

RewardStrucClinicalDataTable.Rmd

Anup Sharma

11/13/2017

Clinical Data Table construction

This script generates a clinical table based on redcap data, clinical QA, and scan availability (T1 and/or restbold)

N=304 (final unique T1 images)

N=126 scans removed as were duplicates (e.g. went with the lowest trait distance or clinical comments)

N=2 unique subjects have been removed for missing clinical diagnosis

N=42 unique subjects have been removed for one of multiple possible reasons (e.g. missing clinical data, scale distances,etc)

```
##### Import REDCAP Data#####
#Import redcap data (all reward protocols)
items_reward<-read.csv("/data/joy/BBL/studies/reward/summaryScores/clinical/subjectData/redcapClinical_Item
Level_1095_050117.csv", header=T,na.strings=c('-9999','NA',''))

#Order by bblid
items_reward.v2<-items_reward[order(items_reward$bblid),]

#Import all available T1 scans (from file system)
T1_all<-read.csv("/data/joy/BBL/studies/reward/summaryScores/clinical/subjectLists/all_T1_Paths_050117.csv"
, header=F)
T1_scans<-strsplit(as.character(T1_all[1:474,]),"/,x")
T1_scans.v2<-matrix(unlist(T1_scans),ncol=14, byrow=TRUE)
T1_scanid<-as.data.frame(T1_scans.v2)[10]
T1_scanid<-T1_scanid[order(T1_scanid$V10),]
T1_scanid<-as.data.frame(T1_scanid)
colnames(T1_scanid)<-"T1"

#Import all available restbold scans (from file system)
RB_all<-read.csv("/data/joy/BBL/studies/reward/summaryScores/clinical/subjectLists/all_RB_Paths_050117.csv"
, header=F)
RB_scans<-strsplit(as.character(RB_all[1:372,]),"/,x")
RB_scans.v2<-matrix(unlist(RB_scans),ncol=14, byrow=TRUE)
RB_scanid<-as.data.frame(RB_scans.v2)[10]
RB_scanid<-RB_scanid[order(RB_scanid$V10),]
RB_scanid<-as.data.frame(RB_scanid)
colnames(RB_scanid)<-"RB"

### Correct N=2 subjects with discrepant scanids
items_reward.v2[which(items_reward.v2$bblid==16336 & items_reward.v2$date_provided==20140827),]$scanid = 90
40
items_reward.v2[which(items_reward.v2$bblid==17491 & items_reward.v2$date_provided==20150908),]$scanid = 98
17

#### Subjects that have T1 Niftis ####
items_reward.v3<-items_reward.v2[which(items_reward.v2$scanid %in% T1_scanid$T1),] #N=474
missing_T1<-items_reward.v2[which(! items_reward.v2$scanid %in% T1_scanid$T1),]
missing_T1.key<-missing_T1[,c("bblid","scanid","date_provided")] #none

##### QA by Clinical Scale/Scale Distance Presence #####
bdi_total_excluded<-items_reward.v3[which(items_reward.v3$bdi_total %in% c(NA)),]
bdi_total_excluded.v2<-items_reward.v3[which(items_reward.v3$bdi_summary_distance %in% c(NA)),]
bdi_total_excluded$comment<-if (nrow (bdi_total_excluded)>0) "Excluded for no BDI_total measures" else charac
ter (0)
bdi_total_excluded.v2$comment<-if (nrow (bdi_total_excluded.v2)>0) "Excluded for no BDI_total distances" else
character (0)
testBdi<-rbind(bdi_total_excluded,bdi_total_excluded.v2)
testBdi.v4<-testBdi[,c(1:4,154:238,293:316)]
#write.csv(testBdi,'/data/joy/BBL/studies/reward/summaryScores/clinical/subjectData/bdiTotalMissing_050117.
csv', row.names=F) #subj missing BDI total

bdiitems_excluded<-items_reward.v3[which(items_reward.v3$bdi1_0==0 & items_reward.v3$bdi1_1==0 & items_rewa
rd.v3$bdi1_2==0 & items_reward.v3$bdi1_3==0 & items_reward.v3$bdi_old1 %in% c(NA,-9999)),]
```

```

bdiitems_excluded.v2<-items_reward.v3[which(items_reward.v3$bdi_distance %in% c(NA) & items_reward.v3$bdiol
d_distance %in% c(NA)),]
bdiitems_excluded$comment<-if (nrow (bdiitems_excluded)>0) "Excluded for no BDI_items measures" else charac
ter (0)
bdiitems_excluded.v2$comment<-if (nrow (bdiitems_excluded.v2)>0) "Excluded for no BDI_items distances" else
character (0)

rpas_scale_excluded<-items_reward.v3[which(items_reward.v3$rpasshortcomb_total %in% c(NA)),]
rpas_scale_excluded$comment<-if (nrow (rpas_scale_excluded) >0) "Excluded for no RPAS measures" else charac
ter (0)

rsas_scale_excluded<-items_reward.v3[which(items_reward.v3$rsasshortcomb_total %in% c(NA)),]
rsas_scale_excluded$comment<-if (nrow (rsas_scale_excluded) >0) "Excluded for no rsas measures" else charac
ter (0)

bisbas_scale_excluded<-items_reward.v3[which(items_reward.v3$bisbas_1 %in% c(NA)),]
bisbas_scale_excluded$comment<-if (nrow (bisbas_scale_excluded) >0) "Excluded for no bisbas measures" else
character (0)

cains_scale_excluded<-items_reward.v3[which(items_reward.v3$fr_anh_amot_all %in% c(NA)),]
cains_scale_excluded.v2<-items_reward.v3[which(items_reward.v3$newfranksummary_distance %in% c(NA)),]
cains_scale_excluded$comment<-if (nrow (cains_scale_excluded) >0) "Excluded for no cains measures" else cha
racter (0)
cains_scale_excluded.v2$comment<-if (nrow (cains_scale_excluded.v2) >0) "Excluded for no cains distances" e
lse character (0)

clinical_excluded_missing_ALL<-rbind(bdiitems_excluded,bdiitems_excluded.v2, rpas_scale_excluded, rsas_scale_
excluded, bisbas_scale_excluded, cains_scale_excluded, cains_scale_excluded.v2)
clinical_excluded_missing_unique<-clinical_excluded_missing_ALL[!duplicated(clinical_excluded_missing_ALL$s
canid),]

items_reward.clinical_present_unique<-items_reward.v3[which(! items_reward.v3$scanid %in% clinical_excluded
_missing_unique$scanid),] #Subjects with key clinical scales present

##### QA by State Mean distance #####
items_reward.clinical_present_unique$abs_bdisummary_distance<-abs(items_reward.clinical_present_unique$bdis
ummary_distance)
items_reward.clinical_present_unique$abs_newfranksummary_distance<-abs(items_reward.clinical_present_unique
$newfranksummary_distance)
items_reward.clinical_present_unique$state_distance<-rowMeans(items_reward.clinical_present_unique[,c("abs_
bdisummary_distance", "abs_newfranksummary_distance")])
state_excluded<-items_reward.clinical_present_unique[abs(items_reward.clinical_present_unique$state_distanc
e)>14,]
state_excluded$comment<-if (nrow (state_excluded)>0) "Excluded for State Mean distance from T1 > 14 days" e
lse character (0)
state_excluded_unique<-state_excluded[!duplicated(state_excluded$scanid),] #subjects excluded due to high
mean state distance

StateSubjRem<-items_reward.clinical_present_unique[which(items_reward.clinical_present_unique$age_at_date_p
rovided %in% state_excluded_unique$age_at_date_provided),] #16 subjects removed

items_reward.state_QA<-items_reward.clinical_present_unique[which(! items_reward.clinical_present_unique$ag
e_at_date_provided %in% state_excluded_unique$age_at_date_provided),]

##### QA by Lowest Trait Mean distance #####
items_reward.st.unique.v1<-data.frame()
items_reward.st.unique.v2<-data.frame()
items_reward.state_QA$abs_bisbassummary_distance<-abs(items_reward.state_QA$bisbassummary_distance)
items_reward.state_QA$abs_rpassshortcombsummary_distance<-abs(items_reward.state_QA$rpassshortcombsummary_dis
tance)
items_reward.state_QA$abs_rsasshortcombsummary_distance<-abs(items_reward.state_QA$rsasshortcombsummary_dis
tance)
items_reward.state_QA$trait_distance<-rowMeans(items_reward.state_QA[,c("abs_bisbassummary_distance", "abs_r
passshortcombsummary_distance", "abs_rsasshortcombsummary_distance")])
bblidList<-unique(items_reward.state_QA$bblid)
for (i in bblidList){
  tmp<-items_reward.state_QA[items_reward.state_QA$bblid==i,]
  if (nrow(tmp) > 1) {
    tmp2<-tmp[which(abs(tmp$trait_distance) == min(abs(tmp$trait_distance))),] #select minimum mean trait d
istance among duplicate subjects
    items_reward.st.unique.v1<-rbind(items_reward.st.unique.v1,tmp2)
  } else {
    items_reward.st.unique.v1<-rbind(items_reward.st.unique.v1,tmp)
  }
}

