How can categorical data be plotted in Pandas?

Authored by

stats writer

June 26, 2024

Categorical data can be easily plotted in Pandas using the built-in function "plot()" which allows for various types of plots such as bar charts, pie charts, and histograms. This function takes in the categorical data as the x-axis and the desired plot type as the y-axis, allowing for a quick and efficient visualization of the data. Additionally, Pandas allows for customization of the plots through the use of parameters, making it a versatile tool for plotting categorical data. With its user-friendly interface and powerful features, Pandas is an ideal choice for visualizing and analyzing categorical data.

Plot Categorical Data in Pandas (With Examples)

There are three common ways to visualize :

Bar ChartsBoxplots by GroupMosaic Plots

The following examples show how to create each of these plots for a pandas DataFrame in Python.

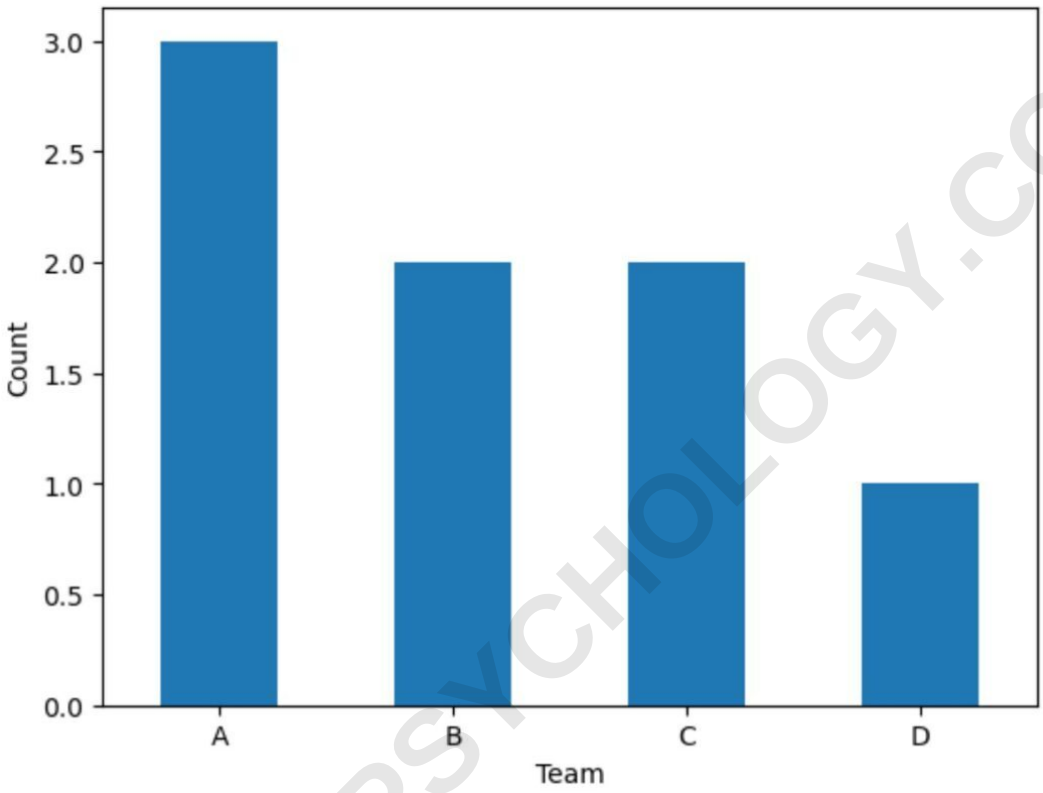
Example 1: Bar Charts

The following code shows how to create a bar chart to visualize the frequency of teams in a certain pandas DataFrame:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'points': })
```

#create bar plot to visualize frequency of each team
df.value_counts().plot(kind='bar', xlabel='Team', ylabel='Count', rot=0)



The x-axis displays each team name and the y-axis shows the frequency of each team in the DataFrame.

Note: The argument rot=0 tells pandas to rotate the x-axis labels to be parallel to the x-axis.

Example 2: Boxplots by Group

Grouped boxplots are a useful way to visualize a

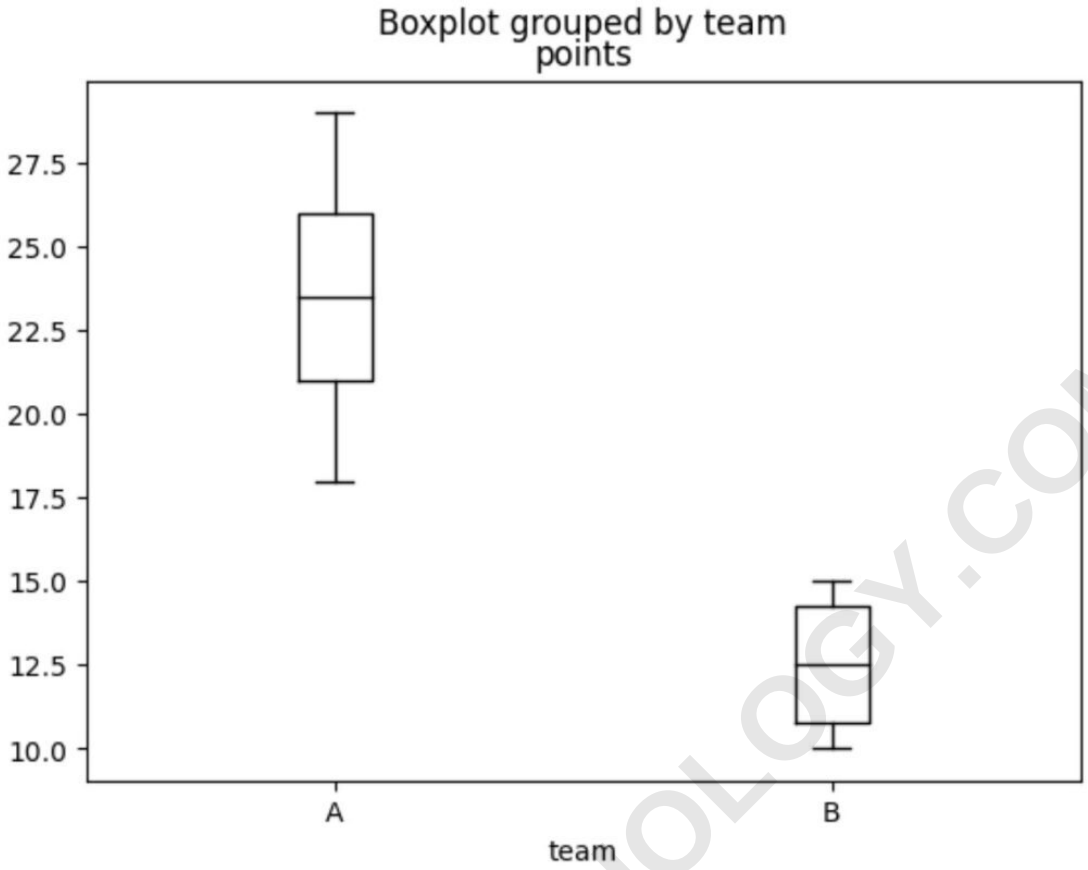
numeric variable, grouped by a categorical variable.

For example, the following code shows how to create boxplots that show the distribution of points scored, grouped by team:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'points': })

#create boxplot of points, grouped by team
df.boxplot(column=, by='team', grid=False, color='black')
```



The x-axis displays the teams and the y-axis displays the distribution of points scored by each team.

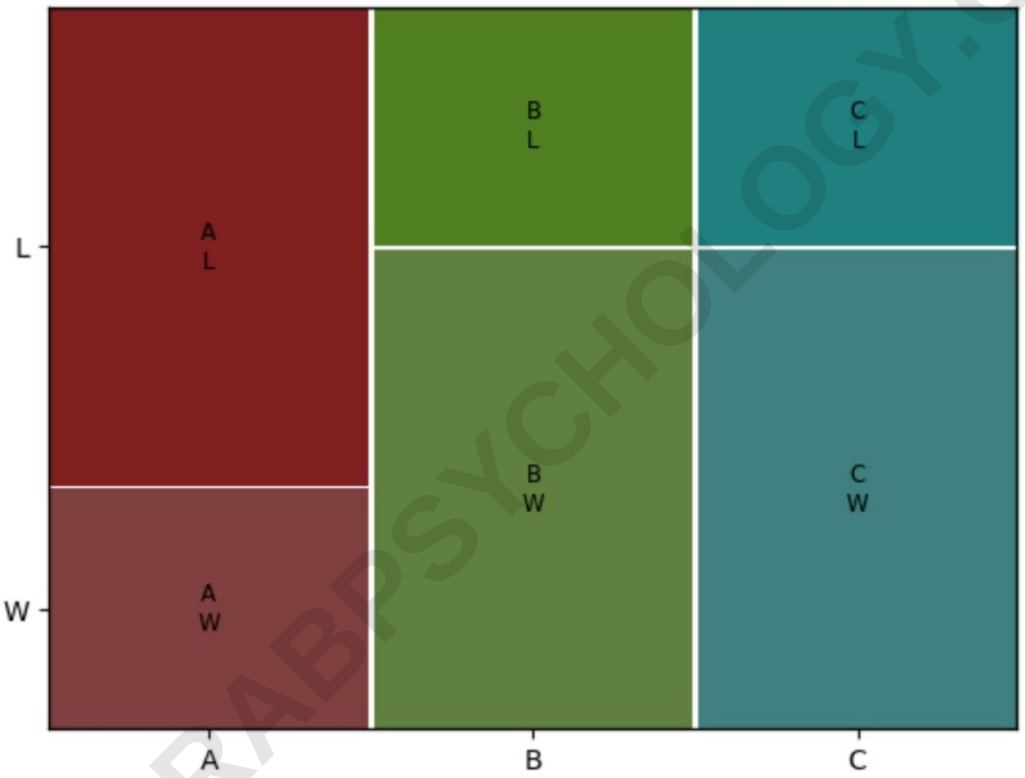
Example 3: Mosaic Plot

A mosaic plot is a type of plot that displays the frequencies of two different categorical variables in one plot.

```
import pandas as pd
from statsmodels.graphics.mosaicplotimport mosaic
```

```
#create DataFrame
df = pd.DataFrame({'team': ,
'result': })

#create mosaic plot
mosaic(df, );
```



The x-axis displays the teams and the y-axis displays the frequency of results for each team.

The following tutorials explain how to perform other common tasks in pandas: