Analyses for JSLHR version

2020-10-17

Contents

History:

- 2020-08-05 final first version
- 2020-10-10 (rehaul),
- latest minor edits 2020-10-17

TODO:

- pipeline is still not transparent
- there are duplicate files across raw and derived data
- there are a bunch of files with similar names
- README is old
- a note said read demo data created by AC from info in paper - should be replaced with real demo data
- add code to print out the results paragraphs
- corage: - adjust margins and remove title repetition?
- chunk-based analyses for age

Read data in

```r
# read datasets

demo_data=read.csv("../Derived_Data/demo-data.tsv",sep="\t")
data_ang <- read.csv("../Derived_Data/classifications_PU_zoon_final17.csv",header=T,sep=",")
data_td <- read.csv("../Derived_Data/classifications_PU_zoon_final.csv")
data_all<-rbind(data_ang, data_td)

#add filenames to demo data
demo_data_fn <- demo_data %>%
    left_join(select(data_all, filename, ChildID), by = c("ChildID"))
```

```
## Warning: Column `ChildID` joining factors with different levels, coercing to
## character vector
```

```r
demo_data_fn<-unique(demo_data_fn)

#remove the word mixed that takes up space and is unnecessary
data_all$Zoon_classif=factor(gsub("Mixed_","",as.character(data_all$Zoon_classif),fixed=T))
#relevel the factor so that it's easier to read
data_all$Zoon_classif=factor(data_all$Zoon_classif, levels=c("Canonical","Non-Canonical",
# create lab column with easier to read correspondance
data_all$lab<-as.character(data_all$Major_Choice)
data_all$lab[data_all$lab=="Non-canonical syllables"]<-"Non-Canonical"
data_all$lab[data_all$lab=="Canonical syllables"]<-"Canonical"
data_all$lab[data_all$lab %in% c("Don't mark","None")]<-"Junk"
data_all$lab=factor(data_all$lab,levels=levels(data_all$Zoon_classif))
#apply same factor levels as zooniverse so that we can do symmetrical confusion matrices
```

Data post-processing

We collected a total of 169,628 judgments provided for 33,880 500-ms chunks, corresponding to 11,984 LENA segments. Nearly a fifth of chunks did not have at least 3 labels in agreement out of the 5 Zooniverse labels (N = 6,585, 19% of all chunks). Of the chunks without a majority agreement, 4341 (66%) contained one or two Junk judgements (out of 5), 6523 (99,9%) had at least two matching judgements (the threshold used for lab-annotated segments), and only 61 (0,01%) had 5 different judgements. Future work may explore different ways of setting the minimal requirement for convergence, but for further analyses here, we focused on the 81% of chunks that did have at least 3 labels in agreement; this represented 136,703 labels for 27,295 chunks, corresponding to 11,593 LENA segments. As the segments average 1.12 seconds in length, this means about 3.8 hours of audio data were annotated by 8 different annotators (3 in the laboratory, 5 on Zooniverse).

```r
# remove non-majority labels
#TODO

# we map the mixed
data_all$Zoon_classif[data_all$Zoon_classif=="Laughing_Canonical"]<-"Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Laughing_Non-Canonical"]<-"Non-Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Laughing_Non-Canonical_Crying"]<-"Non-Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Laughing_Crying"]<-"Crying"
data_all$Zoon_classif[data_all$Zoon_classif=="Non-Canonical_Crying"]<-"Non-Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Non-Canonical_Laughing_Crying"]<-"Non-Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Crying_Canonical"]<-"Canonical"
# +
data_all$Zoon_classif[data_all$Zoon_classif=="Canonical_Crying"]<-"Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Canonical_Laughing"]<-"Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Laughing_Canonical_Crying"]<-"Non-Canonical"

data_all$Zoon_classif[data_all$Zoon_classif=="Crying_Laughing"]<-"Crying"
data_all$Zoon_classif[data_all$Zoon_classif=="Crying_Canonical_Laughing"]<-"Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Crying_Laughing_Non-Canonical"]<-"Non-Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Crying_Non-Canonical"]<-"Non-Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Crying_Non-Canonical_Laughing"]<-"Non-Canonical"
data_all$Zoon_classif[data_all$Zoon_classif=="Non-Canonical_Laughing"]<-"Non-Canonical"

#and reset the factors for cleanliness
data_all$Zoon_classif=factor(data_all$Zoon_classif)
```

```r
data_all$lab=factor(data_all$lab)
sample_data<-cbind(data_all$lab,data_all$Zoon_classif)

table_zoon=round(table(data_all$Zoon)/sum(table(data_all$Zoon))*100)
table_lab=round(table(data_all$lab)/sum(table(data_all$lab))*100)

#get the ns by child, then calculate the linguistic ratio & canonical ratio, separately for zooniverse
ztab=table(data_all$filename,data_all$Zoon_classif)
z_lp=rowSums(ztab[,c("Canonical","Non-Canonical")])/rowSums(ztab[,-which(colnames(ztab) %in% c("Junk"))]
z_cp=ztab[,c("Canonical")]/rowSums(ztab[,c("Canonical","Non-Canonical")])
ltab=table(data_all$filename,data_all$lab)
l_lp=rowSums(ltab[,c("Canonical","Non-Canonical")])/rowSums(ltab[,-which(colnames(ztab) %in% c("Junk"))]
l_cp=ltab[,c("Canonical")]/rowSums(ltab[,c("Canonical","Non-Canonical")])
#put all the proportions together
if(sum(rownames(ztab)==rownames(ltab))==dim(ztab)[1]) proportions=cbind(rownames(ztab),z_lp,z_cp,l_lp,l
colnames(proportions)[1]<-"filename"

#ages=aggregate(data_all$Age,by=list(data_all$ChildID),mean) #this is a weird way of adding ages, since
#improvement: now we merge with a demo data tab, but note this is merged with child id, so the problem

# Created demo_data with filenames. Use filenames instead of childIDs to merge proportions and demo dat

merge(proportions,demo_data_fn,by="filename")->proportions
colnames(proportions)[dim(proportions)[2]]<-"Age"

#cbinding results in text, so we numerize the proportions
for(thisvar in c("z_lp","z_cp","l_lp","l_cp")) proportions[,thisvar]=as.numeric(as.character(proportions
summary(proportions)
```

```
##                      filename       z_lp             z_cp
##   20180206_110905_009463: 1   Min.   :0.5927   Min.   :0.02882
##   20180419_111712_022875: 1   1st Qu.:0.8058   1st Qu.:0.05891
##   20180530_180405_024879: 1   Median :0.8785   Median :0.14065
##   20180530_181101_022875: 1   Mean   :0.8511   Mean   :0.19128
##   20180808_111325_024882: 1   3rd Qu.:0.9313   3rd Qu.:0.28454
##   20180906_133011_022875: 1   Max.   :0.9740   Max.   :0.51084
##   (Other)               :14
##       l_lp             l_cp           ChildID                   Diagnosis
##   Min.   :0.3880   Min.   :0.01488   Length:20          AngelmanSyndrome:10
##   1st Qu.:0.9130   1st Qu.:0.06609   Class :character   Low-RiskControl :10
##   Median :0.9578   Median :0.11115   Mode  :character
##   Mean   :0.9131   Mean   :0.19129
##   3rd Qu.:0.9787   3rd Qu.:0.26529
##   Max.   :0.9968   Max.   :0.60000
##
##   Sex          Age
##   F: 7   Min.   : 4.67
##   M:13   1st Qu.:12.13
##          Median :17.50
##          Mean   :25.04
##          3rd Qu.:43.46
##          Max.   :53.26
##
```

