In [2]:

```
library('Seurat')
library('data.table')
library('Matrix')
library('tidyverse')
library('ggplot2')
library('ggpubr')
library('biomaRt')
library('cluster')
library('pheatmap')
library(org.Hs.eg.db)
library(stringr)
`%notin%` <- Negate(`%in%`)

library(future)
options(future.globals.maxSize = 100000 * 1024^2)
plan("multiprocess", workers = 10)
```

```
Warning message:
“Strategy 'multiprocess' is deprecated in future (>= 1.20.0) [2020-10-30]. Instead, explicitly specify either 'multisession' (recommended) or 'multicore'. In the current R session, 'multiprocess' equ
als 'multicore'.”
```

SPP1+ Mac matrix initialization

In [3]:

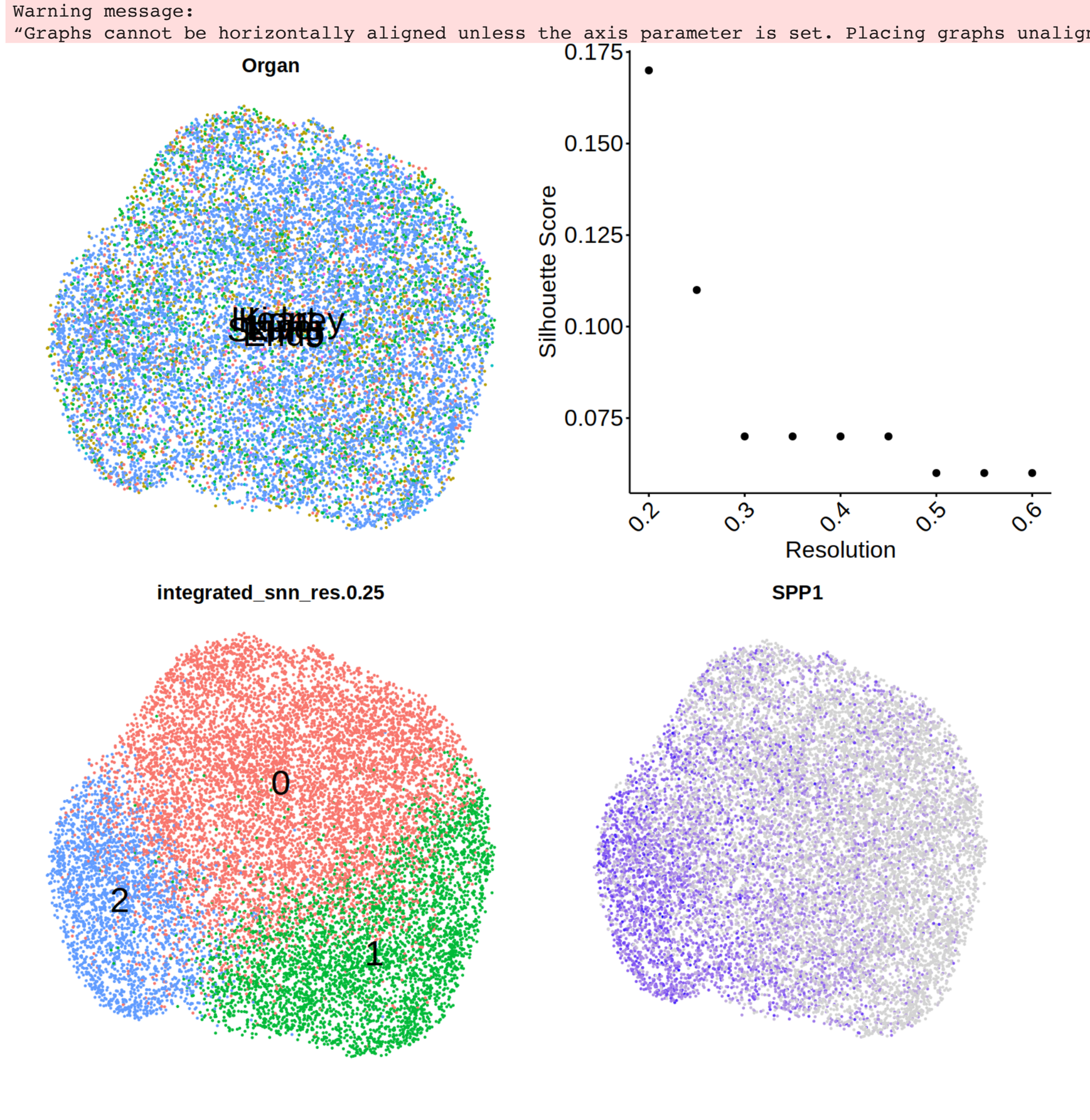
```
Skin_all <- readRDS('~/Skin_Code_Final.RDS')
Kidney_all <- readRDS( 'Kidney_Code_Final.RDS')
All_endo <- readRDS('~/Endo_Final_Code.RDS')
Lung_all <- readRDS('~/Lung_all_Feb_20.RDS')
All_Heart <- readRDS('~/Heart_Code.RDS')
Liver <- readRDS('~/Liver_cd45-.RDS')

transcriptome_SPP1 <- list(All_Heart, Liver, Lung_all, All_endo, Kidney_all, Skin_all)
organ <- c('Heart', 'Liver', 'Lung', 'Endo', 'Kidney', 'Skin')
extract <- c('Diagnosis', 'orig.ident', 'nCount_RNA', 'nFeature_RNA', 'ribo.ratio', 'heam.ratio', 'mito.ratio', 'cluster_d', 'Organ')
