Assignment 2 - Language Development in ASD - Part 1 - Explaining development

Anna, Andrea, Helle, Maria, Mia

[DATE]

# Assignment 2

In this assignment you will have to discuss a few important questions (given the data you have). More details below. The assignment submitted to the teachers consists of: - a report answering and discussing the questions (so we can assess your conceptual understanding and ability to explain and critically reflect) - a link to a git repository with all the code (so we can assess your code)

Part 1 - Basic description of language development - Describe your sample (n, age, gender, clinical and cognitive features of the two groups) and critically assess whether the groups (ASD and TD) are balanced - Describe linguistic development (in terms of MLU over time) in TD and ASD children (as a function of group). - Describe how parental use of language (in terms of MLU) changes over time. What do you think is going on? - Include individual differences in your model of language development (in children). Identify the best model.

Part 2 - Model comparison - Discuss the differences in performance of your model in training and testing data - Which individual differences should be included in a model that maximizes your ability to explain/predict new data? - Predict a new kid’s performance (Bernie) and discuss it against expected performance of the two groups

Part 3 - Simulations to plan a new study - Report and discuss a power analyses identifying how many new kids you would need to replicate the results

The following involves only Part 1.

## Learning objectives

* Summarize and report data and models
* Critically apply mixed effects (or multilevel) models
* Explore the issues involved in feature selection

# Quick recap

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.

This data is in the file you prepared in the previous class.

NB. A few children have been excluded from your datasets. We will be using them next week to evaluate how good your models are in assessing the linguistic development in new participants.

This RMarkdown file includes 1) questions (see above). Questions have to be answered/discussed in a separate document that you have to directly submit on Blackboard. 2) A break down of the questions into a guided template full of hints for writing the code to solve the exercises. Fill in the code and the paragraphs as required. Then report your results in the doc for the teachers.

REMEMBER that you will have to have a github repository for the code and submit the answers to Blackboard without code (but a link to your github/gitlab repository). This way we can check your code, but you are also forced to figure out how to report your analyses :-)

Before we get going, here is a reminder of the issues you will have to discuss in your report:

1- Describe your sample (n, age, gender, clinical and cognitive features of the two groups) and critically assess whether the groups (ASD and TD) are balanced 2- Describe linguistic development (in terms of MLU over time) in TD and ASD children (as a function of group). 3- Describe how parental use of language (in terms of MLU) changes over time. What do you think is going on? 4- Include individual differences in your model of language development (in children). Identify the best model.

# Let’s go

### Loading the relevant libraries

Load necessary libraries : what will you need? - e.g. something to deal with the data - e.g. mixed effects models - e.g. something to plot with

#load packages  
pacman::p\_load(tidyverse, MuMIn, lme4, lmerTest, purrr, effects, broom)

### Define your working directory and load the data

If you created a project for this class and opened this Rmd file from within that project, your working directory is your project directory.

If you opened this Rmd file outside of a project, you will need some code to find the data: - Create a new variable called locpath (localpath) - Set it to be equal to your working directory - Move to that directory (setwd(locpath)) - Load the data you saved last time (use read\_csv(fileName))

#load data  
df <- read\_csv("portfolio1\_data.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## Diagnosis = col\_character(),  
## Ethnicity = col\_character(),  
## Gender = col\_character()  
## )

## See spec(...) for full column specifications.

### Characterize the participants (Exercise 1)

Identify relevant variables: participants demographic characteristics, diagnosis, ADOS, Verbal IQ, Non Verbal IQ, Socialization, Visit, Number of words used, Number of unique words used, mean length of utterance in both child and parents.

Make sure the variables are in the right format.

Describe the characteristics of the two groups of participants and whether the two groups are well matched.

#looking at the data  
str(df)

## Classes 'spec\_tbl\_df', 'tbl\_df', 'tbl' and 'data.frame': 372 obs. of 21 variables:  
## $ X1 : num 1 2 3 4 5 6 7 8 9 10 ...  
## $ ID : num 1 1 1 1 1 1 2 2 2 2 ...  
## $ ADOS1 : num 0 0 0 0 0 0 13 13 13 13 ...  
## $ MullenRaw1 : num 28 28 28 28 28 28 34 34 34 34 ...  
## $ ExpressiveLangRaw1: num 14 14 14 14 14 14 27 27 27 27 ...  
## $ Socialization1 : num 108 108 108 108 108 108 85 85 85 85 ...  
## $ VISIT : num 1 2 3 4 5 6 1 2 3 4 ...  
## $ Diagnosis : chr "TD" "TD" "TD" "TD" ...  
## $ Ethnicity : chr "White" "White" "White" "White" ...  
## $ Gender : chr "M" "M" "M" "M" ...  
## $ Age : num 19.8 23.9 27.7 32.9 35.9 ...  
## $ ADOS : num 0 NA NA NA 0 NA 13 NA NA NA ...  
## $ MullenRaw : num 28 NA NA 33 NA 42 34 NA NA 49 ...  
## $ ExpressiveLangRaw : num 14 NA NA NA NA 44 27 NA NA NA ...  
## $ Socialization : num 108 110 109 102 107 100 85 105 77 75 ...  
## $ MOT\_MLU : num 3.62 3.86 4.32 4.42 5.21 ...  
## $ CHI\_MLU : num 1.25 1.01 1.56 2.25 3.24 ...  
## $ types\_MOT : num 378 403 455 533 601 595 317 307 351 335 ...  
## $ types\_CHI : num 14 18 97 133 182 210 146 171 262 200 ...  
## $ tokens\_MOT : num 1835 2160 2149 2260 2553 ...  
## $ tokens\_CHI : num 139 148 255 321 472 686 461 562 983 674 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. X1 = col\_double(),  
## .. ID = col\_double(),  
## .. ADOS1 = col\_double(),  
## .. MullenRaw1 = col\_double(),  
## .. ExpressiveLangRaw1 = col\_double(),  
## .. Socialization1 = col\_double(),  
## .. VISIT = col\_double(),  
## .. Diagnosis = col\_character(),  
## .. Ethnicity = col\_character(),  
## .. Gender = col\_character(),  
## .. Age = col\_double(),  
## .. ADOS = col\_double(),  
## .. MullenRaw = col\_double(),  
## .. ExpressiveLangRaw = col\_double(),  
## .. Socialization = col\_double(),  
## .. MOT\_MLU = col\_double(),  
## .. CHI\_MLU = col\_double(),  
## .. types\_MOT = col\_double(),  
## .. types\_CHI = col\_double(),  
## .. tokens\_MOT = col\_double(),  
## .. tokens\_CHI = col\_double()  
## .. )

#reformatting as factor  
cols <- c("ID", "VISIT", "Diagnosis", "Gender")  
df[cols] <- lapply(df[cols], as.factor)  
  
#renaming coloumns  
df <- df %>%   
 rename(verbalIQ = ExpressiveLangRaw,  
 nonverbalIQ = MullenRaw,  
 verbalIQ1 = ExpressiveLangRaw1,  
 nonverbalIQ1 = MullenRaw1)  
  
#get summary overview  
df %>%   
 split(df$Diagnosis) %>%   
 map(summary)

## $ASD  
## X1 ID ADOS1 nonverbalIQ1 verbalIQ1   
## Min. : 7.0 2 : 6 Min. : 0.00 Min. :13.00 Min. : 8.00   
## 1st Qu.:116.8 4 : 6 1st Qu.:11.00 1st Qu.:25.00 1st Qu.:11.00   
## Median :184.5 5 : 6 Median :14.00 Median :27.00 Median :16.00   
## Mean :188.4 6 : 6 Mean :14.11 Mean :26.89 Mean :17.58   
## 3rd Qu.:277.2 7 : 6 3rd Qu.:17.00 3rd Qu.:30.00 3rd Qu.:24.50   
## Max. :371.0 18 : 6 Max. :21.00 Max. :42.00 Max. :33.00   
## (Other):140   
## Socialization1 VISIT Diagnosis Ethnicity Gender Age   
## Min. : 64.00 1:31 ASD:176 Length:176 F: 26 Min. :18.77   
## 1st Qu.: 69.00 2:31 TD : 0 Class :character M:150 1st Qu.:36.88   
## Median : 76.00 3:29 Mode :character Median :42.87   
## Mean : 77.20 4:28 Mean :43.17   
## 3rd Qu.: 85.25 5:28 3rd Qu.:50.03   
## Max. :105.00 6:29 Max. :62.40   
## NA's :4   
## ADOS nonverbalIQ verbalIQ Socialization   
## Min. : 0.0 Min. :13.00 Min. : 8.00 Min. : 38.00   
## 1st Qu.:11.0 1st Qu.:27.00 1st Qu.:11.75 1st Qu.: 68.00   
## Median :14.0 Median :31.00 Median :19.50 Median : 74.00   
## Mean :13.8 Mean :33.15 Mean :22.62 Mean : 77.34   
## 3rd Qu.:17.0 3rd Qu.:40.50 3rd Qu.:30.00 3rd Qu.: 85.00   
## Max. :25.0 Max. :50.00 Max. :50.00 Max. :116.00   
## NA's :117 NA's :89 NA's :116 NA's :1   
## MOT\_MLU CHI\_MLU types\_MOT types\_CHI   
## Min. :1.856 Min. :0.000 Min. : 74.0 Min. : 0.00   
## 1st Qu.:3.222 1st Qu.:1.012 1st Qu.:284.2 1st Qu.: 9.00   
## Median :3.699 Median :1.370 Median :340.0 Median : 50.50   
## Mean :3.657 Mean :1.642 Mean :338.7 Mean : 77.21   
## 3rd Qu.:4.100 3rd Qu.:2.158 3rd Qu.:397.8 3rd Qu.:144.75   
## Max. :5.380 Max. :4.302 Max. :585.0 Max. :307.00   
## NA's :10 NA's :10 NA's :10 NA's :10   
## tokens\_MOT tokens\_CHI   
## Min. : 209 Min. : 0.0   
## 1st Qu.:1388 1st Qu.: 58.0   
## Median :1805 Median : 196.0   
## Mean :1780 Mean : 293.6   
## 3rd Qu.:2220 3rd Qu.: 482.0   
## Max. :3182 Max. :1293.0   
## NA's :10 NA's :10   
##   
## $TD  
## X1 ID ADOS1 nonverbalIQ1   
## Min. : 1.00 1 : 6 Min. : 0.000 Min. :17.00   
## 1st Qu.: 81.75 3 : 6 1st Qu.: 0.000 1st Qu.:24.00   
## Median :187.50 8 : 6 Median : 0.000 Median :27.00   
## Mean :184.76 9 : 6 Mean : 0.949 Mean :25.93   
## 3rd Qu.:284.25 10 : 6 3rd Qu.: 1.000 3rd Qu.:29.00   
## Max. :372.00 12 : 6 Max. :15.000 Max. :32.00   
## (Other):160 NA's :1   
## verbalIQ1 Socialization1 VISIT Diagnosis Ethnicity Gender   
## Min. :13.00 Min. : 84.0 1:35 ASD: 0 Length:196 F: 36   
## 1st Qu.:17.00 1st Qu.: 96.0 2:34 TD :196 Class :character M:160   
## Median :19.00 Median :102.0 3:32 Mode :character   
## Mean :20.14 Mean :100.5 4:32   
## 3rd Qu.:22.00 3rd Qu.:104.0 5:32   
## Max. :33.00 Max. :115.0 6:31   
## NA's :1   
## Age ADOS nonverbalIQ verbalIQ   
## Min. :18.07 Min. : 0.000 Min. :17.0 Min. :13.00   
## 1st Qu.:23.96 1st Qu.: 0.000 1st Qu.:28.0 1st Qu.:18.00   
## Median :30.88 Median : 0.000 Median :39.0 Median :28.00   
## Mean :30.60 Mean : 1.273 Mean :35.8 Mean :29.29   
## 3rd Qu.:36.33 3rd Qu.: 1.000 3rd Qu.:44.0 3rd Qu.:40.00   
## Max. :45.07 Max. :15.000 Max. :50.0 Max. :50.00   
## NA's :6 NA's :130 NA's :101 NA's :133   
## Socialization MOT\_MLU CHI\_MLU types\_MOT   
## Min. : 59.0 Min. :2.776 Min. :0.5584 Min. :178.0   
## 1st Qu.: 97.0 1st Qu.:3.805 1st Qu.:1.5577 1st Qu.:305.5   
## Median :102.0 Median :4.117 Median :2.3202 Median :363.5   
## Mean :101.8 Mean :4.150 Mean :2.3064 Mean :368.6   
## 3rd Qu.:107.0 3rd Qu.:4.472 3rd Qu.:2.9095 3rd Qu.:418.8   
## Max. :125.0 Max. :5.744 Max. :4.3648 Max. :601.0   
## NA's :2 NA's :10 NA's :10 NA's :10   
## types\_CHI tokens\_MOT tokens\_CHI   
## Min. : 7.00 Min. : 584 Min. : 16.0   
## 1st Qu.: 72.25 1st Qu.:1503 1st Qu.: 254.2   
## Median :131.00 Median :1844 Median : 435.0   
## Mean :129.31 Mean :1879 Mean : 475.7   
## 3rd Qu.:181.00 3rd Qu.:2264 3rd Qu.: 669.8   
## Max. :298.00 Max. :3077 Max. :1294.0   
## NA's :10 NA's :10 NA's :10

The sample included mostly young (<20) white males …

[REPORT THE RESULTS]

## Let’s test hypothesis 1: Children with ASD display a language impairment (Exercise 2)

### Hypothesis: The child’s MLU changes: i) over time, ii) according to diagnosis

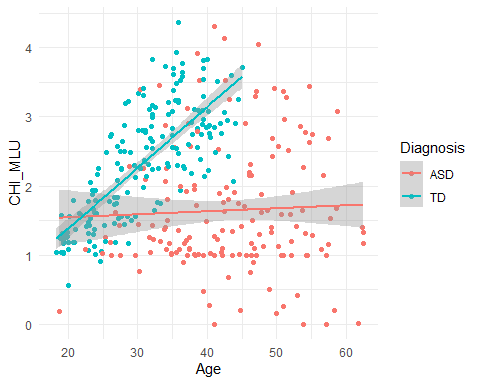
Let’s start with a simple mixed effects linear model

Remember to plot the data first and then to run a statistical test. - Which variable(s) should be included as fixed factors? - Which variable(s) should be included as random factors?

df %>%   
 ggplot() +   
 aes(x = Age, y = CHI\_MLU, color = Diagnosis) +  
 geom\_point() +  
 geom\_smooth(method = lm) +  
 theme\_minimal()

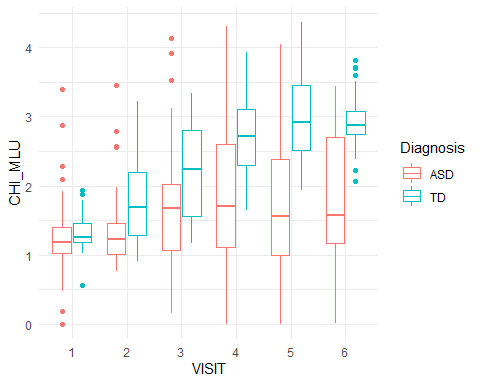
## Warning: Removed 26 rows containing non-finite values (stat\_smooth).

## Warning: Removed 26 rows containing missing values (geom\_point).



df %>%   
 ggplot() +  
 aes(x = VISIT, y = CHI\_MLU, color = Diagnosis) +  
 geom\_boxplot() +   
 theme\_minimal()

## Warning: Removed 20 rows containing non-finite values (stat\_boxplot).



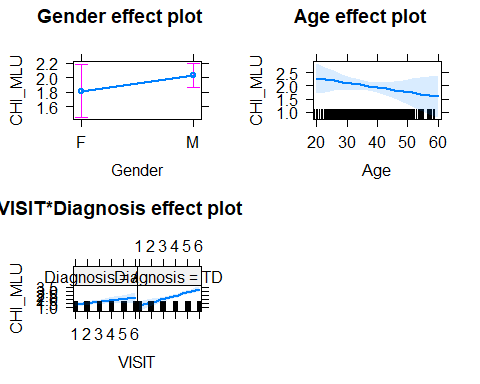
#making visit numerice  
df$VISIT <- as.numeric(df$VISIT)  
  
#make models  
m0 <- lmer(CHI\_MLU ~ Gender + Age + (1 | ID), df, REML = F)  
m1 <- update(m0, .~. + VISIT)  
m2 <- update(m1, .~. + Diagnosis)  
m3 <- update(m2, .~. + VISIT:Diagnosis)  
  
#including random slope  
m02 <- lmer(CHI\_MLU ~ Gender + Age + (1 + VISIT | ID), df, REML = F)  
m12 <- update(m02, .~. + VISIT)  
m22 <- update(m12, .~. + Diagnosis)  
m32 <- update(m22, .~. + VISIT:Diagnosis)

How would you evaluate whether the model is a good model?

