

To complete this challenge, determine the five most common journals and the total articles for each. Next, calculate the mean, median, and standard deviation of the open-access cost per article for each journal.

```
In [1]: import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

%matplotlib inline
```

```
In [61]: journals = pd.read_csv(r'C:\Users\mikel\Documents\Thinkful\WELCOME.csv', encoding = 'ISO-8859-1')
journals.head()
```

Out[61]:

	PMID/PMCID	Publisher	Journal title	Article title	COST (£) charged to Wellcome (inc VAT when charged)
0	NaN	CUP	Psychological Medicine	Reduced parahippocampal cortical thickness in ...	£0.00
1	PMC3679557	ACS	Biomacromolecules	Structural characterization of a Model Gram-ne...	£2381.04
2	23043264 PMC3506128	ACS	J Med Chem	Fumaroylamino-4,5-epoxymorphinans and related ...	£642.56
3	23438330 PMC3646402	ACS	J Med Chem	Orvinols with mixed kappa/mu opioid receptor a...	£669.64
4	23438216 PMC3601604	ACS	J Org Chem	Regioselective opening of myo-inositol orthoes...	£685.88

```
In [62]: journals['Cost'] = journals['COST (£) charged to Wellcome (inc VAT when charged)'].str[1:]
journals.head()
```

Out[62]:

	PMID/PMCID	Publisher	Journal title	Article title	COST (£) charged to Wellcome (inc VAT when charged)	Cost
0	NaN	CUP	Psychological Medicine	Reduced parahippocampal cortical thickness in ...	£0.00	0.00
1	PMC3679557	ACS	Biomacromolecules	Structural characterization of a Model Gram-ne...	£2381.04	2381.04
2	23043264 PMC3506128	ACS	J Med Chem	Fumaroylamino-4,5-epoxymorphinans and related ...	£642.56	642.56
3	23438330 PMC3646402	ACS	J Med Chem	Orvinols with mixed kappa/mu opioid receptor a...	£669.64	669.64
4	23438216 PMC3601604	ACS	J Org Chem	Regioselective opening of myo-inositol orthoes...	£685.88	685.88

```
In [63]: journals['Journal title'] = journals['Journal title'].str.upper()
journals['Journal title'] = journals['Journal title'].str.replace('PLOSONE', 'PLOS ONE')
journals['Journal title'] = journals['Journal title'].str.replace('BMJ OPEN', 'BMJ')
journals['Journal title'] = journals['Journal title'].str.replace('PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES', 'PNAS')

print(journals['Journal title'].value_counts())

PLOS ONE 199
JOURNAL OF BIOLOGICAL CHEMISTRY 53
NEUROIMAGE 29
PNAS 26
PLOS PATHOGENS 24
PLOS GENETICS 24
NUCLEIC ACIDS RESEARCH 23
PLOS NEGLECTED TROPICAL DISEASES 20
HUMAN MOLECULAR GENETICS 19
BMJ 19
NATURE COMMUNICATIONS 19
BRAIN 14
BMC PUBLIC HEALTH 14
MOVEMENT DISORDERS 13
BIOCHEMICAL JOURNAL 12
DEVELOPMENTAL CELL 12
JOURNAL OF NEUROSCIENCE 12
CURRENT BIOLOGY 11
JOURNAL OF GENERAL VIROLOGY 11
PLOS COMPUTATIONAL BIOLOGY 9
CELL REPORTS 9
NEURON 9
PROCEEDINGS OF THE ROYAL SOCIETY B: BIOLOGICAL SCIENCES 8
MALARIA JOURNAL 8
JOURNAL OF STRUCTURAL BIOLOGY 8
EUROPEAN JOURNAL OF IMMUNOLOGY 8
DEVELOPMENT 8
NEUROBIOLOGY OF AGING 8
JOURNAL OF VIROLOGY 8
HEPATOLOGY 8
...
JOURNAL OF MOLECULAR MEDICINE 1
BIOCHIMICA ET BIOPHYSICA ACTA (BBA) - MOLECULAR CELL RESEARCH 1
BRITISH JOURNAL OF OPHTHALMOLOGY 1
JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY 1
INTERNATIONAL JOURNAL OF OBESITY 1
INTERNATIONAL JOURNAL OF GERIATRIC PSYCHIATRY 1
DATABASE 1
JOURNAL OF AUTISM AND DEVELOPMENT DISORDERS 1
CHEMICAL SOCIETY REVIEWS 1
JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY 1
PLOS NTD 1
PARASIT VECTORS. 1
PMEDICINE-D-12-03130 1
JOURNAL OF CELLULAR BIOCHEMISTRY 1
AM J BIOETH 1
FRONTIERS IN INTEGRATIVE NEUROSCIENCE 1
GENESIS: JOURNAL OF GENETICS 1
JOURNAL OF ARCHAEOLOGICAL SCIENCE 1
INFLUENZA OTHER RESPI VIRUSES. 1
ANGEWANDE CHEMIE 1
AMERICAN JOURNAL OF GERIATRIC PSYCHIATRY 1
JOURNAL OF CELL PHYSIOLOGY 1
OSTEOPOROSIS INTERNATIONAL 1
THORAX AN INTERNATIONAL JOURNAL FOR RESPIRATORY MEDICINE 1
GENETICS 1
CURRENT OPINION MICROBIOLOGY 1
NEUROBIOLOGY OF LEARNING AND MEMORY 1
MOLECULAR PAIN 1
METABOLOMICS 1
OXFORD JOURNALS 1
Name: Journal title, Length: 925, dtype: int64
```

