

Import des librairies et des données

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2    3.4.4      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(tibble)
```

```
dg <- read.csv("diagnostics.csv", sep = ";")
pt <- read.csv("patients.csv", sep = ";")
```

Jointure

```
full_df <- merge(x = dg, y = pt, by = "Patient_id", all.x = TRUE, all.y = TRUE)
```

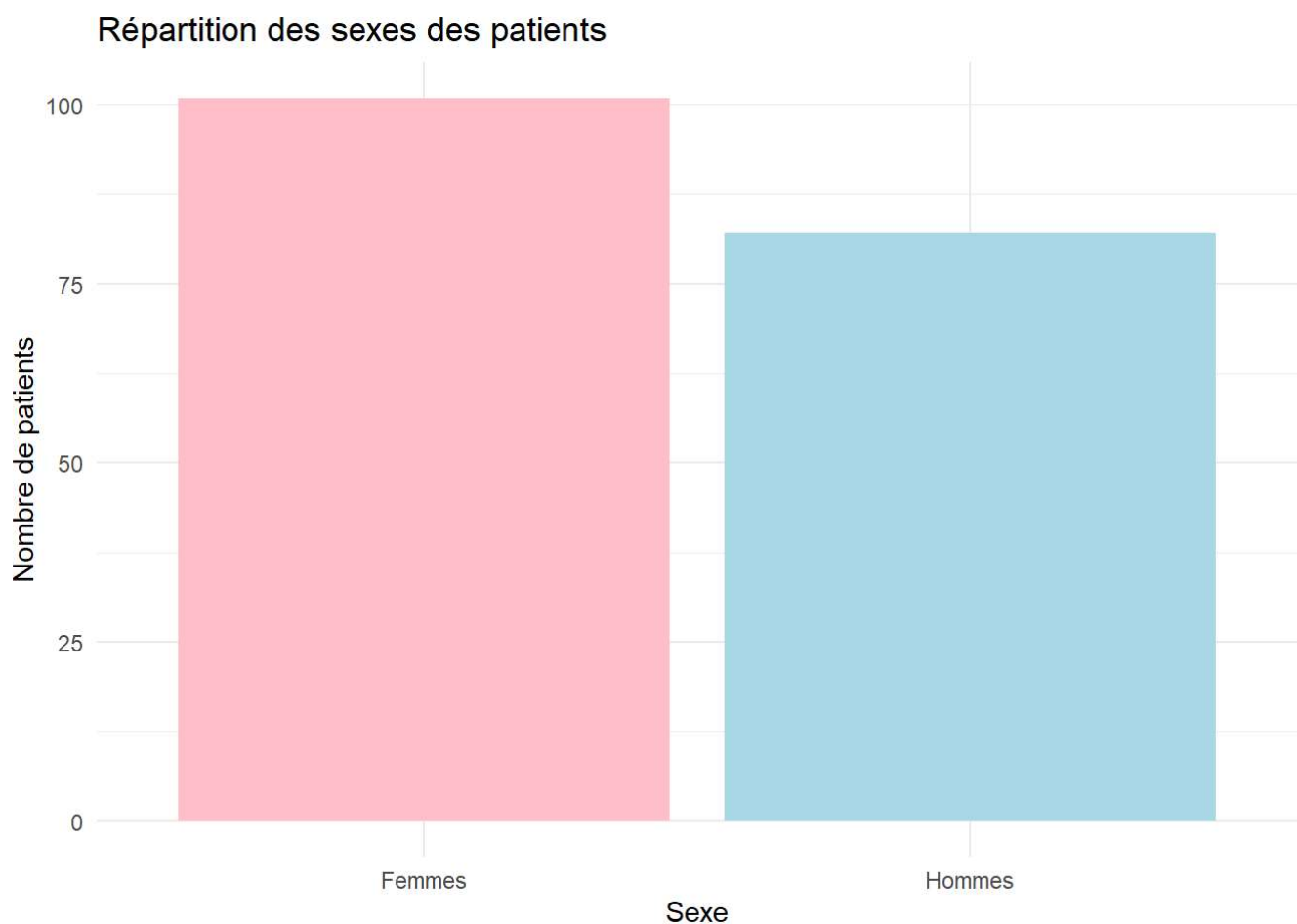
Préparation des labels des epidermolyses

```
codes <- list('257'="Epidermolyse bulleuse simple avec dystrophie musculaire",
              '304'="Epidermolyse bulleuse simple",
              '305'="Epidermolyse bulleuse jonctionnelle",
              '79403'="Epidermolyse bulleuse jonctionnelle avec atrésie pylorique",
              '158684'="Epidermolyse bulleuse simple avec atrésie du pylore",
              '251393'="Epidermolyse bulleuse jonctionnelle localisée")

filiation <- list('305' = c("305", "79403", "251393"),
                 '304' = c("304", "257", "158684"))
```

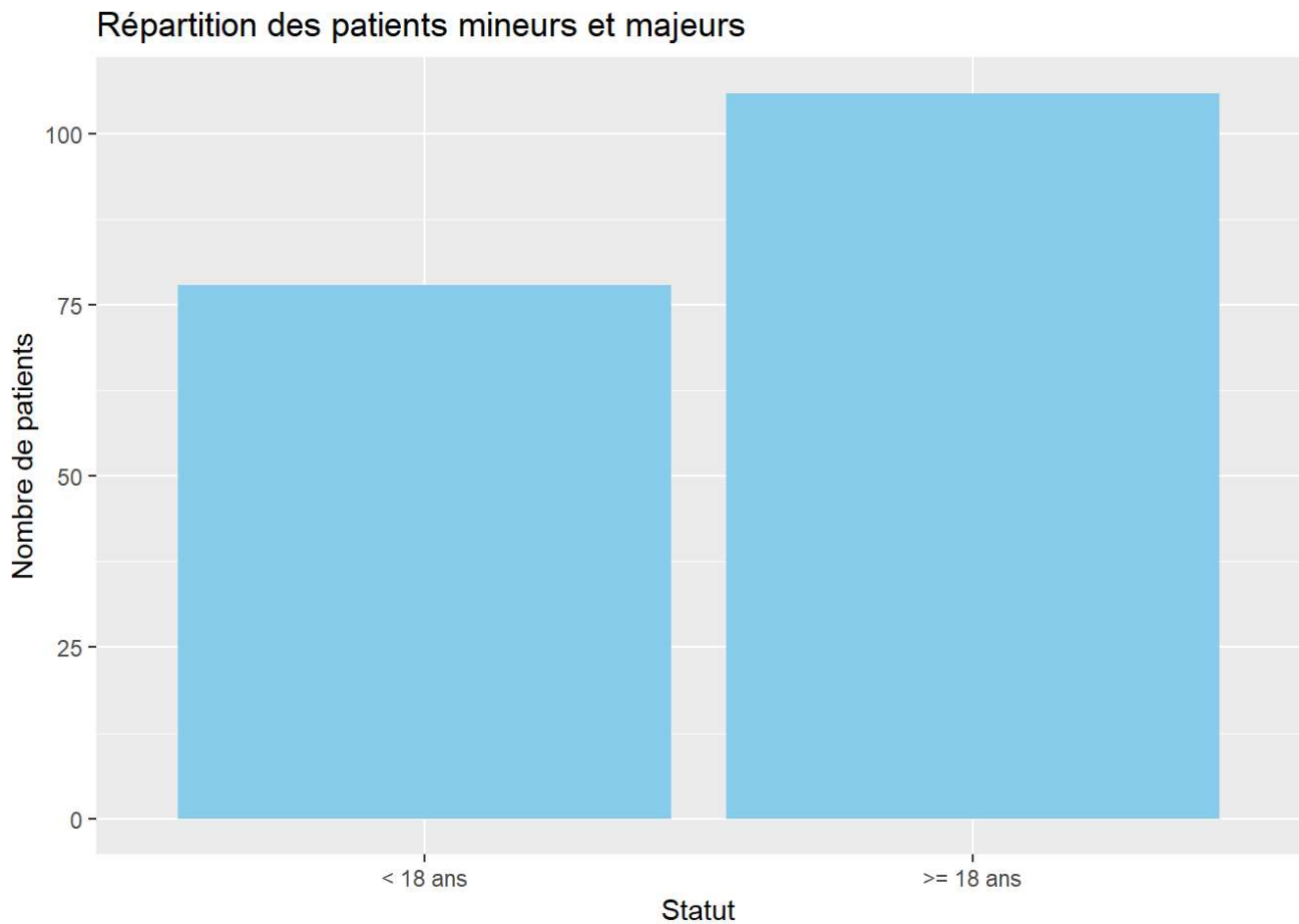
Description par le sexe

```
sex_patients <- full_df %>%  
  select('Patient_id','Sexe') %>%  
  distinct()  
  
ggplot(sex_patients, aes(x = Sexe, fill = Sexe)) +  
  geom_bar() +  
  scale_fill_manual(values = c("M" = "lightblue", "F" = "pink")) +  
  scale_x_discrete(labels = c(  
    "M" = "Hommes",  
    "F" = "Femmes")) +  
  theme_minimal() +  
  theme(legend.position="none")+  
  labs(title = "Répartition des sexes des patients",  
    x = "Sexe",  
    y = "Nombre de patients")
```



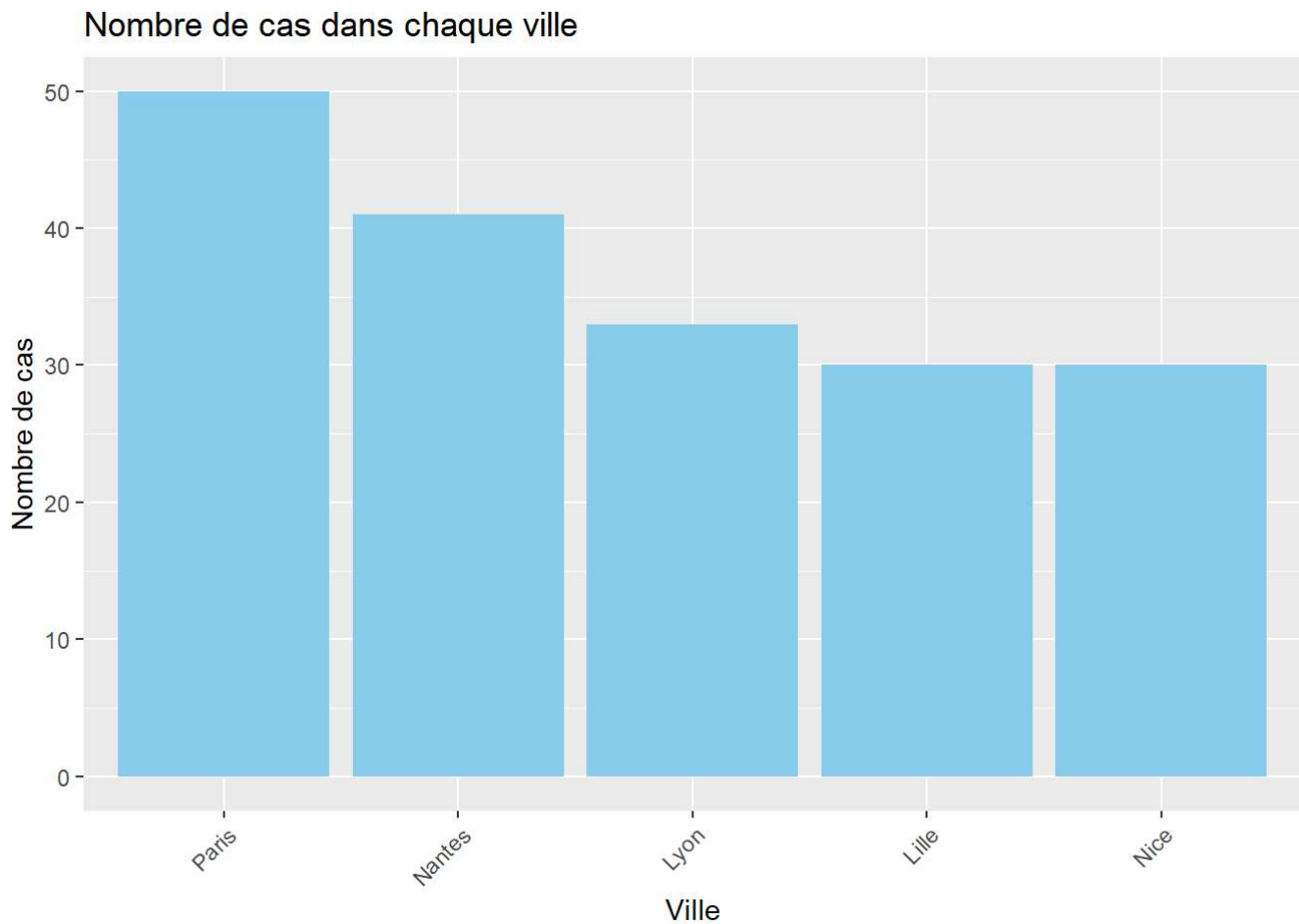
Description par tranche d'âge >< 18

```
age_patients <- full_df %>%  
  select('Patient_id','Annee_naissance') %>%  
  distinct() %>%  
  mutate(age = 2024-Annee_naissance) %>%  
  mutate(majeur = if_else(age >= 18, TRUE, FALSE))  
  
ggplot(age_patients, aes(x = majeur)) +  
  geom_bar(fill = "skyblue") +  
  scale_x_discrete(labels = c("< 18 ans", ">= 18 ans")) +  
  labs(title = "Répartition des patients mineurs et majeurs",  
       x = "Statut",  
       y = "Nombre de patients")
```



Description par ville de prise en charge

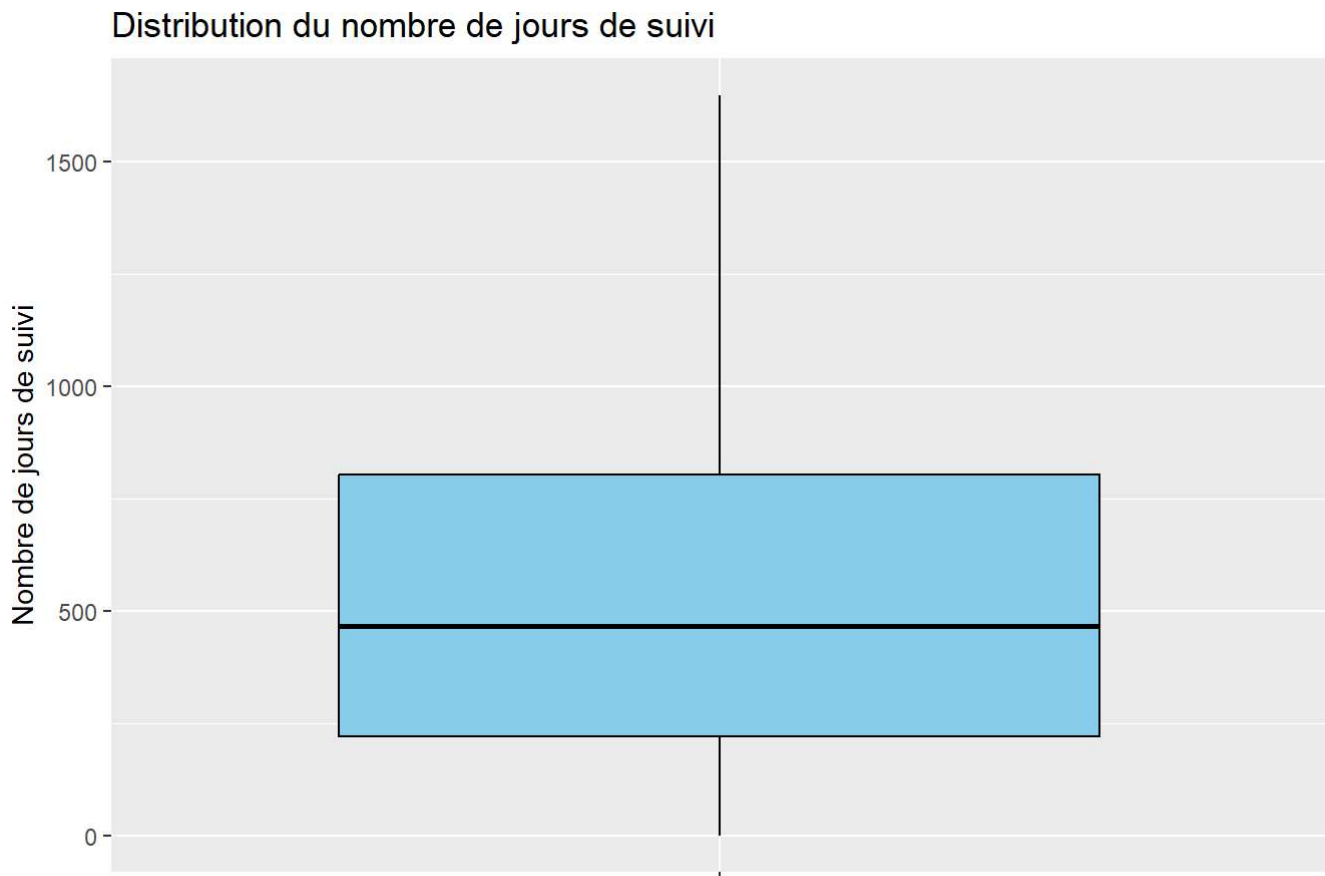
```
ville_patient <- full_df %>%  
  select("Patient_id", "Ville_prise_en_charge") %>%  
  distinct() %>%  
  group_by(Ville_prise_en_charge) %>%  
  summarise(nombre_occurrences = n())  
  
ggplot(ville_patient, aes(x = reorder(Ville_prise_en_charge, -nombre_occurrences), y = nombre_occurrences)) +  
  geom_bar(stat = "identity", fill = "skyblue") +  
  labs(title = "Nombre de cas dans chaque ville",  
       x = "Ville",  
       y = "Nombre de cas") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Description par durée de suivi en jour

```
patient_suivi <- full_df %>%
  select(Patient_id, date_inclusion, date_derniere_activité) %>%
  distinct() %>%
  na.omit() %>%
  mutate(date_inclusion = as.Date(date_inclusion, format = "%d/%m/%Y"),
         date_derniere_activité = as.Date(date_derniere_activité, format = "%d/%m/%Y")) %>%
  arrange(Patient_id, date_inclusion) %>%
  distinct(Patient_id, .keep_all = TRUE) %>%
  mutate(jours = as.numeric(date_derniere_activité - date_inclusion))

ggplot(patient_suivi, aes(x = "", y = jours)) +
  geom_boxplot(fill = "skyblue", color = "black") +
  labs(title = "Distribution du nombre de jours de suivi",
       x = "",
       y = "Nombre de jours de suivi")
```



distribution des cas par groupe d'EBHs

```
codes_df <- as.data.frame(tibble::enframe(codes), stringsAsFactors = FALSE)

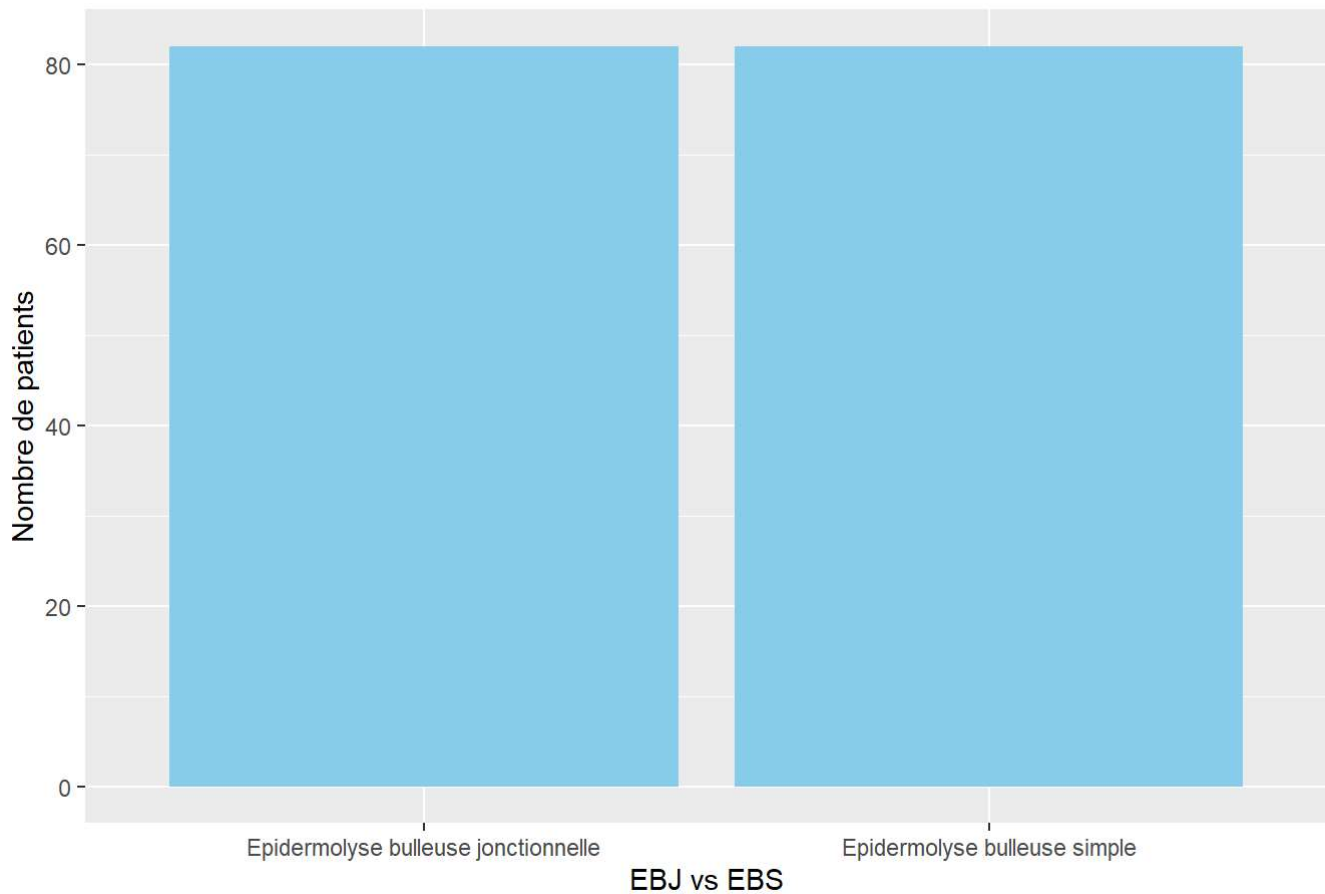
colnames(codes_df) <- c("diagnostic", "label_diag")

patient_diag <- full_df %>%
  select(Patient_id, diagnostic) %>%
  mutate(diagnostic = as.character(diagnostic)) %>%
  na.omit() %>%
  left_join(codes_df, by = "diagnostic") %>%
  mutate(diag_parent = case_when(
    diagnostic %in% filiation[['305']] ~ '305',
    diagnostic %in% filiation[['304']] ~ '304',
    TRUE ~ NA_character_
  )) %>%
  left_join(codes_df, by = c("diag_parent" = "diagnostic")) %>%
  rename(son_label = label_diag.x, ancestor_label = label_diag.y)

patient_diag$ancestor_label<- as.character(trimws(patient_diag$ancestor_label))
patient_diag$son_label<- as.character(trimws(patient_diag$son_label))

ggplot(patient_diag, aes(x = ancestor_label)) +
  geom_bar(fill = "skyblue") +
  labs(title = "Répartition des Epidermolyses Bulleuses et Jonctionnelles",
    x = "EBJ vs EBS",
    y = "Nombre de patients")
```

Répartition des Epidermolyses Bulleuses et Jonctionnelles



```
ggplot(patient_diag, aes(x = son_label)) +  
  geom_bar(fill = "lightgreen") +  
  labs(title = "Répartitions des types d'Epidermolyses",  
        x = "Epidermolyses",  
        y = "Nombre de patients")+  
  theme(axis.text.x = element_text(angle = 60, hjust = 1))
```

