Big Data Primer

Introducing:

- Technology Association of Oregon's event
  Case Studies in Big Data

Allen Grimm
TheGrimmScientist@gmail.com
Data Science Consultant
June 6th, 2014

What is Data Science?

"Delivering on the promise of big data." - Predictive Analytics World

"Making data work" - Strata Conference

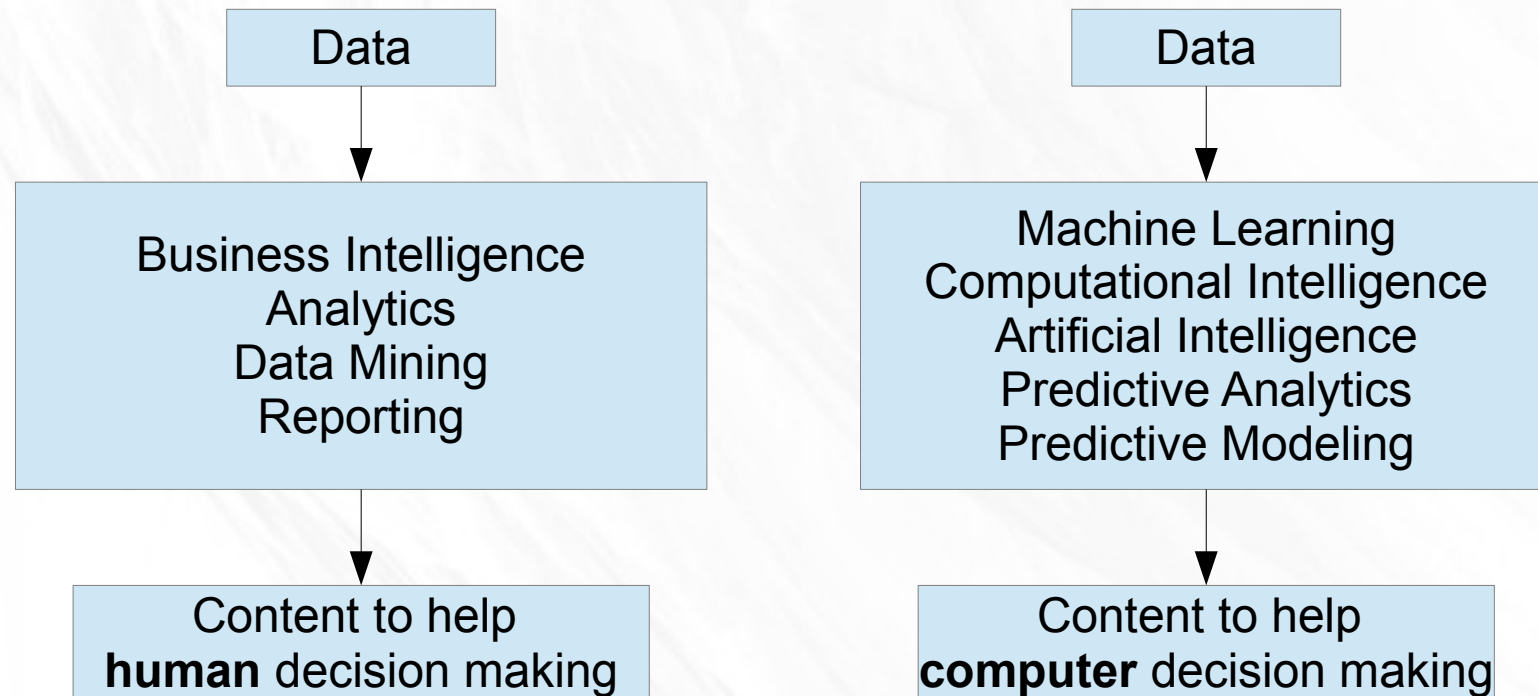"Almost like a Renaissance individual who really wants to learn and bring change to an organization." - IBM

"The study of the generalizable extraction of knowledge from data" - Wikipedia

"Have a deep passion for applying advanced analytic approaches, an eagerness to dig into the data, and a vision for turning disparate data streams into a cohesive view of our global consumers" - Nike job description

Data is Potential Action

Buzzwords, and Where Data Helps

Data

Data

Business Intelligence
Analytics
Data Mining
Reporting

Machine Learning
Computational Intelligence
Artificial Intelligence
Predictive Analytics
Predictive Modeling

Content to help
human decision making

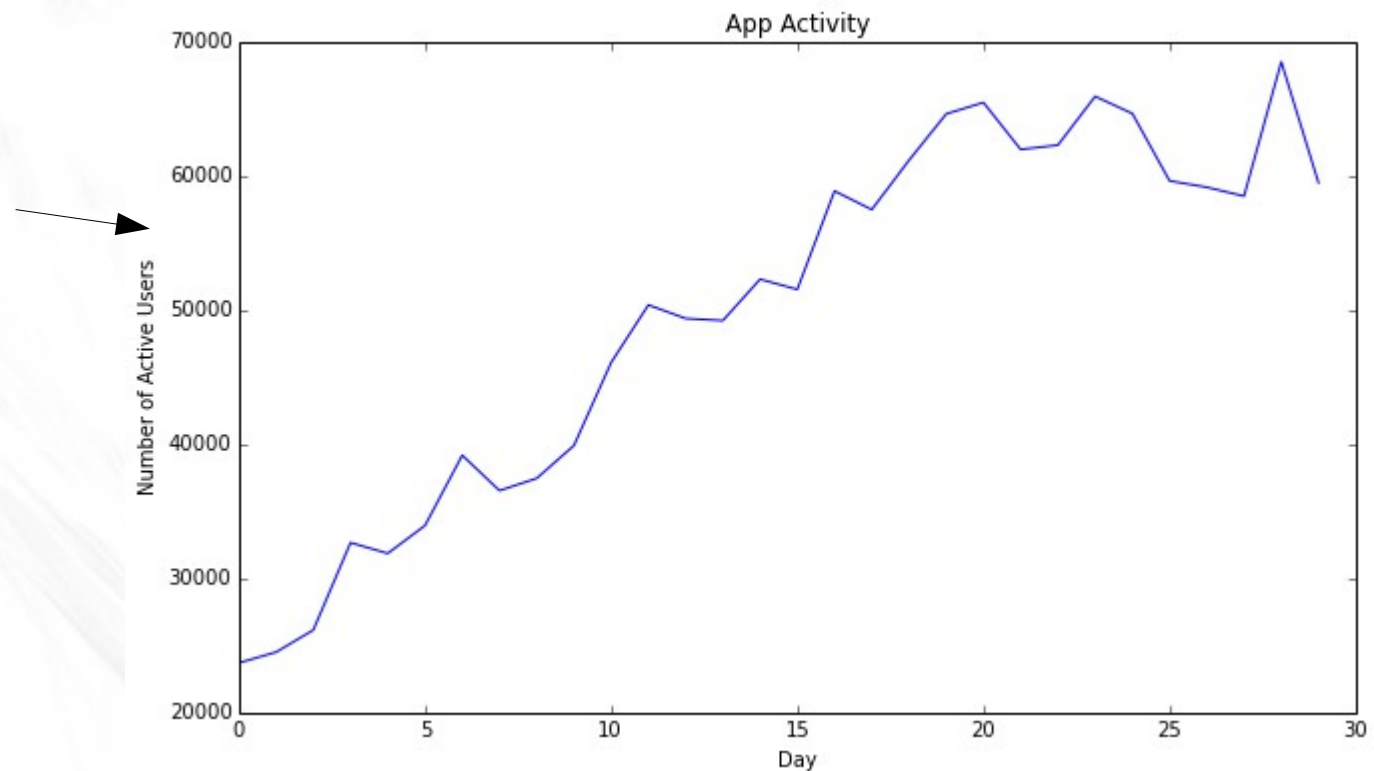Content to help
computer decision making

Data Science Roles

- Liaison between machine learners and business people

- Integrate and use disparate data sources

- Manage / manipulate big data

- Data savvy product manager
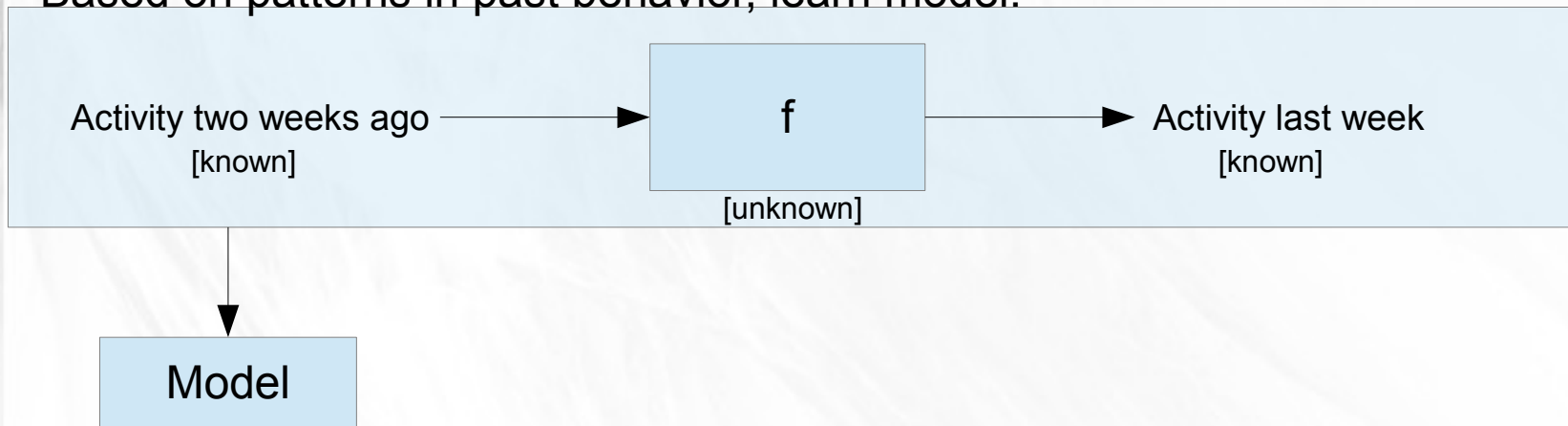
Example: Your App

Raw data:

user	date
1	5/1
1	5/1
1	5/1
1	5/4
1	5/5
1	5/6
2	5/3
2	5/3
3	5/2
3	5/3
…	…

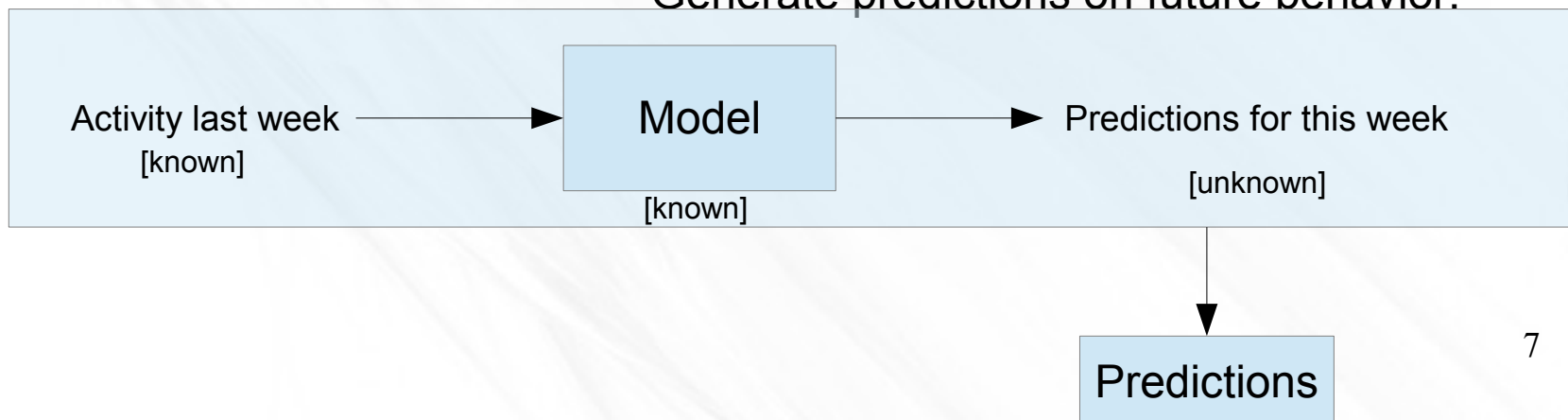App Activity

Churn Prediction

Based on patterns in past behavior, learn model:

Activity two weeks ago [known] → f [unknown] → Activity last week [known]

Model

Generate predictions on future behavior:

Activity last week [known] → Model [known] → Predictions for this week [unknown]

Predictions

AB Testing

Determine the best action to take for a specific group:
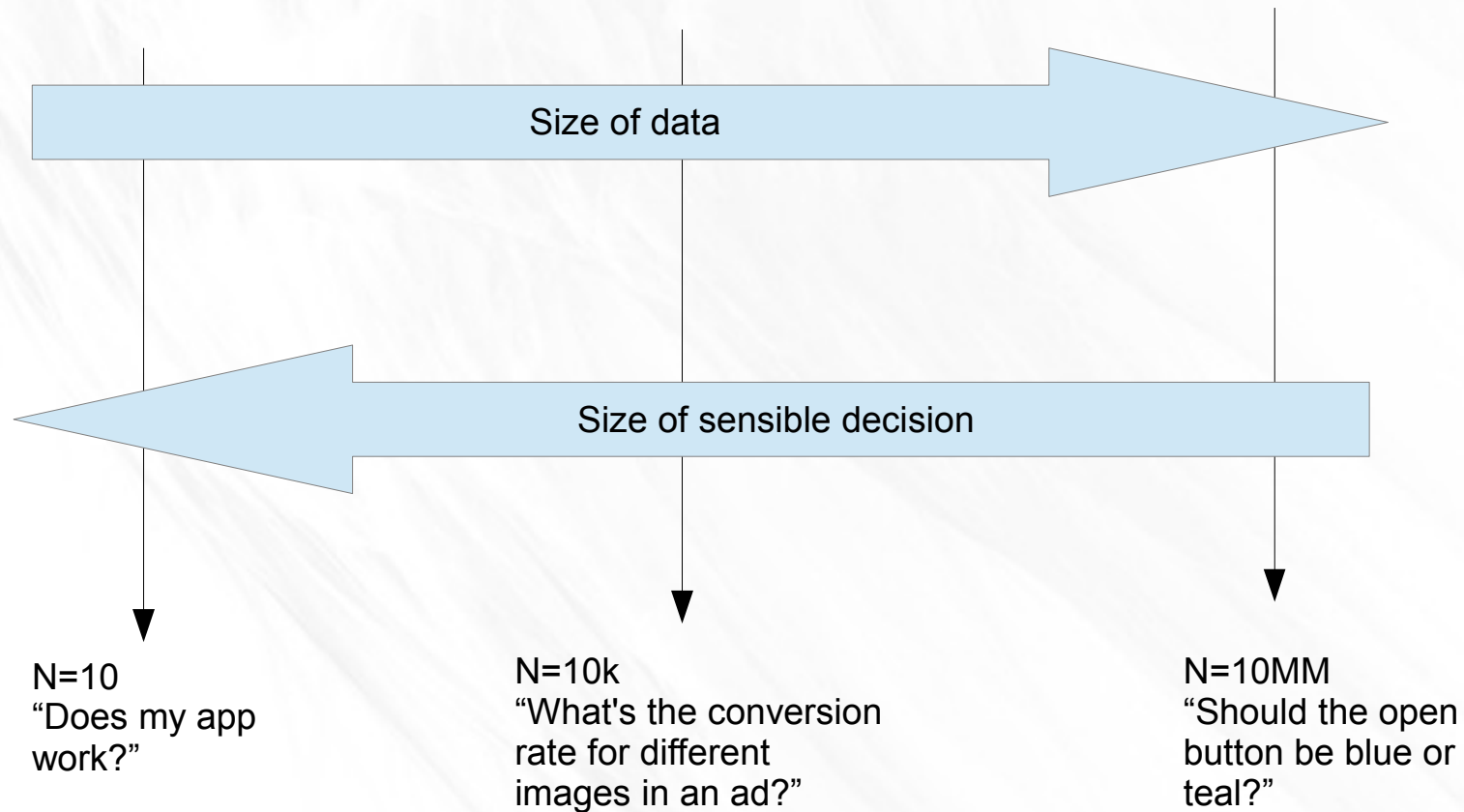
Looking only at users likely to leave:

Treatment:	Original Population:	Population Remaining
Control	4000	756
Email	4000	642
Gift	4000	1108

Other Examples (More Buzzwords)

- Behavior Prediction

- Ad Optimization

- Response Modeling

- Your credit score

- Cohort Analysis

- Clustering

- Time Series Forecasting

- Virality Modeling

- Natural Language Processing

- Fraud / Outlier Detection

- Failure Prediction

Size of Data Needed

Size of data

Size of sensible decision

N=10
"Does my app work?"

N=10k
"What's the conversion rate for different images in an ad?"

N=10MM
"Should the open button be blue or teal?"

For More Info...

Come to the Technology Associate of Oregon's event

- – Case Studies in Big Data
- – see www.techoregon.org for the details

Allen Grimm
TheGrimmScientist@gmail.com
@GrimmScientist
TheGrimmScientist.com

11