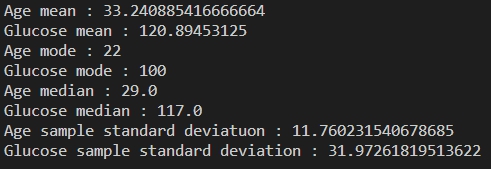
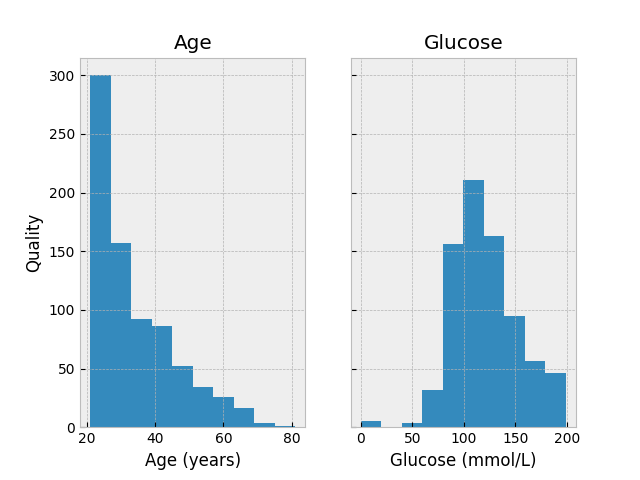
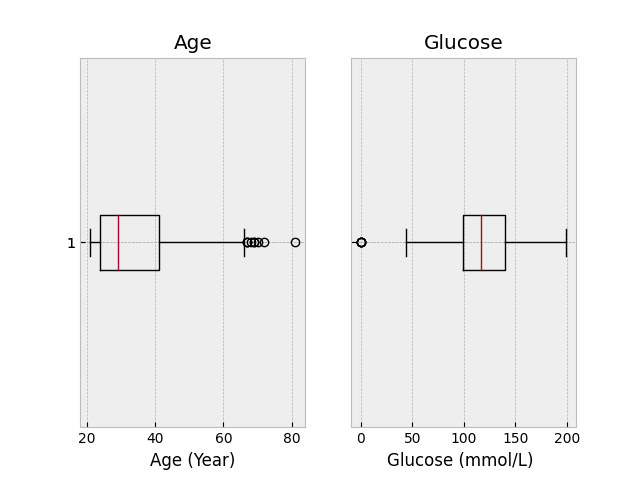
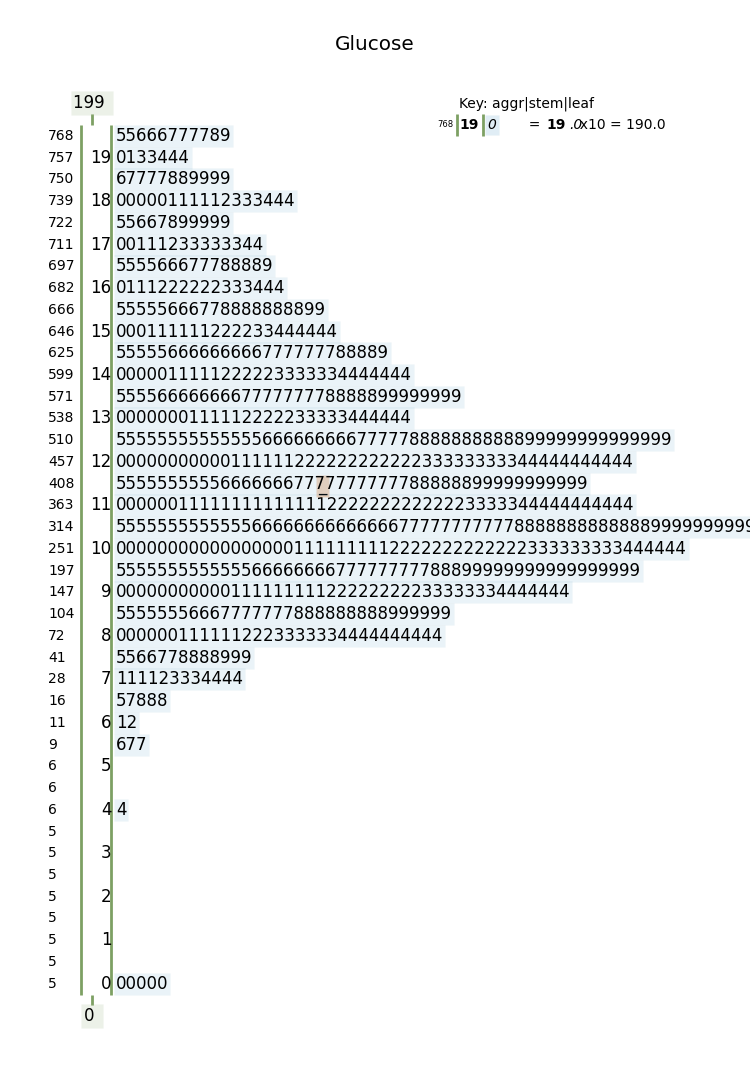
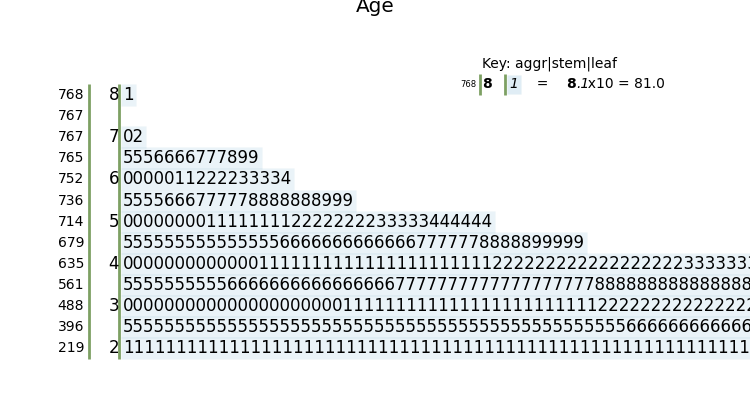
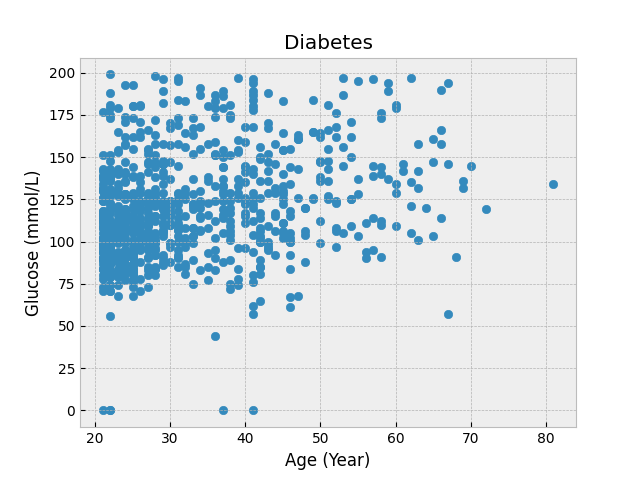
**HomeWork 2**

* ใช้ column : Age-อายุ (Year), Glucose-จำนวนน้ำตาลในเลือด (mmol/L)
* ค่าสถิติต่างๆ



* กราฟต่างๆ
* Histogram
* Box plots
* Stem and Leaf
* Scatter
* ตัวแปรแกน x : อายุของผู้ป่วย
* ตัวแปรแกน y : จำนวนน้ำตาลในเลือด
* เหตุผล
  + เหตุผลที่ใช้ อายุ และ จำนวนน้ำตาลในเลือดเป็น ตัวแปรเพราะต้องการศึกษาว่า ปริมาณน้ำตาลในเลือดเท่าไรถึงมีโอกาสเป็นโรคเบาหวาน
* Outlier
  + ตามแนวแกน x : Age -อายุ 80 ปี
  + ตามแนวแกน y : Glucose -จำนวนน้ำตาลในเลือด 0 mmol/L มีอยู่ 5 จุด
* บทวิเคราะห์ข้อมูล
  + จากข้อมูลที่ได้มาจะเห็นได้ว่า ช่วงที่เป็นโรคเบาหวานมากที่สุดจะอยู่ในช่วง 21-42 ปี และจำนวนน้ำตาลในร่างกายที่ผู้ป่วยส่วนมากมีคือ 100-140 mmol/L นี้แสดงให้เห็นว่าในช่วงอายุ 21-40ปี มีโอกาสเป็นโรคเบาหวานสูง ส่วนคนที่มีน้ำตาลในเส้นเลือดมี น้ำตาลในเส้นเลือดอยู่ระหว่าง 100-120 mmol/L ก็มีโอกาสเป็นโรคเบาหวานสูงเช่นเดียวกัน

จึงวิเคราะห์ได้ว่าอายุไม่มีผลต่อการเป็นโรคเบาหวาน ถ้าน้ำตาลในเส้นเลือดนั้นมีมาก จะทำให้มีอากาสเป็นโรคเบาหวานมากขึ้นไปด้วย

* Source Code
* import statistics as stc
* import matplotlib.pyplot as plt
* import pandas as pd
* plt.style.use('bmh')
* df = pd.read\_csv('diabetes.csv')
* # age glucose BMI
* x = df['Age']
* y = df['Glucose']
* z = df['BMI']
* # convert to list
* age = x.to\_list()
* glucose = y.to\_list()
* bmi = z.to\_list()
* def mean(data):
* print(stc.mean(data))
* def mode(data):
* print(stc.multimode(data))
* def median(data):
* print(stc.median(data))
* # # mean
* # age\_mean = stc.mean(age)
* # print('Age mean : ' + str(age\_mean))
* # glucose\_mean = stc.mean(glucose)
* # print('Glucose mean : ' + str(glucose\_mean))
* # #mode
* # age\_mode = stc.multimode(age)
* # print('Age mode : ' + str(\*age\_mode))
* # glucose\_mode = stc.mode(glucose)
* # print('Glucose mode : ' + str(glucose\_mode))
* # # median
* # age\_median = stc.median(age)
* # print('Age median : ' + str(age\_median))
* # glucose\_median = stc.median(glucose)
* # print('Glucose median : ' + str(glucose\_median))
* # Sample standard deviation
* age\_ssd = stc.stdev(age)
* print('Age sample standard deviatuon : ' + str(age\_ssd))
* glucose\_ssd = stc.stdev(glucose)
* print('Glucose sample standard deviation : ' + str(glucose\_ssd))
* # histogram
* # plt.xlabel('Age')
* # plt.ylabel('Interval')
* # plt.title('Age')
* # plt.hist(age, bins=10)
* # plt.show()
* # fig, ax = plt.subplots(1, 2, sharey=True)
* # ax[0].set\_title('Age (years)')
* # ax[0].hist(age)
* # ax[0].set\_title('Age')
* # ax[0].set\_xlabel('Age (years)')
* # ax[0].set\_ylabel('Quality')
* # ax[1].set\_title('Glucose')
* # ax[1].hist(glucose)
* # ax[1].set\_xlabel('Glucose (mmol/L)')
* # plt.show()
* # box plot
* # fig, ax = plt.subplots(1, 2, sharey=True)
* # ax[0].set\_title('Age')
* # ax[0].boxplot(age, vert=False)
* # ax[0].set\_xlabel('Age (Year)')
* # ax[1].set\_title('Glucose')
* # ax[1].boxplot(glucose, vert=False)
* # ax[1].set\_xlabel('Glucose (mmol/L) ')
* # plt.show()
* # stem and leave
* # ls = [i for i in range(1,10)]
* # print(ls)
* # fig, ax = plt.subplots(2)
* # ax[0].set\_title('Age')
* # ax[0].stem(age, ls)
* # ax[1].set\_title('Glucose')
* # ax[1].stem(glucose)
* # plt.show()
* # import stemgraphic
* # fig, ax = stemgraphic.stem\_graphic(df['Age'])
* # plt.title('Age')
* # plt.show()
* # fig, ax = stemgraphic.stem\_graphic(df['Glucose'])
* # plt.title('Glucose')
* # plt.show()
* # scatter
* # fig, ax = plt.subplots(2)
* # ax[0].set\_title('Age')
* # ax[0].scatter(x, y)
* # ax[1].set\_title('Glucose')
* # ax[1].scatter(x, y)
* # plt.xlabel('Age (Year)')
* # plt.ylabel('Glucose (mmol/L)')
* # plt.title('Diabetes')
* # plt.scatter(x, y)
* # plt.show()

นายสหทัศน์ ลีวัฒนา รหัส 62010922