

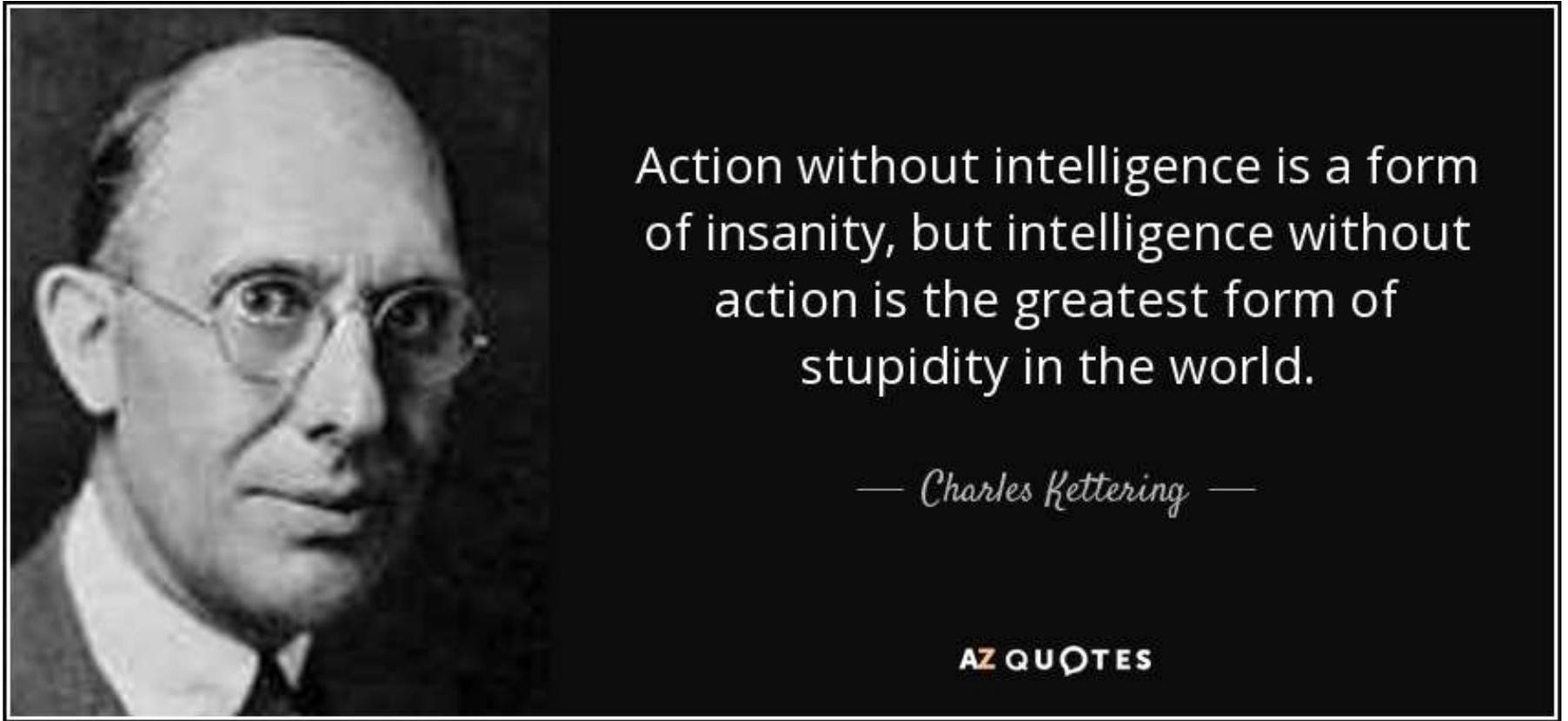


Public Health  
England

Protecting and improving the nation's health

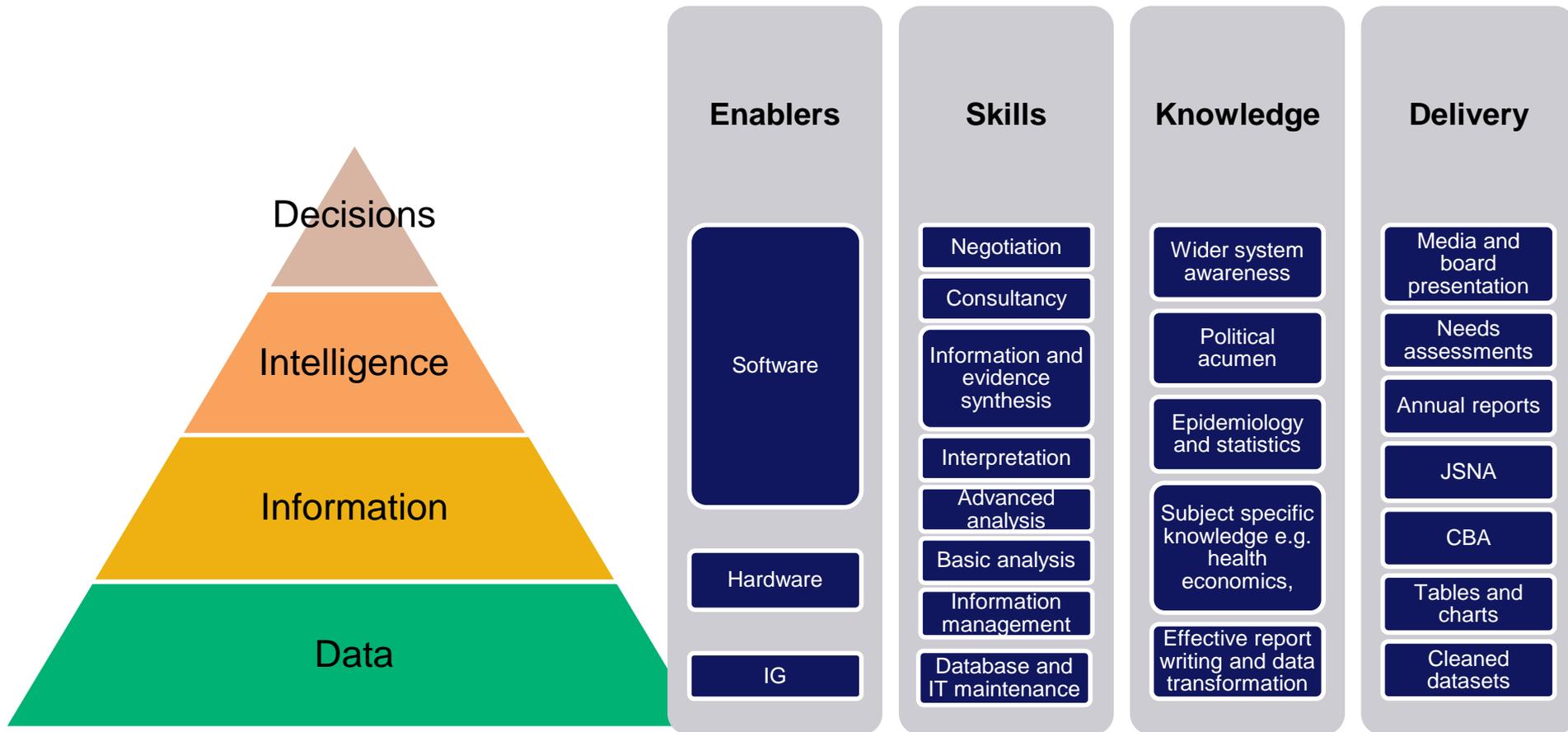
# Intelligent design: creating a blueprint for future intelligence provision

