Problem 1

Create a data frame that includes two columns, one named "Animals" and the other named "Foods". The first column should be this vector (note the intentional repeated values): Dog, Cat, Fish, Fish, Lizard

The second column should be this vector: Bread, Orange, Chocolate, Carrots, Milk

Write your code below:

Import Libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import datetime
from datetime import datetime, timedelta
import scipy.stats
import pandas_profiling
from pandas_profiling import ProfileReport


%matplotlib inline
#sets the default autosave frequency in seconds
%autosave 60
sns.set_style('dark')
sns.set(font_scale=1.2)

plt.rc('axes', titlesize=9)
plt.rc('axes', labelsize=14)
plt.rc('xtick', labelsize=12)
plt.rc('ytick', labelsize=12)

import warnings
warnings.filterwarnings('ignore')

# Use Folium library to plot values on a map.
#import folium

#import feature_engine.missing_data_imputers as mdi
#from feature_engine.outlier_removers import Winsorizer
#from feature_engine import categorical_encoders as ce

pd.set_option('display.max_columns',None)
#pd.set_option('display.max_rows',None)
pd.set_option('display.width', 1000)
pd.option_context('float_format','{:.2f}'.format)

np.random.seed(0)
np.set_printoptions(suppress=True)
```

```
Autosaving every 60 seconds
```

In [2]:
```
columns = ["Animals","Foods"]
```

In [3]:
```
df = pd.DataFrame(data={"Animals":['Dog', 'Cat', 'Fish', 'Fish', 'Lizard'],
                        "Foods":['Bread', 'Orange', 'Chocolate', 'Carrots', 'Milk']}, columns=columns)
```

In [4]:
```
df
```

Out[4]:

	Animals	Foods
0	Dog	Bread
1	Cat	Orange
2	Fish	Chocolate
3	Fish	Carrots
4	Lizard	Milk

Problem 2

Using the data frame created in Problem 2, use the table() command to create a frequency table for the column called "Animals".

Write your code below:

In [5]:
```
df.groupby("Animals").count()
```

Out[5]:

Animals	Foods
Cat	1
Dog	1
Fish	2
Lizard	1

Problem 3

Use read.csv() to import the survey data included in this assignment. Using that data, make a histogram of the column called "pid7".

Write your code below:

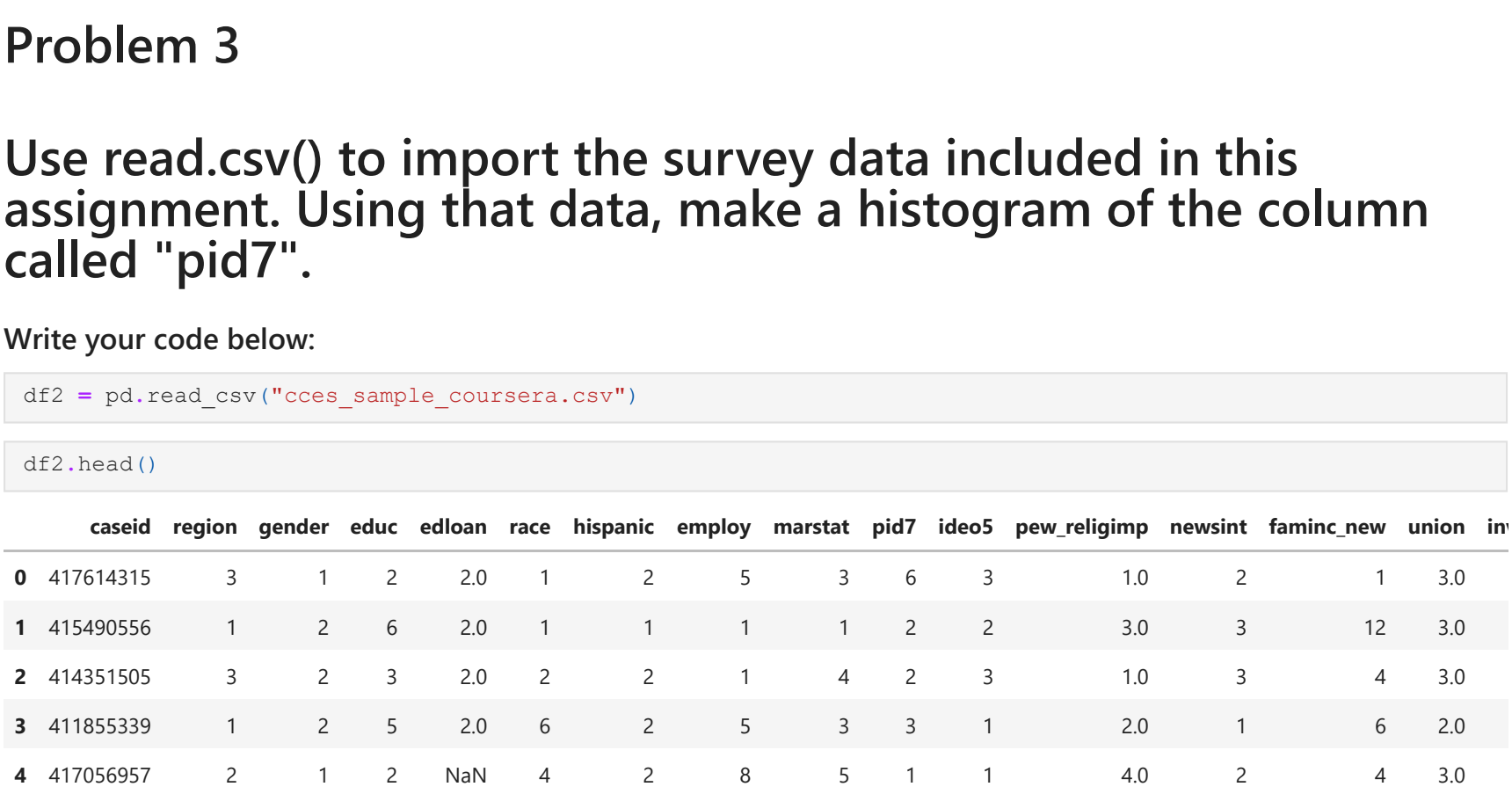
In [6]:
```
df2 = pd.read_csv("cces_sample_coursera.csv")
```

In [7]:
```
df2.head()
```

Out[7]:

	caseid	region	gender	educ	edloan	race	hispanic	employ	marstat	pid7	ideo5	pew_religimp	newsint	faminc_new	union	in
0	417614315	3	1	2	2.0	1	2	5	3	6	3	1.0	2	1	3.0	
1	415490556	1	2	6	2.0	1	1	1	1	2	2	3.0	3	12	3.0	
2	414351505	3	2	3	2.0	2	2	1	4	2	3	1.0	3	4	3.0	
3	411855339	1	2	5	2.0	6	2	5	3	3	1	2.0	1	6	2.0	
4	417056957	2	1	2	NaN	4	2	8	5	1	1	4.0	2	4	3.0	

In [8]:
```
df2["pid7"].hist(bins=50, figsize=(20,10))
plt.suptitle('Histogram of dat$pid7', x=0.5, y=1.02, ha='center', fontsize=20)
plt.xlabel("dat$pid7", fontsize=20)
plt.ylabel("Frequency", fontsize=20)
plt.tight_layout()
plt.show()
```



In [ ]:
```
```