#evaluate using anova function   
anova(m0, m1, m2, m3, m02, m12, m22, m32)

## Data: df  
## Models:  
## m0: CHI\_MLU ~ Gender + Age + (1 | ID)  
## m1: CHI\_MLU ~ Gender + Age + (1 | ID) + VISIT  
## m2: CHI\_MLU ~ Gender + Age + (1 | ID) + VISIT + Diagnosis  
## m02: CHI\_MLU ~ Gender + Age + (1 + VISIT | ID)  
## m3: CHI\_MLU ~ Gender + Age + (1 | ID) + VISIT + Diagnosis + VISIT:Diagnosis  
## m12: CHI\_MLU ~ Gender + Age + (1 + VISIT | ID) + VISIT  
## m22: CHI\_MLU ~ Gender + Age + (1 + VISIT | ID) + VISIT + Diagnosis  
## m32: CHI\_MLU ~ Gender + Age + (1 + VISIT | ID) + VISIT + Diagnosis +   
## m32: VISIT:Diagnosis  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## m0 5 706.67 725.90 -348.33 696.67   
## m1 6 644.67 667.75 -316.34 632.67 63.9991 1 1.245e-15 \*\*\*  
## m2 7 646.07 673.00 -316.03 632.07 0.6007 1 0.4383   
## m02 7 619.45 646.37 -302.72 605.45 26.6205 0 < 2.2e-16 \*\*\*  
## m3 8 579.98 610.75 -281.99 563.98 41.4682 1 1.198e-10 \*\*\*  
## m12 8 592.56 623.33 -288.28 576.56 0.0000 0 1.0000   
## m22 9 594.22 628.84 -288.11 576.22 0.3416 1 0.5589   
## m32 10 562.54 601.00 -271.27 542.54 33.6825 1 6.488e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot(effects:: allEffects(m32))

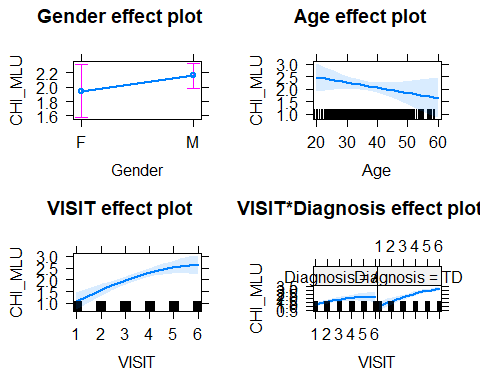


Not too good, right? Let’s check whether a growth curve model is better. Remember: a growth curve model assesses whether changes in time can be described by linear, or quadratic, or cubic (or… etc.) components. First build the different models, then compare them to see which one is better.

#making groth curve models  
growth2 <- lmer(CHI\_MLU ~ Gender + Age + VISIT\*Diagnosis + I(VISIT^2) + (1 + VISIT | ID), df, REML = F)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.00226506 (tol = 0.002, component 1)

growth3 <- lmer(CHI\_MLU ~ Gender + Age + VISIT\*Diagnosis + I(VISIT^3) + (1 + VISIT | ID), df, REML = F)  
  
#plot effects  
plot(effects:: allEffects(growth3))



#compare AIC and BIC scores  
anova(m32, growth2, growth3)

## Data: df  
## Models:  
## m32: CHI\_MLU ~ Gender + Age + (1 + VISIT | ID) + VISIT + Diagnosis +   
## m32: VISIT:Diagnosis  
## growth2: CHI\_MLU ~ Gender + Age + VISIT \* Diagnosis + I(VISIT^2) + (1 +   
## growth2: VISIT | ID)  
## growth3: CHI\_MLU ~ Gender + Age + VISIT \* Diagnosis + I(VISIT^3) + (1 +   
## growth3: VISIT | ID)  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## m32 10 562.54 601.00 -271.27 542.54   
## growth2 11 536.90 579.21 -257.45 514.90 27.6405 1 1.461e-07 \*\*\*  
## growth3 11 536.46 578.77 -257.23 514.46 0.4427 0 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

scoresAIC <- AIC(m32, growth2, growth3)  
scoresBIC <- BIC(m32, growth2, growth3)  
  
scoresAIC

## df AIC  
## m32 10 562.5393  
## growth2 11 536.8988  
## growth3 11 536.4561

scoresBIC

## df BIC  
## m32 10 601.0037  
## growth2 11 579.2096  
## growth3 11 578.7669

#growth3 model is slightly better

Exciting right? Let’s check whether the model is doing an alright job at fitting the data. Plot the actual CHI\_MLU data against the predictions of the model fitted(model).

