Results

Table of contents

Cleaning 1
Baseline model: frequency and strokes 2
Visual complexity penalty 3


```
library(yaml)
library(here)
```

```
here() starts at /root/repo
```

```
fmt3 <- function(x) sprintf("%.3f", x)
fmt6 <- function(x) sprintf("%.6f", x)

cleaning <- yaml::read_yaml(here("outputs", "results", "cleaning.yml"))
base <- yaml::read_yaml(here("outputs", "results", "base_lm.yml"))
complexity <- yaml::read_yaml(here("outputs", "results", "visual_complexity_penalty.yml"))
```

Cleaning

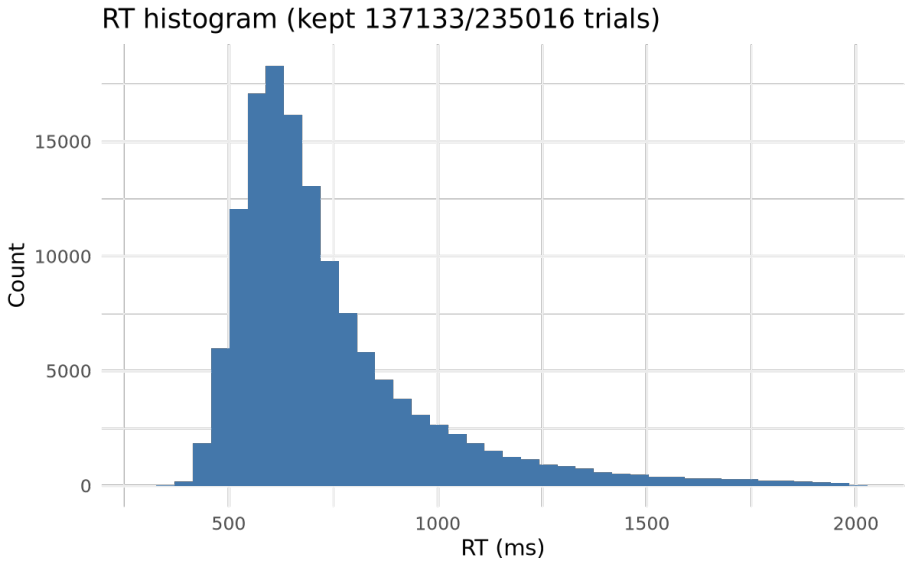
The pipeline kept 137133 of 235016 trials (dropped 97883). Settings: correct-only = TRUE, RT range = 200–2000 ms.

```
data.frame(
  setting = c(
    "correct_only",
    "rt_min_ms",
    "rt_max_ms",
    "total_trials",
    "kept_trials",
    "dropped_trials"
  ),
  value = c(
    as.character(cleaning$trimming$correct_only),
    cleaning$trimming$rt_min_ms,
    cleaning$trimming$rt_max_ms,
    cleaning$counts$total_trials,
```

```
    cleaning$counts$kept_trials,
    cleaning$counts$dropped_trials
  )
)
```

```
         setting  value
1   correct_only   TRUE
2      rt_min_ms    200
3      rt_max_ms   2000
4   total_trials 235016
5    kept_trials 137133
6 dropped_trials  97883
```

```
knitr::include_graphics(here("outputs", "figures", "rt_hist.png"))
```



Baseline model: frequency and strokes

```
data.frame(
  term = c("intercept", "log_freq", "strokes"),
  estimate = c(
    fmt6(as.numeric(base$coefficients$intercept)),
    fmt6(as.numeric(base$coefficients$log_freq)),
    fmt6(as.numeric(base$coefficients$strokes))
  )
)
```

```
       term  estimate
```

```
1 intercept  6.452355
2  log_freq -0.070823
3   strokes  0.013355
```

R^2 0.434; adjusted R^2 0.433; residual sigma 0.099. AIC -6851.160, BIC -6826.134.

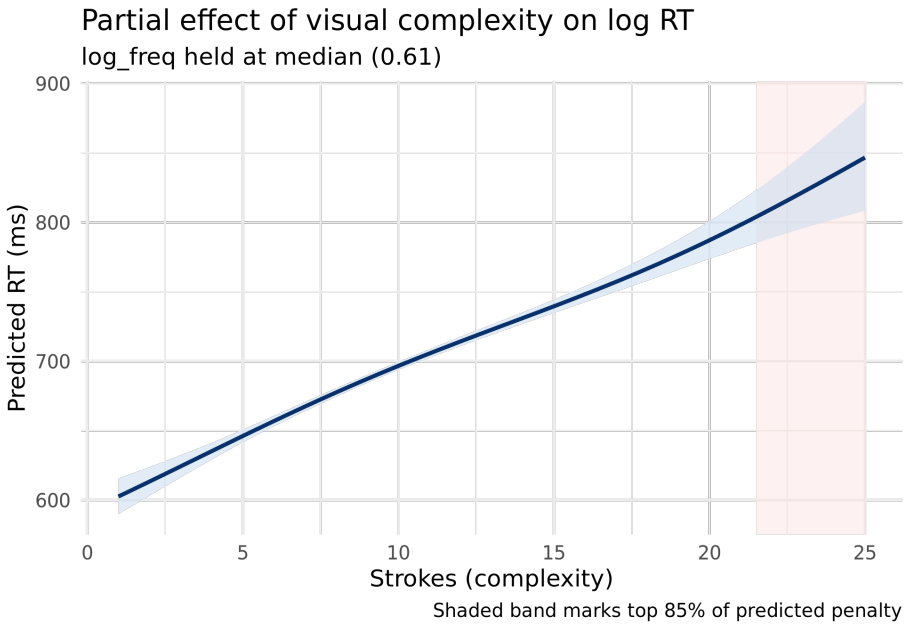
Visual complexity penalty

The partial effect of strokes remains reliable after holding frequency at its median (0.605111). The smooth term uses 2.656 effective degrees of freedom (F = 248.842, p = 0.000000). The predicted range from the least to most complex characters implies a 0.340 increase in log RT (about 244.060 ms). The strongest penalty lies between 21.5 and 25 strokes (top 85% of the effect curve).

```
data.frame(
  metric = c("edf (strokes)", "F statistic", "p-value", "log RT span", "RT span (ms)", "pena
  value = c(
    fmt3(as.numeric(complexity$edf_strokes)),
    fmt3(as.numeric(complexity$f_strokes)),
    fmt6(as.numeric(complexity$p_strokes)),
    fmt6(as.numeric(complexity$penalty$log_rt)),
    fmt3(as.numeric(complexity$penalty$rt_ms)),
    complexity$penalty$range_strokes$min,
    complexity$penalty$range_strokes$max
  )
)
```

```
               metric    value
1       edf (strokes)    2.656
2         F statistic  248.842
3             p-value 0.000000
4         log RT span 0.340020
5        RT span (ms)  244.060
6 penalty strokes min     21.5
7 penalty strokes max       25
```

```
knitr::include_graphics(here("outputs", "figures", "visual_complexity_penalty.png"))
```


Partial effect of visual complexity on log RT
log_freq held at median (0.61)
Predicted RT (ms)
900
800
700
600
0
5
10
15
20
25
Strokes (complexity)
Shaded band marks top 85% of predicted penalty