



Society of Agricultural Scientists, Nepal (SAS-Nepal)

You are cordially invited to attend
Talk Program, Series -11
"Recent Frontiers in Statistical Road from Research
to Publication"

माघ २९, २०८०,
12 Feb, 2024, Monday

National Agricultural Research Institute (NARI) Hall
NARC, Khumaltar, Kathmandu



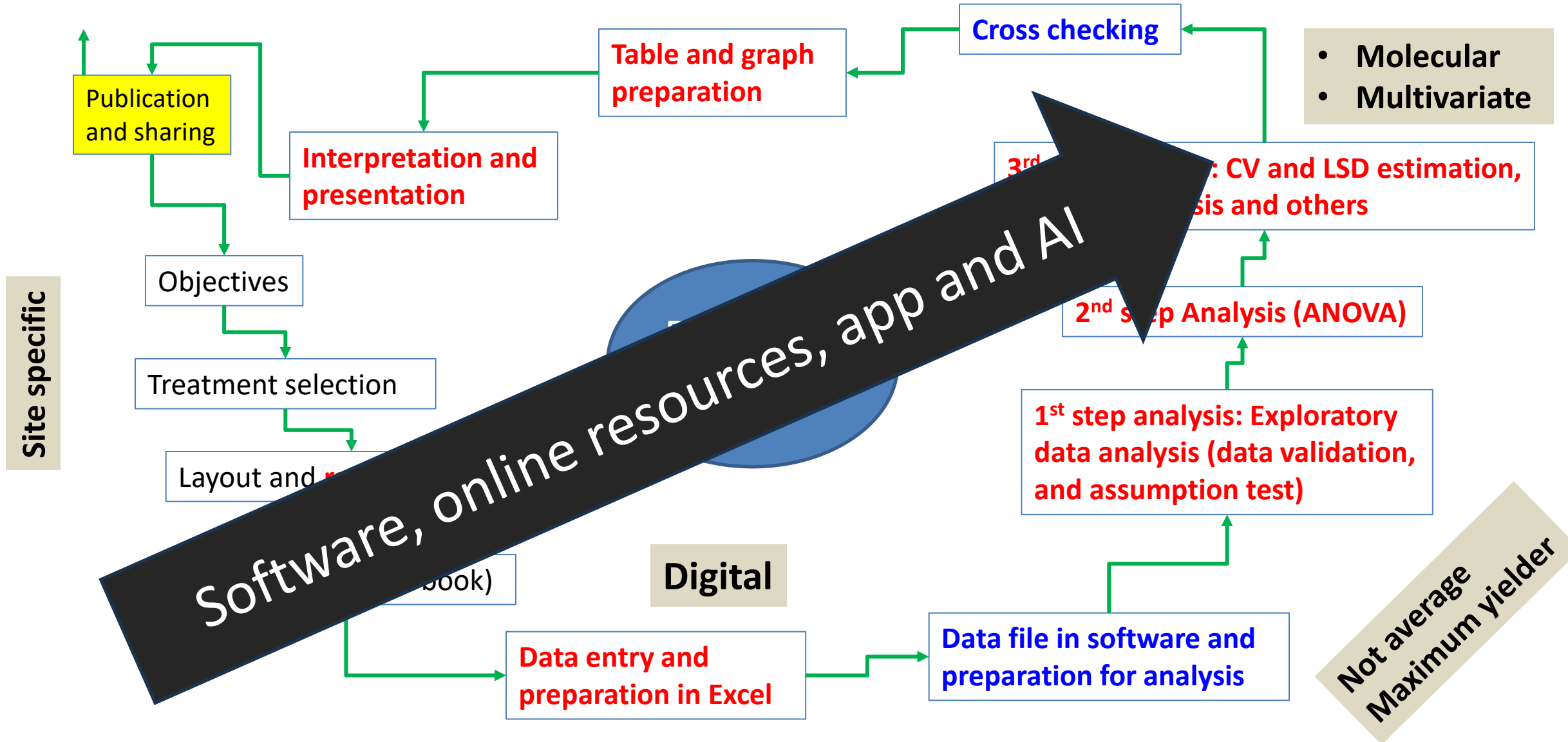
Speaker

Time: 11:00 to 12:00 pm

Dr Bal Krishna Joshi
Senior Scientist, Nepal Agricultural Research Council
Chief, National Agriculture Genetic Resources Center
(Genebank), Khumaltar

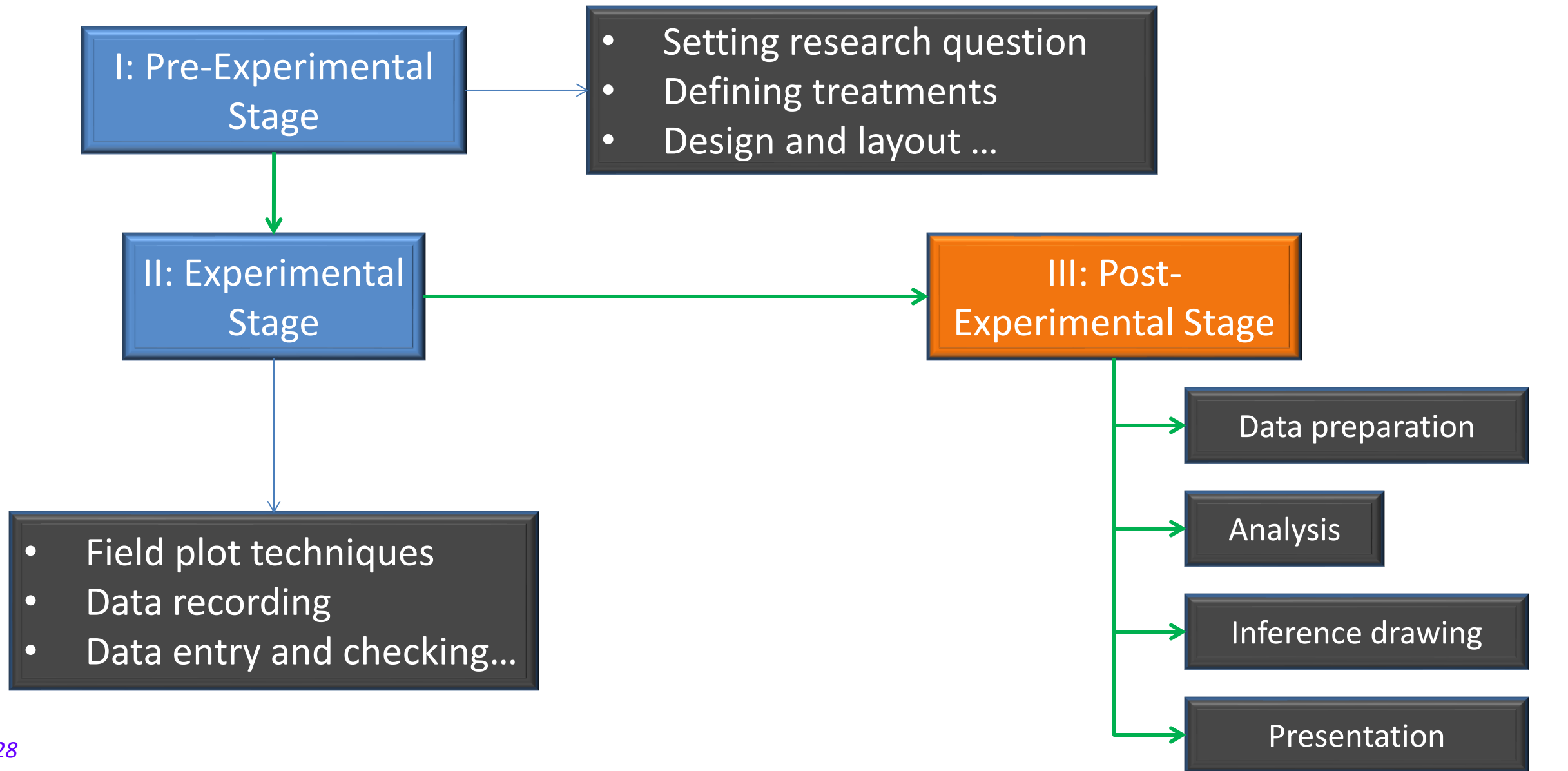


Research Cycle: Statistical Perspective





Research Steps: Statistical Perspective





Research Targets: Only farmers? 5 beneficiaries: FAR-BP)

- Farmers
- Academicians
- Researchers
- Businessperson
- Policy makers



Decision?

- Farmer's decision
- **Biometrical decision**
- Observation based decision



Pre experimental stage

- Sample size calculator: <https://www.calculator.net/sample-size-calculator.html>, <https://www.surveymonkey.com/mp/sample-size-calculator/>
- **Treatment:** Global and national (defining the scope of findings)



Layout and Randomization

■ Randomization

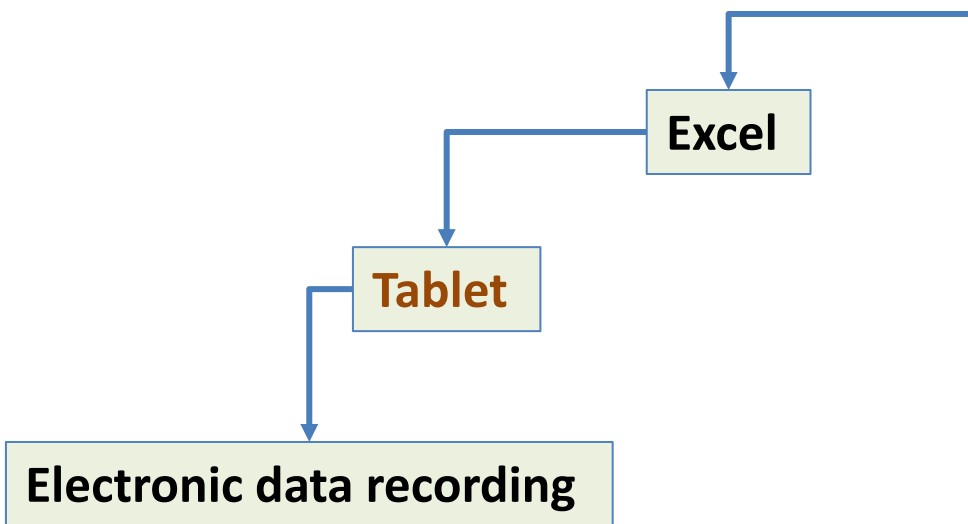
R, randomizr, agricolae

Research Randomizer

List Randomizer

DiGGER

Layout, randomization and field book with DiGGER



Plot Number Name of Variety Entry Number Row Column Replication

UNIT	ID	ENTRY	ROW	RANGE	REP	TRT
1	V46	46	1	1	1	46
2	V10	10	2	1	1	10
3	V14	14	3	1	1	14
4	V37	37	4	1	1	37
5	V34	34	5	1	1	34
6	V49	49	6	1	1	49
7	V4	4	7	1	1	4
8	V33	33	8	1	1	33
9	V24	24	9	1	1	24
10	V36	36	10	1	1	36
11	V2	2	11	1	2	2
12	V21	21	12	1	2	21
13	V27	27	13	1	2	27
14	V6	6	14	1	2	6
15	V26	26	15	1	2	26
16	V35	35	16	1	2	35
17	V33	33	17	1	2	33
18	V16	16	18	1	2	16
19	V22	22	19	1	2	22



Experimental Stage

- Software based experimental layout and design, and field book generation

- R package: agricolae**

- design.ab
- design.alpha
- design.bib
- design.crd
- design.cyclic
- design.dau
- design.graeco
- design.lattice
- design.lsd
- design.mat
- design.rcbd
- design.split
- design.strip
- design.youden

```

7 # factorial 2 x 2 x 2 with 5 replications in completely randomized des
8 trt<-c(2,2,2)
9 outdesign<-design.ab(trt, r=5, serie=2,design="crd")
10 book<-outdesign$book
11 print(book)
12 # factorial 3 x 3 in latin square design.
13 trt <-c(3,3)
14 outdesign<-design.ab(trt, serie=2, design="lsd")
15 book<-outdesign$book
16 print(book)
17 |
17:1 (Top Level)
R Script

```

```

R 4.2.3 . ~/
9      109      2 1 1
10     110      2 2 1
> # factorial 2 x 2 x 2 with 5 replications in completely randomized desig
n.
> trt<-c(2,2,2)
> outdesign<-design.ab(trt, r=5, serie=2,design="crd")
> book<-outdesign$book
> print(book)
      plots r A B C
1      101 1 1 1 2
2      102 1 2 1 2
3      103 1 2 2 2
4      104 2 2 2 2
5      105 1 2 1 1
6      106 1 1 1 1
7      107 1 1 2 2
8      108 2 2 1 2
9      109 2 1 1 1
10     110 1 2 2 1
11     111 3 2 2 2
12     112 3 2 1 2
13     113 2 1 1 2
14     114 2 2 2 1

```

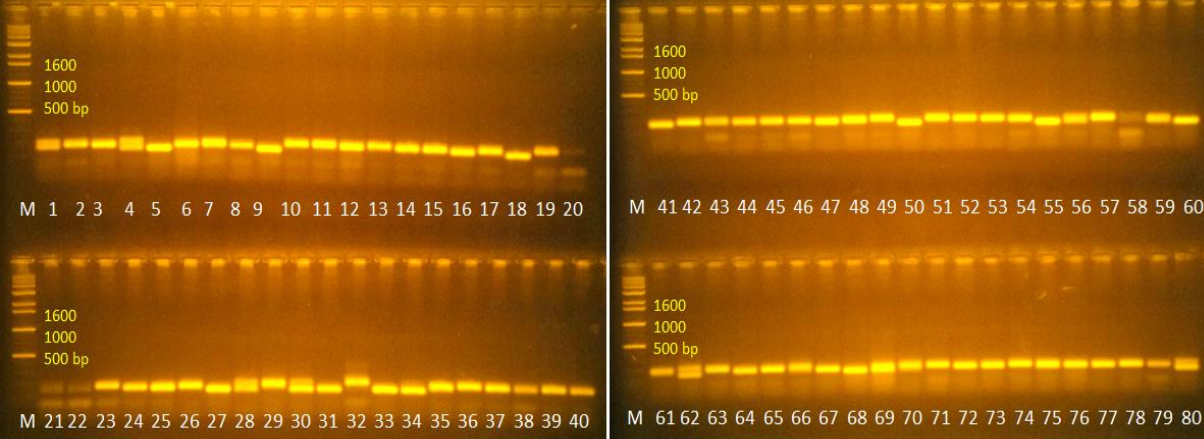


Data Generation: On-station, On-farm and Lab Experiments, and Survey

Data collection app & online survey tools (Electronic media based survey tools) + image analysis (ImageJ)

