# Statistical Analysis for Plant Breeding

**Luis Fernando Delgado Munoz**
luis.delgado@cgiar.org
Faisalabad, November 2023

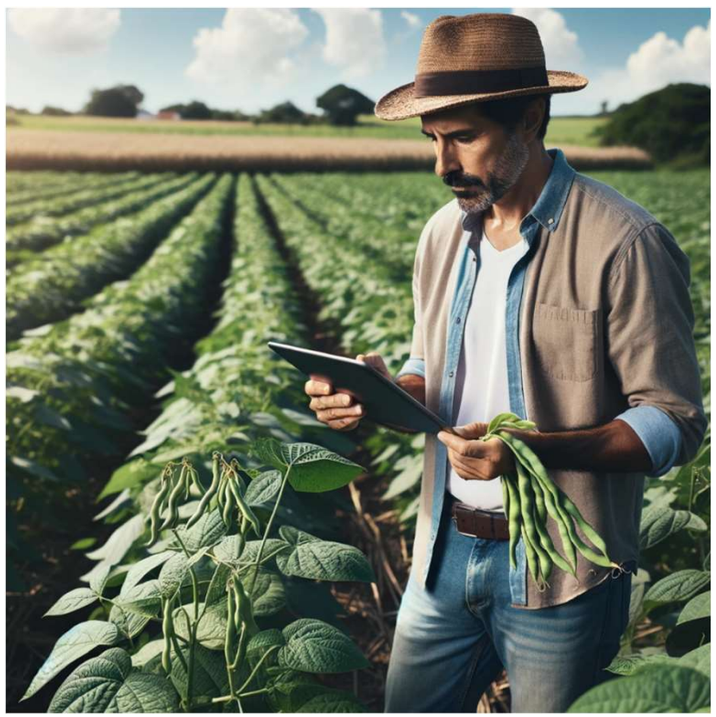
@BiovIntCIAT_eng
@BiovIntCIAT_esp

*#Alliance4Science*

# Introduction

Plant breeders are required to have knowledge in many disciplines including statistics.

At the heart of plant breeding is designing and analyzing yield trials to rank varieties in order of merit, which make **statistics** an essential component in plant breeding.

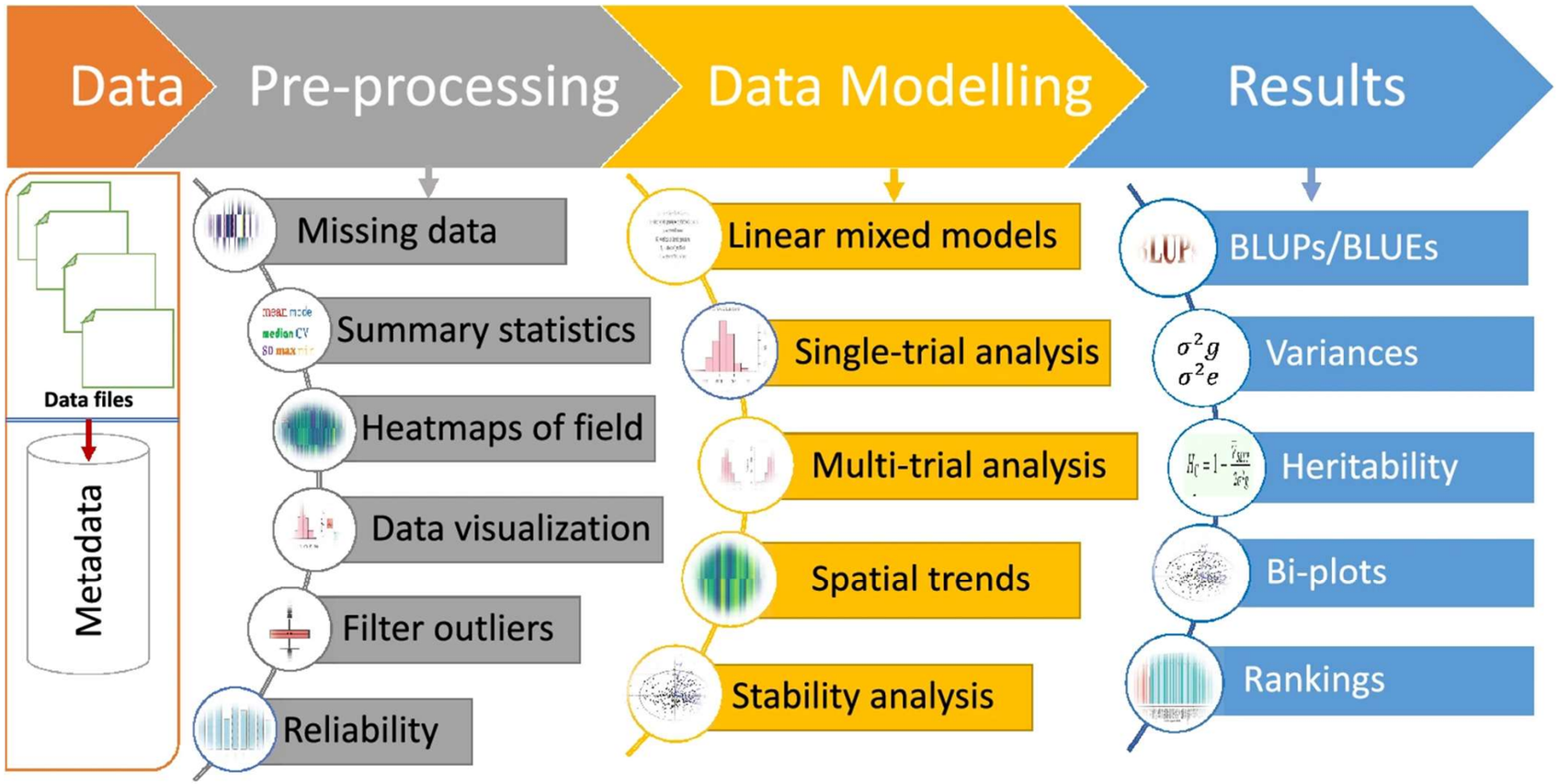


COMMENTARY Open Access

# Open-source analytical pipeline for robust data analysis, visualizations and sharing in crop breeding

doi: https://doi.org/10.1186/s13007-022-00845-7

Waseem Hussain*, Mahender Anumalla, Margaret Catolos, Apurva Khanna, Ma. Teresa Sta. Cruz, Joie Ramos and Sankalp Bhosale



# Importance of statistics in plant breeding

**Systematic Data Collection and Interpretation:**

Statistics plays a central role in the systematic collection and interpretation of numerical data in plant breeding. This includes big data like phenotype, sequence, and pedigree data. By exploiting these data effectively, statistics can help accelerate the breeding process and enhance understanding of the underlying biological mechanisms.

Alliance Bioversity & CIAT

CGIAR

# Phenotyping in bean crop

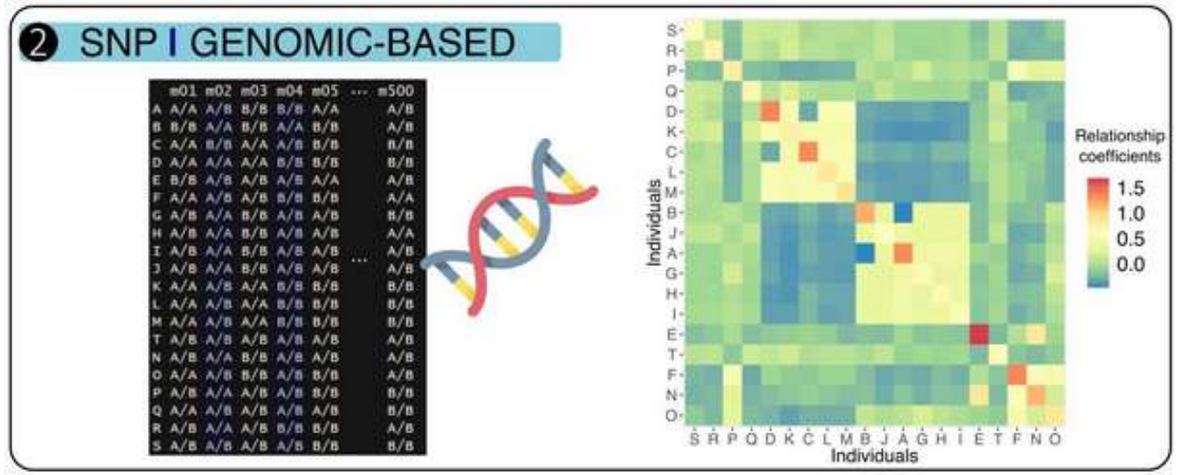

Bean Mattson cooker

Mineral’s rovers – don roverto

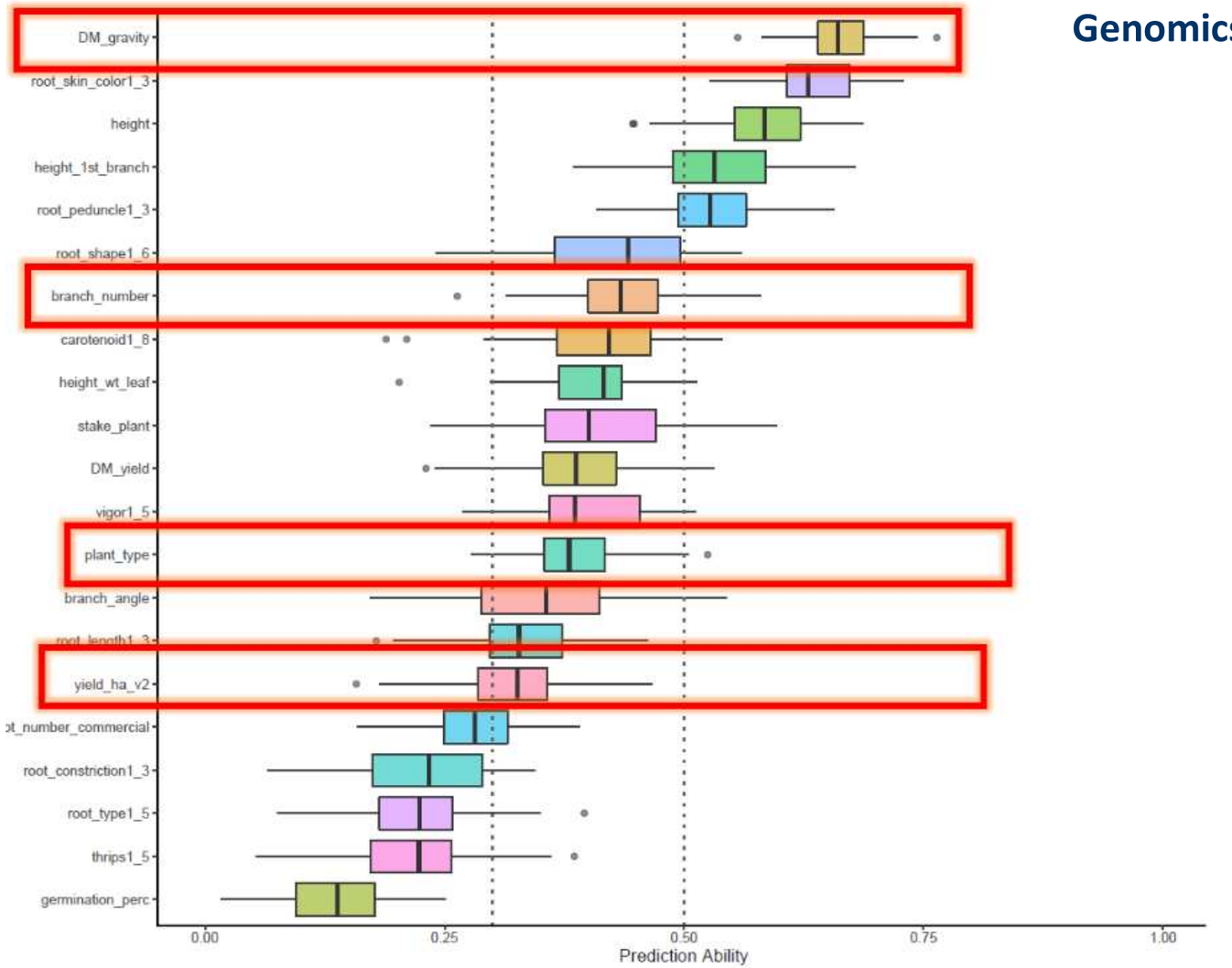
LysipheN - plant´s transpiration rate

## Genomic Selection and Prediction:

The implementation of genomic selection and prediction in plants is increasingly common. Statistics provide a crucial link in these processes, enabling the incorporation of genomic information into plant breeding programs.



# Preliminary genomic selection results in Cassava




**Genomics assisted selection in cassava breeding**

**Challenges in Early-Stage Selection Trials:**

In the initial stages of plant breeding, a significant challenge is identifying a few superior individuals from a large number of cross-pollinated lines.

Due to limitations on planting material and space for field testing.

- Partly replicated genotypes
- Non replicate genotypes (augmented chaecks)
- Single rows.

Statistical methods are essential for making accurate selections under these constraints

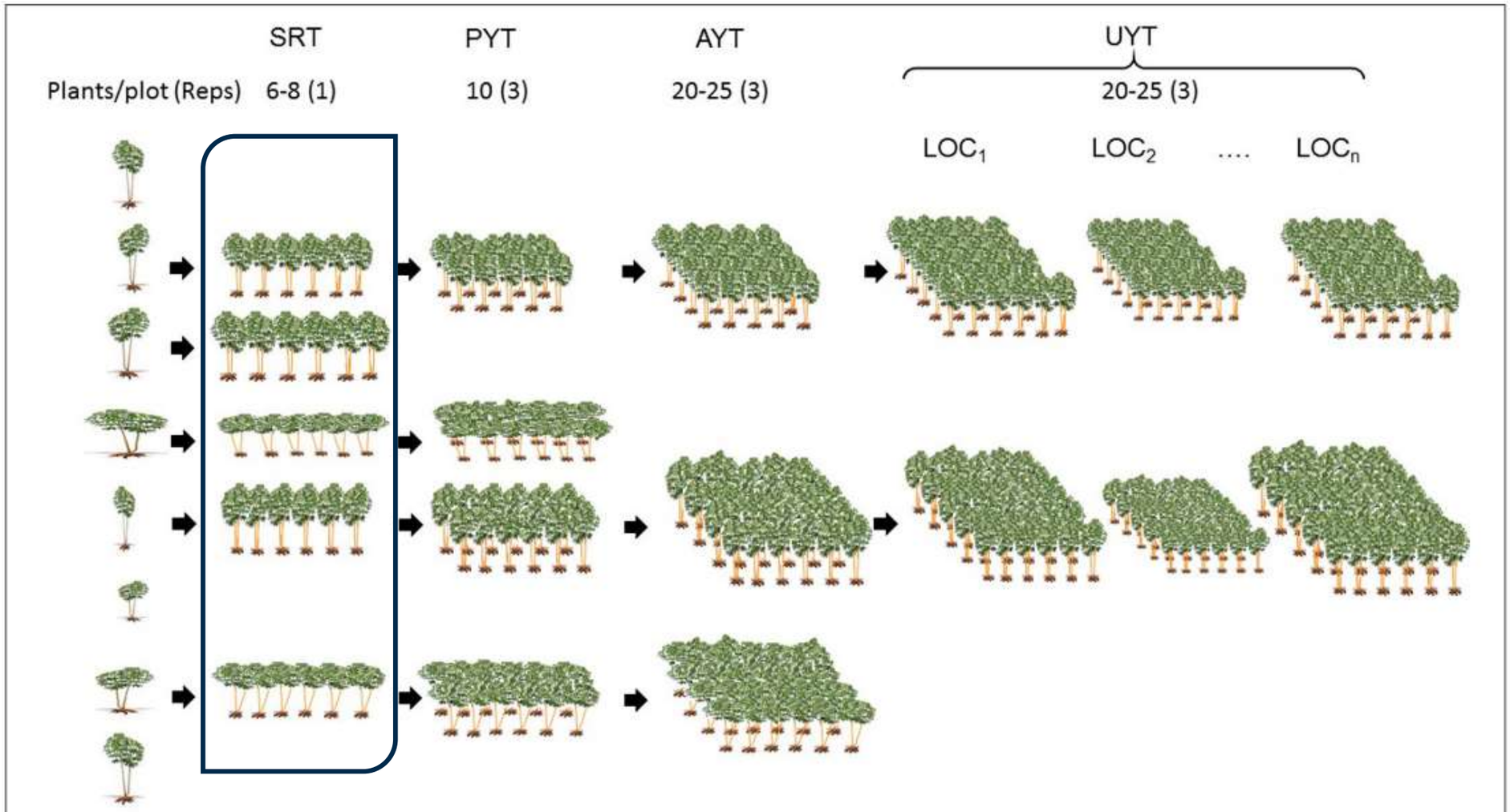


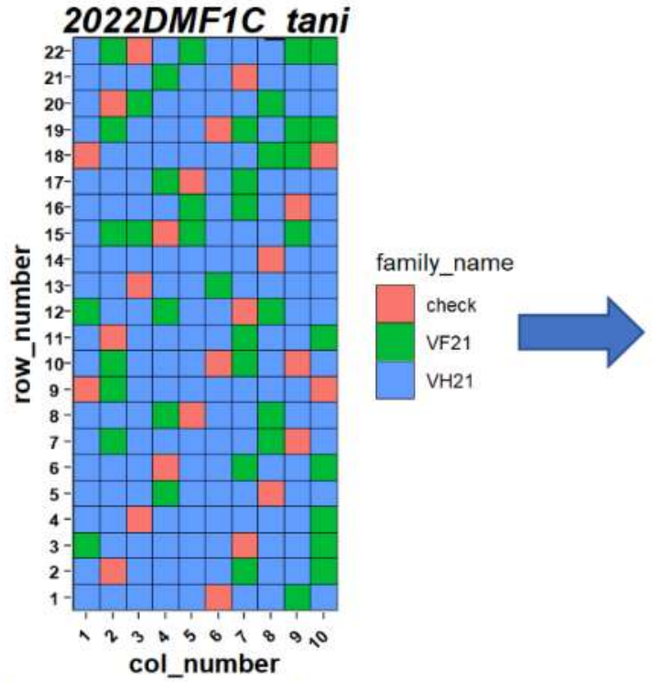

**FIGURE 1 | Illustration of the different stages of a typical evaluation process in cassava breeding.** Plants from germinated seed (seedling plants) are grown in the field and used as the source of clonal planting material (left side). The first evaluation takes place in single row trials (SRT), followed by preliminary (PYT) and advanced (AYT) yield trials. The first multi-location evaluation is in the uniform yield trials (UYT), or sometimes earlier, in the AYTs. Size of plots in UYT has been slightly modified to illustrate the effect of different environments on the growth of cassava.

| | STG 1 | STG 2 | STG 3 | STG 4 | STG 5 |
|---|---|---|---|---|---|
| Augmented design | 1 | | | | |
| P-rep design | | 1.2 | | | |
| RCBD / alpha-lattice | | | 2 | | |
| RCBD / alpha-lattice | | | | 2 - 3 | |
| RCBD / alpha-lattice | | | | | 2 - 3 |

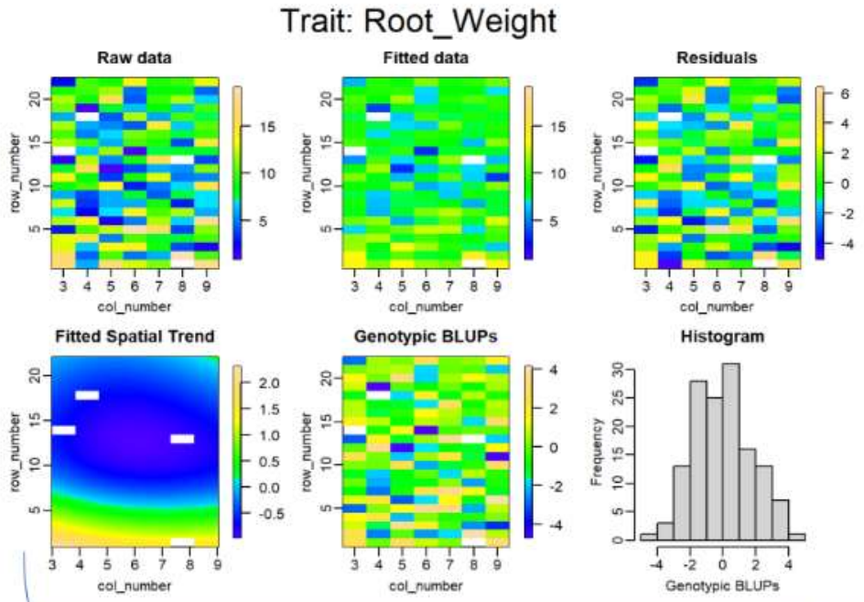
# Locations

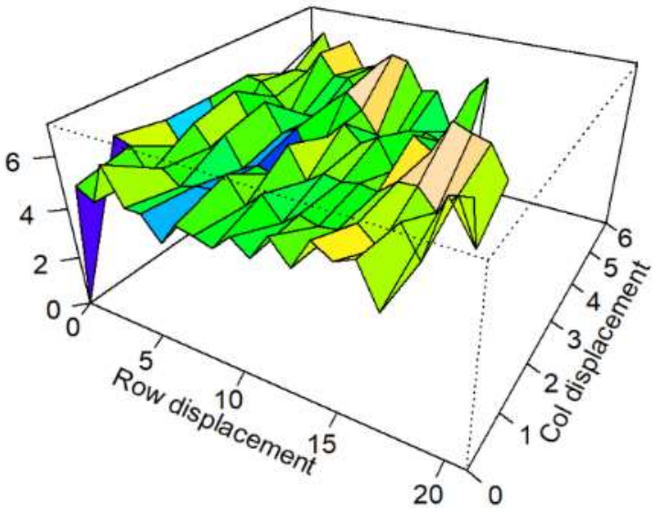
Replications / location

**Augmented designs** are used in early-generation trial




Augmented Trial layout




