fMRI Scans and Clinical Data Analysis

Setup

```
# load packages

packages <- c("here","dplyr","data.table","psych","FSA","ggplot2")

lapply(packages, library, character.only = TRUE)
```

Read in toddler data

```
# read in data
tidy_fMRI_clinical_toddlers <- read.table(here::here("data", "tidy_fMRI_clinical_toddlers.txt"),
                 header = T, sep = "\t",stringsAsFactors = F)

# toddler sample
toddler_sample <- tidy_fMRI_clinical_toddlers[!duplicated(tidy_fMRI_clinical_toddlers$subjid), ]
dim(toddler_sample)[1]
```

```
## [1] 71
```

```
table(toddler_sample$group)
```

```
##
## ASD  TD
##  41  30
```

```
table(toddler_sample$gender)
```

```
##
##  1  2
## 53 18
```

Toddler fMRI scans and sample size

```r
# total scans
Story_scans <- tidy_fMRI_clinical_toddlers[!is.na(tidy_fMRI_clinical_toddlers$Story_Lang),]
Karen_scans <- tidy_fMRI_clinical_toddlers[!is.na(tidy_fMRI_clinical_toddlers$Karen_Lang),]
Motherese_scans <- tidy_fMRI_clinical_toddlers[!is.na(tidy_fMRI_clinical_toddlers$Motherese),]

knitr::kable(rbind(Mild_affect_speech = table(Story_scans$group),
      Moderate_affect_speech = table(Karen_scans$group),
      Motherese = table(Motherese_scans$group)))
```

	ASD	TD
Mild_affect_speech	33	26
Moderate_affect_speech	40	33
Motherese	39	29

```r
# sample size
knitr::kable(rbind(Mild_affect_speech = table(Story_scans$group[!duplicated(Story_scans$subjid)]),
      Moderate_affect_speech = table(Karen_scans$group[!duplicated(Karen_scans$subjid)]),
      Motherese = table(Motherese_scans$group[!duplicated(Motherese_scans$subjid)])))
```

	ASD	TD
Mild_affect_speech	31	23
Moderate_affect_speech	36	28
Motherese	37	25

```r
# retest scans
knitr::kable(rbind(Mild_affect_speech = table(Story_scans$group[duplicated(Story_scans$subjid)]),
      Moderate_affect_speech = table(Karen_scans$group[duplicated(Karen_scans$subjid)]),
      Motherese = table(Motherese_scans$group[duplicated(Motherese_scans$subjid)])))
```

	ASD	TD
Mild_affect_speech	2	3
Moderate_affect_speech	4	5
Motherese	2	4

Adult scans and sample size

```r
# read adult data
adult_scans <- read.table(here::here("data", "tidy_fMRI_adults.txt"),
                 header = T, sep = "\t",stringsAsFactors = F)

# adult sample
adult_sample <- adult_scans[!duplicated(adult_scans$Subj), ]
dim(adult_sample)[1]
```

```
## [1] 14
```

```
table(adult_sample$gender)
```

```
##
## F M
## 8 6
```

```
# adult fMRI scans and sample size
Story_scans_adult <- adult_scans[!is.na(adult_scans$Story_Lang),]
Karen_scans_adult <- adult_scans[!is.na(adult_scans$Karen_Lang),]
Motherese_scans_adult <- adult_scans[!is.na(adult_scans$Motherese),]

knitr::kable(cbind(Mild_affect_speech = dim(Story_scans_adult)[1],
      Moderate_affect_speech = dim(Karen_scans_adult)[1],
      Motherese = dim(Motherese_scans_adult)[1]))
```

Mild_affect_speech	Moderate_affect_speech	Motherese
18	12	11

```
# sample size
knitr::kable(cbind(Mild_affect_speech =
            dim(Story_scans_adult[!duplicated(Story_scans_adult$Subj),])[1],
      Moderate_affect_speech = dim(Karen_scans_adult[!duplicated(Karen_scans_adult$Subj),])[1],
      Motherese = dim(Motherese_scans_adult[!duplicated(Motherese_scans_adult$Subj),])[1]))
```

Mild_affect_speech	Moderate_affect_speech	Motherese
13	12	8

```
# retest scans
knitr::kable(cbind(Mild_affect_speech =
            dim(Story_scans_adult[duplicated(Story_scans_adult$Subj),])[1],
      Moderate_affect_speech = dim(Karen_scans_adult[duplicated(Karen_scans_adult$Subj),])[1],
      Motherese = dim(Motherese_scans_adult[duplicated(Motherese_scans_adult$Subj),])[1]))
```

Mild_affect_speech	Moderate_affect_speech	Motherese
5	0	3

Head motion for each language paradigm

```
# mean and sd in toddlers
Story_mean <- describeBy(Story_scans[, c("group","Story_meanFD")], group = "group",
             mat = TRUE,  digits = 2)
Karen_mean <- describeBy(Karen_scans[, c("group","Karen_meanFD")], group = "group",
             mat = TRUE,  digits = 2)
Motherese_mean <- describeBy(Motherese_scans[, c("group","Motherese_meanFD")],
```

```r
                 group = "group",mat = TRUE,  digits = 2)

# ASD vs. TD toddlers
t_Story <- t.test(Story_scans$Story_meanFD[Story_scans$group == "TD"],
          Story_scans$Story_meanFD[Story_scans$group == "ASD"])
t_Karen <-t.test(Karen_scans$Story_meanFD[Karen_scans$group == "TD"],
         Karen_scans$Story_meanFD[Karen_scans$group == "ASD"])
t_Motherese <- t.test(Motherese_scans$Story_meanFD[Motherese_scans$group == "TD"],
              Motherese_scans$Story_meanFD[Motherese_scans$group == "ASD"])

# mean and sd in adults
adult_Story <- Summarize(Story_scans_adult$Story_meanFD)[2:3]
adult_Karen <- Summarize(Karen_scans_adult$Karen_meanFD)[2:3]
adult_Motherese <-Summarize(Motherese_scans_adult$Motherese_meanFD)[2:3]

# adults vs. TD toddlers
t_Story_vsTD <- t.test(Story_scans_adult$Story_meanFD,
           Story_scans$Story_meanFD[Story_scans$group == "TD"])
t_Karen_vsTD <- t.test(Karen_scans_adult$Karen_meanFD,
       Karen_scans$Karen_meanFD[Karen_scans$group == "TD"])
t_Motherese_vsTD <- t.test(Motherese_scans_adult$Motherese_meanFD,
       Motherese_scans$Motherese_meanFD[Motherese_scans$group == "TD"])

# adults vs. ASD toddlers
t_Story_vsASD <- t.test(Story_scans_adult$Story_meanFD,
               Story_scans$Story_meanFD[Story_scans$group == "ASD"])
t_Karen_vsASD <- t.test(Karen_scans_adult$Karen_meanFD,
                Karen_scans$Karen_meanFD[Karen_scans$group == "ASD"])
t_Motherese_vsASD <- t.test(Motherese_scans_adult$Motherese_meanFD,
                Motherese_scans$Motherese_meanFD[Motherese_scans$group == "ASD"])

# summary of head motion in each group and comparisons
FD_sum <- cbind(rbind(paste0(Story_mean$mean[3:4],"(",Story_mean$sd[3:4],")"),
      paste0(Karen_mean$mean[3:4],"(",Karen_mean$sd[3:4],")"),
      paste0(Motherese_mean$mean[3:4], "(",Motherese_mean$sd[3:4],")")),
      rbind(paste0("t=",round(t_Story$statistic,2), " p=",round(t_Story$p.value,2)),
      paste0("t=",round(t_Karen$statistic,2), " p=",round(t_Karen$p.value,2)),
      paste0("t=",round(t_Motherese$statistic,2), " p=",round(t_Motherese$p.value,2))),
      rbind(paste0(round(adult_Story[1],2),"(",round(adult_Story[2],2),")"),
            paste0(round(adult_Karen[1],2),"(",round(adult_Karen[2],2),")"),
            paste0(round(adult_Motherese[1],2),"(",round(adult_Motherese[2],2),")")),
      rbind(paste0("t=",round(t_Story_vsTD$statistic,2), " p=",round(t_Story_vsTD$p.value,2)),
      paste0("t=",round(t_Karen_vsTD$statistic,2), " p=",round(t_Karen_vsTD$p.value,2)),
      paste0("t=",round(t_Motherese_vsTD$statistic,2), " p=",round(t_Motherese_vsASD$p.value,2))),
      rbind(paste0("t=",round(t_Story_vsASD$statistic,2), " p=",round(t_Story_vsASD$p.value,2)),
      paste0("t=",round(t_Karen_vsASD$statistic,2), " p=",round(t_Karen_vsASD$p.value,2)),
      paste0("t=",round(t_Motherese_vsASD$statistic,2), " p=",round(t_Motherese_vsASD$p.value,2))))

rownames(FD_sum) <- c("Mild affect speech", "Moderate affect speech", "Motherese")
colnames(FD_sum) <- c("ASD", "TD", "TD vs ASD","Adults", "Adults vs TD", "Adults vs ASD")

knitr::kable(FD_sum)
```

	ASD	TD	TD vs ASD	Adults	Adults vs TD	Adults vs ASD
Mild affect speech	0.09(0.06)	0.11(0.1)	t=0.88 p=0.38	0.08(0.02)	t=-1.46 p=0.16	t=-0.81 p=0.42
Moderate affect speech	0.09(0.06)	0.1(0.05)	t=1.28 p=0.21	0.08(0.04)	t=-1.32 p=0.2	t=-0.68 p=0.5
Motherese	0.11(0.14)	0.09(0.06)	t=-0.61 p=0.55	0.07(0.03)	t=-1.31 p=0.1	t=-1.66 p=0.1

Demographic information and clinical test scores

```
colnames(dplyr::select(toddler_sample, contains("final")))
```

```
##  [1] "final_Dx"                   "final_ados_CoSoTot"
##  [3] "final_ados_RRTot"           "final_ados_CoSoTotRRTot"
##  [5] "final_vine_ComTotal_DomStd"  "final_vine_DlyTotal_DomStd"
##  [7] "final_vine_SocTotal_DomStd"  "final_vine_MtrTotal_DomStd"
##  [9] "final_vine_AdapBehav_DomStd" "final_vine_DomStdTotal"
## [11] "final_mullen_VRT"           "final_mullen_FMT"
## [13] "final_mullen_RLT"           "final_mullen_ELT"
## [15] "final_mullen_ELC_Std"
```

```
describeBy(toddler_sample[, c("group", "Gender","scan_age","test_age", colnames(dplyr::select(toddler_sa
       group = "group",mat = TRUE,  digits = 3)
```

```
##                              item group1 vars  n   mean     sd median trimmed
## group*1                         1    ASD    1 41  1.000  0.000    1.0   1.000
## group*2                         2     TD    1 30  1.000  0.000    1.0   1.000
## Gender*1                        3    ASD    2 41  1.854  0.358    2.0   1.939
## Gender*2                        4     TD    2 30  1.600  0.498    2.0   1.625
## scan_age1                       5    ASD    3 41 28.805  9.732   27.0  28.000
## scan_age2                       6     TD    3 30 23.700  5.984   22.0  23.292
## test_age1                       7    ASD    4 41 28.879  8.416   33.0  29.004
## test_age2                       8     TD    4 30 26.267  8.136   27.0  26.458
## final_Dx*1                      9    ASD    5 41  1.024  0.156    1.0   1.000
## final_Dx*2                     10     TD    5 30  4.700  2.535    4.5   4.667
## final_ados_CoSoTot1            11    ASD    6 41 12.854  4.059   13.0  13.000
## final_ados_CoSoTot2            12     TD    6 30  2.700  1.489    3.0   2.708
## final_ados_RRTot1              13    ASD    7 41  5.341  2.128    6.0   5.455
## final_ados_RRTot2              14     TD    7 30  1.233  1.165    1.0   1.125
## final_ados_CoSoTotRRTot1       15    ASD    8 41 18.195  5.372   18.0  18.394
## final_ados_CoSoTotRRTot2       16     TD    8 30  3.933  1.799    4.0   3.833
## final_vine_ComTotal_DomStd1    17    ASD    9 41 82.927 16.626   85.0  84.030
## final_vine_ComTotal_DomStd2    18     TD    9 30 97.167 11.859   96.0  96.583
## final_vine_DlyTotal_DomStd1    19    ASD   10 41 86.366 11.764   85.0  85.758
## final_vine_DlyTotal_DomStd2    20     TD   10 30 97.533 12.227   96.5  97.250
## final_vine_SocTotal_DomStd1    21    ASD   11 41 82.951 12.586   84.0  83.455
## final_vine_SocTotal_DomStd2    22     TD   11 30 98.567 10.311   99.0  98.000
## final_vine_MtrTotal_DomStd1    23    ASD   12 41 89.829 17.943   89.0  91.212
```

```
## final_vine_MtrTotal_DomStd2    24      TD   12 30  94.633 20.769   97.5  97.750
## final_vine_AdapBehav_DomStd1   25     ASD   13 41  82.366 11.510   83.0  81.667
## final_vine_AdapBehav_DomStd2   26      TD   13 30  96.800 10.889   97.5  96.333
## final_vine_DomStdTotal1        27     ASD   14 41 341.829 44.069  344.0 341.030
## final_vine_DomStdTotal2        28      TD   14 30 388.200 34.035  392.0 388.208
## final_mullen_VRT1              29     ASD   15 41  38.610 12.730   40.0  39.000
## final_mullen_VRT2              30      TD   15 30  54.300 11.621   55.0  54.000
## final_mullen_FMT1              31     ASD   16 41  39.951 11.853   42.0  40.455
## final_mullen_FMT2              32      TD   16 30  50.000  8.154   49.0  49.917
## final_mullen_RLT1              33     ASD   17 41  32.293 14.780   26.0  31.667
## final_mullen_RLT2              34      TD   17 30  48.200 11.493   47.5  48.042
## final_mullen_ELT1              35     ASD   18 41  33.098 16.143   30.0  32.879
## final_mullen_ELT2              36      TD   18 30  43.767 12.204   42.0  43.042
## final_mullen_ELC_Std1          37     ASD   19 41  74.073 21.979   72.0  74.818
## final_mullen_ELC_Std2          38      TD   19 30  98.367 16.587   97.5  98.042
##                                  mad min max range   skew kurtosis    se
## group*1                        0.000   1   1     0    NaN      NaN 0.000
## group*2                        0.000   1   1     0    NaN      NaN 0.000
## Gender*1                       0.000   1   2     1 -1.928    1.764 0.056
## Gender*2                       0.000   1   2     1 -0.388   -1.910 0.091
## scan_age1                     10.378  14  55    41  0.727    0.002 1.520
## scan_age2                      6.672  14  38    24  0.511   -0.571 1.092
## test_age1                      4.448  12  51    39 -0.025   -0.205 1.314
## test_age2                     10.378  13  37    24 -0.163   -1.548 1.486
## final_Dx*1                     0.000   1   2     1  5.942   34.145 0.024
## final_Dx*2                     3.706   1   8     7 -0.005   -1.817 0.463
## final_ados_CoSoTot1            4.448   0  20    20 -0.621    0.669 0.634
## final_ados_CoSoTot2            1.483   0   6     6  0.078   -0.361 0.272
## final_ados_RRTot1              2.965   0   9     9 -0.347   -0.711 0.332
## final_ados_RRTot2              1.483   0   4     4  0.444   -0.895 0.213
## final_ados_CoSoTotRRTot1       5.930   5  27    22 -0.334   -0.372 0.839
## final_ados_CoSoTotRRTot2       1.483   1   8     7  0.474   -0.539 0.328
## final_vine_ComTotal_DomStd1   14.826  35 126    91 -0.475    0.875 2.597
## final_vine_ComTotal_DomStd2   10.378  70 122    52  0.231   -0.135 2.165
## final_vine_DlyTotal_DomStd1   14.826  68 116    48  0.469   -0.423 1.837
## final_vine_DlyTotal_DomStd2   11.119  76 122    46  0.208   -1.019 2.232
## final_vine_SocTotal_DomStd1   16.309  57 108    51 -0.205   -0.871 1.966
## final_vine_SocTotal_DomStd2    8.896  79 126    47  0.526    0.463 1.883
## final_vine_MtrTotal_DomStd1   10.378   0 117   117 -2.845   12.703 2.802
## final_vine_MtrTotal_DomStd2    9.637   0 119   119 -3.081   11.578 3.792
## final_vine_AdapBehav_DomStd1   8.896  58 111    53  0.445    0.206 1.798
## final_vine_AdapBehav_DomStd2  10.378  79 128    49  0.587    0.561 1.988
## final_vine_DomStdTotal1       51.891 250 445   195  0.099   -0.515 6.882
## final_vine_DomStdTotal2       24.463 315 483   168  0.188    0.675 6.214
## final_mullen_VRT1             11.861   1  63    62 -0.512    0.241 1.988
## final_mullen_VRT2             13.343  30  77    47  0.078   -0.746 2.122
## final_mullen_FMT1             11.861  20  57    37 -0.495   -1.074 1.851
## final_mullen_FMT2             10.378  35  64    29  0.089   -1.075 1.489
## final_mullen_RLT1             10.378   1  59    58  0.253   -1.095 2.308
## final_mullen_RLT2             12.602  23  72    49  0.093   -0.528 2.098
## final_mullen_ELT1             14.826   1  63    62  0.228   -1.011 2.521
## final_mullen_ELT2             14.085  25  70    45  0.367   -0.802 2.228
## final_mullen_ELC_Std1         22.239   7 115   108 -0.467    0.545 3.433
## final_mullen_ELC_Std2         15.567  71 127    56  0.191   -1.063 3.028
```

```r
# Chi-squared test on gender by group
gender_diff <- table(toddler_sample$Gender, toddler_sample$group)
knitr::kable(gender_diff)
```

	ASD	TD
F	6	12
M	35	18

```r
chisq.test(gender_diff)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  gender_diff
## X-squared = 4.6259, df = 1, p-value = 0.03149
```

```r
# group differences between ASD and TD
lapply(toddler_sample[, c("scan_age","test_age",colnames(dplyr::select(toddler_sample, contains("final"))
       function(x) t.test(x ~ toddler_sample$group, var.equal = TRUE))
```

```
## $scan_age
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = 2.5404, df = 69, p-value = 0.01333
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.096119 9.113638
## sample estimates:
## mean in group ASD  mean in group TD
##          28.80488          23.70000
##
##
## $test_age
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = 1.3098, df = 69, p-value = 0.1946
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.366141  6.589881
## sample estimates:
## mean in group ASD  mean in group TD
##          28.87854          26.26667
##
##
## $final_ados_CoSoTot
##
```

```
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = 13.052, df = 69, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   8.601704 11.705613
## sample estimates:
## mean in group ASD  mean in group TD
##          12.85366           2.70000
##
##
## $final_ados_RRTot
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = 9.5632, df = 69, p-value = 2.845e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.251147 4.965113
## sample estimates:
## mean in group ASD  mean in group TD
##          5.341463          1.233333
##
##
## $final_ados_CoSoTotRRTot
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = 13.956, df = 69, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.22313 16.30044
## sample estimates:
## mean in group ASD  mean in group TD
##         18.195122          3.933333
##
##
## $final_vine_ComTotal_DomStd
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -4.0019, df = 69, p-value = 0.0001559
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.338451  -7.141224
## sample estimates:
## mean in group ASD  mean in group TD
##          82.92683          97.16667
##
##
```

```
## $final_vine_DlyTotal_DomStd
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -3.8862, df = 69, p-value = 0.000231
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.900232  -5.434727
## sample estimates:
## mean in group ASD  mean in group TD
##          86.36585          97.53333
##
##
## $final_vine_SocTotal_DomStd
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -5.5628, df = 69, p-value = 4.692e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.21549 -10.01540
## sample estimates:
## mean in group ASD  mean in group TD
##          82.95122          98.56667
##
##
## $final_vine_MtrTotal_DomStd
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -1.0424, df = 69, p-value = 0.3008
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.997691   4.389561
## sample estimates:
## mean in group ASD  mean in group TD
##          89.82927          94.63333
##
##
## $final_vine_AdapBehav_DomStd
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -5.3386, df = 69, p-value = 1.132e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -19.827972  -9.040321
## sample estimates:
## mean in group ASD  mean in group TD
##          82.36585          96.80000
```

```
##
##
## $final_vine_DomStdTotal
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -4.8061, df = 69, p-value = 8.672e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -65.61853 -27.12293
## sample estimates:
## mean in group ASD  mean in group TD
##          341.8293          388.2000
##
##
## $final_mullen_VRT
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -5.3199, df = 69, p-value = 1.218e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.574061  -9.806427
## sample estimates:
## mean in group ASD  mean in group TD
##          38.60976          54.30000
##
##
## $final_mullen_FMT
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -3.999, df = 69, p-value = 0.0001575
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -15.061763  -5.035798
## sample estimates:
## mean in group ASD  mean in group TD
##          39.95122          50.00000
##
##
## $final_mullen_RLT
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -4.9056, df = 69, p-value = 5.967e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -22.376336  -9.438298
## sample estimates:
```

```
## mean in group ASD  mean in group TD
##          32.29268          48.20000
##
##
## $final_mullen_ELT
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -3.038, df = 69, p-value = 0.003361
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -17.675195  -3.663016
## sample estimates:
## mean in group ASD  mean in group TD
##          33.09756          43.76667
##
##
## $final_mullen_ELC_Std
##
##  Two Sample t-test
##
## data:  x by toddler_sample$group
## t = -5.0833, df = 69, p-value = 3.037e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -33.82757 -14.75943
## sample estimates:
## mean in group ASD  mean in group TD
##          74.07317          98.36667
```

