

Appendix I: Boxplots of numerical explanatory variables

```
In [2]: """Imports necessary packages"""

import itertools
import math
from typing import Dict, Iterable, List, Union

import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import pylab
import scipy
import scipy.stats as stats
import seaborn as sns
import statsmodels.api as sm
from sklearn.model_selection import train_test_split

sns.set_style("whitegrid")
```

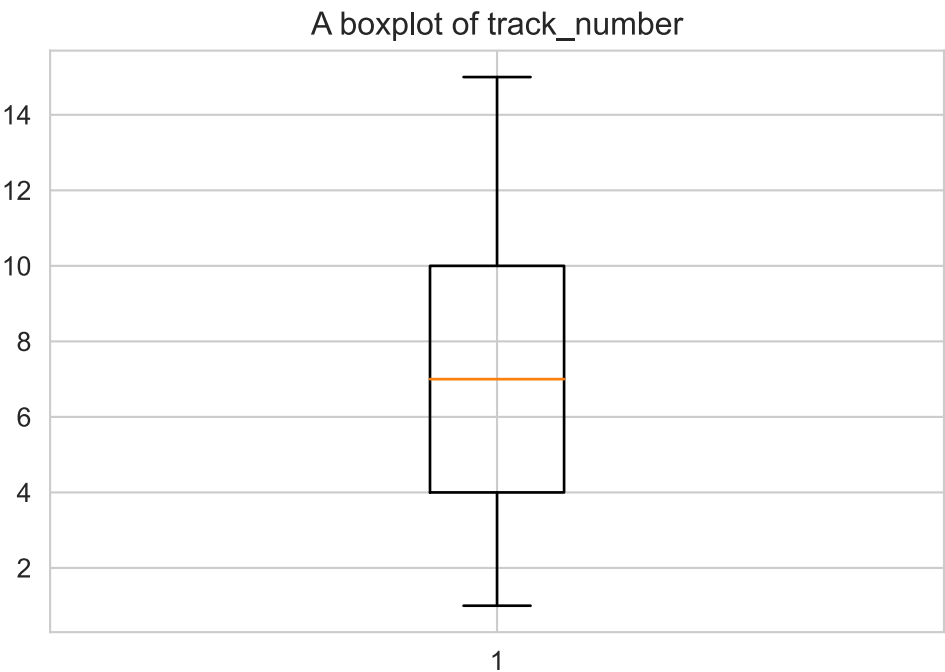
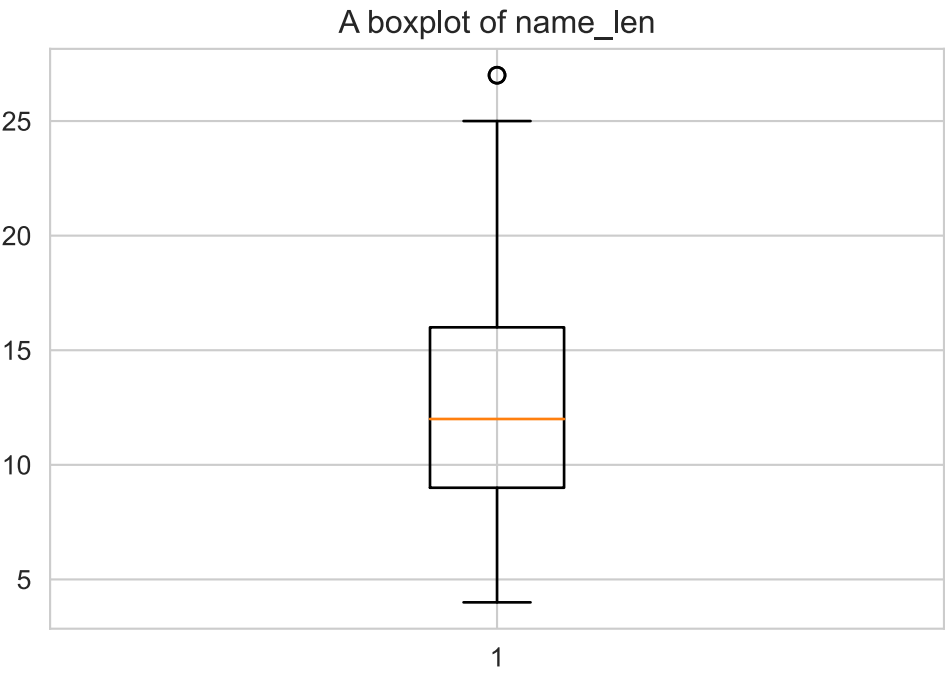
```
In [3]: def make_boxplot(data: Iterable, title: str = "A boxplot"):
        """Prints a box plot.

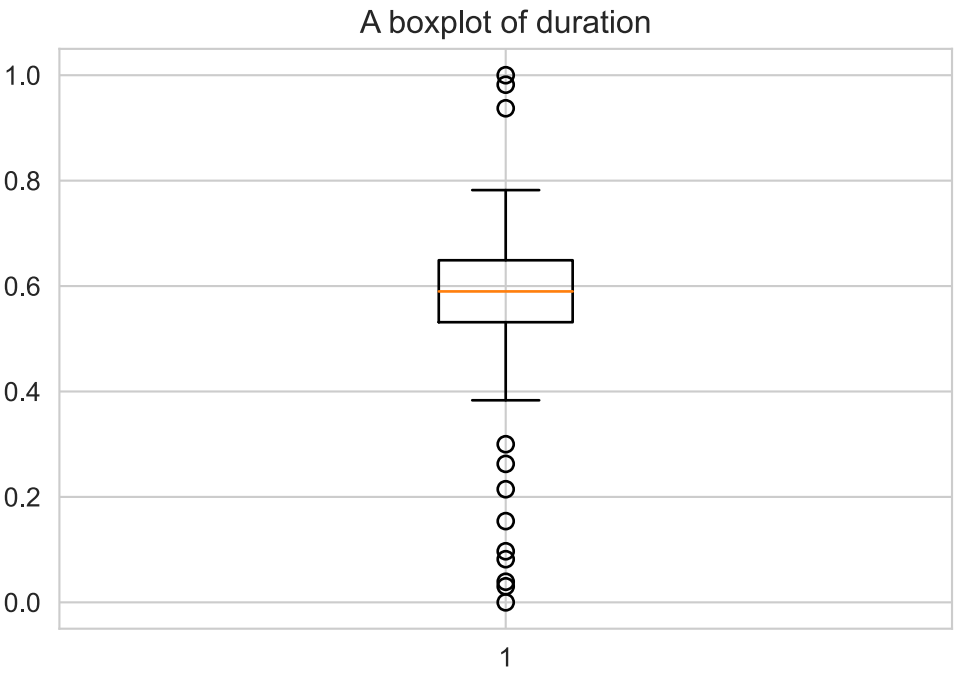
        Args:
            data (Iterable): the one dimensional data to plot.
            title (str, optional): the title of the plot. Defaults to "A boxplot".
        """
        plt.figure()
        plt.boxplot(data)
        plt.title(title)
        plt.show()
```

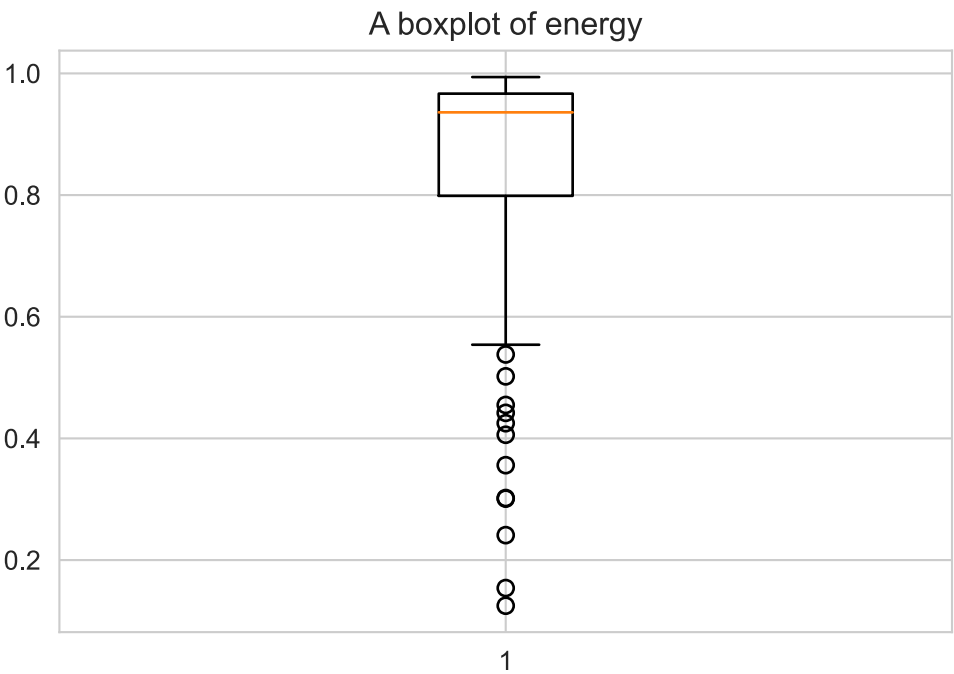
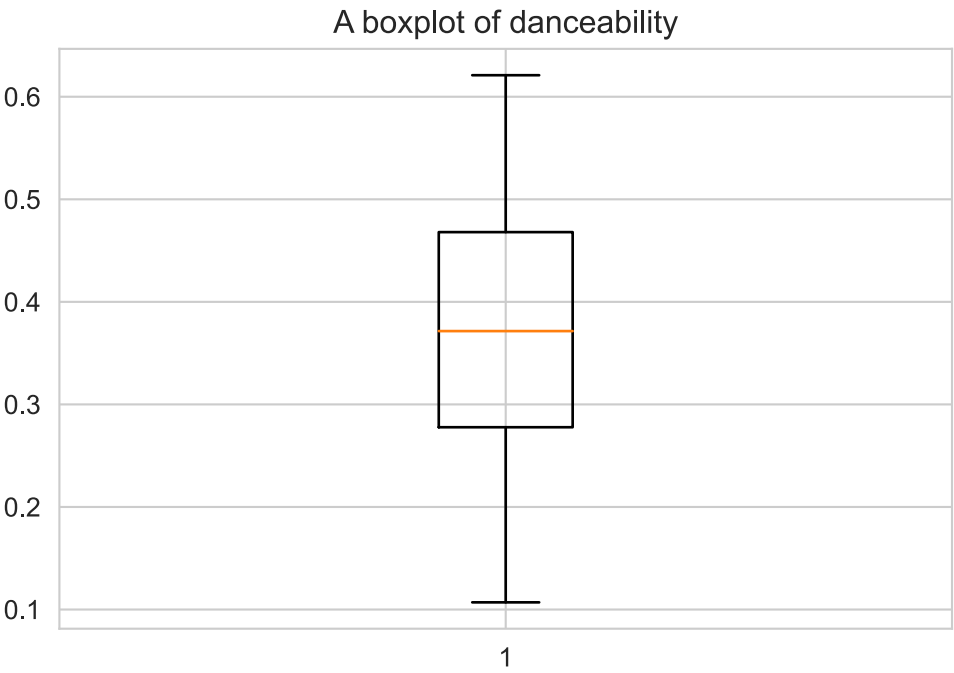
```
In [4]: data = pd.read_csv("D:/School/frequentist-statistics/ITM-song-popularity/database/it
data = data.drop("Unnamed: 0", axis=1)
```

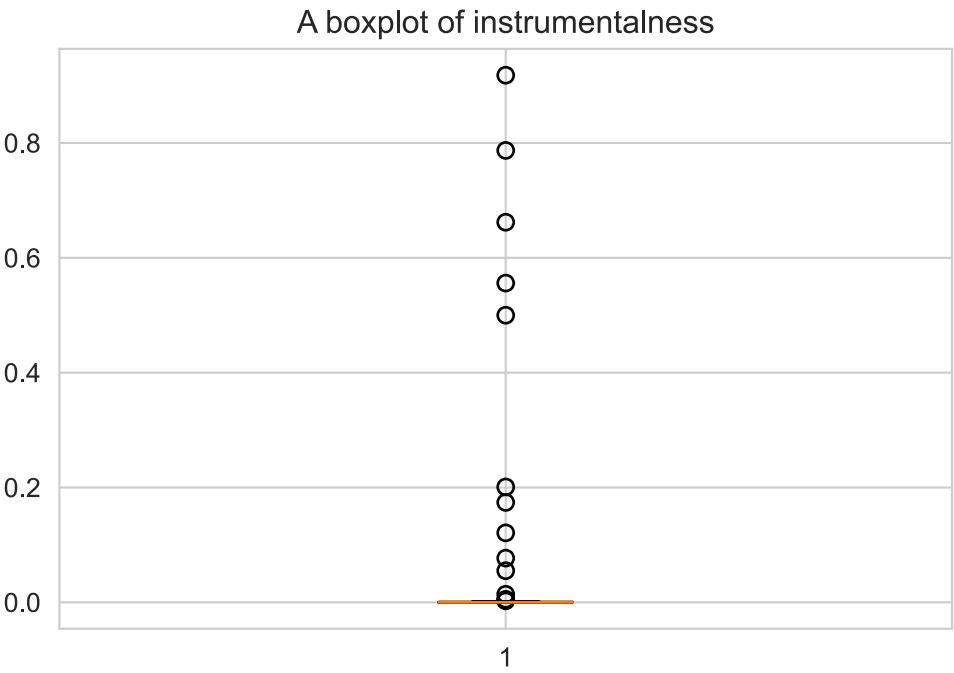
```
In [5]: numerical_variables = ["name_len", "track_number", "duration", "acousticness", "danc
categorical_variables = ["key", "mode", "time_signature", "explicit"]
```

```
In [6]: for variable in numerical_variables:
        make_boxplot(data[variable], title="A boxplot of %s" % variable)
```



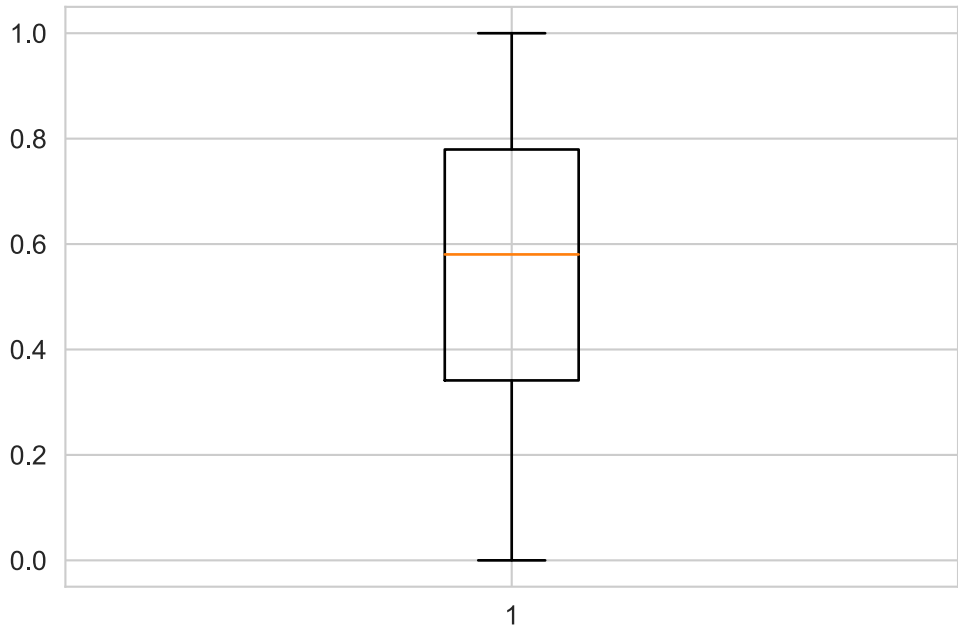








A boxplot of tempo



A boxplot of complexity

