

In [1]:

```
import pandas as pd
import numpy as np
import statistics as stats
import matplotlib.pyplot as plt
import seaborn as sns
```

Sunspots Data

In [2]:

```
df = pd.read_csv('./Sunspots.csv', index_col=0) # First column would be treated as index values which are just records
```

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3235 entries, 0 to 3234
Data columns (total 2 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   Date                                3235 non-null   object
 1   Monthly Mean Total Sunspot Number  3235 non-null   float64
dtypes: float64(1), object(1)
memory usage: 75.8+ KB
```

In [4]:

```
df.head()
```

Out[4]:

	Date	Monthly Mean Total Sunspot Number
0	1749-01-31	96.7
1	1749-02-28	104.3
2	1749-03-31	116.7
3	1749-04-30	92.8
4	1749-05-31	141.7

In [5]:

```
# .loc is an accessor, Filtering first 30 (29 is inclusive) records and Monthly.. col
data = df.loc[:29, 'Monthly Mean Total Sunspot Number'].to_numpy()
# sort doesn't effects mean but nemerical order is necessary for median (for sake of manual calculation)
data = np.sort(data)
print(data)

[ 72.5  75.5  84.5  92.8  94.   96.7 101.2 104.3 105.5 109.5 110.5 116.7
 116.7 122.2 125.7 125.8 126.5 126.5 139.2 141.7 142.   142.3 147.2 148.7
 150.  152.  158.  166.7 171.7 264.3]
```

We would be dealing with first 30 records of Montly Mean Total Sunspot Number

Finding Mean

3830.9 / 30

Mean = 127.6967

Finding Median

Numbers in Numerical Order =>

72.5, 75.5, 84.5, 92.8, 94.0, 96.7, 101.2, 104.3, 105.5, 109.5, 110.5,

116.7, 116.7, 122.2, **125.7, 125.8**, 126.5, 126.5, 139.2, 141.7,

142.0, 142.3, 147.2, 148.7, 150.0, 152.0, 158.0, 166.7, 171.7, 264.3

As we have even length so we take middle two numbers and take their mean

Middle two numbers are : 125.7, 125.8, their mean results in median: $125.7 + 125.8 / 2 \Rightarrow 125.75$

Median = 125.75

Finding Q1

$0.25 \times (30 + 1) - 1 \Rightarrow 6.75$ value which is **103.525**

Finding Q3

$0.75 \times (30 + 1) - 1 \Rightarrow 22.25$ value which is **147.575**

In [6]:

```
print("Mean is :", np.mean(data))
print("Median is:", np.median(data))
print("Quantiles are:", stats.quantiles(data))
# print("Q1 is:", np.quantile(data, q=0.25))
# print("Q3 is:", np.quantile(data, q=0.75)) # It shows different?!
```

Mean is : 127.69666666666667

Median is: 125.75

Quantiles are: [103.52499999999999, 125.75, 147.575]

In [42]:

```
font_size = 13
sns.set_style(style='whitegrid') # {darkgrid, whitegrid, dark, white, ticks}

sns.boxplot(data)
plt.tight_layout() # Spreads a bit

plt.title("Boxplot of Monthly Mean Total Sunspot Number", fontdict={'fontsize':font_size})
plt.xlabel('Monthly Mean Total Sunspot Values', fontsize=font_size)
plt.xticks(fontsize=font_size) # xticks are the values of x-axis

plt.show()
```