Motherese eye-tracking data

```r
# read in data
Motherese_ET <- read.table(here::here("data","tidy_Motherese_ET.txt"),header = T,
               sep = "\t", stringsAsFactors = F)

colnames(Motherese_ET)[2] <- "subj"

# age at Motherese eye-tracking test
table(Motherese_ET$group)
```

```
##
## ASD  TD
##  31  23
```

```r
table(Motherese_ET$group,Motherese_ET$gender)
```

```
##
##        1  2
```

```
##   ASD 26  5
##   TD  15  8
```

```
gender_diff <- table(Motherese_ET$group,Motherese_ET$gender)
knitr::kable(gender_diff)
```

	1	2
ASD	26	5
TD	15	8

```
chisq.test(gender_diff)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  gender_diff
## X-squared = 1.5966, df = 1, p-value = 0.2064
```

```
min(Motherese_ET$ET.Age)
```

```
## [1] 12
```

```
max(Motherese_ET$ET.Age)
```

```
## [1] 42
```

```
mean(Motherese_ET$ET.Age[Motherese_ET$group == "ASD"])
```

```
## [1] 25.32258
```

```
sd(Motherese_ET$ET.Age[Motherese_ET$group == "ASD"])
```

```
## [1] 8.553312
```

```
mean(Motherese_ET$ET.Age[Motherese_ET$group == "TD"])
```

```
## [1] 26.39435
```

```
sd(Motherese_ET$ET.Age[Motherese_ET$group == "TD"])
```

```
## [1] 8.056302
```

```
t.test(Motherese_ET$ET.Age[Motherese_ET$group == "TD"],
         Motherese_ET$ET.Age[Motherese_ET$group == "ASD"])
```

```
##
##  Welch Two Sample t-test
##
## data:  Motherese_ET$ET.Age[Motherese_ET$group == "TD"] and Motherese_ET$ET.Age[Motherese_ET$group ==
## t = 0.47082, df = 49.035, p-value = 0.6399
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.502699  5.646234
## sample estimates:
## mean of x mean of y
##  26.39435  25.32258
```

```
# group differences in Motherese eye-tracking between ASD and TD
eff <- effsize::cohen.d(Motherese_ET$LK_.fixation.Motherese, Motherese_ET$group)
tvalue <- t.test(Motherese_ET$LK_.fixation.Motherese[Motherese_ET$group == "TD"],
         Motherese_ET$LK_.fixation.Motherese[Motherese_ET$group == "ASD"],
         alternative = "greater")

knitr::kable(cbind(effect_size = abs(eff$estimate), p_value = tvalue$p.value))
```

effect_size	p_value
0.8258775	0.00105

```
# Motherese eye-tracking test: how many completed before the scan and how many after the scan
dim(Motherese_ET[as.Date(Motherese_ET$ScanDate) > as.Date(Motherese_ET$LK_Date),])[1]
```

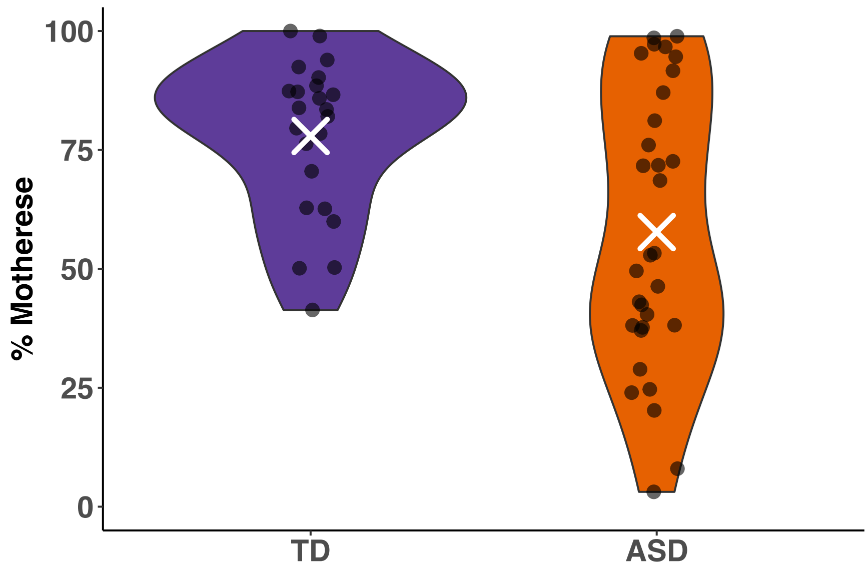
```
## [1] 37
```

```
dim(Motherese_ET[as.Date(Motherese_ET$ScanDate) < as.Date(Motherese_ET$LK_Date),])[1]
```

```
## [1] 17
```

```
# plot Motherese eye-tracking data in ASD and TD
Motherese_ET$group <- factor(Motherese_ET$group, levels = unique(Motherese_ET$group))

ggplot(Motherese_ET, aes(x = group, y = `LK_.fixation.Motherese`)) +
    geom_violin(aes(fill = group), position = "dodge",trim = T) +
    geom_point(aes(fill = group),size = 3, alpha = 0.6, position =
            position_jitterdodge(jitter.width = 0.3)) +
    stat_summary(fun = "mean", geom = "point", shape = 4, size = 7,
             color = "white",stroke = 2) +
    scale_fill_manual(values = c("#5e3c99", "#e66101")) +
    labs(y = "% Motherese", x = "") +
    guides(color = F, fill = F) +
    theme(plot.title = element_text(hjust = 0.5, size = 16, face = "bold"),
          axis.text = element_text(size = 16, face = "bold"),
          axis.title.y = element_text(size = 16, face = "bold"))+
    theme(panel.border = element_blank(),
          panel.background = element_blank(),
          panel.grid = element_blank(),
```

```
        axis.line = element_line(colour = "black")) +
    coord_cartesian(ylim=c(00, 100)) +
    scale_y_continuous(breaks = seq(0, 100, 25))
```



Difference in age between fMRI, clinical, and eye-tracking data collection

```
# read in data
Motherese_ET <- read.table(here::here("data","tidy_Motherese_ET.txt"),header = T,
               sep = "\t", stringsAsFactors = F)

dim(toddler_sample)
```

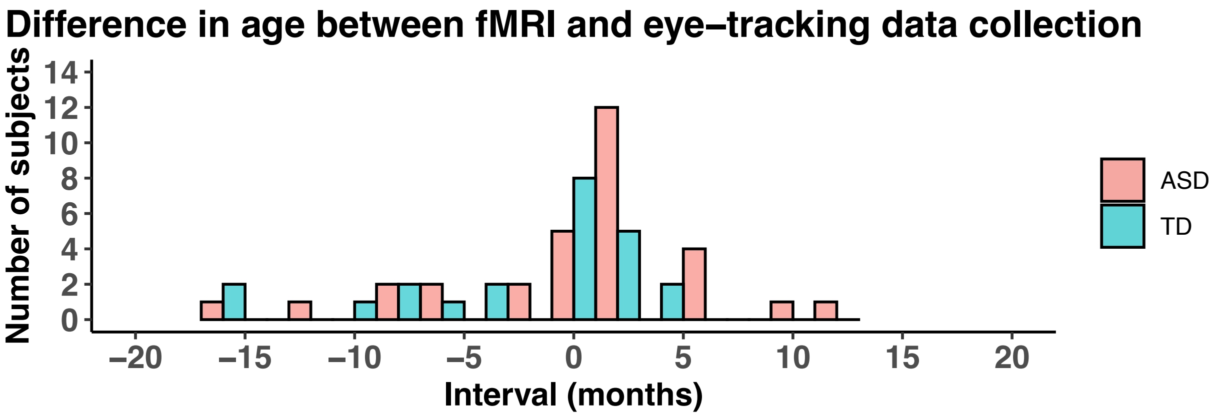
```
## [1] 71 45
```

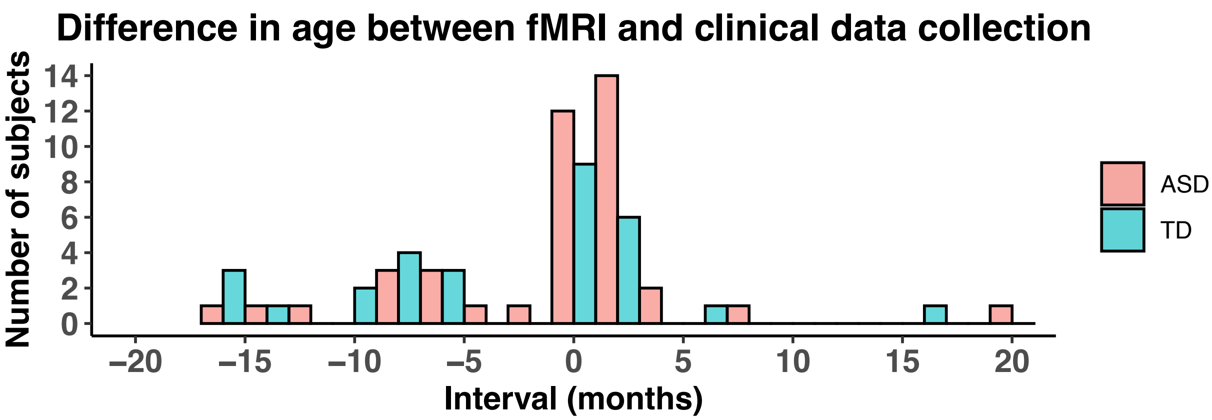
```
Motherese_ET$age_diff <- Motherese_ET$Scan_Age - Motherese_ET$ET.Age


## age distribution between fMRI scan vs. clinical test, and between fMRI scan vs. ET
toddler_sample$age_diff <- toddler_sample$scan_age - toddler_sample$test_age

p1 <- ggplot(toddler_sample, aes(age_diff,fill = group)) +
    geom_histogram(binwidth = 2, alpha = .6,position="dodge",
               color = "black") +
    #geom_density(alpha = .2, fill = "#FF6666") +
    labs(y = "Number of subjects", x = "Interval (months)",
         title = "Difference in age between fMRI and clinical data collection") +
    #guides(fill = F) +
    theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
```

```r
          axis.text = element_text(size = 12, face = "bold"),
          axis.title = element_text(size = 12, face = "bold")) +
    theme(legend.title = element_blank(),
          panel.border = element_blank(),
          panel.background = element_blank(),
          panel.grid = element_blank(),
          axis.line = element_line(colour = "black")) +
    coord_cartesian(xlim=c(-20, 20), ylim = c(0,14)) +
    scale_x_continuous(breaks = seq(-20, 20, 5)) +
    scale_y_continuous(breaks = seq(0, 14, 2))

p2 <- ggplot(Motherese_ET, aes(age_diff, fill = group)) +
    geom_histogram(binwidth = 2, alpha = .6,
               color = "black",position="dodge") +
    #geom_density(alpha = .2, fill = "#FF6666") +
    labs(y = "Number of subjects", x = "Interval (months)",
         title = "Difference in age between fMRI and eye-tracking data collection") +
    #guides(color = F, fill = F) +
    theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
          axis.text = element_text(size = 12, face = "bold"),
          axis.title = element_text(size = 12, face = "bold")) +
    theme(legend.title = element_blank(),
          panel.border = element_blank(),
          panel.background = element_blank(),
          panel.grid = element_blank(),
          axis.line = element_line(colour = "black")) +
    coord_cartesian(xlim=c(-20, 20), ylim = c(0,14)) +
    scale_x_continuous(breaks = seq(-20, 20, 5)) +
    scale_y_continuous(breaks = seq(0, 14, 2))



 library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
grid.arrange(p2,p1,nrow = 2)
```


Difference in age between fMRI and eye−tracking data collection
Number of subjects
Interval (months)
ASD
TD


Difference in age between fMRI and clinical data collection
Number of subjects
Interval (months)
ASD
TD

Characteristics of fMRI paradigms

Setup

```
# load packages
packages <- c("here","dplyr","openxlsx","tidyverse")
lapply(packages, library, character.only = TRUE)
```

Load data

```
Frequency <- xlsx::read.xlsx(here("data", "Characteristics_MRI_Paradigms.xlsx"), sheetIndex = 1)

Beats <- xlsx::read.xlsx(here("data", "Characteristics_MRI_Paradigms.xlsx"), sheetIndex = 2)

Frequency <- as.data.frame(Frequency)
Beats <- as.data.frame(Beats)
```

Mean, sd, and range

```
# peak frequency
frequency_sum <- cbind(rbind(paste0(round(mean(Frequency$Story.language[1:4])),
           "(",round(sd(Frequency$Story.language[1:4])),")"),
        paste0(round(mean(Frequency$Karen.language[1:18])),
              "(",round(sd(Frequency$Karen.language[1:18])),")"),
        paste0(round(mean(Frequency$Motherese[1:12])), "(",
            round(sd(Frequency$Motherese[1:12])),")")),
      rbind(paste0(round(min(Frequency$Story.language[1:4])), "-",
            round(max(Frequency$Story.language[1:4]))),
            paste0(round(min(Frequency$Karen.language[1:18])), "-",
                   round(max(Frequency$Karen.language[1:18]))),
            paste0(round(min(Frequency$Motherese[1:12])), "-",
                   round(max(Frequency$Motherese[1:12])))))

colnames(frequency_sum) <- c("mean (sd)", "range")
rownames(frequency_sum) <- c("Mild affect speech", "Moderate affect speech", "Motherese")

knitr::kable(frequency_sum)
```

	mean (sd)	range
Mild affect speech	275(35)	258-328
Moderate affect speech	236(41)	211-375
Motherese	354(67)	258-469

```
# beats per minutes
beats_sum <- cbind(rbind(paste0(round(mean(Beats$Story.language[1:4])),
           "(",round(sd(Beats$Story.language[1:4])),")"),
        paste0(round(mean(Beats$Karen.language[1:18])),
               "(",round(sd(Beats$Karen.language[1:18])),")"),
        paste0(round(mean(Beats$Motherese[1:12])), "(",
            round(sd(Beats$Motherese[1:12])),")")),
      rbind(paste0(round(min(Beats$Story.language[1:4])), "-",
            round(max(Beats$Story.language[1:4]))),
            paste0(round(min(Beats$Karen.language[1:18])), "-",
                   round(max(Beats$Karen.language[1:18]))),
            paste0(round(min(Beats$Motherese[1:12])), "-",
                   round(max(Beats$Motherese[1:12])))))

colnames(beats_sum) <- c("mean (sd)", "range")
rownames(beats_sum) <- c("Mild affect speech", "Moderate affect speech", "Motherese")

knitr::kable(beats_sum)
```

	mean (sd)	range
Mild affect speech	60(21)	44-88
Moderate affect speech	77(27)	20-119
Motherese	59(21)	20-93

Affect level tests for three language paradigms

Setup

```r
# load packages

packages <- c("here","dplyr","ggplot2","openxlsx","tidyverse","ggpubr")
lapply(packages, library, character.only = TRUE)
```

Load data from two surveys

```r
Survey1 <- xlsx::read.xlsx(here("data", "AffectLevels_testing.xlsx"), sheetIndex = 1)
head(Survey1)
```

```
##   Subject  Paradigm    Score
## 1       1 Motherese 4.333333
## 2       2 Motherese 3.333333
## 3       3 Motherese 4.583333
## 4       4 Motherese 4.416667
## 5       5 Motherese 4.000000
## 6       6 Motherese 4.750000
```

```r
Survey2 <- xlsx::read.xlsx(here("data", "AffectLevels_testing.xlsx"), sheetIndex = 2)
head(Survey2)
```

```
##   Subject  Paradigm    Score
## 1       1 Motherese 2.944444
## 2       2 Motherese 3.000000
## 3       3 Motherese 3.000000
## 4       4 Motherese 3.000000
## 5       5 Motherese 3.000000
## 6       6 Motherese 3.000000
```

```r
Survey1 <- as.data.frame(Survey1)
Survey2 <- as.data.frame(Survey2)
```

