

Shiny Data Visualization

Tutorial

Índice

Introdução

Inicialização

- Sobre o aplicativo
- Salvar imagens
- Layout
- Resumo dos dados

Dados Unidimensionais

Inserção dos dados

- Importar os dados
- Inserir os dados na planilha

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

Checando os dados

- Distribuição de dados normais
- Homogeneidade das variâncias
- Avaliando a esfericidade
- Transformando os dados para a normalidade

Dados Bidimensionais

Inserção dos Dados

- Importar os dados.
- Inserir os dados na planilha

Comparando duas médias

- Teste – T
- Teste Wilcoxon
- Teste do sinal (Sign test)

Comparando múltiplas médias

- ANCOVA
- ANOVA
- ANOVA misturada
- MANOVA

Dados Tridimensionais

Inserção dos dados

- Importar os dados

Gráficos em Mesh

Referências

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 também, a primeira linha são os nomes das colunas e as restantes são os dados.

O aplicativo aceita **3** tipos de dados. Dados **unidimensionais**, **bidimensionais** e **tridimensionais**

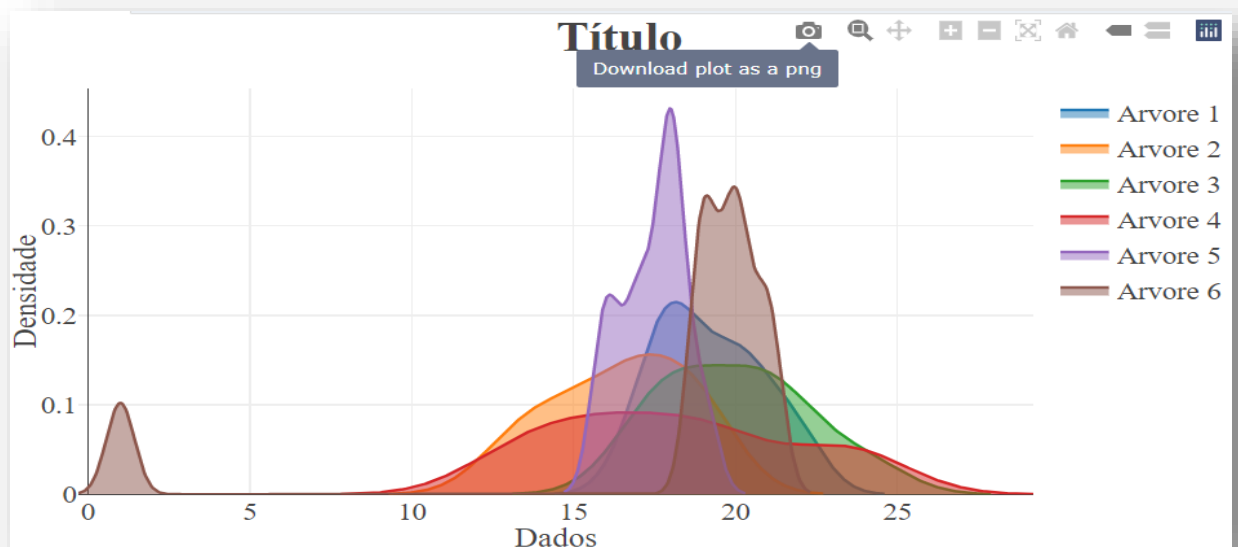
Cada um dos gráficos pode ser **personalizado**, modificando cores, estilos, tamanhos e algoritmos para o cálculo dos gráficos, dependendo de cada um dos gráficos. Para esta função existe uma barra de controle de opções na lateral, específica para cada um dos gráficos.

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

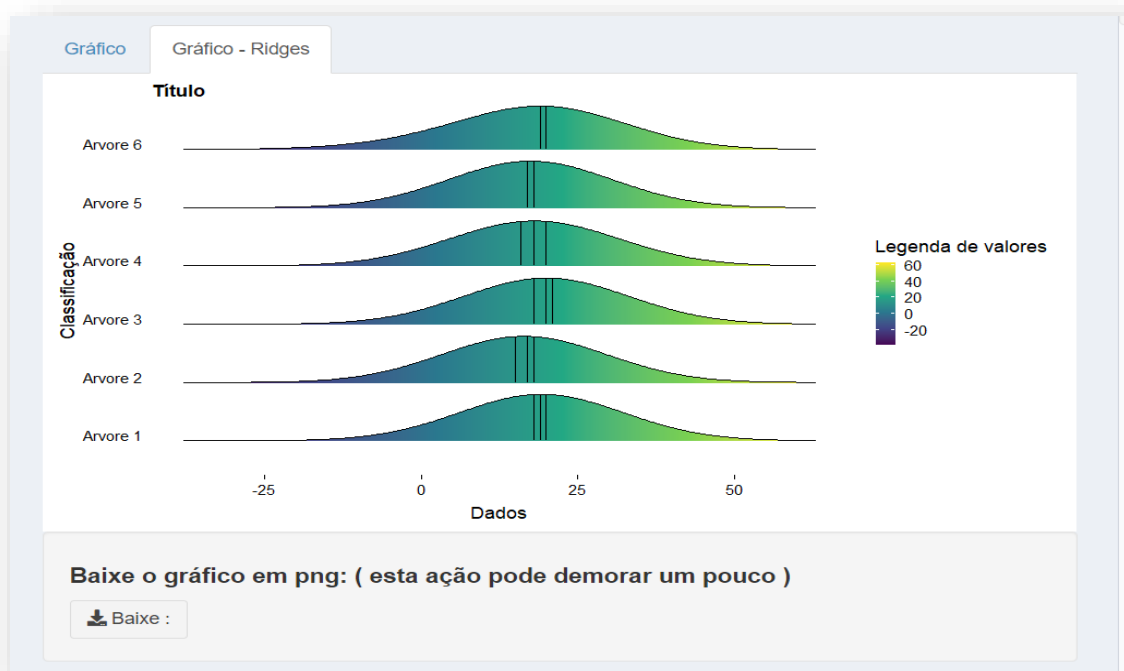
Salvar imagens:

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:



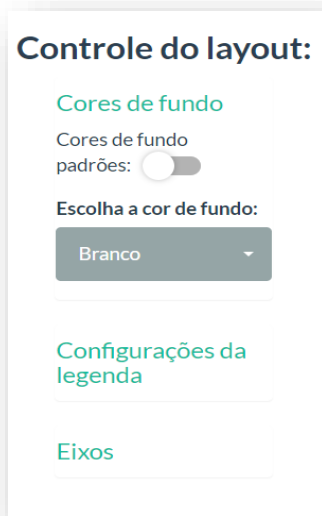
Já para gráficos que **não utilizam o plotly** como ferramenta de visualização, terá um botão embaixo do gráfico escrito “**Baixe o gráfico em png:**” como no exemplo abaixo:



Layout:

Todos os gráficos visuais possuem uma aba, geralmente embaixo das configurações de Controle de opções, chamada **Controle de Layout**.

Nesta aba o usuário pode personalizar um gráfico específico modificando as legendas, as cores e os eixos.



Dados Unidimensionais

A barra de navegação de dados unidimensionais serve para você inserir um ou mais dados unidimensionais, podendo verificar gráficos, distribuições, testes e outros aspectos.

Para os dados unidimensionais, você deve inserir os dados como mostra na página

Inserção dos **dados.**

Com os dados já inseridos, a barra de navegação possui quatro funcionalidades, mostrar um resumo com algumas informações sobre os dados inseridos, um painel contendo gráficos 2D, um painel contendo gráficos 3D e um painel checando as propriedades do gráfico.

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 3D**, **Gráfico em Superfície**, **Gráfico de pontos 3D**, **Gráfico de Densidade 3D** e **Gráfico de Barras 3D**.

Na parte de checagem de dados, existe a checagem se os dados seguem uma **distribuição normal**, a **homogeneidade das variações**,

avaliando a esfericidade e transformando os dados recebidos para uma distribuição normal.

Inserção dos dados

Importar os dados:

Existem duas opções para a **importação** dos dados pelo usuário, dados de exemplo que já vem baixados com o aplicativo e inserir dados de uma planilha Excel.

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

Para os dados exemplares você deve selecionar o painel:

Dados Unidimensionais -> Carregue seus dados -> Importe seus dados -> Usar um arquivo exemplo

Existem 6 opções de exemplos: “**Obs de Laboratório**” que são dados inteiros e aleatórios, cada um com a aleatoriedade começando de um certo valor e terminando em outro.

“**Distribuições aleatórias**” são diversas distribuições com dados aleatórios, possui a distribuição normal, de Poisson, Binomial com 50% para cada. Exponencial e de Chi-Quadrado.

“**Votações na Florida**” mostra o número de pessoas que votaram para cada um dos candidatos a presidente no ano de 2000. Cada uma das observações é de uma determinada cidade da Flórida.

“**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.

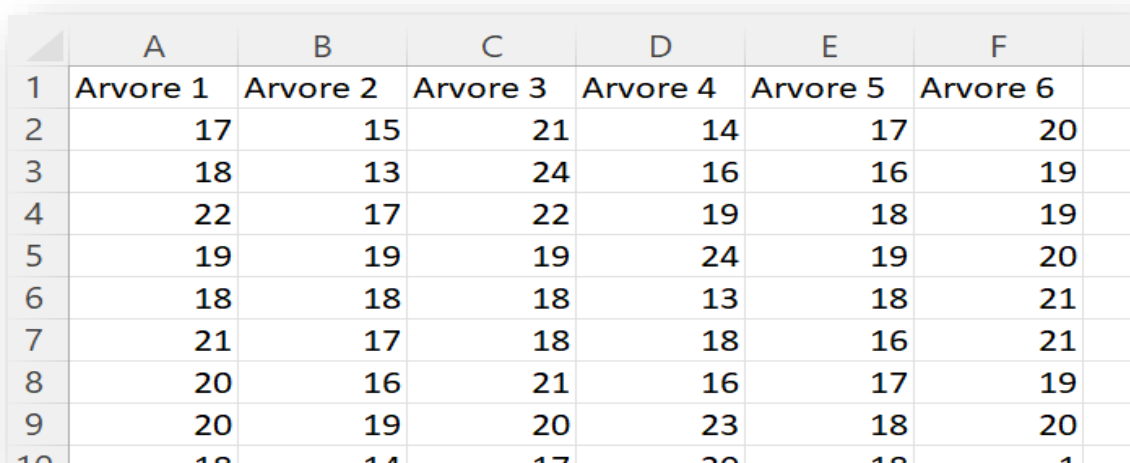
“**Temperatura Lincoln**”, que mostra 3 categorias. A média das temperaturas mínimas, a média das temperaturas médias, e a média das temperaturas máximas de cada mês.

“**demografia – Midwest**” mostra a demografia étnica dos povos no centro-oeste dos Estados Unidos.

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

Dados Unidimensionais -> Carregue seus dados -> Importe 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	

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:

Dados Unidimensionais -> Carregue 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		

De novo, a primeira linha é o nome das colunas e as demais são os seus respectivos dados.

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.

Resumo dos dados:

Após carregar os dados irá surgir um painel escrito “Informações gerais dos dados” na aba de Dados Unidimensionais, e neste painel contém 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:				
	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

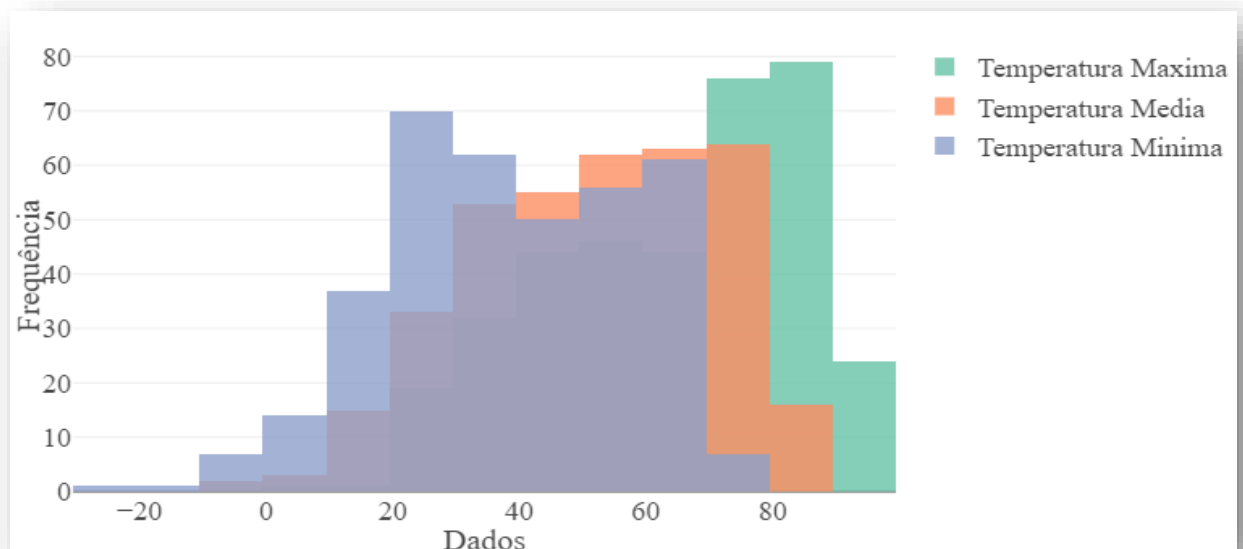
Gráficos 2D

Os gráficos 2D podem ser encontrados na aba de Dados Unidimensionais, após carregar os dados no aplicativo.

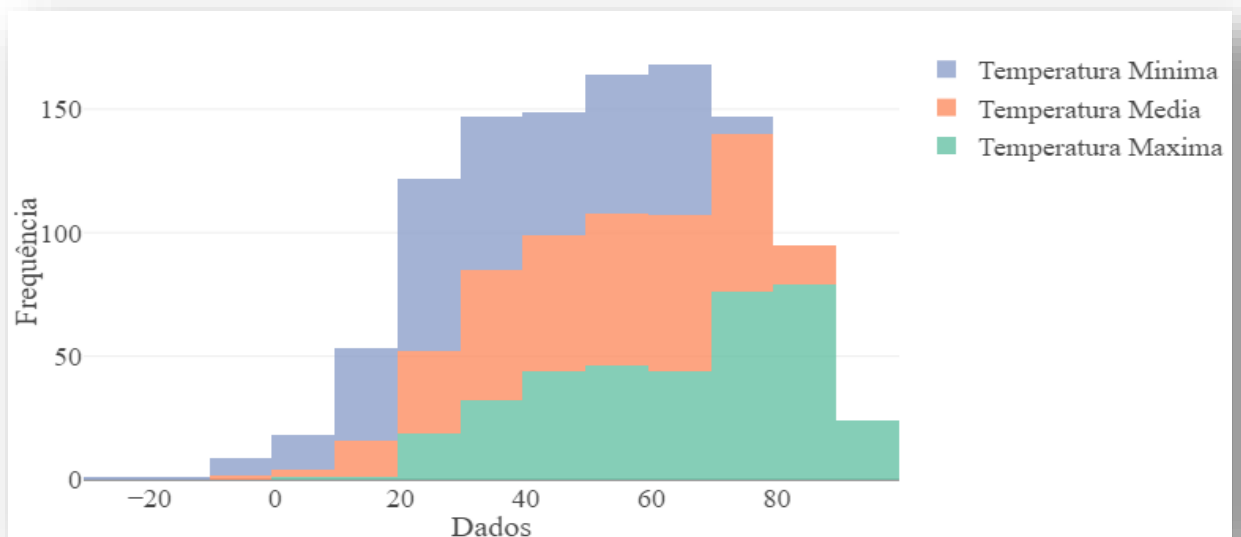
Histograma:

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.

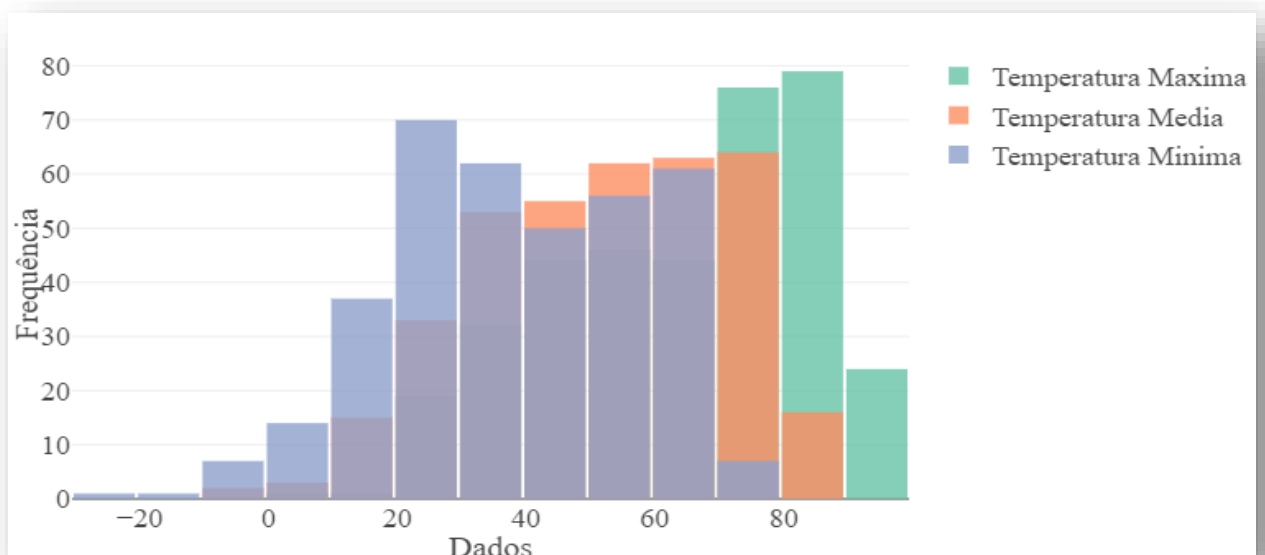
Histograma sem modificações:



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.

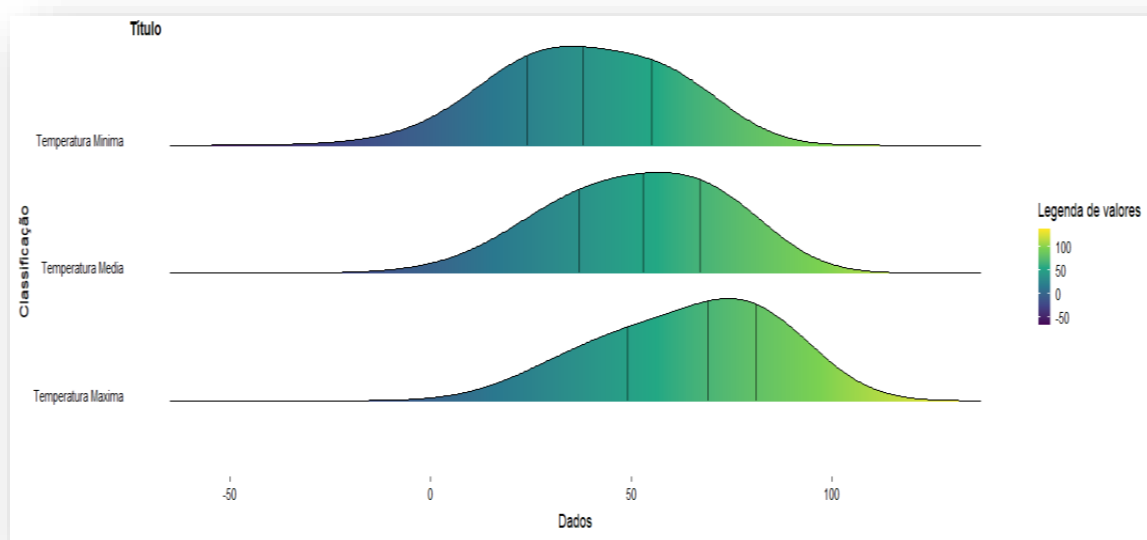
O intervalo padrão do histograma é calculado utilizando a regra de Sturges

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



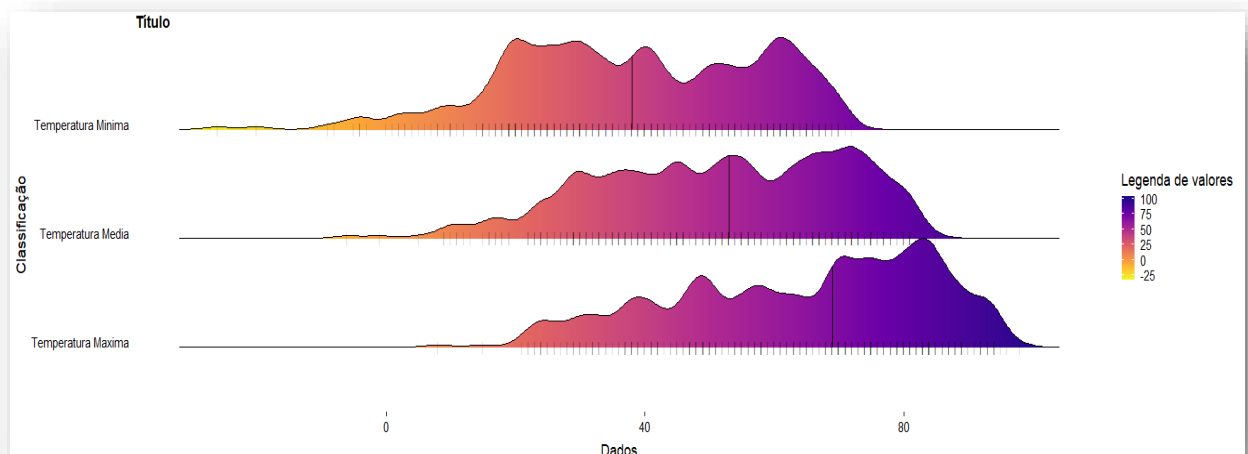
Mais modificações:

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.

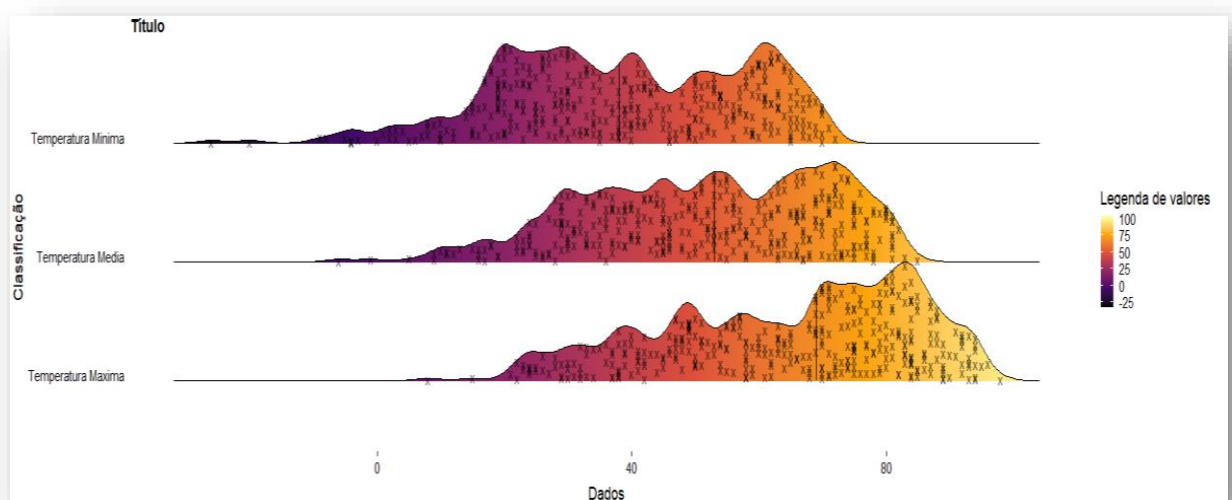
Na configuração dos pontos, o usuário pode escolher em **mostrar os pontos**, e assim poderá escolher seu tamanho, opacidade, formato (linhas, pontos ou cruz) e sua posição. Entre as posições ele poderá escolher a “identidade”, em que todos os pontos vão estar no $y = 0$. “dentro do gráfico”, com os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

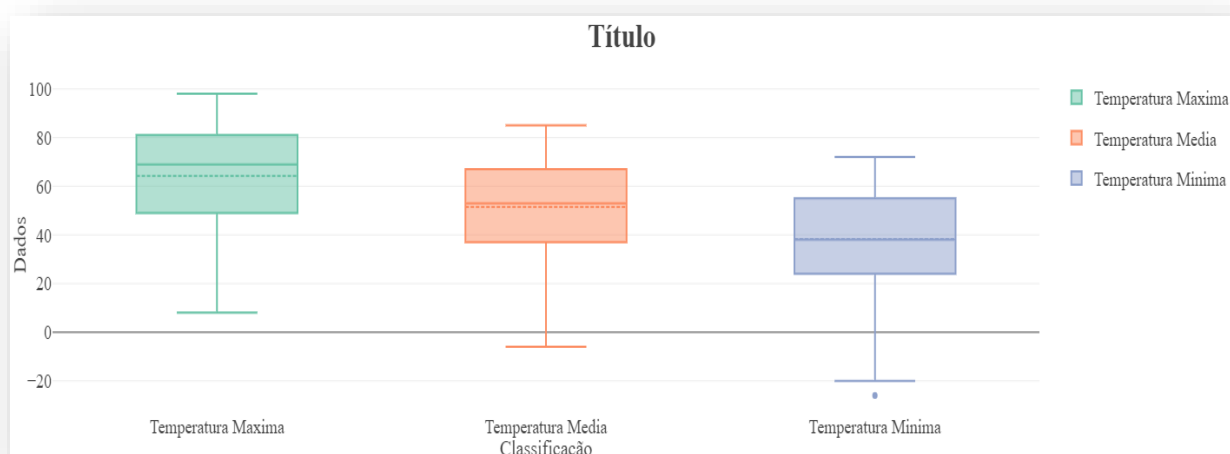


Este gráfico pode demorar mais para ser carregado.

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:



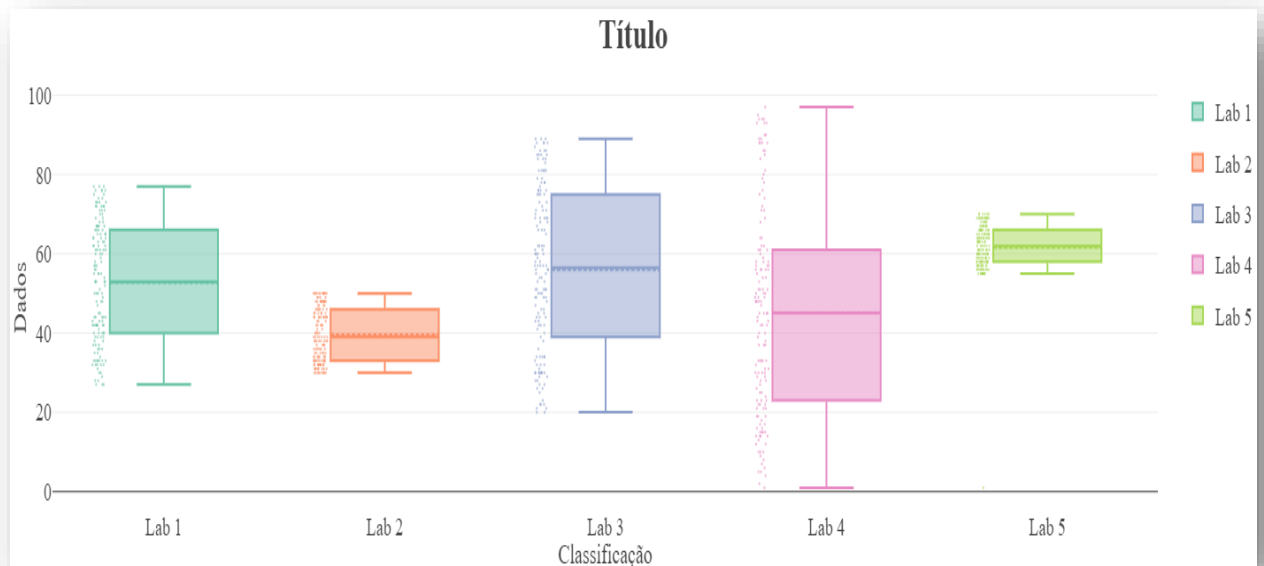
Mais modificações:

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 **mostrar os pontos** 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.

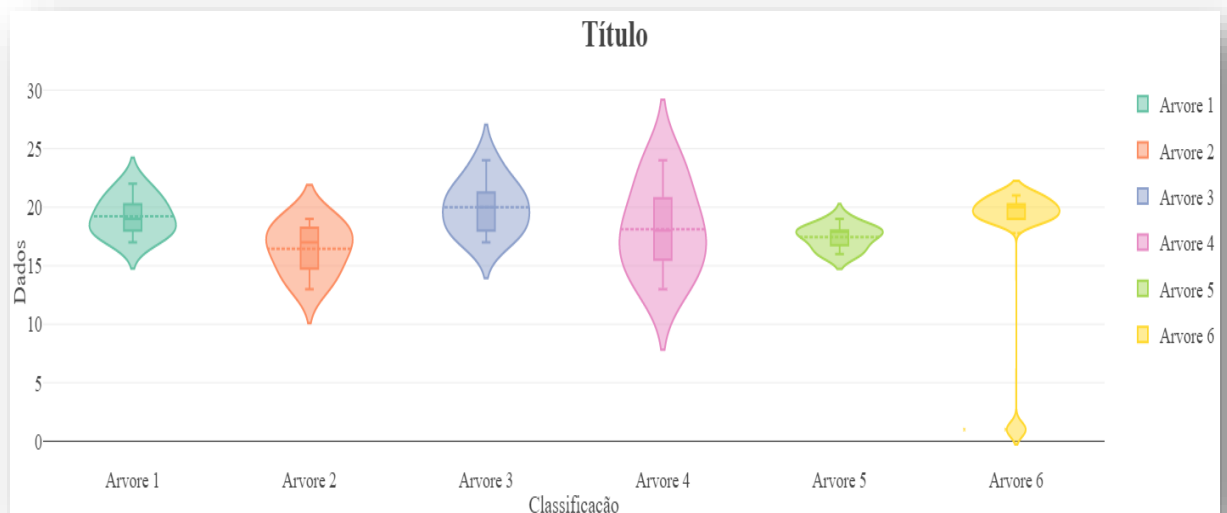
Gráfico com algoritmo de cálculo linear, média visível, com 0.13 de distância entre os pontos, -1.2 da posição dos pontos, 2 de tamanho, 0 de largura das bordas dos pontos, 60% de opacidade com formato de círculo:



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 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 **mostrar os pontos** 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:

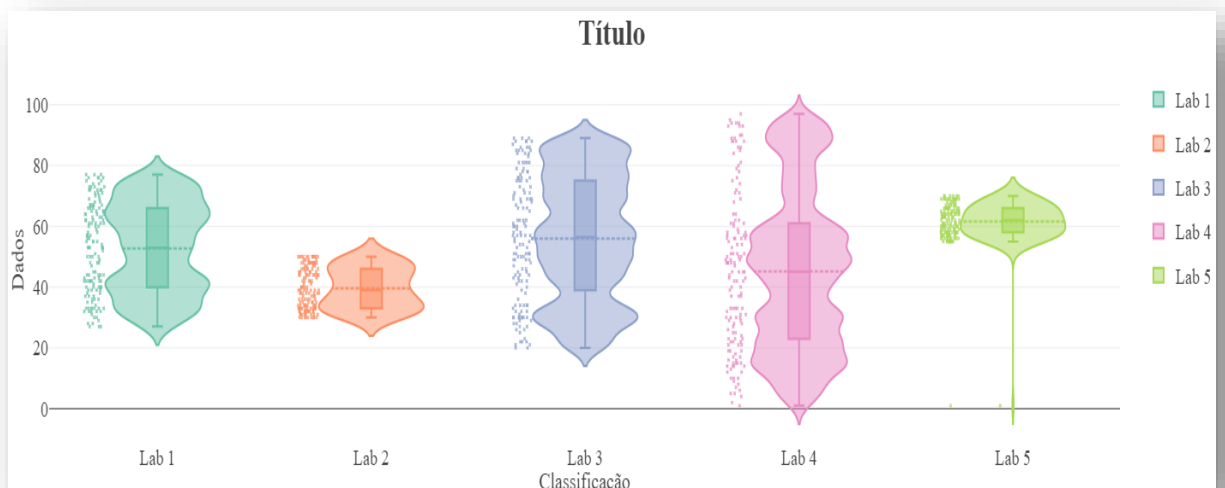


Gráfico de pontos:

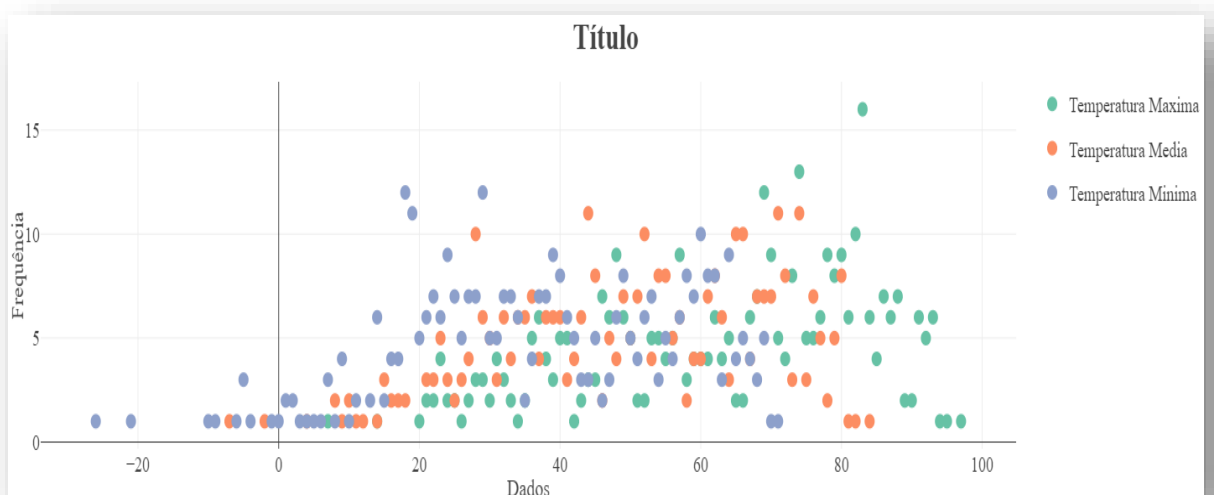
O gráfico de pontos, geralmente quando são dados parecidos, com várias observações e com frequências altas, representam os dados e suas frequências.