Results

Descriptive analyses

In this section, we provide descriptive analyses of our dataset. According to lab annotators, 15% of segments were canonical, 56% non-canonical, 2% laughing, and 5% crying, with the remaining 22% being categorized as "Don't code". Zooniverse data revealed a similar distribution: 15% canonical, 60% non-canonical, 4% laughing, 9% crying, 13% junk. Next, we inspected the relationship between age and child-level derived metrics, of which we had two: i) Linguistic proportion = ("Canonical"+"Non-Canonical")/"All vocalizations" (i.e., we remove junk), and ii) Canonical proportion = "Canonical"/("Canonical"+"Non-Canonical") (i.e. we remove junk + non-linguistic vocalizations). See Figure XX for results.

Figure XX. Correlations between child-level descriptors and age as a function of metric (linguistic ratio in the top row, canonical ratio in the bottom row), annotation method, and child group (Red as TD, and black as AS)

```r
prettynames=c("Linguistic Ratio (Zooniverse)","Canonical Ratio (Zooniverse)",
              "Linguistic Ratio (Lab)","Canonical Ratio (Lab)" )
names(prettynames)<-c("z_lp","z_cp","l_lp","l_cp")
mycols=c("black","red")
names(mycols)<-c("Low-RiskControl","AngelmanSyndrome")
mypch=c(4,20)
names(mypch)<-c("Low-RiskControl","AngelmanSyndrome")

lp_range=range(proportions[,c("z_lp","l_lp")])
cp_range=range(proportions[,c("z_cp","l_cp")])

#jpeg("../Results/corage.jpg",width=20,height=20,units="cm",res=300)
layout(matrix(c(1:4), 2, 2, byrow = F))
par(mar = c(4, 4, 1, 1))

#create matrix to hold r results
myr=matrix(NA,nrow=4,ncol=2)
rownames(myr)<-c("z_lp","z_cp","l_lp","l_cp")
colnames(myr)<-c("AS","LR")
mylb=myhb=myn=myp=myr #this is an ugly way of creating 5 matrices

for(thisvar in c("z_lp","z_cp","l_lp","l_cp")) {

  if(thisvar %in% c("z_lp","l_lp")) myrange=lp_range else myrange=cp_range

  plot(proportions[,thisvar]~proportions$Age, pch=mypch[proportions$Diagnosis],xlab="Age (months)",ylab=
       col=mycols[proportions$Diagnosis],
         ylim=myrange)
  abline(lm(proportions[,thisvar]~proportions$Age,subset=c(proportions$Diagnosis=="AngelmanSyndrome")),

  #get cor properties
  thiscor=cor.test(proportions[proportions$Diagnosis=="AngelmanSyndrome",thisvar],proportions$Age[propo
    myr[thisvar,"AS"]=round(thiscor$estimate,3)
    mylb[thisvar,"AS"]=round(thiscor$conf.int[1],3)
    myhb[thisvar,"AS"]=round(thiscor$conf.int[2],3)
    myn[thisvar,"AS"]=round(thiscor$parameter,3)
    myp[thisvar,"AS"]=round(thiscor$p.value,3)

    #add r to figure
```

```
    text(mean(proportions$Age[proportions$Diagnosis=="AngelmanSyndrome"]),mean(proportions[proportions$Dia

    #repeat for LR
    abline(lm(proportions[,thisvar]~proportions$Age,subset=c(proportions$Diagnosis!="AngelmanSyndrome")

      #get cor properties
  thiscor=cor.test(proportions[proportions$Diagnosis!="AngelmanSyndrome",thisvar],proportions$Age[propor
    myr[thisvar,"LR"]=round(thiscor$estimate,3)
    mylb[thisvar,"LR"]=round(thiscor$conf.int[1],3)
    myhb[thisvar,"LR"]=round(thiscor$conf.int[2],3)
    myn[thisvar,"LR"]=round(thiscor$parameter,3)
    myp[thisvar,"LR"]=round(thiscor$p.value,3)

    #add r to figure

    text(mean(proportions$Age[proportions$Diagnosis!="AngelmanSyndrome"]),mean(proportions[proportions$M
}
```
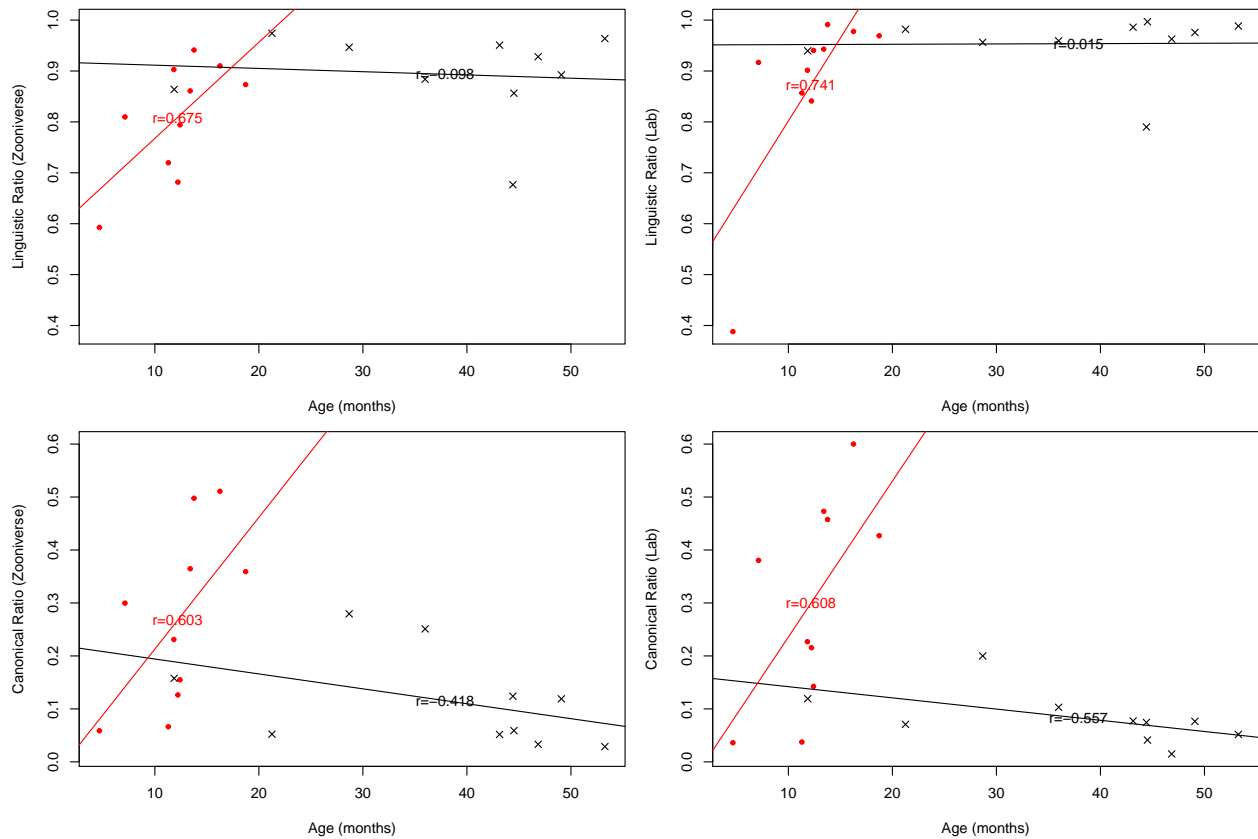


```
#dev.off()
```

Descriptive analyses on the laboratory annotations showed that correlations between the Linguistic proportion and age differed across the groups. There was a near-zero relationship among the older children diagnosed with Angelman Syndrome r(8) = 0.015, CI [-0.621,0.638], p=0.968]; and a significant association among younger low-risk control children r(8) = 0.741, CI [0.209,0.935], p=0.014]. The Canonical proportion exhibited non-significant developmental decreases among older children diagnosed with Angelman Syndrome r(8) = -0.557, CI [-0.879,0.112], p=0.094]; and marginal developmental increases among low-risk control r(8) = 0.608, CI [-0.035,0.895], p=0.062].

5

Using the Zooniverse annotations, we found that the association with age was very weak for children diagnosed with Angelman Syndrome r(8) = -0.098, CI [-0.685,0.567], p=0.788]; whereas younger low-risk control children showed a significant increase with age r(8) = 0.675, CI [0.079,0.916], p=0.032]. Similarly, there were non-significant developmental decreases in the Canonical among children with Angelman Syndrome r(8) = -0.418, CI [-0.829,0.288], p=0.23]; and marginal developmental increases among low-risk control children, r(8) = 0.603, CI [-0.043,0.893], p=0.065].

Given that these derived metrics are defined at the child level, there was one stage in our pre-processing that may not have been necessary, whereby we collapsed judgments across chunks associated to the same segment. We therefore repeated our analyses but deriving our proportions for the Zooniverse data not from the segment-level composite, but rather the individual chunk-level annotations. We found that the Linguistic proportion was not strongly associated with age in the Angelman Syndrome group [r(8) = -.04, CI -.6 .6, p=0.91], but increased with age in the low-risk control group [r(8) = .71, CI .15 .92, p=.02]. Similarly, the Canonical proportion did not exhibit the same pattern across the groups, with developmental decreases found among children with Angelman Syndrome r(8) = -.35, CI -.80 .35, p=.32; and developmental increases among low-risk control r(8) = .57, CI -.08 0.88, p=.83.

Main analyses

Next, we discuss the correspondence between citizen science classifications and the laboratory gold standard, at the level of individual clips. Results were visualized and assessed with a confusion matrix. We report a Precision plot (Fig. 1) and a Recall plot (Fig. 2): the diagonal elements show the number of correct segment-level classifications for each class while the off-diagonal elements show non-matching classifications.

```
mygac=gac(data = sample_data, kat = 5, weight = c("unweighted"),
    conf.level = 0.95)


mycf=confusionMatrix(data_all$lab, data_all$Zoon_classif, dnn = c("Lab","Zooniverse"))
conf_tab=mycf$table
# this package uses sensitivity & specificity
#Sensitivity=recall
#Specificity=precision
#mycf
```

```
colsums=colSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:dim(conf_tab)[2]) my_conf_tab[,i]=my_conf_tab[,i]/colsums[i]
colSums(my_conf_tab)
```

```
##      Canonical Non-Canonical        Crying      Laughing          Junk
##              1             1             1             1             1
```

```
prop_cat=data.frame(my_conf_tab*100) #generates precision because columns
prop_cat$id=paste(prop_cat$Lab,prop_cat$Zooniverse)
colnames(prop_cat)[3]<-"pr"
data.frame(conf_tab)->stall
stall$id=paste(stall$Lab,stall$Zooniverse)
stall=merge(stall,prop_cat[c("id","pr")])
prec_plot = ggplot(data = stall, mapping = aes(y = Lab, x=Zooniverse)) +
 geom_tile(aes(fill= rescale(pr)), colour = "white") +
  geom_text(aes(label = paste(round(pr),"%")), vjust = -1,size=2) +
  geom_text(aes(label = Freq), vjust = 1,size=1) +
  scale_fill_gradient(low = "white", high = "red", name = "Percentage") +
    theme(legend.position = "none") +
  xlab("Zooniverse") + ylab("Lab") +
```

```r
  ggtitle("Precision")+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

# repeat for recall
rowsums=rowSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:dim(conf_tab)[2]) my_conf_tab[,i]=my_conf_tab[,i]/rowsums[i]
rowSums(my_conf_tab)
```
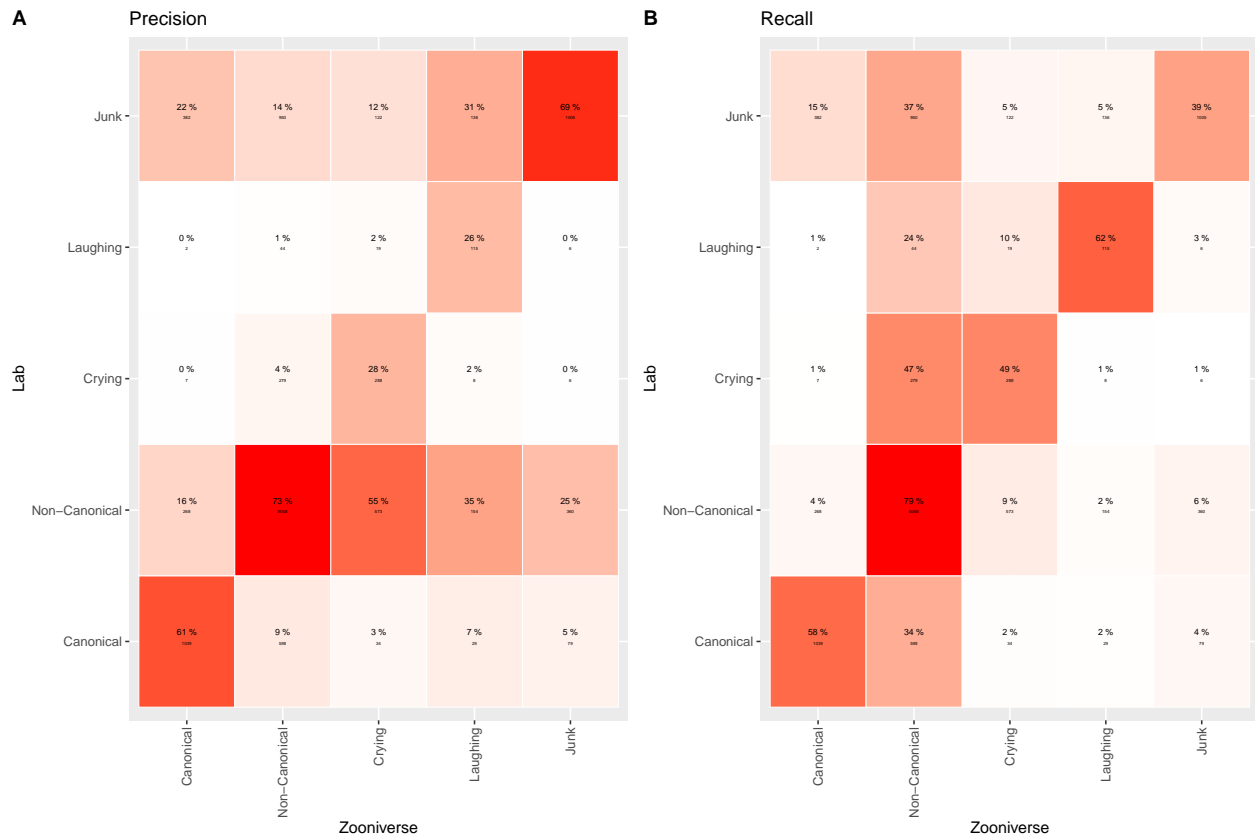
```
##     Canonical Non-Canonical        Crying      Laughing          Junk
##     0.9213192     2.8808609     0.5824913     0.6608792     1.6885824
```

```r
prop_cat=data.frame(conf_tab/rowSums(conf_tab)*100)   #generates recall because rows
prop_cat$id=paste(prop_cat$Lab,prop_cat$Zooniverse)
colnames(prop_cat)[3]<-"rec"
data.frame(conf_tab)->stall
stall$id=paste(stall$Lab,stall$Zooniverse)
stall=merge(stall,prop_cat[c("id","rec")])
rec_plot = ggplot(data = stall, mapping = aes(y = Lab, x=Zooniverse)) +
 geom_tile(aes(fill= rescale(rec)), colour = "white") +
  geom_text(aes(label = paste(round(rec),"%")), vjust = -1,size=2) +
  geom_text(aes(label = Freq), vjust = 1,size=1) +
  scale_fill_gradient(low = "white", high = "red", name = "Percentage") +
    theme(legend.position = "none") +
  xlab("Zooniverse") + ylab("Lab") +
  ggtitle("Recall")+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

ggarrange(prec_plot, rec_plot,
          labels = c("A", "B"),
          ncol = 2, nrow = 1)
```

A Precision

Lab (rows) × Zooniverse (columns)

Lab \ Zooniverse	Canonical	Non-Canonical	Crying	Laughing	Junk
Junk	22 %	14 %	12 %	31 %	69 %
Laughing	0 %	1 %	2 %	26 %	0 %
Crying	0 %	4 %	28 %	2 %	0 %
Non-Canonical	16 %	73 %	55 %	35 %	25 %
Canonical	61 %	9 %	3 %	7 %	5 %

B Recall

Lab \ Zooniverse	Canonical	Non-Canonical	Crying	Laughing	Junk
Junk	15 %	37 %	5 %	5 %	39 %
Laughing	1 %	24 %	10 %	62 %	3 %
Crying	1 %	47 %	49 %	1 %	1 %
Non-Canonical	4 %	79 %	9 %	2 %	6 %
Canonical	58 %	34 %	2 %	2 %	4 %

From both visualizations, it appears that performance is moderate to good, with an overall accuracy of 65%, CI = [64,66], a kappa of 0.426, and a Gwet's AC1 coefficient of 0.587, CI = [0.576,0.597].

Child level descriptors

TODO

- fix size
- point should be dependent on child group

```r
layout(matrix(c(1:2), 1, 2, byrow = F))
par(mar = c(4, 4, 1, 1))

#create matrix to hold r results
myr=matrix(NA,nrow=2,ncol=3)
rownames(myr)<-c("lp","cp")
colnames(myr)<-c("AS","LR","all")
mylb=myhb=myn=myp=myr #this is an ugly way of creating 5 matrices

lims=range(c(proportions[,"z_lp"],proportions[,"l_lp"]))

  plot(proportions[,"z_lp"]~proportions[,"l_lp"], pch=20,xlab=prettynames["l_lp"],ylab=prettynames["z_l
        xlim=lims,ylim=lims,
        col=mycols[proportions$Diagnosis])
  lines(c(0,1),c(0,1),lty=2,col="darkgray")

  abline(lm(proportions[,"z_lp"]~proportions[,"l_lp"],subset=c(proportions$Diagnosis=="AngelmanSyndrome
```

8

```r
#get cor properties
thiscor=cor.test(proportions[proportions$Diagnosis=="AngelmanSyndrome","z_lp"],proportions[proportions
  myr[1,"AS"]=round(thiscor$estimate,3)
  mylb[1,"AS"]=round(thiscor$conf.int[1],3)
  myhb[1,"AS"]=round(thiscor$conf.int[2],3)
  myn[1,"AS"]=round(thiscor$parameter,3)
  myp[1,"AS"]=round(thiscor$p.value,3)


#repeat for LR
  abline(lm(proportions[,"z_lp"]~proportions[,"l_lp"],subset=c(proportions$Diagnosis!="AngelmanSyndron

    #get cor properties
thiscor=cor.test(proportions[proportions$Diagnosis!="AngelmanSyndrome","l_lp"],proportions[proportions
  myr[1,"LR"]=round(thiscor$estimate,3)
  mylb[1,"LR"]=round(thiscor$conf.int[1],3)
  myhb[1,"LR"]=round(thiscor$conf.int[2],3)
  myn[1,"LR"]=round(thiscor$parameter,3)
  myp[1,"LR"]=round(thiscor$p.value,3)

#repeat for all cor

    thiscor=cor.test(proportions$z_lp,proportions$l_lp)
  myr[1,"all"]=round(thiscor$estimate,3)
  mylb[1,"all"]=round(thiscor$conf.int[1],3)
  myhb[1,"all"]=round(thiscor$conf.int[2],3)
  myn[1,"all"]=round(thiscor$parameter,3)
  myp[1,"all"]=round(thiscor$p.value,3)

  # repeat whole for cp

  lims=range(c(proportions[,"z_cp"],proportions[,"l_cp"]))

plot(proportions[,"z_cp"]~proportions[,"l_cp"], pch=20,xlab=prettynames["l_cp"],ylab=prettynames["z_cp
     xlim=lims,ylim=lims,
     col=mycols[proportions$Diagnosis])
lines(c(0,1),c(0,1),lty=2,col="darkgray")

abline(lm(proportions[,"z_cp"]~proportions[,"l_cp"],subset=c(proportions$Diagnosis=="AngelmanSyndrome

#get cor properties
thiscor=cor.test(proportions[proportions$Diagnosis=="AngelmanSyndrome","z_cp"],proportions[proportions
  myr[2,"AS"]=round(thiscor$estimate,3)
  mylb[2,"AS"]=round(thiscor$conf.int[1],3)
  myhb[2,"AS"]=round(thiscor$conf.int[2],3)
  myn[2,"AS"]=round(thiscor$parameter,3)
  myp[2,"AS"]=round(thiscor$p.value,3)


#repeat for LR
  abline(lm(proportions[,"z_cp"]~proportions[,"l_cp"],subset=c(proportions$Diagnosis!="AngelmanSyndron
```
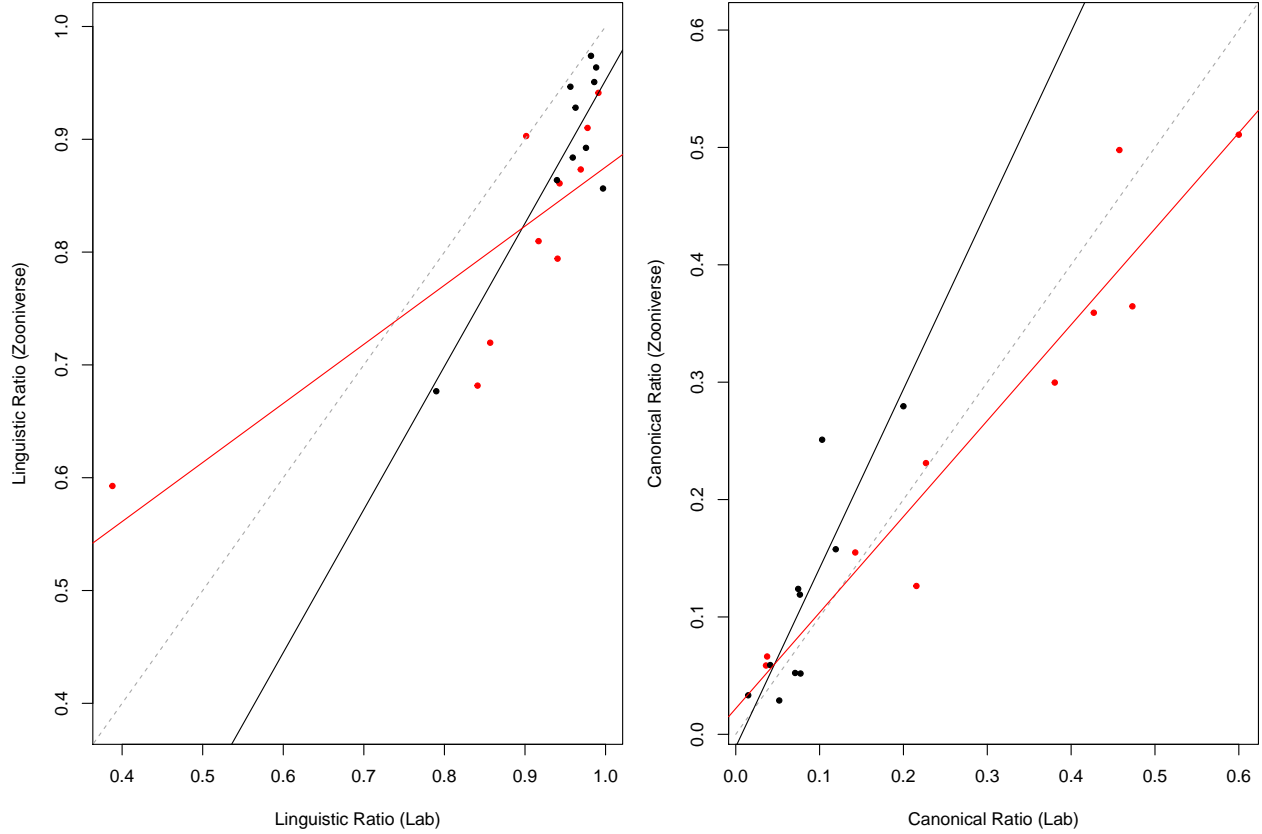
```
    #get cor properties
  thiscor=cor.test(proportions[proportions$Diagnosis!="AngelmanSyndrome","l_cp"],proportions[proportions
    myr[2,"LR"]=round(thiscor$estimate,3)
    mylb[2,"LR"]=round(thiscor$conf.int[1],3)
    myhb[2,"LR"]=round(thiscor$conf.int[2],3)
    myn[2,"LR"]=round(thiscor$parameter,3)
    myp[2,"LR"]=round(thiscor$p.value,3)


#repeat for all cor

    thiscor=cor.test(proportions$z_cp,proportions$l_cp)
    myr[2,"all"]=round(thiscor$estimate,3)
    mylb[2,"all"]=round(thiscor$conf.int[1],3)
    myhb[2,"all"]=round(thiscor$conf.int[2],3)
    myn[2,"all"]=round(thiscor$parameter,3)
    myp[2,"all"]=round(thiscor$p.value,3)
```

Although the classification at the clip level is only moderately accurate, what we are ultimately interested in is whether citizen scientists' classifications are able to provide a reliable snapshot of childrens' individual development. Looking at all 20 children together, we found a strong positive correlation (r (18) = 0.811, CI 0.58-0.92, p<0.001) between Linguistic proportion by child from the Zooniverse and the lab annotators' data. We notice a trend for Linguistic Proportions from Zooniverse data to be lower than that from laboratory data (see Fig. 3). When we split by participant group, correlations remain high [Angelman Syndrome r(8) = .88, CI .55-.97, p<0.001; low-risk control r(8) =.82, CI .4 - .96, p<0.01].
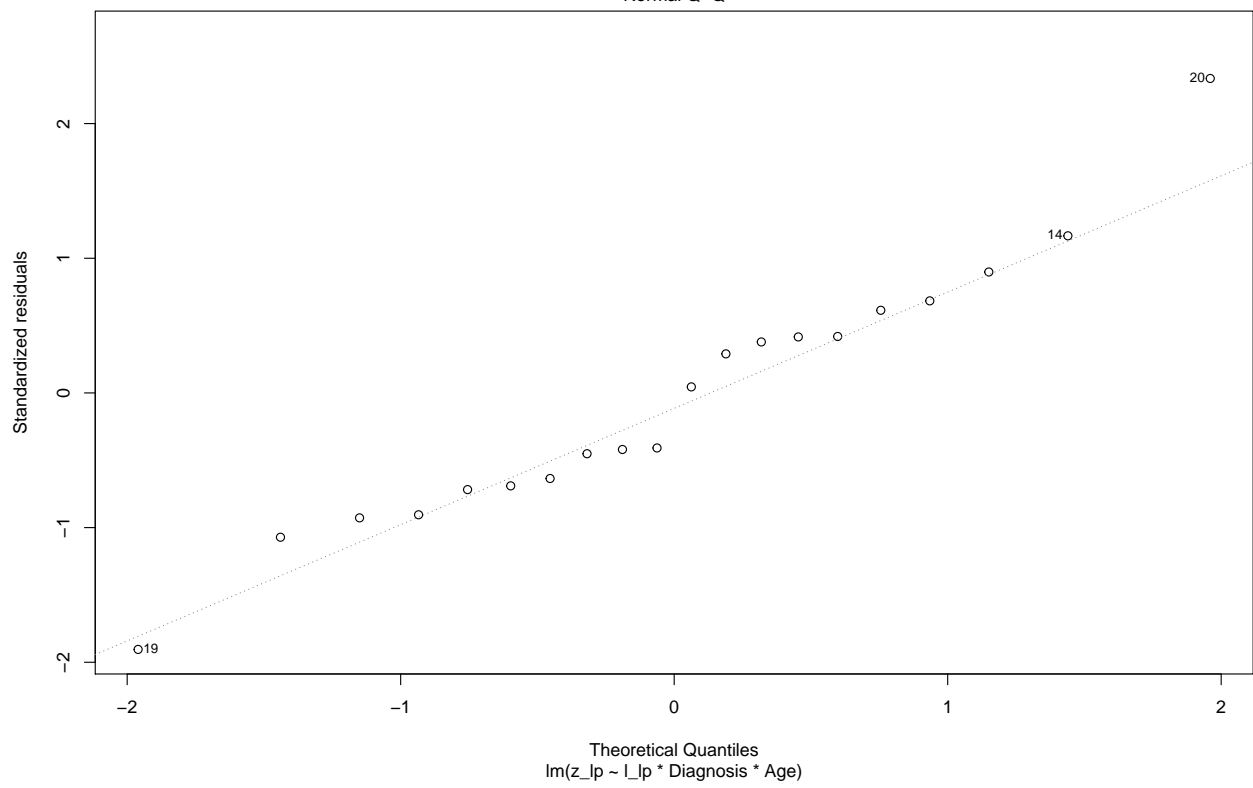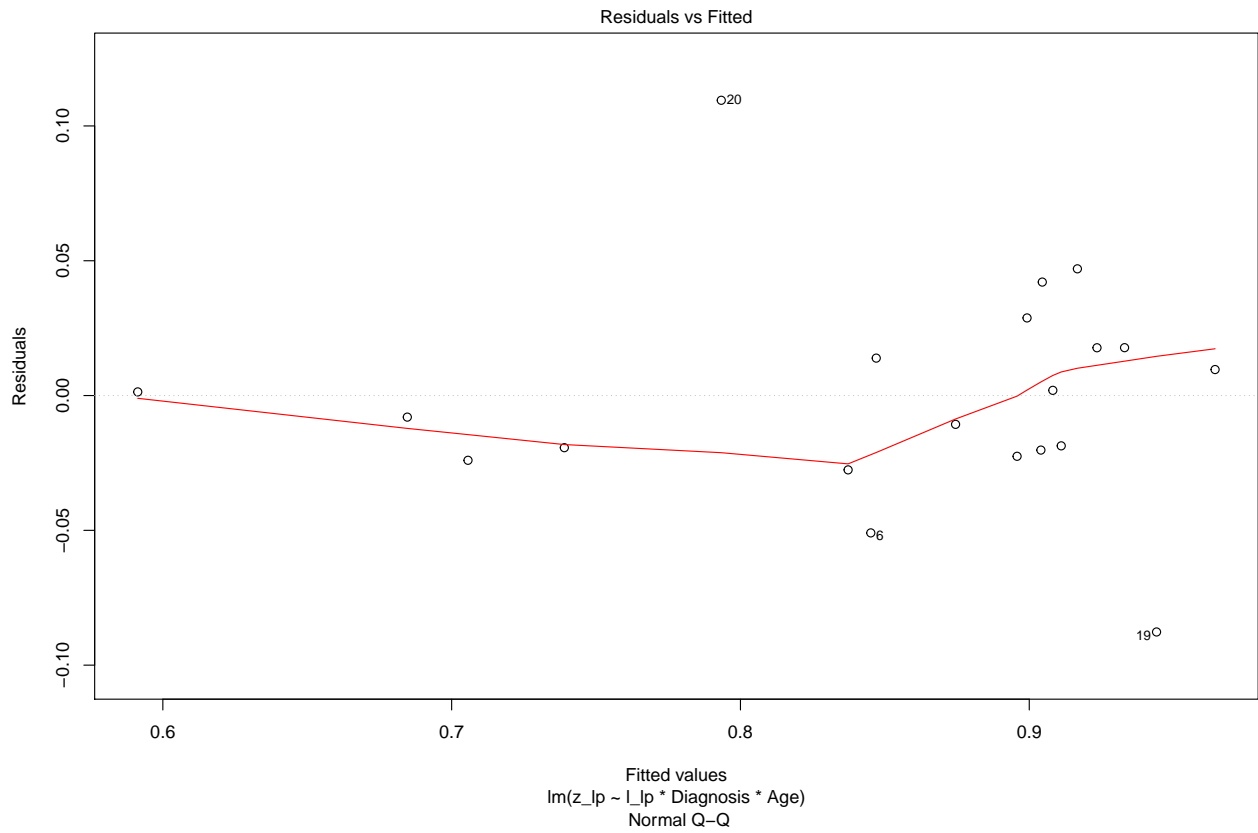
Similarly, a strong positive correlation (r=0.937, 0.84 0.97, p<0.001) is found in the Canonical Proportion. Unlike the linguistic proportion, the error for Canonical Ratio is very small and does not suggest a systematic over- or under-estimation (see Fig. XX). When we split by participant group, correlations remain high
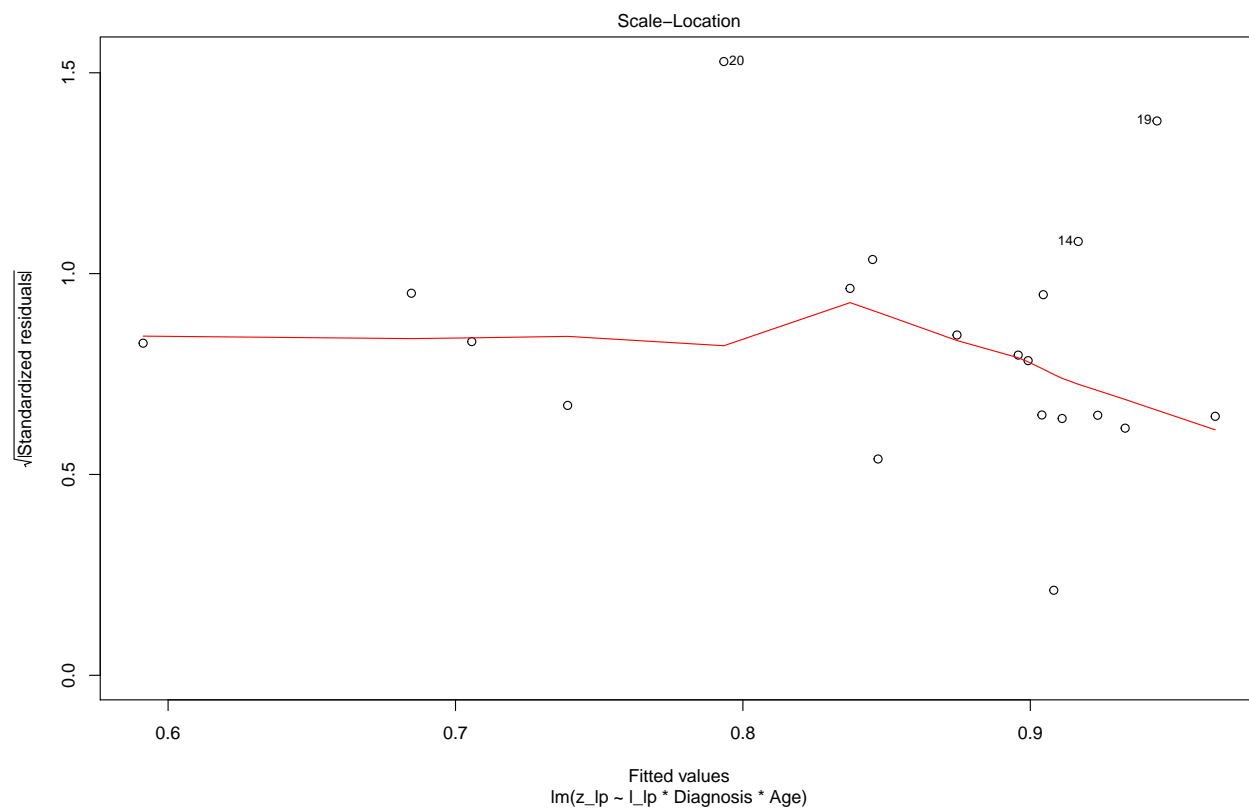
although we do note they are somewhat smaller for the children with Angelman Syndrome, who are also older than the low-risk control children [Angelman Syndrome r(8) = .86, CI 0.5-.96, p<0.01; low-risk control r(8) = .96, CI 0.83-0.99, p<0.001].

When we use chunk-level judgments to derive the our child-level metrics, we observe very similar levels of correlation: linguistic proportion overall r (18) = .84, CI .63 .93, p<0.001 [Angelman Syndrome r(8) = .82, CI .41 .95, p<0.01; low-risk control r(8) = .84, CI .44 0.96, p<0.01]; canonical proportion overall r (18) = .96, CI .90 .98, p<0.001 [Angelman Syndrome r(8) = .85, CI .49 .85, p<0.01; low-risk control r(8) = .97, CI .89 .99, p<0.001].

```
lin_mod=lm(z_lp~l_lp*Diagnosis*Age,data=proportions)
summary(lin_mod)
```

```
##
## Call:
## lm(formula = z_lp ~ l_lp * Diagnosis * Age, data = proportions)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.087697 -0.020805 -0.003327  0.017732  0.109518
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       -1.89796    2.53800  -0.748   0.4690
## l_lp                               2.95125    2.65350   1.112   0.2878
## DiagnosisLow-RiskControl           2.84516    2.54890   1.116   0.2862
## Age                                0.03583    0.05740   0.624   0.5442
## l_lp:DiagnosisLow-RiskControl     -3.03307    2.66494  -1.138   0.2773
## l_lp:Age                          -0.03819    0.05996  -0.637   0.5362
## DiagnosisLow-RiskControl:Age      -0.15256    0.07332  -2.081   0.0595 .
## l_lp:DiagnosisLow-RiskControl:Age  0.16017    0.07535   2.126   0.0550 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05096 on 12 degrees of freedom
## Multiple R-squared:  0.8575, Adjusted R-squared:  0.7744
## F-statistic: 10.32 on 7 and 12 DF,  p-value: 0.0003014
```

```
plot(lin_mod)
```

Residuals vs Fitted

Residuals

Fitted values
lm(z_lp ~ l_lp * Diagnosis * Age)

Normal Q–Q

Standardized residuals

Theoretical Quantiles
lm(z_lp ~ l_lp * Diagnosis * Age)

Scale–Location

lm(z_lp ~ l_lp * Diagnosis * Age)

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```


Residuals vs Leverage

lm(z_lp ~ l_lp * Diagnosis * Age)
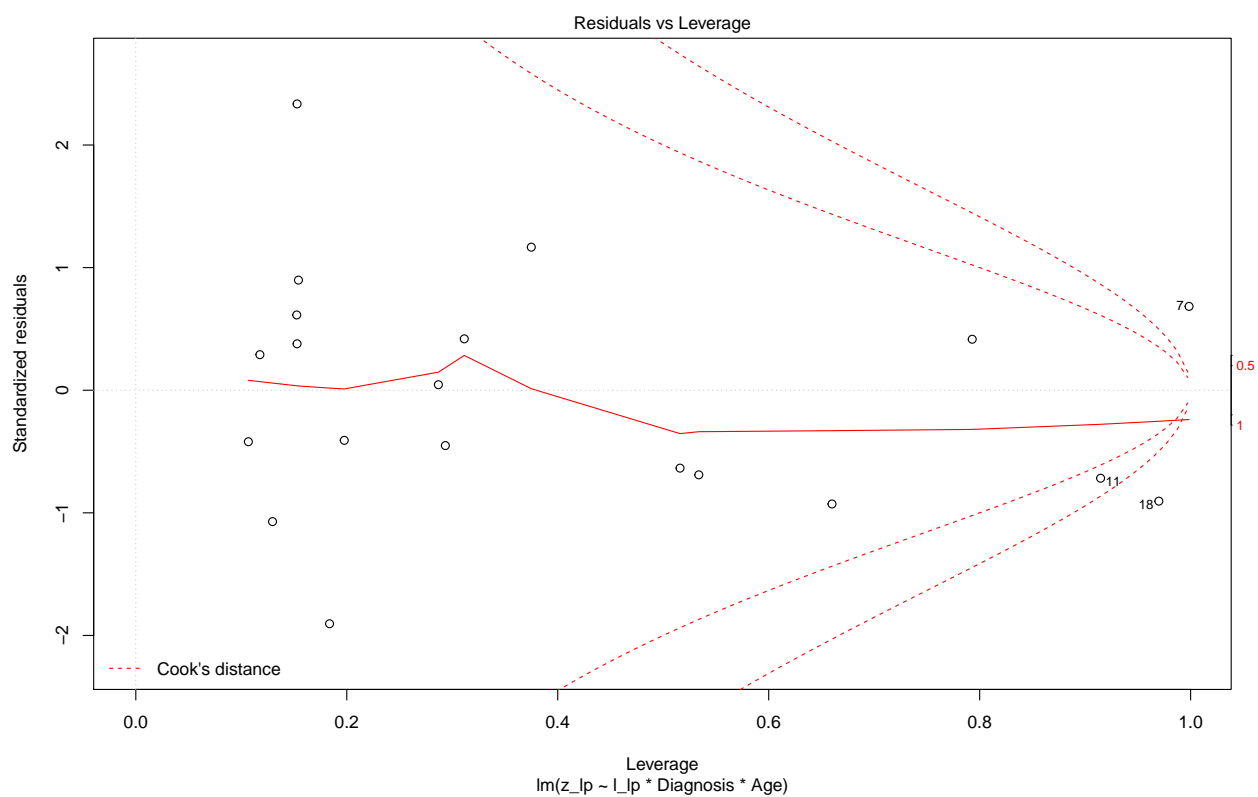
Additional analyses

We additionally explored under what conditions Zooniverse judgments more closely aligned with laboratory judgments:

- take all 5 Zoo judgments or 3 or 1 (may need to run the latter two a couple of times to check for potential variance – if variance is small, we'll just say so, if it's large then we'll have to write a loop and report CIs)
- look at whether kids who have a high "junk" proportion according to Zooniverse
  - lead to lower Zoo-PU correlations
  - have lower levels of concordance across Zoo coders
- look at what happens to TD Zoo-PU correlation when you undersample 100 vocs to 80 and 60 and to the same number of non-junk vocs as in AS