Supplemental File 3 - Data Analysis Report

Dr. F.J. Rodenburg

Contents

0.1 Methods

Statistical analyses were conducted in R, version 4.2.1, using the RStudio IDE.[1,2] Mixed effects models were fitted to account for the repeated measures of individuals using the glmmTMB package.[3] Model diagnostics were conducted with the DHARMa package,[4] and marginal effects plots of the resulting models were generated using the package sjPlot.[5]

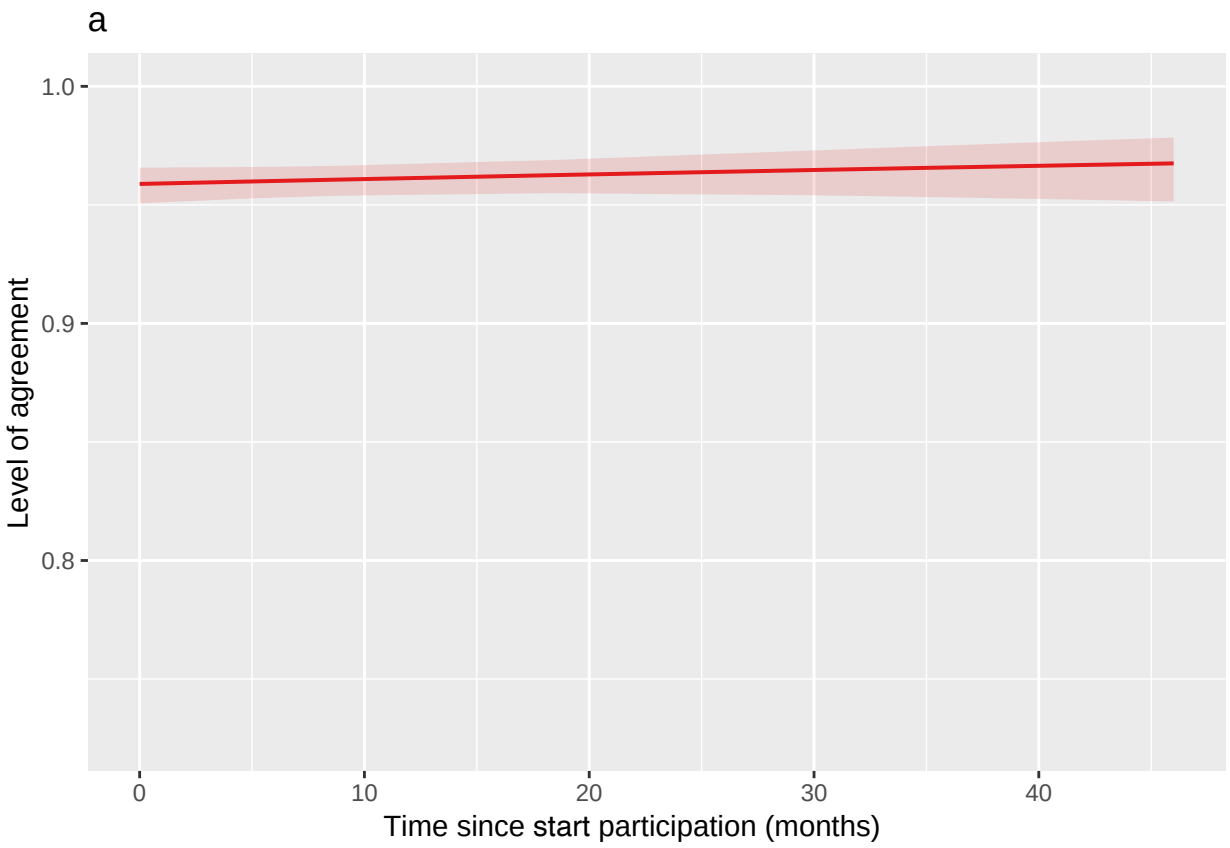
For the identification of question-clusters, conditional independence networks were constructed by penalized inverse covariance estimation, using the ridge penalty, as implemented in the rags2ridges package.[6]

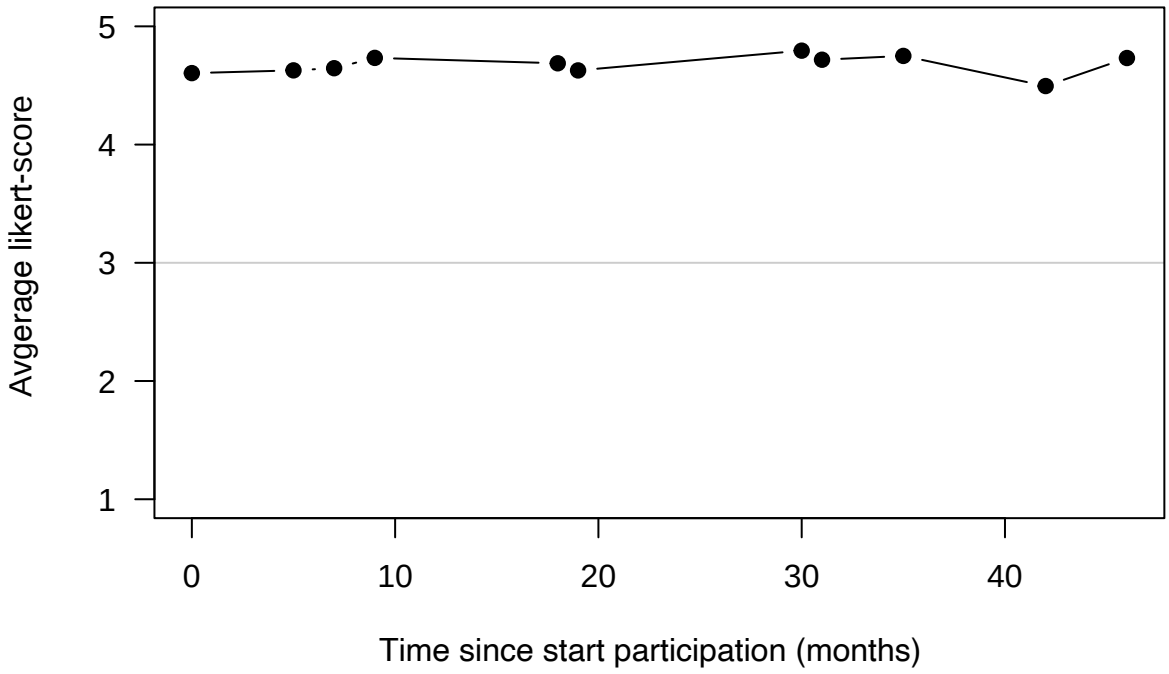
1 Motivation

1.1 Cluster *Environmental*

```
##  Family: binomial  ( logit )
## Formula:
## cbind(summed, summedcomplement) ~ time_since_start + (time_since_start |
##     Subject)
## Data: Mo3
##
##      AIC      BIC   logLik deviance df.resid
##   2519.9   2543.1  -1254.9   2509.9      759
##
## Random effects:
##
## Conditional model:
##  Groups  Name             Variance Std.Dev. Corr
##  Subject (Intercept)      1.835455 1.35479
##          time_since_start 0.002029 0.04505  -0.54
## Number of obs: 764, groups:  Subject, 400
##
## Conditional model:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      3.148016   0.096868   32.50   <2e-16 ***
## time_since_start 0.005339   0.005272    1.01    0.311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```

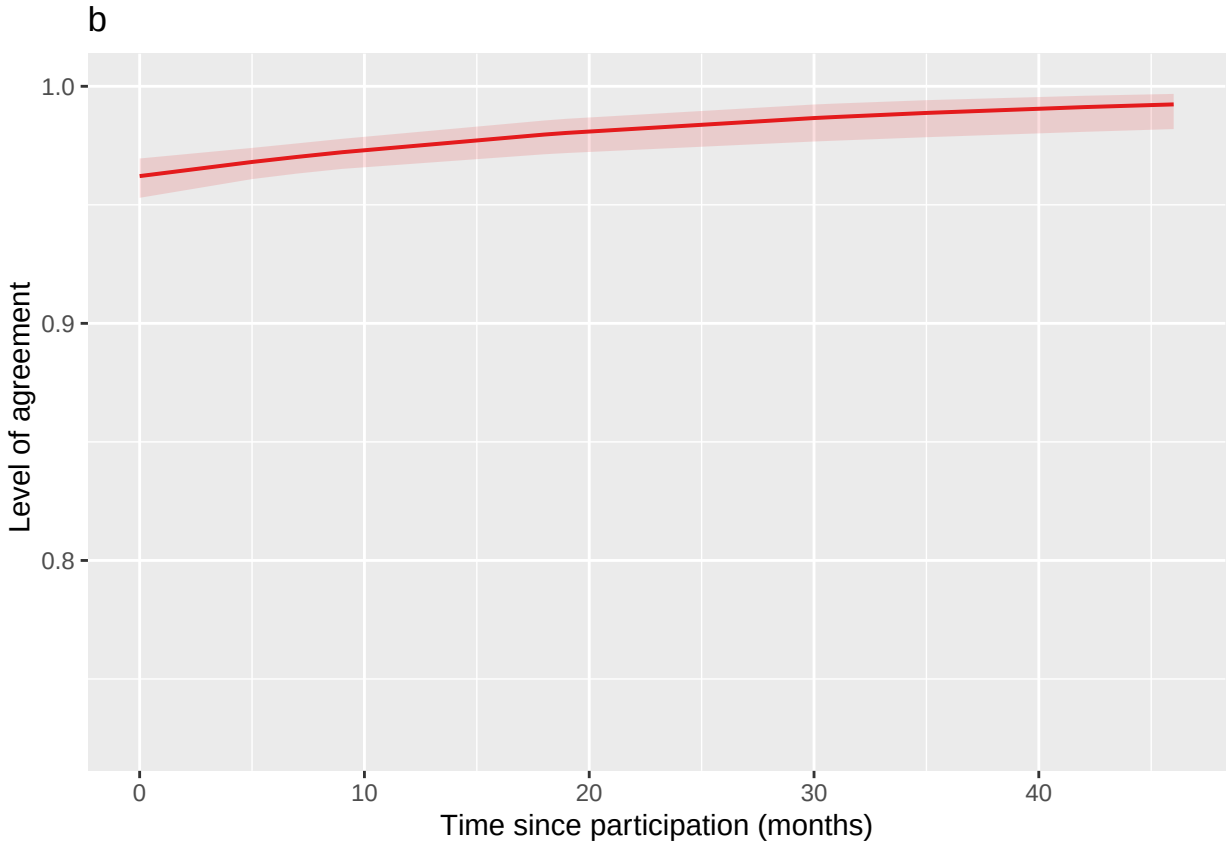

a
1.0
0.9
0.8
Level of agreement
0
10
20
30
40
Time since start participation (months)

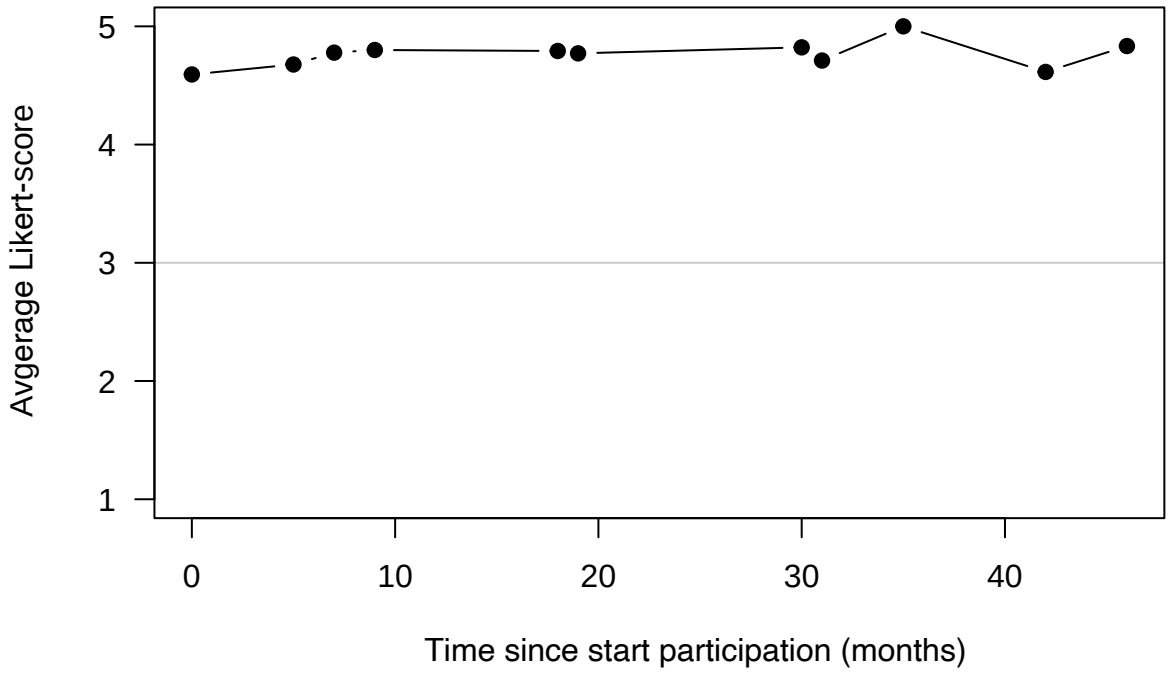


1.2 Cluster *Project action*

```
##  Family: binomial  ( logit )
## Formula:
## cbind(summed, summedcomplement) ~ time_since_start + (time_since_start |
##     Subject)
## Data: Mo2
##
##      AIC      BIC   logLik deviance df.resid
##   2061.4   2084.5  -1025.7   2051.4      752
##
## Random effects:
##
## Conditional model:
##  Groups  Name             Variance Std.Dev. Corr
##  Subject (Intercept)      2.155372 1.46812
##          time_since_start 0.005204 0.07214  -0.46
## Number of obs: 757, groups:  Subject, 396
##
## Conditional model:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.23459    0.11545   28.02  < 2e-16 ***
## time_since_start  0.03549    0.01029    3.45  0.00056 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```

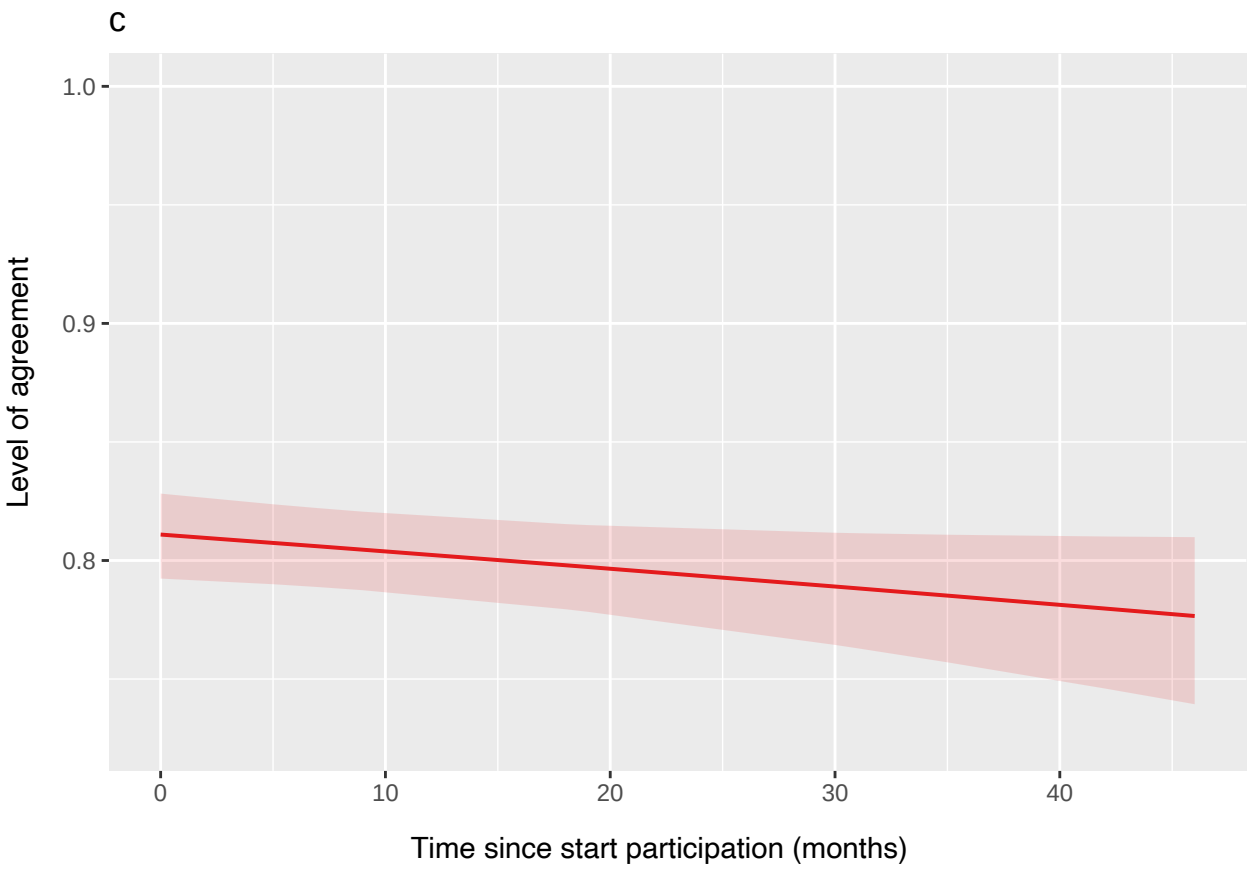


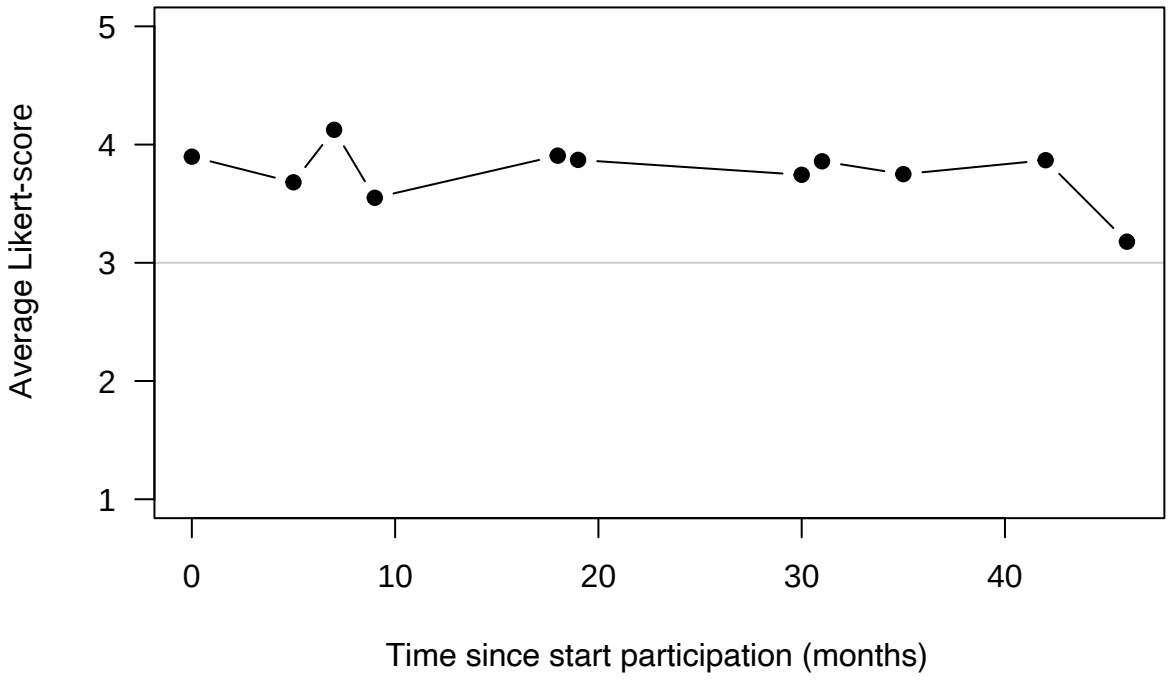


1.3 Cluster *Interest in scientific research*

```
##  Family: binomial  ( logit )
## Formula:          cbind(y, n) ~ time_since_start + (time_since_start | Subject)
## Data: Mo1
##
##      AIC      BIC   logLik deviance df.resid
##   3755.0   3778.1  -1872.5   3745.0      752
##
## Random effects:
##
## Conditional model:
##  Groups  Name             Variance  Std.Dev. Corr
##  Subject (Intercept)      1.0049494 1.00247
##          time_since_start 0.0005453 0.02335  -0.54
## Number of obs: 757, groups:  Subject, 396
##
## Conditional model:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.456150   0.059629  24.420   <2e-16 ***
## time_since_start -0.004572   0.002494  -1.833   0.0668 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```


c
1.0
0.9
0.8
Level of agreement
0
10
20
30
