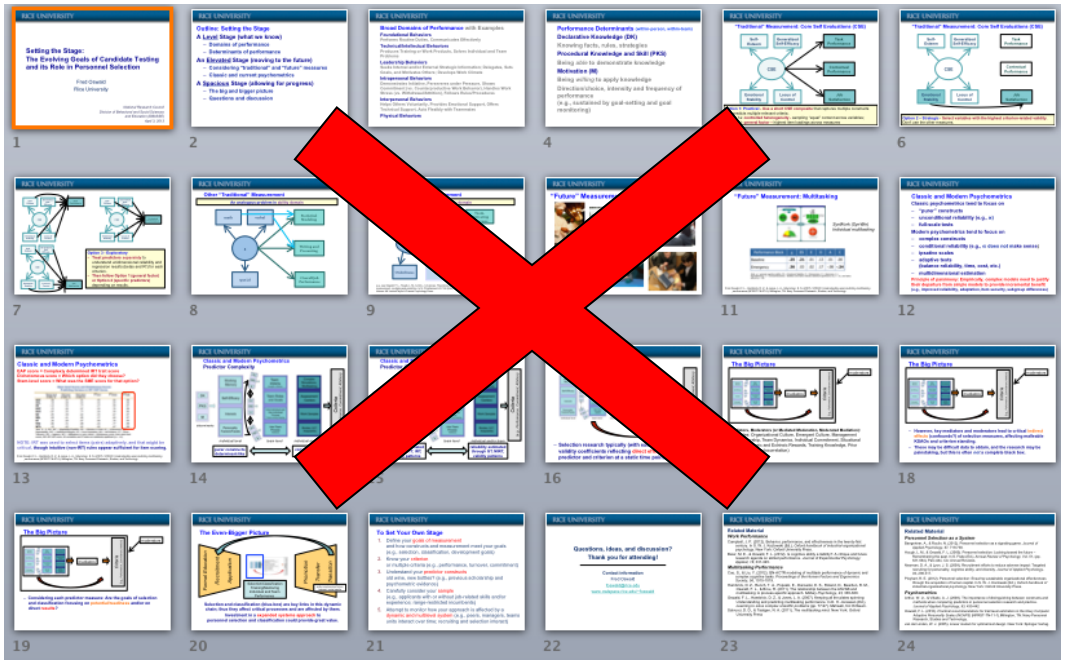
Under the Hood of Big Data in Personnel Selection

Fred Oswald
Rice University

11/7/14
Personnel Testing Council
of Southern California

What this talk <u>was</u> going to be…

Tying Big Data analytics to technological advances in selection, given increasing complexity in construct measurement, response processes, scoring, application...
…a pleasant 9-hour talk….



Instead, I want to start with the type of news we all have read:

Big Data Analytics HR Forum
www.**bigdatahr**forum.com/
Transform Your Organization With **Big Data**. As an experienced **HR** Professional you are facing a perfect storm of global macro competition, shrinking **HR** ...

Big Data for **HR Analytics** | eQuest
www.equest.com/solutions/**big-data**-for-**hr**/
What is **Big Data** for **HR**? Broken down to its most basic form, eQuest's **Big Data Analytics** program provides companies the advantages of utilizing the ...

Talent **analytics** and **big data**: the challenge for **HR ...** - CIPD
www.cipd.co.uk/**hr**.../t... Chartered Institute of Personnel and Development
Discover why talent **analytics** and **big data** are now must-have capabilities in **HR**. As the business world is transformed by the sheer volume, speed and ...

Analysing **analytics**: what does **big data** mean for **HR ...**
www.hrzone.com/feature/technology/...**analytics**...**big-data**...**hr**/142802
Jan 29, 2014 - Dr Tom Calvard analyses the history of **big data** and what it currently means for **HR** professionals.

[PDF] Talent **analytics** and **big data** – the challenge for **HR** - ...
www.oracle.com/.../talent-**analytics**-and-**big-data**-206... Oracle Corporation
Nov 1, 2013 - Talent **analytics** and **big data** – the challenge for **HR**. Talent **analytics** and **big data** are now must-have capabilities in **HR**. As the business world ...

What happened to me last year…

Reading a raft of stories about "big data" in HR and personnel selection…

…without much mention of the (DECADES) of personnel assessment expertise in HR and I-O psychology…

UNBOXED

Big Data, Trying to Build Better Workers

These are some of the startling findings of an emerging field called work-force science. It adds a large dose of data analysis, a k a Big Data, to the field of human resource management, which has traditionally relied heavily on gut feel and established practice to guide hiring, promotion and career planning.

Work-force science, in short, is what happens when Big Data meets H.R.

Adding insult to injury…

LEADERSHIP | 2/17/2013 @ 8:00PM | 15,195 views

Big Data in Human Resources: Talent Analytics Comes of Age

BigData Tells a Story, but We Have to Listen

Companies are loaded with employee, HR, and performance data. For the last 30 years we have captured demographic information, performance information, educational history, job location, and many other factors about our employees. Are we using this data scientifically to make people decisions? Not yet.

This, to me, is the single biggest BigData opportunity in business. If we can apply science to improving the selection, management, and alignment of people, the returns can be tremendous.

FYI, even statisticians note the analysis aspect of the 'big data' movement is not entirely new…

Data Science: The End of Statistics?

As I see newspapers and blogs filled with talk of "Data Science" and "Big Data" I find myself filled with a mixture of optimism and dread. Optimism, because it means statistics is finally a sexy field. Dread, because statistics is being left on the sidelines.

The very fact that people can talk about data science without even realizing there is a field already devoted to the analysis of data — a field called statistics — is alarming. I like what Karl Broman says:

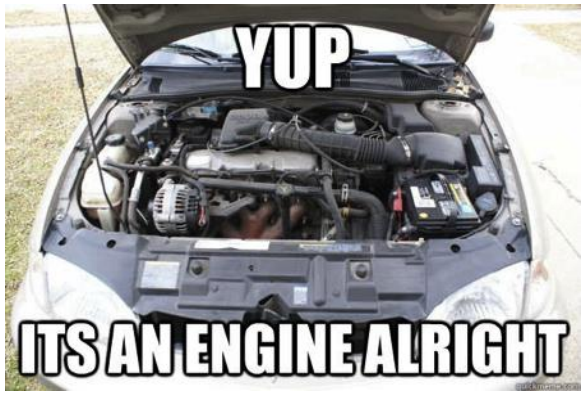
> *When physicists do mathematics, they don't say they're doing "number science". They're doing math.*
>
> *If you're analyzing data, you're doing statistics. You can call it data science or informatics or analytics or whatever, but it's still statistics.*

Well put.

Maybe I am just pessimistic and am just imagining that statistics is getting left out. Perhaps, but I don't think so. It's my impression that the attention and resources are going mainly to Computer Science. Not that I have anything against CS of course, but it is a tragedy if Statistics gets left out of this data revolution.

Thoughts While under the Hood…

1.



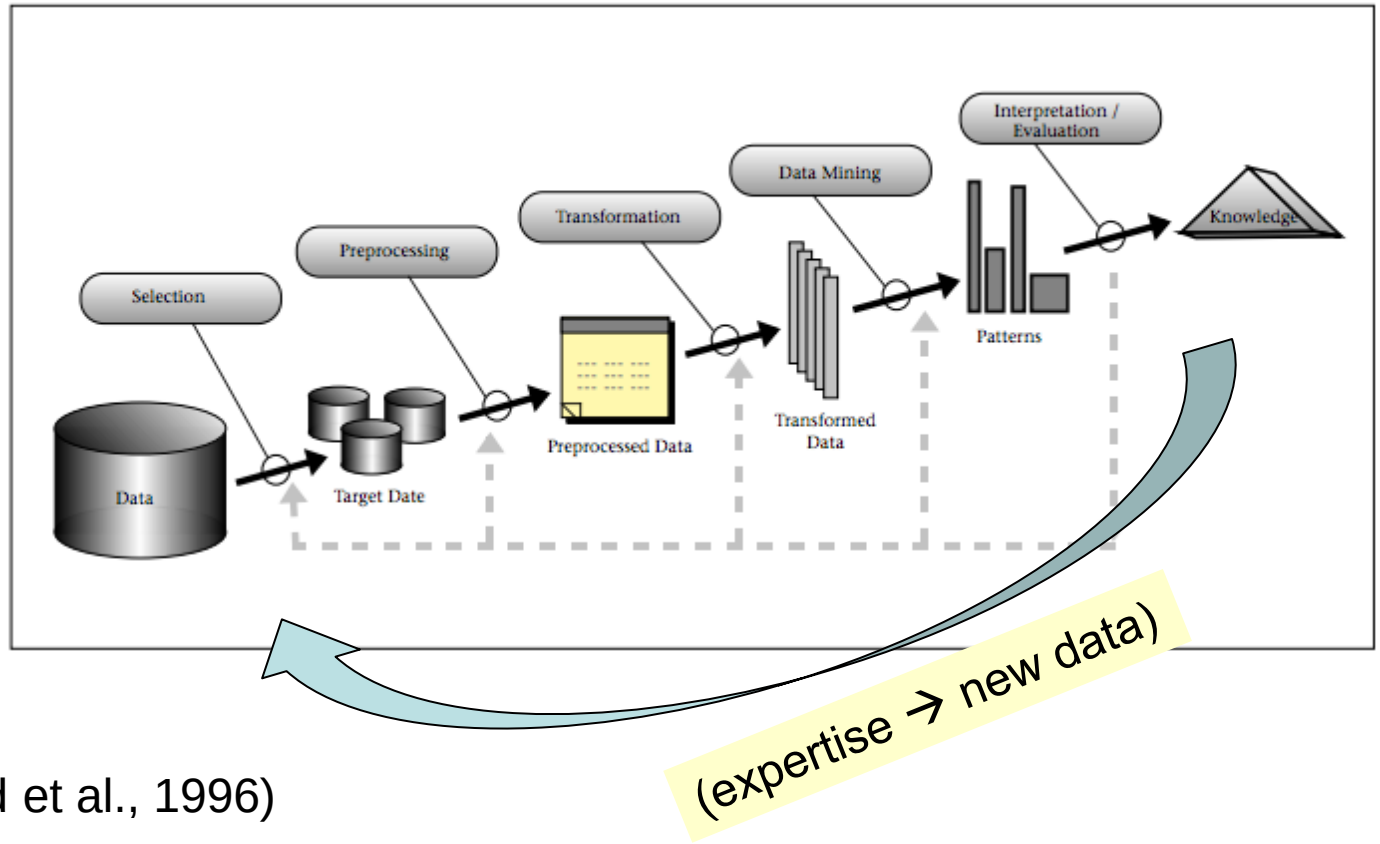
If Big Data only captures the 3Vs on an ever-expanding hard drive, it is useless.

Taylor 2013, HBR Blog: "We can amass all the data in the world, but if it doesn't help to save a life, allocate resources better, fund the organization, or avoid a crisis, what good is it?"

- **Companies indeed have an increasing amount of data on hand.**
- **Some of those data are *directly* relevant to selection (lots of online applications, screening tests).**
- **Other data might be *indirect*, but an argument can be made for selection (resume text mining).**
- **Still other data are indirect but difficult to justify *even if predictive* (e.g., time to complete an application online).**
- **…do something with it! ☺**

Thoughts While under the Hood…

1. **Indirect data could lead to developing new selection measures (given enough development time, testing time, $...).**



(Fayyad et al., 1996)

Let's Stop Asking "What is Big Data"

What Is Big Data? - Forbes
www.forbes.com/sites/lisaarthur/2013/08/15/what
Aug 15, 2013 - **Big data** is new and
very, very scary. No, wait. **Big data**
same old c

What is B
www.weboped
Big data d
unstructure
using tradi

What is Big Data?
www.sas.com/en_us/insig
Big data is a popul
growth and availab
And **big data** may be as importa
Big Data in Big Companies - Thr
- Macys.com

IBM - **What is big data?**
www.ibm.com/software/**data/bigdata/what-i**
Become well versed in **big data**
that should be applied to choosi
knowledge and confidence to build a ...

What is Big Data | Big Da
www.villanovau.com/resources/bi/wl
Big Data analysis can give
internal processes and bus
opportunities for growth.

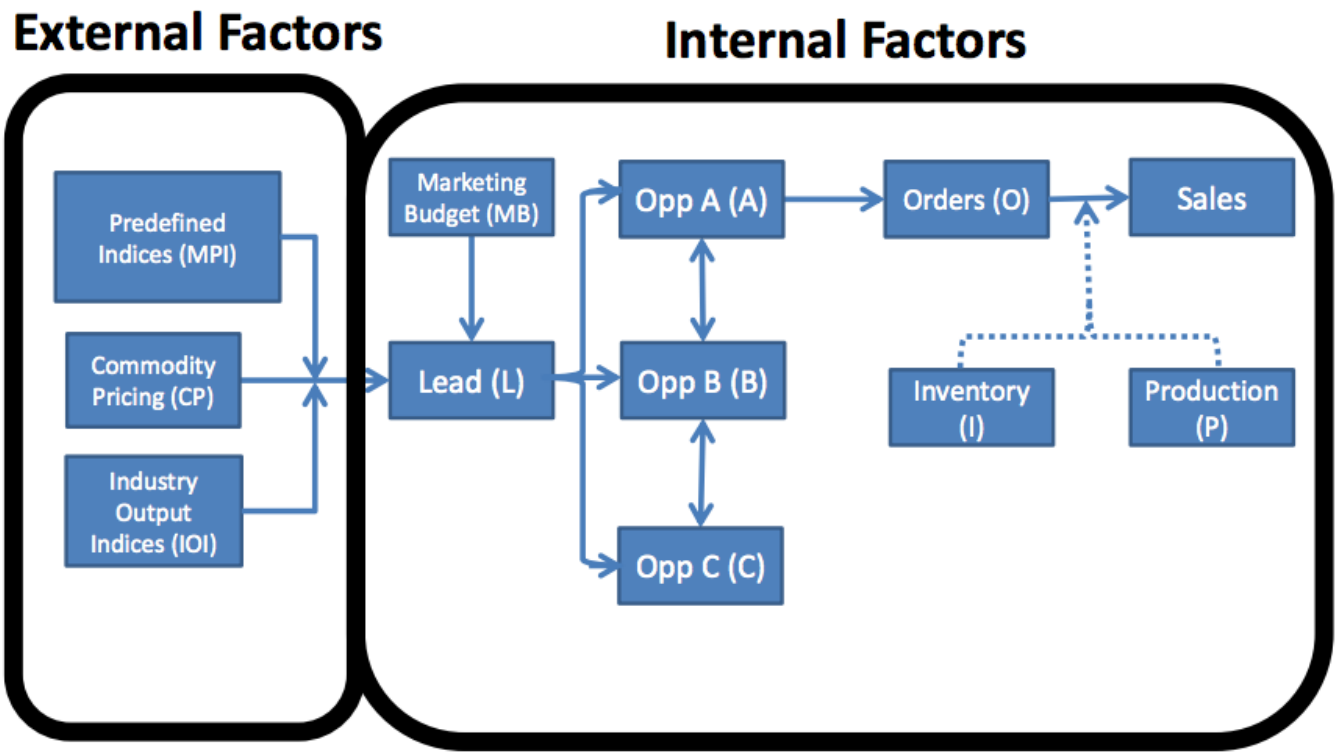
What is big data? - O'Rei
radar.oreilly.com/2012/01/**what-is-b**
Jan 11, 2012 - It's the hot t
does **big data** mean, and how can you exploit it? Strata chair
Edd Dumbill presents an ...

What is Big Data? - Zettaset
www.zettaset.com/index.php/info-center/**what-is-big-data/**
Gartner defines **Big Data** as high volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for ...

What Is Big Data? - datascience@berkeley
datascience.berkeley.edu/**what-is-big-d...** University of California, Berkeley
Sep 3, 2014 - "Big data." It seems like the phrase is everywhere. The term was added to the Oxford English Dictionary in 2013 External link and appeared in ...

What Is "Big Data," and Who's Collecting It? - Lifehacker
lifehacker.com/**what-is-big-data**-and-whos-collecting-it-15957... Lifehacker
Jun 25, 2014 - Dear Lifehacker, I've been hearing more and more about "big data." What is it, and is it something I should be worried about? Is this another ...

Big Data Is as Big Data Does
(build up by examples)

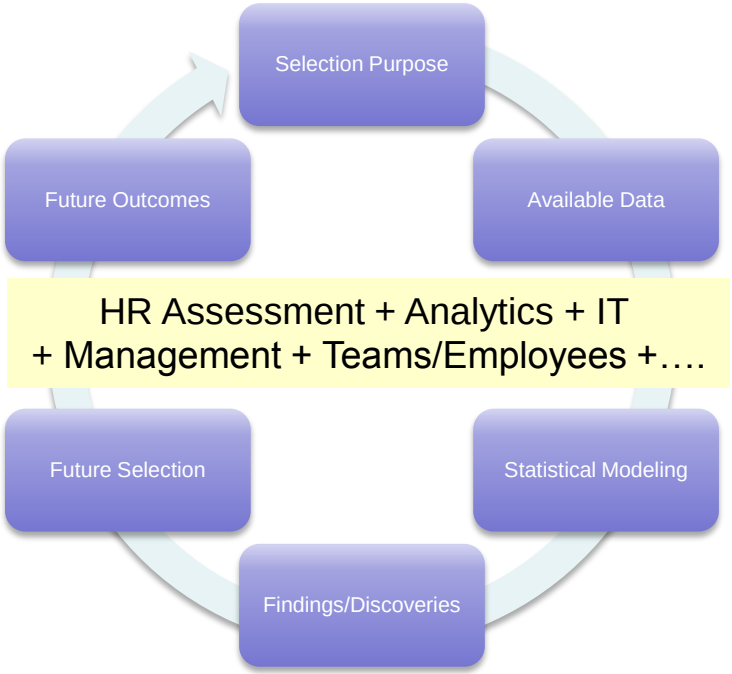


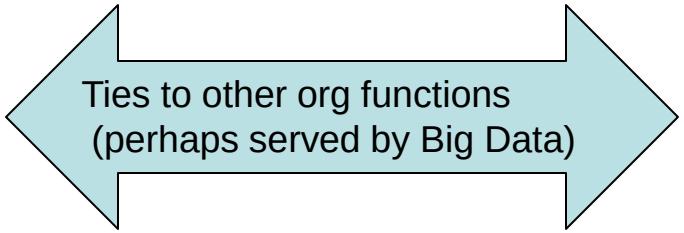
"Our company has *tons* of data….
…I'm embarrassed for other companies….
…but no one has time to analyze data….
…what if we don't find anything?"

Thoughts While under the Hood…

2.

Big Data analytics provides reasons/opportunities to collaborate – if there is a culture for that. (e.g., Kantrowicz, 2014, 25% of 1,406 HR professionals satisfied with how talent data are managed)





Thoughts While under the Hood…

3.

Big Data +
"Knowing Where to Cut"
= Data

Say you have 1,000,000,000,000s of data points for 1000s of employees, monitoring every millisecond of the day.

You could model performance dynamics and what applicant data predict it….how?

e.g., team vs. individual activities for selection, training, real-time intervention

Thoughts While under the Hood…

3.

Big Data +
"Knowing Where to Cut"
= Data

Even with an ambitious goal for prediction, you likely need to summarize the data.

By manager, department, team, season (hello, SQL…)

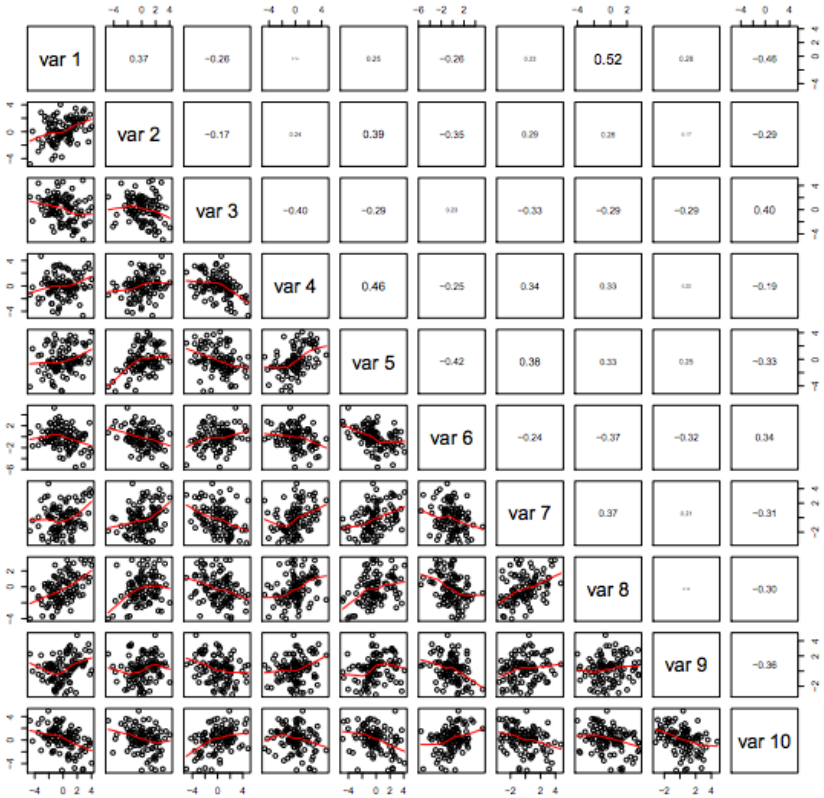
Can borrow some "strength" across the data, but largely we're in some ways back to traditional data analysis.

Big Data = data/visualization, not only analytics

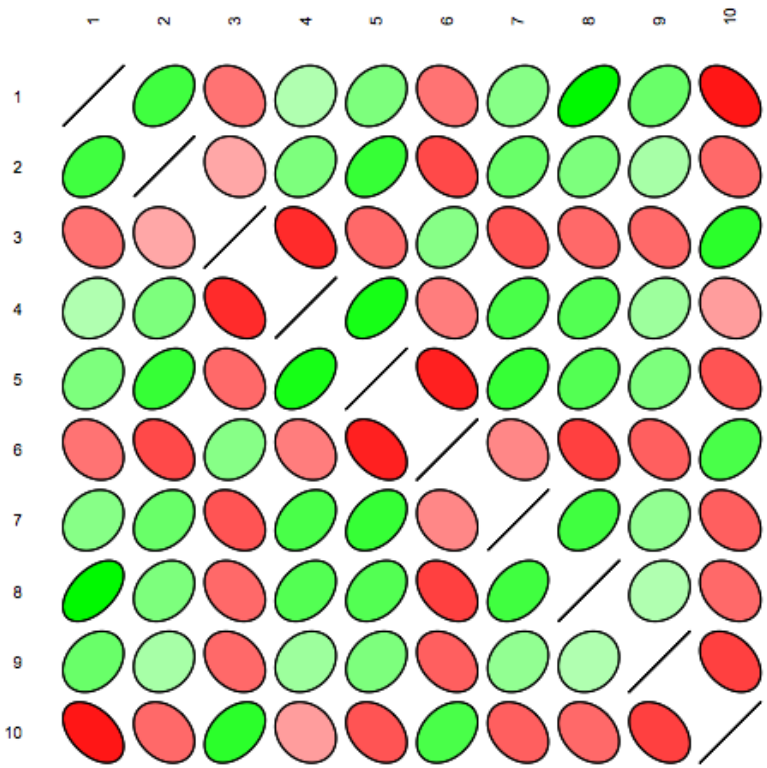
(e.g., Bersin 2013, 86% of 435 US/Canadian orgs only using dashboards, simple analytics)

Big Data Trend: More Tools for Data Visualization (caution: the eyes can deceive, need stability/replication)

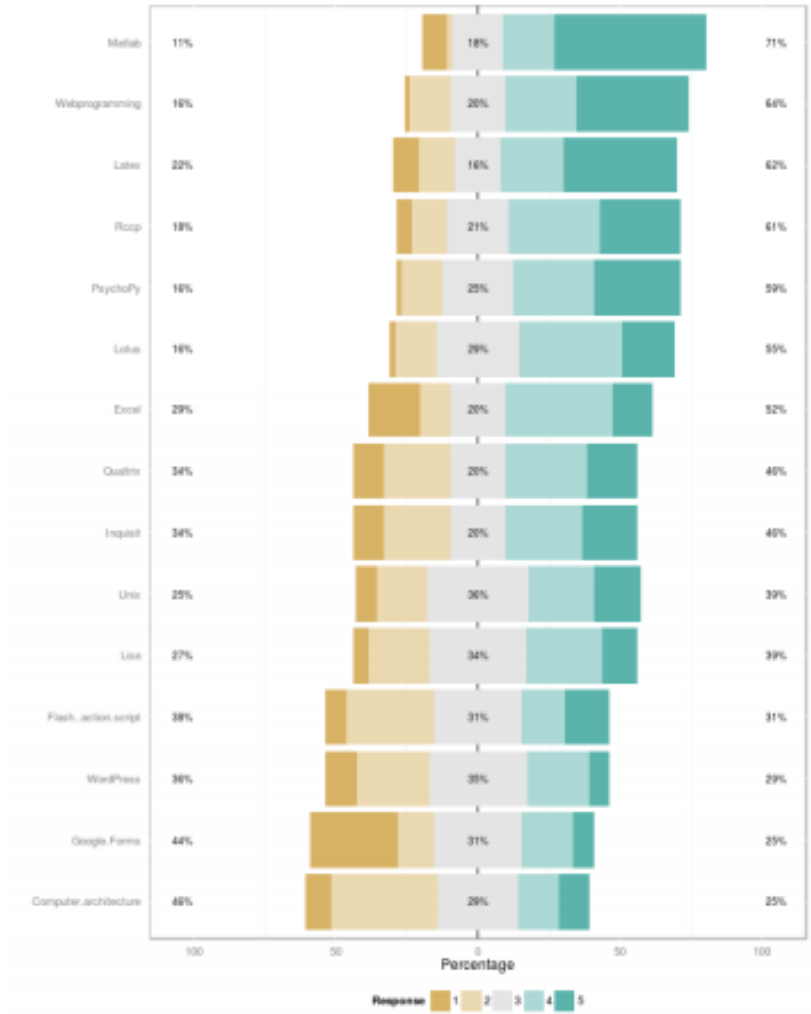
Bivariate Scatter Plots



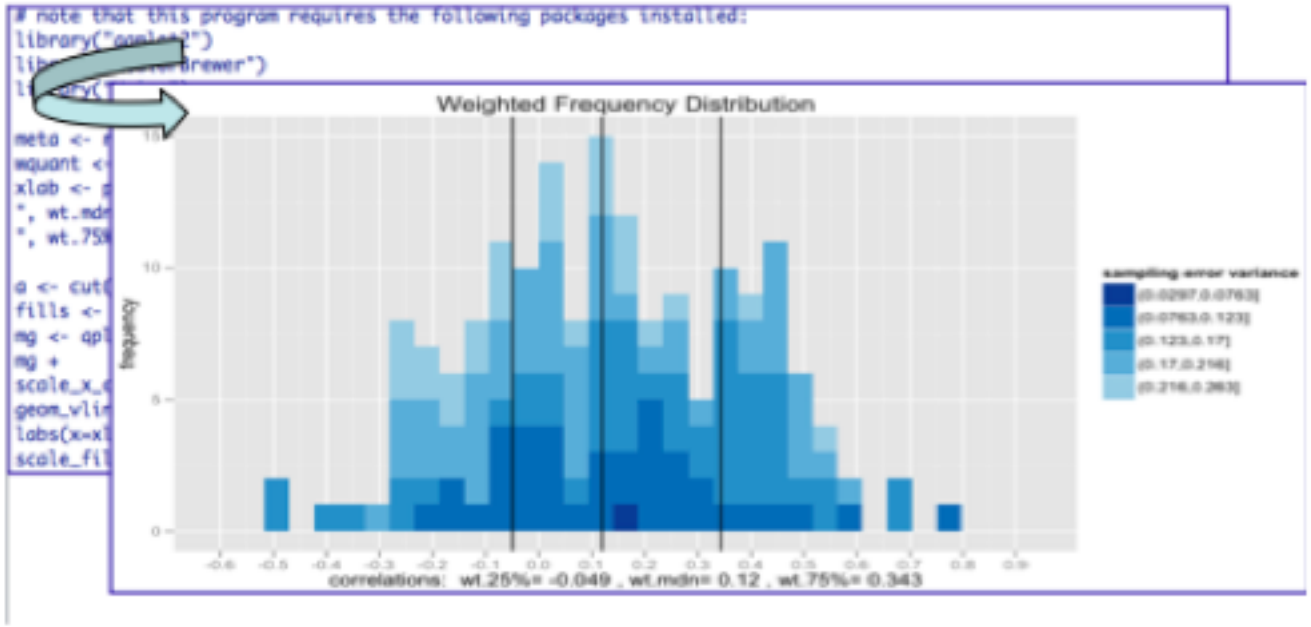
Correlation "Heat Map"



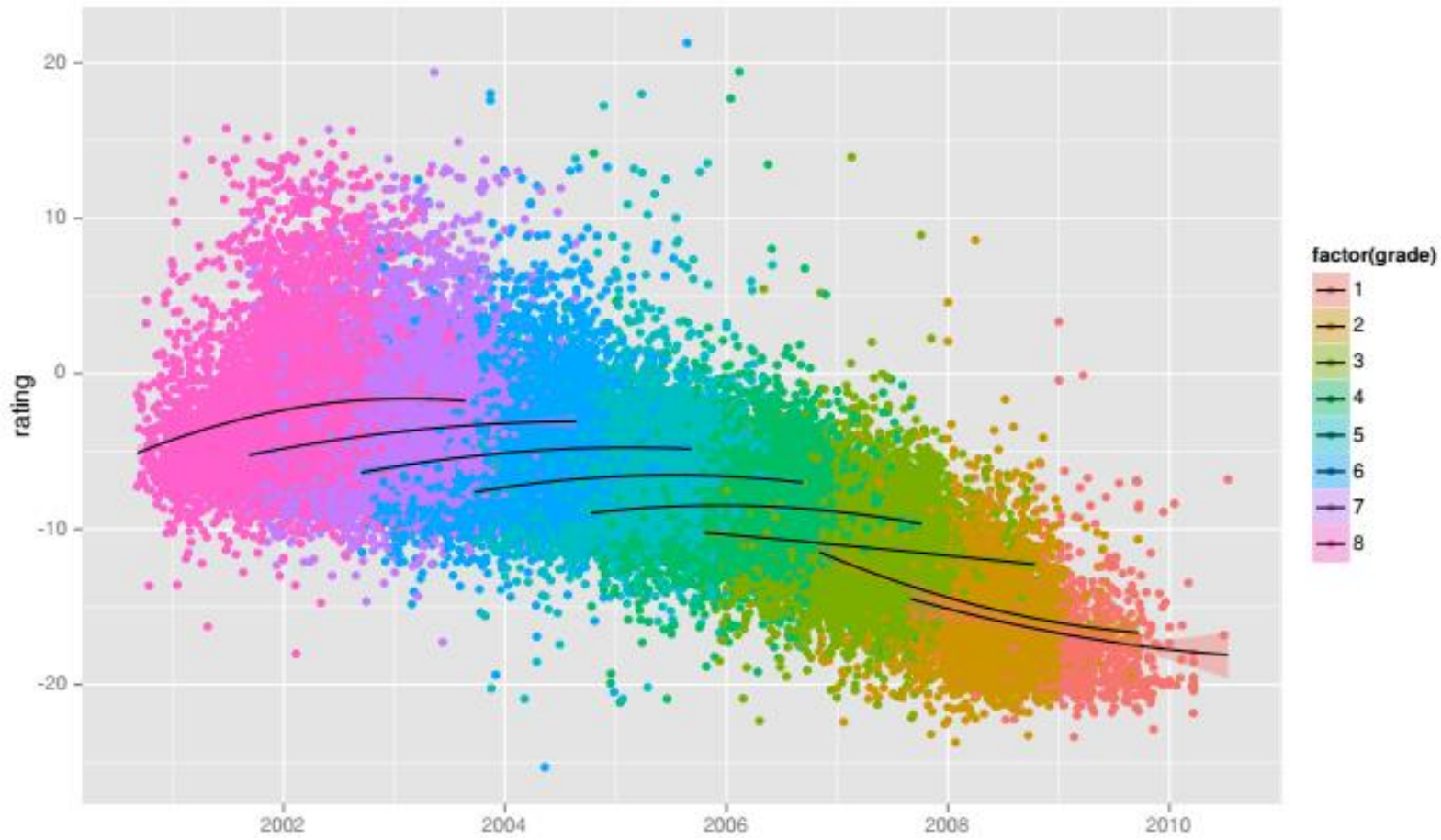
likert package (Bryer & Speerschneider, 2013)



Weighted Histogram for Meta-Analysis (Oswald & Ercan, 2012)



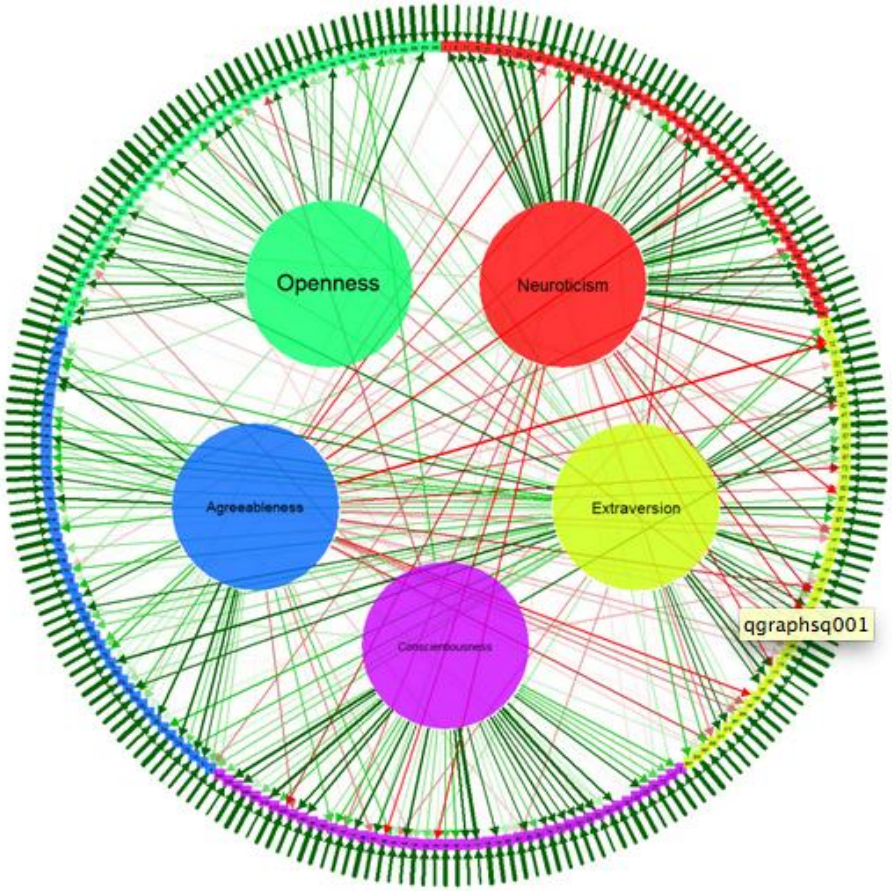
Nonlinear differential prediction (qgplot, R code)



pay grade

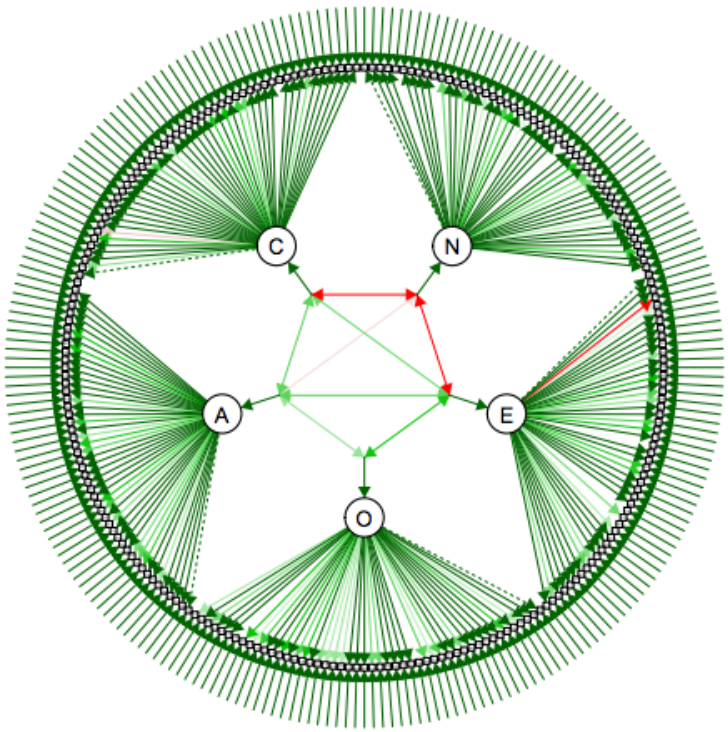
Adapted from http://www.sachaepskamp.com/files/dataVisualization.pdf

EFA with crossloadings



http://sachaepskamp.com/qgraph

CFA Model and Results



Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom (2012).

Model-implied correlations between items

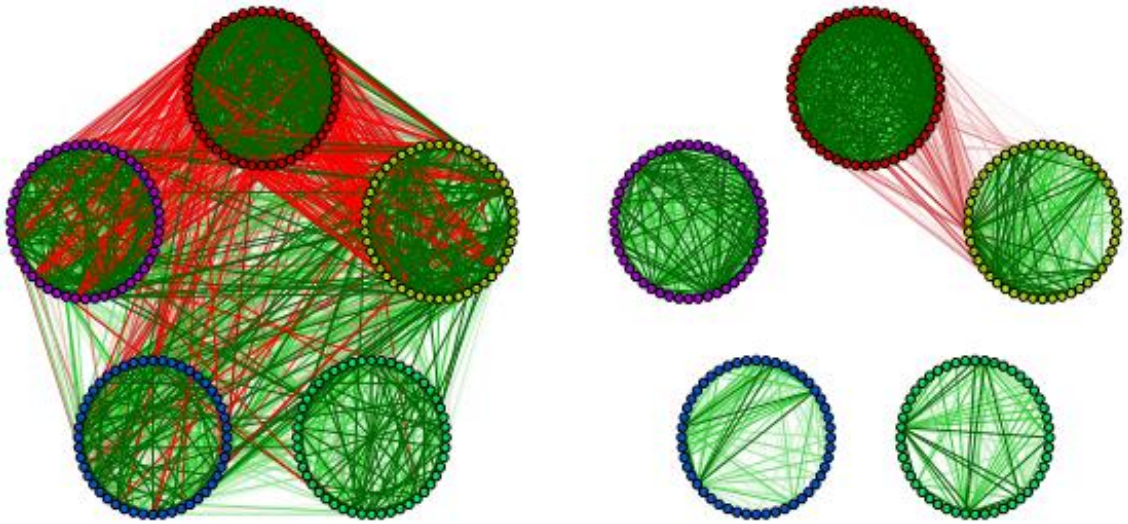


Figure 7: The observed correlations in the NEO-PI-R dataset (left) and the correlations that are implied by the model of Figure 6 (right).

Epskamp, Kramer, Waldorp, Schmittmann, & Borsboom (2012).

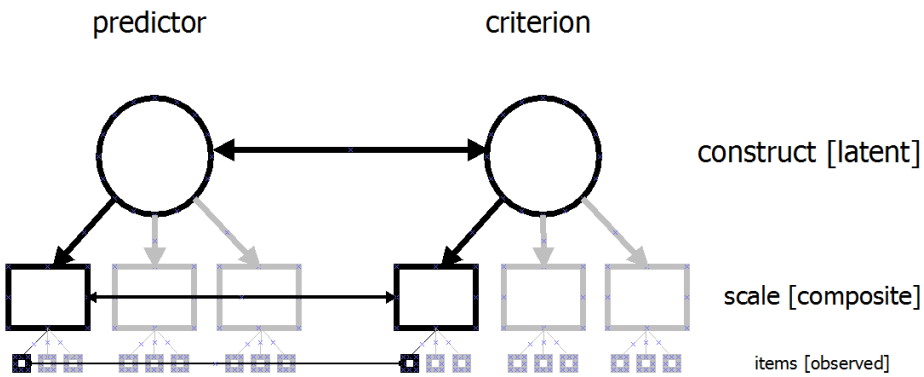
Thoughts While under the Hood…

4.

Big Data analytics is a form of engineering.

OK, so predictions are more *accurate* than ever…

- **If they *are* stable, do you *understand* them?**
  (why do we have reliability; why not just validity?)



Generally, our substantive research focuses on correlations, mean differences, etc. at the factor or composite level, not at the single-item level.

History tells us that item-level analyses can be hard to interpret (e.g., DIF). Surprises hard to find.

- **And what's missing…is it measurable/predictable?**
  (complexities of selection > your model, by a long shot)

Thoughts While under the Hood…

4.

Big Data analytics is a form of engineering.

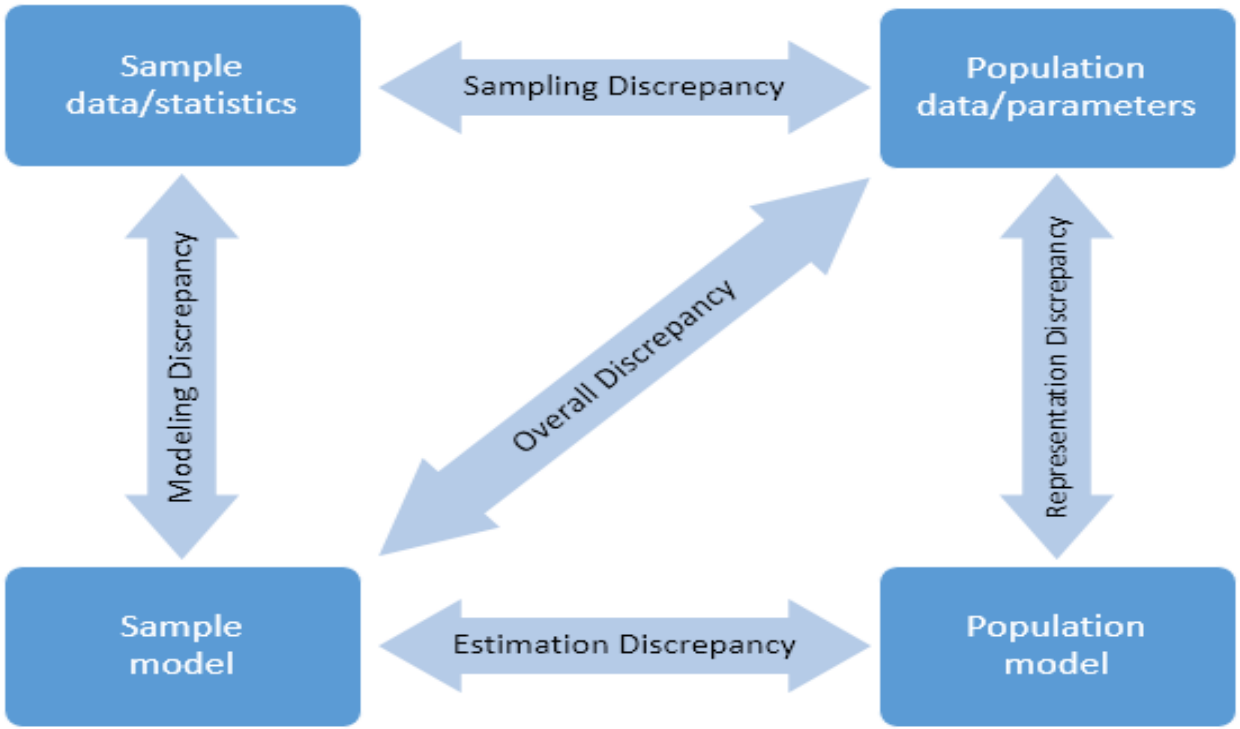
OK, so predictions are more *accurate* than ever…

- **If they are *not* stable across situations and over time, do you *care*?**

Sprinkles = data
Donut = model
What sticks when donut is shaken?

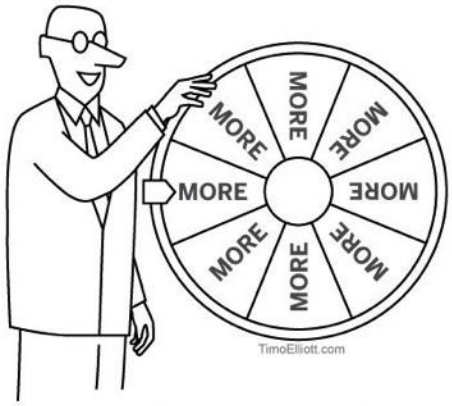
The Glazed Donut:
Big Data and Model Flexibility

Prediction = f(data, model)



Thoughts While under the Hood…

5.



"Let's take the Big Data prediction wheel for a spin..."

Improving your prediction using Big Data is not the same as improving.

Example: With Big Data, you predict applicants who will be your best-performing employees.

- **Questions: What performance dimensions? Over what time period? What measures/metrics?**
- **More questions: What if the best-performing employees are also more likely to turnover? Were you better off "worse"?**

Thoughts While under the Hood…

6.

Why would companies share their Big Data for the sake of science?

Large companies seek benefit from their local data.

Why share innovations in data collection, management, analytics, and Big Data insights?

For companies invested in HC analytics function, is the published research literature (even) less relevant?

- Literature uses "old" techniques, less-relevant data
- Organization has current data, can update frequently
- Complementary? (e.g., Bayes)

In what ways might sharing benefit everyone?

Closing the Hood and Asking You 4 Questions

1. **In what ways might our selection expertise in be supplemented or complemented by "big data" approaches?**

1b. …and how might it be supplanted?

Examples

- *Increasingly detailed employee and team data (…which data? …too much data?)*
- *Measurement (gamifying measures)*
- *Advances in predictive modeling (…what types of selection can--and cannot--be "Moneyball"-ed?)*
- *Closer ties between data, modeling and ROI (e.g., multilevel effects for climate, turnover)*

Closing the Hood and Asking You 4 Questions

2. How might social media* affect or (have affected) selection research and practice

***not merely using Facebook, of course…**

Examples

- *Organizations passively recruiting through information from online references (cybervetting; Zickar, 2013).*
- *Organizations can scour for talent displayed online (personal information about hobbies, listservs).*
- *People may network to form teams for hire (complementary skills, time-split dyads)?*

Closing the Hood and Asking You 4 Questions

3. **How might technological advances in selection affect other organizational practices? And legal issues?**

Examples

- *Selection and recruiting (e.g., what is an applicant?)*
- *Selection and adverse impact (e.g., en masse applicants vs. the 'good applicants')*
- *Selection and outsourcing (employees and HR functions)*

Closing the Hood and Asking You 4 Questions

4. **What phenomena relevant to selection might be more researchable with modern technologies?**

   Examples

   - *Multitasking (and distractibility)*
   - *Real-time teamwork/sociometrics (team badges)*
   - *Bottom-up job analysis from highly detailed data on job activity*

Parting Quotes

"Big Data can be a good thing if the analytics have some foundation in theory/previous research (not just throwing the data to the wall to see what sticks) and if there are practical (actionable) implications for the organization."

- Kyle Lundby
  I/O psychologist, LinkedIn comment, 6/6/13

"Some things might stick to the Big Data wall,
but they may not confess to how they got there."

- Fred Oswald
  I/O psychologist, Big Data enthusiast and skeptic
  Typed on the plane to ORD, 9/18/14

Thank you!

Fred Oswald
foswald@rice.edu