Preliminary summary of results, Maetzig et al ICCM2017

Shravan Vasishth

2/20/2017

I (SV) tried averaging so that we have only one set of parameters per subject:

```r
if(1){
## averaging values for each subject:
control_subj<-unique(results_SR_controls$subj)
results_SR_c_mn <- data.frame()
for(i in control_subj){
    tmp<-subset(results_SR_controls,subj==i)
    temp<-data.frame(GA=mean(tmp$GA),DAT=mean(tmp$DAT),ANS=mean(tmp$ANS),subj=i)
    results_SR_c_mn<-rbind(results_SR_c_mn,
                    temp,make.row.names=FALSE)
}

control_subj<-unique(results_OR_controls$subj)
results_OR_c_mn <- data.frame()
for(i in control_subj){
    tmp<-subset(results_OR_controls,subj==i)
    temp<-data.frame(GA=mean(tmp$GA),DAT=mean(tmp$DAT),ANS=mean(tmp$ANS),subj=i)
    results_OR_c_mn<-rbind(results_OR_c_mn,
                    temp,make.row.names=FALSE)
}

iwa_subj<-unique(results_SR_iwa$subj)
results_SR_i_mn <- data.frame()
for(i in iwa_subj){
    tmp<-subset(results_SR_iwa,subj==i)
    temp<-data.frame(GA=mean(tmp$GA),DAT=mean(tmp$DAT),ANS=mean(tmp$ANS),subj=i)
    results_SR_i_mn<-rbind(results_SR_i_mn,
                    temp,make.row.names=FALSE)
}

iwa_subj<-unique(results_OR_iwa$subj)
results_OR_i_mn <- data.frame()
for(i in iwa_subj){
    tmp<-subset(results_OR_iwa,subj==i)
    temp<-data.frame(GA=mean(tmp$GA),DAT=mean(tmp$DAT),ANS=mean(tmp$ANS),subj=i)
    results_OR_i_mn<-rbind(results_OR_i_mn,
                    temp,make.row.names=FALSE)
}

results_SR_controls<-results_SR_c_mn
results_OR_controls<-results_OR_c_mn
results_SR_iwa<-results_SR_i_mn
results_OR_iwa<-results_OR_i_mn
}
```

```
# coding subjects as factors
results_SR_controls$subj <- as.integer(factor(results_SR_controls$subj))
results_OR_controls$subj <- as.integer(factor(results_OR_controls$subj))

results_SR_iwa$subj <- as.integer(factor(results_SR_iwa$subj))
results_OR_iwa$subj <- as.integer(factor(results_OR_iwa$subj))

## Trying out how plot with subj number (as integers) looks like
## However, not very readable and too much visual information, and clustering
## information is lost
#tmpSRcontrol <- cbind(jitter(results_SR_controls[,1], factor=10),
#                      jitter(results_SR_controls[,2], factor=4),
#                      jitter(results_SR_controls[,3], factor=4) )
#s3d <- scatterplot3d(tmpSRcontrol[, 1:3], pch = "")
#text(s3d$xyz.convert(tmpSRcontrol[, 1:3]), labels = results_SR_controls$subj,
#     col = "steelblue",cex=1.5)}
```

We can do exploratory analyses.

Impressions from plots

- In SR controls, noise seems to be generally low, but in SR iwa noise is higher.
- In SR controls, DAT is generally low, with a few trials have high DAT, but in SR iwa there is higher proportion of high DAT.
- In SR controls, GA tends to be high in controls, bit in SR iwa there is a higher proportion of low GA
- In OR controls, noise is in general low, but in SR iwa noise tends to be higher.
- In OR controls, GA is generally high, in OR iwa there is a large proportion of low GA.
- In OR controls, DAT is generally fast, with a few slow cases, but in OR iwa there is a higher proportion of slow cases.

Correlations:

```
cor(results_SR_controls[,1:3])
```

```
##             GA        DAT        ANS
## GA   1.0000000 -0.6846780 -0.3806761
## DAT -0.6846780  1.0000000  0.4672837
## ANS -0.3806761  0.4672837  1.0000000
```

```
cor(results_SR_iwa[,1:3])
```

```
##             GA        DAT        ANS
## GA   1.0000000 -0.6116764 -0.5962201
## DAT -0.6116764  1.0000000  0.4730951
## ANS -0.5962201  0.4730951  1.0000000
```
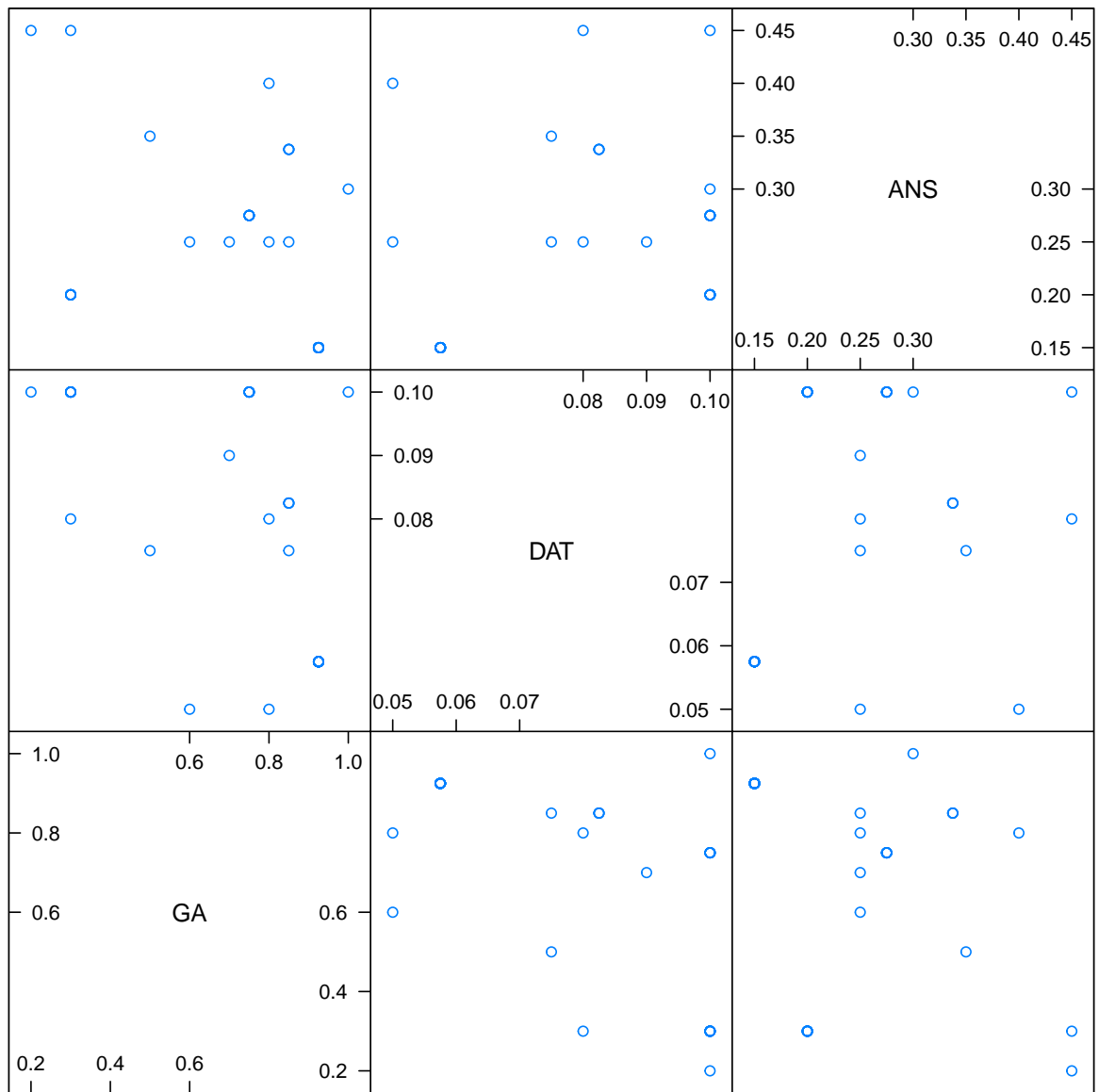
```
cor(results_OR_controls[,1:3])
```

```
##              GA        DAT        ANS
## GA   1.00000000 -0.6631551 0.04525844
## DAT -0.66315513  1.0000000 0.16177672
## ANS  0.04525844  0.1617767 1.00000000
```

```
cor(results_OR_iwa[,1:3])
```
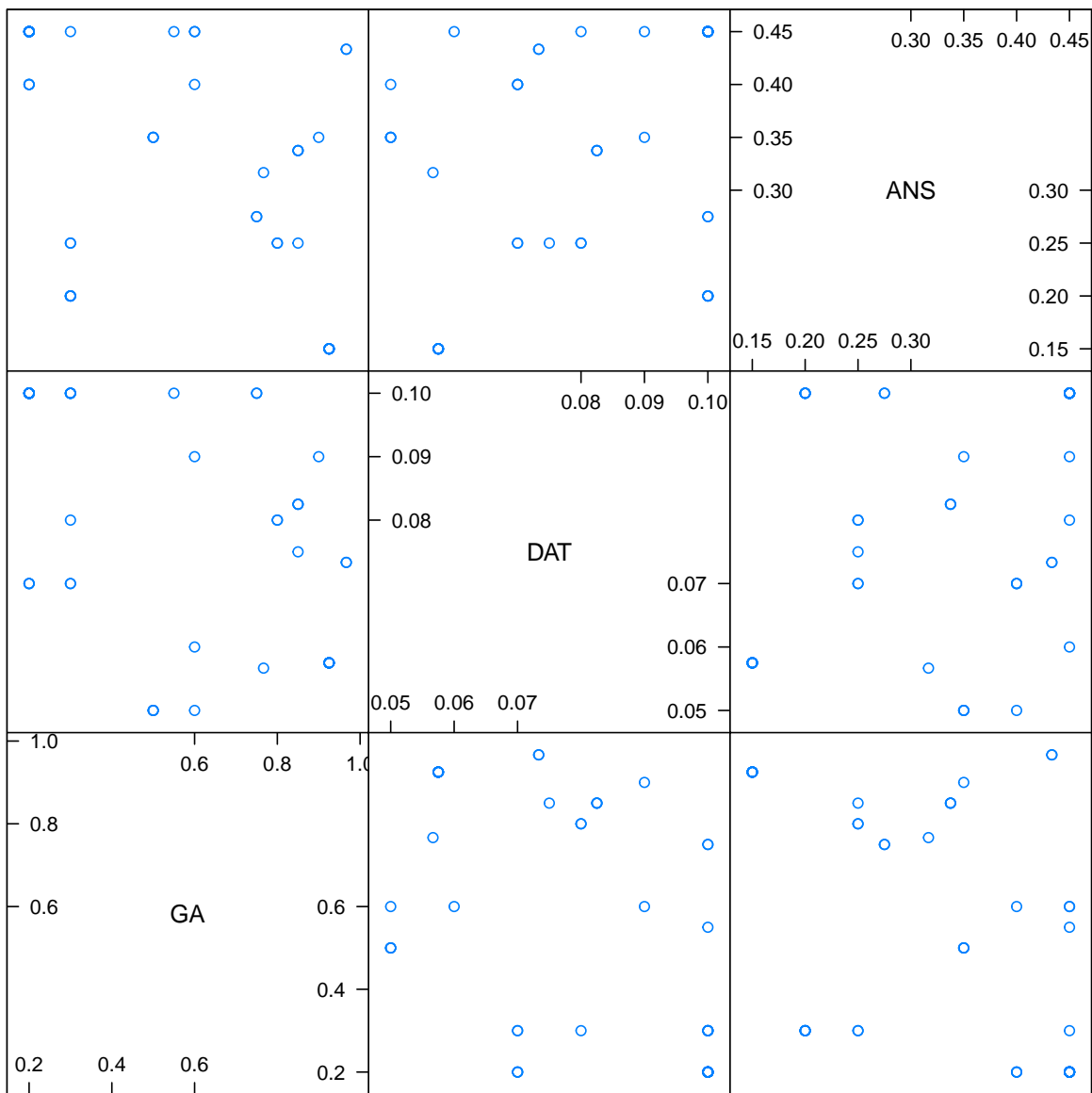
```
##              GA        DAT        ANS
## GA    1.0000000 -0.7612306 -0.7321958
## DAT  -0.7612306  1.0000000  0.6439172
## ANS  -0.7321958  0.6439172  1.0000000
```

```
library(lattice)
splom(results_SR_controls[,1:3])
```
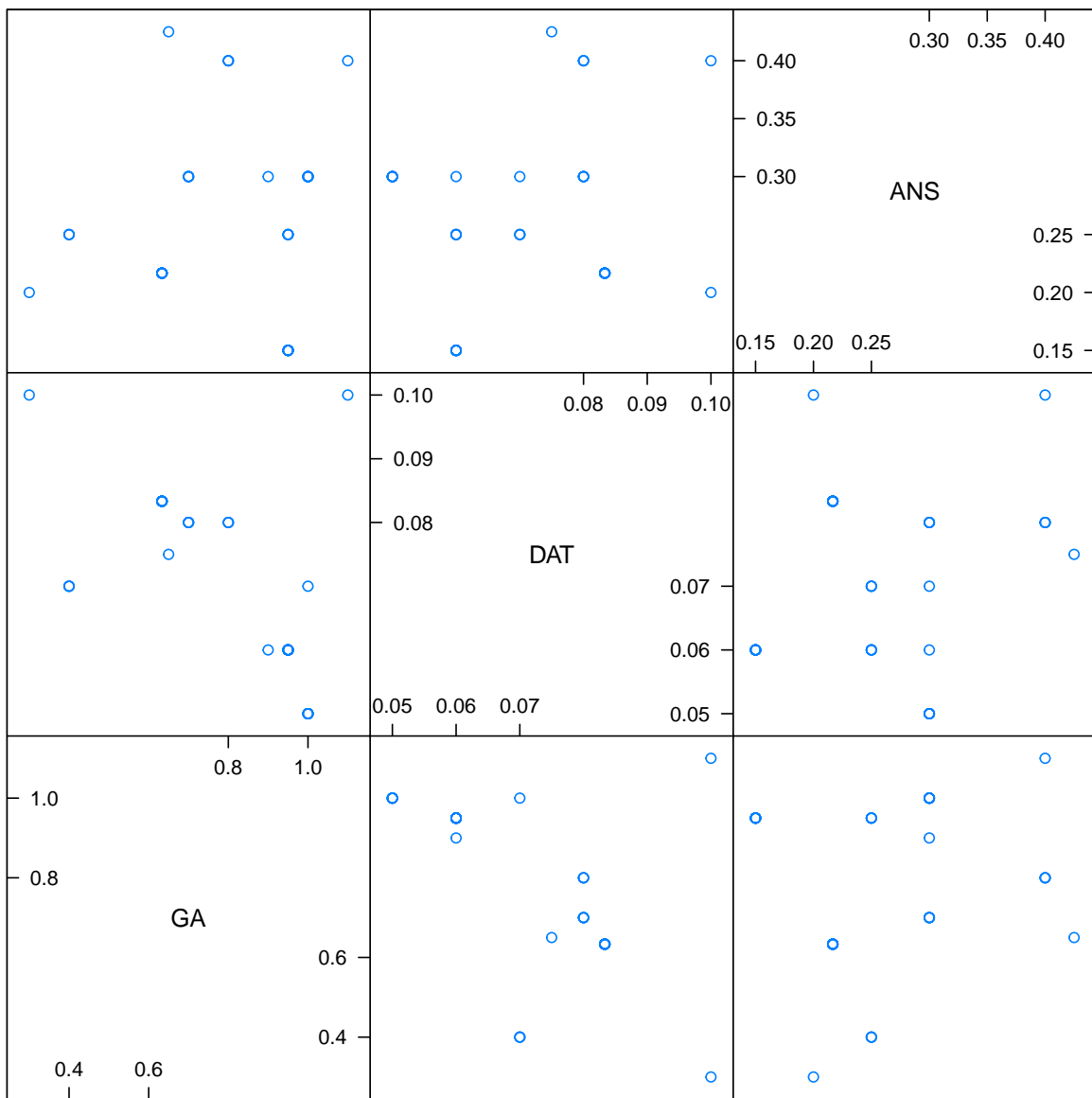

