Descriptives & Zero-order associations — HPT (Czech data)

Means/SDs, distributions, correlations, and school/class variation

HPT and Extremism project

2025-12-09

Contents

1 Purpose

This file establishes transparent baseline patterns—distributions, scale descriptives, zero-order correlations, and between-group variation—before any modeling. It follows the HPT framework (POP/ROA/CONT) and scoring conventions used in prior work, so higher HPT scores reflect better contextualization/agent-sensitive reasoning; FR-LF and KSA reflect ideological/authoritarian agreement; SDR reflects social-desirability responding. These descriptives document the empirical landscape your hypotheses build on. *Interpretation cheatsheet appears under each table/figure.*

1.1 1) Data & packages

```r
# Core
library(dplyr); library(stringr); library(tidyr); library(purrr); library(readxl)
# Tables & plots
library(knitr); library(kableExtra); library(ggplot2); library(scales)
# Correlations
library(psych);           # corr.test
# Multilevel & ICC
library(lme4);            # lmer
suppressPackageStartupMessages(library(performance)) # icc()

theme_set(theme_minimal(base_size = 11))
```

```r
# Load the dataset created in 00_data-preparation
load("normalised_responses.RData")
stopifnot(exists("normalised_responses"))

all_data <- normalised_responses

# Ensure factors as in codebook
all_data <- all_data %>%
  mutate(
    school_id    = as.factor(school_id),
    class_label  = as.factor(class_label),
    school_level = as.factor(school_level),
    school_type  = as.factor(school_type),
    region       = as.factor(region),
    gender       = as.factor(gender)
  )
```

1.2 2) Scoring: constructs & subscores

```r
# Convenience scorer: mean across items with minimum answered requirement
scale_mean <- function(df, items, min_n = ceiling(length(items)/2), na.rm = TRUE) {
  x <- df[, items]
  ok <- rowSums(!is.na(x)) >= min_n
  out <- rowMeans(x, na.rm = na.rm)
  out[!ok] <- NA_real_
```

```r
  out
}

# HPT (1-4): subscores and total (average so all stay on 1-4)
hpt_pop_items  <- c("POP1","POP2","POP3")
hpt_roa_items  <- c("ROA1","ROA2","ROA3")
hpt_cont_items <- c("CONT1","CONT2","CONT3")

# Knowledge (0-6 correct)
kn_items <- paste0("KN", 1:6)

# FR-LF mini (1-5 Likert): RD1-3 + NS1-3
frlf_rd <- paste0("RD", 1:3)
frlf_ns <- paste0("NS", 1:3)

# KSA-3 authoritarianism (1-5 Likert): 9 items
ksa_items <- c(paste0("A",1:3), paste0("U",1:3), paste0("K",1:3))

# SDR-5 (1-5 Likert): SDR2-SDR4 already reversed in the dataset per codebook
sdr_items <- paste0("SDR", 1:5)

dat <- all_data %>%
  mutate(
    HPT_POP   = scale_mean(., hpt_pop_items,  min_n = 2),
    HPT_ROA   = scale_mean(., hpt_roa_items,  min_n = 2),
    HPT_CONT  = scale_mean(., hpt_cont_items, min_n = 2),
    HPT_TOTAL = scale_mean(., c(hpt_pop_items, hpt_roa_items, hpt_cont_items), min_n = 5),
    KN_TOTAL  = rowSums(select(., all_of(kn_items)), na.rm = TRUE),
    FRLF_RD   = scale_mean(., frlf_rd, min_n = 2),
    FRLF_NS   = scale_mean(., frlf_ns, min_n = 2),
    FRLF_MINI = scale_mean(., c(frlf_rd, frlf_ns), min_n = 4),
    KSA_TOTAL = scale_mean(., ksa_items, min_n = 7),
    SDR_TOTAL = scale_mean(., sdr_items, min_n = 4)
  )
```

How to read: • **HPT**: 1-4, higher = better fit of reasoning to historical context/agent constraints (POP/ROA/CONT per instrument). • **Knowledge (KN_TOTAL)**: 0-6 correct. • **FR-LF mini (FRLF_MINI; plus RD, NS)**: 1-5, higher = stronger endorsement (e.g., leader/one-party; NS relativization). • **KSA-3 (KSA_TOTAL)**: 1-5, higher = stronger authoritarianism. • **SDR_TOTAL**: 1-5, higher = stronger social desirability response tendency. Variable names and coding follow the project codebook.

1.3 3) Sample overview

```
vars_context <- c("school_id","class_label","school_level","school_type","region","gender","history_grade")
n_raw <- nrow(all_data); n_anal <- nrow(dat)

kable(data.frame(
  N_raw = n_raw,
  N_after_scoring = n_anal,
  classes = dplyr::n_distinct(dat$class_label),
  schools = dplyr::n_distinct(dat$school_id)
), caption = "Sample counters (after loading and scoring).") %>%
  kable_styling(full_width = FALSE)
```

Table 1: Sample counters (after loading and scoring).

N_raw	N_after_scoring	classes	schools
156	156	11	7

1.4 4) Descriptives (means, SDs, n, range)

```
desc_vars <- c("HPT_POP","HPT_ROA","HPT_CONT","HPT_TOTAL",
               "KN_TOTAL","FRLF_RD","FRLF_NS","FRLF_MINI",
               "KSA_TOTAL","SDR_TOTAL")

desc_tbl <- dat %>%
  summarise(across(all_of(desc_vars),
                   list(n = ~sum(!is.na(.)),
                        mean = ~mean(., na.rm=TRUE),
                        sd = ~sd(., na.rm=TRUE),
                        min = ~min(., na.rm=TRUE),
                        max = ~max(., na.rm=TRUE)))) %>%
  pivot_longer(everything(),
               names_to = c("variable",".value"),
               names_pattern = "([^_]+_[^_]+|[^_]+)_(n|mean|sd|min|max)")

# Keep original order
desc_tbl$variable <- factor(desc_tbl$variable, levels = desc_vars)
```

```r
kable(desc_tbl[order(desc_tbl$variable),], booktabs = TRUE, digits = 2,
      caption = "Scale descriptives (student-level).") %>%
  kable_styling(latex_options = c("striped","hold_position"), full_width = FALSE)
```

Table 2: Scale descriptives (student-level).

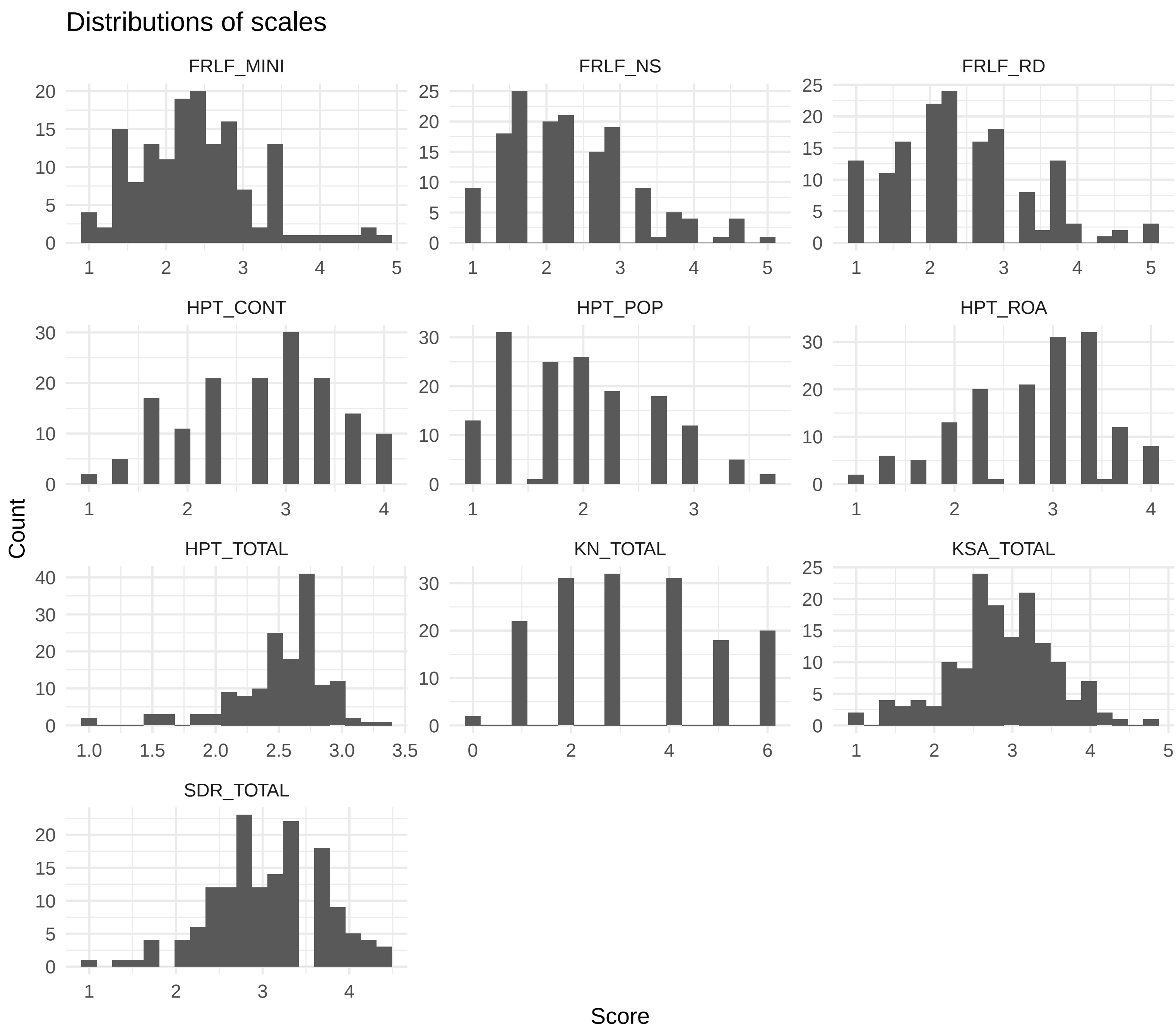
variable	n	mean	sd	min	max
HPT_POP	152	1.99	0.66	1	3.67
HPT_ROA	152	2.82	0.68	1	4.00
HPT_CONT	152	2.73	0.74	1	4.00
HPT_TOTAL	152	2.51	0.39	1	3.33
KN_TOTAL	156	3.29	1.62	0	6.00
FRLF_RD	152	2.45	0.93	1	5.00
FRLF_NS	152	2.35	0.89	1	5.00
FRLF_MINI	151	2.40	0.77	1	4.83
KSA_TOTAL	151	2.87	0.67	1	4.78
SDR_TOTAL	151	3.05	0.64	1	4.40