Matt Hennessey, Associate Director



Our challenge is to build a unified intelligence function that is neither insane nor stupid! All roads lead to a decision or behaviour change

# Data driven decisions: core components



# Sweating the assets: Analytical resource



Are we at a crossroads?

- Cannot afford to be all things to all people
- Maximising transferable **skills** in the system
- Maximising transferable **capacity** in the system
- Setting clear career paths; framework for development
- Understanding where and how to deploy specialist knowledge

# Laying the foundations: data/information

- Big data; new sources, new formats
- Changes to Architecture and IG - cloud
- Opportunities of new technology for data transformation
- Embracing automation
- Managing the risks of automation – data integrity and quality

The new world of 'Data Science' – Managing your Munging

# Say what?!

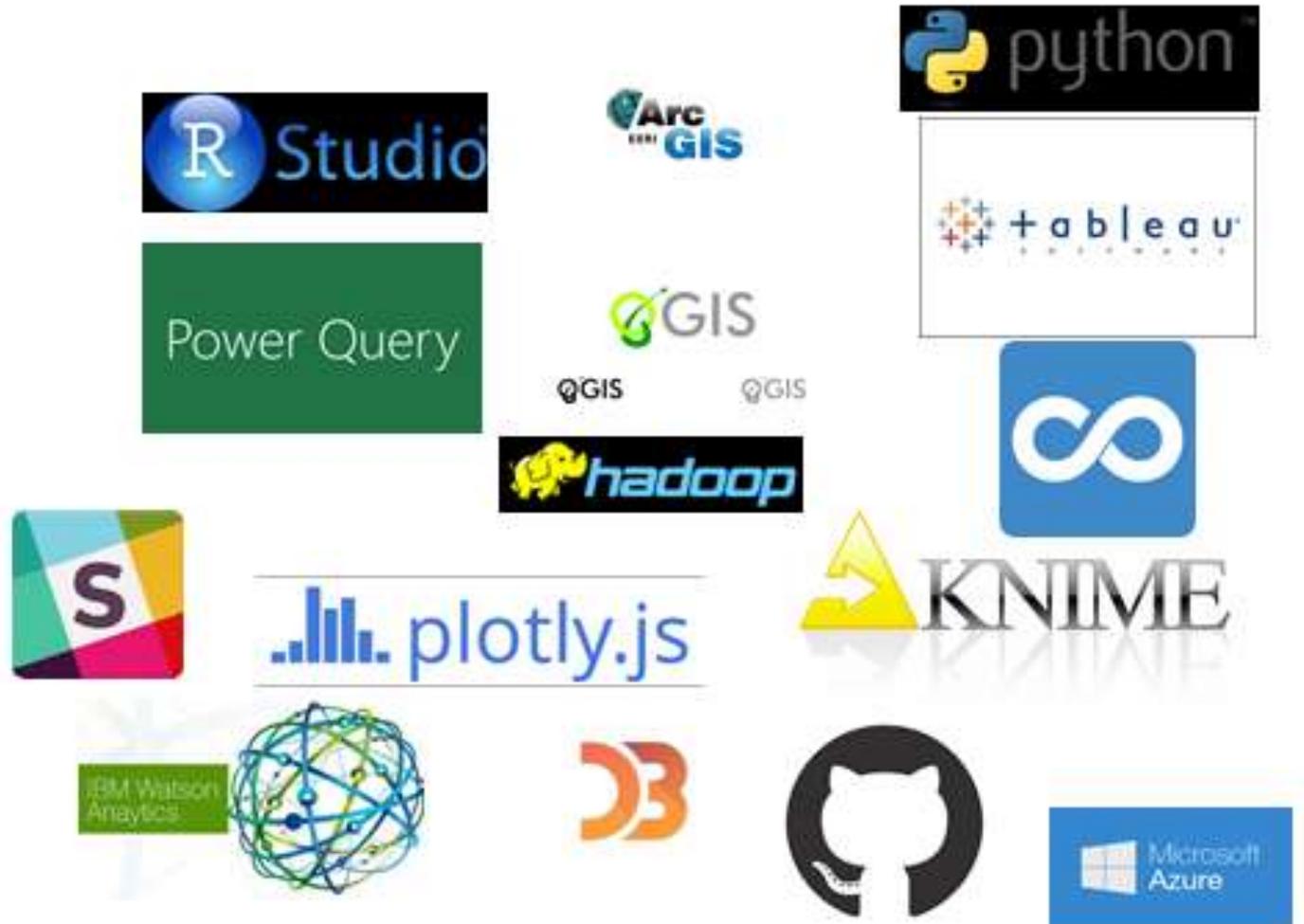
I need an oozie to link my Pig  
and my Hive

How big does my sandbox  
need to be?

