M21 202303 n250 aov

Joanna Morris

2023-03-20

This R script contains the code for analysing the morph 21 erp data for the 200-300 ms time window.

1. First we load the libraries we need

```
library(readr)
library(psych)
library(dplyr)
library(tidyr)
```

Compute PCA

Following Andrews and Lo (2013) this script computes a PCA for our spelling and vocabulary measures. Because the standardised spelling and vocabulary scores were correlated, to facilitate interpretation, two orthogonal measures of individual differences were derived from a principal components analysis. Analysis based on this tutorial

```
library(readr)
library(dplyr)
library(datawizard)
sv_202303 <- read_csv("m21_spell_vocab_raw.csv")
sv_202303.na <- na.omit(sv_202303)
sv_202303.na <- mutate(sv_202303.na, z_ART = standardise(ART_correct), z_vocab = standardise(vocab_corre

cor.test(sv_202303.na$z_vocab, sv_202303.na$z_spell)
```

```
    Pearson's product-moment correlation

data:  sv_202303.na$z_vocab and sv_202303.na$z_spell
t = 1.9352, df = 61, p-value = 0.05761
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.007724833  0.460807138
sample estimates:
      cor
0.2405005
```

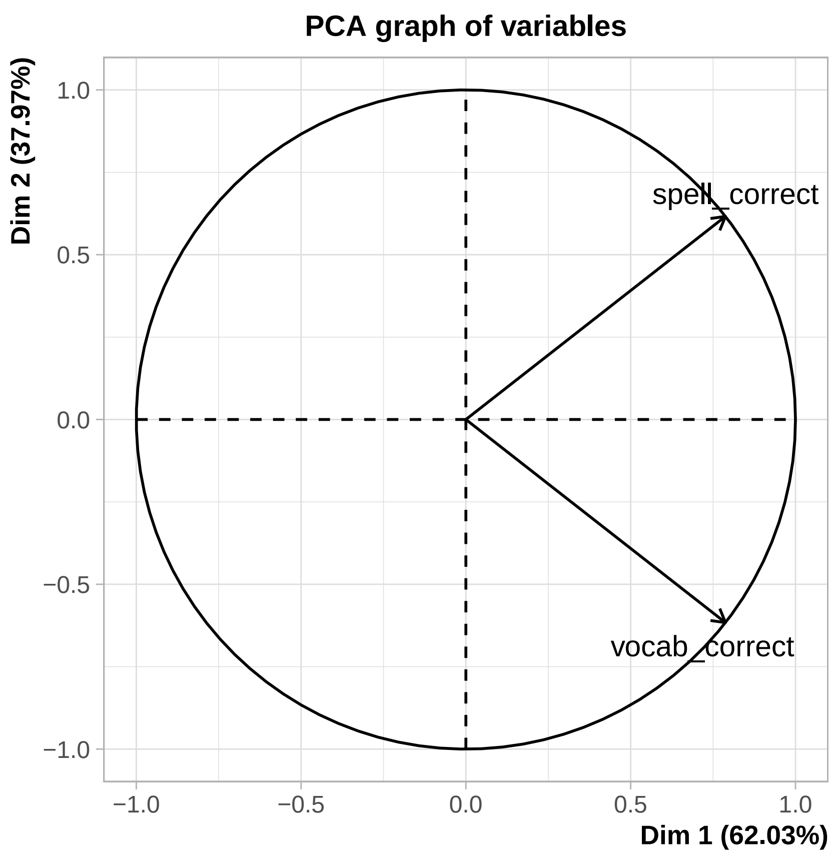
By default, the function `PCA()` in `FactoMineR`, standardizes the data automatically during the PCA; so you don't need do this transformation before the PCA.

- `X`: a data frame. Rows are individuals and columns are numeric variables
- `scale.unit`: a logical value. If TRUE, the data are scaled to unit variance before the analysis. This standardization to the same scale avoids some variables to become dominant just because of their large measurement units. It makes variable comparable.
- `ncp`: number of dimensions kept in the final results.
- `graph`: a logical value. If TRUE a graph is displayed.

The plot shows the relationships between all variables. It can be interpreted as follow:

- Positively correlated variables are grouped together.
- Negatively correlated variables are positioned on opposite sides of the plot origin (opposed quadrants).
- The distance between variables and the origin measures the quality of the variables on the factor map. Variables that are away from the origin are well represented on the factor map.

```
library(FactoMineR)
library(factoextra)

res.pca <- PCA(sv_202303.na[,3:4], scale.unit = TRUE, ncp = 2, graph = FALSE)
plot(res.pca, choix = "varcor", graph.type = c("ggplot"))
```



The eigenvalues measure the amount of variation retained by each principal component. Eigenvalues are large for the first PCs and small for the subsequent PCs. That is, the first PCs corresponds to the directions with the maximum amount of variation in the data set.

We examine the eigenvalues to determine the number of principal components to be considered

```r
(eig.val <- get_eigenvalue(res.pca))
```

```
      eigenvalue variance.percent cumulative.variance.percent
Dim.1  1.2405005         62.02503                    62.02503
Dim.2  0.7594995         37.97497                   100.00000
```

The quality of representation of the variables on factor map is called cos2 (square cosine, squared coordinates). A high cos2 indicates a good representation of the variable on the principal component. In this case the variable is positioned close to the circumference of the correlation circle. A low cos2 indicates that the variable is not perfectly represented by the PCs. In this case the variable is close to the center of the circle. For a given variable, the sum of the cos2 on all the principal components is equal to one. If a variable is perfectly represented by only two principal components (Dim.1 & Dim.2), the sum of the cos2 on these two PCs is equal to one. In this case the variables will be positioned on the circle of correlations.

```r
res.pca$var$cos2
```

```
                  Dim.1     Dim.2
vocab_correct 0.6202503 0.3797497
spell_correct 0.6202503 0.3797497
```

The contributions of variables in accounting for the variability in a given principal component are expressed in percentage. Variables that are correlated with PC1 (i.e., Dim.1) and PC2 (i.e., Dim.2) are the most important in explaining the variability in the data set. Variables that do not correlated with any PC or correlated with the last dimensions are variables with low contribution and might be removed to simplify the overall analysis.

```r
res.pca$var$contrib
```

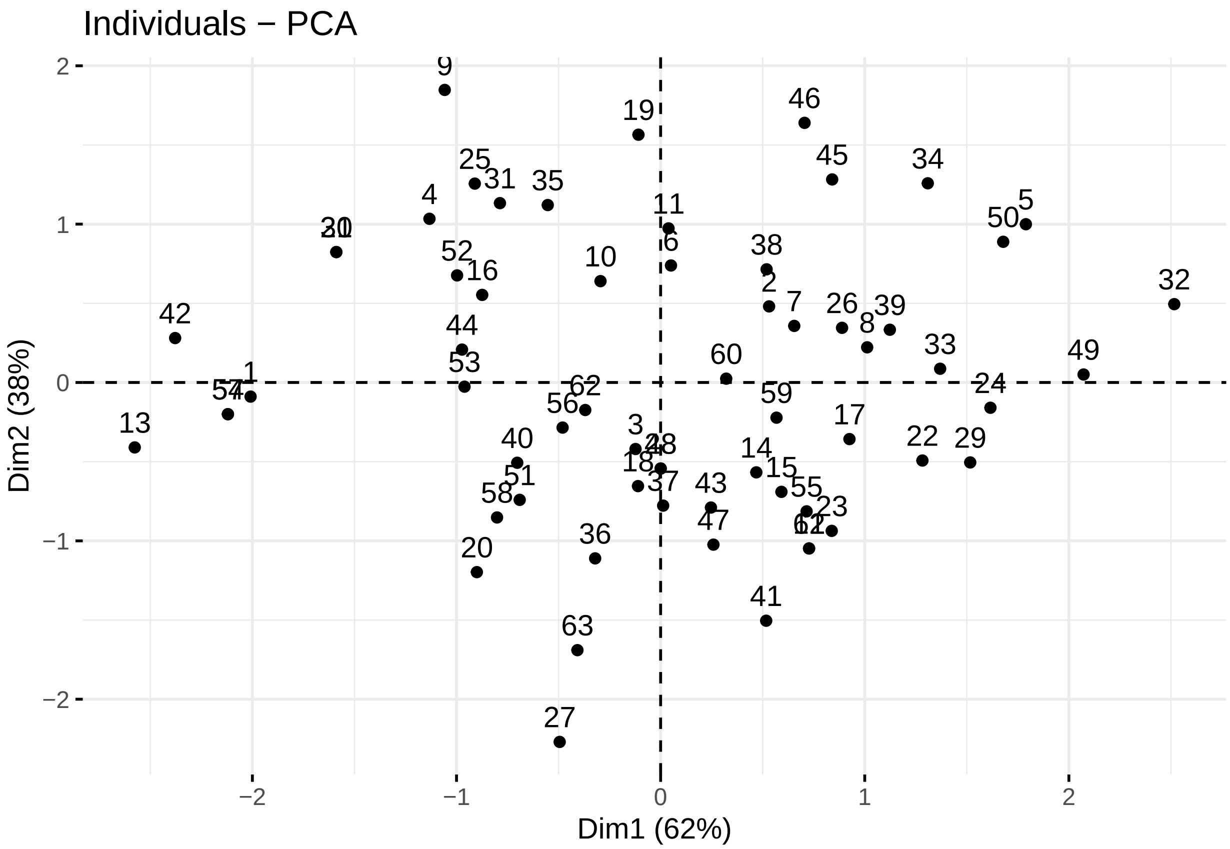
```
              Dim.1 Dim.2
vocab_correct    50    50
spell_correct    50    50
```

```r
(res.desc <- dimdesc(res.pca, axes = c(1,2), proba = 0.05))
```

```
$Dim.1

Link between the variable and the continuous variables (R-square)
=================================================================================
              correlation      p.value
vocab_correct   0.7875597 1.913196e-14
spell_correct   0.7875597 1.913196e-14

$Dim.2

Link between the variable and the continuous variables (R-square)
=================================================================================
              correlation      p.value
spell_correct   0.6162384 7.592937e-08
vocab_correct  -0.6162384 7.592937e-08
```

The fviz_pca_ind() is used to produce the graph of individuals.

```
fviz_pca_ind(res.pca)
```




```
sv_202303.na<-bind_cols(sv_202303.na,res.pca$ind$coord)
```

We load the N250 erp data file and the word and non-word base frequency data

```
n250 <- read_csv("S101-177_n250.csv")
```

Then we join the demographic and erp data files. We will use the `inner_join` rather than the `full_join` function in order to eliminate rows with missing data.

```
n250 <- inner_join(sv_202303.na,n250, by = "SubjID")  #join subject PCA data
```

Divide participants based on median split of Dim2. Higher values on this factor indicate that spelling scores were relatively higher than vocabulary,

```
n250.median <- median(n250$Dim.2)
n250 <- n250 |>
    mutate(lang_type = case_when(
      Dim.2 < n250.median ~ "Semantic",
      Dim.2 > n250.median ~ "Orthographic"
    ))
```

5. Let's save a `.csv` file with the data from the combined dataset

```
write_csv(n250, "202303_sv_n250_rmna.csv")
```

6. For each dataset, we will create a subset with only the electrode sites we will be analysing—F3, Fz, F4, C3, Cz, C4, P3, Pz, P4

```
sites = c(3,2, 25, 7, 20, 21, 12, 11, 16)
n250_9 <- dplyr::filter(n250, chindex %in% sites)
```

7. We then create separate columns, one for each independent variable (anteriority, laterality, morphological family size). To do this we have to use the`mutate` function from the dplyr package along with the `case_when` function. The `case_when` function is a sequence of two-sided formulas. The left hand side determines which values match this case. The right hand side provides the replacement value.

```
n250_9 <- dplyr::mutate(n250_9,
                        anteriority = case_when(grepl("F", chlabel) ~ "Frontal",
                                                grepl("C", chlabel) ~ "Central",
                                                grepl("P", chlabel) ~ "Parietal"))

n250_9 <- dplyr::mutate(n250_9,
                        laterality = case_when(grepl("3", chlabel) ~ "Left",
                                               grepl("z", chlabel) ~ "Midline",
                                               grepl("Z", chlabel) ~ "Midline",
                                               grepl("4", chlabel) ~ "Right"))



n250_9 <- dplyr::mutate(n250_9,
                        fam_size = case_when(grepl("small", binlabel) ~ "Small",
                                             grepl("large", binlabel) ~ "Large"))
```

