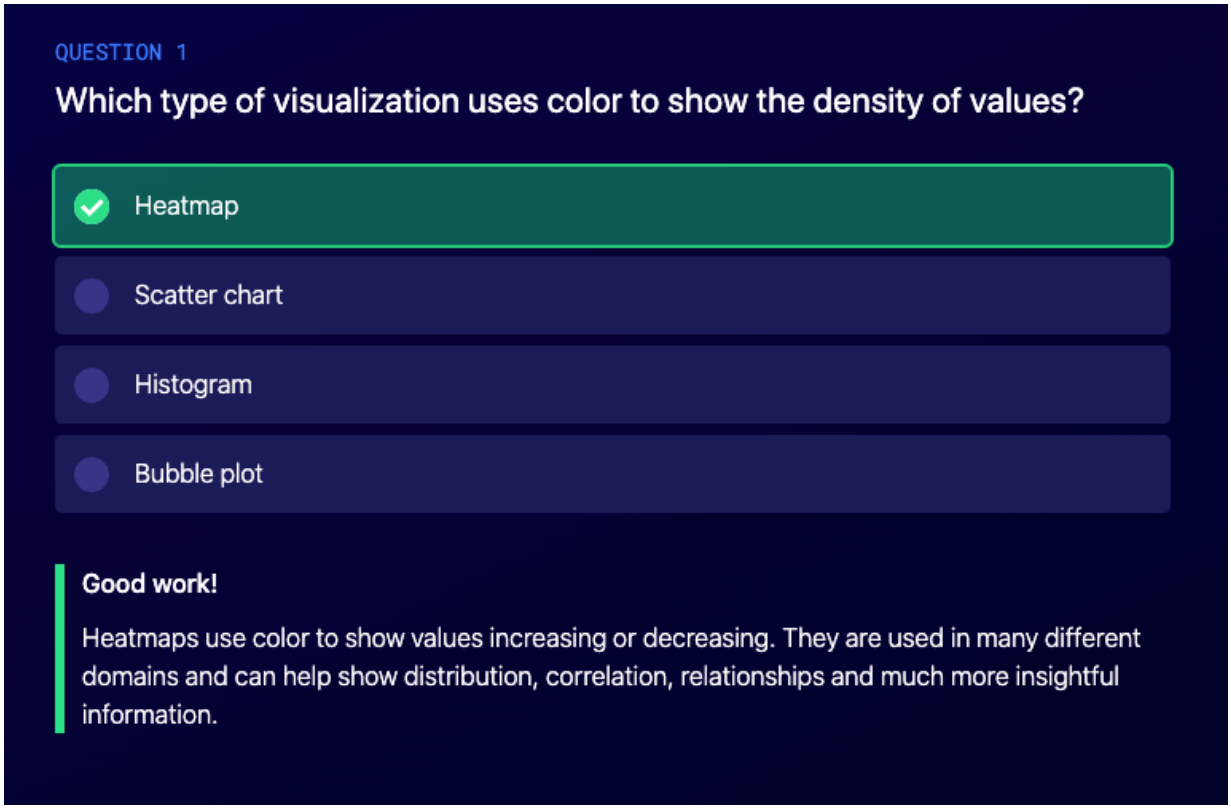
Cloud Guru - 4 - Data Analysis and Visualization Quiz

https://acloud.guru/course/aws-certified-machine-learning-specialty/learn/9ec6163d-ffe3-0975-5904-5d2b2d793493/chapter-5/watch?backUrl=~2Fcourses&backUrl=~2Fcourses&backUrl=~2Fcourses,~2Fcourses



QUESTION 2

Which visualizations help show comparisons?

Choose 2

- [x] Stacked bar chart
- [ ] Scatter plot
- [ ] Histogram
- [x] Stacked area chart
- [ ] Bubble chart
- [x] Line chart
- [x] Bar chart

Sorry!

Correct Answer

Visualizing comparisons in our data is a great way to quickly get an idea on how different values stack up against other values.

QUESTION 3

Which Amazon service allows you to create interactive graphs and charts, and acts as Business Intelligence (BI) tool?

- Athena
- Tableau
- Matplotlib
- ✓ Quicksight

Good work!

QuickSight is a fully managed service in AWS that lets you easily create and publish interactive dashboards. Dashboards can then be accessed from any device, and embedded into your applications, portals, and websites.

QUESTION 4

You are a ML specialist working for a retail organization. You are analyzing data that has different items at different costs. You decide to choose the top 5 most expensive items and visually compare their prices. Which visualization can help you achieve this?

- ✓ Bar chart
- Scatter chart
- Pie chart
- Histogram

Good work!

Bar charts can show single values really well. Each product is represented by a bar extended up the the price of the item.

QUESTION 5

You are a ML specialist designing a regression model to predict the sales for an upcoming festival. The data from the past consists of 1,000 records containing 20 numeric attributes. As you start to analyze the data, you discovered that 30 records have values that are in the far left of a box plot's lower quartile. The festival manager confirmed that those values are unusual, but plausible. There are also 65 records where another numerical value is blank. What should you do to correct these problems?

- Drop the unusual records and fill in the blank values with 0.
- ✗ Use the unusual data and replace the missing values with a separate Boolean variable.
- Drop the unusual records and replace the blank values with separate Boolean values.
- ✓ Drop the unusual records and replace the blank values with the mean value.

Sorry!

Correct Answer

There are many different ways to handle this scenario. We can eliminate the answer that deals with creating separate Boolean. This leaves the two answers with filling in the missing values with 0 or the mean. The mean is going to give us much better results than using 0. We should drop the unusual values and replace the missing values with the mean.

QUESTION 6

Which visualizations help show composition?

Choose 3

- [ ] Bubble chart
- [ ] Bar chart
- [x] Stacked bar chart
- [ ] Box plot
- [x] Stacked area chart
- [ ] Histogram
- [x] Pie chart

Good work!

Visualizing the composition of our data is a great way to show what our data is made of.

QUESTION 7

Which visualizations help show distribution?

Choose 3

- [ ] Line chart
- [ ] Stacked area chart
- [ ] Stacked bar chart
- [x] Histogram
- [x] Box plot
- [x] Scatter chart

Good work!

Distribution can show us how our data is distributed over certain intervals and show us clustering or grouping of data.

QUESTION 8

You are a ML specialist working for a retail organization. You are analyzing customer spending data for particular locations and comparing how it changes over time. You want to visualize the monthly total amount spent at each location over the last 5 years. Which visualization can you use to help you see this?

- Scatter chart
- Histogram
- Line chart ✓
- Bar chart

Good work!

The key words in this question is how the data changes over time. We can sum up the total amount spent by all customers for each month. Place the months on the x axis and the dollar amount on the y axis. Plot a point for each month and connect each point creating a line chart, where each line represent a different location.

QUESTION 9

You are working for a major research university analyzing data about the professors who teach there. The features within the data contain information like employee id, position, department, job description, salary, and tenure. The tenure attribute is binary 0 or 1, whether the professor has tenure or does not have tenure. You need to find the distribution of professors and salaries. What is the best visualization to use to achieve this?

- ✗ Scatter chart
- Pie chart
- ✓ Histogram
- Bubble chart
- Line chart

Sorry!

Correct Answer

Since we are looking for distribution, a histogram is the best visualization to use from these answers.

QUESTION 10

What does the box in a box plot represent?

- [x] The middle 50% of the values.
- [ ] The minimum values.
- [ ] The maximum values.
- [ ] The median value.

Good work!

The box in a box plot represents 50% of the data, or the middle quartile. The line in the box represents the median. The upper/far right of the box plot represents the upper quartile. The lower/left of the box represents the lower quartile.

QUESTION 11

Which visualizations help show relationships?

Choose 2

- [x] Scatter plot
- [ ] Pie chart
- [ ] Stacked area chart
- [ ] Histogram
- [ ] Bar chart
- [ ] Stacked bar chart
- [x] Bubble chart

Good work!

Visualizing relationship in data is important because it shows how different attributes can effect one another. They can also show trends and outliers within our data.

QUESTION 12

You are a ML specialist building a regression model to predict the amount of rainfall for the upcoming year. The data you have contains 18,000 observations collected over the last 50 years. Each observation contains the date, amount of rainfall (in cm), humidity, city, and state. You plot the values in a scatter plot for a given day and amount of rainfall. After plotting points, you find a large grouping of values around 0 cm and 0.2 cm. There is a small grouping of values around 500 cm. What are the reasons for each of these groupings? What should you do to correct these values?

- The groupings around 0 cm and 0.2 cm are extremes and should be removed. The values around 500 cm should be normalized and used once normalized.
- The groupings around 0 cm are days that had no rainfall, the groupings around 0.2 cm are days where it rained, the groupings around 500 cm are days where it snowed. The values should be used as is.
- The groupings around 0 cm are days that had no rainfall, the groupings around 0.2 cm are days where it rained, the groupings around 500 cm are outliers. The values around 500 cm should be normalized so they are on the same scale as the other values.
- ✓ The groupings around 0 cm are days that had no rainfall, the groupings around 0.2 cm are days where it rained, the groupings around 500 cm are outliers. The values around 500 cm should be dropped and the other values should be used as is.

Good work!

Normalizing the values will not help since the values around 500 cm are outliers. There must have been some mistake when the data was created. The groupings around 0 cm are days where it did not rain. The groupings around 0.2 cm are days where it rained. The grouping around 500 cm are extreme values. The values around 500 cm should be dropped and the other values should be used as is.


75%
Congratulations!
You passed AWS Certified Machine Learning - Specialty 2020 - Data Analysis and Visualization Quiz!