Appendix III: Categorical explanatory variables vs numerical explanatory variables

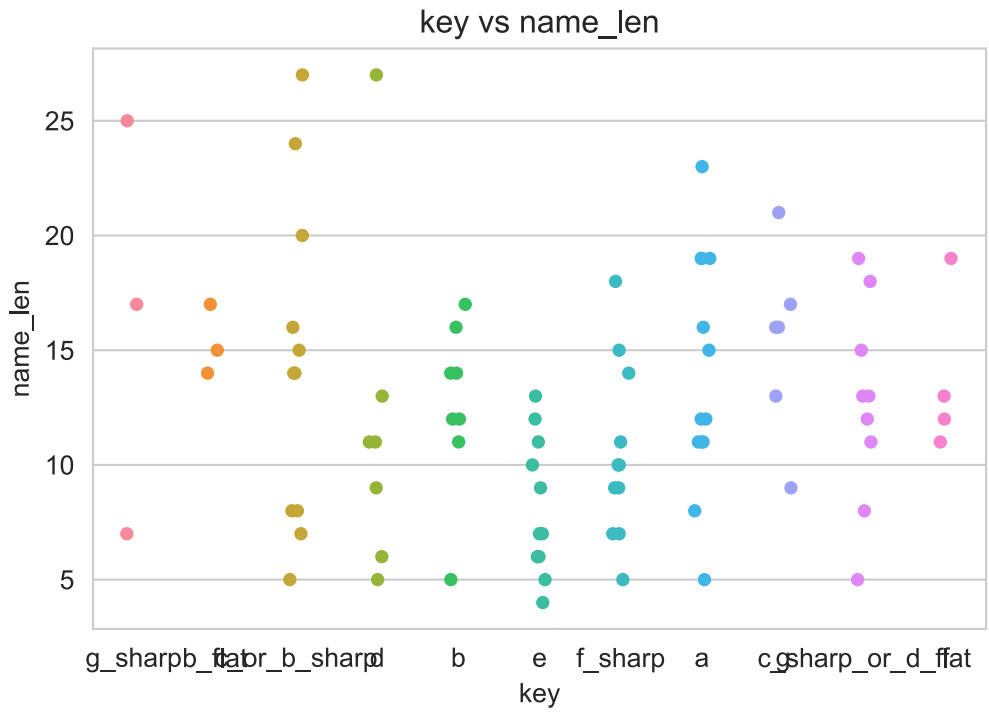
In [1]:
```
"""Imports necessary packages"""

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")
```

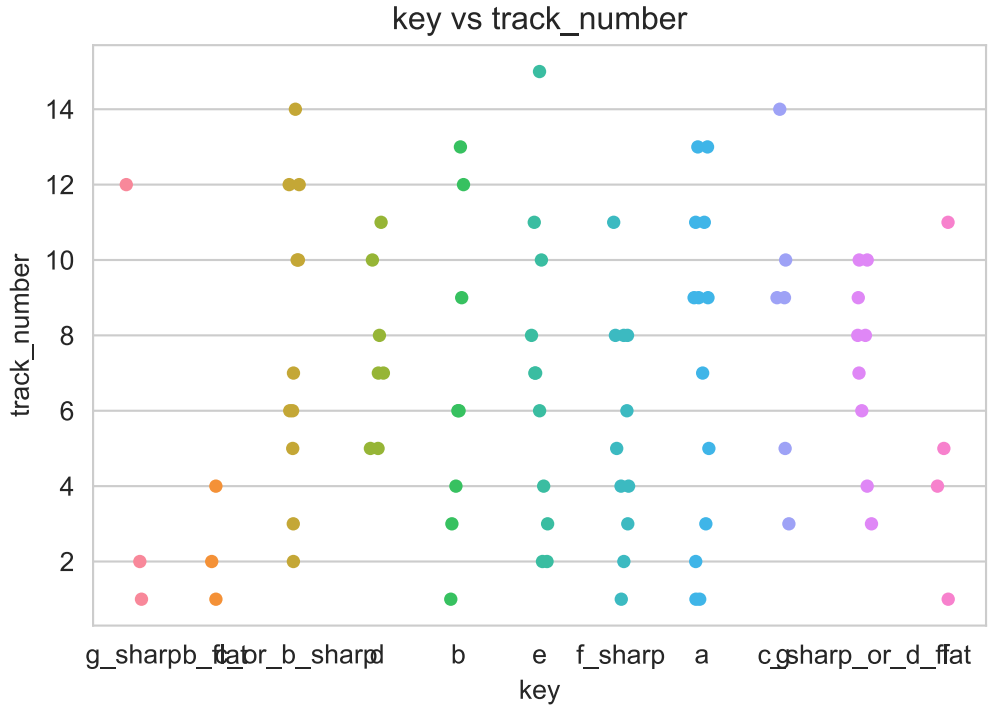
In [2]:
```
def make_stripplot(cat_var: str, num_var: str, data: Iterable, title:str="The distri
    """Prints a scatter plot of a categorical variable's values [x axis] against a n

    Args:
        cat_var (str): the name of the categorical variable.
        num_var (str): the name of the numerical variable.
        data (Iterable): the two dimensional data which includes the categorical and
        title (str, optional): the title of the plot. Defaults to "The distribution"
    """
    plt.figure()
    sns.stripplot(x=cat_var, y=num_var, data=data)
    plt.title(title)
    plt.show()
```

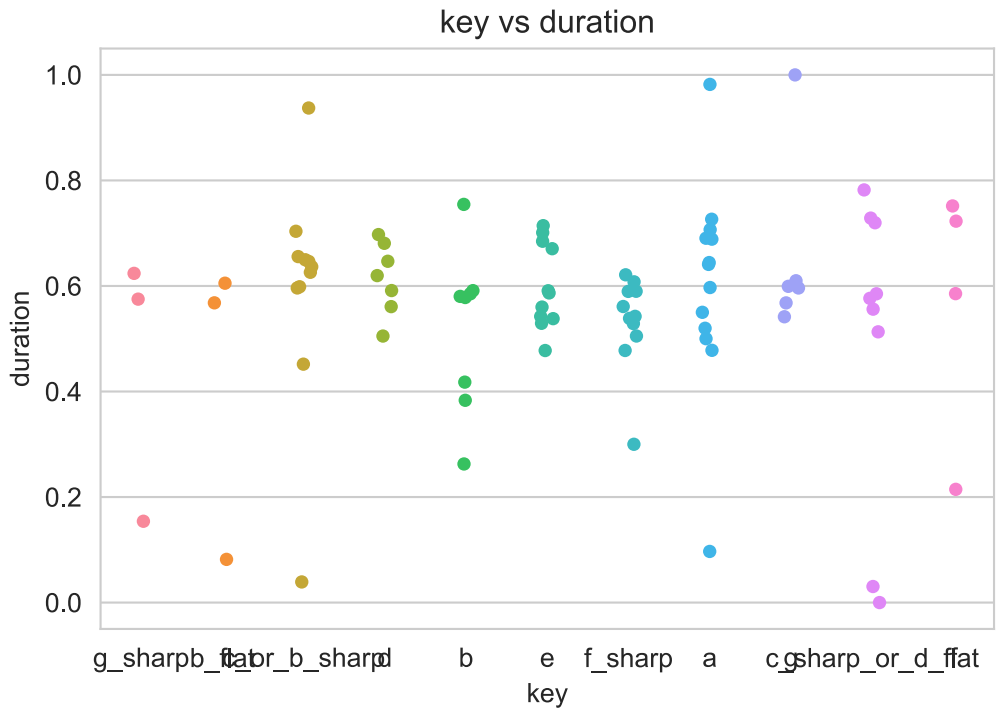
In [3]:
```
data = pd.read_csv("D:/School/frequentist-statistics/ITM-song-popularity/database/it
data = data.drop("Unnamed: 0", axis=1)
```

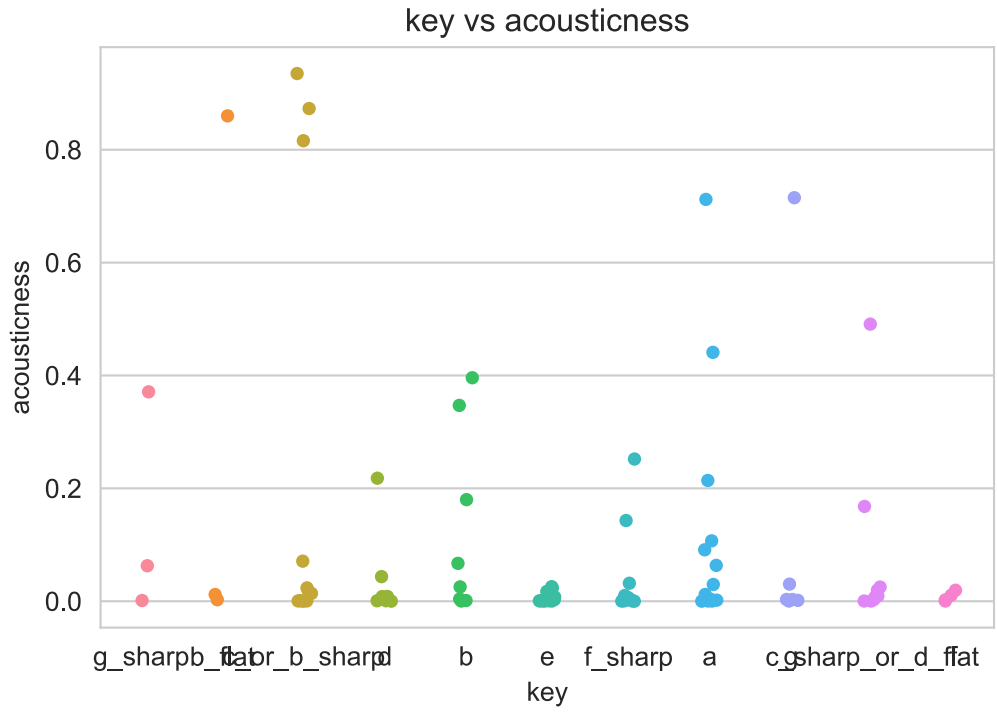
In [4]:
```
numerical_variables = ["name_len", "track_number", "duration", "acousticness", "danc
categorical_variables = ["key", "mode", "time_signature", "explicit"]
```

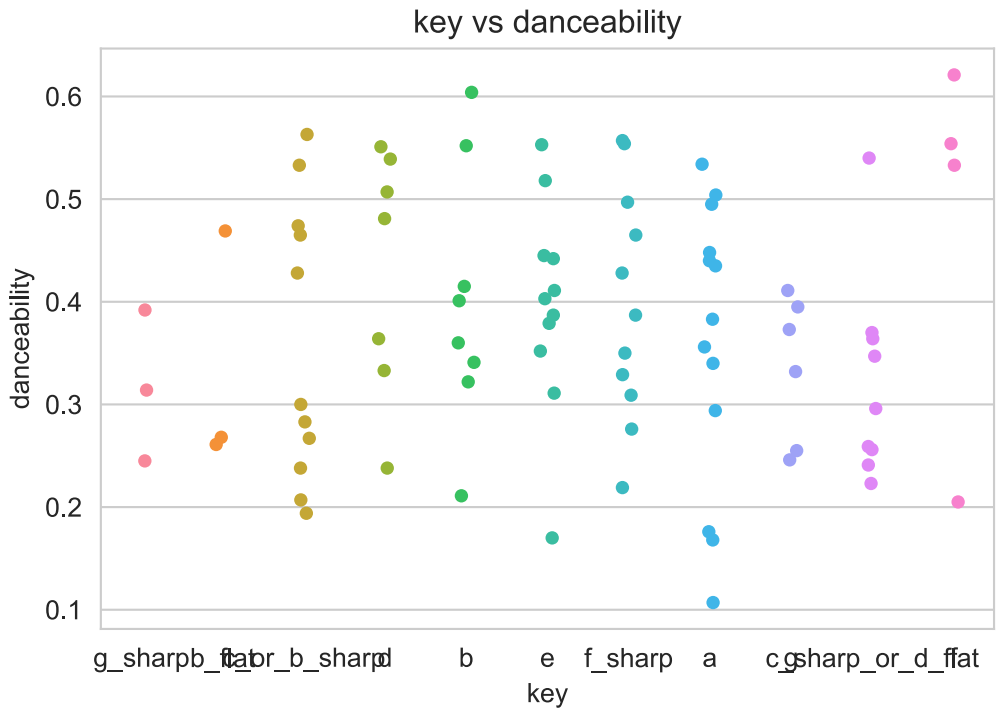
In [5]:
```
for x in categorical_variables:
    for y in numerical_variables:
        make_stripplot(x, y, data, title="%s vs %s" % (x, y))
```

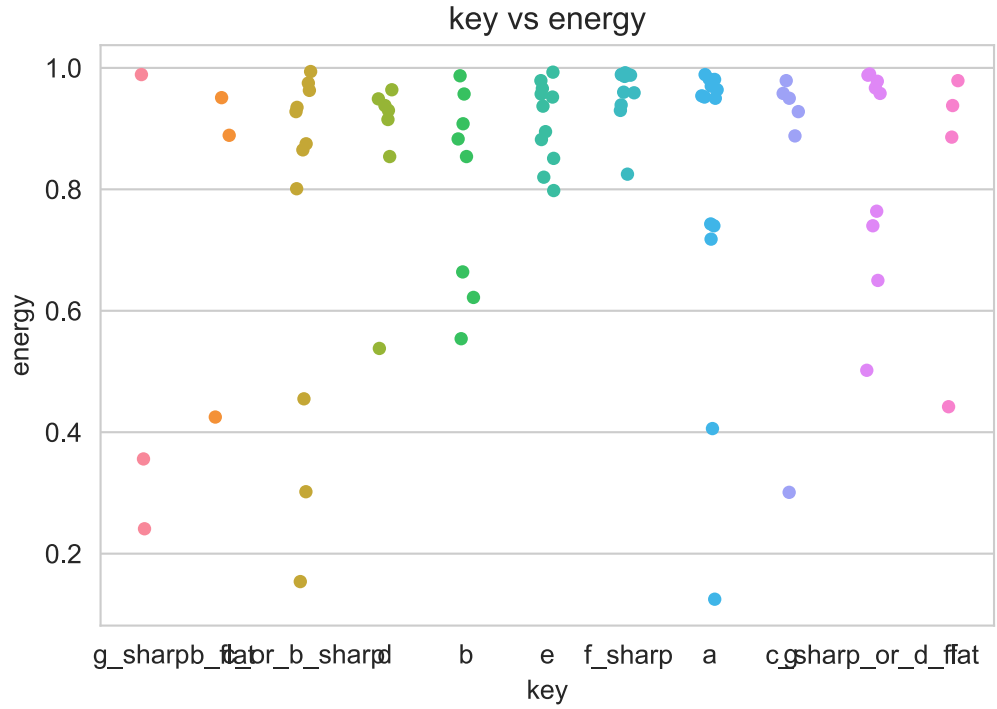

key vs name_len
name_len
25
20
15
10
5
g_sharp
b_flat
a_flat
or_b_sharp
d
b
e
f_sharp
a
c
g_sharp_or_d_flat
key

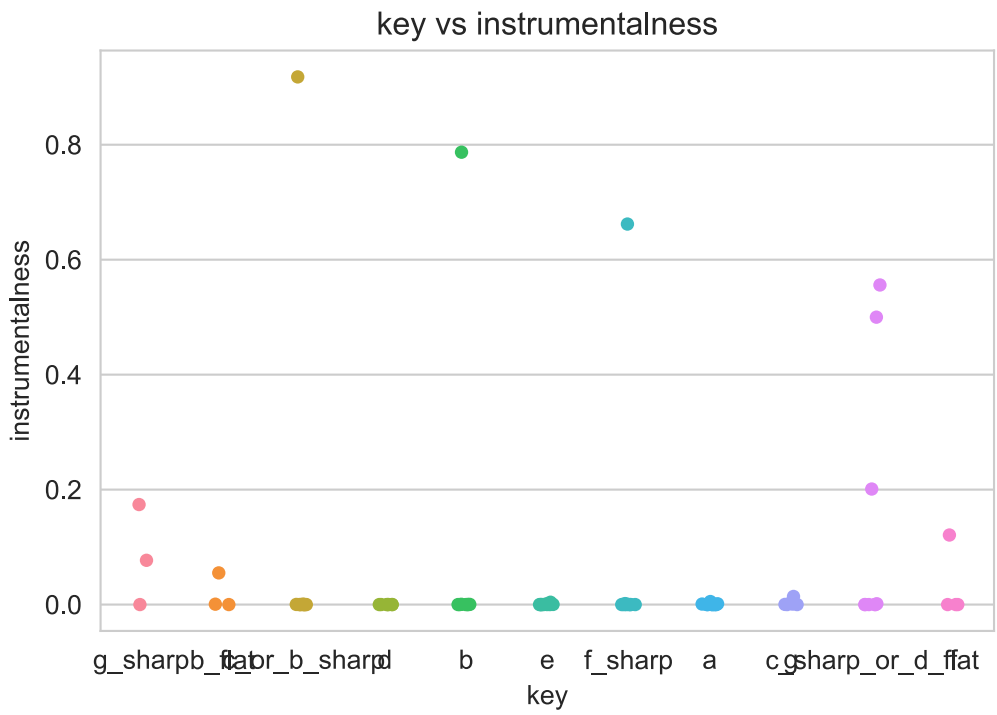

key vs track_number
track_number
14
12
10
8
6
4
2
g_sharp
b_flat
a_flat
or_b_sharp
d
b
e
f_sharp
a
c
g_sharp_or_d_flat
key

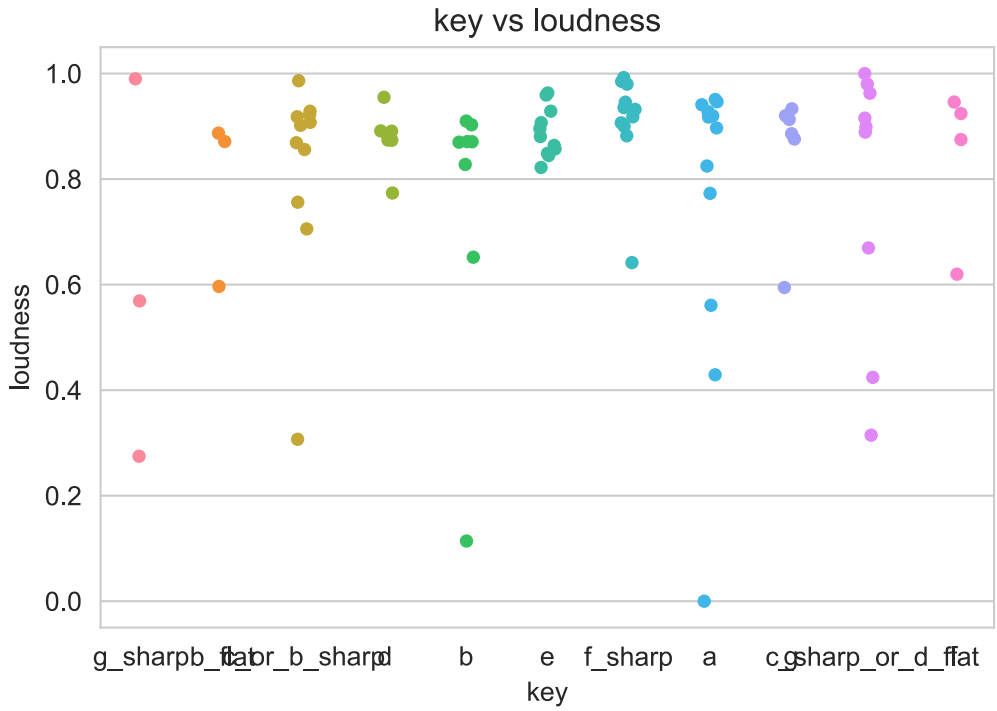

key vs duration
duration
1.0
0.8
0.6
0.4
0.2
0.0
g_sharp
b_flat
or_b_sharp
d
b
e
f_sharp
a
c
g
sharp_or_d_flat
key

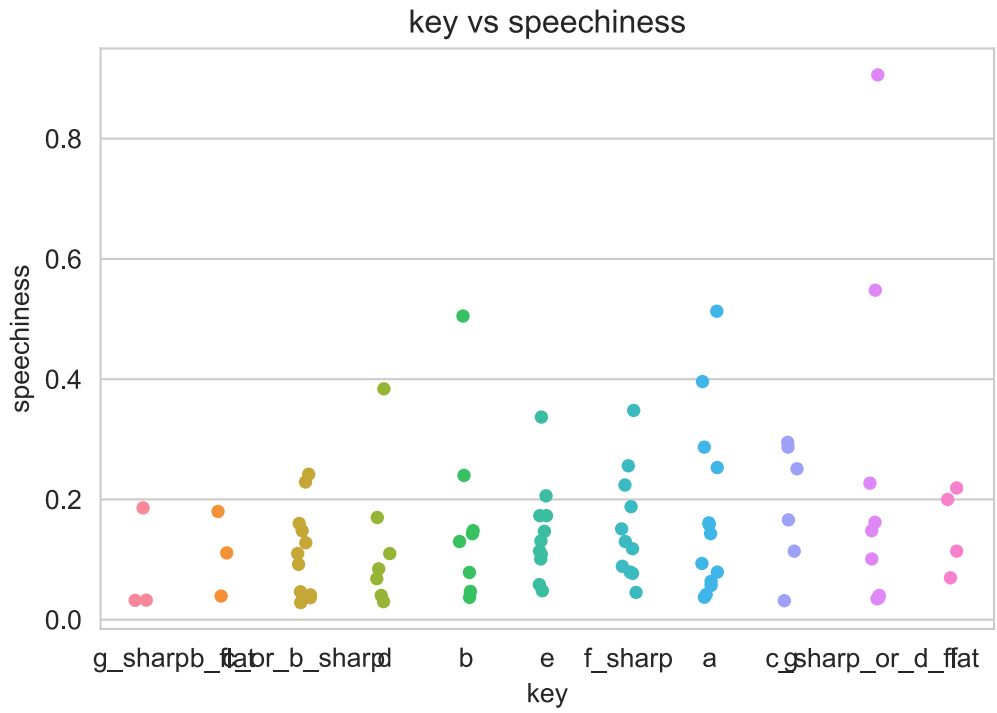

key vs acousticness
acousticness
0.8
0.6
0.4
0.2
0.0
g_sharp
b_flat
or_b_sharp
d
b
e
f_sharp
a
c
g
sharp_or_d_flat
key

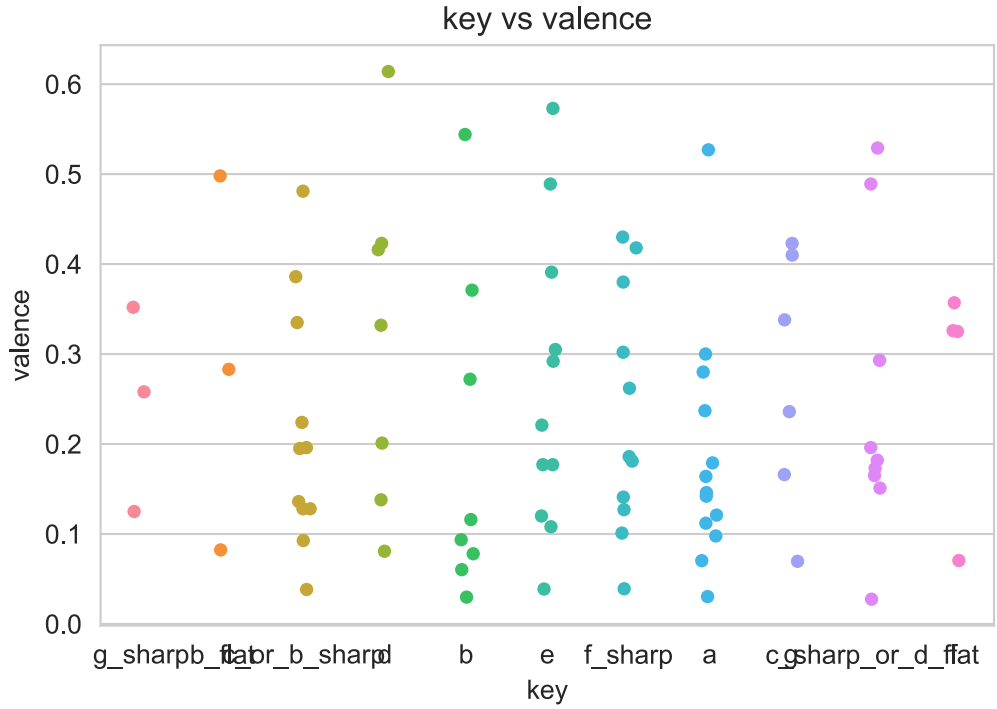

key vs danceability
danceability
0.6
0.5
0.4
0.3
0.2
0.1
g_sharp
b_flat_or_a_sharp
d
b
e
f_sharp
a
c_sharp_or_d_flat
key

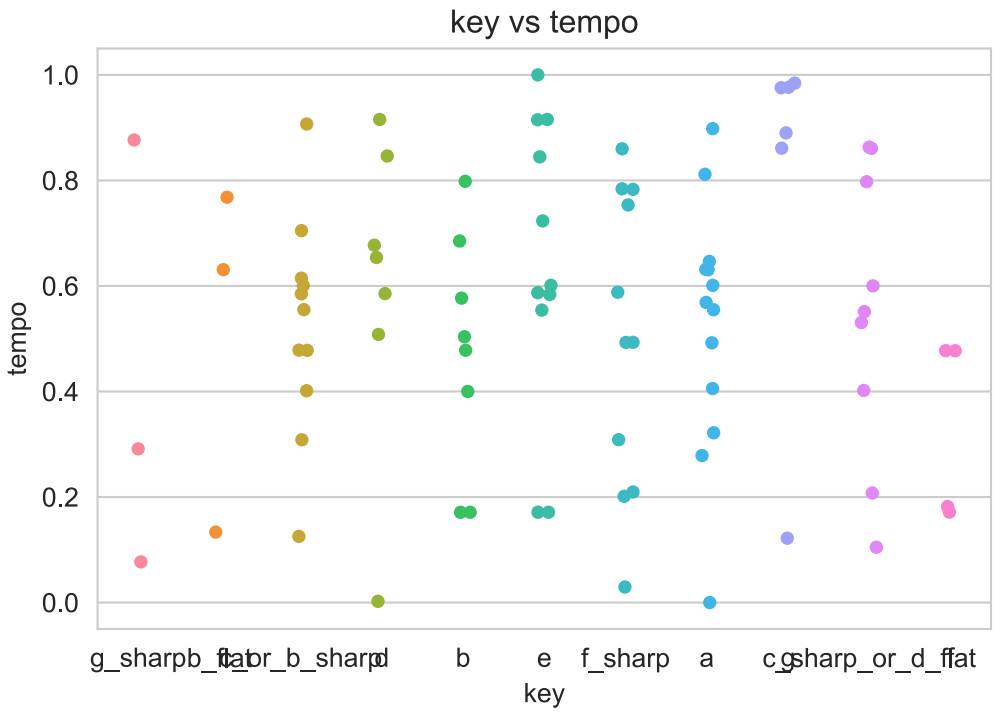

key vs energy
energy
1.0
0.8
0.6
0.4
0.2
g_sharp
b_flat_or_a_sharp
d
b
e
f_sharp
a
c_sharp_or_d_flat
key

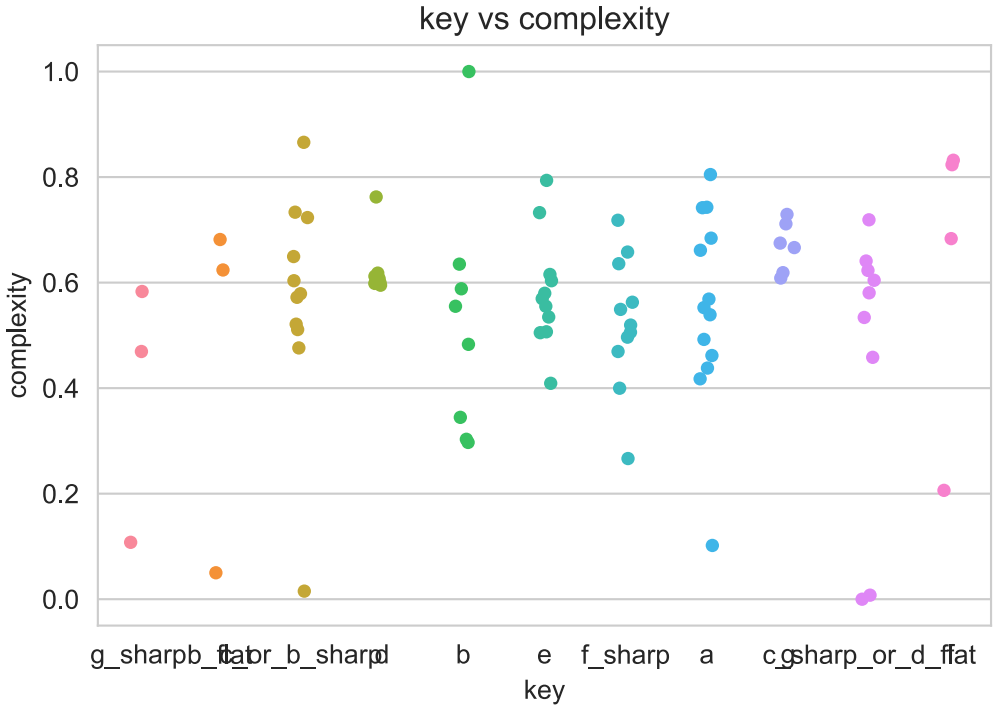

key vs instrumentalness
instrumentalness
0.0
0.2
0.4
0.6
0.8
g_sharp
b_flat_or_b_sharp
d
b
e
f_sharp
a
c_g_sharp_or_d_flat
key

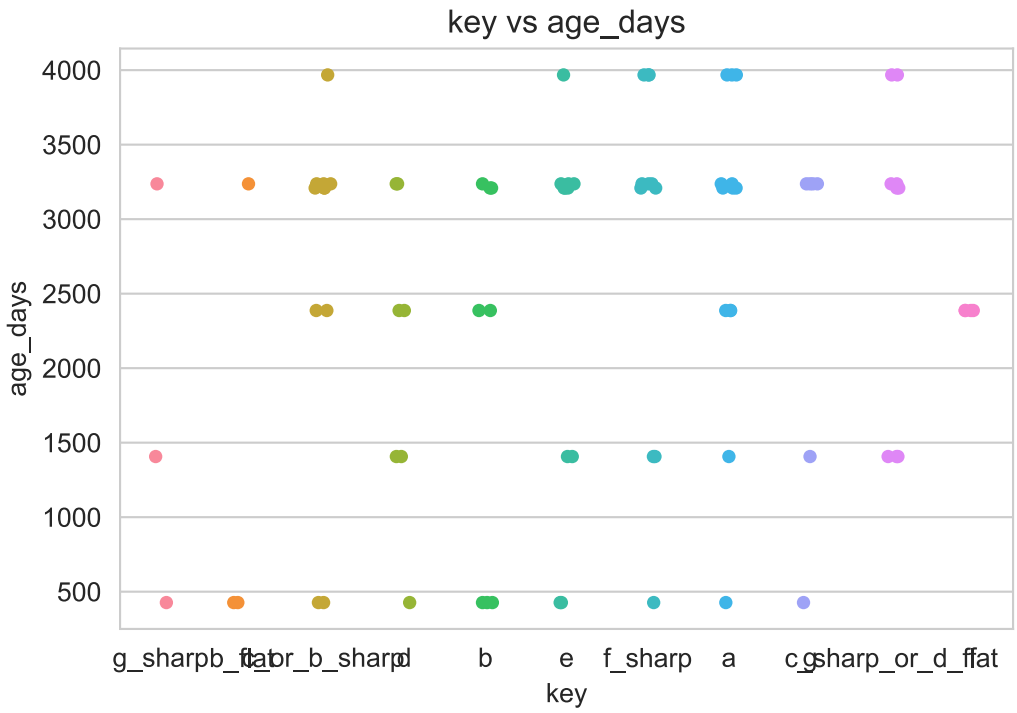

key vs loudness
loudness
0.0
0.2
0.4
0.6
0.8
1.0
g_sharp
b_flat_or_b_sharp
d
b
e
f_sharp
a
c_g_sharp_or_d_flat
key

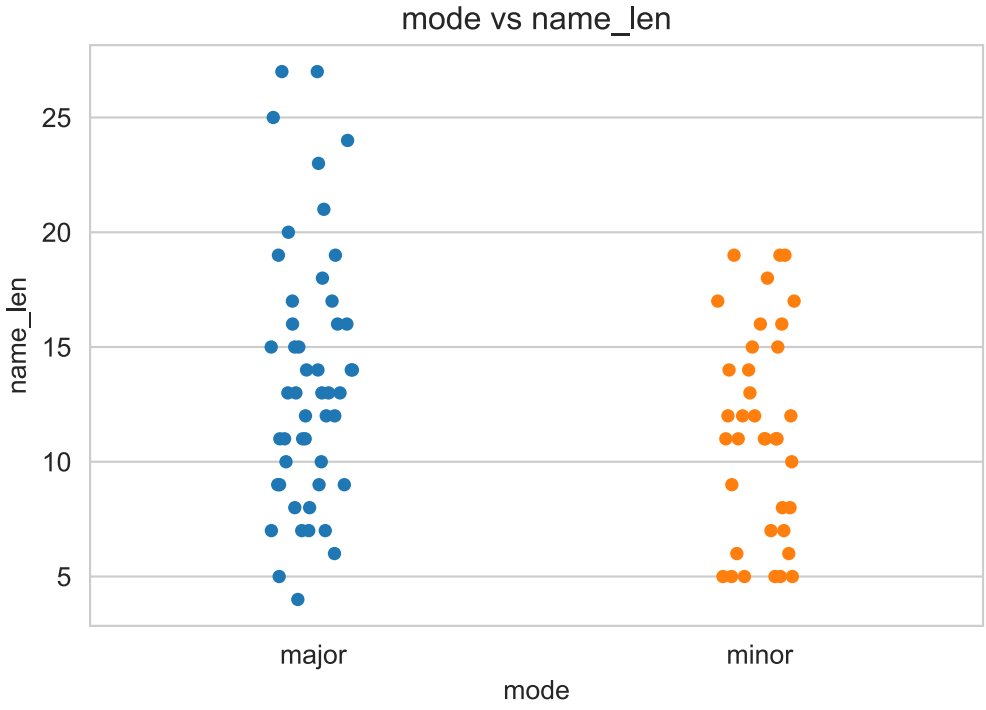

key vs speechiness
speechiness
0.0
0.2
0.4
0.6
0.8
g_sharp
b_flat_or_a_sharp
d
b
e
f_sharp
a
c_sharp_or_d_flat
key

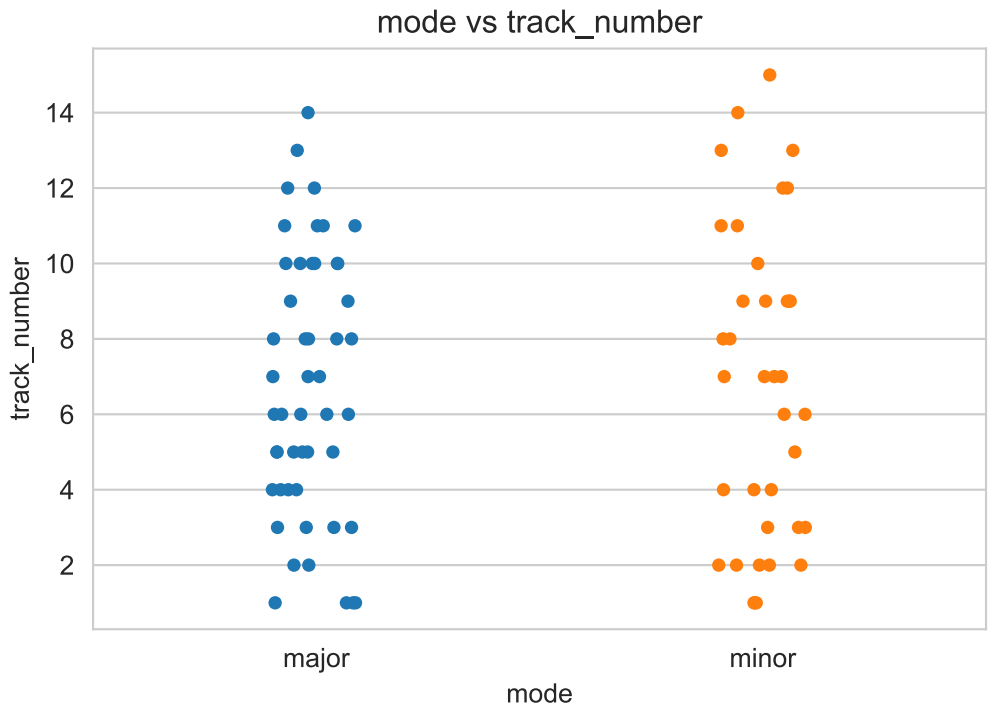

key vs valence
valence
0.0
0.1
0.2
0.3
0.4
0.5
0.6
g_sharp
b_flat_or_a_sharp
d
b
e
f_sharp
a
c_sharp_or_d_flat
key

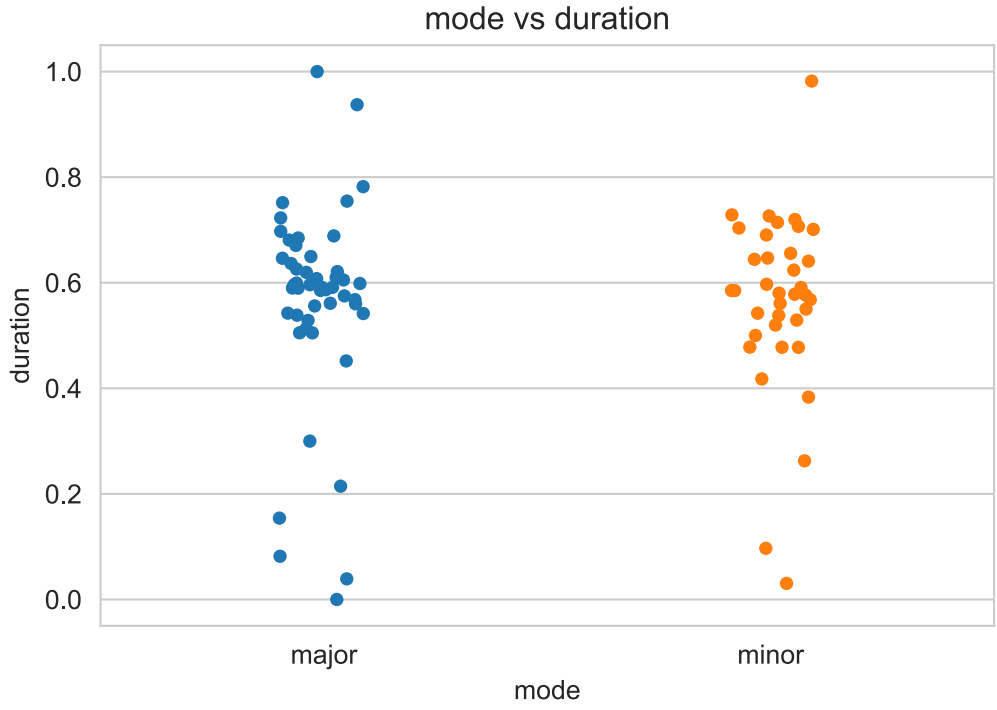

key vs tempo
tempo
1.0
0.8
0.6
0.4
0.2
0.0
g_sharp
b_flat_or_a_sharp
d
b
e
f_sharp
a
c_sharp_or_d_flat
g
key

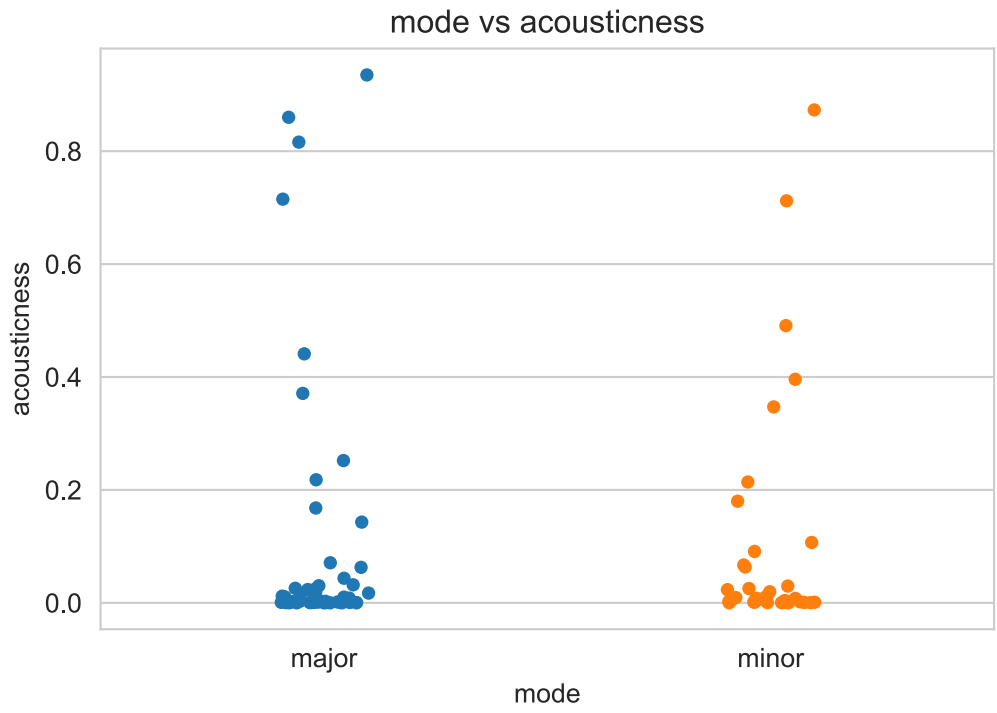

key vs complexity
complexity
1.0
0.8
0.6
0.4
0.2
0.0
g_sharp
b_flat_or_a_sharp
d
b
e
f_sharp
a
c_sharp_or_d_flat
g
key

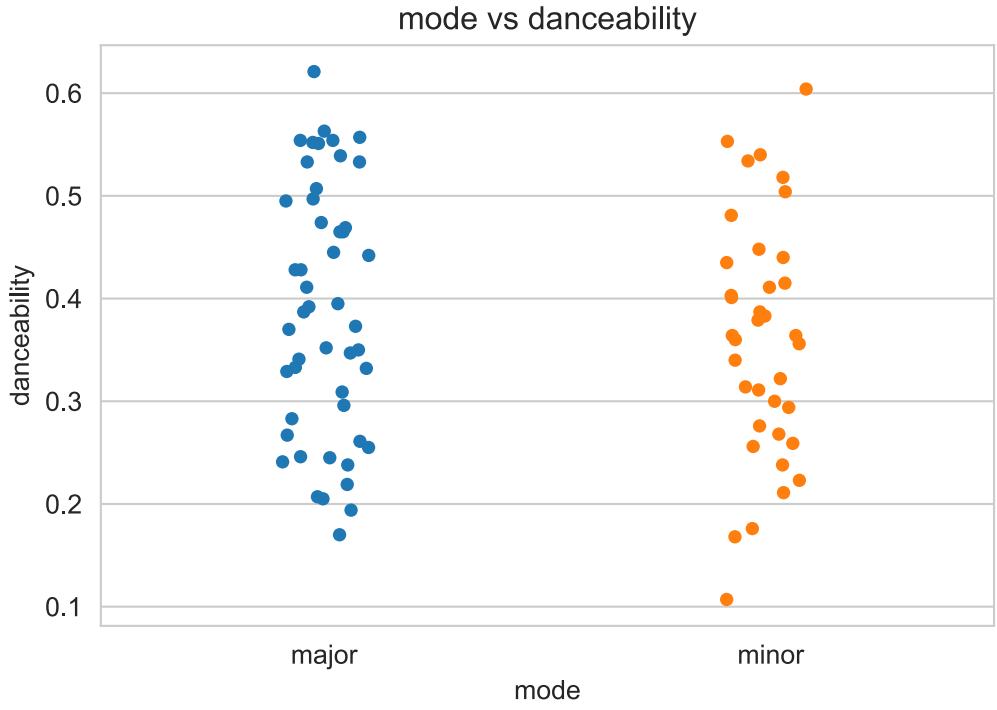

key vs age_days
4000
3500
3000
2500
2000
1500
1000
500
age_days
g_sharp
b_flat
a_flat
b_sharp
d
b
e
f_sharp
a
c
g_sharp
d_flat
key

