Post-Data Simulation ML Pipeline Analysis
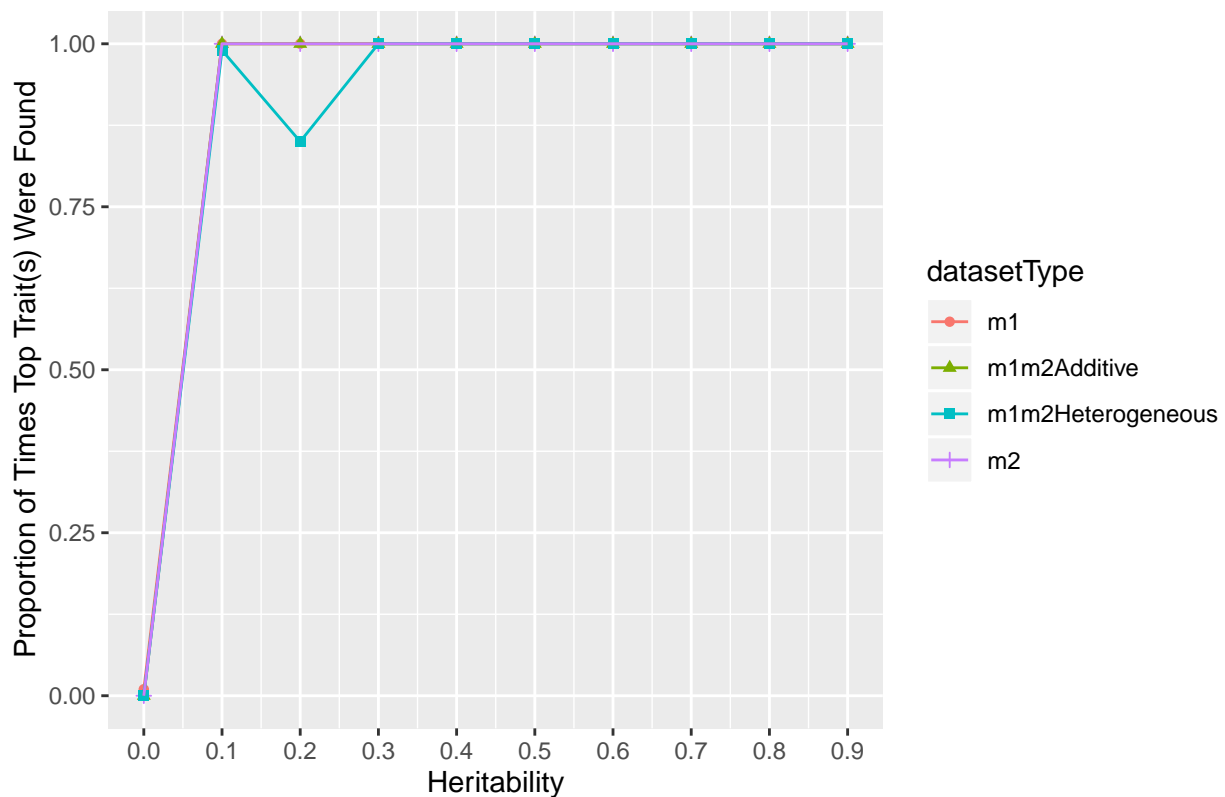
Vivek Sriram

11/22/2019

In this markdown document, we will perform machine learning/feature importance analysis on the results of our GAMETES simulated data. We will then produce plots that give us summary information on the behavior of machine learners on our data, both in terms of accuracy as well as for picking the correct top feature(s).
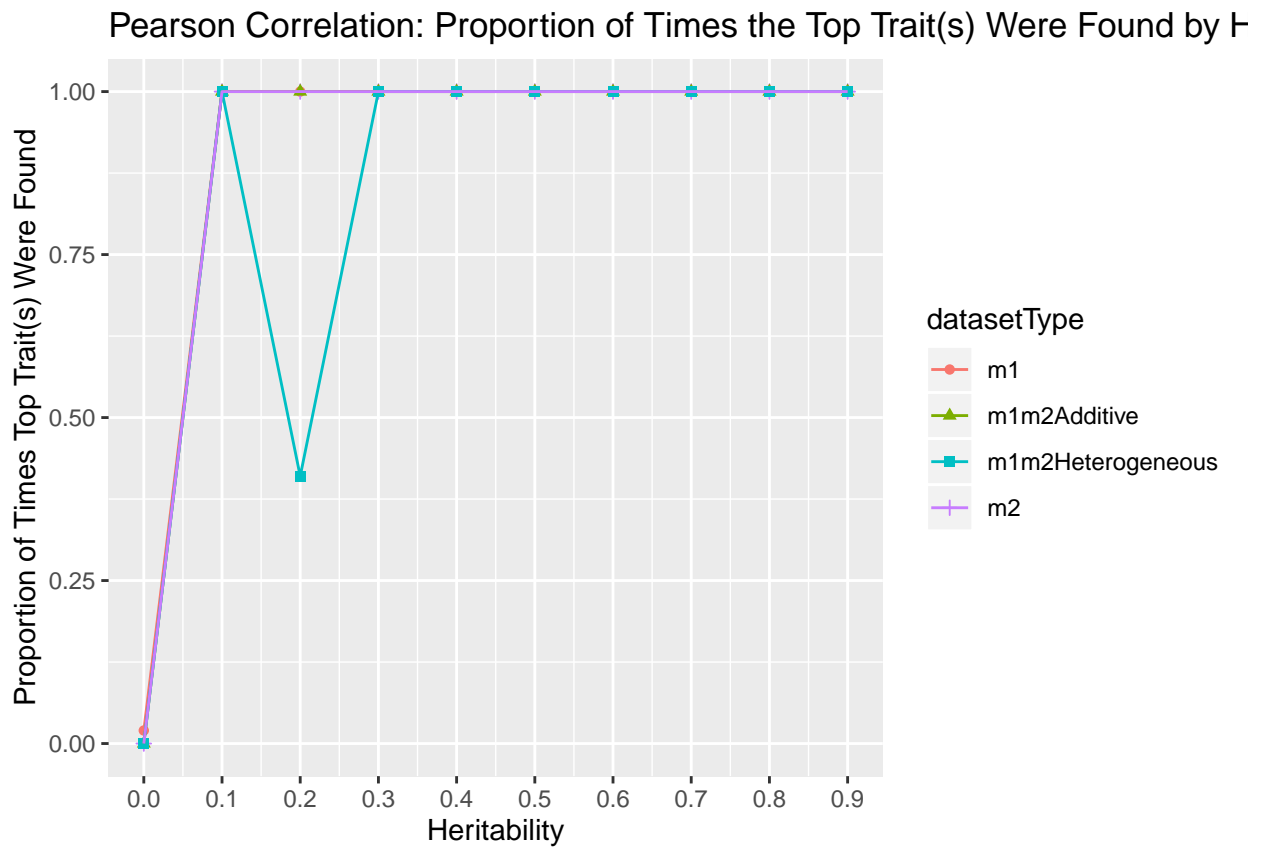
Mann-Whitney Output

Mann–Whitney: Proportion of Times the Top Trait(s) Were Found by Herita
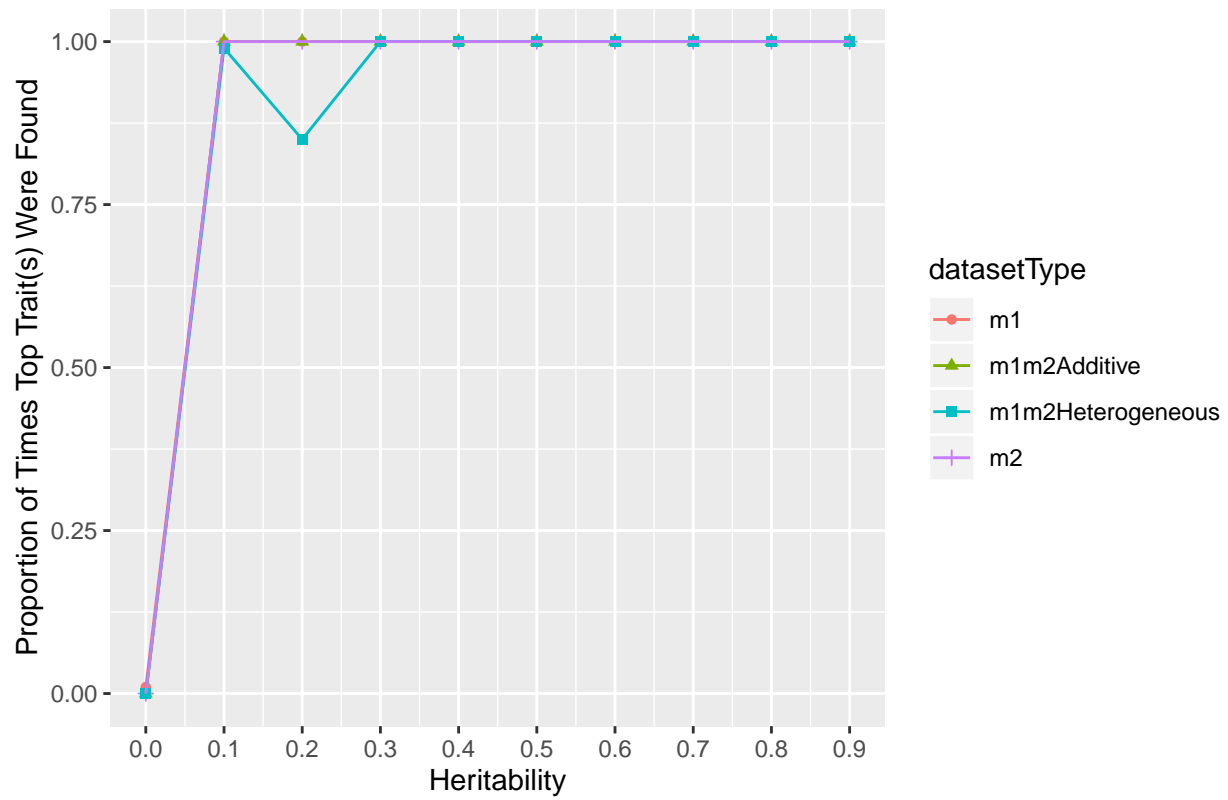
```
## Saving 6.5 x 4.5 in image
```

