EFFECT OF FIELD MANAGEMENT METHODOLOGIES ON RICE YIELD AND NITROGEN USE EFFICIENCY

PRELIMINARY DATA ANALYSIS

APRIL 7, 2022

JAYA BIJESH

COSC593

Table of Contents

Introduction

The aim of this project is to analyse the effect of management strategies such as plant population, sowing methods and fertilizer application rates on rice yield, in addition to studying Nitrogen use efficiency. A detailed graphical and numerical analysis of the various variables in the data set has been conducted with a view to preliminarily determine the behavior of response variables in relation to the predictors, prior to building models.

Analysis of variables

The data set used in this analysis has variables corresponding to rice cultivation metrics over a period of 6 rice growing seasons. It has 1778 observations of 23 variables each. A brief description of the variables is given in Table 1 below:

Table 1: Description of data set headers

Sl. No.	Variable Name	Description
1	Ref	Reference number
2	ExpSite	Name of the site/farmer
3	Year	Rice sown in October and harvested the following year.
4	ExpName	Combination of experiment site and year
5	SowingMethod	Four sowing methods – Drill, Aerial, Delayed Permanent Water (DPW) and DryBC
6	Range	Replicate identification number of the experiment (each treatment has 3 replicates)
7	Row	Row identification number (1 - 24)
8	Rep	Same as Range
9	Variety	Variety of rice
10	Nitrogen	Nitrogen rates (kg N/ha) applied first prior to Permanent Water (PW) and then at Panicle Initiation (PI)
11	NGroup	Groups representing different Nitrogen rates as above
12	Estab	Plant population/Count at establishment (No./m^2)
13	PI_DM	Weight of Dry Matter at PI (g/m^2)
14	PI_Tillers	Count of tillers at PI (No./m^2)
15	PI_NIR	Percentage of Nitrogen in the tissue at PI (% N)

16	PI_N_Uptake	Nitrogen uptake at PI (kg N/ha)
17	PM_DM	Weight of DM at Physiological Maturity (PM) (g/m^2)
18	HarvTillers	Count of tillers at PM (No./m^2)
19	HI	Harvest Index - ratio of grain to total plant dry weight
20	QuadYield	Quad (hand harvested) grain factor (tonnes per hectare at 14% moisture)
21	HeaderYield	Header (machine harvested) grain factor (tonnes per hectare at 14% moisture)
22	Lodging	Score given at harvest to indicate if a crop/plot is standing or lodged (1 = standing, 10 = fully lodged)
23	PlantHeight	Average plant height at harvest (cm)

Figure 1 below shows the response and explanatory variables identified from the data set.
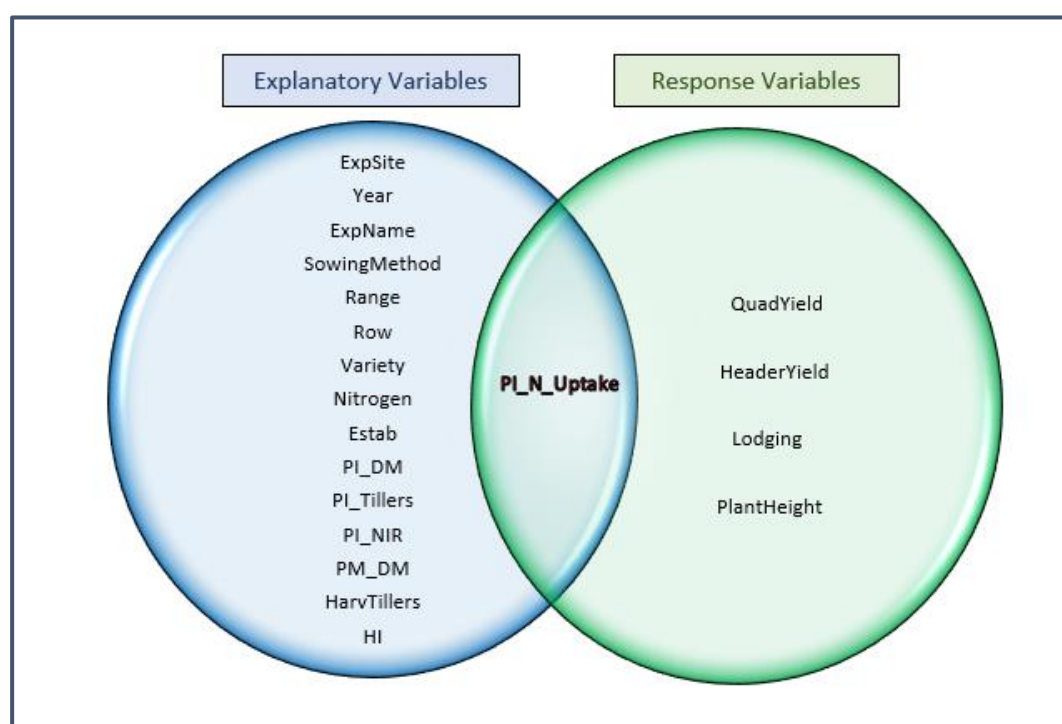

Figure 1: Response and Predictor Variables

Nitrogen Uptake at PI would be a predictor for when determining the effect of management practices on rice yield, while itself being a response variable in assessing Nitrogen use efficiency.

Data Cleansing

As a primary step in data analysis, some of the data cleansing methods that were undertaken are as follows:

- Renamed columns as above.
- Selected only required columns for analysis.
- Dropped rows with missing QuadYield values.
- Trimmed whitespaces from factor variables (e.g., 'Drill' and 'Drill ' were two levels of SowingMethod prior to trimming).
- Removed 2 rows from analysis (a duplicate header row and a blank row from merging the 2020-'21 data into the original data set)
- All categorical variables were converted to factors and any discrepancies in factor levels owing to typos were removed (e.g., In rice varieties, 99 observations had VO37 incorrectly recorded instead of V037, thus making them two different levels).

In the data set, all variables were of type character initially. Each of them was converted to numerical and categorical values accordingly. The Ref column was excluded from the analysis while plotting, as it is only a means of referencing and not a predictor. Similarly, the Range and Rep variables are the same and hence either one of it could be excluded from the analysis.

For Nitrogen rates, many of the observations had been recorded in the format XX as opposed to XX-XX followed by most observations. To examine if XX corresponded to XX-00 implying that Nitrogen was applied only at PW and not at PI, the equivalence of XX and XX-00 was investigated in R by means of a contingency table between Nitrogen rates (Nitrogen) and their groups (NGroup). Their equivalence was evident from the contingency table and hence duplication was avoided to reduce the factor levels of Nitrogen rates by replacing all XX values with XX-00 consistently.

Preliminary analysis of categorical variables

The frequency of each of the categorical variable levels were determined, out of which the plots of NGroup (Figure 2), Nitrogen (Figure 3), SowingMethod (Figure 4), and Variety (Figure 5) are shown below:
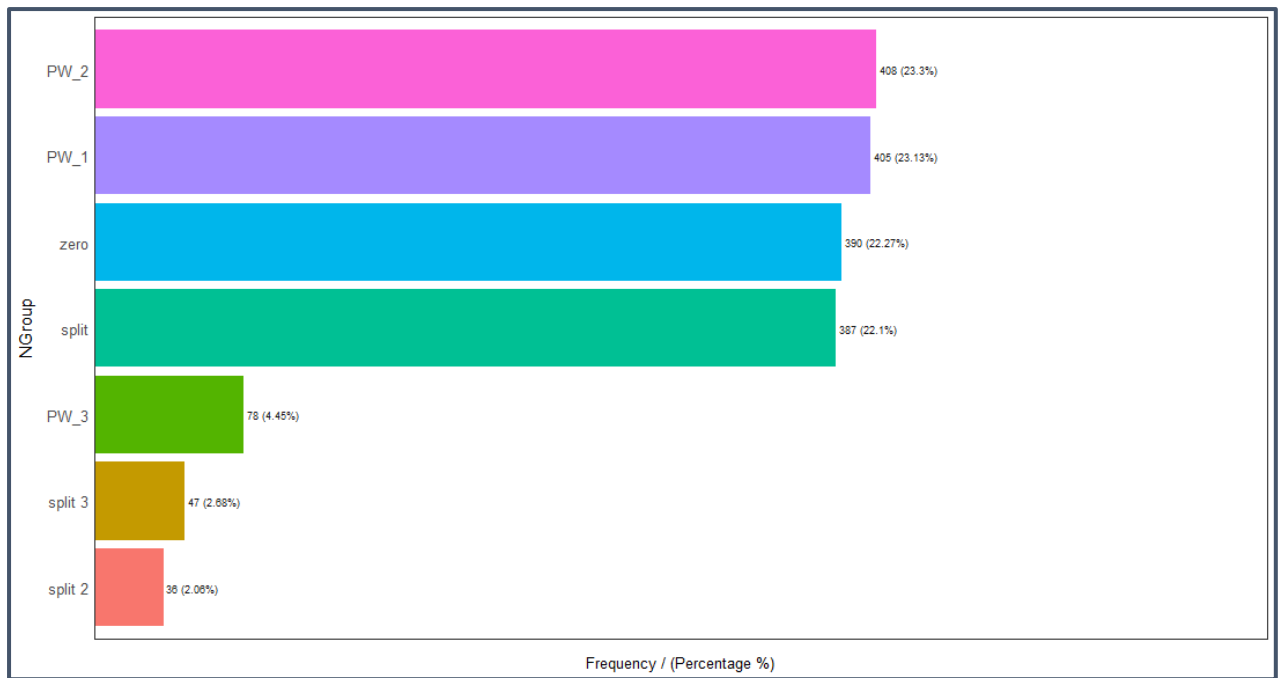
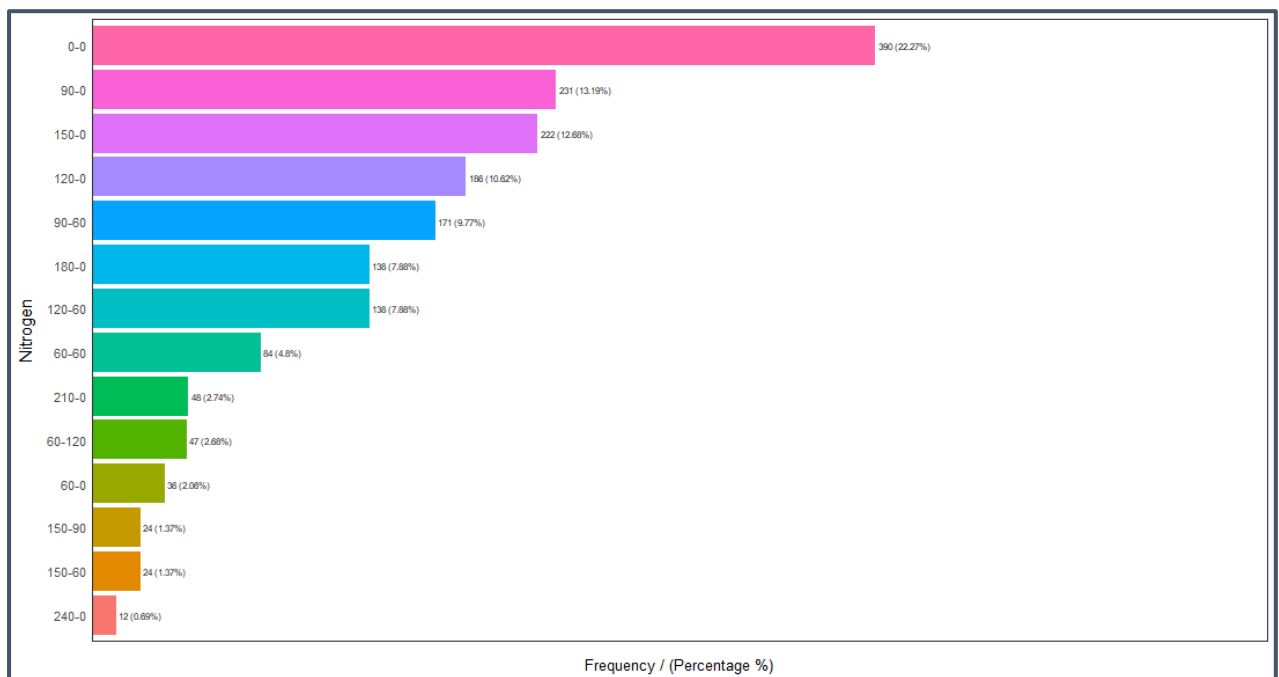Figure 2: Frequency of NGroup in the data set

Figure 3: Frequency of Nitrogen rates in the data set
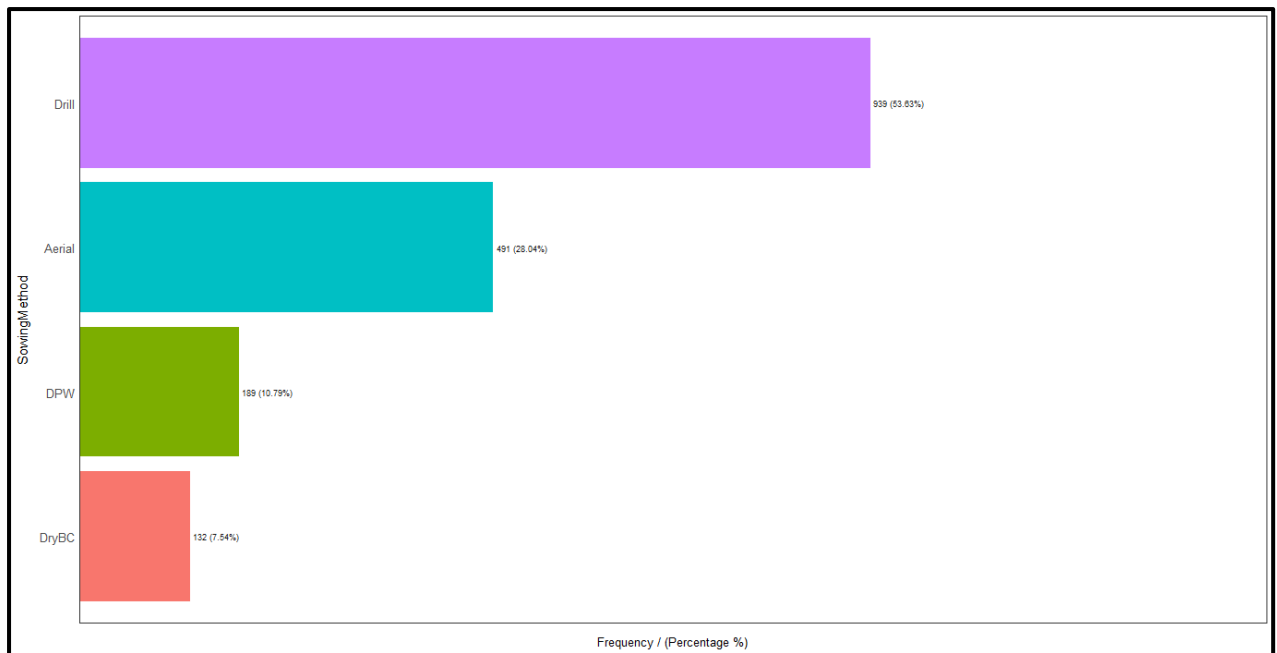
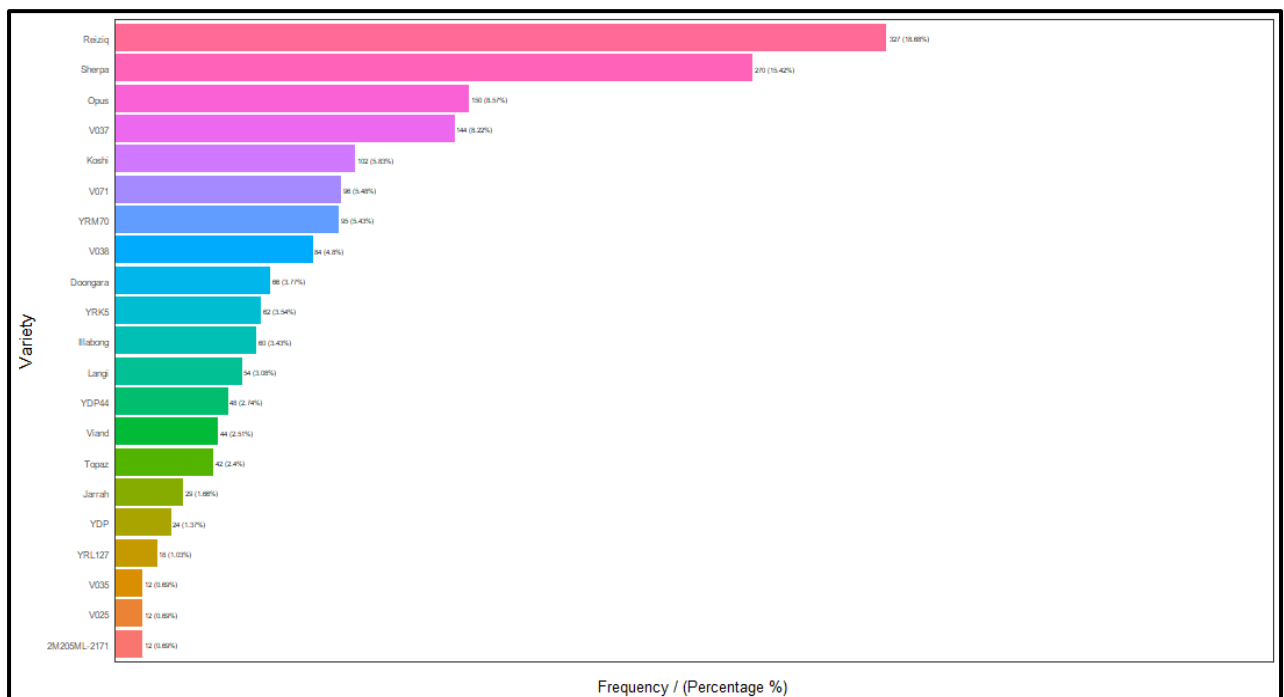Figure 4: Frequency of Sowing Methods in the data set

Figure 5: Frequency of Rice Variety in the data set

The following points were summarized for categorical predictors:

- The highest frequency of Nitrogen rate in the data set is 0-0, accounting for 22.27%, with the least applied Nitrogen rate being 240-0 (0.69%).

- The Reiziq variety of rice has the highest frequency (18.68%), and the lowest number of observations is for 2M205ML- 2171 (0.69%). Reiziq has been set as the base variety to which other rice varieties will be compared to.

- Drill-sown method is the most frequently recorded sowing method in the data set (53.63%), followed by Aerial (28.04%), DPW (10.79%) and DryBC (7.54%).

- In NGroup variable, zero, split, PW_1 and PW_2 have similar frequencies (22- 23%), while the frequency of PW_3, split 3 and split 2 methods together constitute less than 10%.

Correlation analysis

Pairs plot was generated as shown in figure below for all the numerical variables in the data set.
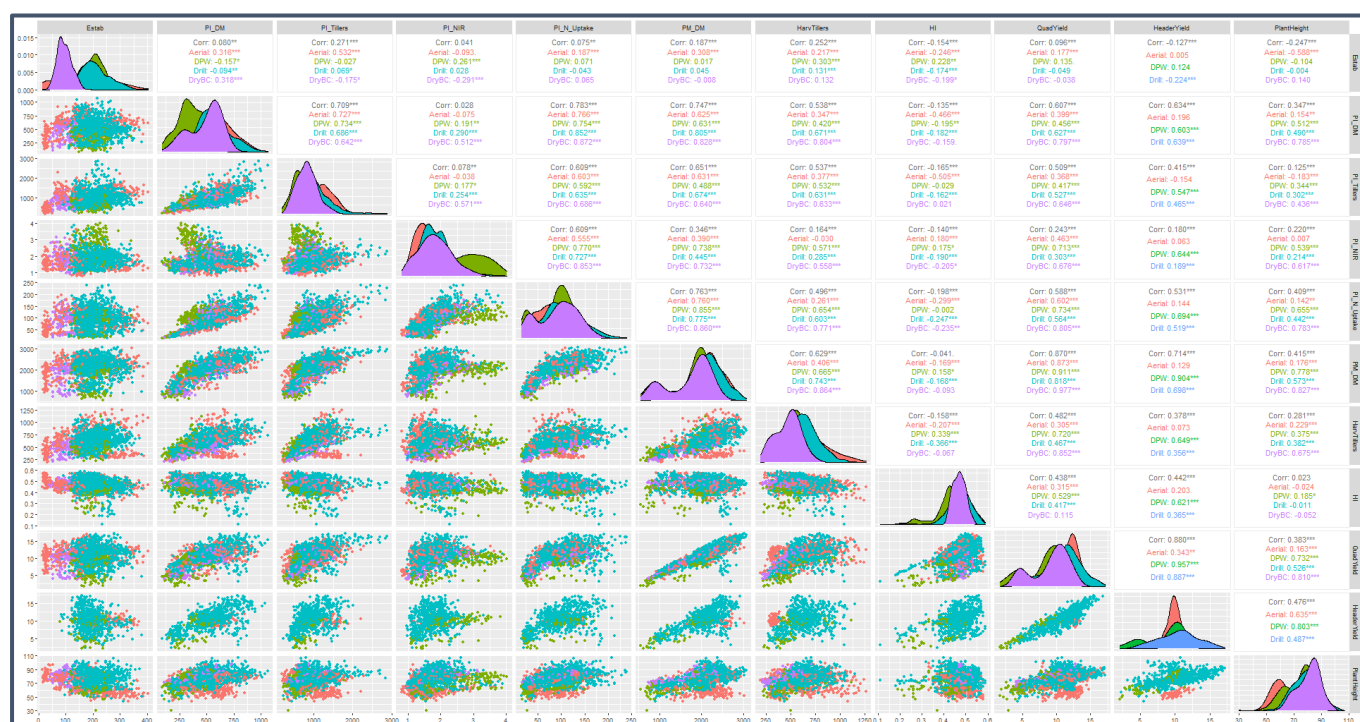

Figure 6: Pairs Plot

In pairs plot, the response variables QuadYield and HeaderYield exhibited a strong positive correlation of 0.88. The correlation between these two response variables was higher for DPW Sown (0.96) and followed by Drill Sown rice (0.89). The other pairs of response variables showed only a weak correlation amongst them.

Other Response-Explanatory variable pairs that had moderate to strong correlation are set out in Table 2 below:

Table 2: Correlation Coefficients

Sl. No.	Response Variable	Explanatory Variable	Correlation Coefficient
1	QuadYield	PI_DM	0.61
		PI_Tillers	0.51
		PI_N_Uptake	0.59
		PM_DM	0.87
		HarvTillers	0.48
		HI	0.44
2	HeaderYield	PI_DM	0.63
		PI_Tillers	0.42
		PI_N_Uptake	0.53
		PM_DM	0.71
		HI	0.44
3	PlantHeight	PI_N_Uptake	0.41
		PM_DM	0.42

The correlation coefficients for response variable pairs are shown in Table 3 below:

Table 3: Correlation Coefficients (Response Variables)

Sl. No.	Variable 1	Variable 2	Correlation Coefficient
1	QuadYield	HeaderYield	0.88
2	QuadYield	PlantHeight	0.38
3	HeaderYield	PlantHeight	0.48

➕ Strong positive correlation between QuadYield and HeaderYield.

➕ PlantHeight is moderately positively correlated with QuadYield and HeaderYield.

