Redução de Dimensionalidade Usando PCA.

Christopher Silva de Pádua
Bacharel em Matemática Computacional
Mestre em Ciência: e Experimentação Agronômica

Cientista de Dados
Semantix

22 de Novembro de 2017

Outline

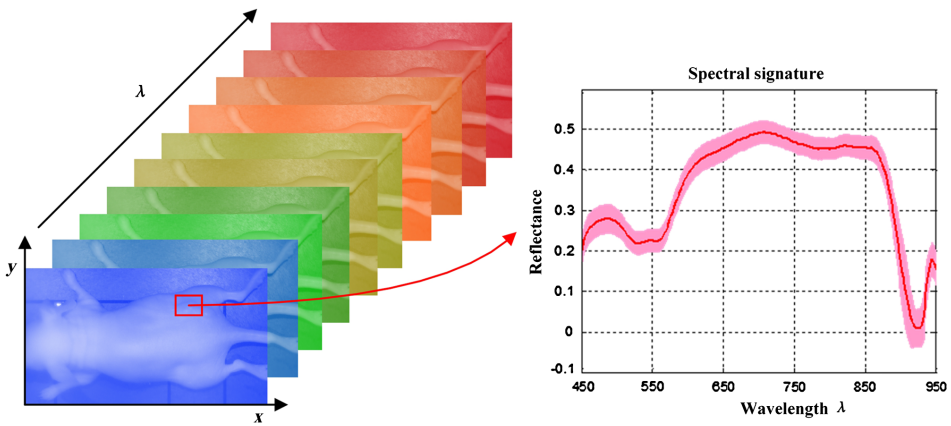
1. Contextualização
2. Conceitos
3. Análise de Componentes Principais
4. Resultados
5. Considerações Finais

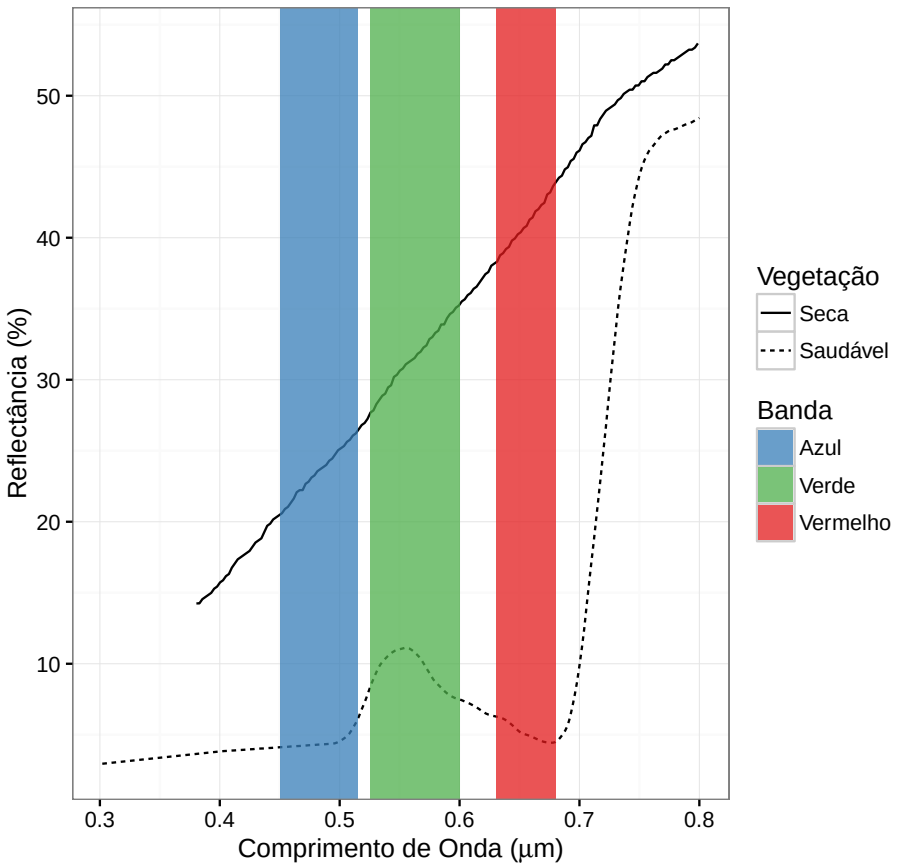
Outline

1. Contextualização
2. Conceitos
3. Análise de Componentes Principais
4. Resultados
5. Considerações Finais

Contextualização

Contextualização

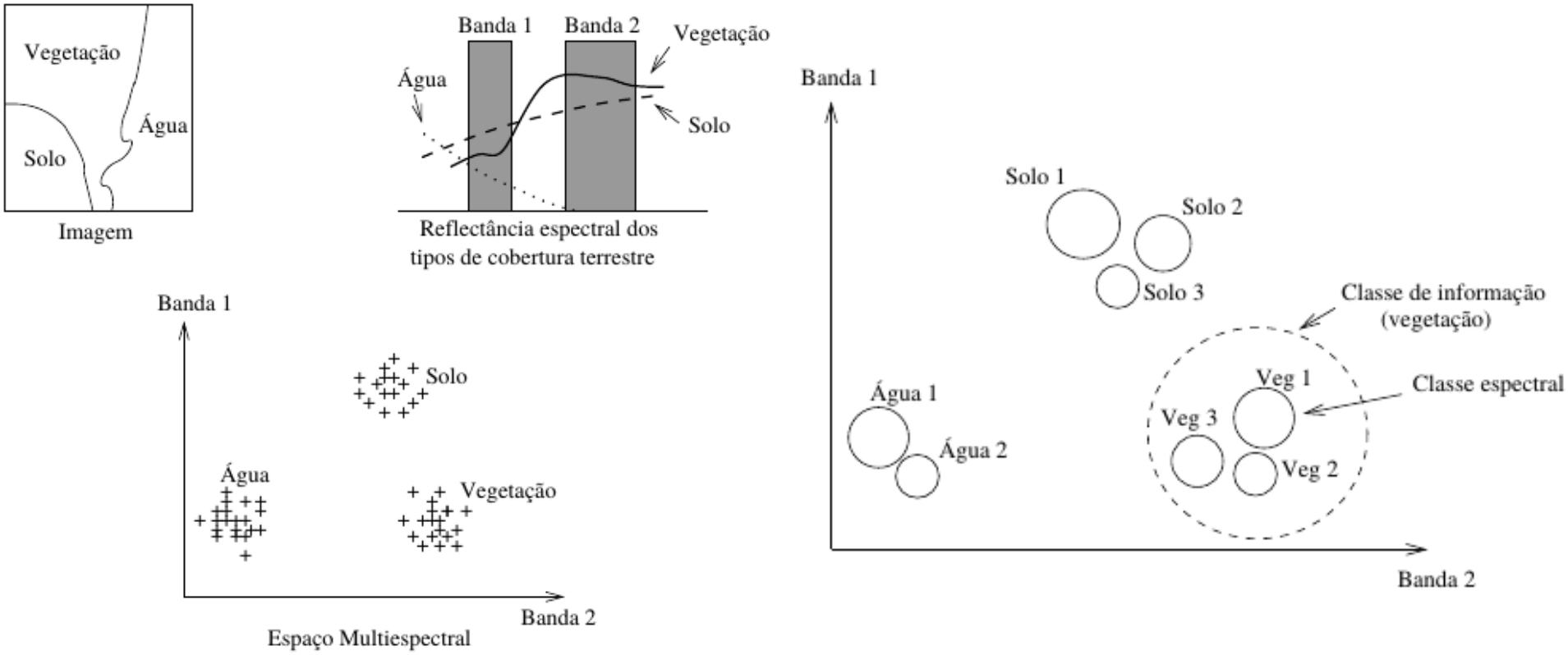




Outline

1. Contextualização
2. Conceitos
3. Análise de Componentes Principais
4. Resultados
5. Considerações Finais

Espaço de atributos



Outline

1. Contextualização
2. Conceitos
3. Análise de Componentes Principais
4. Resultados
5. Considerações Finais

Análise de Componente Principal

$$\mathsf{T}\colon \mathbb{R}^p \to \mathbb{R}^p \Rightarrow \mathbb{R}^d, \quad d < p$$

Componente Principal

Sejam as variáveis $X_1, \ldots, X_p$. As componentes principais $Z_1, \ldots, Z_p$ são combinações lineares destas variáveis.

$$\mathbf{z} = \mathbf{A}\mathbf{x}$$

em que $\mathbf{A}$ é uma matriz de coeficientes e sua i-ésima linha, $\mathbf{a}_i$, satisfaz a seguinte condição:

$$\mathbf{a}_i'\mathbf{a}_i = 1 \quad \forall i \in \{1, \ldots, p\}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 1 & c_{12} & \dots & c_{1p} \\ c_{21} & 1 & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \dots & 1 \end{bmatrix}$$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$$

$$\mathbf{a}_1 = \frac{\mathbf{v}_1}{\| \mathbf{v}_1 \|}, \dots, \mathbf{a}_p = \frac{\mathbf{v}_p}{\| \mathbf{v}_p \|}$$

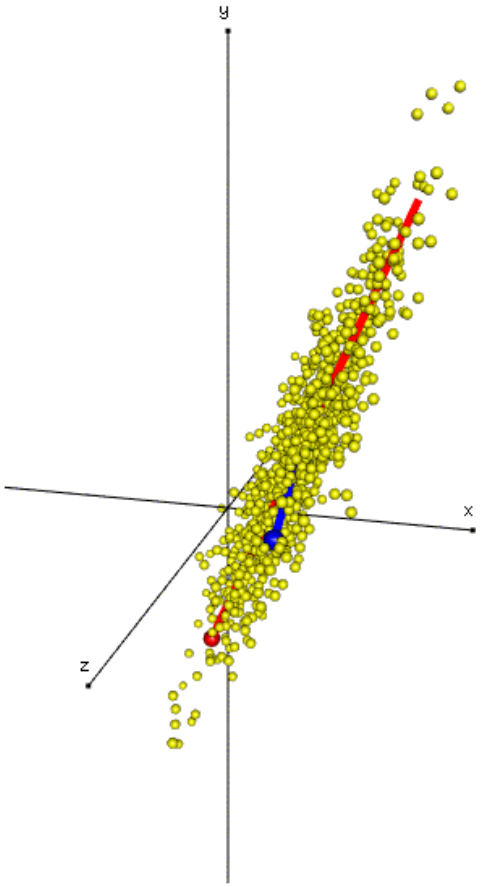
$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_p' \end{bmatrix}$$

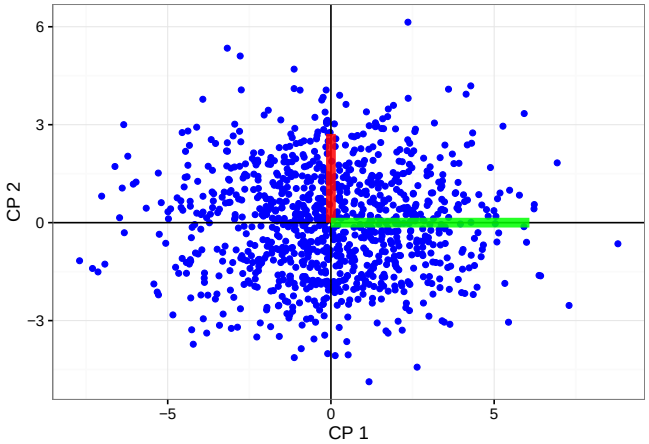
$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 1 & c_{12} & \dots & c_{1p} \\ c_{21} & 1 & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \dots & 1 \end{bmatrix}$$

Propriedades

- $\mathrm{Var}(Z_i) = \lambda_i \quad \forall i \in \{1, \dots, p\}$
- $\mathrm{Cov}(Z_i, Z_j) = 0 \quad \forall i, j \in \{1, \dots, p\} \mid i \neq j$
- $\mathrm{tr}(\mathbf{\Sigma}) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \sigma_{ii}$
- $\mathrm{tr}(\mathbf{C}) = \sum_{i=1}^{p} \lambda_i = p$

Análise de Componente Principal



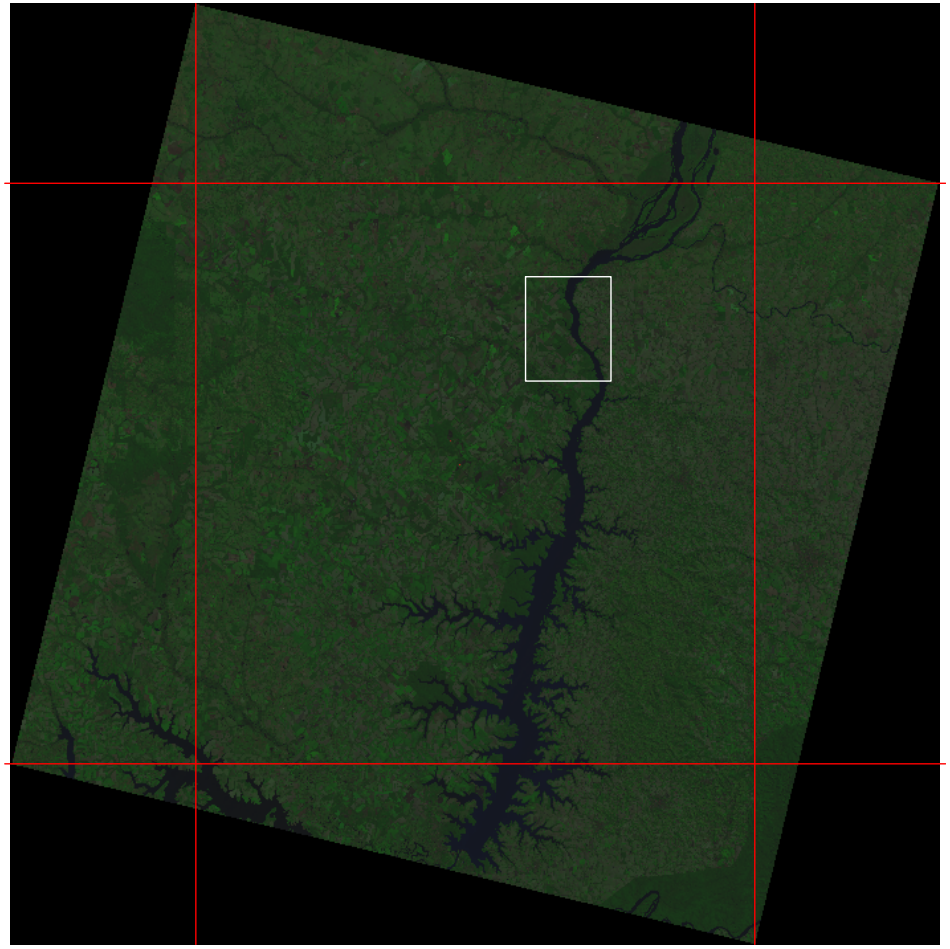


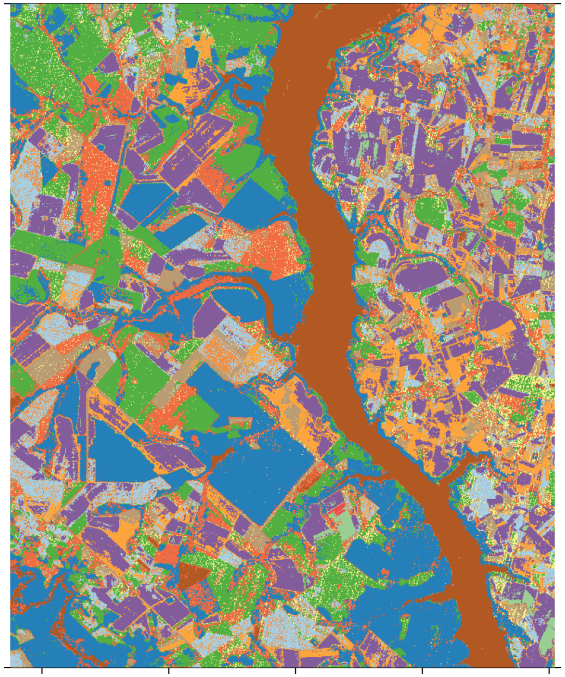
$$\boldsymbol{\Sigma} = \begin{bmatrix} 1.01 & 2.01 & 0.77 \\ 2.01 & 4.93 & 0.73 \\ 0.77 & 0.73 & 2.95 \end{bmatrix} \rightarrow \boldsymbol{\Sigma} = \begin{bmatrix} 6.08 & 0.00 & 0.0 \\ 0.00 & 2.72 & 0.0 \\ 0.00 & 0.00 & 0.1 \end{bmatrix} \rightarrow \boldsymbol{\Sigma} = \begin{bmatrix} 6.08 & 0.00 \\ 0.00 & 2.72 \end{bmatrix}$$

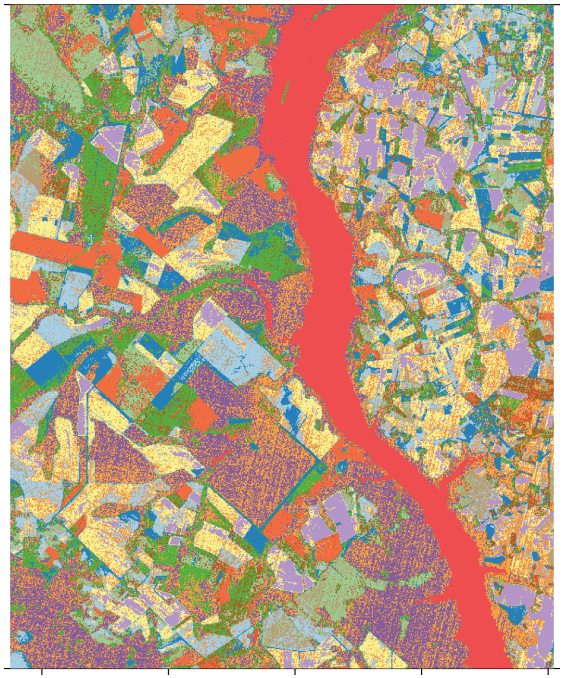
Outline

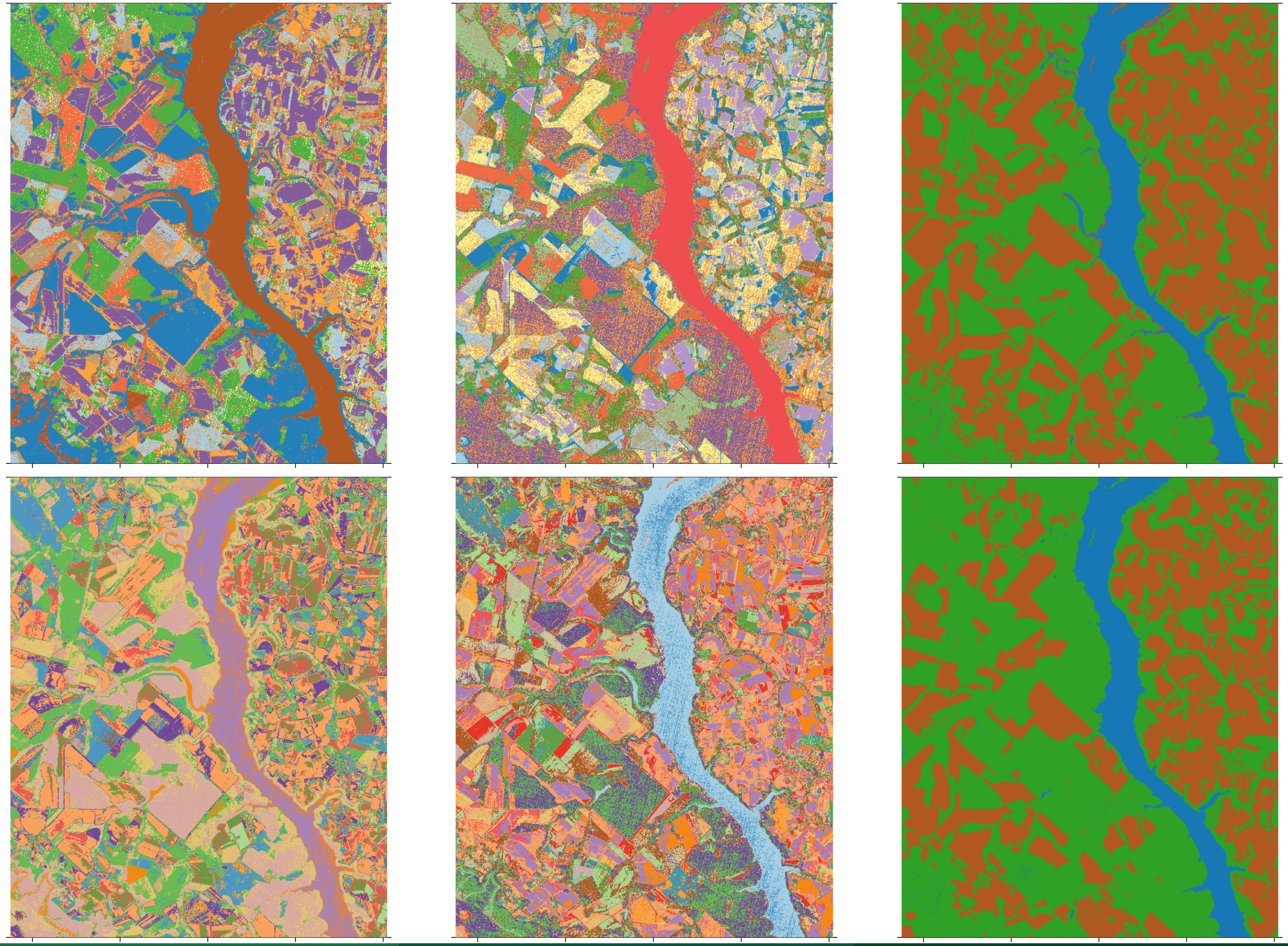
1. Contextualização
2. Conceitos
3. Análise de Componentes Principais
4. Resultados
5. Considerações Finais

Dados









Outline

1. Contextualização
2. Conceitos
3. Análise de Componentes Principais
4. Resultados
5. Considerações Finais

Considerações Finais

Pontos Importantes

- PCA é uma técnica de transformação de dados

Considerações Finais

Pontos Importantes

- PCA é uma técnica de transformação de dados
- Serve para resolver multicolinearidade

Considerações Finais

Pontos Importantes

- PCA é uma técnica de transformação de dados
- Serve para resolver multicolinearidade
- Pode ser usada para reduzir dimensão

Considerações Finais

Pontos Importantes

- PCA é uma técnica de transformação de dados
- Serve para resolver multicolinearidade
- Pode ser usada para reduzir dimensão
- Sempre existe perda de informação

Muito Obrigado!