

# Préparation de données désordonnées

```
rm(list=ls())
```

## Structuration (tidyr)

### Importation et structure

```
# library(rstudioapi)

# current_path <- getActiveDocumentContext()$path
# setwd(dirname(current_path ))

# setwd("C:/Users/Eric/Documents/Symposium/2019/data_workflow/")

raw <- read.csv("data/weather.csv", col.names=c("site_id", "year", "month", "element",
                                                paste("d", 1:31, sep = "")), header = FALSE)

library(knitr)
library(dplyr)
# subset(raw, select=c("year","month","element",paste("d", 1:10, sep = "")))
# kable(subset(raw, select=c("year","month","element",paste("d", 1:10, sep = ""))))
# kable(raw)

library(tidyr)
library(stringr)
library(dplyr)

clean1 <- raw %>% gather(day, value, d1:d31, na.rm = TRUE)
clean1$day <- as.integer(str_replace(clean1$day, "d", ""))
clean1$date <- as.Date(ISOdate(clean1$year, clean1$month, clean1$day))
clean1 <- clean1[c("site_id", "date", "element", "value")]
clean1 <- arrange(clean1, date, element)
clean2 <- spread(clean1, element, value)

outclean <- cbind(id = as.integer(rownames(clean2)), clean2)

write.csv(outclean, "data/weatherclean.csv", row.names = FALSE)

# Changer les permissions du fichier csv pour son importation avec pgadmin
system("icacls sites.csv /grant Everyone:(r)")
```

## Avant-après de la structuration

```
## [1] "Données brutes désordonnées"

##   year month element d1   d2   d3 d4   d5 d6 d7 d8 d9   d10
## 1 2010     1     tmax NA    NA    NA NA    NA NA NA NA NA    NA
## 2 2010     1     tmin NA    NA    NA NA    NA NA NA NA NA    NA
## 3 2010     2     tmax NA 27.3 24.1 NA    NA NA NA NA NA    NA
## 4 2010     2     tmin NA 14.4 14.4 NA    NA NA NA NA NA    NA
## 5 2010     3     tmax NA    NA    NA NA 32.1 NA NA NA NA 34.5
## 6 2010     3     tmin NA    NA    NA NA 14.2 NA NA NA NA 16.8

##
## Premier passage de structuration

##   site_id      date element value
## 1 MX17004 2010-01-30     tmax  27.8
## 2 MX17004 2010-01-30     tmin  14.5
## 3 MX17004 2010-02-02     tmax  27.3
## 4 MX17004 2010-02-02     tmin  14.4
## 5 MX17004 2010-02-03     tmax  24.1
## 6 MX17004 2010-02-03     tmin  14.4

##
## Deuxième passage de structuration

##   site_id      date tmax tmin
## 1 MX17004 2010-01-30 27.8 14.5
## 2 MX17004 2010-02-02 27.3 14.4
## 3 MX17004 2010-02-03 24.1 14.4
## 4 MX17004 2010-02-11 29.7 13.4
## 5 MX17004 2010-02-23 29.9 10.7
## 6 MX17004 2010-03-05 32.1 14.2
```

## Intégration (PostgreSQL)

### Connexion à une base de données

```
#library(rstudioapi)
library(DBI)
library(RPostgreSQL)
library(getPass)

# loads the PostgreSQL driver
drv <- dbDriver("PostgreSQL")

con <- dbConnect(drv, dbname = "postgres",
  host = "localhost", port = 5432,
  # user = rstudioapi::askForPassword("Database user"),
  user = "postgres",
  # password = rstudioapi::askForPassword("Database password"),
  password = "postgres")
# password = getPass())
```

### Création d'une table et insertion de données avec SQL

```
DROP TABLE sites;

CREATE TABLE sites
(
  id integer NOT NULL,
  site_id character varying(50) NOT NULL,
  longitude real,
  latitude real,
  geom geometry,
  site_name character varying(255),
  CONSTRAINT sites_pkey PRIMARY KEY (id)
);

COPY sites(ID,site_id,longitude,latitude,site_name)
FROM 'C:/Users/Eric/Documents/Symposium/2019/data_workflow/data/sites.csv' DELIMITER ',' CSV HEADER;

UPDATE sites SET geom = ST_MakePoint(longitude,latitude);
```

### Exécuter une requête SQL

```
select id, site_id, longitude, latitude, site_name from sites where site_id = 'MX17004';
```

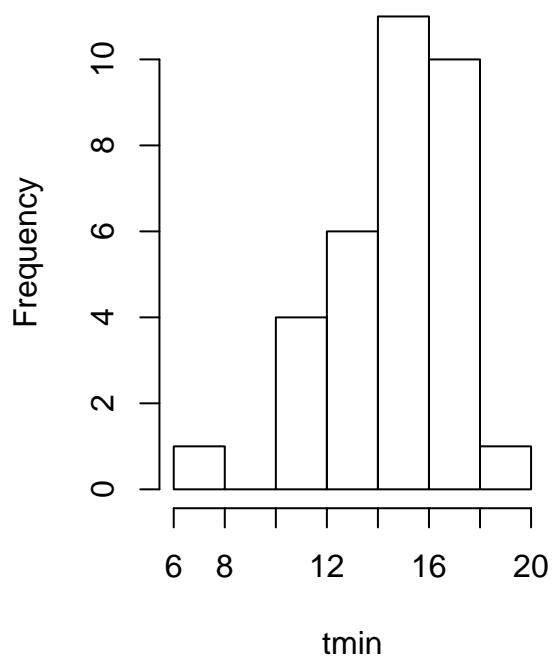
Table 1: 1 records

id	site_id	longitude	latitude	site_name
1	MX17004	-71.1043	42.3151	Mexico

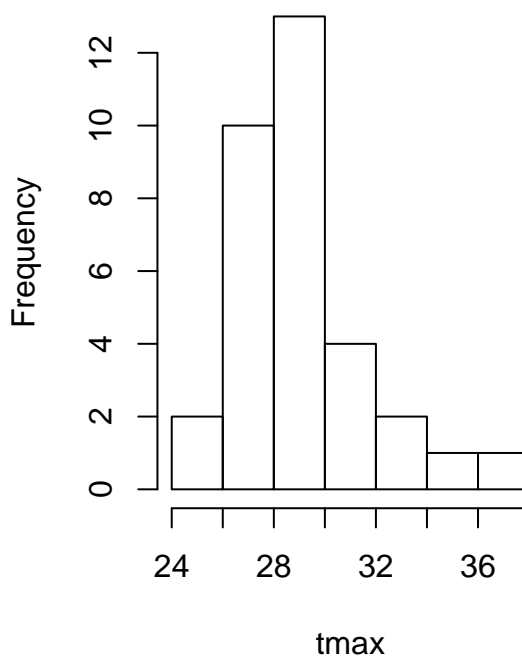
# Analyse des données

## Exploration des données

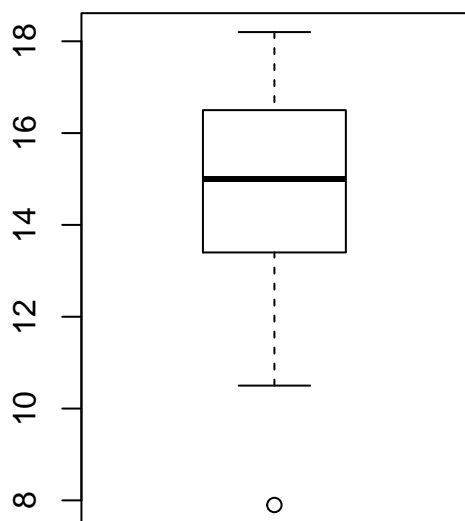
**Histogram of tmin**



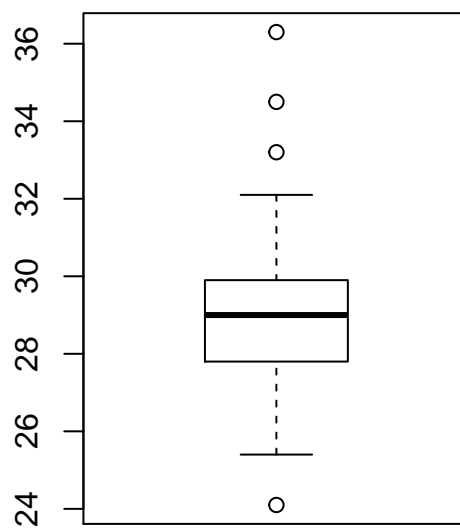
**Histogram of tmax**



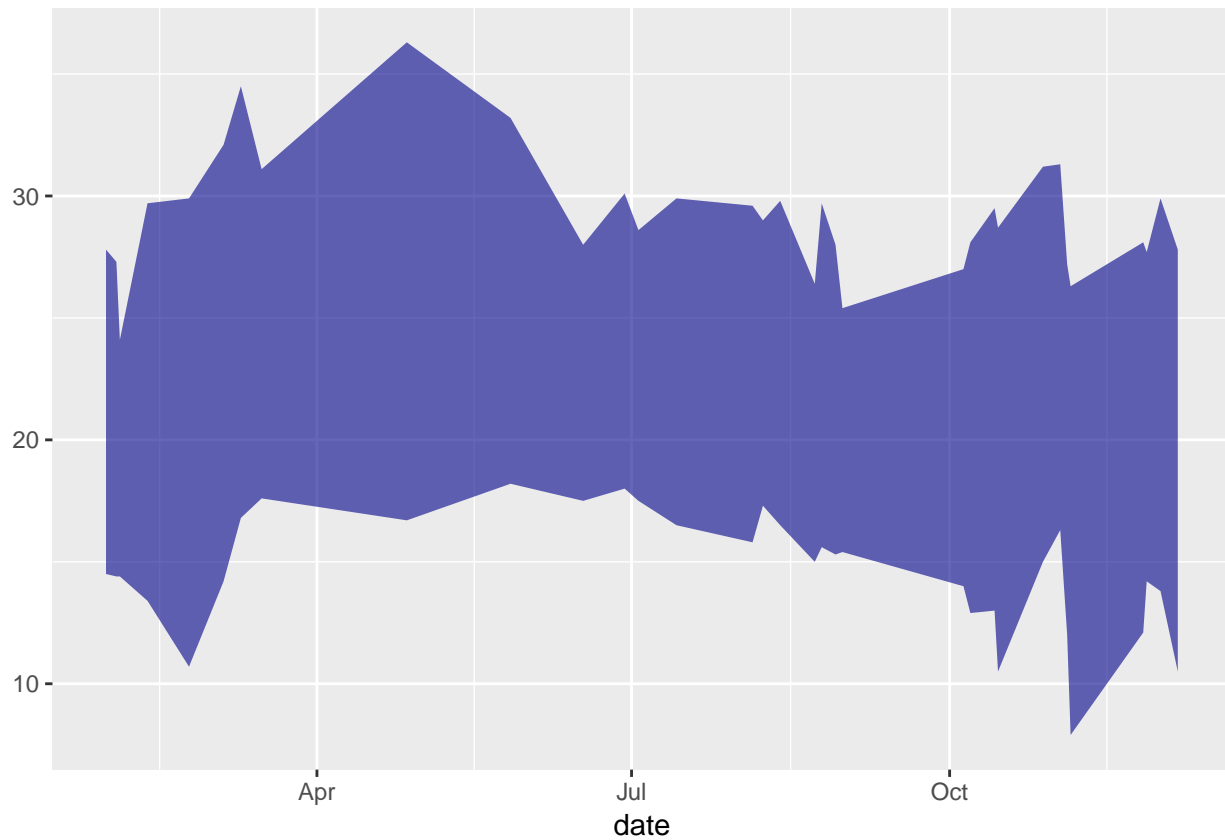
**Boxplot of tmin**



**Boxplot of tmax**



```
## Warning: package 'ggplot2' was built under R version 3.5.3
```



## Gestion de versions

Mise en place :

```
git config --global user.name ebeaulieu
```

```
git config --global user.email e_beaulieu@hotmail.com
```

R Studio Menu Global Options...

Restart RStudio

Onglet Git/SVN :

Git executable : C:/Program Files/Git/bin/git.exe