "DMI", and "about". Below the header is a secondary navigation bar with links: "Home", "Channel Info", "Channel Network", "Video List", "Video Network", "Video Info", and "FAQ". The main content area explains that the tools extract data from YouTube via the YouTube API v3 and provides links to a "blog post" and an "introductory video". A "Modules" section follows, listing four tools: "Channel Info", "Channel Network", "Video List", and "Video Network". Each tool has a brief description and a "launch" button. The "Channel Info" module description states it retrieves information for a channel from a specified channel id. The "Channel Network" module description states it crawls a network of channels connected via "featured channels" and subscriptions. The "Video List" module description states it creates a list of video infos and statistics from one of four sources: channel uploads, playlists, search queries, or video ids.

The screenshot shows a YouTube channel page for "Nacional" by "AGENCIA EFE". The page features a header with the YouTube logo, a search bar, and navigation icons. The main content area displays a video player for a video titled "Telefónica" with a duration of 1:29. Below the player, the channel name "Nacional" is shown, along with statistics: "4804 videos · 9442 visualizaciones · Actualizado ayer". A "SUSCRIBIRME" (Subscribe) button is visible. On the right side, there is a list of six recommended videos, each with a thumbnail, title, and duration. The titles include "Telefónica ganó 886 millones hasta marzo, un 118 % más que en 2020", "Pleno del Congreso", "ERC y JxCAT, enrocados; Comunes entran en juego", "La bolsa española se deja más del 1 % tras la apertura y arriesga los 9.000", "El Gobierno aprueba hoy la ley de 'riders'", and "El inminente acuerdo en pensiones se limitará a su revalorización con el IPC".

A3.YouTube analysis



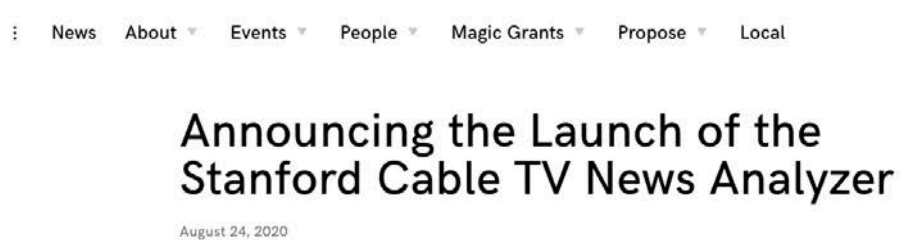
Description of the project

- **Modules employed:**
 - “Channel Info”
 - “Video list”
 - “Video Info and Comments” (Basic info, comments and authors).
- Students learn for the first time how to import **CSV files** into Excel.
- After working with data, they once again have to pick some of the obtained results to **produce a visualization**.

A4. Television news analysis

Description of the project

- Through Stanford Cable TV News Analyzer (<https://tvnews.stanford.edu>), an interactive tool developed by the Brown Institute for Media Innovation, students explore content from a database that incorporates **material from CNN, Fox News and NBC recordings**.