```

```

    items_reward.st.unique.v2<-rbind(items_reward.st.unique.v2,tmp)
  }
}
items_reward.st.unique.all<-rbind(items_reward.st.unique.v1,items_reward.st.unique.v2)    #n=312 scans (n=12
6 removed as are duplicated)
items_reward.st.unique.ord<-items_reward.st.unique.all[order(items_reward.st.unique.all$bblid),]    #at this
point, still have a few (N=5) duplicate subjects

##### QA by Trait Mean distance #####
trait_excluded<-items_reward.st.unique.ord[abs(items_reward.st.unique.ord$trait_distance)>365,]
trait_excluded$comment<-if (nrow (trait_excluded)>0) "Excluded for Trait Mean distance from T1 > 365 days"
else character (0)
trait_excluded_unique<-trait_excluded[!duplicated(trait_excluded$scanid),]
items_reward.st.unique.ord.v2<-items_reward.st.unique.ord[which(! items_reward.st.unique.ord$scanid %in% tr
ait_excluded$scanid),]

#### QA based on resting state motion (choose minimum among duplicates)
items_reward.st.unique.ord.v2 = within (items_reward.st.unique.ord.v2, {
  motion<- ifelse(! is.na(restbold204_meanrelrms), restbold204_meanrelrms, 0)
  motion<- ifelse(! is.na(boldmgh_meanrelrms), boldmgh_meanrelrms, motion )
  motion<- ifelse(! is.na(restbold124_meanrelrms), restbold124_meanrelrms, motion)
})

items_reward.st.final.v1<-data.frame()
items_reward.st.final.v2<-data.frame()
bblidList<-unique(items_reward.st.unique.ord.v2$bblid)
for (i in bblidList){
  tmp3<-items_reward.st.unique.ord.v2[items_reward.st.unique.ord.v2$bblid==i,]
  if (nrow(tmp3) > 1) {
    tmp4<-tmp3[which(abs(tmp3$motion) == min(abs(tmp3$motion))),] #including min motion duplicate subjects
    items_reward.st.final.v1<-rbind(items_reward.st.final.v1,tmp4)
  } else {
    items_reward.st.final.v2<-rbind(items_reward.st.final.v2,tmp3)
  }
}

#### Remove remaining duplicate subjects based on other clinical comments
dup<-items_reward.st.final.v1[,c("bblid","scanid","date_provided", "commentsbisbas","commentsbdi","comments
_socanh","comments_franken","state_distance")]
dup.v2<-dup[which(! dup$date_provided %in% c("20131002","20130920","20131108","20131008","20131031")),]
dup.exclude<-dup[which(dup$date_provided %in% c("20131002","20130920","20131108","20131008","20131031")),]
dup.exclude$comment<-if (nrow (dup.exclude)>0) "Excluded these duplicate subjects based on clinical comment
s" else character (0)

items_reward.st.final.v1<-items_reward.st.final.v1[which(! dup$date_provided %in% c("20131002","20130920","
20131108","20131008","20131031")),]

items_reward.st.final<-rbind(items_reward.st.final.v1,items_reward.st.final.v2)
items_reward.st.final.ord<-items_reward.st.final[order(items_reward.st.final$bblid),]

### Add updated diagnosis information
items_reward.st.final.ord = within (items_reward.st.final.ord, {
  updated_study_gp<- ifelse(primarydiagnosis=="Schizophrenia", "SCZ", 0)
  updated_study_gp<- ifelse(primarydiagnosis=="Schizoaffective", "SCZ", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="schizoaffectiveDepressedType", "SCZ", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="SchizoaffectiveDepressedType", "SCZ", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="schizoaffectiveBipolarType", "SCZ", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="Schizophreniform", "SCZ", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="psychoticDisorderNOS", "SCZ", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="clinicalRisk", "RSK", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="majorDepressiveDisorder", "MDD", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="depressiveDisorderNOS", "MDD", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="bipolarDisorderTypeI", "BPD", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="bipolarDisorderTypeII", "BPD", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="bipolarDisorderNOS", "BPD", updated_study_gp)
  updated_study_gp<- ifelse(primarydiagnosis=="bipoloarDisorderNOS", "BPD", updated_study_gp)
  updated_study_gp<- ifelse(studygroup=="FM" & primarydiagnosis=="noDiagnosis", "RSK", updated_study_gp)
  updated_study_gp<- ifelse(studygroup=="NC" & primarydiagnosis=="noDiagnosis", "NC", updated_study_gp)
  updated_study_gp<- ifelse(studygroup=="Pilot_NC_R01" & primarydiagnosis=="noDiagnosis", "NC", updated_stu
dy_gp)
})

