

Mark Kotanchek, Evolved Analytics

Reducing Barriers

The DataModeler GUI

What is DataModeler?



Need & Perspective Matter

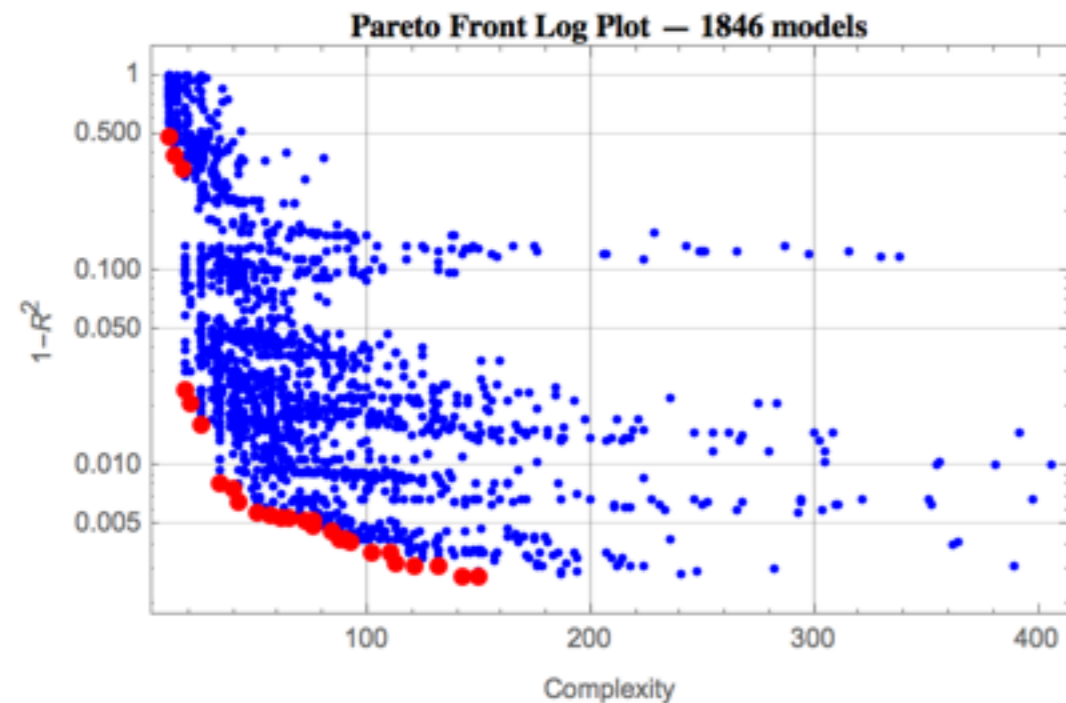


- ❖ Rapid data analysis
- ❖ Human-interpretable models
- ❖ Insight generator
- ❖ Variable selection
- ❖ Trustable models
- ❖ Active design of experiments

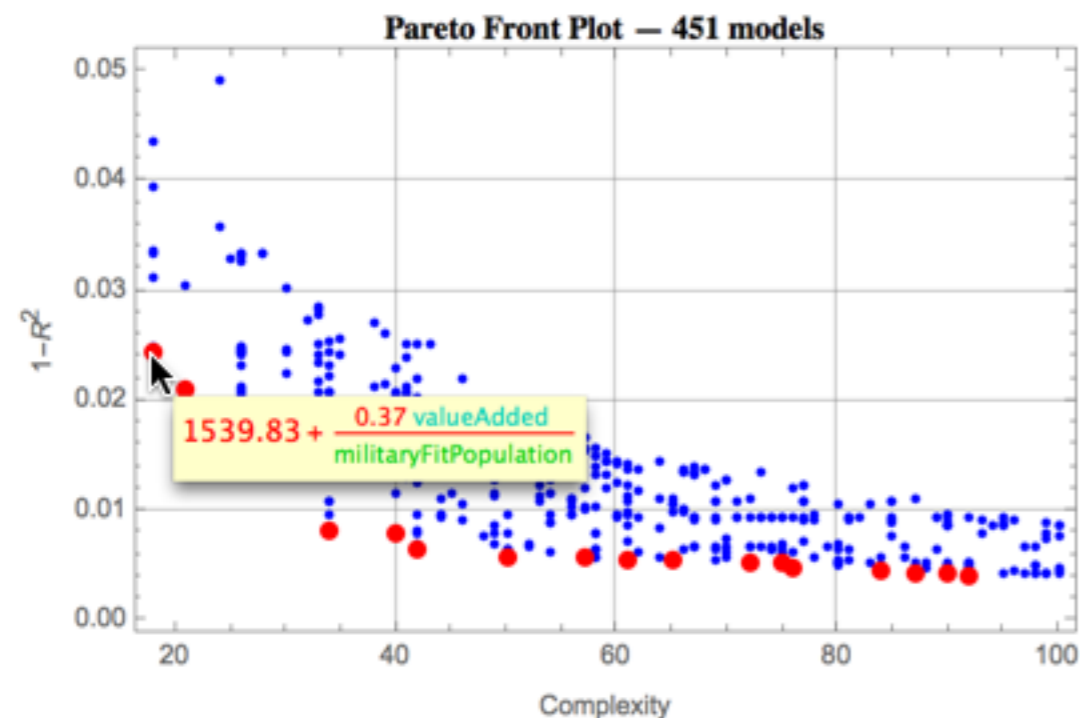
Analysis for
the Real World

Symbolic Regression

```
ParetoFrontLogPlot[  
  round1Archive = RetrieveModelSets[  
    ProjectName -> "GDPPerCapitaModeling",  
    RoundName -> "Round1"  
  ]  
]
```

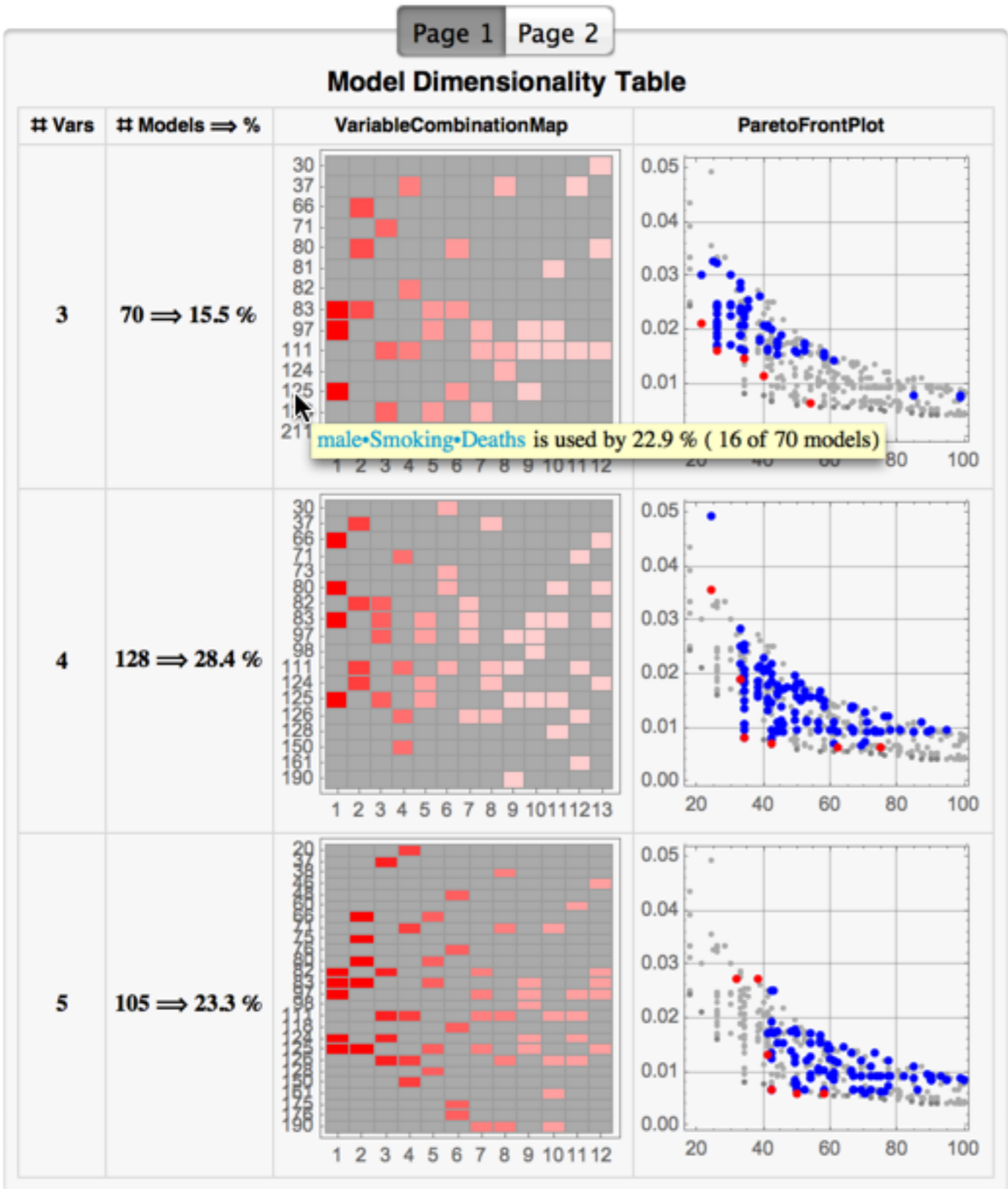


```
interesting1Models = SelectModels[  
  round1Archive,  
  QualityBox -> {100, 0.05}  
];  
ParetoFrontPlot@interesting1Models
```



- ❖ Evolve models from variables, operators and constants
- ❖ Reward simplicity & accuracy
- ❖ Many disparate hypotheses are developed, assessed and refined
- ❖ different model structures
- ❖ different variable combinations

```
ModelDimensionalityTable[interestingModels, MaxRowsPerPage -> 1/2]
```



Optimizer not Optimizer

- ❖ Human provides the initial framework
- ❖ Data provides the story
- ❖ Human extracts the insight and meaning
- ❖ Breaks the conventional hypothesis-testing approach

VariableCombinationTable[interestingModels, SignificanceLevel → 4]

Variable Combination Table

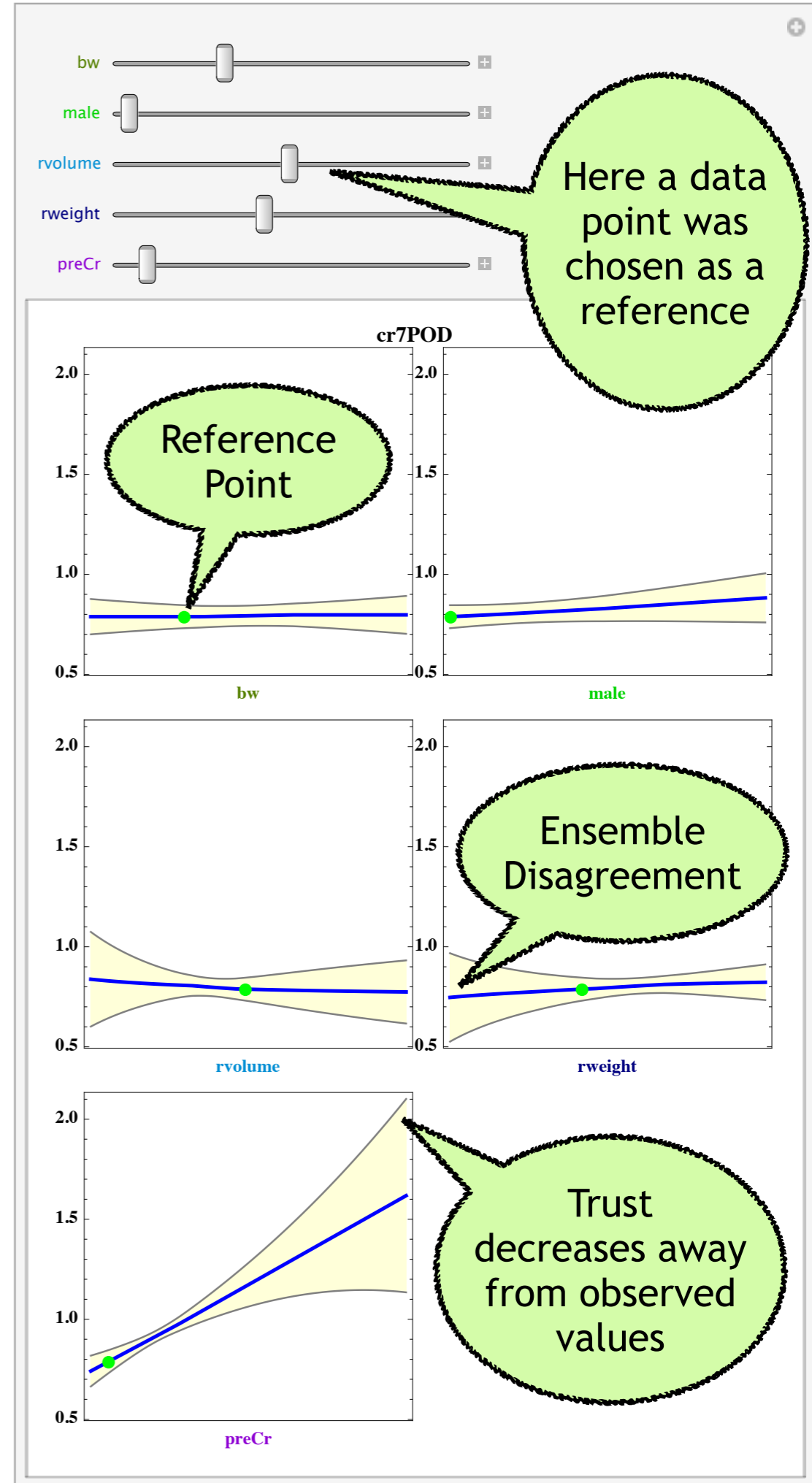
	num ⇒ %	Variables Used	ParetoFrontPlot
1	16 ⇒ 3.5 %	maleElderlyPopulation militaryFitMales nationalIncome male•Smoking•Deaths	
2	12 ⇒ 2.7 %	femalePopulation miscellaneousValueAdded valueAdded world•Bank•Income•Group	
3	10 ⇒ 2.2 %	miscellaneousValueAdded nationalIncome population male•Smoking•Deaths	
4	10 ⇒ 2.2 %	nationalIncome population male•Smoking•Deaths	

Implications

- ❖ Transparency
 - ❖ human-interpretable models
 - ❖ metavariables
- ❖ Trustable Models
 - ❖ agree where constrained by data
 - ❖ diverge when not constrained
- ❖ Active Design-of-Experiments
 - ❖ collect data iteratively
 - ❖ target potential optima
 - ❖ target locations of uncertainty — maximum value experiment!

Trustable Models

- ❖ The impact of variables depend upon the parameter settings
- ❖ Visualization involves choosing a reference and looking at projections of the model variables
- ❖ Moving sliders allows “what if” exploration



Implementation

- ❖ 400+ Functions & Symbols
- ❖ Fully documented
- ❖ Intelligent functions & meta-functions
- ❖ Use notebook interface for analysis & audit trail
- ❖ 100% Mathematica

DataModeler – Wolfram Mathematica – Wolfram Mathematica 10.0

Mathematica >

DATA MODELER GUIDE Functions » Tutorials » URL »

▼ DataModeler

DataModeler is an incredibly powerful tool. The best way to get started is to explore the [tutorials](#) (which include a number of case studies illustrating key features using industrial data). In the short-term, the [function overview](#), [quick start](#) and [FAQ](#) are good places to get started as well as see some of the capabilities.

Finally, www.evolved-analytics.com is a good source for publications and additional information.

▼ Data Exploration

▼ Data Adjustments

- [MakeDataNumeric](#) · [MergeInputResponseData](#) · [SplitInputResponseData](#) · [RescaleData](#) · [AugmentData](#)

▼ Data Visualization

- [SmallPlot](#) · [CorrelationChart](#) · [UnivariatePlot](#) · [BivariatePlot](#) · [CorrelationMatrixPlot](#) · [DataSummaryTable](#) · [DataDistributionPlot](#) · [DataCompletenessPlot](#) · [DataCompletenessMap](#)

▼ Statistics

- [RobustCorrelationMatrix](#) · [ConfidenceEllipsoid](#) · [AbsCorrelation](#) · [MedianAverage](#) · [NoisePower](#) · [ScaleInvariantNoisePower](#) · [SummaryStatistics](#) · [WeightedMean](#) · [WeightedStandardDeviation](#)

▼ Data Subset Selection

- [ExtractDataSubset](#) · [UncorrelatedVariables](#) · [SubSample](#) · [ConfidenceEllipsoidSelection](#) · [ConfidenceEllipsoidSelectionIndices](#) · [NumericDataRecords](#) · [NumericDataRecordIndices](#) · [NonNumericDataRecords](#) · [NonNumericDataRecordIndices](#)

▼ Outlier Detection

- [DataStrangeness](#) · [DataOutlierAnalysis](#) · [DataOutliers](#) · [DataOutlierIndices](#) · [DataOutlierTable](#)

▼ Model Development (Symbolic Regression)

▼ Model Creation

- [BuildFunctionPatterns](#) · [RandomModels](#) · [CreateModelFromExpression](#) · [KeijzerExpansion](#) · [RandomGenomes](#) · [CreateModelFromGenome](#) · [ExtractGenomeSubtrees](#) · [CreateLinearModel](#) · [MetaVariableModels](#)

▼ Evaluating Model Quality

- [EvaluateModelQuality](#) · [UpdateModelQuality](#) · [RearrangeModelQuality](#) · [EvaluateModelQualityVsMultipleDataSets](#) · [UpdateModelQualityVsMultipleDataSets](#)

▼ Selecting Models

- [ParetoFront](#) · [RobustModels](#) · [UncorrelatedModels](#) · [NicheModels](#) · [SelectModels](#) · [UniqueModels](#) · [FitModels](#) · [UniqueFitnessModels](#) · [UniqueAndFitModels](#) · [NumericModelQualityQ](#) · [SelectPopulation](#) · [IndependentEvolutions](#)

▼ Model Selection Strategies

- [ParetoTourneySelect](#) · [ParetoEliteSelect](#) · [ParetoFrontSelect](#) · [TourneySelectWithoutReplacement](#) · [TourneySelect](#) · [RankBasedSelect](#) · [ProportionalSelect](#) · [EliteSelect](#) · [BestUniqueSelect](#) · [BestSelect](#) · [RandomSelectWithoutReplacement](#) · [RandomSelect](#) · [AllModels](#)

▼ Diversity Introduction

- [Clone](#) · [Crossover](#) · [MutateSubtree](#) · [DepthPreservingSubtreeMutation](#)

▼ Evolutionary Model Development

- [SymbolicRegression](#) · [ClassicGP](#) · [ParetoGP](#) · [ESSENCE](#) · [OrdinalGP](#)

▼ Other Model Development Approaches

- [KeijzerExpansion](#) · [CreateLinearModel](#) · [ConvertToFittedModel](#)

