

Modern software is all about
data.

Development environments
should be, too.

Rob DeLine

Microsoft Research

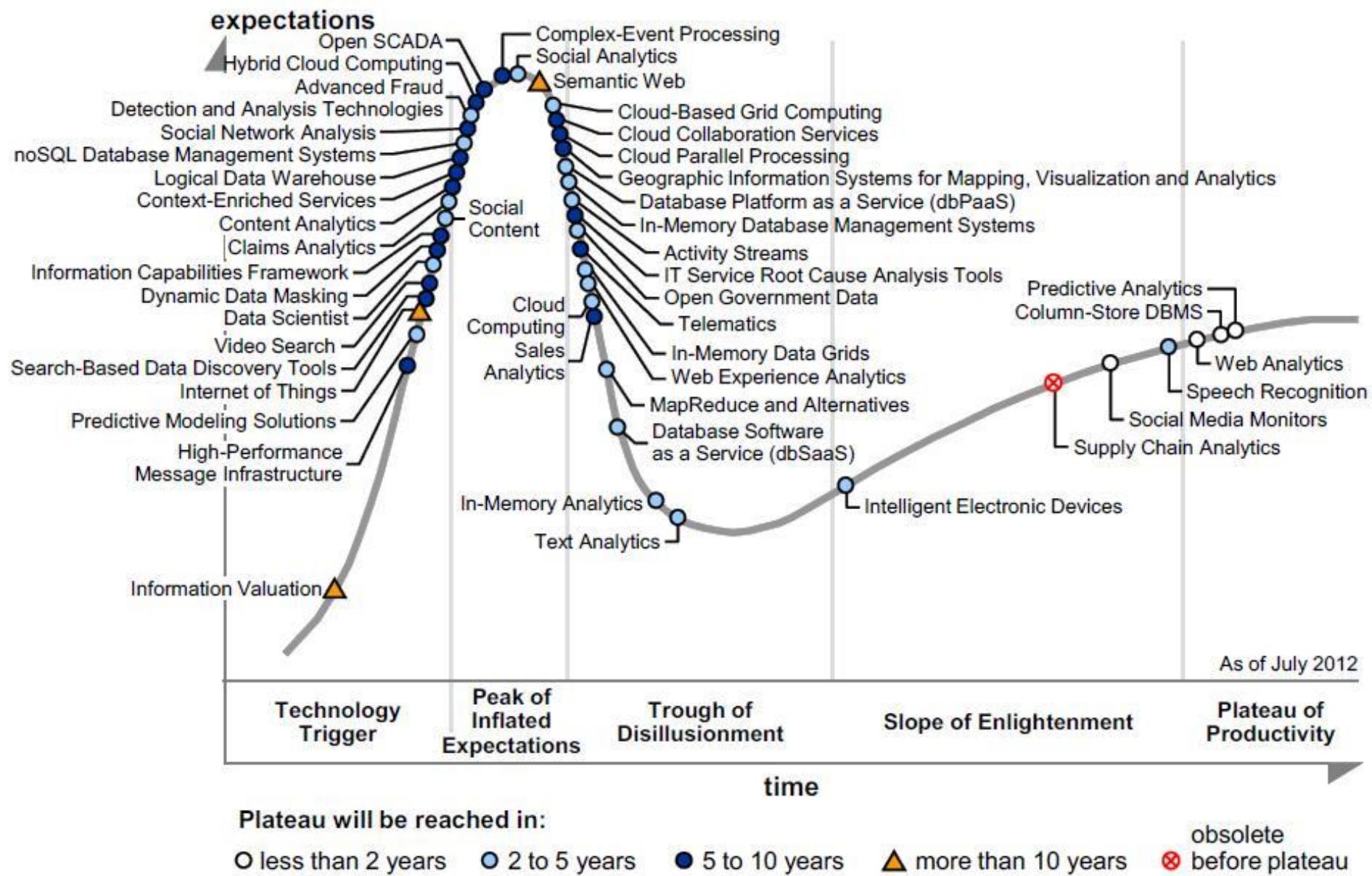
SPLASH 2015



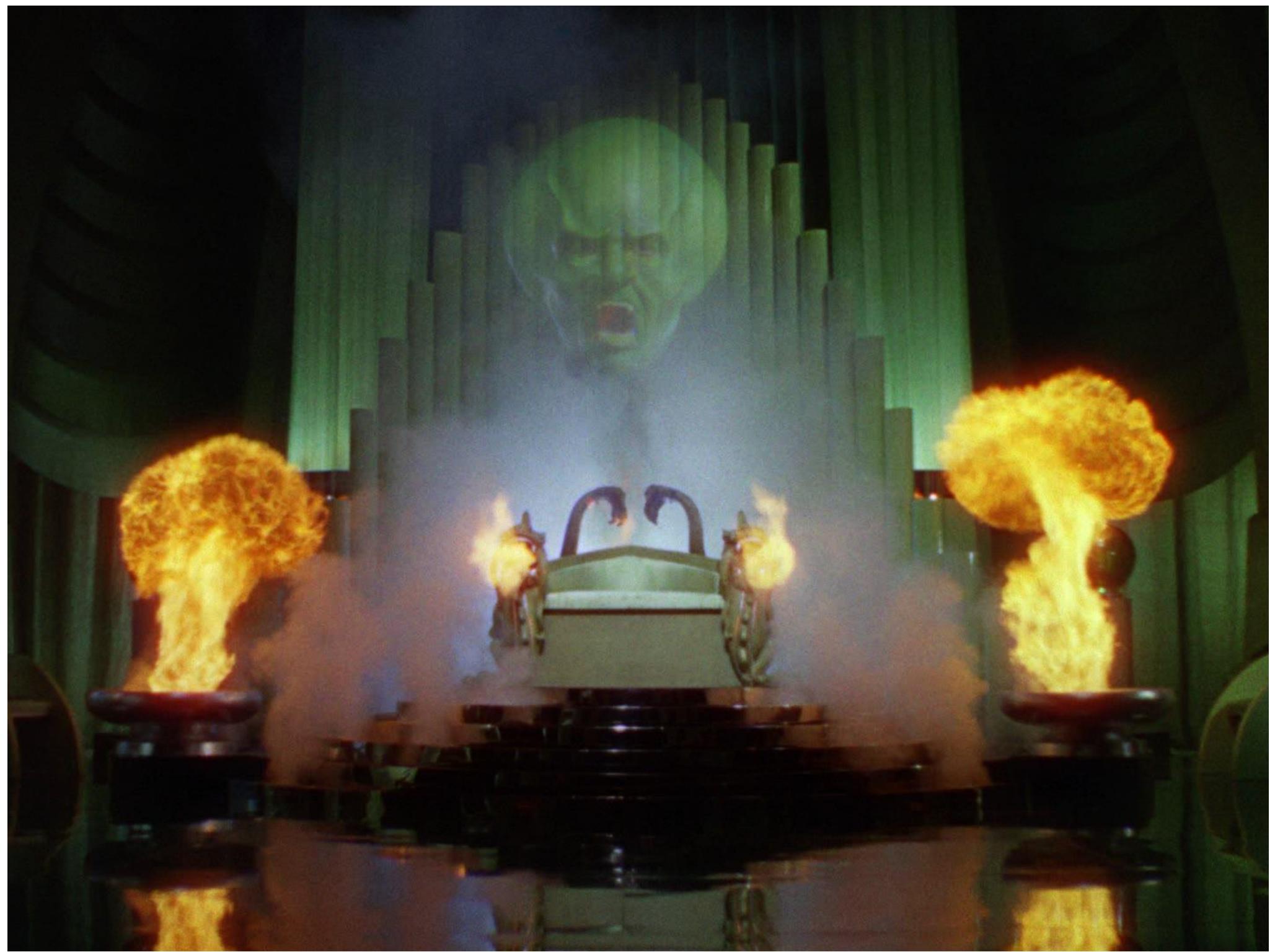


**BIG
DATA**

Figure 1. Hype Cycle for Big Data, 2012



Source: Gartner (July 2012)











What is a data scientist?

We interviewed 16 data scientists at Microsoft.

5 women, 11 men. 3 MS/MBA, 8 PhDs

Ads, Azure, Bing, Exchange, Office, R&D, Skype, Windows, and Xbox

Recruited through snowball sampling



Kim, Zimmermann, DeLine, and Begel,
“The Emerging Role of Data Scientists on Software Development
Teams”

MSR-TR-2015-30 (in submission)

Five working styles of data scientists

insight

team provider
leader

platform builder

model

specialist
polymath

Five working styles of data scientists

insight provider
team leader
platform builder
model specialist
polymath

Coordinate between managers and engineers within a product group
Generate insights and to guide their managers in decision making
Strong communication and coordination skills
P2 got a clear goal for managers and worked with engineers to get data:
I basically tried to eliminate from the vocabulary the notion of "You can just throw the data over the wall ... She'll figure it out." [P2]

Five working styles of data scientists

insight provider
team leader
platform builder
model specialist
polymath

Senior data scientists who typically run their own data science teams

Act as data science “evangelists”

Work with senior company leaders to inform broad business decisions

P10 led a team to do bug estimation:

Sometimes people who are really good with numbers are not as good with words (laughs), so having an intermediary to handle the human interfaces between the data sources and the data scientists, I think, is a way to have a stronger influence.

Five working styles of data scientists

insight provider
team leader
platform builder
model specialist
polymath

Build reusable data engineering platforms

Make trade-offs between engineering and scientific concerns

Strong systems background

P4 makes crash data actionable:

You come up with something called a bucket feed. It is a name of a function most likely responsible for the crash in the small bucket. We found in the source code who touched this function last time. He gets the bug.

Five working styles of data scientists

insight
team provider
platform builder
model specialist
polymath

Act as expert consultants

Build predictive models that can be instantiated as new software features and support other team's data-driven decision making

Strong background in machine learning and statistics.

P7 is an expert in time series analysis:

The [Program Managers] and the Dev Ops from that team... come up with a new set of time series data that they think has the most value and then they will point us to that, and we will try to come up with an algorithm or with a methodology to find the anomalies for that set of time series.

Five working styles of data scientists

insight provider
team leader
platform builder
model specialist
polymath

“Do it all”, from forming a business goal, to data collection, to analysis, to communication

P13 thinks of new ideas for ads:

For months at a time I'll wear a dev hat and I actually really enjoy that, too. ... I spend maybe three months doing some analysis and maybe three months doing some coding that is to integrate whatever I did into the product. ... I love the flexibility that I can go from being developer to being an analyst.

What do data scientists work on?

Performance Regression

Are we getting better in terms of crashes or worse?
[P3]

Requirements Identification

If you see the repetitive pattern where people don't recognize the feature is there. [P3]

Root Cause Analysis

What areas of the product are failing and why? [P3]

Bug Prioritization

Oh, cool. Now we know which bugs we should fix first. Then how can we reproduce this error? [P5]

Server Anomaly Detection

Is this application log abnormal w.r.t. the rest of the data? [P12]

Failure Rate Estimation

Is the beta ready to ship?
[P8]

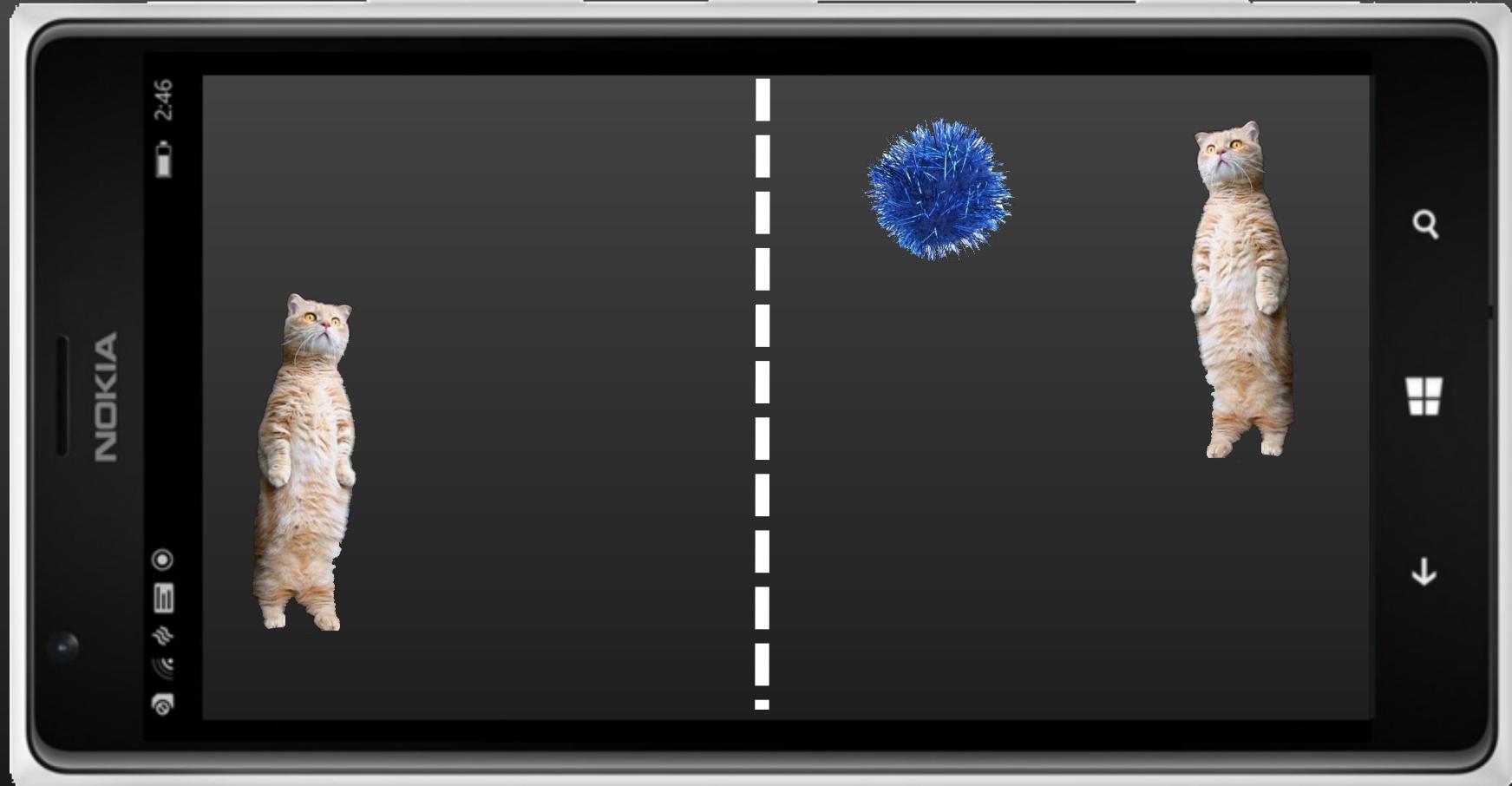
Customer Understanding

How long do our users use the app? [P1]
What are the most popular features? [P4]

Cost Benefit Analysis

How many customer service calls can we prevent if we detect this type of anomaly?
[P9]

What's it like to
be a data
scientist?



Kitteh Pong!

Matching players by skill

- 1 Help the team decide whether to implement this feature (analytics).
- 2 If so, help the team deploy the feature.
- 3 Measure player reaction to the feature (flighting).
- 4 Monitor customer usage of the feature.



John W. Tukey

EXPLORATORY DATA ANALYSIS



la. Grab a manageable sample of the data.

```
Players = LOAD 'player_data';
PSample = SAMPLE Players 0.01;
STORE PSample INTO 'psample';

Games = LOAD 'games_data';
GSample = SAMPLE Games 0.0001;
STORE GSample INTO 'gsample';
```

File Home Insert Page Layout Formulas Data Review View LOAD TEST FOXIT READER PDF Team Tell me Rob De... Share

F9 credit_card

	A	B	C	D	E	F	G	H	I	J
1	Amount.Req	Amount.Fun	Interest_Rate	Loan_Length	Loan_Purpose	Debt_To_Income_Ratio	State	Home_Ownership	Monthly_Income	FICO.Range
2	81174	20000								MORTGAGE 6541.6
3	99592	19200								MORTGAGE 4583.3
4	80059	35000								MORTGAGE 11500
5	15825	10000	9975	9.99%	36 months	debt_consolidation	14.30%	KS		MORTGAGE 3833.3
6	33182	12000	12000	11.71%	36 months	credit_card	18.78%	NJ		RENT 3195
7	62403	6000	6000	15.31%	36 months	other	20.05%	CT		OWN 4891.6
8	48808	10000	10000	7.90%	36 months	debt_consolidation	26.09%	MA		RENT 2916.6
9	22090	33500	33450	17.14%	60 months	credit_card	14.70%	LA		MORTGAGE 13863.4
10	76404	14675	14675	14.33%	36 months	credit_card	26.92%	CA		RENT 3150
11	15867	7000	7000	6.91%	36 m					5000
12	94971	2000	2000	19.72%	36 m					3575
13	36911	10625	10625	14.27%	36 m					4250
14	41200	28000	27975	21.67%	60 m					4166.6
15	83869	35000	34950	8.90%	36 m					9166.6
16	53853	9600	9600	7.62%	36 m					11250
17	21399	25000	24975	15.65%	60 m					5416.6
18	62127	10000	10000	12.12%	36 m					9000
19	23446	14000	13900.25	10.37%	60 m					4333.3
20	44987	10000	10000	9.76%	36 m					2733.3
21	17977	5200	5175	9.99%	60 m					3750
22	86099	22000	21975	21.98%	36 m					6666.6
23	99483	30000	30000	19.05%	60 m					6250
24	28798	6500	6500	17.99%	60 m					4100
25	24168	17400	17400	11.99%	36 m					6833.3
26	10356	4000	4000	16.82%	60 months	vacation	13.71%	GA		MORTGAGE 4500
27	46027	7200	7200	7.90%	36 months	debt_consolidation	24.82%	TX		RENT 5416.6

loansData

lb. Explore the data sample.

Amount Requested

Amount Range	Count
[10, 7810]	196
(7810, 15610]	180
(15610, 23410]	178
(23410, 31210]	204
(31210, 39010]	196
(39010, 46810]	194
(46810, 54610]	201
(54610, 62410]	171
(62410, 70210]	167
(70210, 78010]	200
(78010, 85810]	197
(85810, 93610]	167
(93610, 101410]	165
(101410, ...)	84

Data

World Bank Indicators

Dimensions

- Date (year)
- Location
 - Region
 - Subregion
 - Country / Region
- Measure Names

Measures

- % of world average
- F: Deposit interest rate (%)
- F: GDP (curr \$)
- F: GDP per capita (curr \$)
- F: Lending interest rate (%)
- GDP per capita (weighted)
- H: Health exp (% GDP)
- H: Health exp/cap (curr \$)
- H: Life exp (years)
- P: Population (count)
- Rate spread (difference)
- Latitude (generated)
- Longitude (generated)
- Number of Records
- Measure Values

Pages

Columns AVG(F: GDP per capita (curr \$))

Rows Country / Region

Title Softpedia

Marks

- Automatic
- Color
- Size
- Label
- Detail
- Tooltip

Region

- Europe
- Middle East
- The Americas
- Oceania
- Asia
- Africa

Filters

- YEAR(Date (year)): 201...
- Region
- AVG(F: GDP per capita...

Legend

Show Me

For symbol maps try

- 1 geo dimension
- 0 or more dimensions
- 0 to 2 measures

Country	GDP per capita (curr \$)
Luxembourg	\$104.5K
Bermuda	\$89.7K
Norway	\$85.4K
Qatar	\$72.4K
Switzerland	\$67.6K
Denmark	\$56.3K
Macao SAR, China	\$52.0K
Australia	\$50.7K
Sweden	\$49.3K
United States	\$46.7K
Netherlands	\$46.6K
Canada	\$46.2K
Ireland	\$45.9K
Kuwait	\$45.4K
Austria	\$44.9K
Finland	\$44.1K
Japan	\$43.1K
Belgium	\$42.8K
Singapore	\$42.0K
Germany	\$39.9K
United Arab Emirates	\$39.6K
Iceland	\$39.5K

Bottom Navigation

- GDP per capita
- GDP per capita map
- GDP per capita by region
- GDP per Capita Dashboard
- Health spending vs life expecta...



```
analysis.R * gr * q * q * edge.pct.table * act * edge.cou >> □
```

```
209 Pipeline = act.norm(edge.count.table$Pipeline),
210 Triaging = act.norm(edge.count.table$Triaging),
211 Troubleshooting = act.norm(edge.count.table$Troubleshooting),
212 UX = act.norm(edge.count.table$UX)
213 )
214 colnames(edge.pct.table) <- c("To BUS", "To DS", "To INS", "To MON",
215 row.names =
216
217 write.cs
218
219
220 library(GGally)
221 library(ggplot2)
222 library(corrplot)
223
224 ggparcoord(data = pct.agree, columns = c(1,3:9), groupColumn = 2, sc
225   scale_fill_discrete(breaks = pct.agree$challenge[order(-pct.ag
226
227 corrplot(as.matrix(edge.pct.table),
228   is.corr=FALSE,
229   method="square",
230   addgrid.col = "white",
231   addCoef.col = "black",
232   addCoefasPercent = TRUE,
233   col=(colorRampPalette(c("pink", "#FFFFFF", "#AAAAAA"))(10))
234   sig.level = 0,
235   cl.pos="n")
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
```

225:89 (Top Level) ▾

Console C:/logan/titussurvey/analysis/ ↵

```
Warning message:
Warning: unable to connect securely with the server (possibly due to lack of date certificate files on your system). Falling back to using insecure URL for this mirror.
```

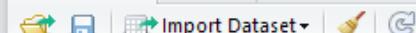
To learn more and/or disable this warning message see the "use secure download method for HTTP" option in Tools -> Global options -> Packages.

```
> source('C:/Logan/TitusSurvey/analysis/analysis.R')
```

```
Warning messages:
```

```
1: package 'GGally' was built under R version 3.0.3
2: package 'ggplot2' was built under R version 3.0.3
3: package 'corrplot' was built under R version 3.0.3
> ggparcoord(data = pct.agree, columns = c(1,3:9), groupColumn = 2, scale=
+   "globalminmax", order=c(4, 7, 6, 3, 9, 5, 1, 8)) +
+   scale_fill_discrete(breaks = pct.agree$challenge[order(-pct.agree$Troubleshooting)])
> |
```

Environment History



Global Environment▼

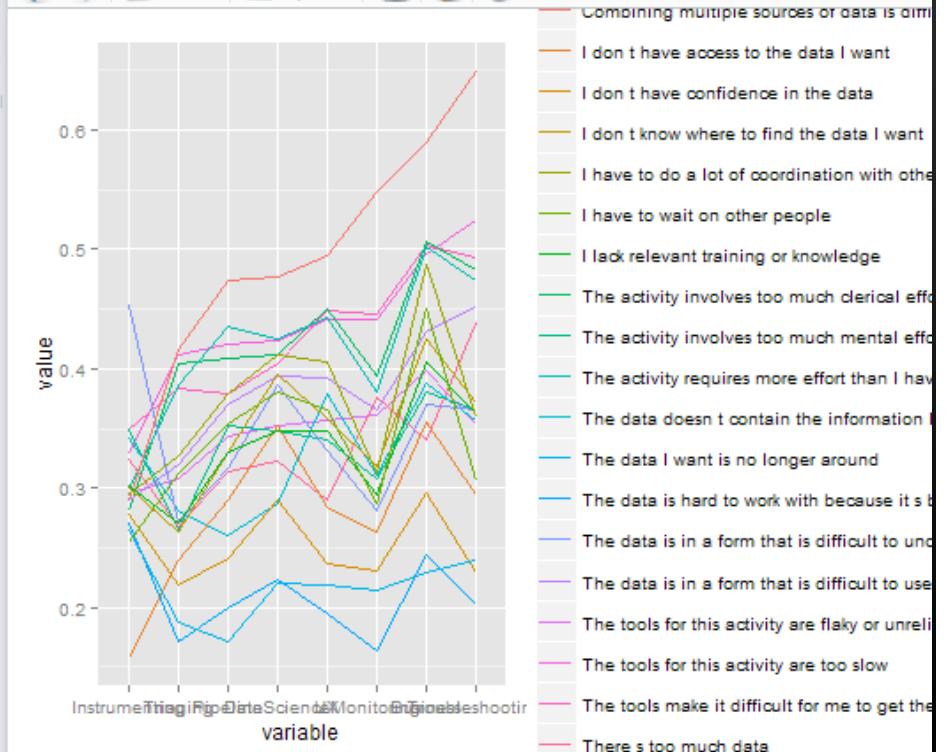
Data

act	1183 obs. of 9 variables
activities	1823 obs. of 9 variables

```
0 0 203 0 ...
0 0 0.19 0 ...
```

edge.pct.table	8 obs. of 8 variables
gr	1183 obs. of 8 variables
pct.agree	19 obs. of 9 variables

Files Plots Packages Help Viewer



IPython Notebook

setsim

Save

Notebook

Actions

New Open
Download ipynb Print

Cell

Actions

Delete
Format Code Markdown
Output Toggle ClearAll
Insert Above Below
Move Up Down
Run Selected All
Autoindent:

Kernel

Actions

Interrupt Restart
Kill kernel upon exit:

Help

Links

Python IPython
NumPy SciPy
MPL Sympy

Shift-Enter : run selected cell

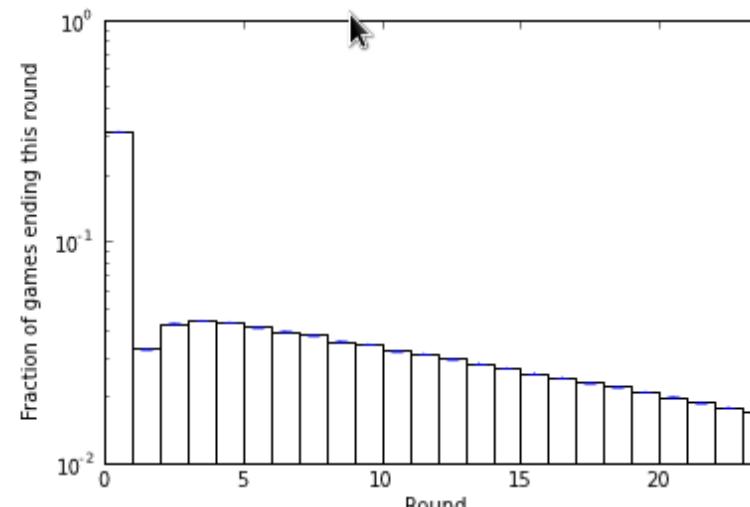
Ctrl-Enter : run in terminal mode

Ctrl-m h : show keyboard shortcuts

Now we build a histogram of the returned numbers, for plotting purposes. This takes a while.

```
In [10]: h = np.bincount([play() for i in range(1000000)])
```

```
In [11]: plt.bar(range(len(h)), h/float(np.sum(h)),
               width=1, yerr=np.sqrt(h)/np.sum(h),
               log=True, fill=False);
plt.xlim(0,len(h));
plt.xlabel("Round");
plt.ylabel("Fraction of games ending this round");
```



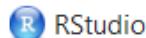
The numbers are a little hard to read because of the log scale, but about 30% of games finish without a "no sim" result. That's without the zero bin, and not on a log scale:

```
In [12]: plt.bar(range(1,len(h)), h[1:]/float(np.sum(h)),
```

1d. Scale out the score computation.

```
REGISTER gametrics.jar;
Player = LOAD 'player_data' AS
    (pid: chararray, region: chararray, signupDate: chararray);
Games = LOAD 'game_data' AS
    (gid: chararray, winner: chararray, winningScore: int, timestamp: chararray);
WinPlayers = JOIN Player BY pid, Games by winner;
Wins = GROUP WinPlayers BY pid;
WinSummary = FOREACH Wins
    GENERATE pid, COUNT($1), MAX($1.timestamp) AS latest
Skills = FOREACH WinSummary GENERATE gametrics.SkillScore(*);
STORE Skills INTO 'skills';
```

```
public class SkillScore extends EvalFunc<String>
{
    public Double exec(Tuple input) throws IOException {
        if (input == null || input.size() == 0)
            return null;
        // COMPUTE SKILL SCORE
        return skillScore;
    }
}
```



File Edit Code View Plots Session Build Debug Tools Help

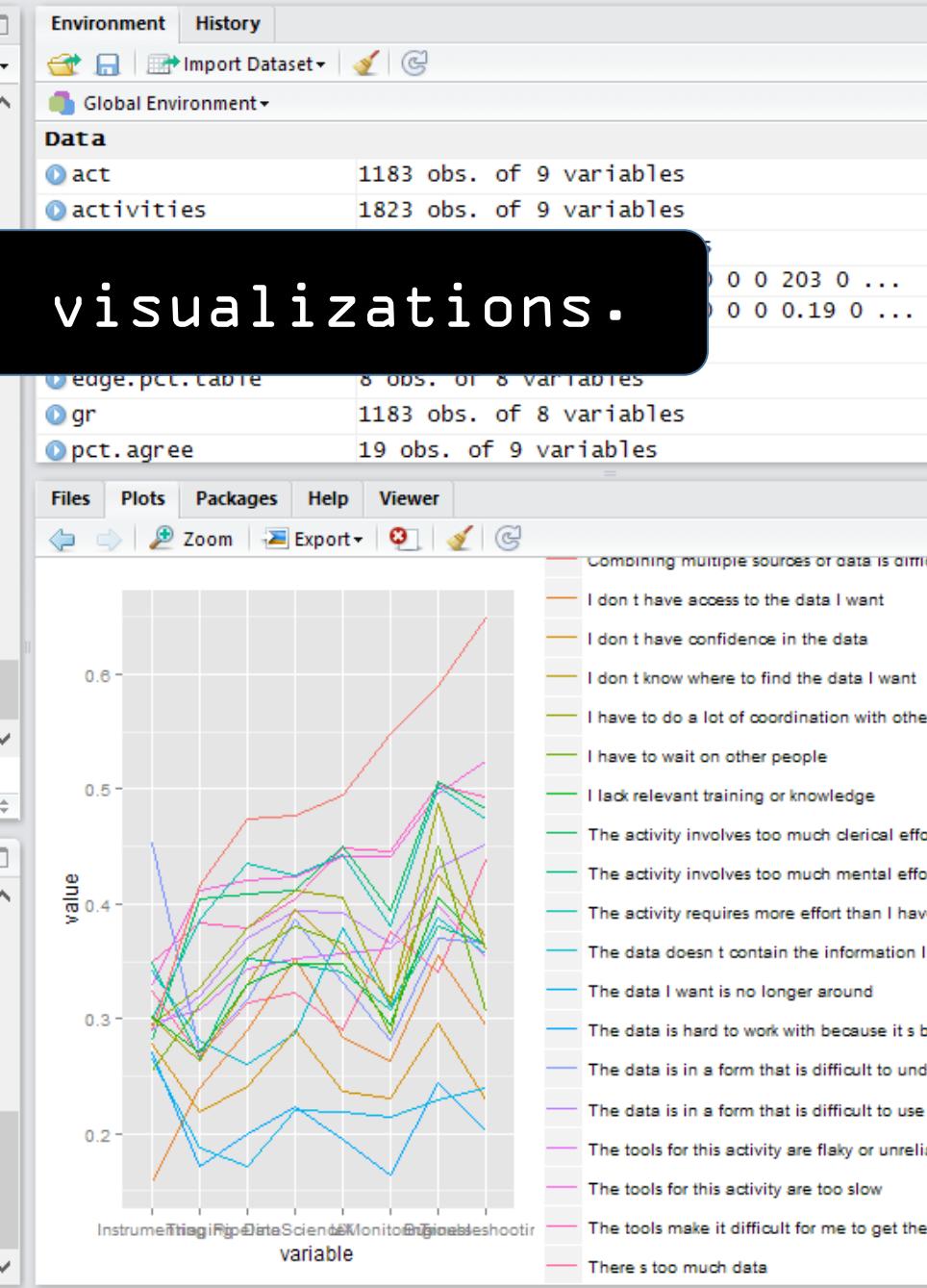


```
analysis.R * gr * q * q * edge.pct.table * act * edge.cou >
Source on Save | Run | Source | Environment | History | Import Dataset | Global Environment | Data | act 1183 obs. of 9 variables | activities 1823 obs. of 9 variables | edge.pct.table 8 obs. of 8 variables | gr 1183 obs. of 8 variables | pct.agree 19 obs. of 9 variables | Files Plots Packages Help Viewer | Zoom | Export | Combing multiple sources of data is difficult | I don't have access to the data I want | I don't have confidence in the data | I don't know where to find the data I want | I have to do a lot of coordination with other people | I have to wait on other people | I lack relevant training or knowledge | The activity involves too much clerical effort | The activity involves too much mental effort | The activity requires more effort than I have | The data doesn't contain the information I need | The data I want is no longer around | The data is hard to work with because it's broken | The data is in a form that is difficult to understand | The data is in a form that is difficult to use | The tools for this activity are flaky or unreliable | The tools for this activity are too slow | The tools make it difficult for me to get the data I need | There's too much data |
```

209 Pipeline = act.norm(edge.count.table\$Pipeline),
210 Triaging = act.norm(edge.count.table\$Triaging),
211 Troubleshooting = act.norm(edge.count.table\$Troubleshooting),
212 UX = act.norm(edge.count.table\$UX)
213)
214 colnames(edge.pct.table) <- c("To BUS", "To DS", "To INS", "To MON", "
215 row.names
216
217 write.cs
218
219
220 library(GGally)
221 library(ggplot2)
222 library(corrplot)
223
224 ggparcoord(data = pct.agree, columns = c(1,3:9), groupColumn = 2, sc
scale_fill_discrete(breaks = pct.agree\$challenge[order(-pct.agree\$
225 challenge)])
226
227 corrplot(as.matrix(edge.pct.table),
is.corr=FALSE,
method="square",
addgrid.col = "white",
addCoef.col = "black",
addCoefasPercent = TRUE,
col=(colorRampPalette(c("pink", "#FFFFFF", "#AAAAAA"))(10))
sig.level = 0,
cl.pos="n")
228
229
230
231
232
233
234
235
225:89 (Top Level) R Script

Console C:/logan/titussurvey/analysis/ ↵
Warning message: In download.file(url, destfile, method = "curl", quiet = !interactive(), progress = FALSE) :
 unable to communicate securely with the server (possibly due to a lack of date certificate files on your system). Falling back to using insecure URL for this mirror.

To learn more and/or disable this warning message see the "use secure download method for HTTP" option in Tools -> Global options -> Packages.
> source('C:/Logan/TitusSurvey/analysis/analysis.R')
Warning messages:
1: package 'GGally' was built under R version 3.0.3
2: package 'ggplot2' was built under R version 3.0.3
3: package 'corrplot' was built under R version 3.0.3
> ggparcoord(data = pct.agree, columns = c(1,3:9), groupColumn = 2, scale= "globalminmax", order=c(4, 7, 6, 3, 9, 5, 1, 8)) +
+ scale_fill_discrete(breaks = pct.agree\$challenge[order(-pct.agree\$
Troubleshooting)])
> |



1. Help the team decide whether to implement player matching (analytics).
2. If so, help the team deploy the feature.
3. Measure player reaction to the feature (flighting).
4. Monitor customer usage of the feature.

Dashboards

* Application Ins

4. Monitor customer usage.



My Favorites

No items in this folder.

Shared Favorites

No items in this folder.

My Dashboards

Customer Production Site

Fabrikam Extranet - Prod

Fabrikam Extranet - Test

Fabrikam Intranet - Prod

Fabrikam Intranet - Test

Shared Dashboards

No items in this folder.

Customer Production Site

Availability

Ticket Browsing Scenario

52.3%



* Availability

Landing Page

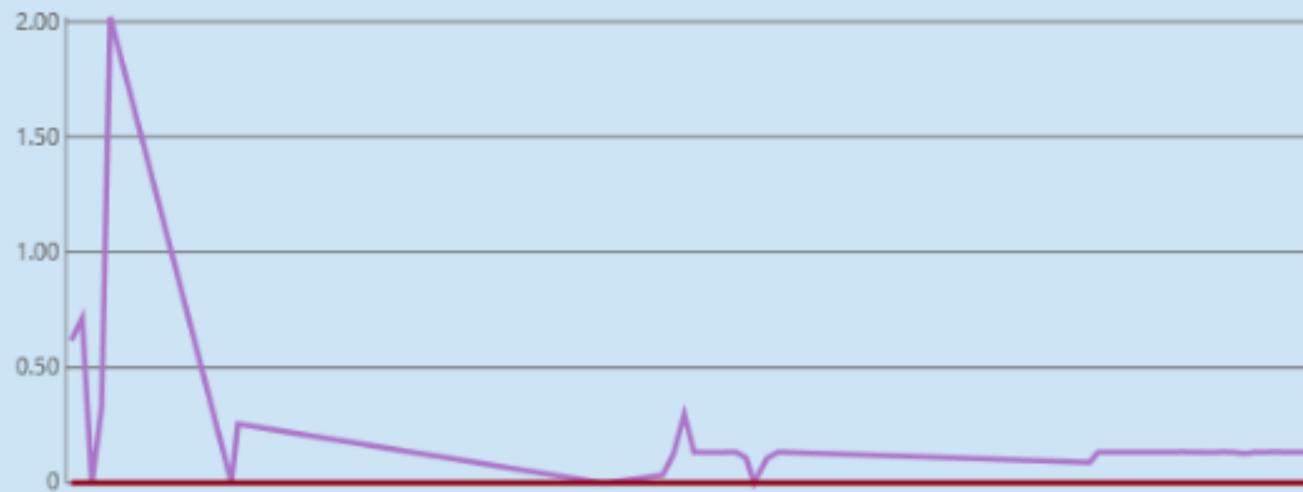
52.9%

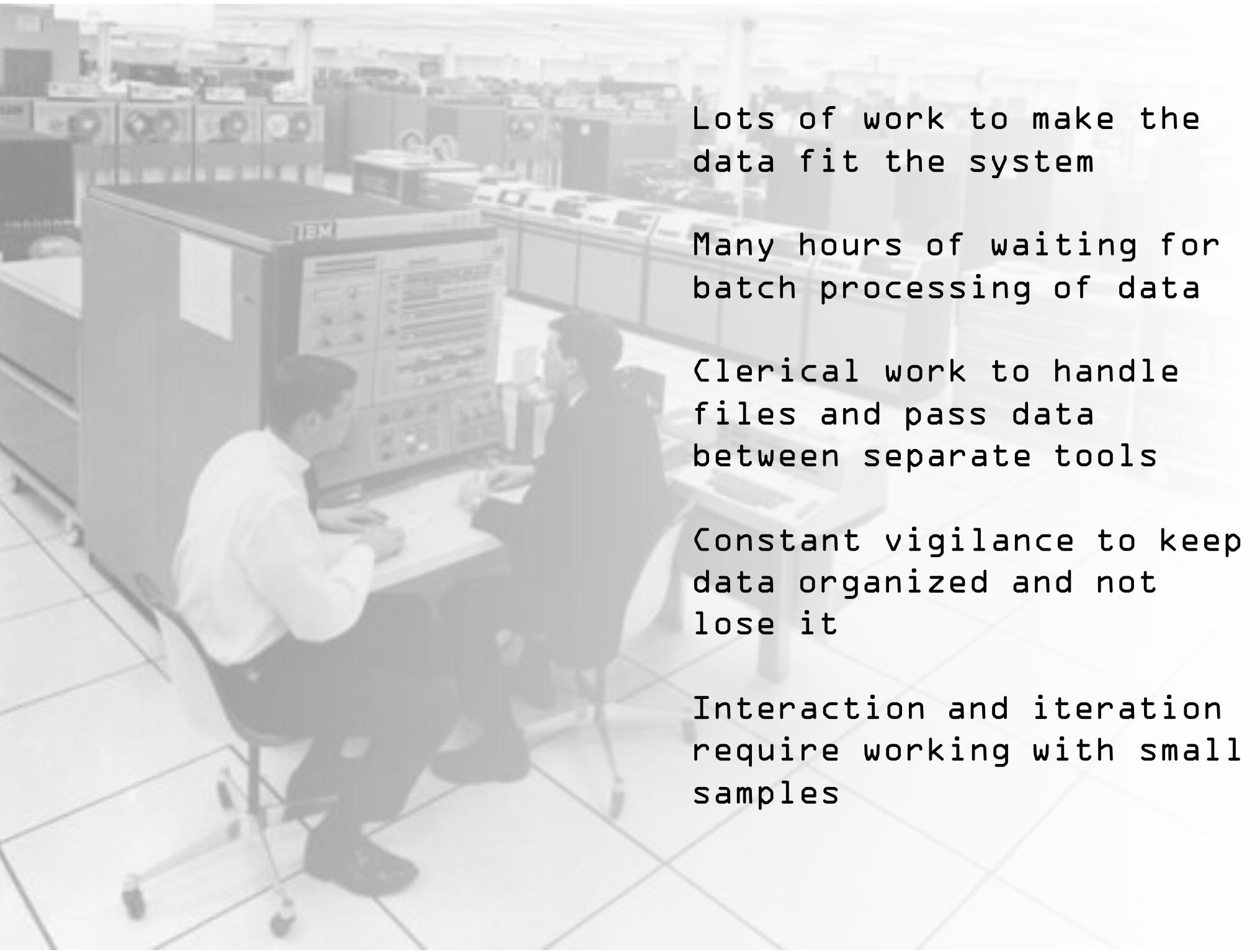


Exception & Request Rate

FF-Extranet-Prod

Request Rate (Total) (/s) Exception Rate (Total) (/s)





Lots of work to make the data fit the system

Many hours of waiting for batch processing of data

Clerical work to handle files and pass data between separate tools

Constant vigilance to keep data organized and not lose it

Interaction and iteration require working with small samples

Tempe

[Add data](#) [Annotate](#)[Restart](#)[Configure](#)[Stop](#)[Clone](#)[Delete](#)[Dashboard](#) [Live Editing](#)

explore loans

Created 10/19/2015 5:52:10 PM by rdeline

Last edited 10/19/2015 5:52:22 PM

LoansData 2,498 rows (100% of the total rows) 2 rows ignored due to parse errors

int	float	string	string	string	string
AmountRequested	AmountFundedByInvestors	InterestRate	LoanLength	LoanPurpose	DebtToIncome
20000	20000	8.90%	36 months	debt_consolidation	14.90%
19200	19200	12.12%	36 months	debt_consolidation	28.36%
35000	35000	21.98%	60 months	debt_consolidation	23.81%
10000	9975	9.99%	36 months	debt_consolidation	14.30%
12000	12000	11.71%	36 months	credit_card	18.78%
1000 - 35000	-0.01 - 35000	10.00% - 9.99%	36 months - 60 months	car - wedding	0% - 9.99%
$\mu=12405.02 \sigma=7802.13$	$\mu=12001.29 \sigma=7743.91$		Categorical		Categorical



Datasets

[BandGyro](#)[BandHeart](#)[EastsideHomesData](#)[LoansData](#)[NASDAQ](#)[StateAbbreviations](#)[Titanic](#)[WhFull](#)

Tempe Tutorials



Test



Demo

[explore loans](#)[New page](#)[New notebook](#)[New copy of the tutorials](#)

http://localhost:43664/twitter

Tempe: twitter

Tempe

Add data ▾ Annotate Restart Configure Stop Clone Delete Dashboard Live Editing

Datasets

- BandGyro
- EastsideHomesData
- NASDAQ
- StateAbbreviations
- Titanic
- WhFull

Tempe Tutorials

- Welcome!
- Writing scripts
- Working with data
- Visualizing data
- Query samples
- Machine learning: using TLC
- Sample: NASDAQ stock data
- Sample: White House visitors
- New page

Demo

- twitter
- stock
- loans
- easy
- signals
- heartbeat
- New page

twitter

Created 10/29/2015 3:32:38 PM by Rob DeLine
Last edited 10/29/2015 3:33:08 PM

```
var myConf = Badrish.GetTwitterConfig();
oauth_consumer_key s5ocdNJ2JnKi87ayC0Fpwto1f
oauth_consumer_secret D7OH70N1QUhCMUYdbIQuqZFx6Dqc1mKiuHljT1Sv7lgh4scXqo
oauth_token 2547031843-C3Dt114eXNSYkZx83Xj1kBkL3AK4F4E76dWMmrD
oauth_token_secret YIX6vSQoofKZp78Zf1j4hxGBiPSQbtmdVOCUjRiA5BDN2
twitter_keywords Office Microsoft, Surface Microsoft, Phone Windows 8, SQL Server, SharePoint, Bing, Skype, XBox, System Center
delay_ms 0
```

```
var inputObservable = TwitterObservable.TwitterStream.Create(myConf);

var twitterStreamable = inputObservable
    .ToStreamable(OnCompletedPolicy.EndOfStream(), PeriodicPunctuationPolicy.Count(1))
    .RepetitiveHoppingWindowLifetime(TimeSpan.FromSeconds(10).Ticks, TimeSpan.FromSeconds(1).Ticks);
```

ID	CreatedAt	UserName	ProfileImageUrl	Text
	2015/10/29 16:14:28 – 16:14:29	659825640011644928	10/29/2015 8:14:26 PM freewebcams.sluts	http://pbs.twimg.com/profile_images/648288720387354624/RV4Udn13_normal.jpg **** https:// **** chat? https:// #kik #c #skype https://

HALO 5

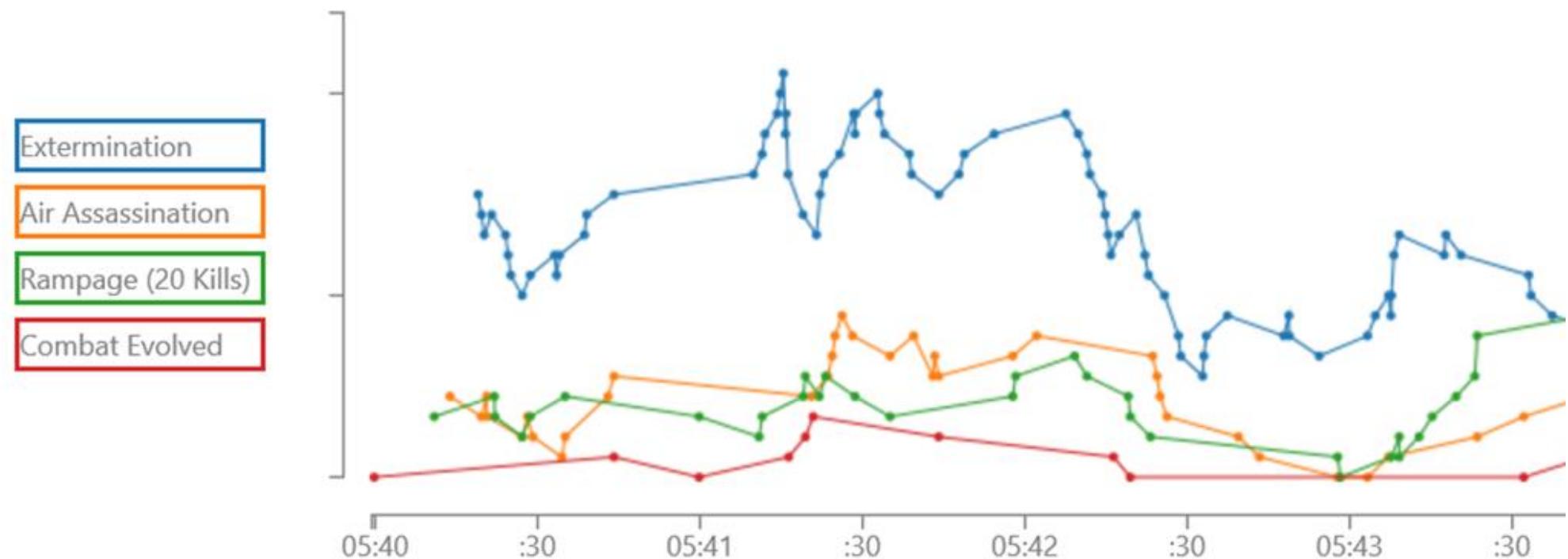
G U A R D I A N S

MULTIPLAYER BETA

12.29.14 - 1.18.15



```
var medalInteresting = medals.Where(val => goodMedalIds.ContainsKey(val.MedalId))
    .AlterEventDuration(TimeSpan.FromMinutes(1).Ticks)
    .GroupApply(val => goodMedalIds[val.MedalId], e => e.Count(),
                (input, aggregateData) => Tuple.Create(input.Key, aggregateData))
    .Vis("Medals per minute Count", "", true, false)
    .GroupBy(val => val.Item1, val => val.Select(e => e.Item2));
```







http://localhost:43664/signals



Tempe: signals



Tempe

Add data ▾ Annotate Restart Configure Stop Clone Delete Dashboard Live Editing

▼ Datasets

BandGyro
EastsideHomesData
NASDAQ
StateAbbreviations
Titanic
WhFull

signals

Created 10/26/2015 3:20:59 PM by Rob DeLine
Last edited 10/26/2015 3:21:21 PM

BandGyro

double double double double double double str
AccelerationX AccelerationY AccelerationZ AngularVelocityX AngularVelocityY AngularVelocityZ Ki

▼ Tempe Tutorials

Welcome!
Writing scripts
Working with data
Visualizing data
Query samples
Machine learning: using TLC
Sample: NASDAQ stock data
Sample: White House visitors
New page

▼ Demo

stock
loans
easy
signals
heartbeat

What Tempe gets right and wrong

- 👍 URL sharing and visualizations support team communication.
- 👍 A single query API supports many scenarios.
- 👍 The use of C# allows code to move between Tempe and the production system.
- 👍 Easy switching between monitoring and ad hoc queries supports “drilling in”.
- 👎 The need for scripting turns developers into gatekeepers.

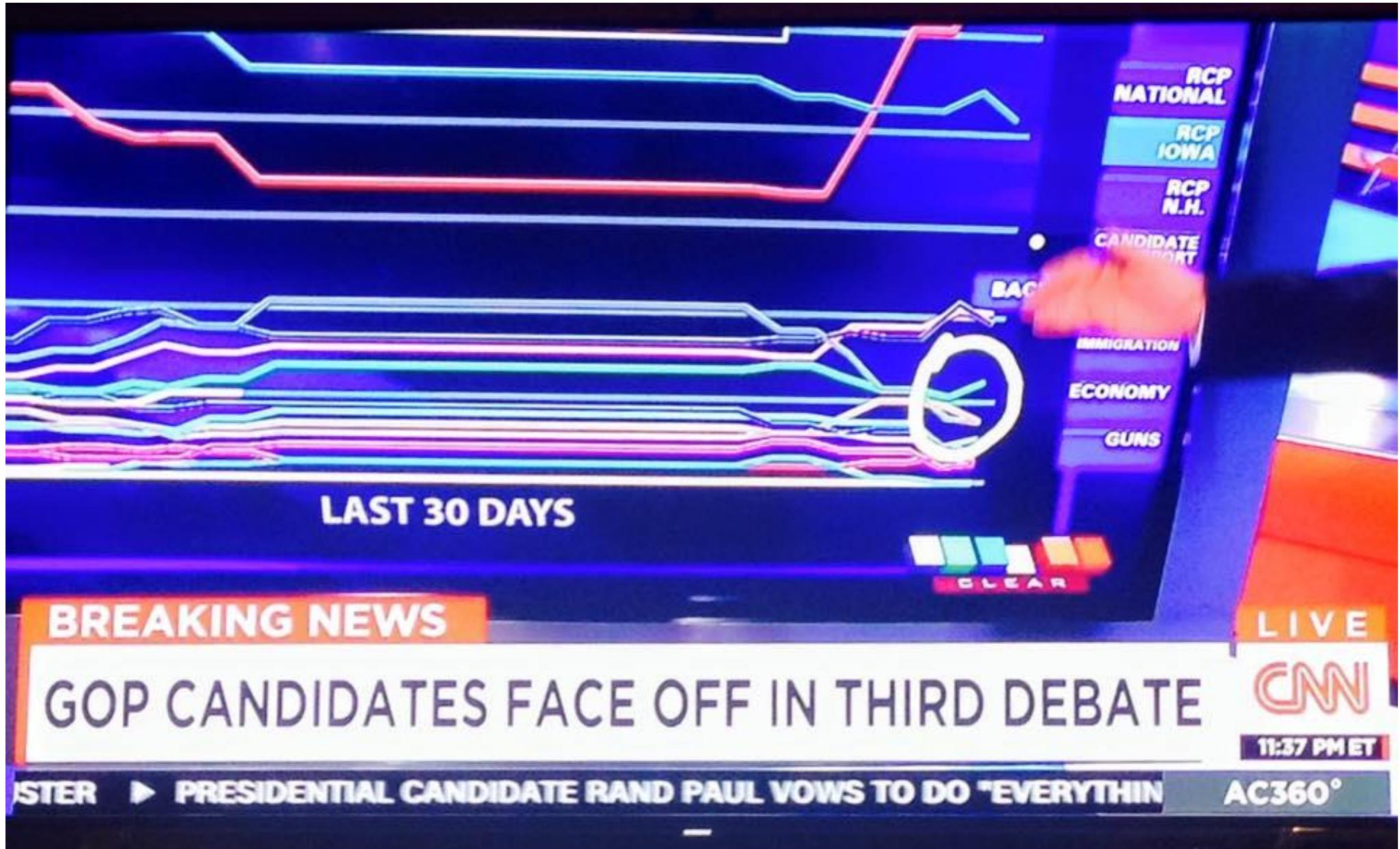
Originally we took the PMs and said here's a quick way to do this. They sort of tried to use it, but they weren't able to, so it fell back to me.

Profession	Primary		Secondary		Survey (n=1823)	
Business Operations (42)	BUS	20%	UX	19%		
Engineering						
Software Engineering (393)	TS	20%	MON	16%	BUS	Making business decisions
Software Development (241)	TS	20%	MON	17%	DS	Data science
Program Management (195)	TRI	16%	TS	15%	INS	Instrumentation
Service Engineering (71)	MON	20%	TS	19%		
Software Testing (23)	TS	21%	TRI	15%	MON	Monitoring
Applied Sciences (18)	DS	20%	PIP	17%	PIP	Data pipeline
Content Publishing (17)	UX	30%	TRI	19%		
IT Operations					TRI	Triaging work
Service Engineering (36)	TS	21%	MON	21%	TS	Troubleshooting
Program Management (31)	TRI	19%	BUS	17%		
Service Operations (28)	TS	24%	TRI	19%	UX	Improving the User Exp.
Software Development (9)	TS	18%	MON	18%		
Solution Management (9)	BUS	25%	TRI	19%		
Research (19)	DS	24%	BUS	18%		
Sales (55)	BUS	29%	UX	25%		
Services						
Support Engineering (127)	TS	28%	MON	19%		
Delivery Management (35)	MON	18%	BUS	16%		
Technical Delivery (34)	TS	21%	MON	19%		
Support Delivery (29)	TS	23%	MON	22%		
Services Leadership (7)	BUS	23%	TS	15%		

Barik, DeLine, Drucker, Fisher

"The Bones of the System: A Case Study of Logging and Telemetry at Microsoft"

MSR-TR-2015-79 (in submission)



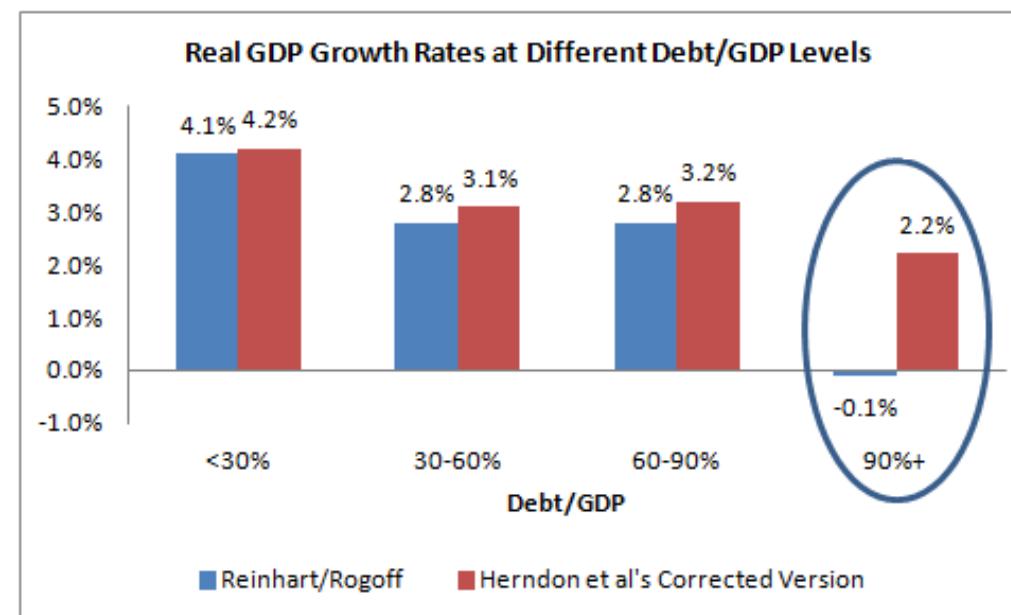
Wonkblog

Is the Excel

[A](#) [🖨️](#)By Brad Plumer [A](#)

One of the more come in for an e below.)

The paper in que Reinhart and Ke famous 2010 stu "Time of Debt," w economic growth when a country's p reaches 90 percent percent figure has c in the past few years as one big



as a pretty revolutionary idea to change America's post



1. Both experts and non-experts want to get answers from data.
2. The conclusions drawn from data are having an increasingly large effect on the world.

How can we help both experts and non-experts be confident in the conclusions they draw from data?

BIG DATA LANDSCAPE, VERSION 3.0

Exited: Acquisition or IPO

Infrastructure



Open Source



Analytics



Applications



Data Sources

