EFA

```r
set.seed(42)

library(rcompanion) # effect size calculation
library(igraph)
```

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```r
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```r
library(QuantPsyc) # for the multivariate normality test
```

```
## Loading required package: boot

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:igraph':
##
##     as_data_frame, groups, union

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Loading required package: purrr

##
## Attaching package: 'purrr'

## The following objects are masked from 'package:igraph':
##
##     compose, simplify

## Loading required package: MASS

##
## Attaching package: 'MASS'
```

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

##
## Attaching package: 'QuantPsyc'

## The following object is masked from 'package:base':
##
##     norm
library(dunn.test)
library(nFactors) # for the scree plot

## Loading required package: lattice

##
## Attaching package: 'lattice'

## The following object is masked from 'package:boot':
##
##     melanoma

##
## Attaching package: 'nFactors'

## The following object is masked from 'package:lattice':
##
##     parallel
library(psych) # for PA FA

##
## Attaching package: 'psych'

## The following object is masked from 'package:boot':
##
##     logit

## The following object is masked from 'package:rcompanion':
##
##     phi
library(caret) # highly correlated features removal

## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v forcats   1.0.0      v stringr   1.5.1
## v lubridate 1.9.3      v tibble    3.2.1
## v readr     2.1.5      v tidyr     1.3.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x lubridate::%--%()       masks igraph::%--%()
## x ggplot2::%+%()          masks psych::%+%()
## x ggplot2::alpha()        masks psych::alpha()
## x tibble::as_data_frame() masks dplyr::as_data_frame(), igraph::as_data_frame()
## x purrr::compose()        masks igraph::compose()
## x tidyr::crossing()       masks igraph::crossing()
## x dplyr::filter()         masks stats::filter()
## x dplyr::lag()            masks stats::lag()
## x caret::lift()           masks purrr::lift()
## x MASS::select()          masks dplyr::select()
## x purrr::simplify()       masks igraph::simplify()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(paletteer) # color palettes

library(conflicted) # to resolve QuantPsyc x dplyr conflicts
conflict_prefer("select", "dplyr")
```

```
## [conflicted] Will prefer dplyr::select over any other package.
```

```r
conflict_prefer("filter", "dplyr")
```

```
## [conflicted] Will prefer dplyr::filter over any other package.
```

Helpers

```r
analyze_distributions <- function(data_factors_long, variable) {
  factors <- levels(data_factors_long$factor)

  print(table(data_factors_long[[variable]], useNA = "ifany") / length(factors))

  plot_all <- data_factors_long %>%
    ggplot(aes(x = factor_score, y = !!sym(variable))) +
    geom_boxplot() +
    facet_grid(factor ~ .) +
    labs(x = "factor score")
  print(plot_all)

  plot <- data_factors_long %>%
    drop_na(!!sym(variable)) %>%
    ggplot(aes(x = factor_score, y = !!sym(variable))) +
    geom_boxplot() +
    facet_grid(factor ~ .) +
    labs(x = "factor score") +
    theme_bw()
  ggsave(paste(c("distr", variable, ".pdf"), collapse = ""))
  print(plot)

  # formula <- reformulate(variable, "factor_score")
```

```r
chi2 <- numeric()
p_val <- numeric()
epsilon2 <- numeric()
epsilon2_lci <- numeric()
epsilon2_uci <- numeric()
min_p_values <- numeric()

for (f in factors) {
  data <- data_factors_long %>% filter(factor == f)

  cat(
    "\nTest for the significance of differences in",
    variable, "over", f, ":\n\n"
  )

  kw <- kruskal.test(data$factor_score, data[[variable]])

  dunn <- dunn.test(
    data$factor_score, data[[variable]],
    altp = TRUE, method = "bonferroni"
  )

  e2_test <- epsilonSquared(data$factor_score, data[[variable]], ci = TRUE)

  e2 <- e2_test[[1]]
  e2_lci <- e2_test[[2]]
  e2_uci <- e2_test[[3]]
  cat("epsilon2 = ", e2, "(95% CI:", e2_lci, "-", e2_uci, ")\n")

  min_p_values <- c(min_p_values, min(dunn$altP.adjusted))
  chi2 <- c(chi2, kw$statistic[[1]])
  p_val <- c(p_val, kw$p.value)
  epsilon2 <- c(epsilon2, e2)
  epsilon2_lci <- c(epsilon2_lci, e2_lci)
  epsilon2_uci <- c(epsilon2_uci, e2_uci)
}

cat("\n")
print(
  data.frame(
    factor = factors,
    chi2 = chi2,
    kruskal_p = p_val,
    epsilon2_lci = epsilon2_lci,
    epsilon2 = epsilon2,
    epsilon2_uci = epsilon2_uci
  ) %>% mutate(
    across(c(epsilon2, epsilon2_lci, epsilon2_uci), ~ round(.x, 3))
  ) %>%
    mutate(across(kruskal_p, ~ case_when(
      .x < 0.0001 ~ "< 0.0001",
      .x < 0.001 ~ "< 0.001",
      .x < 0.01 ~ "< 0.01",
```

```
        .x < 0.05 ~ "< 0.05",
        .default = as.character(round(.x, 2))
      ))) %>%
      mutate(across(chi2, ~ round(.x, 2)))
  )

  cat(
    "\np < 5e-2 found in:",
    factors[min_p_values < 0.05],
    "\np < 1e-2 found in:",
    factors[min_p_values < 0.01],
    "\np < 1e-3 found in:",
    factors[min_p_values < 0.001],
    "\np < 1e-4 found in:",
    factors[min_p_values < 0.0001], "\n"
  )
}

data_factor_bind <- function(data, fa_fit) {
  data_factors <- bind_cols(data, fa_fit$scores %>% as.data.frame())
  colnames(data_factors) <- prettify_feat_name_vector(colnames(data_factors))

  fnames <- colnames(fa_fit$loadings)

  data_factors_long <- data_factors %>%
    pivot_longer(
      any_of(fnames),
      names_to = "factor", values_to = "factor_score"
    ) %>%
    mutate(across(
      factor,
      ~ factor(.x, levels = fnames)
    )) %>%
    select(
      all_of(1:(.firstnonmetacolumn - 1)), factor, factor_score, everything()
    )

  data_factors_longer <- data_factors_long %>% pivot_longer(
    all_of((.firstnonmetacolumn + 2):ncol(data_factors_long)),
    names_to = "feat", values_to = "feat_value"
  )

  return(list(
    data = data_factors,
    long = data_factors_long,
    feat_long = data_factors_longer
  ))
}
```

Load and tidy data

```
pretty_names <- read_csv("../feat_name_mapping.csv")
```

```
## Rows: 85 Columns: 2
## -- Column specification -----------------------------------------------------
## Delimiter: ","
## chr (2): name_orig, name_pretty
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
prettify_feat_name <- function(x) {
  name <- pull(pretty_names %>%
    filter(name_orig == x), name_pretty)
  if (length(name) == 1) {
    return(name)
  } else {
    return(x)
  }
}


prettify_feat_name_vector <- function(x) {
  map(
    x,
    prettify_feat_name
  ) %>% unlist()
}



data <- read_csv("../measurements/measurements.csv")
```

```
## Rows: 753 Columns: 108
## -- Column specification -----------------------------------------------------
## Delimiter: ","
## chr (20): fpath, KUK_ID, FileName, FileFormat, FolderPath, subcorpus, Source...
## dbl (85): RuleAbstractNouns, RuleAmbiguousRegards, RuleAnaphoricReferences, ...
## lgl  (3): ClarityPursuit, SyllogismBased, Bindingness
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
.firstnonmetacolumn <- 17

data_no_nas <- data %>%
  select(!c(
    fpath,
    # KUK_ID,
    # FileName,
    FolderPath,
    # subcorpus,
    DocumentTitle,
    ClarityPursuit,
    Readability,
    SyllogismBased,
    SourceDB
  )) %>%
  # replace -1s in variation coefficients with NAs
  mutate(across(c(
```

```
  `RuleDoubleAdpos.max_allowable_distance.v`,
  `RuleTooManyNegations.max_negation_frac.v`,
  `RuleTooManyNegations.max_allowable_negations.v`,
  `RuleTooManyNominalConstructions.max_noun_frac.v`,
  `RuleTooManyNominalConstructions.max_allowable_nouns.v`,
  `RuleCaseRepetition.max_repetition_count.v`,
  `RuleCaseRepetition.max_repetition_frac.v`,
  `RulePredSubjDistance.max_distance.v`,
  `RulePredObjDistance.max_distance.v`,
  `RuleInfVerbDistance.max_distance.v`,
  `RuleMultiPartVerbs.max_distance.v`,
  `RuleLongSentences.max_length.v`,
  `RulePredAtClauseBeginning.max_order.v`,
  `mattr.v`,
  `maentropy.v`
), ~ na_if(.x, -1))) %>%
# replace NAs with 0s
replace_na(list(
  RuleGPcoordovs = 0,
  RuleGPdeverbaddr = 0,
  RuleGPpatinstr = 0,
  RuleGPdeverbsubj = 0,
  RuleGPadjective = 0,
  RuleGPpatbenperson = 0,
  RuleGPwordorder = 0,
  RuleDoubleAdpos = 0,
  RuleDoubleAdpos.max_allowable_distance.v = 0,
  RuleAmbiguousRegards = 0,
  RuleReflexivePassWithAnimSubj = 0,
  RuleTooManyNegations = 0,
  RuleTooManyNegations.max_negation_frac.v = 0,
  RuleTooManyNegations.max_allowable_negations.v = 0,
  RuleTooManyNominalConstructions.max_noun_frac.v = 0,
  RuleTooManyNominalConstructions.max_allowable_nouns.v = 0,
  RuleFunctionWordRepetition = 0,
  RuleCaseRepetition.max_repetition_count.v = 0,
  RuleCaseRepetition.max_repetition_frac.v = 0,
  RuleWeakMeaningWords = 0,
  RuleAbstractNouns = 0,
  RuleRelativisticExpressions = 0,
  RuleConfirmationExpressions = 0,
  RuleRedundantExpressions = 0,
  RuleTooLongExpressions = 0,
  RuleAnaphoricReferences = 0,
  RuleLiteraryStyle = 0,
  RulePassive = 0,
  RulePredSubjDistance = 0,
  RulePredSubjDistance.max_distance.v = 0,
  RulePredObjDistance = 0,
  RulePredObjDistance.max_distance.v = 0,
  RuleInfVerbDistance = 0,
  RuleInfVerbDistance.max_distance.v = 0,
  RuleMultiPartVerbs = 0,
```

```r
    RuleMultiPartVerbs.max_distance.v = 0,
    RuleLongSentences.max_length.v = 0,
    RulePredAtClauseBeginning.max_order.v = 0,
    RuleVerbalNouns = 0,
    RuleDoubleComparison = 0,
    RuleWrongValencyCase = 0,
    RuleWrongVerbonominalCase = 0,
    RuleIncompleteConjunction = 0
)) %>%
# merge GPs
mutate(
  GPs = RuleGPcoordovs +
    RuleGPdeverbaddr +
    RuleGPpatinstr +
    RuleGPdeverbsubj +
    RuleGPadjective +
    RuleGPpatbenperson +
    RuleGPwordorder
) %>%
select(!c(
  RuleGPcoordovs,
  RuleGPdeverbaddr,
  RuleGPpatinstr,
  RuleGPdeverbsubj,
  RuleGPadjective,
  RuleGPpatbenperson,
  RuleGPwordorder
)) %>%
# norm data expected to correlate with text length
mutate(across(c(
  GPs,
  RuleDoubleAdpos,
  RuleAmbiguousRegards,
  RuleFunctionWordRepetition,
  RuleWeakMeaningWords,
  RuleAbstractNouns,
  RuleRelativisticExpressions,
  RuleConfirmationExpressions,
  RuleRedundantExpressions,
  RuleTooLongExpressions,
  RuleAnaphoricReferences,
  RuleLiteraryStyle,
  RulePassive,
  RuleVerbalNouns,
  RuleDoubleComparison,
  RuleWrongValencyCase,
  RuleWrongVerbonominalCase,
  RuleIncompleteConjunction,
  num_hapax,
  RuleReflexivePassWithAnimSubj,
  RuleTooManyNominalConstructions,
  RulePredSubjDistance,
  RuleMultiPartVerbs,
```

```r
    RulePredAtClauseBeginning
  ), ~ .x / word_count)) %>%
  mutate(across(c(
    RuleTooFewVerbs,
    RuleTooManyNegations,
    RuleCaseRepetition,
    RuleLongSentences,
    RulePredObjDistance,
    RuleInfVerbDistance
  ), ~ .x / sent_count)) %>%
  # replace NAs with medians
  mutate(across(c(
    RuleDoubleAdpos.max_allowable_distance,
    RuleTooManyNegations.max_negation_frac,
    RuleTooManyNegations.max_allowable_negations,
    RulePredSubjDistance.max_distance,
    RulePredObjDistance.max_distance,
    RuleInfVerbDistance.max_distance,
    RuleMultiPartVerbs.max_distance
  ), ~ coalesce(., median(., na.rm = TRUE))))

data_clean <- data_no_nas %>%
  # remove variables identified as text-length dependent
  select(!c(
    RuleTooFewVerbs,
    RuleTooManyNegations,
    RuleTooManyNominalConstructions,
    RuleCaseRepetition,
    RuleLongSentences,
    RulePredAtClauseBeginning,
    syllab_count,
    char_count
  )) %>%
  # remove variables identified as unreliable
  select(!c(
    RuleAmbiguousRegards,
    RuleFunctionWordRepetition,
    RuleDoubleComparison,
    RuleWrongValencyCase,
    RuleWrongVerbonominalCase
  )) %>%
  # remove further variables belonging to the 'acceptability' category
  select(!c(RuleIncompleteConjunction)) %>%
  # remove artificially limited variables
  select(!c(
    RuleCaseRepetition.max_repetition_frac,
    RuleCaseRepetition.max_repetition_frac.v
  )) %>%
  # remove variables with too many NAs
  select(!c(
    RuleDoubleAdpos.max_allowable_distance,
    RuleDoubleAdpos.max_allowable_distance.v
  )) %>%
```

```r
  mutate(across(c(
    class,
    FileFormat,
    subcorpus,
    DocumentVersion,
    LegalActType,
    Objectivity,
    AuthorType,
    RecipientType,
    RecipientIndividuation,
    Anonymized
  ), ~ as.factor(.x)))

# no NAs should be present now
data_clean[!complete.cases(data_clean[.firstnonmetacolumn:ncol(data_clean)]), ]
```

```
## # A tibble: 0 x 77
## # i 77 variables: KUK_ID <chr>, FileName <chr>, FileFormat <fct>,
## #   subcorpus <fct>, SourceID <chr>, DocumentVersion <fct>,
## #   ParentDocumentID <chr>, LegalActType <fct>, Objectivity <fct>,
## #   Bindingness <lgl>, AuthorType <fct>, RecipientType <fct>,
## #   RecipientIndividuation <fct>, Anonymized <fct>, Recipient Type <chr>,
## #   class <fct>, RuleAbstractNouns <dbl>, RuleAnaphoricReferences <dbl>,
## #   RuleCaseRepetition.max_repetition_count <dbl>, ...
```

```r
colnames(data_clean) <- prettify_feat_name_vector(colnames(data_clean))
```

Important features identification

```r
feature_importances <- read_csv("../importance_measures/featcomp.csv")
```

```
## Rows: 61 Columns: 21
## -- Column specification ---------------------------------------------------
## Delimiter: ","
## chr  (2): Variable, Sign
## dbl (15): Importance, p_value, estimate, wilcox_p, wilcox_r, kw_p, kw_chi2, ...
## lgl  (4): selected_pval, wilcox_sel, kw_sel, selected_reg
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
selected_features_names <- feature_importances %>%
  filter(kw_sel) %>%
  pull(Variable)
```

Correlations

See Levshina (2015: 353–54).

```r
analyze_correlation <- function(data) {
  cor_matrix <- cor(data)

  cor_tibble_long <- cor_matrix %>%
```

```r
    as_tibble() %>%
    mutate(feat1 = rownames(cor_matrix)) %>%
    pivot_longer(!feat1, names_to = "feat2", values_to = "cor") %>%
    mutate(abs_cor = abs(cor))

  cor_matrix_upper <- cor_matrix
  cor_matrix_upper[lower.tri(cor_matrix_upper)] <- 0

  cor_tibble_long_upper <- cor_matrix_upper %>%
    as_tibble() %>%
    mutate(feat1 = rownames(cor_matrix)) %>%
    pivot_longer(!feat1, names_to = "feat2", values_to = "cor") %>%
    mutate(abs_cor = abs(cor)) %>%
    filter(feat1 != feat2 & abs_cor > 0)

  list(
    cor_matrix = cor_matrix,
    cor_matrix_upper = cor_matrix_upper,
    cor_tibble_long = cor_tibble_long,
    cor_tibble_long_upper = cor_tibble_long_upper
  )
}

data_purish <- data_clean %>%
  # remove readability metrics as they're conceptually different
  # to the remaining features
  select(!c(ari, cli, fkgl, fre, gf, smog)) %>%
  # remove atl as it heavily reflects phenomena
  # that cannot be influenced by the author
  select(!atl) %>%
  select(any_of(selected_features_names))
```

High correlations

```r
.hcorrcutoff <- 0.9

analyze_correlation(data_purish)$cor_tibble_long %>%
  filter(feat1 != feat2 & abs_cor > .hcorrcutoff) %>%
  arrange(feat1, -abs_cor) %>%
  print(n = 100)
```

```
## # A tibble: 4 x 4
##   feat1     feat2        cor abs_cor
##   <chr>     <chr>      <dbl>   <dbl>
## 1 hpoint    wordcount  0.958   0.958
## 2 maentropy mattr      0.964   0.964
## 3 mattr     maentropy  0.964   0.964
## 4 wordcount hpoint     0.958   0.958
```

```r
high_correlations <- findCorrelation(
  cor(data_purish),
  verbose = TRUE, cutoff = .hcorrcutoff
)
```

```
## Compare row 6  and column  5 with corr  0.958
##   Means:  0.183 vs 0.183 so flagging column 5
## Compare row 19  and column  14 with corr  0.964
##   Means:  0.17 vs 0.183 so flagging column 14
## All correlations <= 0.9
```

```r
names(data_purish)[high_correlations]
```

```
## [1] "hpoint" "mattr"
```

```r
data_pureish_striphigh <- data_purish %>% select(!all_of(high_correlations))

analyze_correlation(data_pureish_striphigh)$cor_tibble_long %>%
  filter(feat1 != feat2 & abs_cor > .hcorrcutoff) %>%
  arrange(feat1, -abs_cor) %>%
  print(n = 100)
```

```
## # A tibble: 0 x 4
## # i 4 variables: feat1 <chr>, feat2 <chr>, cor <dbl>, abs_cor <dbl>
```

Low correlations

```r
# 0.35 instead of 0.3 otherwise the FA bootstrapping would freeze
.lcorrcutoff <- 0.35

low_correlating_features <- analyze_correlation(data_pureish_striphigh)$
  cor_tibble_long %>%
  filter(feat1 != feat2) %>%
  group_by(feat1) %>%
  summarize(max_cor = max(abs_cor)) %>%
  filter(max_cor < .lcorrcutoff) %>%
  pull(feat1)

feature_importances %>%
  filter(Variable %in% low_correlating_features) %>%
  pull(Variable)
```
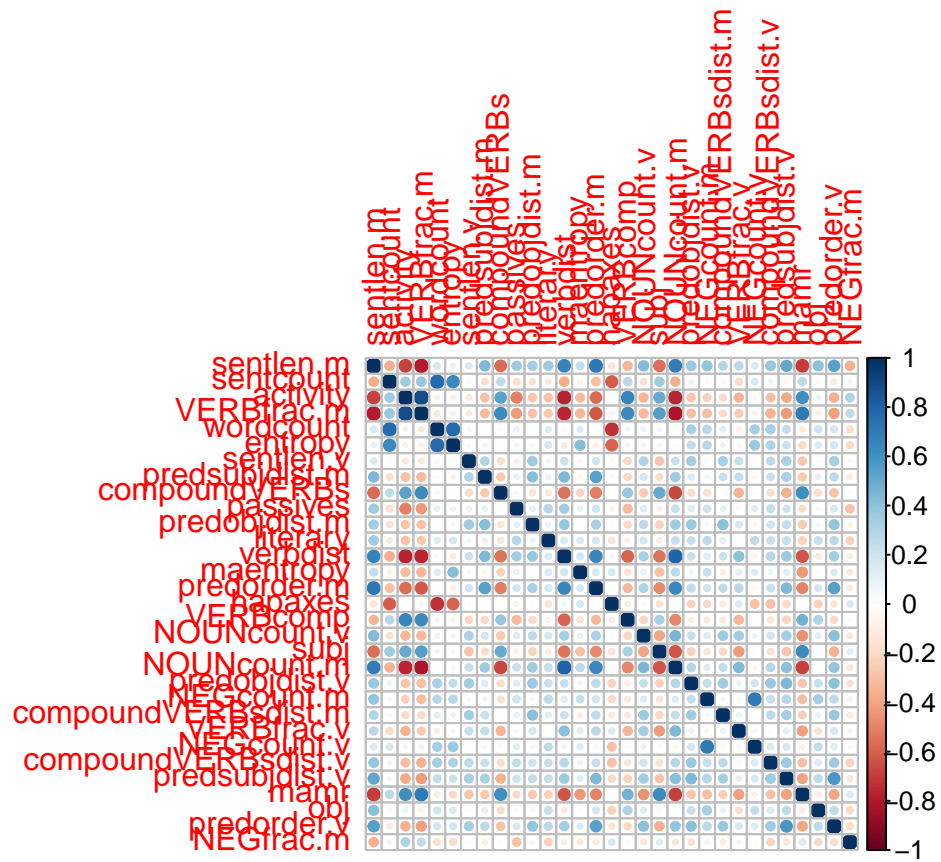
```
## [1] "anaphoricrefs"     "extrcaseexprs"    "caserepcount.v"
## [4] "redundexprs"       "relativisticexprs" "VERBcompdist.m"
## [7] "NOUNfrac.v"        "verbalNOUNs"      "abstractNOUNs"
```

```r
data_pure <- data_pureish_striphigh %>%
  select(!any_of(low_correlating_features))

colnames(data_pure) <- prettify_feat_name_vector(colnames(data_pure))
```
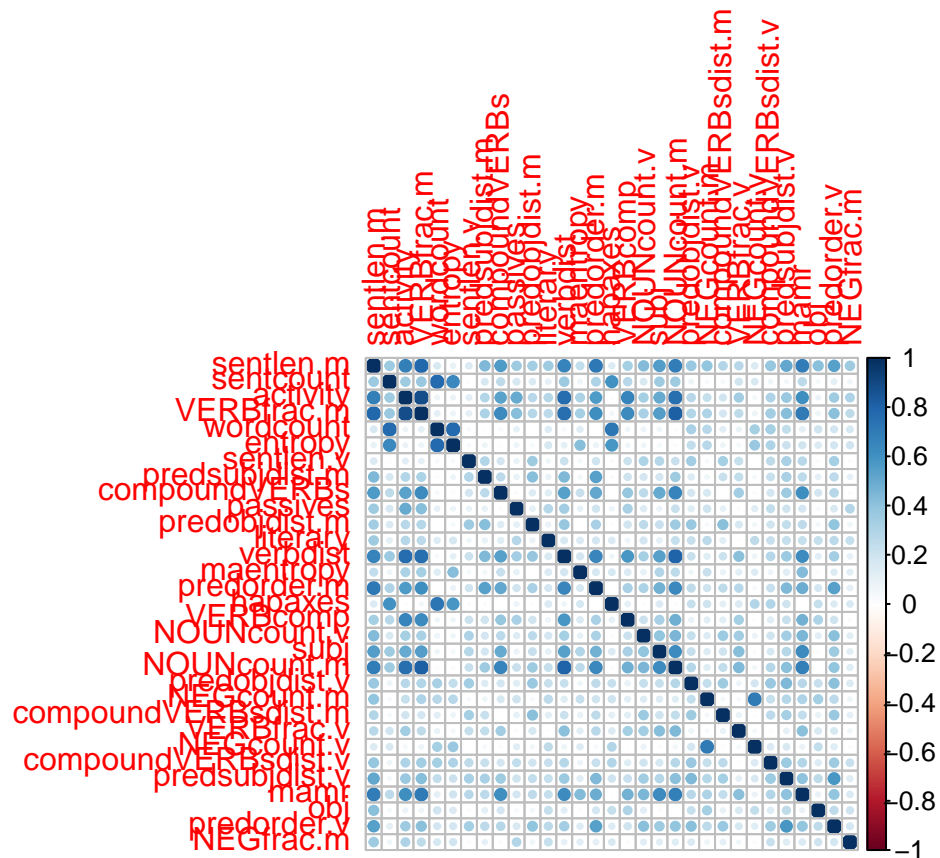
Visualisation

```r
corrplot(cor(data_pure))
```

```
corrplot(abs(cor(data_pure)))
```

```r
my_colors <- paletteer::paletteer_d("ggthemes::Classic_10_Medium")

network_edges <- analyze_correlation(data_pure)$cor_tibble_long_upper %>%
  filter(abs_cor > .lcorrcutoff)

network <- graph_from_data_frame(
  network_edges,
  directed = FALSE
)
E(network)$weight <- network_edges$abs_cor
network_communities <- cluster_optimal(network)

network_membership <- membership(network_communities)

plot(
  network,
  layout = layout.fruchterman.reingold,
  vertex.color = map(
    network_communities$membership,
    function(x) my_colors[x]
  ) %>% unlist(use.names = FALSE),
  vertex.size = 6,
  vertex.label.color = "black",
  vertex.label.cex = 0.7
)
```
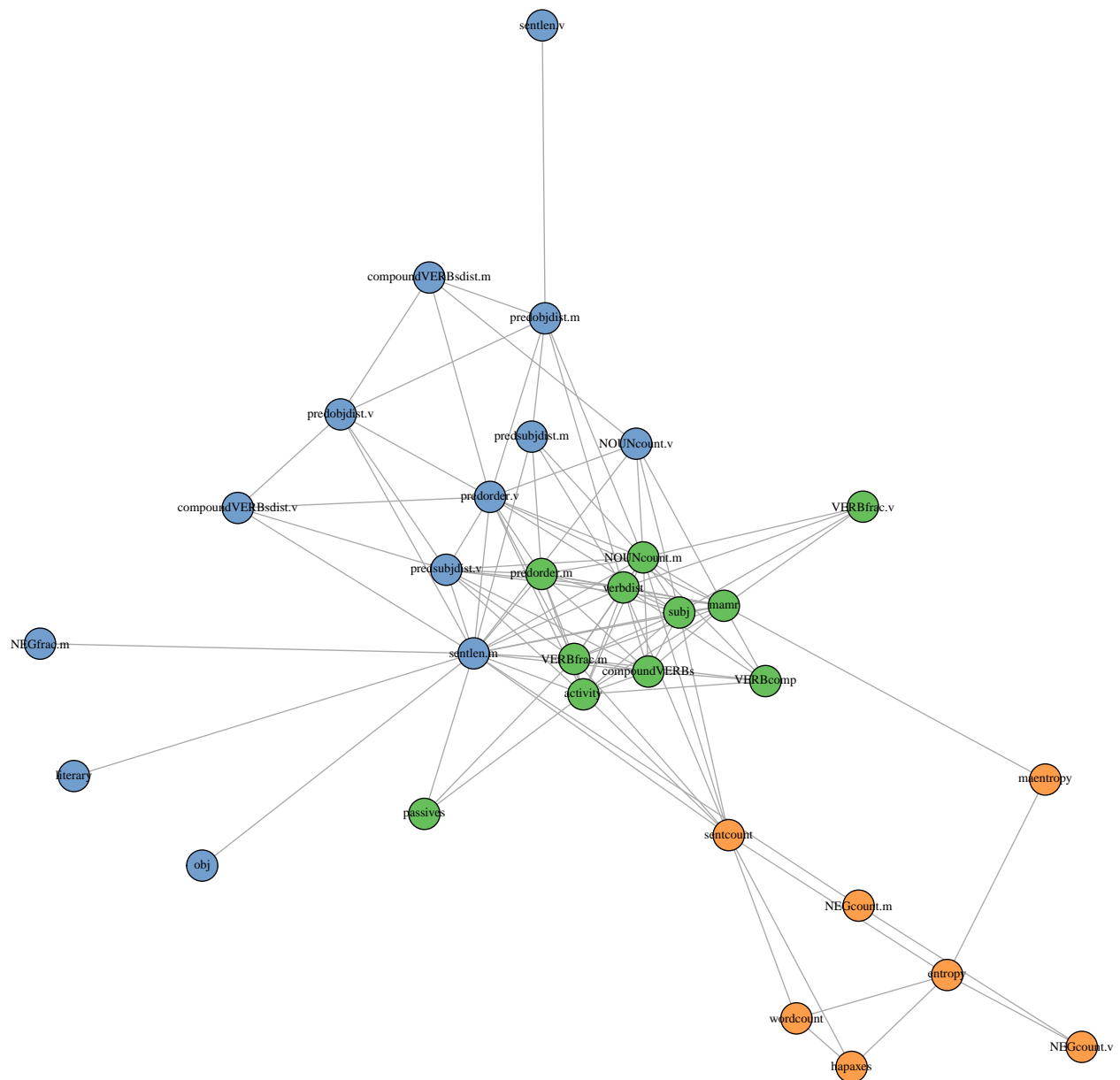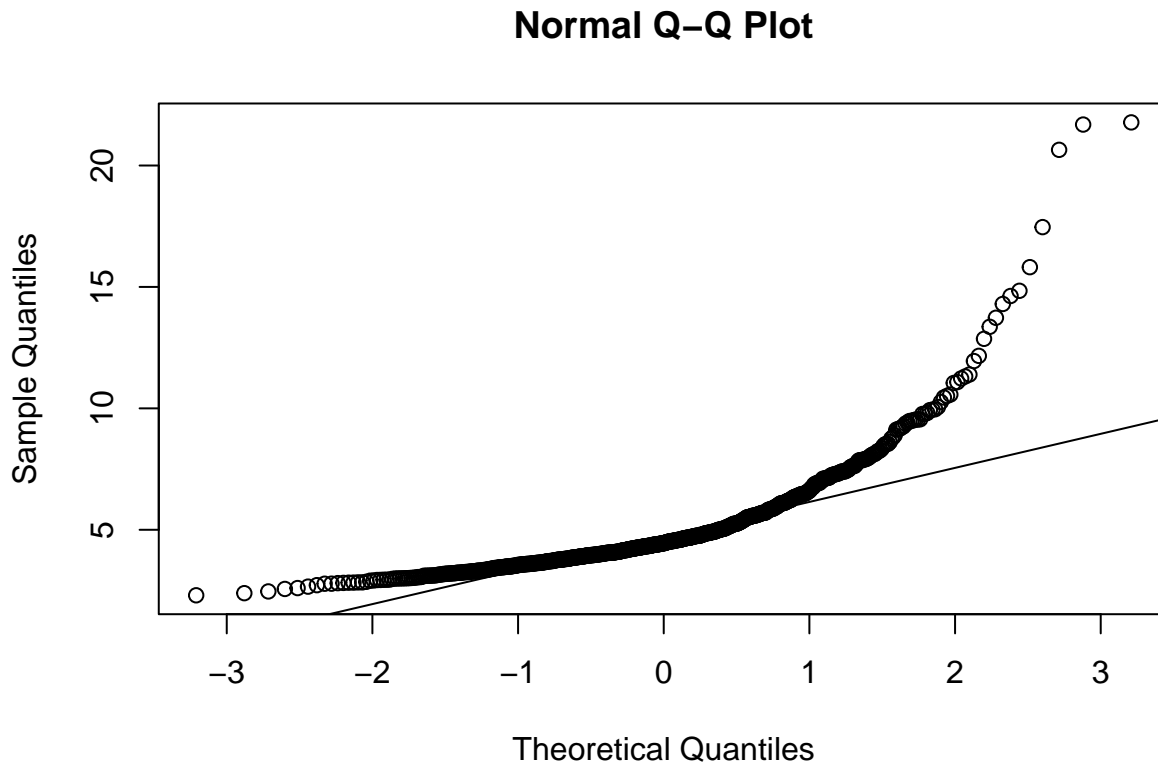
Scaling

```
data_scaled <- data_pure %>%
  mutate(across(seq_along(data_pure), ~ scale(.x)[, 1]))
```

Check for normality

```
mult.norm(data_scaled %>% as.data.frame())$mult.test
```

```
##          Beta-hat      kappa p-val
## Skewness 1006.915 126367.8448     0
## Kurtosis 2532.745    457.9503     0
```

```r
mardia(data_scaled)
```

Normal Q–Q Plot


```
## Call: mardia(x = data_scaled)
##
## Mardia tests of multivariate skew and kurtosis
## Use describe(x) the to get univariate tests
## n.obs = 753   num.vars =  31
## b1p =  1006.92   skew =  126367.8  with probability  <=  0
##   small sample skew =  126902.9  with probability <=  0
## b2p =  2532.75   kurtosis =  457.95  with probability <=  0
```

Low (null) p-values show that we can reject the hypothesis that the data would be in a multivariate normal distribution. I.e. the distribution isn't multivariate normal.

Check for goodness of data

```r
data_scaled %>%
  cor() %>%
  det()
```

```
## [1] 1.192791e-10
```

```r
KMO(data_scaled)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = data_scaled)
## Overall MSA =  0.85
## MSA for each item =
##         sentlen.m          sentcount          activity          VERBfrac.m
##              0.90               0.72              0.90              0.86
```

```
##          wordcount              entropy           sentlen.v      predsubjdist.m
##               0.73                 0.74                0.83                0.81
##       compoundVERBs             passives       predobjdist.m            literary
##               0.89                 0.84                0.82                0.89
##           verbdist            maentropy          predorder.m             hapaxes
##               0.93                 0.59                0.87                0.81
##           VERBcomp          NOUNcount.v                subj         NOUNcount.m
##               0.86                 0.90                0.95                0.90
##        predobjdist.v         NEGcount.m compoundVERBsdist.m          VERBfrac.v
##               0.91                 0.71                0.83                0.81
##         NEGcount.v compoundVERBsdist.v       predsubjdist.v                mamr
##               0.69                 0.93                0.92                0.91
##                obj          predorder.v           NEGfrac.m
##               0.69                 0.87                0.65
```

```
bartlett.test(data_scaled)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  data_scaled
## Bartlett's K-squared = 2.5035e-13, df = 30, p-value = 1
```
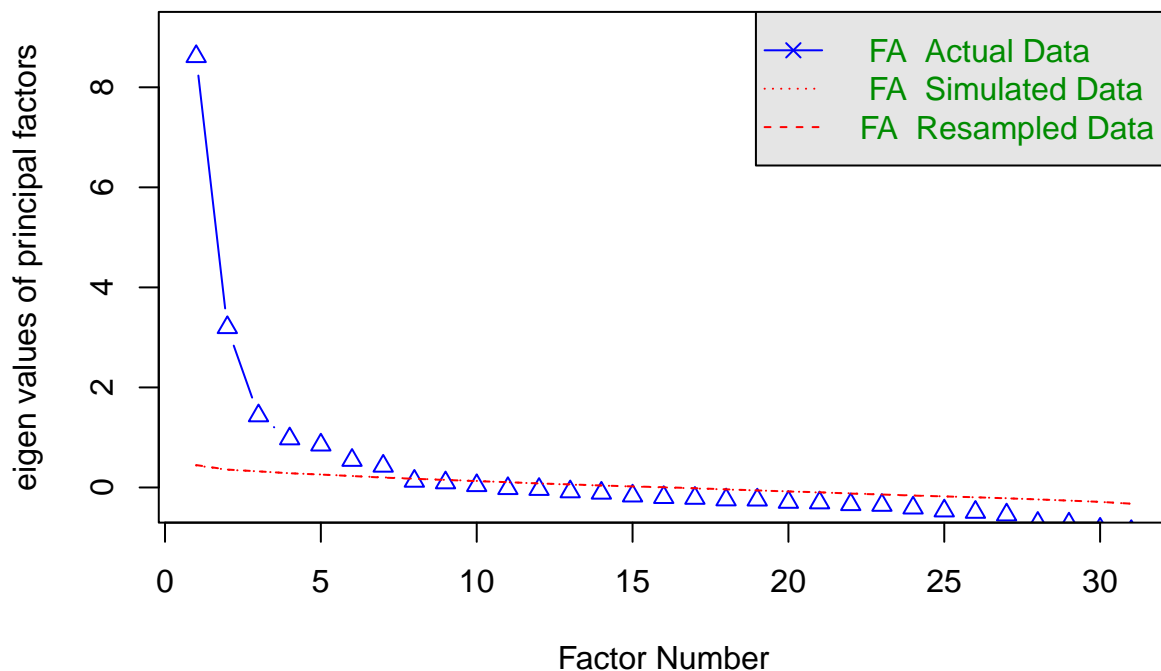
Good and broad FA!

No. of vectors

```
fa_parallel_broad <- fa.parallel(data_scaled, fm = "pa", fa = "fa", n.iter = 20)
```

Parallel Analysis Scree Plots


```
## Parallel analysis suggests that the number of factors =  7  and the number of components =  NA
```

```
fa_parallel_broad_df <- data.frame(
  factor = seq_along(data_scaled),
  actual = fa_parallel_broad$fa.values,
  simulated = fa_parallel_broad$fa.sim
) %>%
  pivot_longer(!factor, names_to = "data", values_to = "eigenvalue")

fa_parallel_broad_df %>%
  ggplot(aes(x = factor, y = eigenvalue, linetype = data)) +
  geom_line() +
  geom_point(
    data = fa_parallel_broad_df %>% filter(data == "actual"),
    mapping = aes(x = factor, y = eigenvalue)
  ) +
  labs(x = "factor number", y = "eigen values of principal factors") +
  theme_bw()
```
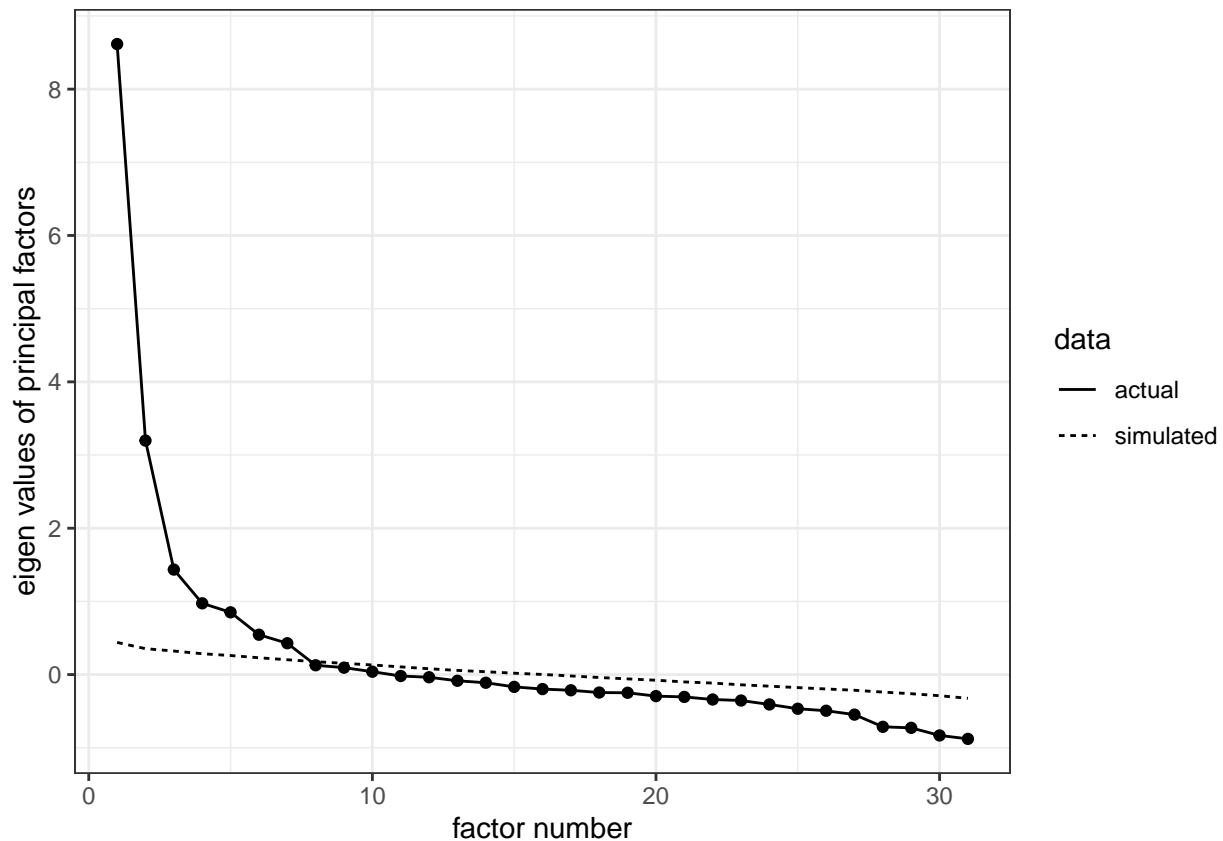


```
ggsave("scree.pdf", height = 4, width = 6)
```

Model

```
set.seed(42)

fa_broad <- fa(
  data_scaled,
  nfactors = 7,
```

```
  fm = "pa",
  rotate = "promax",
  oblique.scores = TRUE,
  scores = "tenBerge",
  n.iter = 100
)
```

```
## Loading required namespace: GPArotation
```

```
fa_broad
```

```
## Factor Analysis with confidence intervals using method = fa(r = data_scaled, nfactors = 7, n.iter =
##       scores = "tenBerge", fm = "pa", oblique.scores = TRUE)
## Factor Analysis using method =  pa
## Call: fa(r = data_scaled, nfactors = 7, n.iter = 100, rotate = "promax",
##       scores = "tenBerge", fm = "pa", oblique.scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                       PA1   PA2   PA3   PA5   PA6   PA4   PA7   h2    u2 com
## sentlen.m           -0.68 -0.05  0.01 -0.21  0.04  0.38 -0.01 0.92 0.080 1.8
## sentcount            0.15  0.98  0.01  0.27 -0.10 -0.18  0.02 0.93 0.065 1.3
## activity             0.76 -0.03  0.10  0.46  0.01  0.29  0.09 0.90 0.100 2.1
## VERBfrac.m           0.89 -0.05  0.19  0.31 -0.03  0.08  0.06 0.90 0.096 1.4
## wordcount           -0.13  0.95  0.00  0.01  0.01 -0.02 -0.07 0.89 0.112 1.1
## entropy              0.09  0.75  0.06 -0.08  0.04 -0.07 -0.45 0.87 0.135 1.7
## sentlen.v            0.07  0.00  0.77  0.26  0.01 -0.14  0.02 0.46 0.535 1.3
## predsubjdist.m      -0.37 -0.01  0.27  0.05 -0.05  0.09  0.30 0.35 0.647 3.0
## compoundVERBs        1.03 -0.13  0.29 -0.36  0.01 -0.22  0.06 0.70 0.296 1.6
## passives            -0.02 -0.09 -0.02 -0.76  0.11 -0.26  0.05 0.56 0.441 1.3
## predobjdist.m       -0.04 -0.08  0.62 -0.04 -0.07 -0.07  0.15 0.39 0.613 1.2
## literary             0.00 -0.05  0.08 -0.30  0.15  0.14 -0.09 0.24 0.758 2.4
## verbdist            -0.86  0.00  0.02 -0.12 -0.06 -0.22  0.10 0.80 0.197 1.2
## maentropy           -0.22  0.02 -0.18 -0.11  0.04 -0.02 -0.64 0.50 0.499 1.5
## predorder.m         -0.71 -0.05  0.09  0.02 -0.04  0.21  0.15 0.63 0.373 1.3
## hapaxes              0.12 -0.79  0.06  0.01 -0.03 -0.09 -0.22 0.68 0.318 1.2
## VERBcomp             0.57  0.02 -0.02  0.15 -0.13  0.52 -0.02 0.60 0.403 2.2
## NOUNcount.v         -0.13 -0.08  0.46  0.00  0.00  0.02 -0.16 0.35 0.654 1.5
## subj                 0.54  0.15 -0.17 -0.10  0.06 -0.03  0.30 0.56 0.436 2.1
## NOUNcount.m         -0.90  0.04  0.02 -0.03 -0.13 -0.05 -0.07 0.81 0.193 1.1
## predobjdist.v        0.04  0.15  0.53 -0.06  0.07  0.05  0.00 0.40 0.604 1.3
## NEGcount.m          -0.06 -0.08 -0.06  0.14  1.00  0.15 -0.01 0.95 0.054 1.1
## compoundVERBsdist.m  0.21 -0.03  0.75 -0.12 -0.07 -0.06  0.09 0.42 0.578 1.3
## VERBfrac.v          -0.44 -0.04  0.17  0.25 -0.02 -0.19 -0.15 0.35 0.651 2.6
## NEGcount.v           0.21  0.07  0.02  0.02  0.74  0.06 -0.07 0.59 0.412 1.2
## compoundVERBsdist.v -0.09  0.23  0.30 -0.19  0.03  0.00  0.03 0.33 0.670 2.9
## predsubjdist.v      -0.21  0.10  0.41 -0.02  0.10  0.14  0.03 0.46 0.536 2.1
## mamr                 0.67 -0.03 -0.09 -0.04 -0.03  0.00  0.36 0.74 0.255 1.6
## obj                  0.02 -0.06 -0.04  0.07  0.15  0.84  0.04 0.69 0.312 1.1
## predorder.v         -0.09 -0.02  0.56 -0.05  0.06  0.17 -0.02 0.53 0.470 1.3
## NEGfrac.m           -0.06 -0.03 -0.03  0.60  0.31 -0.17  0.17 0.41 0.592 1.9
##
##                       PA1  PA2  PA3  PA5  PA6  PA4  PA7
## SS loadings          6.71 3.09 2.77 1.83 1.74 1.53 1.24
## Proportion Var       0.22 0.10 0.09 0.06 0.06 0.05 0.04
## Cumulative Var       0.22 0.32 0.41 0.46 0.52 0.57 0.61
## Proportion Explained 0.35 0.16 0.15 0.10 0.09 0.08 0.07
```

```
## Cumulative Proportion 0.35 0.52 0.66 0.76 0.85 0.93 1.00
##
##   With factor correlations of
##         PA1   PA2   PA3   PA5   PA6   PA4   PA7
## PA1   1.00  0.12 -0.61  0.37 -0.27 -0.13  0.17
## PA2   0.12  1.00  0.15 -0.27  0.31  0.30 -0.08
## PA3  -0.61  0.15  1.00 -0.32  0.26  0.30 -0.12
## PA5   0.37 -0.27 -0.32  1.00 -0.38 -0.34  0.03
## PA6  -0.27  0.31  0.26 -0.38  1.00  0.22 -0.18
## PA4  -0.13  0.30  0.30 -0.34  0.22  1.00 -0.07
## PA7   0.17 -0.08 -0.12  0.03 -0.18 -0.07  1.00
##
## Mean item complexity =  1.6
## Test of the hypothesis that 7 factors are sufficient.
##
## df null model =  465  with the objective function =  22.85 with Chi Square =  16927.71
## df of  the model are 269  and the objective function was  2.74
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## The harmonic n.obs is  753 with the empirical chi square  517.31  with prob <  6.9e-18
## The total n.obs was  753  with Likelihood Chi Square =  2020.55  with prob <  1.4e-265
##
## Tucker Lewis Index of factoring reliability =  0.815
## RMSEA index =  0.093  and the 90 % confidence intervals are  0.089 0.097
## BIC =  238.68
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                      PA1  PA2  PA3  PA5  PA6  PA4
## Correlation of (regression) scores with factors      0.98 0.98 0.93 0.93 0.98 0.93
## Multiple R square of scores with factors             0.97 0.96 0.86 0.87 0.96 0.87
## Minimum correlation of possible factor scores        0.94 0.92 0.73 0.74 0.91 0.75
##                                                      PA7
## Correlation of (regression) scores with factors      0.9
## Multiple R square of scores with factors             0.8
## Minimum correlation of possible factor scores        0.6
##
##   Coefficients and bootstrapped confidence intervals
##                      low  PA1 upper   low  PA2 upper   low  PA3 upper   low
## sentlen.m          -0.76 -0.68 -0.55 -0.10 -0.05 -0.02 -0.05  0.01  0.10 -0.27
## sentcount           0.10  0.15  0.22  0.93  0.98  1.04 -0.02  0.01  0.06  0.21
## activity            0.64  0.76  0.84 -0.06 -0.03  0.01  0.04  0.10  0.14  0.40
## VERBfrac.m          0.72  0.89  0.99 -0.08 -0.05  0.00  0.09  0.19  0.23  0.25
## wordcount          -0.16 -0.13 -0.07  0.91  0.95  0.98 -0.03  0.00  0.05 -0.03
## entropy            -0.05  0.09  0.15  0.71  0.75  0.82 -0.06  0.06  0.12 -0.13
## sentlen.v          -0.06  0.07  0.13 -0.06  0.00  0.09  0.59  0.77  0.89  0.20
## predsubjdist.m     -0.54 -0.37 -0.22 -0.05 -0.01  0.06  0.08  0.27  0.43 -0.03
## compoundVERBs       0.76  1.03  1.17 -0.17 -0.13 -0.04  0.12  0.29  0.37 -0.44
## passives           -0.10 -0.02  0.04 -0.14 -0.09 -0.03 -0.10 -0.02  0.04 -0.84
## predobjdist.m      -0.23 -0.04  0.19 -0.18 -0.08  0.01  0.42  0.62  0.84 -0.16
## literary           -0.11  0.00  0.09 -0.11 -0.05  0.02 -0.04  0.08  0.16 -0.40
## verbdist           -0.95 -0.86 -0.73 -0.04  0.00  0.03 -0.02  0.02  0.08 -0.27
## maentropy          -0.40 -0.22 -0.10 -0.05  0.02  0.12 -0.33 -0.18 -0.06 -0.20
```

```
## predorder.m        -0.81 -0.71 -0.57 -0.12 -0.05  0.01 -0.03  0.09  0.22 -0.08
## hapaxes            -0.04  0.12  0.21 -0.83 -0.79 -0.71 -0.09  0.06  0.15 -0.05
## VERBcomp            0.46  0.57  0.66 -0.03  0.02  0.07 -0.08 -0.02  0.05  0.08
## NOUNcount.v        -0.29 -0.13 -0.01 -0.16 -0.08  0.01  0.27  0.46  0.59 -0.09
## subj                0.44  0.54  0.65  0.08  0.15  0.21 -0.23 -0.17 -0.10 -0.19
## NOUNcount.m        -1.02 -0.90 -0.72 -0.01  0.04  0.08 -0.05  0.02  0.12 -0.10
## predobjdist.v      -0.10  0.04  0.17  0.05  0.15  0.26  0.40  0.53  0.67 -0.15
## NEGcount.m         -0.11 -0.06  0.01 -0.12 -0.08 -0.03 -0.11 -0.06  0.00  0.06
## compoundVERBsdist.m 0.07  0.21  0.34 -0.11 -0.03  0.05  0.58  0.75  0.92 -0.20
## VERBfrac.v         -0.58 -0.44 -0.33 -0.11 -0.04  0.04  0.04  0.17  0.25  0.14
## NEGcount.v          0.13  0.21  0.26  0.02  0.07  0.13 -0.05  0.02  0.07 -0.06
## compoundVERBsdist.v -0.21 -0.09  0.01  0.16  0.23  0.32  0.18  0.30  0.41 -0.30
## predsubjdist.v     -0.35 -0.21 -0.08  0.03  0.10  0.19  0.24  0.41  0.56 -0.10
## mamr                0.56  0.67  0.80 -0.10 -0.03  0.02 -0.16 -0.09  0.00 -0.12
## obj                -0.03  0.02  0.11 -0.10 -0.06 -0.02 -0.10 -0.04  0.06  0.01
## predorder.v        -0.27 -0.09  0.03 -0.09 -0.02  0.06  0.33  0.56  0.72 -0.15
## NEGfrac.m          -0.13 -0.06  0.06 -0.10 -0.03  0.02 -0.11 -0.03  0.06  0.50
##                      PA5 upper   low   PA6 upper   low   PA4 upper   low   PA7
## sentlen.m          -0.21 -0.17  0.01  0.04  0.10  0.32  0.38  0.46 -0.09 -0.01
## sentcount           0.27  0.32 -0.16 -0.10 -0.06 -0.23 -0.18 -0.14 -0.07  0.02
## activity            0.46  0.53 -0.02  0.01  0.06  0.24  0.29  0.38  0.02  0.09
## VERBfrac.m          0.31  0.38 -0.10 -0.03  0.03  0.02  0.08  0.18 -0.06  0.06
## wordcount           0.01  0.06 -0.03  0.01  0.05 -0.06 -0.02  0.02 -0.21 -0.07
## entropy            -0.08 -0.01 -0.02  0.04  0.12 -0.13 -0.07  0.01 -0.59 -0.45
## sentlen.v           0.26  0.36 -0.08  0.01  0.08 -0.20 -0.14 -0.07 -0.07  0.02
## predsubjdist.m      0.05  0.15 -0.18 -0.05  0.06 -0.08  0.09  0.32  0.03  0.30
## compoundVERBs      -0.36 -0.27 -0.09  0.01  0.09 -0.30 -0.22 -0.12 -0.10  0.06
## passives           -0.76 -0.69  0.06  0.11  0.17 -0.33 -0.26 -0.20 -0.02  0.05
## predobjdist.m      -0.04  0.05 -0.16 -0.07  0.02 -0.17 -0.07  0.01 -0.14  0.15
## literary           -0.30 -0.20  0.06  0.15  0.26  0.05  0.14  0.24 -0.18 -0.09
## verbdist           -0.12 -0.03 -0.11 -0.06 -0.02 -0.30 -0.22 -0.17  0.00  0.10
## maentropy          -0.11 -0.01 -0.06  0.04  0.16 -0.12 -0.02  0.10 -0.89 -0.64
## predorder.m         0.02  0.12 -0.17 -0.04  0.10  0.05  0.21  0.40 -0.09  0.15
## hapaxes             0.01  0.11 -0.11 -0.03  0.04 -0.16 -0.09 -0.01 -0.30 -0.22
## VERBcomp            0.15  0.23 -0.19 -0.13 -0.06  0.43  0.52  0.66 -0.13 -0.02
## NOUNcount.v         0.00  0.10 -0.10  0.00  0.11 -0.07  0.02  0.13 -0.33 -0.16
## subj               -0.10 -0.02  0.00  0.06  0.14 -0.10 -0.03  0.03  0.16  0.30
## NOUNcount.m        -0.03  0.03 -0.20 -0.13 -0.07 -0.15 -0.05  0.01 -0.24 -0.07
## predobjdist.v      -0.06  0.03 -0.03  0.07  0.18 -0.06  0.05  0.14 -0.09  0.00
## NEGcount.m          0.14  0.20  0.81  1.00  1.20  0.11  0.15  0.24 -0.16 -0.01
## compoundVERBsdist.m -0.12 -0.05 -0.14 -0.07  0.00 -0.15 -0.06  0.02 -0.02  0.09
## VERBfrac.v          0.25  0.37 -0.13 -0.02  0.09 -0.28 -0.19 -0.08 -0.28 -0.15
## NEGcount.v          0.02  0.10  0.58  0.74  0.97  0.01  0.06  0.16 -0.19 -0.07
## compoundVERBsdist.v -0.19 -0.08 -0.03  0.03  0.11 -0.10  0.00  0.10 -0.05  0.03
## predsubjdist.v     -0.02  0.07  0.02  0.10  0.18  0.06  0.14  0.24 -0.10  0.03
## mamr               -0.04  0.02 -0.11 -0.03  0.03 -0.07  0.00  0.06  0.18  0.36
## obj                 0.07  0.13  0.09  0.15  0.25  0.76  0.84  0.97 -0.07  0.04
## predorder.v        -0.05  0.04 -0.01  0.06  0.14  0.07  0.17  0.31 -0.14 -0.02
## NEGfrac.m           0.60  0.69  0.19  0.31  0.44 -0.28 -0.17 -0.08  0.04  0.17
##                    upper
## sentlen.m           0.04
## sentcount           0.07
## activity            0.21
## VERBfrac.m          0.26
```

```
## wordcount          -0.01
## entropy            -0.34
## sentlen.v           0.16
## predsubjdist.m      0.70
## compoundVERBs       0.33
## passives            0.13
## predobjdist.m       0.49
## literary            0.00
## verbdist            0.17
## maentropy          -0.49
## predorder.m         0.41
## hapaxes            -0.10
## VERBcomp            0.11
## NOUNcount.v         0.00
## subj                0.49
## NOUNcount.m         0.02
## predobjdist.v       0.11
## NEGcount.m          0.10
## compoundVERBsdist.m 0.24
## VERBfrac.v         -0.02
## NEGcount.v          0.03
## compoundVERBsdist.v 0.12
## predsubjdist.v      0.19
## mamr                0.62
## obj                 0.13
## predorder.v         0.10
## NEGfrac.m           0.28
##
##   Interfactor correlations and bootstrapped confidence intervals
##          lower estimate upper
## PA1-PA2 -0.56    0.118  0.44
## PA1-PA3 -1.15   -0.610  0.73
## PA1-PA5 -0.83    0.370  0.61
## PA1-PA6 -0.71   -0.268  0.46
## PA1-PA4 -0.49   -0.132  0.33
## PA1-PA7 -0.40    0.171  0.27
## PA2-PA3 -0.08    0.146  0.35
## PA2-PA5 -0.41   -0.266  0.62
## PA2-PA6 -0.22    0.313  0.63
## PA2-PA4 -0.13    0.297  0.59
## PA2-PA7 -0.27   -0.075  0.37
## PA3-PA5 -0.45   -0.324  0.71
## PA3-PA6 -0.22    0.259  0.65
## PA3-PA4 -0.13    0.301  0.57
## PA3-PA7 -0.24   -0.119  0.35
## PA5-PA6 -0.64   -0.377  0.72
## PA5-PA4 -0.50   -0.336  0.65
## PA5-PA7 -0.28    0.034  0.31
## PA6-PA4 -0.31    0.225  0.53
## PA6-PA7 -0.29   -0.184  0.38
## PA4-PA7 -0.35   -0.070  0.44
```

Healthiness diagnostics

```r
fa_broad$loadings[] %>%
  as_tibble() %>%
  mutate(feat = colnames(data_scaled)) %>%
  select(feat, everything()) %>%
  pivot_longer(!feat) %>%
  mutate(value = abs(value)) %>%
  group_by(feat) %>%
  summarize(maxload = max(value)) %>%
  arrange(maxload)
```

```
## # A tibble: 31 x 2
##    feat               maxload
##    <chr>               <dbl>
##  1 literary            0.303
##  2 compoundVERBsdist.v 0.304
##  3 predsubjdist.m      0.370
##  4 predsubjdist.v      0.409
##  5 VERBfrac.v          0.444
##  6 NOUNcount.v         0.457
##  7 predobjdist.v       0.534
##  8 subj                0.536
##  9 predorder.v         0.555
## 10 VERBcomp            0.567
## # i 21 more rows
```

```r
fa_broad$communality %>% sort()
```

```
##            literary compoundVERBsdist.v         NOUNcount.v          VERBfrac.v
##           0.2416725           0.3296174           0.3462354           0.3490726
##      predsubjdist.m       predobjdist.m       predobjdist.v           NEGfrac.m
##           0.3534706           0.3870510           0.3956853           0.4080319
## compoundVERBsdist.m      predsubjdist.v           sentlen.v           maentropy
##           0.4218770           0.4636951           0.4648207           0.5005586
##          predorder.v            passives                subj           NEGcount.v
##           0.5304832           0.5593128           0.5636361           0.5882451
##            VERBcomp          predorder.m             hapaxes                 obj
##           0.5973404           0.6266506           0.6822421           0.6883013
##        compoundVERBs                mamr             verbdist         NOUNcount.m
##           0.7035859           0.7447957           0.8034091           0.8071553
##             entropy           wordcount            activity          VERBfrac.m
##           0.8654729           0.8878834           0.9002538           0.9036643
##           sentlen.m           sentcount          NEGcount.m
##           0.9199822           0.9349906           0.9460857
```

```r
fa_broad$communality[fa_broad$communality < 0.5] %>% names()
```

```
##  [1] "sentlen.v"           "predsubjdist.m"      "predobjdist.m"
##  [4] "literary"            "NOUNcount.v"         "predobjdist.v"
##  [7] "compoundVERBsdist.m" "VERBfrac.v"          "compoundVERBsdist.v"
## [10] "predsubjdist.v"      "NEGfrac.m"
```

```r
fa_broad$complexity %>% sort()
```

```
##           wordcount         NOUNcount.m                 obj          NEGcount.m
```

```
##        1.050148           1.068934              1.095678         1.111812
##        NEGcount.v         verbdist             predobjdist.m    hapaxes
##        1.216647           1.220211              1.232746         1.243273
##        predobjdist.v compoundVERBsdist.m        predorder.v      sentcount
##        1.257433           1.275774              1.290328         1.294705
##        sentlen.v          predorder.m           passives         VERBfrac.m
##        1.318462           1.321357              1.324733         1.379623
##        NOUNcount.v        maentropy             compoundVERBs    mamr
##        1.476835           1.486361              1.571818         1.596350
##        entropy            sentlen.m             NEGfrac.m        activity
##        1.736212           1.818851              1.916877         2.072686
##        predsubjdist.v     subj                  VERBcomp         literary
##        2.087153           2.116688              2.247216         2.422278
##        VERBfrac.v compoundVERBsdist.v           predsubjdist.m
##        2.639408           2.919631              3.017975
```

```
fa_broad$complexity[fa_broad$complexity > 2] %>% names()
```
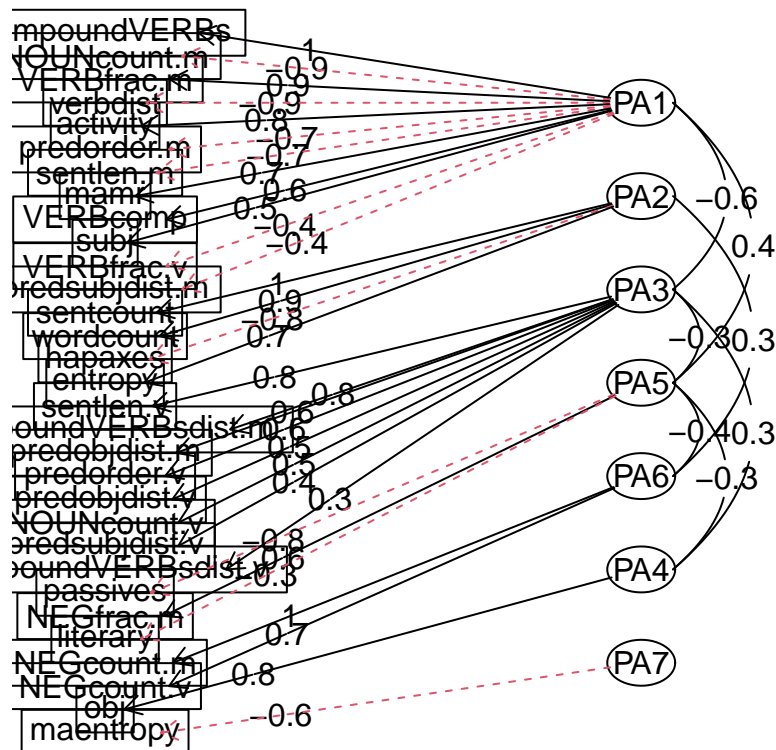
```
## [1] "activity"            "predsubjdist.m"     "literary"
## [4] "VERBcomp"            "subj"               "VERBfrac.v"
## [7] "compoundVERBsdist.v" "predsubjdist.v"
```

