spreadsheets

@JennyBryan

@jennybc

@STAT545

stat545-ubc.github.io

Talk given at

`csv, conf, v2:`
`http://csvconf.com`

2016 May 03/04

Berlin

links, credits, and slides available here

https://github.com/jennybc/2016-05_csvconf-spreadsheets

csv,conf,v1

A conference for data makers everywhere

Berlin, 15 July, 2014

Felienne Hermans

Spreadsheets Are Code

back in Vancouver, I was thinking …

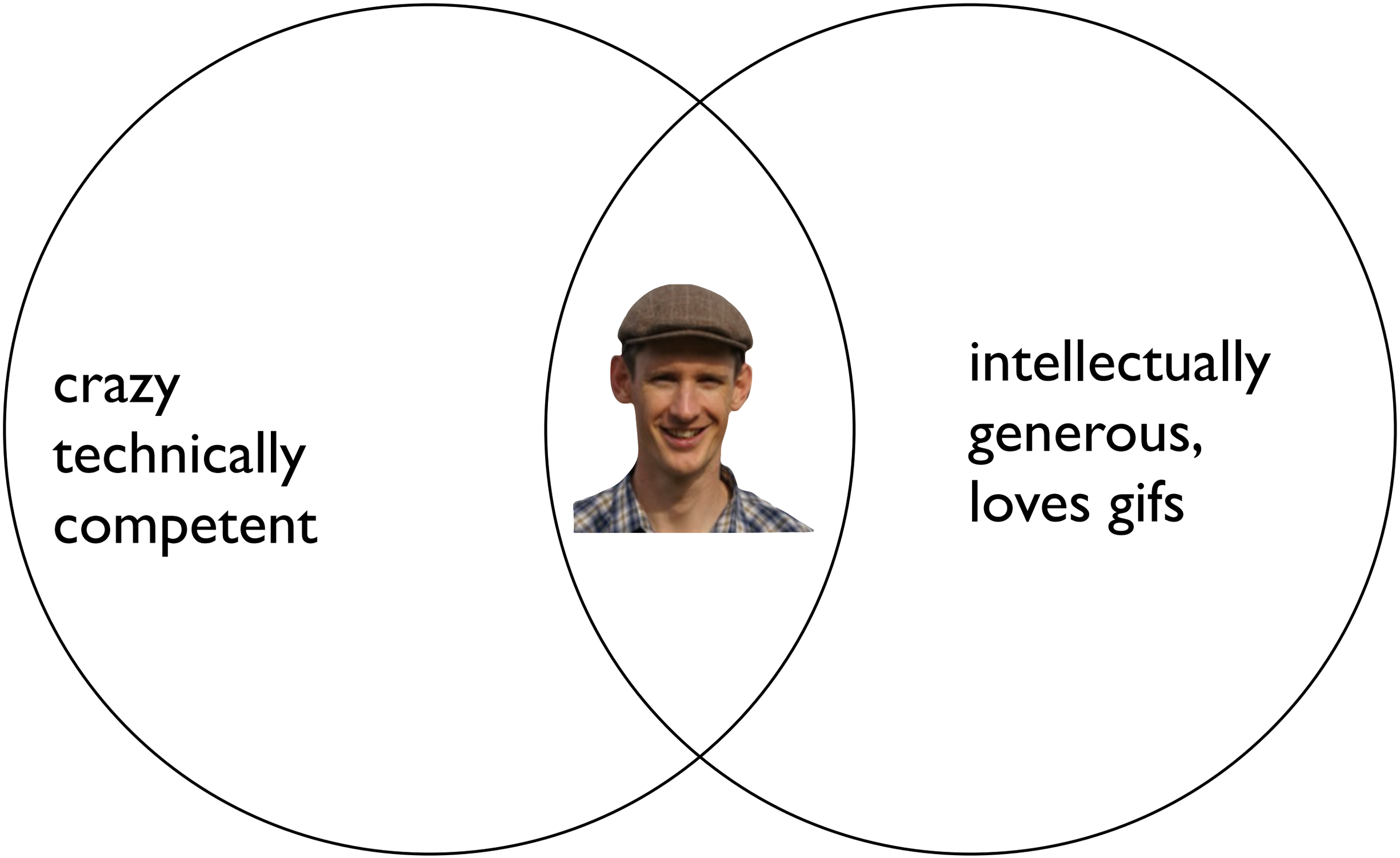
it's ok to care about the mundane* aspects of data!

it's even ok to care about spreadsheets!

these are my people!

* ha!

how to pick people to work with



Rich Fitzjohn web GitHub twitter

jiffy

https://github.com/richfitz/jiffy

build passing

build passing

It's pronounced "yiffy".

Interface to giphy via their API.

Interface to 'Giphy'

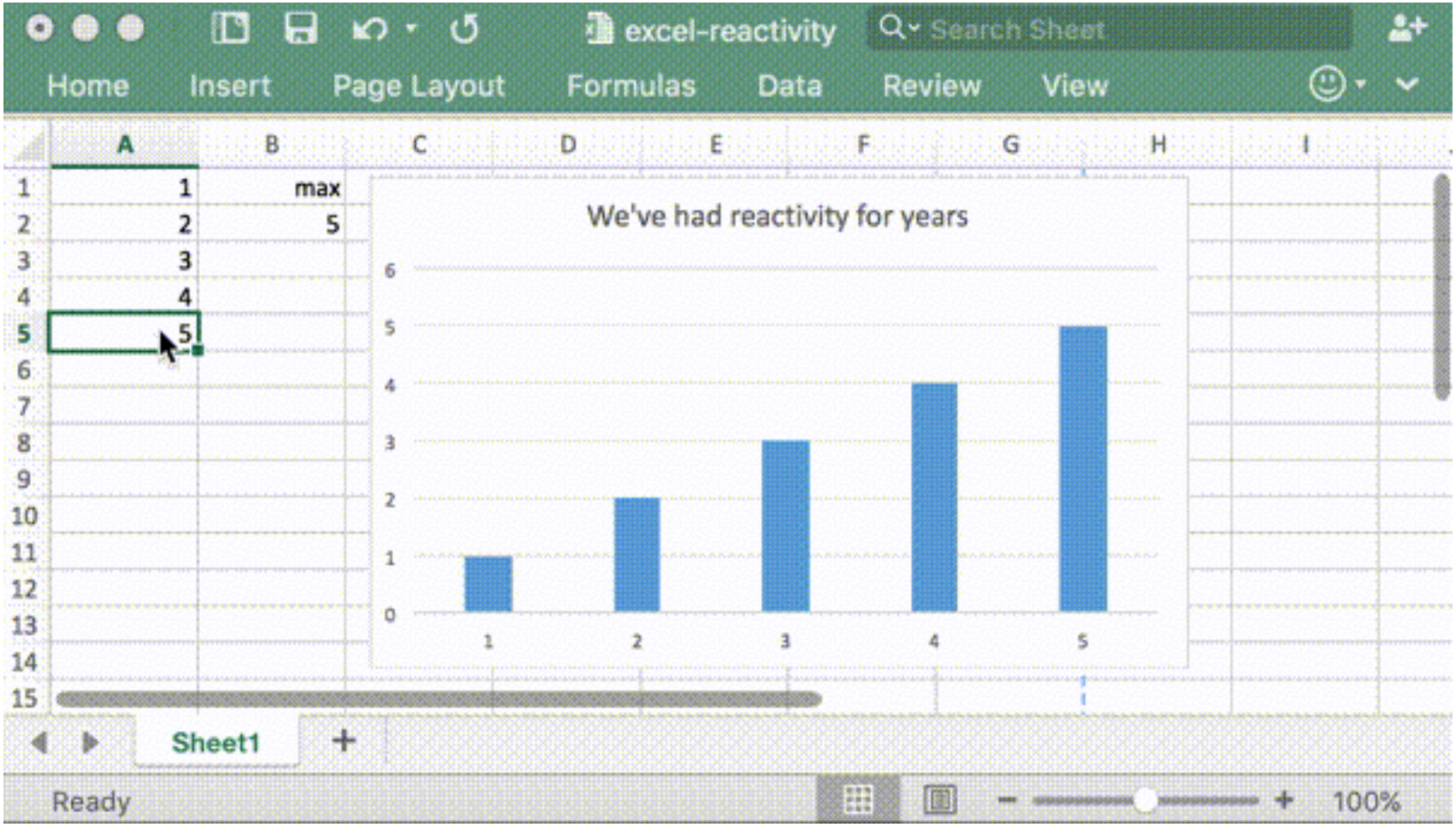
some of my best friends use spreadsheets

I supported myself for ~4 years doing spreadsheets

Tomas Petricek
@tomaspetricek

Follow

Ouch. "50 million accountants use monads in Excel. They just don't go around explaining monads to everyone..." @Felienne #ndclondon


excel-reactivity
Search Sheet
Home
Insert
Page Layout
Formulas
Data
Review
View
A
B
C
D
E
F
G
H
I
max
5
We've had reactivity for years
0
1
2
3
4
5
6
Sheet1
Ready
100%

syntax bullsh!ttery

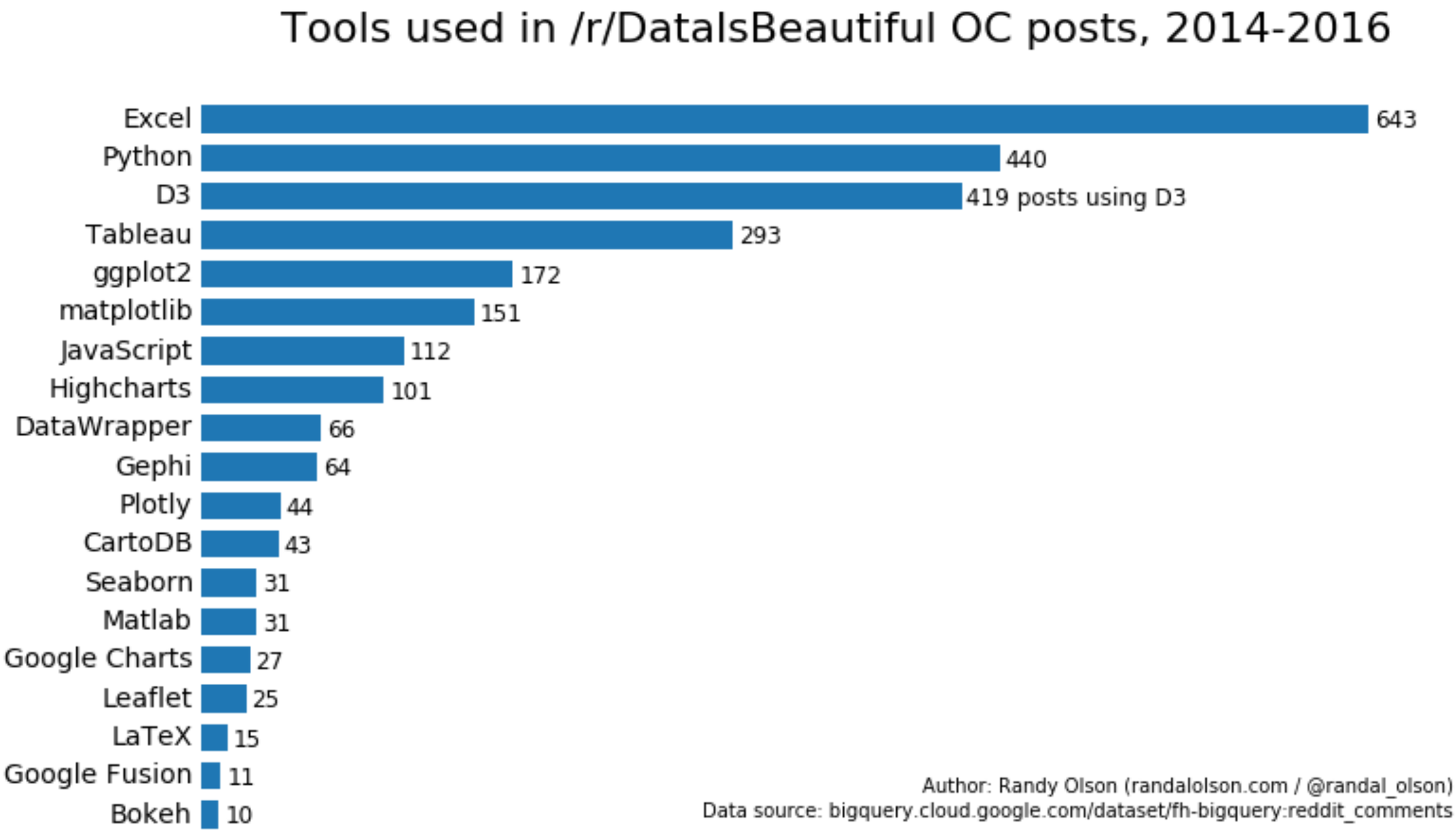
About 1 billion people use Microsoft Office

About 650 million use spreadsheets

Up to half use formulas in spreadsheets

250K - 1 million people use R

1 - 5 million people use Python


Tools used in /r/DataIsBeautiful OC posts, 2014-2016
Excel 643
Python 440
D3 419 posts using D3
Tableau 293
ggplot2 172
matplotlib 151
JavaScript 112
Highcharts 101
DataWrapper 66
Gephi 64
Plotly 44
CartoDB 43
Seaborn 31
Matlab 31
Google Charts 27
Leaflet 25
LaTeX 15
Google Fusion 11
Bokeh 10
Author: Randy Olson (randalolson.com / @randal_olson)
Data source: bigquery.cloud.google.com/dataset/fh-bigquery:reddit_comments

you go into data analysis
with the tools you know,
not the tools you need

Jenny Bryan @JennyBryan · Apr 20

I'm seeking TRUE, crazy spreadsheet stories. Happy to get the actual sheet or just a description of the crazy. Also: I can keep a secret.

Based on surveys made by members of the Staff of Indiana University and supported by the National Research Council with Rockefeller Foundation funds

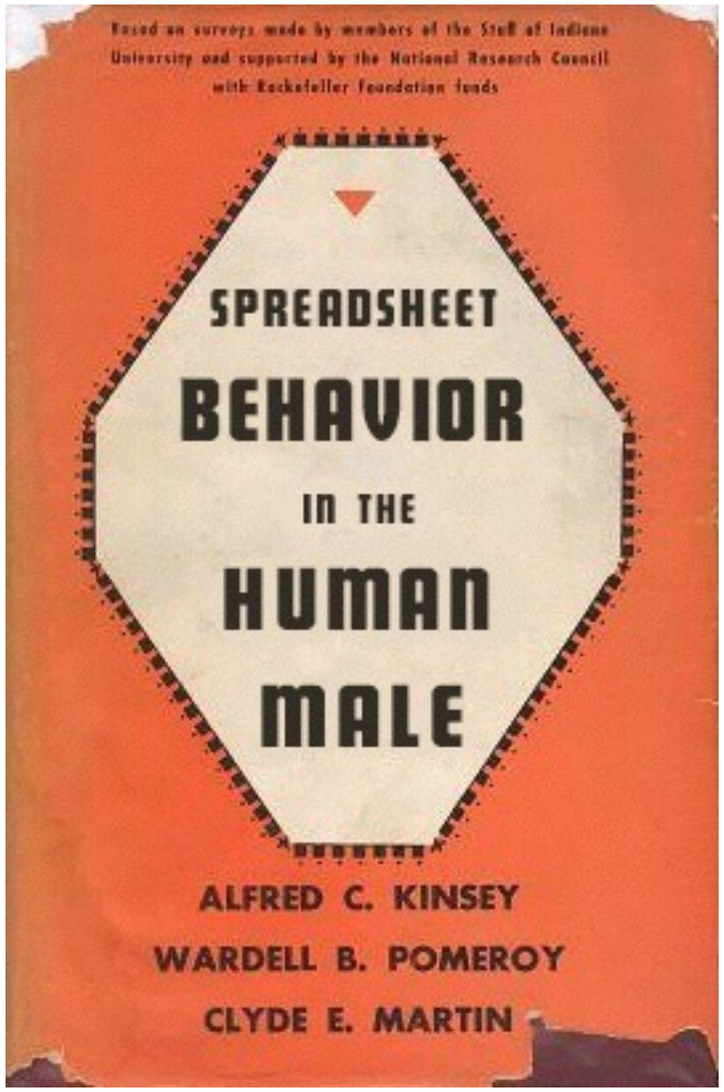
SEXUAL BEHAVIOR IN THE HUMAN MALE

ALFRED C. KINSEY
WARDELL B. POMEROY
CLYDE E. MARTIN

Based on surveys made by members of the Staff of Indiana University and supported by the National Research Council with Rockefeller Foundation funds

