# IT'S NOT ABOUT BIG DATA:

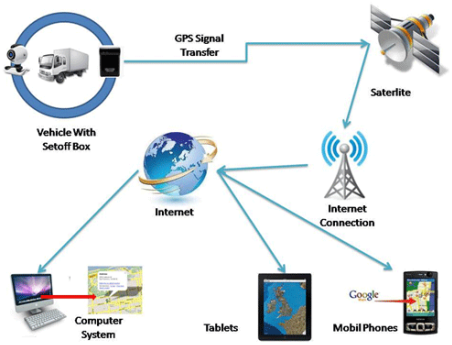
# IT'S ABOUT BIG DATA "COMPRESSION"!

AMA - ECMI - EMAC
Marketing & Innovation
Symposium


**Eric T. Bradlow**
**K.P. Chao Professor; Professor of Marketing, Statistics and Education**
**Co-Director, Wharton Customer Analytics Initiative**
**Vice-Dean and Director, Wharton Doctoral Programs**

# The "Faucet of Big Data" Is Turned On: And, We Are Never Going Back!

GPS Location
("action and space")



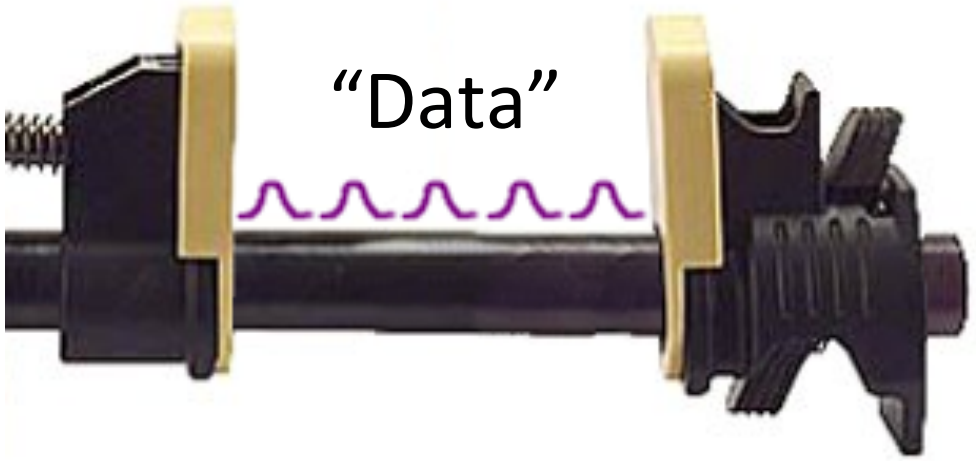

Social Network
("mine and others")

Eye Tracking
("fast and frequent measurement")

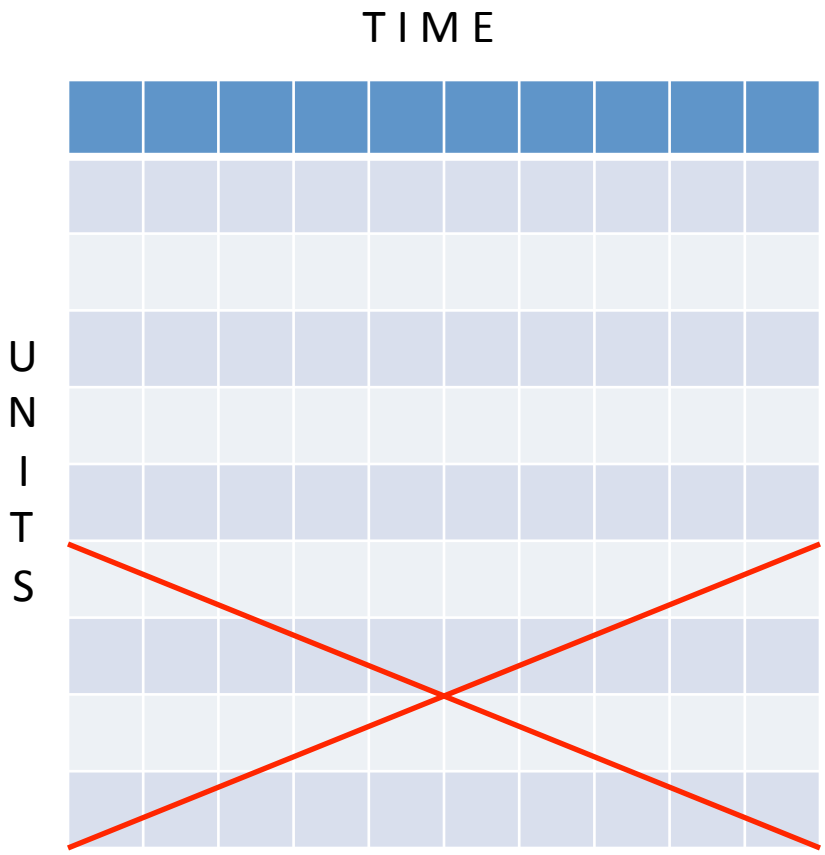
Data Fusion
("Many Variables")

nielsen

# The More We Learn, The More We Forget
# "Smart Data Compression"

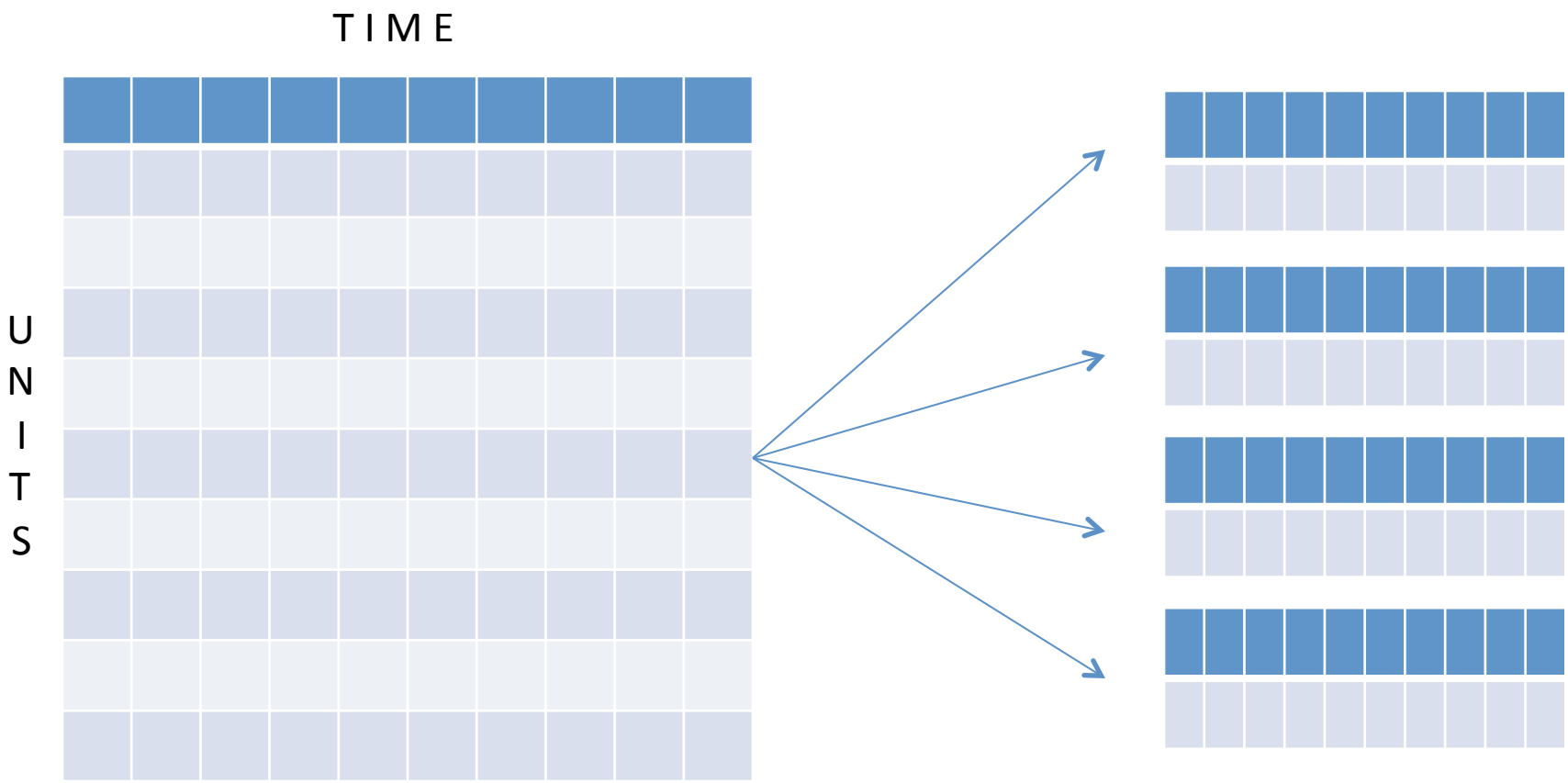


# Sampling: The $\sqrt{N}$ is Your Friend




Big Data Just Became Smaller Data

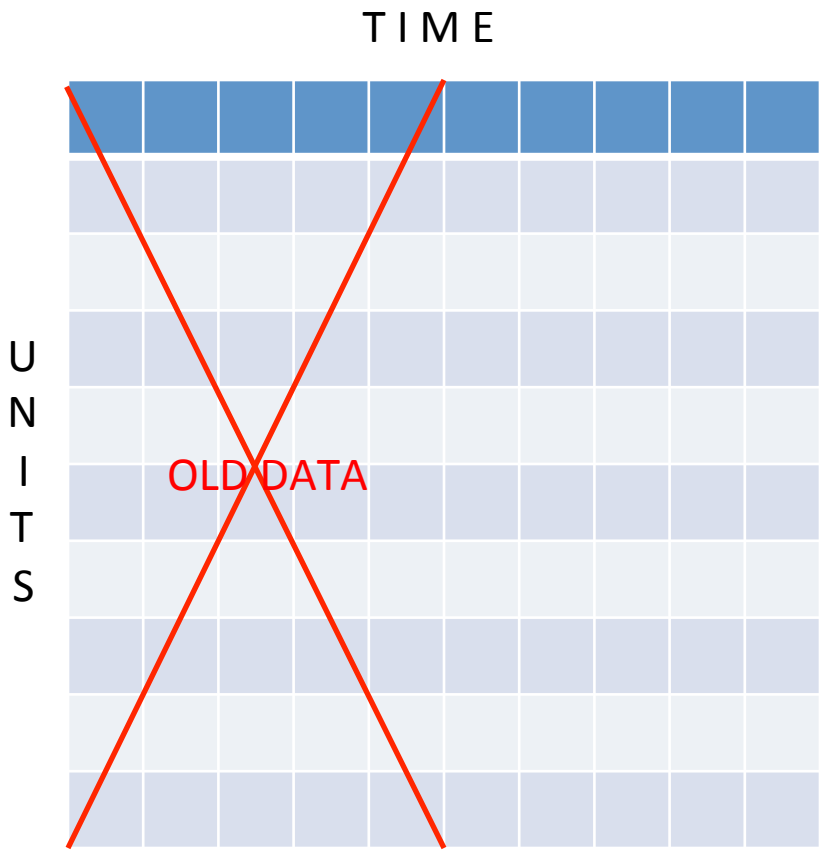
# Meta-Analysis: 1 + 1 < 2 is Your Friend




* Wang, P., Bradlow, E.T., and George, E.I. (2014), "Meta-Analysis Using Information Reweighting: An Application to Online Advertising"

Big Data Just Became Smaller Data

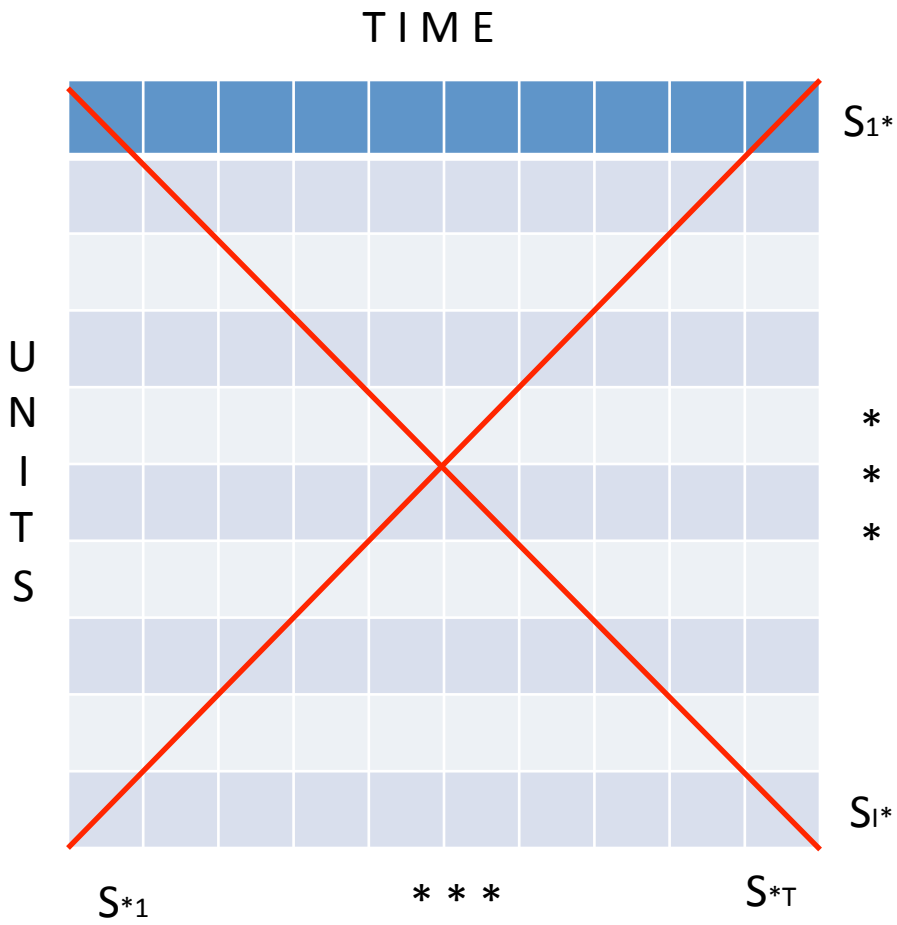
# Getting Rid of Old Data: Time is Your Friend




Big Data Just Became Smaller Data

* Gopalakrishnan, A., Bradlow, E.T., and Fader, P. (2014), "A Cross-Cohort Changepoint Model for Customer-Base Analysis"

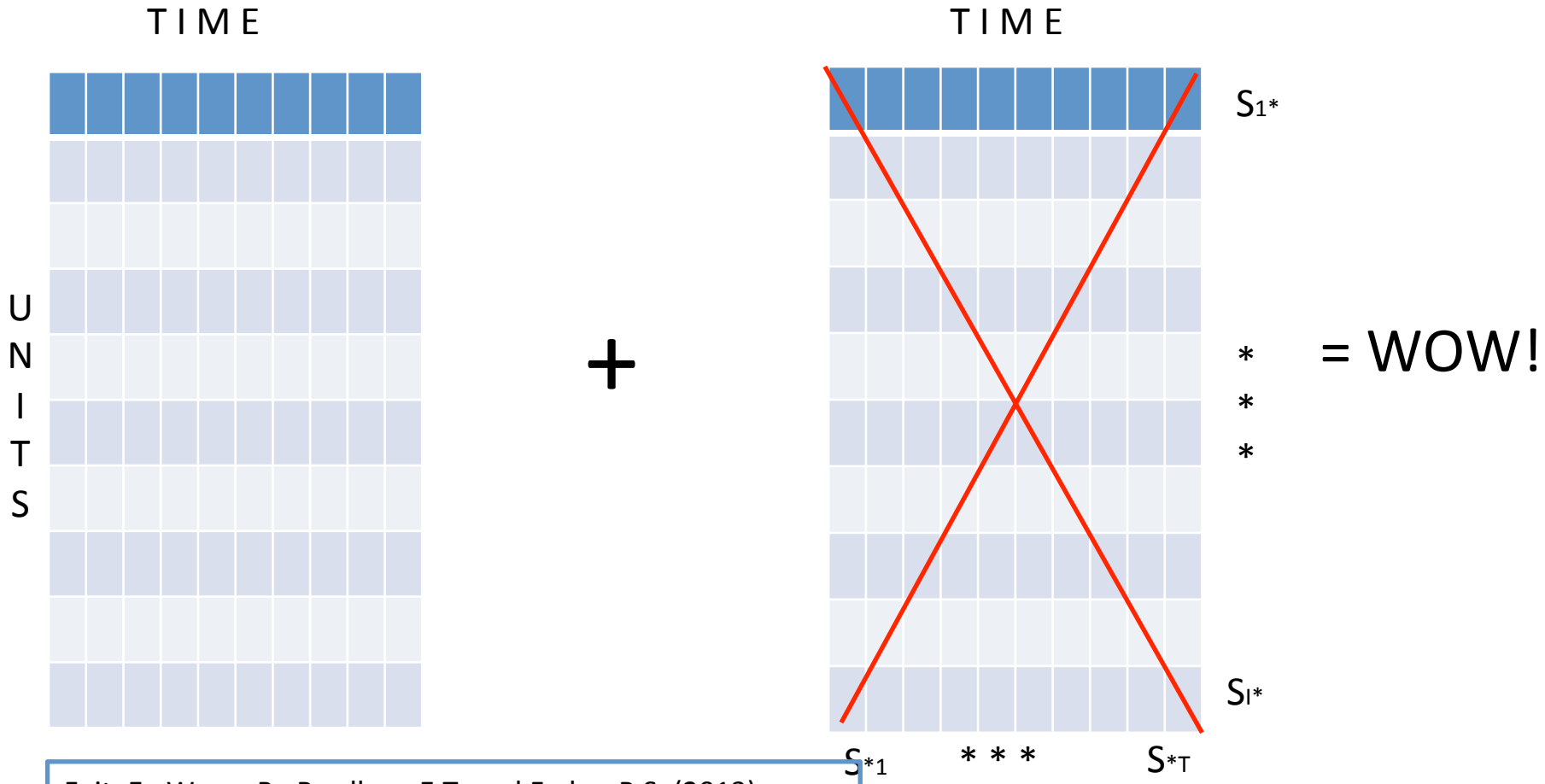
# Data Aggregation:
# The Information in the Margins is Your Friend




Musalem, A., Bradlow, E.T., and Raju, J. (2008), "Who's got the coupon?
Estimating Consumer Preferences and Coupon Usage from Aggregate Information"

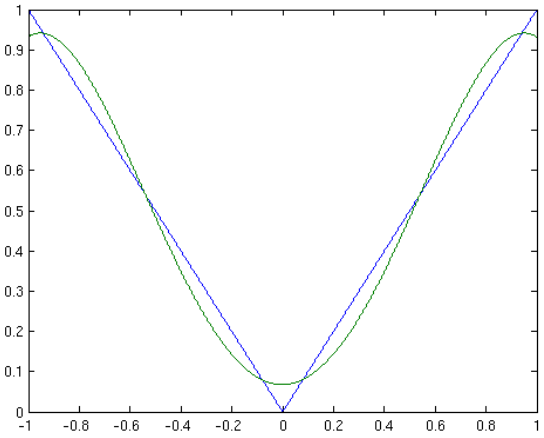
Big Data Just Became Smaller Data
And You Did It On Purpose!

# Data Fusion:
# Grabbing Information "When you can" is Your Friend




Feit, E., Wang, P., Bradlow, E.T. and Fader, P.S. (2013), "Fusing Aggregate & Disaggregate Data with an Application to Multi-Platform Media Consumption"

# Function Approximation: Math Is Your Friend




---- = Truth
Computation = 36 Hours

---- = Approximation
Computation = 2 seconds

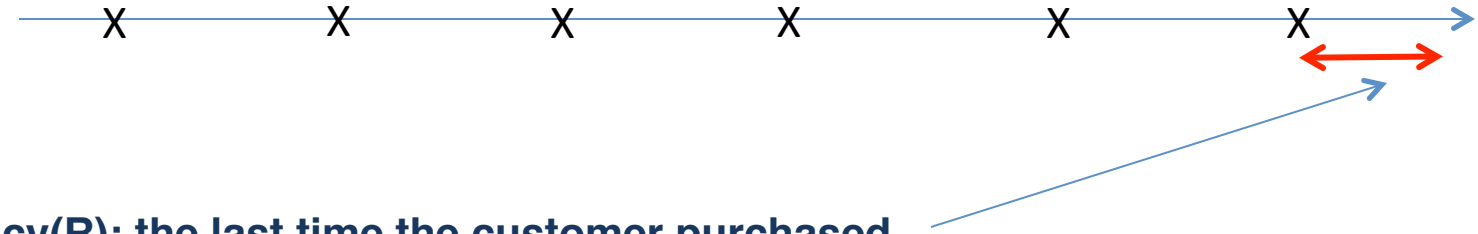
* Bradlow, E.T., Hardie, B.G.S., and Fader, P.S. (2002), "Bayesian Inference for the Negative Binomial Distribution Via Polynomial Expansions"

* Everson, P.J. and Bradlow, E.T. (2002), "Bayesian Inference for the Beta-Binomial Distribution via Polynomial Expansions"

* Miller, S.J., Bradlow, E.T., and Dayartna, K. (2006) "Closed-Form Bayesian Inferences for the Logit Model via Polynomial Expansions"

Time Just Became Smaller

# The Next Frontier: Statistical Sufficiency/Data Compression is Your BEST FRIEND!




**Recency(R): the last time the customer purchased.**

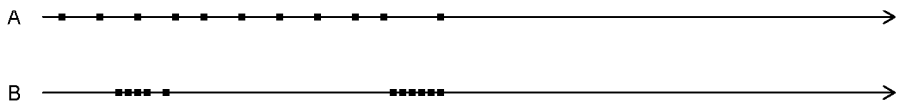
**Frequency(F): number of purchase occasions.**

**Monetary Value(M): average dollar value spent.**

Data = [date stamp, amount] for each transaction

Three Numbers (R,F, and M)

# Be Careful, You May Be Throwing Away Valuable Information



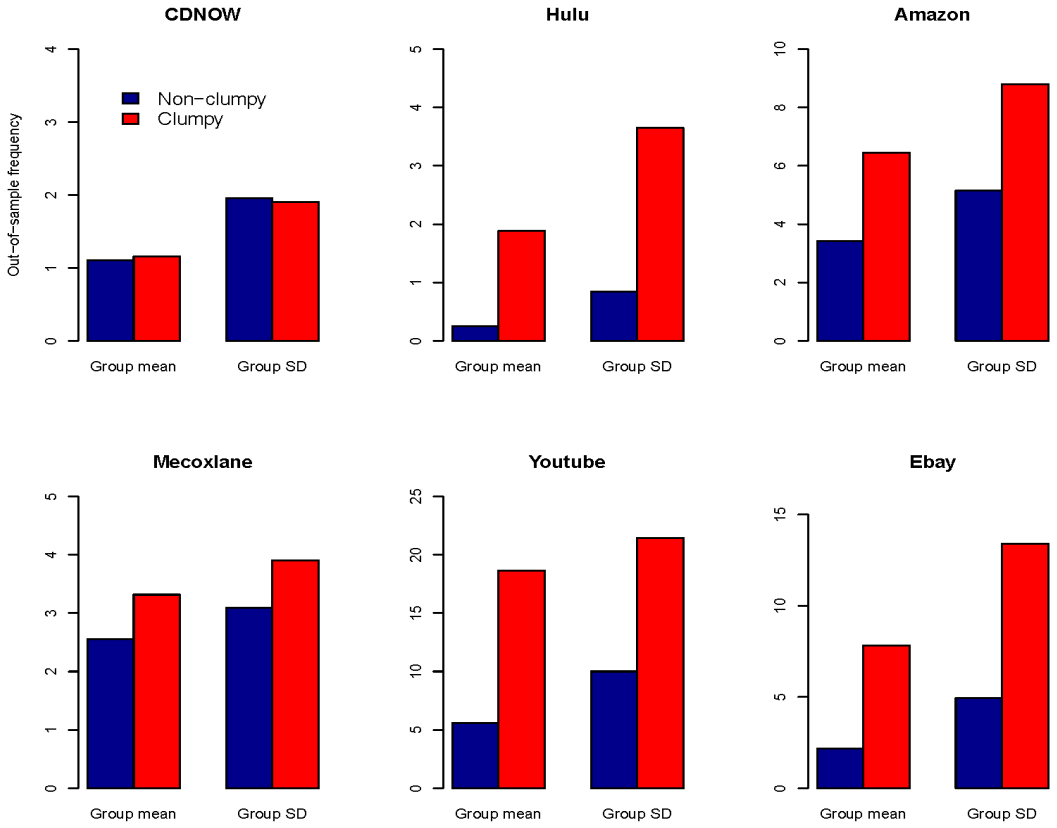

Do these Customers Look the Same to You???

C

* Zhang, Y., Bradlow, E.T., and Small, D.S. (2013), "New Measures of Clumpiness for Incidence Data",

* Zhang, Y., Bradlow, E.T., and Small, D.S. (2014), "Capturing Clumpiness when Valuing Customers: From RFM to RFMC

# Clumpiness Matters: Be Careful!



# What Does The Future of Data Compression Look Like?

**Answer: I don't know, but here is what my future looks like**

* Fast Computational Methods
* Data Sufficiency
* Aggregate Information Analyses (Link to Privacy)

**Thanks!**

**ebradlow@wharton.upenn.edu**