Raport\_Projet\_AcLab

Dahiez Burdy

20/05/2021

# Introduction

Sur ce tp, nous allons annalyser des données météo enregistré par une sonde météorologique conçu par la filière IOT Maker, ces données sont constitué, de la date et heure, l’humidité, pression, temperature, intensité lumineuse et la pluie.avec ces données nous allons étudier quelles sont les facteurs qui annoncent la pluie. comme par exemple si le taux d’humidité augmente avant qu’il pleuve.

* Hypothèse :

Nous émetons l’hypothèse que la diminution de pression est un facteur annonciateur de la pluie.

Nous émetons l’hypothèse que le facteur d’humidité est un facteur annonciateur de la pluie plus conséquent selon la saison.

Nous émetons l’hypothèse que la température diminue avant l’arrivé de la pluie.

Nous émetons l’hypothèse que si nous detectons la présence de la pluie moins de 10 minutes avant la prédiction, la probabilitée qu’il pleuve augmente.

Nous émétons l’hypothèse que la nuit est un facteur annonciateur de la pluie.

Est il possible de creer un algorythme de prédiction d’apparition de la pluie ayant une precision de plus de 0.8 ?

# Analyse de la Pression

Pour etre sûr que cette hypothése est intéréssante nous allons recherché la corrélation entre la Pression atmosphérique et l’apparition de pluie .Nous voulons déja savoir si la temperature à une corrélation avec la pressions athmosphérique . Pour ce faire nous executons ces commandes suivantes :

cor.test(data\_meteo$pressure , data\_meteo$temperature) #-0.31

##   
## Pearson's product-moment correlation  
##   
## data: data\_meteo$pressure and data\_meteo$temperature  
## t = -90.622, df = 72255, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.3259972 -0.3129026  
## sample estimates:  
## cor   
## -0.3194652

la corrélation entre la température et la pression est de -0.31 avec un p-value très faible , la temperarute à donc une influence sur la pression . Sachant que le cycle du soleil joue un grand rôle sur la température relevé nous décidons de ne traiter les données effectué entre 7 heures et 19 heures .

data\_meteo\_season$hour <- strtoi(str\_sub(data\_meteo\_season$date..date, 13, 14))  
  
data\_meteo\_jour<- filter(data\_meteo\_season , hour >= 7)  
  
data\_meteo\_jour <- filter(data\_meteo\_jour , hour <= 21)

On observe que de 8 a 10 heures les données ne sont pas traitées par problème de selection dans R studio .

Maintenant on vas observer la correlation entre la pression et la pluie pour une journée les de 7 heure à 21 heures

cor.test(data\_meteo\_jour$pressure , data\_meteo\_jour$rain) #-0.07924169

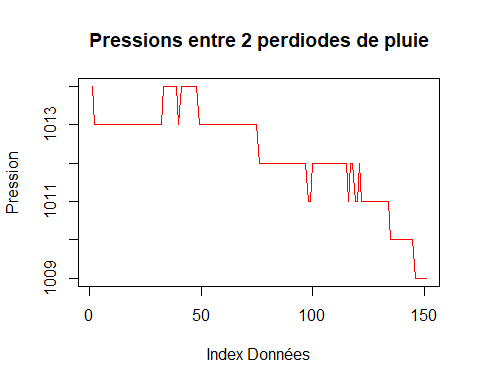
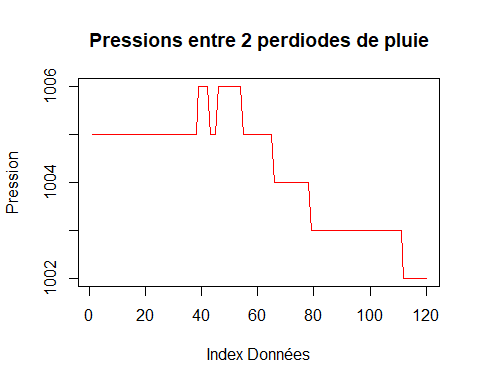
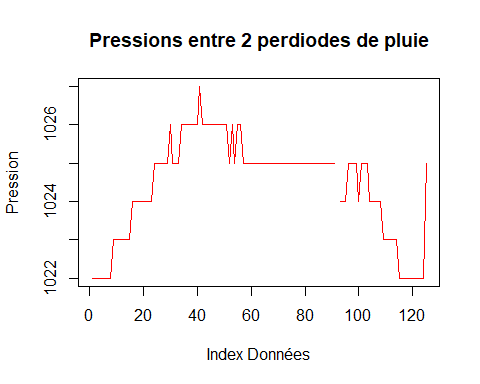
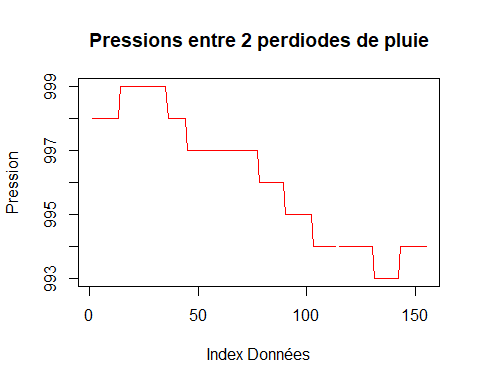
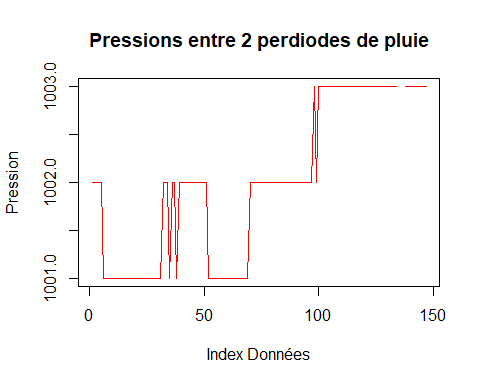
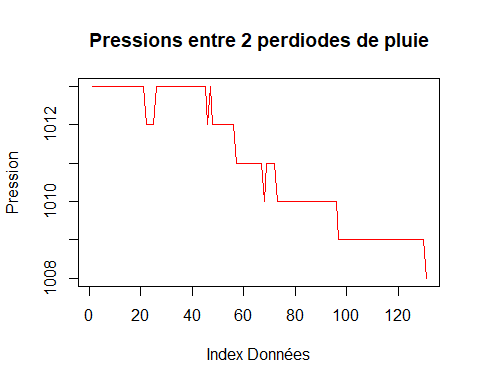
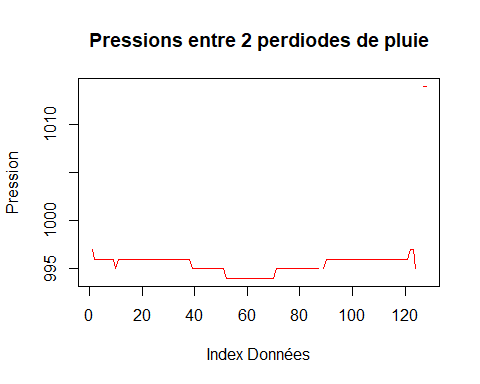
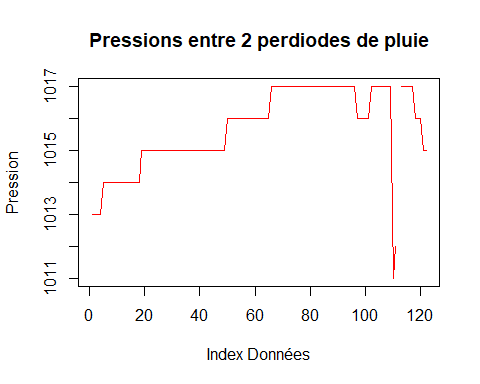
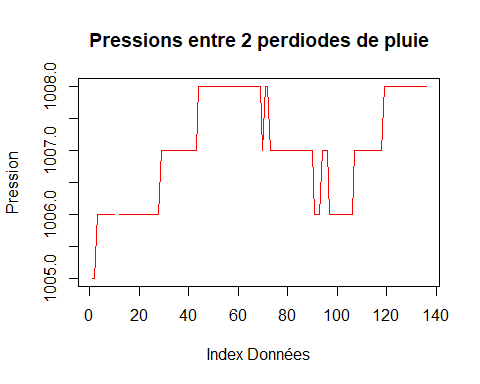
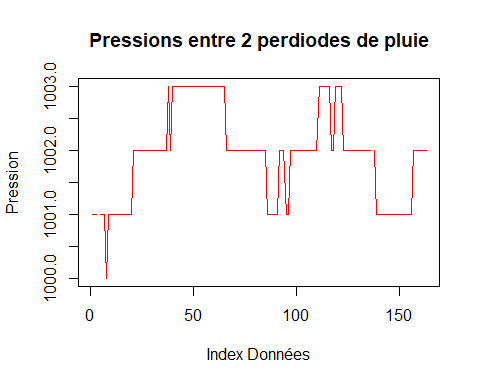
##   
## Pearson's product-moment correlation  
##   
## data: data\_meteo\_jour$pressure and data\_meteo\_jour$rain  
## t = -15.425, df = 42047, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.08451056 -0.06550187  
## sample estimates:  
## cor   
## -0.07501303

