

TaylorFit

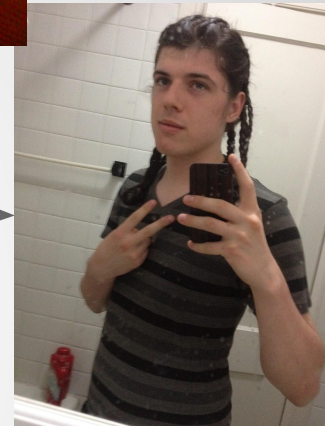
Not your usual fit

The Team

ADAM



CHRIS



BDAN

PROFESSOR VACCARI



TIM



What does Taylorfit do?

- Regression program that makes fitting MPR models easy.
 - Multivariable polynomial regression (MPR)
- Produces a polynomial describing the relationship between data
- Helps to predict non-obvious relationships
- Math-magic

Where Can it be Used?

- Demand forecasting
- Business analytics
- Market research
- Data mining

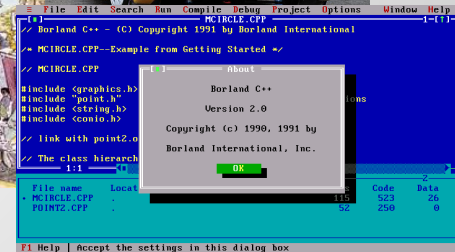
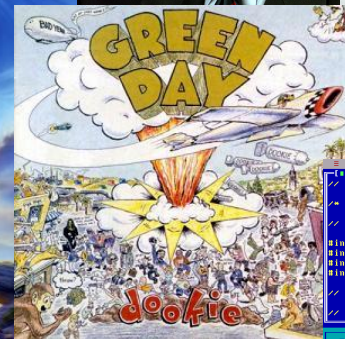
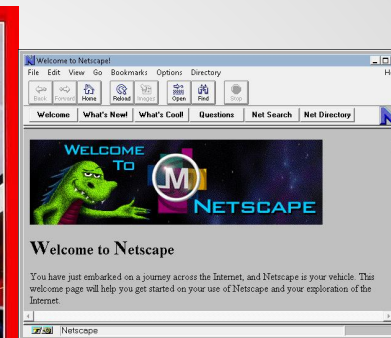
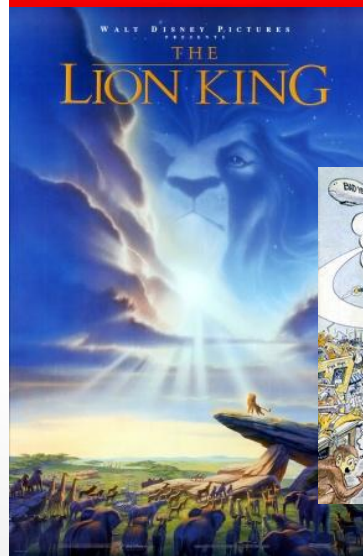
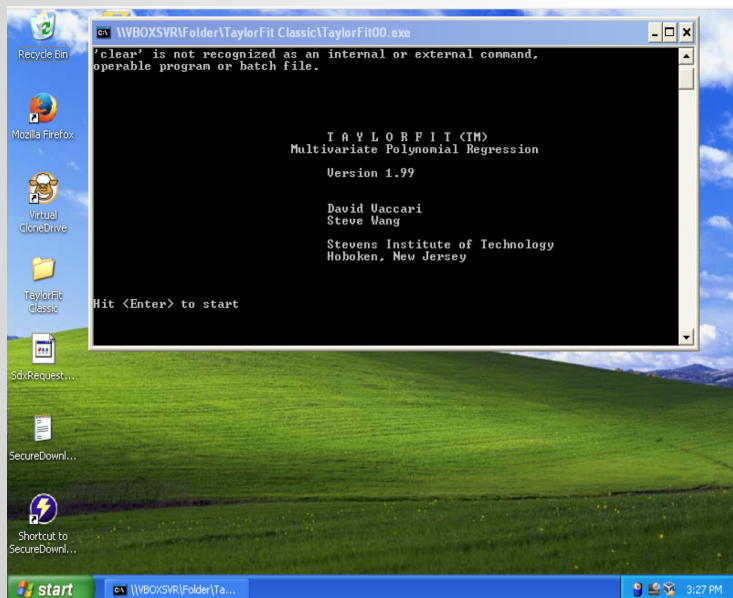
So, what did we do?

- Started off as a DOS program.
 - Would not even compile in modern GCC.
- Modernized
- Extended -- Added functionality
 - Graphs
 - New math
- Ended as a modern QT5 GUI
- Documented EVERYTHING

Before

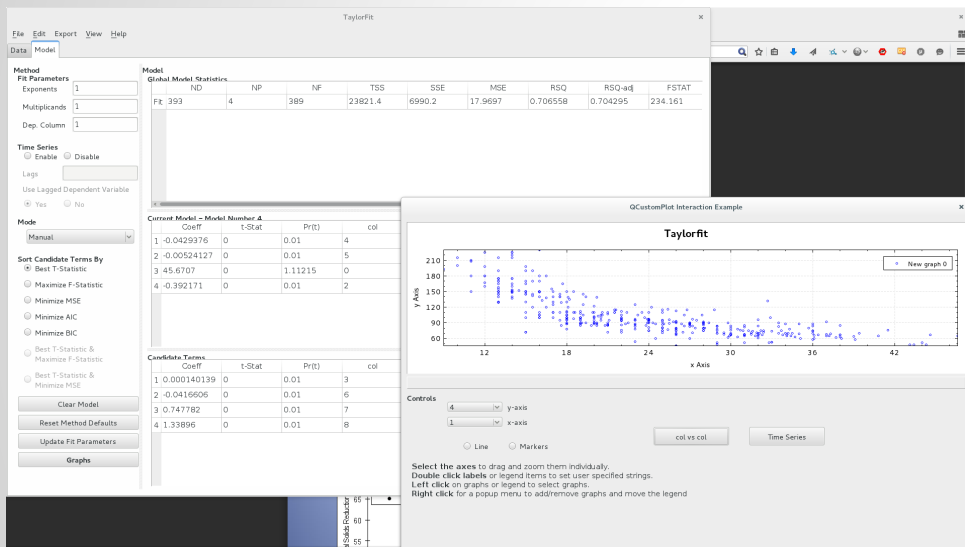
AKA - Things that were still new in 1994

TaylorFit - 1994

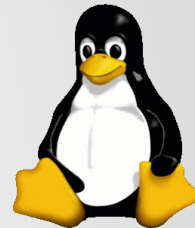
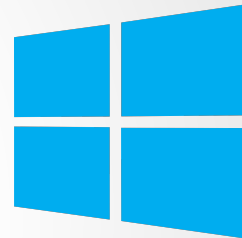


After

TaylorFit 3.0 - 2015



Cross Platform Compatibility

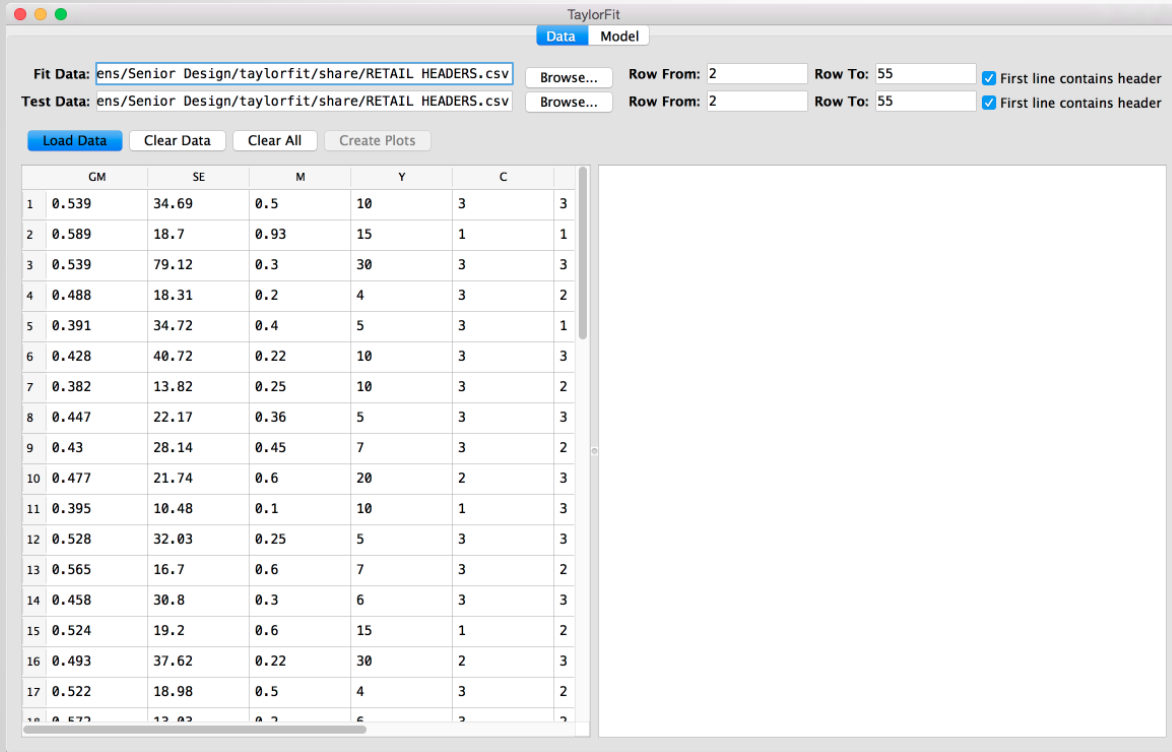


Single Codebase



QCustormPlot

Demonstration



We select RETAIL HEADERS.csv as our fitting data set. The program populates the testing data with the same file if another file is not provided. The program also automatically detects if there is a string header in the file and loads it into the chart. The user can select more specific data ranges if they wish

TaylorFit

Data
Model

Method

Fit Parameters

Exponents

Multiplicands

Dep. Column

Time Series

☐ Enable ☒ Disable

Lags

Use Lagged Dependent Variable

☒ Yes ☐ No

Mode

Manual

Sort Candidate Terms By

☒ Best T-Statistic

☐ Maximize F-Statistic

☐ Minimize MSE

☐ Minimize AIC

☐ Minimize BIC

☐ Best T-Statistic & Maximize F-Statistic

☐ Best T-Statistic & Minimize MSE

Clear Model

Reset Method Defaults

Compute Fit

Create Plots

Model

Global Model Statistics

Current Model

Coeff	t-Stat	Pr(t)	col	lag	exp	SSE-add

Candidate Terms

	Coeff	t-Stat	Pr(t)	col	lag	exp	SSE-add
1	0.813782	65.5947	0.397595	7	0	1	0.161446
2	0.492167	60.7776	0.389439	0	0	0	0.187675
3	0.177656	27.0543	0.251172	6	0	1	0.895874
4	0.183636	23.8833	0.218559	5	0	1	1.12799
5	2.3217	14.7029	0.0666365	8	0	1	2.61245

When we switch to the model tab we see several options to set up the fit. When we compute the fit the table of candidate terms appears sorted by the specified criteria.

TaylorFit

Data
Model

Method

Fit Parameters

Exponents

Multiplicands

Dep. Column

Time Series

☐ Enable ☒ Disable

Lags

Use Lagged Dependent Variable

☒ Yes ☐ No

Mode

Sort Candidate Terms By

☒ Best T-Statistic

☐ Maximize F-Statistic

☐ Minimize MSE

☐ Minimize AIC

☐ Minimize BIC

☐ Best T-Statistic & Maximize F-Statistic

☐ Best T-Statistic & Minimize MSE

Compute Fit

Model

Global Model Statistics

	ND	NP	NF	TSS	SSE	MSE	RSQ	RSQ-adj	FSTAT	AIC	BIC
Fit	54	3	51	0.187676	0.0692744	0.00135832	0.630882	0.616407	29.0558	-2.75589	-2.77075

Current Model – Model Number 3

	Coeff	t-Stat	Prt()	col	lag	exp	SSE-add
1	0.458239	8.08118	0.0001	7	0	1	0.161446
2	0.181038	5.11551	0.0001	0	0	0	0.103211
3	0.129086	4.9984	0.0001	3	0	1	0.0692744

Candidate Terms

	Coeff	t-Stat	Prt()	col	lag	exp	SSE-add
1	0.00101302	1.60692	0.0001	4	0	1	0.0658725
2	0.0105838	1.50316	0.0001	5	0	1	0.0662792
3	-0.00946219	-0.968834	0.0001	6	0	1	0.0679978
4	0.040655	0.665021	0.0001	8	0	1	0.068667
5	-0.000119937	-0.406882	0.0001	2	0	1	0.0690457

Double clicking terms adds them to our model. Here when we add all terms with a T Statistic greater than 2. This is a linear model. Note that our R Squared is .63. Let's see if we can do better.

TaylorFit
Data Model

Method
Fit Parameters
Exponents 1 2 -1
Multiplicands 1 2
Dep. Column 1

Time Series
☐ Enable ☒ Disable
Lags
Use Lagged Dependent Variable
☒ Yes ☐ No

Mode
Manual

Sort Candidate Terms By
☒ Best T-Statistic
☐ Maximize F-Statistic
☐ Minimize MSE
☐ Minimize AIC
☐ Minimize BIC
☐ Best T-Statistic & Maximize F-Statistic
☐ Best T-Statistic & Minimize MSE

Clear Model
Reset Method Defaults
Compute Fit
Create Plots

Model
Global Model Statistics

	ND	NP	NF	TSS	SSE	MSE	RSQ	RSQ-adj	FSTAT	AIC	BIC
Fit	54	3	51	0.187676	0.0692744	0.00135832	0.630882	0.616487	29.8558	-2.75589	-2.77875

Current Model – Model Number 3

	Coeff	t-Stat	P(t)	col	lag	exp	col	lag	exp	SSE-add
1	0.458239	8.08118	0.0001	7	0	1	0	0	0	0.15798
2	0.181038	5.11551	0.0001	0	0	0	0	0	0	0.104819
3	0.129086	4.9984	0.0001	3	0	1	0	0	0	0.103211

Candidate Terms

	Coeff	t-Stat	P(t)	col	lag	exp	col	lag	exp	SSE-add
1	-0.0268353	-4.39282	0.0001	3	0	-1	4	0	-1	0.0499887
2	-0.0287442	-3.34551	0.0001	4	0	-1	6	0	2	0.0566837
3	-0.0082369	-3.01003	0.0001	3	0	-1	8	0	-1	0.0586471
4	-0.291368	-2.93729	0.0001	4	0	-1	5	0	-1	0.0590799
5	-0.00807487	-2.71882	0.0001	4	0	-1	8	0	-1	0.0603976
6	0.006245802	2.68982	0.0001	4	0	1	5	0	2	0.0605173
7	-0.008595822	-2.6062	0.0001	3	0	-1	6	0	2	0.0609892
8	-0.00825784	-2.58416	0.0001	6	0	2	8	0	-1	0.0611124
9	-0.0411696	-2.57452	0.0001	4	0	-1	6	0	1	0.061186

We change our exponents to 1 2 and -1 and our multiplicands to 1 and 2. This results in a polynomial model with interactions including ratio terms.

TaylorFit
Data Model

Method
Fit Parameters
Exponents 1 2 -1
Multiplicands 1 2
Dep. Column 1

Time Series
☐ Enable ☒ Disable
Lags
Use Lagged Dependent Variable
☒ Yes ☐ No

Mode
Manual

Sort Candidate Terms By
☒ Best T-Statistic
☐ Maximize F-Statistic
☐ Minimize MSE
☐ Minimize AIC
☐ Minimize BIC
☐ Best T-Statistic & Maximize F-Statistic
☐ Best T-Statistic & Minimize MSE

Clear Model
Reset Method Defaults
Compute Fit
Create Plots

Model
Global Model Statistics

	ND	NP	NF	TSS	SSE	MSE	RSQ	RSQ-adj	FSTAT	AIC	BIC
Fit	54	6	48	0.187676	0.0331236	0.000690074	0.823506	0.805121	37.3274	-2.93888	-2.96862

Current Model - Model Number 14

	Coeff	t-Stat	P(t)	col	lag	exp	col	lag	exp	SSE-add
1	0.522181	12.3324	0.0109699	7	0	1	0	0	0	0.13643
2	0.148905	5.73744	0.0001	0	0	0	0	0	0	0.0500734
3	0.00567077	2.90645	0.0001	5	0	2	7	0	1	0.0378119
4	-0.0639988	-7.89084	0.0001	3	0	-1	4	0	-1	0.0547769
5	1.42685	5.32289	0.0001	2	0	-1	6	0	-1	0.0442748
6	0.0122863	3.30206	0.0001	4	0	-1	8	0	-1	0.0367418

Candidate Terms

	Coeff	t-Stat	P(t)	col	lag	exp	col	lag	exp	SSE-add
1	-0.0010694	-1.74005	0.0001	7	0	-1	8	0	-1	0.0311189
2	-4.46965e-05	-1.62039	0.0001	2	0	1	8	0	-1	0.031371
3	-0.00159375	-1.56845	0.0001	8	0	-1	0	0	0	0.0314761
4	0.145419	1.4688	0.0001	3	0	2	8	0	1	0.0316699
5	-0.00229787	-1.40935	0.0001	7	0	1	8	0	-1	0.0317805
6	0.00664776	1.40563	0.0001	3	0	2	6	0	2	0.0317873
7	0.0181224	1.39915	0.0001	3	0	2	6	0	1	0.0317991
8	0.385807	1.3891	0.0001	3	0	2	8	0	2	0.0318173
9	-0.000582214	-1.28591	0.0001	6	0	1	8	0	-1	0.0319978

Next we add all terms with t-stat > 2.0 AND remove terms with t-stat < 2.0. We do this iteratively until there are no terms t>2 in the Candidate Terms table and none with t<2 in the Current Model table.

TaylorFit
Data Model

Method
Fit Parameters
Exponents 1 2 -1
Multiplicands 1 2
Dep. Column 1

Time Series
☐ Enable ☒ Disable
Lags
Use Lagged Dependent Variable
☒ Yes ☐ No

Mode
Manual

Sort Candidate Terms By
☒ Best T-Statistic
☐ Maximize F-Statistic
☐ Minimize MSE
☐ Minimize AIC
☐ Minimize BIC
☐ Best T-Statistic & Maximize F-Statistic
☐ Best T-Statistic & Minimize MSE

Clear Model
Reset Method Defaults
Compute Fit
Create Plots

Model
Global Model Statistics

	ND	NP	NF	TSS	SSE	MSE	RSQ	RSQ-adj	FSTAT	AIC	BIC
Fit	54	6	48	0.187676	0.0331236	0.000690074	0.823506	0.805121	37.3274	-2.93888	-2.96862

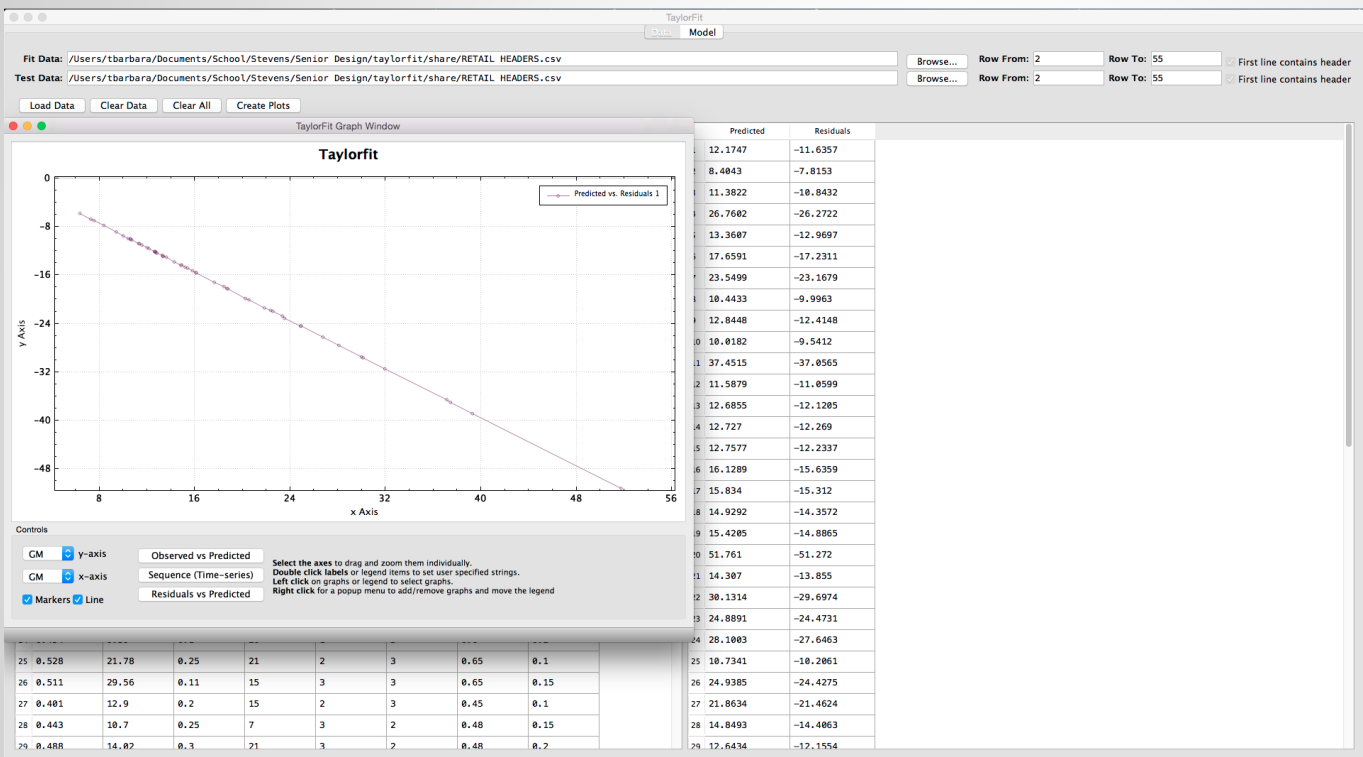
Current Model -- Model Number 14

	Coeff	t-Stat	P(t)	col	lag	exp	col	lag	exp	SSE-add
1	0.522181	12.3324	0.0109699	7	0	1	0	0	0	0.13643
2	0.148985	5.73744	0.0001	0	0	0	0	0	0	0.0500734
3	0.00567077	2.90645	0.0001	5	0	2	7	0	1	0.0378119
4	-0.0639988	-7.89084	0.0001	3	0	-1	4	0	-1	0.0547769
5	1.42685	5.32289	0.0001	2	0	-1	6	0	-1	0.0442748
6	0.0122863	3.30206	0.0001	4	0	-1	8	0	-1	0.0367418

