

This is a graded discussion: 10 points possible

due -

45 45

Why should we care about the number of bins?

So, why should we care?

1. What happens if your histogram has too few bins? Any illustrative examples? When will a histogram with too few bins mislead you?
2. Conversely, what happens if your histogram has too many bins? Any illustrative examples? When will a histogram with too many bins mislead you?

Unread



✓ Subscribed

← Reply



(https://

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



Sep 25, 2023


The number of bins is a crucial parameter that can significantly impact the interpretation of a histogram.

Too few bins can lead to loss of information and misleading representation. It might fail to capture important details and can result in incorrect conclusions about the nature of the data.

An example of this could be a dataset of exam scores ranging from 0 to 100. If there are only 2 or 3 bins, you might get a histogram that shows scores as "low", "medium", "high" which obscures the distribution in each category.

Too many bins can lead to noise and difficulty in comparing distributions. It might cause fluctuations in the data and can be challenging to discern meaningful differences or similarities between the two datasets.

An example could be daily temperature readings for a year, if you use a separate bin for each degree Fahrenheit, the histogram might show excessive detail, making it hard to identify seasonal temperature patterns.

← [Reply](#) 



Dustin Cole (<https://iu.instructure.com/courses/2165942/users/6701715>)

Sep 25, 2023

1. Too few bins in a histogram can oversimplify the data. If you have one or two bins for a decent sized dataset, you don't see enough of the details. Histograms are supposed to show distribution and with too few bins you won't see the actual distribution.

2. Too many bins makes it hard to categorize or simplify the distribution. If you choose too many bins you may as well create an area chart or distribution plot. Especially if you have so many that it looks like you are plotting every point on the x-axis.

← [Reply](#) 




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

Sep 25, 2023

If you have too few bins then it looks like all of the data is clumped together when in reality it's more spread out than expected. With this amount of bins the chart could not give enough detail to get any details from it. This type of histogram could mislead you when a 2 bins merge that hold information that should be kept separate.

If you have too many bins the histogram will look very jagged and look closer to a scatterplot's representation of the data. A histogram with too many bins could mislead when it the extra bins just result in more noise and don't give off the message you are trying to give with bins.

← [Reply](#) 



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

Sep 26, 2023

1. 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.

Ex :

Score and Freq Ex

Score Frequency

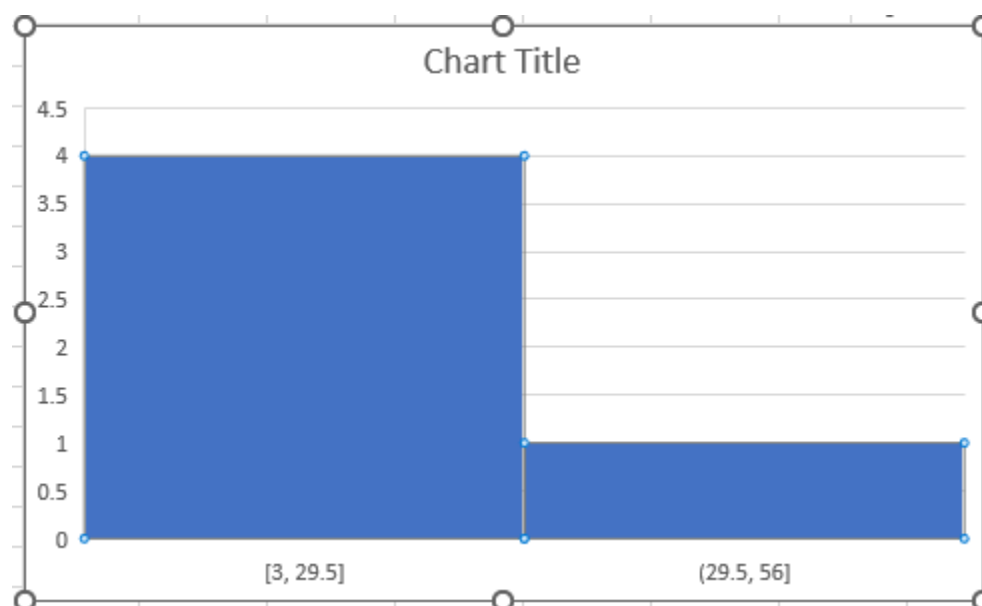
0-199 5

200-399 29

400-599 56

600-799 17

800-999 3



2. If histogram has too many bins, then the data distribution will look rough, and it will be difficult to discern the signal from the noise. On the other hand, with too few bins, the histogram will lack the details needed to discern any useful pattern from the data. Yes, If we defined too many bins like 0.0, 4.2, 8.4, 12.6, 16.8, 21.0 then to understand the frequency of the data will be difficult and can not understand the content of the Graph.

Edited by **Madhuri Patibandla** (<https://iu.instructure.com/courses/2165942/users/6760559>) on Sep 26 at 1:20am

← Reply 👍



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

Sep 26, 2023

1. Too few bins may result in the concealing of ranges with smaller counts/frequencies. Grouping ranges that contain completely different counts/frequencies and displaying some

combination of their numbers will mislead a viewer.

2. A histogram with too many bins can clutter a visualization. If 10 bins contain the same count/frequency, keeping them separated keeps more lines on the visualization that aren't needed - chart junk!

← [Reply](#) 👍

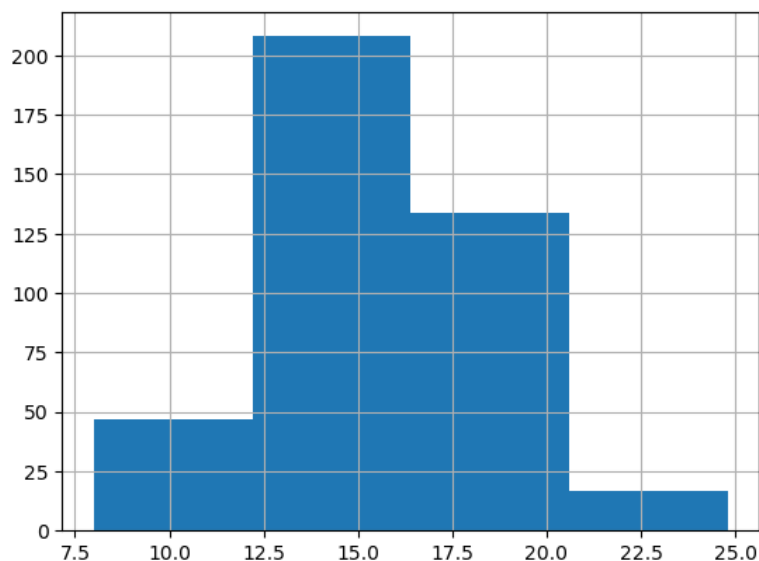


Gary Croke (<https://iu.instructure.com/courses/2165942/users/6706306>)

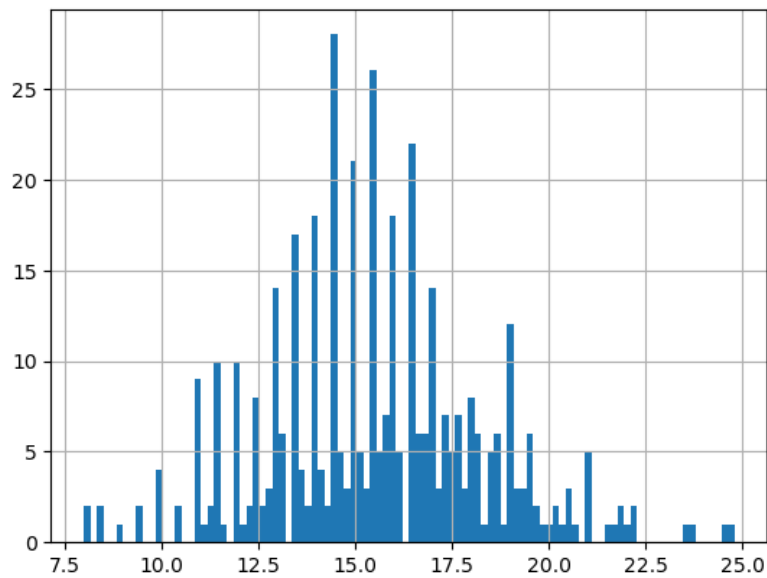
Sep 26, 2023



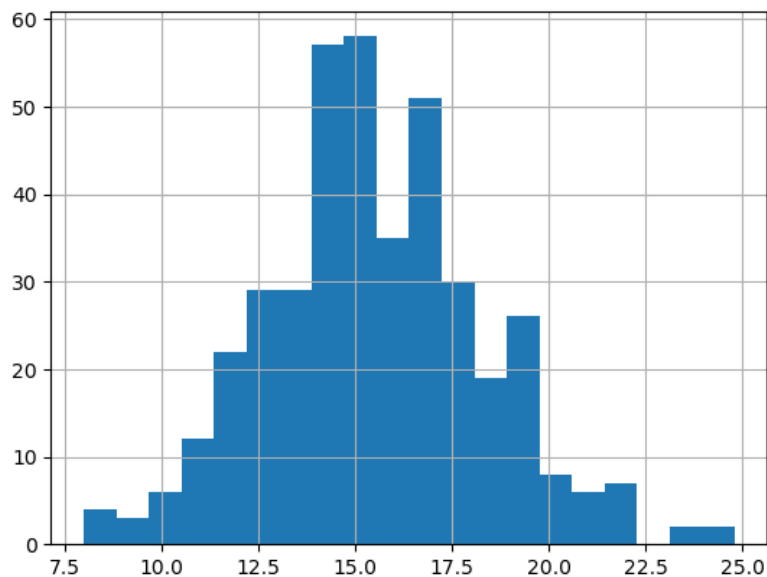
1) The purpose of a histogram is to demonstrate the structure, or distribution of a dataset. When there are too few bins, the histogram is too smooth, and it's difficult to discern the pattern in the data properly, and thus learn much about the dataset. Here's the car acceleration histogram from the previous module, with only 4 bins. We can't tell a lot from this.



2) If a histogram has too many bins, the presentation looks rough and disjointed, and doesn't summarize the distribution optimally. Taken to an extreme, you end up with a 1D scatter plot, which defeats the purpose of a histogram. Here's the same data with 100 bins.



Here's a histogram with 20 bins, which seems about right for this dataset of ~400 points:



← [Reply](#) 👍




Erik Gonzalez (<https://iu.instructure.com/courses/2165942/users/6352173>)

Sep 26, 2023

1. If a histogram has too few bins, then any of the interesting spikes or decreases in frequency may be masked due to not showing the data at a granular enough level. Taking the most extreme example, a histogram with a bin of one shows no useful information about the distribution of data, but instead simply serves as a frequency count. When there is a significant change in frequency at one particular point, having too few bins can lead to important information being masked.

2. Alternatively, having too many bins can lead to the masking of trends that are perhaps more apparent at a higher level of aggregation. For example, imagine a largely uniform distribution of counts between 1 and 100. In this example, let's pretend that the counts are twice as likely to occur between 21 and 30. Leveraging 10 bins would better show this impact than leveraging 100 bins.

← [Reply](#) 



[https://](https://iu.instructure.com/courses/2165942/users/6762945)


Carmen Galgano (<https://iu.instructure.com/courses/2165942/users/6762945>)



Wednesday

1. If the histogram has too few bins, the histogram could lose a certain level of detail. You may not be able to see all the data or may not be able to see the full picture if there aren't enough bins, potentially misleading someone.

2. Too many bins can lead to the histogram being too cluttered, making it difficult to come away with general takeaways. It really doesn't allow viewers to see the patterns and trends the data might be showing.

← [Reply](#) 



[https://](https://iu.instructure.com/courses/2165942/users/6587577)


Yumeng Liang (<https://iu.instructure.com/courses/2165942/users/6587577>)



Wednesday

If we have too less bins, it can mask important details and overgeneralizes data distribution. For example, For data [1, 2, 3, 20, 21, 22], using one bin would combine disparate data points, obscuring underlying patterns and making different data points appear similar. Also, we will miss important features of the data such as multimodality or outliers, potentially leading to incorrect interpretations and conclusions.

If we have too many bins, we will overemphasize minor fluctuations in the data, potentially highlighting noise as significant features. For example, if every unique value in the dataset has its bin, normal minor fluctuations can appear as drastic variations, making the data seem more varied than it truly is.

← [Reply](#) 



(https://

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

Thursday



The number of bins can be chosen by the square root of a number of values.

bin width can be chosen by (Upper limit- Lower Limit)/number of bins

too many bins and too few bins can mislead for sure.

The distribution of data can be shown with the help of the number of bins.

Many data points fall under one bin if there are less bins.

few data points fall under one bin if there are many bins and data distribution looks so rough if there are too many bins.

← Reply



(https://

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

Thursday



1. I think it misses important details and might be confusing to interpret.

2. It probably will be misleading.

← Reply



(https://

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

Thursday



Firstly, if there are too many bins, it may become challenging to analyze the data, and the purpose of visualization might seem to diminish. Conversely, if there are too few bins, it is likely that important data features might be missed. This is because meaningful data could be lumped together within a single, large bin.

← Reply




(https://

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

Thursday



1. if the histogram has too few bins, we can't tell the mean, mode or median very well because height of each bin might be similar.
2. if the histogram has too many bins, many bins would have 0 frequency and some bins have a large frequency to make the plot misleading. And also it's not good to see the trajectory and mean from the plot.

← [Reply](#) 



Onur Tekiner (<https://iu.instructure.com/courses/2165942/users/6758180>)

Friday



Yes, it misleads us because bins show the frequency of particular values. If there are fewer bins, we can get less information about the target value. It makes it harder to see details about the data. Data visualization is supposed to explain the data's facts, not confuse us.


For example, there is an integer number, and you will guess. Which clue would you prefer? 50 is between 1 and 100 and also between 49 and 50.

They're both technically accurate, but the second one gives more details than the first one.

But if there are too many bins, it also makes it difficult for us to see the big picture about data. We should find the optimum bin number depending on axis values.

For example, if we analyze the human lifespan. We don't need 200 bins because we don't need to look at data closer.

Using too many bins makes us look up too close to data, sometimes leading to missing the big picture.

← [Reply](#) 



Shreedeeep Sadasivan Nair (he/him/his) (<https://iu.instructure.com/courses/2165942/users/6813278>)

Friday



1) A histogram too few bins often does not show the appropriate distribution of data, there is a loss of detail as the histogram fails to capture small details about a feature of the dataset also an histogram with few bins gives the impression that there is less variation in the data.

2) Too many bins can also cause the histogram to appear erratic and also leads to overfitting

← Reply 👍



[https://](https://iu.instructure.com/courses/2165942/users/6644229)

Jeevan Deep Mankar (<https://iu.instructure.com/courses/2165942/users/6644229>)



Friday

1. What happens if your histogram has too few bins? Any illustrative examples? When will a histogram with too few bins mislead you?

If the histogram has too few bins then there is a possibility of loss of information or it could be unclear for the viewers to get complete information. It could even depict the wrong information which can even confuse the viewers. Generally bins are there to visualize range of data for example, score of students i.e. 1-100, if we have fewer bins such as just 2 then we will have to divide into 1-50 and 51-100 which will be so less and won't provide the distribution of scores of students.

2. Conversely, what happens if your histogram has too many bins? Any illustrative examples? When will a histogram with too many bins mislead you?

If there are too many bins then it will be difficult to understand as it may not even fit into the chart. And some of the information could be useless. In score of 1-100, if we have bins such as 1-2, 3-4, 5-6, etc ranges then it will be a lot and won't be able to fit into the chart.

← Reply 👍



[https://](https://iu.instructure.com/courses/2165942/users/6443321)

Andrea Chung (<https://iu.instructure.com/courses/2165942/users/6443321>)



Friday

1. Too few bins mean larger bin sizes, which would lack in specificity of patterns or observations from the data. The histogram would become too general to look at distribution of the data. This can be misleading when outliers can influence the shape of the histogram. It may make it seem like there are more outliers than there are. Also, wrong summary statistics can be drawn from the histogram with too few bins.

2. On the other hand, when histogram has too many bins, it can overemphasize the noise. This can mislead by hindering from bringing out meaningful patterns. It can deter from deducing significant peaks from too much fluctuations.

← Reply 👍



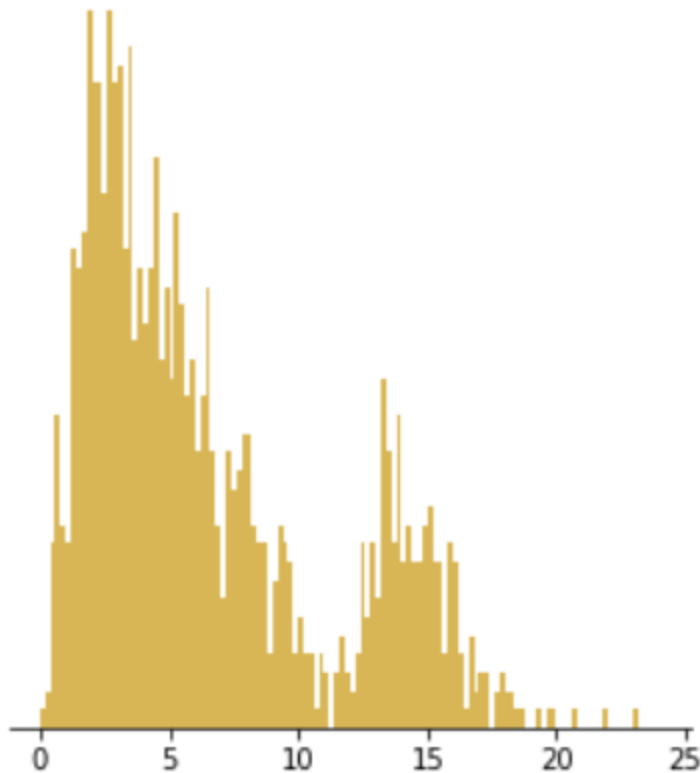
What happens if your histogram has too few bins? Any illustrative examples? When will a histogram with too few bins mislead you?



A histogram with too few bins might mislead by hiding underlying patterns or variations in the data. It can result in a loss of detail and smooth out important features in the data. For instance, if analyzing exam scores of students and using too few bins, one might not notice the distinction between students who barely passed and those who barely failed.

Conversely, what happens if your histogram has too many bins? Any illustrative examples? When will a histogram with too many bins mislead you?

bin size = 0.2



Using too many bins can result in a noisy and cluttered histogram, making it difficult to identify the overall data distribution. An excessive number of bins can overemphasize small fluctuations or random noise in the data. This can lead to an incorrect interpretation of the data's true characteristics. For example, in a financial dataset, too many bins might make it seem like there are numerous price changes when, in reality, they are just minor fluctuations within a stable range

Reference

<https://chartio.com/learn/charts/histogram-complete-guide/#:~:text=If%20you%20have%20too%20many,useful%20pattern%20from%20the%20data.>

← [Reply](#) 



Olufisola Oladipo (<https://iu.instructure.com/courses/2165942/users/6469527>)

Saturday



1. The histogram will not give discernible information as too few bins will obscure the details that can help to read necessary patterns in the data. This histogram will mislead with just about one peak and discernible details are lost in the large bins..
2. When the histogram has too many bins, the data appears rough and contain lots of noise, making it less discernible. This histogram will mislead with several false peaks and troughs because of the bin size.

← [Reply](#) 




<https://iu.instructure.com/courses/2165942/users/6678810>

Saturday



If you use too few bins to create a histogram, it's like trying to see the whole picture with just a few big puzzle pieces. You won't see any small details or patterns in your data. For instance, if you use just one or two bins for ages, you'll get a histogram that shows a simple bar from 0 to 100, and you won't know if there are any specific age groups that stand out. This can mislead you because you might miss important information like whether there are two distinct groups of people in your data.

If you use too many bins, it's like breaking your puzzle into tiny pieces, so each piece represents a single number. Your histogram will look very messy and confusing because there are too many bars, and it's hard to make sense of the data. For instance, if you use hundreds of tiny bins for ages, you'll see a jumbled mess, and it might seem like ages are all over the place.

← [Reply](#) 



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

Saturday




1. Too few bins

If a histogram has too few bins, the data will be aggregated into large groups, which can obscure the underlying distribution of the data. This can lead to misleading conclusions.

2. Too many bins

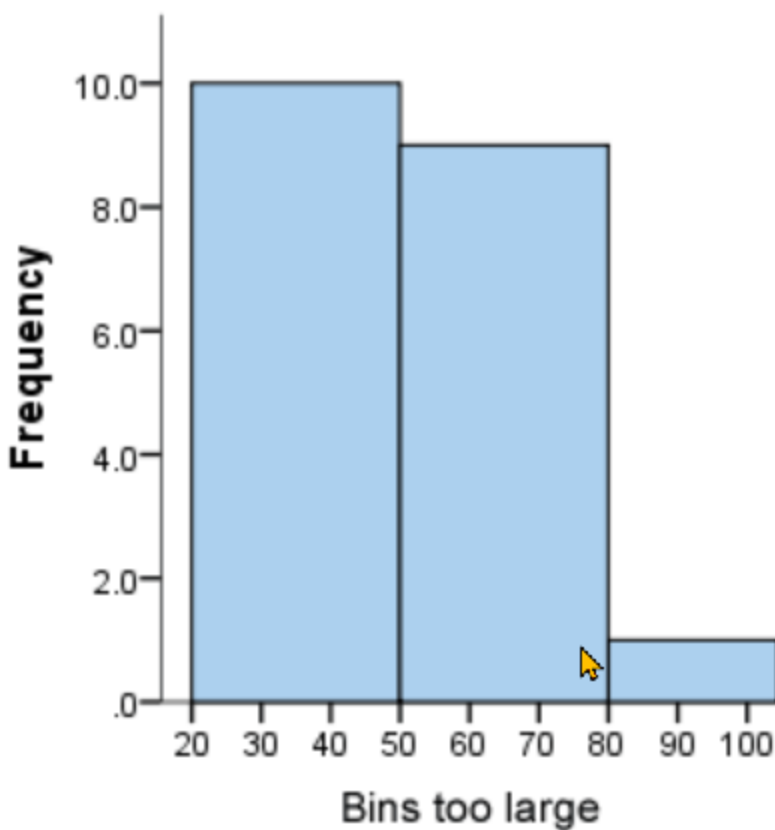
If a histogram has too many bins, the data will be spread out too thinly, and it may be difficult to see any patterns or trends in the data. This is especially true for small datasets.

Edited by **Mothi Gowtham Ashok Kumar** (<https://iu.instructure.com/courses/2165942/users/6683278>) on Sep 30 at 12:02pm

← [Reply](#) 

 <https://iu.instructure.com/courses/2165942/users/6701521> **Robert Perez (he/him/his)** (<https://iu.instructure.com/courses/2165942/users/6701521>)
Saturday

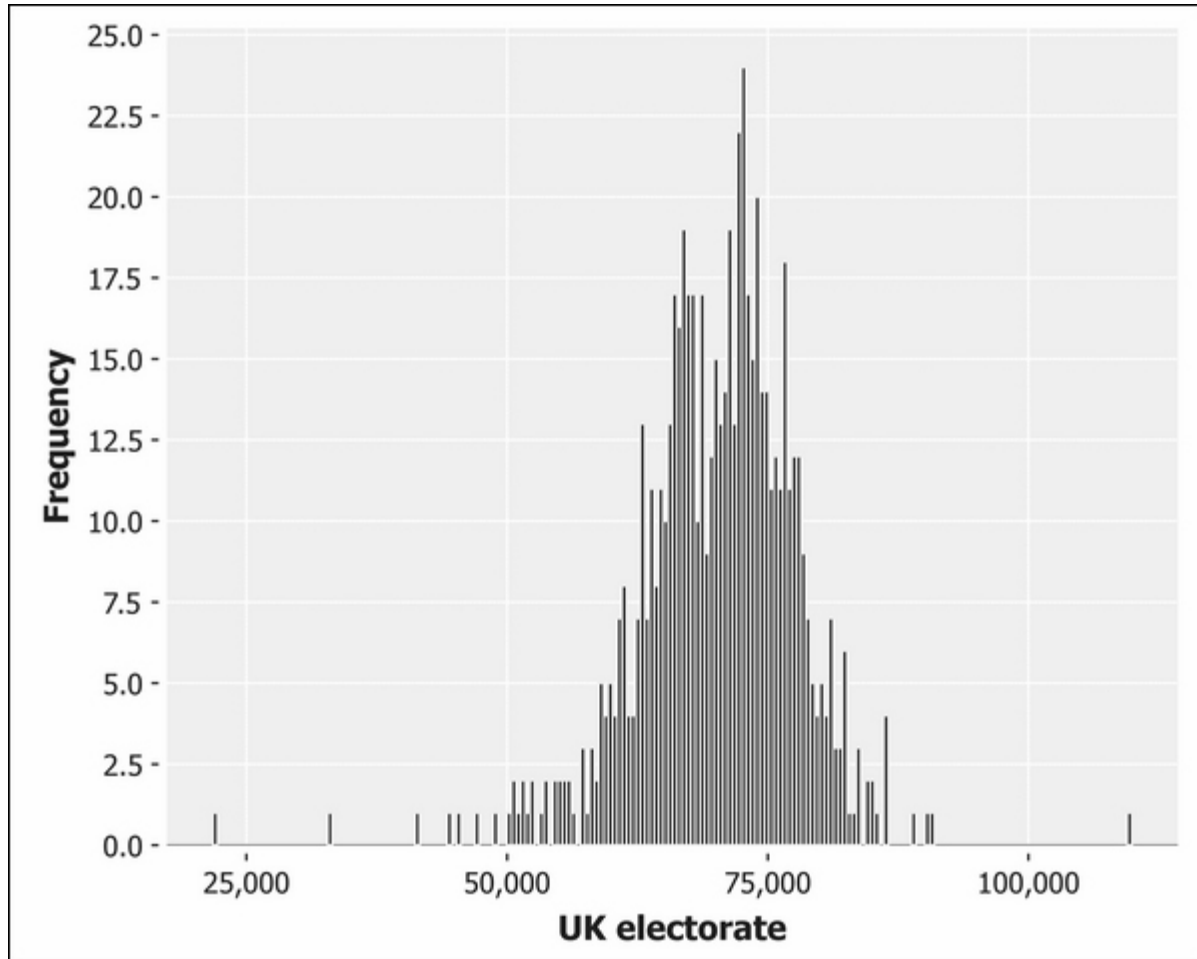
1. If the bins are too large, we lose fine detail about the underlying data. Here is an example:



(Source: <https://statistics.laerd.com/statistical-guides/understanding-histograms.php> 
(<https://statistics.laerd.com/statistical-guides/understanding-histograms.php>)

In the example above, the two largest bars span a range of 30, but we don't know whether there is more nuanced distribution across the 30 units. For example, the first bar spans from 20-50. Perhaps the bulk of these observations fall between 40 and 50. We have no way of knowing that level of detail from these bin sizes.

2. Too many bins creates a situation where the shape of the data is harder to see, particularly around the low ends of the normal distribution bell curve, as show in this example:



(Source:

<https://subscription.packtpub.com/book/data/9781784397180/1/ch01lvl1sec17/histograms> [↗](https://subscription.packtpub.com/book/data/9781784397180/1/ch01lvl1sec17/histograms) (<https://subscription.packtpub.com/book/data/9781784397180/1/ch01lvl1sec17/histograms>)
)

Having fewer bins in this second histogram would allow us to group some of the small values at the extreme ends of the scale while still illustrating the large central distribution of the data. Too many bins could be misleading, therefore, by obscuring the data at the fringes.

[← Reply](#)

[https://](https://iu.instructure.com/courses/2165942/users/6678592)

Mukul Gharpure (<https://iu.instructure.com/courses/2165942/users/6678592>)

Saturday

1. Too few bins in a histogram:

- **Effect: Oversimplifies data distribution, hiding variations and details.**
- **Example: For a dataset of ages, a 50-year bin might mask the distinction between ages like 20s and 40s.**
- **Misleading when: You're seeking specific distributions or patterns; it can suggest a uniform distribution that isn't accurate.**

2. Too many bins in a histogram:

- **Effect: Overcomplicates representation, mistaking random noise for genuine patterns.**
- **Example: For the same age dataset, a 1-year bin can show inconsequential fluctuations.**
- **Misleading when: Trying to grasp general trends; suggests variability not representative of the actual data.**

← Reply 👍



<https://iu.instructure.com/courses/2165942/users/6591977>

Saturday



1. What happens if your histogram has too few bins? Any illustrative examples? When will a histogram with too few bins mislead you?

If there are too few histograms than needed, then the actual visual representation might be less detail as it will not cover enough information about the data distribution. This is over simplification of the data that is being visualized.


Example: If we have data for the 24 hour interval ranges of time interval distributed on persons spent on watching TV and we choose to represent this in two bins(0-12, 12-24), then we might not know on when the peak watch hours or so, as the data is oversimplified.

2. Conversely, what happens if your histogram has too many bins? Any illustrative examples? When will a histogram with too many bins mislead you?

If there are too many histograms than needed, then it is likely that all data elements are plotted. This creates confusion as there is noise as well in the data. Too much data is difficult to interpret. The overall trend cannot be easily spotted out to compare the histograms

Example: If we have data for the 24 hour interval ranges of time interval distributed on persons spent on watching TV and we choose to represent this in 24 bins for each hour, then we might not know the trend and it would be hard to compare the histograms bin. Also, it adds difficult in visual captures.

So, ideal case is to choose right bins, in the above example, choosing bins for every 4 or 6 hours will give us an overall trend in a neat way

[← Reply](#) 

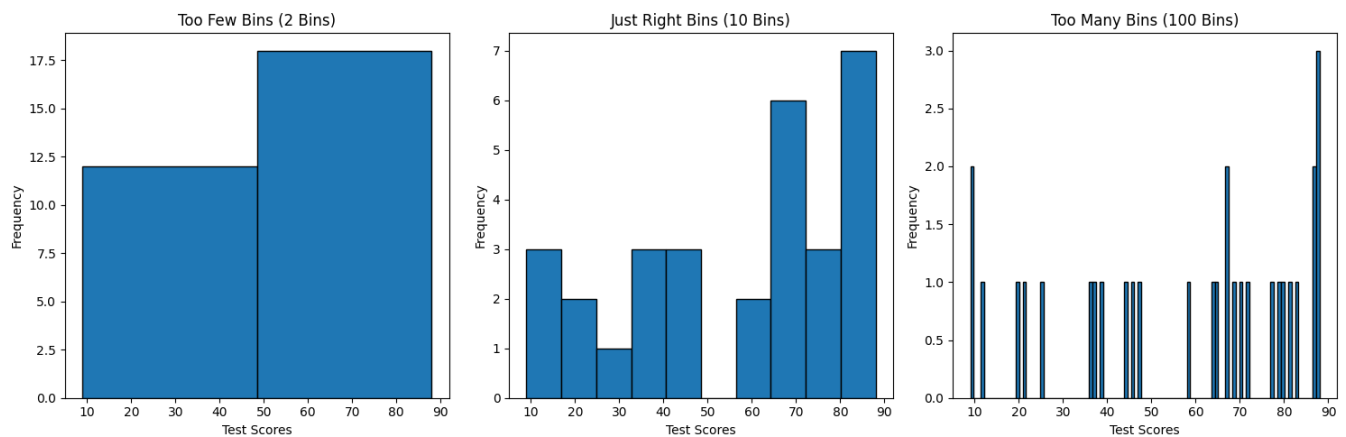


Jash Shah (<https://iu.instructure.com/courses/2165942/users/6684840>)

Sunday

Let us consider with an example of test scores for a class of 30 students, and you decide to create a histogram to see how they performed.

The graphs plotted are used using python code



In this histogram, you can see that some students passed, and some failed, but it doesn't show you the full picture. You can't tell if most students got average scores or if there's a wider range of scores hidden in those two bins. Too few bins can mislead you because it simplifies the data too much.


With many bins:

With so many bins, it becomes challenging to interpret the data. You see lots of tiny bars for individual scores, making it hard to identify any patterns or trends. Too many bins can mislead you because it overwhelms you with detail.

With an exact number of bins

With 10 bins, we strike a balance. we can see that most students scored between 60 and 90, with a few outliers. This histogram gives us a clear overview of the distribution without oversimplifying or overwhelming you.

So, choosing the right number of bins in a histogram is like finding the "sweet spot" that provides enough detail to understand the data's shape and trends while remaining clear and informative.

← [Reply](#) 

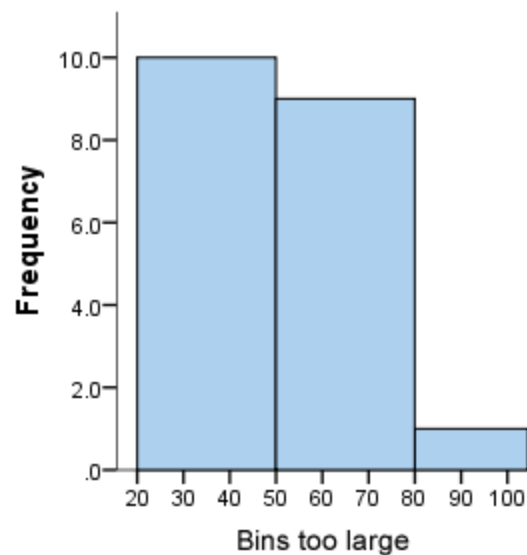
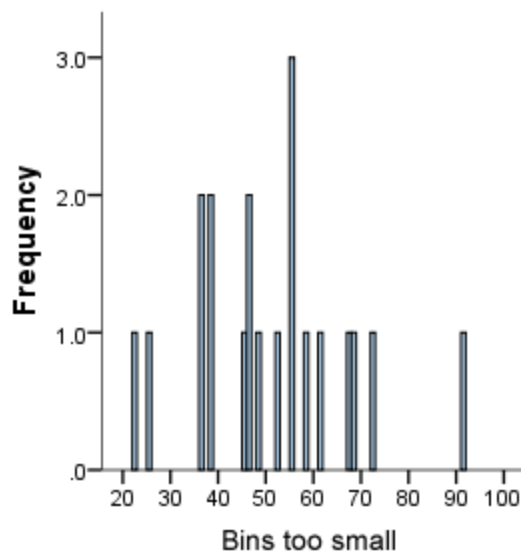
 **Simon Driver** (<https://iu.instructure.com/courses/2165942/users/6818242>)

Sunday

1. If there are too few bins, you lose oversight of the data, and everything is clustered together. You end up basically with just a few large blocks of data which are difficult to interpret, lose nuance, and are not very helpful. A histogram with too few bins can be misleading when we are trying to find out trends or patterns within a narrow range.

2. If the histogram has too many bins, it becomes less helpful because there will be a lot of white space on the chart; not every bin will have an observation associated with it, perhaps, plus there may be so few observations per bin that it is hard to make any comparisons or draw any conclusions.

A great example of both cases is this:



Source: <https://statistics.laerd.com/statistical-guides/understanding-histograms.php>

← [Reply](#) 

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

Sunday

1. If a histogram has too few bins, it can be a bit reductive: it won't show the complexity of the data since the bins represent area across a certain span of values. You may miss a significant "high" or "low" in the data. It could be misleading if half the data in a bin are lower in frequency and the other half are higher. Since the bin represents area, the height could appear like an average between the two, but principally, it would look like there is no change across that bin when in reality there is quite a bit of change. I don't have any specific examples, but if your X bins are original at 1, 2, 3, and 4 (and the frequency at each is 5, 20, 30, 2, respectively), it's clear how blending that data into a single bin of 1-4, though it shows area, would not show the more interesting spike at points 2 and 3.

2. Though I think I'd prefer a histogram with too many bins to too few bins, a downside here could be more along the lines of difficulty to view, or, depending on your view of the data, it may not be focused enough, and you can miss the story in the data. From an analyzing perspective, I wouldn't mind "too many" bins, though if I'm trying to tell a story, it could be confusing or cumbersome to show so much minute change rather than an intentional pattern.

[← Reply](#) [https://](https://iu.instructure.com/courses/2165942/users/5444499)**Maria Klein** (<https://iu.instructure.com/courses/2165942/users/5444499>)

Sunday

Too few bins may disguise difference in frequency from one end of the bin range to the other. For example, when plotting mortality rate in a population by age, a histogram with only two bins representing 0-50 years and 50+ years would mask that mortality has a u-shape, with higher rates in the youngest and oldest ends of the spectrum and lowest rates toward the middle.

Too many bins can also be misleading because it disguises the true concentration of the frequency ratio relative to adjacent bins. For example, say we are looking at a histogram of number of disease cases by age and we represent each year of age with 1 bin. There are 0 cases for ages 0, 1, 2, and 3 and 10 cases at age 4. There are 20 cases each at age 5, 6, 7, 8, and 9. You may only notice the doubling of cases between ages 4 and 5. However, by reducing the number of bins to 2, ages 0-4 and ages 5-9, we see there is actually a 10-fold increase in cases between these ages ranges, from 10 cases to 100.

[← Reply](#)



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

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



Sunday

1. When my histogram has too few bins, I will consider several situation. Firstly, it will make me think about the distribution of the data. Is this histogram hiding some important information? Secondly, I will think about the settings in histogram. Do I set the width or other parameters right? Too few bins in histogram will mislead me when there are subtle variations in the dataset.

2. Similarly, too many bins in my histogram also lead to misunderstanding. Too many bins in the histogram cannot show the real distribution of dataset either. It is possibly that I can merge some groups to get a better figure. For example, when I use histogram to show the distribution of age, it will give me too many bins in the figure. That's not what I want. A histogram with too many bins can mislead me when the data has a simple, clear distribution. The excessive number of bins can create a false impression of complexity or randomness, which may not actually exist in the data.

← [Reply](#)



([https://](https://iu.instructure.com/courses/2165942/users/6056428))

Adam Hume (<https://iu.instructure.com/courses/2165942/users/6056428>)



Sunday

If there are too few bins, there can be loss of detail. The histogram tends to not accurately represent the data when there are too few bins due to over clumping of the data. On the other hand, if there are too many bins the data may also mispresent the data by being too jagged. Single records could be outliers which could easily skew small groupings.

← [Reply](#)



([https://](https://iu.instructure.com/courses/2165942/users/6682743))

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



Sunday

1. A histogram with too few bins can lead to a loss of important trends in the visualization. Decision-making or drawing conclusions based on a misleading histogram can lead to incorrect insights. For example, if the histogram hides the presence of outliers, you might underestimate their impact on a dataset.

2. A high number of bins can amplify random noise or fluctuations in the data, making it challenging to distinguish meaningful patterns from noise. This can lead to misinterpretations and incorrect conclusions about the data.

← Reply 👍



<https://iu.instructure.com/courses/2165942/users/6694681>

Yesterday



The number of bins in a histogram is an important parameter that can significantly impact the interpretation of the data. with too few bins there may be loss of information, there can be misleading interpretation etc.

Example: if there are too less bins in histogram of dataset of students grades of a final year students it might look like most students scored in the same range, hiding variations in performance.

With too many bins the dataset would show some noise or is likely to overfit , there may be loss of efficiency and generalization .

Example: with too many bins you might see small fluctuations in grades that are not practically meaningful and might lead you to incorrect conclusions about the data's distribution.

← Reply 👍



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

Yesterday



Too Few Bins:


- Imagine you're making a bar chart to show how many people have different heights.
- If you use too few bars (bins), like just three, it's like putting people into three groups: short, medium, and tall.
- This might hide the differences in heights within each group, making it look like everyone in the same group has the same height. That can be misleading.

Too Many Bins:

- Now, picture you have too many bars, like hundreds, in the same chart.
- It's like dividing people into super tiny height ranges, say 5.25 feet to 5.26 feet, and so on.

- This can make the chart too detailed and hard to read. You might think there are patterns or gaps in the data that aren't real, just because you've split it into too many pieces.

So, getting the right number of bins is like finding the sweet spot. Not too few, so you don't miss important details, and not too many, so you don't create confusion with too much detail.

← Reply 



<https://iu.instructure.com/courses/2165942/users/6825193>

Yesterday



1} If the histogram had too few bins, it would lead to loss of details. For example, suppose we create a histogram for test score. If we only create two bins such as one for score above 50 and another for below 50, it will create confusion and will not show variations or patterns that will make the illustration irrelevant. Such a histogram can mislead you by not showing the full distribution of the data.

2} If the histogram had too many bins, it would not show clarity and hard to understand. For example, consider a data set for covid cases according to people's ages. if we create 100 bins, each representing a year age range, then this histogram will have many bars with low frequencies, and it may appear jagged and noisy. A histogram with too many bins can mislead you by numerous variability in the data. It can make the data seem more complex than it actually is.

← Reply 



<https://iu.instructure.com/courses/2165942/users/6694525>

Yesterday



For Qn1 - I think if there are too few bins it would become hard to identify the distribution because the histogram could get choppy. This would happen especially if there are not too many data points.

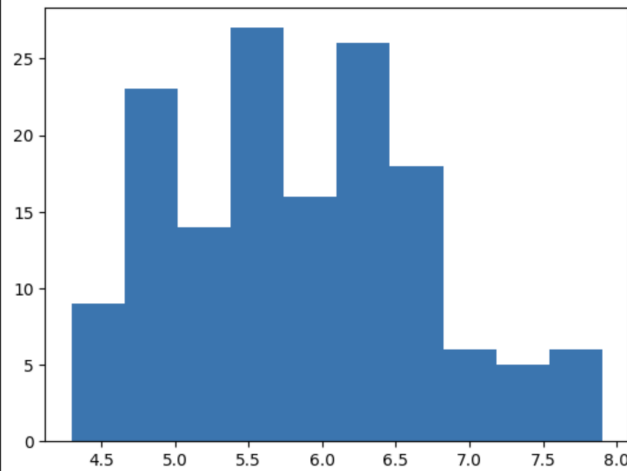
Qn2 - Here we may miss out ranges that are influencing other ranges due to their high frequency. Important inferences can be overlooked if bins are larger.

← [Reply](#) 👍

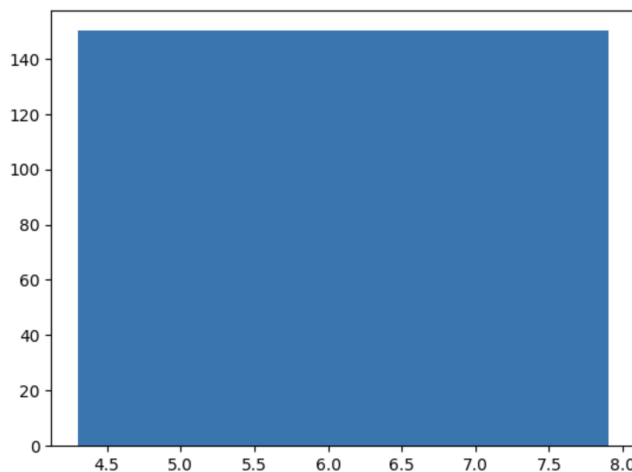


Akash Patil (<https://iu.instructure.com/courses/2165942/users/6699404>)

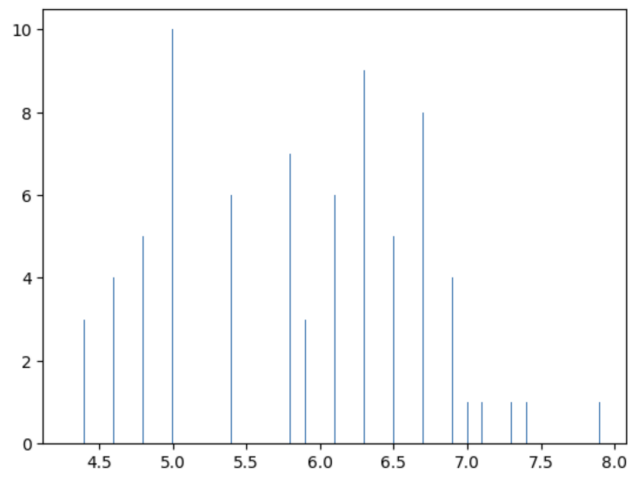
12:51am



1. If you have too few bins, the histogram may oversimplify the data's distribution, and you might lose important details and patterns in the data. For example we plot exam grades using just 1 bin, we will not see any distribution and just 1 bar would be visible. The histogram may make it seem like all data points fall into a single category, which can be misleading when there is actually more variation in the data.



1. Conversely, if we use too many bins, the histogram may appear overly detailed, even capturing noise in the data. For example, if you create a histogram with 1000 bins for a dataset with only a few data points, it may show spurious patterns that don't represent the underlying distribution. It might put more focus on random variations in the data instead of the general underlying trend.

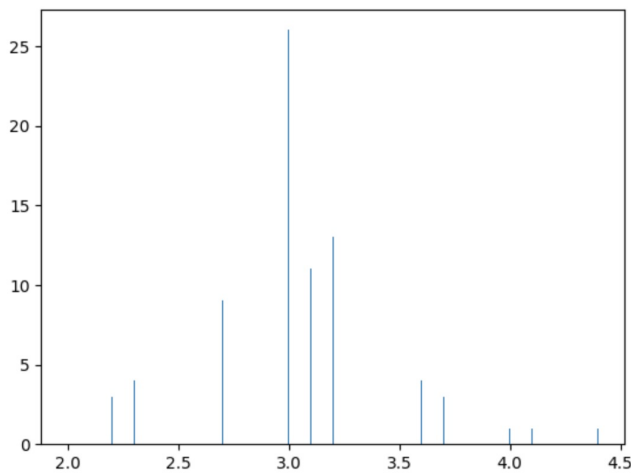
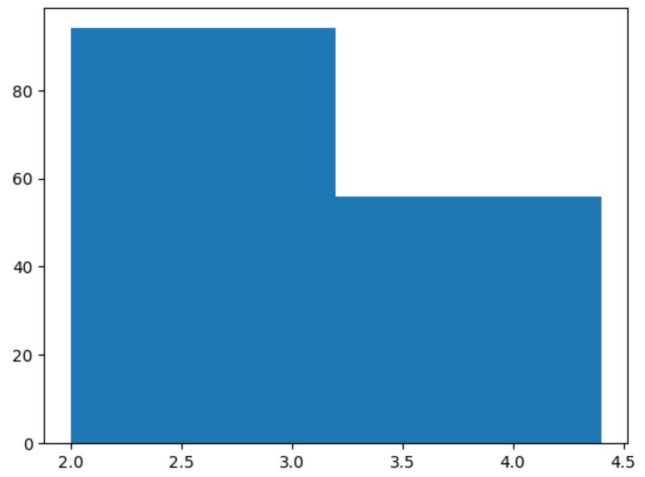
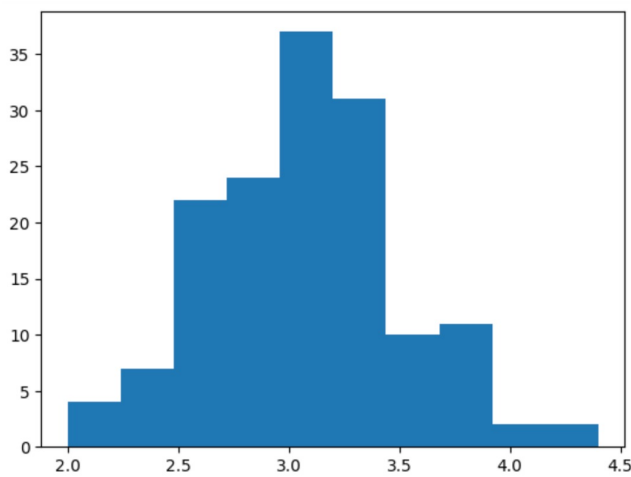


← Reply 👍



Shruthi Senthilmani (<https://iu.instructure.com/courses/2165942/users/6688148>)

12:54am



Appropriate bins
More bins

Less bins

1. If the histogram has too few bins, then it might not capture all of the data correctly and may lead to misrepresentation of data quantities. It will result in a more binned representation that cannot differentiate between data points within each bin. Also, it might simplify the data distribution and many points will have the same bin as it will be generalized due to less number of bins. If we use a histogram of say, the heights of students, even though their values are completely unique, and if we use fewer bins, we cannot see a distribution in the height data over a large variety of students. The histogram may show a broad range of height without revealing finer details, making it difficult to identify whether there are distinct subgroups of tall and short students and this can be misleading when the data is actually varied and has different subgroups.

2. If we use too many bins, the histogram may look over detailed. It might in a few cases capture noise in the data and don't reflect the true characteristics of the data. For example, if we create a histogram with 1000 bins for the same height data with fewer data points, it can be challenging to discern any meaningful patterns and this cannot represent the underlying distribution. We might see spurious fluctuations that are not representative of the true height distribution and this may result in the loss of important details of data.

 [Reply](#) 

 <https://iu.instructure.com/courses/2165942/users/6696028> **Anudeep Devulapally (he/him/his)** <https://iu.instructure.com/courses/2165942/users/6696028> 

11:01am

1.

If a histogram has too few bins, it can obscure important features of the data distribution.

example, if the data is bimodal (i.e has two peaks), a histogram with too few bins may show only one peak.

2.

If a histogram has too many bins, it can make it difficult to see any underlying patterns in the data.

example,collecting data on the weights of people in a population. If we create a histogram with too many bins, we may not be able to see that the distribution of weights is approximately normal

 [Reply](#) 



The number of bins in a histogram is a crucial parameter, and getting it right is essential for an accurate representation of data. Here's why we should care about the number of bins and how it can affect the interpretation of data:

1. Too Few Bins:

- **Loss of Detail:** If you use too few bins, the histogram may oversimplify the distribution, leading to a loss of detail. The resulting bars may be too broad, making it difficult to identify smaller peaks or variations in the data.
- **Misleading Shape:** A histogram with too few bins can mislead you by smoothing out important features in the data. For example, it may make a bimodal distribution appear unimodal or mask the presence of outliers.

Illustrative Example: Imagine you're plotting the distribution of exam scores for a class of students. If you use too few bins, you might end up with a single, wide bar that doesn't show the distinction between students who scored exceptionally well and those who barely passed.

2. Too Many Bins:

- **Noise and Overfitting:** Using too many bins can result in a noisy histogram that captures random fluctuations in the data, rather than meaningful patterns. This is akin to overfitting in machine learning, where the model fits the noise in the data rather than the underlying trend.
- **Loss of Generalization:** With too many bins, you risk losing the ability to generalize the data's overall distribution. Instead, you may focus on small, inconsequential variations.

Illustrative Example: Suppose you're creating a histogram of daily temperature fluctuations over a year. If you use an excessive number of bins, you might have a jagged, irregular histogram that doesn't convey the overall pattern of seasonal temperature changes.

In practice, determining the right number of bins often involves some trial and error, as well as considering the nature of the data and the insights you want to gain. Techniques like Scott's rule or Freedman-Diaconis' rule can provide guidelines for choosing an appropriate number of bins based on data characteristics like sample size and range.

Finding the balance between too few and too many bins is crucial for creating a histogram that effectively communicates the underlying data distribution without oversimplifying or overcomplicating it.

← Reply



Renu Jaiswal (<https://iu.instructure.com/courses/2165942/users/6704404>)



2:12pm

The right number of bins in a histogram is crucial because it affects how we understand data. If we have too few bins, the histogram oversimplifies the data. For example, if we're looking at people's ages and use only two bins, one for children and one for adults, we lose the details of different age groups, like teenagers or seniors. This simplification can mislead us by hiding important patterns in the data.

On the other hand, if we have too many bins, the histogram becomes too detailed. Try creating a histogram of people's ages with a separate bin for each year; it would be overwhelming. Too many bins can also mislead us because we might see patterns or variations that are just random fluctuations in the data. So, choosing the right number of bins helps us find the balance between simplicity and detail, allowing us to understand our data better.

← Reply



Shreyas Aswar (he/him/his) (<https://iu.instructure.com/courses/2165942/users/6676727>)



3:20pm

1. Having too few bins would make the histogram more susceptible to misleading interpretations and loss of patterns in the data. Say we are drawing a histogram for age groups and we decide to have only 2 bins with first bin having age groups till 50's and second bin having from 50's to 90's. It won't be possible to distinguish between 2 age groups like 40's and 60's. This makes the histogram not helpful in understanding the crucial information which might be possible to see with higher number of bins. This could also mislead as one group in the first bin might have a few outlier which makes us conclude that first group is more influencing.

2. Histograms with too many bins are very sensitive to the outliers. Having too many bins is capturing all the points and which means there could be a lot many fluctuations in the visualization. However it might be the case that it is just an outlier and the data does not follow the fluctuation pattern at all. It could mislead as we might conclude that the data is too complex due to having too many bins. We might also get misled and spend more time in analyzing variations which does not really matter.

← Reply





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

3:51pm



1. What happens if your histogram has too few bins? Any illustrative examples? When will a histogram with too few bins mislead you?

Using too few bins in a histogram can lead to misleading interpretations due to less data. The histogram doesn't appear detailed when there aren't enough bins, which makes it difficult to spot trends, variances, or outliers in the data. For example - a year-long dataset's temperature changes could go undetected by a histogram with only a few bins, making it difficult to analyze seasonal trends.

2. Conversely, what happens if your histogram has too many bins? Any illustrative examples? When will a histogram with too many bins mislead you?

Too many bins in a histogram can introduce noise, complicate interpretation, and exaggerate data variability. For instance, making a histogram with too many monthly expense bins may make it difficult to understand significant expense trends and to efficiently compare different expense categories.

[← Reply](#)



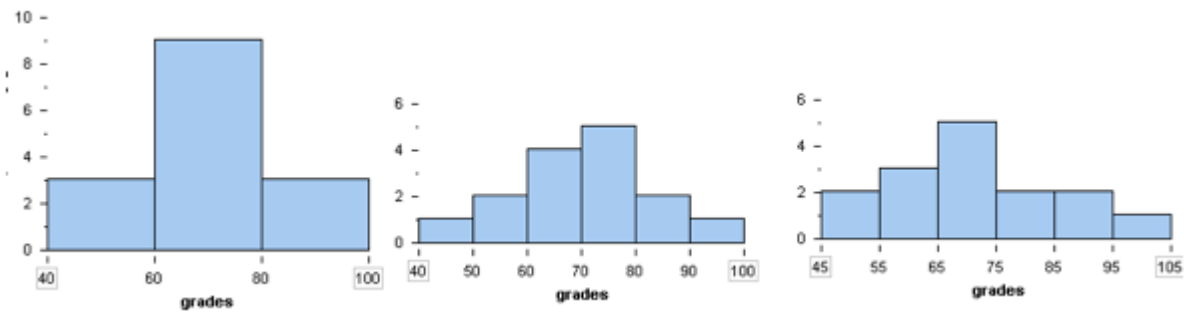
<https://iu.instructure.com/courses/2165942/users/6762824>

3:52pm



1. If a histogram has too few bins, it can mislead the reader into a false pretense that the data is as accurate as it would like to depict. As seen in the image below for the grade example pulled from "<https://courses.lumenlearning.com/wm-concepts-statistics/chapter/histograms-2-of-4/>" you can see that if you have fewer bins you can be misled into thinking that grades were not as bad since most fall between 60-80.

2. If a histogram has too many bins it can also be misleading and give too much information that becomes incomprehensible. As seen in the image below for the grade example pulled from "<https://courses.lumenlearning.com/wm-concepts-statistics/chapter/histograms-2-of-4/>" the chart all the way on the right shows the scores broken down into increments of 5 and skewing the grades to the left looking a little worse than if you have the buckets separated by 10's.



Edited by [David Rosenthal \(https://iu.instructure.com/courses/2165942/users/6762824\)](https://iu.instructure.com/courses/2165942/users/6762824) on Oct 3 at 3:53pm

← [Reply](#)

 <https://iu.instructure.com/courses/2165942/users/6688908> **Jerin Easo Thomas** <https://iu.instructure.com/courses/2165942/users/6688908> 3:54pm

1. Histogram with Insufficient Bins

A histogram with too few bins might oversimplify the structure of the data, perhaps missing significant subtleties and trends. This overly polished representation may cause misunderstandings.

Consider the following scenario: You have a dataset of students' exam scores ranging from 0 to 100. If you only utilize two bins (0-50 and 51-100), you may miss out on the comprehensive distribution of results. You won't be able to see if there are clusters of scores around specific values, such as 75 or 90, or if the scores are distributed uniformly over the range.

Misleading Scenario: In the context of exam scores, employing two bins may lead us to believe that there are just two categories of students: those who scored less than 50 and those who scored more than 50. We lose the granularity of performance levels and are unable to target specific areas for improvement.

2. Histogram with an Excessive Number of Bins

Too many bins in a histogram can generate a lot of noise and make the underlying pattern of the data difficult to comprehend. Granularity might draw attention to random changes rather than actual underlying trends.

Continuing with the students' scores example, if you use a bin for each individual score (i.e., 100 bins for scores ranging from 0 to 100), the histogram may become overly cluttered. Each bar may represent only a few pupils or perhaps individual individuals, making broader patterns difficult to see.

Misleading situation: In this high granularity situation, we may focus too much on single scores and miss out on larger trends. For example, if three students received a 78 and no students

received a 77 or 79, we may mistakenly focus on the "peak" at 78 while failing to recognize a general trend of performance in the 70s range.

Why Should We Be Concerned?

It is critical to select the appropriate number of bins in order to accurately reflect the data distribution. A small number of bins can hide features, while a large number of bins can generate noise. Carefully selecting the number of bins aids in achieving the right balance for effectively understanding the underlying patterns and trends in the data, leading to informed decisions and actions.

← Reply 👍



Alan Varkey (<https://iu.instructure.com/courses/2165942/users/6681532>)

4:53pm



1. Too Few Bins:

Loss of Detail: When you use too few bins, the histogram may oversimplify the data, and you might lose important details in the distribution. The resulting histogram will be too smooth and may not capture local variations in the data.

Misleading Shapes: A histogram with too few bins can mislead you by smoothing out the underlying distribution. For example, in a dataset with multiple modes (peaks), a histogram with too few bins might merge these modes into a single peak, giving an inaccurate representation of the data's structure.

Inadequate for Small Samples: If you have a small sample size, using too few bins can make it challenging to detect patterns or variations in the data. The histogram may not provide enough granularity to make informed decisions.

Illustrative Example: Imagine a dataset of test scores where you use only one or two bins. The histogram will show you little more than a rough average, and you won't see the distribution of scores within specific score ranges.



2. Too Many Bins:

Noise and Fluctuations: Using too many bins can result in a noisy histogram, where individual fluctuations in the data appear as peaks and valleys. This can make it challenging to discern the actual underlying pattern or distribution.

Overemphasis on Outliers: An over-binned histogram may highlight outliers excessively, making it appear as though they are more significant than they are. This can mislead you into thinking that the data has extreme values when it may not.

Loss of General Patterns: With too many bins, you might lose the ability to see broader trends or general patterns in the data. The histogram can become too fragmented.

Illustrative Example: Consider a dataset of people's ages where each age is treated as a separate bin. The resulting histogram will show every age as a spike, making it difficult to understand age distribution trends within broader age groups.

 [Reply](#) 

○



Aditya Sanjay Mhaske (<https://iu.instructure.com/about/6692144>)



5:14pm

- The number of bins in a histogram is important because it affects the accuracy and readability of the visualization. If there are too few bins, the histogram will not capture the accurate distribution of the data. If there are too many bins, the histogram will be noisy and difficult to interpret.
- If a histogram has too few bins, the data will be lumped together, making it difficult to see any patterns or trends. This can lead to misleading conclusions about the data.
- A histogram with too few bins can mislead you if it hides important patterns or trends in the data.
- If a histogram has too many bins, the data will be spread out over many small bars, making it difficult to see any overall patterns or trends. The histogram will also be more susceptible to noise.
- A histogram with too many bins can mislead you if it makes it difficult to see the overall shape of the distribution.
- The best number of bins will depend on the specific data set and the purpose of the visualization. Start with a small number of bins and then increase the number of bins until the histogram is easy to read and interpret. One should also avoid using too many bins, as this can make the histogram noisy and difficult to interpret.

 [Reply](#) 