Developing a Shiny App for Model-Based Clustering and Classification

Harry O'Callaghan
Alan Jeffares

Contents

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

Contents

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

Contents

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

Contents

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

Contents

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

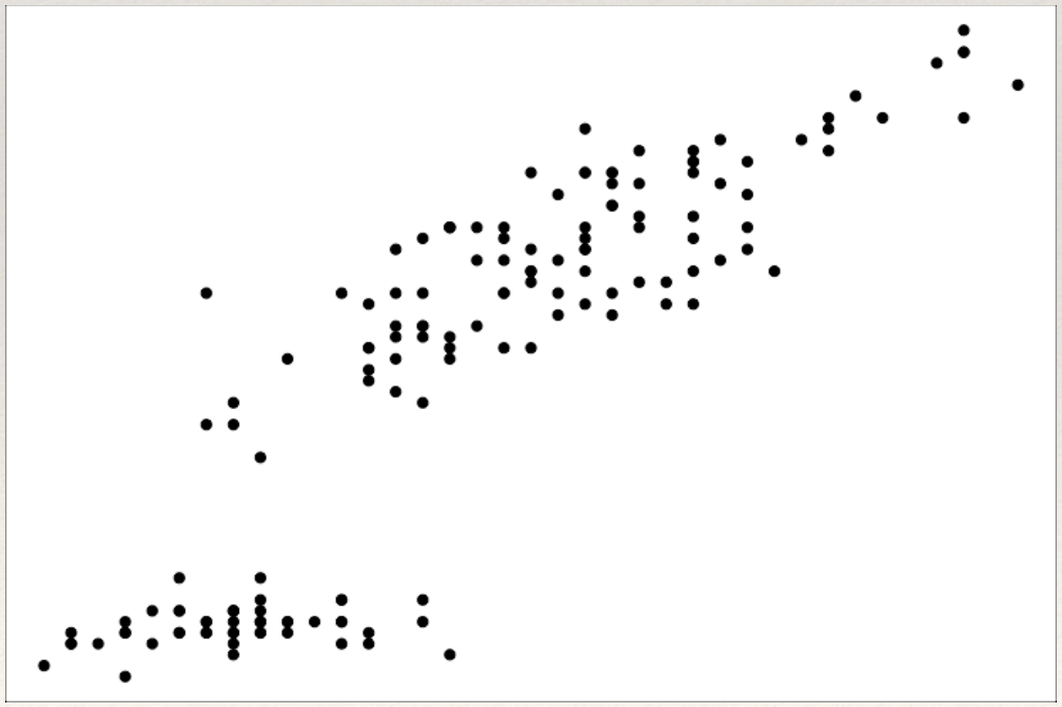
Contents

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

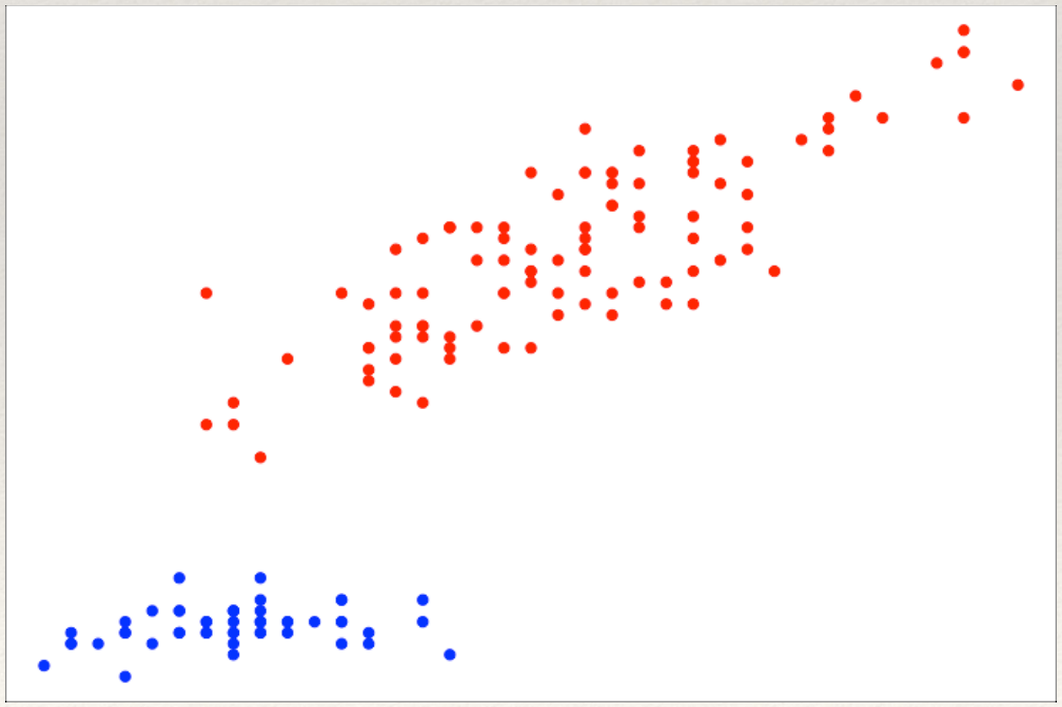
Model-Based Clustering

Clustering - An unsupervised classification technique in which we attempt to divide a set of data objects into clusters



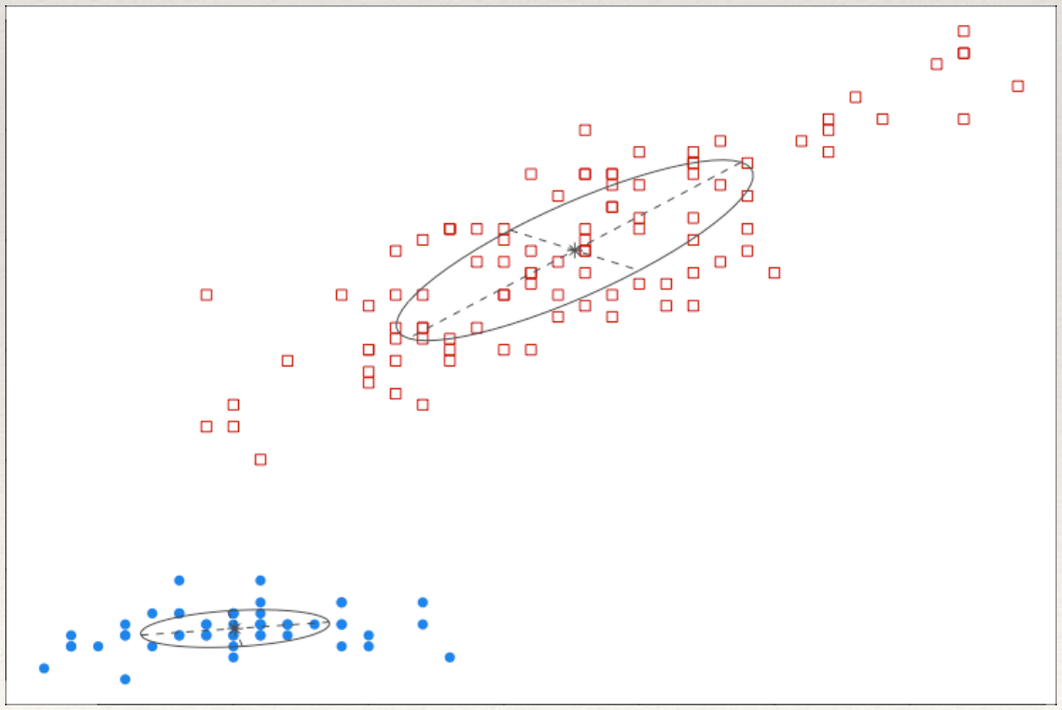
Model-Based Clustering

Clustering - An unsupervised classification technique in which we attempt to divide a set of data objects into clusters



Model-Based Clustering

Model-Based Clustering - A parametric approach in which we assume the data as coming from a distribution that is a mixture of two or more clusters



Model-Based Clustering

We assume the data is generated from a mixture of multivariate normal distributions:

$$f_k(\mathbf{x}_i|\mu_k, \Sigma_k) = \frac{exp\{-\frac{1}{2}(\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1}(\mathbf{x}_i - \mu_k)\}}{(2\pi)^{\frac{p}{2}}|\Sigma_k|^{\frac{1}{2}}}$$

And thus the mixture likelihood approach maximises:

$$L_M(\theta_1, ..., \theta_G; \tau_1, ..., \tau_G|\mathbf{x}) = \prod_{i=1}^{n}\sum_{k=1}^{G}\tau_k f_k(\mathbf{x}_i|\theta_k)$$

Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:

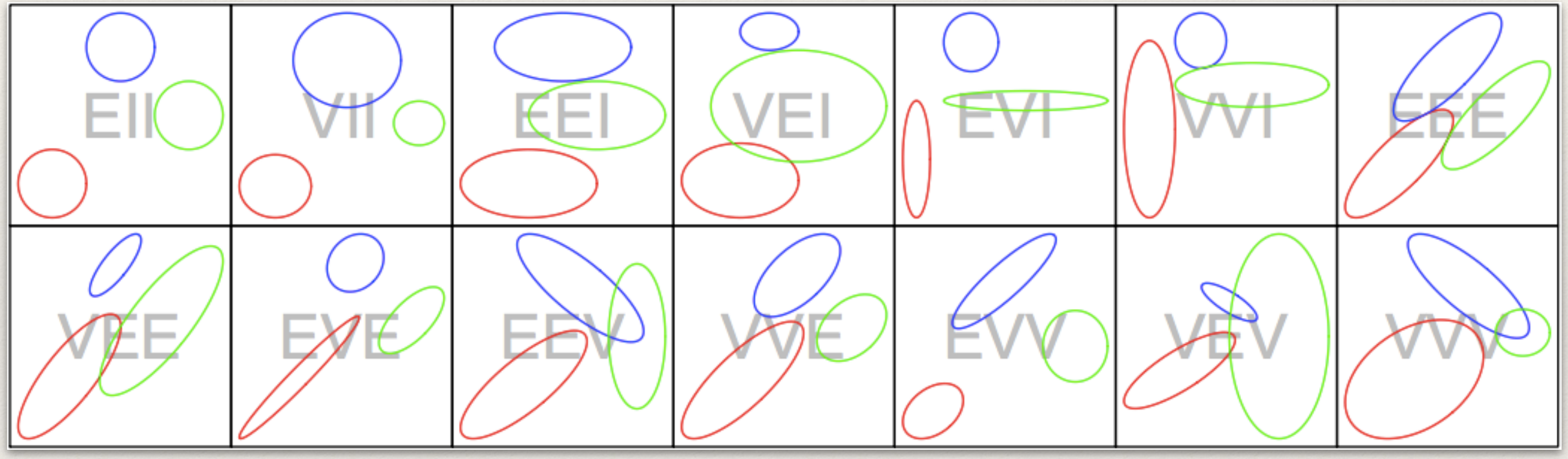
1. Determine the maximum number of clusters to consider and set parameterisations of the Gaussian model to consider

$$\Sigma_k = \lambda_k D_k A_k D_k^T$$

λ	A	D
Volume	Shape	Orientation

Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:



λ Volume

A Shape

D Orientation

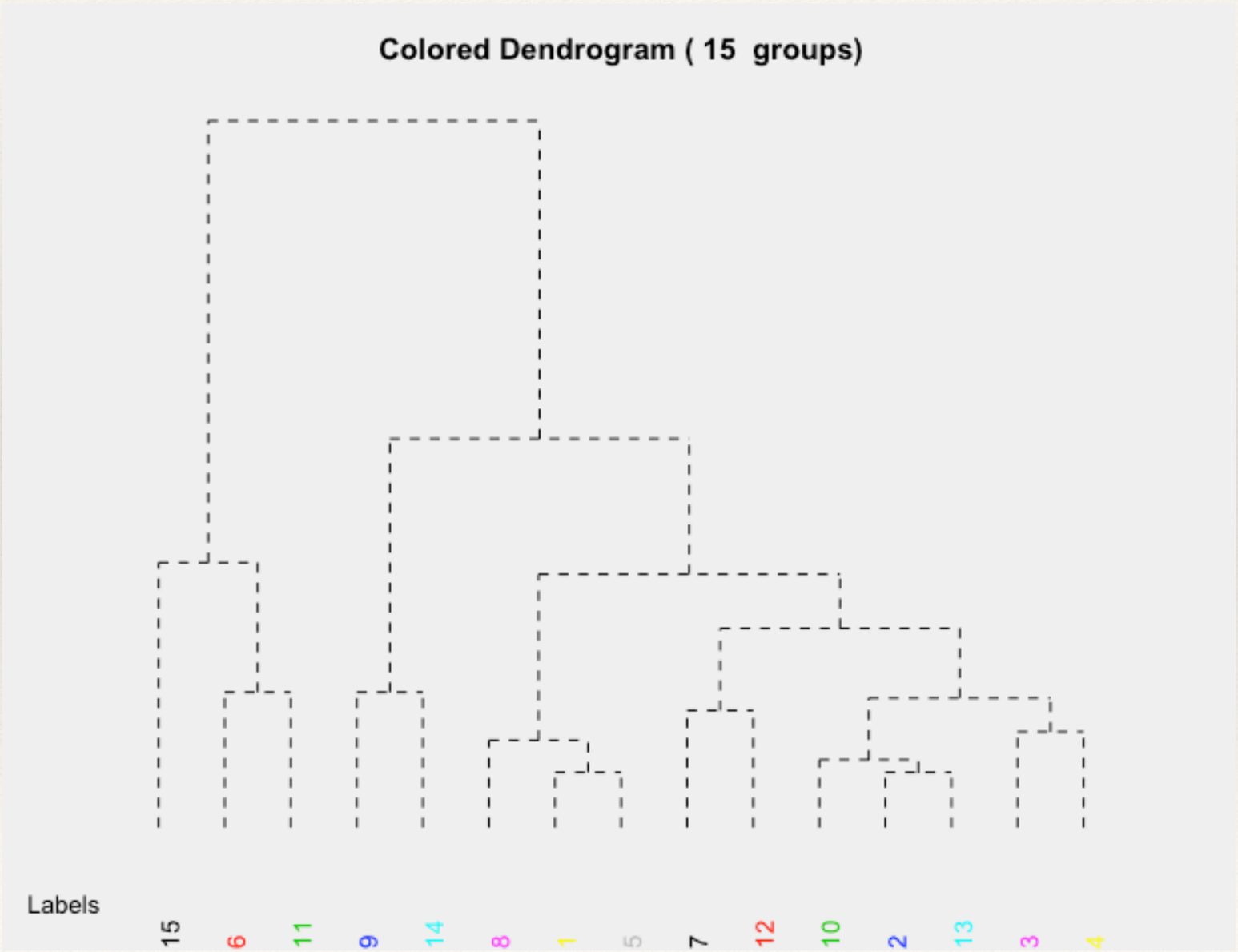
Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:

2. Do agglomerative hierarchical clustering and obtain the initial classification for each possible number of groups

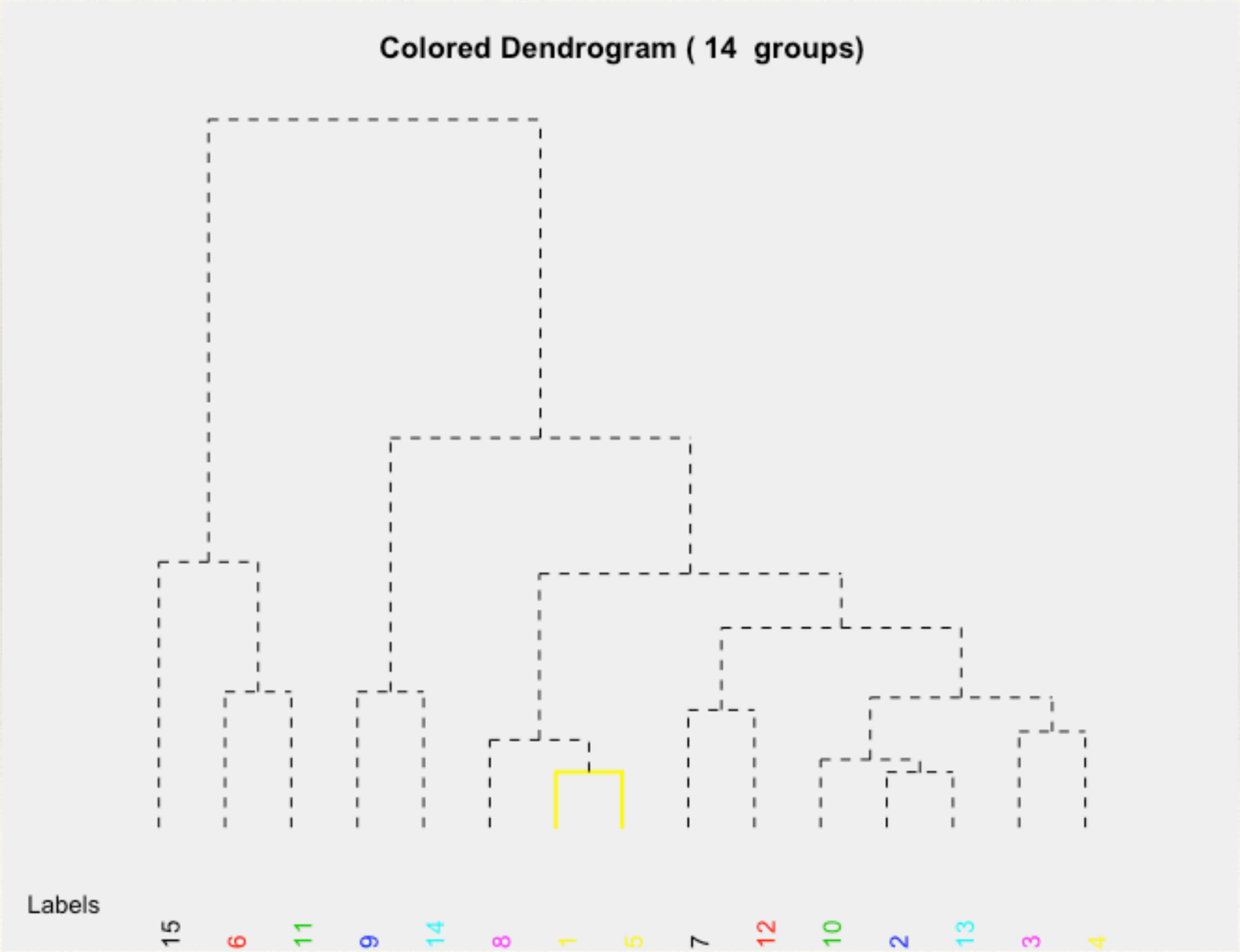
Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:



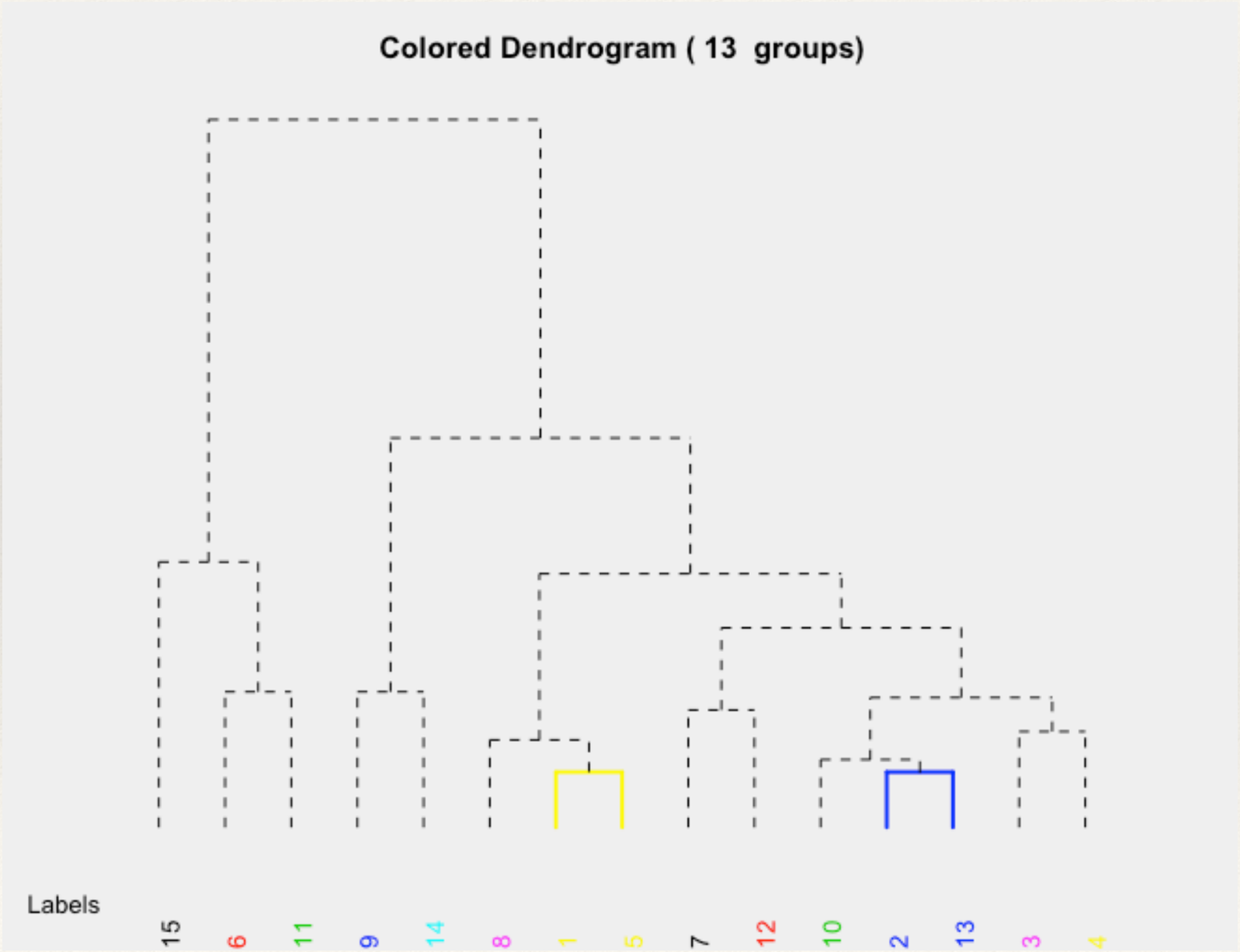
Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:



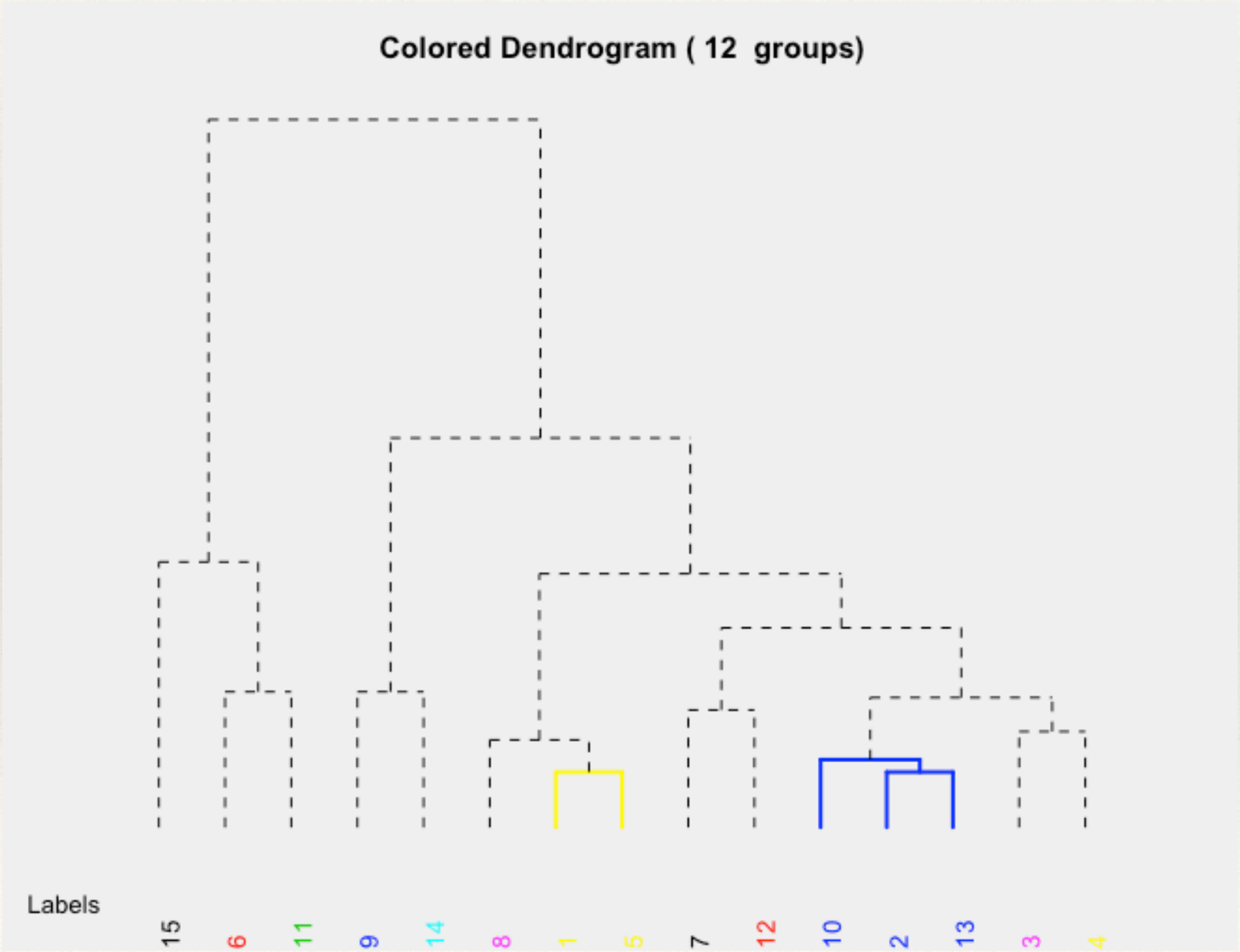
Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:



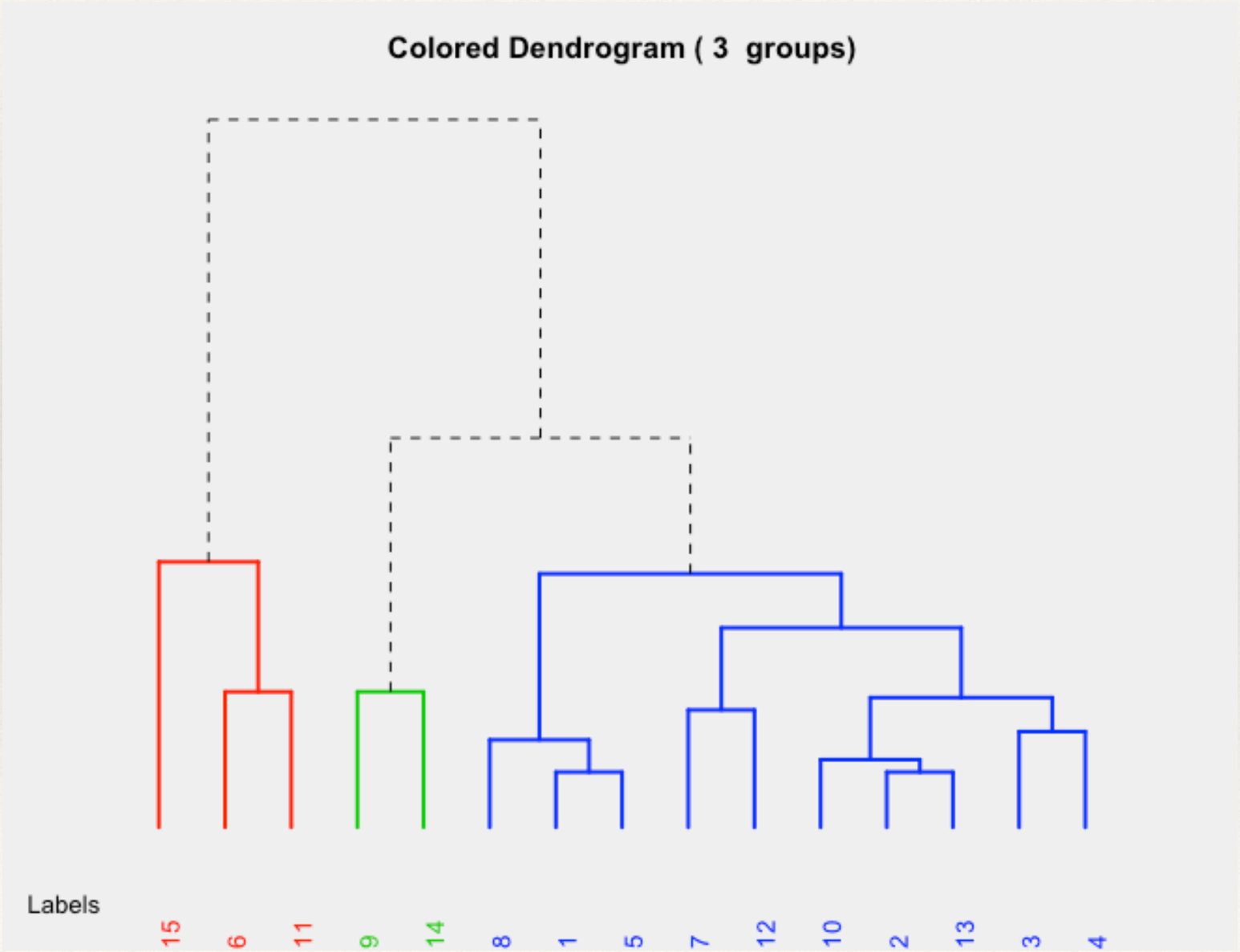
Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:



Model-Based Clustering

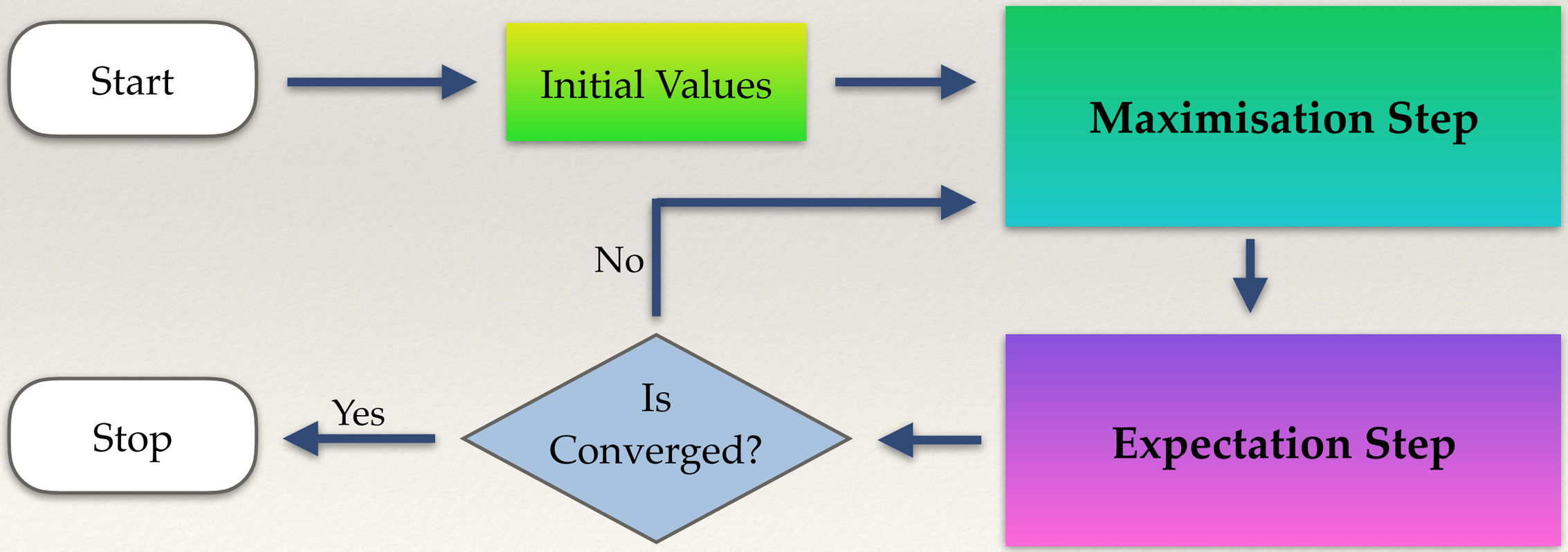
Model-based Strategy for Clustering in 5 Steps:



Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:

3. Apply EM algorithm for each parameterisation and each number of clusters



Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:

<u>M-Step:</u>

$$l(\theta_k, \tau_k, z_{ik}|\mathbf{x}) = \sum_{i=1}^{n}\sum_{k=1}^{G} z_{ik}[log(\tau_k f_k(\mathbf{x}_i|\theta_k))]$$

$$n_k \longleftarrow \sum_{i=1}^{n} \hat{z}_{ik} \qquad \hat{\mu}_k \longleftarrow \frac{\sum_{i=1}^{n} \hat{z}_{ik}\mathbf{x}_i}{n_k}$$

$$\hat{\tau}_k \longleftarrow \frac{n_k}{n} \qquad \hat{\Sigma}_k \longleftarrow \text{Depends on the model}$$

Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:

<u>E-Step:</u>

$$\hat{z}_{ik} \longleftarrow \frac{\hat{\tau}_k f_k(\mathbf{x}_i|\hat{\mu}_k, \hat{\Sigma}_k)}{\sum_{j=1}^{G} \hat{\tau}_j f_j(\mathbf{x}_i|\hat{\mu}_j, \hat{\Sigma}_j)}$$

Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:

4. Compute BIC values corresponding to each possible combination of parameterisation and number of clusters

$$BIC = 2l_M(x, \hat{\theta}) - m_M log(n)$$

Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:

5. Plot the BIC values for each model

Parameterisations

Number of components

	EII	**VII**	**EEI**	**VEI**	**…**
1	-1345	-1046	-821	-762	…
2	-1034	-866	-924	-844	…
3	-1042	-994	-694	-888	…
4	-986	-877	-909	-711	…
…	…	…	…	…	

Model-Based Clustering

Model-based Strategy for Clustering in 5 Steps:

5. Plot the BIC values for each model

Parameterisations

Number of components

	EII	**VII**	**EEI**	**VEI**	**…**
1	-1345	-1046	-821	-762	…
2	-1034	-866	-924	-844	…
3	-1042	-994	-694	-888	…
4	-986	-877	-909	-711	…
…	…	…	…	…	

Mclust

Mclust is an R package that implements model-based clustering and classification

```
library(mclust)
model <- Mclust(data, G, modelNames)
class <- MclustDA(data, class, G, modelNames)
```

R Shiny

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

R Shiny

Shiny is an R package that makes it easy to build interactive web applications straight from R

R Shiny

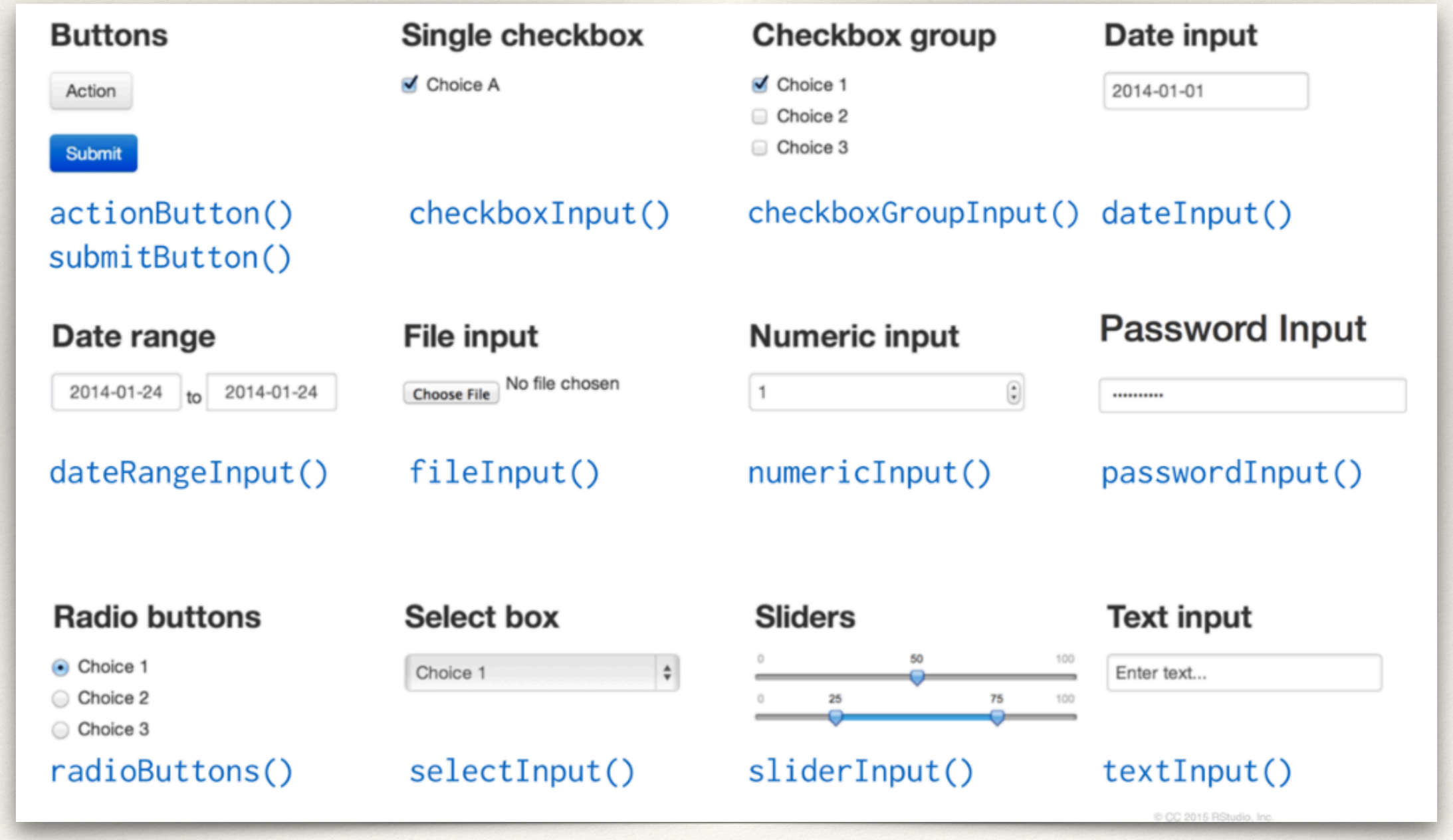
```
library(shiny)
ui <- fluidPage()

server <- function(input, output) {}

shinyApp(ui = ui, server = server)
```

R Shiny

Inputs



R Shiny

Outputs

Function	Inserts
`dataTableOutput()`	an interactive table
`htmlOutput()`	raw HTML
`imageOutput()`	image
`plotOutput()`	plot
`tableOutput()`	table
`textOutput()`	text
`uiOutput()`	a Shiny UI element
`verbatimTextOutput()`	text

R Shiny

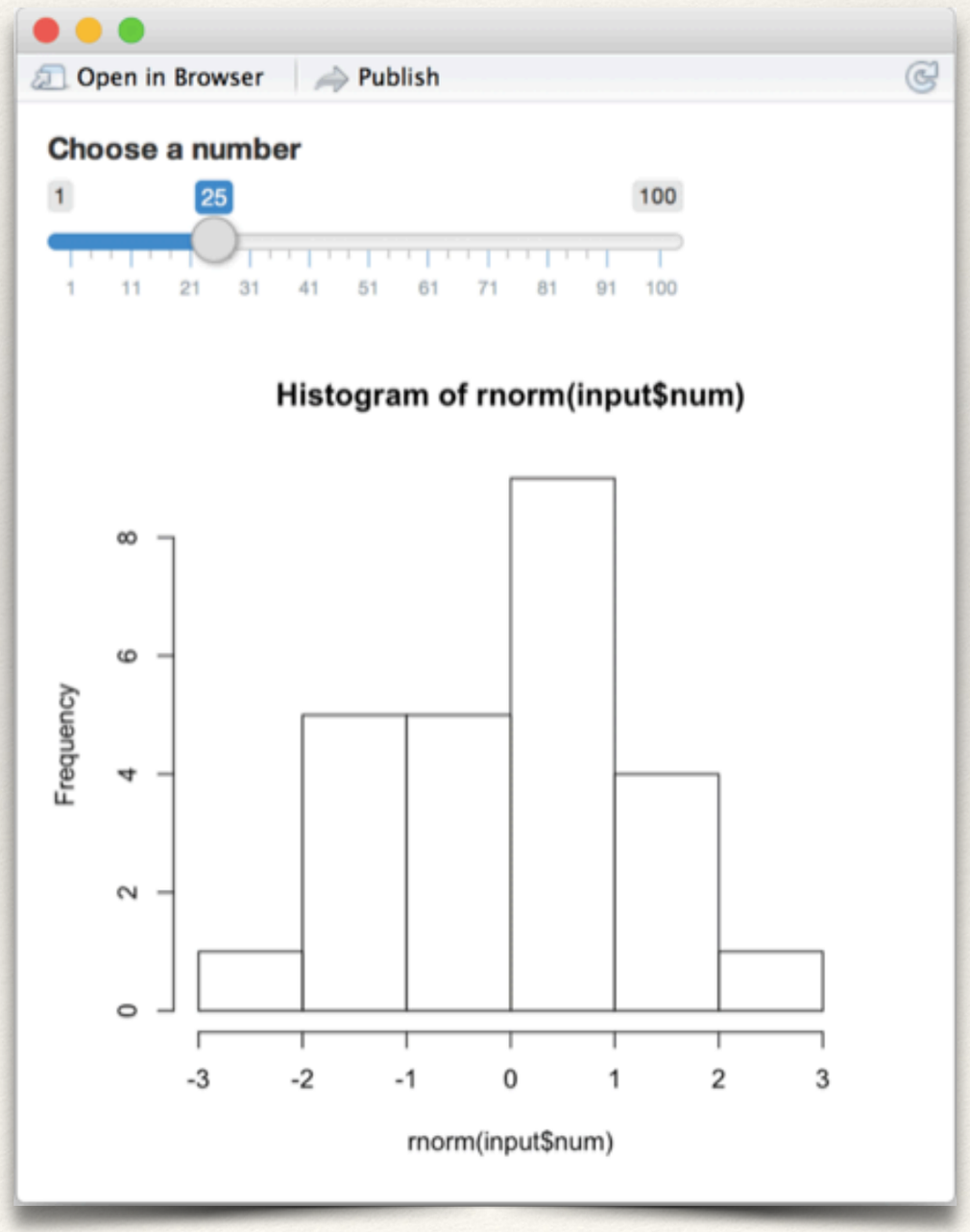
Example

```
library(shiny)

ui <- fluidPage(
  sliderInput(inputId = "num",
    label = "Choose a number",
    value = 25, min = 1, max = 100),
  plotOutput("hist")
)

server <- function(input, output) {
  output$hist <- renderPlot({
    hist(rnorm(input$num))
  })
}

shinyApp(ui = ui, server = server)
```

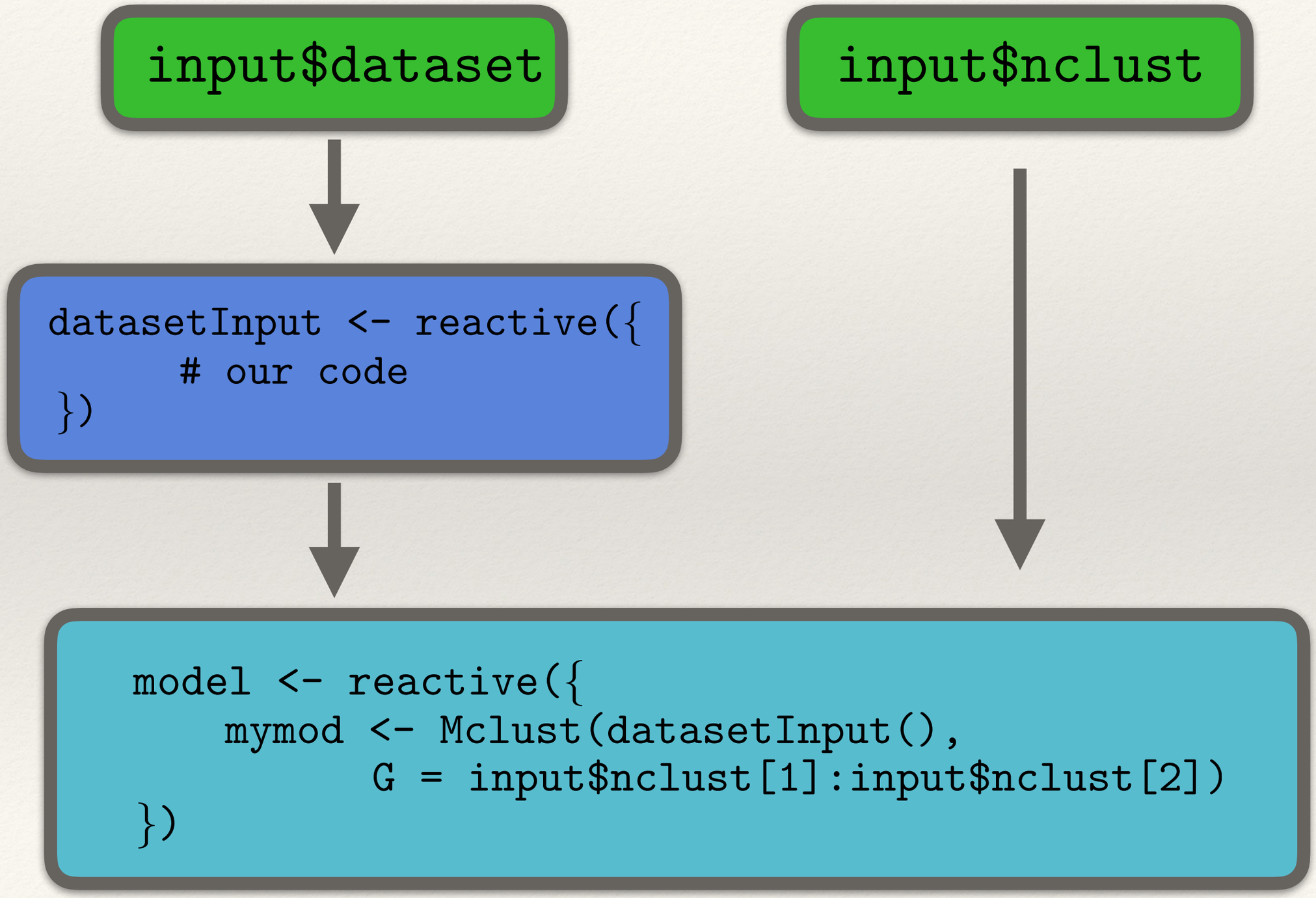


Application (Clustering)

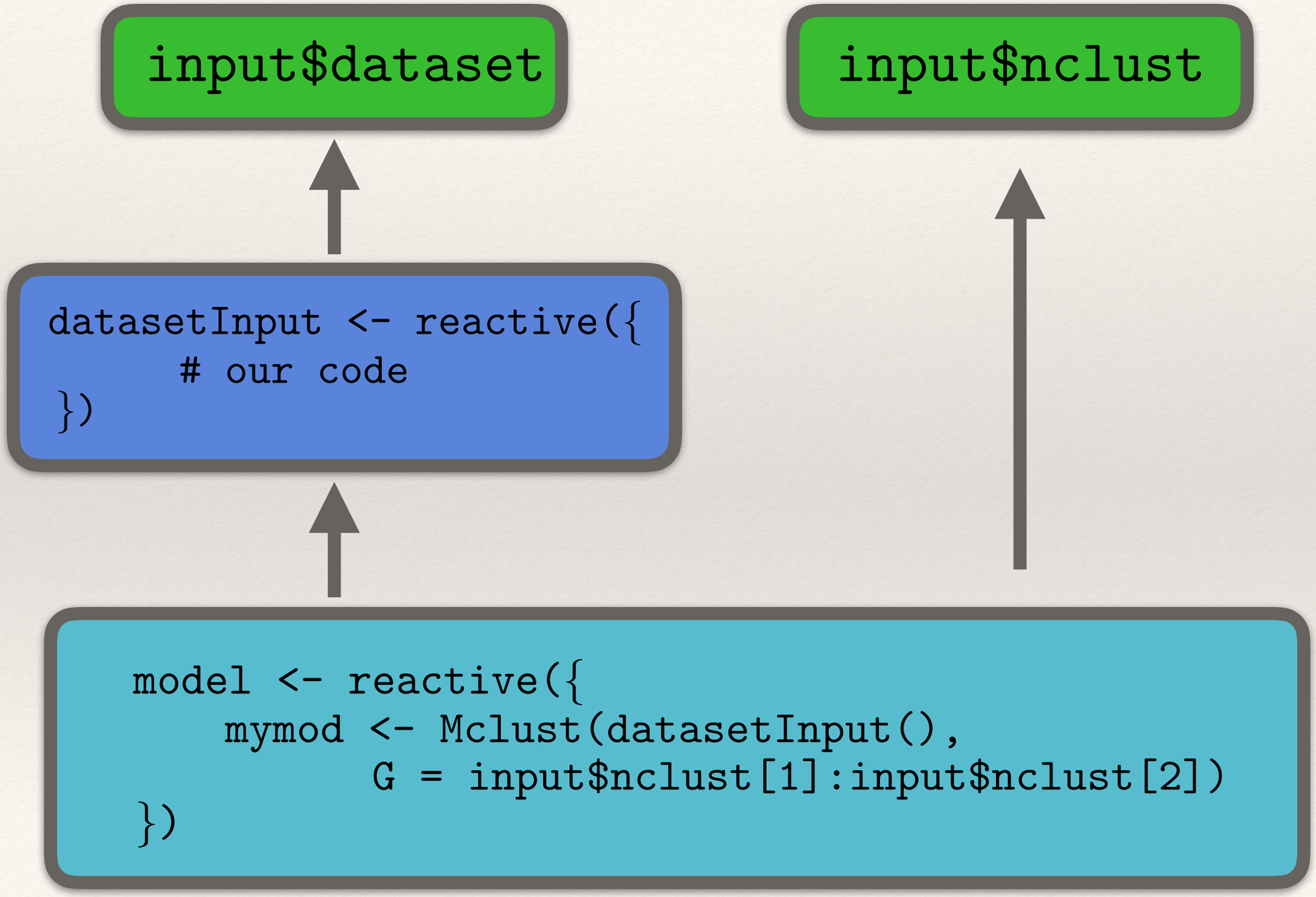
Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

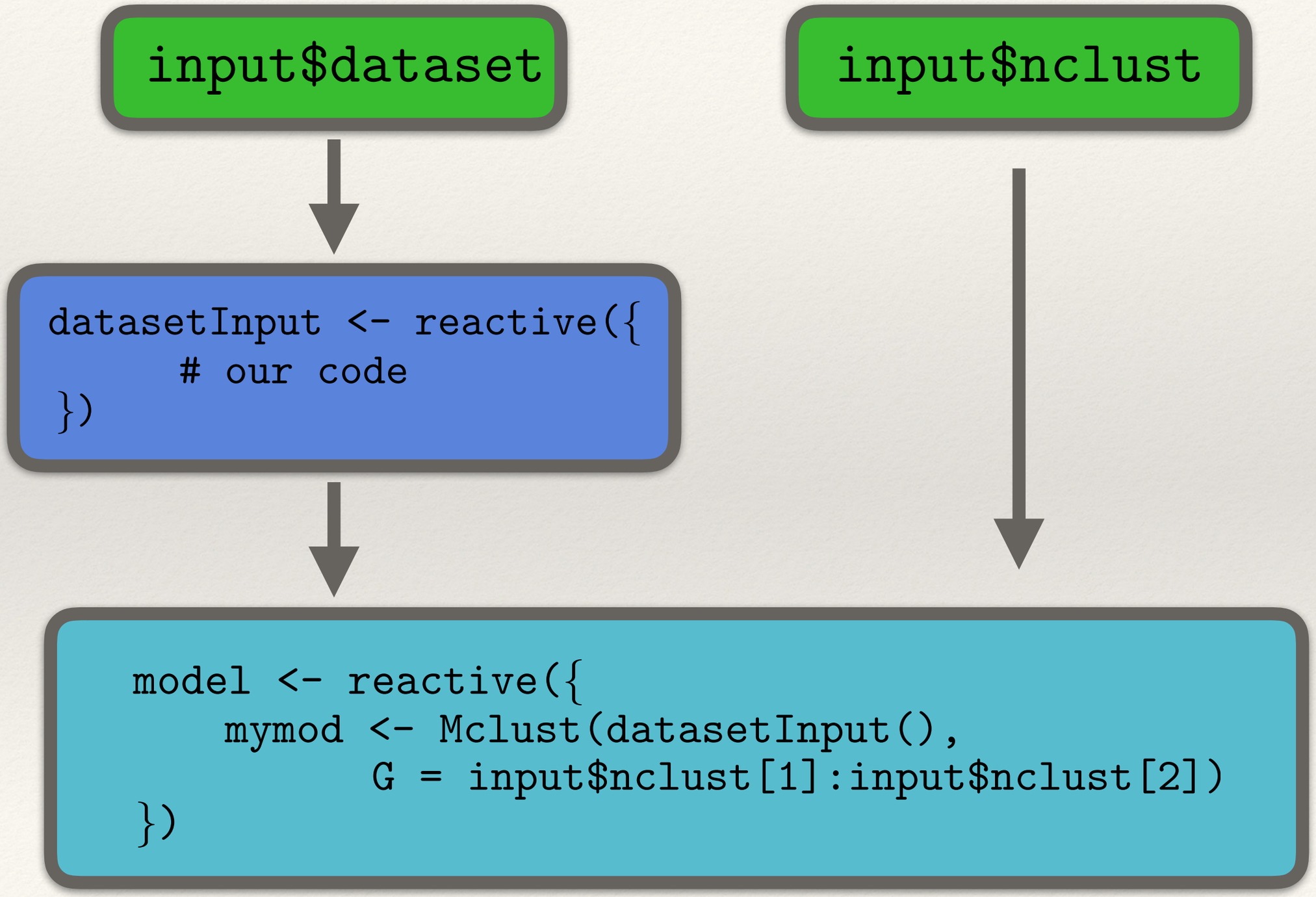
Application (Clustering)



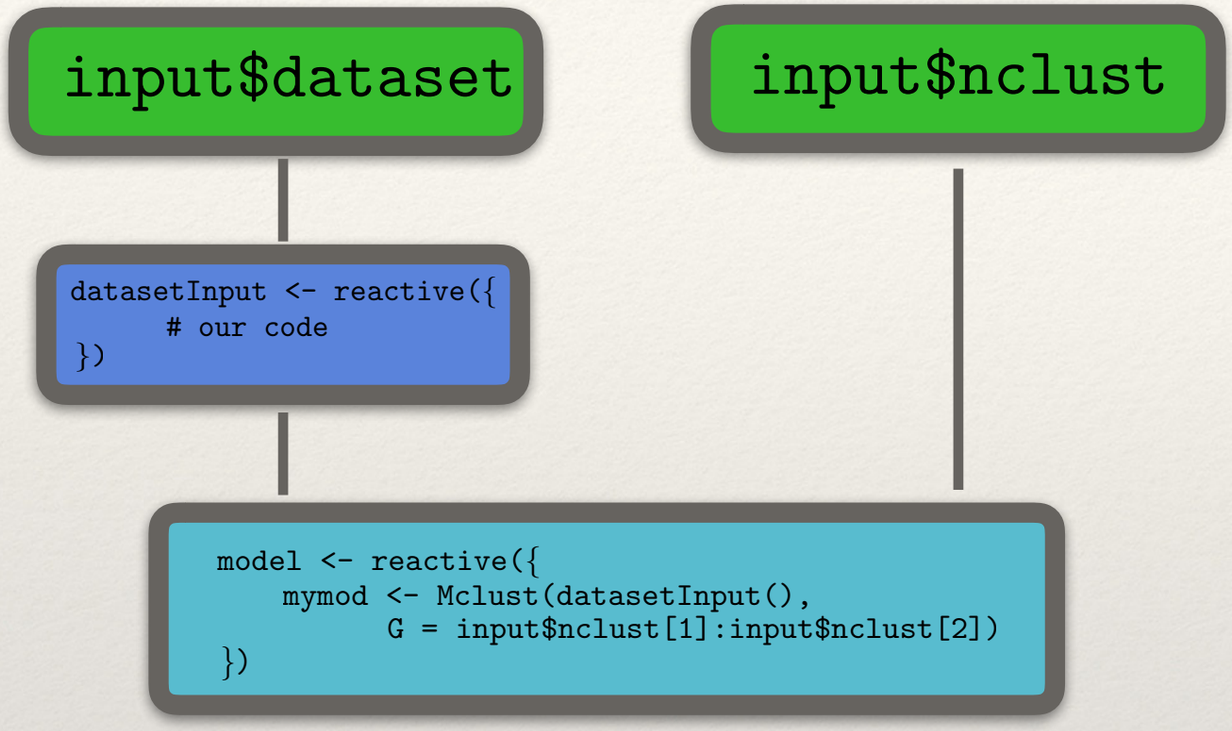
Application (Clustering)



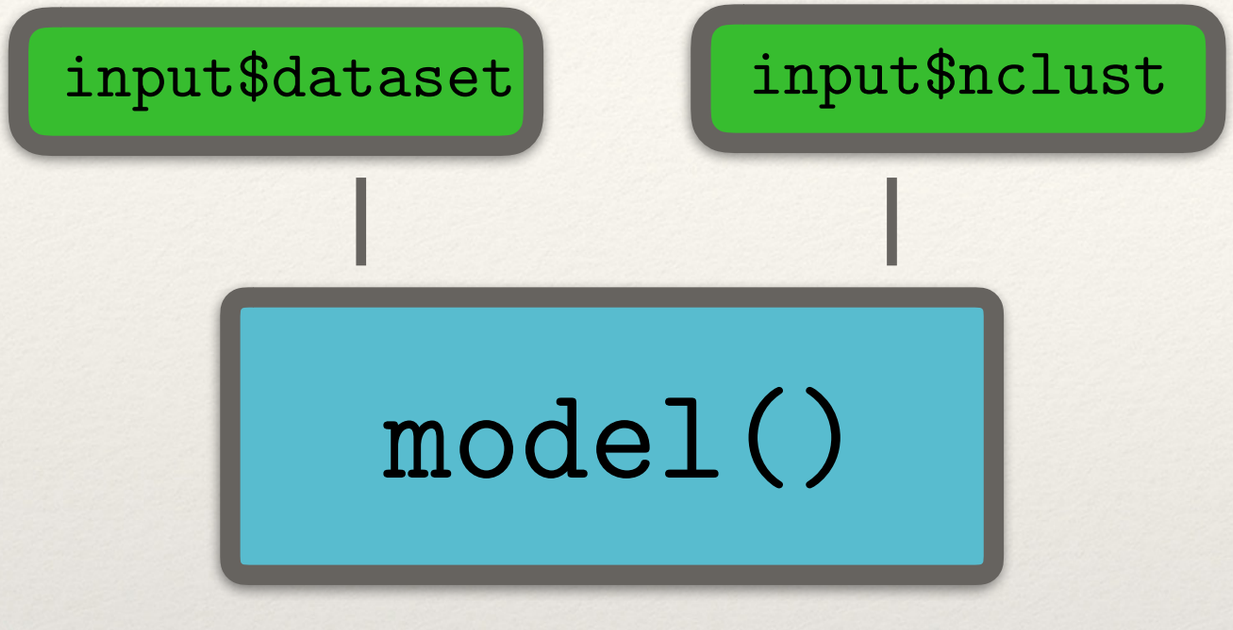
Application (Clustering)



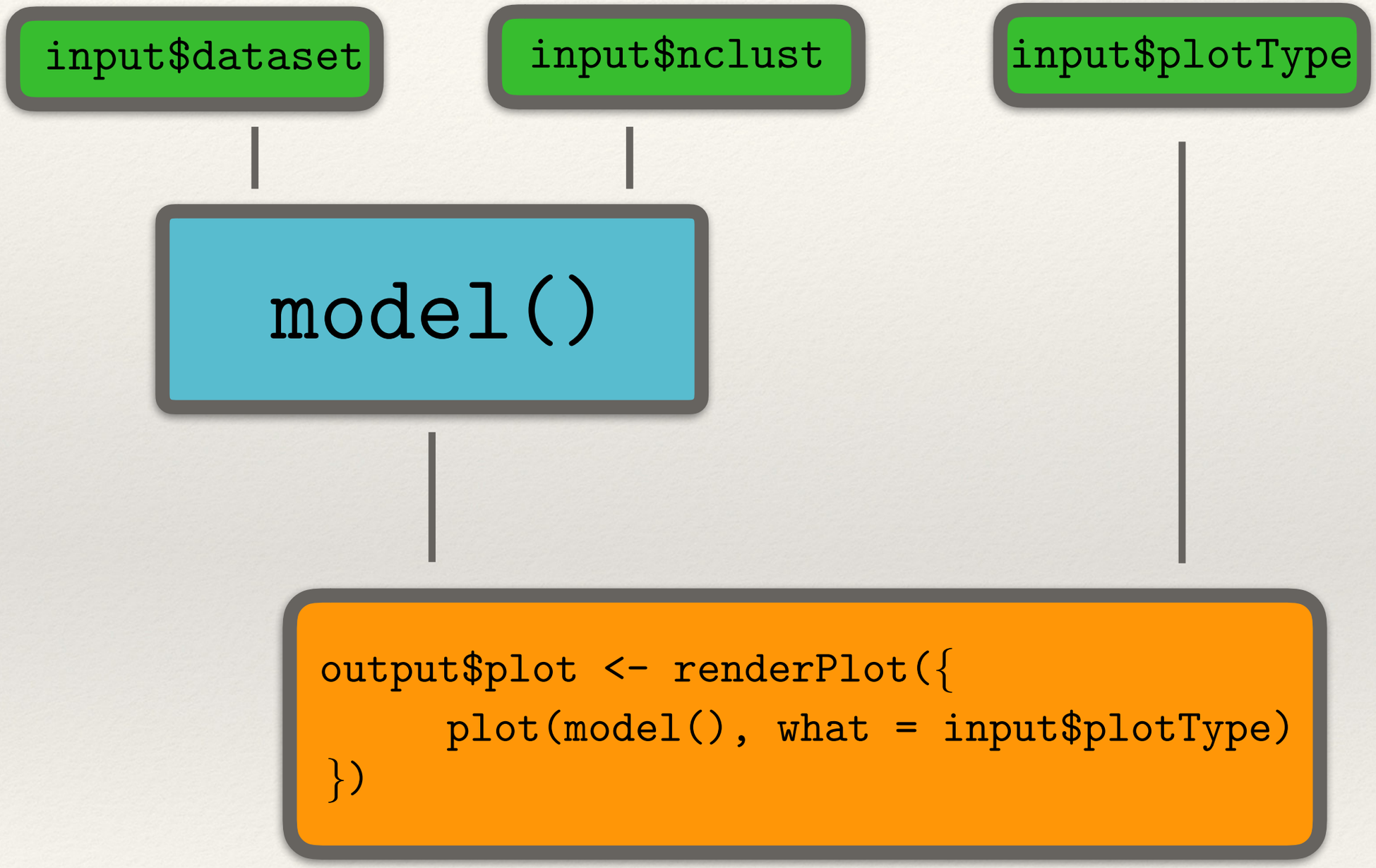
Application (Clustering)



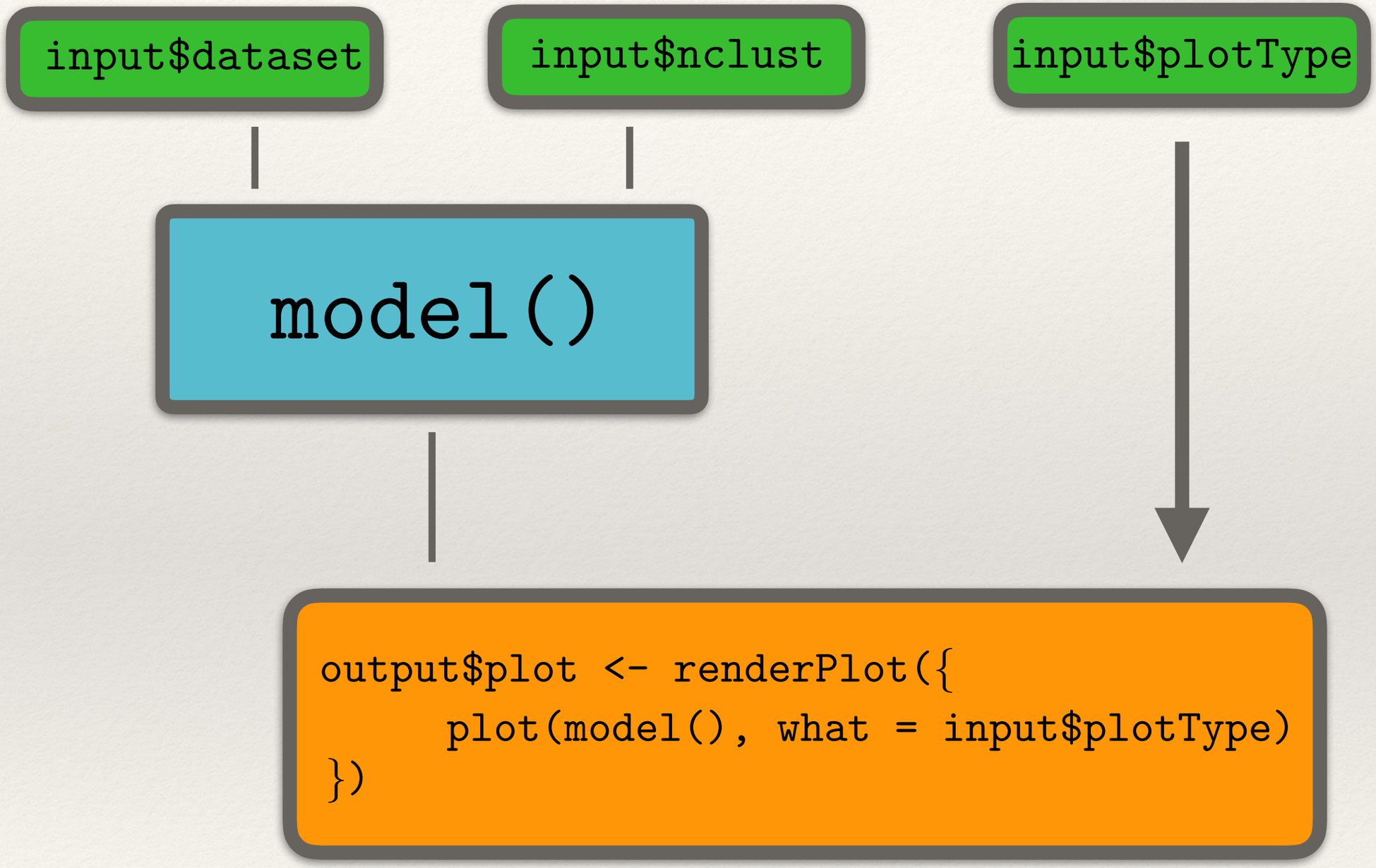
Application (Clustering)



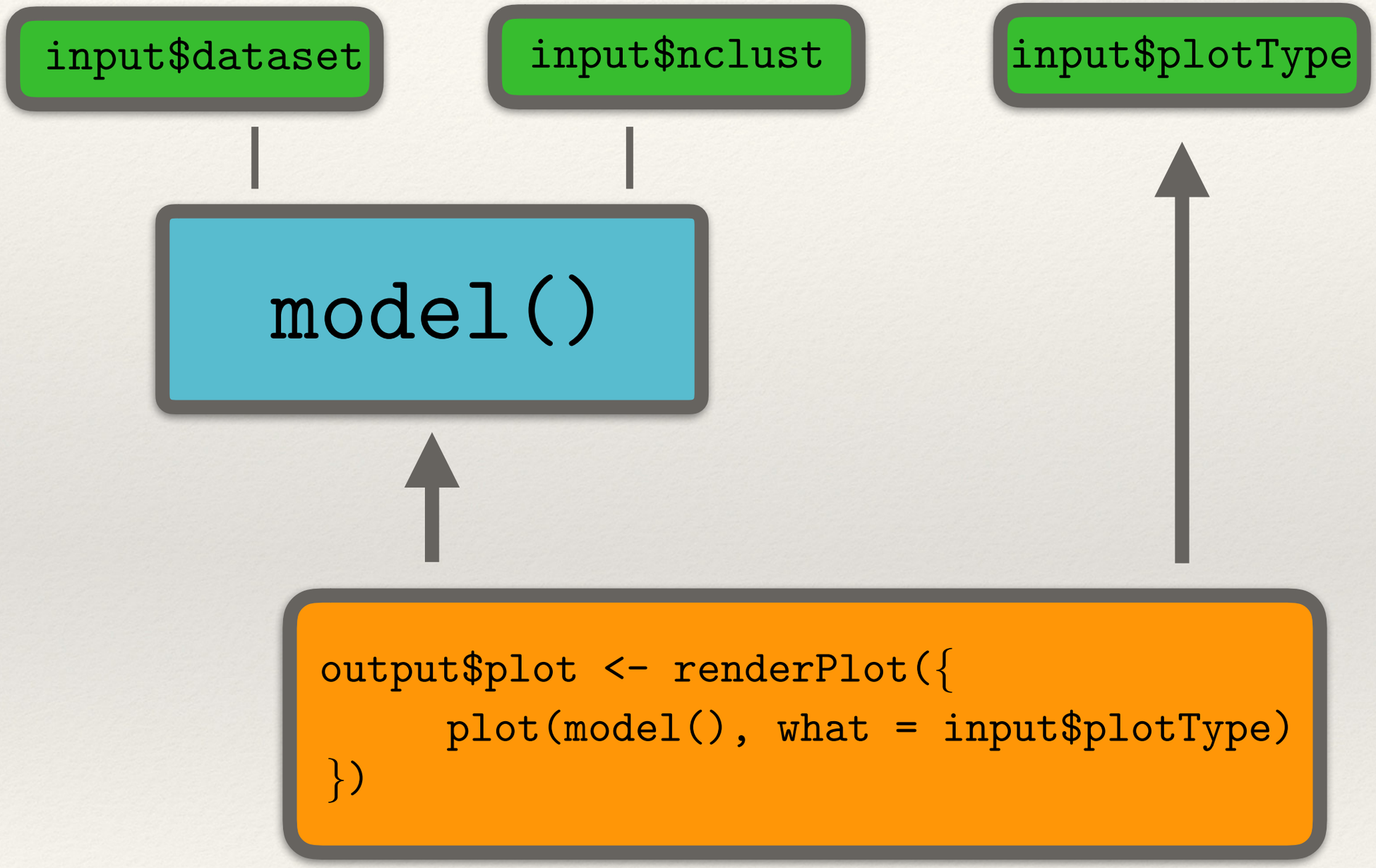
Application (Clustering)



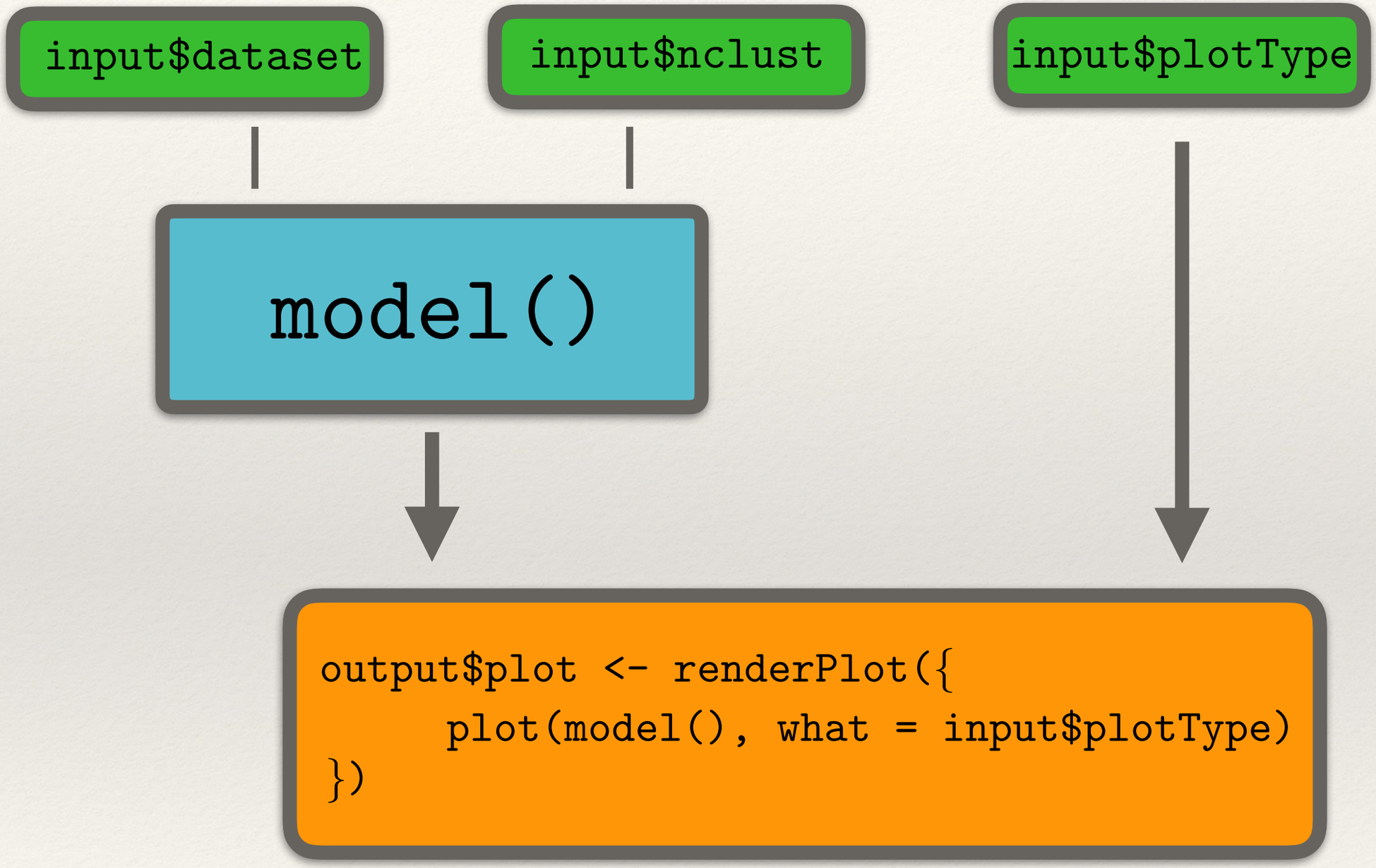
Application (Clustering)



Application (Clustering)



Application (Clustering)



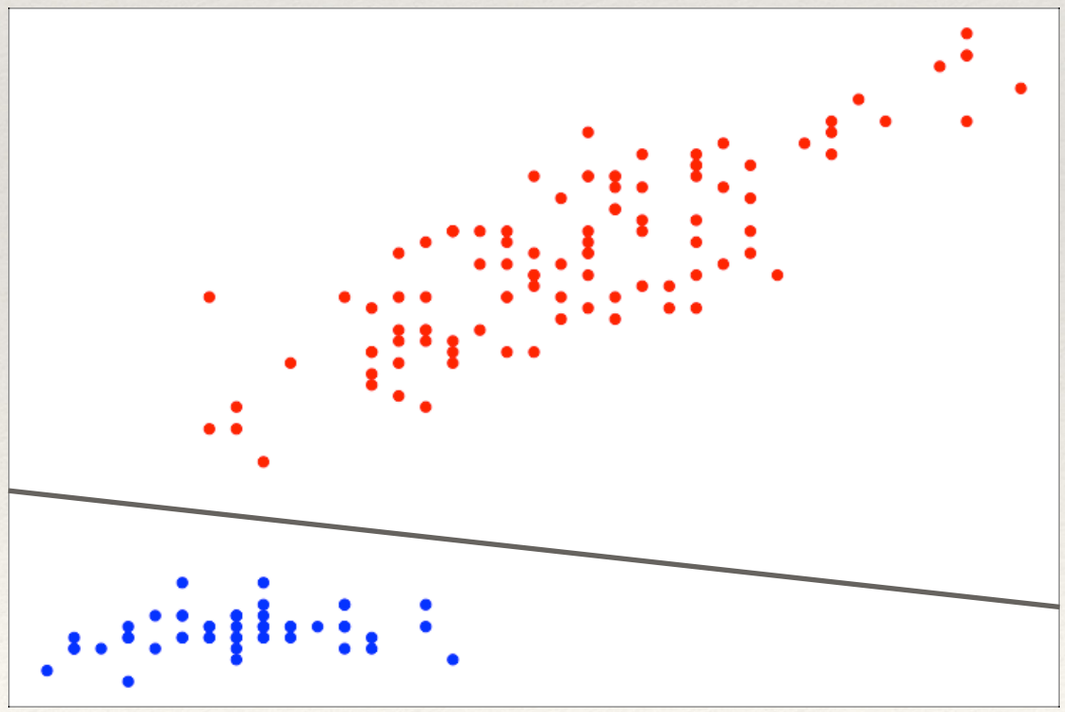
Model-Based Classification

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application

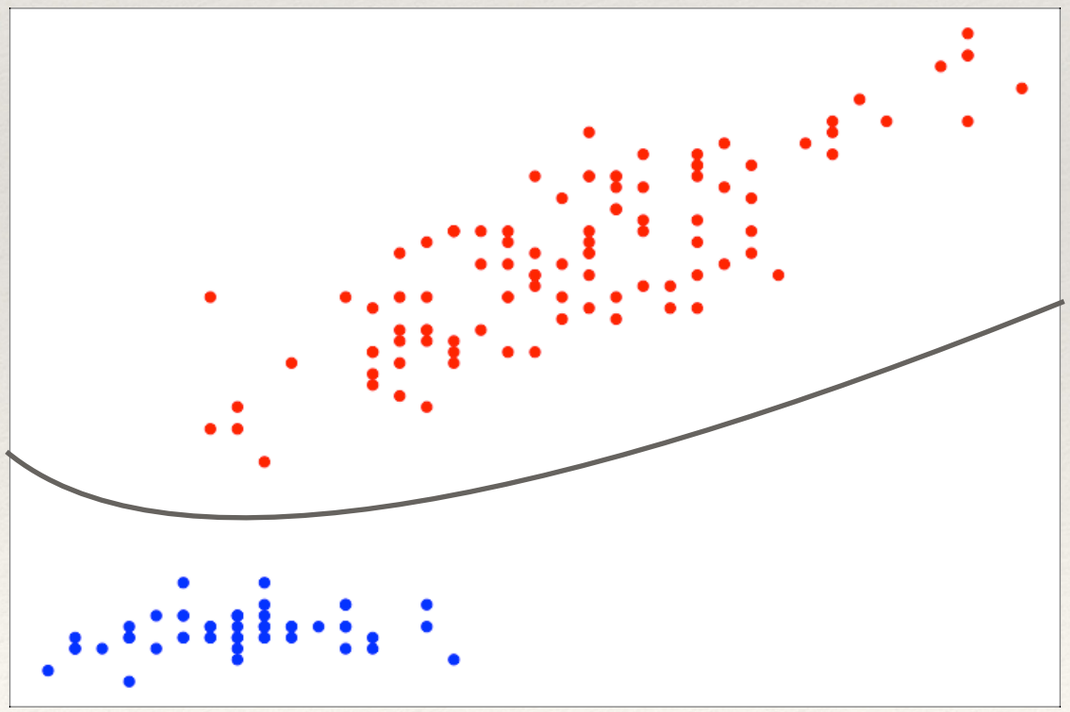
Model-Based Classification

Discriminant Analysis - A supervised statistical technique where the class labels are used to learn the structure of the data



Model-Based Classification

Discriminant Analysis - A supervised statistical technique where the class labels are used to learn the structure of the data



Model-Based Classification

Some Methods for Discriminant Analysis:

- Linear discriminant analysis
- Quadratic discriminant analysis
- Mixture discriminant analysis

Model-Based Classification

Linear Discriminant Analysis

Assume the data is generated from a mixture of multivariate normal distributions

$$\mathbb{P}(g|\mathbf{x}_i) = \frac{\mathbb{P}(g)\mathbb{P}(\mathbf{x}_i|g)}{\sum_{g\prime=1}^{G}\mathbb{P}(g\prime)\mathbb{P}(\mathbf{x}_i|g\prime)} = \frac{\pi_g f_g(\mathbf{x}_i)}{\sum_{g\prime=1}^{G}\pi_{g\prime} f_{g\prime}(\mathbf{x}_i)}$$

$$\pi_g = \frac{n_g}{n}$$

Model-Based Classification

Linear Discriminant Analysis

Assuming equal covariance matrix across groups

$$log(\pi_g f_g(\mathbf{x}_i|\mu_g, \Sigma_g)) = log(\pi_g) + \mathbf{x}_i^T \Sigma^{-1} \mu_g - \frac{1}{2} \mu_g^T \Sigma^{-1} \mu_g = \delta_g(\mathbf{x}_i)$$

Model-Based Classification

Quadratic Discriminant Analysis

Do not assume equal covariance matrices across groups

$$log(\pi_g f_g(\mathbf{x}_i|\mu_g, \Sigma_g)) = log(\pi_g) - \frac{1}{2}log|\Sigma_g| - \frac{1}{2}(\mathbf{x}_i - \mu_g)^T\Sigma_g^{-1}(\mathbf{x}_i - \mu_g) = \delta_g(\mathbf{x}_i)$$

Model-Based Classification

Mixture Discriminant Analysis: A more general case of discriminant analysis in which:

- Can impose constraints on covariance matrices for each class
- Can fit more than one distribution to each class

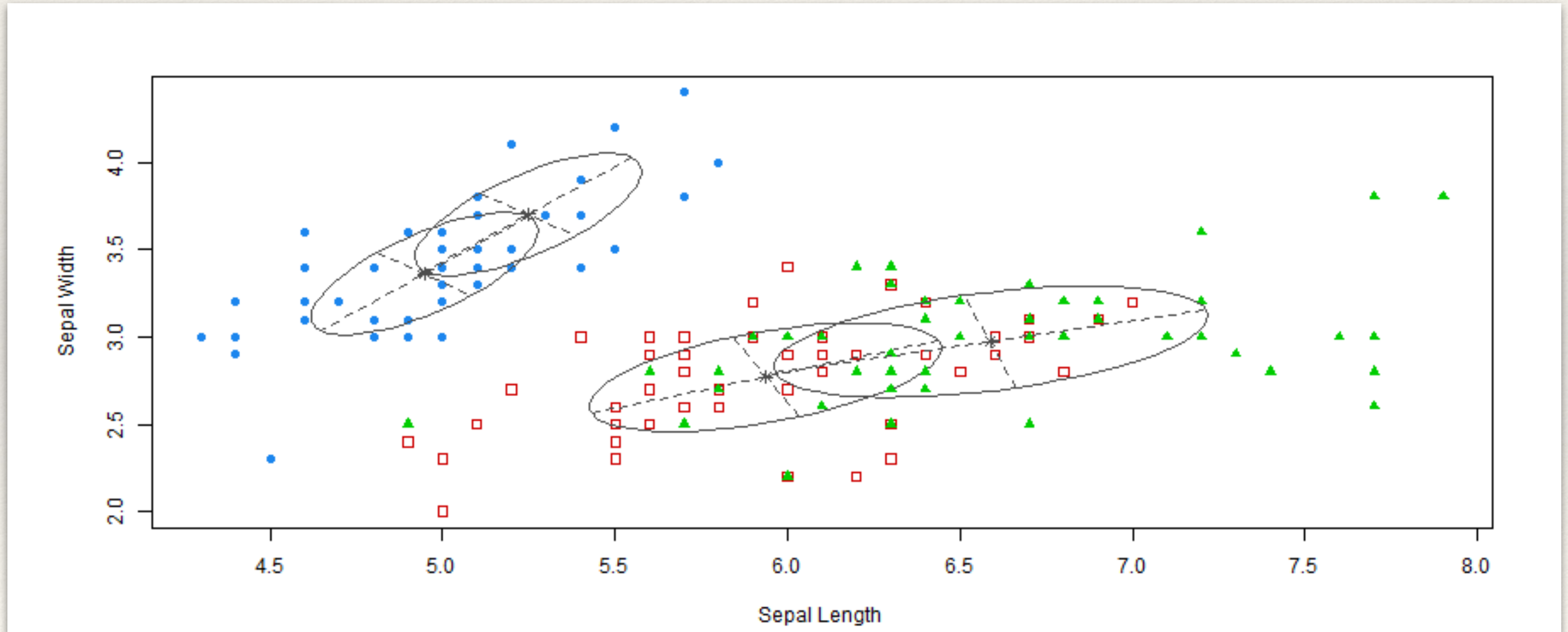
Model-Based Classification

$$\Sigma_k = \lambda_k D_k A_k D_k^T$$

	λ Volume	A Shape	D Orientation
LDA	E	E	E
QDA	V	V	V
MDA	E/V	I/E/V	I/E/V

Model-Based Classification

Mixture Discriminant Analysis



Application (Classification)

Developing a Shiny App for Model-Based Clustering and Classification

- Introduction to model-based clustering and mclust
- Introduction to R Shiny
- Demonstration of model-based clustering application
- Introduction to model-based classification
- Demonstration of model-based classification application