Nanna Kildahl Mathiasen

Céline Carlsen

11/9/2017

Assignment 1, Part 1: Language development in Autism Spectrum Disorder (ASD) - Brushing up your code skills

In this first part of the assignment we will brush up your programming skills, and make you familiar with the data sets you will be analysing for the next parts of the assignment.

In this first part of the assignment you will: 1) Create a Github account and link it to your RStudio 2) Use small nifty lines of code to transform several data sets into just one. The final data set will contain only the variables that are needed for the analysis in the next parts of the assignment 3) Become familiar with the tidyverse package, which you will find handy for later assignments.

0. First an introduction on the data

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.

1. Let's get started on GitHub

Follow the link to a Github tutorial: https://support.rstudio.com/hc/en-us/articles/200532077-Version-Control-with-Git-and-SVN

In the assignments you will be asked to upload your code on Github and the GitHub repositories will be part of the portfolio, therefore all students must make an account and link it to their RStudio (you'll thank us later for this!).

N.B. Create a GitHub repository for the Language Development in ASD set of assignments and link it to a project on your RStudio (including a working directory where you will save all your data and code for these assignments)

2. Now let's take dirty dirty data sets and make them into a tidy one

Set the working directory (the directory with your data and code for these assignments):

```
setwd("~/OneDrive/3 Semester/Experimental Methods 3/Assignment 1")
```

Load the three data sets, after downloading them from dropbox and saving them in your working directory: * Demographic data for the participants: https://www.dropbox.com/s/w15pou9wstgc8fe/demo_train.csv?dl=0 * Length of utterance data: https://www.dropbox.com/s/usyauqm37a76of6/LU_train.csv?dl=0 * Word data: https://www.dropbox.com/s/8ng1civpl2aux58/token_train.csv?dl=0

```r
data1 = read.csv("demo_train.csv")
data2 = read.csv("LU_train.csv")
data3 = read.csv("token_train.csv")
```

Explore the 3 datasets (e.g. visualize them, summarize them, etc.). You will see that the data is messy, since the psychologists collected the demographic data, a linguist analyzed the length of utterance in May 2014 and the same linguist analyzed the words several months later. In particular: - the same variables might have different names (e.g. identifier of the child) - the same variables might report the values in different ways (e.g. visit) Welcome to real world of messy data :-)

Before being able to combine the data sets we need to make sure the relevant variables have the same names and the same kind of values.

So:

2a. Find a way to transform variable names. Tip: Look into the package data.table, or google "how to rename variables in R"

```r
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.3.2
```

```r
library(plyr)

names(data1)[names(data1)=="Child.ID"] <- "ID"
names(data2)[names(data2)=="SUBJ"] <- "ID"
names(data3)[names(data3)=="SUBJ"] <- "ID"

names(data1)[names(data1)=="Visit"] <- "Visit"
names(data2)[names(data2)=="VISIT"] <- "Visit"
names(data3)[names(data3)=="VISIT"] <- "Visit"
```

2b. Find a way to homogeneize the way “visit” is reported. If you look into the original data sets, you will see that in the LU data and the Token data, Visits are called “visit 1” in stead of just 1 (which is the case in the demographic data set). Tip: There is a package called stringr, which will be very handy for you also in future assignments. We will return to this package later, but for now use the str_extract () to extract only the number from the variable Visit in each data set. Tip: type ? str_extract() after loading the library, for examples of how to use it.

```
library(stringr)
data1$Visit = str_extract(data1$Visit, "\\d")
data2$Visit = str_extract(data2$Visit, "\\d")
data3$Visit = str_extract(data3$Visit, "\\d")
```

2c. We also need to make a small adjustment to the content of the Child.ID coloumn in the demographic data. Within this column, names that are not abbreviations do not end with “.” (i.e. Adam), which is the case in the other two data sets (i.e. Adam.). If The content of the two variables isn’t identical the data sets will not be merged sufficiently. We wish to remove the “.” at the end of names in the LU data and the tokens data. To do these a subfunction of apply(), called sapply() can be used.

Tip: Take a look into the gsub() function. Tip: A possible solution has one line of code for each child name that is to be changed. Another combines mutate() and recode()

Tip: You will have to do identical work for both data sets, so to save time on the copy/paste use the cmd+f/ctrl+f function. Add the data frame name (e.g. LU_data) in the first box, and the data frame name (e.g. Tokens_data) you wish to change it to in the other box, and press replace.

```
data1$ID = gsub("\\.","",data1$ID)
data2$ID = gsub("\\.","",data2$ID)
data3$ID = gsub("\\.","",data3$ID)
```

2d. Now that the nitty gritty details of the different data sets are fixed, we want to make a subset of each data set only containig the variables that we wish to use in the final data set. For this we use the tidyverse package, which contain the function select().

The variables we need are: Child.ID, Visit, Ethnicity, Diagnosis, Gender, Age, ADOS, MullenRaw, ExpressiveLangRaw, MOT_MLU, MOT_LUstd, CHI_MLU, CHI_LUstd, types_MOT, types_CHI, tokens_MOT, tokens_CHI.

- ADOS indicates the severity of the autistic symptoms (the higher the worse)
- MullenRaw indicates non verbal IQ
- ExpressiveLangRaw indicates verbal IQ
- MLU stands for mean length of utterance
- types stands for unique words (e.g. even if "doggie" is used 100 times it only counts for 1)
- tokens stands for overall amount of words (if "doggie" is used 100 times it counts for 100)

It would be smart to rename the MullenRaw and ExpressiveLangRaw into something you can remember (i.e. nonVerbalIQ, verbalIQ)

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.3.2
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Warning: package 'readr' was built under R version 3.3.2
```

```
## Warning: package 'purrr' was built under R version 3.3.2
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## arrange():   dplyr, plyr
## between():   dplyr, data.table
## compact():   purrr, plyr
## count():     dplyr, plyr
## failwith():  dplyr, plyr
## filter():    dplyr, stats
## first():     dplyr, data.table
## id():        dplyr, plyr
## lag():       dplyr, stats
## last():      dplyr, data.table
## mutate():    dplyr, plyr
## rename():    dplyr, plyr
## summarise(): dplyr, plyr
## summarize(): dplyr, plyr
## transpose(): purrr, data.table
```

```
data1_sub = select(data1, ID, Visit, Ethnicity, Diagnosis, Gender, Age, ADOS, MullenRaw, Expres
siveLangRaw)
data2_sub = select(data2, ID, Visit, MOT_MLU, MOT_LUstd, CHI_MLU, CHI_LUstd)
data3_sub = select(data3, ID, Visit, types_MOT, types_CHI, tokens_MOT, tokens_CHI)

names(data1_sub)[names(data1_sub)=="MullenRaw"] <- "nonverbalIQ"
names(data1_sub)[names(data1_sub)=="ExpressiveLangRaw"] <- "verbalIQ"
```

2e. Finally we are ready to merge all the data sets into just one. Google “How to merge datasets in R” Tip: Use the merge() function for this. Tip: Merge only works for two data frames at the time. Tip: Check the number of observations in the datasets before and after merging. What is going on?

```
data4 = merge(data1_sub, data2_sub)
finaldata = merge(data4, data3_sub)
```

Are we done yet?

If you look at the data set now, you'll se a lot of NA's in the variables ADOS, nonVerbalIQ (MullenRaw) and verbalIQ (ExpressiveLangRaw). These measures were not taken at all visits. Additionally, we only want these measures for the first visit (Riccardo will explain why in class). So let's make sure that we select only these variables as collected during the first visit for each child and repeat these values throughout all other visits.

Tip: one solution requires you to select only the rows corresponding to visit 1 in a new dataset, to rename the columns of the relevant variables and to merge it back to the old dataset. Tip: subset() and select() might be useful. Tip: the final dataset should have as many rows as the the old one.

```
finaldata_sub = subset(finaldata, Visit == "1")
finaldata_sub = select(finaldata_sub, ID, ADOS, nonverbalIQ, verbalIQ)

finaldata = finaldata[,-7:-9]
finaldata_1 = merge(finaldata, finaldata_sub)
```

Now, we are almost ready to actually start working with the data. However, here are some additional finishing touches:

- in some experiments your participants must be anonymous. Therefore we wish to turn the CHILD.ID into numbers. Tip: as.numeric() might be a useful function, but not alone.
- Note that visit is (probably) not defined as numeric. Turn it into a numeric variable
- In order to make it easier to work with this nice, clean dataset in the future, it is practical to make sure the variables have sensible values. E.g. right now gender is marked 1 and 2, but in two weeks you will not be able to remember, which gender were connected to which number, so change the values from 1 and 2 to F and M in the gender variable. For the same reason, you should also change the values of Diagnosis from A and B to ASD (autism spectrum disorder) and TD (typically developing). Tip: Google "how to rename levels in R".

```
finaldata_1$ID = as.numeric(as.factor(finaldata_1$ID))

finaldata_1$Visit = as.numeric(finaldata_1$Visit)

finaldata_1$Gender = as.factor(finaldata_1$Gender)
finaldata_1 <- transform(finaldata_1,
          Gender=revalue(Gender,c("1"="M", "2"="F")))

finaldata_1 <- transform(finaldata_1,
          Diagnosis=revalue(Diagnosis,c("A"="ASD", "B"="TD")))
```

Save the data set using into a csv file. Hint: look into write.csv()

```
write.csv(finaldata_1, file = "finaldata.csv", sep = ",")
```

```
## Warning in write.csv(finaldata_1, file = "finaldata.csv", sep = ","):
## attempt to set 'sep' ignored
```

3. Now that we have a nice clean data set to use for the analysis next week, we shall play a bit around with it. The following exercises are not relevant for the analysis, but are here so you can get familiar with the functions within the tidyverse package.

Here's the link to a very helpful book, which explains each function: http://r4ds.had.co.nz/index.html

1. USING FILTER List all kids who:

1. have a mean length of utterance (across all visits) of more than 2.7 morphemes.
2. have a mean length of utterance of less than 1.5 morphemes at the first visit
3. have not completed all trials. Tip: Use pipes to solve this

```
filter(finaldata_1, CHI_MLU > 2.7)
```

```
##     ID Visit     Ethnicity Diagnosis Gender   Age  MOT_MLU MOT_LUstd
## 1    1     5         White        TD      M 35.90 5.209615  2.814165
## 2    1     6         White        TD      M 40.13 4.664013  2.765261
## 3    2     2         White       ASD      M 33.17 4.964664  2.499220
## 4    2     3         White       ASD      M 37.07 4.147059  2.803222
## 5    2     4         White       ASD      M 41.07 5.309804  2.842621
## 6    2     6         White       ASD      M 49.70 4.588477  2.783585
## 7    3     4         White        TD      F 35.53 5.301053  2.912026
## 8    3     5         White        TD      F 39.47 4.566038  2.792687
## 9    3     6         White        TD      F 45.07 5.229885  3.014147
## 10   4     3 White/Latino       ASD      M 38.90 3.818681  2.420551
## 11   4     4 White/Latino       ASD      M 43.13 4.301624  2.367015
## 12   4     5 White/Latino       ASD      M 47.40 4.602851  2.611098
## 13   4     6 White/Latino       ASD      M 51.37 3.532374  2.574647
## 14   5     4         White       ASD      M 45.53 4.744000  2.538989
## 15   5     6         White       ASD      M 54.13 4.587179  2.685925
## 16   8     4         White        TD      M 32.07 4.658333  2.519080
## 17   9     4         White        TD      F 31.07 4.262195  2.494374
## 18   9     5         White        TD      F 35.00 4.384946  2.698504
## 19  10     6         White        TD      M 40.43 4.235585  2.511690
## 20  11     6         White        TD      M 40.27 4.287582  2.747960
## 21  12     3         White        TD      M 28.27 4.974684  2.728076
## 22  12     4         White        TD      M 32.07 3.988304  2.626870
## 23  12     5         White        TD      M 35.87 4.910494  2.669214
## 24  12     6         White        TD      M 41.50 4.468493  2.574313
## 25  13     4         White        TD      M 31.03 4.083700  2.429695
## 26  13     5         White        TD      M 35.37 4.487842  2.740111
## 27  13     6         White        TD      M 39.40 4.847418  2.694579
## 28  14     5         White        TD      M 35.10 4.446281  2.396452
```

```
## 29 14      6         White        TD       M 39.43 4.664286  2.732374
## 30 15      6         White        TD       M 40.30 4.347709  2.570906
## 31 16      2         White        TD       M 26.27 4.750455  2.491813
## 32 16      3         White        TD       M 30.63 4.164789  2.751112
## 33 16      5         White        TD       M 38.17 5.433579  2.515307
## 34 16      6         White        TD       M 42.93 4.445872  2.582458
## 35 18      2         White        TD       M 26.13 4.357911  2.588408
## 36 18      3         White        TD       M 30.03 4.116057  2.456142
## 37 18      4         White        TD       M 34.43 4.131579  2.684469
## 38 18      5         White        TD       M 38.70 3.877102  2.561362
## 39 18      6         White        TD       M 44.07 4.013353  2.749934
## 40 20      6         White       ASD       M 37.30 5.379798  2.712708
## 41 22      5         White        TD       F 35.13 4.746606  2.668579
## 42 22      6         White        TD       F 39.23 4.211321  2.707896
## 43 25      3         White        TD       M 30.13 4.127941  2.433186
## 44 25      4         White        TD       M 34.00 5.362500  2.914683
## 45 25      5         White        TD       M 37.93 4.267409  2.686150
## 46 25      6         White        TD       M 42.47 4.472993  2.718080
## 47 26      1         White       ASD       M 30.40 4.690751  2.545488
## 48 26      3         White       ASD       M 38.60 4.316279  2.713869
## 49 26      4         White       ASD       M 42.63 4.857143  2.675294
## 50 26      5         White       ASD       M 46.93 4.345515  2.772711
## 51 26      6         White       ASD       M 51.00 4.111413  2.889464
## 52 28      5         White        TD       M 36.43 3.708934  2.370327
## 53 28      6         White        TD       M 39.43 4.239437  2.588672
## 54 31      3         White        TD       M 32.07 3.552459  2.378463
## 55 31      4         White        TD       M 35.03 3.667638  2.511264
## 56 37      5         White       ASD       M 55.17 4.658802  2.756894
## 57 37      6         White       ASD       M 58.77 4.240798  2.555277
## 58 38      3         White        TD       F 28.07 4.577491  2.647231
## 59 38      5         White        TD       F 36.13 4.917927  2.721358
```

```
## 60 40      4      Lebanese       ASD       M    NA 3.812930  2.397431
## 61 41      6         White        TD       M 39.93 4.061475  2.517218
## 62 42      4         White        TD       M 32.07 4.611247  2.590743
## 63 42      5         White        TD       M 36.40 3.921444  2.453860
## 64 42      6         White        TD       M 40.13 3.391525  2.517921
## 65 43      5         White       ASD       M 53.40 4.232258  2.305867
## 66 44      5         White        TD       M 35.83 4.113846  2.521775
## 67 47      4         White        TD       F 32.13 4.190698  2.622072
## 68 47      5         White        TD       F 39.10 3.673418  2.398771
## 69 47      6         White        TD       F 40.37 4.676101  2.442830
## 70 49      3         White        TD       M 31.63 4.737127  2.743557
## 71 49      4         White        TD       M 35.63 4.880240  2.722280
## 72 49      5         White        TD       M 39.47 5.743772  2.723008
## 73 49      6         White        TD       M 43.40 5.247093  2.733299
## 74 51      4         Asian        TD       F 34.43 4.137014  2.600857
## 75 51      5         Asian        TD       F 37.67 5.185941  2.686471
## 76 51      6         Asian        TD       F 42.10 5.153639  2.756146
## 77 52      3         White        TD       M 30.77 5.288991  2.867721
## 78 52      4         White        TD       M 35.03 5.338462  2.885118
## 79 52      5         White        TD       M 38.60 4.983389  2.948008
## 80 53      2         White        TD       M 28.60 4.557471  2.480875
## 81 53      3         White        TD       M 32.50 4.078292  2.509964
## 82 53      4         White        TD       M 36.40 4.458937  2.765733
## 83 53      5         White        TD       M 40.07 4.857143  2.619025
## 84 53      6         White        TD       M 44.43 3.706790  2.400582
## 85 54      4         White        TD       M 30.83 3.819961  2.452458
## 86 54      5         White        TD       M 35.17 3.750000  2.519183
## 87 54      6         White        TD       M 39.30 4.186161  2.590683
## 88 55      4         White        TD       M 32.03 4.177340  2.596526
## 89 55      6         White        TD       M 41.93 3.957230  2.537740
## 90 59      4         White        TD       M 33.60 4.697624  2.404849
```

```
## 91 59      6         White         TD       M 41.00 4.113158  2.372060
## 92 60      1         White        ASD       M 34.00 3.604140  2.550110
## 93 60      2         White        ASD       M 38.63 4.604341  2.561833
## 94 60      3         White        ASD       M 42.47 4.907591  2.667128
## 95 60      4         White        ASD       M 47.00 4.085409  2.536530
## 96 60      5         White        ASD       M 51.13 4.223572  2.645720
## 97 60      6         White        ASD       M 54.73 4.080446  2.591821
##     CHI_MLU CHI_LUstd types_MOT types_CHI tokens_MOT tokens_CHI ADOS
## 1  3.238095  2.355940       601       182       2553        472    0
## 2  2.865169  2.247884       595       210       2586        686    0
## 3  3.453039  2.267779       307       171       1270        562   13
## 4  3.119318  2.206002       351       262       1445        983   13
## 5  4.302326  2.531521       335       200       1286        674   13
## 6  3.413502  2.323497       304       245        999        698   13
## 7  3.929204  2.673094       449       206       2397        754    1
## 8  3.298578  2.122091       534       207       2672        588    1
## 9  3.710345  2.162110       486       173       2564        460    1
## 10 3.518072  2.674563       388       165       1788        490    8
## 11 3.257143  2.208333       356       163       1711        479    8
## 12 4.043478  2.367869       397       146       2082        539    8
## 13 3.278195  2.537992       410       166       2171        738    8
## 14 2.907258  2.374993       384       187       2685        604    9
## 15 2.766520  2.358437       462       179       3182        538    9
## 16 3.026596  2.045840       375       134       2069        493    5
## 17 2.775641  2.030453       400       121       1934        390    0
## 18 2.835821  1.913011       428       121       1879        346    0
## 19 2.705128  1.868147       400        73       2271        189    3
## 20 2.757143  2.195357       260       168       1138        666    0
## 21 3.185771  2.163540       318       169       1463        733    0
## 22 3.000000  2.175935       197       122        632        243    0
## 23 4.364754  2.577119       290       222       1467        916    0
```

```
## 24 3.504950  2.503952       339       201       1498        640    0
## 25 2.798283  2.025078       463       203       2361        538    0
## 26 3.230114  2.483102       511       247       2668        932    0
## 27 3.701195  2.707956       388       210       1959        864    0
## 28 3.375000  2.348335       390       238       1864        755    0
## 29 3.811404  2.794803       359       178       1802        793    0
## 30 2.869048  2.086121       327       156       1395        410    0
## 31 2.744108  1.909177       391       196       2303        733    0
## 32 2.807692  2.169244       408       200       2675        825    0
## 33 3.109524  2.134168       517       219       2762        583    0
## 34 2.948207  2.324584       491       250       2264        710    0
## 35 2.722034  2.019571       461       160       2687        738    0
## 36 3.340000  2.284236       487       201       2479        940    0
## 37 3.212821  2.422827       565       207       2965       1092    0
## 38 3.090278  2.369989       575       219       2881        769    0
## 39 2.909535  2.244065       516       235       2576       1079    0
## 40 2.902778  2.499035       433       155       2389        590    9
## 41 3.700000  2.282542       436       178       1965        686    0
## 42 3.091195  2.280872       478       221       2044        847    0
## 43 2.804217  2.144111       455       217       2589        826    0
## 44 3.731092  2.895077       553       291       2978       1225    0
## 45 2.741379  2.069361       578       298       2940       1145    0
## 46 3.061453  2.368702       555       237       2895       1010    0
## 47 3.400000  1.722577       278       119       1450        483   11
## 48 3.919689  2.370211       333       307       1668       1293   11
## 49 3.523810  2.537698       398       188       2518        714   11
## 50 3.291990  2.229706       437       261       2410       1154   11
## 51 3.364341  2.596772       452       273       3076       1249   11
## 52 3.585253  1.958878       272       177       1098        622    0
## 53 2.960784  2.244477       391       164       1889        521    0
## 54 2.987526  2.167577       396       233       1872       1246    0
```

```
## 55 2.727273  2.173550       480        186        2233        750    0
## 56 2.746875  2.228015       407        228        2314        815    7
## 57 3.077419  2.352858       429        217        2510        897    7
## 58 2.869048  2.421529       420        175        2146        825    1
## 59 3.518519  2.542533       447        210        1999        800    1
## 60 2.990000  1.989950       430        180        2377        493   13
## 61 2.852632  1.957272       397        213        1741        702    0
## 62 3.830040  2.388784       339        193        1726        820    0
## 63 3.774908  2.596669       358        213        1600        875    0
## 64 3.072797  2.518603       357        219        1764        719    0
## 65 2.866505  2.185376       396        262        2326       1054   14
## 66 3.200000  2.780354       307        131        1207        433    3
## 67 3.162242  2.282469       343        211        1559        973    0
## 68 2.738806  2.181774       315        131        1224        304    0
## 69 2.760000  2.025438       322         34        1371         61    0
## 70 3.112450  1.990793       323        151        1609        671    0
## 71 3.480952  2.573143       340        229        1452        684    0
## 72 3.537143  2.669948       425        209        1525        586    0
## 73 3.595000  2.406444       383        217        1701        646    0
## 74 3.028986  2.349659       363        112        1829        348    1
## 75 2.921569  2.131401       388        108        1986        238    1
## 76 2.761290  2.146517       383        140        1866        395    1
## 77 3.303754  2.271628       474        200        2142        897    0
## 78 3.077551  2.265560       554        177        2585        642    0
## 79 2.832168  2.293152       563        219        2772        723    0
## 80 3.217949  2.000945       197        126         686        449    0
## 81 3.131356  2.255810       219        152        1020        660    0
## 82 3.634069  2.377306       232        217         798        978    0
## 83 3.822581  2.490460       278        217        1067        814    0
## 84 3.243902  2.224554       249        102        1024        358    0
## 85 3.406504  1.899240       332        153        1699        694    1
```

```
## 86 3.607287  2.343832       394        244        1977          725    1
## 87 2.892157  2.215838       437        183        2347          517    1
## 88 2.862069  1.964134       318        205        1524         1051    0
## 89 2.909274  2.178949       367        260        1731         1294    0
## 90 2.714829  2.119628       330        157        1974          637    4
## 91 2.848000  2.169077       303        158        1460          659    4
## 92 2.876344  1.917878       400        149        2587          469   13
## 93 2.784000  2.214801       413        149        2534          670   13
## 94 4.131868  2.436936       459        196        2841          698   13
## 95 3.359833  2.479248       539        214        3163          693   13
## 96 2.965517  2.166879       521        145        3090          357   13
## 97 3.441558  2.434387       505        226        3072          713   13
##    nonverbalIQ verbalIQ
## 1           28       14
## 2           28       14
## 3           34       27
## 4           34       27
## 5           34       27
## 6           34       27
## 7           29       18
## 8           29       18
## 9           29       18
## 10          31       27
## 11          31       27
## 12          31       27
## 13          31       27
## 14          34       27
## 15          34       27
## 16          32       31
## 17          24       18
## 18          24       18
```

```
## 19             27        18
## 20             21        15
## 21             30        16
## 22             30        16
## 23             30        16
## 24             30        16
## 25             25        17
## 26             25        17
## 27             25        17
## 28             23        17
## 29             23        17
## 30             24        15
## 31             29        26
## 32             29        26
## 33             29        26
## 34             29        26
## 35             29        33
## 36             29        33
## 37             29        33
## 38             29        33
## 39             29        33
## 40             26        14
## 41             21        19
## 42             21        19
## 43             29        22
## 44             29        22
## 45             29        22
## 46             29        22
## 47             32        33
## 48             32        33
## 49             32        33
```

```
## 50         32      33
## 51         32      33
## 52         20      16
## 53         20      16
## 54         27      22
## 55         27      22
## 56         33      26
## 57         33      26
## 58         29      28
## 59         29      28
## 60         27      13
## 61         25      17
## 62         24      19
## 63         24      19
## 64         24      19
## 65         42      27
## 66         30      20
## 67         27      20
## 68         27      20
## 69         27      20
## 70         27      27
## 71         27      27
## 72         27      27
## 73         27      27
## 74         22      14
## 75         22      14
## 76         22      14
## 77         29      22
## 78         29      22
## 79         29      22
## 80         30      30
```

```
## 81          30        30
## 82          30        30
## 83          30        30
## 84          30        30
## 85          24        22
## 86          24        22
## 87          24        22
## 88          26        17
## 89          26        17
## 90          29        22
## 91          29        22
## 92          30        30
## 93          30        30
## 94          30        30
## 95          30        30
## 96          30        30
## 97          30        30
```

```
filter(finaldata_1, Visit == 1, CHI_MLU < 1.5)
```

```
##     ID Visit        Ethnicity Diagnosis Gender   Age  MOT_MLU MOT_LUstd
## 1    1     1            White        TD      M 19.80 3.621993  2.164553
## 2    5     1            White       ASD      M 34.03 3.986357  2.500713
## 3    6     1      Bangladeshi       ASD      F 26.17 2.618729  1.935874
## 4    7     1            White       ASD      F 41.00 2.244755  1.510878
## 5    9     1            White        TD      F 18.30 3.544419  2.272387
## 6   10     1            White        TD      M 19.27 4.204846  2.384767
## 7   11     1            White        TD      M 19.23 3.380463  2.214518
## 8   12     1            White        TD      M 20.07 4.195335  2.280551
## 9   14     1            White        TD      M 18.97 3.420315  2.273399
```

```
## 10 15      1            White        TD       M 19.27 3.967078  2.302921
## 11 17      1            White       ASD       M 34.80 3.182390  2.269630
## 12 19      1            White       ASD       M 35.80 2.539823  1.994618
## 13 20      1            White       ASD       M 18.77 2.524740  2.141337
## 14 21      1 African American       ASD       M 27.53 4.390879  3.068753
## 15 22      1            White        TD       F 18.93 3.630476  2.452720
## 16 23      1     White/Latino       ASD       M 27.37 3.024548  2.278008
## 17 24      1            White       ASD       M 37.47 2.917355  2.073392
## 18 27      1            White        TD       M 21.03 3.616867  2.261294
## 19 28      1            White        TD       M 19.93 2.788793  2.055953
## 20 29      1            White       ASD       M 34.87 3.304189  2.372505
## 21 30      1            White       ASD       M 36.53 3.607088  2.340376
## 22 33      1 African American       ASD       F 25.33 2.287293  1.928359
## 23 34      1            White       ASD       M 33.77 2.743455  1.892987
## 24 35      1            White        TD       M 19.30 3.561364  2.229117
## 25 36      1            White        TD       M 19.20 3.921109  2.376179
## 26 37      1            White       ASD       M 39.50 4.135036  2.563076
## 27 38      1            White        TD       F 19.87 3.420975  2.244646
## 28 39      1      White/Asian       ASD       M 33.20 3.298748  1.995123
## 29 40      1         Lebanese       ASD       M 24.90 2.997050  2.202839
## 30 41      1            White        TD       M 19.23 3.093146  2.362590
## 31 42      1            White        TD       M 19.37 4.033333  2.460373
## 32 44      1            White        TD       M 19.77 3.088757  2.098655
## 33 45      1            White        TD       M 20.03 2.776181  1.904606
## 34 46      1            White       ASD       M 36.73 4.883966  2.773678
## 35 47      1            White        TD       F 20.03 3.943005  2.384506
## 36 48      1            White       ASD       M 31.63 3.765528  2.329794
## 37 49      1            White        TD       M 23.07 4.030075  2.695787
## 38 50      1            White       ASD       M 37.47 3.704110  2.276770
## 39 51      1            Asian        TD       F 20.87 3.435743  2.257715
## 40 52      1            White        TD       M 22.57 5.344227  2.917012
```

```
## 41 54      1            White        TD      M 19.10 3.487871  2.432205
## 42 55      1            White        TD      M 19.97 3.509138  2.189292
## 43 56      1            White       ASD      M 35.50 2.548969  1.706620
## 44 57      1            White       ASD      F 41.07 3.833770  2.417727
## 45 58      1            White       ASD      M 26.00 2.747100  1.853343
## 46 59      1            White        TD      M 20.80 3.432387  2.124274
## 47 61      1            White       ASD      M 42.00 3.030405  2.107521
##      CHI_MLU CHI_LUstd types_MOT types_CHI tokens_MOT tokens_CHI ADOS
## 1  1.2522523 0.4739801       378        14       1835        139    0
## 2  1.3947368 0.6897549       324        57       2859        197    9
## 3  1.0000000 0.0000000       212         4        761         29   17
## 4  1.2641509 0.7177255       152        29        578        130   18
## 5  1.0378788 0.1909031       363        36       1408        137    0
## 6  1.0375000 0.1899836       289        15       1808         83    3
## 7  1.2168675 0.4121116       215        24       1136        101    0
## 8  1.0877193 0.2828862       235        17       1262         62    0
## 9  1.0396040 0.1950269       287         7       1625        105    0
## 10 1.1647059 0.4813476       277        27       1643         99    0
## 11 1.0277778 0.1643355       281         9       1418         37   14
## 12 1.3595506 0.7606501       283        89       1019        227   11
## 13 0.1857143 0.4564727       321        16       1787        214    9
## 14 1.0000000 0.0000000       485         8       2826        122   21
## 15 1.2371134 0.4926929       343        36       1698        118    0
## 16 1.4324324 1.1159904       328        41       2138        103   14
## 17 1.0833333 0.3996526       193         6        654         26   20
## 18 1.3661972 0.6763497       317        37       1361         95    0
## 19 1.2758621 0.4719647       178        34        584        111    0
## 20 1.0086207 0.0924466       295         6       1643        117   17
## 21 0.9000000 0.4358899       366        11       2054         21   12
## 22 1.2500000 0.5747670       206        13        788         35   14
## 23 1.3766234 0.6353026       214        67        893        180   10
```

```
## 24 1.2641509 0.5546015       291        24        1344        260    1
## 25 1.2307692 0.5756396       281         8        1631         16    3
## 26 0.4805825 0.7221272       381        39        1988        337    7
## 27 1.3322785 0.7633014       342        96        2035        398    1
## 28 1.2043011 0.5784507       274        36        1537        109   11
## 29 1.0175439 0.1312862       252         9        1827         58   13
## 30 1.0262009 0.2791346       333        32        1547        235    0
## 31 1.3034483 0.7818999       373        47        2334        176    0
## 32 1.2592593 0.4382281       275         9        1417         68    3
## 33 0.5584416 0.7811664       258        20        1188         91    5
## 34 1.1666667 0.5000000       387        10        2144         63   20
## 35 1.0761421 0.3010798       260        13        1347        212    0
## 36 1.2500000 0.6123724       303        17        2147         40   17
## 37 1.4258373 1.0238787       255        73        1452        286    0
## 38 1.1000000 0.4898979       265         8        1215         43   19
## 39 1.1818182 0.4575657       331         7        1503         38    1
## 40 1.4086957 0.8433438       441        92        2267        319    0
## 41 1.3139535 0.7664723       214        32        1118        109    1
## 42 1.1846154 0.3879852       178        11        1144        154    0
## 43 1.0444444 0.2060804       195         9         955         47   14
## 44 0.0000000 0.0000000       386         0        2613          0   15
## 45 1.1809045 0.4454541       338        98        2084        233   15
## 46 1.1830065 1.0317693       345        62        1805        244    4
## 47 1.0375000 0.1899836       303        15        1579        166   15
##    nonverbalIQ verbalIQ
## 1           28       14
## 2           34       27
## 3           20       17
## 4           24       14
## 5           24       18
## 6           27       18
```

```
## 7              21        15
## 8              30        16
## 9              23        17
## 10             24        15
## 11             25        11
## 12             28        20
## 13             26        14
## 14             22         8
## 15             21        19
## 16             25        19
## 17             13        11
## 18             26        18
## 19             20        16
## 20             26        14
## 21             31        13
## 22             25        11
## 23             27        22
## 24             23        21
## 25             19        13
## 26             33        26
## 27             29        28
## 28             26        19
## 29             27        13
## 30             25        17
## 31             24        19
## 32             30        20
## 33             24        20
## 34             21         9
## 35             27        20
## 36             28        10
## 37             27        27
```

```
## 38           17        10
## 39           22        14
## 40           29        22
## 41           24        22
## 42           26        17
## 43           27        11
## 44           28        10
## 45           30        24
## 46           29        22
## 47           27        16
```

```
missed = finaldata_1 %>% group_by(ID) %>% tally()
filter(missed, n < 6)
```

```
## # A tibble: 13 × 2
##        ID     n
##     <dbl> <int>
## 1       2     5
## 2       7     5
## 3       8     5
## 4       9     5
## 5      17     5
## 6      26     5
## 7      38     5
## 8      40     5
## 9      44     5
## 10     45     5
## 11     48     5
## 12     55     5
## 13     56     4
```

USING ARRANGE

1. Sort kids to find the kid who produced the most words on the 6th visit
2. Sort kids to find the kid who produced the least amount of words on the 1st visit.

```
visit_6 = subset(finaldata_1, Visit == "6")
arrange(visit_6, tokens_CHI)
```

```
##    ID Visit        Ethnicity Diagnosis Gender   Age  MOT_MLU MOT_LUstd
## 1  21     6 African American       ASD      M 48.97 3.943636  2.819817
## 2  50     6            White       ASD      M 62.40 3.886842  2.583416
## 3  56     6            White       ASD      M    NA 3.474725  2.095238
## 4  17     6            White       ASD      M 54.43 3.650235  2.613115
## 5  47     6            White        TD      F 40.37 4.676101  2.442830
## 6  57     6            White       ASD      F 61.70 3.241422  2.096168
## 7  33     6 African American       ASD      F 46.17 3.965392  2.431706
## 8  32     6            White       ASD      M 54.63 4.003257  2.537020
## 9  39     6      White/Asian       ASD      M 53.63 4.100707  2.162669
## 10 46     6            White       ASD      M 57.43 3.460548  2.287079
## 11 24     6            White       ASD      M 62.40 3.860795  2.583920
## 12 10     6            White        TD      M 40.43 4.235585  2.511690
## 13 30     6            White       ASD      M 56.43 4.341549  2.704892
## 14  8     6            White        TD      M 40.17 4.366972  2.542581
## 15 29     6            White       ASD      M 55.17 4.349353  2.639446
## 16 61     6            White       ASD      M 62.33 3.514403  2.240388
## 17 23     6     White/Latino       ASD      M 47.50 3.534173  2.459876
## 18 19     6            White       ASD      M 56.73 3.156695  2.184106
## 19  6     6      Bangladeshi       ASD      F 46.53 2.483146  2.109968
## 20 48     6            White       ASD      M    NA 3.370093  2.089905
## 21 53     6            White        TD      M 44.43 3.706790  2.400582
```

```
## 22 51     6          Asian        TD       F 42.10 5.153639  2.756146
## 23 15     6          White        TD       M 40.30 4.347709  2.570906
## 24 31     6          White        TD       M 43.80 3.603473  2.723471
## 25 52     6          White        TD       M 43.03 5.587332  2.899559
## 26  3     6          White        TD       F 45.07 5.229885  3.014147
## 27 54     6          White        TD       M 39.30 4.186161  2.590683
## 28 28     6          White        TD       M 39.43 4.239437  2.588672
## 29 27     6          White        TD       M 41.17 4.254505  2.576855
## 30  5     6          White       ASD       M 54.13 4.587179  2.685925
## 31 36     6          White        TD       M 38.53 4.388235  2.571003
## 32 40     6       Lebanese       ASD       M 46.40 3.926928  2.532082
## 33 20     6          White       ASD       M 37.30 5.379798  2.712708
## 34 34     6          White       ASD       M 53.77 3.370937  2.314655
## 35 35     6          White        TD       M 39.07 4.239198  2.596866
## 36 58     6          White       ASD       M 46.07 3.320000  1.990377
## 37 12     6          White        TD       M 41.50 4.468493  2.574313
## 38 49     6          White        TD       M 43.40 5.247093  2.733299
## 39 59     6          White        TD       M 41.00 4.113158  2.372060
## 40 11     6          White        TD       M 40.27 4.287582  2.747960
## 41  1     6          White        TD       M 40.13 4.664013  2.765261
## 42  2     6          White       ASD       M 49.70 4.588477  2.783585
## 43 41     6          White        TD       M 39.93 4.061475  2.517218
## 44 16     6          White        TD       M 42.93 4.445872  2.582458
## 45 60     6          White       ASD       M 54.73 4.080446  2.591821
## 46 42     6          White        TD       M 40.13 3.391525  2.517921
## 47  4     6   White/Latino       ASD       M 51.37 3.532374  2.574647
## 48 14     6          White        TD       M 39.43 4.664286  2.732374
## 49 22     6          White        TD       F 39.23 4.211321  2.707896
## 50 13     6          White        TD       M 39.40 4.847418  2.694579
## 51 37     6          White       ASD       M 58.77 4.240798  2.555277
## 52 43     6          White       ASD       M 57.37 4.250853  2.596967
```

```
## 53 25      6            White        TD      M 42.47 4.472993  2.718080
## 54 18      6            White        TD      M 44.07 4.013353  2.749934
## 55 26      6            White       ASD      M 51.00 4.111413  2.889464
## 56 55      6            White        TD      M 41.93 3.957230  2.537740
##       CHI_MLU CHI_LUstd types_MOT types_CHI tokens_MOT tokens_CHI ADOS
## 1  0.5000000 0.5000000       388         2       2077          2   21
## 2  1.3333333 0.7453560       370         4       1396          8   19
## 3  1.0588235 0.2352941       284         6       1390         36   14
## 4  1.0000000 0.0000000       281         3       1454         37   14
## 5  2.7600000 2.0254382       322        34       1371         61    0
## 6  0.0156250 0.1240196       444         2       2450         64   15
## 7  0.7536232 0.9387904       326        14       1852         79   14
## 8  2.6315789 1.4585691       585        12       2202        100   21
## 9  1.6447368 1.1888069       330        55       1987        110   11
## 10 1.1610169 0.4312052       351        10       1918        137   20
## 11 1.1680000 0.4855677       309        62       1349        143   20
## 12 2.7051282 1.8681471       400        73       2271        189    3
## 13 1.0843373 0.8877319       425       101       2219        195   12
## 14 2.2258065 2.0060609       322        64       1352        197    5
## 15 1.2883436 0.7889103       454        52       2391        204   17
## 16 1.4012346 0.9125996       311        58       1481        210   15
## 17 1.3052632 0.7549651       372        47       1748        236   14
## 18 2.1450382 1.4148727       256        55        995        274   11
## 19 1.4967320 1.7863834       158        66        536        300   17
## 20 1.4734513 0.8681705       319        98       1565        306   17
## 21 3.2439024 2.2245543       249       102       1024        358    0
## 22 2.7612903 2.1465167       383       140       1866        395    1
## 23 2.8690476 2.0861206       327       156       1395        410    0
## 24 2.0717489 1.7854067       475       185       2363        418    0
## 25 2.4292929 1.9981179       548       163       2887        436    0
## 26 3.7103448 2.1621102       486       173       2564        460    1
```

```
## 27 2.8921569 2.2158381       437        183        2347        517    1
## 28 2.9607843 2.2444772       391        164        1889        521    0
## 29 2.4800000 1.8810872       385        154        1788        530    0
## 30 2.7665198 2.3584374       462        179        3182        538    9
## 31 2.5614754 1.9013346       324        165        1978        568    3
## 32 2.1586207 1.8389497       417        170        2508        571   13
## 33 2.9027778 2.4990353       433        155        2389        590    9
## 34 2.2745098 1.5852065       342        157        1540        605   10
## 35 2.3815789 2.1396452       411        156        2438        611    1
## 36 2.1553398 1.5439030       335        197        2221        618   15
## 37 3.5049505 2.5039524       339        201        1498        640    0
## 38 3.5950000 2.4064445       383        217        1701        646    0
## 39 2.8480000 2.1690772       303        158        1460        659    4
## 40 2.7571429 2.1953569       260        168        1138        666    0
## 41 2.8651685 2.2478838       595        210        2586        686    0
## 42 3.4135021 2.3234966       304        245         999        698   13
## 43 2.8526316 1.9572721       397        213        1741        702    0
## 44 2.9482072 2.3245843       491        250        2264        710    0
## 45 3.4415584 2.4343866       505        226        3072        713   13
## 46 3.0727969 2.5186027       357        219        1764        719    0
## 47 3.2781955 2.5379922       410        166        2171        738    8
## 48 3.8114035 2.7948028       359        178        1802        793    0
## 49 3.0911950 2.2808716       478        221        2044        847    0
## 50 3.7011952 2.7079561       388        210        1959        864    0
## 51 3.0774194 2.3528582       429        217        2510        897    7
## 52 2.6794872 1.8979989       374        219        2227        921   14
## 53 3.0614525 2.3687019       555        237        2895       1010    0
## 54 2.9095355 2.2440646       516        235        2576       1079    0
## 55 3.3643411 2.5967720       452        273        3076       1249   11
## 56 2.9092742 2.1789486       367        260        1731       1294    0
##    nonverbalIQ verbalIQ
```

```
## 1              22        8
## 2              17       10
## 3              27       11
## 4              25       11
## 5              27       20
## 6              28       10
## 7              25       11
## 8              21        9
## 9              26       19
## 10             21        9
## 11             13       11
## 12             27       18
## 13             31       13
## 14             32       31
## 15             26       14
## 16             27       16
## 17             25       19
## 18             28       20
## 19             20       17
## 20             28       10
## 21             30       30
## 22             22       14
## 23             24       15
## 24             27       22
## 25             29       22
## 26             29       18
## 27             24       22
## 28             20       16
## 29             26       18
## 30             34       27
## 31             19       13
```

```
## 32          27       13
## 33          26       14
## 34          27       22
## 35          23       21
## 36          30       24
## 37          30       16
## 38          27       27
## 39          29       22
## 40          21       15
## 41          28       14
## 42          34       27
## 43          25       17
## 44          29       26
## 45          30       30
## 46          24       19
## 47          31       27
## 48          23       17
## 49          21       19
## 50          25       17
## 51          33       26
## 52          42       27
## 53          29       22
## 54          29       33
## 55          32       33
## 56          26       17
```

```
#Kid 55 produced the most words on the 6th visit

visit_1 = subset(finaldata_1, Visit == "1")
arrange(visit_1, tokens_CHI)
```

```
##     ID Visit         Ethnicity Diagnosis Gender   Age  MOT_MLU MOT_LUstd
## 1  57     1             White       ASD      F 41.07 3.833770  2.417727
## 2  32     1             White       ASD      M 34.27 3.686347  2.650865
## 3  36     1             White        TD      M 19.20 3.921109  2.376179
## 4  30     1             White       ASD      M 36.53 3.607088  2.340376
## 5  24     1             White       ASD      M 37.47 2.917355  2.073392
## 6   6     1       Bangladeshi       ASD      F 26.17 2.618729  1.935874
## 7  33     1  African American       ASD      F 25.33 2.287293  1.928359
## 8  17     1             White       ASD      M 34.80 3.182390  2.269630
## 9  51     1             Asian        TD      F 20.87 3.435743  2.257715
## 10 48     1             White       ASD      M 31.63 3.765528  2.329794
## 11 50     1             White       ASD      M 37.47 3.704110  2.276770
## 12 56     1             White       ASD      M 35.50 2.548969  1.706620
## 13 40     1          Lebanese       ASD      M 24.90 2.997050  2.202839
## 14 12     1             White        TD      M 20.07 4.195335  2.280551
## 15 46     1             White       ASD      M 36.73 4.883966  2.773678
## 16 44     1             White        TD      M 19.77 3.088757  2.098655
## 17 10     1             White        TD      M 19.27 4.204846  2.384767
## 18 45     1             White        TD      M 20.03 2.776181  1.904606
## 19 27     1             White        TD      M 21.03 3.616867  2.261294
## 20 15     1             White        TD      M 19.27 3.967078  2.302921
## 21 11     1             White        TD      M 19.23 3.380463  2.214518
## 22 23     1      White/Latino       ASD      M 27.37 3.024548  2.278008
## 23 14     1             White        TD      M 18.97 3.420315  2.273399
## 24 39     1       White/Asian       ASD      M 33.20 3.298748  1.995123
## 25 54     1             White        TD      M 19.10 3.487871  2.432205
## 26 28     1             White        TD      M 19.93 2.788793  2.055953
## 27 29     1             White       ASD      M 34.87 3.304189  2.372505
## 28 22     1             White        TD      F 18.93 3.630476  2.452720
## 29 21     1  African American       ASD      M 27.53 4.390879  3.068753
## 30  7     1             White       ASD      F 41.00 2.244755  1.510878
```

```
## 31  9      1            White        TD        F 18.30 3.544419  2.272387
## 32  1      1            White        TD        M 19.80 3.621993  2.164553
## 33  8      1            White        TD        M 20.10 3.265082  2.151724
## 34 55      1            White        TD        M 19.97 3.509138  2.189292
## 35 61      1            White       ASD        M 42.00 3.030405  2.107521
## 36 42      1            White        TD        M 19.37 4.033333  2.460373
## 37 34      1            White       ASD        M 33.77 2.743455  1.892987
## 38  5      1            White       ASD        M 34.03 3.986357  2.500713
## 39 47      1            White        TD        F 20.03 3.943005  2.384506
## 40 20      1            White       ASD        M 18.77 2.524740  2.141337
## 41 19      1            White       ASD        M 35.80 2.539823  1.994618
## 42 58      1            White       ASD        M 26.00 2.747100  1.853343
## 43 41      1            White        TD        M 19.23 3.093146  2.362590
## 44 59      1            White        TD        M 20.80 3.432387  2.124274
## 45 13      1            White        TD        M 19.00 3.960254  2.133094
## 46 16      1            White        TD        M 21.67 4.910190  2.496744
## 47  3      1            White        TD        F 23.50 3.757269  2.350833
## 48 35      1            White        TD        M 19.30 3.561364  2.229117
## 49  4      1     White/Latino       ASD        M 31.03 3.459370  2.216235
## 50 49      1            White        TD        M 23.07 4.030075  2.695787
## 51 31      1            White        TD        M 23.13 3.494327  2.302219
## 52 52      1            White        TD        M 22.57 5.344227  2.917012
## 53 37      1            White       ASD        M 39.50 4.135036  2.563076
## 54 38      1            White        TD        F 19.87 3.420975  2.244646
## 55 18      1            White        TD        M 20.77 3.587855  2.282255
## 56 25      1            White        TD        M 21.77 4.643200  2.619140
## 57 43      1            White       ASD        M 37.03 3.341418  2.073518
## 58  2      1            White       ASD        M 28.80 4.098446  2.612375
## 59 60      1            White       ASD        M 34.00 3.604140  2.550110
## 60 53      1            White        TD        M 23.90 4.159159  2.509181
## 61 26      1            White       ASD        M 30.40 4.690751  2.545488
```

```
##       CHI_MLU CHI_LUstd types_MOT types_CHI tokens_MOT tokens_CHI ADOS
## 1  0.0000000 0.0000000       386         0       2613          0   15
## 2  1.5000000 0.5000000       228         3        927          3   21
## 3  1.2307692 0.5756396       281         8       1631         16    3
## 4  0.9000000 0.4358899       366        11       2054         21   12
## 5  1.0833333 0.3996526       193         6        654         26   20
## 6  1.0000000 0.0000000       212         4        761         29   17
## 7  1.2500000 0.5747670       206        13        788         35   14
## 8  1.0277778 0.1643355       281         9       1418         37   14
## 9  1.1818182 0.4575657       331         7       1503         38    1
## 10 1.2500000 0.6123724       303        17       2147         40   17
## 11 1.1000000 0.4898979       265         8       1215         43   19
## 12 1.0444444 0.2060804       195         9        955         47   14
## 13 1.0175439 0.1312862       252         9       1827         58   13
## 14 1.0877193 0.2828862       235        17       1262         62    0
## 15 1.1666667 0.5000000       387        10       2144         63   20
## 16 1.2592593 0.4382281       275         9       1417         68    3
## 17 1.0375000 0.1899836       289        15       1808         83    3
## 18 0.5584416 0.7811664       258        20       1188         91    5
## 19 1.3661972 0.6763497       317        37       1361         95    0
## 20 1.1647059 0.4813476       277        27       1643         99    0
## 21 1.2168675 0.4121116       215        24       1136        101    0
## 22 1.4324324 1.1159904       328        41       2138        103   14
## 23 1.0396040 0.1950269       287         7       1625        105    0
## 24 1.2043011 0.5784507       274        36       1537        109   11
## 25 1.3139535 0.7664723       214        32       1118        109    1
## 26 1.2758621 0.4719647       178        34        584        111    0
## 27 1.0086207 0.0924466       295         6       1643        117   17
## 28 1.2371134 0.4926929       343        36       1698        118    0
## 29 1.0000000 0.0000000       485         8       2826        122   21
## 30 1.2641509 0.7177255       152        29        578        130   18
```

```
## 31 1.0378788 0.1909031       363        36        1408         137    0
## 32 1.2522523 0.4739801       378        14        1835         139    0
## 33 1.5600000 1.2273549       288        59        1564         143    5
## 34 1.1846154 0.3879852       178        11        1144         154    0
## 35 1.0375000 0.1899836       303        15        1579         166   15
## 36 1.3034483 0.7818999       373        47        2334         176    0
## 37 1.3766234 0.6353026       214        67         893         180   10
## 38 1.3947368 0.6897549       324        57        2859         197    9
## 39 1.0761421 0.3010798       260        13        1347         212    0
## 40 0.1857143 0.4564727       321        16        1787         214    9
## 41 1.3595506 0.7606501       283        89        1019         227   11
## 42 1.1809045 0.4454541       338        98        2084         233   15
## 43 1.0262009 0.2791346       333        32        1547         235    0
## 44 1.1830065 1.0317693       345        62        1805         244    4
## 45 1.5740741 0.6460283       329        69        2139         249    0
## 46 1.5988372 0.7206254       375        90        2585         254    0
## 47 1.8776978 0.9630619       334        51        2674         260    1
## 48 1.2641509 0.5546015       291        24        1344         260    1
## 49 2.0972222 1.2819226       379       102        2009         269    8
## 50 1.4258373 1.0238787       255        73        1452         286    0
## 51 1.5333333 1.0666667       322        91        1870         319    0
## 52 1.4086957 0.8433438       441        92        2267         319    0
## 53 0.4805825 0.7221272       381        39        1988         337    7
## 54 1.3322785 0.7633014       342        96        2035         398    1
## 55 1.7948718 1.0664941       467       108        2555         406    0
## 56 1.5827338 0.8388008       373        71        2740         433    0
## 57 1.9144981 1.3034604       271       101        1591         450   14
## 58 2.2768595 1.4437818       317       146        1428         461   13
## 59 2.8763441 1.9178784       400       149        2587         469   13
## 60 1.9397163 1.2603827       236       120        1170         473    0
## 61 3.4000000 1.7225773       278       119        1450         483   11
```

```
##    nonverbalIQ verbalIQ
## 1           28       10
## 2           21        9
## 3           19       13
## 4           31       13
## 5           13       11
## 6           20       17
## 7           25       11
## 8           25       11
## 9           22       14
## 10          28       10
## 11          17       10
## 12          27       11
## 13          27       13
## 14          30       16
## 15          21        9
## 16          30       20
## 17          27       18
## 18          24       20
## 19          26       18
## 20          24       15
## 21          21       15
## 22          25       19
## 23          23       17
## 24          26       19
## 25          24       22
## 26          20       16
## 27          26       14
## 28          21       19
## 29          22        8
## 30          24       14
```

```
## 31            24        18
## 32            28        14
## 33            32        31
## 34            26        17
## 35            27        16
## 36            24        19
## 37            27        22
## 38            34        27
## 39            27        20
## 40            26        14
## 41            28        20
## 42            30        24
## 43            25        17
## 44            29        22
## 45            25        17
## 46            29        26
## 47            29        18
## 48            23        21
## 49            31        27
## 50            27        27
## 51            27        22
## 52            29        22
## 53            33        26
## 54            29        28
## 55            29        33
## 56            29        22
## 57            42        27
## 58            34        27
## 59            30        30
## 60            30        30
## 61            32        33
```

```
#Kid 57 produced the most words on the 6th visit
```

USING SELECT

1. Make a subset of the data including only kids with ASD, mlu and word tokens
2. What happens if you include the name of a variable multiple times in a select() call?

```
ASD_only = subset(finaldata_1, Diagnosis == "ASD")
ASD_only = select(ASD_only, MOT_MLU, CHI_MLU, tokens_MOT, tokens_CHI)

ASD_only2 = select(ASD_only, MOT_MLU, CHI_MLU, tokens_MOT, tokens_CHI, tokens_CHI)
#The varible will only appear one time even though you have written it two times.
```

USING MUTATE, SUMMARISE and PIPES 1. Add a column to the data set that represents the mean number of words spoken during all visits. 2. Use the summarise function and pipes to add an column in the data set containing the mean amount of words produced by each trial across all visits. HINT: group by Child.ID 3. The solution to task above enables us to assess the average amount of words produced by each child. Why don’t we just use these average values to describe the language production of the children? What is the advantage of keeping all the data?

```
finaldata_2 = finaldata_1 %>% group_by(Visit) %>% mutate(Mean_words_visit = sum(tokens_CHI)/len
gth(tokens_CHI))

finaldata_3 = finaldata_2 %>% group_by(ID) %>% mutate(Mean_words_id = sum(tokens_CHI)/length(to
kens_CHI))
finaldata_3
```

```
## Source: local data frame [352 x 19]
## Groups: ID [61]
##
```

```
##        ID Visit Ethnicity Diagnosis Gender   Age  MOT_MLU MOT_LUstd
##     <dbl> <dbl>    <fctr>    <fctr> <fctr> <dbl>    <dbl>     <dbl>
## 1       1     1     White        TD      M 19.80 3.621993  2.164553
## 2       1     2     White        TD      M 23.93 3.857367  2.417939
## 3       1     3     White        TD      M 27.70 4.321881  2.517464
## 4       1     4     White        TD      M 32.90 4.415330  2.449573
## 5       1     5     White        TD      M 35.90 5.209615  2.814165
## 6       1     6     White        TD      M 40.13 4.664013  2.765261
## 7       2     1     White       ASD      M 28.80 4.098446  2.612375
## 8       2     2     White       ASD      M 33.17 4.964664  2.499220
## 9       2     3     White       ASD      M 37.07 4.147059  2.803222
## 10      2     4     White       ASD      M 41.07 5.309804  2.842621
## # ... with 342 more rows, and 11 more variables: CHI_MLU <dbl>,
## #   CHI_LUstd <dbl>, types_MOT <int>, types_CHI <int>, tokens_MOT <int>,
## #   tokens_CHI <int>, ADOS <int>, nonverbalIQ <int>, verbalIQ <int>,
## #   Mean_words_visit <dbl>, Mean_words_id <dbl>
```

```
#We don't just use the average values because we end up not taking the previous visit into acco
unt. We would violate the assumption of independence. We cannot regard multiple recordings from
the same subject as independent from each other.
```