THE BROWN INSTITUTE
FOR MEDIA INNOVATION

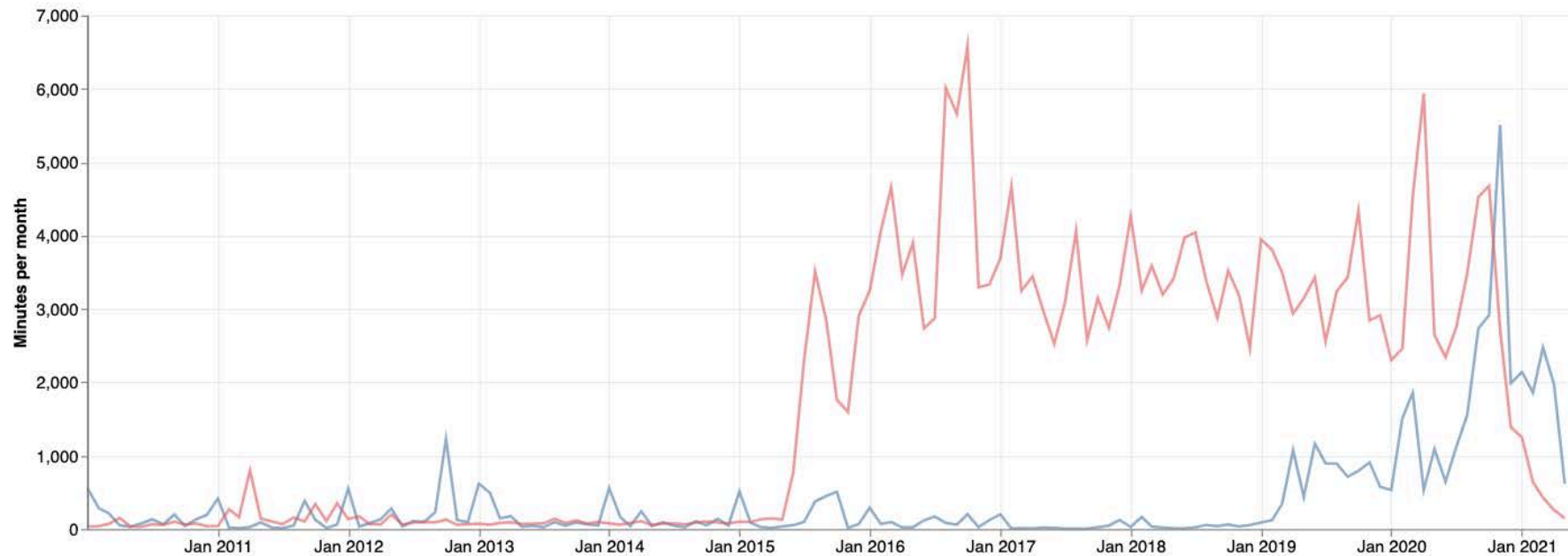


Google Cloud





Use this tool to count the screen time of **who** and **what** is in cable TV news. ([help](#), [more examples](#))



[Copy url](#), [embed](#) chart, [save](#) image, or [download](#) the data.

By introducing certain parameters into the search engine, the tool **automatically generates visualizations** such as line charts, which can be saved or embedded.

Libro1					
Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista					
<div> <div> <div>Cortar</div> <div>Copiar</div> <div>Pegar</div> </div> <div> <div>Calibri (Cuer...)</div> <div>12</div> <div>A A</div> </div> <div> <div> <div></div> <div></div> <div></div> </div> <div> <div></div> <div></div> <div></div> </div> </div> <div> <div>Ajustar texto</div> <div>Combinar y centrar</div> </div> <div> <div>General</div> <div>%</div> <div>000</div> <div>0.00</div> <div>0.00</div> </div> <div> <div>Formato condicional</div> <div>Dar formato como tabla</div> </div> <div> <div>Normal</div> <div>Bueno</div> <div>Incorrecto</div> <div>Neutral</div> <div>Cálculo</div> </div> <div> <div>Celda de com...</div> <div>Celda vinculada</div> <div>Entrada</div> <div>Notas</div> <div>Salida</div> </div> <div> <div>Insertar</div> <div>Eliminar</div> <div>Formato</div> </div> <div> <div>Autosuma</div> <div>Rellenar</div> <div>Borrar</div> </div> <div> <div>Ordenar y filtrar</div> </div> </div>					
A2	fx name="joe Biden"				
	A	B	C	D	E
1	Query	Time	Value	Unit	
2	name="joe Biden"	1/1/10	33.216.823	seconds	
3	name="joe Biden"	1/2/10	17.148.546.999.999.900	seconds	
4	name="joe Biden"	1/3/10	12.933.139.000.000.000	seconds	
5	name="joe Biden"	1/4/10	28.778.939.999.999.900	seconds	
6	name="joe Biden"	1/5/10	19.581.960.000.000.000	seconds	
7	name="joe Biden"	1/6/10	4.550.574.999.999.990	seconds	
8	name="joe Biden"	1/7/10	7.831.436.000.000.000	seconds	
9	name="joe Biden"	1/8/10	3.740.222.000.000.000	seconds	
10	name="joe Biden"	1/9/10	11.884.775.000.000.000	seconds	
11	name="joe Biden"	1/10/10	23.809.500.000.000.000	seconds	
12	name="joe Biden"	1/11/10	7.957.403.999.999.990	seconds	
13	name="joe Biden"	1/12/10	11.498.256.000.000.000	seconds	
14	name="joe Biden"	1/1/11	24.901.630.000.000.000	seconds	

However, students must **download raw data and work with it** on Excel. Later, they produce the visualizations of their choice, whether it is with Flourish or with another tool.