- **Field Book:** An open-source application for field data collection on Android.
- **Phenobook:** a software tool that can be easily implemented in collaborative research and development projects involving data collecting and forward analyses
- **FieldLab:** An application for Android tablet that used for data collection in the field.
- **ODK:** Open Data Kit, open-source software for collecting, managing and using data in resource-constrained environments.

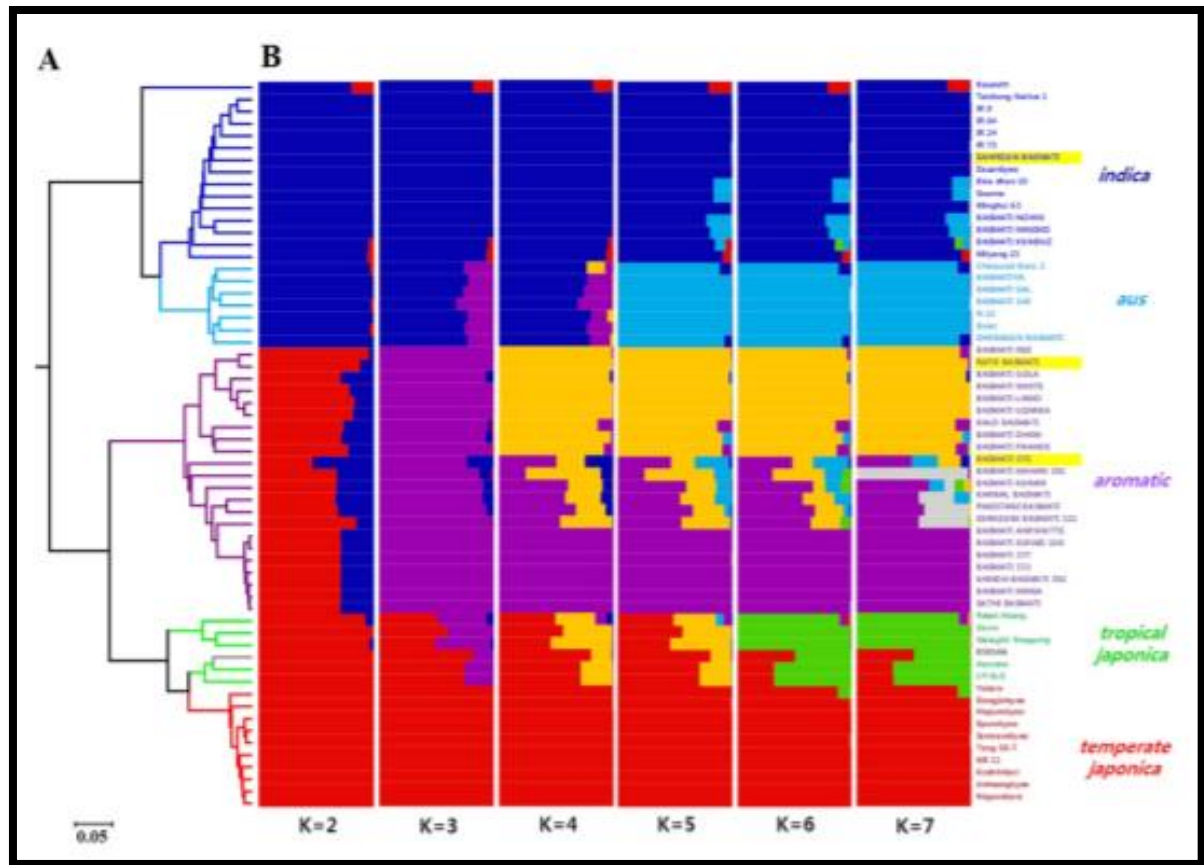
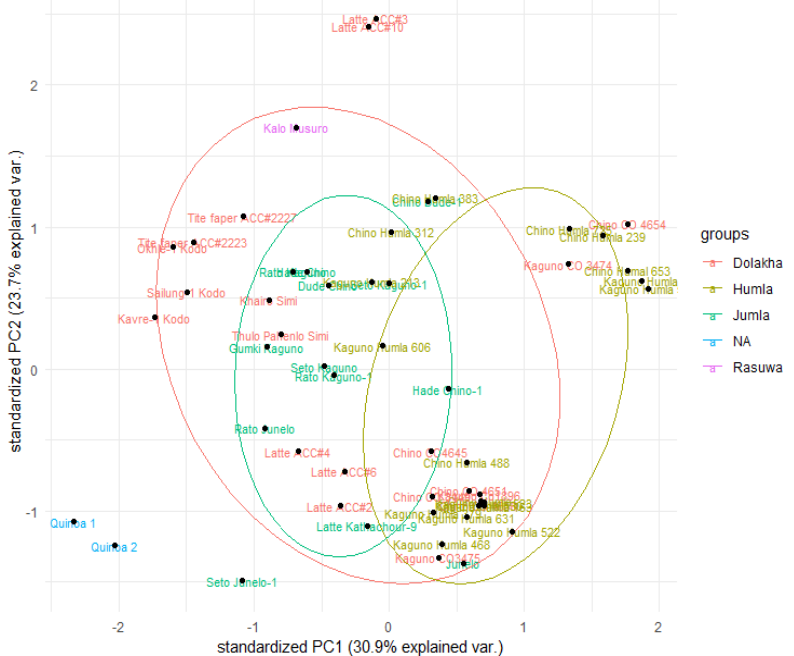
- **SurveyMonkey** (<https://www.surveymonkey.com/>): A cloud-based survey tool that helps users create, send and analyze surveys.
- **Google forms** (<https://www.google.com/forms/about/>): Create online forms and surveys
- **SoGoSurvey** (<https://experience.sogosurvey.com/>): A cloud-based platform that enables creation, distribution and multilingual analysis of surveys, forms, polls, quizzes and assessments.
- **mWater Portal** (<https://portal.mwater.co/#/>): Free platform for data collection, data visualizations, and data-driven management of infrastructure in emerging economies.



