Analysing eye-movement data using R

Andrew Stewart
Division of Neuroscience and Experimental Psychology

Andrew.Stewart@manchester.ac.uk


@ajstewart_lang

https://github.com/ajstewartlang/RUM_talk_12-9-18

- We conduct our research in the context of Open Science, open data and analysis sharing…

- Since ~ 2017 we have been making our data available via GitHub and OSF repositories.

- And we (try to) follow the advice of Andrew Gelman…

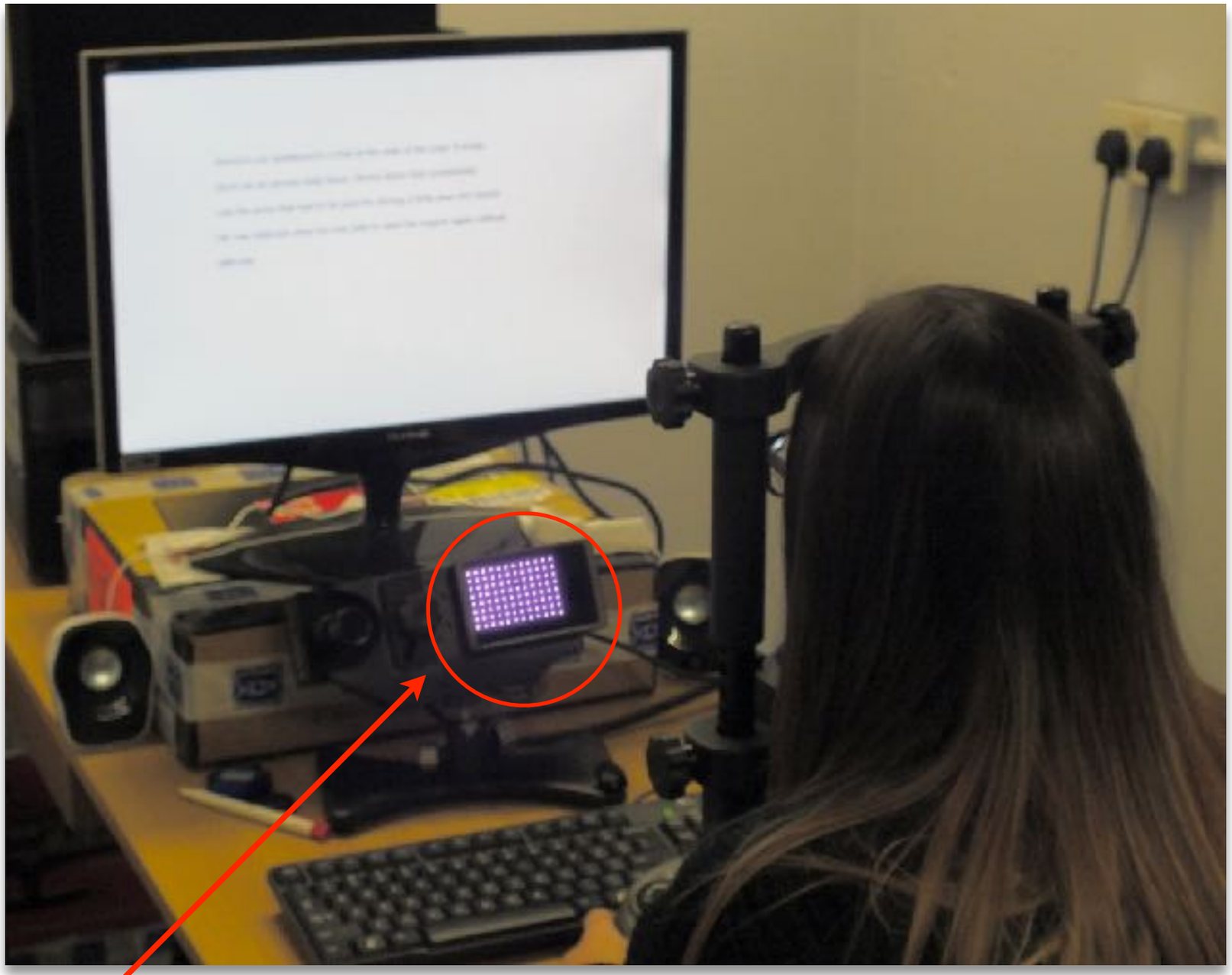
https://github.com/ajstewartlang/RUM_talk_12-9-18

http://www.stat.columbia.edu/~gelman/

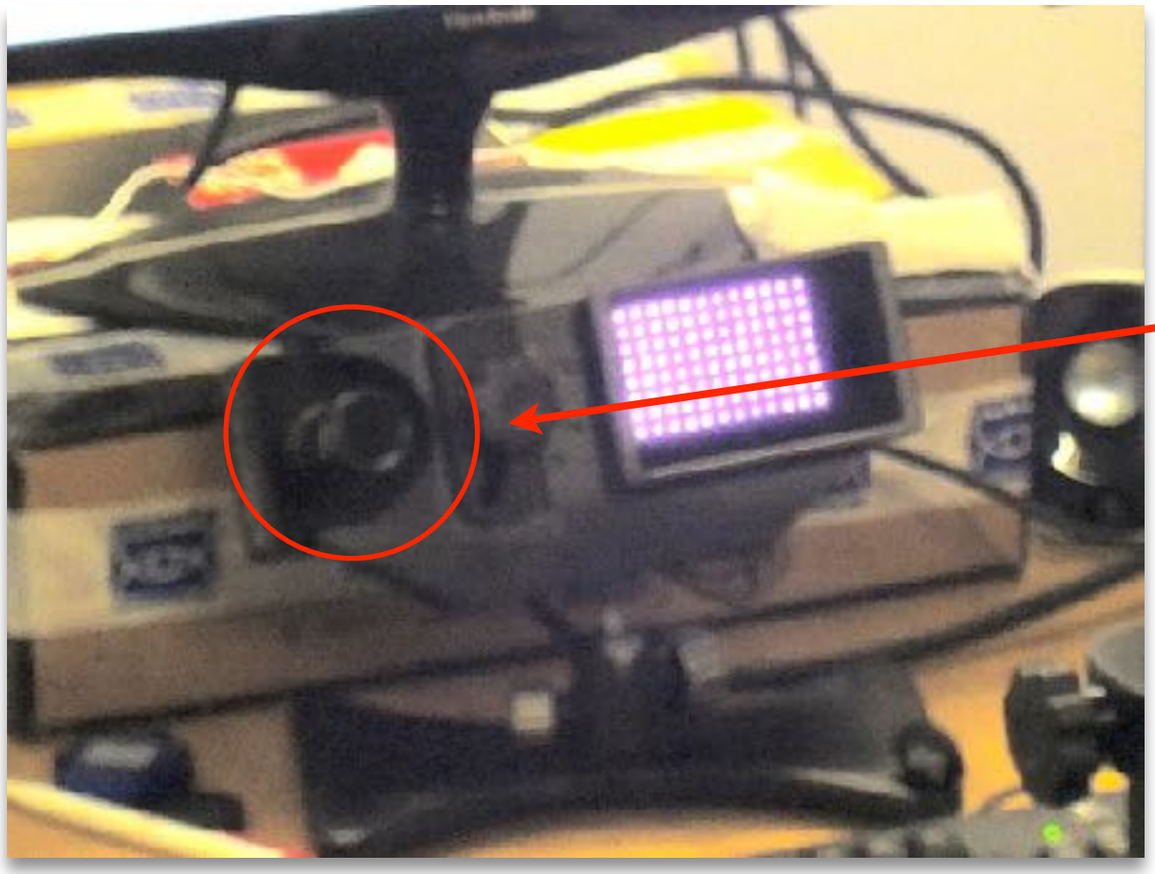
Andrew Gelman gives the following recommendations to researchers:

- Analyze all your data.
- Present all your comparisons.
- Make your data public.
- Put in the effort to take accurate measurements (low bias, low variance, and a large enough sample size).
- Do repeated-measures comparisons where possible.

