Data Models Questions Insights Understanding Actions

Breaking Down Corporate Barriers

Mark Kotanchek
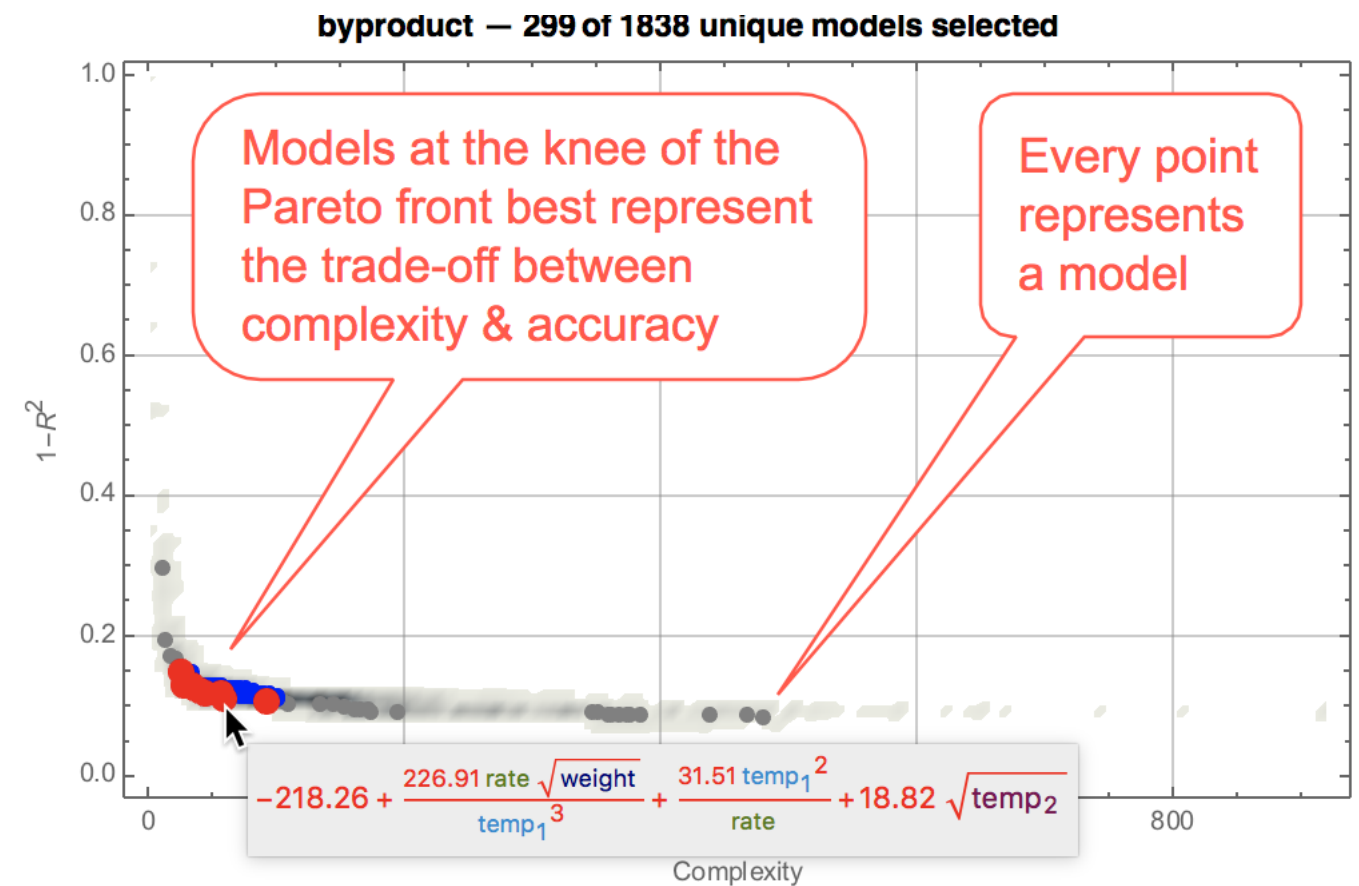Evolved Analytics LLC
mark@evolved-analytics.com

