Supplementary materials to: Describing vocalizations in young children: A big data approach through citizen science annotations

Contents

History:

- 2020-11-03 first version

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"))
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 (using LENA segmentation). Unlike in the main paper, here we will show results before applying the ordered rules that give prevalence to canonical, non-canonical, laughing, crying (in that order).

```
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

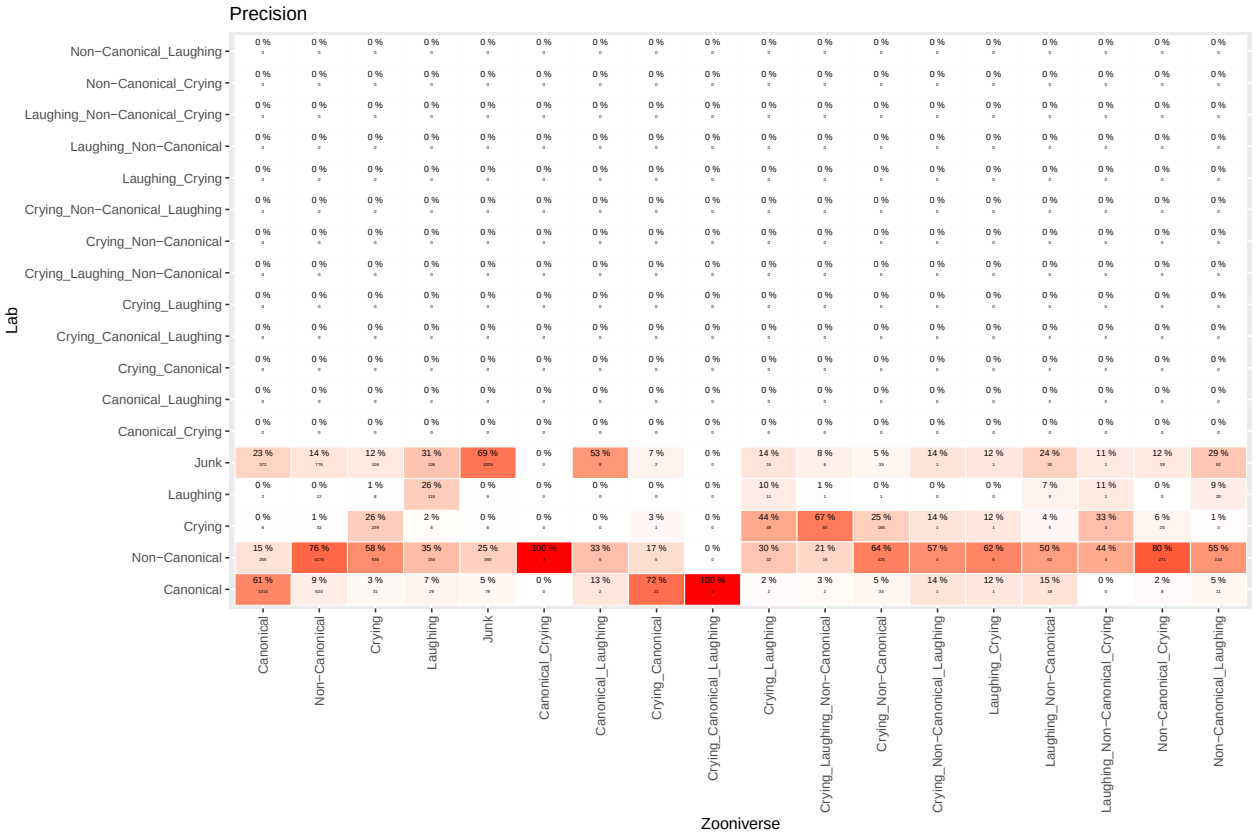
Precision means: If a segment was called X by zooniverse coders, what proportion of the time was it called X by lab coders?

