

# **Shiny Data Visualization**

## Tutorial

# Índice

## Inicialização

- Sobre o aplicativo
- Inserir os dados
- Utilizar os dados de exemplo
- Importar os dados de um arquivo Excel
- Inserir os dados na planilha
- Salvar imagens
- Layout
- Resumo dos dados

## Gráficos 2D

- Histograma
- Box Plot
- Violino
- Gráfico de pontos
- Gráfico de Densidade
- Gráfico de Erro

## Gráficos 3D

- Histograma 3D
- Gráfico de Densidade
- Gráfico de pontos
- Gráfico de Barras

# Inicialização

## Sobre o aplicativo:

Este aplicativo tem função de visualizar e comparar vários dados, com uma única variável discreta, do mesmo tipo.

O aplicativo aceita dados em arquivos .xlsx, no qual a primeira linha são os nomes dos dados, e as n linhas restantes são os dados de cada uma das colunas. O aplicativo também possui uma interface com uma planilha interativa onde ele é válido, apenas que os espaços para os nomes das colunas já estão definidos.

Com os dados já inseridos, o aplicativo possui três funcionalidades, mostrar um resumo com algumas informações sobre os dados inseridos, um painel contendo gráficos 2D e um painel contendo gráficos 3D.

Existem 6 tipos de gráficos 2D: Histogramas, Box Plot, Violino, Gráfico de pontos, Gráfico de Densidade e Gráfico de Erros, e existem 5 tipos de gráficos 3D: Histogramas, Gráfico em Superfície, Gráfico de pontos, Gráfico de Densidade e Gráfico de Barras.

Cada um dos gráficos pode ser personalizado, modificando cores, estilos, tamanhos e algoritmos para o cálculo dos gráficos, dependendo de cada um dos gráficos. Para esta função existe uma barra de controle de opções na lateral, específica para cada um dos gráficos.

Por fim, todos os gráficos podem ser baixados em um arquivo .png.

## Inserir os dados:

Existem três opções para a inserção dos dados pelo usuário, dados de exemplo que já vem baixados com o aplicativo, inserir dados excel e inserir dados utilizando a planilha.

Para modificar o título, o usuário deve escrever um título abaixo do texto “Digite o título”.

Todos os dados inseridos devem ser palavras (para os títulos ou os nomes das colunas) ou números (para os dados restantes).

## Utilizar os dados de exemplo:

Para os dados exemplares você deve selecionar o painel:

Seus Dados - > Carregue seus dados - > Usar um arquivo exemplo

Existem 3 opções de exemplos: “Construções”, que mostra o número de construções de cada uma das quatro regiões do Centro-Oeste dos Estados Unidos, para cada um dos meses, desde janeiro de 2018 até setembro de 2018. “Temperatura Lincoln”, que mostra 3 categorias. A média das temperaturas mínimas, a média das temperaturas médias, e a média das temperaturas máximas de cada mês. “demografia – Midwest” mostra a demografia étnica dos povos no centro-oeste dos Estados Unidos.

## Importar os dados de um arquivo Excel:

Para inserir uma planilha Excel com os dados você deve selecionar o painel:

Seus Dados - > Carregue seus dados - > Importar um arquivo .xlsx

Os dados devem estar neste formato:

	A	B	C	D	E	F	
1	Arvore 1	Arvore 2	Arvore 3	Arvore 4	Arvore 5	Arvore 6	
2	17	15	21	14	17	20	
3	18	13	24	16	16	19	
4	22	17	22	19	18	19	
5	19	19	19	24	19	20	
6	18	18	18	13	18	21	
7	21	17	18	18	16	21	
8	20	16	21	16	17	19	
9	20	19	20	23	18	20	
10	18	14	17	20	18	1	
11							

Com a primeira linha contendo o nome de cada uma das colunas e as demais linhas contendo os dados de cada uma das colunas. O nome das colunas é essencial

## **Inserir os dados na planilha:**

Para inserir uma planilha Excel com os dados você deve selecionar o painel:

Seus Dados - > Digite seus dados na planilha

Os dados devem estar neste formato:

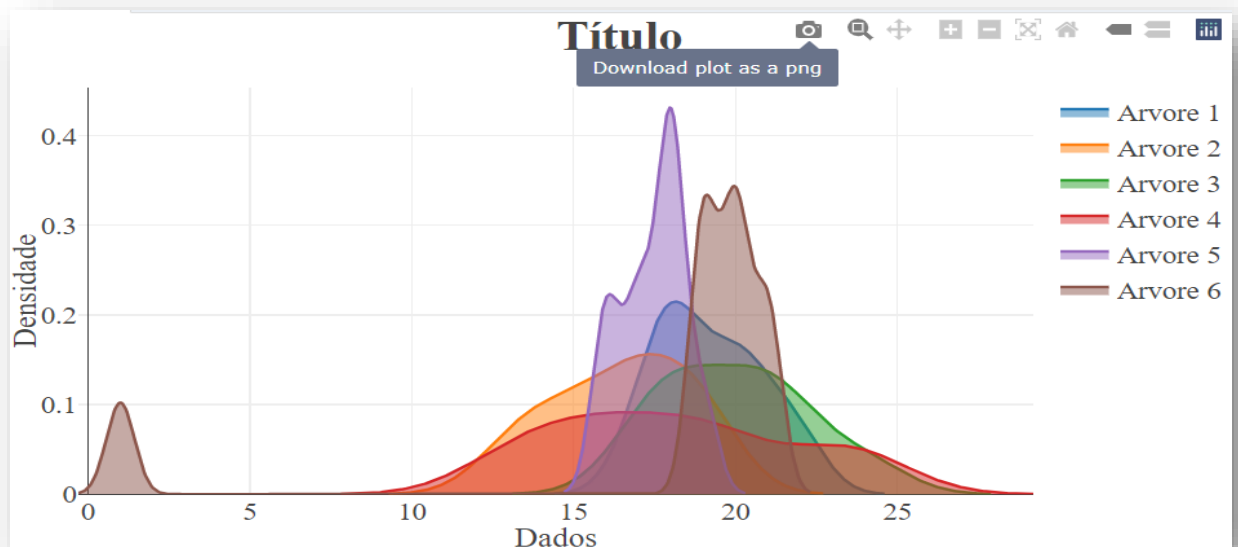
Arvore 1	Arvore 2	Arvore 3	Arvore 4	Arvore 5	Arvore 6	V7	
17	15	21	14	17	20		
18	13	24	16	16	19		
22	17	22	19	18	19		
19	19	19	24	19	20		
18	18	18	13	18	21		
21	17	18	18	16	21		
20	16	21	16	17	19		
20	19	20	23	18	20		
18	14	17	20	18	1		

Devendo modificar cada um dos nomes das colunas, e inserindo os dados de cada uma das colunas. Opções de formatação como copiar e colar dados excel funcionam na planilha, além de outras funcionalidades como selecionando um conjunto retangular de células e arrastando no canto inferior direito, fazendo uma segunda seleção copiando a primeira seleção e colando na segunda.

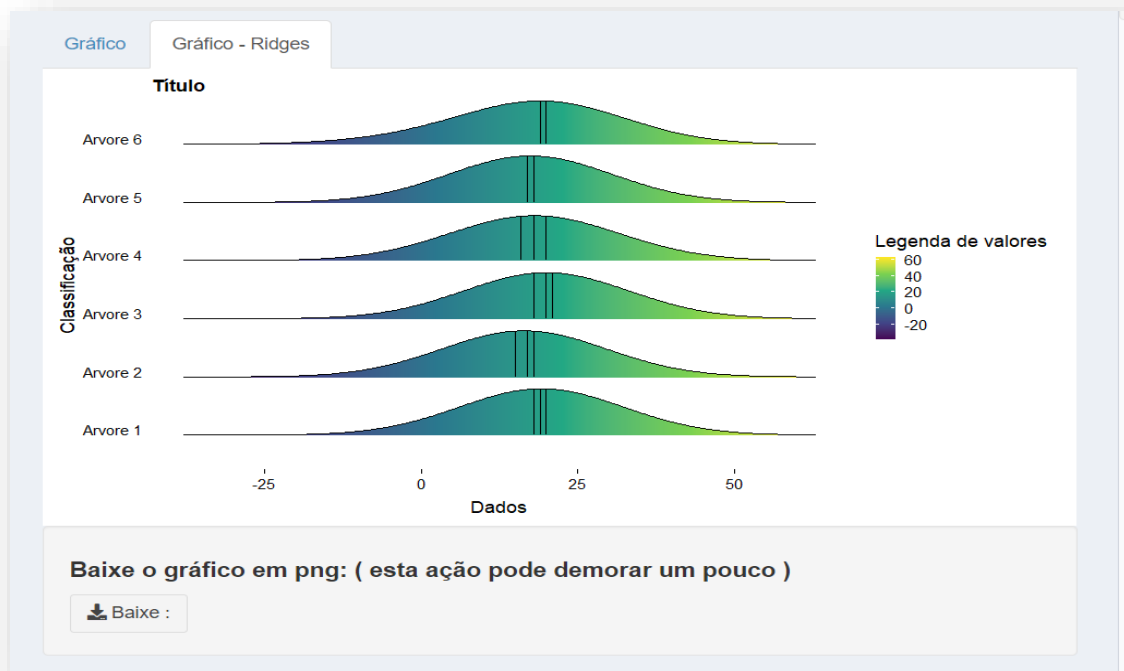
## Salvar imagens:

Todos os gráficos possuem pelo menos um meio para salvar os gráficos em uma imagem .png

Para gráficos no modelo plotly, deve-se clicar no botão “Download plot as a png” como no exemplo abaixo:

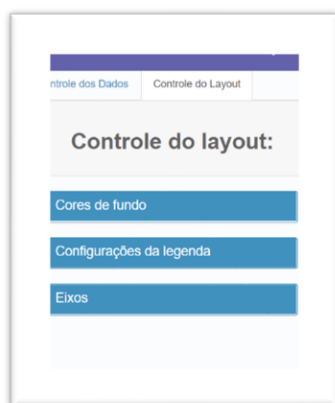


Já para gráficos que não utilizam o plotly como ferramenta de visualização, terá um botão embaixo do gráfico escrito “Baixe o gráfico em png:” como no exemplo abaixo:



## Layout:

Todos os gráficos possuem uma aba dentro do painel de controles, ao lado da aba Controle de Dados, chamada Controle de Layout. Nesta aba o usuário pode personalizar um gráfico específico modificando as legendas, as cores e os eixos.



## Resumo dos dados:

Após carregar os dados irá surgir um painel escrito “Informações gerais dos dados” na barra lateral esquerda, e neste painel tem uma tabela com os dados carregados e uma outra tabela abaixo com as informações dos dados, como a média, desvio padrão, valores mínimos e máximos, a variância, e o primeiro e terceiro quadrante. A segunda tabela

pode ser transposta na barra de controle de opções, ao clicar em: “Transcrever tabela de dados estatísticos”

Dados estatísticos:								
how	10	entries		Search: <input type="text"/>				
	Média	Desvio Padrão	Mínimo	1º Quadrante	Mediana	3º Quadrante	Máximo	Variância
Lab 1	52.72	14.84	27	40	53	65.75	77	220.2256
Lab 2	39.573	6.634	30	33	39	46	50	44.009956
Lab 3	55.953	20.379	20	39.25	56.5	74.5	89	415.303641
Lab 4	45.16	26.087	1	23	45	61	97	680.531569
Lab 5	61.573	6.768	1	58	62	66	70	45.805824
howing 1 to 5 of 5 entries							Previous	1 Next

## Gráficos 2D

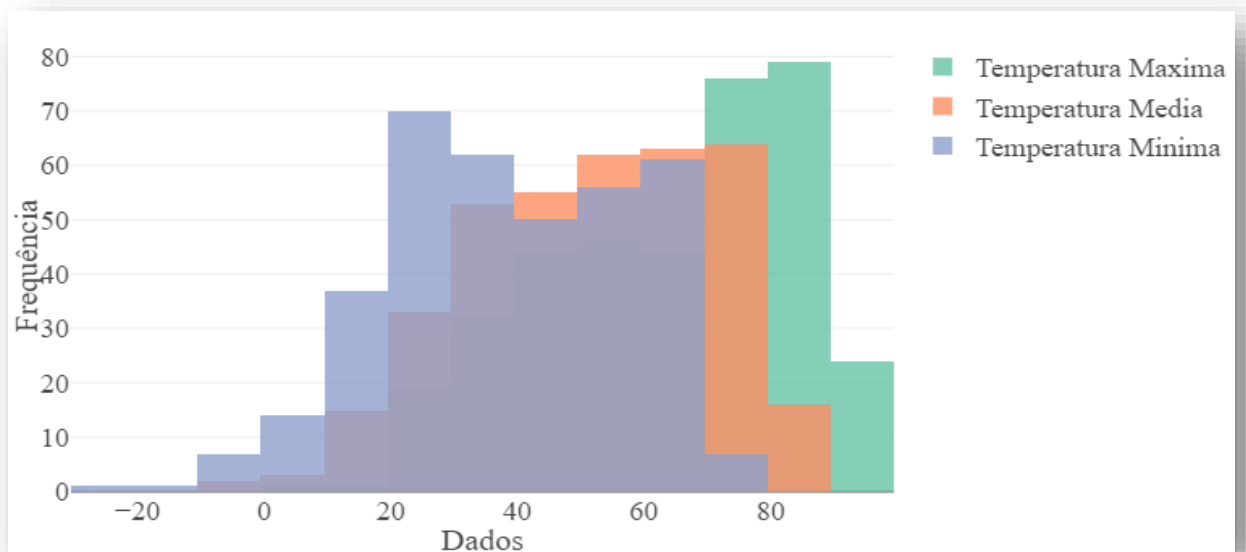
Os gráficos 2D podem ser encontrados na barra lateral esquerda após carregar os dados no aplicativo.

### Histograma:

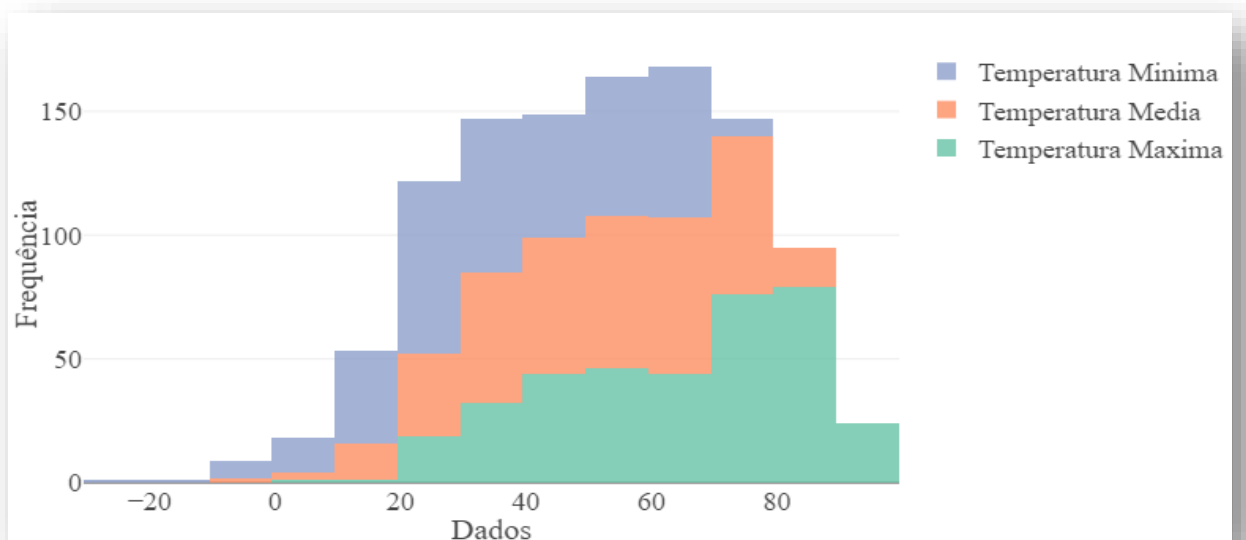
O painel Histograma possui dois tipos de histogramas, o primeiro está em formato de barras, da frequência sobre os dados numéricos. No controle de opções o usuário pode modificar o gráfico para um histograma empilhado e/ou com espaçamento entre as barras.

Histograma sem modificações:

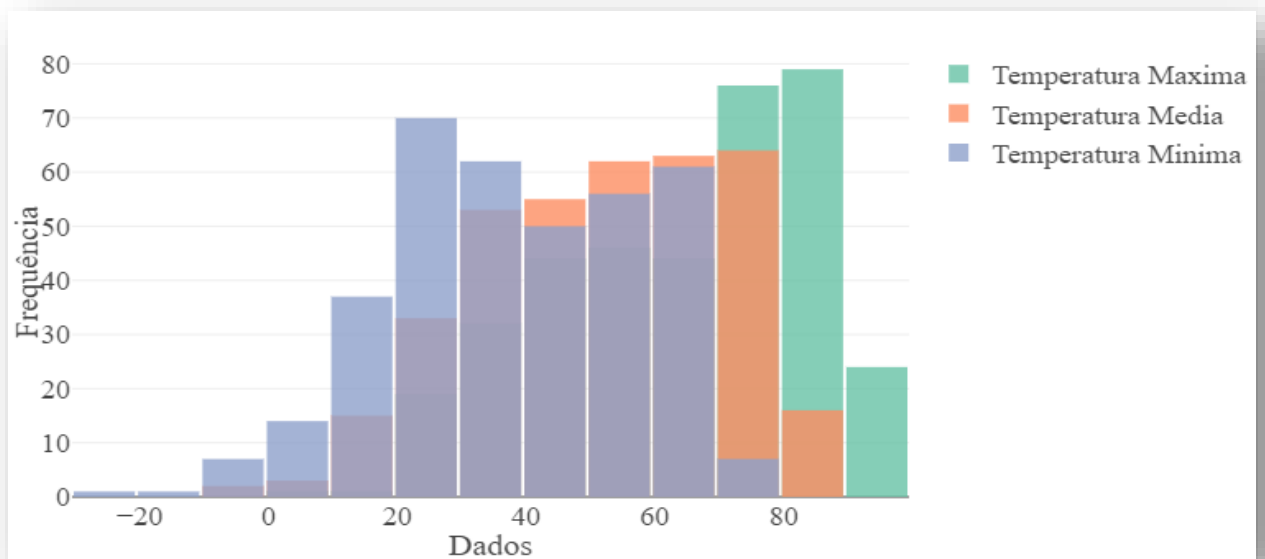




Histograma empilhado:



Histograma com espaçamento:



### Mais modificações:

O usuário pode modificar a quantidade de bandas, ou seja, aumentar o número de barras representativas do gráfico correspondente ao seu tamanho, desativando o intervalo padrão do histograma; e também a opacidade das barras.

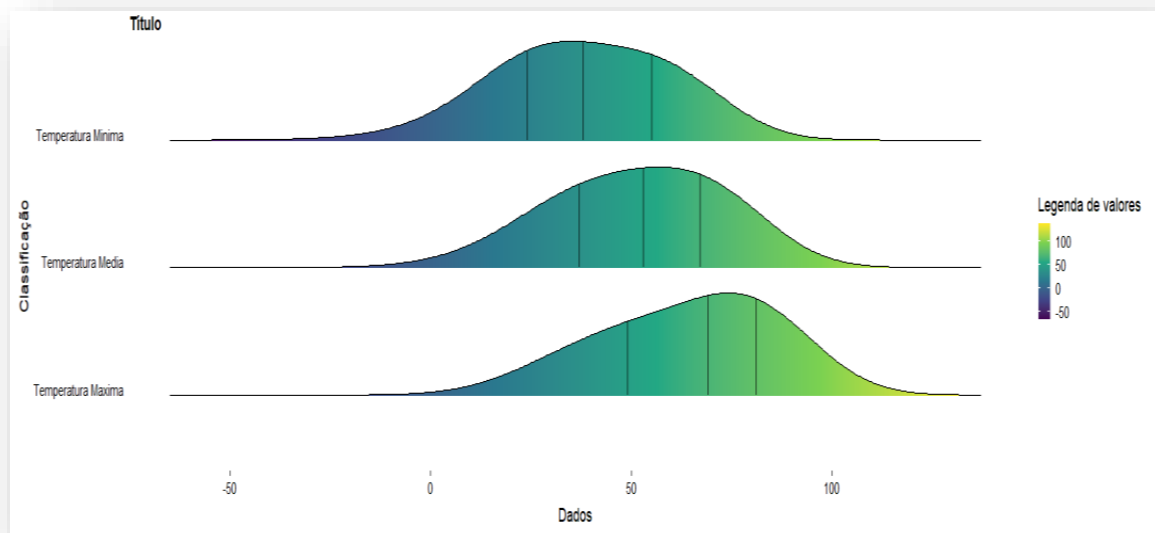
O intervalo padrão do histograma é calculado utilizando a regra de Sturges

O tamanho do espaçamento pode ser modificado caso selecione ativar os espaçamentos.

Um histograma de densidade pode ser inserido, podendo modificar a escala do histograma, a opacidade do histograma e a opacidade da área do histograma.