Loadings

Comrey and Lee (1992): loadings excelent > .70 > very good > .63 > good > .55 > fair > .45 > poor > .32

```
fa.diagram(fa_broad)
```

Factor Analysis


```
fa_broad$loadings
```

```
## 
## Loadings:
##                        PA1    PA2    PA3    PA5    PA6    PA4    PA7
## sentlen.m           -0.679                -0.215         0.376
## sentcount            0.153  0.982         0.267        -0.179
## activity             0.758         0.103  0.460         0.290
## VERBfrac.m           0.892         0.191  0.311
## wordcount           -0.126  0.946
## entropy                     0.748                              -0.448
## sentlen.v                          0.773  0.261        -0.139
## predsubjdist.m      -0.370         0.265                       0.298
## compoundVERBs        1.026 -0.126  0.287 -0.363        -0.221
## passives                                 -0.761  0.112 -0.262
## predobjdist.m                      0.615                       0.153
## literary                                 -0.303  0.152  0.138
## verbdist            -0.863                -0.125        -0.224  0.102
## maentropy           -0.219        -0.176 -0.115                -0.638
## predorder.m         -0.706                              0.207  0.152
## hapaxes              0.119 -0.789                              -0.217
## VERBcomp             0.567                0.147 -0.129  0.520
## NOUNcount.v         -0.129         0.457                       -0.156
## subj                 0.536  0.148 -0.170                       0.297
## NOUNcount.m         -0.902                       -0.129
## predobjdist.v               0.151  0.534
## NEGcount.m                               0.141  0.997  0.149
## compoundVERBsdist.m  0.214         0.754 -0.118
## VERBfrac.v          -0.444         0.165  0.249        -0.188 -0.145
## NEGcount.v           0.210                       0.740
## compoundVERBsdist.v         0.234  0.304 -0.191
## predsubjdist.v      -0.208  0.101  0.409         0.100  0.140
## mamr                 0.671                                    0.361
## obj                                      0.146  0.844
## predorder.v                        0.555                0.166
## NEGfrac.m                                0.599  0.309 -0.167  0.172
## 
##                      PA1   PA2   PA3   PA5   PA6   PA4   PA7
## SS loadings        6.625 3.220 2.930 1.874 1.786 1.668 1.150
## Proportion Var     0.214 0.104 0.095 0.060 0.058 0.054 0.037
## Cumulative Var     0.214 0.318 0.412 0.473 0.530 0.584 0.621
```

```r
for (i in 1:fa_broad$factors) {
  cat("\n----", colnames(fa_broad$loadings)[i], "----\n")

  loadings <- fa_broad$loadings[, i]
  load_df <- data.frame(loading = loadings)

  load_df_filtered <- load_df %>%
    mutate(abs_l = abs(loading)) %>%
    mutate(strng = case_when(
      abs_l > 0.70 ~ "*****",
      abs_l <= 0.70 & abs_l > 0.63 ~ "**** ",
      abs_l <= 0.63 & abs_l > 0.55 ~ "***  ",
      abs_l <= 0.55 & abs_l > 0.45 ~ "**   ",
      abs_l <= 0.45 & abs_l > 0.32 ~ "*    ",
```

```
    .default = ""
  )) %>%
  arrange(-abs_l) %>%
  filter(abs_l > 0.1)

load_df_filtered %>%
  mutate(across(c(loading, abs_l), ~ round(.x, 3))) %>%
  print()

cat("\n")
}
```

```
##
## ----- PA1 -----
##                        loading abs_l strng
## compoundVERBs            1.026 1.026 *****
## NOUNcount.m            -0.902 0.902 *****
## VERBfrac.m              0.892 0.892 *****
## verbdist               -0.863 0.863 *****
## activity                0.758 0.758 *****
## predorder.m            -0.706 0.706 *****
## sentlen.m              -0.679 0.679 ****
## mamr                    0.671 0.671 ****
## VERBcomp                0.567 0.567 ***
## subj                    0.536 0.536 **
## VERBfrac.v             -0.444 0.444 *
## predsubjdist.m         -0.370 0.370 *
## maentropy              -0.219 0.219
## compoundVERBsdist.m     0.214 0.214
## NEGcount.v              0.210 0.210
## predsubjdist.v         -0.208 0.208
## sentcount               0.153 0.153
## NOUNcount.v            -0.129 0.129
## wordcount              -0.126 0.126
## hapaxes                 0.119 0.119
##
##
## ----- PA2 -----
##                        loading abs_l strng
## sentcount               0.982 0.982 *****
## wordcount               0.946 0.946 *****
## hapaxes                -0.789 0.789 *****
## entropy                 0.748 0.748 *****
## compoundVERBsdist.v     0.234 0.234
## predobjdist.v           0.151 0.151
## subj                    0.148 0.148
## compoundVERBs          -0.126 0.126
## predsubjdist.v          0.101 0.101
##
##
## ----- PA3 -----
##                        loading abs_l strng
## sentlen.v               0.773 0.773 *****
## compoundVERBsdist.m     0.754 0.754 *****
```

```
## predobjdist.m          0.615 0.615 ***
## predorder.v            0.555 0.555 ***
## predobjdist.v          0.534 0.534 **
## NOUNcount.v            0.457 0.457 **
## predsubjdist.v         0.409 0.409 *
## compoundVERBsdist.v    0.304 0.304
## compoundVERBs          0.287 0.287
## predsubjdist.m         0.265 0.265
## VERBfrac.m             0.191 0.191
## maentropy             -0.176 0.176
## subj                  -0.170 0.170
## VERBfrac.v             0.165 0.165
## activity               0.103 0.103
##
##
## ----- PA5 -----
##                        loading abs_l strng
## passives              -0.761 0.761 *****
## NEGfrac.m              0.599 0.599 ***
## activity               0.460 0.460 **
## compoundVERBs         -0.363 0.363 *
## VERBfrac.m             0.311 0.311
## literary              -0.303 0.303
## sentcount              0.267 0.267
## sentlen.v              0.261 0.261
## VERBfrac.v             0.249 0.249
## sentlen.m             -0.215 0.215
## compoundVERBsdist.v   -0.191 0.191
## VERBcomp               0.147 0.147
## NEGcount.m             0.141 0.141
## verbdist              -0.125 0.125
## compoundVERBsdist.m   -0.118 0.118
## maentropy             -0.115 0.115
##
##
## ----- PA6 -----
##              loading abs_l strng
## NEGcount.m     0.997 0.997 *****
## NEGcount.v     0.740 0.740 *****
## NEGfrac.m      0.309 0.309
## literary       0.152 0.152
## obj            0.146 0.146
## VERBcomp      -0.129 0.129
## NOUNcount.m   -0.129 0.129
## passives       0.112 0.112
## predsubjdist.v 0.100 0.100
##
##
## ----- PA4 -----
##              loading abs_l strng
## obj            0.844 0.844 *****
## VERBcomp       0.520 0.520 **
## sentlen.m      0.376 0.376 *
## activity       0.290 0.290
```

```
## passives        -0.262 0.262
## verbdist        -0.224 0.224
## compoundVERBs   -0.221 0.221
## predorder.m      0.207 0.207
## VERBfrac.v      -0.188 0.188
## sentcount       -0.179 0.179
## NEGfrac.m       -0.167 0.167
## predorder.v      0.166 0.166
## NEGcount.m       0.149 0.149
## predsubjdist.v   0.140 0.140
## sentlen.v       -0.139 0.139
## literary         0.138 0.138
##
##
## ----- PA7 -----
##              loading abs_l strng
## maentropy      -0.638 0.638 ****
## entropy        -0.448 0.448 *
## mamr            0.361 0.361 *
## predsubjdist.m  0.298 0.298
## subj            0.297 0.297
## hapaxes        -0.217 0.217
## NEGfrac.m       0.172 0.172
## NOUNcount.v    -0.156 0.156
## predobjdist.m   0.153 0.153
## predorder.m     0.152 0.152
## VERBfrac.v     -0.145 0.145
## verbdist        0.102 0.102
```

hypotheses:

Uniquenesses

```r
fa_broad$uniquenesses %>% round(3)
```

```
##         sentlen.m          sentcount            activity          VERBfrac.m
##             0.080              0.065               0.100               0.096
##         wordcount            entropy            sentlen.v        predsubjdist.m
##             0.112              0.135               0.535               0.647
##      compoundVERBs           passives         predobjdist.m            literary
##             0.296              0.441               0.613               0.758
##          verbdist          maentropy          predorder.m             hapaxes
##             0.197              0.499               0.373               0.318
##          VERBcomp         NOUNcount.v               subj          NOUNcount.m
##             0.403              0.654               0.436               0.193
##     predobjdist.v          NEGcount.m compoundVERBsdist.m          VERBfrac.v
##             0.604              0.054               0.578               0.651
##        NEGcount.v compoundVERBsdist.v      predsubjdist.v                mamr
##             0.412              0.670               0.536               0.255
##               obj         predorder.v           NEGfrac.m
##             0.312              0.470               0.592
```

Distributions over factors

```
broad_data <- data_factor_bind(data_clean, fa_broad)
broad_data$data %>% write_csv("data_w_factors.csv")

broad_data$long %>%
  group_by(factor) %>%
  summarize(shapiro = shapiro.test(factor_score)$p.value)
```

```
## # A tibble: 7 x 2
##   factor  shapiro
##   <fct>     <dbl>
## 1 PA1    2.98e-13
## 2 PA2    2.39e-14
## 3 PA3    7.87e-33
## 4 PA5    1.32e- 3
## 5 PA6    6.04e-12
## 6 PA4    1.43e-14
## 7 PA7    1.69e-11
```

```
broad_data$long %>%
  ggplot(aes(x = factor_score, y = class)) +
  facet_grid(factor ~ .) +
  theme(legend.position = "bottom") +
  geom_jitter(width = 0, height = 0.1, alpha = 0.2)
```
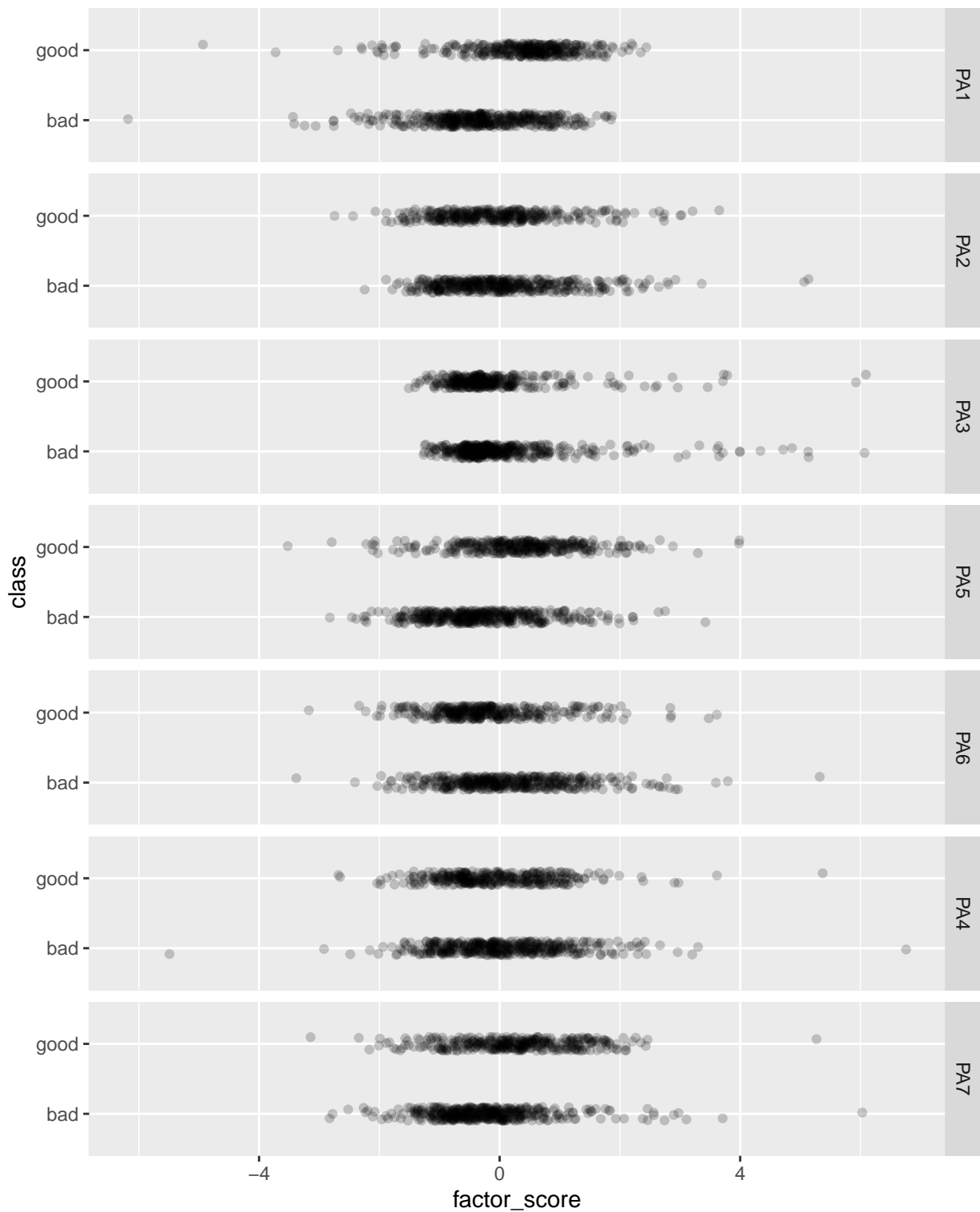
class

```
analyze_distributions(broad_data$long, "class")
```

##

```
## bad good
## 414 339

## Saving 6.5 x 4.5 in image
```
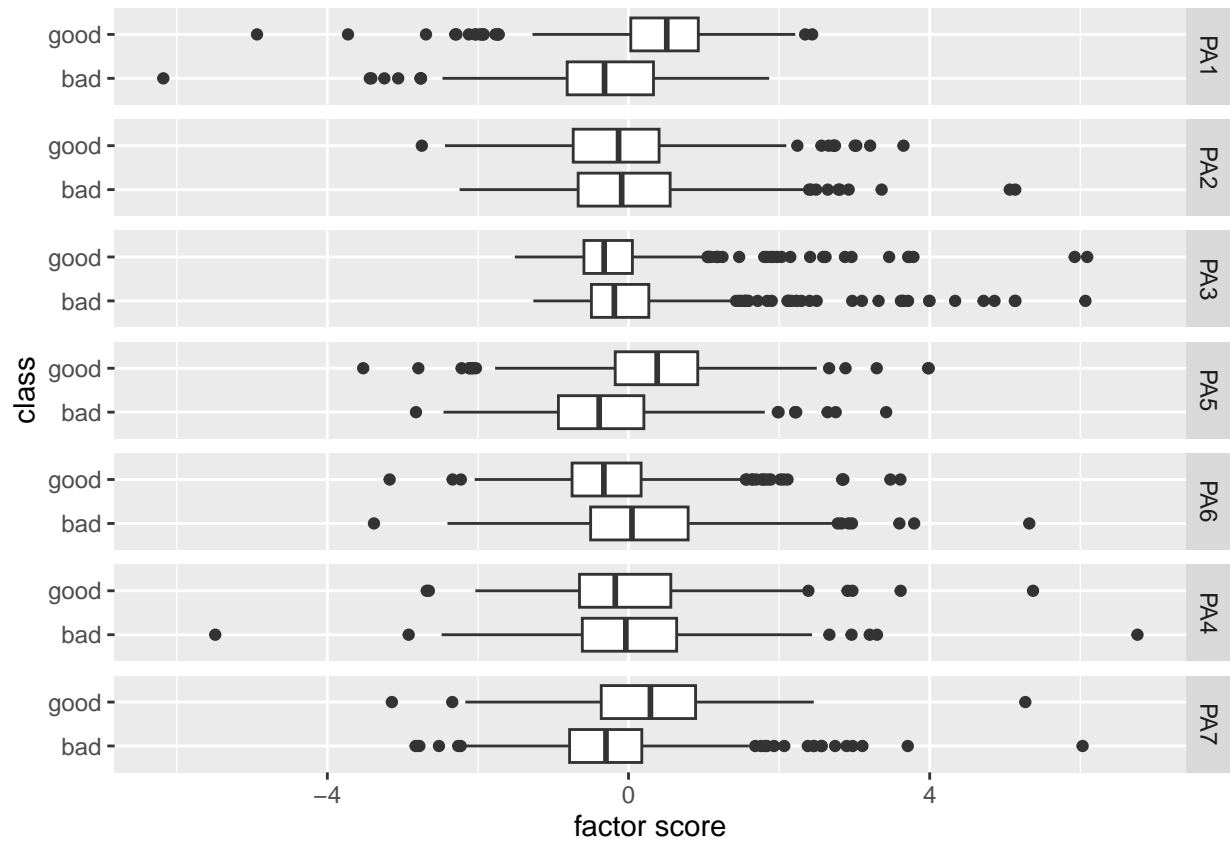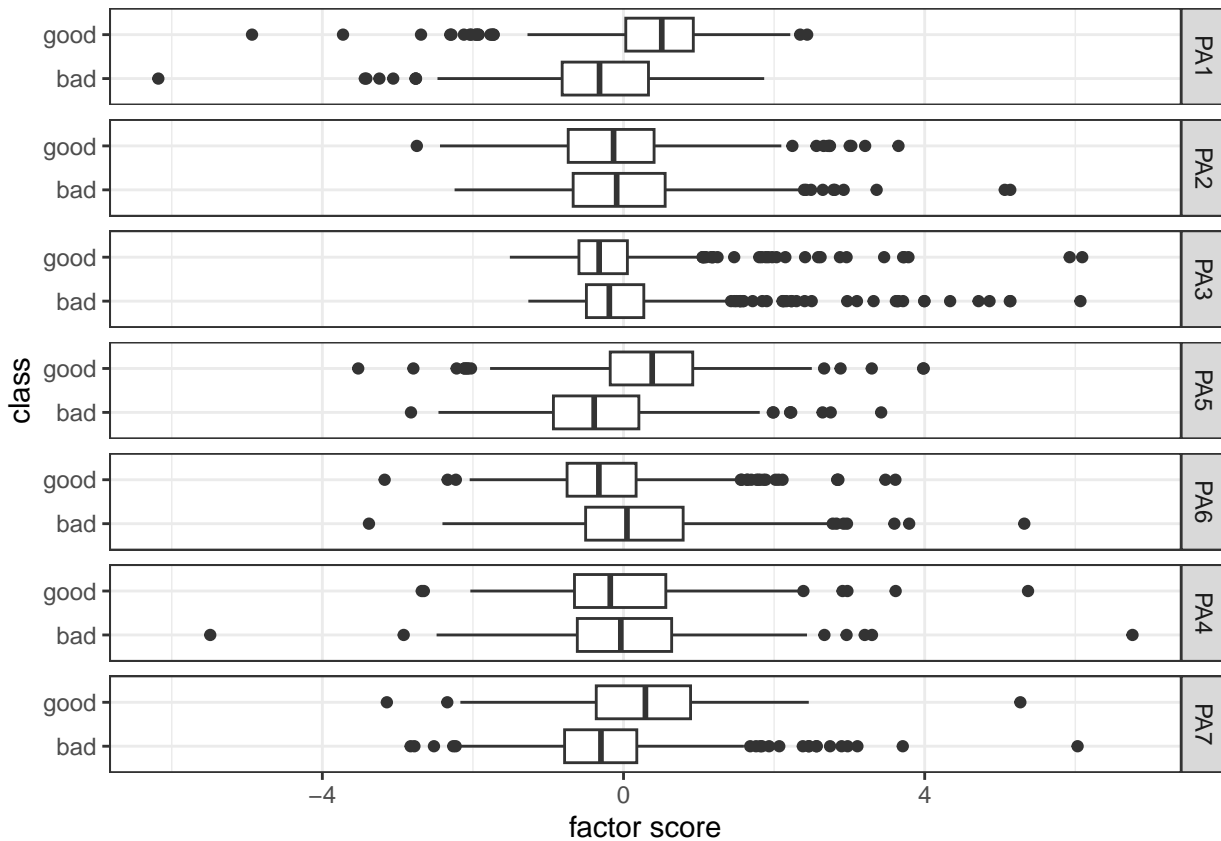
```
##
## Test for the significance of differences in class over PA1 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 123.4655, df = 1, p-value = 0
##
##
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |        bad
## ---------+-----------
##     good |  -11.11150
##          |    0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.164 (95% CI: 0.115 - 0.218 )
##
## Test for the significance of differences in class over PA2 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 0.866, df = 1, p-value = 0.35
```

```
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |        bad
## ---------+-----------
##     good |   0.930602
##          |     0.3521
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.00115 (95% CI: 3.55e-06 - 0.011 )
##
## Test for the significance of differences in class over PA3 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 12.2358, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |        bad
## ---------+-----------
##     good |   3.497969
##          |     0.0005*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0163 (95% CI: 0.00315 - 0.0391 )
##
## Test for the significance of differences in class over PA5 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 97.8011, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |        bad
## ---------+-----------
##     good |  -9.889444
##          |     0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.13 (95% CI: 0.0861 - 0.179 )
##
```

```
## Test for the significance of differences in class over PA6 :
##
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 32.3171, df = 1, p-value = 0
##
##
##                               Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |        bad
## ---------+-----------
##     good |   5.684810
##          |     0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   0.043 (95% CI: 0.0189 - 0.0776 )
##
## Test for the significance of differences in class over PA4 :
##
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 2.5333, df = 1, p-value = 0.11
##
##
##                               Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |        bad
## ---------+-----------
##     good |   1.591639
##          |      0.1115
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   0.00337 (95% CI: 3.02e-05 - 0.0165 )
##
## Test for the significance of differences in class over PA7 :
##
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 64.2257, df = 1, p-value = 0
##
##
##                               Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |        bad
## ---------+-----------
##     good |  -8.014095
```

```
##              |   0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0854 (95% CI: 0.0507 - 0.126 )
##
##   factor   chi2 kruskal_p epsilon2_lci epsilon2 epsilon2_uci
## 1    PA1 123.47  < 0.0001        0.115    0.164        0.218
## 2    PA2   0.87      0.35        0.000    0.001        0.011
## 3    PA3  12.24   < 0.001        0.003    0.016        0.039
## 4    PA5  97.80  < 0.0001        0.086    0.130        0.179
## 5    PA6  32.32  < 0.0001        0.019    0.043        0.078
## 6    PA4   2.53      0.11        0.000    0.003        0.016
## 7    PA7  64.23  < 0.0001        0.051    0.085        0.126
##
## p < 5e-2 found in: PA1 PA3 PA5 PA6 PA7
## p < 1e-2 found in: PA1 PA3 PA5 PA6 PA7
## p < 1e-3 found in: PA1 PA3 PA5 PA6 PA7
## p < 1e-4 found in: PA1 PA5 PA6 PA7
```

