

# The impact of sentiment on school inspection reports in England

 Christian Bokhove, Samuel Sims

# INTRO

- We look at the inspection regime in England.
- The Office for Standards in Education (Ofsted) was created in 1992. Since then there have been multiple Chief Inspectors.
- We wondered whether sentiments of secondary schools' inspection reports changed with these Chief Inspectors.

## METHODS

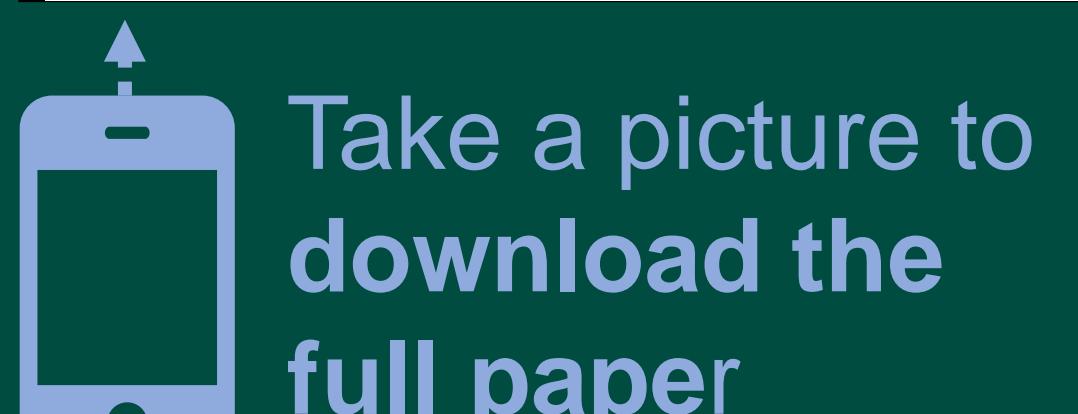
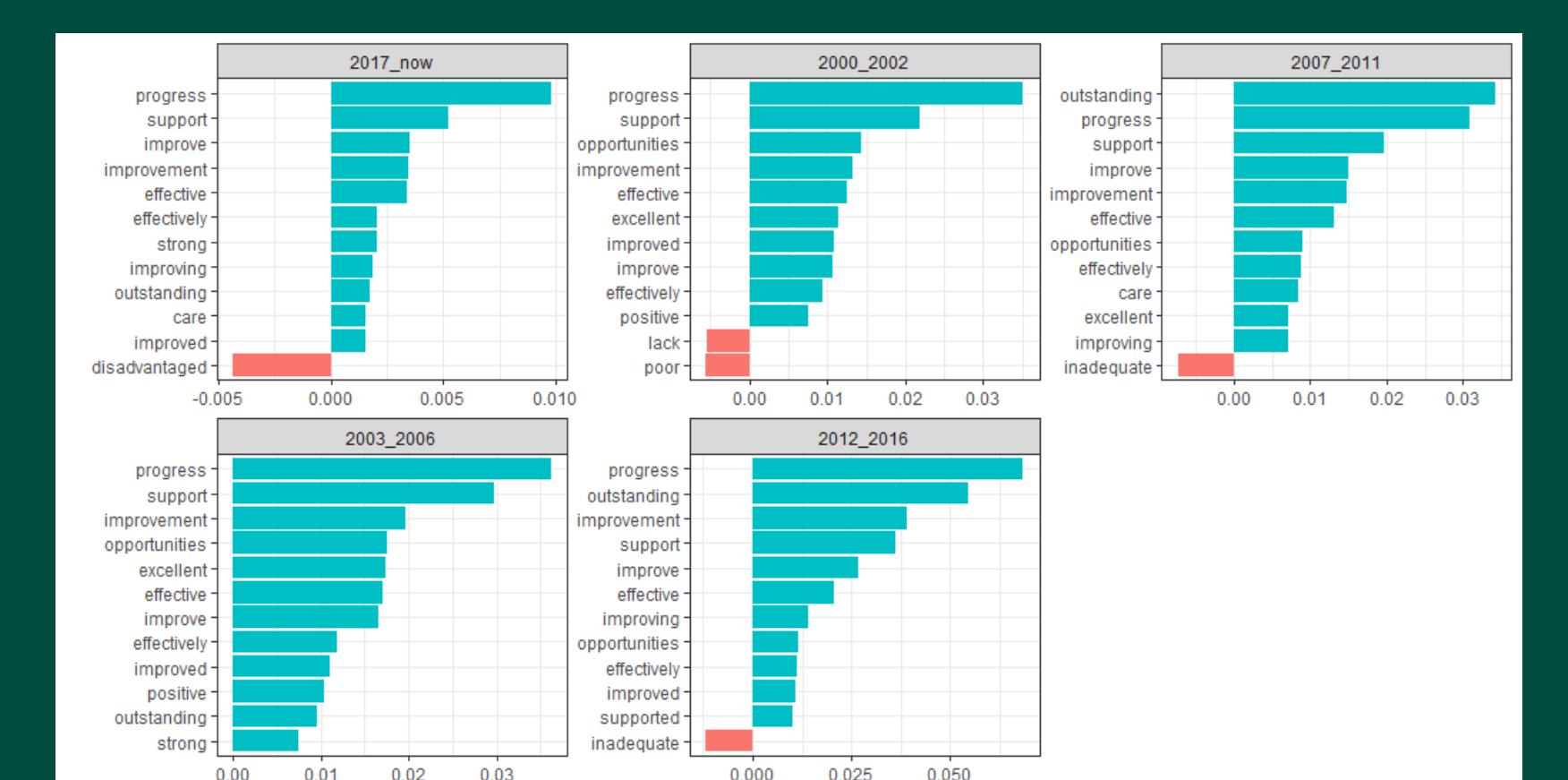
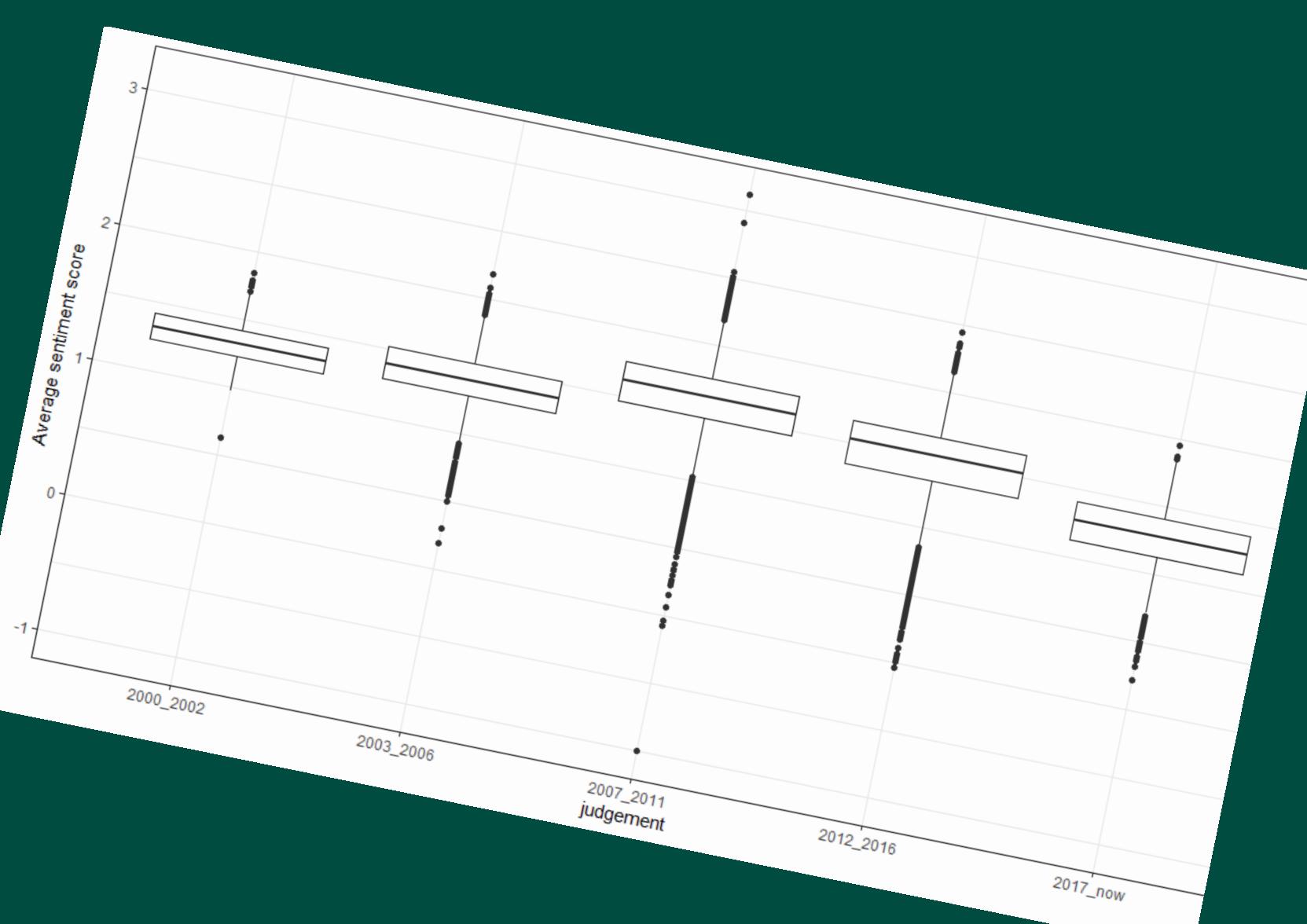
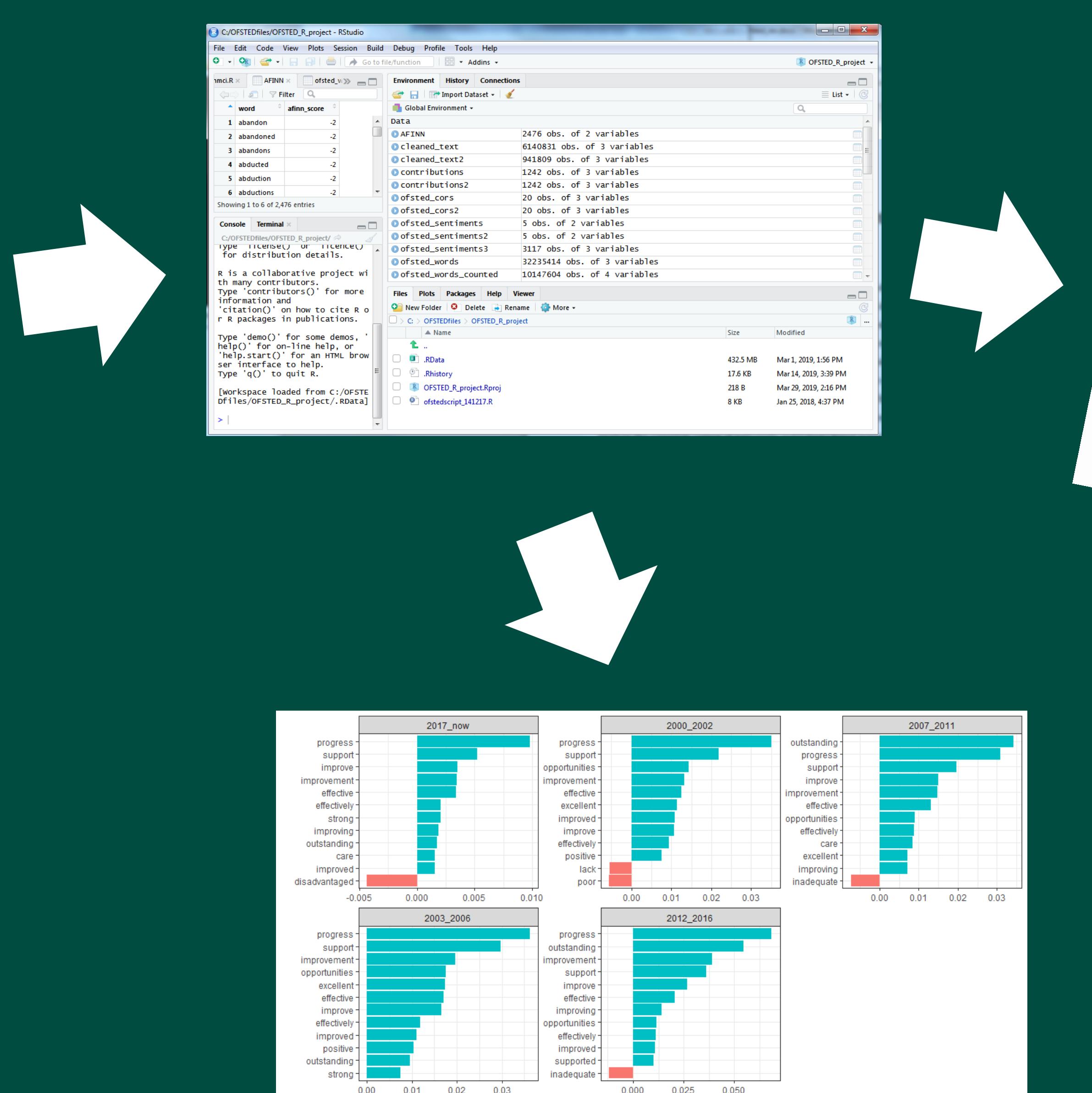
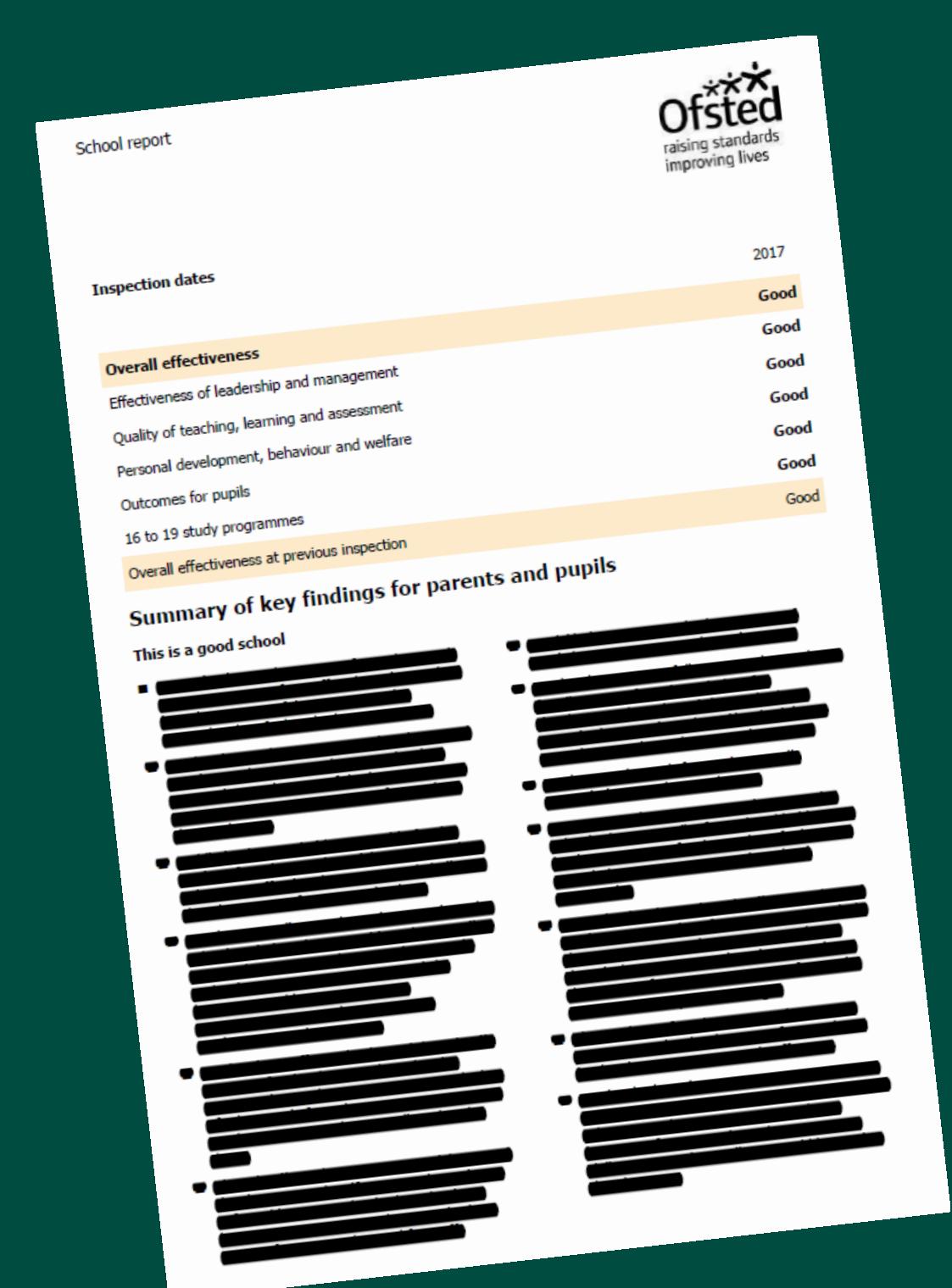
- We ‘scraped’ all inspection documents available from 2000 to end 2017 from the Ofsted website <https://reports.ofsted.gov.uk/>
- This resulted in a corpus of 17,212 documents, 2.49 GB of data.
- All documents were pre-processed to a ‘tidy text format’ (Silge and Robinson, 2017), a format with 32,235,414 separate words. The documents were grouped per Chief Inspector (period of time).
- We then used sentiment analysis, applying a general-purpose AFINN lexicon (Nielsen, 2011) which assigns words with a score that runs between -5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment.
- We used R in Rstudio to do these analyses.