No segundo tipo de histograma, mostra um gráfico contendo um histograma único para cada uma das colunas, cada um dos histogramas empilhados.

Histograma com as configurações padrões



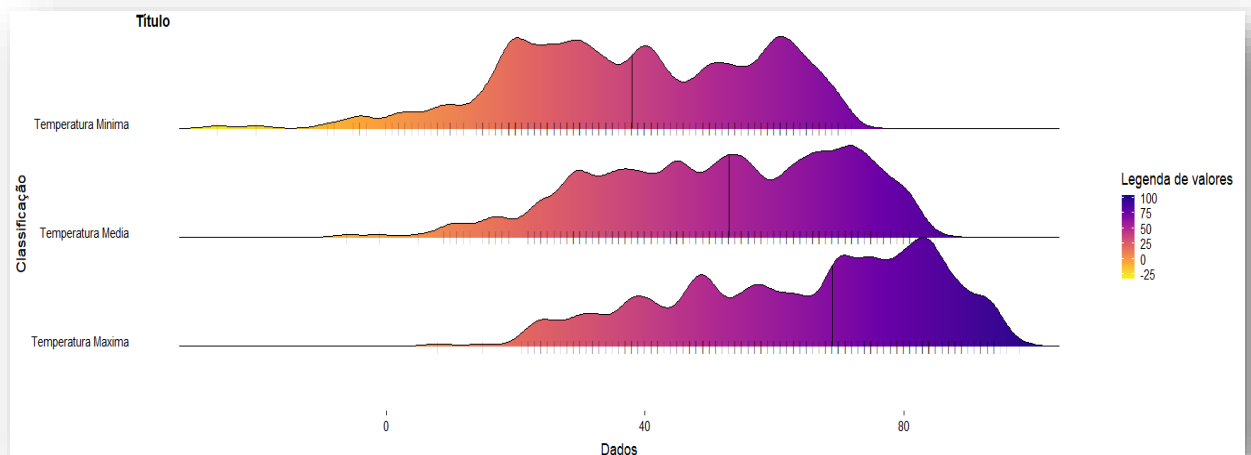
### Mais modificações:

Na configuração do gráfico, o usuário pode escolher em mostrar os quartis e selecionar sua quantidade, modificar a largura de banda padrão do histograma, e modificar a sua altura.

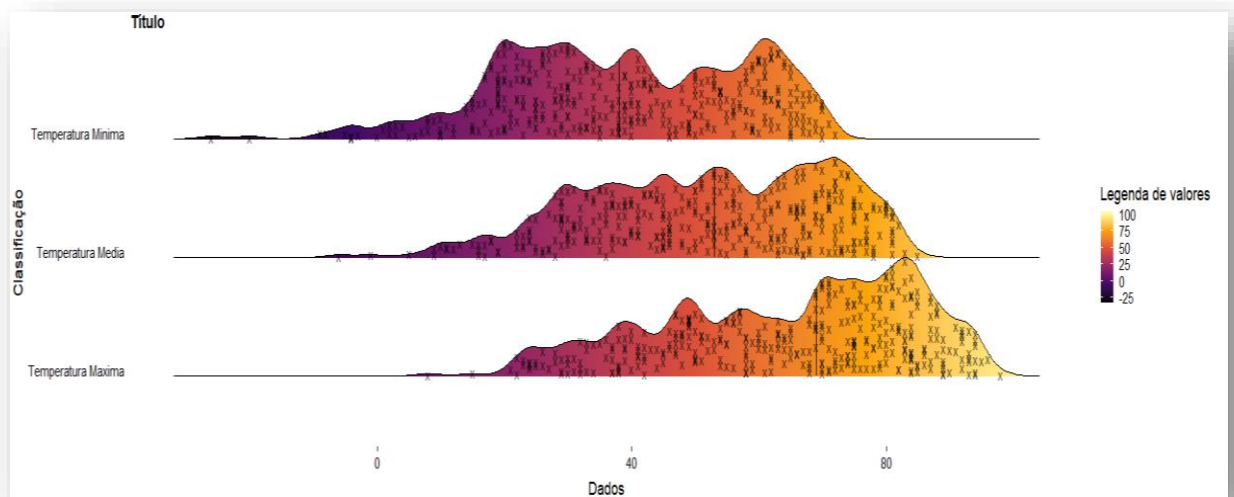
Na configuração dos pontos, o usuário pode escolher em mostrar os pontos, e assim poderá escolher seu tamanho, opacidade, formato (linhas, pontos ou cruz) e sua posição. Entre as posições ele poderá escolher a “identidade”, em que todos os pontos vão estar no  $y = 0$ . “dentro do gráfico”, com os pontos dentro do gráfico, “espalhados”, com os pontos espalhados, e “abaixo do gráfico” com os pontos abaixo do gráfico.

Na configuração das cores do histograma, o usuário pode escolher diversas paletas de cores para o histograma e o sentido da paleta de cores (reverso ou não).

Histograma apenas mostrando a mediana, largura de banda reduzida, altura aumentada, pontos em linhas com opacidade baixa, tamanho aumentado na posição de identidade. Paleta de cores D reversa:



Mesmo histograma, agora com os pontos em formato de cruz, opacidade em 70%, e dentro do gráfico. Paleta de cores C:

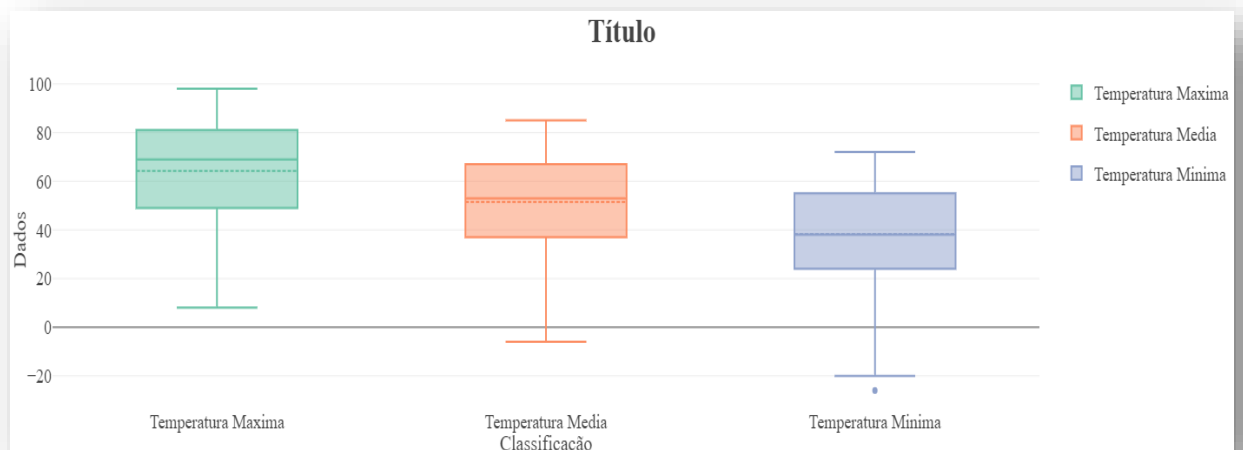


Este gráfico pode demorar mais para ser carregado.

## Box Plot:

O painel Box Plot mostra um gráfico Box Plot, e no eixo X contem as categorias e no eixo Y contendo os Dados.

Gráfico Box Plot sem modificações:



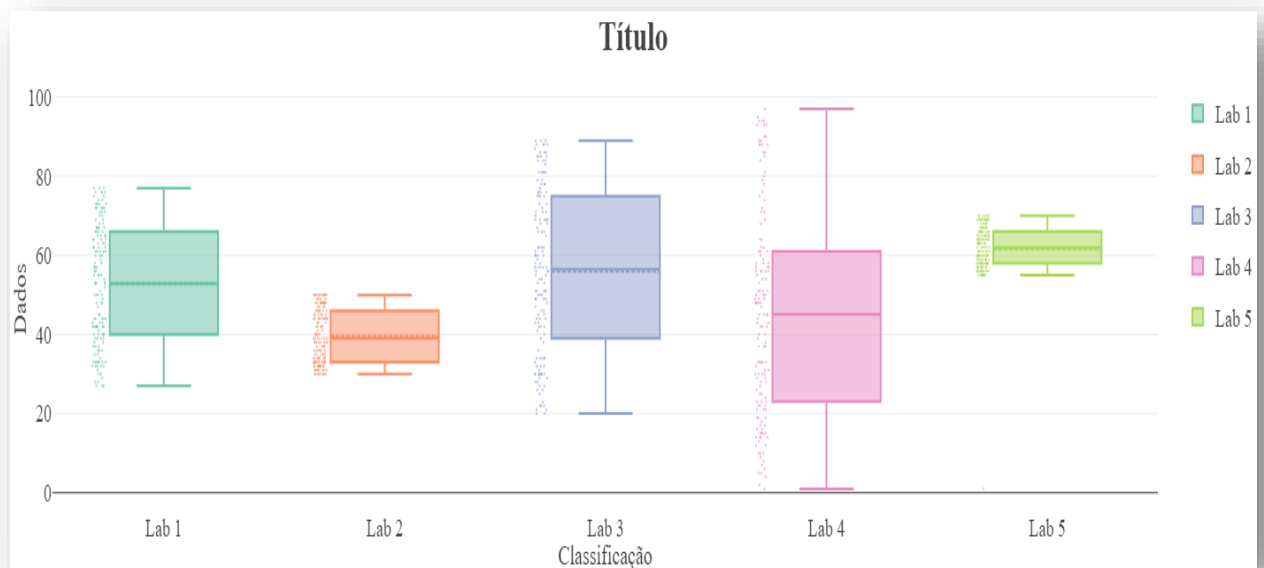
### Mais modificações:

O usuário pode escolher o algoritmo do cálculo dos quartis (Linear, Inclusivo e Exclusivo), e se deseja mostrar uma linha tracejada que é a média de cada um dos dados.

Escolhendo o método “linear”, o programa irá utilizar a interpolação linear, já nos métodos “inclusivo” e “exclusivo” utiliza a mediana para dividir os dados em dois. Caso a quantidade de dados seja ímpar, no inclusivo não inclui a mediana em nenhuma das metades, já no metodo exclusivo inclui a mediana em ambas as metades. E assim o primeiro quarteto é a mediana da primeira metade, e o segundo quarteto é a mediana da segunda metade.

Na configuração dos pontos o usuário pode aumentar ou reduzir a distância entre os pontos, a posição dos pontos conforme cada uma das colunas, o tamanho dos pontos, a largura das linhas dos pontos e a opacidade dos pontos.