Can you data wrangle with a  
Panda?

Fork me on GitHub

# This slide is already out of date!



# Deciphering the jargon!

The screenshot shows a blog post on the GOV.UK website. The page header includes the GOV.UK logo and the Public Health England logo. The main title of the post is 'An A-Z of Public Health Data Science' by Mary E Black, dated 11 November 2015. The post content features a dark blue graphic with various data science and technology terms: Python, Query, Git, Yottabyte, Tiny data, Stack, Wrangling, API, and HADOOP. Below the graphic, the text begins with 'This is a new language that all public health folks will require. There are lots of encyclopedias for Big Data: Dutch company Dataflog do a good one and the Big Data Made Simple portal has a useful one. What the generic lists do not do is...'

**GOV.UK**

Blog

**Public health matters**

Public Health England

Organisations: Public Health England

**An A-Z of Public Health Data Science**

Mary E Black, 11 November 2015 — Data blog, Public health data

Python Query Git Yottabyte Tiny data Stack Wrangling API HADOOP

This is a new language that all public health folks will require. There are lots of encyclopedias for Big Data: Dutch company Dataflog do a good one and the Big Data Made Simple portal has a useful one. What the generic lists do not do is...

**Public health matters**

The official blog of Public Health England, providing expert insight on the organisation's work and all aspects of public health. [More about this blog.](#)

**Categories**

Select Category

**Sign up for updates**

Email Atom

**Follow PHE**

Twitter Facebook YouTube

# Automating data analysis – Example HAY

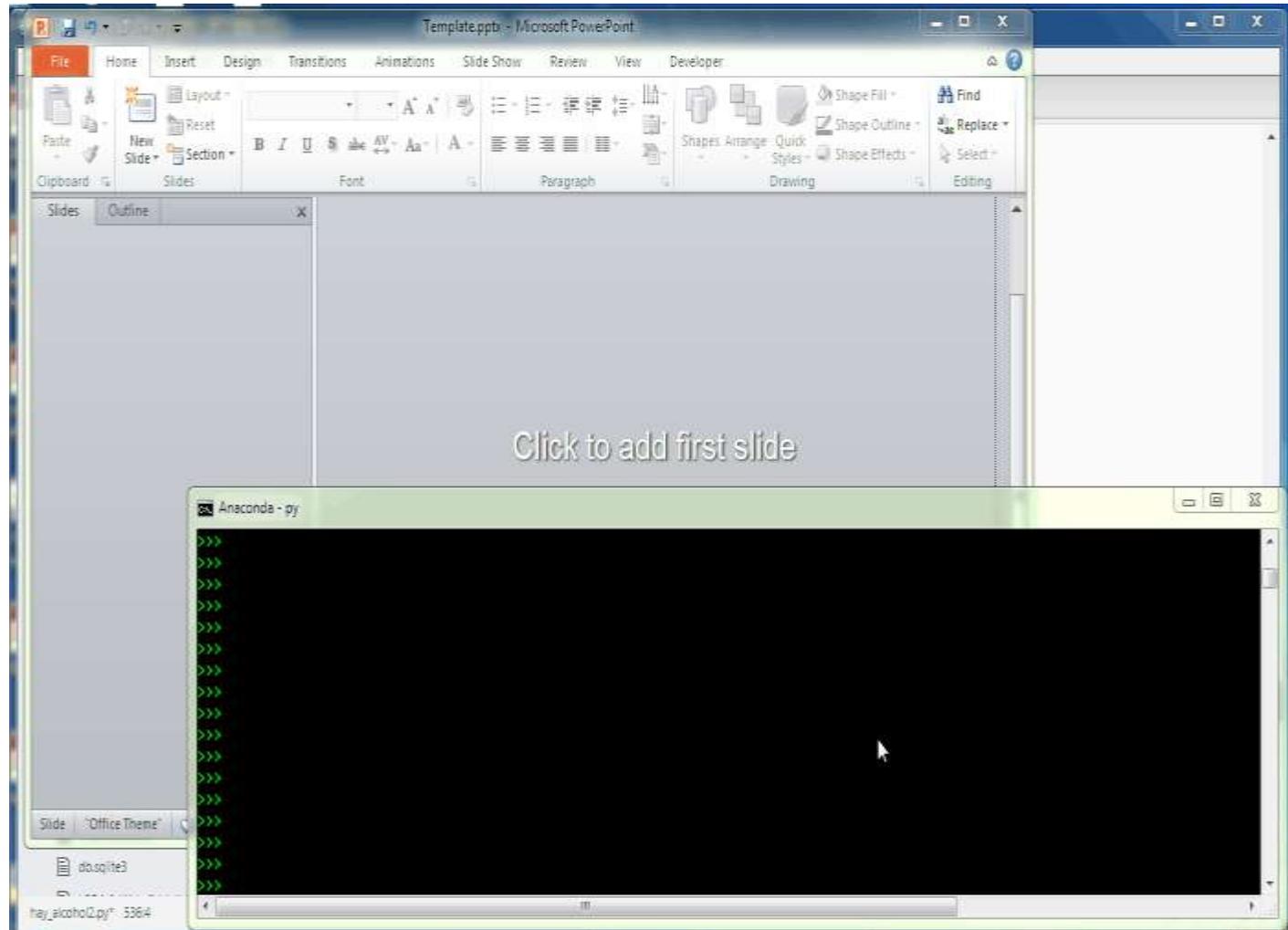
Approx 1.8 million rows

Python used to merge daily .csv files

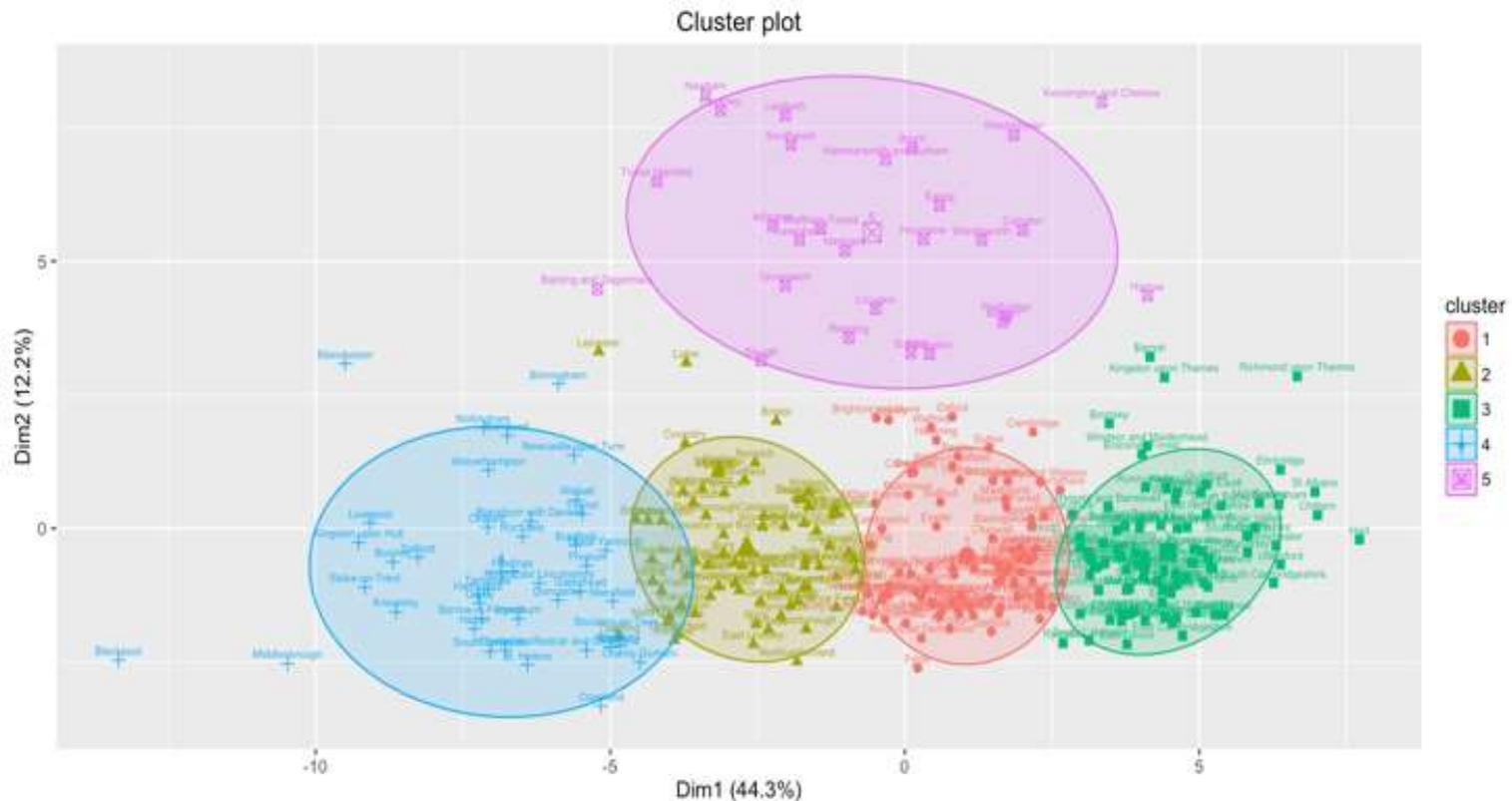
Pandas used to analyse dataframes – lightning quick to organise data in to groups, such as age, alcohol consumption, smoking status

Win32com acts as a bridge between Python and MS Office tools (Excel and Powerpoint)

Result – repeatable, high quality



# New insight? Clustering of local authorities in health profiles – what do they tell us?





Department  
for Transport

# Data Science Public consultations



Tom Ewing



David Pavitt

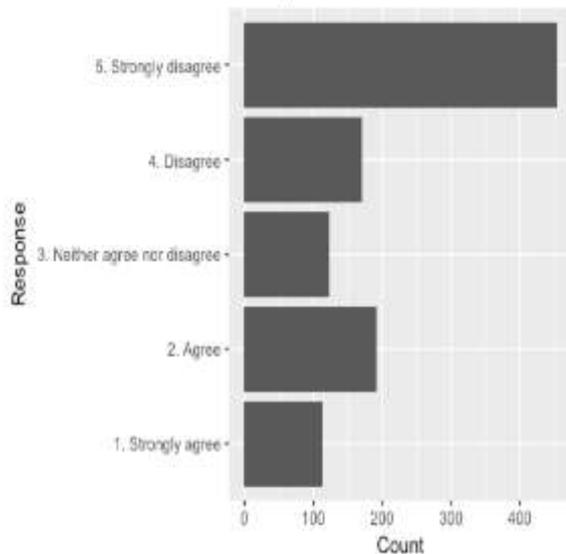
# Basic stats and word counts

Q8. How strongly do you agree or disagree with our proposed environmental objective for the next regime?

## Response strength

The average response to the question is 3.6 where 1 = strongly agree and 5 = strongly disagree.

All responses

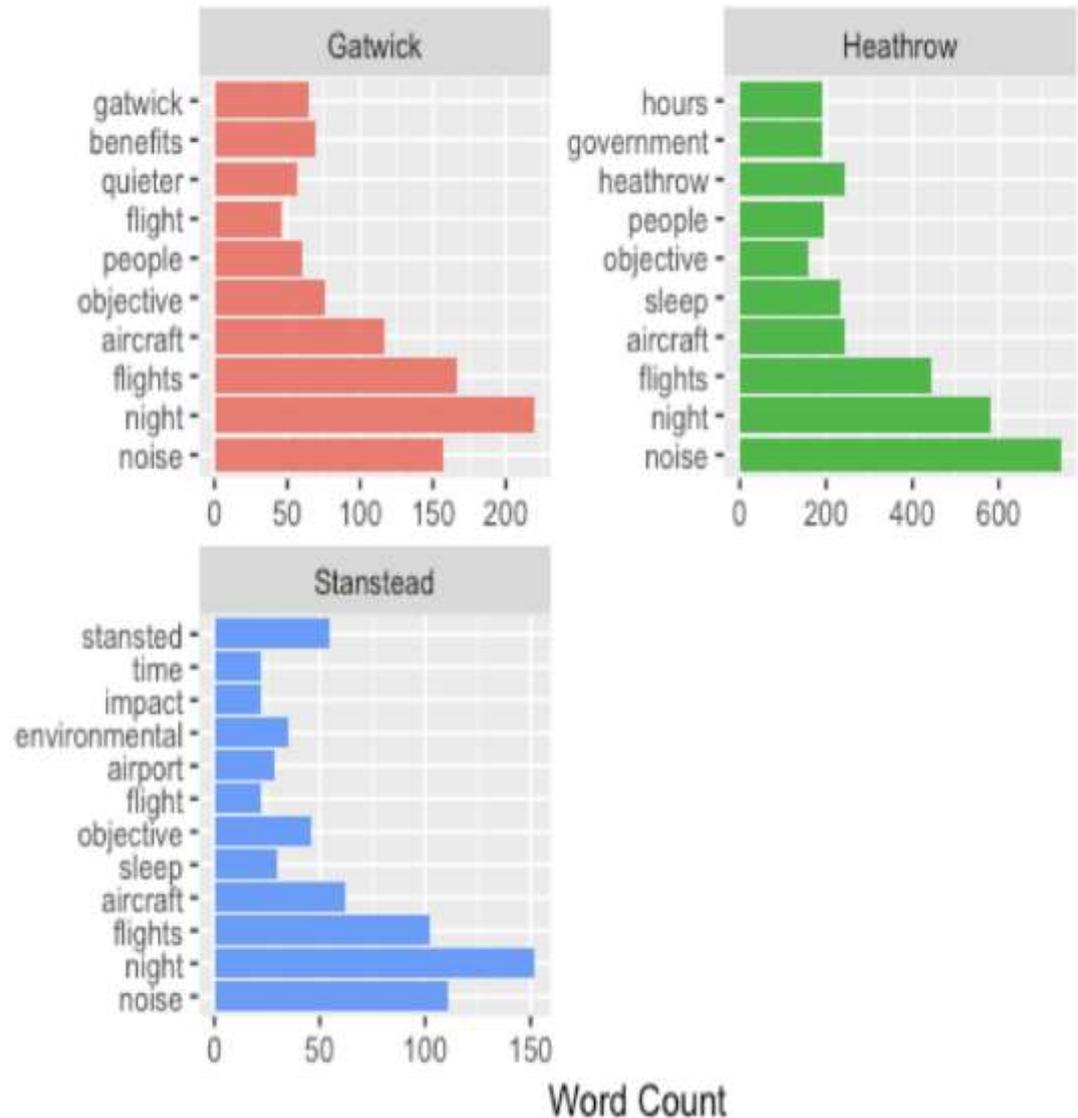


## Breakdown

### Agreement score by organisation type

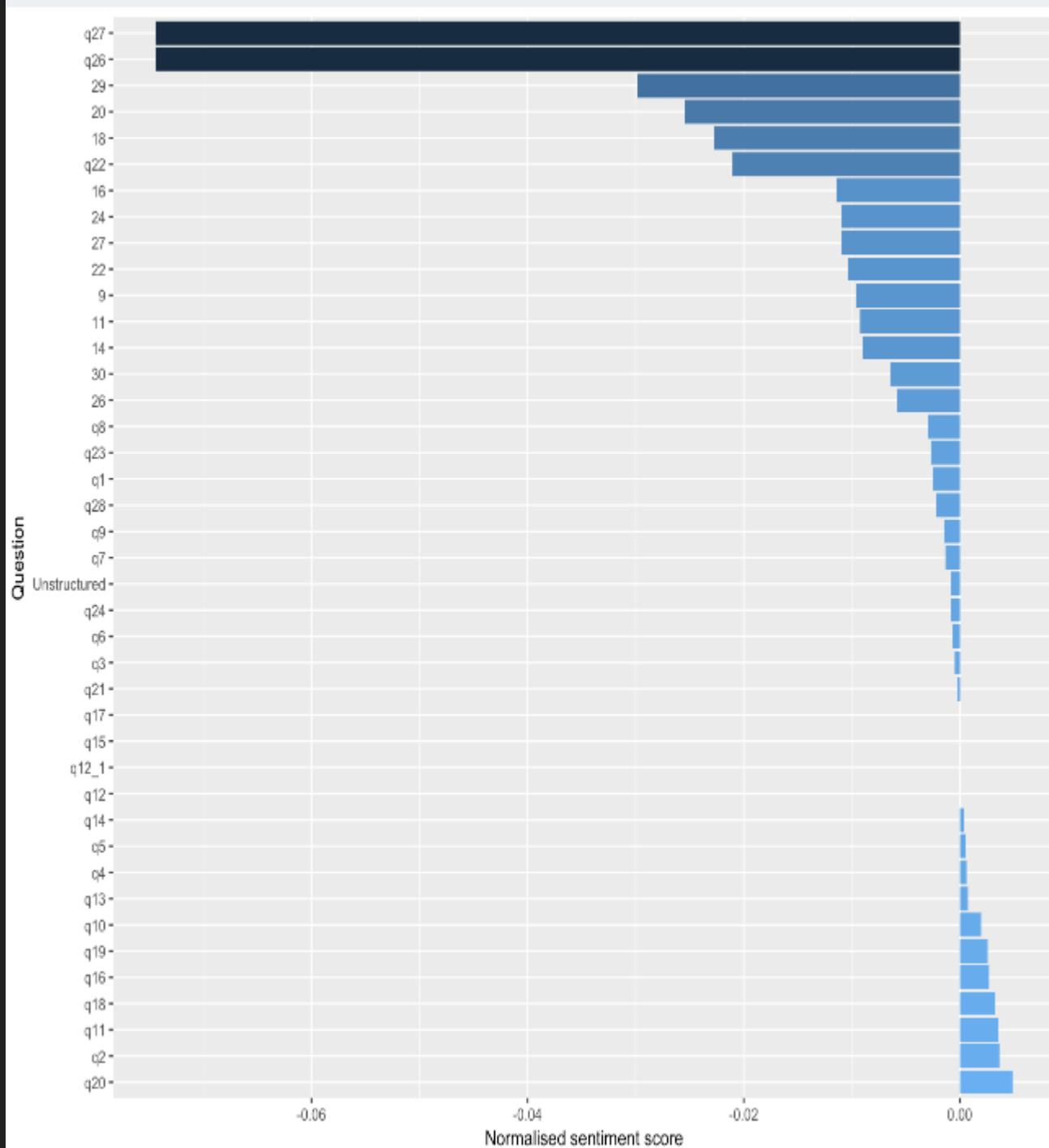
Organisation Type	Average Score
None	3.7
Public Body	3.4
Noise or environmental group	4.1
Airport or Airline	1.9
Business or Trade Union	2.1
Other community group or charity	2.8
Local Authority	3.5

## Chart showing most common used words by airport



# Sentiment

Across the  
different  
consultation  
questions



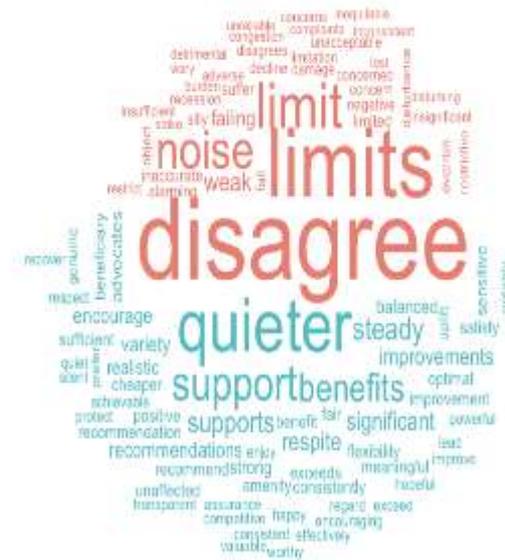
# Sentiment

## Word cloud v2.0

Select Question

q9

negative



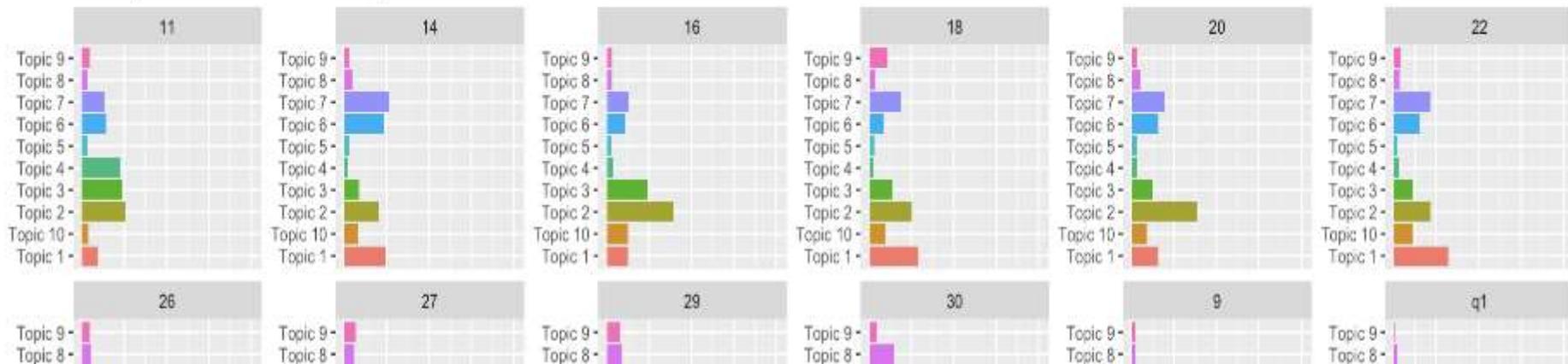
positive



# Topic modelling

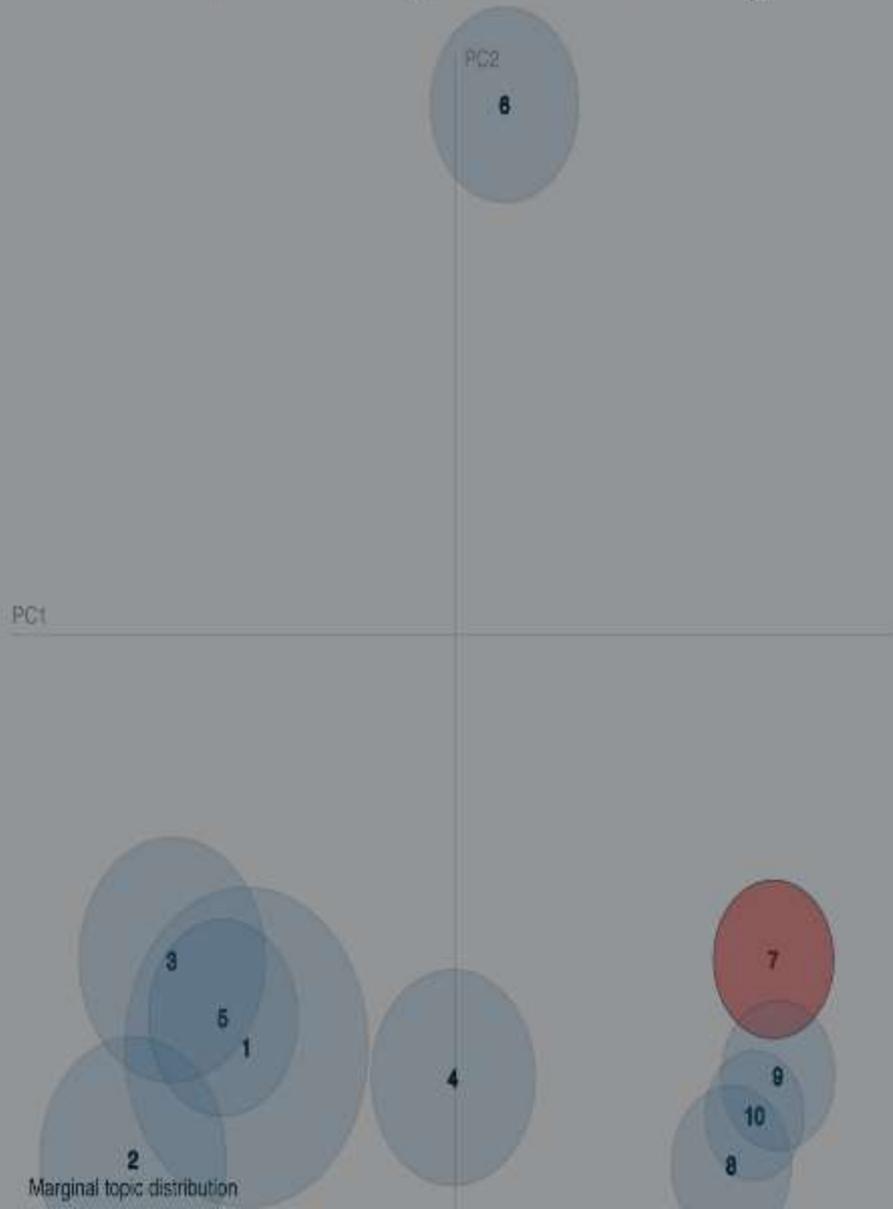
Topic 1	Topic 10	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
planes	agree	flights	flights	regime	email	noise	aircraft	agree	government
west	strongly	sleep	ban	environmental	www	pollution	quota	section	noise
gatwick	proposal	pm	heathrow	objective	visit	db	noise	aircraft	disagree
noise	disagree	hours	time	length	security	government	airlines	strongly	quotas
flying	movement	ban	period	proposed	scanned	sleep	summer	proposals	suggestion
fly	comments	stansted	scheduled	benefits	response	people	reduction	disagree	proposal
communities	aircraft	residents	cost	period	symanteccloud	residents	winter	limits	gatwick
affected	additional	disturbance	economic	opportunity	cloud	levels	movements	additional	people