extract_liver <- c('Diagnosis', 'orig.ident', 'nCount_RNA', 'nFeature_RNA', 'ribo.ratio', 'heam.ratio', 'mito.ratio', 'cluster_d', 'Organ', 'Patient')

names(transcriptome_SPP1) <- organ
transcriptome_SPP1$Liver$orig.ident <- transcriptome_SPP1$Liver$Patient
for(i in 1:length(transcriptome_SPP1)) {
    transcriptome_SPP1[[i]]$Organ <- organ[[i]]

    if( i != 2 ) {
        transcriptome_SPP1[[i]]@meta.data <- transcriptome_SPP1[[i]]@meta.data[, extract]
    } else {
        transcriptome_SPP1[[i]]@meta.data <- transcriptome_SPP1[[i]]@meta.data[, extract_liver]
    }
    transcriptome_SPP1[[i]] <- RenameCells(object = transcriptome_SPP1[[i]], add.cell.id = paste0(organ[[i]], '_'))
}

common_g <- rownames(transcriptome_SPP1[[1]]$RNA@counts)
for(i in 2:length(transcriptome_SPP1)) {
    common_g <- intersect( common_g , rownames(transcriptome_SPP1[[i]]$RNA@counts) )
}

SPP1 <- list()
for(i in 1:length(transcriptome_SPP1)) {
    SPP1[[i]] <- transcriptome_SPP1[[i]][, transcriptome_SPP1[[i]]$cluster_d %in%  c('SPP1', 'TREM2')]
    DefaultAssay(SPP1[[i]]) <- 'RNA'
    SPP1[[i]] <- SPP1[[i]][ common_g, ]

}
```

Integrative SPP1+ macophage atlas

In [82]:

```
# set.seed(8)
# SPP1[[7]] <- subset( SPP1[[3]], downsample = dim(SPP1[[3]])[2]* 2/ 3)
# SPP1[[3]] <- SPP1[[3]][, colnames(SPP1[[3]]) %notin% colnames(SPP1[[7]]) ]

# for(i in 1:length(SPP1)) {
#     SPP1[[i]] <- SCTransform( SPP1[[i]], vars.to.regress = c('orig.ident', 'mito.ratio'))
# }

# features <- SelectIntegrationFeatures(object.list = SPP1, nfeatures = 3000)
# SPP1 <- PrepSCTIntegration(object.list = SPP1, anchor.features = features)

# anchors <- FindIntegrationAnchors(object.list = SPP1, normalization.method = "SCT", anchor.features = features, reference = c(1, 2, 3, 4, 5, 6), dims = 1:30)
# SPP1 <- IntegrateData(anchorset = anchors, normalization.method = "SCT")
# SPP1 <- RunPCA(SPP1)

# DefaultAssay( SPP1) <- "integrated"
# SPP1 <- FindNeighbors(SPP1, dims = 1:9, k.param = 60L)
# SPP1 <- FindClusters(SPP1, resolution = seq(from = 0.2, to = .6, by = .05))

# DefaultAssay(SPP1) <- 'integrated'
# resolution <- colnames( SPP1@meta.data )[ startsWith(x =colnames( SPP1@meta.data ), prefix = 'integrated_snn_res')]

# silhouetteee <- list()
# i <- 1
# dist.matrix <- dist(x = Embeddings(object = SPP1[['pca']])[,1:9])
# for(sil in sort(resolution)) {
#     s <- silhouette(x = as.numeric(x = as.factor(x = SPP1@meta.data[, sil])), dist = dist.matrix)
#     silhouetteee[[i]] <- summary(s)
#     i <- i + 1
# }

# sil_score <- c()
# for( sil in silhouetteee) {
#     sil_score <- c(sil_score, mean(sil$clus.avg.widths))
# }

# to_plot <- data.frame('Score' = sil_score[1: length(seq(from = 0.2, to = .6, by = .05))], 'Res' = seq(from = 0.2, to = .6, by = .05))
# to_plot$Score <- round(to_plot$Score,digits = 2)

p1 <- ggpubr::ggscatter(data = to_plot, x = 'Res', y = 'Score', xlab = 'Resolution', ylab = 'Silhouette Score') +
 rotate_x_text(angle = 45) +
   theme(text = element_text(size = 19), plot.title = element_text(hjust = 0.5))

# SPP1 <- RunUMAP(SPP1, dims = 1:8, verbose = F)

p3 <- DimPlot(SPP1, group.by = 'Organ', label = T, label.size = 10) + NoLegend() + NoAxes()
p2 <- DimPlot(SPP1, group.by = 'integrated_snn_res.0.25', label = T, label.size = 10) + NoLegend() + NoAxes()

DefaultAssay(SPP1) <- 'RNA'
p4 <-FeaturePlot(SPP1, 'SPP1') + NoLegend() + NoAxes()

options(repr.plot.width=12, repr.plot.height=12)
CombinePlots(list(p3, p1, p2, p4), ncol = 2)
```

```
Warning message:
“CombinePlots is being deprecated. Plots should now be combined using the patchwork system.”
Warning message:
“Graphs cannot be vertically aligned unless the axis parameter is set. Placing graphs unaligned.”
Warning message:
“Graphs cannot be horizontally aligned unless the axis parameter is set. Placing graphs unaligned.”
```



Signature of matrisome-associated macrophage (MAM)

In [111…

```
# DefaultAssay(SPP1) <- 'RNA'
# Idents(SPP1) <- SPP1$integrated_snn_res.0.25

# i <- 1
# MAM.marker <- list()
# for( org in organ) {
#     SPP1_tmp <- SPP1[, SPP1$Organ == org]
#     DefaultAssay( SPP1_tmp) <- 'RNA'
#     SPP1_tmp <- NormalizeData(SPP1_tmp)
#     MAM.marker[[i]] <- FindMarkers( SPP1_tmp, ident.1 = 2, logfc.threshold = -Inf, min.pct = -Inf, min.diff.pct = -Inf)
#     i <- i + 1
# }

MAM_FC_avg <-  MAM.marker[[1]]$avg_log2FC
MAM_diff <-  MAM.marker[[1]]$pct.1 -  MAM.marker[[1]]$pct.2
gene <- rownames( MAM.marker[[1]])

for(i in 2:6) {
    MAM.marker[[i]] <- MAM.marker[[i]][ gene, ]
    MAM_FC_avg  <- cbind(MAM_FC_avg, MAM.marker[[i]][, 'avg_log2FC'])
    MAM_diff  <- cbind(MAM_diff, MAM.marker[[i]]$pct.1 -  MAM.marker[[i]]$pct.2)
}
rownames(MAM_FC_avg) <- gene

MAM_signature <- data.frame( 'log2_FC_avg' = rowMeans( MAM_FC_avg), 'Diff_prop' = rowMeans( MAM_diff))
MAM_signature$Diff_prop_z <- ( MAM_signature$Diff_prop  - mean(MAM_signature$Diff_prop )) /
        sd(MAM_signature$Diff_prop )

MAM_signature_FINAL <- MAM_signature[ MAM_signature$Diff_prop_z > 1.645 & MAM_signature$log2_FC_avg > .25, ]
MAM_signature_FINAL <- MAM_signature_FINAL[ order(MAM_signature_FINAL$log2_FC_av, decreasing = T),]

# SPP1.m <- FindMarkers( SPP1, ident.1 = 2, logfc.threshold = -Inf, min.pct = -Inf, min.diff.pct = -Inf)
# quantile(SPP1.m$p_val_adj)
```

Top 10 MAM markers by fold change (log2)

In [50]:

```
MAM_signature_FINAL <- MAM_signature_FINAL[ order(MAM_signature_FINAL$log2_FC_avg, decreasing = T),]
MAM_signature_FINAL[1:10, ]
```

A data.frame: 10 × 3

	log2_FC_avg	Diff_prop	Diff_prop_z
	<dbl>	<dbl>	<dbl>
SPP1	2.5392168	0.3968333	8.093762
MMP9	1.6347682	0.2300000	4.401996
FABP5	1.3050846	0.1883333	3.479977
LPL	1.2916951	0.3096667	6.164897
TIMP3	1.0913111	0.1148333	1.853535
IL1RN	1.0131095	0.2721667	5.335080
MATK	0.9908850	0.3268333	6.544769
RGCC	0.9600027	0.2210000	4.202840
FBP1	0.9038893	0.2365000	4.545832
TM4SF19	0.8970557	0.2281667	4.361428

Gene-set pathway enrichment analysis of MAM

In [44]:

```
# GO <- rbind( clusterProfiler::read.gmt('/mnt/storage/projects/huang_macrophage/gene_set/Human/Pathways/c2.cp.v7.5.1.symbols.gmt.txt'),
#   clusterProfiler::read.gmt('/mnt/storage/projects/huang_macrophage/gene_set/Human/Pathways/h.all.v7.5.1.symbols.gmt.txt'))
# GO$prefix <- str_split_fixed(GO$term, pattern = '_', n = 2)[,1]
# GO <- GO[GO$prefix %notin% c( 'WP'), ]
# GO$prefix <- NULL
# Term <- c()
# term <- str_split_fixed(GO$term, pattern = '_', n = 20)
# for( i  in  1:dim(term )[1]) {
#     Term <- c(Term, str_trim(paste0(term[i, 2:20], collapse = ' ')))
# }
# GO$term <- Term


# MAM_signature <- MAM_signature[ order( MAM_signature$log2_FC_avg, decreasing = T), ]
# FC_avg <- MAM_signature$log2_FC_avg
# names(FC_avg) <- rownames(MAM_signature)
# FC_avg <- sort(FC_avg[!startsWith(names(FC_avg), prefix =  'PSM')], decreasing = T)


# set.seed(15)
# mac_1 <- clusterProfiler::GSEA(geneList= FC_avg, TERM2GENE = GO, pAdjustMethod = 'BH',
#                 minGSSize = 5, maxGSSize = 500, pvalueCutoff = 0.1, verbose = TRUE)
# mac_1@result  <- mac_1@result[ order(mac_1@result$NES, decreasing = T), ]
mac_1@result[, c('NES', 'p.adjust')][1:10, ]
```

A data.frame: 10 × 2

	NES	p.adjust
	<dbl>	<dbl>
VIBRIO CHOLERAE INFECTION	2.141615	8.066225e-07
EPH EPHRIN SIGNALING	2.141253	4.340037e-07
RESPIRATORY ELECTRON TRANSPORT ATP SYNTHESIS BY CHEMIOSMOTIC COUPLING AND HEAT PRODUCTION BY UNCOUPLING PROTEINS	2.131265	3.572562e-07
OXIDATIVE PHOSPHORYLATION	2.124419	1.261250e-08
RESPIRATORY ELECTRON TRANSPORT	2.115609	2.993429e-07
DEGRADATION OF THE EXTRACELLULAR MATRIX	2.076534	1.509505e-05
L1CAM INTERACTIONS	2.074211	1.175911e-05
EXTRACELLULAR MATRIX ORGANIZATION	2.060036	5.476962e-07
MYC TARGETS V1	2.053015	1.261250e-08
GOLGI TO ER RETROGRADE TRANSPORT	2.051959	1.989478e-06

SPP1+ MAM+ enriched matrisome processes (NABA lab, UIC)

In [112…

```
# NABA <- rbind( clusterProfiler::read.gmt('/mnt/storage/projects/huang_macrophage/gene_set/Human/Pathways/c2.cp.v7.5.1.symbols.gmt.txt'),
#   clusterProfiler::read.gmt('/mnt/storage/projects/huang_macrophage/gene_set/Human/Pathways/h.all.v7.5.1.symbols.gmt.txt'))
# NABA <- NABA[ str_split_fixed( NABA$term , pattern = '_', n = 2)[,1] == 'NABA',]
# NABA$term <- as.character(NABA$term)
# NABA <- NABA[ NABA$term %notin% c('NABA_MATRISOME', 'NABA_MATRISOME_ASSOCIATED', 'NABA_CORE_MATRISOME'), ]


mam_naba <- rownames( MAM_signature )[ MAM_signature$log2_FC_avg > .25 & MAM_signature$Diff_prop_z > 1.645]
mam_naba <- clusterProfiler::enricher(gene = mam_naba, TERM2GENE = NABA, pAdjustMethod = 'BH', universe = rownames( MAM_signature ),
                  pvalueCutoff = 1)

mam_n_naba <- rownames( MAM_signature )[ MAM_signature$log2_FC_avg < -.25 & MAM_signature$Diff_prop_z < -1.975]
mam_n_naba <- clusterProfiler::enricher(gene = mam_n_naba, TERM2GENE = NABA, pAdjustMethod = 'BH',  universe = rownames( MAM_signature ),
                  pvalueCutoff = 1)

mam_naba@result[, c('qvalue','pvalue', 'geneID')]
```

A data.frame: 4 × 3

	qvalue	pvalue	geneID
	<dbl>	<dbl>	<chr>
NABA_ECM_REGULATORS	0.09265265	0.02934001	TGM2/CD109/MMP9/FAM20C/TIMP3/MMP19/ADAM9/CTSZ/SERPINE1/CTSL/ST14
NABA_ECM_AFFILIATED	0.28605479	0.18116804	GPC4/ANXA2/SDC2/ANXA4/ANXA11/LGALS9/CLEC5A
NABA_ECM_GLYCOPROTEINS	0.78791409	0.75804575	SPP1/CRELD2/EMILIN2
NABA_SECRETED_FACTORS	0.78791409	0.99802452	IL1RN/CCL2

SPP1+ MAM- enriched matrisome processes (NABA lab, UIC)

In [113…

```
mam_n_naba@result[, c('qvalue','pvalue', 'geneID')]
```

A data.frame: 4 × 3

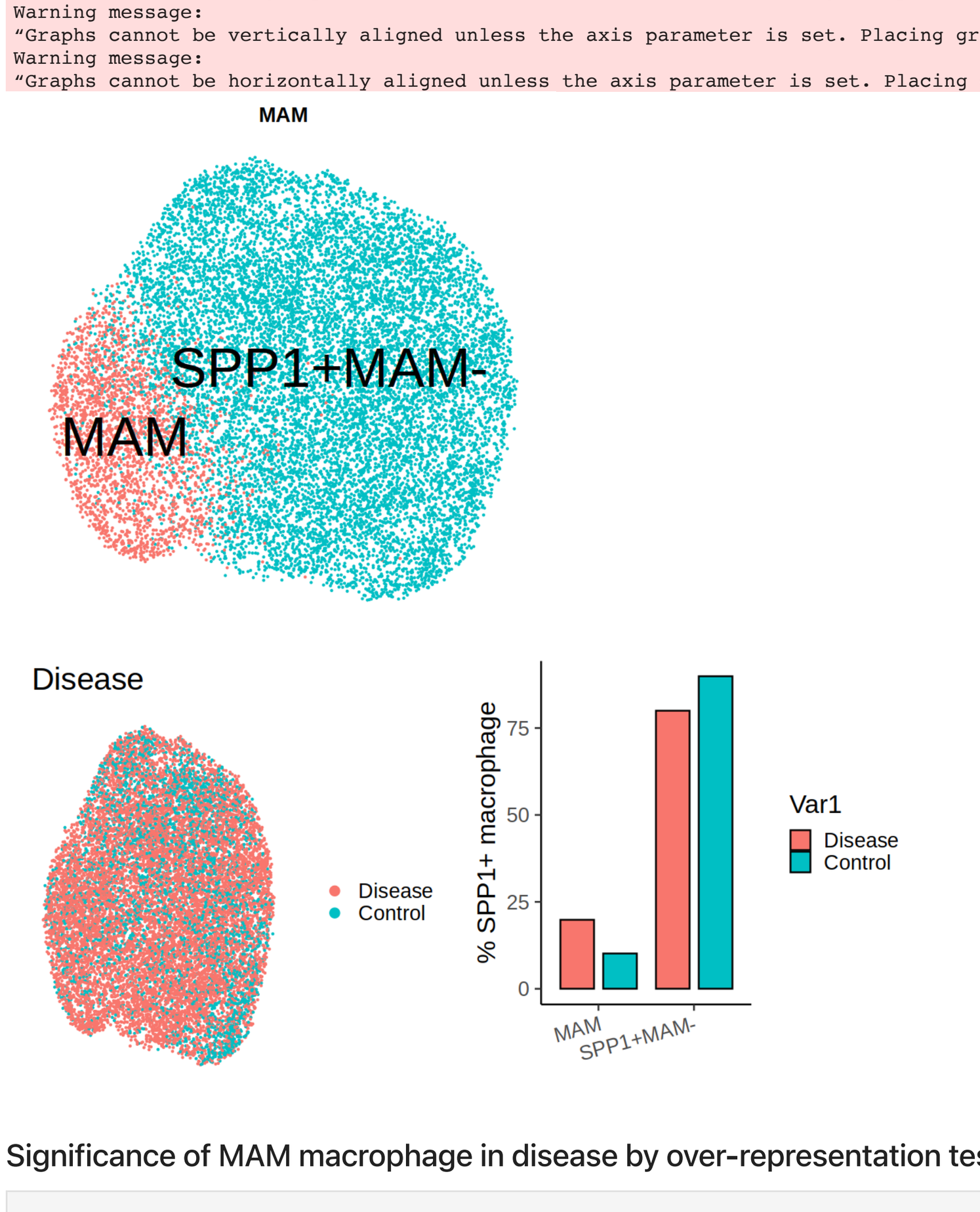
	qvalue	pvalue	geneID
	<dbl>	<dbl>	<chr>
NABA_ECM_AFFILIATED	0.1055209	0.03341495	CLEC10A/C1QA/CLEC4E/CLEC12A/C1QC/C1QB
NABA_SECRETED_FACTORS	0.7280748	0.59313516	CXCL2/IL1B/EREG/IL18
NABA_ECM_GLYCOPROTEINS	0.7280748	0.90568664	FGL2
NABA_ECM_REGULATORS	0.7280748	0.92222814	SERPINF1/SERPING1

MAM in disease v. control

In [67]:

```
# SPP1$Disease <- SPP1$Diagnosis
# SPP1$Disease[ SPP1$Disease!= 'Control'] <- 'Disease'
# SPP1$Disease <- factor( SPP1$Disease, levels = c('Disease', 'Control'))

# SPP1$MAM <- 'MAM'
# SPP1$MAM[SPP1$integrated_snn_res.0.25 != 2 ] <- 'SPP1+MAM-'


# p1 <- DimPlot(SPP1, group.by = 'Disease') + theme_classic( base_line_size = .5, base_size = 20) +  NoAxes()

# to_plot <- melt(table(SPP1$Disease, SPP1$MAM) / rowSums(table(SPP1$Disease, SPP1$MAM)) * 100)
p2 <- ggbarplot(data = to_plot, x= 'Var2', y = 'value', fill = 'Var1', position = position_dodge(0.9)) +
    xlab('') + ylab('% SPP1+ macrophage') + theme_classic( base_line_size = .5, base_size = 20) +
    rotate_x_text(15)

options(repr.plot.width=6, repr.plot.height=6)
DimPlot(SPP1, group.by = 'MAM', label = T, label.size = 15) + NoLegend() + NoAxes()

options(repr.plot.width=10, repr.plot.height=5)
CombinePlots(list( p1, p2), ncol = 2)
```

```
Warning message:
“CombinePlots is being deprecated. Plots should now be combined using the patchwork system.”
Warning message:
“Graphs cannot be vertically aligned unless the axis parameter is set. Placing graphs unaligned.”
Warning message:
“Graphs cannot be horizontally aligned unless the axis parameter is set. Placing graphs unaligned.”
```



Significance of MAM macrophage in disease by over-representation test:

In [66]:

```
over_t <- phyper( sum( SPP1$Disease == 'Disease' & SPP1$MAM == 'MAM') -1, sum(SPP1$Disease == 'Disease'),
       length(SPP1$MAM ) -sum(SPP1$Disease == 'Disease'), sum(SPP1$MAM == 'MAM'),lower.tail= FALSE)

sprintf('p = %s', over_t)
```