Pearson Correlation Output

Pearson Correlation: Proportion of Times the Top Trait(s) Were Found by H


```
## Saving 6.5 x 4.5 in image
```
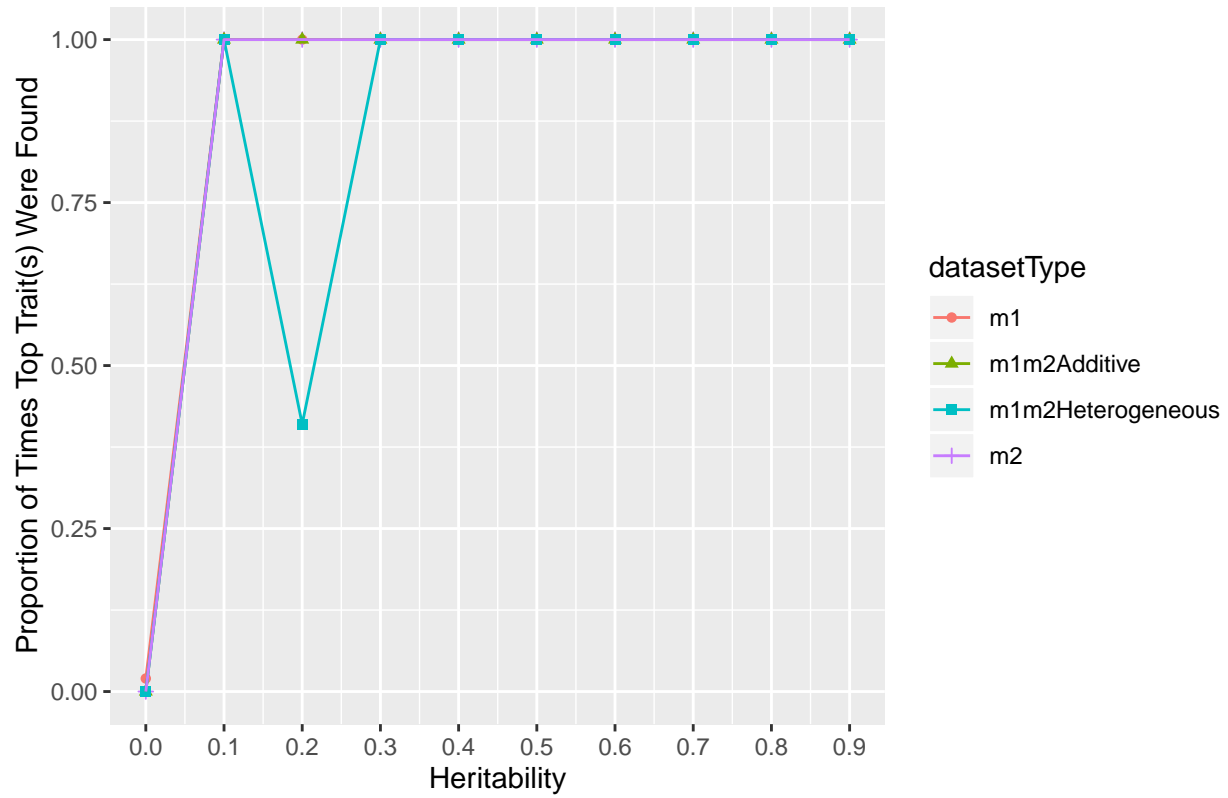
Spearman Correlation Output

Spearman Correlation: Proportion of Times the Top Trait(s) Were Found by
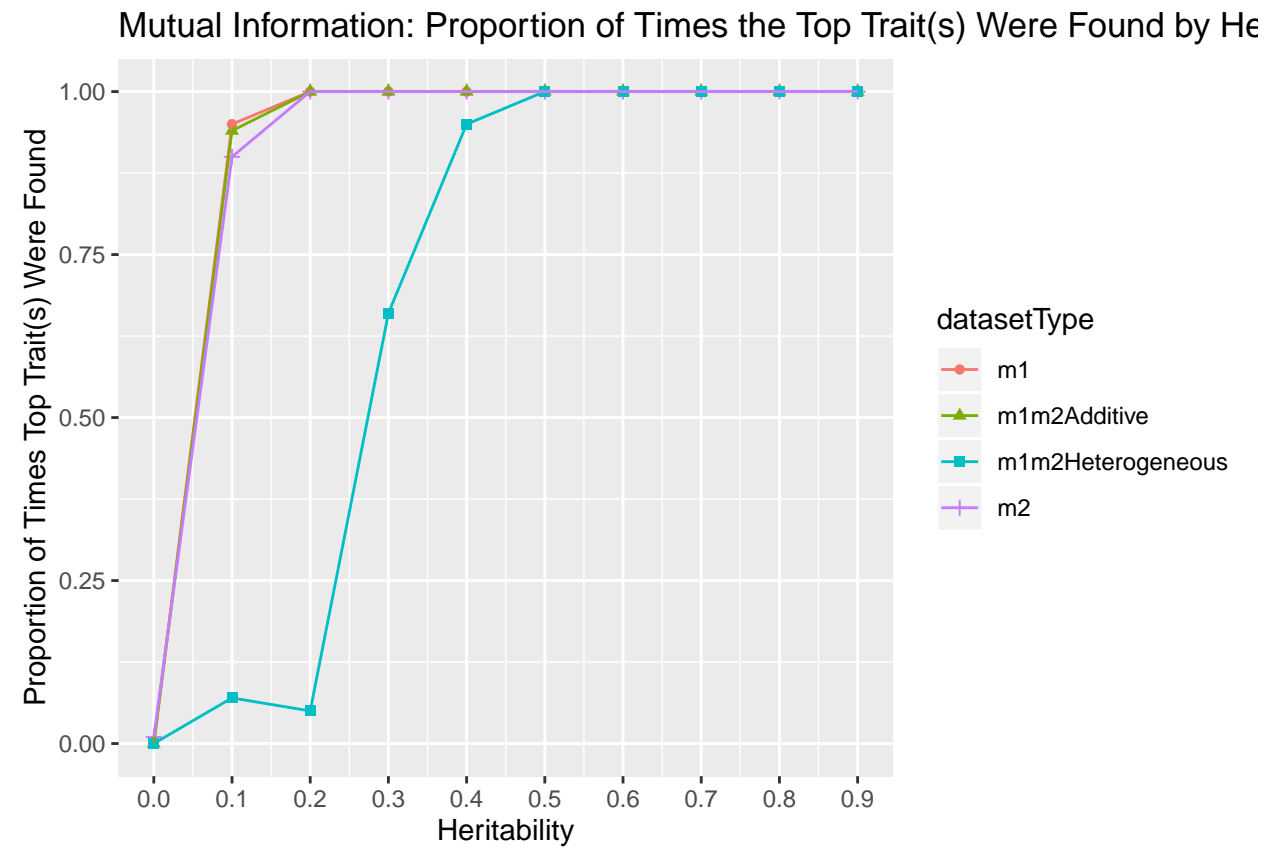


```
## Saving 6.5 x 4.5 in image
```

Logistic Regression Output

Logistic Regression: Proportion of Times the Top Trait(s) Were Found by H
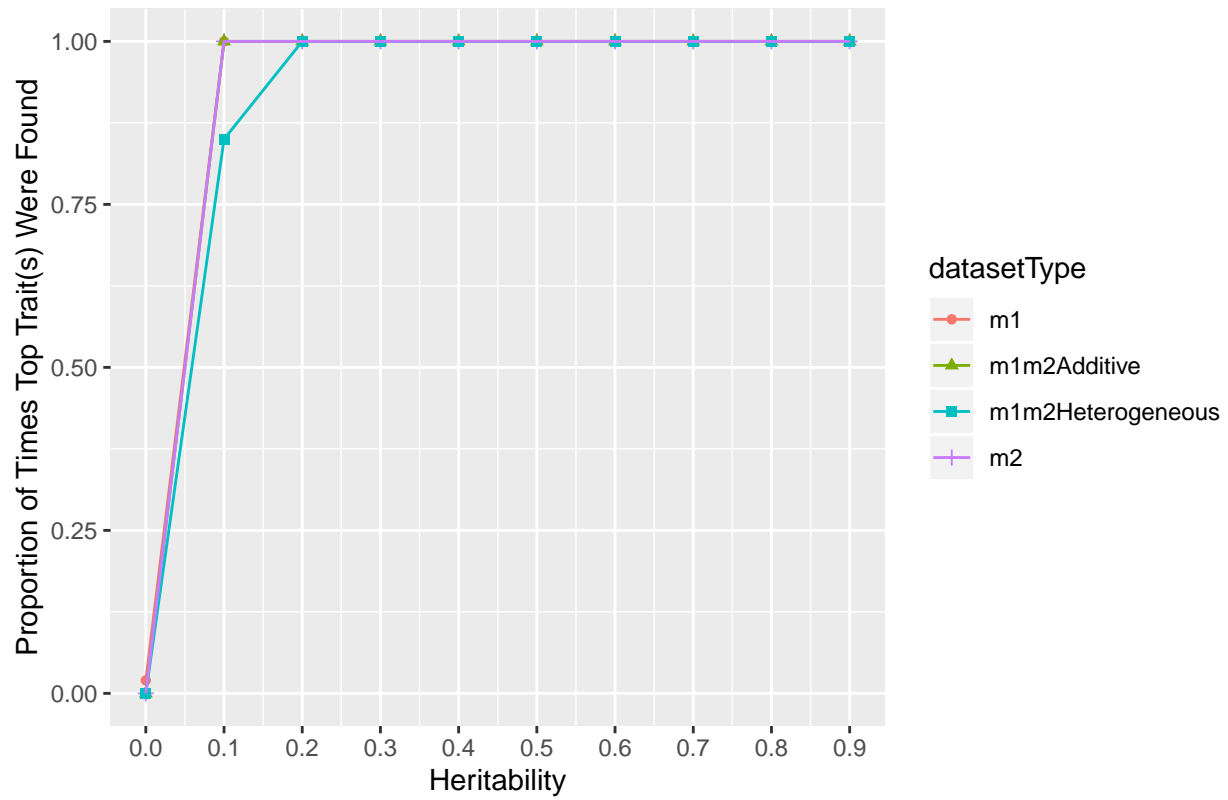


```
## Saving 6.5 x 4.5 in image
```

Mutual Information Output

Mutual Information: Proportion of Times the Top Trait(s) Were Found by He


```
## Saving 6.5 x 4.5 in image
```
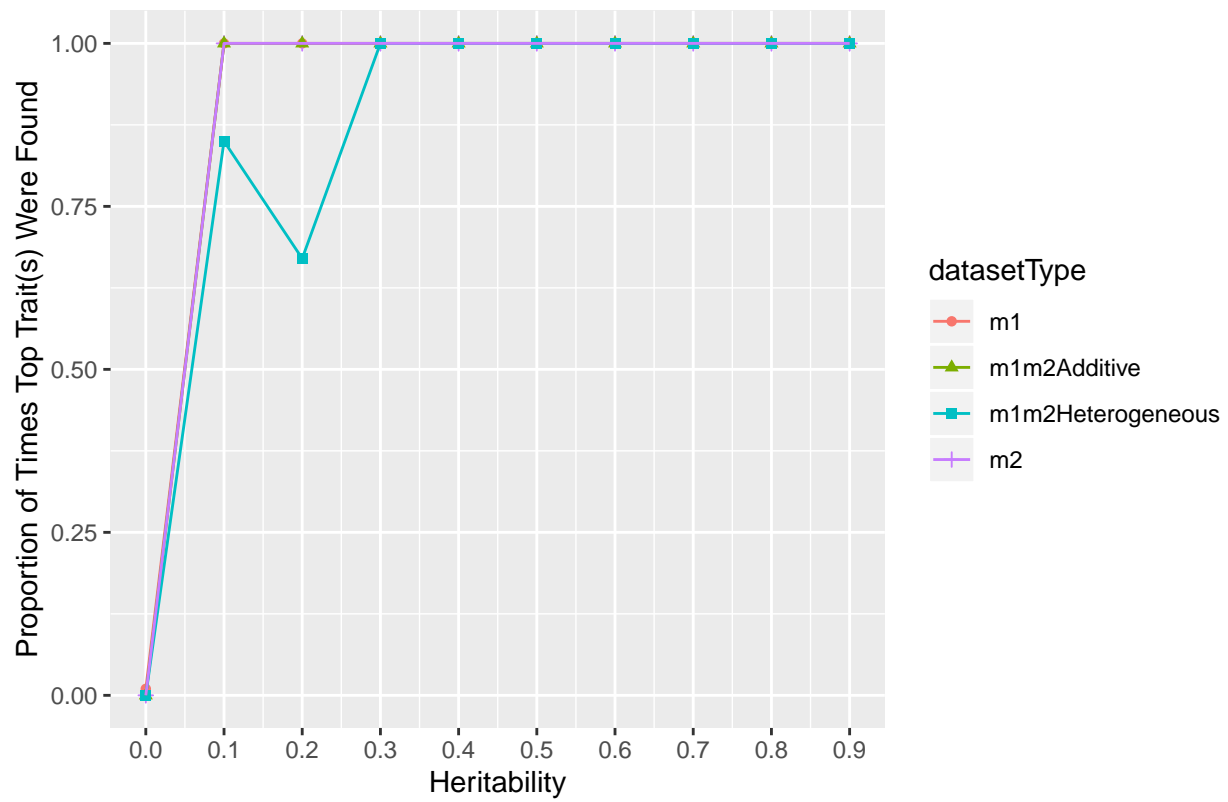
Relief Feature Selection

ReliefF: Proportion of Times the Top Trait(s) Were Found by Heritability

```
## Saving 6.5 x 4.5 in image
```
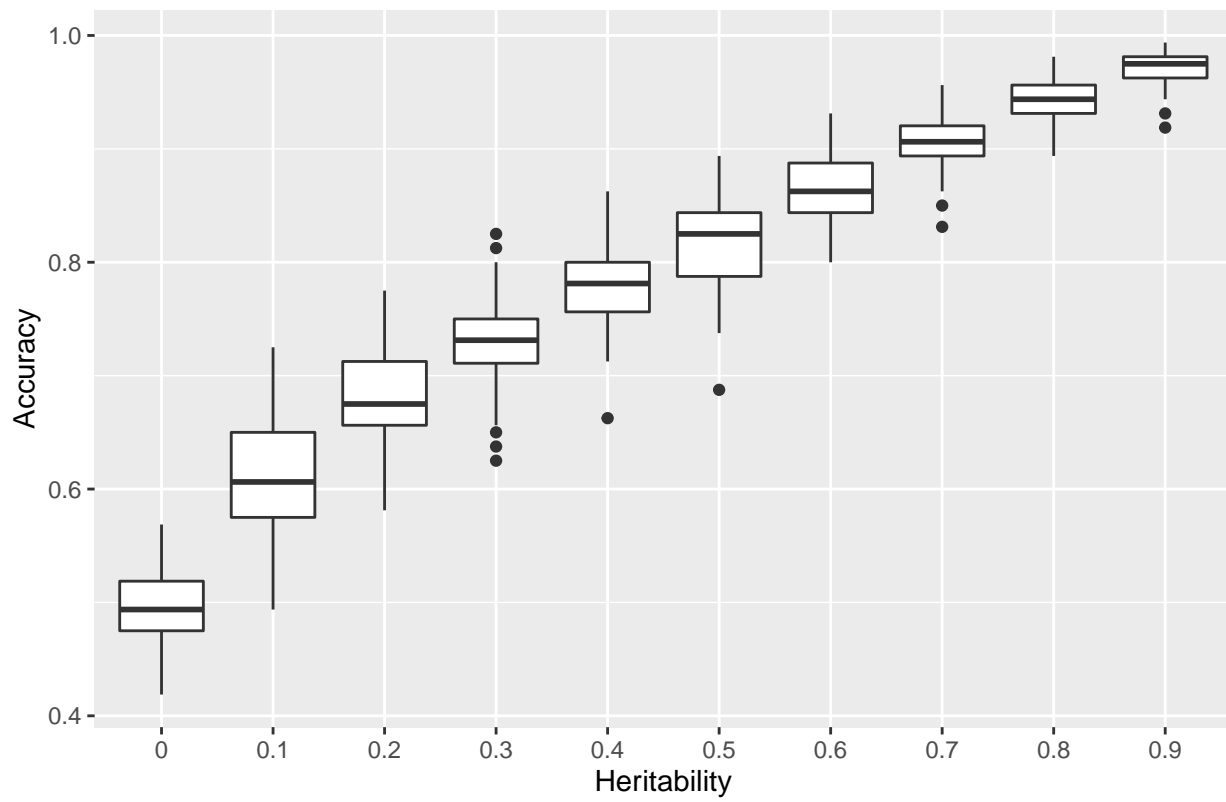
Decision Trees ML