Scatter Plot Matrix
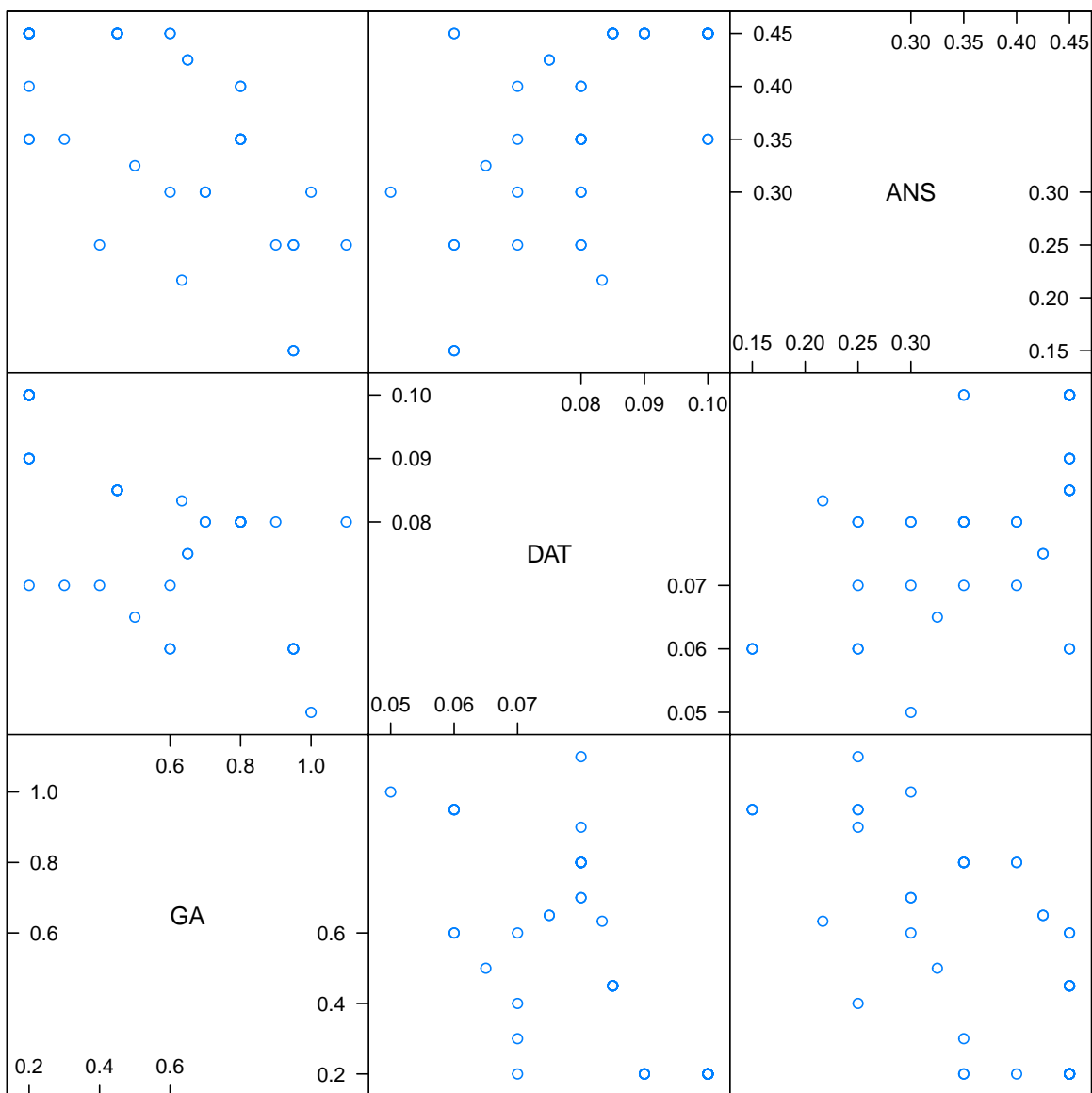
```
splom(results_SR_iwa[,1:3])
```

Scatter Plot Matrix

```
splom(results_OR_controls[,1:3])
```
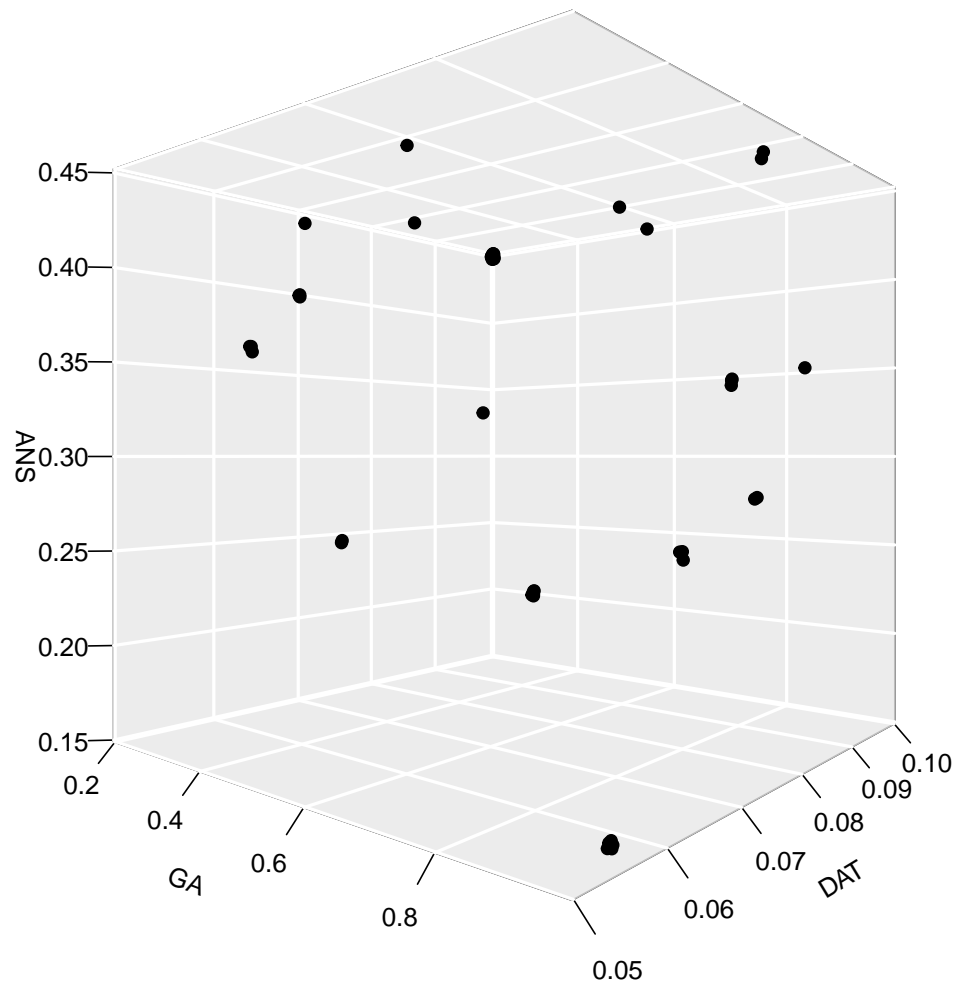
Scatter Plot Matrix

```
splom(results_OR_iwa[,1:3])
```
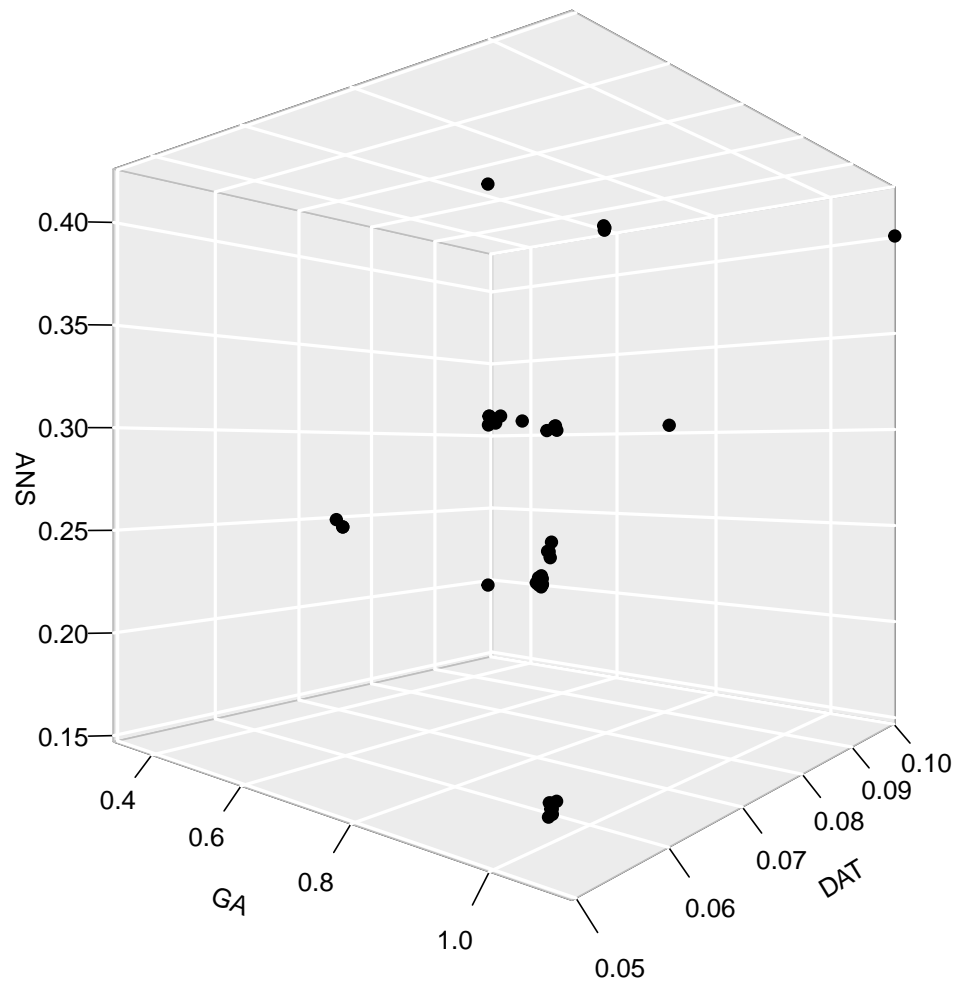
Scatter Plot Matrix

