

# Classificação em Arqueologia e análise de cerâmica arqueológica: processos e métodos para compreender a mudança cultural

Dr. Glauco Constantino Perez (MAE/USP)

Dr. Leandro Elias Canaan Mageste (UNIVASF)

Dr. Astolfo Araujo (MAE/USP)





# Objetivos do curso

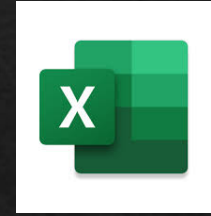
- ◆ Sistematização de dados em arqueologia
- ◆ Banco de dados;
- ◆ Tabelas de frequência – Excel;
- ◆ Selecionando atributos;
- ◆ Seleção de atributos importantes para a identificação de mudanças artefatuais;
- ◆ A aplicação do método Multivariado:
  - ▢ Produção de clusters;
  - ▢ Índices de similaridade, agrupamentos,
  - ▢ Análise de Componentes Principais (PCA) através do PAST.

# Sistematização de dados – banco de dados

- ◆ O que você pode falar sobre o seu banco de dados?
  - ❑ Software utilizado;
  - ❑ Como foi pensado;
  - ❑ Distribuição dos dados;
  - ❑ Intenção de elaboração do banco – motivo da pesquisa;
  - ❑ Erros que possam ter surgido ao longo da elaboração;
  - ❑ Componentes, funcionamento, estrutura do Sistema de Gerenciamento de Banco de Dados (SGBD);
  - ❑ Apresentação de fontes, bases de dados, documentos utilizados para a construção do banco de dados;

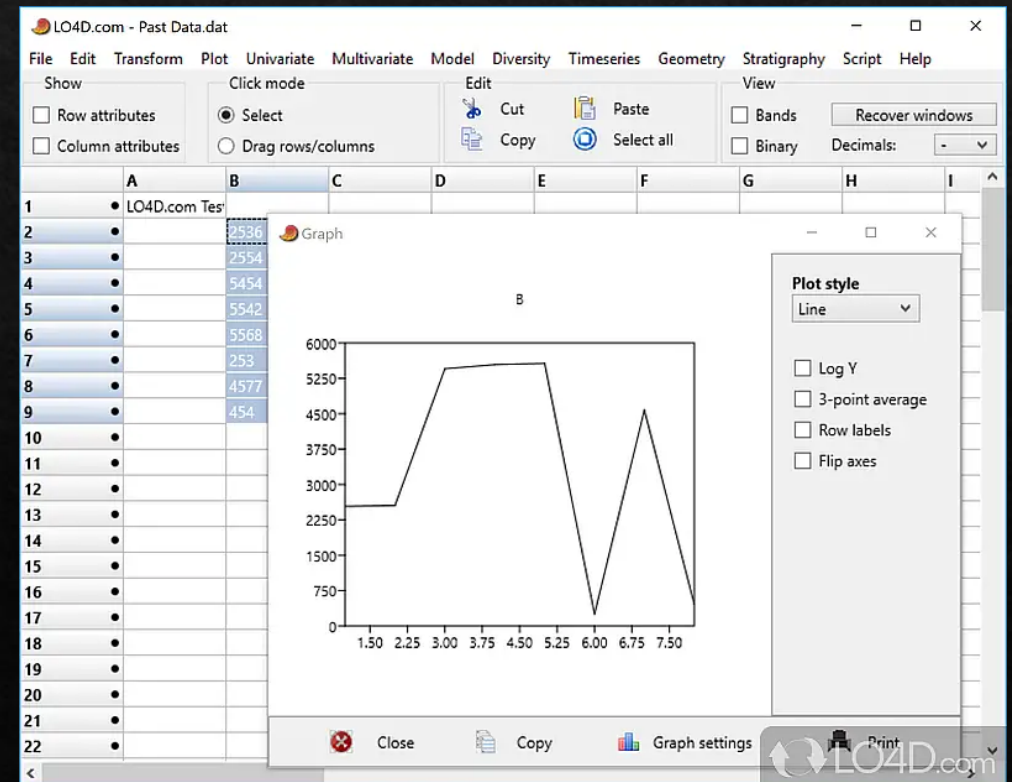


# Software



-Excel;

- PAST - <https://past.en.lo4d.com/windows>



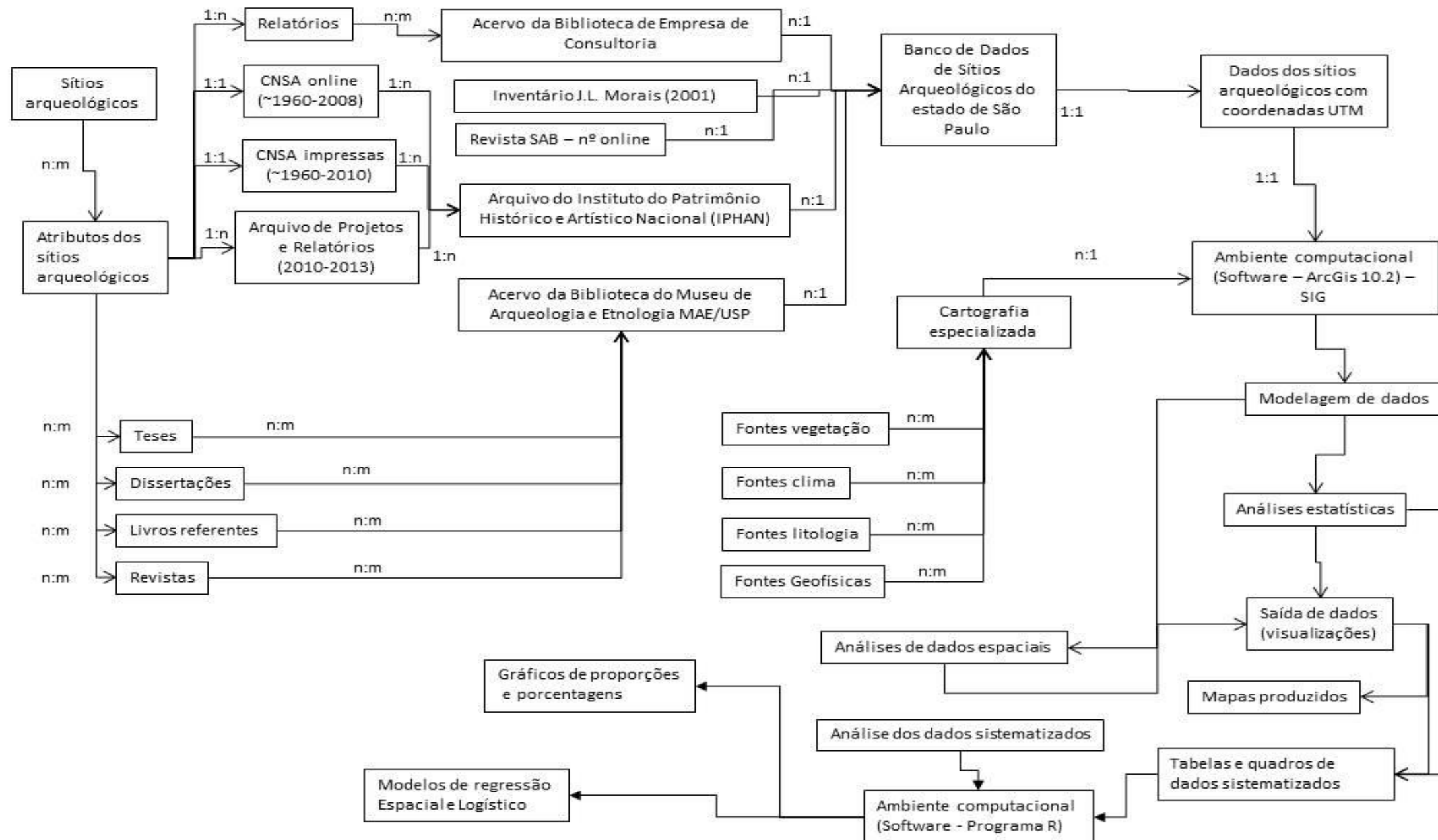
# Relação entre os componentes do SGBD

- ◆ Relacionamento entre entidades e objetos do banco de dados pode ser entendido como uma associação de atributos:
- ◆ Relação 1:1 - um sítio arqueológico como um objeto e estabelecer o relacionamento com outro sítio;
- ◆ Relação 1:n - uma sítio e seus diversos atributos componenciais (tradição, datação, dimensão, entre outros) - um para vários atributos;
- ◆ Relação n:m - uma relação entre os atributos característicos dos diversos sítios englobados no Banco de dados, isso é, uma relação de vários para vários;

# Relação entre os componentes do SGBD

- ♦ As relações entre as entidades e objetos na pesquisa podem ser:
  - ▢ **Hierárquicas:** em que um registro principal (pai) está associado a registros secundários (filhos ou detalhados), como se fossem uma árvore de dados;
  - ▢ **Redes:** em que a ideia de registros “pai” e “filhos” deixa de fazer sentido, pois todos estão relacionados a todos, numa relação **n:m**;
  - ▢ **Relacional:** em que os dados e atributos estão organizados em forma de tabelas e o relacionamento se dá através de códigos de ligação (elos) entre as diversas tabelas ou simplesmente através dos registros armazenados em campos comuns de tabelas distintas - esse último tipo de relação permite que novas tabelas sejam geradas através da associação entre elas dentro do SIG;





# Seleção de atributos

- ♦ Seleção de atributos deve ser pensado para o objetivo da pesquisa;
- ♦ Criando TABELA DINÂMICA:
- ♦ Excel > Banco de dados > Inserir > Tabela dinâmica
- ♦ Sítio Itaguá\_Ubatuba\_SP
- ♦ Sítios Piraju\_ análise de cluster, similaridade e PCA



# Análises multivariadas – agrupamentos e similaridade entre objetos

- ♦ ANÁLISE MULTIVARIADA ou MVA (Multivariate Data Analysis) é uma técnica analítica que usa informações de várias fontes simultaneamente para obter uma síntese completa do que está sendo observado;
- ♦ ANÁLISE DE AGRUPAMENTOS é uma técnica da estatística multivariada e tem como objetivo entender como pode um objeto ser semelhante a outros dentro do agrupamento com base em um conjunto de características escolhidas;
- ♦ Os agrupamentos resultantes devem exibir homogeneidade interna dentro dos agrupamentos e heterogeneidade externa, entre agrupamentos, isto é, se o agrupamento for bem sucedido, os objetos dentro dos agrupamentos estarão próximos quando representados graficamente, e diferentes agrupamentos estarão distantes;
- ♦ A questão de similaridade dentro dos agrupamentos são normas da estatística, isso deve ficar realçado, pois com as análises de agrupamento esse é um dos resultados a que se espera observar entre conjuntos analisados

# Análises multivariadas – agrupamentos e similaridade entre objetos

- ◆ a similaridade entre objetos é uma medida empírica de correspondências, ou semelhança, entre objetos a serem agrupados e pode ser medida de diversas maneiras, sendo:
  - ◆ Tipos de medida de distância:
    - ▢ Distância Euclidiana é a medida mais comumente reconhecida, muitas vezes chamada de distância em linha reta;  
  
No gráfico: a Distância Euclidiana quanto mais próximo do zero  $[0]$  mais similares são as estruturas, pois elas estão próximas delas mesmas e podem variar ao infinito  $[\infty]$ , com suas distâncias expressas de maneira cartesiana.
    - ▢ Coeficiente de Jaccard : varia entre  $[0 \text{ e } 1]$  - quanto mais próximo do 1, mais serão semelhantes;



# Índices de similaridade – dendrogramas de similaridade

- ◆ Os dados devem estar organizados com espécies (na coluna A) e as amostras ou variáveis em colunas adjacentes;
- ◆ informações ausentes devem ser preenchidas com zero e para inserir os zeros (vá em substituir, substituir “nada” por “zeros” – isso em uma tabela de frequência dentro de uma planilha Excel);
- ◆ Os dados destas planilhas devem ser copiados (Ctrl +C) e utilizando os softwares PAST os dados devem ser transferidos;
- ◆ Ativa-se os itens “row attributes” e o “column attributes” e ao lado da coluna A, e acima da linha 1 os dados devem ser colados no PAST (Ctrl +V) – dessa forma, todos os atributos da planilha do Excel serão transportados para o software. Esse procedimento é para a edição da planilha no PAST. Após a inserção dos dados no PAST, os itens “row attributes” e o “column attributes” devem ser desabilitados;
- ◆ É indicado que nesse momento seja feita a passagem dos itens de colunas para linhas e linhas transpostas por colunas, através do procedimento: Edit > Rearrange> Transpose;
- ◆ Nesse rearranjo, as informações que eram Sítios (linhas) X atributos (colunas) se tornam Atributos (linhas) X Sítios (colunas); esse procedimento é pertinente, pois o Excel apresenta colunas limitadas, enquanto a quantidade de linhas é ilimitado, e dessa forma a quantidade de colunas seria limitada (o número de sítios é limitado) e a quantidade de atributos (linhas) ilimitado; o que facilita a apresentação dos dados.
- ◆ No PAST, toda a planilha deve ser selecionada clicando no canto superior esquerdo da planilha, logo acima da primeira linha – a célula em cinza. Após isso: Multivariate > Similarity and distance índices e automaticamente o procedimento será feito.



# Calculando clusters ou agrupamentos

- ◆ No PAST:
- ◆ procedimento: Multivariate> Cluster> Classical.
- ◆ O cluster é elaborado automaticamente e atentar-se para o algoritmo utilizado é o Paired group (UPGMA)
  - grupos pareado, com o index de similaridade Euclidiano.
- ◆ Em ‘Graph settings’, o eixo Y deve ser colocado em 0 (ZERO) e o X em 1 (UM), para que o índice de similaridade seja entre 0-1 como o que é representado através do teste de distância feito anteriormente.

# Análise de Componentes Principais (ACP) ou *Principal Component Analysis* (PCA)

- ♦ é um procedimento estatístico que utiliza uma transformação ortogonal para converter um conjunto de observações de variáveis possivelmente correlacionadas num conjunto de valores de variáveis linearmente não correlacionadas chamadas de componentes principais (MINGOTI, 2005);
- ♦ Essas projeções são apresentadas a partir dos dados de frequência analisados e podem apresentar redundância na interpretação junto aos clusters feitos a partir da Distância Euclidiana, pois apresentam os mesmos dados crus (*Raw*) utilizados para a construção de ambas as visualizações;
- ♦ A Distância Euclidiana já é uma forma de visualização dos dados de maneira cartesiana, mas o uso da PCA auxilia na interpretação desses dados;
- ♦ A Análise de Componentes Principais (ACP) ou *Principal Component Analysis* (PCA) pode ser calculado a partir do *software* Past que exige a implementação dos dados crus, como os inseridos para a elaboração dos clusters. O procedimento é: Multivariate> Ordination> Principal Components (PCA).

Uma janela a parte é aberta com os componentes principais apresentados a partir de seus valores e a variância (%) dentro da aba Summary.

- ♦ Na aba seguinte Scatter Plot é apresentado um gráfico de dispersão em que se pode manipular inserindo elipses com 95% de representatividade, etiquetas para identificar os dados apresentados no gráfico;



# Referências

HAIR JR., Joseph F.; BLACK, William C.; BABIN, Barry J.; ANDERSON, Rolph E.; TATHAM, Ronald L. 2009. *Análise Multivariada de Dados*. Porto Alegre. Bookman.

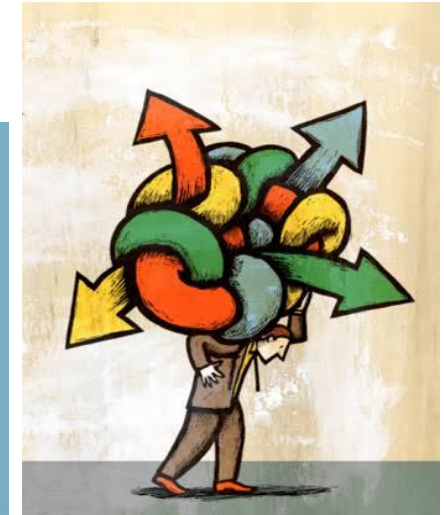
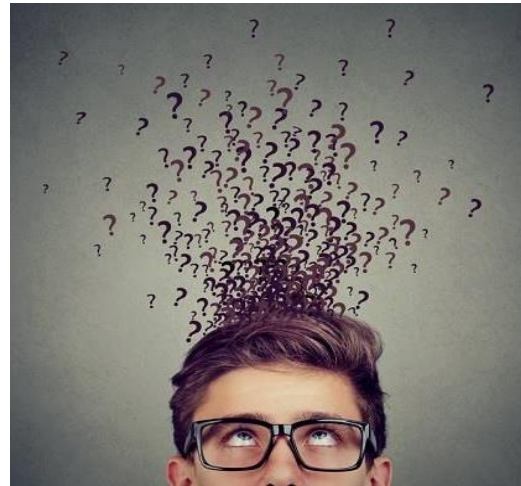
BANNING, Edward B. 2020. *The Archaeologist's Laboratory. The analysis of archaeological evidence*. Second Edition. Springer.

MINGOTI, Sueli Aparecida. 2005. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG.

PEREZ, Glauco Constantino. 2018. *Arqueologia Paulista e o marcador cerâmico como delimitador de fronteira étnica: um estudo das regiões sul e oeste do Estado de São Paulo*. Tese de doutorado. MAE/USP.

PEREZ, Glauco Constantino. 2021. *Relatório de Atividades. O Estabelecimento de redes de transmissão cultural entre os grupos Tupiguarani no sul do Estado de São Paulo*. Relatório CNPq. Não Publicado.





# Obrigado pela atenção!

## Contatos:

E-mail: [glaucou113@gmail.com](mailto:glaucou113@gmail.com)

Instagram: @glaucou113

Facebook: <https://web.facebook.com/glauco.constantino.9/>

=]

## Mais informações do professor:

Plataforma Lattes: <http://lattes.cnpq.br/6873863102873855>

Google Scholars: <https://scholar.google.com.br/citations?hl=pt-BR&user=tbnVvmkAAAAJ>

ORCID: <https://orcid.org/0000-0002-1472-3360>

ReasearcherID: <http://www.researcherid.com/rid/P-4191-2018>

## Agradecimentos

Fundação de Amparo à Pesquisa do Estado de São Paulo - Bolsa de Pós-doutorado Processo 2021/04562-0. "As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do(s) autor(es) e não necessariamente refletem a visão da FAPESP".