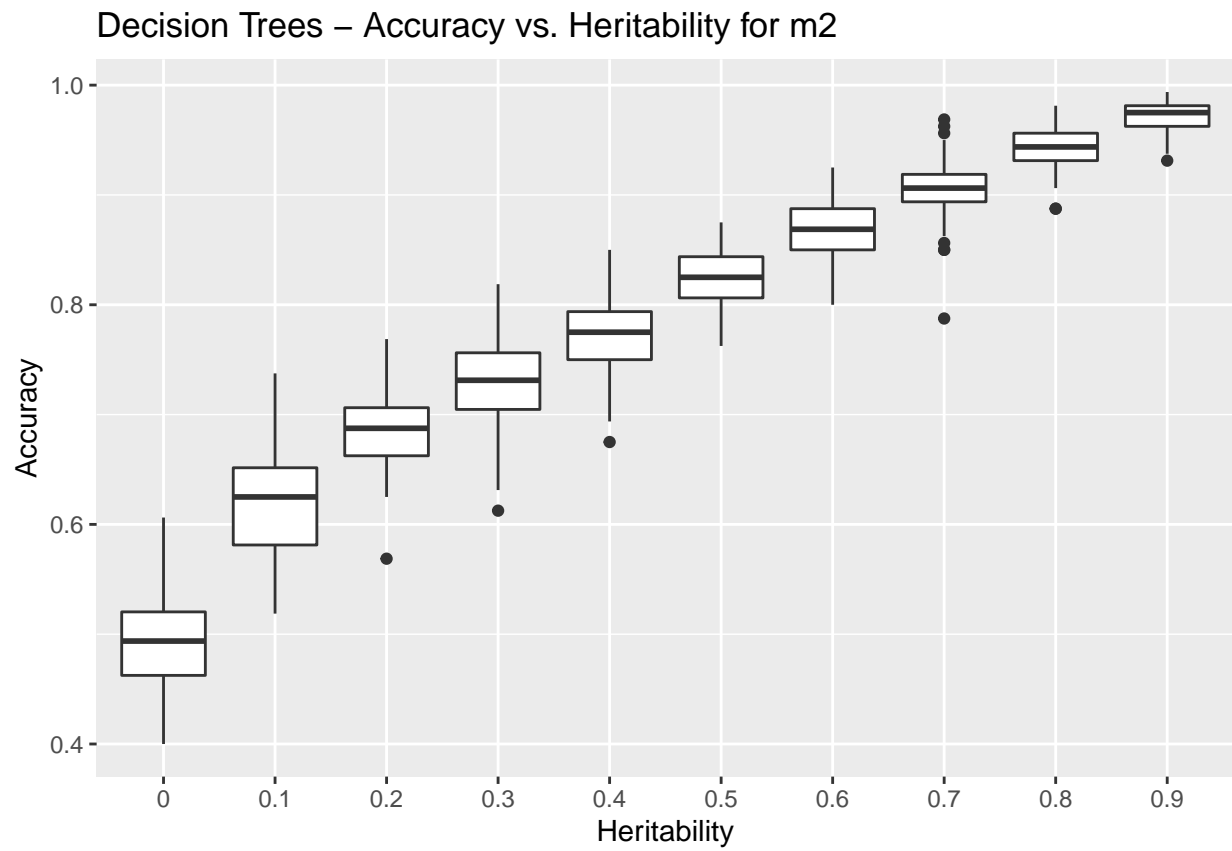## Decision Trees: Proportion of Times the Top Trait(s) Were Found by Herital


```
## Saving 6.5 x 4.5 in image
```
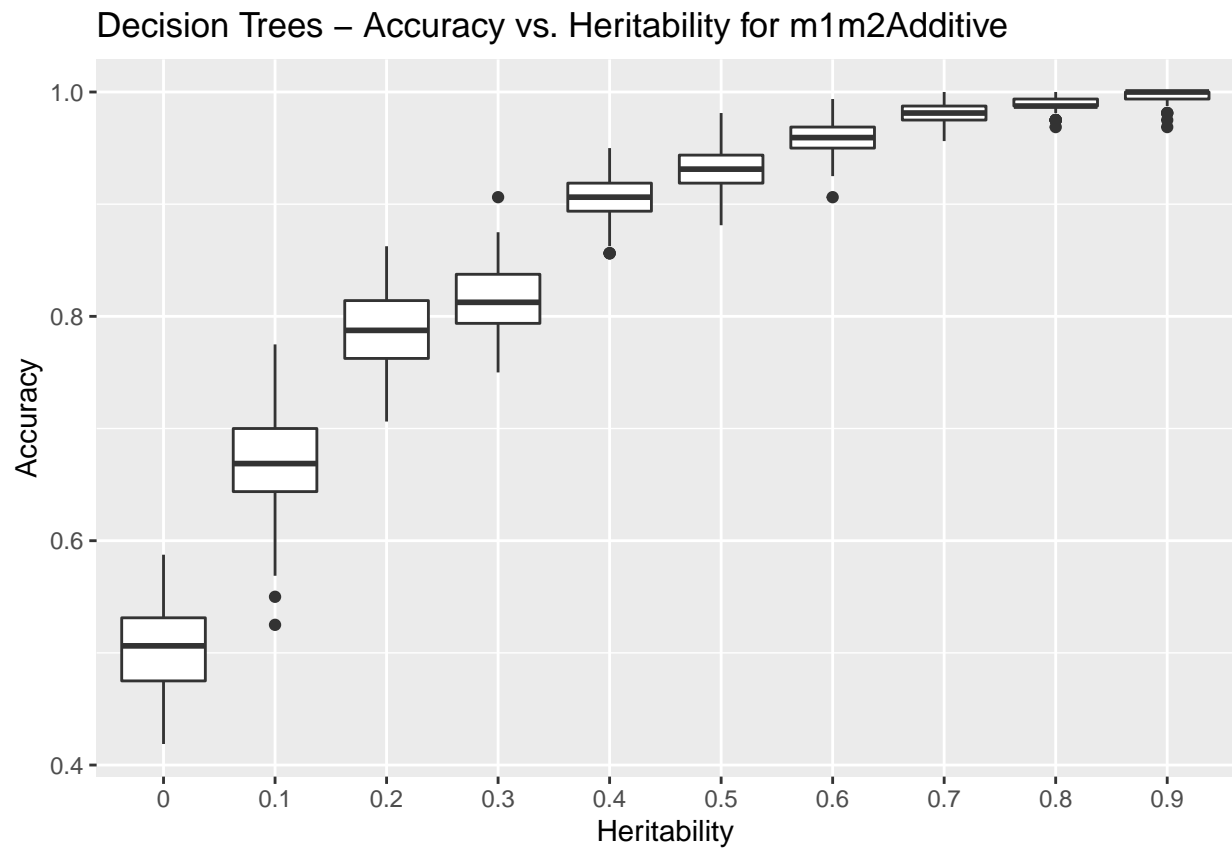
Accuracy Plots

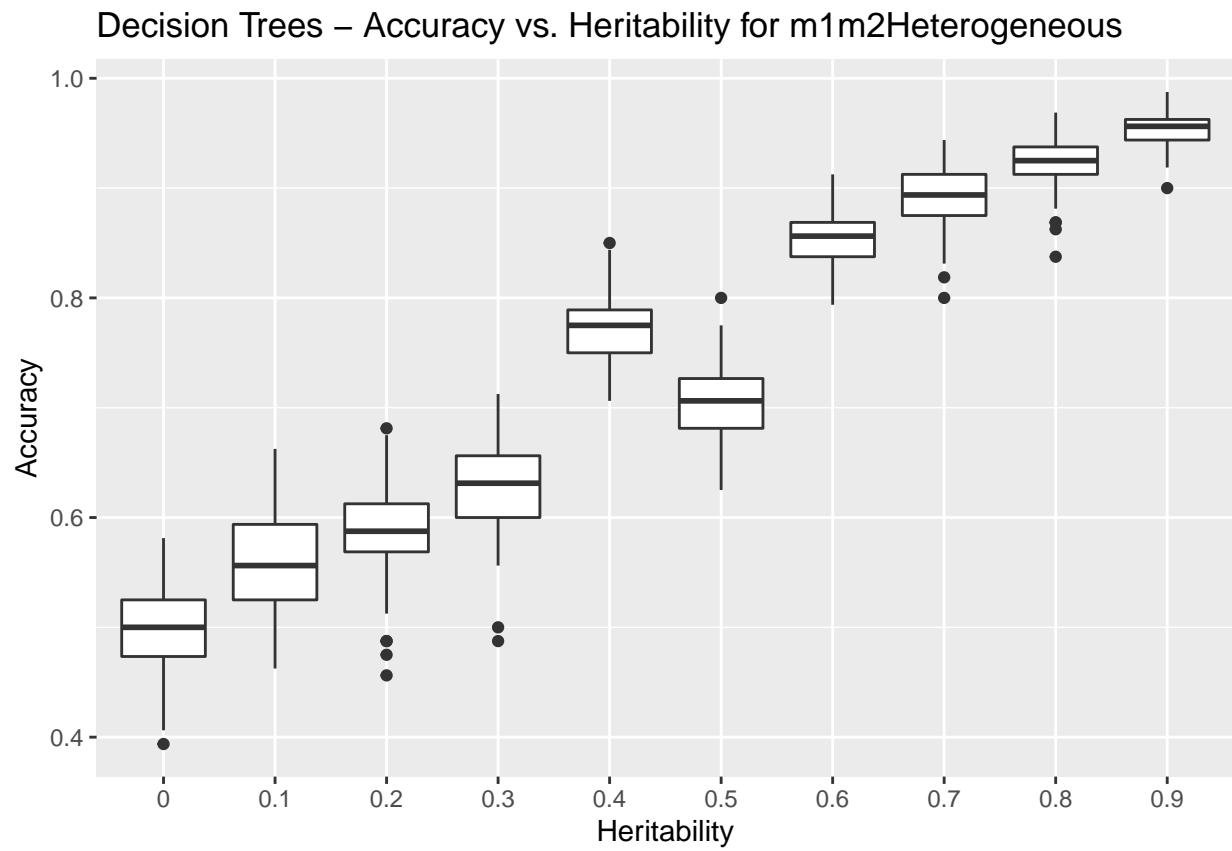## Decision Trees – Accuracy vs. Heritability for m1


```
## Saving 6.5 x 4.5 in image
```

