

Manipulation of library Pandas

(Djaouida Dahmani-Zaouche)

1. Write a python script which fills a file, named data.csv. This latter has the following columns: nameW (the name of the wife), yearW (her year of birth), nameH (the name of the husband), yearH (his year of birth), yearM (the year of their marriage), deptMarriage (the department of their marriage) and nbChild (number of their children). The number of lines must be equal to 2000. Your data must reflect the reality.

For each question, execute and explain.

2.

```
df = pd.read_csv("data.csv")

list=pd.concat([df.head(n=2), df.tail(n=3)])
print(list)
print(df)
```

What is the type of df?

4.

```
df1=df.groupby('yearM')
print(df1)
df1.groups
```

What means groupby?

5.

```
df2=df1.count()
print(df2)
```

What represents df2?

6.

```
df3=df1.max()
print(df3)
```

What represents df3?

7.

```
df4=df[['deptMarriage', 'yearM']].groupby('yearM').count()
print(df4)
```

what represents df4?

8.

```
df5=df4.reset_index()
```

```
print(df5)
```

```
9. current_year=2019
   df['ageH']=current_year-df['yearH']
   df['ageW']=current_year-df['yearW']
   print(df)
```

Give a sense to columns ageH and ageW.

```
10.   df['nb']=0
      df6=df[['nb','yearM']].groupby('yearM').count().reset_
      index()
      print(df6)
      Give a sense to columns nb.
```

```
10.
max=0
year=0
for index, row in df6.iterrows():
    if (max< row["nb"]):
        max=row["nb"]
        year=row["yearM"]
print ('(' ,max, ',' ,year, ')')
```

Give a sense to the couple displayed.

11. Write a piece of code which gives, for each year appearing in the marriage date column, the average of children of children whose parents were married in year y1.

12. Display graphically the following relations:

- (average_number_children, year)
- (number_of_marriages, year)
- (number_of_couples, difference_age)

Try to use different kinds of charts and customize them.

