# **Breakout Sessions Handout**

## **Breakout Session 1:**

|  |
| --- |
| *# Box Plot Exercise*  ## Overview  # In this exercise, you'll create and interpret box plots to analyze temperature distributions across different months.  ## Instructions  # 1. Use the provided mock temperature data in the starter code  # 2. Create box plots to visualize the temperature distribution by month  # 3. Customize your box plot with appropriate labels and title  # 4. Answer the questions below based on your visualization  ## Questions  # 1. Which month shows the highest median temperature?  # 2. Which month has the widest temperature range (most variability)?  # 3. Are there any outliers visible in the data? If so, in which months?  # 4. Compare summer and winter months. How do their temperature distributions differ?  # 5. Based on the box plot, which month appears to have the most consistent temperatures?  ## Starter Code  # Import necessary libraries  import pandas as pd  import matplotlib.pyplot as plt  import seaborn as sns  import numpy as np  # Create mock temperature data  # This dictionary contains temperature readings for each month  # Each month has 30 temperature readings in degrees Celsius  monthly\_temps = {  'Jan': [1.3, -2.5, 0.2, -1.5, 2.3, 0.8, -3.2, -2.5, 1.2, -0.8,  -1.5, 2.1, 0.5, -2.8, -0.4, -1.9, 0.7, -3.5, -2.2, 0.9,  -1.1, -0.3, 1.8, -2.1, 0.3, 1.1, -1.7, -0.9, -15.0, 0.4],  'Feb': [0.2, 2.5, -1.3, 3.8, 1.7, -2.1, 4.2, 0.8, -3.5, 2.9,  -0.7, 5.1, 1.3, -2.8, 3.3, 0.5, -1.9, 4.7, 2.1, -0.3,  5.5, 3.2, 0.1, -2.4, 4.3, 1.5, -3.1, 3.7, 0.3, 2.8],  'Mar': [4.8, 7.2, 5.5, 3.1, 8.4, 6.7, 9.3, 5.2, 4.0, 7.9,  6.3, 8.8, 5.7, 3.5, 9.1, 7.5, 4.3, 6.8, 8.2, 5.9,  3.8, 7.3, 9.5, 6.1, 4.9, 8.0, 5.3, 7.7, 4.5, 6.5],  'Apr': [12.3, 9.8, 11.5, 14.2, 10.7, 13.0, 9.5, 12.8, 15.0, 11.2,  14.5, 10.3, 13.7, 11.9, 9.2, 14.8, 12.1, 10.5, 13.3, 15.2,  11.7, 14.0, 10.1, 12.5, 15.5, 13.8, 10.9, 14.3, 11.3, 12.7],  'May': [16.5, 18.2, 17.0, 19.8, 16.7, 18.5, 17.3, 19.0, 16.2, 18.8,  17.5, 19.3, 16.9, 18.0, 17.7, 19.5, 16.3, 18.7, 17.2, 19.2,  16.8, 18.3, 17.8, 19.7, 16.1, 18.9, 17.4, 19.1, 16.6, 18.1],  'Jun': [21.3, 23.7, 20.5, 24.2, 22.8, 26.0, 21.5, 25.3, 22.0, 24.8,  23.2, 20.7, 25.5, 21.7, 24.5, 23.0, 26.3, 22.5, 20.2, 25.8,  23.5, 21.0, 24.0, 22.3, 25.0, 23.8, 21.8, 24.7, 22.7, 20.9],  'Jul': [25.8, 28.3, 27.0, 30.5, 29.2, 26.5, 31.0, 27.5, 29.8, 28.7,  26.2, 30.0, 28.5, 27.2, 29.5, 31.2, 28.0, 26.8, 30.7, 29.0,  27.8, 31.5, 29.3, 26.3, 28.8, 30.2, 27.3, 29.7, 38.0, 28.2],  'Aug': [26.3, 29.8, 28.0, 24.5, 27.3, 30.2, 28.7, 25.5, 29.0, 27.7,  24.0, 28.3, 30.5, 27.0, 29.3, 26.5, 28.8, 25.0, 29.5, 27.5,  30.0, 26.0, 28.5, 24.8, 27.8, 29.2, 26.8, 30.7, 28.2, 25.8],  'Sep': [22.0, 20.7, 23.5, 21.3, 24.0, 22.8, 21.5, 23.2, 20.5, 22.5,  23.8, 21.0, 22.3, 20.2, 23.0, 21.8, 24.2, 22.2, 20.8, 23.7,  21.7, 24.5, 22.7, 21.2, 23.3, 20.3, 22.8, 24.7, 21.5, 23.0],  'Oct': [15.7, 13.2, 16.5, 12.8, 14.3, 17.0, 13.8, 15.2, 14.0, 16.3,  12.5, 15.0, 13.5, 16.8, 14.7, 12.2, 15.5, 17.2, 13.0, 14.8,  16.0, 12.7, 15.3, 14.5, 17.7, 13.3, 16.2, 12.0, 14.2, 15.8],  'Nov': [7.2, 9.8, 6.5, 10.3, 8.7, 5.0, 9.5, 7.8, 11.0, 6.3,  8.2, 10.5, 7.0, 9.2, 5.8, 8.5, 6.8, 10.8, 9.0, 7.5,  5.5, 8.0, 10.0, 7.3, 9.7, 6.0, 8.8, 7.7, 5.3, 9.3],  'Dec': [2.7, -0.5, 3.8, 1.0, -1.8, 2.3, 0.2, -2.5, 3.0, 1.5,  -1.0, 2.0, -3.2, 1.2, 3.5, 0.8, -2.2, 2.5, 0.5, -1.5,  3.2, 1.8, -0.3, 2.2, 0.0, -2.8, 15.0, 1.3, -1.3, 2.8]  }  # Convert the dictionary to a DataFrame  data = []  for month, temps in monthly\_temps.items():  for temp in temps:  data.append({'month': month, 'temperature': temp})  temperatures = pd.DataFrame(data)  # Make sure months are in correct order (Jan to Dec)  month\_order = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun',  'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec']  temperatures['month'] = pd.Categorical(temperatures['month'],  categories=month\_order,  ordered=True)  # Your code begins here:  # 1. Create a box plot of the temperature by month  # 2. Add appropriate labels and title  # 3. Display the plot  plt.savefig('plot.png') |

## **Breakout Session 2:**

|  |
| --- |
| *# Histogram Exercise*  ## Overview  # In this exercise, you'll create and interpret histograms to analyze daily temperature distributions across different seasons.  ## Instructions  # 1. Use the provided mock temperature data in the starter code  # 2. Create histograms to visualize temperature distributions  # 3. Experiment with different bin sizes and normalization  # 4. Create separate histograms to compare seasonal distributions  # 5. Answer the questions below based on your visualizations  ## Questions  # 1. What shape does the overall temperature distribution have? Is it normal, skewed, bimodal, or something else?  # 2. How does changing the number of bins affect your interpretation of the data?  # 3. How do the temperature distributions differ between winter and summer?  # 4. Which season shows the widest spread of temperatures?  # 5. Based on the histogram, what temperature range occurs most frequently in the dataset?  ## Starter Code  # Import necessary libraries  import matplotlib.pyplot as plt  import numpy as np  import pandas as pd  # Create mock temperature data  # This dictionary contains temperature readings for each season  # Each season has 100 temperature readings in degrees Celsius  seasonal\_temps = {  'Winter': [  -5.2, -3.8, -7.1, -2.5, -6.3, -4.7, -0.9, -5.8, -3.1, -6.7,  -2.0, -4.5, -8.3, -1.7, -5.0, -3.5, -7.8, -2.3, -6.0, -4.1,  -1.0, -5.5, -3.3, -7.5, -2.8, -6.5, -0.5, -4.3, -8.0, -3.7,  0.2, -2.1, -6.8, -1.5, -5.3, -3.0, -7.2, -4.8, -2.6, -0.3,  -3.8, -5.7, -1.2, -6.2, -4.0, -8.5, -3.2, -0.8, -5.0, -2.4,  1.8, -4.6, -2.9, -6.7, -0.1, -5.4, -3.6, -7.9, -1.9, -4.2,  0.8, -3.4, -7.3, -1.8, -5.1, -2.7, -0.4, -6.3, -3.9, -8.2,  -2.2, -5.8, -1.1, -4.5, -7.0, -3.5, -0.7, -5.2, -2.3, -6.4,  0.5, -4.8, -1.4, -7.6, -3.2, -5.5, -0.2, -6.1, -3.7, -8.8,  -1.6, -4.9, -2.6, -5.9, -3.4, -7.1, -0.5, -6.5, -2.0, -4.4  ],  'Spring': [  8.3, 12.1, 10.5, 7.8, 13.6, 9.2, 11.8, 6.7, 14.3, 10.0,  7.5, 12.7, 9.8, 11.2, 8.0, 13.4, 10.3, 6.9, 14.7, 9.5,  12.2, 7.1, 11.5, 13.8, 8.7, 10.9, 6.4, 14.0, 9.1, 12.5,  7.9, 11.7, 10.1, 14.5, 8.5, 13.2, 6.8, 12.0, 9.6, 7.2,  14.8, 10.6, 8.2, 11.9, 13.0, 7.6, 12.4, 9.0, 15.1, 10.7,  8.9, 13.5, 11.0, 7.0, 12.8, 9.4, 14.2, 10.2, 6.5, 11.3,  9.7, 14.9, 8.1, 12.6, 7.4, 10.8, 13.9, 9.3, 11.6, 8.4,  14.6, 10.4, 7.3, 13.1, 11.4, 8.8, 12.9, 6.6, 14.4, 9.9,  11.1, 7.7, 13.3, 10.5, 12.3, 8.6, 15.0, 9.2, 11.8, 7.5,  14.1, 10.0, 8.3, 13.7, 6.4, 12.1, 7.8, 14.5, 11.3, 9.6  ],  'Summer': [  23.5, 26.2, 28.9, 25.0, 30.5, 27.8, 24.3, 29.1, 26.6, 22.0,  28.2, 25.7, 31.0, 23.8, 27.3, 29.8, 25.3, 22.7, 28.4, 26.0,  31.5, 24.7, 29.2, 27.0, 23.0, 30.7, 25.8, 28.6, 22.5, 26.4,  29.5, 24.0, 27.5, 32.0, 25.2, 29.7, 26.9, 23.3, 30.2, 28.0,  25.6, 22.2, 29.4, 26.8, 31.3, 23.7, 27.2, 29.9, 25.4, 22.8,  28.3, 24.5, 30.9, 26.7, 23.2, 29.0, 25.9, 32.5, 27.7, 24.2,  28.7, 26.3, 23.9, 31.8, 25.1, 27.6, 30.0, 22.9, 29.6, 26.5,  24.8, 32.3, 28.1, 25.5, 22.3, 30.4, 27.4, 23.6, 26.1, 28.8,  31.2, 24.9, 27.9, 30.8, 25.0, 22.6, 29.3, 26.2, 23.4, 30.6,  28.5, 24.6, 32.8, 27.1, 23.1, 26.3, 29.9, 25.5, 31.7, 24.1  ],  'Fall': [  18.3, 15.7, 12.9, 17.0, 14.2, 11.5, 16.1, 13.6, 10.8, 18.7,  15.0, 12.2, 16.8, 13.1, 17.5, 14.7, 11.0, 19.2, 16.3, 13.8,  10.5, 17.2, 14.5, 11.8, 15.5, 12.0, 18.0, 14.9, 11.3, 16.6,  13.2, 18.5, 15.3, 12.5, 17.8, 14.3, 11.6, 16.2, 19.0, 15.9,  12.7, 18.2, 14.0, 10.2, 17.3, 13.4, 15.8, 12.3, 16.5, 18.8,  14.6, 11.9, 17.4, 13.7, 10.1, 15.2, 18.9, 16.0, 12.8, 14.1,  17.9, 13.5, 16.4, 11.7, 18.6, 15.4, 12.6, 16.9, 13.9, 10.4,  17.6, 14.8, 11.2, 16.7, 15.1, 12.4, 18.1, 14.4, 19.5, 13.0,  17.1, 15.6, 12.1, 18.4, 13.3, 10.9, 15.0, 16.5, 19.3, 14.7,  11.4, 17.7, 13.6, 15.2, 10.7, 18.5, 14.0, 11.8, 16.3, 12.9  ]  }  # Convert the dictionary to a DataFrame  data = []  for season, temps in seasonal\_temps.items():  for temp in temps:  data.append({'season': season, 'temperature': temp})  temperatures = pd.DataFrame(data)  # Set the order of seasons for plotting  season\_order = ['Winter', 'Spring', 'Summer', 'Fall']  temperatures['season'] = pd.Categorical(temperatures['season'],  categories=season\_order,  ordered=True)  # Your code begins here:  # 1. Create a histogram of overall temperature distribution  # 2. Create histograms with different bin sizes  # 3. Create histograms comparing seasonal distributions  # 4. Save your figures (use savefig instead of plt.show())  # plt.savefig('histogram\_filename.png') |