SPREADSHEET BEHAVIOR IN THE HUMAN MALE

ALFRED C. KINSEY
WARDELL B. POMEROY
CLYDE E. MARTIN



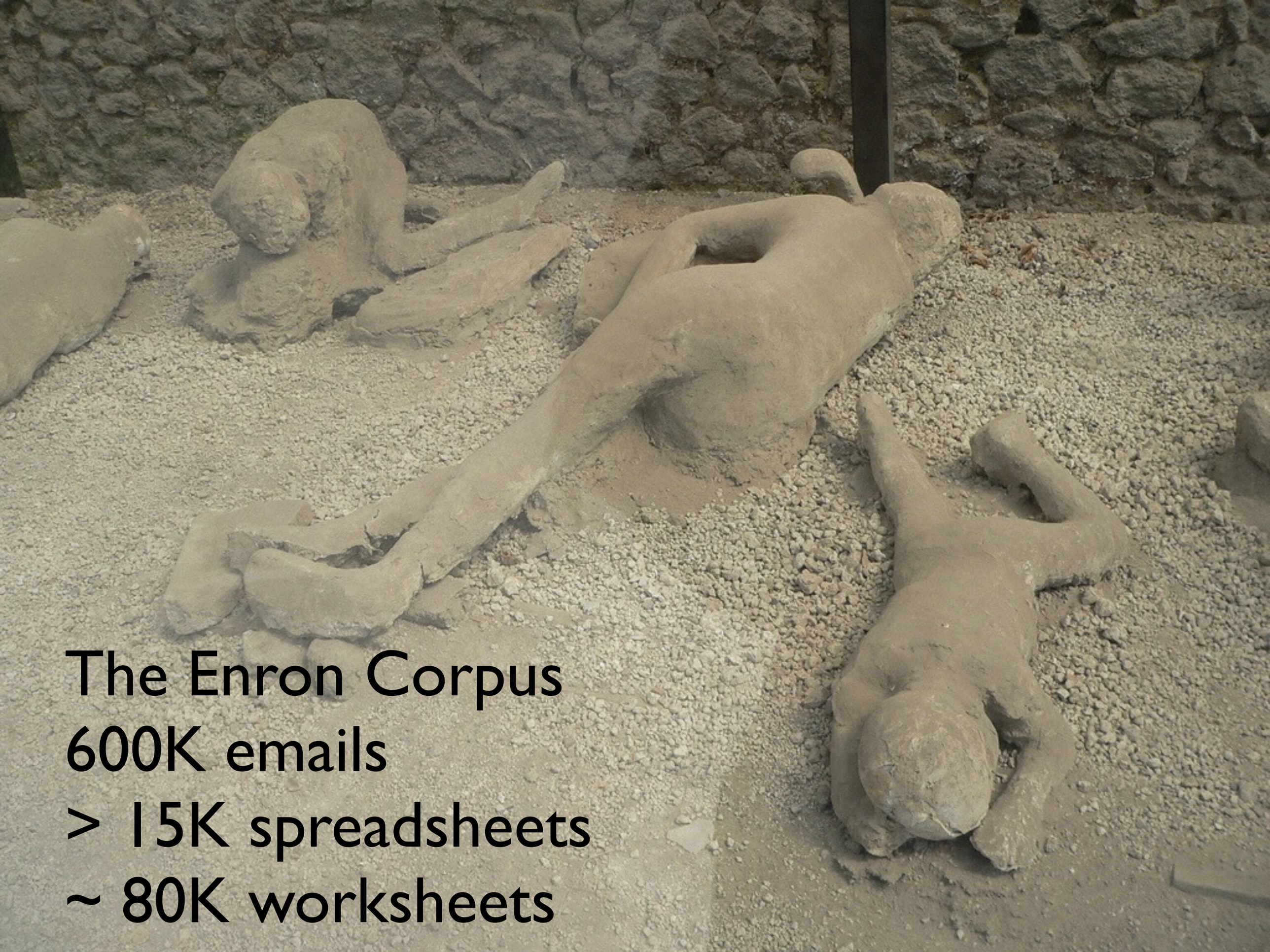
what you THINK people are doing

!=

what you think people SHOULD be doing

!=

what people ARE ACTUALLY doing



The Enron Corpus
600K emails
> 15K spreadsheets
~ 80K worksheets

Enron North America - West Gas

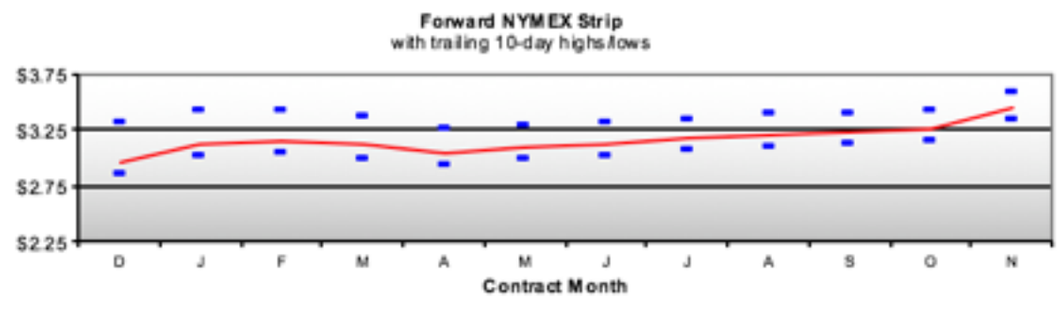
November 9, 2001

ENA - West Gas Contacts

Houston Office		Regional Offices		
Barry Tycholiz	(713) 853-1587	Mark Whitt	(303) 575-6473	Denver
Kim Ward	(713) 853-0685	Paul Lucci	(303) 575-6474	Denver
Stephanie Miller	(713) 853-1688	Tyrell Harrison	(303) 575-6478	Denver
Philip Polsky	(713) 853-5181	Dave Fuller	(503) 464-3732	Portland

Forward Prices (US$/MMBtu)

	NYMEX	
	SETTLE	Δ
Cash		
ROM		
Dec-01	2.960	0.090
Dec-01 to Mar-02	3.088	0.083
Apr-02 to Oct-02	3.166	0.084
Nov-02 to Mar-03	3.651	0.090
One Year Strip*	3.165	0.084



IF NWPL Rocky Mountains			
Fixed Price		Basis	
BID	OFFER	BID	OFFER
1.890	1.910		
2.060	2.080		
2.395	2.415	(0.565)	(0.545)
2.594	2.614	(0.494)	(0.474)
2.581	2.601	(0.585)	(0.565)
3.356	3.376	(0.295)	(0.275)
2.634	2.654	(0.530)	(0.510)

	IF CIG Rocky Mountains				IF EL Paso Permian				IF EL Paso San Juan			
	Fixed Price		Basis		Fixed Price		Basis		Fixed Price		Basis	
	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER
Cash	1.940	1.960			2.375	2.395			2.450	2.470		
ROM	1.960	1.980			2.420	2.440			2.350	2.370		
Dec-01	2.345	2.365	(0.615)	(0.595)	2.700	2.720	(0.260)	(0.240)	2.560	2.580	(0.400)	(0.380)
Dec-01 to Mar-02	2.548	2.568	(0.540)	(0.520)	2.855	2.875	(0.233)	(0.213)	2.743	2.763	(0.345)	(0.325)
Apr-02 to Oct-02	2.471	2.491	(0.695)	(0.675)	3.009	3.029	(0.158)	(0.138)	2.801	2.821	(0.365)	(0.345)
Nov-02 to Mar-03	3.311	3.331	(0.340)	(0.320)	3.499	3.519	(0.153)	(0.133)	3.421	3.441	(0.230)	(0.210)
One Year Strip*	2.551	2.571	(0.614)	(0.594)	2.982	3.002	(0.182)	(0.162)	2.817	2.837	(0.347)	(0.327)

	AECO / NIT				IF NWPL Canadian Border (Sumas)				IF PEPL TX-OK			
	Fixed Price		Basis		Fixed Price		Basis		Fixed Price		Basis	
	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER
Cash	2.376	2.396			2.480	2.500			2.530	2.550		
ROM	2.398	2.418			2.460	2.480			2.530	2.550		
Dec-01	2.552	2.572	(0.408)	(0.388)	2.800	2.820	(0.160)	(0.140)	2.828	2.848	(0.133)	(0.113)
Dec-01 to Mar-02	2.616	2.636	(0.472)	(0.452)	2.892	2.912	(0.196)	(0.176)	2.958	2.978	(0.130)	(0.110)
Apr-02 to Oct-02	2.661	2.681	(0.505)	(0.485)	2.796	2.816	(0.370)	(0.350)	3.046	3.066	(0.120)	(0.100)
Nov-02 to Mar-03	3.216	3.236	(0.435)	(0.415)	3.706	3.726	0.055	0.075	3.531	3.551	(0.120)	(0.100)
One Year Strip*	2.676	2.696	(0.488)	(0.468)	2.880	2.900	(0.285)	(0.265)	3.041	3.061	(0.123)	(0.103)

2001 NFL POOL
WEEK FIFTEEN

VISITOR		HOME	POINTS
Saturday 12/15/01			
Arizona	at	NY Giants	6
Oakland	at	San Diego	14
Sunday 12/16/01			
Atlanta	at	Indianapolis	10
Cincinnati	at	NY Jets	15
Denver	at	Kansas City	2
Jacksonville	at	Cleveland	3
Minnesota	at	Detroit	4
New England	at	Buffalo	13
Philadelphia	at	Washington	7
Tampa Bay	at	Chicago	5
Miami	at	San Francisco	11
Dallas	at	Seattle	12
Green Bay	at	Tennessee	8
Pittsburgh	at	Baltimore	1
Monday 12/17/01			
St. Louis	at	New Orleans	9
			120

OPEN:
Carolina

Submit picks by 11:00 AM Saturday
15 Games: Rank games 1 to 15 (Total Points = 120)

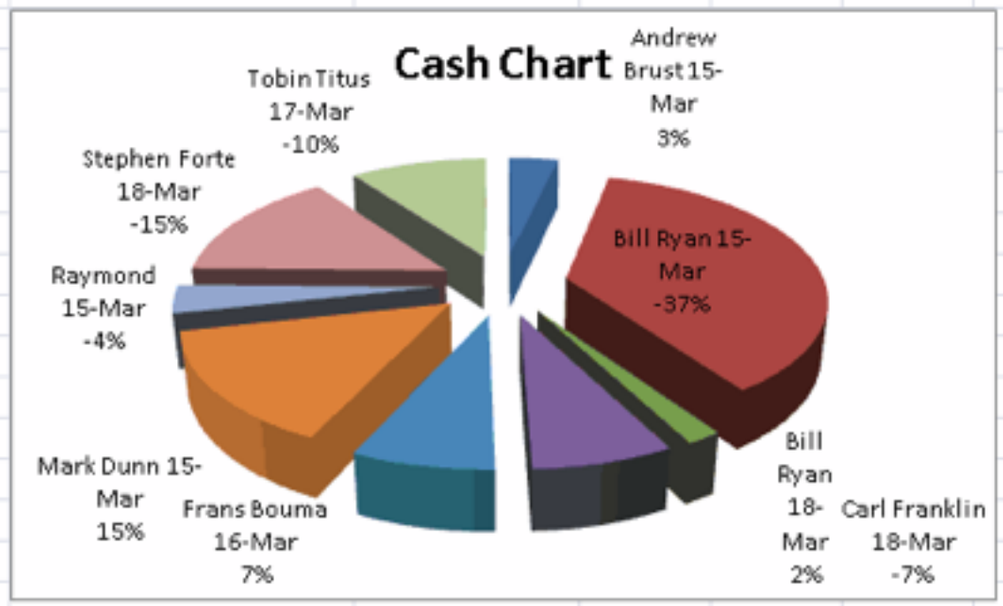
	EUSES	Enron
Number of spreadsheets analyzed	4,447	15,770
Number of spreadsheets with formulas	1,961	9,120
Number of worksheets	16,853	79,983
Maximum number of worksheets	106	175
Number of non-empty cells	8,209,095	97,636,511
Average number of non-empty cells per spreadsheet	1,846	6,191
Number of formulas	730,186	20,277,835
Average of formulas per spreadsheet with formulas	372	2,223
Number of unique formulas	65,143	913,472
Number of unique formulas per spreadsheet with formulas	33	100

from Hermans, Murphy-Hill

CashExchange.xlsx - Microsoft Excel

Person	Date	Amount
Andrew Brust	15-Mar	$10.00
Bill Ryan	15-Mar	-$100.00
Bill Ryan	18-Mar	$5.00
Carl Franklin	18-Mar	-$20.00
Frans Bouma	16-Mar	$20.00
Mark Dunn	15-Mar	$40.00
Raymond	15-Mar	-$10.00
Stephen Forte	18-Mar	-$40.00
Tobin Titus	17-Mar	-$27.38

Total CashFlow: -$122.38



oldpcnail2	peak force (kN/m^2)		developed force (kN		time to pf (ms)		time to relax (ms)	
Time (m	Control	AAPH	Control	AAPH	Control	AAPH	Control	AAPH
0	15.51	N/A	15.34	N/A	362	N/A	251	N/A
5	15.79	11.65	12.53	10.25	346	335	244	278
8	15.15	13.34	12.02	11.77	343	349	239	278
12	14.47	14.77	10.72	11.3	358	353	280	298
18	12.06	15.77	8.954	11.34	337	371	313	317
24	11.57	16.33	6.679	11.7	334.4	369	324.1	317
40	8.429	16.34	1.239	12.02	319.1	413	385.1	336
50	6.465	15.78	-2.161	13.97	309.5	429	423.3	349

	net work (mJ/g)		Shortening work (mJ/g)	
Time (m	Control	AAPH	Control	AAPH
10	0.225	0.215	0.555	0.823
15	0.151	0.207	0.856	0.936
20	0.167	0.195	0.734	1.00
30	0.105	0.205	0.638	0.989
45	0.007	0.196	0.395	0.559

oldpcnail2	peak force (kN/m^2)		developed force (kN		time to pf (ms)		time to relax (ms)	
Time (m	Control	AAPH	Control	AAPH	Control	AAPH	Control	AAPH
0	9.371	N/A	8.965	N/A	429	N/A	545	N/A
5	6.764	3.239	4.924	2.549	357	341	569	558
8	5.66	3.446	4.318	2.772	344	353	497	563
12	6.394	5.176	3.838	3.571	362	349	656	570
18	5.13	5.553	2.558	3.678	352	382	789	622
24	3.565	5.248	0.351	3.331	316	360	826.2	585
40	0.333	4.649	-4.477	3.558	261.2	381	1050	594
50	-1.687	5.902	-7.494	4.211	227	392	1189	581

	net work (mJ/g)		Shortening work (mJ/g)	
Time (m	Control	AAPH	Control	AAPH
10	0.16	0.084	0.759	0.30
15	0.113	0.1	0.703	0.555
20	0.085	0.091	0.584	0.561
30	0.007	0.102	0.417	0.674
45	-0.106	0.112	0.153	0.68

oldpcnail2	peak force (kN/m^2)		developed force (kN		time to pf (ms)		time to relax (ms)	
Time (m	Control	AAPH	Control	AAPH	Control	AAPH	Control	AAPH
0	6.946	N/A	6.842	N/A	440	N/A	459	N/A
5	7.551	4.366	6.427	3.835	361	365	301	393
8	6.886	4.514	6.012	3.974	381	379	314	403
12	5.769	5.123	4.596	3.939	374	369	370	426
18	4.329	4.957	3.355	3.835	357	340	424	491
24	3.768	4.385	2.236	3.43	326.3	330	375.6	480
40	1.141	3.886	-1.089	3.415	267.7	350	377.7	465
50	-0.501	4.043	-3.167	3.443	231.2	350	379	436

	net work (mJ/g)		Shortening work (mJ/g)	
Time (m	Control	AAPH	Control	AAPH
10	0.109	0.081	0.445	0.352
15	0.09	0.108	0.384	0.402
20	0.074	0.101	0.342	0.402
30	0.019	0.091	0.236	0.398
45	-0.059	0.108	0.082	0.357

oldpcnail2	peak force (kN/m^2)		developed force (kN		time to pf (ms)		time to relax (ms)	
Time (m	Control	AAPH	Control	AAPH	Control	AAPH	Control	AAPH
0	21.82	N/A	20.78	N/A	325	N/A	321	N/A
5	16.95	14.78	15.35	11.55	307	348	331	338
8	15.77	15.7	13.48	12.18	313	344	334	330
12	13.43	15.77	9.345	9.741	343	365	452	436
18	11.12	13.51	7.929	8.269	359	366	491	624
24	6.935	11.79	2.205	7.023	366.7	346	550.3	644
40	-2.294	12.81	-9.401	8.722	405.4	385	721.2	531
50	-8.062	12.33	-16.66	7.079	429.5	421	828	689

	net work (mJ/g)		Shortening work (mJ/g)	
Time (m	Control	AAPH	Control	AAPH
10	0.258	0.243	1.324	1.368
15	0.211	0.214	1.013	1.305
20	0.154	0.18	0.866	1.022
30	0.078	0.19	0.401	1.083
45	-0.078	0.195	-0.257	1.132

oldpcnail2	peak force (kN/m^2)		developed force (kN		time to pf (ms)		time to relax (ms)	
Time (m	Control	AAPH	Control	AAPH	Control	AAPH	Control	AAPH
0	6.673	N/A	6.254	N/A	532	N/A	627	N/A
5	7.446	3.778	6.151	2.673	510	539	385	612
8	6.475	3.591	5.372	2.565	516	545	393	644
12	6.642	3.82	4.357	2.392	532	512	426	628
18	5.352	3.563	3.685	2.156	573	518	617	695
24	5.269	3.044	3.007	1.844	571.6	467	508.9	687
40	3.971	2.912	0.643	1.334	612.1	510	528.9	643
50	3.16	2.676	-0.835	1.719	637.5	513	541.4	679

	net work (mJ/g)		Shortening work (mJ/g)	
Time (m	Control	AAPH	Control	AAPH
10	0.122	0.051	0.528	0.33
15	0.12	0.061	0.602	0.37
20	0.111	0.048	0.533	0.34
30	0.102	0.043	0.562	0.284
45	0.085	0.046	0.569	0.29

oldpcnail2	peak force (kN/m^2)		developed force (kN		time to pf (ms)		time to relax (ms)	
Time (m	Control	AAPH	Control	AAPH	Control	AAPH	Control	AAPH
0	9.826	N/A	9.734	N/A	385	N/A	443	N/A
5	8.673	7.365	8.16	5.434	362	416	334	433
8	8.927	8.393	8.102	5.945	378	417	324	425
12	8.829	10.01	6.878	5.829	382	396	344	412
18	7.194	9.576	5.216	5.029	395	382	479	509
24	6.76	9.514	3.873	4.83	394.4	390	423.1	589
40	4.758	9.514	-0.018	6.353	409	404	462.9	445
50	3.507	9.327	-2.45	5.887	418.2	440	487.8	547

	net work (mJ/g)		Shortening work (mJ/g)	
Time (m	Control	AAPH	Control	AAPH
10	0.177	0.152	0.619	0.77
15	0.172	0.144	0.682	0.7
20	0.147	0.131	0.639	0.921
30	0.121	0.152	0.677	0.936
45	0.076	0.181	0.707	0.78

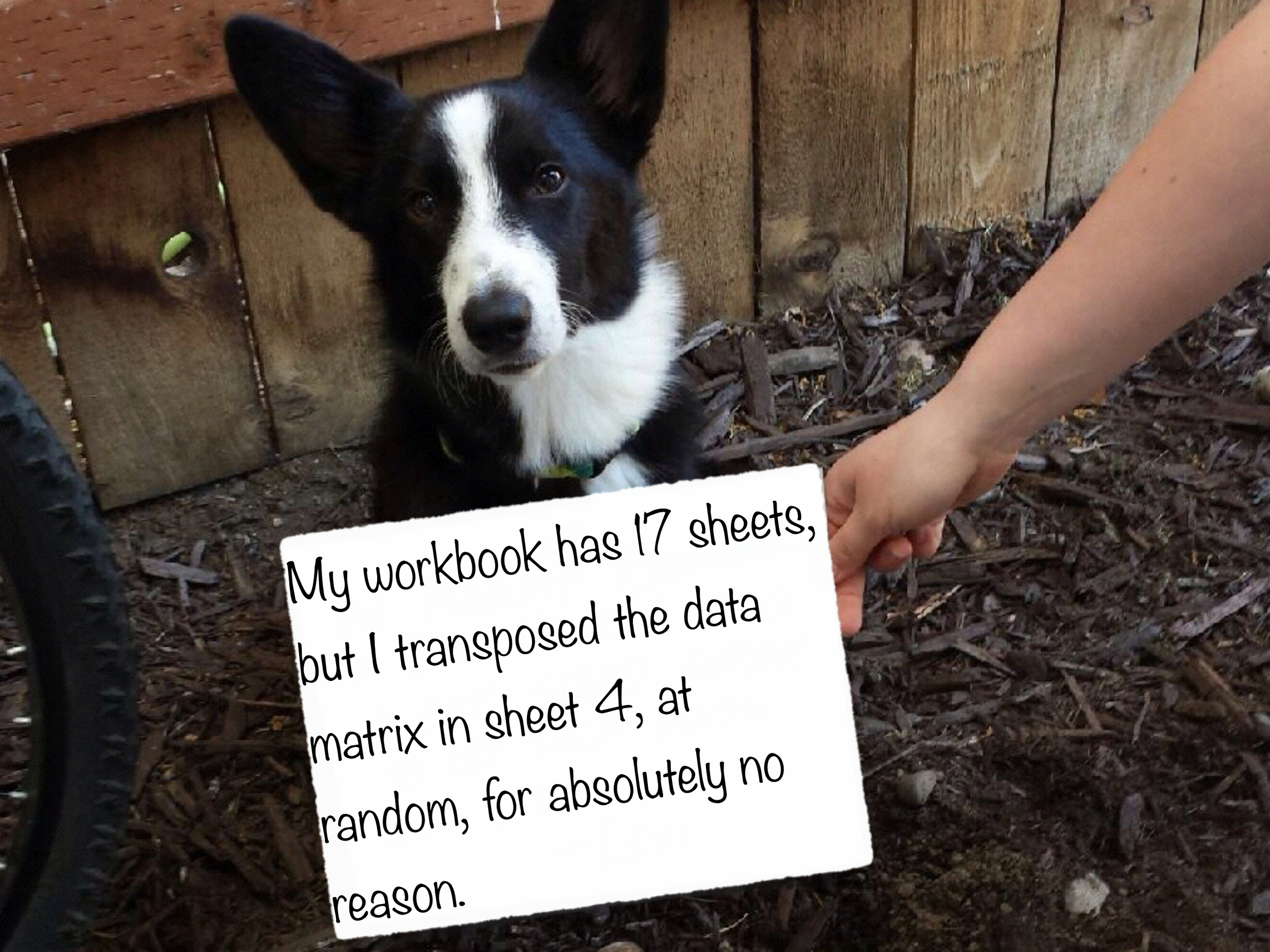
EXTRAPOLATED FROM TRIALS

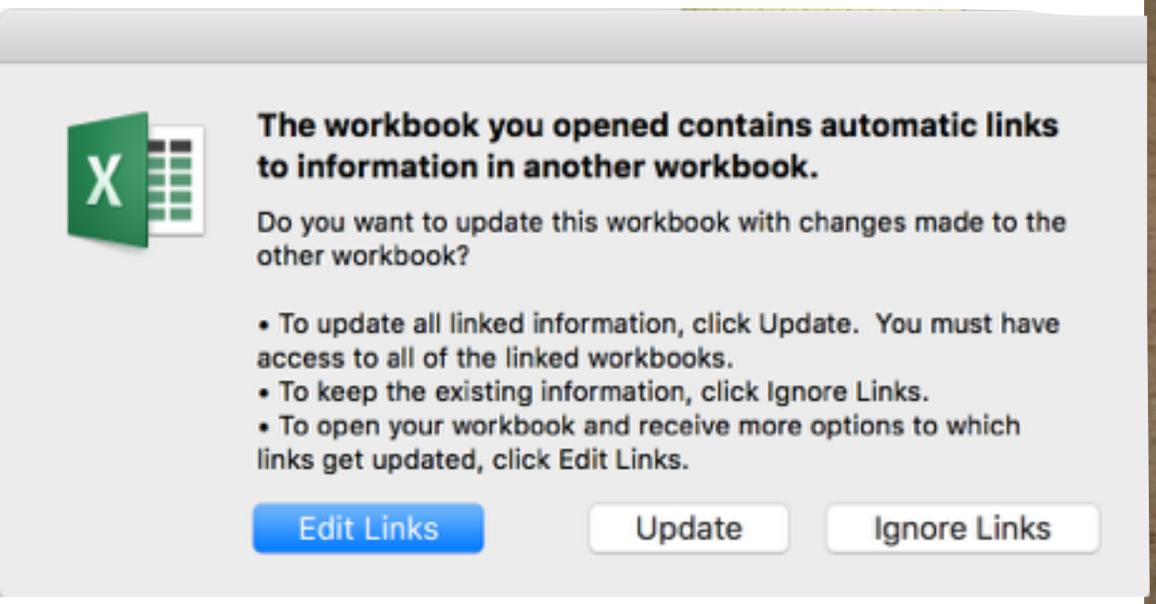
	Mean Peak Force (kN		Mean Dev. Force (kN		Mean Time to P.F. (m		Mean time to relax (ms)	
Time (mir	Control	AAPH	Control	AAPH	Control	AAPH	Control	AAPH
0	11.69	N/A	11.32	N/A	412.2	N/A	441	N/A
5	10.53	7.53	8.92	6.25	373.8	390.7	360.7	435.3
8	9.829	8.165	8.218	6.534	379.2	397.8	350.2	440.5
12	9.255	9.111	6.723	6.129	391.8	390.7	421.3	461.7
18	7.54	8.822	5.341	5.717	395.5	393.2	518.8	543
24	6.311	8.351	3.059	5.375	384.9	377	501.4	550.3
40	2.723	8.353	-2.184	6.008	379.1	407.2	587.6	502.3
50	0.48	8.343	-5.46	6.052	375.5	424.2	641.4	546.8

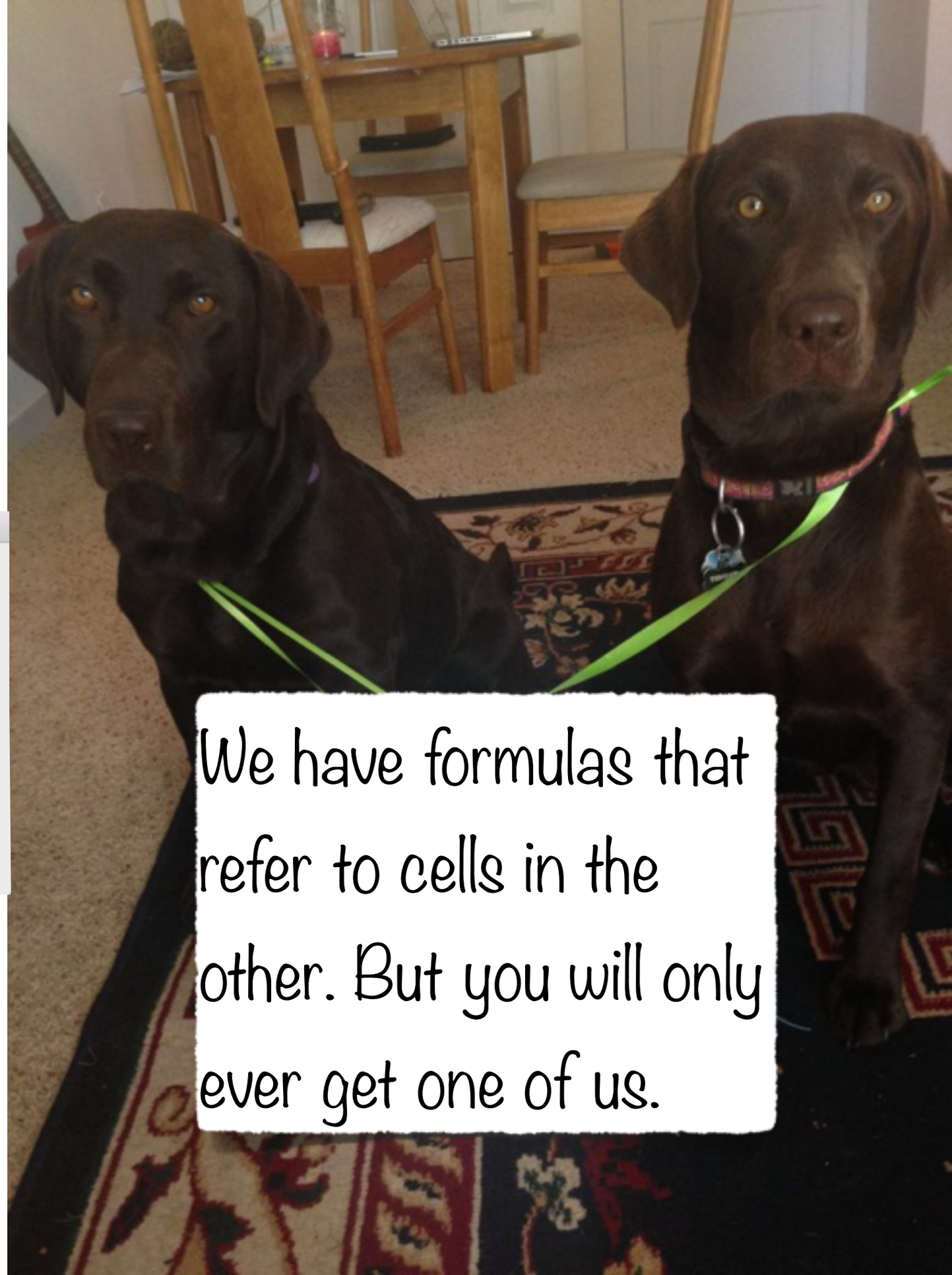
	Mean Net Work (mJ/g		Mean Shortening Work (mJ/g)	
Time (mir	Control	AAPH	Control	AAPH
10	0.182	0.14	0.769	0.668
15	0.15	0.139	0.713	0.75
20	0.13	0.126	0.626	0.711
30	0.072	0.131	0.489	0.727
45	-0.01	0.141	0.275	0.634

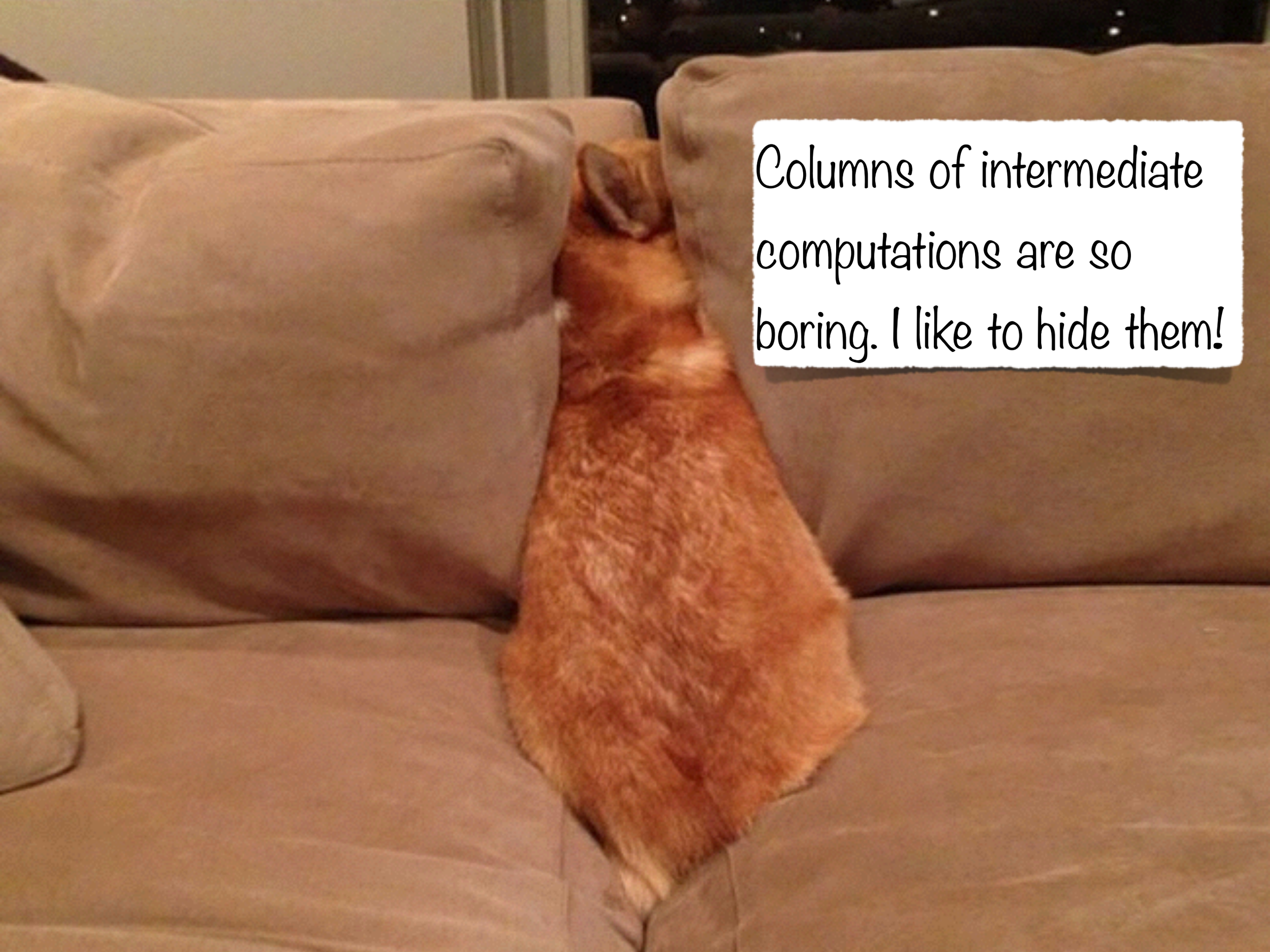
	Peak Force SE		Dev. Force SE		Time to P.F. SE		Time to relax SE	
Time (mir	Control	AAPH	Control	AAPH	Control	AAPH	Control	AAPH
0	2.407	N/A	2.306	N/A	29.59	N/A	56.64	N/A
5	1.87	1.933	1.683	1.601	28.5	31.98	45.74	52.37
8	1.842	2.162	1.531	1.789	29.24	31.45	35.59	56.77
12	1.549	2.133	1.138	1.474	28.57	25.19	53.11	48.51
18	1.341	2.034	1.038	1.402	36.35	25.75	67.42	55.03
24	1.202	2.095	0.866	1.451	39.16	19.83	73.33	54.63
40	1.545	2.223	1.663	1.558	53.48	22.41	105.7	45.23
50	2.088	2.072	2.422	1.759	63.41	22.14	127.1	54.85

