Report of ClimMob project chocolatesnepal

ClimMob.net

13 August, 2020

You are reading a report generated by ClimMob, a software to design, manage, analyze trial data generated by crowdsourced citizen science.

# Introduction

In agriculture, the local environmental conditions determine to a large degree which technological solutions are the most suitable. In dry soils, for example, drought-resistant crop varieties will outperform other varieties, but in wet soils these same varieties may do worse than most. Not only drought, but an entire range of problems including excessive heat, floods, new pests and diseases tend to intensify under climate change. This multitude of limiting factors requires multiple technological solutions, tested in diverse environments.

Citizen science is based on the cooperation of citizen scientist or observers (paid or unpaid). Researchers assign microtasks (observations, experiments) that, once completed and gathered, contribute with a great amount of information to science. One of the advantages of citizen science is that agricultural researchers can get access to many environments by crowdsourcing their experiments. As farmers contribute with their time, skills and knowledge to the investigation, researchers are able to do more tests than in a traditional setup. Also citizen scientists acquire new knowledge, abilities and information useful for future challenges of their work.

## ClimMob

The primary goal of ClimMob is to support the selection of innovative technologies (e.g. new crop varieties, management, new product). ClimMob was created as part of Bioversity International’s research in the CGIAR Research Programme on Climate Change, Agriculture, and Food Security (CCAFS). It serves to prepare and analyze citizen science experiments in which a large number of participants observe and compare different technological options under a wide range of conditions (*1*).

ClimMob software assigns a limited number of items (typically 3) to each participant, who will compare their performance. Each participant gets a different combination of items drawn from a much larger set of items (typically 15-25). Comparisons of this kind are thought to be a very reliable way to obtain data. Once the results of the microtasks have been collected, ClimMob builds an image of the whole set of assigned objects, combining all observations. ClimMob not only reconstructs the overall ordering of items, but also takes into account differences and similarities between participants and the conditions under which they observe (e.g. socio-economic and plot environmental characteristics). It assigns similar participants to groups that each corresponds among different group profiles. Groups are created on the basis of whichever items which have been collected, that are found to be significantly linked to the observed rankings.

ClimMob uses Plackett-Luce models to analyze ranking data with the R (*2*) package ‘PlackettLuce’ (*3*). It automatically generates analytical reports, as well as individualized information sheets for each participant using the R packages ‘knitr’ (*4*) and ‘rmarkdown’ (*5*). Organizing the data relies on packages ‘ClimMobTools’ (*6*), ‘gosset’ (*7*), ‘gtools’ (*8*), ‘jsonlite’ (*9*), ‘partykit’ (*10*), ‘psychotools’ (*11*) and ‘qvcalc’ (*12*). Summaries and data visualization are supported by packages ‘igraph’ (*13*), ‘ggparty’ (*14*), ‘ggplot2’ (*15*), ‘ggrepel’ (*16*), ‘leaflet’ (*17*), ‘multcompView’ (*18*), ‘patchwork’ (*19*), and ‘pls’ (*20*).

## How to cite

If you publish any results generated with ClimMob, you should cite a number of articles as the package builds on various contributions. Van Etten et al. (2019) (*1*) introduced the crowdsourcing philosophy behind ClimMob. It is important to mention that ClimMob is implemented in R, a free, open-source analysis software (*2*). Methodologically, if you report on the Plackett-Luce tree results, you should mentioned that ClimMob applies the Plackett-Luce model published by Turner et al. 2020 (*3*). To cite ClimMob itself, mention van Etten et al. (2020) (*21*).

# Section 1: Headline results

Overall there were 22 participants contributing to this project. Each participant assessed 3 different colours and ranked them in order of its overall performance. In addition they also provided rankings for 4 additional characteristic(s):

Table 1.1. Summary of characteristics assessed in this project and valid answers used in this report.

|  |  |  |
| --- | --- | --- |
| Characteristic | Question | Number of valid answers |
| Overall Characteristic | Overall, which option performed better? | 22 |
| Sweetness | Which one is more sweet? | 22 |
| Bitterness | Which one more bitter? | 21 |
| Cocoa content | Which one has more cocoa mass? | 22 |
| Dark color | Which one is more dark? | 22 |

Table 1.2 shows the colours assessed within this project, with the frequency and percentage of participants who assessed each colour. If the colours has large names (> 10 characters) an abbreviation was applied across the figures in this report.

Table 1.2. Frequency of colours assessed.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Colour | Abbreviation | Freq | Relative freq | Man (n=8) | Woman (n=14) |
| Amul Milk | AmlM | 17 | 77.3% | 6 | 11 |
| Cadbury Dairy Milk | CdDM | 16 | 72.7% | 6 | 10 |
| Lindt 90 | Lindt 90 | 17 | 77.3% | 5 | 12 |
| Nestle Classic | NstC | 16 | 72.7% | 7 | 9 |

Figure 1.1 shows that the colours are all connected to each other. That means that they all co-occurred at least once in an incomplete block of 3 colours.

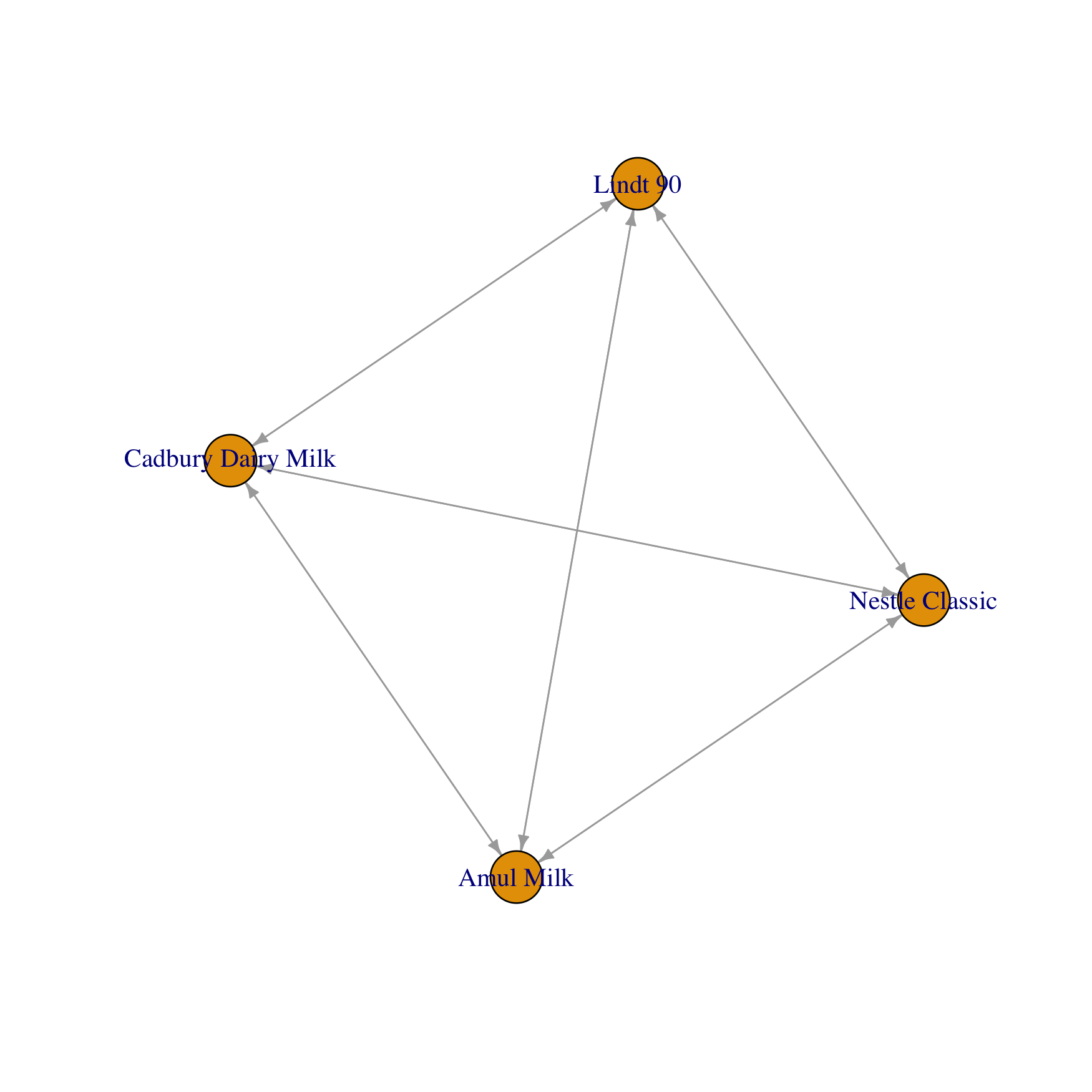


Figure 1.1. Network representation of colours tested in this project.

## Overall differences in rankings

There were statistically significant differences found in the rankings of colours in the overall performance (p = 0.00418057). The best ranked colours overall were Nestle Classic, Cadbury Dairy Milk, Amul Milk. Statistically significant differences were also found in the characteristic(s) Overall Characteristic, Sweetness, Cocoa content, Dark color

A summary of the p-values testing the hypothesis that there were differences in the rankings for each characteristic, and the list of colours which were significantly highest and lowest ranked overall, are summarised in Table 1.3.

Table 1.3. Summary of differences found in colours by characteristic.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ranking | Best Ranked | Worst Ranked | p.value |  |
| Overall | Nestle Classic, Cadbury Dairy Milk, Amul Milk | Lindt 90, Amul Milk | 0.00418057 | \*\* |
| Sweetness | Nestle Classic, Cadbury Dairy Milk, Amul Milk | Lindt 90 | 0.00012904 | \*\*\* |
| Bitterness | Lindt 90 | Nestle Classic, Cadbury Dairy Milk, Amul Milk | 1.5740e-08 | \*\*\* |
| Cocoa content | Lindt 90, Cadbury Dairy Milk, Amul Milk | Nestle Classic, Amul Milk, Cadbury Dairy Milk | 0.21742575 |  |
| Dark color | No significant difference | No significant difference | 5.1926e-11 | \*\*\* |

## Effect of covariates

None of the covariates tested were found to have a statistically significant relationship to the overall performance with an alpha = 0.1.

Table 1.4. Summary of univariate p-values for first split in Plackett-Luce tree model for the ‘overall performance’.

|  |  |
| --- | --- |
| Covariate | p.value |
| What is the gender ? | 0.979 |

## Correlation between the overall performance and the other characteristics

Table 1.6 shows, for each characteristic in the project, the frequency for which the rankings matched with the overall performance. Best and worst agreement represents the percentage for which the best and worst colour for the characteristic matched the overall best and worst. Complete ranking agreement shows the proportion of correlation on the full ranking with the ‘overall performance’ as baseline using the Kendall correlation coefficient (*22*).

Table 1.6. Correlation between the ‘overall performance’ and the others characteristics assessed in this report.

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | Complete Ranking Agreement | Agreement with Overall Best | Agreement with Overall Worst |
| Sweetness | 30.2% | 52.4% | 57.1% |
| Bitterness | -23.8% | 23.8% | 19% |
| Cocoa content | 4.8% | 33.3% | 42.9% |
| Dark color | -30.2% | 19% | 9.5% |

The characteristic which had the strongest relationship with the ‘overall performance’ was Sweetness. Overall, the rankings for Sweetness matched the rankings for the overall performance 30% of the time. The characteristic which had the weakest relationship with the ‘overall performance’ was Dark color. Overall the rankings for Dark color matched the rankings for the ‘overall performance’ -30% of the time. The correlation with the other characteristics is also shown in Figure 1.2.

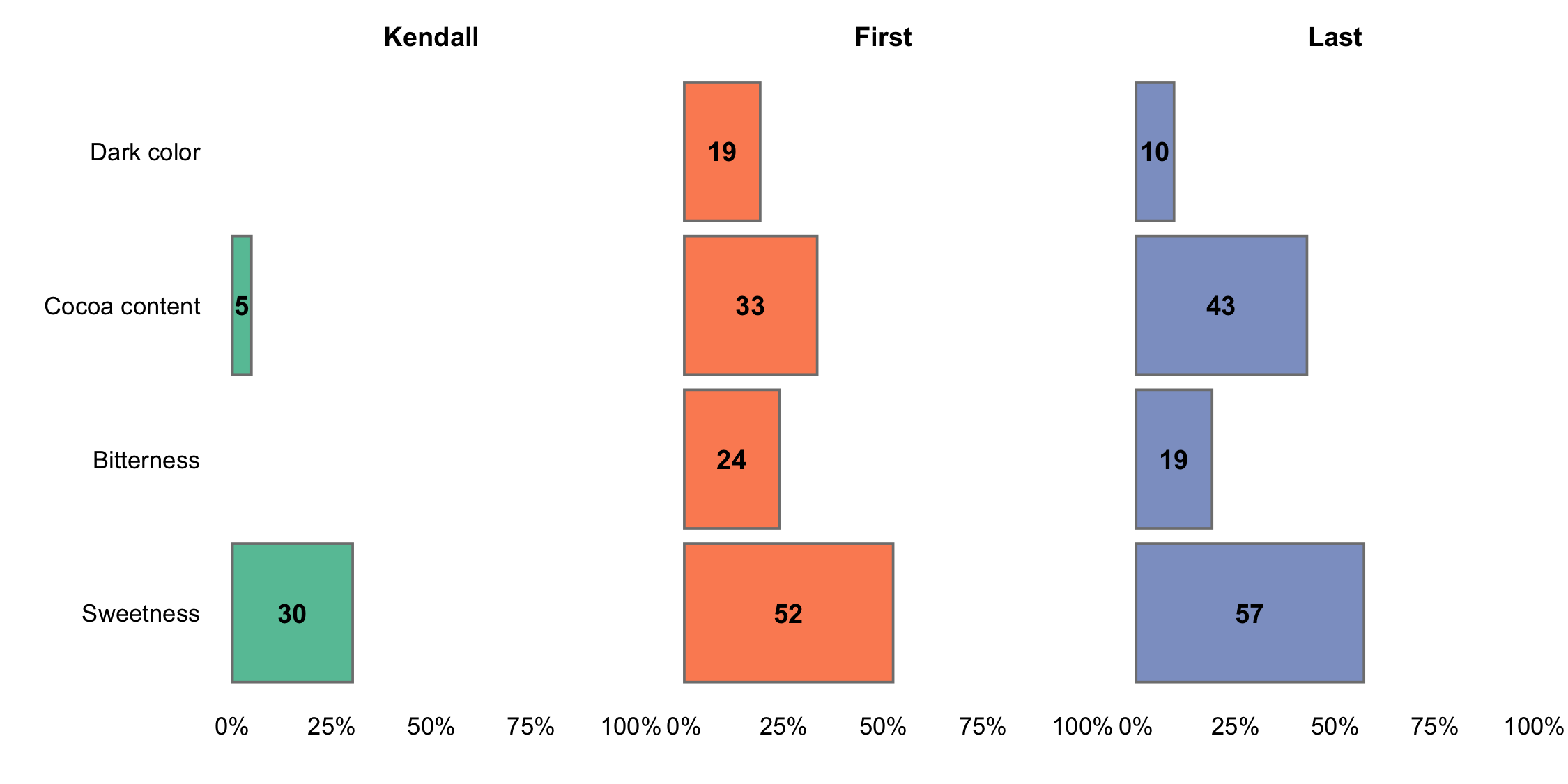


Figure 1.2. Correlation between individual characteristic and the ‘overall performance’.

Partial least squares regression was used to determine relationship between the specific characteristics and the ‘overall performance’. The first two components recombining the specific characteristics are able to explain the variability in the ‘overall performance’. The dashed line represents the overall performance, with an increase in performance as the x and y increase. The colour positioned close to the dashed line will be performing equally across all characteristics; the colour positioned further away from the dashed line, on either side, will have varying performance in different characteristics. Better performance in characteristics will correspond with arrows pointing in the direction away from the dashed line and worse performance in characteristics directed on the opposite side.

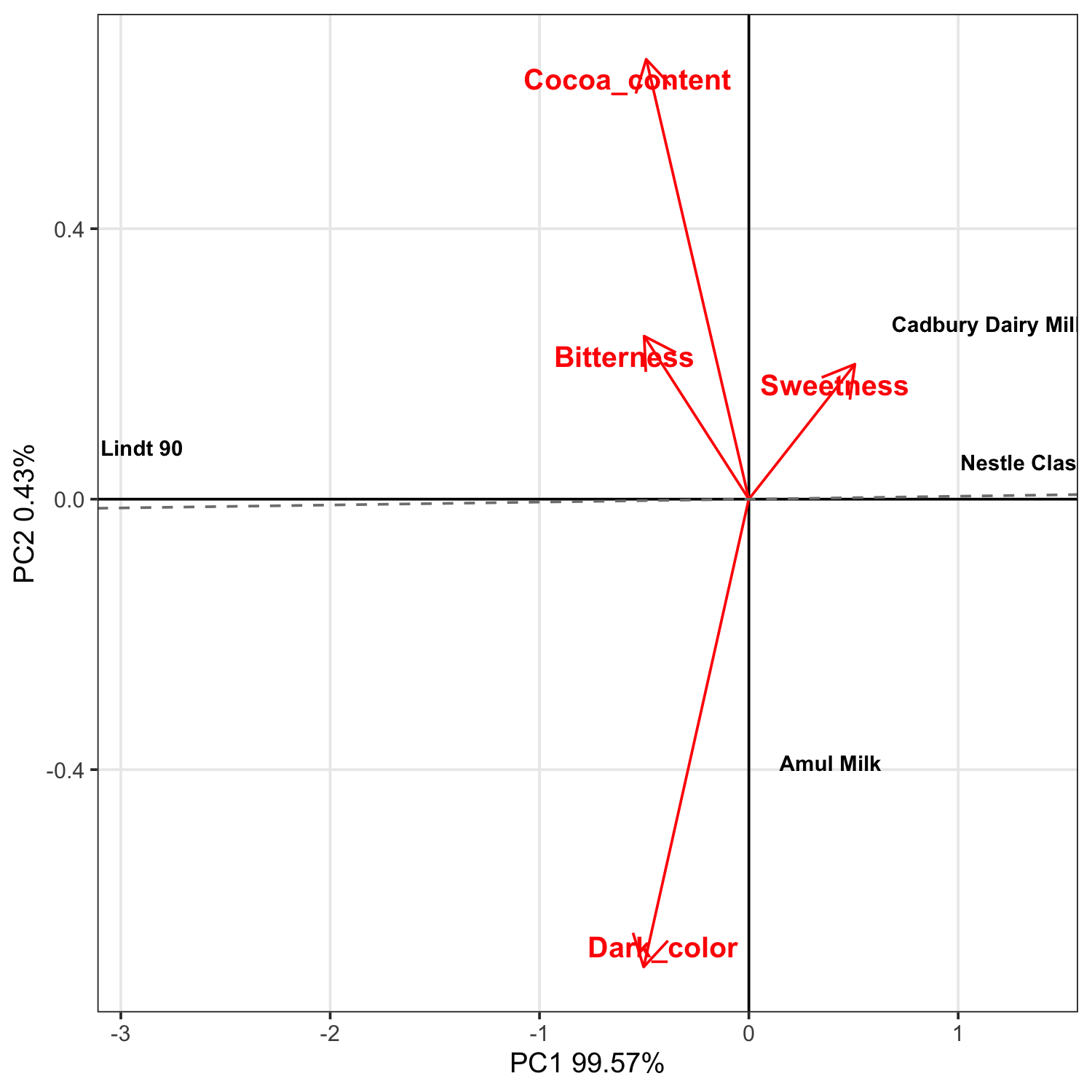


Figure 1.3. Partial least squares biplot of relationship between other characteristics and overall performance.

# Section 2: Data summary and exploratory analysis of characteristics

## Assessment of colours

Exploratory analysis within the following section summarises results from the data directly. Given the structure of a ClimMob trail, where each participant only assessed 3 of the 4 possible colours these results may be skewed if certain colours were randomly assigned to face worse colours than others. This is particularly a potential issue within a smaller trial, as due to the randomisation process the potential for an unbalanced assignment decreases as the sample size increases. Results from other sections, and in the overall summary use Plackett-Luce models (*3*), to adjust for any imbalance.

### Overall performance

Overall performance of each of the colours is summarised in Table 2.1.

Table 2.1. Favourability scores for overall performance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Colour | N | Top Ranked | Bottom Ranked | Net Favourability Score |
| Nestle Classic | 16 | 50% | 18.8% | 31.2 |
| Cadbury Dairy Milk | 16 | 37.5% | 6.2% | 31.2 |
| Amul Milk | 17 | 29.4% | 29.4% | 0.0 |
| Lindt 90 | 17 | 17.6% | 76.5% | -58.8 |

This shows the percentage of participants who assessed the colours as the best among the 3 colours they were provided, the percentage of participants who included the colour as their worst, the percentage of ‘head to head contests’ for which the colour won and the net *favourability* score. A score of +100 indicates the colour won all ‘contests’ it was involved in, a score of 0 indicates an equal number of wins and losses, a score of -100 indicates the colour lost all contests.

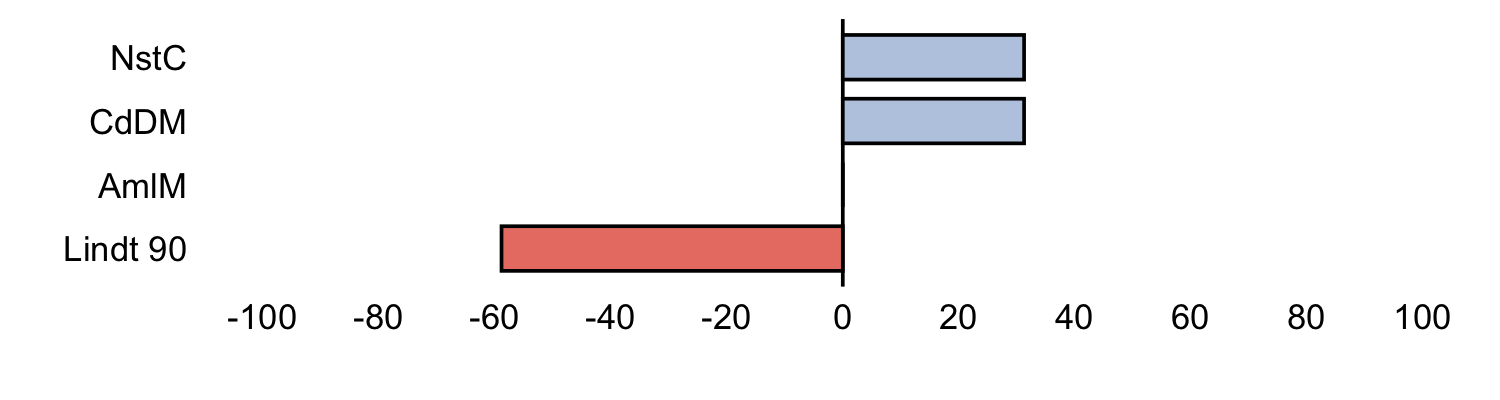


Figure 2.1. Net favourability score for ‘overall performance’.

The colour Nestle Classic was the ‘best’ colour overall being ranked highest by 50% of the 22 participants who assessed this colour .

### Other characteristics

Net favourability scores are shown below for theother characteristics assessed in this project.

**Sweetness**

Table 2.1.1. Favourability scores for ‘sweetness’

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Colour | N | Top Ranked | Bottom Ranked | Net Favourability Score |
| Lindt 90 | 17 | 0% | 82.4% | -82.4 |
| Amul Milk | 17 | 35.3% | 29.4% | 5.9 |
| Cadbury Dairy Milk | 16 | 50% | 12.5% | 37.5 |
| Nestle Classic | 16 | 50% | 6.2% | 43.8 |

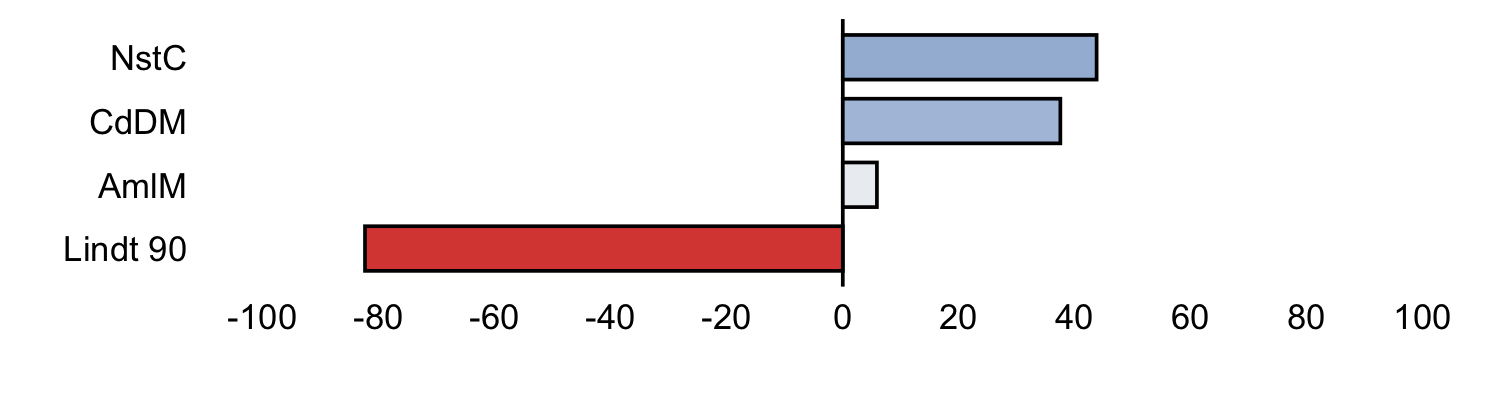


Figure 2.1.1. Net favourability score for ‘sweetness’.

**Bitterness**

Table 2.1.2. Favourability scores for ‘bitterness’

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Colour | N | Top Ranked | Bottom Ranked | Net Favourability Score |
| Nestle Classic | 16 | 0% | 68.8% | -68.8 |
| Cadbury Dairy Milk | 15 | 6.7% | 33.3% | -26.7 |
| Amul Milk | 16 | 25% | 31.2% | -6.2 |
| Lindt 90 | 16 | 100% | 0% | 100.0 |

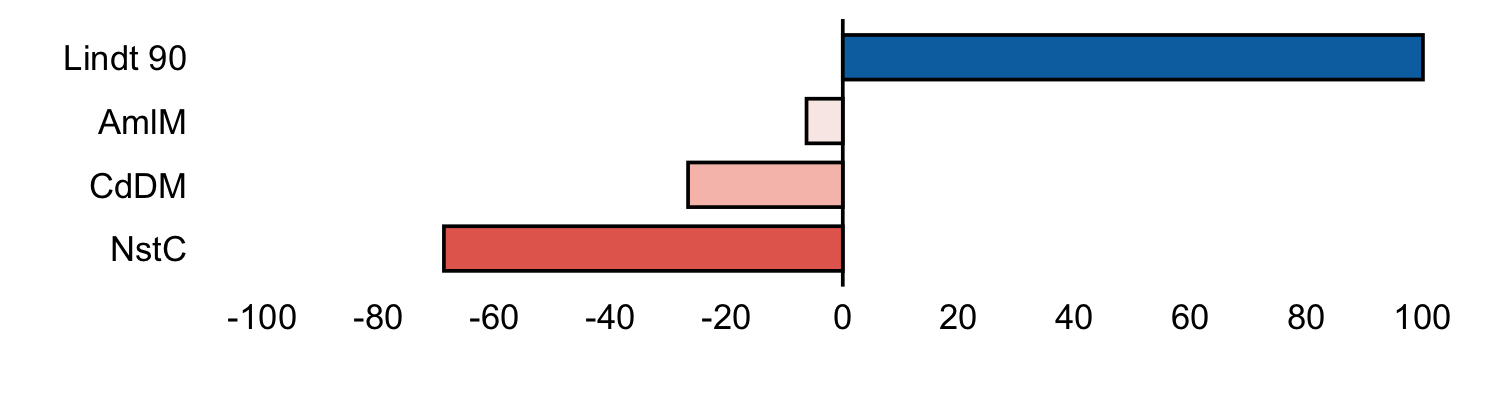


Figure 2.1.2. Net favourability score for ‘bitterness’.

**Cocoa content**

Table 2.1.3. Favourability scores for ‘cocoa content’

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Colour | N | Top Ranked | Bottom Ranked | Net Favourability Score |
| Nestle Classic | 16 | 18.8% | 37.5% | -18.8 |
| Amul Milk | 17 | 17.6% | 35.3% | -17.6 |
| Cadbury Dairy Milk | 16 | 25% | 37.5% | -12.5 |
| Lindt 90 | 17 | 70.6% | 23.5% | 47.1 |

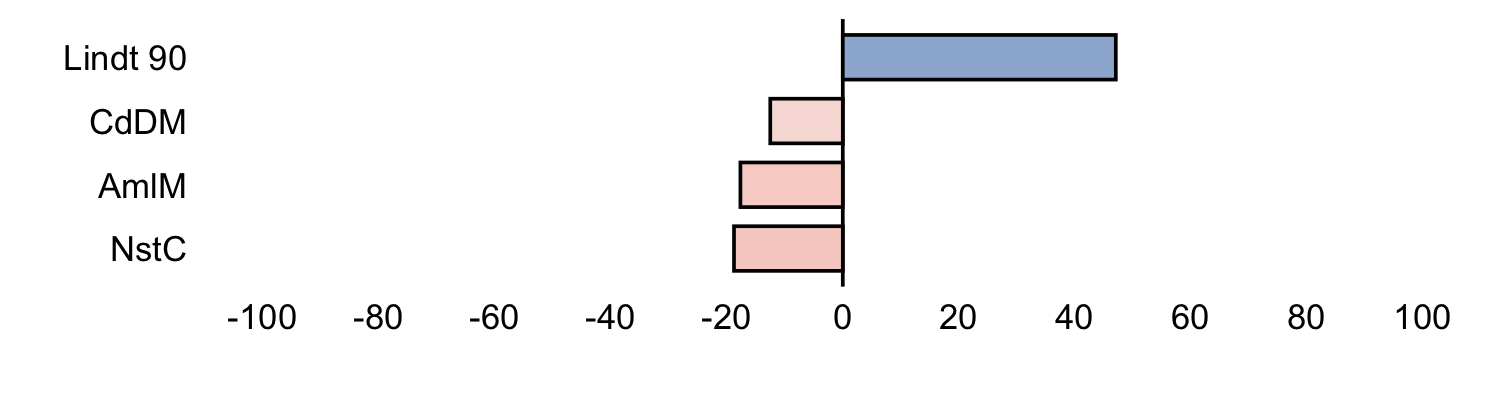


Figure 2.1.3. Net favourability score for ‘cocoa content’.

**Dark color**

Table 2.1.4. Favourability scores for ‘dark color’

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Colour | N | Top Ranked | Bottom Ranked | Net Favourability Score |
| Cadbury Dairy Milk | 16 | 0% | 68.8% | -68.8 |
| Nestle Classic | 16 | 0% | 56.2% | -56.2 |
| Amul Milk | 17 | 29.4% | 11.8% | 17.6 |
| Lindt 90 | 17 | 100% | 0% | 100.0 |

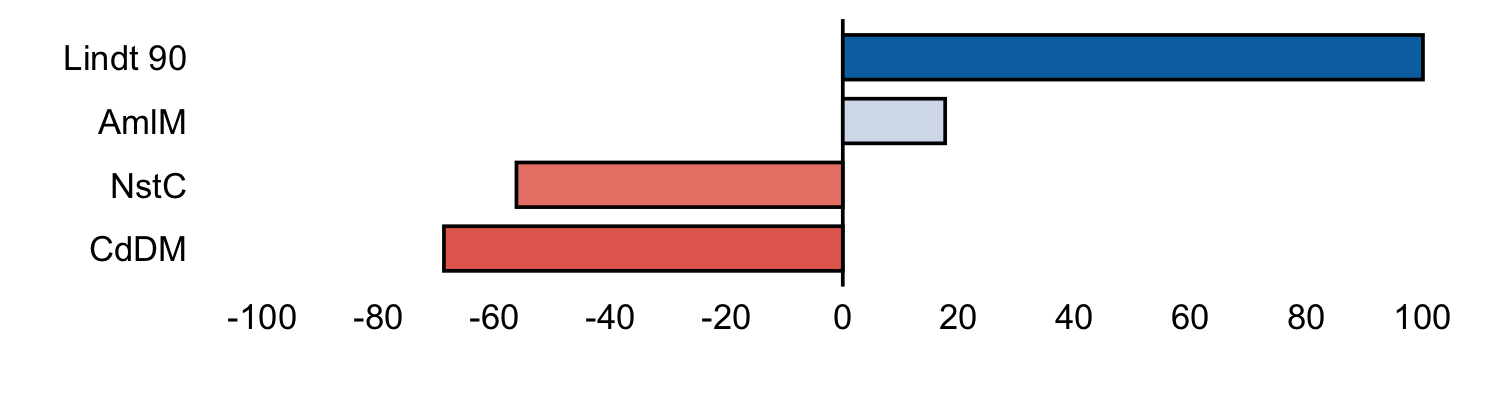


Figure 2.1.4. Net favourability score for ‘dark color’.

## Pairwise contests

### Overall performance

Figure 2.2 shows the outcomes of all pairwise contests between the colours included in the project for ‘overall performance’. Each panel shows the performance of one colour against all the other colours, and shows the percentage of the times in which the panelled colour was ranked above the other colours shown as bars.

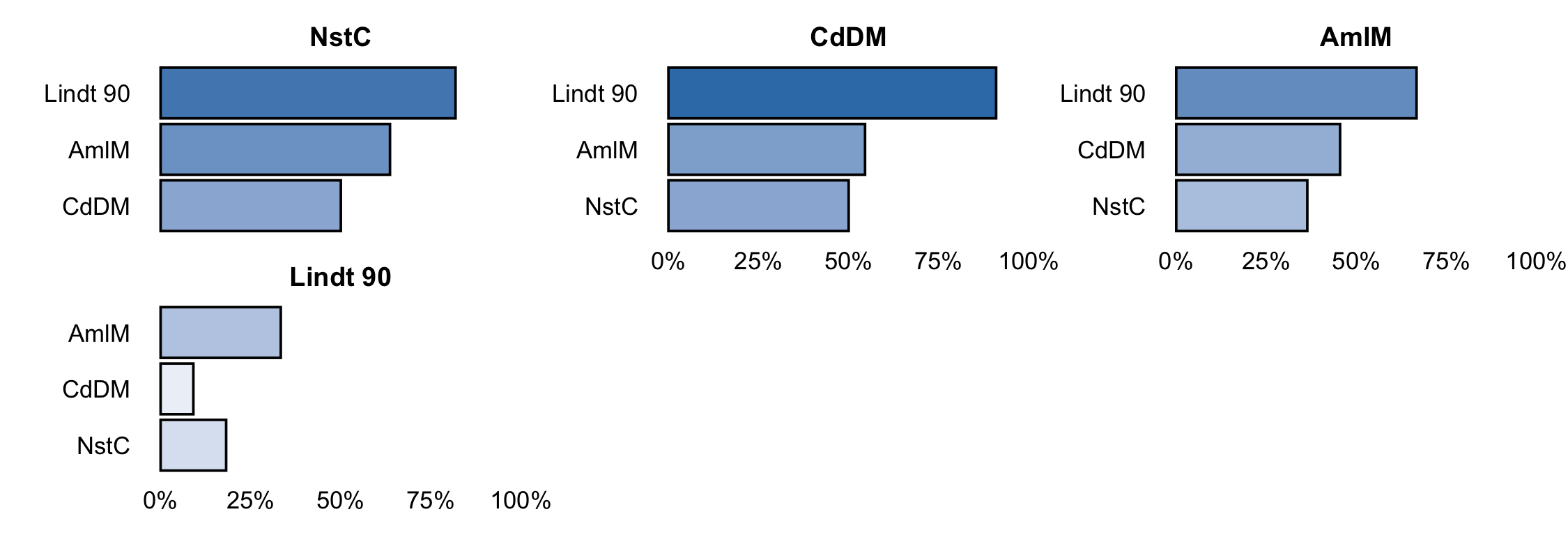
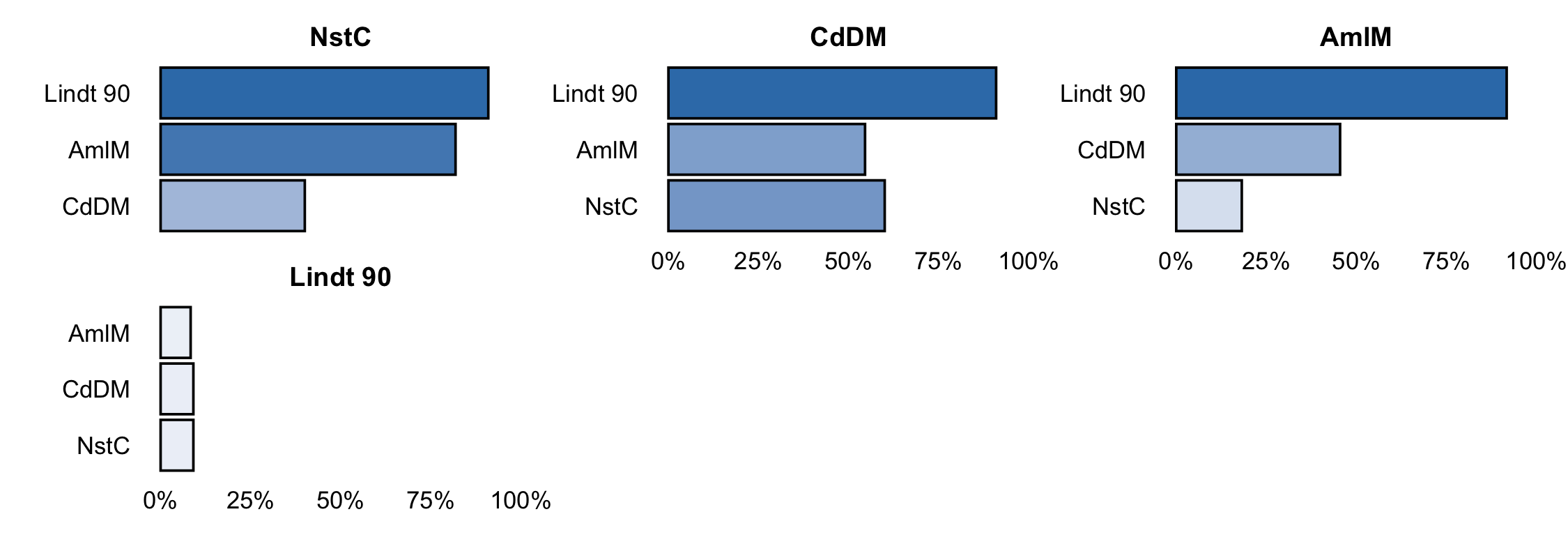


Figure 2.2. Head to head performance for ‘overall performance’.

### Other characteristics

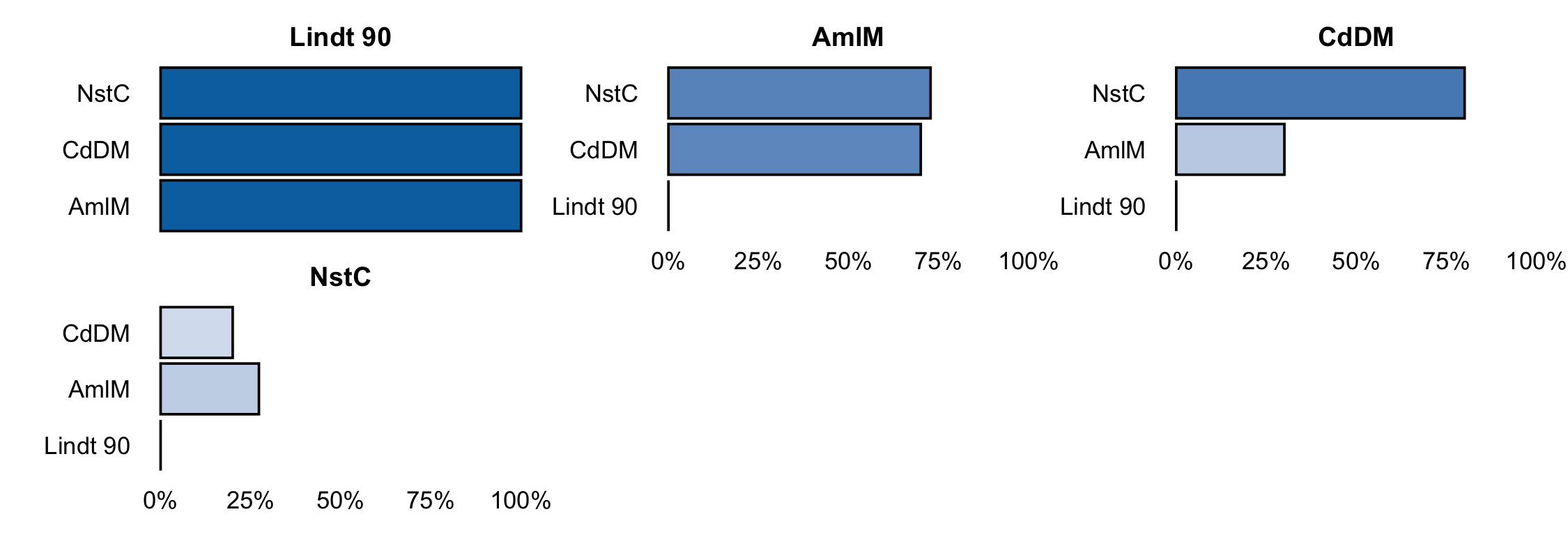
Results from the pairwise contests of the otherindividual characteristics assessed are shown below.

**Sweetness**



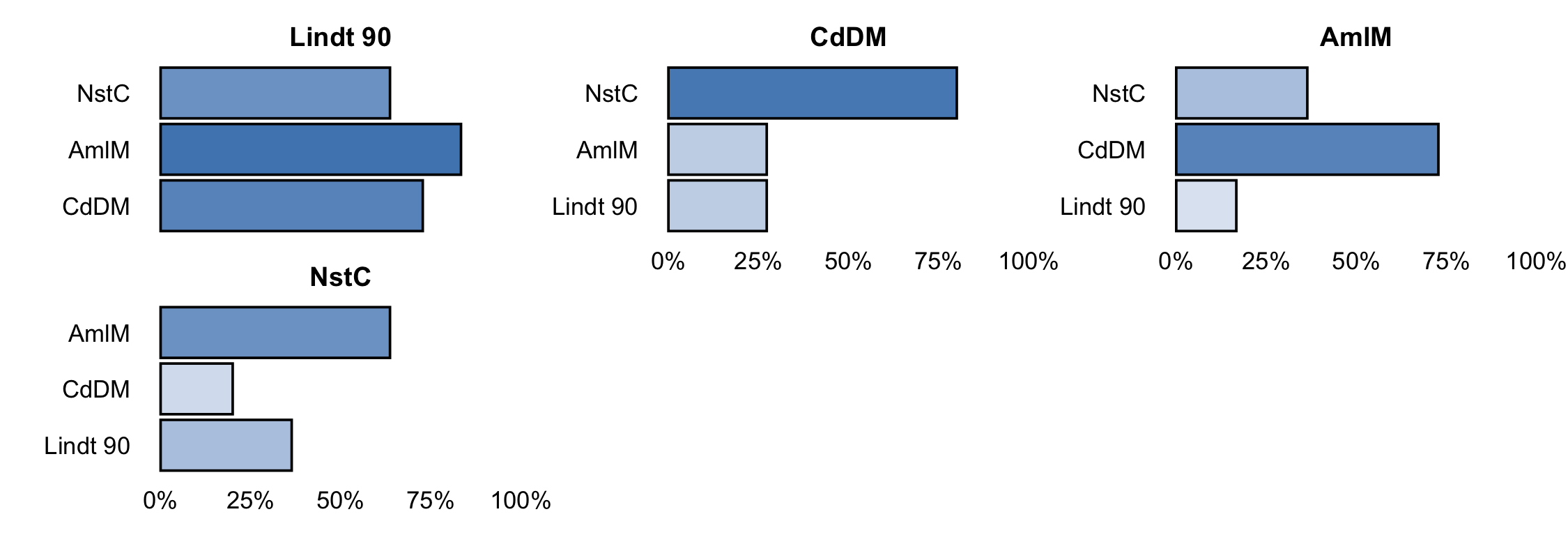
2.2.1. Head to head performance for ‘sweetness’.

**Bitterness**



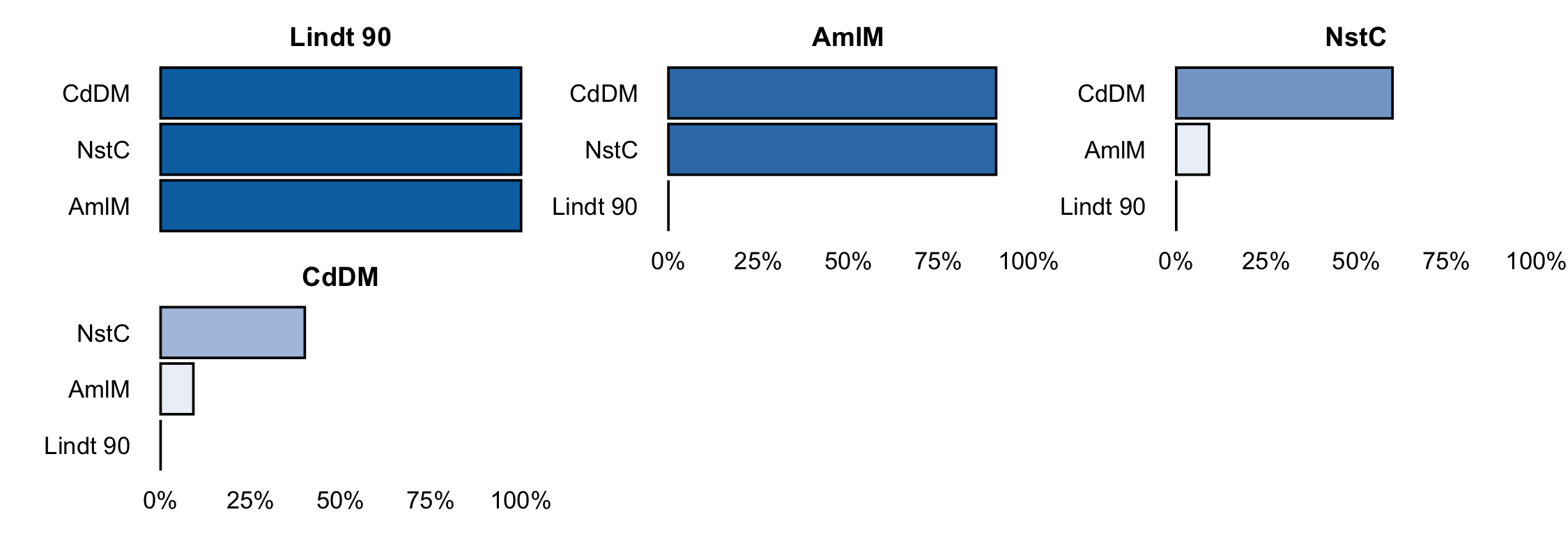
2.2.2. Head to head performance for ‘bitterness’.

**Cocoa content**



2.2.3. Head to head performance for ‘cocoa content’.

**Dark color**



2.2.4. Head to head performance for ‘dark color’.

# Section 3: Data summary and exploratory analysis of covariates

The effect of covariates on the rankings of colours was explored with Plackett-Luce trees using the following covariates (Table 3.1)

Table 3.1. Covariates used in this project.

|  |  |
| --- | --- |
| Short name | Question |
| registration\_REG\_gender | What is the gender ? |

## Overall performance

Table 3.2 shows the results from the likelihood ratio test from the Plackett-Luce model for overall performance of the different colours. The hypothesis being tested is that there is no difference in the assessments of any of the different colours.

Table 3.2. Likelihood ratio test results from fitted Plackett-Luce model with rankings from ‘overall performance’.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | logLikelihood | DF | Statistic | Pr(>Chisq) |
| NULL | -39.41871 | 66 | NA | NA |
| Overall performance | -32.80778 | 63 | 13.22187 | 0.0041806 \*\* |

Figure 3.1 shows the estimates of the model coefficients with 84% confidence intervals. The purpose of this graph is to be able to best distinguish between the relative strength of each of the colours assessed. As such the coefficient estimates themselves are not directly interpretable, but it can be concluded that a higher value for the coefficient indicates that a colour has been ranked as best more often. The 84% confidence width is chosen so that non-overlapping confidence intervals could be interpreted as indicating significant differences at the alpha = 0.1. This may not match exactly with the mean separation groupings, as these groupings also take into account multiple testing through the Benjamini and Hochberg adjustment (*23*).

Mean separation analysis was also conducted to indicate, using letters, which colours are significantly more preferred than others; when colours have at least one letter in common, there is not enough evidence from the experiment to be confident about their relative order at the alpha = 0.1.

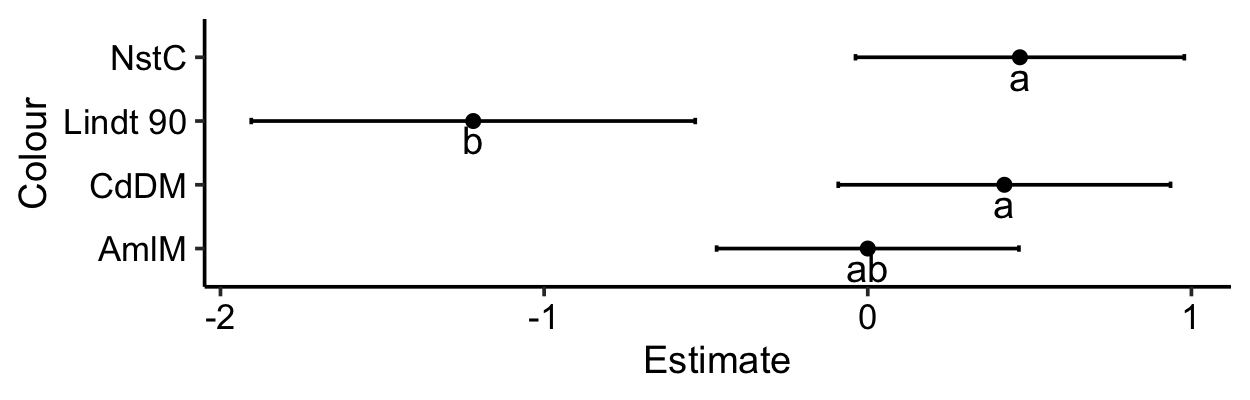


Figure 3.1. Model coefficients and mean separation of Plackett-Luce model for ‘overall performance’ with 84% confidence intervals.

The same information as Figure 3.1 is shown in Table 3.3 below.

Table 3.3. Model coefficients and mean separation of colours with an alpha = 0.1.

|  |  |  |  |
| --- | --- | --- | --- |
| Colour | Estimate | quasiSE | Group |
| 1 | 0.4701 | 0.3615 | a |
| 2 | 0.4222 | 0.3653 | a |
| 3 | 0.0000 | 0.3324 | ab |
| 4 | -1.2190 | 0.4882 | b |

Table 3.4 and Figure 3.2 use the coefficients from the Plackett-Luce model to estimate the probability of each colour being considered to be the top ranked colour in a direct comparison between all of the possible colours.

Table 3.4. Percentage probability of being the best ranked for ‘overall performance’.

|  |  |
| --- | --- |
| Colour | Win probability |
| Nestle Classic | 36.2% |
| Cadbury Dairy Milk | 34.5% |
| Amul Milk | 22.6% |
| Lindt 90 | 6.7% |

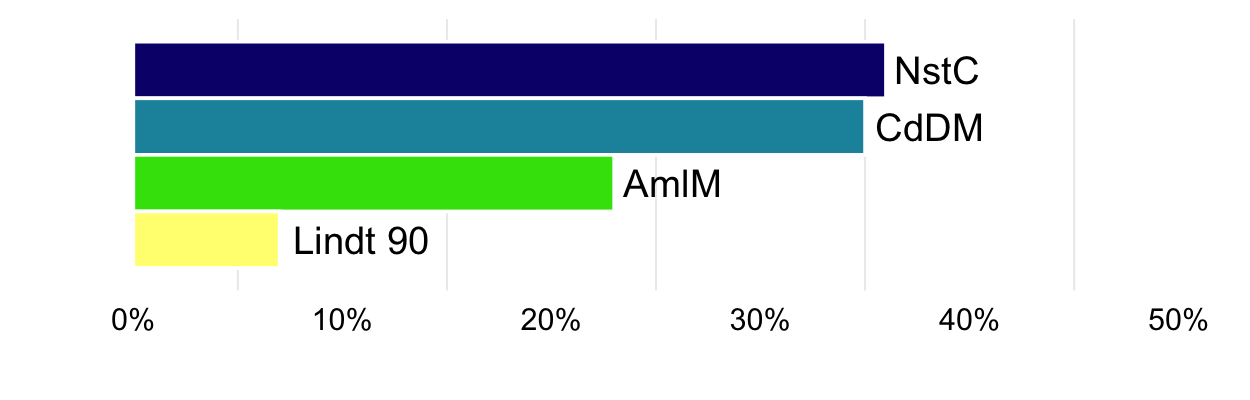


Figure 3.2. Probability of being the best ranked for ‘overall performance’.

# Section 4: Plackett-Luce models of other characteristics

This section shows the Plackett-Luce models without covariates built using the rankings for each of the 4 other characteristics that were evaluated by the participants in this project.

**Sweetness**

Table 4.1.1. Likelihood ratio test results from fitted Plackett-Luce model with rankings from ‘sweetness’.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | logLikelihood | DF | Statistic | Pr(>Chisq) |
| NULL | -39.41871 | 66 | NA | NA |
| Sweetness | -29.13163 | 63 | 20.57415 | 0.00012904 \*\*\* |

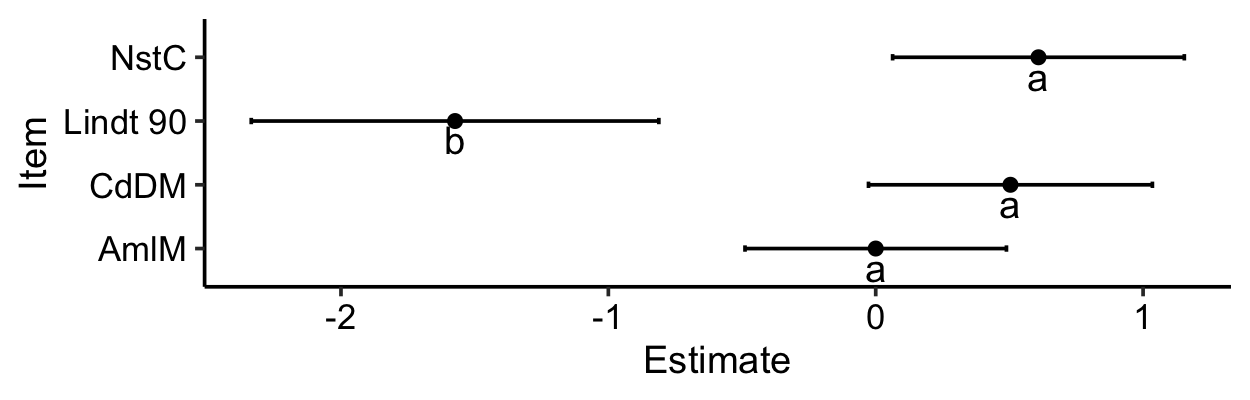


Figure 4.1.1. Model coefficients and mean separation of Plackett-Luce model for ‘sweetness’ with 84% confidence intervals.

Table 4.1.2. Model coefficients and mean separation of Plackett-Luce model for ‘sweetness’ with 84% confidence intervals.

|  |  |  |  |
| --- | --- | --- | --- |
| Colour | Estimate | quasiSE | Group |
| Nestle Classic | 0.6087 | 0.3882 | a |
| Cadbury Dairy Milk | 0.5039 | 0.3779 | a |
| Amul Milk | 0.0000 | 0.3480 | a |
| Lindt 90 | -1.5734 | 0.5425 | b |

Table 4.1.3 Percetage probability of being highest ranked for ‘sweetness’.

|  |  |
| --- | --- |
| Colour | Win probability |
| Nestle Classic | 39.1% |
| Cadbury Dairy Milk | 35.2% |
| Amul Milk | 21.3% |
| Lindt 90 | 4.4% |

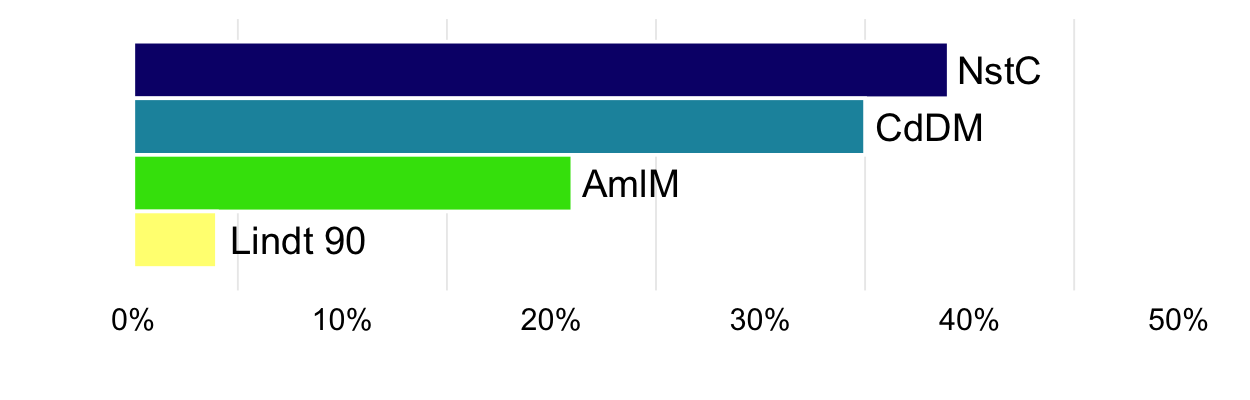


Figure 4.1.2. Probability of being the best ranked for ‘sweetness’.

**Bitterness**

Table 4.2.1. Likelihood ratio test results from fitted Plackett-Luce model with rankings from ‘bitterness’.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | logLikelihood | DF | Statistic | Pr(>Chisq) |
| NULL | -37.62695 | 63 | NA | NA |
| Bitterness | -18.02676 | 60 | 39.20037 | 1.574e-08 \*\*\* |

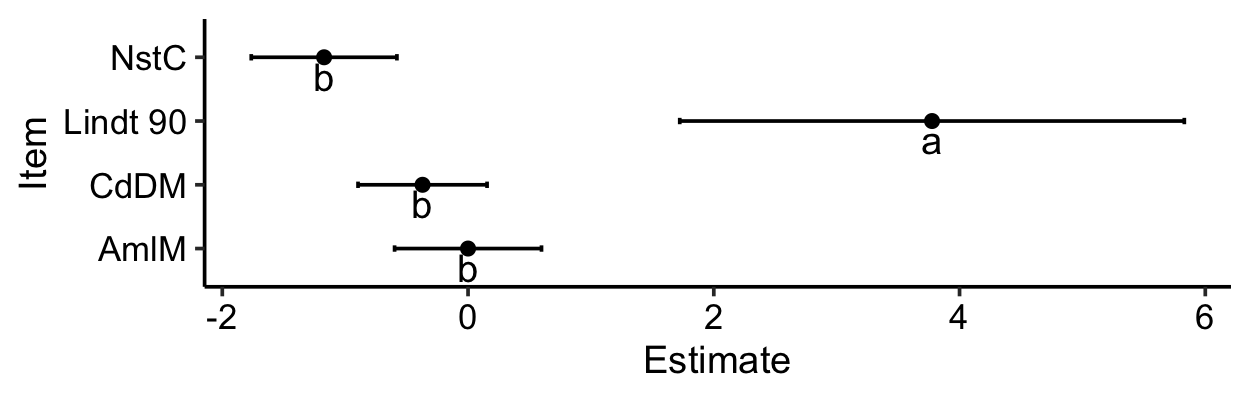


Figure 4.2.1. Model coefficients and mean separation of Plackett-Luce model for ‘bitterness’ with 84% confidence intervals.

Table 4.2.2. Model coefficients and mean separation of Plackett-Luce model for ‘bitterness’ with 84% confidence intervals.

|  |  |  |  |
| --- | --- | --- | --- |
| Colour | Estimate | quasiSE | Group |
| Lindt 90 | 3.7761 | 1.4613 | a |
| Amul Milk | 0.0000 | 0.4255 | b |
| Cadbury Dairy Milk | -0.3703 | 0.3737 | b |
| Nestle Classic | -1.1716 | 0.4218 | b |

Table 4.2.3 Percetage probability of being highest ranked for ‘bitterness’.

|  |  |
| --- | --- |
| Colour | Win probability |
| Lindt 90 | 95.6% |
| Amul Milk | 2.2% |
| Cadbury Dairy Milk | 1.5% |
| Nestle Classic | 0.7% |

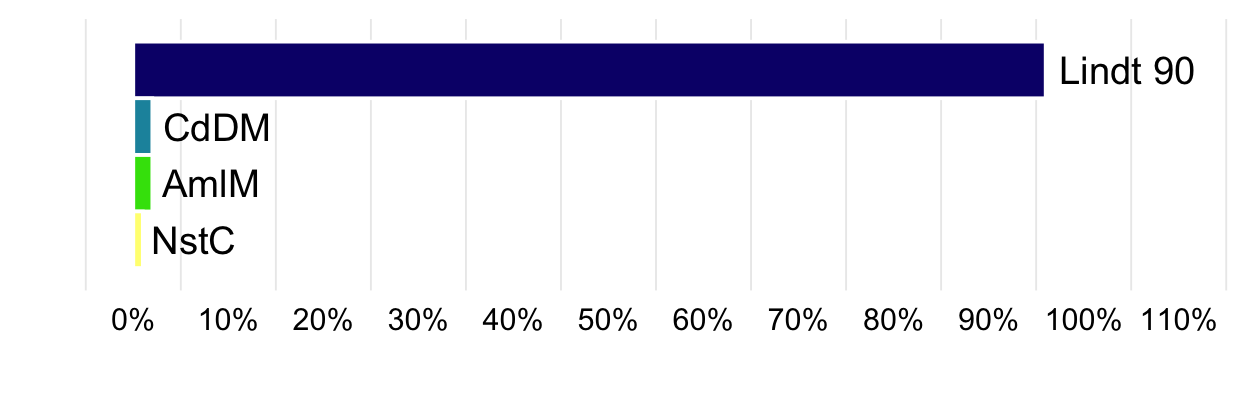


Figure 4.2.2. Probability of being the best ranked for ‘bitterness’.

**Cocoa content**

Table 4.3.1. Likelihood ratio test results from fitted Plackett-Luce model with rankings from ‘cocoa content’.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | logLikelihood | DF | Statistic | Pr(>Chisq) |
| NULL | -39.41871 | 66 | NA | NA |
| Cocoa content | -37.19718 | 63 | 4.44306 | 0.21743 |

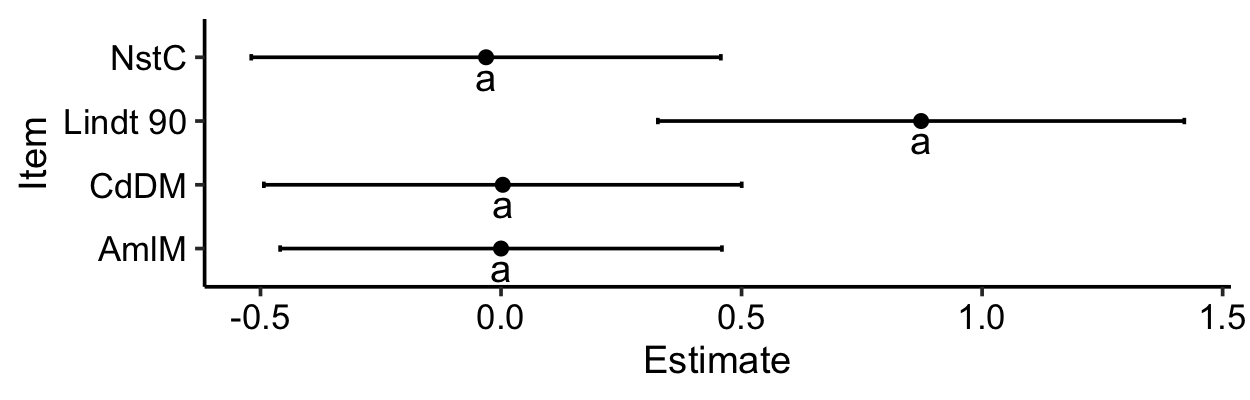


Figure 4.3.1. Model coefficients and mean separation of Plackett-Luce model for ‘cocoa content’ with 84% confidence intervals.

Table 4.3.2. Model coefficients and mean separation of Plackett-Luce model for ‘cocoa content’ with 84% confidence intervals.

|  |  |  |  |
| --- | --- | --- | --- |
| Colour | Estimate | quasiSE | Group |
| Lindt 90 | 0.8732 | 0.3894 | a |
| Cadbury Dairy Milk | 0.0036 | 0.3535 | a |
| Amul Milk | 0.0000 | 0.3268 | a |
| Nestle Classic | -0.0311 | 0.3474 | a |

Table 4.3.3 Percetage probability of being highest ranked for ‘cocoa content’.

|  |  |
| --- | --- |
| Colour | Win probability |
| Lindt 90 | 44.6% |
| Cadbury Dairy Milk | 18.7% |
| Amul Milk | 18.6% |
| Nestle Classic | 18.1% |

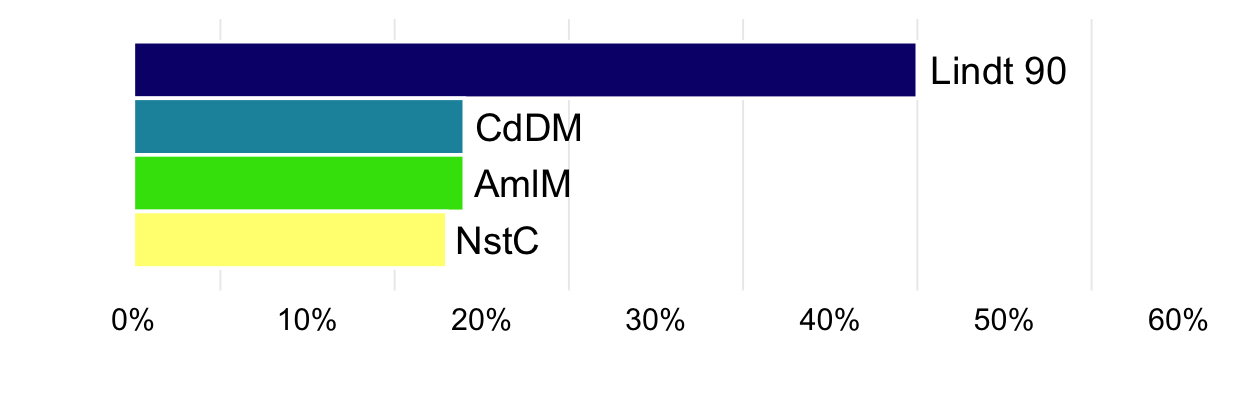


Figure 4.3.2. Probability of being the best ranked for ‘cocoa content’.

**Dark color**

Table 4.4.1. Likelihood ratio test results from fitted Plackett-Luce model with rankings from ‘dark color’.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | logLikelihood | DF | Statistic | Pr(>Chisq) |
| NULL | -39.41871 | 66 | NA | NA |
| Dark color | -13.97947 | 63 | 50.87847 | 5.1926e-11 \*\*\* |

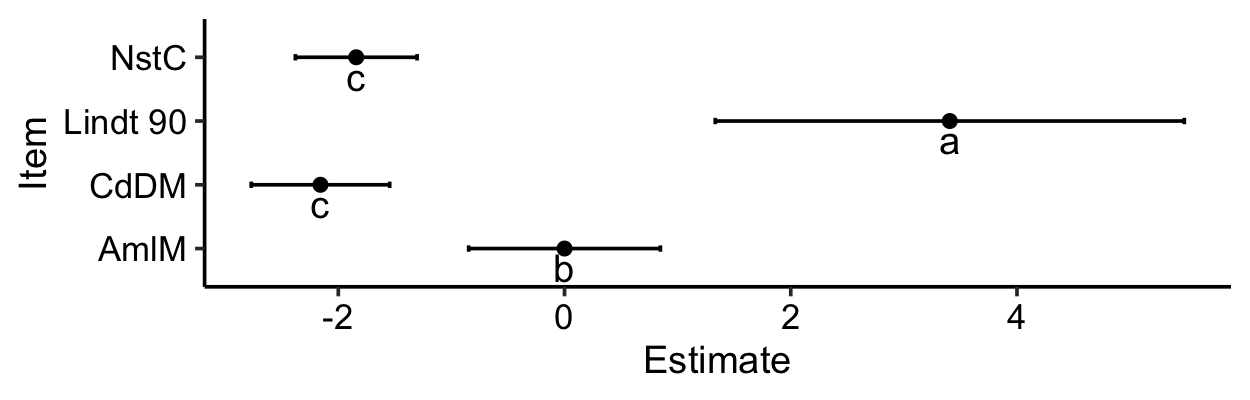


Figure 4.4.1. Model coefficients and mean separation of Plackett-Luce model for ‘dark color’ with 84% confidence intervals.

Table 4.4.2. Model coefficients and mean separation of Plackett-Luce model for ‘dark color’ with 84% confidence intervals.

|  |  |  |  |
| --- | --- | --- | --- |
| Colour | Estimate | quasiSE | Group |
| Lindt 90 | 3.4059 | 1.4756 | a |
| Amul Milk | 0.0000 | 0.6032 | b |
| Nestle Classic | -1.8416 | 0.3830 | c |
| Cadbury Dairy Milk | -2.1577 | 0.4354 | c |

Table 4.4.3 Percetage probability of being highest ranked for ‘dark color’.

|  |  |
| --- | --- |
| Colour | Win probability |
| Lindt 90 | 95.9% |
| Amul Milk | 3.2% |
| Nestle Classic | 0.5% |
| Cadbury Dairy Milk | 0.4% |

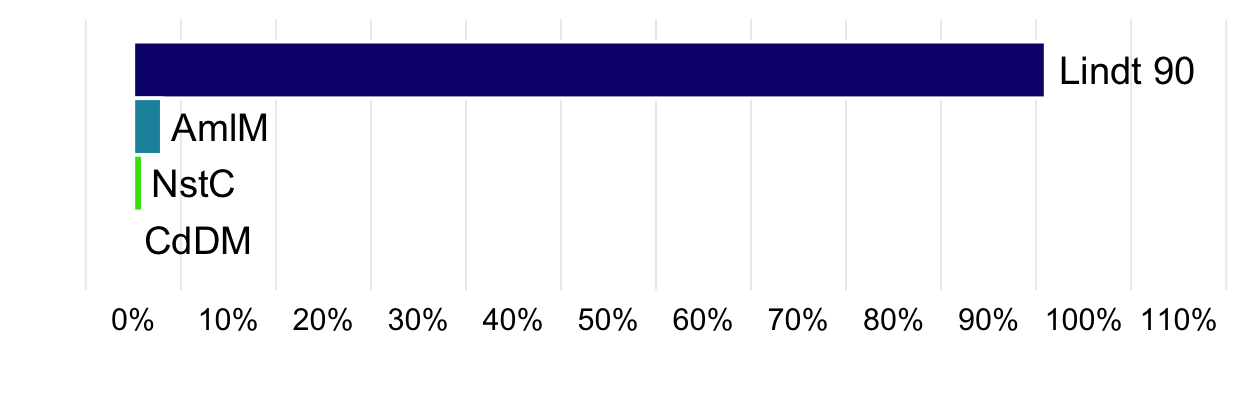


Figure 4.4.2. Probability of being the best ranked for ‘dark color’.

# Section 5: Plackett-Luce models with covariates

## Overall performance

A model-based recursive partitioning method (*24*) was used to determine which of the explanatory variables, if any, had significant relationships with the rankings. This approach identifies sub-groups in the data for which the rankings of the different varieties are significantly different to each other.

Table 5.1. Univariate p-values for first split in Plackett-Luce tree model for the overall ranking.

|  |  |
| --- | --- |
| Covariate | p.value |
| What is the gender ? | 0.979 |

No split was found for the Plackett-Luce model using the selected covariates. Figure 5.1 shows the model for the rankings provided in this project, showing that none of the tested covariates had an influence on the rankings, whith an alpha = 0.1.

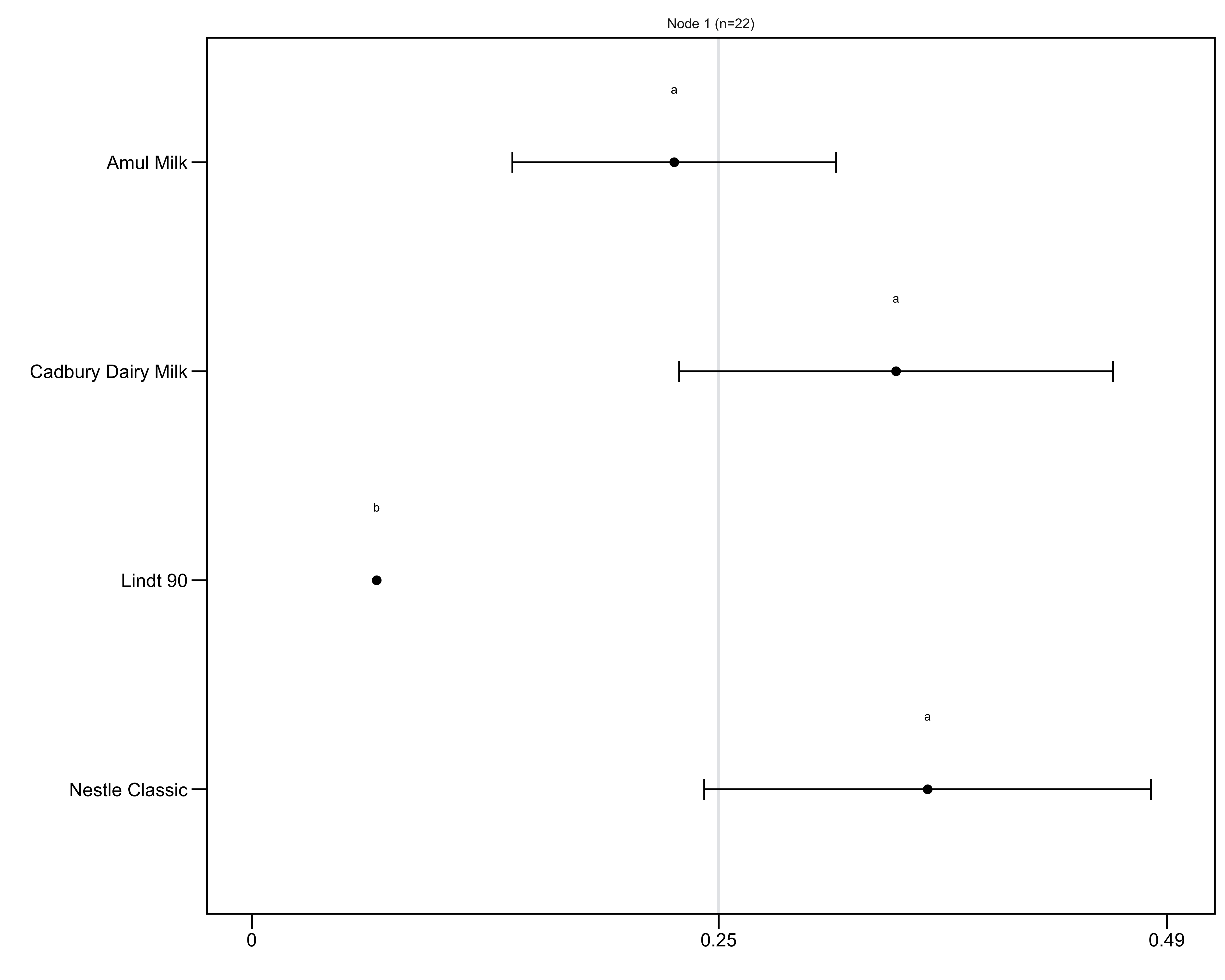


Figure 5.1. Plackett-Luce tree for the overall performance. The horizontal axis is the probability of winning. Error bars show quasi-SEs. The gray vertical line indicates the average probability of winning (1/number of items).

# References

1. J. van Etten, E. Beza, L. Calderer, K. Van Duijvendijk, C. Fadda, B. Fantahun, Y. G. Kidane, J. van de Gevel, A. Gupta, D. K. Mengistu, D. Kiambi, P. N. Mathur, L. Mercado, S. Mittra, M. J. Mollel, J. C. Rosas, J. Steinke, J. G. Suchini, K. S. Zimmerer, First experiences with a novel farmer citizen science approach: crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (tricot). *Experimental Agriculture*. **55**, 275–296 (2019).

2. R Core Team, R: A language and environment for statistical computing. version 3.6.2. (2019), (available at <https://r-project.org/>).

3. H. L. Turner, J. van Etten, D. Firth, I. Kosmidis, Modelling rankings in R: the PlackettLuce package. *Computational Statistics* (2020), doi:[10.1007/s00180-020-00959-3](https://doi.org/10.1007/s00180-020-00959-3).

4. Y. Xie, *Dynamic Documents with R and knitr* (Chapman; Hall/CRC, Boca Raton, Florida, ed. 2nd, 2015; <https://yihui.org/knitr/>).

5. Y. Xie, J. J. Allaire, G. Grolemund, *R Markdown: The Definitive Guide* (Chapman; Hall/CRC, Boca Raton, Florida, 2018; <https://bookdown.org/yihui/rmarkdown>).

6. K. de Sousa, J. van Etten, ClimMobTools: API Client for the ’ClimMob’ platform (2020), (available at <https://CRAN.R-project.org/package=ClimMobTools>).

7. K. de Sousa, J. van Etten, S. Dumble, N. Greliche, J. Steinke, gosset: Modelling Metadata and Crowdsourced Citizen Science (2020), (available at <https://agrobioinfoservices.github.io/gosset/>).

8. G. R. Warnes, B. Bolker, T. Lumley, gtools: Various R Programming Tools (2018), (available at <https://CRAN.R-project.org/package=gtools>).

9. J. Ooms, The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. *arXiv:1403.2805 [stat.CO]* (2014) (available at <https://arxiv.org/abs/1403.2805>).

10. A. Zeileis, T. Hothorn, K. Hornik, Model-Based Recursive Partitioning. *Journal of Computational and Graphical Statistics*. **17**, 492–514 (2008).

11. J. Kopf, C. Strobl, A. Zeileis, Anchor Selection Strategies for DIF Analysis: Review, Assessment, and New Approaches. *Educational and Psychological Measurement*. **75**, 22–56 (2015).

12. D. Firth, qvcalc: Quasi Variances for Factor Effects in Statistical Models (2020), (available at <https://CRAN.R-project.org/package=qvcalc>).

13. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal*. **Complex Systems**, 1695 (2006).

14. M. Borkovec, N. Madin, *ggparty: ’ggplot’ Visualizations for the ’partykit’ Package* (2019; <https://CRAN.R-project.org/package=ggparty>).

15. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016; <https://ggplot2.tidyverse.org>).

16. K. Slowikowski, *ggrepel: Automatically Position Non-Overlapping Text Labels with ’ggplot2’* (2020; <https://CRAN.R-project.org/package=ggrepel>).

17. J. Cheng, B. Karambelkar, Y. Xie, *leaflet: Create Interactive Web Maps with the JavaScript ’Leaflet’ Library* (2019; <https://CRAN.R-project.org/package=leaflet>).

18. S. Graves, H.-P. Piepho, L. S. with help from Sundar Dorai-Raj, multcompView: Visualizations of Paired Comparisons (2019), (available at <https://CRAN.R-project.org/package=multcompView>).

19. T. L. Pedersen, *patchwork: The Composer of Plots* (2019; <https://CRAN.R-project.org/package=patchwork>).

20. B.-H. Mevik, R. Wehrens, K. H. Liland, pls: Partial Least Squares and Principal Component Regression (2019), (available at <https://CRAN.R-project.org/package=pls>).

21. J. van Etten, C. Quirós, K. de Sousa, J. Steinke, S. Dumble, B. Madriz, A. Coto, J. van de Gevel, N. Greliche, N. Murchan, J. Hook, ClimMob: Software to support experimental citizen science in agriculture (2020), (available at <https://climmob.net/climmob3>).

22. M. G. Kendall, A new measure of ranking correlation. *Biometrika*. **30**, 81–93 (1938).

23. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. **57**, 289–300 (1995).

24. C. Strobl, J. Malley, G. Tutz, An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods*. **14**, 323–48 (2009).