Candidate Terms

	Coeff	t-Stat	P(t)	col	lag	exp	col	lag	exp	SSE-add
1	-0.0010694	-1.74005	0.0001	7	0	-1	8	0	-1	0.0311189
2	-4.46965e-05	-1.62039	0.0001	2	0	1	8	0	-1	0.031371
3	-0.00159375	-1.56845	0.0001	8	0	-1	0	0	0	0.0314761
4	0.145419	1.4688	0.0001	3	0	2	8	0	1	0.0316699
5	-0.00229787	-1.40935	0.0001	7	0	1	8	0	-1	0.0317805
6	0.00664776	1.40563	0.0001	3	0	2	6	0	2	0.0317873
7	0.0181224	1.39915	0.0001	3	0	2	6	0	1	0.0317991
8	0.385807	1.3891	0.0001	3	0	2	8	0	2	0.0318173
9	-0.000582214	-1.28591	0.0001	6	0	1	8	0	-1	0.0319978

Note that R Squared is now .82. Since the amount of error is proportional to $1 - \text{RSQ}$, this means adding polynomial interaction terms decreased the relative error from 0.37 to 0.18, cutting the error by more than half.



TaylorFit can also be used to create graphs of models. The user may choose any of the columns to plot on the x or y axis and may rename and delete graphs. Graphs can be exported as images and pdfs.

Who are the competitors?

- R
- STATA
- SAS
- SPSS
- General purpose
 - Not just modeling
- Difficult to learn
- Limited to linear modeling and other constrained approaches

How does it compare to others?

- Others designed for Design of Experiment.
- TaylorFit allows fitting of Nth degree polynomials
- TaylorFit selection of terms using stepwise regression.
- Has a pleasant GUI.

Why No Artificial Neural Network?

- Good for accurate fits
 - Prone to overfit
- Can describe complex relationships
 - But the structure must be determined a priori

Where Did We Stumble?

- Engine is confusing and has lots of unneeded abstractions.
- Difficulty supporting all major OSs.
- Coordinating code was tough at first but git made that easy.

TaylorFit 4.0?

- Time Series
- Probability calculations
- History capability
- Sensitivity analysis capability
- Swap out core engine
 - <http://www.astro.umd.edu/~bjw/software/boycottnr.html>
- Update save format
 - (it's really temperamental right now)

Acknowledgements (1/2) (in no order)

- David Adam Perez
- Brandon Schur
- Chris Barry
- Tim Barbra
- Professor Vaccari
- Professor Burlick
- StackOverflow
- Neal
- Mamouns
- That corner store next to Mamouns
- Qt Developers
- Nick Bevacqua
- Xebra Corporation
- Adam Cotenoff
- Mark Fleres
- Marc Kaminsky
- Victoria Albright
- Bier Garten
- Qdoba
- Perl (the human)
- John Coltrane
- Professor Gabarró
- Shamwow Guy
- Erich Polixa
- Radhika Kasabwala
- ???
- Randall Suliga
- Dan Coullias
- Jess Berg
- Christian Benitez
- The fairy wand
- Blase Feeney
- Matt Coccia
- Fred Florio
- GCC / Clang teams
- Kevin

Acknowledgements (2/2) (cont.)

- Bacon Pancakes
- Trello
- Slack
- Snoop Dogg
- Jake the Dog and Finn the Human
- The dog I once played with while walking home
- Chip (dog)
- Sonny and Wilson (dogs)
- Off Center
- Neal
- Neil deGrasse Tyson
- Neil Patrick Harris
- Neil Diamond
- Neil Young
- Neil Armstrong
- Neal again
- Thor
- Tom Kenny
- Kenny McCormick (RIP)
- Kenny Loggins
- The Danger Zone and the brave souls who built that highway
- Jake from State Farm
- The Mr. Game! team
- The Game Grumps
- Roosterteeth
- The T.V. Series, “Too Many Cooks”
- Richard Stallman
- Richard Stallman’s Beard
- Steve Jobs
- The City of Hoboken
- Professor Duchamp
- Bill Nye the Science Guy
- Bill Gates (post Microsoft)
- Mom
- Professor Klapholz

EOF