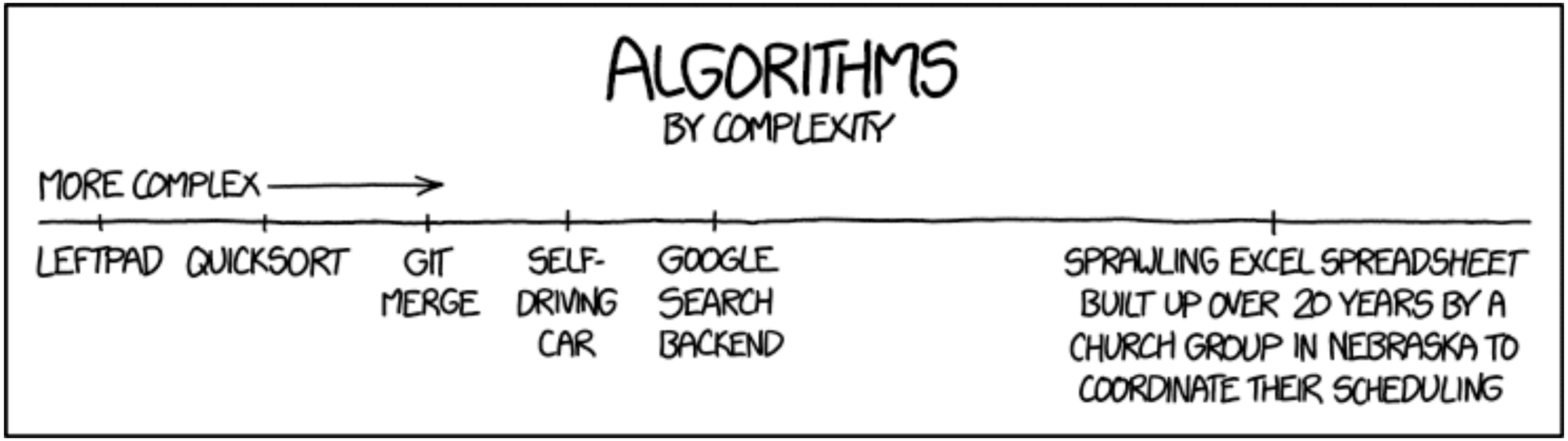
	Net Work SE		Shortening Work SE	
Time (mir	Control	AAPH	Control	AAPH
10	0.029	0.031	0.13	0.167
15	0.02	0.025	0.09	0.147
20	0.019	0.023	0.077	0.126
30	0.02	0.025	0.068	0.135
45	0.033	0.025	0.144	0.125


My workbook has 17 sheets,
but I transposed the data
matrix in sheet 4, at
random, for absolutely no
reason.


The workbook you opened contains automatic links to information in another workbook.
Do you want to update this workbook with changes made to the other workbook?
• To update all linked information, click Update. You must have access to all of the linked workbooks.
• To keep the existing information, click Ignore Links.
• To open your workbook and receive more options to which links get updated, click Edit Links.
Edit Links
Update
Ignore Links


We have formulas that refer to cells in the other. But you will only ever get one of us.


Columns of intermediate computations are so boring. I like to hide them!


ALGORITHMS
BY COMPLEXITY
MORE COMPLEX
LEFTPAD
QUICKSORT
GIT MERGE
SELF-DRIVING CAR
GOOGLE SEARCH BACKEND
SPRAWLING EXCEL SPREADSHEET BUILT UP OVER 20 YEARS BY A CHURCH GROUP IN NEBRASKA TO COORDINATE THEIR SCHEDULING

machine readable

&

human readable

code

can be
machine & human readable

data

can be

machine & human readable

csv,conf,v2

A community conference for data makers everywhere, again!

Featuring some of the brightest minds in science, journalism, government and open source.

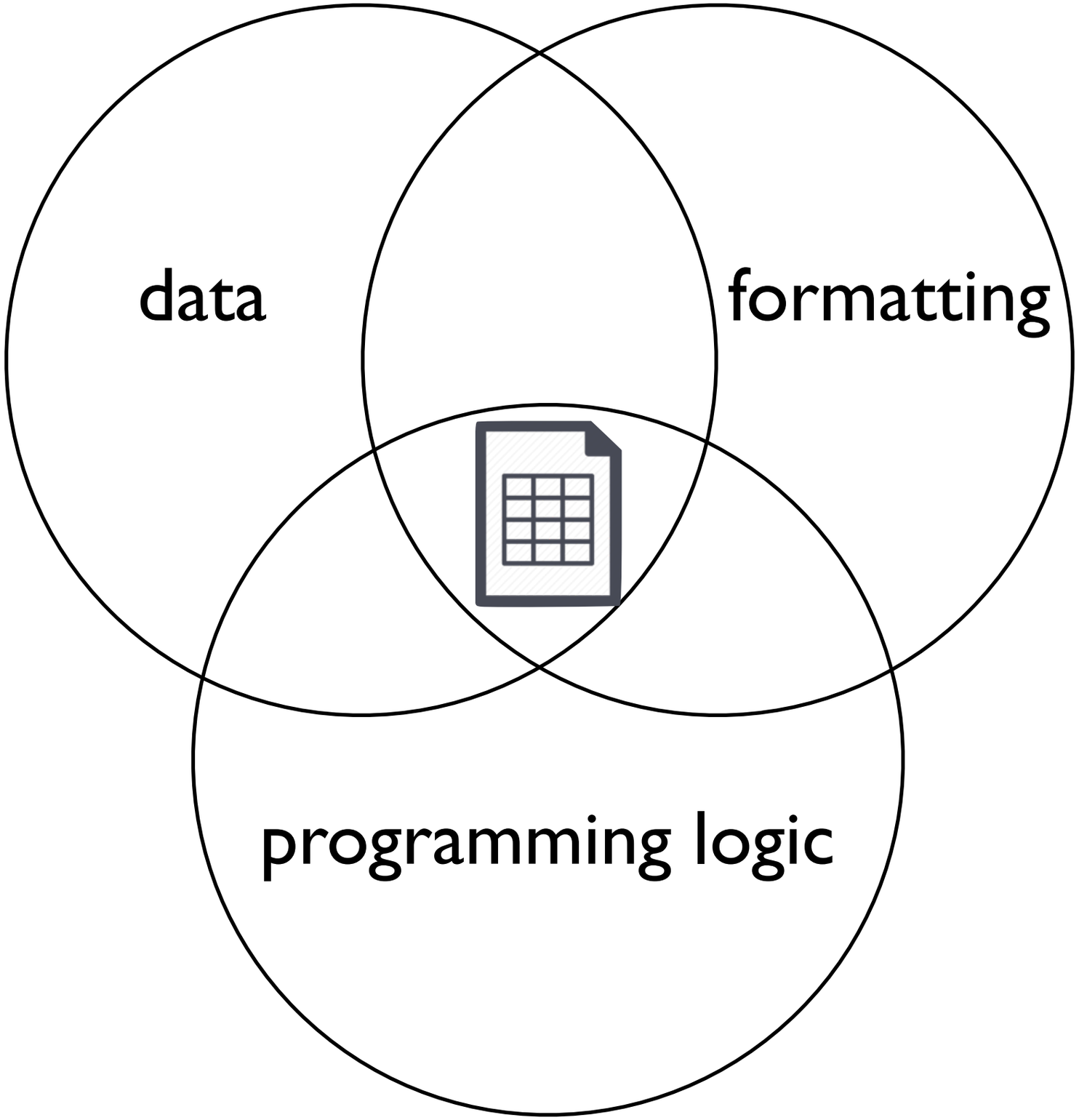
3-4 May 2016 in Berlin, Germany.

GET YOUR TICKET

Pay What You Can, Early Bird Ends April 15th

Coming to Berlin early? Sign up for csv,conf,pre,event 2 May 2016!

a spreadsheet
is often neither
machine nor human readable


data
formatting
programming logic

what are the problems?

which ones can we solve?

via training

via tooling

be realistic,
be fair,
be precise

let 1,000 flowers bloom!

Two angles on the Spreadsheet Problem:

Create new spreadsheet applications
that use, e.g., R or Python, for computation
and visualization.

Accept spreadsheets as they are.
Create tools to get goodies out
and into, e.g. R or Python.
Maybe write back into sheets?

Stencila

Sheets

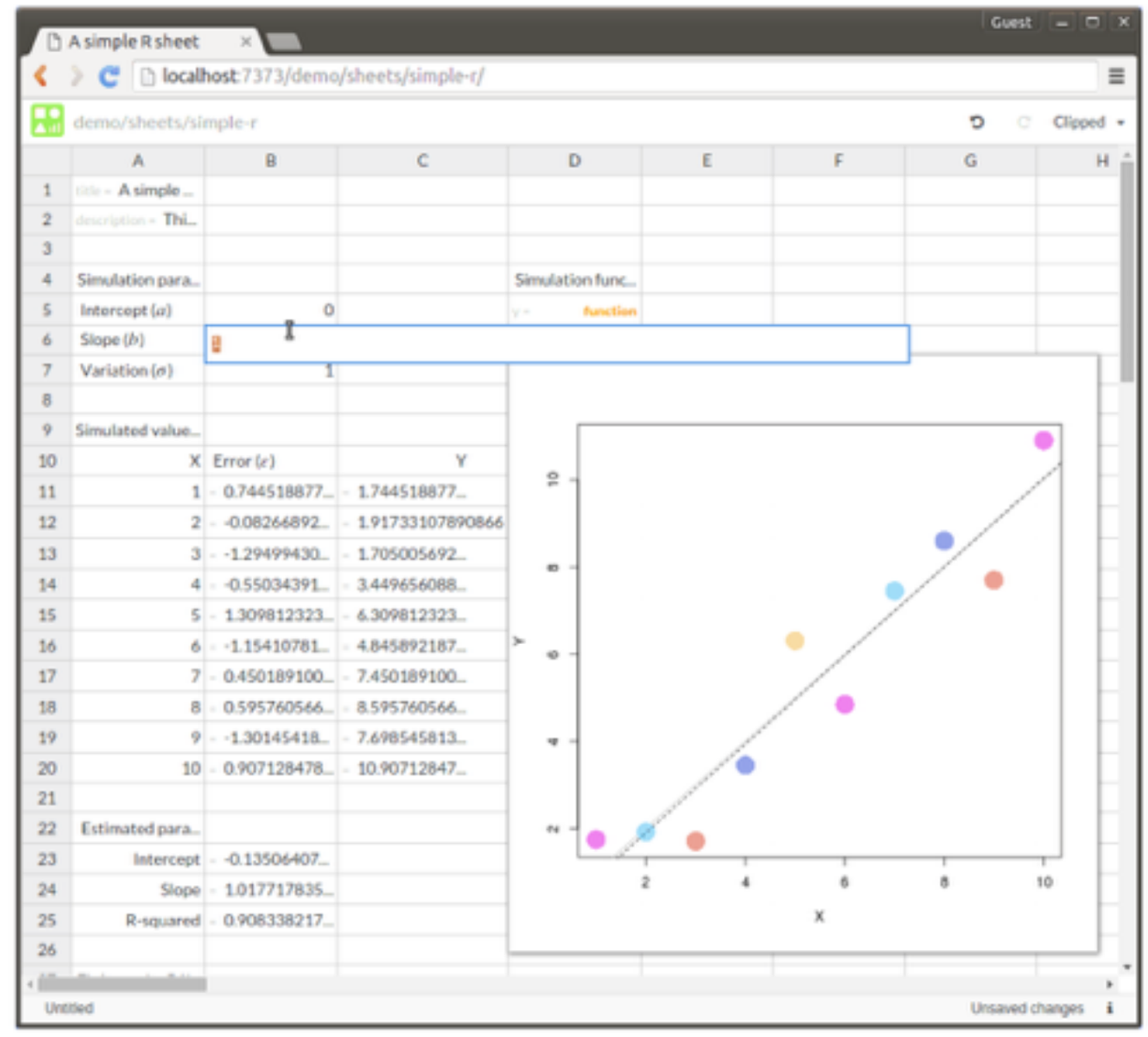
Spreadsheets are the probably the most widely used environment for data analysis and end user programming. Their ubiquity comes from the intuitive simplicity of a live, reactive, see-your-data-and-do-stuff with it interface.

Stencila sheets combine the benefits of the spreadsheet interface with power of languages like R and Python.

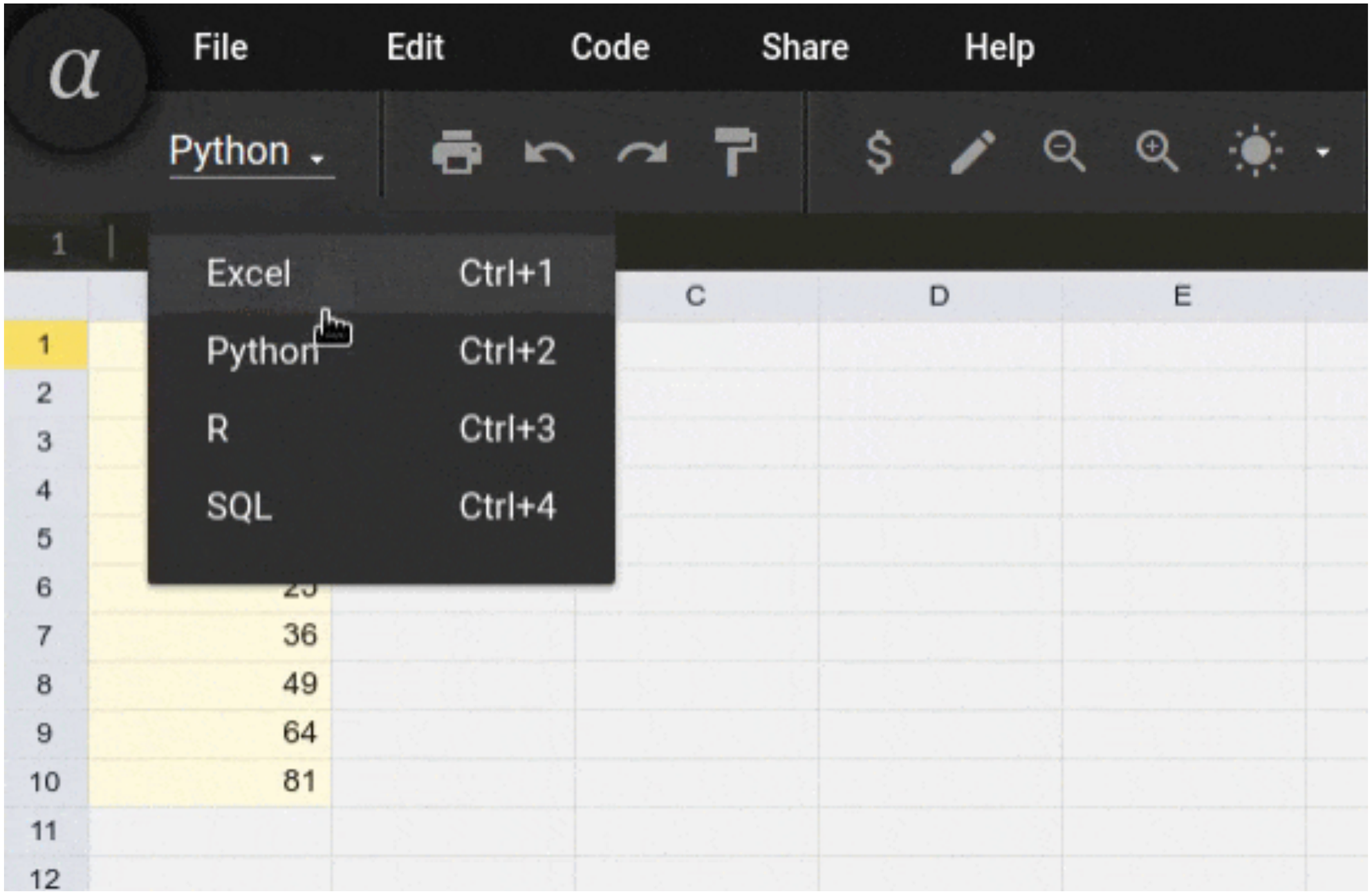
Built from the ground up for transparency, testability and version control, we think Stencila sheets are the next step in the evolution of the spreadsheet.

More:

Spreadsheets are dead, long live reactive programming environments!
Getting under Stencila sheets
A spreadsheet file format for humans

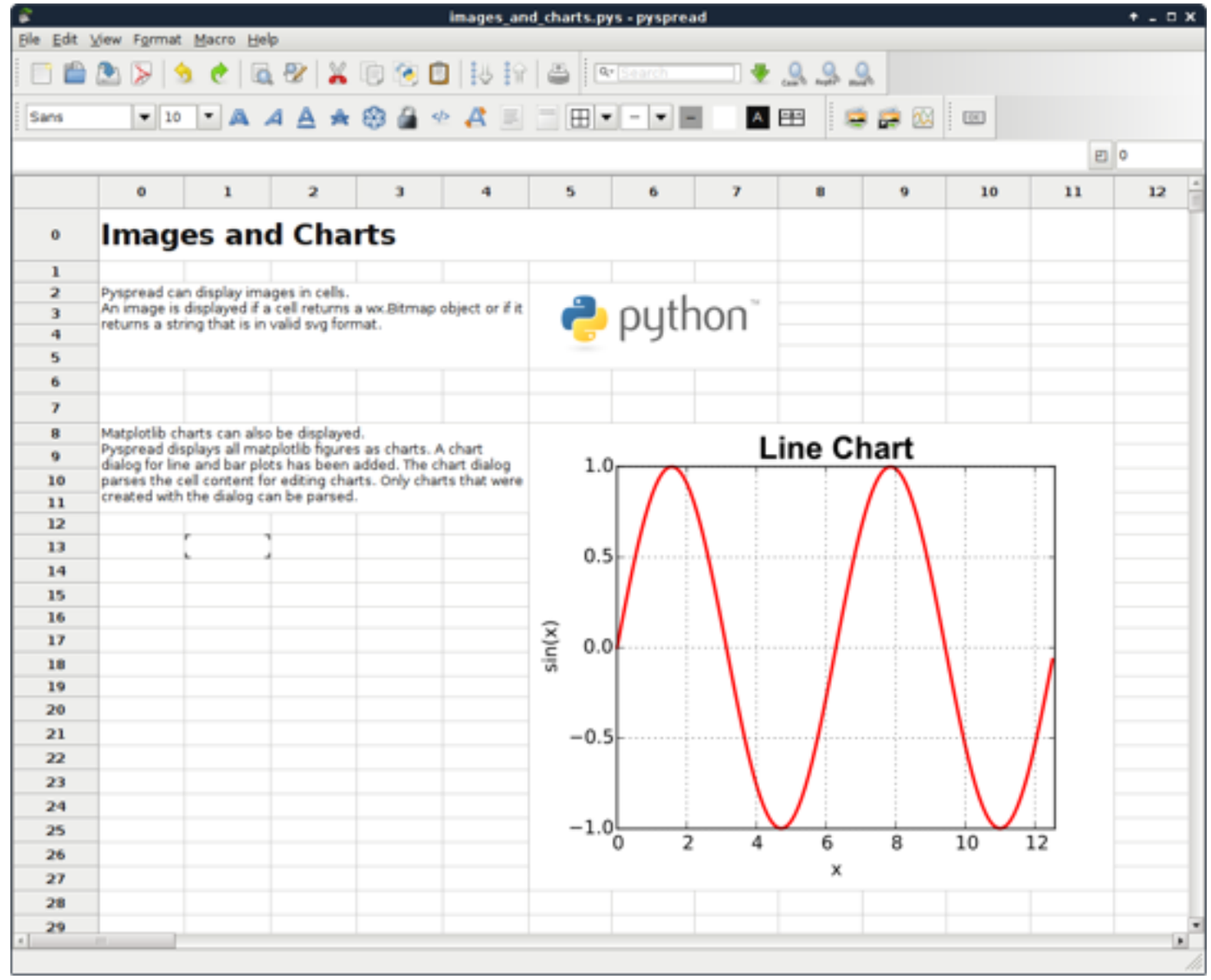


AlphaSheets
“collaborative, programmable spreadsheets”



pyspread

Python power for your tables



Pyspread is a non-traditional spreadsheet application that is based on and written in the programming language Python.

The goal of pyspread is to be the most pythonic spreadsheet.

Pyspread expects Python expressions in its grid cells, which makes a spreadsheet specific language obsolete. Each cell returns a Python object that can be accessed from other cells. These objects can represent anything including lists or matrices.

Pyspread is free software. It is released under the GPL v3. You can find the sorce code at github.

Jean Adams
@JeanVAdams

If your collaborator asks, “In what form would you like the data?” you should respond, “In its current form.” via @kwbroman

Till Doehmen

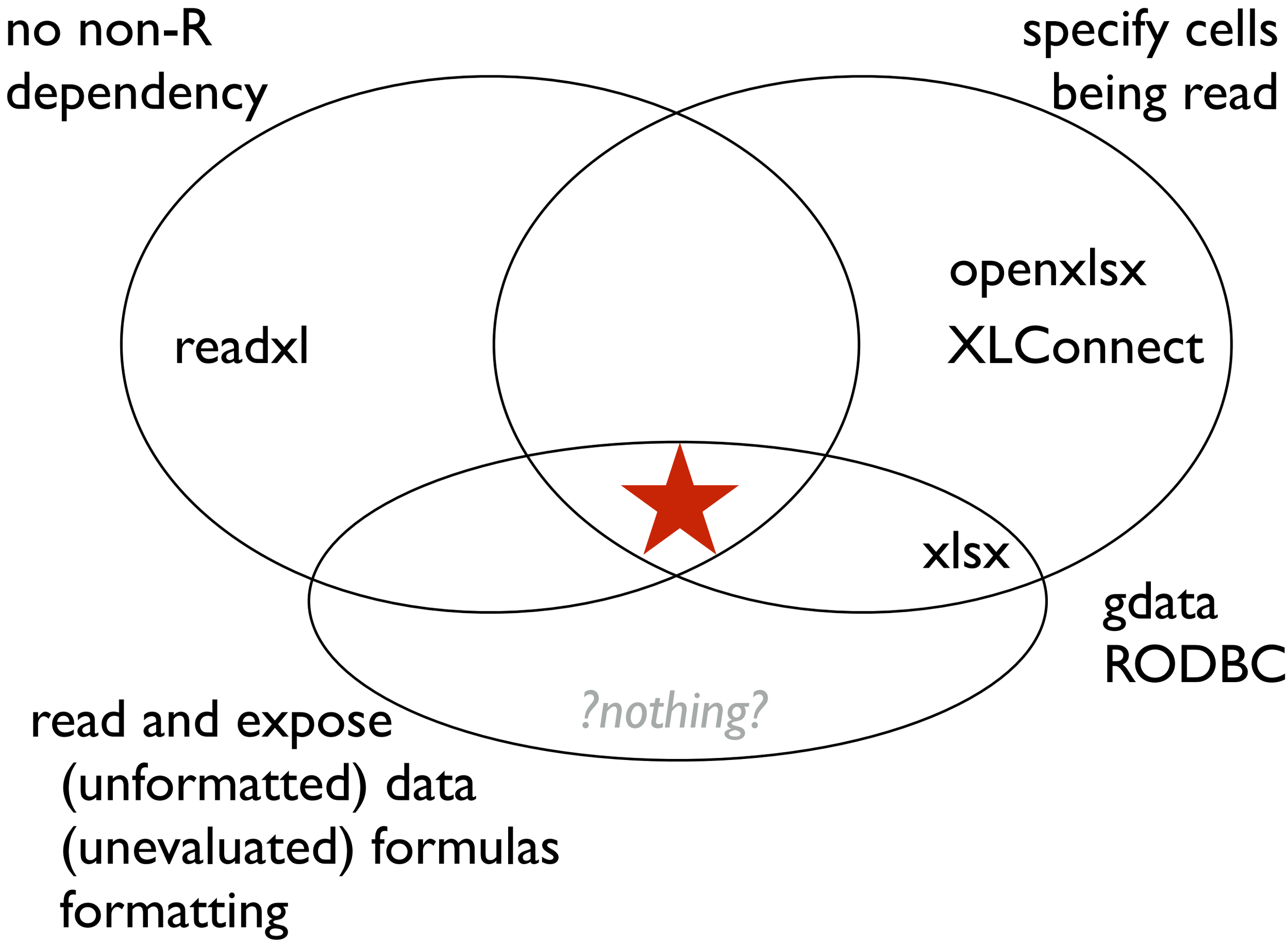
There and back again – Automatic detection and conversion of logical table structures

Darren Barnes

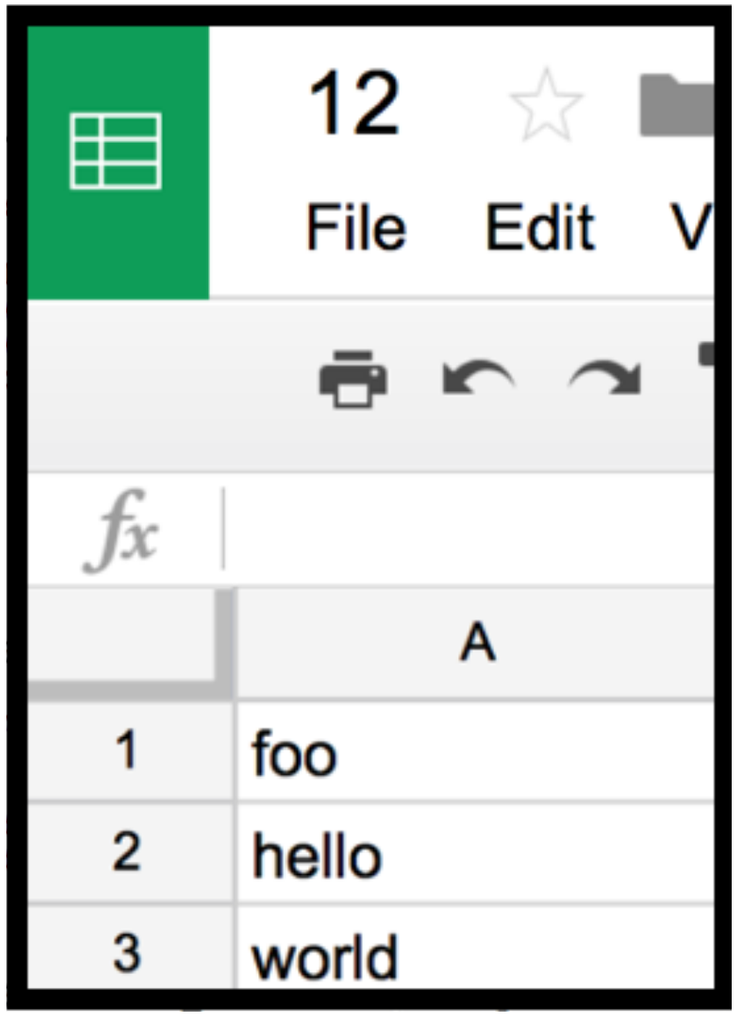
ONS Databaker: from 'pretty spreadsheets' to useful CSVs

Also: R package TableToLongForm

readxl: CRAN, GitHub
openxlsx: CRAN, GitHub
XLConnect: CRAN, GitHub
gdata: CRAN, R-Forge
xlsx: CRAN, GitHub
RODBC: CRAN
… and more

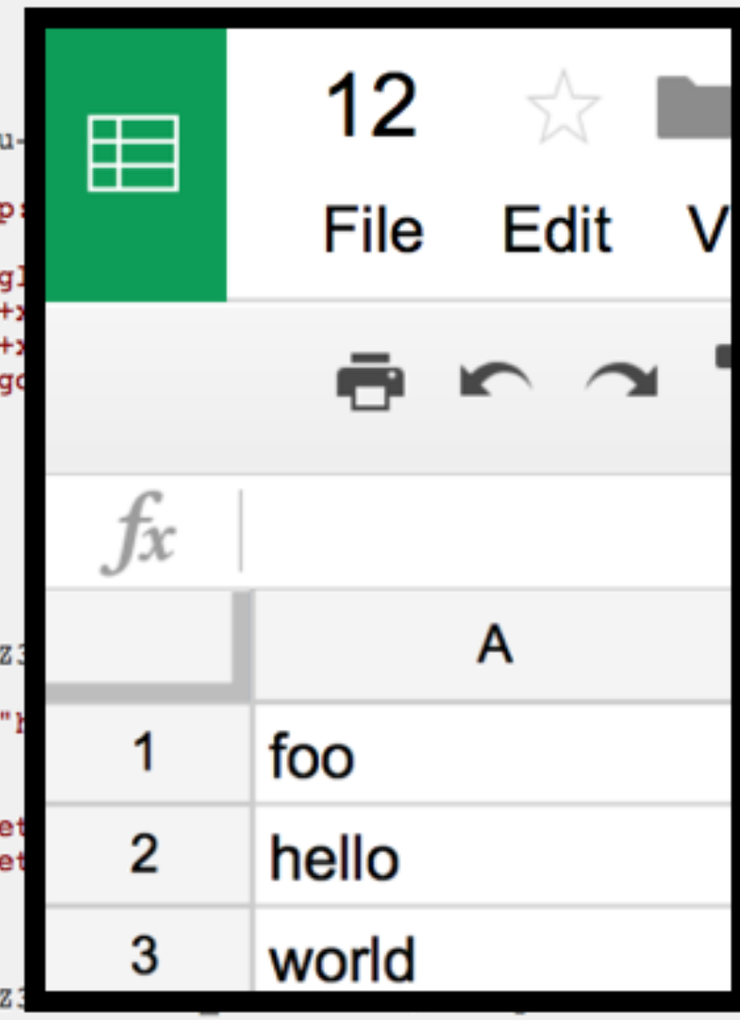

no non-R dependency
specify cells being read
readxl
openxlsx
XLConnect
xlsx
gdata
RODBC
?nothing?
read and expose
(unformatted) data
(unevaluated) formulas
formatting

This is your spreadsheet.



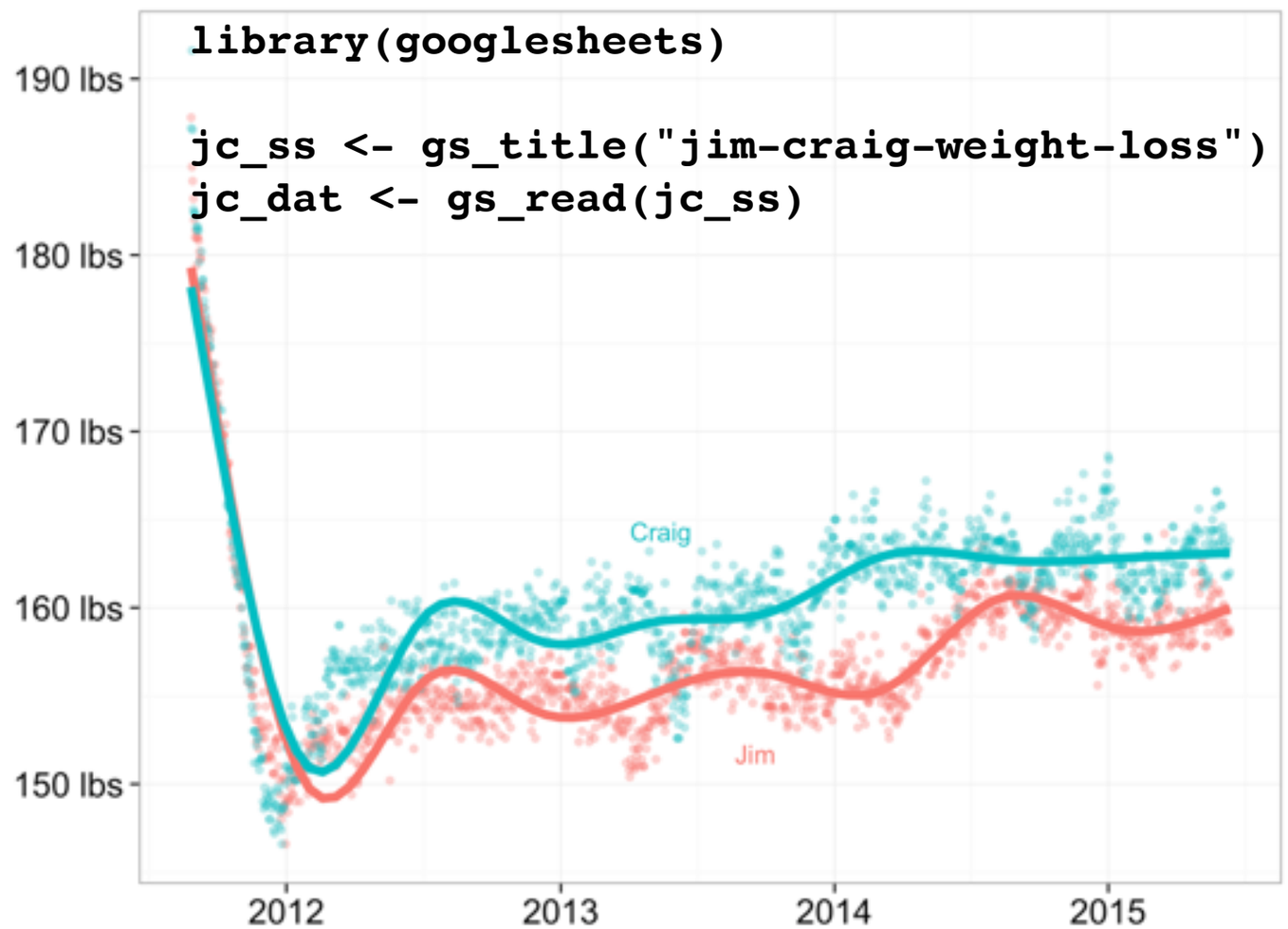
This is your spreadsheet on XML.

```
<?xml version="1.0" encoding="UTF-8"?>
<feed
    xmlns="http://www.w3.org/2005/Atom"
    xmlns:openSearch="http://a9.com/-/spec/opensearchrss/1.0/"
    xmlns:gsx="http://schemas.google.com/spreadsheets/2006/extended">
    <id>https://spreadsheets.google.com/feeds/list/1oLRq_aSy5A6uh60-Hk3EZ3bcu-
    <updated>2016-02-20T04:25:10.497Z</updated>
    <category scheme="http://schemas.google.com/spreadsheets/2006" term="http:
    <title type="text">Sheet1</title>
    <link rel="alternate" type="application/atom+xml" href="https://docs.googl
    <link rel="http://schemas.google.com/g/2005#feed" type="application/atom+x
    <link rel="http://schemas.google.com/g/2005#post" type="application/atom+x
    <link rel="self" type="application/atom+xml" href="https://spreadsheets.go
    <author>
        <name>gspreadr</name>
        <email>gspreadr@gmail.com</email>
    </author>
    <openSearch:totalResults>2</openSearch:totalResults>
    <openSearch:startIndex>1</openSearch:startIndex>
    <entry>
        <id>https://spreadsheets.google.com/feeds/list/1oLRq_aSy5A6uh60-Hk3EZ3
        <updated>2016-02-20T04:25:10.497Z</updated>
        <category scheme="http://schemas.google.com/spreadsheets/2006" term="h
        <title type="text">hello</title>
        <content type="text"/>
        <link rel="self" type="application/atom+xml" href="https://spreadsheet
        <link rel="edit" type="application/atom+xml" href="https://spreadsheet
        <gsx:foo>hello</gsx:foo>
    </entry>
    <entry>
        <id>https://spreadsheets.google.com/feeds/list/1oLRq_aSy5A6uh60-Hk3EZ3
        <updated>2016-02-20T04:25:10.497Z</updated>
        <category scheme="http://schemas.google.com/spreadsheets/2006" term="http://schemas.google.com/spreadsheets/2006#l
        <title type="text">world</title>
        <content type="text"/>
        <link rel="self" type="application/atom+xml" href="https://spreadsheets.google.com/feeds/list/1oLRq_aSy5A6uh60-Hk3
        <link rel="edit" type="application/atom+xml" href="https://spreadsheets.google.com/feeds/list/1oLRq_aSy5A6uh60-Hk3
        <gsx:foo>world</gsx:foo>
    </entry>
</feed>
```



<u>googlesheets</u>

```
library(googlesheets)

jc_ss <- gs_title("jim-craig-weight-loss")
jc_dat <- gs_read(jc_ss)
```



	A	B	C	D	E	F
1	Point Events		Pulling team:	Seattle Rainmakers	Pulling team lineup	Receiving team lineup
2	Offense	Defense	Period:	1	2	78
3		42P	Clock before point:	3:12:00	29	98
4	21pu		Clock after point:	1:49:00	31	47
5		29pu			14	8
6		31			22	21
7		29			42	14
8		31				12
9		14			Pulling team TO 1	Receiving team TO 1
10		22			22	17
11	47pu				29	19
12	98td				7	78
13		29pu			6	11
14		31TO			16	8
15		6pu			3	45
16		29			24	13
17		16			Pulling team TO 2	Receiving team TO 2
18		29				
19		16G				

	A	B	C	D	E	F
1	Point Events		Pulling team:	Seattle Rainmakers	Pulling team lineup	Receiving team lineup
2	Offense	Defense	Period:	1	2	78
3		42P	Clock before point:	3:12:00	29	98
4	21pu		Clock after point:	1:49:00	31	47
5		29pu			14	8
6		31			22	21
7		29			42	14
8		31				12
9		14			Pulling team TO 1	Receiving team TO 1
10		22			22	17
11	47pu				29	19
12	98td				7	78
13		29pu			6	11
14		31TO			16	8
15		6pu			3	45
16		29			24	13
17		16			Pulling team TO 2	Receiving team TO 2
18		29				
19		16G				

seaRM at vanNH 2015-06-06

seaRM 28

vanNH 27

game is complete

Go to ...

* Scoring progression
* Player stats via figures
* Player stats for seaRM
* Player stats for vanNH
* Data on goals, assists, D's
* Full raw data

I wanted live stats!

Scoring progression:

point	period	begin	end	pt_duration	desc	seaRM	vanNH
59	4	0:18:00	0:00:00	00:18	- no goal -	28	27
58	4	0:52:00	0:18:00	00:34	seaRM-27-pape to seaRM-29-feeley	28	27
57	4	1:10:00	0:52:00	00:18	vanNH-14-eyrich to vanNH-8-hibbert	27	27
56	4	1:36:00	1:10:00	00:26	seaRM-3-el-salaam to seaRM-24-lynch	27	26
55	4	1:48:00	1:36:00	00:12	vanNH-41-hochhalter to vanNH-6-chu	26	26
54	4	2:29:00	1:48:00	00:41	vanNH-8-hibbert to vanNH-14-eyrich	26	25

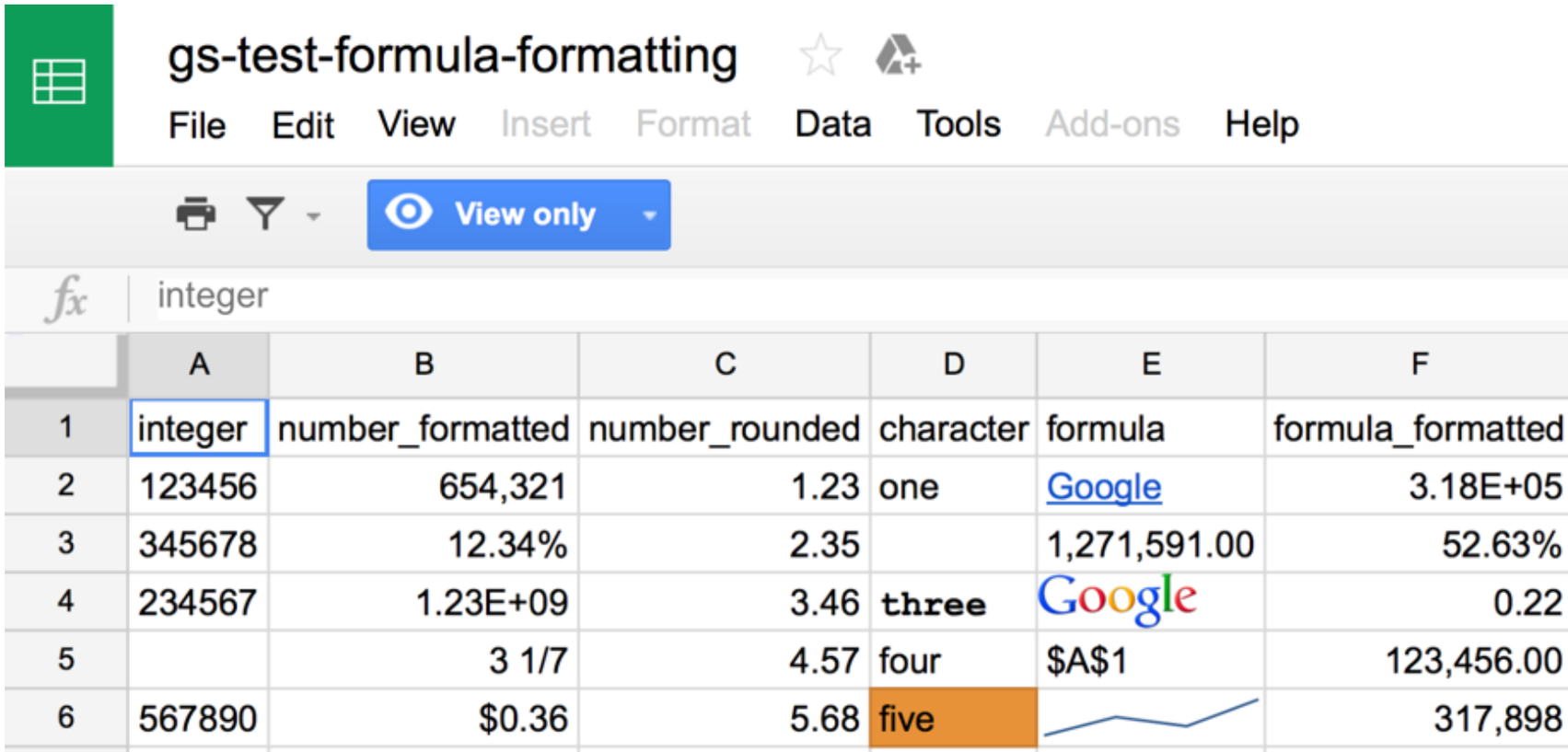
gs-test-formula-formatting

File Edit View Insert Format Data Tools Add-ons Help

View only

fx integer

	A	B	C	D	E	F
1	integer	number_formatted	number_rounded	character	formula	formula_formatted
2	123456	654,321	1.23	one	Google	3.18E+05
3	345678	12.34%	2.35		1,271,591.00	52.63%
4	234567	1.23E+09	3.46	three	Google	0.22
5		3 1/7	4.57	four	A1	123,456.00
6	567890	$0.36	5.68	five		317,898



	A	B	C	D	E	F
1	integer	number_formatted	number_rounded	character	formula	formula_formatted
2	123456	654,321	1.23	one	Google	3.18E+05
3	345678	12.34%	2.35		1,271,591.00	52.63%
4	234567	1.23E+09	3.46	three	Google	0.22
5		3 1/7	4.57	four	A1	123,456.00
6	567890	$0.36	5.68	five		317,898

```
gs_ff() %>%
  gs_read() %>%
  select(-integer)
#> Accessing worksheet titled 'Sheet1'.
#> No encoding supplied: defaulting to UTF-8.
#> Source: local data frame [5 x 5]
#>
#>   number_formatted number_rounded character      formula formula_formatted
#>              (chr)          (dbl)     (chr)        (chr)             (chr)
#> 1          654,321           1.23       one       Google          3.18E+05
#> 2           12.34%           2.35        NA 1,271,591.00           52.63%
#> 3         1.23E+09           3.46     three           NA              0.22
#> 4            3 1/7           4.57      four         $A$1        123,456.00
#> 5            $0.36           5.68      five           NA           317,898
```

	A	B	C	D	E
1	Number	Number_wFormat	Character	Formulas	Formula_wFormat
2	123456	654,321	one	Google	3.18E+05
3	345678	12.34%		1,271,591	52.63%
4	234567	1.23E+09	three	Google	0.22
5		3 1/7	four	A1	123,456.00
6	567890	$0.36	five		317,898

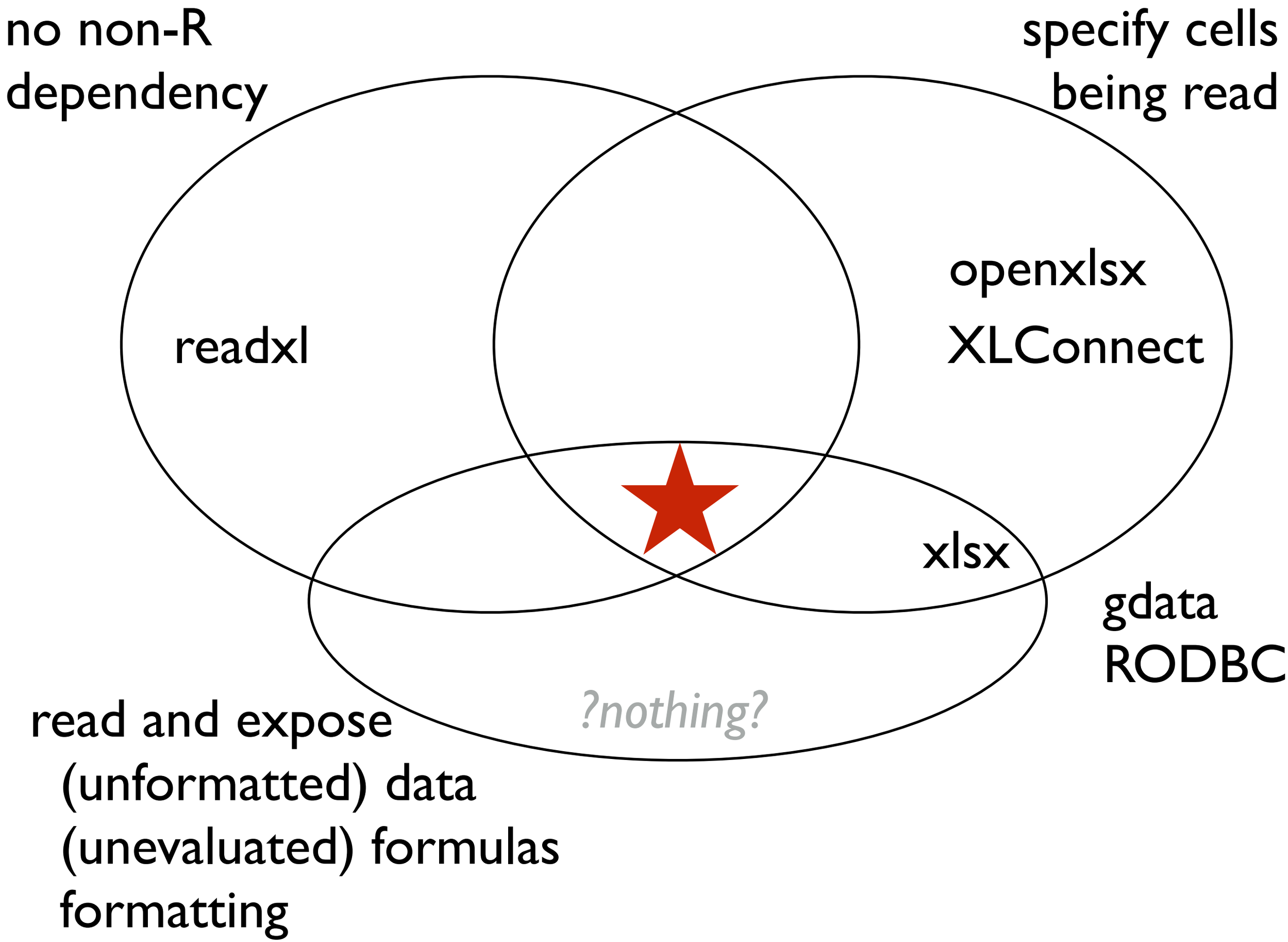
```
#> Source: local data frame [6 x 3]
#>
#>    literal_value    input_value numeric_value
#>            (chr)          (chr)         (chr)
#> 1 Number_wFormat Number_wFormat            NA
#> 2        654,321         654321      654321.0
#> 3         12.34%         12.34%        0.1234
#> 4       1.23E+09     1234567890  1.23456789E9
#> 5          3 1/7  3.14159265359 3.14159265359
#> 6          $0.36           0.36          0.36
```

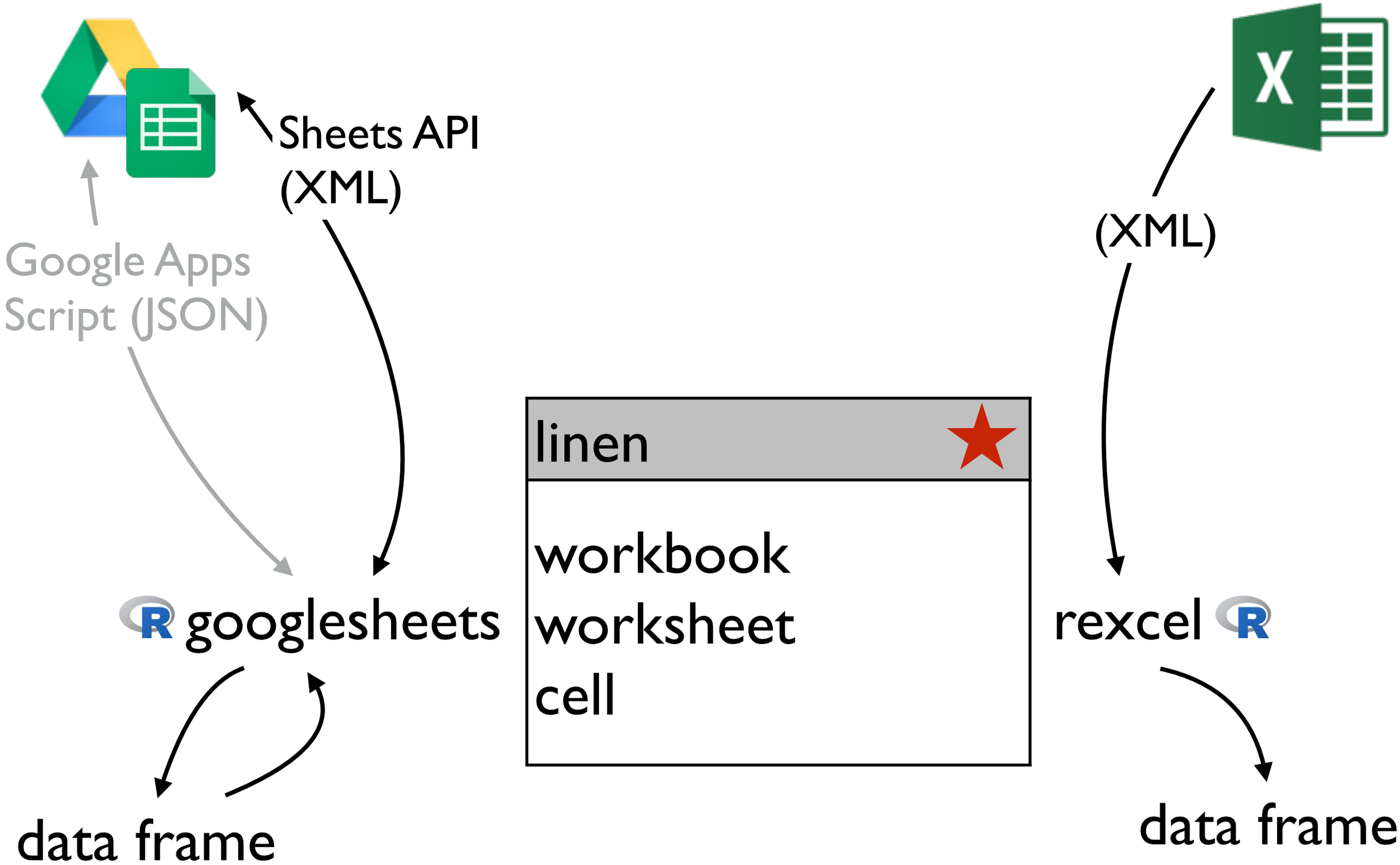
	A	B	C	D	E
1	Number	Number_wFormat	Character	Formulas	Formula_wFormat
2	123456	654,321	one	Google	3.18E+05
3	345678	12.34%		1,271,591	52.63%
4	234567	1.23E+09	three	Google	0.22
5		3 1/7	four	A1	123,456.00
6	567890	$0.36	five		317,898

```
#> Source: local data frame [6 x 3]
#>
#>   literal_value                                                  input_value
#>           (chr)                                                        (chr)
#> 1      Formulas                                                     Formulas
#> 2        Google                =HYPERLINK("http://www.google.com/","Google")
#> 3  1,271,591.00                                   =sum(R[-1]C[-3]:R[3]C[-3])
#> 4                =IMAGE("https://www.google.com/images/srpr/logo3w.png"
#> 5          $A$1                                                =ADDRESS(1,1)
#> 6                                     =SPARKLINE(R[-4]C[-3]:R[0]C[-3])
#> Variables not shown: numeric_value (chr).
```

	A	B	C	D	E
1	Number	Number_wFormat	Character	Formulas	Formula_wFormat
2	123456	654,321	one	Google	3.18E+05
3	345678	12.34%		1,271,591	52.63%
4	234567	1.23E+09	**three**	Google	0.22
5		3 1/7	four	A1	123,456.00
6	567890	$0.36	five		317,898

```
#> Source: local data frame [6 x 3]
#>
#>     literal_value                    input_value       numeric_value
#>             (chr)                          (chr)               (chr)
#> 1 Formula_wFormat                Formula_wFormat                  NA
#> 2        3.18E+05 =average(R[0]C[-4]:R[4]C[-4])           317897.75
#> 3          52.63%          =R[-1]C[-4]/R[1]C[-4]  0.5263144432081239
#> 4            0.22          =R[-2]C[-4]/R[2]C[-4] 0.2173942136681399 6
#> 5      123,456.00     =min(R[-3]C[-4]:R[1]C[-4])            123456.0
#> 6         317,898            =average(R2C1:R6C1)           317897.75
```

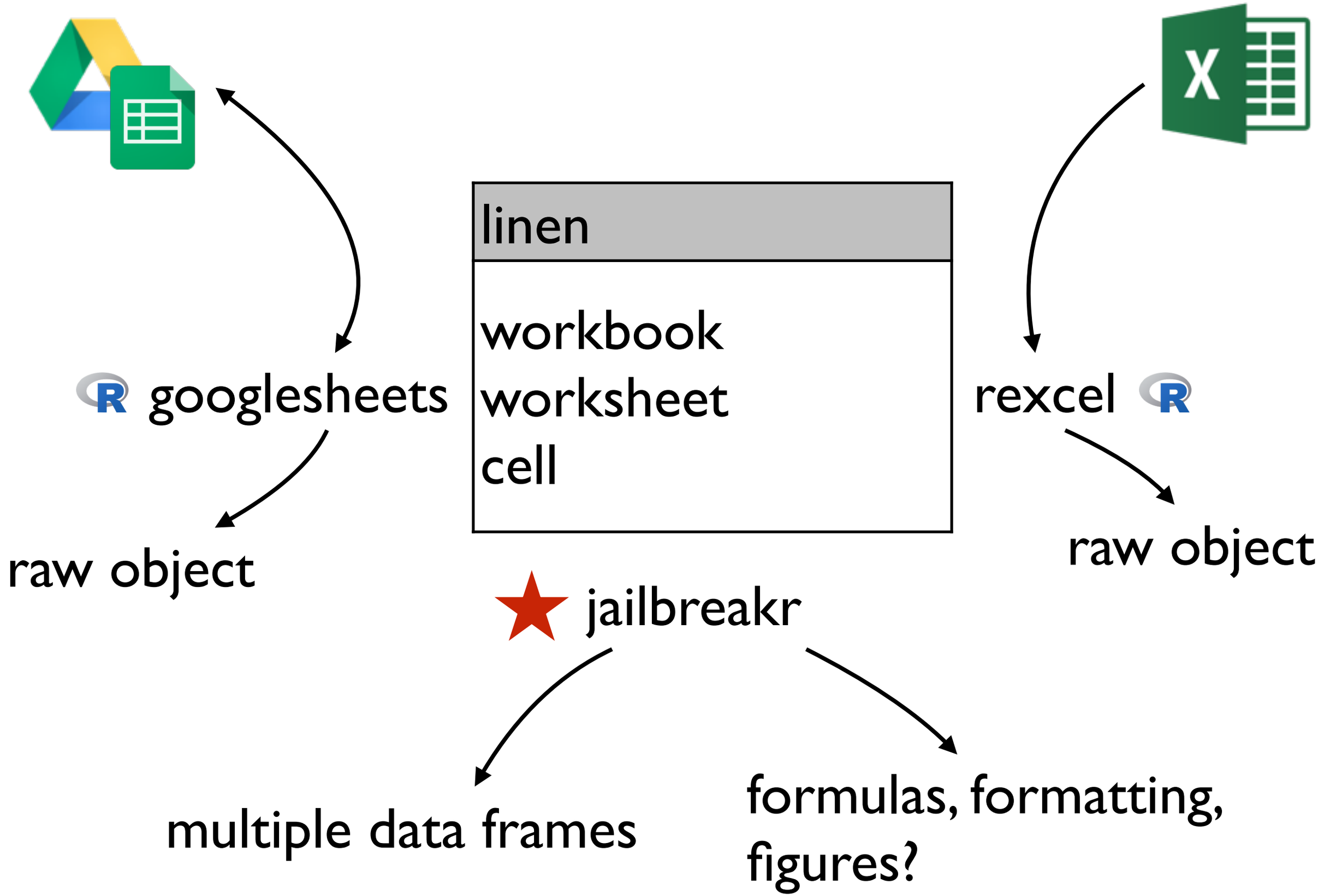

no non-R dependency
specify cells being read
readxl
openxlsx
XLConnect
xlsx
gdata
RODBC
?nothing?
read and expose
(unformatted) data
(unevaluated) formulas
formatting


Sheets API
(XML)
Google Apps
Script (JSON)
(XML)
linen
workbook
worksheet
cell
googlesheets
rexcel
data frame
data frame

Why do we want to get and expose so much?

- ★ jailbreakr:
  - create “views” into a worksheet
  - parse formulas, translate

Expose the linen object for anyone to use.


linen
workbook
worksheet
cell
googlesheets
raw object
rexcel
raw object
jailbreakr
multiple data frames
formulas, formatting, figures?

Enron North America - West Gas

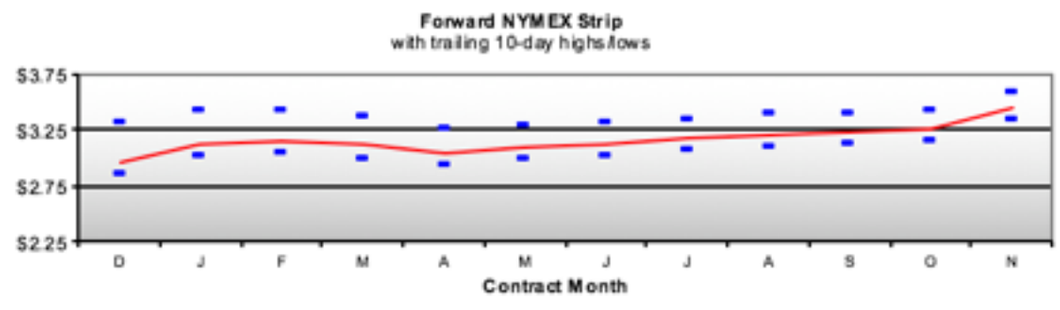
November 9, 2001

ENA - West Gas Contacts

Houston Office		Regional Offices		
Barry Tycholiz	(713) 853-1587	Mark Whitt	(303) 575-6473	Denver
Kim Ward	(713) 853-0685	Paul Lucci	(303) 575-6474	Denver
Stephanie Miller	(713) 853-1688	Tyrell Harrison	(303) 575-6478	Denver
Philip Polsky	(713) 853-5181	Dave Fuller	(503) 464-3732	Portland

Forward Prices (US$/MMBtu)

	NYMEX	
	SETTLE	Δ
Cash		
ROM		
Dec-01	2.960	0.090
Dec-01 to Mar-02	3.088	0.083
Apr-02 to Oct-02	3.166	0.084
Nov-02 to Mar-03	3.651	0.090
One Year Strip*	3.165	0.084



IF NWPL Rocky Mountains			
Fixed Price		Basis	
BID	OFFER	BID	OFFER
1.890	1.910		
2.060	2.080		
2.395	2.415	(0.565)	(0.545)
2.594	2.614	(0.494)	(0.474)
2.581	2.601	(0.585)	(0.565)
3.356	3.376	(0.295)	(0.275)
2.634	2.654	(0.530)	(0.510)

	IF CIG Rocky Mountains				IF EL Paso Permian				IF EL Paso San Juan			
	Fixed Price		Basis		Fixed Price		Basis		Fixed Price		Basis	
	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER
Cash	1.940	1.960			2.375	2.395			2.450	2.470		
ROM	1.960	1.980			2.420	2.440			2.350	2.370		
Dec-01	2.345	2.365	(0.615)	(0.595)	2.700	2.720	(0.260)	(0.240)	2.560	2.580	(0.400)	(0.380)
Dec-01 to Mar-02	2.548	2.568	(0.540)	(0.520)	2.855	2.875	(0.233)	(0.213)	2.743	2.763	(0.345)	(0.325)
Apr-02 to Oct-02	2.471	2.491	(0.695)	(0.675)	3.009	3.029	(0.158)	(0.138)	2.801	2.821	(0.365)	(0.345)
Nov-02 to Mar-03	3.311	3.331	(0.340)	(0.320)	3.499	3.519	(0.153)	(0.133)	3.421	3.441	(0.230)	(0.210)
One Year Strip*	2.551	2.571	(0.614)	(0.594)	2.982	3.002	(0.182)	(0.162)	2.817	2.837	(0.347)	(0.327)

	AECO / NIT				IF NWPL Canadian Border (Sumas)				IF PEPL TX-OK			
	Fixed Price		Basis		Fixed Price		Basis		Fixed Price		Basis	
	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER	BID	OFFER
Cash	2.376	2.396			2.480	2.500			2.530	2.550		
ROM	2.398	2.418			2.460	2.480			2.530	2.550		
Dec-01	2.552	2.572	(0.408)	(0.388)	2.800	2.820	(0.160)	(0.140)	2.828	2.848	(0.133)	(0.113)
Dec-01 to Mar-02	2.616	2.636	(0.472)	(0.452)	2.892	2.912	(0.196)	(0.176)	2.958	2.978	(0.130)	(0.110)
Apr-02 to Oct-02	2.661	2.681	(0.505)	(0.485)	2.796	2.816	(0.370)	(0.350)	3.046	3.066	(0.120)	(0.100)
Nov-02 to Mar-03	3.216	3.236	(0.435)	(0.415)	3.706	3.726	0.055	0.075	3.531	3.551	(0.120)	(0.100)
One Year Strip*	2.676	2.696	(0.488)	(0.468)	2.880	2.900	(0.285)	(0.265)	3.041	3.061	(0.123)	(0.103)

Enron North America - West Gas

November 9, 2001

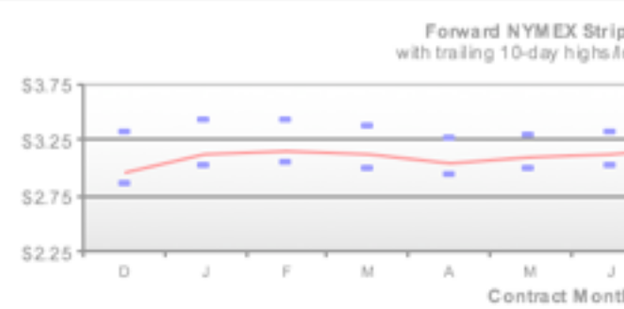
ENA - West Gas Contacts

Houston Office	
Barry Tycholiz	(713) 853-1587
Kim Ward	(713) 853-0685
Stephanie Miller	(713) 853-1688
Philip Polsky	(713) 853-5181

Regional Offices		
Mark Whitt	(303) 575-6473	Denver
Paul Lucci	(303) 575-6474	Denver
Tyrell Harrison	(303) 575-6478	Denver
Dave Fuller	(503) 464-3732	Portland

Forward Prices (US$/MMBtu)

	NYMEX SETTLE	NYMEX Δ
Cash ROM		
Dec-01	2.960	0.090
Dec-01 to Mar-02	3.088	0.083
Apr-02 to Oct-02	3.166	0.084
Nov-02 to Mar-03	3.651	0.090
One Year Strip*	3.165	0.084



IF CIG Rocky Mountains	Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	1.940	1.960		
ROM	1.960	1.980		
Dec-01	2.345	2.365	(0.615)	(0.595)
Dec-01 to Mar-02	2.548	2.568	(0.540)	(0.520)
Apr-02 to Oct-02	2.471	2.491	(0.695)	(0.675)
Nov-02 to Mar-03	3.311	3.331	(0.340)	(0.320)
One Year Strip*	2.551	2.571	(0.614)	(0.594)

AECO / NIT	Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	2.376	2.396		
ROM	2.398	2.418		
Dec-01	2.552	2.572	(0.408)	(0.388)
Dec-01 to Mar-02	2.616	2.636	(0.472)	(0.452)
Apr-02 to Oct-02	2.661	2.681	(0.505)	(0.485)
Nov-02 to Mar-03	3.216	3.236	(0.435)	(0.415)
One Year Strip*	2.676	2.696	(0.488)	(0.468)

IF NWPL Rocky Mountains

```
> sheet <- rexcel::rexcel_read("enron.xlsx")
> sheet
<worksheet: 235 x 99>
   : ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMN
1  : a←←←
```

	AC	AD	AE	AF	AG
115					
116			Congratulations to Dan Foulston		
117			of Canadian Western Natural Gas!!!		
118			He is the winner of our weekend getaway		
119			to Lake Louise. Dan's guess of $1.375 C/GJ		
120			was the closest to the average Aeco		
121			day gas price for the first		
122			15 days of December.		

```
113:
114:
115:
116:                                    a
117:                                    a
118:                                    a
119:                                    a
120:                                    a
121:                                    a
122:                                    a
```

Enron North America - West Gas

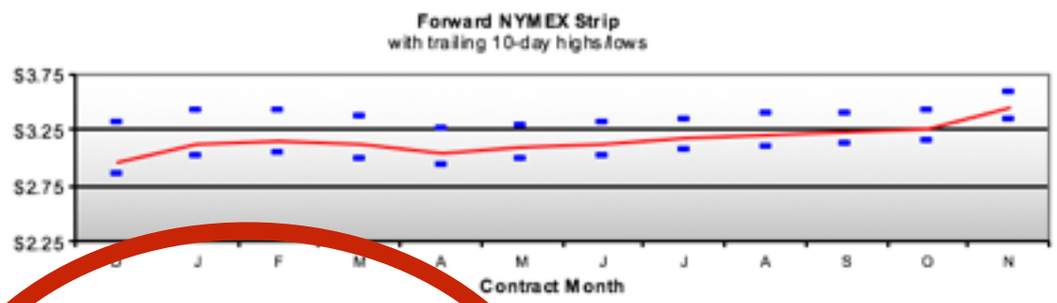
November 9, 2001

ENA - West Gas Contacts

Houston Office		Regional Offices		
Barry Tycholiz	(713) 853-1587	Mark Whitt	(303) 575-6473	Denver
Kim Ward	(713) 853-0685	Paul Lucci	(303) 575-6474	Denver
Stephanie Miller	(713) 853-1688	Tyrell Harrison	(303) 575-6478	Denver
Philip Polsky	(713) 853-5181	Dave Fuller	(503) 464-3732	Portland

Forward Prices (US$/MMBtu)

	NYMEX SETTLE	NYMEX Δ
Cash		
ROM		
Dec-01	2.960	0.090
Dec-01 to Mar-02	3.088	0.083
Apr-02 to Oct-02	3.166	0.084
Nov-02 to Mar-03	3.651	0.090
One Year Strip*	3.165	0.084



	IF NWPL Rocky Mountains Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	1.890	1.910		
ROM	2.060	2.080		
Dec-01	2.395	2.415	(0.565)	(0.545)
Dec-01 to Mar-02	2.594	2.614	(0.494)	(0.474)
Apr-02 to Oct-02	2.581	2.601	(0.585)	(0.565)
Nov-02 to Mar-03	3.356	3.376	(0.295)	(0.275)
One Year Strip*	2.634	2.654	(0.530)	(0.510)

	IF CIG Rocky Mountains Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	1.940	1.960		
ROM	1.960	1.980		
Dec-01	2.345	2.365	(0.615)	(0.595)
Dec-01 to Mar-02	2.548	2.568	(0.540)	(0.520)
Apr-02 to Oct-02	2.471	2.491	(0.695)	(0.675)
Nov-02 to Mar-03	3.311	3.331	(0.340)	(0.320)
One Year Strip*	2.551	2.571	(0.614)	(0.59

	IF EL Paso Permian Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	2.375	2.395		
ROM	2.420	2.440		
Dec-01	2.700	2.720	(0.260)	(0.240)
Dec-01 to Mar-02	2.855	2.875	(0.233)	(0.213)
Apr-02 to Oct-02	3.009	3.029	(0.158)	(0.138)
Nov-02 to Mar-03	3.499	3.519	(0.153)	(0.133)
One Year Strip*	2.982	3.002	(0.182)	(0.162)

	IF EL Paso San Juan Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	2.450	2.470		
ROM	2.350	2.370		
Dec-01	2.560	2.580	(0.400)	(0.380)
Dec-01 to Mar-02	2.743	2.763	(0.345)	(0.325)
Apr-02 to Oct-02	2.801	2.821	(0.365)	(0.345)
Nov-02 to Mar-03	3.421	3.441	(0.230)	(0.210)
One Year Strip*	2.817	2.837	(0.347)	(0.327)

	AECO / NIT Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	2.376	2.396		
ROM	2.398	2.418		
Dec-01	2.552	2.572	(0.408)	(0.388)
Dec-01 to Mar-02	2.616	2.636	(0.472)	(0.452)
Apr-02 to Oct-02	2.661	2.681	(0.505)	(0.485)
Nov-02 to Mar-03	3.216	3.236	(0.435)	(0.415)
One Year Strip*	2.676	2.696	(0.488)	(0.468)

	IF NWPL Canadian Border (Sumas) Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	2.480	2.500		
ROM	2.460	2.480		
Dec-01	2.800	2.820	(0.160)	(0.140)
Dec-01 to Mar-02	2.892	2.912	(0.196)	(0.176)
Apr-02 to Oct-02	2.796	2.816	(0.370)	(0.350)
Nov-02 to Mar-03	3.706	3.726	0.055	0.075
One Year Strip*	2.880	2.900	(0.285)	(0.265)

	IF PEPL TX-OK Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER
Cash	2.530	2.550		
ROM	2.530	2.550		
Dec-01	2.828	2.848	(0.133)	(0.113)
Dec-01 to Mar-02	2.958	2.978	(0.130)	(0.110)
Apr-02 to Oct-02	3.046	3.066	(0.120)	(0.100)
Nov-02 to Mar-03	3.531	3.551	(0.120)	(0.100)
One Year Strip*	3.041	3.061	(0.123)	(0.103)

Enron North America - West Gas

November 9, 2001

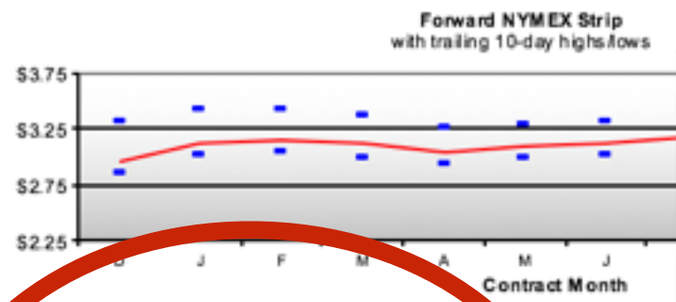
ENA - West Gas Contacts

Houston Office	
Barry Tycholiz	(713) 853-1587
Kim Ward	(713) 853-0685
Stephanie Miller	(713) 853-1688
Philip Polsky	(713) 853-5181

Regional Offices		
Mark Whitt	(303) 575-6473	Denver
Paul Lucci	(303) 575-6474	Denver
Tyrell Harrison	(303) 575-6478	Denver
Dave Fuller	(503) 464-3732	Portland

Forward Prices (US$/MMBtu)

	NYMEX SETTLE	NYMEX Δ
Cash ROM		
Dec-01	2.960	0.090
Dec-01 to Mar-02	3.088	0.083
Apr-02 to Oct-02	3.166	0.084
Nov-02 to Mar-03	3.651	0.090
One Year Strip*	3.165	0.084



	IF CIG Rocky Mountains Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER	IF NWPL Rocky Mountains Fixed Price BID
Cash ROM					
Dec-01	1.940	1.960			2.375
Dec-01 to Mar-02	1.960	1.980			2.420
Apr-02 to Oct-02	2.345	2.365	(0.615)	(0.595)	2.700
Nov-02 to Mar-03	2.548	2.568	(0.540)	(0.520)	2.855
One Year Strip*	2.471	2.491	(0.695)	(0.675)	3.009
	3.311	3.331	(0.340)	(0.320)	3.499
	2.551	2.571	(0.614)	(0.59	2.982

	AECO / NIT Fixed Price BID	Fixed Price OFFER	Basis BID	Basis OFFER	IF NW Fixed P BID
Cash ROM					
Dec-01	2.376	2.396			2.480
Dec-01 to Mar-02	2.398	2.418			2.460
Apr-02 to Oct-02	2.552	2.572	(0.408)	(0.388)	2.800
Nov-02 to Mar-03	2.616	2.636	(0.472)	(0.452)	2.892
One Year Strip*	2.661	2.681	(0.505)	(0.485)	2.796
	3.216	3.236	(0.435)	(0.415)	3.706
	2.676	2.696	(0.488)	(0.468)	2.880

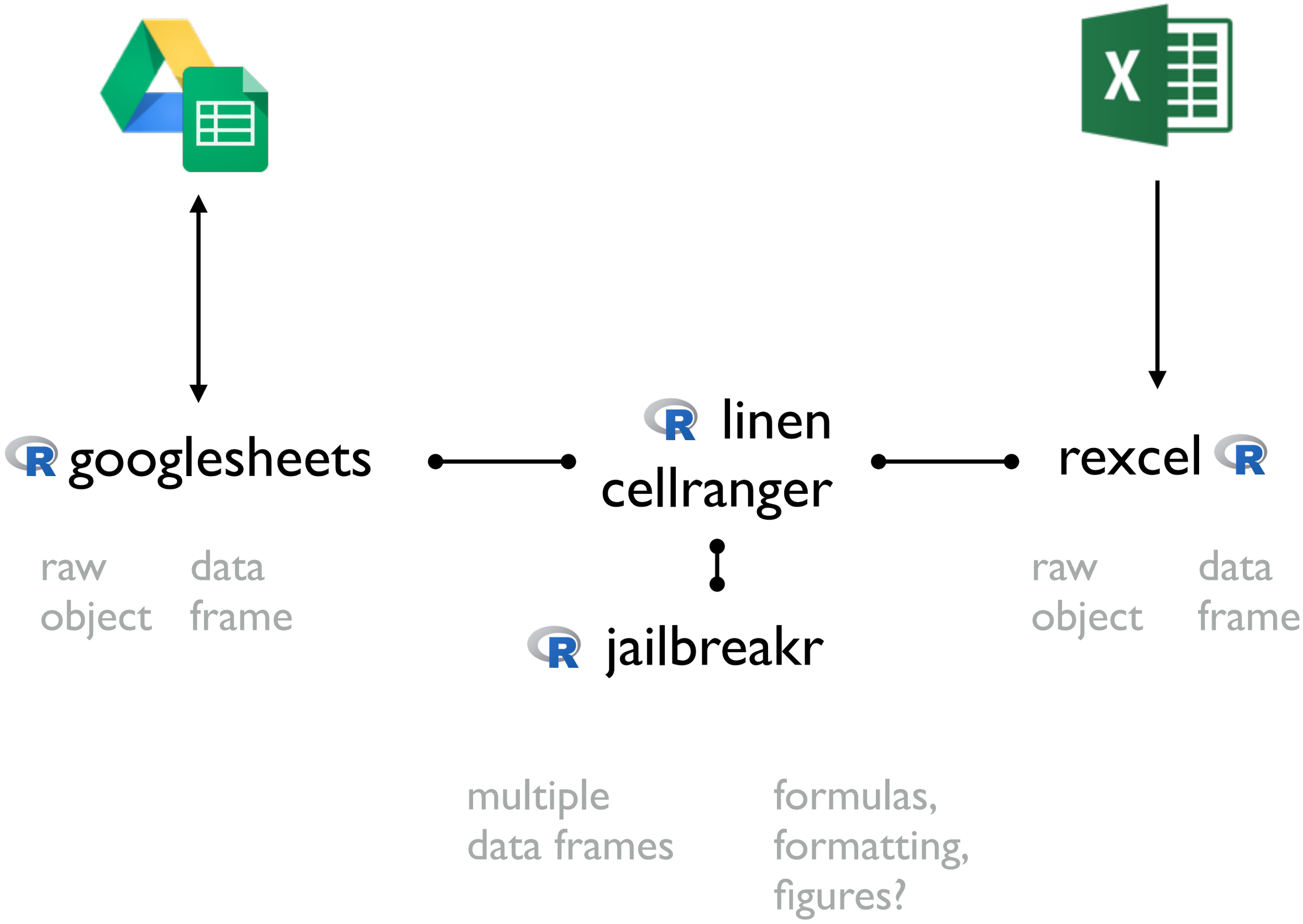
```
> views <- jailbreakr::split_sheet(sheet)
> length(views)
[1] 40
> head(views, 3)
[[1]]
<worksheet_view: 2 x 20 (of 235 x 99)>
  : BCDEFGHIJKLMNOPQRSTU
1: a
2: 0

[[2]]
<worksheet_view: 1 x 20 (of 235 x 99)>
  : BCDEFGHIJKLMNOPQRSTU
4: a

[[3]]
<worksheet_view: 1 x 20 (of 235 x 99)>
  : BCDEFGHIJKLMNOPQRSTU
12: a
```

IF CIG Rocky Mountains			
Fixed Price		Basis	
BID	OFFER	BID	OFFER
1.940	1.960		
1.960	1.980		
2.345	2.365	(0.615)	(0.595)
2.548	2.568	(0.540)	(0.520)
2.471	2.491	(0.695)	(0.675)
3.311	3.331	(0.340)	(0.320)
2.551	2.571	(0.614)	(0.594)

```
> x2 <- jailbreakr::split_metadata(x1, FALSE, min_data_block=1)
> x3 <- jailbreakr::split_headers_apply(x2, 2)
> x3
<worksheet_view: 7 x 4 (of 235 x 99)>
(with headers)
(with data: metadata)
  : GHIJ
28: 00
29: 00
30: 0000
31: 0000
32: 0000
33: 0000
34: 0000
```

IF CIG Rocky Mountains			
Fixed Price		Basis	
BID	OFFER	BID	OFFER
1.940	1.960		
1.960	1.980		
2.345	2.365	(0.615)	(0.595)
2.548	2.568	(0.540)	(0.520)
2.471	2.491	(0.695)	(0.675)
3.311	3.331	(0.340)	(0.320)
2.551	2.571	(0.614)	(0.594)

```
> x3$values()
     Fixed Price:BID Fixed Price:OFFER Basis:BID Basis:OFFER
[1,] 1.94            1.96              NA        NA
[2,] 1.96            1.98              NA        NA
[3,] 2.345           2.365             -0.615    -0.595
[4,] 2.54775         2.56775           -0.54     -0.52
[5,] 2.471286        2.491286          -0.695    -0.675
[6,] 3.311           3.331             -0.34     -0.32
[7,] 2.55075         2.57075           -0.61375  -0.59375
```


googlesheets
raw object
data frame
linen
cellranger
jailbreakr
multiple data frames
formulas, formatting, figures?
rexcel
raw object
data frame

Master of Data Science

http://mds.science.ubc.ca