Phenomics + Genomics



Genetic parameters: Justification for your pheno results





Partition of Variation

1. ANOVA
2. AMOVA: Analysis of MOlecular VAriance
3. MANOVA: Multivariate analysis of variance

4. ANCOVA
5. MANCOVA

Reporting exact P value

AMOVA = Analysis of molecular variance
 N = Total number of individuals
 df = degree of freedom
 SSD = Sum of squared deviations
 EMS= Expected mean square

Source of variance	df	SS	EMS
Among groups, species	G-1	SSD (AG)	$n''\sigma^2_a + n'\sigma^2_b + 2\sigma^2_c + \sigma^2_d$
Among populations/ within groups	P-G	SSD (AP/WG)	$n\sigma^2_b + 2\sigma^2_c + \sigma^2_d$
Among individuals/ within populations	N-P	SSD (AI/WP)	$2\sigma^2_c + \sigma^2_d$
Within individuals	N	SSD (WI)	σ^2_d

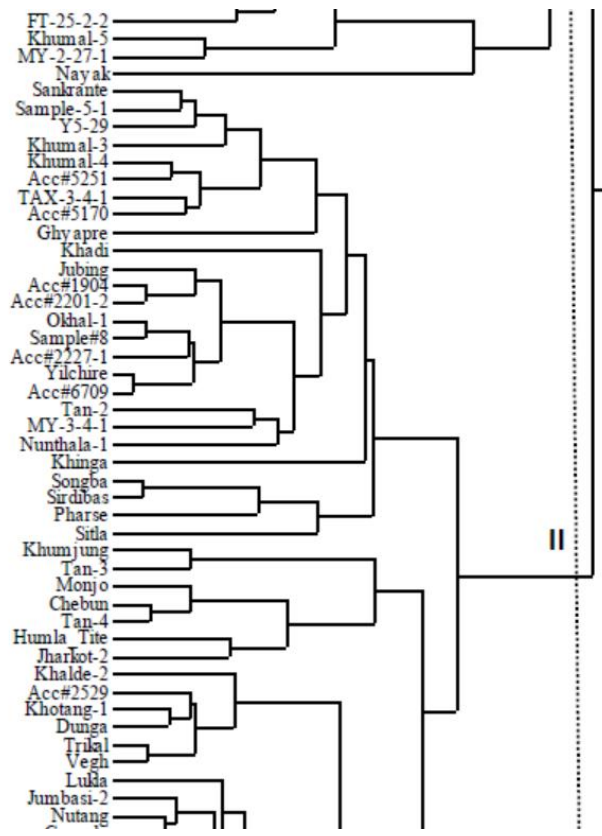


Multivariate Analysis and Multiple Regression/ nonlinear reg

1. Cluster analysis: Landraces study based on DNA markers

2. Principal Component Analysis

Many others

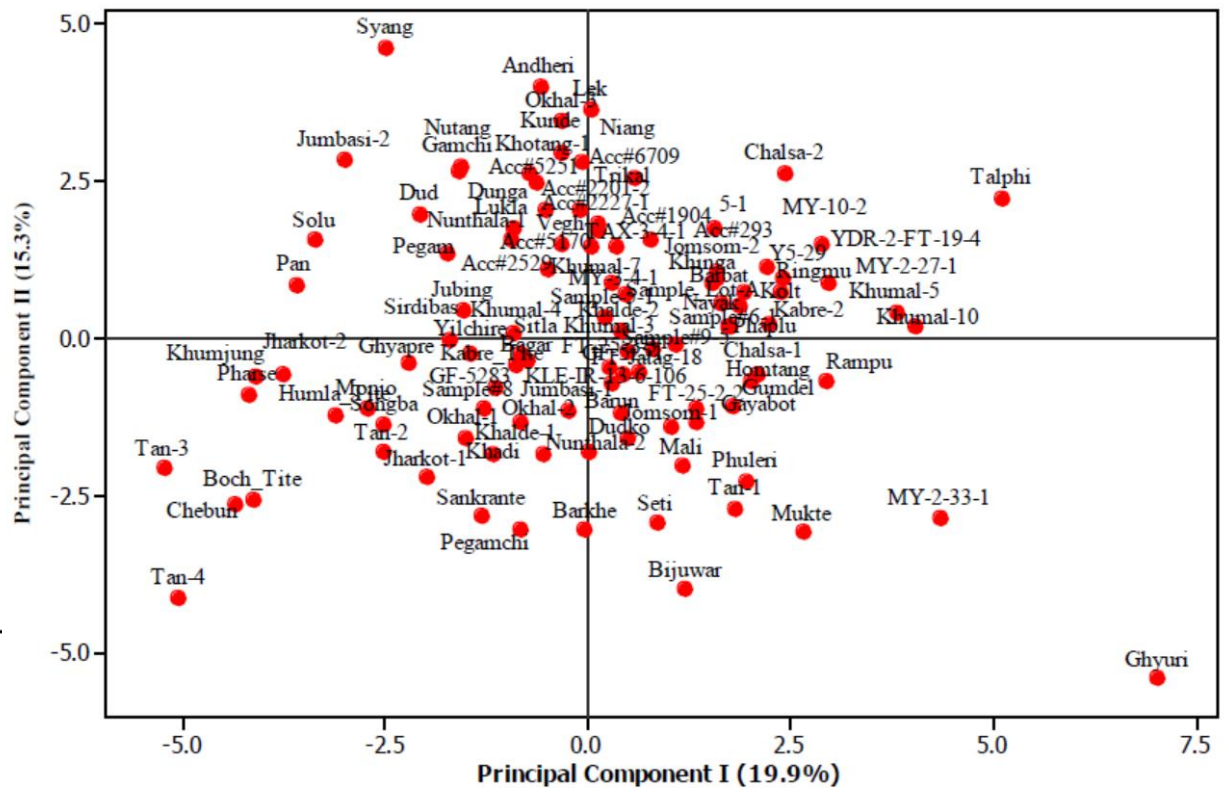


II (53): Leaf number, cyme length, grain/plant, grain yield, straw yield

III (1): Primary branch, leaf number, cyme/plant, flowering days, grain/plant, grain yield, seed width; Seed/cyme, seed/cluster

IV (1): Cyme length, seed/cyme, seed/cluster, unfilled grain, straw yield; Primary branch, flowering days