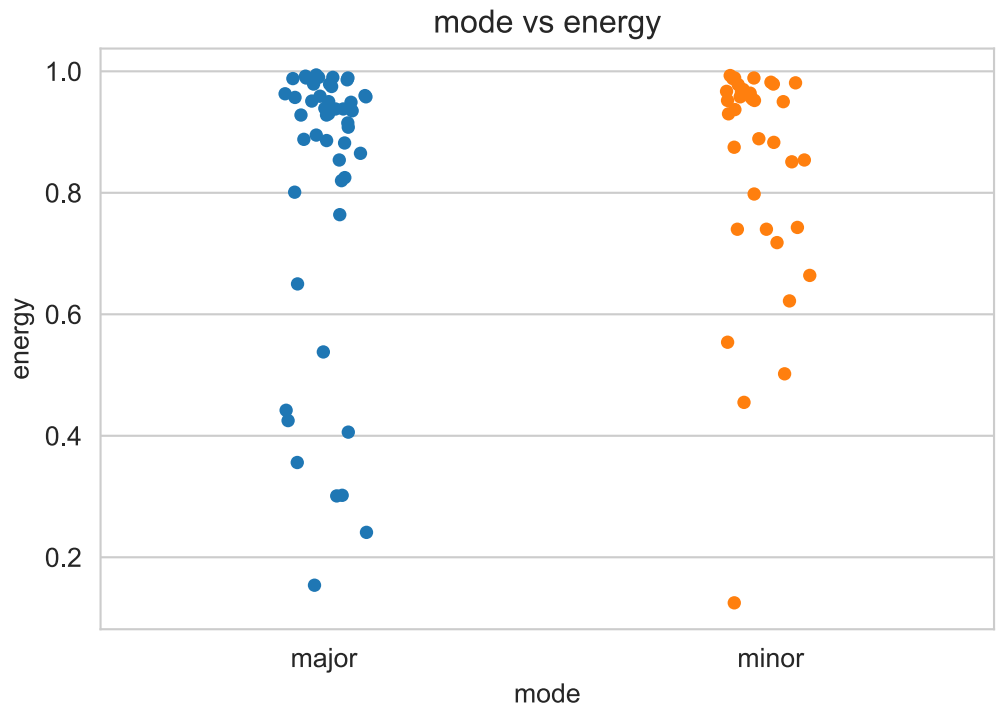

mode vs name_len
25
20
15
10
5
name_len
major
minor
mode

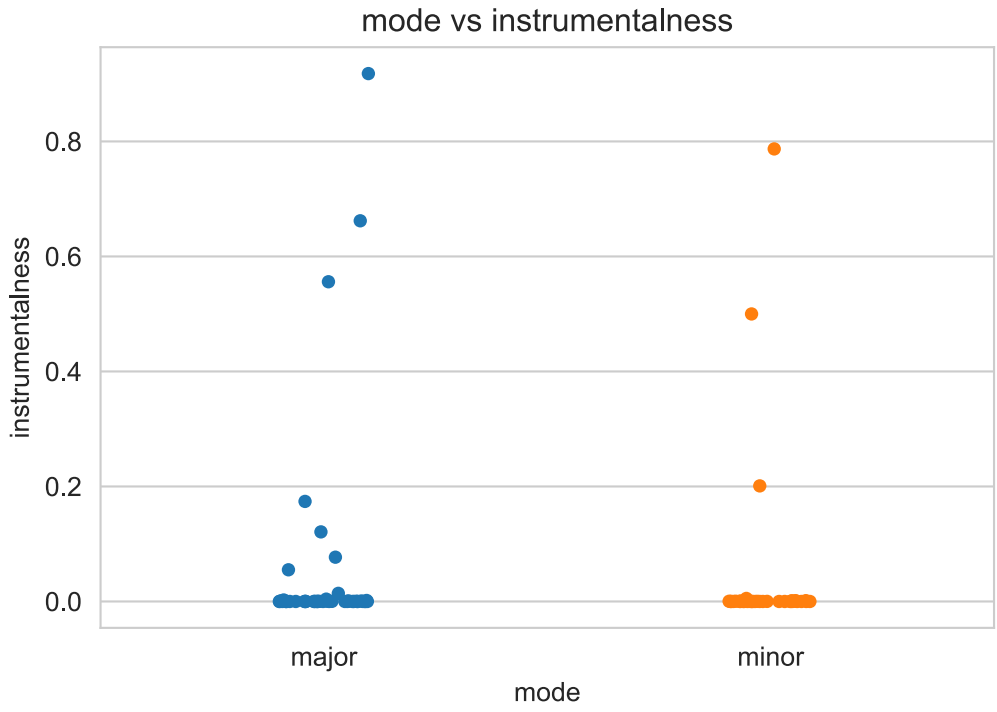

mode vs track_number
track_number
14
12
10
8
6
4
2
major
minor
mode

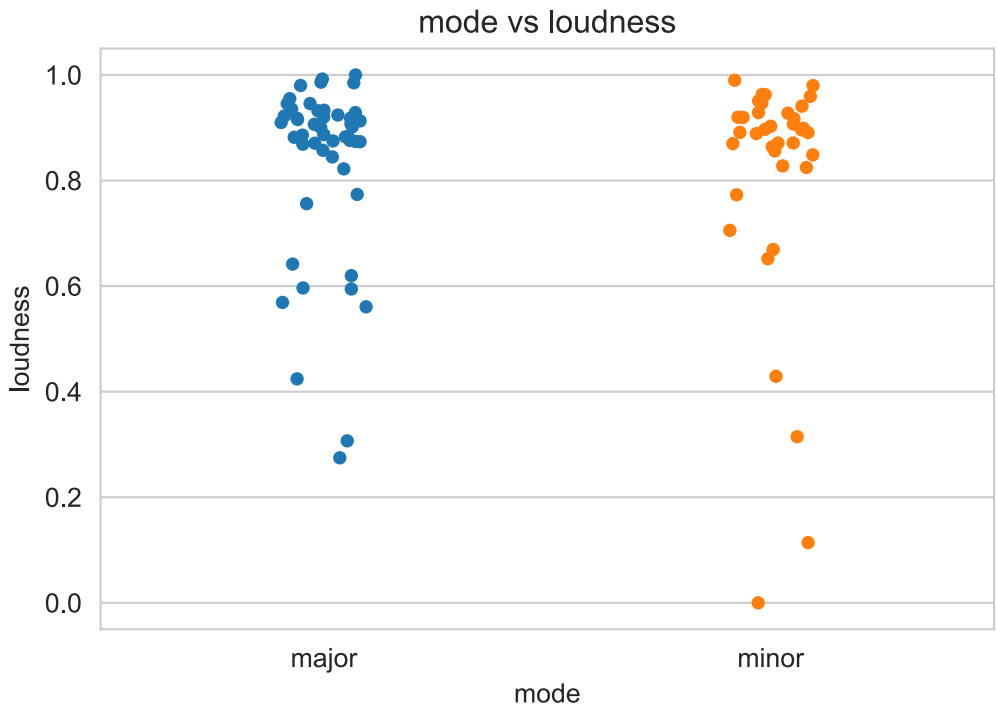

mode vs duration
duration
1.0
0.8
0.6
0.4
0.2
0.0
major
minor
mode

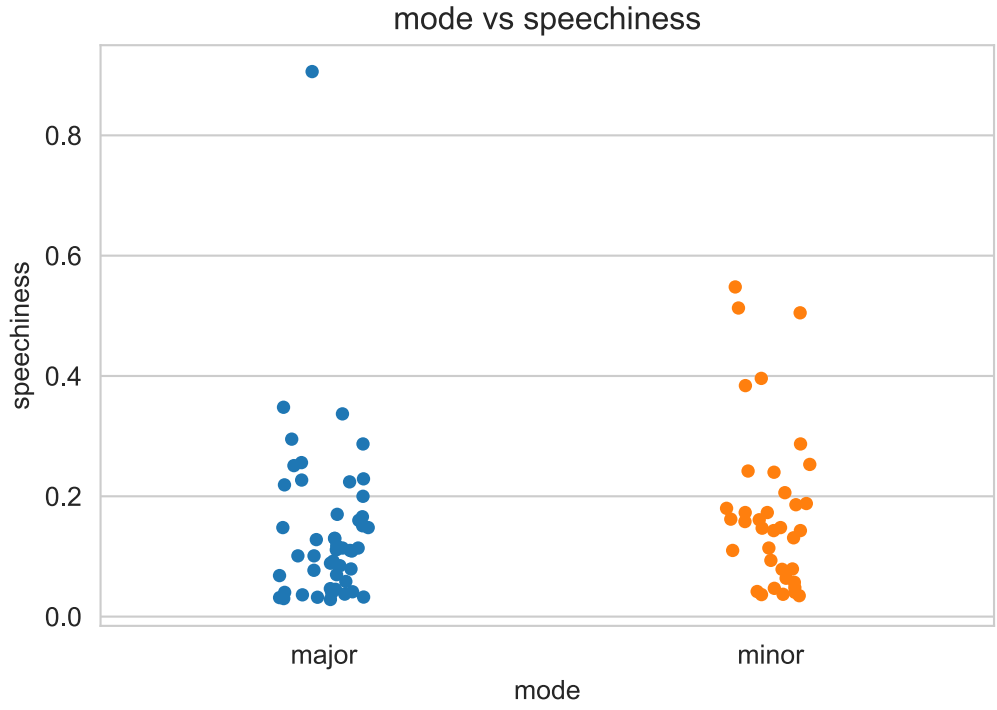

mode vs acousticness
acousticness
0.0
0.2
0.4
0.6
0.8
major
minor
mode

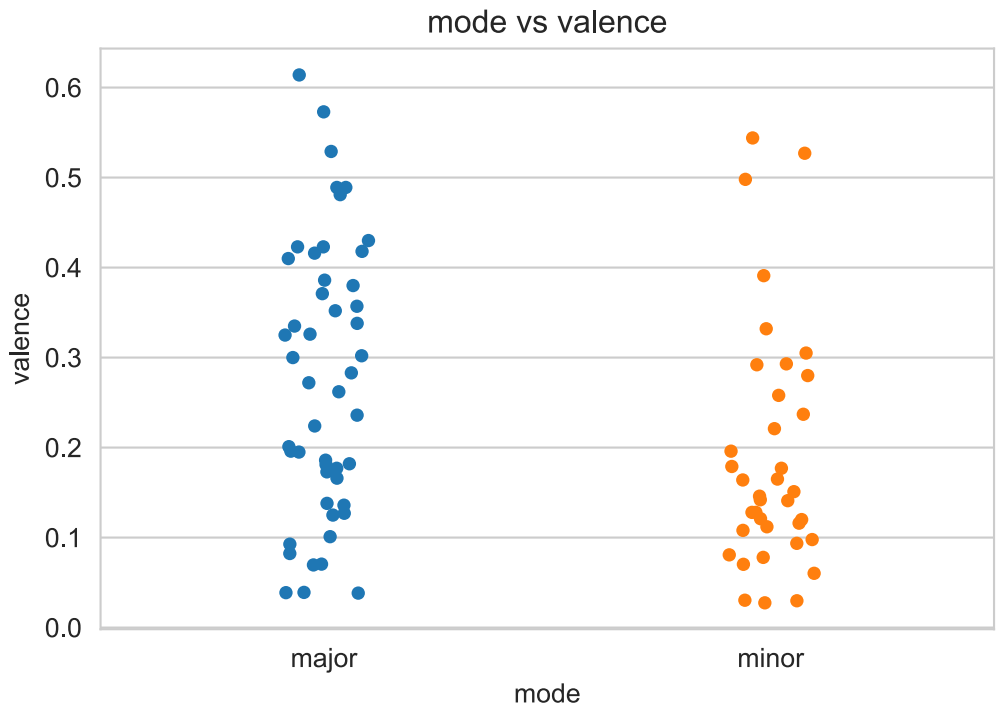

mode vs danceability
danceability
0.1
0.2
0.3
0.4
0.5
0.6
major
minor
mode

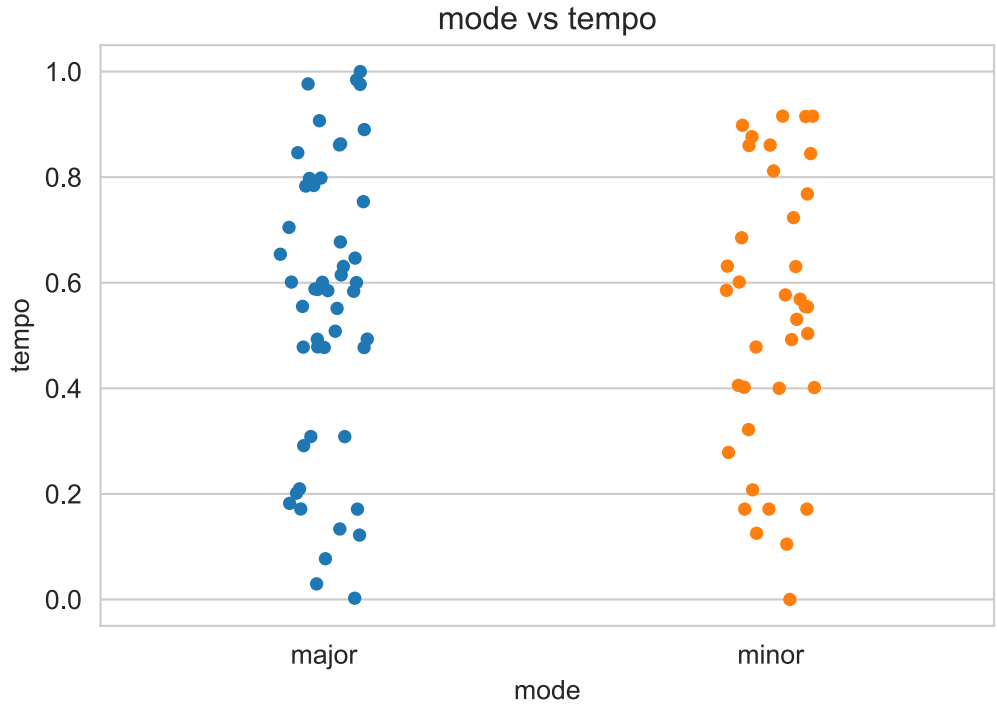

mode vs energy
1.0
0.8
0.6
0.4
0.2
energy
major
minor
mode

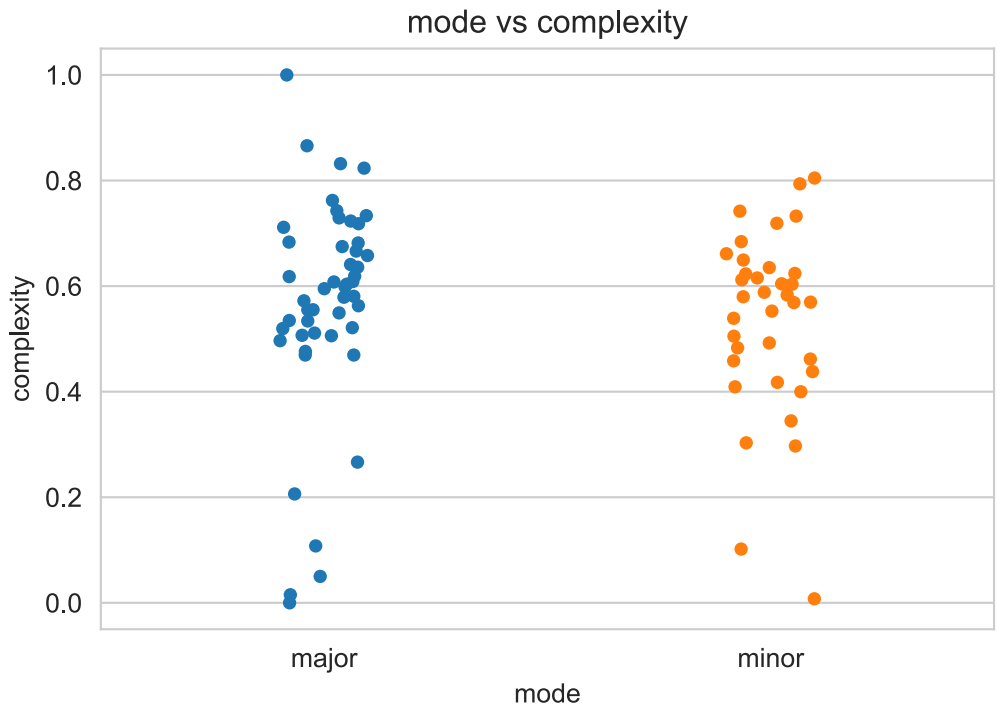

mode vs instrumentalness
0.8
0.6
0.4
0.2
0.0
instrumentalness
major
minor
mode

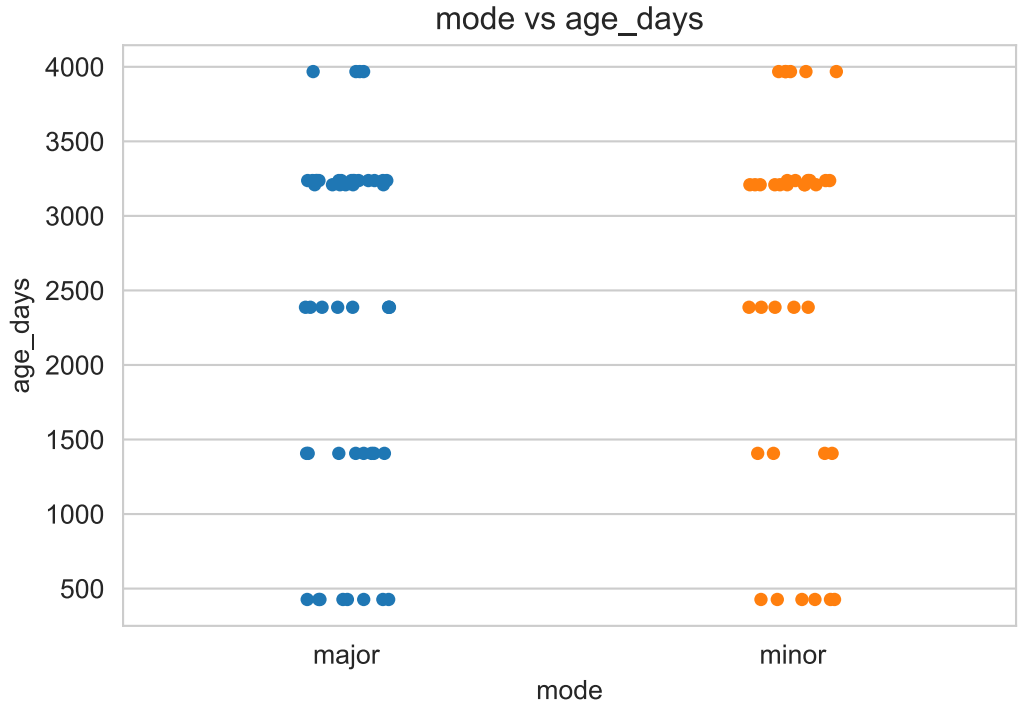

mode vs loudness
loudness
1.0
0.8
0.6
0.4
0.2
0.0
major
minor
mode

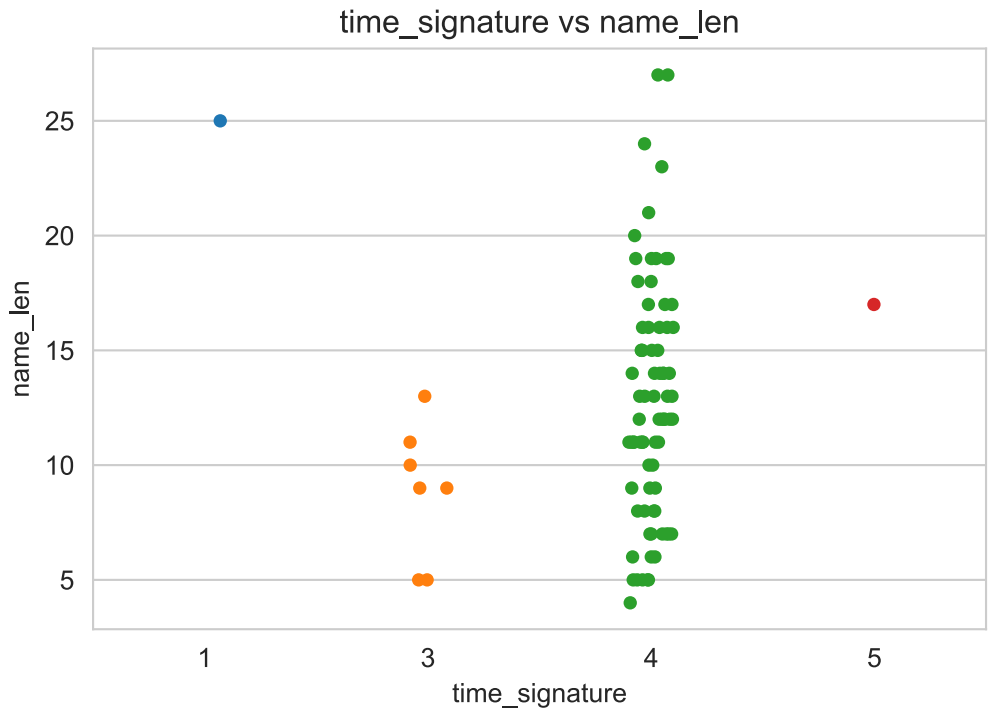

mode vs speechiness
speechiness
0.8
0.6
0.4
0.2
0.0
major
minor
mode

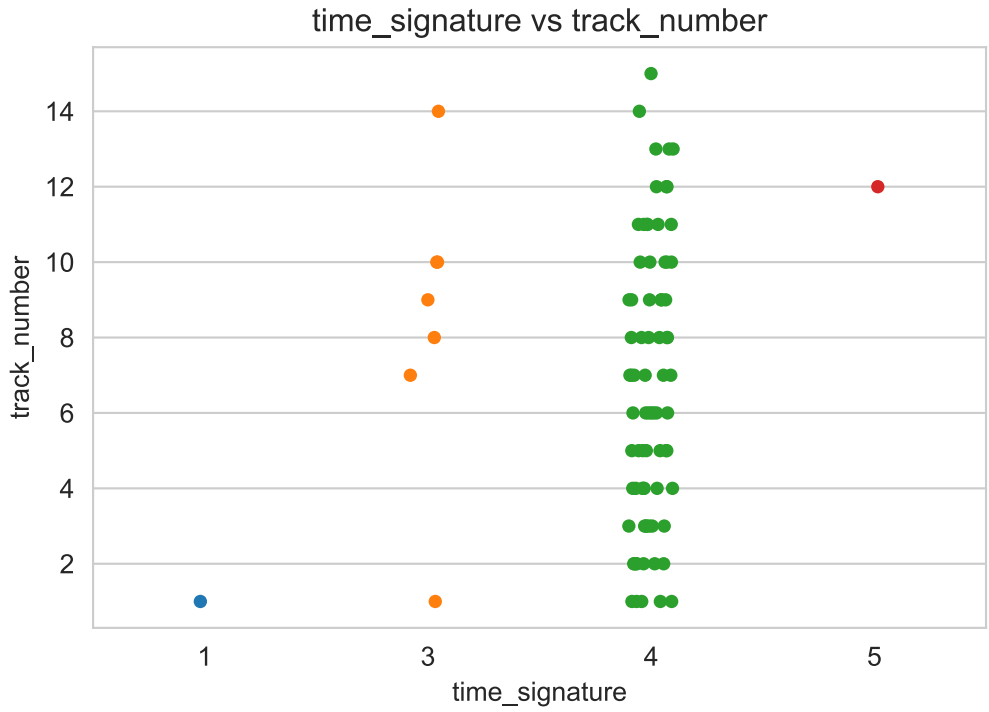

mode vs valence
valence
0.0
0.1
0.2
0.3
0.4
0.5
0.6
major
minor
mode

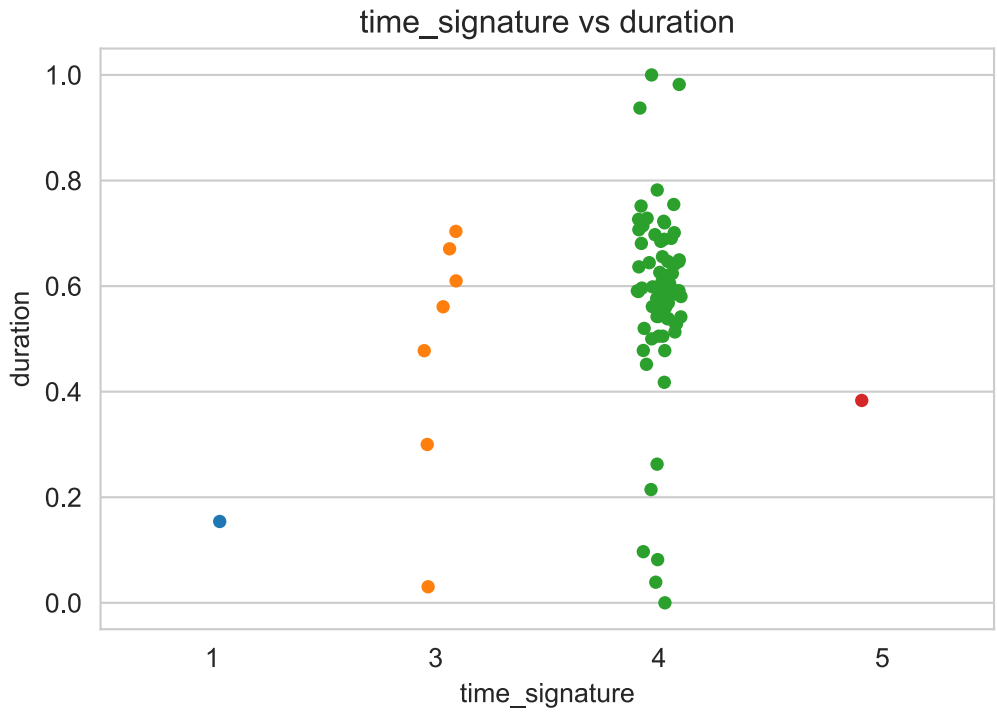

mode vs tempo
tempo
0.0
0.2
0.4
0.6
0.8
1.0
major
minor
mode

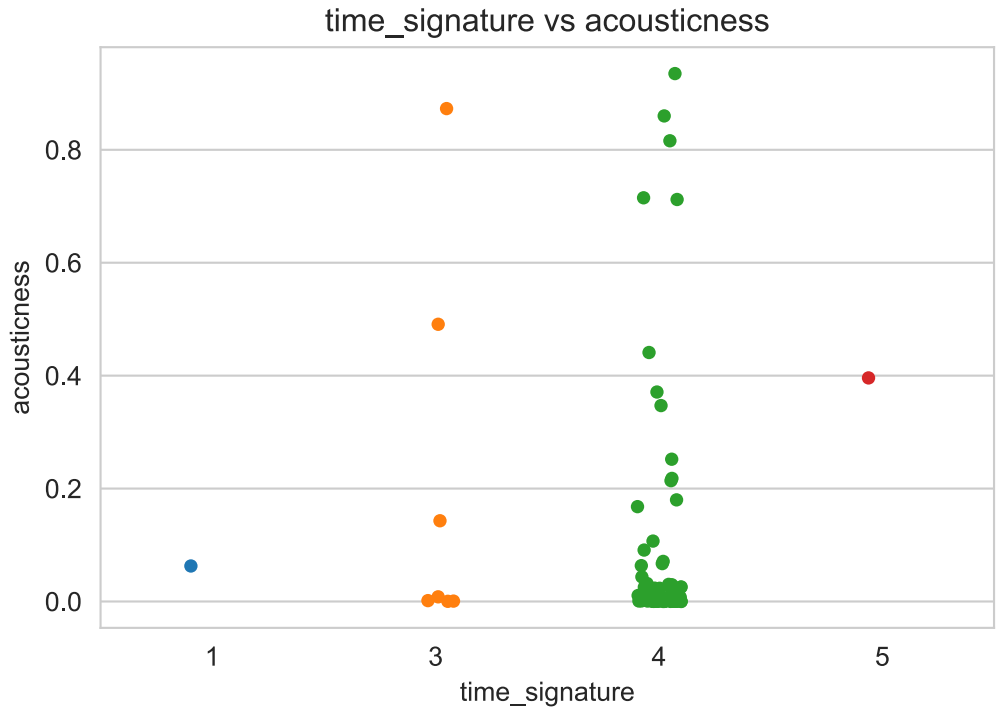

mode vs complexity
1.0
0.8
0.6
0.4
0.2
0.0
complexity
major
minor
mode

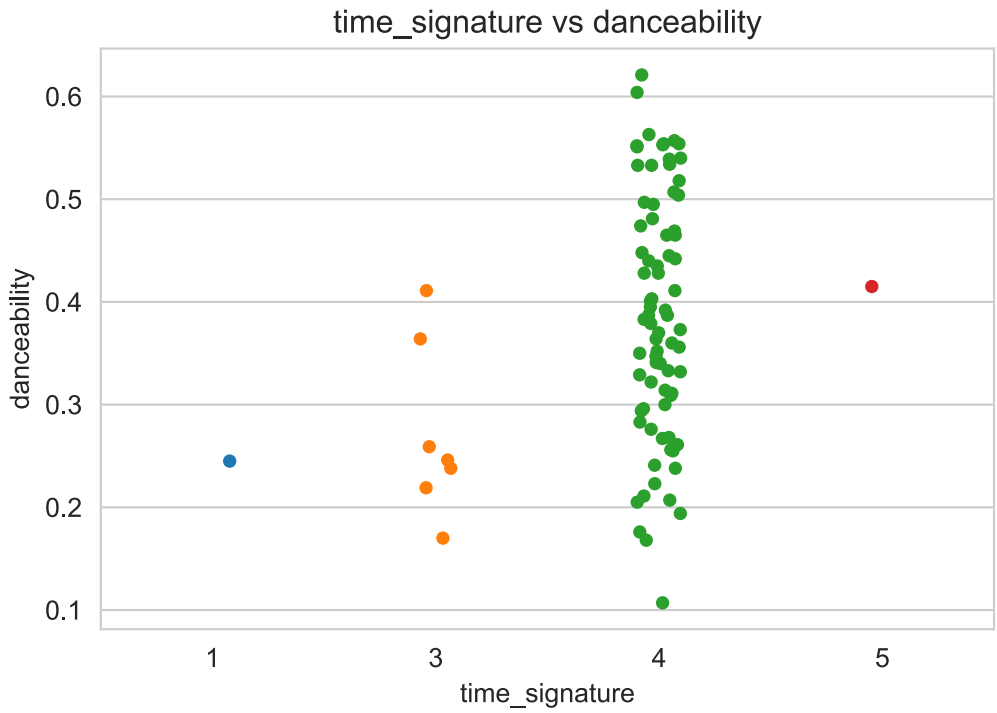

mode vs age_days
4000
3500
3000
2500
2000
1500
1000
500
age_days
major
minor
mode

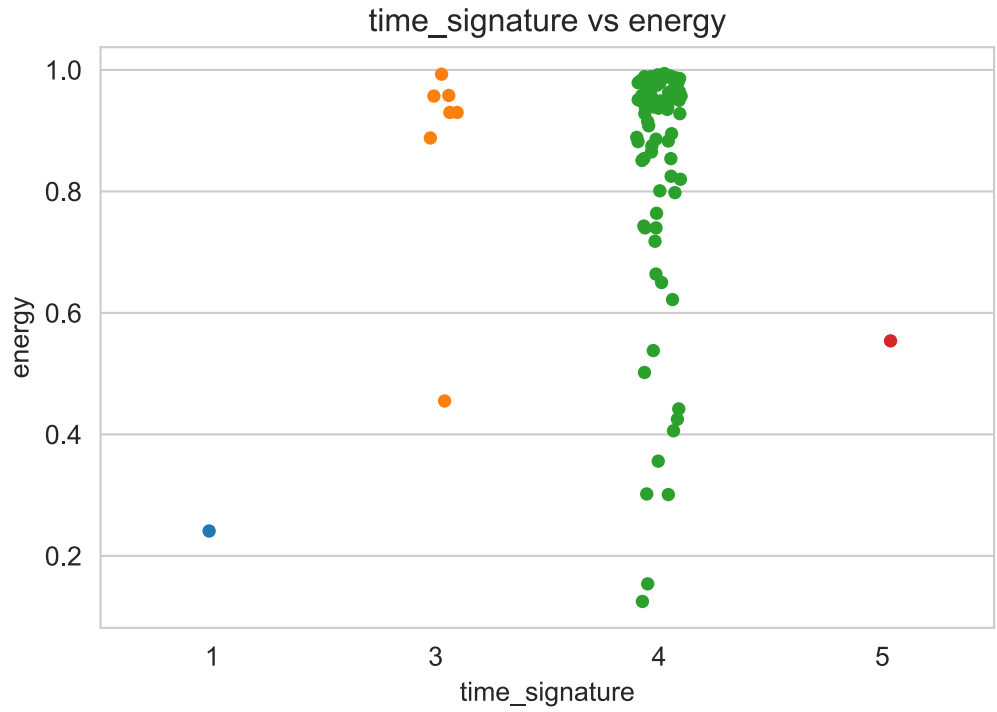

time_signature vs name_len
name_len
25
20
15
10
5
1
3
4
5
time_signature

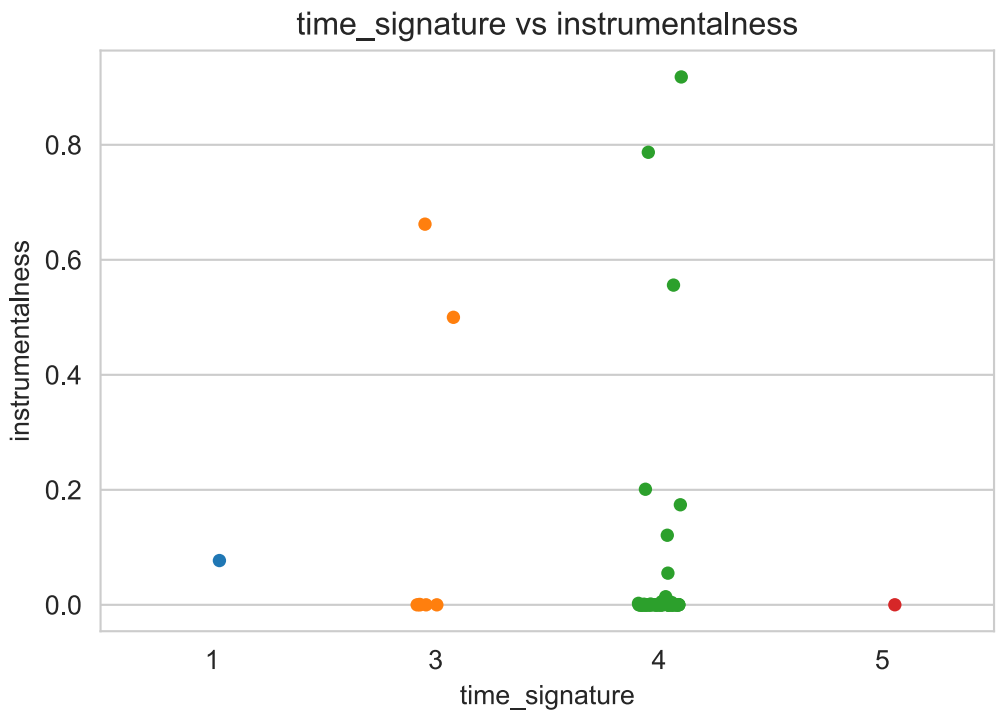

time_signature vs track_number
track_number
14
12
10
8
6
4
2
1
3
4
5
time_signature

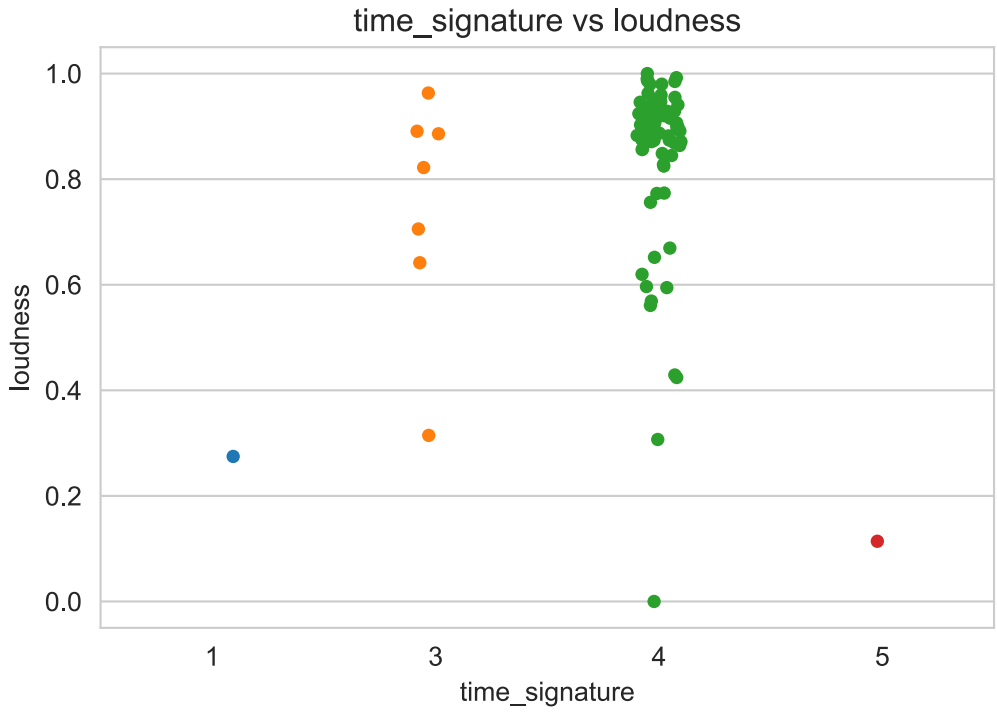

time_signature vs duration
duration
1.0
0.8
0.6
0.4
0.2
0.0
1
3
4
5
time_signature

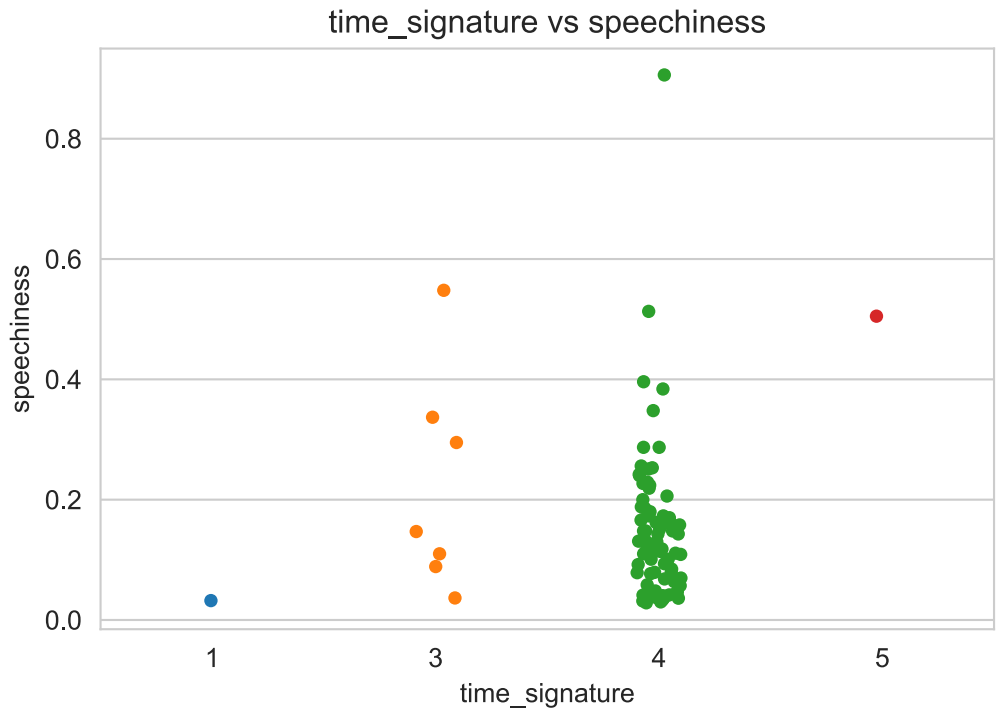

time_signature vs acousticness
acousticness
0.8
0.6
0.4
0.2
0.0
1
3
4
5
time_signature

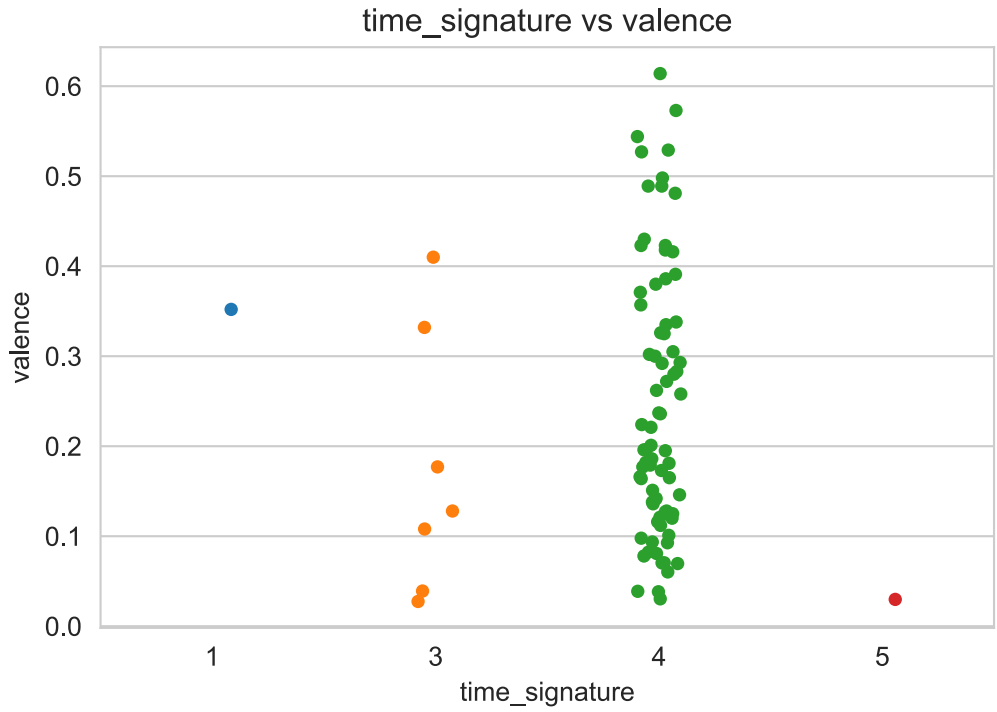

time_signature vs danceability
danceability
0.6
0.5
0.4
0.3
0.2
0.1
1
3
4
5
time_signature

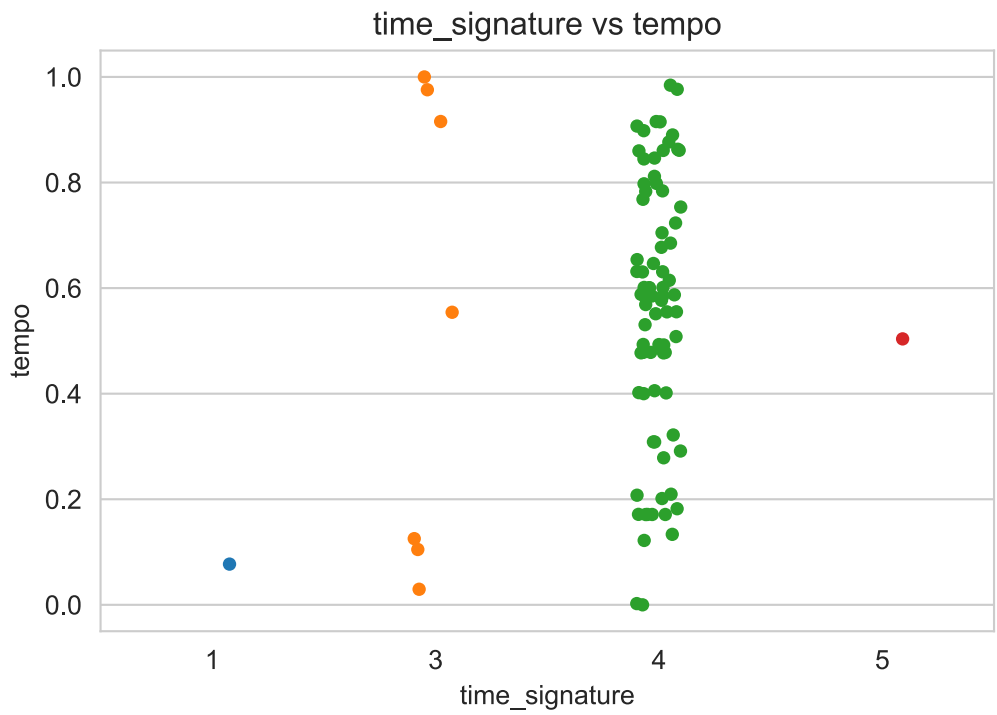

time_signature vs energy
energy
1.0
0.8
0.6
0.4
0.2
1
3
4
5
time_signature

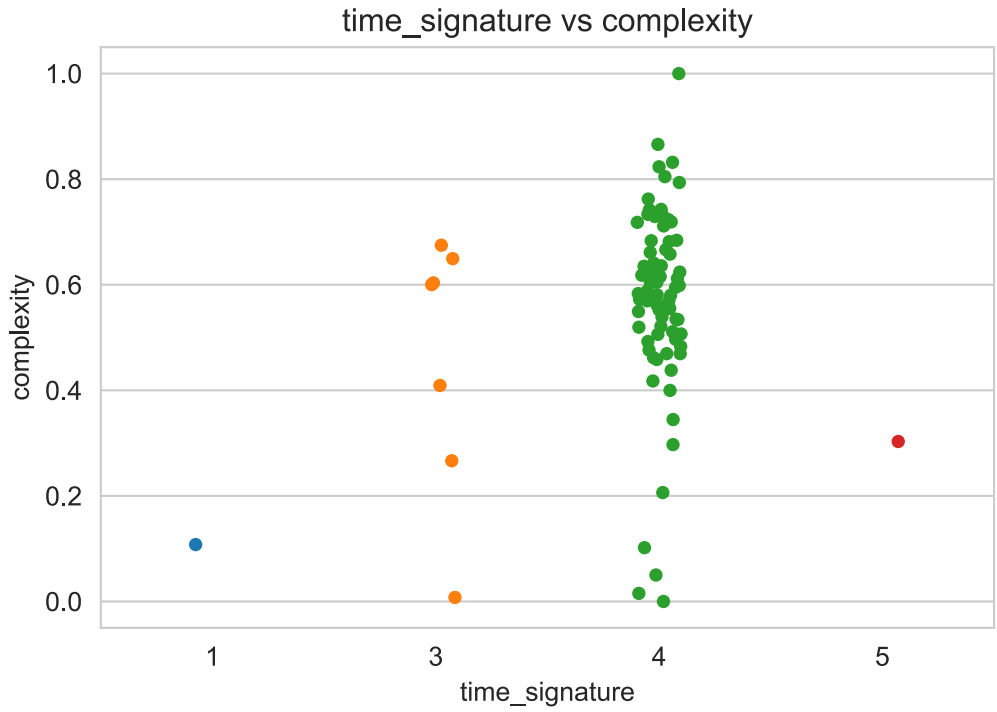

time_signature vs instrumentalness
instrumentalness
0.0
0.2
0.4
0.6
0.8
1
3
4
5
time_signature