subcorpus

```r
analyze_distributions(broad_data$long, "subcorpus")
```

```
##
##      CzCDC       FrBo       KUKY     LiFRLaw OmbuFlyers
##        211        307        194           3         38

## Saving 6.5 x 4.5 in image
```
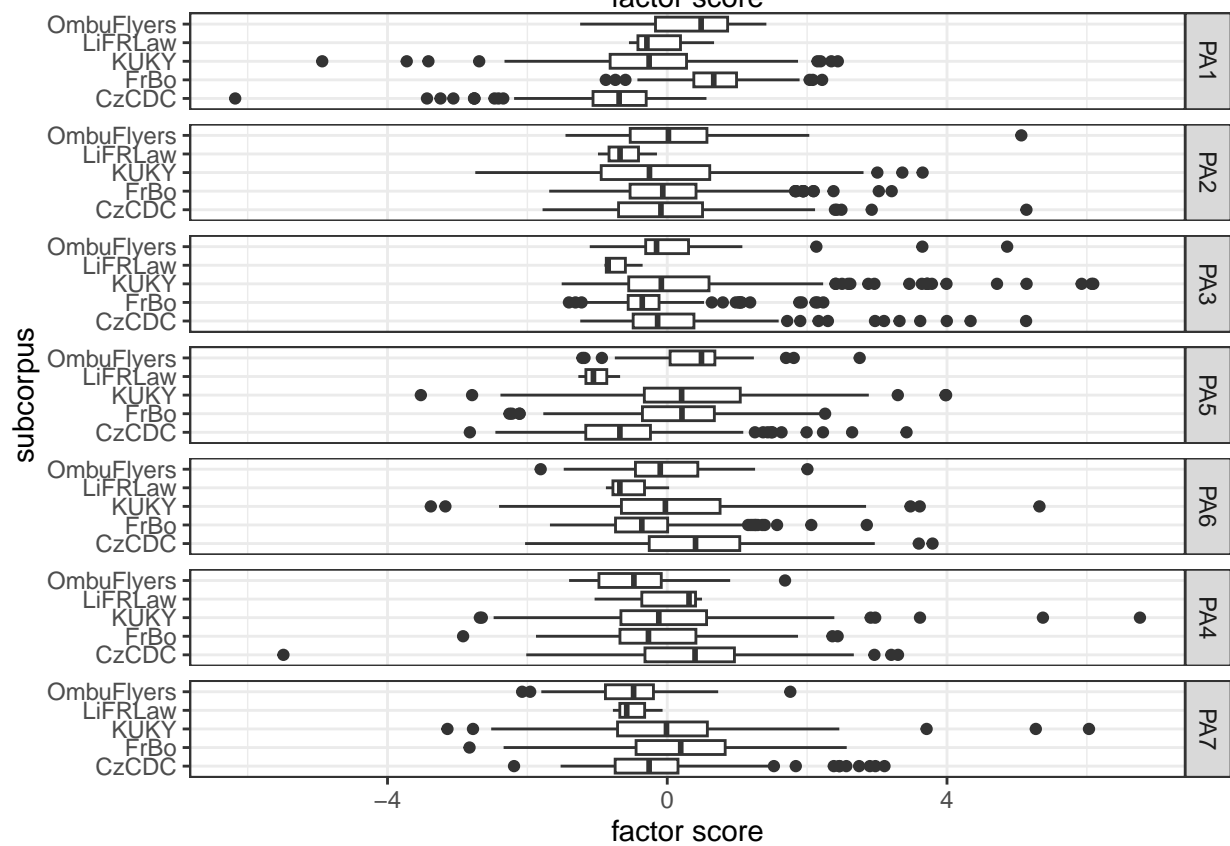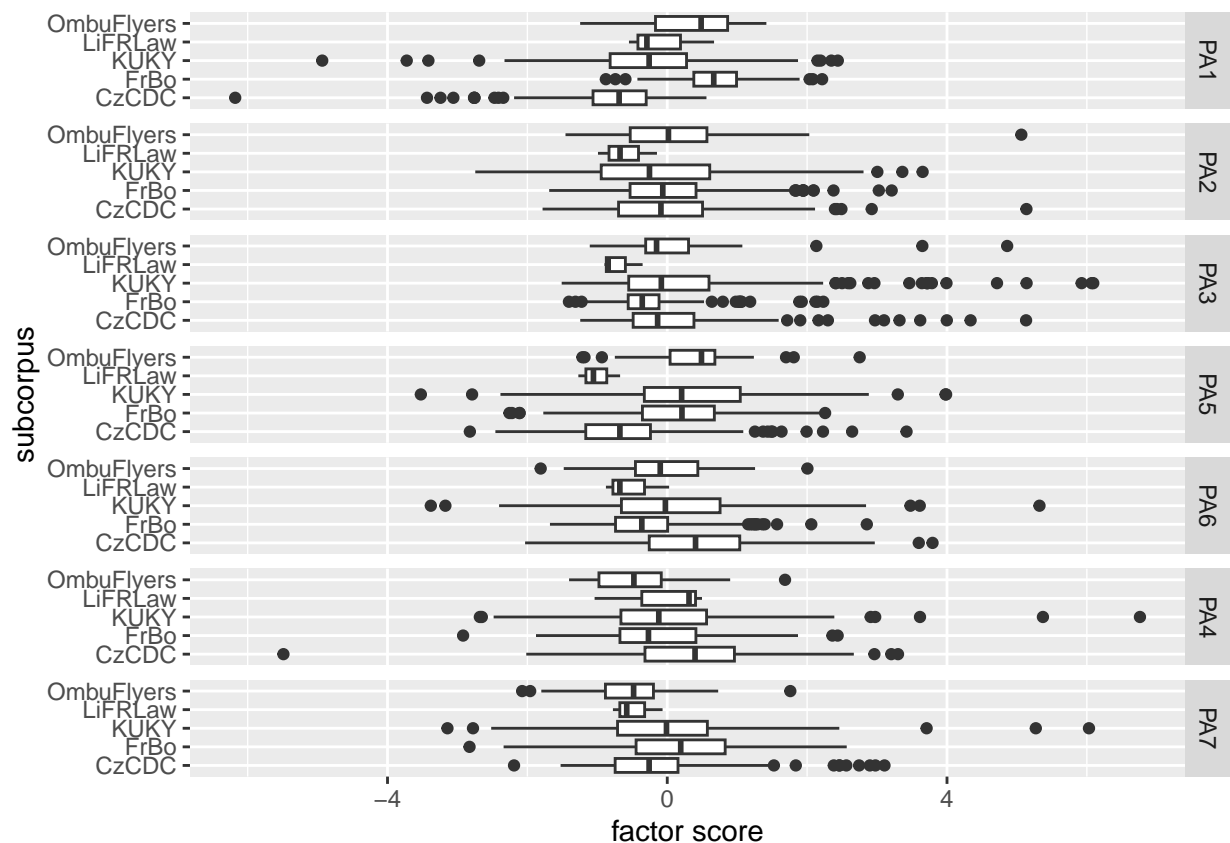
##

```
## Test for the significance of differences in subcorpus over PA1 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 377.3425, df = 4, p-value = 0
##
##
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |     CzCDC       FrBo       KUKY    LiFRLaw
## ---------+--------------------------------------------
##     FrBo |  -18.60055
##          |    0.0000*
##          |
##     KUKY |  -5.567728   12.09728
##          |    0.0000*    0.0000*
##          |
##  LiFRLaw |  -1.250078    1.614183  -0.297410
##          |    1.0000     1.0000     1.0000
##          |
## OmbuFlye |  -7.027404    2.471010  -3.859030  -0.853008
##          |    0.0000*    0.1347     0.0011*    1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.502 (95% CI: 0.462 - 0.551 )
##
## Test for the significance of differences in subcorpus over PA2 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 6.7889, df = 4, p-value = 0.15
##
##
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |     CzCDC       FrBo       KUKY    LiFRLaw
## ---------+--------------------------------------------
##     FrBo |  -0.547372
##          |    1.0000
##          |
##     KUKY |   1.439325    2.094663
##          |    1.0000     0.3620
##          |
##  LiFRLaw |   1.234977    1.322062   0.988141
##          |    1.0000     1.0000     1.0000
##          |
## OmbuFlye |  -0.747903  -0.481779  -1.549989  -1.417131
##          |    1.0000     1.0000     1.0000     1.0000
##
```

```
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.00903 (95% CI: 0.00286 - 0.0315 )
##
## Test for the significance of differences in subcorpus over PA3 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 45.1488, df = 4, p-value = 0
##
##
##                             Comparison of x by group
##                                   (Bonferroni)
## Col Mean-|
## Row Mean |    CzCDC       FrBo       KUKY     LiFRLaw
## ---------+-------------------------------------------
##     FrBo |   4.943535
##          |     0.0000*
##          |
##     KUKY |  -0.564893  -5.432595
##          |     1.0000     0.0000*
##          |
##  LiFRLaw |   2.166509   1.409297   2.261758
##          |     0.3027     1.0000     0.2371
##          |
## OmbuFlye |  -0.590289  -3.175552  -0.269642  -2.273973
##          |     1.0000     0.0150*     1.0000     0.2297
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.06 (95% CI: 0.0359 - 0.0987 )
##
## Test for the significance of differences in subcorpus over PA5 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 143.9294, df = 4, p-value = 0
##
##
##                             Comparison of x by group
##                                   (Bonferroni)
## Col Mean-|
## Row Mean |    CzCDC       FrBo       KUKY     LiFRLaw
## ---------+-------------------------------------------
##     FrBo | -10.26999
##          |     0.0000*
##          |
##     KUKY |  -9.824054  -0.641089
##          |     0.0000*     1.0000
##          |
##  LiFRLaw |   0.858009   2.442870   2.537078
##          |     1.0000     0.1457     0.1118
```

```
##           |
## OmbuFlye |  -6.398949   -1.216896   -0.848195   -2.712212
##          |     0.0000*      1.0000      1.0000      0.0668
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.191 (95% CI: 0.147 - 0.246 )
##
## Test for the significance of differences in subcorpus over PA6 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 82.3697, df = 4, p-value = 0
##
##
##                          Comparison of x by group
##                                 (Bonferroni)
## Col Mean-|
## Row Mean |     CzCDC        FrBo       KUKY     LiFRLaw
## ---------+-------------------------------------------
##     FrBo |   8.966160
##          |     0.0000*
##          |
##     KUKY |   3.984597   -4.420660
##          |     0.0007*     0.0001*
##          |
##  LiFRLaw |   1.823567    0.445571    1.141211
##          |     0.6822      1.0000      1.0000
##          |
## OmbuFlye |   2.617962   -1.979672    0.366449   -0.998729
##          |     0.0885      0.4774      1.0000      1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.11 (95% CI: 0.0764 - 0.16 )
##
## Test for the significance of differences in subcorpus over PA4 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 51.7167, df = 4, p-value = 0
##
##
##                          Comparison of x by group
##                                 (Bonferroni)
## Col Mean-|
## Row Mean |     CzCDC        FrBo       KUKY     LiFRLaw
## ---------+-------------------------------------------
##     FrBo |   6.203533
##          |     0.0000*
##          |
##     KUKY |   4.012379   -1.696963
```

```
##          |    0.0006*        0.8970
##          |
##  LiFRLaw |   0.700048    -0.254598      0.013631
##          |      1.0000        1.0000        1.0000
##          |
## OmbuFlye |   5.319030     2.224822      3.034058      0.884278
##          |    0.0000*        0.2609       0.0241*        1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0688 (95% CI: 0.0418 - 0.113 )
##
## Test for the significance of differences in subcorpus over PA7 :
##
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 43.9067, df = 4, p-value = 0
##
##
##                               Comparison of x by group
##                                    (Bonferroni)
## Col Mean-|
## Row Mean |     CzCDC         FrBo         KUKY      LiFRLaw
## ---------+-------------------------------------------------
##     FrBo |  -5.358556
##          |     0.0000*
##          |
##     KUKY |  -2.336286     2.690861
##          |     0.1948        0.0713
##          |
##  LiFRLaw |   0.629436     1.456765      1.028479
##          |     1.0000        1.0000        1.0000
##          |
## OmbuFlye |   1.968306     4.803489      3.265242     -0.031874
##          |     0.4903        0.0000*       0.0109*        1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0584 (95% CI: 0.0352 - 0.0991 )
##
##    factor   chi2 kruskal_p epsilon2_lci epsilon2 epsilon2_uci
## 1     PA1 377.34  < 0.0001        0.462    0.502        0.551
## 2     PA2   6.79      0.15        0.003    0.009        0.032
## 3     PA3  45.15  < 0.0001        0.036    0.060        0.099
## 4     PA5 143.93  < 0.0001        0.147    0.191        0.246
## 5     PA6  82.37  < 0.0001        0.076    0.110        0.160
## 6     PA4  51.72  < 0.0001        0.042    0.069        0.113
## 7     PA7  43.91  < 0.0001        0.035    0.058        0.099
##
## p < 5e-2 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-2 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-3 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-4 found in: PA1 PA3 PA5 PA6 PA4 PA7
```
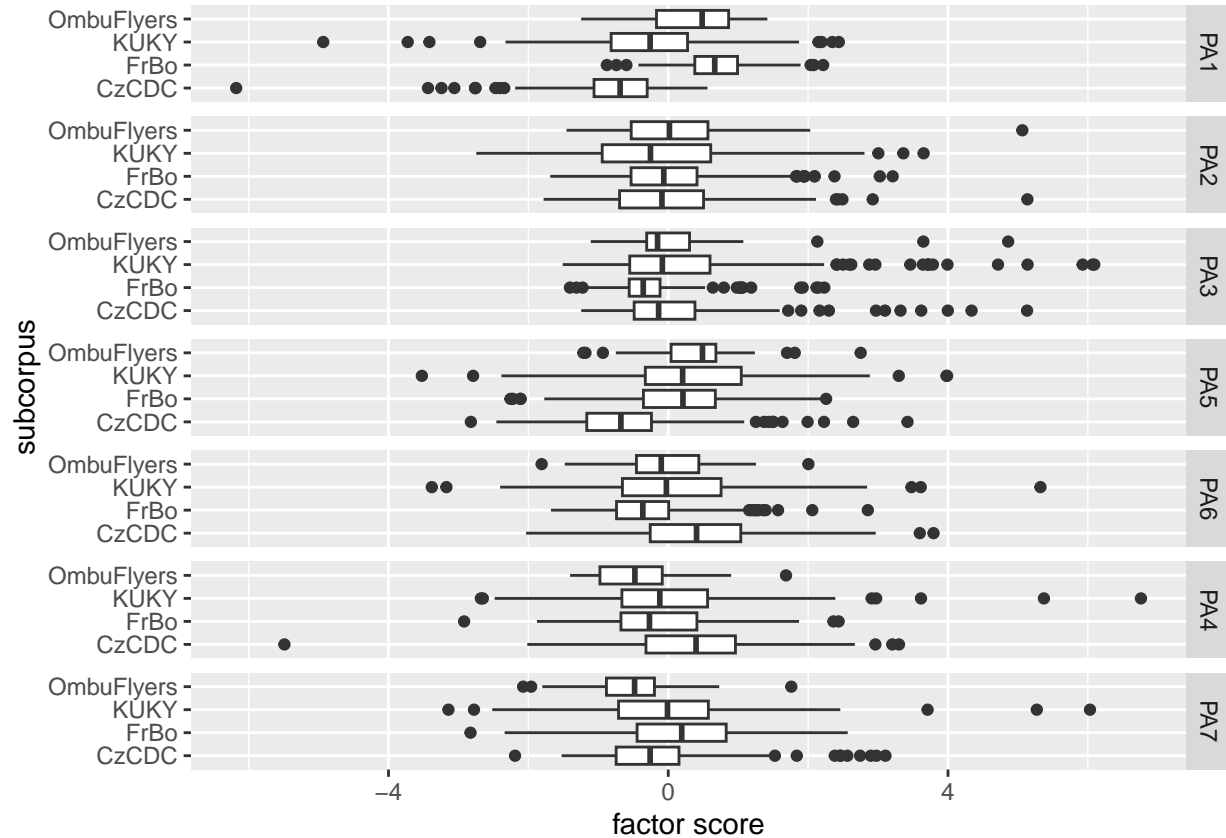
subcorpus wo/ LiFRLaw
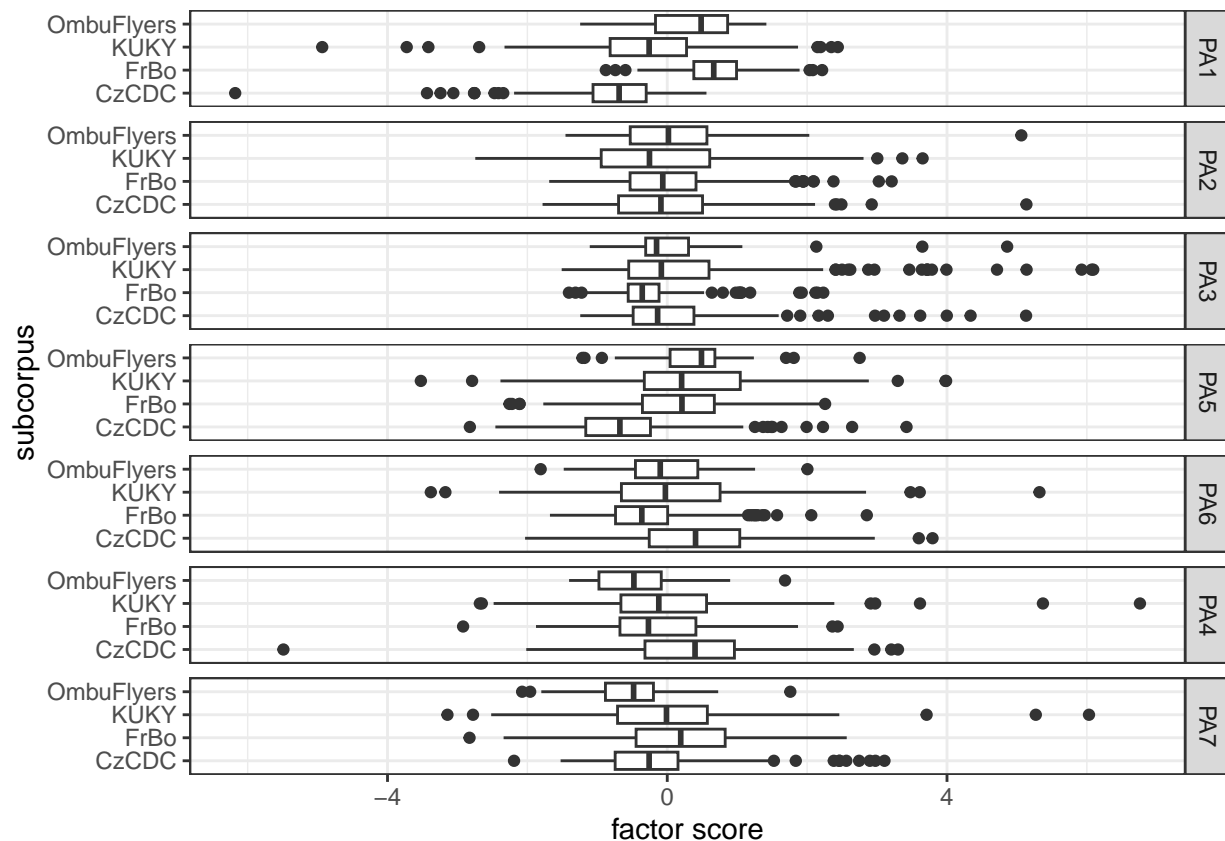
```
analyze_distributions(
  broad_data$long %>% filter(subcorpus != "LiFRLaw"), "subcorpus"
)
```

```
##
##      CzCDC       FrBo       KUKY    LiFRLaw OmbuFlyers
##        211        307        194          0         38
```

```
## Saving 6.5 x 4.5 in image
```

```
##
## Test for the significance of differences in subcorpus over PA1 :
##
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 376.5495, df = 3, p-value = 0
##
##
##                         Comparison of x by group
##                               (Bonferroni)
## Col Mean-|
## Row Mean |     CzCDC        FrBo        KUKY
## ---------+------------------------------------
##     FrBo | -18.58246
##          |    0.0000*
##          |
##     KUKY | -5.560238    12.08776
##          |    0.0000*     0.0000*
##          |
## OmbuFlye | -7.018545    2.470679   -3.854430
##          |    0.0000*     0.0809     0.0007*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.503 (95% CI: 0.461 - 0.549 )
##
```

```
## Test for the significance of differences in subcorpus over PA2 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 5.238, df = 3, p-value = 0.16
##
##
##                              Comparison of x by group
##                                    (Bonferroni)
## Col Mean-|
## Row Mean |      CzCDC        FrBo        KUKY
## ---------+-------------------------------------
##     FrBo |   -0.543387
##          |       1.0000
##          |
##     KUKY |    1.431965    2.082795
##          |       0.9129      0.2236
##          |
## OmbuFlye |   -0.745973   -0.481873   -1.543944
##          |       1.0000      1.0000      0.7356
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.00699 (95% CI: 0.00108 - 0.0268 )
##
## Test for the significance of differences in subcorpus over PA3 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 41.7665, df = 3, p-value = 0
##
##
##                              Comparison of x by group
##                                    (Bonferroni)
## Col Mean-|
## Row Mean |      CzCDC        FrBo        KUKY
## ---------+-------------------------------------
##     FrBo |    4.955310
##          |       0.0000*
##          |
##     KUKY |   -0.570035   -5.449652
##          |       1.0000      0.0000*
##          |
## OmbuFlye |   -0.589885   -3.181261   -0.266358
##          |       1.0000      0.0088*      1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0558 (95% CI: 0.0306 - 0.0912 )
##
## Test for the significance of differences in subcorpus over PA5 :
##
```

```
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 139.9113, df = 3, p-value = 0
##
##
##                           Comparison of x by group
##                                   (Bonferroni)
## Col Mean-|
## Row Mean |      CzCDC         FrBo        KUKY
## ---------+-------------------------------
##     FrBo |  -10.27201
##          |     0.0000*
##          |
##     KUKY |  -9.835261   -0.651277
##          |     0.0000*      1.0000
##          |
## OmbuFlye |  -6.402968   -1.219965   -0.845903
##          |     0.0000*      1.0000      1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.187 (95% CI: 0.136 - 0.244 )
##
## Test for the significance of differences in subcorpus over PA6 :
##
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 81.3197, df = 3, p-value = 0
##
##
##                           Comparison of x by group
##                                   (Bonferroni)
## Col Mean-|
## Row Mean |      CzCDC         FrBo        KUKY
## ---------+-------------------------------
##     FrBo |   8.968444
##          |     0.0000*
##          |
##     KUKY |   3.982588   -4.425067
##          |     0.0004*      0.0001*
##          |
## OmbuFlye |   2.617627   -1.981203    0.367243
##          |     0.0531       0.2854      1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.109 (95% CI: 0.0737 - 0.16 )
##
## Test for the significance of differences in subcorpus over PA4 :
##
##    Kruskal-Wallis rank sum test
##
```

```
## data: x and group
## Kruskal-Wallis chi-squared = 51.6735, df = 3, p-value = 0
##
##
##                                Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |     CzCDC        FrBo       KUKY
## ---------+--------------------------------
##     FrBo |   6.201584
##          |     0.0000*
##          |
##     KUKY |   4.010016  -1.697626
##          |     0.0004*     0.5375
##          |
## OmbuFlye |   5.316390   2.223129   3.032760
##          |     0.0000*     0.1572     0.0145*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.069 (95% CI: 0.0404 - 0.113 )
##
## Test for the significance of differences in subcorpus over PA7 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 42.7952, df = 3, p-value = 0
##
##
##                                Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |     CzCDC        FrBo       KUKY
## ---------+--------------------------------
##     FrBo |  -5.358239
##          |     0.0000*
##          |
##     KUKY |  -2.336550   2.690265
##          |     0.1168     0.0428*
##          |
## OmbuFlye |   1.966417   4.801387   3.263513
##          |     0.2955     0.0000*     0.0066*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0571 (95% CI: 0.032 - 0.0962 )
##
##   factor   chi2 kruskal_p epsilon2_lci epsilon2 epsilon2_uci
## 1    PA1 376.55  < 0.0001        0.461    0.503        0.549
## 2    PA2   5.24      0.16        0.001    0.007        0.027
## 3    PA3  41.77  < 0.0001        0.031    0.056        0.091
## 4    PA5 139.91  < 0.0001        0.136    0.187        0.244
## 5    PA6  81.32  < 0.0001        0.074    0.109        0.160
```
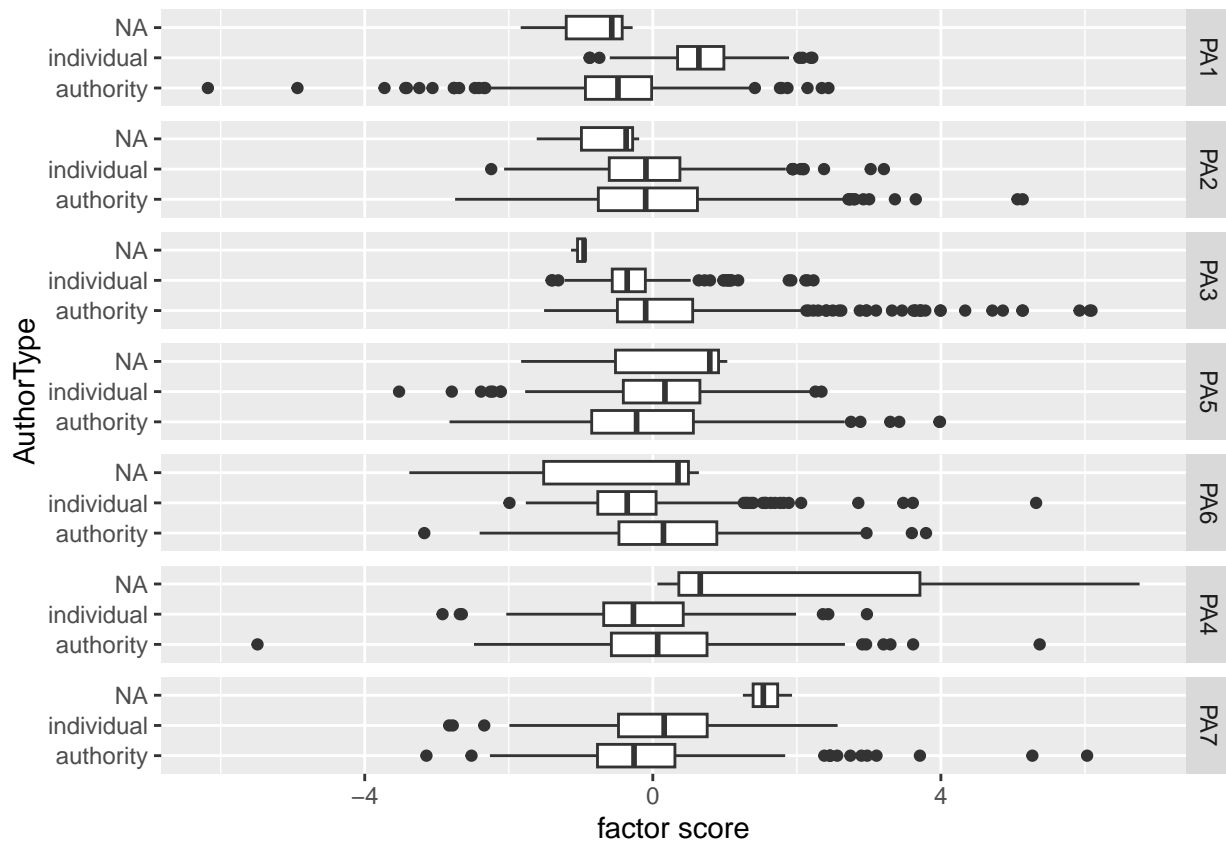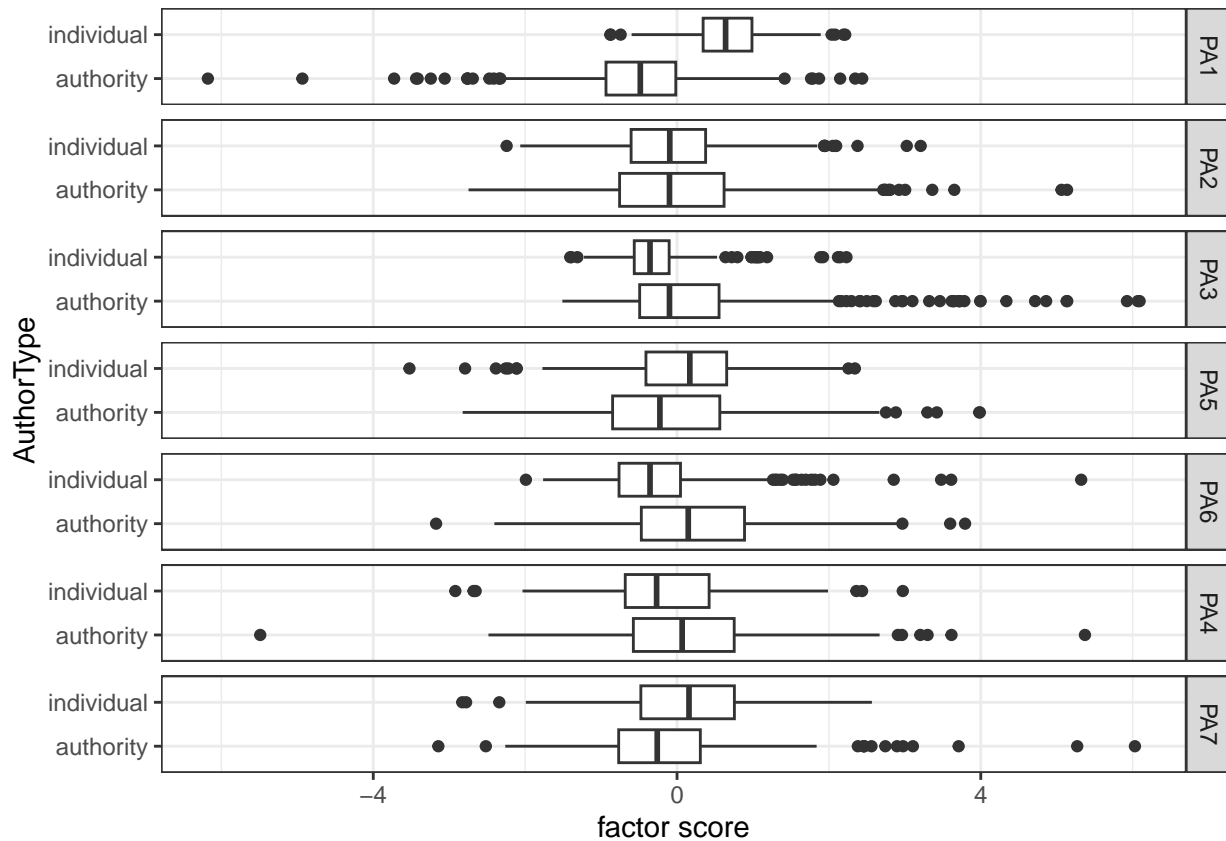
```
## 6    PA4  51.67  < 0.0001            0.040    0.069          0.113
## 7    PA7  42.80  < 0.0001            0.032    0.057          0.096
##
## p < 5e-2 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-2 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-3 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-4 found in: PA1 PA3 PA5 PA6 PA4 PA7
```

AuthorType

```
analyze_distributions(broad_data$long, "AuthorType")
```

```
##
##  authority individual       <NA>
##        411        339          3
```

```
## Saving 6.5 x 4.5 in image
```

```
##
## Test for the significance of differences in AuthorType over PA1 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 322.2485, df = 1, p-value = 0
##
##
##                         Comparison of x by group
##                                (Bonferroni)
## Col Mean-|
## Row Mean |   authorit
## ---------+-----------
## individu |  -17.95127
##          |     0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   0.429 (95% CI: 0.374 - 0.488 )
##
## Test for the significance of differences in AuthorType over PA2 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 0.0091, df = 1, p-value = 0.92
```

```
##
##
##                              Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |   authorit
## ---------+-----------
## individu |   0.095329
##          |      0.9241
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  1.21e-05 (95% CI: 1.31e-06 - 0.00731 )
##
## Test for the significance of differences in AuthorType over PA3 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 45.0354, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |   authorit
## ---------+-----------
## individu |   6.710839
##          |      0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0599 (95% CI: 0.0318 - 0.097 )
##
## Test for the significance of differences in AuthorType over PA5 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 17.197, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |   authorit
## ---------+-----------
## individu |  -4.146927
##          |      0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0229 (95% CI: 0.00703 - 0.0486 )
##
```

```
## Test for the significance of differences in AuthorType over PA6 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 54.5596, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |   authorit
## ---------+-----------
## individu |   7.386444
##          |    0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0726 (95% CI: 0.0409 - 0.113 )
##
## Test for the significance of differences in AuthorType over PA4 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 17.8605, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |   authorit
## ---------+-----------
## individu |   4.226171
##          |    0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0238 (95% CI: 0.00677 - 0.0485 )
##
## Test for the significance of differences in AuthorType over PA7 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 31.3267, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |   authorit
## ---------+-----------
## individu |  -5.597022
```

```
##            |    0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0417 (95% CI: 0.0184 - 0.0737 )
##
##    factor    chi2 kruskal_p epsilon2_lci epsilon2 epsilon2_uci
## 1     PA1 322.25  < 0.0001        0.374    0.429        0.488
## 2     PA2   0.01      0.92        0.000    0.000        0.007
## 3     PA3  45.04  < 0.0001        0.032    0.060        0.097
## 4     PA5  17.20  < 0.0001        0.007    0.023        0.049
## 5     PA6  54.56  < 0.0001        0.041    0.073        0.113
## 6     PA4  17.86  < 0.0001        0.007    0.024        0.048
## 7     PA7  31.33  < 0.0001        0.018    0.042        0.074
##
## p < 5e-2 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-2 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-3 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-4 found in: PA1 PA3 PA5 PA6 PA4 PA7
```

RecipientType

```r
analyze_distributions(broad_data$long, "RecipientType")
```

```
##
##       combined    legal person natural person         <NA>
##            304              23            413           13

## Saving 6.5 x 4.5 in image
```

##
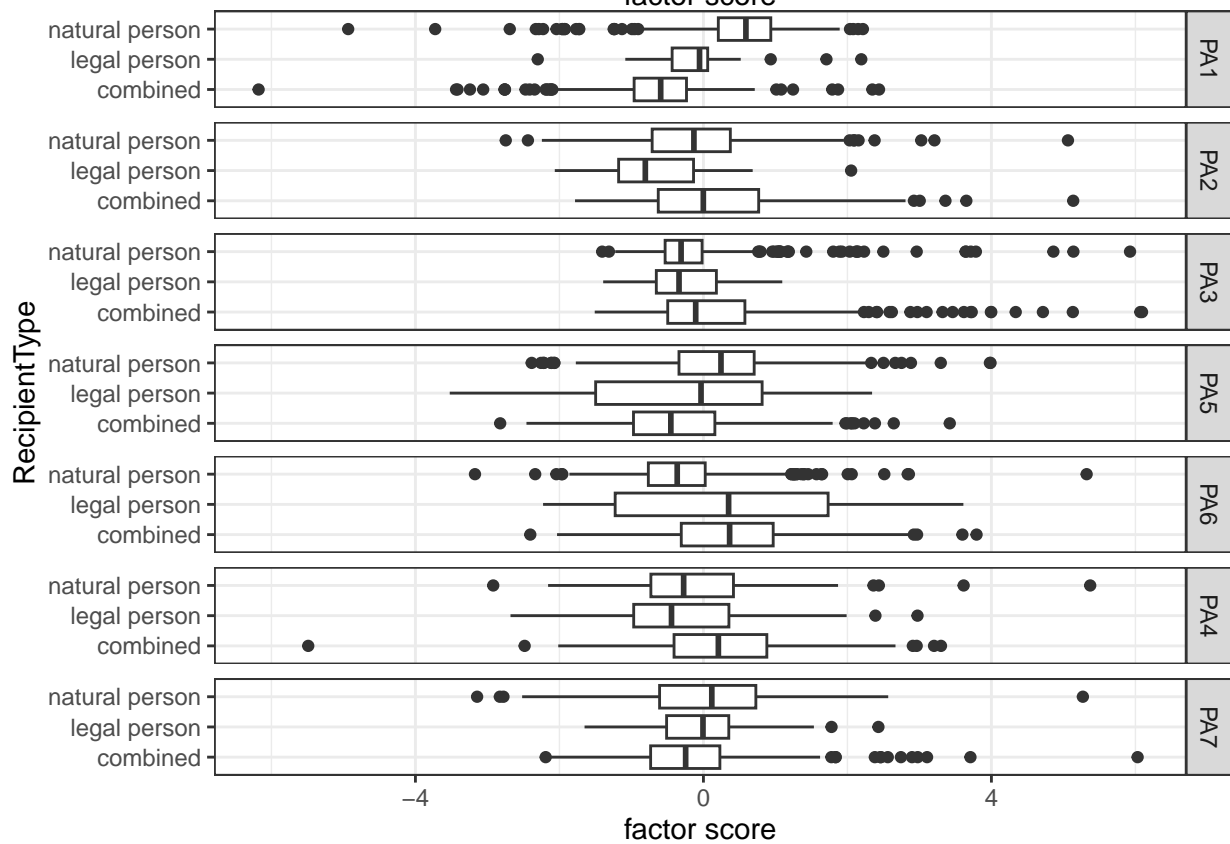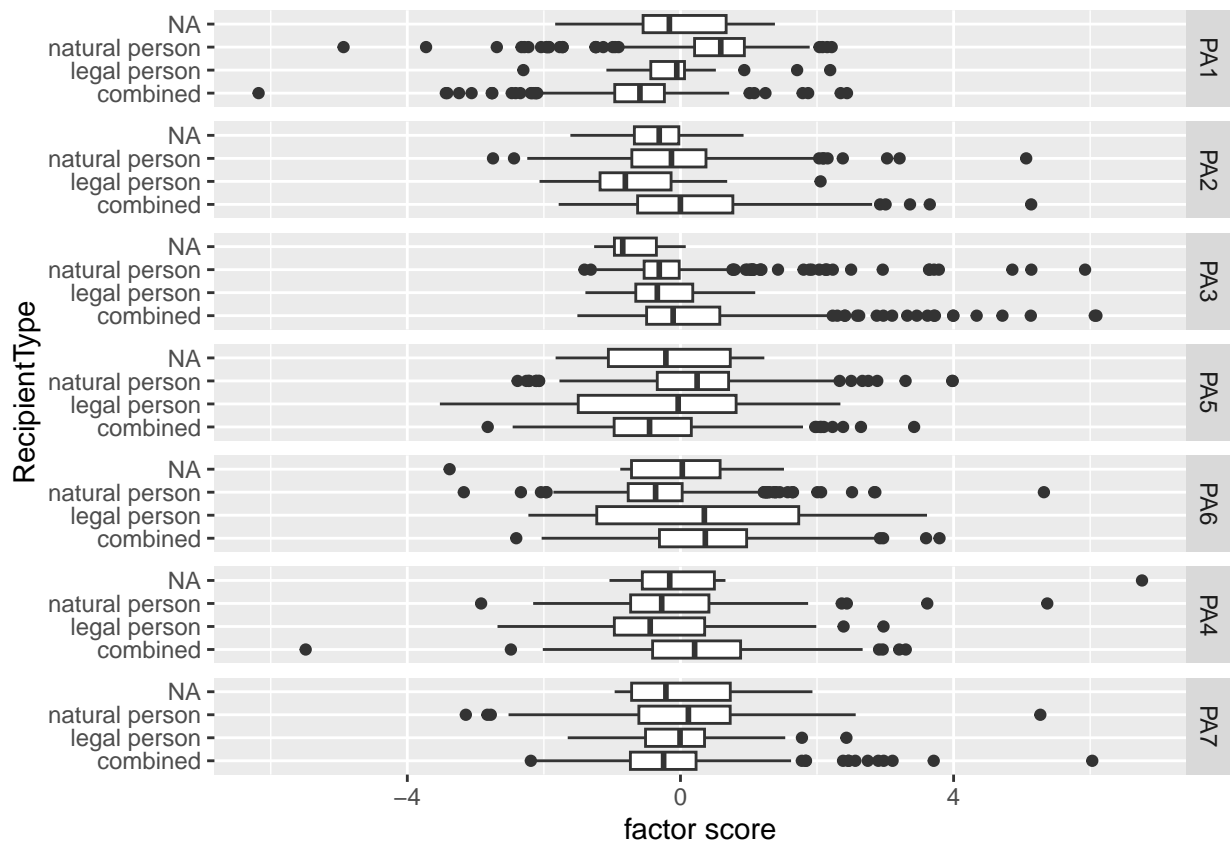
```
## Test for the significance of differences in RecipientType over PA1 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 314.5305, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                    (Bonferroni)
## Col Mean-|
## Row Mean |   combined   legal pe
## ---------+----------------------
## legal pe |  -2.565495
##          |     0.0309*
##          |
## natural  |  -17.70569  -3.655701
##          |     0.0000*    0.0008*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.418 (95% CI: 0.359 - 0.478 )
##
## Test for the significance of differences in RecipientType over PA2 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 16.3093, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                    (Bonferroni)
## Col Mean-|
## Row Mean |   combined   legal pe
## ---------+----------------------
## legal pe |   3.658195
##          |     0.0008*
##          |
## natural  |   2.412131  -2.841796
##          |     0.0476*    0.0135*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0217 (95% CI: 0.00646 - 0.0505 )
##
## Test for the significance of differences in RecipientType over PA3 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 20.0099, df = 2, p-value = 0
##
##
```

```
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |    combined    legal pe
## ---------+----------------------
## legal pe |    1.654730
##          |       0.2939
##          |
## natural  |    4.403938   -0.116900
##          |     0.0000*     1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0266 (95% CI: 0.00874 - 0.0583 )
##
## Test for the significance of differences in RecipientType over PA5 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 74.4874, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |    combined    legal pe
## ---------+----------------------
## legal pe |   -0.546314
##          |       1.0000
##          |
## natural  |   -8.546976   -2.463325
##          |     0.0000*     0.0413*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0991 (95% CI: 0.0622 - 0.147 )
##
## Test for the significance of differences in RecipientType over PA6 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 92.3301, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |    combined    legal pe
## ---------+----------------------
## legal pe |    1.077348
##          |       0.8440
##          |
```

```
## natural  |   9.569695    2.288037
##          |    0.0000*     0.0664
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.123 (95% CI: 0.0825 - 0.174 )
##
## Test for the significance of differences in RecipientType over PA4 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 35.0423, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                    (Bonferroni)
## Col Mean-|
## Row Mean |   combined   legal pe
## ---------+----------------------
## legal pe |   2.301815
##          |      0.0640
##          |
## natural  |   5.805495  -0.275704
##          |    0.0000*    1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0466 (95% CI: 0.0214 - 0.0831 )
##
## Test for the significance of differences in RecipientType over PA7 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 19.2854, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                    (Bonferroni)
## Col Mean-|
## Row Mean |   combined   legal pe
## ---------+----------------------
## legal pe |  -1.109359
##          |      0.8018
##          |
## natural  |  -4.385409  -0.427067
##          |    0.0000*    1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0256 (95% CI: 0.00909 - 0.0532 )
##
##   factor   chi2 kruskal_p epsilon2_lci epsilon2 epsilon2_uci
```

```
## 1    PA1 314.53  < 0.0001         0.359    0.418        0.478
## 2    PA2 16.31   < 0.001          0.006    0.022        0.051
## 3    PA3 20.01   < 0.0001         0.009    0.027        0.058
## 4    PA5 74.49   < 0.0001         0.062    0.099        0.147
## 5    PA6 92.33   < 0.0001         0.082    0.123        0.174
## 6    PA4 35.04   < 0.0001         0.021    0.047        0.083
## 7    PA7 19.29   < 0.0001         0.009    0.026        0.053
##
## p < 5e-2 found in: PA1 PA2 PA3 PA5 PA6 PA4 PA7
## p < 1e-2 found in: PA1 PA2 PA3 PA5 PA6 PA4 PA7
## p < 1e-3 found in: PA1 PA2 PA3 PA5 PA6 PA4 PA7
## p < 1e-4 found in: PA1 PA3 PA5 PA6 PA4 PA7
```

court decisions often with `RecipientType = combined`.

RecipientIndividuation

```
analyze_distributions(broad_data$long, "RecipientIndividuation")
```

```
##
##      bulk individual     public      <NA>
##       69        356        319         9
```

```
## Saving 6.5 x 4.5 in image
```
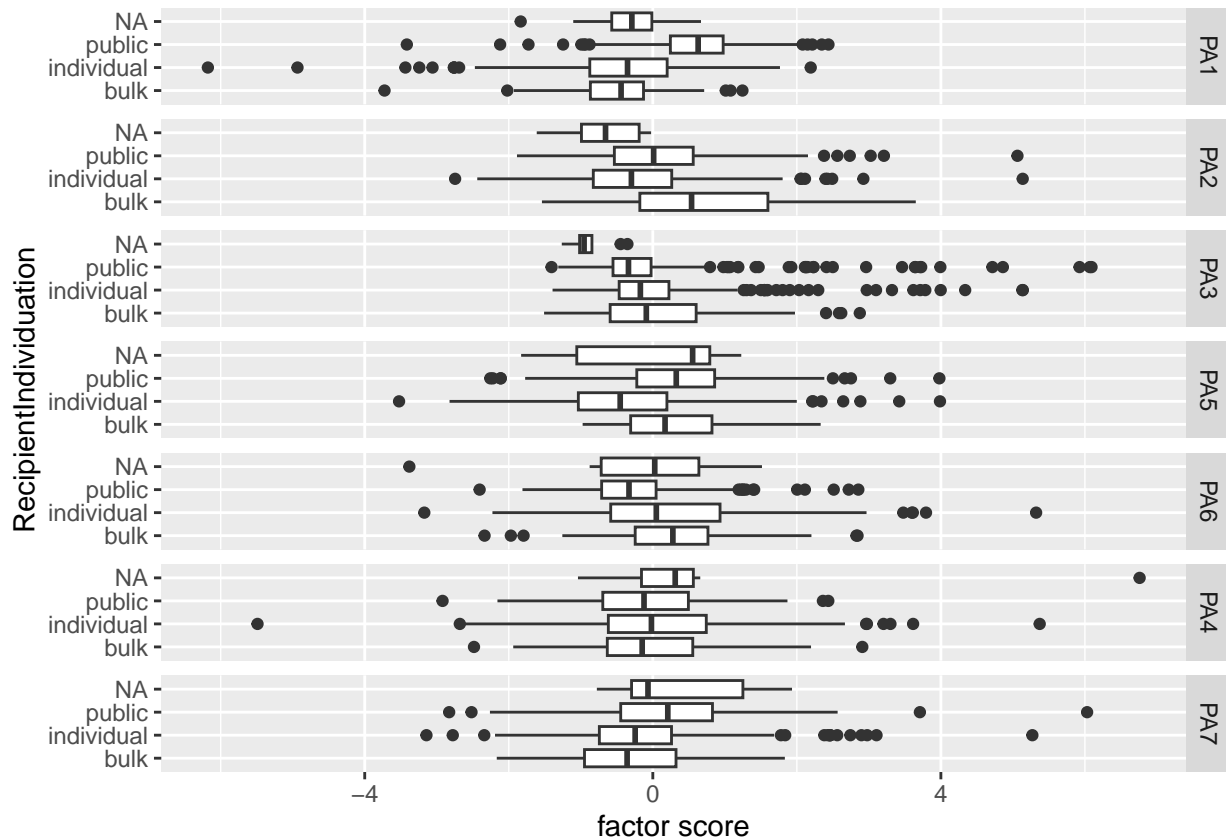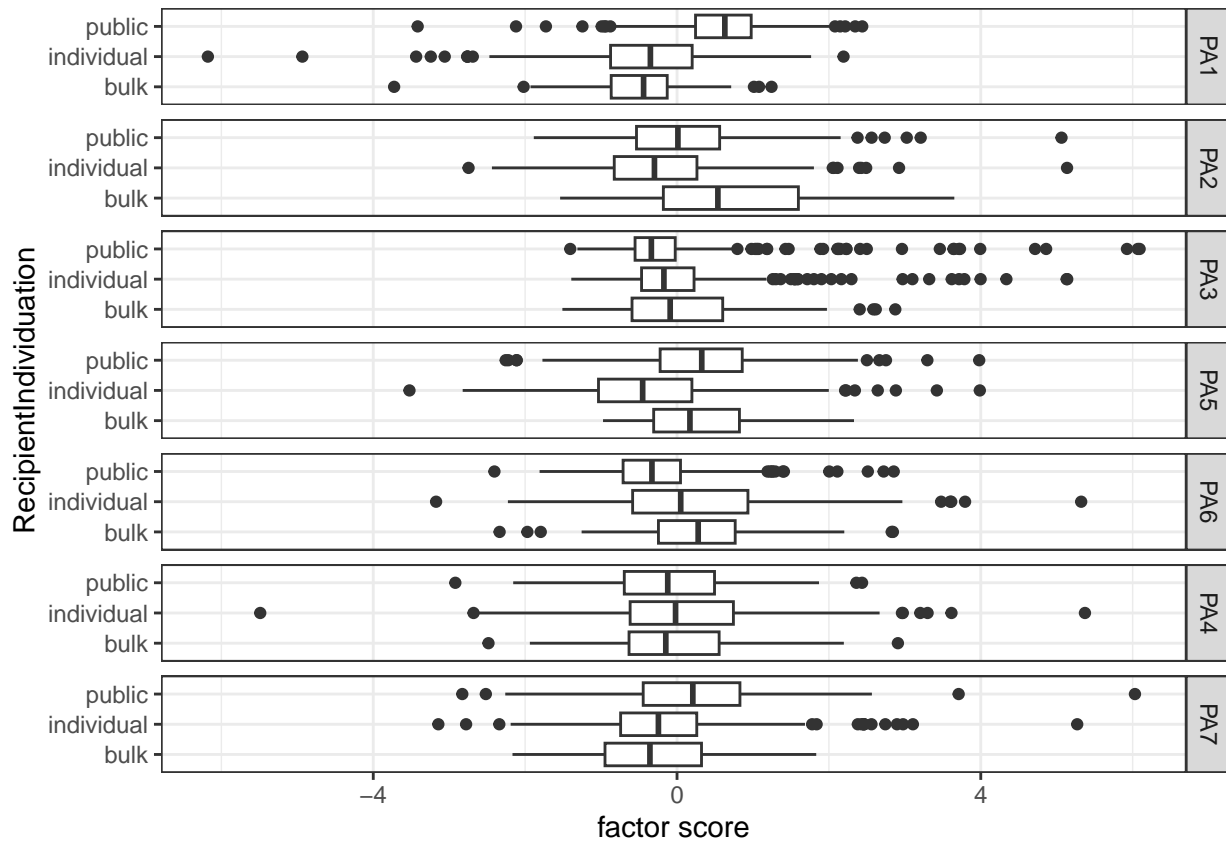
```
## 
## Test for the significance of differences in RecipientIndividuation over PA1 :
## 
##    Kruskal-Wallis rank sum test
## 
## data: x and group
## Kruskal-Wallis chi-squared = 233.132, df = 2, p-value = 0
## 
## 
##                          Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |      bulk    individu
## ---------+----------------------
## individu |  -1.103793
##          |     0.8090
##          |
##    public |  -9.412970  -14.32708
##          |     0.0000*     0.0000*
## 
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.31 (95% CI: 0.257 - 0.363 )
## 
## Test for the significance of differences in RecipientIndividuation over PA2 :
## 
##    Kruskal-Wallis rank sum test
```

```
##
## data: x and group
## Kruskal-Wallis chi-squared = 41.5502, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |       bulk    individu
## ---------+----------------------
## individu |   5.864716
##          |      0.0000*
##          |
##   public |   3.374456   -4.194765
##          |      0.0022*      0.0001*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   0.0553 (95% CI: 0.0276 - 0.092 )
##
## Test for the significance of differences in RecipientIndividuation over PA3 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 13.9732, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |       bulk    individu
## ---------+----------------------
## individu |   0.492146
##          |      1.0000
##          |
##   public |   2.475948   3.424222
##          |      0.0399*      0.0018*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   0.0186 (95% CI: 0.0049 - 0.0471 )
##
## Test for the significance of differences in RecipientIndividuation over PA5 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 108.2741, df = 2, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
```

```
## Row Mean |       bulk    individu
## ---------+----------------------
## individu |    5.699792
##          |      0.0000*
##          |
##   public |  -0.127232  -9.943723
##          |     1.0000     0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.144 (95% CI: 0.0989 - 0.199 )
##
## Test for the significance of differences in RecipientIndividuation over PA6 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 42.0919, df = 2, p-value = 0
##
##
##                                Comparison of x by group
##                                        (Bonferroni)
## Col Mean-|
## Row Mean |       bulk    individu
## ---------+----------------------
## individu |    1.618330
##          |      0.3168
##          |
##   public |    4.848507   5.588641
##          |      0.0000*     0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.056 (95% CI: 0.0323 - 0.0963 )
##
## Test for the significance of differences in RecipientIndividuation over PA4 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 4.0916, df = 2, p-value = 0.13
##
##
##                                Comparison of x by group
##                                        (Bonferroni)
## Col Mean-|
## Row Mean |       bulk    individu
## ---------+----------------------
## individu |   -0.714174
##          |      1.0000
##          |
##   public |    0.463258   2.016266
##          |      1.0000     0.1313
##
```

```
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.00544 (95% CI: 0.000386 - 0.0235 )
##
## Test for the significance of differences in RecipientIndividuation over PA7 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 42.8594, df = 2, p-value = 0
##
##
##                             Comparison of x by group
##                                    (Bonferroni)
## Col Mean-|
## Row Mean |      bulk   individu
## ---------+----------------------
## individu |  -0.544438
##          |      1.0000
##          |
##   public |  -4.091940  -6.117944
##          |     0.0001*    0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.057 (95% CI: 0.0306 - 0.099 )
##
##   factor   chi2 kruskal_p epsilon2_lci epsilon2 epsilon2_uci
## 1    PA1 233.13  < 0.0001        0.257    0.310        0.363
## 2    PA2  41.55  < 0.0001        0.028    0.055        0.092
## 3    PA3  13.97   < 0.001        0.005    0.019        0.047
## 4    PA5 108.27  < 0.0001        0.099    0.144        0.199
## 5    PA6  42.09  < 0.0001        0.032    0.056        0.096
## 6    PA4   4.09      0.13        0.000    0.005        0.024
## 7    PA7  42.86  < 0.0001        0.031    0.057        0.099
##
## p < 5e-2 found in: PA1 PA2 PA3 PA5 PA6 PA7
## p < 1e-2 found in: PA1 PA2 PA3 PA5 PA6 PA7
## p < 1e-3 found in: PA1 PA2 PA5 PA6 PA7
## p < 1e-4 found in: PA1 PA2 PA5 PA6 PA7
```
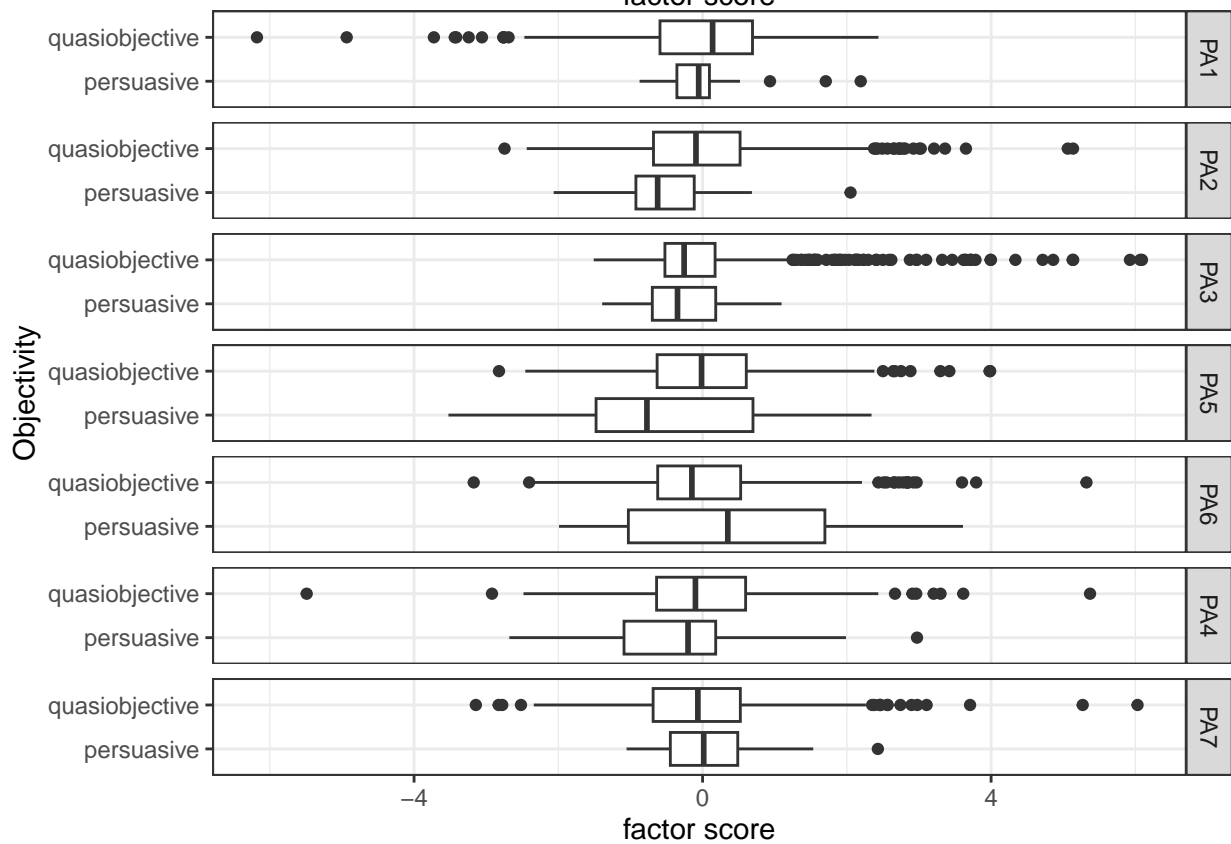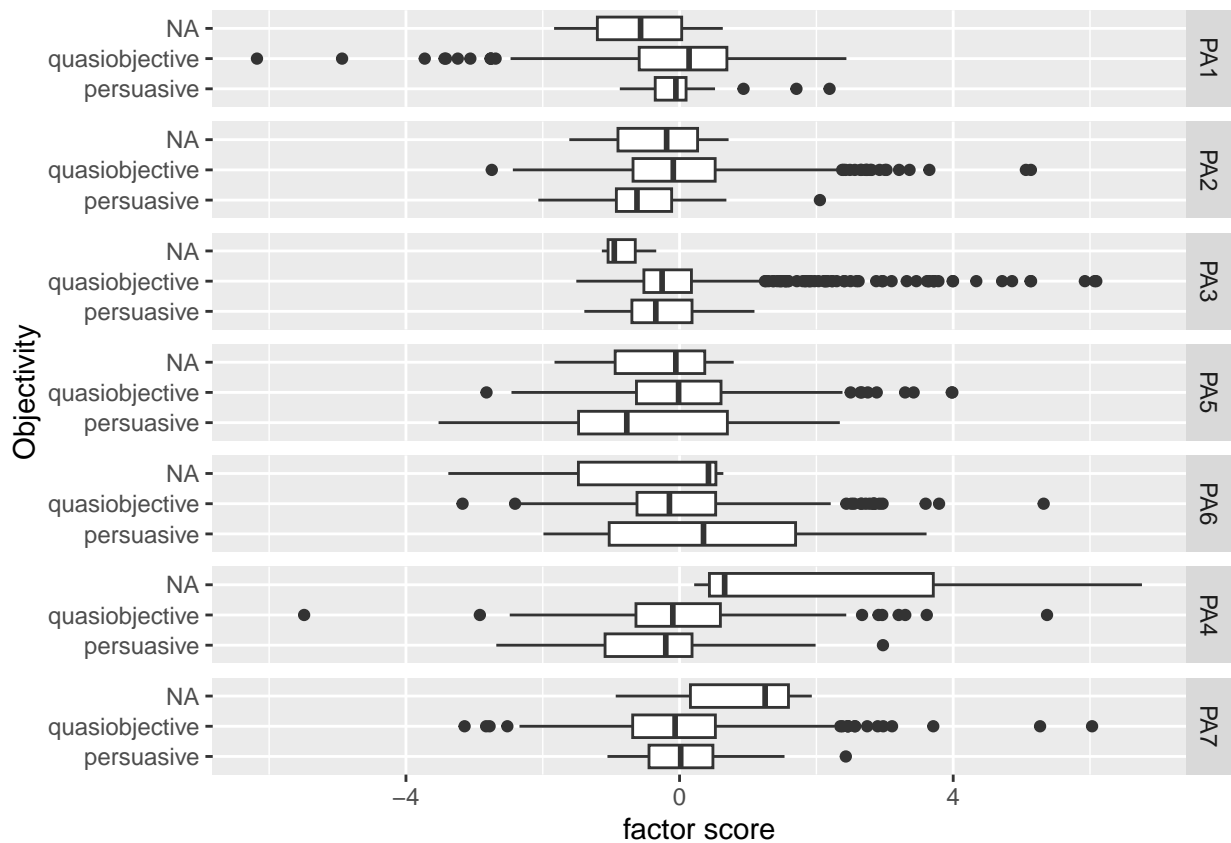
Objectivity

```
analyze_distributions(broad_data$long, "Objectivity")
```

```
##
##     persuasive quasiobjective           <NA>
##             21            729              3

## Saving 6.5 x 4.5 in image
```

##

```
## Test for the significance of differences in Objectivity over PA1 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 0.3232, df = 1, p-value = 0.57
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |   persuasi
## ---------+-----------
## quasiobj |  -0.568541
##          |      0.5697
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   0.00043 (95% CI: 1.1e-06 - 0.00582 )
##
## Test for the significance of differences in Objectivity over PA2 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 5.9196, df = 1, p-value = 0.01
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |   persuasi
## ---------+-----------
## quasiobj |  -2.433032
##          |      0.0150*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   0.00787 (95% CI: 0.000313 - 0.0233 )
##
## Test for the significance of differences in Objectivity over PA3 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 0.9549, df = 1, p-value = 0.33
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |   persuasi
## ---------+-----------
## quasiobj |  -0.977197
```

```
##            |     0.3285
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.00127 (95% CI: 2.82e-06 - 0.0121 )
##
## Test for the significance of differences in Objectivity over PA5 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 2.8261, df = 1, p-value = 0.09
##
##
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |   persuasi
## ---------+-----------
## quasiobj |  -1.681106
##          |      0.0927
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.00376 (95% CI: 1.17e-05 - 0.0251 )
##
## Test for the significance of differences in Objectivity over PA6 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 0.7532, df = 1, p-value = 0.39
##
##
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |   persuasi
## ---------+-----------
## quasiobj |   0.867881
##          |      0.3855
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.001 (95% CI: 6.13e-06 - 0.0181 )
##
## Test for the significance of differences in Objectivity over PA4 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 1.0469, df = 1, p-value = 0.31
##
##
```

```
##                          Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |   persuasi
## ---------+-----------
## quasiobj |  -1.023170
##          |      0.3062
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.00139 (95% CI: 6.34e-06 - 0.0163 )
##
## Test for the significance of differences in Objectivity over PA7 :
##
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 0.6277, df = 1, p-value = 0.43
##
##
##                          Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |   persuasi
## ---------+-----------
## quasiobj |   0.792280
##          |      0.4282
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.000835 (95% CI: 2.17e-06 - 0.00722 )
##
##    factor chi2 kruskal_p epsilon2_lci epsilon2 epsilon2_uci
## 1    PA1 0.32      0.57            0    0.000        0.006
## 2    PA2 5.92    < 0.05            0    0.008        0.023
## 3    PA3 0.95      0.33            0    0.001        0.012
## 4    PA5 2.83      0.09            0    0.004        0.025
## 5    PA6 0.75      0.39            0    0.001        0.018
## 6    PA4 1.05      0.31            0    0.001        0.016
## 7    PA7 0.63      0.43            0    0.001        0.007
##
## p < 5e-2 found in: PA2
## p < 1e-2 found in:
## p < 1e-3 found in:
## p < 1e-4 found in:
```
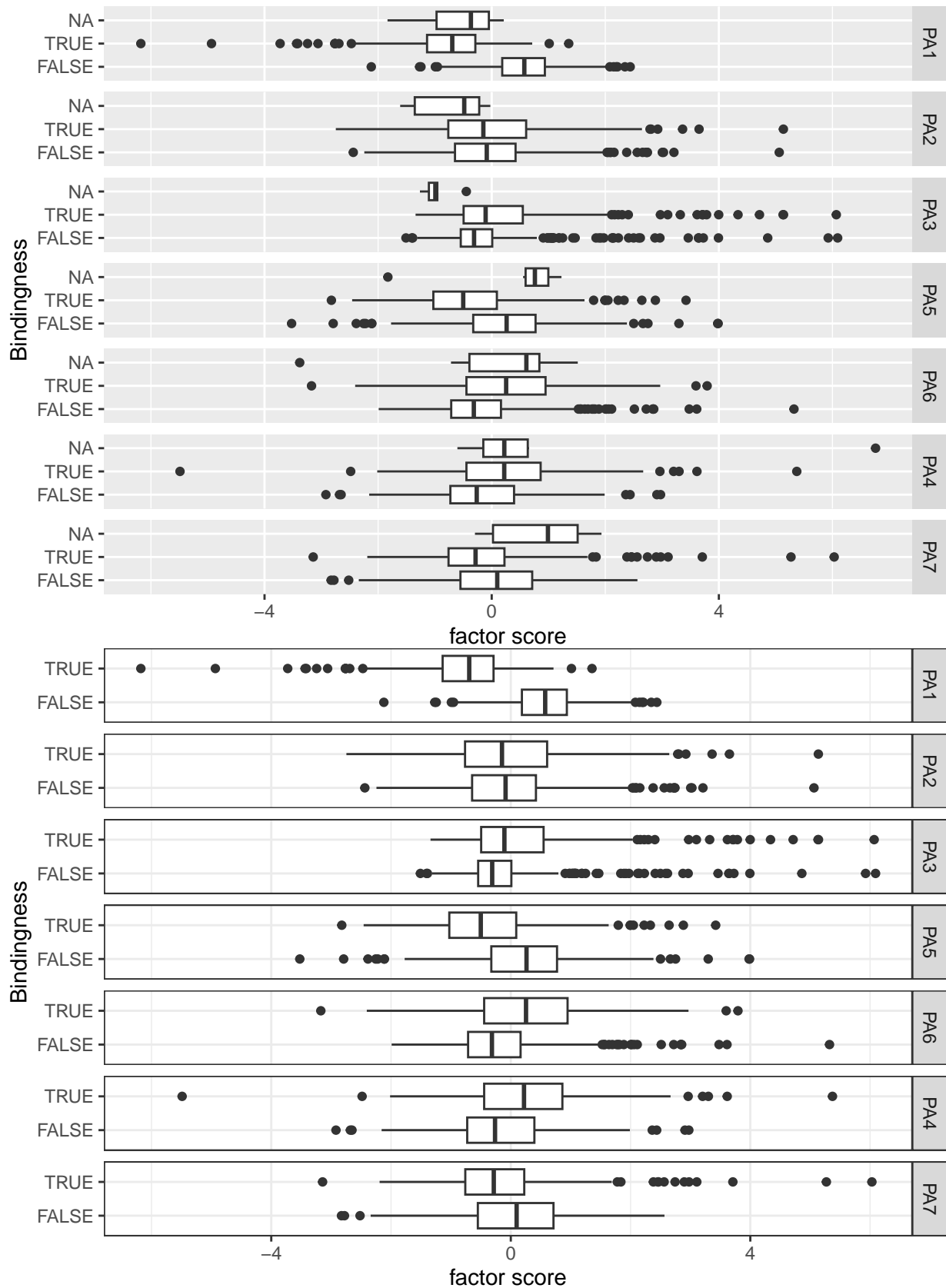
Bindingness

```r
analyze_distributions(broad_data$long, "Bindingness")
```

```
##
## FALSE   TRUE   <NA>
##   444    303      6
```

```
## 
## Test for the significance of differences in Bindingness over PA1 :
## 
##    Kruskal-Wallis rank sum test
## 
## data: x and group
## Kruskal-Wallis chi-squared = 389.7403, df = 1, p-value = 0
## 
## 
##                               Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |      FALSE
## ---------+-----------
##     TRUE |   19.74184
##          |     0.0000*
## 
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   0.518 (95% CI: 0.468 - 0.564 )
## 
## Test for the significance of differences in Bindingness over PA2 :
## 
##    Kruskal-Wallis rank sum test
## 
## data: x and group
## Kruskal-Wallis chi-squared = 0.0271, df = 1, p-value = 0.87
## 
## 
##                               Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |      FALSE
## ---------+-----------
##     TRUE |   0.164719
##          |     0.8692
## 
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =   3.61e-05 (95% CI: 2.76e-06 - 0.00677 )
## 
## Test for the significance of differences in Bindingness over PA3 :
## 
##    Kruskal-Wallis rank sum test
## 
## data: x and group
## Kruskal-Wallis chi-squared = 19.5469, df = 1, p-value = 0
## 
## 
##                               Comparison of x by group
##                                       (Bonferroni)
## Col Mean-|
## Row Mean |      FALSE
## ---------+-----------
```

```
##      TRUE |  -4.421185
##           |    0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.026 (95% CI: 0.00776 - 0.0531 )
##
## Test for the significance of differences in Bindingness over PA5 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 100.7037, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                     (Bonferroni)
## Col Mean-|
## Row Mean |      FALSE
## ---------+-----------
##      TRUE |   10.03512
##           |    0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.134 (95% CI: 0.0905 - 0.183 )
##
## Test for the significance of differences in Bindingness over PA6 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 41.3619, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                     (Bonferroni)
## Col Mean-|
## Row Mean |      FALSE
## ---------+-----------
##      TRUE |  -6.431318
##           |    0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.055 (95% CI: 0.026 - 0.0925 )
##
## Test for the significance of differences in Bindingness over PA4 :
##
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 31.9676, df = 1, p-value = 0
##
```

```
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |      FALSE
## ---------+-----------
##     TRUE |  -5.653993
##          |     0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0425 (95% CI: 0.0177 - 0.0789 )
##
## Test for the significance of differences in Bindingness over PA7 :
##
##    Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 23.2128, df = 1, p-value = 0
##
##
##                              Comparison of x by group
##                                      (Bonferroni)
## Col Mean-|
## Row Mean |      FALSE
## ---------+-----------
##     TRUE |   4.817963
##          |     0.0000*
##
## alpha = 0.05
## Reject Ho if p <= alpha
## epsilon2 =  0.0309 (95% CI: 0.0116 - 0.0609 )
##
##   factor   chi2 kruskal_p epsilon2_lci epsilon2 epsilon2_uci
## 1    PA1 389.74  < 0.0001        0.468    0.518        0.564
## 2    PA2   0.03      0.87        0.000    0.000        0.007
## 3    PA3  19.55  < 0.0001        0.008    0.026        0.053
## 4    PA5 100.70  < 0.0001        0.090    0.134        0.183
## 5    PA6  41.36  < 0.0001        0.026    0.055        0.092
## 6    PA4  31.97  < 0.0001        0.018    0.043        0.079
## 7    PA7  23.21  < 0.0001        0.012    0.031        0.061
##
## p < 5e-2 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-2 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-3 found in: PA1 PA3 PA5 PA6 PA4 PA7
## p < 1e-4 found in: PA1 PA3 PA5 PA6 PA4 PA7
```

Feature-factor correlations

```r
broad_data_factors_corr <- broad_data$feat_long %>%
  group_by(feat, factor) %>%
  summarize(correlation = cor(feat_value, factor_score))
```

```
## `summarise()` has grouped output by 'feat'. You can override using the
```

```
## `.groups` argument.
broad_data_factors_corr %>%
  filter(feat %in% rownames(fa_broad$loadings)) %>%
  ggplot(aes(
    x = factor,
    y = feat,
    fill = correlation,
    label = round(correlation, 2)
  )) +
  geom_tile() +
  geom_text() +
  scale_fill_gradient2(limits = c(-1, 1))
```
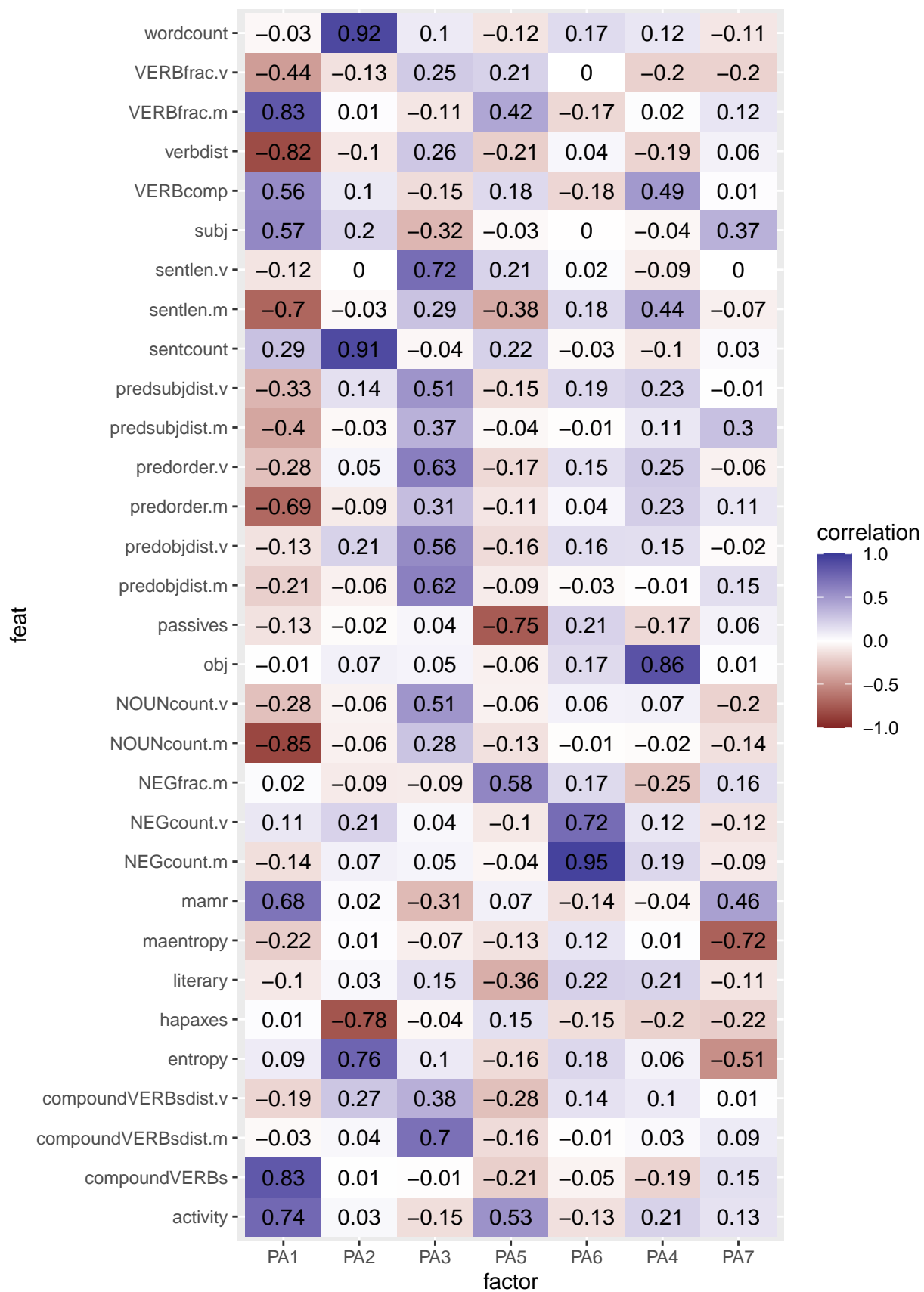
```
broad_data_factors_corr %>%
  filter(!(feat %in% rownames(fa_broad$loadings))) %>%
  ggplot(aes(
    x = factor,
    y = feat,
    fill = correlation,
    label = round(correlation, 2)
  )) +
  geom_tile() +
  geom_text() +
  scale_fill_gradient2(limits = c(-1, 1)) +
  labs(x = "factors", y = "variables") +
  theme_minimal()
```
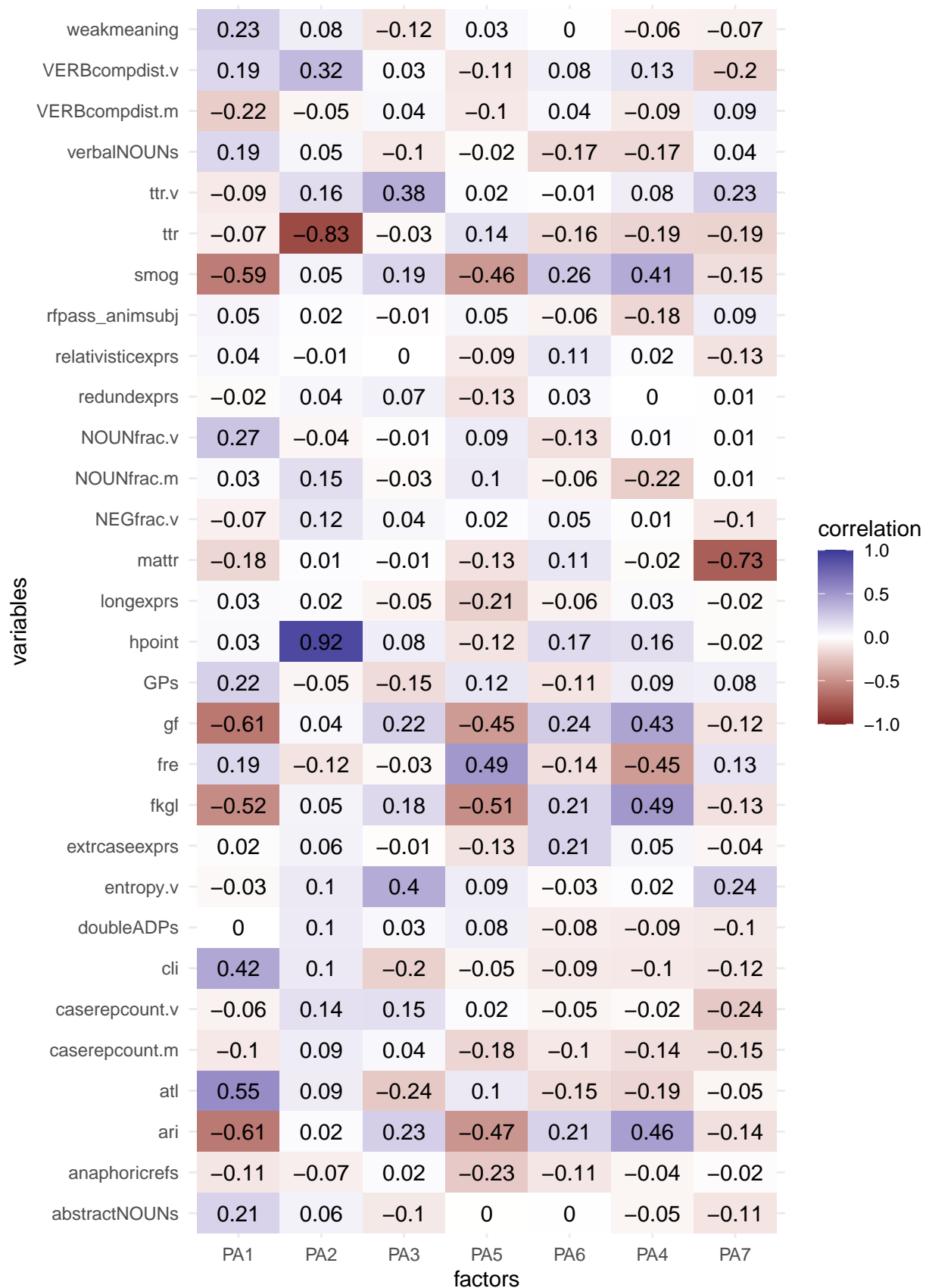
```
ggsave("varfactcorr.pdf")
```

```
## Saving 6.5 x 9 in image
```