DATAFRAME TUTO

Maxime & Lucas

I. Créer un dataframe R vs Python

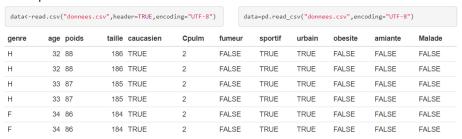
Comment créer un tableau qui contient différents type de données.



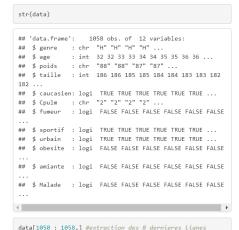
Nom=["Aline","Bertrand","Charlie","Adrien"] Sexe=["F","H","H"] Age=[15,20,58,32] df=pd.DataFrame({"Nom":Nom,"Sexe":Sexe,"Age":Age})							
hea	d(py\$df)						
	Nom	Sexe	Age				
	NOITI	Jeke	Age				
	<chr></chr>	<chr></chr>	<dbl></dbl>				
1			•				
1 2	<chr></chr>	<chr></chr>	<dbl></dbl>				
	<chr></chr>	<chr></chr>	<dbl></dbl>				

II. Traitement de données

1. Import de Dataframe



2. Structure de données





data.info()								
## <class 'pandas.core.frame.dataframe'=""></class>								
##	## RangeIndex: 1058 entries, 0 to 1057							
##	Data	columns (t	otal 12 columns)	:				
##	#	Column	Non-Null Count	Dtype				
##								
##	0	genre	1058 non-null	object				
##	1	age	1058 non-null	int64				
##	2	poids	1058 non-null	object				
##	3	taille	1058 non-null	int64				
##	4	caucasien	1058 non-null	bool				
##	5	Cpulm	1058 non-null	object				
##	6	fumeur	1058 non-null	bool				
##	7	sportif	1058 non-null	bool				
##	8	urbain	1058 non-null	bool				
##	9	obesite	1058 non-null	bool				
##	10	amiante	1058 non-null	bool				
##	11	Malade	1058 non-null	bool				
##	dtyp	es: bool(7)	, int64(2), obje	ect(3)				
##	memo	ry usage: 4	8.7+ KB					

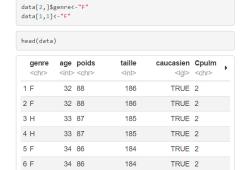
#data[1049:1058] py\$data[1050:1058,]

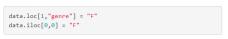
	genre <chr></chr>	a poids dbl><chr></chr>	taille <dbl></dbl>	caucasien < g >	Cpulm <chr></chr>
1050	F	49 90	183	TRUE	1,9
1051	F	49 79	173	TRUE	1,9
1052	Н	49 92	183	TRUE	1,9
1053	Н	49 92	183	TRUE	1,9
1054	Н	74 63,8	184	TRUE	1,94
1055	F	74 65,5	178	TRUE	1,92
1056	F	74 70	185	TRUE	2,11
1057	F	74 66,4	178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

3. Résumé statistiques

summary(data)#permet d'obtenir un resume statistiques

```
## genre age poids
## Length:1058 Min. : 32.0 Length:1058
## Class :character 1st Qu.: 61.0 Class :charac
                                                                                                                           taille
Min. :147.0
                                                                                                                           1st Qu.:173.0
                                                                                   Class :character
  ## Mode :character Median : 66.0
## Mean : 63.2
                                                                                                                           Median :178.0
Mean :176.9
                                                                                   Mode :character
  ##
                                                  3rd Qu.: 69.0
                                                                                                                            3rd Qu.:182.0
  ##
                                                Max. :110.0
Cpulm
                                                                                                                        Max. :192.0 sportif
  ## caucasien
  ## Mode :logical Length:1058
## FALSE:194 Class :charac
                                                                                Mode :logical Mode :logical
                                           Class :character FALSE:668
  ## TRUE :864
                                           Mode :character TRUE :390
                                                                                                                     TRUE :679
  ##
              urbain
                                              obesite
                                                                                amiante
                                                                                                                   Malade
  ## Mode:logical Mo
   ## TRUE :825
                                            TRUE :17
                                                                             TRUE :12
                                                                                                               TRUE :422
  ##
  ##
  \textbf{def} \ \mathsf{count}(\mathsf{val}) \colon
                                                                                                                                            data["sportif"].value_counts()
      for i in val
           print(data[i].value_counts())
                                                                                                                                            ## True
  count(["fumeur","sportif"])
                                                                                                                                            ## Name: sportif, dtype: int64
  ## False 668
  ## True
                          390
  ## Name: fumeur, dtype: int64
                     679
  ## True
   ## False
                         379
   ## Name: sportif, dtype: int64
  data.describe()
  ## count 1058.000000 1058.000000
  ## mean 63.196597
                                                  176.942344
  ## std
                            9 452020
                                                       5 954215
  ## min
                           32.000000
                                                     147.000000
  ## 25%
                           61.000000
                                                    173.000000
  ## 50%
                           66.000000 178.000000
                           69.000000
                                                  182.000000
  ## max
                        110.000000 192.000000
  data.info()
  ## <class 'pandas.core.frame.DataFrame'>
  ## RangeIndex: 1058 entries, 0 to 1057
  ## Data columns (total 12 columns):
  ## # Column
                                      Non-Null Count Dtype
                                     1058 non-null object
  ## 0 genre
  ## 1 age
                                         1058 non-null int64
  ## 2
                  poids
                                         1058 non-null
                                                                           object
                                         1058 non-null int64
  ## 3 taille
                  caucasien 1058 non-null
  ## 5 Cpulm 1058 non-null
## 6 fumeur 1058 non-null
                                                                           object
                                          1058 non-null
                                                                           bool
  ## 7
                  sportif
                                         1058 non-null
                                                                           boo1
                                         1058 non-null
  ## 8 urbain
                                                                           bool
                  obesite
                                         1058 non-null
  ## 10 amiante 1058 non-null
                                                                           bool
  ## 11 Malade
                                         1058 non-null
  ## dtypes: bool(7), int64(2), object(3)
  ## memory usage: 48.7+ KB
4. Modification d'une valeur de la table
```





head(pv\$data)

genre <chr></chr>	age poids <dbl><chr></chr></dbl>	taille <dbl></dbl>	caucasien < g >	Cpulm <chr></chr>
1 F	32 88	186	TRUE	2
2 F	32 88	186	TRUE	2
3 H	33 87	185	TRUE	2
4 H	33 87	185	TRUE	2
5 F	34 86	184	TRUE	2
6 F	34 86	184	TRUE	2
6 rows 1-	7 of 13 columns			

5. Ajout d'une colonne

6 rows | 1-7 of 13 columns

taille_m = data\$tai	lle/100	
n_data<-cbind(data,	taille m)	
#head(n data)		
wnedd(n_ddcd)		

<pre>taille_m = data["taille"]/100 data["taille_m"] = taille_m #data.head()</pre>

genre	age poids	taille caucasien	Cpulm	fumeur	sportif	urbain	obesite	amiante	Malade	taille_m
F	32 88	186 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	1.86

F	32 88	186 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	1.86
Н	33 87	185 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	1.85
Н	33 87	185 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	1.85
F	34 86	184 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	1.84
F	34 86	184 TRUE	2	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	1.84

6. Suppression d'une colonne

	1.1								
n_data <-subset(n_data, select=-c(taille_m))					ta = data.	drop(["tail	le_m"], axi	s=1)	
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