Interpretation: HPT scales should center above 2.5 if students generally avoid presentism and contextualize; FR-LF/KSA are not "good/bad" in isolation but provide ideological context for later modeling; SDR alerts to response bias.

1.5 5) Distributions (histograms, same axes where possible)

```r
long_scales <- dat %>%
  select(all_of(desc_vars)) %>%
  pivot_longer(everything(), names_to = "scale", values_to = "value")

ggplot(long_scales, aes(x = value)) +
  geom_histogram(bins = 20) +
  facet_wrap(~scale, scales = "free", ncol = 3) +
  labs(title = "Distributions of scales", x = "Score", y = "Count")
```



Distributions of scales
FRLF_MINI
FRLF_NS
FRLF_RD
HPT_CONT
HPT_POP
HPT_ROA
HPT_TOTAL
KN_TOTAL
KSA_TOTAL
SDR_TOTAL
Count
Score

Interpretation: Watch for spikes at bounds (e.g., KN at 0 or max), floor/ceiling on FR-LF/KSA, and skew in HPT subscores—useful cues for later transformation or robust modeling.

1.6 6) Zero-order correlations (student level)

```r
corr_df <- dat %>% select(all_of(desc_vars)) %>% drop_na()
ct <- psych::corr.test(corr_df, use = "pairwise")
round(ct$r, 2)
```

```
##           HPT_POP HPT_ROA HPT_CONT HPT_TOTAL KN_TOTAL FRLF_RD FRLF_NS FRLF_MINI
## HPT_POP      1.00   -0.14    -0.36      0.25    -0.30    0.08    0.10      0.11
## HPT_ROA     -0.14    1.00     0.43      0.77     0.27   -0.07   -0.06     -0.08
## HPT_CONT    -0.36    0.43     1.00      0.67     0.25   -0.10   -0.02     -0.07
## HPT_TOTAL    0.25    0.77     0.67      1.00     0.15   -0.05    0.01     -0.03
## KN_TOTAL    -0.30    0.27     0.25      0.15     1.00   -0.13   -0.21     -0.20
## FRLF_RD      0.08   -0.07    -0.10     -0.05    -0.13    1.00    0.41      0.84
## FRLF_NS      0.10   -0.06    -0.02      0.01    -0.21    0.41    1.00      0.83
## FRLF_MINI    0.11   -0.08    -0.07     -0.03    -0.20    0.84    0.83      1.00
## KSA_TOTAL    0.12    0.07     0.08      0.17    -0.01    0.48    0.38      0.51
## SDR_TOTAL    0.10    0.00     0.06      0.10     0.02   -0.04   -0.21     -0.14
##           KSA_TOTAL SDR_TOTAL
## HPT_POP        0.12      0.10
## HPT_ROA        0.07      0.00
## HPT_CONT       0.08      0.06
## HPT_TOTAL      0.17      0.10
## KN_TOTAL      -0.01      0.02
## FRLF_RD        0.48     -0.04
## FRLF_NS        0.38     -0.21
## FRLF_MINI      0.51     -0.14
## KSA_TOTAL      1.00     -0.13
## SDR_TOTAL     -0.13      1.00
```

```r
corr_tab <- as.data.frame(round(ct$r, 2))
corr_tab$Var1 <- rownames(corr_tab)
corr_tab <- corr_tab %>% relocate(Var1)

kable(corr_tab, booktabs = TRUE, caption = "Zero-order Pearson correlations among constructs (pairwise).") %>%
  kable_styling(latex_options = c("hold_position"))
```

Table 3: Zero-order Pearson correlations among constructs (pairwise).

	Var1	HPT_POP	HPT_ROA	HPT_CONT	HPT_TOTAL	KN_TOTAL	FRLF_RD	FRLF_NS	FRLF_MINI	KSA_TOTA
HPT_POP	HPT_POP	1.00	-0.14	-0.36	0.25	-0.30	0.08	0.10	0.11	0.
HPT_ROA	HPT_ROA	-0.14	1.00	0.43	0.77	0.27	-0.07	-0.06	-0.08	0.
HPT_CONT	HPT_CONT	-0.36	0.43	1.00	0.67	0.25	-0.10	-0.02	-0.07	0.
HPT_TOTAL	HPT_TOTAL	0.25	0.77	0.67	1.00	0.15	-0.05	0.01	-0.03	0.
KN_TOTAL	KN_TOTAL	-0.30	0.27	0.25	0.15	1.00	-0.13	-0.21	-0.20	-0.
FRLF_RD	FRLF_RD	0.08	-0.07	-0.10	-0.05	-0.13	1.00	0.41	0.84	0.
FRLF_NS	FRLF_NS	0.10	-0.06	-0.02	0.01	-0.21	0.41	1.00	0.83	0.
FRLF_MINI	FRLF_MINI	0.11	-0.08	-0.07	-0.03	-0.20	0.84	0.83	1.00	0.
KSA_TOTAL	KSA_TOTAL	0.12	0.07	0.08	0.17	-0.01	0.48	0.38	0.51	1.
SDR_TOTAL	SDR_TOTAL	0.10	0.00	0.06	0.10	0.02	-0.04	-0.21	-0.14	-0.

Interpretation: Correlations locate broad relationships your hypotheses rely on. For example, if **FRLF_MINI** correlates positively with **HPT_CONT/HPT_TOTAL**, this suggests possible ideological alignment inflating contextualization—an effect to test with controls in models (knowledge, SDR) and with item-level checks later.

1.7 7) Between-group variation: school/class ICCs

We estimate the variance attributable to schools and classes with random-intercept models. ICC proportion of total variance that is between clusters.

```r
# --- Robust ICC helpers (replace previous version) ---

# Safely pull ICC_adjusted (fallback to ICC or first numeric) for different object types
get_icc_value <- function(ic) {
  # ic can be data.frame/tibble, list, or numeric
  if (is.null(ic)) return(NA_real_)
  if (is.data.frame(ic)) {
    # new performance::icc often returns a data.frame with ICC and ICC_adjusted
    if ("ICC_adjusted" %in% names(ic)) return(suppressWarnings(as.numeric(ic$ICC_adjusted[1])))
    if ("ICC" %in% names(ic))          return(suppressWarnings(as.numeric(ic$ICC[1])))
    # else, first numeric column
    num_cols <- which(vapply(ic, is.numeric, logical(1)))
    if (length(num_cols)) return(as.numeric(ic[[ num_cols[1] ]][1]))
    return(NA_real_)
  }
  if (is.list(ic)) {
    if (!is.null(ic$ICC_adjusted)) return(suppressWarnings(as.numeric(ic$ICC_adjusted)))
    if (!is.null(ic$ICC))          return(suppressWarnings(as.numeric(ic$ICC)))
```

```r
    # first numeric element
    nums <- unlist(ic[ vapply(ic, is.numeric, logical(1)) ], use.names = FALSE)
    if (length(nums)) return(as.numeric(nums[1]))
    return(NA_real_)
  }
  if (is.atomic(ic) && is.numeric(ic)) return(as.numeric(ic[1]))
  NA_real_
}

fit_icc <- function(v) {
  # guard: need some variance and at least 2 clusters
  if (!v %in% names(dat)) return(NULL)
  f_cls <- as.formula(paste0(v, " ~ 1 + (1|school_id/class_label)"))
  f_sch <- as.formula(paste0(v, " ~ 1 + (1|school_id)"))
  list(
    class_in_school = tryCatch(lme4::lmer(f_cls, data = dat, REML = TRUE, na.action = na.omit),
                               error = function(e) NULL),
    school_only     = tryCatch(lme4::lmer(f_sch, data = dat, REML = TRUE, na.action = na.omit),
                               error = function(e) NULL)
  )
}

extract_icc <- function(fm) {
  if (is.null(fm)) return(list(ICC = NA_real_, N = NA_integer_, clusters = NA_integer_))
  # Try ICC; handle different output types
  ic <- tryCatch(performance::icc(fm), error = function(e) NULL)
  icc_val <- get_icc_value(ic)
  # Sample size and cluster count
  N <- tryCatch(nobs(fm), error = function(e) NA_integer_)
  # Number of levels of the *first* grouping factor in the model
  clusters <- tryCatch({
    fl <- lme4::getME(fm, "flist")
    length(levels(fl[[1]]))
  }, error = function(e) NA_integer_)
  list(ICC = icc_val, N = N, clusters = clusters)
}

targets <- c("HPT_TOTAL","HPT_POP","HPT_ROA","HPT_CONT",
             "FRLF_MINI","KSA_TOTAL","KN_TOTAL","SDR_TOTAL")

icc_rows <- purrr::map(targets, function(sc) {
  mods <- fit_icc(sc)
```

```r
  # extract once per model
  cls <- extract_icc(mods$class_in_school)
  sch <- extract_icc(mods$school_only)
  data.frame(
    scale = sc,
    ICC_class_in_school = round(cls$ICC, 3),
    N_class_in_school   = as.integer(cls$N),
    clusters_classes    = as.integer(cls$clusters),
    ICC_school_only     = round(sch$ICC, 3),
    N_school_only       = as.integer(sch$N),
    clusters_schools    = as.integer(sch$clusters)
  )
})
icc_res <- dplyr::bind_rows(icc_rows)
```

```r
kable(icc_res, booktabs = TRUE,
      caption = "Intraclass correlations (ICCs): class (nested in school) and school.") %>%
  kable_styling(latex_options = c("striped","hold_position"), full_width = FALSE)
```

Table 4: Intraclass correlations (ICCs): class (nested in school) and school.

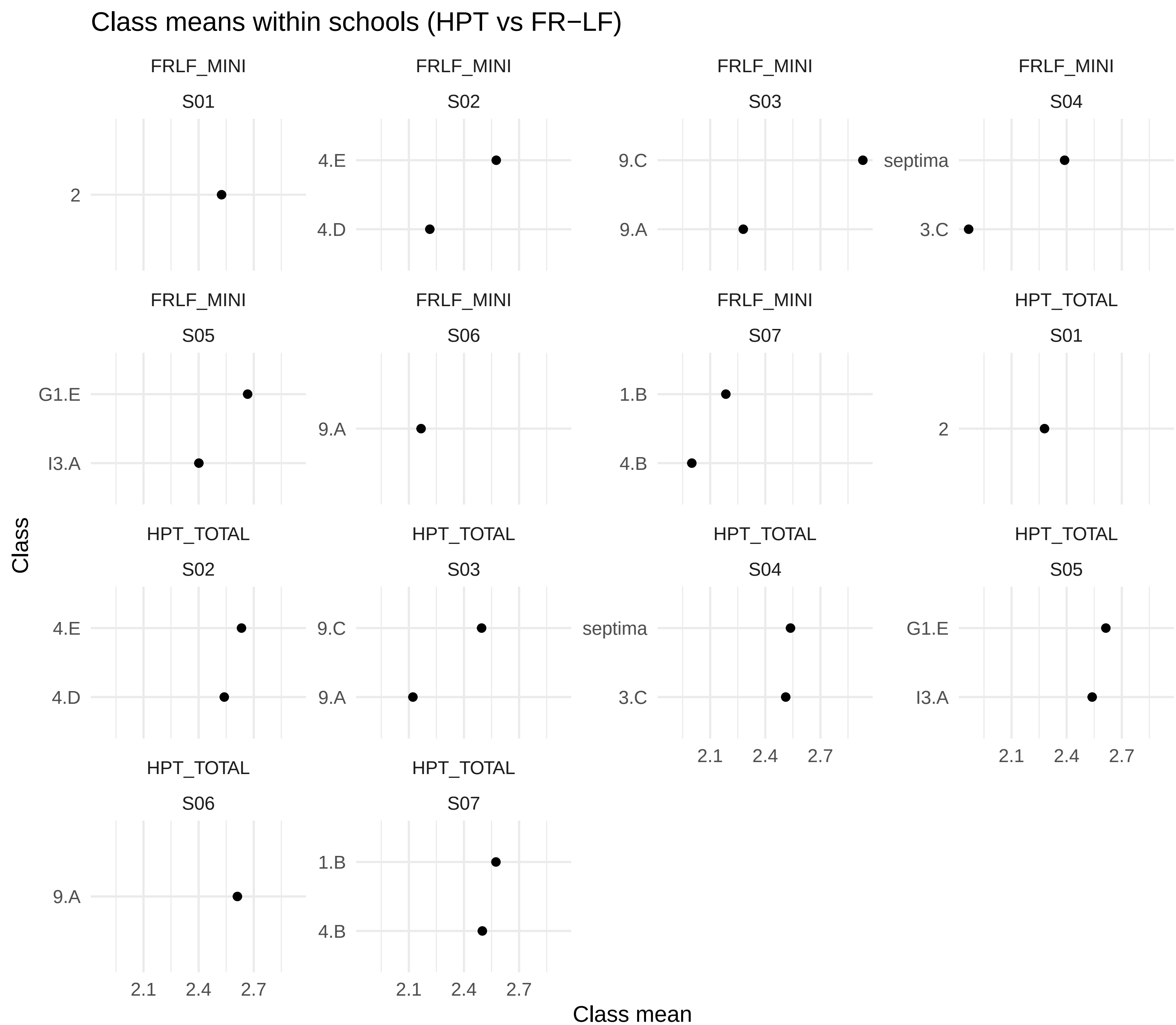
scale	ICC_class_in_school	N_class_in_school	clusters_classes	ICC_school_only	N_school_only	clusters_schools
HPT_TOTAL	0.043	152	12	0.036	152	7
HPT_POP	NA	152	12	0.038	152	7
HPT_ROA	0.046	152	12	0.043	152	7
HPT_CONT	0.060	152	12	0.047	152	7
FRLF_MINI	0.077	151	12	0.052	151	7
KSA_TOTAL	0.161	151	12	0.124	151	7
KN_TOTAL	NA	156	12	0.030	156	7
SDR_TOTAL	NA	151	12	NA	151	7

Interpretation: • **ICC** ~ **0.00-0.05** → little clustering (ordinary regression OK). • **ICC** ~ **0.05-0.15** → modest clustering; use cluster-robust SE or multilevel models. • **ICC > 0.15** → strong clustering; multilevel modeling recommended. Compare **HPT** vs **FR-LF/KSA**: larger ICCs for ideology might reflect school milieu; larger ICCs for HPT may signal classroom task/context effects.

1.8 8) Quick sanity plots by class/school (optional)

```r
p1 <- dat %>%
  group_by(school_id, class_label) %>%
  summarise(n = n(),
            HPT_TOTAL = mean(HPT_TOTAL, na.rm=TRUE),
            FRLF_MINI = mean(FRLF_MINI, na.rm=TRUE),
            .groups = "drop") %>%
  pivot_longer(c(HPT_TOTAL, FRLF_MINI), names_to="scale", values_to="mean") %>%
  ggplot(aes(x = reorder(class_label, mean), y = mean)) +
  geom_point() +
  facet_wrap(scale ~ school_id, scales = "free_y") +
  coord_flip() +
  labs(x = "Class", y = "Class mean", title = "Class means within schools (HPT vs FR-LF)")

p1
```


Class means within schools (HPT vs FR−LF)
FRLF_MINI
S01
2
FRLF_MINI
S02
4.E
4.D
FRLF_MINI
S03
9.C
9.A
FRLF_MINI
S04
septima
3.C
FRLF_MINI
S05
G1.E
I3.A
FRLF_MINI
S06
9.A
FRLF_MINI
S07
1.B
4.B
HPT_TOTAL
S01
2
HPT_TOTAL
S02
4.E
4.D
HPT_TOTAL
S03
9.C
9.A
HPT_TOTAL
S04
septima
3.C
HPT_TOTAL
S05
G1.E
I3.A
HPT_TOTAL
S06
9.A
HPT_TOTAL
S07
1.B
4.B
2.1
2.4
2.7
Class
Class mean

Interpretation: Visual check for unusually high/low classes can inform later sensitivity checks (e.g., re-running models without extreme classes).

1.9 9) Reproducibility appendix

```r
sessionInfo()
```

```
## R version 4.4.2 (2024-10-31)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 24.04.3 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.12.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.12.0
##
## locale:
##  [1] LC_CTYPE=C.UTF-8           LC_NUMERIC=C
##  [3] LC_TIME=cs_CZ.UTF-8        LC_COLLATE=C.UTF-8
##  [5] LC_MONETARY=cs_CZ.UTF-8    LC_MESSAGES=C.UTF-8
##  [7] LC_PAPER=cs_CZ.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=cs_CZ.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Prague
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] performance_0.15.1 lme4_1.1-38        Matrix_1.7-1       psych_2.4.12
##  [5] scales_1.4.0       ggplot2_4.0.1      kableExtra_1.4.0   knitr_1.50
##  [9] readxl_1.4.3       purrr_1.1.0        tidyr_1.3.1        stringr_1.5.1
## [13] dplyr_1.1.4
##
## loaded via a namespace (and not attached):
##  [1] generics_0.1.3     xml2_1.3.6         stringi_1.8.4      lattice_0.22-5
##  [5] digest_0.6.37      magrittr_2.0.3     evaluate_1.0.5     grid_4.4.2
##  [9] RColorBrewer_1.1-3 fastmap_1.2.0      cellranger_1.1.0   viridisLite_0.4.2
## [13] textshaping_0.4.1  reformulas_0.4.1   Rdpack_2.6.4       mnormt_2.1.1
## [17] cli_3.6.5          rlang_1.1.6        rbibutils_2.3      splines_4.4.2
```

```
## [21] withr_3.0.2       yaml_2.3.10       tools_4.4.2       parallel_4.4.2
## [25] nloptr_2.2.1      minqa_1.2.8       boot_1.3-31       vctrs_0.6.5
## [29] R6_2.6.1          lifecycle_1.0.4   MASS_7.3-61       insight_1.4.2
## [33] pkgconfig_2.0.3   pillar_1.10.0     gtable_0.3.6      Rcpp_1.0.13-1
## [37] glue_1.8.0        systemfonts_1.3.1 xfun_0.54         tibble_3.2.1
## [41] tidyselect_1.2.1  rstudioapi_0.17.1 farver_2.1.2      htmltools_0.5.8.1
## [45] nlme_3.1-166      labeling_0.4.3    rmarkdown_2.29    svglite_2.2.2
## [49] compiler_4.4.2    S7_0.2.1
```

1.9.1 Notes & interpretation pointers

- **HPT scales (1-4):** Higher indicates better contextualization/agent-sensitive reasoning—consistent with the three-part structure (POP/ROA/CONT) used in prior validation work. Consider that ideological alignment can mimic contextualization; correlations here are descriptive only and motivate the multilevel models planned in the main analysis.
- **FR-LF mini (1-5):** Short right-wing authoritarian/Nazi relativization composite; higher = stronger endorsement. Use primarily as a predictor/covariate and for DIF checks later.
- **Knowledge:** Treat as a covariate; it often shows small but non-zero links to HPT.
- **SDR:** Use to check attenuation/amplification of sensitive attitudes.