On observe ici une très légére correlation négative , est un p-value très faible . Nous en déduisons donc que l’analyse de l’évolution de la préssion athmosphérique ne peut pas nous permettre de determiner fiablement si la pluie risque de tomber dans l’heure .

Pour confirmer graphiquement que la Pression athmosphérique n’est pas un bon indicateur nous avons extrait les données se trouvant juste avant une période de pluie . Pour faire plus simple nous selectionnons les données se trouvant entre chaque péridode de pluie grace un un code python.

#path\_to\_python <- "C:/Python39/python.exe"  
#use\_python(path\_to\_python, required = TRUE)  
# C:\\Users\\sburd\\OneDrive\\Bureau\\Semestre2\\Projet\_Data\Stat\\Stat\_Projet\_Data\\RecupPluieTrue.py  
py\_install("pandas")  
  
source\_python("C:\\Users\\sburd\\OneDrive\\Bureau\\Semestre2\\Projet\_Data\\Stat\\Stat\_Projet\_Data\\RecupPluieTrue.py")  
#monCsv <- openMonCsv("C:\\Users\\sburd\\OneDrive\\Bureau\\Semestre2\\Projet\_Data\\DataV2\\dataTrainSortiPluie1.csv")  
monCsv <- openMonCsv("C:\\Users\\sburd\\OneDrive\\Bureau\\Semestre2\\Projet\_Data\\DataV2\\dataTrainSorti.csv")  
heureAdd <- AddHour(monCsv)  
maTabPluie <- trie(heureAdd)  
tabDataFrame <- SubsetData(maTabPluie)  
newCollection <- TrieTaille(120,180,tabDataFrame , heureAdd)

La varable NewCollection est une liste de dataFrame de periode d’avant pluie comprise entre 120 et 150 données ( correspondant aproximativement à une journée de 24 heures ).



Dans l’objective nous voudrions voir une augmentation ou une diminution de clair de la pression pouvant nous permettre de prédire la pluie sans trop de risuqe de ce tromper. Ces graphiques montre bien la très faible corrélation trouvé ci dessus.

# Analyse de la température :

Toujour dans le même objectif de trouver une variable présentant une forte corrélation avec la varibale pluie, nous décidons maintenant d’observer la corelation entre la température et la pluie.

Pour limiter les variation de température lié au cycle jour / nuit du soleil nous décidons de traiter des données effectué entre 7:00 heure et 21:00 heures.

cor.test(data\_meteo\_jour$temperature , data\_meteo\_jour$rain)

##   
## Pearson's product-moment correlation  
##   
## data: data\_meteo\_jour$temperature and data\_meteo\_jour$rain  
## t = -51.184, df = 38867, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.2605810 -0.2419536  
## sample estimates:  
## cor   
## -0.2512906

on observe une corrélation négative relativement forte entre la temperature et la présense de pluie , ce qui semble logique , la pluis diminue forcément la température extérieure .

Essayons de regarder maintenant si l’arrivé de la pluie est précédé par une diminution de la température .

#   
# for (elem in newCollection){  
#   
#   
# plot( elem[" temperature"], main="temperature entre 2 perdiodes de pluie",  
# ylab="Temperature ( en degré ° )",  
# xlab="Index Données",  
# type="l",  
# col="blue" )  
#   
#   
# }

La température semble effectivement etre assez basse à l’arret de la pluie et semble dans la plupart des cas léérement diminué lors de l’arrivée de la pluie mais ces resultats ne semble pas rééllement etre un moyen de prédire efficassement la pluie .

