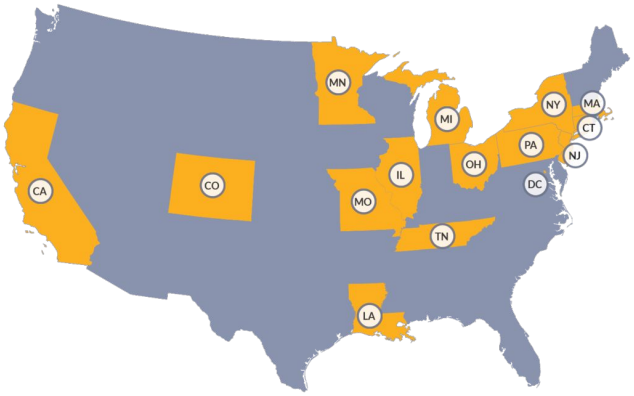
April 28, 2016

Building an Analytical Toolkit Beyond Excel

Alan Zhao
Jeremy James

Achievement Network (ANet): At a glance

- ANet is a **non-profit organization** that partners with ~700 K-8 schools around the country. We provide multiple services to our partners, including:
    - **Leadership** training and professional development
    - Formative, interim **assessments** aligned to state standards
    - Online tools that provide **resources** for standards based and data driven instruction
- **TL;DR**. We help schools through assessments, coaching and resources. Data is behind each aspect of our program model.



Who are we?

Alan Zhao
Senior Analyst,
Analysis & Reporting
(external facing analysis)

Jeremy James
Manager,
Analysis & Insights
(internal facing analysis)

What are our tools?

But what about in 2013?

VLOOKUP (lookup_value, table_array, col_index_num, [range_lookup])

For example:

- =VLOOKUP(105,A2:C7,2,TRUE)
- =VLOOKUP("Fontana",B2:E7,2,FALSE)

1

	A	B	C
1	**Sport**	**Quarter**	**Sales**
2	Golf	Qtr3	$1,500
3	Golf	Qtr4	$2,000
4	Tennis	Qtr3	$600
5	Tennis	Qtr4	$1,500
6	Tennis	Qtr3	$4,070
7	Tennis	Qtr4	$5,000
8	Golf	Qtr3	$6,430

2

3

E	F	G
Sum of Sales	Quarter	
Sport	Qtr3	Qtr4
Golf	$7,930	$2,000
Tennis	$4,670	$6,500
Grand Total	$12,600	$8,500

4

"When all you have is a hammer, everything looks like a nail"

"When all you have is Excel, everything looks like a pivot table"

Thousands of hours formatting reports

Student Performance, A1 6th (by Grade) - Page 1
Bold Thinking Academy 2015-2016

QUICK STATS: 26 questions, 28 points possible, 5 students
Overall Performance: Mean: 42%, Median: 32%

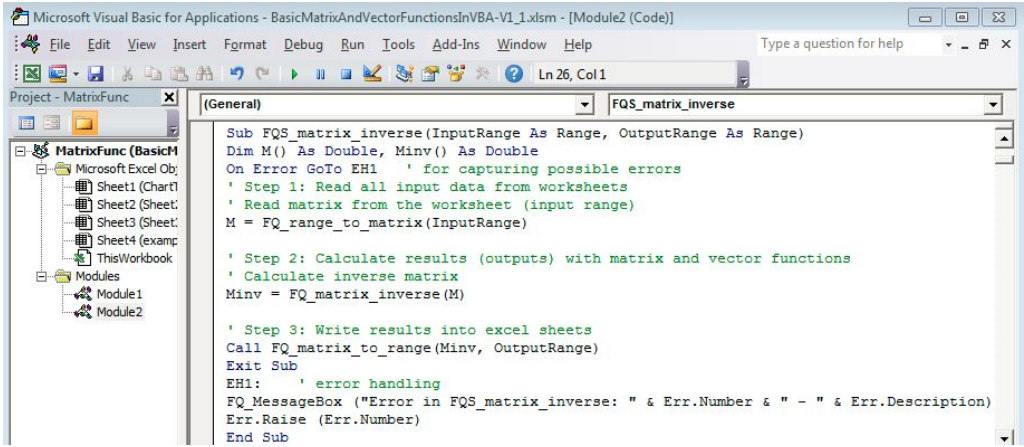
Item Type KeySR - Selected ResponseMS - Multi-SelectSA - Short Answe					Major: 6.NS.A.1 Interpret, Compute, and Solve Word Problems with			6.RP.A.1 Understand Ratios and Describe Ratio Relationships			6.RP.A.2 Understand Unit Rates			6.RP.A.3a Use Tables and Graphs to Make Equivalent Ratios and Compare		
Weighted Average by Standard (Grade)					30%			35%			47%			60%		
% Points Possible Earned by Item (Grade)					40%	0%	40%	60%	60%	10%	60%	40%	40%	80%	60%	40%
Question #					10	17	25	12	19	24	14	21	26	2	4	6
Correct Response (Points Possible)					C	C	AE (2)	B	(1)	BE (2)	D	B	AC (1)	A	B	(1)
Item Type					SR	SR	MS	SR	TEI	MS	SR	SR	MS	SR	SR	TEI
Student	Teacher	Period	Raw Score	Weighted Average												
Souliere, Tyrone	Kamaka, Qu	6-Math Kamaka	5	18%	A	B	BCE (0)	C	0	BCD (0)	C	A	ABC (0)	*	A	0
Dunavant, Kandis	Kamaka, Qu	6-Math Kamaka	6	21%	A	B	A (1)	D	0	C (0)	B	C	AC (1)	*	A	1
Deshaies, Shalond	Kamaka, Qu	6-Math Kamaka	9	32%	B	A	AC (1)	*	1	BCD (0)	*	A	AE (0)	B	*	0
Hamby, Jalisa	Kamaka, Qu	6-Math Kamaka	17	61%	*	A	A (1)	*	1	BCD (0)	*	*	A (0)	*	*	1
Cargo, Adrianna	Kamaka, Qu	6-Math Kamaka	22	79%	*	B	A (1)	*	1	B (1)	*	*	AC (1)	*	*	0
Number of Students Selecting Each Answer Choice					A	2	2	4*	0	correct	0	0	2	5*	4*	2	correct
*denotes Correct Response					B	1	3	1	3*	partial	4*	1	2*	1	1	3*	partial
					C	2*	0*	2	1	ncorrect	4	1	1	3*	0	0	ncorrect
Multi-Select: total # of student selections may exceed total students for a given question					D	0	0	0	1		3	3*	0	0	0	0	
					E	0	0	1*	0		0*	0	0	1	0	0	
					Omitted	0	0	0	0	0	0	0	0	0	0	0	0

rt Answer					Major: 6.NS.A.1 Interpret, Compute, and Solve Word Problems with Division of Fractions			6.RP.A.1 Understand Ratios and Describe Ratio Relationships			6.RP.A.2 Understand Unit Rates			6.RP.A.3a Use Tables and Graphs to Make Equivalent Ratios and Compare Ratios		
andard (Grade)					30%			35%			47%			60%		
by Item (Grade)					40%	0%	40%	60%	60%	10%	60%	40%	40%	80%	60%	40%
Question #					10	17	25	12	19	24	14	21	26	2	4	6
Correct Response (Points Possible)					C	C	AE (2)	B	(1)	BE (2)	D	B	AC (1)	A	B	(1)
Item Type					SR	SR	MS	SR	TEI	MS	SR	SR	MS	SR	SR	TEI
Student	Teacher	Period	Raw Score	Weighted Average												
Souliere, Tyrone	Kamaka, Qu	6-Math Kamaka	5	18%	A	B	BCE (0)	C	0	BCD (0)	C	A	ABC (0)	*	A	0
Dunavant, Kandis	Kamaka, Qu	6-Math Kamaka	6	21%	A	B	A (1)	D	0	C (0)	B	C	AC (1)	*	A	1
Deshaies, Shalonda	Kamaka, Qu	6-Math Kamaka	9	32%	B	A	AC (1)	*	1	BCD (0)	*	A	AE (0)	B	*	0
Hamby, Jalisa	Kamaka, Qu	6-Math Kamaka	17	61%	*	A	A (1)	*	1	BCD (0)	*	*	A (0)	*	*	1
Cargo, Adrianna	Kamaka, Qu	6-Math Kamaka	22	79%	*	B	A (1)	*	1	B (1)	*	*	AC (1)	*	*	0
Number of Students Selecting Each Answer Choice					A	2	2	4*	0	3 correct	0	0	2	5*	4*	2	2 correct
*denotes Correct Response					B	1	3	1	3*	0 partial	4*	1	2*	1	1	3*	0 partial
					C	2*	0*	2	1	2 incorrect	4	1	1	3*	0	0	3 incorrect
Multi-Select: total # of student selections may exceed total students for a given question					D	0	0	0	1		3	3*	0	0	0	0	
					E	0	0	1*	0		0*	0	0	1	0	0	
					Omitted	0	0	0	0	0	0	0	0	0	0	0	0

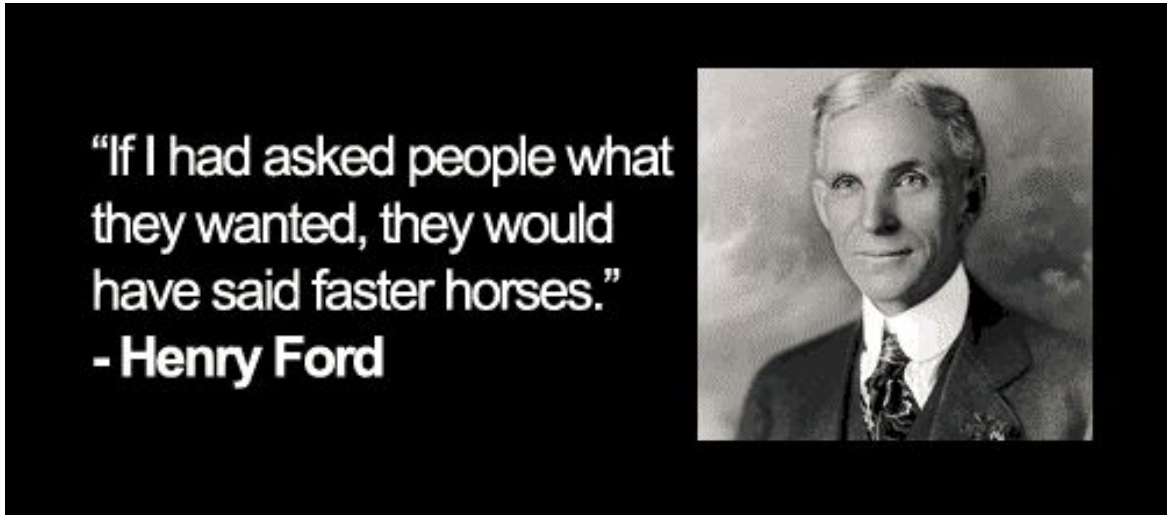
Excel VBA to the rescue

Visual Basic for Applications (VBA) is a complete programming language that powers Excel.

- Every action in the Excel interface corresponds to code
- At ANet, we applied VBA to automate the formatting process
- Archaic language, but effective



Excel VBA to excess



The Data Pyramid

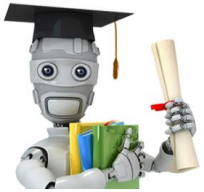
What drives student performance gains?

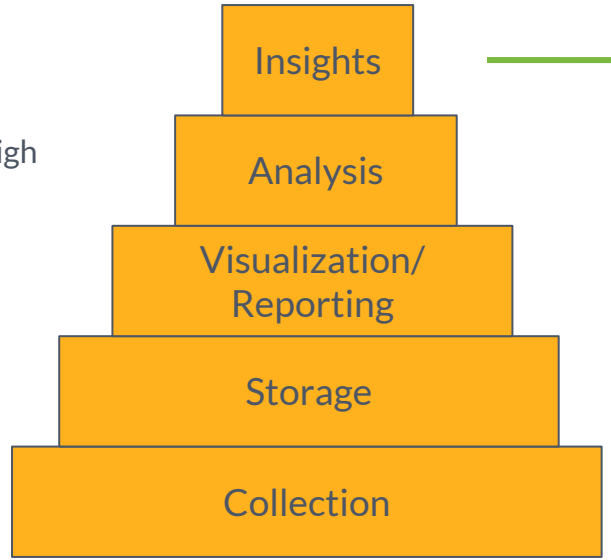
What are the differences between high performers and low performers?

Who are the high performers and low performers?

How do we access performance data?

Where can we find school performance data? What about school categorical data?





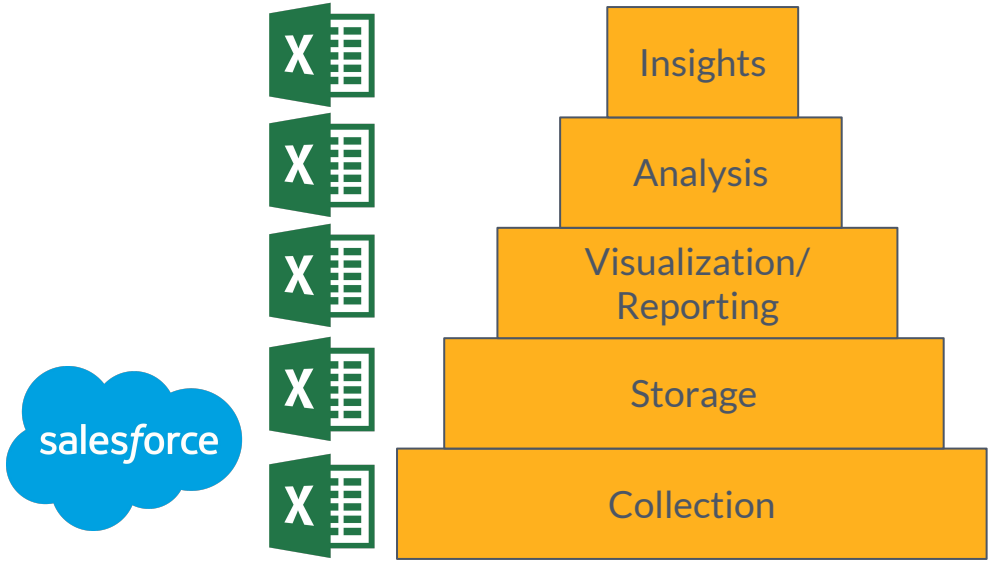
Machine learning lives here??

What your boss cares about

What analysts do

Where your data lives

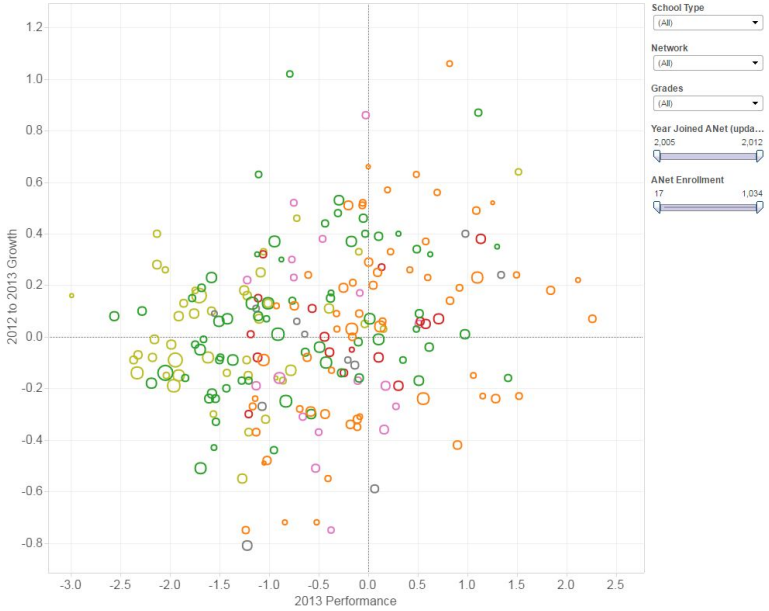
The Data Pyramid, circa 2012-13



So we started with Tableau

- Tableau, data visualization software
- Like a pivot table, but feeds directly into visualizations
- Primary focus on visual analysis of data, not statistical analysis, no ability to edit data

Tableau, the good



My first dashboard!

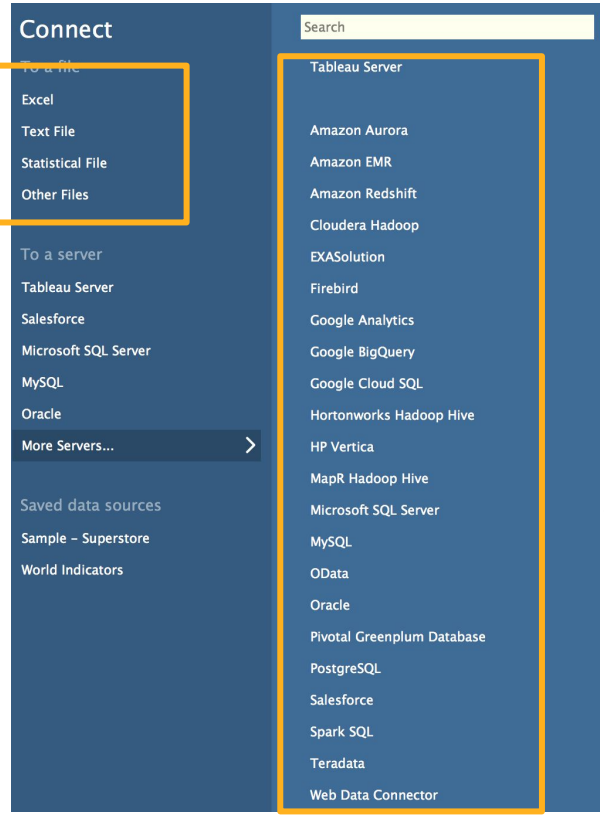
Some great advantages to Tableau

- Easy-ish to learn
- Can probably work with your current datasets
- Interactivity provided an immediate win

Tableau, the bad

At its heart, a visualization platform

- Works with spreadsheets
  - But what if your data is too big for Excel?
- Most powerful when you have access across all your data
  - But what if your data is separated into Excel spreadsheets?



4 types of static files!

20 + types of databases!

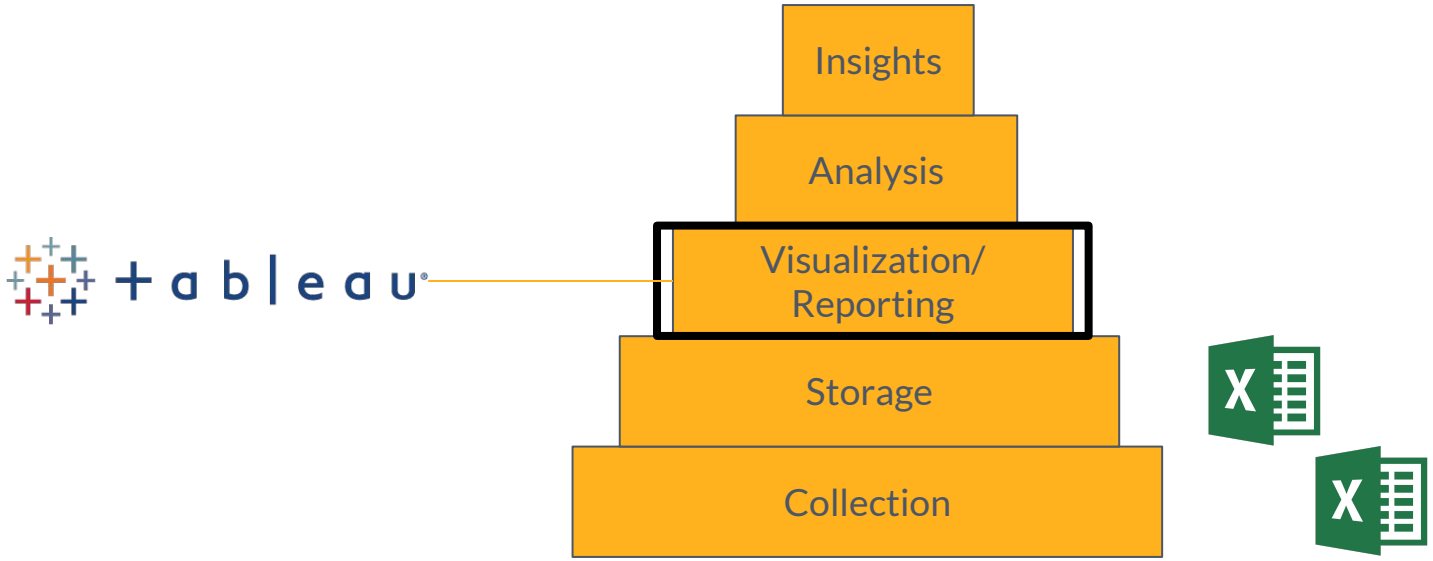
Tableau's own words!

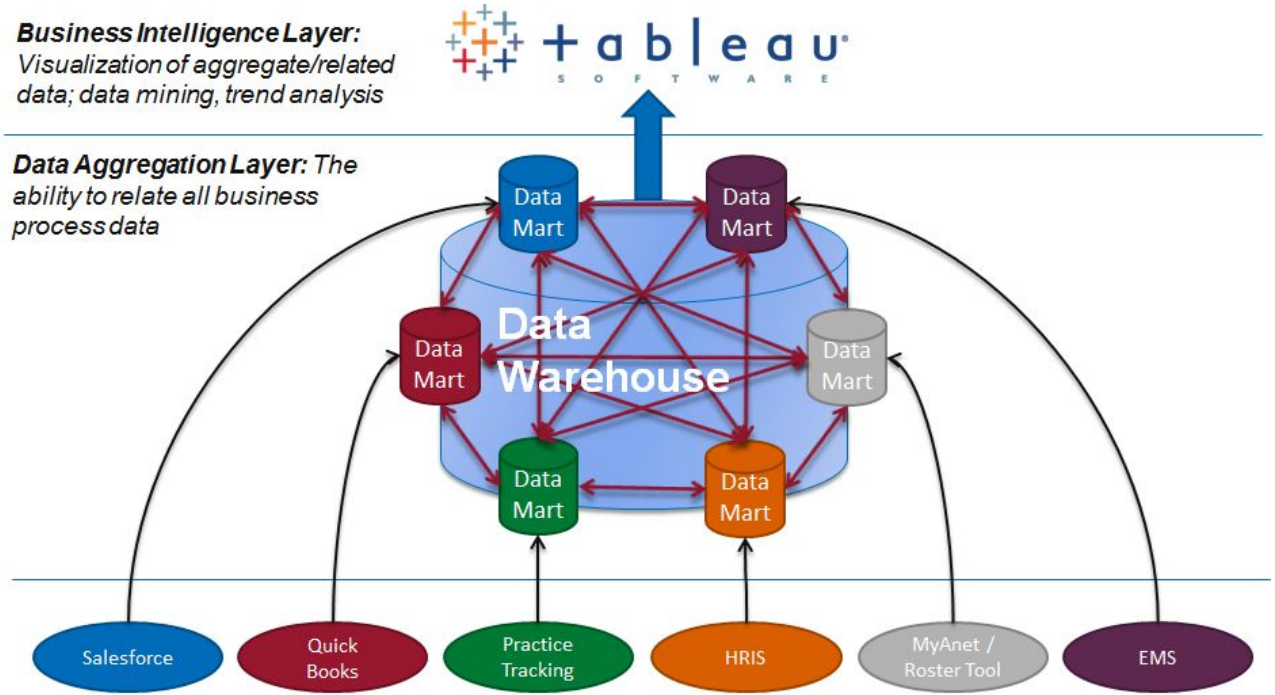
"A Live, Direct Data Connection is Better…

Avoid data silos and ensure a single point of truth by pointing your analyses at a single, optimized database. You can give business people the power to ask and answer questions of massive data sets just by pointing to the source data."

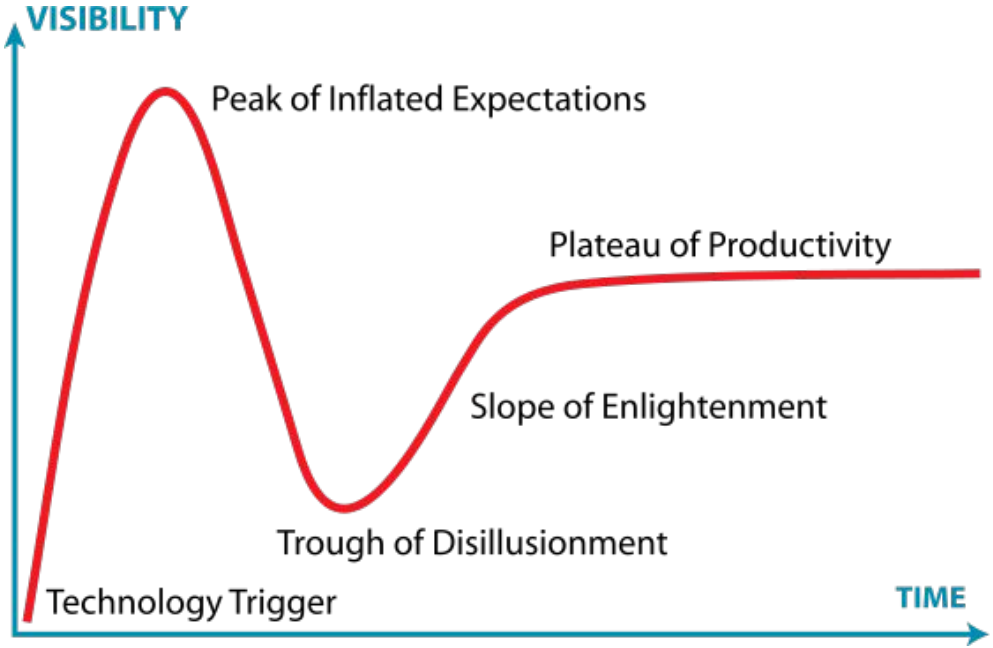
Ellie Fields, Tableau VP of Product Marketing.
Tableau Whitepaper: In-Memory or Live Data: Which Is Better
http://www.tableau.com/learn/whitepapers/memory-or-live-data

Tableau (without a database), the ugly

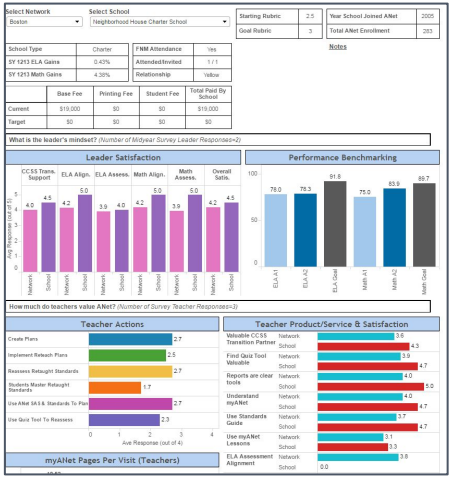



Business Intelligence Layer: Visualization of aggregate/related data; data mining, trend analysis
tableau SOFTWARE
Data Aggregation Layer: The ability to relate all business process data
Data Mart
Data Mart
Data Mart
Data Warehouse
Data Mart
Data Mart
Data Mart
Salesforce
Quick Books
Practice Tracking
HRIS
MyAnet / Roster Tool
EMS
Foundation Layer: Operational Data Systems aligned with specific business processes

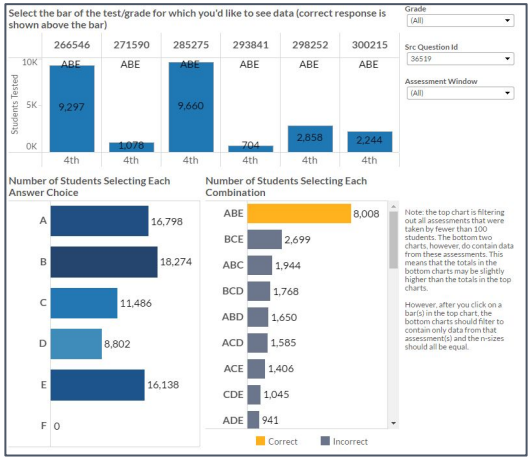
The march towards data infrastructure - two years

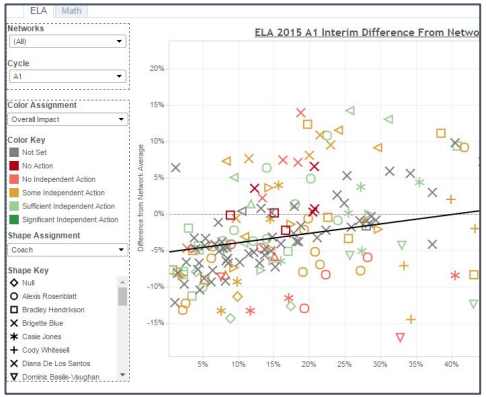


This new infrastructure allowed us to automate some previously very difficult tasks

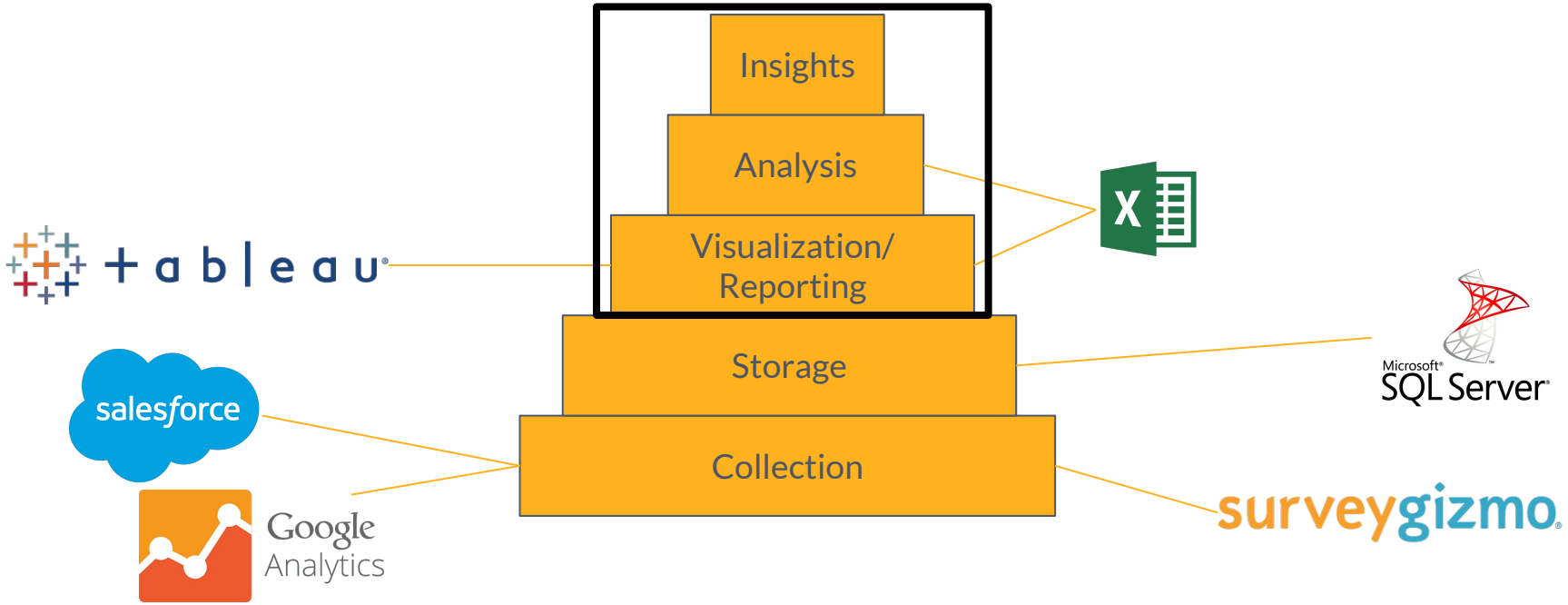


Less Repetition!





The Data Pyramid, circa 2014



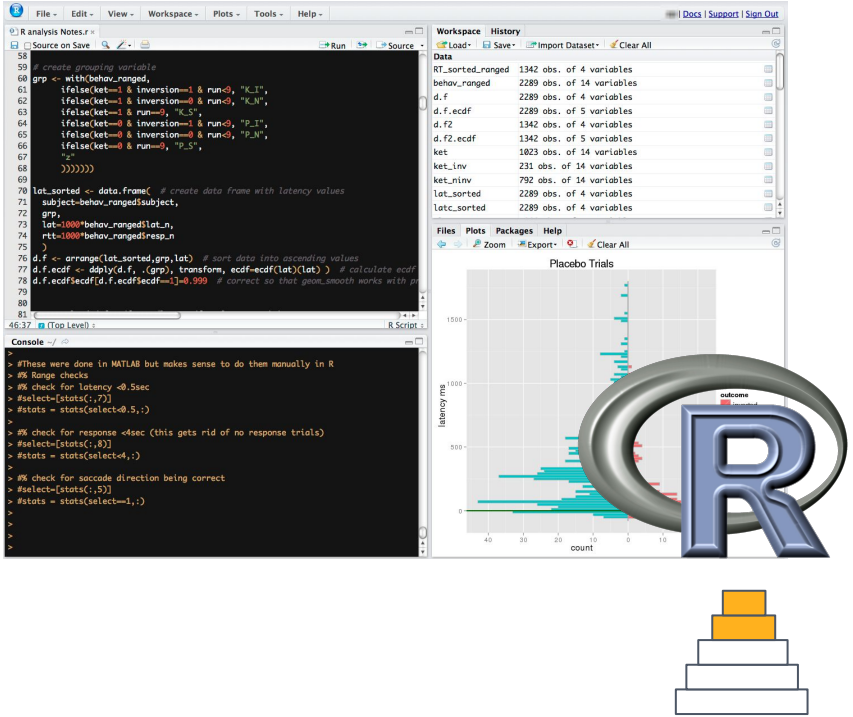
Back end innovations vastly improved our analytical capacity

Some great advantages to R

- Amazingly powerful/flexible
- Can be used to apply a higher level of rigor to your analysis
- Can handle very complex calculations on very large datasets

Some disadvantages to R

- Not as easy to pick up for a beginner



Leveraging Python to Automate Reporting (and more)

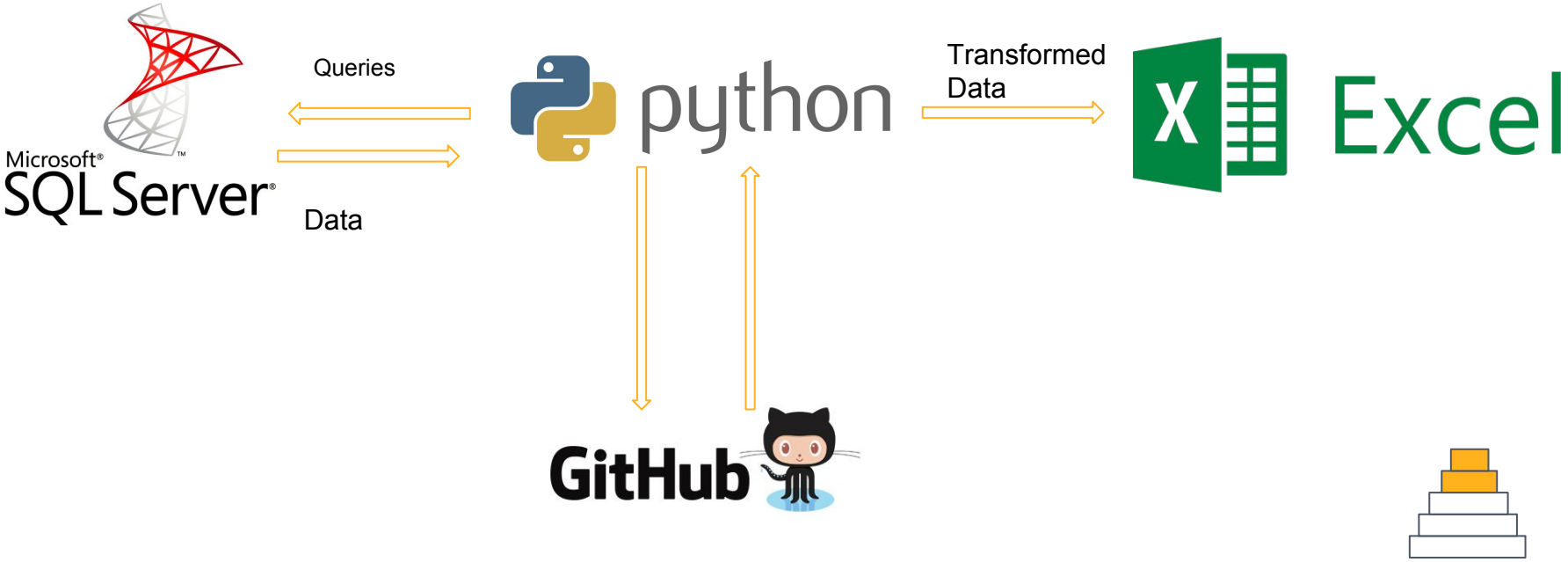
Some great advantages to Python

- Most popular first programming language
- General programming language, powerful/flexible
- Tons of packages

Some disadvantages to Python

- Slower for “big” data
- Learning curve

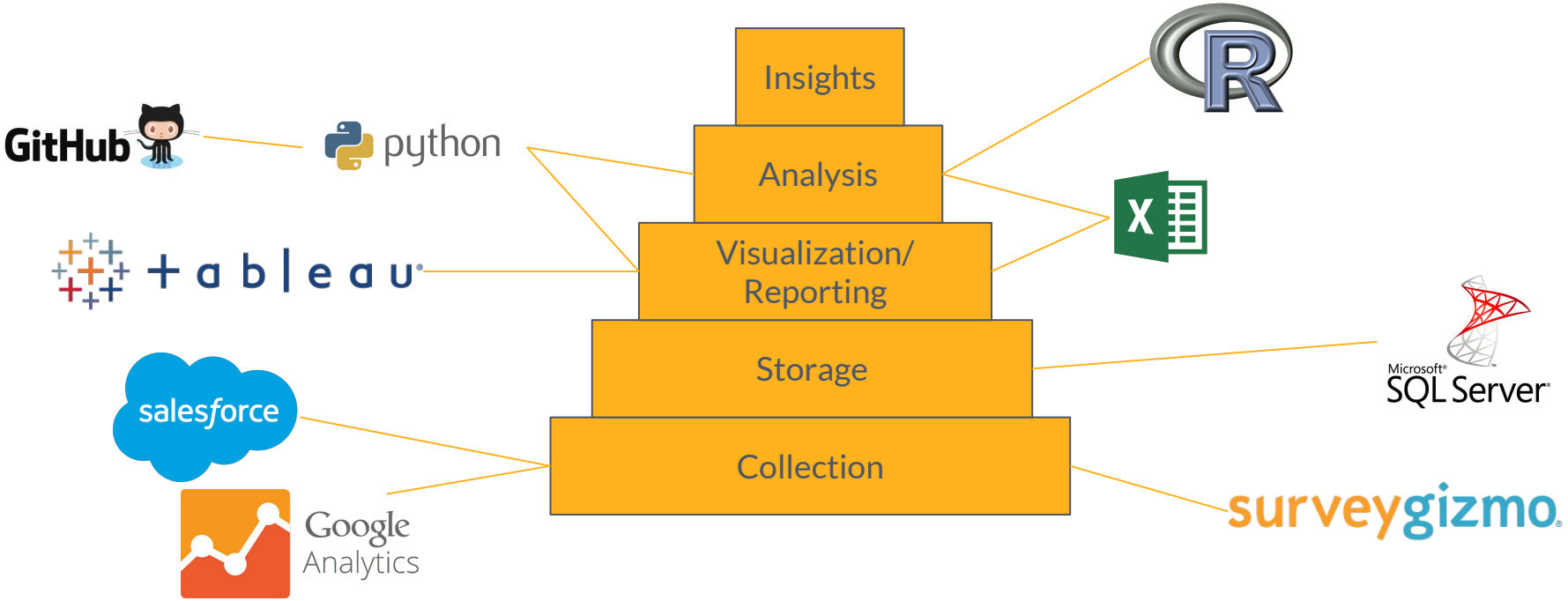
New Tools, same end product



Developing analytic human capital in house

Anyone can program, with curiosity and **relevant application**.

The Data Pyramid, present day



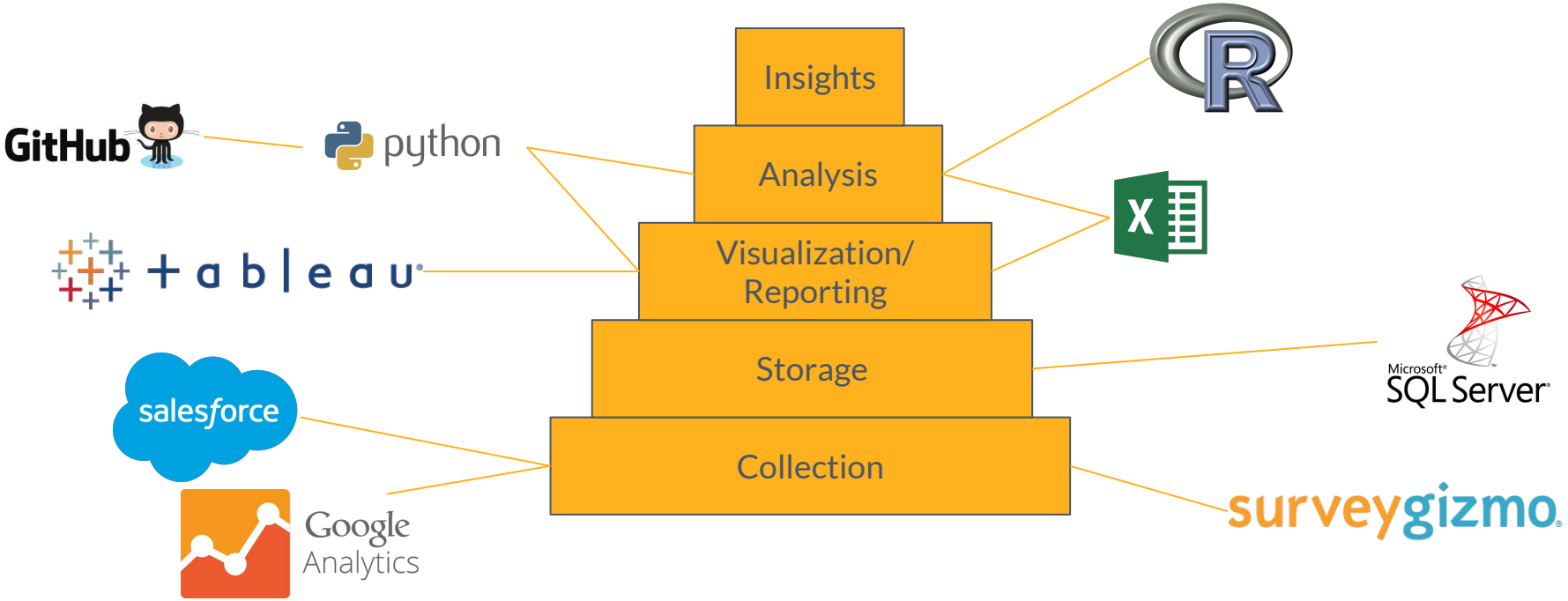
Always innovating - our path forward

- Moving more datasets into our central database
- Finding ways to collect the new data we come in contact with
- Automating analysis for standard requests

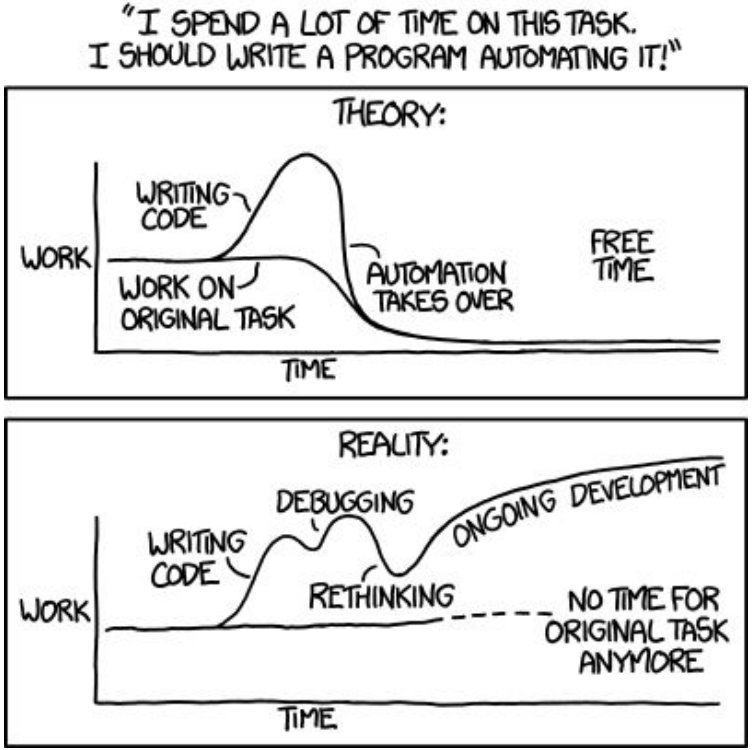
Match the Right Tool to each Problem

	Common Problems	Possible Tools
Collection	"I don't know if this data **is right**" "I know what data I want, I just **don't have it**."	SurveyGizmo, Salesforce, Google Analytics
Storage	"It takes me **hours** to compile the data" "My data is **too big** for excel" "I spend days downloading data to put it in excel"	Salesforce, SQL, EXCEL!
Visualization/Reporting	"My presentations are **boring** pivot tables" "I'm spending all my time formatting and emailing out excel reports" "I create the same report **over and over.**"	Tableau, R, EXCEL!
Analysis	"Excel's formulas are making my **computer crash.**" "Excel's statistical functions aren't cutting it." "I run the **same analysis daily** for different samples"	Tableau, R, Python, Stata, SAS, EXCEL!
Insights	"I have data, but it's **not actionable.**"	Humans, "machine learning" (with humans)

It’s a pyramid - start at the bottom.



Change takes time.



With apologies to...

We're hiring!!

www.achievementnetwork.org/careers

Alan: azhao@achievementnetwork.org
Jeremy: jerjames13@gmail.com