# Analysing hashtags

The objective of this analysis is to see which hashtag is more popular in a time stamp, using `tweepy` tweets stream.

First we collected the data using the file `stream.py` for about ten minuts. Now let's anaylise it!

In [29]:

```python
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

Let's read our data

In [43]:

```python
df = pd.read_csv('hashtags.csv', header=None)

df.columns = ['time', 'hashtag']
df['time'] = pd.to_datetime(df['time'])

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 185 entries, 0 to 184
Data columns (total 2 columns):
time       185 non-null datetime64[ns]
hashtag    185 non-null object
dtypes: datetime64[ns](1), object(1)
memory usage: 3.0+ KB
```

How much which hashtag were tweet?

In [44]:

```python
df['hashtag'].value_counts()
```

Out[44]:

```
elenao      82
bolsonaro   65
forapt      27
haddad      11
Name: hashtag, dtype: int64
```

If we plot the results it's easier to see the differences!

In [45]:

```
plt.bar(df['hashtag'].unique(), df['hashtag'].value_counts())
```

Out[45]:

<BarContainer object of 4 artists>