



This slide unintentionally left blank



The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

Chris Anders,
Wired, 06/23/08

Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

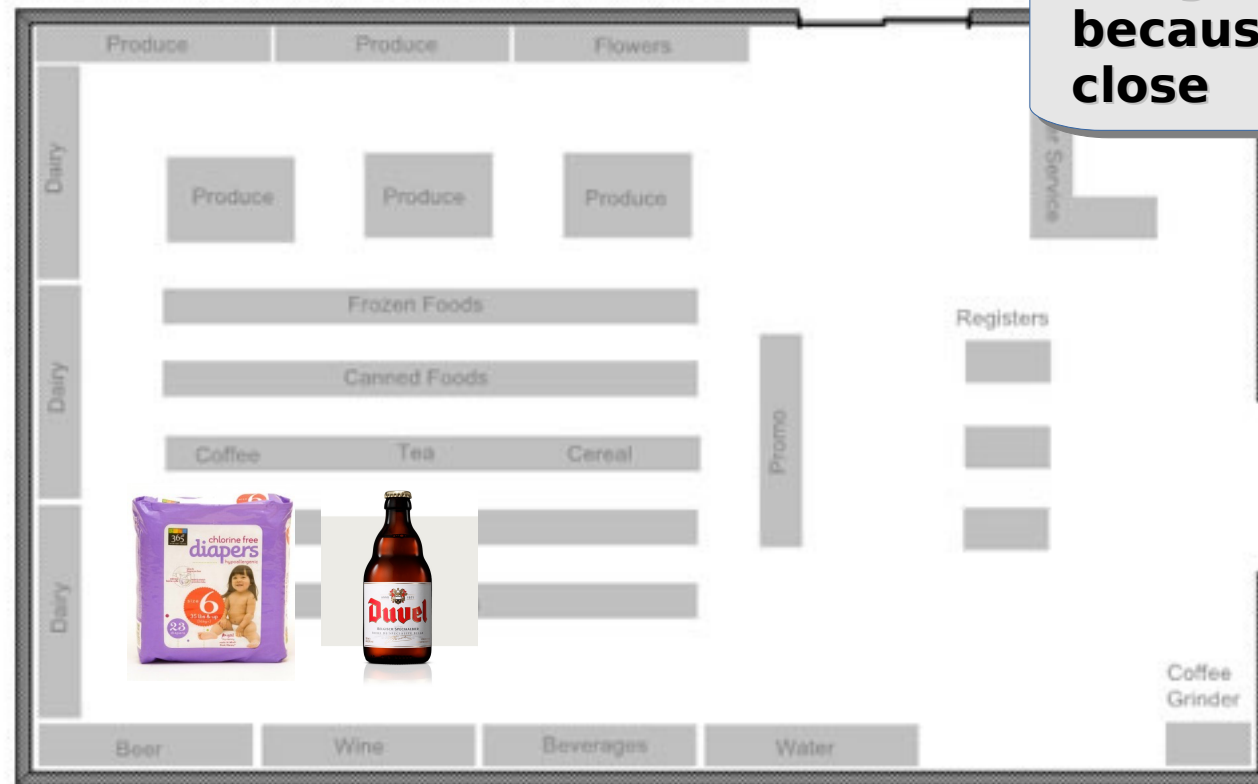
Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.

What does itemset mining tell us?



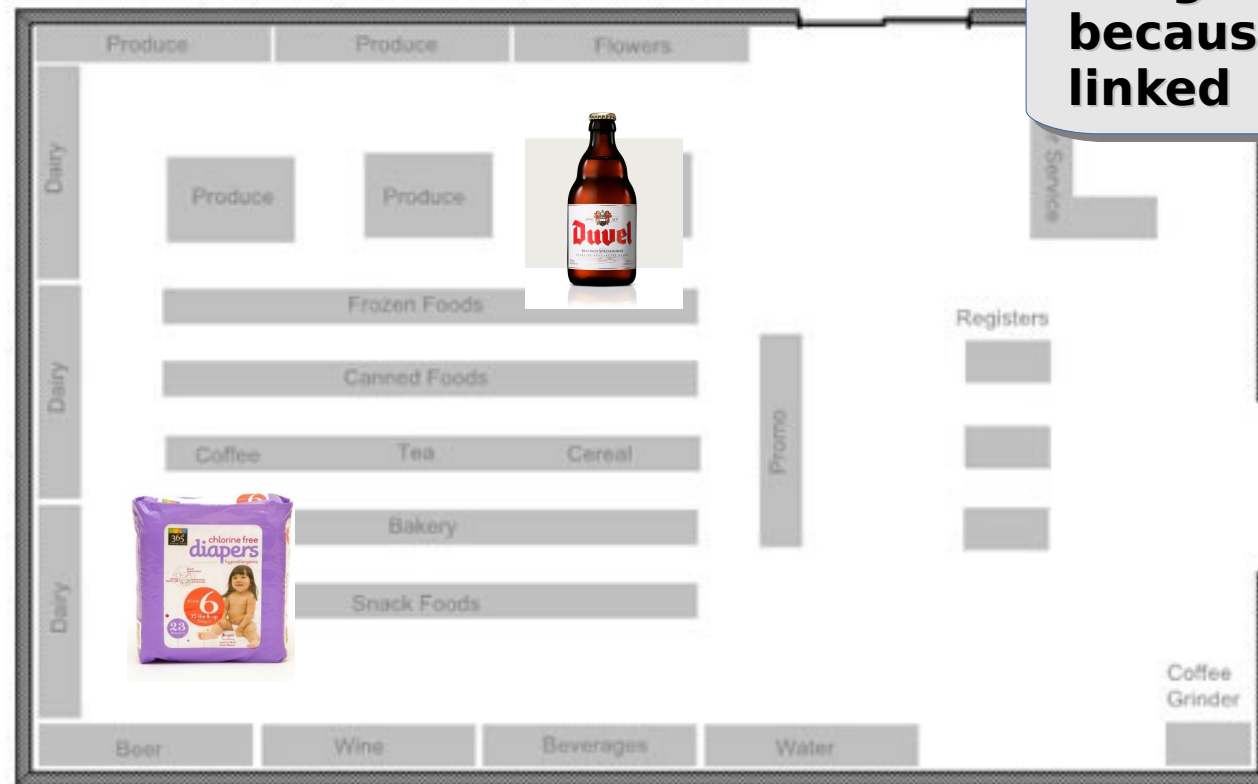
**Bought together
because they are
close**

What does itemset mining tell us?



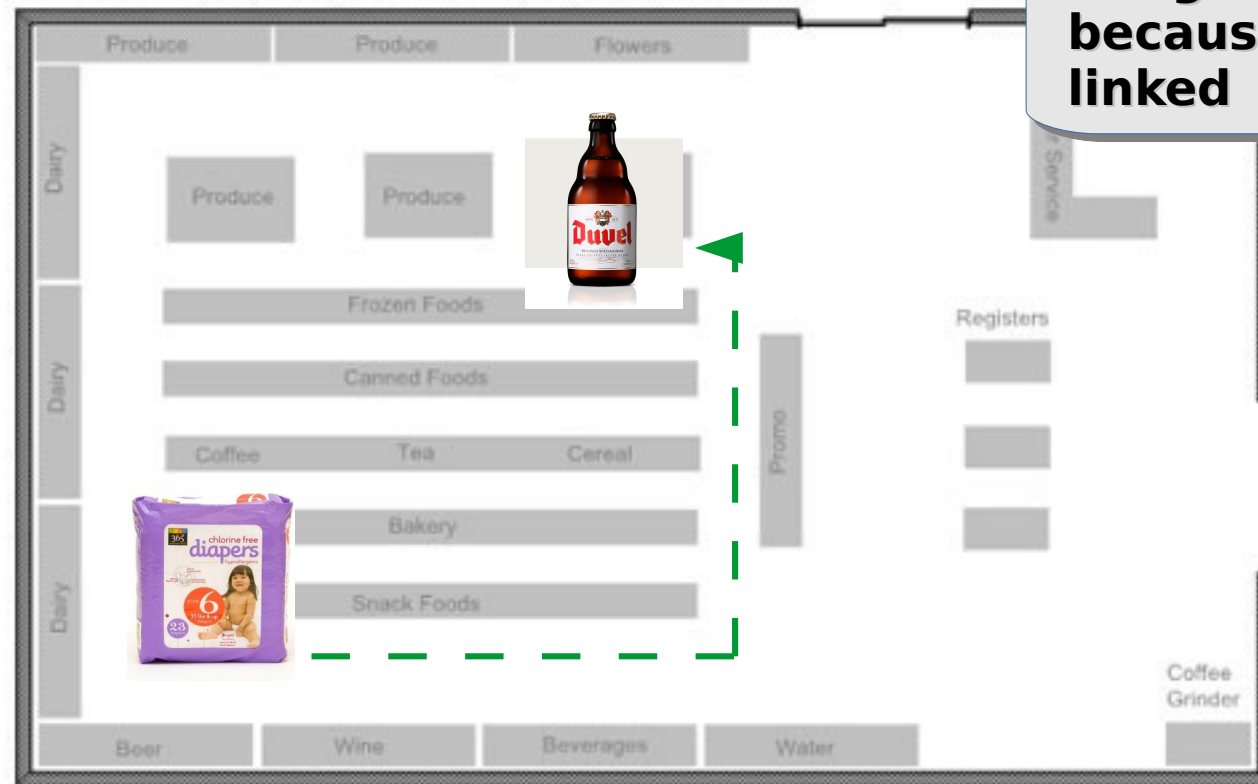
**Bought together
because they are
close**

What does itemset mining tell us?



**Bought together
because they're
linked**

What does itemset mining tell us?



**Bought together
because they're
linked**

More/new data

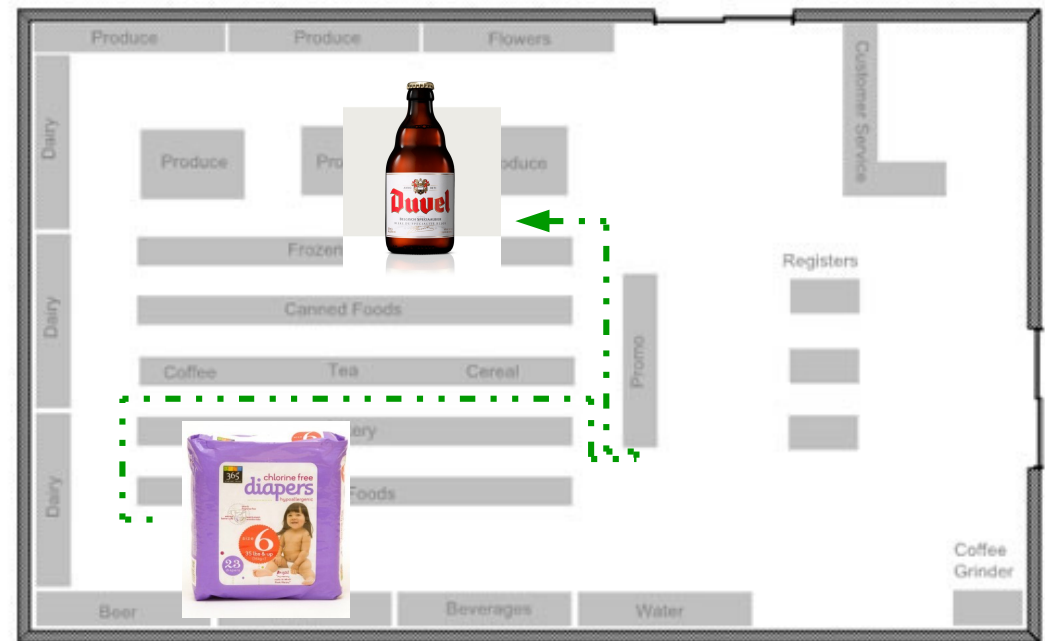
1) User **movements** in addition to basket

- Not enough
- Mobile self-check in?

2) Compare “rest” of basket to supermarket **layout**

- What if no rest?

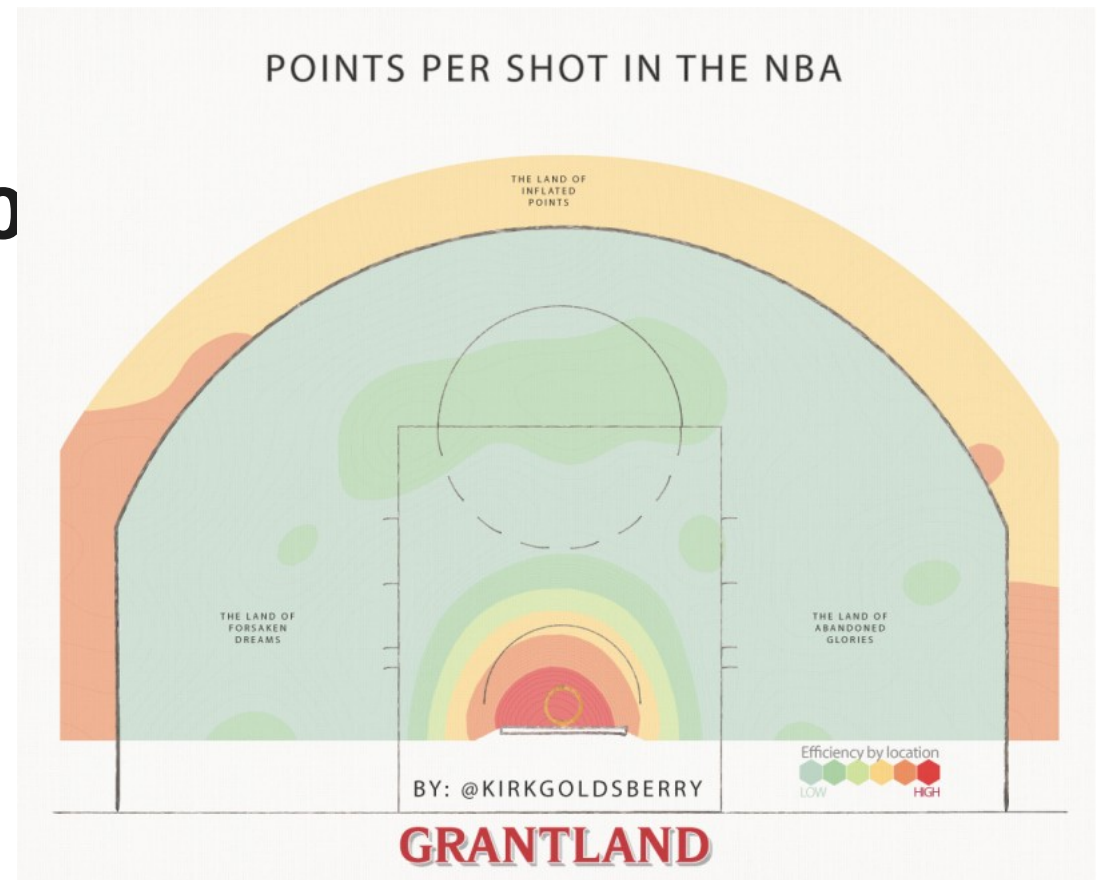
3) **Different** layouts



Changing tack - value of shots in the modern NBA

3 pointer = $1.5 \times$ 2 pointer \rightarrow 2P% should be at least $1.5 \times$ 2P%

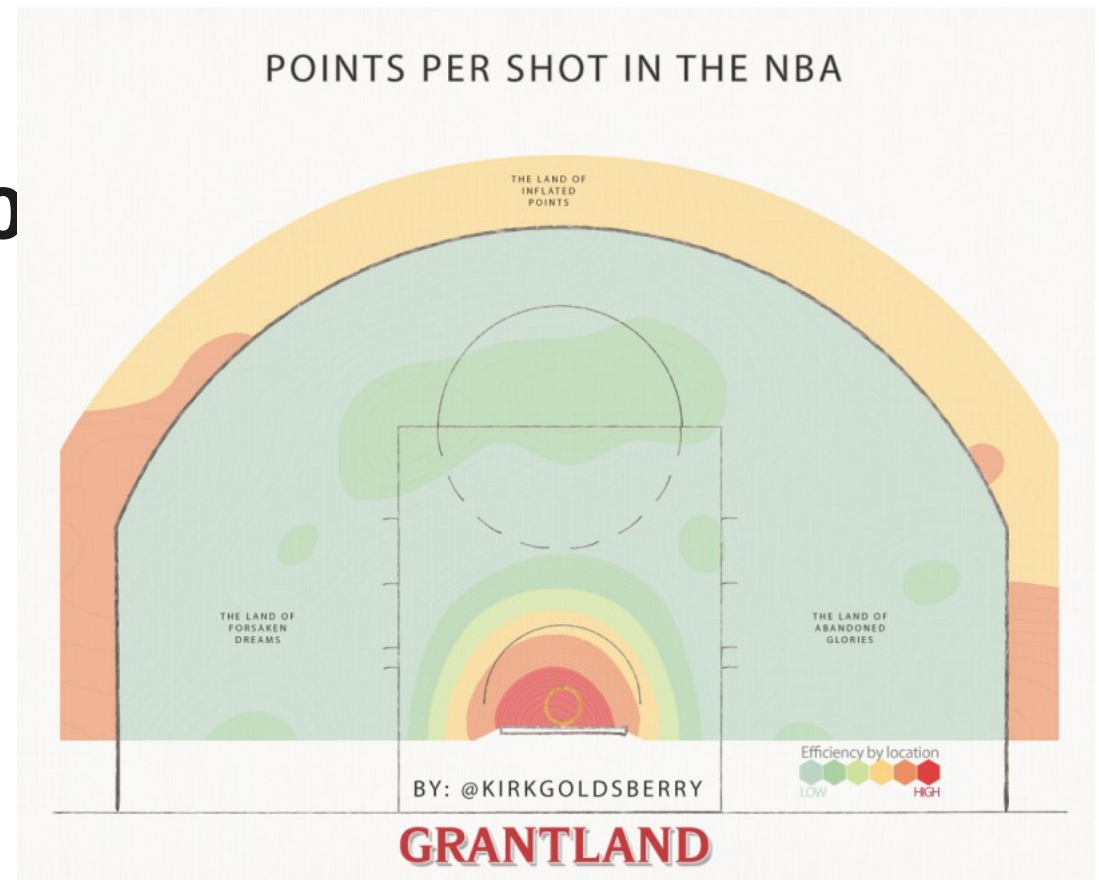
Aggregate 3P%: .360



Changing tack - value of shots in the modern NBA

3 pointer = $1.5 \times$ 2 pointer \rightarrow 2P% should be at least $1.5 \times$ 2P%

Aggregate 3P%: .360



Changing tack - value of shots in the modern NBA

3 pointer = $1.5 \times$ 2 pointer \rightarrow 2P% should be at least $1.5 \times$ 2P%

Aggregate 3P%: .360

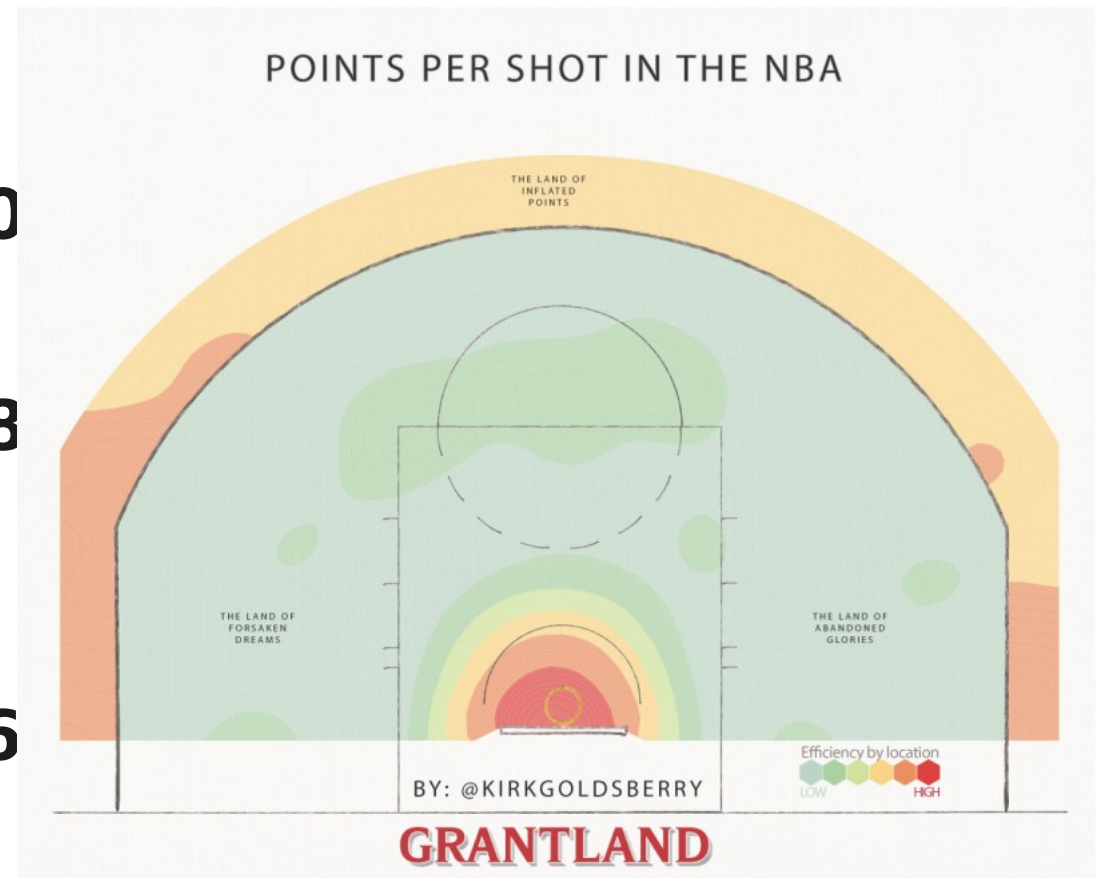
Corner: .390

Aggregate 2P%: .488

0-3 feet: .636

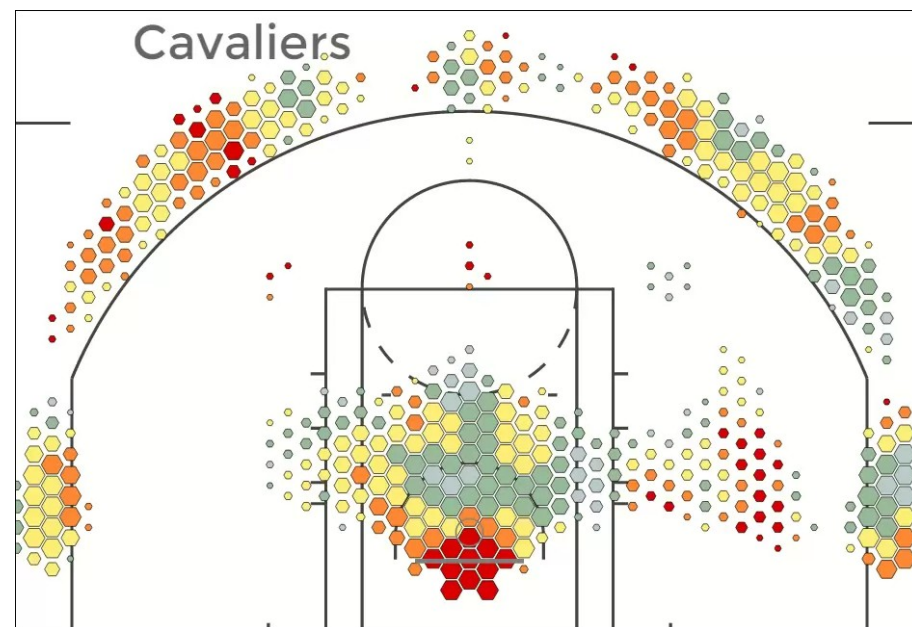
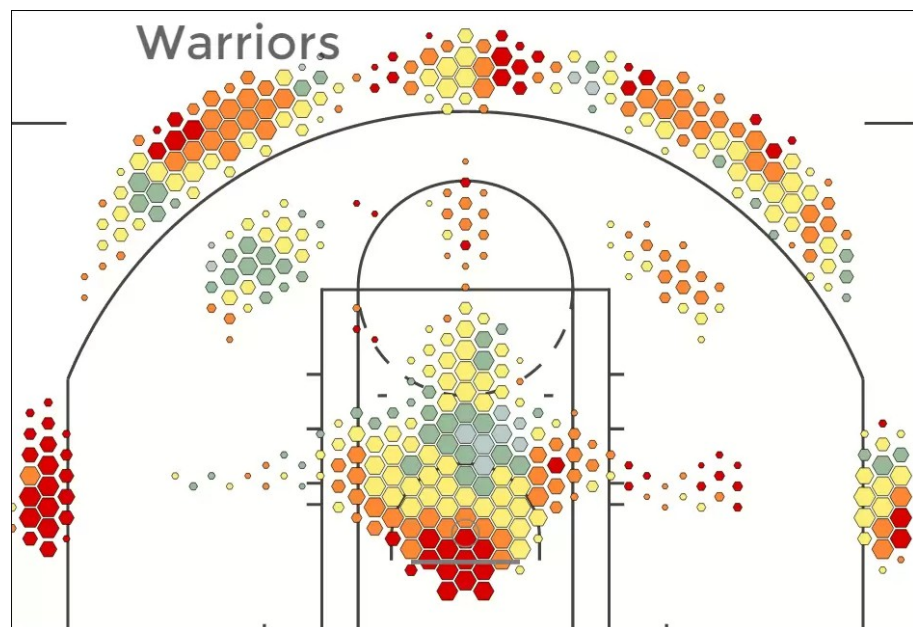
3+ feet: $\sim .4$

Aggregate FT%: .756

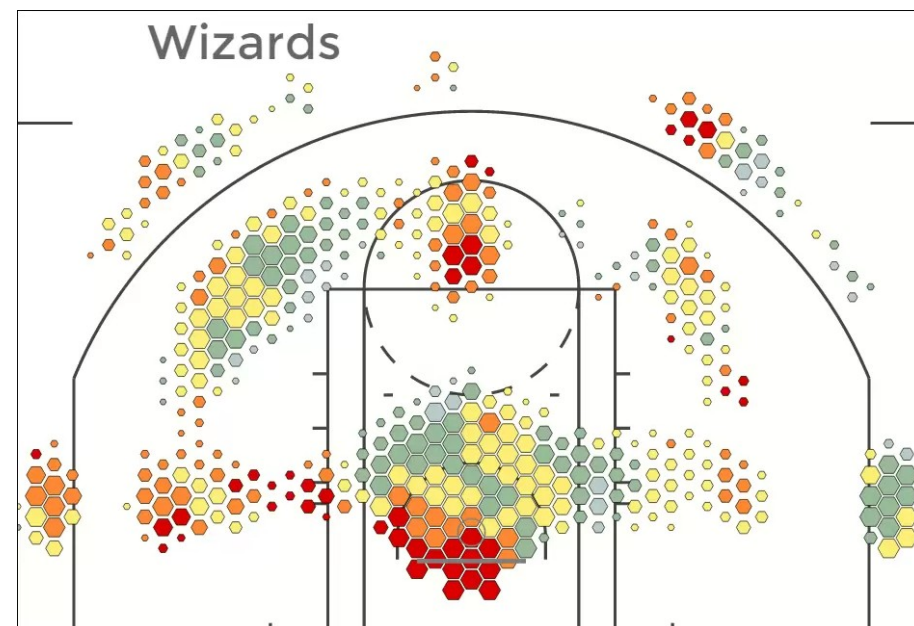
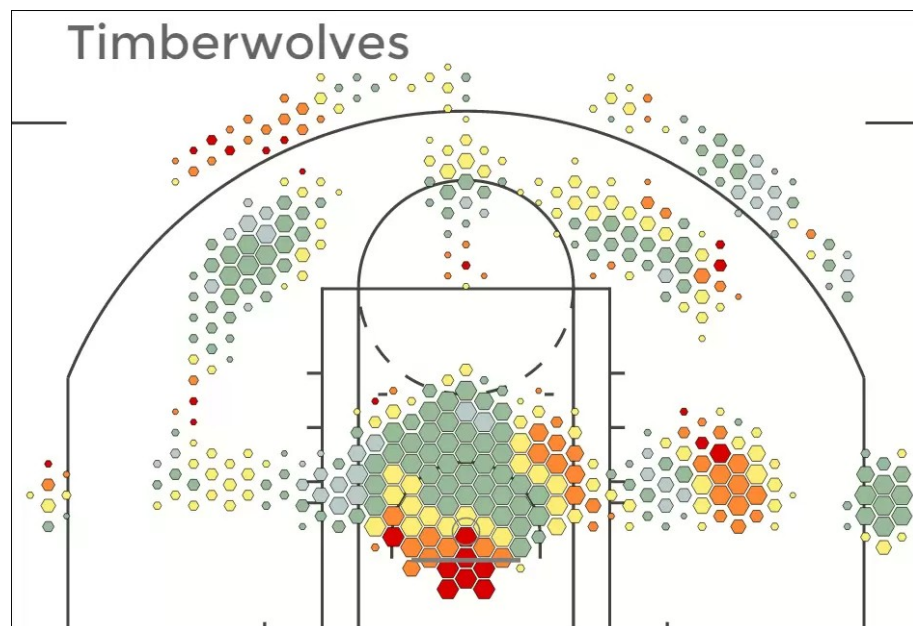


Winning teams adapt

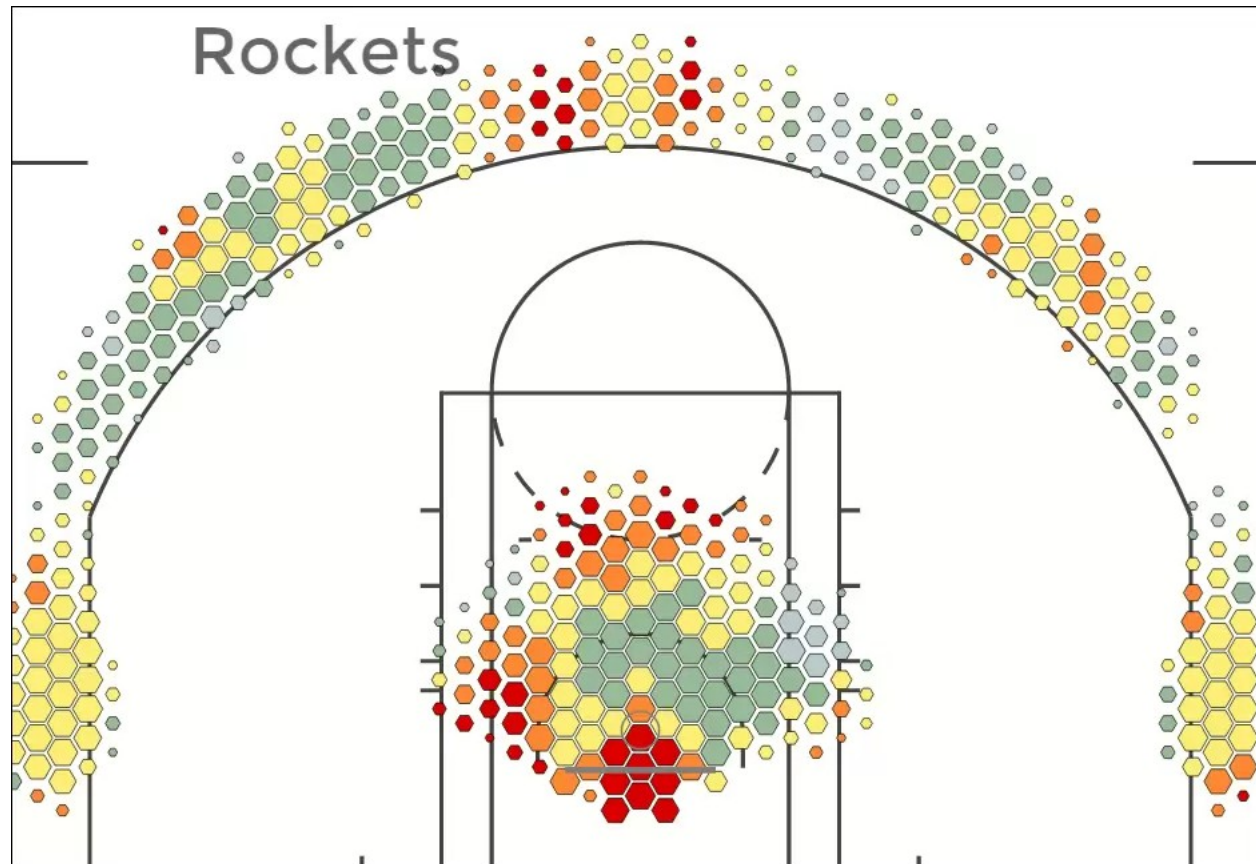
NBA Shot Charts
BabyS10thWithA1 @ imgur



Losing teams don't



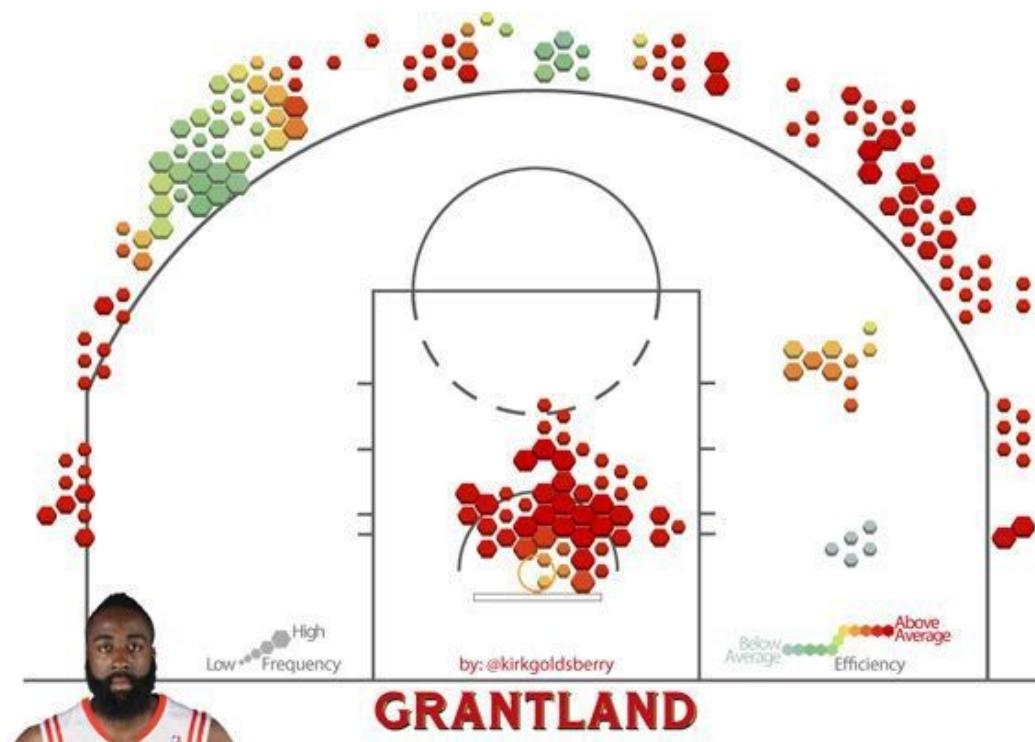
Stat geek purity



“The Future of Basketball Is Here, and It Looks a Lot Like James Harden”

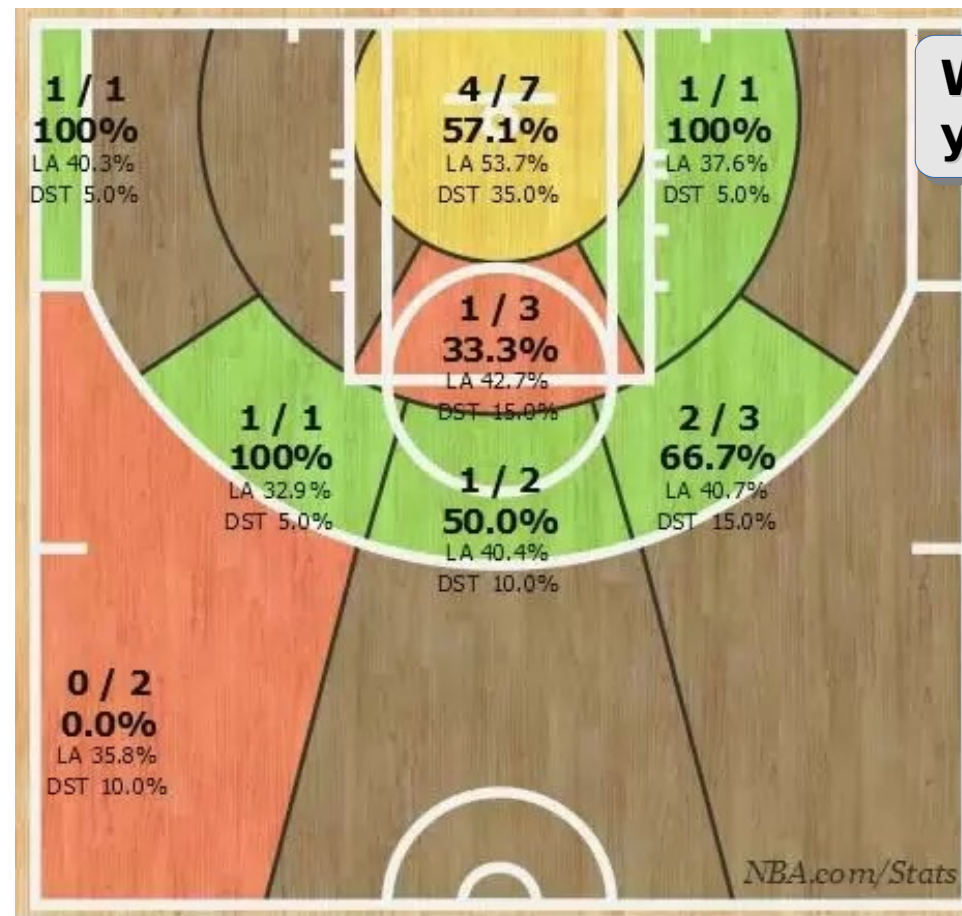
Grantland, 01/07/15

James Harden



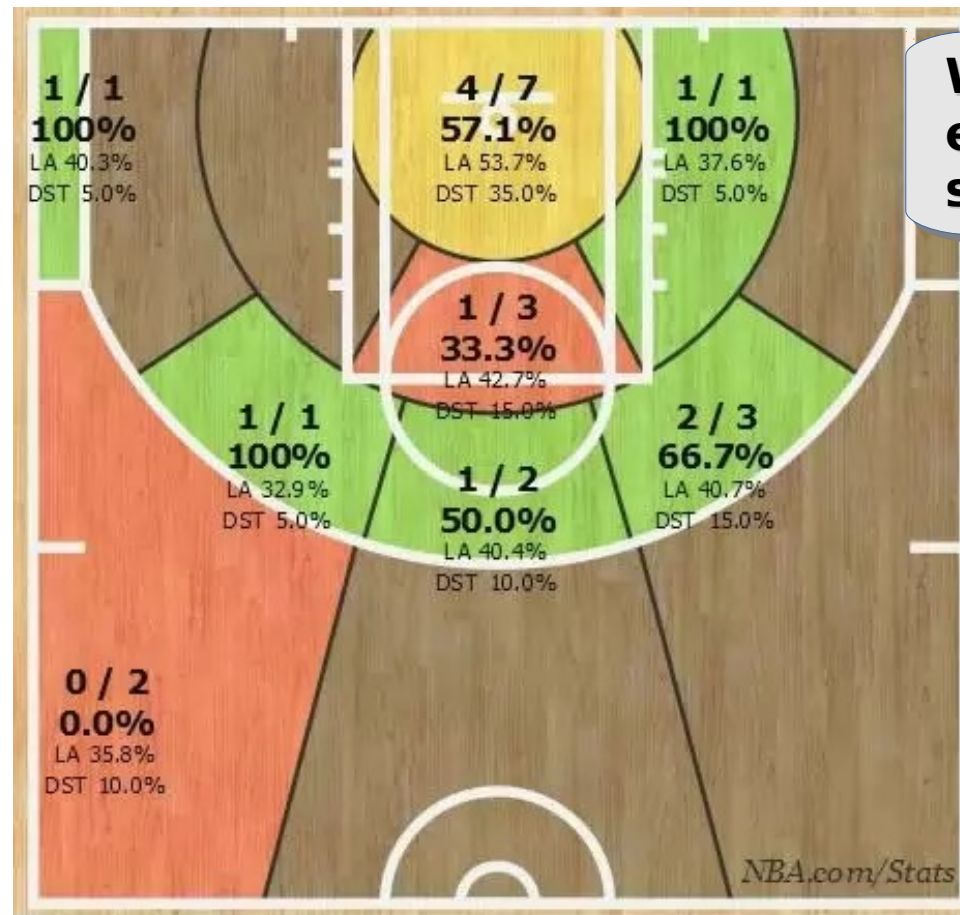
James Harden becomes the Washington Wizards

Greg Price
International Business Times,
05/20/15



James Harden becomes the Washington Wizards

Greg Price
International Business Times,
05/20/15



Why is the efficiency of 3P so high?

Hypotheses

Arbitrary line

Teams get **specialists** (“3 and D”)

3 pointers are **consciously** set up:

- .837 assisted vs
- .518 for 2 pointers

Maybe less well **defended** during regular season, witness **Kyle Korver**:

- .492 regular season
- .355 playoffs

**Can't replay
the match!**

Digging into details

Detailed play-by-play data:

- Temporal sequence of events
- Seconds of dribbling
- Nearest defender...

Exploring **counter-factuals**: Graphical Model for Basketball Match Simulation (Sloan Conference '15)

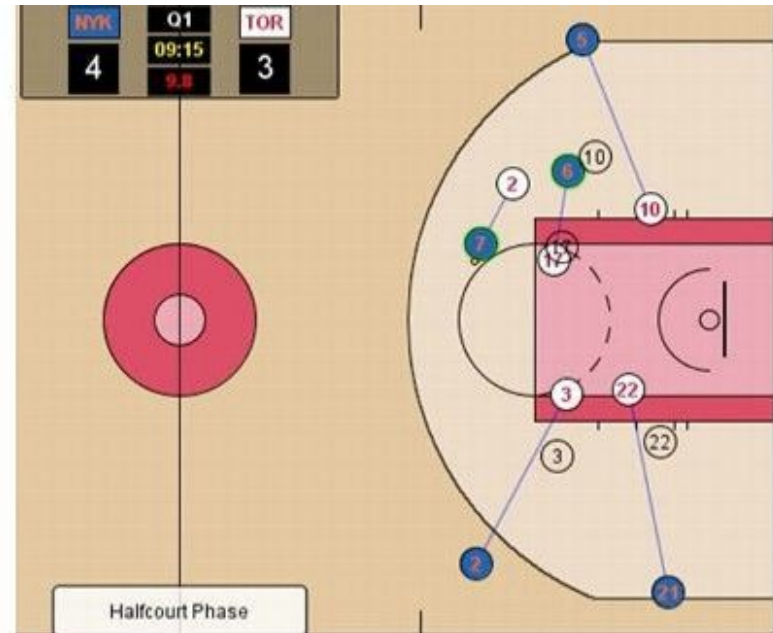
- Probabilistic graphical model
- Team-specific parameters for shots, passes, defensive actions...
- Change line-ups, simulate matches

Digging into details (2)

Lights, Camera, Revolution
Grantland
03/19/2013

SportVU NBA **tracking** data Exploring counter-factuals:

- Where was the defender/offensive player?
- Where should/could he have been?
- How does this change the outcome?



I really, really wanna generate data!

Define shoppers (behavior), layouts, re-run loads of time

- W/input from psychologists, marketers

Define people, have them make friends → Social networks

Constructive learning

- Much more primitive than Fabrizio's

...

No time, no funding, but a big need, I believe!