TP N°0

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 2008.csv)**

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

# 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 “/”

1. **Voir les 6 premières lignes (vols) :** *head(myDf)*
2. **Voir les 6 premières valeurs de l’attribut *Origin* :** *head(myDf$Origin)*
3. **Voir les 6 dernières valeurs de l’attribut *Origin* :** *tail(myDf$Origin)*
4. **Voir quelles sont les 6 premiers vols tels que *Origin égal à*** *"****IND****"* **:**

head(myDf$Origin=="IND")

## Total des vols ayant pour origine l’aéroport de la ville "IND" (Origin= "IND") :

sum(myDf$Origin=="IND")

1. ***Total des vols ayant pour* destination *l’aéroport de la ville*** *"****IND****"* ***(Dest=*** *"****IND****")* **:**

sum(myDf$Dest=="IND")

## Total des vols ayant pour origine l’aéroport de la ville "ORD" (Origin= "ORD") :

sum(myDf$Origin=="ORD")

1. ***Total des vols ayant pour* destination *l’aéroport de la ville*** *"****ORD****"* ***(Dest=*** *"****ORD****")* **:**

sum(myDf$Dest=="ORD")

# Affiche le help de la fonction AND : *??AND*

## Total des vols tels que Origine égale "IND" et Destination égale "ORD" :

sum(myDf$Origin=="IND" & myDf$Dest=="ORD")

# Créer un sous-ensemble des vols ayant pour origine "IND" :

myIndyOrigins<-subset(myDf,myDf$Origin=="IND")

# Créer un sous-ensemble des vols ayant pour destination "ORD" :

myIndyDest<-subset(myDf,myDf$Dest=="IND")

1. *head(myIndyOrigins)*
2. *head(myIndyDest)*
3. **Créer une table qui récapitule le total des vols par mois :** *table(myIndyOrigins$Month)*
4. **Afficher le graphique de la table des vols par mois :** *plot(table(myIndyOrigins$Month))*
5. *sum(myDf$Origin=="TUP")*

# Donner la moyenne du nombre de vols ayant pour origine la ville "TUP" :

mean(myDf$Origin=="TUP")

1. *myTUPOrigins<-subset(myDf, myDf$Origin=="TUP")*
2. *mean(myTUPOrigins$DepDelay)*
3. **Donner le total des vols dont l’heure de départ est inférieur à 6h00min sans tenir compte les vols dont l’heure de départ n’est pas mentionnée (na.rm=TRUE) :** *sum(myIndyOrigins$DepTime<600, na.rm=TRUE)*
4. *sum(myIndyOrigins$DepTime<1200, na.rm=TRUE)*
5. *sum(myIndyOrigins$DepTime<1800, na.rm=TRUE)*
6. *sum(myIndyOrigins$DepTime<2400, na.rm=TRUE)*

# Donner le nombre des vols dont l’heure de départ n’est pas mentionnéee (égale NA):

sum(is.na(myIndyOrigins$DepTime))

1. *sum(myDf$Dest=="LAX")*
2. *sum(myDf$Origin=="ATL" & myDf$Dest=="LAX")*
3. *myATLandLAX<-subset(myDf, myDf$Origin=="ATL" & myDf$Dest=="LAX")*
4. *sum(myATLandLAX$DepTime<1200, na.rm=TRUE)*
5. *table(myATLandLAX$DepTime)*
6. *plot(table(myATLandLAX$DepTime))*