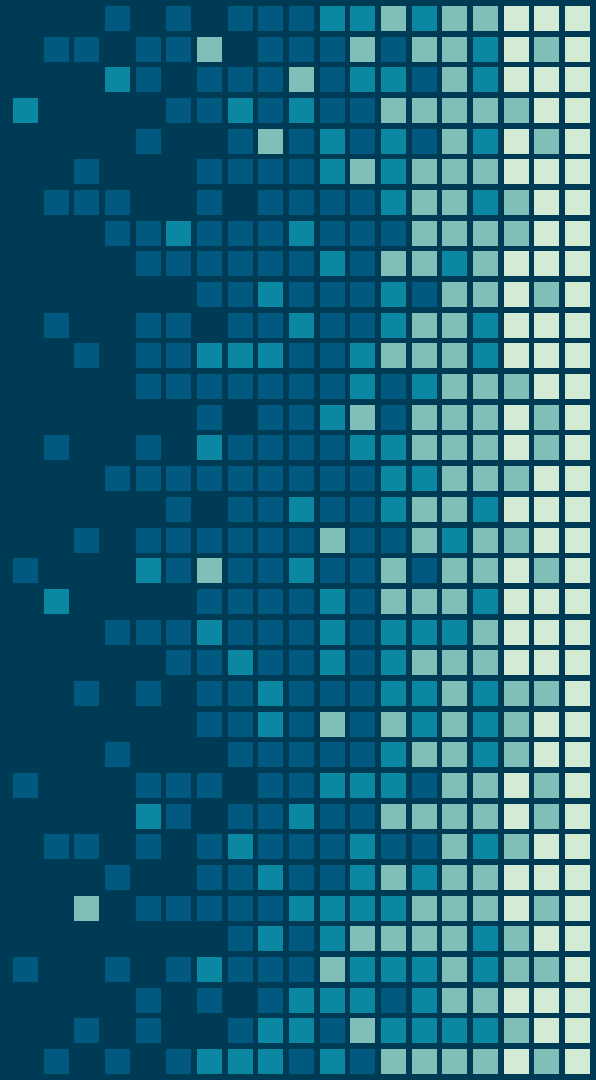


# Comparing groups of texts

Joanna Byszuk, Artjoms Šeļa and  
Maciej Eder

Budapest workshop



# Craig's Zeta

in Stylo



# About Craig's Zeta



# ~~Frequencies matter~~

- Zeta focuses on CONSISTENCY of appearance of a word/feature in the text



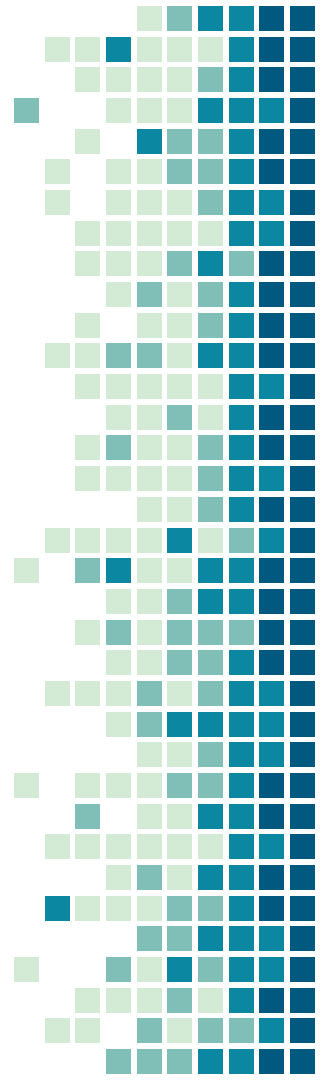
# Distribution matters



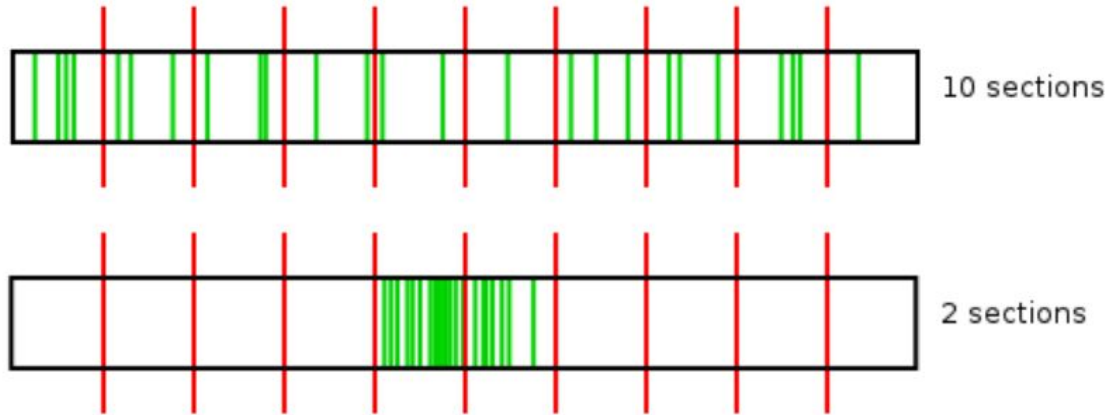
25 times



25 times



# Distribution matters



# Craig's Zeta – how to

- Take a corpus, split it in two you hypothesize are different in a significant way
- Slice each text into smaller sections
- See if the word is there



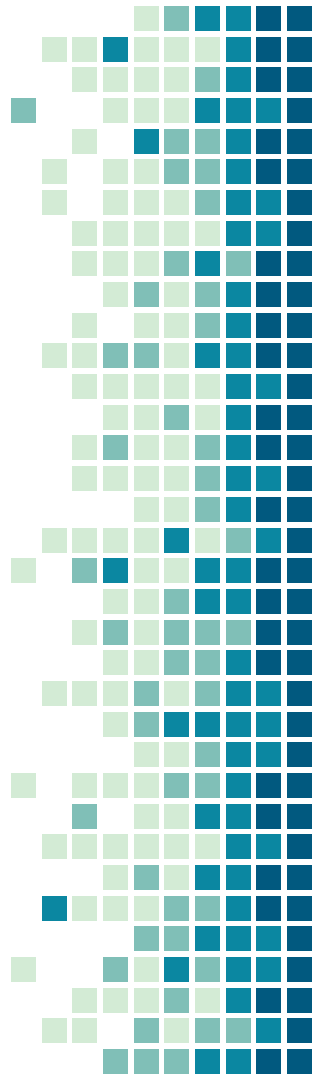
# Craig's Zeta

$$\zeta_{(a,b)} = \left( \frac{f_{(a)} - f_{(b)}}{100} \right) + 1$$

$$\zeta_{(a,b)} = \frac{f_{(a)} - f_{(b)}}{f_{(a)} + f_{(b)}}$$

$f(a)$  – frequency in section 'a'

$f(b)$  – frequency in section 'b'



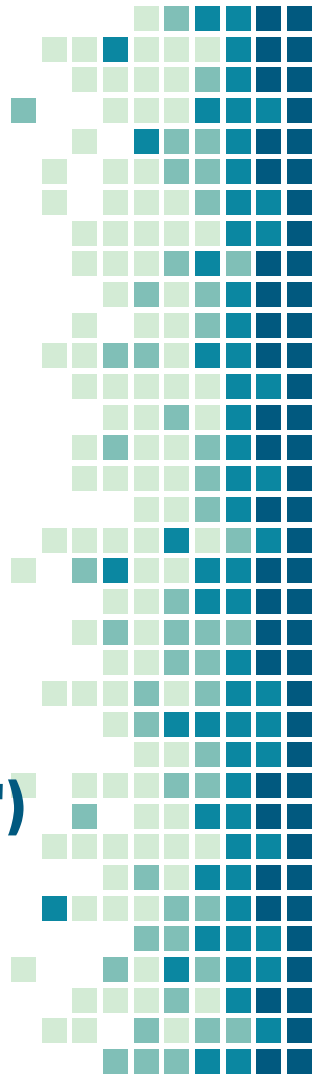


# Calling the function

**oppose()**

Or if it's not in English:

**oppose(encoding = "UTF-8", corpus.lang = "Spanish")**



# Calling the function

## Parameters to consider

- Slice length: size (in words) of the samples (5000)
- Slice overlap: (0)
- Method: (Craig's Zeta)
- Visualization: type of graph (Markers / Words)



# Comparing corpora in lexis

## **primary\_set**

One group of texts (e.g. texts by one author)



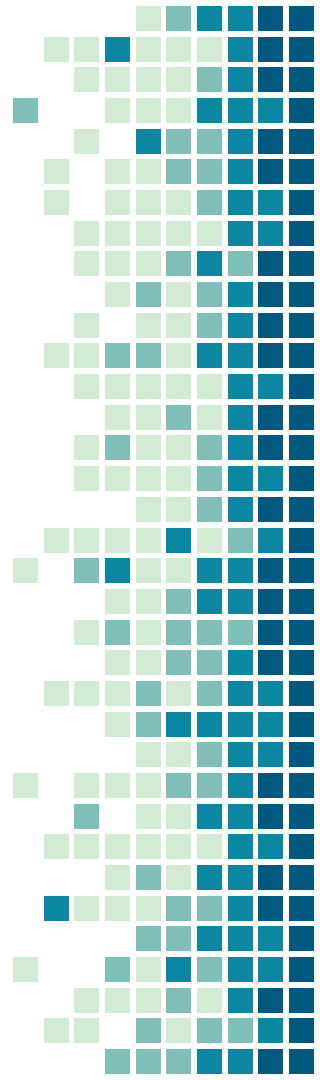
words\_preferred

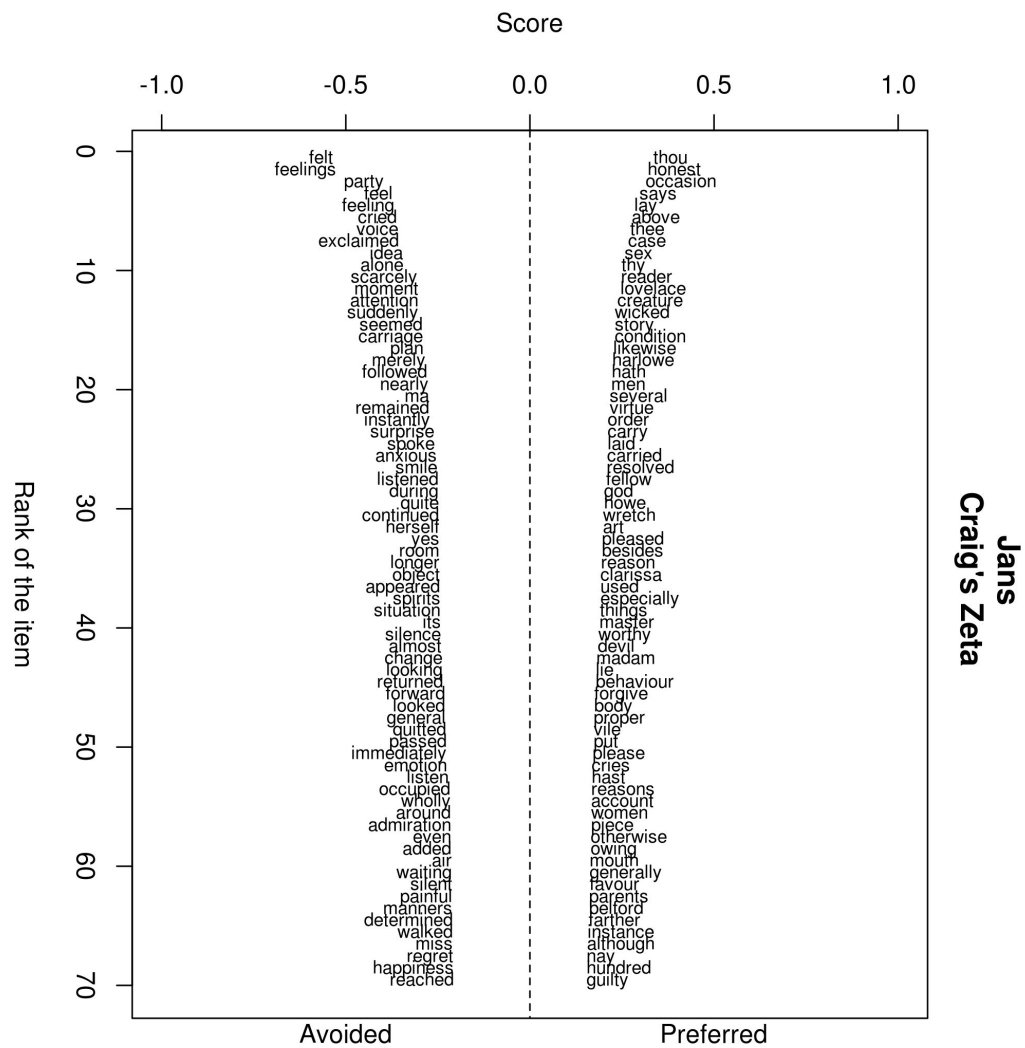
## **secondary\_set**

Another group of texts (e.g. different author)

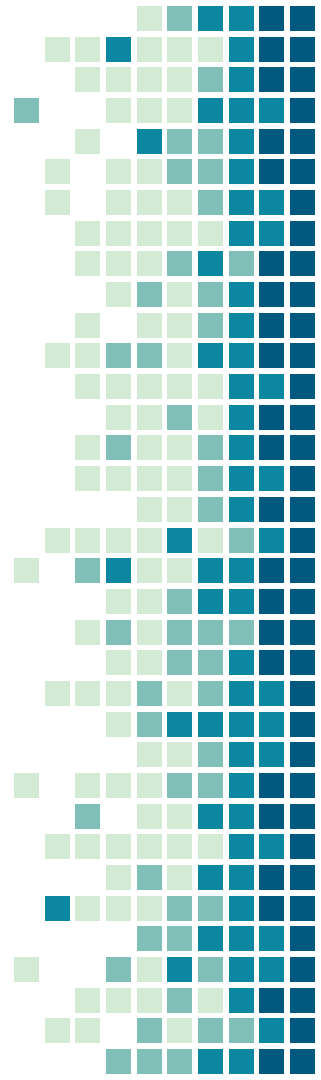


words\_avoided





Comparing  
corpora in  
lexis



# Comparing corpora in markers

## **primary\_set**

One group of texts (e.g. texts by one author)



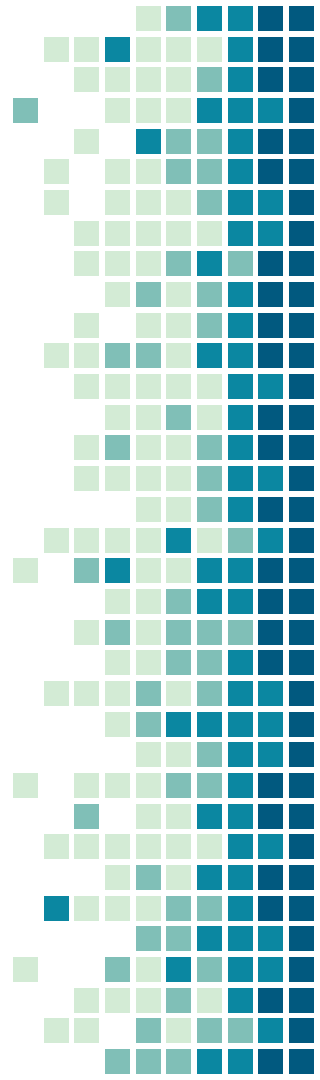
A group  
of marker points

## **secondary\_set**

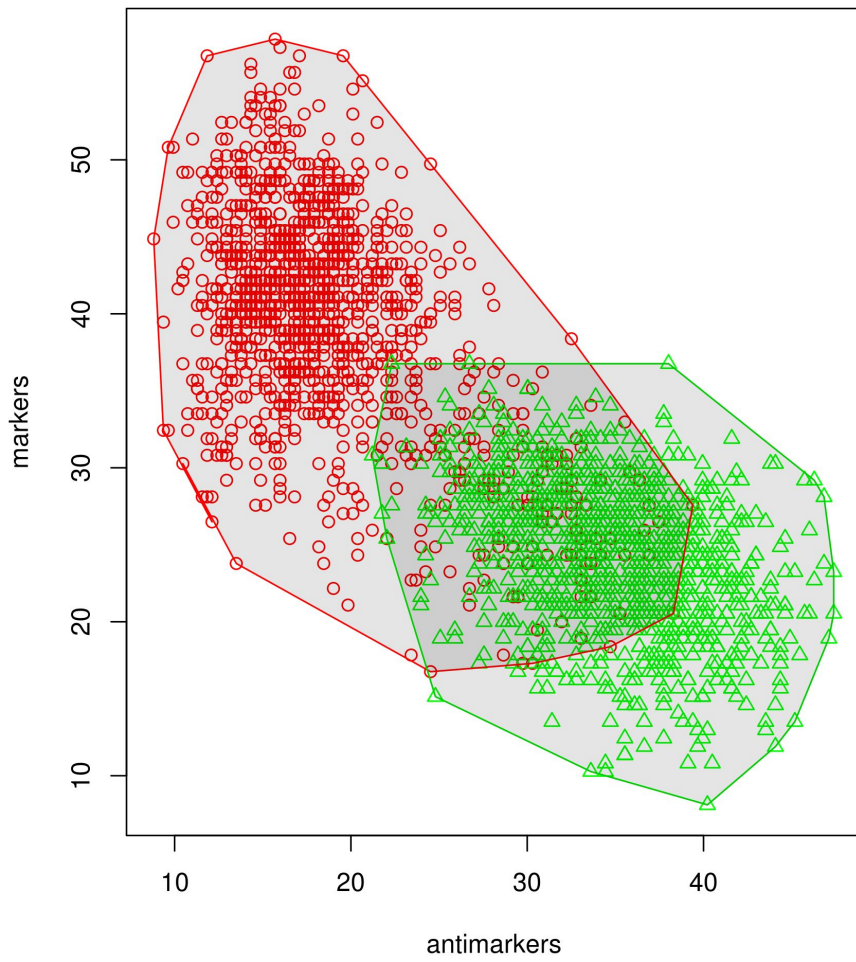
Another group of texts (e.g. different author)



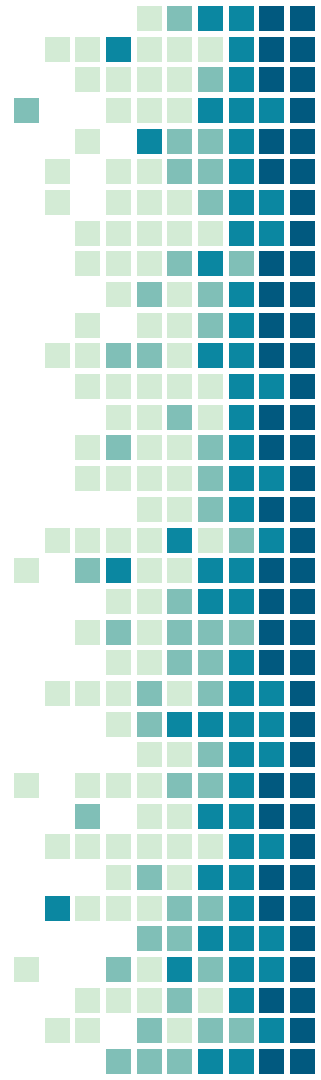
Another group  
of marker points



Jans  
Craig's Zeta



Comparing  
corpora in  
markers



# Applications



# Applications

- Comparing discourse markers
- Seeing how distinct two groups *really* are





# Case study: the tale of two showrunners

Discourse in TV shows



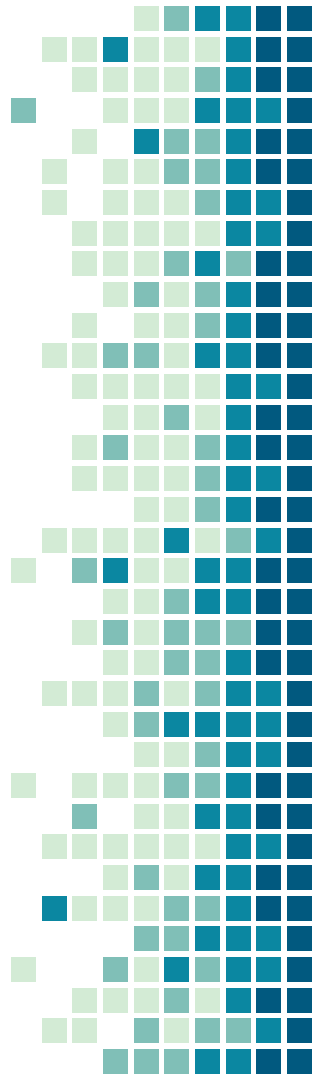
# Who writes TV shows?

9 seasons of Doctor Who = 26 screenwriters

What about directors?

Script editors?

... actors?



Showrunner – the producer who actually [holds] ultimate management and creative authority for the program (...) the boss, the hard-nosed operational manager responsible for all creative aspects of a show and who runs interference with the studios and the networks.

Shane Bannon

<https://www.theaustralian.com.au/arts/getting-the-run-of-the-place/news-story/78b66658bc660728f73a2def314a3788>



# Doctor Who (1963 – )

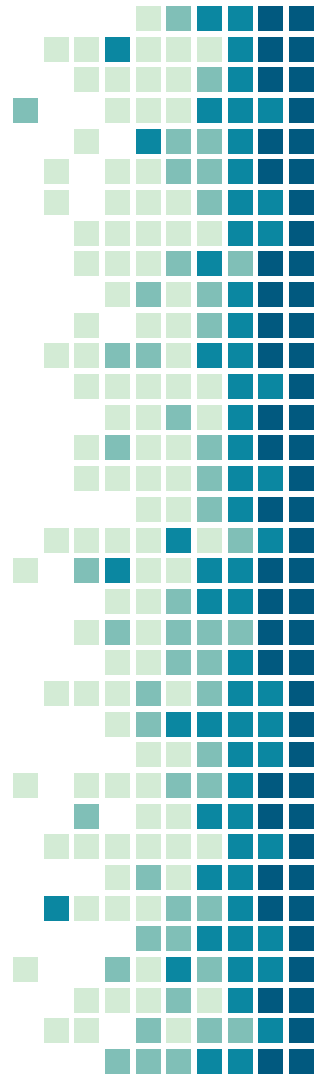
1963-89 'Classic' series

- focus on the main character

1996 Film

2005- ? revival of the show: New/Nu series

- transition to authorial – American model, increased role of a showrunner

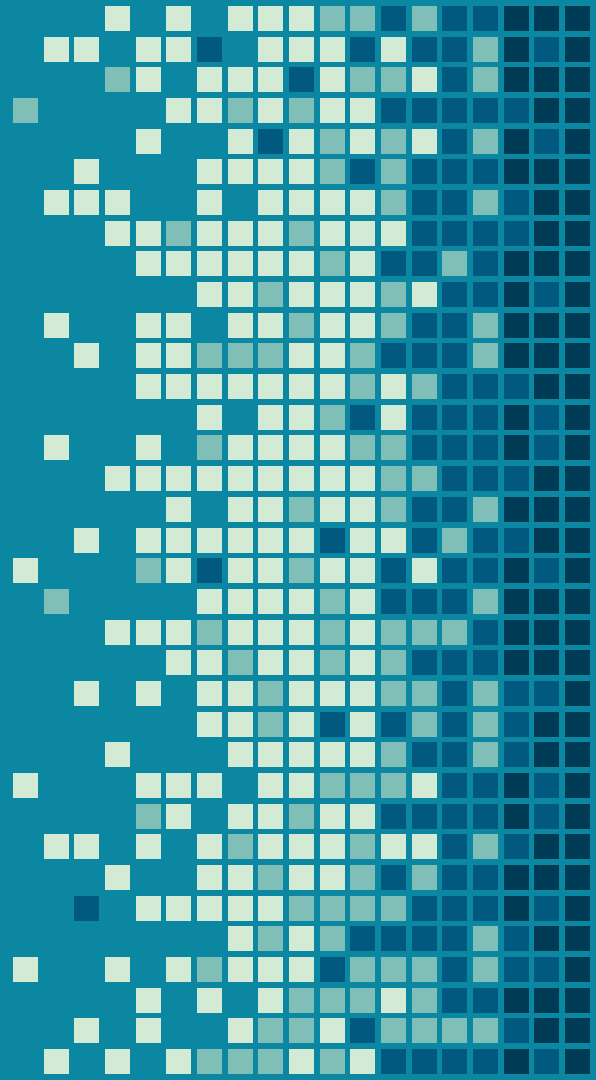


“ I might change at least 30 per cent of the material, often 60 per cent, sometimes almost 100 per cent. I go over every line of dialogue, either adding new stuff or refining what's there; sometimes that means enhancing a line that the original writer hasn't realised is good. I'll bring out themes, punch up moments, add signature dialogue, clarify stage directions and make cuts. To every single scene, if need be. Usually, the basic shape remains intact, but sometimes I'll invent brand new characters and subplots... while at the same time remaining faithful to the original writer. I'll even impersonate them.

Russell T Davies

“ what you’re there to do is to say in  
different ways every single day, to send  
out a consistent message: this is the show  
we’re trying to make, it’s like this.

Steven Moffat

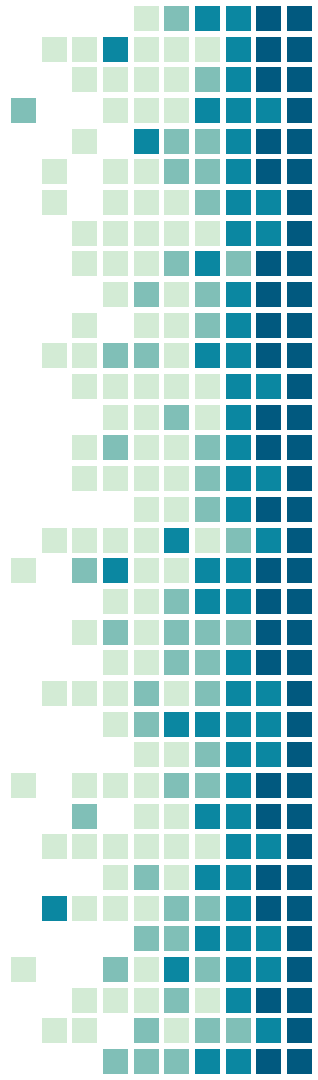


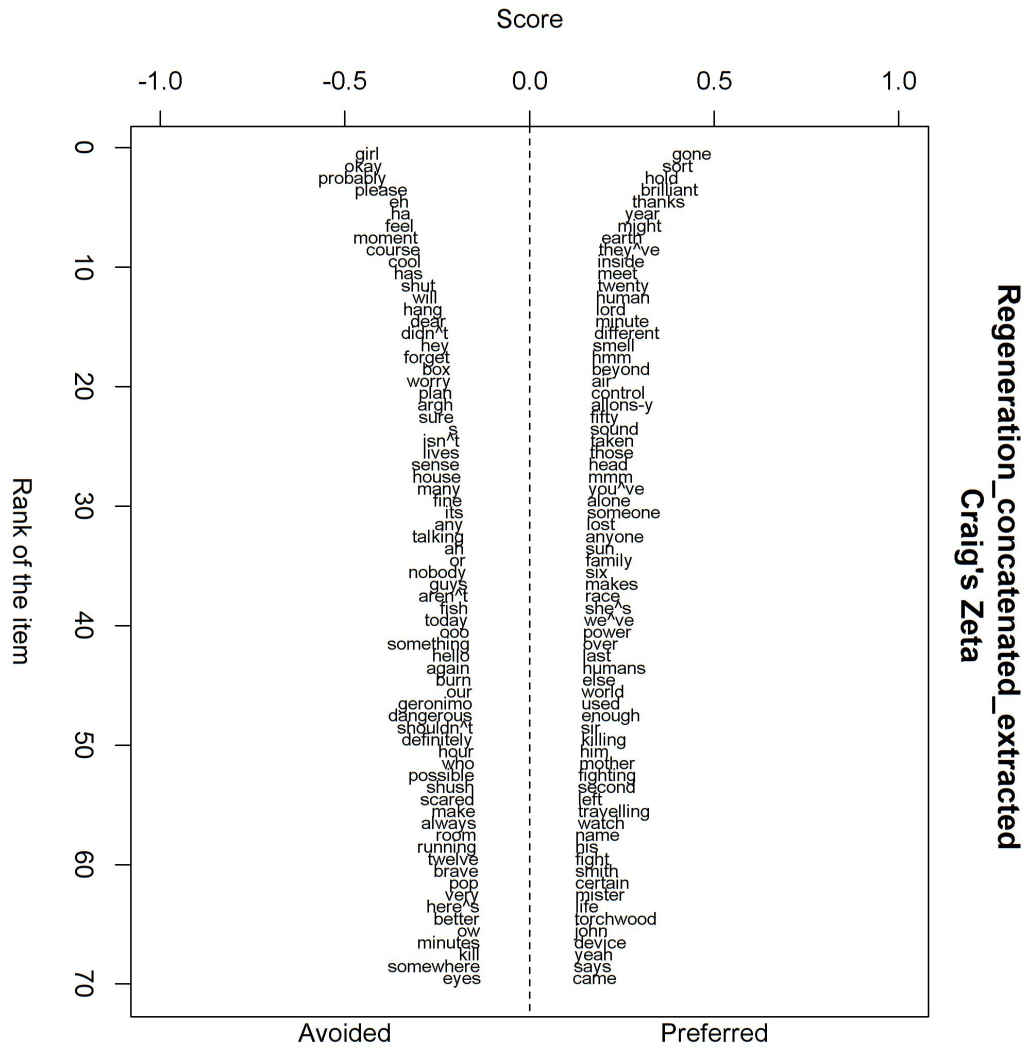
# An episode dissected

1 episode of Doctor Who - from 4 to 7 thousand words,

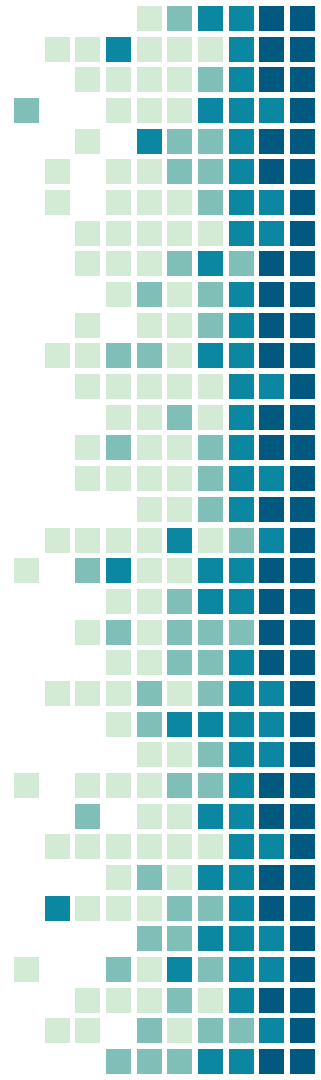
Doctor's lines - on average 1,5 to 2,8 thousand words,

A companion's lines - on average 500 to 1 thousand words.





Davies's vs  
Moffat's  
Doctor





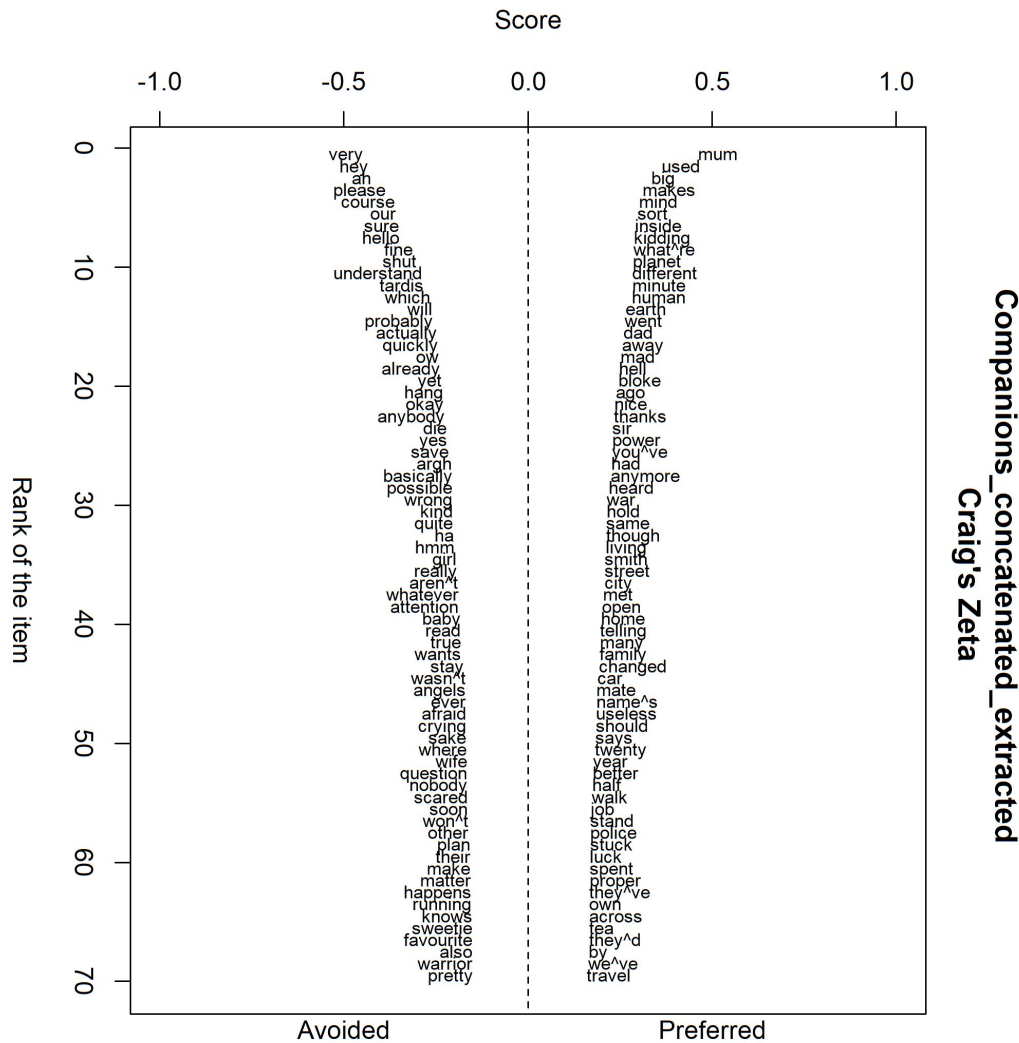
# Davies vs Moffat

gone, sort, hold brilliant, thanks,  
year, might, earth, they've,  
inside, meet, twenty, human,  
lord, minute, smell, hmm,  
beyond, air, control, allons-y,  
fifty, sound, taken, those, head,  
mmm, you've, alone, someone,  
lost, anyone, sun, family, six,  
makes, race, she's, we've,  
power, over, last, humans, else,  
world, used, enough, sir, killing,  
him, mother, fighting, second,  
left, travelling, watch, name, bis,  
fight, smith, certain, mister, life,  
torchwood, john, device, yeah,  
says, came

girl, okay, probably, please, eh,  
ha, feel, moment, course, cool,  
has, shut, will, hang, dear, didn't,  
hey, forget, box, worry, plan,  
argh, sure, s, isn't, lives, sense,  
house, many, fine, its any,  
taking, ah, or, nobody, guys,  
aren't, fish, today, ooo,  
something, hello, again, burn,  
our, geronimo, dangerous,  
shouldn't, definitely, hour, who,  
possible, shush, scared, make,  
always, room, running, twelve,  
brave, pop, very, here's better,  
ow, minutes, kill, somewhere,  
eyes



# Companions



# Davies vs Moffat

mum, used, big, makes, mind,  
sort, inside, kidding, what're,  
planet, different, minute,  
human, earth, went, dad, away,  
mad, hell, bloke, ago, nice,  
thanks, sir, power, you've had,  
anymore, heard, war, hold,  
same, though, living, smith,  
street, city, met, open, home,  
telling, many, family, changed,  
car, mate, name's, useless,  
should, says, twenty, year,  
better, half, walk, job, stand,  
police, stuck, luck, spent, propr,  
they've own, across, tea, they'd,  
by, we've, travel

very, hey, ah, please, course, our,  
sure, hello, fine, shut, understand,  
tardis, which, will, probably,  
actually, quickly, ow, already, yet,  
hang, okay, anybody, die, yes,  
save, argh, basically, possible,  
wrong, kind, quite, ha, hmm, girl,  
really, aren't, whatever, attention,  
baby, read, true, wants, stay,  
wasn't, angels, over, afraid, crying,  
sake, where, wife, question,  
nobody, scared, soon, won't,  
other, plan, their, make, matter,  
happens, running, knows, sweetie,  
favourite, also, warrior, pretty

