Final Project Data Exploration

Denis Ostroushko

```
[1] "Distribution of Available time Points per subject"

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    3.00    4.00    3.44    4.00    4.00

[1] "Most Subjects >= 3 time points, but data shows some imbalance"

[1] "Distribution of Unique Counts of Homecog - confirmed cluster invariant variable"


  1
405

[1] "Distribution of Unique Counts of Homeemo - confirmed cluster invariant variable"


  1
405
```

Primary Outcome

Figure 1 shows:

- Linear Trend in Reading ability against time
- High degree of variation overall, variation increases as time goes on

Figure 2 shows:

- Likely a strong linear trend persists
- Individual effect counts for a large portion of variation
- Individual variation is strong in a setting with both time variant predictors

Figure 3: hard t determine effect of `homecog` on the reading ability

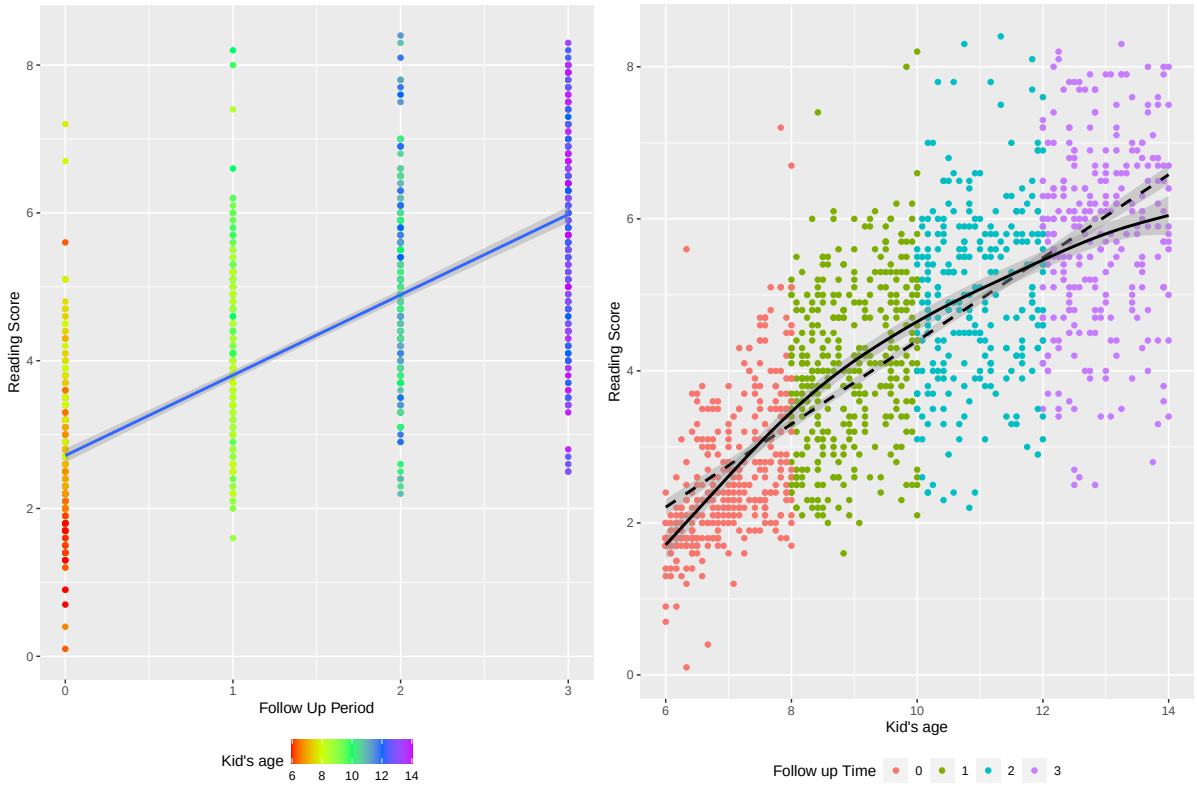


Figure 1: Trend of Reading ability vs Follow Up Time and Kid's age

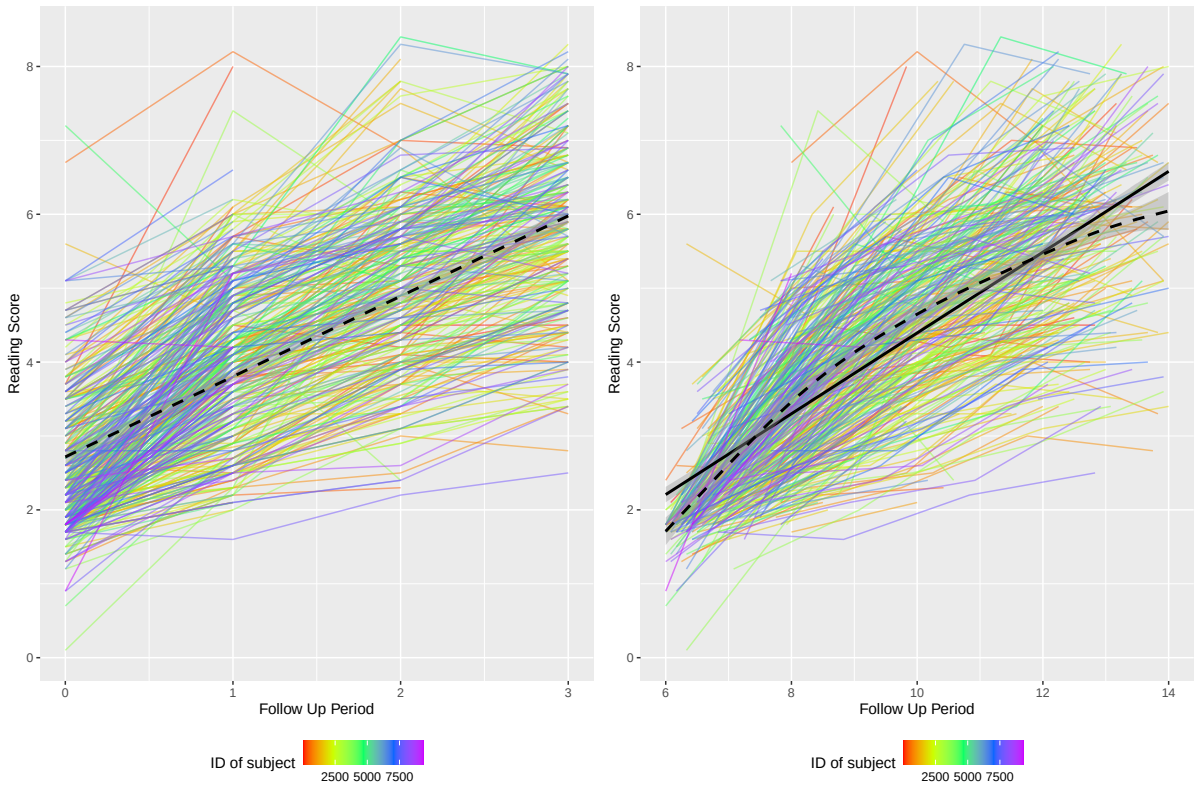


Figure 2: Trend of Reading ability vs Follow Up Time and Kid's age for Individuals

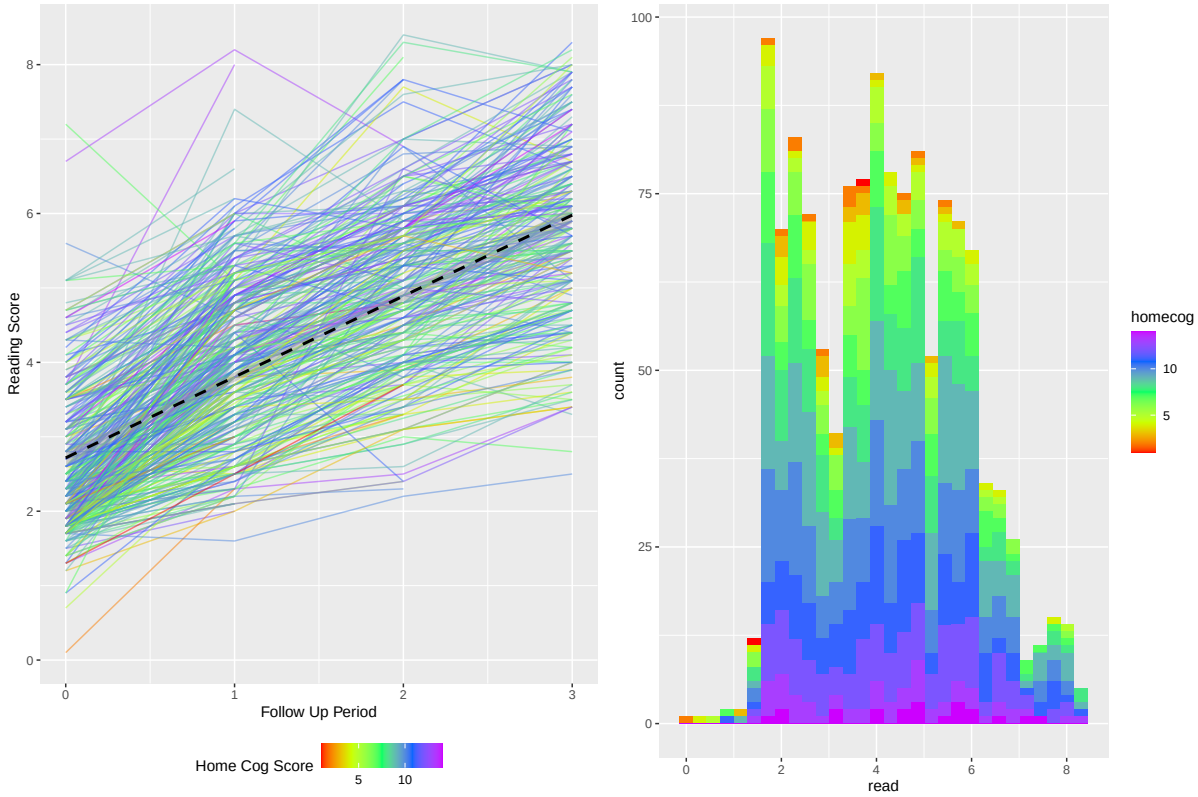


Figure 3: Trend of Reading ability by Homecog Score

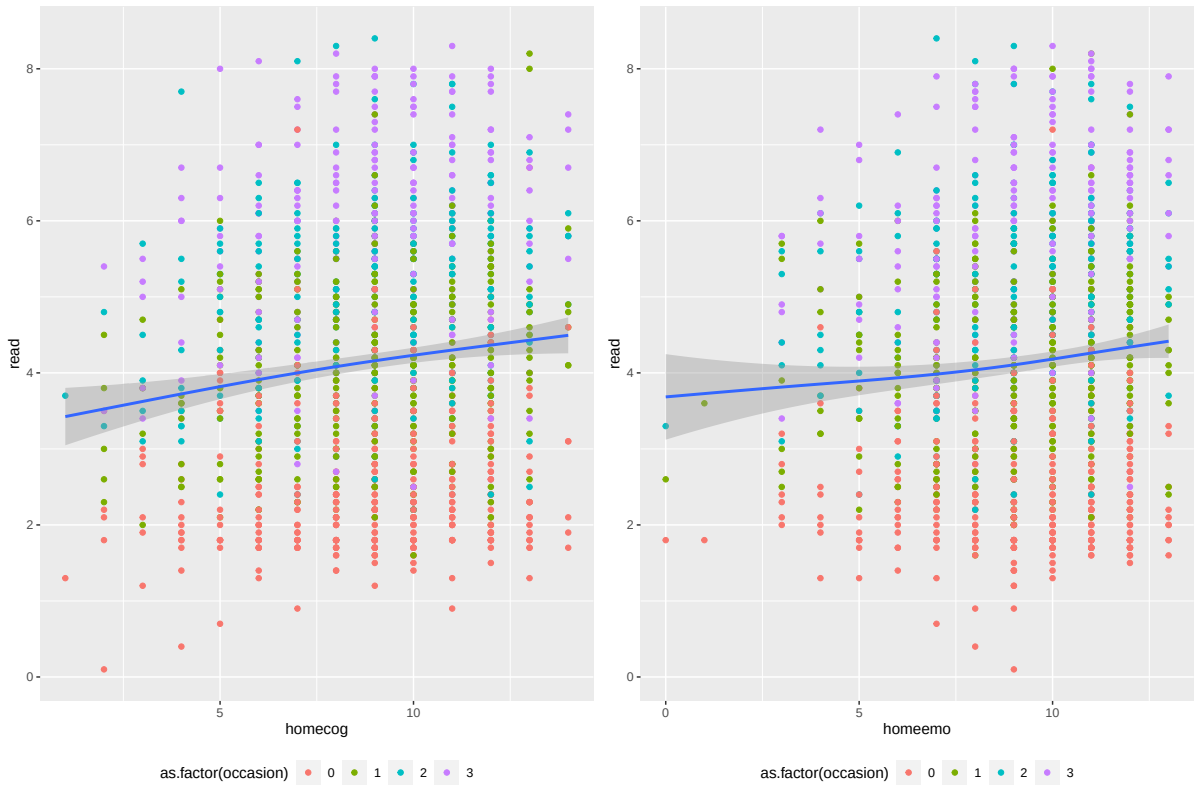
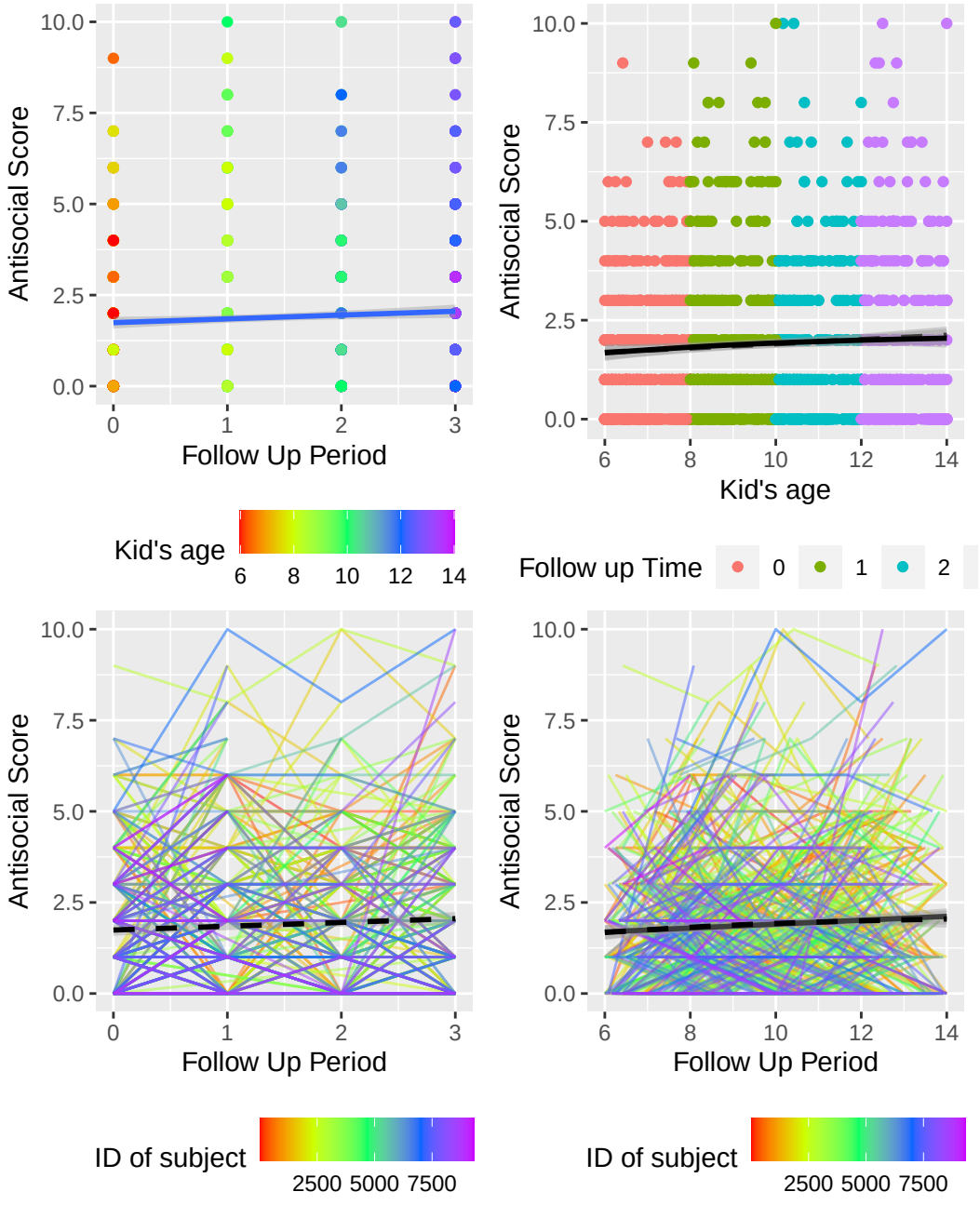
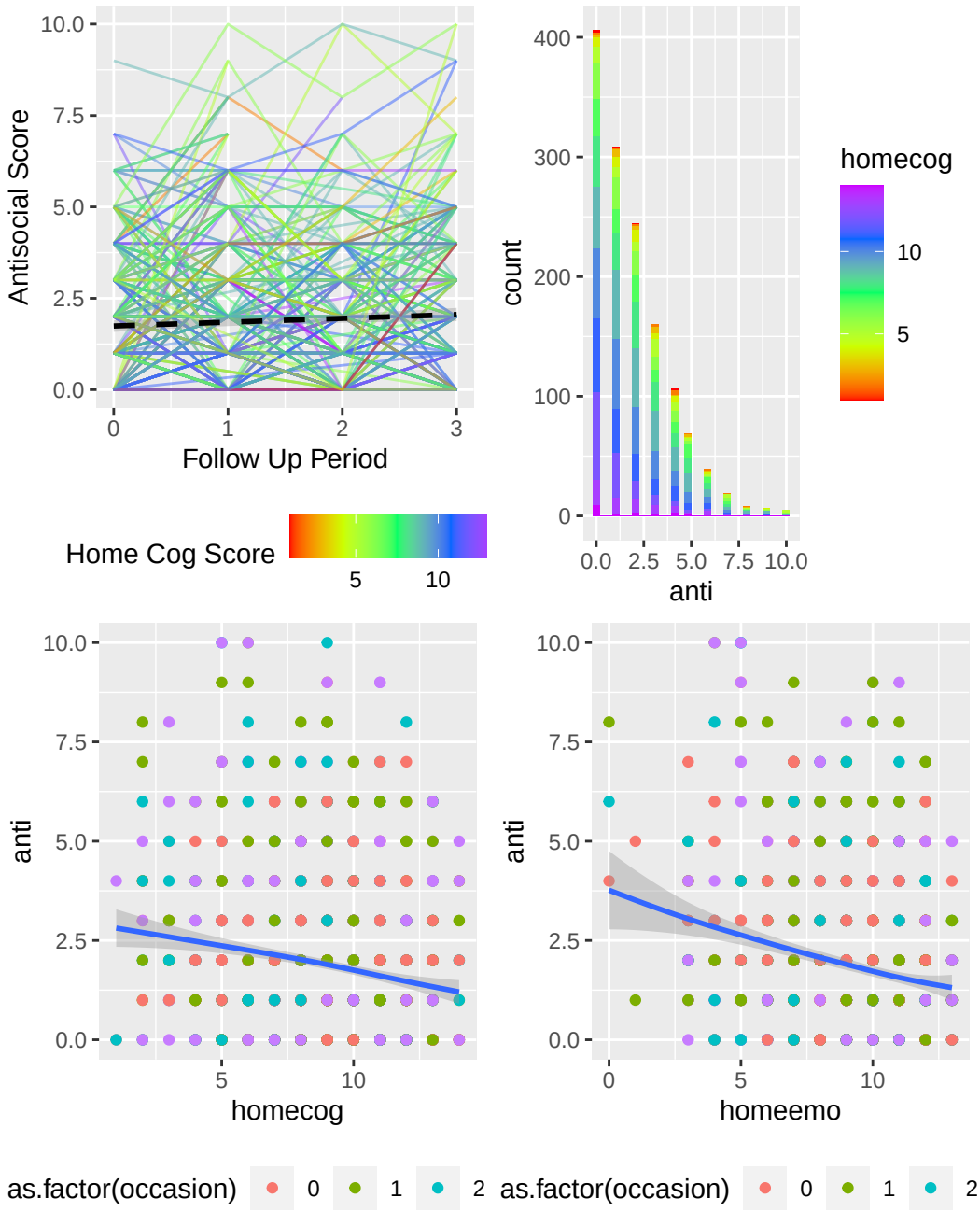


Figure 4: Relationship between Baseline Home variables and Reading ability

Secondary Outcome




Antisocial Score
Follow Up Period
Home Cog Score
count
anti
homecog
homeemo
as.factor(occasion)
0
1
2

Conclusions

- Might want to use mixed effects model to account for a large amount of individual variation
- `Homecog` and `Homeemo` might have a weak effect, but after accounting for individual random effect, they might be significant after all
- No visual evidence that we need to consider higher order predictors, no apparent non-linear trends
- Variance increases as time increases, we will consider a poisson, or some other model to account for this effect of variance
- Antisocial scores vary quite a lot in the population
- As HOME variables increase, antisocial score decreases
- Antisocial score fluctuates quite a lot, however, it seems that on average it is higher for kids with lower HOME variables scores