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 entre os pontos, e também 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%:

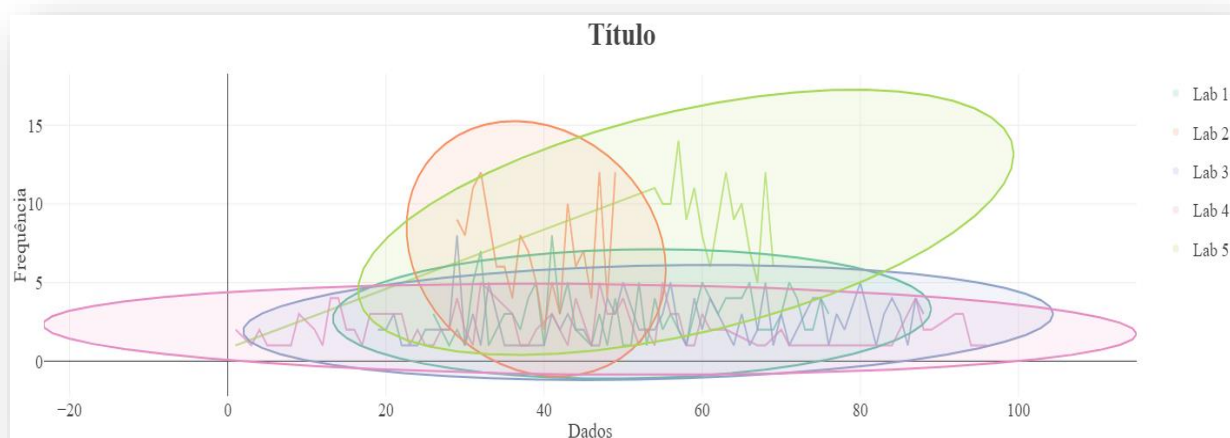


Gráfico com os pontos em formato de círculo, com linhas nos pontos, tamanho 10, opacidade em 80% e intervalo entre valores 1. Sem linhas e com uma elipse com 15% de opacidade e intervalo de confiança de 90%:

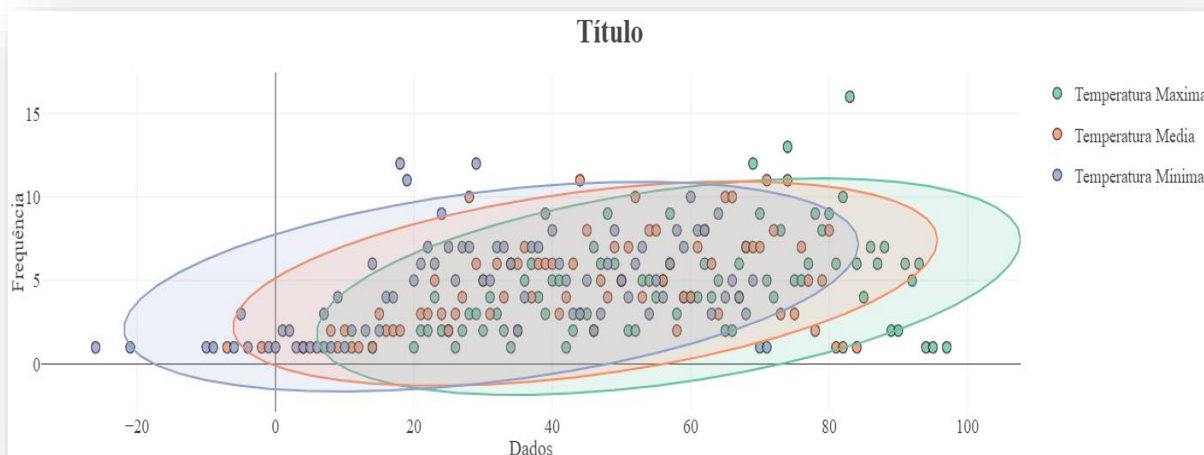
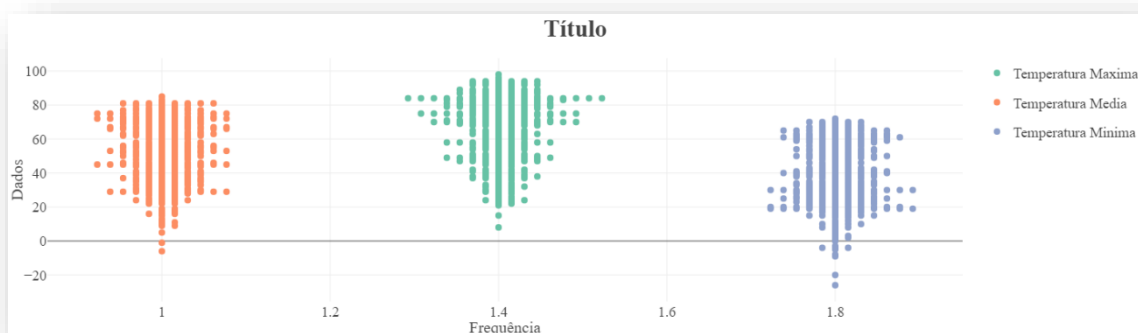


Gráfico de pontos em bee swarm:

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”:

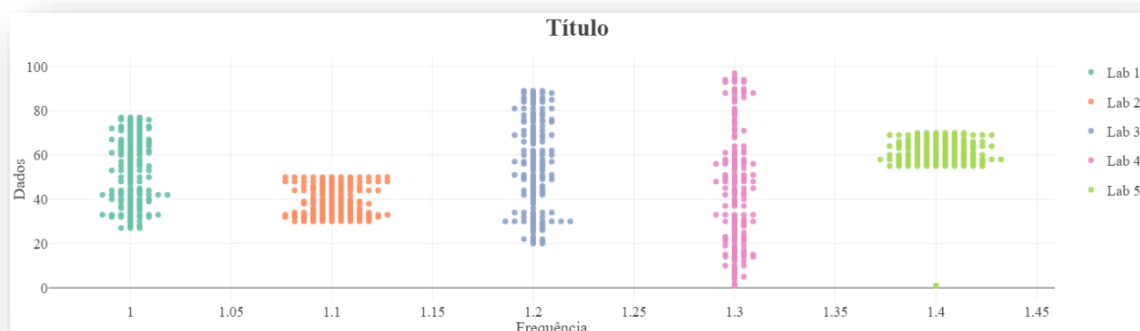
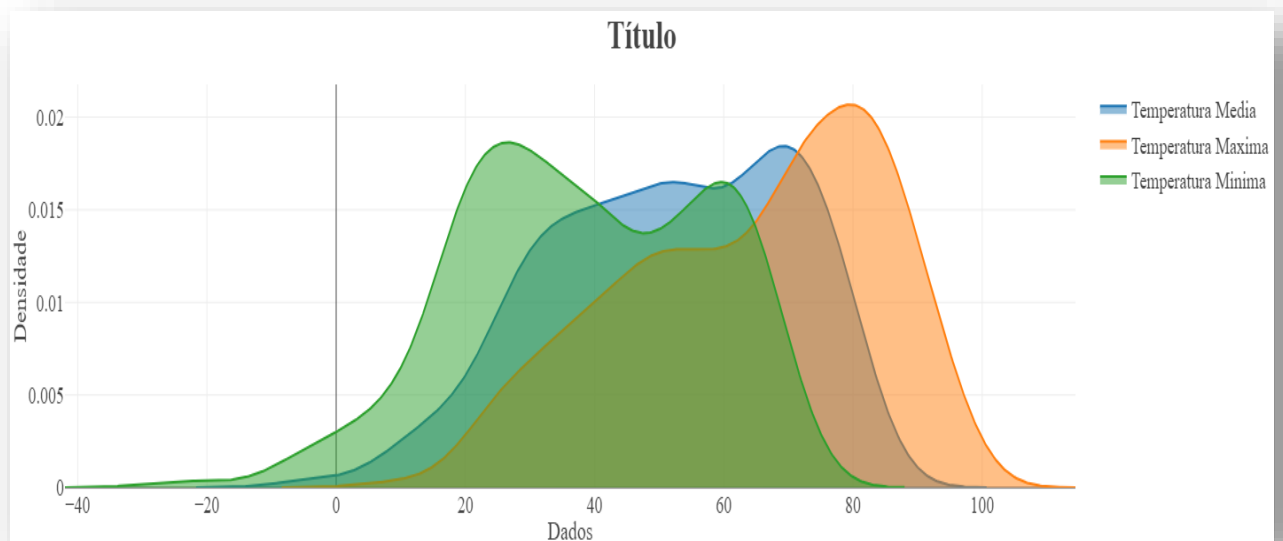


Gráfico de Densidade:

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, e 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

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

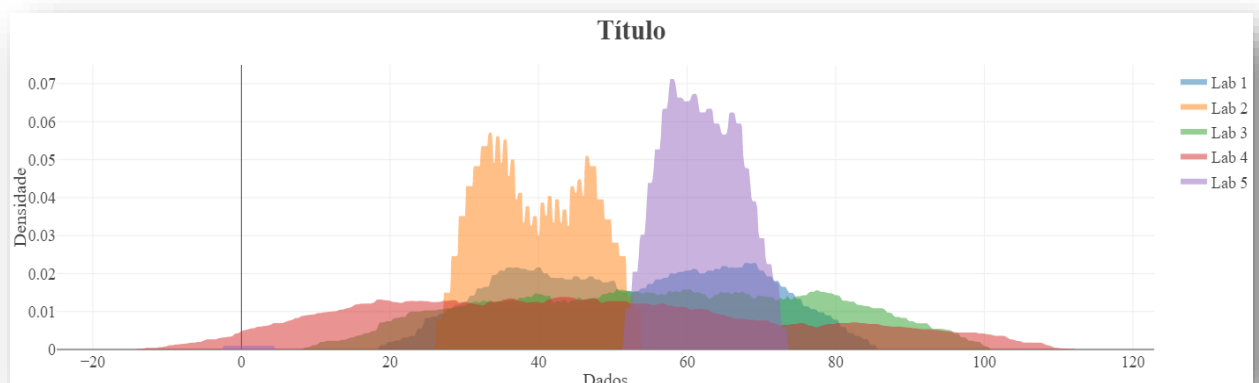
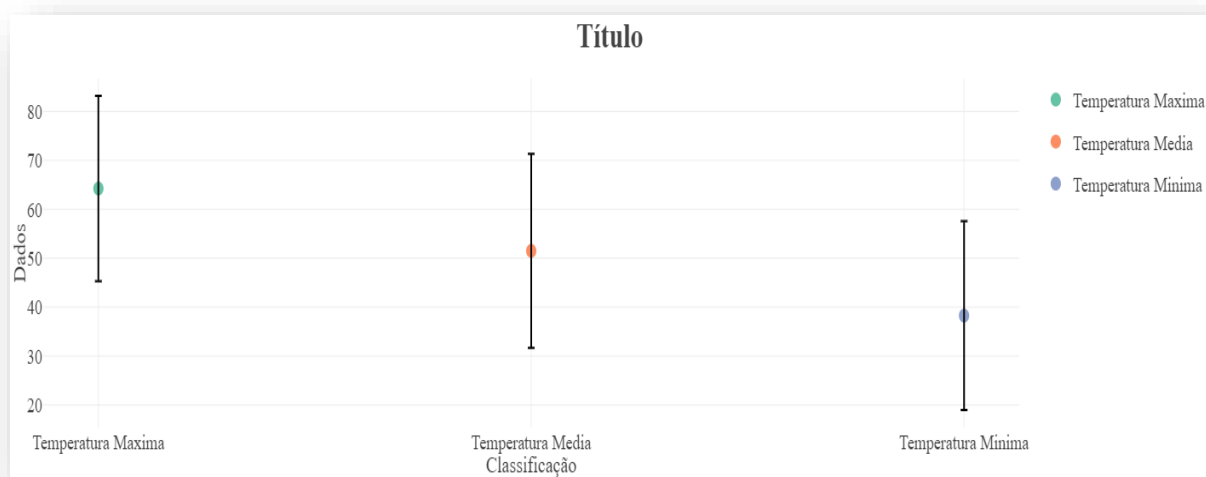


Gráfico de Erro:

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 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.

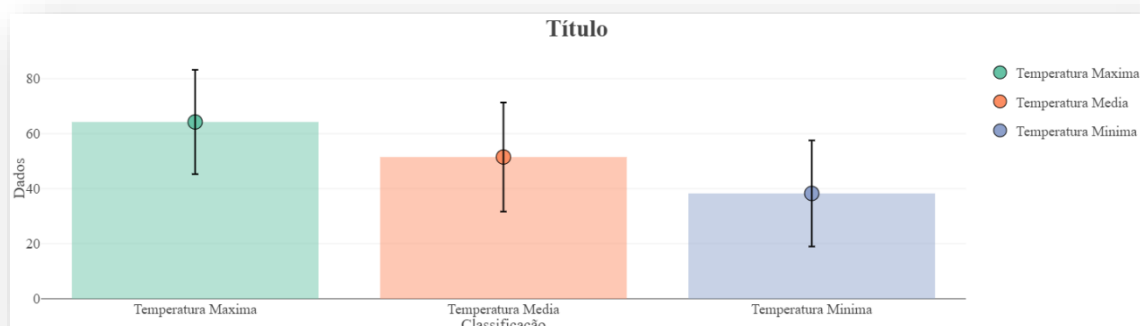
Para calcular o intervalo de confiança basta multiplicar o erro padrão pelo percentil associado ao nível de significância observado em

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

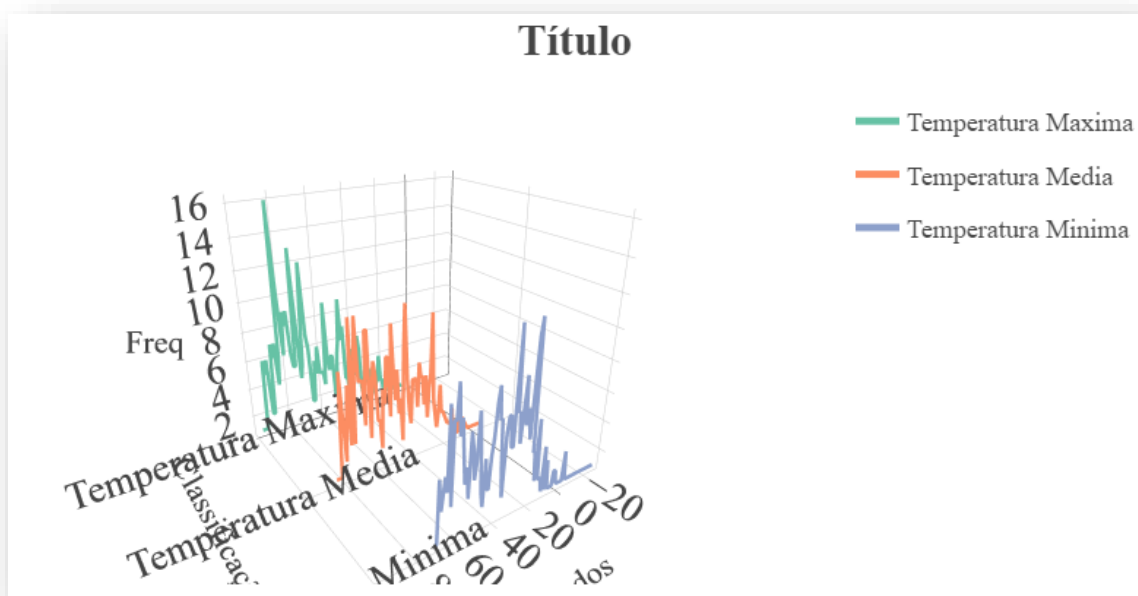
Os gráficos 3D podem ser encontrados na aba Dados Unidimensionais após carregar os dados no aplicativo.

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 nos controles de opções.

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.

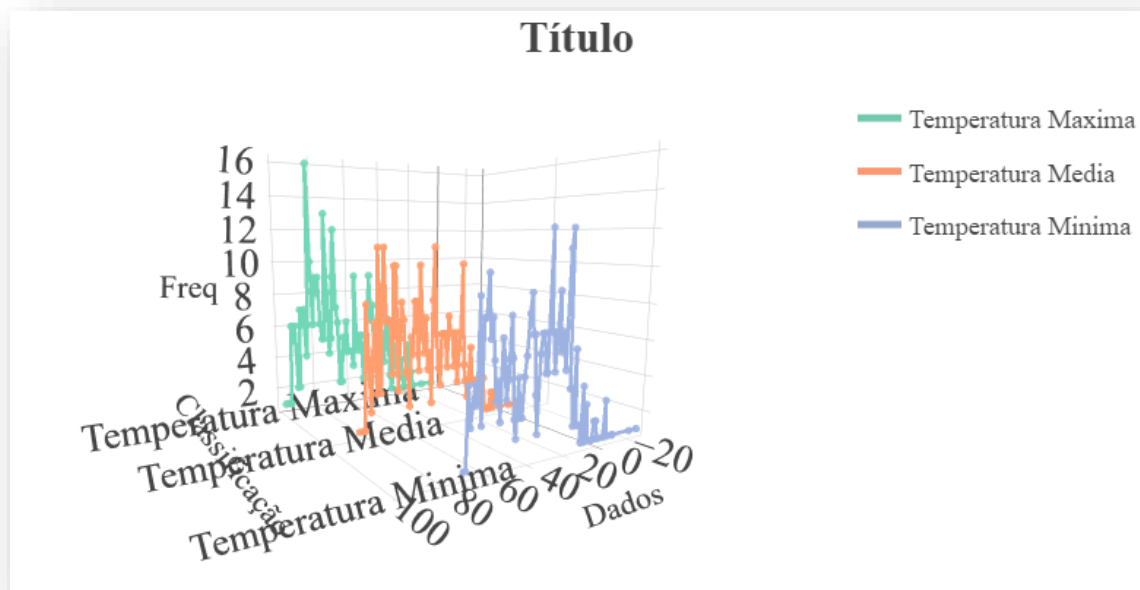
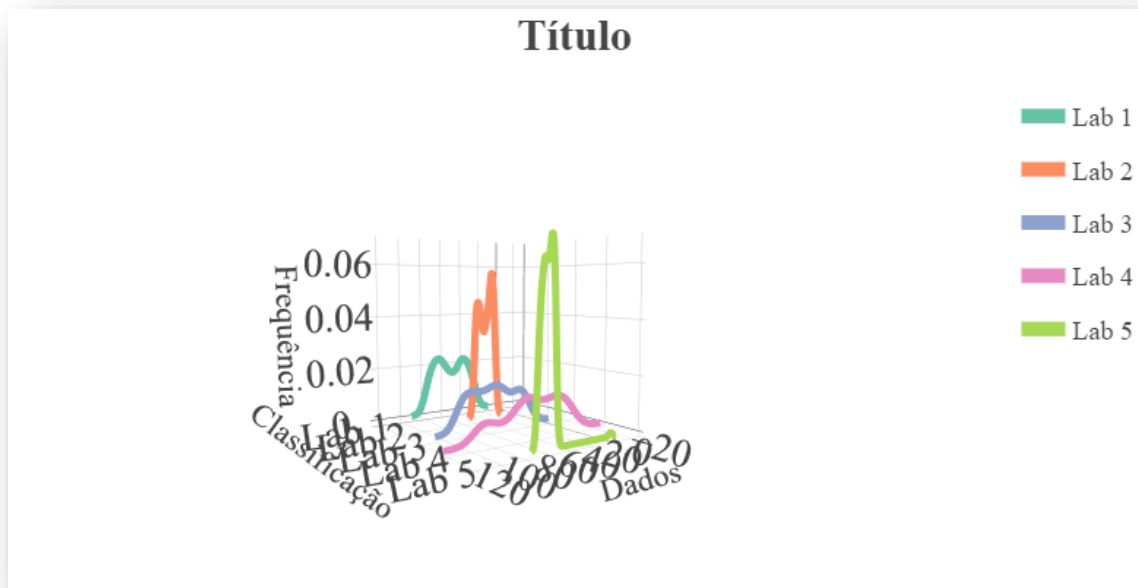


Gráfico de Densidade:

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

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

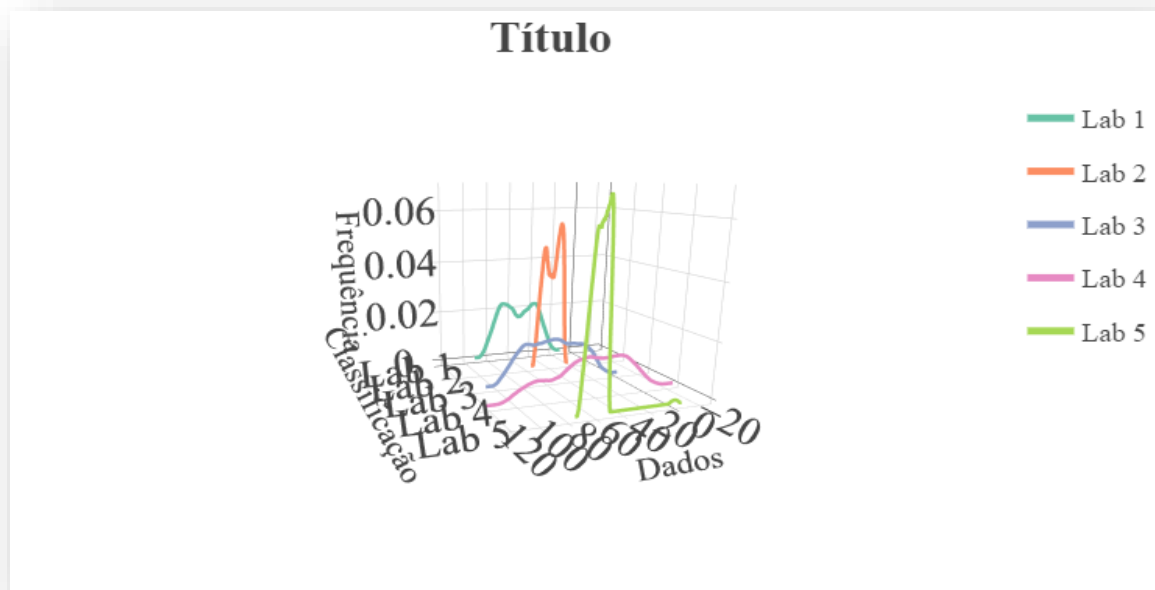


Gráfico de pontos

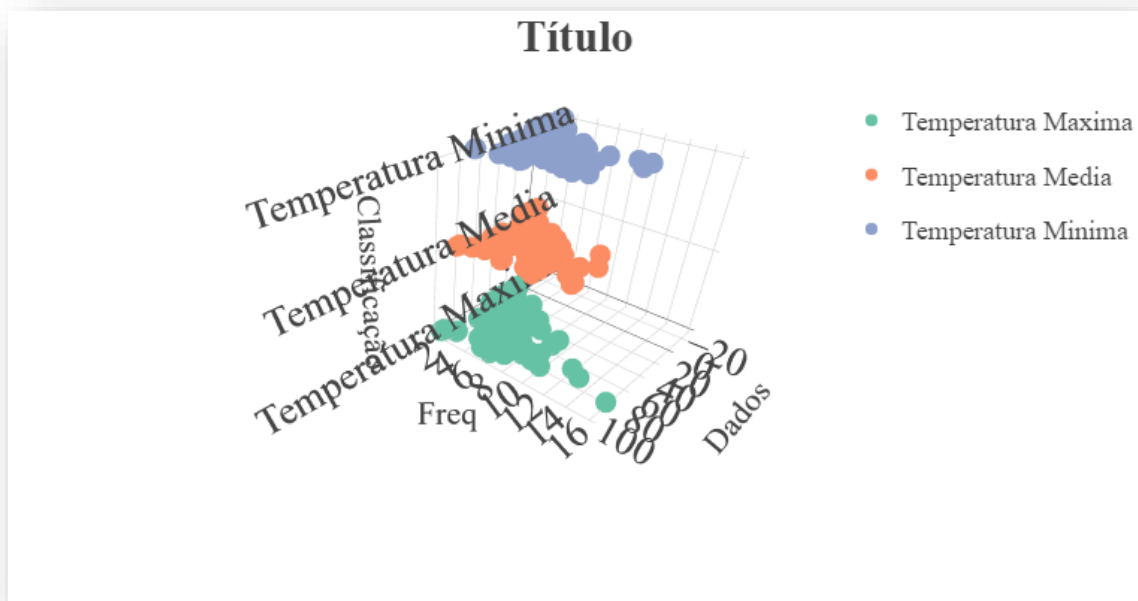
O gráfico de pontos, geralmente quando são dados parecidos, com várias observações e com frequências altas, representam os dados e suas frequências.

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 **linhas 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%:

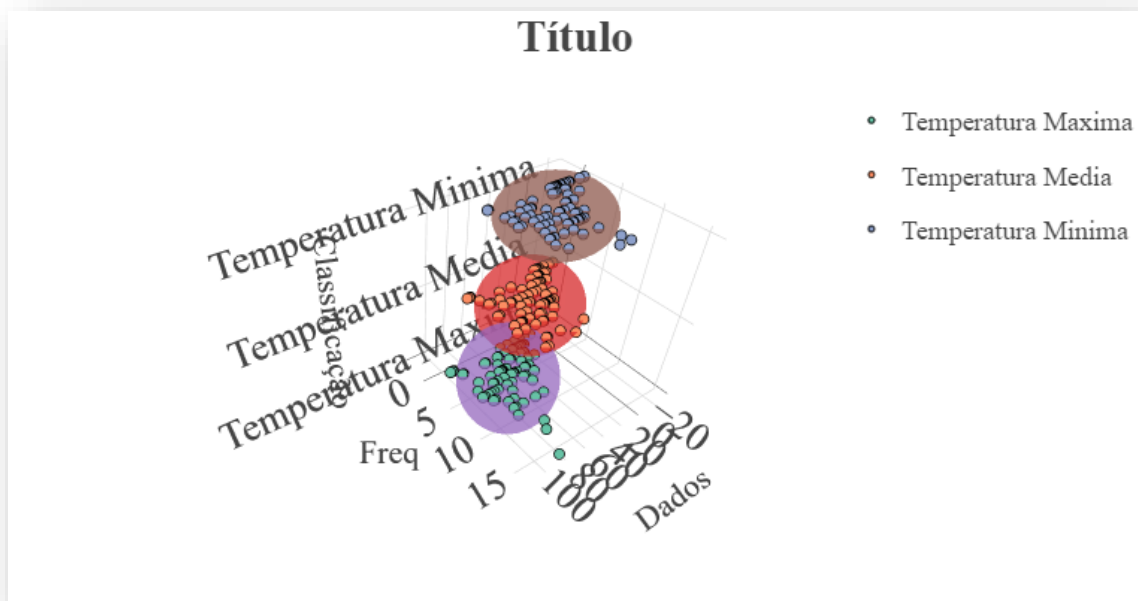


Gráfico com pontos com opacidade em 100% e intervalo entre valores 1, sem linha de borda, e uma elipse mostrando o contorno da elipse, com opacidade em 100% e com contorno igual a 2,5 e mostrando a área com 20% de opacidade e com um intervalo de confiança de 87%:

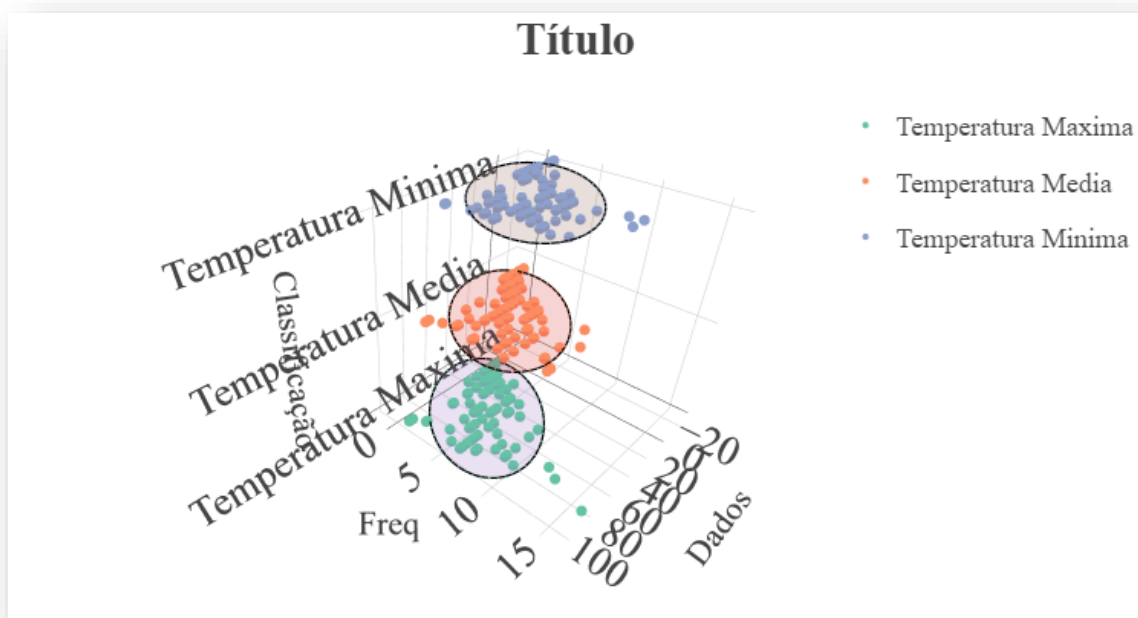
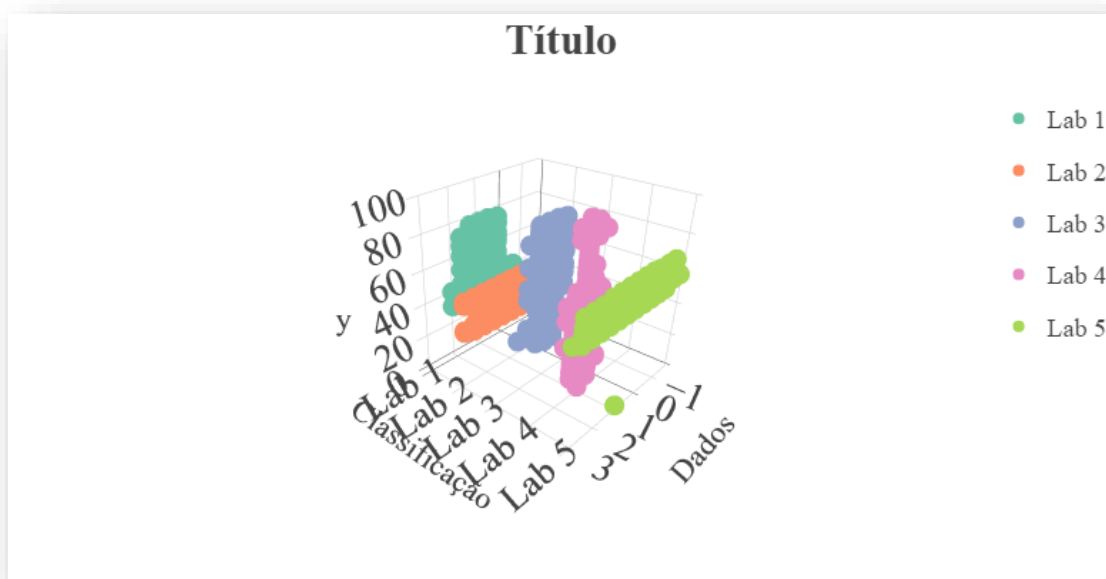


Gráfico de pontos 3D em Bee swarm:

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%

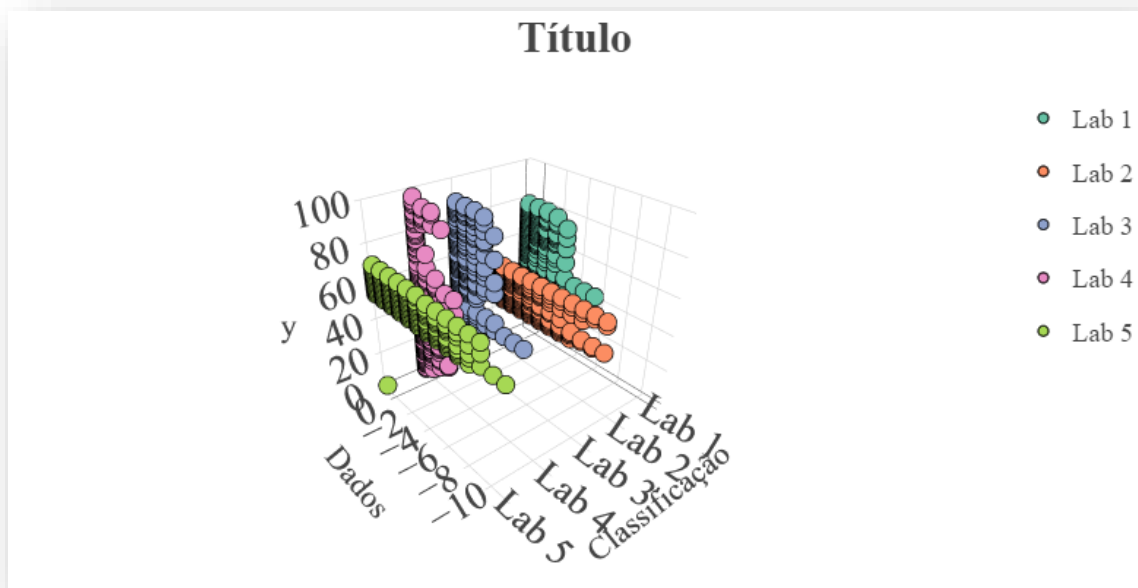
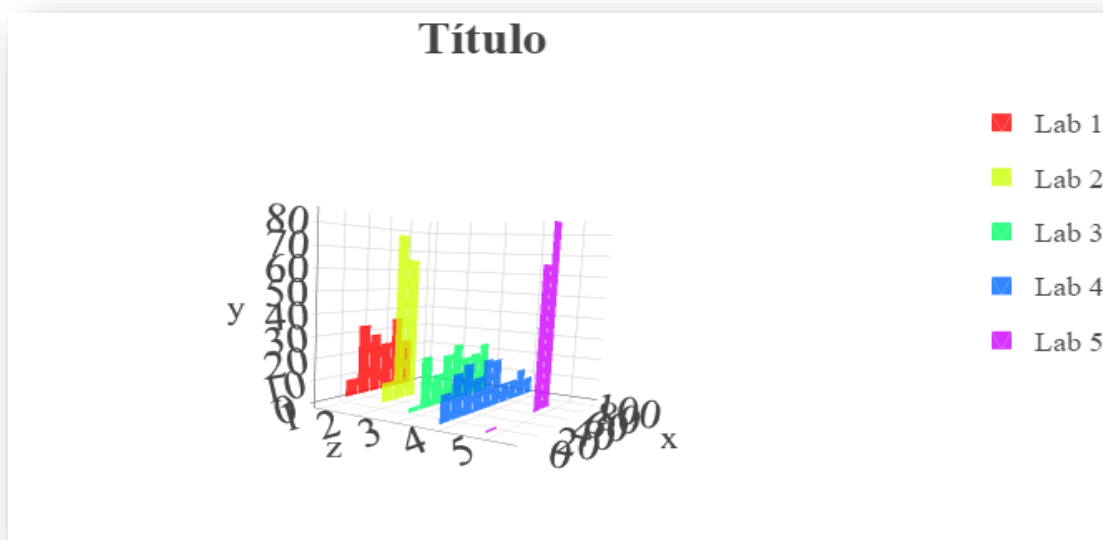


Gráfico de Barras:

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:

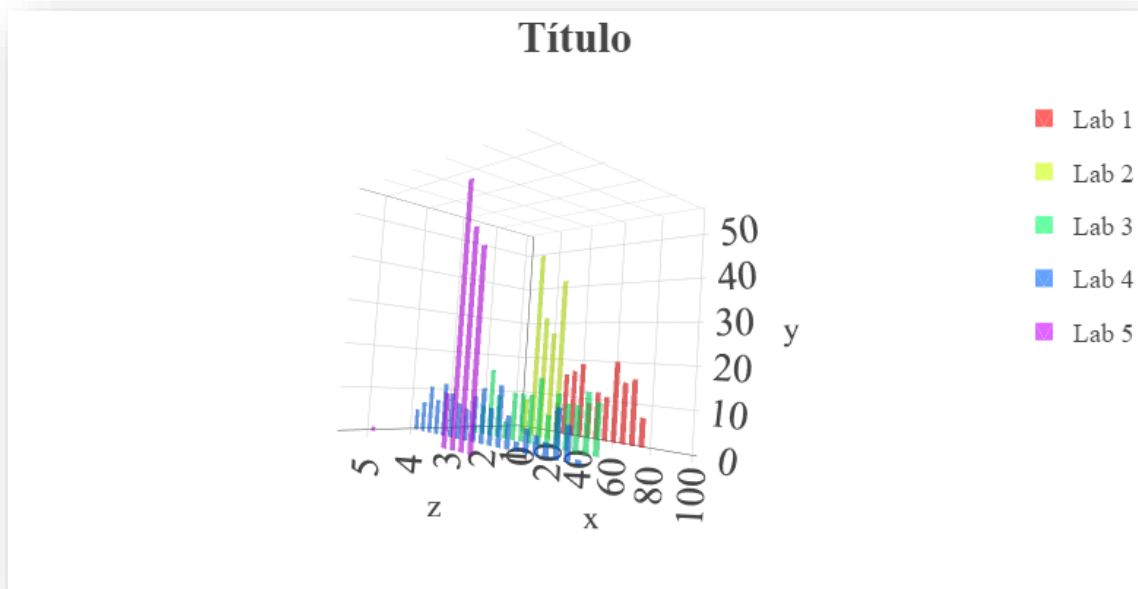
Nas configurações dos algoritmos pode-se escolher o **algoritmo para o cálculo do número de intervalos**, pode se escolher entre o “Sturges” (padrão), “Scott”, e o “Freedman-Diaconis”.

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

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%:



Checando os dados

Nesta página você pode checar os dados inseridos para verificar as **estatísticas, normalidades, distribuições, testes e homogeneidades**, além de poder realizar a **transformação** de certos dados.

Esta página está localizada em:

Dados unidimensionais -> Checando os dados

Após o usuário já ter carregado os dados

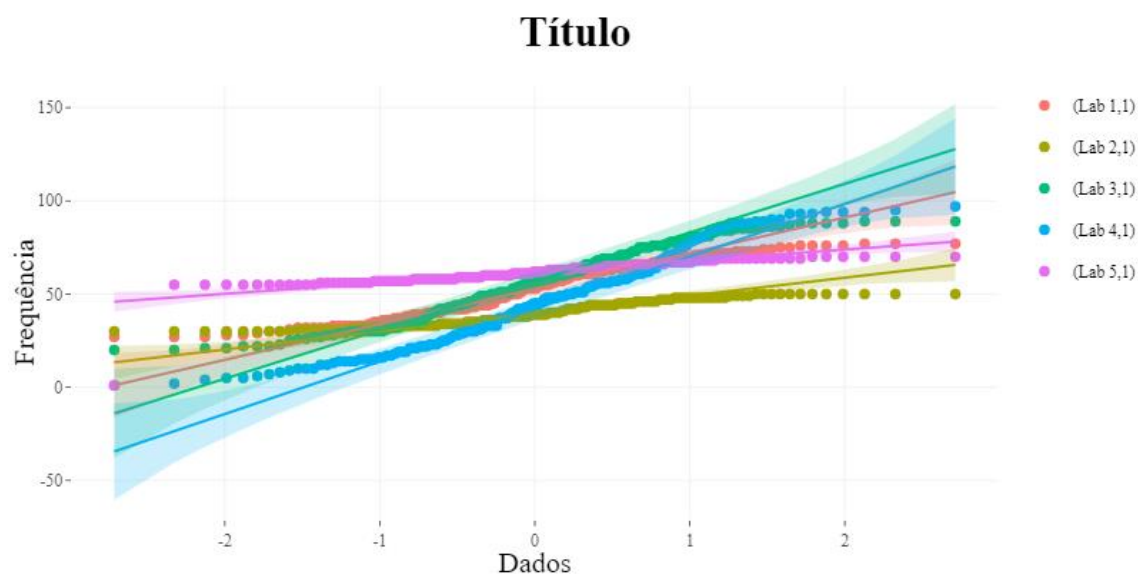
Distribuição de dados normais

Para verificar se os dados são normais podemos utilizar métodos visuais e estatísticos.

Neste painel mostra um **gráfico de densidade**, um método visual para o usuário verificar a normalidade dos dados, assim o usuário pode julgar se a distribuição está no formato de sino (curva normal).



Também está o **gráfico Q-Q** (Gráfico Quantil – Quantil) que é um método visual para comparar duas ou mais distribuições de probabilidade, traçando seus quantis uns contra os outros. Uma linha de referência também é mostrada. No gráfico Q-Q, cada observação é plotada como um ponto único. Se os dados são normais os pontos formam uma linha reta.



Por fim, como os testes visuais não são muito confiáveis existe os testes de normalidade para verificar a normalidade dos dados como o teste de **Shapiro-Wilk** e o teste de normalidade de **Kolmogorov-Smirnov (K-S)**.

Para ambos os testes, a hipótese nula é que a população está normalmente distribuída, assim, se o valor de p for menor que o nível alfa escolhido, a hipótese nula é rejeitada e há evidências de que os dados testados não são normalmente distribuídos

O teste de Shapiro-Wilk é amplamente recomendado para teste de normalidade e fornece melhores resultados do que o teste K-S. É baseado na correlação entre os dados e o valor normal correspondente.

Já o teste de normalidade de Kolmogorov-Smirnov é recomendado para amostras de dimensão igual ou maior do que 30. Para amostras com dimensões menores, é recomendado o teste de Shapiro-Wilk.

Show entries
Search:

	Teste Shapiro-Wilk	Decisão - teste Shapiro-Wilk	Teste Kolmogorov-Smirnov	Decisão - teste Kolmogorov-Smirnov
Normal	0.929	Normal	0.7642	Normal
Poisson	0.001718	Não normal	0	Não normal
Binomial	2.33e-10	Não normal	0	Não normal
Exponencial	3.824e-10	Não normal	0	Não normal
Chi Quadrado	1.779e-11	Não normal	0	Não normal

Showing 1 to 5 of 5 entries

Previous
1
Next

Dados Bidimensionais

A barra de navegação de dados bidimensionais serve para você inserir um ou mais dados bidimensionais, podendo verificar gráficos, distribuições, testes e outros aspectos.

Inserção dos Dados

Importar os dados:

Existem duas opções para a inserção dos dados pelo usuário, dados de exemplo que já vem baixados com o aplicativo e inserir dados de uma planilha Excel.

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

Para os dados exemplares você deve selecionar o painel:

Dados Bidimensionais -> Carregue seus dados -> Importe seus dados -> Usar um arquivo exemplo

Existem 3 opções de exemplos: “**Combustível**” são três colunas, a primeira chamada “Velocity.ratio”, que contém a velocidade de cada um dos carros. “Err” que é o erro, e “Fluido” que é o grupo, ou a variável independente, com as opções “gás”, “querosene”, “Siptech_34_cSt”, “Siptech_46_cSt” e “Siptech_90_cSt”.

O segundo exemplo é a base de dados “**Ansiedade**”, da base de dados “datarium::anxiety” com 5 colunas. “Id” que representa a identificação de cada uma das linhas, “Group” que é o grupo, ou a variável independente com as opções “grp1”, “grp2” e “grp3”. E “T1”, “T2” e “T3” que são os mesmos testes de ansiedade, porém medido em 3 tempos diferentes. Todos os “T1”, “T2” e “T3” podem ser representados como variáveis dependentes ou covariáveis.

O terceiro exemplo é a base de dados “**Escolaridade**” que contém 6 colunas. “Sujeito” que representa a identificação de cada uma das linhas, Gênero que pode ser a variável independente, ou de grupo que podem ser M ou F. “Grau_Instrução” que também pode ser a variável independente, ou de grupo com as opções “Superior”, “Ensino Médio” e “Ensino Fundamental”. “n_Filhos” é a quantidade de filhos de cada sujeito, “Idade” que é a idade de cada um dos sujeitos, e “Salário” que contém o salário de cada um dos sujeitos em mil. Tanto “n_Filhos”, “Idade” e “Salário” podem ser a variável dependente ou a covariável.

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

Dados Bidimensionais -> Carregue seus dados -> Importe seus dados -> Importar um arquivo .xlsx

Os dados devem estar neste formato:

	A	B	C	D	E	F	G	
1	Sujeito	Gênero	Grau_Instr	n_Filhos	Idade	Salário		
2	1	M	Superior	1	31	8.2		
3	2	F	Superior	0	25	5.3		
4	3	M	Superior	0	33	9.4		
5	4	M	Ens Fundar	0	20	2.9		
6	5	F	Ensino Mé	0	23	3.7		
7	6	F	Superior	2	37	4.4		
8	7	M	Ens Fundar	0	38	4.7		
9	8	M	Superior	0	37	5.4		
10	9	M	Superior	1	34	5.8		
11	10	M	Ensino Mé	2	40	3.2		
12	11	F	Superior	0	41	6		

Com a primeira linha sendo os nomes das colunas, e as demais os seus respectivos valores.