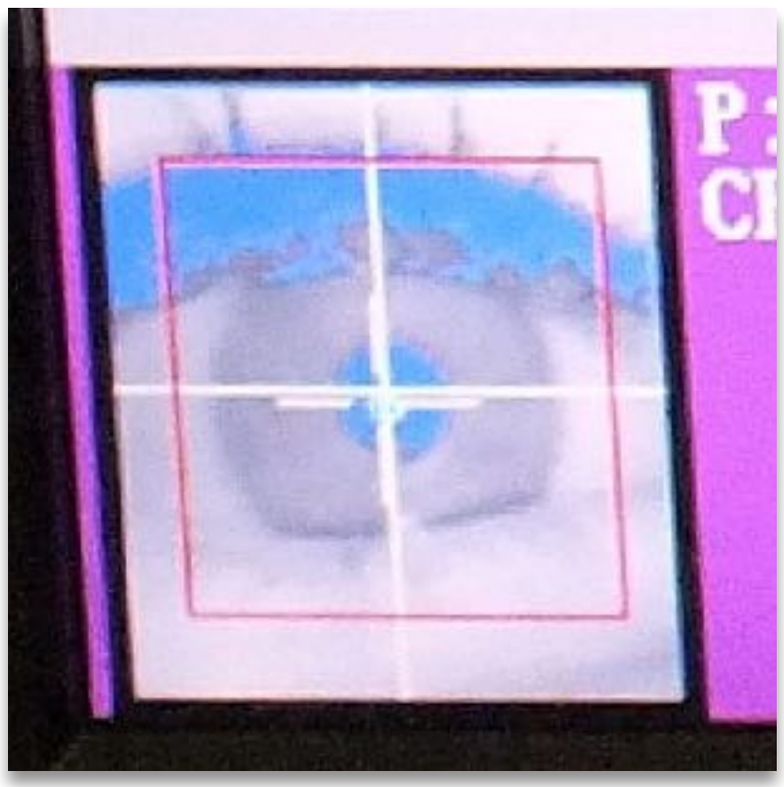
Eye Tracking

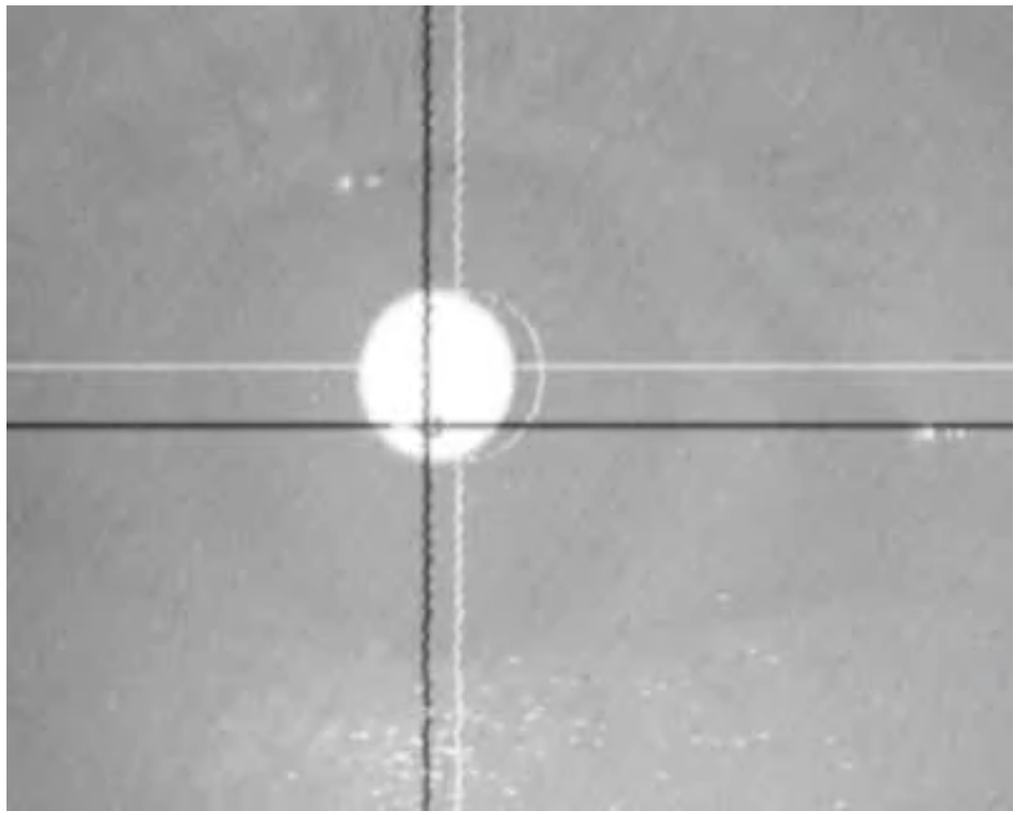


Infrared light is shone from the illuminator into the eye

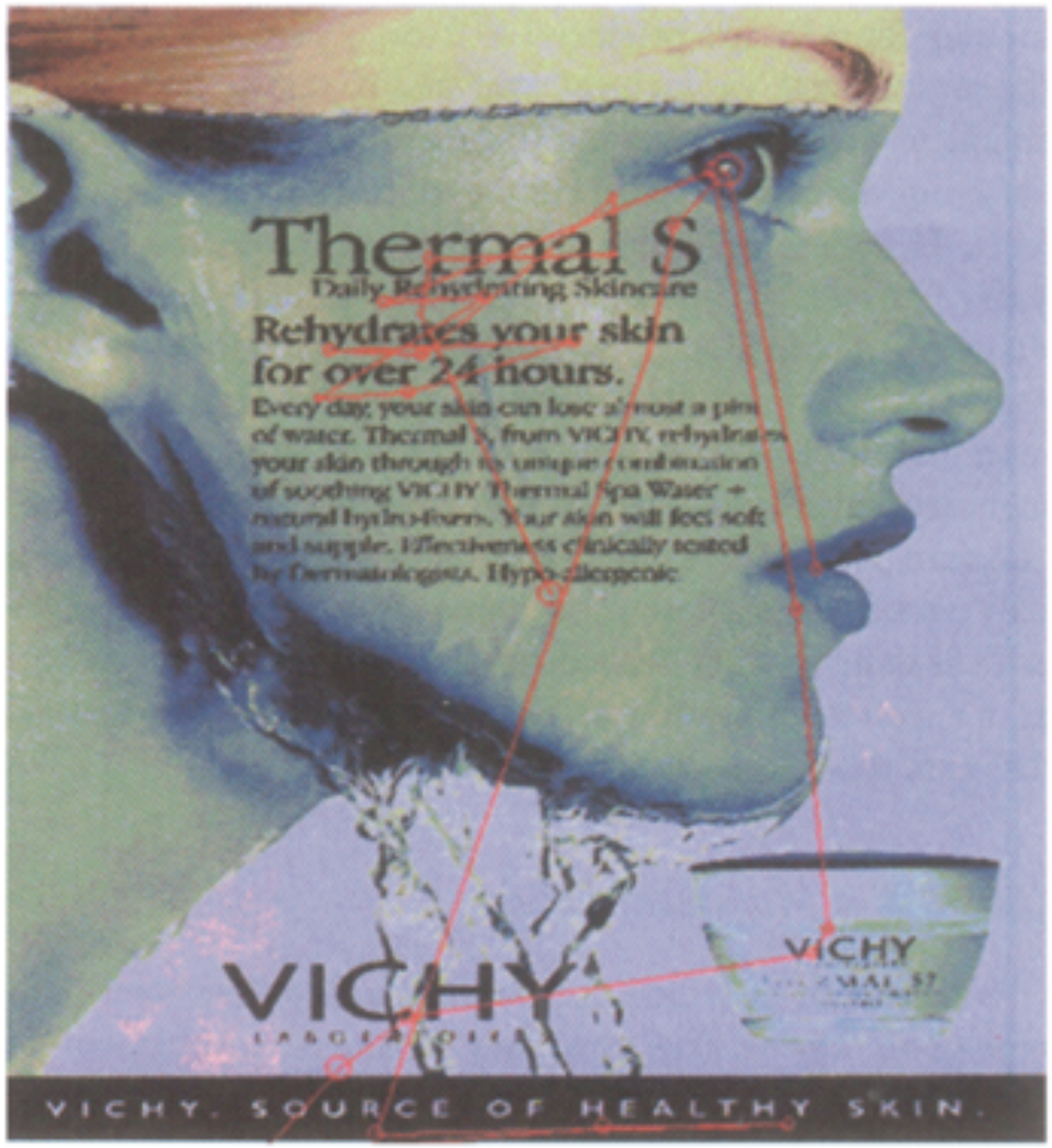


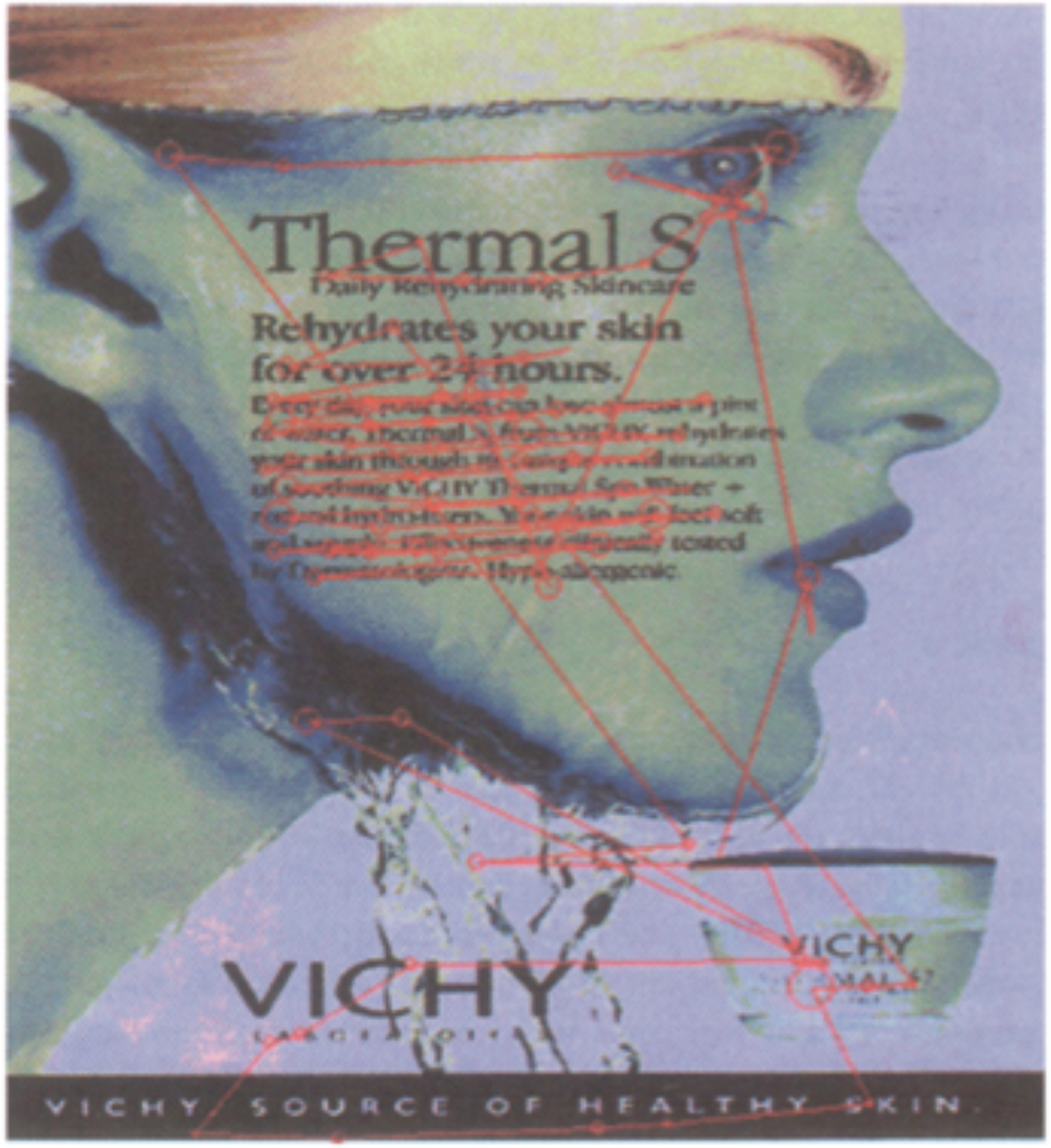
Reflections of the infrared light from the eye are detected by the camera and overlaid on the image of the eye





Two reflections result - by measuring how these reflections move relative to each other, it’s possible to calculate what the eye is looking at.





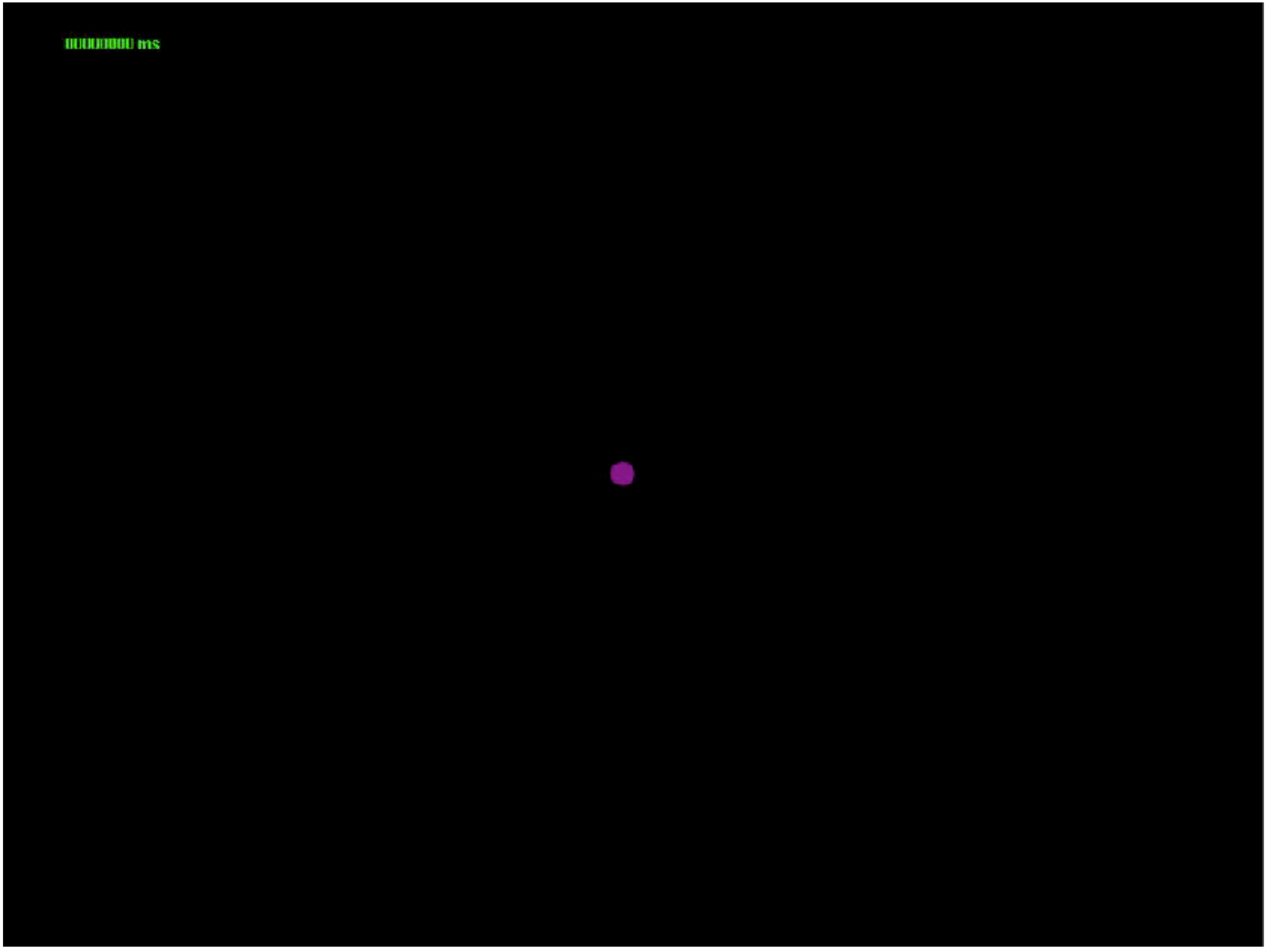
Rayner, K., Rotello, C., Stewart, A.J., Keir, J., & Duffy, S.A. (2001). Integrating text and pictorial information: eye movements when looking at print advertisements. Journal of Experimental Psychology: Applied, 7, 219–226.

An eye-tracking in reading primer

- Eye-movements during reading consist of fixations (about 250 msec. each) and saccades (where the eye jumps from one location to another).

- During reading 10-15% of all eye-movements are backwards (called regressions). They allow the reader to (re)look at previously read text.

- When fixating at a point in a word, you can see about 4 characters to the left and about 12-15 to the right of fixation.



Video courtesy of BUPsychTech

Reading measures map onto cognitive processes that are temporally distinct:

- First Pass/Gaze Duration - sum of all fixations within a region of text before the eye exits to the left or to the right.

- Regression Path - sum of all fixations within a region of text before the eye exits to the right (incl. re-reading of previous regions).

- Total Time - sum of all fixations within a region of text.

- Regressions In/Out - regressions into/out of a region of text.

Eye-Movements and Frequency Effects

```
The concerned steward calmed the child.

The concerned student calmed the child.
```

The only difference between the two sentences is ‘steward’ (a low frequency word) appears in the first example and ‘student’ (a high frequency word) appears in the second.

Increased fixation durations on ‘steward’ *vs.* ‘student’.

R Packages

We use mainly the following packages:

`Tidyverse` for data tidying, wrangling, and visualisation.

`lme4` for mixed effects models.

`lmerTest` for p-value estimates of fixed effect parameters.

`emmeans` for pairwise comparisons following interactions and main effects of factors with > 2 levels.

Data in long (tidy) format with usually a minimum of 8 observations per person per condition - always repeated measures.

Deviation contrast coding used when we have more than one factor.

Indirect Replies

- People don’t like giving other people negative information that could be face threatening (Brown & Levinson, 1987).

- Face management argued to motivate the use of indirect replies.

- Indirect replies typically violate relevance (Holtgraves, 1998) - this violation triggers a search for a possible negative meaning.

- Face management is arguably a complex social process - are people sensitive to face management needs when reading conversations between two interlocutors?

Negative Situation.

Roberta and Andy are friends. Roberta is taking introductory chemistry this semester and is struggling on her course. Andy asked “How are you doing in chemistry?” She replied “The exams are not fair.” Andy planned to take the same course the following year. He was hopeful the course would be interesting.

Positive Situation.

Roberta and Andy are friends. Roberta is taking introductory chemistry this semester and is excelling on her course. Andy asked “How are you doing in chemistry?” She replied “The exams are not fair.” Andy planned to take the same course the following year. He was hopeful the course would be interesting.

Neutral Situation.

Roberta and Andy are friends. Roberta is taking introductory chemistry this semester that she attends on Tuesday afternoons. Andy asked “How are you doing in chemistry?” She replied “The exams are not fair.” Andy planned to take the same course the following year. He was hopeful the course would be interesting.

We manipulated whether the context was Negative, Positive, or Neutral.

This gives us a 1 factor with 3-levels repeated measures design.

- Twenty four participants.
- Twenty four experimental vignettes.
- Twenty four filler vignettes.
- Eye movements recorded using Eyelink 1000.

Two key analysis regions - Region 3 is the critical region, and Region 4 the post-critical:

She replied |**“The exams are not fair.”**|critical

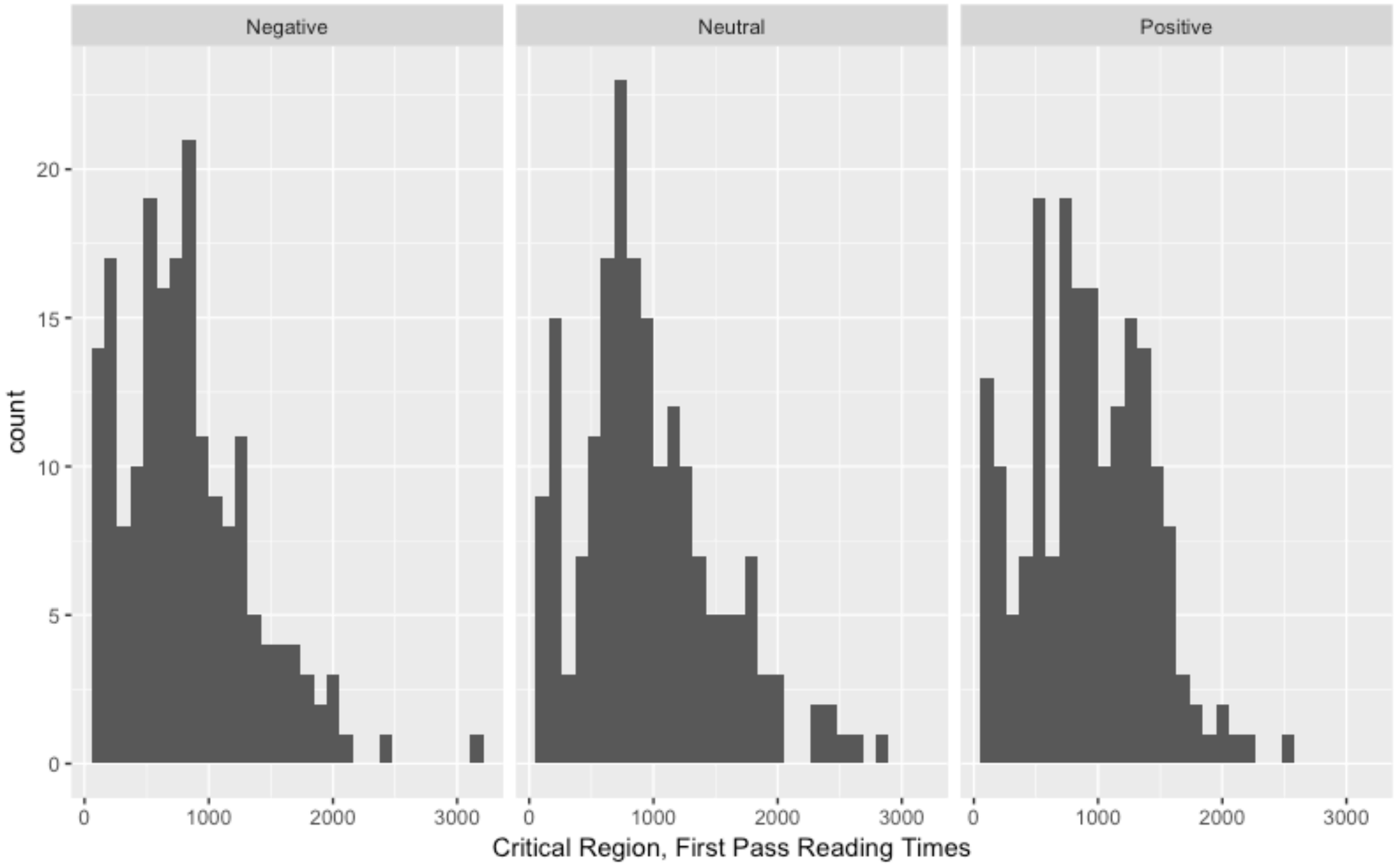
|**Andy planned to take the same course the following year.** |post-critical

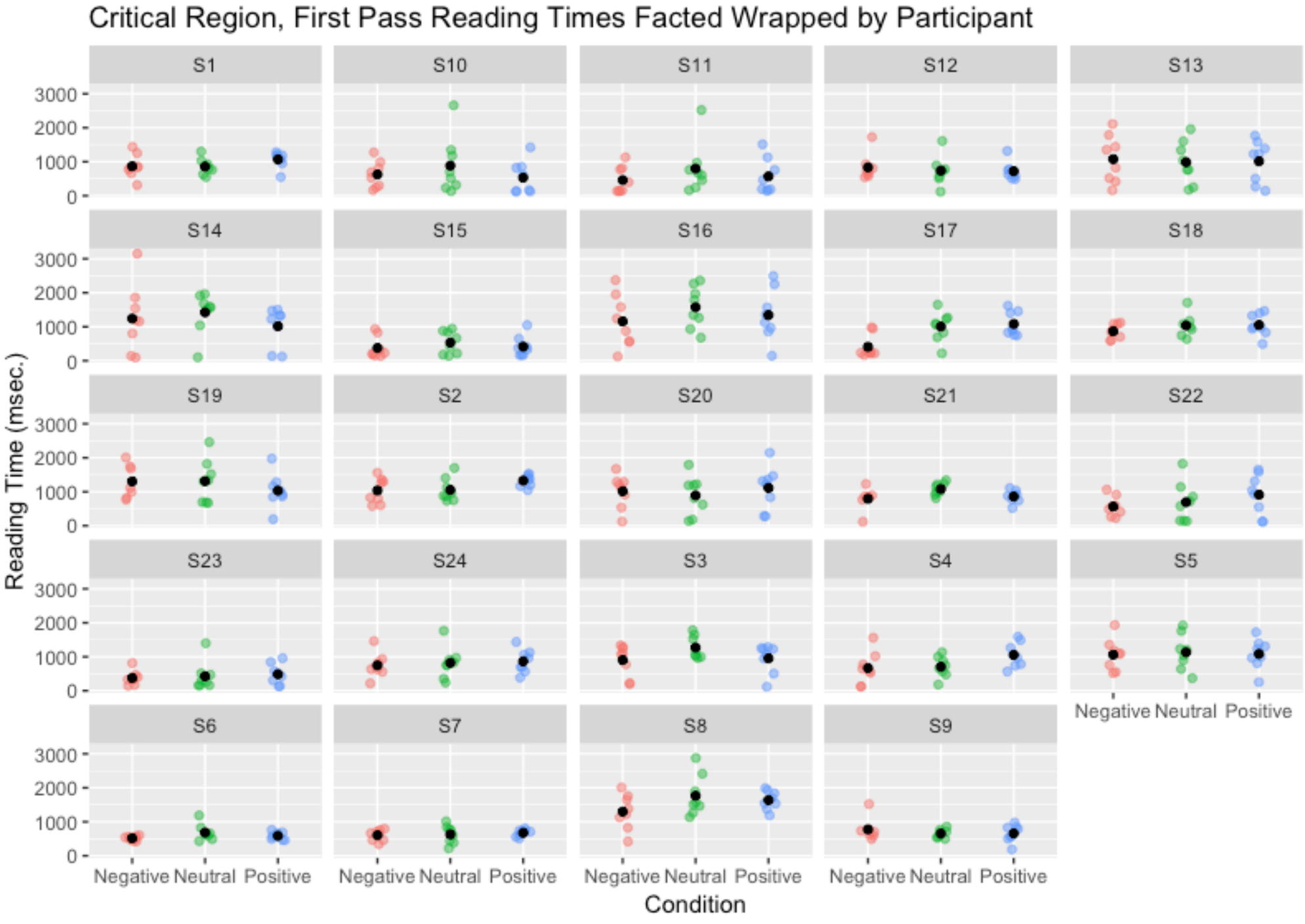
Our critical region is Region 3 of the text.

```
> data_R3
# A tibble: 572 x 7
   subj  item  cond       seq meas  reg       DV
   <fct> <fct> <fct>    <int> <chr> <chr> <int>
 1 S1    I1    Neutral     60 FP    R3      867
 2 S1    I2    Positive    44 FP    R3     1061
 3 S1    I3    Negative    30 FP    R3      771
 4 S1    I4    Neutral     15 FP    R3      626
 5 S1    I5    Positive    55 FP    R3     1283
 6 S1    I6    Negative    18 FP    R3      846
 7 S1    I7    Neutral     32 FP    R3      547
 8 S1    I8    Positive    70 FP    R3     1135
 9 S1    I9    Negative    37 FP    R3     1254
10 S1    I10   Neutral      1 FP    R3      926
# ... with 562 more rows
```

Our `subj` and `item` variables are our random effects, our `cond` variable our fixed effect and `DV` our outcome measure.