Discussion

- In each of these four projects, students learn the **basic functionalities of Microsoft Excel**: entering data, opening CSV files, applying formulas in spreadsheets, sorting and filtering data and calculating percentages and averages.
 - Those operations provide students with numerical skills that are “no longer optional” (Lewis, McAdams & Stalph, 2020, p. 19).





Data journalism is the new punk

Simon Rogers

Discussion

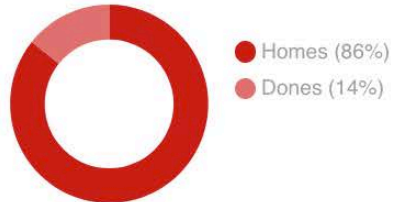
- By embracing a **DIY approach** (Rogers, 2014), students **experiment with visualization tools** such as Flourish, Google Data GIF Maker and Datawrapper.
 - These tools are of heightened importance since they “have the potential to empower journalists, designers and news developers alike while democratising the immense power of data visualisation” (De Lima-Santos, Schapals & Bruns, 2020, p. 6-7).
- Throughout the process, we stimulate students to **think how they can apply the skills** they acquire towards their own data journalism practice.
 - They soon start to apply what they have learned to produce visualizations for other courses and for *Diari de Barcelona*, an online newspaper developed and hosted by the UPF (<https://www.diaridebarcelona.cat>).

Diari de  Barcelona

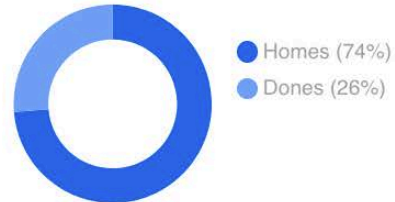
Presència de dones en els espais del F.C. Barcelona

MARÇ 2021

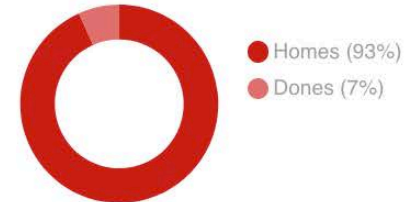
A l'anterior directiva de Josep Maria Bartomeu



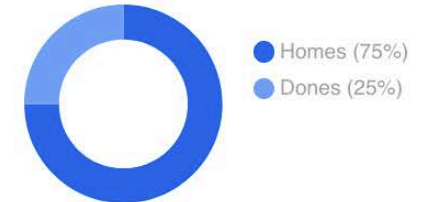
A la candidatura de Víctor Font



A la candidatura de Joan Laporta

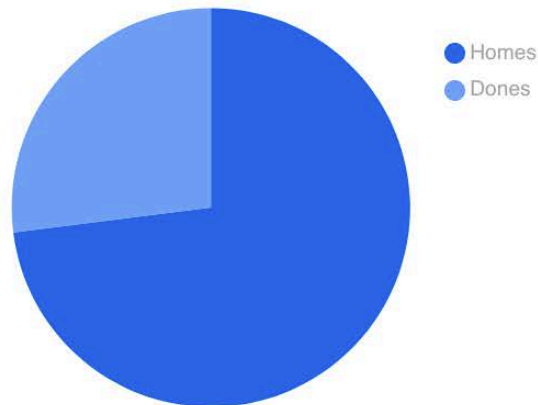


A la candidatura de Toni Freixa



Actualment les sòcies representen **gairebé el 27% de la massa social barcelonista**. Tot i que el percentatge ha millorat respecte a molts anys enrere, **més del 40% dels socis són homes** i tenen aproximadament **més de 65 anys**.

