# ClimMob Reporting

# Structure of Code

The master file is Climmob.R, which is the only file which will need to be executed directly on the server, all other files are called through this main master file. The structure of the calls to the other script files is as follows:

* **1. dependencies.R**

This loads all relevant R libraries on which the scripts depend on. Mostly these are available on CRAN but two github only packages are called and it also requires phantomjs to be installed to embed the map images. Code to install the non- CRAN packages is included in the script but commented out.

* **2. functions.R**

This loads the custom functions written for the analysis

* **3. params.R**

This takes in user specified or context determined parameters for the reporting process. These user specified parameters should be linked into the Climmob interface for building the report. This is the only file which needs to be dynamically linked to the Climmob interface.

* **4. data.R**

Using the information from params.R about which data to access this connects to the Climmob server, downloads the data, and then various data formatting and manipulation tasks are carried out so that the data can be used within the analysis functions

* **5. analysis.R**

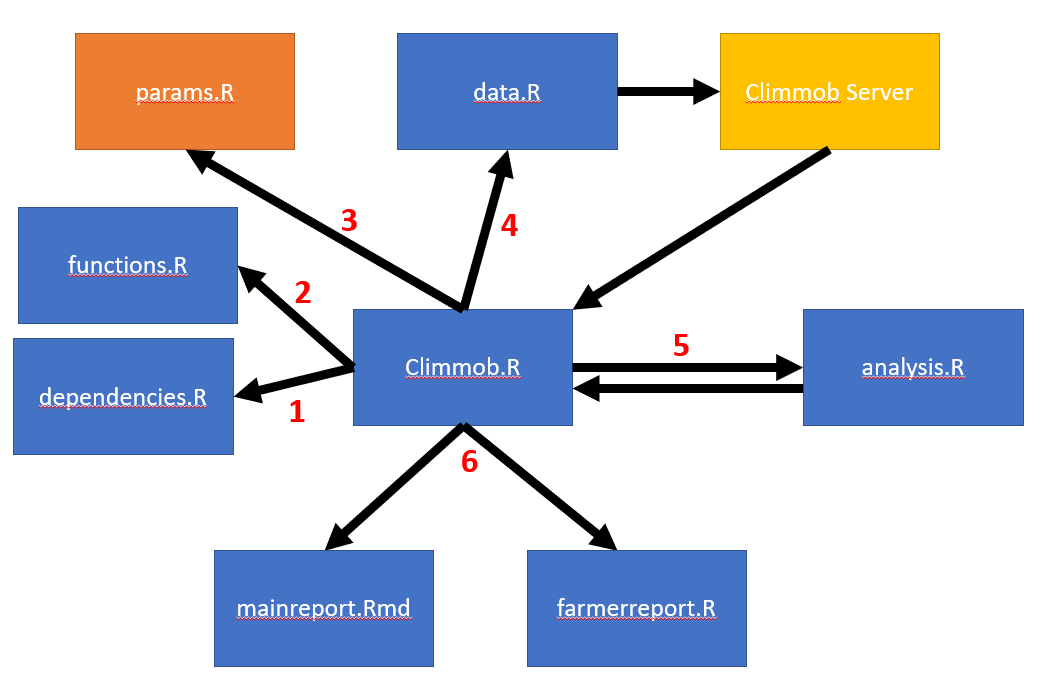
The analysis of the data is carried out and the outputs stored as objects in R

* **6. Mainreport.Rmd** and/or F**armerreport.R**

The results of the analysis are brought into the reporting format and then rendered as a word/pdf/html file as requested.

Various sections of Mainreport.Rmd require separate Rmd files as they require iterations looping over multiple columns.

Farmerreport.R script creates some temporary png images, before calling to an Rmd file multiple times to produce a file for each farmer.



# Customisable Input Parameters

|  |  |  |
| --- | --- | --- |
| Parameter | Default | Possible Options |
| key |  | The Climmob project key |
| projname |  | The Climmob project name |
| forcesplit | “gender” | The variable which will be used as a forced split in certain analysis. Must be a 2 -level factor |
| output | “both” | “both”,”summary”, “farmer”  Whether to produce both reports, only the summary report or only the farmer report |
| extension | “docx” | “docx”,”pdf”,”html” |
| ranker | “farmer” | The word(s) to be used in the report about each of the people who are ranking options. |
| option | “variety” | The word(s) to be used in the report about each of the options that are being ranked |
| coordset | NULL | A number.  If more than one set of coordinates is recorded, which is the set that should be used in the analysis and for maps. |
| sig.level | 0.05 | Level of statistical significance to use. |
| ci.adjust | “BH” | Method for adjusting p-values – can be any of the methods implemented in the p.adjust function in R. |
| ci.level | 0.84 | Level for confidence intervals in plots. Set to 84% to match the mean separation approach. |
| missper | 0.2 | The maximum threshold for missing data in a variable before it is excluded. |
| minsplit | 30 | Minimum sample size in a tree node before it attempts a split. |
| info.table.items | “” | The list of options |
| info.table.info | “” | The list of associated information |
| info.table.typeinfo | “” | “expert advice”, “info”, etc. |
| ranker.ids | List of all farmer ids | List of specific ids |

# Main Report

### Section 1: Headline Results

|  |  |  |
| --- | --- | --- |
| # | Name | Description |
| Table 1 | Frequency of Options Assessed | Frequency table for each of the options assessed |
| Table 1.1 | Summary of Differences Found in Options by Trait | Overview of which of the traits had significant differences identified overall, with the “best” and “worst” options highlighted |
| Table 1.2.1 | Summary of univariate p-values for first split in Plackett-Luce tree model for the overall ranking | Summary of the which of each of the covariates had significant relationships to the overall ranking |
| Table 1.2.2 | Summary of different subgroups identified by multivariate Plackett-Luce tree model | Table summarising how the variables affect the ranking based on what subgroups are created |
| Table 1.3 | Relationship between individual trait assessment and overall assessment | Table summarising the strength of relationship between the individual traits and the overall ranking illustrating which traits have the strongest link. |

### Section 2: Data Summary and Exploratory Analysis of Traits

|  |  |  |
| --- | --- | --- |
| # | Name | Description |
| Table 2.1 | Summary of Overall Performance | “Wins”, “Losses” and “Favourability Scores” for each of the varieties. Although this is not adjusted for different comparisons these numbers are much easier to understand/interpret as compared to the model parameter estimates. |
| Figure 2.1 | Net Favourability Score for Overall Performance | Plot showing the results from table 2.1 |
| Table 2.2.i | Summary of Performance for [trait i] | Same as table 2.1, but repeated for each of the other traits which are included in the data. |
| Figure 2.2.i | Net Favourability Score for [trait j] | Same as figure 2.1, but repeated for each of the other traits which are included in the data. |
| Figure 2.3 | Head to head performance of All Varieties Using Overall Trait Preference | Head to head plots showing “wins” and “losses” of all pairwise comparisons. Presented visually to try to make it easier to process information - numbers are included in Appendix B. |
| Figure 2.4.i | Head to head performance of All Varieties Using [trait j] | Same as figure 2.3 but repeated for all individual traits. |
| Table 2.5 | Relationship between individual trait assessment and overall assessment | Table summarising the % agreement between rankings on the overall variable against each of the individual trait rankings. |
| Figure 2.5 | Relationship between individual trait assessment and overall assessment | Same information as table 2.5 presented in a chart. |

### Section 3: Data Summary and Exploratory Analysis of Covariates

|  |  |  |
| --- | --- | --- |
| # | Name | Description |
| Figure 3.0 | Map showing location of respondents | Map produced based on coordinate set. Will automatically adjust and scale and label. |
| Table 3.i.1 | Missing values in [covariate i] | Summary of completeness for a possible covariate |
| Table 3.i.2 | Summary statistics for [covariate i] | Summary stats for a possible covariate – table produced varies depending on whether covariate is a number, category or date. |
| Figure 3.i | Summary plot for [covariate i] | Useful summary plot of possible covariate. Either a bar chart or a histogram based on the type of variable |

### Section 4: Plackett-Luce Models of Ranking Differences in Overall Trait

|  |  |  |
| --- | --- | --- |
| # | Name | Description |
| Table 4.1 | Likelihood ratio test results from overall model | Extracting likelihood ratios and degrees of freedom from Plackett-Luce models and creating an ANOVA like result for a hypothesis of whether there is an overall difference between the rankings of varieties |
| Figure 4.1 | Overall Ranking: Model Coefficients and Mean Separation | Extracting coefficients, confidence intervals, and construction of mean separation of varieties to determine “best” and “worst” varieties |
| Table 4.2 | Model Coefficients and Mean Separation | Same information as Fig 4.1 in a table. |
| Table 4.3 | Percentage probability of being the highest ranked overall | Conversion of model parameters into a more interpretable number – the % probability that each variety would be the highest ranked overall in a contest between all possible varieties. |
| Figure 4.2 | Overall Ranking: Probability of Being the Highest Ranked Overall | Same as Table 4.3 but in a figure. |

### Section 5: Plackett-Luce Models of Ranking Differences In Other Traits

|  |  |  |
| --- | --- | --- |
| # | Name | Description |
| Table 5.i.1 | Likelihood ratio test results from model for [trait i] | As Table 4.1; but looped through each individual trait |
| Figure 5.i.1 | Model Coefficients and Mean Separation for [trait i] | As Figure 4.1; but looped through each individual trait |
| Table 5.i.2 | Model Coefficients and Mean Separation for [trait i] | As Table 4.2; but looped through each individual trait |
| Table 5.i.3 | Table 4.3: Percentage probability of being the highest ranked for [trait i] | As Table 4.3; but looped through each individual trait |
| Figure 5.i.2 | Probability of Being The Highest Ranked Overall for [trait i] | As Figure 4.2; but looped through each individual trait |

### Section 6: Plackett-Luce Models With Covariates

|  |  |  |
| --- | --- | --- |
| # | Name | Description |
| Table 6.1 | Univariate p-values for first split in Plackett-Luce tree model for the overall ranking | p-values from the univariate Plackett-Luce trees |
| Figure 6.1 | Overall Ranking Plackett-Luce Tree Considering All Possible Covariate | Overall tree produced from model including all covariates |
| Table 6.2 | Summary of Performance in Each Node | Summarising the results within each of the terminal nodes – equivalent to Table 4.2 |
| Figure 6.2 | Coefficient Estimates Within Each Identified Terminal Node Subgroup | Summarising the results within each of the terminal nodes – equivalent to Fig 4.1, but overlayed with different colours for each node |
| Table 6.3 | p-values for effect of each covariate at each node | Further summaries detailing each split and the effects of covariates within each of these splits |
| Figure 6.4.i | Overall Ranking Plackett-Luce Tree Considering [covariate i] | For each possible covariate the tree produced from a univariate analysis |
| Figure 6.5.i | Coefficient Estimates Within Each Identified Terminal Node Subgroup [covariate i] | For each possible covariate the terminal node plots equivalent to Fig 6.2 |

### Section 7: Plackett-Luce Models With Covariates

|  |  |  |
| --- | --- | --- |
| # | Name | Description |
| Table 7.1 | Partial Least Squares Biplot of Relationship Between Traits and Overall Performance | Biplot of varieties and traits in relation to overall preferences as determined through PLS (Partial Least Squares) methods. |
| Table 7.1 | Partial Least Squares Regression Model of Traits Against Overall Ranking | Summary of relationships between individual traits and the overall trait in for of a Partial least Squares model, with parameters showing direction of relationship and p-values indicating level of significance. Because of strong correlation between all traits this will need some careful interpretations. |

### Issues/Limitations/Next Steps

Not been able to spend enough to time on how the classification tree plots will appear in reliably useful ways. I know Kaue has done some work on this so maybe worthwhile he links in his work on this.

Resolution of graphs is robust to varying numbers of treatments and different treatment names; but very long treatment names may cause graphs to be a problem, as would a very large number of varieties being compared.

Looping section 6 over each individual trait would be an obvious thing to add

Grammar/explanatory text can be improved in lots of places to aid interpretation of the results.

Noticed Kaue has made updates to his package(s) which mean some functions/options have been renamed. Code still uses the old function and argument names. Everything still works for now but may need updating calls to these functions to reflect new naming conventions from these packages.

# Farmer Feedback Reports

The farmer feedback reports are one page feedback sheets showing the most to least often favourite varieties based on overall ratings, together with the farmer’s own ranking, in a way that should be accessible to the farmers.

We also have at the top of the sheet, a short info box with farmer’s name, number of farmers who participated, and the names of the three options that the farmer had to rank.

At the back of the sheet, we have some additional information about some of the varieties, or just the list of varieties with bank column info that can be filled by the farmer if no additional information has been added by the user at the input stage (default).

A farmer feedback report will be produced for each of the farmers whose id has been included in the *ranker.ids* parameter (should likely be defaulted to all ids)

### Issues/Limitations

Only rankings based on overall rankings is available at this stage.

The sheet can handle a fairly large number of options, but it will start being unreadable when we go over 150 options. Similarly, it handles fairly long option names, but if some names get very long, some texts may start to overlap.

Information concerning varieties should be limited in length. No more than 95 characters (80 if there are lots of capital letters). One line break can be added when there are less than 25 options for which information is provided.