Linear Mixed Effects Models - 2 - results

Kiri Koppelgaard

September 14, 2018

Language development in Autism Spectrum Disorder (ASD)

Background: Autism Spectrum Disorder is often related to language impairment. However, this phenomenon has not been empirically traced in detail: i) relying on actual naturalistic language production ii) over extended periods of time

We therefore videotaped circa 30 kids with ASD and circa 30 comparison kids (matched by linguistic performance at visit 1) for ca. 30 minutes of naturalistic interactions with a parent. We repeated the data collection 6 times per kid, with 4 months between each visit. We transcribed the data and counted: i) the amount of words that each kid uses in each video. Same for the parent. ii) the amount of unique words that each kid uses in each video. Same for the parent. iii) the amount of morphemes per utterance (Mean Length of Utterance) displayed by each child in each video. Same for the parent.

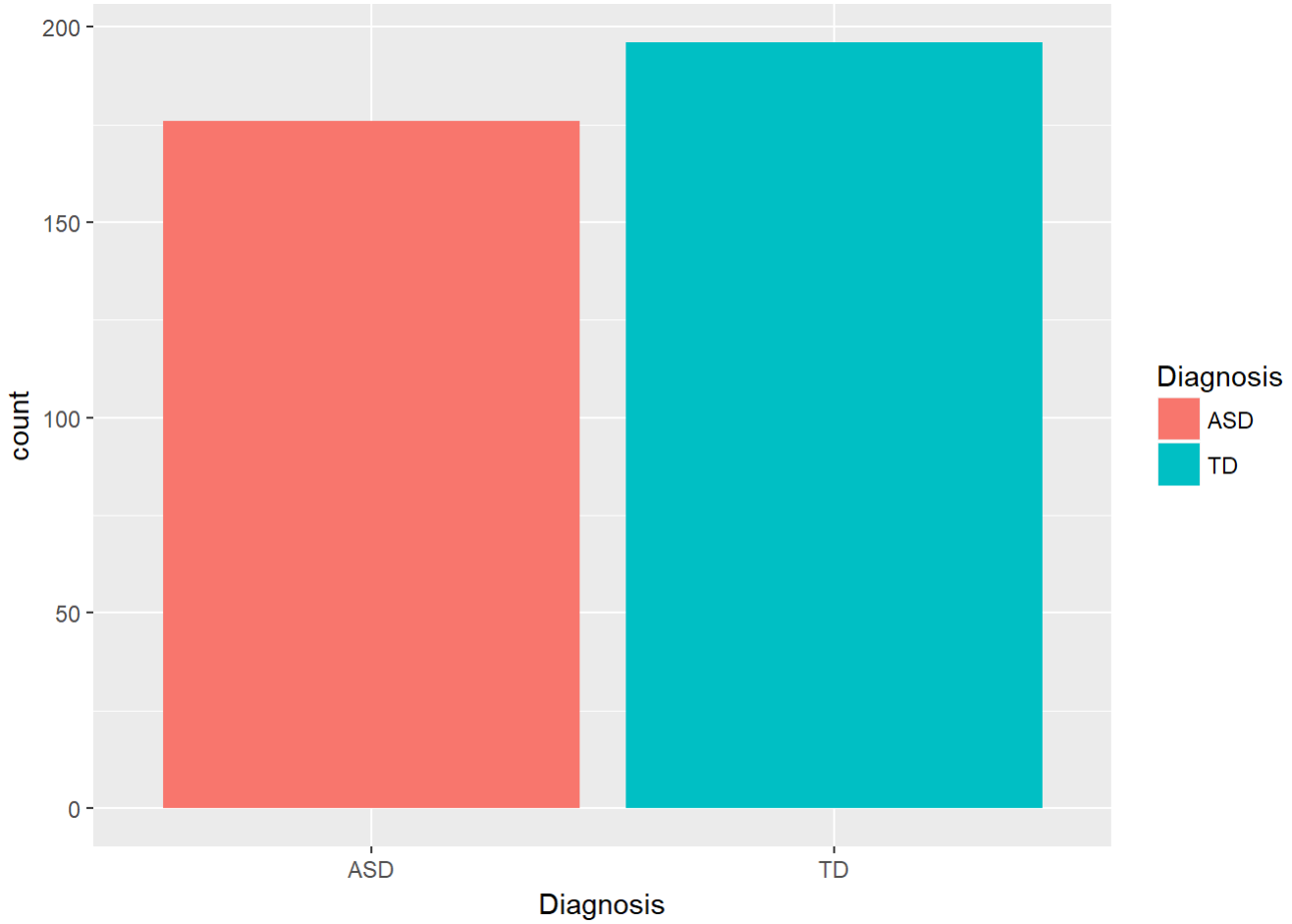
We then want to test the language trajectory of child and parent over time.

Exercise 1) Preliminary Data Exploration

Describe the participant samples in the dataset (e.g. by diagnosis, age, etc.). Do you think the two groups are well balanced? If not, what do you think was the reason?

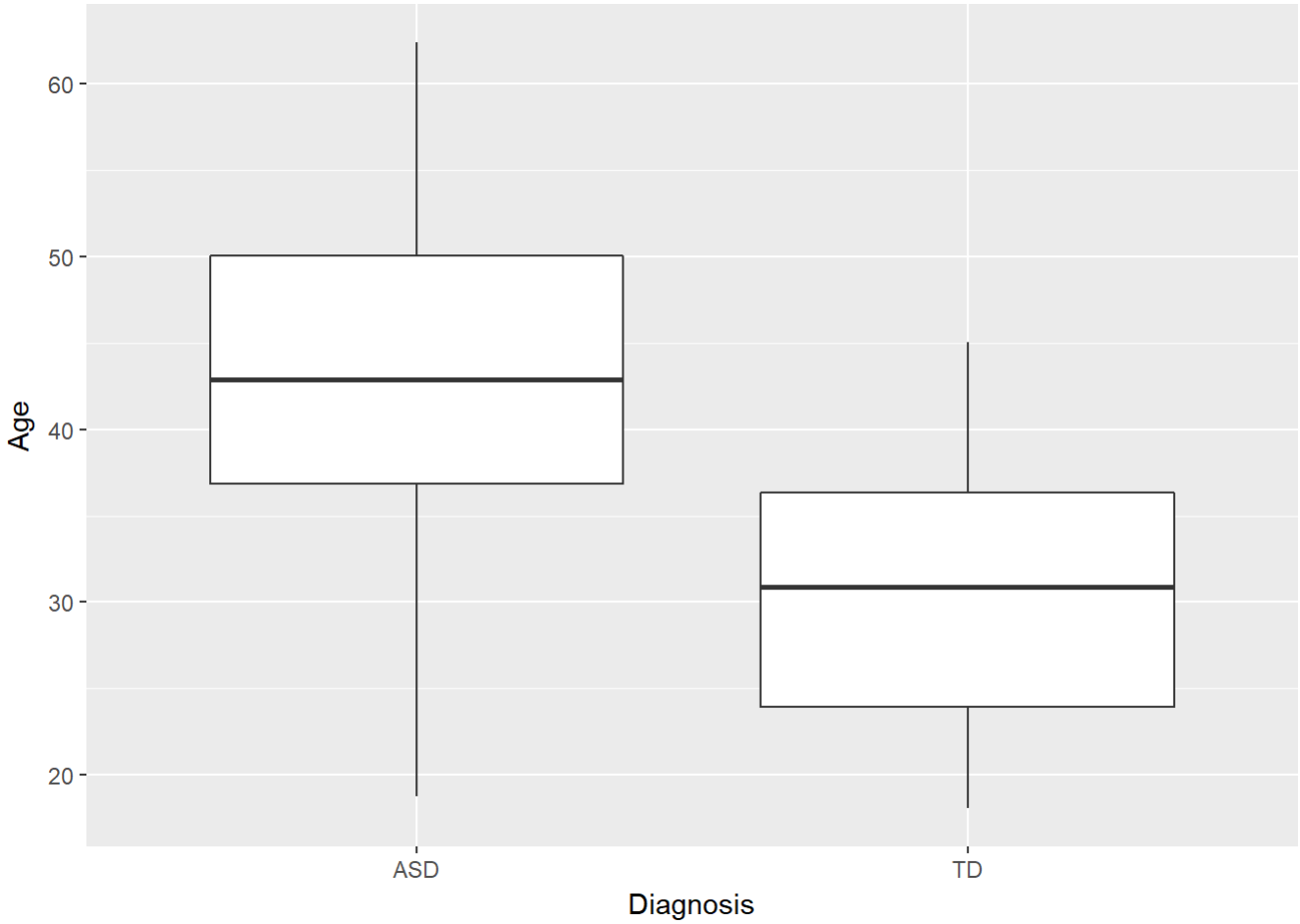
Generally, the sample of participants are proportionally distributed according to the most relevant variables. This enables application of statistical tools which will produce relatively reliable predictions of the mean length of utterance of children predicted by time and their diagnosis. However, there are several issues I would like to point to.

As seen below the distribution of typically developed children and children with ASD is relatively equal

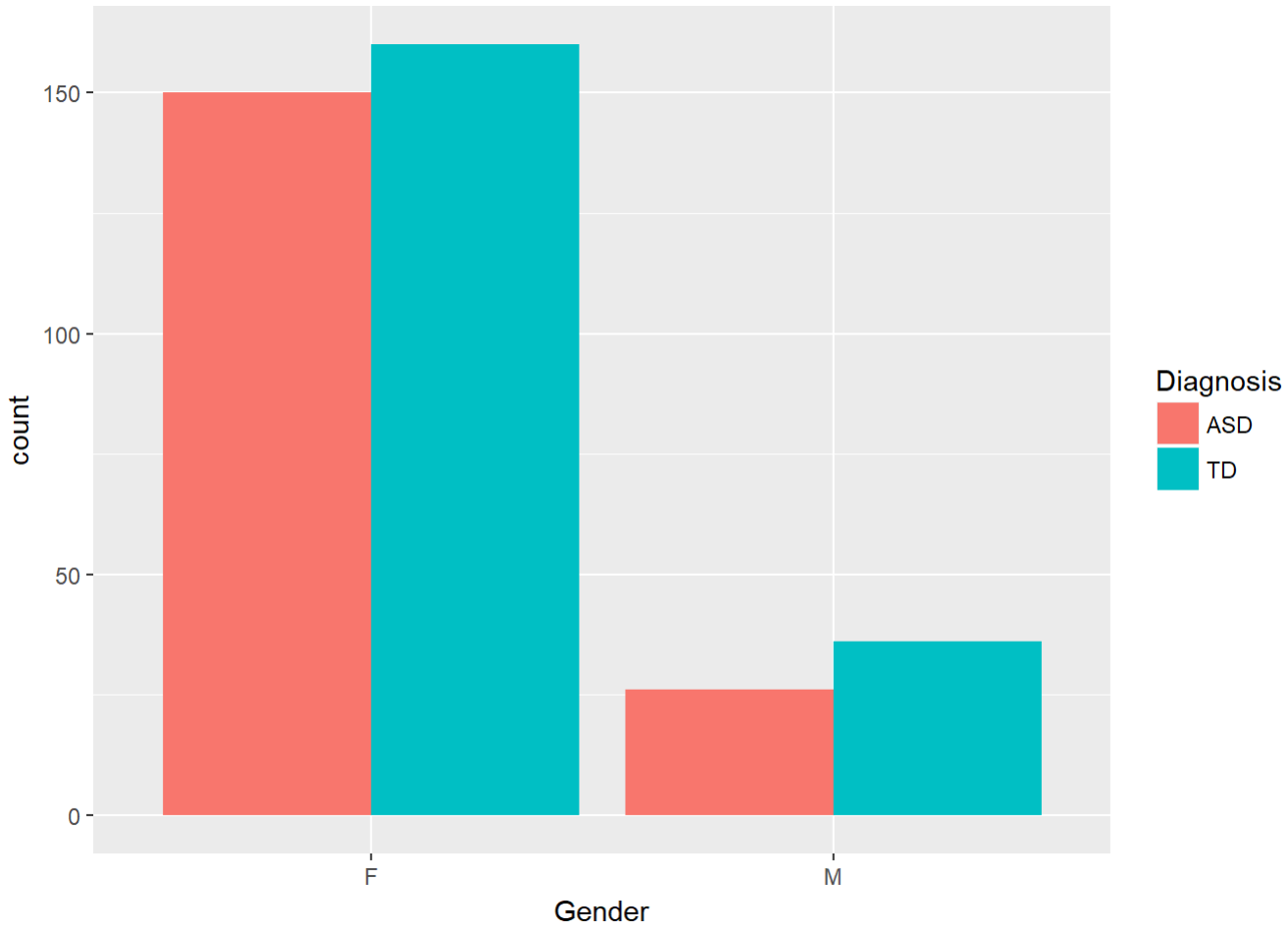


A comment to be made in relation to the distribution of typically developing children compared to children with ASD is that this proportion does not map onto the population. However, since the variable we wish to measure is the mean length of utterance in relation to time and diagnosis, the distribution of diagnosis play little role.

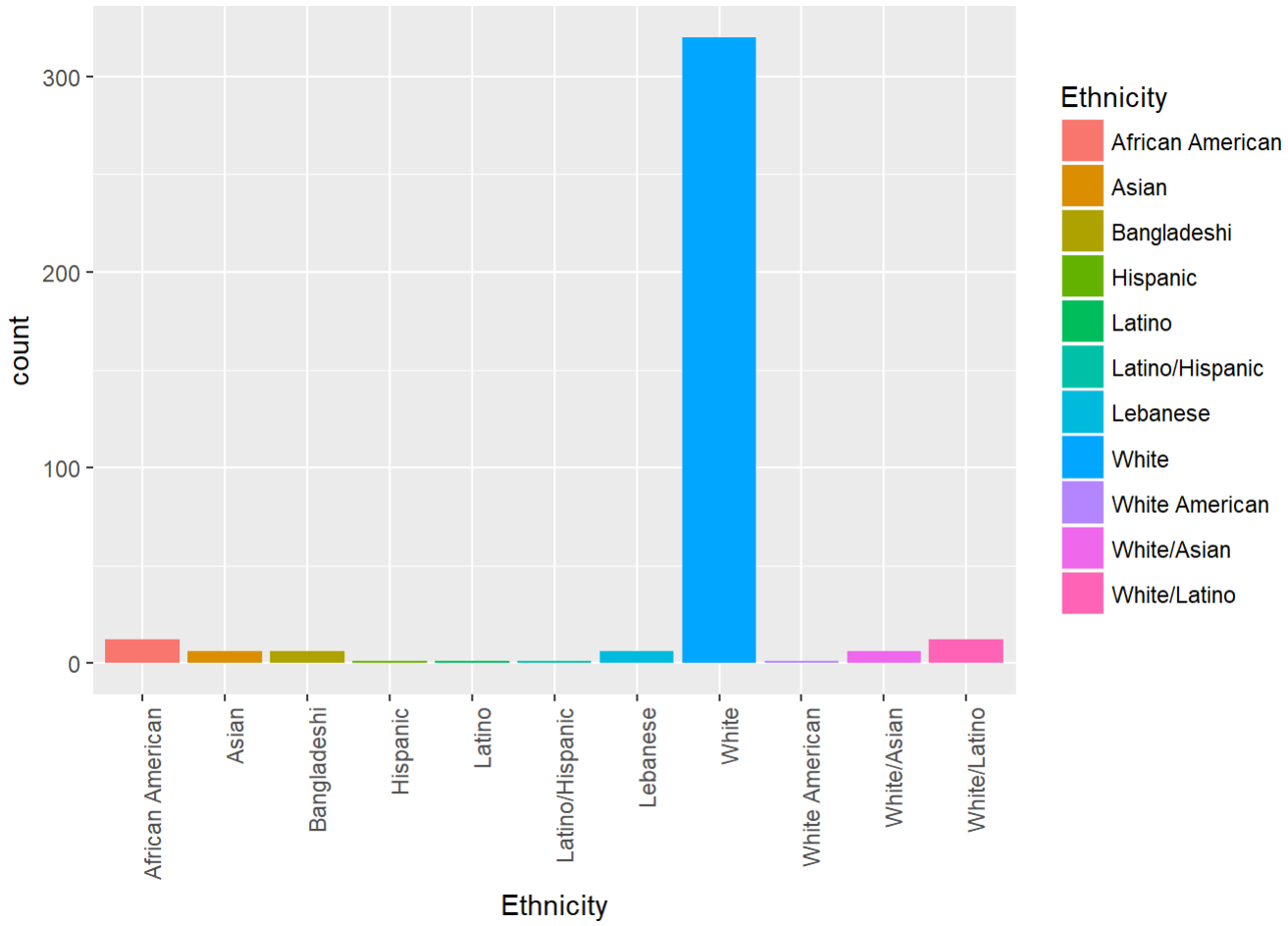
A comment can be made in relation to the distribution of age of in relation to diagnosis. As seen in the plot below, the sample has a disproportionate age distribution. It appears that the sampled children with ASD are slightly older than typically developed children. Since we use visits as a predictor of mean length of utterance rather than age, this might have the effect that the sampled children with ASD will be further along in their development without us taking into account that they were older in the first place.

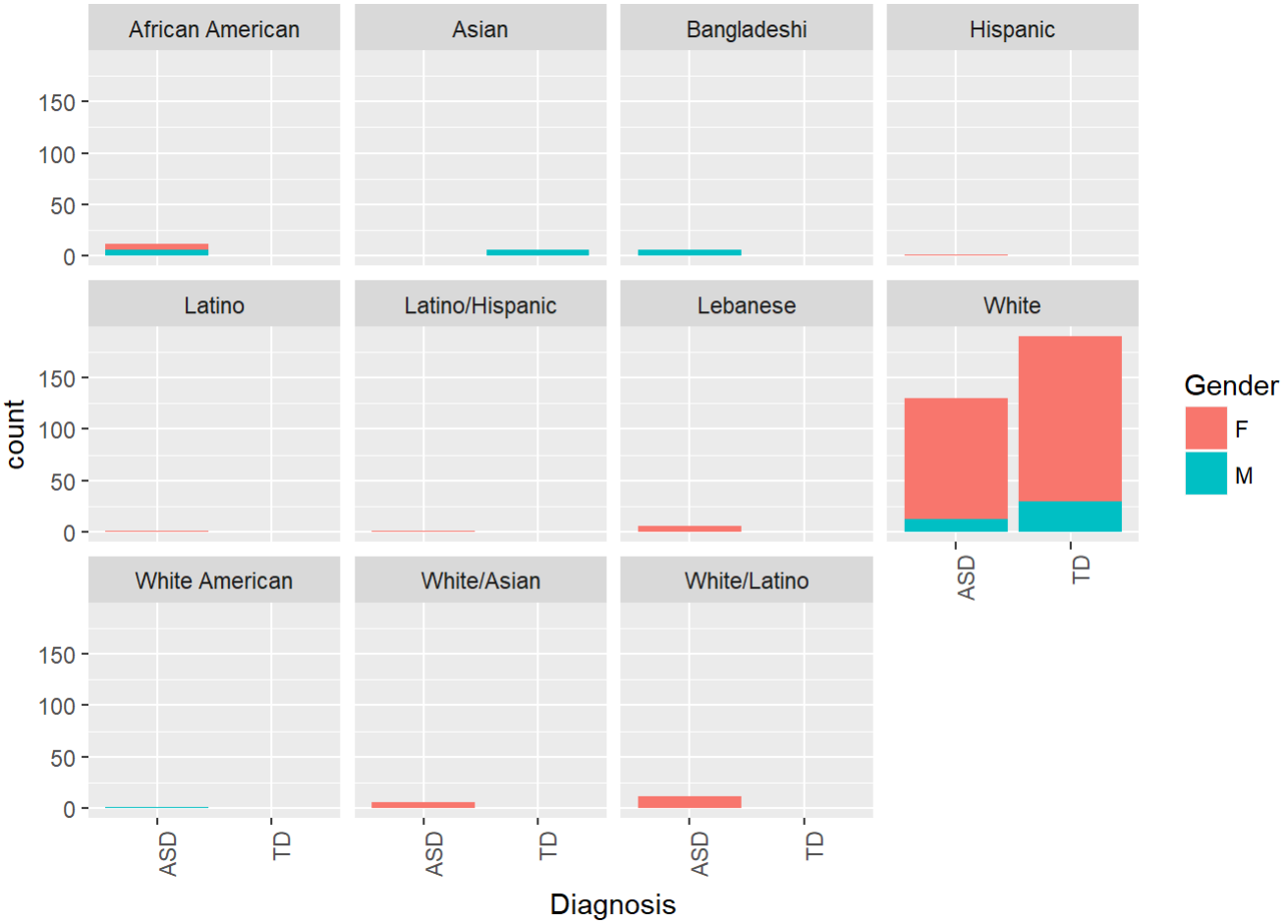


The distribution of gender is pretty skewed. The high amount of females seems strange having in mind mostly males are diagnosed with ASD. However, we do not wish to predict anything about gender, therefore we will be fine, though it might be a good idea to include gender as a random effect.

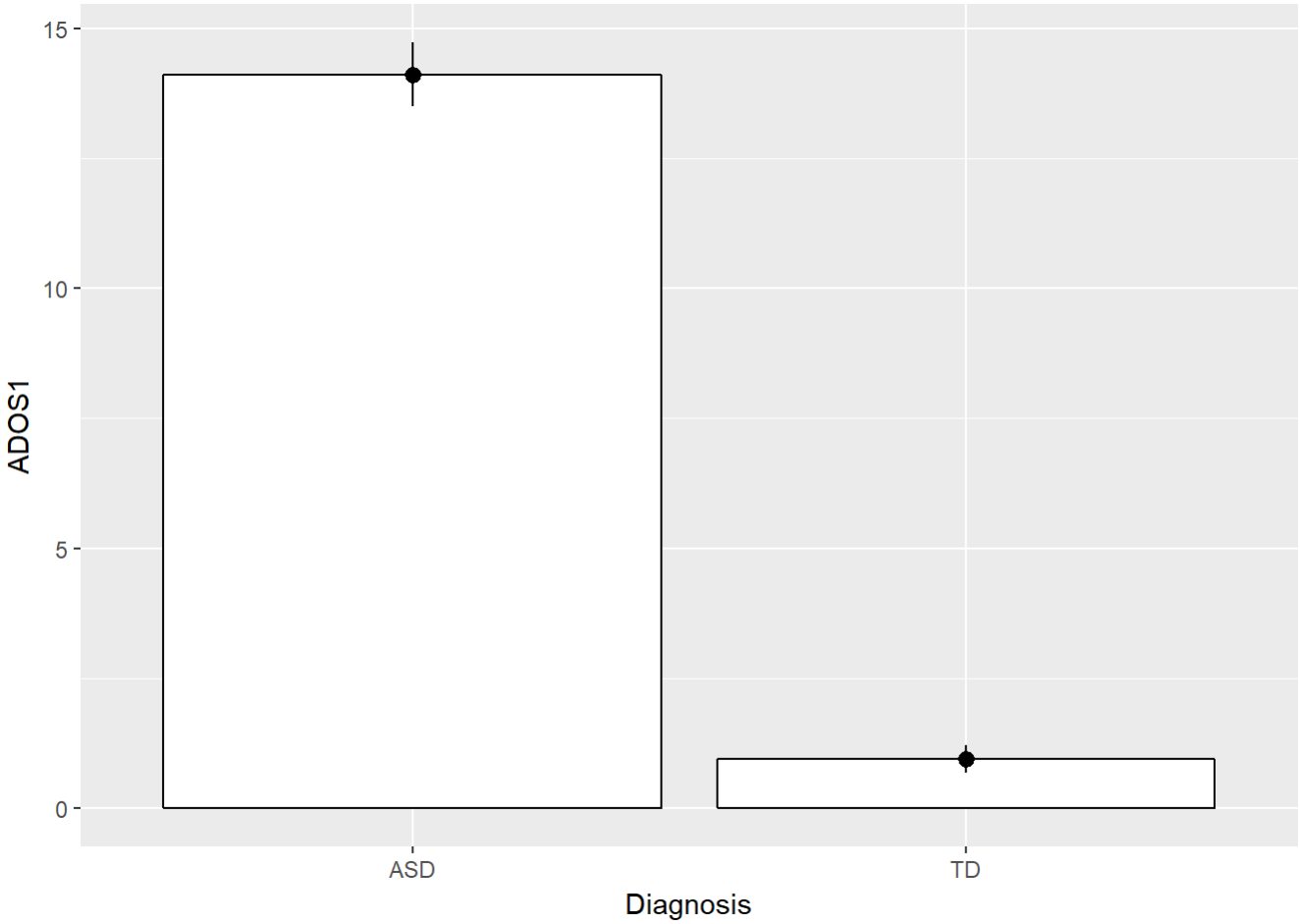


The distribution of ethnicity shows a high amount of whites compared to other ethnicities. However, since we have little interest in the effect of ethnicity on mean length of utterance, this is not too alarming. Especially, since the distribution of diagnosis within the catgories of ethnicity show a equal distribution.


count
300
200
100
0
African American
Asian
Bangladeshi
Hispanic
Latino
Latino/Hispanic
Lebanese
White
White American
White/Asian
White/Latino
Ethnicity
Ethnicity
African American
Asian
Bangladeshi
Hispanic
Latino
Latino/Hispanic
Lebanese
White
White American
White/Asian
White/Latino

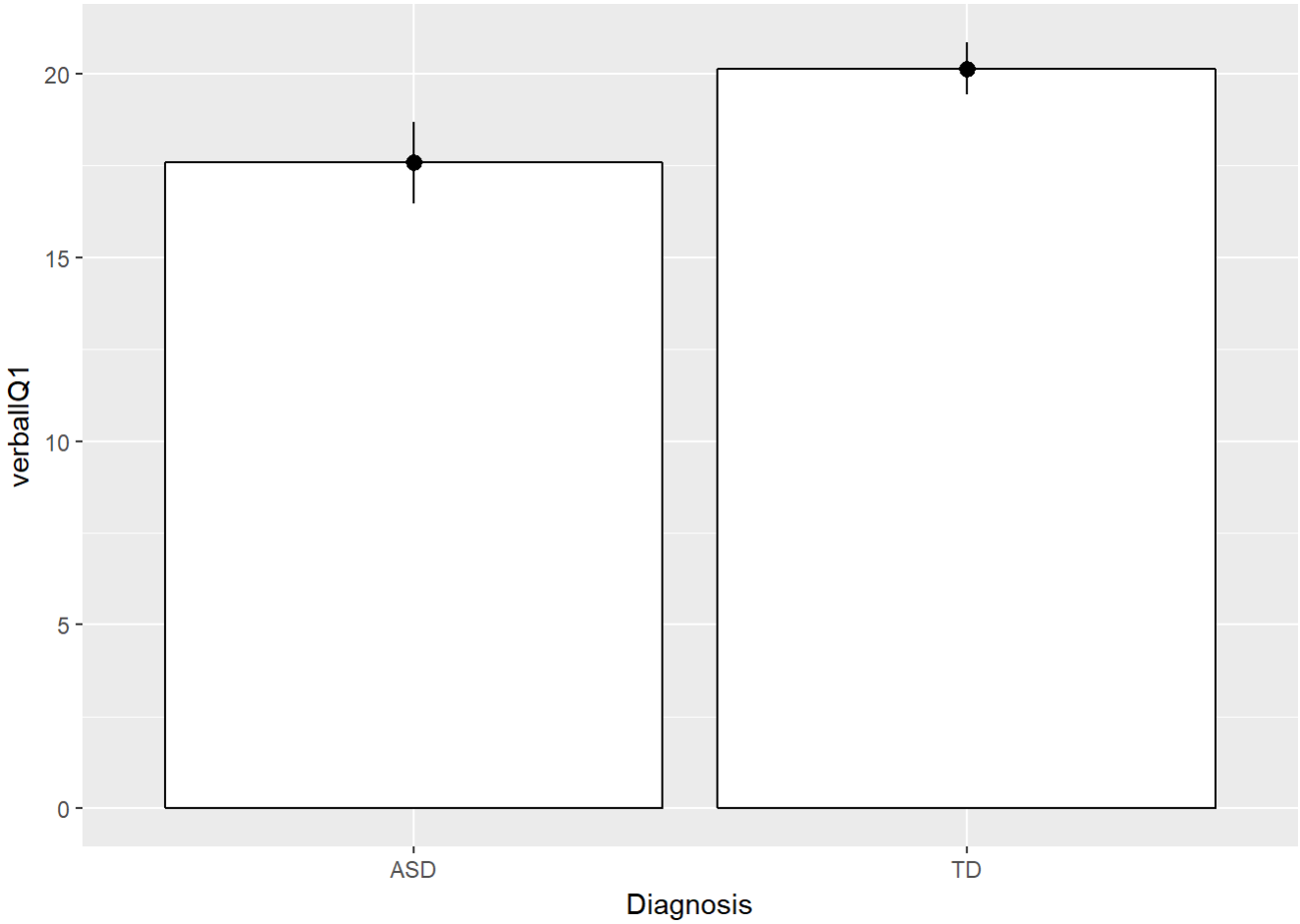


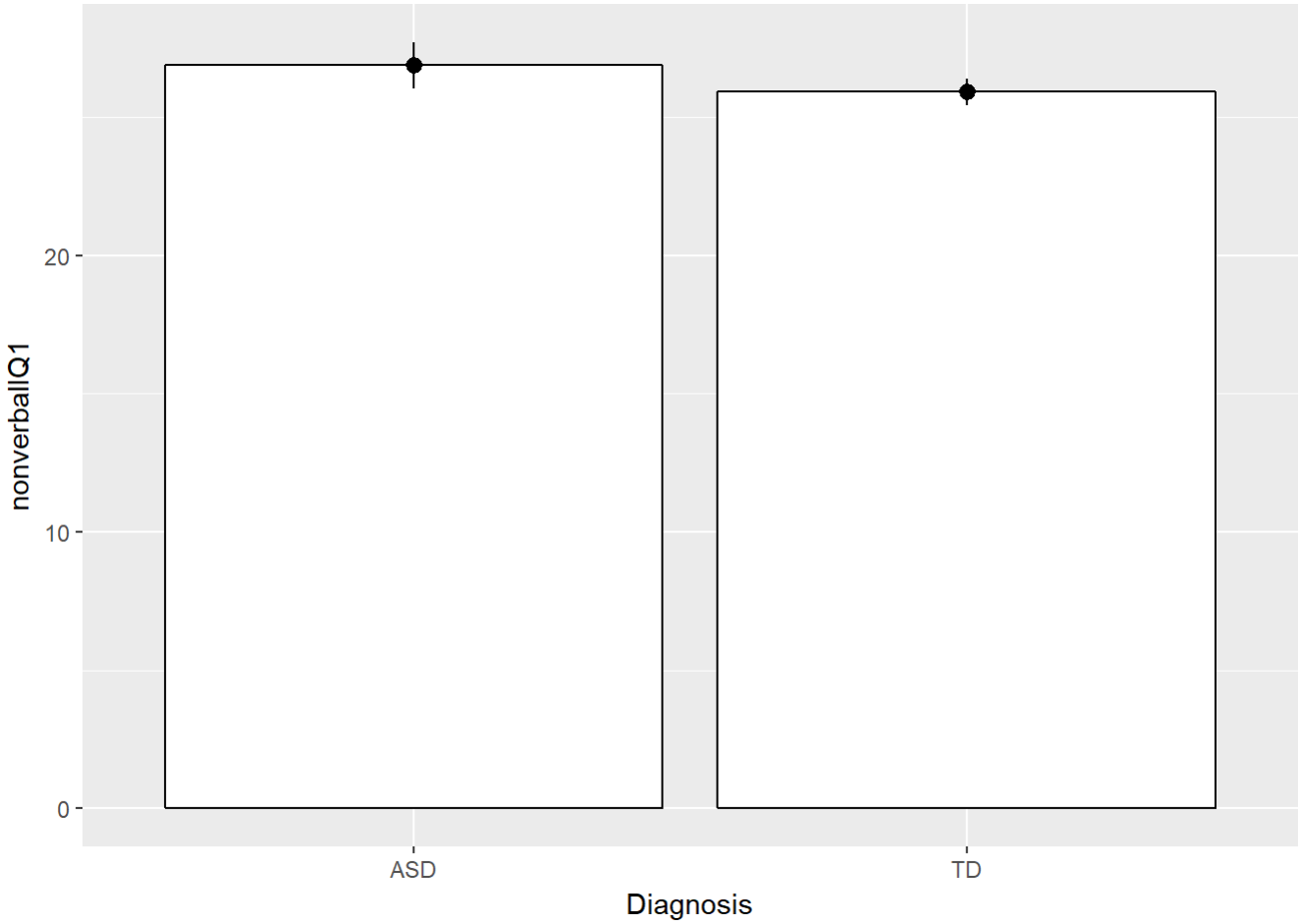
The following plots are highly relevant for assessing, whether the measures of diagnosing ASD are any good. The plot below visualize that the mean ADOS value (an instrument for diagnosing and assessing autism) is significantly higher for children with ASD than typically developing children - thus, the measure works. However, the diagnosis of the children is probably based on this measure, therefore it might in reality not tell us much.



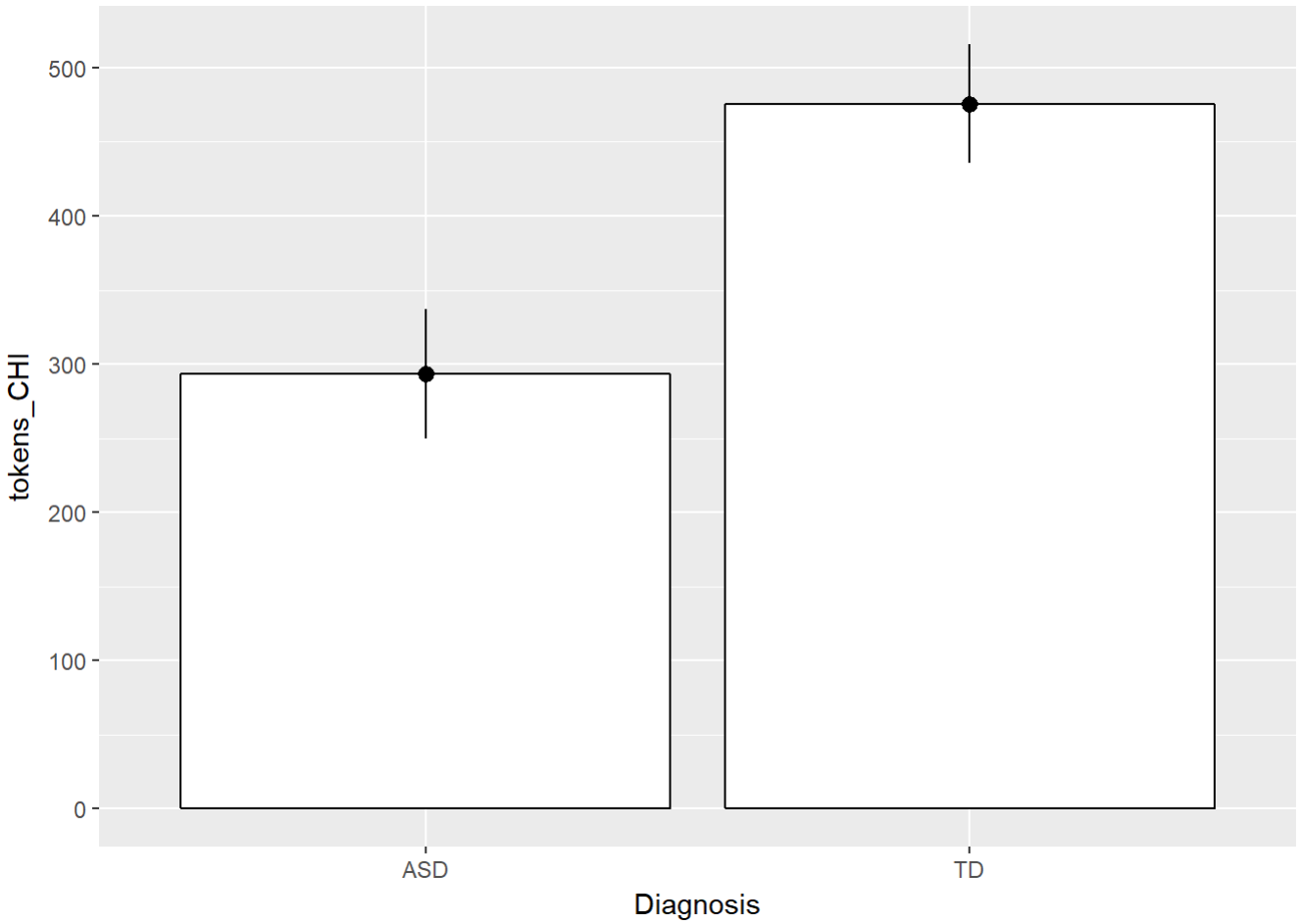
The linguistic intelligence of the children are measured using a verbal IQ measure. The verbal IQ would we expect to be lower for children with ASD, since we hypothesise children with ASD in general use fewer words, also unique words and speak less in coherence. The plot supports this hypothesis, since typically developing childern has a higher verbal IQ.

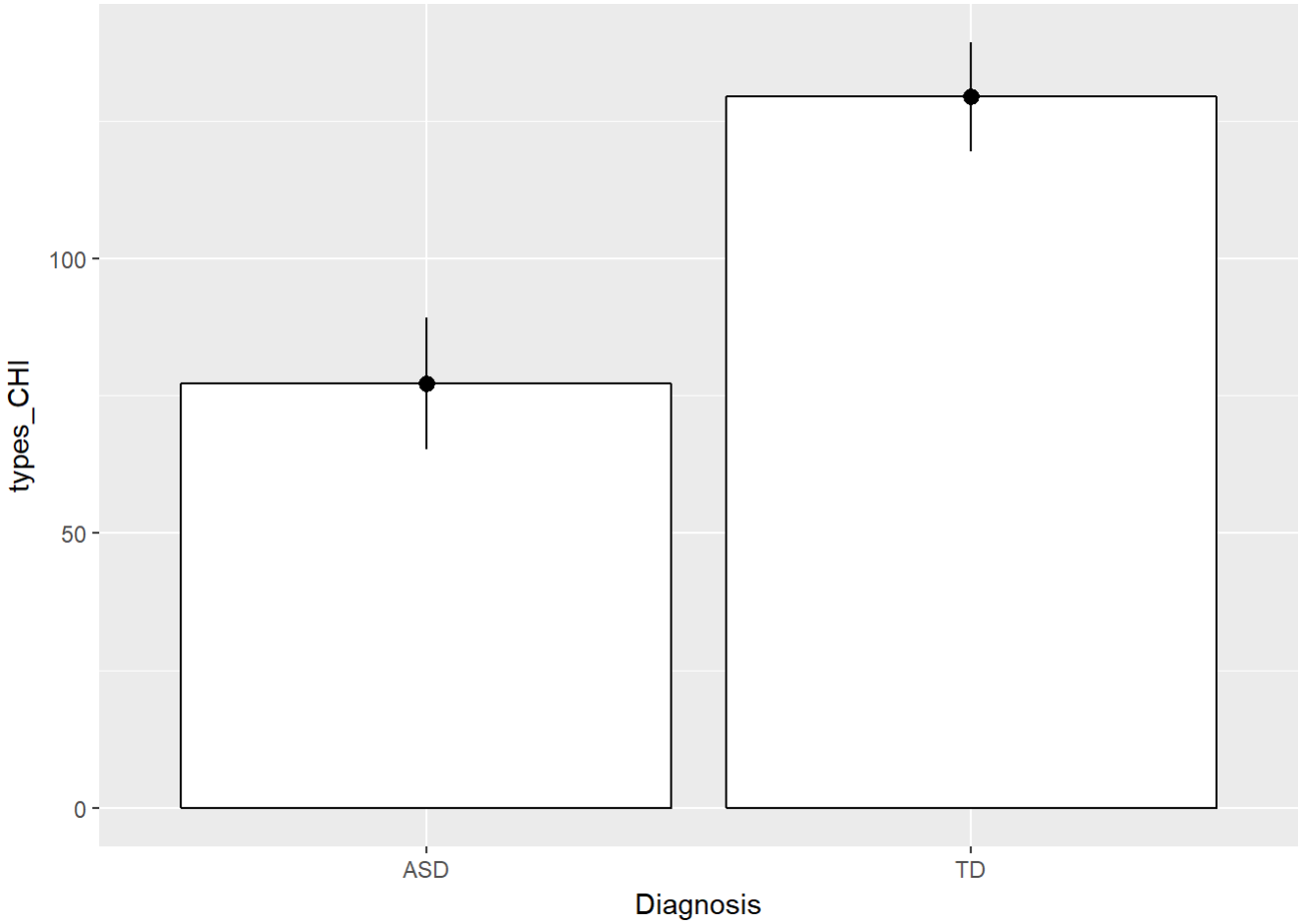
An interesting observation to be made is, that nonverbal IQ appears to be higher for children with ASD.


20
15
10
5
0
verbalIQ1
ASD
TD
Diagnosis

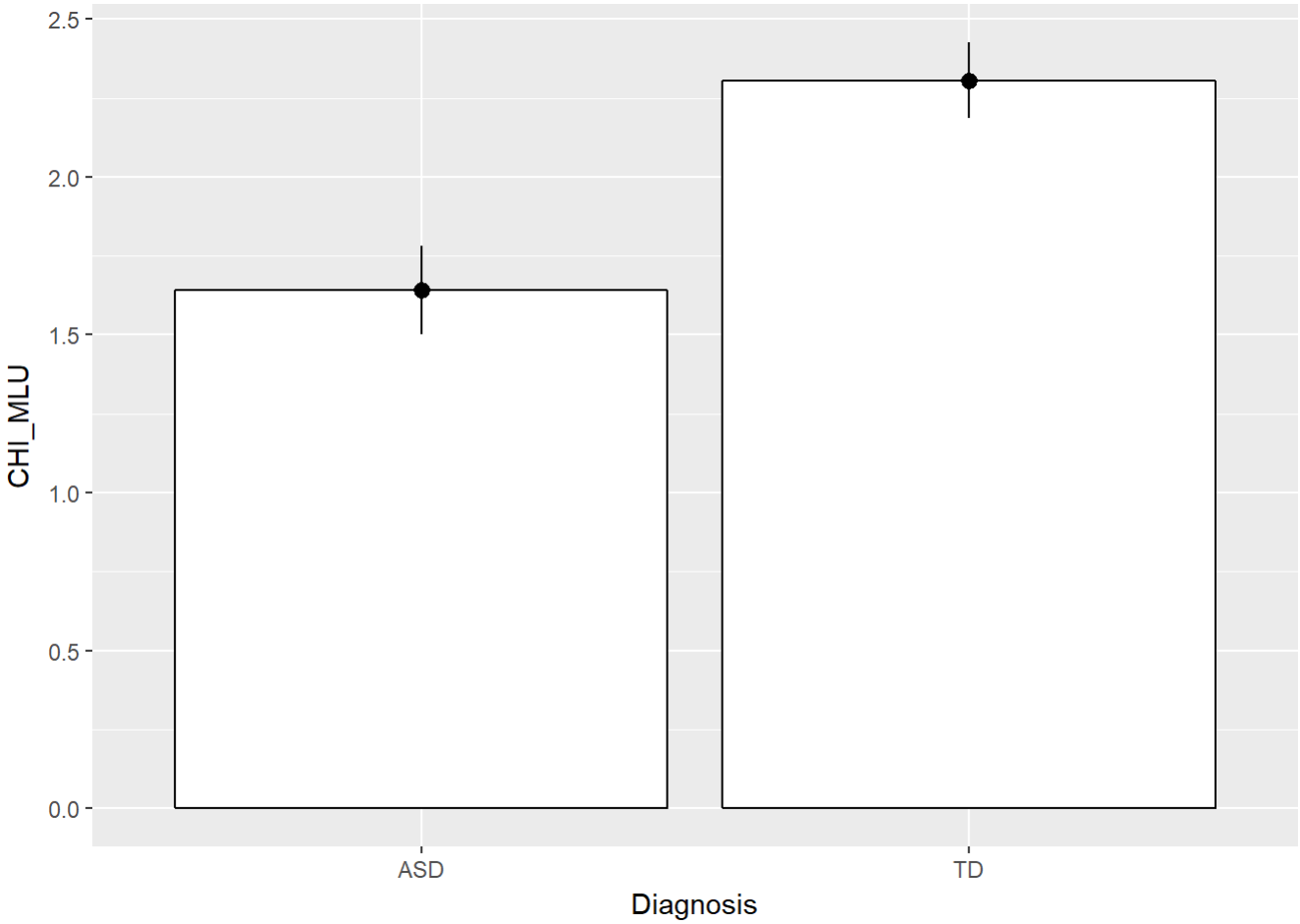


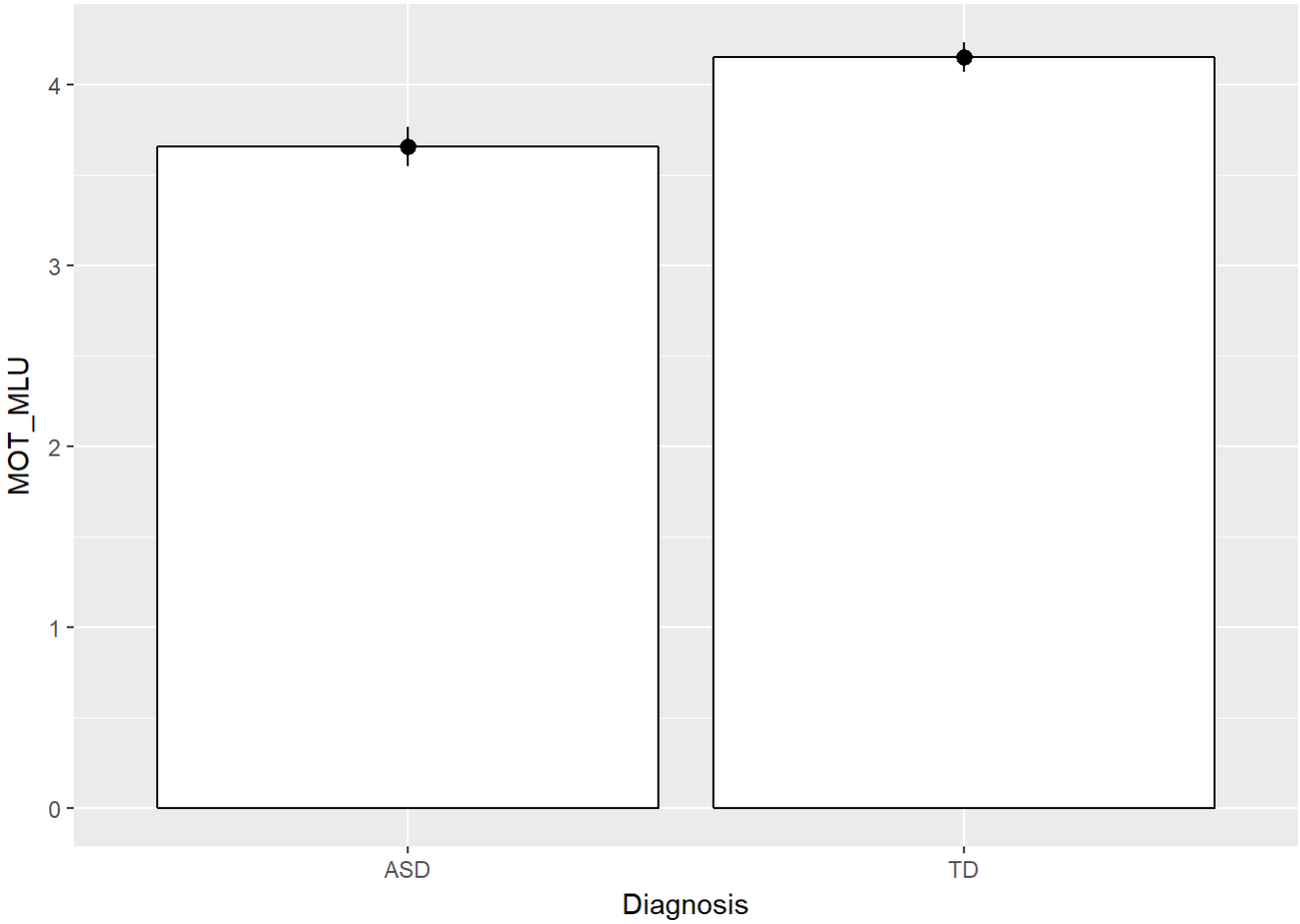
The following plots show a distribution of how many words the sample's children utter and how many unique words they utter. In line with our beliefs, children with ASD usually have smaller vocabulary and generally speak less.


500
400
300
200
100
0
tokens_CHI
ASD
TD
Diagnosis



Lastly reported are barplots of the children and their mother's s mean length of utterance. This shows that children with ASD tend to talk less in coherence - a pattern which is reflected in their mothers. Generally mothers to children with ASD also talk less in coherence. An interesting matter to investigate is whether this is an effect of the children speaking less or if the children speak less because their mothers do.


2.5
2.0
1.5
1.0
0.5
0.0
CHI_MLU
ASD
TD
Diagnosis



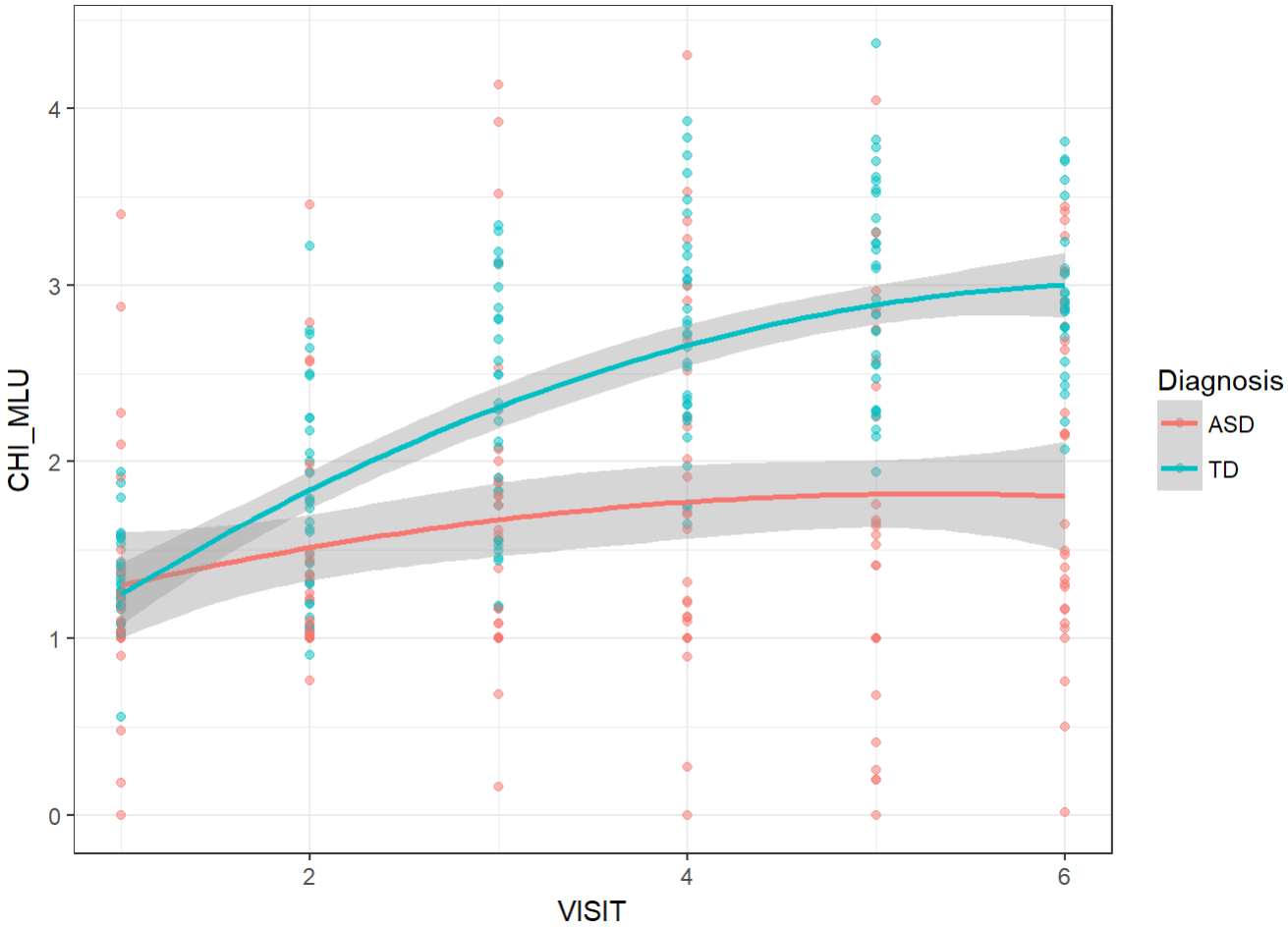
Exercise 2) Children learning language: the effects of time and ASD

Describe linguistic development in TD and ASD children in terms of Mean Length of Utterance (MLU)?

Linguistic development of childrens' MLU is affected by visit (time) and diagnosis (MLU ~ visit + diagnosis).

Using lme4 to perform a quadratic mixed effects analysis of the relationship between the children's mean length of utterance as it evolves over time and their diagnosis. As fixed effects, we entered diagnosis (ASD; autism spectrum disorder and TP; typically developing) and visit (without interaction term) into the model. As random effects, we had an intercept for subjects, as well as a by-subject random slope for the effect of visit. P-value was obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question; children's diagnosis and number of visit affected mean length of utterance ($\chi 2(5)=64.089$, $p<0.05$), we observe a significant linear increase (0.54, SE = 0.07, $p < 0.05$) accompanied by a significant quadratic component of the

growth curve (-0.04, SE = 0.009, $p < 0.05$). As can be observed in figure which indicates a slow down of the increase over time.



Exercise 3) Child directed speech as a moving target

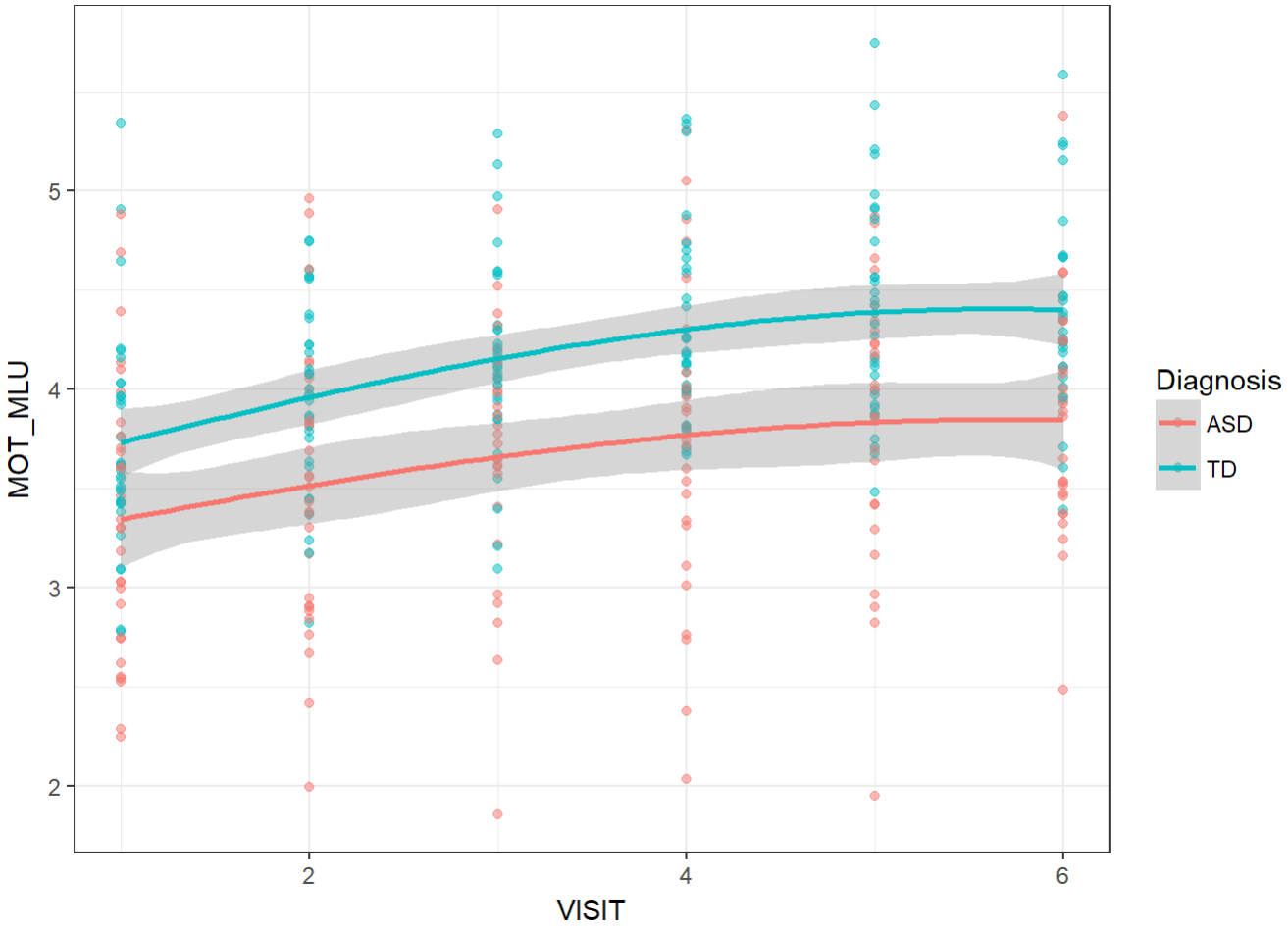
Describe how parental use of language changes over time in terms of MLU. What do you think is going on?

The parental use of language is affected by visit and diagnosis (MLU ~ visit + diagnosis) in line with their children.

Using lme4 to perform a cubic mixed effects analysis of the relationship between the mothers' mean length of utterance as it evolves over time and their childrens diagnosis. As fixed effects, we entered diagnosis (ASD; autism spectrum disorder and TP; typically developing) and visit (without interaction term) into the model. As random effects, we had an intercept for subjects, as well as a by-subject random slope for the effect of visit. P-value was obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question; children's diagnosis and number of visit significantly affected parental mean length of utterance ($\chi2(3)=24.949$, $p<0.05$), we

observe a significant linear increase for the effect of diagnosis (0.49, SE = 0.11, p > 0.05) and for the effect of visit (0.24, SE = 0.20, p < 0.05) accompanied by a significant quadratic component of the growth curve equally for the effect of visit (-0.01, SE = 0.06, p < 0.05) and a significant cubic component of the growth curve accordingly (-0.001, SE = 0.006, p < 0.05). As can be observed in figure, this indicates a slow down of the increase over time.

The similar effect of shorter mean length of utterances for parents there have children with ASD is probably a result of an adjustment to the child's progress.



Exercise 4) Looking into "individual differences" (demographic, clinical or cognitive profiles)

The dataset contains some additional variables characterizing the kids' cognitive and clinical profile: ADOS (autism severity), MSEL EL (Expressive Language, that is, verbal IQ, or linguistic skills at first visit as assessed by a psychologist using Mullen Scales of Early Learning), MSEL VR (Visual Reception, used as a proxy for non verbal IQ at first visit), Age, Gender, Ethnicity. Would it make sense to add any of them to your model of linguistic trajectories? Create the best possible model (the

one that best explain the data, with MLU as outcome). Next time your model will be tested on new participants, and we will proclaim a winner. Describe your strategy to select the best models (how did you choose the variables to include?) and send the code to Riccardo and Malthe.

As mentioned previously including ethnicity and gender as random effects will probably be beneficial. Furthermore, it could be interesting to include fixed effects such as ADOS, verbal IQ and nonverbal IQ measures.

The best model found by a serious amount of p-hacking was: Childrens mean length of utterance ~ visit + I(visit^2) + diagnosis + ados + verbalIQ + nonverbalIQ1 +(1+visit+ I(visit^2)|subject)+ (1|Gender)+ (1|Ethnicity)

As fixed effects, we entered diagnosis (ASD; autism spectrum disorder and TP; typically developing), ados (an instrument for diagnosing and assessing autism)), verbal IQ, nonverbal IQ and visit (without interaction term) into the model. As random effects, we had an intercept for subjects plus a random intercept for gender and ethnicity, as well as a by-subject random slope for the effect of visit.

P-value was obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question; children's diagnosis, ados, verbal IQ, nonverbal IQ and number of visits affected the mean length of utterance ($\chi 2(10)=28.728$, $p<0.05$), we observe a significant linear increase for the effect of visit (0.54, SE = 0.07, $p < 0.05$) accompanied by a significant quadratic component of the growth curve for the effect of visit (-0.044, SE = 0.009, $p < 0.05$). As can be observed in figure, this indicates a slow down of the increase over time. Equally, a significant linear increase is observed for the effect of diagnosis (0.04, SE=0.25, $p > 0.05$), the effect of ados (-0.003, SE = 0.017, $p > 0.05$), verbal IQ (0.06, SE = 0.009, $p > 0.05$) and nonverbal IQ (0.005, SE = 0.01, $p > 0.05$).

Here is the link to my repository: https://github.com/KiriKoppelgaard/Portfolios