

Machine Learning for Robotics Lab 1

Name : C Uday Karthik

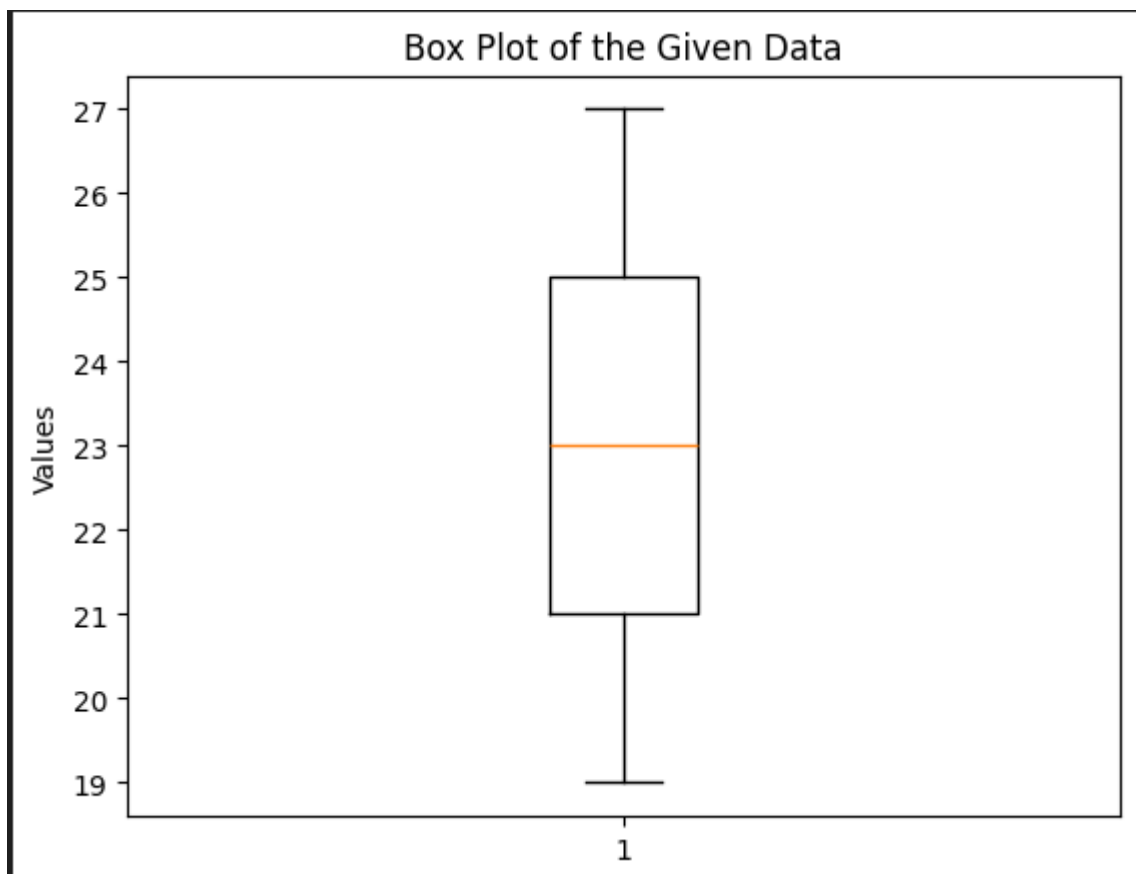
Reg.No : 23BRS1046

Example 1:

Code:

```
#Example 1
import matplotlib.pyplot as plt
data = [20, 19, 21, 22, 23, 24, 25, 27, 26]
plt.boxplot(data)
plt.title("Box Plot of the Given Data")
plt.ylabel("Values")
plt.show()
```

Output:



Example 2:

Code:

```
#Example 2

import pandas as pd
import matplotlib.pyplot as plt

Data=pd.read_csv('Lab1_Vanilla.csv')

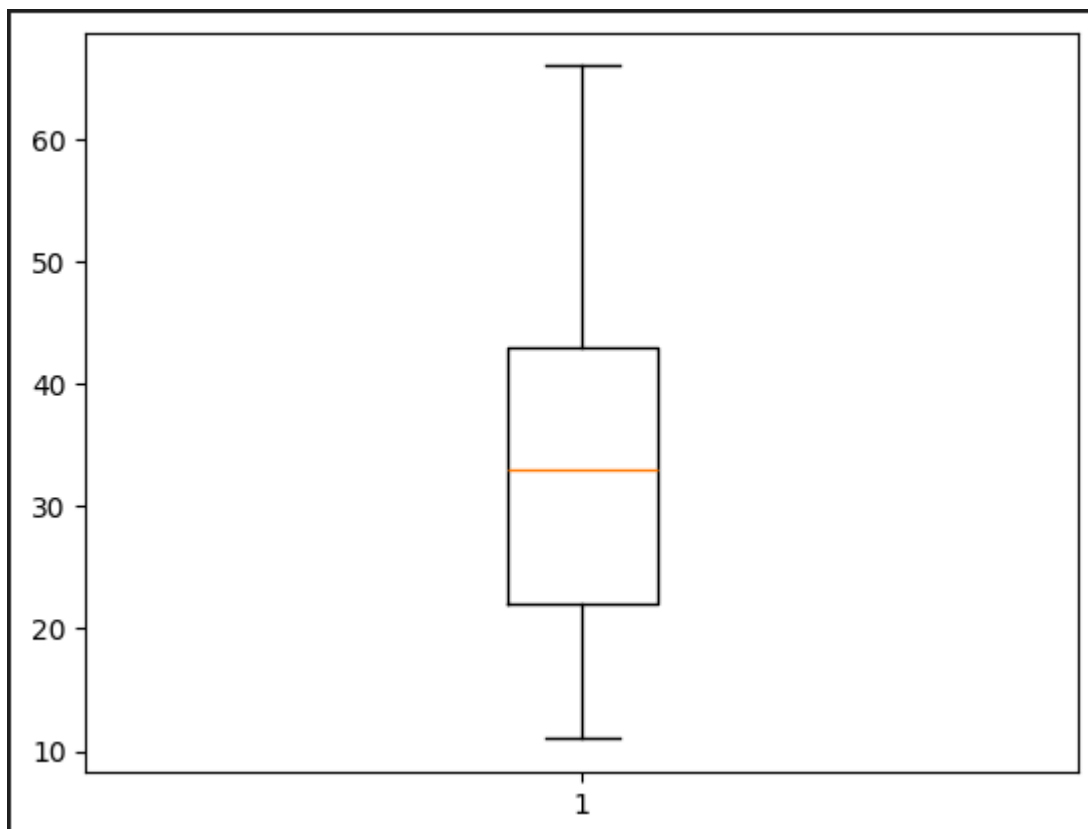
# Creating plot
plt.boxplot(Data)

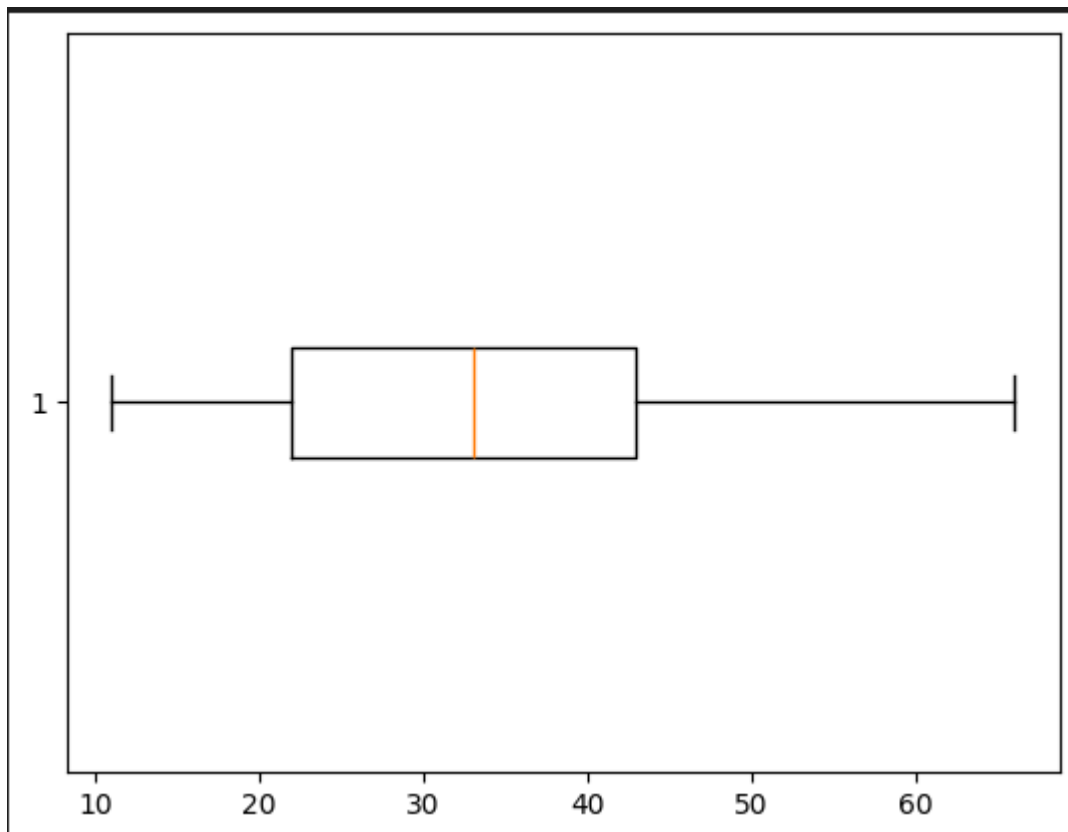
plt.show()

plt.boxplot(Data, vert=False)

# show plot
plt.show()
```

Output:



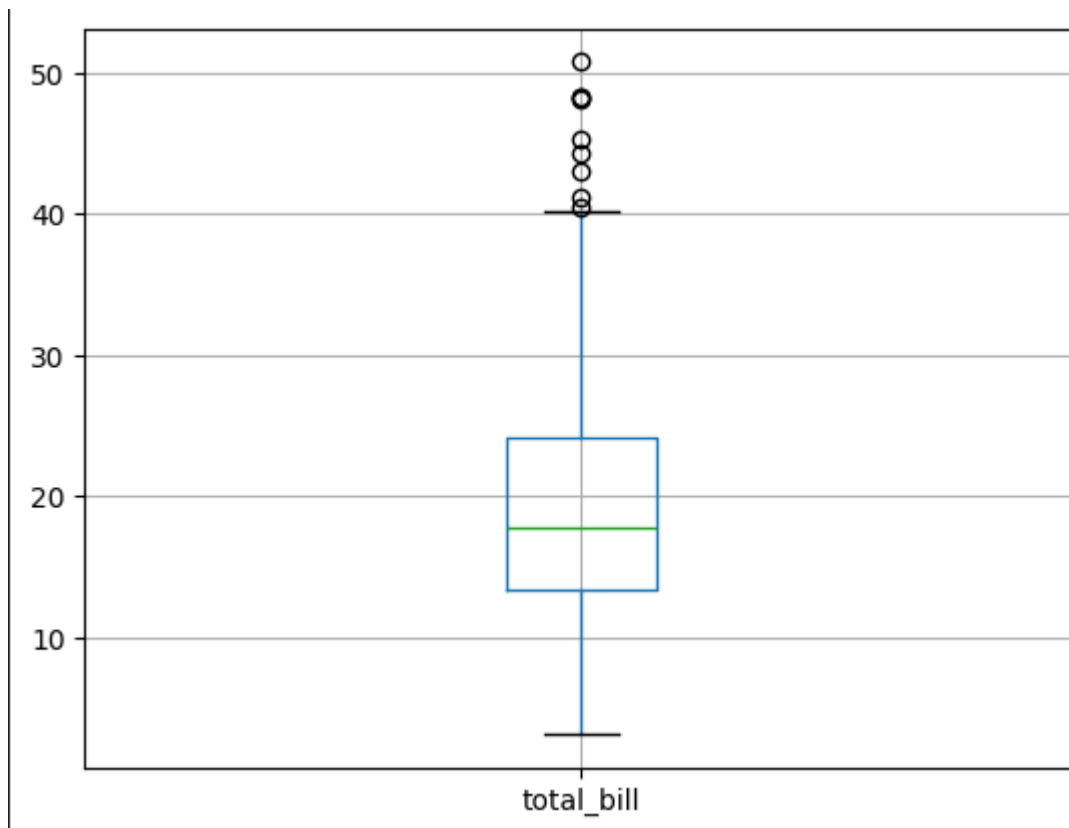


Example 3:

Code:

```
#Example 3
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
Data =pd.read_csv('Lab1_Tips.csv')
Data.boxplot(column =['total_bill'])
plt.show()
```

Output:



Example 4:

Code:

#Example 4

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
Data =pd.read_csv('Lab1_Tips.csv')
```

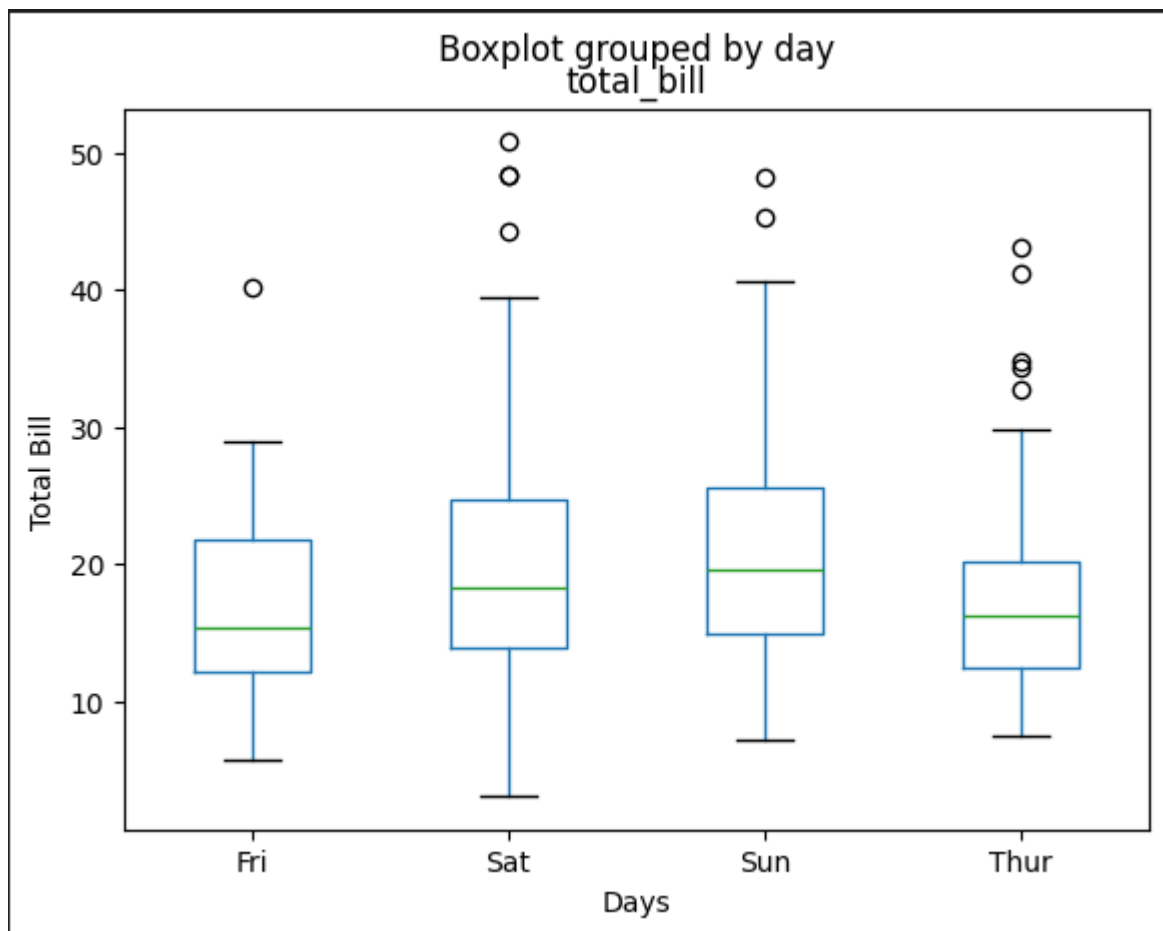
```
Data.boxplot(by ='day', column =['total_bill'], grid=False)
```

```
plt.ylabel('Total Bill')
```

```
plt.xlabel('Days')
```

```
plt.show()
```

Output:



Example 5:

Code:

```
#Example 5
```

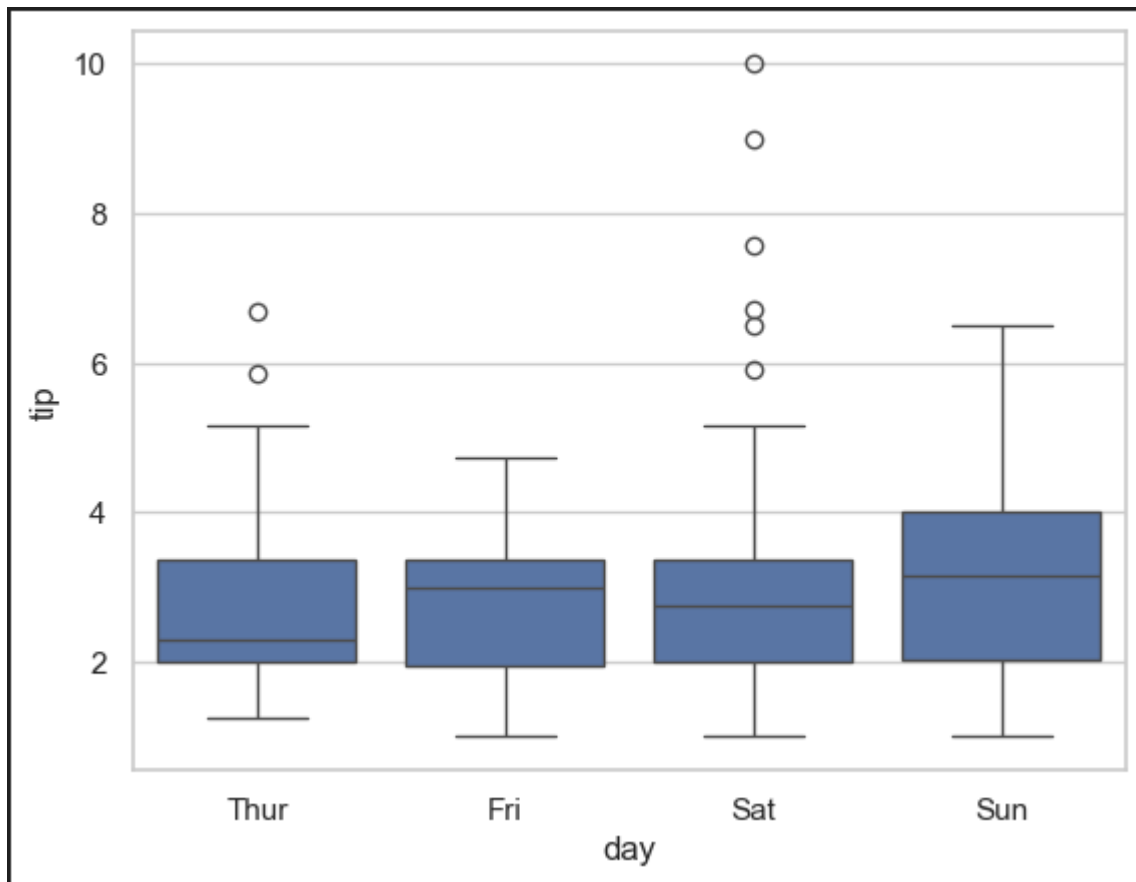
```
import seaborn
```

```
seaborn.set_theme(style='whitegrid')
```

```
tip = seaborn.load_dataset('tips')
```

```
seaborn.boxplot(x='day', y='tip', data=tip)
```

Output:



Example 6:

Code:

```
#Example 6

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# -----
# 1. Load Kaggle dataset
# -----

df = pd.read_csv("Lab1_StudentsPerformance.csv")
print("First 5 rows:")
print(df.head())

# -----
```

```

# 2. Select numeric features
# -----
numeric_cols = ["math score", "reading score", "writing score"]
df_numeric = df[numeric_cols]
print("\nSummary Statistics:")
print(df_numeric.describe())
# -----

# 3. Box plots to show variability
# -----

plt.figure(figsize=(12, 5))
for i, col in enumerate(numeric_cols):
    plt.subplot(1, 3, i+1)
    sns.boxplot(y=df[col])
    plt.title(f"Variability in {col}")
plt.tight_layout()
plt.show()
# -----

# 4. Interpretation (printed)
# -----

print("\nInterpretation Guide:")
print("- The height of each box shows the spread of the middle 50% of scores.")
print("- Whiskers show the overall range of performance.")
print("- Dots (if any) indicate outliers—students with unusually high/low scores.")
print("- Seeing unequal spreads helps understand variability before modeling.")

```

Output:

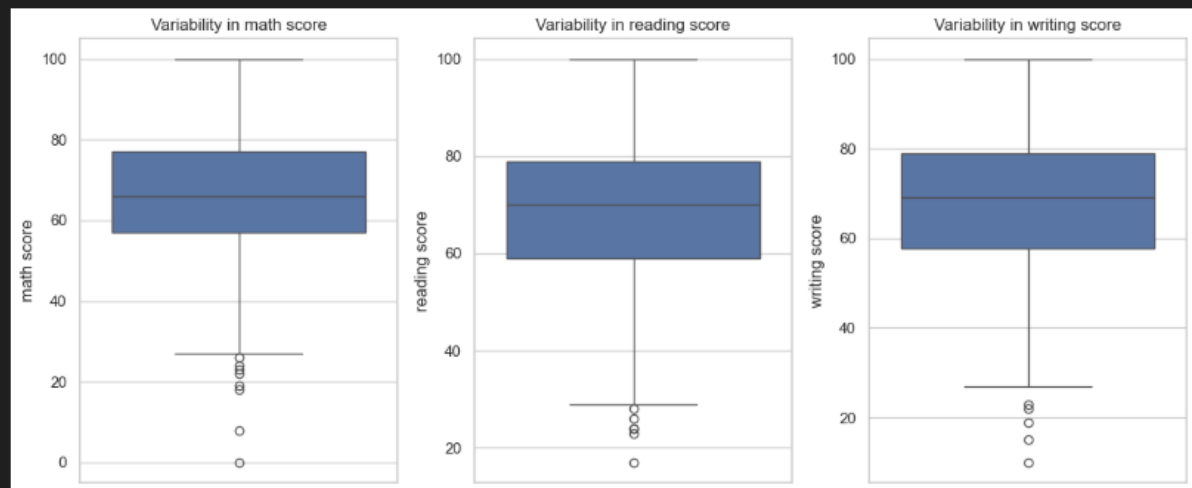
First 5 rows:

	gender	race/ethnicity	parental level of education	lunch
0	female	group B	bachelor's degree	standard
1	female	group C	some college	standard
2	female	group B	master's degree	standard
3	male	group A	associate's degree	free/reduced
4	male	group C	some college	standard

	test preparation course	math score	reading score	writing score
0	none	72	72	74
1	completed	69	90	88
2	none	90	95	93
3	none	47	57	44
4	none	76	78	75

Summary Statistics:

	math score	reading score	writing score
count	1000.00000	1000.00000	1000.00000
mean	66.08900	69.16900	68.05400
std	15.16308	14.600192	15.195657
min	0.00000	17.00000	10.00000
25%	57.00000	59.00000	57.75000
50%	66.00000	70.00000	69.00000
75%	77.00000	79.00000	79.00000
max	100.00000	100.00000	100.00000



Interpretation Guide:

- The height of each box shows the spread of the middle 50% of scores.
- Whiskers show the overall range of performance.
- Dots (if any) indicate outliers—students with unusually high/low scores.
- Seeing unequal spreads helps understand variability before modeling.

Exercise 1:

Code:

```
#Exercise 1

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# -----

# 1. Load Kaggle dataset
# -----

df = pd.read_csv("Lab1_tips.csv")
# -----

# 2. Select numeric features
# -----

numeric_cols = ["total_bill", "tip", "size"]
df_numeric = df[numeric_cols]
print("\nSummary Statistics:")
print(df_numeric.describe())
# -----

# 3. Box plots to show variability
# -----

plt.figure(figsize=(12, 5))
for i, col in enumerate(numeric_cols):
    plt.subplot(1, 3, i+1)
    sns.boxplot(y=df[col])
    plt.title(f"Variability in {col}")
plt.tight_layout()
plt.show()
```

Output:

Summary Statistics:

	total_bill	tip	size
count	244.000000	244.000000	244.000000
mean	19.785943	2.998279	2.569672
std	8.902412	1.383638	0.951100
min	3.070000	1.000000	1.000000
25%	13.347500	2.000000	2.000000
50%	17.795000	2.900000	2.000000
75%	24.127500	3.562500	3.000000
max	50.810000	10.000000	6.000000