Background

- 2000: Autonomous Analysis Initiative

  - Needed to improve the efficiency & efficacy of data analysis

  - Focus→ Predictive Analytics

- 2005: Founded Evolved Analytics

  - DataModeler

  - Wide variety of application domains
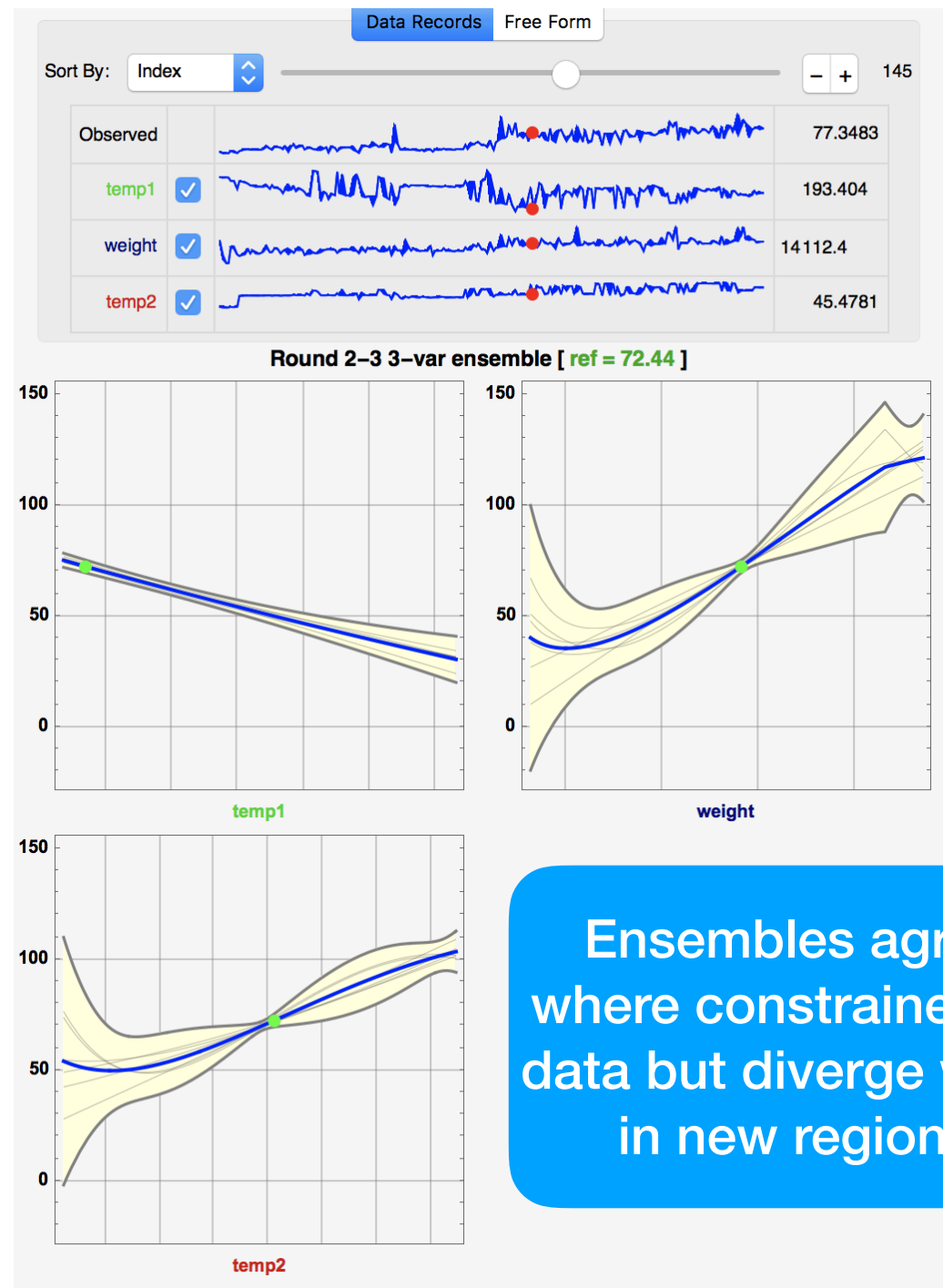
  - Clients around the world

Basic Concept

- Use evolution to develop predictive models from data

- Reward simplicity & accuracy

- Transparent human-interpretable models

- Many (thousands) of developed models are "good" (accurate & simple) and different

byproduct — 299 of 1838 unique models selected

Models at the knee of the Pareto front best represent the trade-off between complexity & accuracy

Every point represents a model

$$-218.26 + \frac{226.91\, rate\, \sqrt{weight}}{temp_1{}^3} + \frac{31.51\, temp_1{}^2}{rate} + 18.82\, \sqrt{temp_2}$$

Complexity

Automated hypothesis definer & refiner

Implications

Ensembles agree where constrained by data but diverge when in new regions

- Can handle ill-conditioned data arrays (e.g., fat or correlated variables)

- Multiple paths to quality models

- We can mine the diversity of "good enough" models for insight & guidance

- Ensembles of diverse good models $\implies$ Trustable Models

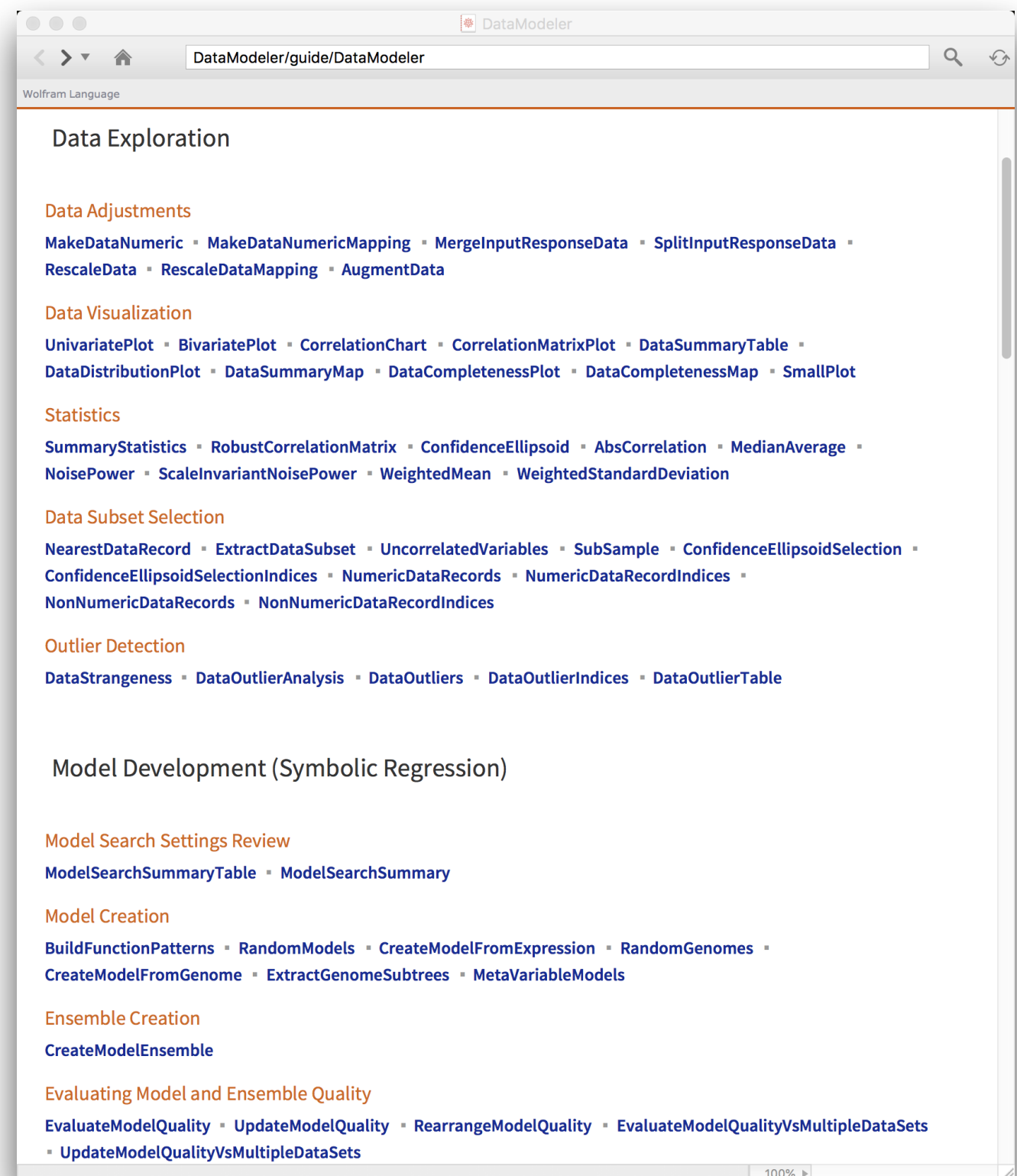- Trustable Models $\implies$ Active DOE
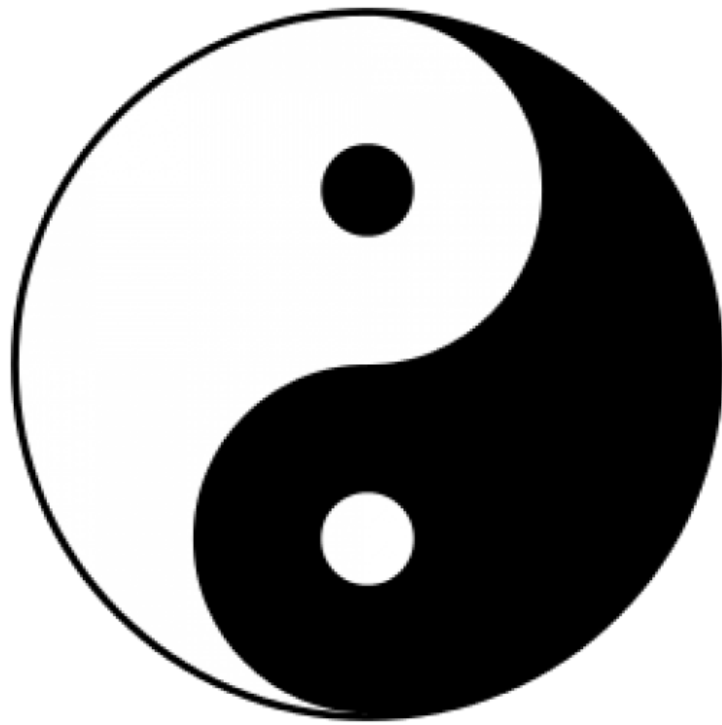
Philosophy

AI $\implies$ Augmented Intelligence

- Humans *have* to be in the loop

- Trust established by ensembles AND context

Automation is Good

- The machine should serve — not the converse

- Missing data adaption

- Automatic categorical conversion

- Outlier identification

In the Beginning was The Package

- A thing of beauty!

- ~460 Functions — spanning entire data analytics workflow

- Extensive & complete documentation

- Total control

- Custom workflows
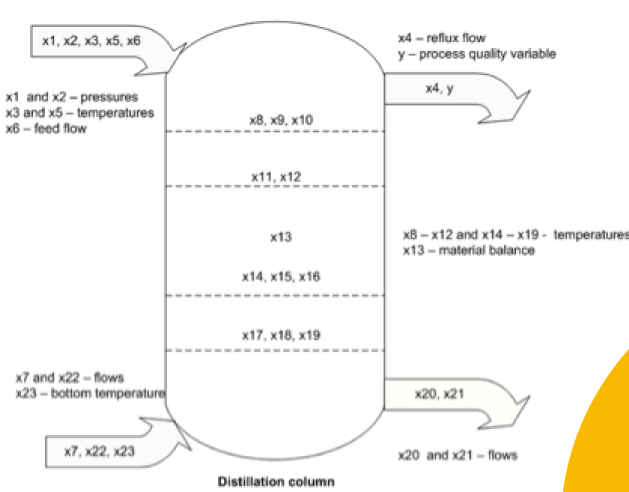
- Limited market acceptance

Package Issues

Notebooks are linear

Insight & understanding is nonlinear

- Users were coming from outside the Mathematica community

- Needed to train users in Mathematica as well as analytics

- Shift-enter is a remarkable barrier

… and then there was the GUI

- Ease of use

- Beautiful interface

- Supported best-practice workflow

- Robust (but pushes Dynamic limits)

- Layered complexity

- New clients had no Mathematica experience or interest
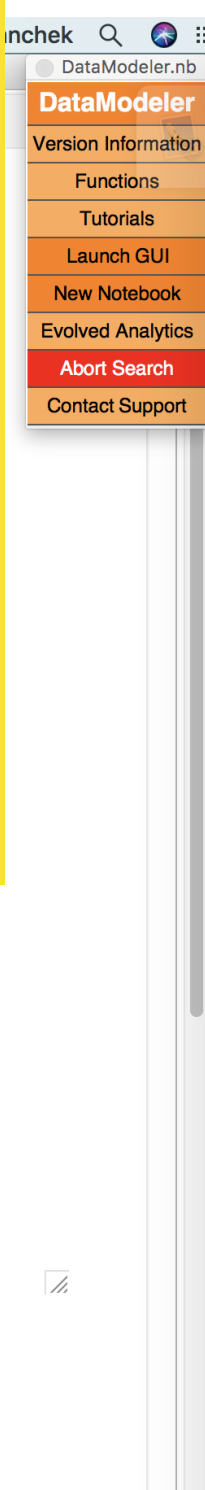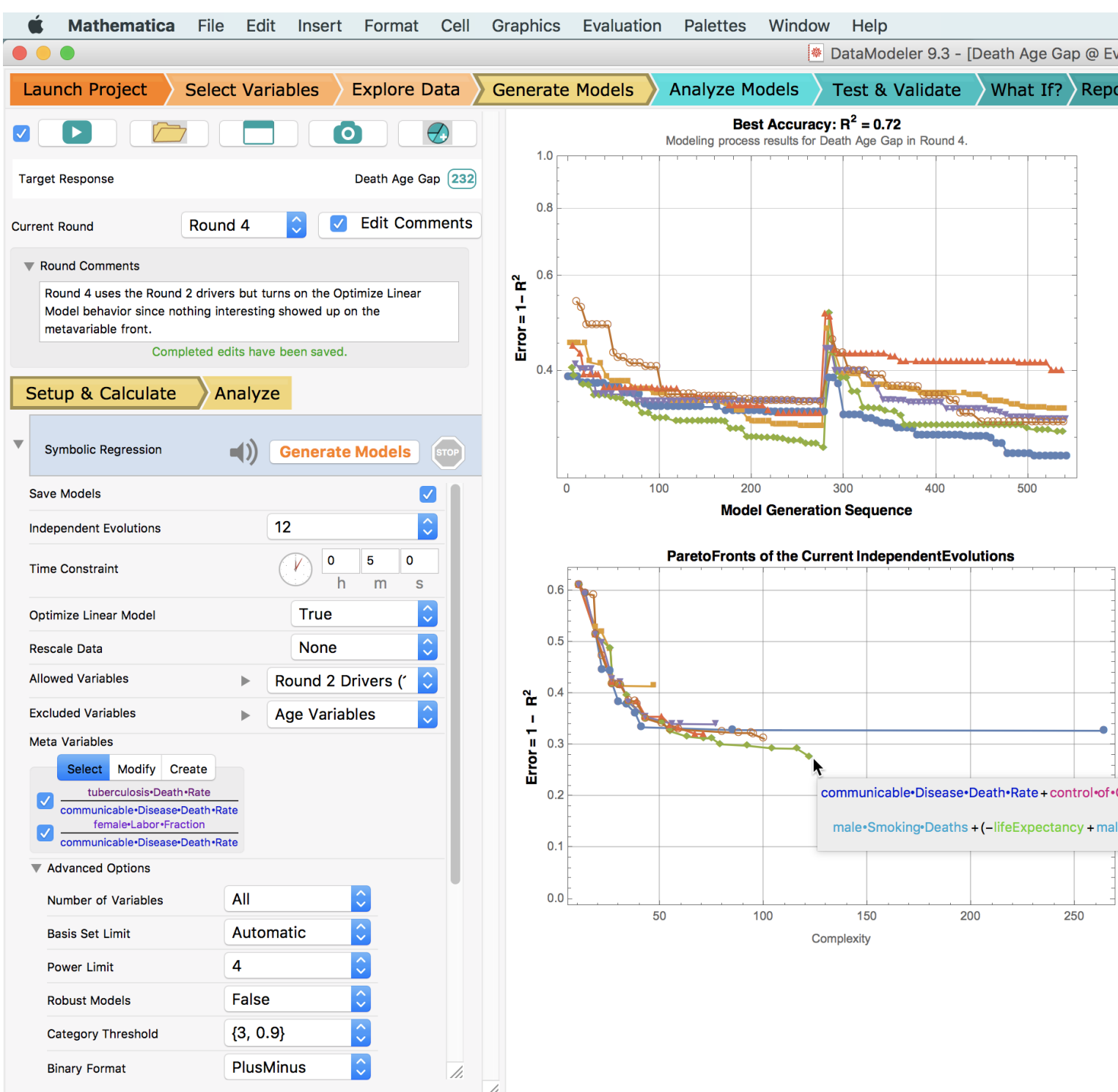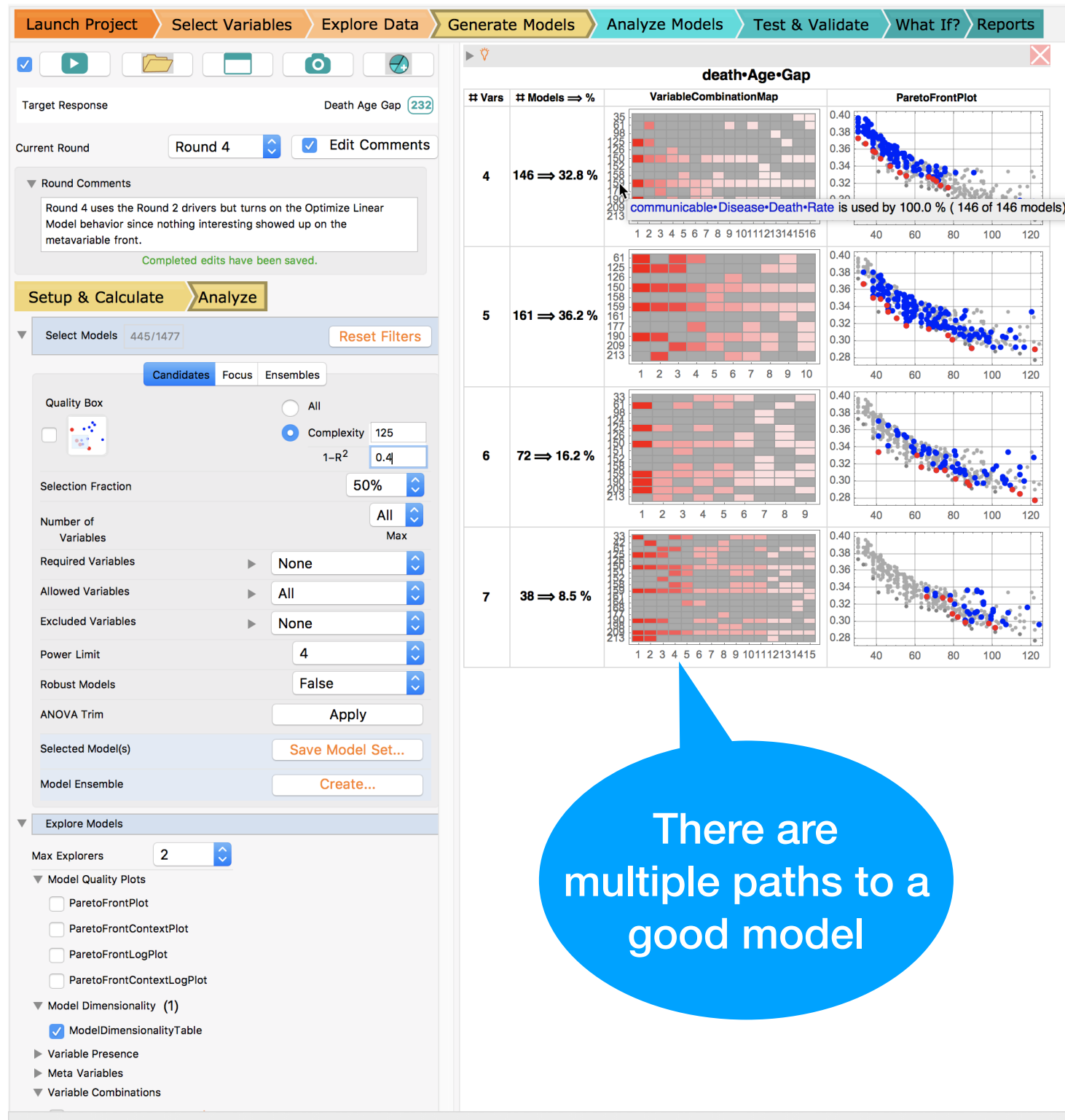
Package + GUI Issues

- Mathematica required

- WRI required to demo

- Concurrent licenses

  - Flexera the chosen vendor

  - Reprise is an alternative

  - Neither are stellar in terms of customer support

Enter the App

App versions are available for MacOS, Windows & Linux

- GUI-only (same as package version)

- Created with Enterprise Mathematica

- CDF Player embedded

- Code signing

  - Apple & DigiCert

- OS-specific installers — MacOS, Windows, Linux

- Behaves (mostly) like a conventional app

- People seem to care!

App Demo

Page 1 | Page 2

DataSummaryTable

Col	Label	Type	Uniformity	Class	Unique	Distribution Plot	Zero-Cross	Min	Mean	Median	Max
1	cap-shape	ABC	100%		6			bell	convex	convex	sunken
2	cap-surface	ABC	100%		4			fibrous	scaly	scaly	smooth
3	cap-color	ABC	100%		10	fibrous		brown	brown	brown	yellow
4	bruises?	ABC	100%		2			False	False	False	True
5	odor	ABC	100%		9			almond	none	none	spicy
6	gill-attachment	ABC	100%		2			attached	free	free	free
7	gill-spacing	ABC	100%		2			close	close	close	crowded
8	gill-size	ABC	100%		2			broad	broad	broad	narrow
9	gill-color	ABC	100%		12			black	buff	buff	yellow
10	stalk-shape	ABC	100%		2			enlarging	tapering	tapering	tapering
11	stalk-root	ABC	100%		5			bulbous	bulbous	bulbous	rooted
12	stalk-surface-above-ring	ABC	100%		4			fibrous	smooth	smooth	smooth
13	stalk-surface-below-ring	ABC	100%		4			fibrous	smooth	smooth	smooth
14	stalk-color-above-ring	ABC	100%		9			brown	white	white	yellow
15	stalk-color-below-ring	ABC	100%		9			brown	white	white	yellow
16	veil-type	ABC	100%		1		•	partial	partial	partial	partial
17	veil-color	ABC	100%		4			brown	white	white	yellow
18	ring-number	A									
19	ring-type	A									
20	spore-print-color	A									