Nombre de socis per sexe l'any 2019



Nombre de socis l'any 2019



Percentage of irregular migrants in the Mediterranean by country

Spain Greece Italy Others

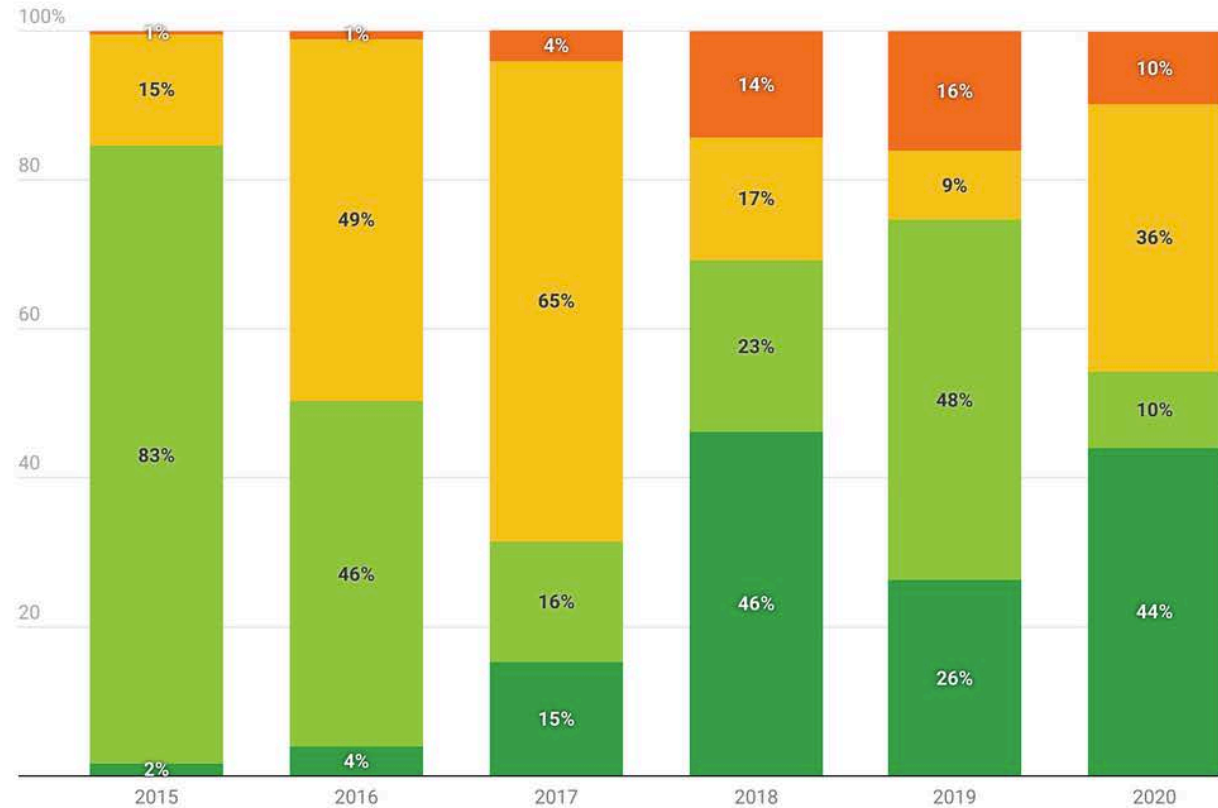


Chart: Àlex Gutierrez - Berta Codina- Berta Posas -Pau Torres • Source: United Nations High Commissioner for Refugees • Created with Datawrapper

