Analysis of Homeless Datasets (from HUD website)
Python version of an original analysis done using R

Source code & Data files on Github at:
https://github.com/susub31/Homelessness-Analysis/tree/master/Python-Files

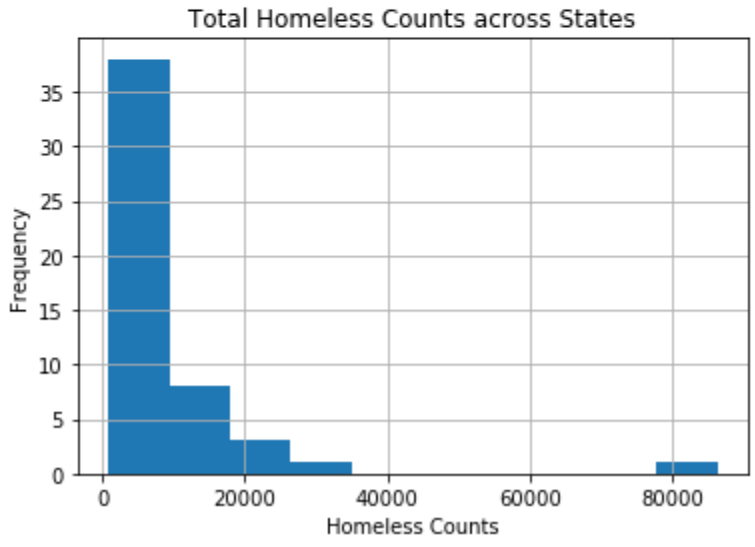
```python
"""
Homeless Data Analysis using Python
[Reproducing in Python, an original analysis that was done using R]

@author: sudha
"""

from pandas import read_csv
import pandas as pd
from pandas import set_option
from matplotlib import pyplot as plt

# Read data from local file - that holds CoC names and the lat/lon
COCNames = ['CoCNumber', 'lon', 'lat']
filename = 'COCNumWithGeoCodes.csv'
dataframe = read_csv(filename, names=COCNames)
array = dataframe.values

# Perform some data cleaning
# Remove rows with NAs in the lat / lon columns
dfCoC = read_csv(filename, names=COCNames)
dfCoC.dropna(subset=['lat'], inplace=True)
dfCoC.dropna(subset=['lon'], inplace=True)

HomelessNames = ['CoCNumber', 'Indv', 'ShelteredIndv', 'UnshelteredIndv', 'PeopleFamilies',
                 'ShelteredPeopleFamilies', 'UnshelteredPeopleFamilies', 'ShelteredVeterans',
                 'UnshelteredVeterans', 'ShelteredYouth', 'UnshelteredYouth']
HomelessFilename = 'HomelessData2016.csv'
dfHomeless = read_csv(HomelessFilename, names=HomelessNames)

# Combine values in specific columns to aggregate counts in each category and form new columns
dfHomeless['Sheltered'] = dfHomeless['ShelteredIndv'] + dfHomeless['ShelteredPeopleFamilies']
dfHomeless['Unsheltered'] = dfHomeless['UnshelteredIndv'] + dfHomeless['UnshelteredPeopleFamilies']

# Aggregate counts to get totals in each category and form new columns
dfHomeless['Total'] = dfHomeless['Sheltered'] + dfHomeless['Unsheltered']
dfHomeless['Veterans'] = dfHomeless['ShelteredVeterans'] + dfHomeless['UnshelteredVeterans']
dfHomeless['Youth'] = dfHomeless['ShelteredYouth'] + dfHomeless['UnshelteredYouth']

set_option('display.width', 100)
set_option('precision', 3)
description=dfHomeless.describe()
print(description)

# Select aggregate columns into a separate dataset
data = dfHomeless[['Total', 'Veterans', 'Youth']]
data.Total.plot(kind='box', subplots=True, layout=(3,3), sharex=False, sharey=False)
plt.show()

subdata = data[data['Total'] < 20000]
subdata.Total.plot(kind='box', subplots=True, layout=(3,3), sharex=False, sharey=False)
plt.show()

subdata.Total.hist()
plt.show()

# Merge the datasets - similar to inner join in Database, based on a key column, CoCNumber in this case
# Purpose is to associate the Lat and lon values with the homeless dataset
newdf = pd.merge(dfHomeless, dfCoC, on='CoCNumber')
```

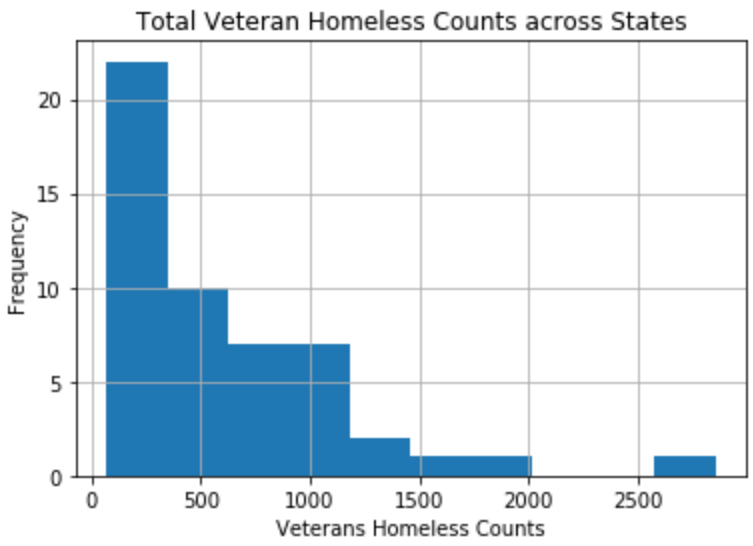
```
60
61 # Extract state name from the CoCNumber as a separate column
62 newdf['State'] = newdf['CoCNumber'].str[:2]
63
64 # Read States Dataset
65 StatesNames = ['StateName', 'State']
66 StateFilename = 'StateNames.csv'
67 dfStates = read_csv(StateFilename, names=StatesNames)
68
69 # Now get totals in each state and sort in descending order
70 #grouped = newdf.groupby('State').size().to_frame(name='Total').reset_index()
71 group1 = newdf.groupby(by=['State'])['Total'].sum().sort_values(ascending=False)
72 group2 = newdf.groupby(by=['State'])['Veterans'].sum().sort_values(ascending=False)
73 group3 = newdf.groupby(by=['State'])['Youth'].sum().sort_values(ascending=False)
74
75 df1 = group1.rename(None).to_frame()
76 df2 = group2.rename(None).to_frame()
77 df3 = group3.rename(None).to_frame()
78
79 # form new columns with names for merging into one
80 df1['State'] = df1.index
81 df1['Total'] = df1[[0]]
82 df2['State'] = df2.index
83 df2['Veterans'] = df2[[0]]
84 df3['State'] = df3.index
85 df3['Youth'] = df3[[0]]
86
87
88 # Merge State names into the newly formed dataset
89 grouped = pd.merge(df1, df2, on='State')
90 grouped = pd.merge(grouped, df3, on='State')
91 grouped.drop(grouped.index[0], inplace=True)
92
93 # Reset index and group only columns of interest that captues total in each category
94 grouped = grouped.reset_index(drop=True)
95 grouped = grouped[['State', 'Total', 'Veterans', 'Youth']]
96
97 # Display to console top 10 states with high homeless counts
98 print("Below are top 10 states with high homeless counts:")
99 print(group1.head(10))
100
101
102 # Display to console top 10 states with high homeless counts among Veterans
103 print("Below are top 10 states with high homeless Veteran counts:")
104 print(group2.head(10))
105
106 # Display to console top 10 states with high homeless counts (Total)
107 print("Below are top 10 states with high homeless counts:")
108 print(grouped.head(10))
109
110 grouped.Total.hist()
111 plt.title("Total Homeless Counts across States")
112 plt.xlabel("Homeless Counts")
113 plt.ylabel("Frequency")
114 plt.show()
115
116 grouped.Veterans.hist()
117 plt.title("Total Veteran Homeless Counts across States")
118 plt.xlabel("Veterans Homeless Counts")
119 plt.ylabel("Frequency")
120 plt.show()
121
122 grouped.Youth.hist()
123 plt.title("Total Youth Homeless Counts across States")
124 plt.xlabel("Youth Homeless Counts")
125 plt.ylabel("Frequency")
126 plt.show()
127
```

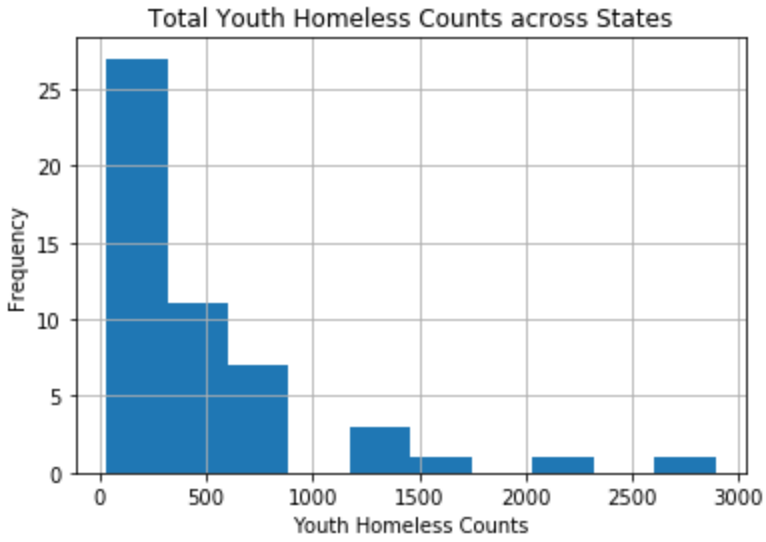
Output:

```
In [36]: runfile('C:/Users/sudha/Documents/Sudha/Analytics/Python-Related/Homeless-Data-Analysis/Homeless-Data-Analysis.py',
wdir='C:/Users/sudha/Documents/Sudha/Analytics/Python-Related/Homeless-Data-Analysis')
             Indv  ShelteredIndv  Unsheltered Indv  PeopleFamilies  ShelteredPeopleFamilies  \
count     402.000        402.000           402.000         402.000                  402.000
mean      883.612        492.557           391.055         484.368                  436.724
std      2526.263       1418.567          1654.549        2286.520                 2271.470
min        10.000          1.000             0.000           7.000                    4.000
25%       180.500        117.000            29.250          89.250                   81.250
50%       360.500        212.000            75.000         191.500                  166.500
75%       860.500        493.250           282.500         357.750                  317.750
max     37726.000      26127.000         30950.000       44558.000                44558.000

       UnshelteredPeopleFamilies  ShelteredVeterans  UnsheltedVeterans  ShelteredYouth  \
count                    402.000            402.000            402.000         402.000
mean                      47.644             65.682             32.505          47.731
std                      167.772            107.390             97.559         103.194
min                        0.000              0.000              0.000           0.000
25%                        0.000              9.000              1.000          10.000
50%                        4.000             28.500              6.000          23.500
75%                       25.000             83.750             26.000          50.750
max                     1831.000           1243.000           1485.000        1653.000

       UnshelteredYouth  Sheltered  Unsheltered     Total  Veterans     Youth
count           402.000    402.000      402.000    402.00   402.000   402.000
mean             41.040    929.281      438.699   1367.98    98.187    88.771
std             163.626   3654.467     1761.696   4425.29   187.166   226.251
min               0.000     11.000        0.000     20.00     0.000     0.000
25%               1.000    209.750       32.000    296.25    14.000    15.000
50%               6.000    377.000       84.000    550.50    41.500    36.500
75%              25.750    841.500      338.500   1267.75   117.750    85.000
max            2272.000  70685.000    32781.000  73523.00  2728.000  3086.000
```

```
Below are top 10 states with high homeless counts:
  State  Total  Veterans  Youth
0    NY  86352      1248   2889
1    FL  32964      2853   2064
2    TX  23122      1768   1309
3    WA  20827      1484   1307
4    MA  19608       949    374
5    PA  15339      1136    868
6    OR  13238      1341   1175
7    GA  12909      1055    725
8    IL  11212       937    651
9    CO  10550      1181    653
```




Total Veteran Homeless Counts across States
Frequency
0
5
10
15
20
0
500
1000
1500
2000
2500
Veterans Homeless Counts


Total Youth Homeless Counts across States
Frequency
0
5
10
15
20
25
0
500
1000
1500
2000
2500
3000
Youth Homeless Counts