# Supplementary Material

## Aims of Analysis

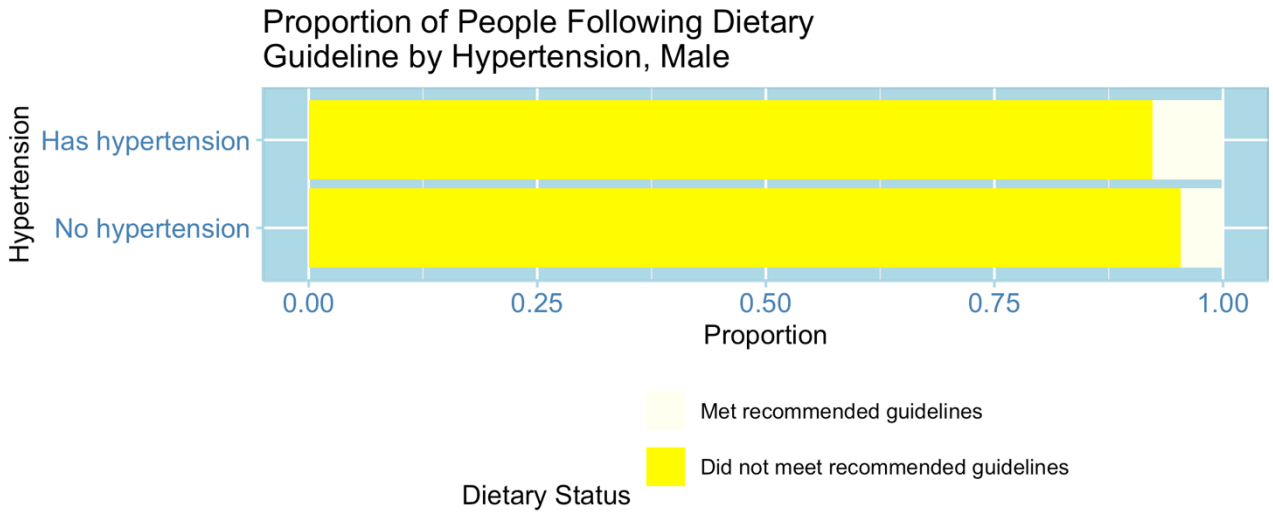
The aim of the following analysis was to obtain importance scores for a range of biomarkers and nutritional data in relation to Australians aged over 60 both with and without hypertension.

## Description of Data

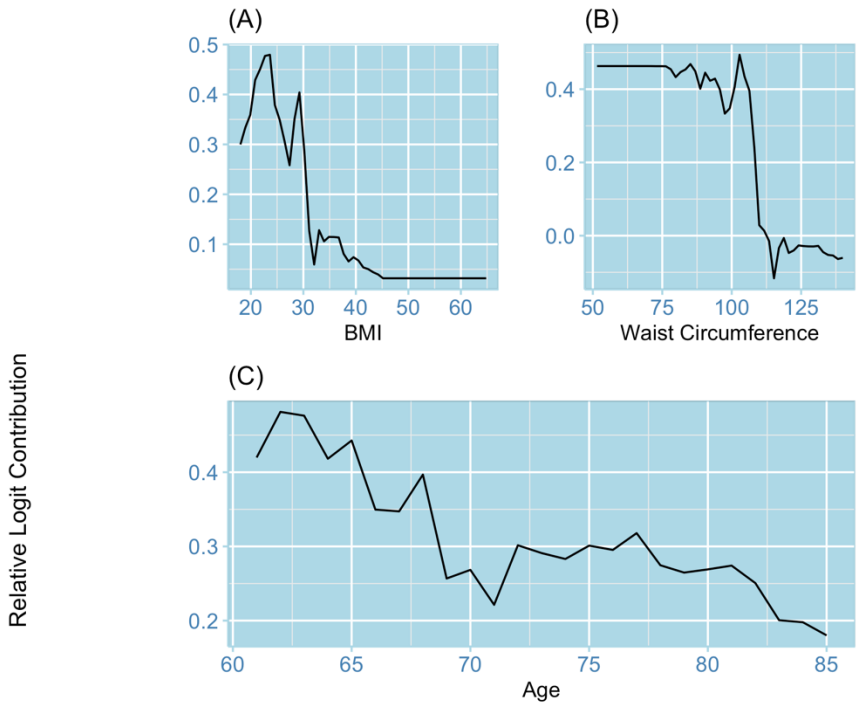
Data was obtained from the Australian Health Survey 2019 (Australian Bureau of Statistics). It contained a variety of biomedical factors and nutritional details obtained from participants. The data was cleaned then the sample size reduced to include only participants over 60 years old.

## Supplementary Figures

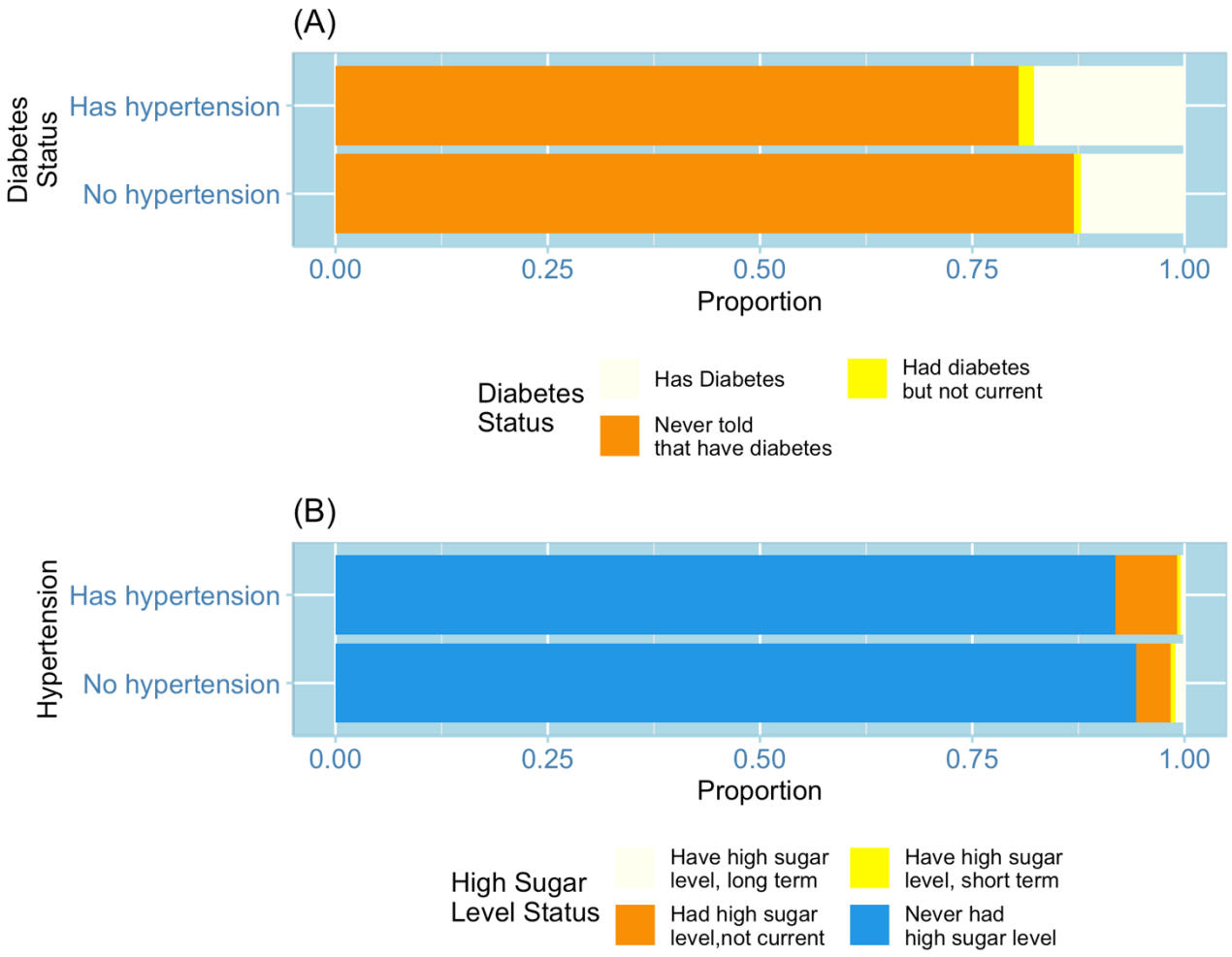
### Male Biomedical Data



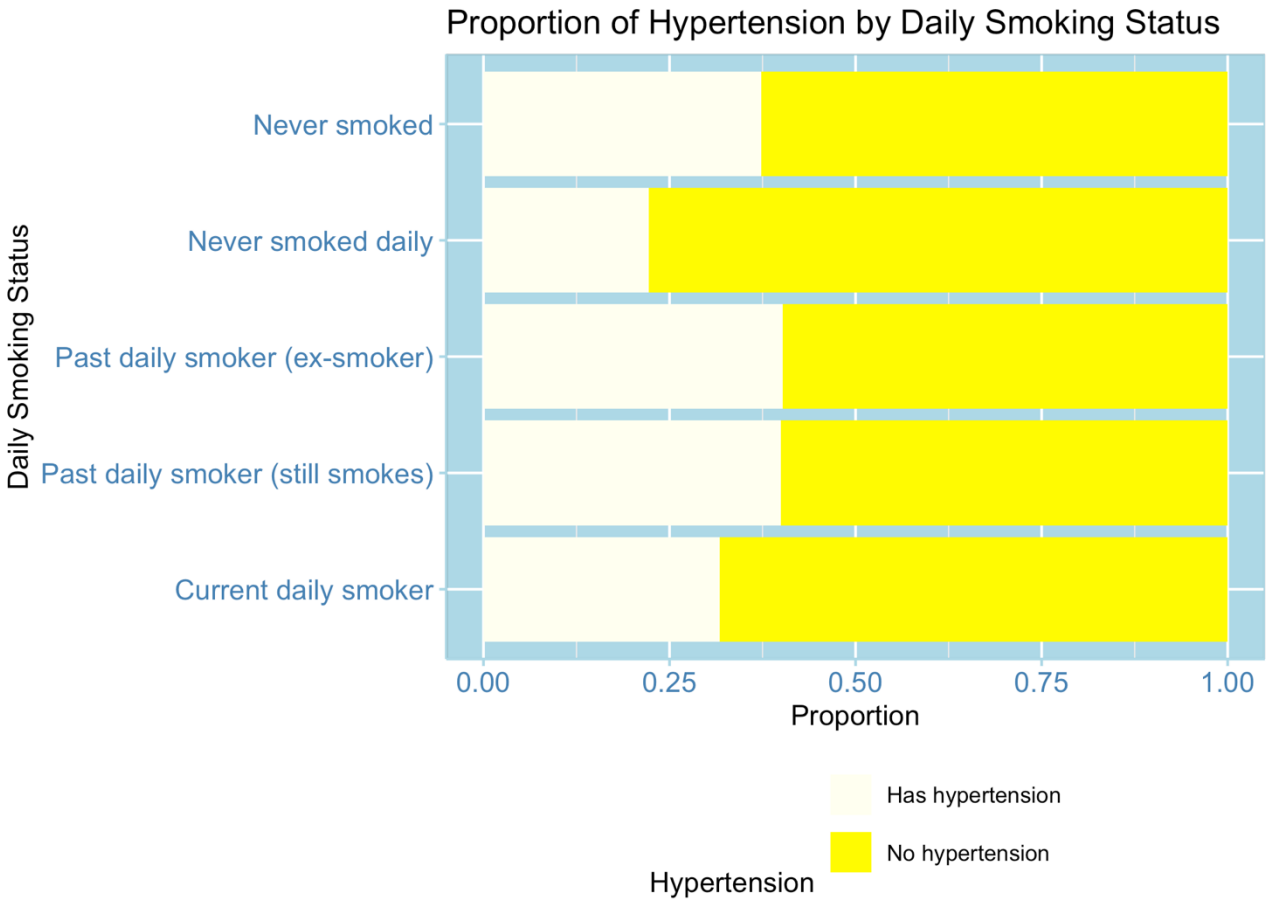

*Supplementary Figure 1. Proportion of Males following dietary guidelines. This plot shows the proportion of males following dietary guidelines by hypertension status. Regardless of having hypertension, most males did not follow guidelines.*



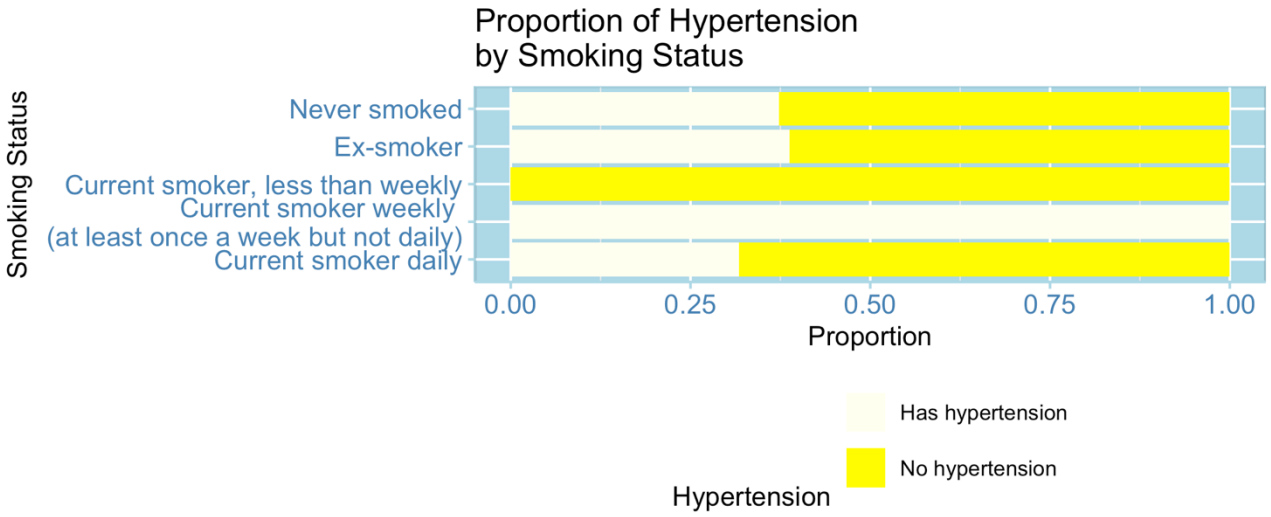

*Supplementary Figure 2. Partial Dependence Plot – BMI, WC, and Age. This partial dependence plot explains the relative contribution of variables to having no hypertension. A higher value indicates less likelihood of being classified with hypertension. Hence, all three plots show that with larger values, the less hypertension can be avoided.*



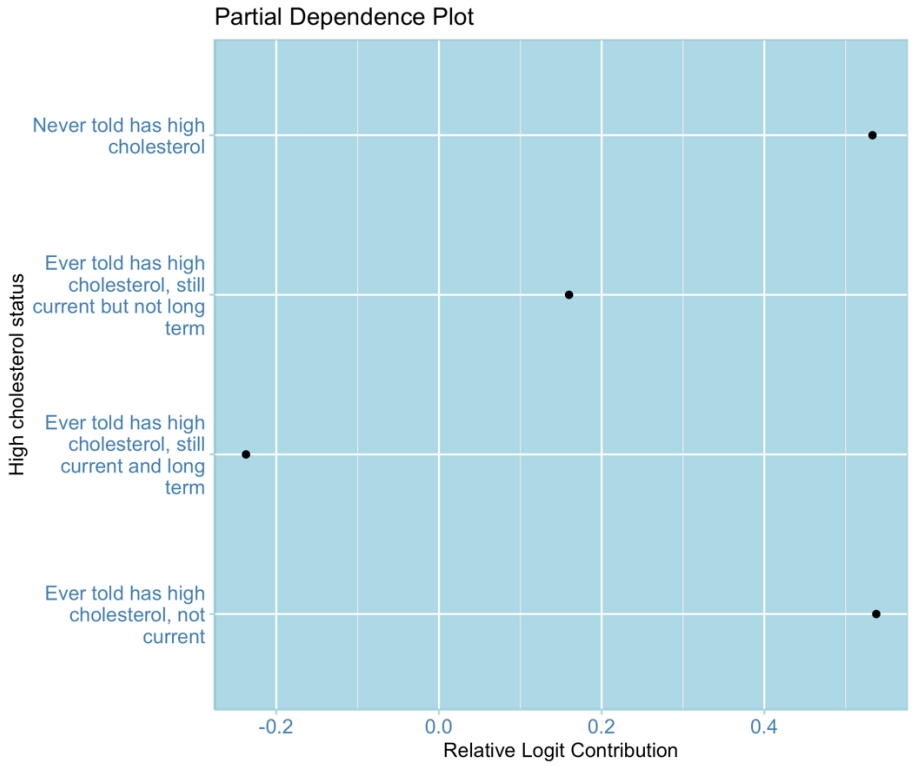

*Supplementary Figure 3. Proportion plots by Hypertension – Diabetes Status and High Sugar Level Status. These two plots aid in understanding why diabetes and sugar level have low importance in the classification model. Regardless of hypertension, most survey participants were never told of having diabetes or high sugar levels.*



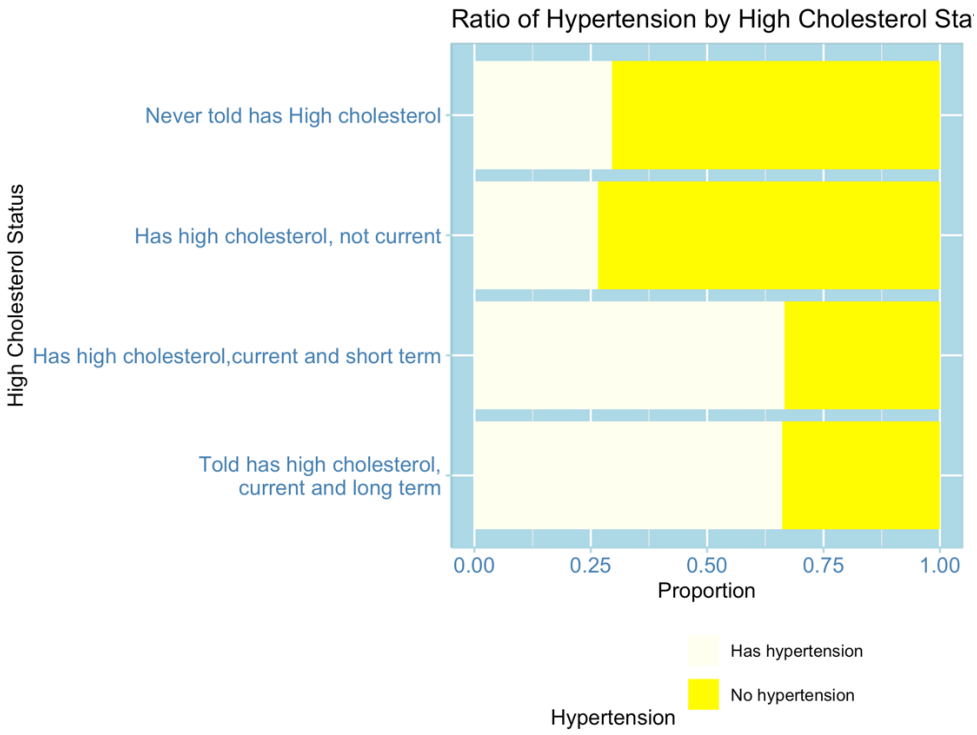

*Supplementary Figure 4. Proportion of people having Hypertension by Daily Smoking Status. This plot explains the proportion of participants' daily smoking status. In all categories, there are more people with hypertension but no significant differences between groups.*



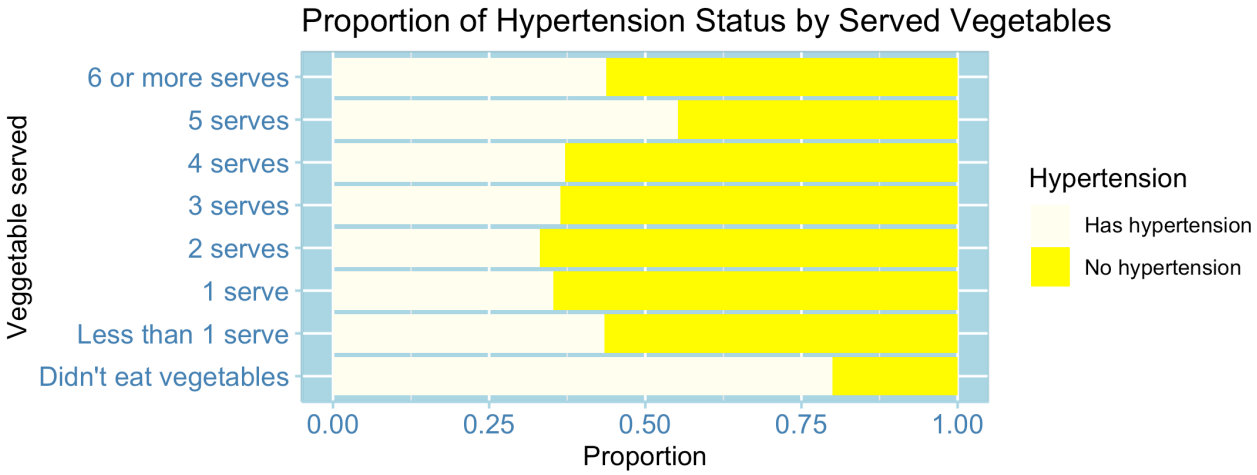

*Supplementary Figure 5. Proportion of people with Hypertension by Smoking Status. This plot shows the ratio of survey participants' smoking status. While less than weekly smokers had no hypertension, weekly smokers all have hypertension.*



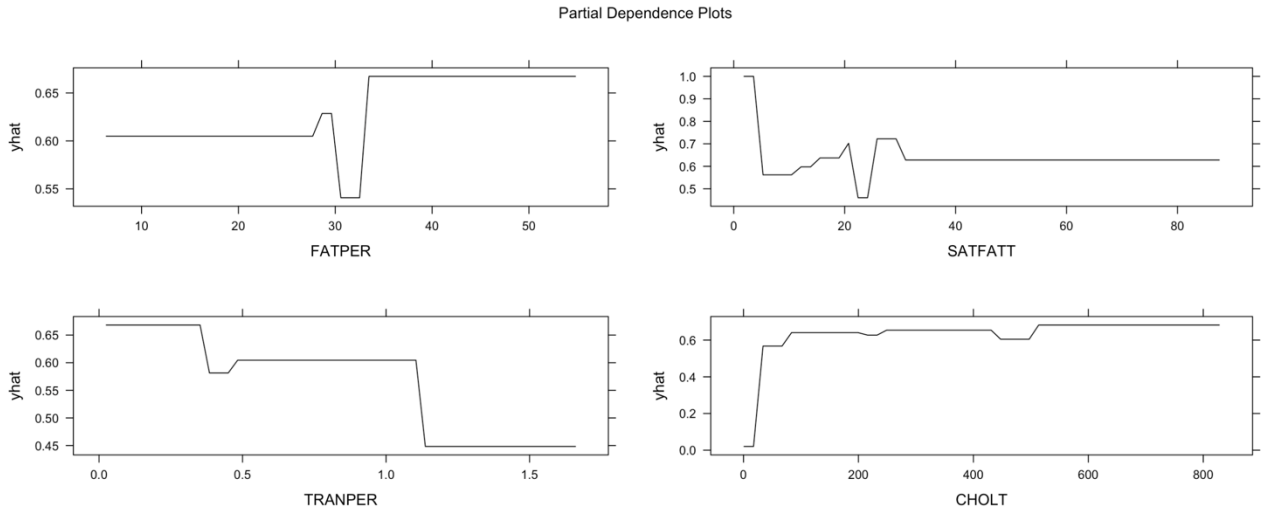

*Supplementary Figure 6. Partial Dependence Plot of High Cholesterol Status. This plot explains how high cholesterol contributes to the prevention of hypertension. People who did not have high cholesterol or had it short term were less likely to have hypertension.*



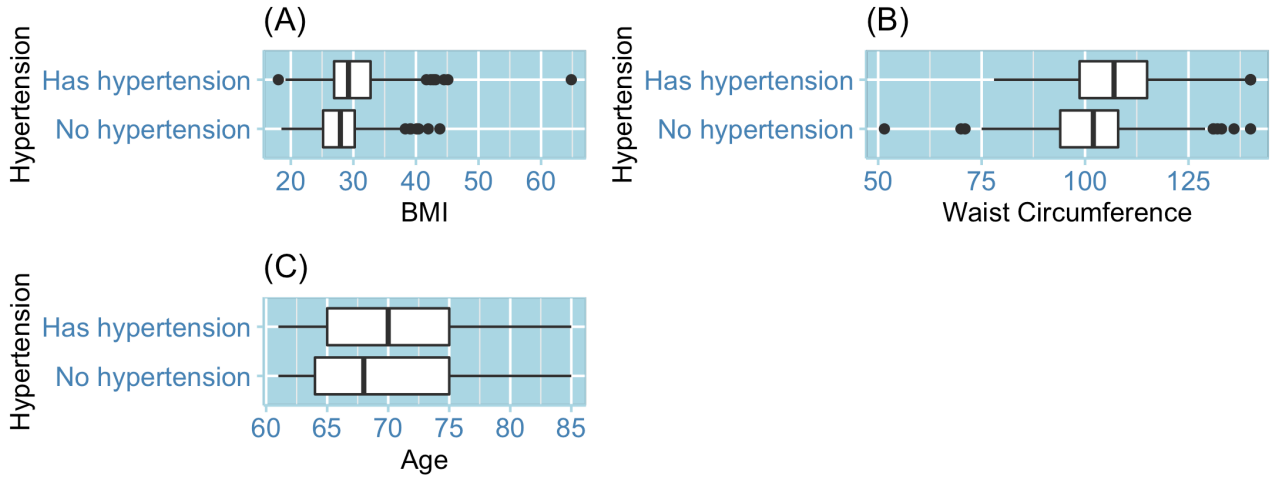

*Supplementary Figure 7. Proportion of Hypertension by Cholesterol Status. This plot displays the ratio of people having hypertension by cholesterol status. It shows that the proportion of people having hypertension and high cholesterol is large. There could be a correlation between the two.*




*Supplementary Figure 8. Proportion of Hypertension by Served Vegetables. This plot shows the ratio of participants having hypertension by the vegetable servings eaten. People eating no vegetables or five serves per day had a higher proportion of hypertension.*



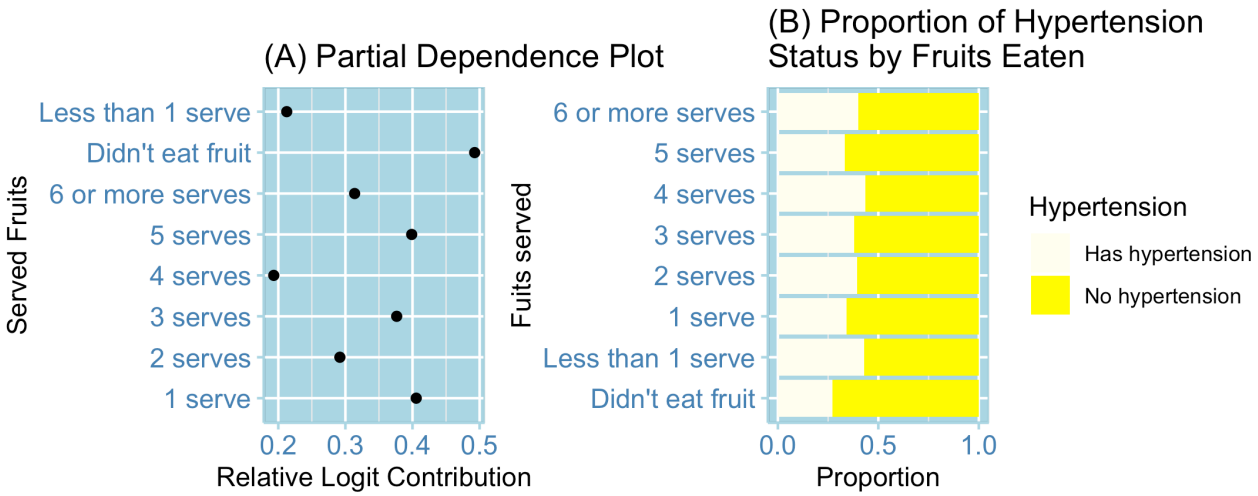

*Supplementary Figure 9. Partial Dependence Plots. Individual dependence relationships between the predictor variable and the outcome variable. An increase in the percentage of energy from fats and increase in cholesterol show an increased likelihood of being classified with no hypertension. The reverse is true for saturated fat consumption and trans fatty acid consumption. These both showed decreasing trends. These plots were made from a rpart object that was optimised using the train function from Caret. Partial from the PDP package was then used to for the above plots.*




*Supplementary Figure 10. Boxplot of Numerical Biomarkers, Male – BMI, Age, WC. The boxplots display some differences between the two groups of three variables. People with hypertension have higher BMI, WC, and are somewhat older (median values are larger).*

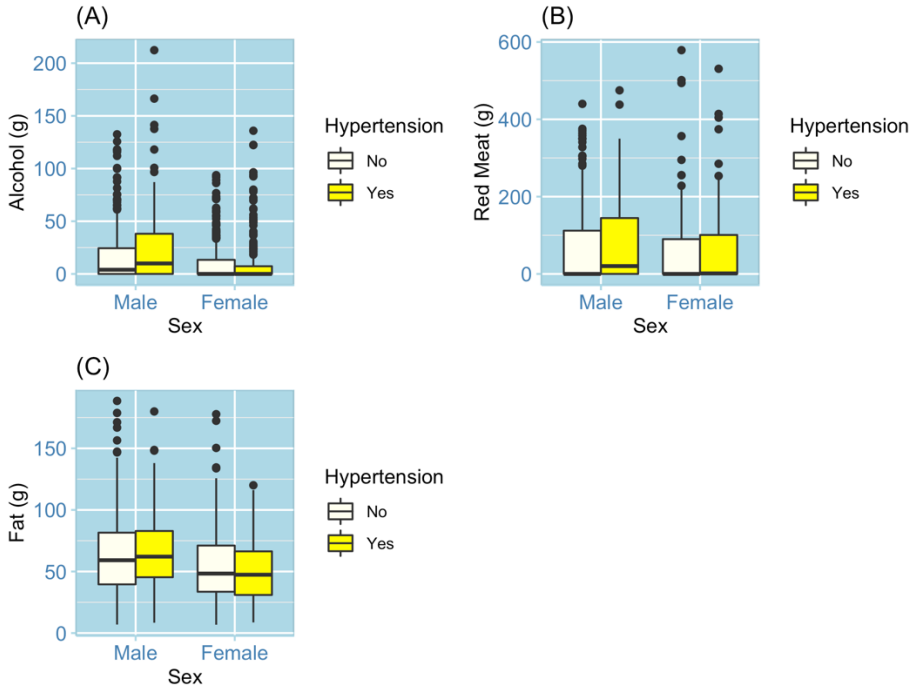
*Supplementary Table 1. Hypothesis Testing Results from Wilcoxon Sum-Rank test of mean nutrition intake between people with and without. Null hypothesis assumes equal mean intake between the groups while alternative hypothesis claims people with hypertension have higher intake.*