Five most common journals are:

- PLOS ONE
- JOURNAL OF BIOLOGICAL CHEMISTRY
- NEUROIMAGE
- PNAS
- PLOS GENETICS

```
In [71]: plos_one = journals.loc[journals['Journal title'] == 'PLOS ONE']
plos_one['Cost'] = pd.to_numeric(plos_one['Cost'])

print("The mean for Plos One is ", np.mean(plos_one['Cost']))
print("The median of Plos One is ", np.median(plos_one['Cost']))
print("The standard deviation of Plos One is ", np.std(plos_one['Cost']))

The mean for Plos One is 42058.791306532665
The median of Plos One is 897.19
The standard deviation of Plos One is 196518.0451146453

C:\Users\mikel\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
In [74]: JBC = journals.loc[journals['Journal title'] == 'JOURNAL OF BIOLOGICAL CHEMISTRY']
JBC['Cost'] = pd.to_numeric(JBC['Cost'])
print("The number of articles for JBC is ", JBC['Article title'].count())

print("The mean for JBC is ", np.mean(JBC['Cost']))
print("The median of JBC is ", np.median(JBC['Cost']))
print("The standard deviation of JBC is ", np.std(JBC['Cost']))

The number of articles for JBC is 53
The mean for JBC is 20264.63396226415
The median of JBC is 1314.53
The standard deviation of JBC is 135865.31224533138

C:\Users\mikel\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
In [76]: neuroimage = journals.loc[journals['Journal title'] == 'NEUROIMAGE']
neuroimage['Cost'] = pd.to_numeric(neuroimage['Cost'])
print("The number of articles for Neuroimage is ", neuroimage['Article title'].count())

print("The mean for Neuroimage is ", np.mean(neuroimage['Cost']))
print("The median of Neuroimage is ", np.median(neuroimage['Cost']))
print("The standard deviation of Neuroimage is ", np.std(neuroimage['Cost']))

The number of articles for Neuroimage is 29
The mean for Neuroimage is 2215.168275862069
The median of Neuroimage is 2326.43
The standard deviation of Neuroimage is 262.0161331841066

C:\Users\mikel\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
In [77]: pnas = journals.loc[journals['Journal title'] == 'PNAS']
pnas['Cost'] = pd.to_numeric(pnas['Cost'])
print("The number of articles for PNAS is ", pnas['Article title'].count())

print("The mean for PNAS is ", np.mean(pnas['Cost']))
print("The median of PNAS is ", np.median(pnas['Cost']))
print("The standard deviation of PNAS is ", np.std(pnas['Cost']))

The number of articles for PNAS is 26
The mean for PNAS is 39141.821153846155
The median of PNAS is 728.825
The standard deviation of PNAS is 192171.52167056446

C:\Users\mikel\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
In [80]: plos_genetics = journals.loc[journals['Journal title'] == 'PLOS GENETICS']
plos_genetics['Cost'] = pd.to_numeric(plos_genetics['Cost'])
print("The number of articles for Plos Genetics is ", plos_genetics['Article title'].count())

print("The mean for Plos Genetics is ", np.mean(plos_genetics['Cost']))
print("The median of Plos Genetics is ", np.median(plos_genetics['Cost']))
print("The standard deviation of Plos Genetics is ", np.std(plos_genetics['Cost']))

The number of articles for Plos Genetics is 24
The mean for Plos Genetics is 84839.435
The median of Plos Genetics is 1718.3899999999999
The standard deviation of Plos Genetics is 275931.0282404746

C:\Users\mikel\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```