

Actionable Data: the Wolfram Approach

Matthew Day
matthewday@wolfram.com

Manager, Data Repository
Wolfram Partnerships
Wolfram Research

Introduction to Wolfram Research

- ***Mathematica*** is a computation engine.
- **Wolfram|Alpha** is a knowledge engine. Natural language queries + data + computation.
- **CDF** is a technology for creating and sharing user-friendly, interactive documents.

If a question can be answered on the basis of human knowledge, a computer should be able to provide that answer.

Wolfram | Alpha

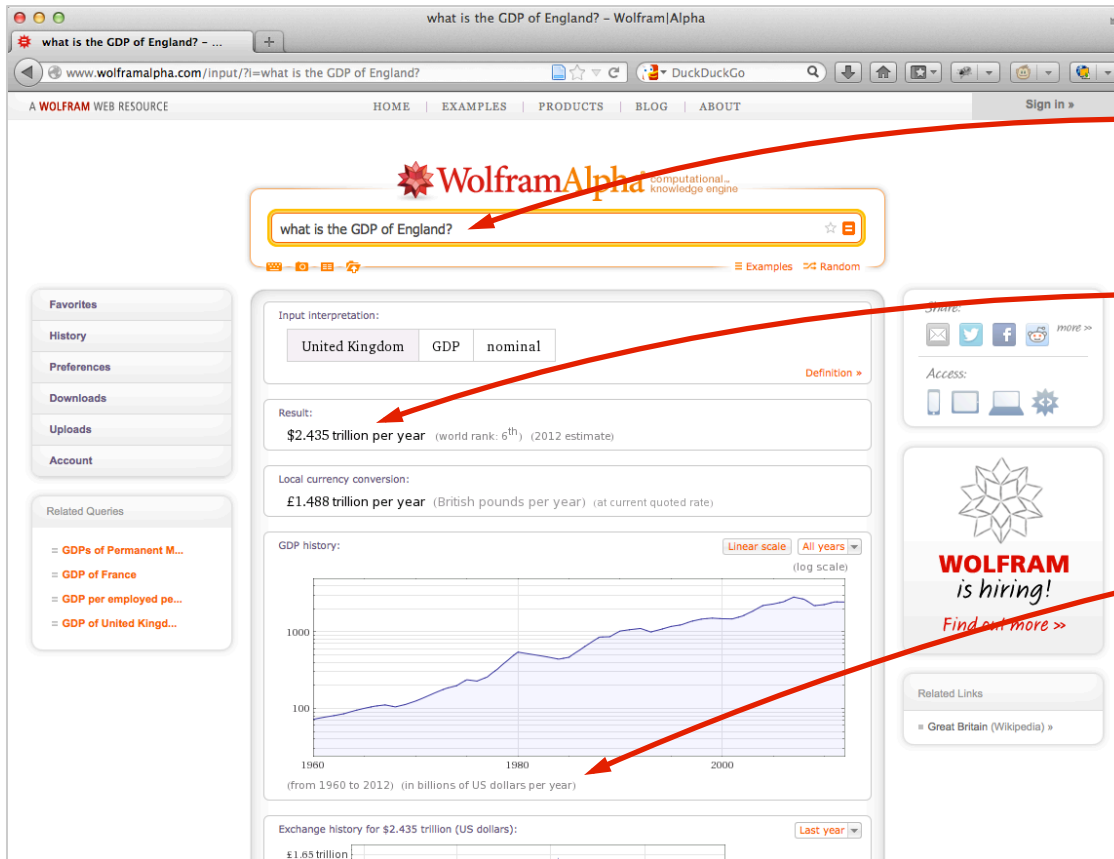
A knowledge engine for working with:

- 1. Curated facts.**
- 2. Models.**
- 3. Random pieces of data.**

Simple facts

What is the GDP of England?

wolframalpha.com/input/?i=what is the GDP of England?

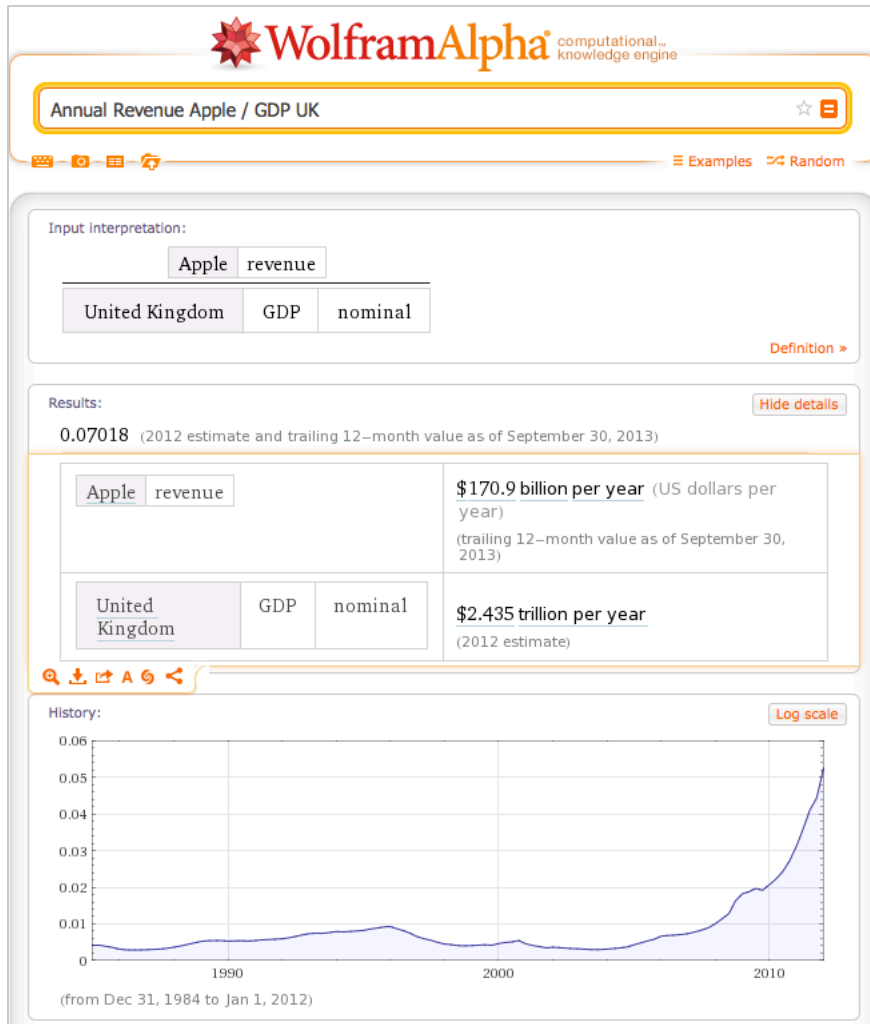


natural language question

the answer

an automated report giving context to the answer

Computed facts



Revenue of Apple / GDP of the UK?

To answer this, Wolfram|Alpha needs to know about currencies and exchange rates.

Everything is computed on the fly

The screenshot shows the WolframAlpha PRO interface. At the top, the WolframAlpha logo is displayed. Below it, a search bar contains the query "number of goats in Spain / distance to the sun". The interface is divided into several sections:


- Input Interpretation:** This section shows the query broken down into "Spain", "livestock population", "goats", "Sun", and "distance from Earth".
- Result:** The result is 2.3×10^6 au (reciprocal astronomical units) (2009 estimate). A "Show details" button is visible.
- History:** A line graph showing the number of goats in Spain from 1961 to 2009. The y-axis represents the number of goats in millions of reciprocal astronomical units, ranging from 2.5 to 4. The x-axis represents years from 1970 to 2000. The graph shows a fluctuating trend with a notable peak around 1990.
- Unit conversions:** A section with a "More" button listing various unit conversions: $15.39 \mu\text{m}^2/\text{mL}$ (square micrometers per milliliter), $0.2043 \text{ ft}/\text{ac}$ (feet per acre), $3.87 \times 10^{-5} \text{ mi}/\text{ac}$ (miles per acre), $0.02477 \text{ mi}/\text{mi}^2$ (miles per square mile), and $0.0681 \text{ yd}/\text{ac}$ (yards per acre).

At the bottom, it states "Computed by Wolfram Mathematica" and provides links for "Sources" and "Download page".

Number of goats in Spain /
distance to the sun.

We didn't foresee this
question, but it is
answerable.

Models of the world

 **WolframAlpha** | PRO

Position of the sun at 3pm next Thursday in Blackpool? ☆

☰ Examples ↗ Random

Input interpretation:

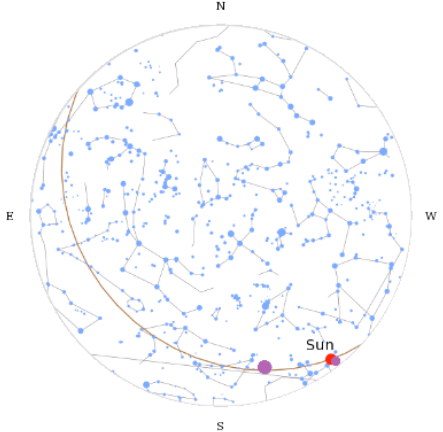
Sun	location	next Thursday at 3:00 pm
	location	Blackpool, Blackpool

Result: [Show decimal](#)

right ascension	$18^{\text{h}} 21^{\text{m}} 48^{\text{s}}$
declination	$-23^{\circ} 20' 35''$

[Units >](#)

Sky position from Blackpool at 3:00 pm December 26: [Show mesh](#) [Zoom](#) [Show decimal](#)



altitude 5° (—)

azimuth $217^{\circ} 30'$ (SW)

next rise 8:31 am GMT/BST | Friday, December 27, 2013

next set 3:57 pm GMT/BST | Thursday, December 26, 2013

constellation Sagittarius

Local sidereal time:
 $21^{\text{h}} 7^{\text{m}} 12.28^{\text{s}}$

Computed by [Wolfram Mathematica](#) [Sources](#) [Download page](#)

“Position of the sun” ?
“3pm next Thursday” ?
“in Blackpool” ?

Wolfram | Alpha needs to know quite a lot about the world to compute this answer.

A biomedical model

WolframAlpha PRO

heart risk for male 43 years old not diabetic

- LDL cholesterol: 111 mg/dL
- HDL cholesterol: 54 mg/dL
- systolic blood pressure: 120 mmHg
- diastolic blood pressure: 80 mmHg
- smoker: no

Input information:

heart disease	
gender	male
diabetic	no
age	43 years
LDL cholesterol	111 mg/dL (milligrams per deciliter)
HDL cholesterol	54 mg/dL (milligrams per deciliter)
systolic blood pressure	120 mmHg (millimeters of mercury)
diastolic blood pressure	80 mmHg (millimeters of mercury)
smoker	no

10-year risk of developing coronary heart disease:
≈ 3.9% (1 in 26)
(based on the Framingham Heart Study, predominantly Caucasian US population)

Impact of cholesterol and blood pressure on risk: Disable interpolation

LDL cholesterol (mg/dL)

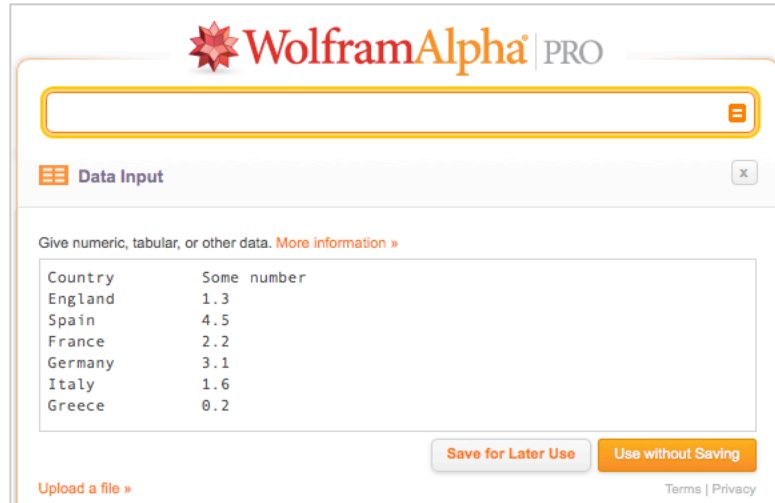
HDL cholesterol (mg/dL)

Wolfram|Alpha has been taught a model of heart disease risk in humans (from the Framingham Heart Study).

I gave some of the required information in the question ...

and editable assumptions have been provided for the other variables.

Random data



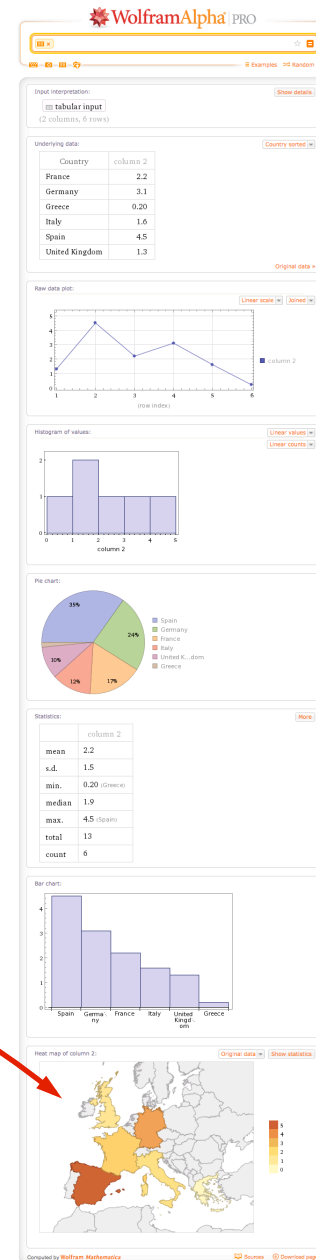
The screenshot shows the WolframAlpha PRO interface for data input. At the top, the WolframAlpha logo and 'PRO' are visible. Below is a search bar. The 'Data Input' section is active, showing a table with the following data:

Country	Some number
England	1.3
Spain	4.5
France	2.2
Germany	3.1
Italy	1.6
Greece	0.2

Buttons for 'Save for Later Use' and 'Use without Saving' are present, along with an 'Upload a file' link and 'Terms | Privacy' text.

I gave Wolfram|Alpha some random data ...

... and I got back something useful, including a nice coloured map that I could use in a publication.



The screenshot shows the results of the data input in WolframAlpha PRO. The interface includes several visualizations and statistical summaries:

- Underlying data:** A table showing the input data for 'column 2'.
- Raw data plot:** A line graph showing the values for 'column 2' across six indices.
- Histogram of values:** A bar chart showing the distribution of values for 'column 2'.
- Pie chart:** A pie chart showing the percentage distribution of values for 'column 2'.
- Statistics:** A table of statistical measures for 'column 2'.
- Bar chart:** A bar chart showing the distribution of values for 'column 2' across six categories.
- Heat map of column 2:** A map of Europe with countries colored according to their values for 'column 2'. A red arrow points to this map from the text below.

Country	column 2
France	2.2
Germany	3.1
Greece	0.20
Italy	1.6
Spain	4.5
United Kingdom	1.3

column 2	
mean	2.2
s.d.	1.5
min.	0.20 (Greece)
median	1.9
max.	4.5 (Spain)
total	13
count	6

Another data example

	A	B	C	D	E	F	G	H	I	J
1	MyField	MyCategory	ISI Category	# Articles	Total Cites	Median IF	Agg. IF	Agg. Immed. Index	Agg. Cited 1/2 Life	# Journals
35	Med	Clinical	EMERGENCY MEDICINE	3463	67394	1.081	1.76	0.355	6.7	24
36	Med	Clinical	TROPICAL MEDICINE	2979	62811	1.107	2.096	0.372	6.9	20
37	Med	Clinical	MEDICAL LABORATORY TECHNOLOGY	2960	76496	1.319	2.103	0.438	7.4	32
38	Med	Clinical	INTEGRATIVE & COMPLEMENTARY MEDICINE	2628	41091	1.329	2.195	0.291	6.1	22
39	Med	Clinical	MEDICAL INFORMATICS	2083	46210	1.516	1.782	0.438	7.1	23
40	Med	Clinical	AUDIOLGY & SPEECH-LANGUAGE PATHOLOGY	1955	64894	1.55	1.744	0.274	>10.0	21
41	Med	Clinical	MEDICINE, LEGAL	1660	25546	1.229	1.82	0.28	6.4	15
42	Med	Clinical	SUBSTANCE ABUSE	1649	57124	2.359	2.847	0.598	7.2	14
43	Med	Clinical	PRIMARY HEALTH CARE	1084	19994	1.269	1.462	0.744	6.8	14
44	Med	Clinical	MEDICAL ETHICS	676	8391	0.917	1.264	0.731	5.1	17
45	Med	Clinical	ANDROLOGY	477	9840	1.535	2.222	0.499	6.8	6
46	Sci	Agriculture	FOOD SCIENCE & TECHNOLOGY	18470	421932	1.165	1.898	0.315	7.2	128
47	Sci	Agriculture	AGRONOMY	7485	171793	1	1.452	0.281	8.5	79
48	Sci	Agriculture	AGRICULTURE, MULTIDISCIPLINARY	6068	125325	0.495	1.439	0.25	7.8	57
49	Sci	Agriculture	AGRICULTURE, DAIRY & ANIMAL SCIENCE	5494	121768	0.698	1.296	0.228	8.6	55
50	Sci	Agriculture	FISHERIES	4661	129802	1.177	1.548	0.301	9.1	48
51	Sci	Agriculture	FORESTRY	4068	96892	0.868	1.557	0.324	8.1	59
52	Sci	Agriculture	SOIL SCIENCE	3659	132592	1.53	1.78	0.365	9.8	33
53	Sci	Agriculture	AGRICULTURAL ENGINEERING	3344	63520	1.01	3.193	0.616	4.9	12
54	Sci	Agriculture	HORTICULTURE	3156	71876	0.75	1.28	0.202	9.1	31
55	Sci	Agriculture	LIMNOLOGY	1840	67019	1.207	1.843	0.503	>10.0	19
56	Sci	Agriculture	AGRICULTURAL ECONOMICS & POLICY	536	10575	0.893	1.069	0.17	8.7	14
57	Sci	Astronomy	ASTRONOMY & ASTROPHYSICS	16525	680837	1.683	4.242	1.439	6.8	56
58	Sci	Bio/BioMed	BIOCHEMISTRY & MOLECULAR BIOLOGY	51425	2893583	2.857	4.276	0.873	7.7	289
59	Sci	Bio/BioMed	PHARMACOLOGY & PHARMACY	33464	1085783	2.153	2.943	0.541	6.7	261
60	Sci	Bio/BioMed	NEUROSCIENCES	33223	1666914	2.759	3.951	0.782	7.4	243
61	Sci	Bio/BioMed	CELL BIOLOGY	23181	1581141	3.271	5.779	1.186	6.9	180

This example is of some more extensive, real-life data on journal impact factors.

We get back a lot of automated analysis, much of which is genuinely useful. It includes some natural language statements that can be drawn from the data.



The screenshot shows a detailed statistical analysis interface with the following sections:

- Statistical statements:** A list of 10 statements regarding correlations between variables, all marked as "100% confidence".
- Regression results:** A table showing the regression model $y = a + b_1x + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6$ and the estimated coefficients for each variable.
- Statistical statements (continued):** A list of 10 statements regarding the relationship between MyField and various numeric columns, with confidence levels ranging from < 90% to > 99%.
- One-way ANOVA:** A table showing the F-ratio (14.5) and p-value (1.18×10^{-6}) for the comparison of # Journals across MyField categories.
- Grouped bar chart:** A bar chart showing the mean values of # Articles, Total Cites, Median IF, Agg. IF, Agg. Immed. Index, and # Journals for each MyField category.

Publishing interactive content

CDF (Computable Document Format)

wolfram.com/cdf

The screenshot shows the top navigation bar with 'WOLFRAM' and links for PRODUCTS, SOLUTIONS, PURCHASE, SUPPORT, COMMUNITY, COMPANY, and OUR SITES. Below is the 'CDF' logo and the text 'Computable Document Format Documents come alive with the power of computation'. A secondary navigation bar includes 'Overview', 'Uses & Examples', 'Why CDF?', 'Adopting CDF', 'Resources', and 'FAQs'. The main content area features a section titled 'Uses and Examples of the Computable Document Format (CDF)' with a 'Get the Free CDF InfoKit' button. Below this is a grid of icons representing various application categories: Consulting Reports, Investment Statements, Management Presentations, Automated Reports, Infographics and Blogs, Textbooks, Journal Articles, Knowledge Apps, and Data Applications. A 'View Full Example' button is also present.

CDF Demonstrations

demonstrations.wolfram.com

The screenshot shows the 'Wolfram Demonstrations Project' website. The top navigation bar includes 'EXPLORE', 'LATEST', 'ABOUT', 'PARTICIPATE', and 'AUTHORING AREA'. The main content area features a large banner with the text 'Bring ideas to life from recreation & education to research & industry' and a 'Start exploring' button. Below the banner is a 'FEATURED DEMONSTRATIONS' section with a grid of 24 small thumbnail images. At the bottom, there is a 'BROWSE TOPICS' section with icons for Mathematics, Business & Social Systems, Creative Arts, Computation, Systems, Models & Methods, Kids & Fun, Physical Sciences, Engineering & Technology, Mathematics Functionality, Life Sciences, Our World, and Browse by US Common Core State Standards. A 'FEATURED CONTRIBUTORS' section lists Karl Scherer, Abraham Gadalla, and Marc Brodie.

The future ...

- **More powerful data analysis, yet still easy to use.**
- **New mechanisms for publishing actionable data.**
- **Cloud-based, plug-in free.**