| Variable | P value | Keep Null Hypothesis | Sex |
|---|---|---|---|
| BMI | 1.8e-07 | FALSE | male |
| Waist Circumference | 2e-08 | FALSE | male |
| Age | 0.0348 | FALSE | male |
| BMI | 0.0014 | FALSE | female |
| Age | 4e-04 | FALSE | female |
| Waist Circumference | 0.0011 | FALSE | female |
| Red Meat | 0.0205 | FALSE | male |
| Fat (g) | 0.1226 | TRUE | male |
| Alcohol (g) | 7e-04 | FALSE | male |
| Sodium | 0.3442 | TRUE | male |
| Caffeine | 0.0461 | FALSE | male |
| Cholesterol | 0.4264 | TRUE | male |
| Saturated Fat | 0.1037 | TRUE | male |
| Polyunsaturated Fat (g) | 0.3828 | TRUE | male |
| Trans Fatty Acid | 0.368 | TRUE | male |
| Fat % | 0.2326 | TRUE | male |
| Alcohol % | 5e-04 | FALSE | male |
| Saturated Fat % | 0.3787 | TRUE | male |
| Trans Fatty Acid % | 0.6231 | TRUE | male |
| Polyunsaturated Fat % | 0.6413 | TRUE | male |
| Water | 0.2921 | TRUE | male |
| Fat | 0.8232 | TRUE | female |
| Red Meat | 0.1441 | TRUE | female |
| Alcohol | 0.9782 | TRUE | female |
| Sodium | 0.3493 | TRUE | female |
| Caffeine | 0.964 | TRUE | female |
| Cholesterol | 0.673 | TRUE | female |
| Saturated Fat | 0.8991 | TRUE | female |
| Polyunaturated Fat | 0.4424 | TRUE | female |
| Trans Fatty Acid | 0.6259 | TRUE | female |
| Fat % | 0.4427 | TRUE | female |
| Alcohol % | 0.9766 | TRUE | female |
| Saturated Fat % | 0.6225 | TRUE | female |
| Trans Fatty Acid % | 0.2631 | TRUE | female |
| Polyunsaturated Fat % | 0.1057 | TRUE | female |
| Water | 0.4568 | TRUE | female |

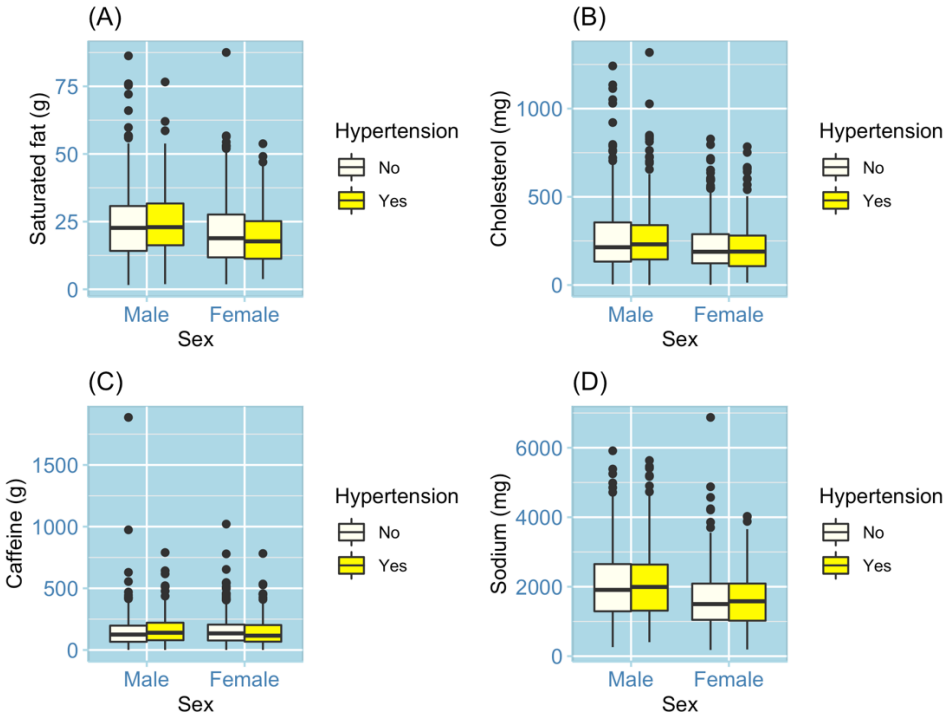



*Supplementary Figure 11. Partial Dependence Plot and Proportion Plot for Number of Fruit Servings.* ***A.*** *Partial dependence plot shows the relative contribution of eaten fruits to the prevention of hypertension. No unique pattern could be concluded, possibly due to the quality of the data.* ***B****. There are no significant differences between the two groups as the proportions of people having hypertension are similar.*

**Male Nutrition Data**

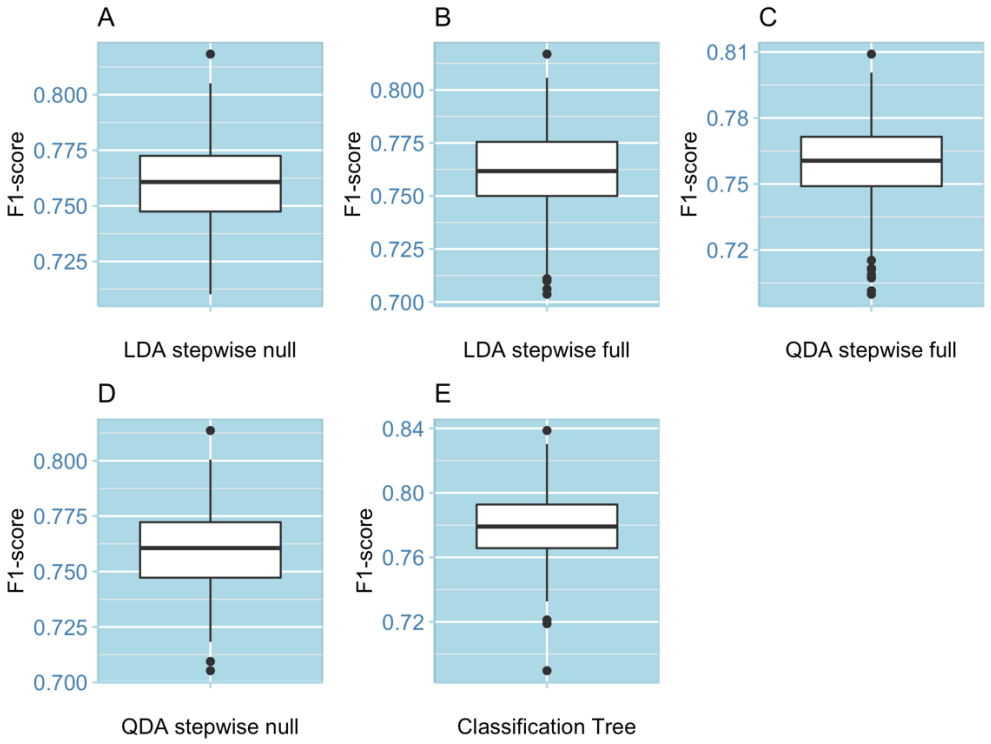



*Supplementary Figure 12. Boxplot of Red Meat, Alcohol, and Fat for Males and Female. The boxplots compare consumption of red meat, alcohol, and fat; sex; and people with and without hypertension. The boxplots' position, size, and median value suggest that males consume more of all three nutrition variables. Moreover, males with hypertension had higher consumption of all these.*

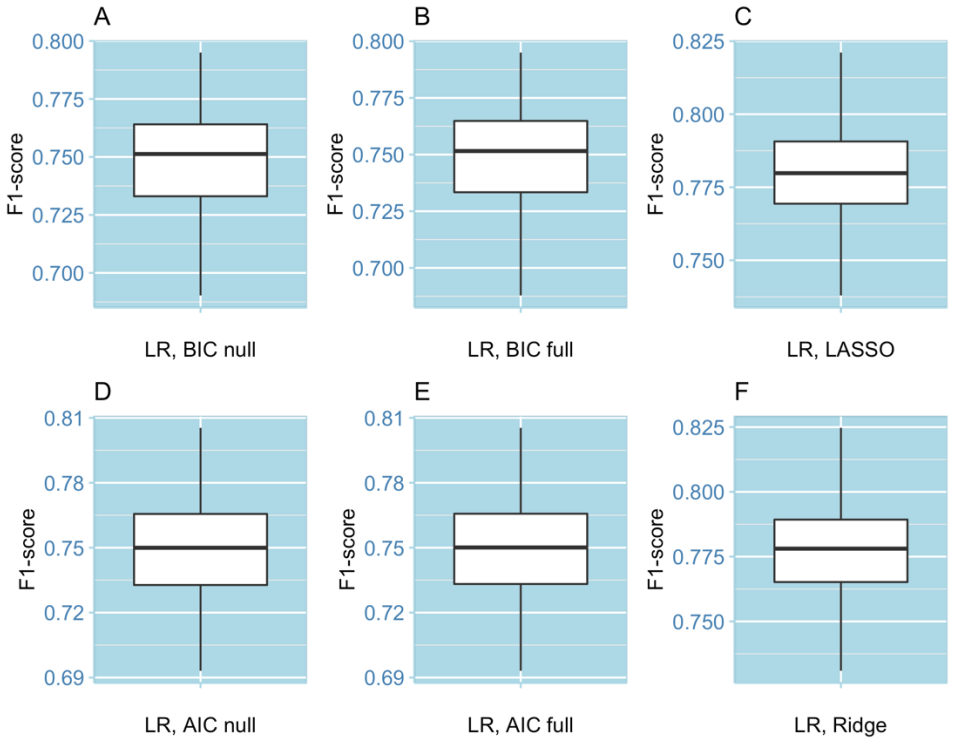



*Supplementary Figure 13. Boxplots of Saturated Fat, Cholesterol, Caffeine, and Sodium for males and Females. The boxplots compare consumption of these with sex and hypertension status. The position, size, and median values suggest that males consume more of these four items. Moreover, males with hypertension have higher consumption values.*

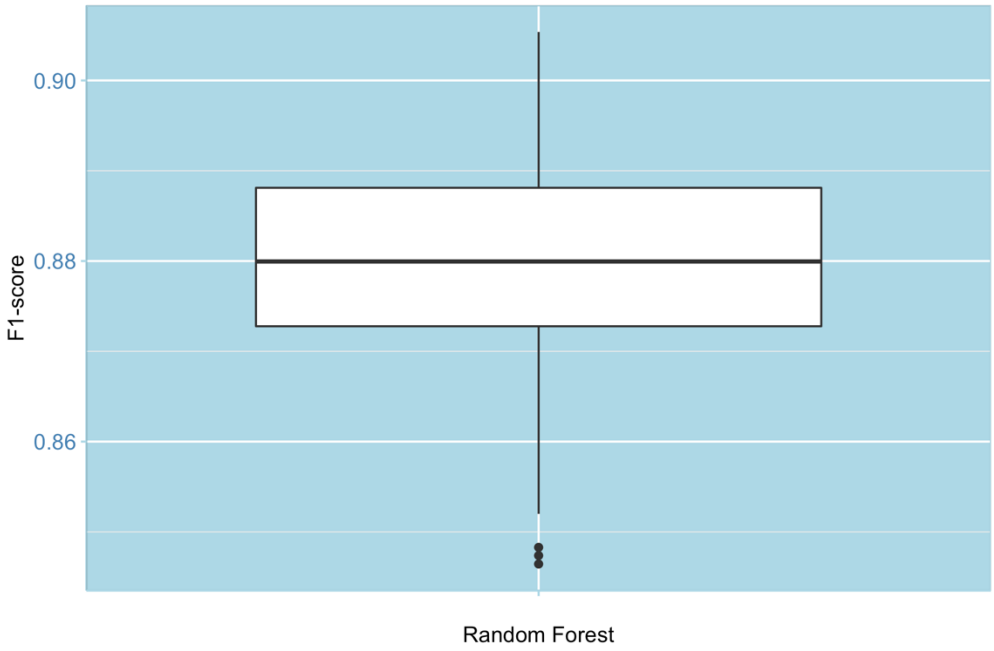
**Model Performance**



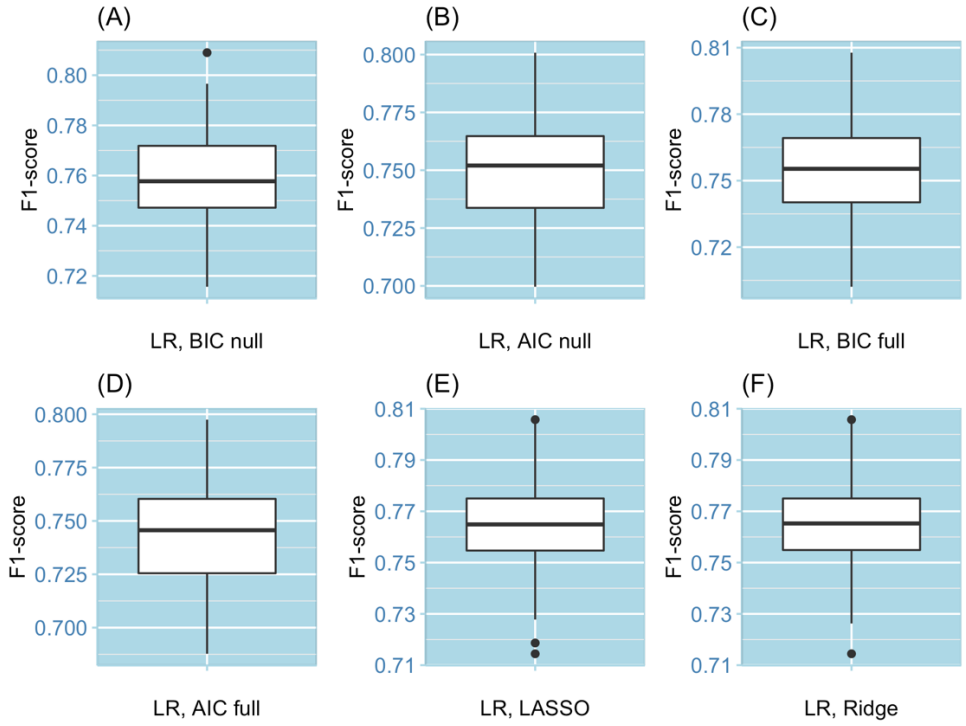

*Supplementary Figure 14. Model Performance on Male Biomedical Data 1. Comparing different classification models by F1-score. Repeated five cross-fold validation was used 200 times and, in each iteration, mean F1-score was kept. The boxplots show the distribution of those values.*



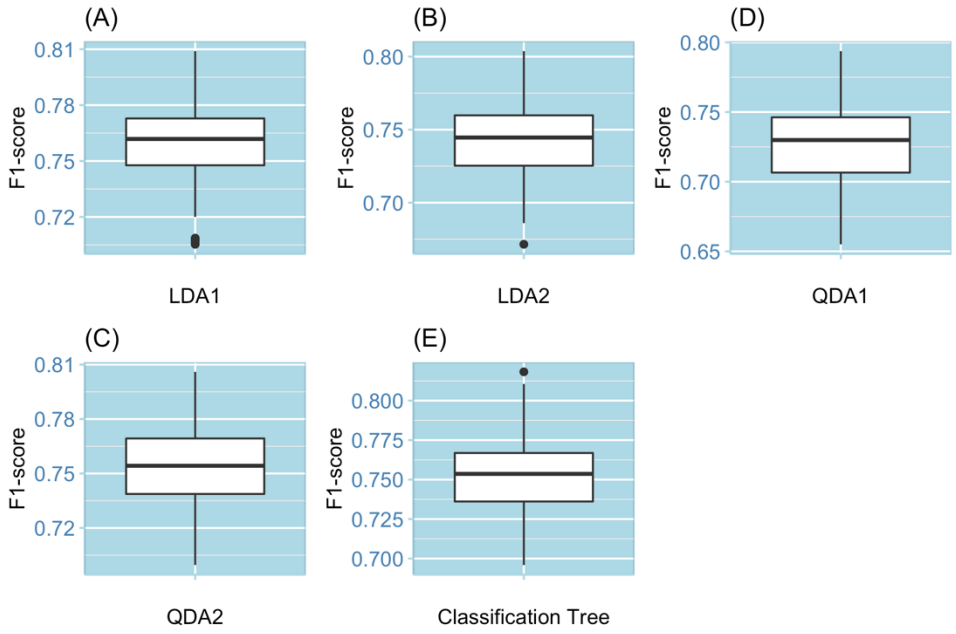

*Supplementary Figure 15. Model Performance on Male Biomedical Data 2. Comparing different classification models by F1-score. Repeated five cross-fold validation was used 200 times and, in each iteration, mean F1-score was kept. The boxplots show the distribution of those values.*



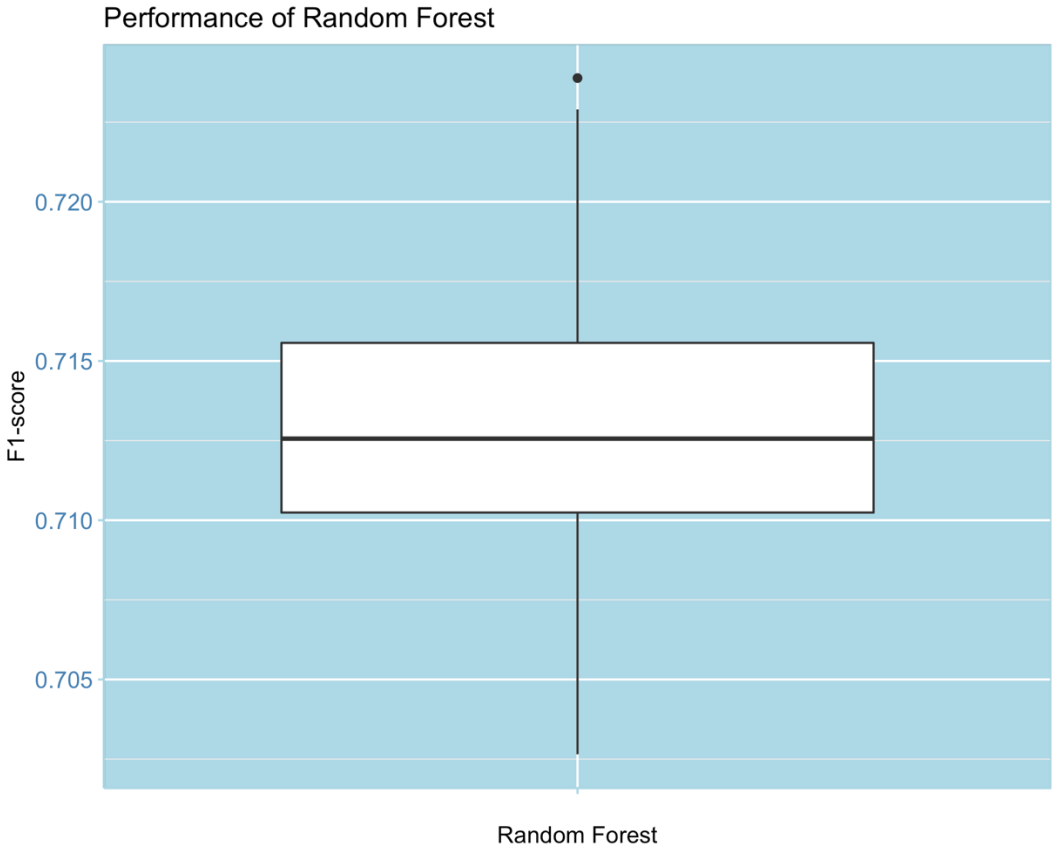

*Supplementary Figure 16. Random Forest Model Performance on Male Biomedical Data. Boxplot explains the performance of random forest classification model. Performance was measured with 200 repeated five cross-fold validation. The median accuracy is 88%, and excluding the outliers, the model performed about 86% consistently.*




*Supplementary Figure 17. Model Performance on Nutrition Data – Logistic Regression. Comparing different classification models by F1-score. Repeated five cross-fold validation was used 200 times and, in each iteration, mean F1-score was kept. The boxplots show the distribution of those values.*

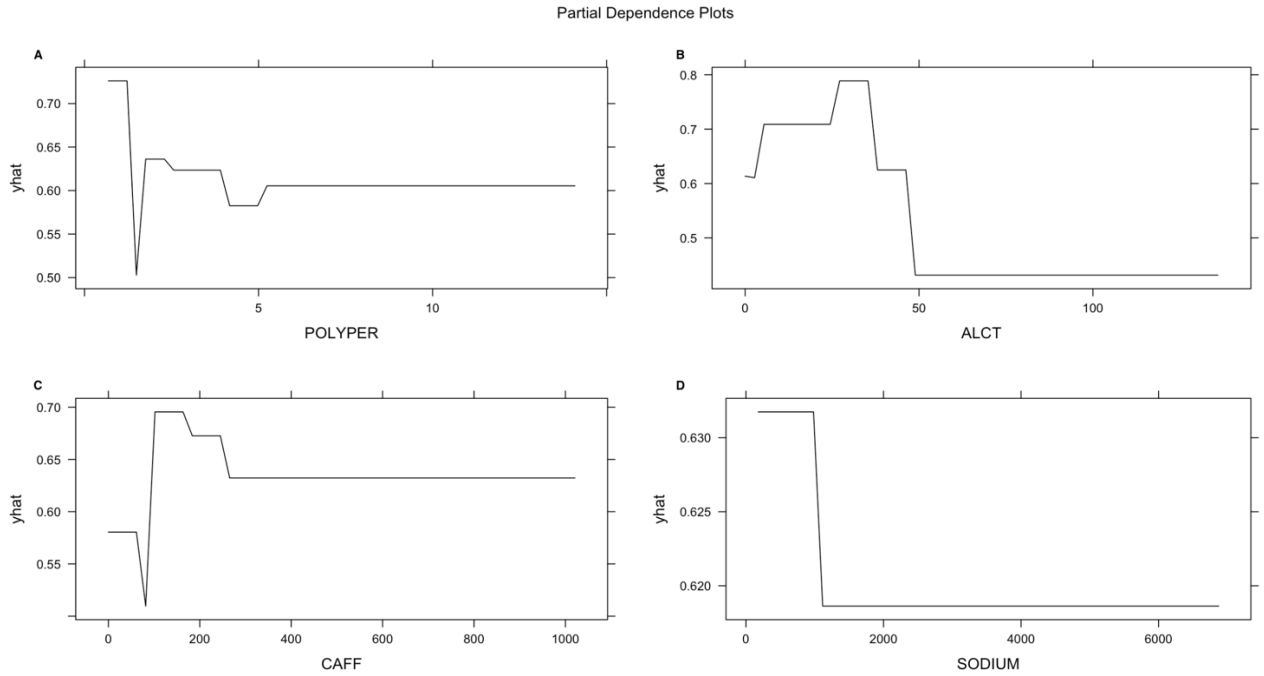



*Supplementary Figure 18. Model Performance on Nutrition Data – Linear Discriminant Analysis, Quadratic Discriminant Analysis, Logistic Regression. Repeated five cross-fold validation was used 200 times and, in each iteration, mean F1-score was kept. The boxplots show the distribution of those values.*

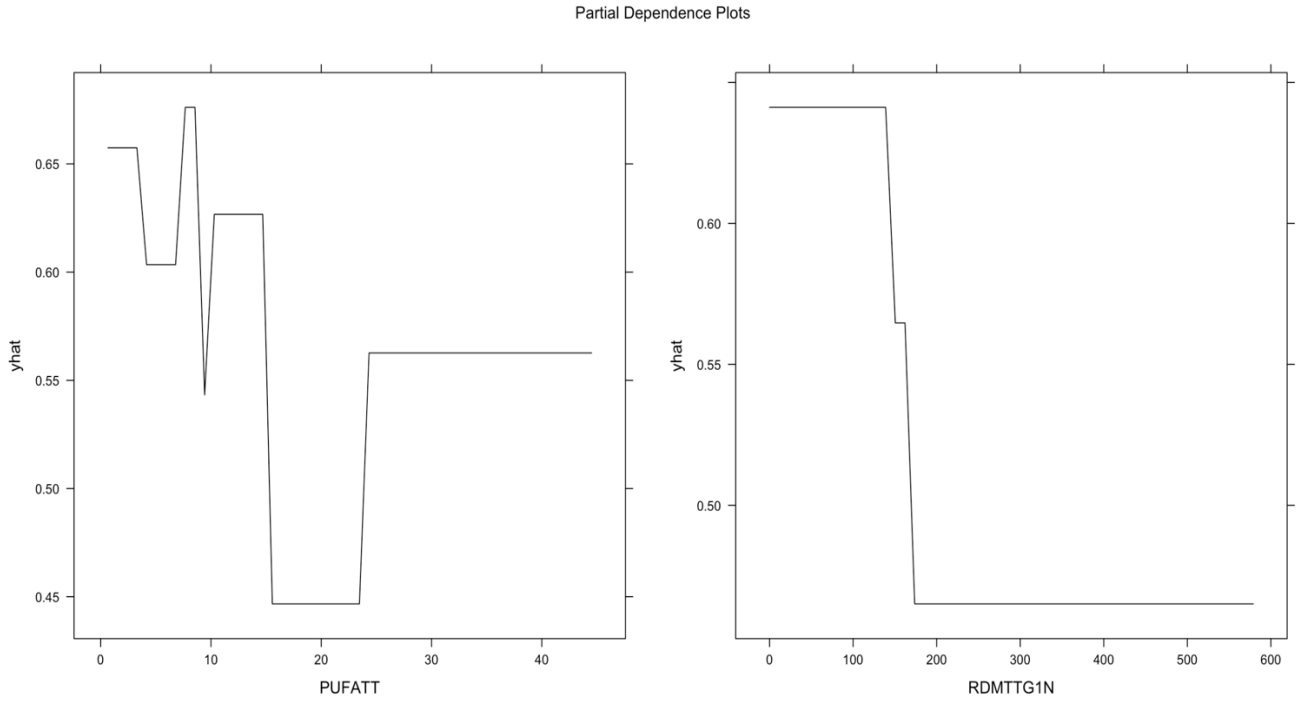



*Supplementary Figure 19. Random Forest Performance on Nutrition Data, Males. Boxplot explains the performance of the random forest classification model on nutrition data. Performance was measured with 200-time repeated five cross-fold validation. Accuracy was 77%.*

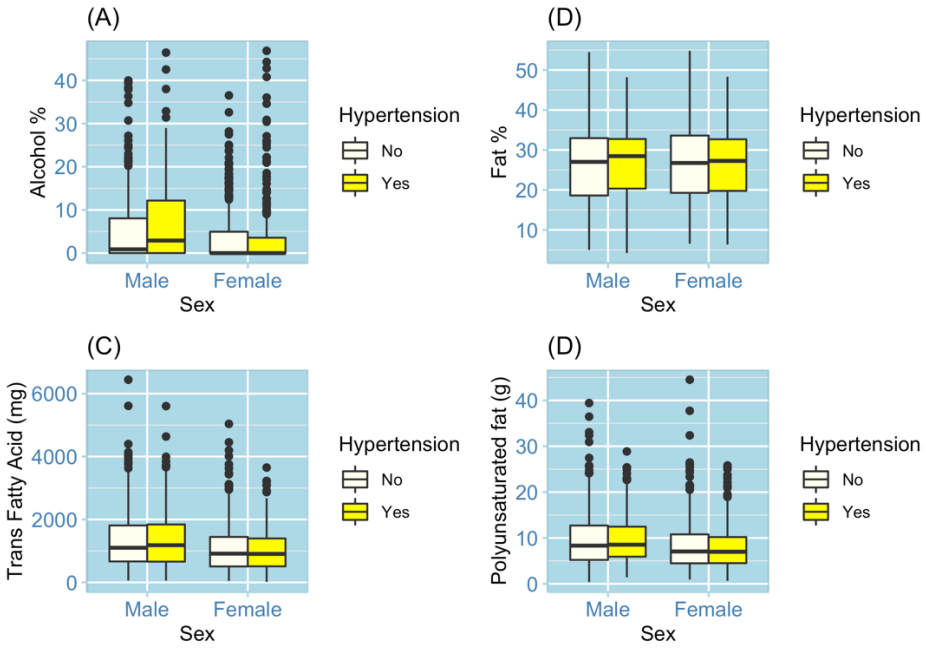
**Female Nutrition Data**



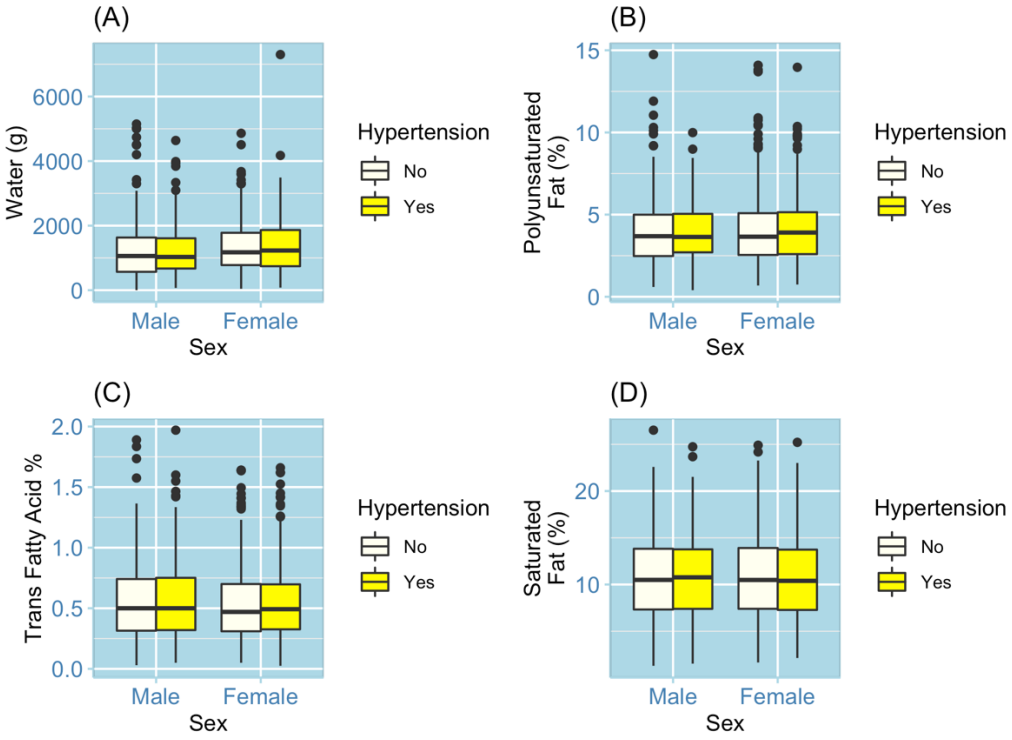

*Supplementary Figure 20. Female Nutrition Data. A, B, and D reflect a decreasing likelihood of hypertension as a result of increases in the percentage of polyunsaturated fats, alcohol, and sodium consumed. C – increased consumption of caffeine tended towards a greater likelihood of hypertension compared to no caffeine consumption. These plots were created using the Partial function from the PDP package, using a rpart object as the argument.*




*Supplementary Figure 21. Female Nutrition Data. Partial dependence plots serving the same purpose as Fig S20. They show relationships between the feature of interest and the predictor variables. Created using rpart object in conjunction with Partial from the PDP package. As red meat and polyunsaturated fat consumption increase, there is a decrease in likelihood of being classified with hypertension.*




*Supplementary Figure 22. Boxplot of Health Survey Data – Alcohol (%), Fat (%), Trans Fatty Acid (mg), Polyunsaturated Fat (g). The boxplots compare nutrition consumption of sexes and people with/without hypertension from the Australian Health Survey data. Overall, the boxplots' position, size and median value suggest that males consume more or even of the above-mentioned nutrients. Moreover, males with hypertension have higher consumption values.*




*Supplementary Figure 23. Boxplots of Water (g), Polyunsaturated Fat (%), Trans Fatty Acid (%), and Saturated Fat (%). Boxplots compare consumption of these with sexes and hypertension status. Boxplot position, size, and median value suggest that males consume more than females for the above nutrients. Moreover, males with hypertension have higher consumption values.*