Analyses for JSLHR version

2020-10-17

Contents

History: . . . 1
TODO: . . . 1
Read data in . . . 1
Correspondence between lab & zooniverse annotation at the level of segments . . . 2
Precision . . . 8
Recall . . . 10
repeat collapsing . . . 11
Child level descriptors . . . 17
Correlations with age . . . 18
Correlations across zooniverse and lab . . . 19
Code junkyard . . . 23

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

Read data in

```
# 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
```

```
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
```

Correspondence between lab & zooniverse annotation at the level of segments

Here we look at to what extent zooniverse and lab annotations match at the level of individual segments. Each data point is one segment (one "vocalization").

```
table(data_all$lab)
```

```
##
##                     Canonical                  Non-Canonical
##                          1779                           6423
##                        Crying                      Laughing
##                           588                            186
##                          Junk              Canonical_Crying
##                          2595                             0
##             Canonical_Laughing              Crying_Canonical
##                             0                             0
##     Crying_Canonical_Laughing               Crying_Laughing
##                             0                             0
## Crying_Laughing_Non-Canonical          Crying_Non-Canonical
##                             0                             0
## Crying_Non-Canonical_Laughing               Laughing_Crying
##                             0                             0
##        Laughing_Non-Canonical Laughing_Non-Canonical_Crying
##                             0                             0
##          Non-Canonical_Crying        Non-Canonical_Laughing
##                             0                             0
```

```
table(data_all$Zoon_classif)
```

```
##
##                     Canonical                  Non-Canonical
##                          1665                           5525
##                        Crying                      Laughing
##                           920                            442
##                          Junk              Canonical_Crying
##                          1456                             3
##             Canonical_Laughing              Crying_Canonical
##                            15                            30
```

```
##     Crying_Canonical_Laughing               Crying_Laughing
##                             2                           108
## Crying_Laughing_Non-Canonical          Crying_Non-Canonical
##                            75                           655
## Crying_Non-Canonical_Laughing               Laughing_Crying
##                             7                             8
##        Laughing_Non-Canonical Laughing_Non-Canonical_Crying
##                           124                             9
##          Non-Canonical_Crying        Non-Canonical_Laughing
##                           338                           211
```

```
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
```

```
## Confusion Matrix and Statistics
##
##                               Zooniverse
## Lab                            Canonical Non-Canonical Crying Laughing Junk
##   Canonical                         1014           524     31       29   79
##   Non-Canonical                      255          4176    536      154  360
##   Crying                               6            32    239        8    6
##   Laughing                             2            12      8      115    6
##   Junk                               372           776    106      136 1005
##   Canonical_Crying                     0             0      0        0    0
##   Canonical_Laughing                   0             0      0        0    0
##   Crying_Canonical                     0             0      0        0    0
##   Crying_Canonical_Laughing            0             0      0        0    0
##   Crying_Laughing                      0             0      0        0    0
##   Crying_Laughing_Non-Canonical        0             0      0        0    0
##   Crying_Non-Canonical                 0             0      0        0    0
##   Crying_Non-Canonical_Laughing        0             0      0        0    0
##   Laughing_Crying                      0             0      0        0    0
##   Laughing_Non-Canonical               0             0      0        0    0
##   Laughing_Non-Canonical_Crying        0             0      0        0    0
##   Non-Canonical_Crying                 0             0      0        0    0
##   Non-Canonical_Laughing               0             0      0        0    0
##                               Zooniverse
## Lab                            Canonical_Crying Canonical_Laughing
##   Canonical                                   0                  2
##   Non-Canonical                               3                  5
##   Crying                                      0                  0
##   Laughing                                    0                  0
##   Junk                                        0                  8
##   Canonical_Crying                            0                  0
##   Canonical_Laughing                          0                  0
##   Crying_Canonical                            0                  0
##   Crying_Canonical_Laughing                   0                  0
##   Crying_Laughing                             0                  0
##   Crying_Laughing_Non-Canonical               0                  0
##   Crying_Non-Canonical                        0                  0
##   Crying_Non-Canonical_Laughing               0                  0
```

```
##   Laughing_Crying                              0                    0
##   Laughing_Non-Canonical                       0                    0
##   Laughing_Non-Canonical_Crying                0                    0
##   Non-Canonical_Crying                         0                    0
##   Non-Canonical_Laughing                       0                    0
##                               Zooniverse
## Lab                            Crying_Canonical Crying_Canonical_Laughing
##   Canonical                                   21                         2
##   Non-Canonical                                5                         0
##   Crying                                       1                         0
##   Laughing                                     0                         0
##   Junk                                         2                         0
##   Canonical_Crying                             0                         0
##   Canonical_Laughing                           0                         0
##   Crying_Canonical                             0                         0
##   Crying_Canonical_Laughing                    0                         0
##   Crying_Laughing                              0                         0
##   Crying_Laughing_Non-Canonical                0                         0
##   Crying_Non-Canonical                         0                         0
##   Crying_Non-Canonical_Laughing                0                         0
##   Laughing_Crying                              0                         0
##   Laughing_Non-Canonical                       0                         0
##   Laughing_Non-Canonical_Crying                0                         0
##   Non-Canonical_Crying                         0                         0
##   Non-Canonical_Laughing                       0                         0
##                               Zooniverse
## Lab                            Crying_Laughing Crying_Laughing_Non-Canonical
##   Canonical                                  2                            2
##   Non-Canonical                             32                           16
##   Crying                                    48                           50
##   Laughing                                  11                            1
##   Junk                                      15                            6
##   Canonical_Crying                           0                            0
##   Canonical_Laughing                         0                            0
##   Crying_Canonical                           0                            0
##   Crying_Canonical_Laughing                  0                            0
##   Crying_Laughing                            0                            0
##   Crying_Laughing_Non-Canonical              0                            0
##   Crying_Non-Canonical                       0                            0
##   Crying_Non-Canonical_Laughing              0                            0
##   Laughing_Crying                            0                            0
##   Laughing_Non-Canonical                     0                            0
##   Laughing_Non-Canonical_Crying              0                            0
##   Non-Canonical_Crying                       0                            0
##   Non-Canonical_Laughing                     0                            0
##                               Zooniverse
## Lab                            Crying_Non-Canonical
##   Canonical                                    34
##   Non-Canonical                               420
##   Crying                                      165
##   Laughing                                      1
##   Junk                                         35
##   Canonical_Crying                              0
##   Canonical_Laughing                            0
```

```
##   Crying_Canonical                                0
##   Crying_Canonical_Laughing                       0
##   Crying_Laughing                                 0
##   Crying_Laughing_Non-Canonical                   0
##   Crying_Non-Canonical                            0
##   Crying_Non-Canonical_Laughing                   0
##   Laughing_Crying                                 0
##   Laughing_Non-Canonical                          0
##   Laughing_Non-Canonical_Crying                   0
##   Non-Canonical_Crying                            0
##   Non-Canonical_Laughing                          0
##                               Zooniverse
## Lab                            Crying_Non-Canonical_Laughing Laughing_Crying
##   Canonical                                                1               1
##   Non-Canonical                                            4               5
##   Crying                                                   1               1
##   Laughing                                                 0               0
##   Junk                                                     1               1
##   Canonical_Crying                                         0               0
##   Canonical_Laughing                                       0               0
##   Crying_Canonical                                         0               0
##   Crying_Canonical_Laughing                                0               0
##   Crying_Laughing                                          0               0
##   Crying_Laughing_Non-Canonical                            0               0
##   Crying_Non-Canonical                                     0               0
##   Crying_Non-Canonical_Laughing                            0               0
##   Laughing_Crying                                          0               0
##   Laughing_Non-Canonical                                   0               0
##   Laughing_Non-Canonical_Crying                            0               0
##   Non-Canonical_Crying                                     0               0
##   Non-Canonical_Laughing                                   0               0
##                               Zooniverse
## Lab                            Laughing_Non-Canonical
##   Canonical                                        18
##   Non-Canonical                                    62
##   Crying                                            5
##   Laughing                                          9
##   Junk                                             30
##   Canonical_Crying                                  0
##   Canonical_Laughing                                0
##   Crying_Canonical                                  0
##   Crying_Canonical_Laughing                         0
##   Crying_Laughing                                   0
##   Crying_Laughing_Non-Canonical                     0
##   Crying_Non-Canonical                              0
##   Crying_Non-Canonical_Laughing                     0
##   Laughing_Crying                                   0
##   Laughing_Non-Canonical                            0
##   Laughing_Non-Canonical_Crying                     0
##   Non-Canonical_Crying                              0
##   Non-Canonical_Laughing                            0
##                               Zooniverse
## Lab                            Laughing_Non-Canonical_Crying
##   Canonical                                                0
```

```
##   Non-Canonical                                          4
##   Crying                                                 3
##   Laughing                                               1
##   Junk                                                   1
##   Canonical_Crying                                       0
##   Canonical_Laughing                                     0
##   Crying_Canonical                                       0
##   Crying_Canonical_Laughing                              0
##   Crying_Laughing                                        0
##   Crying_Laughing_Non-Canonical                          0
##   Crying_Non-Canonical                                   0
##   Crying_Non-Canonical_Laughing                          0
##   Laughing_Crying                                        0
##   Laughing_Non-Canonical                                 0
##   Laughing_Non-Canonical_Crying                          0
##   Non-Canonical_Crying                                   0
##   Non-Canonical_Laughing                                 0
##                                Zooniverse
## Lab                             Non-Canonical_Crying Non-Canonical_Laughing
##   Canonical                                        8                     11
##   Non-Canonical                                  271                    115
##   Crying                                          20                      3
##   Laughing                                         0                     20
##   Junk                                            39                     62
##   Canonical_Crying                                 0                      0
##   Canonical_Laughing                               0                      0
##   Crying_Canonical                                 0                      0
##   Crying_Canonical_Laughing                        0                      0
##   Crying_Laughing                                  0                      0
##   Crying_Laughing_Non-Canonical                    0                      0
##   Crying_Non-Canonical                             0                      0
##   Crying_Non-Canonical_Laughing                    0                      0
##   Laughing_Crying                                  0                      0
##   Laughing_Non-Canonical                           0                      0
##   Laughing_Non-Canonical_Crying                    0                      0
##   Non-Canonical_Crying                             0                      0
##   Non-Canonical_Laughing                           0                      0
##
## Overall Statistics
##
##                Accuracy : 0.566
##                  95% CI : (0.5569, 0.575)
##     No Information Rate : 0.4771
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.3621
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Canonical Class: Non-Canonical Class: Crying
## Sensitivity                    0.61492               0.7565       0.25978
## Specificity                    0.92290               0.6287       0.96723
```

```
## Pos Pred Value               0.56998              0.6502       0.40646
## Neg Pred Value               0.93515              0.7389       0.93800
## Prevalence                   0.14251              0.4771       0.07951
## Detection Rate               0.08763              0.3609       0.02066
## Detection Prevalence         0.15375              0.5551       0.05082
## Balanced Accuracy            0.76891              0.6926       0.61351
##                      Class: Laughing Class: Junk Class: Canonical_Crying
## Sensitivity                 0.260181     0.69025                0.0000000
## Specificity                 0.993620     0.84281                1.0000000
## Pos Pred Value              0.618280     0.38728                      NaN
## Neg Pred Value              0.971278     0.94975                0.9997407
## Prevalence                  0.038199     0.12583                0.0002593
## Detection Rate              0.009939     0.08686                0.0000000
## Detection Prevalence        0.016075     0.22427                0.0000000
## Balanced Accuracy           0.626901     0.76653                0.5000000
##                      Class: Canonical_Laughing Class: Crying_Canonical
## Sensitivity                           0.000000                0.000000
## Specificity                           1.000000                1.000000
## Pos Pred Value                             NaN                     NaN
## Neg Pred Value                        0.998704                0.997494
## Prevalence                            0.001296                0.002506
## Detection Rate                        0.000000                0.000000
## Detection Prevalence                  0.000000                0.000000
## Balanced Accuracy                     0.500000                0.500000
##                      Class: Crying_Canonical_Laughing Class: Crying_Laughing
## Sensitivity                                 0.0000000               0.000000
## Specificity                                 1.0000000               1.000000
## Pos Pred Value                                    NaN                     NaN
## Neg Pred Value                              0.9998272               0.990666
## Prevalence                                  0.0001728               0.009334
## Detection Rate                              0.0000000               0.000000
## Detection Prevalence                        0.0000000               0.000000
## Balanced Accuracy                           0.5000000               0.500000
##                      Class: Crying_Laughing_Non-Canonical
## Sensitivity                                      0.000000
## Specificity                                      1.000000
## Pos Pred Value                                        NaN
## Neg Pred Value                                   0.993518
## Prevalence                                       0.006482
## Detection Rate                                   0.000000
## Detection Prevalence                             0.000000
## Balanced Accuracy                                0.500000
##                      Class: Crying_Non-Canonical
## Sensitivity                              0.00000
## Specificity                              1.00000
## Pos Pred Value                               NaN
## Neg Pred Value                           0.94339
## Prevalence                               0.05661
## Detection Rate                           0.00000
## Detection Prevalence                     0.00000
## Balanced Accuracy                        0.50000
##                      Class: Crying_Non-Canonical_Laughing
## Sensitivity                                      0.000000
## Specificity                                      1.000000
```

```
## Pos Pred Value                                    NaN
## Neg Pred Value                               0.999395
## Prevalence                                   0.000605
## Detection Rate                               0.000000
## Detection Prevalence                         0.000000
## Balanced Accuracy                            0.500000
##                      Class: Laughing_Crying Class: Laughing_Non-Canonical
## Sensitivity                       0.0000000                        0.00000
## Specificity                       1.0000000                        1.00000
## Pos Pred Value                          NaN                            NaN
## Neg Pred Value                    0.9993086                        0.98928
## Prevalence                        0.0006914                        0.01072
## Detection Rate                    0.0000000                        0.00000
## Detection Prevalence              0.0000000                        0.00000
## Balanced Accuracy                 0.5000000                        0.50000
##                      Class: Laughing_Non-Canonical_Crying
## Sensitivity                                     0.0000000
## Specificity                                     1.0000000
## Pos Pred Value                                        NaN
## Neg Pred Value                                  0.9992222
## Prevalence                                      0.0007778
## Detection Rate                                  0.0000000
## Detection Prevalence                            0.0000000
## Balanced Accuracy                               0.5000000
##                      Class: Non-Canonical_Crying Class: Non-Canonical_Laughing
## Sensitivity                              0.00000                        0.00000
## Specificity                              1.00000                        1.00000
## Pos Pred Value                               NaN                            NaN
## Neg Pred Value                           0.97079                        0.98176
## Prevalence                               0.02921                        0.01824
## Detection Rate                           0.00000                        0.00000
## Detection Prevalence                     0.00000                        0.00000
## Balanced Accuracy                        0.50000                        0.50000
```

Precision

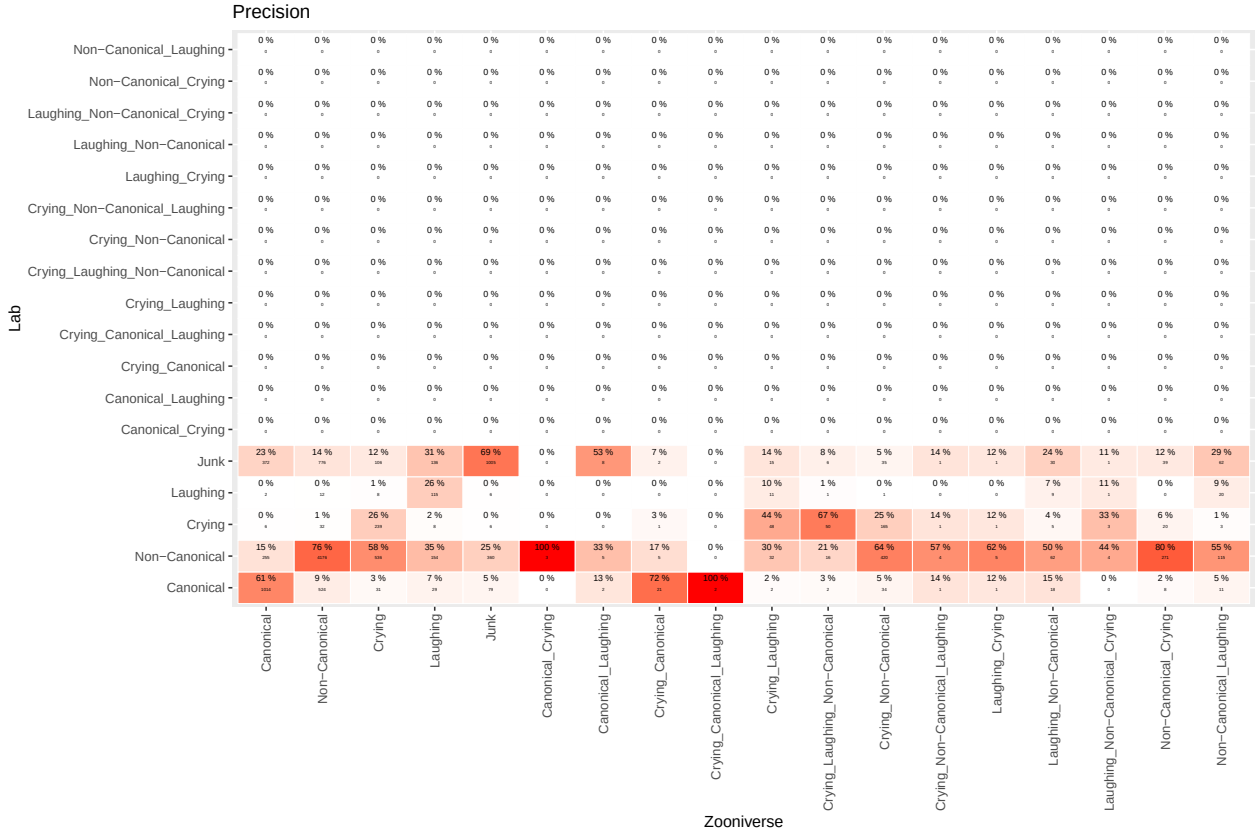
Precision means: If a segment was called X by zooniverse coders, what proportion of the time was it called X by lab coders?

```
colsums=colSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:18) my_conf_tab[,i]=my_conf_tab[,i]/colsums[i]
colSums(my_conf_tab)
```

```
##                     Canonical                 Non-Canonical
##                             1                             1
##                        Crying                      Laughing
##                             1                             1
##                          Junk              Canonical_Crying
##                             1                             1
##            Canonical_Laughing              Crying_Canonical
##                             1                             1
##     Crying_Canonical_Laughing               Crying_Laughing
##                             1                             1
```

```
## Crying_Laughing_Non-Canonical          Crying_Non-Canonical
##                              1                             1
## Crying_Non-Canonical_Laughing              Laughing_Crying
##                              1                             1
##        Laughing_Non-Canonical Laughing_Non-Canonical_Crying
##                              1                             1
##          Non-Canonical_Crying        Non-Canonical_Laughing
##                              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")])
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") +
  ggtitle("Precision")+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



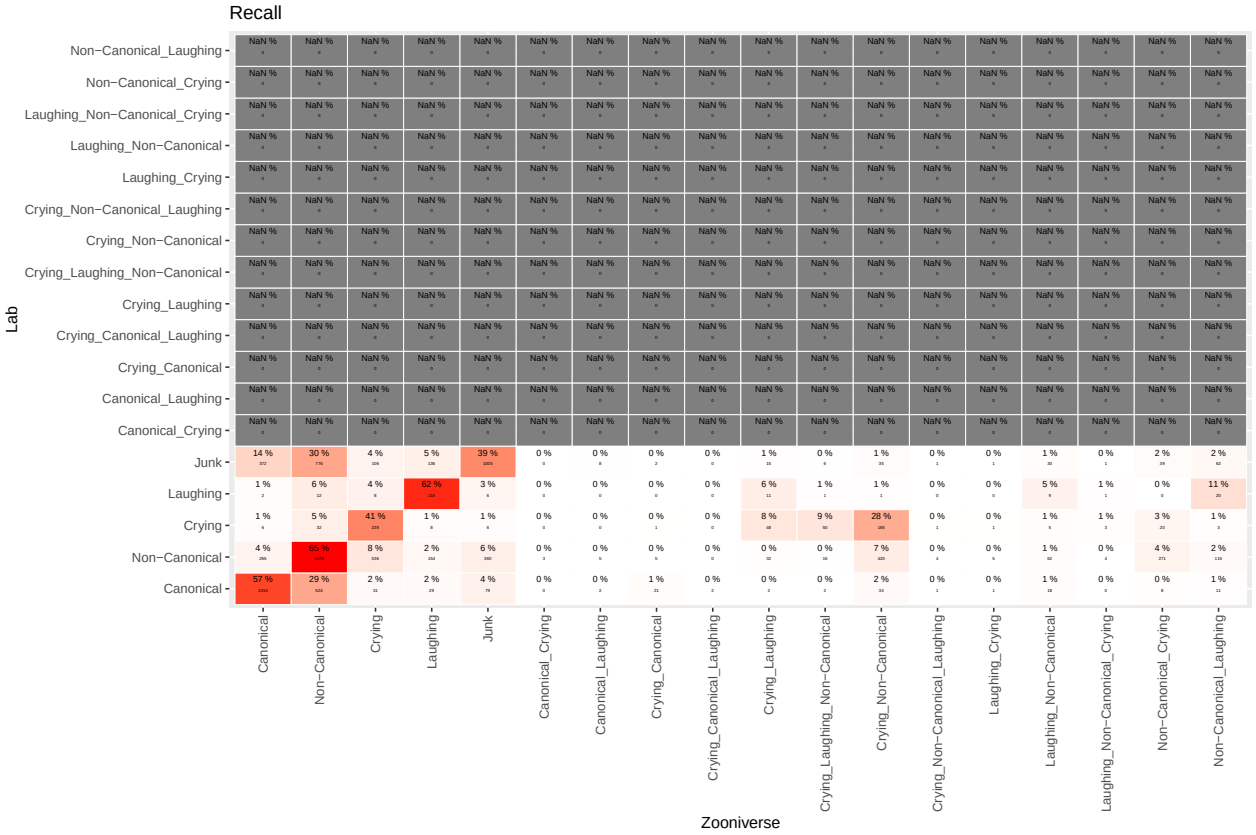
Recall

Recall means: If a segment was called X by lab coders, what proportion of the time was it called X by zooniverse coders?

```
rowsums=rowSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:18) my_conf_tab[,i]=my_conf_tab[,i]/rowsums[i]
rowSums(my_conf_tab)
```

```
##                      Canonical                  Non-Canonical
##                            NaN                            NaN
##                         Crying                       Laughing
##                            NaN                            NaN
##                          Junk              Canonical_Crying
##                            NaN                            NaN
##            Canonical_Laughing              Crying_Canonical
##                            NaN                            NaN
##     Crying_Canonical_Laughing               Crying_Laughing
##                            NaN                            NaN
## Crying_Laughing_Non-Canonical          Crying_Non-Canonical
##                            NaN                            NaN
## Crying_Non-Canonical_Laughing               Laughing_Crying
##                            NaN                            NaN
##        Laughing_Non-Canonical Laughing_Non-Canonical_Crying
##                            NaN                            NaN
##          Non-Canonical_Crying        Non-Canonical_Laughing
##                            NaN                            NaN
```

```
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")])
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))
```



repeat collapsing

```
#given results above, 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)
data_all$lab=factor(data_all$lab)
sample_data<-cbind(data_all$lab,data_all$Zoon_classif)
```

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

```
## Call:
## gac(data = sample_data, kat = 5, weight = c("unweighted"), conf.level = 0.95)
##
##        Estimate    StdErr   LowerCB UpperCB
## Const 0.5866689 0.0053816 0.5761201  0.5972
##
## Confidence level = 95%
## Sample size = 11571
```

```
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
```

```
## Confusion Matrix and Statistics
##
##                Zooniverse
## Lab             Canonical Non-Canonical Crying Laughing Junk
##   Canonical          1039           598     34       29   79
##   Non-Canonical       268          5068    573      154  360
##   Crying                7           279    288        8    6
##   Laughing              2            44     19      115    6
##   Junk                382           950    122      136 1005
##
## Overall Statistics
##
##                Accuracy : 0.6495
##                  95% CI : (0.6407, 0.6582)
##     No Information Rate : 0.5997
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.4265
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: Canonical Class: Non-Canonical Class: Crying
## Sensitivity                   0.61190               0.7304       0.27799
## Specificity                   0.92505               0.7075       0.97152
## Pos Pred Value                0.58404               0.7890       0.48980
## Neg Pred Value                0.93270               0.6366       0.93189
## Prevalence                    0.14675               0.5997       0.08953
## Detection Rate                0.08979               0.4380       0.02489
## Detection Prevalence          0.15375               0.5551       0.05082
## Balanced Accuracy             0.76847               0.7189       0.62476
##                      Class: Laughing Class: Junk
## Sensitivity                 0.260181     0.69025
## Specificity                 0.993620     0.84281
```

```
## Pos Pred Value              0.618280     0.38728
## Neg Pred Value              0.971278     0.94975
## Prevalence                  0.038199     0.12583
## Detection Rate              0.009939     0.08686
## Detection Prevalence        0.016075     0.22427
## Balanced Accuracy           0.626901     0.76653
```

```
pdf("../Results/precision_final.pdf",height=10,width=10)
colsums=colSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:5) 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 columms
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")])
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=8) +
  geom_text(aes(label = Freq), vjust = 1,size=8) +
  scale_fill_gradient(low = "white", high = "red", name = "Proportion") +
     theme(legend.position = "none") +
  xlab("Zooniverse") + ylab("Lab") +
  ggtitle("Precision")+theme(text = element_text(size=20),
        axis.text.x = element_text(angle=90, hjust=1))
dev.off()
```

```
## pdf
##   2
```

```
pdf("../Results/recall_final.pdf",height=10,width=10)
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")])
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=8) +
  geom_text(aes(label = Freq), vjust = 1,size=8) +
  scale_fill_gradient(low = "white", high = "red", name = "Proportion") +
     theme(legend.position = "none") +
  xlab("Zooniverse") + ylab("Lab") +
  ggtitle("Recall")+theme(text = element_text(size=20),
        axis.text.x = element_text(angle=90, hjust=1))
dev.off()
```

```
## pdf
##   2
```

```
data_as_td<-left_join(data_all,demo_data,on="ChildID")
```

```
## Joining, by = c("ChildID", "Age")
```

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

```
# CM with just AS kids
data_AS<-subset(data_as_td, Diagnosis=="AngelmanSyndrome")
mycf=confusionMatrix(data_AS$lab, data_AS$Zoon_classif, dnn = c("Lab","Zooniverse"))
conf_tab=mycf$table
mycf
```

```
## Confusion Matrix and Statistics
##
##                Zooniverse
## Lab             Canonical Non-Canonical Crying Laughing Junk
##   Canonical            90           165      2       13   15
##   Non-Canonical       100          2984    116       92  115
##   Crying                1            39     15        2    1
##   Laughing              0            16      4       57    2
##   Junk                239           618     34       92  462
##
## Overall Statistics
##
##                Accuracy : 0.6841
##                  95% CI : (0.6714, 0.6966)
##     No Information Rate : 0.7247
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.3624
##
##  Mcnemar's Test P-Value : <2e-16
##
## Statistics by Class:
##
##                      Class: Canonical Class: Non-Canonical Class: Crying
## Sensitivity                   0.20930               0.7807      0.087719
## Specificity                   0.95974               0.7087      0.991574
## Pos Pred Value                0.31579               0.8758      0.258621
## Neg Pred Value                0.93185               0.5512      0.970092
## Prevalence                    0.08153               0.7247      0.032423
## Detection Rate                0.01706               0.5658      0.002844
## Detection Prevalence          0.05404               0.6460      0.010997
## Balanced Accuracy             0.58452               0.7447      0.539646
##                      Class: Laughing Class: Junk
## Sensitivity                  0.22266      0.7765
## Specificity                  0.99562      0.7899
## Pos Pred Value               0.72152      0.3197
## Neg Pred Value               0.96169      0.9653
## Prevalence                   0.04854      0.1128
## Detection Rate               0.01081      0.0876
## Detection Prevalence         0.01498      0.2740
## Balanced Accuracy            0.60914      0.7832
```

```
pdf("../Results/precision_AS.pdf",height=10,width=10)
colsums=colSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:5) 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 columms
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")])
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=8) +
  geom_text(aes(label = Freq), vjust = 1,size=8) +
  scale_fill_gradient(low = "white", high = "red", name = "Proportion") +
     theme(legend.position = "none") +
  xlab("Zooniverse") + ylab("Lab") +
  ggtitle("Precision")+theme(text = element_text(size=20),
        axis.text.x = element_text(angle=90, hjust=1))
dev.off()
```

```
## pdf
##   2
```

```
pdf("../Results/recall_AS.pdf",height=10,width=10)
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")])
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=8) +
  geom_text(aes(label = Freq), vjust = 1,size=8) +
  scale_fill_gradient(low = "white", high = "red", name = "Proportion") +
     theme(legend.position = "none") +
  xlab("Zooniverse") + ylab("Lab") +
  ggtitle("Recall")+theme(text = element_text(size=20),
        axis.text.x = element_text(angle=90, hjust=1))
dev.off()
```

```
## pdf
##   2
```

```
# CM with just TD kids
data_TD<-subset(data_as_td, Diagnosis=="Low-RiskControl")
mycf=confusionMatrix(data_TD$lab, data_TD$Zoon_classif, dnn = c("Lab","Zooniverse"))
conf_tab=mycf$table
mycf
```

```
## Confusion Matrix and Statistics
##
##                Zooniverse
## Lab             Canonical Non-Canonical Crying Laughing Junk
##   Canonical           949           433     32       16   64
##   Non-Canonical       168          2084    457       62  245
##   Crying                6           240    273        6    5
##   Laughing              2            28     15       58    4
##   Junk                143           332     88       44  543
##
## Overall Statistics
##
##                Accuracy : 0.6205
##                  95% CI : (0.6083, 0.6325)
##     No Information Rate : 0.495
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.4403
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: Canonical Class: Non-Canonical Class: Crying
## Sensitivity                    0.7484                0.6686       0.31561
## Specificity                    0.8916                0.7069       0.95269
## Pos Pred Value                 0.6352                0.6910       0.51509
## Neg Pred Value                 0.9336                0.6852       0.89735
## Prevalence                     0.2014                0.4950       0.13737
## Detection Rate                 0.1507                0.3310       0.04335
## Detection Prevalence           0.2373                0.4790       0.08417
## Balanced Accuracy              0.8200                0.6878       0.63415
##                      Class: Laughing Class: Junk
## Sensitivity                 0.311828     0.63066
## Specificity                 0.991982     0.88834
## Pos Pred Value              0.542056     0.47217
## Neg Pred Value              0.979321     0.93822
## Prevalence                  0.029538     0.13673
## Detection Rate              0.009211     0.08623
## Detection Prevalence        0.016992     0.18263
## Balanced Accuracy           0.651905     0.75950
```

```
pdf("../Results/precision_TD.pdf",height=10,width=10)
colsums=colSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:5) 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")])
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=8) +
  geom_text(aes(label = Freq), vjust = 1,size=8) +
  scale_fill_gradient(low = "white", high = "red", name = "Proportion") +
     theme(legend.position = "none") +
  xlab("Zooniverse") + ylab("Lab") +
  ggtitle("Precision")+theme(text = element_text(size=20),
        axis.text.x = element_text(angle=90, hjust=1))
dev.off()
```

```
## pdf
##   2
```

```
pdf("../Results/recall_TD.pdf",height=10,width=10)
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")])
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=8) +
  geom_text(aes(label = Freq), vjust = 1,size=8) +
  scale_fill_gradient(low = "white", high = "red", name = "Proportion") +
     theme(legend.position = "none") +
  xlab("Zooniverse") + ylab("Lab") +
  ggtitle("Recall")+theme(text = element_text(size=20),
        axis.text.x = element_text(angle=90, hjust=1))
dev.off()
```

```
## pdf
##   2
```

Child level descriptors

Although there may be errors at the level of the segment, what we really care about is whether Zooniverse annotations give a reliable image of the child's individual development. This is what we look at in this section. In all of these graphs, red points correspond to children diagnosed with Angelman Syndrome, black for low-risk control.

```
#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_lr=rowSums(ztab[,c("Canonical","Non-Canonical")])/rowSums(ztab[,-which(colnames(ztab) %in% c("Junk"))]
z_cr=ztab[,c("Canonical")]/rowSums(ztab[,c("Canonical","Non-Canonical")])
ltab=table(data_all$filename,data_all$lab)
l_lr=rowSums(ltab[,c("Canonical","Non-Canonical")])/rowSums(ltab[,-which(colnames(ztab) %in% c("Junk"))]
l_cr=ltab[,c("Canonical")]/rowSums(ltab[,c("Canonical","Non-Canonical")])
#put all the ratios together
if(sum(rownames(ztab)==rownames(ltab))==dim(ztab)[1]) ratios=cbind(rownames(ztab),z_lr,z_cr,l_lr,l_cr)
colnames(ratios)[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 ratios and demo data.

merge(ratios,demo_data_fn,by="filename")->ratios
colnames(ratios)[dim(ratios)[2]]<-"Age"
#cbinding results in text, so we numerize the ratios
for(thisvar in c("z_lr","z_cr","l_lr","l_cr")) ratios[,thisvar]=as.numeric(as.character(ratios[,thisvar]
summary(ratios)
```

```
##                   filename       z_lr             z_cr
##  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_lr             l_cr           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
##
```

Correlations with age

We first look generally at two measures that have been found to relate to age:

- linguistic ratio = ("Canonical"+"Non-Canonical")/"All vocalizations" (i.e. we remove junk)
- canonical ratio = "Canonical"/("Canonical"+"Non-Canonical") (i.e. we remove junk + non-linguistic vocalizations)

TODO

- make margins smaller
- remove title repetition (age only at the bottom)

```
prettynames=c("Linguistic Ratio (Zooniverse)","Canonical Ratio (Zooniverse)",
              "Linguistic Ratio (Lab)","Canonical Ratio (Lab)" )
names(prettynames)<-c("z_lr","z_cr","l_lr","l_cr")
mycols=c("black","red")
names(mycols)<-c("Low-RiskControl","AngelmanSyndrome")
```

```
mypch=c(4,20)
names(mypch)<-c("Low-RiskControl","AngelmanSyndrome")

jpeg("../Results/corage.jpg",width=20,height=20,units="cm",res=300)
layout(matrix(c(1:4), 2, 2, byrow = F))
for(thisvar in c("z_lr","z_cr","l_lr","l_cr")) {

  plot(ratios[,thisvar]~ratios$Age, pch=mypch[ratios$Diagnosis],xlab="Age (months)",ylab=prettynames[thi
       col=mycols[ratios$Diagnosis])
  abline(lm(ratios[,thisvar]~ratios$Age,subset=c(ratios$Diagnosis=="AngelmanSyndrome")),col="black")
  myr=paste0("r=",round(cor.test(ratios[ratios$Diagnosis=="AngelmanSyndrome",thisvar],ratios$Age[ratios$
  text(mean(ratios$Age[ratios$Diagnosis=="AngelmanSyndrome"]),mean(ratios[ratios$Diagnosis=="AngelmanSyn
    abline(lm(ratios[,thisvar]~ratios$Age,subset=c(ratios$Diagnosis!="AngelmanSyndrome")),col="red")
    myr=paste0("r=",round(cor.test(ratios[ratios$Diagnosis!="AngelmanSyndrome",thisvar],ratios$Age[ratio
    text(mean(ratios$Age[ratios$Diagnosis!="AngelmanSyndrome"]),mean(ratios[ratios$Diagnosis!="AngelmanS
}
dev.off()
```

```
## pdf
##   2
```

Correlations across zooniverse and lab

But the key thing for us: Are Zooniverse annotations describing children similar to lab annotations? The answer is clearly yes.

TODO:

- add code to print out the results paragraphs
- change figure generation to have lines fitted to each group

```
#Ling ratio
pdf("../Results/ling_rat_z_vs_l_final.pdf",height=5,width=5)
lims=range(c(ratios[,"z_lr"],ratios[,"l_lr"]))
  myr=round(cor.test(ratios[,"z_lr"],ratios[,"l_lr"])$estimate,3)
  plot(ratios[,"z_lr"]~ratios[,"l_lr"], pch=20,xlab=prettynames["l_lr"],ylab=prettynames["z_lr"],main=pa
       xlim=lims,ylim=lims,
       col=mycols[ratios$Diagnosis])
  abline(lm(ratios[,"z_lr"]~ratios[,"l_lr"]))
  lines(c(0,1),c(0,1),lty=2,col="darkgray")
dev.off()
```

```
## pdf
##   2
```

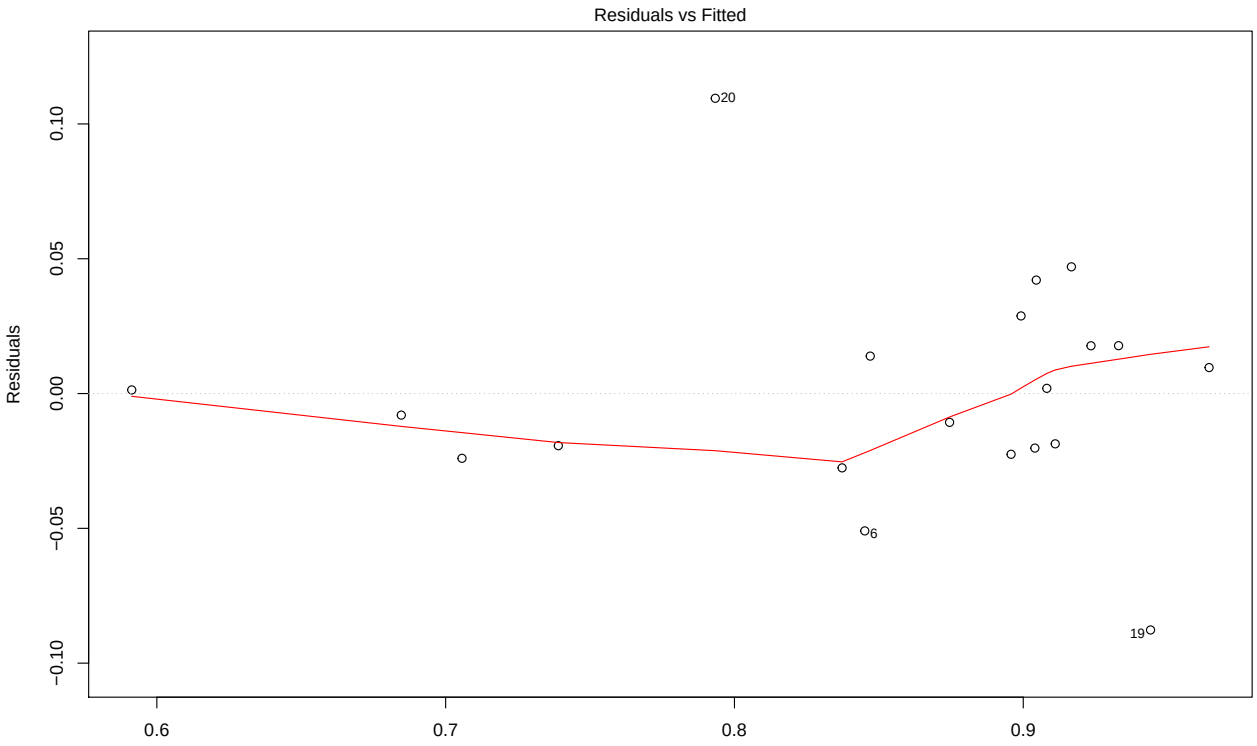
```
  #CR
pdf("../Results/can_rat_z_vs_l_final.pdf",height=5,width=5)
lims=range(c(ratios[,"z_cr"],ratios[,"l_cr"]))
    myr=round(cor.test(ratios[,"z_cr"],ratios[,"l_cr"])$estimate,3)
  plot(ratios[,"z_cr"]~ratios[,"l_cr"], pch=20,xlab=prettynames["l_cr"],ylab=prettynames["z_cr"],main=pa
       xlim=lims,ylim=lims,
       col=mycols[ratios$Diagnosis])
  abline(lm(ratios[,"z_cr"]~ratios[,"l_cr"]),col="darkgray")
    lines(c(0,1),c(0,1),lty=2,col="darkgray")
dev.off()
```

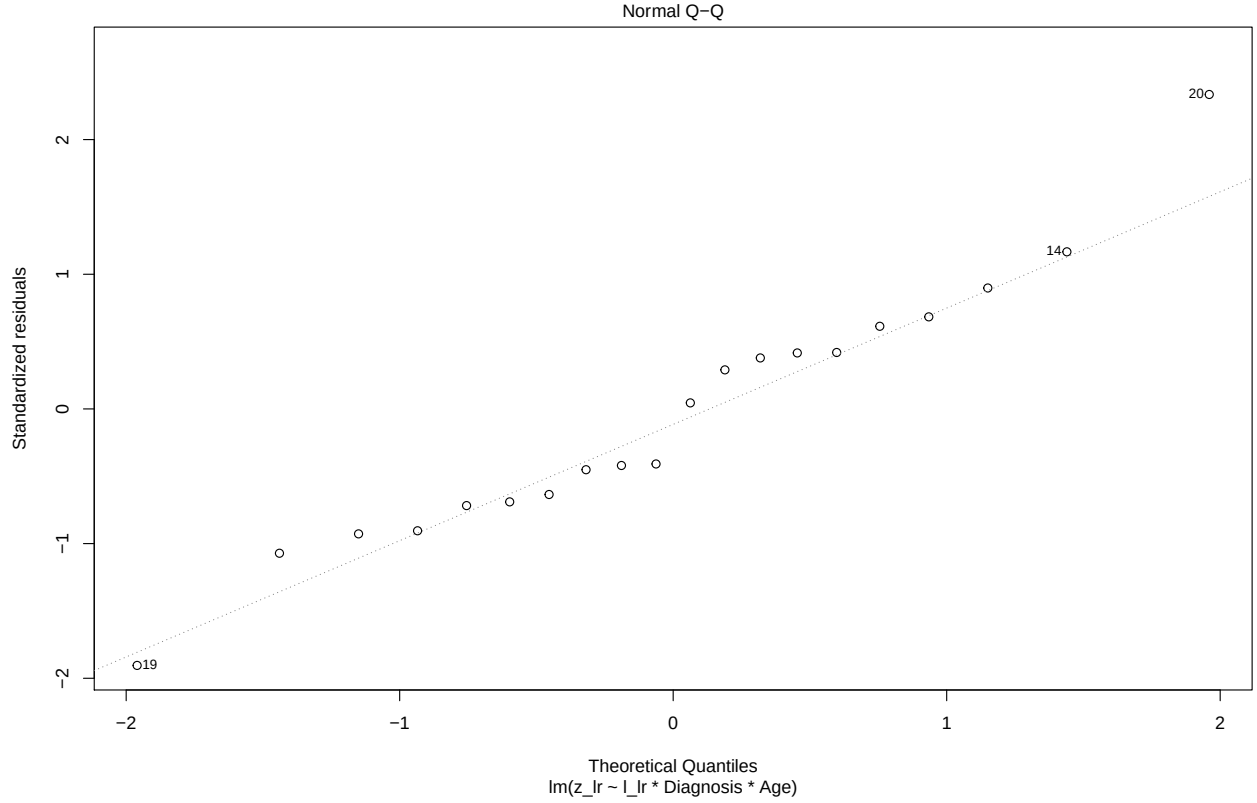
```
## pdf
##   2
```

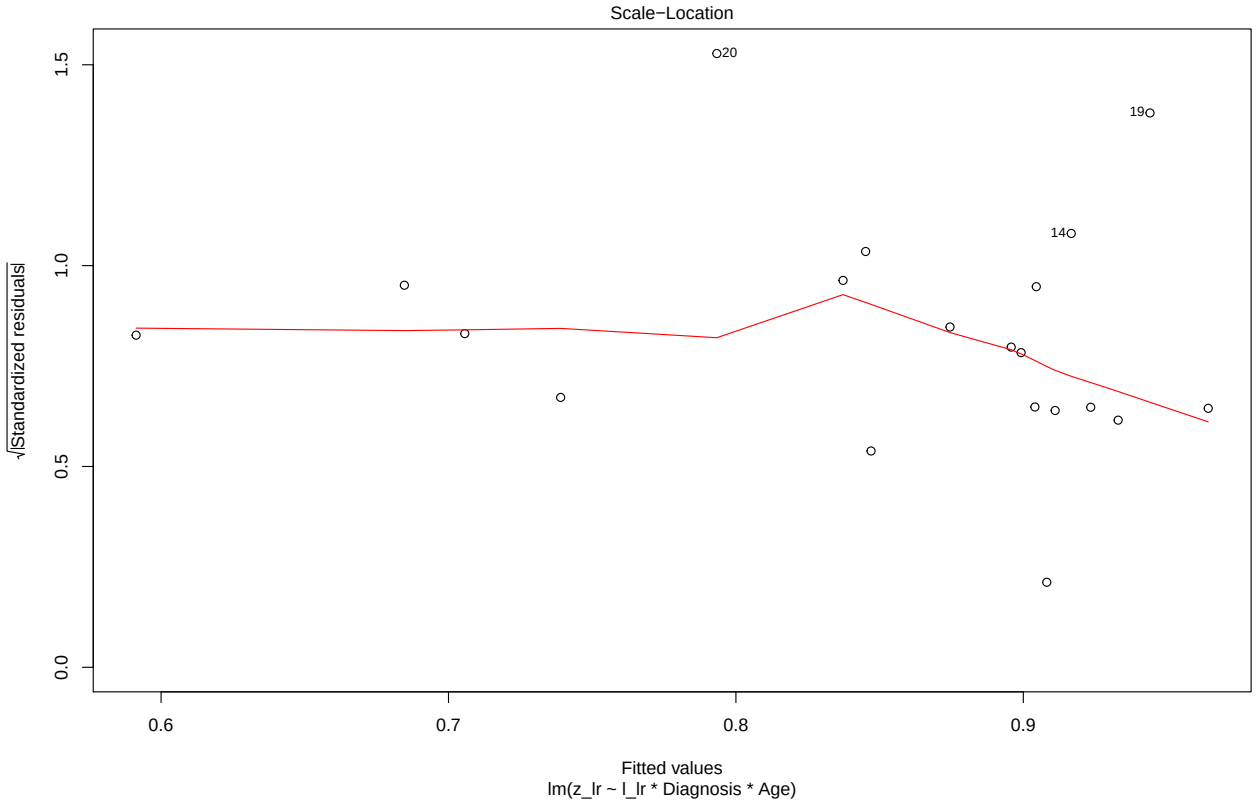
```
lin_mod=lm(z_lr~l_lr*Diagnosis*Age,data=ratios)
summary(lin_mod)
```

