

HomeWork 2

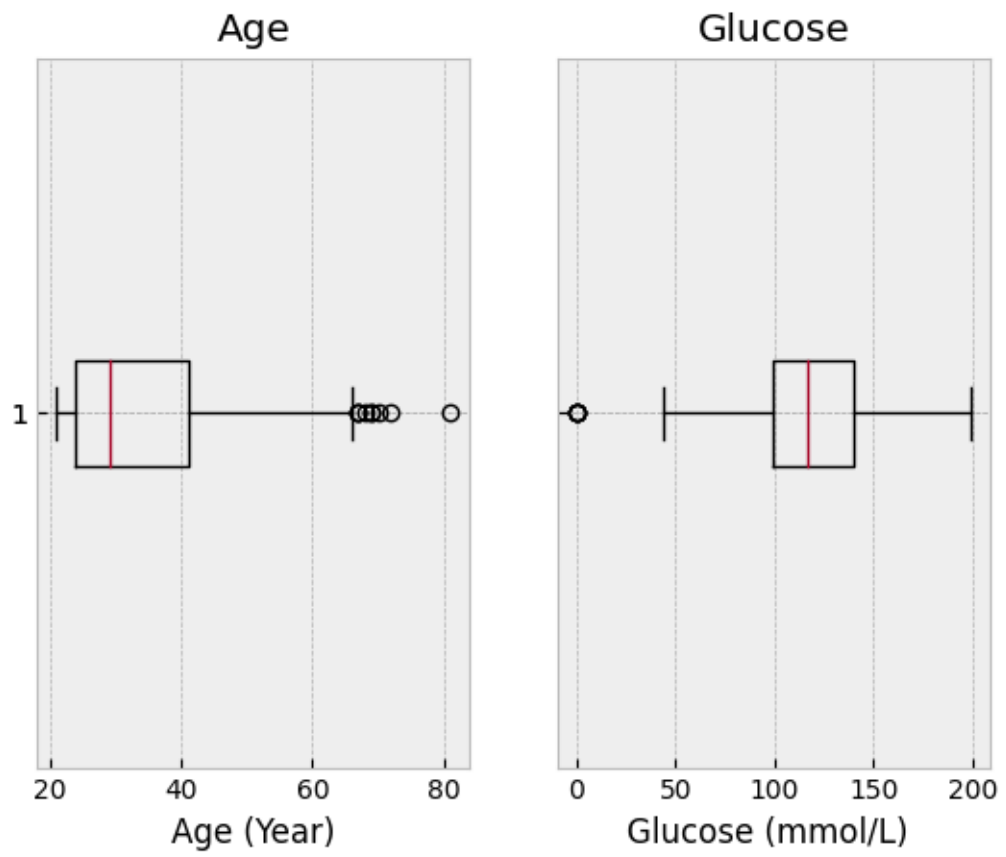
- ใช้ column : Age-อายุ (Year), Glucose-จำนวนน้ำตาลในเลือด (mmol/L)
- ค่าสถิติต่างๆ

```
Age mean : 33.240885416666664
Glucose mean : 120.89453125
Age mode : 22
Glucose mode : 100
Age median : 29.0
Glucose median : 117.0
Age sample standard deviatuon : 11.760231540678685
Glucose sample standard deviation : 31.97261819513622
```

- กราฟต่างๆ
- Histogram



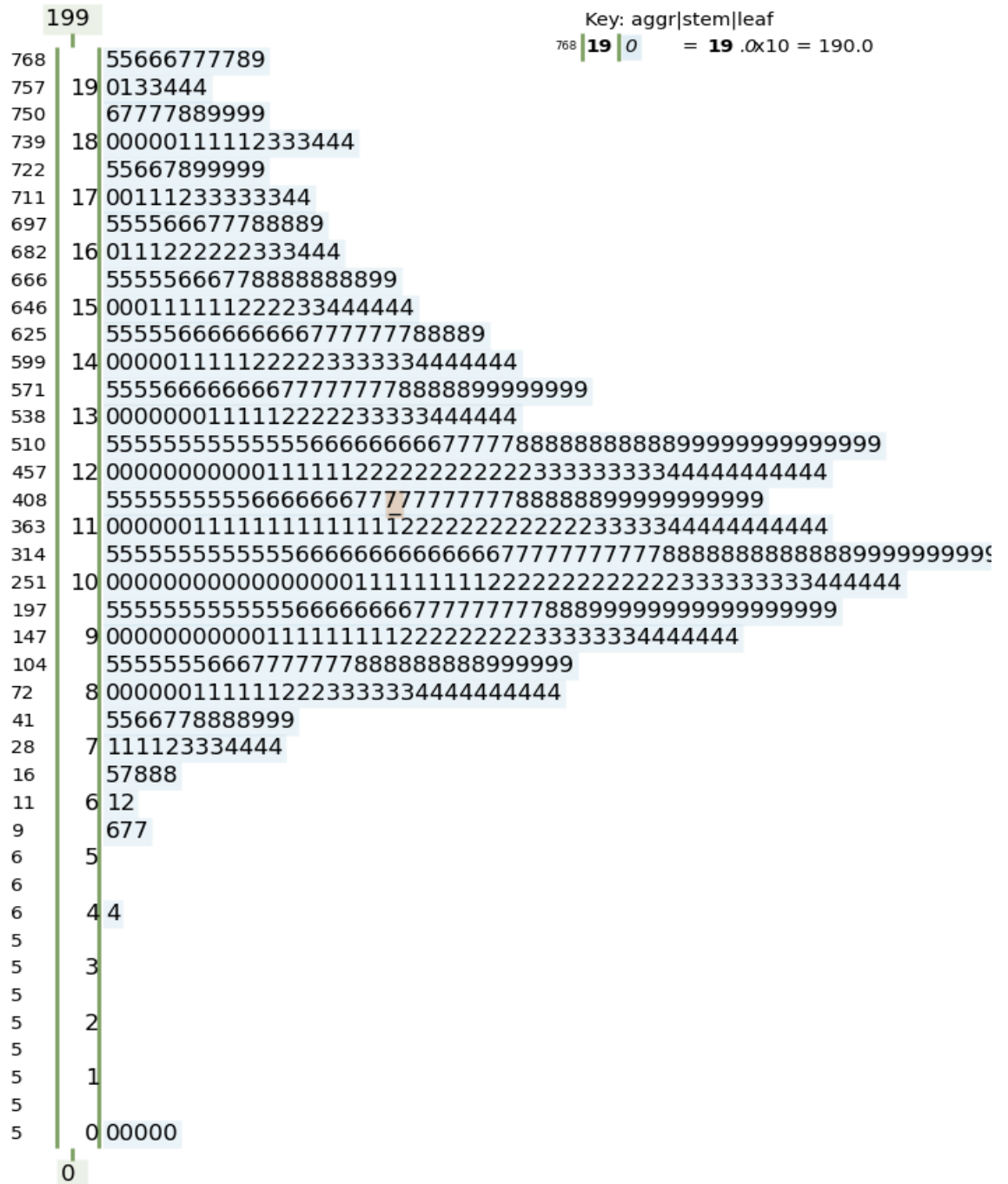
- Box plots



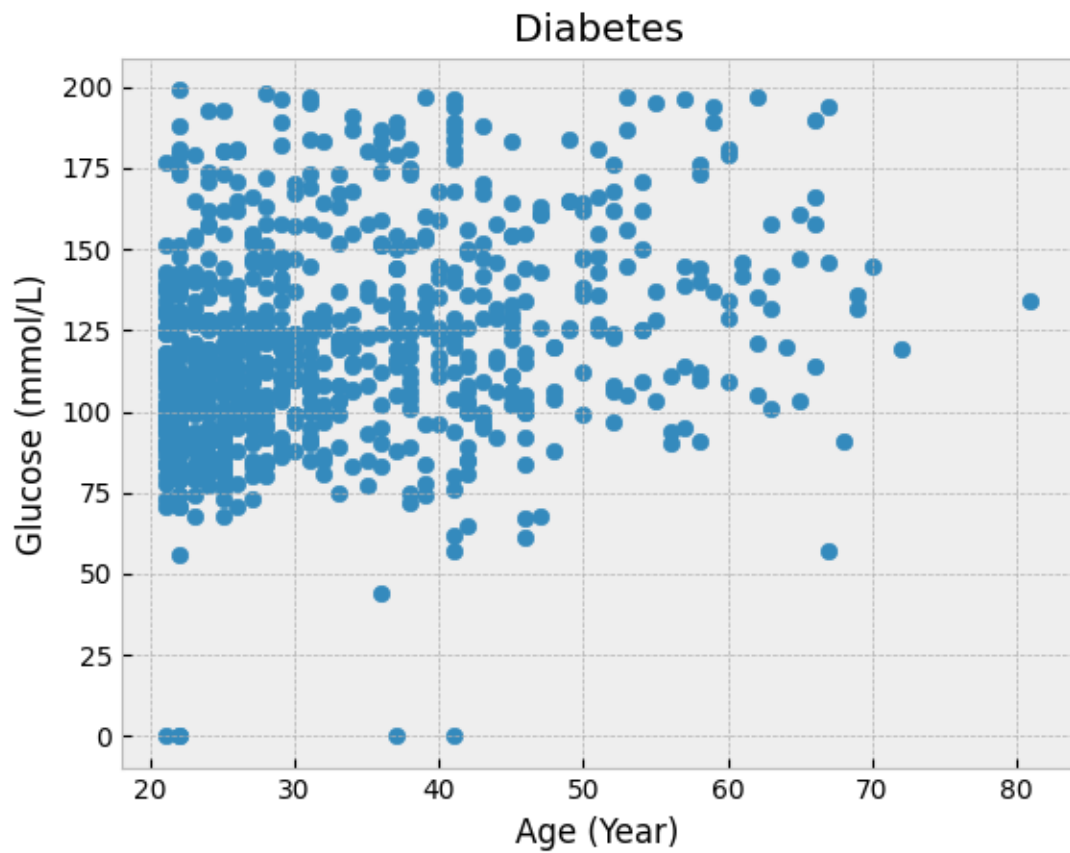
- Age

$$768 \quad \mathbf{8} \quad \mathbf{1} = 8.1 \times 10 = 81.0$$
[illegible]

Glucose



- 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 21-40 ปี มีโอกาสเป็นโรคเบาหวานสูง ส่วนคนที่มีย่านตาลในเส้นเลือดมี น้ำตาลในเส้นเลือดอยู่ระหว่าง 100-120 mmol/L ก็มีโอกาสเป็นโรคเบาหวานสูงเช่นเดียวกัน

จึงวิเคราะห์ได้ว่าอายุไม่มีผลต่อการเป็นโรคเบาหวาน ถ้าน้ำตาลในเส้นเลือดนั้นมีมาก จะทำให้มีโอกาสเป็นโรคเบาหวานมากขึ้นไปด้วย

- 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()

```