Assignment 3
Seaborn and the Heatmap

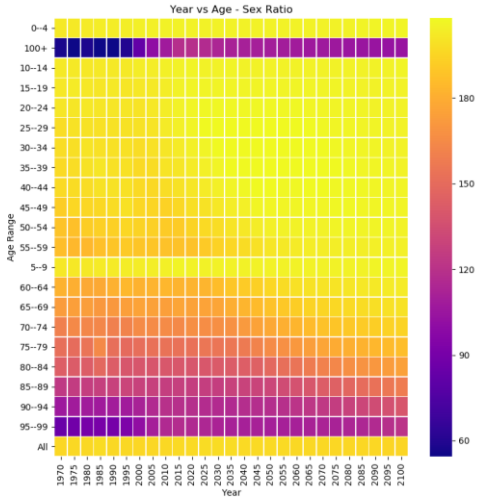
The data given to me was vast and we need to group that data inorder to plot a heatmap. There were duplicates which were important to drop and also at the same time we had to do pivoting. I have done this in two ways.

In first step I used excel and performed the pivoting and then I used python to generate the heat map.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
exf=pd.ExcelFile("f2m.xlsx")
df=exf.parse('xh')
df1=df[:]
del df1["Age"]
pd.isnull(np.array(df1, dtype=float))
cols=df.columns
plt.figure(figsize=(9,9))
ax=sns.heatmap(df1,fmt="g", linewidths=.5,yticklabels=df["Age"],cmap="plasma")
plt.title("Year vs Age - Sex Ratio")
plt.ylabel("Age Range")
plt.xlabel("Year")
plt.savefig("Save1.png")
plt.show()
```

I read the sheet as taught previously and had imported the required libraries already and then I dropped the age range column from data frame and made sure that the type of conversion is float using numpy and further I used heatmap function in which I passed the df[“Age”] as the labels and plasma as the color scheme. Added some of the details to the chart and the result was done.

I could have shown the data using annot function but it caused lot of messiness in chart.



Second approach was using the python and doing pivoting as taught in the course.

I performed groupby operation along the age and year and then used the pivot function to make the pivot as specified the code is here

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
exf=pd.ExcelFile("f2m.xlsx")
df=exf.parse('f2m_ratios')
#print(df)
df1=df.groupby(["Age","Year"],as_index=False).sum()
#print(df1)
dfu=df1.pivot("Age","Year","Ratio")
dfu.reset_index()
flat=pd.DataFrame(dfu.to_records())
print(flat["Age"])
dfli=flat[:]
del dfli["Age"]
pd.isnull(np.array(dfli, dtype=float))
print(dfli)
plt.figure(figsize=(8,8))
ax=sns.heatmap(dfli,fmt="g", linewidths=.5,yticklabels=flat["Age"],cmap="plasma")
plt.title("Year vs Age - Sex Ratio")
plt.ylabel("Age Range")
plt.xlabel("Year")
plt.savefig("Save2.png")
plt.show()
```

Instead of finding the heatmap for each area differently I took their sum as the parameter for the years and age wise so that it gets independent of the area. The results are below.