NAKERAKNTI SRAVAN KUMAR
RA2111047010189
DATA ANALYSIS PROJECT
2ND SEM

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

In [2]:
```
data = pd.read_csv('global_internet.csv')
data.head()
```

Out[2]:

	country	incomeperperson	internetuserate	urbanrate
0	Afghanistan	NaN	3.654122	24.04
1	Albania	1914.996551	44.989947	46.72
2	Algeria	2231.993335	12.500073	65.22
3	Andorra	21943.339900	81.000000	88.92
4	Angola	1381.004268	9.999954	56.70

In [3]:
```
print(data.isnull().sum())
```

```
country             0
incomeperperson    23
internetuserate    21
urbanrate          10
dtype: int64
```

In [4]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 213 entries, 0 to 212
Data columns (total 4 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   country          213 non-null    object 
 1   incomeperperson  190 non-null    float64
 2   internetuserate  192 non-null    float64
 3   urbanrate        203 non-null    float64
dtypes: float64(3), object(1)
memory usage: 6.8+ KB
```

In [5]:
```
data.describe()
```

Out[5]:

	incomeperperson	internetuserate	urbanrate
count	190.000000	192.000000	203.000000
mean	8740.966076	35.632716	56.769360
std	14262.809083	27.780285	23.844933
min	103.775857	0.210066	10.400000
25%	748.245151	9.999604	36.830000
50%	2553.496056	31.810121	57.940000
75%	9379.891166	56.416046	74.210000
max	105147.437700	95.638113	100.000000

In [6]:
```
data.groupby('country').sum()
```

Out[6]:

country	incomeperperson	internetuserate	urbanrate
Afghanistan	0.000000	3.654122	24.04
Albania	1914.996551	44.989947	46.72
Algeria	2231.993335	12.500073	65.22
Andorra	21943.339900	81.000000	88.92
Angola	1381.004268	9.999954	56.70
...	...	...	...
Vietnam	722.807559	27.851822	27.84
West Bank and Gaza	0.000000	36.422772	71.90
Yemen, Rep.	610.357367	12.349750	30.64
Zambia	432.226337	10.124986	35.42
Zimbabwe	320.771890	11.500415	37.34

In [7]:
```
data.value_counts('urbanrate')
```

Out[7]:
```
urbanrate
100.00    6
65.58     2
36.84     2
27.84     2
61.34     2
         ..
46.72     1
46.78     1
46.84     1
47.04     1
57.18     1
Length: 194, dtype: int64
```

In [8]:
```
data = data.drop('incomeperperson',axis=1)
```

In [9]:
```
for col in data.columns:
    print(col)
    print(data[col].value_counts())
    print("         ----------          ")
    print()
```

```
country
Afghanistan             1
Nauru                   1
Netherlands             1
Netherlands Antilles    1
New Caledonia           1
                       ..
Grenada                 1
Guadeloupe              1
Guam                    1
Guatemala               1
Zimbabwe                1
Name: country, Length: 213, dtype: int64
         ----------

internetuserate
3.654122     1
44.989947    1
51.958038    1
49.000632    1
4.170136     1
            ..
9.196775     1
26.297251    1
82.526898    1
9.549931     1
11.500415    1
Name: internetuserate, Length: 192, dtype: int64
         ----------

urbanrate
100.00    6
27.84     2
36.84     2
65.58     2
```

```
61.34     2
          ..
83.52     1
30.84     1
93.16     1
48.58     1
37.34     1
Name: urbanrate, Length: 194, dtype: int64
         ----------
```

In [10]:
```
len(data.internetuserate.unique())
```

Out[10]:
```
193
```

In [11]:
```
data.isnull().sum()
```

Out[11]:
```
country            0
internetuserate    21
urbanrate          10
dtype: int64
```

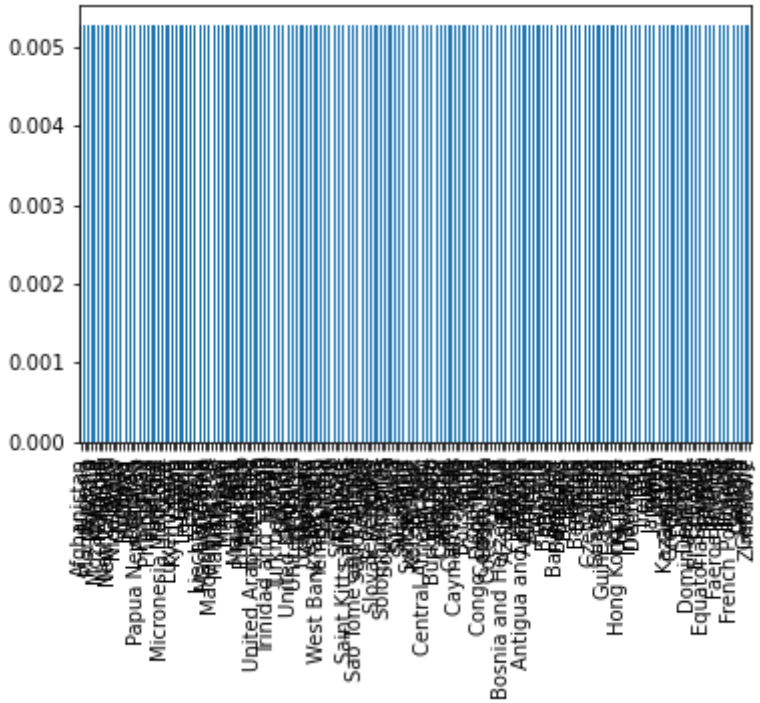
In [12]:
```
data = data.dropna(axis=0)
```

In [13]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 190 entries, 0 to 212
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   country          190 non-null    object
 1   internetuserate  190 non-null    float64
 2   urbanrate        190 non-null    float64
dtypes: float64(2), object(1)
memory usage: 5.9+ KB
```

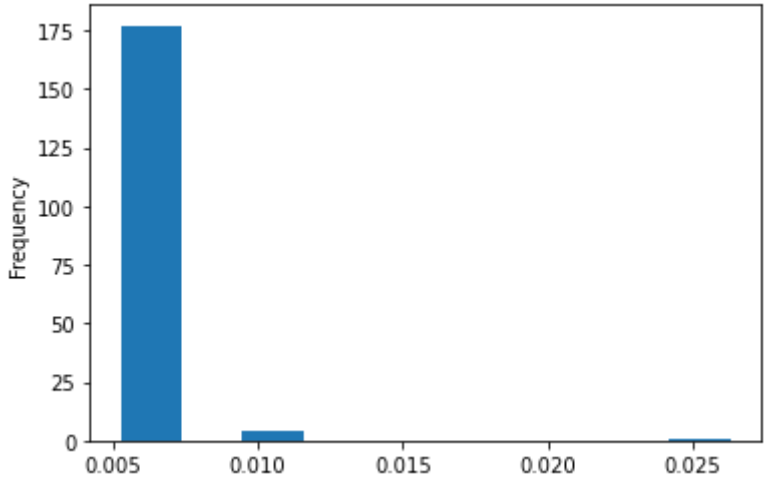
In [14]:
```
m = data.country.value_counts(normalize=True)
m.plot.bar()
```

Out[14]: <AxesSubplot:>



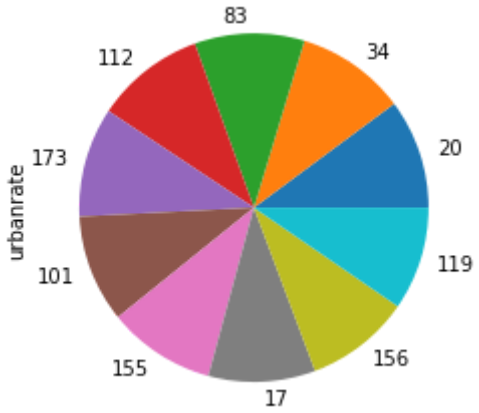
In [15]:
```
m = data.urbanrate.value_counts(normalize=True)
m.plot.hist()
```

Out[15]: <AxesSubplot:ylabel='Frequency'>



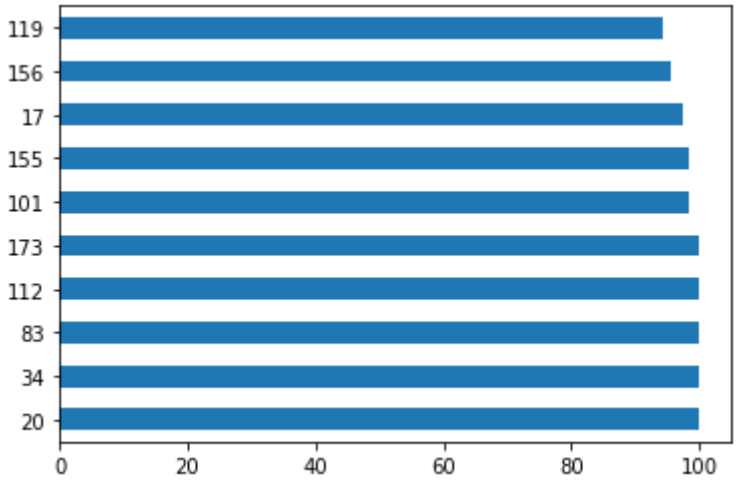
In [16]:
```
data["urbanrate"].nlargest(10).plot(kind='pie')
```

Out[16]: <AxesSubplot:ylabel='urbanrate'>



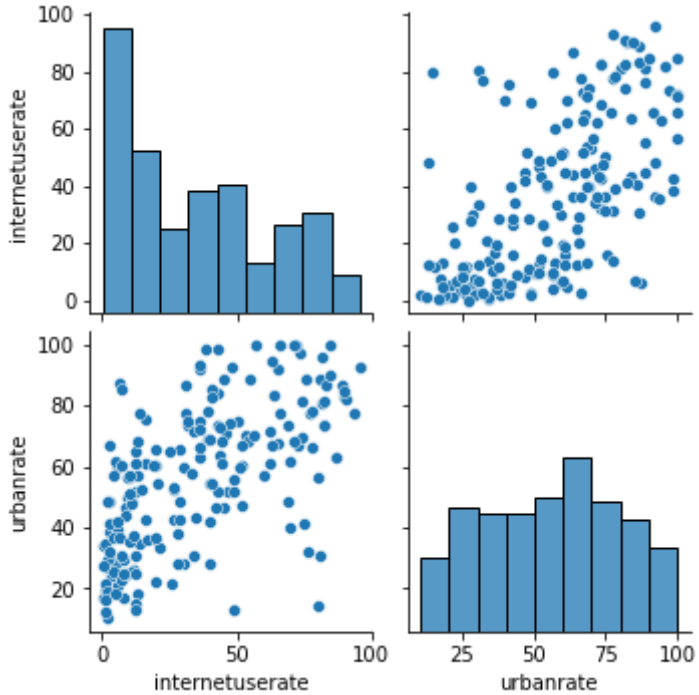
In [17]:
```
data["urbanrate"].nlargest(10).plot(kind='barh')
```

Out[17]: <AxesSubplot:>



In [18]:
```
import seaborn as sns
sns.pairplot(data)
```

Out[18]: <seaborn.axisgrid.PairGrid at 0x14ccebc9a90>



In [ ]: