Statistical Learning and Learning to Read

Davide Crepaldi, Valentina Pescuma, Maria Ktori,
Benedetta Cevoli, Francesca Franzon, Eleonora Lomi
[davide.crepaldi@sissa.it]
[http://lrlac.sissa.it]

International School for Advanced Studies (SISSA), Trieste

AIP, Bari, 21 September 2017

Reading is a human wonder

Reading is outside of our genetic endowment:

- Not observed universally
- Not learned spontaneously

Nearly all readers are astonishingly efficient:

- 8–letter words in ~35ms (e.g., Forster & Davis, 1984)
- ~20 letters every ~250ms (e.g., Rayner, 1998)

Arbitrariness

- elephant
- table
- heat
- drum
- ...

Arbitrariness. Really?

- elephant
- table
- heat
- drum
- ...

- preheat
- juicer

- bioweapon
- guesstimate
- untweet (?)

Statistical learning

Children learning to read

- Natural reading
  - Stories (=connected text)
  - Just read and understand (=no strange task to carry out)
- Eye tracking
- Many children, create a database to share
- Across a natural spectrum of age
- Across a natural spectrum of reading proficiency
- Check sensitivity to statistical regularities

For today

- Data from 22 kids (out of the 80 tested so far)
- nGrams of different size
- Average nGram frequency across whole words

ALBERO

- 2grams: AL, LB, BE, ER, RO
- 3grams: ALB, LBE, BER, ERO
- 4grams: ALBE, LBER, BERO

Brains At Work

Brains At Work

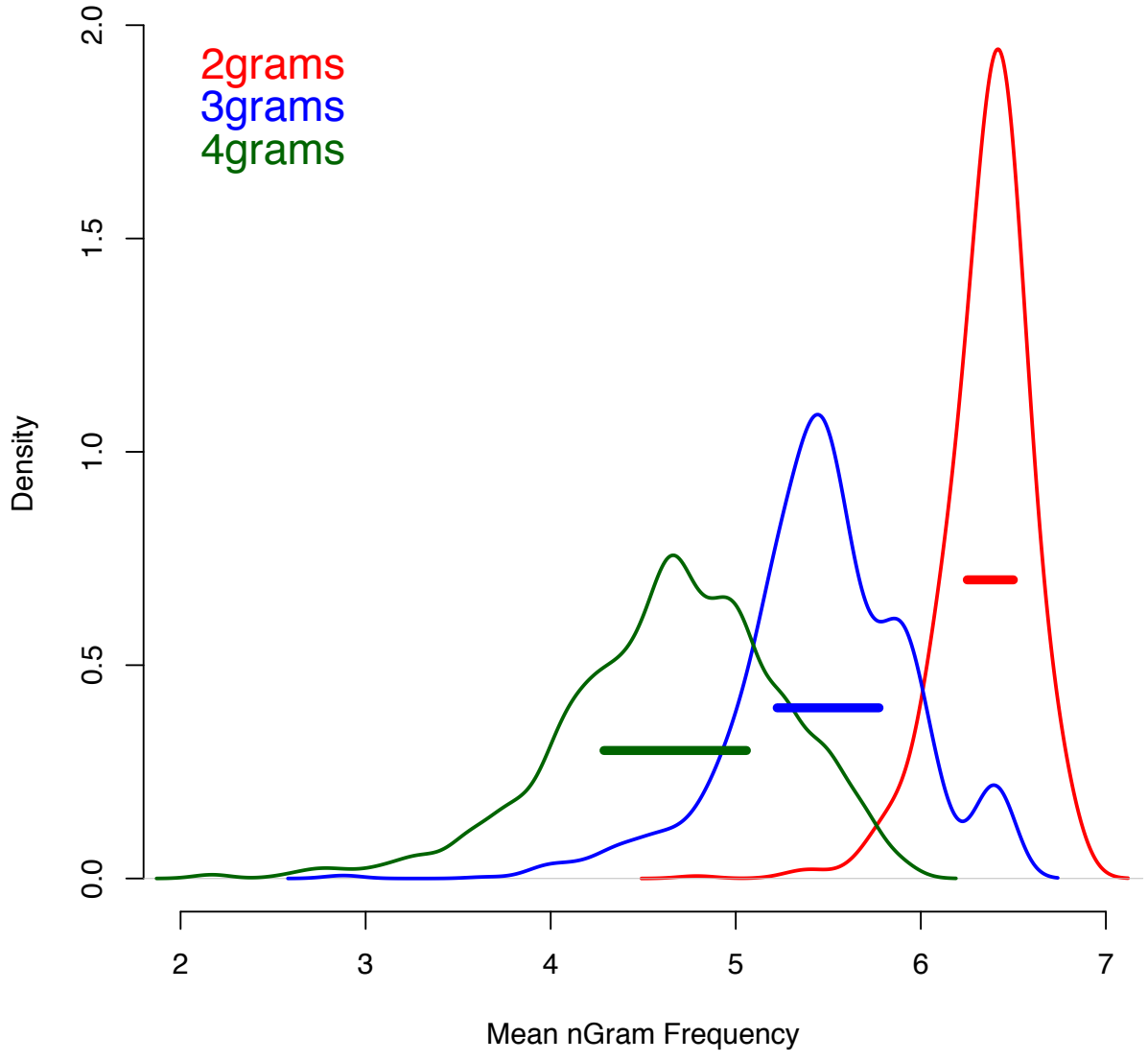
Brains At Work

Brains At Work

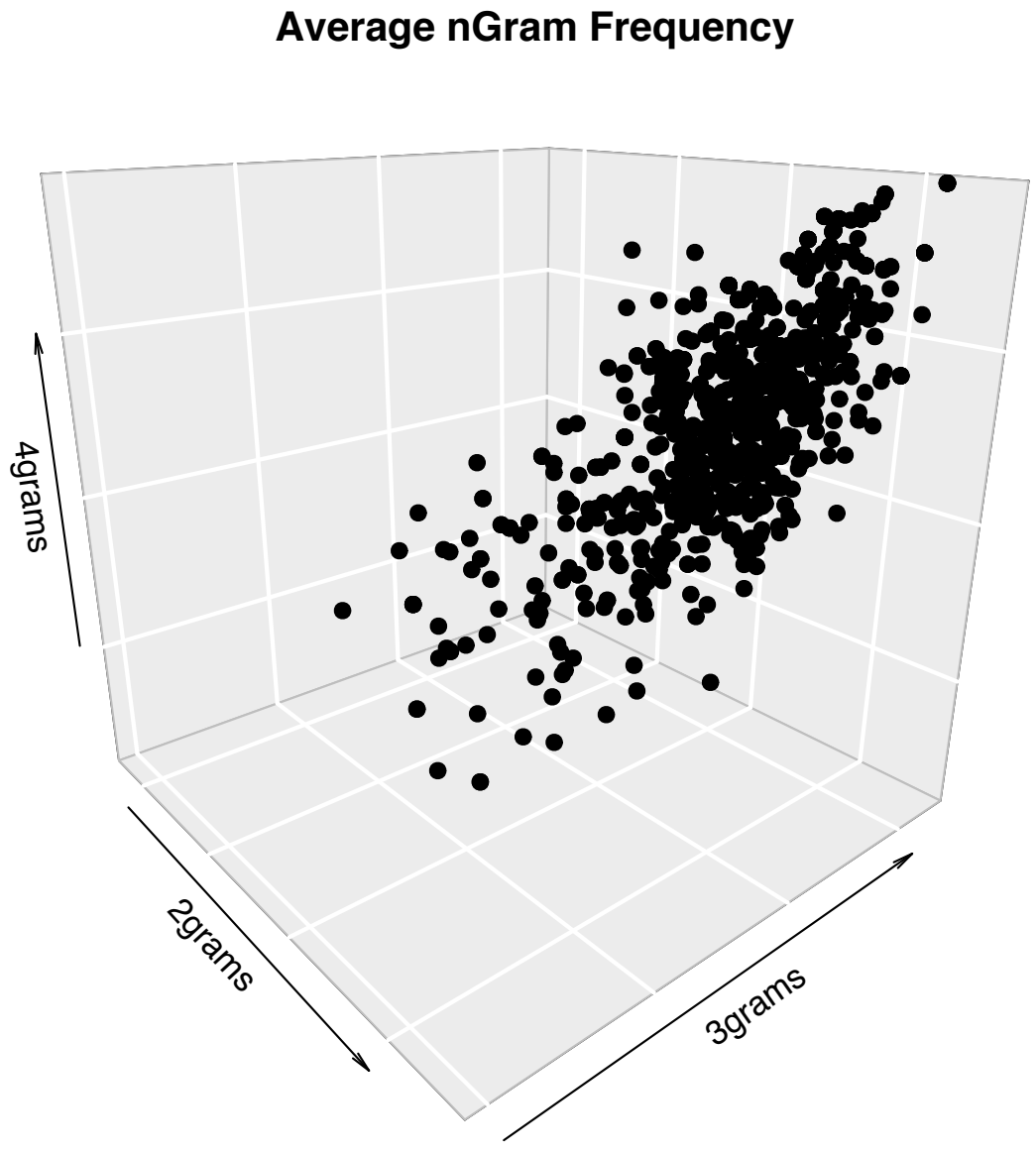
Word sample

- 1745 tokens, from 728 different words, across 12 short stories

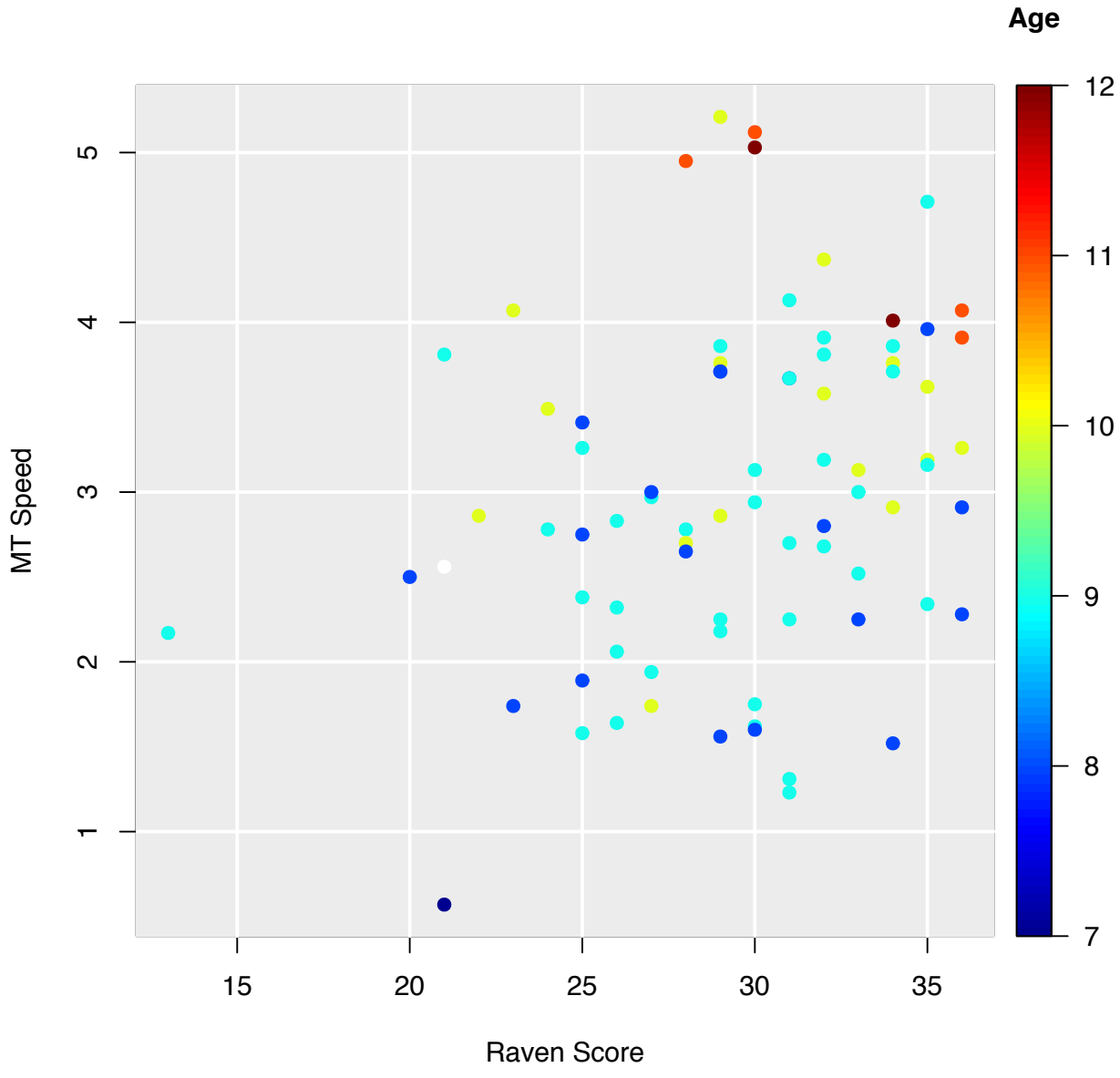
nGrams distribution



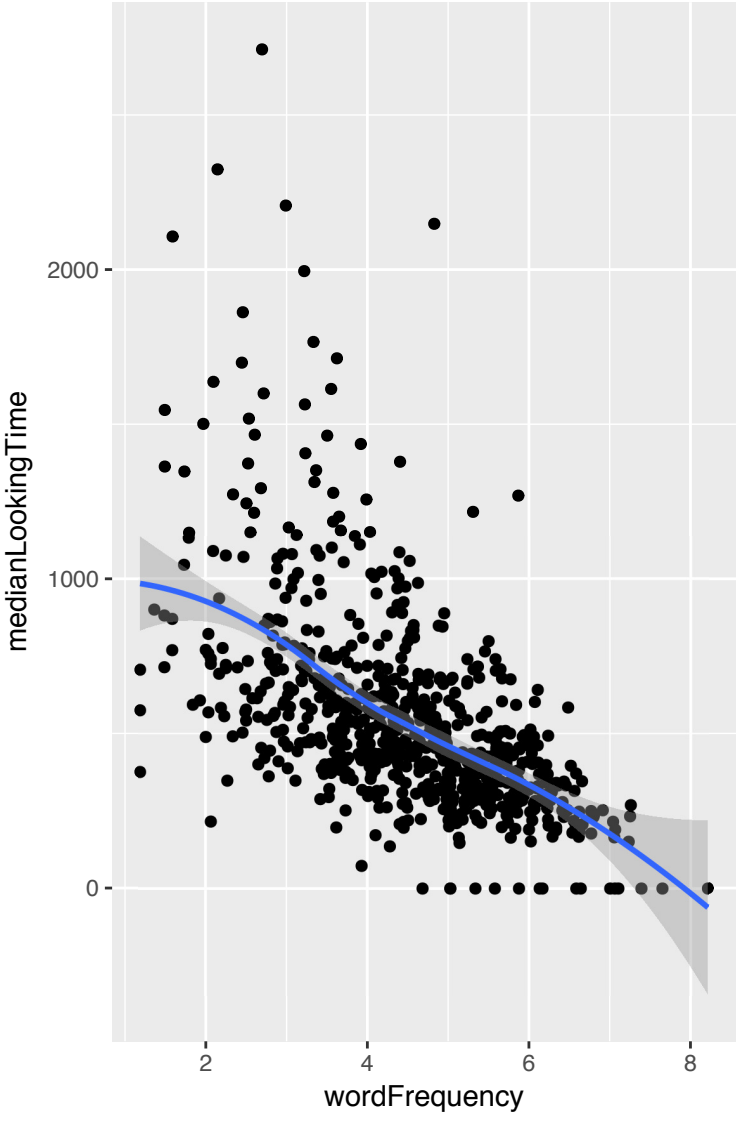
nGrams correlation

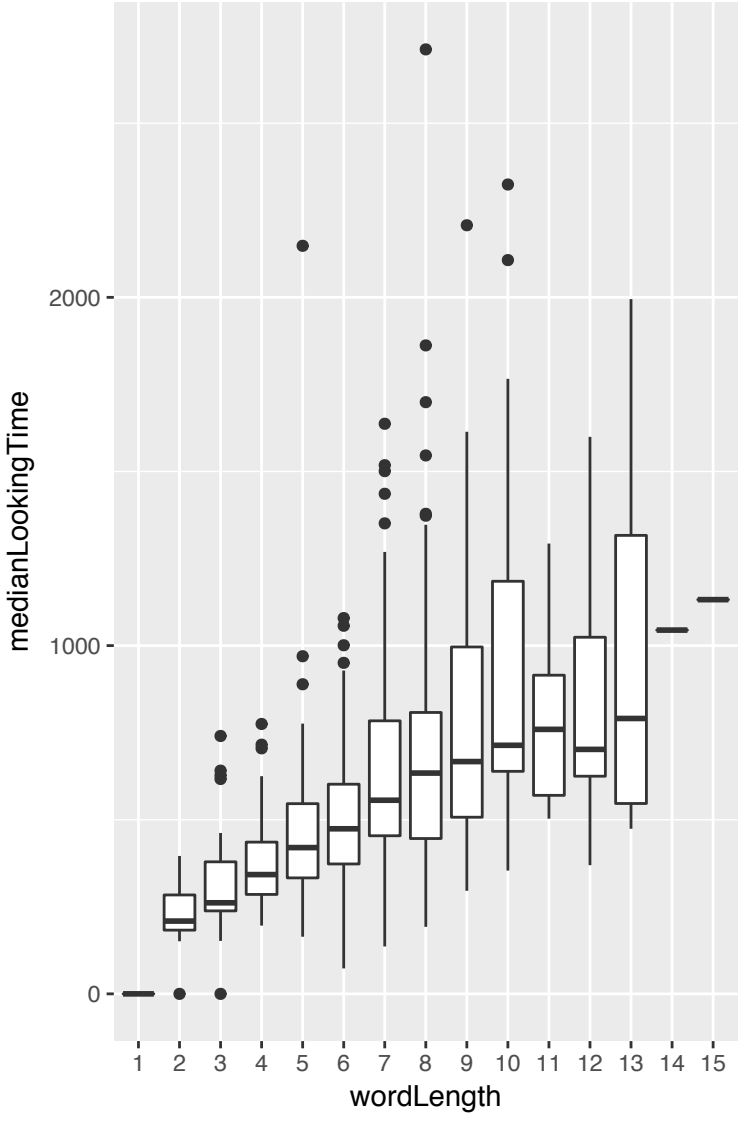


Participant sample

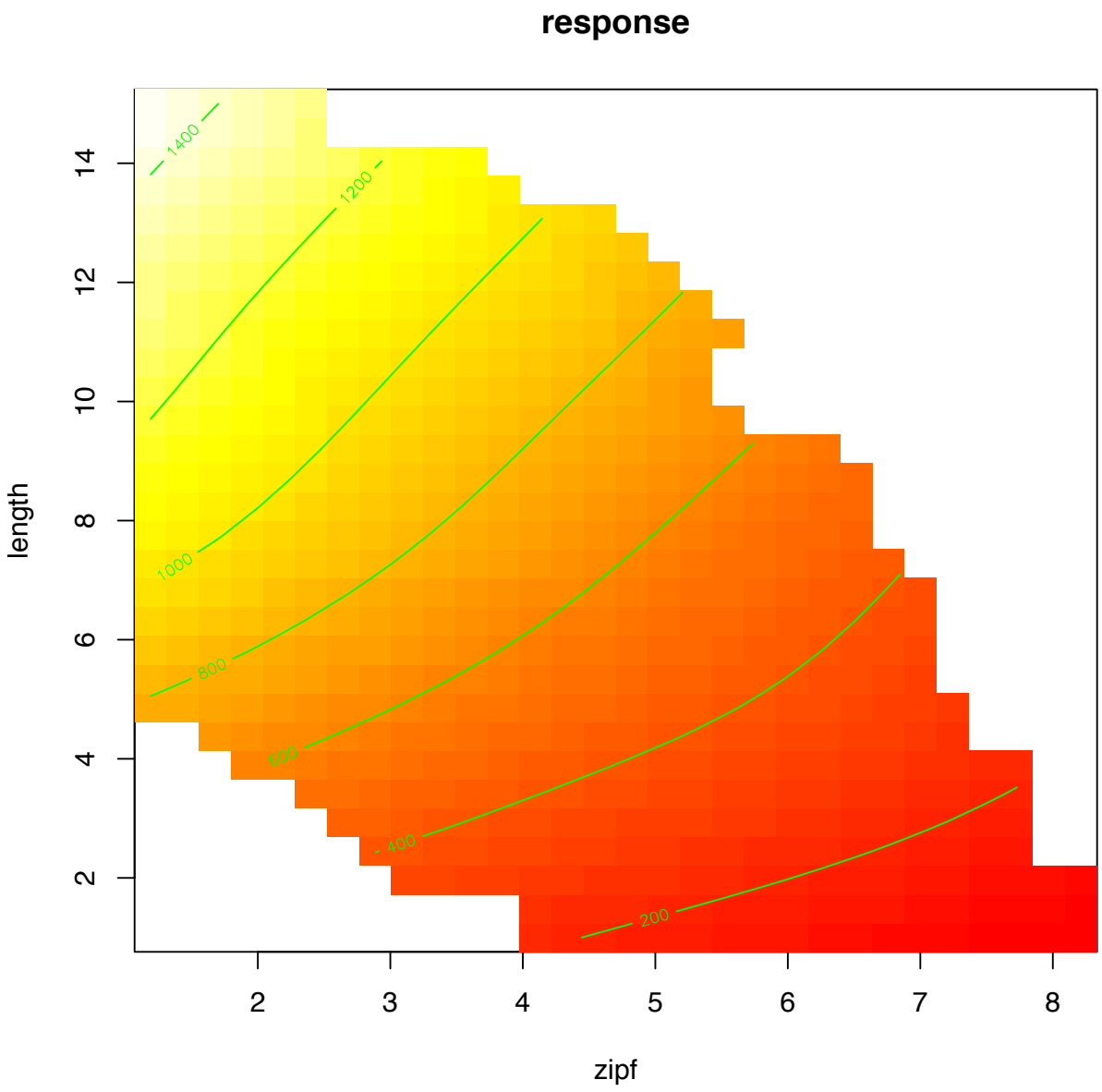


Frequency and length

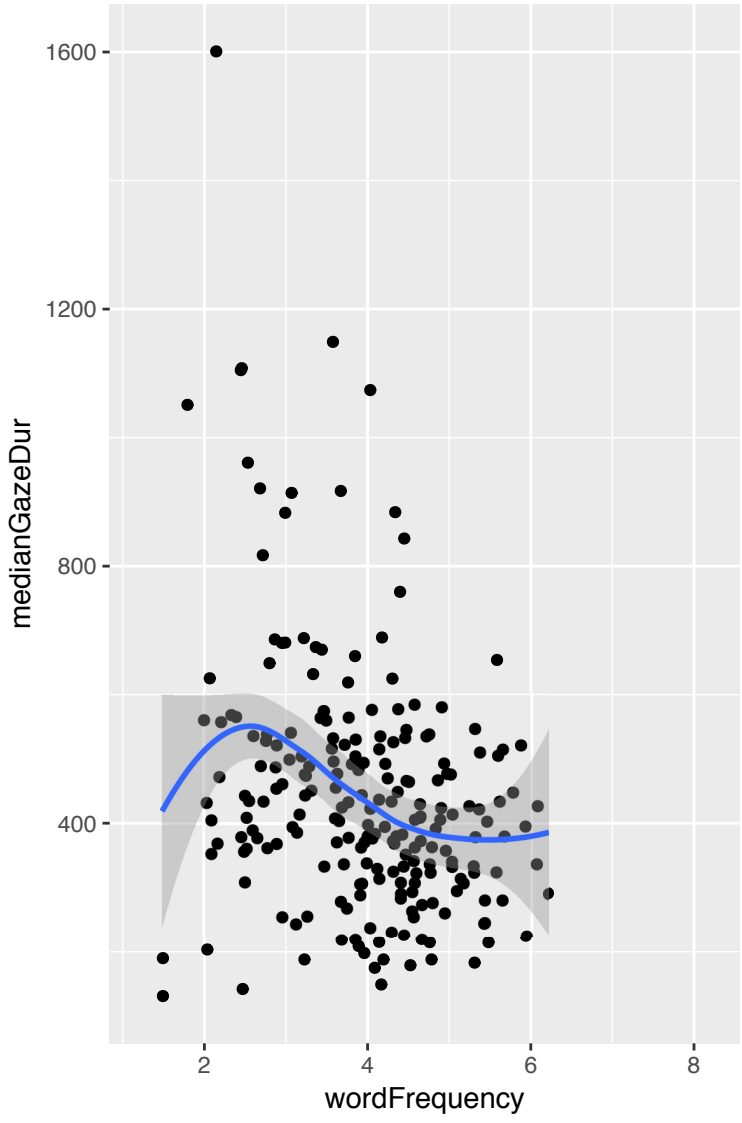


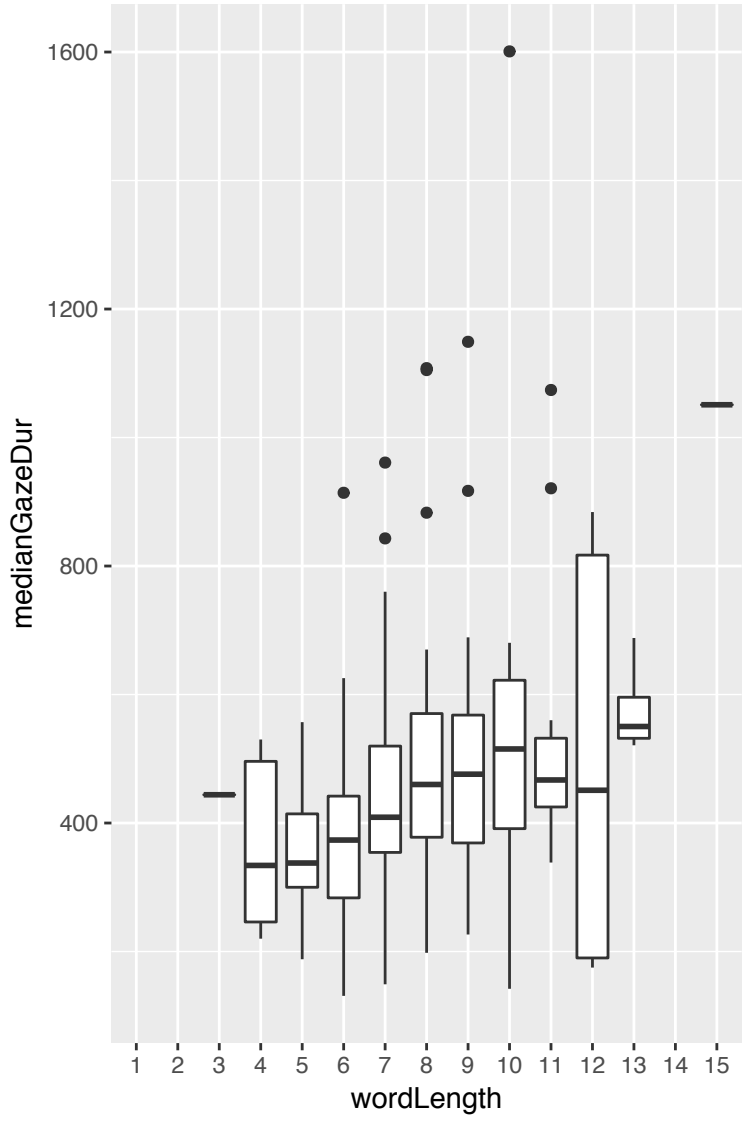


Frequency and length

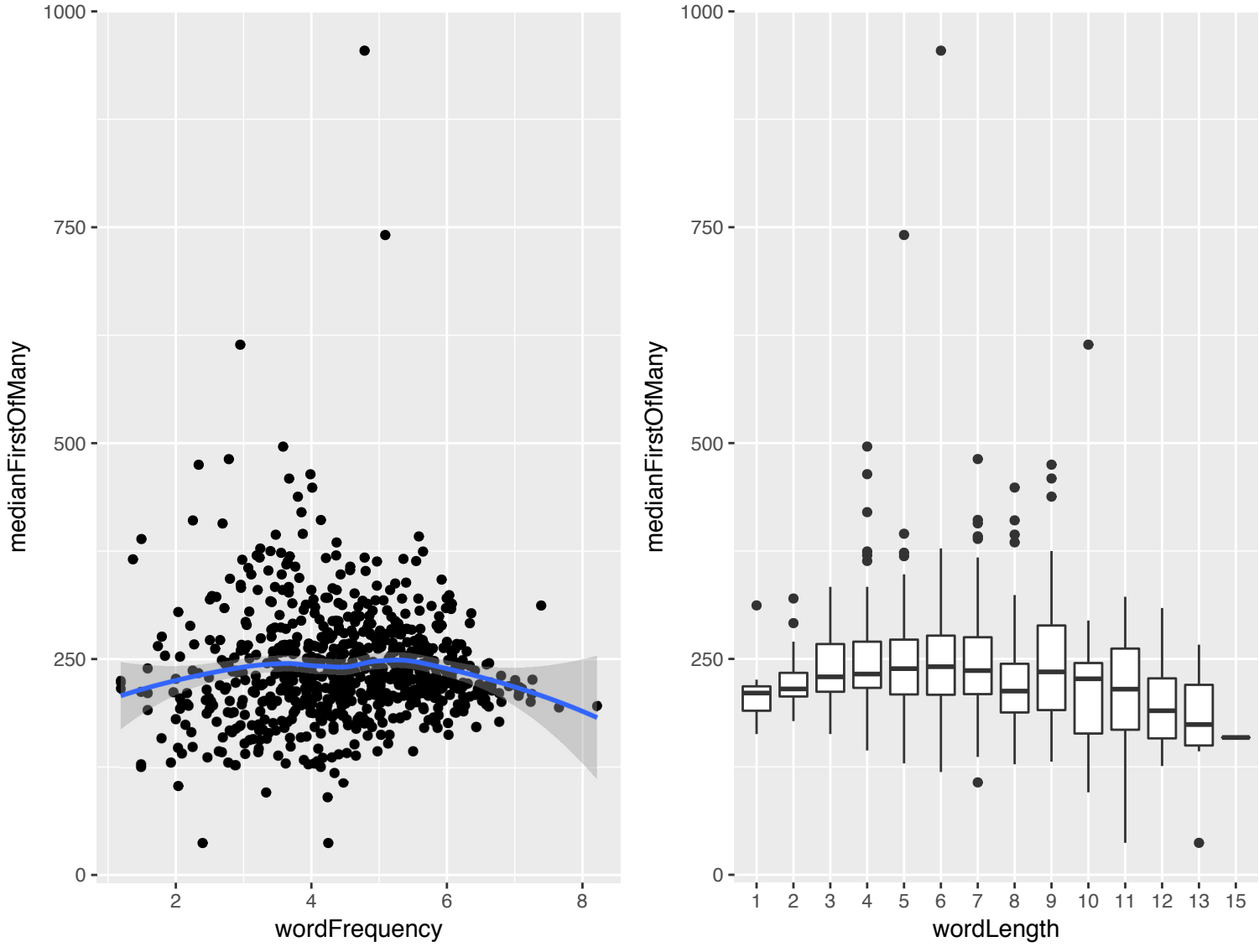


Early on?

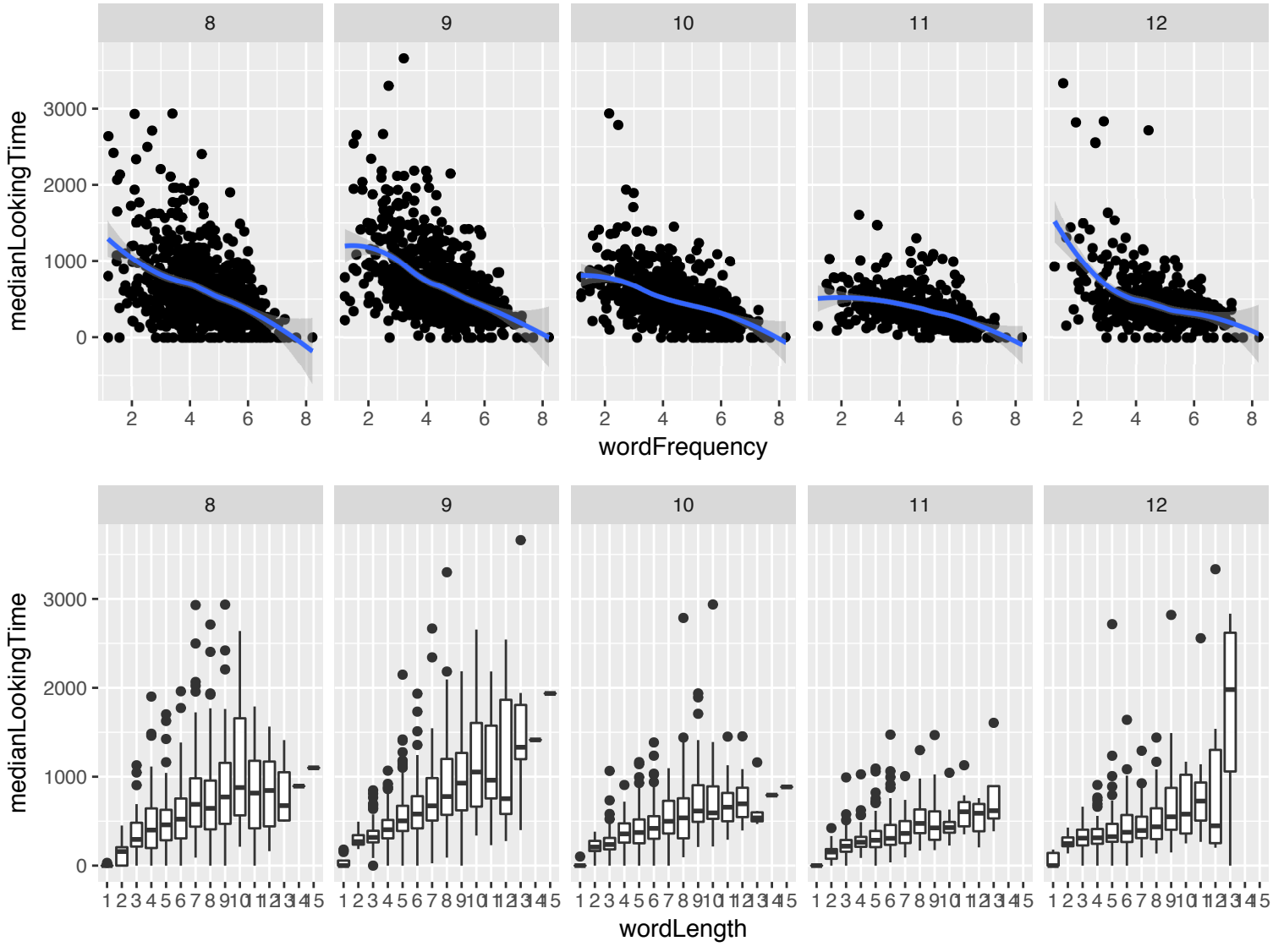

1600
1200
800
400
medianGazeDur
2
4
6
8
wordFrequency


1600
1200
800
400
medianGazeDur
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
wordLength

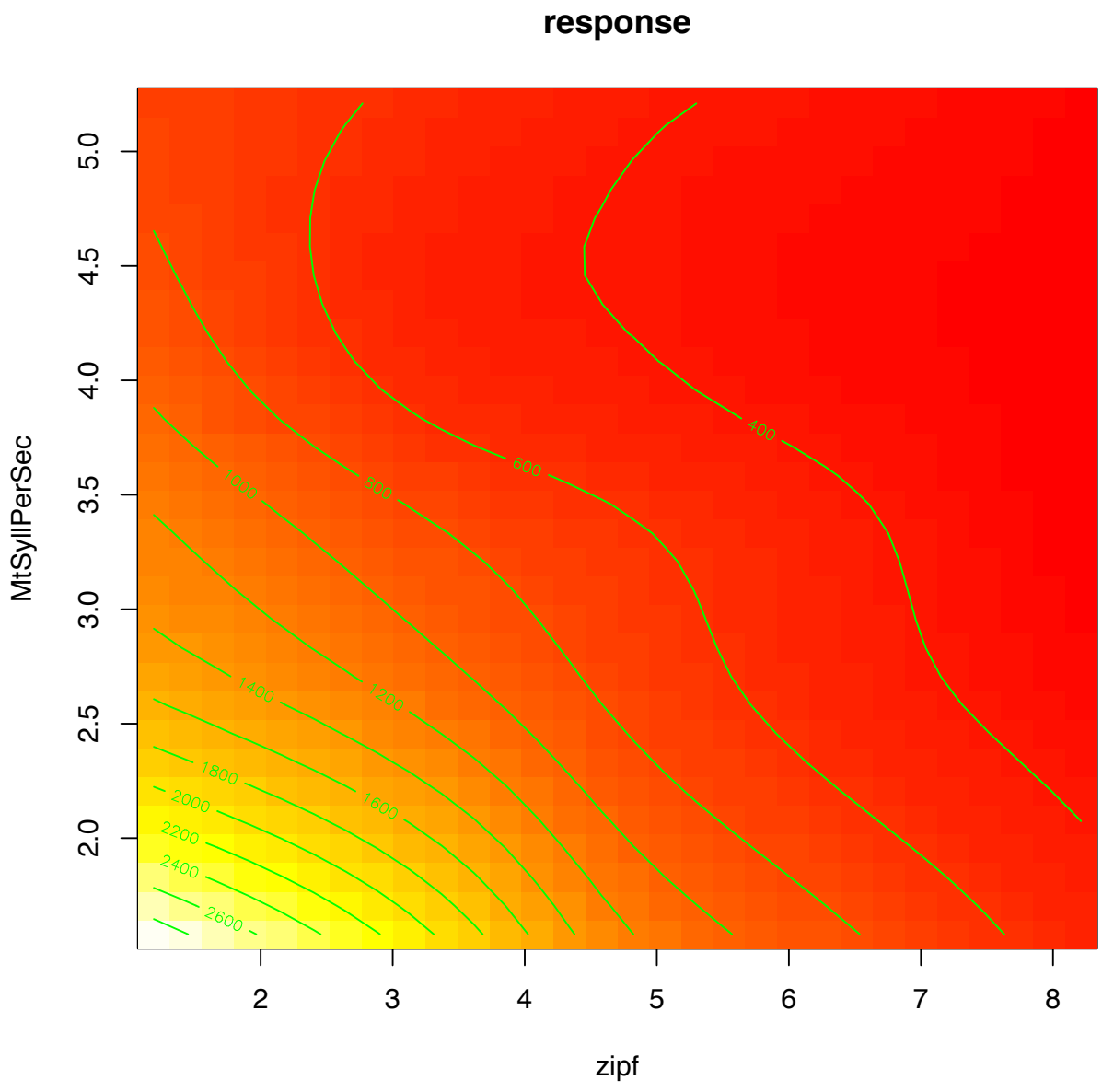
Really early on?



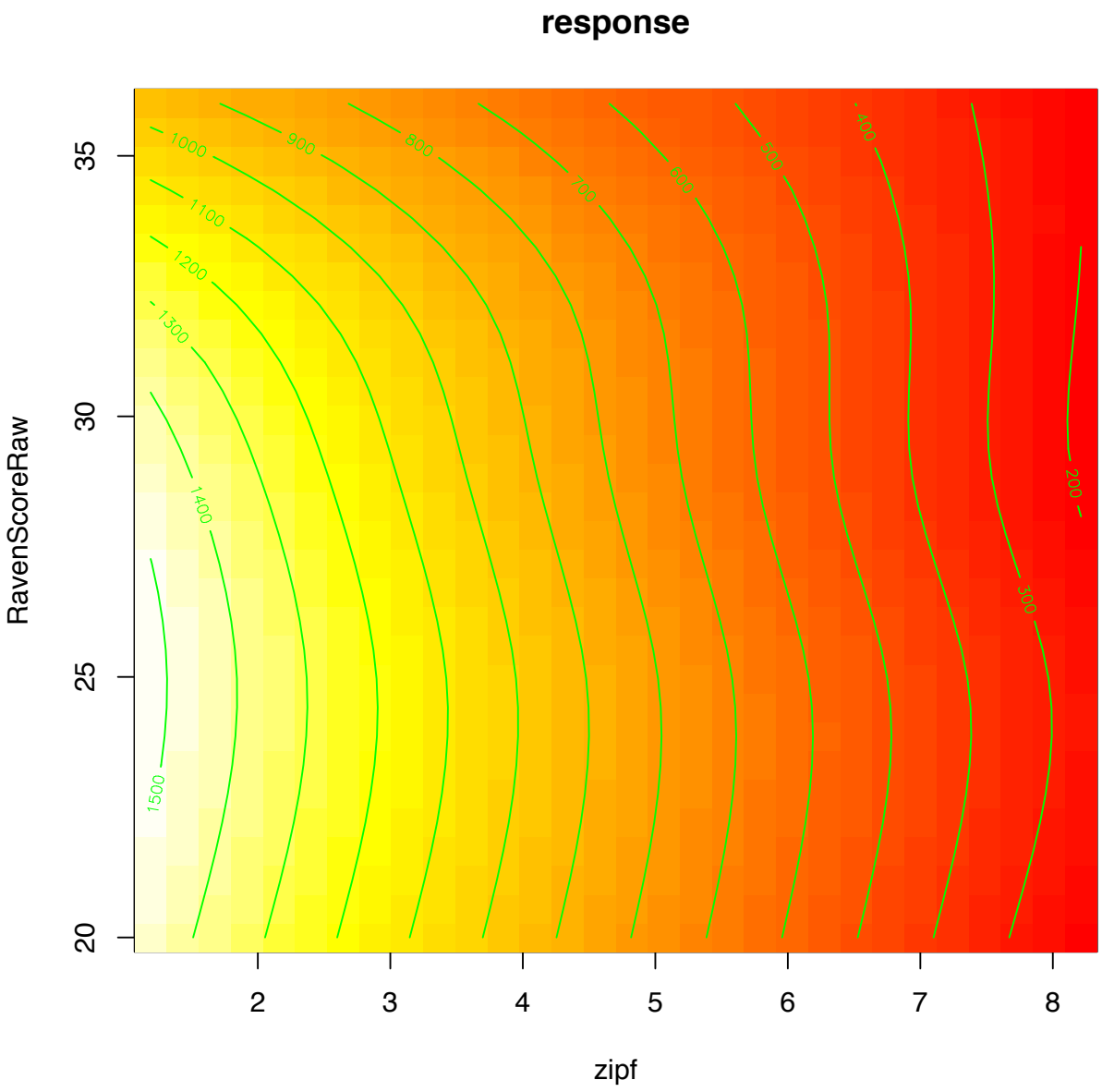
Age effects


8
9
10
11
12
medianLookingTime
3000
2000
1000
0
2
4
6
8
wordFrequency
wordLength
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

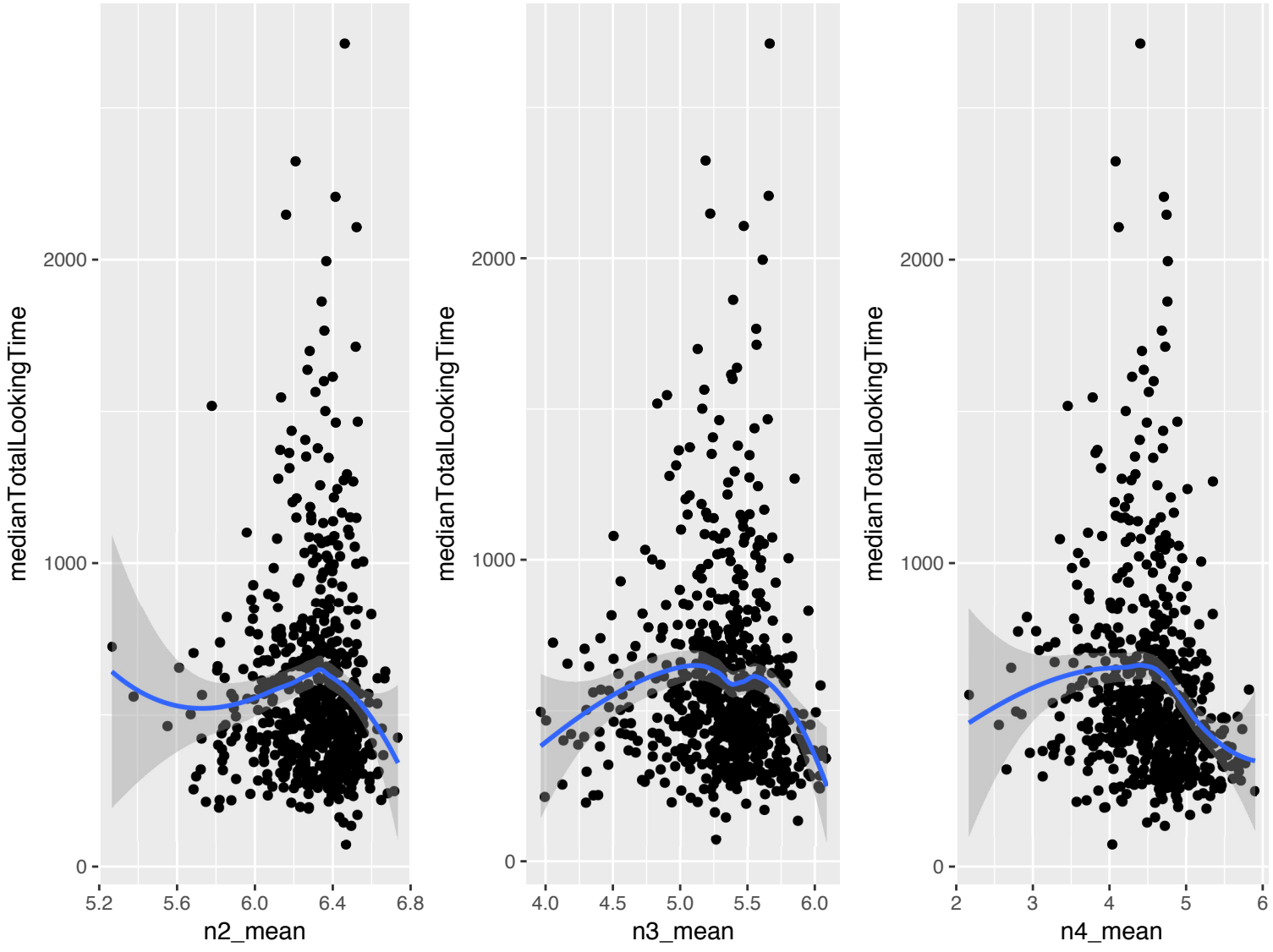
Overall reading speed



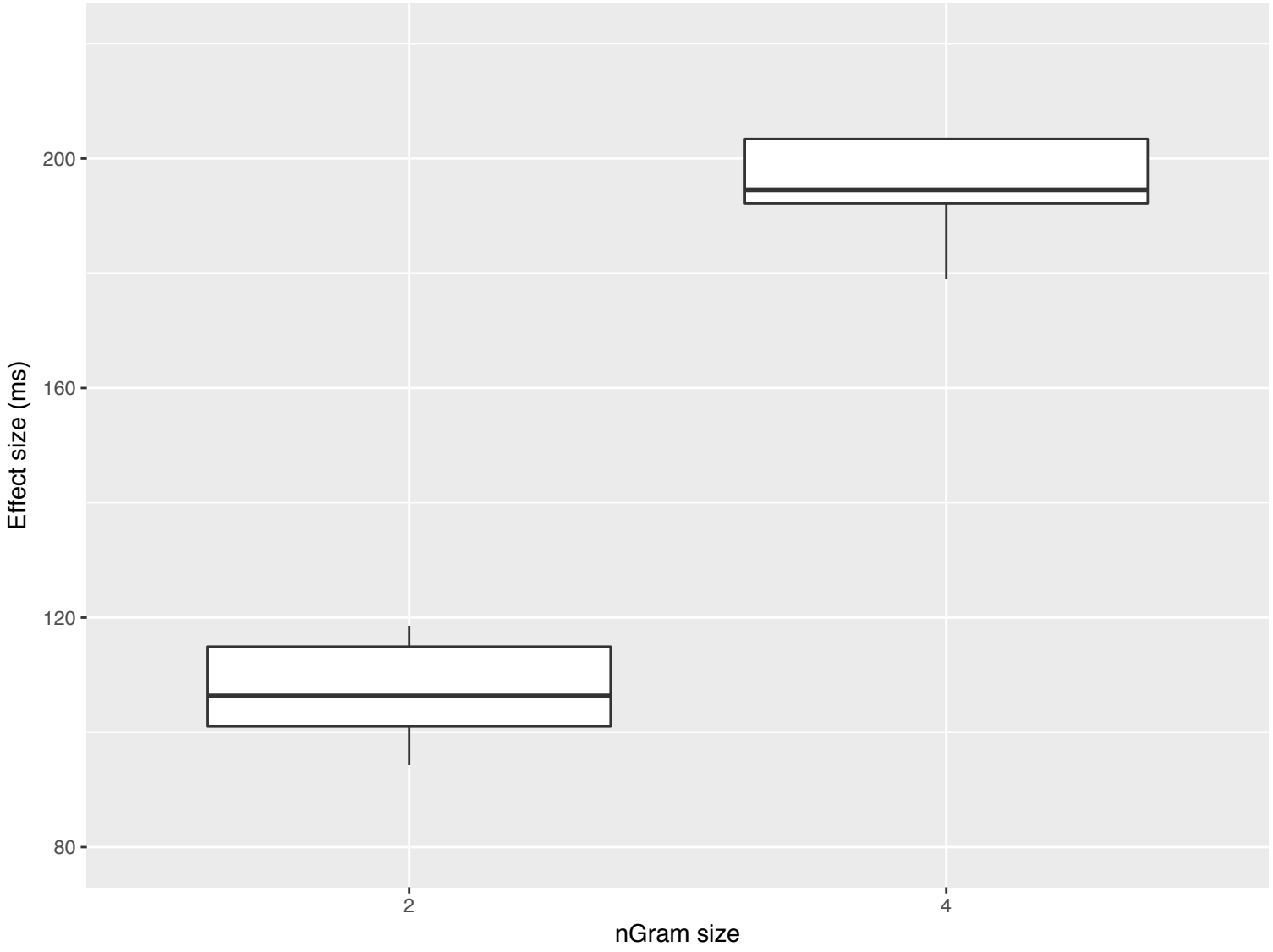
Specific to reading?



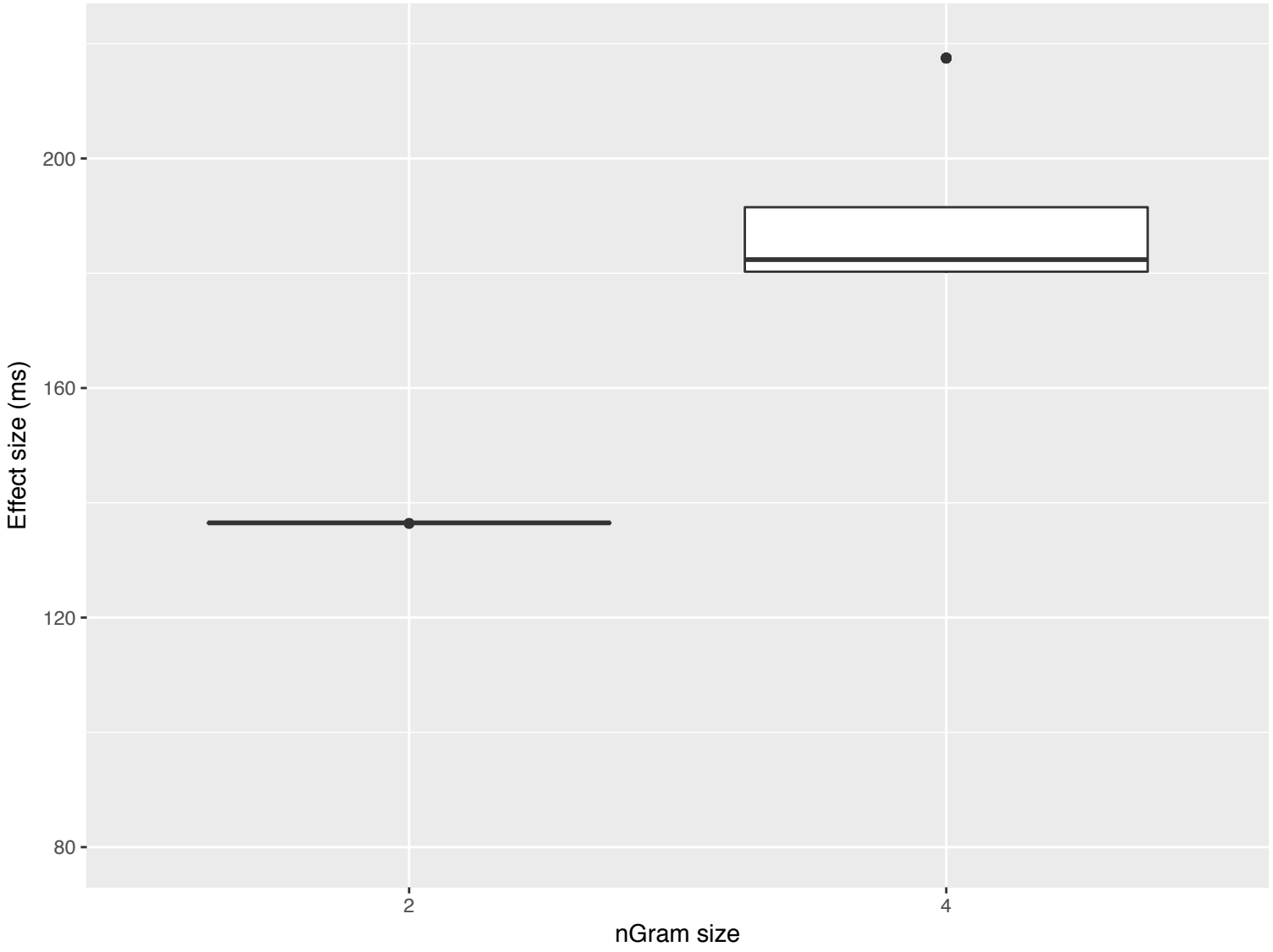
nGrams, finally



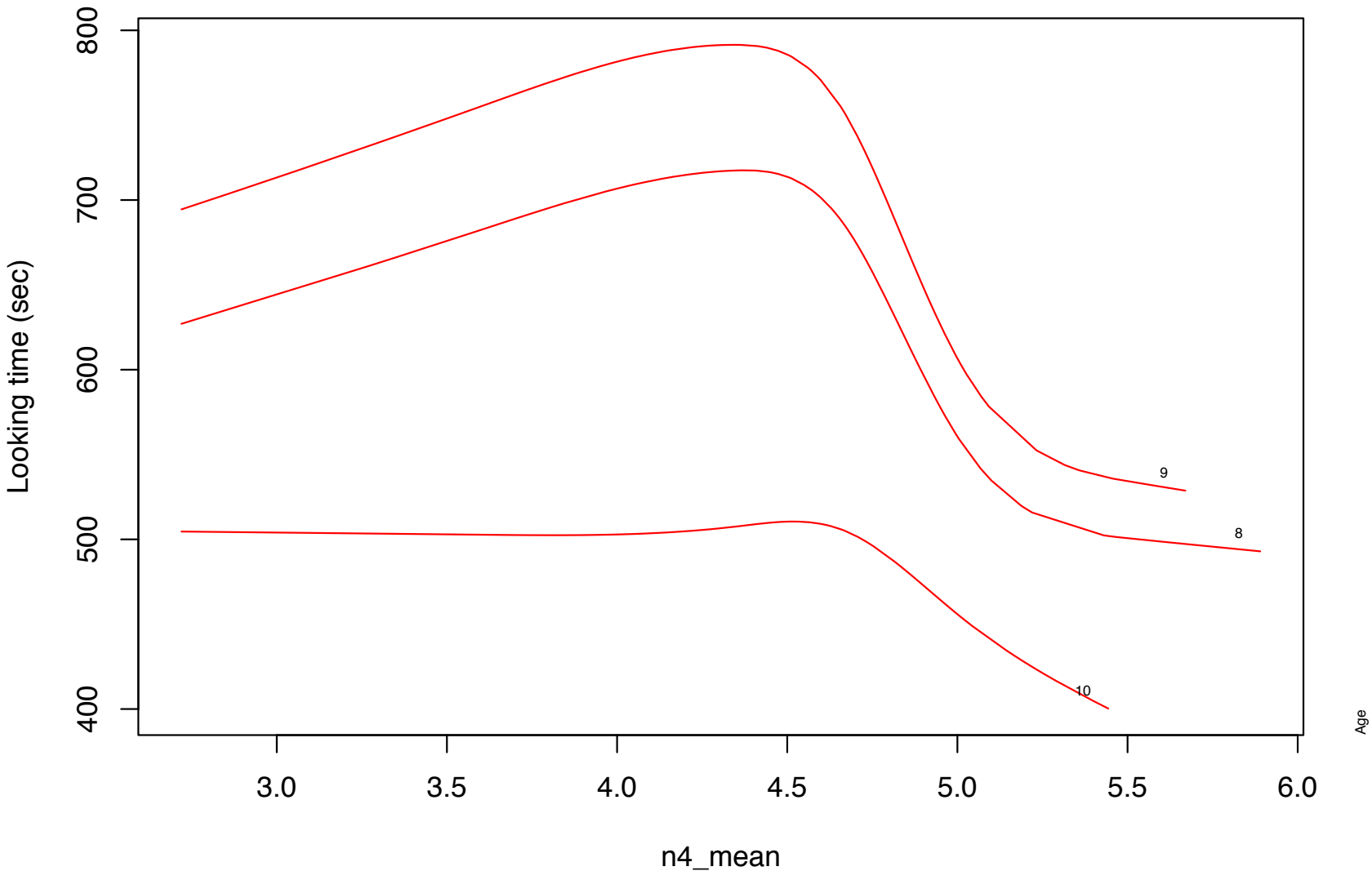
Size matters?



Size matters. But with caution.



Do we see age effects, already?



To wrap up

- 2grams more characteristics of words, thus good to **distinguish words from non–existing strings**; but also more variable across words, thus **ineffective to identify specific words**.
- Frequency effects **(which is statistical learning!)** in very young kids, in early measures of processing, and moderated by reading speeding, but (probably) not in a theoretically relevant way.
- nGram frequency seems to affect eye movements in children, even early on.
- Children seem to track better the stats of larger chunks (jumping to lexicality?).

- The logic behind the experiment seems to work
- The logistics behind the experiment seem to work

Down the line

- Check morphology
- A 'sliding window' analysis
- Word predictability in context (corpora, cloze task with kids)
- Explore other types of statistical regularities (e.g., transitional probabilities, long–distance relationships)
- Consider spaces, which may be critical for its perceptual salience
- Takes care of predictor correlation more seriously

A new approach to reading

- Scripts can be seen as fully–fledged visual systems
- They can be studied as such (without language)
- The way we learn to deal with them can be captured through statistical learning
- The way we learn to map them onto language can be captured through statistical learning

A new approach to reading

- Scripts can be seen as fully–fledged visual systems
- They can be studied as such (without) language
- The way we learn to deal with them can be captured through statistical learning
- The way we learn to map them onto language can be captured through statistical learning

Acknowledgments

- Valentina Parma (SISSA) and Simona Cerrato (SISSA Medialab).

Statistical Learning and Learning to Read

Davide Crepaldi, Valentina Pescuma, Maria Ktori,
Benedetta Cevoli, Francesca Franzon, Eleonora Lomi
[davide.crepaldi@sissa.it]
[http://lrlac.sissa.it]

International School for Advanced Studies (SISSA), Trieste

AIP, Bari, 21 September 2017