

---

# Hello!

*I am Xavi Giménez*

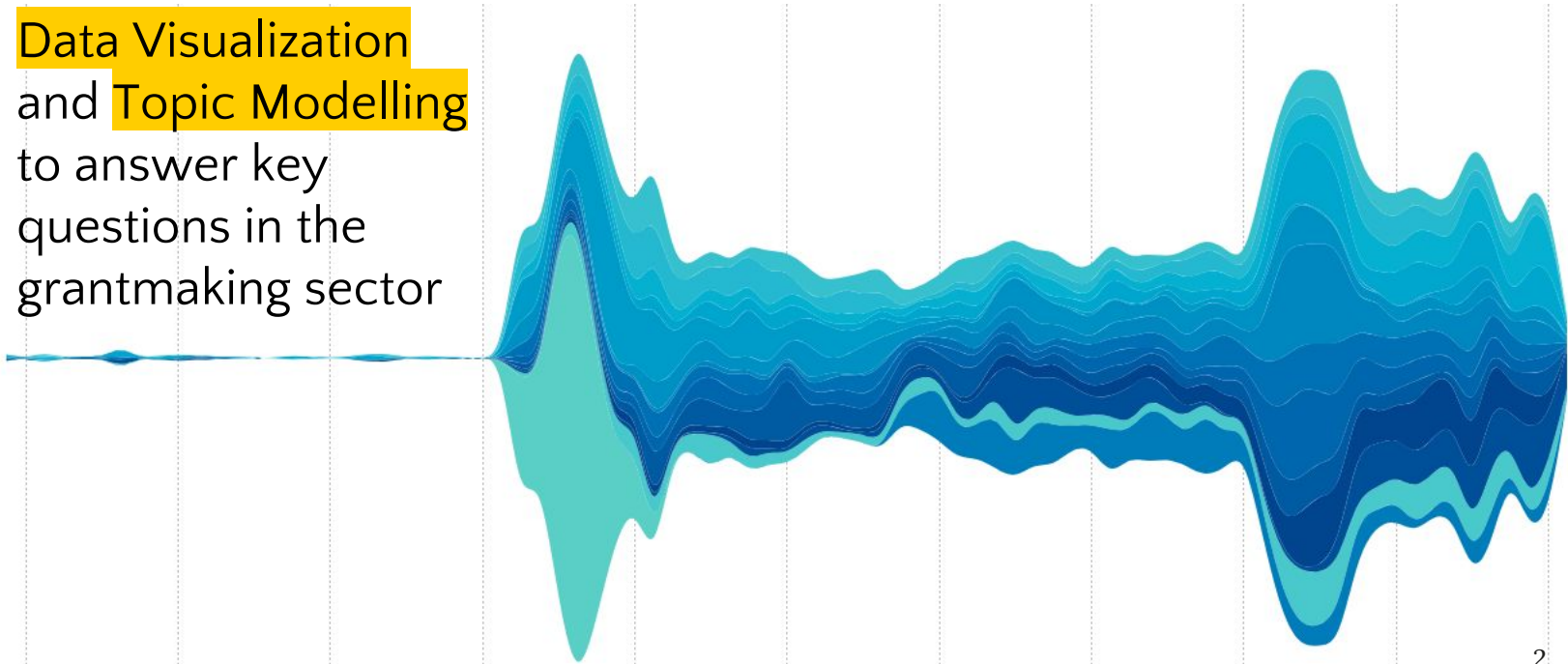
Data Visualization Engineer at SIRIS Academic

@xavigimenez



## Who has funded what themes throughout the years?

Data Visualization and Topic Modelling to answer key questions in the grantmaking sector





## “Digging the data”: Data Visualization Challenge

---

### **Vision**

UK grantmaking to become more evidence-based, impactful and strategic.

### **Mission**

support UK grantmakers to publish information on who, where and what they fund in an open, standardised format to build a better picture of the funding landscape and boost its impact



## “Digging the data”: Data Visualization Challenge

### An Standard

360Giving standard:

Open Data Driven, easy to use and comprehensive.

### The challenge

More than 30 entries: from dashboards and analytical tools to creative, physical representations of the data

### Key questions

1. Who has funded what themes throughout the years?
2. User-led organisations: Who funds them, in what area, how much funding do they receive



## Who has funded what themes throughout the years?

### Themes?

The 360giving standard gives a lot of info regarding a grant:

- Grant details
- Applied funding / granted funding
- Planned dates
- Details of funder and recipient

But not too much on categorization...



## Who has funded what themes throughout the years?

---

### Themes?

That question is open in the sense of what a theme is, since there isn't a standard categorization in the grantmaking sector

**What if we can extract them from the data itself?**

### Topic Modelling

aims to automatically discover the hidden thematic structure in a large corpus of text documents



# Topic Modelling

## Topic 1

Basketball, Lebron, NBA,...

## Topic 2

NFL, Football, American,...

## Topic 3

Trump, president, Clinton,...

LeBron James says President Trump 'trying to Divide through sport'

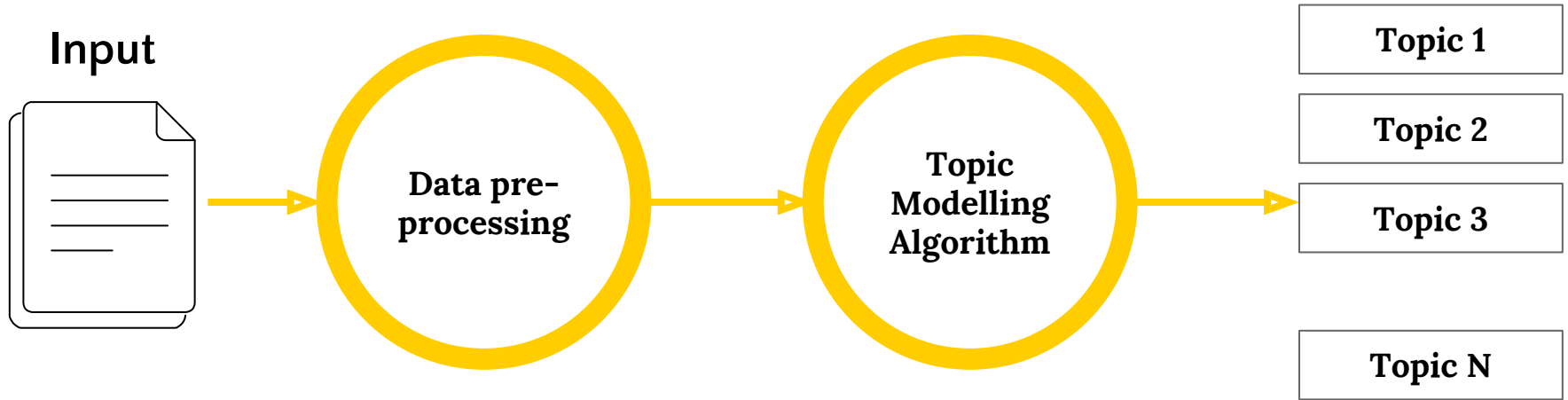
Basketball star LeBron James has praised the American football players who have protested against Donald Trump, and accused the US president of "using sports to try and divide us".

Trump said that NFL players who fail to stand during the national anthem should be sacked or suspended.

James praised the players' unity, and said: "The people run this country." James, who plays for the Cleveland Cavaliers and has won three NBA championships, campaigned for Hillary Clinton, Trump's rival, during the 2016 presidential election campaign.



# Topic Modelling







## Topic Modelling: Input

### Input



**290K**  
**grants**

Title	Description
Homeless Advice and Resilience Project (H.A.R.P)	A project to support people in hardship including those with chaotic lifestyles and multiple needs such as poor health, homelessness, substance misuse, offending, victimisation, poor social relationships, financial difficulties, lack of employability skills, exclusion or lack of engagement with services. it will provide access to hot meals, foodbanks, fuel, clothing, washing and laundry facilities, furniture and financial support for those moving from benefits into work.
Coton Community Great British Family Festival Fun Weekend	Coton Community Great British Family Festival Fun Weekend
A new van for Scrapstore	A new van for Scrapstore



## Topic Modelling: data pre-processing

---

**Data pre-processing**

- Minimum term length
- Case conversion (lowercase)
- Stop-word filtering
- Minimum frequency filtering
- Maximum frequency filtering
- Steeming (compute, computing, computer = comput)



## Topic Modelling: Algorithms

**Topic  
Modelling  
Algorithm**

### **Probabilistic approaches**

Words and documents get probability scores for each topic.

E.g. **Latent Dirichlet Allocation (LDA)**

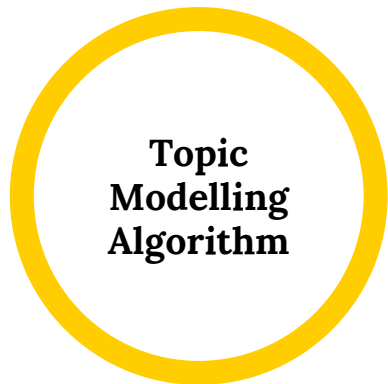
### **Matrix Factorisation approaches**

Apply methods from linear algebra to decompose a single matrix (e.g. document-term matrix) into a set of smaller matrices.