IWA, SR
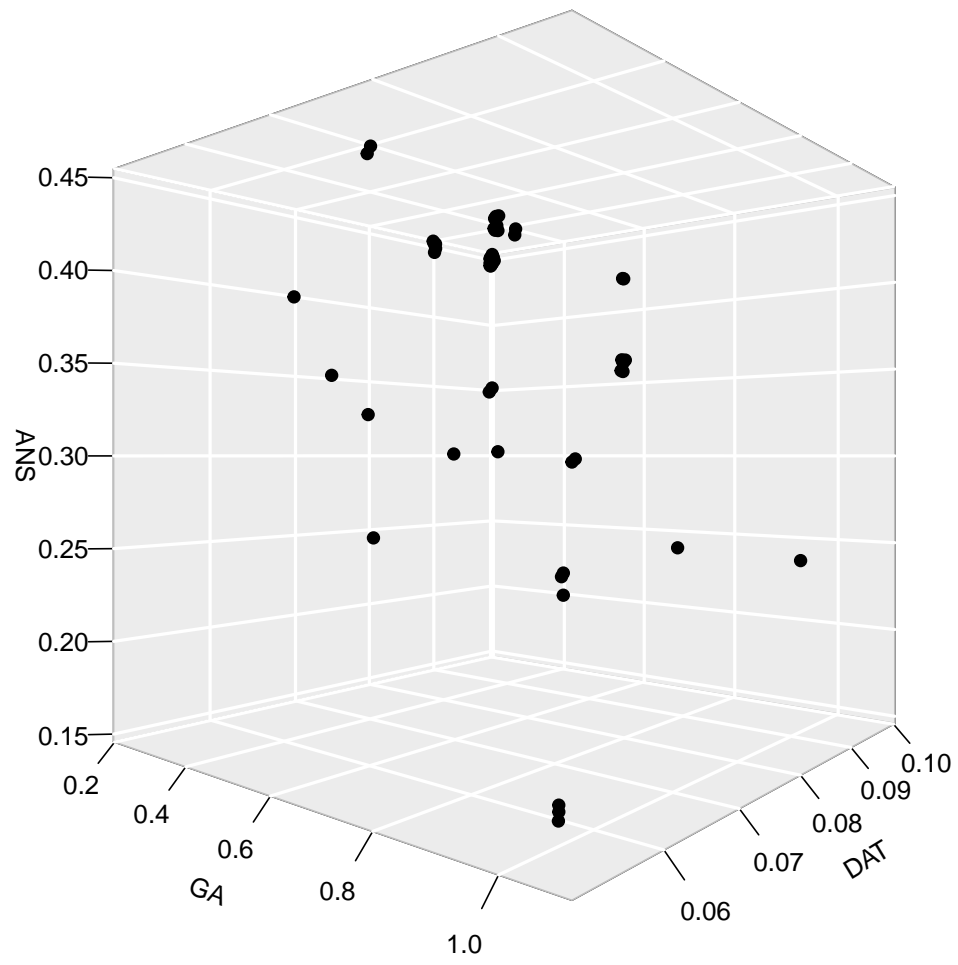
Controls, SR

IWA, OR

Subject relatives   **Subject relatives**   **Subject relatives**
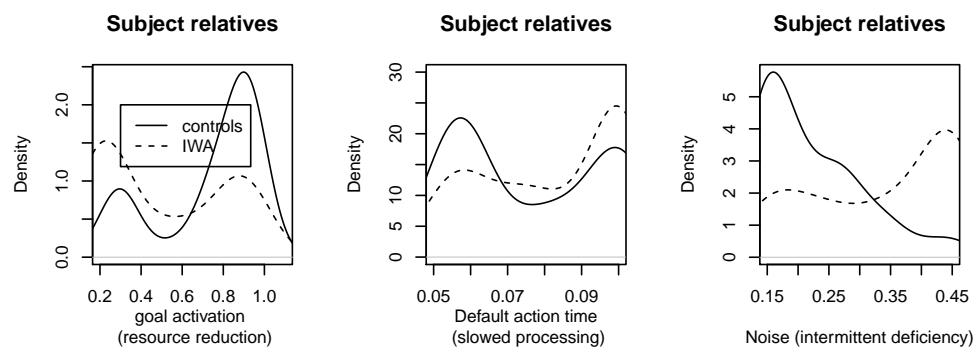
```
op<-par(
    mar=c(2,4.5,4,2),
    mfrow=c(1,3),pty="s",
    oma = c(2,10,2,2)
    )

#plot(density(results_SR_iwa$GA),xlab="GA",main="IWA, SR")
#plot(density(results_SR_iwa$DAT),xlab="DAT",main="IWA, SR")
#plot(density(results_SR_iwa$ANS),xlab="ANS",main="IWA, SR")

#op<-par(mfrow=c(2,3),pty="s")
plot(density(results_OR_controls$GA),xlab="goal activation \n (resource reduction)",main="Object relativ
lines(density(results_OR_iwa$GA),xlab="GA",lty=2)
legend(x=.3,y=3,legend=c("controls","patients"),
```

```
        lty=1:2)

plot(density(results_OR_controls$DAT),xlab="Default action time \n (slowed processing)",main="Object rel
lines(density(results_OR_iwa$DAT),xlab="DAT",lty=2)

plot(density(results_OR_controls$ANS),xlab="Noise (intermittent deficiency)",main="Object relatives",xl
lines(density(results_OR_iwa$ANS),xlab="ANS",lty=2)
```
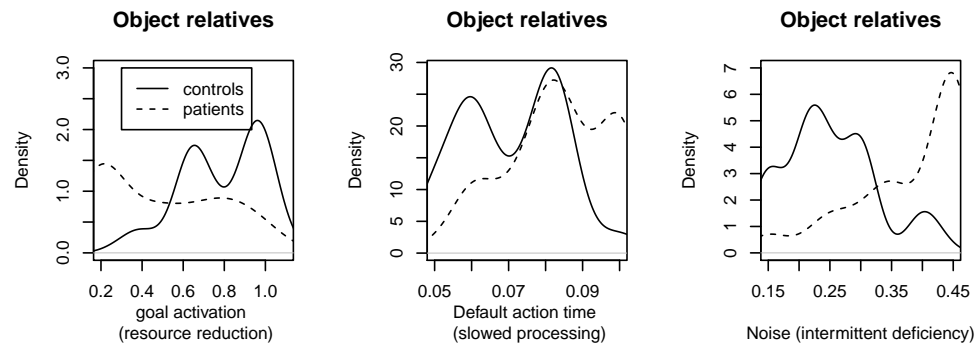


```
#plot(density(results_OR_iwa$GA),xlab="GA",main="IWA, OR")
#plot(density(results_OR_iwa$DAT),xlab="DAT",main="IWA, OR")
#plot(density(results_OR_iwa$ANS),xlab="ANS",main="IWA, OR")
```