# Analyse de l’humidité :

Nous pensons que l’humidité doit être croissante avant l’arrivé de la pluie . Nous cherchons tout d’abord à s’avoir si une correlation existe entre l’humidité et la quantité de Pluie. Toujours limiter les variation d’humidité lié au cycle jour / nuit du soleil et de la rosée nous décidons de traiter des données effectué entre 7:00 heure et 21:00 heures.

cor.test(data\_meteo\_jour$humidity , data\_meteo\_jour$rain)# 0.30

##   
## Pearson's product-moment correlation  
##   
## data: data\_meteo\_jour$humidity and data\_meteo\_jour$rain  
## t = 65.671, df = 41922, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.2967093 0.3140684  
## sample estimates:  
## cor   
## 0.3054142

On observe étonnament une corrélation plus faible que nous ne le pensions entre la pluie et le pourcentage d’humidité . Nous nous attendions à trouver un corélation positive et très élévé .

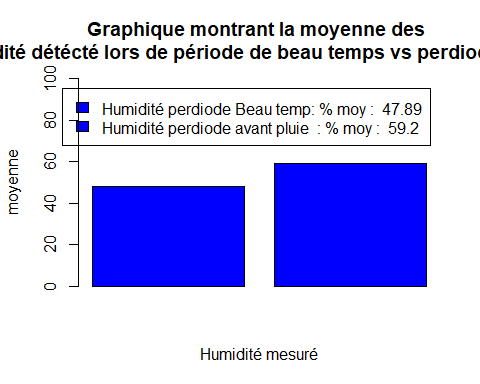
Nous allons donc observer les graphs sur les quelques échantillons correspondants.

for (elem in newCollection){  
   
  
 plot( elem[" humidity"], main="% Humidité entre 2 perdiodes de pluie",  
 ylab="humidity ( en degré ° )",  
 xlab="Index Données",  
 type="l",  
 col="blue" )  
  
   
}

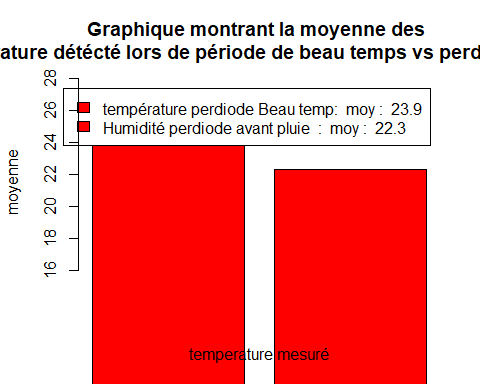
On observe que le pourcentage d’humidité est toujours très élevé à la fin d’une période de pluie par contre l’augmentation de se pourcentage ne semble pas toujours mené à une pédiode de pluie.

## Moyenn millieu perdiode sans pluie vs juste avant la pluie

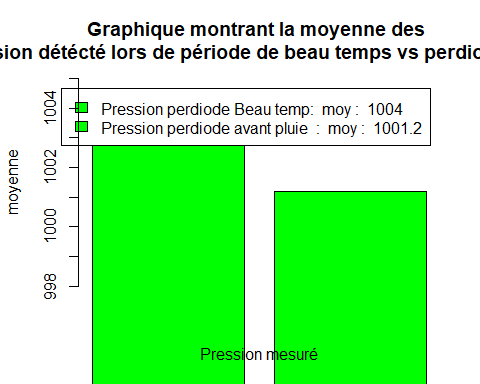
# source\_python("C:\\Users\\sburd\\OneDrive\\Bureau\\Semestre2\\Projet\_Data\\Stat\\Stat\_Projet\_Data\\RecupPluieTrue.py")  
# monCsv <- openMonCsv("C:\\Users\\sburd\\OneDrive\\Bureau\\Semestre2\\Projet\_Data\\DataV2\\dataTrainSorti.csv")  
# #heureAdd <- AddHour(monCsv)  
# maTabPluie <- trie(monCsv)  
# tabDataFrame <- SubsetData(maTabPluie)  
# newCollection <- TrieTaille(100,800,tabDataFrame , monCsv)  
# TakeDonnePeriode <- MoyenneMidPerdiodeFalsePluieVsBeforePluie(newCollection)  
# resMillieu <- resMoyenneMillieu(TakeDonnePeriode[0])  
# resBefore <- resMoyenneBefore(TakeDonnePeriode[1])  
  
  
# {'HumidityM': 47.888888888888886, 'TemperatureM': 23.9, 'PressureM': 1004.0, 'LightM': 529.6666666666666}  
# {'HumidityB': 59.2, 'TemperatureB': 22.3, 'PressureB': 1001.4, 'LightB': 153.6}  
  
# vM <- c(round(47.888888888888886,2), round(23.9,2), round(1004.0,2), round(529.6666666666666,2))  
# valeurMoyenneMillieu<-as.matrix(vM)  
#   
# vB <- c(round(59.2,1), round(22.3,1), round(1001.4,1), round(153.6,1))  
# valeurMoyenneBefore<-as.matrix(vB)  
#   
#   
# barplotvM <- barplot(vM,  
# col = c("blue", "green", "yellow", "red"),  
# legend.text = c(paste("Humidité : % moy : " , round(59.2,2)) ,  
# paste("Temperature : température moy : " , round(22.3,2) ,"°C"),  
# paste("Pression : pressions moy" , round(1001.4,2)),  
# paste("Lumière : moy : " , round(153.6,2))   
# ),  
# main = paste("Graphique montrant la moyenne des \n variables détécté lors de période de beau temps"),  
# ylim = c(0,1010),  
# xlab = "Varaible mesurées",  
# ylab = "moyenne"  
#   
# )  
#   
#   
# barplotvB <- barplot(vB,  
# col = c("blue", "green", "yellow", "red"),  
# legend.text = c(paste("Humidité : % moy : " , round(47.888888888888886,2)) ,  
# paste("Temperature : température moy : " , round(23.9,2) ,"°C"),  
# paste("Pression : pressions moy" , round(1004.0,2)),  
# paste("Lumière : moy : " , round(529.6666666666666,2))   
# ),  
# main = paste("Graphique montrant la moyenne des \n variables détécté quelque seconde avant un épisode de pluie"),  
# ylim = c(0,1010),  
# xlab = "Varaible mesurées",  
# ylab = "moyenne"  
#   
# )  
  
  
barplotvM <- barplot(c(round(47.888888888888886,2) , round(59.2,2)),  
 col = c("blue", "blue"),  
 legend.text = c(paste("Humidité perdiode Beau temp: % moy : " ,round(47.888888888888886,2) ) ,  
 paste("Humidité perdiode avant pluie : % moy : " , round(59.2,2))),  
 main = paste("Graphique montrant la moyenne des \n de l'humidité détécté lors de période de beau temps vs perdiode de pré pluie"),  
 ylim = c(0,100),  
 xlab = "Humidité mesuré",  
 ylab = "moyenne"  
)



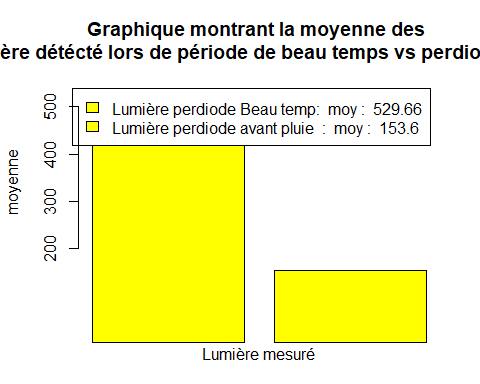
barplotvM <- barplot(c(23.9 , 22.3),  
 col = c("red", "red"),  
 legend.text = c(paste("température perdiode Beau temp: moy : " , 23.9 ) ,  
 paste("Humidité perdiode avant pluie : moy : " , 22.3)),  
 main = paste("Graphique montrant la moyenne des \n de la temperature détécté lors de période de beau temps vs perdiode de pré pluie"),  
 ylim = c(15,28),  
 xlab = "temperature mesuré",  
 ylab = "moyenne"  
)



barplotvM <- barplot(c(1004.0 , 1001.2),  
 col = c("green", "green"),  
 legend.text = c(paste("Pression perdiode Beau temp: moy : " , 1004.0 ) ,  
 paste("Pression perdiode avant pluie : moy : " , 1001.2)),  
 main = paste("Graphique montrant la moyenne des \n de la Pression détécté lors de période de beau temps vs perdiode de pré pluie"),  
 ylim = c(998,1005),  
 xlab = "Pression mesuré",  
 ylab = "moyenne"  
)



barplotvM <- barplot(c(529.6666666666666 , 153.6),  
 col = c("yellow", "yellow"),  
 legend.text = c(paste("Lumière perdiode Beau temp: moy : " , 529.66 ) ,  
 paste("Lumière perdiode avant pluie : moy : " , 153.6)),  
 main = paste("Graphique montrant la moyenne des \n de la Lumière détécté lors de période de beau temps vs perdiode de pré pluie"),  
 ylim = c(120,560),  
 xlab = "Lumière mesuré",  
 ylab = "moyenne"  
)

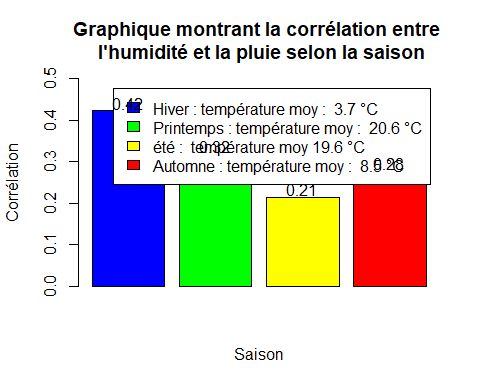


On observe que la moyenne d’humidité juste avant une période de pluie est plus élévé de presque 10 point . La température moyenne perd presque 1 degré juste avant une période de pluie . La pression atmosphérique moyenne est au alentour de 1001 juste avant une période de pluie alors qu’elle est au alentour de 1004 lors de période de beau temps . La lumière détécté est quand à elle presque 4 fois supérieure lors de période de beau temps . c’est information semble montrer qu’une période de pluie peut finalement etre prédit lorsque l’on regarde les moyennes .

## Facteur d’humidité en hiver

Nous émetons l’hipothèse que le facteur d’humidité est un facteur annonciateur de la pluie plus conséquent en hiver que pendant les autres saisons.

data\_meteo\_season$month <- strtoi(str\_sub(data\_meteo$date..date, 7, 8))  
  
data\_meteo\_season$season<-"summer"  
data\_meteo\_season$season[data\_meteo\_season$month>9&data\_meteo\_season$month<=12]<-"autumn"  
data\_meteo\_season$season[data\_meteo\_season$month>1&data\_meteo\_season$month<=3]<-"winter"  
data\_meteo\_season$season[data\_meteo\_season$month>3&data\_meteo\_season$month<=6]<-"spring"  
data\_meteo\_season$season<-factor(data\_meteo\_season$season,levels=c("summer","spring","winter","autumn"))  
  
  
summer <- filter(data\_meteo\_season, season == "summer", rain != "NA", humidity != "NA", temperature != "NA")  
autumn <- filter(data\_meteo\_season, season == "autumn", rain != "NA", humidity != "NA", temperature != "NA")  
spring <- filter(data\_meteo\_season, season == "spring", rain != "NA", humidity != "NA", temperature != "NA")  
winter <- filter(data\_meteo\_season, season == "winter", rain != "NA", humidity != "NA", temperature != "NA")  
  
corWin = cor( winter$humidity, winter$rain, method=c("pearson", "kendall", "spearman")) #0.4774225  
corSum = cor( summer$humidity, summer$rain, method=c("pearson", "kendall", "spearman")) #0.2675425  
corAut = cor( autumn$humidity, autumn$rain, method=c("pearson", "kendall", "spearman")) #0.313826  
corSpr = cor( spring$humidity, spring$rain, method=c("pearson", "kendall", "spearman")) # 0.3957939  
  
  
  
  
CorSaisVal<-as.matrix(c(corWin, corSpr, corSum, corAut))  
  
  
  
barplotsais <- barplot(c(corWin, corSpr, corSum, corAut),  
 col = c("blue", "green", "yellow", "red"),  
 legend.text = c(paste("Hiver : température moy : " , round(mean(winter$temperature), 1),"°C") ,  
 paste("Printemps : température moy : " , round(mean(spring$temperature), 1),"°C"),  
 paste("été : température moy" , round(mean(summer$temperature), 1),"°C"),  
 paste("Automne : température moy : " , round(mean(autumn$temperature), 1),"°C")   
 ),  
 main = paste("Graphique montrant la corrélation entre \n l'humidité et la pluie selon la saison"),  
 ylim = c(0,0.5),  
 xlab = "Saison",  
 ylab = "Corrélation"  
   
)  
  
text(barplotsais ,CorSaisVal+0.02,labels=round(CorSaisVal,2))

 D’après les données recueillit, nous pouvons constater que plus la saison est froide, plus le taux d’humidité est en corrélation avec l’arrivé de la pluie.

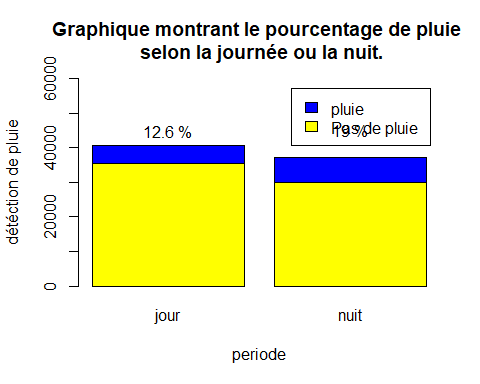
L’hypothèse suggèrant que le facteur d’humidité est un facteur plus conséquent selon la saison est validée, sur le graphique ci-dessus, nous constatons que la corrélation entre l’humidité et la pluie est prêts de deux fois supérieur en hiver que pendant une autre saison.

À partir de ces résultats, nous pourrons donner plus d’importances à la valeur d’humidité selon la saison et selon la température.

## Le cycle jour/nuit jour sur la pluie

Nous émétons l’hypothèse que la nuit est un facteur annonciateur de la pluie.

data\_meteo\_rain <- data\_meteo  
data\_meteo\_rain <- filter(data\_meteo\_rain, rain != "NA" & light != "NA")  
  
data\_meteo\_rain$pluie[data\_meteo\_rain$rain>0]<-"oui"  
data\_meteo\_rain$pluie[data\_meteo\_rain$rain==0]<-"non"  
data\_meteo\_rain$jour[data\_meteo\_rain$light>0]<-"jour"  
data\_meteo\_rain$jour[data\_meteo\_rain$light==0]<-"nuit"  
  
data\_meteo\_jour <- filter(data\_meteo\_rain, jour == "jour")  
data\_meteo\_nuit <- filter(data\_meteo\_rain, jour == "nuit")  
  
effectifsjour <- table(data\_meteo\_jour$pluie)  
racajour=prop.table(effectifsjour) \* 100  
effectifsnuit <- table(data\_meteo\_nuit$pluie)  
racanuit=prop.table(effectifsnuit) \* 100  
  
pcrjp7<-as.matrix(c(racajour["oui"], racanuit["oui"]))  
pcrjp<-as.matrix(c(pcrjp7[1], pcrjp7[2]))  
  
  
barplotjp = barplot(table(data\_meteo\_rain$pluie, data\_meteo\_rain$jour),  
 col = c("yellow", "blue"),  
 legend.text = c(paste("Pas de pluie ") ,  
 paste("pluie ")  
 ),  
 main = paste("Graphique montrant le pourcentage de pluie \n selon la journée ou la nuit."),  
 ylim = c(0,60000),  
 xlab = "periode",  
 ylab = "détéction de pluie"  
 )  
  
text(barplotjp ,pcrjp+45000,labels=as.character(paste(round(pcrjp,1),"%")))



D’après les données recueillit, non pouvons constater sur le graphique précédant qu’il n’y pas de différence significative entre le jour et la nuit.

L’hypothèse signifiant que le le jour ou la nuit sont des facteurs plus conséquents, est donc fausse malgré le fait que nous avons une petite augmentation de pluie la nuit.

D’après ces résultats, nous savons que le cycle jour nuit n’est pas un facteur assez déterminant pour le prendre en compte lors de la mise en place de l’algorithme de détection de pluie.