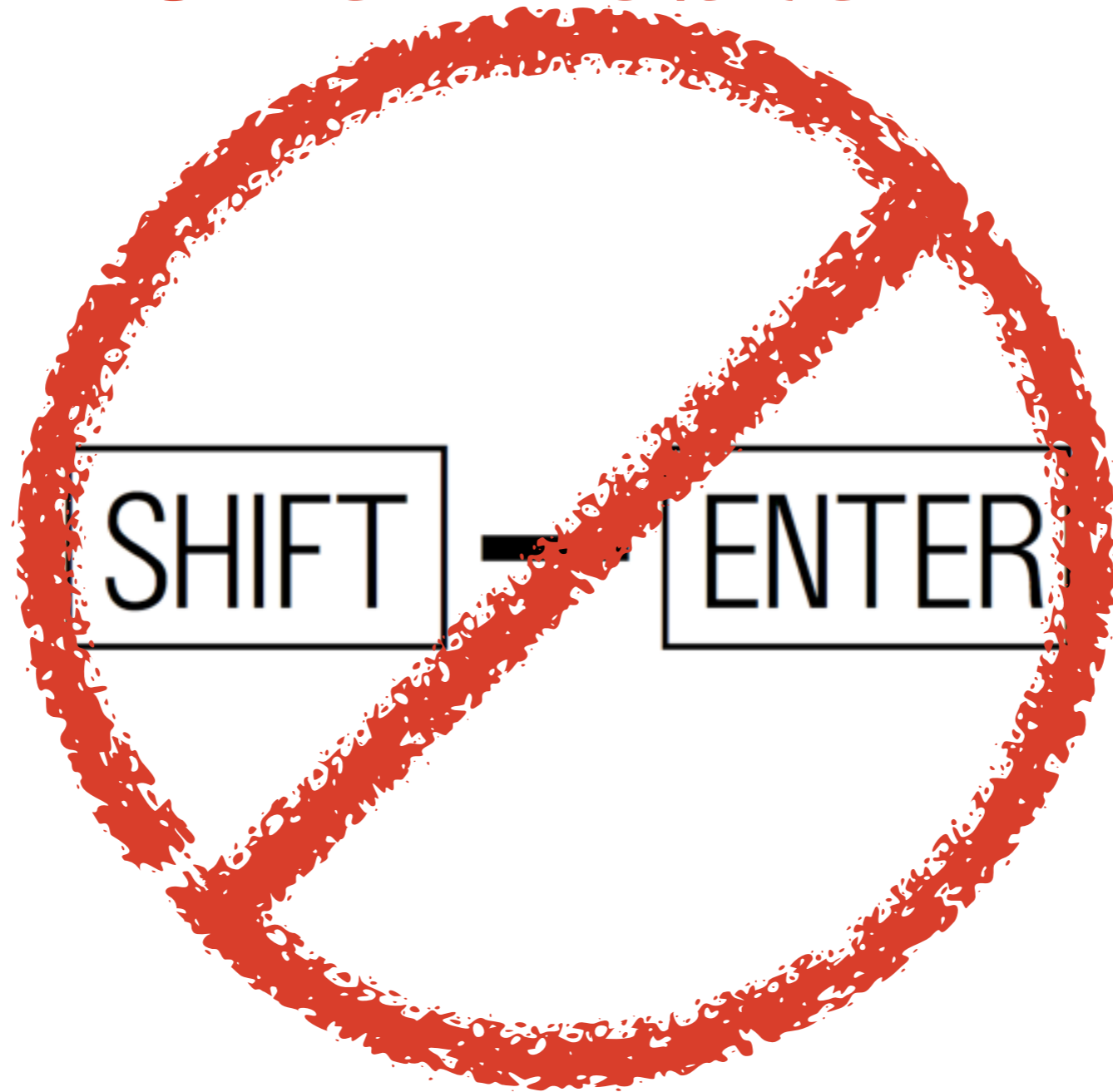
▼ Pre-defined SymbolicRegression Option Sets

100% ▸

One Problem

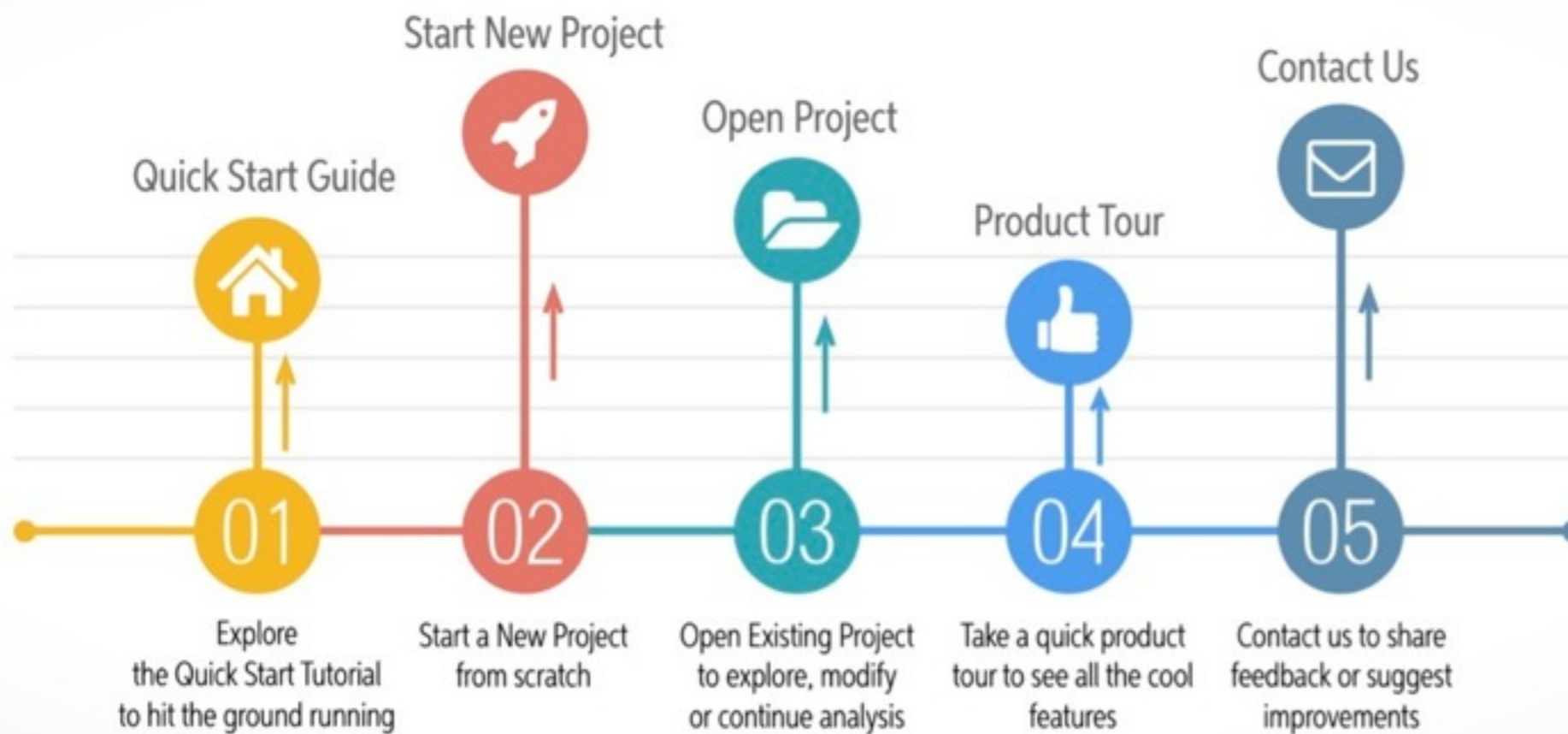
SHIFT — ENTER

One Problem



DataModelerUI

Welcome to Evolved Analytics' **DataModeler**



- ❖ No Shift-Enter
- ❖ Balance ease-of-use vs. power
- ❖ 100% Mathematica
- ❖ Data-centric

Demonstration

Data can contain non-numeric

The screenshot shows the DataModeler 1.0 interface. The main window displays a dataset with 240 rows and 231 columns. The current view shows the first 25 rows and 10 columns. The columns are: ID, Country, Continent, AdultPopulation, AgriculturalValueAdded, Airports, AMRadioStations, AnnualBirths, AnnualDeaths, ArableLandArea, and ArableLandFraction. The data includes various countries and their corresponding values for these metrics.

ID	Country	Continent	AdultPopulation	AgriculturalValueAdded	Airports	AMRadioStations	AnnualBirths	AnnualDeaths	ArableLandArea	ArableLandFraction
1	Alghanistan	Asia	1.80538 × 10 ⁷	4.74724 × 10 ⁹	51.	21.	1.03398 × 10 ⁶	238 866.	79 115.5	0.1213
2	Albania	Europe	2.22943 × 10 ⁶	2.30567 × 10 ⁹	5.	13.	41 355.	22 257.	5507.	0.201
3	Algeria	Africa	2.54063 × 10 ⁷	1.32765 × 10 ¹⁰	143.	25.	951 014.	236 729.	75 501.2	0.0317
4	American Samoa	Oceania	38 337.	.	3.	2.	1221.	186.	19.9	0.1
5	Andorra	Europe	60 280.	1.94356 × 10 ⁷	0.	0.	611.	492.	10.	0.0213
6	Angola	Africa	1.07557 × 10 ⁷	2.92319 × 10 ⁹	192.	21.	964 778.	301 325.	33 037.5	0.0265
7	Anguilla	NorthAmerica	10 765.	3.57407 × 10 ⁶	3.	2.	163.	55.	0.	0.
8	Antigua and Barbuda	NorthAmerica	61 414.	3.09111 × 10 ⁷	3.	4.	1485.	555.	80.5	0.1818
9	Argentina	SouthAmerica	2.69058 × 10 ⁷	3.00401 × 10 ¹⁰	1130.	260.	693 496.	322 033.	274 490.	0.1003
10	Armenia	Asia	2.13897 × 10 ⁶	1.86928 × 10 ⁹	12.	9.	40 163.	26 220.	4732.5	0.1678
11	Aruba	NorthAmerica	78 049.	1.0626 × 10 ⁷	1.	2.	1055.	880.	19.	0.1053
12	Australia	Oceania	1.54674 × 10 ⁷	2.50581 × 10 ¹⁰	464.	262.	310 621.	155 895.	468 503.	0.0615
13	Austria	Europe	5.69293 × 10 ⁶	6.26523 × 10 ⁹	55.	2.	81 684.	80 299.	13 677.6	0.1659
14	Azerbaijan	Asia	6.86262 × 10 ⁶	2.69443 × 10 ⁹	34.	10.	165 640.	66 270.	17 038.1	0.2062
15	Bahamas	NorthAmerica	252 744.	1.15257 × 10 ⁸	62.	3.	5820.	2354.	58.1	0.0058
16	Bahrain	Asia	1.06341 × 10 ⁶	7.21965 × 10 ⁷	3.	2.	20 021.	3121.	20.9	0.0282
17	Bangladesh	Asia	1.01777 × 10 ⁸	1.45734 × 10 ¹⁰	17.	15.	3.12231 × 10 ⁶	886 581.	72 100.1	0.5539
18	Barbados	NorthAmerica	196 849.	9.86743 × 10 ⁷	1.	2.	3607.	2590.	160.	0.3721
19	Belarus	Europe	6.74209 × 10 ⁶	5.34357 × 10 ⁹	65.	28.	101 941.	145 734.	54 316.3	0.2677
20	Belgium	Europe	7.02345 × 10 ⁶	3.45699 × 10 ⁹	45.	7.	129 565.	111 472.	8302.2	0.2742
21	Belize	NorthAmerica	207 475.	1.44595 × 10 ⁸	42.	1.	7837.	1509.	695.6	0.0305
22	Benin	Africa	5.16177 × 10 ⁶	2.15606 × 10 ⁹	5.	1.	381 648.	99 652.	26 029.4	0.2353
23	Bermuda	NorthAmerica	46 147.	5.40807 × 10 ⁷	1.	5.	718.	514.	10.8	0.2
24	Bhutan	Asia	511 739.	2.79354 × 10 ⁸	2.	0.	14 691.	4941.	883.1	0.023
25	Bolivia	SouthAmerica	6.26694 × 10 ⁶	1.744 × 10 ⁹	952.	171.	276 382.	76 339.	30 115.8	0.02

Launch Project

Select Variables

Explore Data

Generate Models

Analyze Models

Test & Validate

What If?

Reports



Add currently displaying output to the items in project Reports.

Target Response

FemaleLiteracyFraction

Target Dataset

 Project Test & Validate What If?

Variable Set

All

Exploring Methods

Max Explorers

1

 DataSummaryTable UnivariatePlot DataDistributionPlot DataCompletenessMap DataCompletenessPlot BivariatePlot CorrelationChart CorrelationMatrixPlot

DataSummaryTable

1	Country	ABC	100%		240	Lots of different values		
2	Continent	ABC	100%		6			
3	AdultPopulation	123	96%		232			
4	AgriculturalValueAdded	123	86%		208			
5	Airports	123	99%		119			
6	AMRadioStations	123	95%		71			
7	AnnualBirths	123	95%		230			
8	AnnualDeaths	123	95%		228			
9	ArableLandArea	123	97%		214			
10	ArableLandFraction	123	97%		209			
11	Area	123	100%		239			
12	BirthRateFraction	123	94%		221			
13	BoundaryLength	123	100%		236			
14	CellularPhones	123	95%		220			
15	ChildPopulation	123	96%		232			
16	CoastlineLength	123	100%		179			
17	ConstructionValueAdded	123	87%		209			
18	CropsLandArea	123	98%		200			
19	CropsLandFraction	123	98%		166			
20	DeathRateFraction	123	94%		202			

We have the ability to output selected graphics to a report

DataModeler 1.0 - [FemaleLiteracyFraction]

Launch Project | Select Variables | Explore Data | Generate Models | Analyze Models | Test & Validate | What If? | Reports

Target Response: FemaleLiteracyFraction (35)

Target Dataset: Project (selected), Test & Validate, What If?

Variable Set: All

Exploring Methods

- Max Explorers: 1
- DataSummaryTable
- UnivariatePlot
- DataDistributionPlot
- DataCompletenessMap
- DataCompletenessPlot
- BivariatePlot
- CorrelationChart
- CorrelationMatrixPlot

1 Country

Max Rows Per Page: 8

Sort By: FemaleLiteracyFraction

Image Size: 500

Aspect Ratio: Default

Ticks: None

Graphics Array Columns: 5

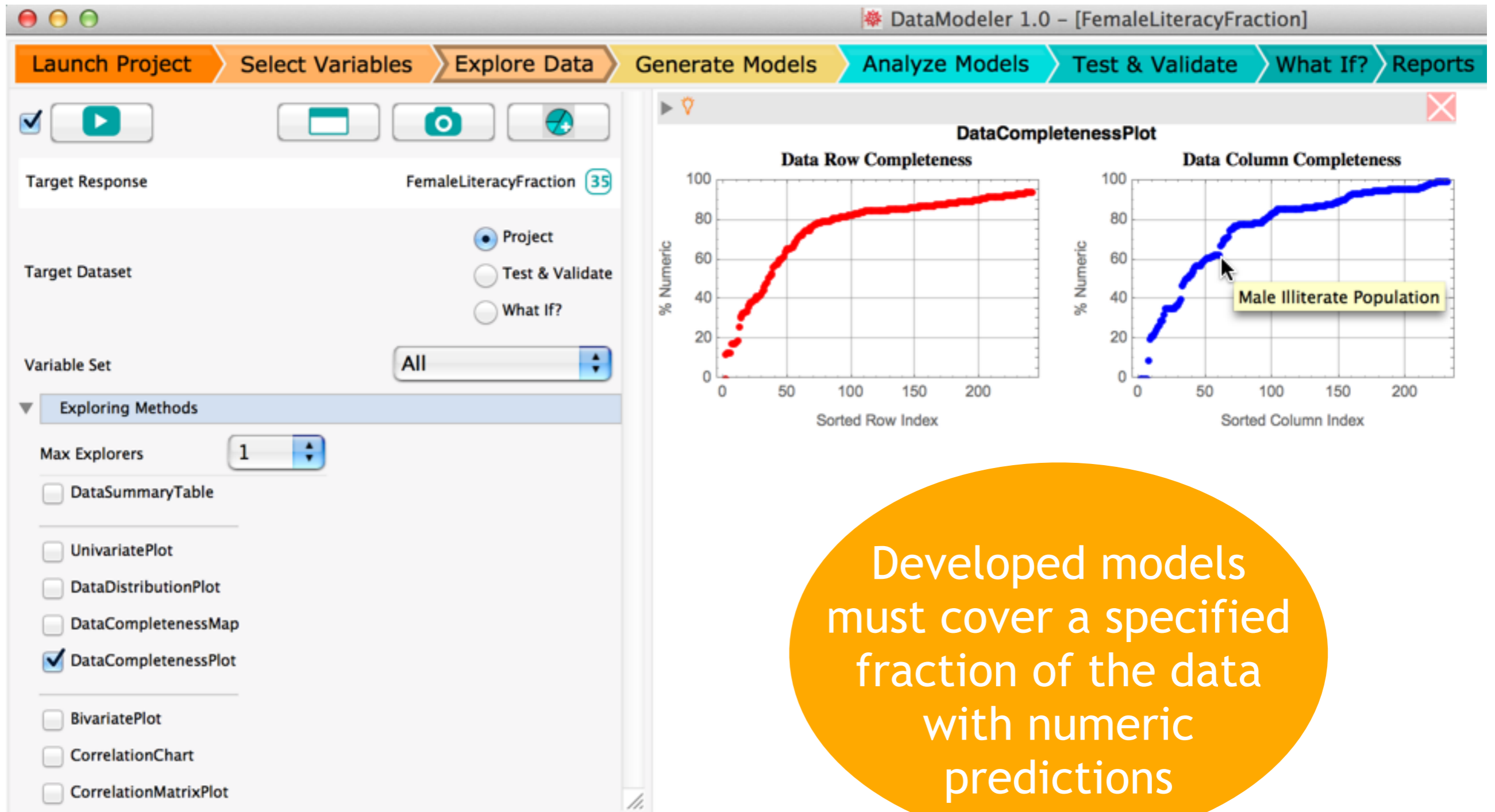
Reset <<<

UnivariatePlot

AdultPopulation	AgriculturalValueAdd	Airports	AMRadioStations	AnnualBirths
AnnualDeaths	ArableLandArea	ArableLandFraction	Area	BirthRateFraction
BoundaryLength	CellularPhones	ChildPopulation	CoastlineLength	ConstructionValueAdd
CropsLandArea	CropsLandFraction	DeathRateFraction	EconomicAid	ElderlyPopulation
ElectricityConsumption	ElectricityExports	ElectricityImports	ElectricityProduction	ExchangeRate

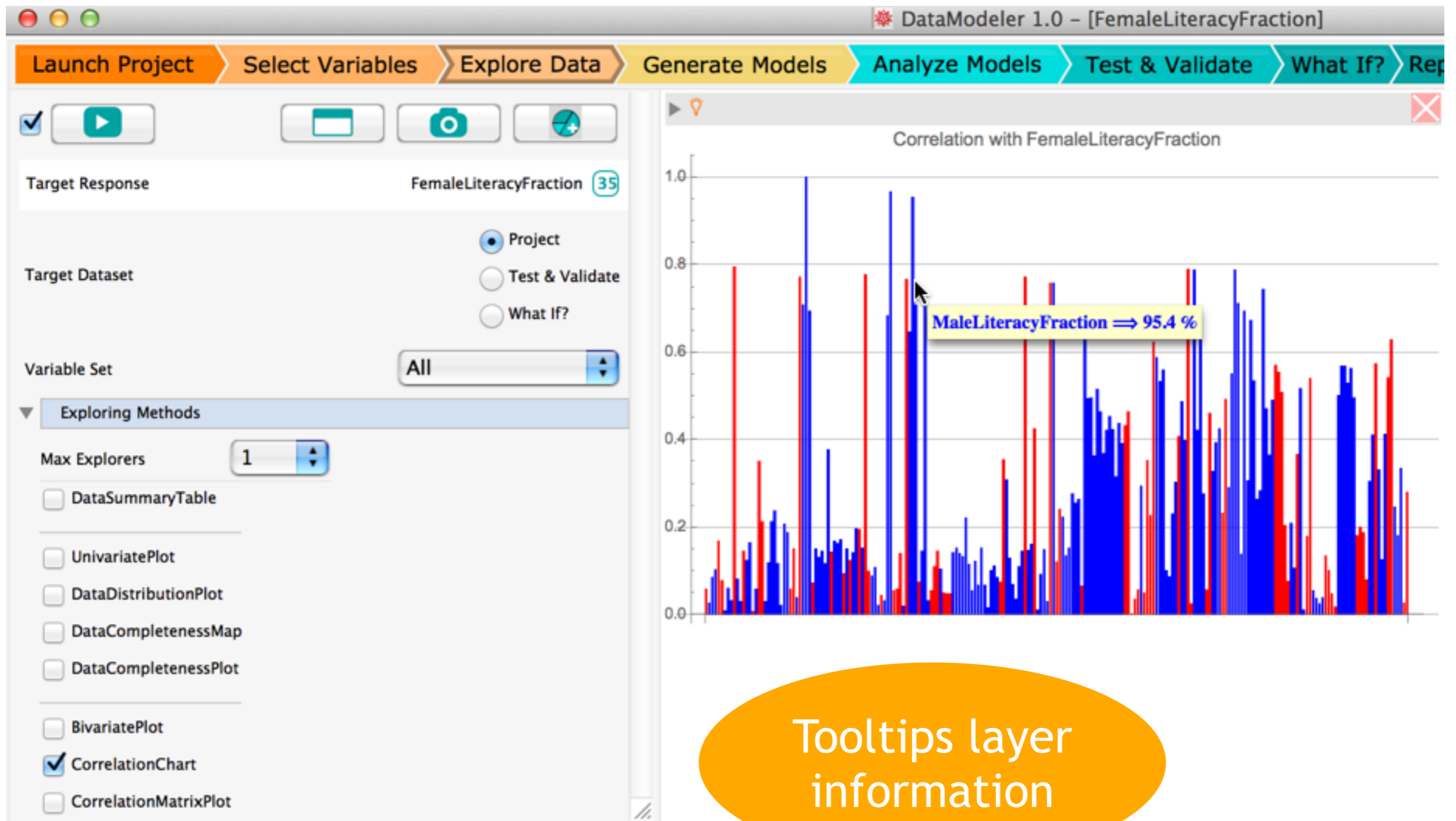
Options for explorers provide fine control but intelligent defaults

Data Completeness



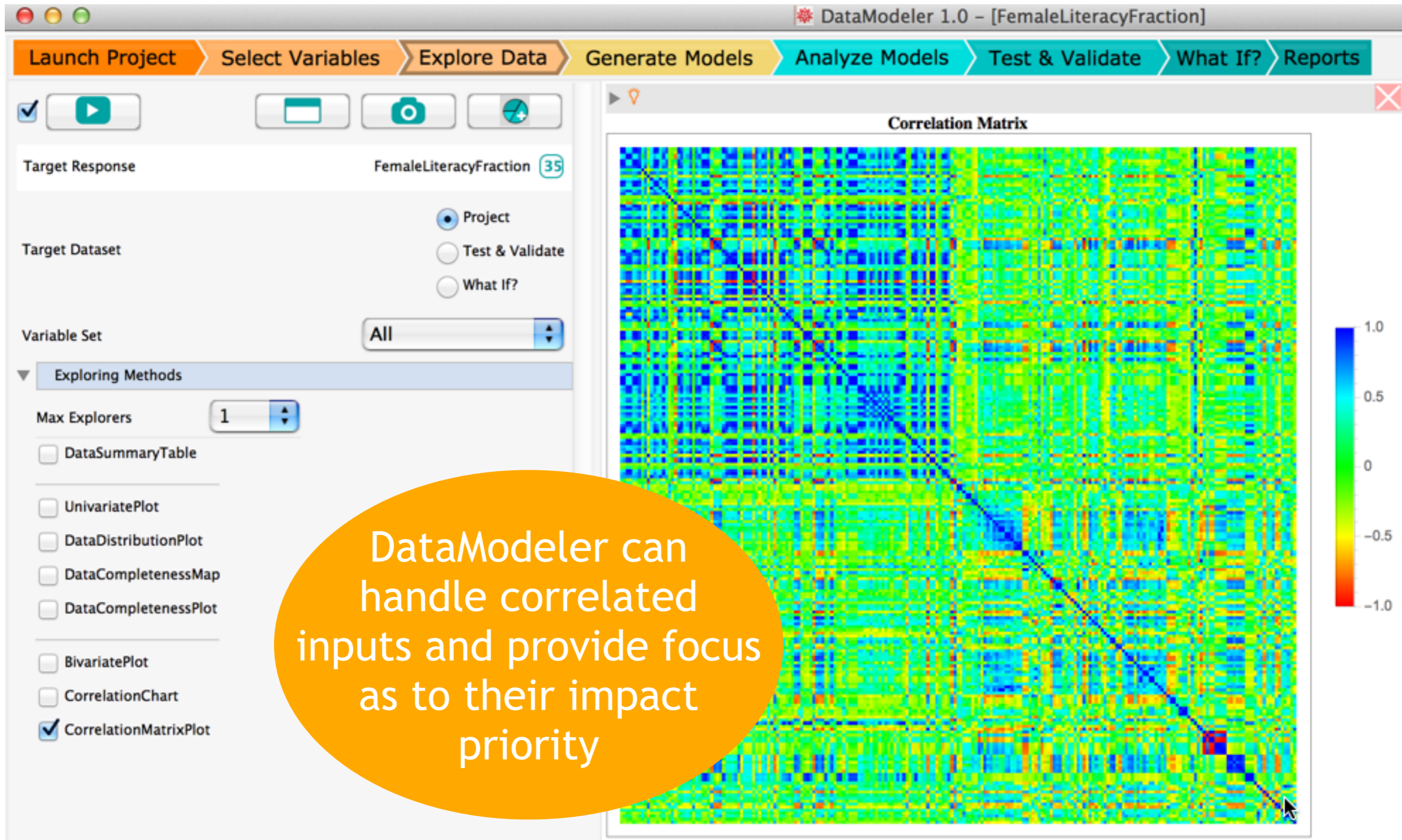
Developed models must cover a specified fraction of the data with numeric predictions

Correlation Chart



Tooltips layer information

Correlation Matrix Plot



DataModeler can handle correlated inputs and provide focus as to their impact priority

DataModeler 1.0 - [FemaleLiteracyFraction]

Launch Project | Select Variables | Explore Data | Generate Models | Analyze Models | Test & Validate | What If? | Reports

Calculating. Please wait.

Best Accuracy Irrespective of Complexity: $R^2 = 0.96$

Target Response: FemaleLiteracyFraction (35)

Current Round: Round 1 | Edit Comments

Round Comments: Round 1 is open-ended using all available inputs with short (2-minute) searches. Saved

Setup & Calculate | Analyze

Symbolic Regression Calculate

Save Models:

Data Source: Full Dataset

Independent Evolutions: 4

Time Constraint: 0 h 21 m 0 s

Optimize Linear Model: Automatic

Rescale Data: None

Allowed Variables: All

Excluded Variables: None

Meta Variables: Select Modify

1 cigarettesPerCapita

Function Patterns

Advanced Options

Number of Variables: Automatic

Basis Set Limit: Automatic

Power Limit: None

Robust Models:

Numeric Prediction Requirement: 75%

Error = 1 - R²

Model Generation Sequence

Total models for FemaleLiteracyFraction: 45

1 - R²

Complexity

9.35 + femaleMedianAge - literacyFraction^{1/3} + literacyFraction - $\frac{1}{\text{maleLiteracyFraction}}$ - youth * Cigarette * Use

Inspectors provide feedback during the model search

Round comments let us track our assumptions and actions during the modeling workflow

On a multi-core CPU, multiple independent evolutions will run in parallel

As we develop understanding, we want to adjust the settings used in the model search

DataModeler 1.0 - [FemaleLiteracyFraction]

Launch Project > Select Variables > Explore Data > Generate Models > Analyze Models > Test & Validate > What If? > Reports

Target Response: FemaleLiteracyFraction (35)

Retrieve > Analyze

1. Round Name

2. Advanced Selection

Selected Models: 66 out of 2147

Delete

1. FemaleLiteracyFraction_35-femaleLiteracyFractio

2. FemaleLiteracyFraction_35-femaleLiteracyFractio

3. FemaleLiteracyFraction_35-femaleLiteracyFractio

4. FemaleLiteracyFraction_35-femaleLiteracyFractio

5. FemaleLiteracyFraction_35-femaleLiteracyFractio

6. FemaleLiteracyFraction_35-femaleLiteracyFractio

7. FemaleLiteracyFraction_U-interesting rd 3 models

8. FemaleLiteracyFraction_35-femaleLiteracyFractio

9. FemaleLiteracyFraction_35-femaleLiteracyFractio

10. FemaleLiteracyFraction_35-femaleLiteracyFractio

All Round Names
 Single-Model or Ensemble
 Single-Model Sets
 Model Ensembles
 User Selected Models
 E-finalEnsemble
 E-quid
 E-round5 Ensemble
 Round1
 Round2
 Round3
 Round4
 Round5
 S-anotherModel
 S-One Model
 U-finalModelCandidates
 U-interesting5Models
 U-interesting rd 3 models
 More...

Save Models

Pareto Front Context LogPlot - 2147 of 2539 selected

1-R²

Complexity

Many models are returned from each independent evolution (modeling run)

By default, each independent evolution (modeling run) archives a model file

Target Response: FemaleLiteracyFraction (35)

Retrieve Analyze

Select Models: 903/4260 [Reset]

Candidates Focus Ensembles

Quality Box: All (10 100) (0.092 0.25)

Selection Fraction: 50%

Number of Variables: All

Required Variables: None

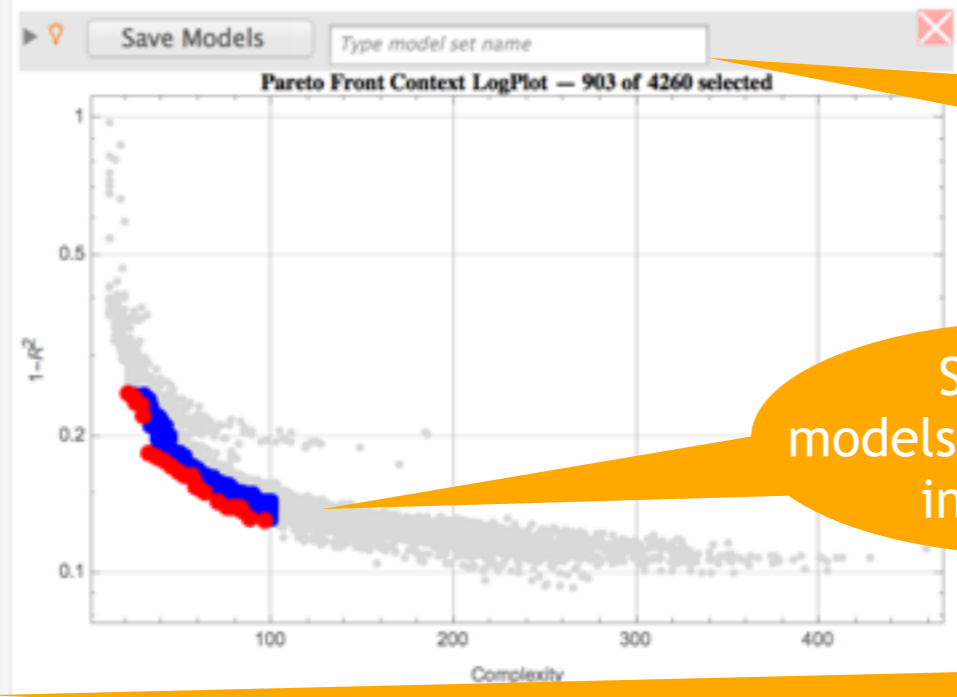
Allowed Variables: All

Excluded Variables: None

Power Limit: 3

Robust Models:

Model Ensemble: round 5 ensemble [Save]



Selected models can be archived

Selected models can be shown in context

We can apply filters to select subsets of models

MetaVariableDistributionTable

	Rank	# models	MetaVariable	# Evolutions	% Evolutions	Max Count		
++	1	640	communicable-Disease-Death-Rate countryLatitude	16	80.0	76	98.4	57.6
++	2	329	$\sqrt{\text{total}}$	19	95.0	50	71.4	31.7
++	3	76	female-Labor-Fraction improved-Sanitation-Access	6	30.0	50	58.1	5.3
++	4	222	$\sqrt[3]{\text{total}}$	17	85.0	38	100.0	27.2
++	5	114	cigarettes-Per-Capita communicable-Disease-Death-Rate countryLatitude	9	45.0	36	58.1	9.9
++	6	170	communicable-Disease-Death-Rate ²	16	80.0	34	57.6	14.9
++	7	51	communicable-Disease-Death-Rate female-Labor-Fraction	8	40.0	30	41.1	4.6
++	8	27	$\frac{1}{2.7+\text{total}}$	1	5.0	27	52.9	2.6
--	9	85	$\frac{1}{\text{cigarettes-Per-Capita}}$	12			56.7	10.2
++	10	38						
++	11	49	$\sqrt[3]{\text{recorded-consumption}}$					

The number of simultaneous explorers is adjustable

Interesting metavariables can be available in subsequent modeling as a building block

Target Response FemaleLiteracyFraction 35

Retrieve Analyze

Select Models 903/4260 Reset

Explore Models

Max Explorers 2

- ParetoFrontPlot
- ParetoFrontContextPlot
- ParetoFrontLogPlot
- ParetoFrontContextLogPlot
- ModelDimensionalityTable
- VariablePresenceChart
- VariablePresenceTable
- VariablePresenceDistributionChart
- VariablePresenceMap
- MetaVariableTable
- MetaVariableDistributionTable
- MetaVariableDistributionChart
- ModelSelectionReport
- ModelSelectionTable
- SelectionVariables
- VariableCombinationTable

Save round 4 interesting inputs

Driver Variables Off

Variables to Plot 0.1

Image Size 500

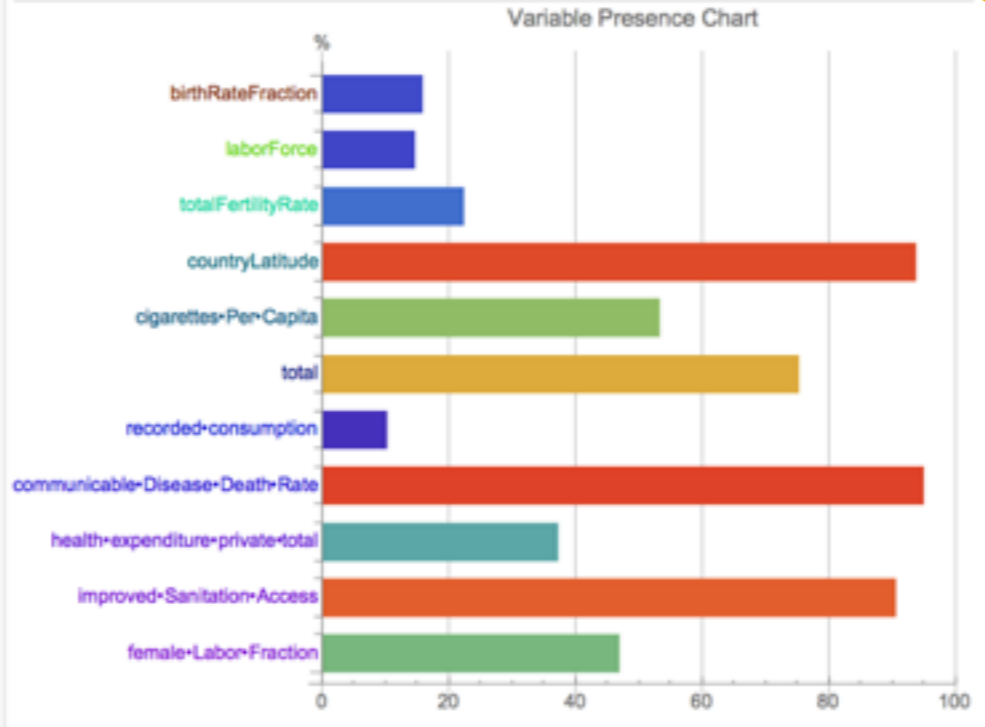
Aspect Ratio 1

Bar Origin Left

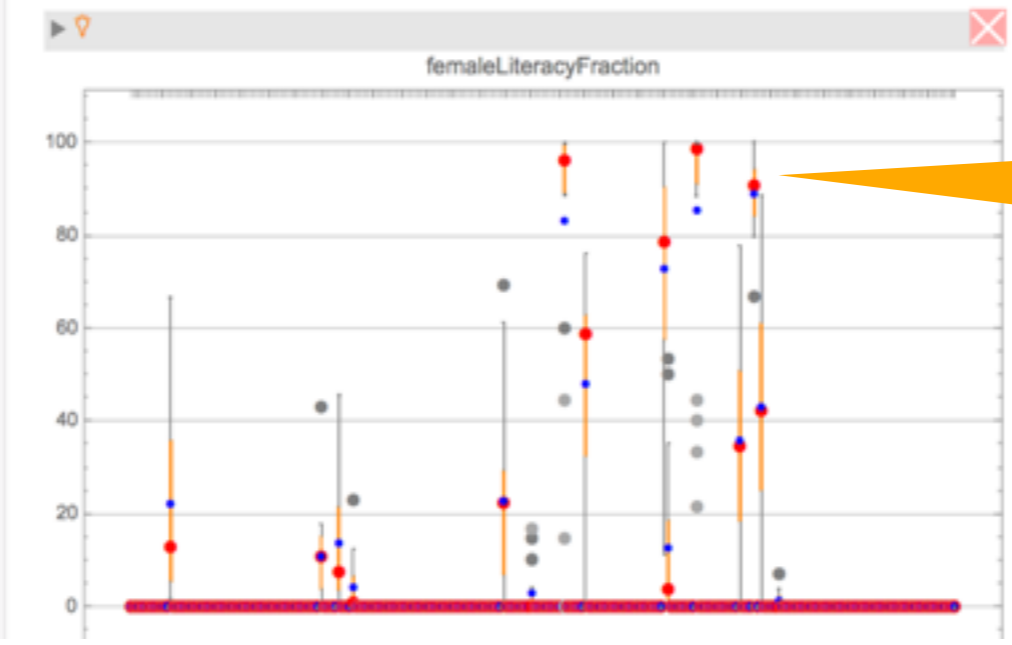
Reset <<<

We can apply filters and save the resulting variable set

Here we are looking at the presence of models in at least 10% of the models



The Variable Presence Distribution Chart shows the diversity of presence across independent model searches



Target Response FemaleLiteracyFraction 35

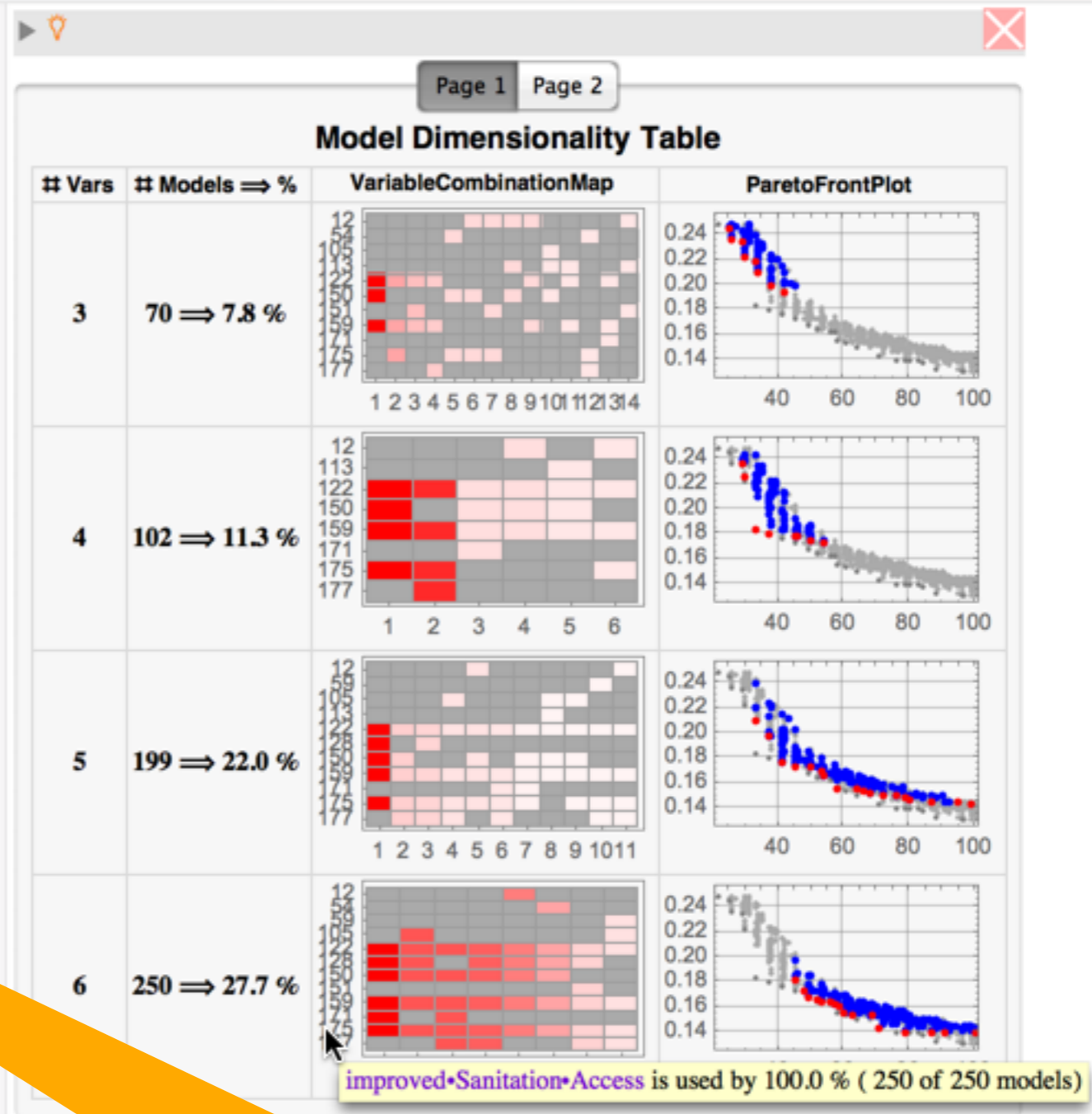
Retrieve Analyze

Select Models 903/4260 Reset

Explore Models

Max Explorers 1

- ParetoFrontPlot*
- ParetoFrontContextPlot*
- ParetoFrontLogPlot*
- ParetoFrontContextLogPlot*
- ModelDimensionalityTable
- VariablePresenceChart*
- VariablePresenceTable
- VariablePresenceDistributionChart
- VariablePresenceMap
- MetaVariableTable
- MetaVariableDistributionTable*
- MetaVariableDistributionChart
- ModelSelectionReport
- ModelSelectionTable
- SelectionVariables
- VariableCombinationTable*



The model dimensionality table lets us see how changing the number of variables affects the modeling potential and which variable combinations are most popular for a specific number of inputs

Target Response: FemaleLiteracyFraction 35

Retrieve Analyze

Select Models 903/4260 [Reset]

Explore Models

Max Explorers: 1

- ParetoFrontPlot*
- ParetoFrontContextPlot*
- ParetoFrontLogPlot*
- ParetoFrontContextLogPlot*
- ModelDimensionalityTable
- VariablePresenceChart*
- VariablePresenceTable
- VariablePresenceDistributionChart
- VariablePresenceMap
- MetaVariableTable
- MetaVariableDistributionTable*
- MetaVariableDistributionChart
- ModelSelectionReport
- ModelSelectionTable
- SelectionVariables
- VariableCombinationTable*
- VariableCombinationMap
- VariableCombinationChart

Max Rows Per Page: 5
Significance Level: 5
Reset: <<<

Variable Combination Table

num => %	Variables Used	ParetoFrontPlot
1 90 => 10.0 %	countryLatitude cigarettes-Per-Capita total communicable-Disease-Death-Rate improved-Sanitation-Access	
2 51 => 5.6 %	countryLatitude cigarettes-Per-Capita total communicable-Disease-Death-Rate health-expenditure-private-total improved-Sanitation-Access	
3 35 => 3.9 %	countryLatitude cigarettes-Per-Capita total communicable-Disease-Death-Rate health-expenditure-private-total improved-Sanitation-Access female-Labor-Fraction	
4 34 => 3.8 %	totalFertilityRate countryLatitude cigarettes-Per-Capita total communicable-Disease-Death-Rate improved-Sanitation-Access	
5 34 => 3.8 %	countryLatitude total communicable-Disease-Death-Rate health-expenditure-private-total improved-Sanitation-Access female-Labor-Fraction	

This function lets us look at the most popular combinations of inputs in the models being explored

We have the ability to define variable sets of interest from this view

Select these variables only in the Variables Manager:
TotalFertilityRate, CountryLatitude, Cigarettes Per Capita, Total, Communicable Disease Death Rate, Improved Sanitation Access

Target Response FemaleLiteracyFraction 35

Retrieve > Analyze

Select Models 1/1

Candidates Focus Ensembles

Model Picker

Approx. Max Number of Models

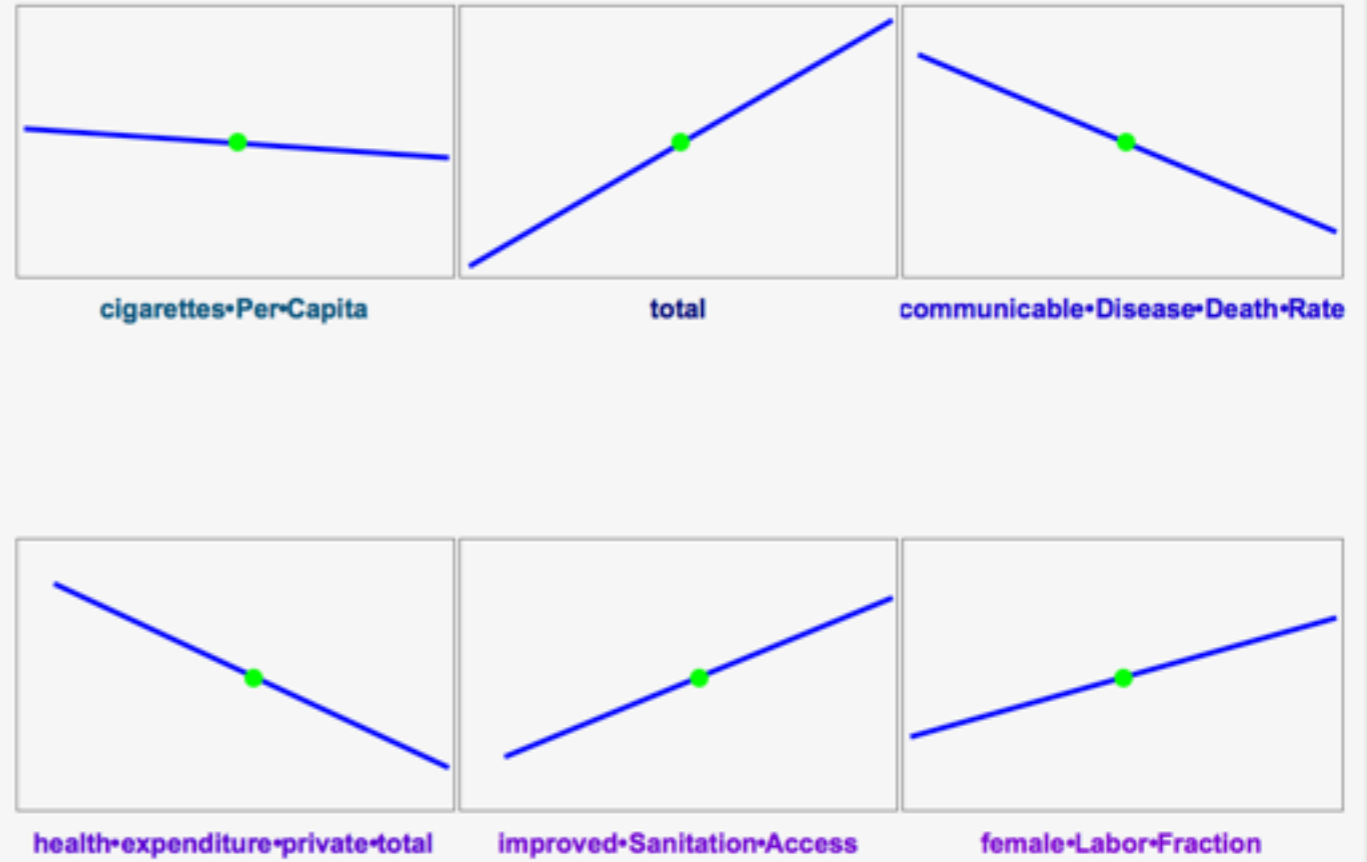
Approx. Min Number of Models

Explore Models

- Max Explorers
- ModelPredictionPlot
 - ModelPredictionComparisonPlot
 - ResponsePlot
 - ResponsePlotExplorer
 - ModelResidualPlot
-
- DataOutlierTable
-
- ParetoFrontPlot*
 - ParetoFrontContextPlot*
 - ParetoFrontLogPlot*
 - ParetoFrontContextLogPlot*

Cigarettes Per Capita	<input type="range"/>	1435.
Total	<input type="range"/>	8.8
Communicable Disease Death Rate	<input type="range"/>	41.3596
Health expenditure, private (total)	<input type="range"/>	49.0461
Improved Sanitation Access	<input type="range"/>	54.5
Female Labor Fraction	<input type="range"/>	32.9557

anotherModel [ref = 0.70]



The Response Plot Explorer lets us interactively explore the response behavior from a specified point



Target Response FemaleLiteracyFraction 35

Retrieve Analyze

Select Models 1/0

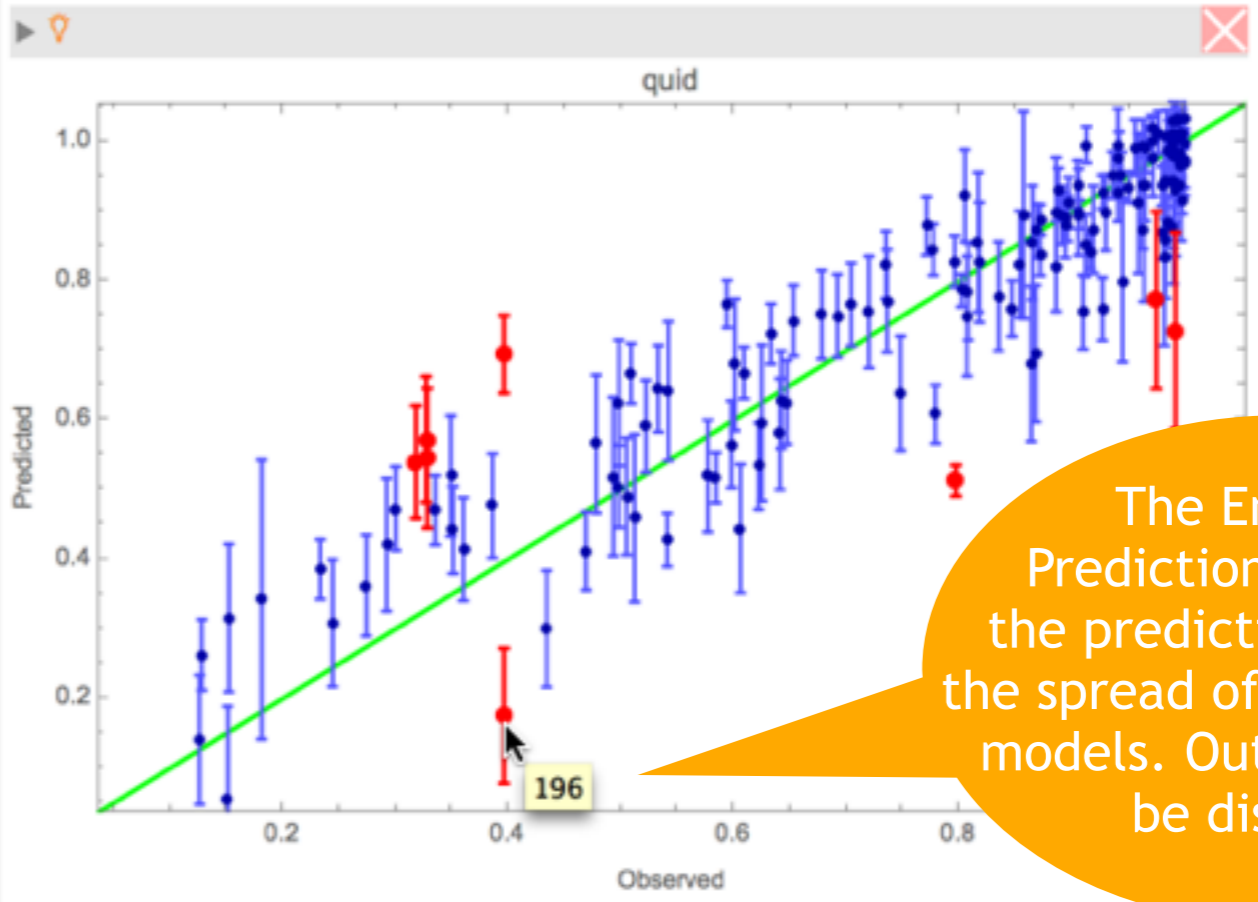
Candidates Focus Ensembles

quid

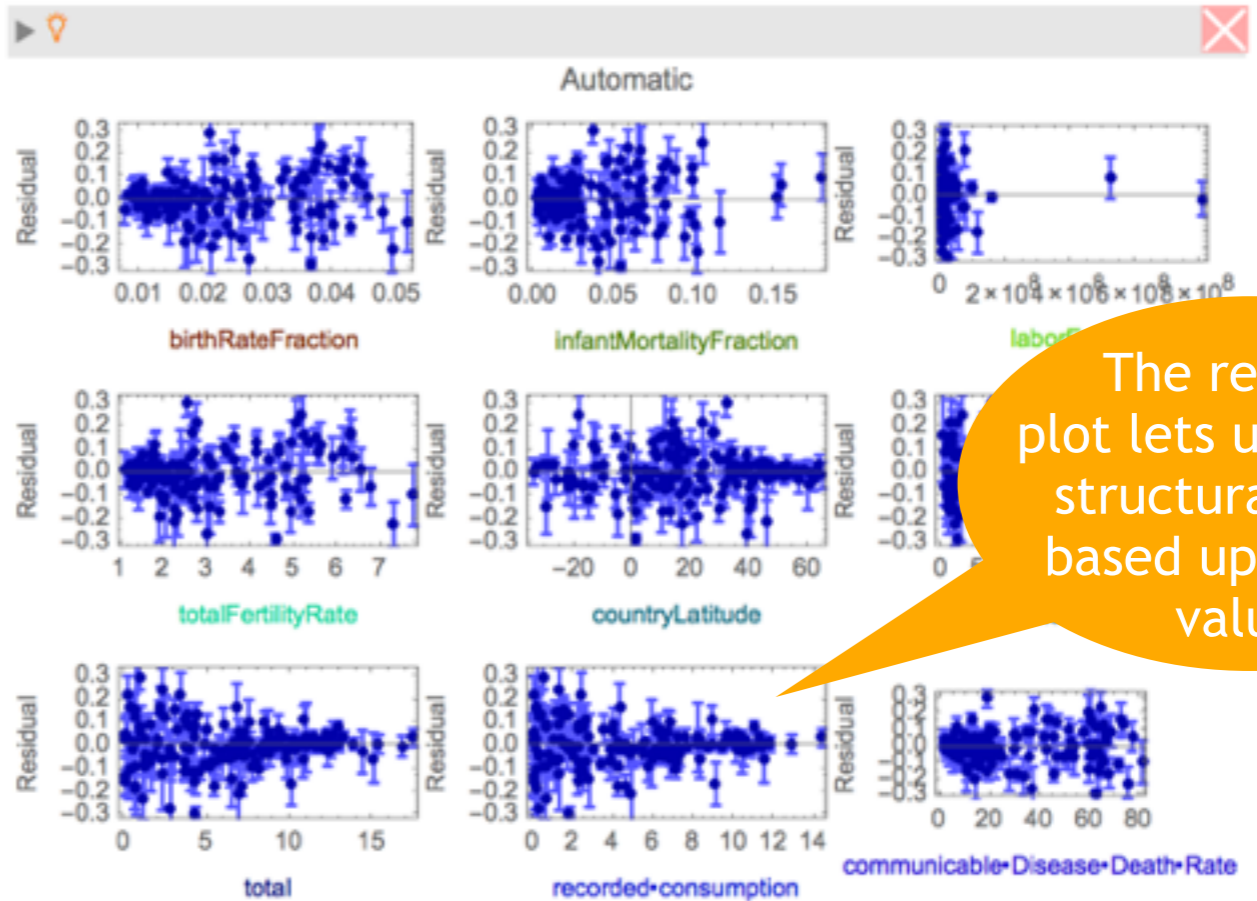
Explore Models

Max Explorers 2

- EnsemblePredictionPlot
- EnsembleResidualPlot
- ModelPredictionPlot
- ModelPredictionComparisonPlot
- ModelResidualPlot
- DataOutlierTable
- ParetoFrontPlot
- ParetoFrontContextPlot
- ParetoFrontLogPlot
- ParetoFrontContextLogPlot
- ModelDimensionalityTable
- VariablePresenceChart
- VariablePresenceTable
- VariablePresenceDistributionChart
- VariablePresenceMap
- MetaVariableTable
- MetaVariableDistributionTable
- MetaVariableDistributionChart



The Ensemble Prediction plot shows the prediction as well as the spread of the embedded models. Outliers can also be displayed



The residual plot lets us look for structural errors based upon input values



Target Response FemaleLiteracyFraction 35

Test Data > Models > Test

Select Models

- Models
 - another good simple model
- Model Ensembles
 - No ensembles retrieved.

Explore Models

Max Explorers 1

- ModelStatisticsTable
- ParetoFrontPlot
- ParetoFrontLogPlot
- ModelPredictionPlot
- ModelPredictionComparisonPlot
- ModelResidualPlot
- ModelSelectionReport
- ModelSelectionTable
- DataOutlierTable
- EnsemblePredictionPlot
- EnsembleResidualPlot

Fitted Model Properties

- Model Expression
- Model Phenotype
- C-Form Model
- Fortran-Form Model
- R Squared
- Adjusted R-Squared
- ANOVA Table
- Parameter Table
- Prediction Data Report

Reset <<<

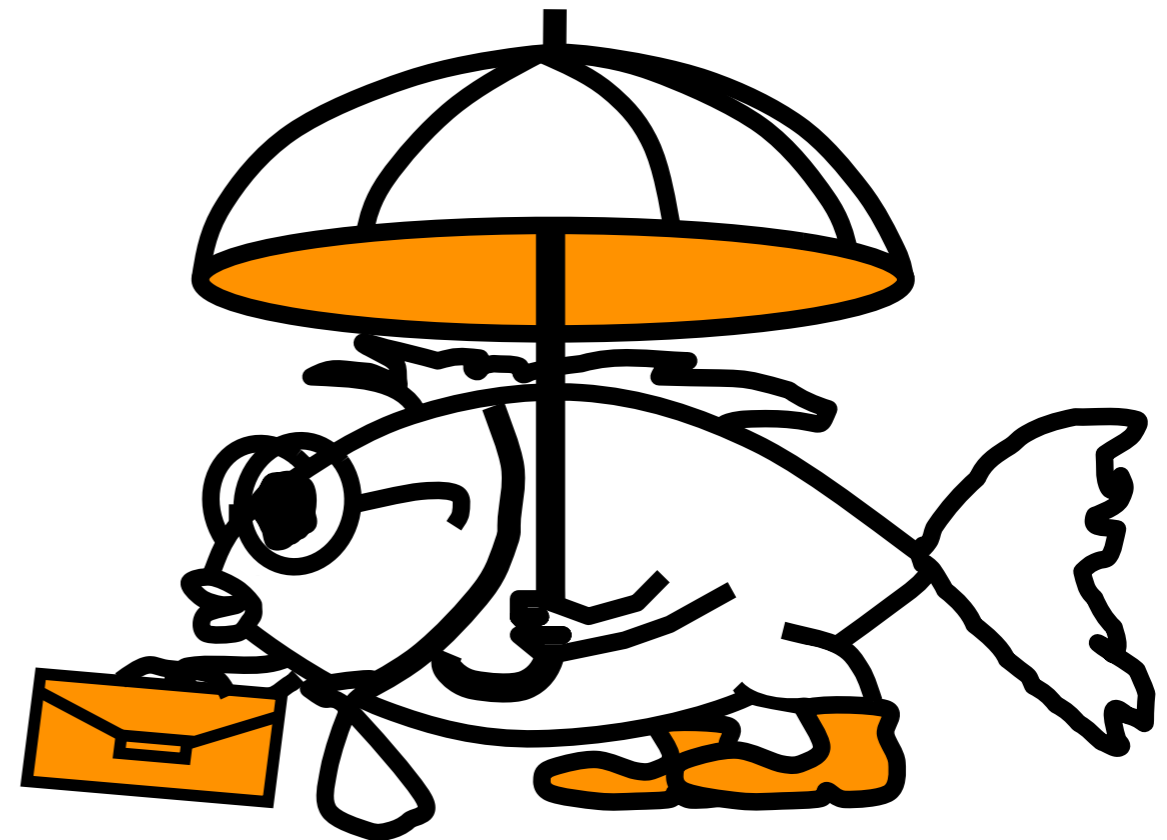
We can also evaluate the developed models using conventional statistical metrics

Model Properties for another good simple model

Model Expression	$0.70 - (9.74 \times 10^{-4}) \text{communicable} \cdot \text{Disease} \cdot \text{Death} \cdot \text{Rate} \sqrt{\text{health} \cdot \text{expenditure} \cdot \text{private} \cdot \text{total}} + 0.15 \text{total}^{1/3}$					
R Squared	0.759645					
Adjusted R-Squared	0.7568					
ANOVA Table		DF	SS	MS	F-Statistic	P-Value
	communicable·Disease·Death·Rate $\sqrt{\text{health} \cdot \text{expenditure} \cdot \text{private} \cdot \text{total}}$	1	6.52751	6.52751	469.303	1.21595×10^{-50}
	total ^{1/3}	1	0.901629	0.901629	64.8237	1.39604×10^{-13}
	Error	169	2.35061	0.013909		
	Total	171	9.77975			

Further Information

- ❖ www.evolved-analytics.com
- ❖ info@evolved-analytics.com
- ❖ Evolved Analytics LLC
3411 Valley Drive
Midland, MI 48640
+1-989-785-0513
- ❖ Evolved Analytics Europe BVBA
Antoine Coppenslaan, 27/11,
2300 Turnhout
Belgium
Phone: +32-148-00-414



Evolved Analytics
Evolving Daily. Learning Always.

Acknowledgments

- ❖ Evolved Analytics
 - ❖ Arthur Kordon
 - ❖ Katya Vladislavleva
 - ❖ Theresa Kotancheck
- ❖ Pronto Analytics
 - ❖ Ariel Sepúlveda