V (2): Plant height, cluster/plant

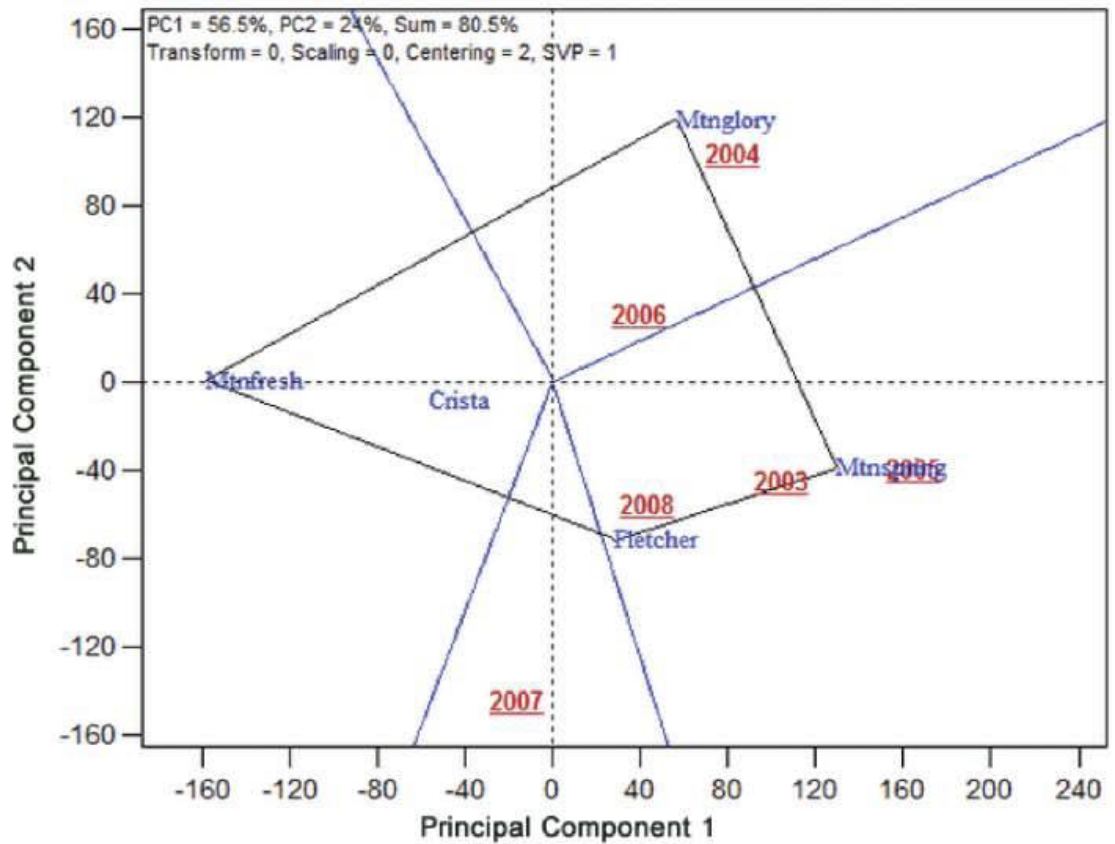
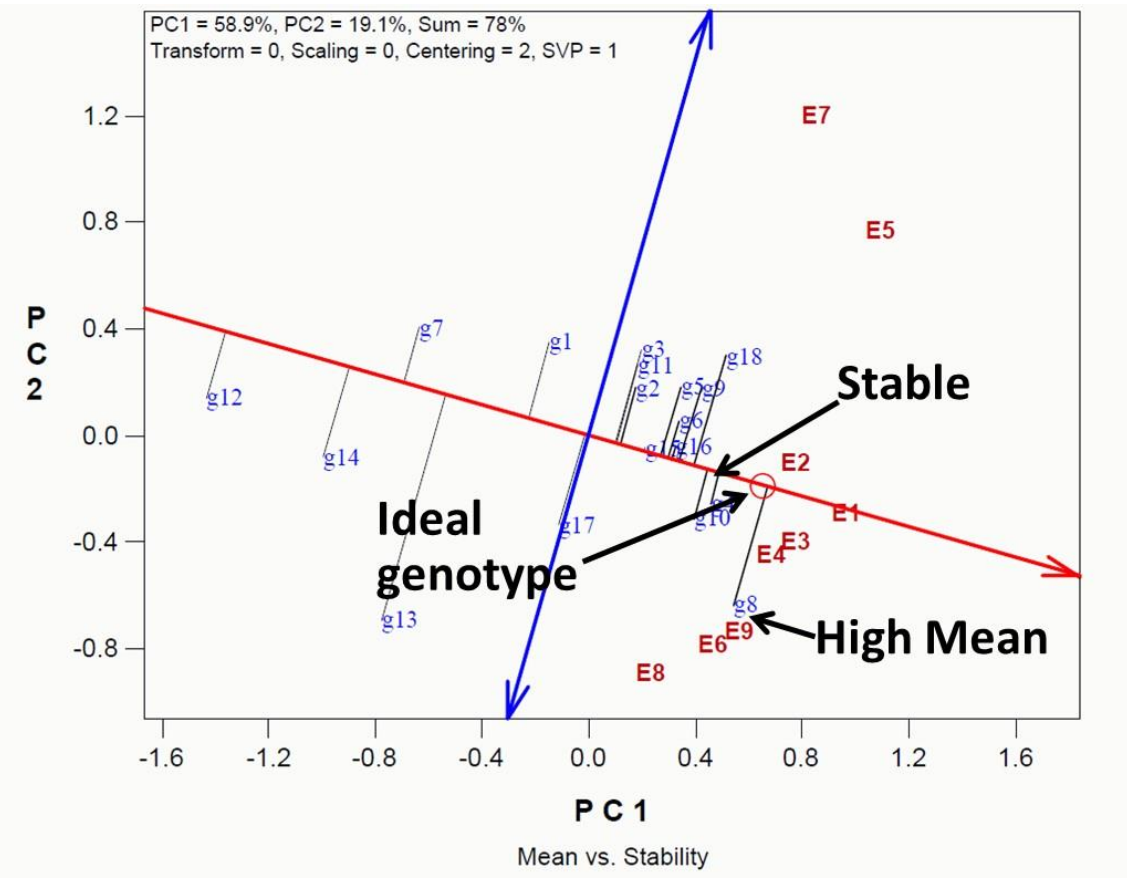


Values of collinearity increasing



Site Specific Variety: GGE Biplot (not a univariate)

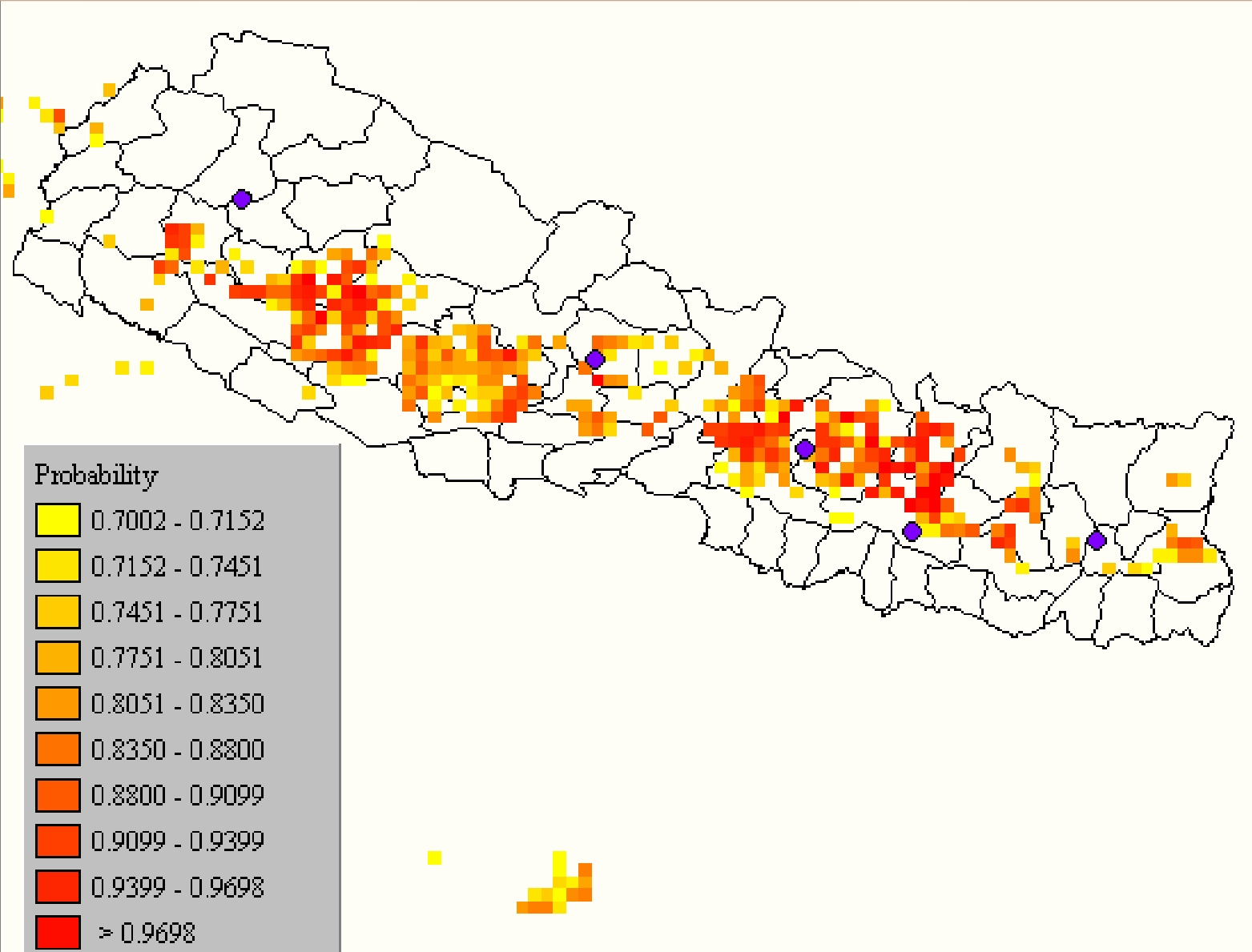
- Not average performers, go for maximum yielder
- Repeat experiment over the years on same site





Analog site based Recommendations of Technology (climate smart tech)

- Current climate data from Worldclim (representative of 1950-2000, 19 bioclimatic variables): <http://www.worldclim.org>
- Climate (current and future) data: www.diva-gis.org/climate, <http://gisweb.ciat.cgiar.org/GCMPPage/>
- GIS Data (country level, global climate): www.diva-gis.org/data





Multidisciplinary Design and Multidisciplinary Multivariate Analysis?

Why this treatment is best?
Reason at molecular genetics



Post experimental stage



Data Entry and Saving in Excel

- Check data before, during and after entering in the Excel
- **Freezing** or splitting panes (under view menu)/ Filter / Sort
- **Validation:** Data is rejected from the system at entry
 - **Drop down** lists (for repeating list of text string): Data – Validation – List
 - **Data validation** (for range checks for numerical data): Data – Validation – Allow
 - Decimal – Minimum and maximum
 - Adding **comments** to cells (for any unusual observations or questions): Review – New comment

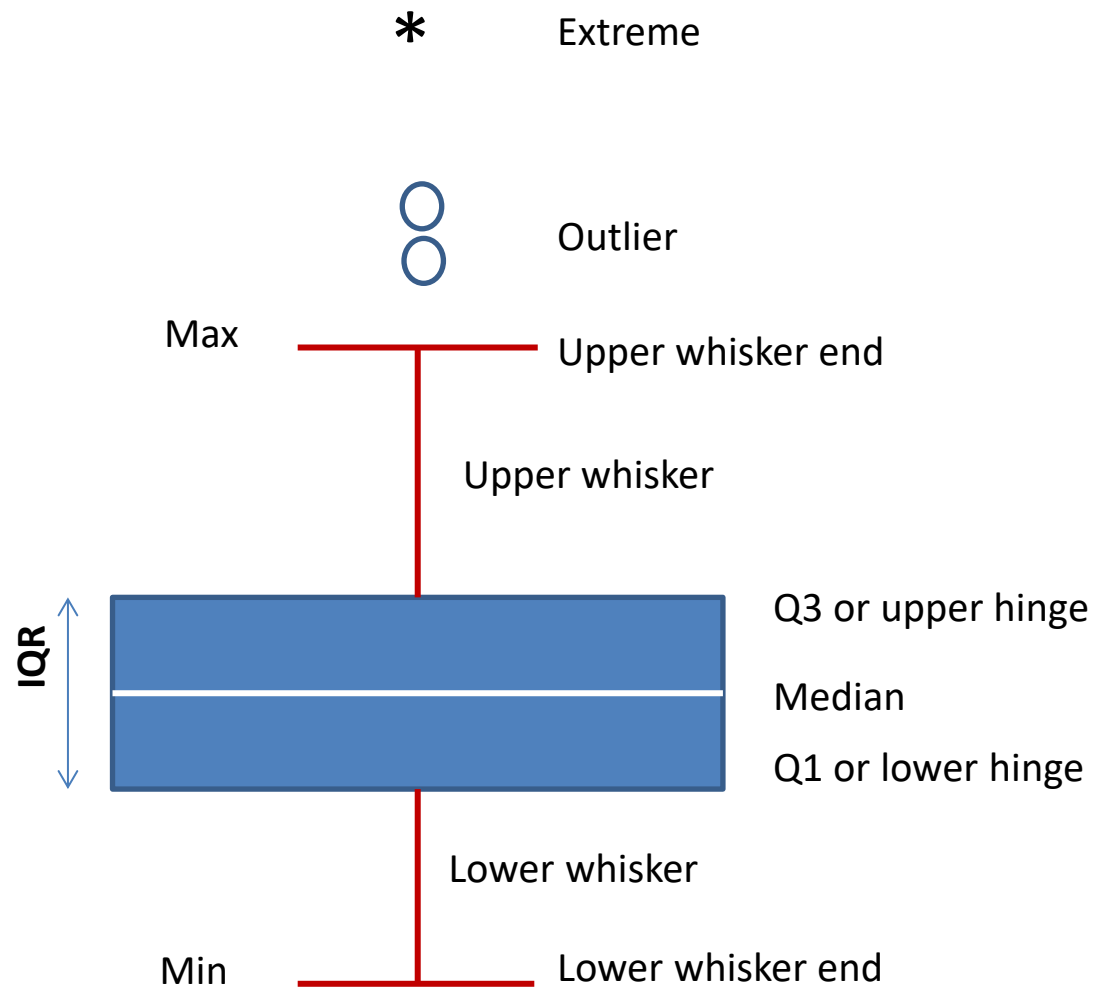


Exploratory data analysis (EDA): Data validation and assumption test

- Getting to know your data
- Quality control for mistakes in data entry
- Distribution assumption testing
- Finding outliers
- Missing data?

- Command menu
 - *Analyze – descriptive statistics – explore*
- Histogram/ Normal Q-Q plot/box plot/Shapiro-Wilk test

- Relate data value with visual observation and perception



Box plot



Software Data Analysis and R Packages

- **R and Rstudio:** For statistical computing and graphics
- **Past:** for data analysis, with functions for data manipulation, plotting, univariate and multivariate statistics,
- **Minitab:** Simple and general statistical package
- **SPSS:** Statistical Package for the Social Sciences
- **SAS:** A statistical software for data management, advanced analytics, multivariate analysis,
- **DivaGIS:** For mapping and geographic data analysis (a geographic information system (GIS))
- **STAR:** Statistical tool for agricultural research
- **NTSYS:** Numerical taxonomy and multivariate analysis system
- **Genstat:** A statistical software package with data analysis capabilities, particularly in the field of agriculture
- **Agrobase:** Data management, experiment management, and statistical analysis for plant breeders and agronomists
- **Statistica:** A data analysis and visualization program
- **MSTATC:** For the design, management, and analysis of agronomic research experiments
- **Instat:** A general statistical package
- **GGEBiplot:** For biplot analysis, conventional analysis, and decision making based on univariate and multivariate data
- **vegan:** For community ecologists with multivariate and diversity analysis and other functions
- **agricolae:** Statistical Procedures for Agricultural Research, offers extensive functionality on experimental design especially for agricultural and plant breeding experiments and other statistical analysis
- **BIO-R:** Biodiversity analysis using molecular data















Online statistical tools

- Statistical software: <https://stats.blue/>
- DATAtab is a web-based statistics software: <https://datatab.net/>
- Statistics online: <https://www.statskingdom.com/>
- For t value: <http://www.ttable.org/student-t-value-calculator.html>
 - <https://goodcalculators.com/student-t-value-calculator/>
- Free tools: <https://www.duplichecker.com/free-tools.php>

Free Tools

Text Analysis Tools

 Plagiarism Checker	 Paraphrasing Tool	 Grammar Checker
 Word Counter	 Spell Check	 MD5 Generator
 Change Text Case	 Merge Words	 Image To Text
 Translate English To English	 Text to Speech	 Reverse Text Generator



Software for Reference Management and Style

- Reference manager
- Software tool for publishing and managing bibliographies, citations and references
 - Zotero (<https://www.zotero.org/>) **NARC Journal Module: NARC Style**
 - Mendeley (<https://www.mendeley.com/>)
 - Refworks (www.refworks.com)
 - Endnote (<http://endnote.com/>), etc
- Add in for Word
- NARC own module in Zotero



Google Docs: Edit together in real-time

- An online word processor, free
- Edit together in real-time with easy sharing, and use comments, suggestions, and action items to keep things moving. <https://www.google.com/docs/>

Grammarly

- Typing assistant that reviews spelling, grammar, punctuation, clarity, engagement, and delivery mistakes in English language text. It also detects plagiarism. It uses artificial intelligence to identify and search for an appropriate replacement for the located error.
- Get suggestions from Grammarly while you write or after write in MS Word
- <https://www.grammarly.com/>



Plagiarism

- The practice of taking someone else's work or ideas and passing them off as one's own
- What is plagiarism: <http://www.plagiarism.org/plagiarism-101/what-is-plagiarism/>
- Plagiarism checker free online: <https://www.duplichecker.com/>
- iThenticate: <https://www.ithenticate.com/>
- Plagiarism checker by Grammarly: https://www.grammarly.com/plagiarism-checker?q=plagiarism&utm_source
- **Research Square**: a multidisciplinary preprint and author services platform. share your work early in the form of a preprint, gain feedback from the community, and use our tools and services to improve your paper



Image editing soft

PhoXo

- A free and powerful image editing software. It's tiny, fast, easy to use, can be treated as a mini PhotoShop
- For labeling: Insert name in image and give file name by accession number or collection number
- <http://www.phoxo.com/en/>

Adobe photoshop

- <https://www.adobe.com/products/photoshop.html>

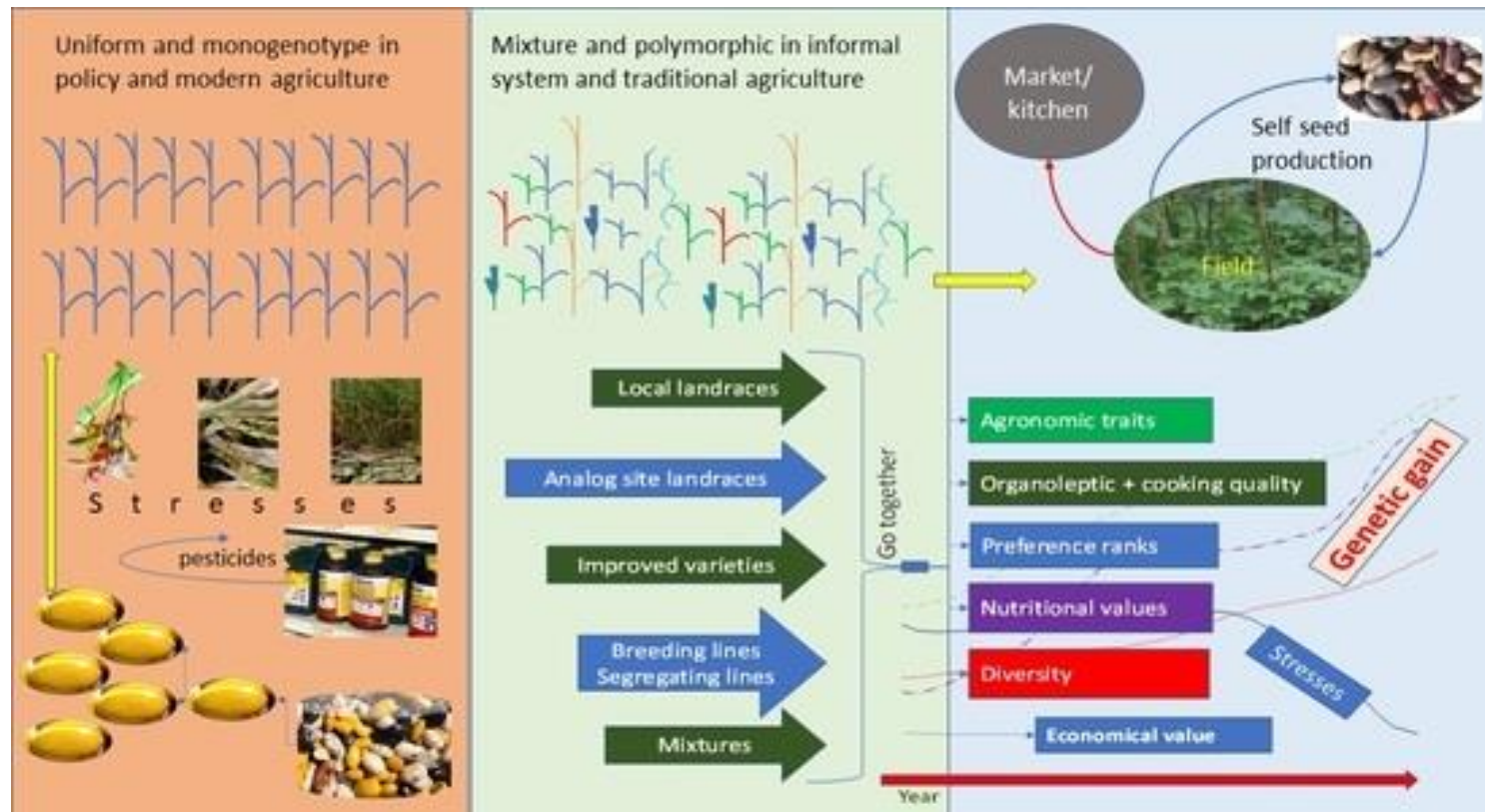


Online Submission and Graphical Abstract

- Nepal Journals Online (NepJOL) is a service to provide online publication of Nepalese journals
- **Graphical abstract:** Single, concise, pictorial and visual summary of the main findings of the article

- PowerPoint
- Inkscape
- Adobe Illustrator

Diversity Journal MDPI: Approaches and Advantages of Increased Crop Genetic Diversity in the Fields





General Mistakes and Errors...

- Studied genotypes and research findings not available for ever (**accessioning system** not in practice)
- **Materials not listed**, method not described (Methodology not complete, objectives to expected outputs)
- **Local check**: Just Local (not available later)/ national/ global check?
- Nepali Devanagari: kaa, aa; Dhaan
- **Translating Nepalese noun** (Bitter for Tite, Local for Sthaniya)
- **Year**: 2014/15; In Nepali calendar (2078... in English proposal)
- Not same materials in repeated experiments
- Poor data from collaborator, outlier and then manipulate



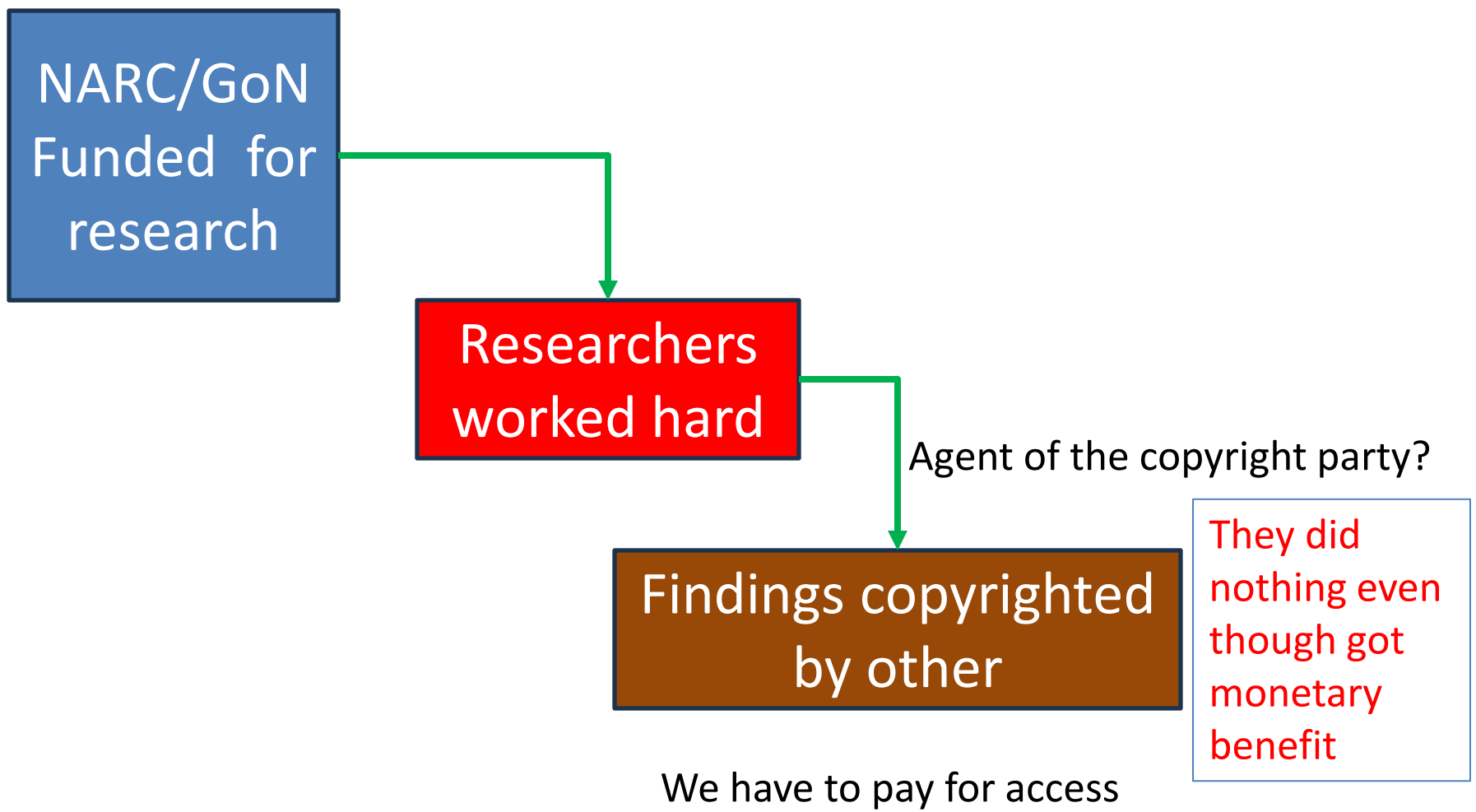
General issues other than proposal writing

- Experiment गर्दा ध्यान नदिने पछि गएर Research results वा Data लाइ Manipulation गर्ने
- Local check with higher yield, then manipulate data
- **Blocking** without any criteria and Limited experimental plot
- **Different researchers/ laborers** in a single (same) experiment
- **Data and layout by junior/ labor.** Researchers visit to field: Very low
- **Experiment in one design**, analysis in another (alpha, split plot –RCBD)
- Experiment in small plot and implication turn to larger (sampling error)
- Limited knowledge on sampling methods (materials from outside country)
- **No data checking and exploratory analysis** (checking for assumption)
- **Not suitable machine** for data recording (hand written vs electronic record)
- **Research findings always perceived positive** (if not significant, then tried to have it)
- Research findings not in academic courses

SAS-Nepal: Research quality assessment and publication ranking



Where goes the benefit?





Thank you all

Special thanks to

- SAS-Nepal
- Dr Shova Shrestha
- Dr Ram Ghimire
- Dr Krishna Timisina

Many missing?

Let's think agricultural biodiversity in agricultural problems first.

Let's make native technologies statistically significant