'p = 3.92479174287732e-46'

by Organ

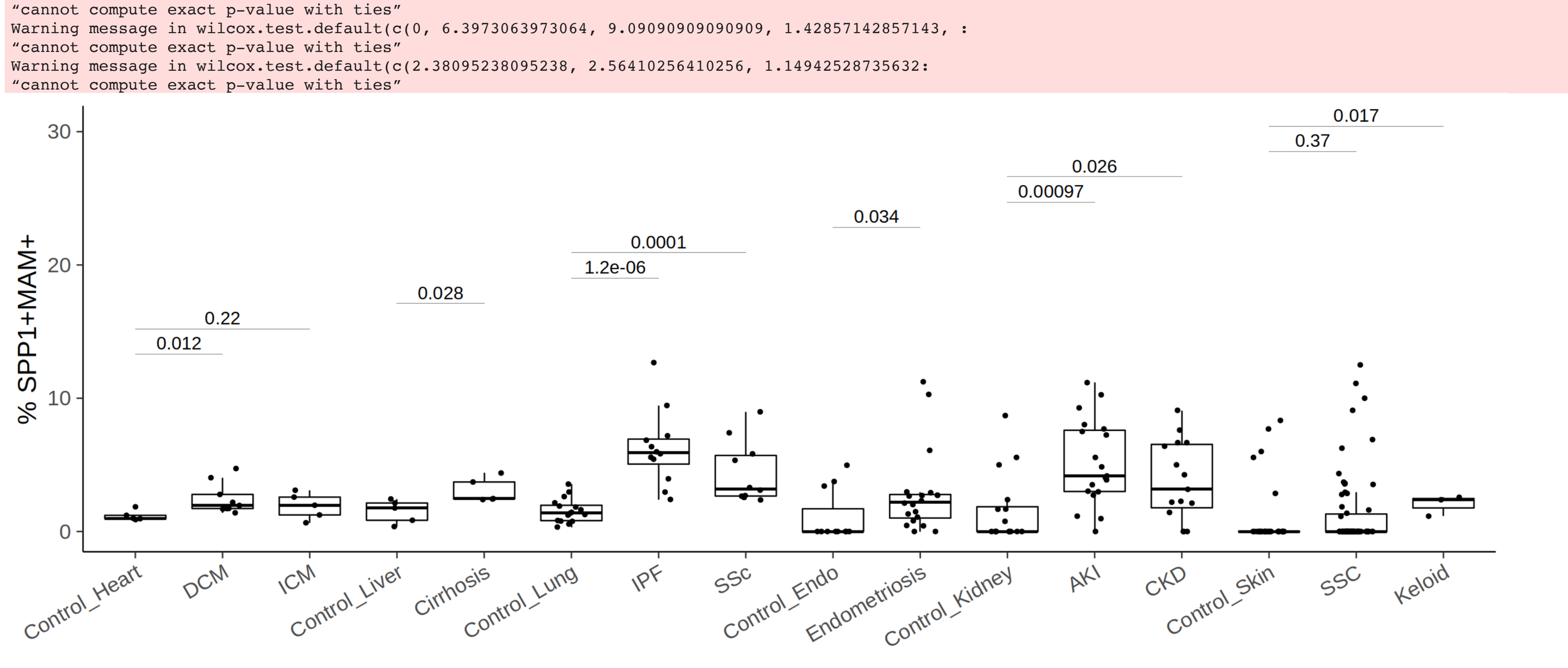
In [24]:

```
SPP1$MAM <- 'MAM'
SPP1$MAM[SPP1$integrated_snn_res.0.25 != 2] <- 'SPP1+MAM-'
```

In [28]:

```
for(i in 1:length(transcriptome_SPP1)) {
    transcriptome_SPP1[[i]]$cluster_d <- as.character( transcriptome_SPP1[[i]]$cluster_d)
    transcriptome_SPP1[[i]]$cluster_d [ colnames(transcriptome_SPP1[[i]]) %in% colnames(SPP1)[ SPP1$MAM == 'MAM']] <- 'MAM'

}

to_plot <- list()
N <- c(0, 0, 0, 10, 10, 5)
D <- list( c('Control', 'DCM', 'ICM'), c('Control', 'Cirrhosis'), c('Control', 'IPF', 'SSc'),
         c('Control', 'Endometriosis'), c('Control', 'AKI', 'CKD'), c('Control', 'SSC', 'Keloid'))


D <- list( c('Control', 'DCM', 'ICM'), c('Control', 'Cirrhosis'), c('Control', 'IPF', 'SSc'),
         c('Control', 'Endometriosis'), c('Control', 'AKI', 'CKD'), c('Control', 'SSC', 'Keloid'))

for( i in 1: length(transcriptome_SPP1) ) {

    MAM <- (table( transcriptome_SPP1[[i]]$orig.ident, transcriptome_SPP1[[i]]$cluster_d) / rowSums(table(transcriptome_SPP1[[i]]$orig.ident, transcriptome_SPP1[[i]]$cluster_d)))[table(transcriptome_
    Diagnosis <-  plyr::mapvalues( names(table( transcriptome_SPP1[[i]]$orig.ident ))[ table(transcriptome_SPP1[[i]]$orig.ident) > N[[i]] ],
        from = transcriptome_SPP1[[i]]$orig.ident[ !duplicated( transcriptome_SPP1[[i]]$orig.ident )], to = transcriptome_SPP1[[i]]$Diagnosis[ !duplicated(transcriptome_SPP1[[i]]$orig.ident)])
    tmp <- data.frame('MAM' = MAM, 'Diagnosis' = Diagnosis)
    tmp <- tmp[ tmp$Diagnosis %in% D[[i]], ]
    tmp$Diagnosis[ tmp$Diagnosis == 'Control'] <- paste0(tmp$Diagnosis[ tmp$Diagnosis == 'Control'], '_', organ[[i]])

    if(i == 1) {
        all_pat <- tmp
    } else {
        all_pat <- rbind(all_pat, tmp)
    }
}

D <- c('Control_Heart', 'DCM', 'ICM','Control_Liver', 'Cirrhosis', 'Control_Lung', 'IPF', 'SSc',
 'Control_Endo', 'Endometriosis', 'Control_Kidney', 'AKI', 'CKD','Control_Skin', 'SSC', 'Keloid')

Com <- list( c('DCM', 'Control_Heart'), c('ICM', 'Control_Heart'),
            c('Cirrhosis', 'Control_Liver'), c('IPF', 'Control_Lung'), c('SSc', 'Control_Lung'),
            c('Endometriosis', 'Control_Endo'), c('AKI', 'Control_Kidney'), c('CKD', 'Control_Kidney'),
            c('SSC', 'Control_Skin'), c('Keloid', 'Control_Skin'))

all_pat$Diagnosis <- factor( all_pat$Diagnosis, levels = D)
all_pat$MAM <- 100 * all_pat$MAM

options(repr.plot.width=20, repr.plot.height=8)
ggboxplot( data = all_pat, x = 'Diagnosis', y = 'MAM',  add = "jitter" ) +
    theme_classic(base_line_size = .5, base_size = 25) + ylab('% SPP1+MAM+') + rotate_x_text(30) + xlab('')  +
    stat_compare_means( comparisons = Com, method = 'wilcox.test', size = 6, bracket.size = 0.1, tip.length = 0, step.increase = 0.15)
```

```
The following `from` values were not present in `x`: 1

The following `from` values were not present in `x`: 28-10051, PRE038, PRE19-025, PRE19-05, PRE98sc, Sample1157-EO2, Sample1157-EO3, Sample1158-EO1, Sample1158-EO2, Sample1158-EO3, Sample1162-EO1, CD
m9, CDm11, CDp1, CDp2

The following `from` values were not present in `x`: GVHD_pt005, Ctrl002, Ctrl007, pt02003, pt03001, pt02004, Ctrl008, pt02006, pt01019, pt03016, pt01014, pt01046

Warning message in wilcox.test.default(c(2.4390243902439, 2.39596469104666, 2.46679316888046, :
“cannot compute exact p-value with ties”
Warning message in wilcox.test.default(c(2.9126213592233, 2.02020202020202, 2.65957446808511, :
“cannot compute exact p-value with ties”
Warning message in wilcox.test.default(c(4.84848484848485, 4.16666666666667, 7.5, :
“cannot compute exact p-value with ties”
Warning message in wilcox.test.default(c(0, 6.3973063973064, 9.09090909090909, 1.42857142857143, :
“cannot compute exact p-value with ties”
Warning message in wilcox.test.default(c(2.38095238095238, 2.56410256410256, 1.14942528735632:
“cannot compute exact p-value with ties”
```



In [118…

```
saveRDS( SPP1, '~/MAM_FEB_24_2023.RDS')
```

In [120…

```
SPP1 <- NULL
memory.profile()
```

NULL: 1 **symbol:** 59615 **pairlist:** 2998492 **closure:** 63300 **environment:** 15117 **promise:** 58069 **language:** 821622 **special:** 45 **builtin:** 692 **char:** 398942 **logical:** 104301 **integer:** 566954 **double:** 44733 **complex:** 62 **character:** 1007269 **...:** 163 **any:** 0 **list:** 276049 **expression:** 4 **bytecode:** 240919 **externalptr:** 12116 **weakref:** 3068 **raw:** 3091 **S4:** 11353