Simple Multi-environment data analysis

Contents

Install and load packages 1

Data Preparation 1

Data Overview 2

Dataset Structure . . . 2

Missing Values . . . 3

Statistical Analysis 3

Perform ANOVA for each environment (RCBD Design) . . . 3

Joint ANOVA for Multi-Environment Trial (RCBD Design) . . . 8

LSD (Least Significant Difference) Test . . . 9

Variance Components and Genetic Parameters from Multi-Environment Trial 11

Correlations between ENVIRONMENTS 13

Phenotypic (Based on observed averages) . . . 13

Genetic (Based on adjusted genotypic effects) . . . 14

Environmental (Based on non-genetic effects) . . . 15

Environmental effects for each location 16

Install and load packages

```
suppressMessages({
  if (!require("readxl")) install.packages("readxl")
  if (!require("tidyverse")) install.packages("tidyverse")
  if (!require("ggpubr")) install.packages("ggpubr")
  if (!require("knitr")) install.packages("knitr")
  if (!require("kableExtra")) install.packages("kableExtra")
  if (!require("agricolae")) install.packages("agricolae")
  if (!require("broom")) install.packages("broom")
  if (!require("ggplot2")) install.packages("ggplot2")
  if (!require("metan")) install.packages("metan")
```

```
  if (!require("ggcorrplot")) install.packages("ggcorrplot")
  if (!require("corrplot")) install.packages("corrplot")
  if (!require("reshape2")) install.packages("reshape2")

  library(dplyr)
  library(readxl)
  library(tidyverse)
  library(ggpubr)
  library(knitr)
  library(kableExtra)
  library(tidyverse)
  library(agricolae)
  library(broom)
  library(ggplot2)
  library(metan)
  library(ggcorrplot)
  library(corrplot)
  library(reshape2)
})
```

Data Preparation

```
field_data <- read_excel("field_data.xlsx") %>%
  mutate(
    Genotype = as.factor(Genotype),
    Environment = as.factor(Environment),
    Rep = as.factor(Rep)
  )
```

Data Overview

Dataset Structure

```
visualize_data <- function(data) {
  # Start formatted output
  cat("\\begin{shaded}\n")
  cat("\\footnotesize\n")

  # Dataset dimensions
  cat("\\textbf{Dataset Dimensions:} ",
      nrow(data), " observations $\\times$ ", ncol(data), " variables\n\n")

  # Data structure overview
  cat("\\textbf{Data Structure:}\n")
  cat("\\begin{verbatim}\n")
  glimpse_out <- capture.output(dplyr::glimpse(data))
  cat(paste(glimpse_out, collapse = "\n"))
```

```
  cat("\\end{verbatim}\n")

  # Variable types summary
  cat("\n\\textbf{Variable Types:}\n")
  cat("\\begin{verbatim}\n")
  var_types <- tibble::tibble(
    Variable = names(data),
    Type = sapply(data, class),
    `Unique Values` = sapply(data, function(x) length(unique(x))),
    `Missing Values` = sapply(data, function(x) sum(is.na(x)))
  )
  print(var_types)
  cat("\\end{verbatim}\n")

  # End formatted output
  cat("\\end{shaded}\n")
}

# Usage:
visualize_data(field_data)
```

Dataset Dimensions: 900 observations × 4 variables

Data Structure:

```
Rows: 900
Columns: 4
$ Genotype    <fct> G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1~
$ Environment <fct> E1, E1, E1, E2, E2, E2, E3, E3, E3, E4, E4, E4, E5, E5, E5~
$ Rep         <fct> 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2~
$ Yield       <dbl> 7.721363, 8.098298, 8.735440, 10.665111, 10.135101, 10.532~
```

Variable Types:

```
# A tibble: 4 x 4
  Variable    Type    `Unique Values` `Missing Values`
  <chr>       <chr>             <int>            <int>
1 Genotype    factor               20                0
2 Environment factor               15                0
3 Rep         factor                3                0
4 Yield       numeric             900                0
```

