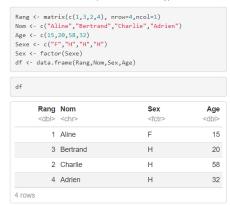
DATAFRAME TUTO

Maxime & Lucas

I. Créer une dataframe R vs Python

Comment créer un tableau qui contient différents type de données.



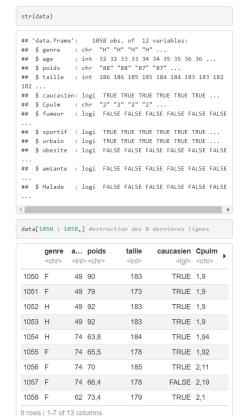


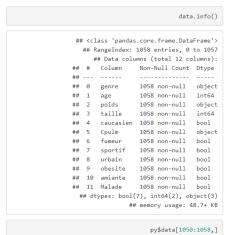
II. Traitement de données

1. Import de Dataframe

data<-read.csv("donnees.csv",header=TRUE,encoding="UTF-8")					<pre>data=pd.read_csv("donnees.csv",encoding="UTF-8")</pre>				
genre	age poids	taille caucasien	Cpulm	fumeur	sportif	urbain	obesite	amiante	Malade
Н	32 88	186 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
Н	32 88	186 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
Н	33 87	185 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
Н	33 87	185 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
F	34 86	184 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
F	34 86	184 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE

2. Structure de données





	-	a poids			
	<chr></chr>	<dbl><chr></chr></dbl>	<dbl></dbl>	< g >	<chr></chr>
1050	F	49 90	183	TRUE	1,9
1051	F	49 79	173	TRUE	1,9
1052	Н	49 92	183	TRUE	1,9
1053	Н	49 92	183	TRUE	1,9
1054	Н	74 63,8	184	TRUE	1,94
1055	F	74 65,5	178	TRUE	1,92
1056	F	74 70	185	TRUE	2,11
1057	F	74 66,4	178	FALSE	2,19
1058	F	62 73,4	179	TRUE	2,1
			9) rows 1-7 of	13 colur

rgzvrvtetertbzetrztnet

3. Résumé statistiques

summary(data)#permet d'obtenir un resume statistiques									
	##	genre	age	poids	taille				

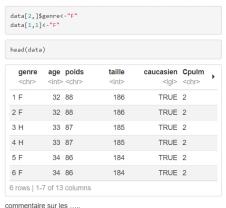
```
Class :character 1st Qu.: 61.0 Class :character 1st Qu.:173.0
## Mode :character Median : 66.0 Mode :character Median :178.0
##
                     Mean : 63.2
                                                      Mean :176.9
                    . 69.0
max. :110.0
Cpulm
                     3rd Qu.: 69.0
                                                      3rd Qu.:182.0
                                                     Max. :192.0
sportif
##
##
   caucasien
## Mode :logical Length:1058
                                    Mode :logical Mode :logical
## FALSE:194
                 Class :character FALSE:668
                                                   FALSE: 379
## TRUE :864
                  Mode :character TRUE :390
##
##
     urbain
                   obesite
##
                                   amiante
                                                   Malade
## Mode:logical Mode:logical Mode:logical Mode:logical
                  FALSE:1041
                                 FALSE:1046
## FALSE:233
                                                 FALSE:636
   TRUE :825
##
data.describe()
## count 1058.000000 1058.000000
          63.196597
                      176.942344
## std
           9.452020
                        5.954215
## min
           32.000000
                      147.000000
## 25%
           61.000000
                      173 000000
## 50%
          66.000000
                     178.000000
## 75%
           69.000000
                      182.000000
## max
          110.000000
                      192,000000
data["sportif"].value_counts()
## True
           679
## Name: sportif, dtype: int64
def count(val):
   print(data[i].value_counts())
count(["fumeur","sportif"])
## False
## True
          390
## Name: fumeur, dtype: int64
        679
## True
## False
          379
## Name: sportif, dtype: int64
```

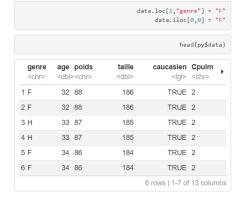
Min. :147.0

Min. : 32.0 Length:1058

Length:1058

4. MODIFICATION D'UNE VALEUR DE LA TABLE DE DONNEES





5. AJOUT D'UNE OU PLUSIEURS COLONNES A LA TABLE DE DONNEES



6. SUPPRESSION D'UNE COLONNE

genre	age poids	taille caucasien	Cpulm	fumeur	sportif	urbain	obesite	amiante	Malade
F	32 88	186 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
F	32 88	186 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
Н	33 87	185 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
Н	33 87	185 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
F	34 86	184 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
F	34 86	184 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE