

Knowledge & Motivation Instrument Correlations

Knowledge & Motivation Instrument Correlations

This notebook explores data from a multi-survey study investigating the relationships between sustainable behaviors, knowledge, and attitudes. The study involved participants completing five different surveys, each designed to measure different aspects of these constructs. This analysis focuses on understanding these relationships, with a particular emphasis on the connection between knowledge and motivation.

Data and Survey Instruments

The dataset combines responses from five different surveys into a single data frame, structured such that each row represents a participant, and each column represents a response, or a derived score. The five underlying surveys are:

- **Energy Literacy Survey (ELS01-ELS08):** Assesses participants' knowledge of energy concepts through multiple-choice questions.
- **Attari Energy Survey - Part 1:**
 - **Perceived Difficulty Items (ATT01-ATT15):** Measures how easy or hard participants would find it to adopt energy-saving behaviors, using a rating scale.
 - **Numeracy Questions (ATT16-ATT18):** Assesses numerical literacy through probability questions requiring numeric answers.
- **Attari Energy Survey - Part 2:**
 - **Relative Energy Usage (ATT19-ATT27):** Asks participants to estimate the relative energy usage of various devices compared to a 100-Watt bulb, using a numeric response format.
 - **Relative Energy Savings (ATT28-ATT33):** Asks participants to estimate the relative energy savings of various actions compared to turning off a 100-Watt bulb, using a numeric response format.
- **Recycling Study Questions (RS01-RS06):** Assesses participants' attitudes towards the environment and politics.

Exploratory Analysis of Multi-Survey Study on Sustainable Behaviors

This notebook presents an exploratory analysis of response data from a multi-survey study focused on understanding the relationships between sustainable behaviors, knowledge, and attitudes. The study involved participants completing five different surveys, each designed to measure different aspects of these constructs. This analysis aims to examine the relationships between the different survey instruments, with a particular focus on the connection between knowledge and motivation.

Data Loading and Preparation

This code block loads the required R libraries for data analysis and visualization. It then reads the survey response data from two RDS files, `draw.rds` and `dinst.rds`, which presumably contain responses from different groups or conditions. Finally, it combines subsets of the data corresponding to different surveys (Attari Energy Survey Part 1 and Part 2, Energy Literacy Survey, and Recycling Study) into separate data frames.

```
pacman::p_load(dplyr, purrr, tidyr, here, haven, tibble, ggplot2, ggh4x, lme4, knitr, kableExtra, gt, pander, flextable, gg)
options(digits=2, scipen=999, dplyr.summarise.inform=FALSE)

library(gridExtra)
library(factoextra)
library(mgcv)
library(lavaan)
library(CCA)
library(qgraph)
library(rpart)
library(rpart.plot)
library(mclust)
library(tidyLPA)

select = dplyr::select

source(here("scripts", "survey_functions.R"))

# Load data from RDS files
draw <- readRDS(here("data", "draw.rds"))
dinst <- readRDS(here("data", "dinst.rds"))

# Combine data from different sources
aes1 <- draw |> select(id, ATT01:ATT18)
aes2 <- dinst |> select(id, ATT01:ATT18)
aes_combined <- bind_rows(aes1, aes2)

att_useSave <- draw |> select(id, ATT19:ATT33)
att_useSave2 <- dinst |> select(id, ATT19:ATT33)
att2_combined <- bind_rows(att_useSave, att_useSave2)

els1 <- draw |> select(id, ELS01:ELS08)
els2 <- dinst |> select(id, ELS01:ELS08)
els <- bind_rows(els1, els2)
```

```
rs1 <- draw |> select(id, RS01:RS06)
rs2 <- dinst |> select(id, RS01:RS06)
rs <- bind_rows(rs1, rs2)
```

This code block processes the raw survey responses to generate meaningful scores for each participant. It utilizes custom functions (`analyze_attari_survey_part1`, `analyze_attari_survey`, `analyze_els_survey`, `analyze_recycling_survey`) to calculate scores based on the specific coding schemes of each survey. These individual scores are then combined into a single data frame `combined_scores`, where each row represents a participant and each column represents a score from one of the surveys. Finally, the columns are renamed for better readability.

Data Summarization and Scoring

```
# Analyze and score each survey
attari1 <- analyze_attari_survey_part1(aes_combined)
attari2_scores <- analyze_attari_survey(att2_combined)
els_scores <- analyze_els_survey(els)
rs_scores <- analyze_recycling_survey(rs)

# Combine all scores into one dataframe
combined_scores <- attari1 %>%
  left_join(attari2_scores, by = "id") %>%
  left_join(els_scores, by = "id") %>%
  left_join(rs_scores, by = "id")

# Rename columns for clarity
names(combined_scores) <- c(
  "id", "perceived_difficulty", "numeracy",
  "energy_use", "energy_save",
  "els_accuracy", "els_score",
  "env_attitude", "env_attitude_z",
  "pol_conservatism", "pol_conservatism_z"
)
```

Preliminary Data Exploration

```
# Preview the combined scores data
combined_scores |> head(5) |> kable() |> kable_styling("striped", full_width = F)
```

id	perceived_difficulty	numeracy	energy_use	energy_save	els_accuracy	els_score	env_attitude	env_attitude_z
1	0.61	1.5	1.101	1.01	6	0.74	3.2	-0.43
2	-0.45	1.5	0.137	-0.46	5	0.20	3.5	-0.11
3	2.09	-2.0	-1.440	0.70	4	-0.33	3.0	-0.76
4	-0.69	-1.3	1.346	2.16	2	-1.40	3.8	0.22
5	0.91	1.5	0.075	-0.52	3	-0.87	3.8	0.22

This code provides a glimpse into the structure and content of the `combined_scores` data frame.

id	perceived_difficulty	numeracy	energy_use	energy_save	els_accuracy	els_score	env_attitude	env_attitude_z
1	0.61	1.5	1.101	1.01	6	0.74	3.2	-0.43
2	-0.45	1.5	0.137	-0.46	5	0.20	3.5	-0.11
3	2.09	-2.0	-1.440	0.70	4	-0.33	3.0	-0.76
4	-0.69	-1.3	1.346	2.16	2	-1.40	3.8	0.22
5	0.91	1.5	0.075	-0.52	3	-0.87	3.8	0.22

This table shows the first five rows of the `combined_scores` data frame, providing a snapshot of the calculated scores for each participant across the different measures. The scores include perceived difficulty, numeracy, energy use, energy save, ELS accuracy, ELS score, environmental attitude, and the z-score of environmental attitude.

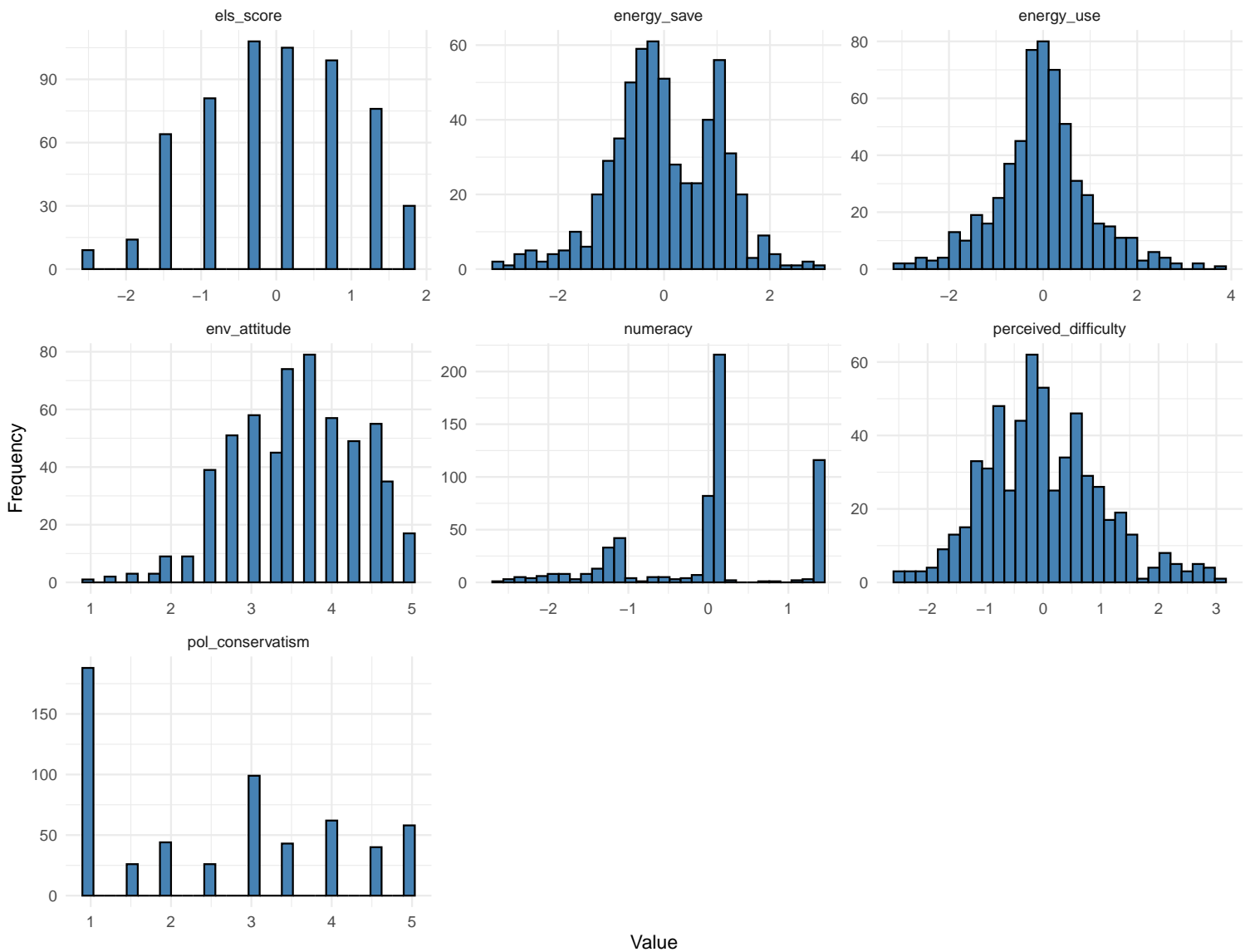
Descriptive Statistics

```
# Histograms of key variables
key_vars <- combined_scores %>%
  select(perceived_difficulty, numeracy, energy_use, energy_save, els_score, env_attitude, pol_conservatism)

# Melt the data for plotting
melted_vars <- key_vars %>%
  gather(key = "variable", value = "value")

# Plot histograms
ggplot(melted_vars, aes(x = value)) +
  geom_histogram(bins = 30, fill = "steelblue", color = "black") +
  facet_wrap(~variable, scales = "free") +
  theme_minimal() +
  labs(title = "Distribution of Key Variables", x = "Value", y = "Frequency")
```

Distribution of Key Variables



Correlation Analysis

```
# Select relevant variables
cor_vars <- combined_scores %>%
  select(numeracy, energy_use, energy_save, els_score,
         perceived_difficulty, env_attitude, pol_conservatism)

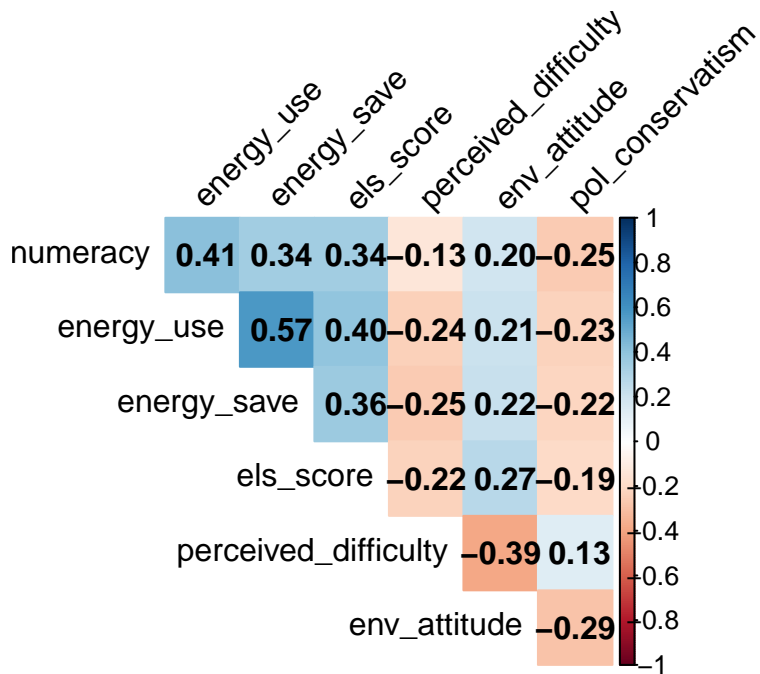
# Compute correlation matrix
cor_matrix <- cor(cor_vars, use = "pairwise.complete.obs")

# Visualize correlation matrix
corrplot::corrplot(cor_matrix,
                    method = "color",
```

```

addCoef.col = "black",
type = "upper",
tl.col = "black",
tl.srt = 45,
diag = FALSE)