Inserir os dados na planilha

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

Dados Bidimensionais -> Carregue seus dados -> Digite seus dados na planilha

Os dados devem estar neste formato:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	Sujeito	Gênero	Grau_Instrução	n_Filhos	Idade	Salário					
2	1	M	Superior	1	31	8.2					
3	2	F	Superior	0	25	5.3					
4	3	M	Superior	0	33	9.4					
5	4	M	Ens Fundamental	0	20	2.9					
6	5	F	Ensino Médio	0	23	3.7					
7	6	F	Superior	2	37	4.4					
8	7	M	Ens Fundamental	0	38	4.7					
9	8	M	Superior	0	37	5.4					
10	9	M	Superior	1	34	5.8					
11	10	M	Ensino Médio	2	40	3.2					
12	11	F	Superior	0	41	6					
13	12	F	Ensino Médio	3	46	6.9					
14	13	F	Ensino Médio	0	26	2					
15	14	F	Ensino Médio	0	44	2.4					

De novo, a primeira linha é o nome das colunas e as demais são os seus respectivos dados.

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.

Comparando duas médias.

Neste painel, após a inserção dos dados, você pode comparar dois dados em diversos testes e distribuições

Comparando múltiplas médias

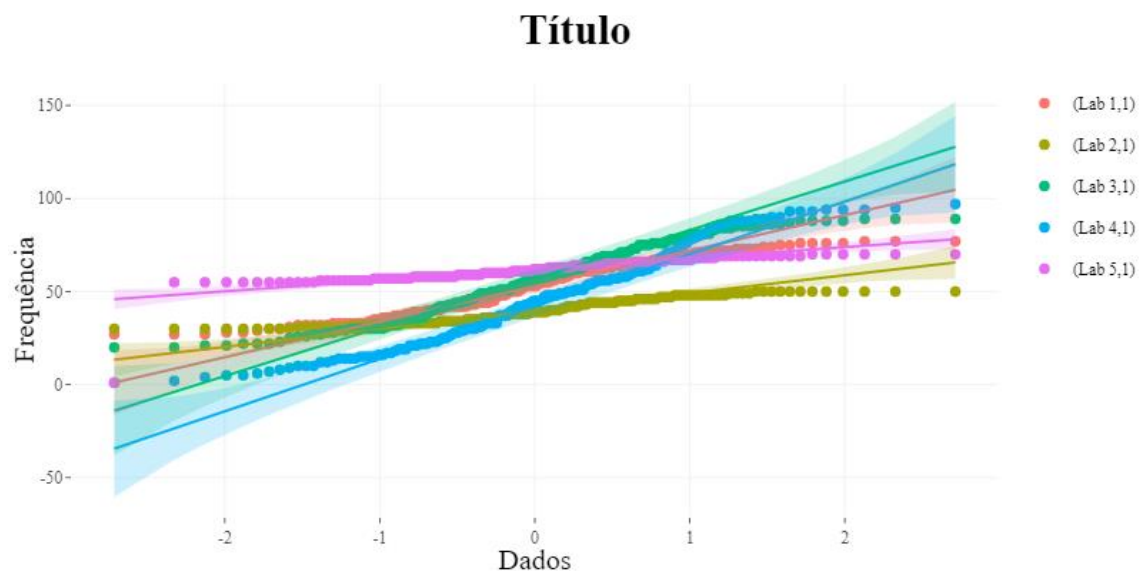
Neste painel, após a inserção dos dados, você pode comparar os diversos dados.

ANCOVA

A ANCOVA (Análise de covariância) é um é um modelo linear geral que combina análise de variância e regressão.

A ANCOVA avalia se as médias de uma variável dependente (VD) são iguais entre os diferentes níveis de uma variável independente categórica (VI), muitas vezes chamada de tratamento, apesar de estatisticamente controlar os efeitos de outras variáveis contínuas que não são de interesse primário, conhecidas como covariáveis (CV) ou variáveis incômodas.

Na ANCOVA o usuário define qual coluna é a VD, CV e VI e retornará um **gráfico de regressão linear**, com pontos e retas, mostrando o coeficiente angular de cada uma das variáveis independentes além de uma reta mostrando o coeficiente angular de todos os pontos.



Além disso, embaixo do gráfico tem uma tabela ANCOVA **removendo o efeito da covariável**. Após o ajuste o usuário consegue observar estatisticamente a diferença.

ANCOVA

Show entries Search:

	Soma.de.quadrados	DF	F	p
Velocity.ratio	1.3859	1	408.6387	0
Fluido	0.0443	4	3.2659	0.0138
Residuos	0.4375	129		

Showing 1 to 3 of 3 entries Previous Next

Abaixo desta existem duas tabelas com os **testes de Levene** e de **Shapiro-Wilk** mostrando respectivamente a normalidade dos resíduos e a Homogeneidade das Variâncias.

Verificação de suposição

Teste de Levene					Teste de Shapiro-Wilk		
Show	10	entries	Search:		Show	10	entries Search:
	F	Df1	Df2	p		Estatística	p
Teste de Levene	2.673	4	130	0.0349	Teste de Shapiro-Wilk	0.9912	0.5612
Showing 1 to 1 of 1 entries					Showing 1 to 1 of 1 entries		
Previous					Previous		
1					1		
Next					Next		

Por fim existe a tabela **PostHoc** que compara cada uma das combinações 2 a 2 da variável independente, e sua significância de acordo com o valor de P;

Tabela Posthoc:

Show

10

▼

entries

Search:

	Grupo 1	Grupo 2	df	Estatística	p	p.adj	Significância
1	gas	kerosene	129	2.3375	0.021	0.2095	ns
2	gas	Siptech_34_cSt	129	2.0817	0.0393	0.3935	ns
3	gas	Siptech_46_cSt	129	2.6342	0.0095	0.0947	ns
4	gas	Siptech_90_cSt	129	0.7978	0.4264	1	ns
5	kerosene	Siptech_34_cSt	129	0.2362	0.8136	1	ns
6	kerosene	Siptech_46_cSt	129	0.8601	0.3913	1	ns
7	kerosene	Siptech_90_cSt	129	-0.6196	0.5366	1	ns
8	Siptech_34_cSt	Siptech_46_cSt	129	0.8873	0.3766	1	ns
9	Siptech_34_cSt	Siptech_90_cSt	129	-1.1184	0.2655	1	ns
10	Siptech_46_cSt	Siptech_90_cSt	129	-2.0744	0.04	0.4003	ns

Showing 1 to 10 of 10 entries

Previous

1

Next

Dados Tridimensionais

A barra de navegação de dados tridimensionais serve para você inserir um ou mais dados tridimensionais, e recebe um ou mais gráficos em Mesh tridimensional para a visualização e comparação dos dados.

Inserção dos dados

Importar os dados:

Existem duas opções para a inserção dos dados pelo usuário, dados de exemplo que já vem baixados com o aplicativo e inserir dados de uma planilha Excel.

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

Para os dados exemplares você deve selecionar o painel:

Dados Tridimensionais -> Carregue seus dados -> Importe seus dados -> Usar um arquivo exemplo

Existem 2 opções de exemplos: “zoneA” e “zoneB”, que são uma lista de matrizes bem grandes.

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

Dados Tridimensionais -> Carregue seus dados -> Importe seus dados -> Importar um arquivo .xlsx

Os dados devem estar neste formato:

	A	B	C	D	E	
		<u>Ref. Vol. Flow</u>	<u>Fluid Kin.</u>			
1	<u>%Err (Ref.Vol) (%)</u>	<u>(litre/s)</u>	<u>Viscosity</u>			
			<u>(cSt)</u>			
2	<u>1.23</u>	<u>129.937452</u>	<u>10.32</u>			
3	<u>1.21</u>	<u>129.9202346</u>	<u>10.31</u>			
4	<u>1.23</u>	<u>129.9914902</u>	<u>10.36</u>			
5	<u>1.22</u>	<u>130.0470201</u>	<u>10.38</u>			
6	<u>1.25</u>	<u>129.9913216</u>	<u>10.42</u>			
7	<u>1.25</u>	<u>130.0443494</u>	<u>10.44</u>			
8	<u>1.20</u>	<u>89.76443802</u>	<u>10.65</u>			
9	<u>1.25</u>	<u>90.04478425</u>	<u>10.71</u>			
10	<u>1.28</u>	<u>89.98729684</u>	<u>10.69</u>			
11	<u>1.18</u>	<u>90.07862331</u>	<u>10.67</u>			
12	<u>1.22</u>	<u>90.08810392</u>	<u>10.65</u>			
13	<u>1.34</u>	<u>65.52886776</u>	<u>10.62</u>			
14	<u>1.28</u>	<u>65.51356611</u>	<u>10.61</u>			
15	<u>1.34</u>	<u>65.49131991</u>	<u>10.59</u>			
16	<u>1.32</u>	<u>65.48238396</u>	<u>10.58</u>			
17	<u>1.26</u>	<u>65.47099372</u>	<u>10.55</u>			
18	<u>1.38</u>	<u>39.37665209</u>	<u>10.55</u>			
19	<u>1.33</u>	<u>39.37155512</u>	<u>10.56</u>			
20	<u>1.36</u>	<u>39.36693941</u>	<u>10.56</u>			
21	<u>1.36</u>	<u>39.36451462</u>	<u>10.57</u>			
<div> <div> <div></div> <div></div> </div> <div> <div>Siptech46cSt</div> <div>Siptech 35 cSt</div> <div>Siptech 90 cSt</div> <div>Kerosene 3 cSt</div> </div> </div>						

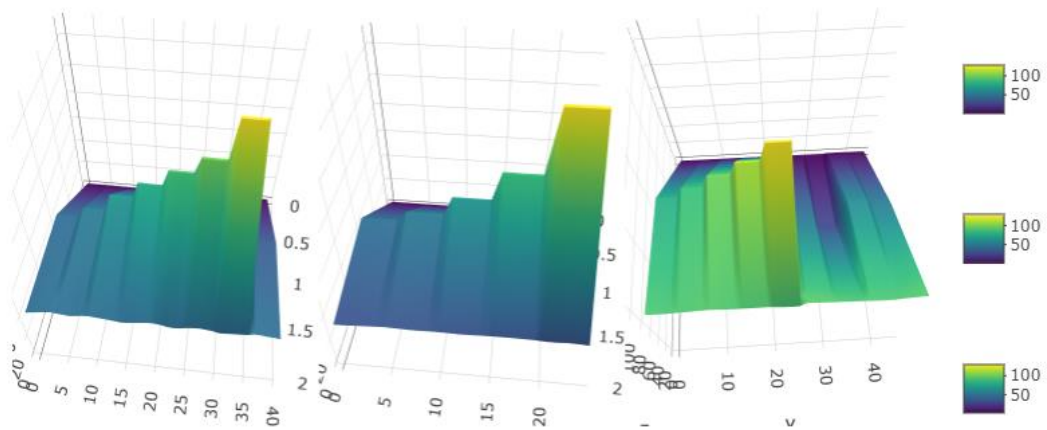
Criando diversas planilhas, para cada planilha, a primeira linha são os nomes das colunas, e as demais os seus respectivos valores.

Em seguida você deve informar o número de planilhas que existem no seu arquivo .xlsx. Vale ressaltar que não é necessário que todas as planilhas tenham o mesmo número de linhas e colunas.

Gráficos em Mesh

Os gráficos em Mesh são uma comparação de diversos Gráficos em Mesh de dados tridimensionais. A página está em Dados Tridimensionais após o usuário carregar os dados tridimensionais.

Exemplo de 3 gráficos em Mesh comparando de lado, um a um.



Dado os dados inseridos, o usuário pode escolher quais gráficos mesh ele deseja mostrar, podendo definir as dimensões que ele deseja mostrar o gráfico completo.

Vale ressaltar que as posições dos gráficos em Mesh vão ser reiniciadas se o usuário adicionar mais um gráfico em Mesh.

Referências

- Practical Statistics in R II - Alboukadel Kassambara