E.g. **Non-negative Matrix Factorisation (NMF)**



# Topic Modelling: Algorithms



Document-term matrix

	w1	w2	w2	w4	w5	w6	w7	w8	w9	w10
Grant 1	10	20	7	18	7	0	3	77	3	0
Grant 2	30	15	10	20	11	0	19	8	11	38
Grant 3	5	24	16	6	9	2	0	12	70	32
Grant 4	38	50	0	18	0	7	0	31	35	14
Grant 5	13	12	2	21	4	6	0	18	14	0
Grant 6	7	3	0	5	1	10	19	22	3	9



# Topic Modelling: Algorithms

**Topic  
Modelling  
Algorithm**

	Topic 1	Topic 2	Topic 3
Grant 1	0	1	1
Grant 2	0.3	0	1
Grant 3	0.5	1	0
Grant 4	0.6	0.5	0
Grant 5	0	0.5	0
Grant 6	0	0.5	0.3

	Topic 1	Topic 2	Topic 3
research	0	0	1
school	0	0.1	0.5
education	0	0	1
disease	0	0.6	0
patient	0	0.7	0
health	0	1	0
budget	0.3	0.1	0.2
finance	0.6	0	0
banking	0.7	0	0
funding	0.3	0	0



## Topic Modelling: Output

spouse carer september  
second-world-war  
france return widow heroes october  
year travel awarded  
veteran visit italy costs  
commemorative

organisation opportunities  
people skills  
youth  
aged  
learning  
older disabled  
deliver education  
disabilities help  
disadvantaged  
confidence self  
behaviour life workshops

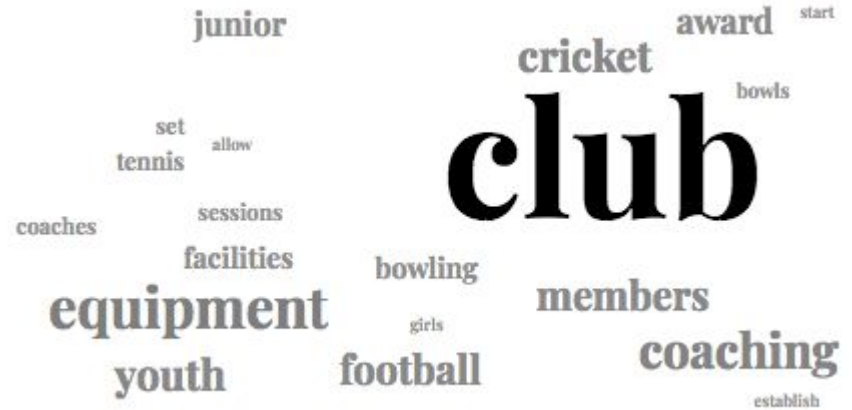


## Topic Modelling: Output





## Topic Modelling: Output





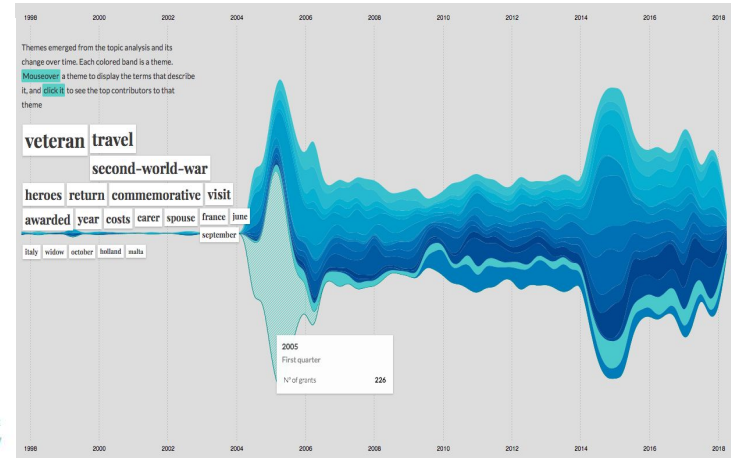
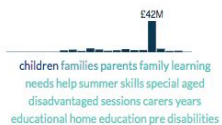
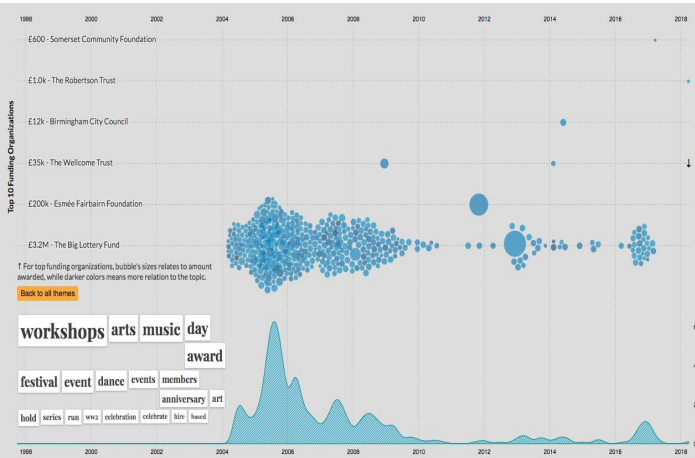


## Topic Modelling: Output





# Data Visualization



<http://www.xavigimenez.net/360giving-grantmaking-themes>



# Thanks!

*Any* **questions** ?

You can find me at

- @xavigimenez
- xavi@xavigimenez.net