```
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

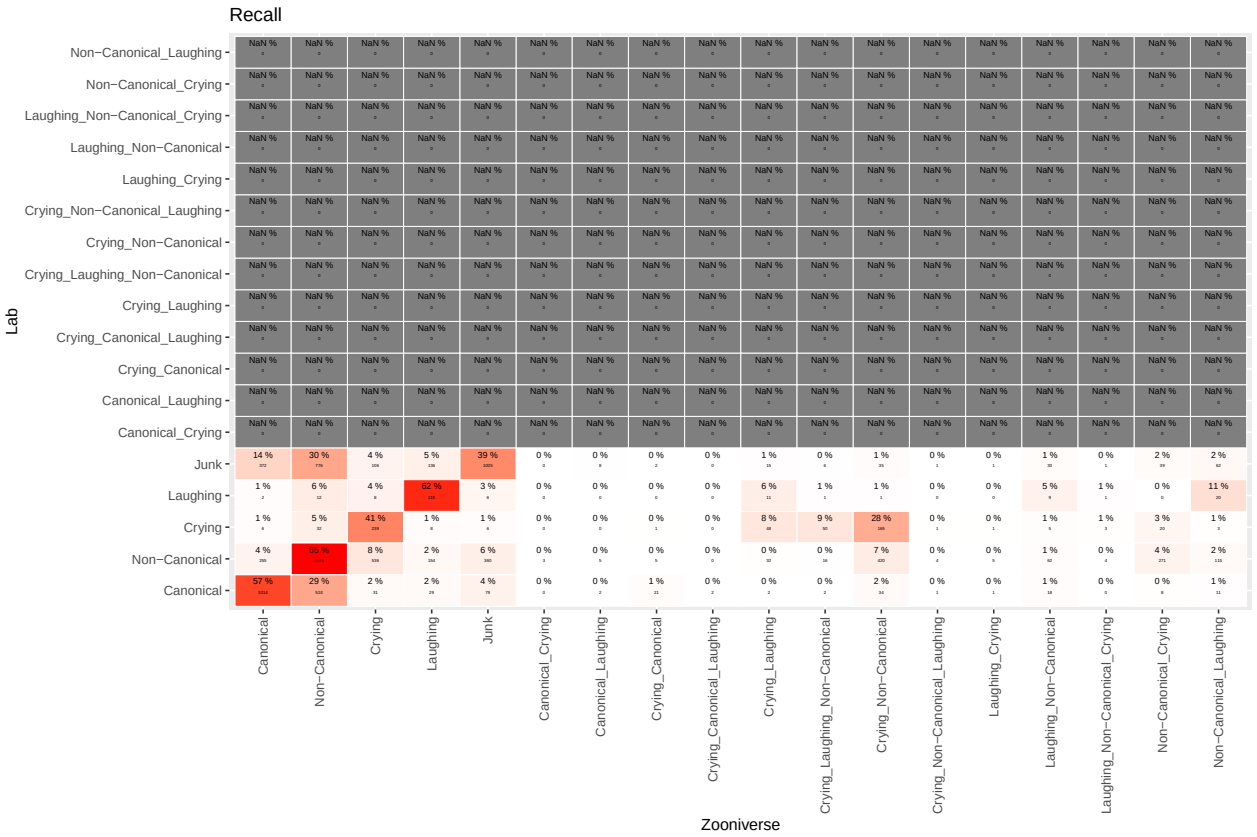
Recall means: If a segment was called X by lab coders, what proportion of the time was it called X by zooniverse coders?

```
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))
```



Collapse across "mixed"

```
#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)
```

Separate confusion matrices for Angelman syndrome children

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

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

```
# 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
```