```

```

#Remove 2 subjects due to unclear dx
unclearDiag<-items_reward.st.final.ord[which(items_reward.st.final.ord$updated_study_gp %in% c("0")),]
items_reward.st.final.ord<-items_reward.st.final.ord[which(! items_reward.st.final.ord$updated_study_gp %in% c("0")),]

#Group as factor
items_reward.st.final.ord$updated_study_gp<-factor(x=items_reward.st.final.ord$updated_study_gp, levels=c("NC", "BPD", "MDD", "SCZ", "RSK"))

#Final excluded
c<-clinical_excluded_missing_ALL[,c("bblid", "scanid", "date_provided", "comment")]
s<-state_excluded[,c("bblid", "scanid", "date_provided", "comment")]
t<-trait_excluded[,c("bblid", "scanid", "date_provided", "comment")]
d<-dup.exclude[,c("bblid", "scanid", "date_provided", "comment")]
all_excluded<-rbind(c,s,t,d)
all_excluded_unique<-all_excluded[!duplicated(all_excluded$scanid),]

#write.csv(all_excluded_unique, '/data/joy/BBL/studies/reward/summaryScores/clinical/subjectData/all_excluded_unique_050117.csv', row.names=F)

#Final T1 data
#write.csv(items_reward.st.final.ord, '/data/joy/BBL/studies/reward/summaryScores/clinical/subjectData/n304_T1_integratedData_050117.csv', row.names=F)

#With restbold
items_reward.rb.final<-items_reward.st.final.ord[which(items_reward.st.unique.ord$scanid %in% RB_scanid$RB),]
items_reward.rb.final.ord<-items_reward.rb.final[order(items_reward.rb.final$bblid),]

#write.csv(items_reward.rb.final.ord, '/data/joy/BBL/studies/reward/summaryScores/clinical/subjectData/n255_RB_integratedData_050117.csv', row.names=F)

```

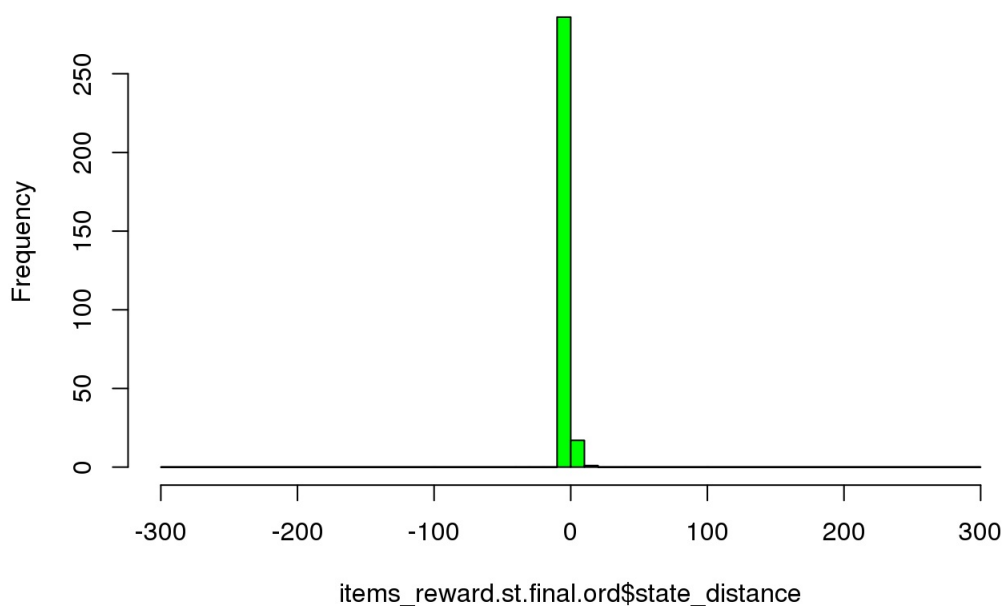
Embeded Plots

```

## Warning in plot.histogram(r, freq = freq1, col = col, border = border,
## angle = angle, : the AREAS in the plot are wrong -- rather use 'freq =
## FALSE'

```

Histogram of items_reward.st.final.ord\$state_distance

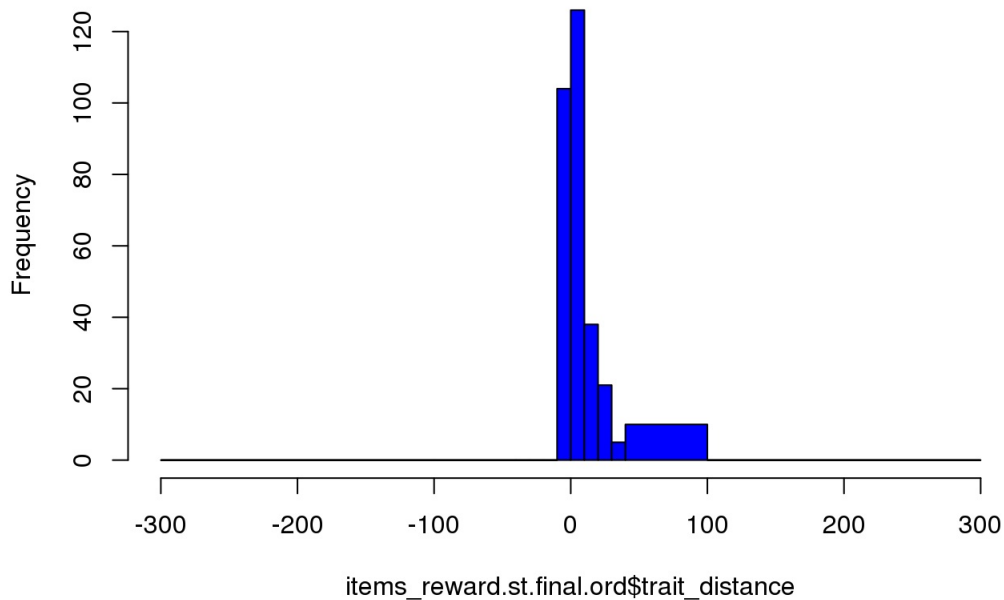


```

## Warning in plot.histogram(r, freq = freq1, col = col, border = border,
## angle = angle, : the AREAS in the plot are wrong -- rather use 'freq =
## FALSE'

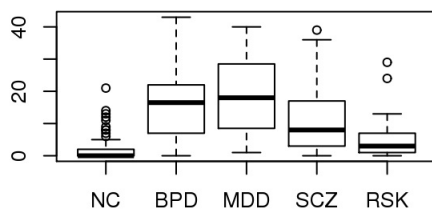
```

Histogram of items_reward.st.final.ord\$trait_distance

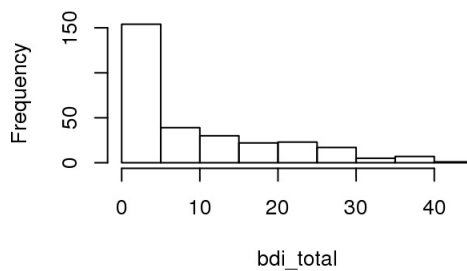


```
## The following object is masked from package:base:
##
## version
```

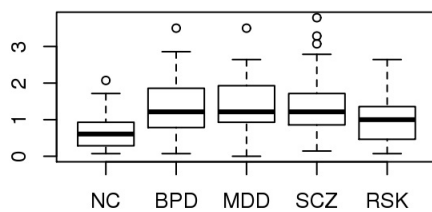
bdi_total x diagnosis



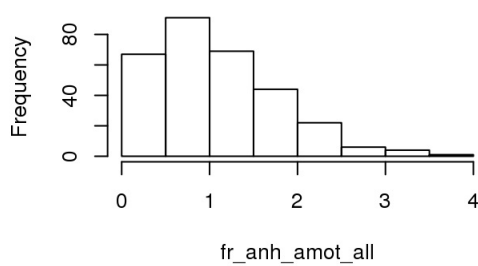
histogram of bdi_total



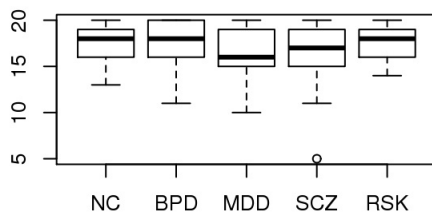
fr_anh_amot_all x diagnosis



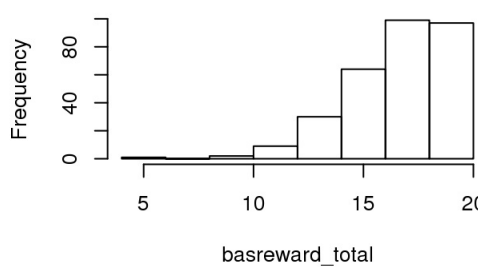
histogram of fr_anh_amot_all



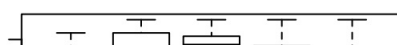
basreward_total x diagnosis



histogram of basreward_total

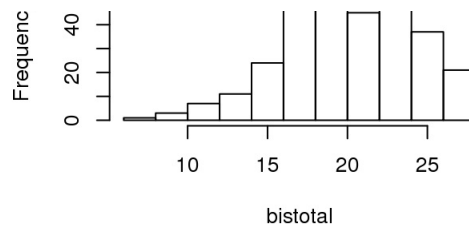
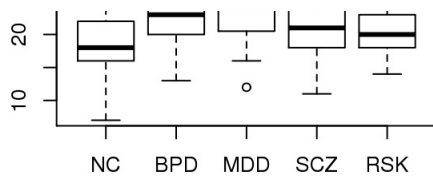


bistotal x diagnosis

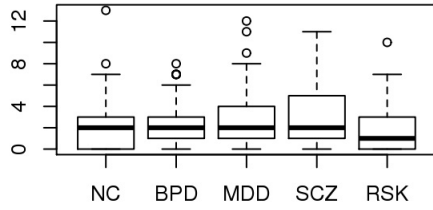


histogram of bistotal

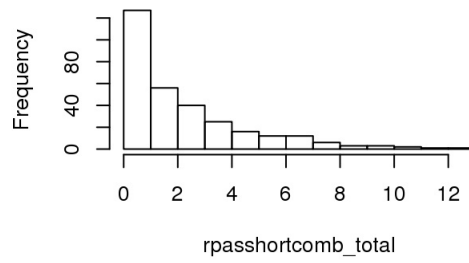




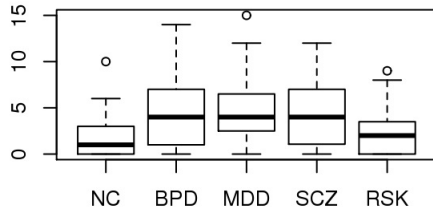
rpassshortcomb_total x diagnosis



histogram of rpassshortcomb_total



rsassshortcomb_total x diagnosis



histogram of rsassshortcomb_total