Exploring the mixing distributions for GA

The mixture estimates don't really tell us much. In ORs I would have expected a higher proportion of low GA in iwa. But the mixture model is clearly failing to identify the mixing distributions.

```
# mixtools
library(mixtools)
```

```
## mixtools package, version 1.1.0, Released 2017-03-10
## This package is based upon work supported by the National Science Foundation under Grant No. SES-0518
```

```
## In SRs, higher proportion of low GA in iwa than controls:
mixmodSRcGA <- normalmixEM(results_SR_controls$GA,
                           lambda = c(1/2), ## inits
                           mu = c(.3,.9), ## inits
                           sigma = 1)
```

```
## number of iterations= 489
```

```
summary(mixmodSRcGA)
```

```
## summary of normalmixEM object:
##           comp 1    comp 2
## lambda 0.2417378 0.7582622
## mu     0.3122557 0.8666193
## sigma  0.0843551 0.0843551
## loglik at estimate:  23.42587
```

```
mixmodSRiGA <- normalmixEM(results_SR_iwa$GA,
                           lambda = c(1/2), ## inits
                           mu = c(.2,.9), ## inits
                           sigma = 1)
```

```
## number of iterations= 328
```

```
summary(mixmodSRiGA)
```

```
## summary of normalmixEM object:
##          comp 1   comp 2
## lambda 0.565813 0.434187
## mu     0.262029 0.837874
## sigma  0.109359 0.109359
## loglik at estimate:  8.5583
```

```
## In ORs iwa have lower goal activations and more variance than controls:
mixmodORcGA <- normalmixEM(results_OR_controls$GA,
                           lambda = c(1/2), ## inits
                           mu = c(.04,.08), ## inits
                           sigma = 1)
```

```
## number of iterations= 3
```

```
summary(mixmodORcGA)
```

```
## summary of normalmixEM object:
##          comp 1   comp 2
## lambda 0.492709 0.507291
## mu     0.789076 0.789183
## sigma  0.197424 0.197424
```

```
## loglik at estimate:  9.359407
```

```r
mixmodORiGA <- normalmixEM(results_OR_iwa$GA,
                           lambda = c(1/2), ## inits
                           mu = c(.04,.1), ## inits
                           sigma = 1)
```

```
## number of iterations= 4
```

```r
summary(mixmodORiGA)
```

```
## summary of normalmixEM object:
##           comp 1   comp 2
## lambda 0.493542 0.506458
## mu     0.499795 0.501375
## sigma  0.290404 0.290404
## loglik at estimate:  -10.21771
```
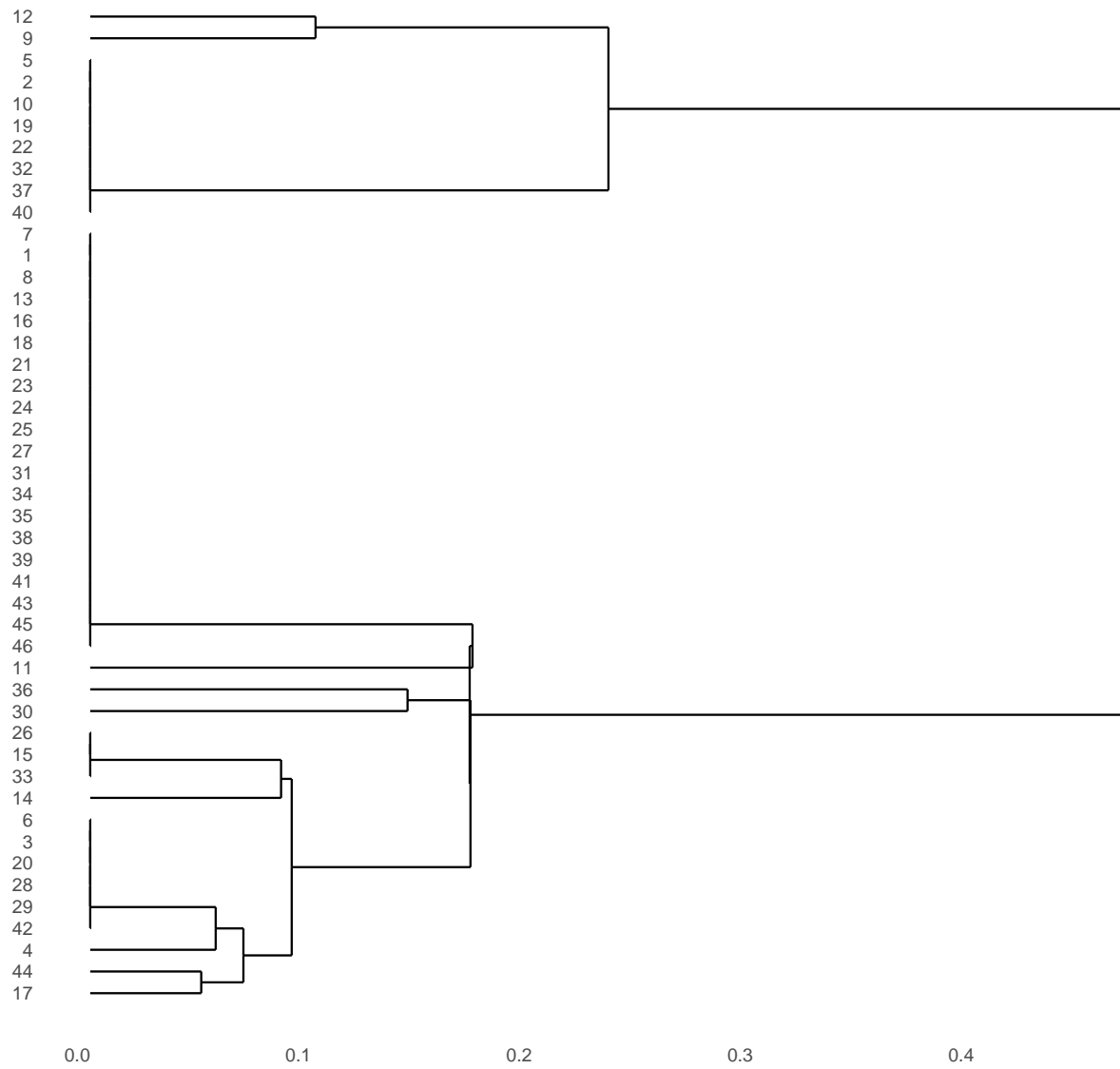
Hierarchical clustering

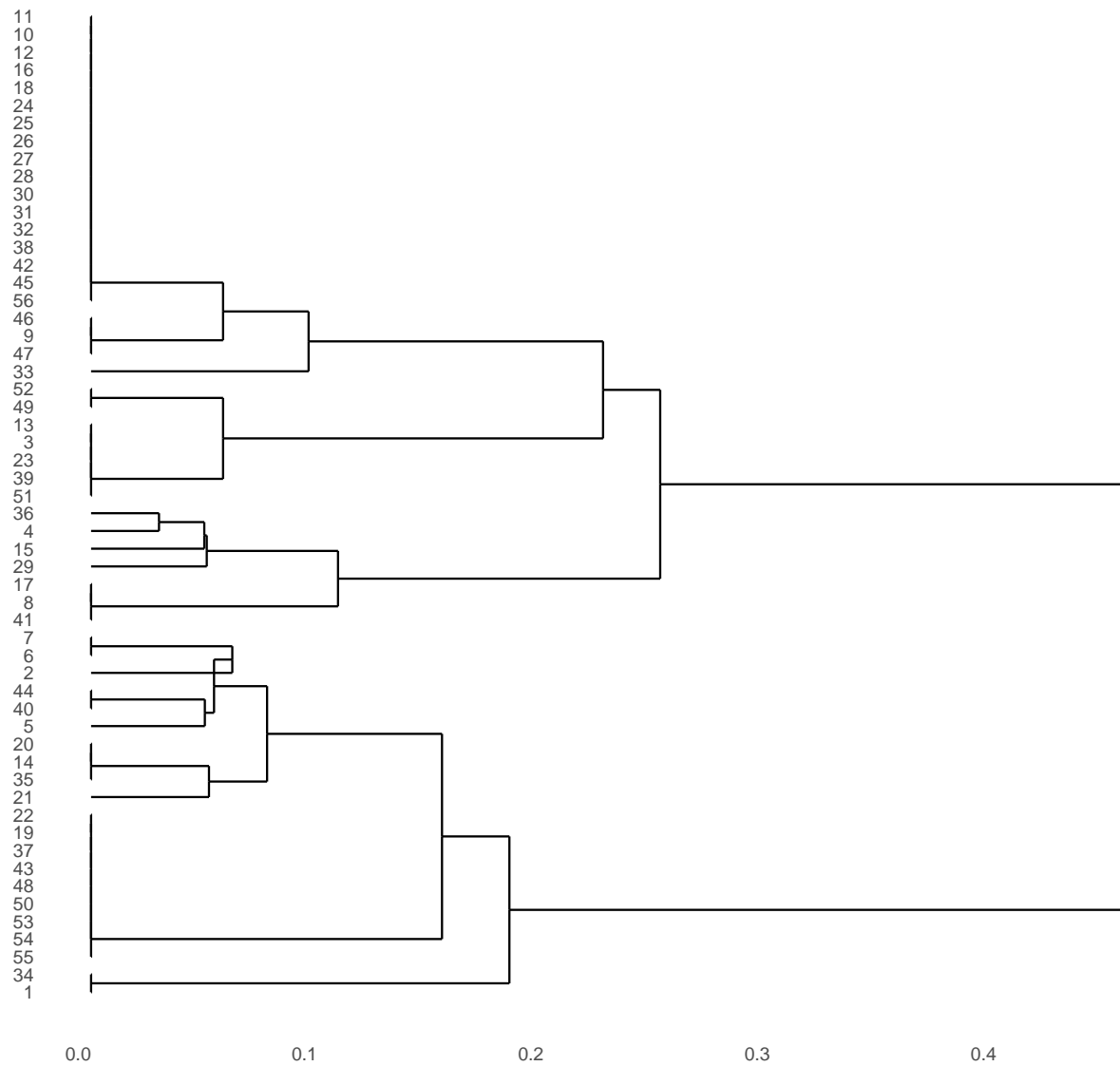I don't see much of a difference between the two groups in the clustering.

In SRs I chose centroid as the method because that gives the best discrimination ability between controls and iwa.

Hierarchical clustering, SR controls

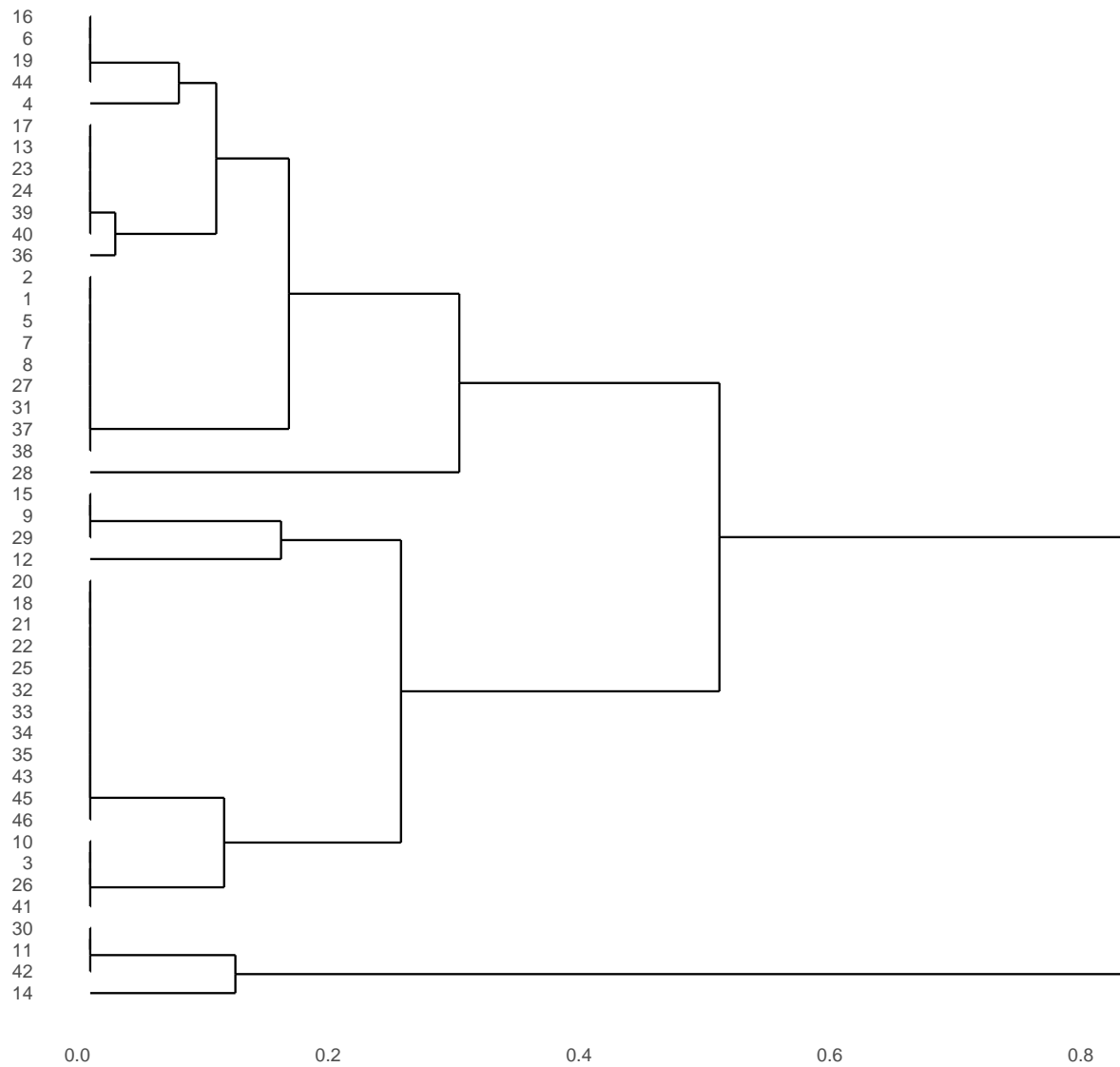## Hierarchical clustering, SR iwa

For ORs I chose complete as method because it gives the best discriminative performance between controls and iwa.

Hierarchical clustering, OR controls

Hierarchical clustering, OR iwa
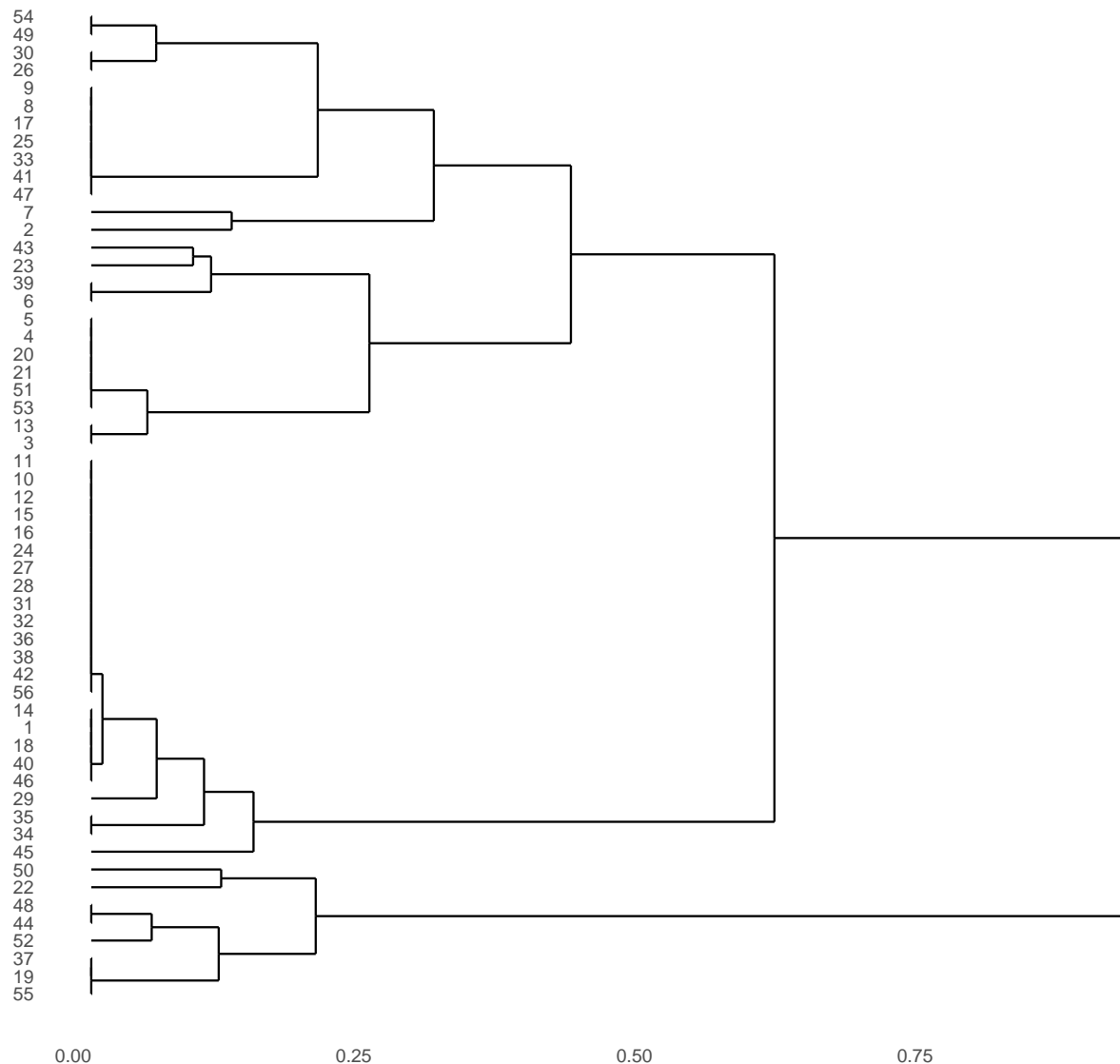
Principal components analysis

We can see whether we can discriminate between controls and aphasics by decomposing the three dimensions into its principal components.

```
## combining controls and patient data:
results_SR_controls$type<-"control"
results_SR_iwa$type<-"iwa"
results_SR<-rbind(results_SR_controls,results_SR_iwa)
#summary(results_SR)

## we will use correlations, because the scales are different:
pcSR<-prcomp(results_SR[,1:3],scale=TRUE)
```

```
summary(pcSR)
```

```
## Importance of components%s:
##                           PC1    PC2    PC3
## Standard deviation     1.4649 0.7283 0.5688
## Proportion of Variance 0.7153 0.1768 0.1079
## Cumulative Proportion  0.7153 0.8921 1.0000
```

```
#screeplot(pcSR)
#pcSR$sdev
```

```
round(head(pcSR$rotation, 5), 2)
```

```
##       PC1   PC2  PC3
## GA  -0.61  0.08 0.79
## DAT  0.57 -0.65 0.50
## ANS  0.55  0.75 0.35
```

The first component separates GA from the ANS, DAT, the third component separates ANS and DAT, GA.

```
results_OR_controls$type<-"control"
results_OR_iwa$type<-"iwa"
results_OR<-rbind(results_OR_controls,results_OR_iwa)

pcOR<-prcomp(results_OR[,1:3],scale=TRUE)
summary(pcOR)
```

```
## Importance of components%s:
##                           PC1    PC2     PC3
## Standard deviation     1.5240 0.6734 0.47331
## Proportion of Variance 0.7742 0.1512 0.07467
## Cumulative Proportion  0.7742 0.9253 1.00000
```

```
round(head(pcOR$rotation, 5), 2)
```

```
##       PC1   PC2  PC3
## GA  -0.60  0.27 0.75
## DAT  0.59 -0.49 0.64
## ANS  0.54  0.83 0.14
```

The first component separates GA and ANS, DAT, and the second component separates ANS and GA, DAT.

Discrimination ability using hierarchical clustering

The discrimination ability is OK for controls, but below 50% for iwa, suggesting that the three parameters don't give us much ability to discriminate controls and iwa.

```
clusters <- hclust(dist(results_SR[, 1:3]),method="centroid")
clusterCut <- cutree(clusters, 2)
## not able to identify aphasics well
table(clusterCut, results_SR$type)
```

```
##
## clusterCut control iwa
##         1      34   21
##         2      12   35
```

```
96/(96+25)
```

```
## [1] 0.7933884
```

```
45/(45+57)
```

```
## [1] 0.4411765
```

Consider ORs. In ORs, the discrimination ability is a bit better than SRs, but not really good. The three parameters do allow us to discriminate between the two groups.

```
clusters <- hclust(dist(results_OR[, 1:3]),method="centroid")
clusterCut <- cutree(clusters, 2)
## so-so discrimination ability for aphasics:
table(clusterCut, results_OR$type)
```

```
## 
## clusterCut control iwa
##          1      42  24
##          2       4  32
```

Discrimination ability using k-means

I'm using the `factoextra` package for an automatised workflow for k-means and their implementation of the gap statistic (cf. Hastie, Tibshirani & Friedman, 2008 2nd ed, p. 519f.). Using the `eclust` function with method k-means and a k equal to NULL results in estimation of optimal cluster size using the gap statistic.

```
library(factoextra)
SR_clusters <- eclust(results_SR[, 1:3], k=2)
```

```
table(SR_clusters$cluster, results_SR$type)
```

```
##
##      control iwa
##   1       11  35
##   2       35  21
```

```
# SR, controls
(101/(101+23))
```

```
## [1] 0.8145161
```

```
# SR, iwa
(42/(60+42))
```

```
## [1] 0.4117647
```
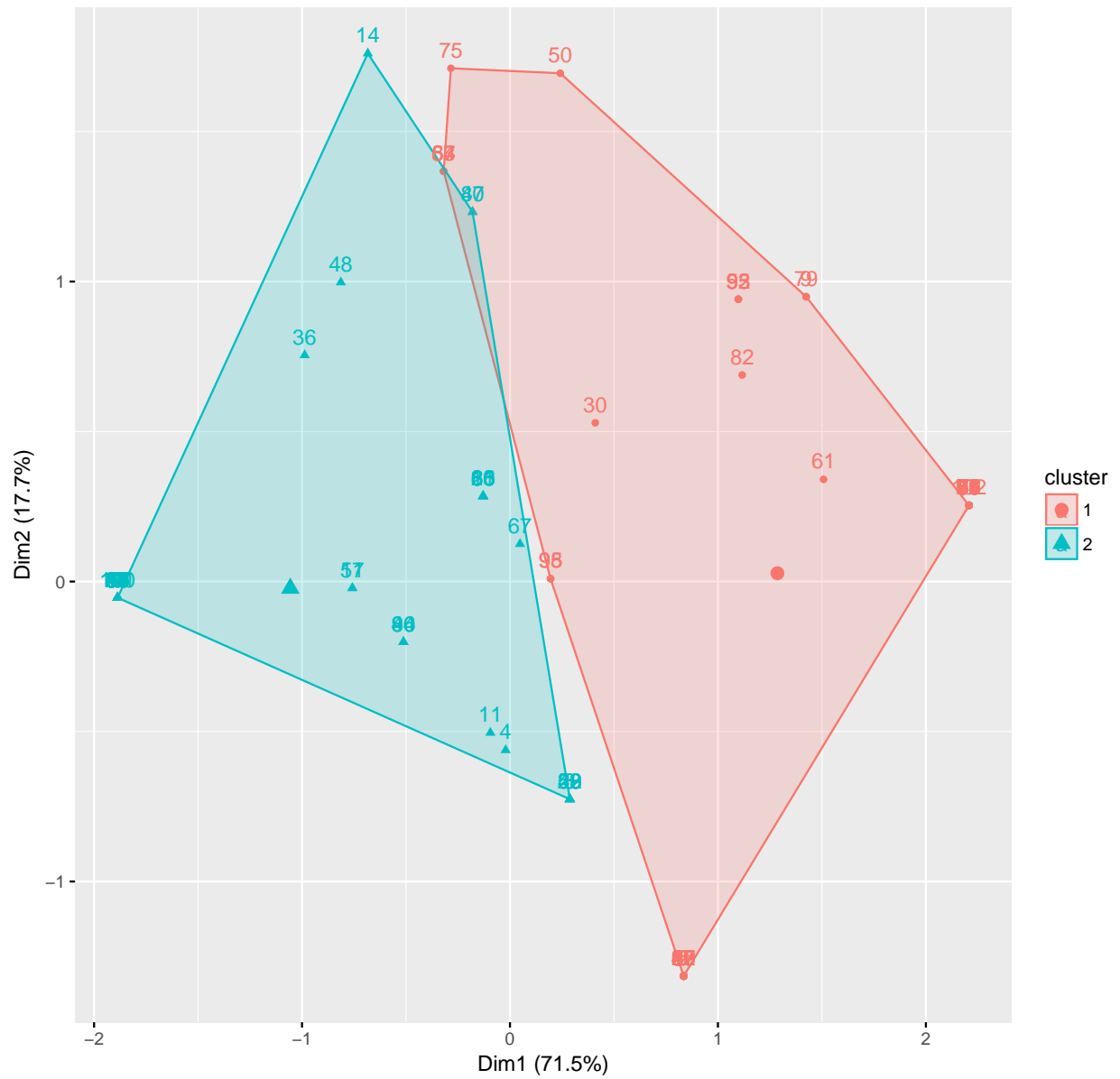
```r
OR_clusters <- eclust(results_OR[, 1:3], k=2)
```