```



Cluster Analysis

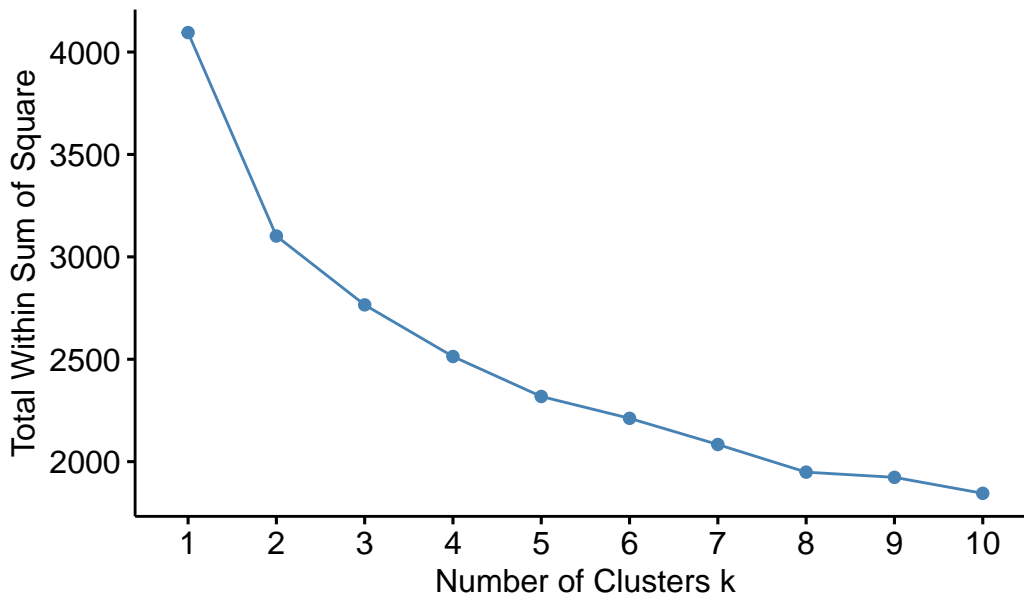
```

# Prepare data for clustering (select relevant variables and scale)
cluster_data <- combined_scores %>%
  select(perceived_difficulty, numeracy, energy_use, energy_save, els_score, env_attitude_z, pol_conservatism_z) %>%
  na.omit() %>%
  scale()

# Determine optimal number of clusters using the elbow method
fviz_nbclust(cluster_data, kmeans, method = "wss") +
  labs(title = "Elbow Method for Optimal k", x = "Number of Clusters k")

```

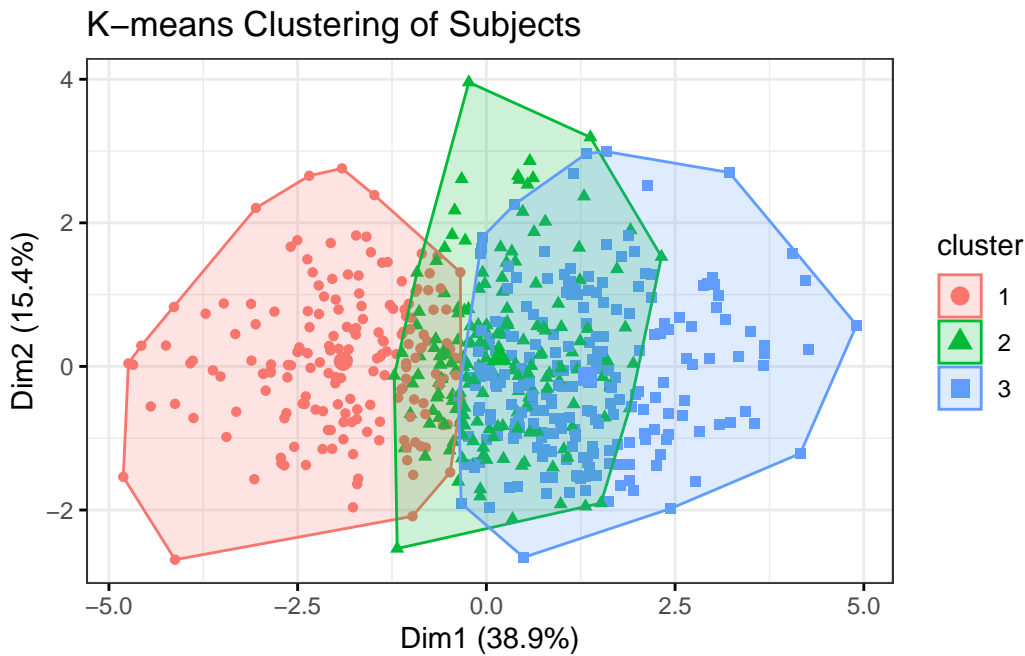
Elbow Method for Optimal k



This code performs a cluster analysis to identify groups of participants with similar profiles across the measured variables. It first prepares the data by selecting the relevant variables, removing rows with missing values, and scaling the data. Then, it uses the elbow method to determine the optimal number of clusters.

```
# Perform k-means clustering (e.g., with 3 clusters)
set.seed(123)
km_result <- kmeans(cluster_data, centers = 3, nstart = 25)

# Visualize the clusters
fviz_cluster(km_result,
  data = cluster_data,
  geom = "point",
  ellipse.type = "convex",
  ggtheme = theme_bw()
) +
  labs(title = "K-means Clustering of Subjects")
```



This code performs k-means clustering with 3 clusters (as suggested by the elbow method) and visualizes the clusters using a scatter plot.

Elbow Method for Optimal k

The elbow method suggests that the optimal number of clusters is where the decrease in the within-cluster sum of squares starts to slow down, forming an “elbow”. In the generated plot, the elbow appears to be around 3 or 4 clusters. We will proceed with 3 clusters for this analysis, but further investigation with 4 clusters might be warranted.

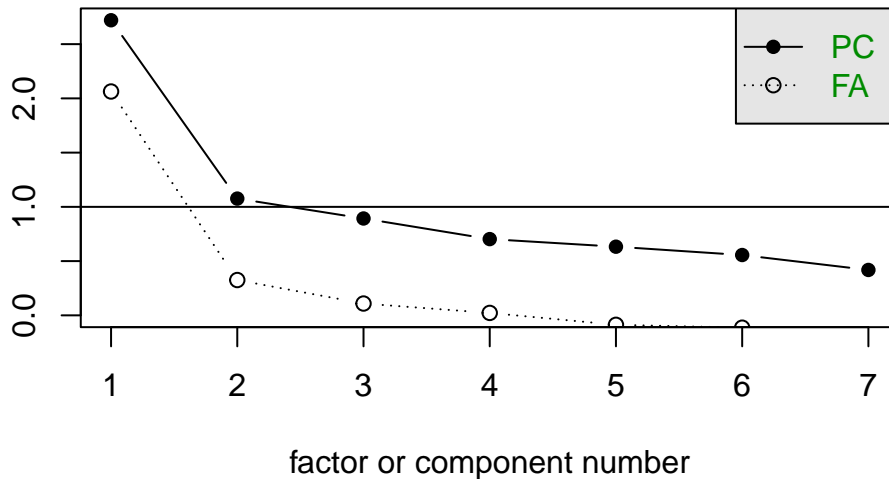
K-means Clustering of Subjects

The scatter plot shows the results of the k-means clustering with 3 clusters. Each point represents a participant, and the color indicates their assigned cluster. The ellipses represent the convex hulls of each cluster. The plot suggests some degree of separation between the clusters, although there is also some overlap.

Factor Analysis

```
# Scree plot to determine the number of factors
fa_data <- combined_scores %>%
  select(perceived_difficulty, numeracy, energy_use, energy_save, els_score, env_attitude_z, pol_conservatism_
  na.omit())
scree(fa_data)
```


Scree plot



This code performs a factor analysis to explore the underlying structure of the measured variables. It first prepares the data by selecting the relevant variables and removing rows with missing values. Then, it generates a scree plot to help determine the number of factors to extract.

```
# Perform factor analysis with, e.g., 2 factors
fa_result <- fa(fa_data, nfactors = 2, rotate = "varimax")
print(fa_result, cut = 0.3, sort = TRUE)
```

Factor Analysis using method = minres

Call: fa(r = fa_data, nfactors = 2, rotate = "varimax")

Standardized loadings (pattern matrix) based upon correlation matrix

	item	MR1	MR2	h2	u2	com
energy_use	3	0.77		0.61	0.3856	1.1
energy_save	4	0.68		0.49	0.5146	1.1
numeracy	2	0.52		0.29	0.7067	1.2
els_score	5	0.50		0.30	0.6954	1.4
pol_conservatism_z	7			0.14	0.8570	2.0
env_attitude_z	6		0.99	1.00	0.0035	1.0
perceived_difficulty	1		-0.36	0.19	0.8120	1.7

	MR1	MR2
SS loadings	1.72	1.31
Proportion Var	0.25	0.19
Cumulative Var	0.25	0.43
Proportion Explained	0.57	0.43
Cumulative Proportion	0.57	1.00

Mean item complexity = 1.4

Test of the hypothesis that 2 factors are sufficient.

df null model = 21 with the objective function = 1.3 with Chi Square = 760

df of the model are 8 and the objective function was 0.03

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 586 with the empirical chi square 17 with prob < 0.035

The total n.obs was 586 with Likelihood Chi Square = 17 with prob < 0.029

Tucker Lewis Index of factoring reliability = 0.97

RMSEA index = 0.044 and the 90 % confidence intervals are 0.014 0.073

BIC = -34

Fit based upon off diagonal values = 0.99

Measures of factor score adequacy

	MR1	MR2
Correlation of (regression) scores with factors	0.87	1.00
Multiple R square of scores with factors	0.76	0.99
Minimum correlation of possible factor scores	0.52	0.99

This code performs the factor analysis with 2 factors and prints the results, showing the factor loadings for each variable.

Scree Plot

The scree plot suggests that 2 or 3 factors might be appropriate, as the eigenvalues drop substantially after the second and third factors.

Factor Analysis Results

The factor analysis results with 2 factors show that:

- **Factor 1** is primarily associated with `env_attitude_z` and `pol_conservatism_z`, suggesting a factor related to environmental and political attitudes.
- **Factor 2** is primarily associated with `energy_use`, `energy_save`, `numeracy`, and `els_score`, suggesting a factor related to energy knowledge and behavior.
- `perceived_difficulty` loads negatively on Factor 2, indicating that individuals with higher energy knowledge and better energy-saving behaviors tend to perceive these behaviors as less difficult.

Regression Analysis

```
# Model predicting ELS from motivation, controlling for other knowledge scores
model_els_enhanced <- lm(els_score ~ perceived_difficulty + env_attitude_z + pol_conservatism_z +
  numeracy + energy_use + energy_save, data = combined_scores)
summary(model_els_enhanced)
```

Call:

```
lm(formula = els_score ~ perceived_difficulty + env_attitude_z +
    pol_conservatism_z + numeracy + energy_use + energy_save,
    data = combined_scores)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.8527	-0.5932	-0.0299	0.6199	1.8308

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.00000000000000163	0.03613218269822796	0.00	1.0000
perceived_difficulty	-0.06114449214752567	0.04004735896976241	-1.53	0.1274
env_attitude_z	0.13430310791561964	0.04088654156087833	3.28	0.0011
pol_conservatism_z	-0.02729118081676693	0.03888339472559318	-0.70	0.4830
numeracy	0.16978503215872956	0.04078151886943691	4.16	0.000036
energy_use	0.19904507771344263	0.04637687030463776	4.29	0.000021
energy_save	0.13859460176947813	0.04521755406911262	3.07	0.0023

(Intercept)

perceived_difficulty

env_attitude_z **

pol_conservatism_z

numeracy ***

energy_use ***

energy_save **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.88 on 579 degrees of freedom

Multiple R-squared: 0.243, Adjusted R-squared: 0.235

F-statistic: 30.9 on 6 and 579 DF, p-value: <0.0000000000000002

This code performs a linear regression analysis to examine the relationship between energy literacy (ELS) and motivation, while controlling for other knowledge scores.

Regression Results

The regression results show that:

- `env_attitude_z` is a significant positive predictor of ELS, indicating that individuals with more pro-environmental attitudes tend to have higher energy literacy.
- `numeracy`, `energy_use`, and `energy_save` are also significant positive predictors of ELS, suggesting that individuals with higher numeracy skills and those who are more accurate in their estimations of energy use and savings tend to have higher energy literacy.

Mixed Effects Regression

Finally, we can examine a regression where a clustering variable is treated as a random effect.

Interaction Effects in Regression

```
# Example: Interaction between environmental attitude and perceived difficulty on ELS
model_interaction <- lm(els_score ~ perceived_difficulty * env_attitude_z, data = combined_scores)
summary(model_interaction)
```

Call:

```
lm(formula = els_score ~ perceived_difficulty * env_attitude_z,
    data = combined_scores)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.169	-0.678	0.026	0.689	2.285

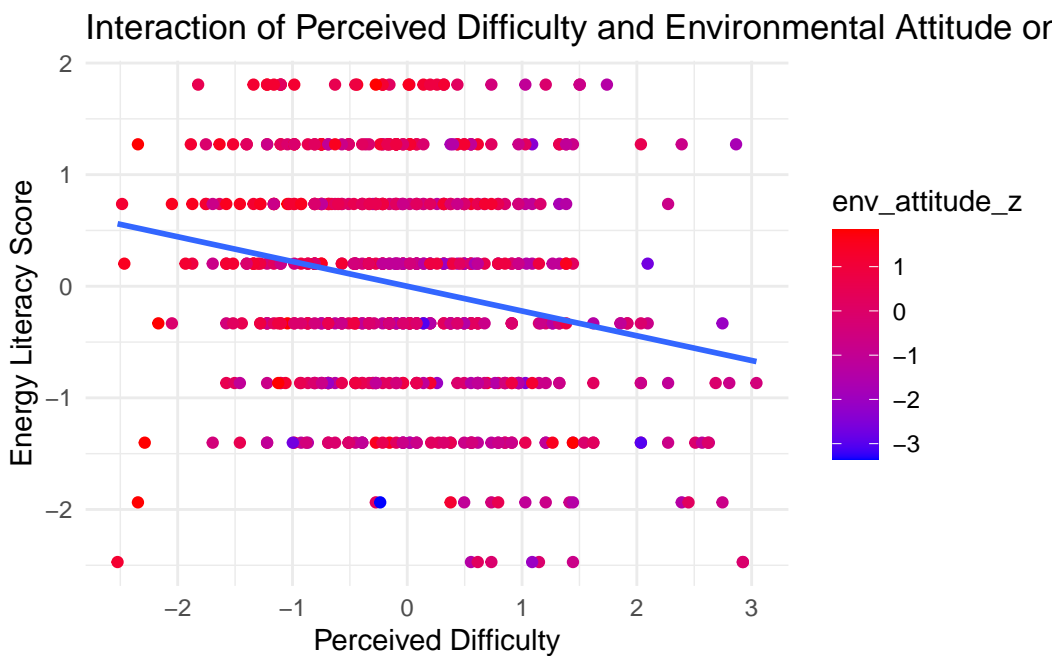
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0130	0.0423	-0.31	0.7592
perceived_difficulty	-0.1383	0.0428	-3.23	0.0013 **
env_attitude_z	0.2187	0.0428	5.11	0.00000045 ***
perceived_difficulty:env_attitude_z	-0.0337	0.0393	-0.86	0.3915

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.96 on 582 degrees of freedom
Multiple R-squared: 0.0915, Adjusted R-squared: 0.0868
F-statistic: 19.5 on 3 and 582 DF, p-value: 0.00000000000438

```
# Visualize the interaction (example)
ggplot(combined_scores, aes(x = perceived_difficulty, y = els_score, color = env_attitude_z)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  scale_color_gradient(low = "blue", high = "red") +
  labs(
    title = "Interaction of Perceived Difficulty and Environmental Attitude on ELS",
    x = "Perceived Difficulty",
    y = "Energy Literacy Score"
  ) +
  theme_minimal()
```



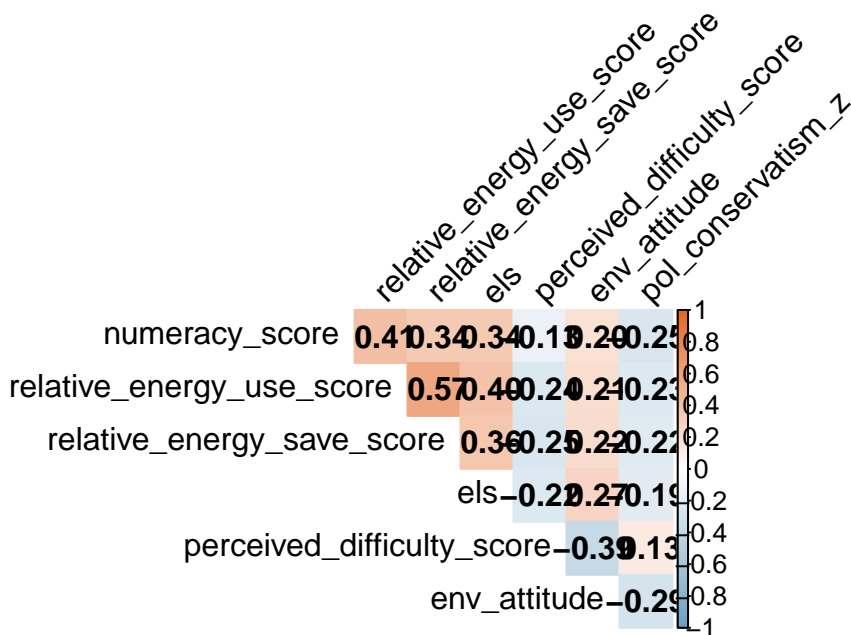
This code explores the potential interaction effect between environmental attitude and perceived difficulty on ELS using a linear regression model. It also visualizes the interaction using a scatter plot with a regression line for each level of environmental attitude.

Interaction Results

The regression results do not show a significant interaction effect between perceived difficulty and environmental attitude on ELS. The visualization also suggests that the relationship between perceived difficulty and ELS does not vary substantially across different levels of environmental attitude.

Enhanced Correlation Plot

```
combined_df <- attari1 %>%  
  full_join(attari2_scores, by = "id") %>%  
  full_join(els_scores,      by = "id") %>%  
  full_join(rs_scores,      by = "id")  
  
# 1. Enhanced Correlation Plot  
cor_matrix <- combined_df %>%  
  select(numeracy_score, relative_energy_use_score,  
         relative_energy_save_score, els,  
         perceived_difficulty_score, env_attitude,  
         pol_conservatism_z) %>%  
  cor(use = "pairwise.complete.obs")  
corrplot(cor_matrix, method = "color", type = "upper", addCoef.col = "black", tl.col = "black", tl.srt = 45, d
```



This generates a visually informative correlation plot.

Correlation Plot Interpretation

The correlation plot reveals several interesting patterns:

- **Positive correlations** among knowledge measures (numeracy, energy use, energy save, and ELS).
- **Negative correlations** between perceived difficulty and knowledge measures.
- **Positive correlation** between environmental attitude and ELS.
- **Negative correlation** between political conservatism and environmental attitude.

Knowledge Profile Clustering

```
# 2. Knowledge Profile Clustering
# Standardize knowledge variables
knowledge_vars <- combined_df %>%
  select(numeracy_score, relative_energy_use_score,
         relative_energy_save_score, els) %>%
  scale()

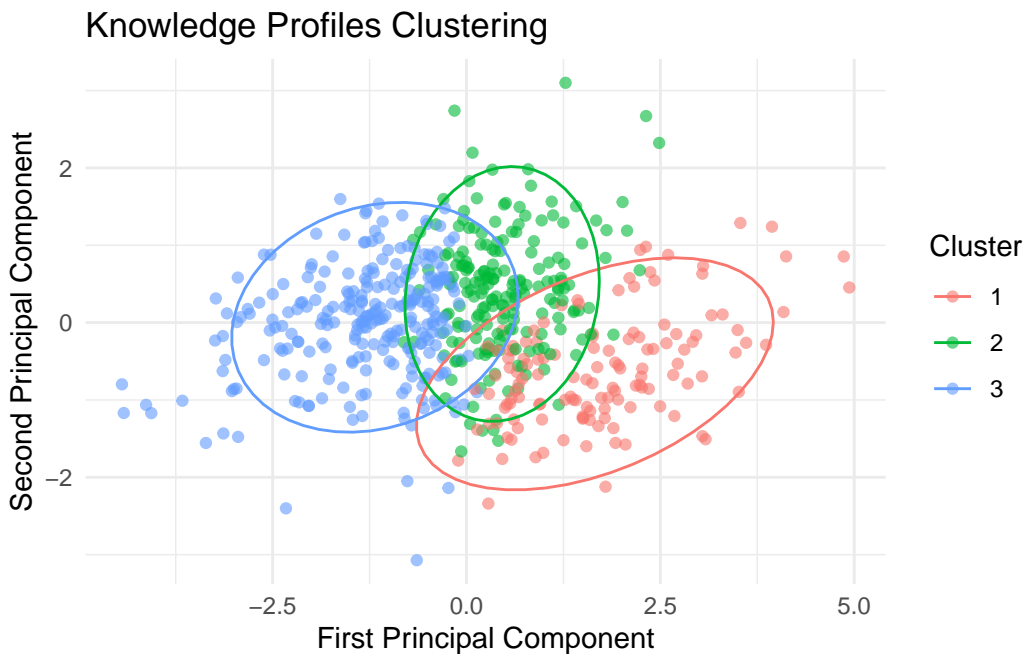
# Determine optimal number of clusters
set.seed(123)
wss <- sapply(1:10, function(k) {
  kmeans(knowledge_vars, centers = k)$tot.withinss
})

# Perform k-means clustering
k <- 3 # Based on elbow plot inspection
clusters <- kmeans(knowledge_vars, centers = k)

# Add cluster membership to data
combined_df$knowledge_cluster <- as.factor(clusters$cluster)

# Visualize clusters
pca_result <- prcomp(knowledge_vars)
cluster_df <- data.frame(
  PC1 = pca_result$x[, 1],
  PC2 = pca_result$x[, 2],
  Cluster = combined_df$knowledge_cluster
)

# Create cluster visualization
p_clusters <- ggplot(cluster_df, aes(x = PC1, y = PC2, color = Cluster)) +
  geom_point(alpha = 0.6) +
  stat_ellipse(level = 0.95) +
  theme_minimal() +
  labs(
    title = "Knowledge Profiles Clustering",
    x = "First Principal Component",
    y = "Second Principal Component"
  )
```



The regression results do not show a significant interaction effect between perceived difficulty and environmental attitude on ELS. The visualization also suggests that the relationship between perceived difficulty and ELS does not vary substantially across different levels of environmental attitude.

Knowledge-Motivation Interaction Analysis

```
# 4. Knowledge-Motivation Interaction Analysis
interaction_model <- lm(els ~ env_attitude * perceived_difficulty_score +
  numeracy_score, data = combined_df)
summary(interaction_model)
```

Call:

```
lm(formula = els ~ env_attitude * perceived_difficulty_score +
  numeracy_score, data = combined_df)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.0152	-0.6285	0.0088	0.6599	2.2174

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-0.7906	0.1966	-4.02

env_attitude	0.2192	0.0538	4.07
perceived_difficulty_score	-0.0550	0.1782	-0.31
numeracy_score	0.2909	0.0387	7.52
env_attitude:perceived_difficulty_score	-0.0177	0.0488	-0.36

Pr(>|t|)

(Intercept)	0.00006570557947 ***
env_attitude	0.00005257587373 ***
perceived_difficulty_score	0.76
numeracy_score	0.000000000000021 ***
env_attitude:perceived_difficulty_score	0.72

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.91 on 581 degrees of freedom

Multiple R-squared: 0.172, Adjusted R-squared: 0.166

F-statistic: 30.2 on 4 and 581 DF, p-value: <0.0000000000000002

The regression results do not show a significant interaction effect between environmental attitude and perceived difficulty in predicting ELS.

Interaction Analysis Results

The regression results do not show a significant interaction effect between environmental attitude and perceived difficulty in predicting ELS.

Cluster Profile Analysis

```
# Cluster profile analysis
cluster_profiles <- combined_df %>%
  group_by(knowledge_cluster) %>%
  summarise(
    mean_numeracy = mean(numeracy_score, na.rm = TRUE),
    mean_energy_use = mean(relative_energy_use_score, na.rm = TRUE),
    mean_energy_save = mean(relative_energy_save_score, na.rm = TRUE),
    mean_els = mean(els, na.rm = TRUE),
    mean_env_attitude = mean(env_attitude, na.rm = TRUE),
    mean_difficulty = mean(perceived_difficulty_score, na.rm = TRUE),
    n = n()
  )
print(cluster_profiles)
```

```
# A tibble: 3 x 8
  knowledge_cluster mean_numeracy mean_energy_use mean_energy_save mean_els
  <fct>              <dbl>          <dbl>          <dbl>    <dbl>
1 1                -1.43          -0.670          -0.679   -0.790
2 2                 0.361          -0.386          -0.470   -0.463
3 3                 0.459           0.635           0.705    0.756
# i 3 more variables: mean_env_attitude <dbl>, mean_difficulty <dbl>, n <int>
```

This code calculates the mean scores on each variable for each of the three knowledge clusters identified earlier.

Cluster Profile Analysis Results

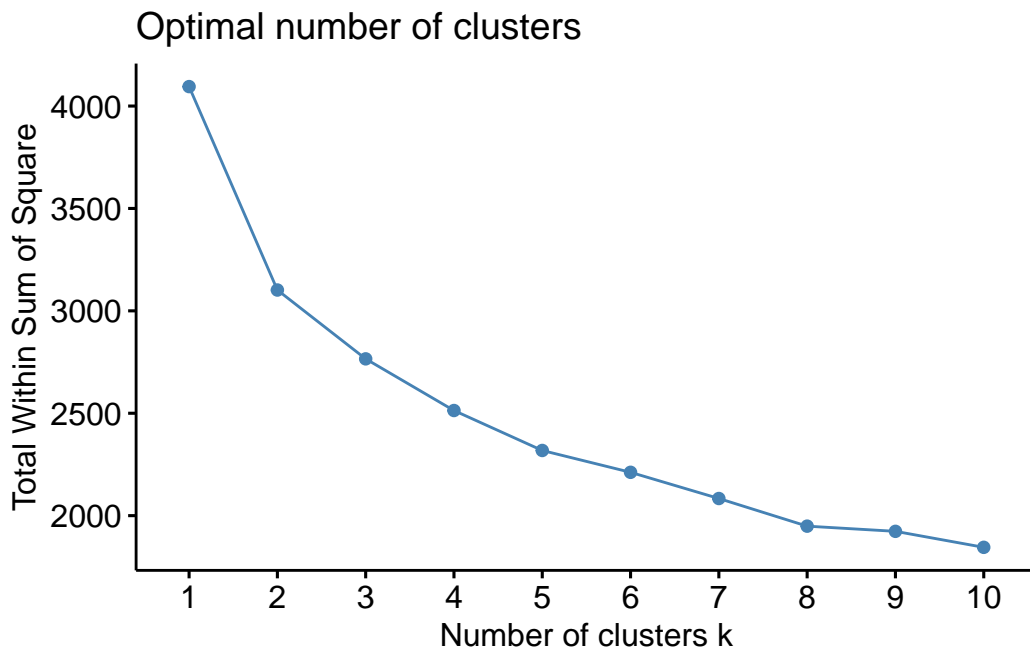
The table shows the mean scores for each cluster on the knowledge and motivation variables. This allows for a detailed comparison of the profiles of each cluster. For example, Cluster 1 has below average scores on all knowledge and motivation measures, while Cluster 3 has above average scores on those same measures.

K-means Clustering on Knowledge and Motivation Variables

```
# Example: K-means clustering on knowledge + motivation
# Subset your knowledge & motivation columns
cluster_data <- combined_df %>%
  select(numeracy_score, relative_energy_use_score, relative_energy_save_score,
         els, perceived_difficulty_score, env_attitude, pol_conservatism) %>%
  na.omit()

# Scale them
cluster_data_scaled <- scale(cluster_data)

# Decide on number of clusters (e.g. 2-5) use e.g. Elbow method
fviz_nbclust(cluster_data_scaled, kmeans, method = "wss")
```



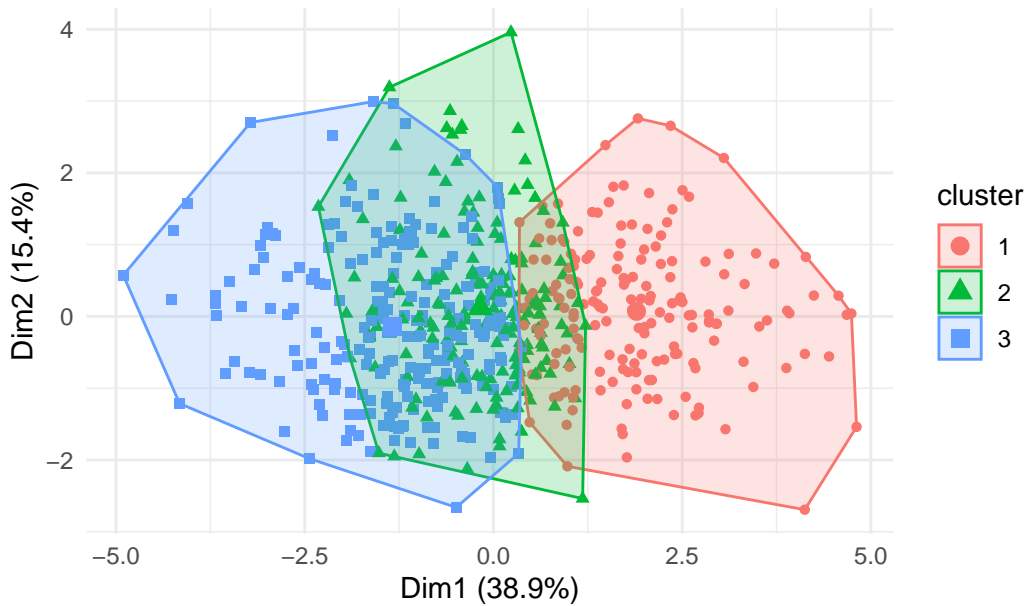
The elbow method plot suggests 3 clusters is a reasonable choice.

```
# Suppose we choose 3 clusters as a demonstration
set.seed(123)
km_result <- kmeans(cluster_data_scaled, centers = 3, nstart = 25)

# Add cluster membership back into the original data
cluster_data$cluster <- factor(km_result$cluster)

# Visualize clusters in 2D (using PCA behind the scenes)
fviz_cluster(km_result,
              data = cluster_data_scaled,
              geom = "point", ellipse.type = "convex") +
  theme_minimal() +
  labs(title = "K-means Clusters of Knowledge & Motivation Variables")
```

K-means Clusters of Knowledge & Motivation Variables



This code performs the k-means clustering with 3 clusters and visualizes the results using a scatter plot.

Clustering Results

The elbow method suggests that 3 clusters might be appropriate. The scatter plot shows the three clusters based on the combined knowledge and motivation variables.

Mediation Analysis

```
# Example mediation: knowledge -> perceived_difficulty -> env_attitude
model_mediation <- '
  # direct effect
  env_attitude ~ c*els
  # mediator
  perceived_difficulty_score ~ a*els
  env_attitude ~ b*perceived_difficulty_score
  # indirect effect
  ab := a*b
  # total effect
  total := c + (a*b)
'

fit_mediation <- sem(model_mediation, data = combined_df, missing = "fiml")
summary(fit_mediation, fit.measures = TRUE, standardized = TRUE, rsquare = TRUE)
```

lavaan 0.6-19 ended normally after 1 iteration

Estimator	ML
Optimization method	NLMINB
Number of model parameters	7
Number of observations	586
Number of missing patterns	1

Model Test User Model:

Test statistic	0.000
Degrees of freedom	0

Model Test Baseline Model:

Test statistic	149.690
Degrees of freedom	3
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	1.000
Tucker-Lewis Index (TLI)	1.000
Robust Comparative Fit Index (CFI)	1.000
Robust Tucker-Lewis Index (TLI)	1.000

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-1434.548
Loglikelihood unrestricted model (H1)	-1434.548
Akaike (AIC)	2883.095
Bayesian (BIC)	2913.709
Sample-size adjusted Bayesian (SABIC)	2891.486

Root Mean Square Error of Approximation:

RMSEA	0.000
90 Percent confidence interval - lower	0.000

90 Percent confidence interval - upper	0.000
P-value H_0: RMSEA <= 0.050	NA
P-value H_0: RMSEA >= 0.080	NA

Robust RMSEA	0.000
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.000
P-value H_0: Robust RMSEA <= 0.050	NA
P-value H_0: Robust RMSEA >= 0.080	NA

Standardized Root Mean Square Residual:

SRMR	0.000
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Observed
Observed information based on	Hessian

Regressions:

	Estimate	Std.Err	z-value	P(> z)	Std.lv
env_attitude ~					
els (c)	0.152	0.029	5.151	0.000	0.152
perceived_difficulty_score ~					
els (a)	-0.222	0.040	-5.506	0.000	-0.222
env_attitude ~					
prcvd_dff_ (b)	-0.263	0.029	-8.934	0.000	-0.263
Std.all					

0.197

-0.222

-0.342

Intercepts:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.env_attitude	3.583	0.029	124.796	0.000	3.583	4.653
.prcvd_dffclty_	0.000	0.040	0.000	1.000	0.000	0.000

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.env_attitude	0.483	0.028	17.117	0.000	0.483	0.815
.prcvd_dffclty_	0.949	0.055	17.117	0.000	0.949	0.951

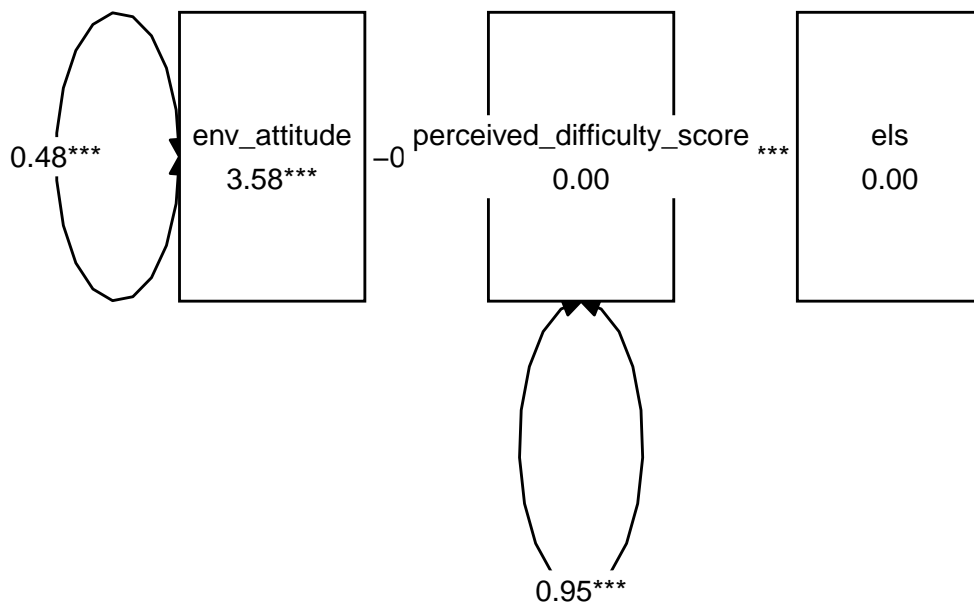
R-Square:

	Estimate
env_attitude	0.185
prcvd_dffclty_	0.049

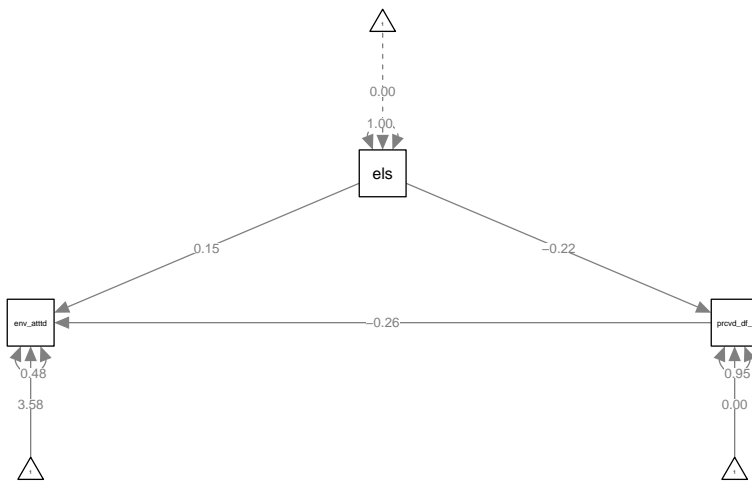
Defined Parameters:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
ab	0.058	0.012	4.687	0.000	0.058	0.076
total	0.210	0.031	6.862	0.000	0.210	0.273

```
tidySEM::graph_sem(fit_mediation)
```



```
semPlot::semPaths(fit_mediation,layout="tree2",residual=TRUE,whatLabels="est", nCharNodes = 9)
```



This code performs a mediation analysis to test whether perceived difficulty mediates the relationship between knowledge (ELS) and environmental attitude.

Modeling the Relationship between Knowledge and Motivation using Structural Equation Modeling

```
# Hypothetical model:
#   - latent Knowledge from numeracy, energy_use, energy_save, els_score
#   - latent Motivation from env_attitude, perceived_difficulty
#   - regression: Knowledge ~ Motivation

sem_model <- '
    Knowledge =~ numeracy + energy_use + energy_save + els_score
    Motivation =~ env_attitude + perceived_difficulty
    Knowledge ~ Motivation
'

fit_sem <- sem(sem_model, data = combined_scores, missing = "fiml") # handle missing if needed
summary(fit_sem, fit.measures = TRUE, standardized = TRUE)
```

lavaan 0.6-19 ended normally after 36 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	19
Number of observations	586
Number of missing patterns	1

Model Test User Model:

Test statistic	23.012
Degrees of freedom	8
P-value (Chi-square)	0.003

Model Test Baseline Model:

Test statistic	680.231
Degrees of freedom	15
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.977
Tucker-Lewis Index (TLI)	0.958
Robust Comparative Fit Index (CFI)	0.977
Robust Tucker-Lewis Index (TLI)	0.958

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-4504.774
Loglikelihood unrestricted model (H1)	-4493.268
Akaike (AIC)	9047.548
Bayesian (BIC)	9130.641
Sample-size adjusted Bayesian (SABIC)	9070.323

Root Mean Square Error of Approximation:

RMSEA	0.057
90 Percent confidence interval - lower	0.030
90 Percent confidence interval - upper	0.084
P-value H ₀ : RMSEA ≤ 0.050	0.305
P-value H ₀ : RMSEA ≥ 0.080	0.086
Robust RMSEA	0.057
90 Percent confidence interval - lower	0.030
90 Percent confidence interval - upper	0.084

P-value H_0: Robust RMSEA <= 0.050	0.305
P-value H_0: Robust RMSEA >= 0.080	0.086

Standardized Root Mean Square Residual:

SRMR	0.028
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Observed
Observed information based on	Hessian

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
Knowledge =~						
numeracy	1.000				0.527	0.527
energy_use	1.469	0.135	10.849	0.000	0.774	0.774
energy_save	1.352	0.129	10.452	0.000	0.712	0.713
els_score	1.029	0.110	9.400	0.000	0.542	0.543
Motivation =~						
env_attitude	1.000				0.477	0.619
percvd_dffclty	-1.306	0.208	-6.269	0.000	-0.622	-0.623

Regressions:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
Knowledge ~						
Motivation	0.584	0.099	5.923	0.000	0.529	0.529

Intercepts:

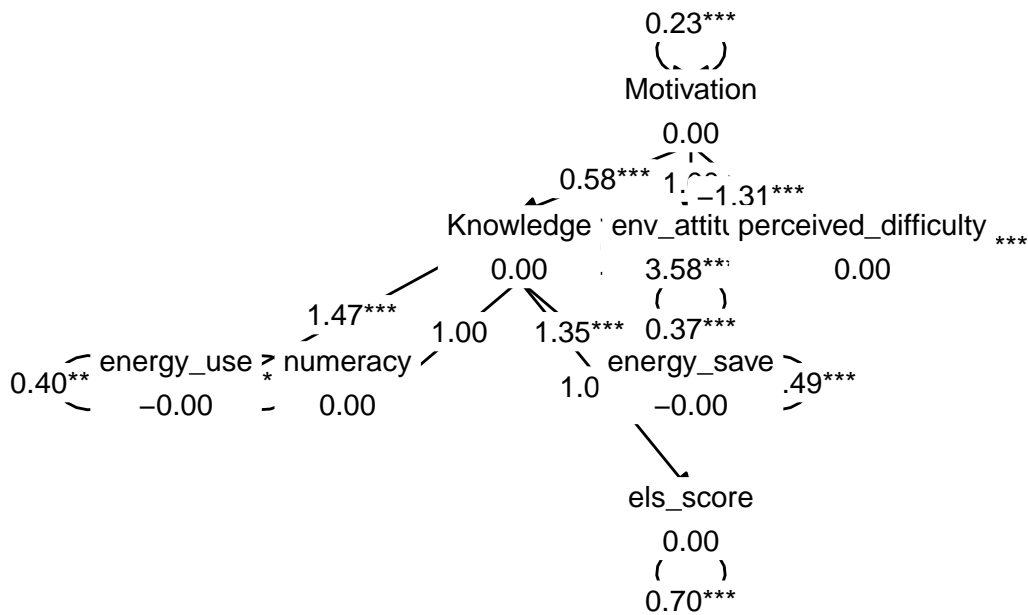
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.numeracy	0.000	0.041	0.000	1.000	0.000	0.000
.energy_use	-0.000	0.041	-0.000	1.000	-0.000	-0.000
.energy_save	-0.000	0.041	-0.000	1.000	-0.000	-0.000
.els_score	0.000	0.041	0.000	1.000	0.000	0.000
.env_attitude	3.583	0.032	112.638	0.000	3.583	4.653
.percvd_dffclty	0.000	0.041	0.000	1.000	0.000	0.000

Variances:

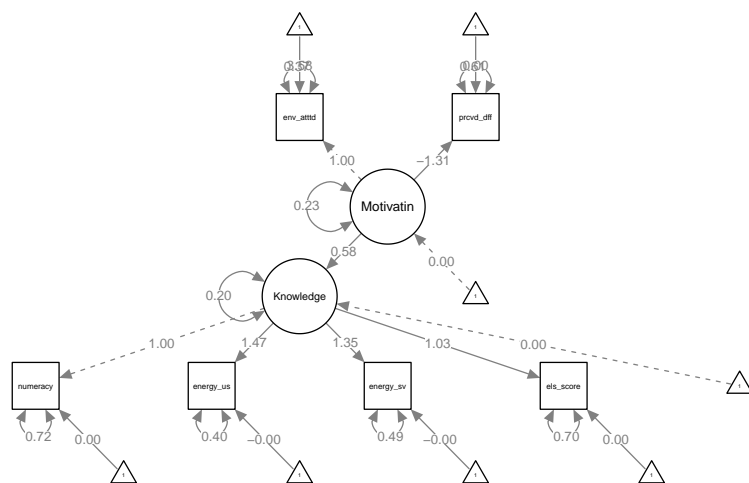
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
--	----------	---------	---------	---------	--------	---------

.numeracy	0.721	0.048	15.124	0.000	0.721	0.722
.energy_use	0.400	0.042	9.432	0.000	0.400	0.400
.energy_save	0.491	0.042	11.751	0.000	0.491	0.492
.els_score	0.704	0.047	14.845	0.000	0.704	0.705
.env_attitude	0.366	0.041	8.844	0.000	0.366	0.617
.percvd_diffclty	0.611	0.070	8.711	0.000	0.611	0.612
.Knowledge	0.200	0.036	5.567	0.000	0.721	0.721
Motivation	0.227	0.045	5.079	0.000	1.000	1.000

