Data Analysis

José-Clemente Hernández-Hernández

Homework 1

September 24, 2020

1. Import the 2018 Storm Events data and determine on which months have the most tornadoes.

Libraries to use in this homework:

```
In [1]: import numpy as np

   ...: import pandas as pd

   ...: import matplotlib.pyplot as plt
```

Importing data:

```
In [2]: df = pd.read_csv("StormEvents_2018.csv")
```

Which months have the most tornadoes:

```
In [3]: df[df["Event_Type"] == "Tornado"]["Month"].value_counts()

Out [1]:

     May          184          September    121          December      68
     June         159          November     108          March         66
     April        150          July          97          February      55
     October      134          August        90          January       16
```

The months that have the most tornados are May, June, April, and October.

2. Using the same data set, which is the third most frequent event type?

```
In [4]: df["Event_Type"].value_counts()

Out [2]:

                      Thunderstorm Wind            14585
                      Hail                          7861
                      Flood                         4715
                      Winter Weather                4478
                      Flash Flood                   4358
                      Winter Storm                  3375
```

The third most frequent event type is “Flood”.

3. Import the file named "hw1.csv".

Importing:

```
In [5]: df2 = pd.read_csv("hw1.csv")
```

3.1 Identify the data types on each column:

```
In [6]: df2.dtypes
```

```
Out [3]:
```

```
   Unnamed: 0     object      palmitoleic    object      linolenic      object
   region          int64      "stearic"      object      arachidic      object
   area            int64      oleic          object      eicosenoic     object
   palmitic        int64      linoleic       object          dtype: object
```

3.2 Plot the values of columns 5 through 11, each column individually, beginning from the second row.

First, get a new `DataFrame` with the columns and rows:

```
In [7]: new_df = df2.iloc[1:, 4:11]
```

Second, the values of the attributes are not integers, so for this, the values will be change from string to int:

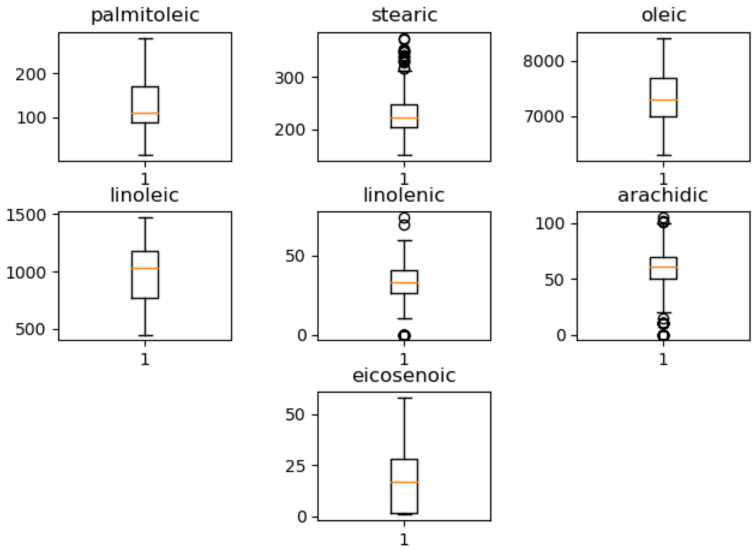
```
In [8]: for column in new_df.columns:

   ...:      for value in range(len(new_df)):

   ...:            new_df[column].iloc[value] = new_df[column].iloc[value]
           .replace('\"', "")

In [9]: new_df["stearic"] = new_df['\"stearic\"'].values

   ...: new_df = new_df.drop('\"stearic\"', axis = 1)

   ...: new_df = new_df.astype(int)

   ...: new_df  =  new_df.reindex(columns  =  ["palmitoleic",  "stearic",  "oleic",
           "linoleic", "linolenic", "arachidic", "eicosenoic"])
```

In the input 9 the “Stearic” column the values were changed to another column with the name well written.

Finally, the plotting of the values:

```
In [10]: fig, axs = plt.subplots(3, 3)
    ...: axs[0, 0].boxplot(new_df[new_df.columns[0]])
    ...: axs[0, 0].set_title(new_df.columns[0])
    ...: axs[0, 1].boxplot(new_df[new_df.columns[1]])
    ...: axs[0, 1].set_title(new_df.columns[1])
    ...: axs[0, 2].boxplot(new_df[new_df.columns[2]])
    ...: axs[0, 2].set_title(new_df.columns[2])
    ...: axs[1, 0].boxplot(new_df[new_df.columns[3]])
    ...: axs[1, 0].set_title(new_df.columns[3])
    ...: axs[1, 1].boxplot(new_df[new_df.columns[4]])
    ...: axs[1, 1].set_title(new_df.columns[4])
    ...: axs[1, 2].boxplot(new_df[new_df.columns[5]])
    ...: axs[1, 2].set_title(new_df.columns[5])
    ...: axs[2, 1].boxplot(new_df[new_df.columns[6]])
    ...: axs[2, 1].set_title(new_df.columns[6])
    ...: fig.delaxes(axs[2, 0])
    ...: fig.delaxes(axs[2, 2])
    ...:   fig.subplots_adjust(left=0.08,   right=0.98,   bottom=0.05,   top=0.9,
hspace=0.4, wspace=0.5)
```

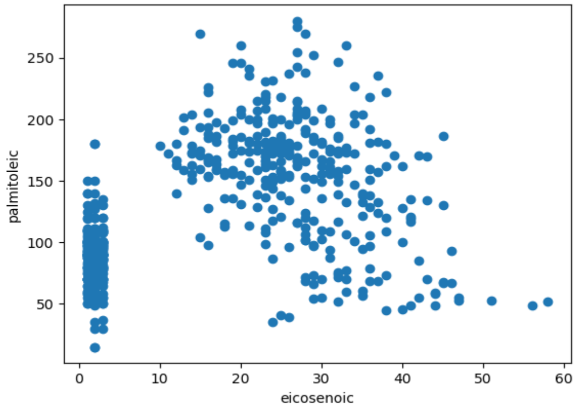
```
Out [4]:
```



3.3 Plot the values of columns 11 and 5 together. The plot should look like this:

```
In [11]: plt.scatter(new_df["eicosenoic"], new_df["palmitoleic"])

    ...: plt.xlabel("eicosenoic")

    ...: plt.ylabel("palmitoleic")

    ...: plt.show()
```

```
Out [5]:
```


palmitoleic
eicosenoic