Spatial correction analysis

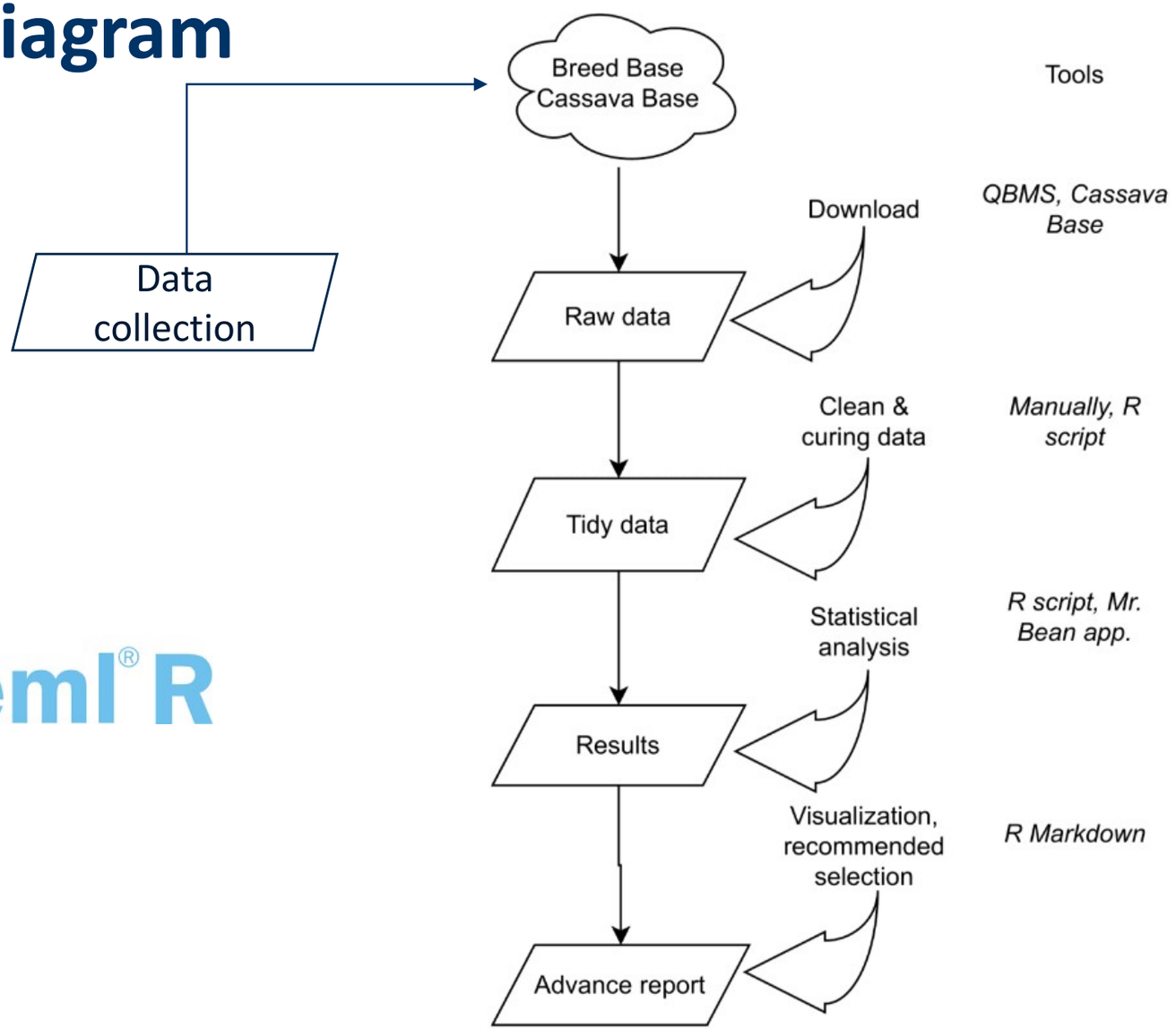
**Data Analysis in Plant Breeding:**

Classical univariate and multivariate statistics are widely used for data analysis in plant breeding and biotechnology.

These statistical methods are essential for evaluating genetic diversity, classifying plant genotypes, analyzing yield components, assessing yield stability, evaluating biotic and abiotic stresses, and more.

# Data analysis flow diagram



# Data Analysis

**Clean data**

**Analysis**

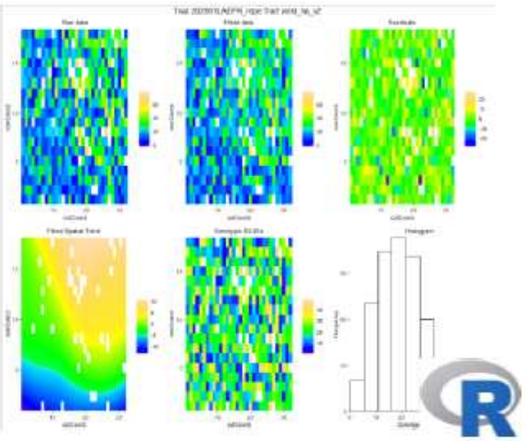
**Results**

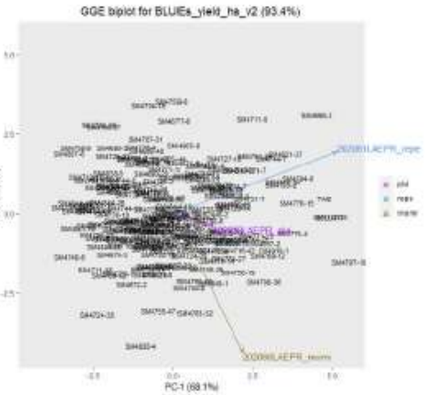
**Individual trial**
- Spatial adjustment
- Broad-sense heritability ($V_G$)
- BLUE and seBLUE
- LSD, mean, CV

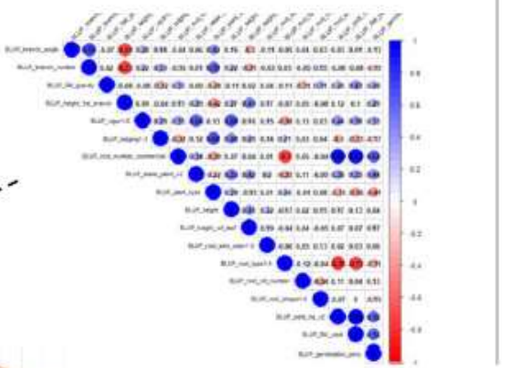
**GxE**
- AMMI and SREG analysis
- Stability of each genotype
- Genetic correlation among trials

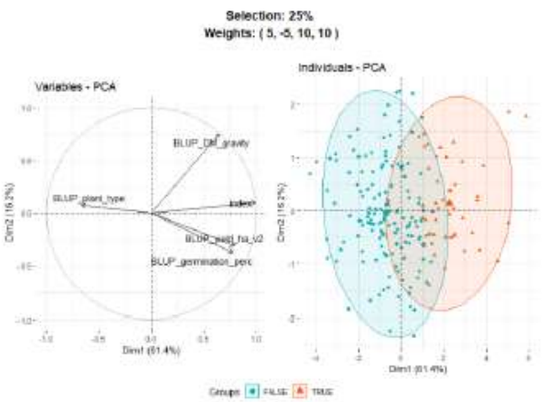
**BLUP**
- BLUP
- Genetic correlation among traits
- BLUP and BLUE in a table

**Selection index**
- Trait, weight, % of selection








**Tools:** MrBean **& BreedBase & R script**

*“To consult the statistician after an experiment is finished is often merely to ask him to conduct a postmortem examination. He can perhaps say what the experiment died of.”*

**Ronald Fisher. 1938.**



# Thank you!