Welcome to the Evaluation of AccuStripes

Techniqual Requirements & Instructions

If you want participate in this study, you will have to use the following tools/devices:

The monitor/device must have a resolution of at least 1920x1080 or higher.

Please execute this study on a commerically available monitor, preferably with a resolution of 1920x1080 pixel or higher.

The system will not allow you to do this study more than once on your machine.

Before you can start with the study these requirements are tested automatically.

Use a tested browser: Firefox, Chrome, Edge, Opera, or Safari.

Using a browser other than those listed may cause technical problems and you may not be able to complete the study. For example, the study is not executable with Internet Explorer!

If you want to participate in this study, you will have to follow these instructions:

You must have your browser window in full screen.

You are not allowed to reduce the size of the browser window during the study.

You must not click the back-button, forward-button, or the refresh button of the browser. You are only allowed to click on the buttons mentioned in the next instructions.

All these instructions are regularly tested during the study.

Violating any of the instructions will result in immediate termination of the study and no payment will be made.

All collected data (your responses to this questionnaire) will be fully anonymous.

You can cancel the study at any time. In that case, we will not use your data and no payment will be made.

Additional Notes:

Don't be frustrated if you don't recognize the answer - this is no test of your competence.

In this study, different forms of representations are investigated. For this reason, it is not always obvious which answer is the correct one. If you are unsure, this is just a sign that the representation format is not ideal.

This study requires you to assess colors.

A color-blind friendly color map is used, but please use a device/monitor, which is capable of displaying colors.

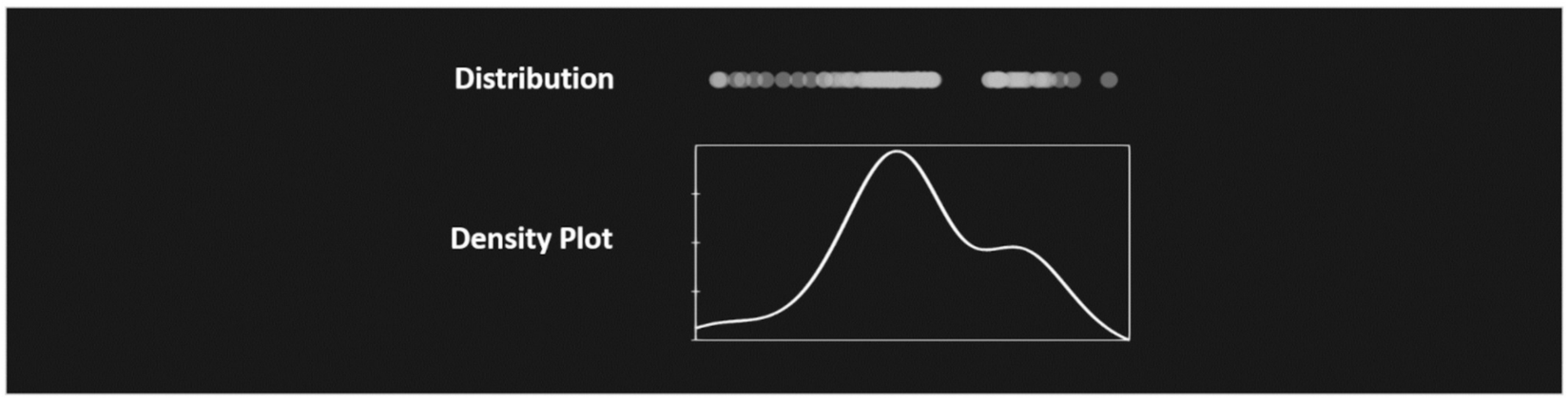
If your device fullfils the requirements and you agree with the stated instructions, please provide your Worker ID and click on the button "Start". Then you will be given the next instructions.

Your Worker ID:

Start

Evaluation - AccuStripes

This study is concerned with investigating different visual encodings for the representation of distributions. We encounter distributions all the time in everyday life, in the news, in weather forecasts, etc. Generally, a distribution is a collection of data points, each of which has a certain value. Distributions are often visualized as density plots in which the shape of the distribution is shown in detail.



Situations may arise where an observer wants to identify and compare multiple distributions simultaneously. To support these two tasks equally, we investigate different visual encodings for distributions in this study. We refer to the collection of visualization compositions as AccuStripes.

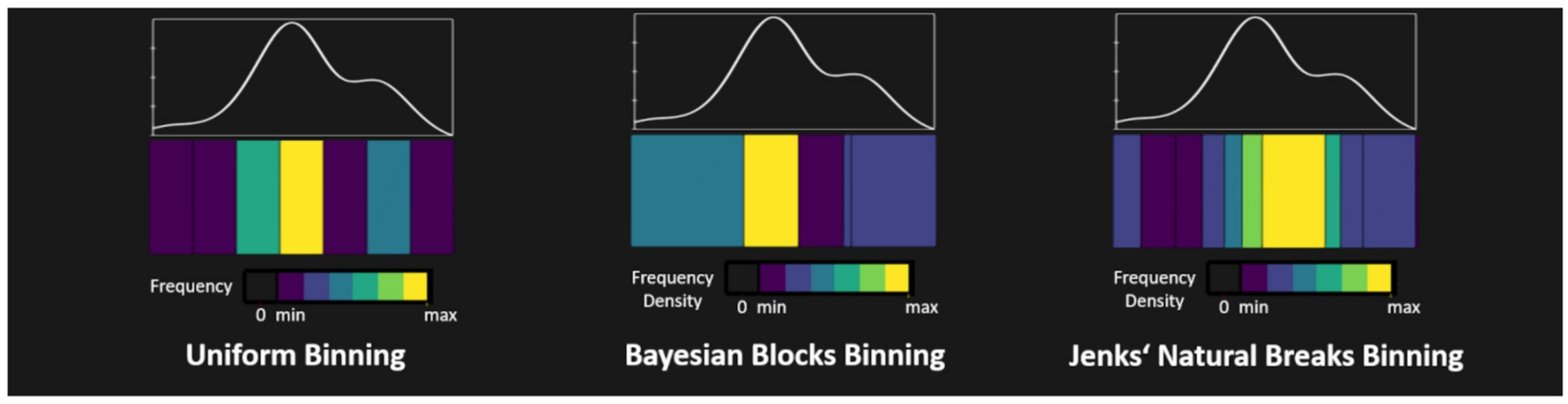
AccuStripes

AccuStripes display a distribution in the form of colored rectangles computed through a binning technique. AccuStripes are determined by three different binning techniques: Uniform Binning, Bayesian Blocks Binning, and Jenks' Natural Breaks Binning.

Binning Techniques:

The binning techniques provide an abstraction of the distribution by **dividing it into individual rectangles, i.e., bins**. The bins are colored according to the number of points that lie inside the bin. **The more points are located in the bin, the brighter** or more yellowish is its color, **the fewer points are located in it, the darker** or more violet is its color. Therefore, AccuStripes can adopt three different representation for a single distribution, depending on the binning technique applied.

In the following image you can inspect the three different AccuStripes representations for the same distribution.



Uniform Binning:

The distribution is divided into bins of **equal width**. The bins are colored according to the number of points inside, called frequency.

Uniform Binning follows the strategy to **divide the distribution into a fixed number of bins**. We use the Sturge's Rule to compute the number of bins.

Adaptive Binning:

The distribution is divided into bins of **varying width**. The bins are colored according to the number of points inside divided by the width of the bins, called frequency density.
Bayesian Blocks and Jenks' Natural Breaks Binning are both adaptive binning techniques.

Bayesian Blocks Binning: This method groups sections of the curve into bins that have minimal or no changes in height.

Jenks' Natural Breaks Binning: This method groups the data into bins by recognizing "natural" breaks (i.e., divisions) in the distribution.

Evaluation Procedure

If you decide to participate in this research, you will be asked to answer questions about fictional data.

In total you will answer 15 main questions and for each of those one supplemental question.
Each question has to be answered in order to move on to the next one. You cannot change your answer for a particular question after you have pressed the button "Continue".

We expect this study to take about 10-15 minutes.

This study consists of 3 parts: the Identification Challenge, the Comparison Challenge, and the Flaw Detection Challenge. For each challenge you will first get a short introduction with a description what you will have to do and an example question. Then you will have to answer some questions. The time it takes you to answer the main question is measured. Please, do not feel stressed! Be as accurate as possible. After answering the main question, there will be time to rest.

Anytime inbetween the study we included some trivial questions or attention tasks to filter out people who just randomly click through the study. Be aware of them. **If you answer them incorrectly, your answers will be removed from the study and no payment will be made.**

Supplementary Question

After each main question, you will have to state how confident you are with your answer. Hereby you can choose from "very unconfident" to "very confident". For example if you do not know the correct answer for the main question, tick your best guess and check "very unconfident" at this supplementary question.

How confident are you about your answer? (example)

- ○ Very Unconfident
- ○ Unconfident
- ○ Confident
- ○ Very Confident

Demographic Information

In the following we kindly ask you to give us a little information about your person. All data will be treated absolutely anonymously. An assignment of your answers to your person is not possible and also not intended. The evaluation will only be carried out according to scientific criteria.

What is your gender?

- ○ Male
- ○ Female
- ○ Diverse
- ○ Decline to state

How old are you? (years)

- ○ < 20
- ○ 20 - 29
- ○ 30 - 39
- ○ 40 - 49
- ○ > 50

What is your highest level of education?

- ○ Some high school
- ○ High school degree
- ○ Some college
- ○ College degree
- ○ Graduate degree

Do you have experience with charts and graphs?

- ○ most unfamiliar (never worked with them before)
- ○ unfamiliar (once or twice)
- ○ neutral (have some experience)
- ○ familiar (working regularly with them)
- ○ most familiar (working with them daily)

When you are creating charts, do you experiment with its parameters? (i.e. changing the number of bins drawn in an histogram, changing the curvature in a density chart, etc.)

- ○ most unfamiliar (never changed them before)
- ○ unfamiliar (once or twice)
- ○ neutral (have some experience)
- ○ familiar (working regularly with them)
- ○ most familiar (working with them daily)

Continue >>

0/16

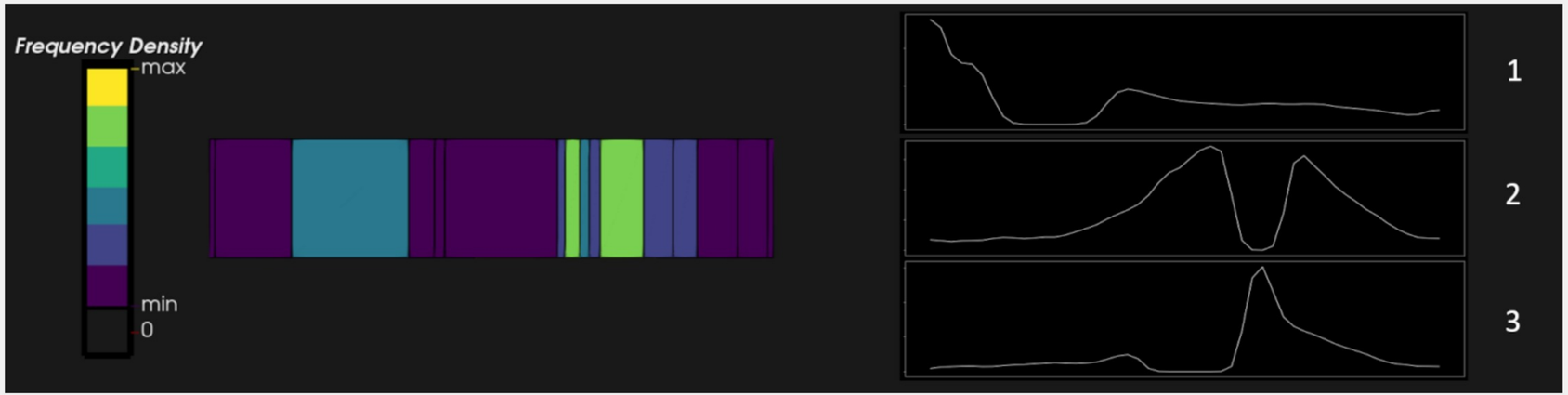
End Study

Identification Challenge - Explanation

In this challenge your task is to select the density plot (1,2,3), that is shown in the AccuStripe.

In the following you will see a sample question. You can choose from three answer possibilities. One answer is correct.

Select the density plot that displays the distribution in the AccuStripe: (example)



- 1
- 2
- 3

Continue >>

End Study

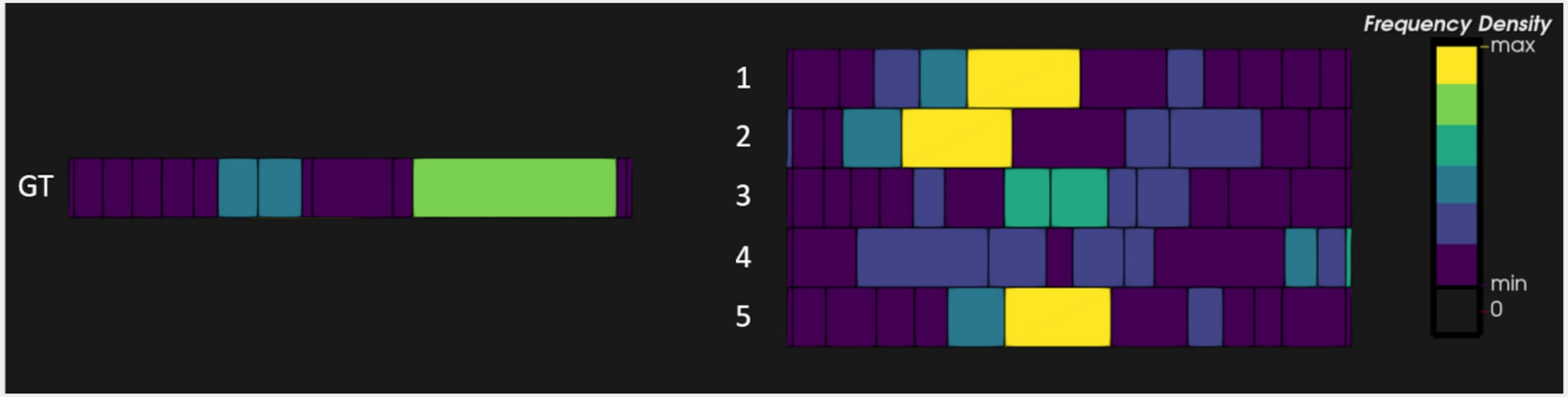
Comparison Challenge - Explanation

In this second part of the study your task will be to compare various distributions (1 - 5) with one ground truth distribution. Select one distribution (1 - 5), which you think is most similar to the ground truth distribution (GT).

The most similar distribution to ground truth is the distribution that can be transformed into ground truth by the least amount of effort.

In the following you will see a sample question. You can choose from five answer possibilities. One answer is correct.

Choose the distribution, which is the most similar one to the groundTruth (GT). (example)



- 1
- 2
- 3
- 4
- 5

Continue >>

End Study

Flaw Detection Challenge

In this part you will look for flaws in data charts. You will be shown a table of charts, each containing sample data. Most of the charts are random data, but one of them will have a flaw. Your task is to select the chart that has the flaw. In some cases, this flaw is hard to spot. In these cases, make your best guess. There can be three different types of flaws: Gaps, Spikes, and Outlier.

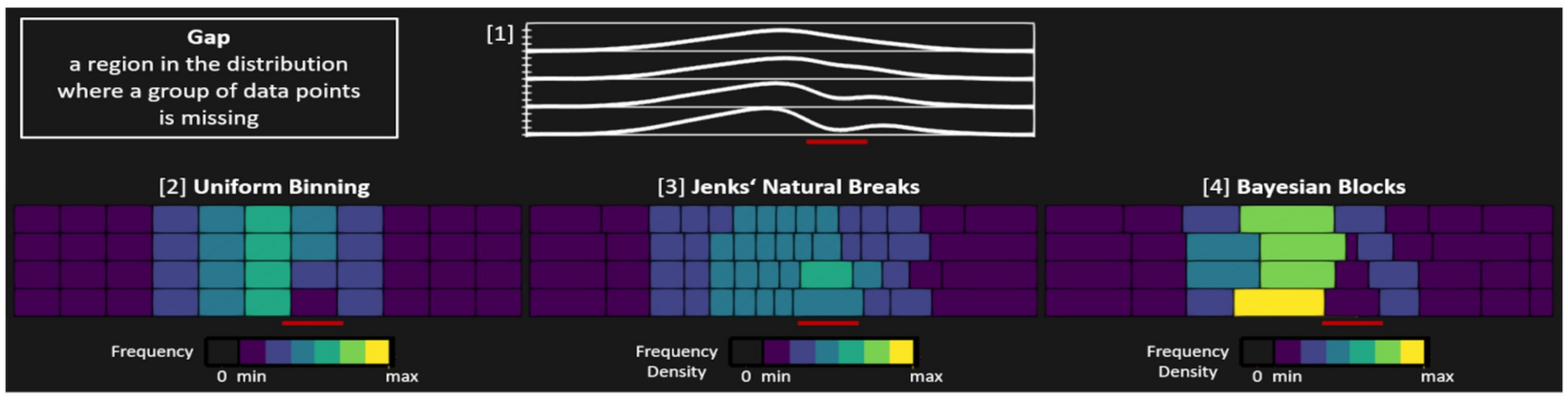
Data without flaws: The distributions were generated by randomly sampling. Even data without flaws can look different from other data, just due to random chance.

Gaps: For data with gaps, all of the data points in a particular region will be removed.

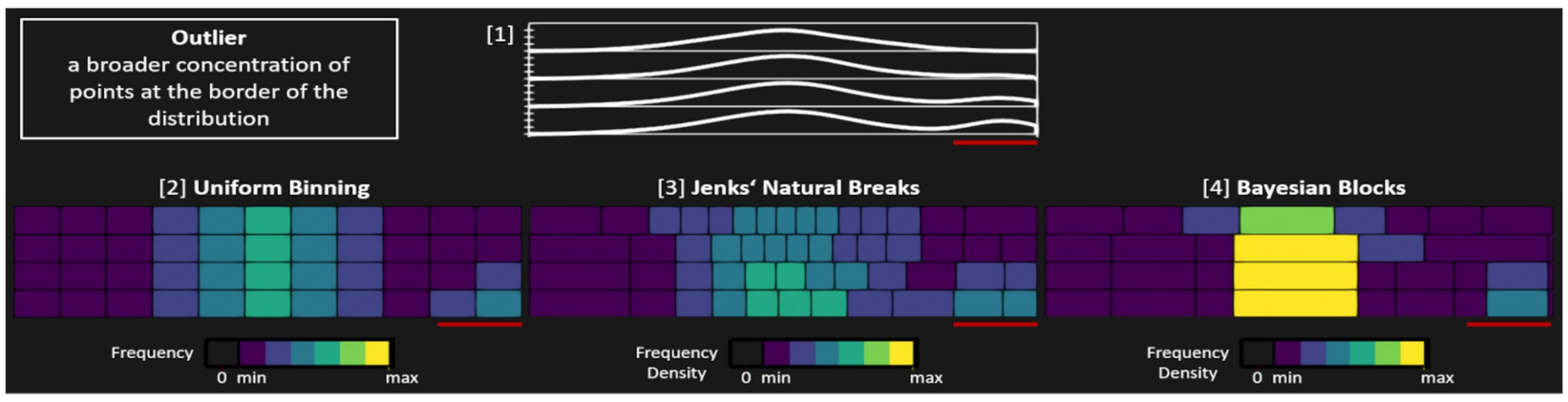
Spikes: For data with spikes, there will be a lot of data points with the exact same value.

Outliers: For data with outliers, there will be a large number of data points that are much lower, or much higher, than most of the rest of the data points.

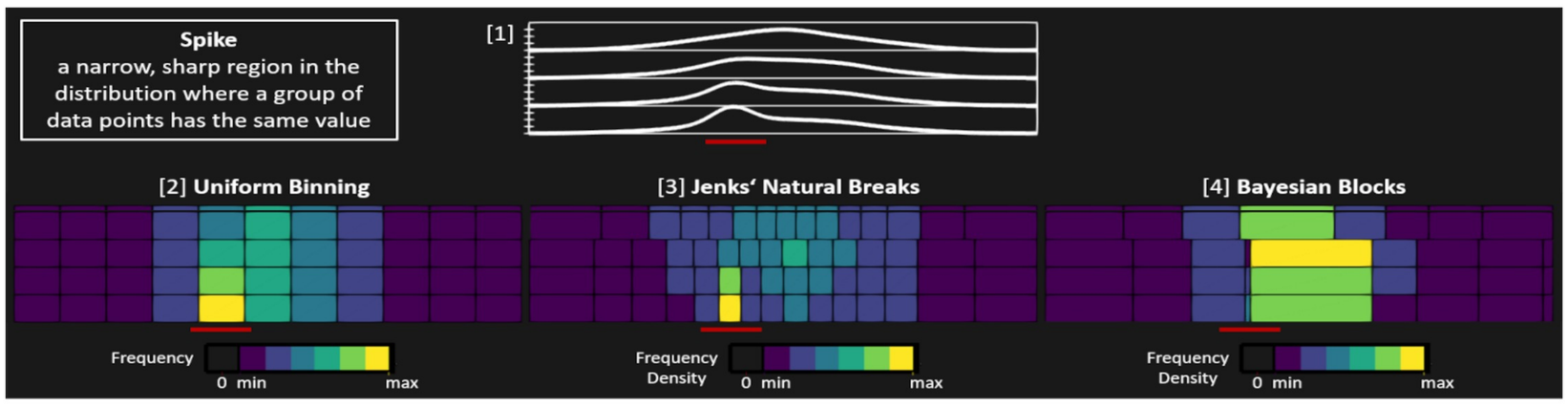
In the following image you will see how **gaps** are represented. There are 4 distributions, the first distribution does not contain a gap, the following 3 distributions do have a gap of increasing severity at the same region. In [1] the 4 distributions are shown in a density plot. In [2] the distributions are binned by Uniform Binning, in [3] they are binned by Jenks' Natural Breaks and in [4] by Bayesian Blocks. The red lines indicate where the gaps are located in the distributions and how the different binning techniques represent the gaps.



In the next image you will you will see how **outliers** are represented. Again, there are 4 distributions, the first one does not contain a data flaw, the following 3 distributions do have an outlier at the right side of the distribution of increasing severity. Again, [1] shows the 4 distributions in a density plot, [2] binned through Uniform Binning, [3] in Jenks' Natural Breaks and [4] in Bayesian Blocks. The red lines indicate where the outlier is located in the distributions.



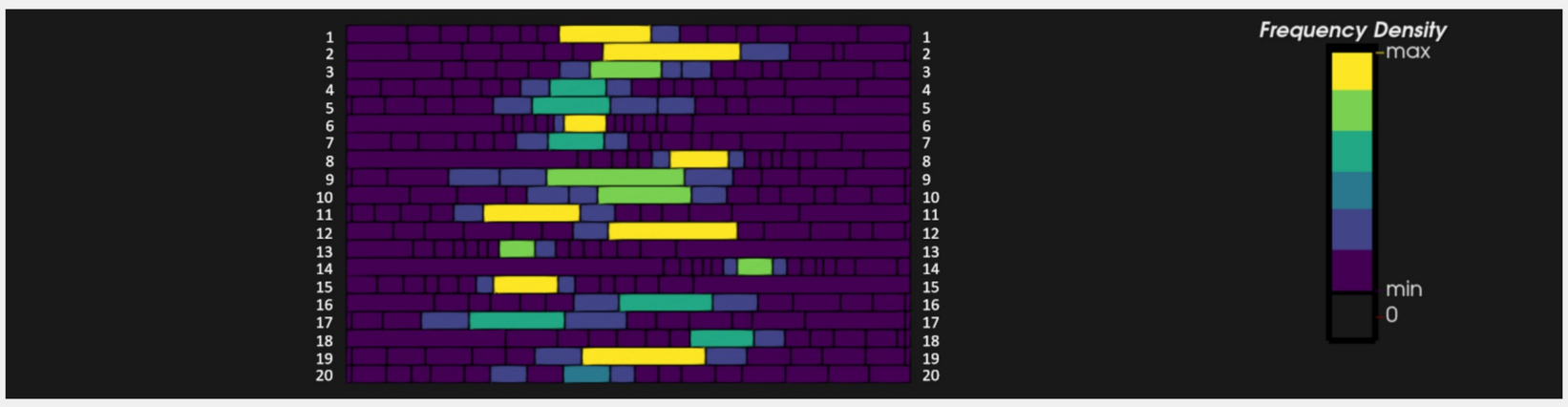
In the following image you will see how **spikes** are represented. Again, the first distribution does not contain a spike, the following 3 distributions have a spike of increasing severity, slightly left of the center. [1] shows the 4 distributions in a density plot, [2] binned through Uniform Binning, [3] in Jenks' Natural Breaks and [4] in Bayesian Blocks. The red lines indicate where the spikes are located and how the spikes are represented via the different binnings.



In this challenge your task is to select the one distribution out of 20 distributions, which has a data flaw. You will see one of the three types of flaws in each question. You will be told what kind of flaw to look for.

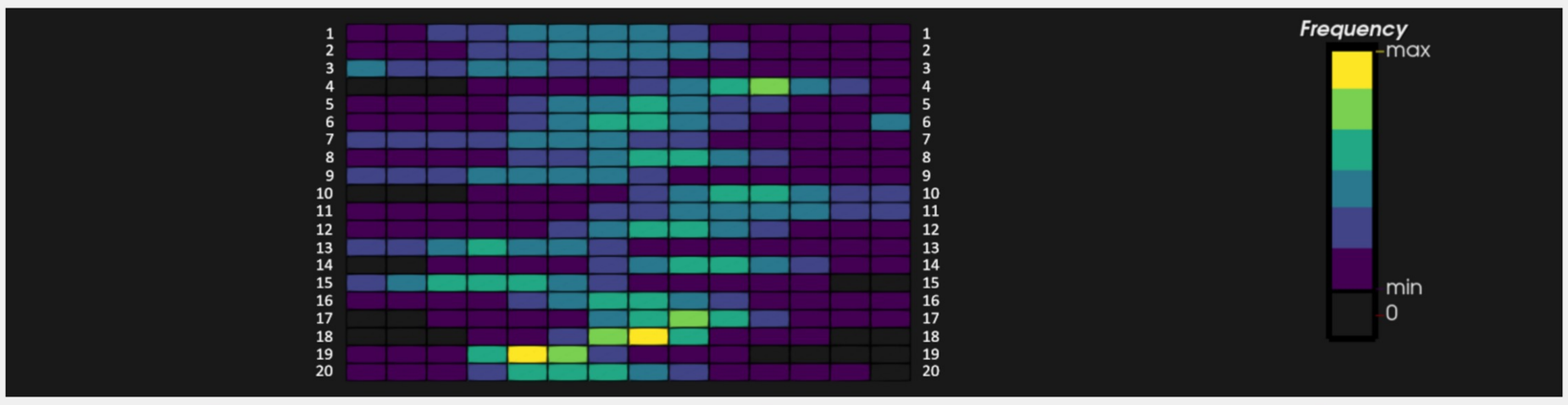
In the following you will see three sample questions, one for each data flaw type.

Choose the distribution, which has a **GAP** in it. (example)



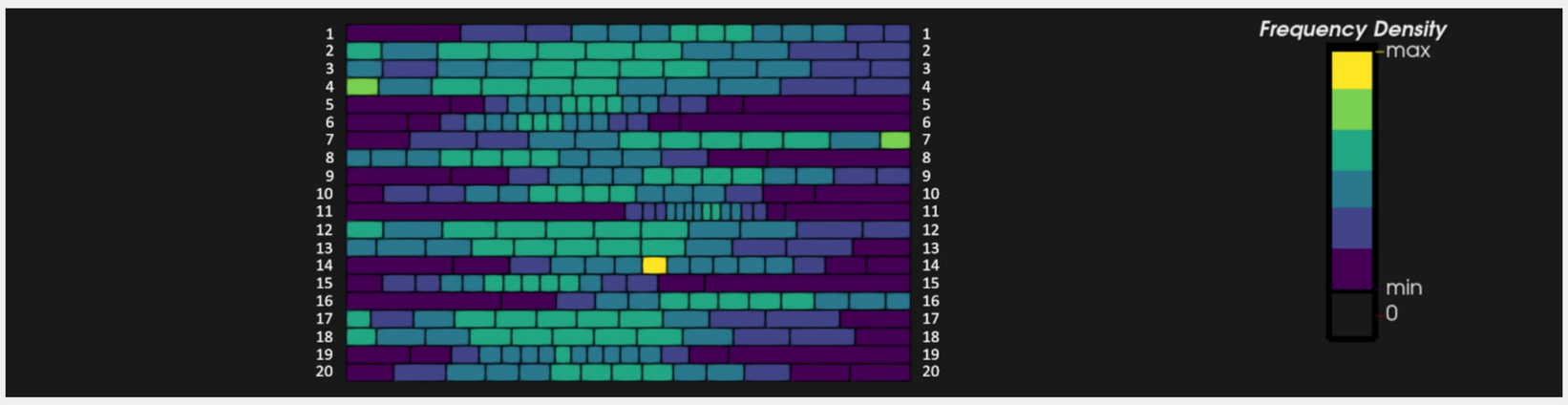
○ 1	○ 6	○ 11	○ 16
○ 2	○ 7	○ 12	○ 17
○ 3	○ 8	○ 13	○ 18
○ 4	○ 9	○ 14	○ 19
○ 5	○ 10	○ 15	○ 20

Choose the distribution, which has a **OUTLIER** in it. (example)



○ 1	○ 6	○ 11	○ 16
○ 2	○ 7	○ 12	○ 17
○ 3	○ 8	○ 13	○ 18
○ 4	○ 9	○ 14	○ 19
○ 5	○ 10	○ 15	○ 20

Choose the distribution, which has a **SPIKE** in it. (example)

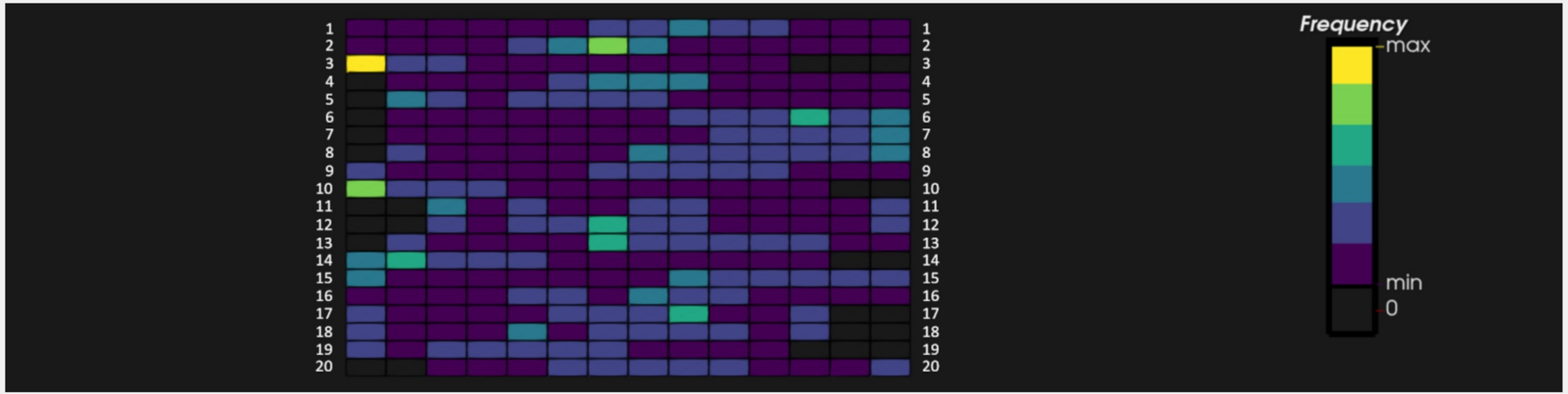


○ 1	○ 6	○ 11	○ 16
○ 2	○ 7	○ 12	○ 17
○ 3	○ 8	○ 13	○ 18
○ 4	○ 9	○ 14	○ 19
○ 5	○ 10	○ 15	○ 20

Continue >>

End Study

Click answer **16!**



- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20

Continue >>

End Study