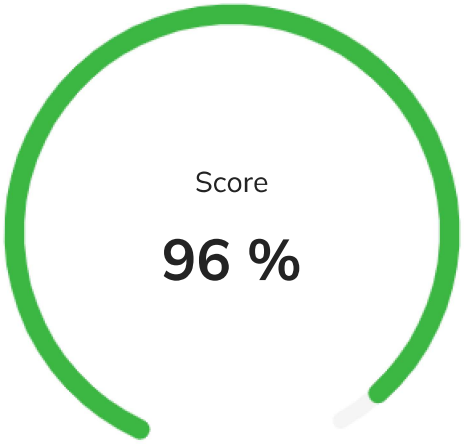
final test 01



Congratulations!

You completed this test on 05/04/2025 at 16:49

Passed

Which of these is an example of ordinal data?

- ✗ Blood type (A, B, AB, O)
- ✓ Grades in school (A, B, C)
- ○ Number of pets owned
- ○ Temperature in Celsius

How do you reshape a NumPy array arr to have 3 rows and 4 columns?

- ○ arr.resize(3,4)
- ✓ arr.reshape(3,4)
- ✗ arr.reshape((3,4))
- ○ arr.shape(3,4)

How do you create an array with values ranging from 1 to 10 in NumPy?

- [ ] np.range(1,10)
- [x] np.arange(1,11)
- [ ] np.linspace(1,10)
- [ ] np.list(1,10)

The normal distribution is:

- [ ] Skewed left
- [x] Bell-shaped and symmetric
- [ ] Uniformly distributed
- [ ] Bimodal

Which distribution shape can a histogram help identify?

- [ ] Normal
- [ ] Skewed
- [ ] Bimodal
- [x] All of the above

A dataset with multiple modes is called:

- [ ] Unimodal
- [ ] Bimodal
- [x] Multimodal
- [ ] Nonmodal

The mode represents:

- [ ] The average value of a dataset
- [x] The value that occurs most frequently
- [ ] The middle value of the dataset
- [ ] The spread of data

In a normal distribution, the mean, median, and mode are:

- [ ] Different
- [x] Equal
- [ ] Random
- [ ] Always zero

How can you assign a default value to a function argument in R?

- [ ] By assigning it in the function body
- [ ] Using the default() function
- [x] Assigning a value in the argument list
- [ ] Using the set() function

In ggplot2, which function is used for a histogram?

- [x] geom_histogram()
- [ ] geom_col()
- [ ] geom_bar()
- [ ] geom_density()

What is the output of if (FALSE) print("Hello")?

- [ ] Hello
- [ ] FALSE
- [ ] NULL
- [x] No output

In ggplot2, which geom function is used to create a line plot?

- [ ] geom_bar()
- [x] geom_line()
- [ ] geom_histogram()
- [ ] geom_col()

A null hypothesis is:

- [x] A statement of no effect or no difference
- [ ] Always true
- [ ] A claim of significant effect
- [ ] Randomly chosen

How do you create an infinite loop in R?

- [ ] while(TRUE) { ... }
- [ ] for (i in 1:Inf) { ... }
- [ ] repeat { ... }
- [x] All of the above

In base R, which argument in heatmap() controls clustering?

- [ ] scale
- [ ] clustering
- [x] hclustfun
- [ ] col

What does np.array([1, 2, 3]) return?

- [ ] A list
- [x] A NumPy array
- [ ] A tuple
- [ ] A dictionary

Which type of plot is most useful for detecting outliers?

- [x] Box plot
- [ ] Line plot
- [ ] Histogram
- [ ] Scatter plot

Type I error occurs when:

- [x] Rejecting a true null hypothesis
- [ ] Accepting a true null hypothesis
- [ ] Rejecting a false null hypothesis
- [ ] No error is made

In ggplot2, how do you convert a bar chart into a pie chart?

- [x] Add coord_polar(theta = "y")
- [ ] Use geom_pie()
- [ ] Apply facet_wrap()
- [ ] Change geom_col() to geom_point()

What visualization is best for checking if a dataset follows a normal distribution?

- [x] Histogram
- [ ] Scatter plot
- [ ] Bar chart
- [ ] Pie chart

Scenario: Employee Age Study

A survey is conducted to study the age distribution of employees in a company. The ages are measured in whole years (e.g., 25, 30, 35).

Question: What type of data is represented by the ages of employees?

- [ ] Continuous
- [ ] Nominal
- [x] Discrete
- [ ] Ordinal

Which chart should be used to analyze the relationship between three numerical variables?

- [x] Scatter plot with color mapping
- [ ] Pie chart
- [ ] Histogram
- [ ] Box plot

If you want to visualize the proportion of missing values in a dataset, which type of plot is most useful?

- [ ] Bar chart
- [x] Heatmap
- [ ] Histogram
- [ ] Line plot

Data such as "Yes" or "No" is:

- [ ] Discrete data
- [ ] Continuous data
- [x] Nominal data
- [ ] Ordinal data

Which Seaborn function is best for visualizing categorical data?

- [ ] sns.barplot()
- [ ] sns.countplot()
- [ ] sns.scatterplot()
- [x] Both a) and b)

How do you generate a random number between 0 and 1 in NumPy?

- [ ] np.random.rand()
- [ ] np.random.random()
- [ ] np.random.randint(0,1)
- [x] Both a) and b)

Which visualization is best for showing the distribution of a numerical variable?

- [x] Histogram
- [ ] Bar Chart
- [ ] Line Plot
- [ ] Scatter Plot

What is the best visualization for correlation between multiple numerical variables?

- [x] Heatmap
- [ ] Pie Chart
- [ ] Line Plot
- [ ] Histogram

How do you add a legend to a Matplotlib plot?

- [x] plt.legend()
- [ ] plt.add_legend()
- [ ] plt.show_legend()
- [ ] plt.make_legend()

Which function is used to create violin plots in Seaborn?

- [x] sns.violinplot()
- [ ] sns.boxplot()
- [ ] sns.stripplot()
- [ ] sns.scatterplot()

Which method is used to drop rows with missing values?

- [x] df.dropna()
- [ ] df.fillna()
- [ ] df.remove_na()
- [ ] df.dropna(axis=1)

What does sns.pairplot(df) do?

- [x] Creates scatter plots for all pairwise relationships
- [ ] Plots a single histogram
- [ ] Shows a bar chart of categorical values
- [ ] Draws a heatmap

How do you change the color palette in Seaborn?

- [x] sns.set_palette("pastel")
- [ ] sns.set_theme("colorful")
- [ ] sns.color_map("red")
- [ ] sns.set_color("blue")

What is the best plot for time-series data?

- [x] Line Chart
- [ ] Bar Chart
- [ ] Pie Chart
- [ ] Scatter Plot

How do you generate a random integer between 10 and 100?

- [x] np.random.randint(10,100)
- [ ] np.random.random(10,100)
- [ ] np.random.uniform(10,100)
- [ ] np.random.normal(10,100)

What does plt.xlabel("X-axis") do?

- Adds a title
- Labels the X-axis
- Labels the Y-axis
- Adds a legend

Which function creates a heatmap in Seaborn?

- sns.heatmap()
- sns.correlationplot()
- sns.matrixplot()
- sns.gridplot()

What argument is used to change the line color in plt.plot()?

- color
- fill
- linecolor
- shade

Which Seaborn function is used to create a histogram?

- sns.histplot()
- sns.distplot()
- sns.barplot()
- sns.scatterplot()

What does arr[1:4] return in NumPy?

- [ ] Elements from index 1 to 4
- [x] Elements from index 1 to 3
- [ ] Elements from index 0 to 3
- [ ] Elements from index 2 to 4

How do you reset the index of a Pandas DataFrame?

- [ ] df.index_reset()
- [ ] df.reindex()
- [x] df.reset_index()
- [ ] df.drop_index()

How do you load built-in datasets in Seaborn?

- [ ] sns.datasets.load_dataset()
- [x] sns.load_dataset()
- [ ] sns.get_data()
- [ ] sns.read_data()

Which argument in geom_density() controls the transparency of the curve?

- [x] alpha
- [ ] color
- [ ] size
- [ ] linetype

In base R, what function is used to create multiple box plots in one plot?

- [x] boxplot(var1, var2, ...)
- [ ] plot.boxplot()
- [ ] multi.boxplot()
- [ ] box(var1, var2, ...)

What does the diagonal in a pair plot represent?

- [ ] Box plots
- [x] Histograms of each variable
- [ ] Correlation values
- [ ] Scatter plots

Which chart is best suited for showing time-series data?

- [ ] Heatmap
- [x] Line chart
- [ ] Scatter plot
- [ ] Bar chart

How do you select a single column from a Pandas DataFrame?

- [ ] df.column_name
- [x] df['column_name']
- [ ] df.column['name']
- [ ] df[[column_name]]

Which method displays the first 5 rows of a DataFrame?

- [x] df.head()
- [ ] df.first()
- [ ] df.display()
- [ ] df.show()

Which parameter controls point size in geom_point()?

- [x] size
- [ ] pointsize
- [ ] width
- [ ] alpha

Which chart is best suited for showing trends over time?

- [ ] Histogram
- [x] Line plot
- [ ] Pie chart
- [ ] Box plot

Which function is used to create a pie chart in base R?

- [ ] barplot()
- [x] pie()
- [ ] hist()
- [ ] plot()

What function is used to create a bar chart in base R?

- [x] barplot()
- [ ] hist()
- [ ] plot()
- [ ] pie()

What will the following code output?

```
greet <- function(name = "Guest") {
  paste("Hello,", name)
}
greet()
```

- [ ] Error
- [x] Hello, Guest
- [ ] Null
- [ ] Guest

What is the output of the following code?

```
add <- function(x, y) { x + y }
add(3, 5)
```

- [x] 8
- [ ] 15
- [ ] Error
- [ ] 3

Which of these returns the first conditionally true expression?

- ifelse()
- switch()
- case_when()
- else

What will be the result of 3^2 + 2 * 3 in R?

- 15
- 18
- 21
- 27

What is the data type of c(TRUE, FALSE, TRUE)?

- Numeric
- Logical
- Character
- Complex

What does df.fillna(0) do?

- Replaces all missing values with 0
- Removes all missing values
- Deletes the entire DataFrame
- Drops rows with missing values

What function in Seaborn is used for KDE (Kernel Density Estimation) plots?

- [x] sns.kdeplot()
- [ ] sns.histplot()
- [ ] sns.densityplot()
- [ ] sns.scatterplot()

What function returns the shape of a NumPy array?

- [ ] shape()
- [x] arr.shape
- [ ] arr.size
- [ ] arr.dimension

What is the main advantage of a scatter plot?

- [ ] Shows categorical relationships
- [x] Displays correlations between two numerical variables
- [ ] Highlights median values
- [ ] Represents time series data

What is the correct function for density plots in ggplot2?

- [x] geom_density()
- [ ] geom_histogram()
- [ ] geom_boxplot()
- [ ] geom_col()

Which function is used in ggplot2 for bar charts?

- [x] geom_bar()
- [ ] geom_point()
- [ ] geom_line()
- [ ] geom_histogram()

What does the return() function do in R?

- [ ] Exits the program
- [x] Exits the function and returns a value
- [ ] Returns to the start of a loop
- [ ] Returns nothing

What happens when break is used in a loop?

- [ ] Skips to the next iteration
- [x] Exits the loop
- [ ] Stops the R session
- [ ] Restarts the loop

What does the %in% operator do in R?

- [ ] Performs element-wise addition
- [x] Checks for membership
- [ ] Combines two vectors
- [ ] Assigns a value

Which function converts a numeric vector into a character vector?

- [ ] as.numeric()
- [x] as.character()
- [ ] as.logical()
- [ ] as.vector()

Which method creates an array of zeros in NumPy?

- [x] np.zeros()
- [ ] np.ones()
- [ ] np.empty()
- [ ] np.full()

What is the correct syntax for a for loop in R?

- [x] for (i in 1:5) { print(i) }
- [ ] for i in range(1:5):
- [ ] for i from 1 to 5:
- [ ] loop (i in 1:5) { print(i) }

Which measure is most affected by outliers?

- [x] Mean
- [ ] Median
- [ ] Mode
- [ ] Interquartile range

Which function creates a box plot in base R?

- [ ] hist()
- [x] boxplot()
- [ ] barplot()
- [ ] density()

Which is not a measure of central tendency?

- [ ] Mean
- [ ] Median
- [ ] Mode
- [x] Standard deviation

The alternative hypothesis represents:

- [ ] The status quo
- [x] The presence of an effect or difference
- [ ] No relationship in data
- [ ] A sample statistic

Simple random sampling ensures:

- [x] Equal chance for every population member to be selected
- [ ] Selection based on convenience
- [ ] Grouping data into clusters
- [ ] Proportional selection of subgroups

Which of the following is an example of a random variable?

- [x] Number of heads in 10 coin tosses
- [ ] A fixed value like 3.14
- [ ] A qualitative description like "red"
- [ ] None of the above

In a normal distribution, about 99.7% of data falls within how many standard deviations?

- [ ] 1
- [ ] 2
- [x] 3
- [ ] 4

What is the total area under a normal distribution curve?

- [ ] 0.5
- [x] 1
- [ ] 2
- [ ] 10

The standard normal distribution has a mean of:

- [ ] 1
- [x] 0
- [ ] -1
- [ ] Undefined

Approximately what percentage of data falls within 1 standard deviation of the mean in a normal distribution?

- [ ] 50%
- [x] 68%
- [ ] 95%
- [ ] 99%

Which measure of dispersion is most robust to outliers?

- [ ] Standard deviation
- [ ] Range
- [ ] Variance
- [x] Interquartile range (IQR)

Range is defined as:

- [x] The difference between the highest and lowest values
- [ ] The average of the dataset
- [ ] The most frequently occurring value
- [ ] The middle value of the dataset

The classification of data into male and female is an example of:

- [x] Nominal data
- [ ] Ordinal data
- [ ] Interval data
- [ ] Continuous data

Which measure of dispersion is most sensitive to outliers?

- [x] Range
- [ ] Interquartile range
- [ ] Standard deviation
- [ ] Median

If the mean of 10 numbers is 15, the sum of the numbers is:

- [x] 150
- [ ] 15
- [ ] 10
- [ ] 100

What is IBM Watson Studio primarily used for?

- [ ] Cloud storage
- [x] Data science and AI model development
- [ ] Web hosting
- [ ] File management

Which programming languages are supported in IBM Watson Studio for data visualization?

- [x] Python and R
- [ ] Java and C++
- [ ] HTML and CSS
- [ ] Swift and Kotlin

Which tool in IBM Watson Studio is specifically used for interactive data visualization?

- [ ] Watson Assistant
- [x] Data Refinery
- [ ] Watson Discovery
- [ ] AutoAI

IBM Watson Studio uses which popular Python libraries for visualization?

- [x] Matplotlib and Seaborn
- [ ] NumPy and Pandas
- [ ] TensorFlow and PyTorch
- [ ] SQL and MongoDB

What is the primary advantage of using IBM Watson for data visualization?

- [ ] Only experts can use it
- [x] No programming is required for basic visualizations
- [ ] It only works with structured data
- [ ] It does not support interactive charts

Which type of chart is best for showing trends over time in Watson Studio?

- [ ] Bar Chart
- [x] Line Chart
- [ ] Pie Chart
- [ ] Scatter Plot

When analyzing the distribution of a single numeric variable, which visualization should you use?

- [ ] Histogram
- [ ] Pie Chart
- [ ] Box Plot
- [x] Both a and c

What type of visualization is most effective for comparing multiple categories in IBM Watson?

- [x] Bar Chart
- [ ] Scatter Plot
- [ ] Heatmap
- [ ] Violin Plot

Which type of visualization is best for showing relationships between two continuous variables?

- [x] Scatter Plot
- [ ] Pie Chart
- [ ] Bar Graph
- [ ] Treemap

Heatmaps in IBM Watson Studio are commonly used for:

- [ ] Showing relationships between categorical variables
- [x] Visualizing correlation between numerical variables
- [ ] Creating pie charts
- [ ] Displaying time-series data

IBM Watson Studio can integrate data from which sources?

- [ ] Cloud databases
- [ ] CSV and Excel files
- [ ] APIs and IoT devices
- [x] All of the above

What is an advantage of using Watson's AI-powered visualizations?

- [x] It predicts patterns in the data
- [ ] It replaces human analysts completely
- [ ] It does not require any data preparation
- [ ] It only supports pre-defined charts

IBM Watson can suggest the best visualization type based on:

- [x] Data structure and relationships
- [ ] Random selection
- [ ] User preferences only
- [ ] Pre-defined templates

Can users customize visualizations in Watson Studio?

- [x] Yes, users can modify colors, labels, and axes
- [ ] No, visualizations are auto-generated
- [ ] Only developers can modify them
- [ ] It depends on the Watson plan

What type of visualization is recommended for detecting outliers?

- [x] Box Plot
- [ ] Pie Chart
- [ ] Line Chart
- [ ] Area Chart

In Watson Studio, what feature allows users to create dashboards with multiple charts?

- [ ] Watson Assistant
- [x] IBM Cognos Analytics
- [ ] Data Refinery
- [ ] AutoML