## RESULTS

- There were differences in sentiment over time.
- From 2000 sentiment rose to its highest in 2007-2011, after which sentiment went down again, changing with Chief Inspectors.
- There were commonalities and differences in the sentiments of words that contributed to the total sentiment of Ofsted documents.
- The development of sentiment can cautiously be linked to policy changes in the inspection regime.

## DISCUSSION

- However, interpretation of sentiment is tricky.
- We say ‘cautiously’ because correlation/causation remains a challenge.
- There are semantic challenges, for example double negations.
- We are now looking at differences in sentiments for reports with different judgements: inadequate, requiring improvement (formerly ‘satisfactory’), good, outstanding.

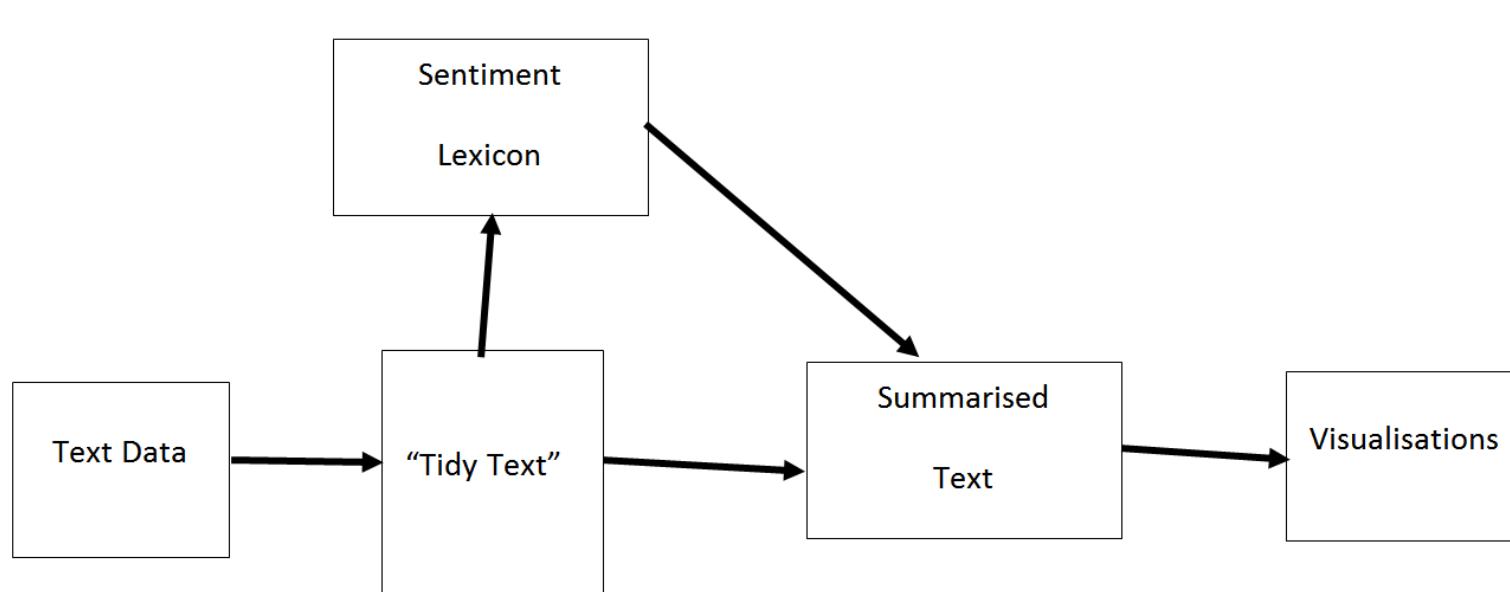


Take a picture to  
download the  
full paper

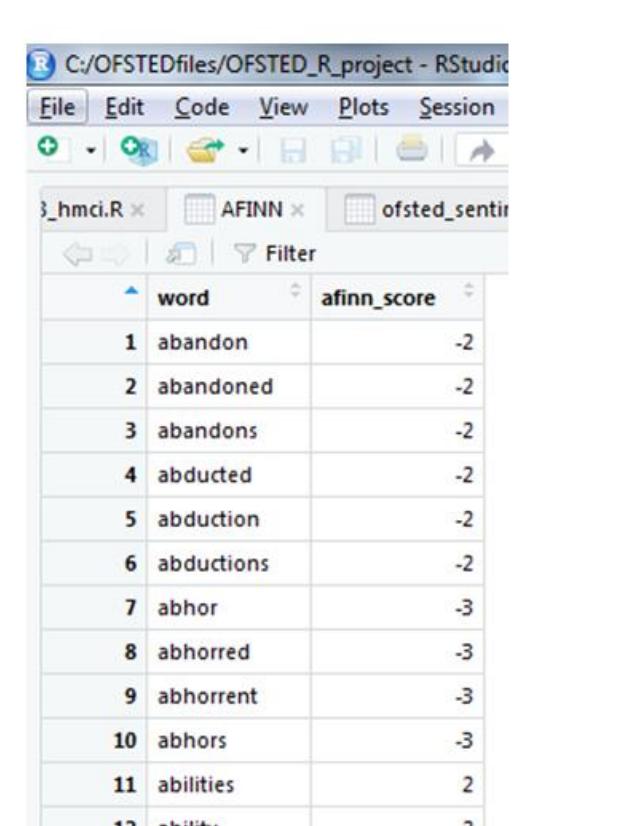
HMCI	In office	Grouping	#
Mike Tomlinson	2000-2002	2000-2002	712
Sir David Bell	2002–2006	2003-2006	1492
Maurice Smith	January 2006–October 2006 (acting)		
Christine Gilbert	2006–2011	2007-2011	5220
Miriam Rosen	July 2011–December 2011 (acting)		
Sir Michael Wilshaw	January 2012–December 2016	2012-2016	8881
Amanda Spielman	January 2017–present	2017	907
		Total	17,21