```
##
## Call:
## lm(formula = z_lr ~ l_lr * Diagnosis * Age, data = ratios)
##
## 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_lr                               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_lr:DiagnosisLow-RiskControl     -3.03307    2.66494  -1.138   0.2773
## l_lr:Age                          -0.03819    0.05996  -0.637   0.5362
## DiagnosisLow-RiskControl:Age      -0.15256    0.07332  -2.081   0.0595 .
## l_lr: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)
```

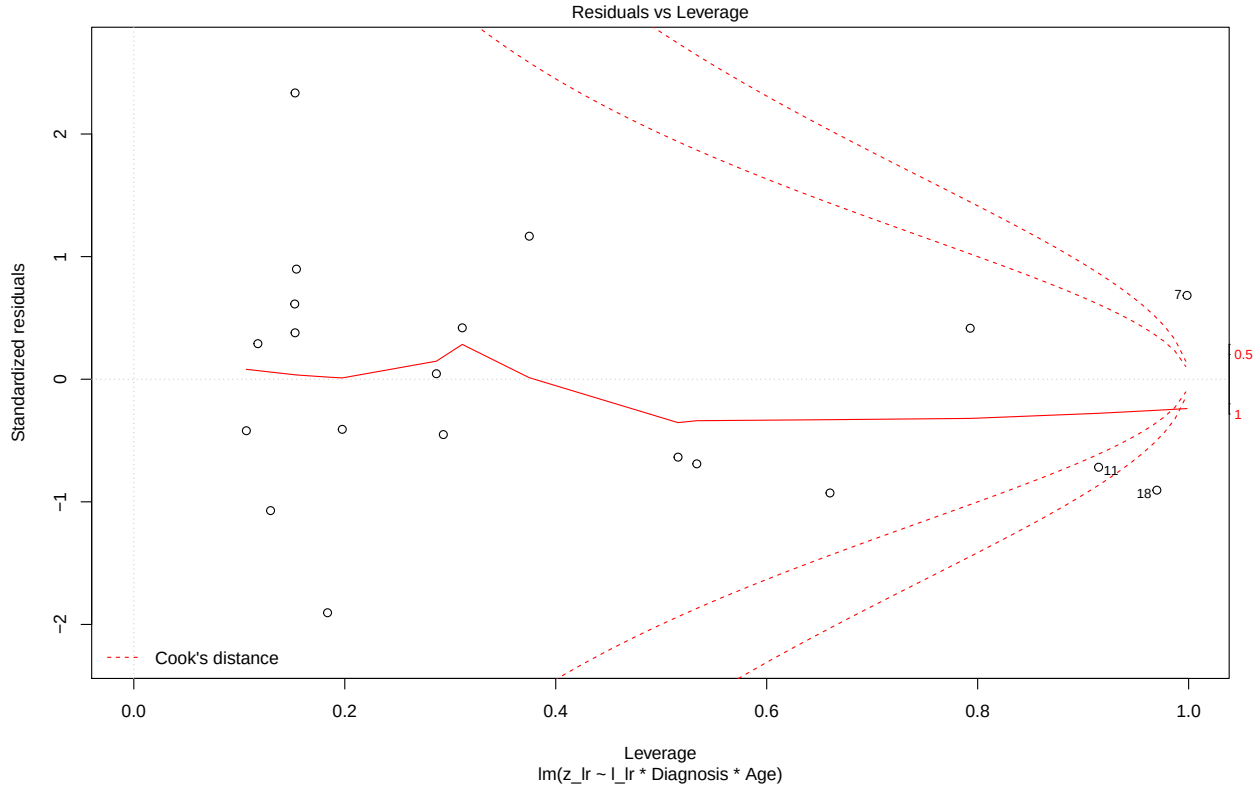

Residuals vs Fitted
Residuals
0.10
0.05
0.00
−0.05
−0.10
20
6
19
0.6
0.7
0.8
0.9
Fitted values
lm(z_lr ~ l_lr * Diagnosis * Age)


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
20
14
19
−2
−1
0
1
2
Theoretical Quantiles
lm(z_lr ~ l_lr * Diagnosis * Age)




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

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



Code junkyard

This is redundant code

Different pdfs for diff groups

```
prettynames=c("AS Linguistic Ratio (Zooniverse)","AS Canonical Ratio (Zooniverse)",
              "AS Linguistic Ratio (Lab)","AS Canonical Ratio (Lab)" )
names(prettynames)<-c("z_lr","z_cr","l_lr","l_cr")

AS_ratios<-subset(ratios, Diagnosis=="AngelmanSyndrome")
TD_ratios<-subset(ratios, Diagnosis=="Low-RiskControl")

for(thisvar in c("z_lr","z_cr","l_lr","l_cr")) {
  myr=round(cor.test(AS_ratios[,thisvar],AS_ratios$Age)$estimate,3)
  plot(AS_ratios[,thisvar]~AS_ratios$Age, pch=20,xlab="Age (months)",ylab=prettynames[thisvar],main=pas
  abline(lm(AS_ratios[,thisvar]~AS_ratios$Age))
  }
```

```
prettynames=c("TD Linguistic Ratio (Zooniverse)","TD Canonical Ratio (Zooniverse)",
              "TD Linguistic Ratio (Lab)","TD Canonical Ratio (Lab)" )
for(thisvar in c("z_lr","z_cr","l_lr","l_cr")) {
  myr=round(cor.test(TD_ratios[,thisvar],TD_ratios$Age)$estimate,3)
  plot(TD_ratios[,thisvar]~TD_ratios$Age, pch=20,xlab="Age (months)",ylab=prettynames[thisvar],main=pas
  abline(lm(TD_ratios[,thisvar]~TD_ratios$Age))
  }
```

```
#Ling ratio
pdf("../Results/td_ling_rat_z_vs_l_final.pdf",height=5,width=5)
lims=range(c(TD_ratios[,"z_lr"],TD_ratios[,"l_lr"]))
  myr=round(cor.test(TD_ratios[,"z_lr"],TD_ratios[,"l_lr"])$estimate,3)
  plot(TD_ratios[,"z_lr"]~TD_ratios[,"l_lr"], pch=20,xlab=prettynames["l_lr"],ylab=prettynames["z_lr"],m
       xlim=lims,ylim=lims,
       col=mycols[TD_ratios$Diagnosis])
  abline(lm(TD_ratios[,"z_lr"]~TD_ratios[,"l_lr"]))
  lines(c(0,1),c(0,1),lty=2,col="darkgray")
dev.off()
  #CR
pdf("../Results/td_can_rat_z_vs_l_final.pdf",height=5,width=5)
lims=range(c(TD_ratios[,"z_cr"],TD_ratios[,"l_cr"]))
    myr=round(cor.test(TD_ratios[,"z_cr"],TD_ratios[,"l_cr"])$estimate,3)
  plot(TD_ratios[,"z_cr"]~TD_ratios[,"l_cr"], pch=20,xlab=prettynames["l_cr"],ylab=prettynames["z_cr"],m
       xlim=lims,ylim=lims,
       col=mycols[TD_ratios$Diagnosis])
  abline(lm(TD_ratios[,"z_cr"]~TD_ratios[,"l_cr"]),col="darkgray")
    lines(c(0,1),c(0,1),lty=2,col="darkgray")
dev.off()
```

```
#Ling ratio
pdf("../Results/AS_ling_rat_z_vs_l_final.pdf",height=5,width=5)
lims=range(c(AS_ratios[,"z_lr"],AS_ratios[,"l_lr"]))
  myr=round(cor.test(AS_ratios[,"z_lr"],AS_ratios[,"l_lr"])$estimate,3)
  plot(AS_ratios[,"z_lr"]~AS_ratios[,"l_lr"], pch=20,xlab=prettynames["l_lr"],ylab=prettynames["z_lr"],m
       xlim=lims,ylim=lims,
       col=mycols[AS_ratios$Diagnosis])
  abline(lm(AS_ratios[,"z_lr"]~AS_ratios[,"l_lr"]))
```

```
  lines(c(0,1),c(0,1),lty=2,col="darkgray")
dev.off()
  #CR
pdf("../Results/AS_can_rat_z_vs_l_final.pdf",height=5,width=5)
lims=range(c(AS_ratios[,"z_cr"],AS_ratios[,"l_cr"]))
    myr=round(cor.test(AS_ratios[,"z_cr"],AS_ratios[,"l_cr"])$estimate,3)
  plot(AS_ratios[,"z_cr"]~AS_ratios[,"l_cr"], pch=20,xlab=prettynames["l_cr"],ylab=prettynames["z_cr"],
       xlim=lims,ylim=lims,
       col=mycols[AS_ratios$Diagnosis])
  abline(lm(AS_ratios[,"z_cr"]~AS_ratios[,"l_cr"]),col="darkgray")
    lines(c(0,1),c(0,1),lty=2,col="darkgray")
dev.off()
```

combined pdf

```
#COMBINED to save space
pdf("../Results/combined_final.pdf",height=5,width=5)
lims=range(c(ratios[,"z_lr"],ratios[,"l_lr"]),c(ratios[,"z_cr"],ratios[,"l_cr"]))
  #myr=round(cor.test(ratios[,"z_lr"],ratios[,"l_lr"])$estimate,3)
  plot(ratios[,"z_lr"]~ratios[,"l_lr"],xlab="Laboratory annotations",ylab="Zooniverse annotations",
       xlim=lims,ylim=lims,
       pch=20,col=mycols[ratios$Diagnosis])
    points(ratios[,"z_cr"]~ratios[,"l_cr"], pch=2, col=mycols[ratios$Diagnosis])
  abline(lm(ratios[,"z_cr"]~ratios[,"l_cr"]))
  abline(lm(ratios[,"z_lr"]~ratios[,"l_lr"]),lty=3)
 # lines(c(0,1),c(0,1),lty=2,col="darkgray")
dev.off()
```