



This is a graded discussion: 10 points possible

due -

28 37

When to use what?

We have been discussing about 1D scatterplot (and its variants), boxplot, histogram, and other variants such as cumulative histogram or CDF. Now my question is: when to use what?

When do you want to use 1D scatterplot? Why? When do you want to use boxplot or histogram? Why? Can you write a brief guideline for choosing a method for 1D data visualization?

To answer this question, think about key limitations and benefits of each method. Also think about what each methods show or hide.



← Reply



([https://](https://iu.instructure.com/courses/2165942/users/6679606))

Sneha Satish (<https://iu.instructure.com/courses/2165942/users/6679606>)


Sep 25, 2023



The following are some guidelines for choosing a visualization:

- A scatterplot is used to visualize the relationship between two variables. It helps identify correlation as well as any outliers in the data.
- A boxplot is used to summarize the distribution of the data. It provides information about median, quartiles, and potential outliers in the data. It is effective for comparing the distributions of multiple variables side by side.
- A histogram is used to visualize the distribution of continuous data. It provides a visual representation of the data's frequency distribution by dividing it into bins.

Some other guidelines include data type, data size, distribution characteristics, etc.


← Reply 



[https://](https://iu.instructure.com/courses/2165942/users/6701599)**Thomas Jablenski** (<https://iu.instructure.com/courses/2165942/users/6701599>)


Sep 25, 2023

I would use 1D scatterplots when I have a small dataset that I would like to see the relationship between the data points. I would use a histograms when I am looking to see what the distribution of large datasets data looks like. Note that the amount of bins is crucial in giving a good example of the distribution. Boxplots would be used to get the general idea behind how much data is in each percentile but if I wanted to see an even better representation of that data I would use a cumulative histogram.

[← Reply](#) [https://](https://iu.instructure.com/courses/2165942/users/6701715)**Dustin Cole** (<https://iu.instructure.com/courses/2165942/users/6701715>)


Sep 26, 2023

Scatterplots are best to compare two variables against each other. 1D scatterplots are helpful when you have one array with few values, but typically they are not the best option. Box plots are great for exploratory analysis to get a quick view of the distribution and easily see IQR, Median, and outliers. Histograms again are helpful with exploratory analysis and they are simple and easy to understand. They provide a general idea of the distribution but if you want a more accurate display you would use density functions or CDFs.

[← Reply](#) [https://](https://iu.instructure.com/courses/2165942/users/6587577)**Yumeng Liang** (<https://iu.instructure.com/courses/2165942/users/6587577>)

Wednesday


When I do not need to see each data point, I want to look at the trend more I can use the 1D scatterplot. But if I really want to know each data point, I think I can consider the histogram. Also, the boxplot can be used to look at the median, mean, etc. It gives an overview of this data set instead of providing the details.

[← Reply](#) [https://](https://iu.instructure.com/courses/2165942/users/6352173)**Erik Gonzalez** (<https://iu.instructure.com/courses/2165942/users/6352173>)

Thursday

The most appropriate plot will be driven by the end purpose behind creating a visualization. I'd recommend using the below rules of thumb:

- 1D scatterplot (and variants) - Appropriate when you desire to visualize the specific location of every individual data point. Displaying each point can lead to a lot of clutter, particularly when there is a large dataset. Too much data can render these graphs illegible
- Boxplots - Most useful when you'd like to understand key statistics for a dataset (such as medians, quartiles, quartile spread), or identify outliers, but does come at the expense of masking data point locations
- Histogram - Ideal for visualizing the distribution of data, but selecting bins (determining whether to use constant or variable bin width, and the number of bins to use) can be crucially important. Incorrectly selecting bin details can cause a histogram to be accurate but misleading. If using variable bin widths, interpretation relies on area and can confuse less familiar members of the audience
- eCDFs are useful for displaying how a distribution aggregates and doing so without requiring bins (and the inherent tradeoffs of bins). If there is too much data, it can be difficult to determine how much the cdf changes from point to point visually. These do require a little more effort for audience members to process

← [Reply](#) 



<https://iu.instructure.com/courses/2165942/users/6762945>

Thursday

A scatterplot will be best when trying to see a relationship in the data, especially to see if there are outliers and what they might be. Doesn't show any statistical information like median and quartiles.

Boxplot is easiest to show summary statistics like median and quartiles but could cause issues when looking at data distribution.

histogram is best for data distribution and for seeing frequency of data. doesn't show summary statistics much however.

cumulative histogram does a better job at showing summary statistics while also showing data distribution. Adding percentile lines can help be the best of both worlds.

← [Reply](#) 

<https://iu.instructure.com/courses/2165942/users/6705680>**Andi Mai** (<https://iu.instructure.com/courses/2165942/users/6705680>)

Friday

When someone wants to visualize the distribution of individual data points along a single dimension, they can use 1D scatterplot.

boxplot: when you want to visualize the summary statistics and key characteristics of a univariate dataset.

histogram: when you want to visualize the frequency or density distribution of a continuous variable


Guidelines for Choosing a 1D Data Visualization Method:

Think about whether your data is continuous or discrete;

Small datasets with distinct data points fit for 1D scatterplots;

If you need to compare multiple groups or categories, boxplots are often a better choice.

Boxplots are robust to outliers, while histograms may require careful binning.

← Reply 

<https://iu.instructure.com/courses/2165942/users/6684610>**Shantanu Dixit** (<https://iu.instructure.com/courses/2165942/users/6684610>)


Friday

When I need to visualize 1D data, I would use a 1D Scatterplot when I want to see each individual data point, but I would be cautious as it can get cluttered with overlapping points.

I would opt for a Boxplot when I want a quick snapshot of the data's central tendency, spread, and any outliers, but it won't show me the data's detailed distribution.

For seeing how often values occur, I would choose a Histogram, but I'm aware its look can change based on bin choices.

And if I want to know how much data falls below a certain value, I go for a Cumulative Histogram or CDF. It's great for understanding data's cumulative nature, but it doesn't give a granular view of the data distribution.

[← Reply](#) **Onur Tekiner** (<https://iu.instructure.com/courses/2165942/users/6758180>)

Friday

⋮


1D scatter plot we use where data is a few. It is effortless to interpret. It is inefficient when data is too large because there is only one-dimensional space to show details. Also, there is an overlay problem too. There are no visual differences between one data point of three or 100 data points of three. There are all overlay each other.

Boxplot is when we need to find out the median and quartiles. It is also very efficient to show outliers.

The histogram shows the frequency and distribution of data.

The cumulative histogram shows cumulative probabilities and distribution with estimated quartiles.

Box plots, histograms, and cumulative histograms do not show exact patterns like scatter plots.

[← Reply](#) **Shreedeeep Sadasivan Nair (he/him/his)** (<https://iu.instructure.com/courses/2165942/users/6813278>)

Saturday

⋮

A) Scatterplot : We typically use 1D scatterplot when it is required to plot something along a single axis .

why: 1D scatterplot is used to display each individual data points and also shows the distribution of the data points.

Benefits: It shows outliers in a distribution and also shows each every point.

B)Boxplot: We use boxplot when we require to see the different quantiles of a distribution.

why: It gives a summary about the distribution of the data

Benefits: Identifies outliers and shows the summary about the distribution


Limitation: Does not show exact individual data points

C) Histogram: we use histogram to display the frequency of the data points

why:- Histogram provides visual representation of data's spread.

Benefits: Provides information about the density of the data

limitation: Exact values are not shown


← Reply 



Sangzun Park (<https://iu.instructure.com/courses/2165942/users/6703376>)

Saturday

1D scatterplot is advantageous when the sample size is small, and is superior to any other plot for quickly understanding data. Additionally, outliers can be quickly identified. Histogram shows better performance than scatter plot when there is a relatively large amount of data. It also allows you to check the distribution of data more quickly and reliably, and is especially specialized for checking data by section. boxplot is a tool that allows you to check key indicators of data at once. It includes various key indicators such as median value, quartile value, outlier value, maximum value, and minimum value. This is especially useful when making comparisons between multiple pieces of data. Lastly, I recommend using 1D scatterplot early in data analysis. Because it is possible to quickly understand the data, analysts can quickly plan the next step.

← Reply 



Prem Amal (<https://iu.instructure.com/courses/2165942/users/6684842>)

Saturday

1D Scatterplot:

We can use 1D scatterplots when we want to visualize individual data points to understand their distribution and identify outliers or anomalies. 1D scatterplots show each data point, making them suitable for exploring the exact values and detecting any unusual observations. With a large number of data points, 1D scatterplots can become cluttered and less interpretable so we should avoid using 1D scatter plots in such scenarios. If we have a dataset of customer transaction amounts and want to identify any unusually large or small transactions, a 1D scatterplot allows us to visualize each transaction's value individually, if the

dataset is too large, it will create occlusion issue, so choice depends on considering various scenarios and requirement.


Boxplot:

We can use boxplots when we want to visualize the summary statistics of the data, including the median, quartiles, and potential outliers. Boxplots provide a concise summary of key statistics and help us quickly identify the central tendency, spread, and presence of outliers. Boxplots do not show individual data points, so they lack the granularity of scatterplots or histograms. In a study comparing the salaries of employees in different departments of a company, we can use boxplots to compare the median and spread of salaries for each department.

Histogram:

We can use histograms when we want to understand the distribution of data across predefined bins or intervals. Histograms show the frequency or density of data within specific ranges, making them useful for visualizing data distribution patterns. The appearance of the histogram can vary based on the choice of bin size, potentially leading to different interpretations. To examine the distribution of customer ages in a retail dataset, we can use a histogram to gain insights into whether the customer base is predominantly young, middle-aged, or older.

By considering data type, distribution, the need for outlier detection, and the level of detail required, we can select the most suitable method for data visualization while being aware of each method's pros and cons.

← [Reply](#) 



Mothi Gowtham Ashok Kumar (he/him/his) (<https://iu.instructure.com/courses/2165942/users/6683278>)

Saturday

1D scatterplot

- **Use when:** we want to see the individual data points and their relationship to each other. You want to identify outliers or patterns in the data.
- **Benefits:** Shows the individual data points and their relationship to each other. Can be used to identify outliers or patterns in the data.
- **Limitations:** Can be difficult to interpret if there are many data points or if the data is not evenly distributed. Cannot be used to compare data across different groups.

Boxplot

- **Use when:** we want to summarize the distribution of a dataset and identify the median, quartiles, and outliers.
- **Benefits:** Provides a good summary of the distribution of a dataset. Easy to interpret.
- **Limitations:** Cannot show the individual data points. Does not show the density of the data.

Histogram

- **Use when:** we want to see the distribution of a dataset and identify the most common values and the range of values.
- **Benefits:** Shows the distribution of a dataset and the density of the data. Easy to interpret.
- **Limitations:** Cannot show the individual data points. Does not show the relationship between the data points.

Cumulative histogram

- **Use when:** we want to see the distribution of a dataset and identify the percentiles.
- **Benefits:** Shows the distribution of a dataset and the percentiles. Easy to interpret.
- **Limitations:** Cannot show the individual data points. Does not show the relationship between the data points.

KDE plot

- **Use when:** we want to see a smooth estimate of the distribution of a dataset.
- **Benefits:** Shows a smooth estimate of the distribution of a dataset. Easy to interpret.
- **Limitations:** Cannot show the individual data points. Does not show the relationship between the data points.

Which method to choose?

The best method to choose for visualizing 1D data will depend on your specific needs and goals. If we want to see the individual data points and their relationship to each other, then a 1D scatterplot is a good option. If we want to summarize the distribution of a dataset and identify the median, quartiles, and outliers, then a boxplot is a good option. If we want to see the distribution of a dataset and identify the most common values and the range of values, then a histogram is a good option. If we want to see the distribution of a dataset and identify the percentiles, then a cumulative histogram or KDE plot is a good option.

Edited by **Mothi Gowtham Ashok Kumar** (<https://iu.instructure.com/courses/2165942/users/6683278>) on Sep 30 at 12:47pm

← Reply 

**Vedant Tapadia** (<https://iu.instructure.com/courses/2165942/users/6678810>)

Saturday

Boxplot. - We can use boxplots when we want to summarize the distribution of our data, especially in terms of its central tendency, spread, and the presence of outliers. It can easily visualize quartiles, medians, and the overall shape of the distribution.

Histogram - We can use histograms when we want to visualize the frequency distribution of continuous data, particularly when we want to see how data points are distributed across different value ranges. We can easily see the shape and density of the data distribution.

1D Scatterplot - Use 1D scatterplots when you want to visualize individual data points along a single axis. This is only useful when you have a relatively small dataset.

[← Reply](#) **Robert Perez (he/him/his)** (<https://iu.instructure.com/courses/2165942/users/6701521>)

Saturday

For the first question, when you only have one variable that you're plotting (even if you're comparing more than one single variable against a common axis), that's the situation to use a 1-D Plot. I think a scatterplot works so long as there isn't so much data as to create obfuscation. Box plots are great when you want to see a simple representation of the distribution of data. Histograms work best when you want more detail than the box plot provides for different ranges of the data distribution.

[← Reply](#) **Mukul Gharpure** (<https://iu.instructure.com/courses/2165942/users/6678592>)

Saturday

Choosing the right visualization method is very important for articulating the characteristics and nuances of the data. Given below is a brief guideline for choosing a method for 1D data visualization:

1. 1D Scatterplot (Strip Plot):

- When to Use: When we have a moderate number of data points and want to show individual data values.
- Benefits:
 - Displays exact data points without aggregation.
 - Useful for spotting individual outliers or specific value concentrations.
- Limitations:
 - Occlusion can be an issue with many overlapping points.
 - Not scalable for very large datasets.

2. Boxplot:

- When to Use: When we want a summary of data distribution with key statistics like median, quartiles, and potential outliers.
- Benefits:
 - Compact representation of data distribution.
 - Shows central tendency, spread, and potential outliers.
 - Useful for comparing distributions across categories.
- Limitations:
 - Hides the exact distribution and modality (e.g., doesn't show bimodal distributions).
 - Does not display individual data points (except potential outliers).

3. Histogram:

- When to Use: When we want to understand the distribution of the data, especially with large datasets.
- Benefits:
 - Shows the data distribution and allows for identification of modes.
 - Useful for spotting skewness in data.
- Limitations:
 - Bin size can heavily influence the appearance and interpretation.
 - Doesn't show individual data points.


4. Cumulative Histogram or CDF:

- When to Use: When we want to understand percentiles and the proportion of data below a certain value.
- Benefits:
 - Displays percentiles clearly.
 - Useful for understanding how data accumulates.
- Limitations:
 - Does not show frequency distribution in the same way a histogram does.
 - Can be less intuitive than a standard histogram for some audiences.

General Guideline:

1. Individual Data Points: If we want to show each individual data point and don't have a large dataset, go with a 1D scatterplot. Add jitter if there's overlap.
2. Data Distribution: For an overview of data distribution (especially with larger datasets), use a histogram.
3. Summary Statistics: If we're more interested in central tendency, spread, and outliers, or if we're comparing distributions across categories, use a boxplot.
4. Percentiles: If we're keen on showing data percentiles, the cumulative histogram or CDF can be better option.

To sum up, the choice often depends on the specific context and the audience. It's a good practice to experiment with multiple visualizations to find the most informative and clear representation of the data being used.

← [Reply](#) 



<https://iu.instructure.com/courses/2165942/users/6443321>

Saturday

For different types of data points we have, we should think about which plot is appropriate for visualization. 1D scatterplot is usually used when visualizing individual data points. It is good when looking at distribution and outliers since direct observation is possible. However, it is difficult when there is A LOT of data points. It should be used with small to moderate-sized datasets where individual data points can be looked at. Box plots are useful when looking at summary of key statistics like median, quartiles, and outliers. Also, skewness and spread of the data can be easily observed. It is simple plot that can easily be used to compare multiple dataset but it does not show individual data points or full distribution, which can be lacking in fine details. Histogram is useful when looking at data density, patterns, and distributions. However, bin size can affect the outcome of the visualization and does not show summary statistics and individual data points. Lastly, CDF is useful when looking at data distribution and helps understanding the percentiles. However, it also does not show fine-details of the distribution.

← [Reply](#) 



[Jeevan Deep Mankar \(https://iu.instructure.com/courses/2165942/users/6644229\)](https://iu.instructure.com/courses/2165942/users/6644229)

Sunday

1D scatterplot

When to use: To identify outliers and patterns in the data, or to show the relationship between two variables.

Limitations: 1D scatterplots can be difficult to interpret when there are many data points.

What it shows/hides: 1D scatterplots show the individual data points. They hide the distribution of the data.

Boxplot

When to use: Use a boxplot to show the distribution of the data, including the median, quartiles, and outliers.

Key limitations: Boxplots cannot be used to show the individual data points.

What it shows/hides: Boxplots show the median, quartiles, and outliers of the data. They hide the individual data points.

Histogram


When to use: Use a histogram to show the distribution of the data, including the shape, center, and spread of the data. You can also use histograms to compare the distributions of two or more groups of data.

Key limitations: Histograms can be difficult to interpret when there are many data points.

Additionally, they can be sensitive to the choice of bin size.

What it shows/hides: Histograms show the distribution of the data. They hide the individual data points.

Edited by **Jeevan Deep Mankar** (<https://iu.instructure.com/courses/2165942/users/6644229>) on Oct 1 at 1:31am

← Reply 



([https://](https://iu.instructure.com/courses/2165942/users/6684840)

Jash Shah (<https://iu.instructure.com/courses/2165942/users/6684840>)



Sunday

We can use

1. 1D Scatterplot:

We Use a 1D scatterplot when we want to visualize individual data points in a simple and unaggregated manner. It's particularly useful when you have a relatively small dataset.

1D scatterplots provide a clear view of data distribution, highlighting each data point's position.

They are great for identifying outliers and understanding data density.

The benefits are Transparency in data points and precise data representation.

Limitations are Limited ability to visualize data density in case of overlapping points or large datasets.

2. Boxplot:

We use a boxplot when we want to get a summary of the data's central tendency, and spread, and identify potential outliers. It's useful for medium to large datasets.

Boxplots provide a concise summary of the data distribution, including median, quartiles, and potential outliers, making it easy to compare multiple distributions.

Benefits are Efficient summary, outlier detection, and comparison between multiple datasets.

Limitations are Less detailed than histograms for understanding underlying data shape.

3. Histogram:

We use a histogram when we want a detailed view of the data distribution and when the data is continuous or can be divided into bins. It's suitable for large datasets.

Histograms show the data's frequency distribution, revealing patterns, modes, skewness, and potential gaps or outliers.

Benefits are Detailed visualization, clear shape of the distribution, easy to spot multiple modes.

Limitations are Choice of bin width impacts interpretation, less efficient for small datasets.

4. Cumulative Histogram (CDF):

We use a CDF when we want to understand cumulative probabilities, percentiles, or the overall distribution shape.

CDFs provide insight into how data accumulates across the range. They are useful for comparing datasets and assessing the likelihood of specific values.

Benefits are Visualization of cumulative probabilities, easy comparison, clear representation of percentiles.

Limitations are Less intuitive for some users can obscure fine details.

So we look at size of dataset, The complexity and we can decide how we want to visualize the end user datapoints

← Reply 👍



([https://](https://iu.instructure.com/courses/2165942/users/6760559))

Madhuri Patibandla (she/her/hers) (<https://iu.instructure.com/courses/2165942/users/6760559>)

Sunday

⋮

1D Scatter plot is also called as sweep charts and simplicity and explicit in nature. when we have more data then we can visualize in 2D scatter plot using Jitter, Bee swarm plot.

when we have millions of data , data will get overlap and data points are not transparent.

when we have less amount of data the 1D or 2D scatter plot

Box plot: **Box plots** may oversimplify complex data distributions and provide limited information about individual data points. Additionally, they may misrepresent the data if the sample size is too small or if the data distribution is highly skewed or has multiple modes. It is easy to identify median and 25th and IQR percentiles, and it's hard to understand the outliers.


if we need to understand the data distribution of data based on percentiles then box plot

Histogram: if histogram has too few bins, then the data distribution will not be to see much details and we have more data it will mislead the data and if data too many bins, then the data distribution will look rough, and it will be difficult to discern the signal from the noise. if we have more bins then Bin size varies height and data does not correspond to frequency, if we have constant bins sizes and height correspond to frequency ranges and histogram losses simplicity.

If we have less bin sizes the frequency ranges and area distribution of data will be good and easy to understand.

CDF: sum up the data points based on the bin size and easy to improvise the frequency of the data sets and to identify the percentiles in histogram.

Edited by **Madhuri Patibandla** (<https://iu.instructure.com/courses/2165942/users/6760559>) on Oct 1 at 2:28am

← Reply 

 <https://iu.instructure.com/courses/2165942/users/6818242> **Simon Driver** (<https://iu.instructure.com/courses/2165942/users/6818242>)
Sunday

1. I think I would use a 1D scatterplot when I am interested in visualizing a small data set, and each data point has a particular impact. For instance, maybe I would want to visualize the number of major earthquakes in the United States in the past 100 years. This would be a small data set where each data point is important and I want to see exactly the year. The limitations of the 1-D scatterplot are that it is hard to read off any measures of central tendency (eg such as the median) and it work poorly with large data sets.

2. I would use a boxplot or histogram for when I care about a large dataset I need to see trends in. For instance, if I wanted to visualize the amount of rainfall in Colorado on any given day in the past 50 years, I would use a histogram or box plot. In this data set, I want to know things like: the median, the spread of the data, and be able to visualize the overall trend or shape of the data. The limitations for this type of visualization are: outliers can be tricky to handle, and I have to be careful when designing my histogram to choose the right number of

bins. I also have to have enough data points, otherwise my visualization may end up being skewed. I also lose the ability to see any specific data point, like I could in the 1D scatterplot.

In essence, if I have a small data set where each point of data is important, a 1D scatterplot is good. If I have a large dataset with many points of data, and I want to visualize trends and measures of central tendency, a histogram or boxplot would be an ideal fit.

← Reply 👍



[https://](https://iu.instructure.com/courses/2165942/users/5444499)

Maria Klein (<https://iu.instructure.com/courses/2165942/users/5444499>)

Sunday



1D visualization selection guide:

When you have a small number of data points:

- 1D scatterplot
 - allows you to see each data point individually

When you have a large number of data points and you are interested in measures of central tendency and dispersion:

- Box plot
 - shows median, 25/75th percentile, IQR, and min/max/outliers, depending on how the plot is specified

When you have a large number of data points and you are interested visually inspecting distribution patterns:

- Histogram
 - bin selection is important

When you have a large number of data points and you still need information on individual data points:

- Cumulative density function
 - removes bins so that individual information is not lost

← Reply 👍



[https://](https://iu.instructure.com/courses/2165942/users/6703098)

Ao Zhang (<https://iu.instructure.com/courses/2165942/users/6703098>)

Sunday



1. When to use 1D scatterplot?

When the number of data points is moderate to draw and the overlap of data points is not serious. Since we cannot draw an ideal scatterplot if the data points are numerous. And if the overlap of data points is frequent, the figure cannot show the distribution of dataset properly. However, there are still some limitations of this method. Firstly, it cannot provide a summary statistics for dataset. And then, it will become clustered when data points are great.

2. When to use boxplot or histogram?

For boxplot, I will use it when I want to see the different contents of the data. For instance, when I want to check the median, IQR, maximum value, and minimum value. Additionally, when I want to check outliers, the boxplot will be my first choice. However, the boxplot cannot show the exact data points or the density of data.

For histogram, I will use it when I have large number of data points and I am interested in the distribution of the data. It performance well in showing density, skewness, and outliers. Nevertheless, it cannot show the statistics of the dataset like median, mean, maximum and minimum values. Moreover, it cannot show the exact data points either.

In the nutshell, there are some generalized rules for utilization of figure methods.

1. Data Size. For small data size, scatterplot is more informative. However, histogram is more fitting when the data size is big.

2. Outliers Detection. Boxplot will be a good choice if we want to detect the outliers based on figures.

← Reply 👍



[https://](https://iu.instructure.com/courses/2165942/users/6679250)

Hymavathi Gummudala (<https://iu.instructure.com/courses/2165942/users/6679250>)

Yesterday

Scatter plot -> When we want to visualize the individual data points.

Limitation: Occlusion with data points

Box plot -> When we are interested in finding the median, min, max, 25%, and 75%.


Limitation: Box plots cannot show percentiles and distribution patterns.

Histogram -> when you want to know what is the distribution about (normal, skewed towards right or left).

Limitation: Bin Choice might effect sometimes.

Histograms cannot show every percentile.

Cumulative Histogram/CDF -> when you want to find each percentile.

← Reply 




<https://iu.instructure.com/courses/2165942/users/6056428>

Yesterday



A scatterplot is best used when there are discrete data points whose individual value matters. This can be especially important if you expect there to be a trendline. A boxplot is best used when the descriptive statistics matter the most such as median, quartiles, and outliers. Histograms are best used when the distribution of the data is most important. If the data is likely unimodal or bimodal, a histogram can help to show this.

← Reply 



<https://iu.instructure.com/courses/2165942/users/6469527>


Yesterday



1D scatterplot is ideal when comparing relationship between two variables. The main reason for scatterplot is to provide a visual and statistical means to test the strength of relationship between two variables.

Boxplot is a statistical graph that shows variation in a set of data - displaying min, max, lower and upper quartile, median and outlier. The reason for boxplot is that it provides a useful way to compare different sets of data as several boxplots can be drawn per graph.

Histogram is a statistical graph that shows the frequency of numerical data or quantitative data. It shows general distributional features of data points variables, where values are concentrated within the distribution, the location of extreme values, possible gaps and unusual values.



← Reply 

 [https://](https://iu.instructure.com/courses/2165942/users/6682743)**Shubham Agarwal** (<https://iu.instructure.com/courses/2165942/users/6682743>)

Yesterday

When communicating the underlying distribution of the data is not the goal and the only goal is to visualize how the data is scattered : Scatter plot



When the goal is to communicate the underlying distribution of the data: Histogram

 [Reply](#) 

 [https://](https://iu.instructure.com/courses/2165942/users/6694681)**Vaibhav Piyushkumar Lodhiya** (<https://iu.instructure.com/courses/2165942/users/6694681>)

Yesterday

we can use a 1D scatterplot when you want to explore individual data points and their distribution. we can use a boxplot when you need a concise summary of central tendency, spread, and outliers. we can use a histogram when you want to understand the shape and density of the data distribution. we can use cumulative histogram (CDF) when you need precise information about percentiles and cumulative probabilities.


 [Reply](#) 

 [https://](https://iu.instructure.com/courses/2165942/users/5667580)**Sarah Biggs** (<https://iu.instructure.com/courses/2165942/users/5667580>)

Yesterday

1. 1D Scatterplot: This form of visualization could be useful if you have a single dimension to analyze without many associated datapoints. To review numerous datapoints, you could jitter them or create a bee plot, but this could become cumbersome. While it shows each individual datapoint, it doesn't give a nice display of the distribution as nicely as other options, and it doesn't display summary data.
2. Boxplot: A boxplot is perfect for showing the summary of your data in a neat, confined, and simple way. You can't see the shape of your distribution as well, but you can easily find your summary data: median, interquartile range, minimum and maximum.
3. Histogram: This form of visualization can be manipulated to show some summary data, but the best feature, in my opinion, is its ability to show the shape of the data. You can easily see if and how the data are skewed, and you can see multiple modalities very easily. This is a great visualization to see non-normal data, but literally seeing the curve is immediately beneficial for most data distributions, I find, especially when considering forms of analysis.

4. Cumulative histogram/CDF: These types of visualizations are very similar, and what they do extremely well is show the percentiles in the data, visually what percent of the data has been represented as you move along the X axis. The curve of this visualization is also another way to view the shape of the data. However, I do find that this is not as simple to look at and consider compared to a normal histogram.

← [Reply](#) 

○




<https://iu.instructure.com/courses/2165942/users/6706306>

Yesterday

⋮

If a single attribute dataset has a small number of points, a 1D scatterplot may be best. The data doesn't need to be summarized or aggregated in any way to be understood, it can be viewed directly. As the number of points grow, techniques such as jittering or beeswarm plots can be employed to maintain a direct visualization. Beyond a certain point, occlusion becomes an insurmountable problem, and some form of summarization or aggregation becomes necessary. If a statistical analysis or summary is most important, the boxplot is a good choice. If viewing a rough distribution of the data is desired, a histogram with a well-chosen number of bins is useful. Finally a CDF can be used to view how a distribution changes without losing fidelity to binning the data.

← [Reply](#) 

○



<https://iu.instructure.com/courses/2165942/users/6677399>

12:08am

⋮

1D Scatterplot:

When: Use a 1D scatterplot when you want to visualize individual data points in a simple and unaggregated manner. It's especially useful when you have a small to moderate number of data points and you want to see their exact values.

Why: 1D scatterplots provide a clear view of individual data points, making them suitable for data exploration and outlier detection. They show the distribution of data without any aggregation.

Boxplot:

When: Use a boxplot when you want to visualize the summary statistics of a dataset, including the median, quartiles, and potential outliers. Boxplots are helpful for comparing distributions

across different categories or groups.

Why: Boxplots provide a compact summary of the data distribution, making it easy to identify the center, spread, and skewness of the data. They are particularly useful for identifying outliers.

Histogram:

When: Use a histogram when you want to visualize the entire data distribution in detail, including its shape, mode, and skewness. Histograms are excellent for understanding data density.

Why: Histograms provide a granular view of data distribution by dividing it into bins. They help in identifying patterns, modes, and potential outliers.


Cumulative Histogram:

When: Use a cumulative histogram when you want to visualize the cumulative distribution of data. It's ideal for understanding percentiles and comparing datasets.

Why: CDFs show how the data accumulates from the lowest to the highest values, making it easy to identify percentiles and compare data distributions.

Guidelines for Choosing:

1. **Data Exploration:** If you're exploring data and need to examine individual data points, start with 1D scatterplots.
2. **Summary Statistics:** When summarizing data distribution or identifying outliers, boxplots are a good choice.
3. **Detailed Distribution:** If you want a detailed view of data distribution, especially for large datasets, histograms are effective.
4. **Percentiles and Comparisons:** For understanding percentiles and comparing datasets, use cumulative histograms.

← [Reply](#) 

○



[https://](https://iu.instructure.com/courses/2165942/users/6825193)

Harsh Patel (he/him/his) (<https://iu.instructure.com/courses/2165942/users/6825193>)

12:41am

⋮

1D Scatterplot:

When to Use: We can use a 1D scatterplot when you want to visualize individual data points along a single axis.

Benefits:

- Shows individual data points, making it easy to identify outliers.
- Provides a sense of data density and distribution.
- Suitable for small to moderately sized datasets.

