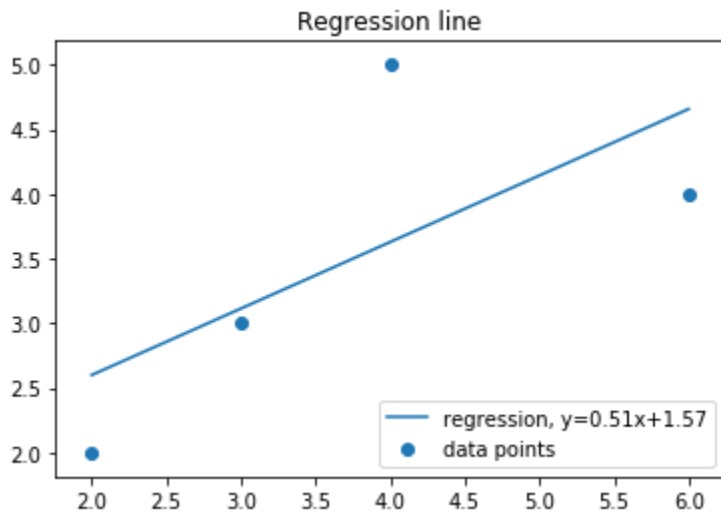# HW3 Solution for Selected Problems

Part I

```
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
import pandas as pd
import seaborn as sn

x = np.array([2,3,4,6])
y = np.array([2,3,5,4])
w, b = np.polyfit(x, y, 1)

plt.plot(x, w*x + b,label="regression, y="+str(round(w,2))+"x+"+str(round(b,2)))
plt.scatter(x,y,label="data points")
plt.title("Regression line")
plt.legend()
plt.show()
```



```
covMatrix = np.cov(x,y)
print(covMatrix)

[[2.91666667 1.5        ]
 [1.5        1.66666667]]
```
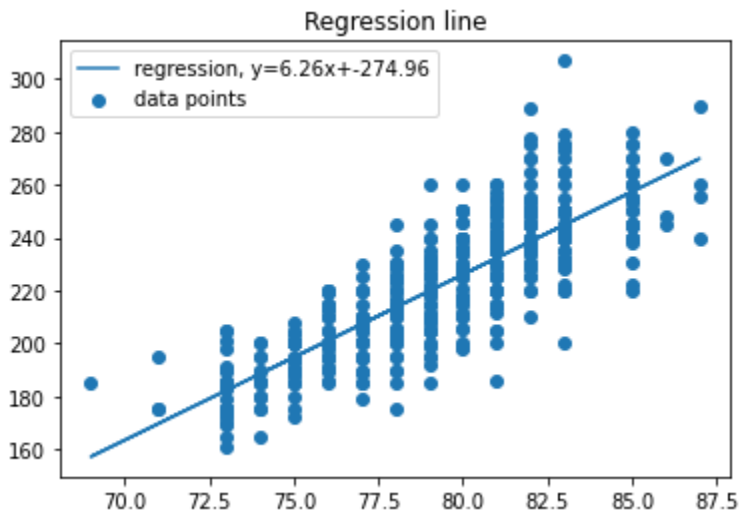
```
print(stats.pearsonr(x,y)[0])

0.680336051416609
```

```
df = pd.read_csv("./nba.csv")
y =  df["Weight"]
x =  df["Height"]
w, b = np.polyfit(x, y, 1)
plt.plot(x, w*x + b,label="regression, y="+str(round(w,2))+"x+"+str(round(b,2)))

plt.scatter(x,y,label="data points")
plt.title("Regression line")
plt.legend()
plt.show()
```



```
x = 91
print( w*x + b)
```

```
294.9809756723818
```
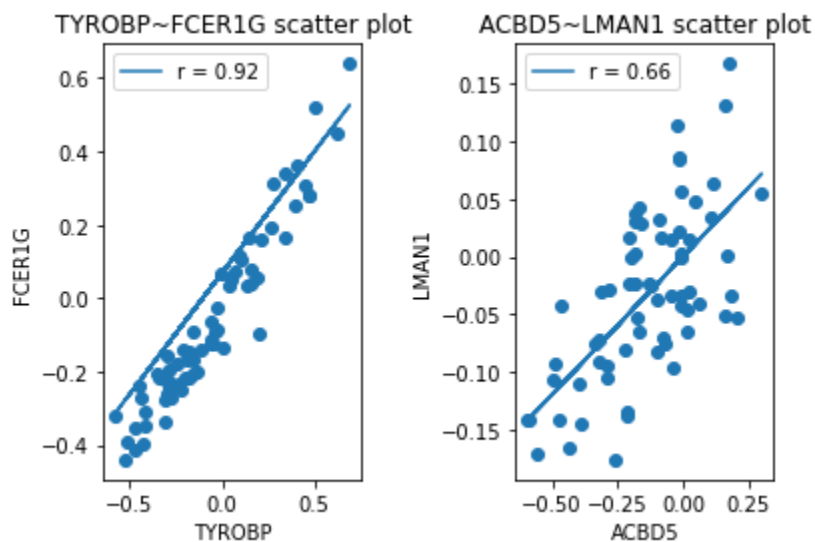
Part II

```
df = pd.read_csv("./GSE44768_CR_alz_female_reduced.csv")

TYROBPvalue = df["TYROBP"]
DOCK2value = df["DOCK2"]
GSTA4value = df["GSTA4"]

w, b = np.polyfit(TYROBPvalue, DOCK2value, 1)
r = stats.pearsonr(TYROBPvalue,DOCK2value)[0]
plt.plot(TYROBPvalue, w*TYROBPvalue + b,label="r = "+str(round(r,2)))
plt.scatter(TYROBPvalue,DOCK2value,label="data points")
plt.title("TYROBP~DOCK2 scatter plot")
plt.legend()
plt.show()

w, b = np.polyfit(TYROBPvalue, GSTA4value, 1)
r = stats.pearsonr(TYROBPvalue,GSTA4value)[0]
plt.plot(TYROBPvalue, w*TYROBPvalue + b,label="r = "+str(round(r,2)))
plt.scatter(TYROBPvalue,GSTA4value,label="data points")
plt.title("TYROBP~GSTA4 scatter plot")
