Homework8

May 29, 2023

1 Stats 21 - HW 8 - Due 6/3/2023 by 11:59PM

1.1 Please modify this line by replacing it with your name and SID thank you.

Homework is generally an opportunity to practice coding and to train your problem solving and critical thinking skills. Putting Python to use is where learning happens.

Copying and pasting another's solutions takes away your learning opportunities. It is also academic dishonesty.

ChatGPT is always allowed in this class, but do remember, it is not foolproof and if your solution looks too much like another submission, I am required to file a report

Please use this document as your homework template and submit both the modified .ipynb file and a PDF OR HTML export.

1.2 Introduction

The data were derived from the US Centers for Disease Control 2010. It can also be found in Tableau. https://www.cdc.gov/obesity/data/index.html

1.3 Description of the Data

- County: Name of location
- Region: Region of the US
- State: State Name
- State_ABB: State Abbrevation (e.g., CA)
- Adult Obesity: Percentage Obese (BMI $>$ 30)
- Adult Smokers: Percentage Smokers
- Children in Poverty: Percentage in Poverty (under age 18)
- Diabetic: Percentage Diabetic (all ages)
- Food Insecure: Percentage reporting difficulty having enough food to eat (all ages)
- Physically Inactive: Percentage Physically Inactive (all ages)

1.4 Problem 1: read the data

Correctly read the data from the CSV file named "Obesity.csv" using pandas and provide evidence (e.g., dimensions, data summary) that it was correctly read. Some of the data values may require

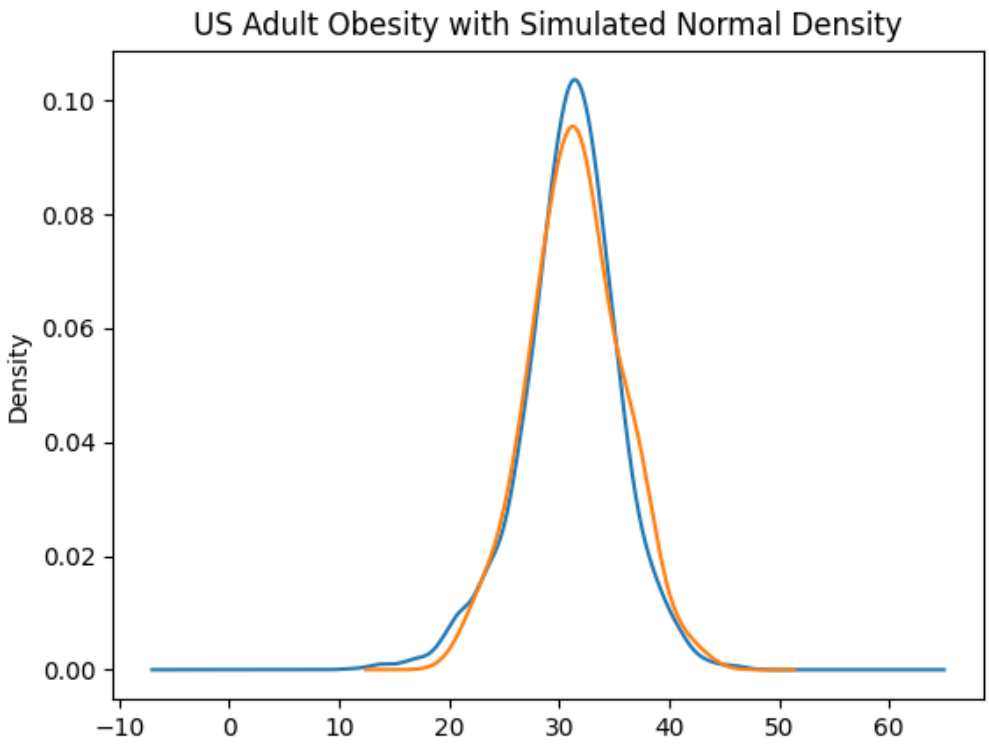
cleaning/correction before they can be used properly.

[20]:
```
## reserved for your answer
## read the data, remove the % and convert columns to numeric

import numpy as np
import pandas as pd

df = pd.read_csv("Obesity.csv")

df = df.dropna()
df['Adult Obesity'] = df['Adult Obesity'].str.replace('%', '').astype(float)
df['Adult Smokers'] = df['Adult Smokers'].str.replace('%', '').astype(float)
df['Children in Poverty'] = df['Children in Poverty'].str.replace('%', '').
 ↪astype(float)
df['Diabetic'] = df['Diabetic'].str.replace('%', '').astype(float)
df['Food Insecure'] = df['Food Insecure'].str.replace('%', '').astype(float)
df['Physically Inactive'] = df['Physically Inactive'].str.replace('%', '').
 ↪astype(float)


print(df.describe())

print(df.shape)
```

```
       Adult Obesity  Adult Smokers  Children in Poverty     Diabetic
count    3140.000000    3140.000000          3140.000000  3140.000000  \
mean       31.001911      18.433758            23.712739    11.177389
std         4.465720       3.811303             8.958772     2.310977
min        11.000000       0.000000             0.000000     0.000000
25%        29.000000      16.000000            17.000000    10.000000
50%        31.000000      18.000000            23.000000    11.000000
75%        34.000000      21.000000            29.000000    13.000000
max        47.000000      41.000000            66.000000    23.000000

       Food Insecure  Physically Inactive
count    3140.000000          3140.000000
mean       15.113057            27.406688
std         3.935967             5.410562
min         4.000000             9.000000
25%        13.000000            24.000000
50%        15.000000            28.000000
75%        17.000000            31.000000
max        33.000000            42.000000
(3140, 10)
```

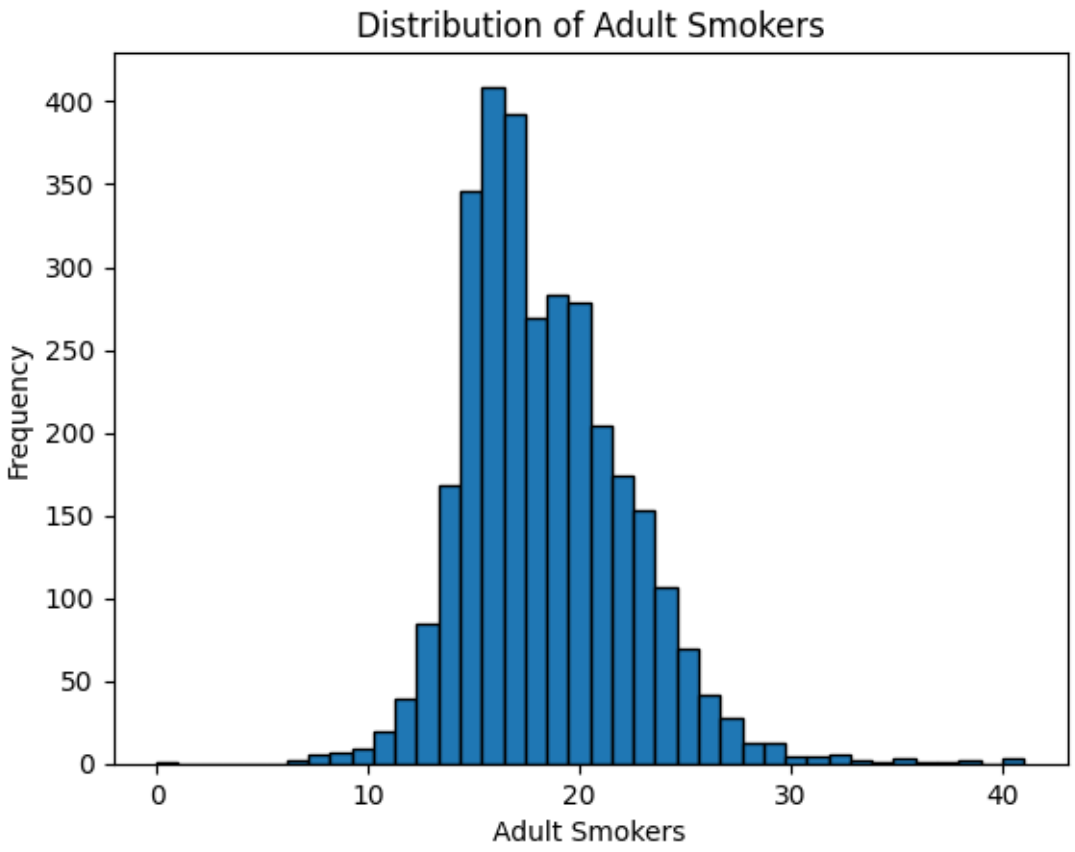
1.5 Problem 2: Choose any one numeric variable and create a simple graphic

In a comment, please offer an interpretation of what you are seeing in the data that is illustrated/made visible with your graphic. For example here is a graphic I made:



It appears that the distribution of obesity rates by state is close to normally distributed around a mean of 31% Obese

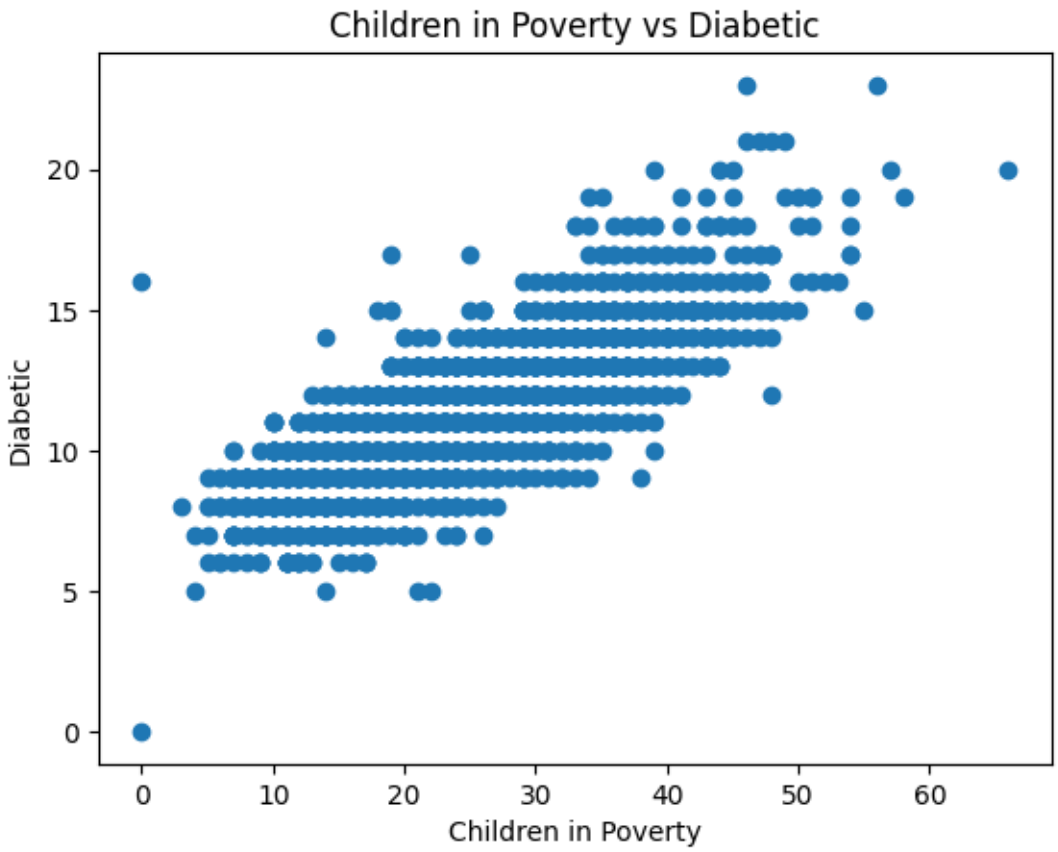
[38]:
```
## reserved for your answer

import matplotlib.pyplot as plt

plt.hist(df['Adult Smokers'], bins=40, edgecolor='black')
plt.xlabel('Adult Smokers')
plt.ylabel('Frequency')
plt.title('Distribution of Adult Smokers')
plt.show()
```



1.6 Problem 3: Choose any two numeric variables and create a graphic

Similar to #2, in a comment, please offer an interpretation of what you are seeing in the data that is illustrated/made visible with your graphic.

[44]:
```
## reserved for your answer

plt.scatter(df['Children in Poverty'], df['Diabetic'])

# Set labels and title
plt.xlabel('Children in Poverty')
plt.ylabel('Diabetic')
plt.title('Children in Poverty vs Diabetic')

# Show the plot
plt.show()
```



[44]:

	County	Region	State	State_ABB	Adult Obesity	Adult Smokers	\
0	Adams	Midwest	Illinois	IL	35.0	16.0	
1	Alexander	Midwest	Illinois	IL	32.0	24.0	
2	Bond	Midwest	Illinois	IL	31.0	17.0	
3	Boone	Midwest	Illinois	IL	34.0	16.0	
4	Brown	Midwest	Illinois	IL	32.0	16.0	
…	…	…	…	…	…	…	
3135	Fremont	West	Wyoming	WY	26.0	19.0	
3136	Goshen	West	Wyoming	WY	28.0	18.0	
3137	Platte	West	Wyoming	WY	27.0	17.0	
3138	Washakie	West	Wyoming	WY	25.0	16.0	
3139	Hot Springs	West	Wyoming	WY	25.0	16.0	

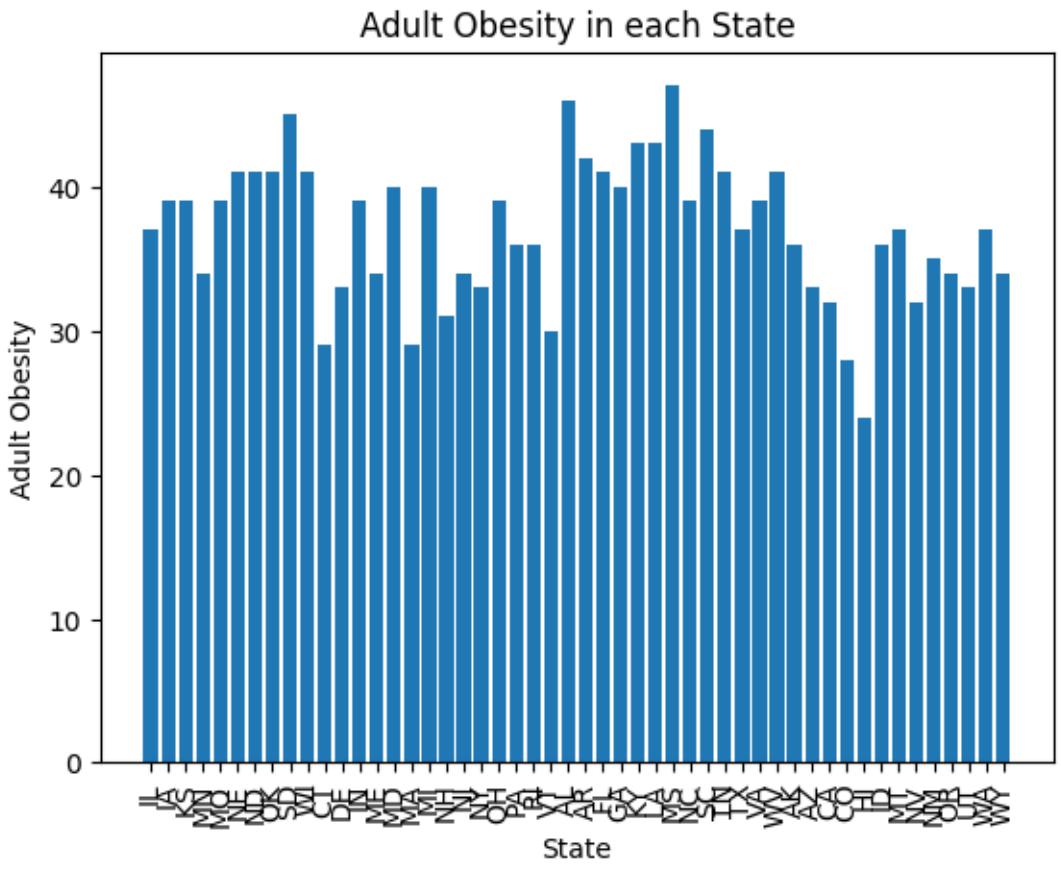
	Children in Poverty	Diabetic	Food Insecure	Physically Inactive
0	20.0	10.0	13.0	24.0
1	52.0	16.0	22.0	30.0
2	21.0	10.0	13.0	26.0
3	15.0	10.0	11.0	24.0
4	15.0	8.0	12.0	27.0
…	…	…	…	…

```
3135                20.0      10.0          14.0                25.0
3136                19.0      10.0          14.0                27.0
3137                19.0      10.0          13.0                27.0
3138                16.0      11.0          12.0                22.0
3139                17.0      11.0          14.0                26.0

[3140 rows x 10 columns]
```

1.7 Problem 4: Choose one non-numeric variable and one numeric variable a create a graphic

Similar to #2, in a comment, please offer an interpretation of what you are seeing in the data that is illustrated/made visible with your graphic.

[46]:
```
## reserved for your answer



plt.bar(df['State_ABB'],df['Adult Obesity'])
plt.xlabel('State')
plt.ylabel('Adult Obesity')
plt.title('Adult Obesity in each State')
plt.xticks(rotation='vertical')
plt.show()
```



1.8 Problem 5: Your choice of a custom plot

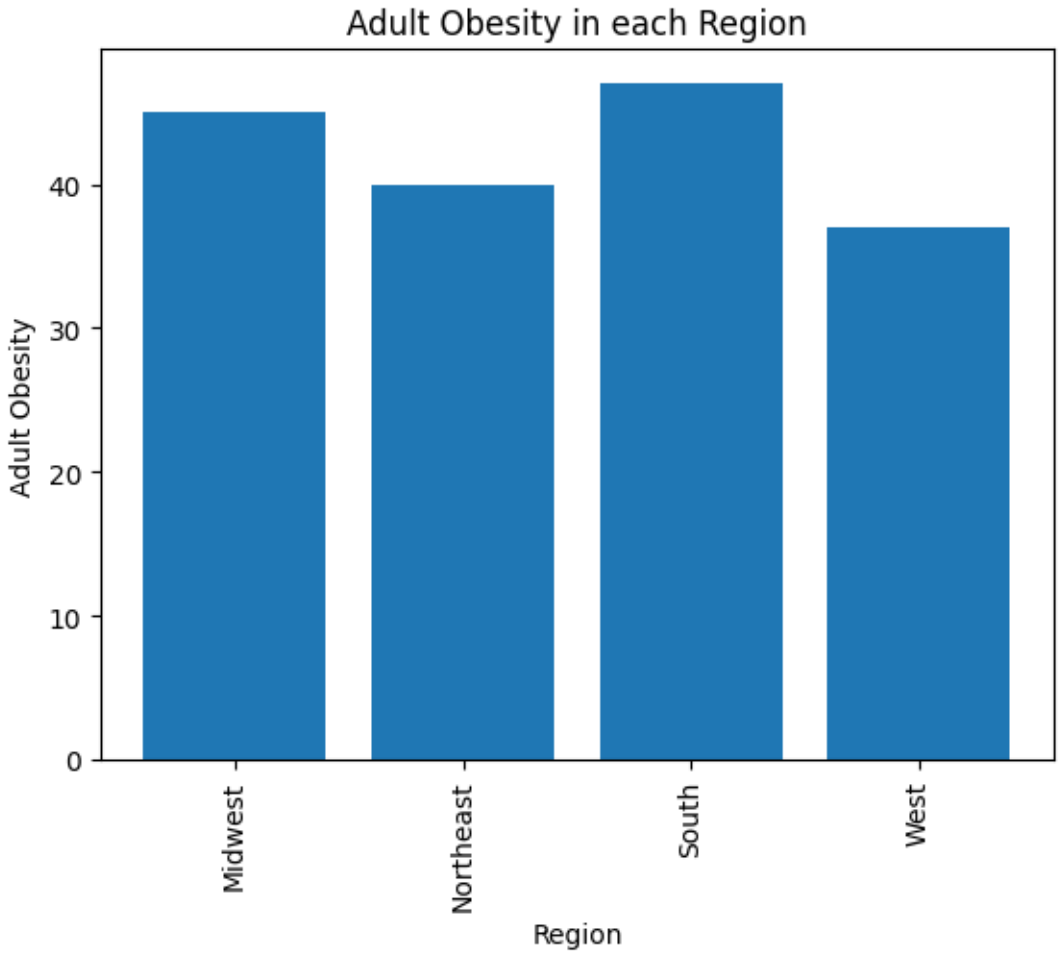
This is your opportunity to be creative. Use whatever module/package/library you want.

Make sure you label your axes with descriptive names and give a title to the graphic. Make sure your graph displays in your PDF or HTML submission

Please write a few sentences telling us about your decision of graphic type. For example, suppose you decide to create a map, we would like to know your justification, like "Oh, I thought it would be easy for anyone to understand because…"

[47]:
```
## reserved for your answer



plt.bar(df['Region'],df['Adult Obesity'])
plt.xlabel('Region')
plt.ylabel('Adult Obesity')
plt.title('Adult Obesity in each Region')
plt.xticks(rotation='vertical')
```

```
plt.show()
```


Adult Obesity in each Region
Adult Obesity
0
10
20
30
40
Midwest
Northeast
South
West
Region