Decision Trees – Accuracy vs. Heritability for m2

```
## Saving 6.5 x 4.5 in image
```

Decision Trees – Accuracy vs. Heritability for m1m2Additive


```
## Saving 6.5 x 4.5 in image
```

Decision Trees – Accuracy vs. Heritability for m1m2Heterogeneous

```
## Saving 6.5 x 4.5 in image
```
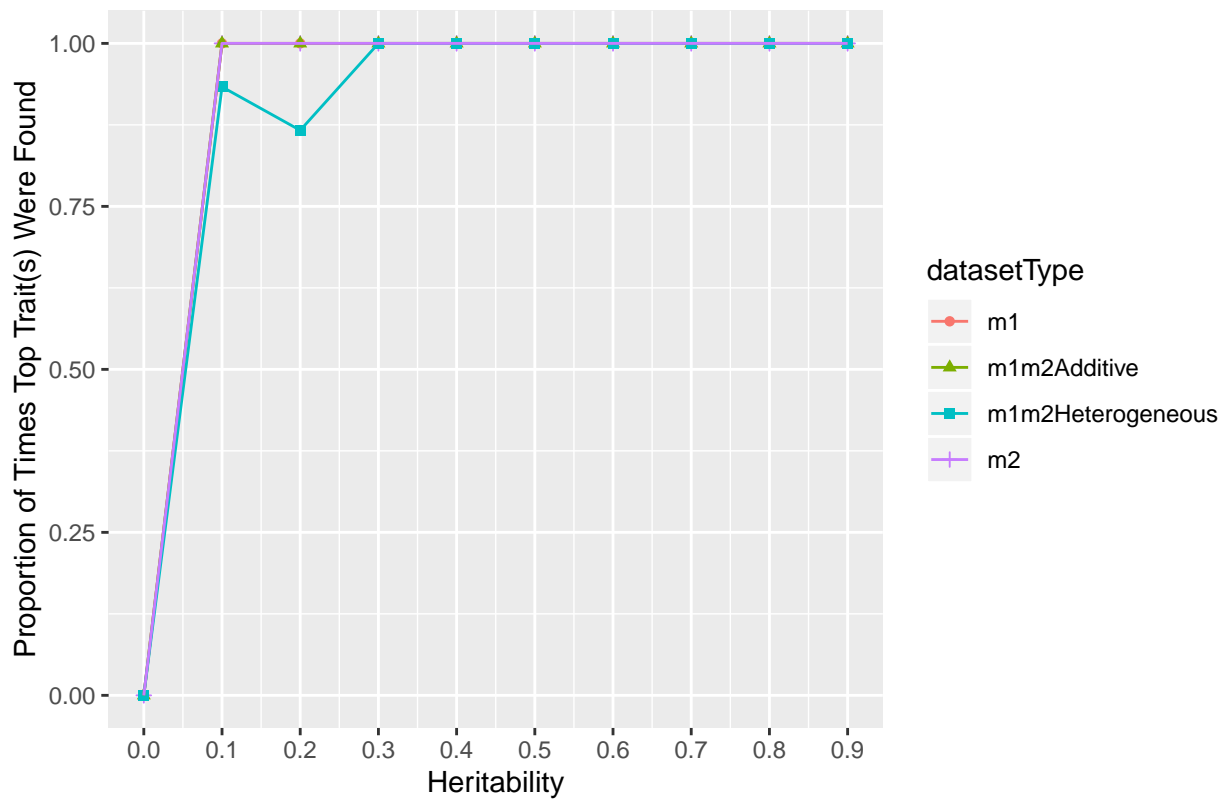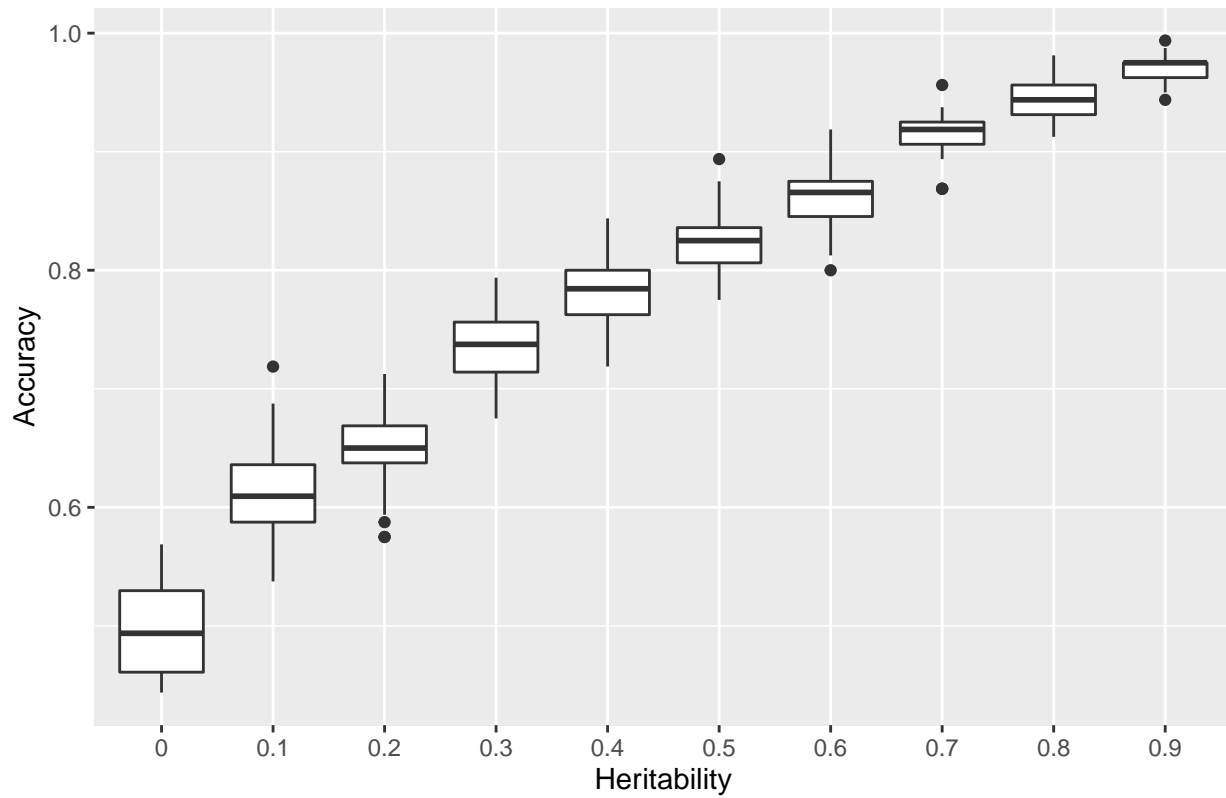
Random Forest ML