Comparisons in affect levels between language paradigms

```r
# Survey 1
Survey1$Paradigm <- factor(Survey1$Paradigm, levels = c("Motherese", "Karen_Lang", "Story_Lang"))

Survey1 %>%
  group_by(Paradigm) %>%
  summarise_at(vars(Score), list(mean = mean, sd = sd))
```

```
## # A tibble: 3 x 3
##   Paradigm    mean    sd
##   <fct>      <dbl> <dbl>
## 1 Motherese   4.32 0.451
## 2 Karen_Lang  2.50 0.454
## 3 Story_Lang  1.45 0.598
```

```r
res.aov <- aov(Score ~ Paradigm, data = Survey1)
anova(res.aov)
```

```
## Analysis of Variance Table
##
## Response: Score
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Paradigm   2 79.977  39.989   156.3 < 2.2e-16 ***
## Residuals 54 13.816   0.256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(res.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Score ~ Paradigm, data = Survey1)
##
## $Paradigm
##                            diff       lwr        upr p adj
## Karen_Lang-Motherese  -1.812521 -2.208020 -1.4170227 0e+00
## Story_Lang-Motherese  -2.868421 -3.263920 -2.4729223 0e+00
## Story_Lang-Karen_Lang -1.055900 -1.451398 -0.6604008 1e-07
```

```r
Survey1_compare_means <- compare_means(Score ~ Paradigm, data = Survey1,
                       method = "t.test", paired = TRUE,
                       p.adjust.method = "fdr")

data <- Survey1[Survey1$Paradigm !="Story_Lang",]
data$Paradigm <- droplevels(data$Paradigm )
comparison1 <- t.test(Score ~ Paradigm, data = data, paired = TRUE)
eff1 <- effsize::cohen.d(data$Score, data$Paradigm, paired=TRUE)

data <- Survey1[Survey1$Paradigm !="Karen_Lang",]
data$Paradigm <- droplevels(data$Paradigm )
eff2 <- effsize::cohen.d(data$Score, data$Paradigm, paired=TRUE)
```

```r
comparison2 <- t.test(Score ~ Paradigm, data = data, paired = TRUE)

data <- Survey1[Survey1$Paradigm !="Motherese",]
data$Paradigm <- droplevels(data$Paradigm )
eff3 <- effsize::cohen.d(data$Score, data$Paradigm, paired=TRUE)
comparison3 <- t.test(Score ~ Paradigm, data = data, paired = TRUE)

summary_Survey1 <- cbind.data.frame(contrast = c("MotheresevsKaren_Lang","MotheresevsStory_Lang",
                         "Karen_LangvsStory_Lang"),
                         df = rep(length(Survey1$Subject[!duplicated(Survey1$Subject)])-1,3),
                    t = c(comparison1$statistic, comparison2$statistic,
                      comparison3$statistic),
                    Survey1_compare_means$p.format, Survey1_compare_means$p.adj,
                      d = c(eff1$estimate,eff2$estimate,eff3$estimate))

colnames(summary_Survey1)[4:5] <- c("p-value","p-adjusted")
summary_Survey1
```

```
##                 contrast df        t p-value p-adjusted        d
## 1  MotheresevsKaren_Lang 18 20.52712 6.1e-14    1.2e-13 4.006231
## 2  MotheresevsStory_Lang 18 20.20427 8.1e-14    1.2e-13 5.362117
## 3 Karen_LangvsStory_Lang 18 11.74070 7.2e-10    7.2e-10 1.883057
```

```r
# Survey 2
Survey2$Paradigm <- factor(Survey2$Paradigm, levels = c("Motherese", "Karen_Lang", "Story_Lang"))

Survey2 %>%
  group_by(Paradigm) %>%
  summarise_at(vars(Score), list(mean = mean, sd = sd))
```

```
## # A tibble: 3 x 3
##   Paradigm    mean     sd
##   <fct>      <dbl>  <dbl>
## 1 Motherese   2.99 0.0330
## 2 Karen_Lang  1.92 0.0780
## 3 Story_Lang  1.09 0.0831
```

```r
res.aov <- aov(Score ~ Paradigm, data = Survey2)
anova(res.aov)
```

```
## Analysis of Variance Table
##
## Response: Score
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Paradigm   2 27.0004 13.5002  2876.5 < 2.2e-16 ***
## Residuals 42  0.1971  0.0047
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(res.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Score ~ Paradigm, data = Survey2)
##
## $Paradigm
##                               diff        lwr        upr p adj
## Karen_Lang-Motherese  -1.0629630 -1.1237380 -1.0021879     0
## Story_Lang-Motherese  -1.8925926 -1.9533676 -1.8318176     0
## Story_Lang-Karen_Lang -0.8296296 -0.8904047 -0.7688546     0
```

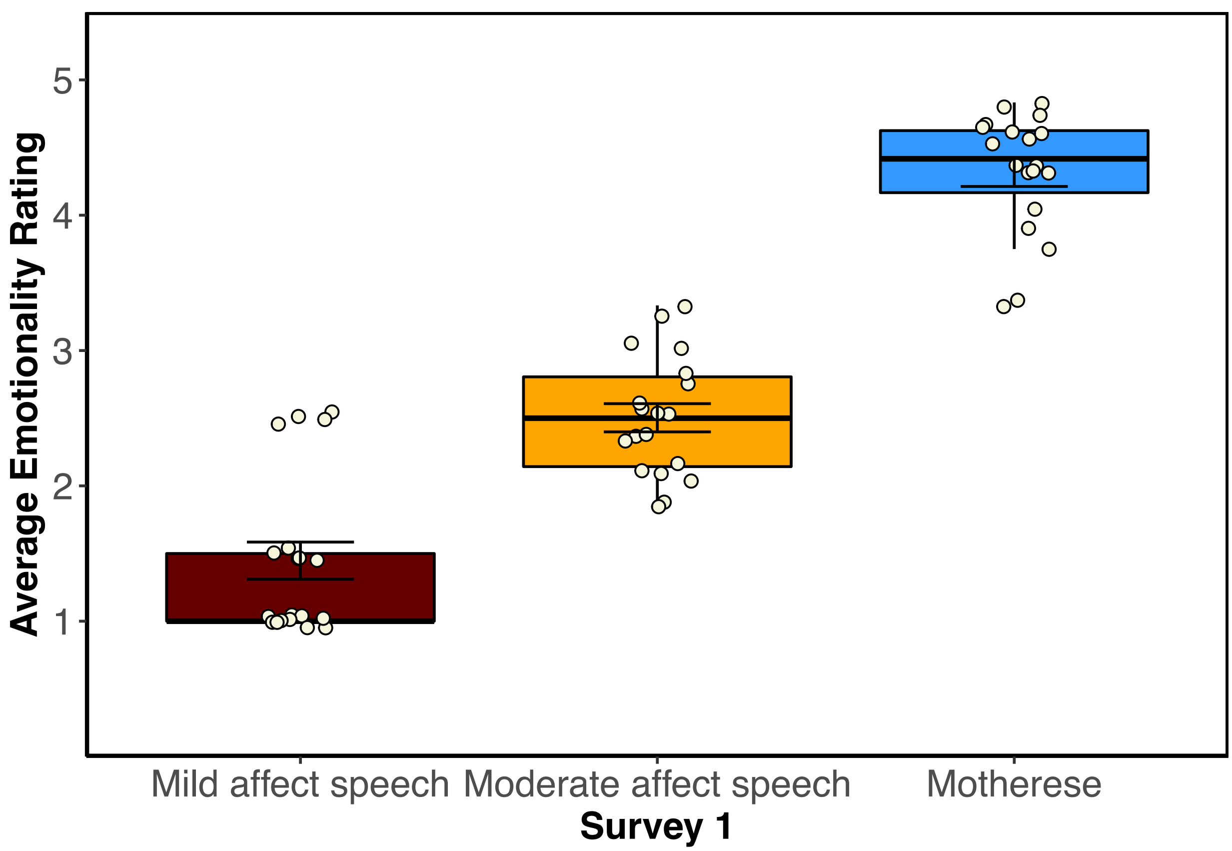
```
Survey2_compare_means <- compare_means(Score ~ Paradigm, data = Survey2, method = "t.test",
                       paired = TRUE, p.adjust.method = "fdr")

data <- Survey2[Survey2$Paradigm !="Story_Lang",]
data$Paradigm <- droplevels(data$Paradigm )
comparison1 <- t.test(Score ~ Paradigm, data = data, paired = TRUE)
eff1 <- effsize::cohen.d(data$Score, data$Paradigm, paired=TRUE)

data <- Survey2[Survey2$Paradigm !="Karen_Lang",]
data$Paradigm <- droplevels(data$Paradigm )
eff2 <- effsize::cohen.d(data$Score, data$Paradigm, paired=TRUE)
comparison2 <- t.test(Score ~ Paradigm, data = data, paired = TRUE)

data <- Survey2[Survey2$Paradigm !="Motherese",]
data$Paradigm <- droplevels(data$Paradigm )
eff3 <- effsize::cohen.d(data$Score, data$Paradigm, paired=TRUE)
comparison3 <- t.test(Score ~ Paradigm, data = data, paired = TRUE)

summary_Survey2 <- cbind.data.frame(contrast = c("MotheresevsKaren_Lang","MotheresevsStory_Lang",
                       "Karen_LangvsStory_Lang"),
                       df = rep(length(Survey2$Subject[!duplicated(Survey2$Subject)])-1,3),
                    t = c(comparison1$statistic, comparison2$statistic,
                      comparison3$statistic),
                    Survey2_compare_means$p.format, Survey2_compare_means$p.adj,
                      d = c(eff1$estimate,eff2$estimate,eff3$estimate))

colnames(summary_Survey2)[4:5] <- c("p-value","p-adjusted")
summary_Survey2
```

```
##                 contrast df        t p-value p-adjusted        d
## 1  MotheresevsKaren_Lang 14 47.73871 < 2e-16    1.0e-16 17.87426
## 2  MotheresevsStory_Lang 14 73.64698 < 2e-16    4.7e-19 31.22883
## 3 Karen_LangvsStory_Lang 14 20.36364 8.4e-12    8.4e-12 10.29827
```

Boxplots for surveys 1 and 2

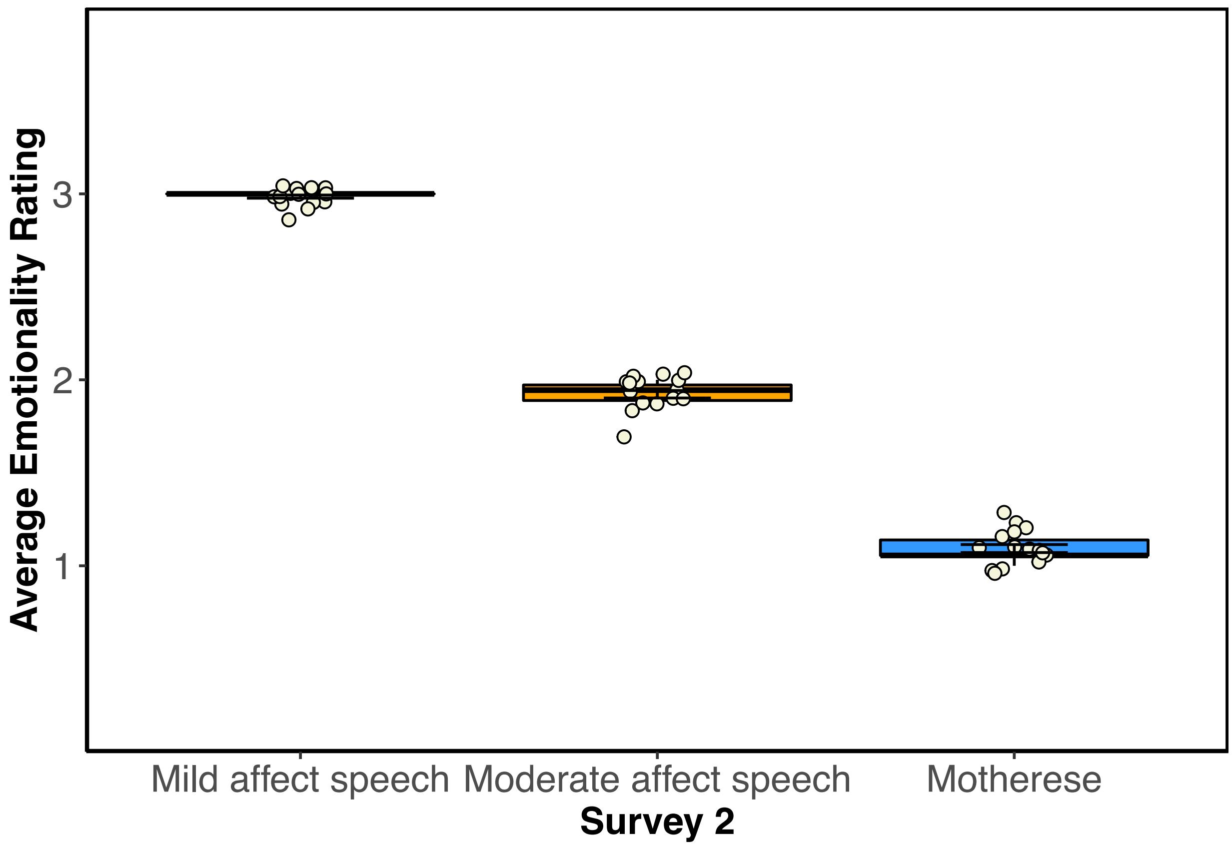
```
# Survey 1
Survey1$Paradigm <- factor(Survey1$Paradigm, levels = c("Story_Lang", "Karen_Lang","Motherese"))

ggplot(Survey1,aes(x=Paradigm, y=Score), fill=Paradigm) +
```

```r
    geom_boxplot(fill=c("#660000","orange","#3399FF"), colour="black", outlier.shape = NA) +
    geom_jitter(size=2, position = position_jitter(width=0.1, height=0.05),
        shape=21, colour="black", fill="beige") +
    stat_summary(geom = "errorbar", fun.data = mean_se,
         position = position_dodge(width=0.65), width=0.3) +
    labs(x = "Survey 1", y = "Average Emotionality Rating") +
    scale_y_continuous(expand = c(0,0), limits=c(0,5.5), breaks=c(1,2,3,4,5)) +
    scale_x_discrete(labels =  c("Mild affect speech", "Moderate affect speech","Motherese")) +
    theme(axis.text=element_text(size=14),
        axis.title=element_text(size=14,face="bold"),
        axis.line = element_line(colour = "black"),
        panel.border = element_rect(colour = "black", fill=NA, size=1),
        panel.background = element_blank())
```




```r
# Survey 2
Survey1$Paradigm <- factor(Survey1$Paradigm, levels = c("Story_Lang", "Karen_Lang","Motherese"))
ggplot(Survey2,aes(x=Paradigm, y=Score), fill=Paradigm) +
    geom_boxplot(fill=c("#660000","orange","#3399FF"), colour="black", outlier.shape = NA) +
    geom_jitter(size=2, position = position_jitter(width=0.1, height=0.05),
        shape=21, colour="black", fill="beige") +
  stat_summary(geom = "errorbar", fun.data = mean_se,
         position = position_dodge(width=0.65), width=0.3) +
    labs(x = "Survey 2", y = "Average Emotionality Rating") +
    scale_y_continuous(expand = c(0,0), limits=c(0,4), breaks=c(1,2,3)) +
    scale_x_discrete(labels =  c("Mild affect speech", "Moderate affect speech","Motherese")) +
```

```
    theme(axis.text=element_text(size=14),
        axis.title=element_text(size=14,face="bold"),
        axis.line = element_line(colour = "black"),
        panel.border = element_rect(colour = "black", fill=NA, size=1),
        panel.background = element_blank())
```



ROI Analysis

Setup

```
# load packages
packages <- c("here","dplyr","WGCNA","factoextra","ggplot2","effsize","data.table",
          "lme4","lmerTest","psych","FSA","irr","lpSolve")
source(here::here("code","Mods2table.R"))
source(here::here("code","ROI_psc_plot.R"))
source(here::here("code","ROI_behavior_plot.R"))
source(here::here("code","test_retest_plot.R"))
lapply(packages, library, character.only = TRUE)
```

Read in toddler data

```
# read in data
tidy_fMRI_clinical_toddlers <- read.table(here::here("data", "tidy_fMRI_clinical_toddlers.txt"),
                 header = T, sep = "\t",stringsAsFactors = F)

# scans for each language paradigm
Story_scans <- tidy_fMRI_clinical_toddlers[!is.na(tidy_fMRI_clinical_toddlers$Story_Lang),]
Karen_scans <- tidy_fMRI_clinical_toddlers[!is.na(tidy_fMRI_clinical_toddlers$Karen_Lang),]
Motherese_scans <- tidy_fMRI_clinical_toddlers[!is.na(tidy_fMRI_clinical_toddlers$Motherese),]
```

Read in adult data

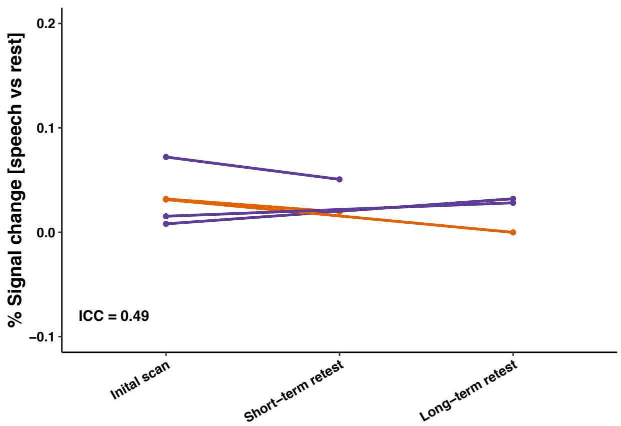
```
# read in data
adult_scans <- read.table(here::here("data", "tidy_fMRI_adults.txt"),
                 header = T, sep = "\t",stringsAsFactors = F)

# scans for each language paradigm
Story_scans_adult <- adult_scans[!is.na(adult_scans$Story_Lang),]
Karen_scans_adult <- adult_scans[!is.na(adult_scans$Karen_Lang),]
Motherese_scans_adult <- adult_scans[!is.na(adult_scans$Motherese),]
```

Plots for test-retest percent siginal change in each language paradigm

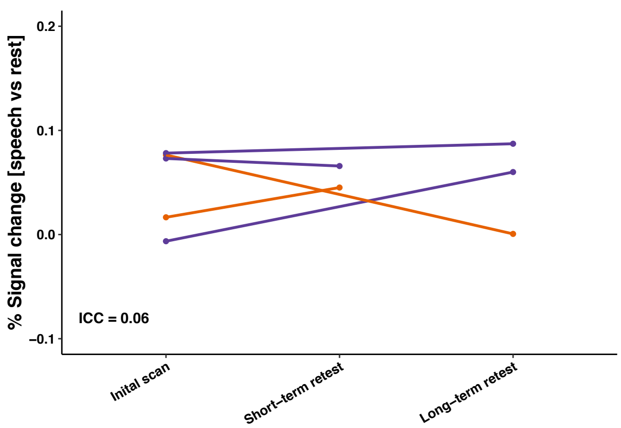
```r
# organize data file
Story_tmp <- Story_scans$subjid[duplicated(Story_scans$subjid)]
Story_retest <- Story_scans[Story_scans$subjid %in% Story_tmp,
colnames(Story_retest)[4:5] <- c("LHtemporal_psc","RHtemporal_psc")
Story_retest$task <- "Story_Lang"
Story_retest$grp <- rep(1:length(Story_tmp), each=2)

Karen_tmp <- Karen_scans$subjid[duplicated(Karen_scans$subjid)]
Karen_retest <- Karen_scans[Karen_scans$subjid %in% Karen_tmp,
        c("subjid","group","scan_age","Karen_LHtemporal_psc","Karen_RHtemporal_psc")]
colnames(Karen_retest)[4:5] <- c("LHtemporal_psc","RHtemporal_psc")
Karen_retest$task <- "Karen_Lang"
Karen_retest$grp <- rep(1:length(Karen_tmp), each=2)

Motherese_tmp <- Motherese_scans$subjid[duplicated(Motherese_scans$subjid)]
Motherese_retest <- Motherese_scans[Motherese_scans$subjid %in% Motherese_tmp,
    c("subjid","group","scan_age","Motherese_LHtemporal_psc","Motherese_RHtemporal_psc")]
colnames(Motherese_retest)[4:5] <- c("LHtemporal_psc","RHtemporal_psc")
Motherese_retest$task <- "Motherese"
Motherese_retest$grp <- rep(1:length(Motherese_tmp), each=2)

combined_retest <- rbind.data.frame(Story_retest,Karen_retest,Motherese_retest)

# add test-retest scan interval
for (i in seq(2, length(combined_retest$subjid), 2)) {
    combined_retest$interval[i-1] <- "Inital scan"
    combined_retest$interval[i] <- combined_retest$scan_age[i] -
        combined_retest$scan_age[i-1]
}

Summarize(as.numeric(combined_retest$interval[seq(2, length(combined_retest$subjid), 2)]),
      digits = 2)
```

```
##      n   mean     sd    min     Q1 median     Q3    max
##  20.00   7.60   5.58   1.00   3.50   4.00  13.00  15.00
```

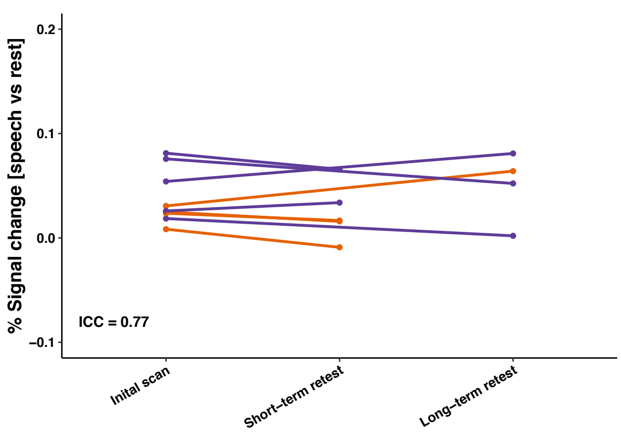
```r
# group retest scans into short-term and long-term scans
combined_retest$scan_group <- combined_retest$interval
combined_retest$scan_group[which(as.numeric(combined_retest$interval) <=4)] <- "Short-term retest"
combined_retest$scan_group[which(as.numeric(combined_retest$interval) >4)] <- "Long-term retest"

# plot line graphs with intraclass correlation coefficients
combined_retest$scan_group <- factor(combined_retest$scan_group,
                     levels = c("Inital scan", "Short-term retest",
                           "Long-term retest"))

combined_retest$group <- as.factor(combined_retest$group)
test_retest_plot(combined_retest, "Story_Lang", "LHtemporal_psc")
```



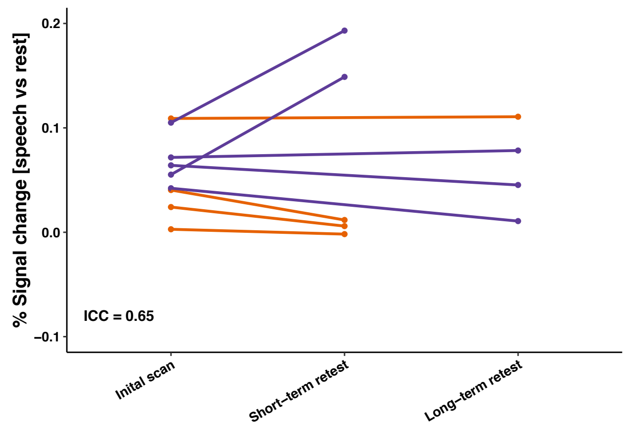

```
test_retest_plot(combined_retest, "Story_Lang", "RHtemporal_psc")
```



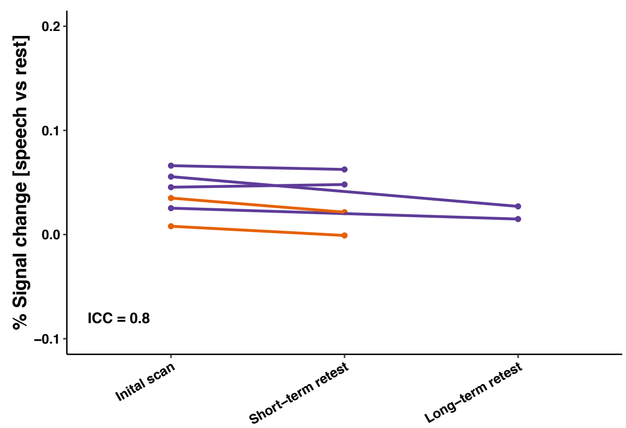

```
test_retest_plot(combined_retest, "Karen_Lang", "LHtemporal_psc")
```



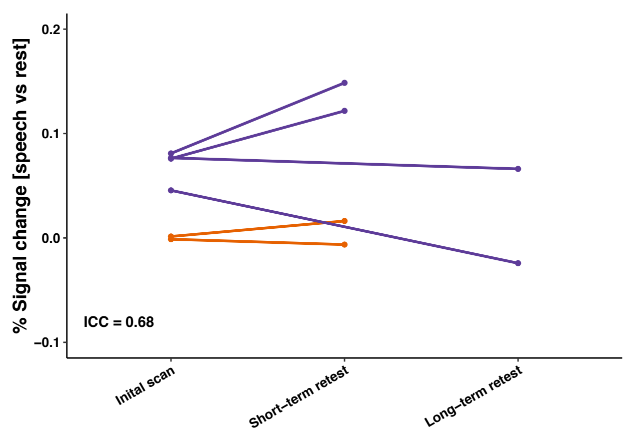

```
test_retest_plot(combined_retest, "Karen_Lang", "RHtemporal_psc")
```




```r
test_retest_plot(combined_retest, "Motherese", "LHtemporal_psc")
```




```r
test_retest_plot(combined_retest, "Motherese", "RHtemporal_psc")
```



Percent signal change in TD and ASD across three language paradigms

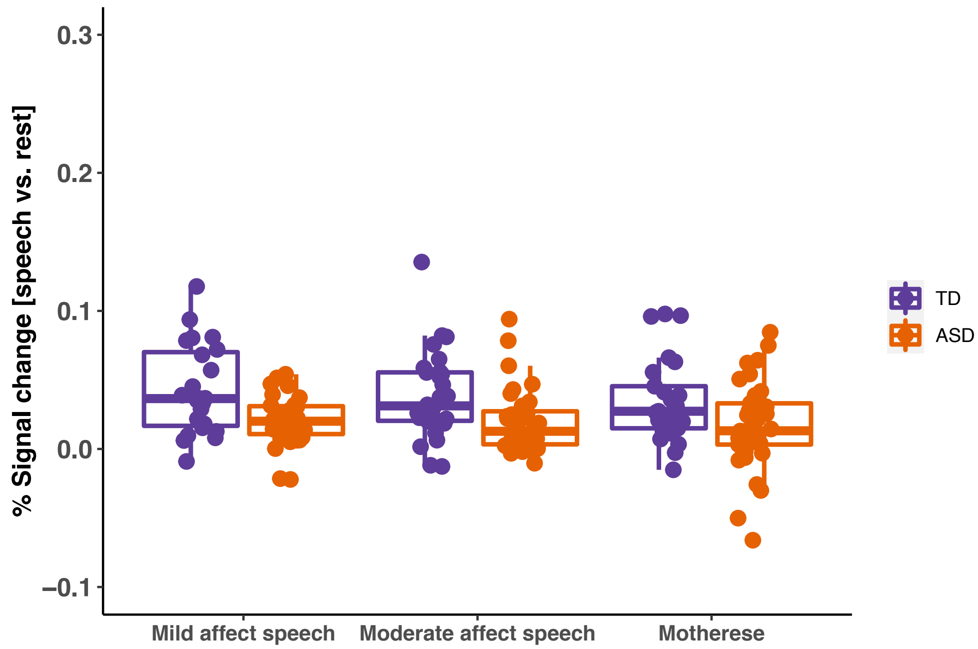
```r
# organize data file
Story_psc <- Story_scans[,c("subjid","scan_age","group","Story_LHtemporal_psc",
                "Story_RHtemporal_psc")]
colnames(Story_psc)[4:5] <- c("LHtemporal_psc","RHtemporal_psc")
Story_psc$task <- "Mild affect speech"

Story_psc <- Story_psc[!duplicated(Story_psc$subjid), ]

Karen_psc <- Karen_scans[,c("subjid","scan_age","group","Karen_LHtemporal_psc",
                "Karen_RHtemporal_psc")]
colnames(Karen_psc)[4:5] <- c("LHtemporal_psc","RHtemporal_psc")
Karen_psc$task <- "Moderate affect speech"

Karen_psc <- Karen_psc[!duplicated(Karen_psc$subjid), ]

Motherese_psc <- Motherese_scans[,c("subjid","scan_age","group","Motherese_LHtemporal_psc",
                    "Motherese_RHtemporal_psc")]
colnames(Motherese_psc)[4:5] <- c("LHtemporal_psc","RHtemporal_psc")
Motherese_psc$task <- "Motherese"

Motherese_psc <- Motherese_psc[!duplicated(Motherese_psc$subjid), ]

combined_psc <- rbind.data.frame(Story_psc, Karen_psc, Motherese_psc)

combined_psc$task <- factor(combined_psc$task, levels = unique(combined_psc$task))
combined_psc$group <- factor(combined_psc$group, levels = unique(combined_psc$group))

aggregate(LHtemporal_psc ~ group + task, FUN=mean, data=combined_psc)
```

```
##   group                   task LHtemporal_psc
## 1    TD     Mild affect speech     0.04286353
## 2   ASD     Mild affect speech     0.02030916
## 3    TD Moderate affect speech     0.03839828
## 4   ASD Moderate affect speech     0.01992519
## 5    TD              Motherese     0.03516916
## 6   ASD              Motherese     0.01691953
```

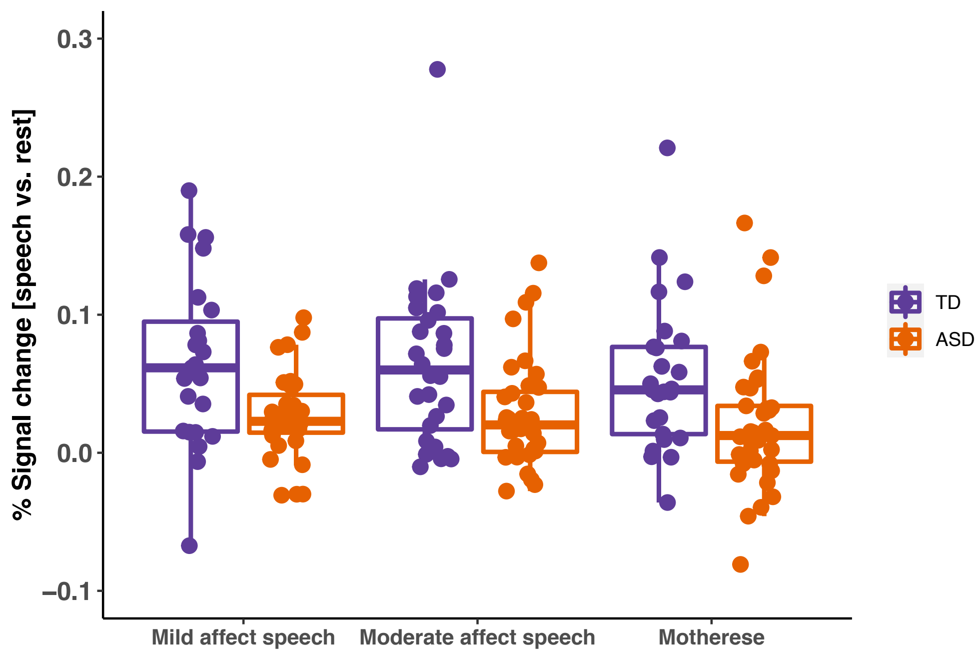
```r
dim(combined_psc)
```

```
## [1] 180   6
```

```r
# boxplots
ROI_psc_plot(combined_psc, "LHtemporal_psc", "TDvsASD")
```




```r
ROI_psc_plot(combined_psc, "RHtemporal_psc", "TDvsASD")
```




```r
# t-tests and effect sizes
tasks <- c("Mild affect speech", "Moderate affect speech", "Motherese")

es_mat <- matrix(1:12,nrow = 3, ncol = 4)
rownames(es_mat) <- c("Mild affect speech", "Moderate affect speech", "Motherese")
colnames(es_mat) <- c("Left temporal", "Right temporal","Left temporal", "Right temporal")
```

```r
i <- 0
for (task in tasks) {
    i <- i + 1
    aa <- effsize::cohen.d(combined_psc[combined_psc$task == task, "LHtemporal_psc"],
                     combined_psc[combined_psc$task == task, "group"],
                     pooled = T)

    bb <- effsize::cohen.d(combined_psc[combined_psc$task == task, "RHtemporal_psc"],
                     combined_psc[combined_psc$task == task, "group"],
                     pooled = T)

    tt1 <- t.test(combined_psc[combined_psc$task == task & combined_psc$group == "TD",
                   "LHtemporal_psc"],
              combined_psc[combined_psc$task == task & combined_psc$group == "ASD",
                     "LHtemporal_psc"])

    tt2 <- t.test(combined_psc[combined_psc$task == task & combined_psc$group == "TD",
                   "RHtemporal_psc"],
              combined_psc[combined_psc$task == task & combined_psc$group == "ASD",
                     "RHtemporal_psc"])

    es_mat[i, 1] <- round(abs(aa$estimate),2)
    es_mat[i, 2] <- round(abs(bb$estimate),2)
    es_mat[i, 3] <- round(abs(tt1$p.value),3)
    es_mat[i, 4] <- round(abs(tt2$p.value),3)
}

# t-values
knitr::kable(es_mat[,1:2])
```

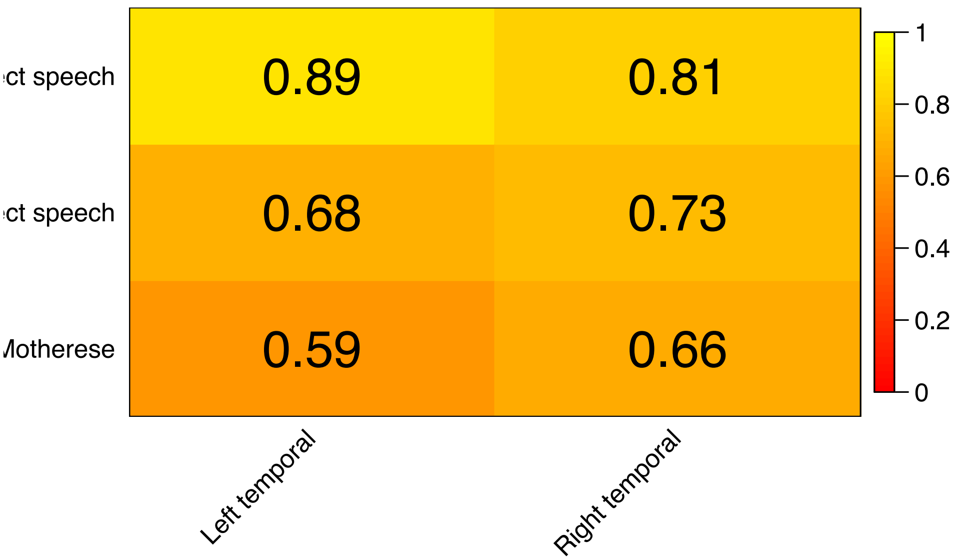
	Left temporal	Right temporal
Mild affect speech	0.89	0.81
Moderate affect speech	0.68	0.73
Motherese	0.59	0.66

```r
# p-values
knitr::kable(es_mat[,3:4])
```

	Left temporal	Right temporal
Mild affect speech	0.005	0.011
Moderate affect speech	0.012	0.009
Motherese	0.026	0.017

```r
# plot the effect size matrix as a heatmap
eff <- es_mat[,1:2]
colfunc <- colorRampPalette(c("red", "yellow"))

WGCNA::labeledHeatmap(Matrix = eff, xLabels = colnames(eff),
              yLabels = rownames(eff), ySymbols = NULL, colorLabels = F,
```

```
              colors = colfunc(50), textMatrix = round(eff, digits = 2),
              setStdMargins = F, cex.text = 2, zlim = c(0, 1))
```



Group differences in percent signal change between TD toddlers vs adults

```
# organize data file: TD toddler
combined_psc_TD <- combined_psc[combined_psc$group == "TD",
                c("subjid","LHtemporal_psc","RHtemporal_psc", "task")]
combined_psc_TD$group <- "TD Toddlers"

# organize data file: adults
Story_adult_psc <- Story_scans_adult[, c("fMRI_Subj","Subj","Story_LHtemporal_psc",
                     "Story_RHtemporal_psc")]
colnames(Story_adult_psc)[3:4] <- c("LHtemporal_psc","RHtemporal_psc")
Story_adult_psc$task <- "Mild affect speech"

Story_adult_psc <- Story_adult_psc[!duplicated(Story_adult_psc$Subj),]

Karen_adult_psc <- Karen_scans_adult[, c("fMRI_Subj","Subj","Karen_LHtemporal_psc",
                     "Karen_RHtemporal_psc")]
colnames(Karen_adult_psc)[3:4] <- c("LHtemporal_psc","RHtemporal_psc")
Karen_adult_psc$task <- "Moderate affect speech"

Karen_adult_psc <- Karen_adult_psc[!duplicated(Karen_adult_psc$Subj),]

Motherese_adult_psc <- Motherese_scans_adult[,c("fMRI_Subj","Subj","Motherese_LHtemporal_psc",
```

```r
                        "Motherese_RHtemporal_psc")]
colnames(Motherese_adult_psc)[3:4] <- c("LHtemporal_psc","RHtemporal_psc")
Motherese_adult_psc$task <- "Motherese"

Motherese_adult_psc <- Motherese_adult_psc[!duplicated(Motherese_adult_psc$Subj),]

combined_psc_adults <- rbind.data.frame(Story_adult_psc, Karen_adult_psc, Motherese_adult_psc)
colnames(combined_psc_adults)
```

```
## [1] "fMRI_Subj"      "Subj"           "LHtemporal_psc" "RHtemporal_psc"
## [5] "task"
```

```r
combined_psc_adults$group <- "TD Adults"

mean(combined_psc_adults$LHtemporal_psc[combined_psc_adults$task == "Mild affect speech"])
```

```
## [1] 0.103344
```

```r
mean(combined_psc_adults$RHtemporal_psc[combined_psc_adults$task == "Mild affect speech"])
```

```
## [1] 0.3077167
```

```r
mean(combined_psc_adults$LHtemporal_psc[combined_psc_adults$task == "Moderate affect speech"])
```

```
## [1] 0.1290372
```

```r
mean(combined_psc_adults$RHtemporal_psc[combined_psc_adults$task == "Moderate affect speech"])
```

```
## [1] 0.2953956
```

```r
mean(combined_psc_adults$LHtemporal_psc[combined_psc_adults$task == "Motherese"])
```

```
## [1] 0.1416974
```

```r
mean(combined_psc_adults$RHtemporal_psc[combined_psc_adults$task == "Motherese"])
```

```
## [1] 0.339525
```

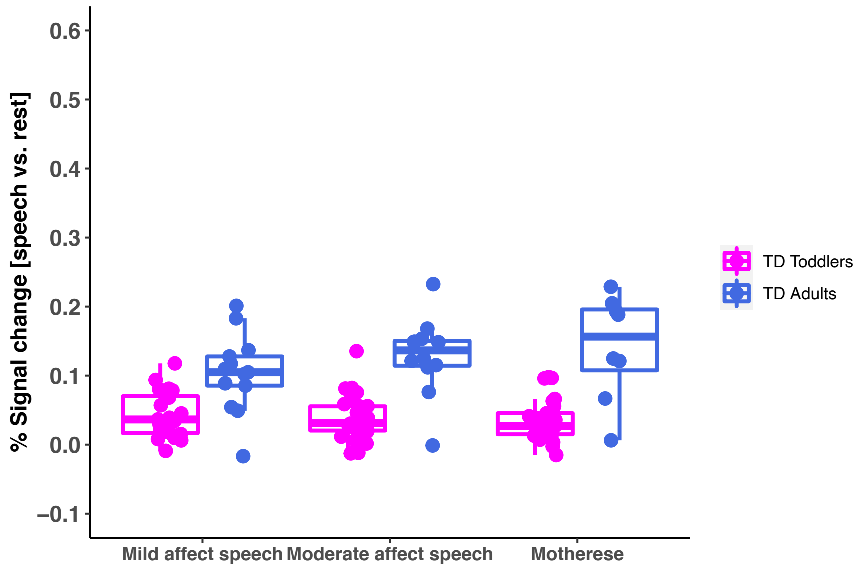
```r
# combine TD toddlers and adults
colnames(combined_psc_adults)[1] <- "subjid"
combined_psc_all <- rbind(combined_psc_TD, combined_psc_adults[,-2])
colnames(combined_psc_TD)
```

```
## [1] "subjid"         "LHtemporal_psc" "RHtemporal_psc" "task"
## [5] "group"
```

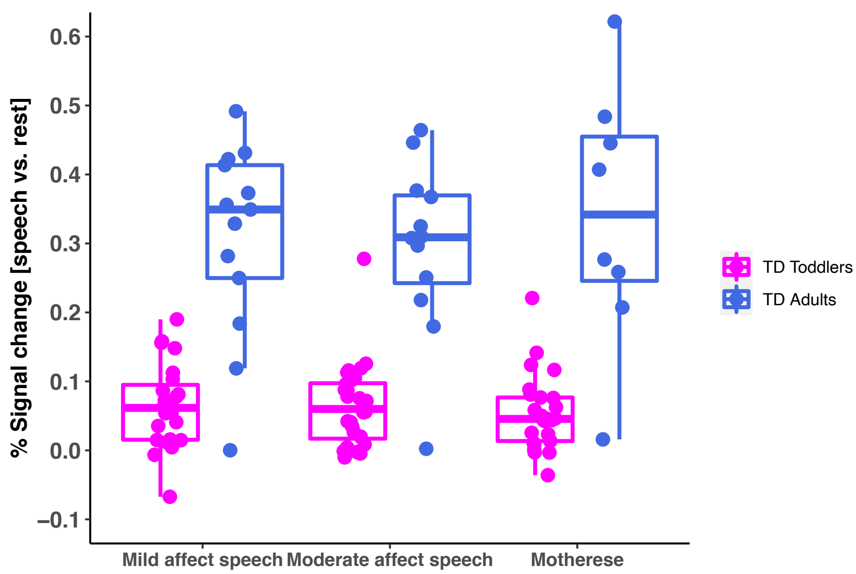
```r
colnames(combined_psc_adults[,-2])
```

```
## [1] "subjid"         "LHtemporal_psc" "RHtemporal_psc" "task"
## [5] "group"
```

```r
combined_psc_all$task <- factor(combined_psc_all$task, levels = unique(combined_psc_all$task))
combined_psc_all$group <- factor(combined_psc_all$group, levels = unique(combined_psc_all$group))

# boxplots
ROI_psc_plot(combined_psc_all, "LHtemporal_psc", "TDvsAdults")
```




```r
ROI_psc_plot(combined_psc_all, "RHtemporal_psc", "TDvsAdults")
```



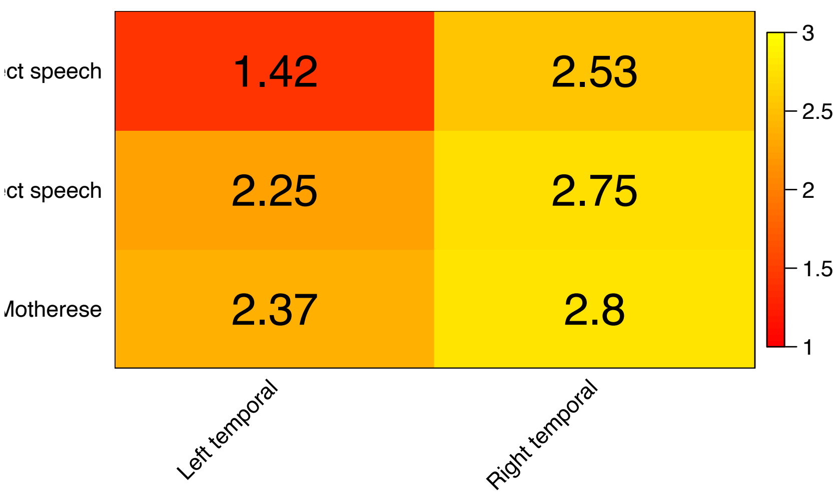
```r
## t-test and effect sizes
tasks <- c("Mild affect speech", "Moderate affect speech", "Motherese")

es_mat <- matrix(1:12,nrow = 3, ncol = 4)
rownames(es_mat) <- c("Mild affect speech", "Moderate affect speech", "Motherese")
colnames(es_mat) <- c("Left temporal", "Right temporal","Left temporal", "Right temporal")

i <- 0
for (task in tasks) {
    i <- i + 1
    aa <- effsize::cohen.d(combined_psc_all[combined_psc_all$task == task, "LHtemporal_psc"],
                     combined_psc_all[combined_psc_all$task == task, "group"],
                     pooled = T)

    bb <- effsize::cohen.d(combined_psc_all[combined_psc_all$task == task, "RHtemporal_psc"],
                     combined_psc_all[combined_psc_all$task == task, "group"],
                     pooled = T)

    tt1 <- t.test(combined_psc_all[combined_psc_all$task == task &
                          combined_psc_all$group == "TD Toddlers",
                    "LHtemporal_psc"],
              combined_psc_all[combined_psc_all$task == task &
                          combined_psc_all$group == "TD Adults",
                     "LHtemporal_psc"])

    tt2 <- t.test(combined_psc_all[combined_psc_all$task == task &
                          combined_psc_all$group == "TD Toddlers",
                    "RHtemporal_psc"],
              combined_psc_all[combined_psc_all$task == task &
                          combined_psc_all$group == "TD Adults",
                     "RHtemporal_psc"])

    es_mat[i, 1] <- round(abs(aa$estimate),2)
    es_mat[i, 2] <- round(abs(bb$estimate),2)
    es_mat[i, 3] <- round(abs(tt1$p.value),4)
    es_mat[i, 4] <- round(abs(tt2$p.value),4)
}

#  t-values
knitr::kable(es_mat[,1:2])
```

	Left temporal	Right temporal
Mild affect speech	1.42	2.53
Moderate affect speech	2.25	2.75
Motherese	2.37	2.80

```r
#  p-values
knitr::kable(es_mat[,3:4])
```

	Left temporal	Right temporal
Mild affect speech	0.0026	0.0000
Moderate affect speech	0.0001	0.0000
Motherese	0.0053	0.0035

```
# plot the effect size matrix as a heatmap
eff <- es_mat[,1:2]
colfunc <- colorRampPalette(c("red", "yellow"))

WGCNA::labeledHeatmap(Matrix = eff, xLabels = colnames(eff),
              yLabels = rownames(eff), ySymbols = NULL, colorLabels = F,
              colors = colfunc(50), textMatrix = round(eff, digits = 2),
              setStdMargins = F, cex.text = 2, zlim = c(1, 3))
```



Mixed effects model analysis

```
# organize data file
Story_datafile <- Story_scans[,c("subj","subjid","scan_age","gender","group","Story_meanFD",
        "Story_LHtemporal_psc","Story_RHtemporal_psc","final_vine_ComTotal_DomStd",
        "final_vine_SocTotal_DomStd")]
Story_datafile$task <- "Story_Lang"
colnames(Story_datafile)[6:8] <- c("meanFD","LHtemporal_psc","RHtemporal_psc")

Karen_datafile <- Karen_scans[,c("subj","subjid","scan_age","gender","group","Karen_meanFD",
            "Karen_LHtemporal_psc","Karen_RHtemporal_psc","final_vine_ComTotal_DomStd",
            "final_vine_SocTotal_DomStd")]
Karen_datafile$task <- "Karen_Lang"
colnames(Karen_datafile)[6:8] <- c("meanFD","LHtemporal_psc","RHtemporal_psc")

Motherese_datafile <- Motherese_scans[,c("subj","subjid","scan_age","gender","group",
```

```r
                       "Motherese_meanFD","Motherese_LHtemporal_psc",
                       "Motherese_RHtemporal_psc","final_vine_ComTotal_DomStd",
                       "final_vine_SocTotal_DomStd")]
Motherese_datafile$task <- "Motherese"
colnames(Motherese_datafile)[6:8] <- c("meanFD","LHtemporal_psc","RHtemporal_psc")

combined_datafile <- rbind.data.frame(Story_datafile, Karen_datafile, Motherese_datafile)

# run mixed effects models
ROIs <- c("LHtemporal_psc","RHtemporal_psc")
clins <- c("final_vine_ComTotal_DomStd", "final_vine_SocTotal_DomStd")
cnames <- c("Estimate", "Std. Error", "df","t value", "p value", "R2")

mixed_effects <- Mods2table(combined_datafile, ROIs, clins, cnames) %>%
    as.data.frame()

knitr::kable(mixed_effects)
```

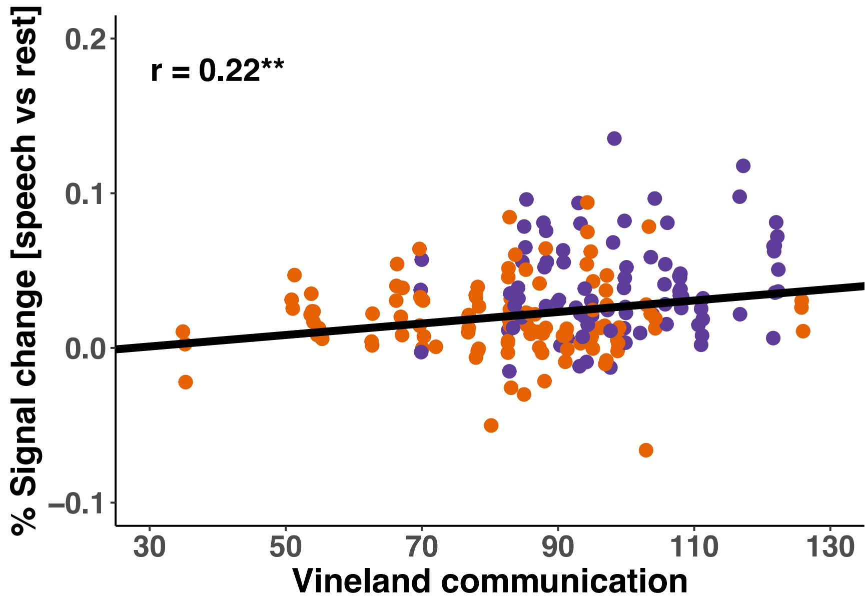
ROI	Variables	Estimate	Std. Error	df	t value	p value	R2
Left temporal	Communication scores	0.00037	0.00016	48.236	2.397	0.02	0.068
Left temporal	scan_age	-7e-05	0.00029	78.495	-0.26	0.795	0.068
Left temporal	gender	0.00707	0.00602	55.391	1.173	0.246	0.068
Left temporal	meanFD	-0.02732	0.02545	159.588	-1.073	0.285	0.068
Left temporal	Social scores	5e-04	0.00018	49.539	2.727	0.009	0.08
Left temporal	scan_age	-3e-05	0.00029	77.449	-0.106	0.916	0.08
Left temporal	gender	0.00513	0.00592	54.593	0.866	0.39	0.08
Left temporal	meanFD	-0.02635	0.02525	157.384	-1.044	0.298	0.08
Right temporal	Communication scores	0.00081	0.00032	49.888	2.577	0.013	0.094
Right temporal	scan_age	-7e-04	0.00056	90.451	-1.267	0.209	0.094
Right temporal	gender	0.00515	0.0121	55.451	0.426	0.672	0.094
Right temporal	meanFD	-0.02706	0.04586	178.257	-0.59	0.556	0.094
Right temporal	Social scores	0.00118	0.00037	50.654	3.227	0.002	0.125
Right temporal	scan_age	-0.00058	0.00055	87.498	-1.056	0.294	0.125
Right temporal	gender	0.00065	0.01174	54	0.056	0.956	0.125
Right temporal	meanFD	-0.02335	0.04528	174.624	-0.516	0.607	0.125

```r
#write.csv(mixed_effects, "SNF_results/mixed_effects.csv")

# fdr correction
p.adjust(as.numeric(Mods2table(combined_datafile, ROIs, clins, cnames)[,"p value"][c(1,5,9,13)]),
     method = "fdr")
```

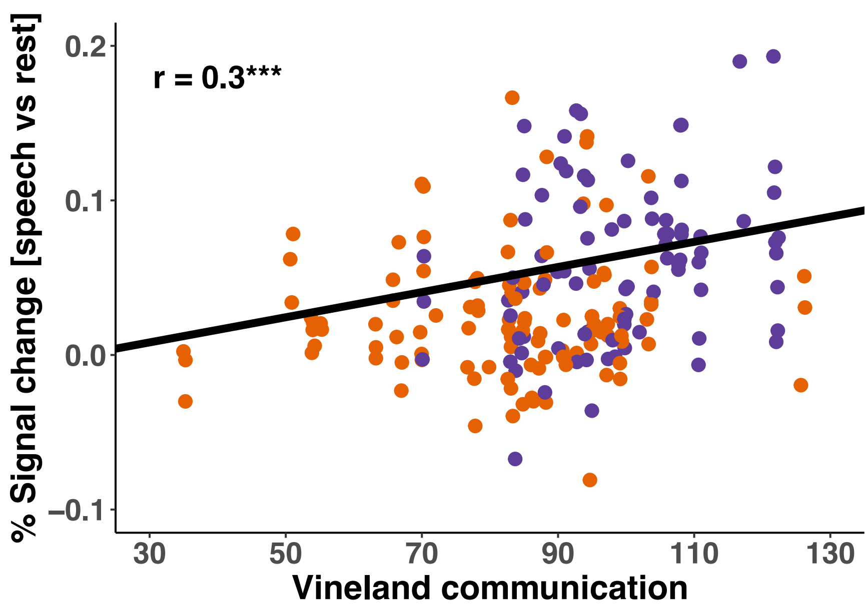
```
## [1] 0.02000000 0.01733333 0.01733333 0.00800000
```

Scatterplots: ROI activation and Vineland communication and social scores

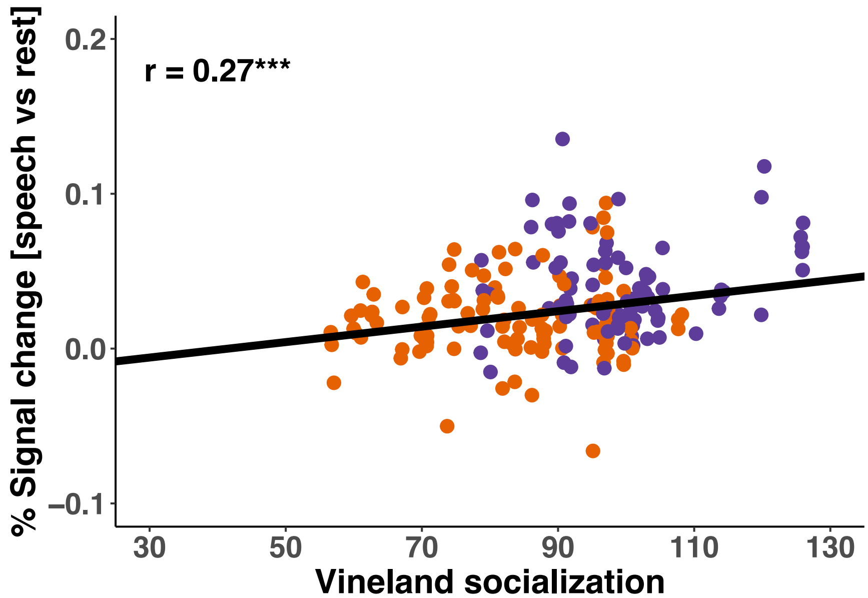
```r
dat <- combined_datafile

ROI_behavior_plot(dat, "final_vine_ComTotal_DomStd", "LHtemporal_psc",
          "Vineland communication")
```



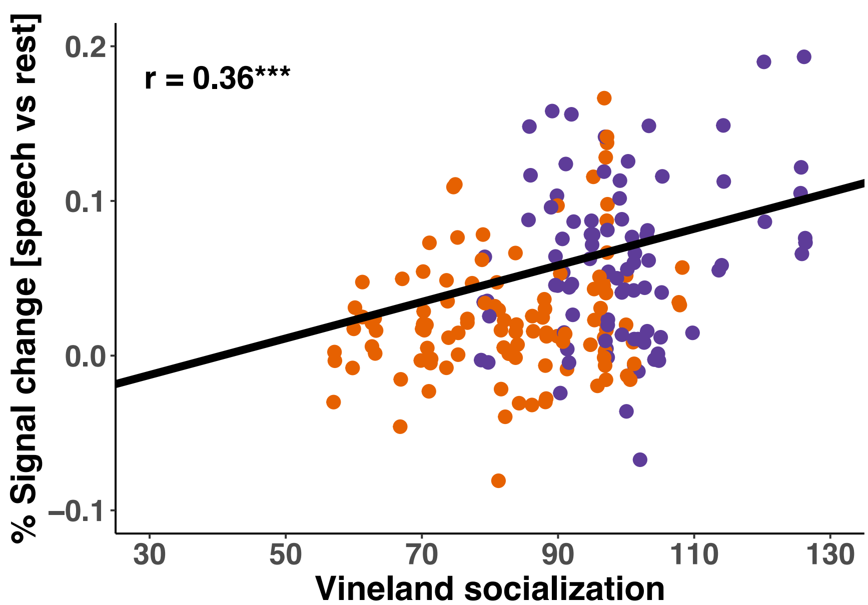

```r
ROI_behavior_plot(dat, "final_vine_ComTotal_DomStd", "RHtemporal_psc",
          "Vineland communication")
```




```r
ROI_behavior_plot(dat, "final_vine_SocTotal_DomStd", "LHtemporal_psc",
          "Vineland socialization")
```




```
ROI_behavior_plot(dat, "final_vine_SocTotal_DomStd", "RHtemporal_psc",
          "Vineland socialization")
```



SNF/Clustering and Motherese Eye-Tracking Analysis

Setup

```
# load packages
packages <- c("here","tidyr","WGCNA","dplyr","SNFtool","ggplot2","effsize","data.table",
          "psych","FSA","rcompanion","rstatix")
lapply(packages, library, character.only = TRUE)

source(here::here("code","SNF_Louvain.R"))
source(here::here("code","clusters_plot.R"))
source(here::here("code","ET_clusters.R"))
set.seed(2019)
```

Read in toddler data

```
tidy_fMRI_clinical_toddlers <- read.table(here::here("data", "tidy_fMRI_clinical_toddlers.txt"),
                 header = T, sep = "\t",stringsAsFactors = F)

# subjects with all three language paradigms
fMRI_clinical_all <- tidy_fMRI_clinical_toddlers[!is.na(tidy_fMRI_clinical_toddlers$Story_Lang) &
                        !is.na(tidy_fMRI_clinical_toddlers$Karen_Lang) &
                         !is.na(tidy_fMRI_clinical_toddlers$Motherese), ]

# repeated fMRI scans
knitr::kable(fMRI_clinical_all[duplicated(fMRI_clinical_all$subjid), c(1:4,8)])
```

	subj	subjid	scan_age	group	Gender
11	B6C2P_02	B6C2P	46	TD	M
51	K6E5T_01b	K6E5T	23	TD	F

```
# exclude repeated fMRI scans
dat <- fMRI_clinical_all[!duplicated(fMRI_clinical_all$subjid), ]
```

Run Similarity Network Fusion analysis

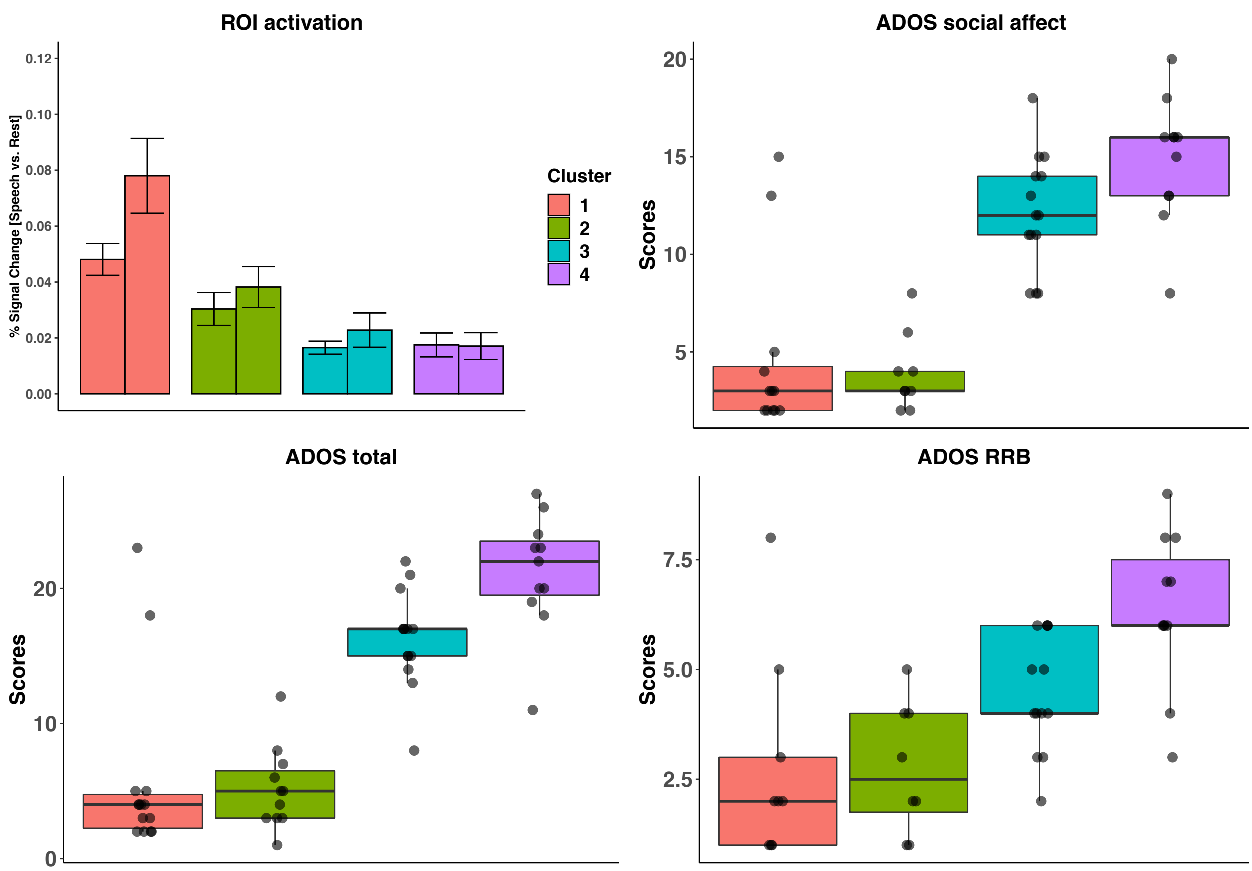
```
ROI_var <- colnames(dplyr::select(fMRI_clinical_all, contains("psc")))
clinic_var <- colnames(dplyr::select(fMRI_clinical_all, contains("final")))[-1]

cluster_results <- SNF_Louvain(dat, ROI_var, clinic_var)
```

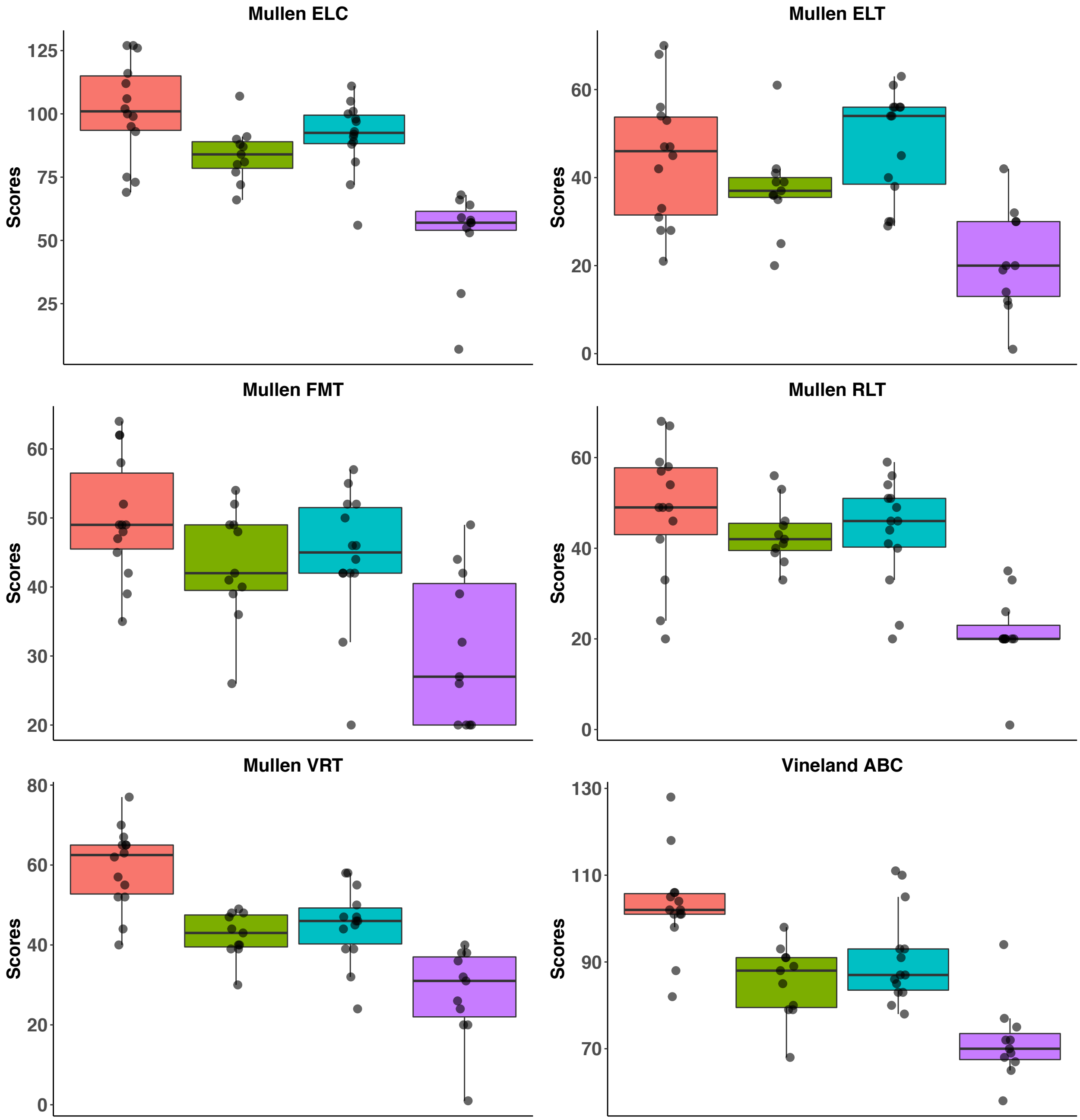
```
## [1] "There are 4 clusters"
```

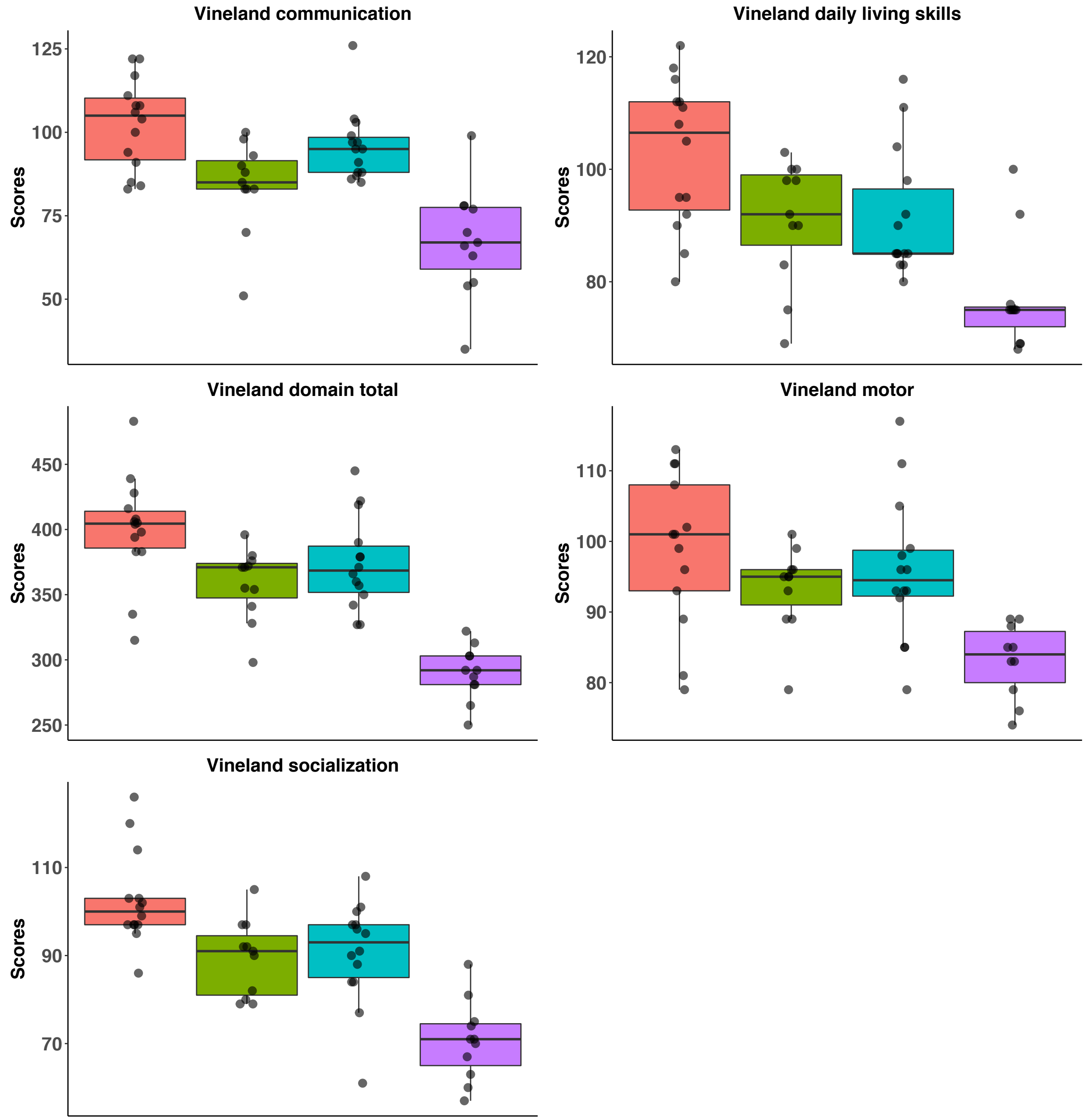
```r
SNF_clusters <- cluster_results[[2]]

# save matrix and clustering results for visualization
cluster_results[[1]][3:6]$Weight <- as.numeric(cluster_results[[1]][3:6]$Weight)
```

Plot fMRI/ROI data and clinical scores across clusters

```r
# plot fMRI and clinical data across clusters
ROI_clinic_clusters = clusters_plot(fMRI_clinical_all, SNF_clusters, ROI_var, clinic_var)
```




Mullen ELC
Mullen ELT
Mullen FMT
Mullen RLT
Mullen VRT
Vineland ABC
Scores
125
100
75
50
25
60
40
20
0
60
50
40
30
20
60
40
20
0
80
60
40
20
0
130
110
90
70




```r
# distribution of clusters
ab <- table(ROI_clinic_clusters[[2]]$index,ROI_clinic_clusters[[2]]$group.x)
knitr::kable(ab)
```

ASD	TD
2	12
3	8
14	0
11	0

```r
# percent of TD subjects in TD cluters
(colSums(ab[1:2,])[2]/colSums(ab)[2])*100
```

```
##  TD
## 100
```

```r
# percent of ASD subjects in ASD cluters
(colSums(ab[3:4,])[1]/colSums(ab)[1])*100
```

```
##      ASD
## 83.33333
```

```r
# ASD in TD clusters
tmp0 <- ROI_clinic_clusters[[2]]

# ASD subjects in TD clusters
tmp0[which(tmp0$index == 1 & tmp0$group.x == "ASD"), 1:22]
```

```
##        subj group.x Story_LHtemporal_psc Story_RHtemporal_psc
## 26 G4M3R_01     ASD             0.045812             0.087231
## 39 M2E2P_01     ASD             0.054112             0.097860
##    Karen_LHtemporal_psc Karen_RHtemporal_psc Motherese_LHtemporal_psc
## 26             0.003444             0.066628                 0.084549
## 39             0.094039             0.137590                 0.075005
##    Motherese_RHtemporal_psc final_ados_CoSoTot final_ados_RRTot
## 26                  0.16645                 15                8
## 39                  0.14142                 13                5
##    final_ados_CoSoTotRRTot final_vine_ComTotal_DomStd
## 26                      23                         83
## 39                      18                         94
##    final_vine_DlyTotal_DomStd final_vine_SocTotal_DomStd
## 26                         95                         97
## 39                        105                         97
##    final_vine_MtrTotal_DomStd final_vine_AdapBehav_DomStd
## 26                        108                          88
## 39                        102                          98
##    final_vine_DomStdTotal final_mullen_VRT final_mullen_FMT final_mullen_RLT
## 26                    383               44               45               20
## 39                    398               40               49               24
##    final_mullen_ELT final_mullen_ELC_Std
## 26               33                   73
## 39               21                   69
```

```r
tmp0[which(tmp0$index == 2 & tmp0$group.x == "ASD"), 1:22]
```

```
##        subj group.x Story_LHtemporal_psc Story_RHtemporal_psc
## 4  A4Q8J_01     ASD             0.031846            0.0165600
## 31 H7R5P_01     ASD             0.018560            0.0052373
## 41 N3C4G_01     ASD             0.047103            0.0782720
##    Karen_LHtemporal_psc Karen_RHtemporal_psc Motherese_LHtemporal_psc
## 4              0.024894             0.040457               -0.0030474
```

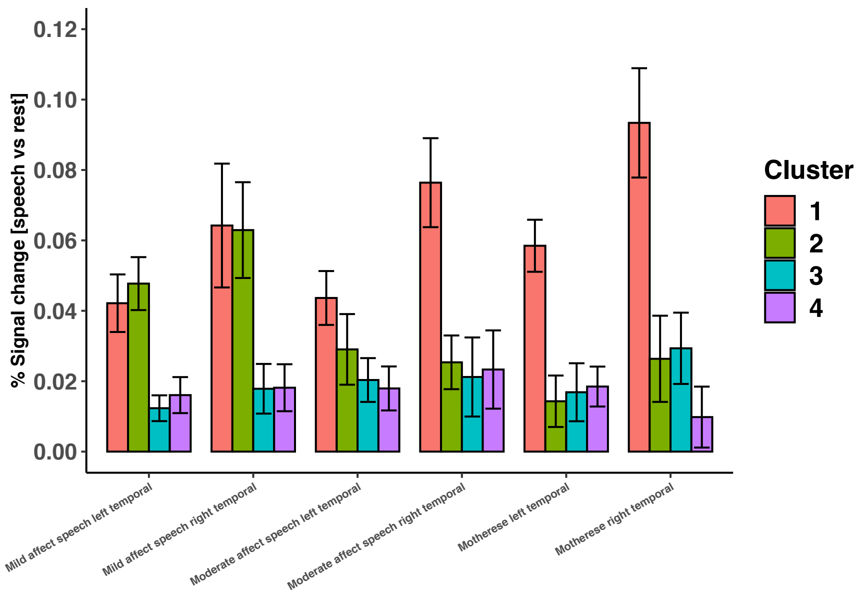
```
## 31             0.014264            0.016447               -0.0257340
## 41             0.031161            0.061965                0.0254680
##    Motherese_RHtemporal_psc final_ados_CoSoTot final_ados_RRTot
## 4                 -0.015587                  6                2
## 31                -0.021631                  0                5
## 41                 0.033967                  8                4
##    final_ados_CoSoTotRRTot final_vine_ComTotal_DomStd
## 4                        8                         83
## 31                       5                         83
## 41                      12                         51
##    final_vine_DlyTotal_DomStd final_vine_SocTotal_DomStd
## 4                          69                         97
## 31                        100                         82
## 41                         75                         79
##    final_vine_MtrTotal_DomStd final_vine_AdapBehav_DomStd
## 4                          79                          80
## 31                         89                          85
## 41                         93                          68
##    final_vine_DomStdTotal final_mullen_VRT final_mullen_FMT final_mullen_RLT
## 4                     328               39               36               40
## 31                    354               49               40               56
## 41                    298               48               26               33
##    final_mullen_ELT final_mullen_ELC_Std
## 4                37                   77
## 31               36                   91
## 41               20                   66
```

```
# mean ROI activation across paradigms
```

barplots for percent signal change

```
# barplot: percent signal change for each language paradigm and each cluster
tmp <- ROI_clinic_clusters[[3]]
tmp$test <- factor(tmp$test, levels = levels(tmp$test)[c(1,4,2,5,3,6)])
tmp$gr[grep("Story*", tmp$test)] <- "Story"
tmp$gr[grep("Karen*", tmp$test)] <- "Karen"
tmp$gr[grep("Motherese*", tmp$test)] <- "Motherese"
tmp$gr <- as.factor(tmp$gr)
ggplot(tmp, aes(x = test, y = values, fill = index)) +
        geom_bar(width = 0.8, stat = "summary",fun = "mean", color = "black",
             position = position_dodge(width = 0.8))+
    geom_errorbar(width = 0.6, stat = "summary", fun.data = "mean_se",
             position = position_dodge(width = 0.8)) +
        labs(y = "% Signal change [speech vs rest]", x = "") +
        theme(legend.title = element_text(colour="black", size=14, face="bold"),
              legend.text = element_text(colour="black", size=14, face="bold")) +
        theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))+
        theme(axis.text.y = element_text(size = 12, face = "bold"),
              axis.text.x = element_text(size = 6, face = "bold", angle = 30,
                            hjust = 1),
              axis.title.y = element_text(size = 10, face = "bold"),
```

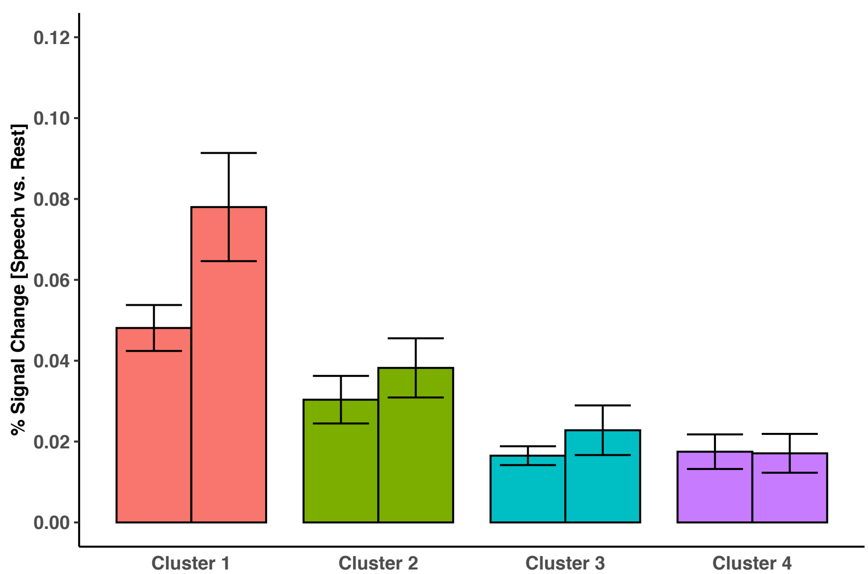
```
              axis.title.x = element_blank()) +
        theme(panel.background = element_blank(),
              panel.grid = element_blank(),
              panel.border = element_blank(),
              axis.line = element_line(colour = "black")) + # remove background
        scale_fill_hue(name = "Cluster") +
        scale_x_discrete(labels=c("Story_LHtemporal_psc" = "Mild affect speech left temporal",
                      "Story_RHtemporal_psc" = "Mild affect speech right temporal",
                      "Karen_LHtemporal_psc" = "Moderate affect speech left temporal",
                      "Karen_RHtemporal_psc" = "Moderate affect speech right temporal",
                      "Motherese_LHtemporal_psc" = "Motherese left temporal",
                      "Motherese_RHtemporal_psc" = "Motherese right temporal")) +
        coord_cartesian(ylim=c(0,0.12)) +
        scale_y_continuous(breaks = seq(0,0.12, 0.02))
```




```
# barplot: average percent signal change across language paradigms for each cluster
tmp1 <- aggregate(values ~ subj + index, FUN = mean, tmp[grep("LHtemporal",tmp$test),])
tmp2 <- aggregate(values ~ subj + index, FUN = mean, tmp[grep("RHtemporal",tmp$test),])

tmp3 <- cbind(tmp1,tmp2[3])

colnames(tmp3)[3:4] <- c("LHtemporal", "RHtemporal")

tmp4 <- gather(tmp3, test, values, 3:4)

tmp5 <- tmp4[order(tmp4$subj),]
tmp5$cluster[tmp5$index == 1] <- "Cluster 1"
tmp5$cluster[tmp5$index == 2] <- "Cluster 2"
tmp5$cluster[tmp5$index == 3] <- "Cluster 3"
tmp5$cluster[tmp5$index == 4] <- "Cluster 4"
tmp5$cluster <- as.factor(tmp5$cluster)

ggplot(tmp5, aes(x = cluster, y = values, group = test, fill = cluster)) +
```

```
        geom_bar(width = 0.8, stat = "summary",fun = "mean", color = "black",
             position = position_dodge(width = 0.8)) +
    geom_errorbar(width = 0.6, stat = "summary", fun.y = "mean_se",
              position = position_dodge(width = 0.8)) +
        labs(y = "% Signal Change [Speech vs. Rest]", x = "") +
    guides(fill = 'none') +
        theme(legend.title = element_text(colour="black", size=14, face="bold"),
              legend.text = element_text(colour="black", size=14, face="bold")) +
        theme(plot.title = element_text(hjust = 0.5, size = 16, face = "bold"))+
        theme(axis.ticks.x = element_blank(),
              axis.text = element_text(size = 10, face = "bold"),
              axis.title.y = element_text(size = 10, face = "bold")) +
        scale_fill_manual(values = c("#F8766D","#7CAE00","#00BFC4","#C77CFF")) +
        theme(panel.background = element_blank(),
              panel.grid = element_blank(),
              panel.border = element_blank(),
              axis.line = element_line(colour = "black")) + # remove background
        coord_cartesian(ylim=c(0.00,0.12)) +
        scale_y_continuous(breaks = seq(0.00,0.12, 0.02))
```

```
## No summary function supplied, defaulting to `mean_se()`
```



Statistical analysis for fMRI activation between clusters

```
# data file
ROI_clinic_new <- ROI_clinic_clusters[[2]]

# 3 way ANOVA: cluster, paradigm, hemisphere
temporal_Cluster <- gather(ROI_clinic_new, ROIs, values,
               Story_LHtemporal_psc:Motherese_RHtemporal_psc)
```

```r
temporal_Cluster$hemisphere[grep("*RHtemporal*", temporal_Cluster$ROIs)] <- "RH"
temporal_Cluster$hemisphere[grep("*LHtemporal*", temporal_Cluster$ROIs)] <- "LH"
temporal_Cluster$paradigm[grep("*Story*", temporal_Cluster$ROIs)] <- "Story"
temporal_Cluster$paradigm[grep("*Karen*", temporal_Cluster$ROIs)] <- "Karen"
temporal_Cluster$paradigm[grep("*Motherese*", temporal_Cluster$ROIs)] <- "Motherese"
temporal_Cluster$hemisphere <- as.factor(temporal_Cluster$hemisphere)
temporal_Cluster$paradigm <- as.factor(temporal_Cluster$paradigm)
temporal_Cluster$cluster <- temporal_Cluster$index
kk <- get_anova_table(anova_test(data = temporal_Cluster, dv = values,
                 wid = subj, between = cluster,
                 within = c(paradigm, hemisphere)))
knitr::kable(kk)
```

Effect	DFn	DFd	F	p	p<.05	ges
cluster	3	46	11.746	7.80e-06	*	0.239000
paradigm	2	92	0.162	8.51e-01		0.001000
hemisphere	1	46	9.502	3.00e-03	*	0.025000
cluster:paradigm	6	92	2.833	1.40e-02	*	0.066000
cluster:hemisphere	3	46	3.683	1.90e-02	*	0.029000
paradigm:hemisphere	2	92	0.315	7.31e-01		0.000554
cluster:paradigm:hemisphere	6	92	1.473	1.96e-01		0.008000

```r
# % signal change in Motherese between clusters using two-sample t-tests
ROI <- "Motherese_RHtemporal_psc"
diff_clusters <- as.data.frame(matrix(0,3,5))
# Cluster 1 vs 2
tt <- t.test(temporal_Cluster$values[temporal_Cluster$index == 1 &
                        temporal_Cluster$ROIs == ROI],
       temporal_Cluster$values[temporal_Cluster$index == 2 &
                    temporal_Cluster$ROIs == ROI])

ef <- effsize::cohen.d(temporal_Cluster$values[temporal_Cluster$index == 1 &
                            temporal_Cluster$ROIs == ROI],
         temporal_Cluster$values[temporal_Cluster$index == 2 &
                        temporal_Cluster$ROIs == ROI])

diff_clusters[1,1] <- "Cluster 1 vs 2"
diff_clusters[1,2:5] <- cbind(round(tt$statistic,2),round(tt$p.value,5),
                  round(ef$estimate,2),
                 paste0("[",round(ef$conf.int[1],2),",",
                 round(ef$conf.int[2],2),"]"))

# Cluster 1 vs 3
tt <- t.test(temporal_Cluster$values[temporal_Cluster$index == 1 &
                        temporal_Cluster$ROIs == ROI],
       temporal_Cluster$values[temporal_Cluster$index == 3 &
                    temporal_Cluster$ROIs == ROI])

ef<- effsize::cohen.d(temporal_Cluster$values[temporal_Cluster$index == 1 &
                            temporal_Cluster$ROIs == ROI],
         temporal_Cluster$values[temporal_Cluster$index == 3 &
                        temporal_Cluster$ROIs == ROI])
```

```r
diff_clusters[2,1] <- "Cluster 1 vs 3"
diff_clusters[2,2:5] <- cbind(round(tt$statistic,2),round(tt$p.value,5),
                  round(ef$estimate,2),
                 paste0("[",round(ef$conf.int[1],2),",",
                 round(ef$conf.int[2],2),"]"))
# Cluster 1 vs 4
tt <- t.test(temporal_Cluster$values[temporal_Cluster$index == 1 &
                        temporal_Cluster$ROIs == ROI],
       temporal_Cluster$values[temporal_Cluster$index == 4 &
                    temporal_Cluster$ROIs == ROI])

ef <- effsize::cohen.d(temporal_Cluster$values[temporal_Cluster$index == 1 &
                            temporal_Cluster$ROIs == ROI],
         temporal_Cluster$values[temporal_Cluster$index == 4 &
                        temporal_Cluster$ROIs == ROI])

diff_clusters[3,1] <- "Cluster 1 vs 4"
diff_clusters[3,2:5] <- cbind(round(tt$statistic,2),round(tt$p.value,5),
                  round(ef$estimate,2),
                 paste0("[",round(ef$conf.int[1],2),",",
                 round(ef$conf.int[2],2),"]"))

colnames(diff_clusters) <- c("Contrast","t value","p value","Cohen's d", "95% CI")
knitr::kable(diff_clusters)
```

Contrast	t value	p value	Cohen's d	95% CI
Cluster 1 vs 2	3.39	0.00255	1.31	[0.39,2.23]
Cluster 1 vs 3	3.45	0.00222	1.31	[0.45,2.16]
Cluster 1 vs 4	4.7	0.00014	1.76	[0.78,2.73]

```r
# FDR correction
p.adjust(diff_clusters$`p value`)
```

```
## [1] 0.00444 0.00444 0.00042
```

Cluster differences between motherese vs. non-motherse speech

```r
# differences in Motherese vs Karen language between Cluster 1 vs 4
ROI_clinic_new$ratio_LHMotherese_Karen <-
    ROI_clinic_new$Motherese_LHtemporal_psc-ROI_clinic_new$Karen_LHtemporal_psc

ROI_clinic_new$ratio_RHMotherese_Karen <-
    ROI_clinic_new$Motherese_RHtemporal_psc-ROI_clinic_new$Karen_RHtemporal_psc

# test if normally distributed
shapiro.test(ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 1])
```

```
##
```

```
##  Shapiro-Wilk normality test
##
## data:  ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 1]
## W = 0.93355, p-value = 0.3419
```

```r
shapiro.test(ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 4])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 4]
## W = 0.8518, p-value = 0.04505
```

```r
shapiro.test(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 1])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 1]
## W = 0.92838, p-value = 0.2896
```

```r
shapiro.test(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 4])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 4]
## W = 0.84276, p-value = 0.03433
```

```r
## non-parametric tests
# Cluster 1 vs 4: right temporal ROI
wilcox.test(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 1],
        ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 4],
        alternative = "greater")
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 1] and ROI_clinic_new$ratio_RHM
## W = 114, p-value = 0.02211
## alternative hypothesis: true location shift is greater than 0
```

```r
wilcoxonZ(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 1],
        ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 4])
```

```
##    z
## 2.03
```

```r
wilcox_effsize(ROI_clinic_new[ROI_clinic_new$index == 1 | ROI_clinic_new$index == 4,],
           ratio_RHMotherese_Karen ~ index, alternative = "greater",
           conf.level = 0.95,ci = TRUE)
```

```
## # A tibble: 1 x 9
##   .y.              group1 group2 effsize    n1    n2 conf.low conf.high magnitude
## * <chr>            <chr>  <chr>    <dbl> <int> <int>    <dbl>     <dbl> <ord>
## 1 ratio_RHMother~ 1      4        0.405    14    11     0.05      0.71 moderate
```

```r
# Cluster 1 vs 4: left temporal ROI
wilcox.test(ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 1],
        ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 4],
        alternative = "greater")
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 1] and ROI_clinic_new$ratio_LHM
## W = 91, p-value = 0.2334
## alternative hypothesis: true location shift is greater than 0
```

```r
wilcoxonZ(ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 1],
        ROI_clinic_new$ratio_LHMotherese_Karen[ROI_clinic_new$index == 4])
```

```
##     z
## 0.766
```

```r
wilcox_effsize(ROI_clinic_new[ROI_clinic_new$index == 1 | ROI_clinic_new$index == 4,],
           ratio_LHMotherese_Karen ~ index, alternative = "greater",
           conf.level = 0.95,ci = TRUE)
```

```
## # A tibble: 1 x 9
##   .y.              group1 group2 effsize    n1    n2 conf.low conf.high magnitude
## * <chr>            <chr>  <chr>    <dbl> <int> <int>    <dbl>     <dbl> <ord>
## 1 ratio_LHMother~ 1      4        0.153    14    11     0.01      0.51 small
```

```r
# Cluster 3 vs 4: right temporal ROI
wilcox.test(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 3],
        ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 4],
        alternative = "greater")
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 3] and ROI_clinic_new$ratio_RHM
## W = 104, p-value = 0.07461
## alternative hypothesis: true location shift is greater than 0
```

```r
wilcoxonZ(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 3],
        ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index == 4])
```

```
##    z
## 1.48
```

```r
wilcox_effsize(ROI_clinic_new[ROI_clinic_new$index == 3 | ROI_clinic_new$index == 4,],
           ratio_RHMotherese_Karen ~ index, alternative = "greater",
           conf.level = 0.95,ci = TRUE)
```

```
## # A tibble: 1 x 9
##   .y.              group1 group2 effsize    n1    n2 conf.low conf.high magnitude
## * <chr>            <chr>  <chr>    <dbl> <int> <int>    <dbl>     <dbl> <ord>
## 1 ratio_RHMother~ 3      4        0.296    14    11     0.02      0.65 small
```

```r
# greater/less activation to motherese relative to moderate affect speech (Karen language)
# Cluster 1
l <- sum(ROI_clinic_new$Motherese_RHtemporal_psc[ROI_clinic_new$index == 1]-
        ROI_clinic_new$Karen_RHtemporal_psc[ROI_clinic_new$index == 1])/
    sum(ROI_clinic_new$Karen_RHtemporal_psc[ROI_clinic_new$index == 1]) * 100

print(paste0("Cluster 1 had ", round(l),
         "% greater activation to motherese than to moderate affect speech"))
```

```
## [1] "Cluster 1 had 22% greater activation to motherese than to moderate affect speech"
```

```r
# Cluster 3
l <- sum(ROI_clinic_new$Motherese_RHtemporal_psc[ROI_clinic_new$index == 3]-
        ROI_clinic_new$Karen_RHtemporal_psc[ROI_clinic_new$index == 3])/
    sum(ROI_clinic_new$Karen_RHtemporal_psc[ROI_clinic_new$index == 3]) * 100

print(paste0("Cluster 3 had ", round(l),
         "% greater activation to motherese than to moderate affect speech"))
```

```
## [1] "Cluster 3 had 38% greater activation to motherese than to moderate affect speech"
```

```r
# Cluster 4
l <- sum(ROI_clinic_new$Motherese_RHtemporal_psc[ROI_clinic_new$index == 4]-
        ROI_clinic_new$Karen_RHtemporal_psc[ROI_clinic_new$index == 4])/
    sum(ROI_clinic_new$Karen_RHtemporal_psc[ROI_clinic_new$index == 4]) * 100

print(paste0("Cluster 4 had ", abs(round(l)),
         "% less activation to motherese than to moderate affect speech"))
```

```
## [1] "Cluster 4 had 58% less activation to motherese than to moderate affect speech"
```

```r
# % subjects had greater/less activation to motherese relative to moderate affect speech
nn <- length(which(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index==1] >0))/
    length(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index==1])

print(paste0(round(nn*100),
         "% of Cluster 1 toddlers had greater activation to motherese vs moderate affect speech"))
```

```
## [1] "71% of Cluster 1 toddlers had greater activation to motherese vs moderate affect speech"
```

```r
nn <- length(which(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index==4] <0))/
    length(ROI_clinic_new$ratio_RHMotherese_Karen[ROI_clinic_new$index==4]<0)

print(paste0(round(nn*100),
         "% of Cluster 4 toddlers had less activation to motherese vs moderate affect speech"))
```

```
## [1] "82% of Cluster 4 toddlers had less activation to motherese vs moderate affect speech"
```

```r
# Chi-squared analysis
diff_psc <- "ratio_RHMotherese_Karen"
m <- length(which(ROI_clinic_new[ROI_clinic_new$index==1, diff_psc]>0))
n <- length(which(ROI_clinic_new[ROI_clinic_new$index==1, diff_psc]<0))
j <- length(which(ROI_clinic_new[ROI_clinic_new$index==4, diff_psc]>0))
k <- length(which(ROI_clinic_new[ROI_clinic_new$index==4, diff_psc]<0))
chisq.test(matrix(c(m,n,j,k),nrow =2))
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  matrix(c(m, n, j, k), nrow = 2)
## X-squared = 5.0265, df = 1, p-value = 0.02496
```

```r
knitr::kable(cramerV(matrix(c(m,n,j,k),nrow =2),ci=TRUE))
```

Cramer.V	lower.ci	upper.ci
0.529	0.1667	0.8397

```r
# motherese vs moderate affect speech across clusters
ROI_clinic_clusters_ratio <- gather(ROI_clinic_new, tests,values, ratio_LHMotherese_Karen:ratio_RHMothe

ROI_clinic_clusters_ratio$tests <- factor(ROI_clinic_clusters_ratio$tests, levels = c("ratio_LHMotheres

ROI_clinic_clusters_ratio$index <- factor(ROI_clinic_clusters_ratio$index)
ROI_clinic_clusters_ratio$values
```

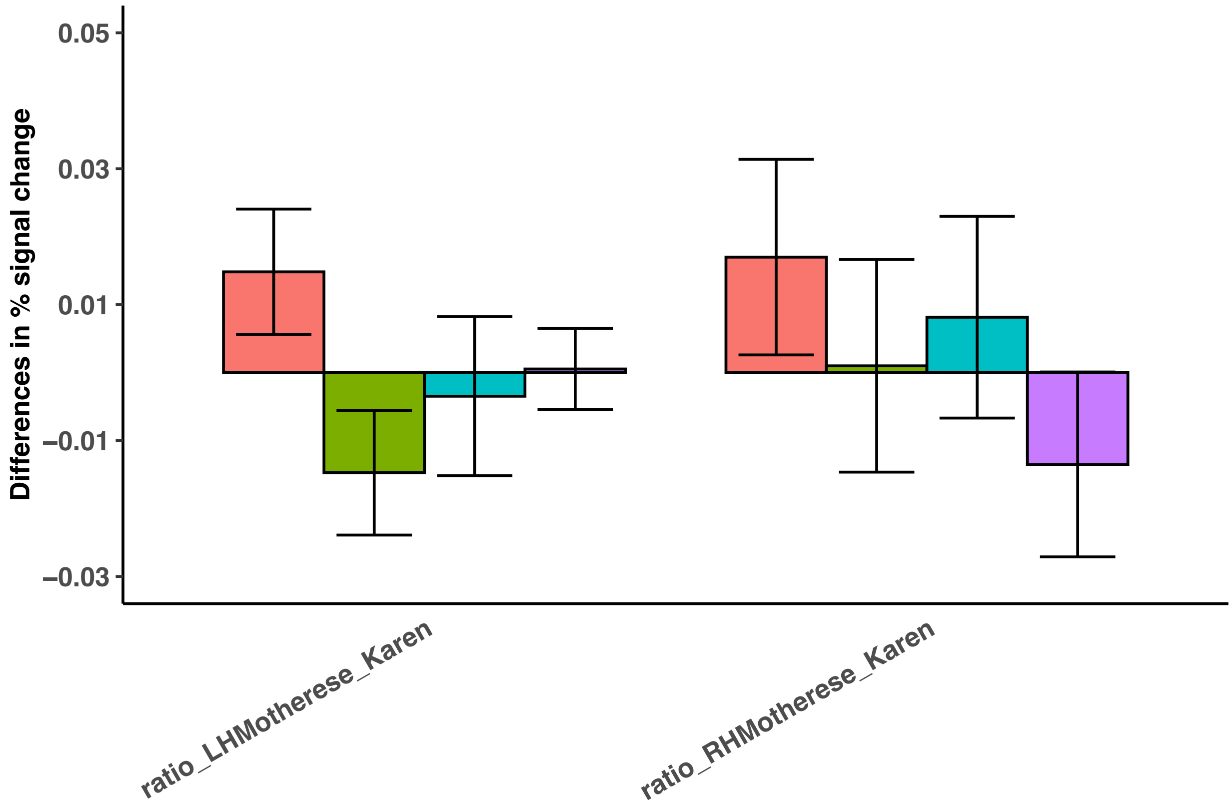
```
##   [1]  0.07598600  0.02941430 -0.01919600 -0.02794140  0.01411600  0.00382400
##   [7]  0.00676900 -0.00467700 -0.01847200 -0.00306000 -0.04332000 -0.00101500
##  [13] -0.00098600  0.01138100 -0.04519300 -0.14451100 -0.02665700  0.01253510
##  [19]  0.00770600 -0.04955600 -0.02012300  0.01292482 -0.00566413  0.01996300
##  [25]  0.03400200  0.08110500  0.00245430  0.03813000 -0.00659191 -0.00271400
##  [31] -0.03999800 -0.01504700 -0.00040890  0.01032300  0.02779000 -0.04018760
##  [37]  0.00416400  0.03050280 -0.01903400 -0.00685050 -0.00569300  0.00811330
##  [43]  0.02374100 -0.03978500  0.00511490  0.04028000  0.01971800  0.02935880
##  [49]  0.03790500  0.00224960  0.13431600  0.03539710 -0.04399700 -0.05604400
##  [55] -0.03703800  0.00351400  0.03452800  0.02088100  0.02254600 -0.02426000
##  [61]  0.01775800 -0.01541900  0.00293600 -0.02274120 -0.03959000 -0.10853560
##  [67]  0.02986710  0.12956820  0.02247000 -0.09594000 -0.01853600  0.01025010
##  [73] -0.03059000 -0.00495240  0.05072550  0.09982200 -0.00701590  0.06765120
```

```
## [79]  0.00826710 -0.02527580 -0.03807800 -0.02904800 -0.00410540  0.02134230
## [85]  0.02305200 -0.03748220 -0.09040100  0.09599800  0.00383000 -0.00398300
## [91] -0.02799800  0.00559170  0.01067240 -0.01572600 -0.01219700  0.02885200
## [97]  0.05022200  0.11966210 -0.01350500 -0.03295100
```

```
ggplot(ROI_clinic_clusters_ratio, aes(x = tests, y = values, group = index, fill = index)) +
    geom_bar(width = 0.8,stat = "summary", fun = "mean", color = "black",
             position = position_dodge(width = 0.8)) +
    geom_errorbar(width = 0.6, stat = "summary", fun.y = "mean_se",
              position = position_dodge(width = 0.8)) +
        labs(y = "Differences in % signal change", x = "") +
    guides(fill = 'none') +
        theme(legend.title = element_text(colour="black", size=14, face="bold"),
              legend.text = element_text(colour="black", size=14, face="bold")) +
        theme(plot.title = element_text(hjust = 0.5, size = 16, face = "bold"))+
        theme(axis.text.x = element_text(size = 10, face = "bold",angle = 30,
                         hjust = 1),
              axis.ticks.x = element_blank(),
              axis.text.y = element_text(size = 10, face = "bold"),
              axis.title.y = element_text(size = 10, face = "bold")) +
        scale_fill_manual(values = c("#F8766D","#7CAE00","#00BFC4","#C77CFF")) +
        theme(panel.background = element_blank(),
              panel.grid = element_blank(),
              panel.border = element_blank(),
              axis.line = element_line(colour = "black")) +
        coord_cartesian(ylim=c(-0.03,0.05)) +
        scale_y_continuous(breaks = seq(-0.03,0.05, 0.02))
```

```
## Warning: Ignoring unknown parameters: fun.y
```

```
## No summary function supplied, defaulting to ‘mean_se()‘
```



Correlations between eye-tracking and fMRI activation in ASD and TD

```
# load motherese data
Motherese_ET <- read.table(here::here("data","tidy_Motherese_ET.txt"),header = T,
                sep = "\t", stringsAsFactors = F)

colnames(Motherese_ET)[2] <- "subj"

# organize data file
tidy_fMRI_Motherese_ET <- merge(tidy_fMRI_clinical_toddlers,Motherese_ET, by = "subj")
dim(tidy_fMRI_Motherese_ET)
```

```
## [1] 54 57
```

```
which(duplicated(tidy_fMRI_Motherese_ET$subjid.x))
```

```
## integer(0)
```

```
# correlation analysis (one-tailed test)
corrL_TD <- cor.test(tidy_fMRI_Motherese_ET$Motherese_LHtemporal_psc[
```

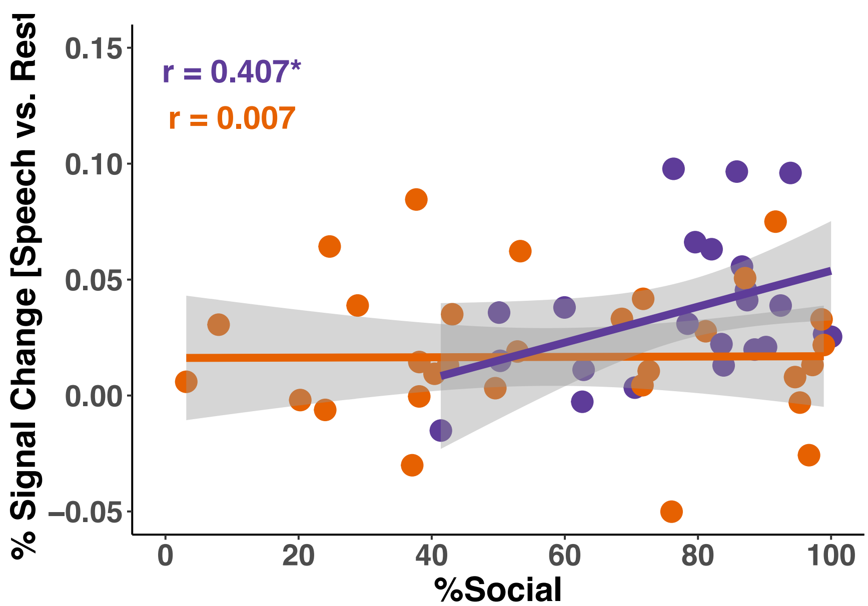
```
    tidy_fMRI_Motherese_ET$group.x == "TD"],
     tidy_fMRI_Motherese_ET$LK_.fixation.Motherese[
        tidy_fMRI_Motherese_ET$group.x == "TD"],
     alternative = "greater")

corrR_TD <- cor.test(tidy_fMRI_Motherese_ET$Motherese_RHtemporal_psc[
    tidy_fMRI_Motherese_ET$group.x == "TD"],
     tidy_fMRI_Motherese_ET$LK_.fixation.Motherese[
        tidy_fMRI_Motherese_ET$group.x == "TD"],
     alternative = "greater")

corrL_ASD <- cor.test(tidy_fMRI_Motherese_ET$Motherese_LHtemporal_psc[
    tidy_fMRI_Motherese_ET$group.x == "ASD"],
     tidy_fMRI_Motherese_ET$LK_.fixation.Motherese[
        tidy_fMRI_Motherese_ET$group.x == "ASD"],
     alternative = "greater")

corrR_ASD <- cor.test(tidy_fMRI_Motherese_ET$Motherese_RHtemporal_psc[
    tidy_fMRI_Motherese_ET$group.x == "ASD"],
     tidy_fMRI_Motherese_ET$LK_.fixation.Motherese[
        tidy_fMRI_Motherese_ET$group.x == "ASD"],
     alternative = "greater")

cor_ET_fMRI <- as.data.frame(matrix(0,2,2))
cor_ET_fMRI <- rbind.data.frame(cbind(paste0("t(",corrL_TD$parameter,")=",round(corrL_TD$estimate,3),
                  ", p=",round(corrL_TD$p.value),3),
               paste0("t(",corrL_ASD$parameter,")=",round(corrL_ASD$estimate,3),
                      ", p=",round(corrL_ASD$p.value,2))),
               cbind(paste0("t(",corrR_TD$parameter,")=",round(corrR_TD$estimate,3),
                     ", p=",round(corrR_TD$p.value,2)),
               paste0("t(",corrR_ASD$parameter,")=",round(corrR_ASD$estimate,3),
                      ", p=",round(corrR_ASD$p.value,2))))

colnames(cor_ET_fMRI) <- c("ASD","TD")
rownames(cor_ET_fMRI) <- c("LHtemporal","RHtemporal")
knitr::kable(cor_ET_fMRI)
```

	ASD	TD
LHtemporal	t(21)=0.407, p=03	t(29)=0.007, p=0.49
RHtemporal	t(21)=0.186, p=0.2	t(29)=-0.097, p=0.7

```
# scatter plots
ROI <- "Motherese_LHtemporal_psc"
p <- ggplot(tidy_fMRI_Motherese_ET, aes_string(x = "LK_.fixation.Motherese", y = ROI)) +
    geom_point(aes(color = group.x),position = "jitter", size = 5) +
    geom_smooth(aes(color = group.x), size = 2, method = "lm") +
    guides(color = F) +
    labs(x = "%Social", y = "% Signal Change [Speech vs. Rest]") +
    scale_color_manual(values = c("#e66101", "#5e3c99")) +
    theme(legend.title = element_text(colour="black", size=16, face="bold"),
          legend.text = element_text(colour="black", size=16, face="bold")) +
    theme(plot.title = element_text(hjust = 0.5))+
```

```
    theme(axis.text = element_text(size = 16, face = "bold"),
          axis.title = element_text(size = 18, face = "bold")) +
    theme(panel.background = element_blank(),
          panel.border = element_blank(),
          panel.grid = element_blank(),
          axis.line = element_line(colour = "black")) +
    coord_cartesian(ylim=c(-0.05,0.15), xlim=c(0, 100)) +
    scale_y_continuous(breaks = seq(-0.05,0.15, 0.05)) +
    scale_x_continuous(breaks = seq(0,100, 20))

    p1 <- p + annotate(geom = "text", x = 10, y = 0.14, size = 6,fontface="bold",
                  label = paste0("r = ",round(corrL_TD$estimate[[1]],3),"*"),
              color = "#5e3c99") +
          annotate(geom = "text", x = 10, y = 0.12, size = 6,fontface="bold",
                  label = paste0("r = ",round(corrL_ASD$estimate[[1]],3)),
            color = "#e66101")

    print(p1)
```

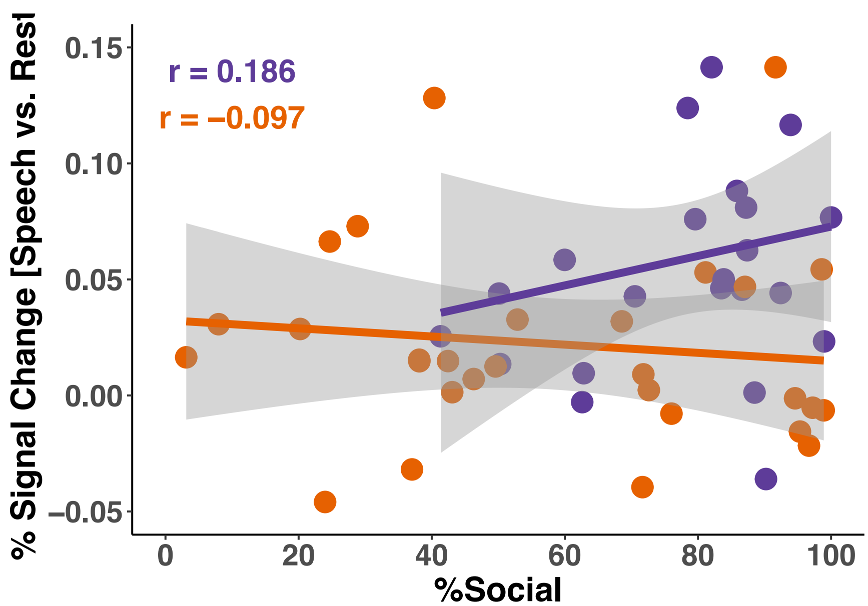
```
## `geom_smooth()` using formula 'y ~ x'
```




```
ROI <- "Motherese_RHtemporal_psc"
p <- ggplot(tidy_fMRI_Motherese_ET, aes_string(x = "LK_.fixation.Motherese", y = ROI)) +
    geom_point(aes(color = group.x),position = "jitter", size = 5) +
    geom_smooth(aes(color = group.x), size = 2, method = "lm") +
    guides(color = F) +
    labs(x = "%Social", y = "% Signal Change [Speech vs. Rest]") +
    scale_color_manual(values = c("#e66101", "#5e3c99")) +
    theme(legend.title = element_text(colour="black", size=16, face="bold"),
          legend.text = element_text(colour="black", size=16, face="bold")) +
    theme(plot.title = element_text(hjust = 0.5))+
    theme(axis.text = element_text(size = 16, face = "bold"),
          axis.title = element_text(size = 18, face = "bold")) +
```

```r
    theme(panel.background = element_blank(),
          panel.border = element_blank(),
          panel.grid = element_blank(),
          axis.line = element_line(colour = "black")) +
    coord_cartesian(ylim=c(-0.05,0.15), xlim=c(0, 100)) +
    scale_y_continuous(breaks = seq(-0.05,0.15, 0.05)) +
    scale_x_continuous(breaks = seq(0,100, 20))

    p1 <- p + annotate(geom = "text", x = 10, y = 0.14, size = 6,fontface="bold",
                   label = paste0("r = ",round(corrR_TD$estimate[[1]],3)),
               color = "#5e3c99") +
          annotate(geom = "text", x = 10, y = 0.12, size = 6,fontface="bold",
                   label = paste0("r = ",round(corrR_ASD$estimate[[1]],3)),
             color = "#e66101")

    print(p1)
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



Associations of clusters and gaze preference for motherese

```r
# organize data file
clusters <- ROI_clinic_clusters[[2]][,c("subj","Clustering","index")]
dim(clusters)
```

```
## [1] 50  3
```

```r
colnames(Motherese_ET)[2] <- "subj"
Motherese_ET_clusters <- merge(Motherese_ET, clusters, by = "subj")
```

```r
dim(Motherese_ET_clusters)[1]
```

```
## [1] 43
```

```r
Motherese_ET_clusters$index <- as.factor(Motherese_ET_clusters$index)

# number of subjects in each cluster
table(Motherese_ET_clusters$group)
```

```
##
## ASD  TD
##  24  19
```

```r
table(Motherese_ET_clusters$index)
```

```
##
##  1  2  3  4
## 13 10 12  8
```

```r
# % Motherese in each cluster
k <- aggregate(LK_.fixation.Motherese ~ index, FUN = mean, Motherese_ET_clusters)
knitr::kable(k[1:4,])
```

index	LK_.fixation.Motherese
1	79.28466
2	78.83265
3	61.75918
4	40.99659

```r
# difference between Cluster 4 vs 1
tt1 <- t.test(Motherese_ET_clusters$LK_.fixation.Motherese[Motherese_ET_clusters$index == 4],
         Motherese_ET_clusters$LK_.fixation.Motherese[Motherese_ET_clusters$index == 1],
         alternative = "less")
cohens1 <- effsize::cohen.d(Motherese_ET_clusters$LK_.fixation.Motherese[
    Motherese_ET_clusters$index == 1],
               Motherese_ET_clusters$LK_.fixation.Motherese[
                Motherese_ET_clusters$index == 4])

knitr::kable(cbind(t_value = round(tt1$statistic[[1]],2), pvalue = round(tt1$p.value,4),
           d = round(cohens1$estimate,2),paste0("95% CI [",round(cohens1$conf.int[1],2),
                            ",", round(cohens1$conf.int[2],2),"]")))
```

t_value	pvalue	d	
-3.95	9e-04	1.86	95% CI [0.75,2.98]

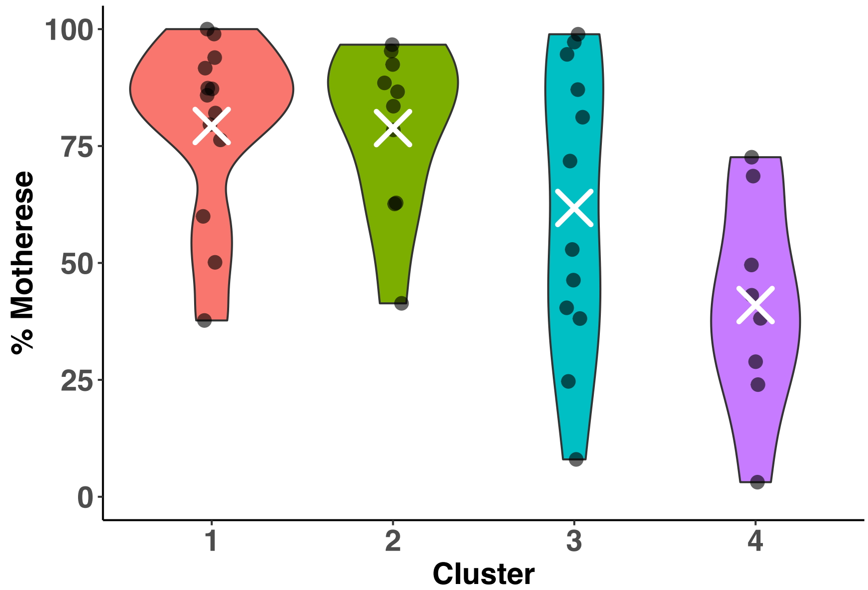
```r
# difference between Cluster 4 vs 2
tt2 <- t.test(Motherese_ET_clusters$LK_.fixation.Motherese[Motherese_ET_clusters$index == 4],
         Motherese_ET_clusters$LK_.fixation.Motherese[Motherese_ET_clusters$index == 2],
         alternative = "less")
cohens2 <- effsize::cohen.d(Motherese_ET_clusters$LK_.fixation.Motherese[
    Motherese_ET_clusters$index == 2],
               Motherese_ET_clusters$LK_.fixation.Motherese[
                Motherese_ET_clusters$index == 4])

knitr::kable(cbind(t_value = round(tt2$statistic[[1]],2), pvalue = round(tt2$p.value,4),
          d = round(cohens2$estimate,2),paste0("95% CI [",round(cohens2$conf.int[1],2),
                            ",", round(cohens2$conf.int[2],2),"]")))
```

t_value	pvalue	d	
-3.82	0.0011	1.86	95% CI [0.66,3.07]

```r
# difference between Cluster 4 vs 3
tt3 <- t.test(Motherese_ET_clusters$LK_.fixation.Motherese[Motherese_ET_clusters$index == 4],
         Motherese_ET_clusters$LK_.fixation.Motherese[Motherese_ET_clusters$index == 3],
         alternative = "less")
cohens3 <- effsize::cohen.d(Motherese_ET_clusters$LK_.fixation.Motherese[
    Motherese_ET_clusters$index == 3],
    Motherese_ET_clusters$LK_.fixation.Motherese[Motherese_ET_clusters$index == 4])

knitr::kable(cbind(t_value = round(tt3$statistic[[1]],2), pvalue = round(tt3$p.value,4),
          d = round(cohens3$estimate,2),paste0("95% CI [",round(cohens3$conf.int[1],2),
                            ",", round(cohens3$conf.int[2],2),"]")))
```

t_value	pvalue	d	
-1.72	0.0513	0.74	95% CI [-0.25,1.73]

```r
# violin plots
ET_clusters(Motherese_ET_clusters)
```




```
# t-tests and effect sizes
    cohen <- matrix(1:16, 4, 4)
    ttest <- matrix(1:16, 4, 4)
    rownames(cohen) <- c("Cluster 1", "Cluster 2", "Cluster 3","Cluster 4")
    colnames(cohen) <- c("Cluster 1", "Cluster 2", "Cluster 3","Cluster 4")
    rownames(ttest) <- c("Cluster 1", "Cluster 2", "Cluster 3","Cluster 4")
    colnames(ttest) <- c("Cluster 1", "Cluster 2", "Cluster 3","Cluster 4")

    for (i in 1:4) {
        for (j in 1:4) {
            if (i == j) {
                cohen[i, j] <- NA
                ttest[i, j] <- NA
            } else {
                aa <-
                effsize::cohen.d(Motherese_ET_clusters$LK_.fixation.Motherese[
                    Motherese_ET_clusters$index == i],
                         Motherese_ET_clusters$LK_.fixation.Motherese[
                            Motherese_ET_clusters$index == j])

                bb <- t.test(Motherese_ET_clusters$LK_.fixation.Motherese[
                    Motherese_ET_clusters$index == i],
                         Motherese_ET_clusters$LK_.fixation.Motherese[
                            Motherese_ET_clusters$index == j],alternative = "greater")

                cohen[i, j] <- abs(round(aa$estimate,2))
                ttest[i, j] <- abs(round(bb$p.value,4))
            }

        }
    }

# effect size: Cohen's d
knitr::kable(cohen)
```

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	NA	0.02	0.69	1.86
Cluster 2	0.02	NA	0.66	1.86
Cluster 3	0.69	0.66	NA	0.74
Cluster 4	1.86	1.86	0.74	NA

```r
# p-values from t-tests
ttest[lower.tri(as.matrix(ttest))] <- NA
knitr::kable(ttest)
```

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	NA	0.477	0.0535	0.0009
Cluster 2	NA	NA	0.0612	0.0011
Cluster 3	NA	NA	NA	0.0513
Cluster 4	NA	NA	NA	NA

```r
# plot the effect size matrix as a heatmap: 4 clusters
colfunc <- colorRampPalette(c("red", "yellow"))

WGCNA::labeledHeatmap(Matrix = cohen[1:4,1:4], xLabels = colnames(cohen)[1:4],
              yLabels = rownames(cohen)[1:4], ySymbols = NULL, colorLabels = F,
              colors = colfunc(50), textMatrix = round(cohen[1:4,1:4], digits = 2),
              setStdMargins = F, cex.text = 2, zlim = c(0, 2))
```