Click Here

All BlogsStatistics ResourcesFrequency PolygonData VisualizationBasics of Data VisualizationData VisualizationDot PlotStem And Leaf PlotChart TypesFrequency PolygonControl ChartTree DiagramVenn DiagramData VisualizationGET: WSM ALL COURSES ACCESSTable Of ContentsA frequency polygon in statistics is a graph that displays the distribution of a set of continuous data. It aims to visually represent a group of consistent data issuance, making it easier to interpret and analyse by plotting the frequency of each interval on the vertical axis and the midpoint of each interval on the horizontal axis.It thus shows the overall pattern of the data distribution, highlighting any trends or patterns that may exist. It can be used in various fields, including statistics, economics, biology, and social sciences, to analyze and interpret data sets. They are handy for identifying outliers, detecting patterns and trends, and comparing data sets.Frequency polygons are line graphs representing a continuous dataset's frequency distribution. We can get it by dividing the data into intervals or bins. This calculates the frequency for each interval and plots the midpoint of each interval.It can help locate patterns and trends in large datasets, detect outliers, and allow easy comparisons between datasets.They are handy for displaying data sets with many observations or values. They are helpful in various fields, including statistics, economics, biology, and social sciences.The frequency polygon represents the frequency distribution of continuous data graphically. It was first introduced by English statistician Karl Pearson in the late 19th century. He proposed it based on the work of other early statisticians such as Francis Galton and Adolphe Quetelet.Its relevance lies in its ability to visually represent data, allowing for more straightforward interpretation and analysis. As a result, they are a valuable tool in statistics, helping researchers to identify patterns and trends in large data sets.One of the main advantages of it is that they are handy for displaying data sets with many observations or values. By dividing the data into equal intervals and plotting the frequency of each container, the polygon provides a clear and concise representation of the data distribution.These are also helpful in identifying outliers or anomalies in the data set, as these are in the form of points that fall outside the normal range of the distribution. Again, this makes detecting and investigating unusual or unexpected data points easier.Drawing a frequency polygon graph involves several steps:Organize data: Start by organizing your data into a frequency table showing the observations or values in each interval or bin. Determine the interval: Decide the break or bin size used to group the data. This should be within the range of values in the data set and the number of observations. A general rule is to have around 10-20 intervals, each equal. Calculate the midpoint of each interval: Once you have determined the interval size, calculate the midpoint by taking the average of the upper and lower boundaries of the gap. Calculate the frequency for each interval: Using the frequency table, calculate the frequency or number of observations for each interval. Plot the points: Plot the midpoint of each interval on the horizontal axis and the frequency for each interval on the vertical axis. Then, connect the ends with a straight line to form the polygon.Let us understand the topic better with the help of some examples:Suppose Harry is surveying the number of hours people spend using their smartphones daily. He collects data from a sample of 50 people and organizes it into the following frequency table:Interval (hours)Frequency0-151-282-3153-4124-565-626-717-81To visualize this data, he could create a frequency polygon by plotting the midpoint of each interval on the horizontal axis and the frequency for each interval on the vertical axis.A recent example of a frequency polygon in the news was the visualization of the COVID-19 vaccination rates in different countries. In addition, many news organizations and public health agencies have used polygons to show the distribution of vaccination rates across other age groups, regions, and ethnic groups. The US Centers for Disease Control and Prevention used polygons to show the distribution of vaccination rates by age group, with a clear peak in the age group of 65 and older.Some of the advantages of frequency polygon are:Easy to interpret: These provide a clear visual representation of the data distribution, making it easy to interpret and analyze. Highlights patterns and trends:It can help patterns in large datasets that may be difficult to identify by simply looking at the raw data. Displays large datasets: These are particularly useful for displaying large datasets, as they group the data into intervals or bins, making it easier to see the overall pattern. Detects outliers: It can help detect outliers or unusual data points affecting the overall data distribution. Allows for comparison: These compare different datasets, allowing for insights and comparisons between other groups or populations.Some of the disadvantages of frequency polygon are:May oversimplify the data: These group data into intervals or bins, which may oversimplify the data and hide essential details or nuances. Can be misleading: If the intervals or bins are not chosen carefully, it can be deceptive and lead to correct conclusions. May not be suitable for small datasets: These are best for larger datasets, as they may need to provide more detail for smaller datasets. Subjective binning: Choosing the size and number of intervals or bins is personal and can affect the resulting polygon, potentially leading to biased or inaccurate interpretations. Limited to continuous data: These usually represent constant data, making them unsuitable for datasets with discrete data points or categories.Some differences between histograms and frequency polygons:Data representation:Histograms represent data using bars, while the latter represent data using lines. Type of data: Histograms usually display continuous data, while frequency polygons can be used for continuous and discrete data. The shape of distribution:Histograms are best for displaying the form of the data distribution. At the same time, frequency polygons show the trends and patterns in the data. Bin size: In histograms, the width of the bars represents the bin size, while in frequency polygons, the bin size is not represented visually. Range of data: Histograms typically show the full range of the data, while frequency polygons may only show a portion of the data range depending on the chosen bin size and intervals.Points comparing frequency polygons and frequency curves:Data representation: Frequency polygons represent data using straight line segments, while frequency curves represent data using a smooth curve. Type of data: Frequency polygons are typically used for displaying continuous data, while frequency curves can be used for serial and discrete data. Bin size: Frequency polygons require intervals or bins to be defined, while frequency curves do not require intervals or bins. Continuity of data: Frequency curves provide a continuous representation of the data, while frequency polygons may not clearly show the data's continuity. Smoothing: Frequency curves can be smoothed to reduce the impact of outliers, while frequency polygons cannot be filed.Some differences between frequency polygons, line graphs, and ogive graphs are as follows:Data representation: Frequency polygons represent data using straight line segments, and line graphs represent data using a continuous line. In contrast, ogive graphs represent data using consecutive connected lines. Type of data: Frequency polygons and ogive graphs are typically used for displaying continuous data, while line graphs can be used for serial and discrete data. Y-axis: Frequency polygons and ogive graphs typically use the frequency or cumulative frequency on the y-axis, while line graphs may use a range of variables on the y-axis. X-axis: All three graphs use a continuous variable on the x-axis, typically with intervals or bins defined for frequency polygons and ogive graphs. Smoothing: Line graphs can be ground to reduce the impact of noise or outliers in the data, while frequency polygons and ogive graphs cannot be filed. A histogram in which most of the data falls to the right of the graph's peak is known as a right-skewed histogram. It is also known as a positively skewed histogram. A right-skewed histogram has a definite relationship between its mean, median, and mode which can be written as mean > median > mode. Let us learn more about it in detail in this article. What is a Right Skewed Histogram? A right-skewed histogram or a positively skewed distribution is a type of histogram that is not symmetrical and in which the peak of the graph lies to the left of the middle value or the median. Let us draw a histogram showing the given data: 2, 3, 3, 4, 3, 8, 3, 3, 2, 5, 7, 4, 5, 4, 6.When we draw a histogram showing the above data, we found that it is skewed to the right. Or we can say that a portion of the right side seems to be chopped or cut off. This is an example of a right-skewed histogram. Now, let us learn about the relation between mean, median, and mode of a right-skewed histogram in the next section.Right Skewed Histogram Mean Median Mode Before learning about the mean, median, and mode of a right-skewed histogram, let us quickly go through the meaning of these terms:Mean: It is the average of the data found by dividing the sum of the observations by the total number of observations.Median: It is the middle value of the data or the observation that lies in the mid or center of all the given values.Mode: It is the most frequently occurred observation in the data. For example, in the above example, 3 is the mode as it has occurred 5 times which is the maximum as compared to the other values.In a histogram skewed to the right, the value of mode is to the left of the value of the median as in a right-skewed histogram the peak of the graph (mode) occurs to the left of the center of the graph (median). And, most of the values or observations fall on the right side of the graph, therefore the value of mean comes to the right of the center or the median of the right-skewed histogram distribution. Therefore, the relation between the mean, median, and mode of a right-skewed histogram is given as Mean > Median > Mode. In other words, we can say that in such a histogram, the value of mean is the highest followed by the value of median and then the mode.Observe the graph of a right-skewed histogram given above representing its mean, median, and mode.Right Skewed Histogram Vs Left Skewed There are three types of histogram possible based on its skewness which are symmetrical histogram, right-skewed histogram, and left-skewed histogram. Let us do an activity to understand the difference between the right-skewed and left-skewed histogram. Draw a histogram to represent the following data: 5, 3, 3, 6, 4, 3, 5, 4, 7, 3, 3, 5, 3, 6, 4, 3, 4, and then draw a histogram to represent the following data: 7, 4, 6, 7, 5, 7, 6, 3, 4, 7, 5, 6, 6, 7, 7, 5, 7. Observe the two histograms given below to represent each of the above data sets.The first data is taking a form of a right-skewed histogram, while the second data is taking the form of a left-skewed histogram. Let us learn some basic differences between a right-skewed histogram and a left-skewed histogram in the table below.Right Skewed HistogramLeft Skewed HistogramAlso known as a positively skewed histogram.Also known as a negatively skewed histogram.The relation between right skewed histogram mean, median, and mode is Mean > Median > Mode.The relation between mean, median, and mode in a left-skewed histogram is Mean < Median < Mode.The peak of the graph lies on the left side of the center.The peak of the graph lies on the right side of the center. Related ArticlesCheck these interesting articles related to the concept of a right-skewed histogram in statistics.Frequency PolygonsLine GraphBar GraphExample 1: Draw a right-skewed histogram to represent the following data: 9, 7, 8, 7, 6, 7, 8, 10, 8, 7, 6, 8, 8, 7, 10, 6, 7, 7, 8, 9, 6, 7.Solution: To represent the given data, the histogram skewed to the right is shown below:Example 2: What is the mode of the right-skewed histogram given in Example 1?Solution: The mode of a right-skewed histogram can be found out by looking at the peak of the graph. In the given histogram, the value of mode is 7 as it is the most repetitive observation.View Answer >go to slidego to slideHave questions on basic mathematical concepts?Become a problem-solving champ using logic, not rules. Learn the why behind math with our certified experts.Book a Free Trial ClassFAQs on Right Skewed Histogram A histogram skewed to the right means that the peak of the graph lies to the left side of the center. On the right side of the graph, the frequencies of observations are lower than the frequencies of observations to the left side.What does a Right-Skewed Histogram Look Like?A right-skewed histogram looks like a graph that reaches the maximum point of its slope before the center point of the graph. It looks like a slope that moves up fastly, and then gradually moves down towards the x-axis. The mode of the right-skewed histogram is smaller than its median and mean, and lies to the left of the median.How do you Interpret a Right-Skewed Histogram?The right-skewed histogram interpretation is done by looking at the values of its mean, median, and mode. The peak of the graph or the mode lies to the left of the median or the center of the right-skewed histogram.What is the Center of a Right-Skewed Histogram?If the histogram is skewed to the right, then the median shows the center of the histogram. It lies between the mean and the mode.What Makes a Histogram Skewed Right?In a right-skewed histogram, the frequency of observation lies to the left is the highest and most of the observations or data points lie to the right of the graph which makes it a positively skewed histogram or a right-skewed histogram. This type of histogram reaches its peak point in one of the initial values only and the tail of the graph is usually extended having low frequencies.Written by: Roger BReviewed by: Dan FinlayUpdated on 8 May 2024Frequency polygons are a very simple way of showing frequencies for continuous, grouped dataThey give a quick guide to how frequencies change from one class interval to the nextApart from plotting and joining up points with straight lines there are 3 rules for frequency polygons:Plot points at the midpoint of class intervalsUnless one of the frequencies is 0 do not join the frequency polygon to the x-axisDo not join the first point to the last oneThe result is not actually a polygonIt's more of an 'open' polygon that floats in mid-air!You may be asked to draw a frequency polygon and/or use it to make comments and compare dataThis is most easily shown by an examplee.g. The lengths of 59 songs, in seconds, are recorded in the table belowSong lengtht secondsFrequency120 t < 1504150 t < 18010180 t < 21024210 t < 24018240 t < 2703Frequencies are plotted at the midpoints of the class intervals, so in this case we would plot the points (135, 4), (165, 10), (195, 24), (225, 18) and(255, 3).Join these up with straight lines (but do not join the last to the first!)If you have a histogram with equal class widthsthen a frequency polygon can be drawnby marking the points at the centres of the tops of all the histogram barsand joining those points together to make the 'polygon'Think about what you could you say about the data above, particularly by looking at the diagram onlyThe two things to look for are averages and spreadThe modal class is 180 t < 210Because the graph reaches its highest point thereIt would be acceptable to say that 195 seconds is (an estimate of) the modal song lengthThe diagram (rather than the table) shows (an estimate of) the range of song lengths is 255 135 = 120 secondsi.e. midpoint of highest class interval minus midpoint of lowest class intervalIf 2 frequency polygons are drawn on the same graph comparisons between the 2 sets of data can be madeJot down the midpoints next to the frequencies so you are not trying to work them out in your head while also concentrating on actually plotting the pointsA local council ran a campaign to encourage households to waste less food.To compare the impact of the campaign the council recorded the weight of food waste produced by 30 households in a week both before and after the campaign.The results are shown in the table below.Food wastew kgFrequency(before campaign)Frequency(after campaign)1 w < 1.4351.4 w < 1.8481.8 w < 2.28142.2 w < 2.61032.6 w < 351(a) On the same diagram, draw two frequency polygons, one for before the councils campaign and one for after.Remember to include a key to show which frequency polygon is which.(b) Comment on whether you think the councils campaign has been successful or not and give a reason why.Remember to look for average(s) and/or spreadThe mode (average) is appropriate in this case.The council campaign has been successful as the modal amount of waste has decreased from 2.4 kg of food waste per week to 2 kgA histogram is similar to a bar chart, but there are important differencesBar charts are used for discrete (and possibly non-numerical) data In a bar chart, the height (or length) of a bar determines the frequencyThere are usually gaps between the barsHistograms are used with continuous data, grouped into class intervalsIf the class intervals are of equal width then the height of a bar indicates the frequencyBut there are no gaps between the barsNote that the situation is much more complex for histograms with unequal class widthsBut these are not included on your Foundation courseIf you know the frequencies and class boundaries, then the bars (rectangles) for each class interval can be drawn The class boundaries are marked on the horizontal (x) axisThe height of each bar (the frequency) is measured on the vertical (y) axisAs the data is continuous, the bars will be touching (no gaps)You may be asked to complete an incomplete histogram using information from a grouped frequency tableJust draw in the missing bars based on the information in the tableYou may be asked to complete an incomplete grouped frequency table using information from a histogramUse the information from the histogram bars to complete the tableThe class boundaries can be read off the horizontal axisThe frequencies can be read off the vertical axisYou may be asked to draw a frequency polygon on the histogramMark the points at the centres of the tops of all the histogram barsJoin those points together to make the 'polygon'You may be asked to comment on the skew shown on a histogramSee the 'Skewness' revision noteYou can also use histograms for two data sets to compare the data distributionsbut only if they have the same class intervals and the same frequency scalesRemember that there should not be any gaps between the bars on a histogramLeaving gaps can lose you marksThe table below and the corresponding histogram show the weight, in kg, of some newborn bottlenose dolphins.Weight (w kg)Frequency4 w < 858 w < 122312 w < 161816 w < 20820 w < 2424 w < 28(a) Use the histogram to complete the table.The frequencies for the 20 w < 24 and 24 w < 28 class intervals are missingThese can both be found from the heights of the bars, using the vertical scale on the diagramWeight (w kg)Frequency4 w < 858 w < 122312 w < 161816 w < 20820 w < 24424 w < 282(b) Use the table to complete the histogram.The bars for the 12 w < 16 and 16 w < 20 class intervals are missingWe can get the frequencies for these from the table The bar for the 12 w < 16 class interval should have a height of 18, and go between 12 and 16 on the horizontal axisThe bar for the 16 w < 20 class interval should have a height of 8, and go between 16 and 20 on the horizontal axisMake sure not to leave any gaps between the barsDid this page help you?TutrsStringent selection, robust training, and continuous upskilling.To match your childs unique personality and learning style.Exam prep, Homework help, Advanced learning, and Remedial support.Helping 200,000+ students succeed!Received prestigious President's Education Awards Program from the President of US. Tops her class with an outstanding score of 77.5/80.Received prestigious Pradhan Mantri Rashtriya Bal Puraskar from the Prime Minister of India.Got Level 5 in the STAAR exam at the Renaissance Institute for Competitive Exams.Secured Rank 1 at SOF IMO Level 1 2023, by scoring an outstanding 100/100! Received prestigious President's Education Awards Program from the President of US. Tops her class with an outstanding score of 77.5/80.Received prestigious Pradhan Mantri Rashtriya Bal Puraskar from the Prime Minister of India.Got Level 5 in the STAAR exam at the Renaissance Institute for Competitive Exams.Secured Rank 1 at SOF IMO Level 1 2023, by scoring an outstanding 100/100! Received prestigious President's Education Awards Program from the President of US. Tops her class with an outstanding score of 77.5/80.Received prestigious Pradhan Mantri Rashtriya Bal Puraskar from the Prime Minister of India.Got Level 5 in the STAAR exam at the Renaissance Institute for Competitive Exams.My son started Cuemath in Grade 1 & now he is in Grade 7. All these years, I have been reassured for math subject! I'm sure he will continue with Cuemath till it serves!Cuemath has helped my kids learn math concepts and practice them in an online setting. It is a great online platform with 1:1 learning experience.Our daughter was losing interest in math. After 4-5 classes, I could see her asking for homework. She started liking math again and has now developed a lot of interest.Cuemath keeps introducing new methods, systems, & make it interesting for learners. Unlike the traditional teaching system, it has innovated a different way of teaching.My son has been taking coaching from Cuemath and is showing consistent improvement. It is mainly because of the standard curriculum, mentoring, supervision, & teaching.Have been a great platform with multiple avenues to augment my 8yr olds math skills. Good support from teacher too!My son started Cuemath in Grade 1 & now he is in Grade 7. All these years, I have been reassured for math subject! I'm sure he will continue with Cuemath till it serves!Cuemath has helped my kids learn math concepts and practice them in an online setting. It is a great online platform with 1:1 learning experience.Our daughter was losing interest in math. After 4-5 classes, I could see her asking for homework. She started liking math again and has now developed a lot of interest.Cuemath keeps introducing new methods, systems, & make it interesting for learners. Unlike the traditional teaching system, it has innovated a different way of teaching.My son has been taking coaching from Cuemath and is showing consistent improvement. It is mainly because of the standard curriculum, mentoring, supervision, & teaching.Have been a great platform with multiple avenues to augment my 8yr olds math skills. Good support from teacher too!We had a great experience with Cuemath. He started in 2021 and was quite weak but since joining Cuemath he has been getting better grades.Cuemath's app facilitates teacher-student interaction. The teacher in India understands our Australian math curriculum. We couldn't find such a teacher even locally.Private 1-to-1 tutoring that just works1-3 classes per week, with hassle-free scheduling.Customized learning plan for every child.Get regular insights on your child's progress.What is the frequency and duration of your classes?Typically, the number of classes is two per week for grades K to 8, and three per week for high school. But the schedule is flexible, according to your childs requirements and availability. Also, each class runs for 55 minutes, extendable to an hour.What devices do I need for attending your classes?A desktop or laptop computer that supports video calling is necessary for attending our classes. We also highly recommend a writing tablet for the best learning experience.My child has specific learning requirements. Is your program flexible enough?Absolutely. Our tutors will always customize the classes according to what your child needs - be it homework help, exam or test prep, remedial support for past gaps, or advanced learning.Can your tutors teach the topics covered in my childs school or curriculum?Our tutors are trained to teach according to various curricula across countries. Further, we have a fully customizable curriculum, tailored to your childs needs.Can my child join anytime of the year?Yes. Our tutors always customize the learning plan according to your childs needs, and the time left in the current academic year. If you wish to cover additional topics in the same time, you can always schedule extra classes.What if I dont like the tutor?In the rare case that happens, please raise a ticket with our helpdesk. Well be happy to diagnose the issue, and find you a different tutor that aligns better with your childs needs.What if I do not like your classes after I enroll? Will I get my money back?We have a no questions asked refund policy. If youre unhappy with the experience, you can cancel anytime for a full refund of the unused classes.What happens if my child misses a Cuemath class?We have a flexible leave policy that allows for both planned and unplanned leaves. Just keep your tutor informed.How can I keep track of my child's maths progress?We have a dedicated parent app, that lets you track the progress of your child, and also lets you connect with their tutor.How do I enroll for your classes?Please tap on the Get Started button. Well ask you a few questions about your child to understand their needs better. Once we receive the details, our admissions counselor will call you to match your child with the right tutor, and schedule a free trial class as per your availability. If you like the experience, you can choose a plan and make the payment to begin your classes.Affordable and personalized. Try a class for free.All BlogsStatistics ResourcesFrequency PolygonData VisualizationBasics of Data VisualizationData VisualizationDot PlotStem And Leaf PlotChart TypesFrequency PolygonControl ChartTree DiagramVenn DiagramData VisualizationGET: WSM ALL COURSES ACCESSTable Of ContentsA frequency polygon in statistics is a graph that displays the distribution of a set of continuous data. It aims to visually represent a group of consistent data issuance, making it easier to interpret and analyse by plotting the frequency of each interval on the vertical axis and the midpoint of each interval on the horizontal axis.It thus shows the overall pattern of the data distribution, highlighting any trends or patterns that may exist. It can be used in various fields, including statistics, economics, biology, and social sciences, to analyze and interpret data sets. They are handy for identifying outliers, detecting patterns and trends, and comparing data sets.Frequency polygons are line graphs representing a continuous dataset's frequency distribution. We can get it by dividing the data into intervals or bins. This calculates the frequency for each interval and plots the midpoint of each interval.It can help locate patterns and trends in large datasets, detect outliers, and allow easy comparisons between datasets.They are handy for displaying data sets with many observations or values. They are helpful in various fields, including statistics, economics, biology, and social sciences.The frequency polygon represents the frequency distribution of continuous data graphically. It was first introduced by English statistician Karl Pearson in the late 19th century. He proposed it based on the work of other early statisticians such as Francis Galton and Adolphe Quetelet.Its relevance lies in its ability to visually represent data, allowing for more straightforward interpretation and analysis. As a result, they are a valuable tool in statistics, helping researchers to identify patterns and trends in large data sets.One of the main advantages of it is that they are handy for displaying data sets with many observations or values. By dividing the data into equal intervals and plotting the frequency of each container, the polygon provides a clear and concise representation of the data distribution.These are also helpful in identifying outliers or anomalies in the data set, as these are in the form of points that fall outside the normal range of the distribution. Again, this makes detecting and investigating unusual or unexpected data points easier.Drawing a frequency polygon graph involves several steps:Organize data: Start by organizing your data into a frequency table showing the observations or values in each interval or bin. Determine the interval: Decide the break or bin size used to group the data. This should be within the range of values in the data set and the number of observations. A general rule is to have around 10-20 intervals, each equal. Calculate the midpoint of each interval: Once you have determined the interval size, calculate the midpoint by taking the average of the upper and lower boundaries of the gap. Calculate the frequency for each interval: Using the frequency table, calculate the frequency or number of observations for each interval. Plot the points: Plot the midpoint of each interval on the horizontal axis and the frequency for each interval on the vertical axis. Then, connect the ends with a straight line to form the polygon.Let us understand the topic better with the help of some examples:Suppose Harry is surveying the number of hours people spend using their smartphones daily. He collects data from a sample of 50 people and organizes it into the following frequency table:Interval (hours)Frequency0-151-282-3153-4124-565-626-717-81To visualize this data, he could create a frequency polygon by plotting the midpoint of each interval on the horizontal axis and the frequency for each interval on the vertical axis.A recent example of a frequency polygon in the news was the visualization of the COVID-19 vaccination rates in different countries. In addition, many news organizations and public health agencies have used polygons to show the distribution of vaccination rates across other age groups, regions, and ethnic groups. The US Centers for Disease Control and Prevention used polygons to show the distribution of vaccination rates by age group, with a clear peak in the age group of 65 and older.Some of the advantages of frequency polygon are:Easy to interpret: These provide a clear visual representation of the data distribution, making it easy to interpret and analyze. Highlights patterns and trends:It can help patterns in large datasets that may be difficult to identify by simply looking at the raw data. Displays large datasets: These are particularly useful for displaying large datasets, as they group the data into intervals or bins, making it easier to see the overall pattern. Detects outliers: It can help detect outliers or unusual data points affecting the overall data distribution. Allows for comparison: These compare different datasets, allowing for insights and comparisons between other groups or populations.Some of the disadvantages of frequency polygon are:May oversimplify the data: These group data into intervals or bins, which may oversimplify the data and hide essential details or nuances. Can be misleading: If the intervals or bins are not chosen carefully, it can be deceptive and lead to correct conclusions. May not be suitable for small datasets: These are best for larger datasets, as they may need to provide more detail for smaller datasets. Subjective binning: Choosing the size and number of intervals or bins is personal and can affect the resulting polygon, potentially leading to biased or inaccurate interpretations. Limited to continuous data: These usually represent constant data, making them unsuitable for datasets with discrete data points or categories.Some differences between histograms and frequency polygons:Data representation:Histograms represent data using bars, while the latter represent data using lines. Type of data: Histograms usually display continuous data, while frequency polygons can be used for continuous and discrete data. The shape of distribution:Histograms are best for displaying the form of the data distribution. At the same time, frequency polygons show the trends and patterns in the data. Bin size: In histograms, the width of the bars represents the bin size, while in frequency polygons, the bin size is not represented visually. Range of data: Histograms typically show the full range of the data, while frequency polygons may only show a portion of the data range depending on the chosen bin size and intervals.Points comparing frequency polygons and frequency curves:Data representation: Frequency polygons represent data using straight line segments, while frequency curves represent data using a smooth curve. Type of data: Frequency polygons are typically used for displaying continuous data, while frequency curves can be used for serial and discrete data. Bin size: Frequency polygons require intervals or bins to be defined, while frequency curves do not require intervals or bins. Continuity of data: Frequency curves provide a continuous representation of the data, while frequency polygons may not clearly show the data's continuity. Smoothing: Frequency curves can be smoothed to reduce the impact of outliers, while frequency polygons cannot be filed.Some differences between frequency polygons, line graphs, and ogive graphs are as follows:Data representation: Frequency polygons represent data using straight line segments, and line graphs represent data using a continuous line. In contrast, ogive graphs represent data using consecutive connected lines. Type of data: Frequency polygons and ogive graphs are typically used for displaying continuous data, while line graphs can be used for serial and discrete data. Y-axis: Frequency polygons and ogive graphs typically use the frequency or cumulative frequency on the y-axis, while line graphs may use a range of variables on the y-axis. X-axis: All three graphs use a continuous variable on the x-axis, typically with intervals or bins defined for frequency polygons and ogive graphs. Smoothing: Line graphs can be ground to reduce the impact of noise or outliers in the data, while frequency polygons and ogive graphs cannot be filed.For most of the work you do in this book, you will use a histogram to display the data. One advantage of a histogram is that it can readily display large data sets. A rule of thumb is to use a histogram when the data set consists of 100 values or more.A histogram consists of contiguous (adjoining) boxes. It has both a horizontal axis and a vertical axis. The horizontal axis is labeled with what the data represents (for instance, distance from your home to school). The vertical axis is labeled either frequency or relative frequency (or percent frequency or probability). The graph will have the same shape with either label. The histogram (like the stemplot) can give you the shape of the data, the center, and the spread of the data.The graph for quantitative data looks similar to a bar graph, except there are some major differences. First, in a bar graph the categories can be put in any order on the horizontal axis. There is no set order for these data values. You cant say how the data is distributed based on the shape, since the shape can change just by putting the categories in different orders. With quantitative data, the data are in specific orders, since you are dealing with numbers. With quantitative data, you can talk about a distribution, since the shape only changes a little bit depending on how many categories you set up. This is called a frequency distribution.This leads to the second difference from bar graphs. In a bar graph, the categories that you made in the frequency table were determined by you. In quantitative data, the categories are numerical categories, and the numbers are determined by how many categories (or what are called classes) you choose. If two people have the same number of categories, then they will have the same frequency distribution. Whereas in qualitative data, there can be many different categories depending on the point of view of the author.The third difference is that the categories touch with quantitative data, and there will be no gaps in the graph. The reason that bar graphs have gaps is to show that the categories do not continue on, like they do in quantitative data. Since the graph for quantitative data is different from qualitative data, it is given a new name. The name of the graph is a histogram. To create a histogram, you must first create the frequency distribution. The idea of a frequency distribution is to take the interval that the data spans and divide it up into equal subintervals called classes. Find the range = largest value smallest value Pick the number of classes to use. Usually the number of classes is between five and twenty. Five classes are used if there are a small number of data points and twenty classes if there are a large number of data points (over 1000 data points). (Note: categories will now be called classes from now on.) Class width = \(\frac{\text { range }}{\# \text { classes }}\) Always round up to the next integer (even if the answer is already a whole number go to the next integer). If you dont do this, your last class will not contain your largest data value, and you would have to add another class just for it. If you round up, then your largest data value will fall in the last class, and there are no issues. Create the classes. Each class has limits that determine which values fall in each class. To find the class limits, set the smallest value as the lower class limit for the first class. Then add the class width to the lower class limit to get the next lower class limit. Repeat until you get all the classes. The upper class limit for a class is one less than the lower limit for the next class. In order for the classes to actually touch, then one class needs to start where the previous one ends. This is known as the class boundary. To find the class boundaries, subtract 0.5 from the lower class limit and add 0.5 to the upper class limit. Sometimes it is useful to find the class midpoint. The process is Midpoint \(=\frac{\text { lower limit +upper limit }}{2}\) To figure out the number of data points that fall in each class, go through each data value and see which class boundaries it is between. Utilizing tally marks may be helpful in counting the data values. The frequency for a class is the number of data values that fall in the class.Note The above description is for data values that are whole numbers. If you data value has decimal places, then your class width should be rounded up to the nearest value with the same number of decimal places as the original data. In addition, your class boundaries should have one more decimal place than the original data. As an example, if your data have one decimal place, then the class width would have one decimal place, and the class boundaries are formed by adding and subtracting 0.05 from each class limit.Example \\(\PageIndex{1}\) creating a frequency table Table 2.2.1 contains the amount of rent paid every month for 24 students from a statistics course. Make a relative frequency distribution using 7 classes. 1500 1350 350 1200 850 900 1500 1150 1500 900 1400 1100 1250 600 610 960 890 1325 900 800 2550 495 1200 690 Table 2.2.1: Data of Monthly Rent Solution: Find the range: largest value - smallest value \(= 2550-350=2200\) Pick the number of classes: The directions to say to use 7 classes. Find the class width: width \(=\frac{\text { range }}{7}=\frac{2200}{7} \approx 314.286\) Round up to 315 Always round up to the next integer even if the width is already an integer. Find the class limits: Start at the smallest value. This is the lower class limit for the first class. Add the width to get the lower limit of the next class. Keep adding the width to get all the lower limits. \(350+315=665,665+315=980,980+315=1295 \rightleftharpoons\), The upper limit is one less than the next lower limit: so for the first class the upper class limit would be \(665-1=664\). When you have all 7 classes, make sure the last number, in this case the 2550, is at least as large as the largest value in the data. If not, you made a mistake somewhere. Find the class boundaries: Subtract 0.5 from the lower class limit to get the class boundaries. Add 0.5 to the upper class limit for the last class's boundary. \(350-0.5=349.5, \quad 665-0.5=664.5,\quad 980-0.5=979.5, \quad 1295-0.5=1294.5 \rightleftharpoons\) Every value in the data should fall into exactly one of the classes. No data values should fall right on the boundary of two classes. Find the class midpoints: midpoint \(=\frac{\text { lower limit }+\text { upper limit }}{2}\) \(\frac{350+664}{2}=507, \frac{665+979}{2}=822, \rightleftharpoons\) Tally and find the frequency of the data: Go through the data and put a tally mark in the appropriate class for each piece of data by looking to see which class boundaries the data value is between. Fill in the frequency by changing each of the tallies into a number. Class Limits Class Boundaries Class Midpoint Tally Frequency 350-664 349.5-664.5 507 |||| 4 665-979 664.5-979.5 822 \(\cancel{||||}\) ||| 8 980-1294 979.5-1294.5 1137 \(\cancel{||||}\) 5 1295-1609 1294.5-1609.5 1452 \(\cancel{||||}\) | 6 1610-1924 1609.5-1924.5 1767 0 1925-2239 1924.5-2239.5 2082 0 2240-2554 2239.5-2554.5 2397 | 1 Table 2.2.2: Frequency Distribution for Monthly Rent Make sure the total of the frequencies is the same as the number of data points. It is difficult to determine the basic shape of the distribution by looking at the frequency distribution. It would be easier to look at a graph. The graph of a frequency distribution for quantitative data is called a frequency histogram or just histogram for short.Definition \(\PageIndex{1}\) Histogram: a graph of the frequencies on the vertical axis and the class boundaries on the horizontal axis. Rectangles where the height is the frequency and the width is the class width are drawn for each class.Example \(\PageIndex{2}\) drawing a histogram Draw a histogram for the distribution from Example 2.2.1. Solution: The class boundaries are plotted on the horizontal axis and the frequencies are plotted on the vertical axis. You can plot the midpoints of the classes instead of the class boundaries. Graph 2.2.1 was created using the midpoints because it was easier to do with the software that created the graph. Graph 2.2.1: Histogram for Monthly Rent Notice the graph has the axes labeled, the tick marks are labeled on each axis, and there is a title. It is important that your graphs (all graphs) are clearly labeled. Reviewing the graph you can see that most of the students pay around $750 per month for rent, with about $1500 being the other common value. You can see from the graph, that most students pay between $600 and $1600 per month for rent. Of course, these values are just estimates from the graph. There is a large gap between the $1500 class and the highest data value. This seems to say that one student is paying a great deal more than everyone else. This value could be considered an outlier. An outlier is a data value that is far from the rest of the values. It may be an unusual value or a mistake. It is a data value that should be investigated. In this case, the student lives in a very expensive part of town, thus the value is not a mistake, and is just very unusual. There are other aspects that can be discussed, but first some other concepts need to be introduced. Frequencies are helpful, but understanding the relative size each class is to the total is also useful. To find this you can divide the frequency by the total to create a relative frequency. If you have the relative frequencies for all of the classes, then you have a relative frequency distribution. Definition \(\PageIndex{2}\) Relative Frequency Distribution A variation on a frequency distribution is a relative frequency distribution. Instead of giving the frequencies for each class, the relative frequencies are calculated. Relative frequency \(=\frac{\text { frequency }}{\# \text { of data points }}\) This gives you percentages of data that fall in each class. Example \(\PageIndex{3}\) creating a relative frequency table Find the relative frequency for the grade data. Solution: From Example 2.2.1, the frequency distribution is reproduced in Table 2.2.2. Class Limits Class Boundaries Class Midpoint Frequency 350-664 349.5-664.5 507 4 665-979 664.5-979.5 822 8 980-1294 979.5-1294.5 1127 5 1295-1609 1294.5-1609.5 1452 6 1610-1924 1609.5-1924.5 1767 0 1925-2239 1924.5-2239.5 2082 0 2240-2554 2239.5-2554.5 2397 1 Table 2.2.2: Frequency Distribution for Monthly Rent Divide each frequency by the number of data points. \(\frac{4}{24}=0.17, \frac{8}{24}=0.33, \frac{5}{24}=0.21, \rightleftharpoons\) Class Limits Class Boundaries Class Midpoint Frequency Relative Frequency 350-664 349.5-664.5 507 4 0.17 665-979 664.5-979.5 822 8 0.33 980-1294 979.5-1294.5 1127 5 0.21 1295-1609 1294.5-1609.5 1452 6 0.25 1610-1924 1609.5-1924.5 1767 0 0 1925-2239 1924.5-2239.5 2082 0 0 2240-2554 2239.5-2554.5 2397 1 0.04 Total 24 1 Table 2.2.3: Relative Frequency Distribution for Monthly Rent The relative frequencies should add up to 1 or 100%. (This might be off a little due to rounding errors.) The graph of the relative frequency is known as a relative frequency histogram. It looks identical to the frequency histogram, but the vertical axis is relative frequency instead of just frequencies. Example \(\PageIndex{4}\) drawing a relative frequency histogram Draw a relative frequency histogram for the grade distribution from Example 2.2.1. Solution: The class boundaries are plotted on the horizontal axis and the relative frequencies are plotted on the vertical axis. (This is not easy to do in R, so use another technology to graph a relative frequency histogram.) Graph 2.2.2: Relative Frequency Histogram for Monthly Rent Notice the shape is the same as the frequency distribution. Another useful piece of information is how many data points fall below a particular class boundary. As an example, a teacher may want to know how many students received below an 80%, a doctor may want to know how many adults have cholesterol below 160, or a manager may want to know how many stores gross less than $2000 per day. This is known as a cumulative frequency. If you want to know what percent of the data falls below a certain class boundary, then this would be a cumulative relative frequency. For cumulative frequencies you are finding how many data values fall below the upper class limit. To create a cumulative frequency distribution, count the number of data points that are below the upper class boundary, starting with the first class and working up to the top class. The last upper class boundary should have all of the data points below it. Also include the number of data points below the lowest class boundary, which is zero. Example \(\PageIndex{5}\) creating a cumulative frequency distribution Create a cumulative frequency distribution for the data in Example 2.2.1. Solution: The frequency distribution for the data is in Table 2.2.2. Class Limits Class Boundaries Class Midpoint Frequency 350-664 349.5-664.5 507 4 665-979 664.5-979.5 822 8 980-1294 979.5-1294.5 1127 5 1295-1609 1294.5-1609.5 1452 6 1610-1924 1609.5-1924.5 1767 0 1925-2239 1924.5-2239.5 2082 0 2240-2554 2239.5-2554.5 2397 1 Table 2.2.2: Frequency Distribution for Monthly Rent Now ask yourself how many data points fall below each class boundary. Below 349.5, there are 0 data points. Below 664.5 there are 4 data points, below 979.5, there are 4 + 8 = 12 data points, below 1294.5 there are 4 + 8 + 5 = 17 data points, and continue this process until you reach the upper class boundary. This is summarized in Table 2.2.4. Class Limits Class Boundaries Class Midpoint Frequency Cumulative Frequency 350-664 349.5-664.5 507 4 4 665-979 664.5-979.5 822 8 12 980-1294 979.5-1294.5 1127 5 17 1295-1609 1294.5-1609.5 1452 6 23 1610-1924 1609.5-1924.5 1767 0 23 1925-2239 1924.5-2239.5 2082 0 23 2240-2554 2239.5-2554.5 2397 1 24 Table 2.2.4: Cumulative Distribution for Monthly Rent Again, it is hard to look at the data the way it is. A graph would be useful. The graph for cumulative frequency is called an ogive (o-jive). To create an ogive, first create a scale on both the horizontal and vertical axes that will fit the data. Then plot the points of the class upper class boundary versus the cumulative frequency. Make sure you include the point with the lowest class boundary and the 0 cumulative frequency. Then just connect the dots. Example \(\PageIndex{6}\) drawing an ogive Draw an ogive for the data in Example 2.2.1. Solution: Using the upper class boundary and its corresponding cumulative frequency, plot the points as ordered pairs on the axes. Then connect the dots. You should have a line graph that rises as you move from left to right. Graph 2.2.3: Ogive for Monthly Rent The usefulness of a ogive is to allow the reader to find out how many students pay less than a certain value, and also what amount of monthly rent is paid by a certain number of students. As an example, suppose you want to know how many students pay less than $1500 a month in rent, then you can go up from the $1500 until you hit the graph and then you go over to the cumulative frequency axes to see what value corresponds to this value. It appears that around 20 students pay less than $1500. (See Graph 2.2.4.) Graph 2.2.4: Ogive for Monthly Rent with Example Also, if you want to know the amount that 15 students pay less than, then you start at 15 on the vertical axis and then go over to the graph and down to the horizontal axis where the line intersects the graph. You can see that 15 students pay less than about $1200 a month. (See Graph 2.2.5.) Graph 2.2.5: Ogive for Monthly Rent with Example If you graph the cumulative relative frequency then you can find out what percentage is below a certain number instead of just the number of people below a certain value. Shapes of the distribution: When you look at a distribution, look at the basic shape. There are some basic shapes that are seen in histograms. Realize though that some distributions have no shape. The common shapes are symmetric, skewed, and uniform. Another interest is how many peaks a graph may have. This is known as modal. Symmetric means that you can fold the graph in half down the middle and the two sides will line up. You can think of the two sides as being mirror images of each other. Skewed means one tail of the graph is longer than the other. The graph is skewed in the direction of the longer tail (backwards from what you would expect). A uniform graph has all the bars the same height. Modal refers to the number of peaks. Unimodal has one peak and bimodal has two peaks. Usually if a graph has more than two peaks, the modal information is not longer of interest. Other important features to consider are gaps between bars, a repetitive pattern, how spread out is the data, and where the center of the graph is. This graph is roughly symmetric and unimodal: Graph 2.2.6: Symmetric, Unimodal Graph This graph is symmetric and bimodal: Graph 2.2.7: Symmetric, Bimodal Graph This graph is skewed to the right: Graph 2.2.8: Skewed Right Graph This graph is skewed to the left and has a gap: Graph 2.2.9: Skewed Left Graph This graph is uniform since all the bars are the same height: Graph 2.2.10: Uniform Graph Example \(\PageIndex{7}\) creating a frequency distribution, histogram, and ogive The following data represents the percent change in tuition levels at public, fouryear colleges (inflation adjusted) from 2008 to 2013 (Weissmann, 2013). Create a frequency distribution, histogram, and ogive for the data. 19.5% 40.8% 57.0% 15.1% 17.4% 5.2% 13.0% 15.6% 51.5% 15.6% 14.5% 22.4% 19.5% 31.3% 21.7% 27.0% 13.1% 26.8% 24.3% 38.0% 21.1% 9.3% 46.7% 14.5% 78.4% 67.3% 21.1% 22.4% 5.3% 17.3% 17.5% 36.6% 72.0% 63.2% 15.1% 2.2% 17.5% 36.7% 2.8% 16.2% 20.5% 17.8% 30.1% 63.6% 17.8% 23.2% 25.3% 21.4% 28.5% 9.4% Table 2.2.5: Data of Tuition Levels at Public, Four-Year Colleges Solution: Find the range: largest value - smallest value = \(78.4\)% \(-2.2\)% \(=76.2\)% Pick the number of classes: Since there are 50 data points, then around 6 to 8 classes should be used. Let's use 8. Find the class width: width \(=\frac{\text { range }}{8}=\frac{76.2 \%}{8} \approx 9.525 \%\) Since the data has one decimal place, then the class width should round to one decimal place. Make sure you round up. width \(=9.6\)% Find the class limits: \(2.2 \%+9.6 \%=11.8 \%, 11.8 \%+9.6 \%=21.4 \%, 21.4 \%+9.6 \%=31.0 \%, \leftrightharpoons\) Find the class boundaries: Since the data has one decimal place, the class boundaries should have two decimal places, so subtract 0.05 from the lower class limit to get the class boundaries. Add 0.05 to the upper class limit for the last classs boundary. \(2.2-0.05=2.15 \%, 11.8-0.05=11.75 \%, 21.4-0.05=21.35 \% \leftrightharpoons\) Every value in the data should fall into exactly one of the classes. No data values should fall right on the boundary of two classes. Find the class midpoints: midpoint \(=\frac{\text { lower limt }+\text { upper limit }}{2}\) \(\frac{2.2+11.7}{2}=6.95 \%, \frac{11.8+21.3}{2}=16.55 \%, \leftrightharpoons\) Tally and find the frequency of the data: Class Limits Class Boundaries Class Midpoint Tally Frequency Relative Frequency Cumulative Frequency 2.2-11.7 2.15-11.75 6.95 \(\cancel{||||} |\) 6 0.12 6 11.8-21.3 11.75-21.35 16.55 \(\cancel{||||} \cancel{||||} \cancel{||||} \cancel{||||}\) 20 0.40 26 21.4-30.9 21.35-30.95 26.15 \(\cancel{||||} \cancel{||||} |\) 11 0.22 37 31.0-45.0 30.95-40.55 35.75 \(|||| \) 4 0.08 41 40.6-50.1 40.55-50.15 45.35 \(|| \) 2 0.04 43 50.2-59.7 50.15-59.75 54.95 \(|| \) 2 0.04 45 59.8-69.3 59.75-69.35 64.55 \(||| \) 3 0.06 48 69.4-78.9 69.35-78.95 74.15 \(|| \) 2 0.04 50 Table 2.2.6: Frequency Distribution for Tuition Levels at Public, Four-Year Colleges Make sure the total of the frequencies is the same as the number of data points. Graph 2.2.11: Histogram for Tuition Levels at Public, Four-Year Colleges This graph is skewed right, with no gaps. This says that most percent increases in tuition were around 16.55%, with very few states having a percent increase greater than 45.35%. Graph 2.2.12: Ogive for Tuition Levels at Public, Four-Year Colleges Looking at the ogive, you can see that 30 states had a percent change in tuition levels of about 25% or less. There are occasions where the class limits in the frequency distribution are predetermined. Example 2.2.8 demonstrates this situation. Example \(\PageIndex{8}\) creating a frequency distribution and histogram The following are the percentage grades of 25 students from a statistics course. Make a frequency distribution and histogram. 62 87 81 69 87 62 45 95 76 76 62 71 65 67 72 80 40 77 87 58 84 73 93 64 89 Table 2.2.7: Data of Test Grades Solution: Since this data is percent grades, it makes more sense to make the classes in multiples of 10, since grades are usually 90 to 100%, 80 to 90%, and so forth. It is easier to not use the class boundaries, but instead use the class limits and think of the upper class limit being up to but not including the next classes lower limit. As an example the class 80 90 means a grade of 80% up to but not including a 90%. A student with an 89.9% would be in the 80-90 class. Class Limit Class Midpoint Tally Freqeuncy 40-50 45 \(|| \) 2 50-60 55 \(| \) 1 60-70 65 \(\cancel{||||} || \) 7 70-80 75 \(\cancel{||||} | \) 6 80-90 85 \(\cancel{||||} || \) 7 90-100 95 \(|| \) 2 Table 2.2.8: Frequency Distribution for Test Grades Graph 2.2.13: Histogram for Test Grades It appears that most of the students had between 60 to 90%. This graph looks somewhat symmetric and also bimodal. The same number of students earned between 60 to 70% and 80 to 90%. There are other types of graphs for quantitative data. They will be explored in the next section. In this histogram, height is divided into bins of equal width (5 feet)Graphical representation of the distribution of numerical dataFor the histogram used in digital image processing, see Image histogram and Color histogram.A histogram is a visual representation of the distribution of quantitative data. To construct a histogram, the first step is to "bin" (or "bucket") the range of values divide the entire range of values into a series of intervalsand then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) are adjacent and are typically (but not required to be) of equal size.[1]Histograms give a rough sense of the density of the underlying distribution of the data, and often for density estimation: estimating the probability density function of the underlying variable. The total area of a histogram used for probability density is always normalized to 1. If the length of the intervals on the x-axis are all 1, then a histogram is identical to a relative frequency plot.Histograms are sometimes confused with bar charts. In a histogram, each bin is for a different range of values, so altogether the histogram illustrates the distribution of values. But in a bar chart, each bar is for a different category of observations (e.g., each bar might be for a different population), so altogether the bar chart can be used to compare different categories. Some authors recommend that bar charts always have gaps between the bars to clarify that they are not histograms.[2][3]The term "histogram" was first introduced by Karl Pearson, the founder of mathematical statistics, in lectures delivered in 1892 at University College London. Pearson's term is sometimes incorrectly said to combine the Greek root (gramma) = "figure" or "drawing" with the root (historia) = "inquiry" or "history". Alternatively the root (histion) is also proposed, meaning "web" or "tissue" (as in histology, the study of biological tissue). Both of these etymologies are incorrect, and in fact Pearson, who knew Ancient Greek well, derived the term from a different if homophonous Greek root, = "something set upright", "mast", referring to the vertical bars in the graph. Pearson's new term was embedded in a series of other analogous neologisms, such as "stigmogram" and "radiogram".[4] Pearson himself noted in 1895 that although the term "histogram" was new, the type of graph it designates was "a common form of graphical representation".[5]In fact the technique of using a bar graph to represent statistical measurements was devised by the Scottish economist, William Playfair, in his Commercial and political atlas (1786).[4]This is the data for the histogram to the right, using 500 items:Bin/IntervalCount/Frequency3.5 to 2.5192.5 to 1.51321.5 to 0.511090.5 to 0.491800.5 to 1.491321.5 to 2.49342.5 to 3.494The words used to describe the patterns in a histogram are: "symmetric", "skewed left" or "right", "unimodal", "bimodal" or "multimodal".Symmetric, unimodalSkewed rightSkewed leftBimodalMultimodalSymmetricIt is a good idea to plot the data using several different bin widths to learn more about it. Here is an example on tips given in a restaurant.Tips using a $1 bin width, skewed right, unimodalTips using a 10c bin width, still skewed right, multimodal with modes at $ and 50c amounts, indicates rounding, also some outliersThe U.S. Census Bureau found that there were 124 million people who work outside of their homes.[6] Using their data on the time occupied by travel to work, the table below shows the absolute number of people who responded with travel times "at least 30 but less than 35 minutes" is higher than the numbers for the categories above and below it. This is likely due to people rounding their reported journey time.[citation needed] The problem of reporting values as somewhat arbitrarily rounded numbers is a common phenomenon when collecting data from people.[citation needed]Histogram of travel time (to work), US 2000 census. Area under the curve equals the total number of cases. This diagram uses Q/width from the table.Data by absolute numbersIntervalWidthQuantityQuantity/width05418083655136872737105186183723155196343926205179813596255719014383051636932733553212642405412282445159200613603064612159060343557This histogram shows the number of cases per unit interval as the height of each block, so that the area of each block is equal to the number of people in the survey who fall into its category. The area under the curve represents the total number of cases (124 million). This type of histogram shows absolute numbers, with Q in thousands.Histogram of travel time (to work), US 2000 census. Area under the curve equals 1. This diagram uses Q/total/width (crowding) from the table. The height of a block represents crowding which is defined as - percentage per horizontal unit.Data by proportionIntervalWidthQuantity (Q)Q/total/width0541800.006755136870.0221105186180.0300155196340.0316205179810.029025571900.0116305163690.026435532120.005240541220.0066451592000.0049603064610.0017906034350.0005This histogram differs from the first only in the vertical scale. The area of each block is the fraction of the total that each category represents, and the total area of all the bars is equal to 1 (the fraction meaning "all"). The curve displayed is a simple density estimate. This version shows proportions, and is also known as a unit area histogram.In other words, a histogram represents a frequency distribution by means of rectangles whose widths represent class intervals and whose areas are proportional to the corresponding frequencies: the height of each is the average frequency density for the interval. The intervals are placed together in order to show that the data represented by the histogram, while exclusive, is also contiguous. (E.g., in a histogram it is possible to have two connecting intervals of 10.520.5 and 20.533.5, but not two connecting intervals of 10.520.5 and 22.532.5. Empty intervals are represented as empty and not skipped.)[7]An ordinary and a cumulative histogram of the same data. The data shown is a random sample of 10,000 points from a normal distribution with a mean of 0 and a standard deviation of 1.The data used to construct a histogram are generated via a function mi that counts the number of observations that fall into each of the disjoint categories (known as bins). Thus, if we let n be the total number of observations and k be the total number of bins, the histogram data mi meet the following conditions: n = i = 1 k m i . {\displaystyle n=\sum _{i=1}^{k}{m_{i}}.} A histogram can be thought of as a simplistic kernel density estimation, which uses a kernel to smooth frequencies over the bins. This yields a smoother probability density function, which will in general more accurately reflect distribution of the underlying variable. The density estimate could be plotted as an alternative to the histogram, and is usually drawn as a curve rather than a set of boxes. Histograms are nevertheless preferred in applications, when their statistical properties need to be modeled. The correlated variation of a kernel density estimate is very difficult to describe mathematically, while it is simple for a histogram where each bin varies independently.An alternative to kernel density estimation is the average shifted histogram,[8]which is fast to compute and gives a smooth curve estimate of the density without using kernels.A cumulative histogram: a mapping that counts the cumulative number of observations in all of the bins up to the specified bin. That is, the cumulative histogram Mi of a histogram mj can be defined as: M i = j = 1 i m j . {\displaystyle M_{i}=\sum _{j=1}^{i}{m_{j}}.} There is no "best" number of bins, and different bin sizes can reveal different features of the data. Grouping data is at least as old as Graunt's work in the 17th century, but no systematic guidelines were given[9] until Sturges's work in 1926.[10]Using wider bins where the density of the underlying data points is low reduces noise due to sampling randomness; using narrower bins where the density is high (so the signal drowns the noise) gives greater precision to the density estimation. Thus varying the bin-width within a histogram can be beneficial. Nonetheless, equal-width bins are widely used.Some theoreticians have attempted to determine an optimal number of bins, but these methods generally make strong assumptions about the shape of the distribution. Depending on the actual data distribution and the goals of the analysis, different bin widths may be appropriate, so experimentation is usually needed to determine an appropriate width. There are, however, various useful guidelines and rules of thumb.[11]The number of bins k can be assigned directly or can be calculated from a suggested bin widthh as:Histogram data represented with different bin widths k = max x min x h . {\displaystyle k=\left\lceil {\frac {\max x-\min x}{h}}\right\rceil .} The braces indicate the ceiling function. k = n {\displaystyle k=\lceil {\sqrt {n}}\rceil \,} which takes the square root of the number of data points in the sample and rounds to the next integer. This rule is suggested by a number of elementary statistics textbooks [12] and widely implemented in many software packages.[13]Sturges's rule[10] is derived from a binomial distribution and implicitly assumes an approximately normal distribution. k = log 2 n + 1 , {\displaystyle k=\lceil \log _{2}n\rceil +1,\,} Sturges's formula implicitly bases bin sizes on the range of the data, and can perform poorly if n