Summaries of response variables

1. QuadYield

QuadYield (hand harvested grain factor) is the primary response variable in this analysis. The numerical summary of QuadYield is in Table 4 below:

Table 4: Numerical Summary of QuadYield

Numerical Summary	
Mean	10.66
Median	11.11
Max	17.16
Min	1.7

The relationship of QuadYield with numerical and categorical predictors have been analyzed and summaries generated as follows:
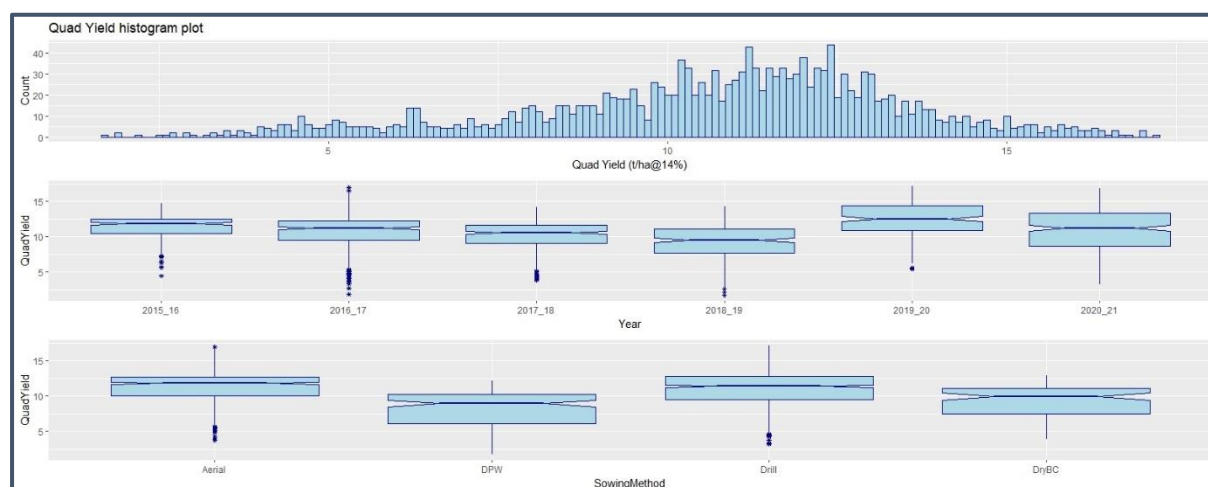

Figure 7: QuadYield plots

The top plot in Figure 7 shows the histogram distribution of QuadYield. The middle QuadYield-Year boxplot shows that the median QuadYield decreased from 2015-'16 to 2018-'19, with the highest median QuadYield in 2019-'20 (12.6 t/ha). Similarly, the QuadYield-SowingMethod boxplot indicates that the median yield is higher for aerial sown rice compared to other sowing methods.
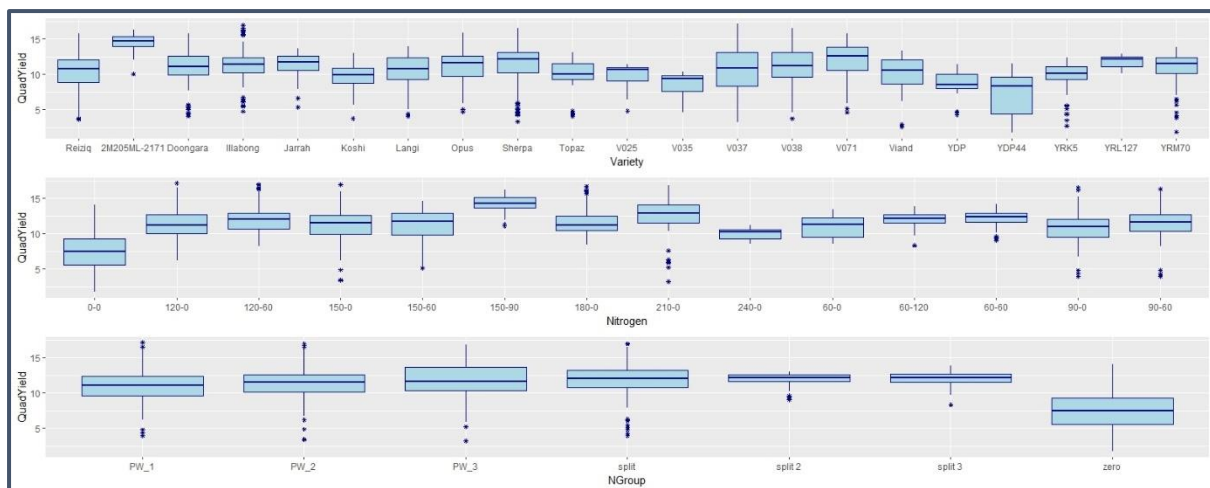
Figure 8: QuadYield Boxplots

The QuadYield-Variety boxplot in Figure 8 shows that the mean QuadYield (approximately 14.8) is higher in 2M205ML-2171 compared to other varieties with very little variability. The lowest QuadYield in the analysis period is in YDP44 variety, also with the most variance.

The middle plot in Figure 8 corresponds to the relationship between yield and Nitrogen rates. The highest yield has been consistently produced by 150-90 nitrogen rates compared to other application rates. The yield is the lowest in the absence of fertilizer application (0-0). The boxplots show that the application of fertilizer at PW and then at PI have a greater yield produce compared to single application of fertilizer at PW only.

The bottom boxplot in Figure 8 shows the effect of Nitrogen groups on QuadYield. The split nitrogen rates have higher median yields compared to PW groups. The least yield is when no fertilizer is applied.
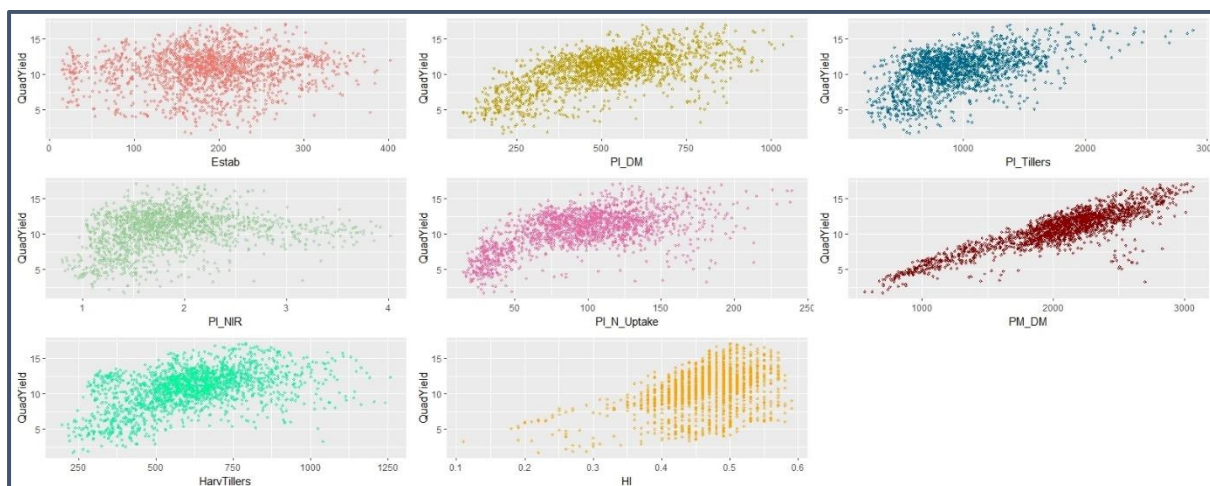

Figure 9: Scatter plots

The scatterplot grid shown in Figure 9 above shows that QuadYield has moderate to strong correlation with its predictors. The plant biomass sample taken at harvest (PM_DM) shows strong positive correlation with yield. The correlation coefficients are shown in table below:

Table 5: QuadYield- Correlation Coefficients

Explanatory Variable	Correlation Coefficient
PI_DM	0.61
PI_Tillers	0.51
PI_N_Uptake	0.59
PM_DM	0.87
HarvTillers	0.48
HI	0.44

The figures 10-13 below are facet plots showing the summary of the effect of

- Nitrogen rates per sowing method – the highest yield has been produced by drill-sown rice at 150-90 nitrogen rates closely followed by aerial sown at 150-90 nitrogen application (Figure 10).
- Sowing method per rice variety – the highest median yield of approximately 14.5 t/ha has been produced by the drill-sown Doongara variety of rice (Figure 11).
- Year and population per sowing method- the median yield was highest in the season 2020-'21 for aerial sown rice. Similarly, the highest median yield is for aerial and drill-sown rice for a given plant population (Figure 12).
- Rice variety per year- for each variety of rice across all seasons, the maximum median yield was approximately 14.5 in 2019-'20 for 2M205ML-2171 (Figure 13).
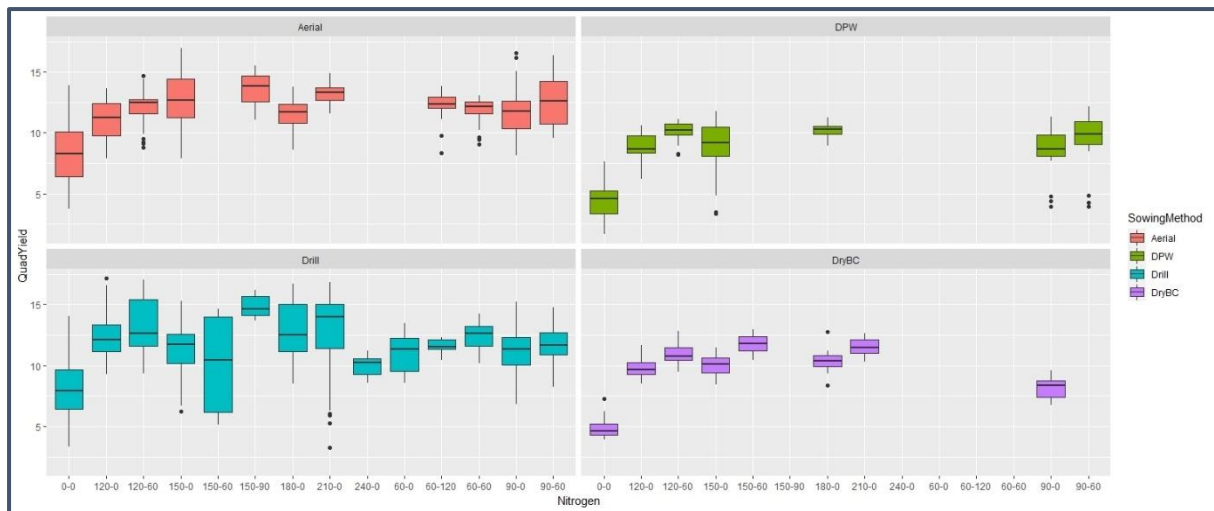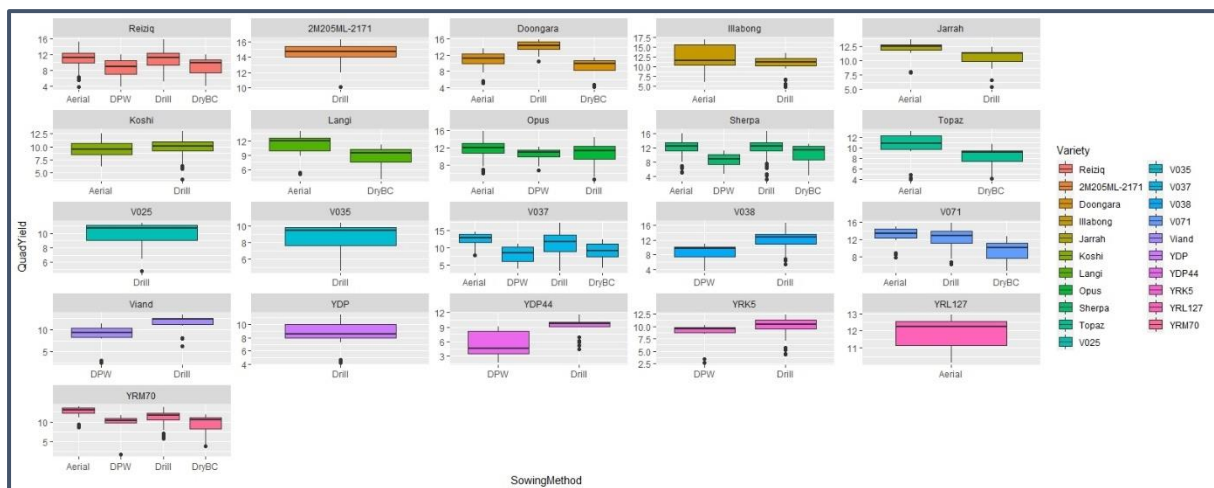
Figure 10: QuadYield-Nitrogen rates facet

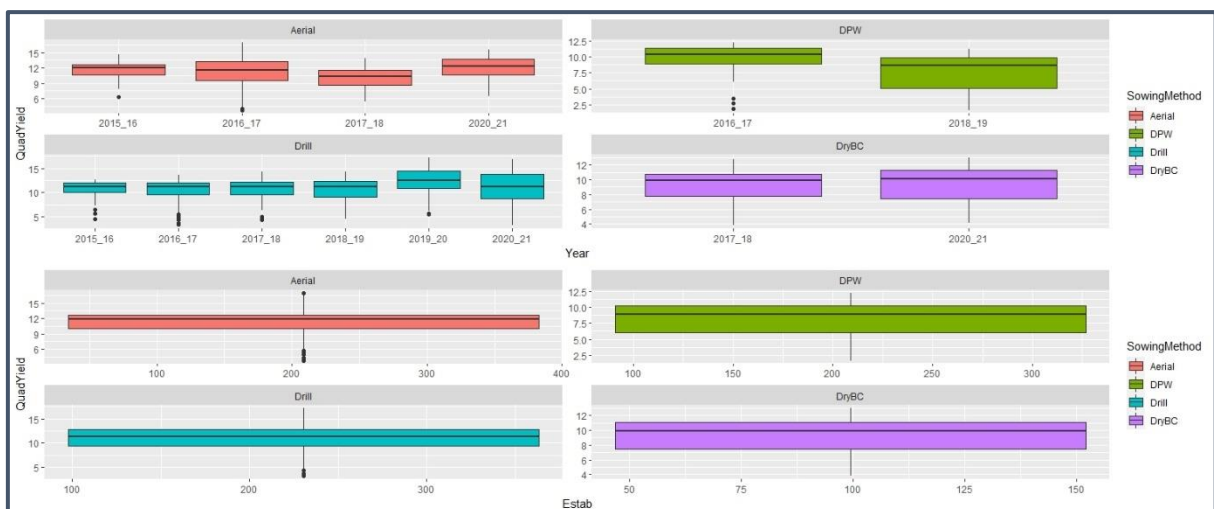Figure 11: QuadYield-Sowing Methods facet
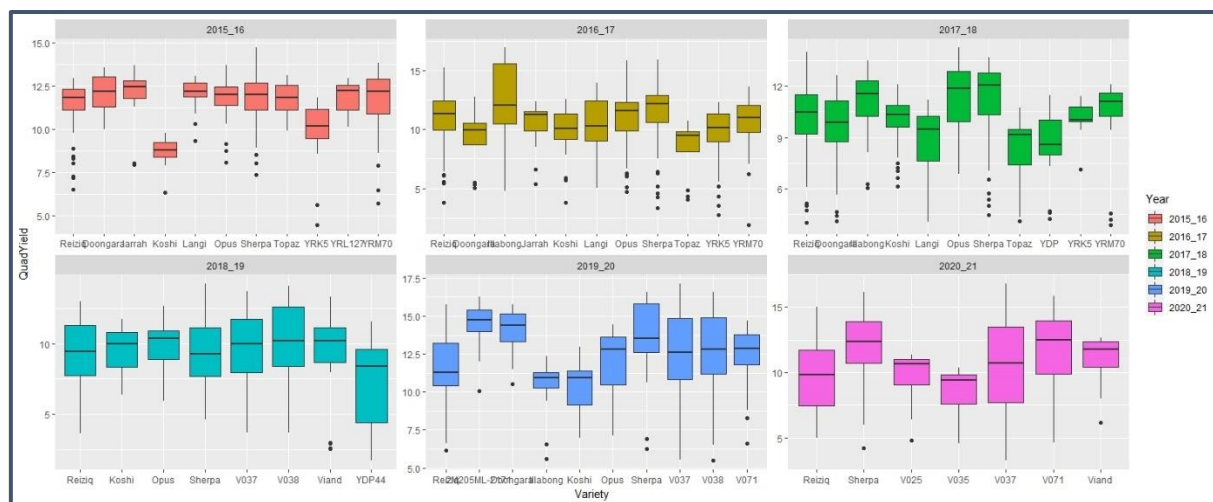

Figure 12: QuadYield-Population facet

Figure 13: QuadYield-Variety facet

2.   HeaderYield

HeaderYield is a measure of machine harvested grains in tonnes/hectare at 14% moisture and is a response variable in this analysis. The numerical summary of HeaderYield is in Table 6 below:

Table 6: Numerical Summary of Header Yield

Numerical Summary	
Mean	10.15
Median	10.32
Max	17.18
Min	2.02

The relationship of HeaderYield with numerical and categorical predictors have been analyzed and summaries generated as follows:
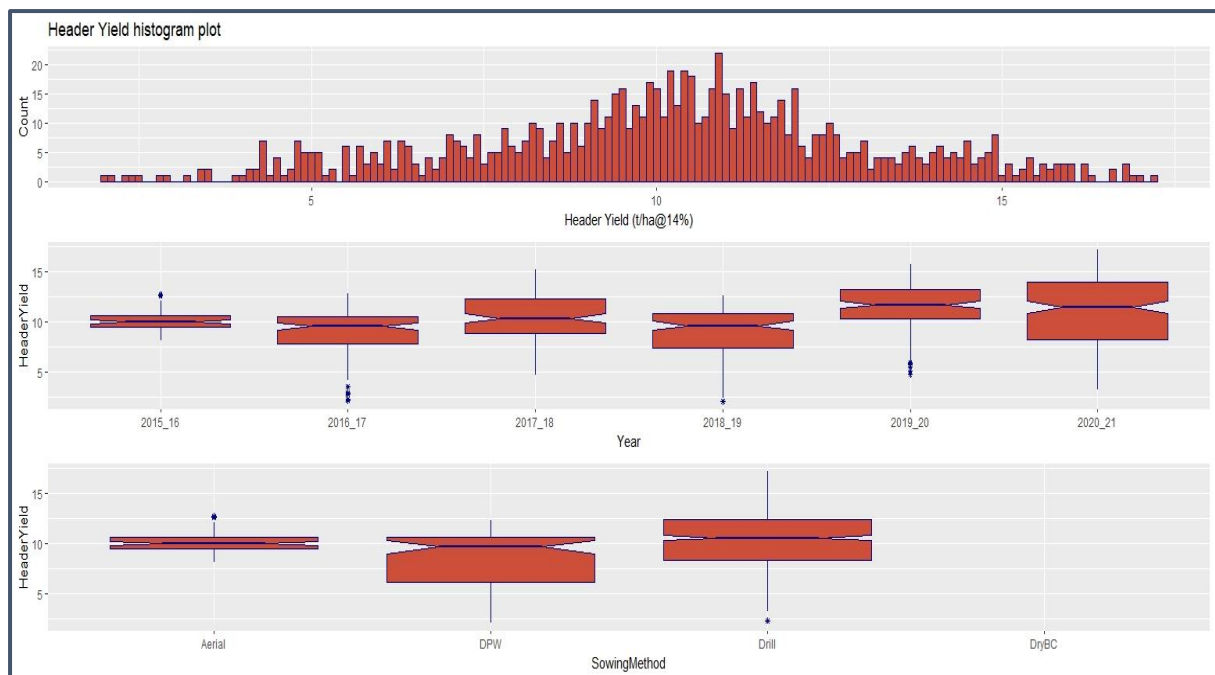
Figure 14: HeaderYield Plots

The top plot in Figure 14 above is the histogram plot showing the distribution of HeaderYield values in the data set. The middle plot shows that there is a considerable increase in the median header yield values for the last two seasons. There is larger variability in the data corresponding to the 2020-'21 season. The bottom boxplot shows that there is not much difference in the median header yield values with respect to the different sowing methods.
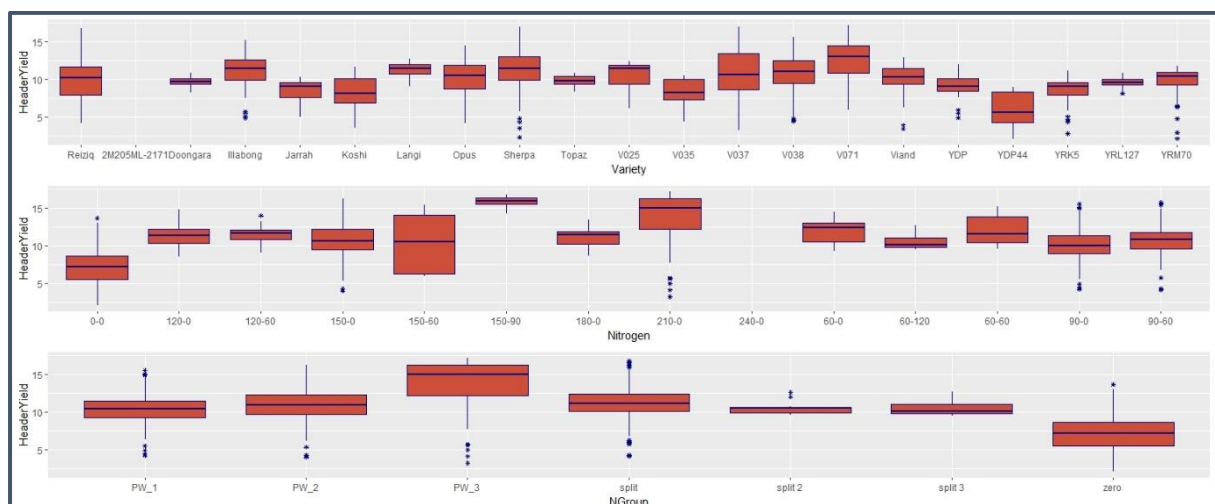

Figure 15: HeaderYield Boxplots

From the above set of boxplots in figure 15, the top plot shows that compared to the Reiziq variety, some varieties of rice such as Illabong, Langi, Opus, Sherpa, V025, V037, V038, V071, Viand and YRM70 have higher yields, whereas the median header yields are lower for the other varieties compared to Reiziq, across all experiments. Similar to QuadYield, the highest median header yield is for 150-90 Nitrogen rates closely followed by 210-0. The bottom plot represents header yield in response to the different Nitrogen application groups. In general, the PW groups have higher median header yields than split groups.
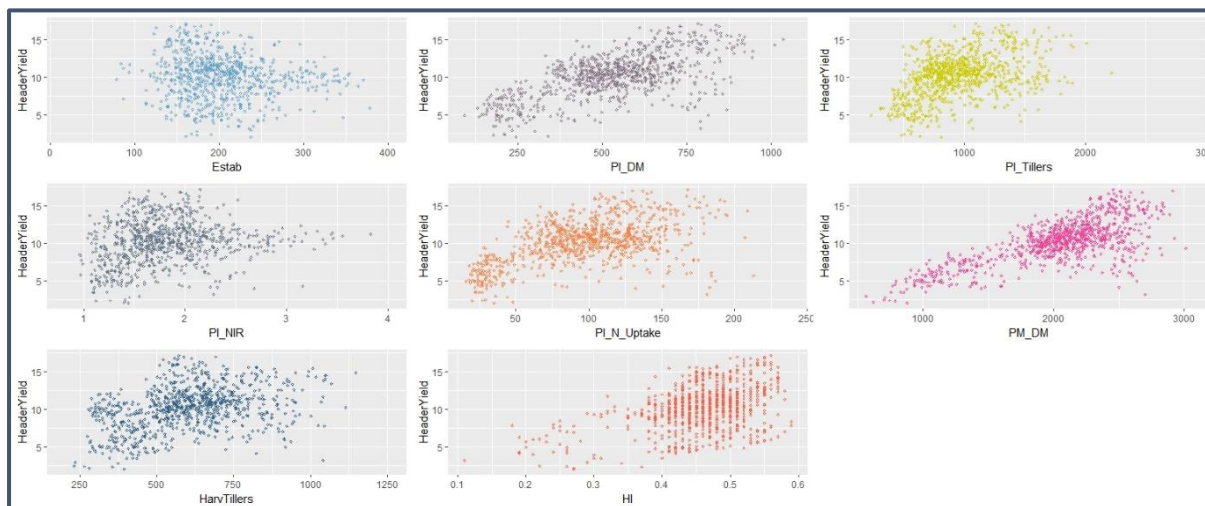

Figure 16: HeaderYield Scatter Plots

The scatter plots in Figure 16 above indicate that header yield has moderate to strong correlation with several predictors, the strongest correlation of 0.71 being with the plant biomass at physiological maturity (PM_DM).

The correlation coefficients are shown in Table 7 below:

Table 7: HeaderYield Correlation Coefficients

Explanatory Variable	Correlation Coefficient
PI_DM	0.63
PI_Tillers	0.42
PI_N_Uptake	0.53
PM_DM	0.71
HI	0.44

Figures 17-19 below shows a series of facet plots depicting the summary of:

🞣 Header yield per season for all sowing methods- of all the sowing methods across all seasons, the drill sown rice in 2019='20 had the highest median header yield.

🞣 Header yield per nitrogen rates for all sowing methods- for aerial sown rice, there is consistence in header yield (median approximately between 8 and 11) for all nitrogen rates. The drill sown rice shows considerable variance for each nitrogen rate application with the most variance in yield for 150-60 rate. The highest median yield was produced by 150-90 application rate with very less variability.

🞣 Header yield per variety of rice for all seasons- for each season, there is considerable variability in rice yields across varieties.

No data was present for header yield for DryBC sowing method in the data set.
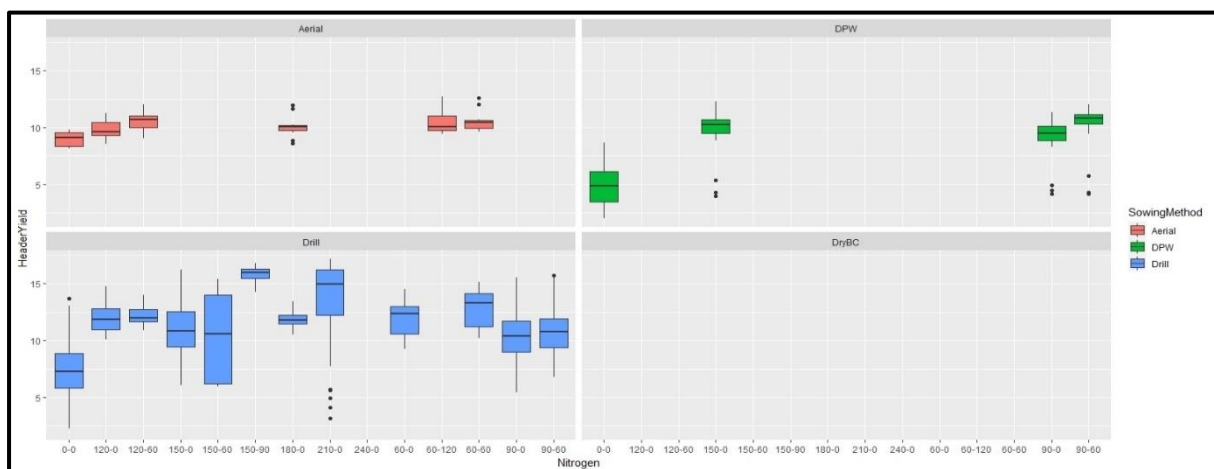

Figure 17: HeaderYield facet plots

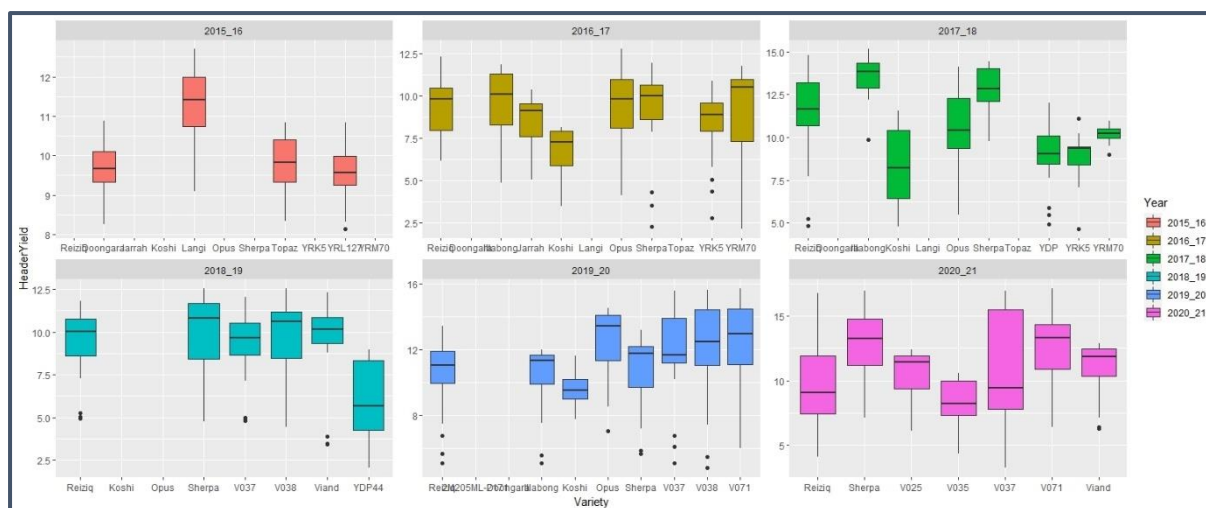Figure 18: HeaderYield – Nitrogen rates facet plot

Figure 19: HeaderYield –Rice Variety facet plot

3.   PI_N_Uptake as a response variable

The nitrogen uptake at Panicle Initiation can be used to determine nitrogen use efficiency. The numerical summary of PI_N_Uptake (kgN/ha) is given in Table 8 below:

Table 8: Nitrogen Uptake Numerical Summary

Numerical Summary	
Mean	95.11
Median	96.00
Max	239.00
Min	15.00

The top plot in Figure 20 below shows that compared to the base variety of Reiziq, the mean uptake of nitrogen at PI was significantly higher in 2M205ML-2171 variety, with other varieties such as Langi, topaz, V037, V038, V071, YRK5, YRL127 and YRM70 also showing high fertilizer intake. The middle plot corresponds to PI_N_Uptake in response to different Nitrogen rates. The highest median uptake is for 240-0 application (PW_3 group) and the lowest is for no application of Nitrogen (zero group). The bottom plot in Figure 20 also demonstrates that the uptake was higher in PW_3 group compared to other nitrogen groups.
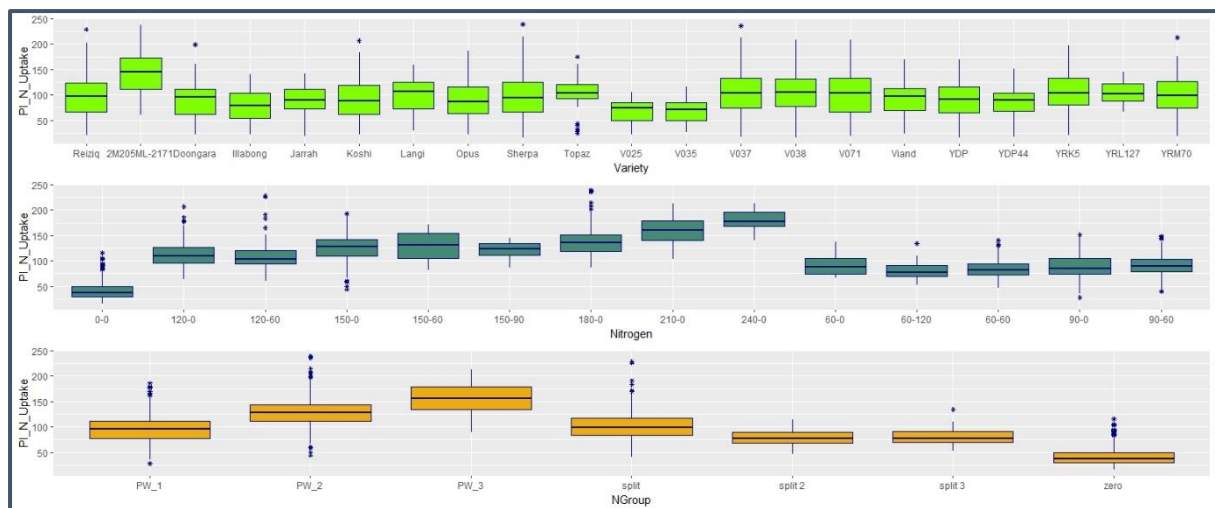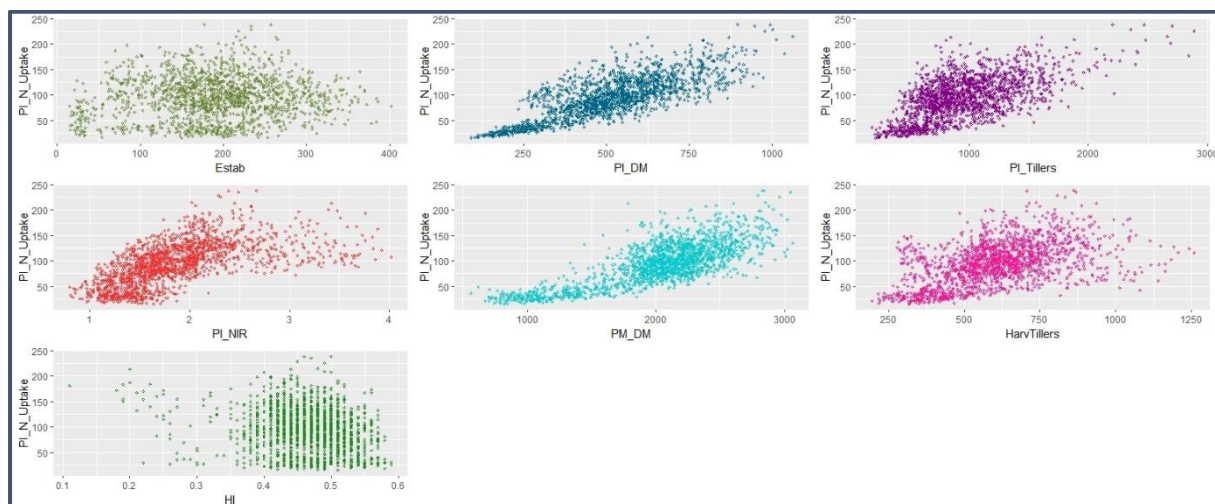
Figure 20: Boxplots- Nitrogen Uptake at PI

Figure 21: Scatter plots- Nitrogen Uptake at PI

The above scatterplot grid in Figure 21 shows the correlation of PI_N_Uptake with all the other numerical predictor variables. PI_N_Uptake has moderate-strong correlation with certain variables as shown in the table below:

Table 9: Nitrogen Uptake Correlation Coefficients

Explanatory Variable	Correlation Coefficient
PI_DM	0.78
PI_Tillers	0.61
PI_NIR	0.61

PM_DM	0.76
HarvTillers	0.5

Figures 22 and 23 below shows a series of facet plots depicting the summary of:

➕ Nitrogen uptake for all sowing methods with respect to the Nitrogen rates- the highest nitrogen uptake has been exhibited by nitrogen rates falling in the PW groups for each sowing method. The highest nitrogen uptake is for drill sown rice at 240-0 nitrogen application rates across all varieties of rice.

➕ Nitrogen uptake per sowing method for each variety of rice – the plot shows the comparison of nitrogen uptake for different sowing methods for all varieties. With approximately 135 kgN/ha uptake of Nitrogen, the aerial sown V071 variety of rice has the highest median response.
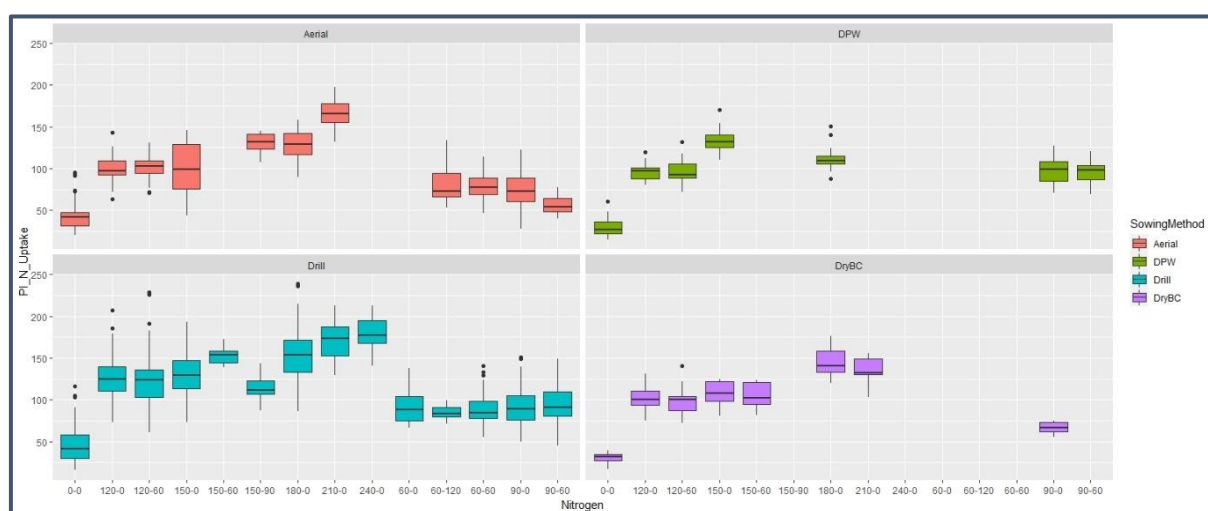

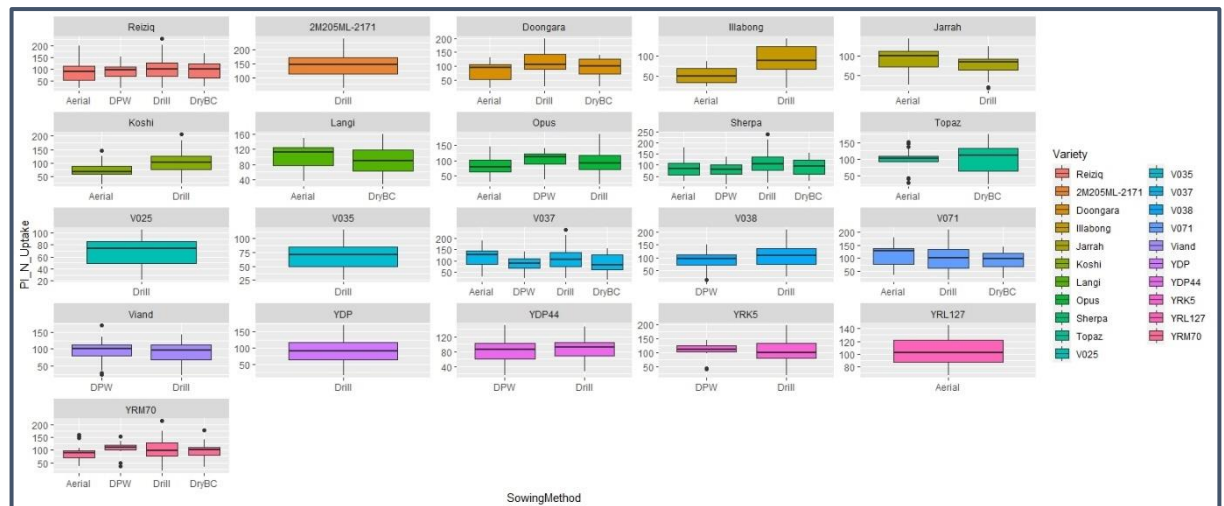Figure 22: Facet plots- Nitrogen Uptake at PI vs Nitrogen rates

Figure 23: Facet plots- Nitrogen Uptake at PI vs Sowing methods

Summary of Analysis

The summary of the preliminary analysis of the data set are as follows:

- For any variety of rice, the sowing method undertaken produces a difference in yields. Aerial and drill sown rice are likely to produce higher yields compared to the other two sowing methods.

- The split methods of Nitrogen application rates produce consistently higher yields with lesser variability compared to PW methods (split methods account only for approximately 27% of the observations). PW_3 method is likely to produce higher yields compared to other methods.

- Compared to Reiziq variety of rice, 2M205ML-2171 is likely to produce higher yields, while the other varieties only have slightly different medians with respect to Reiziq.

- Significantly higher nitrogen intake has been shown by 2M205ML-2171 variety in comparison to all other varieties.

- The Split methods of Nitrogen application tend to result in lower Nitrogen intake at PI.

- PW methods also show higher N intake rates for each of the sowing methods.