

April 28, 2016

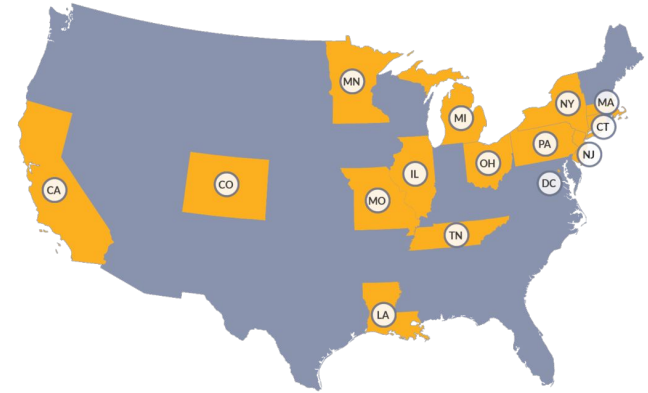
# Building an Analytical Toolkit Beyond Excel



Alan Zhao  
Jeremy James

# Achievement Network (ANet): At a glance

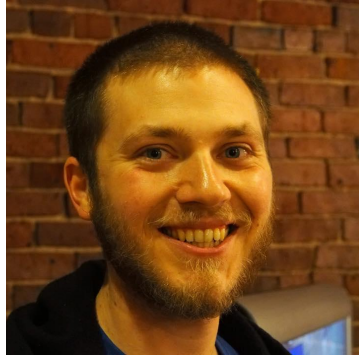
- ANet is a **non-profit organization** that partners with ~700 K-8 schools around the country. We provide multiple services to our partners, including:
  - **Leadership** training and professional development
  - Formative, interim **assessments** aligned to state standards
  - Online tools that provide **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?



*“When all you have is a hammer, everything looks like a nail”*

**VLOOKUP (lookup\_value, table\_array, col\_index\_num, [range\_lookup])**

For example:

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

Diagram illustrating VLOOKUP usage with two tables and numbered callouts:

**Table 1 (Top):** A 7x3 table with columns A, B, and C. Row 1 is the header: Sport, Quarter, Sales. Rows 2-7 contain data. A red vertical line is drawn between columns B and C, and a blue horizontal line is drawn between rows 7 and 8. Callout 1 points to the header row, and callout 2 points to the data rows.

	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

**Table 2 (Bottom):** A 5x3 table with columns E, F, and G. Row 1 is the header: Sum of Sales, Quarter, and a blank cell. Rows 2-5 contain data. A blue horizontal line is drawn between rows 4 and 5. Callout 3 points to the header row, and callout 4 points to the data rows.

	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	

*“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%											

					Major												
Item Type KeySR - Selected ResponseMS - Multi-SelectSA - Short Answer	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, Kandise	Kamaka, Qu	6-Math Kamaka	6	21%	A	B	A (1)	D	0	C (0)	B	C	AC (1)	*	A	1			
Deshaiies, Shaloni	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	incorrect	4	1	1	3*	0	0	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	0				
					E	0	0	1*	0	0*	0	0	1	0	0	0				
					Omitted	0	0	0	0	0	0	0	0	0	0	0				

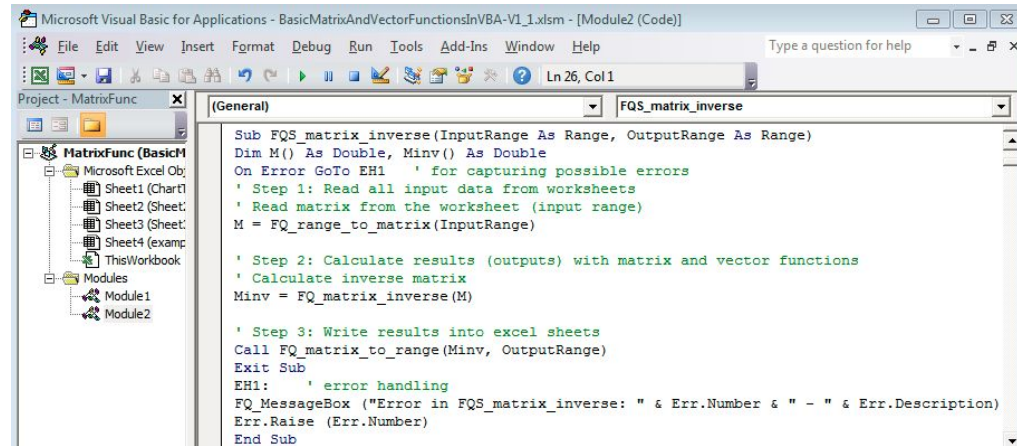
Student	Teacher	Period	Raw Score	Weighted Average															
Souliere, Tyrone	Kamaka, Qu	6-Math Kamaka	5	18%	A	B	BCE (0)	C	0	BCD (0)	C	A	ABC (0)	*	A	0			
Dunavant, Kandise	Kamaka, Qu	6-Math Kamaka	6	21%	A	B	A (1)	D	0	C (0)	B	C	AC (1)	*	A	1			
Deshaiies, Shaloni	Kamaka, Qu	6-Math Kamaka	9	32%	B	A	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



The screenshot shows the Microsoft Visual Basic for Applications editor window. The title bar reads "Microsoft Visual Basic for Applications - BasicMatrixAndVectorFunctionsInVBA-V1\_1.xlsm - [Module2 (Code)]". The menu bar includes File, Edit, View, Insert, Format, Debug, Run, Tools, Add-Ins, Window, and Help. The status bar at the bottom indicates "Ln 26, Col 1". On the left, the Project Explorer shows a tree view with "MatrixFunc (BasicM)" expanded, containing "Microsoft Excel Obj:", "Sheet1 (Chart)", "Sheet2 (Sheet:", "Sheet3 (Sheet:", "Sheet4 (examp", and "ThisWorkbook". Under "Modules", "Module1" and "Module2" are listed. The main window displays the code for the "FQS\_matrix\_inverse" subroutine. The code is as follows:

```
Sub FQS_matrix_inverse(InputRange As Range, OutputRange As Range)
    Dim M() As Double, Minv() As Double
    On Error GoTo EH1 ' for capturing possible errors
    ' Step 1: Read all input data from worksheets
    ' Read matrix from the worksheet (input range)
    M = FQ_range_to_matrix(InputRange)

    ' Step 2: Calculate results (outputs) with matrix and vector functions
    ' Calculate inverse matrix
    Minv = FQ_matrix_inverse(M)

    ' Step 3: Write results into excel sheets
    Call FQ_matrix_to_range(Minv, OutputRange)
Exit Sub
EH1: ' error handling
    FQ_MessageBox ("Error in FQS_matrix_inverse: " & Err.Number & " - " & Err.Description)
    Err.Raise (Err.Number)
End Sub
```

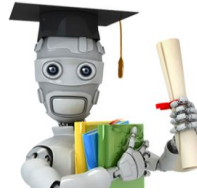
## Excel VBA to excess

“If I had asked people what they wanted, they would have said faster horses.”  
- **Henry Ford**





# 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?

Insights

Analysis

Visualization/  
Reporting

Storage

Collection

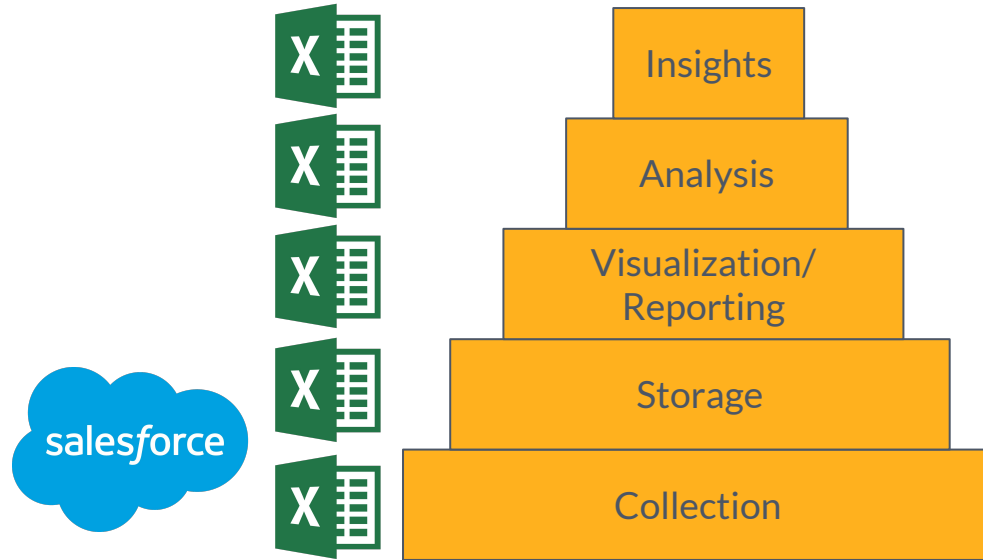
Machine learning lives here??

What your boss cares about

What analysts do

Where your data lives

# The Data Pyramid, circa 2012-13



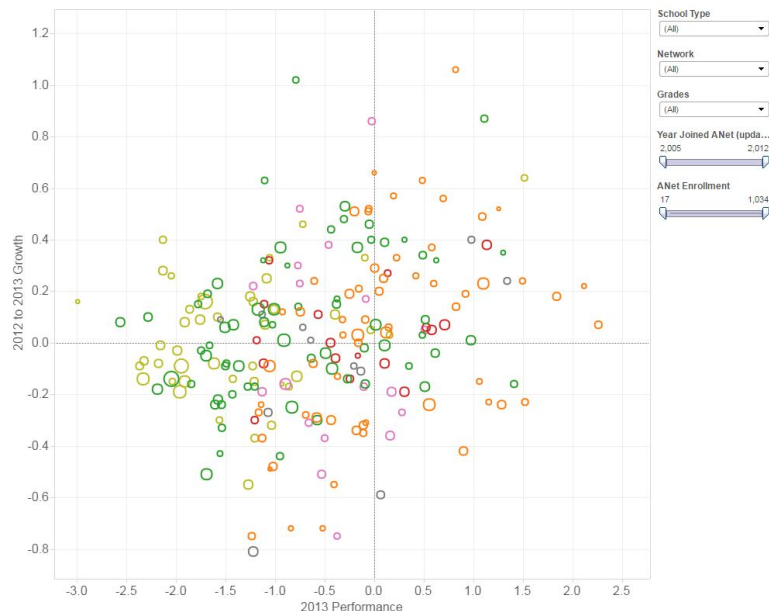
## So we started with Tableau



- 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

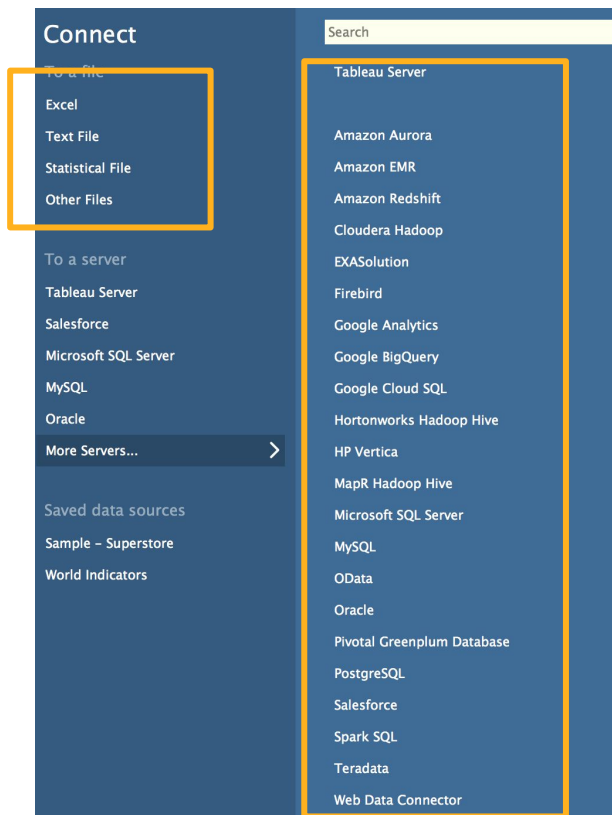
- Easy-ish to learn
- Can probably work with your current datasets
- Interactivity provided an immediate win



# 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.”

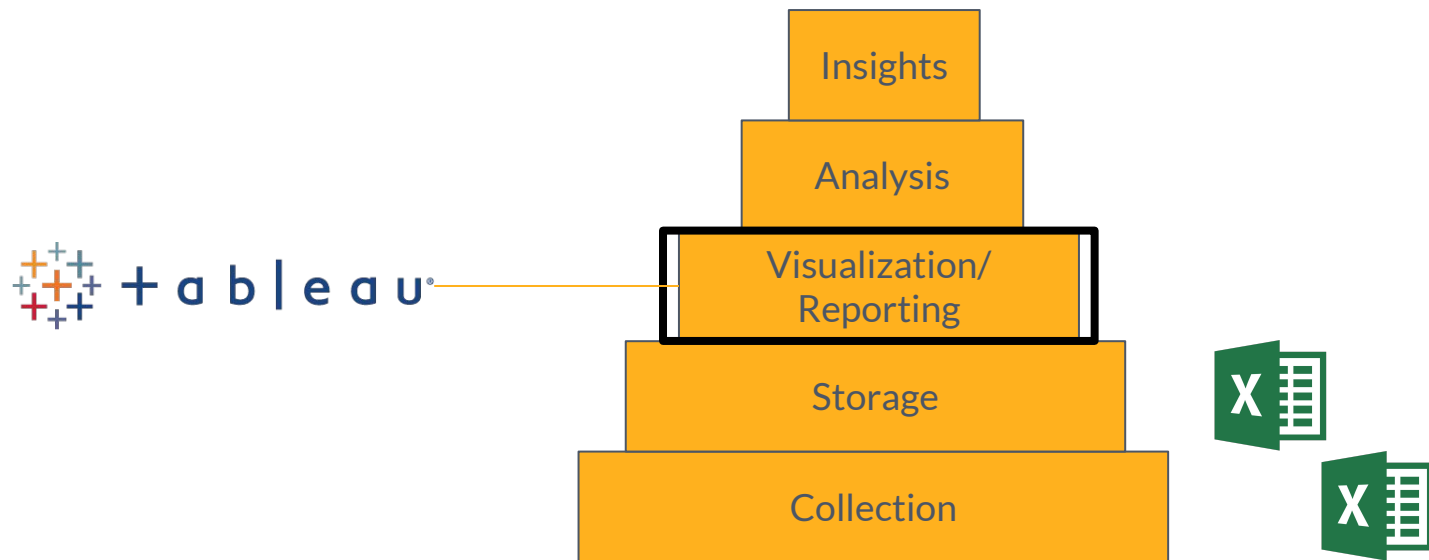
**Ellie Fields, Tableau VP of Product Marketing.**

**Tableau Whitepaper: In-Memory or Live Data: Which Is Better**

**<http://www.tableau.com/learn/whitepapers/memory-or-live-data>**



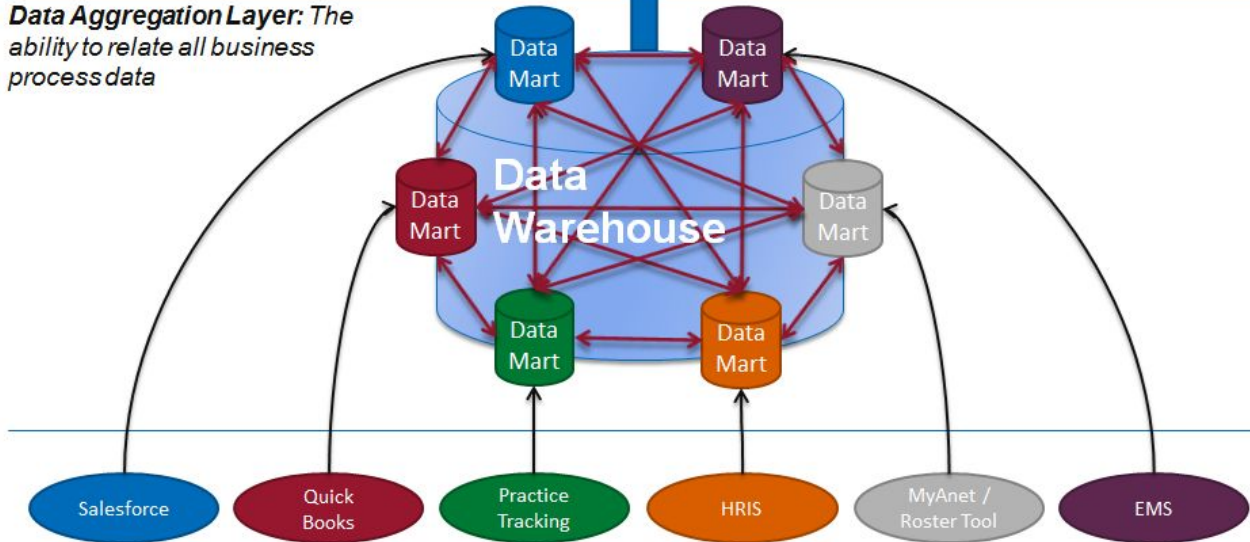
## Tableau (without a database), the ugly



**Business Intelligence Layer:**  
*Visualization of aggregate/related  
data; data mining, trend analysis*



**Data Aggregation Layer:** *The  
ability to relate all business  
process data*

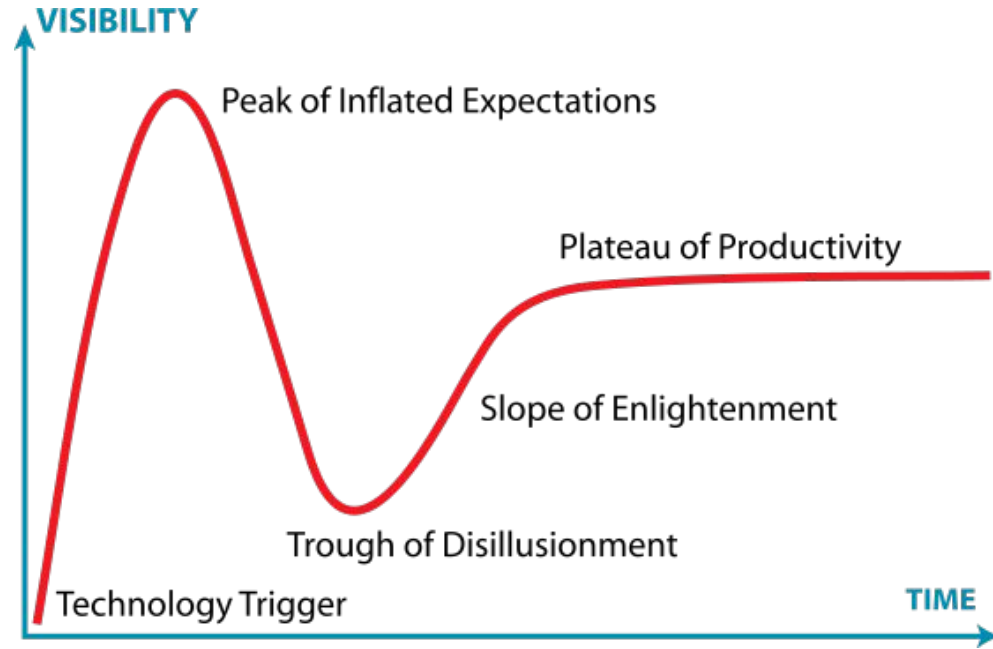


**Foundation Layer:** *Operational Data Systems aligned with specific business  
processes*



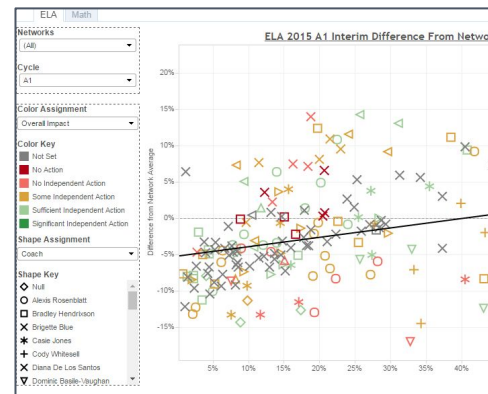
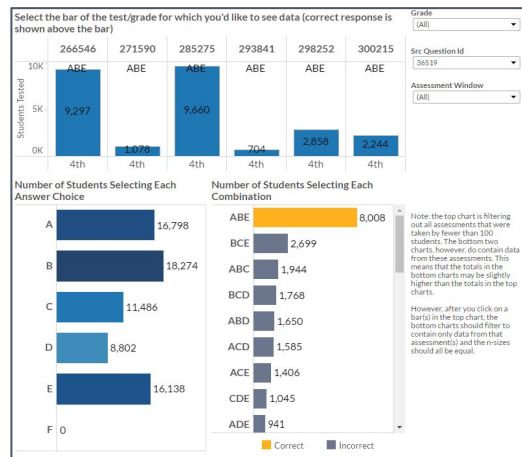
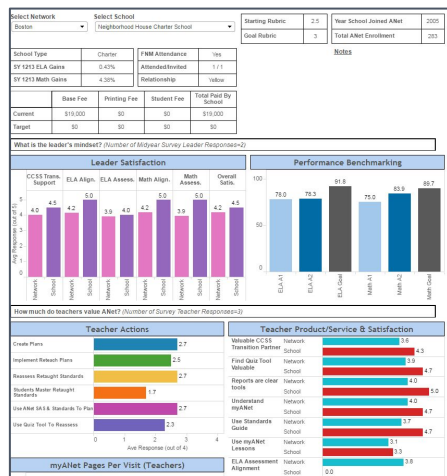


# The march towards data infrastructure - two years

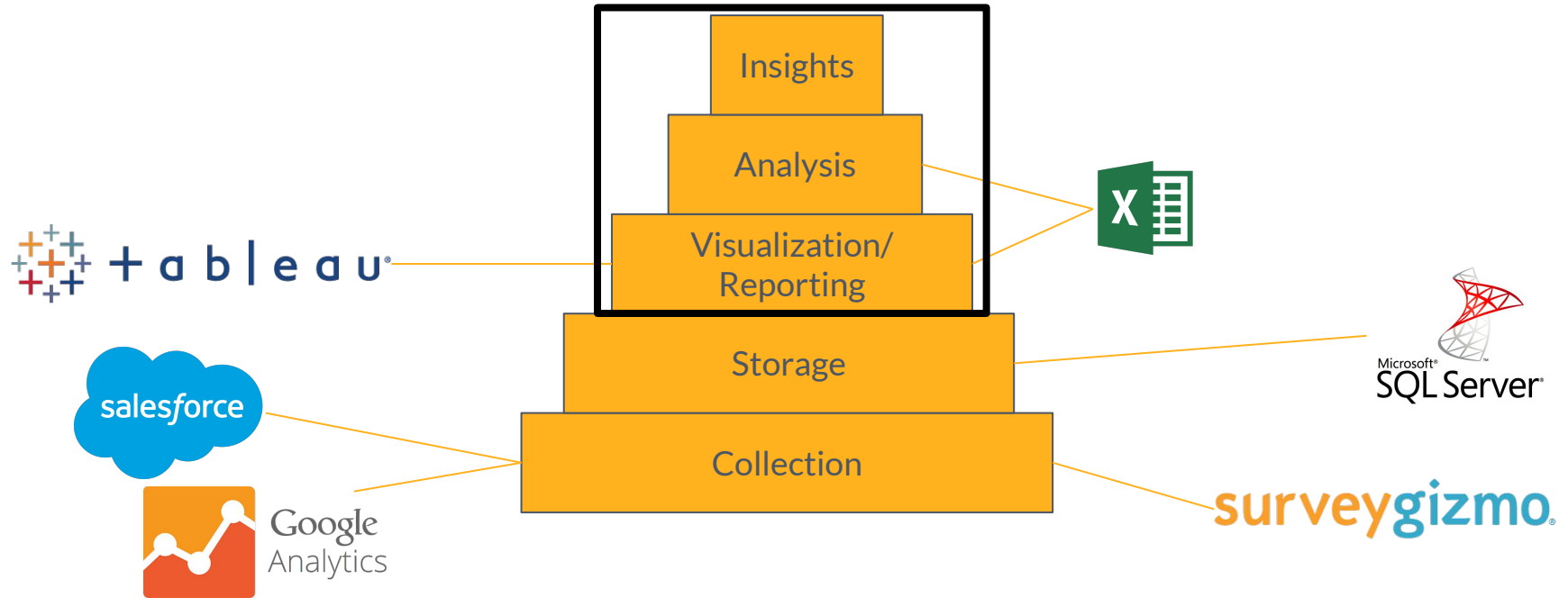


# This new infrastructure allowed us to automate some previously very difficult tasks

## Less Repetition!



# The Data Pyramid, circa 2014



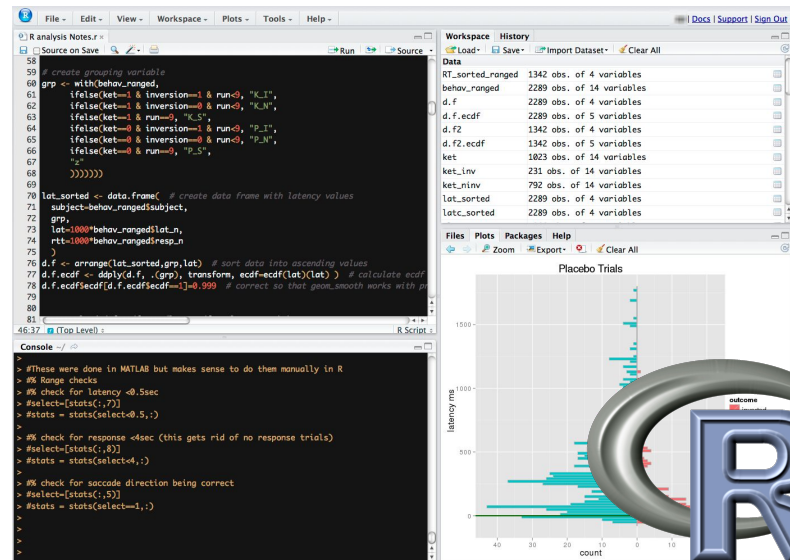
# Back end innovations vastly improved our analytical capacity

## 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

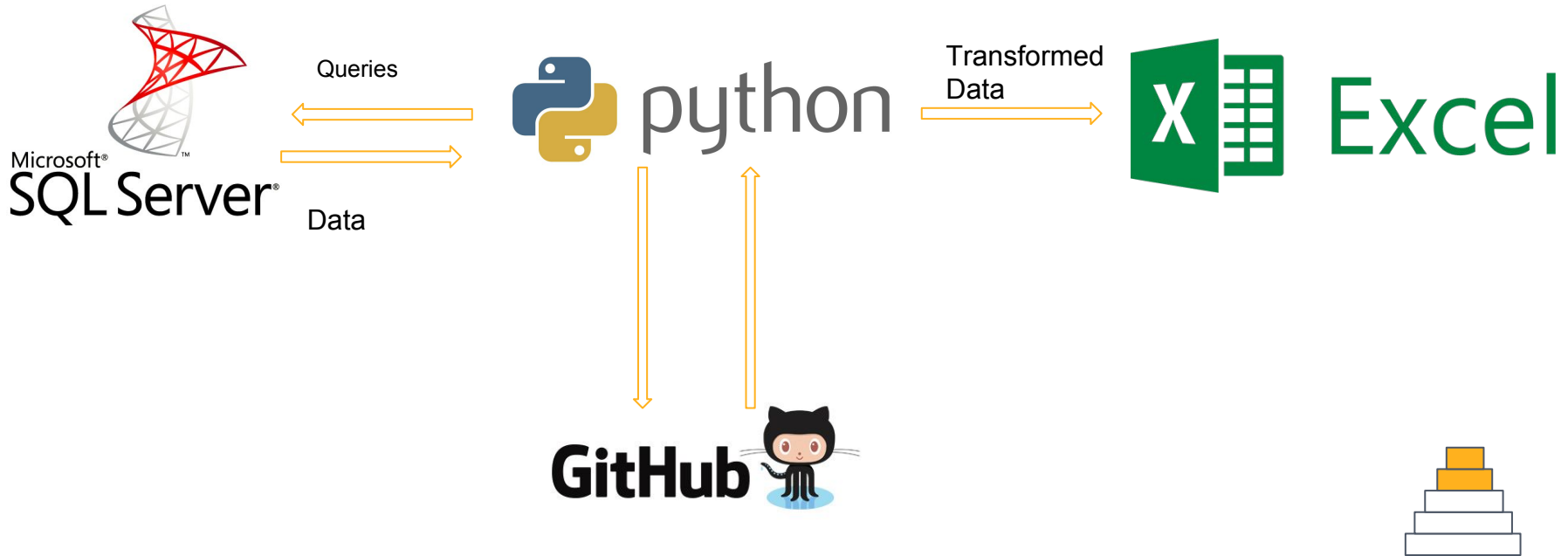
- Most popular first programming language
- General programming language, powerful/flexible
- Tons of packages

## Some disadvantages to Python

- Slower for “big” data
- Learning curve



## New Tools, same end product

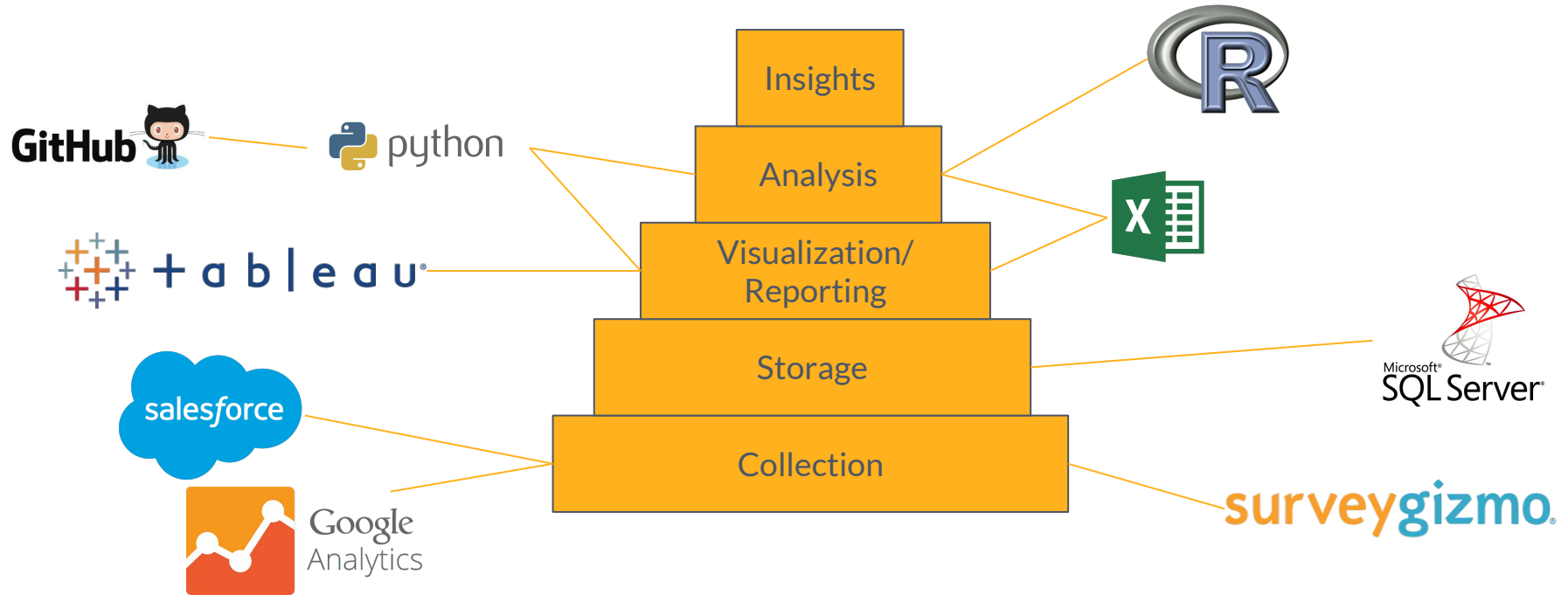


## Developing analytic human capital in house



Anyone can program, with curiosity and relevant application.

# The Data Pyramid, present day





# Always innovating - our path forward

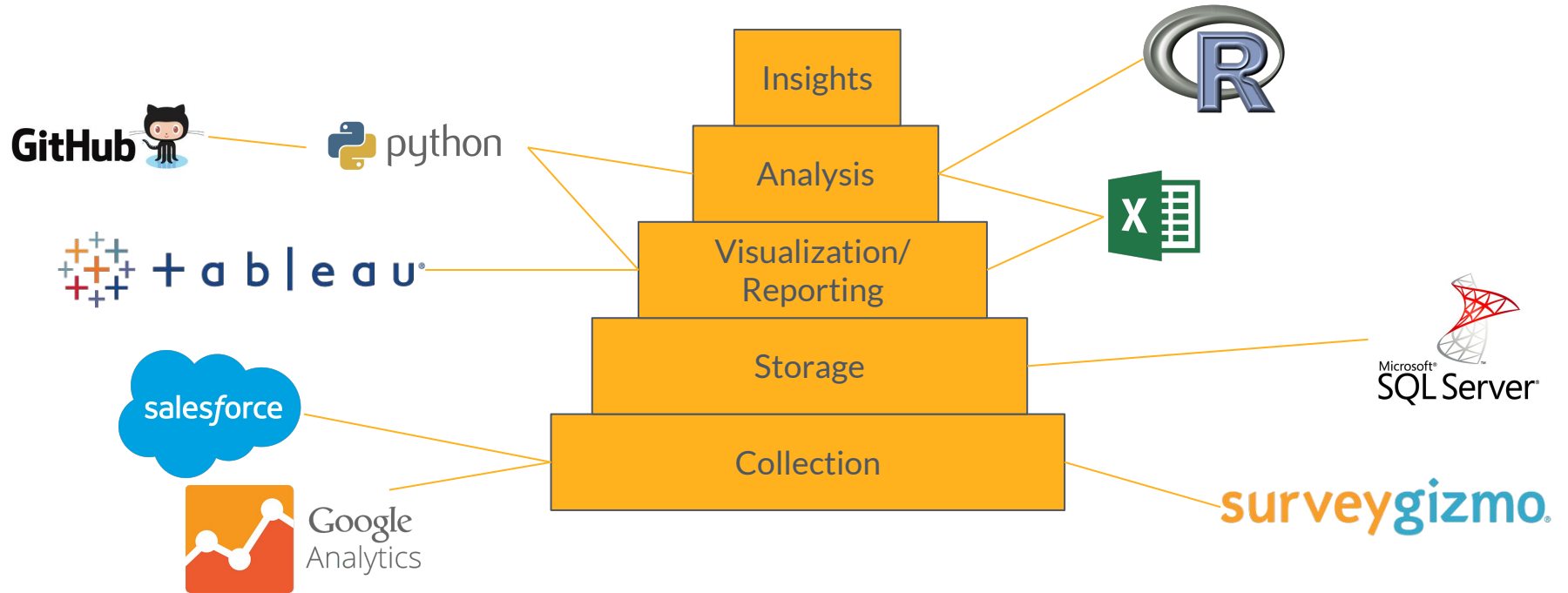
- Moving more datasets into our central database
- Finding ways to collect the new data we come in contact with
- Automating analysis for standard requests



# Match the Right Tool to each Problem

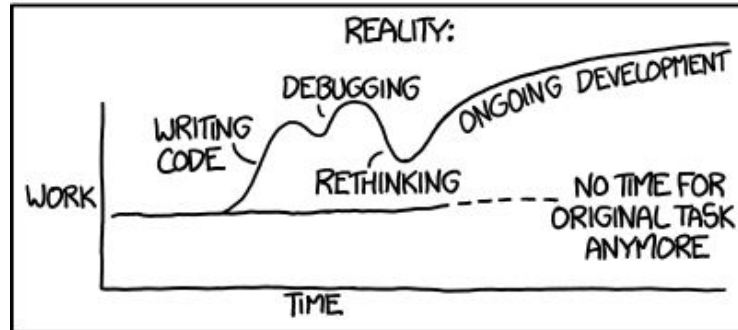
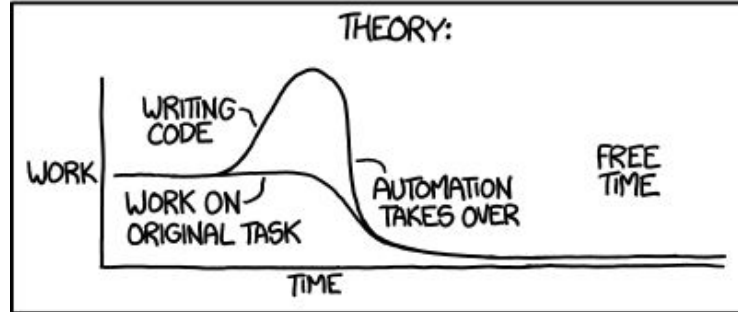
	Common Problems	Possible Tools
Collection	"I don't know if this data <b>is right</b> " "I know what data I want, I just <b>don't have it.</b> "	SurveyGizmo, Salesforce, Google Analytics
Storage	"It takes me <b>hours</b> to compile the data" "My data is <b>too big</b> for excel" "I spend days downloading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

It's a pyramid - start at the bottom.



## Change takes time.

"I SPEND A LOT OF TIME ON THIS TASK.  
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



With apologies to...



We're hiring!!

[www.achievementnetwork.org/careers](http://www.achievementnetwork.org/careers)

Alan: [azhao@achievementnetwork.org](mailto:azhao@achievementnetwork.org)

Jeremy: [jerjames13@gmail.com](mailto:jerjames13@gmail.com)