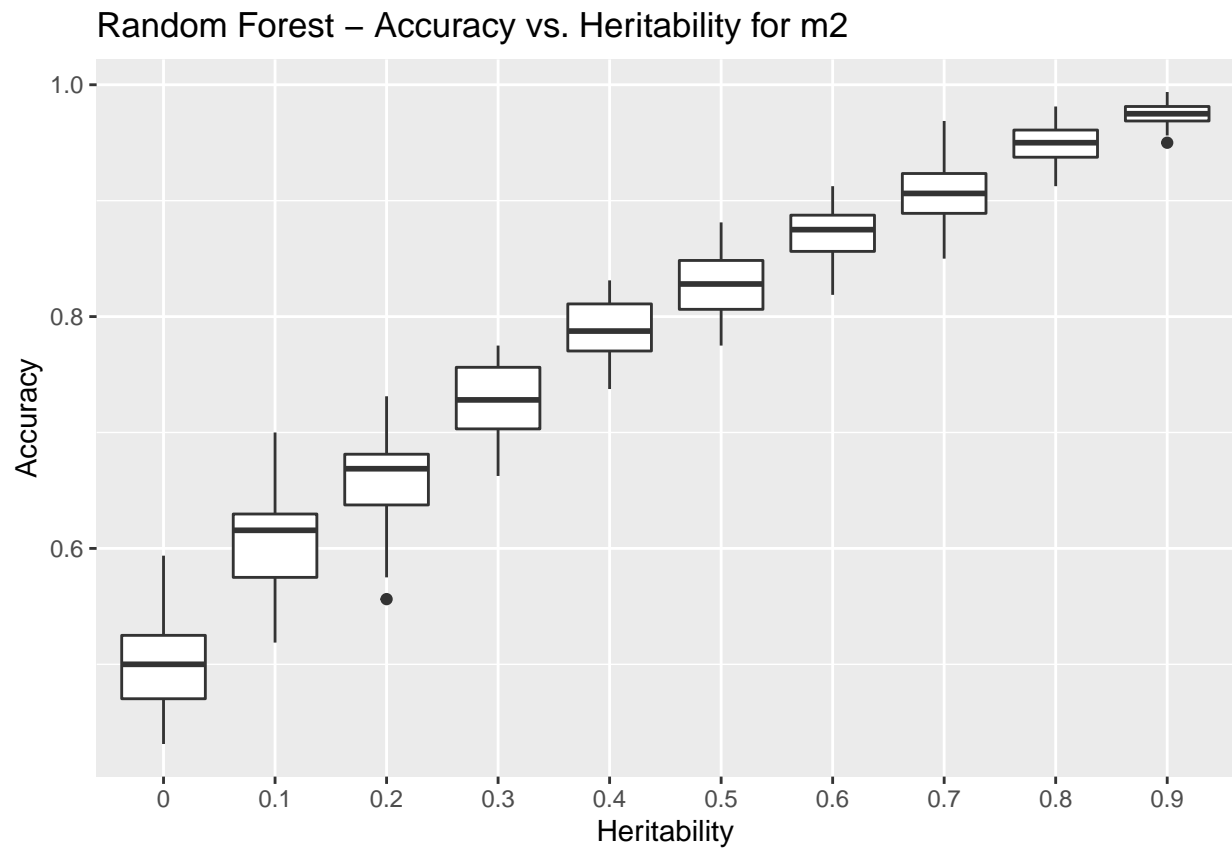Random Forest: Proportion of Times the Top Trait(s) Were Found by Herita
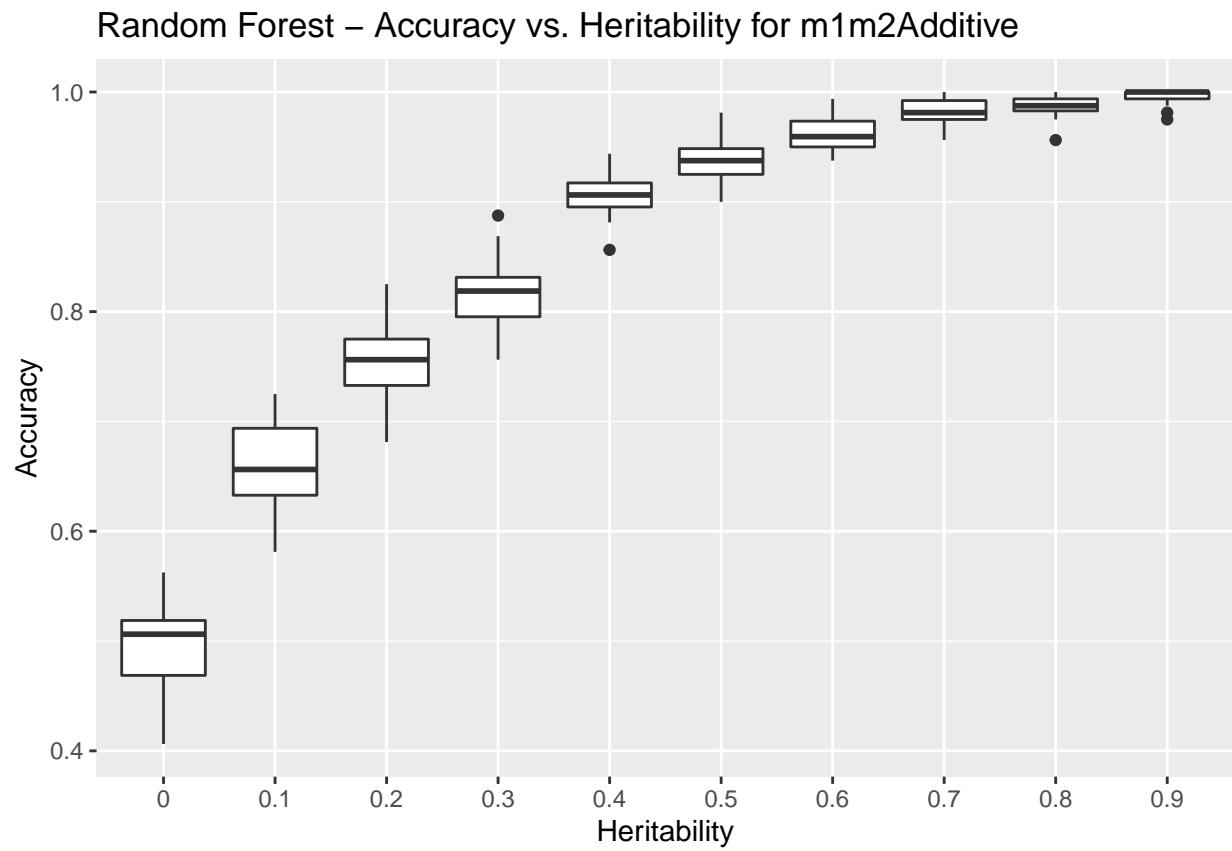
```
## Saving 6.5 x 4.5 in image
```

Accuracy Plots

Random Forest – Accuracy vs. Heritability for m1


```
## Saving 6.5 x 4.5 in image
```

Random Forest – Accuracy vs. Heritability for m2

```
## Saving 6.5 x 4.5 in image
```

Random Forest – Accuracy vs. Heritability for m1m2Additive

```
## Saving 6.5 x 4.5 in image
```

Random Forest – Accuracy vs. Heritability for m1m2Heterogeneous

```
## Saving 6.5 x 4.5 in image
```

XGBoost ML

XGBoost: Proportion of Times the Top Trait(s) Were Found by Heritability

```
## Saving 6.5 x 4.5 in image
```

Accuracy Plots

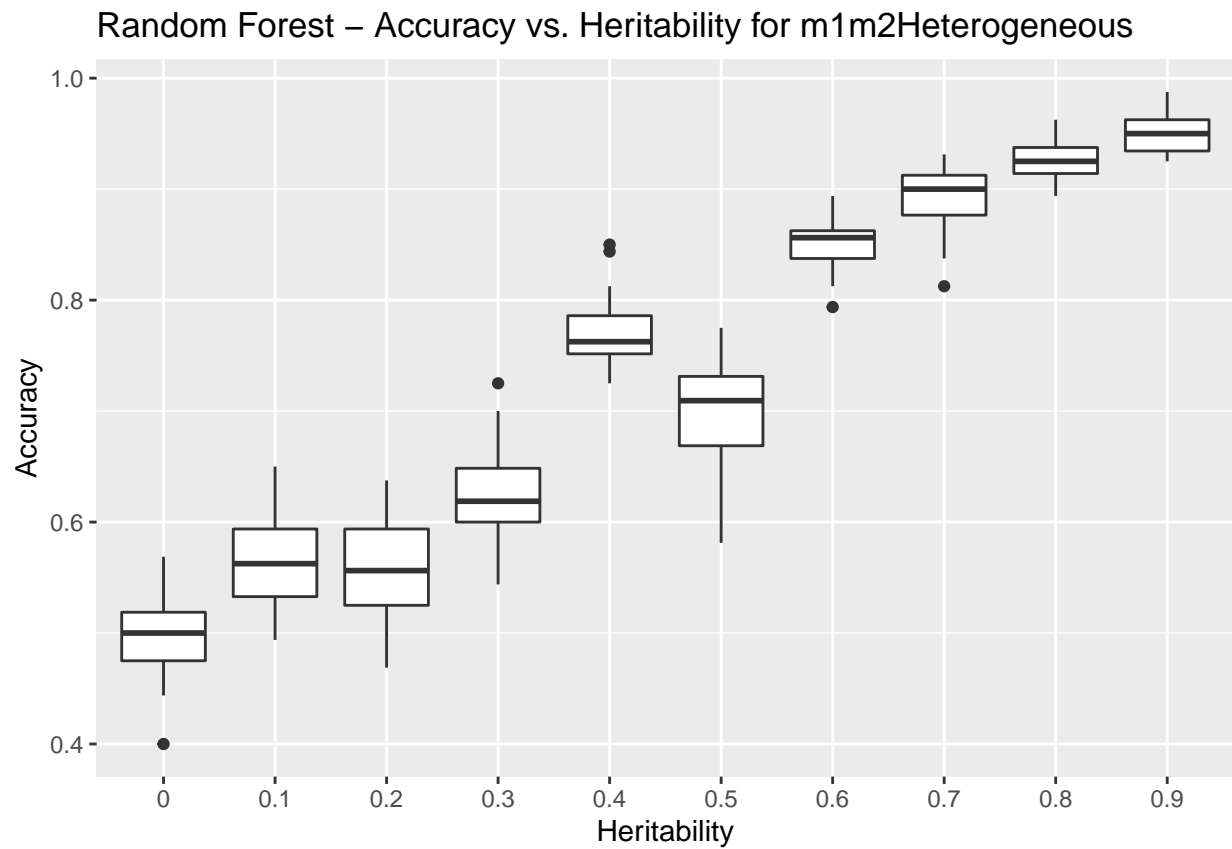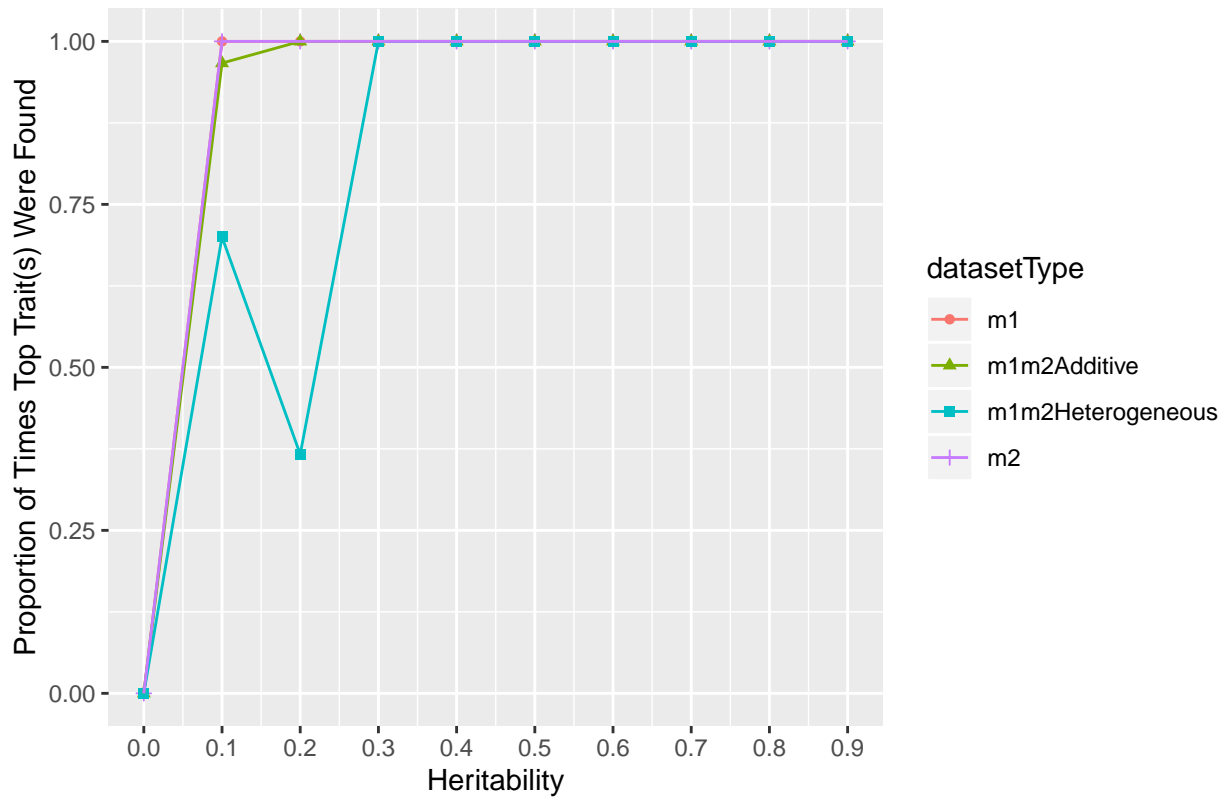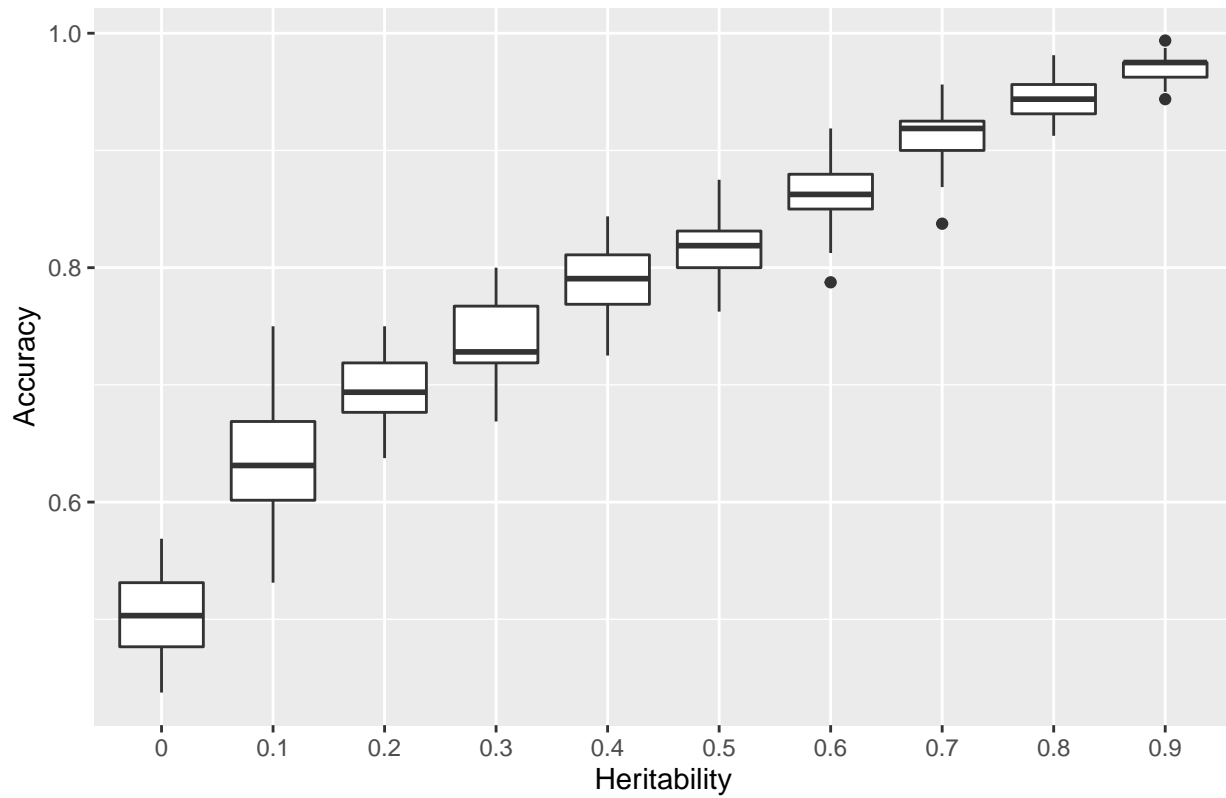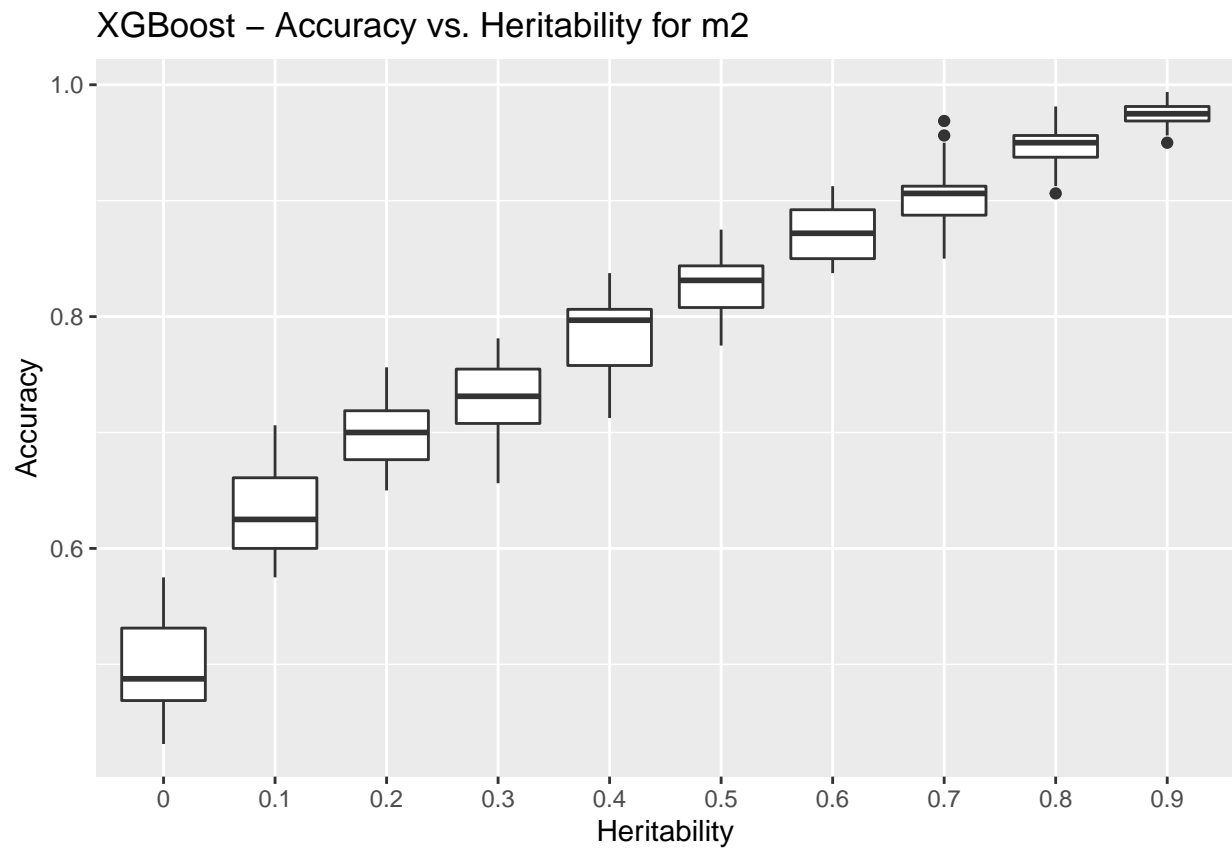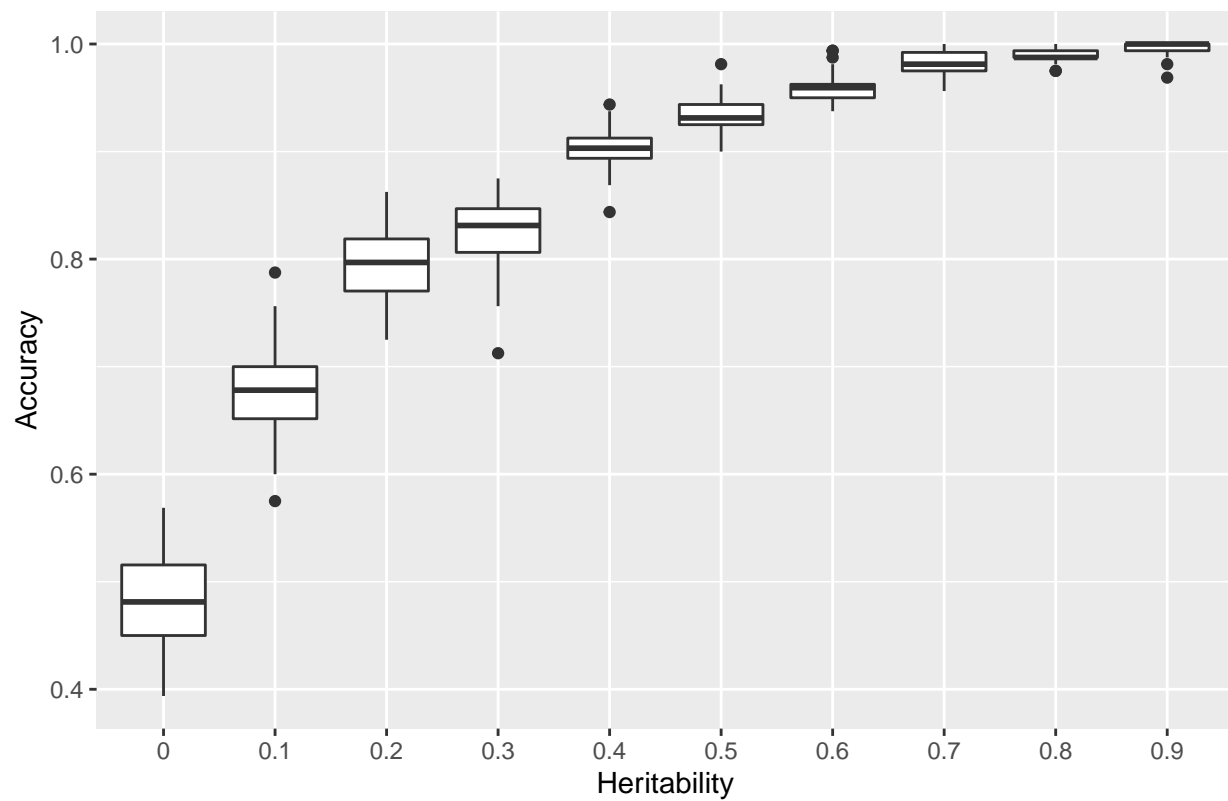## XGBoost – Accuracy vs. Heritability for m1


```
## Saving 6.5 x 4.5 in image
```
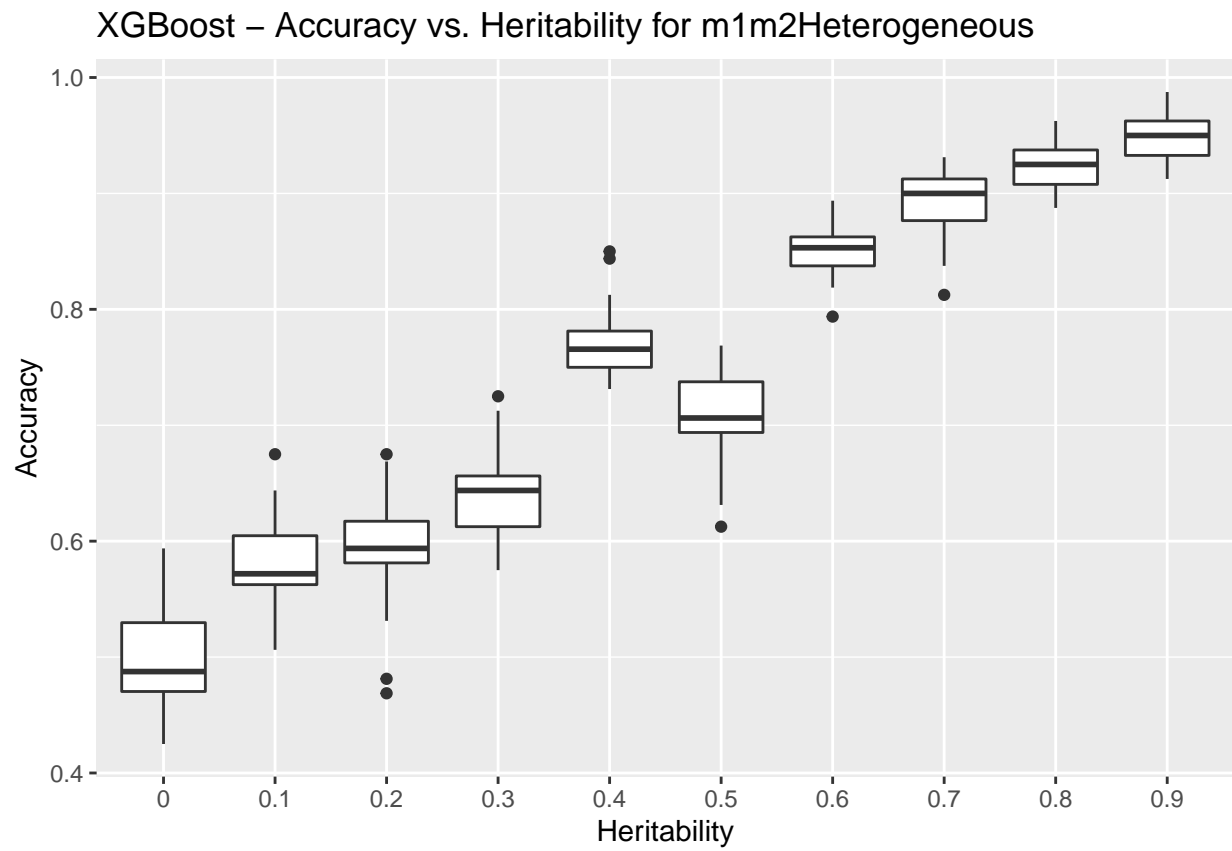
XGBoost – Accuracy vs. Heritability for m2

```
## Saving 6.5 x 4.5 in image
```

XGBoost – Accuracy vs. Heritability for m1m2Additive

```
## Saving 6.5 x 4.5 in image
```

XGBoost – Accuracy vs. Heritability for m1m2Heterogeneous

```
## Saving 6.5 x 4.5 in image
```