Now it’s time to report our results. Remember to report: - the estimates for each predictor (beta estimate, standard error, p-value) - A plain word description of the results - A plot of your model’s predictions (and some comments on whether the predictions are sensible)

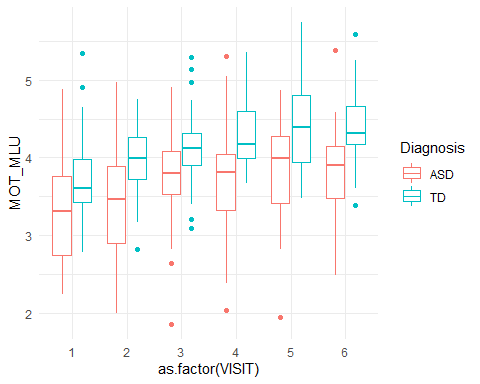
[REPORT THE RESULTS] Linguistic development of children MLU is affected by … [COMPLETE]

## Let’s test hypothesis 2: Parents speak equally to children with ASD and TD (Exercise 3)

### Hypothesis: Parental MLU changes: i) over time, ii) according to diagnosis

#plotting in relation to the hypothesis   
#boxplots  
df %>%   
 ggplot() +   
 aes(x = as.factor(VISIT), y = MOT\_MLU, color = Diagnosis) +   
 geom\_boxplot() +   
 theme\_minimal()

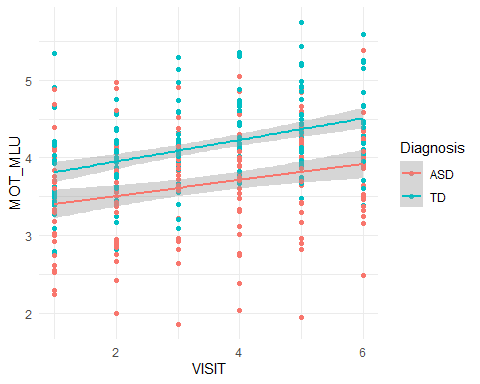
## Warning: Removed 20 rows containing non-finite values (stat\_boxplot).



#over time  
df %>%   
 ggplot() +   
 aes(x = VISIT, y = MOT\_MLU, color = Diagnosis) +  
 geom\_point() +   
 geom\_smooth(method = "lm") +  
 theme\_minimal()

## Warning: Removed 20 rows containing non-finite values (stat\_smooth).

## Warning: Removed 20 rows containing missing values (geom\_point).



#making models   
p0 <- lmer(MOT\_MLU ~ VISIT + (1 | ID), df, REML = F )  
p1 <- update(p0, .~. + Diagnosis)  
p2 <- update(p1, .~. + VISIT:Diagnosis)  
  
#with random slopes  
p02 <- lmer(MOT\_MLU ~ VISIT + (1 + VISIT | ID), df, REML = F)  
p12 <- update(p02, .~. + Diagnosis)  
p22 <- update(p12, .~. + VISIT:Diagnosis) #fails to converge

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.033954 (tol = 0.002, component 1)

#comparing  
anova(p0, p02, p1, p12, p2, p22) #p12 is best

## Data: df  
## Models:  
## p0: MOT\_MLU ~ VISIT + (1 | ID)  
## p1: MOT\_MLU ~ VISIT + (1 | ID) + Diagnosis  
## p02: MOT\_MLU ~ VISIT + (1 + VISIT | ID)  
## p2: MOT\_MLU ~ VISIT + (1 | ID) + Diagnosis + VISIT:Diagnosis  
## p12: MOT\_MLU ~ VISIT + (1 + VISIT | ID) + Diagnosis  
## p22: MOT\_MLU ~ VISIT + (1 + VISIT | ID) + Diagnosis + VISIT:Diagnosis  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## p0 4 541.93 557.39 -266.97 533.93   
## p1 5 527.46 546.78 -258.73 517.46 16.4731 1 4.934e-05 \*\*\*  
## p02 6 527.44 550.62 -257.72 515.44 2.0249 1 0.1547   
## p2 6 527.45 550.64 -257.73 515.45 0.0000 0 1.0000   
## p12 7 512.71 539.75 -249.35 498.71 16.7452 1 4.275e-05 \*\*\*  
## p22 8 513.48 544.39 -248.74 497.48 1.2267 1 0.2681   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#trying to add growth to the p12 model  
pgrowth2 <- lmer(MOT\_MLU ~ VISIT + Diagnosis + I(VISIT^2) + (1 + VISIT | ID), df, REML = F)  
pgrowth3 <- lmer(MOT\_MLU ~ VISIT + Diagnosis + I(VISIT^3) + (1 + VISIT | ID), df, REML = F)  
anova(p12, pgrowth2, pgrowth3) #the cubic growth is very slightly better than the squared

## Data: df  
## Models:  
## p12: MOT\_MLU ~ VISIT + (1 + VISIT | ID) + Diagnosis  
## pgrowth2: MOT\_MLU ~ VISIT + Diagnosis + I(VISIT^2) + (1 + VISIT | ID)  
## pgrowth3: MOT\_MLU ~ VISIT + Diagnosis + I(VISIT^3) + (1 + VISIT | ID)  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## p12 7 512.71 539.75 -249.35 498.71   
## pgrowth2 8 506.53 537.44 -245.26 490.53 8.1791 1 0.004237 \*\*   
## pgrowth3 8 506.51 537.42 -245.26 490.51 0.0168 0 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Parent MLU is affected by … but probably not … [REPORT THE RESULTS]

### Adding new variables (Exercise 4)

Your task now is to figure out how to best describe the children linguistic trajectory. The dataset contains a bunch of additional demographic, cognitive and clinical variables (e.g.verbal and non-verbal IQ). Try them out and identify the statistical models that best describes your data (that is, the children’s MLU). Describe how you selected the best model and send the code to run the model to Victor and Byurakn.

#newdf <- df[complete.cases(df[, 1:19]), ] #remove NA values in specified columns  
v1<-lmer(CHI\_MLU ~ VISIT+Gender+Diagnosis+Age+(1 | ID) , df)  
v2<-lmer(CHI\_MLU ~ VISIT+Gender+Diagnosis+Age+Socialization+(1| ID) , df, REML = F)  
v3<-lmer(CHI\_MLU ~ VISIT+Gender+Diagnosis+Age+nonverbalIQ+(1 | ID) , df, REML = F)  
v4<-lmer(CHI\_MLU ~ VISIT+Gender+Diagnosis+Age+verbalIQ+(1| ID) , df, REML = F)  
v5<-lmer(CHI\_MLU ~ VISIT+Gender+Diagnosis+Age+verbalIQ+Socialization+(1| ID) , df, REML = F)  
v6<-lmer(CHI\_MLU ~ VISIT+Gender+Diagnosis+Age+verbalIQ+as.factor(ADOS1)+(1| ID) , df, REML = F)

## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

v7<-lmer(CHI\_MLU ~ VISIT+Gender+Diagnosis+Age+verbalIQ+Socialization+as.factor(ADOS1)+(1| ID) , df, REML = F)

## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

AIC(v1,v2,v3,v4,v5,v6,v7)

## Warning in AIC.default(v1, v2, v3, v4, v5, v6, v7): models are not all fitted to  
## the same number of observations

## df AIC  
## v1 7 664.8765  
## v2 8 638.4102  
## v3 8 303.9276  
## v4 8 156.4871  
## v5 9 154.4844  
## v6 25 159.3132  
## v7 26 160.7302

BIC(v1,v2,v3,v4,v5,v6,v7)

## Warning in BIC.default(v1, v2, v3, v4, v5, v6, v7): models are not all fitted to  
## the same number of observations

## df BIC  
## v1 7 691.8015  
## v2 8 669.1585  
## v3 8 329.1076  
## v4 8 178.3767  
## v5 9 179.1102  
## v6 25 227.7182  
## v7 26 231.8714

In addition to …, the MLU of the children is also correlated with … Using AIC / nested F-tests as a criterium, we compared models of increasing complexity and found that … .

[REPORT THE RESULTS]