The topic distribution for each question

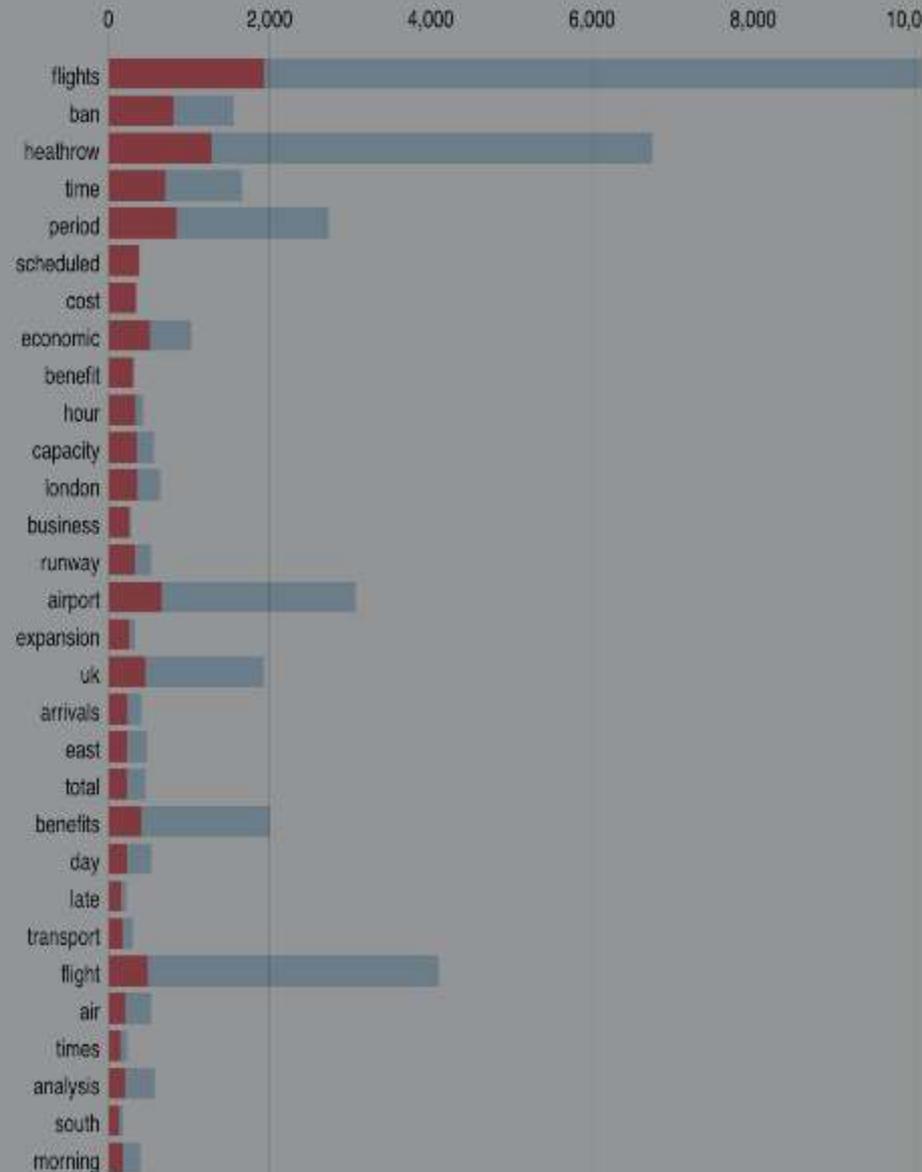


# Topic modelling - clustering and visualisation

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 7 (5.8% of tokens)



# Topic modelling - similarity, machine learning?

**Enter a consultation response and find similar responses**

I think night flights are ...

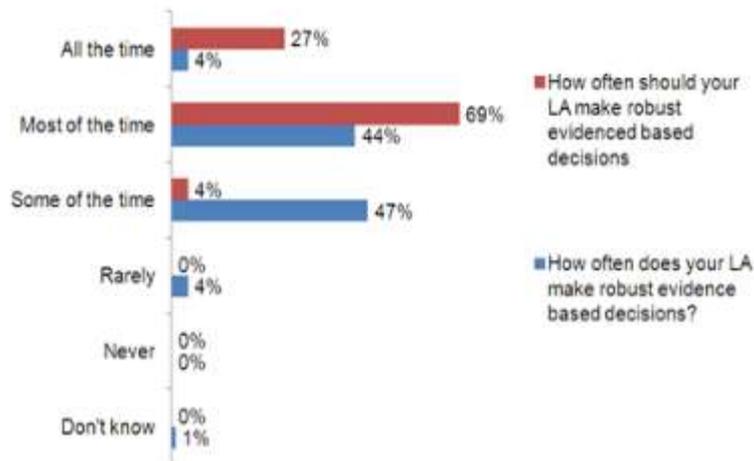
Go!

# Mobilising our knowledge: intelligence to decision

## SOLACE REPORT: The Role of Evidence in Local Government Decision Making

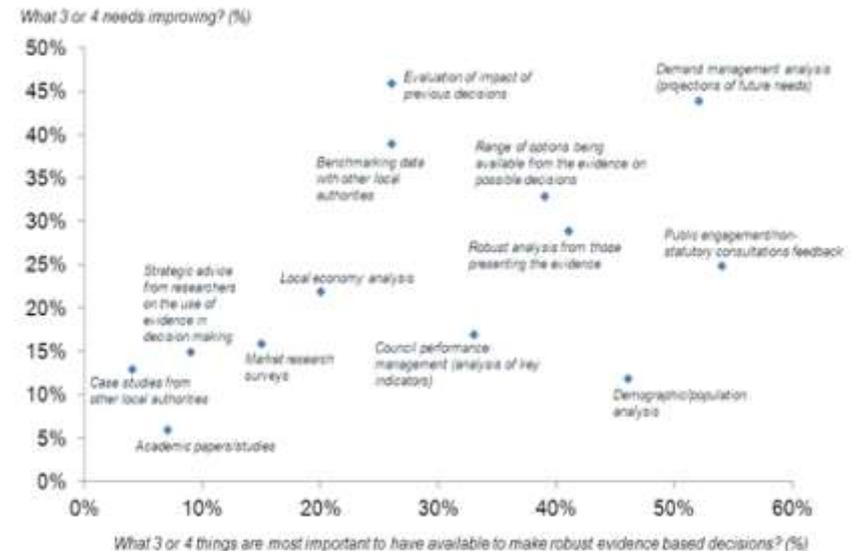
\* Evidence is “data in context” – context being established by a hypothesis or belief

### Making robust evidence based decisions



Source: 37 SOLACE members who responded to the survey online between September and October 2012

### What needs improving?



Source: 37 SOLACE members who responded to the survey online between September and October 2012

# Models for a system or place-based approach?

Population hub:



Created by Eric Arne from Noun Project

Shared investment



Created by Oksana Lutsyshina from Noun Project

Consultancy



Created by Round Pixel from Noun Project

Digital platform



Created by Icons8 from Noun Project

# How do we get people to act on our intel?



Breckon and  
Dodson model

Spend time with your 'audience'