March 2019

February 2019

January 2019

January 2018

Categories

Blogpost

Fact-check

False

Media analysis

Mostly false

Mostly true

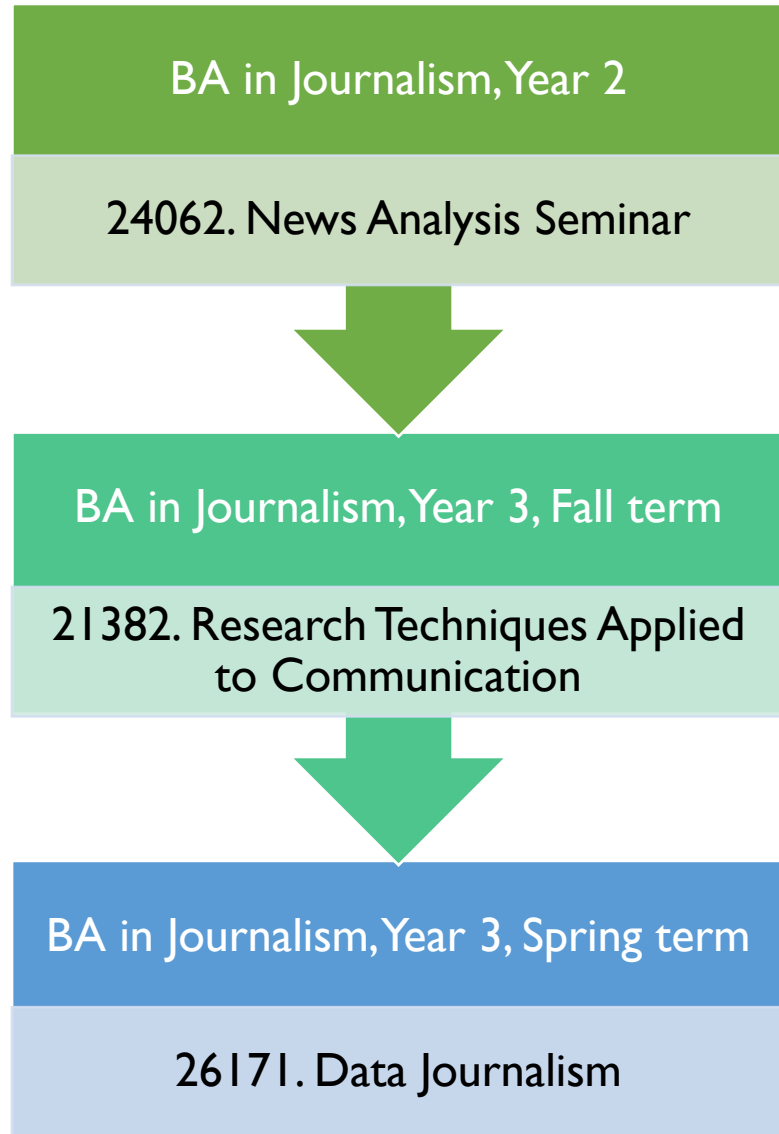
True

Uncheckable

<https://eufactcheck.eu/factcheck/mostly-false-spanish-government-pull-effect-makes-spain-the-main-european-entrance-of-illegal-immigrants/>



Concluding remarks



- The activities developed in the course ‘Research Techniques Applied to Communication’ play a **bridging role** between basic courses on news analysis and more advanced courses, such as the ‘Data Journalism’ elective course.
- This case study illuminates that **introductory workshops** within research methods courses can be a **useful gateway** to train students with no previous experience in data analysis and visualization.
- Such initiatives can help to **challenge students’ preconceptions while motivating them** to enrol in advanced courses in the field.



Thank you!

Dr Xavier Ramon

xavier.ramon@upf.edu // @xramonv



Universitat
Pompeu Fabra
Barcelona

References

- Bradshaw, P. (2018). Data Journalism Teaching, Fast and Slow. *Asia Pacific Media Educator*, 28(1), 55–66. <https://doi.org/10.1177/1326365X18769395>
- De-Lima-Santos, M.-F., Schapals, A. K., & Bruns, A. (2020). Out-of-the-box versus in-house tools: how are they affecting data journalism in Australia? *Media International Australia*. <https://doi.org/10.1177/1329878X20961569>
- Lewis, N. P. (2021). Defining and Teaching Data Journalism: A Typology. *Journalism & Mass Communication Educator*, 76(1), 78–90. <https://doi.org/10.1177/1077695820924309>
- Lewis, N. P., McAdams, M., & Stalph, F. (2020). Data Journalism. *Journalism & Mass Communication Educator*, 75(1), 16–21. <https://doi.org/10.1177/1077695820904971>
- Rogers, S. (2014). Data journalism is the new punk. *British Journalism Review*, 25(2), 31–34. <https://doi.org/10.1177/0956474814538181>

References

- Rogers, S. (2018). Flourish helps journalists create easy data visualizations. Retrieved from: <https://blog.google/outreach-initiatives/google-news-initiative/flourish-helps-journalists-create-easy-data-visualizations/>
- Rogers, S. (2019). Using Data GIF Maker to compare data and tell stories. Retrieved from: <https://blog.google/outreach-initiatives/google-news-initiative/using-data-gif-maker-compare-data-and-tell-stories/>
- Stanford News Service (2020). Stanford launches AI-powered TV news analyser. Retrieved from: <https://news.stanford.edu/2020/08/27/stanford-launches-ai-powered-tv-news-analyzer/>
- Weiss, A. S., & Retis, J. (2018). 'I Don't Like Maths, That's Why I am in Journalism': Journalism Student Perceptions and Myths about Data Journalism. *Asia Pacific Media Educator*, 28(1), 1–15. <https://doi.org/10.1177/1326365X18780418>