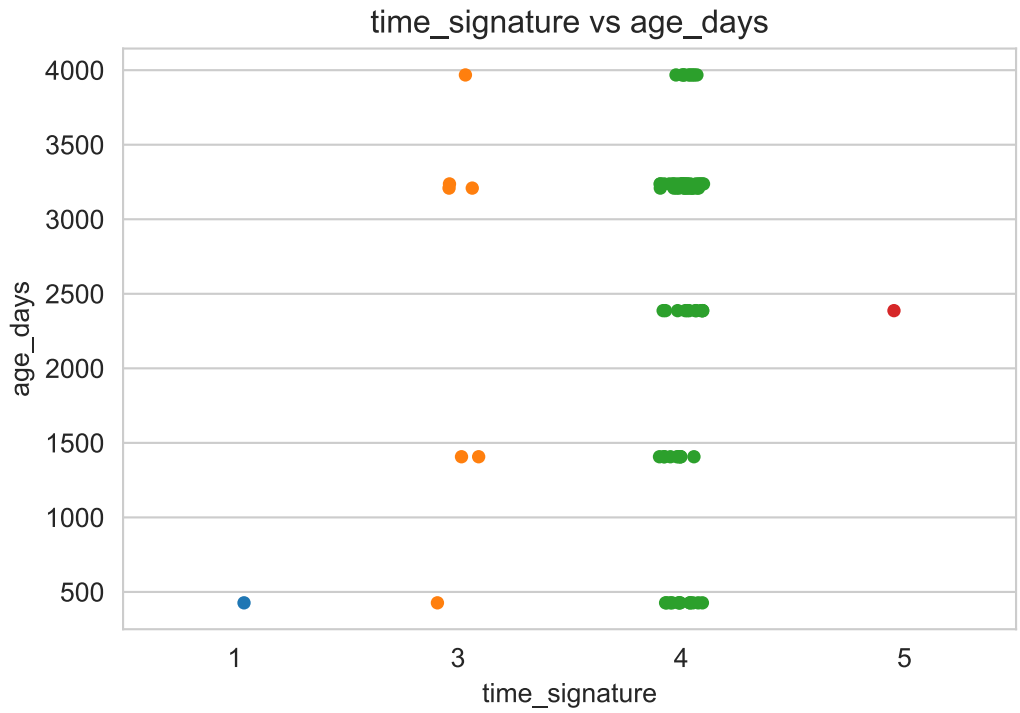

time_signature vs loudness
loudness
0.0
0.2
0.4
0.6
0.8
1.0
1
3
4
5
time_signature

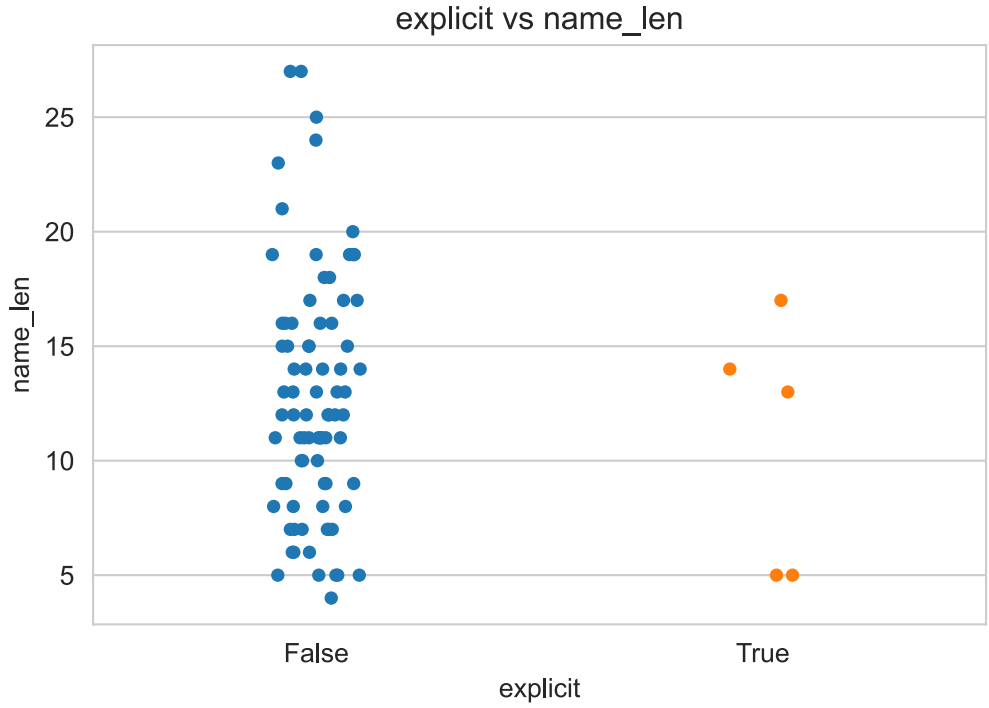

time_signature vs speechiness
speechiness
0.0
0.2
0.4
0.6
0.8
1
3
4
5
time_signature

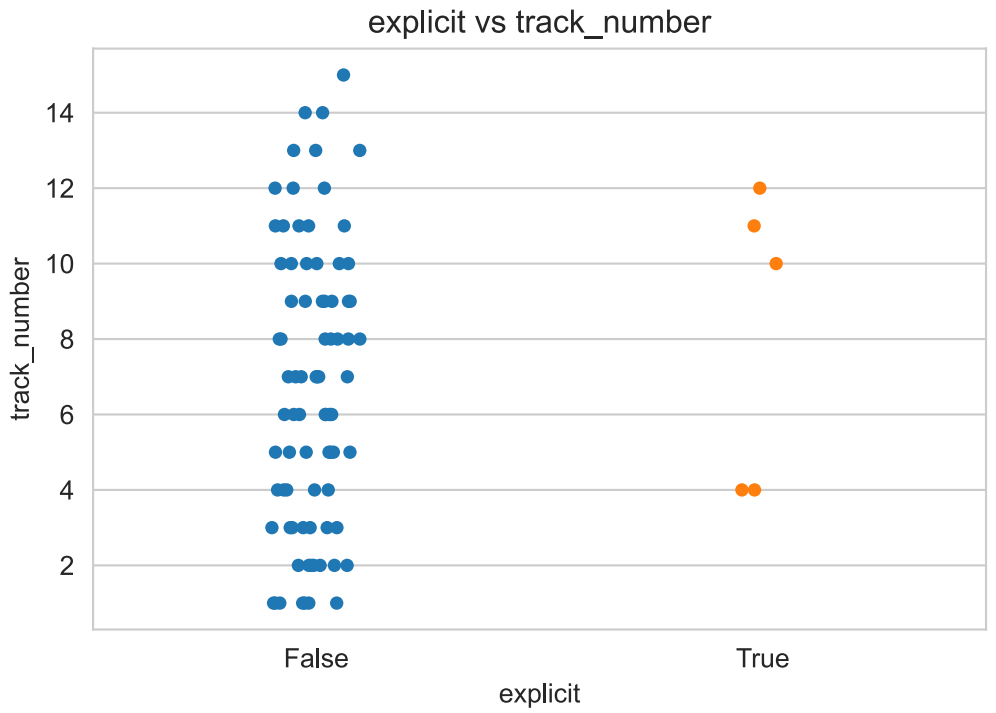

time_signature vs valence
valence
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
3
4
5
time_signature

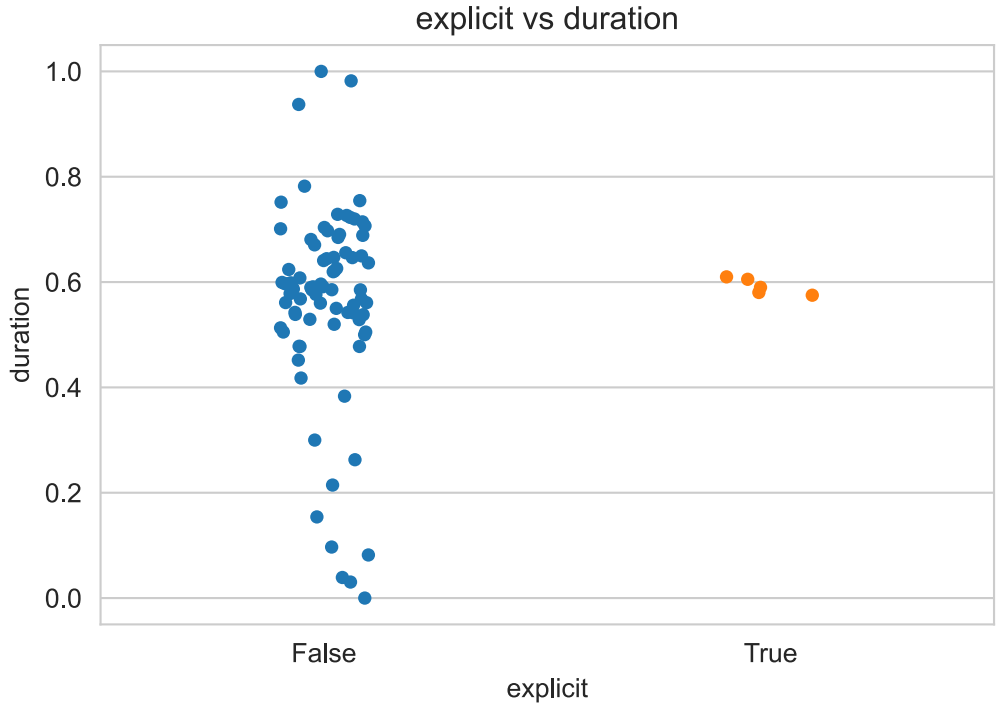

time_signature vs tempo
tempo
1.0
0.8
0.6
0.4
0.2
0.0
1
3
4
5
time_signature

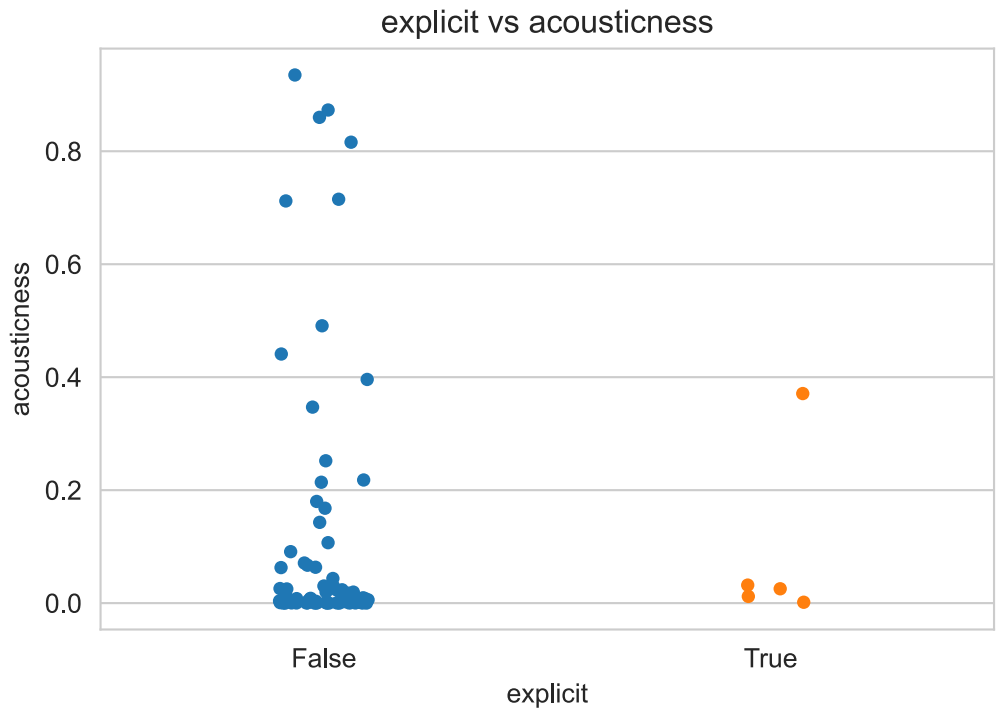

time_signature vs complexity
complexity
1.0
0.8
0.6
0.4
0.2
0.0
1
3
4
5
time_signature

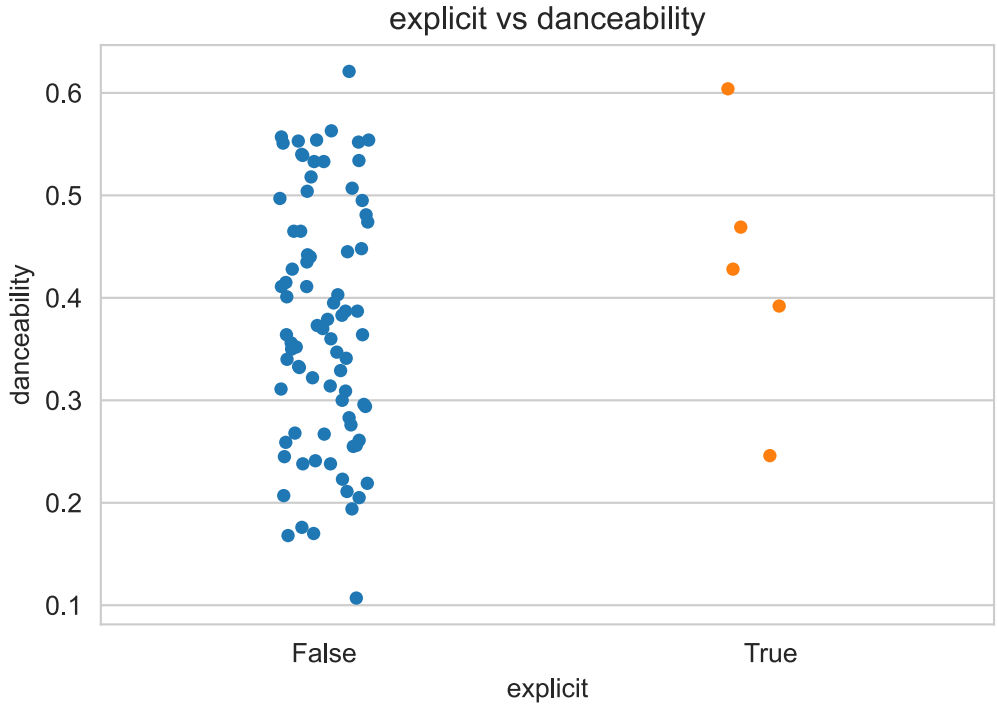

time_signature vs age_days
4000
3500
3000
2500
2000
1500
1000
500
age_days
1
3
4
5
time_signature

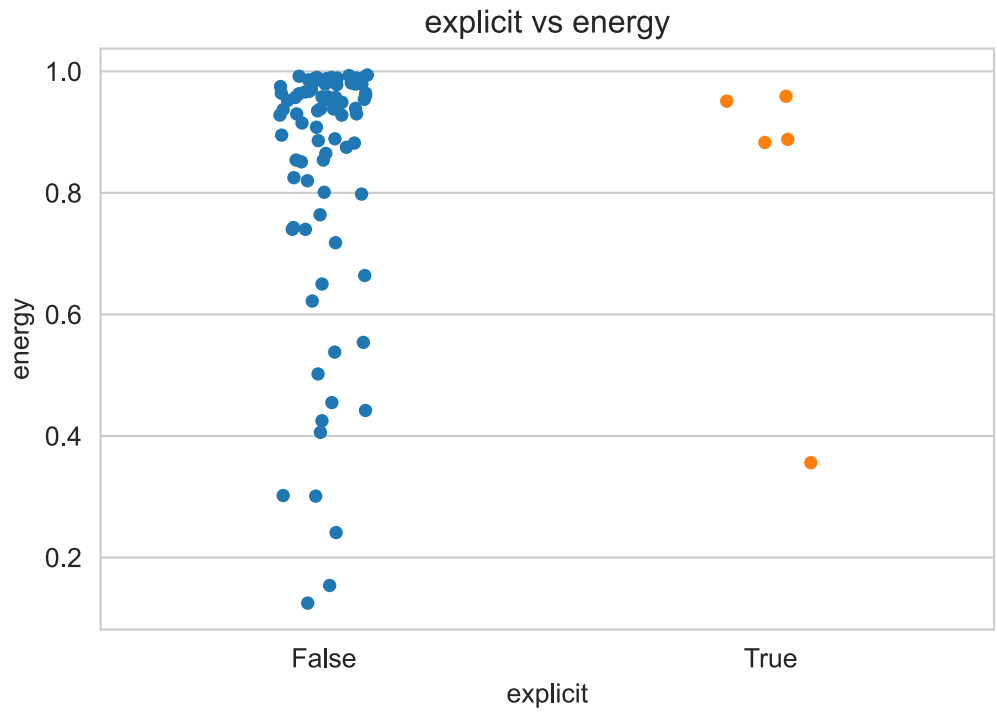

explicit vs name_len
25
20
15
10
5
name_len
False
True
explicit

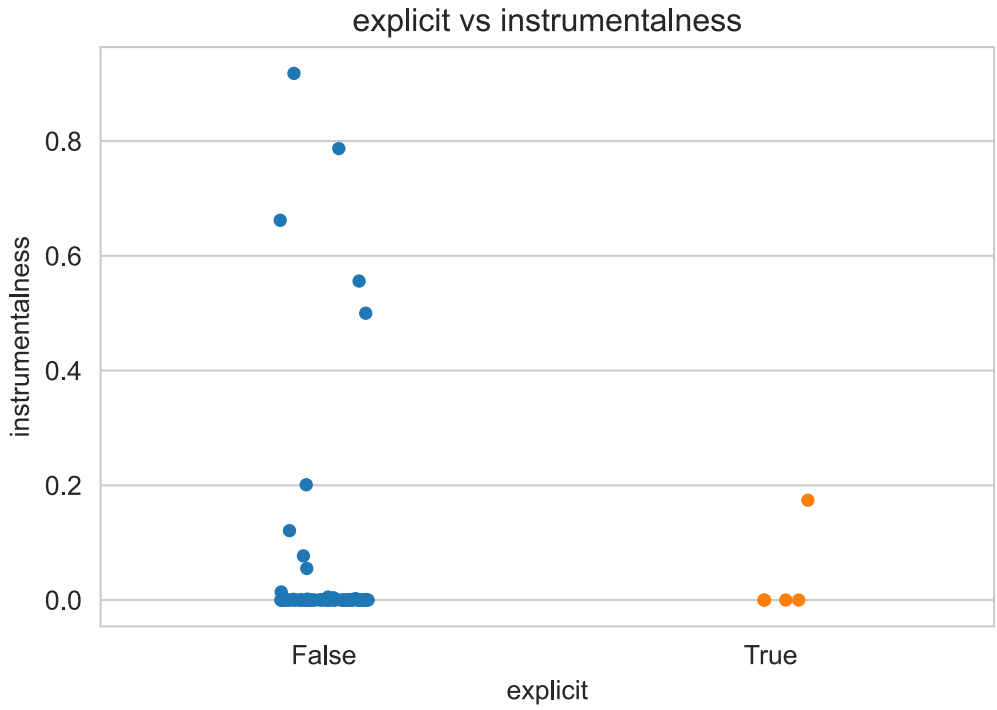

explicit vs track_number
track_number
14
12
10
8
6
4
2
False
True
explicit


explicit vs duration
duration
1.0
0.8
0.6
0.4
0.2
0.0
False
True
explicit


explicit vs acousticness
acousticness
0.0
0.2
0.4
0.6
0.8
False
True
explicit

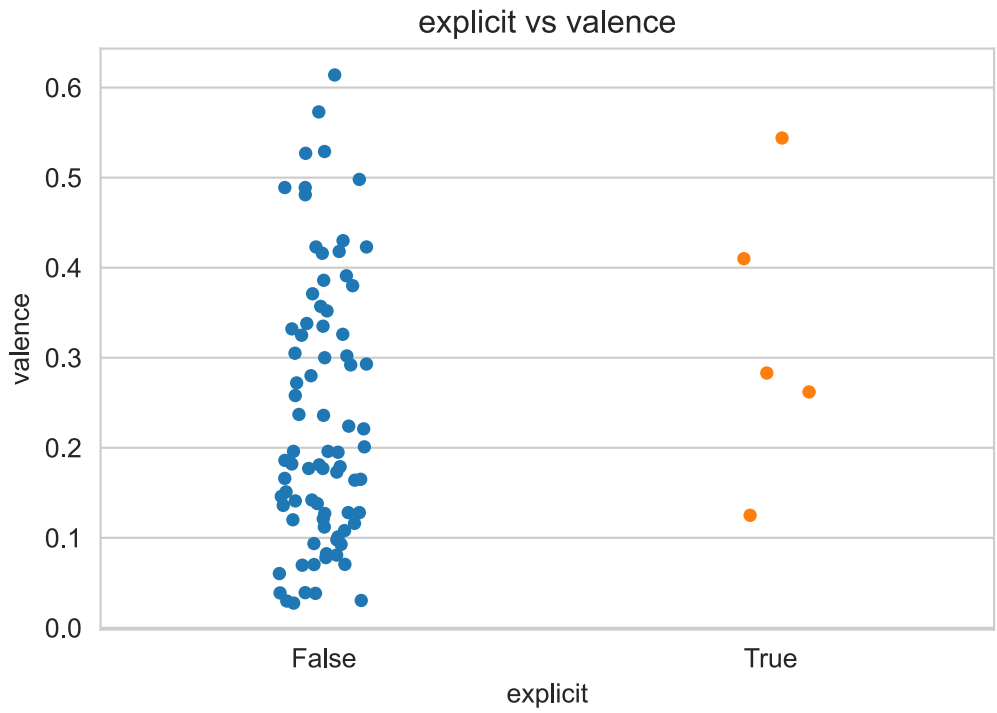

explicit vs danceability
danceability
0.1
0.2
0.3
0.4
0.5
0.6
False
True
explicit

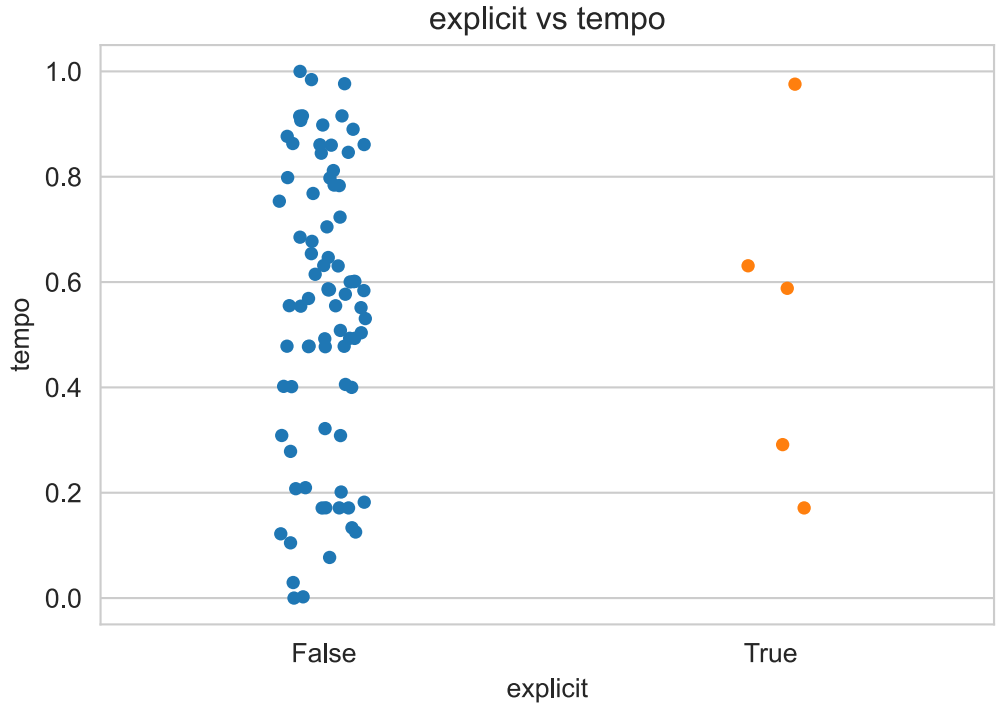

explicit vs energy
1.0
0.8
0.6
0.4
0.2
energy
False
True
explicit

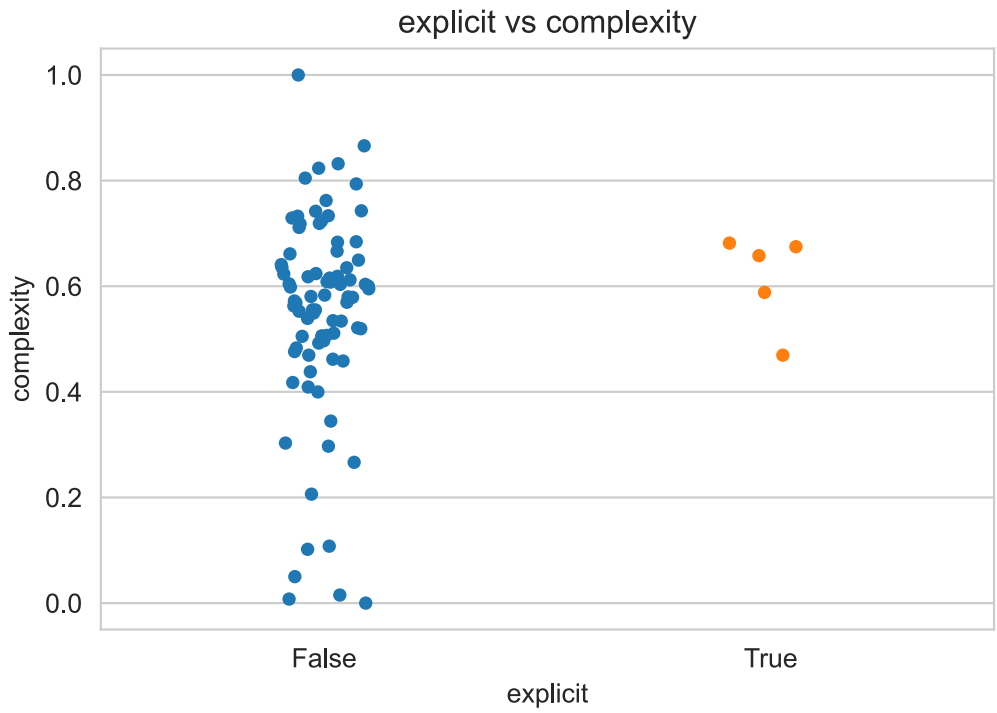

explicit vs instrumentalness
0.8
0.6
0.4
0.2
0.0
instrumentalness
False
True
explicit

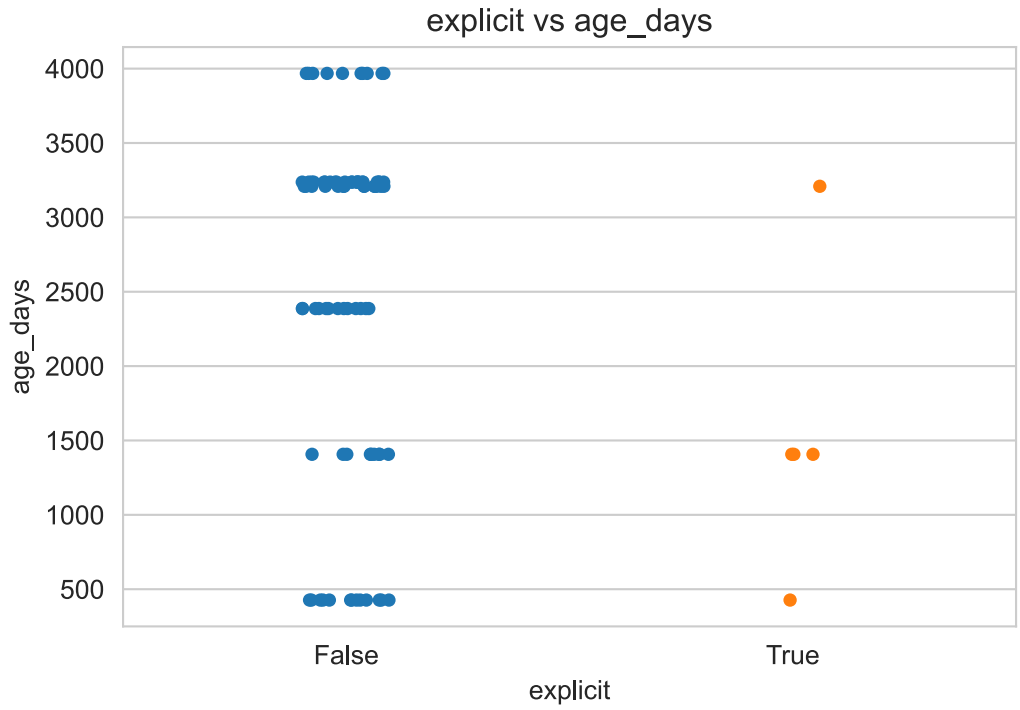

explicit vs loudness
1.0
0.8
0.6
0.4
0.2
0.0
loudness
False
True
explicit


explicit vs speechiness
0.8
0.6
0.4
0.2
0.0
speechiness
False
True
explicit


explicit vs valence
valence
0.0
0.1
0.2
0.3
0.4
0.5
0.6
False
True
explicit


explicit vs tempo
tempo
0.0
0.2
0.4
0.6
0.8
1.0
False
True
explicit


explicit vs complexity
1.0
0.8
0.6
0.4
0.2
0.0
complexity
False
True
explicit


explicit vs age_days
4000
3500
3000
2500
2000
1500
1000
500
age_days
False
True
explicit