

Batch effects and confounders

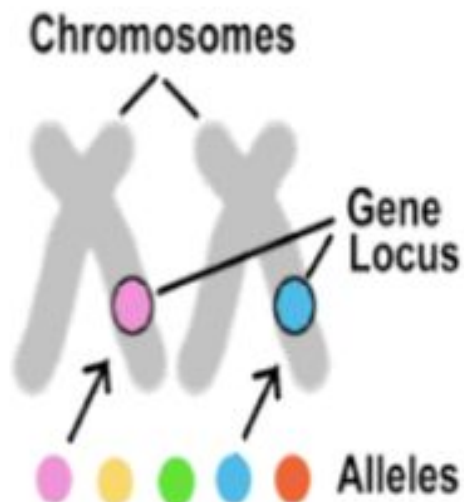
Jeff Leek

@jtleek

www.jtleek.com



**External Factors
(like environment)**

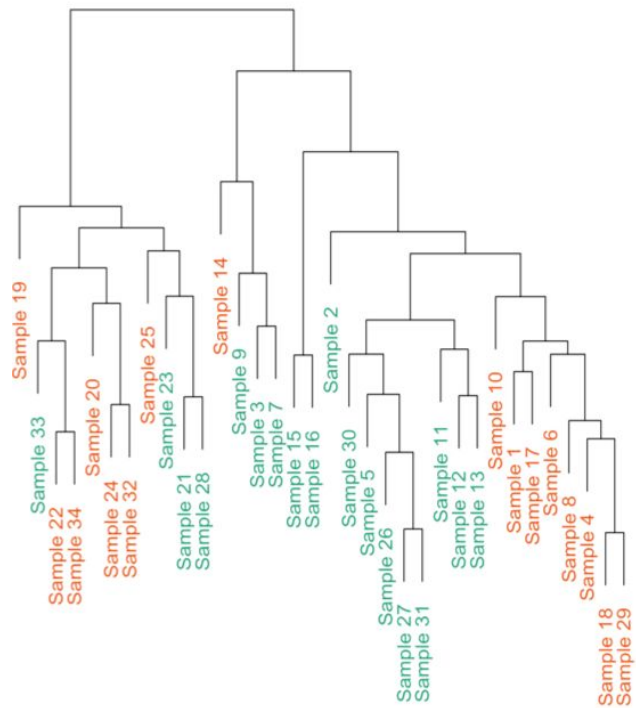


Genetics/Epigenetics

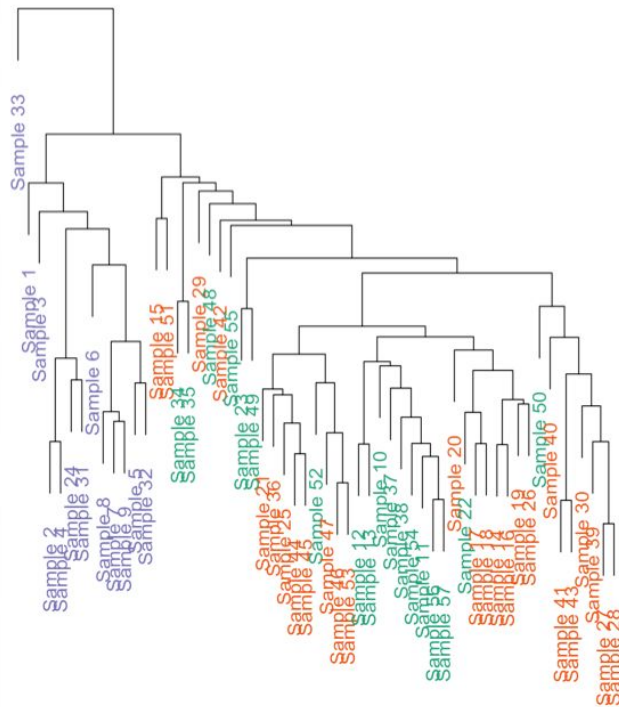


Technical Factors

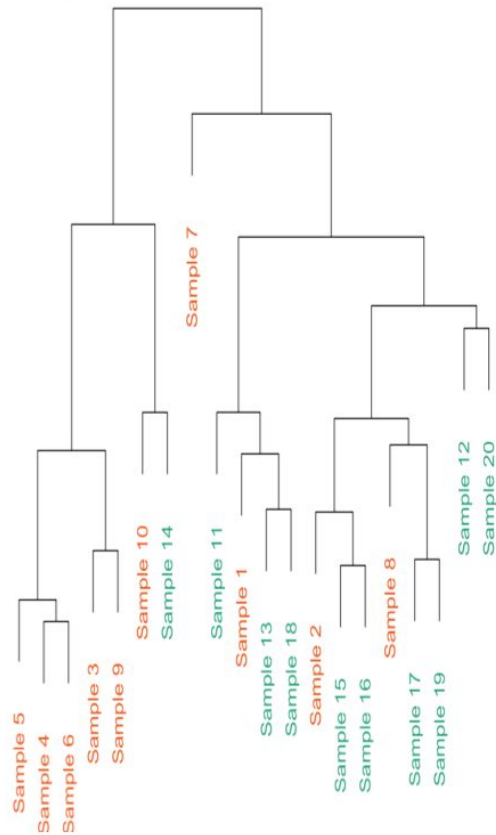
Color = Environment
(Idaghdour et al. 2008)

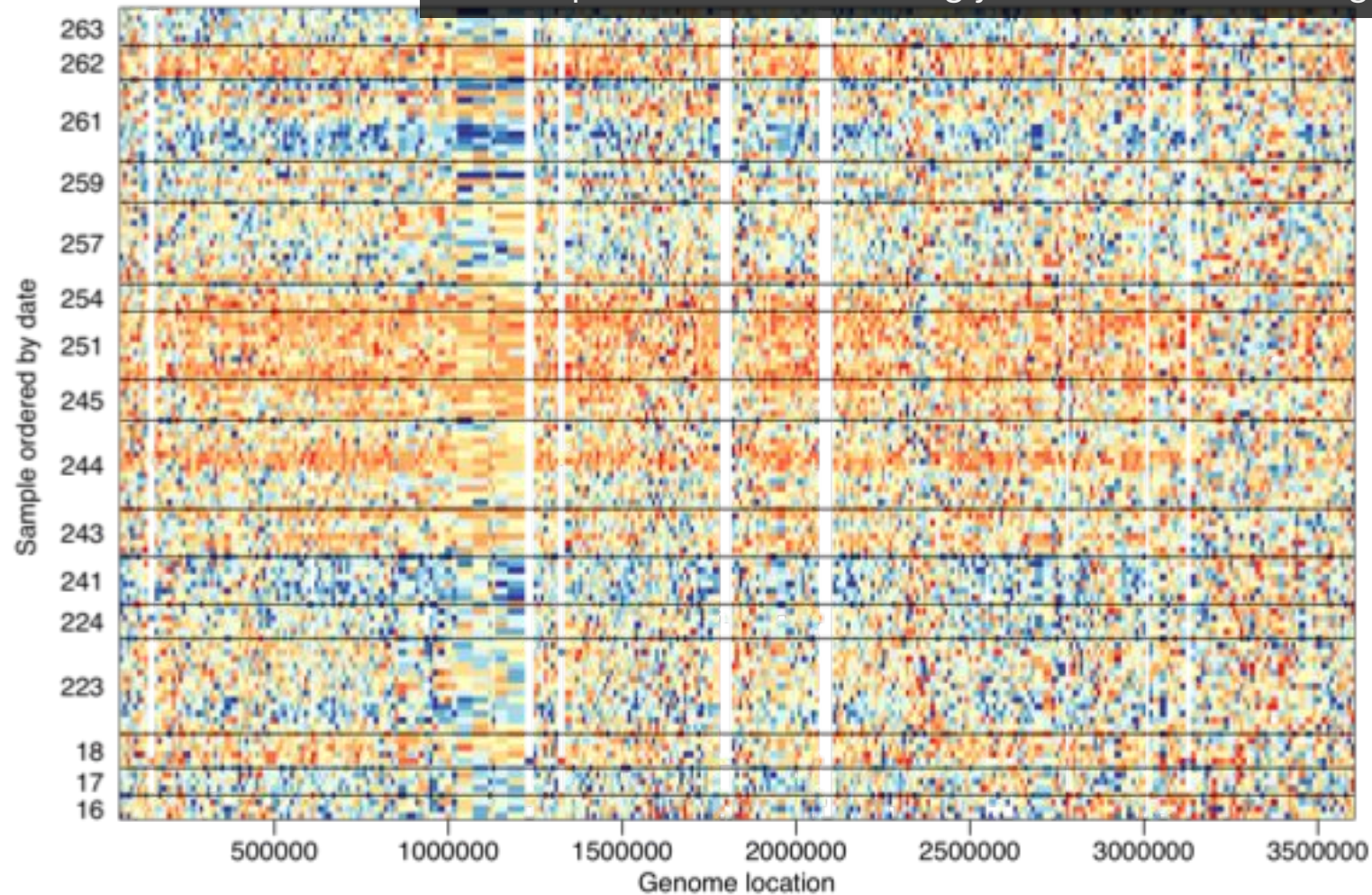


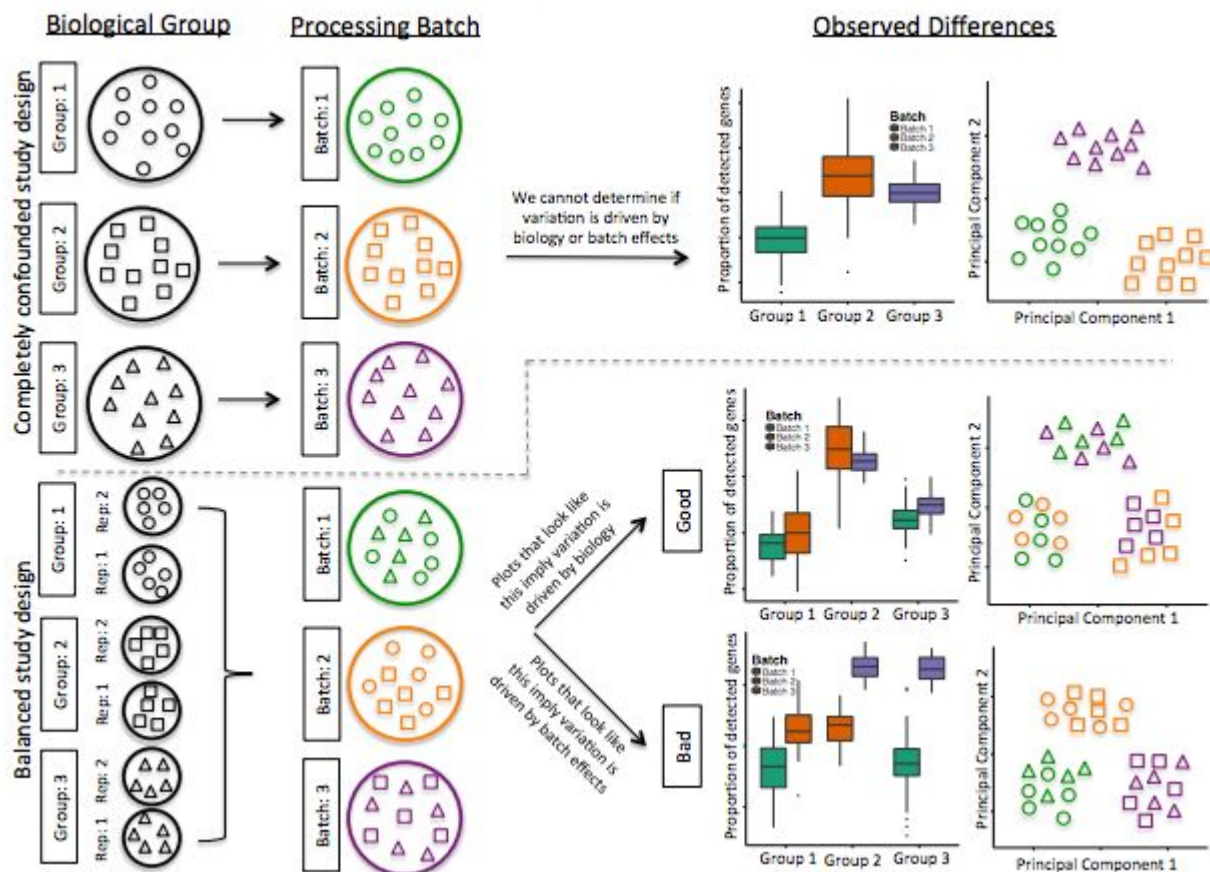
Color = Processing Year
(Cheung et al. 2008)



Color = Allele
(Brem et al. 2005)



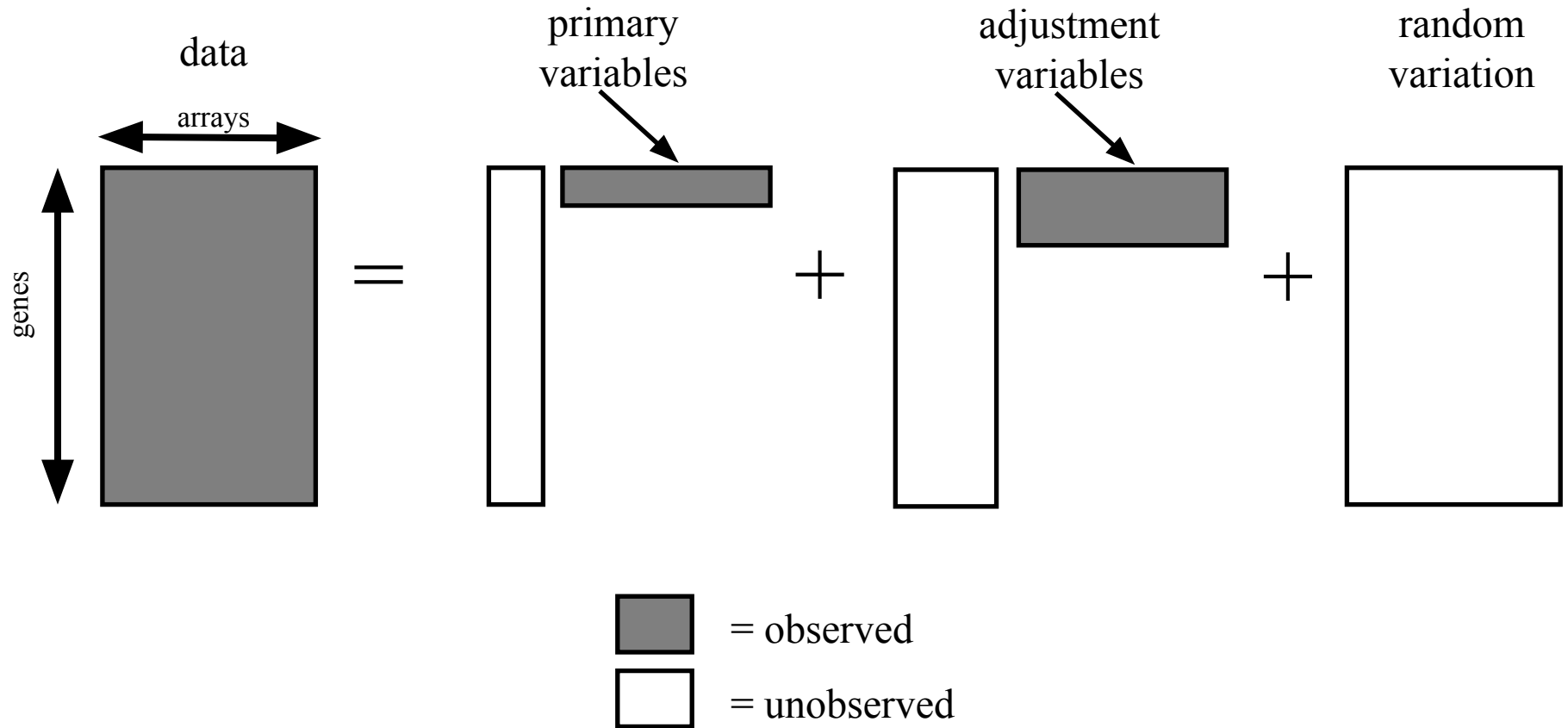




$$Y = b_0 + b_1 P + b_2 B + e$$

P = Phenotype you care about

B = Batch



Adjusting batch effects in microarray expression data using empirical Bayes methods

W. Evan Johnson and Cheng Li*

*To whom correspondence should be addressed.

Ariel Rabinovic

+ Author Affiliations

Received February 28, 2006.
Revision received April 14, 2006.
Accepted April 14, 2006.

Abstract

Non-biological experimental variation or "batch effects" are commonly observed across multiple batches of microarray experiments, often rendering the task of combining data from these batches difficult. The ability to combine microarray data sets is advantageous to researchers to increase statistical power to detect biological phenomena from studies where logistical considerations restrict sample size or in studies that require the sequential hybridization of arrays. In general, it is inappropriate to combine data sets without adjusting for batch effects. Methods have been proposed to filter batch effects from data, but these are often complicated and require large batch sizes (> 25) to implement. Because the majority of microarray studies are conducted using much smaller sample sizes, existing methods are not sufficient. We propose parametric and non-parametric empirical Bayes frameworks for adjusting

« Previous | Next Article »

Table of Contents

This Article

Biostat (2007) 8 (1): 118-127.
doi: 10.1093/biostatistics/kxj037
First published online: April 21, 2006

» Abstract **Free**
Full Text (HTML) **Free**
Full Text (PDF) **Free**
Data Supplement

All Versions of this Article:
kxj037v1
8/1/118 *most recent*

- Classifications

Article

- Services

Alert me when cited
Alert me if corrected
Find similar articles
Similar articles in PubMed
Add to my archive
Download citation
Request Permissions
Disclaimer

+ Citing Articles

+ Google Scholar

Advanced »

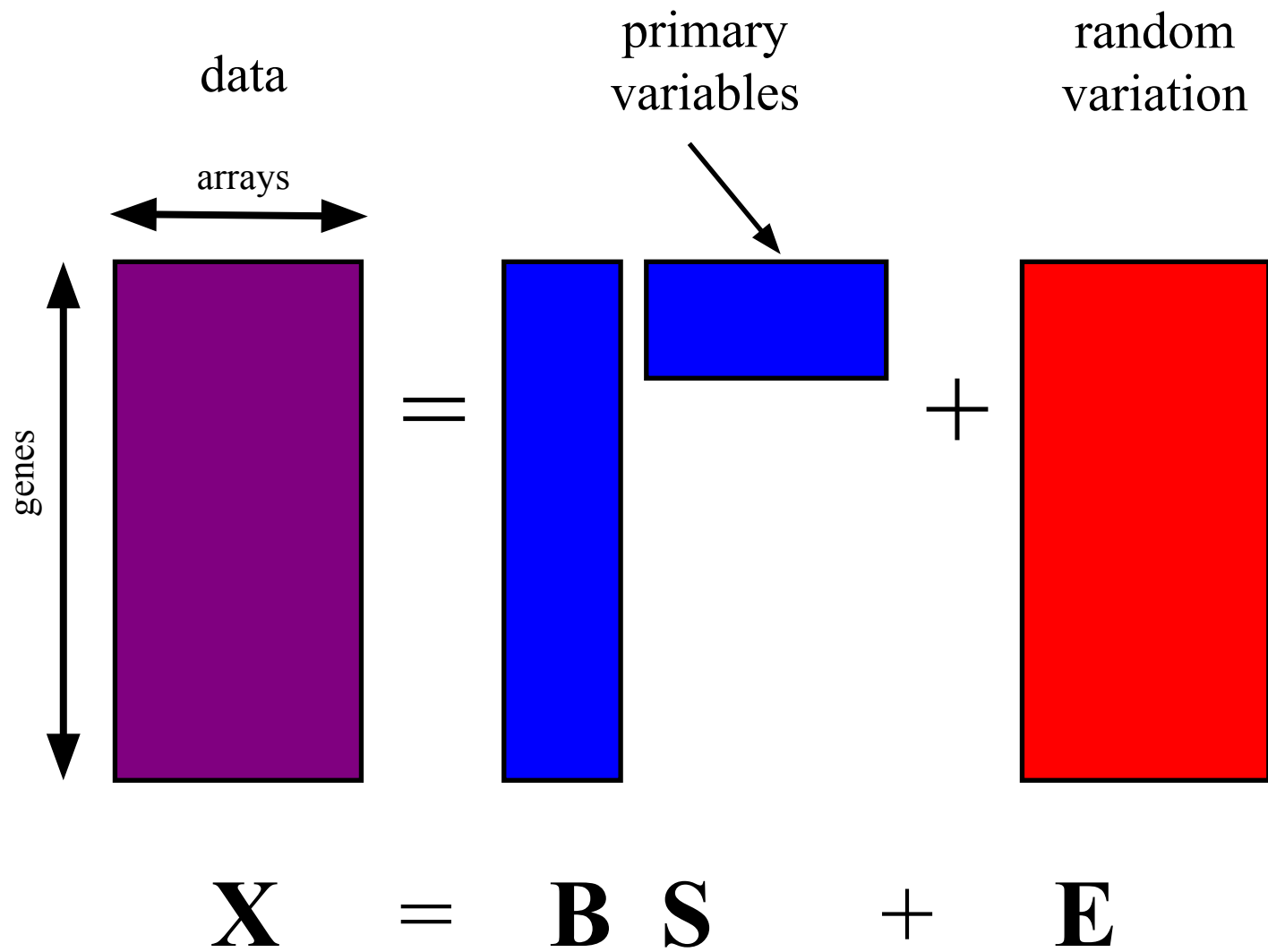
Current Issue

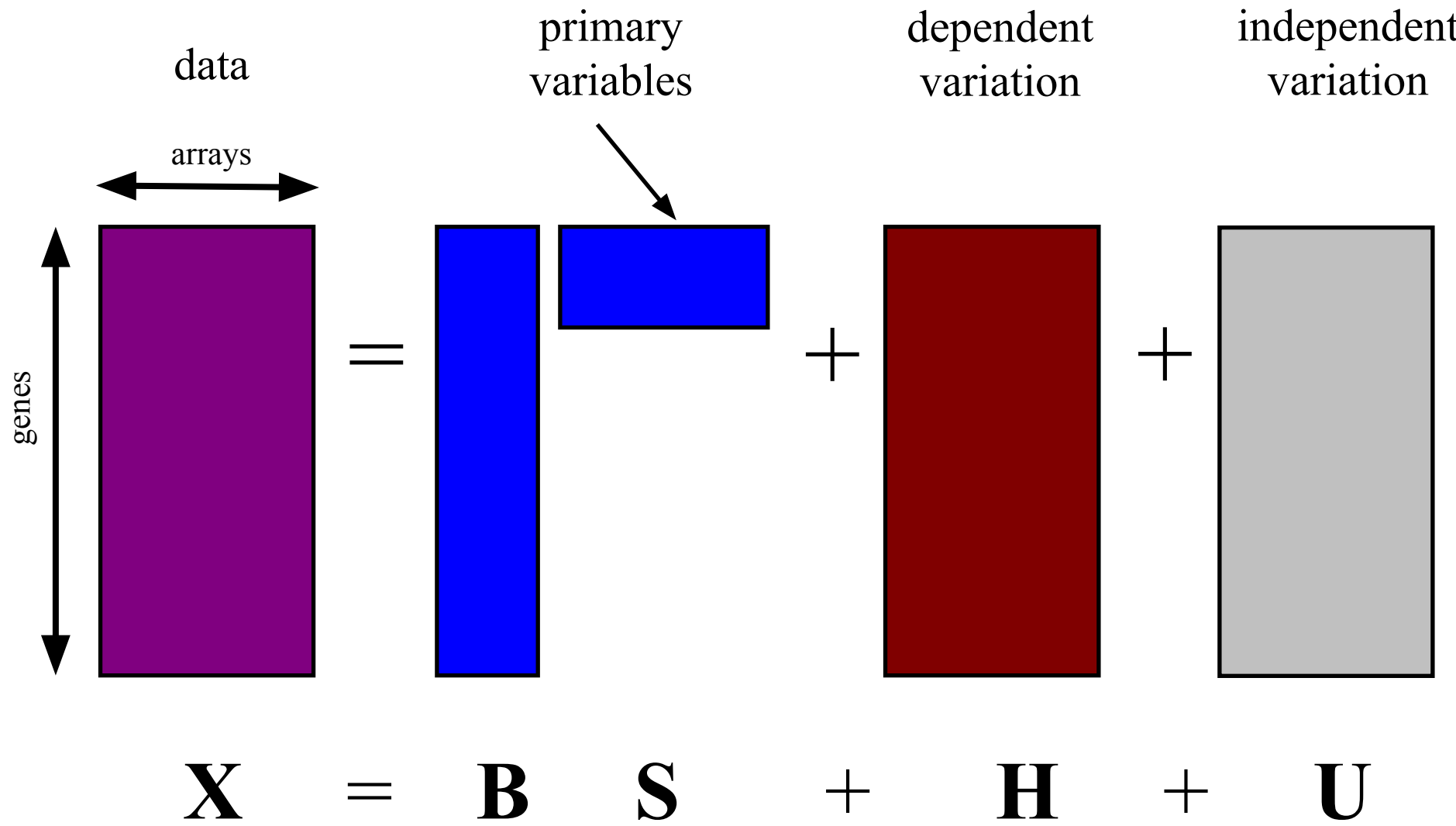
July 2015 16 (3)

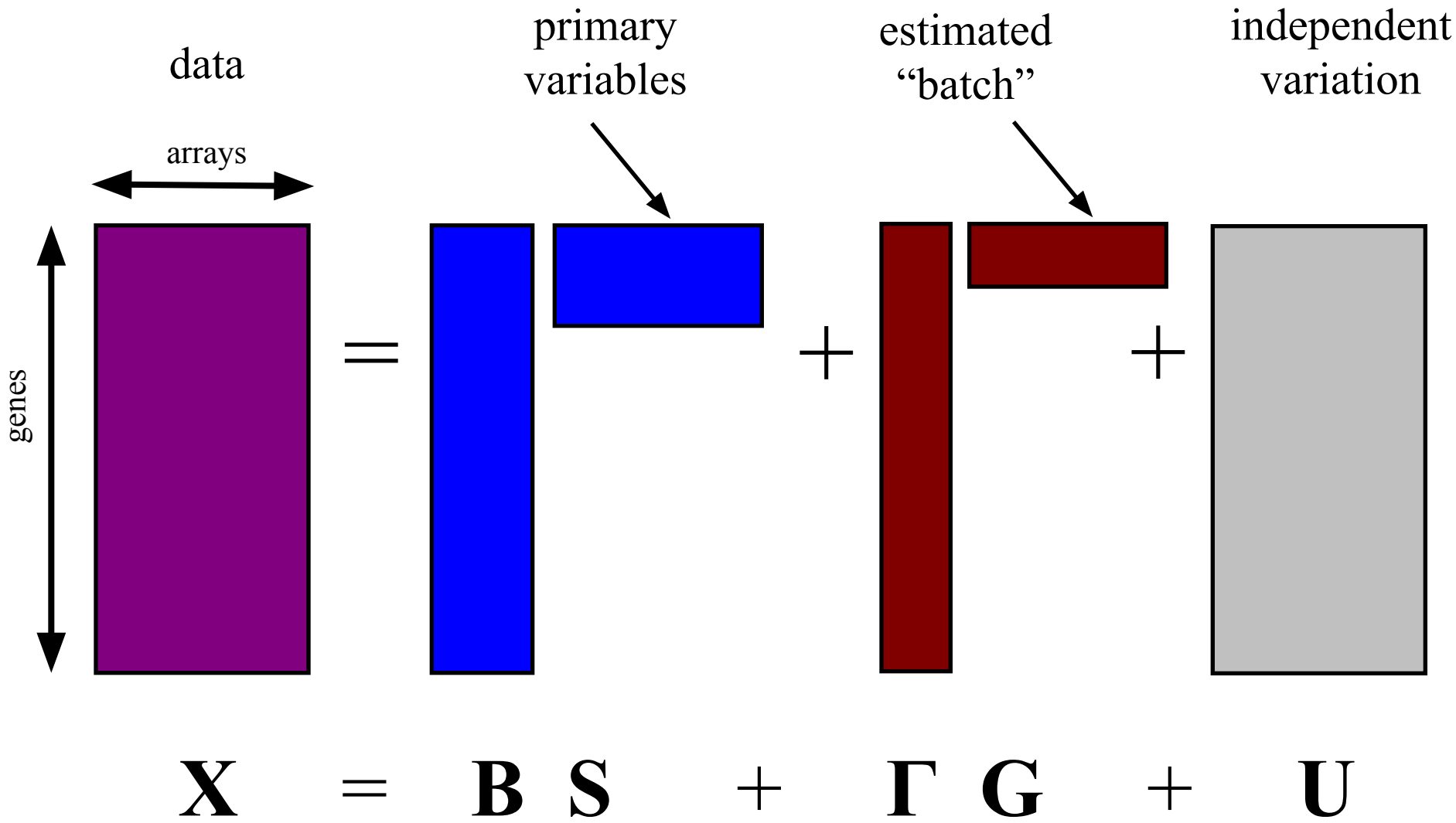


Alert me to new issues

SEE WHAT
THE
AGENTS
ARE
SAYING!



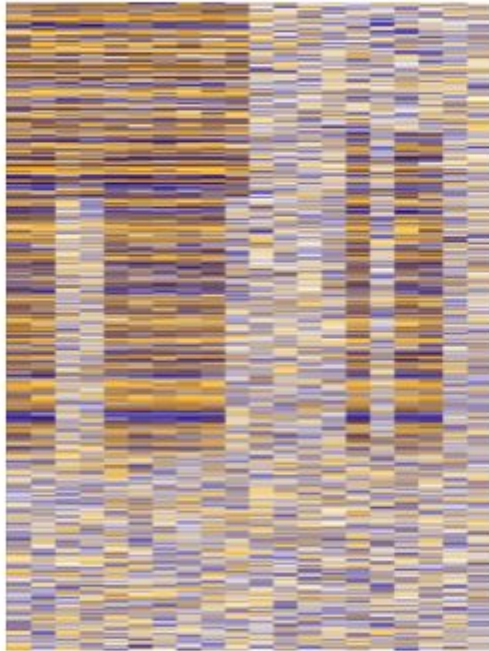




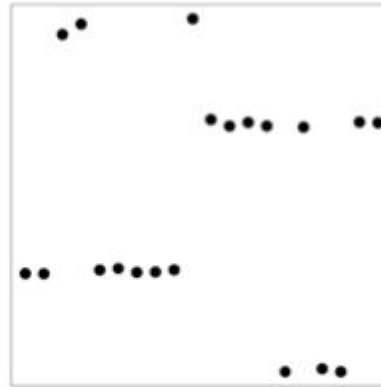
Surrogate Variable Analysis

The Data

$\Pr(\text{Group} \& \text{Batch})$



Estimate of Batch



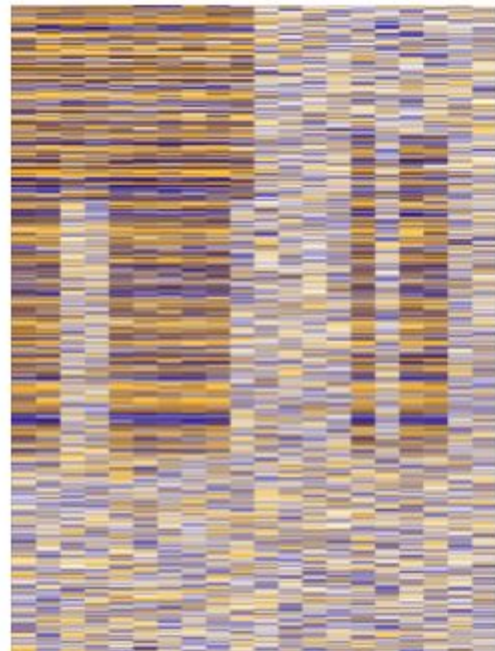
True Batch



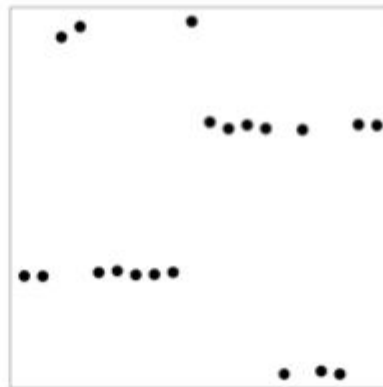
Surrogate Variable Analysis

The Data

$\Pr(\text{!Group} \ \& \ \text{Batch})$



Estimate of **Batch**



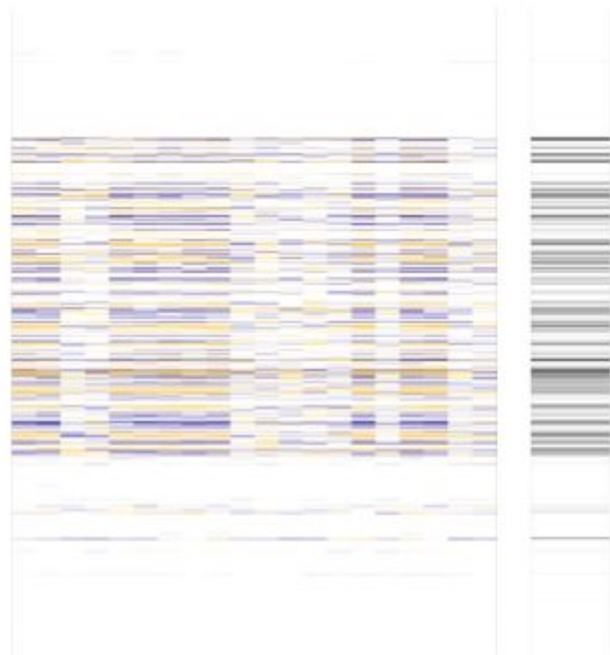
True **Batch**



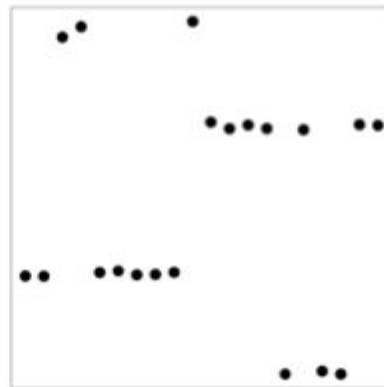
Surrogate Variable Analysis

The Data

$\Pr(\text{!Group \& Batch})$



Estimate of **Batch**



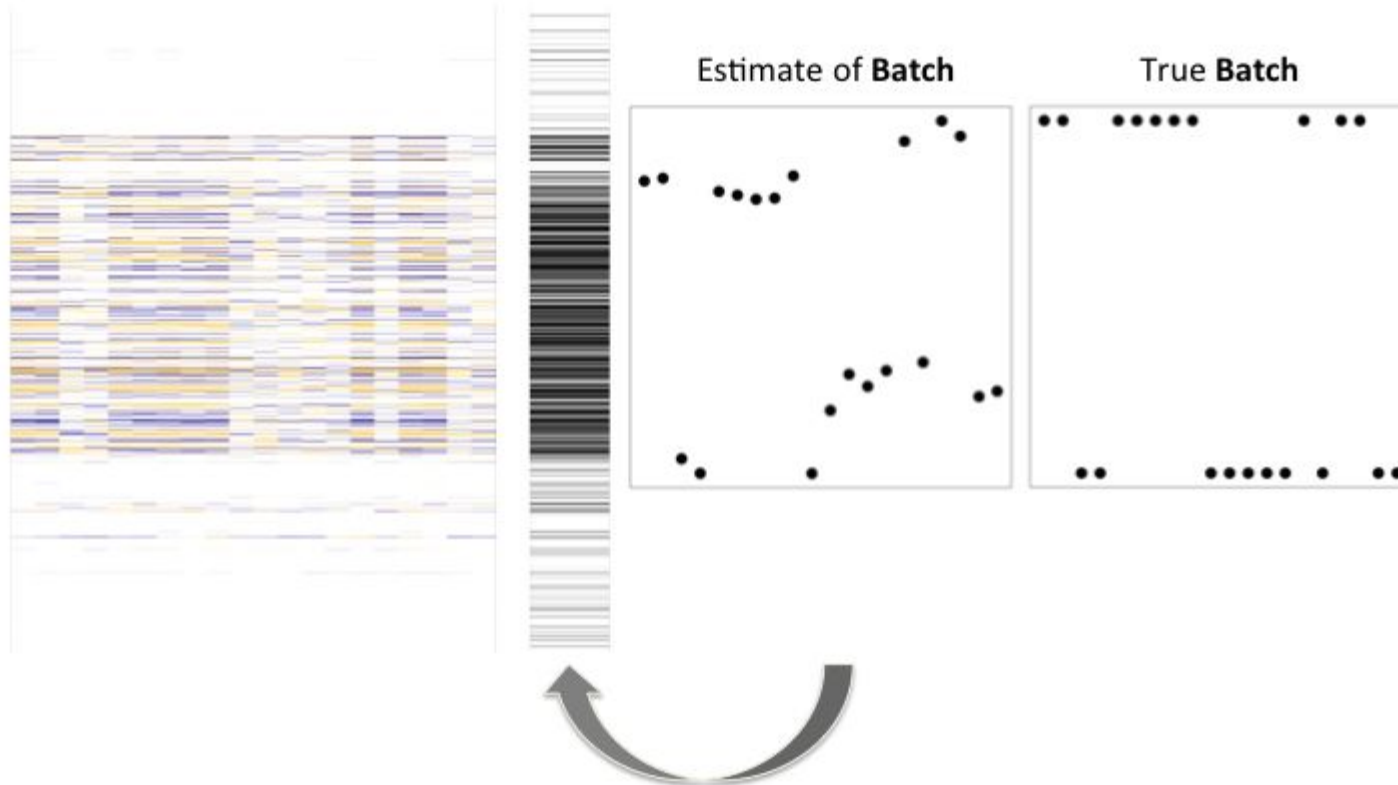
True **Batch**



Surrogate Variable Analysis

The Data

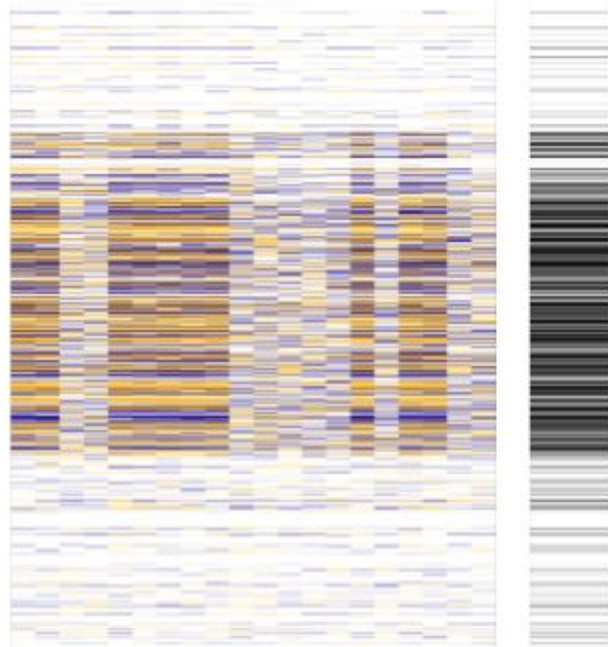
$\Pr(\text{Group} \& \text{Batch})$



Surrogate Variable Analysis

The Data

$\Pr(!\text{Group} \& \text{Batch})$



Estimate of **Batch**



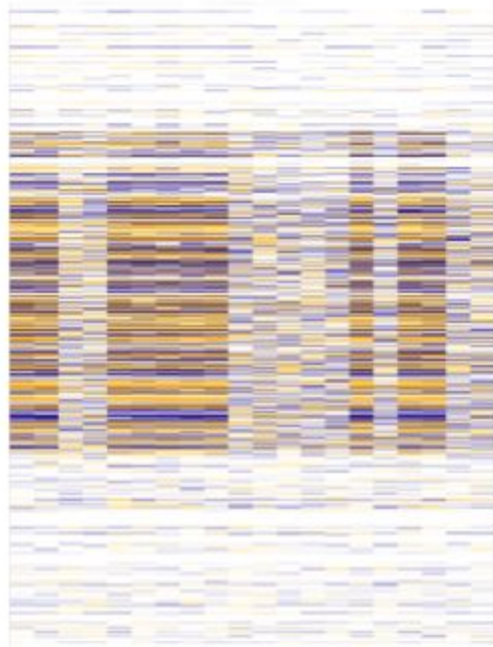
True **Batch**



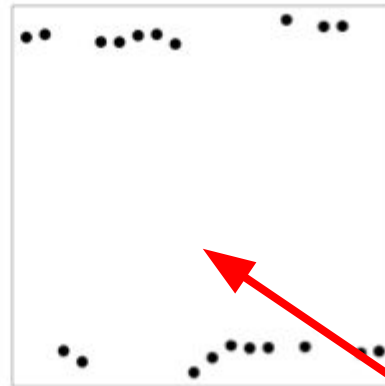
Surrogate Variable Analysis

The Data

$\Pr(\text{Group} \& \text{Batch})$



Estimate of Batch



True Batch



Adjust for this!



Notes and further reading

- Introduction to batch effects
 - <http://www.nature.com/nrg/journal/v11/n10/abs/nrg2825.html>
- Introduction to batch effects with linear models
 - <http://biostatistics.oxfordjournals.org/content/8/1/118.abstract>
- Surrogate variable analysis
 - <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.0030161>