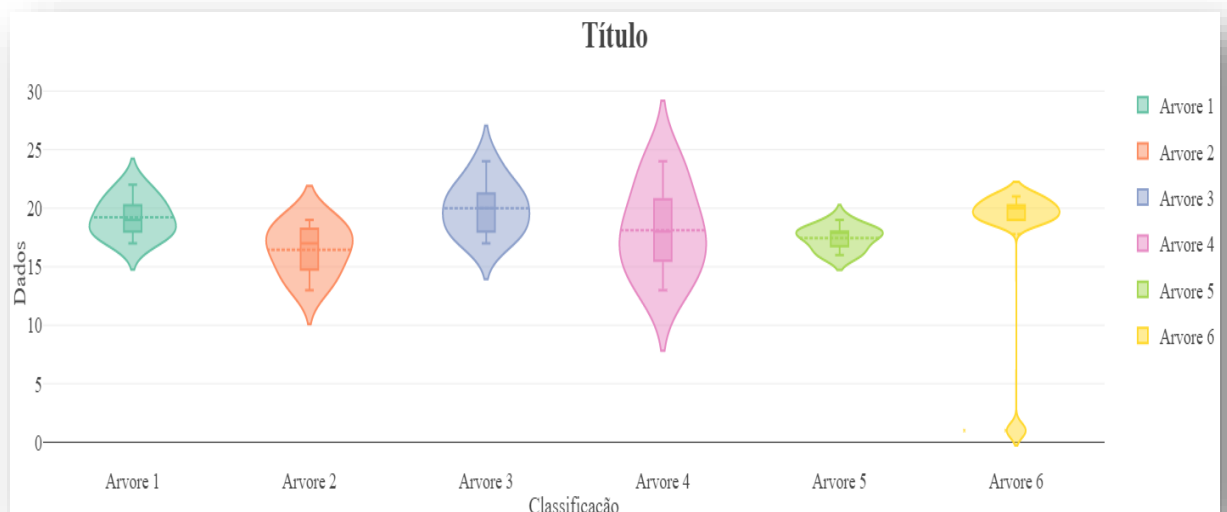
Gráfico com algoritmo de cálculo linear, média visível, com 0.13 de distância entre os pontos, -1.2 da posição dos pontos, 2 de tamanho, 0 de largura das bordas dos pontos, 60% de opacidade com formato de circulo:



## Violino:

O gráfico em violino mostra um gráfico parecido com o gráfico Box Plot porém com as curvas no entorno do gráfico representando a função de densidade. No eixo X contem as categorias e no eixo Y contendo os Dados.

Gráfico em Violino sem modificações:



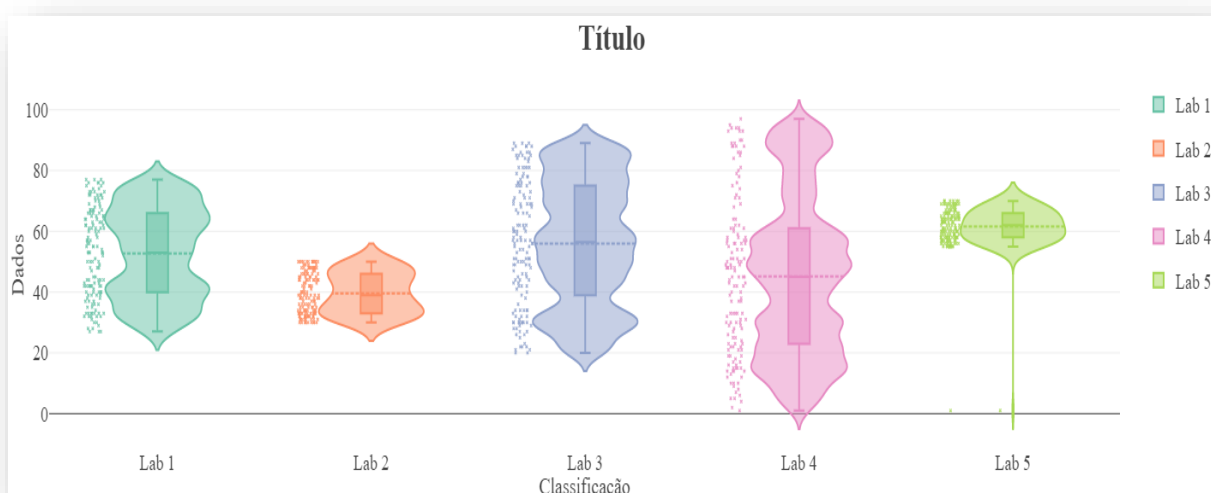
Mais modificações:

O usuário pode escolher o algoritmo do cálculo dos quartis (Linear, Inclusivo e Exclusivo), e se deseja mostrar uma linha tracejada que é a média de cada um dos dados. Ele também pode modificar as curvas de densidade do gráfico ao desativar a largura de banda padrão.

Escolhendo o método “linear”, o programa irá utilizar a interpolação linear, já nos métodos “inclusivo” e “exclusivo” utiliza a mediana para dividir os dados em dois. Caso a quantidade de dados seja ímpar, no inclusivo não inclui a mediana em nenhuma das metades, já no método exclusivo inclui a mediana em ambas as metades. E assim o primeiro quarteto é a mediana da primeira metade, e o segundo quarteto é a mediana da segunda metade.

Na configuração dos pontos o usuário pode aumentar ou reduzir a distância entre os pontos, a posição dos pontos conforme cada uma das colunas, o tamanho dos pontos, a largura das linhas dos pontos e a opacidade dos pontos.

Gráfico com algoritmo de cálculo inclusivo, média visível, largura de banda não padrão (1), com 0.2 de distância entre os pontos, -1.2 da posição dos pontos, 3 de tamanho, 0 de largura das bordas dos pontos, 60% de opacidade com formato de cruz:



**Gráfico de pontos:**

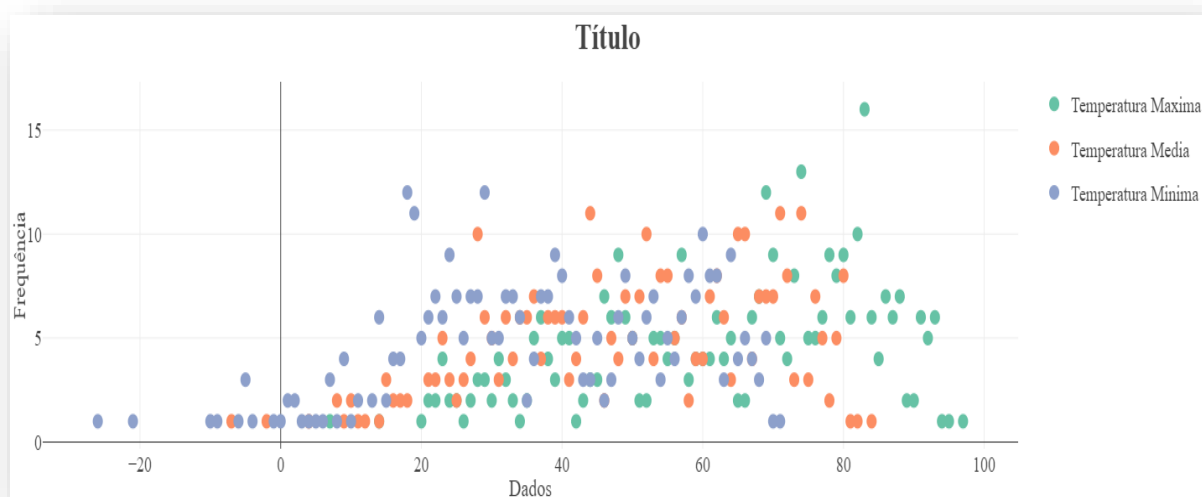
O gráfico de pontos, geralmente quando são dados parecidos, com várias observações e com frequências altas, representam os dados e suas frequências.

O painel do gráfico de pontos possui dois tipos de gráficos, o gráfico de pontos simples e o gráfico de pontos em formato bee swarm.

### Gráfico de pontos Simples:

O gráfico de pontos simples possui os dados no eixo X e as frequências dos dados no eixo Y.

Gráfico de pontos simples sem modificações:



### Mais modificações:

Nas configurações dos pontos, o usuário pode escolher o formato dos pontos, mostrar linhas nas bordas dos pontos, o tamanho dos pontos, a opacidade dos pontos e o intervalo entre os valores (tamanho das bandas).

Nas configurações das linhas, o usuário ao ativar as linhas pode escolher a opacidade das linhas.

Nas configurações das elipses, ao mostrar as elipses pode escolher em mostrar a área da elipse e a sua opacidade, e o intervalo de confiança CI da elipse.

A elipse é calculada sobrepondo os contornos de probabilidade normal sobre um gráfico de dispersão de dados.



Gráfico com pontos com opacidade 0 e intervalo entre valores 1, com linhas com 70% de opacidade e uma elipse com 10% de opacidade e com um intervalo de confiança de 95%:

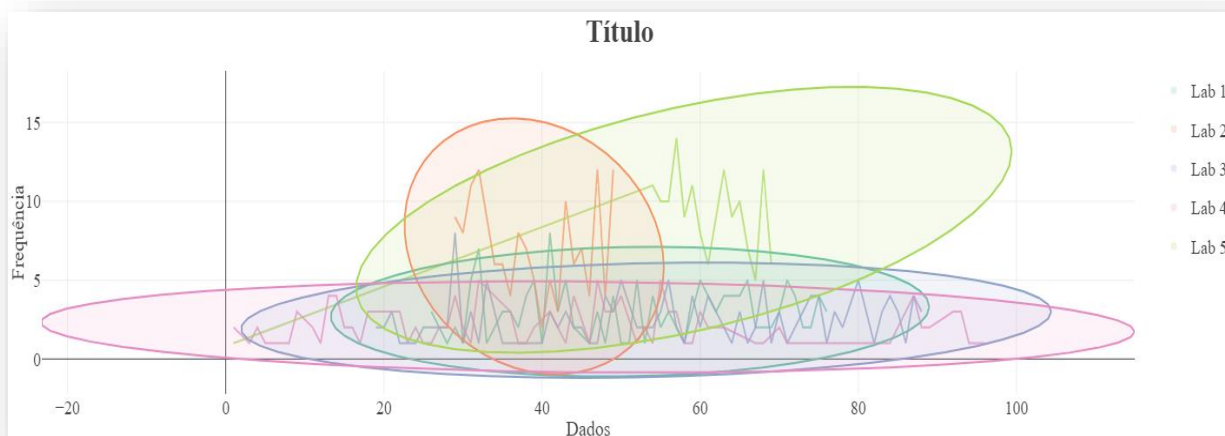
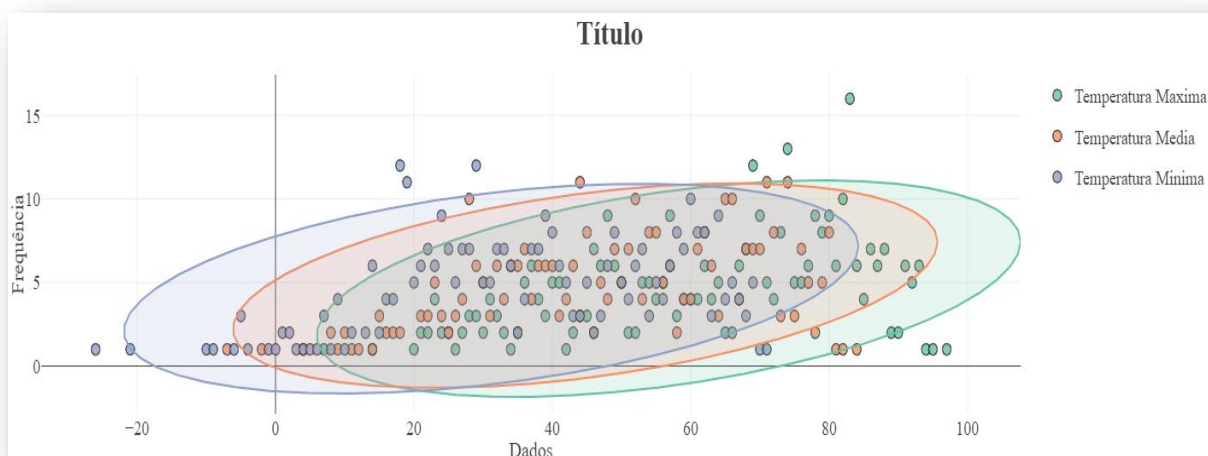