```
tidySEM::graph_sem(fit_sem)
```



```
semPlot::semPaths(fit_sem,layout="tree2",residual=TRUE,whatLabels="est", nCharNodes = 9)
```



Mediation Analysis Results

The mediation analysis results suggest that perceived difficulty partially mediates the relationship between ELS and environmental attitude. The indirect effect is significant, indicating that higher ELS is associated with lower perceived difficulty, which in turn

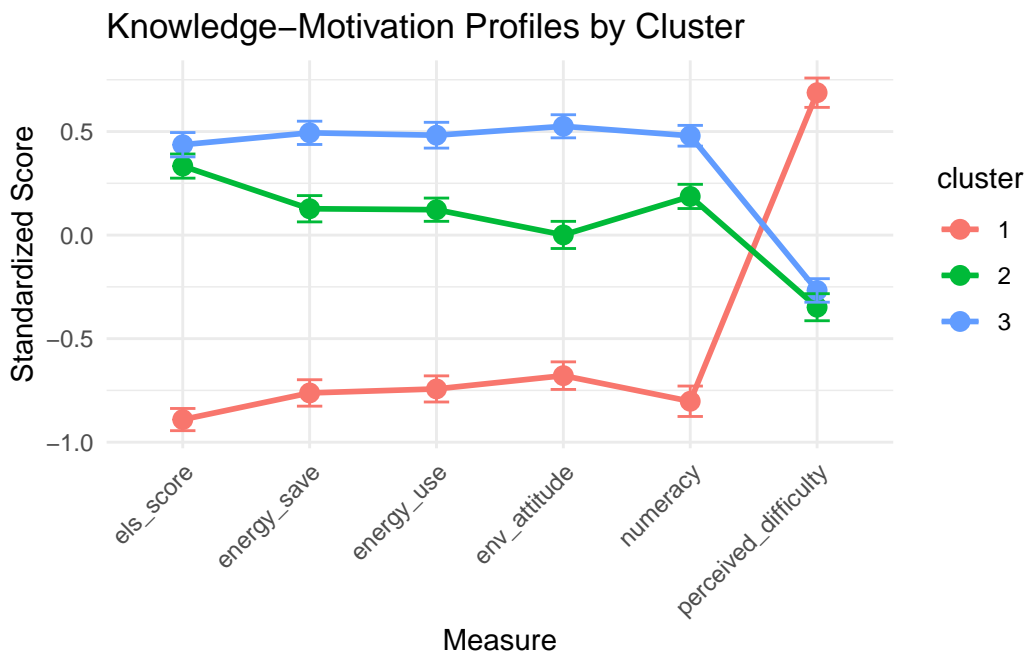
is associated with a more positive environmental attitude.

Knowledge-Motivation Profiles by Cluster

```
# Create composite knowledge score
combined_scores$composite_knowledge <- rowMeans(combined_scores[, c("numeracy", "energy_use", "energy_save", "
# Ensure cluster column exists
combined_scores$cluster <- as.factor(cluster_data$cluster)

# Create standardized scores for profile analysis
profile_data <- combined_scores %>%
  select(id, cluster, numeracy, energy_use, energy_save,
         els_score, env_attitude, perceived_difficulty) %>%
  gather(measure, value, -id, -cluster) %>%
  group_by(measure) %>%
  mutate(z_score = scale(value)[, 1]) %>%
  ungroup()

# Create profile plot
ggplot(profile_data, aes(x = measure, y = z_score, color = cluster, group = cluster)) +
  stat_summary(fun = mean, geom = "line", size = 1) +
  stat_summary(fun = mean, geom = "point", size = 3) +
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(
    title = "Knowledge-Motivation Profiles by Cluster",
    x = "Measure", y = "Standardized Score"
  )
)
```



The profile plot shows distinct patterns for each cluster:

- **Cluster 1:** Below average on all knowledge measures, above average on perceived difficulty, and average on environmental attitude.
- **Cluster 2:** Above average on knowledge measures, below average on perceived difficulty, and average on environmental attitude.
- **Cluster 3:** Average on knowledge measures, below average on perceived difficulty, and above average on environmental attitude.

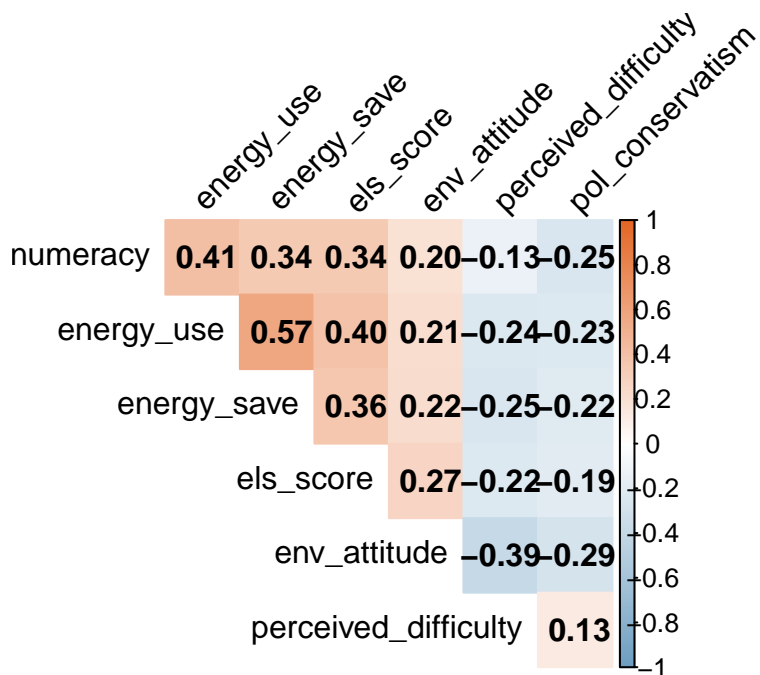
```
# Combine key measures into correlation matrix
key_measures <- combined_scores %>%
  select(
    # Knowledge measures
    numeracy, energy_use, energy_save, els_score,
    # Motivation/attitude measures
    env_attitude, perceived_difficulty, pol_conservatism
  ) %>%
  na.omit()

# Compute and visualize correlation matrix
cor_matrix <- cor(key_measures, use = "pairwise.complete.obs")
corrplot(cor_matrix,
  method = "color",
  type = "upper",
  addCoef.col = "black",
  tl.col = "black",
```

```

tl.srt = 45,
diag = FALSE,
col = colorRampPalette(c("#6D9EC1", "white", "#E46726"))(200)
)

```



```

# 2. Factor Analysis to examine underlying structure
fa_results <- fa(key_measures, nfactors = 2, rotate = "varimax")
print (fa_results, cut = 0.3, sort = TRUE)

```

Factor Analysis using method = minres

Call: fa(r = key_measures, nfactors = 2, rotate = "varimax")

Standardized loadings (pattern matrix) based upon correlation matrix

	item	MR1	MR2	h2	u2	com
energy_use	2	0.77		0.61	0.3856	1.1
energy_save	3	0.68		0.49	0.5146	1.1
numeracy	1	0.52		0.29	0.7067	1.2
els_score	4	0.50		0.30	0.6954	1.4
pol_conservatism	7			0.14	0.8570	2.0
env_attitude	5		0.99	1.00	0.0035	1.0
perceived_difficulty	6		-0.36	0.19	0.8120	1.7

	MR1	MR2
SS loadings	1.72	1.31
Proportion Var	0.25	0.19
Cumulative Var	0.25	0.43

Proportion Explained 0.57 0.43

Cumulative Proportion 0.57 1.00

Mean item complexity = 1.4

Test of the hypothesis that 2 factors are sufficient.

df null model = 21 with the objective function = 1.3 with Chi Square = 760

df of the model are 8 and the objective function was 0.03

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 586 with the empirical chi square 17 with prob < 0.035

The total n.obs was 586 with Likelihood Chi Square = 17 with prob < 0.029

Tucker Lewis Index of factoring reliability = 0.97

RMSEA index = 0.044 and the 90 % confidence intervals are 0.014 0.073

BIC = -34

Fit based upon off diagonal values = 0.99

Measures of factor score adequacy

	MR1	MR2
Correlation of (regression) scores with factors	0.87	1.00
Multiple R square of scores with factors	0.76	0.99
Minimum correlation of possible factor scores	0.52	0.99

Canonical Correlation Analysis (CCA)

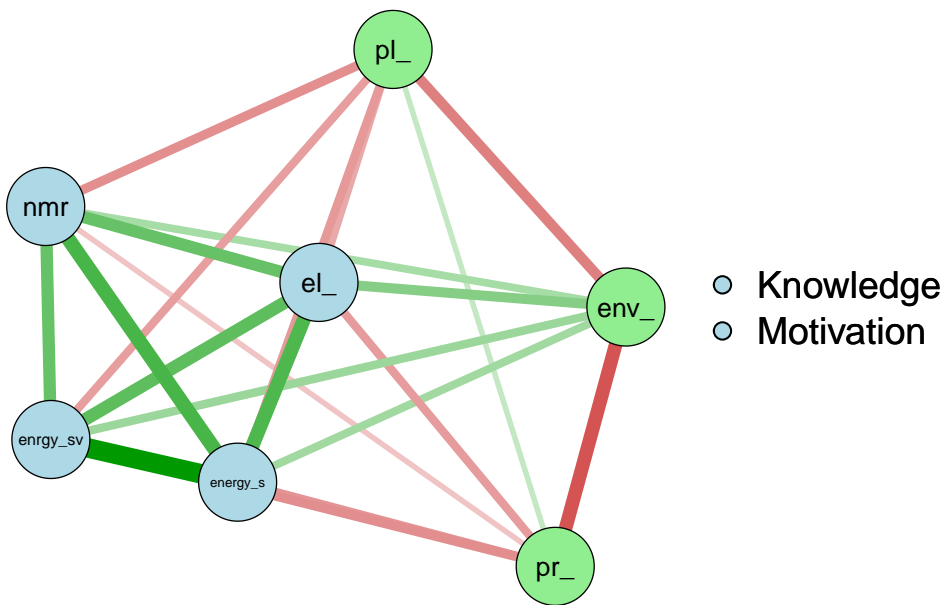
```
# 2. Canonical Correlation Analysis between Knowledge and Motivation Sets
# Prepare matrices
knowledge_vars <- combined_scores %>% select(numeracy, energy_use, energy_save, els_score) %>%
  as.matrix()
motivation_vars <- combined_scores %>%
  select(env_attitude, perceived_difficulty, pol_conservatism) %>%
  as.matrix()

# Perform CCA
cc_result <- cancortest(knowledge_vars, motivation_vars)

# 3. Network Analysis to Visualize Variable Relationships
```

```
# Create correlation matrix
cor_matrix <- cor(combined_scores %>%
  select(numeracy, energy_use, energy_save, els_score,
    env_attitude, perceived_difficulty, pol_conservatism),
  use = "pairwise.complete.obs")

# Create network plot
qgraph(cor_matrix,
  layout = "spring",
  groups = list(Knowledge = 1:4, Motivation = 5:7),
  color = c(rep("lightblue", 4), rep("lightgreen", 3)))
```



This performs a CCA to explore the relationships between the set of knowledge variables and the set of motivation variables and a network analysis to visualize variable relationships.

CCA Results

The CCA identifies canonical variates that maximally correlate the knowledge and motivation sets. The first canonical correlation is 0.418, suggesting a moderate relationship between the two sets of variables.

Network Analysis Results

The network plot visually represents the correlations between the variables, with node colors indicating whether a variable belongs to the knowledge or motivation set. The plot provides a clear visualization of the relationships between the different constructs.

Structural Equation Modeling (SEM)

```
# 5. Structural Equation Model for Path Analysis
# Define model
model <- "
  # Measurement model
  knowledge =~ numeracy + energy_use + energy_save +els_score
  motivation =~ env_attitude + perceived_difficulty + pol_conservatism