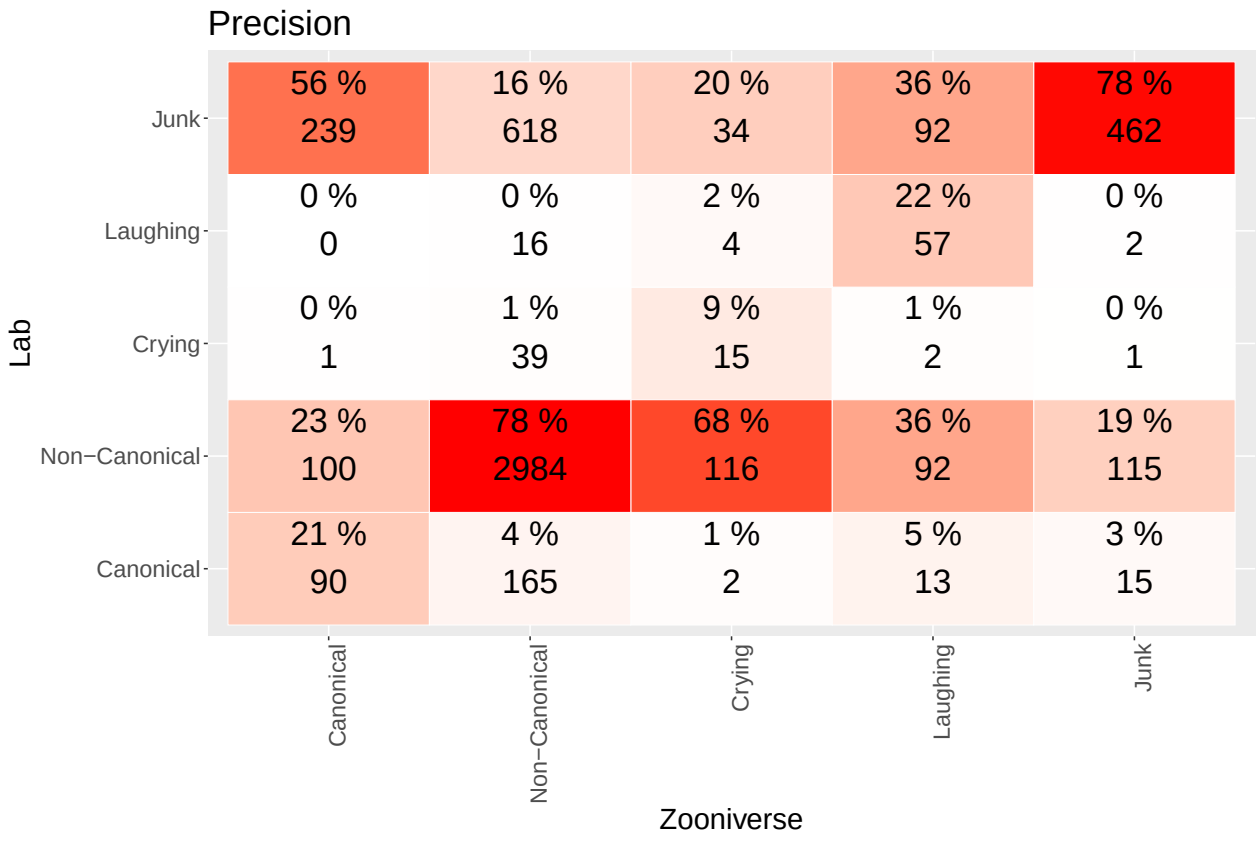
```
## Confusion Matrix and Statistics
##
##                Zooniverse
## Lab             Canonical Non-Canonical Crying Laughing Junk
##   Canonical            90           165      2       13   15
##   Non-Canonical       100          2984    116       92  115
##   Crying                1            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
```

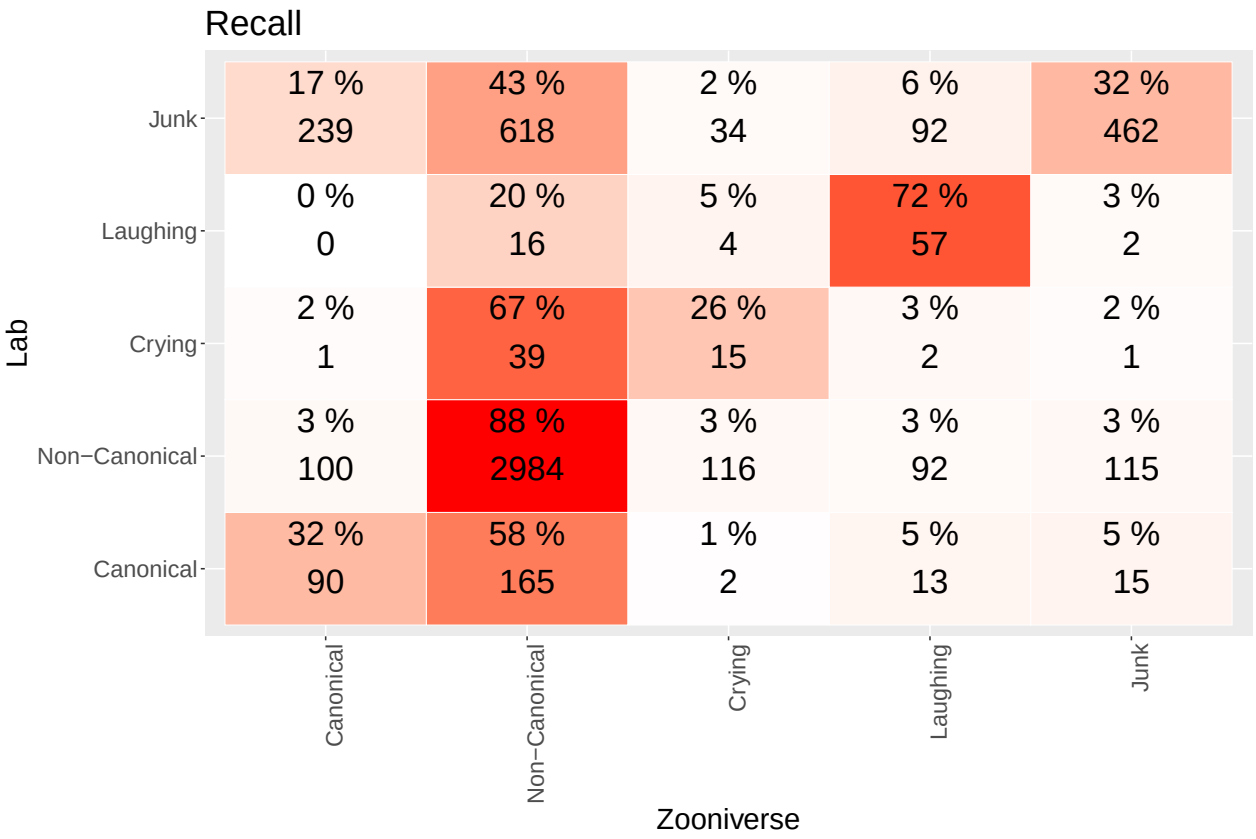
```
colsums=colSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:5) my_conf_tab[,i]=my_conf_tab[,i]/colsums[i]
colSums(my_conf_tab)
```

```
##     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))
```




```
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))
```