8. We then create a smaller dataset with only the columns we need

```
n250_9b <- dplyr::select(n250_9,
                             SubjID,
                             lang_type,
                             anteriority,
                             laterality,
                             fam_size,
                             value,
                             chlabel,
                             binlabel)
```

9. We then divide dataset into 3 separate ones—for "words", "simple nonwords" and "complex nonwords"

```
n250_words <- dplyr::filter(n250_9b, grepl("Critical_word",binlabel))
n250_nwsmpl <- dplyr::filter(n250_9b, grepl("simple",binlabel))
n250_nwcplx <- dplyr::filter(n250_9b, grepl("complex",binlabel))
```

#Plot Means

Get condition means

```r
#Define standard error of the mean function

sem <- function(x) sd(x)/sqrt(length(x))

(cw.cond.means <- n250_words |>
   group_by(fam_size, lang_type) |>
   summarise(mean = mean(value),
             se = sem(value),
             num_stim = n()))
```

```
# A tibble: 4 x 5
# Groups:   fam_size [2]
  fam_size lang_type      mean    se num_stim
  <chr>    <chr>         <dbl> <dbl>    <int>
1 Large    Orthographic  0.208 0.224      252
2 Large    Semantic     -0.666 0.167      252
3 Small    Orthographic  0.509 0.216      252
4 Small    Semantic     -0.504 0.201      252
```

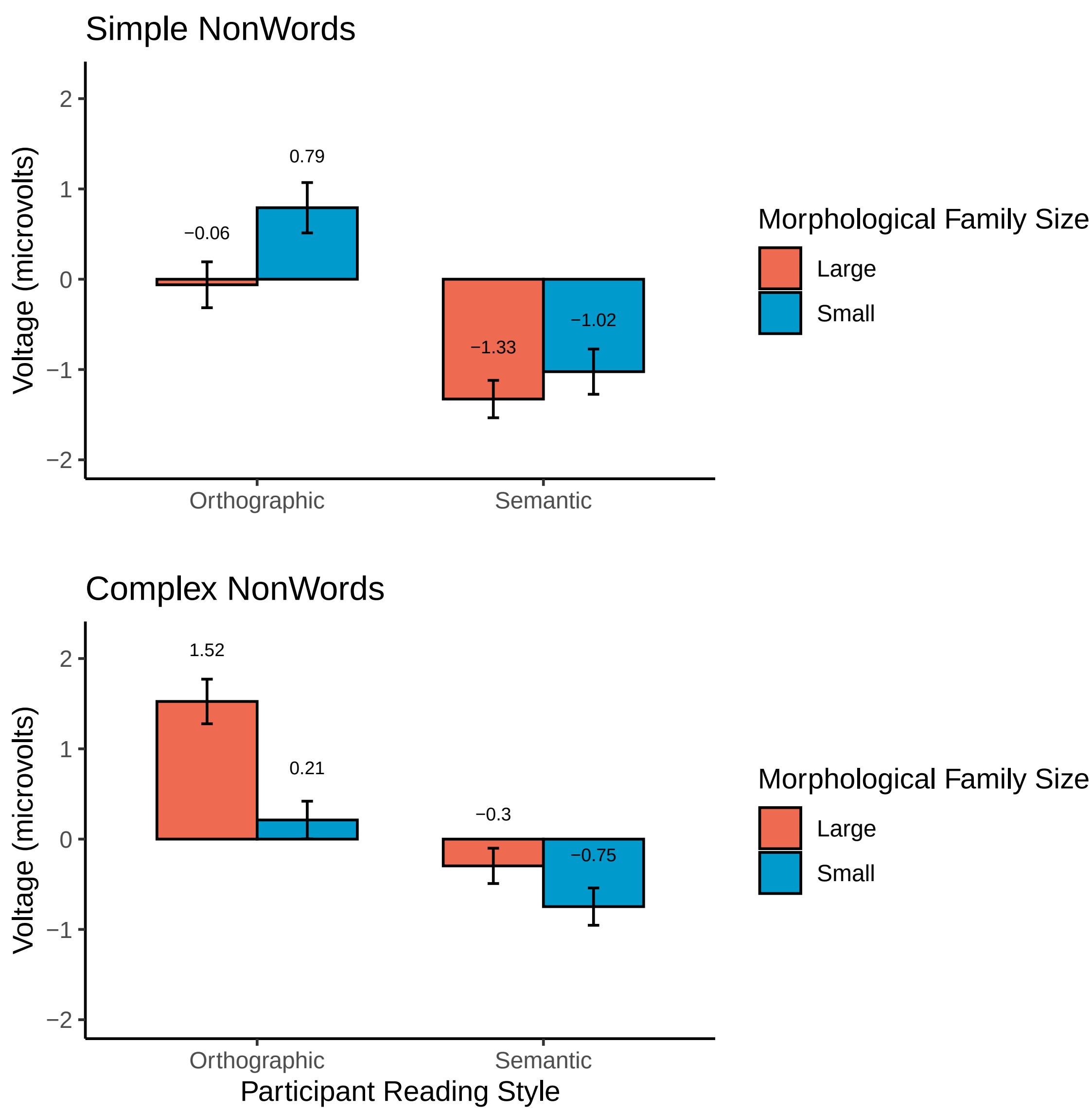
```r
(nw_smp.cond.means <- n250_nwsmpl |>
    group_by(fam_size, lang_type) |>
    summarise(mean = mean(value),
              se = sem(value),
              num_stim = n()))
```

```
# A tibble: 4 x 5
# Groups:   fam_size [2]
  fam_size lang_type       mean    se num_stim
  <chr>    <chr>          <dbl> <dbl>    <int>
1 Large    Orthographic -0.0614 0.254      252
2 Large    Semantic     -1.33   0.207      252
3 Small    Orthographic  0.792  0.279      252
4 Small    Semantic     -1.02   0.250      252
```

```r
(nw_cpx.cond.means <- n250_nwcplx |>
    group_by(fam_size, lang_type) |>
    summarise(mean = mean(value),
              se = sem(value),
              num_stim = n()))
```

```
# A tibble: 4 x 5
# Groups:   fam_size [2]
  fam_size lang_type      mean    se num_stim
  <chr>    <chr>         <dbl> <dbl>    <int>
1 Large    Orthographic  1.52  0.247      252
2 Large    Semantic     -0.296 0.196      252
3 Small    Orthographic  0.211 0.208      252
4 Small    Semantic     -0.748 0.206      252
```

Barplots



10. Now we can compute the ANOVA for each of the three datasets.

```
library(ez)

ezANOVA(data = n250_words,
        dv = value,
        wid = SubjID,
        within = .(fam_size, anteriority, laterality),
        within_full = .(fam_size, anteriority, laterality, chlabel),
        between = lang_type,
        type = 3)
```

```
Warning: Converting "SubjID" to factor for ANOVA.

Warning: Converting "fam_size" to factor for ANOVA.

Warning: Converting "anteriority" to factor for ANOVA.
```

```
Warning: Converting "laterality" to factor for ANOVA.

Warning: Converting "lang_type" to factor for ANOVA.

Warning: Data is unbalanced (unequal N per group). Make sure you specified a
well-considered value for the type argument to ezANOVA().

Warning: Collapsing data to cell means first using variables supplied to
"within_full", then collapsing the resulting means to means for the cells
supplied to "within".

$ANOVA
                                  Effect DFn DFd          F         p p<.05
2                              lang_type   1  53 1.34052920 0.2521318
3                               fam_size   1  53 0.44055247 0.5097330
5                            anteriority   2 106 3.54036213 0.0324800     *
7                             laterality   2 106 1.86750178 0.1595571
4                     lang_type:fam_size   1  53 0.02668433 0.8708615
6                  lang_type:anteriority   2 106 0.17547672 0.8393002
8                   lang_type:laterality   2 106 0.01315056 0.9869371
9                   fam_size:anteriority   2 106 0.78886227 0.4570104
11                   fam_size:laterality   2 106 0.05490062 0.9466061
13                anteriority:laterality   4 212 1.22349077 0.3018535
10         lang_type:fam_size:anteriority   2 106 1.37578636 0.2571147
12          lang_type:fam_size:laterality   2 106 2.15737581 0.1206787
14       lang_type:anteriority:laterality   4 212 1.09698544 0.3590346
15        fam_size:anteriority:laterality   4 212 0.81524150 0.5166772
16 lang_type:fam_size:anteriority:laterality   4 212 1.15610296 0.3312942
            ges
2  1.538873e-02
3  1.230218e-03
5  9.016435e-03
7  9.871855e-04
4  7.460070e-05
6  4.507590e-04
8  6.958375e-06
9  4.283027e-04
11 1.164972e-05
13 5.331596e-04
10 7.467276e-04
12 4.575837e-04
14 4.780588e-04
15 9.993534e-05
16 1.417135e-04

$`Mauchly's Test for Sphericity`
                                  Effect         W            p p<.05
5                            anteriority 0.4058906 6.586204e-11     *
6                  lang_type:anteriority 0.4058906 6.586204e-11     *
7                             laterality 0.9804224 5.980593e-01
8                   lang_type:laterality 0.9804224 5.980593e-01
9                   fam_size:anteriority 0.4879182 7.889015e-09     *
10        lang_type:fam_size:anteriority 0.4879182 7.889015e-09     *
```

```
11                         fam_size:laterality 0.5493066 1.718639e-07     *
12              lang_type:fam_size:laterality 0.5493066 1.718639e-07     *
13                     anteriority:laterality 0.5443236 2.681494e-04     *
14           lang_type:anteriority:laterality 0.5443236 2.681494e-04     *
15            fam_size:anteriority:laterality 0.7448430 8.716674e-02
16 lang_type:fam_size:anteriority:laterality 0.7448430 8.716674e-02

$`Sphericity Corrections`
                                      Effect       GGe      p[GG] p[GG]<.05
5                                anteriority 0.6273095 0.05493706
6                      lang_type:anteriority 0.6273095 0.73326668
7                                 laterality 0.9807984 0.16042064
8                       lang_type:laterality 0.9807984 0.98595966
9                       fam_size:anteriority 0.6613399 0.41054799
10            lang_type:fam_size:anteriority 0.6613399 0.25335612
11                       fam_size:laterality 0.6893255 0.88669122
12             lang_type:fam_size:laterality 0.6893255 0.13828384
13                    anteriority:laterality 0.7707591 0.30308999
14          lang_type:anteriority:laterality 0.7707591 0.35295625
15           fam_size:anteriority:laterality 0.8855715 0.50422514
16 lang_type:fam_size:anteriority:laterality 0.8855715 0.33047616
         HFe      p[HF] p[HF]<.05
5  0.6353169 0.05432862
6  0.6353169 0.73635212
7  1.0181179 0.15955713
8  1.0181179 0.98693714
9  0.6717135 0.41231797
10 0.6717135 0.25365302
11 0.7017170 0.89014371
12 0.7017170 0.13759516
13 0.8237879 0.30306897
14 0.8237879 0.35473182
15 0.9567943 0.51214862
16 0.9567943 0.33106571
```

```r
ezANOVA(data = n250_nwsmpl,
        dv = value,
        wid = SubjID,
        within = .(fam_size, anteriority, laterality),
        within_full = .(fam_size, anteriority, laterality, chlabel),
        between = lang_type,
        type = 3)
```

```
Warning: Converting "SubjID" to factor for ANOVA.

Warning: Converting "fam_size" to factor for ANOVA.

Warning: Converting "anteriority" to factor for ANOVA.

Warning: Converting "laterality" to factor for ANOVA.

Warning: Converting "lang_type" to factor for ANOVA.
```

```
Warning: Data is unbalanced (unequal N per group). Make sure you specified a
well-considered value for the type argument to ezANOVA().

Warning: Collapsing data to cell means first using variables supplied to
"within_full", then collapsing the resulting means to means for the cells
supplied to "within".

$ANOVA
                                       Effect DFn DFd           F          p
2                                   lang_type   1  53 3.542030221 0.06532792
3                                    fam_size   1  53 0.644034405 0.42583771
5                                 anteriority   2 106 4.626995884 0.01184252
7                                  laterality   2 106 1.194697087 0.30683935
4                          lang_type:fam_size   1  53 0.039047341 0.84411083
6                       lang_type:anteriority   2 106 1.197556263 0.30598260
8                        lang_type:laterality   2 106 0.001765241 0.99823635
9                        fam_size:anteriority   2 106 0.202446342 0.81704535
11                        fam_size:laterality   2 106 1.057745855 0.35087359
13                     anteriority:laterality   4 212 1.120027994 0.34800591
10             lang_type:fam_size:anteriority   2 106 0.005155040 0.99485847
12              lang_type:fam_size:laterality   2 106 0.797918163 0.45295096
14           lang_type:anteriority:laterality   4 212 0.333479506 0.85523982
15            fam_size:anteriority:laterality   4 212 0.551179728 0.69836176
16 lang_type:fam_size:anteriority:laterality   4 212 2.874532340 0.02389738
   p<.05          ges
2        3.627006e-02
3        2.835068e-03
5      * 7.943032e-03
7        5.517226e-04
4        1.723471e-04
6        2.067986e-03
8        8.156545e-07
9        1.481022e-04
11       2.770158e-04
13       5.133859e-04
10       3.771780e-06
12       2.089831e-04
14       1.529117e-04
15       1.010517e-04
16     * 5.267842e-04

$`Mauchly's Test for Sphericity`
                                       Effect         W            p p<.05
5                                 anteriority 0.4979627 1.340038e-08     *
6                       lang_type:anteriority 0.4979627 1.340038e-08     *
7                                  laterality 0.9704010 4.578588e-01
8                        lang_type:laterality 0.9704010 4.578588e-01
9                        fam_size:anteriority 0.3768510 9.558976e-12     *
10             lang_type:fam_size:anteriority 0.3768510 9.558976e-12     *
11                        fam_size:laterality 0.9079524 8.121582e-02
12              lang_type:fam_size:laterality 0.9079524 8.121582e-02
13                     anteriority:laterality 0.6944730 2.752149e-02     *
14           lang_type:anteriority:laterality 0.6944730 2.752149e-02     *
15            fam_size:anteriority:laterality 0.7589110 1.161195e-01
```

```
16 lang_type:fam_size:anteriority:laterality 0.7589110 1.161195e-01

$`Sphericity Corrections`
                                      Effect       GGe      p[GG] p[GG]<.05
5                                anteriority 0.6657624 0.02495448         *
6                      lang_type:anteriority 0.6657624 0.29232836
7                                laterality 0.9712519 0.30607721
8                       lang_type:laterality 0.9712519 0.99791877
9                       fam_size:anteriority 0.6160864 0.70574406
10            lang_type:fam_size:anteriority 0.6160864 0.96787810
11                       fam_size:laterality 0.9157110 0.34629651
12             lang_type:fam_size:laterality 0.9157110 0.44334128
13                    anteriority:laterality 0.8312614 0.34496999
14          lang_type:anteriority:laterality 0.8312614 0.82115825
15           fam_size:anteriority:laterality 0.8968233 0.67930081
16 lang_type:fam_size:anteriority:laterality 0.8968233 0.02891560         *
         HFe      p[HF] p[HF]<.05
5  0.6764506 0.02436629         *
6  0.6764506 0.29299050
7  1.0076405 0.30683935
8  1.0076405 0.99823635
9  0.6233343 0.70857606