Missing Values

```
cat("Missing values:")
```

Missing values:

```
colSums(is.na(field_data))
```

Genotype Environment Rep Yield 0 0 0 0

Statistical Analysis

Perform ANOVA for each environment (RCBD Design)

```
anova_rcbd_pro <- function(data) {
  environments <- unique(data$Environment)

  # Create a list to store all ANOVA results
  anova_results <- list()

  for(env in environments) {
    # Filter data for current environment
    env_data <- data %>% filter(Environment == env)

    # Fit ANOVA model
    model <- aov(Yield ~ Genotype + Rep, data = env_data)

    # Extract and format results
    anova_table <- broom::tidy(model) %>%
      filter(term != "Residuals") %>%
      transmute(
        Source = case_when(
          term == "Genotype" ~ "Genotype",
          term == "Rep" ~ "Block",
          TRUE ~ term
        ),
        DF = df,
        `Sum Sq` = round(sumsq, 2),
        `Mean Sq` = round(meansq, 2),
        `F value` = round(statistic, 2),
        `Pr(>F)` = ifelse(p.value < 0.001, "<0.001", round(p.value, 3))
      )

    # Store results
    anova_results[[env]] <- anova_table

    # Print formatted table
    cat("\n\\subsection*{Environment: ", env, "}\n")
    print(
      knitr::kable(
        anova_table,
        caption = paste("ANOVA Table for", env, "(RCBD Design)"),
        digits = c(0, 0, 2, 2, 2, 3),
        align = c("l", "r", "r", "r", "r", "r"),
        booktabs = TRUE
      ) %>%
        kableExtra::kable_styling(
          latex_options = c("striped", "hold_position"),
          full_width = FALSE,
          font_size = 10
        ) %>%
        kableExtra::footnote(
```

```
        general = "RCBD: Randomized Complete Block Design",
        general_title = "Note:",
        footnote_as_chunk = TRUE
      )
    )
    cat("\n\\vspace{5mm}\n")
  }

  # Return all results invisibly
  invisible(anova_results)
}

# Apply the function
anova_results <- anova_rcbd_pro(field_data)
```

Environment: E1

Table 1: ANOVA Table for E1 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	111.94	5.89	14.05	<0.001
Block	2	0.23	0.12	0.28	0.758

Note: RCBD: Randomized Complete Block Design

Environment: E2

Table 2: ANOVA Table for E2 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	114.95	6.05	25.63	<0.001
Block	2	0.59	0.30	1.25	0.298

Note: RCBD: Randomized Complete Block Design

Environment: E3

Table 3: ANOVA Table for E3 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	87.81	4.62	16.22	<0.001
Block	2	0.51	0.26	0.90	0.414

Note: RCBD: Randomized Complete Block Design

Environment: E4

Table 4: ANOVA Table for E4 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	115.59	6.08	20.17	<0.001
Block	2	0.90	0.45	1.49	0.239

Note: RCBD: Randomized Complete Block Design

Environment: E5

Table 5: ANOVA Table for E5 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	106.97	5.63	18.34	<0.001
Block	2	1.55	0.78	2.53	0.093

Note: RCBD: Randomized Complete Block Design

Environment: E6

Table 6: ANOVA Table for E6 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	164.69	8.67	33.32	<0.001
Block	2	0.87	0.44	1.67	0.201

Note: RCBD: Randomized Complete Block Design

Environment: E7

Table 7: ANOVA Table for E7 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	121.92	6.42	22.75	<0.001
Block	2	0.15	0.07	0.26	0.77

Note: RCBD: Randomized Complete Block Design

Environment: E8

Table 8: ANOVA Table for E8 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	108.82	5.73	24.20	<0.001
Block	2	3.10	1.55	6.56	0.004

Note: RCBD: Randomized Complete Block Design

Environment: E9

Table 9: ANOVA Table for E9 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	108.69	5.72	12.73	<0.001
Block	2	1.74	0.87	1.94	0.158

Note: RCBD: Randomized Complete Block Design

Environment: E10

Table 10: ANOVA Table for E10 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	122.73	6.46	28.84	<0.001
Block	2	0.20	0.10	0.46	0.638

Note: RCBD: Randomized Complete Block Design

Environment: E11

Table 11: ANOVA Table for E11 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	98.32	5.17	14.74	<0.001
Block	2	0.58	0.29	0.83	0.446

Note: RCBD: Randomized Complete Block Design

Environment: E12

Table 12: ANOVA Table for E12 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	99.40	5.23	14.09	<0.001
Block	2	1.27	0.64	1.72	0.193

Note: RCBD: Randomized Complete Block Design

Environment: E13

Table 13: ANOVA Table for E13 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	79.23	4.17	13.57	<0.001
Block	2	2.62	1.31	4.26	0.021

Note: RCBD: Randomized Complete Block Design

Environment: E14

Table 14: ANOVA Table for E14 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	150.27	7.91	24.82	<0.001
Block	2	0.40	0.20	0.63	0.536

Note: RCBD: Randomized Complete Block Design

Environment: E15

Table 15: ANOVA Table for E15 (RCBD Design)

Source	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	19	130.96	6.89	20.75	<0.001
Block	2	0.16	0.08	0.24	0.785

Note: RCBD: Randomized Complete Block Design

Joint ANOVA for Multi-Environment Trial (RCBD Design)

```
# Simple Joint ANOVA Analysis
model <- aov(Yield ~ Environment + Genotype + Environment:Rep + Environment:Genotype,
             data = field_data)

# Get ANOVA table
anova_table <- broom::tidy(model) %>%
  filter(term != "Residuals") %>%
  transmute(
    Source = case_when(
      term == "Environment" ~ "Environment",
      term == "Genotype" ~ "Genotype",
      term == "Environment:Genotype" ~ "GxE Interaction",
      term == "Environment:Rep" ~ "Block(Environment)",
      TRUE ~ term
    ),
    DF = df,
    SS = round(sumsq, 2),
    MS = round(meansq, 2),
    F = round(statistic, 2),
    p.value = ifelse(p.value < 0.001, "<0.001", round(p.value, 3))
  )

# Print formatted table
knitr::kable(
  anova_table,
  caption = "Joint ANOVA Results (RCBD Design)",
  digits = c(0, 0, 2, 2, 2, 3),
  align = c("l", "r", "r", "r", "r", "r")
)
```

Table 16: Joint ANOVA Results (RCBD Design)

Source	DF	SS	MS	F	p.value
Environment	14	617.98	44.14	141.43	<0.001
Genotype	19	1302.50	68.55	219.64	<0.001
Block(Environment)	30	14.89	0.50	1.59	0.025
GxE Interaction	266	419.81	1.58	5.06	<0.001

LSD (Least Significant Difference) Test

```
# Perform LSD test
lsd_test <- LSD.test(y = field_data$Yield,
                     trt = field_data$Genotype,
                     DFerror = model$df.residual,
                     MSerror = deviance(model)/model$df.residual,
                     alpha = 0.05,
                     group = TRUE,
```

```
                console = FALSE)

# Prepare clean results
lsd_results <- data.frame(
  Genotype = rownames(lsd_test$groups),
  Mean = lsd_test$groups[,1],
  Group = lsd_test$groups[,2]
) %>%
  arrange(desc(Mean))

# Professional table output
lsd_results %>%
  kable(
    caption = "Genotype Mean Comparison by LSD Test (  = 0.05)",
    col.names = c("Genotype", "Mean Yield", "Significance Group"),
    align = c("l", "r", "c"),
    digits = 2,
    booktabs = TRUE
  ) %>%
  kable_styling(
    latex_options = c("striped", "hold_position"),
    full_width = FALSE,
    font_size = 12
  ) %>%
  column_spec(2, bold = TRUE) %>%
  footnote(
    general = "Means followed by the same letter are not significantly different (p > 0.05)",
    general_title = "Note:",
    footnote_as_chunk = TRUE
  )
```

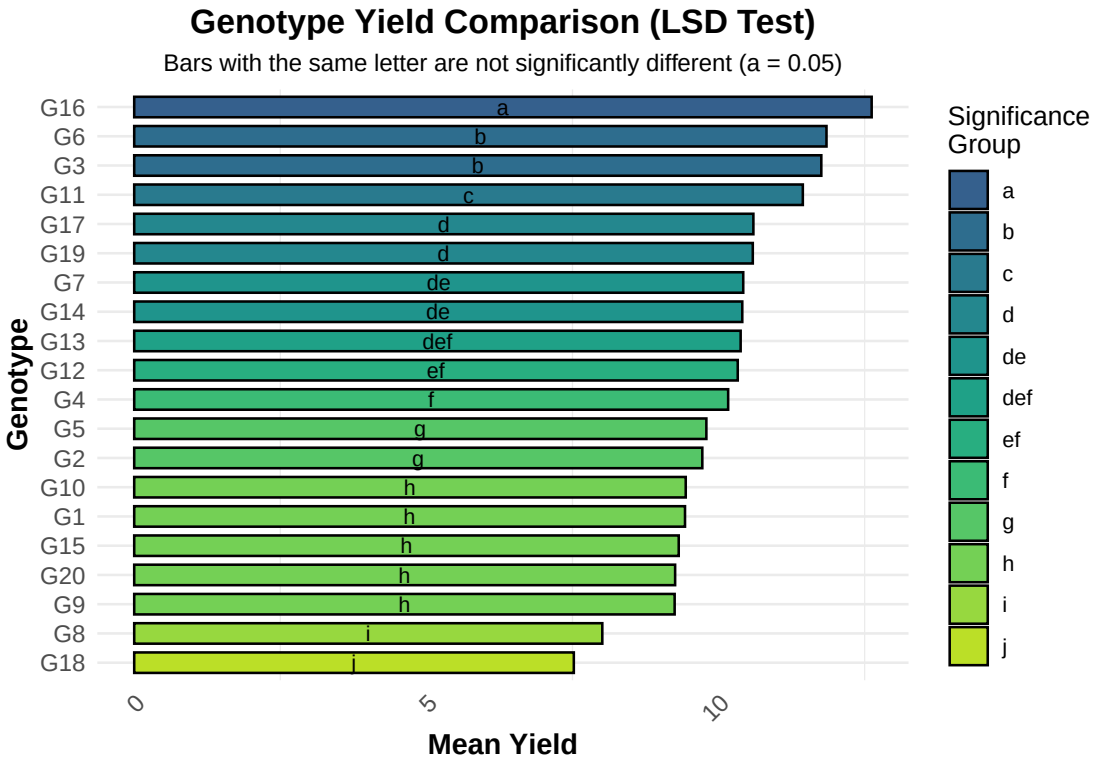
Table 17: Genotype Mean Comparison by LSD Test (= 0.05)

Genotype	Mean Yield	Significance Group
G16	**12.62**	a
G6	**11.85**	b
G3	**11.76**	b
G11	**11.44**	c
G17	**10.60**	d
G19	**10.59**	d
G7	**10.42**	de
G14	**10.41**	de
G13	**10.38**	def
G12	**10.33**	ef
G4	**10.16**	f
G5	**9.79**	g
G2	**9.72**	g

G10	**9.44**	h
G1	**9.43**	h
G15	**9.32**	h
G20	**9.26**	h
G9	**9.25**	h
G8	**8.01**	i
G18	**7.52**	j

Note: Means followed by the same letter are not significantly different (p > 0.05)

```
# Prepare data for plotting
plot_data <- lsd_results %>%
  mutate(Genotype = factor(Genotype, levels = Genotype[order(Mean)])) %>%
  arrange(Mean)

# Create the plot
ggplot(plot_data, aes(x = Genotype, y = Mean, fill = Group)) +
  geom_bar(stat = "identity", width = 0.7, color = "black") +
  geom_text(aes(label = Group),
            position = position_stack(vjust = 0.5),
            size = 3, color = "black") +
  scale_fill_viridis_d(option = "D", begin = 0.3, end = 0.9) +
  labs(title = "Genotype Yield Comparison (LSD Test)",
       subtitle = "Bars with the same letter are not significantly different (  = 0.05)",
       y = "Mean Yield",
       x = "Genotype",
       fill = "Significance\nGroup") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 10),
    axis.text.y = element_text(size = 10),
    axis.title = element_text(size = 12, face = "bold"),
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 10, hjust = 0.5),
    legend.position = "right",
    panel.grid.major.x = element_blank()
  ) +
  coord_flip()  # Horizontal bars for better readability
```



Variance Components and Genetic Parameters from Multi-Environment Trial

```
# Calculate components
ms_gen <- summary(model)[[1]]["Genotype", "Mean Sq"]
ms_gxe <- summary(model)[[1]]["Environment:Genotype", "Mean Sq"]
ms_res <- summary(model)[[1]]["Residuals", "Mean Sq"]

n_rep <- length(unique(field_data$Rep))
n_env <- length(unique(field_data$Environment))
grand_mean <- mean(field_data$Yield)

var_g <- (ms_gen - ms_gxe)/(n_rep*n_env)
var_gxe <- (ms_gxe - ms_res)/n_rep
var_e <- ms_res/n_rep
var_p <- var_g + var_gxe/n_env + var_e/(n_rep*n_env)
h2 <- (var_g/var_p)*100
cv <- (sqrt(var_p)/grand_mean)*100

# Create table with ASCII characters only
results <- data.frame(
  Parameter = c("Genotypic variance (s2g)",
                "GxE interaction variance (s2gxe)",
                "Environmental variance (s2e)",
                "Phenotypic variance (s2p)",
```

```
            "Broad-sense heritability (H2%)",
            "Phenotypic CV (%)"),
  Estimate = c(var_g, var_gxe, var_e, var_p, h2, cv)
)

# Print table
knitr::kable(
  results,
  caption = "Genetic Variance Components",
  digits = 4,
  col.names = c("Variance Component", "Estimate"),
  align = c("l", "r")
) |>
kableExtra::kable_styling(
  bootstrap_options = c("striped", "hover"),
  full_width = FALSE,
  font_size = 14
)
```

Table 18: Genetic Variance Components

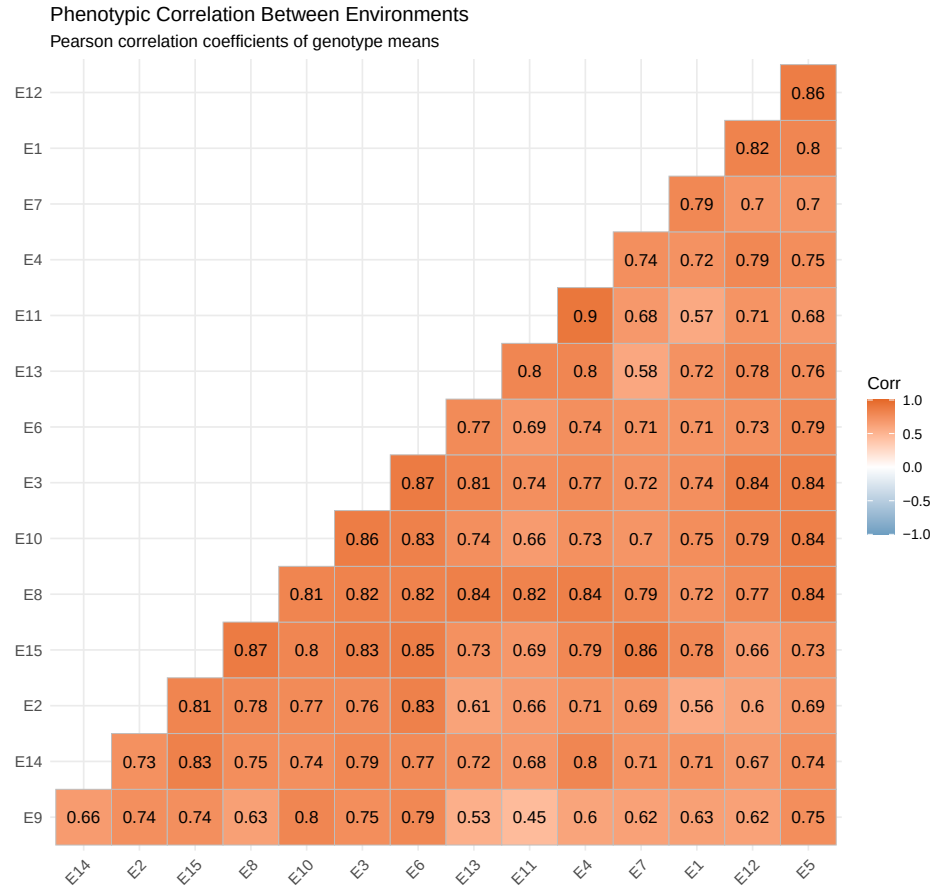
Variance Component	Estimate
Genotypic variance (s2g)	1.4883
GxE interaction variance (s2gxe)	0.4220
Environmental variance (s2e)	0.1040
Phenotypic variance (s2p)	1.5188
Broad-sense heritability (H2%)	97.9952
Phenotypic CV (%)	12.1849

