```
Name: Borris A. Esplanada
Course and Section: CPE019 - CPE32S1
Date of Submission: 06/07/2024
Instructor: Engr. Roman M. Richard
```

PART 1

```
#code cell 1

import pandas as pd
brainFile = './sample_data/brainsize.txt'
brainFrame = pd.read_csv(brainFile, delimiter= '\t')
```

```
#code cell 2
brainFrame.head()
```

	Gender\tFSIQ\tVIQ\tPIQ\tWeight\tHeight\tMRI_Cour
0	Female\t133\t132\t124\t118\t64.5\t81693
1	Male\t140\t150\t124\tNA\t72.5\t100112
2	Male\t139\t123\t150\t143\t73.3\t103843
3	Male\t133\t129\t128\t172\t68.8\t96535
4	Female\t137\t132\t134\t147\t65.0\t95154

Next steps: Generate code with brainFrame View recommended plots

```
# Code cell 3
brainFrame.describe()
```

	Gender\tFSIQ\tVIQ\tPIQ\tWeight\tHeight\tMRI
count	
unique	
top	Female\t133\t132\t124\t118\t64.5\t
freq	

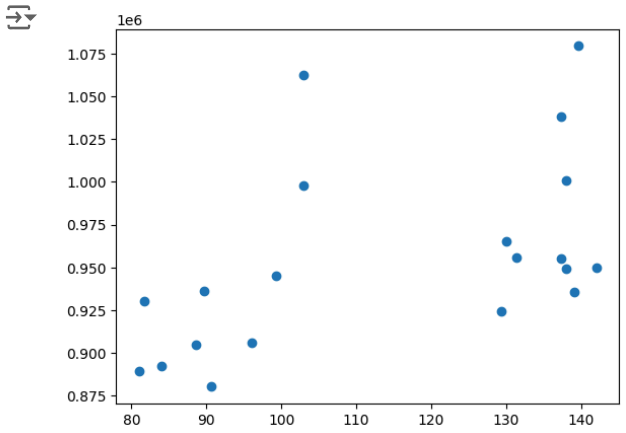
```
# Code cell 4
import numpy as np
import matplotlib.pyplot as plt
```

```
# Code cell 5
menDf = brainFrame[(brainFrame.Gender == 'Male')]
womenDf = brainFrame[(brainFrame.Gender == 'Female')]
```

```
# Code cell 6
menMeanSmarts = menDf[["PIQ", "FSIQ", "VIQ"]].mean(axis=1)
plt.scatter(menMeanSmarts, menDf["MRI_Count"])
plt.show()
%matplotlib inline
```

brainsize.txt

```
Gender  FSIQ  VIQ PIQ Weight  Height  MRI_Cou
Female  133 132 124 118 64.5  816932
Male  140 150 124 NA  72.5  1001121
Male  139 123 150 143 73.3  1038437
Male  133 129 128 172 68.8  965353
Female  137 132 134 147 65.0  951545
Female  99  90  110 146 69.0  928799
Female  138 136 131 138 64.5  991305
Female  92  90  98  175 66.0  854258
Male  89  93  84  134 66.3  904858
Male  133 114 147 172 68.8  955466
Female  132 129 124 118 64.5  833868
Male  141 150 128 151 70.0  1079549
Male  135 129 124 155 69.0  924059
Female  140 120 147 155 70.5  856472
Female  96  100 90  146 66.0  878897
Female  83  71  96  135 68.0  865363
Female  132 132 120 127 68.5  852244
Male  100 96  102 178 73.5  945088
Female  101 112 84  136 66.3  808020
Male  80  77  86  180 70.0  889083
Male  83  83  86  NA  NA  892420
Male  97  107 84  186 76.5  905940
Female  135 129 134 122 62.0  790619
Male  139 145 128 132 68.0  955003
Female  91  86  102 114 63.0  831772
Male  141 145 131 171 72.0  935494
Female  85  90  84  140 68.0  798612
Male  103 96  110 187 77.0  1062462
Female  77  83  72  106 63.0  793549
Female  130 126 124 159 66.5  866662
Female  133 126 132 127 62.5  857782
Male  144 145 137 191 67.0  949589
Male  103 96  110 192 75.5  997925
Male  90  96  86  181 69.0  879987
Female  83  90  81  143 66.5  834344
Female  133 129 128 153 66.5  948066
Male  140 150 124 144 70.5  949395
Female  88  86  94  139 64.5  893983
Male  81  90  74  148 74.0  930016
Male  89  91  89  179 75.5  935863
```




```
# Code cell 7
plt.show()
%matplotlib inline
```

```
 # Code cell 8
numeric_brainFrame = brainFrame.select_dtypes(include=[np.number])
correlation_matrix = numeric_brainFrame.corr(method='pearson')
print(correlation_matrix)
```

```
                FSIQ       VIQ       PIQ    Weight    Height  MRI_Count
FSIQ       1.000000  0.946639  0.934125 -0.051483 -0.086002   0.357641
VIQ        0.946639  1.000000  0.778135 -0.076088 -0.071068   0.337478
PIQ        0.934125  0.778135  1.000000  0.002512 -0.076723   0.386817
Weight    -0.051483 -0.076088  0.002512  1.000000  0.699614   0.513378
Height    -0.086002 -0.071068 -0.076723  0.699614  1.000000   0.601712
MRI_Count  0.357641  0.337478  0.386817  0.513378  0.601712   1.000000
```

```
# Code cell 9
numeric_womenDf = womenDf.select_dtypes(include=[np.number])
correlation_matrix = numeric_womenDf.corr(method='pearson')
print(correlation_matrix)
```

```
                FSIQ       VIQ       PIQ    Weight    Height  MRI_Count
FSIQ       1.000000  0.955717  0.939382  0.038192 -0.059011   0.325697
VIQ        0.955717  1.000000  0.802652 -0.021889 -0.146453   0.254933
PIQ        0.939382  0.802652  1.000000  0.113901 -0.001242   0.396157
Weight     0.038192 -0.021889  0.113901  1.000000  0.552357   0.446271
Height    -0.059011 -0.146453 -0.001242  0.552357  1.000000   0.174541
MRI_Count  0.325697  0.254933  0.396157  0.446271  0.174541   1.000000
```

```
 # Code cell 10
numeric_df = menDf.select_dtypes(include=[float, int])
correlation_matrix = numeric_df.corr(method='pearson')
print(correlation_matrix)
```

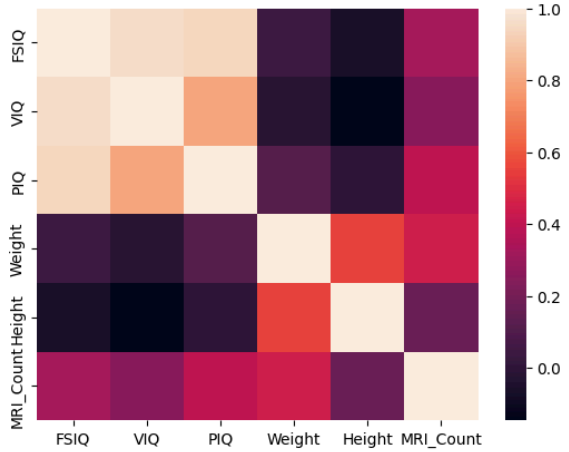
```
                FSIQ       VIQ       PIQ    Weight    Height  MRI_Count
FSIQ       1.000000  0.944400  0.930694 -0.278140 -0.356110   0.498369
VIQ        0.944400  1.000000  0.766021 -0.350453 -0.355588   0.413105
PIQ        0.930694  0.766021  1.000000 -0.156863 -0.287676   0.568237
Weight    -0.278140 -0.350453 -0.156863  1.000000  0.406542  -0.076875
Height    -0.356110 -0.355588 -0.287676  0.406542  1.000000   0.301543
MRI_Count  0.498369  0.413105  0.568237 -0.076875  0.301543   1.000000
```

```
# Code cell 11
!pip install seaborn
```

```
Requirement already satisfied: seaborn in /usr/local/lib/python3.10/dist-p
Requirement already satisfied: numpy!=1.24.0,>=1.20 in /usr/local/lib/pyth
Requirement already satisfied: pandas>=1.2 in /usr/local/lib/python3.10/di
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in /usr/local/lib/p
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/d
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.1
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/pyth
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/d
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
```

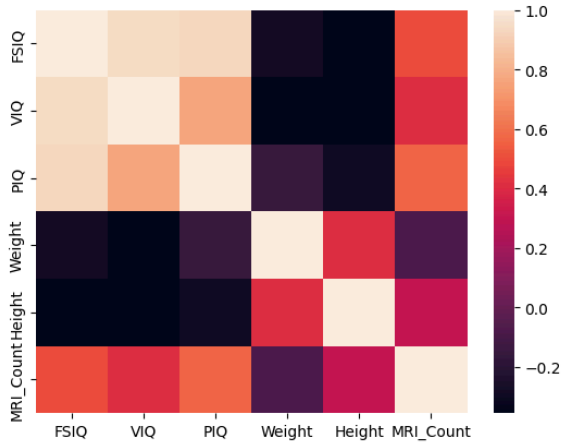
```
# Code cell 12
numeric_women_df = womenDf.select_dtypes(include=[float, int])
wcorr = numeric_women_df.corr(method='pearson')
sns.heatmap(wcorr)
```

```
<Axes: >
```




```
# Code cell 14
numeric_men_df = menDf.select_dtypes(include=[float, int])
mcorr = numeric_men_df.corr(method='pearson')
sns.heatmap(mcorr)
```

<Axes: >



Many variable pairs present correlation close to zero. What does that mean?

- The reason why there is a lot of the same variable pairs that are presented to correlation that is close to zero because

Why separate the genders?

- The reason why we need to separate the genders is that to be able to find out the difference between those in the graph and they have different results by genders so that we need to seperate the genders

What variables have stronger correlation with brain size (MRI_Count)? Is that expected? Explain.

- The Varriables that have stronger correlation with brain size is the FSIQ, VIQ, and PIQ. The reason why it has the stronger correlation due it is closer to the value of 1

*** SUPPLEMENTARY ***

```
import pandas as pd
brainFile = 'nss15.csv'
brainFrame = pd.read_csv(brainFile, delimiter= '\t')


#code cell 2
brainFrame.head()
```

	caseNumber,treatmentDate,statWeight,stratum,age,sex,race,diagnosis,body
0	1507331
1	15073472
2	1508174
3	1507177
4	15072169

```
# Code cell 3
brainFrame.describe()
```

	caseNumber,treatmentDate,statWeight,stratum,age,sex,race,diagnosis
count	
unique	
top	15(
freq	

```
# Code cell 4
import numpy as np
import matplotlib.pyplot as plt


print(brainFrame.columns)
```

```
Index(['caseNumber,treatmentDate,statWeight,stratum,age,sex,race,diagnosis
```

```
brainFrame.rename(columns=lambda x: x.strip(), inplace=True)


# Code cell 5
menDf = brainFrame[(brainFrame.sex == 'Male')]
womenDf = brainFrame[(brainFrame.sex == 'Female')]
```

```
---------------------------------------------------------------------------
-
AttributeError                            Traceback (most recent call
last)
<ipython-input-82-d9e881111f69> in <cell line: 2>()
      1 # Code cell 5
----> 2 menDf = brainFrame[(brainFrame.sex == 'Male')]
      3 womenDf = brainFrame[(brainFrame.sex == 'Female')]

/usr/local/lib/python3.10/dist-packages/pandas/core/generic.py in
__getattr__(self, name)
   5987         ):
   5988             return self[name]
-> 5989         return object.__getattribute__(self, name)
   5990
   5991     def __setattr__(self, name: str, value) -> None:
```

Next steps: Explain error

```
 # Code cell 8
numeric_brainFrame = brainFrame.select_dtypes(include=[np.number])
correlation_matrix = numeric_brainFrame.corr(method='pearson')
print(correlation_matrix)
```