40
Time since start participation (months)



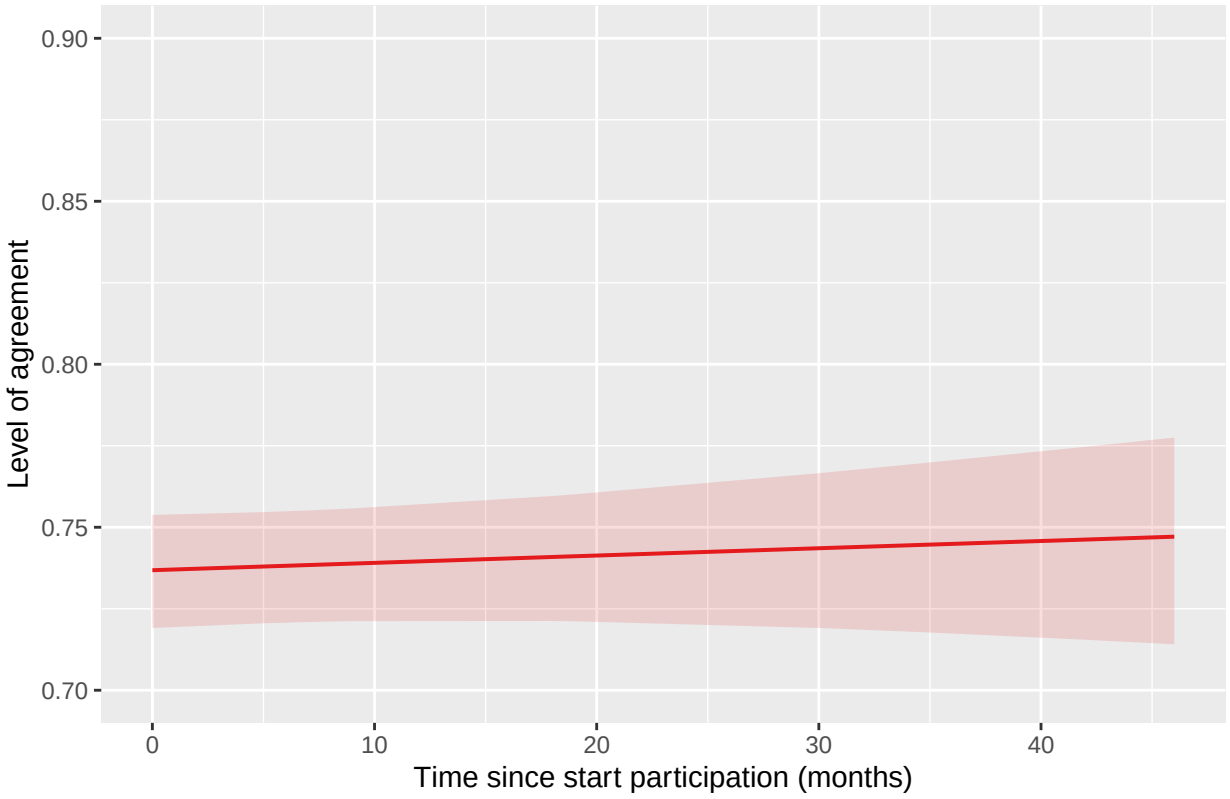
2 Attitude

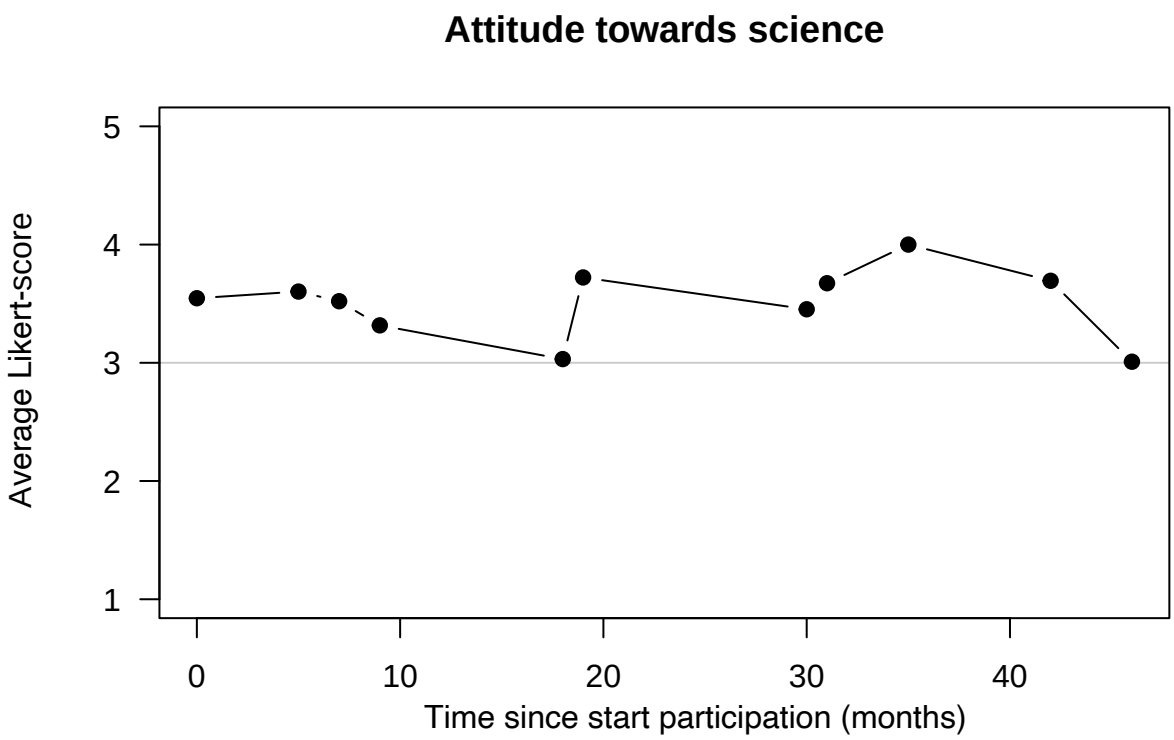
2.1 Cluster *Science*

```
##  Family: binomial  ( logit )
## Formula:
## cbind(summed, 8 * 5 - summed) ~ time_since_start + (time_since_start |
##     Subject)
## Data: AtS
##
##      AIC      BIC   logLik deviance df.resid
##   4740.4   4763.6  -2365.2   4730.4      753
##
## Random effects:
##
## Conditional model:
##  Groups  Name             Variance  Std.Dev. Corr
##  Subject (Intercept)      0.6938197 0.83296
##          time_since_start 0.0002305 0.01518  -0.06
## Number of obs: 758, groups:  Subject, 399
##
## Conditional model:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.029509   0.045626  22.564   <2e-16 ***
## time_since_start 0.001169   0.001749   0.668    0.504
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a

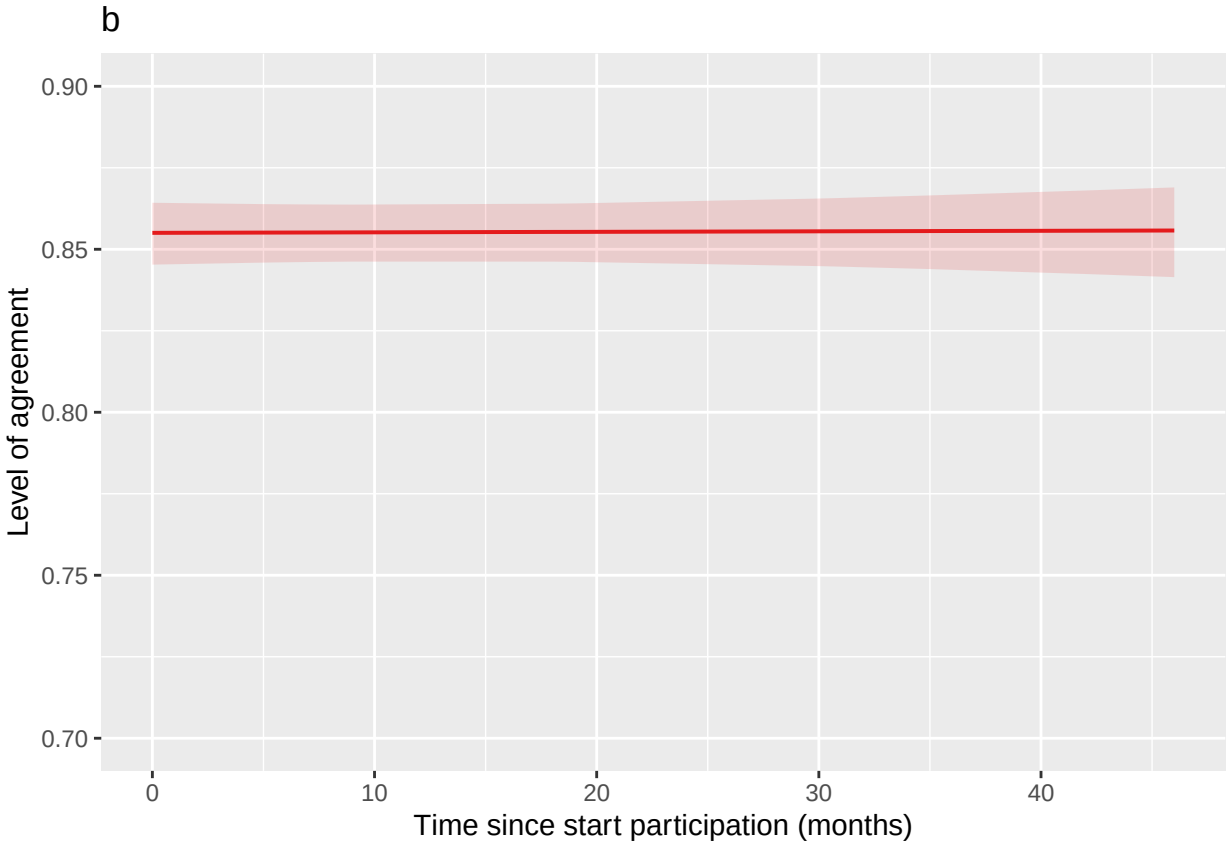


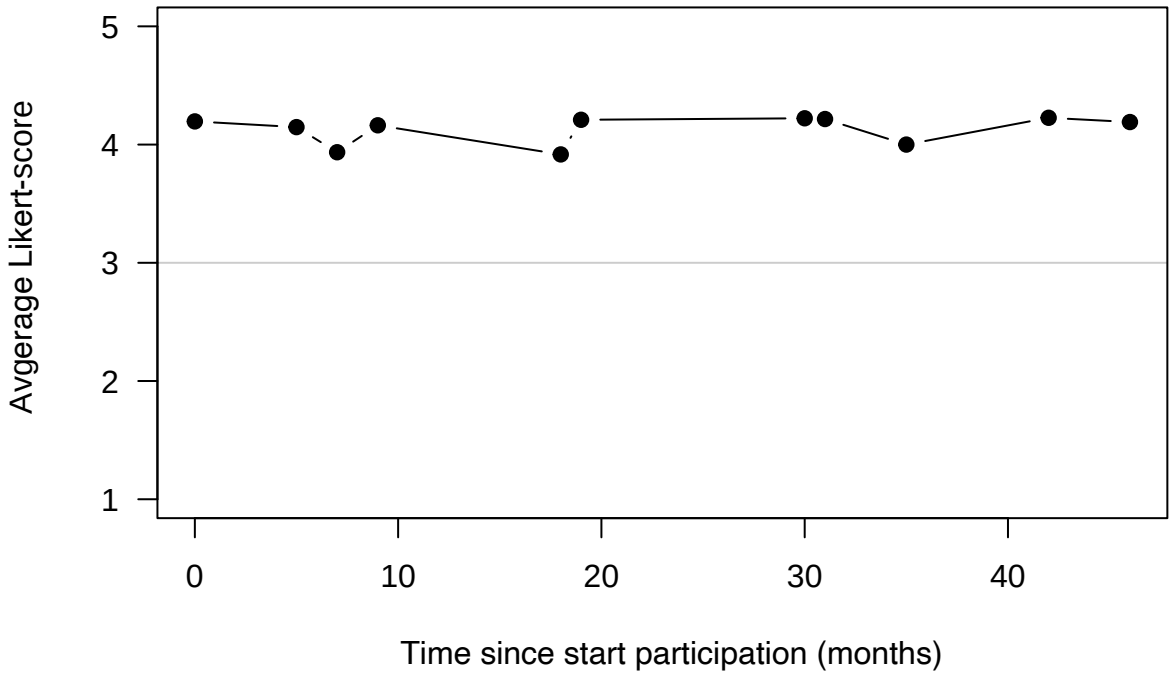


2.2 Cluster *Nature*

```
##  Family: binomial  ( logit )
## Formula:
## cbind(summed, 9 * 5 - summed) ~ time_since_start + (time_since_start |
##     Subject)
## Data: AtN
##
##      AIC      BIC   logLik deviance df.resid
##   4119.4   4142.5  -2054.7   4109.4      754
##
## Random effects:
##
## Conditional model:
##  Groups  Name             Variance  Std.Dev. Corr
##  Subject (Intercept)      4.316e-01 0.656934
##          time_since_start 3.193e-05 0.005651 -0.75
## Number of obs: 759, groups:  Subject, 399
##
## Conditional model:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.7746798  0.0390539   45.44   <2e-16 ***
## time_since_start 0.0001265  0.0012989    0.10    0.922
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```





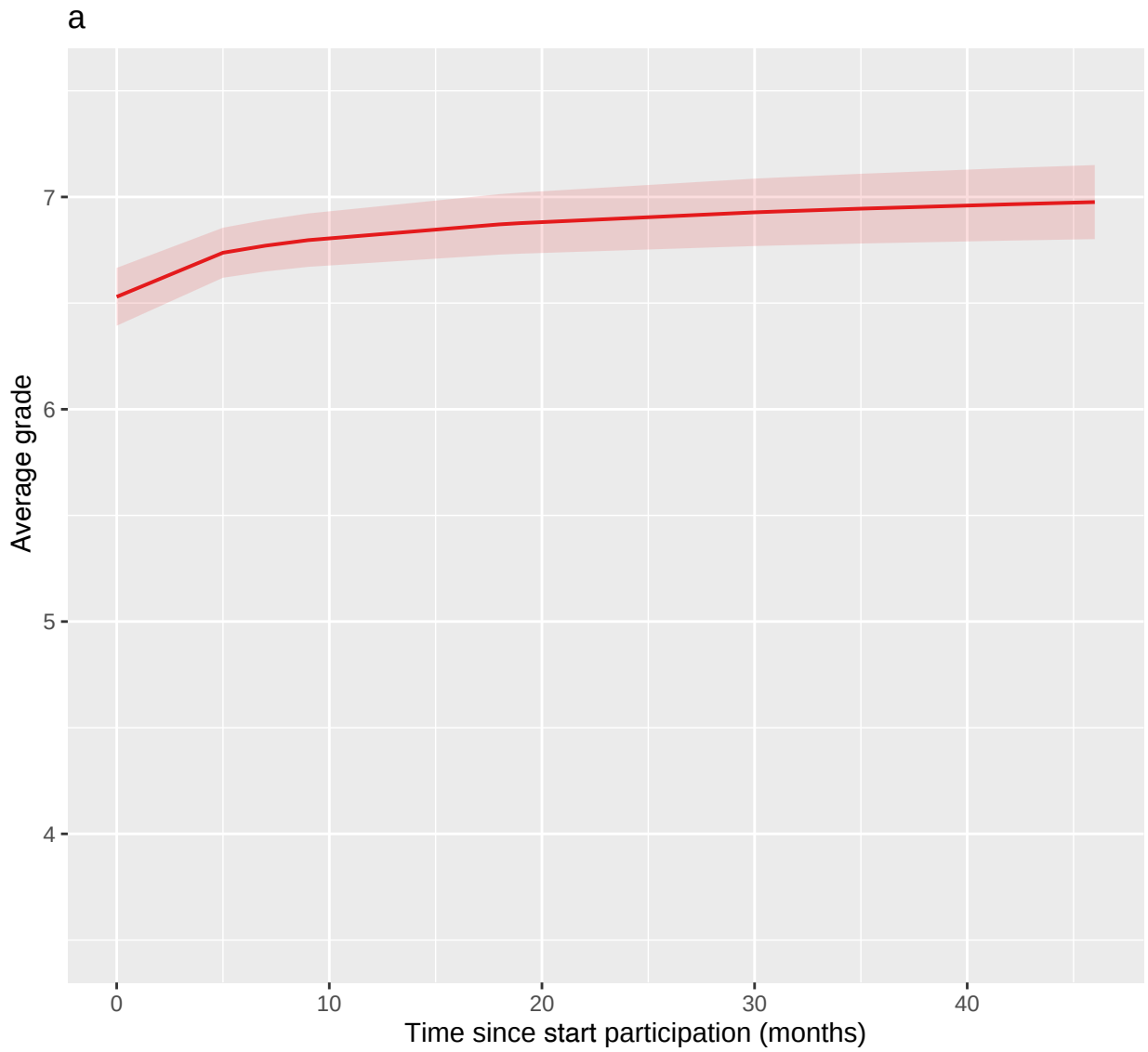
3 Knowledge

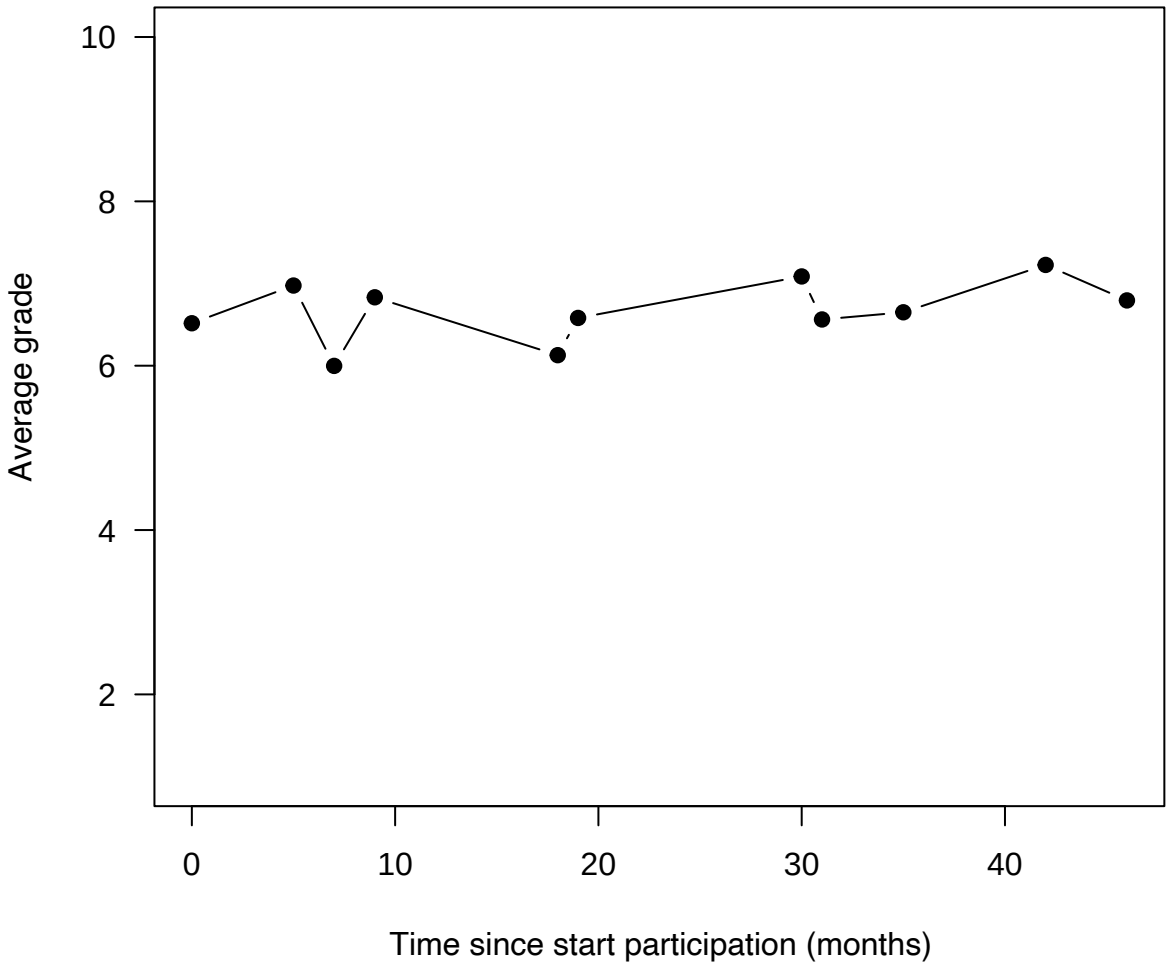
3.1 Cluster *Plastic pollution*

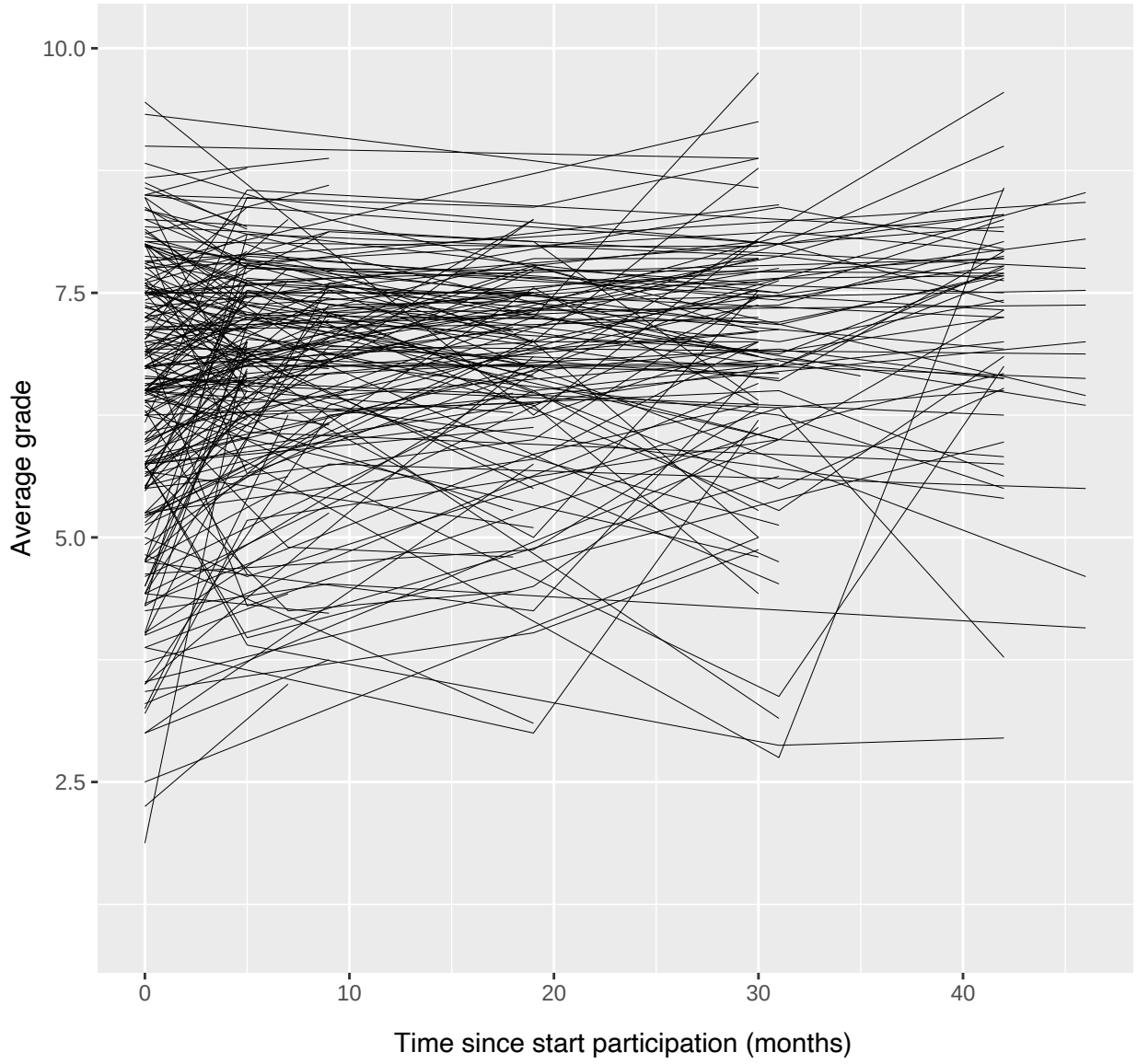
```
##  Family: gaussian  ( identity )
## Formula:
## meangrade ~ log(time_since_start + 1) + (log(time_since_start +
##     1) | Subject)
## Data: Gr1
##
##      AIC      BIC   logLik deviance df.resid
##   2462.0   2489.8  -1225.0   2450.0      749
##
## Random effects:
##
## Conditional model:
##  Groups   Name                        Variance Std.Dev. Corr
##  Subject  (Intercept)                 1.23174  1.1098
##           log(time_since_start + 1) 0.05674  0.2382   -0.62
##  Residual                             0.72118  0.8492
## Number of obs: 755, groups:  Subject, 398
##
## Dispersion estimate for gaussian family (sigma^2): 0.721
##
## Conditional model:
```

```
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 6.52970    0.06947   94.00  < 2e-16 ***
## log(time_since_start + 1)   0.11587    0.02683    4.32 1.57e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```




10
8
6
4
2
Average grade
0
10
20
30
40
Time since start participation (months)

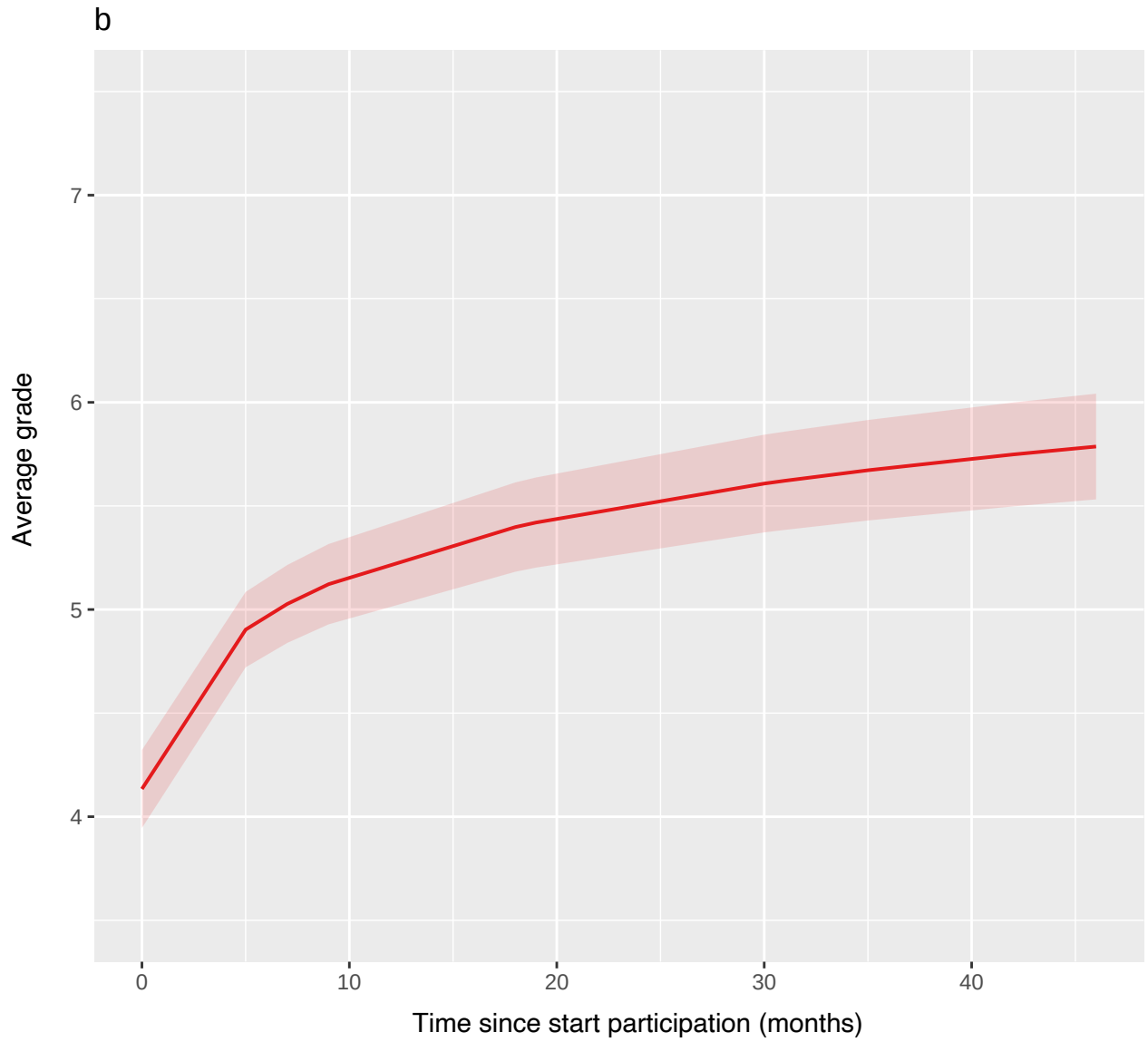


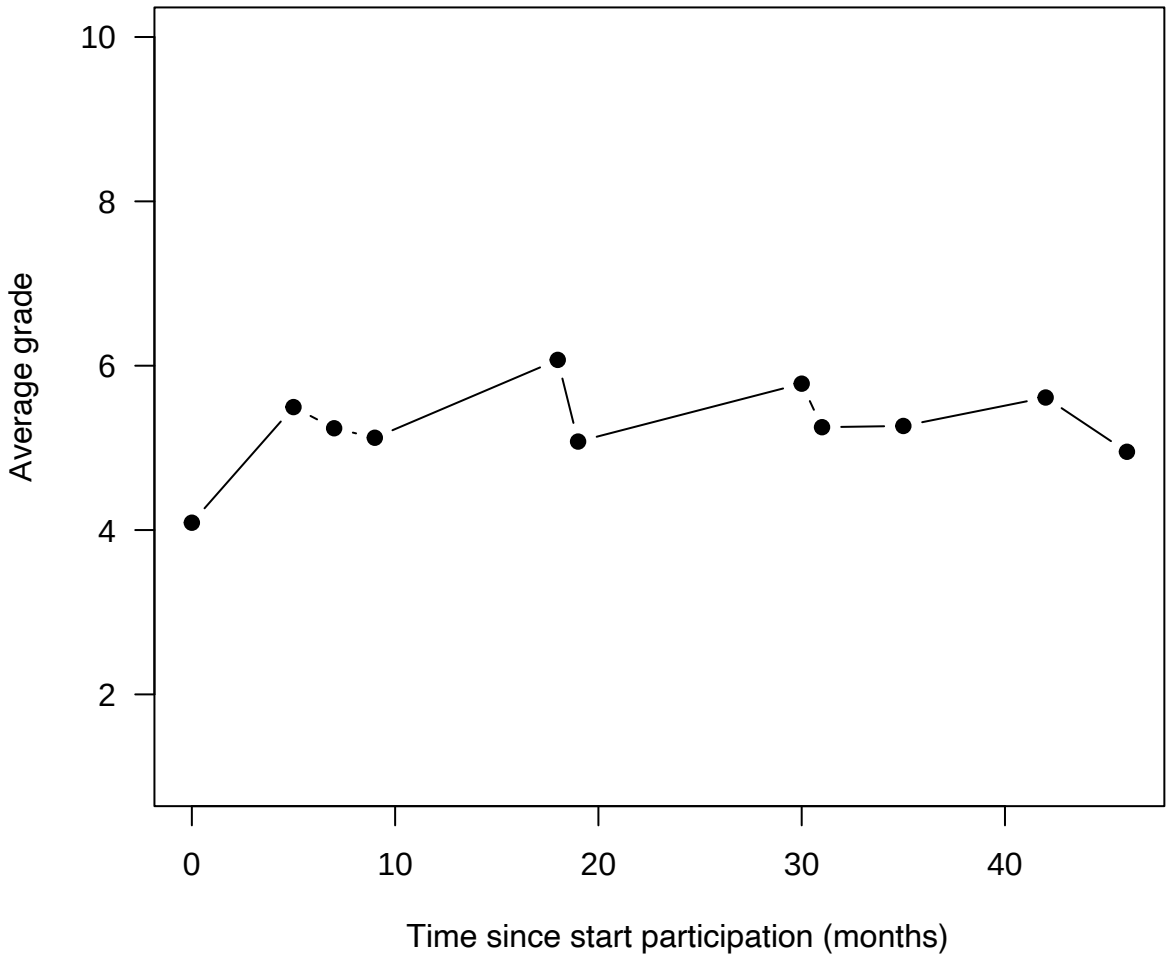
3.2 Cluster *Research*

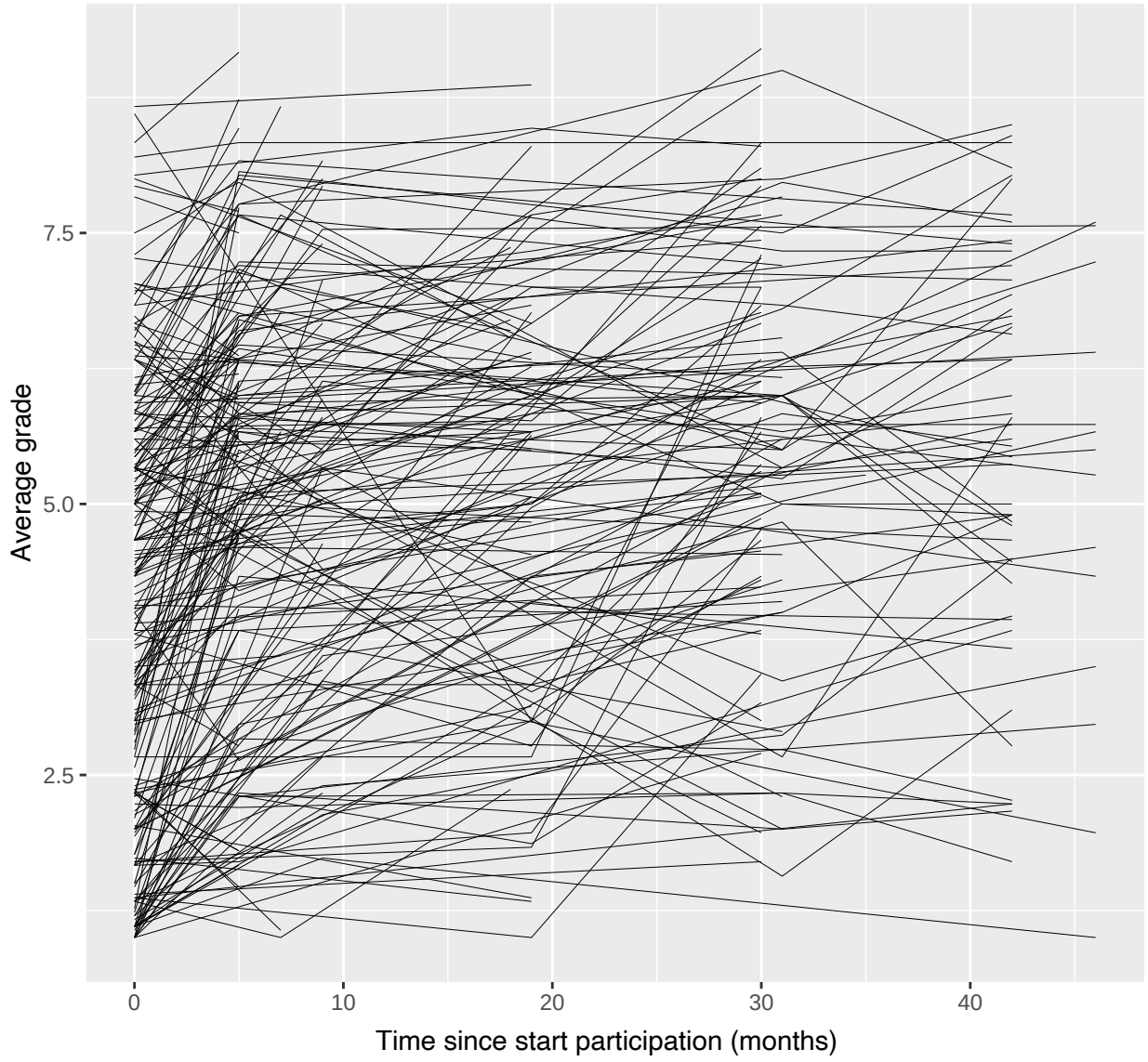
```
##  Family: gaussian  ( identity )
## Formula:
## meangrade ~ log(time_since_start + 1) + (log(time_since_start +
##     1) | Subject)
## Data: Gr2
##
##      AIC      BIC   logLik deviance df.resid
##   2897.2   2924.9  -1442.6   2885.2      749
##
## Random effects:
##
## Conditional model:
##  Groups   Name                        Variance Std.Dev. Corr
##  Subject  (Intercept)                 2.70313  1.6441
##           log(time_since_start + 1) 0.07812  0.2795   -0.33
```

```
##  Residual                             1.00349  1.0017
## Number of obs: 755, groups:  Subject, 398
##
## Dispersion estimate for gaussian family (sigma^2):    1
##
## Conditional model:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                4.13374    0.09596   43.08   <2e-16 ***
## log(time_since_start + 1)  0.42933    0.03368   12.75   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```




10
8
6
4
2
Average grade
0
10
20
30
40
Time since start participation (months)


Average grade
7.5
5.0
2.5
0
10
20
30
40
Time since start participation (months)

References

1. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2023).
2. Posit team. *RStudio: Integrated development environment for r*. (Posit Software, PBC, 2023).
3. Brooks, M. E. *et al.* glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* **9**, 378–400 (2017).
4. Hartig, F. *DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models*. (2022).
5. Lüdecke, D. *sjPlot: Data visualization for statistics in social science*. (2023).
6. Peeters, C. F. W., Bilgrau, A. E. & van Wieringen, W. N. rags2ridges: A one-stop-ℓ_2-shop for graphical modeling of high-dimensional precision matrices. *Journal of Statistical Software* **102**, 1–32 (2022).