Correlations between ENVIRONMENTS

Phenotypic (Based on observed averages)

```
# Phenotypic means matrix
pheno_matrix <- field_data %>%
  group_by(Genotype, Environment) %>%
  summarise(Yield_Mean = mean(Yield, na.rm = TRUE), .groups = 'drop') %>%
  pivot_wider(names_from = Environment, values_from = Yield_Mean) %>%
  column_to_rownames("Genotype")

# Correlation matrix with improved visualization
cor_pheno <- cor(pheno_matrix, use = "complete.obs")

ggcorrplot(cor_pheno,
           hc.order = TRUE,
           type = "lower",
```

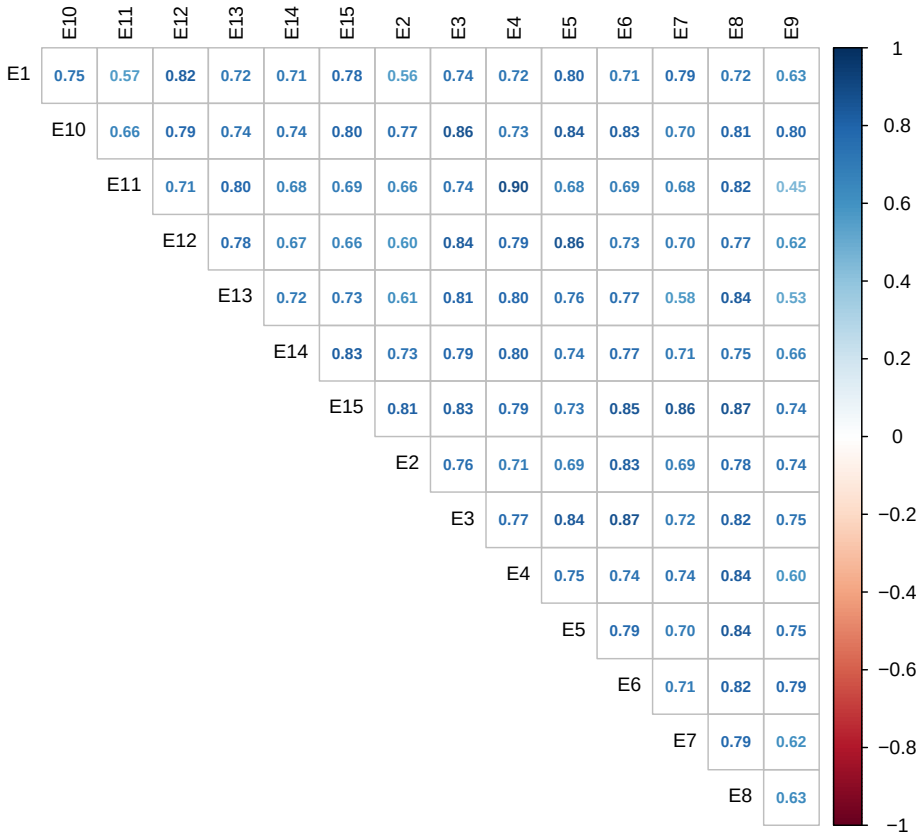
```
        lab = TRUE,
        lab_size = 4,
        digits = 2,
        tl.cex = 10,
        colors = c("#6D9EC1", "white", "#E46726"),
        outline.color = "gray",
        ggtheme = theme_minimal()) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 10),
        axis.text.y = element_text(size = 10),
        legend.position = "right") +
  labs(title = "Phenotypic Correlation Between Environments",
       subtitle = "Pearson correlation coefficients of genotype means")
```



Genetic (Based on adjusted genotypic effects)

```
# Calculate genotype means per environment
env_means <- field_data %>%
  group_by(Genotype, Environment) %>%
  summarise(Mean_Yield = mean(Yield, na.rm = TRUE), .groups = 'drop')

# Convert to wide format
env_matrix <- env_means %>%
```

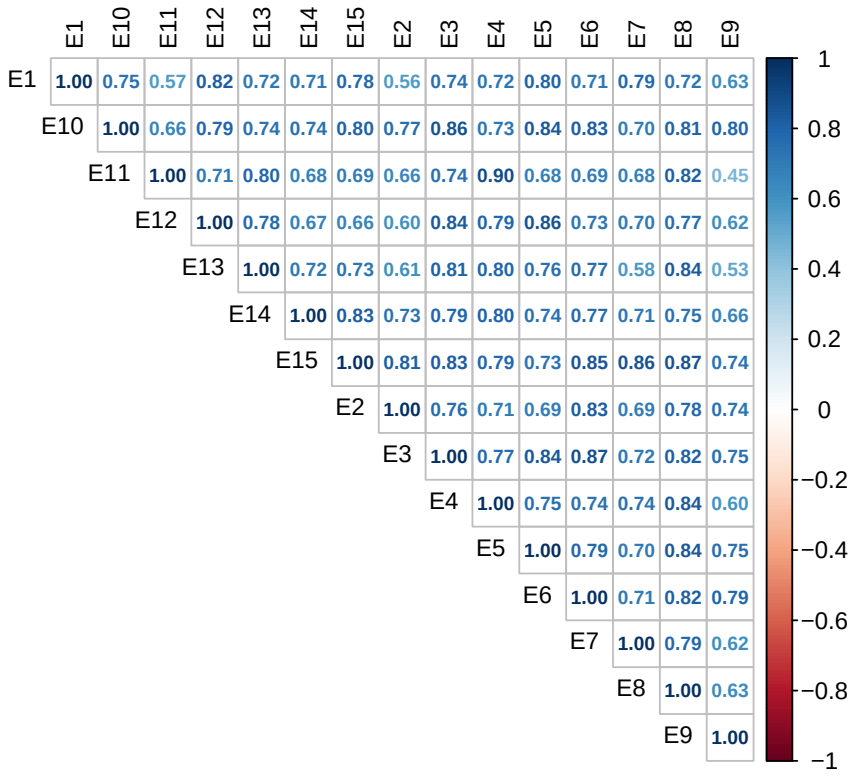
```
  pivot_wider(names_from = Environment, values_from = Mean_Yield) %>%
  as.data.frame()

# Set genotypes as row names
rownames(env_matrix) <- env_matrix$Genotype
env_matrix <- env_matrix[-1]  # Remove Genotype column

# Calculate correlation matrix
env_cor_matrix <- cor(env_matrix, use = "complete.obs")

# Visualize correlations with identical formatting
corrplot(env_cor_matrix,
         method = "number",     # Show correlation coefficients as numbers
         type = "upper",        # Display only upper triangle
         tl.col = "black",      # Black text for environment names
         tl.cex = 0.8,          # Text label size
         number.cex = 0.7,      # Correlation number size
         mar = c(0, 0, 0, 0), # Margins (bottom, left, top, right)
         diag = FALSE)          # Hide diagonal (1's)
```



Environmental (Based on non-genetic effects)

```
# Calculate genotype means per environment
env_means <- field_data %>%
```

```
  group_by(Genotype, Environment) %>%
  summarise(Mean_Yield = mean(Yield, na.rm = TRUE), .groups = 'drop')

# Convert to wide format (genotypes as rows, environments as columns)
env_matrix <- env_means %>%
  pivot_wider(names_from = Environment, values_from = Mean_Yield) %>%
  as.data.frame()

# Set genotypes as row names and remove the column
rownames(env_matrix) <- env_matrix$Genotype
env_matrix <- env_matrix[-1]  # Remove Genotype column

# Calculate correlation matrix
env_cor_matrix <- cor(env_matrix, use = "complete.obs")

# Visualize correlations
corrplot(env_cor_matrix,
         method = "number",
         type = "upper",
         tl.col = "black",
         tl.cex = 0.8,
         number.cex = 0.7)
```



Environmental effects for each location

```
# 1. Calculate environmental effects for each location
env_effects <- field_data %>%
  group_by(Environment) %>%
  summarise(Env_Effect = mean(Yield, na.rm = TRUE) - mean(field_data$Yield, na.rm = TRUE))

# 2. Create a bar plot instead of correlation plot (more appropriate for this data)
ggplot(env_effects, aes(x = reorder(Environment, Env_Effect), y = Env_Effect)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Environmental Effects by Location",
       x = "Environment",
       y = "Effect (Deviation from Grand Mean)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red")
```