UnivariatePlot

cap-color | bruises? | odor | gill-spacing

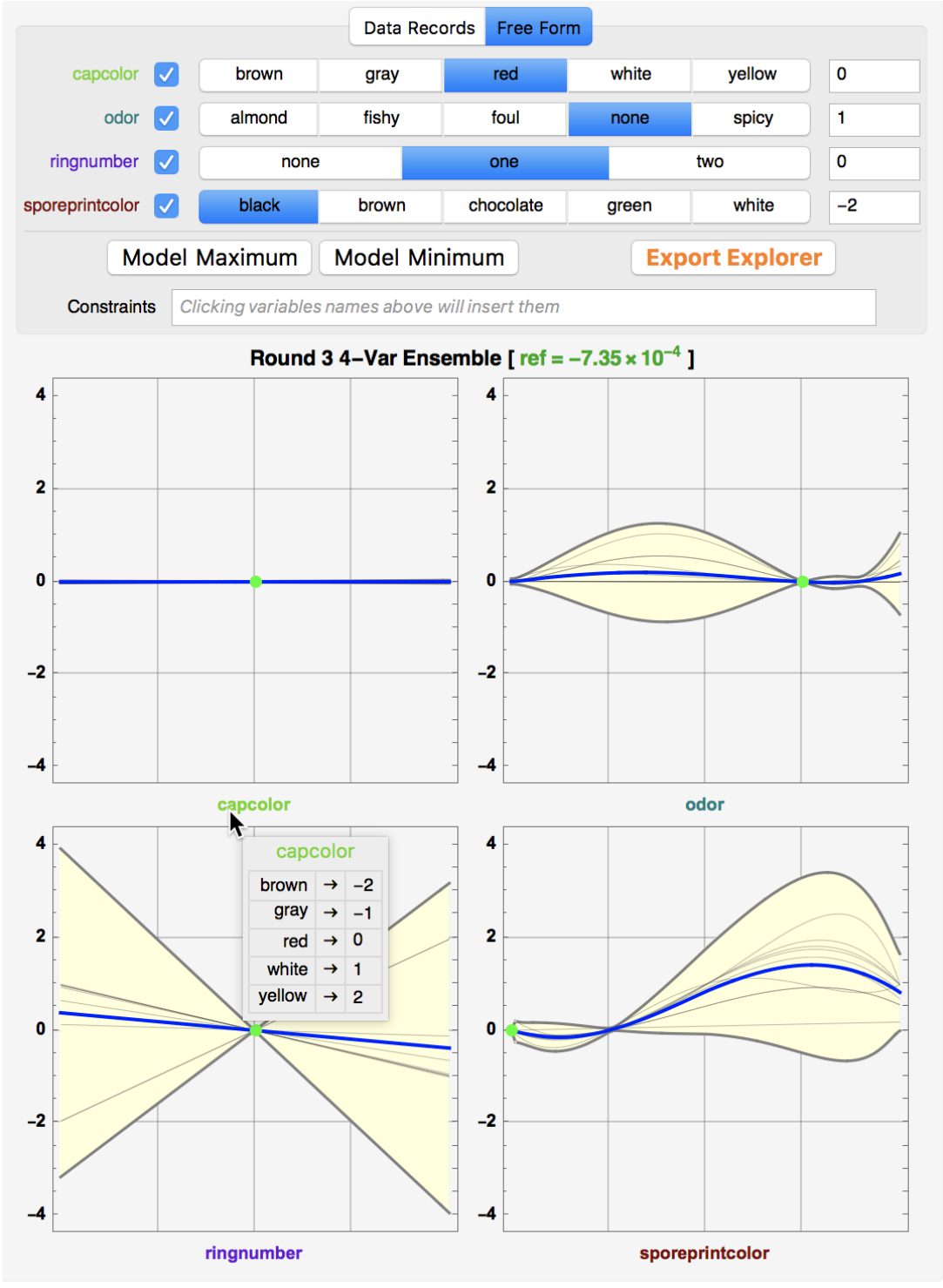gill-size | talk-surface-above-rin | stalk-color-above-ring | ring-number

ring-type | spore-print-color | population | habitat

ushroom (either edible o

spore-print-color

black	→ −2
brown	→ −1
chocolate	→ 0
green	→ 1
white	→ 2

	Data Records	Free Form

capcolor ☑	brown	gray	**red**	white	yellow	0
odor ☑	almond	fishy	foul	**none**	spicy	1
ringnumber ☑	none	**one**		two		0
sporeprintcolor ☑	**black**	brown	chocolate	green	white	−2

Model Maximum | Model Minimum | **Export Explorer**

Constraints | *Clicking variables names above will insert them*

Round 3 4-Var Ensemble [ref = −7.35 × 10⁻⁴]

capcolor | odor

capcolor

brown	→	−2
gray	→	−1
red	→	0
white	→	1
yellow	→	2

ringnumber | sporeprintcolor

Remaining Barriers

- "Most Effective Technology"

  - Think Different is a barrier

  - Corporate Software Certification (IT Restrictions)

- Excel xlsx Import

  - Mathematica implementation is poor

- Dynamic limitations

  - There are dragons here

  - Imposed limits on data sizes

  - Subkernels

- CDF issues

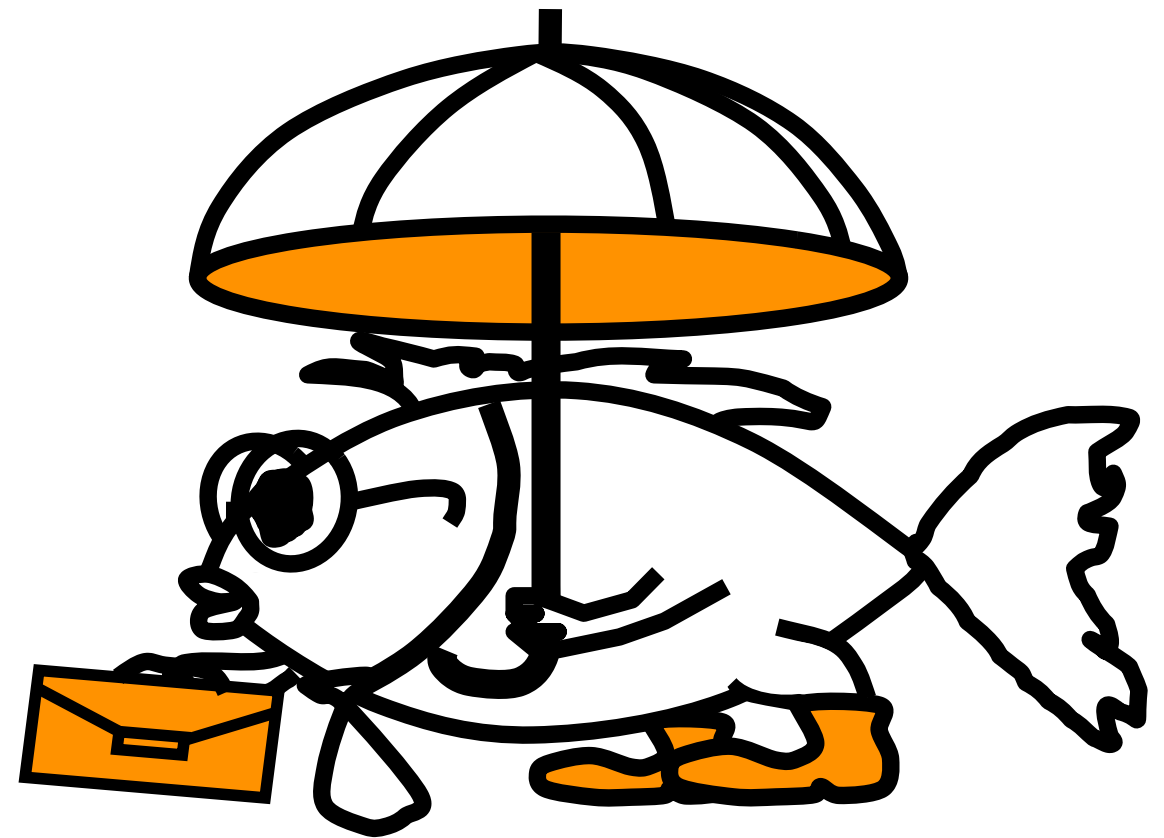 - Audit trail / Cut & Paste

Acknowledgements

- **GUI**: Ariel Sepúlveda (ariel.sepulveda@prontoanalytics.com)

- **App Conversion**: Aureliano Giraud

- **CDF Build**: Andre Kuzniarek & Wolfram Team

- **Interface Design**: Theresa Kotanchek & Katya Vladislavleva

Questions?

- www.evolved-analytics.com

- mark@evolved-analytics.com

Evolved Analytics

Evolving Daily. Learning Always.