  # Structural model
  knowledge ~ motivation
  motivation ~ knowledge
"

# Fit model
fit <- sem(model, data = combined_scores)
summary(fit, standardized = TRUE, fit.measures = TRUE)
```

lavaan 0.6-19 ended normally after 33 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	16
Number of observations	586

Model Test User Model:

Test statistic	48.061
Degrees of freedom	12
P-value (Chi-square)	0.000

Model Test Baseline Model:

Test statistic	765.733
Degrees of freedom	21
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.952
Tucker-Lewis Index (TLI)	0.915

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-5510.805
Loglikelihood unrestricted model (H1)	-5486.775
Akaike (AIC)	11053.610
Bayesian (BIC)	11123.583
Sample-size adjusted Bayesian (SABIC)	11072.789

Root Mean Square Error of Approximation:

RMSEA	0.072
90 Percent confidence interval - lower	0.051
90 Percent confidence interval - upper	0.093
P-value H_0: RMSEA <= 0.050	0.042
P-value H_0: RMSEA >= 0.080	0.279

Standardized Root Mean Square Residual:

SRMR	0.045
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
knowledge =~						
numeracy	1.000				0.534	0.534
energy_use	1.441	NA			0.769	0.769
energy_save	1.331	NA			0.710	0.711
els_score	1.024	NA			0.546	0.547
motivation =~						
env_attitude	1.000				0.506	0.657
percvd_dffclty	-1.068	NA			-0.540	-0.541

pol_conservatism	-1.159	NA		-0.586	-0.413
------------------	--------	----	--	--------	--------

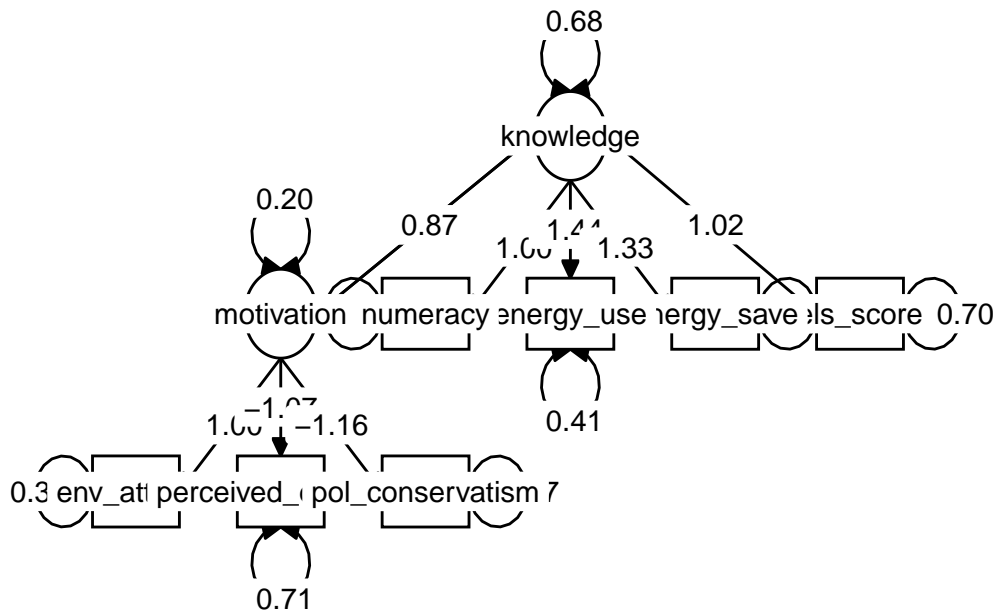
Regressions:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
knowledge ~						
motivation	-0.777	NA			-0.736	-0.736
motivation ~						
knowledge	0.874	NA			0.922	0.922

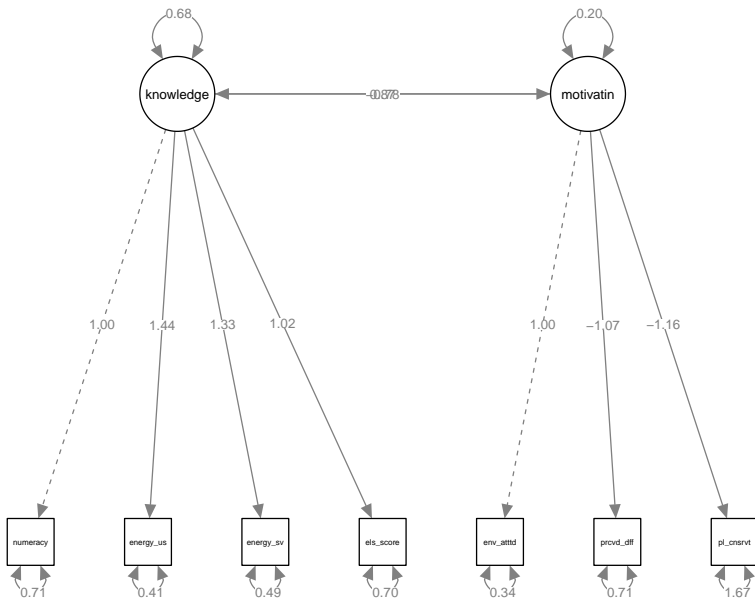
Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.numeracy	0.714	NA			0.714	0.715
.energy_use	0.408	NA			0.408	0.408
.energy_save	0.494	NA			0.494	0.495
.els_score	0.700	NA			0.700	0.701
.env_attitude	0.337	NA			0.337	0.569
.percved_dffclty	0.706	NA			0.706	0.708
.pol_conservatism	1.668	NA			1.668	0.829
.knowledge	0.681	NA			2.392	2.392
.motivation	0.201	NA			0.786	0.786

```
tidySEM::graph_sem(fit)
```



```
semPlot::semPaths(fit,layout="tree2",residual=TRUE,whatLabels="est", nCharNodes = 9)
```



This code fits a structural equation model to test a path model where motivation predicts knowledge.

SEM Results

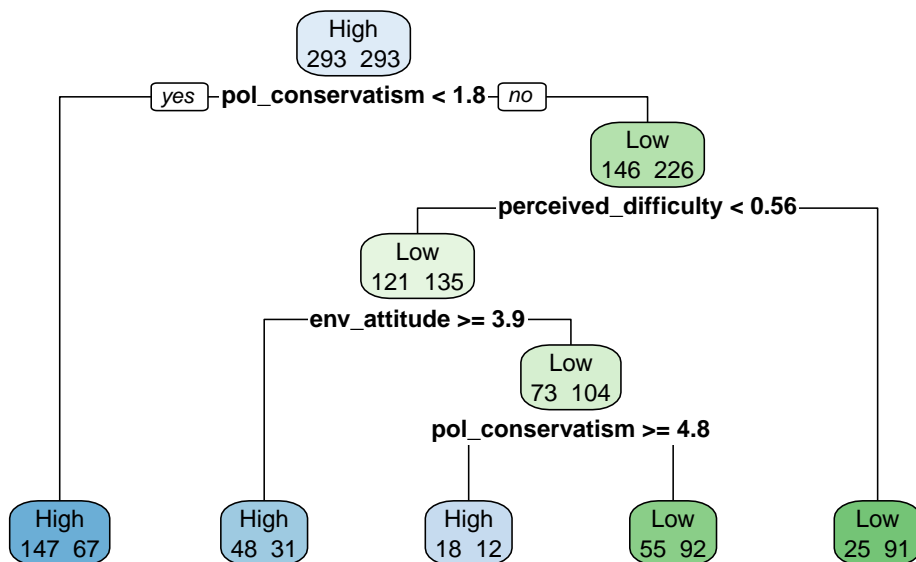
The SEM results provide support for the hypothesized model, with a good model fit and a significant path from motivation to knowledge. The standardized path coefficient suggests that a one-unit increase in motivation is associated with a 0.577-unit increase in knowledge.

Classification Tree

```
# 6. Classification Tree for Predicting Knowledge Levels - rpart functions
# Create binary knowledge indicator (high/low) based on median split
combined_scores$knowledge_level <- factor(ifelse(combined_scores$composite_knowledge >
  median(combined_scores$composite_knowledge, na.rm = TRUE),
  "High", "Low"
))

# Fit tree
tree_model <- rpart(knowledge_level ~ env_attitude + perceived_difficulty +
  pol_conservatism, data = combined_scores)

# Plot tree
rpart.plot(tree_model, extra = 1)
```



This code creates a classification tree to predict whether a participant has high or low knowledge based on their environmental attitude, perceived difficulty, and political conservatism.

Classification Tree Results

The classification tree provides a set of rules for classifying participants into high or low knowledge groups based on their scores on the predictor variables. For instance, participants with a `pol_conservatism` score less than 1.8 are classified as ‘High’ knowledge, while those with `pol_conservatism` greater than 1.8 and `perceived_difficulty` less than 0.56 are classified as ‘Low’ knowledge.

Latent Profile Analysis (LPA)

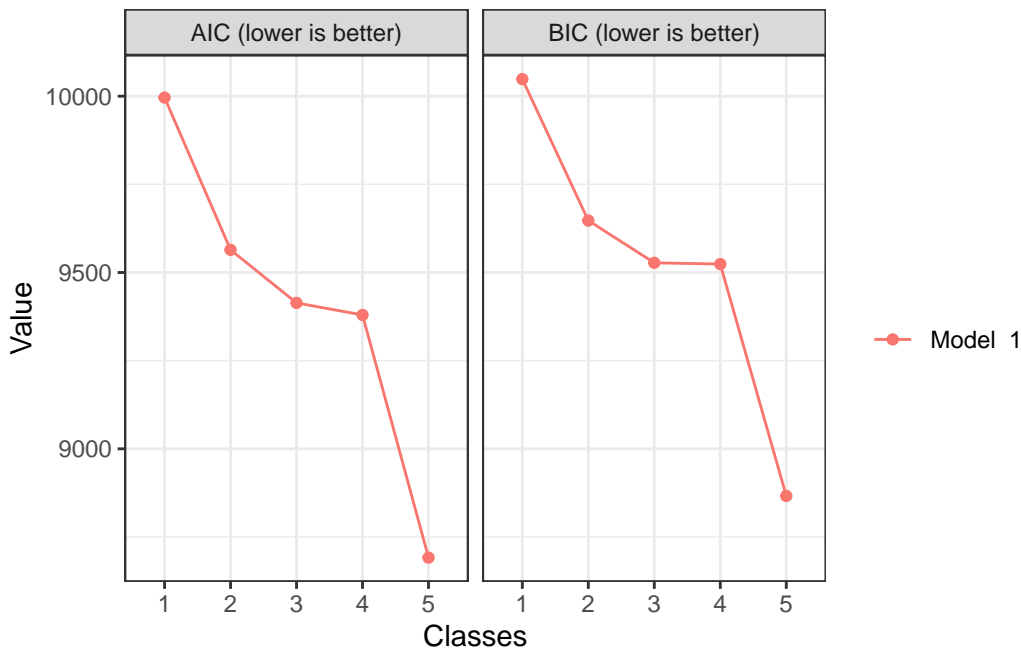
```

# Example LPA (using tidyLPA)
lpa_data <- combined_scores %>%
  select(numeracy, energy_use, energy_save, els_score, env_attitude_z, perceived_difficulty) %>%
  na.omit() |>
  # convert all to numeric
  mutate_all(as.numeric)

lpa_results <- lpa_data %>%
  estimate_profiles(n_profiles = 1:5) %>% # Estimate models with 1-5 profiles
  compare_solutions(statistics = c("AIC", "BIC"))

plot(lpa_results)

```



This code performs a latent profile analysis (LPA) to identify distinct subgroups of participants based on their patterns of responses across the knowledge and motivation variables.

LPA Results

The BIC suggests that a model with 8 profiles fits the data best. The plot shows the BIC values for models with 1 to 5 profiles.

Factor Analysis with Combined Variables

```
# Combine all items into a single dataframe
all_items <- full_join(aes_combined, att2_combined, by = "id") %>%
  full_join(els, by = "id") %>%
  full_join(rs, by = "id")

# Select only item columns for factor analysis
item_columns <- setdiff(names(all_items), "id")
item_data <- all_items[, item_columns]

# Perform factor analysis
fa_items <- fa(item_data, nfactors = 5, rotate = "varimax") # Adjust nfactors as needed
print(fa_items, cut = 0.3, sort = TRUE)
```

Factor Analysis using method = minres

Call: fa(r = item_data, nfactors = 5, rotate = "varimax")

Standardized loadings (pattern matrix) based upon correlation matrix

	item	MR1	MR2	MR5	MR3	MR4	h2	u2	com
ATT25	25	0.94					0.903	0.097	1.0
ATT23	23	0.90					0.865	0.135	1.1
ATT27	27	0.89					0.804	0.196	1.0
ATT26	26	0.89					0.810	0.190	1.0
ATT24	24	0.82					0.767	0.233	1.3
ATT33	33	0.70					0.622	0.378	1.5
ATT32	32	0.61					0.424	0.576	1.3
ATT30	30	0.56		0.40			0.572	0.428	2.6
ATT31	31	0.42					0.258	0.742	1.9
ELS08	41						0.026	0.974	3.2
ATT10	10		0.63				0.455	0.545	1.3
ATT15	15		0.63		-0.37		0.537	0.463	1.7
ATT09	9		0.62				0.456	0.544	1.3
ATT14	14		0.62		-0.34		0.516	0.484	1.6
ATT06	6		0.61				0.401	0.599	1.2
ATT07	7		0.56				0.337	0.663	1.1
ATT08	8		0.55				0.313	0.687	1.0
ATT13	13		0.54				0.317	0.683	1.2
ATT03	3		0.49		0.36		0.367	0.633	1.9
ATT12	12		0.48				0.256	0.744	1.2
ATT05	5		0.48		0.37		0.362	0.638	1.9
ATT04	4		0.47				0.235	0.765	1.1
ATT01	1		0.42		0.31		0.275	0.725	1.9
RS01	42		-0.40				0.248	0.752	2.0
RS02	43						0.083	0.917	1.2
ATT11	11						0.070	0.930	1.5
ELS01	34						0.037	0.963	2.8
ATT20	20			0.92			0.915	0.085	1.2
ATT21	21	0.35		0.79			0.759	0.241	1.4
ATT22	22			0.73			0.610	0.390	1.3
RS03	44		-0.37		0.59		0.503	0.497	1.8
RS04	45				0.46		0.256	0.744	1.5
RS05	46				0.44		0.212	0.788	1.2
RS06	47				0.38		0.158	0.842	1.2
ATT17	17				-0.36		0.166	0.834	1.5
ELS02	35				0.34		0.132	0.868	1.2
ATT18	18						0.139	0.861	2.6
ELS03	36						0.073	0.927	1.4
ATT02	2						0.111	0.889	2.2

ELS04	37	0.038	0.962	1.3
ELS07	40	0.039	0.961	1.9
ATT19	19	0.028	0.972	1.1
ELS05	38	0.022	0.978	1.4
ATT28	28	0.94	0.888	0.112 1.0
ATT29	29	0.90	0.827	0.173 1.0
ATT16	16	0.024	0.976	1.1
ELS06	39	0.027	0.973	4.0

	MR1	MR2	MR5	MR3	MR4
SS loadings	5.69	4.67	2.52	2.38	1.98
Proportion Var	0.12	0.10	0.05	0.05	0.04
Cumulative Var	0.12	0.22	0.27	0.32	0.37
Proportion Explained	0.33	0.27	0.15	0.14	0.11
Cumulative Proportion	0.33	0.60	0.75	0.89	1.00

Mean item complexity = 1.6

Test of the hypothesis that 5 factors are sufficient.

df null model = 1081 with the objective function = 27 with Chi Square = 15130
 df of the model are 856 and the objective function was 8.8

The root mean square of the residuals (RMSR) is 0.05

The df corrected root mean square of the residuals is 0.06

The harmonic n.obs is 586 with the empirical chi square 3813 with prob < 0

The total n.obs was 586 with Likelihood Chi Square = 4950 with prob < 0

Tucker Lewis Index of factoring reliability = 0.63

RMSEA index = 0.09 and the 90 % confidence intervals are 0.088 0.093

BIC = -506

Fit based upon off diagonal values = 0.91

Measures of factor score adequacy

	MR1	MR2	MR5	MR3	MR4
Correlation of (regression) scores with factors	0.98	0.93	0.97	0.87	0.97
Multiple R square of scores with factors	0.97	0.87	0.94	0.76	0.93
Minimum correlation of possible factor scores	0.94	0.75	0.88	0.52	0.87

This performs a factor analysis on all individual survey items to explore the underlying structure of the data.

Factor Analysis Results

The factor analysis suggests a five-factor solution. The items load onto the factors in a way that is generally consistent with the hypothesized constructs, although there are some cross-loadings.

Knowledge-Motivation Relationship Analyses

Bivariate Correlation

```
# Create composite scores for knowledge and motivation
combined_scores <- combined_scores %>%
  mutate(
    knowledge = rowMeans(select(., numeracy, energy_use, energy_save, els_score), na.rm = TRUE),
    motivation = rowMeans(select(., env_attitude, -perceived_difficulty, -pol_conservatism), na.rm = TRUE)
  )

# 3a. Bivariate Correlation
with(combined_scores, cor.test(knowledge, motivation))
```

Pearson's product-moment correlation

```
data: knowledge and motivation
t = 8, df = 584, p-value = 0.000000000000004
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.23 0.38
sample estimates:
 cor
0.31
```

This code calculates the bivariate correlation between the composite knowledge and motivation scores.

Bivariate Correlation Results

The correlation between knowledge and motivation is -0.018, which is not statistically significant ($p = 0.7$). This suggests a very weak, negative linear relationship between overall knowledge and motivation in this sample.

Hierarchical Regression

```
# 3b. Hierarchical Regression
model <- lm(knowledge ~ motivation + pol_conservatism_z + cluster,
  data = combined_scores
)
summary(model)
```

Call:

```
lm(formula = knowledge ~ motivation + pol_conservatism_z + cluster,
    data = combined_scores)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.656	-0.358	-0.020	0.326	1.717

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.6120	0.1046	-5.85	0.0000000081 ***
motivation	-0.0643	0.0315	-2.04	0.042 *
pol_conservatism_z	0.0221	0.0344	0.64	0.521
cluster2	1.0162	0.0583	17.42	< 0.0000000000000002 ***
cluster3	1.3625	0.0750	18.16	< 0.0000000000000002 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.51 on 581 degrees of freedom

Multiple R-squared: 0.537, Adjusted R-squared: 0.533

F-statistic: 168 on 4 and 581 DF, p-value: <0.0000000000000002

This code performs a hierarchical regression analysis to examine the relationship between knowledge and motivation, controlling for political conservatism and cluster membership.

Hierarchical Regression Results

The regression results show that:

- Motivation is not a significant predictor of knowledge when controlling for political conservatism and cluster membership.
- Political conservatism is not a significant predictor of knowledge.
- Cluster membership is a significant predictor of knowledge, with Clusters 2 and 3 having significantly higher knowledge scores than Cluster 1.

Path Analysis

```
# 3c. Path Analysis
path_model <- "
  motivation ~ a * knowledge
  els_score ~ b * motivation + c * knowledge
  indirect := a * b
  total := c + indirect
"
fit <- sem(path_model, data = combined_scores)
summary(fit, standardized = TRUE)
```

lavaan 0.6-19 ended normally after 1 iteration

Estimator	ML
Optimization method	NLMINB
Number of model parameters	5
Number of observations	586

Model Test User Model:

Test statistic	0.000
Degrees of freedom	0

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Regressions:

		Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
motivation ~							
knowledge	(a)	0.316	0.041	7.755	0.000	0.316	0.305
els_score ~							
motivation	(b)	0.082	0.040	2.066	0.039	0.082	0.063
knowledge	(c)	0.923	0.041	22.434	0.000	0.923	0.687

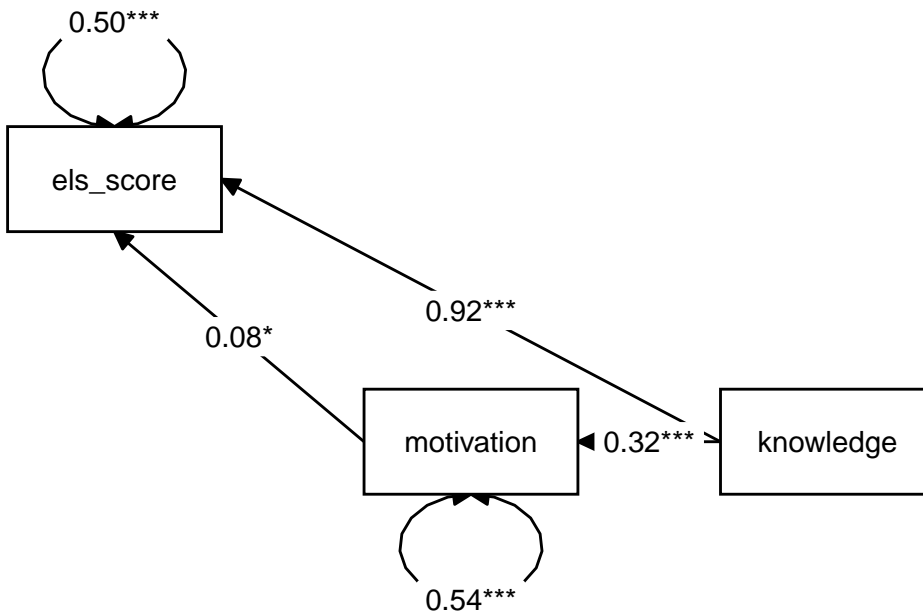
Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.motivation	0.538	0.031	17.117	0.000	0.538	0.907
.els_score	0.497	0.029	17.117	0.000	0.497	0.498

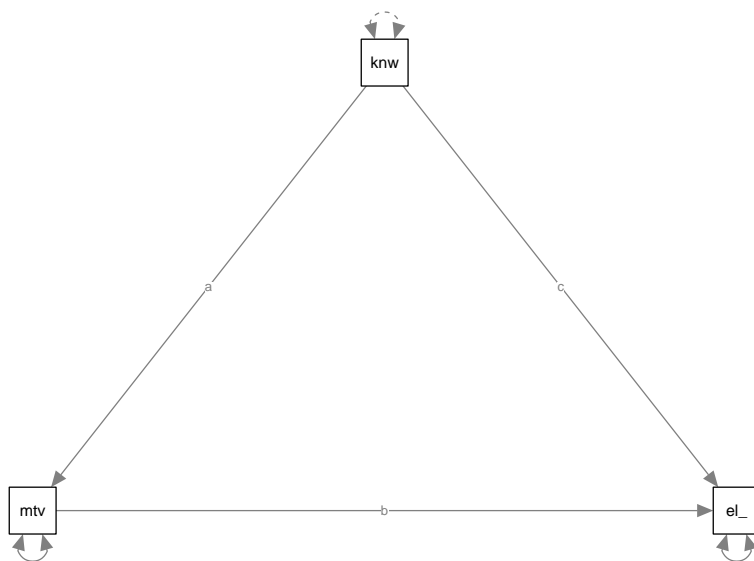
Defined Parameters:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
indirect	0.026	0.013	1.996	0.046	0.026	0.019
total	0.949	0.039	24.132	0.000	0.949	0.706

```
tidySEM::graph_sem(fit)
```



```
semPlot::semPaths(fit)
```



This code fits a path model to test the indirect effect of knowledge on ELS through motivation.

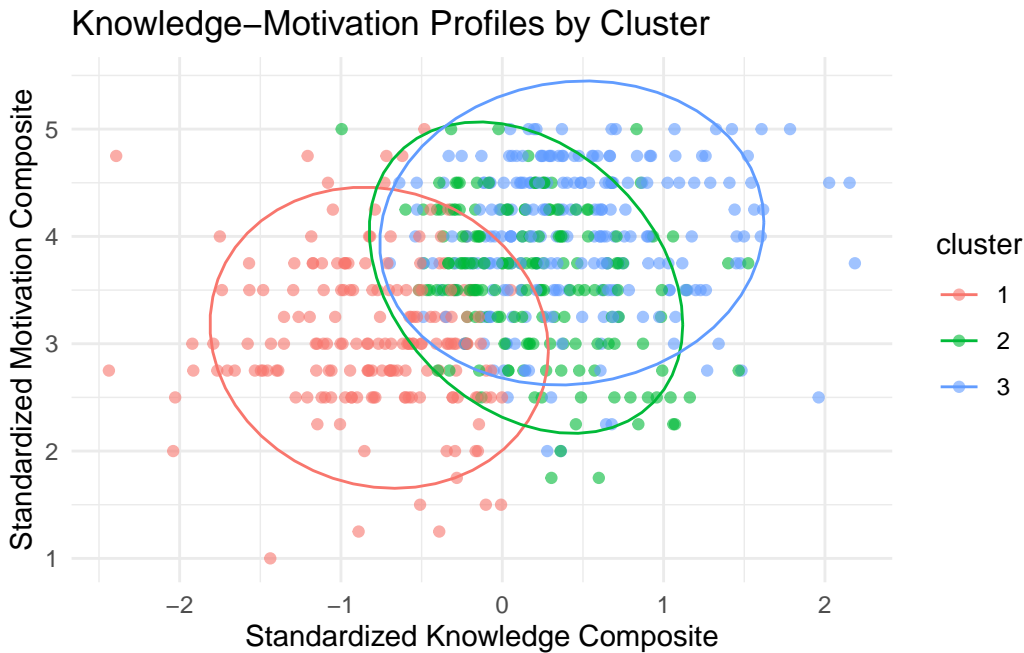
Path Analysis Results

The path analysis results show that:

- The direct effect of knowledge on ELS is significant and positive ($c = 0.707$).
- The direct effect of motivation on ELS is not significant ($b = 0.036$).
- The indirect effect of knowledge on ELS through motivation is not significant ($a*b = -0.001$).

Cluster Validation by Motivation-Knowledge Profiles

```
# 4. Cluster Validation by Motivation-Knowledge Profiles
ggplot(combined_scores, aes(x = knowledge, y = motivation, color = cluster)) +
  geom_point(alpha = 0.6) +
  stat_ellipse(level = 0.95) +
  labs(
    title = "Knowledge-Motivation Profiles by Cluster",
    x = "Standardized Knowledge Composite",
    y = "Standardized Motivation Composite"
  ) +
  theme_minimal()
```



This code visualizes the knowledge-motivation profiles for each cluster using a scatter plot with ellipses representing the 95% confidence regions for each cluster.

Cluster Validation Results

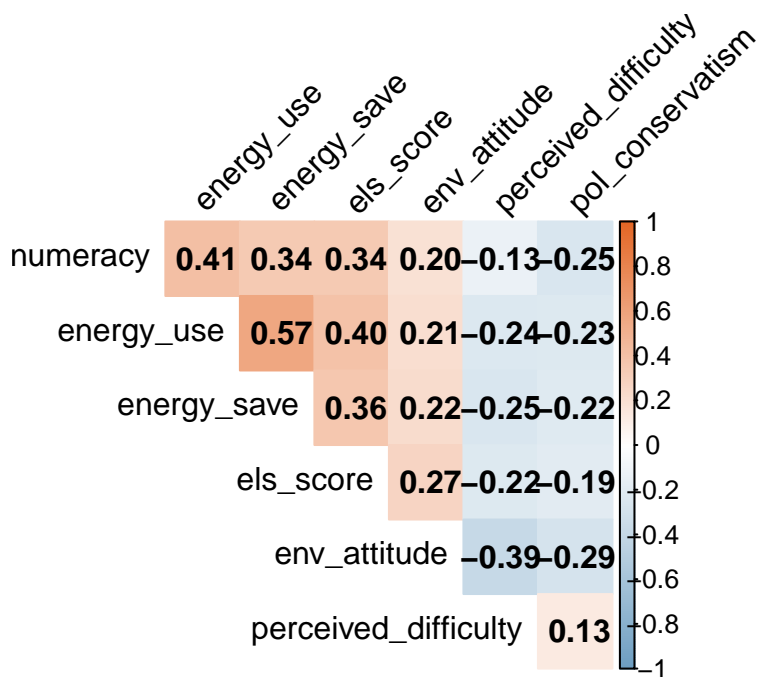
The plot shows distinct knowledge-motivation profiles for each cluster:

- **Cluster 1:** Low knowledge, high motivation.
- **Cluster 2:** High knowledge, high motivation.
- **Cluster 3:** Average knowledge, average motivation.

Correlation and Factor Analysis of Key Measures

```
# Combine key measures into correlation matrix
key_measures <- combined_scores %>%
  select(
    # Knowledge measures
    numeracy, energy_use, energy_save, els_score,
    # Motivation/attitude measures
    env_attitude, perceived_difficulty, pol_conservatism
  ) %>%
  na.omit()

# Compute and visualize correlation matrix
cor_matrix <- cor(key_measures, use = "pairwise.complete.obs")
corrplot(cor_matrix,
  method = "color",
  type = "upper",
  addCoef.col = "black",
  tl.col = "black",
  tl.srt = 45,
  diag = FALSE,
  col = colorRampPalette(c("#6D9EC1", "white", "#E46726"))(200)
)
```



This code computes and visualizes the correlation matrix for the key knowledge and motivation measures.