plt.legend()
plt.show()
```
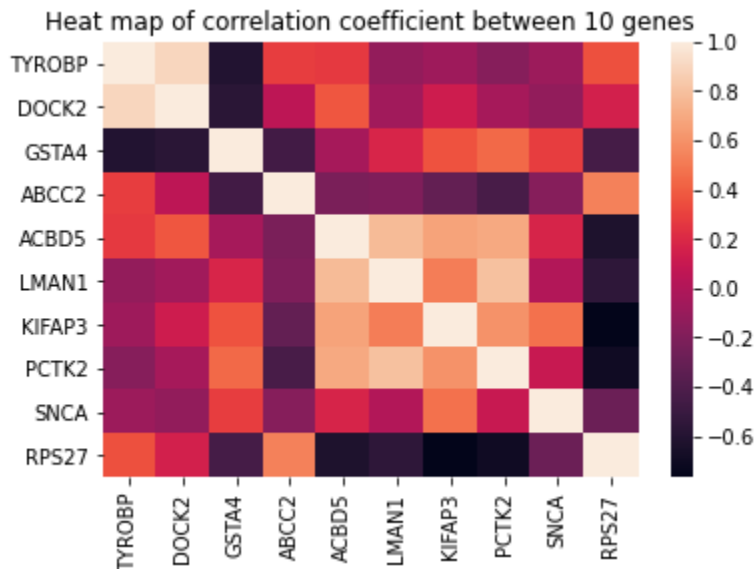
Part III

```
# genelist : TYROBP, DOCK2, GSTA4, ABCC2, ACBD5, LMAN1, KIFAP3, PCTK2, SNCA,
RPS27
df =
pd.read_csv("./GSE44768_CR_nd_female_reduced.csv",usecols=['TYROBP','DOCK2','GSTA
4',"ABCC2","ACBD5","LMAN1","KIFAP3","PCTK2","SNCA","RPS27"])
corrMatrix2 = df.corr()
print (corrMatrix2)
sn.heatmap(corrMatrix2, annot=False)
plt.title("Heat map of correlation coefficient between 10 genes")
plt.show()
```

```
fig, axes = plt.subplots(1, 2,figsize = (30,15))

sn.heatmap(corrMatrix1, annot=False, ax=axes[0])
axes[0].set_title("GSE44768 CR alz female reduced")
sn.heatmap(corrMatrix2, annot=False, ax=axes[1])
axes[1].set_title("GSE44768 CR nd female reduced")
plt.subplots_adjust(wspace=0.1)
```
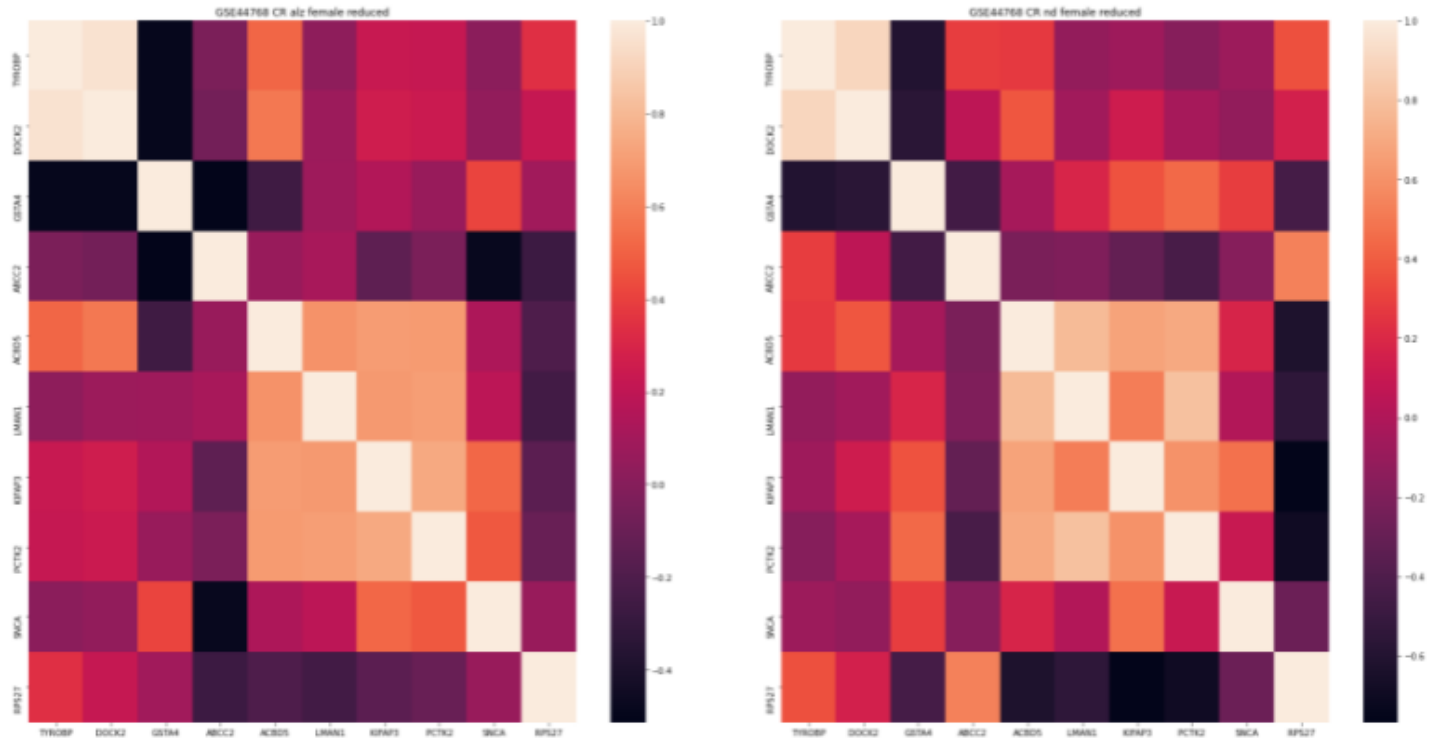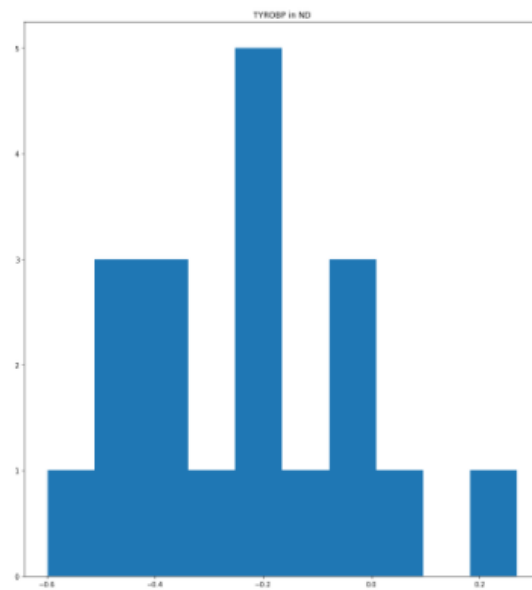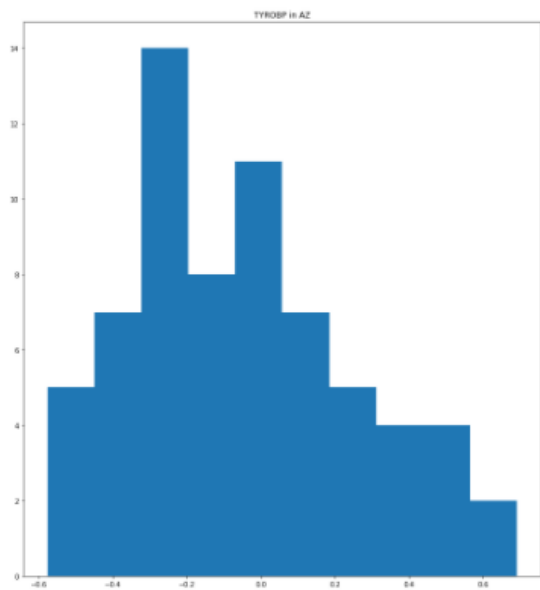

Part IV

```
df1 = pd.read_csv("./GSE44768_CR_alz_female_reduced.csv")
df2 = pd.read_csv("./GSE44768_CR_nd_female_reduced.csv")

TYROBPaz = df1["TYROBP"];
TYROBPnd = df2["TYROBP"];


fig, axes = plt.subplots(1, 2,figsize = (30,15))

axes[0].hist(TYROBPaz);
axes[0].set_title("TYROBP in AZ");
axes[1].hist(TYROBPnd);
axes[1].set_title("TYROBP in ND");
```
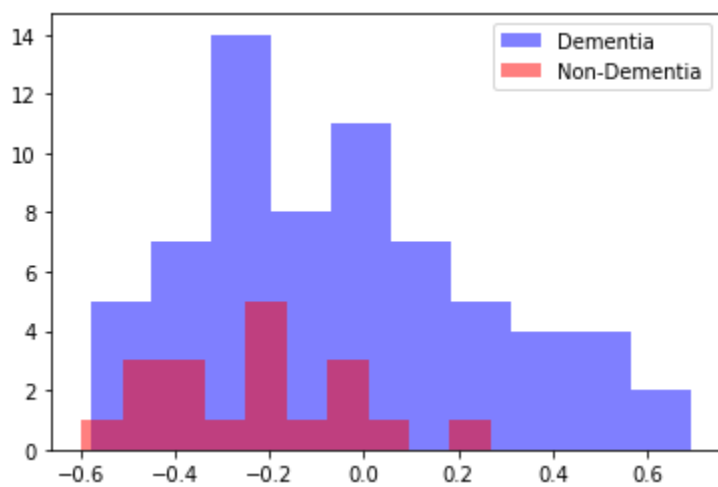
TYROBP in AZ | TYROBP in ND

```
plt.hist(TYROBPaz,alpha = 0.5,color = 'blue', label = "Dementia")
plt.hist(TYROBPnd,alpha = 0.5,color = 'red', label = "Non-Dementia")
plt.legend()
plt.show()
```

```
fig, axes = plt.subplots(1, 2,figsize = (30,15))
axes[0].violinplot(TYROBPaz);
axes[0].set_title("TYROBP in AZ");
axes[1].violinplot(TYROBPnd);
axes[1].set_title("TYROBP in ND");
```