## Total number and size of inspection documents from 2000 to 2017

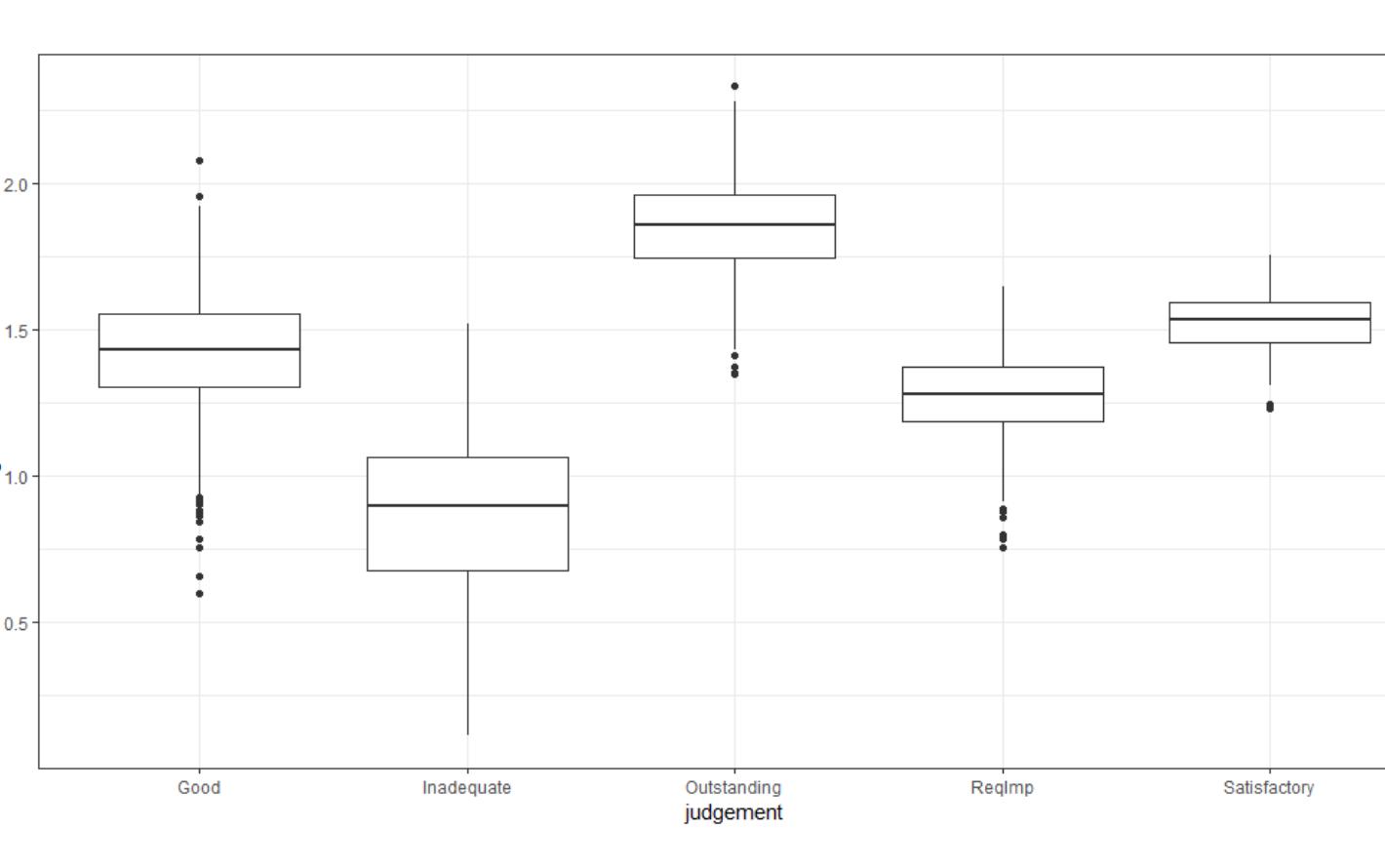
Year	Number of documents	Size in Mb
2000	228	36.9
2001	287	40.4
2002	197	33.8
2003	212	37.8
2004	303	39.1
2005	310	51.7
2006	667	60.3
2007	961	133.0
2008	888	140.0
2009	939	135.0
2010	1113	159.0
2011	1319	223.0
2012	2521	324.0
2013	2125	331.0
2014	1740	274.0
2015	1419	225.0
2016	1076	170.0
2017	907	138.0
Total	17212	2.49 Gb



**Depiction of the flowchart of a typical text analysis that uses tidytext for sentiment analysis.**  
Adapted from Silge and Robinson (2017).



## Experiments of the AEINN architecture



## REFERENCES

## REFERENCES

Nielsen, F.Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages* (pp. 93-98). <http://arxiv.org/abs/1103.2903>

Silge, J., & Robinson, D. (2018). *Text Mining with R - A Tidy Approach*. O'Reilly Media: Sebastopol, CA.

The layout based on [@mikemorrison](#)'s #betterposter ideas