

TP N°1

Introduction à R

Objectif du TP :

- Savoir comment importer un fichier de données.
- Savoir comment afficher le contenu d'un fichier de données.
- Savoir comment afficher les données filtrées.
- Savoir comment récapituler des données et les afficher dans un tableau
- Savoir comment afficher un graphique.

Etapes du TP

1- Ensemble de données (le fichier fromage.txt)

L'ensemble de données contient les informations sur les différents vols d'un aéroport.

Year	Month	DayofMonth	DayOfWeek	DepTime	CRSDepTime	ArrTime	CRSArrTime	UniqueCarrier	FlightNum	Ta
2008	1	3	4	2003,1955,2211,2225	WN,335,N712SW	128,150,116,-14,8	IAD,TPA	810,4,8,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	754,735,1002,1000	WN,3231,N772SW	128,145,113,2,19	IAD,TPA	810,5,10,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	628,620,804,750	WN,448,N428WN	96,90,76,14,8	IND,BWI	515,3,17,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	926,930,1054,1100	WN,1746,N612SW	88,90,78,-6,-4	IND,BWI	515,3,7,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	1829,1755,1959,1925	WN,3920,N464WN	90,90,77,34,34	IND,BWI	515,3,10,0,,0	2,0,0,0,32	
2008	1	3	4	1940,1915,2121,2110	WN,378,N726SW	101,115,87,11,25	IND,JAX	688,4,10,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	1937,1830,2037,1940	WN,509,N763SW	240,250,230,57,67	IND,LAS	1591,3,7,0,,0	10,0,0,0,47	
2008	1	3	4	1039,1040,1132,1150	WN,535,N428WN	233,250,219,-18,-1	IND,LAS	1591,7,7,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	617,615,652,650	WN,11,N689SW	95,95,70,2,2	IND,MCI	451,6,19,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	1620,1620,1639,1655	WN,810,N648SW	79,95,70,-16,0	IND,MCI	451,3,6,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	706,700,916,915	WN,100,N690SW	130,135,106,1,6	IND,MCO	828,5,19,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	1644,1510,1845,1725	WN,1333,N334SW	121,135,107,80,94	IND,MCO	828,6,8,0,,0	8,0,0,0,72	
2008	1	3	4	1426,1430,1426,1425	WN,829,N476WN	60,55,39,1,-4	IND,MDW	162,9,12,0,,0	NA,NA,NA,NA,NA	NA
2008	1	3	4	715,715,720,710	WN,1016,N765SW	65,55,37,10,0	IND,MDW	162,7,21,0,,0	NA,NA,NA,NA,NA	NA

2- Import des données des véhicules (2008.csv)

```
myDf<-read.csv("I:\\chemin du fichier/2008.csv")
```

myDf est le nom de la variable qui contient les données du fichier 2008.csv.

read.csv: l'instruction qui permet de charger les données stockées dans un fichier .csv

"I:\\chemin du fichier/2008.csv" chemin complet et nom du fichier.

Remarque: commencer par introduire le nom du disque (ici "I"), ensuite ajouter ":\\", puis introduire le chemin en séparant les dossiers par slash "/"

3- Voir les 6 premières lignes (vols) : `head(myDf)`

4- Voir les 6 premières valeurs de l'attribut **Origin** : `head(myDf$Origin)`

5- Voir les 6 dernières valeurs de l'attribut **Origin** : `tail(myDf$Origin)`

- 6- **Voir quelles sont les 6 premiers vols tels que *Origin* égal à "IND" :**
`head(myDf$Origin=="IND")`
- 7- **Total des vols ayant pour origine l'aéroport de la ville "IND" (*Origin*= "IND") :**
`sum(myDf$Origin=="IND")`
- 8- **Total des vols ayant pour destination l'aéroport de la ville "IND" (*Dest*= "IND") :**
`sum(myDf$Dest=="IND")`
- 9- **Total des vols ayant pour origine l'aéroport de la ville "ORD" (*Origin*= "ORD") :**
`sum(myDf$Origin=="ORD")`
- 10- **Total des vols ayant pour destination l'aéroport de la ville "ORD" (*Dest*= "ORD") :**
`sum(myDf$Dest=="ORD")`
- 11- **Affiche le help de la fonction AND : ??AND**
- 12- **Total des vols tels que *Origine* égale "IND" et *Destination* égale "ORD" :**
`sum(myDf$Origin=="IND" & myDf$Dest=="ORD")`
- 13- **Créer un sous-ensemble des vols ayant pour origine "IND" :**
`myIndyOrigins<-subset(myDf,myDf$Origin=="IND")`
- 14- **Créer un sous-ensemble des vols ayant pour destination "ORD" :**
`myIndyDest<-subset(myDf,myDf$Dest=="IND")`
- 15- `head(myIndyOrigins)`
- 16- `head(myIndyDest)`
- 17- **Créer une table qui récapitule le total des vols par mois : `table(myIndyOrigins$Month)`**
- 18- **Afficher le graphique de la table des vols par mois : `plot(table(myIndyOrigins$Month))`**
- 19- `sum(myDf$Origin=="TUP")`
- 20- **Donner la moyenne du nombre de vols ayant pour origine la ville "TUP" :**
`mean(myDf$Origin=="TUP")`
- 21- `myTUPOrigins<-subset(myDf, myDf$Origin=="TUP")`
- 22- `mean(myTUPOrigins$DepDelay)`
- 23- **Donner le total des vols dont l'heure de départ est inférieure à 6h00min sans tenir compte les vols dont l'heure de départ n'est pas mentionnée (*na.rm*=TRUE) :**
`sum(myIndyOrigins$DepTime<600, na.rm=TRUE)`
- 24- `sum(myIndyOrigins$DepTime<1200, na.rm=TRUE)`
- 25- `sum(myIndyOrigins$DepTime<1800, na.rm=TRUE)`
- 26- `sum(myIndyOrigins$DepTime<2400, na.rm=TRUE)`
- 27- **Donner le nombre des vols dont l'heure de départ n'est pas mentionnée (égale NA):**
`sum(is.na(myIndyOrigins$DepTime))`
- 28- `sum(myDf$Dest=="LAX")`
- 29- `sum(myDf$Origin=="ATL" & myDf$Dest=="LAX")`
- 30- `myATLandLAX<-subset(myDf, myDf$Origin=="ATL" & myDf$Dest=="LAX")`
- 31- `sum(myATLandLAX$DepTime<1200, na.rm=TRUE)`
- 32- `table(myATLandLAX$DepTime)`
- 33- `plot(table(myATLandLAX$DepTime))`