Limitations:

- Cannot use for large datasets as points may overlap.
- Lacks summary statistics like central tendency and dispersion.
- Doesn't provide a smooth representation of data distribution.

Boxplot:

When to Use: Use a boxplot when you want to identify outliers in your dataset. It's useful for larger datasets or when you want to compare distributions across different categories.

Benefits:

- Provides a clear summary of key statistics: median, quartiles, and potential outliers.
- Robust to outliers, making it suitable for larger datasets.
- Facilitates easy comparison between multiple datasets or categories.

Limitations:

- Doesn't show the detailed data distribution within quartiles.
- May not capture the complete picture of data skewness or multimodality.

Histogram:


When to Use: Use a histogram when you want to visualize the data distribution, especially when you want to understand the shape, central tendency, and spread of the data.

Benefits:

- Provides a detailed view of data distribution, including shape and skewness.
- Allows you to see how data is distributed across bins.
- Suitable for exploring the characteristics of continuous data.

Limitations:

- Bin width choice can impact interpretation.
- May not be as effective for small datasets.

← [Reply](#) 

<https://iu.instructure.com/courses/2165942/users/6819877>**Sydney Dicks** (<https://iu.instructure.com/courses/2165942/users/6819877>)

12:43am



1D scatterplot

- When to use:
 - When working with an objectively small dataset - good for showing each individual data point
 - When working with a dataset with significant spread - good for quick views of a distribution in a dataset
- Limitations:
 - Very susceptible to obscuring data with even slightly large datasets or datasets with grouping

Boxplot

- When to use:
 - When looking to see quick summary statistics of a dataset
 - When working with a large, relatively normally distributed dataset
- Limitations:
 - Doesn't handle outliers or skewness very well

Histogram

- When to use:
 - When working with a large, relatively normally distributed dataset
 - When slight skewness appears in a dataset
- Limitations:
 - Non-normally distributed data can be more likely perceived incorrectly


[← Reply](#)<https://iu.instructure.com/courses/2165942/users/6694525>**Rohan Isaac** (<https://iu.instructure.com/courses/2165942/users/6694525>)

1:43pm



A 1D scatterplot can be used to perform a simple visualization of your data especially when the number of data points available is not too high. A boxplot can be used in situations where there is more importance given to visualizing the percentiles of the data like median, quartiles as well as outliers. Histograms can be used to visualize the distribution of the data and when we want to check ranges of data where the frequency of data points is more.

Each method has its own advantages and disadvantages. For example boxplots can mask multimodal data and not show important insights. It could also not be appropriate for datasets with few points. Histograms also have the tendency to mask data if bins are not tuned appropriately.

← [Reply](#) 

○



<https://iu.instructure.com/courses/2165942/users/6696028>

3:05pm

⋮

Scatterplots can show a lot of detail about the data, including the distribution of the data and the relationship between the two variables. They can also be used to identify outliers and patterns in the data.


But Scatterplots can be difficult to interpret when there is a lot of data. They can also be biased towards larger data points.

where as Boxplots are easy to interpret and can provide a lot of information about the distribution of the data, including the, spread, outliers.

but, Boxplots can be difficult to use when there is a lot of data

Histograms are easy to interpret

but,Histograms can be difficult to use when there is a lot of data or when the data is not normally distributed

← [Reply](#) 

○



<https://iu.instructure.com/courses/2165942/users/6692441>

4:16pm


⋮

Use a **1D scatterplot** when you want to visualize individual data points while maintaining a sense of their distribution. Good for highlighting outliers, spotting clusters, or understanding data point density along a single dimension.

Opt for a **boxplot** when you need a concise summary of the data's central tendency, spread, and the presence of outliers. Good for comparing multiple datasets, revealing quartiles, and identifying potential outliers.

Choose a **histogram** when you want to explore the overall shape and frequency distribution of your data. Good for revealing data patterns, skewness, and multimodality.

Edited by **Yashada Nikam** (<https://iu.instructure.com/courses/2165942/users/6692441>) on Oct 3 at 4:16pm

← Reply 

○



David Rosenthal (<https://iu.instructure.com/courses/2165942/users/6762824>)

5:08pm

⋮

You want to use a 1D scatterplot when you have a small number of data points and want to see where those individual points sit. This will help ensure that you get an accurate picture of what is going on, such as, clusters, patterns, or outliers.

You use a boxplot and/or histogram when you want to summarize the data and find quick quartiles and the median. These are used to provide a summary and are easy to detect outliers in the data. You can also see good distribution on these types of charts.

← Reply 