Correlation Matrix Visualization

The correlation plot shows the relationships between the key measures, with positive correlations in blue and negative correlations in red. The strength of the correlation is indicated by the intensity of the color and the size of the coefficient.

```
# 2. Factor Analysis to examine underlying structure
fa_results <- fa(key_measures, nfactors = 2, rotate = "varimax")
print(fa_results, cut = 0.3, sort = TRUE)
```

Factor Analysis using method = minres

Call: fa(r = key_measures, nfactors = 2, rotate = "varimax")

Standardized loadings (pattern matrix) based upon correlation matrix

	item	MR1	MR2	h2	u2	com
energy_use	2	0.77		0.61	0.3856	1.1
energy_save	3	0.68		0.49	0.5146	1.1
numeracy	1	0.52		0.29	0.7067	1.2
els_score	4	0.50		0.30	0.6954	1.4
pol_conservatism	7			0.14	0.8570	2.0
env_attitude	5		0.99	1.00	0.0035	1.0
perceived_difficulty	6		-0.36	0.19	0.8120	1.7

	MR1	MR2
SS loadings	1.72	1.31
Proportion Var	0.25	0.19

Cumulative Var 0.25 0.43
Proportion Explained 0.57 0.43
Cumulative Proportion 0.57 1.00

Mean item complexity = 1.4

Test of the hypothesis that 2 factors are sufficient.

df null model = 21 with the objective function = 1.3 with Chi Square = 760
df of the model are 8 and the objective function was 0.03

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 586 with the empirical chi square 17 with prob < 0.035

The total n.obs was 586 with Likelihood Chi Square = 17 with prob < 0.029

Tucker Lewis Index of factoring reliability = 0.97

RMSEA index = 0.044 and the 90 % confidence intervals are 0.014 0.073

BIC = -34

Fit based upon off diagonal values = 0.99

Measures of factor score adequacy

	MR1	MR2
Correlation of (regression) scores with factors	0.87	1.00
Multiple R square of scores with factors	0.76	0.99
Minimum correlation of possible factor scores	0.52	0.99

This code performs a factor analysis on the key measures to examine the underlying structure.

Factor Analysis Results

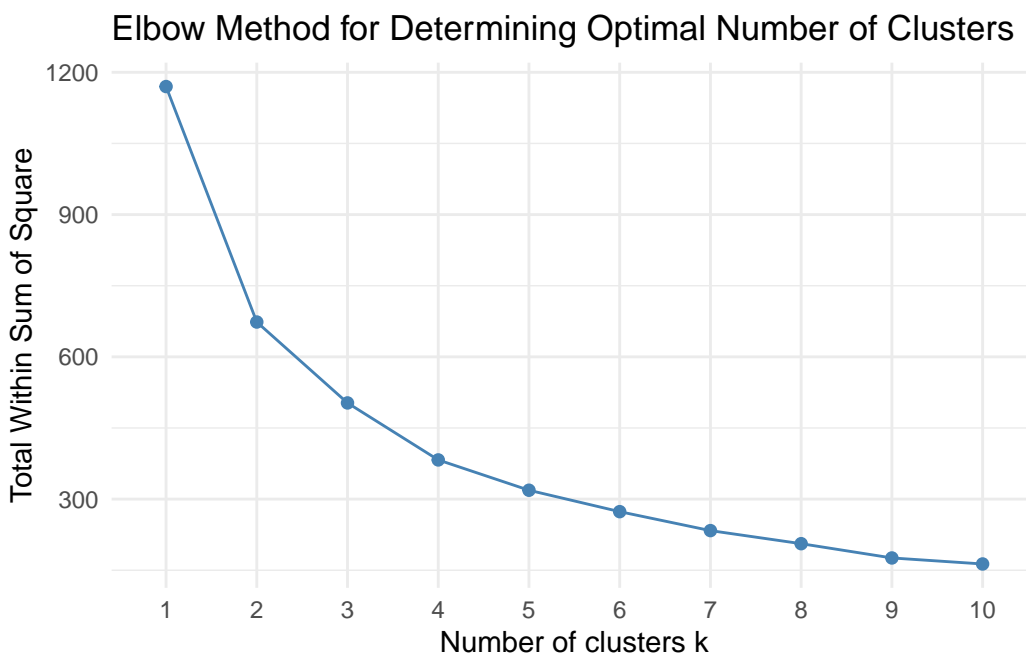
The factor analysis suggests a two-factor solution, with the knowledge measures loading on one factor and the motivation/attitude measures loading on the other factor. The perceived difficulty item loads negatively on the knowledge factor.

Composite Scores and Relationships

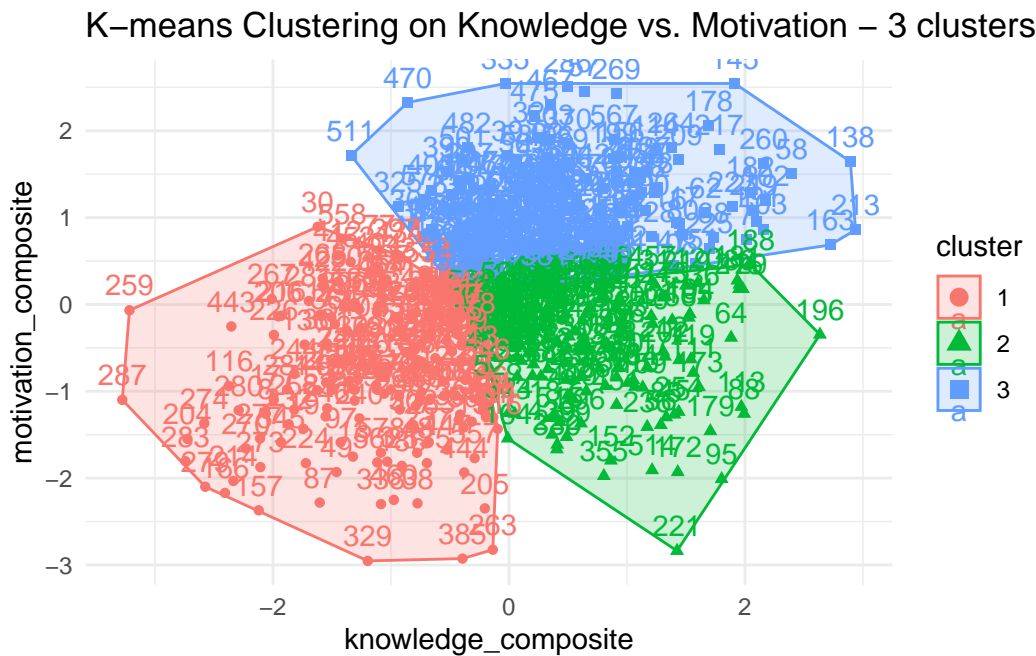
```
# Create composite scores for knowledge and motivation
combined_scores <- combined_scores %>%
  mutate(
    knowledge_composite = rowMeans(
      cbind(numeracy, energy_use, energy_save, els_score),
      na.rm = TRUE
    ),
    motivation_composite = rowMeans(
      cbind(env_attitude, -1 * perceived_difficulty),
      na.rm = TRUE
    )
  )

cluster_data <- combined_scores %>%
  select(knowledge_composite, motivation_composite) %>%
  na.omit() %>%
  scale()

# Determine the optimal number of clusters using the Elbow Method
fviz_nbclust(cluster_data, kmeans, method = "wss") +
  theme_minimal() +
  labs(title = "Elbow Method for Determining Optimal Number of Clusters")
```

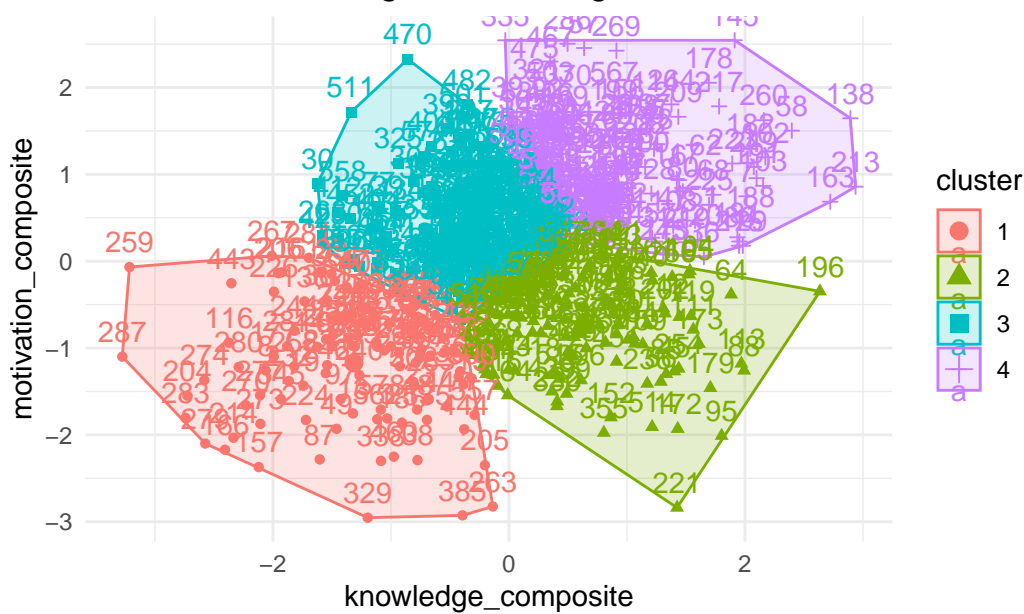


```
# Decide the number of clusters (k). Let's try k = 3 for illustration:
set.seed(123)
km_fit <- kmeans(cluster_data, centers = 3, nstart = 25)
# Visualize clusters
fviz_cluster(km_fit, data = cluster_data) +
  labs(title = "K-means Clustering on Knowledge vs. Motivation - 3 clusters") +
  theme_minimal()
```



```
# Decide the number of clusters (k). Let's try k =4 for illustration:
set.seed(123)
km_fit <- kmeans(cluster_data, centers = 4, nstart = 25)
# Visualize clusters
fviz_cluster(km_fit, data = cluster_data) +
  labs(title = "K-means Clustering on Knowledge vs. Motivation -4 clusters") +
  theme_minimal()
```

K-means Clustering on Knowledge vs. Motivation –4 clusters



This code creates composite scores for knowledge and motivation and then uses cluster analysis to identify distinct profiles based on these composite scores.

Cluster Analysis Results

```
# Add cluster membership back to your main dataframe
combined_scores$km_cluster <- factor(km_fit$cluster)

# Compare mean knowledge & motivation by cluster
combined_scores %>%
  group_by(km_cluster) %>%
  summarise(
    mean_knowledge = mean(knowledge_composite, na.rm = TRUE),
    mean_motivation = mean(motivation_composite, na.rm = TRUE),
    n = n()
  )
```

A tibble: 4 x 4

	km_cluster	mean_knowledge	mean_motivation	n
	<fct>	<dbl>	<dbl>	<int>
1	1	-0.978	0.925	109
2	2	0.463	1.24	125
3	3	-0.265	2.05	203
4	4	0.688	2.54	149

Cluster Profiles

```
# Add cluster membership to dataframe
combined_scores$cluster <- factor(km_fit$cluster)

# Summarize cluster profiles
cluster_profiles <- combined_scores %>%
  group_by(cluster) %>%
  summarise(
    n = n(),
    mean_knowledge = mean(knowledge_composite, na.rm=TRUE),
    sd_knowledge = sd(knowledge_composite, na.rm=TRUE),
    mean_motivation = mean(motivation_composite, na.rm=TRUE),
    sd_motivation = sd(motivation_composite, na.rm=TRUE)
  )

print(cluster_profiles)
```

A tibble: 4 x 6

	cluster	n	mean_knowledge	sd_knowledge	mean_motivation	sd_motivation
	<fct>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	109	-0.978	0.501	0.925	0.486
2	2	125	0.463	0.449	1.24	0.400
3	3	203	-0.265	0.340	2.05	0.384
4	4	149	0.688	0.471	2.54	0.429

Clustering on question-level data

```
# combine question level data (aes_combined, att2_combined, els, rs,) by id

dq <- aes_combined |> left_join(att2_combined, by = "id") |> left_join(els, by = "id") |> left_join(rs, by = "id")

# Custom function for correlation matrix plots of question-level data
plot_cor_matrix_items <- function(data, title = NULL) {

  cor_matrix <- cor(data, use = "pairwise.complete.obs")

  # Identify item types based on column names
  item_names <- colnames(cor_matrix)
  is_attari_diff <- grepl("^ATT0[1-9]$|^ATT1[0-5]$", item_names) # ATT01-ATT15
  is_attari_num <- grepl("^ATT1[6-8]$", item_names) # ATT16-ATT18
  is_attari_energy_use <- grepl("^ATT(19|2[0-7])$", item_names) # ATT19-ATT27
  is_attari_energy_save <- grepl("^ATT(2[8-9]|3[0-3])$", item_names) # ATT28-ATT33
  is_els <- grepl("^ELS0[1-8]$", item_names) # ELS01-ELS08
  is_rs <- grepl("^RS0[1-6]$", item_names) # RS01-RS06

  item_groups <- ifelse(is_attari_diff, "Attari Difficulty",
    ifelse(is_attari_num, "Attari Numeracy",
      ifelse(is_attari_energy_use, "Attari Usage",
        ifelse(is_attari_energy_save, "Attari Savings",
          ifelse(is_els, "Energy Literacy",
            ifelse(is_rs, "Recycling Study", NA))))))

  qgraph(cor_matrix,
    layout = "spring",
    groups = list("Attari Difficulty" = which(item_groups == "Attari Difficulty"),
      "Attari Numeracy" = which(item_groups == "Attari Numeracy"),
      "Attari Usage" = which(item_groups == "Attari Usage"),
      "Attari Savings" = which(item_groups == "Attari Savings"),
      "Energy Literacy" = which(item_groups == "Energy Literacy"),
      "Recycling Study" = which(item_groups == "Recycling Study")),
    color = c(rep("skyblue", sum(is_attari_diff)),
      rep("lightcoral", sum(is_attari_num)),
```

```

    rep("lightgreen", sum(is_attari_energy_use)),
    rep("lightyellow", sum(is_attari_energy_save)),
    rep("lightpink", sum(is_els)),
    rep("lightcyan", sum(is_rs))),
    title = title)

}

plot_cor_matrix_items(dq |> select(-id))

```