KMEANS Clustering
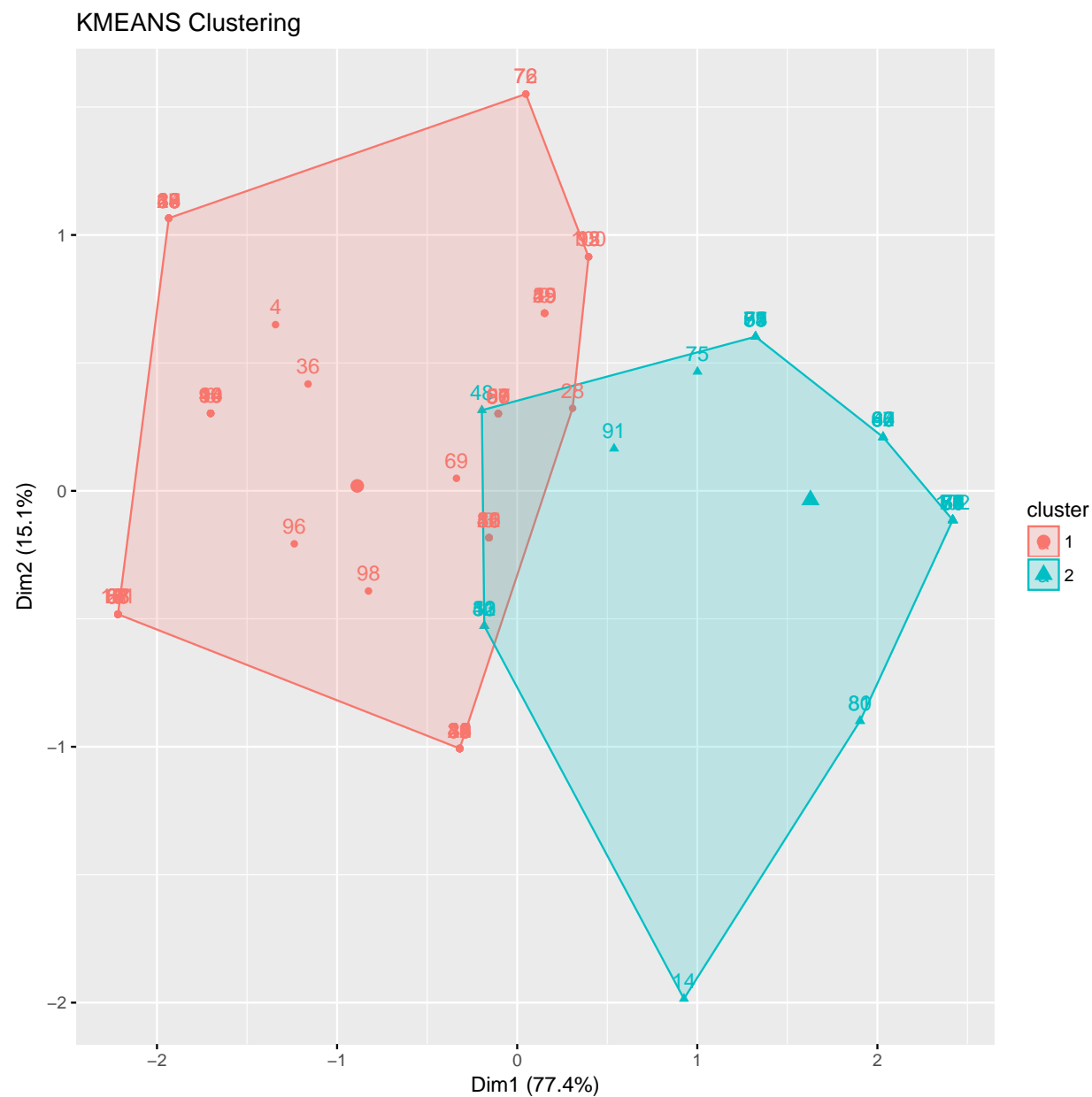

```r
table(OR_clusters$cluster, results_OR$type)
```

```
##
##     control iwa
##   1      42  24
##   2       4  32
```

```r
(103/(102+17))
```

```
## [1] 0.8655462
```

```r
(45/(41+45))
```

```
## [1] 0.5232558
```

21

Exploratory analysis of sub-clusters in SR/OR + controls/iwa

I disabled the code chunks for this section, because I think this kind of exploratory analysis will make more sense once we have data from more constructions.

This doesn't return any results but the fact that kmeans might not be the best idea to estimate the number of clusters for this problem. They are estimated with the gap statistic with the `firstSEmax` method, which looks for the smallest k such that $f(k)$ is not more than 1 standard error away from the first local maximum.

This is same again, but this time with the gap statistic proposed by Tibshirani et al (2001). It always returns that k_{max} is the best cluster size. K-means does not discriminate different clusters in the 4 datasets (SR-controls, SR-iwa, OR-controls, OR-iwa), and also, according to these results, there are not less and clearer clusters in the controls vs. the iwa datasets.

Distributions of subjects

```
results_SR$RCType<-"SR"
results_OR$RCType<-"OR"
results_all<-rbind(results_SR,results_OR)

results_all$DATnormal<-ifelse(results_all$DAT > .06,
      "elevated","normal")
xtabs(~DATnormal+type+RCType,results_all)
```

```
## , , RCType = OR
##
##           type
## DATnormal  control iwa
##   elevated      26  48
##   normal        20   8
##
## , , RCType = SR
##
##           type
## DATnormal  control iwa
##   elevated      24  41
##   normal        22  15
```

```
results_all$ANSnormal<-ifelse(results_all$ANS > .2,
      "elevated","normal")
xtabs(~ANSnormal+type+RCType,results_all)
```

```
## , , RCType = OR
##
##           type
## ANSnormal  control iwa
##   elevated      36  53
##   normal        10   3
##
## , , RCType = SR
##
##           type
## ANSnormal  control iwa
```

```
##    elevated       18  42
##    normal         28  14
```

```
results_all$GAnormal<-ifelse(results_all$GA < .8,
        "elevated","normal")
xtabs(~GAnormal+type+RCType,results_all)
```

```
## , , RCType = OR
##
##           type
## GAnormal    control iwa
##    elevated       21  40
##    normal         25  16
##
## , , RCType = SR
##
##           type
## GAnormal    control iwa
##    elevated       19  38
##    normal         27  18
## two impairments:
```

```
results_all$DATGAnormal<-ifelse(results_all$DAT > .06 & results_all$GA < .8,
        "elevated","normal")
xtabs(~DATGAnormal+type+RCType,results_all)
```

```
## , , RCType = OR
##
##              type
## DATGAnormal control iwa
##     elevated       21  38
##     normal         25  18
##
## , , RCType = SR
##
##              type
## DATGAnormal control iwa
##     elevated       18  32
##     normal         28  24
```

```
results_all$ANSGAnormal<-ifelse(results_all$ANS > .2 & results_all$GA < .8,
        "elevated","normal")
xtabs(~ANSGAnormal+type+RCType,results_all)
```

```
## , , RCType = OR
##
##              type
## ANSGAnormal control iwa
##     elevated       20  40
##     normal         26  16
##
## , , RCType = SR
##
##              type
## ANSGAnormal control iwa
##     elevated       11  33
```

```
##     normal        35  23
```
```
results_all$DATANSnormal<-ifelse(results_all$DAT > .06 & results_all$ANS > .2,
        "elevated","normal")
xtabs(~DATANSnormal+type+RCType,results_all)
```
```
## , , RCType = OR
##
##              type
## DATANSnormal control iwa
##     elevated     25  48
##     normal       21   8
##
## , , RCType = SR
##
##              type
## DATANSnormal control iwa
##     elevated     16  36
##     normal       30  20
```
```
## all three
results_all$GADATANSnormal<-ifelse(results_all$GA < 0.8 & results_all$DAT > .06 & results_all$ANS > .2,
        "elevated","normal")
xtabs(~GADATANSnormal+type+RCType,results_all)
```
```
## , , RCType = OR
##
##                type
## GADATANSnormal control iwa
##       elevated     20  38
##       normal       26  18
##
## , , RCType = SR
##
##                type
## GADATANSnormal control iwa
##       elevated     10  27
##       normal       36  29
```

Proportion of normal values for controls

proportion of controls, SR with normal ANS: 0.6086957

proportion of controls, SR with normal DAT: 0.4782609

proportion of controls, SR with normal GA: 0.5869565

proportion of controls, OR with normal ANS: 0.2173913

proportion of controls, OR with normal DAT: 0.4347826

proportion of controls, OR with normal GA: 0.5434783

Proportion of normal values for IWA

proportion of IWA, SR with normal ANS: 0.25

proportion of IWA, SR with normal DAT: 0.2678571

proportion of IWA, SR with normal GA: 0.3214286

proportion of IWA, OR with normal ANS: 0.0535714

proportion of IWA, OR with normal DAT: 0.1428571

proportion of IWA, OR with normal GA: 0.2857143

Differences in parameter estimates between SR and OR

We estimated parameters for SR and OR separately, which is an error – in the paper, this will have to be fixed. However, considerable differences across many participants only show up in the estimates for the goal activation parameter, see the following table. This should be mentioned in the discussion.

```r
param_differences <- abs(results_SR[c('GA', 'DAT', 'ANS')] - results_OR[c('GA', 'DAT', 'ANS')])
summary(param_differences)
```

```
##       GA               DAT               ANS
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0500   1st Qu.:0.00250   1st Qu.:0.0000
##  Median :0.2000   Median :0.01667   Median :0.0500
##  Mean   :0.2346   Mean   :0.01596   Mean   :0.0594
##  3rd Qu.:0.3500   3rd Qu.:0.02583   3rd Qu.:0.1000
##  Max.   :0.7667   Max.   :0.05000   Max.   :0.2750
```