Project Yellow – Bringing Data Types and Functional Programming to Excel

Virtual Master Class – Andy Gordon Senior Principal Research Manager Microsoft Research

London Tech Centre – JPMorgan Chase April 28, 2020







Project Yellow

Excel as a Programming Language
A long-term partnership between Excel and MSR Cambridge

Presented by Andy Gordon, MSR Cambridge - adg@microsoft.com



Aim

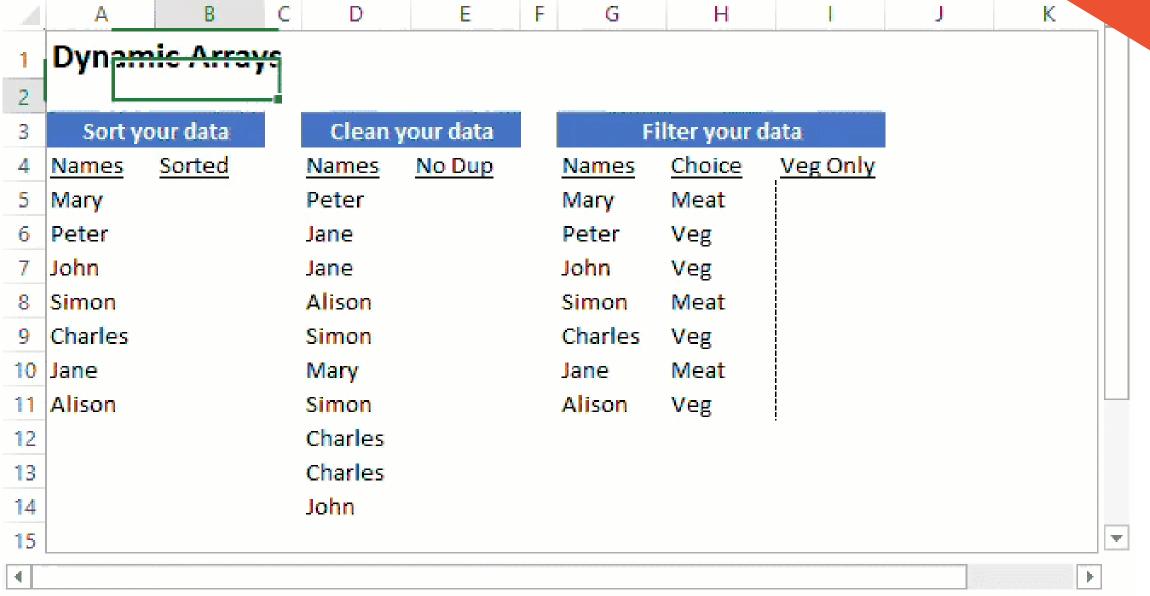
Remove the "glass ceiling" that limits the scope and reach of what a domain expert can do with Excel:

 Make Excel functions reflect the abstractions of our end users, by allowing end-users to define new functions using an ordinary worksheet.

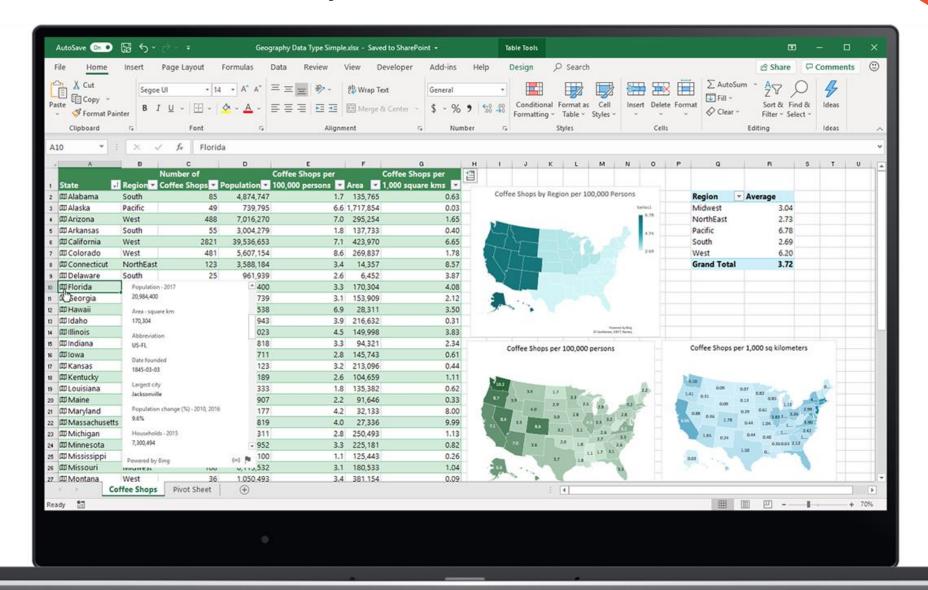
 Make Excel's data values reflect the datatypes of our end users' domains, by adding arrays, vectors, records, and even domainspecific data types implemented by third parties.

So far, we've taken Excel beyond text and numbers

19nite 2010



So far, we've taken Excel beyond text and numbers



り、 🖰 v 🔻 Nutrition -	Saved •	∠ Search				Entire
Insert Page Layout Form	nulas Data R	eview View	Help			Willes E
f_{κ}						0/5/2
С	D	E	F	G	Н	Entities fro
Gloria's Food	n Daily Goals					
#MondayMotivation Keeping up with my nutrition goals = #livingmybestlife	32.1%	24.4%	46.6%	0		
	Calories	Carbohydrates	Fat			
Today	641.4 cal	61.1g	23.3g			
Goal	2000	250	50			
My food 🖣						
food	calories 🔻	arbohydrates 🔻 f	fat 🔻			
Chicken breast	110.04	1.00	2.29			
🖔 boiled egg (1 serving)	210.80	1.52	14.43			
🖔 strawberry (4 items)	15.10	3.78	0.10			
Õ banana (1 item)	105.02	26.95	0.39			
oatmeal cookie (2 items)	134.10	19.92	5.37			
🖔 yogurt (3 ounces)	66.34	7.94	0.75			
ф.						



Starting today, the LET function is available to Excel users in the Insiders Ring to preview. The LET function allows you to define and use formula scoped names. Calculating a weighted average: =LET(values,A1:A4, weights,B1:B4, SUM(values*weights)/SUM(weights)) §

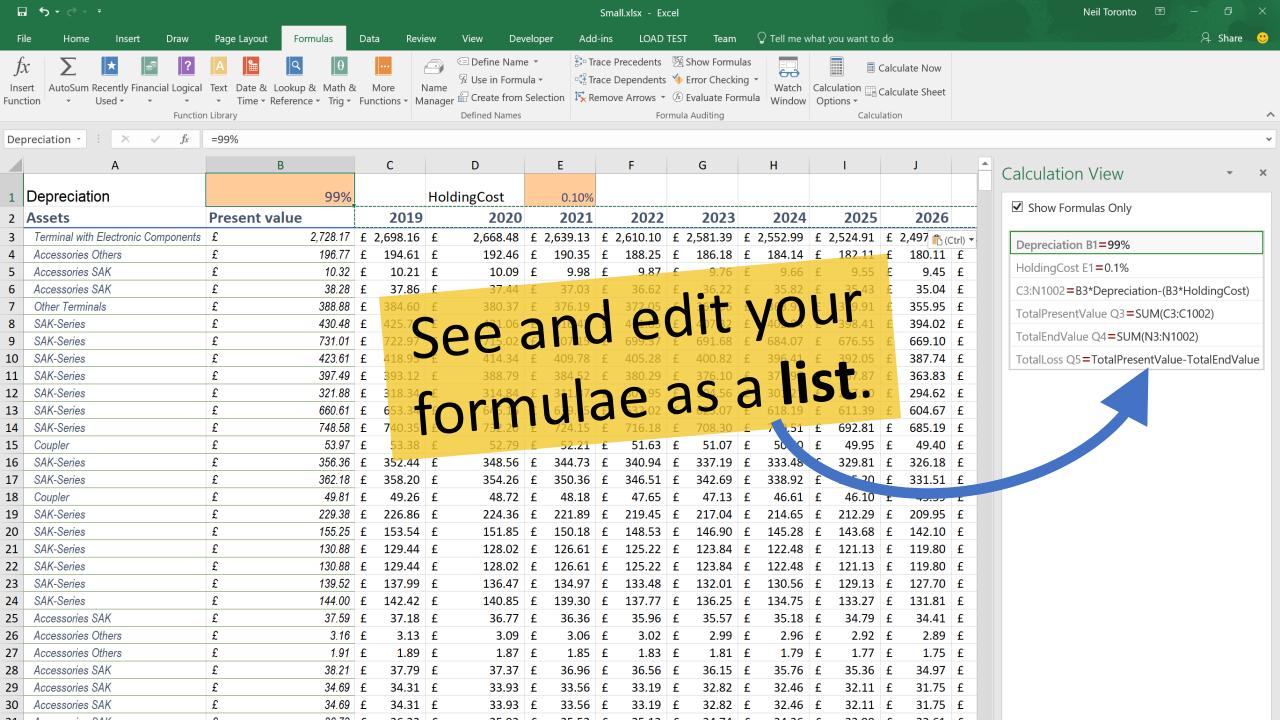
6:16 PM · Mar 17, 2020 · Twitter Web App

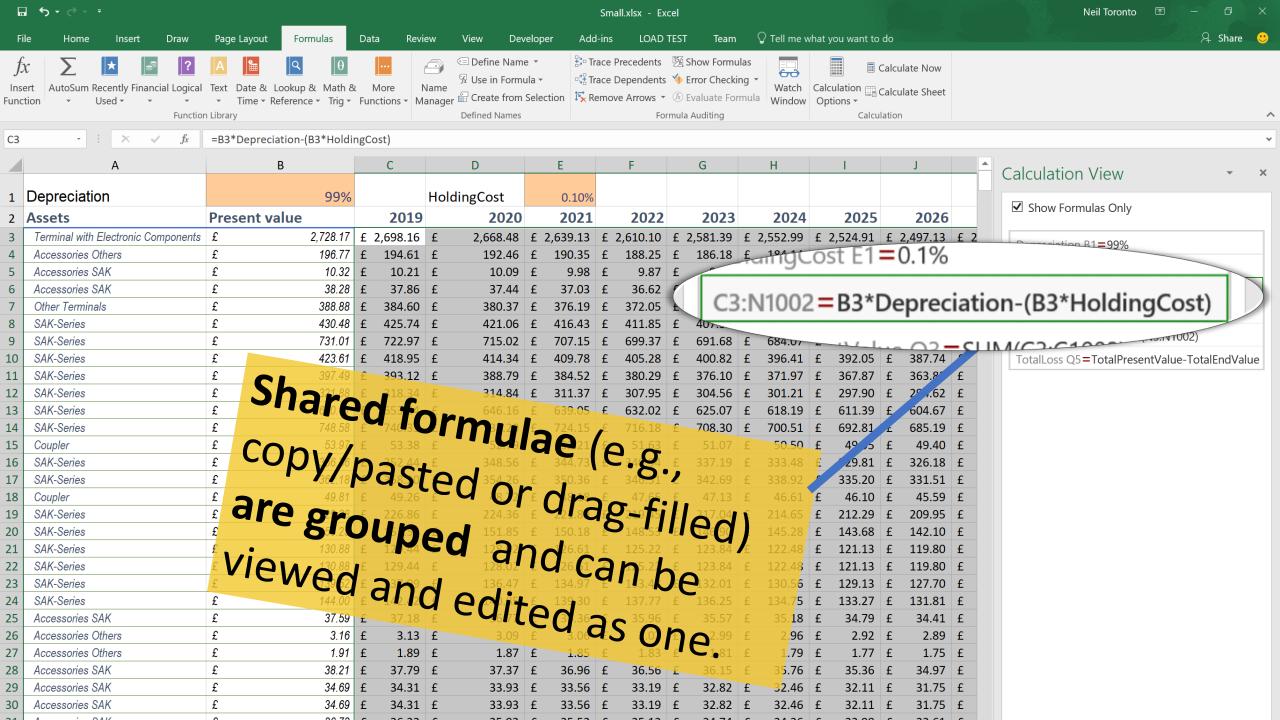
28 Retweets 93 Likes

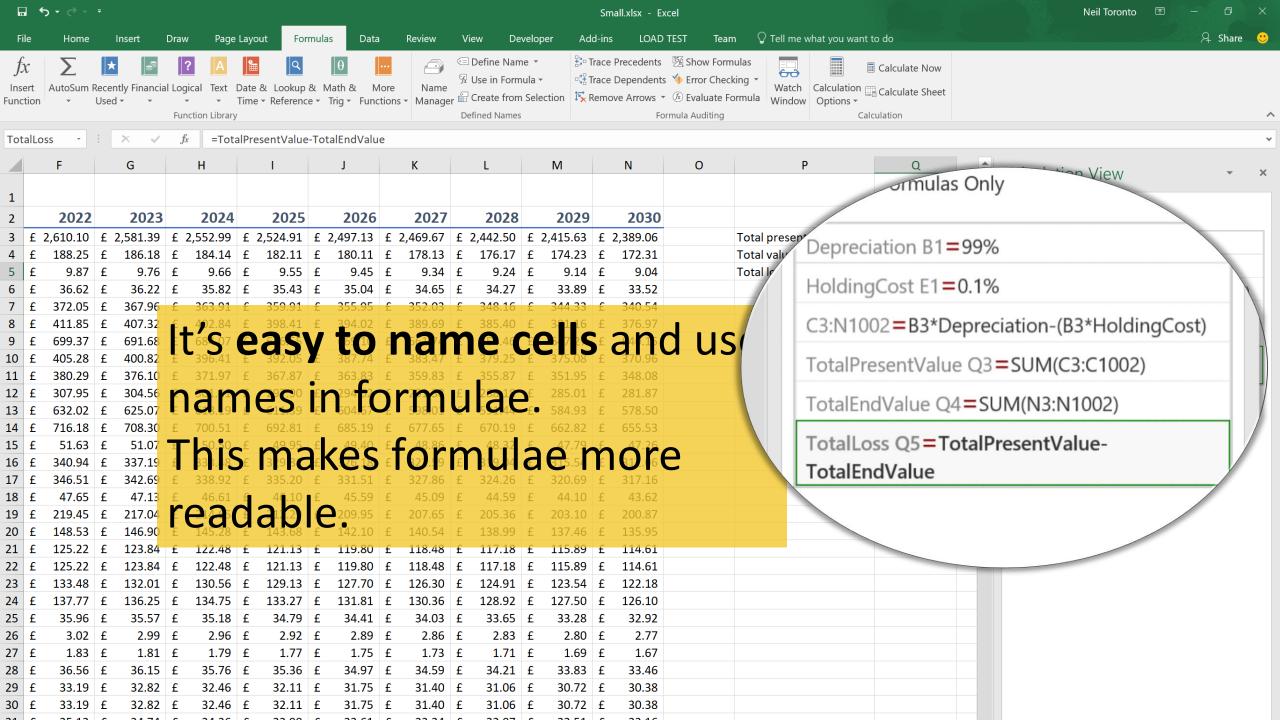


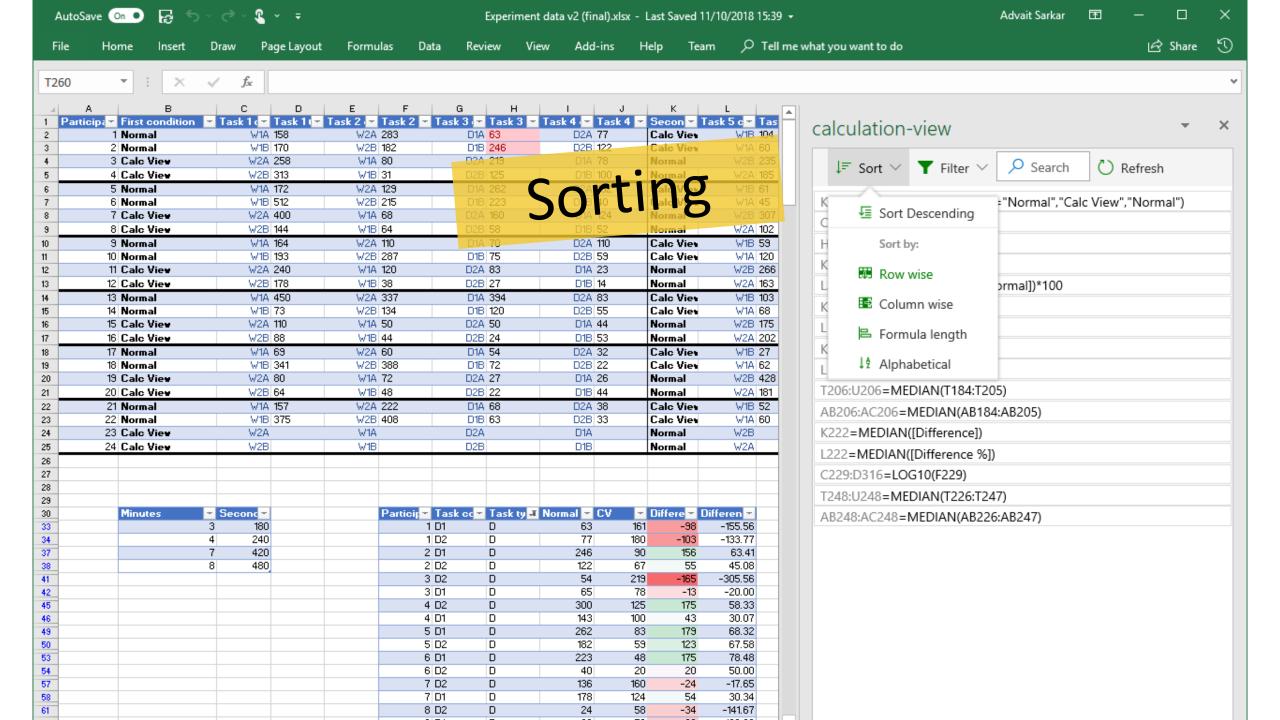
Calculation View

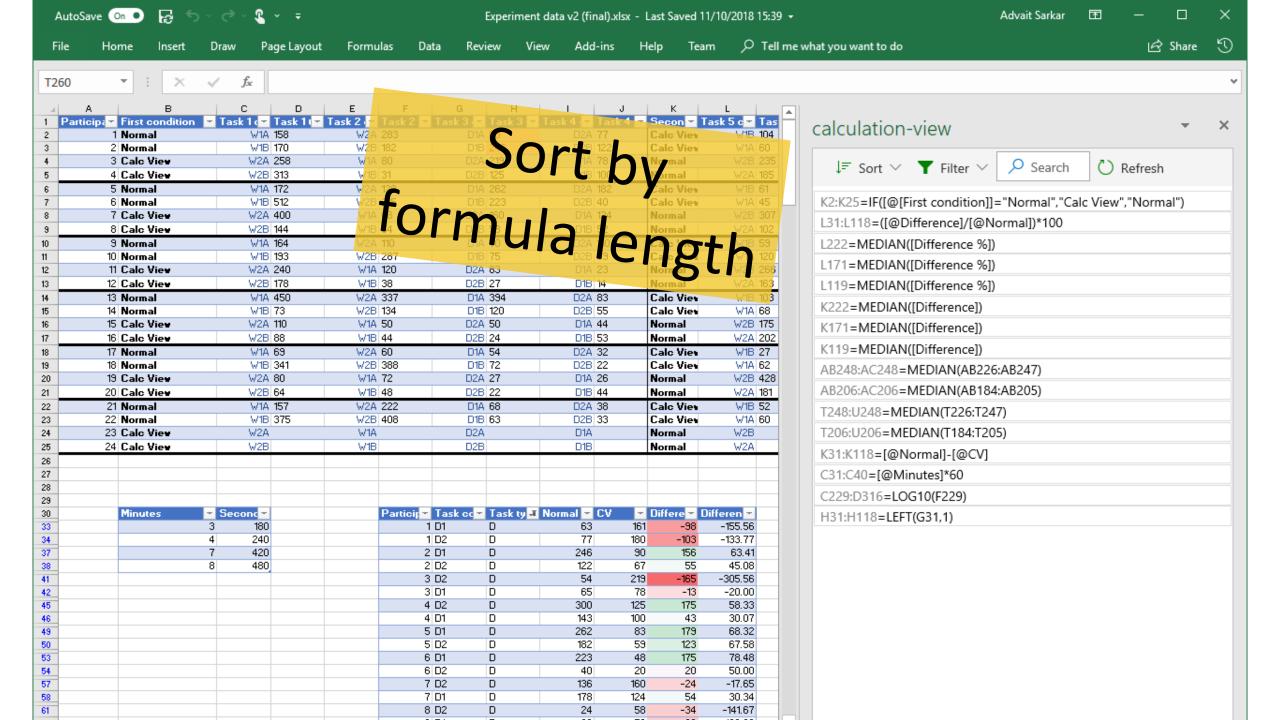
What if we offered a code-centric view of the grid?

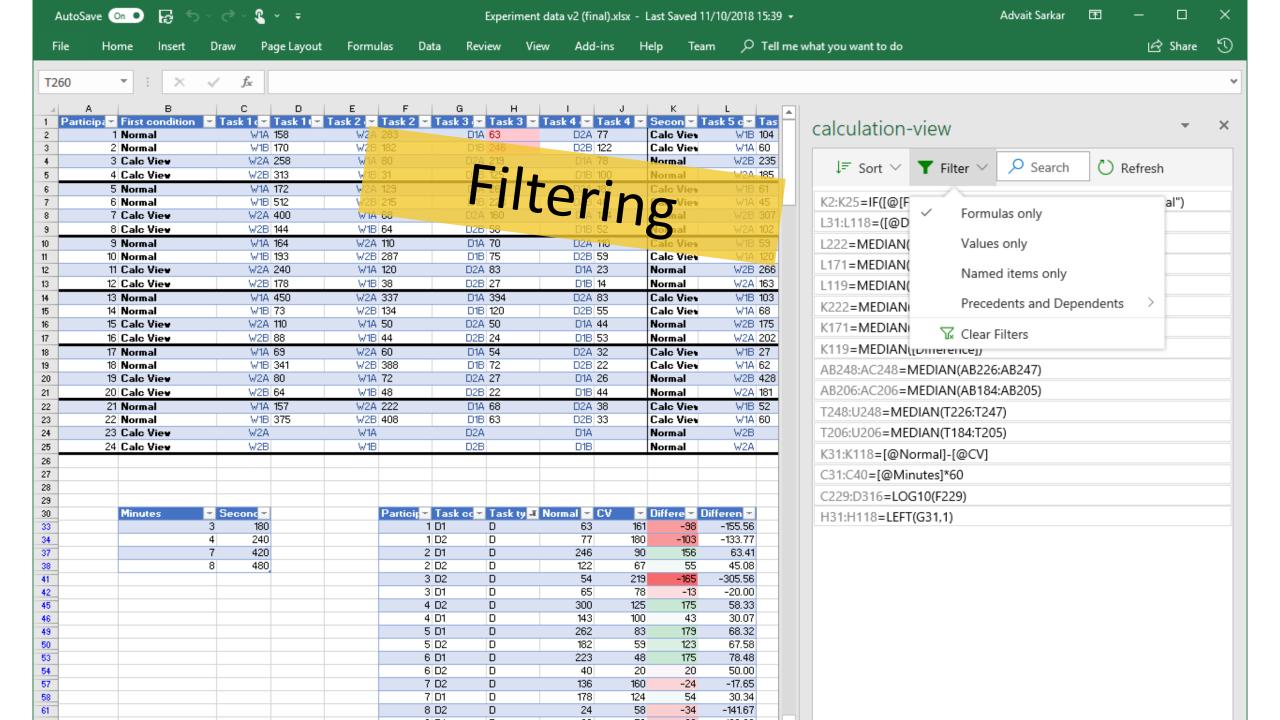


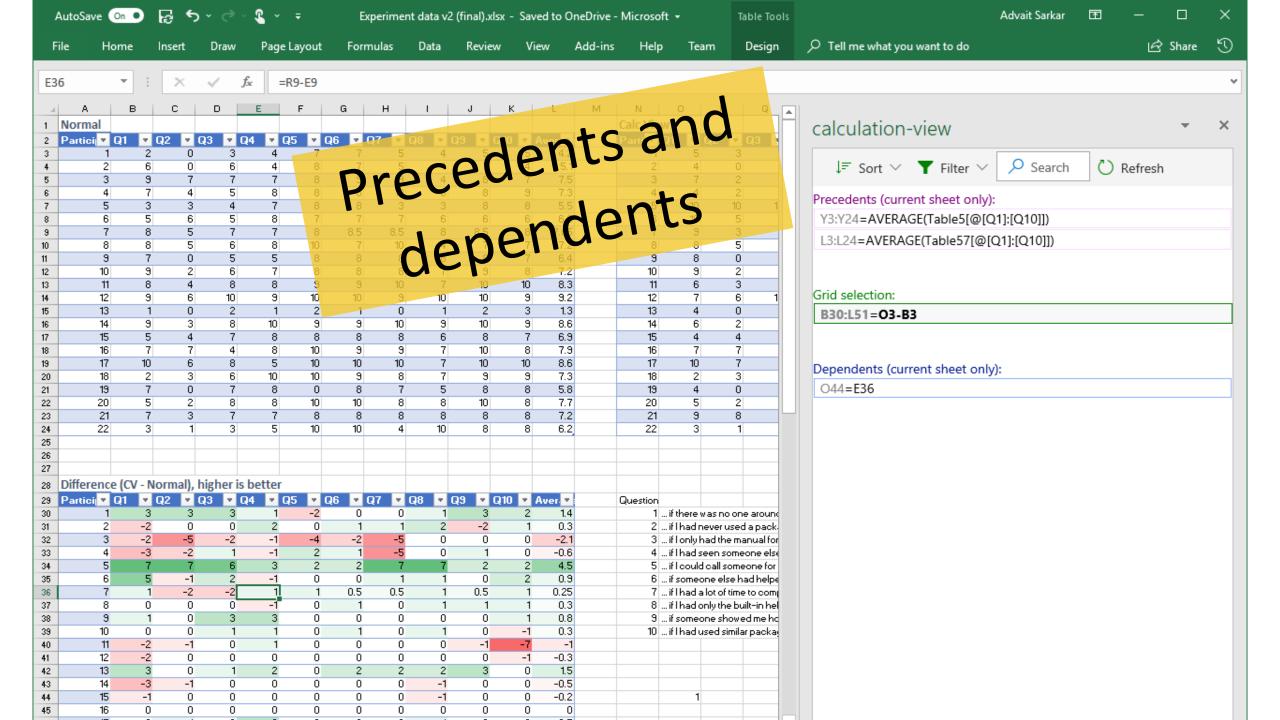


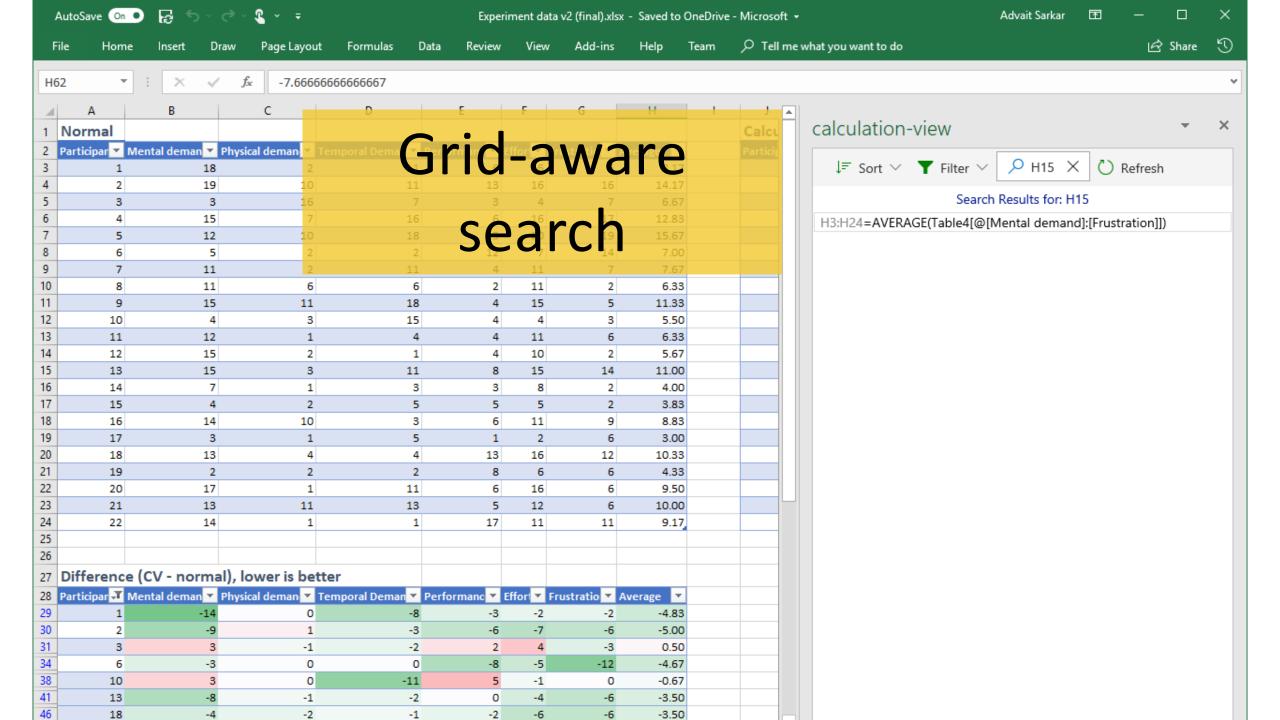










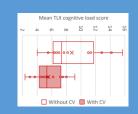


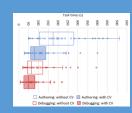
Studying calculation view



Authoring & debugging

- 20 participants, variety of backgrounds & expertise
- Improvement in task time and cognitive load







Comprehension

- 14 participants, spreadsheet comprehension test
- Improvement in comprehension and cognitive load in half the participants



Longitudinal deployment

- 7 MVPs, using CV for day to day work, ongoing reporting back
- Many uses, most positive feedback around auditing and comprehension

MVP feedback

Solves the problem that formula view is aimed at

• P6: "[The current show formulas feature] is useless to me. [...] One of the most particular things that accountants do is get column size exactly right. And I need to see what the numbers are. I'm going to say words I've never said before. I like the fact that [Calc View] is in a task pane. That makes me feel dirty inside. But I love that I don't have to manipulate the sheet to flip it into formula view. Don't have to flip the grid. You guys are absolutely on the right track."

Very helpful for auditing and debugging

- P17: "Calc View has already been extremely useful to me in identifying spurious formulae, no longer needed, and the situation where not all formulae in a row had been amended to the newer [...] version. Calc View is already a winner!!"
- P9: "Just plain text search is really good."
- P6: "Cool to have name, location and formula in one place. Formula evaluation tool is kind of weak. Calc View is a logical place to enable stepping from one thing to another. Maybe don't want to make calc view as rich as formula bar. Maybe just need a quick way to get to the formula bar."

Calculation View: multiple-representation editing in spreadsheets

Advait Sarkar*, Andrew D. Gordon*†, Simon Peyton Jones*, Neil Toronto*

*Microsoft Research, 21 Station Road, Cambridge, United Kingdom

†University of Edinburgh School of Informatics, 10 Crichton Street, Edinburgh, United Kingdom

{advait,adg,simonpj,netoront}@microsoft.com

- Results from over a year of testing, across three studies, with ~40 total participants from a diverse range of backgrounds.
- A powerful successor to 'show formulas'.

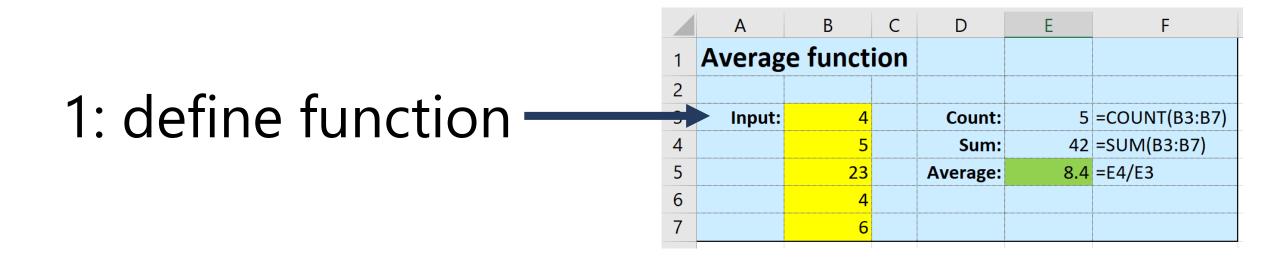
Calculation View helps people author, debug, and comprehend their spreadsheets.

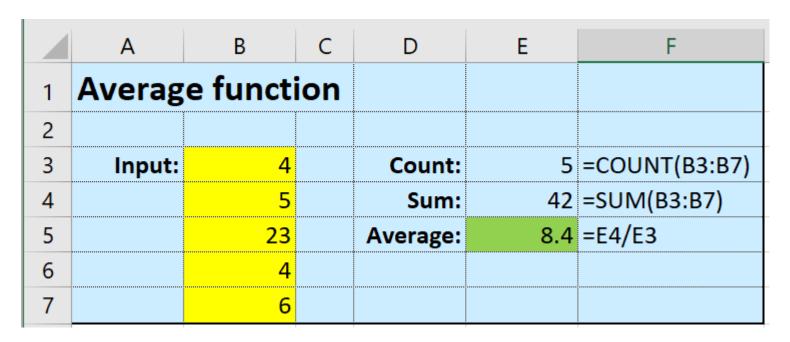
Elastic Sheet-Defined Functions

Generalising Spreadsheet Functions to Variable-Size Input Arrays

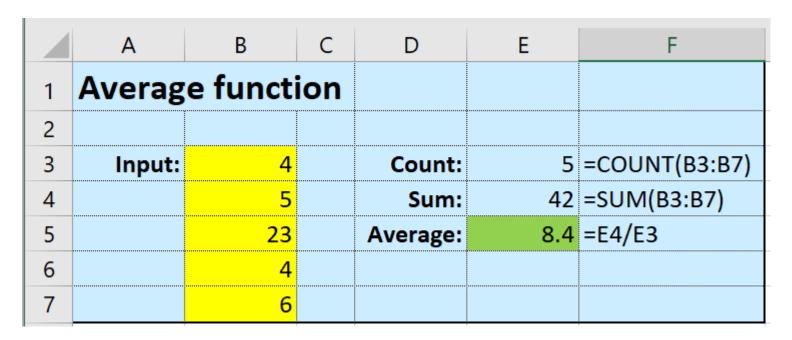
Sheet defined functions

Peyton Jones, S.L., Blackwell, A.F., Burnett, M.M.: A user-centred approach to functions in Excel. ICFP 165–176 (2003)





```
In CalcView:
function AVERAGE( B3:B7 ) returns E5 {
  E3 = COUNT( B3:B7 )
  E4 = SUM( B3:B7 )
  E5 = E4/E3
}
```



```
In CalcView:
function AVERAGE( B3:B7 ) returns E5 {
  E3 = COUNT( B3:B7 )
  E4 = SUM( B3:B7 )
  E5 = E4/E3
}
```

Problem:

- =AVERAGE(X5:X7) too small!
- =AVERAGE(G2:G200) too big!

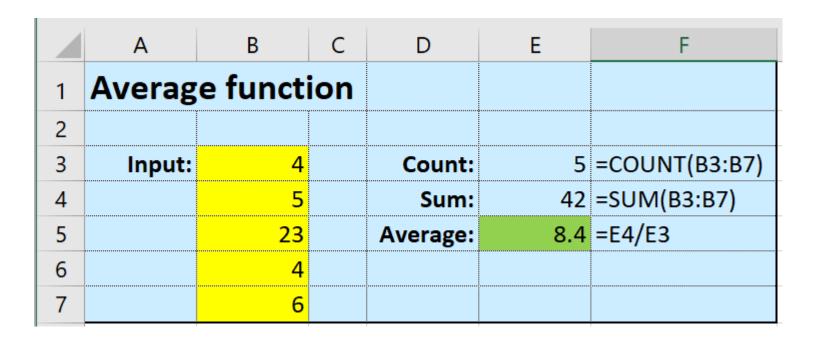
Solution:

 Somehow generalize what the user wrote

	Α	В	С	D	Е	F
1	Averag					
2						
3	Input:	4		Count:	5	=COUNT(B3:B7)
4		5		Sum:	42	=SUM(B3:B7)
5		23		Average:	8.4	=E4/E3
6		4				
7		6				

```
function AVERAGE( B3:B7 ) returns
E5 {
    E3 = COUNT( B3:B7 )
    E4 = SUM( B3:B7 )
    E5 = E4/E3
}
```

```
function AVERAGE < α > ( B3:B{3+α} ) returns
E5 {
    E3 = COUNT( B3:B{3+α} )
    E4 = SUM( B3:B{3+α} )
    E5 = E4/E3
}
```



```
function AVERAGE < α > (B3:B{3+α}) returns
E5 {
E3 = COUNT(B3:B{3+α})
E4 = SUM(B3:B{3+α})
E5 = E4/E3
}
```

Elastic SDF can be called with any $\alpha \ge 0$ =AVERAGE(G2:G200)

- Figure out $\alpha = 198$ from the arguments
- Instantiate the SDF sheet with $\alpha = 198$
- Calc it
- Return the result
- Discard the instantiated sheet

TIMESHEET														
First na	Last na	Catego	Rate	Date	01/04/2017	02/04/2017	03/04/2017	04/04/2017	05/04/2017	*****	******	Total hours	Payment	
John	Wire	Full-Time	20	1	0	· O	4	8	0	4	0	16	320	
Sophie	Gallaghe	Contract	25		0	0	0	0	8	8	4	20	500	
												Total to be pai	820	
ı J	1	1	1	1										

First na	Last na	Catego	Rate	Date	01/08/2017	02/08/2017	03/08/2017	04/08/2017	05/08/2017	*****	******	08/08/2017	******	*******	11/08/2017	*******	*******	14/08/2017	Total hours	Payment
David	Green	Full-Time	20		0	0	3	0	0	5	0	9	4	0	0	5	0	0		
Sarah	Jones	Full-Time	20		5	7	8	9	0	5	5	10	0	0	4	4	4	4		
Edward	Muller	Contract	25		0	0	0	0	0	8	10	0	5	0	0	8	0	0		
Hannah	Kirk	Part-Tim	15		8	0	0	4	0	8	4	0	3	5	3	0	8	8		
Jo	Carter	Contract	25		8	8	8	8	8	0	0	8	8	8	9	0	8	0		
Max	Wells	Full-Time	20		8	0	0	0	8	8	8	8	0	0	0	8	8	8		
																			Total to be p	3:S18))

 $=\mathcal{E}(PAYMENT(D13:D18,F13:S18))$

User's eye view

- · Write a function with fixed-size inputs, using familiar copy/paste
- Magic happens
- The function works on input of arbitrary size

Main point:

we think that automatically inferred elasticity will dramatically broaden the audience that can use SDFs effectively.

The magic

- · It really is quite amazing to
 - · Take a single, concrete function working on fixed size inputs
 - · And generalise it to arbitrary size inputs
 - · In a predictable way
- · But we can do it.
- · We can even prove that the generalisation we find is the "best generalisation" in a very precise sense. Not just a set of heuristics.

Findings from user study (N=20, 7 female)

People perceived **significantly lower cognitive workload** for elastic SDFs than with SDFs based on map/reduce.

I think elastic functions are easier to work with, also with the "mental model" that you have of Excel, because you can more just follow your normal Excel **workflow**. – P9

It'd be nice to have this kind of **middle ground**, of not having to write the same things over and over again, but not having to persuade someone to make a macro either. – P7

Elastic Sheet-Defined Functions

Automatic generalization of map/reduce programs from example

Theorem: Elasticization yields most general generalization

User study: Easier for end-users than map/reduce programs

Elastic Sheet-Defined Functions: Generalising Spreadsheet Functions to Variable-Size Input Arrays*

Matt McCutchen[†]
Massachusetts Institute of Technology
Cambridge, US
matt@mattmccutchen.net

Judith Borghouts[†]
University College London
London, UK
judith.borghouts.14@ucl.ac.uk

Andrew D. Gordon
Microsoft Research
Cambridge, UK
University of Edinburgh
Edinburgh, UK
adg@microsoft.com

Simon Peyton Jones Microsoft Research Cambridge, UK simonpj@microsoft.com Advait Sarkar Microsoft Research Cambridge, UK advait@microsoft.com

aka.ms/CalcIntel

Aim



Remove the "glass ceiling" that limits the scope and reach of what a domain expert can do with Excel:

 Make Excel functions reflect the abstractions of our end users, by allowing end-users to define new functions using an ordinary worksheet.

 Make Excel's data values reflect the datatypes of our end users' domains, by adding arrays, vectors, records, and even domainspecific data types implemented by third parties.

Yellow Related Links

aka.ms/CalcIntel

Dynamic Arrays and New Functions in Excel! - Office Insider Blog

Video tour of Microsoft Research Cambridge

(gif2xlsx: Convert GIFs to XLSX format)