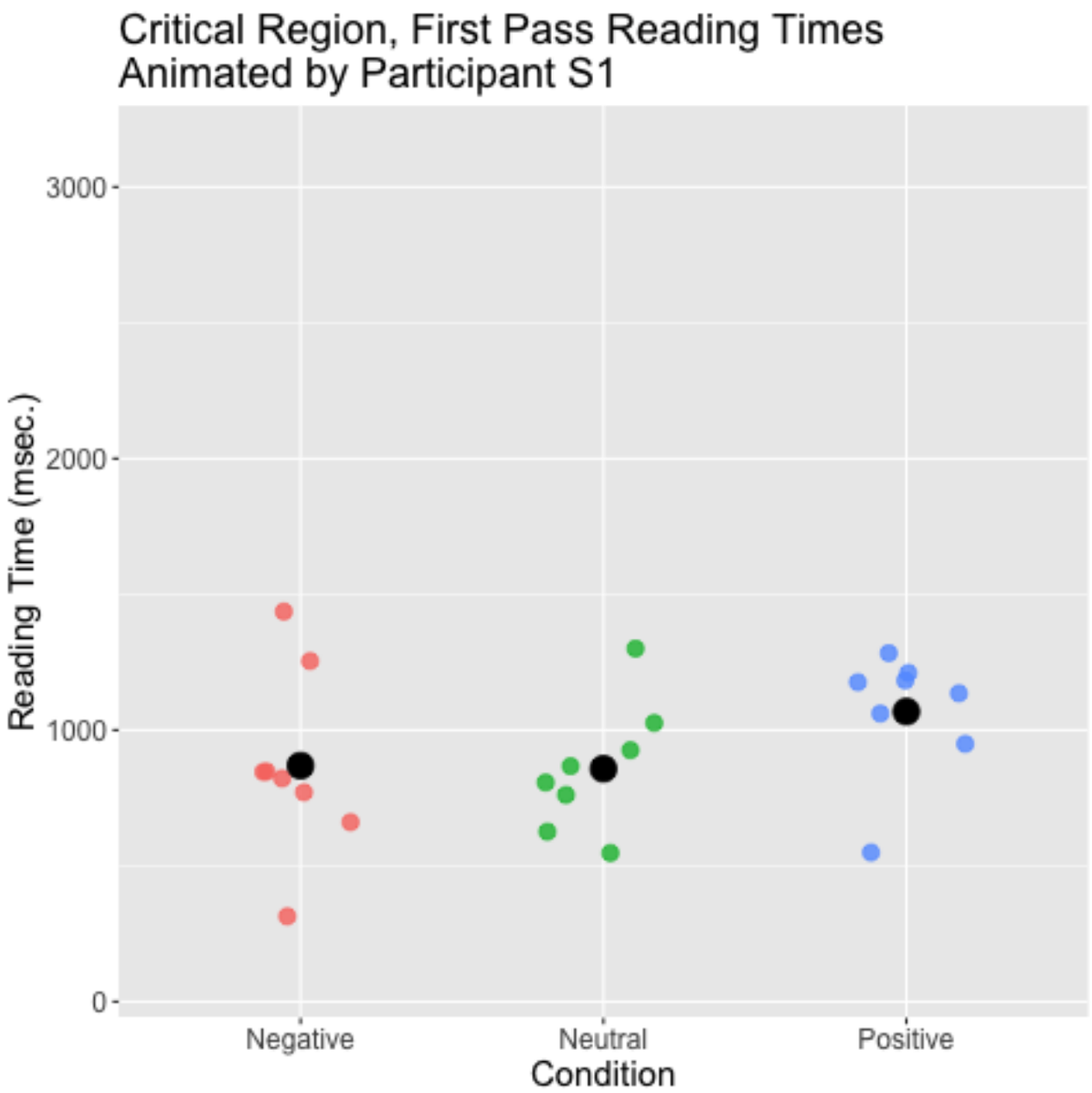
Exploring the Data

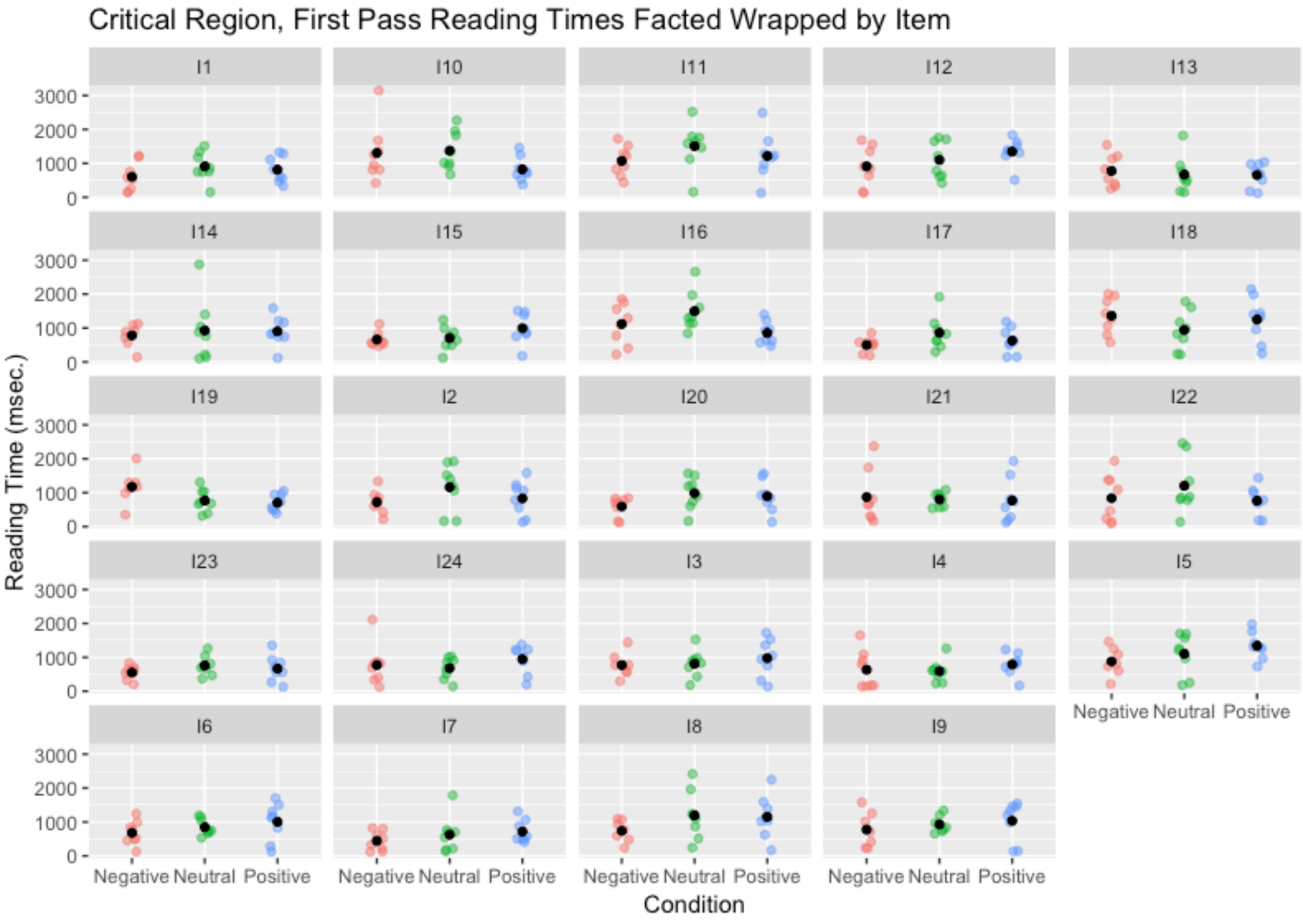




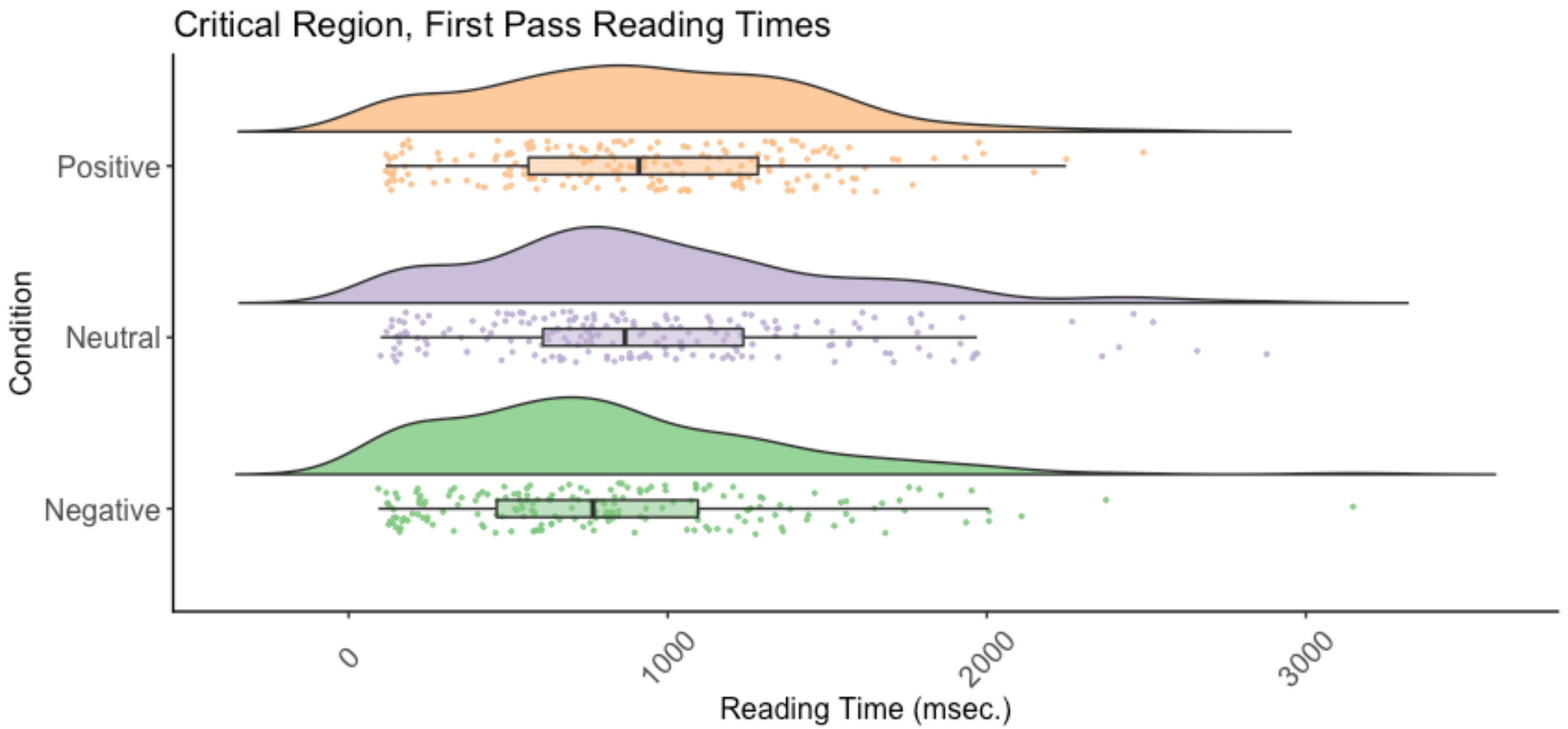

```
> sum (data_agg[data_agg$cond=="Negative",]$mean < data_agg[data_agg$cond=="Positive",]$mean)
[1] 18
> sum (data_agg[data_agg$cond=="Negative",]$mean < data_agg[data_agg$cond=="Neutral",]$mean)
[1] 19
```

Using the new version of gganimate:






```
> sum (data_agg[data_agg$cond=="Negative",]$mean < data_agg[data_agg$cond=="Positive",]$mean)
[1] 17
> sum (data_agg[data_agg$cond=="Negative",]$mean < data_agg[data_agg$cond=="Neutral",]$mean)
[1] 18
```

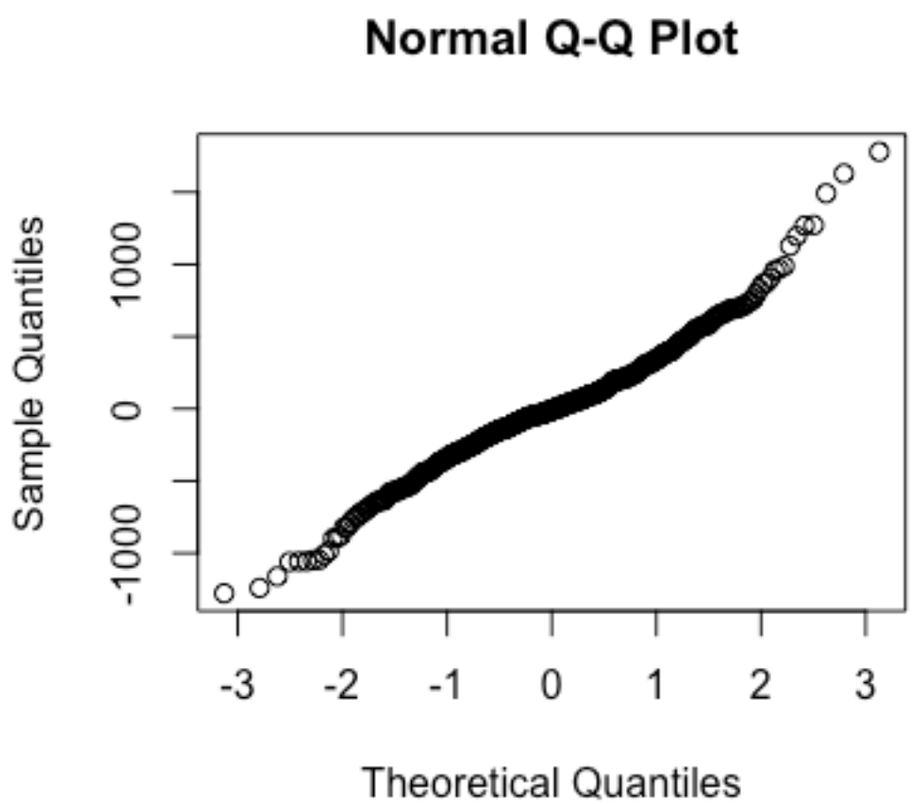


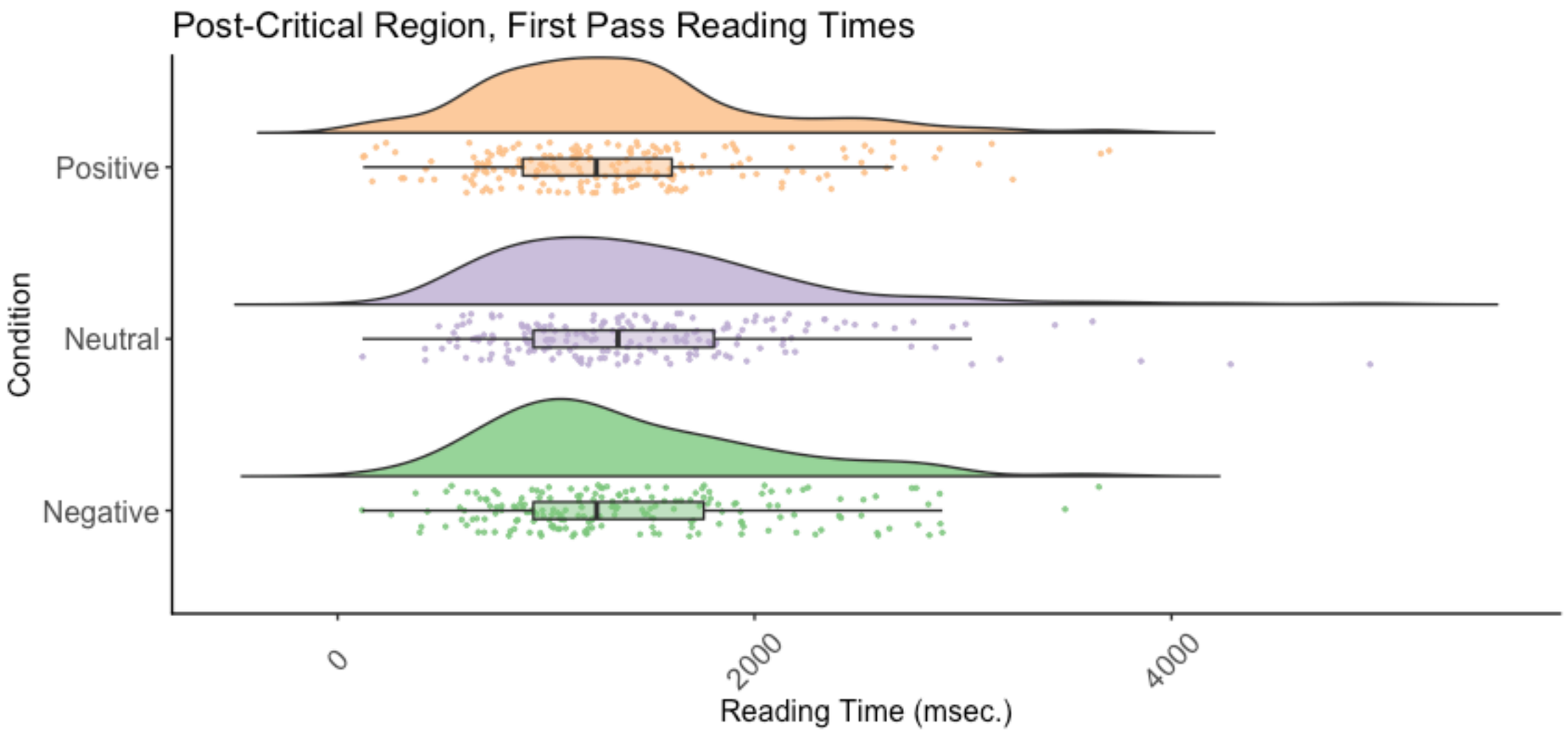

```
> data_R3 %>% group_by(cond) %>% summarise (mean=mean(DV), sd=sd(DV))
# A tibble: 3 x 3
  cond      mean    sd
  <fct>    <dbl> <dbl>
1 Negative  814.  513.
2 Neutral   957.  559.
3 Positive  920.  489.
```

```
> model <- lmer(DV ~ cond + (1+cond|subj) + (1+cond|item), data_R3, REML=TRUE)

> summary (model)

Fixed effects:
             Estimate Std. Error     df t value Pr(>|t|)
(Intercept)    815.42      71.57  32.26  11.393 7.61e-13 ***
condNeutral    142.03      43.66 143.18   3.253  0.00142 **
condPositive   104.46      44.81  29.43   2.331  0.02679 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residuals looks fairly ok.
We conclude that the reply is read most quickly in the Negative context, but more slowly in the Neutral and Positive contexts.

But what if they didn’t look ok?





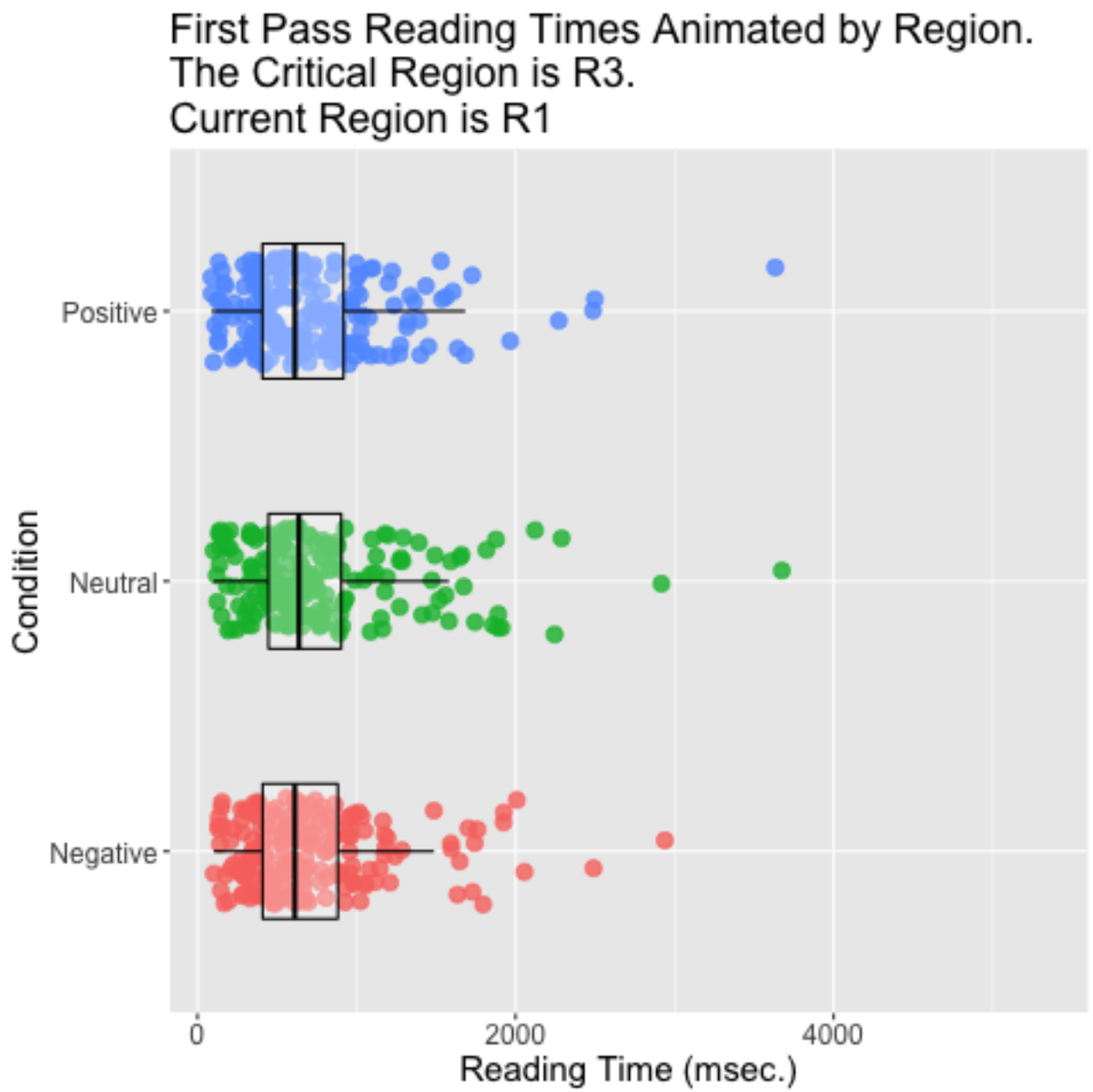

```
> data_R4 %>% group_by(cond) %>% summarise (mean=mean(DV), sd=sd(DV))
# A tibble: 3 x 3
  cond      mean    sd
  <fct>    <dbl> <dbl>
1 Negative 1398.  652.
2 Neutral  1462.  727.
3 Positive 1346.  658.
```

```
> model <- lmer(DV ~ cond + (1|subj) + (1|item), data_R4, REML=TRUE)

> summary (model)

Fixed effects:
             Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   1395.09      92.77   45.32  15.039   <2e-16 ***
condNeutral     66.92      56.33  522.13   1.188    0.235
condPositive   -49.05      56.26  522.14  -0.872    0.384
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> anova (model)

Type III Analysis of Variance Table with Satterthwaite's method
      Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
cond 1297047  648524     2 522.1  2.1585 0.1165
```

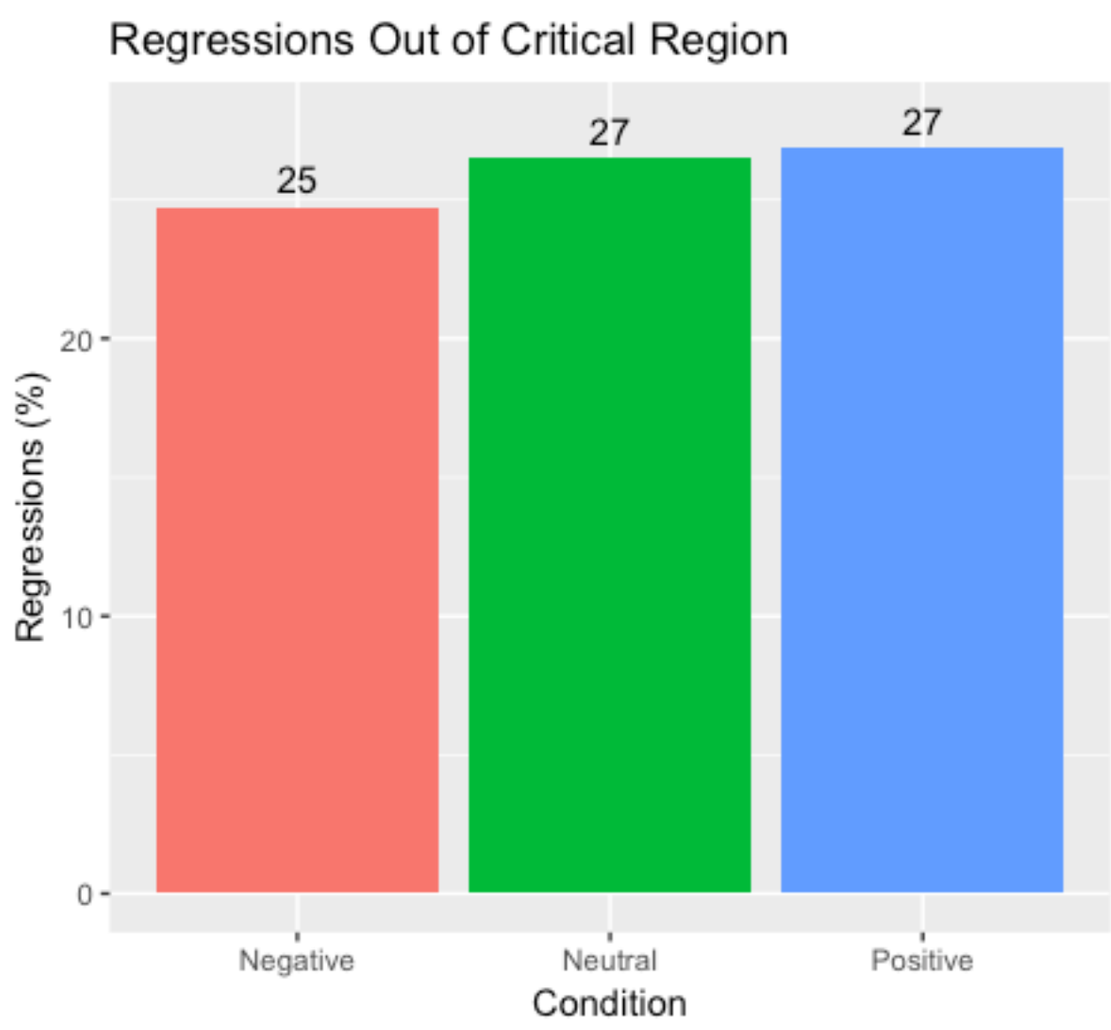
The effect that we had observed on the critical region and has dissipated by the time we get to the post-critical region.

We can view the first pass reading times across the five regions of the text - remember, Region 3 (R3) is our key one.



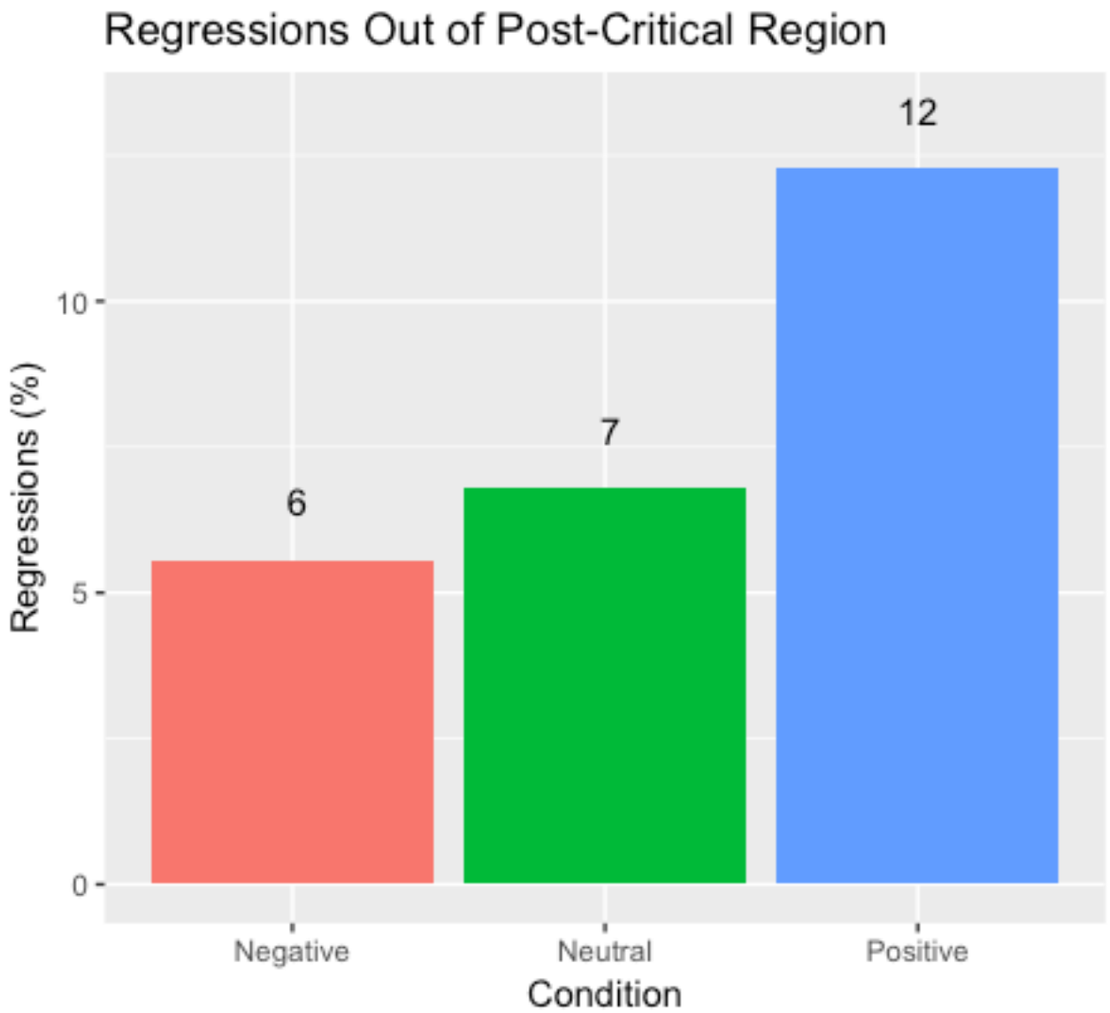
```
> data_R3
# A tibble: 553 x 7
   subj  item  cond       seq meas  reg       DV
   <chr> <chr> <fct>    <int> <chr> <chr> <dbl>
 1 S1    I2    Positive    44 RO    R3        0
 2 S1    I3    Negative    30 RO    R3        0
 3 S1    I4    Neutral     15 RO    R3        0
 4 S1    I5    Positive    55 RO    R3        0
 5 S1    I6    Negative    18 RO    R3        0
 6 S1    I7    Neutral     32 RO    R3        1
 7 S1    I8    Positive    70 RO    R3        0
 8 S1    I9    Negative    37 RO    R3        0
 9 S1    I10   Neutral      1 RO    R3        0
10 S1    I11   Positive    57 RO    R3        0
# ... with 543 more rows
```

Our measure is binomial as people either did (1) or did not (0) make a regression eye movement.



```
> model <- glmer (DV ~ cond + (1|subj) + (1|item), data_R3, family=binomial)
> summary (model)

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.3095     0.2647  -4.947 7.55e-07 ***
condNeutral    0.1093     0.2539   0.431    0.667
condPositive   0.1162     0.2515   0.462    0.644
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Not much going on here - note, we had to drop slopes from our random effects terms in order to find a model that converged.



Again, our measure is binomial as people either did (1) or did not (0) make a regression eye movement.

```
> model <- glmer (DV ~ cond + (1|subj) + (1|item), data_R4, family=binomial)

> summary (model)

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.5034     0.5094  -6.877 6.12e-12 ***
condNeutral    0.2484     0.5076   0.489   0.6246
condPositive   1.0495     0.4607   2.278   0.0227 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

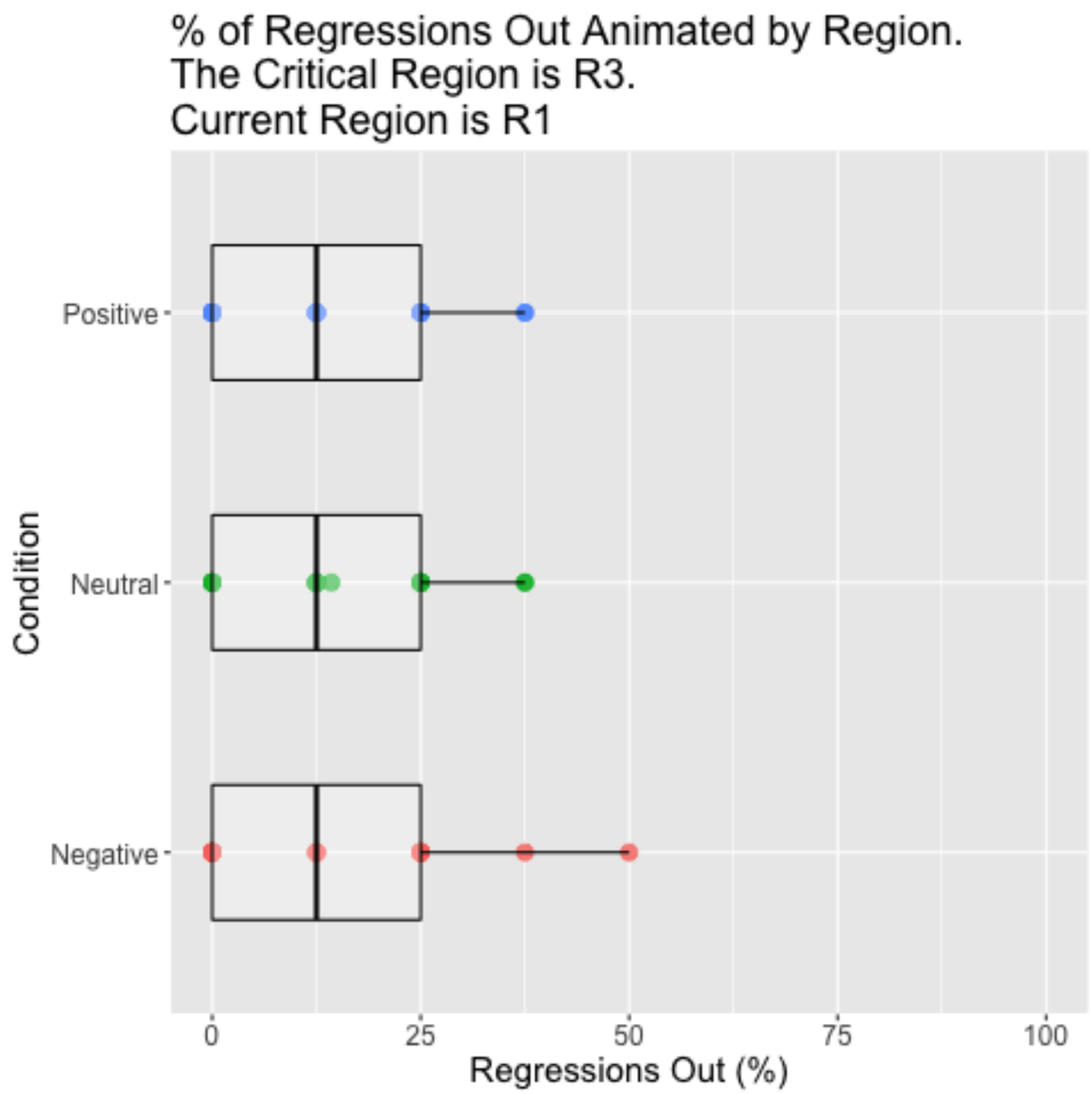
Looks like we have more regression in the Positive condition than in the Negative. We can also request the model in ANOVA format which suggest an effect of condition (cond).

```
> anova (model)

Analysis of Variance Table
     Df Sum Sq Mean Sq F value
cond  2 6.2806  3.1403  3.1403
```

```
> emmeans (model, pairwise~cond, adjust="none", type="response")
$emmeans
 cond           prob         SE  df  asymp.LCL  asymp.UCL
 Negative 0.02921439 0.01444844 Inf 0.01096581 0.07551245
 Neutral  0.03714549 0.01697488 Inf 0.01498975 0.08908695
 Positive 0.07914715 0.02932761 Inf 0.03759123 0.15904989

Confidence level used: 0.95
Intervals are back-transformed from the logit scale

$contrasts
 contrast            odds.ratio        SE  df z.ratio p.value
 Negative / Neutral   0.7800601 0.3959661 Inf  -0.489  0.6246
 Negative / Positive  0.3501293 0.1612980 Inf  -2.278  0.0227
 Neutral / Positive   0.4488491 0.1940912 Inf  -1.853  0.0640

Tests are performed on the log odds ratio scale
```

The p-value of our Negative vs. Positive comparison is the same p-value as our parameter estimate in the model, while the Neutral vs. Positive comparison is 0.064 without any correction for familywise error.

We can view the % of Regression out across the regions of the text - remember, Region 3 (R3) is our critical one and Region 4 (R4) our post-critical one.



Overall, we conclude there is an initial slowdown when reading a face-saving reply in contexts that do not support a possible negative meaning of the reply.

This disruption is short-lasting in the Neutral condition, but persists in the Positive condition as reflected in more backwards eye-movements in reading of the post-critical region.

She replied |**“The exams are not fair.”**|critical

|**Andy planned to take the same course the following year.** |post-critical

Pupillometry

- Changes in pupil size driven by locus coeruleus (reflected in BOLD response) and linked to autonomic arousal and alertness.
- Cognitive decline in animal models found where there are declines in neural function via locus coeruleus projections to prefrontal cortex.
- Widely used in exp. psychology to examine attention and memory (e.g., Beatty, 1982; Kafkas & Montaldi, 2015; Kahneman, 1973).

Using Pupillometry to Measure Effort

- Sustained listening results in self-reported fatigue in those with hearing loss.
- Can we measure fatigue objectively?
- If so, clear clinical application in treating hearing loss.

McGarrigle et al., (2017a). *Psychophysiology.*

- 24 participants took part in experiment (0.8 power).

- Listened to 96 short narratives under two levels of background noise:

  - “Easy” listening with SNR of +15 dB

  - “Hard” listening with SNR of -8 dB

- Task was to respond to a picture and to determine whether it depicted an object that appeared in the preceding narrative. For the following example, the picture was that of a pigeon…

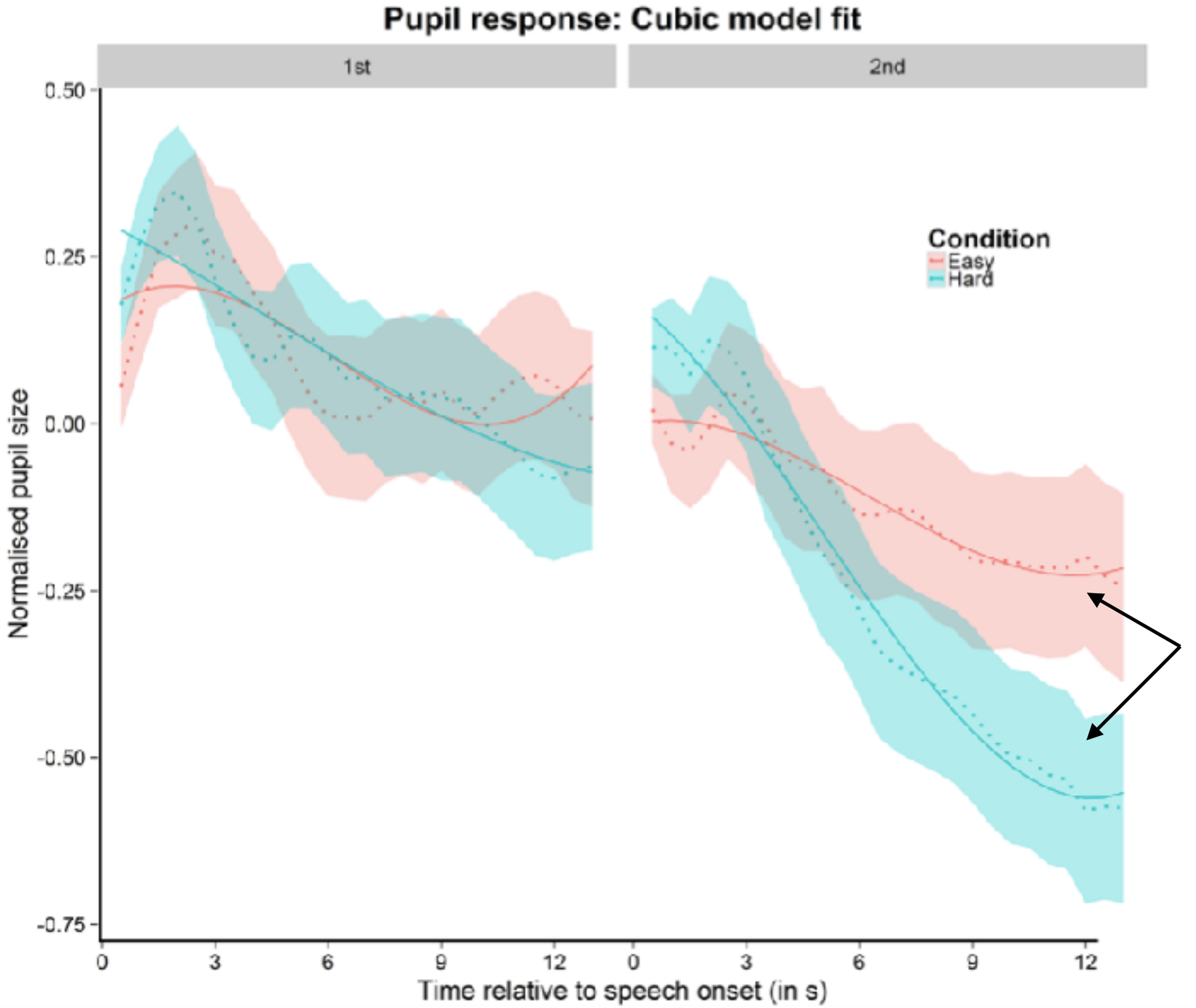
McGarrigle et al., (2017a). *Psychophysiology.*

Easy listening condition

McGarrigle et al., (2017a). *Psychophysiology.*

Hard listening condition

McGarrigle et al., (2017a). *Psychophysiology.*



GCA using LMM with crossed random effects of subjects and items.

Reduced level of arousal in 2nd half of experiment - more pronounced in “hard” listening condition.

McGarrigle et al., (2017a). *Psychophysiology.*

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Chapman and Hall / CRC.

Summary

- Eye movements in terms of fixation durations and regressions reveal moment-by-moment processing (and associated disruption).

- Pupillometry can provide an objective measure of fatigue that reflects a reduction in arousal in a sustained listening task.

- The meaning behind our data was uncovered with the help of the wonderfulness of R (and associated packages)!

Andrew.Stewart@manchester.ac.uk

@ajstewart_lang

https://github.com/ajstewartlang/RUM_talk_12-9-18