**Pandas**

**การ Plot Graph ในรูปแบบต่างๆ สำหรับการทำ Data Visualization เบื้องต้น**

การ Plot Graph เป็นการนำเสนอข้อมูลในรูปแบบภาพ เพื่อให้ง่ายต่อการทำความเข้าใจข้อมูล จากตัวอย่างต่อไปนี้เป็นการสร้างข้อมูลแบบ DataFrames ใน Pandas แล้วเก็บไว้ที่ตัวแปรชื่อ df (ที่ผ่านมานักเรียนจะโหลดข้อมูลจากไฟล์เข้าสู่ DataFrames ชื่อว่า zoo หรือ zoo\_eat เป็นต้น)

import pandas as pd

df **=** pd**.**DataFrame({

'name':['john','mary','peter','jeff','bill','lisa','jose'],

'age':[23,78,22,19,45,33,20],

'gender':['M','F','M','M','M','F','M'],

'state':['california','dc','california','dc','california','texas','texas'],

'num\_children':[2,0,0,3,2,1,4],

'num\_pets':[5,1,0,5,2,2,3]

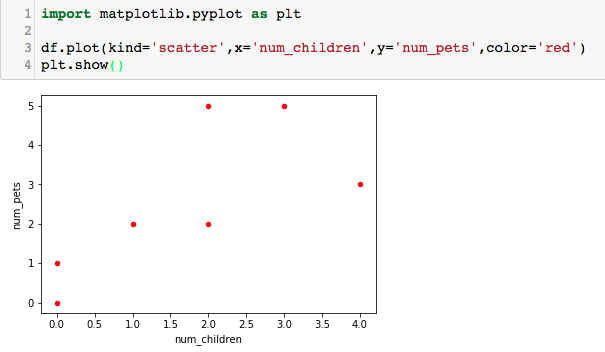
})

1. **ให้นักเรียนสร้างข้อมูลตามตัวอย่างข้างล่าง และพิมพ์ df กด run เพื่อดูข้อมูลที่สร้างขึ้นมา ตามตัวอย่างด้านล่าง (ข้อมูลจะมีทั้งหมด 6 columns และ index อีกหนึ่ง column)**

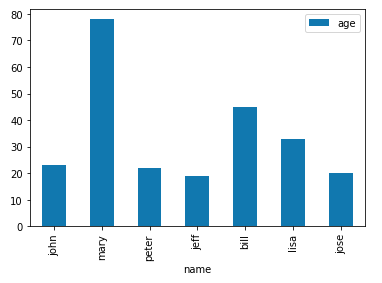


ในการ plot graph ภาษา Python จะมี Modules หลายตัวที่รองรับฟังก์ชันการ plot ข้อมูล module ที่เป็นที่นิยมคือ matplotlib ซึ่งเราต้องทำการโหลดมาก่อนที่จะเรียกใช้ฟังก์ชัน plot ต่างๆ

1. **ให้นักเรียนลองคำสั่งตามตัวอย่างด้านล่าง** (ตัวอย่างนี้เป็น plot เป็น scatter plot เพื่อดูความสัมพันธ์ระหว่างข้อมูลสองตัว ในตัวอย่างนี้คือ ดูว่าคนที่มี children เท่านี้คน จะมี pets หรือสัตว์เลี้ยงเท่าใด ข้อมูลแสดง trend หรือแนวโน้มของข้อมูลว่าเป็นอย่างไร)

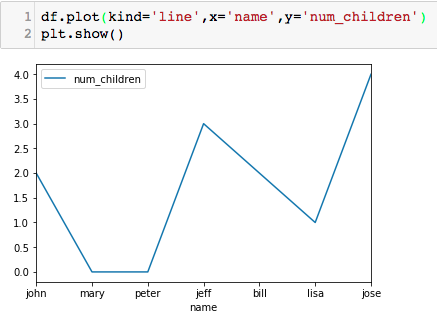


1. **ให้นักเรียนลองเปลี่ยนข้อมูลแกน x เป็น ‘age’ และ plot อีกครั้ง**
2. **ให้นักเรียนลอง plot ข้อมูลแบบ bar โดยแกน x เป็น ‘name’ และ แกน y เป็น ‘age’ (ไม่ต้องกำหนดสี และถ้านักเรียนทำถูกต้องจะได้กราฟดังต่อไปนี้)**

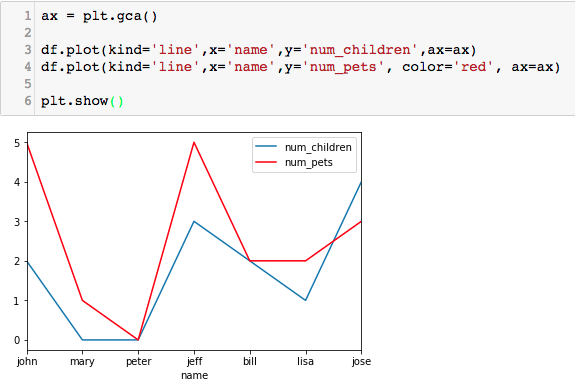
****

1. **ให้นักเรียนกำหนด kind = ‘barh’ โดยใช้ข้อมูลในแกน x และ y ตามข้อ 4 และ plot กราฟ**
2. **ให้นักเรียนลอง plot กราฟแท่งสำหรับข้อมูลอื่นๆ เช่น แต่ละคนมีลูกกี่คน หรือ มีสัตว์เลี้ยงกี่ตัว**

จากตัวอย่างที่ผ่านมานักเรียนได้เรียนรู้วิธีการ plot ข้อมูลในรูปแบบ scatter และ แบบแท่ง ต่อไปจะเป็นวิธีการ plot กราฟเส้น



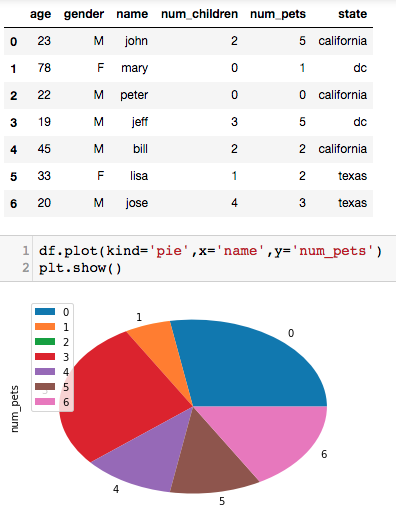
ในกรณีที่เราต้องการ plot กราฟเส้นหลายๆ เส้นพร้อมกัน ต้องทำการกำหนด argument ให้กับ matplotlib เช่น ax = plt.gca() โดยที่ gca ย่อมาจาก get current axis เป็นการระบุให้ matplotlib ใช้แกนเดิมในการ plot กราฟ



จากกราฟที่ plot ออกมาหากนักเรียนต้องการเซฟรูปก็ให้ใช้คำสั่ง plt.savefig(‘ชื่อไฟล์ที่ต้องการ’) แทน plt.show()

1. **ให้นักเรียนทำการเซฟกราฟทั้งหมดที่ได้ทำมาก่อนหน้านี้**

ต่อไปจะเป็นการ plot แบบ pie chart หลักการก็แบบเดียวกัน นักเรียนสามารถกำหนดชนิดของกราฟให้เป็น pie และระบุข้อมูลที่ต้องการ plot เช่น



กราฟ pie แสดงให้เห็นอัตราส่วนของ num\_pets สำหรับแต่ละชื่อ อย่างไรก็ตาม legend ในกราฟจะเป็นค่า index แทนที่จะเป็นชื่อคน เราจำเป็นต้องระบุชื่อให้แต่ละ index



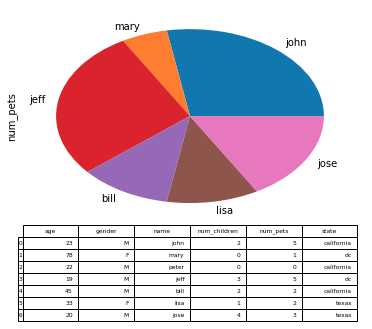
1. **ให้นักเรียนลองคำสั่งข้างและแสดงกราฟออกมา (จะเห็นได้ว่าแต่ละ pie จะมี label กำกับตามที่กำหนด)**

ในบางครั้ง legend ในปรากฏอยู่ในกราฟอาจจะบดบังตัวกราฟ เราสามารถไม่แสดง legend ได้ โดยใส่ argument ในคำสั่งเพิ่มคือ legend = False

1. **ให้นักเรียนลองใส่ legend = False ในคำสั่ง plot**

เนื่องจากการ plot กราฟตามตัวอย่างที่ผ่านมาเป็นการ plot อย่างเบื้องต้น ยังไม่สามารถเพิ่มรายละเอียดข้อมูล เช่น ใน pie ชิ้นนึงมีค่าเท่าไหร่ เพื่อให้การแสดงกราฟสามารถเข้าใจได้อย่างง่ายโดยแสดงข้อมูลใน DataFrame ควบคู่ไปด้วย เราสามารถเพิ่ม argument ในคำสั่งเพิ่มเติมคือ table = df (df คือ ตัวแปร DataFrame ที่เราได้สร้างไว้ ในกรณีที่เราโหลดข้อมูลเข้าไว้ในตัวแปร zoo การใช้ argument นี้ก็ต้องเป็น table = zoo)

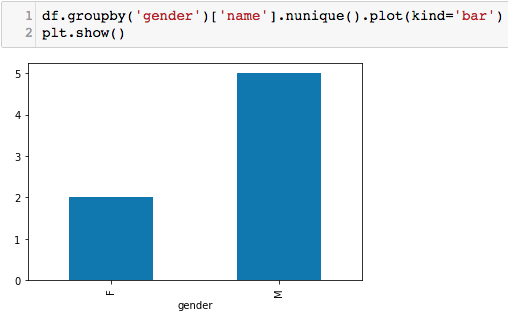
1. **ให้นักเรียนเพิ่ม table = df (หากนักเรียนทำถูกต้องจะได้ผลลัพธ์ตามภาพด้านล่าง)**

****

การปรับขนาดภาพให้ใหญ่ขึ้นสามารถทำได้โดยเพิ่ม figsize = (10,10) ค่า 10 คือขนาดที่ต้องการ

1. **ให้นักเรียนลองปรับขนาดกราฟ เพื่อให้ข้อมูลในตารางเห็นได้ชัดเจนขึ้น**

หากเราต้องการ plot กราฟของข้อมูลแบบจัดกลุ่มเราสามารถใช้คำสั่ง .groupby() เพื่อทำการจัดกลุ่มข้อมูล และกำหนดว่าจะให้ plot อะไร เช่น



จากตัวอย่างคำสั่งข้างบนเราทำการจัดกลุ่มตามเพศ และกำหนดให้ name เป็น index ที่มีค่า unique และกำหนดชนิดของกราฟเป็นกราฟแท่ง

1. **ให้นักเรียนลองเป็นกราฟแบบ pie**
2. **ให้นักเรียนลองจัดกลุ่มตาม state และ plot กราฟแบบแท่งและ pie**

ตัวอย่างต่อไปเป็นการ plot กราฟแบบ histogram เฉพาะข้อมูลใน column ใด column หนึ่ง ในตัวอย่างนี้เราเลือก column age และทำการ plot แบบ histogram โดยมี bin (หรือช่วง) ทั้งหมด 5 ช่วง อาทิ 0-19, 20-39, … และกำหนดให้ขนาดความกว้างของ bar อยู่ที่ 0.7 ซึ่งกราฟแท่งนี้ไม่ได้ระบุอายุ แต่เป็นจำนวนของผู้มีอายุที่ตกอยู่ในช่วงนั้นๆ



1. **ให้นักเรียนทำการ plot กราฟสำหรับ num\_children และ num\_pets**
2. **ให้นักเรียนโหลดข้อมูลจาก zoo และ zoo\_eats ทำการ merge ข้อมูล และ plot กราฟในรูปแบบต่างๆ จากข้อมูลที่ได้ merge มาแล้ว**