10 0.6233343 0.96899505
11 0.9468388 0.34806292
12 0.9468388 0.44699517
13 0.8935714 0.34630701
14 0.8935714 0.83479611
15 0.9699622 0.69302247
16 0.9699622 0.02525749         *
```

```r
ezANOVA(data = n250_nwcplx,
        dv = value,
        wid = SubjID,
        within = .(fam_size, anteriority, laterality),
        within_full = .(fam_size, anteriority, laterality, chlabel),
        between = lang_type,
        type = 3)
```

```
Warning: Converting "SubjID" to factor for ANOVA.

Warning: Converting "fam_size" to factor for ANOVA.

Warning: Converting "anteriority" to factor for ANOVA.

Warning: Converting "laterality" to factor for ANOVA.

Warning: Converting "lang_type" to factor for ANOVA.

Warning: Data is unbalanced (unequal N per group). Make sure you specified a
well-considered value for the type argument to ezANOVA().
```

```
Warning: Collapsing data to cell means first using variables supplied to
"within_full", then collapsing the resulting means to means for the cells
supplied to "within".

$ANOVA
                                      Effect DFn DFd         F           p
2                                  lang_type   1  53 3.3436494 0.073094421
3                                   fam_size   1  53 4.6270020 0.036050679
5                                anteriority   2 106 5.7346857 0.004317225
7                                 laterality   2 106 1.6938118 0.188754330
4                         lang_type:fam_size   1  53 0.9233787 0.340953002
6                      lang_type:anteriority   2 106 1.4888084 0.230321048
8                       lang_type:laterality   2 106 0.3198610 0.726948474
9                       fam_size:anteriority   2 106 3.6236036 0.030043695
11                       fam_size:laterality   2 106 0.5137177 0.599749053
13                    anteriority:laterality   4 212 1.8778827 0.115463833
10             lang_type:fam_size:anteriority   2 106 1.8738263 0.158585377
12              lang_type:fam_size:laterality   2 106 0.2350721 0.790924821
14          lang_type:anteriority:laterality   4 212 1.3512588 0.252096900
15           fam_size:anteriority:laterality   4 212 1.5984002 0.175911774
16 lang_type:fam_size:anteriority:laterality   4 212 0.9403921 0.441455408
   p<.05          ges
2        0.0332048269
3      * 0.0156023905
5      * 0.0135747170
7        0.0010131866
4        0.0031530383
6        0.0035599750
8        0.0001914885
9      * 0.0026908737
11       0.0002320468
13       0.0012165335
10       0.0013933065
12       0.0001061957
14       0.0008756738
15       0.0005226730
16       0.0003075721

$`Mauchly's Test for Sphericity`
                                      Effect         W            p p<.05
5                                anteriority 0.4901323 8.874556e-09     *
6                      lang_type:anteriority 0.4901323 8.874556e-09     *
7                                 laterality 0.8896657 4.785176e-02     *
8                       lang_type:laterality 0.8896657 4.785176e-02     *
9                       fam_size:anteriority 0.4073456 7.228354e-11     *
10            lang_type:fam_size:anteriority 0.4073456 7.228354e-11     *
11                       fam_size:laterality 0.8204428 5.824457e-03     *
12             lang_type:fam_size:laterality 0.8204428 5.824457e-03     *
13                    anteriority:laterality 0.6779696 1.808040e-02     *
14          lang_type:anteriority:laterality 0.6779696 1.808040e-02     *
15           fam_size:anteriority:laterality 0.5500376 3.314039e-04     *
16 lang_type:fam_size:anteriority:laterality 0.5500376 3.314039e-04     *

$`Sphericity Corrections`
```

	Effect	GGe	p[GG]	p[GG]<.05	HFe	p[HF]	p[HF]<.05
5	anteriority	0.6623097	0.01222813	*	0.6727521	0.01183915	*
6	lang_type:anteriority	0.6623097	0.23188373		0.6727521	0.23199151	
7	laterality	0.9006296	0.19200026		0.9303745	0.19105868	
8	lang_type:laterality	0.9006296	0.70423076		0.9303745	0.71131368	
9	fam_size:anteriority	0.6278826	0.05195645		0.6359290	0.05135199	
10	lang_type:fam_size:anteriority	0.6278826	0.17379468		0.6359290	0.17353983	
11	fam_size:laterality	0.8477757	0.57001230		0.8728266	0.57524085	
12	lang_type:fam_size:laterality	0.8477757	0.75426524		0.8728266	0.76084643	
13	anteriority:laterality	0.8277485	0.12892882		0.8895010	0.12393409	
14	lang_type:anteriority:laterality	0.8277485	0.25764137		0.8895010	0.25574985	
15	fam_size:anteriority:laterality	0.8234969	0.18718235		0.8845777	0.18324619	
16	lang_type:fam_size:anteriority:laterality	0.8234969	0.42889859		0.8845777	0.43359041	