Gráfico com os pontos em formato de círculo, com linhas nos pontos, tamanho 10, opacidade em 80% e intervalo entre valores 1. Sem linhas e com uma elipse com 15% de opacidade e intervalo de confiança de 90%:

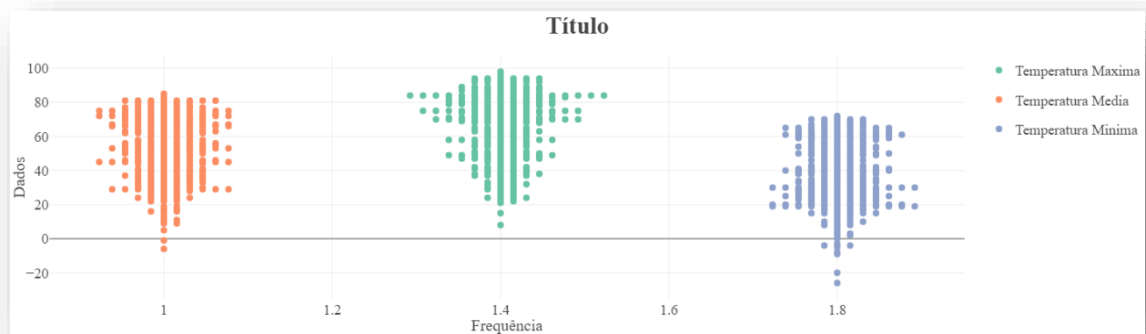


### Gráfico de pontos em bee swarm:

O gráfico de pontos simples possui a dificuldade de visualização quando os pontos, de categorias diferentes, possuem mesma frequência e mesmos valores, ou até na mesma categoria, com frequências e valores muito parecidos.

Este problema pode ser concertado alterando a opacidade, o formato dos pontos, adicionando as bordas dos pontos ou até adicionando linhas entre os pontos. Ou também pode ser concertado utilizando o gráfico em bee swarm.

Gráfico em bee swarm modificando apenas o espaçamento entre os pontos para 10:

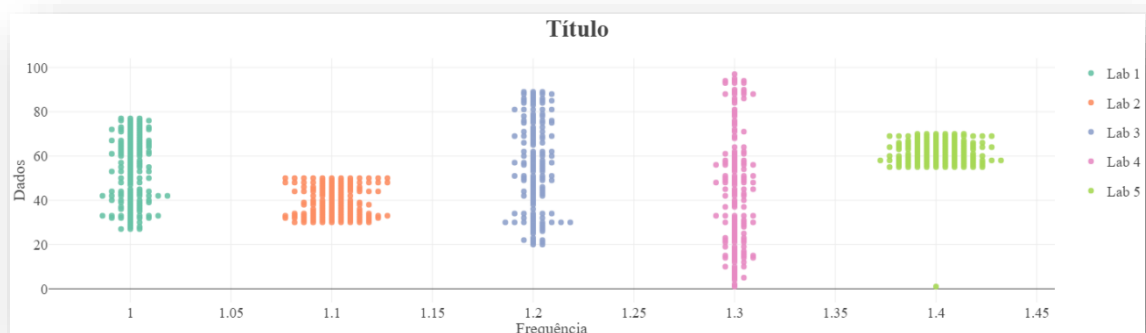


### Mais modificações:

Nas configurações do gráfico o usuário pode modificar o tamanho (escala) do gráfico, o espaçamento entre os pontos, se deseja que o gráfico fique de lado, o método de espalhamento dos pontos, e a prioridade de encaixe dos pontos.

Nas configurações dos pontos, o usuário pode modificar o formato dos pontos (círculos, quadrados, triângulos ou formas diversas).

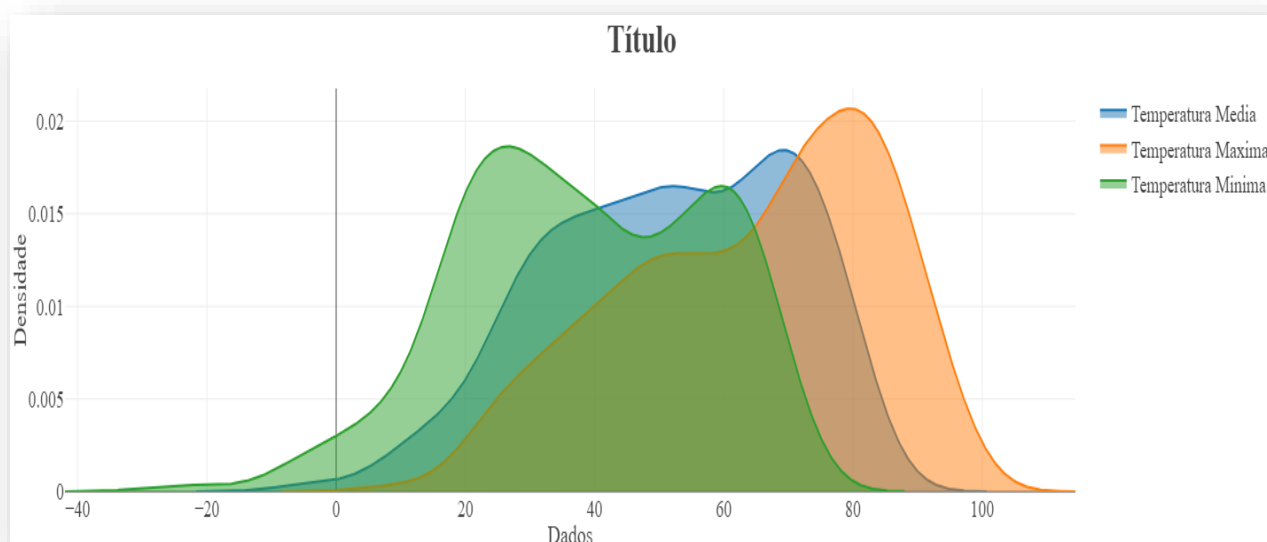
Gráfico com tamanho 1, espaçamento entre os pontos 4, espaçamento entre os gráficos 0.1, com o método "swarm" e com a prioridade "crescente":



## Gráfico de Densidade:

O gráfico de densidade faz um cálculo de densidade, com o eixo X sendo os dados e o eixo Y a densidade dos dados.

Gráfico de densidade sem modificações:



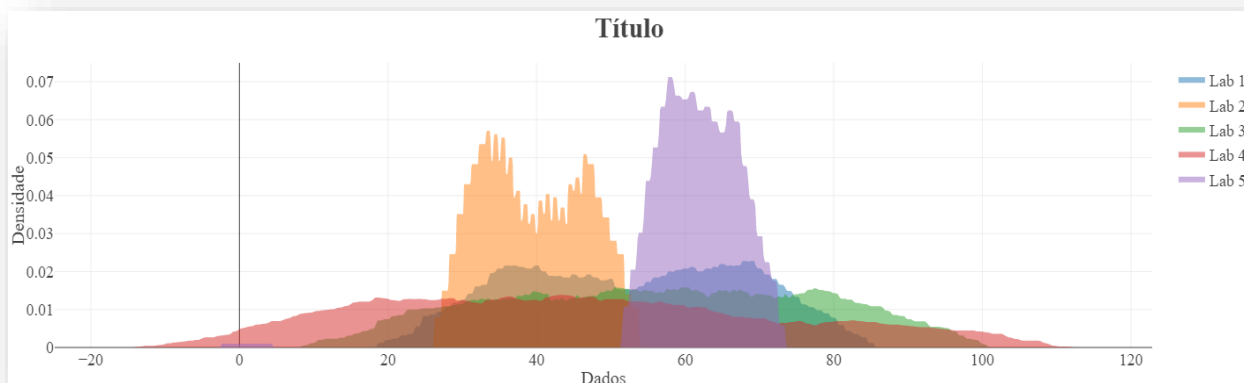
Mais modificações:

As configurações do gráfico de densidade permitem a adição da área ou das linhas do gráfico de densidade, ou a escolha de uma estimativa de densidade kernel (EDK) dentre gaussiana, retangular, triangular, Epanechnikov, biweight, cosseno e optcosseno, sendo o padrão gaussiana.

A estimativa de densidade kernel é uma forma não-paramétrica para estimar a Função densidade de probabilidade (FDP) de uma variável aleatória. Para mais informações:

[https://pt.wikipedia.org/wiki/Estimativa\\_de\\_densidade\\_kernel](https://pt.wikipedia.org/wiki/Estimativa_de_densidade_kernel)

Gráfico mostrando a área, porém não as linhas, com a estimativa de densidade de kernel igual a “Retangular”:

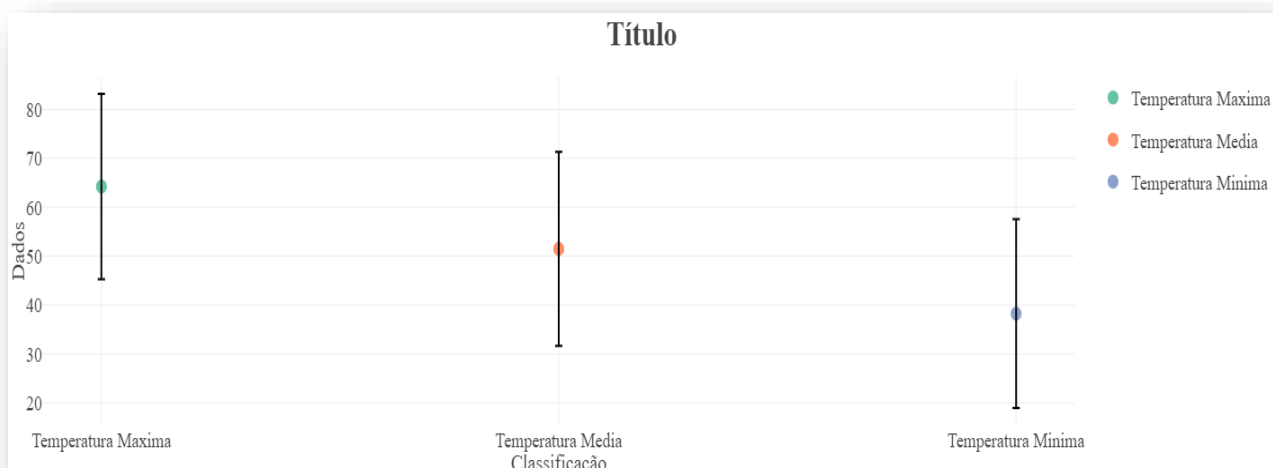


## Gráfico de Erro:

O gráfico de erro é um gráfico que o eixo X são as colunas dos dados, e o eixo Y são as médias de cada um dos dados e os seus respectivos erros.

A linha de erro é realizada com o desvio padrão, e o usuário consegue modificar no painel de controle de opções.

Gráfico de erro sem modificações:



## Mais modificações:

Nas configurações do gráfico de erro, o usuário pode remover a linha de erro, modificar a opacidade da linha de erro e alterar o algoritmo para o cálculo de erro (Desvio padrão, Erro padrão, ou selecionar um intervalo de confiança).

O Desvio Padrão é uma medida que indica a dispersão dos dados dentro de uma amostra com relação à média calculada pela raiz quadrada do somatório de todas as observações subtraídas pela média das observações ao quadrado, dividido pelo número de observações.

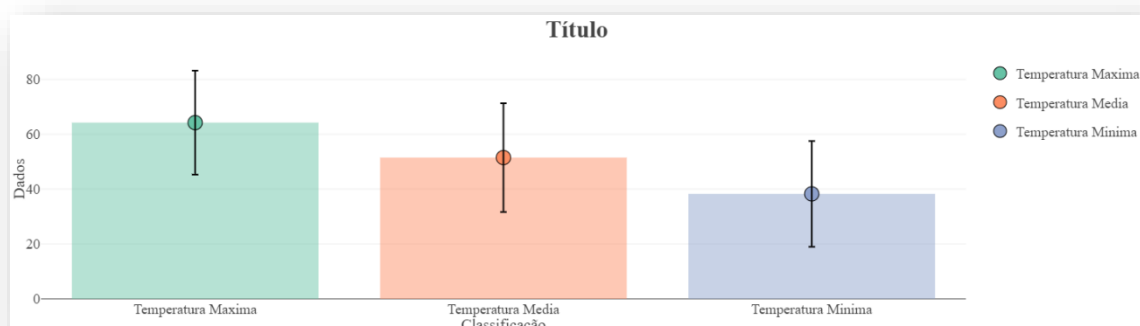
O erro padrão é uma medida de variação de uma média amostral em relação à média da população calculada dividindo o desvio padrão pela raiz quadrada do número de observações.

Para calcular o intervalo de confiança basta multiplicar o erro padrão pelo percentil associado ao nível de significância observado em uma distribuição normal padrão, ou seja, que apresenta média 0 e desvio-padrão igual a 1.

Nas configurações das barras, o usuário pode adicionar uma barra e alterar a sua opacidade.

Nas configurações dos pontos o usuário pode modificar o formato dos pontos (Círculos, Quadrados, Triângulos ou formas diversas), adicionar uma linha de borda nos pontos, alterar o tamanho dos pontos e a sua opacidade.

Gráfico com linha de erro, opacidade em 100%, algoritmo utilizado: desvio padrão, com barras com 60% de opacidade, pontos no formato de círculos, com linhas de borda nos pontos, tamanho 16 e opacidade 100%



## Gráficos 3D

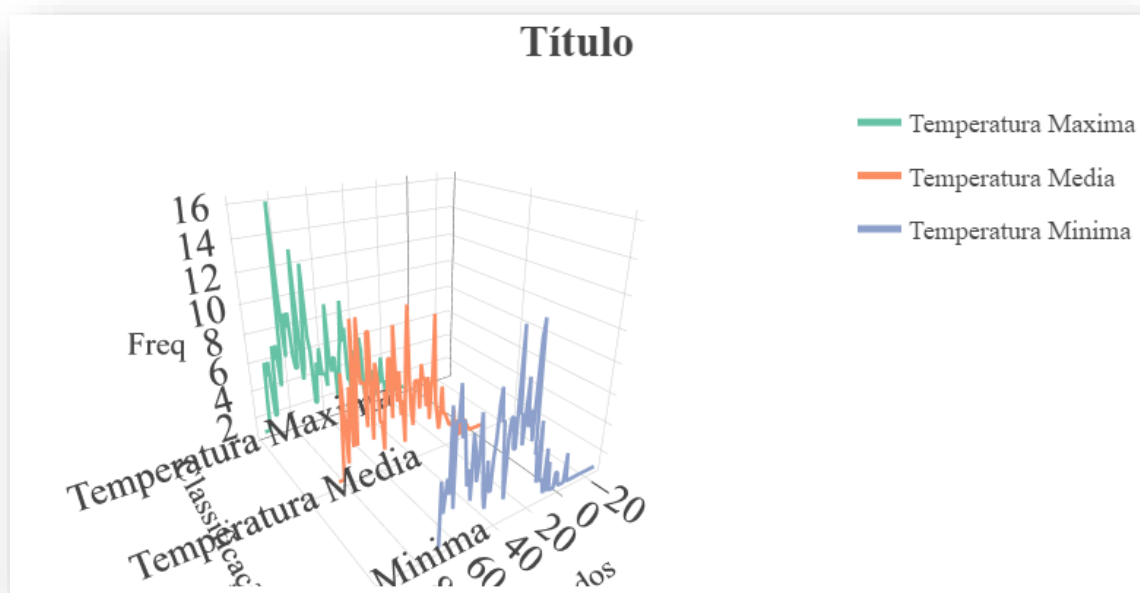
Os gráficos 2D podem ser encontrados na barra lateral esquerda após carregar os dados no aplicativo.

Os gráficos 3D podem ser controlados pelo usuário, podendo rotacionar utilizando o botão esquerdo do mouse, e ser deslocado utilizando o botão direito do mouse. No canto superior direito, existem duas opções de rotação, turntable rotation que o usuário pode rotacionar o gráfico em relação à um dos eixos (eixo z), e orbital rotation que o usuário pode fazer uma rotação orbital.

### Histograma 3D:

O histograma 3D, possui este formato de linhas representando os dados sobre sua frequência, podendo modificar o intervalo de frequências no painel de controle.

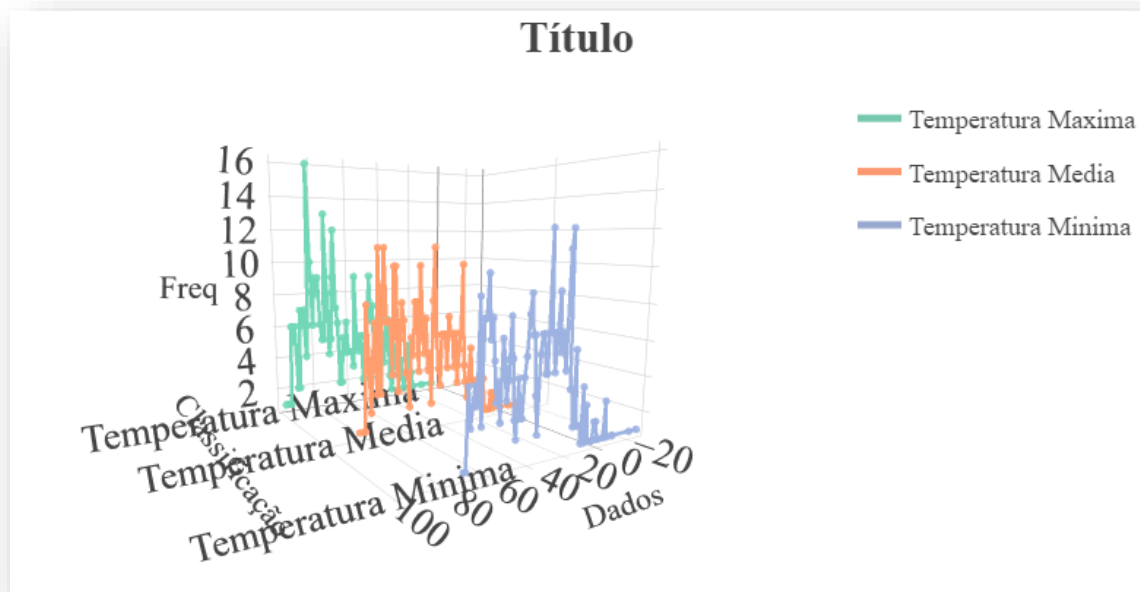
#### Histograma 3D sem modificações:



#### Mais modificações:

Na configuração do gráfico, o usuário pode modificar a quantidade de intervalos de frequência, nas configurações de linha o usuário pode modificar o comprimento das linhas e a sua opacidade, e nas configurações de pontos o usuário pode determinar o tamanho dos pontos caso deseje mostrá-los.

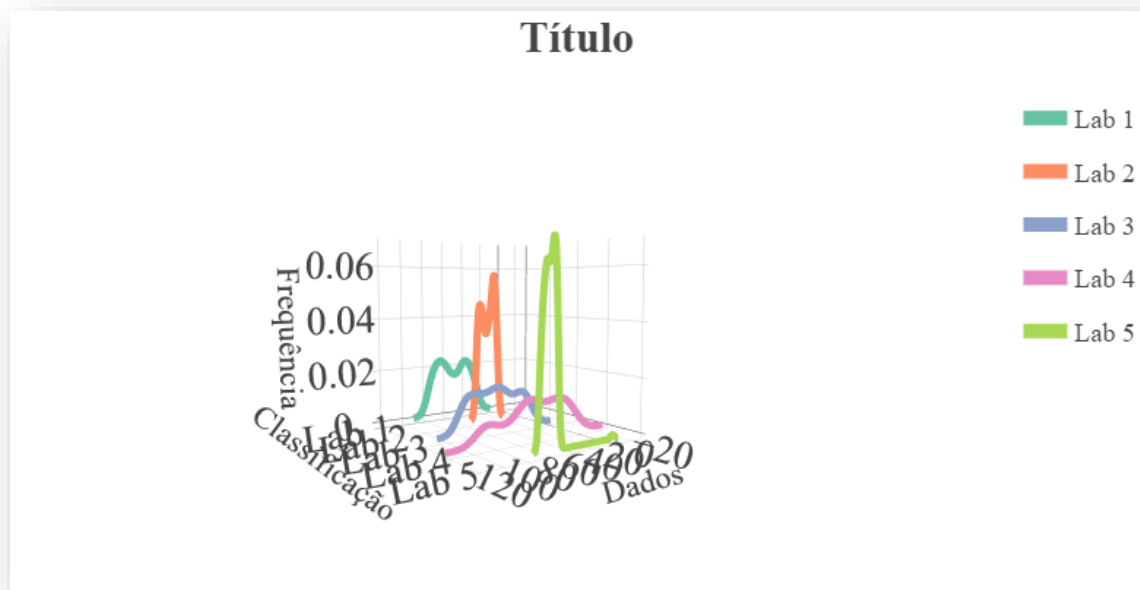
Gráfico com comprimento de linha 5, opacidade em 90% e com pontos no tamanho 3.



### Gráfico de Densidade:

O gráfico de densidade faz um cálculo de densidade, com o eixo X sendo os dados e o eixo Z a densidade dos dados, com as classificações no eixo Y.

Gráfico de densidade sem modificações:



### Mais modificações:

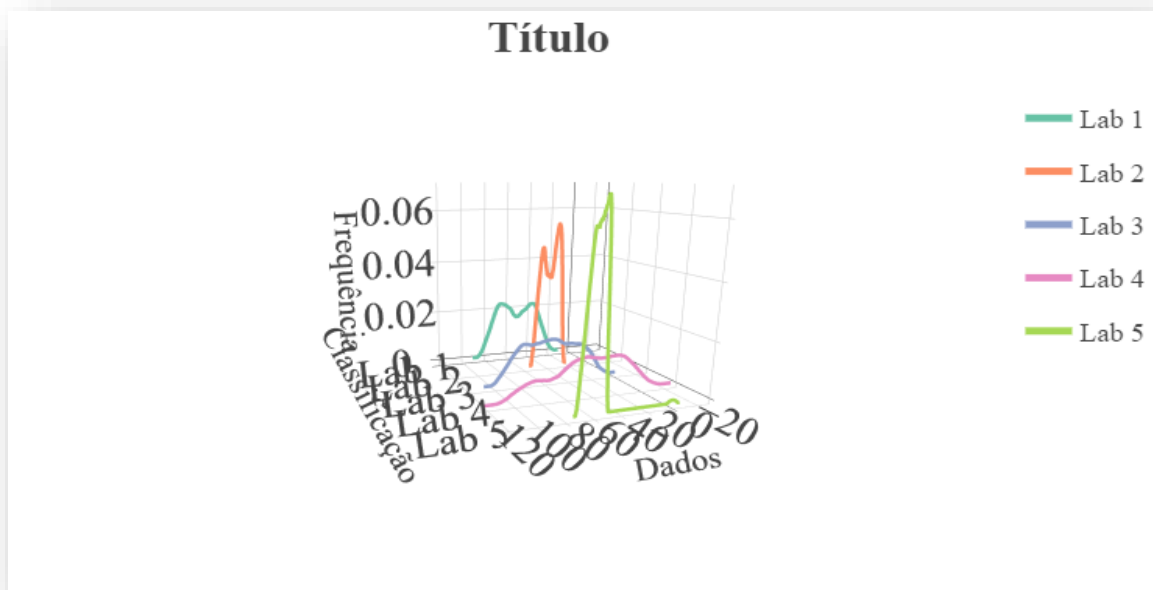
As configurações do gráfico de densidade permitem modificação da largura das linhas ou a escolha de uma estimativa de densidade kernel (EDK) dentre gaussiana, retangular, triangular, Epanechnikov, biweight, cosseno e optcosseno, sendo o padrão gaussiana.

A estimativa de densidade kernel é uma forma não-paramétrica para estimar a Função densidade de probabilidade (FDP) de uma variável aleatória. Para mais informações:

[https://pt.wikipedia.org/wiki/Estimativa\\_de\\_densidade\\_kernel](https://pt.wikipedia.org/wiki/Estimativa_de_densidade_kernel)

Gráfico com as linhas com largura 5, com a estimativa de densidade de kernel igual a “Epanechnikov”:





## Gráfico de pontos:

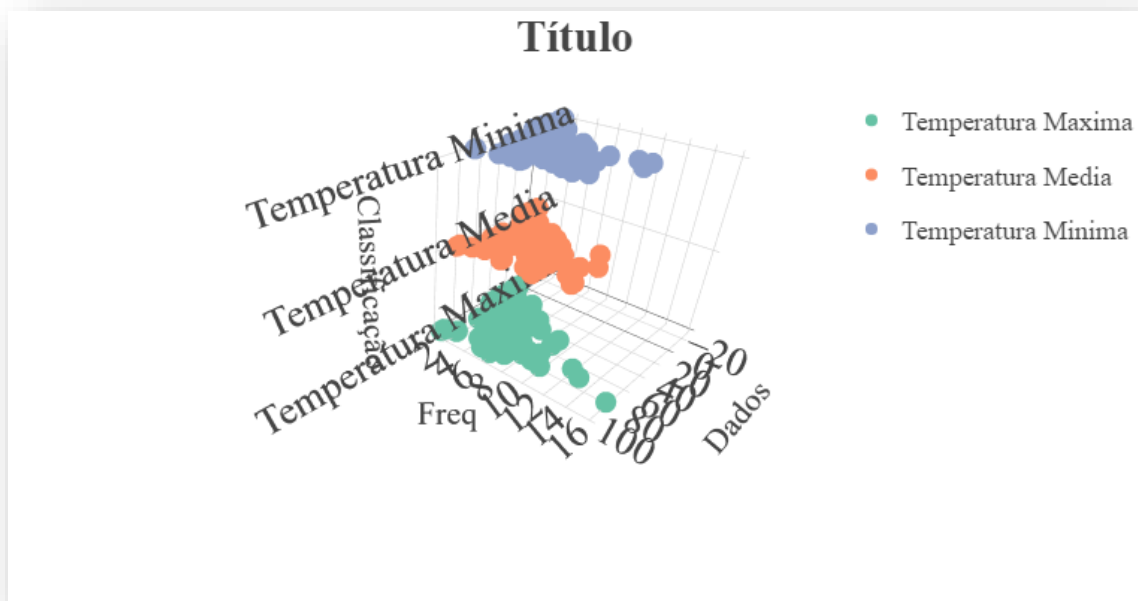
O gráfico de pontos, geralmente quando são dados parecidos, com várias observações e com frequências altas, representam os dados e suas frequências.

O painel do gráfico de pontos possui dois tipos de gráficos, o gráfico de pontos simples e o gráfico de pontos em formato bee swarm.

## Gráfico de pontos 3D Simples:

O gráfico de pontos simples possui os dados no eixo X, as frequências dos dados no eixo Z, e cada uma das classificações no eixo Y.

Gráfico de pontos 3D simples sem modificações:



### Mais modificações:

Nas configurações dos pontos, o usuário pode escolher o tamanho dos pontos, a opacidade dos pontos, adicionar uma linha de borda preta nos pontos e o intervalo entre os valores (tamanho das bandas).

Nas configurações das elipses, ao mostrar as elipses pode escolher em mostrar a área da elipse e a sua opacidade, o contorno da elipse (contorno preto), sua largura e sua opacidade e o intervalo de confiança CI da elipse.

A elipse é calculada sobrepondo os contornos de probabilidade normal sobre um gráfico de dispersão de dados.

Gráfico com pontos com opacidade em 100% e intervalo entre valores 1, com linha de borda, e uma elipse sem mostrar o contorno da elipse e mostrando a área com 75% de opacidade e com um intervalo de confiança de 95%:

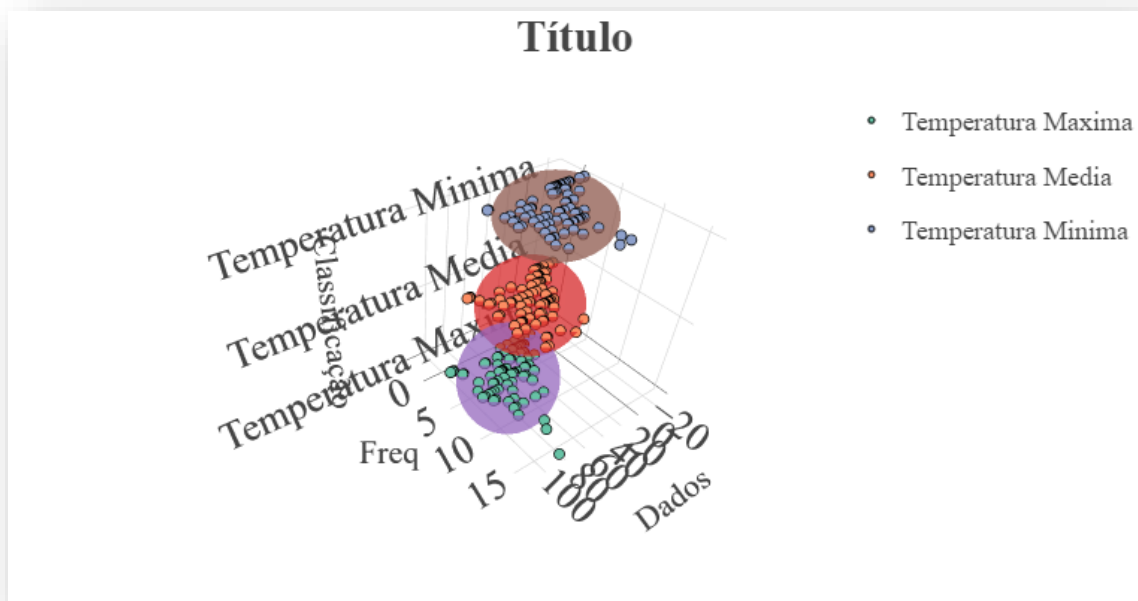
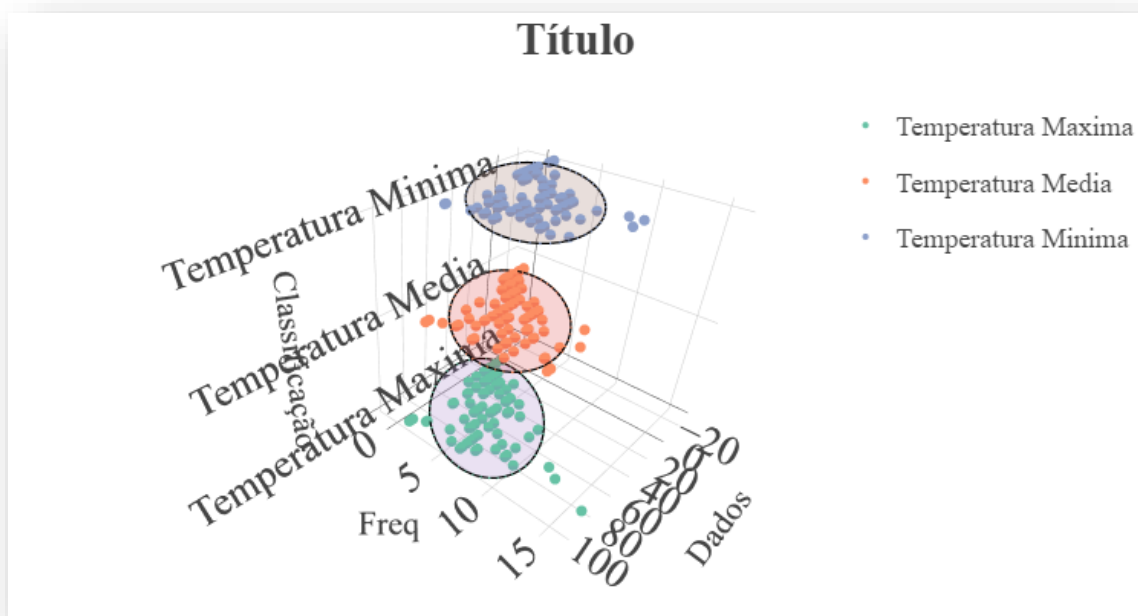


Gráfico com pontos com opacidade em 100% e intervalo entre valores 1, sem linha de borda, e uma elipse mostrando o contorno da elipse, com opacidade em 100% e com contorno igual a 2,5 e mostrando a área com 20% de opacidade e com um intervalo de confiança de 87%:

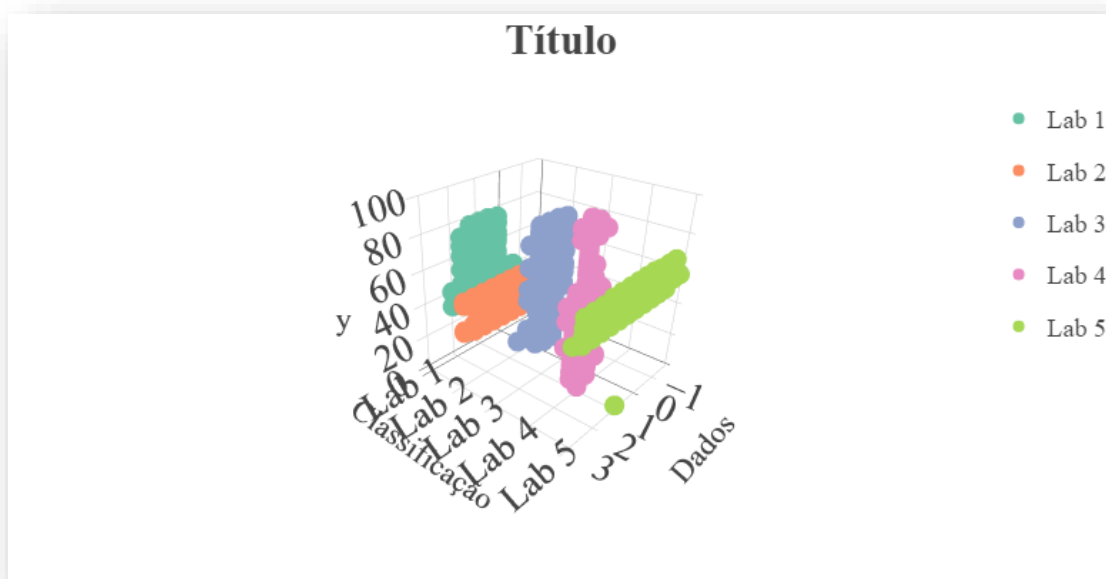


**Gráfico de pontos 3D em Bee swarm:**

O gráfico de pontos simples possui a dificuldade de visualização quando os pontos, de categorias diferentes, possuem mesma frequência e mesmos valores, ou até na mesma categoria, com frequências e valores muito parecidos.

Este problema pode ser concertado alterando a opacidade, o formato dos pontos, adicionando as bordas dos pontos ou até adicionando linhas entre os pontos. Ou também pode ser concertado utilizando o gráfico em bee swarm.

Gráfico de bee swarm sem modificações:

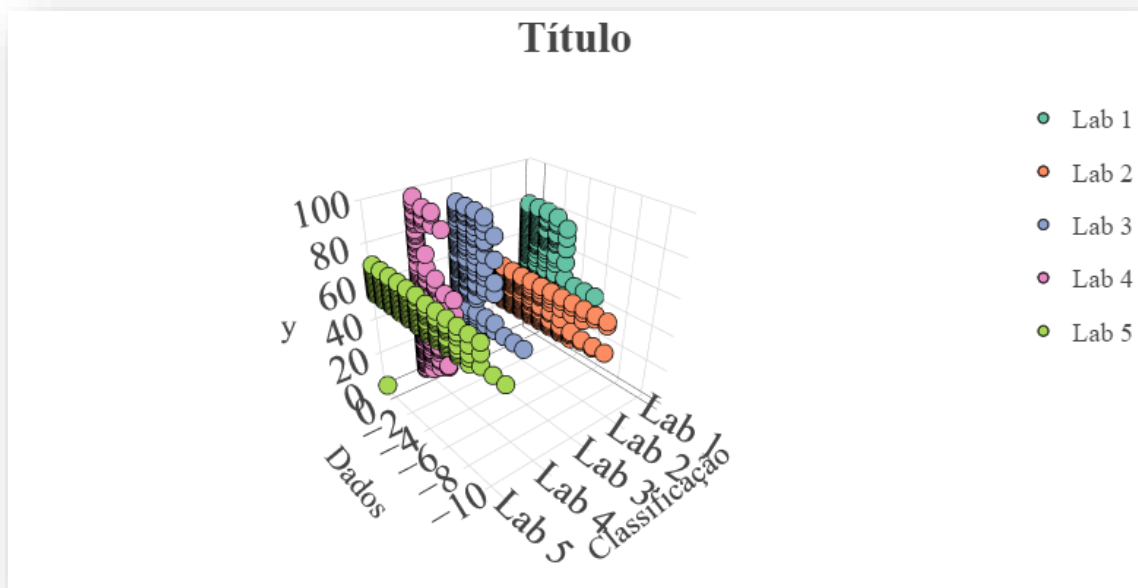


### Mais modificações:

Nas configurações do gráfico o usuário pode modificar o espaçamento entre os pontos, se deseja que o gráfico fique de lado, o método de espalhamento dos pontos, e a prioridade de encaixe dos pontos.

Nas configurações dos pontos o usuário pode adicionar linhas nos pontos, escolher o tamanho dos pontos e escolher a opacidade deles

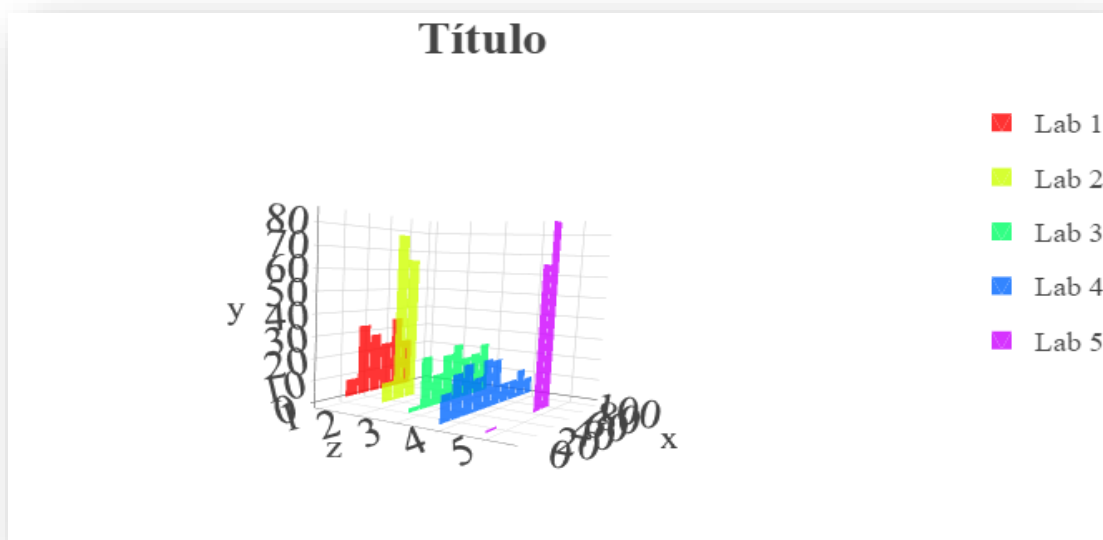
Gráfico com 10 de espaçamento entre os pontos, de lado, com método Swarm na prioridade Crescente, pontos com bordas pretas, com tamanho igual a 7 e opacidade em 100%



## Gráfico de Barras:

O gráfico de barras é um gráfico de histograma em barras com os dados no eixo X, as Frequências no eixo Y e as classificações no eixo Z. O gráfico de barras utiliza o mesmo algoritmo que os histogramas.

Gráfico de barras sem modificações:



Mais modificações:

Nas configurações dos algoritmos para o cálculo do número de intervalos, pode se escolher o “Sturges” (padrão), “Scott”, e o “Freedman-Diaconis”.

Para saber mais sobre estes algoritmos (Contém os 3 algoritmos na página):

[https://pt.wikipedia.org/wiki/Histograma#Escolha\\_do\\_n%C3%BAmero\\_de\\_barras\\_pela\\_f%C3%B3rmula\\_de\\_Sturges](https://pt.wikipedia.org/wiki/Histograma#Escolha_do_n%C3%BAmero_de_barras_pela_f%C3%B3rmula_de_Sturges)

Nas configurações das barras o usuário pode modificar o espaçamento entre as barras e a opacidade das barras.

Gráfico utilizando o algoritmo de Scott, com 1 de espaçamento entre as barras e com opacidade em 60%:

