Trajectory Modeling by Shape

Nicholas P. Jewell
Departments of Statistics &
School of Public Health (Biostatistics)
University of California, Berkeley

Victorian Centre for Biostatistics

Murdoch Children's Research Institute

March 6, 2014

Thanks

- Joint work with Brianna Heggeseth (Williams College)

References

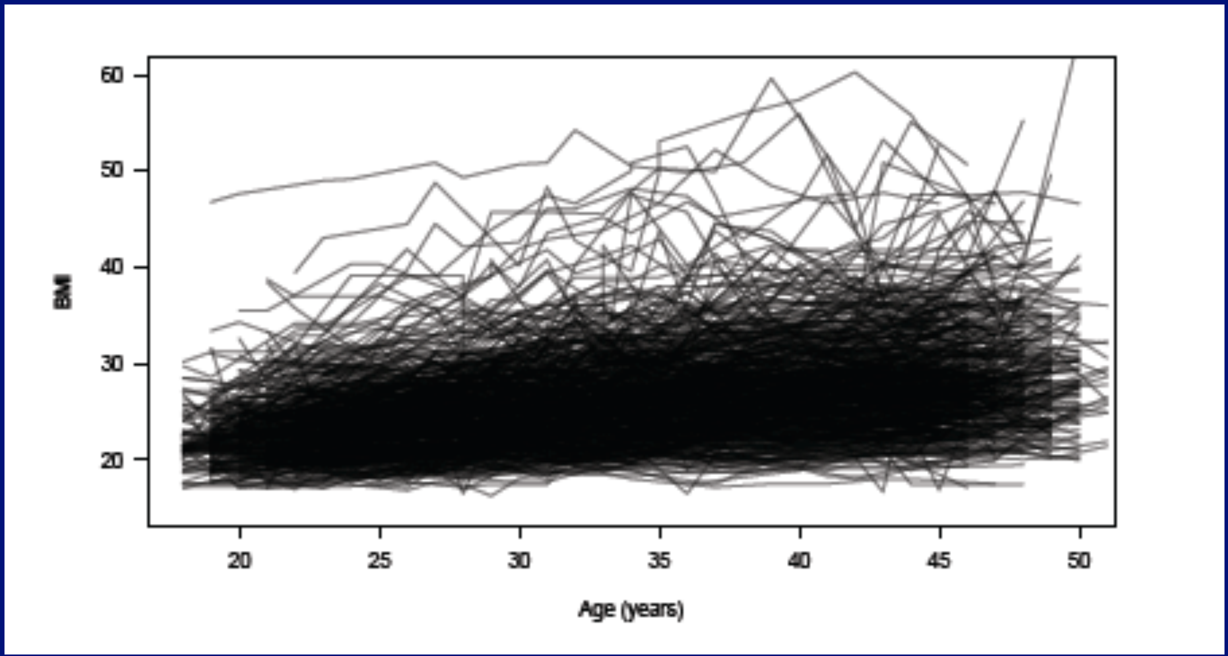
- Heggeseth, BC and Jewell, NP. The impact of covariance misspecification in multivariate Gaussian mixtures on estimation and inference: an application to longitudinal modeling. *Statistics in Medicine*, 2013, 32, 2790-2803 .

- Heggeseth, BC and Jewell, NP. Vertically shifted mixture models for clustering longitudinal data by shape. Submitted for publication.

"Understanding our world requires conceptualizing the similarities and differences between the entities that compose it"

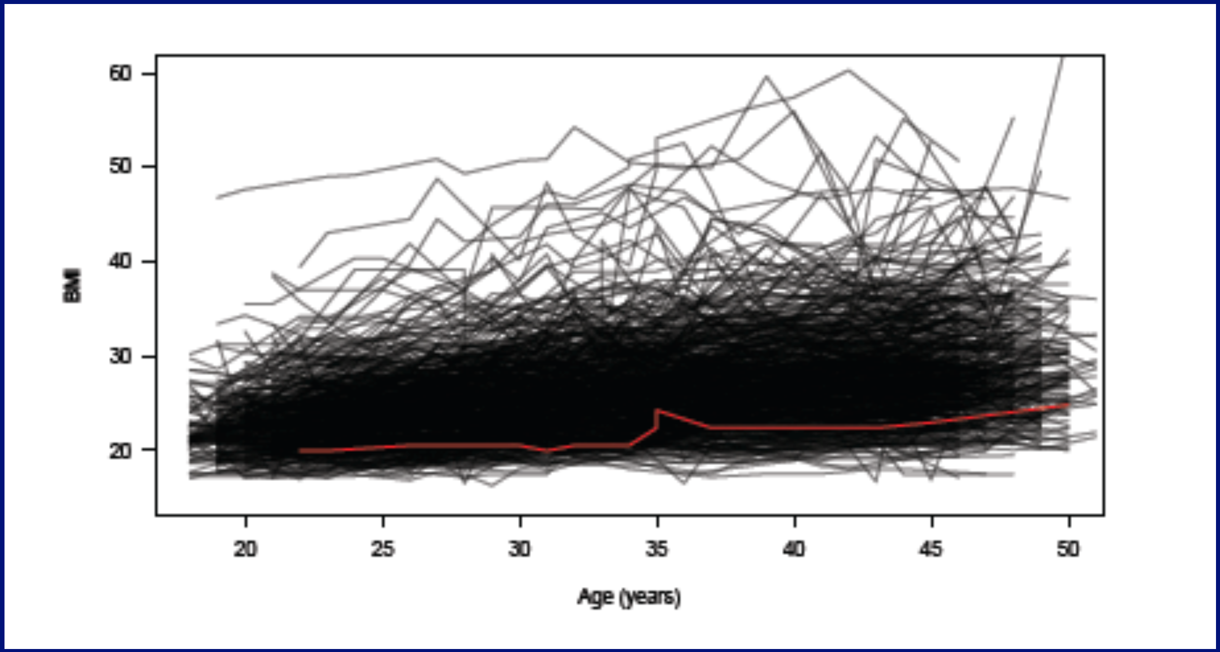
Robert Tryon and Daniel Bailey, 1970

How does BMI change with age?



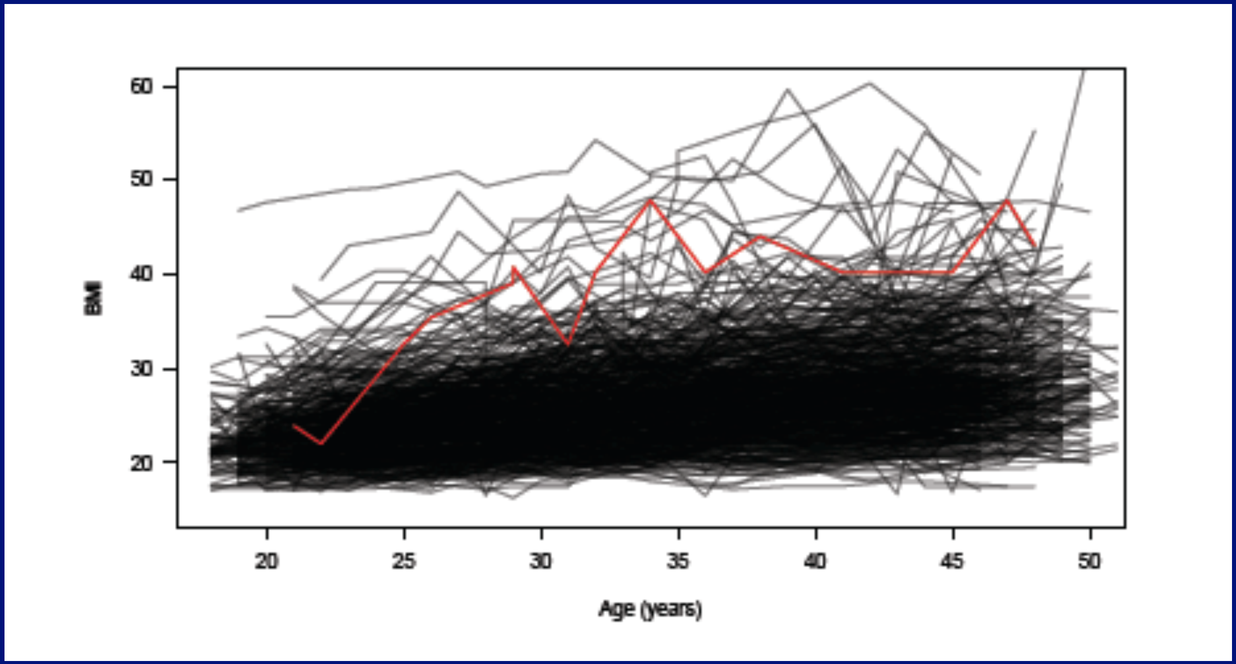
National Longitudinal Study of Youth (NLSY) from 1979 - 2008.

How does BMI change with age?



National Longitudinal Study of Youth (NLSY) from 1979 - 2008.

How does BMI change with age?



National Longitudinal Study of Youth (NLSY) from 1979 - 2008.

Typical Longitudinal Analysis

- Use Generalized Estimating Equations (GEE) to estimate the mean outcome, and how it changes over time, adjusting for covariates
  - regression parameter estimation is consistent despite potential covariance misspecification
  - efficiency can be gained through use of a more appropriate working correlation structure
  - robust (sandwich) standard error estimators available

- But, with a heterogeneous population,
  - BMI does not change much for some people as they age
  - BMI changes considerably for some people as they age

- We don't wish to average out these separate trajectories by modeling the mean over time

Finite Mixture Models

- **Data for *n* individuals:** $\mathbf{y_i} = (\mathbf{y_{i1}}, \ldots, \mathbf{y_{im}})$ **measured at times** $t_i = (t_{i1}, \ldots, t_{im_i})$

- **We assume *K* latent trajectories in the population that are distributed with frequencies:** $\pi_1, \ldots, \pi_K$ **where** $\pi_k > 0$ **and** $\Sigma_{k=1}^{K} \pi_k = 1$**.**

$$f(\mathbf{y}|\mathbf{t}, \theta) = \pi_{\mathbf{1}} \mathbf{f}(\mathbf{y}|\mathbf{t}, \beta_{\mathbf{1}}, \mathbf{\Sigma_1}) + \cdots + \pi_{\mathbf{K}} \mathbf{f}(\mathbf{y}|\mathbf{t}, \beta_{\mathbf{K}}, \mathbf{\Sigma_K})$$

- **The (conditional) mixture density is** $f(\mathbf{y}|\mathbf{t}, \beta_{\mathbf{k}}, \mathbf{\Sigma_k})$**, a multivariate Gaussian with mean** μ_k **and covariance** $\sum_k$ **.**

- **In most trajectory software, (conditional) independence is assumed as a working correlations structure:** $(\Sigma_k = \sigma_k^2 I)$.

$$\theta = (\pi_1, \ldots, \pi_K; \beta_1, \ldots, \beta_K; \Sigma_1, \ldots, \Sigma_K)$$

Finite Mixture Models

- The mean vector μ_k is related to the observation times as follows:
  - Linear: $(\mu_k)_j = \beta_0 + \beta_1 t_{ij}$
  - Quadratic: $(\mu_k)_j = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2$
  - Splines in observation times

  where the regression model (and coefficients) are assumed the same for each cluster, and t_{ij} is the j^{th} observation for the i^{th} individual where $1 \leq j \leq m_i$

Finite Mixture Models

- **Group membership:** $$\pi_k = \frac{exp(\gamma_k z)}{\sum_{j=1}^{K} exp(\gamma_j z)}$$

Z **is set of same or different covariates**

This expands θ **to include the** γ **s also**

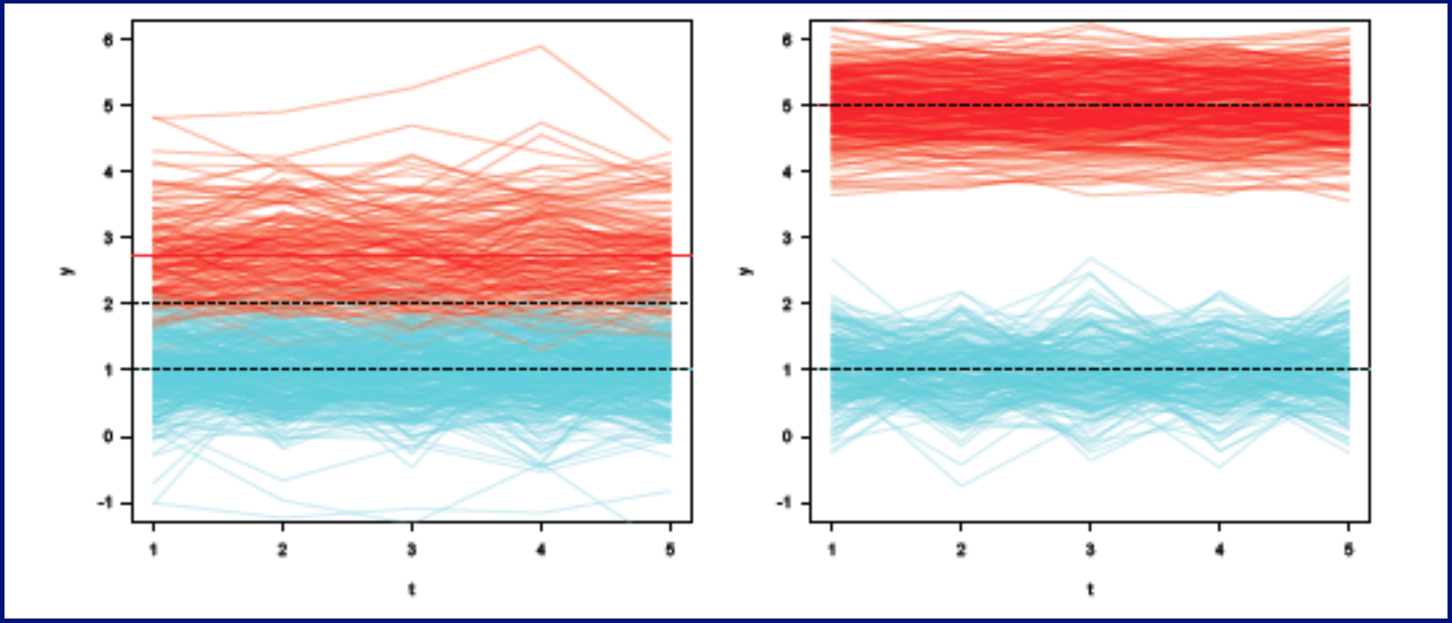
Estimation for Mixture Models

- **Maximum likelihood estimation for $\boldsymbol{\theta}$ via the EM algorithm**

- ***K*** **is pre-specified; can be chosen using the BIC**
- **Parameter estimators are not consistent under covariance misspecification (White, 1982; Heggeseth and Jewell, 2013).**
- **Robust (sandwich) standard error estimators are available.**
- **How bad can the bias in regression estimators be? What influences its size?**

Mispecified Covariance Structure Bias and Separation of Trajectories

- Separated components lead to little bias even when you wrongly assume independence.

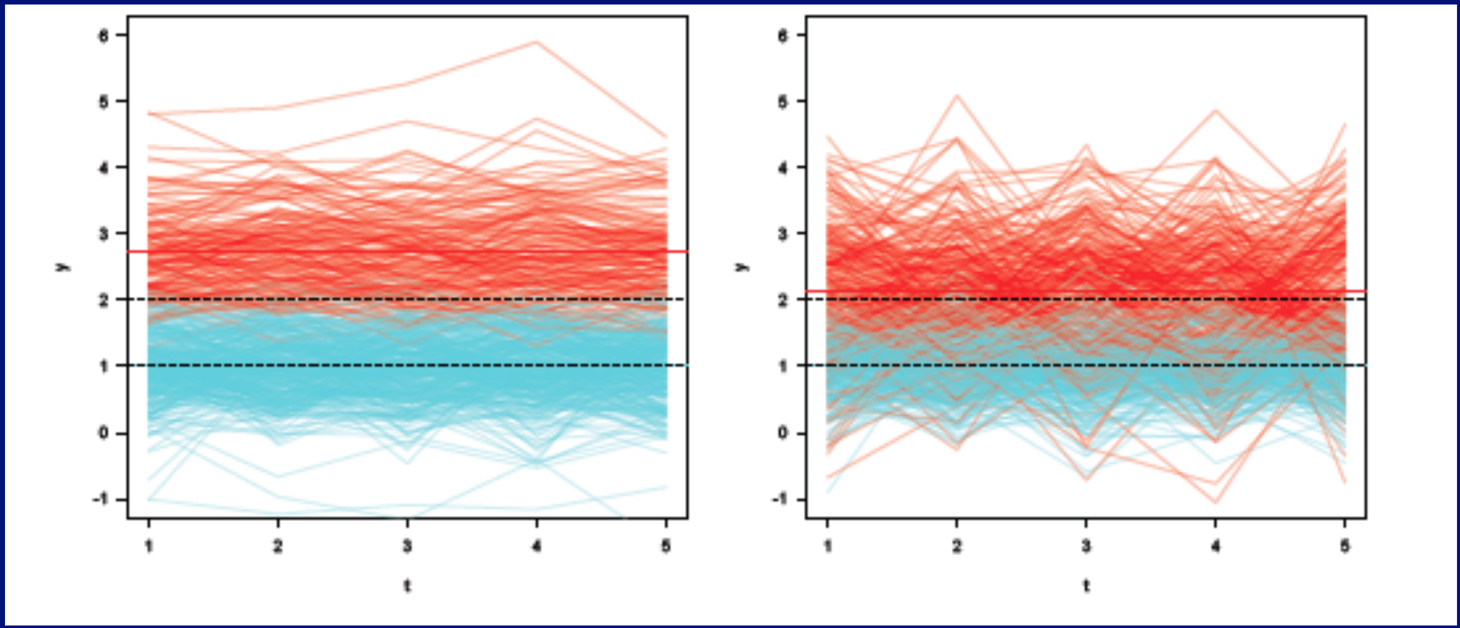


Black dashed -- true means, Solid lines – estimated means

$\hat{SE}_I(\beta_{01}) = 0.02, \hat{SE}_R(\beta_{01}) = 0.06$ $\hat{SE}_I(\beta_{01}) = 0.01, \hat{SE}_R(\beta_{01}) = 0.01$

Mispecified Covariance Structure Bias and Level of Dependence

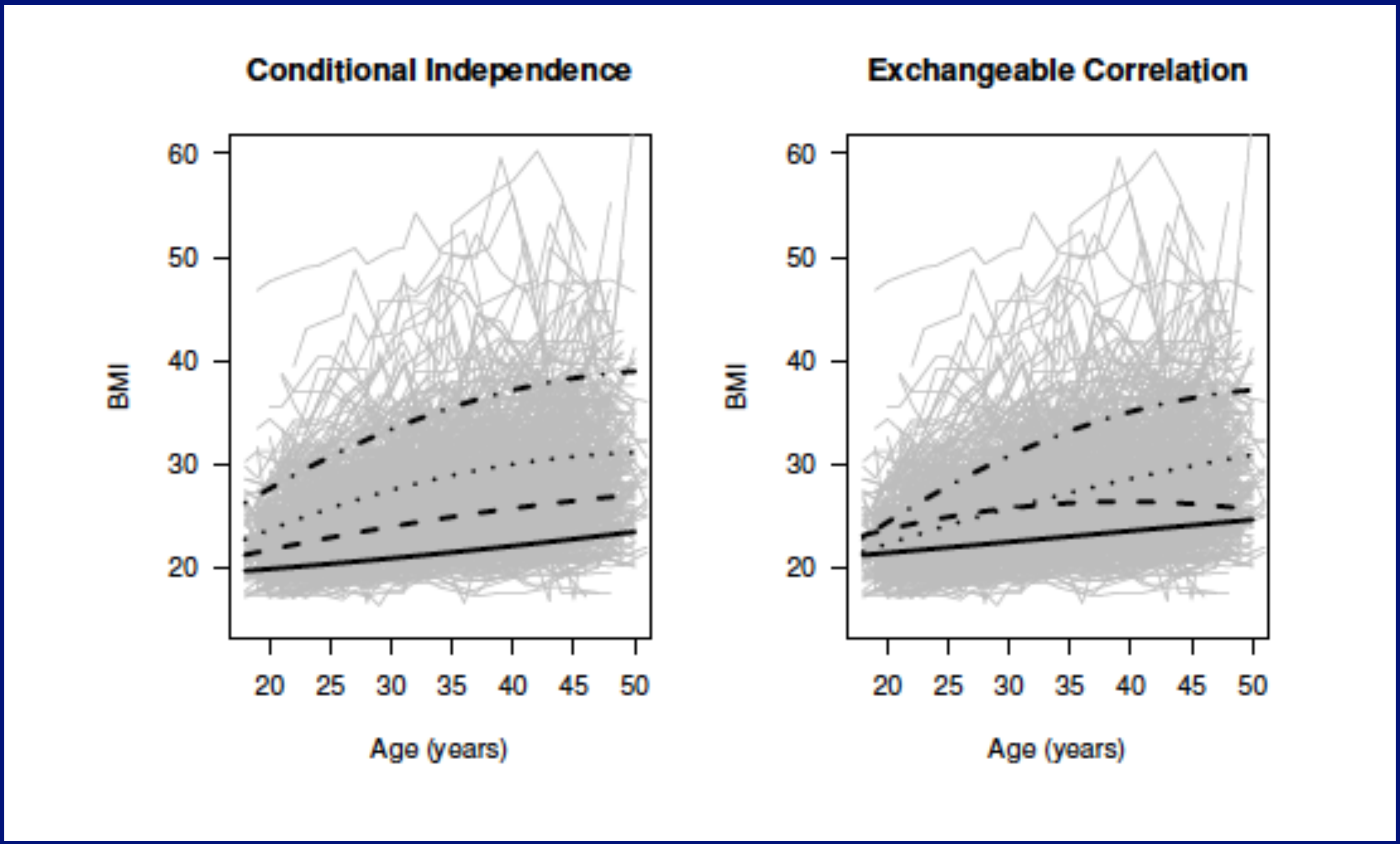
- **Components with little dependence lead to small bias even when you wrongly assume independence.**



Black dashed -- true means, Solid lines – estimated means

$\hat{SE}_I(\beta_{01}) = 0.02, \hat{SE}_R(\beta_{01}) = 0.06$ $\hat{SE}_I(\beta_{01}) = 0.03, \hat{SE}_R(\beta_{01}) = 0.04$

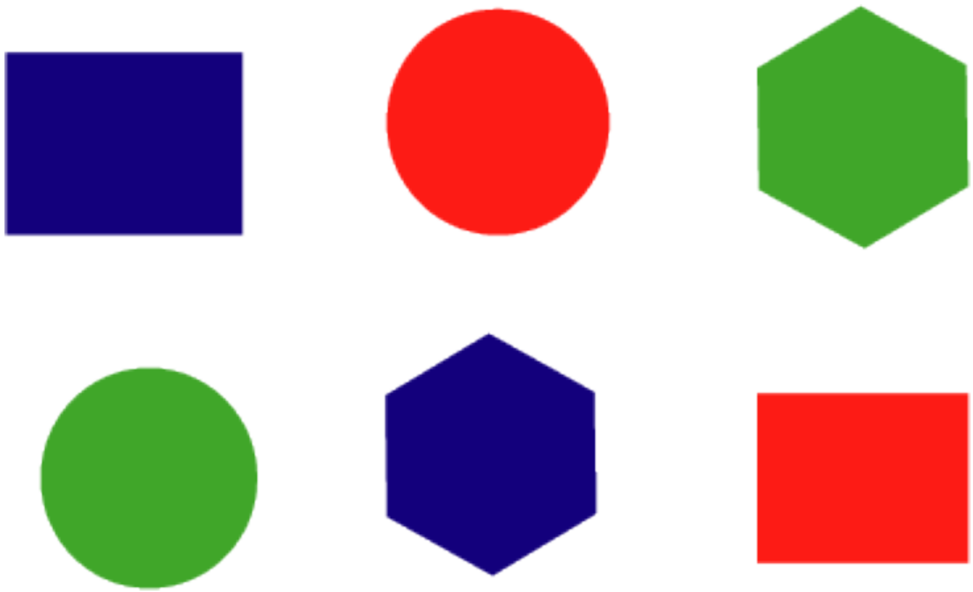
NLSY Data Analysis



Covariance makes a difference to the trajectories

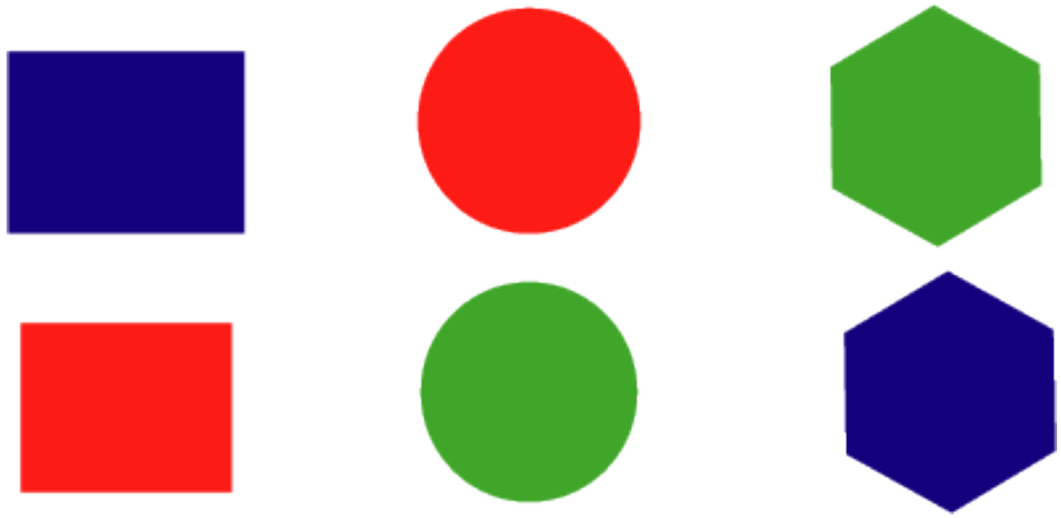
- **hard to estimate bias from mispecified covariance**

How Do We Group These Blocks?

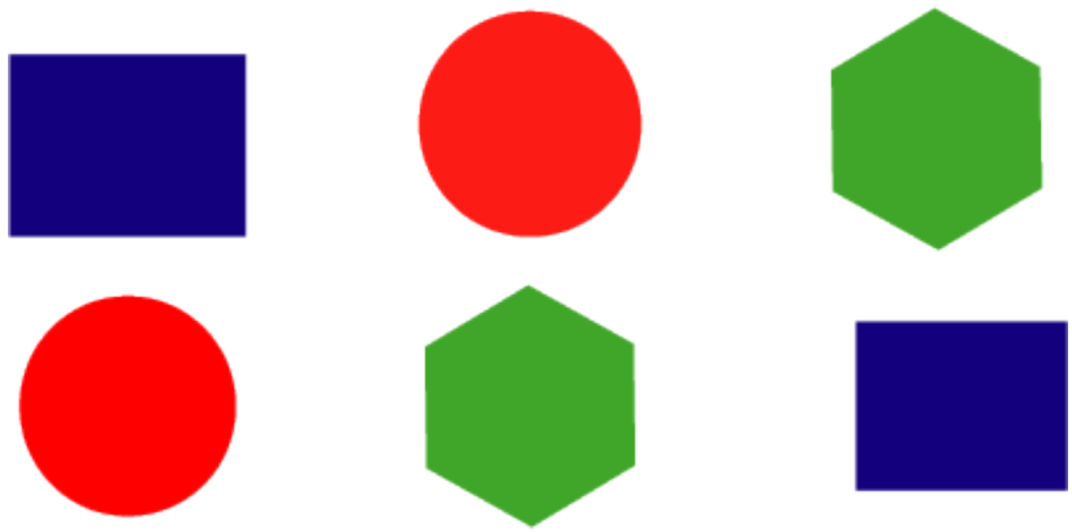


Group by Color

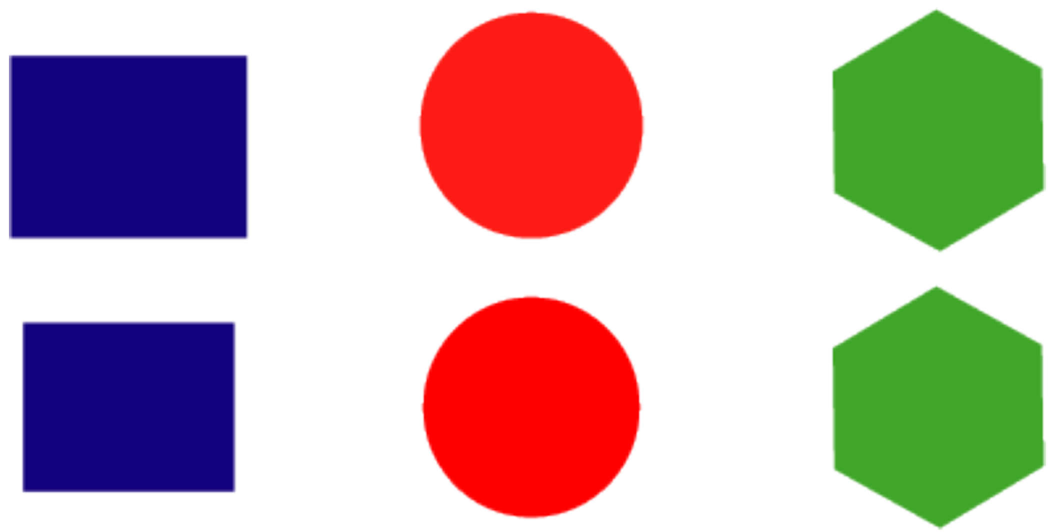
Group by Shape



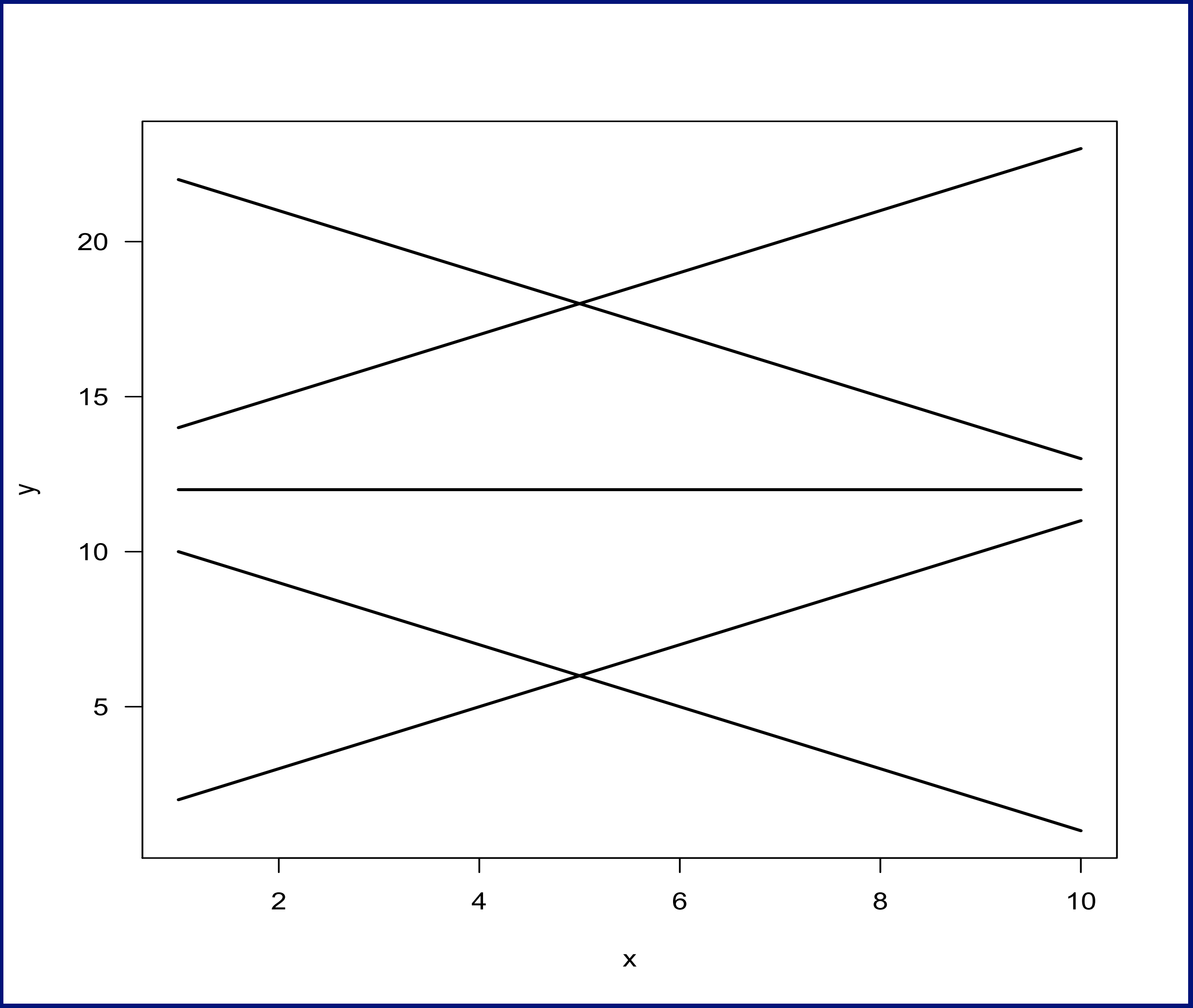
How Do We Group These Blocks?



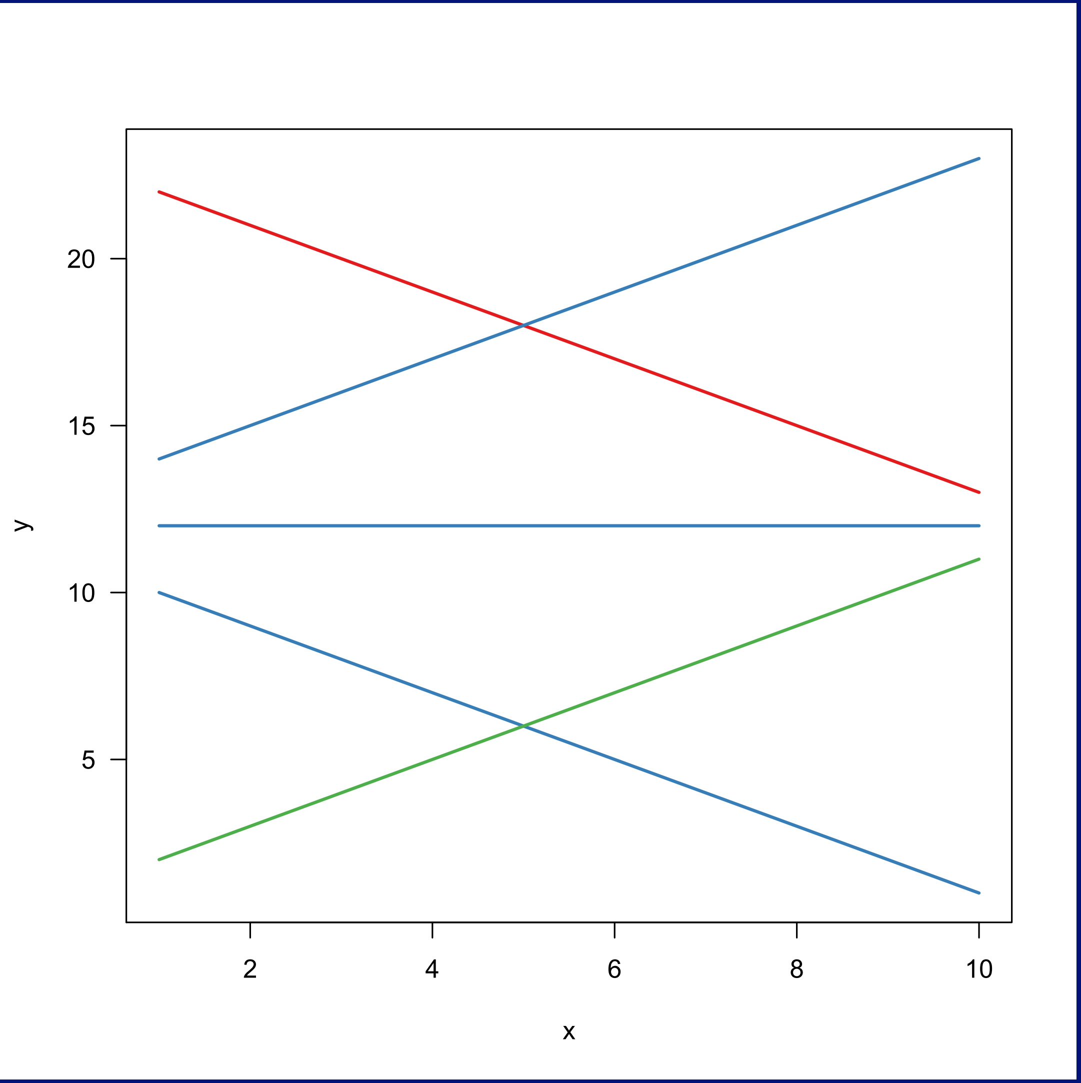
Group by Color or Shape



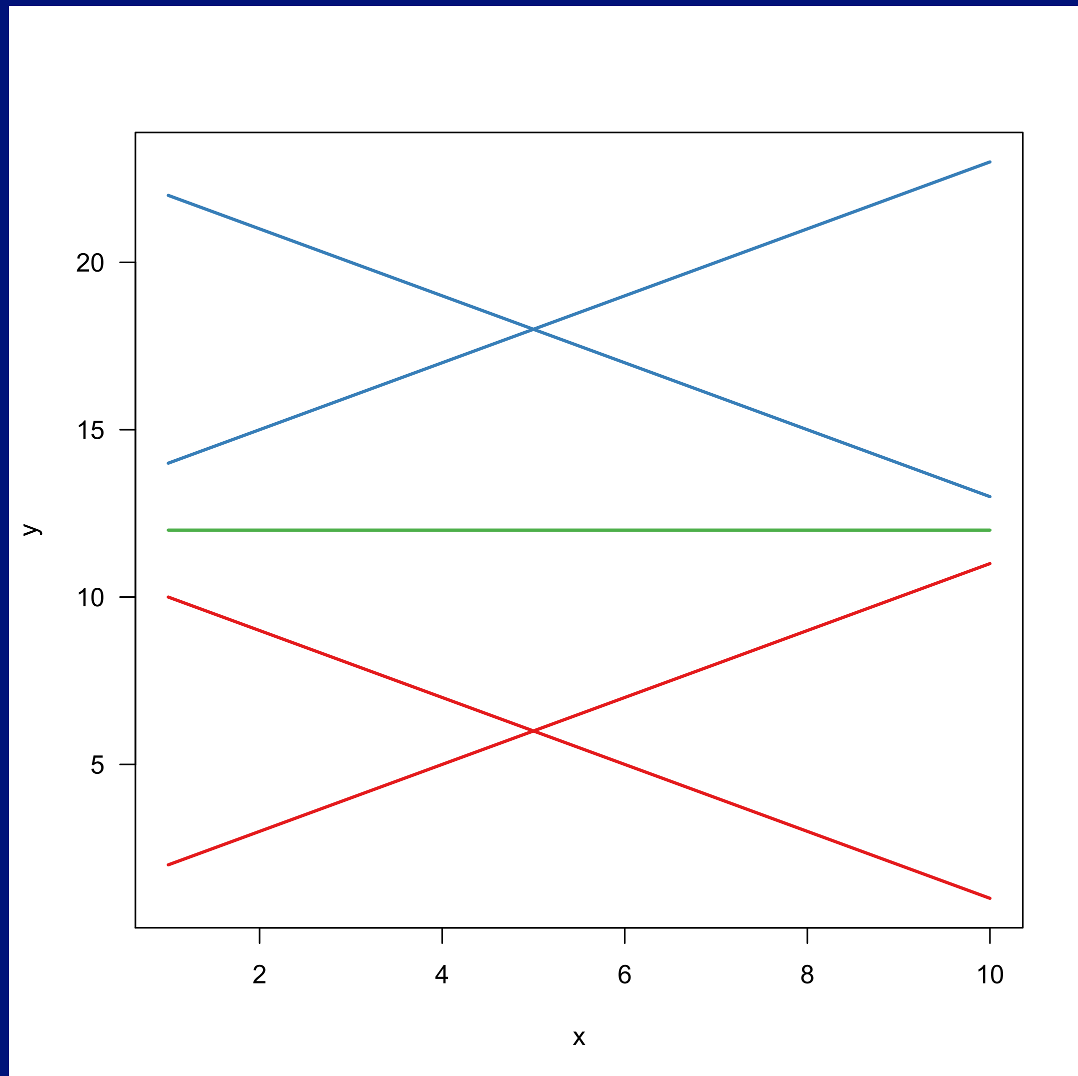
How Do We Group These (Regression) Lines?



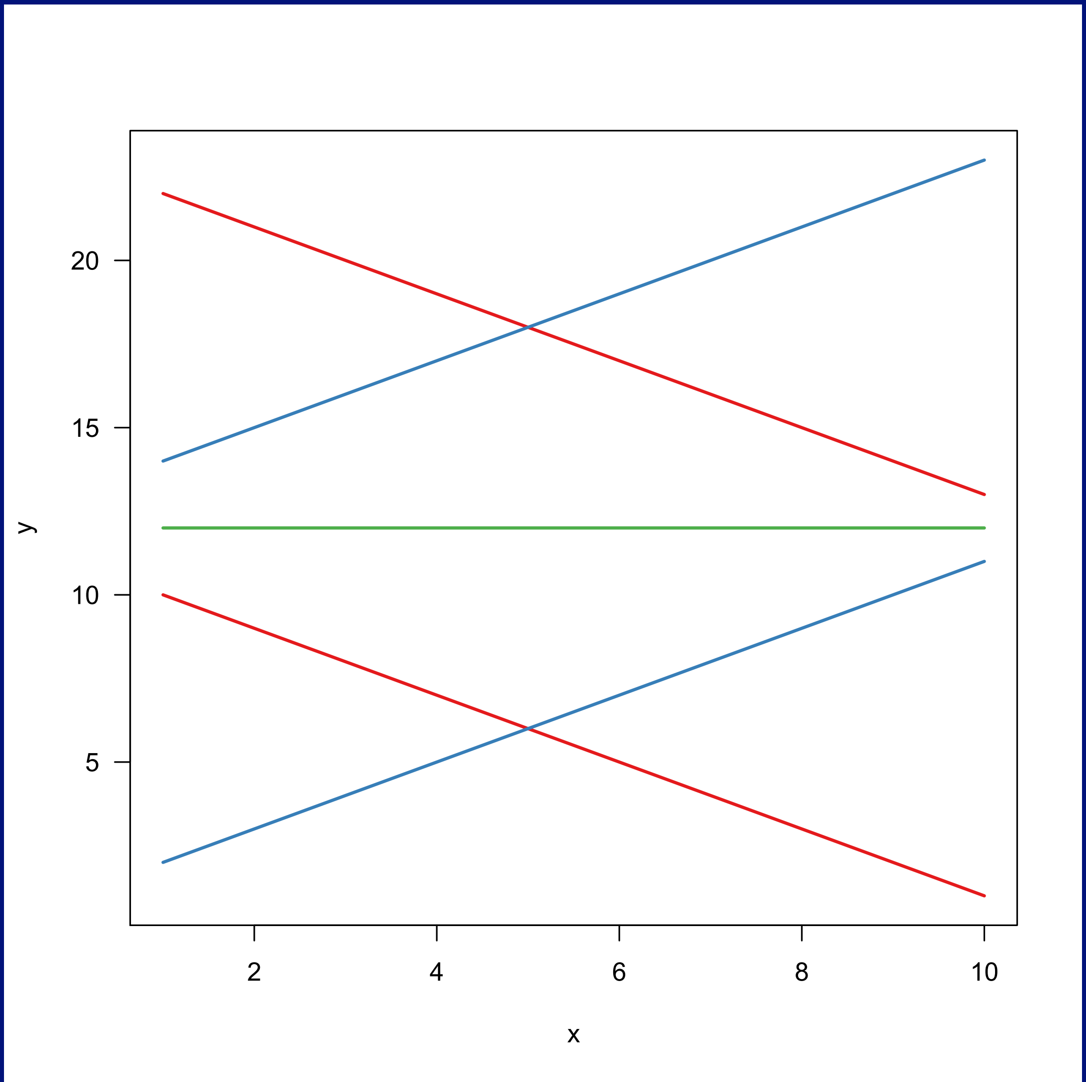
Group by Intercept



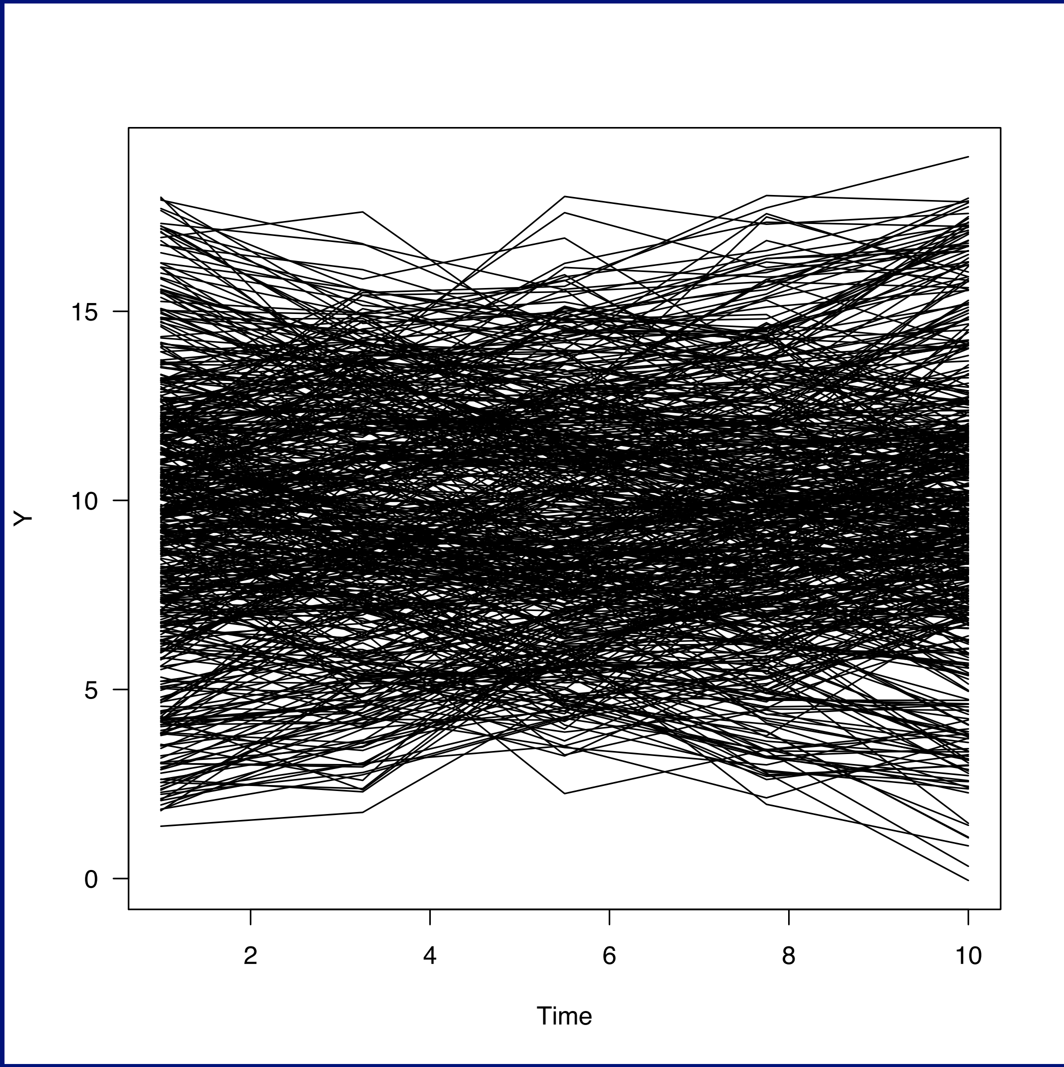
Group by Level



Group by Shape (Slope)

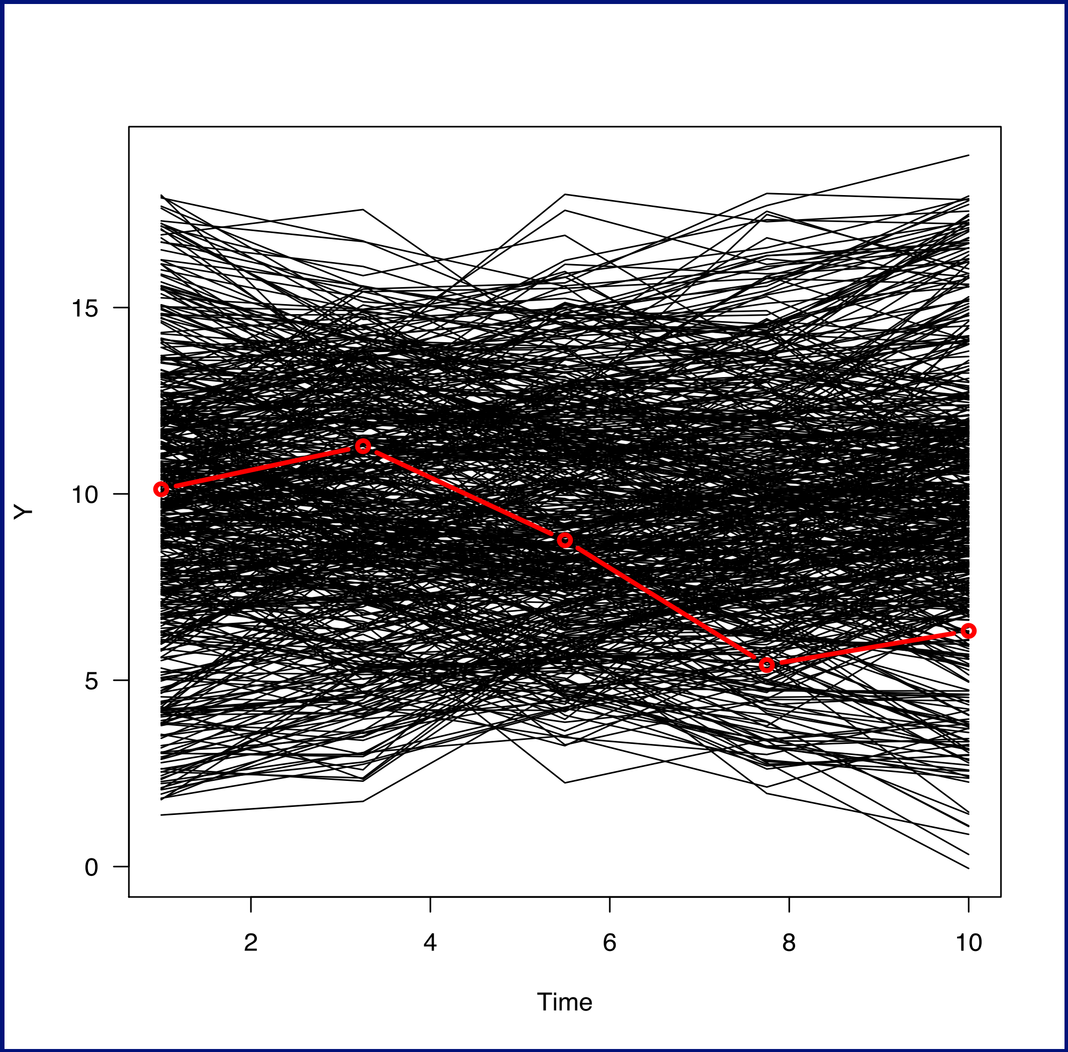


Simulated Data



How could we group these individuals?

Simulated Data



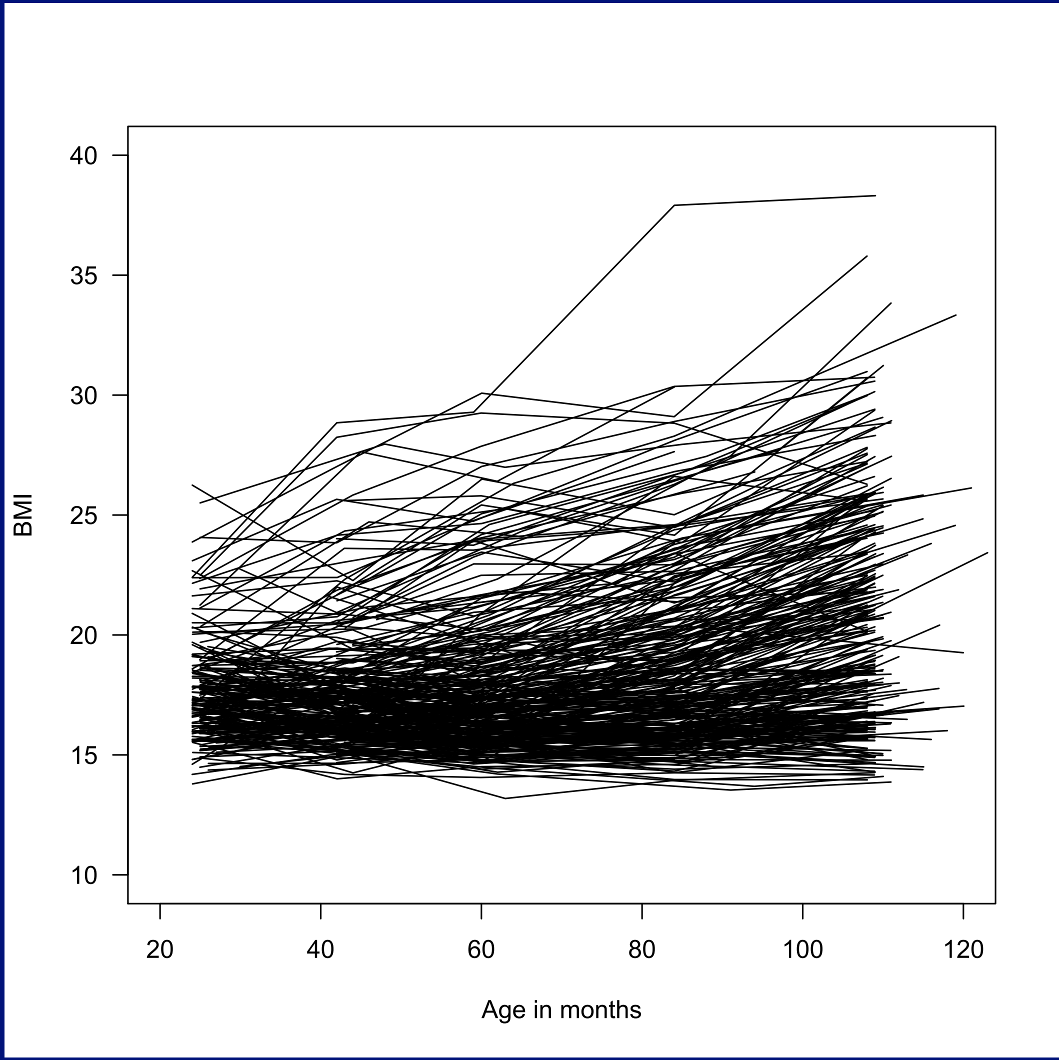
How could we group these individuals?

Real Longitudinal Data

- Center for the Health Assessment of Assessment of Mothers and Children of Salinas (CHAMACOS) Study
  - In 1999-2000, enrolled 601 pregnant women in agricultural Salinas Valley, CA.
  - Mostly Hispanic, agricultural workers.
  - Determine if exposure to pesticides and other chemicals impact children's growth patterns (BMI, neurological measures etc_.

- First, focus on studying/estimating the growth patterns of children.
- Second, determine if early life predictors are related to the patterns
  - pesticide/chemical exposure in utero
    - ODT, PDT, PDE, BPA (bisphenol A)

CHAMACOS Data



How could we group these individuals?

Cluster Analyses

- Clustering is the task of assigning a set of objects into groups so that the objects in the same group are more similar to each other than to those in other groups.
- What does it mean for objects to be more similar or more dissimilar?
  - Distance matrix
- Why do we cluster objects?

Standard Clustering Methods

- Partition methods
  - Partition objects into K groups so that an objective function of dissimilarities is minimized or maximized.
  - Example: K-means Algorithm

- Model-based methods
  - Assume a model that includes a grouping structure and estimate parameters.
  - Example: Finite Mixture Models

K-means algorithm

- Input: Data for *n* individuals in vector form. For individual *i* , the observed data vector is

$$\mathbf{y_i} = (\mathbf{y_{1i}}, \ldots, \mathbf{y_{im}}).$$

- Measure of Dissimilarity: Squared Euclidean distance. The dissimilarity between the 1st and 2nd individuals is

$$d(\mathbf{y_1} - \mathbf{y_2}) = \|\mathbf{y_1} - \mathbf{y_2}\|^{\mathbf{2}} = (\mathbf{y_{11}} - \mathbf{y_{12}})^{\mathbf{2}} + \cdots + (\mathbf{y_{im}} - \mathbf{y_{2m}})^{\mathbf{2}}$$

K-means Algorithm

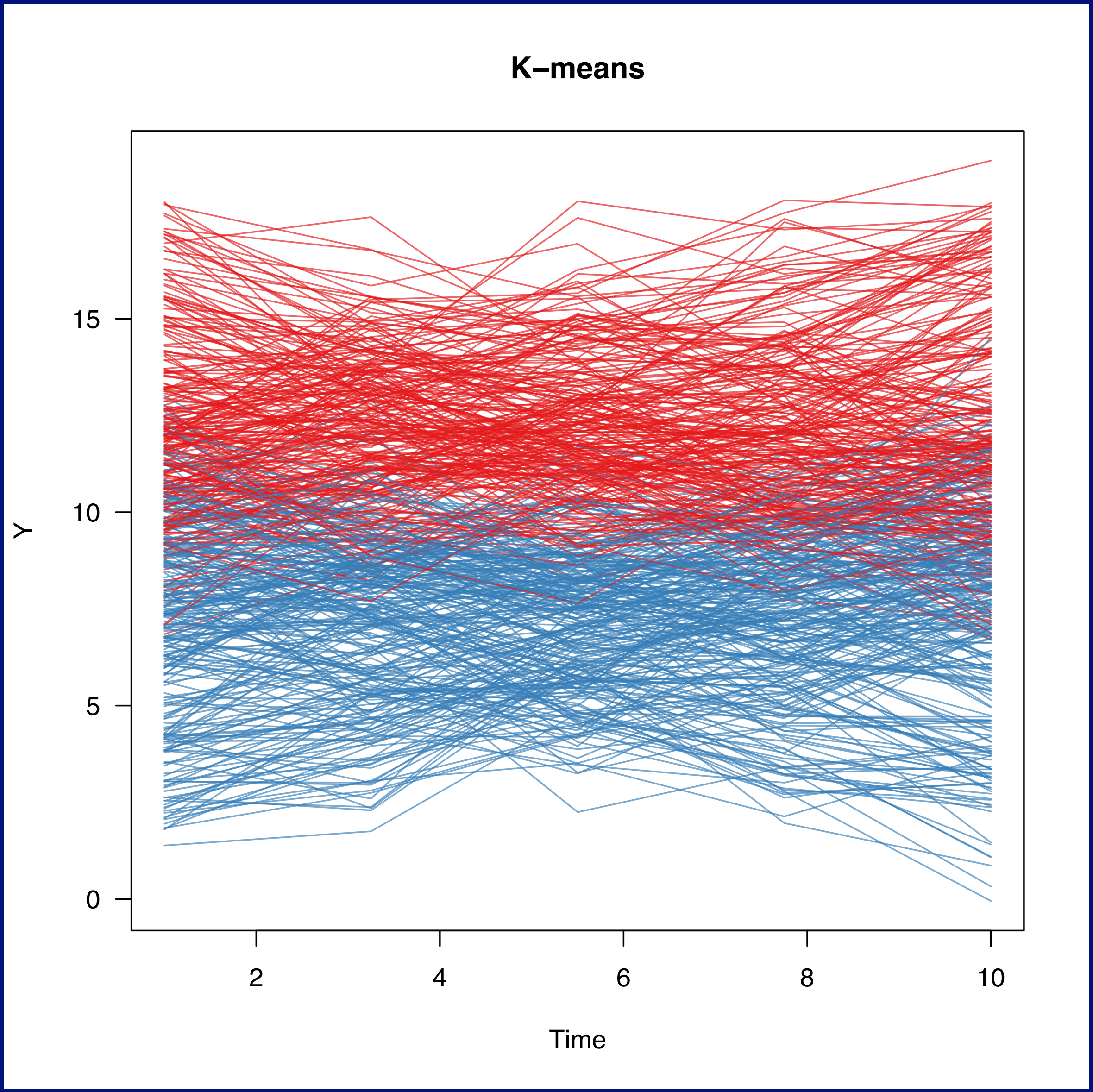
- Goal: Partition individuals into *K* sets $\mathbf{C} = \{\mathbf{C_1}, \mathbf{C_2}, \ldots, \mathbf{C_K}\}$ so as to minimize the within-cluster sum of squares

$$\sum_{k=1}^{K} \sum_{\mathbf{y_i} \in \mathbf{C_k}} \|\mathbf{y_i} - \mu_\mathbf{k}\|^\mathbf{2}$$

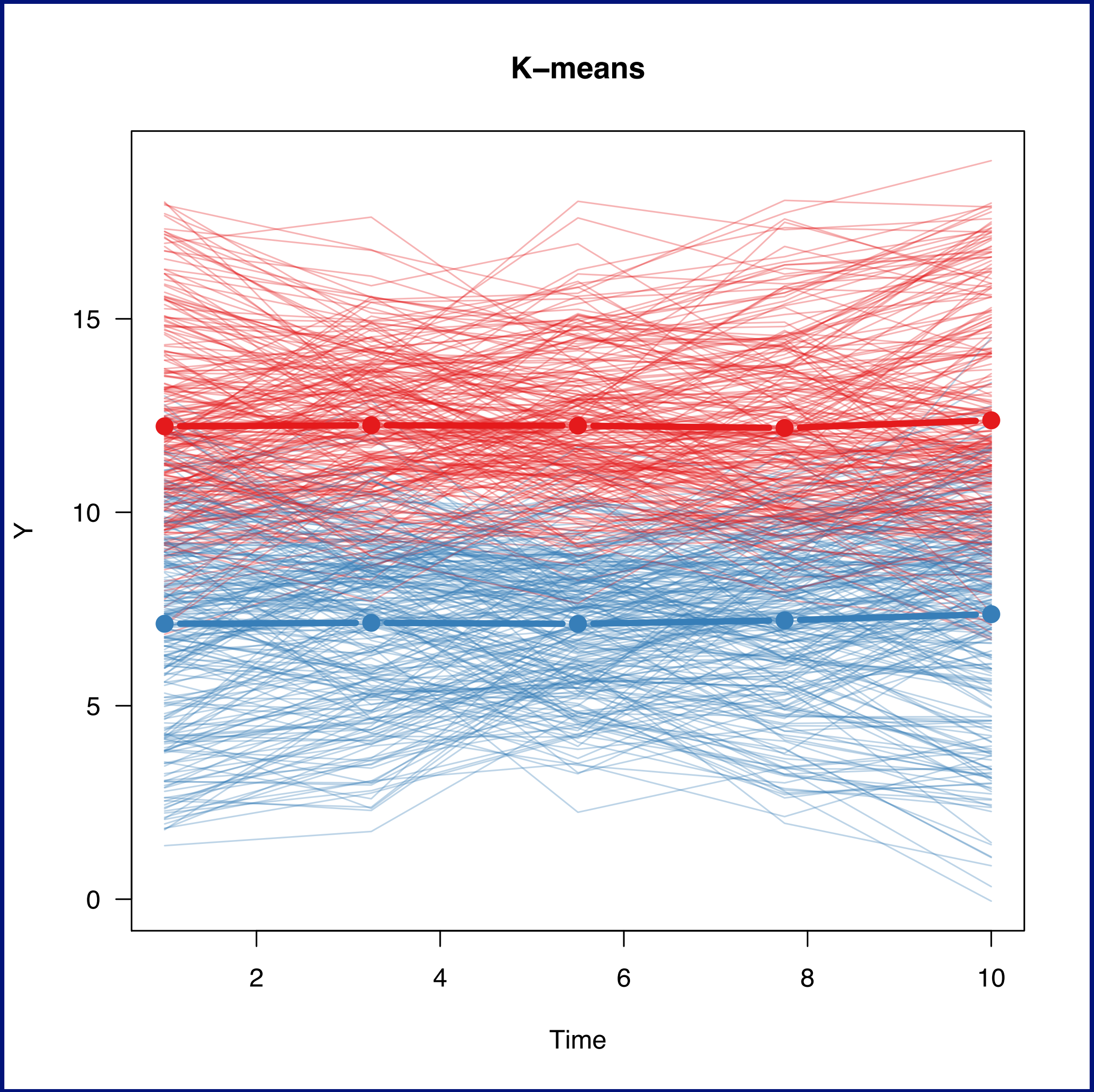
where μ_k is the mean vector of individuals in C_k.

(*K* must be known before starting *K*-means. There are many ways to choose *K* from the data that try to minimize the dissimilarity within each cluster while maximizing the dissimilarity between clusters: for example, the use of *silhouettes*.)

Application to Simulated Data

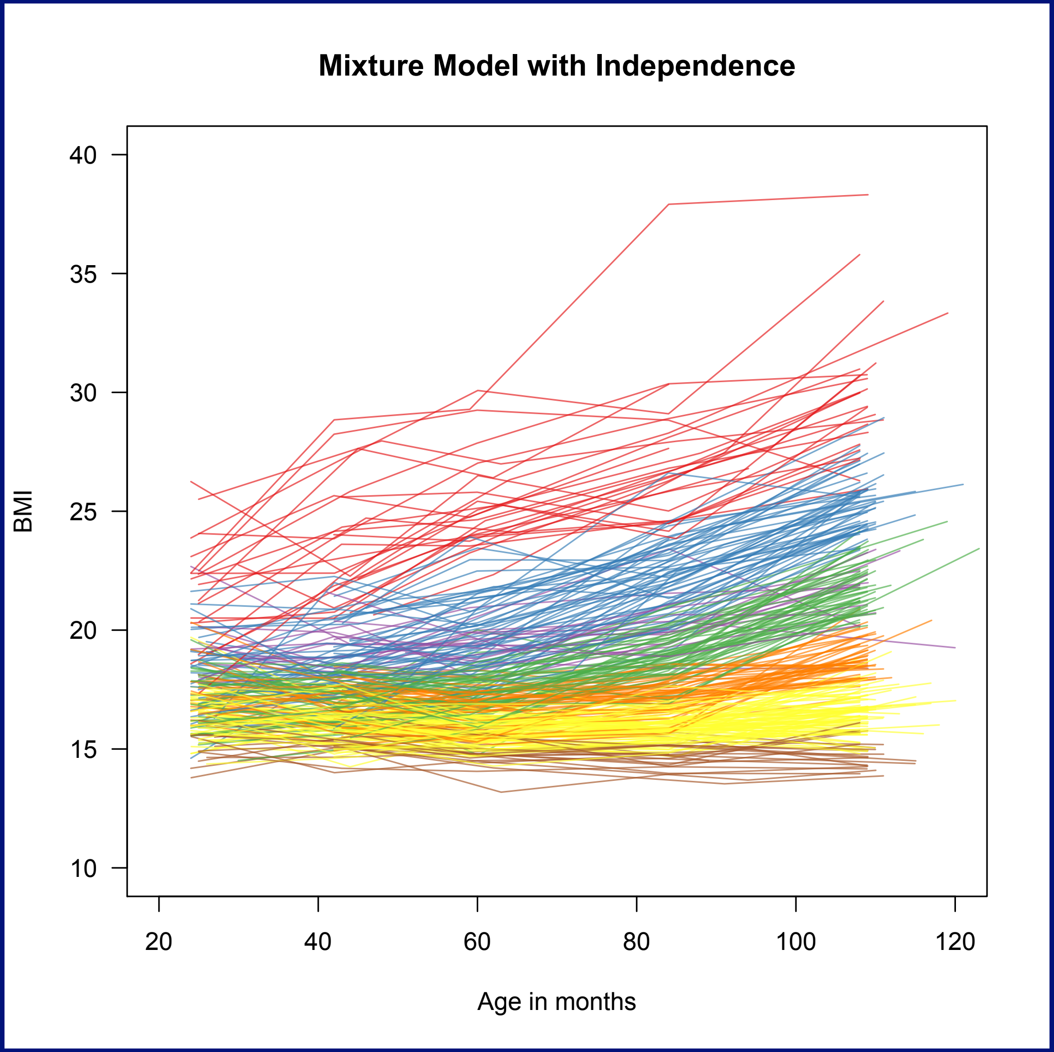


Application to Simulated Data

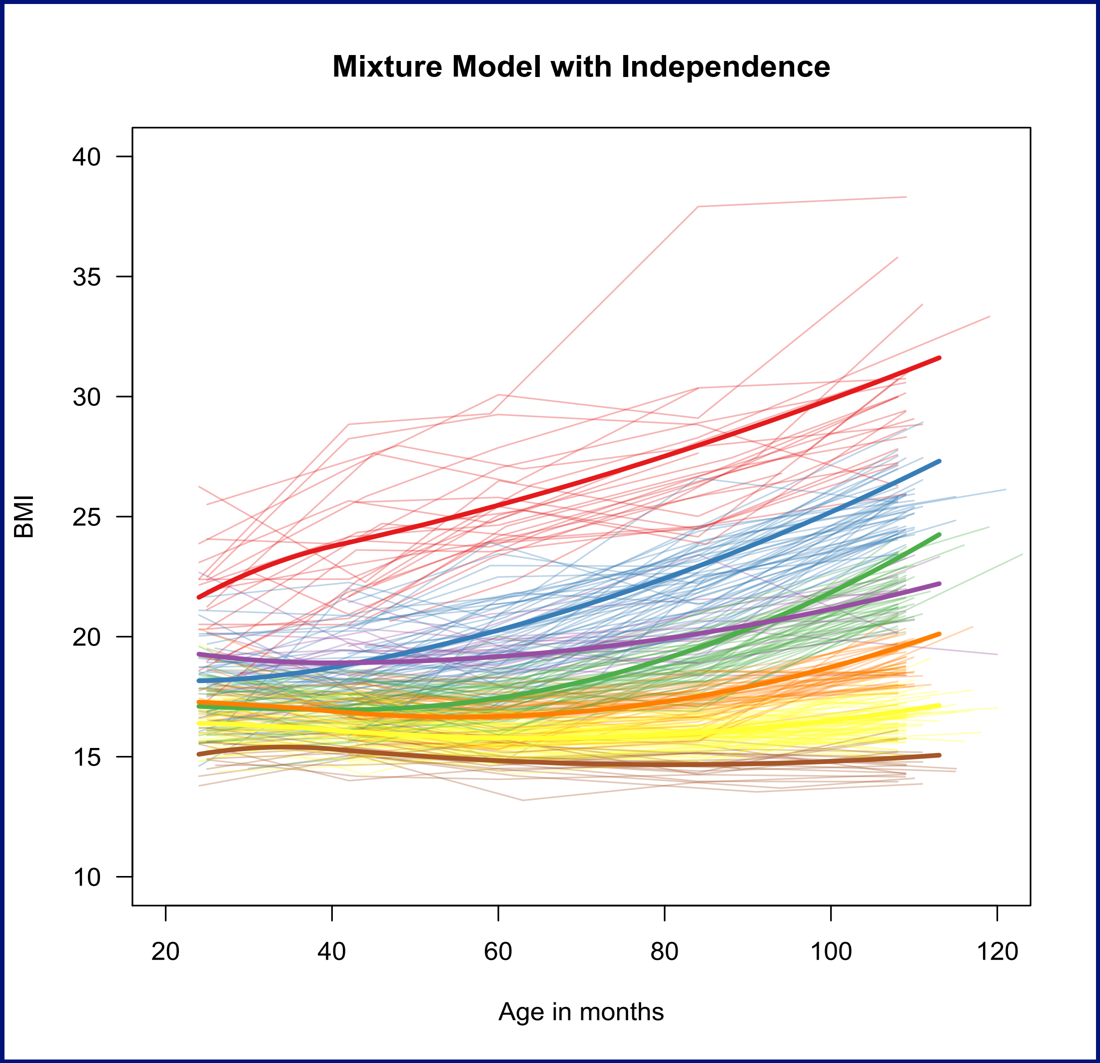


How would you describe—interpret—the group trajectories?

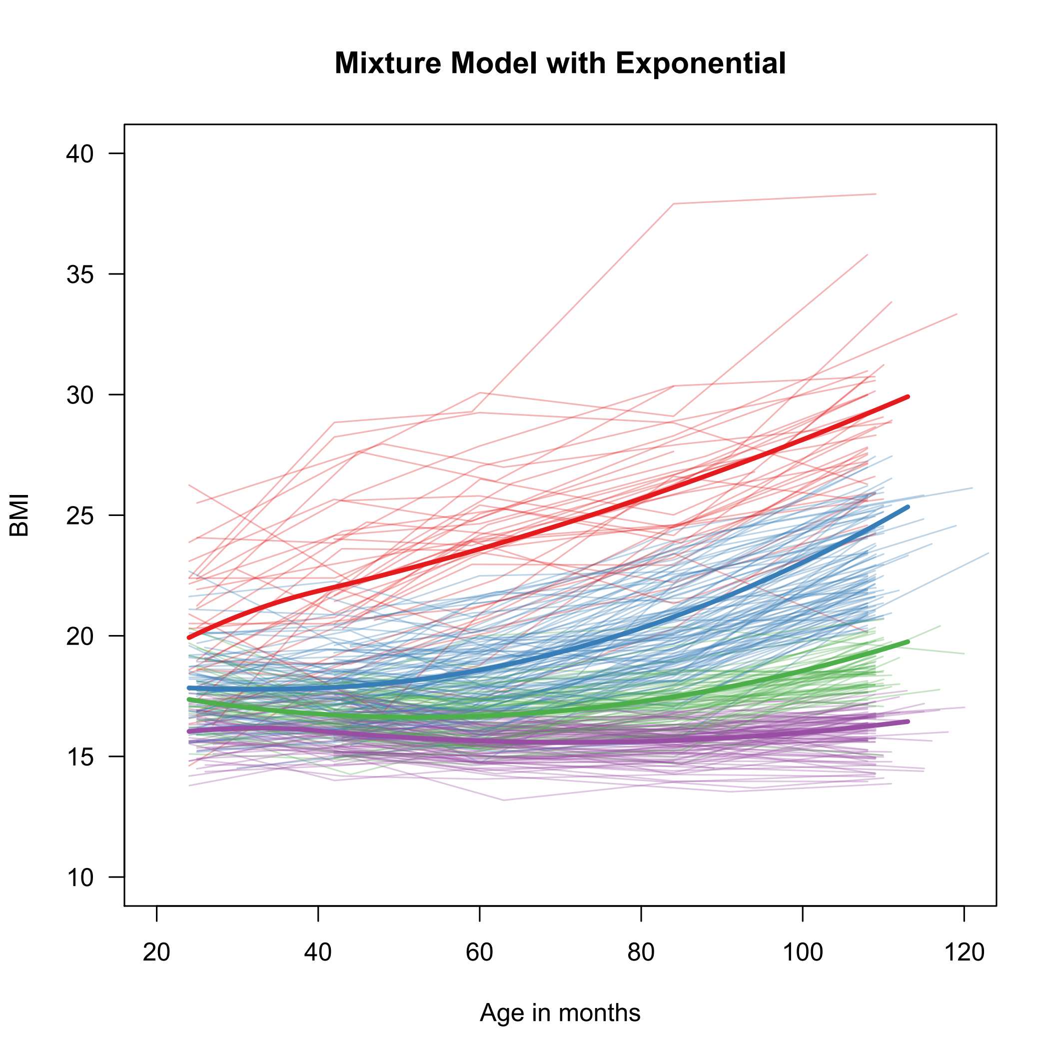
Finite Mixture Model Applied to CHAMACOS Data



Finite Mixture Model Applied to CHAMACOS Data



Finite Mixture Model Applied to CHAMACOS Data



Clustering by Shape

- Interested in *shape* not just level (which appears to dominate clustering techniques)
- Want a method that:
  - Works with irregularly sampled data
  - Includes a way to estimate the relationship between baseline risk factors and group membership
  - Groups individuals according to the outcome pattern over time ignoring the level

Clustering by Shape Options

- Estimate slopes between neighboring observations and cluster on the "derived" observations
- Fit splines for each individual, differentiate, and cluster on coefficients of resulting derivative
- Use partition based cluster methods (like PAM) but use (i) the Pearson coefficient as a distance or dissimilarity measure

$$d_{corr}(\mathbf{x}, \mathbf{y}) = 1 - Corr(\mathbf{x}, \mathbf{y})$$

or the cosine-angle measure of dissimilarity

$$d_{cos}(\mathbf{x}, \mathbf{y}) = 1 - \frac{\Sigma_{j=1}^{m} x_j y_j}{(\Sigma_{j=1}^{m} x_j^2)(\Sigma_{j=1}^{m} y_j^2)}$$

- Vertical shifting individual trajectories

Vertical Shifting

- For each individual, calculate

$$y_i^* = y_i - m_i^{-1} \sum_{j=1}^{m_i} y_{ij}$$

- Each individual now has mean zero and so level is removed from any resulting clustering

- Apply clustering technique to shifted data, e.g. finite mixture model

Correlation Models for Vertical Shifted Data

- **Without specifying group, suppose**

$$\mathbf{y}_i^* = \lambda_i I_{m_i} + \mu_i + \epsilon_i, \lambda \sim F_\lambda, \epsilon \sim N(), \Sigma)$$

where I_{m_i} **is an** m_i **length vector of 1s, and**

$\mu_{ij} = \mu_k(t_{ij})$ **is the *j*th element of the vector of mean values for the *k*th group evaluated at the observation times** t_i **. Thus,**

$$\mathbf{y}_i^* = \mathbf{A}_i \mathbf{y}_i = \mu_i - \bar{\mu_i} + \epsilon_i - \bar{\epsilon_i}$$

Correlation Models for Vertical Shifted Data

$$Cov(\mathbf{Y}_i^* - \mu_i) = Cov((A - \mathbf{I}_{m_i})\mu_i + \mathbf{A}\epsilon)$$

Two components of the covariance

- One induced by the averaging process
- One induced by (random) observation times

Correlation Models for Vertical Shifted Data

Observation Times Fixed

$$Cov(\mathbf{Y}^* - \mu) = \mathbf{A}\Sigma\mathbf{A}^T$$

suppressing the individual/group indices for simplicity (Σ is allowed to vary across clusters)

This covariance matrix is singular since $\det(\mathbf{A}) = 0$

This naturally reflects the "loss" of one dimension

Correlation Models for Vertical Shifted Data

Observation Times Fixed

$$Cov(\mathbf{Y}^* - \mu) = \mathbf{A}\Sigma\mathbf{A}^T$$

- If $\Sigma = \sigma^2 \mathbf{I}$ (conditional independence with constant variance, then the induced covariance is exchangeable with negative correlation given by $-1/(m-1)$ and variance decreases to $\sigma^2(\frac{m-1}{m})$
- If original covariance is exchangeable with constant variance and correlation ρ then the induced covariance remains exchangeable with negative correlation and reduced variance $\sigma^2(1-\rho)(\frac{m-1}{m})$

Correlation Models for Vertical Shifted Data

Observation Times Fixed

$$Cov(\mathbf{Y}^* - \mu) = \mathbf{A}\Sigma\mathbf{A}^T$$

If $\Sigma = \sigma^2\mathbf{I}$ (conditional independence with constant variance, then the induced covariance is exchangeable with negative correlation given by $-1/(m-1)$ and variance decreases to $\sigma^2(\frac{m-1}{m})$

This induced exchangeable correlation is the lower bound for correlation in an exchangeable matrix

Thus, if “estimated”, the (true) parameter is on the boundary of the parameter space

Correlation Models for Vertical Shifted Data

Observation Times Random
(μ is random)

$$Cov(\mathbf{Y}^* - \mu) = m^{-2}\left(\Sigma_{j=1}^{m} Var(t_j)[\mu'(E(t_j))]^2\right)\mathbf{1}\mathbf{1}^T + \mathbf{A}\Sigma\mathbf{A}^T$$

Sum of two non-invertible matrices, but the positive magnitude of the first matrix may counteract the negative correlations of the second.

Correlation Models for Vertical Shifted Data

Observation Times Random
(μ is random)

$$Cov(\mathbf{Y}^* - \mu) = m^{-2}\left(\Sigma_{j=1}^{m} Var(t_j)[\mu'(E(t_j))]^2\right)\mathbf{1}\mathbf{1}^T + \mathbf{A}\Sigma\mathbf{A}^T$$

500 simulations of $\mathbf{Y}_i = \mu_i + \epsilon_i$ $\quad\epsilon_i \sim N(0, \Sigma_\rho)$

where the error covariance matrix is of exponential form with range ρ

$$\mathbf{t} = \mathbf{T} + \tau \qquad \tau \sim N(0, \sigma_\tau^2 \mathbf{I})$$

$$\mathbf{T} = (1, 2, \ldots, 9, 10)$$

$$\mu_{ij} = \mu(t_{ij})$$

Correlation Models for Vertical Shifted Data

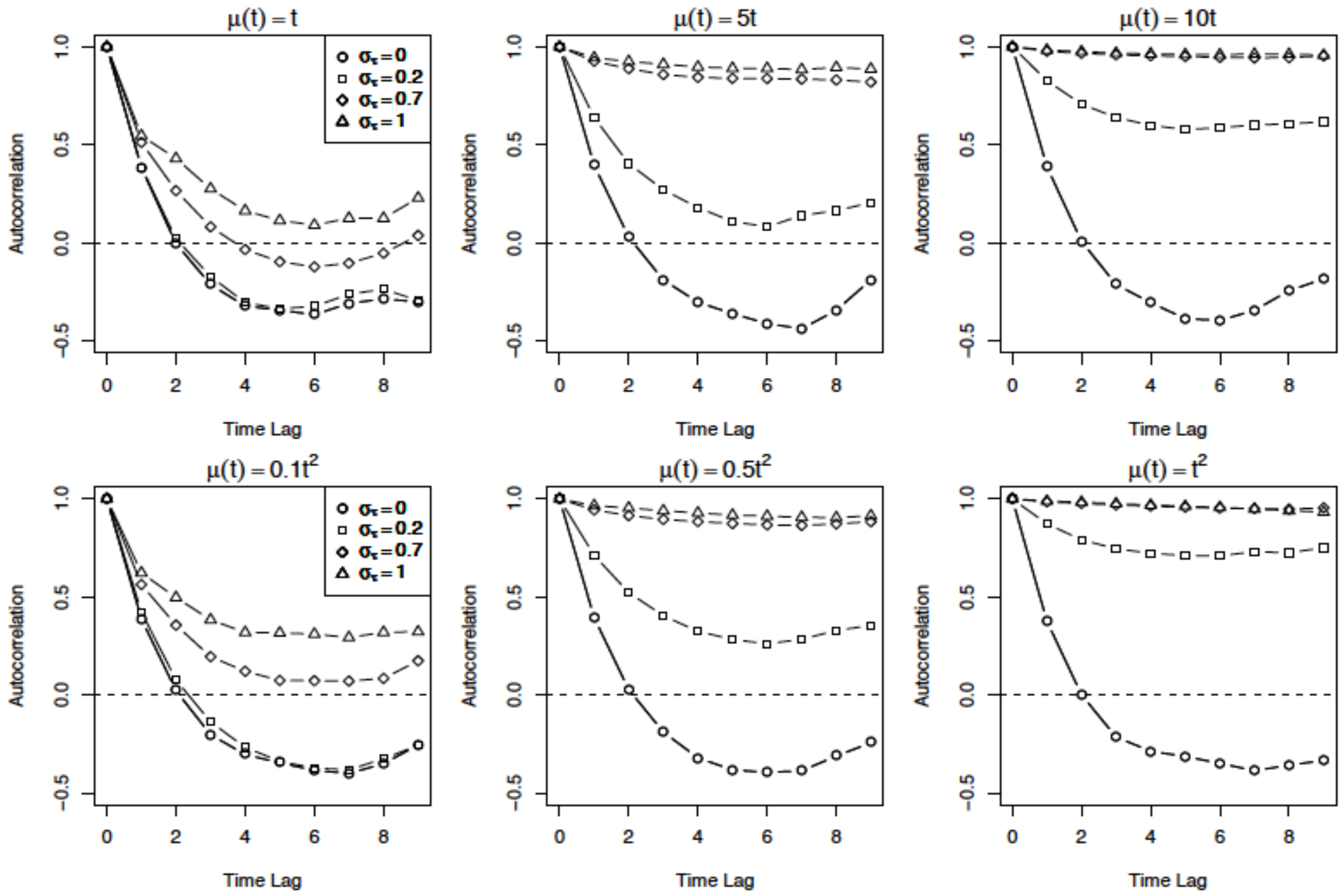
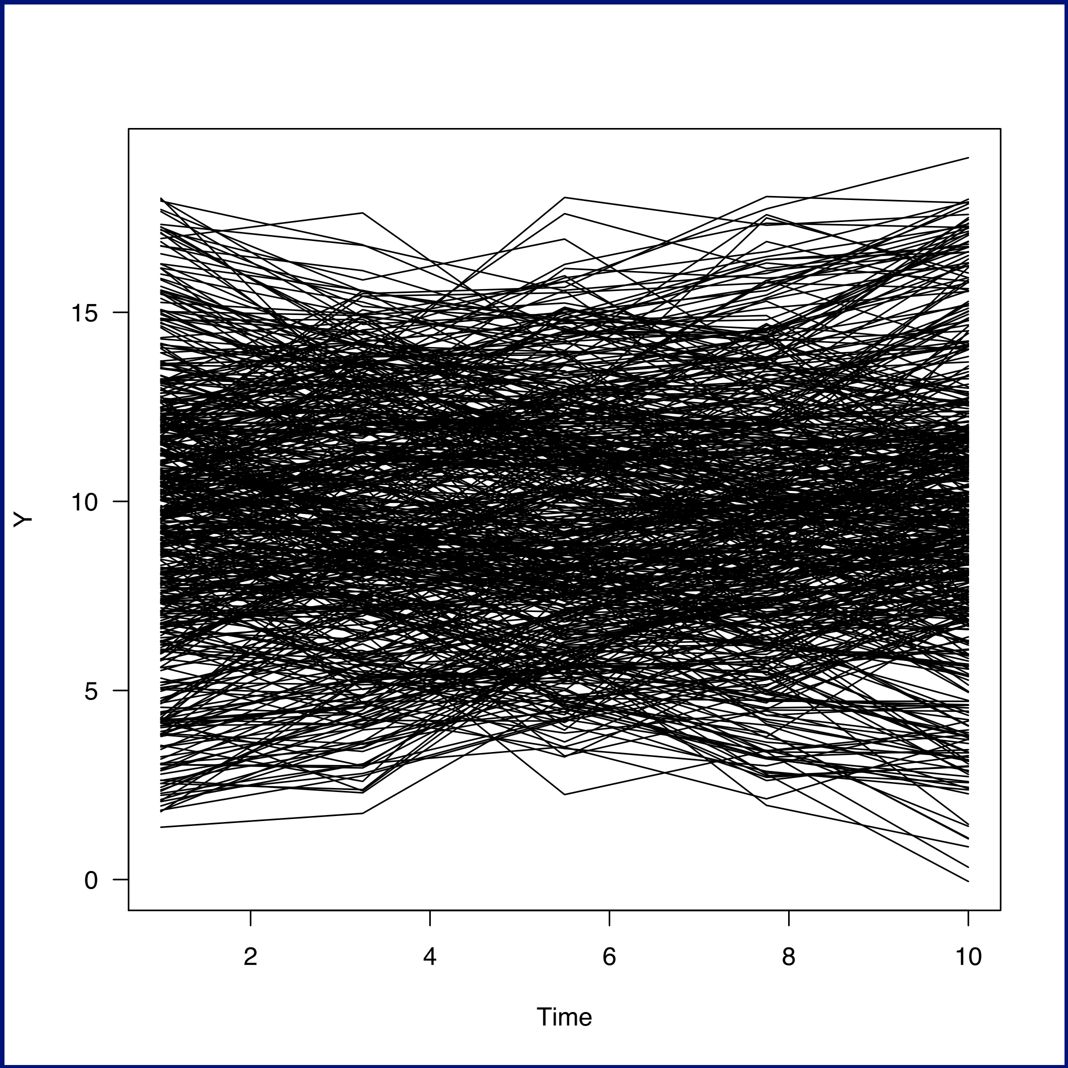


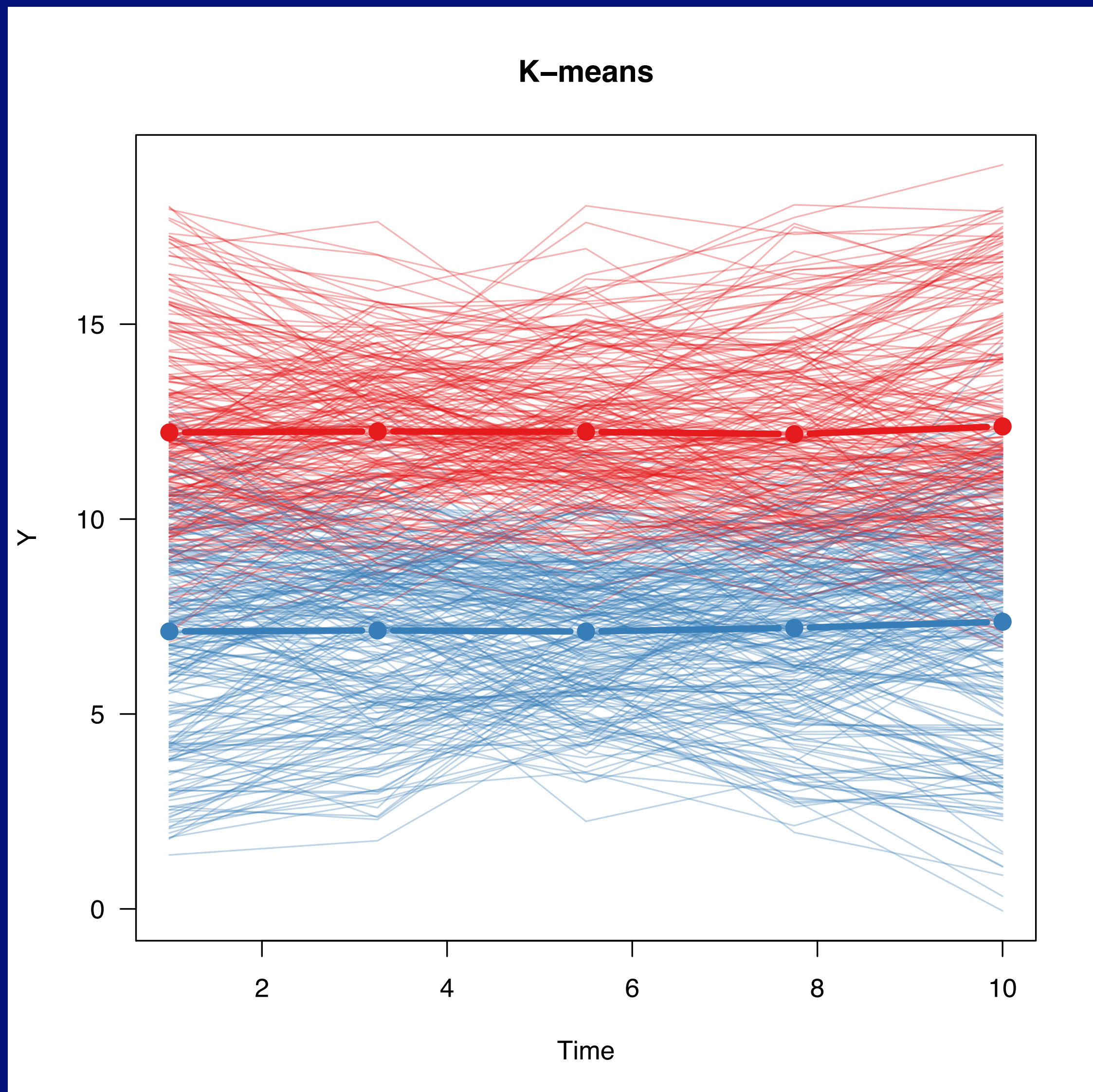
Figure 1: Estimated autocorrelation functions of the deviations from the mean from data generated with an exponential correlation error structure and random observation times under different mean functions, $\mu(t)$, and standard deviations of the random time perturbations, σ_τ.

Simulated Data



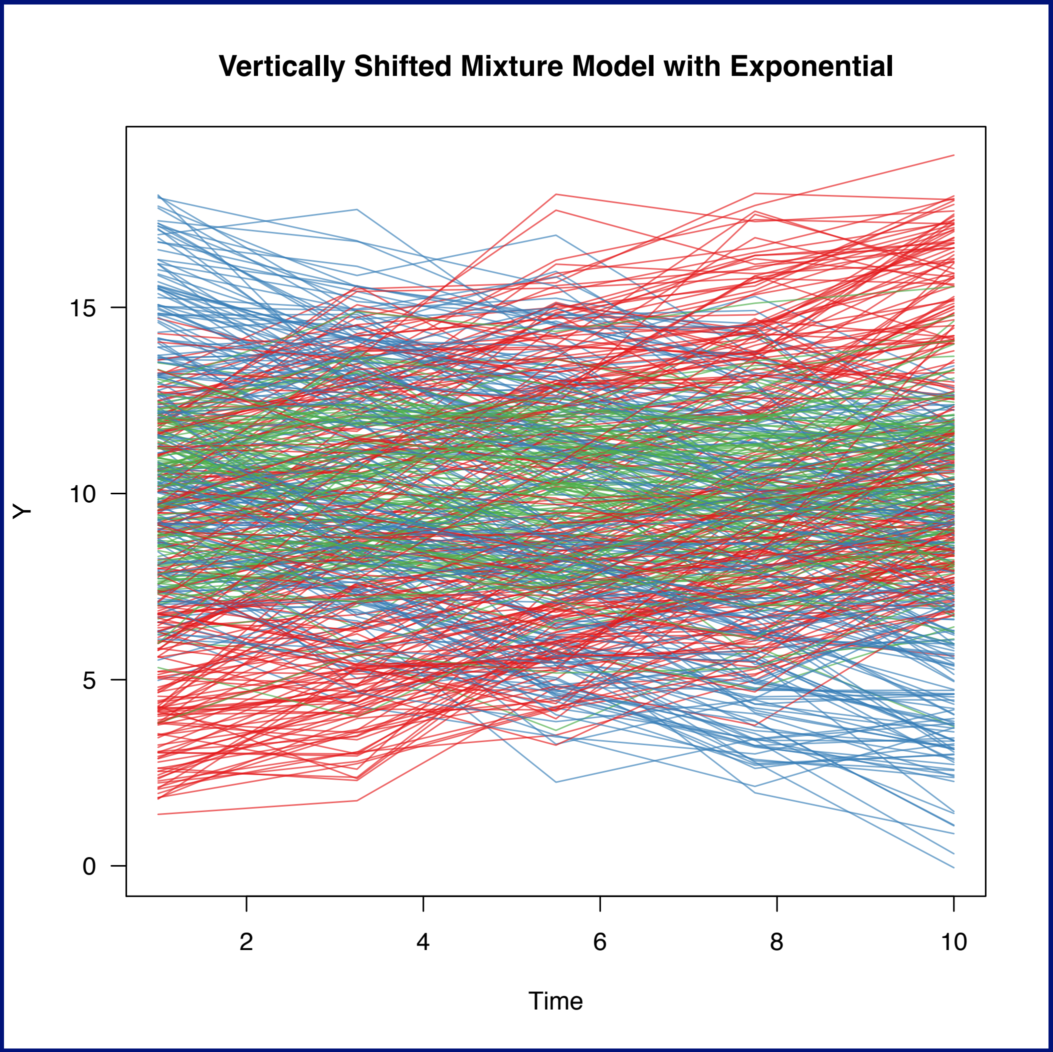
How could we group these individuals?

Application to Simulated Data

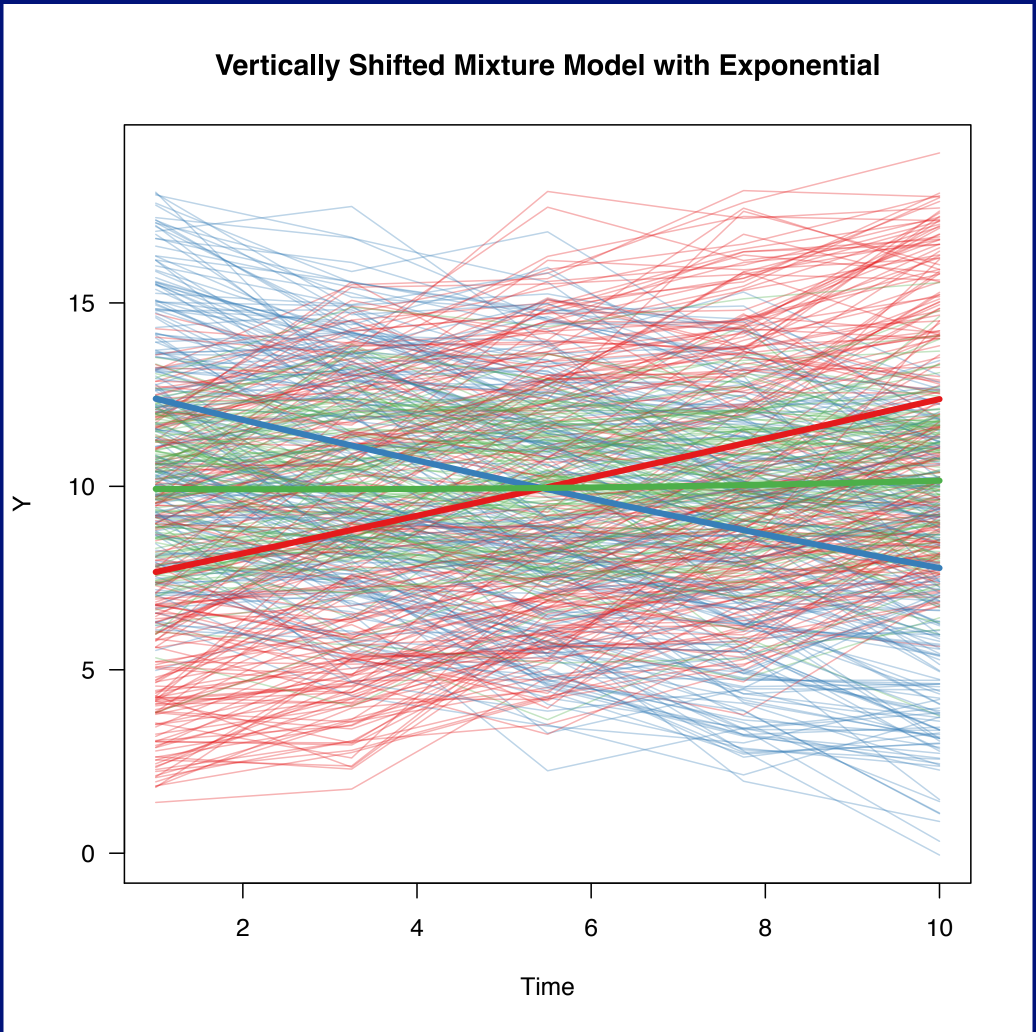


How would you describe—interpret—the group trajectories?

Vertical Shifting Applied to Simulated Data



Vertical Shifting Applied to Simulated Data



500 Simulations

$$\mu_1(t) = -1 - t$$

negative slope, low level

$$\mu_2(t) = 11 - t$$

negative slope, high level

$$\mu_3(t) = 0$$

zero slope middle level

$$\mu_4(t) = -11 + t$$

positive slope, low level

$$\mu_5(t) = 1 + t$$

positive slope, high level

Mean functions evaluated at five equidistant points that span [1,10}
Including ends of the interval

500 Simulations

$$\mu_1(t) = -1 - t$$ negative slope, low level

$$\mu_2(t) = 11 - t$$ negative slope, high level

$$\mu_3(t) = 0$$ zero slope middle level

$$\mu_4(t) = -11 + t$$ positive slope, low level

$$\mu_5(t) = 1 + t$$ positive slope, high level

Two components to noise: random individual level perturbation $N(0, \sigma_\lambda^2)$

random measurement error across times $N(0, \sigma_\epsilon^2)$

(exchangeable correlation)

500 Simulations

σ_ϵ	σ_λ	$K=2$	$K=3$	$K=4$	$K=5$	MR
Independent Gaussian Mixture						
0.50	2.00	0.00	0.00	10.00	490.00	0.41
2.00	2.00	0.00	0.00	24.00	476.00	0.39
0.50	3.00	0.00	0.00	1.00	499.00	0.45
2.00	3.00	0.00	0.00	5.00	495.00	0.45
K-means on Difference Quotients						
0.50	2.00	0.00	500.00	0.00	0.00	0.00
2.00	2.00	483.00	17.00	0.00	0.00	0.38
0.50	3.00	0.00	500.00	0.00	0.00	0.00
2.00	3.00	483.00	17.00	0.00	0.00	0.38
Correlation-based PAM						
0.50	2.00	403.00	0.00	0.00	97.00	0.25
2.00	2.00	500.00	0.00	0.00	0.00	0.27
0.50	3.00	403.00	0.00	0.00	97.00	0.25
2.00	3.00	500.00	0.00	0.00	0.00	0.27
Vertically Shifted Independent Mixture						
0.50	2.00	0.00	499.00	0.00	1.00	0.00
2.00	2.00	0.00	498.00	2.00	0.00	0.05
0.50	3.00	0.00	499.00	0.00	1.00	0.00
2.00	3.00	0.00	498.00	2.00	0.00	0.05
Vertically Shifted Exponential Mixture						
0.50	2.00	0.00	500.00	0.00	0.00	0.00
2.00	2.00	0.00	499.00	1.00	0.00	0.05
0.50	3.00	0.00	500.00	0.00	0.00	0.00
2.00	3.00	0.00	499.00	1.00	0.00	0.05

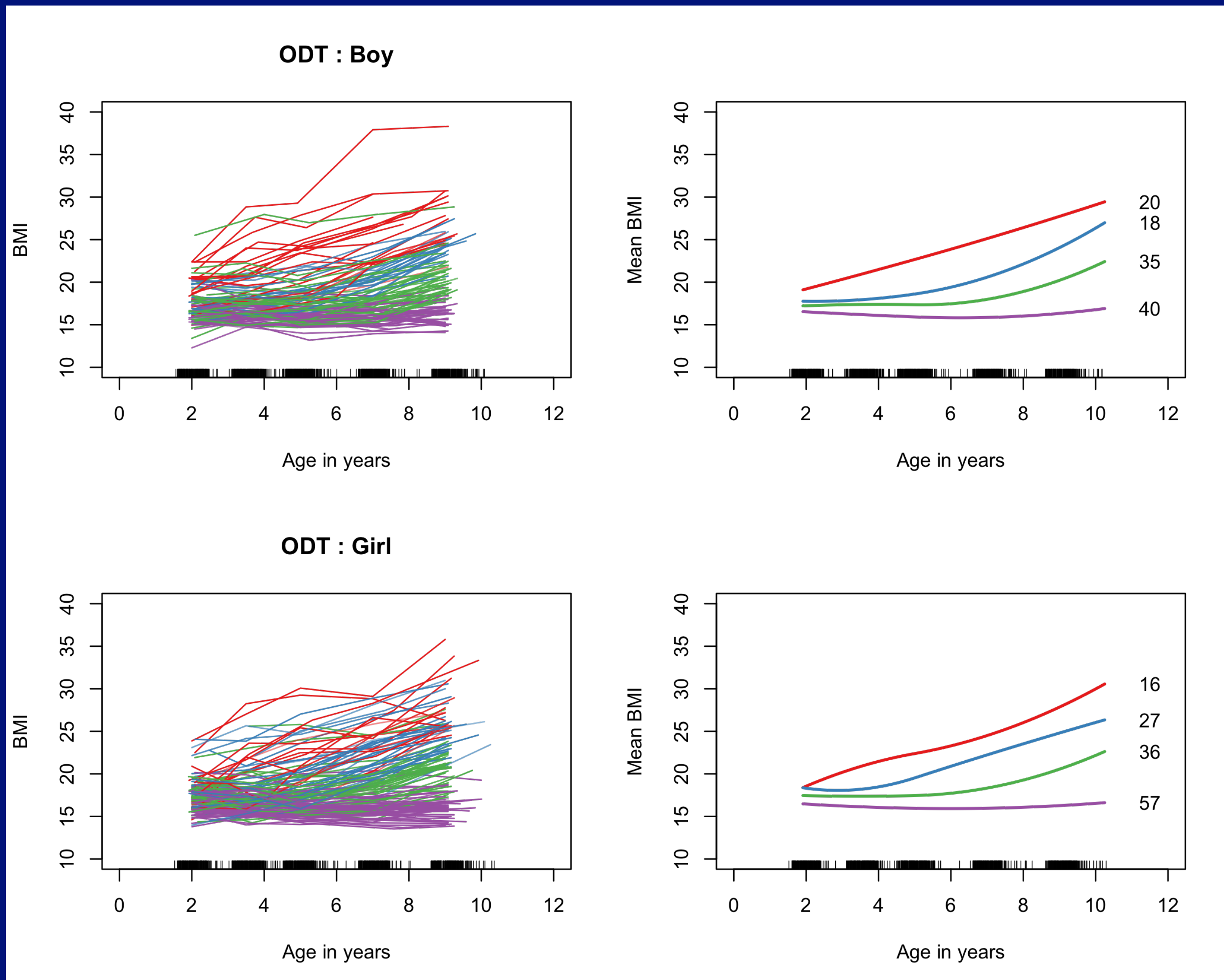
Table 1: Frequency table of the number of groups chosen and average misclassification rate (MR) ($K = 3$) for 500 replications of clustering methods applied to data generated under different values for σ_ϵ and σ_λ.

Vertical Shifting with CHAMACOS

- Two-part models
  - First, use standard regression models to relate baseline predictors to BMI
  - Then, use vertically shifted shape clustering with (same or different baseline predictors for shape groups)

- For BMI in the CHAMACOS data
  - Works with irregularly sampled data
  - Includes a way to estimate the relationship between baseline risk factors and group membership
  - Groups individuals according to the outcome pattern over time ignoring the level

Vertical Shifting with CHAMACOS



Vertical Shifting with CHAMACOS

	Estimate	95% CI	$P > \|z\|$
Red v. Purple : bmicat_preOverweight	2.494	(0.506, 12.294)	0.261
Red v. Purple : bmicat_preObese	34.068	(4.945, 234.699)	0
Red v. Purple : lng_lod2_ODT_pg	1.743	(0.343, 8.857)	0.503
Red v. Purple : yrsusa	0.97	(0.887, 1.062)	0.514
Blue v. Purple : bmicat_preOverweight	2.4	(0.408, 14.129)	0.333
Blue v. Purple : bmicat_preObese	8.636	(0.97, 76.902)	0.053
Blue v. Purple : lng_lod2_ODT_pg	6.679	(1.067, 41.812)	0.042
Blue v. Purple : yrsusa	1.048	(0.955, 1.15)	0.324
Green v. Purple : bmicat_preOverweight	1.45	(0.428, 4.917)	0.551
Green v. Purple : bmicat_preObese	15.561	(2.306, 105.031)	0.005
Green v. Purple : lng_lod2_ODT_pg	4.028	(0.598, 27.113)	0.152
Green v. Purple : yrsusa	0.953	(0.866, 1.048)	0.322

Table 1: Estimated odds ratios for Vertically Shifted Model with ODT for Boy

	Estimate	95% CI	$P > \|z\|$
Red v. Purple : bmicat_preOverweight	6.258	(0.193, 203.281)	0.302
Red v. Purple : bmicat_preObese	64.22	(2.1, 1963.475)	0.017
Red v. Purple : lng_lod2_ODT_pg	0.358	(0.07, 1.846)	0.22
Red v. Purple : yrsusa	0.955	(0.785, 1.162)	0.646
Blue v. Purple : bmicat_preOverweight	1.785	(0.464, 6.86)	0.399
Blue v. Purple : bmicat_preObese	7.121	(1.069, 47.437)	0.042
Blue v. Purple : lng_lod2_ODT_pg	0.846	(0.276, 2.599)	0.771
Blue v. Purple : yrsusa	1.011	(0.922, 1.109)	0.815
Green v. Purple : bmicat_preOverweight	0.452	(0.15, 1.36)	0.158
Green v. Purple : bmicat_preObese	4.335	(0.998, 18.837)	0.05
Green v. Purple : lng_lod2_ODT_pg	0.637	(0.329, 1.233)	0.181
Green v. Purple : yrsusa	0.958	(0.885, 1.038)	0.295

Table 2: Estimated odds ratios for Vertically Shifted Model with ODT for Girl

Further Thoughts

- **Time-dependent covariates**