Separate confusion matrices with just the low risk controls

```
# 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
```

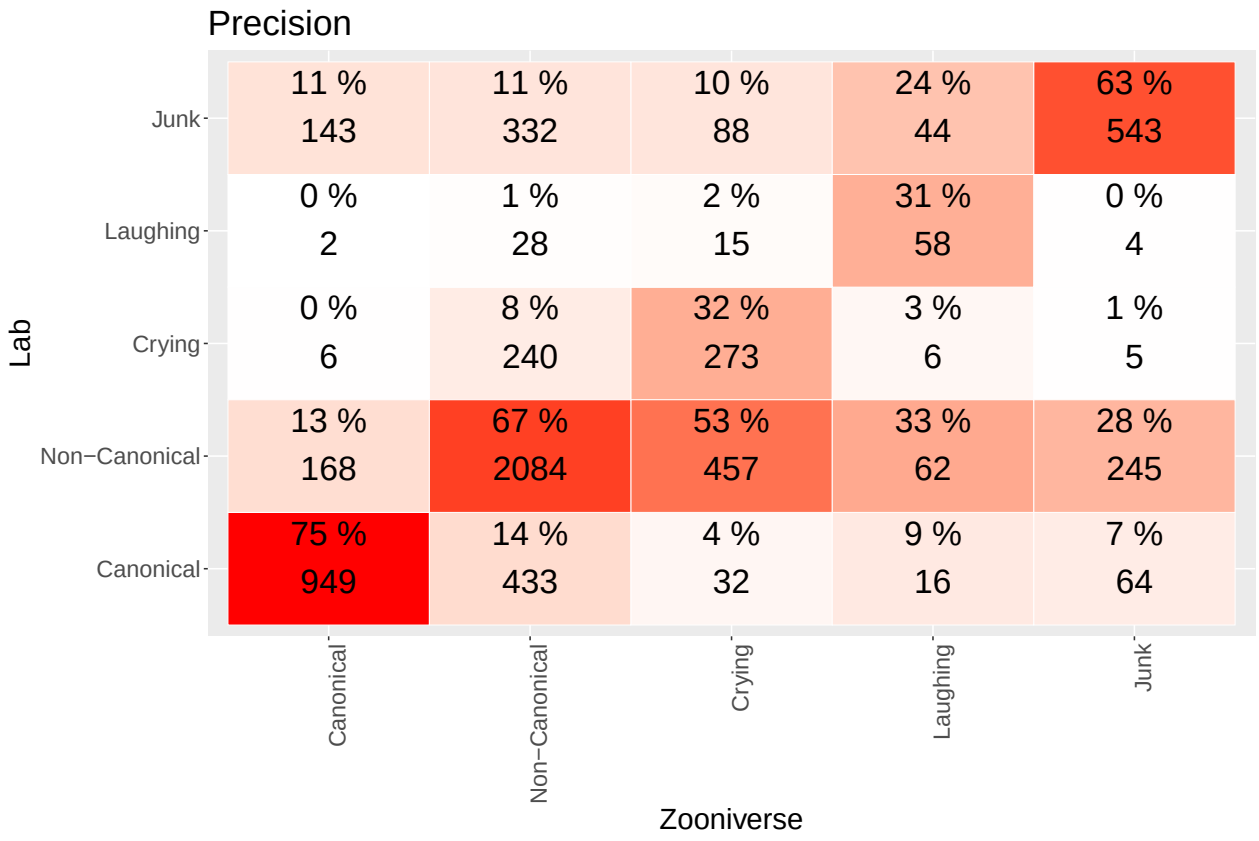
```
## Confusion Matrix and Statistics
##
##                Zooniverse
## Lab             Canonical Non-Canonical Crying Laughing Junk
##   Canonical           949           433     32       16   64
##   Non-Canonical       168          2084    457       62  245
##   Crying                6           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
```

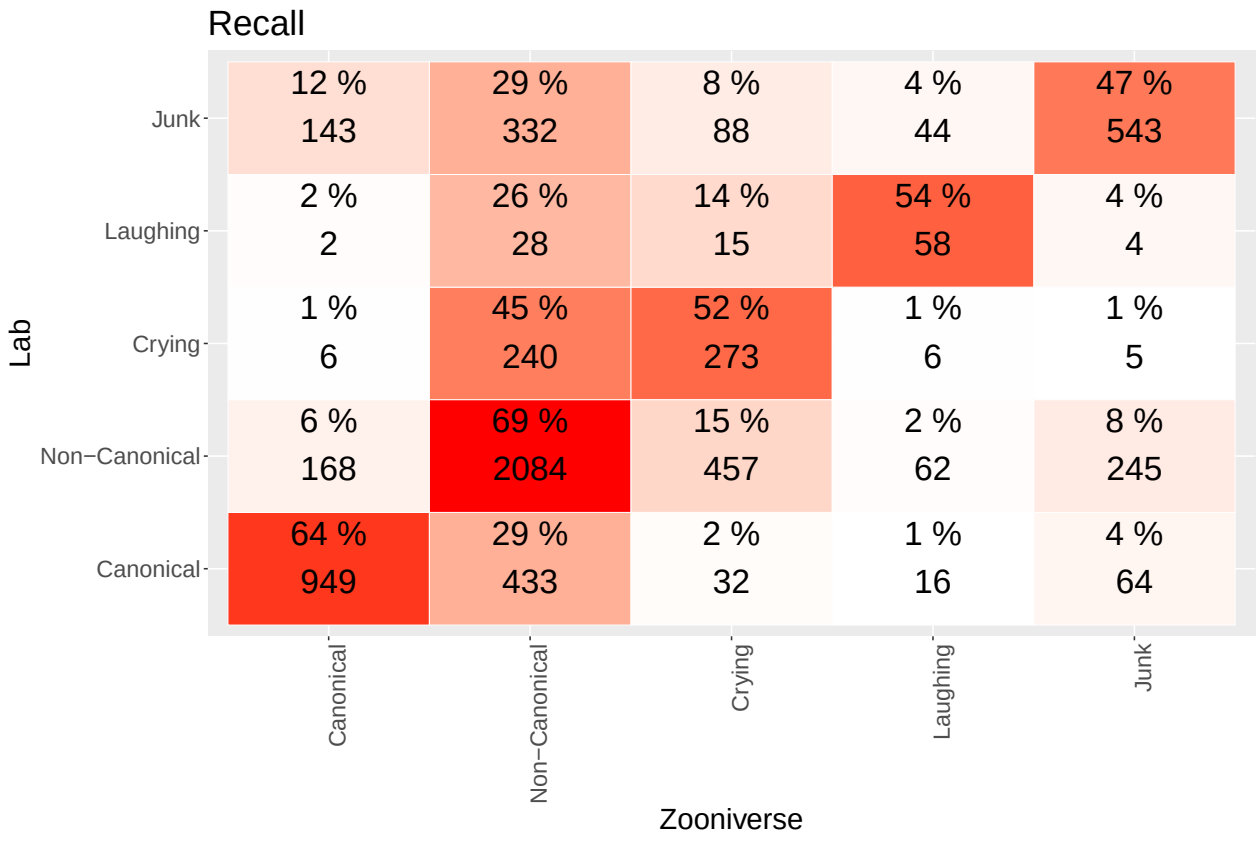
```
colsums=colSums(conf_tab)
my_conf_tab=conf_tab
for(i in 1:5) my_conf_tab[,i]=my_conf_tab[,i]/colsums[i]
colSums(my_conf_tab)
```

```
##     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))
```




```
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))
```


Recall
Junk
Laughing
Crying
Non−Canonical
Canonical
Lab
12 %
143
29 %
332
8 %
88
4 %
44
47 %
543
2 %
2
26 %
28
14 %
15
54 %
58
4 %
4
1 %
6
45 %
240
52 %
273
1 %
6
1 %
5
6 %
168
69 %
2084
15 %
457
2 %
62
8 %
245
64 %
949
29 %
433
2 %
32
1 %
16
4 %
64
Canonical
Non−Canonical
Crying
Laughing
Junk
Zooniverse