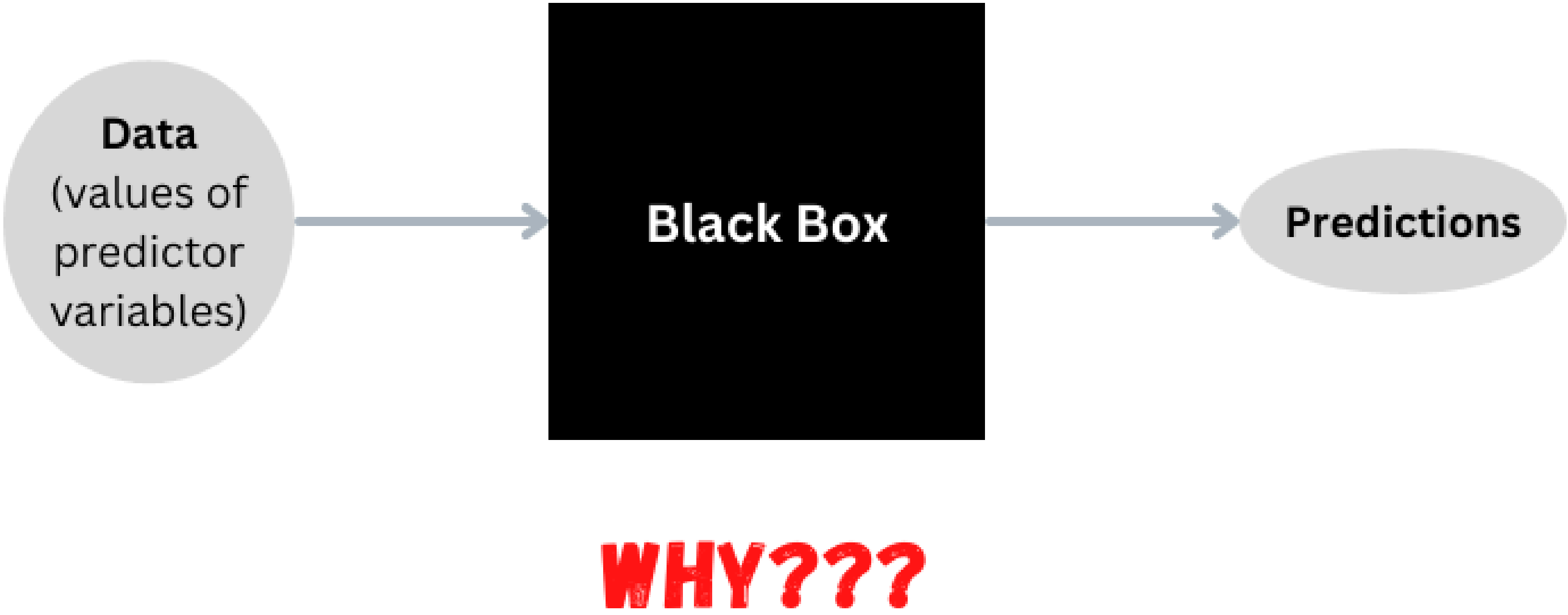
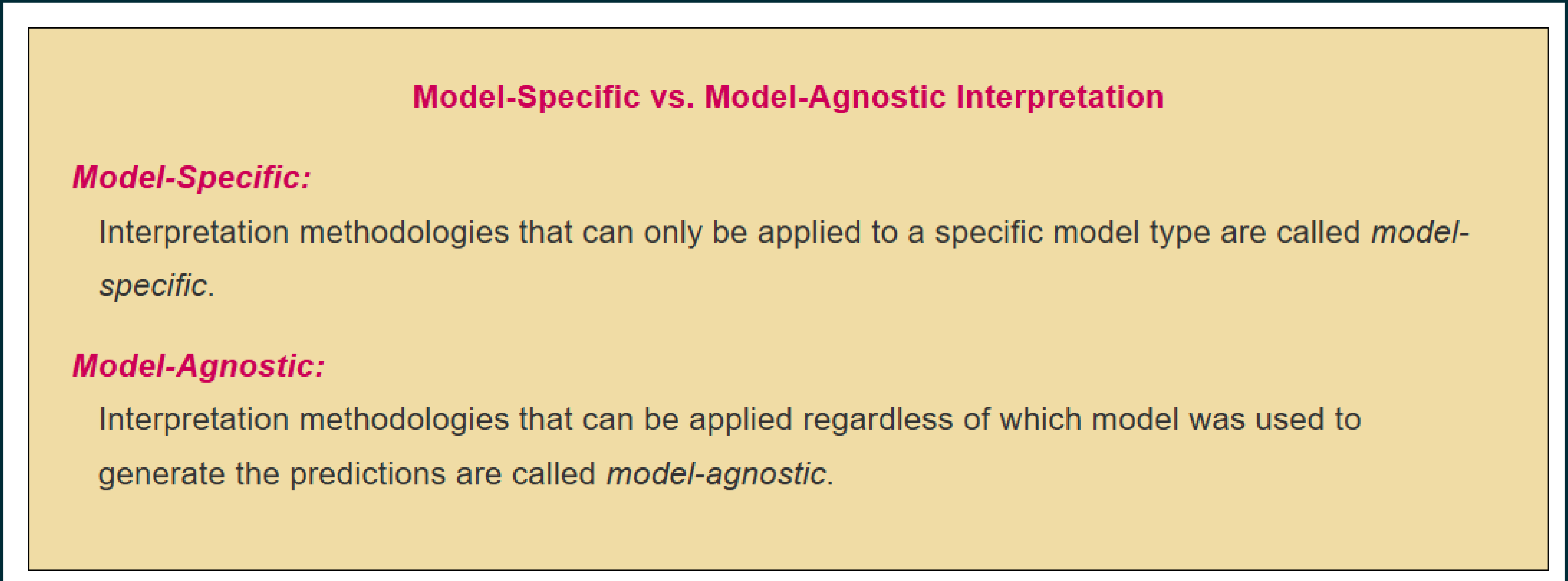
INTERPRETATION OF MACHINE LEARNING RESULTS

https://ai.lange-analytics.com/

FROM BLACKBOX TO INTERPRETATION



TYPES OF INTERPRETATION



Model-Specific vs. Model-Agnostic Interpretation

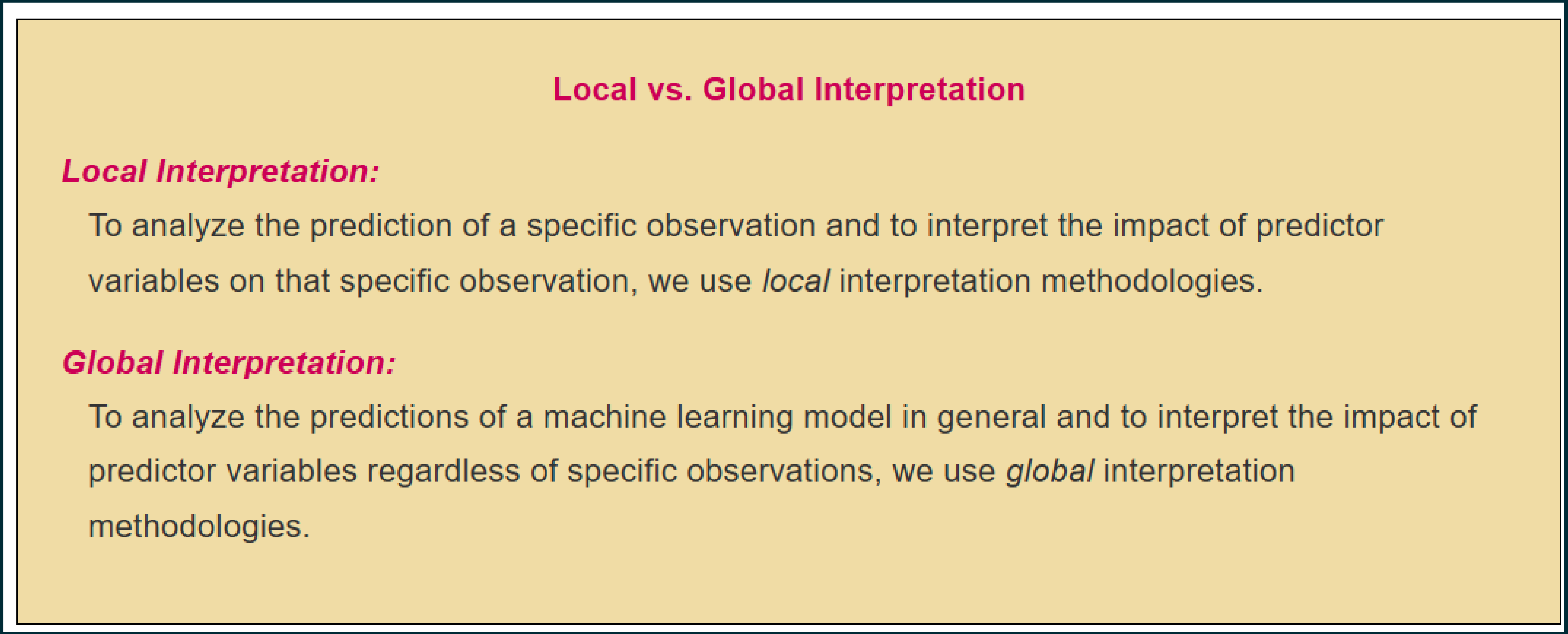
Model-Specific:

Interpretation methodologies that can only be applied to a specific model type are called *model-specific*.

Model-Agnostic:

Interpretation methodologies that can be applied regardless of which model was used to generate the predictions are called *model-agnostic*.

TYPES OF INTERPRETATION



Local vs. Global Interpretation

Local Interpretation:

To analyze the prediction of a specific observation and to interpret the impact of predictor variables on that specific observation, we use *local* interpretation methodologies.

Global Interpretation:

To analyze the predictions of a machine learning model in general and to interpret the impact of predictor variables regardless of specific observations, we use *global* interpretation methodologies.

CETERIS PARIBUS PLOT (LOCAL AND MODEL AGNOSTIC)

LOADING THE DATA AND LIBRARIES

▶ Code

CETERIS PARIBUS PLOT (LOCAL AND MODEL AGNOSTIC)

GENERATING TRAINING/TESING DATA AND IDENTIFYING A OBSERVATION OF INTEREST

▶ Code

```
  Wage Educ Tenure Female
1 3.10   11      0      1
2 3.24   12      2      1
3 3.00   11      0      0
4 3.25    8      1      1
5 3.00   14      3      1
6 2.93   12      2      1
```

Variable of Interest (Helga)

```
# A tibble: 1 × 4
   Wage  Educ Tenure Female
  <dbl> <dbl>  <dbl> <fct>
1   8.9    17     18 1
```

CETERIS PARIBUS PLOT (LOCAL AND MODEL AGNOSTIC)

ESTIMATING WAGE WITH A RANDOM FOREST MODEL

▶ Code

CETERIS PARIBUS PLOT (LOCAL AND MODEL AGNOSTIC)

IDEA

```
# A tibble: 1 × 4
   Wage  Educ Tenure Female
  <dbl> <dbl>  <dbl> <fct>
1   8.9    17     18 1
```

What is the impact of Tenure for this observation?

Idea: Substitute Tenure with different values by leaving the other values the same and predicting for each value the wage using the trained (Random Forest) model.

CETERIS PARIBUS PLOT (LOCAL AND MODEL AGNOSTIC)

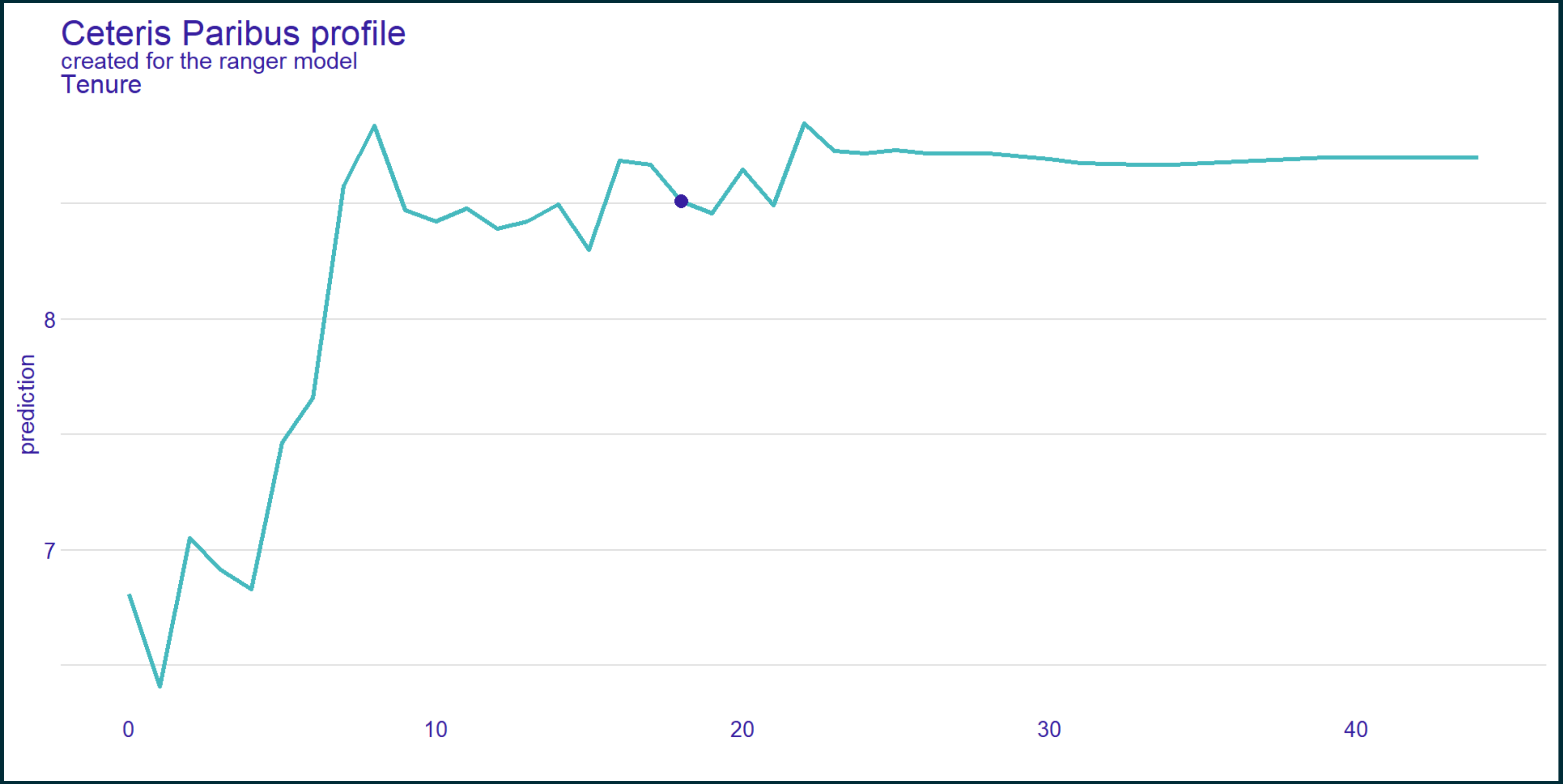
RESULT

▶ Code

CETERIS PARIBUS PLOT (LOCAL AND MODEL AGNOSTIC)

RESULT

▶ Code



PARTIAL DEPENDENCY PLOT (GLOBAL AND MODEL AGNOSTIC)

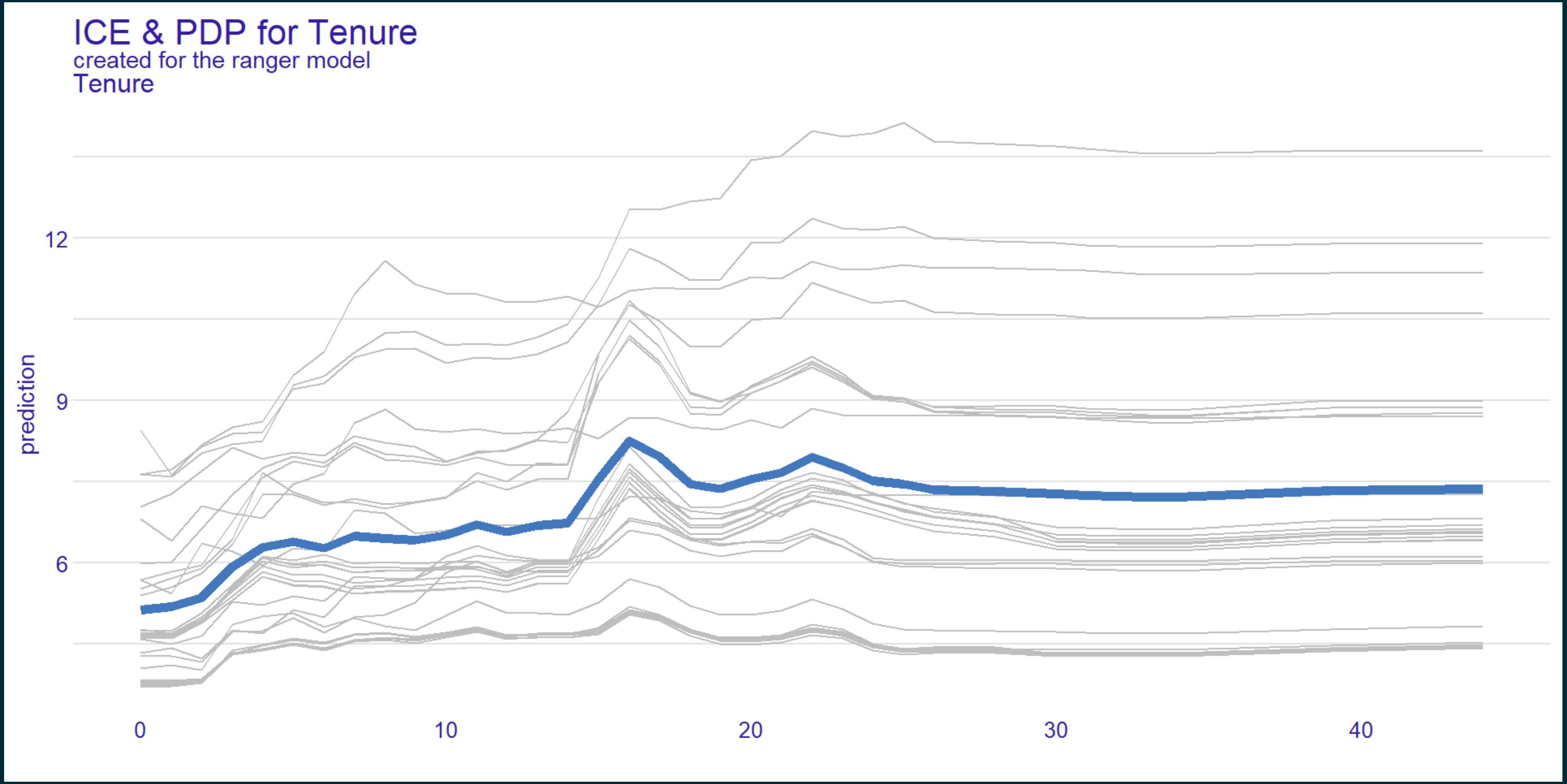
IDEA

The idea behind *Partial Dependence Plots* is first to create a *Ceteris Paribus Plot* for every or at least many observations from the training dataset and then create an average of these individual plots.

PARTIAL DEPENDENCY PLOT (GLOBAL AND MODEL AGNOSTIC)

RESULT

▶ Code



VARIABLE IMPORTANCE (GLOBAL)

LOADING THE DATA AND LIBRARIES

```r
library(tidymodels);library(rio);library(vip);library(DALEXtra); library(kableExtra)
DataVaxFull=import("https://ai.lange-analytics.com/data/DataVax.rds") %>%
            mutate(RowNum=row.names(.))

DataVax= DataVaxFull%>%
                              select(PercVacFull, PercRep,
                                      PercAsian, PercBlack,PercHisp,
                                      PercYoung25, PercOld65,
                                      PercFoodSt, Population) %>%
                       mutate(Population=frequency_weights(Population))
set.seed(2021)
Split85=DataVax %>% initial_split(prop = 0.85,
                                 strata = PercVacFull,
                                 breaks = 3)

DataTrain=training(Split85)
DataTest=testing(Split85)
```

GLOBAL/MODEL SPECIFIC: INTERPRETING COEFFICIENTS AND T-VALUES

LINEAR REGRESIION

▶ Code

term	estimate	std.error	statistic	p.value
(Intercept)	0.9433753	0.0240164	39.2804433	0.0000000
PercRep	-0.5515592	0.0164025	-33.6264704	0.0000000
PercBlack	-0.3771981	0.0207927	-18.1408963	0.0000000
PercYoung25	-0.5999597	0.1060435	-5.6576772	0.0000000
PercHisp	0.0645242	0.0136853	4.7148525	0.0000026
PercFoodSt	-0.1887113	0.0503212	-3.7501339	0.0001813
PercOld65	0.1945055	0.0608728	3.1952800	0.0014165
PercAsian	0.0407740	0.0435763	0.9356931	0.3495327

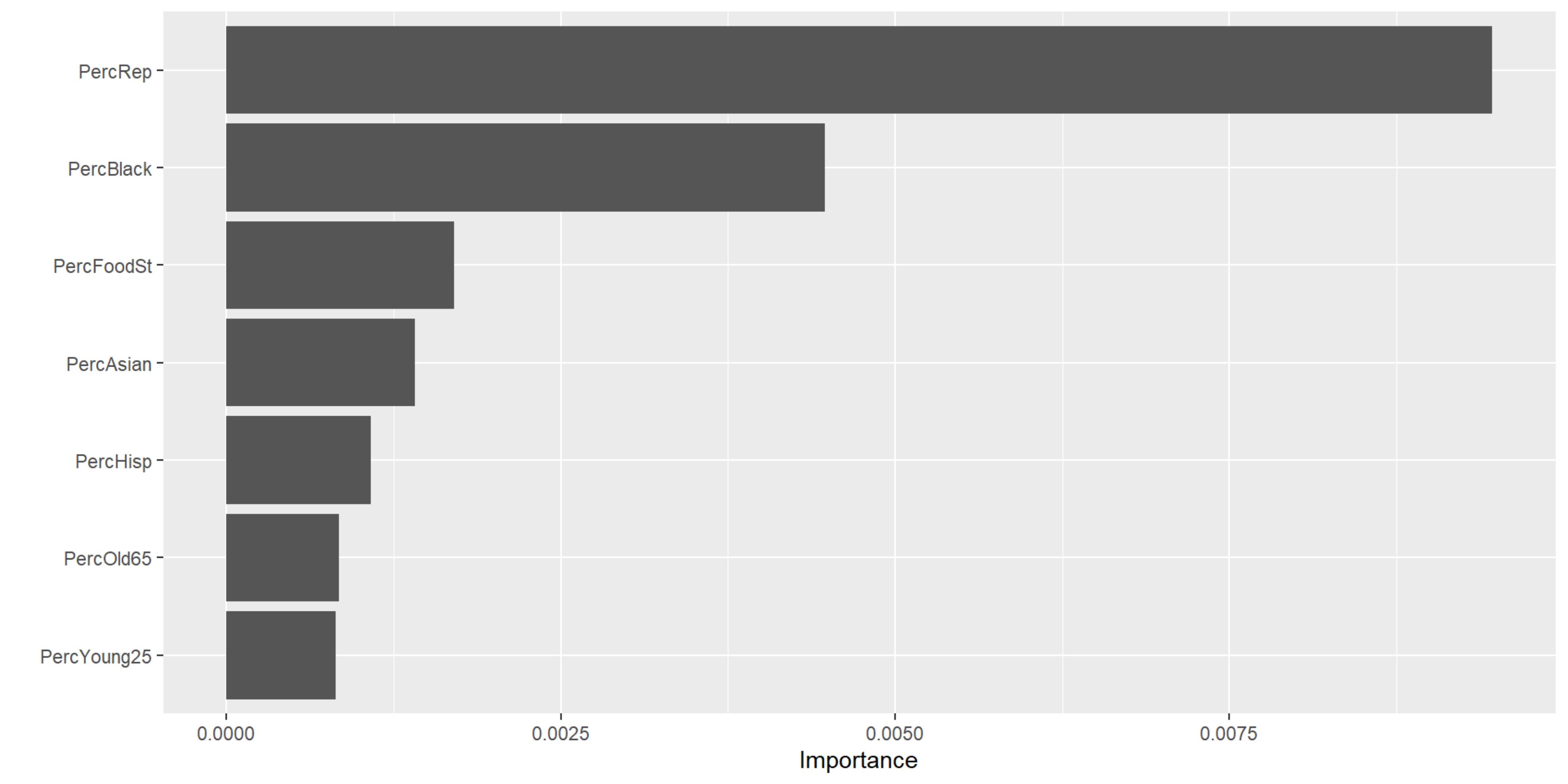
GLOBAL/MODEL SPECIFIC: VARIABLE IMPORTANCE – PERMUTATIONS

RANDOM FOREST

1. Use Out-of-Bag data from each tree.
2. Predict with all variables and record MSE_{all}.
3. Scramble values of first variable and record MSE_1.
4. The (standardized) difference between the MSE_{all} and the MSE_1 of the scrambled version is the variable importance for the first variable.
5. Repeat Steps 3 – 4 for the second, third variables and so on.
6. Plot Variable Importance.

VARIABLE IMPORTANCE PLOT – PERMUTATIONS

- Code



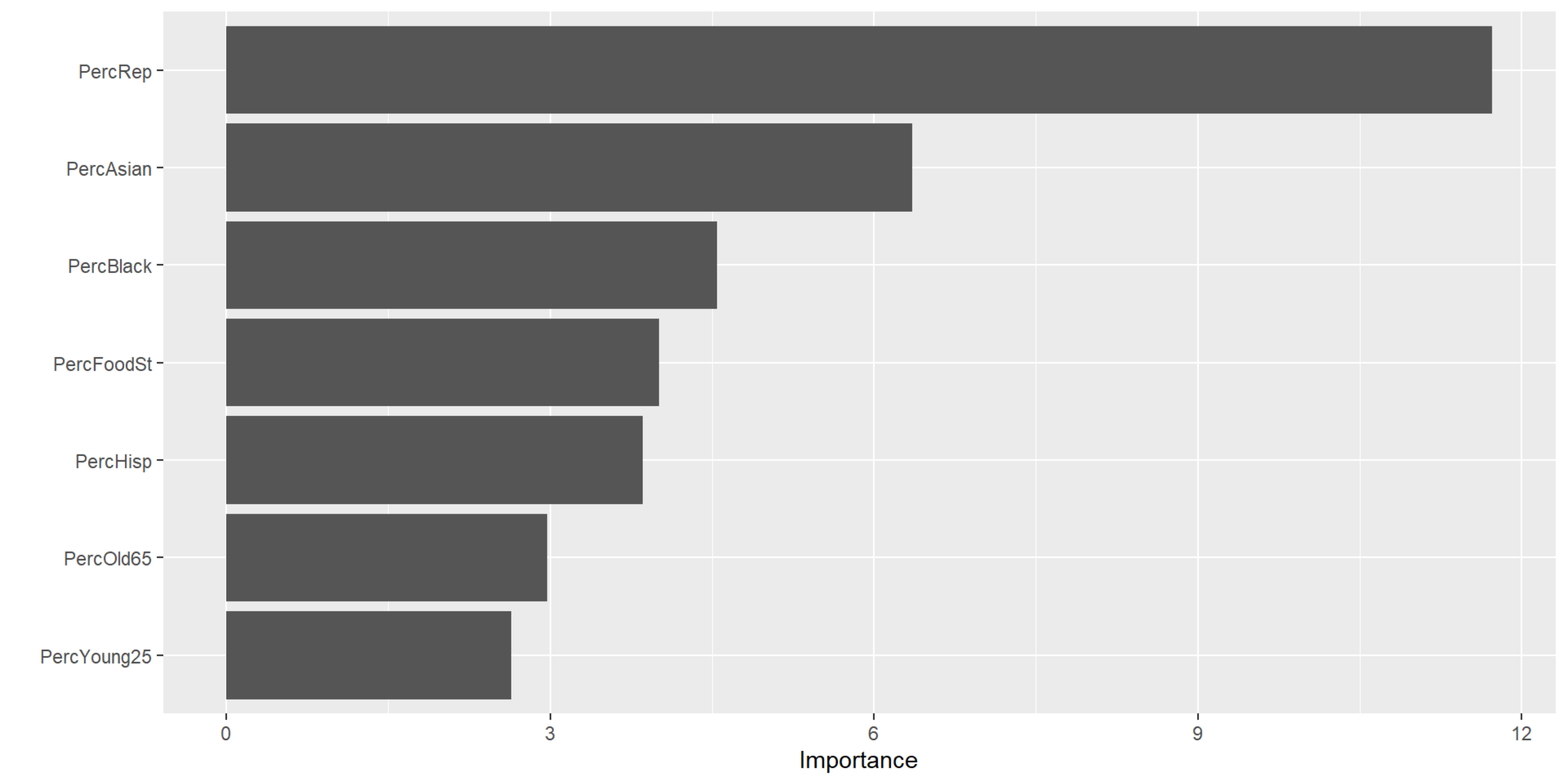
GLOBAL/MODEL SPECIFIC: VARIABLE IMPORTANCE – IMPURITY:

RANDOM FOREST

1. Use Out-of-Bag data from each tree.
2. Calculate the decrease of *Variance Impurity* for regression (*Gini* for classification) for each split in each tree where the first variable is involved.
3. Calculate the (weighted) average of all decreases for the first variable considering alls trees of the *Random Forest*.
4. Repeat steps 2 – 3 for the second, third variables and so on.
5. Plot Variable Importance.

VARIABLE IMPORTANCE PLOT – IMPURITY:

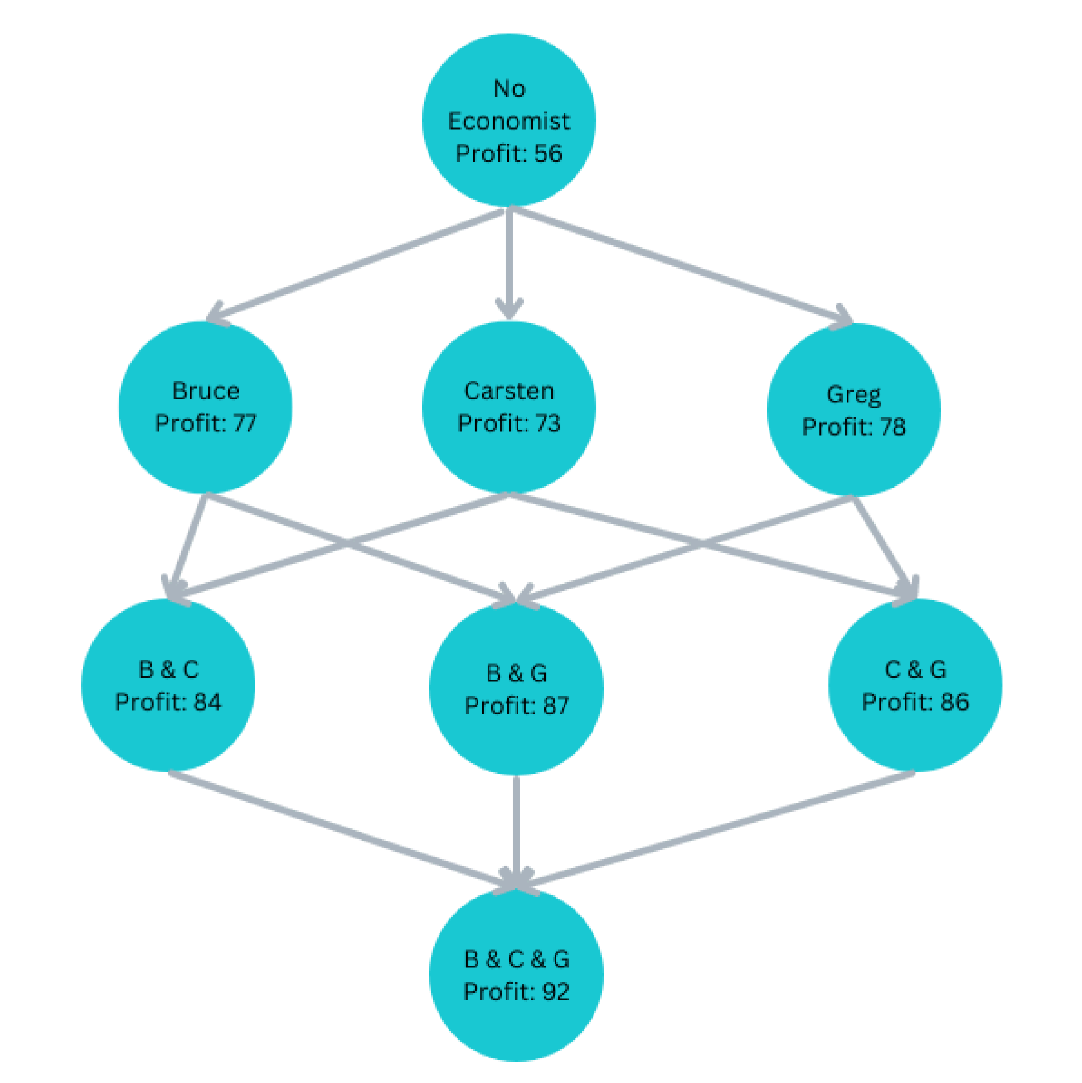
- Code



https://ai.lange-analytics.com/

SHAPLEY VALUES FROM A GAME THEORY APPROACH

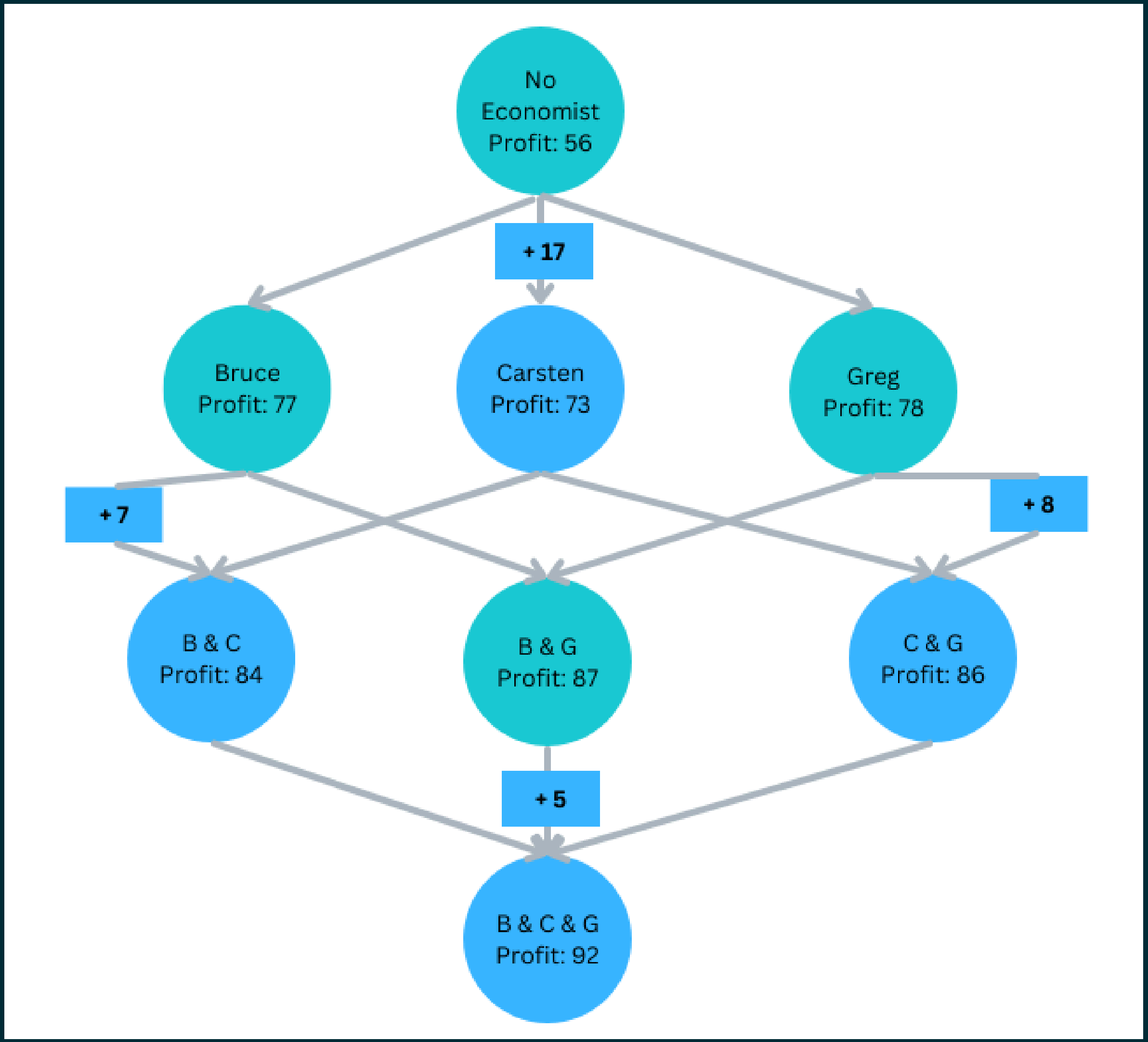
BRUCE, CARSTEN, AND GREG ARE PLAYERS CONTRIBUTING TO THE PROFIT



https://ai.lange-analytics.com/

SHAPLEY VALUES FROM A GAME THEORY APPROACH

ESTIMATING CARSTEN'S CONTRIBUTION

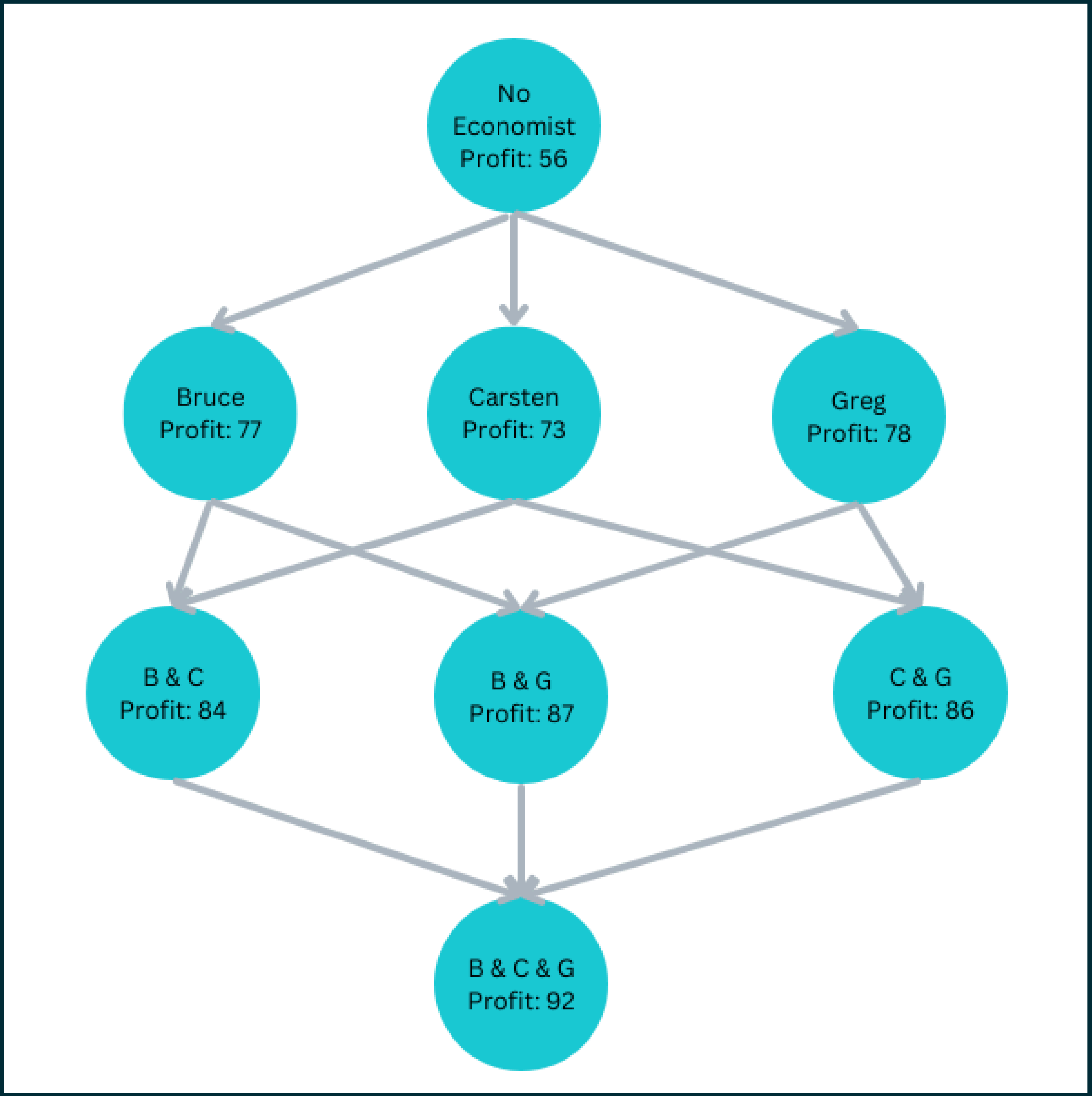


We do not know at which level Carsten joins:

- Carsten joins last: $\Delta P^{rofit}_{BGC} = 5$
- Carsten joins second: $\Delta P^{rofit}_{BC} = 7$ or $\Delta^{rofit}_{GC} = 8$
- Carsten joins first: $\Delta P^{rofit}_{C} = 17$
- Carsten’s average contribution (Shapley value): $S^{hap}_{C} = \frac{1}{3}17 + \frac{1}{6}7 + \frac{1}{6}8 + \frac{1}{3}5 = 9.83$

SHAPLEY VALUES FROM A GAME THEORY APPROACH

ESTIMATING BRUCES’S CONTRIBUTION



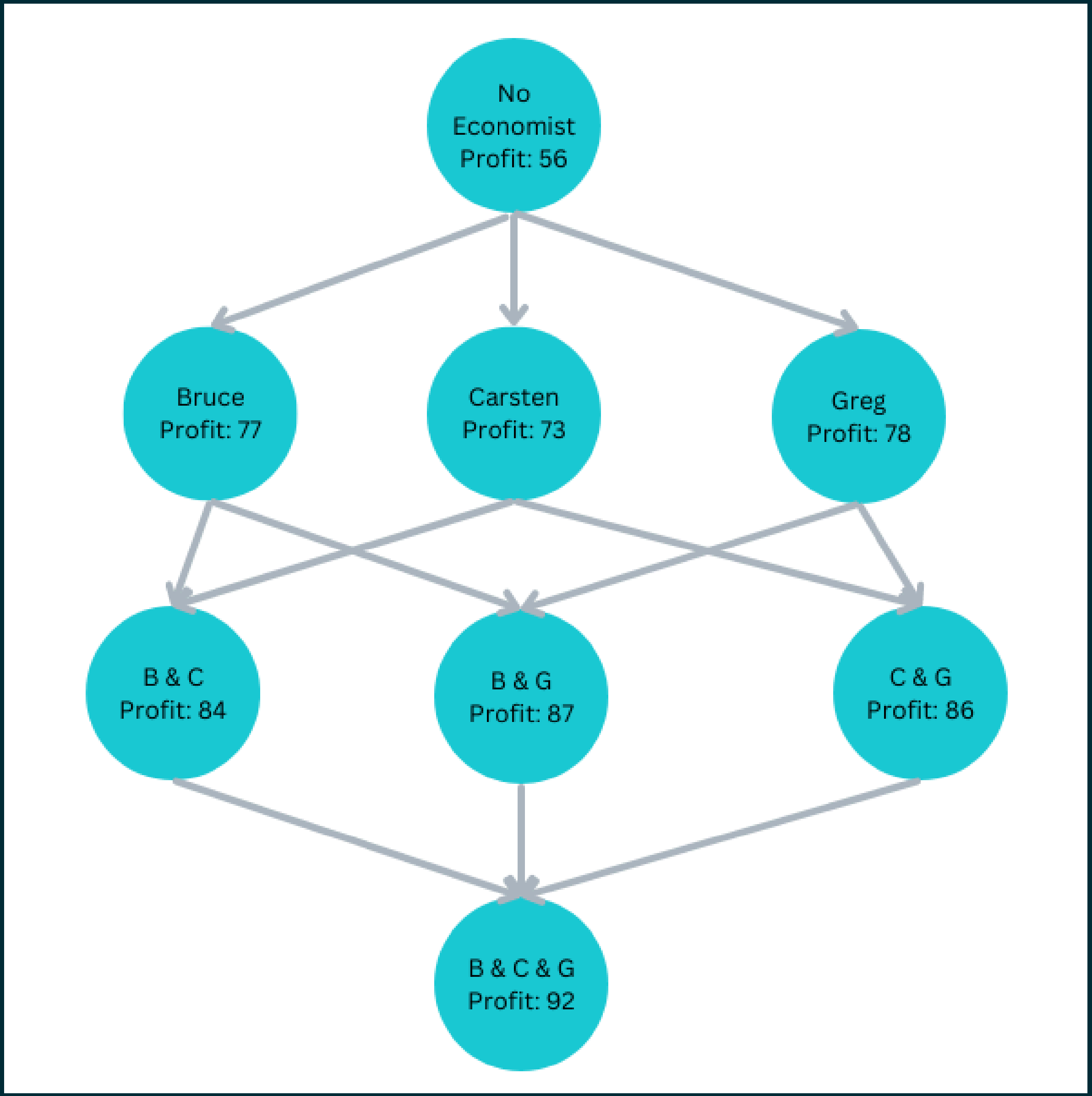
Calculate Bruce' contribution as an exercise:

- Spoiler Alert (below is the result):
- Bruces's average contribution (Shapley value): $S_B^{hap} = 12.33$

https://ai.lange-analytics.com/

SHAPLEY VALUES FROM A GAME THEORY APPROACH

ESTIMATING GREG’S CONTRIBUTION

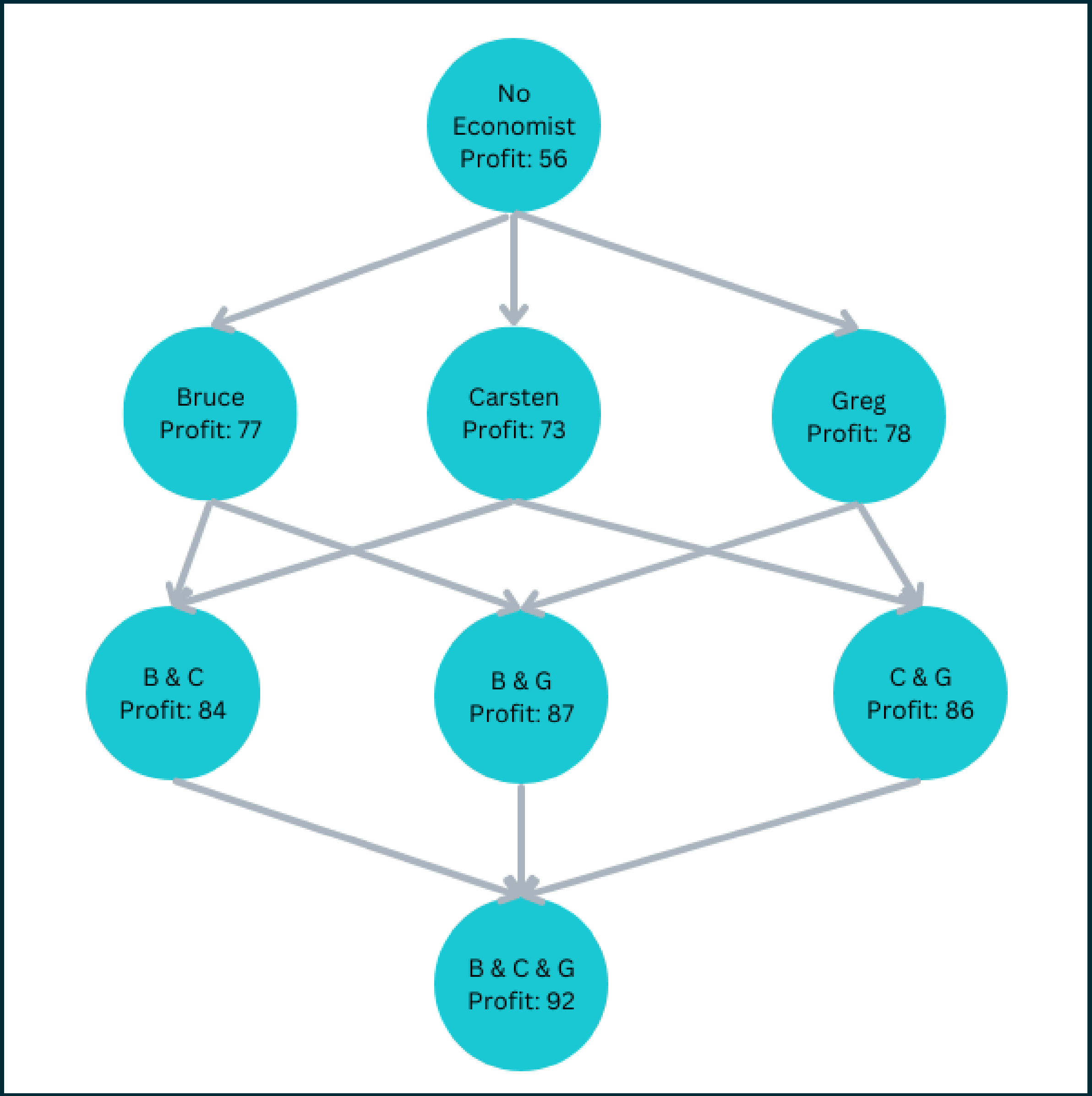


Calculate Greg' contribution as an exercise:

- Spoiler Alert (below is the result):
- Greg's average contribution (Shapley value):
  $S_G^{hap} = 13.83$

SHAPLEY VALUES FROM A GAME THEORY APPROACH

STARTING PROFIT – SHAP VALUES – FINAL PROFIT



Initial Profit: $Profit_0 = 56$

Bruce's SHAP: $S_B^{hap} = 12.33$

Carsten's SHAP: $S_B^{hap} = 9.83$

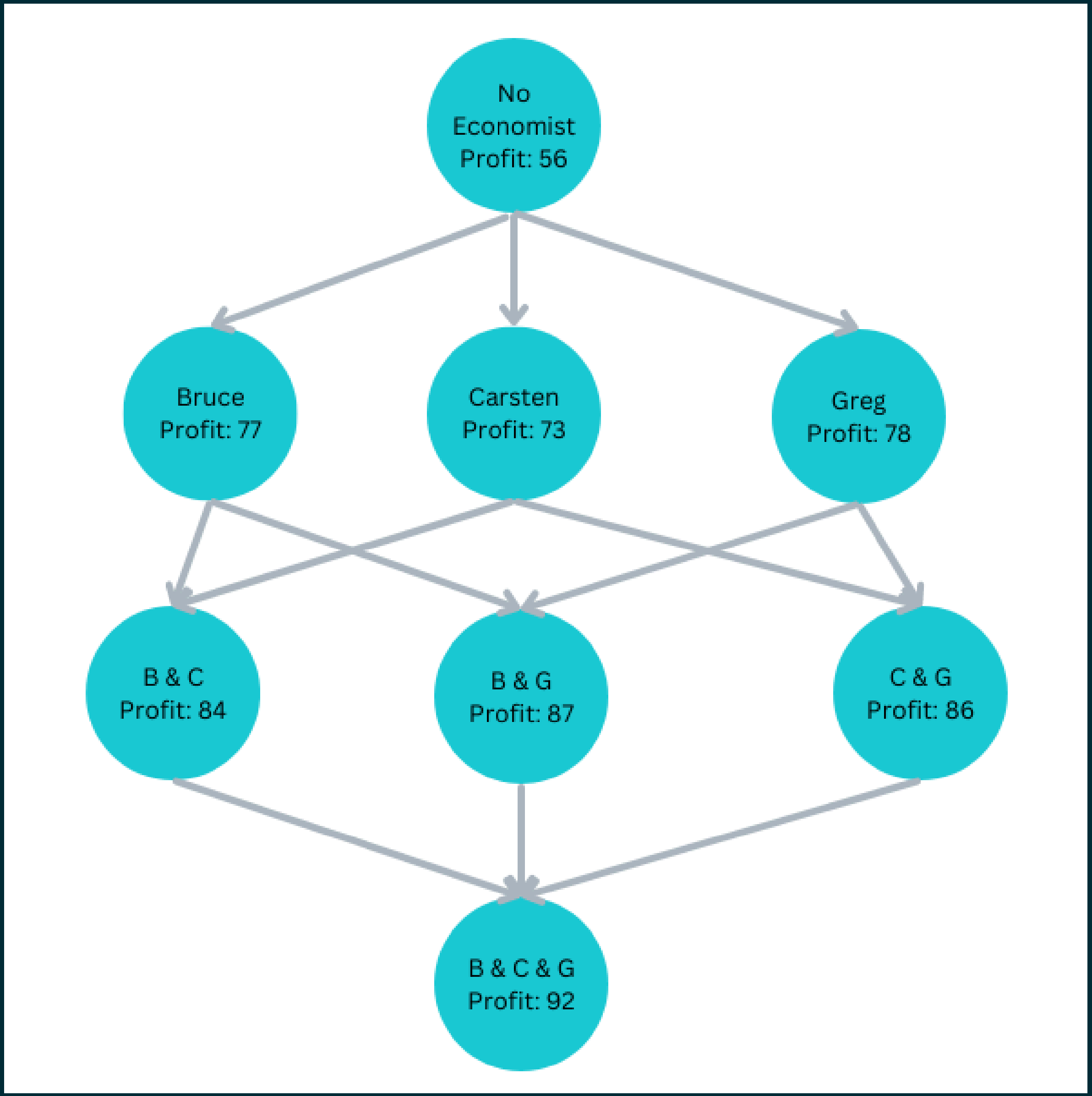
Greg's SHAP: $S_G^{hap} = 13.83$

Profit with 3: $Profit_3 = 92$

https://ai.lange-analytics.com/

SHAPLEY VALUES FROM A GAME THEORY APPROACH

COALITIONS AND PERMUTATIONS



- Number of coalitions: $N_{coalition} = 2^k = 2^3 = 8$
- Number of *joining* scenarios: $k! = 3! = 1 \cdot 2 \cdot 3 = 6$ BCG,BGC, CBG, CGB, GBC, GCB
- If all 6 scenarios are calculated, one can can calculate *SHAPley* values for all contributors (variables).
- If less than 6 scenarios are calculated, one can can estimate *SHAPley* values for all contributors (variables).

SHAPLEY VS SHAP

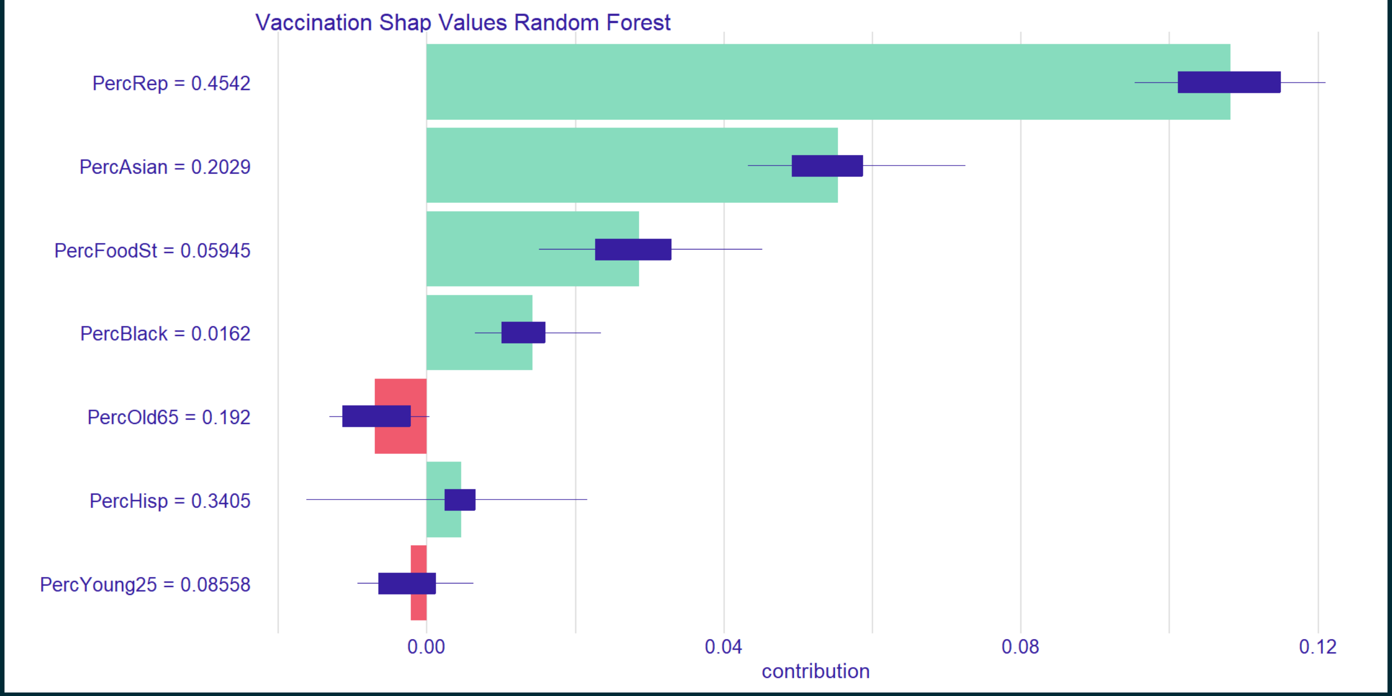
- **SHAPley Values** estimate contribution of players
- **SHAP** is a computer implementation to estimate *SHAPley* values for predictors.
  - The predictors become the players.
  - The contributions become the *SHAP* values.
  - For convinience we will use the terms *SHAPley Values* and *SHAP Values* interchangeable.
  - We will use the term **SHAP Values** to measure the contribution of specific variables-value combinations.
  - The average prediction (average outcome training data) plus all *SHAP* values equals the final prediction for the observation..

Example: The fact that the Republican vote was 62% in the analyzed county, lowered the predicted vaccination rate by 3%. Note, we use variable (`Rep` **and** value 62). The related *SHAP* value would be -0.03 »

LOCAL/MODEL AGNOSTIC: SHAP VALUES FOR ORANGE COUNTY

▶ Code

```
County: Orange, CA, Pred. Vac.: 0.72
```



```
Pred. Vac. Rate (all U.S. counties): 0.62
```

Mean PercRep: `r round(weighted.mean(DataTrain$P`
`as.numeric(DataTrain$Populatio`

Mean PercAsian: 0.05

Mean PercFoodSt: 0.11

Mean PercBlack: 0.11

Mean PercOld65: 0.21

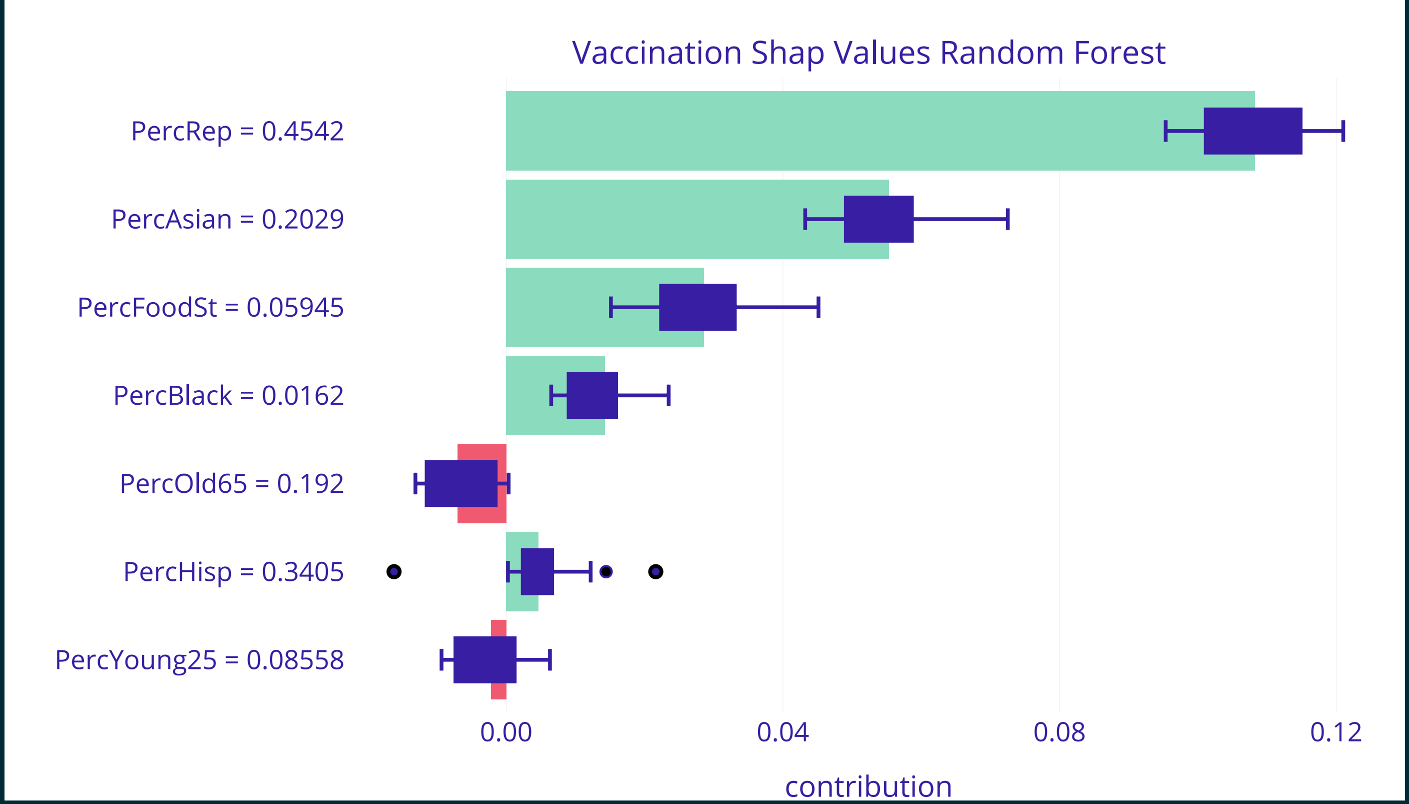
Mean PercOld65: 0.19

Mean PercYoung25: 0.09

https://ai.lange-analytics.com/

SHAP VALUES FOR ORANGE COUNTY

```
County: Orange, CA, Pred. Vac.: 0.72
```



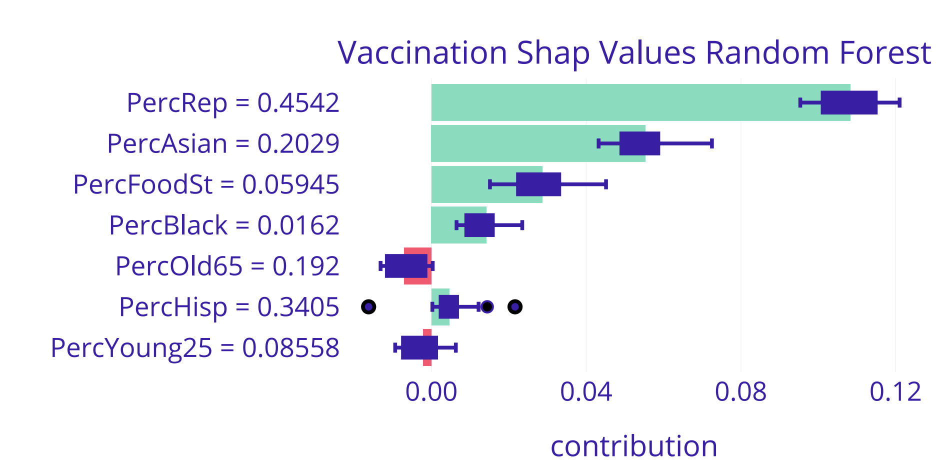
```
Pred. Vac. Rate (all U.S. counties; unweighted): 0.51
```

```
[1] "PercYoung25 = 0.08558"
[1] "PercHisp = 0.3405"
[1] "PercBlack = 0.0162"
[1] "PercFoodSt = 0.05945"
[1] "PercAsian = 0.2029"
[1] "PercRep = 0.4542"
```

🤓 Below is a link to an R script that allows you to create your own SHAP values in R.

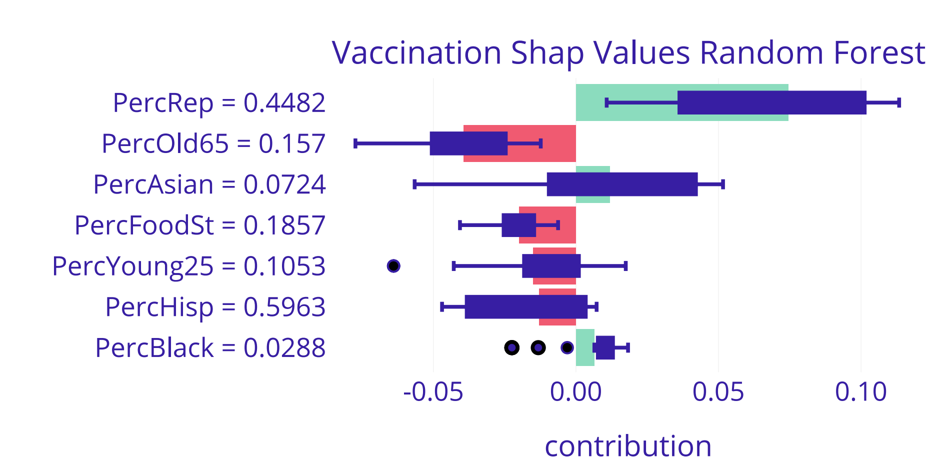
SHAP VALUES FOR TWO COUNTIES (DIFFERENT POLITIC. IMPACT)

County: Orange, CA, Pred. Vac.: 0.72



Pred. Vac. Rate (all U.S. counties): 0.62

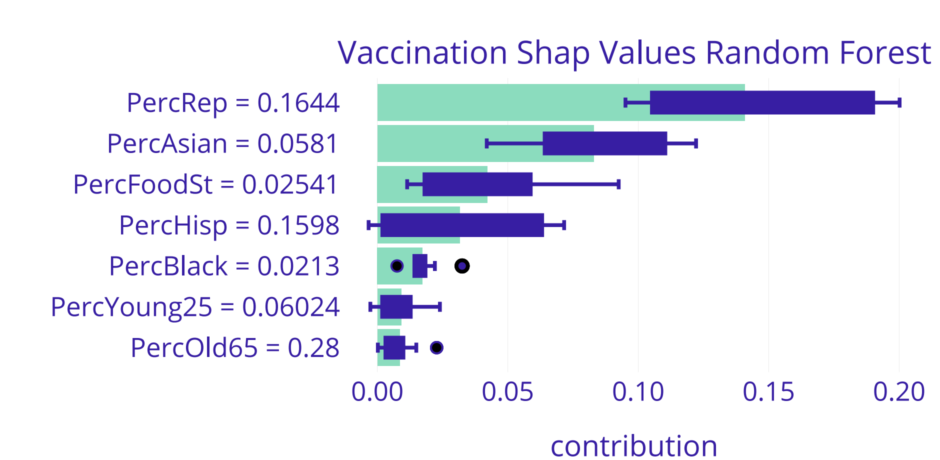
County: Merced, CA, Pred. Vac.: 0.52



Pred. Vac. Rate (all U.S. counties): 0.62

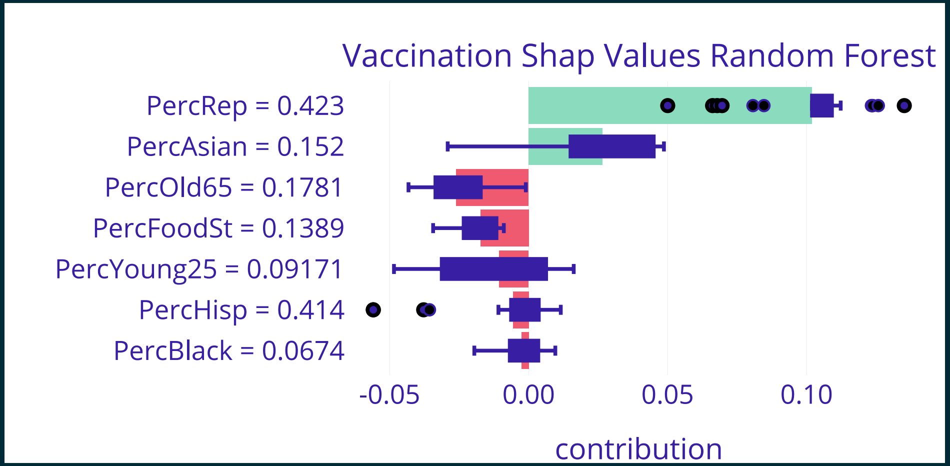
SHAP VALUES FOR TWO COUNTIES (DIFFERENT ASIAN IMPACT)

County: Marin, CA, Pred. Vac.: 0.85



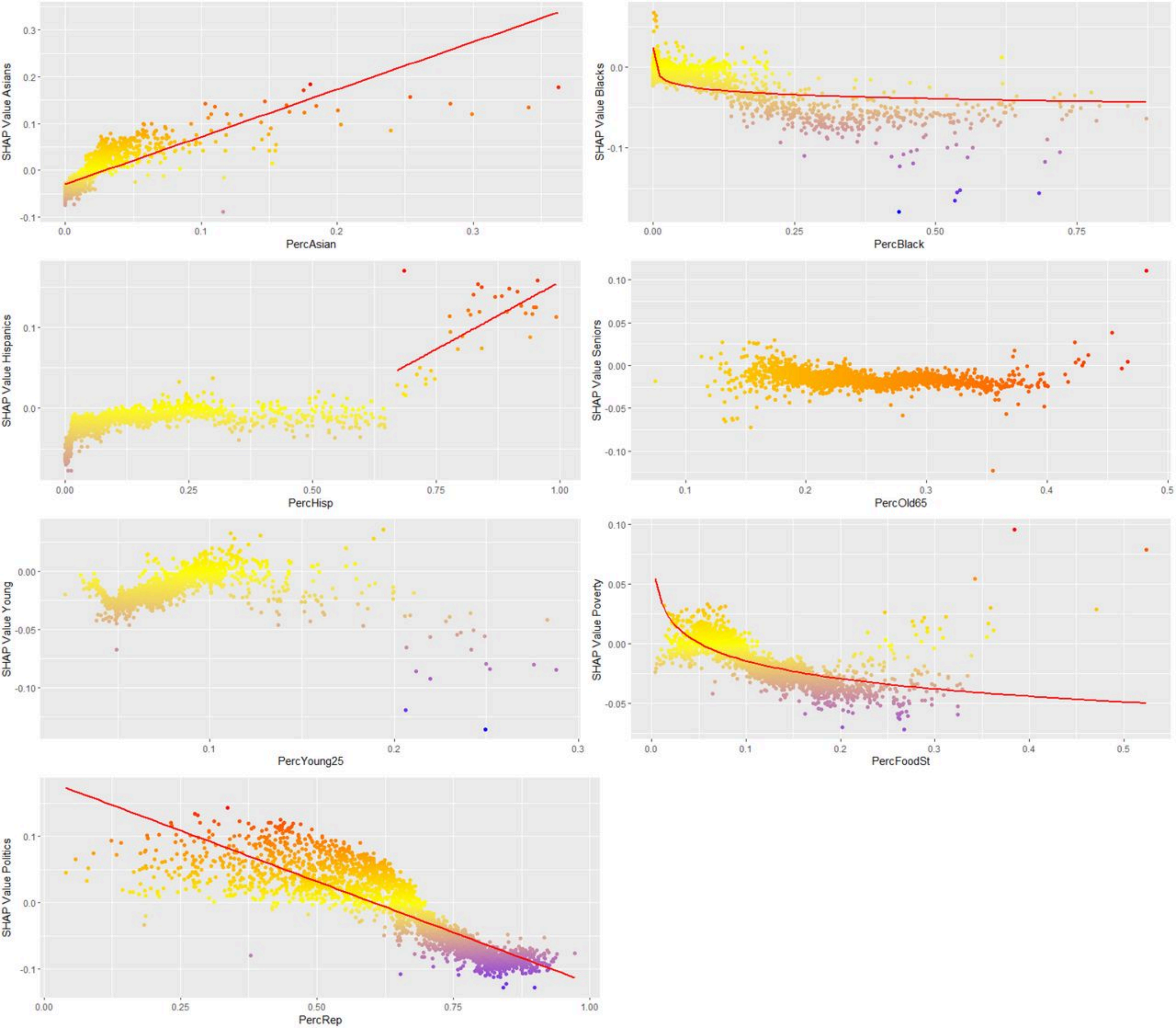
Pred. Vac. Rate (all U.S. counties): 0.62

County: San Joaquin, CA, Pred. Vac.: 0.58



Pred. Vac. Rate (all U.S. counties): 0.62

SHAP VALUES BY VARIABLE



https://ai.lange-analytics.com/

https://ai.lange-analytics.com/