

analyse_valentine

Valentine BERTHIER

2024-01-16

0. LIBRARIES

#Etude du jeu de données “AMR” ##1. Importation du jeu de données

#2. Quelles sont les variables de ce jeu de données

##3. Y-a-t-il des valeurs manquantes ?

Il y n’y a pas de valeurs manquantes

4. Quelles sont les catégories étudiées ? animale, humaine

5. Quelles sont les familles étudiées ? volaille-abattoir, Volaille, ville, veau, poulet-abattoir, Poule-poulet, porc-abattoir, Porc, Ovin, Lapin, ES-hemoculture, ES, EHPAD, dinde-abattoir, Dinde, Chien, Cheval, Chat, Caprin, bovin-adulte, bovin-abattoir‘

6. Lesquelles sont associées à l’étude des humains ? ville, ES-hemoculture, ES, EHPAD

7.

7.2 Pour une étude au sein des régions

Etude des données manquantes

Cette table nous montre le nombre de régions représentées (ou Nationale) par années pour chaque population.

Pour chaque population par année, le nombre maximale d’observation est de `length(unique(AMR$region))` correspondant au nombre de région présentent dans l’étude + l’étude au niveau National. On remarque bien que certaines populations ne sont pas, ou très peu étudiées, pour certaines années.

La dernière ligne du tableau correspond au nombre de population étudiée (toutes régions confondues) par année. On voit bien que le dispositif a beaucoup évolué car en 2012, seulement 65 populations furent étudiées. Ce nombre atteint 152 en 2021.

On ne voit pas grand chose sur ce graphique mais, on peut se rendre compte que certaines populations de sont pas très bien représentées au niveau régional.

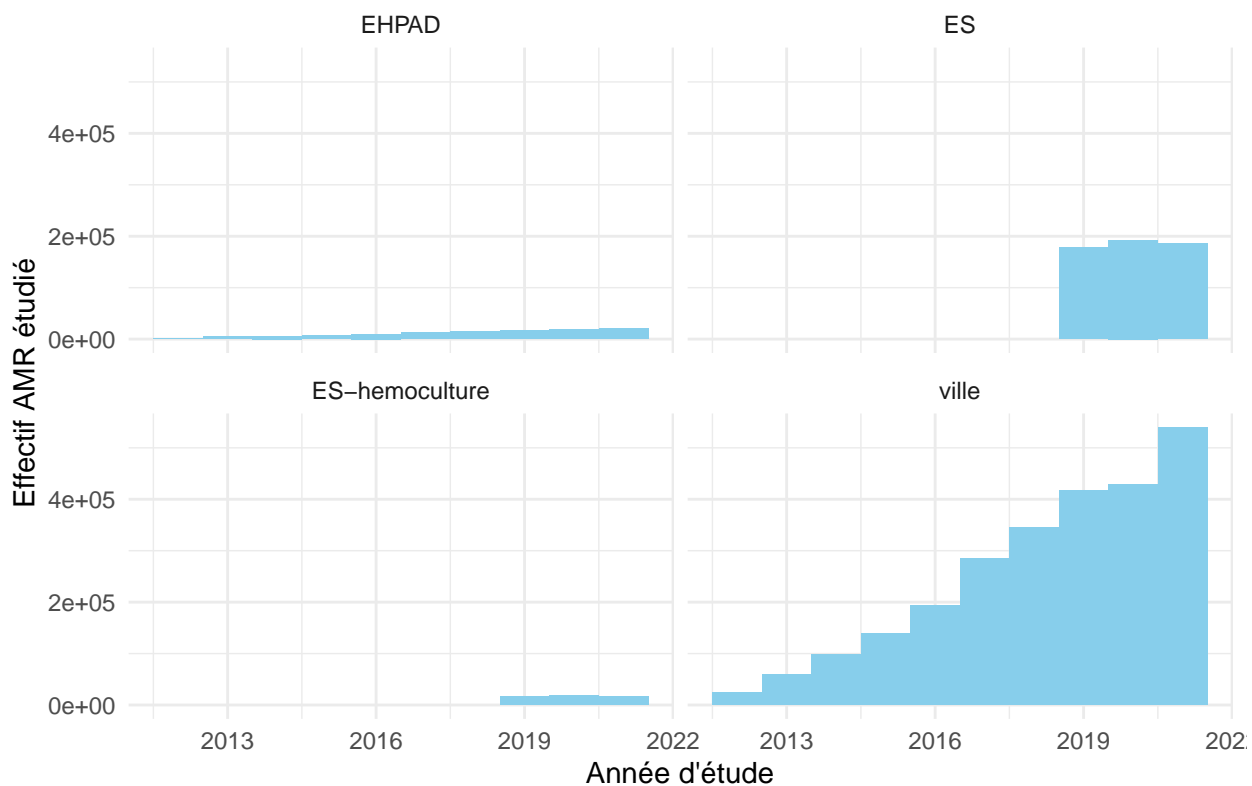
Cette table nous renvoie le nombre d’années d’études (sans savoir lesquelles) des populations par régions

La population la moins étudiée par année est le *porc-abattoir* avec seulement 8 études (régions et années confondues, cela peut être 8 régions sur une même année ou bien 8 années différentes sur des régions différentes ou non) et par région, c’est la Guadeloupe qui est la moins représentée avec seulement 8 populations étudiées (ici seule la famille humaine a été étudiée)

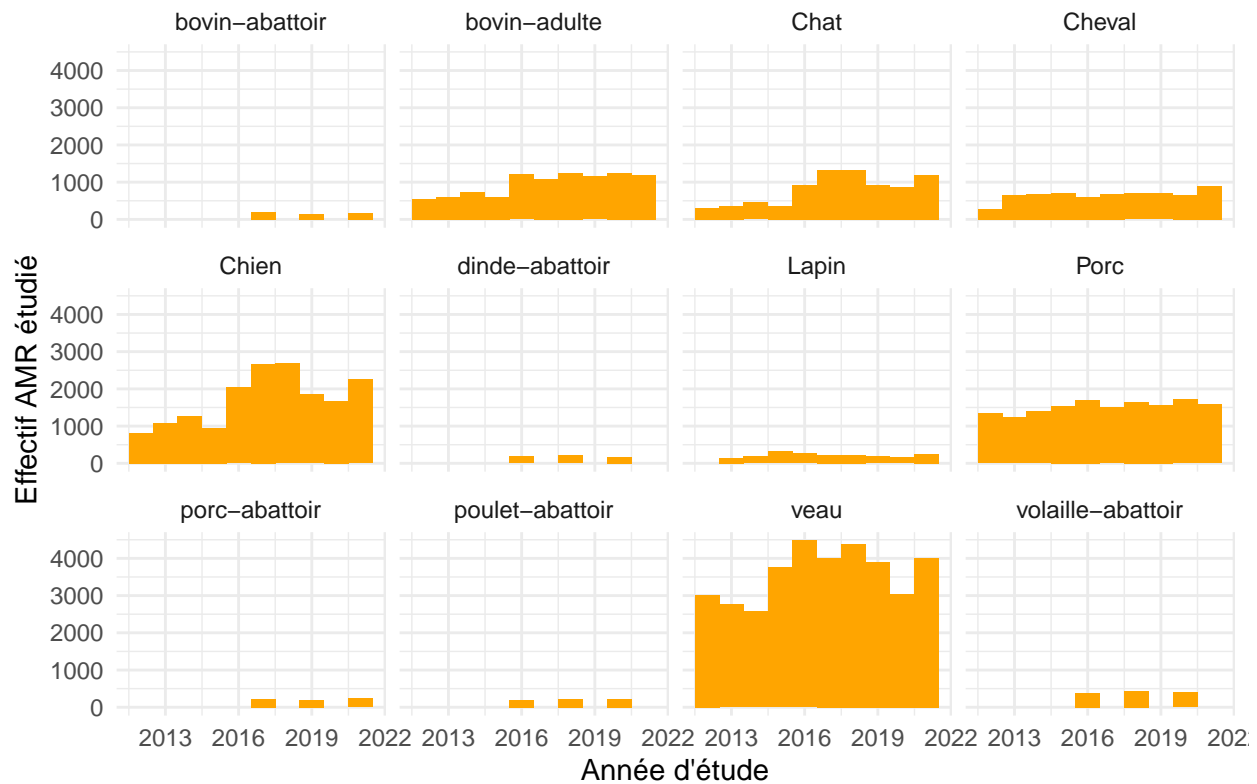
tableau d'effectif de population étudié (par année et par région)

On sépare les populations humaines et animales pour permettre une meilleure comparaison car les effectifs sont d'ordres différents.

Effectif total AMR par population humaine en fonction de l'année



Effectif total AMR par population animale en fonction de l'année



Observations : 1. Dans l'étude des populations humaines, pour les ES, les observations commencent seulement à l'année 2019. 2. En abattoir l'effectif AMR est toujours plus faible, et la résistance aux antibio est testée seulement tous les deux ans. Est ce que c'est plus faible car justement on ne test qu'une petite partie de tous les animaux en abattoir ?

Etudes des données PROMISE

chargement des données, reformattage et filtration (sans SARM) Les variables qui nous intéressent ici sont :

- population
- region
- annee
- profil_AMR
- effectif_AMR

Réseau	Dispositif	Région	population	bactérie(s) étudiée(s)	Profils AMR
Resapath1_sansSARM	Resapath	Auvergne- Rhône-Alpes, Bourgogne- Franche-Comté, Bretagne, Centre-Val de Loire, Grand Est, Hauts-de- France, Île-de-France, Normandie, Nouvelle- Aquitaine, Occitanie, Pays de la Loire, Provence- Alpes-Côte d'Azur	bovin-adulte, Chat, Cheval, Chien, Lapin, Porc, veau, Volaille	Escherichia coli	amoxi-clav-R, C3G-R, carba-R, FQ-R, pan-sensible
ResapathDindes_sansSARM	Resapath	Auvergne- Rhône-Alpes, Bretagne, Centre-Val de Loire, Normandie, Nouvelle- Aquitaine, Pays de la Loire	Dinde	Escherichia coli	amoxi-clav-R, C3G-R, carba-R, FQ-R, pan-sensible
ResapathFrance_sansSARM	Resapath		bovin-adulte, Chat, Cheval, Chien, Dinde, Lapin, Porc, Poule-poulet, veau	Escherichia coli	amoxi-clav-R, C3G-R, carba-R, FQ-R, pan-sensible
ResapathOviCap_sansSARM	Resapath	Auvergne- Rhône-Alpes, Bourgogne- Franche-Comté, Nouvelle- Aquitaine, Occitanie, Pays de la Loire, Provence- Alpes-Côte d'Azur	Caprin, Ovin	Escherichia coli	amoxi-clav-R, C3G-R, carba-R, FQ-R, pan-sensible
Modele_LNR_sansSARM	MNR	bretagne, NA, nouvelle aquitaine, occitanie, pays de la loire	dinde-abattoir, porc-abattoir, poulet-abattoir, veau-abattoir, volaille-abattoir	Escherichia coli	FQ-R, carba-R, C3G-R, pan-sensible

On remarque que les données se superposent car plusieurs régions. Les régions concernées étant

`unique(Collecte$region)` Comment faire ? Peut-être faire une moyenne des effectifs par région en groupant par année ?

Comment est-ce possible que la population des lapins soit si grande ? études des observations de la variable `region` La table `Collecte` est formée des données issues des tables : `* donnees__AMR * Resapath * ResapathDindes * ResapathFrance * ResapathOviCap`

L'erreur est tout à fait logique puisque les données sont doublées ... Dans le réseau `ResapathFrance`, on étudie les populations suivantes : `levels(ResapathFrance$population)` En revanche, pourquoi on a une autre table `ResapathDindes` ?

Modèle Resapath

VALIDATION DE L'INDICATEUR

Est-il possible de calculer l'indicateur pour toutes les caractéristiques ?

(dans `Pour_rapport`) L'indicateur est-il basé sur un nombre d'isolats plus faible (significativement ?) que d'autres indicateurs de résistance ?

```
table1 <- ResapathFrance_sansSARM %>%
  filter(`profil AMR` == "pan-sensible") %>%
  mutate(eff_pansensible = effectif_AMR) %>%
  select(annee, population, eff_pansensible)

table2 <- ResapathFrance_sansSARM %>%
  filter(`profil AMR` == "C3G-R") %>%
  mutate(eff_C3G = effectif_AMR) %>%
  select(annee, population, eff_C3G)

table_join <- merge(table1, table2, by = c("annee", "population"), all = TRUE)

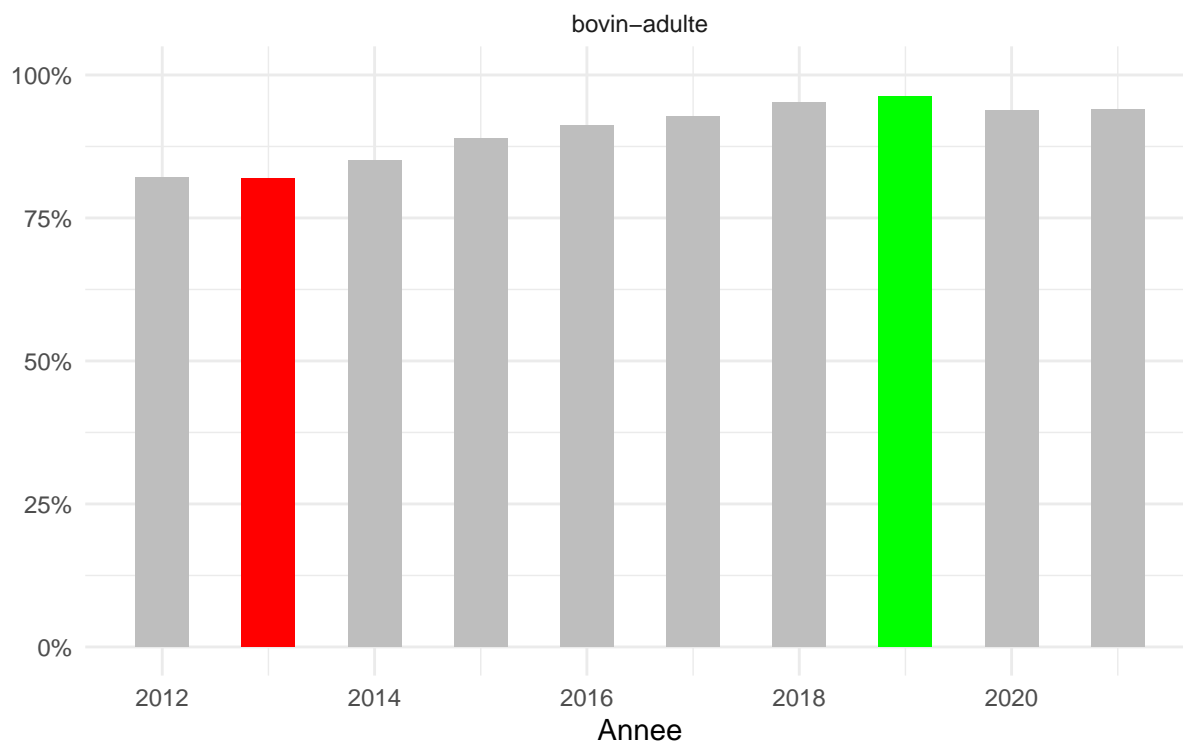
for (pop in levels(table_join$population)){
  table_join1 <- table_join %>%
    filter(population == pop) %>%
    mutate(taux_eff = round((eff_pansensible * 100) / eff_C3G, 2),
           min_max_value = ifelse(taux_eff == max(taux_eff), "MAX", ifelse(taux_eff == min(taux_eff), "MIN", "none")))
}

moyenne = round(mean(table_join1$taux_eff), 2)

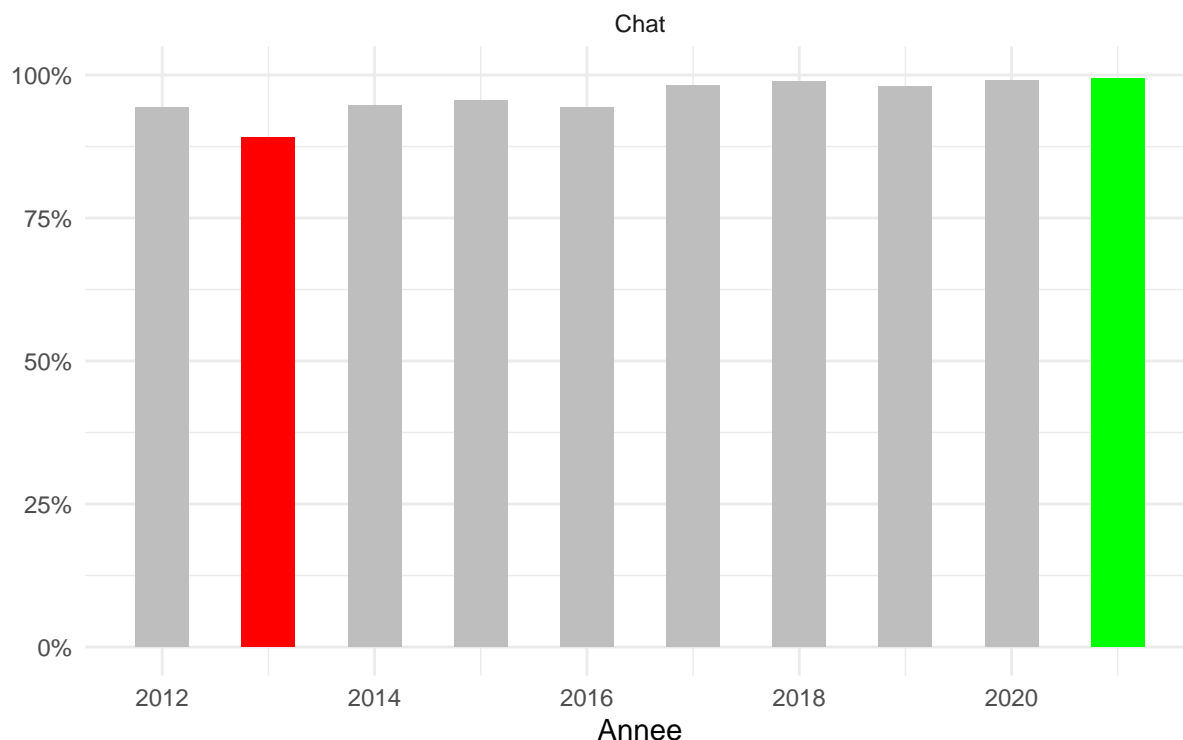
plot <- ggplot(data = table_join1, aes(x = annee, y = taux_eff, fill = min_max_value)) +
  geom_col(width = 0.5) +
  theme_minimal() +
  labs(x = "Annee",
       y = "",
       title = "Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée.",
       subtitle = str_glue("En moyenne, le taux est de {moyenne}%")) +
  facet_wrap(~population) +
  scale_y_continuous(labels = scales::number_format(suffix="%"), limits = c(0,100)) +
  scale_fill_manual(values = c("MIN" = "red", "MAX" = "green", "none" = "grey"), guide = "none")
```

```
plot(plot)  
}
```

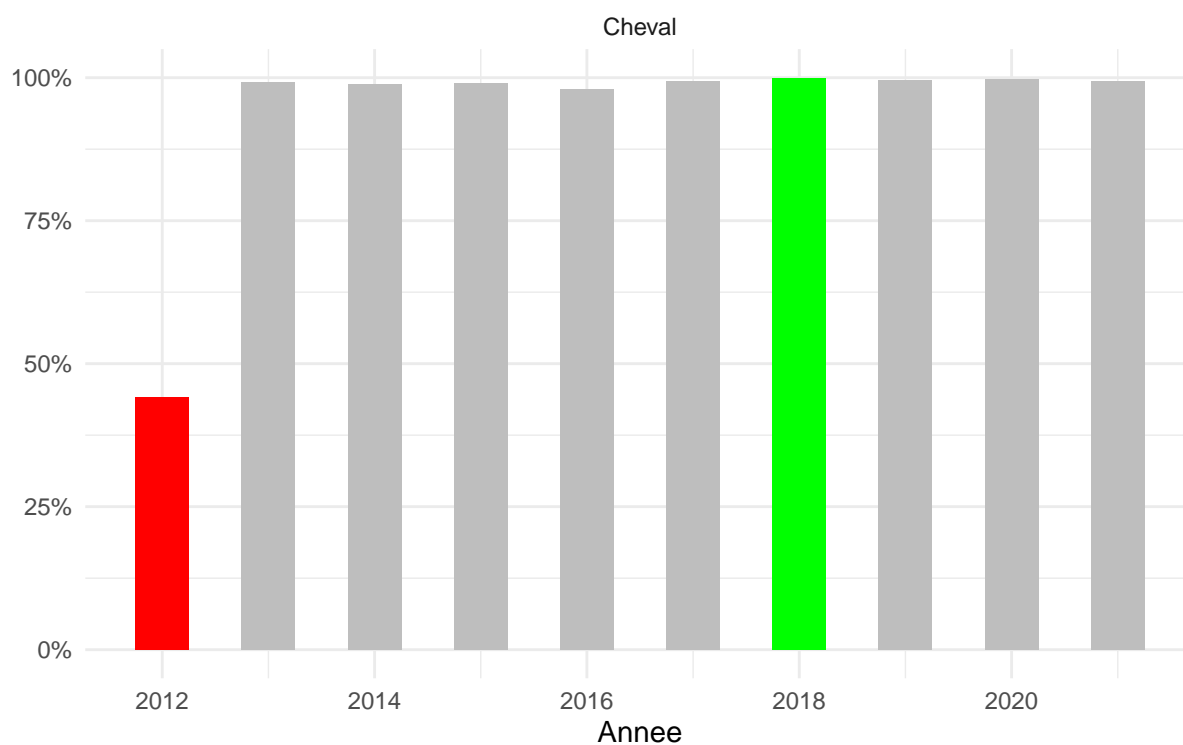
Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 90.11%



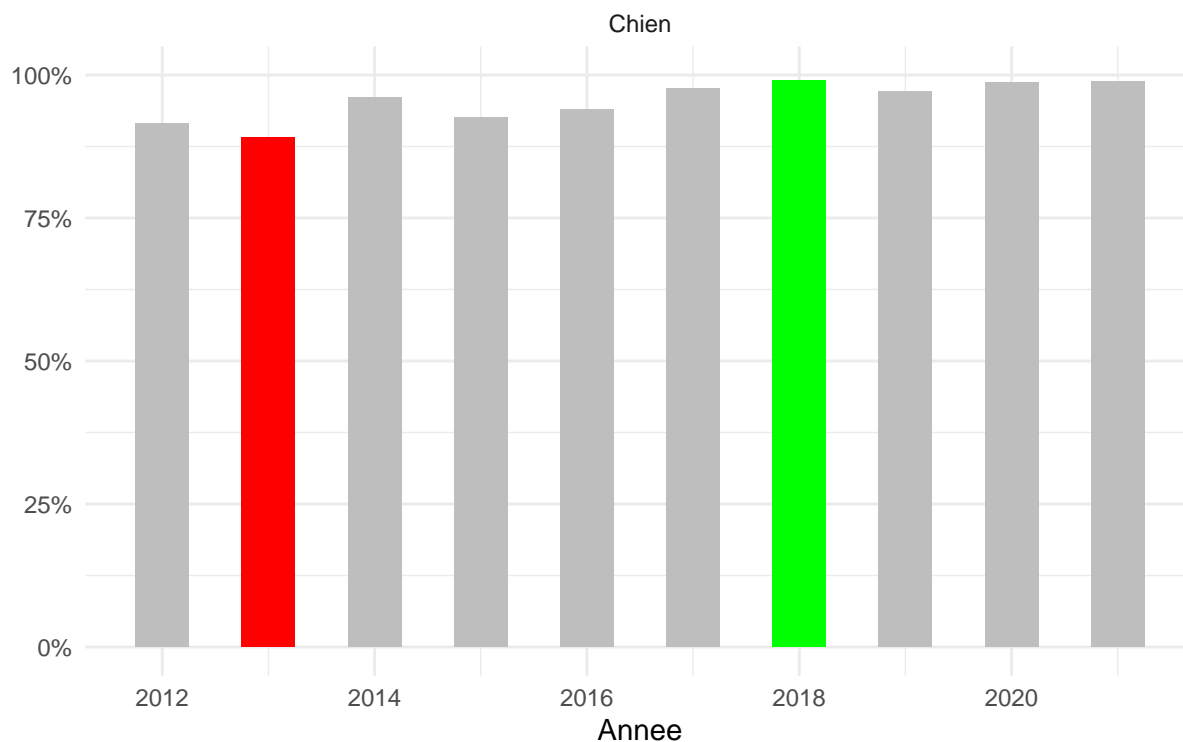
Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 96.12%



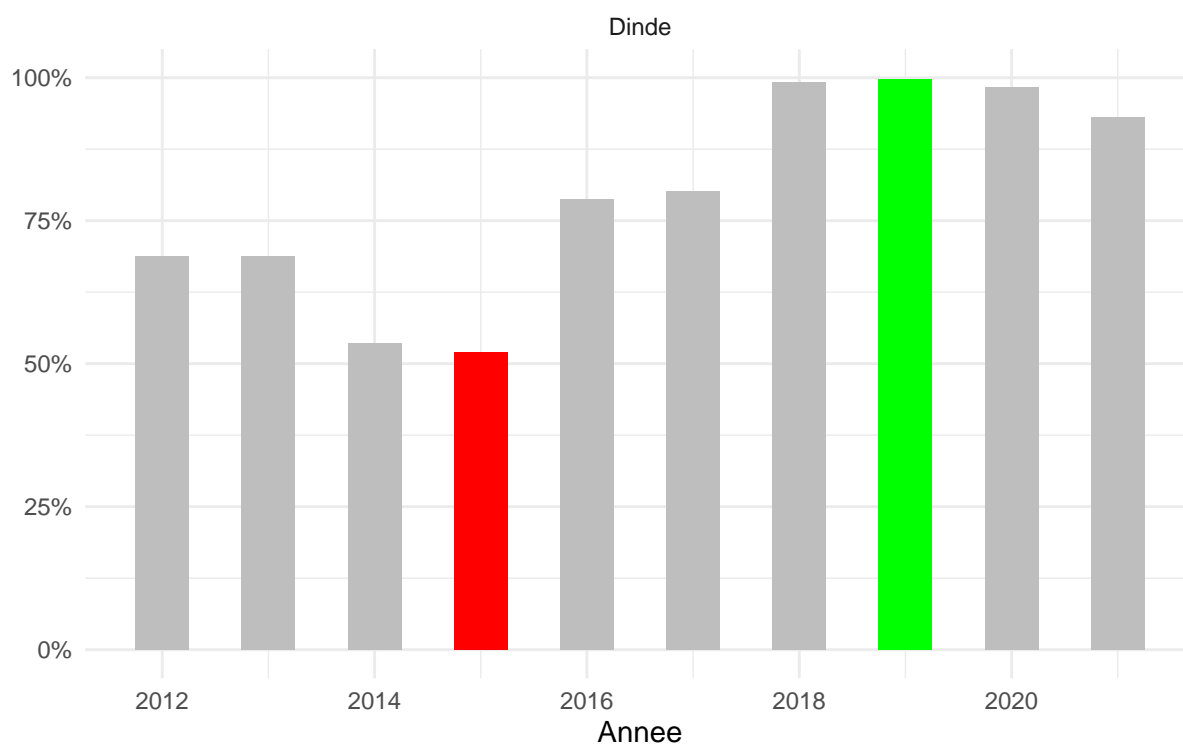
Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 93.71%



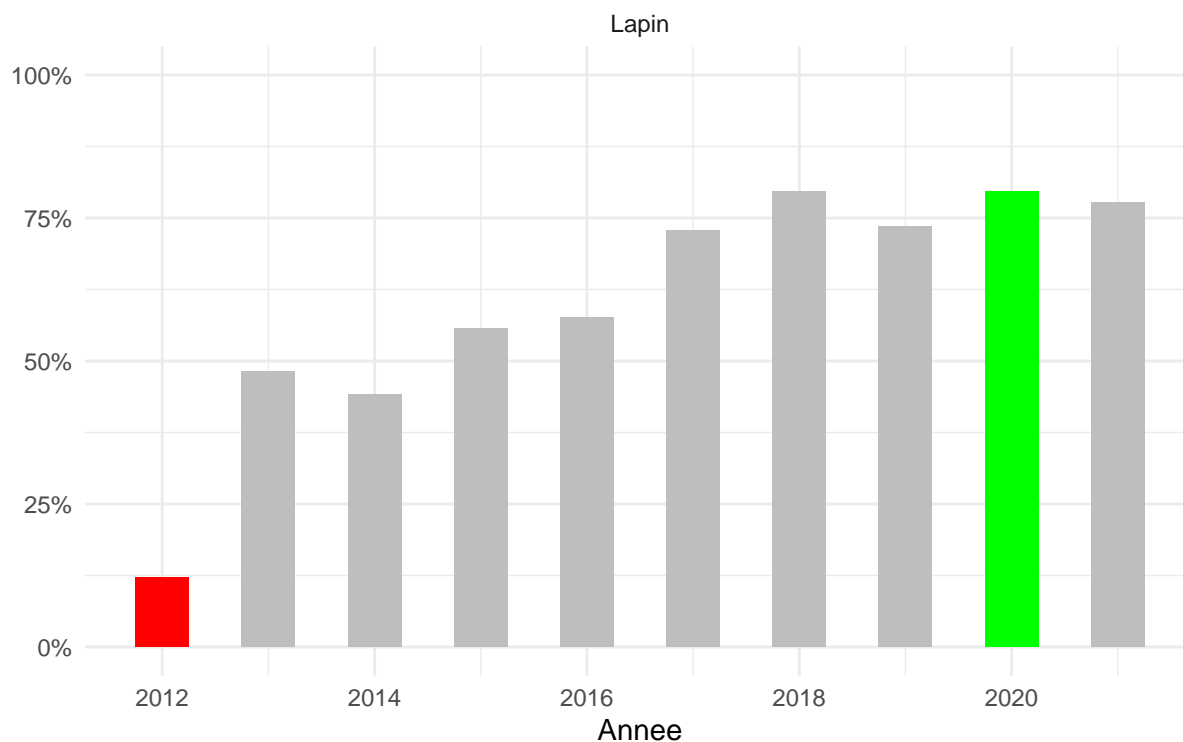
Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 95.45%



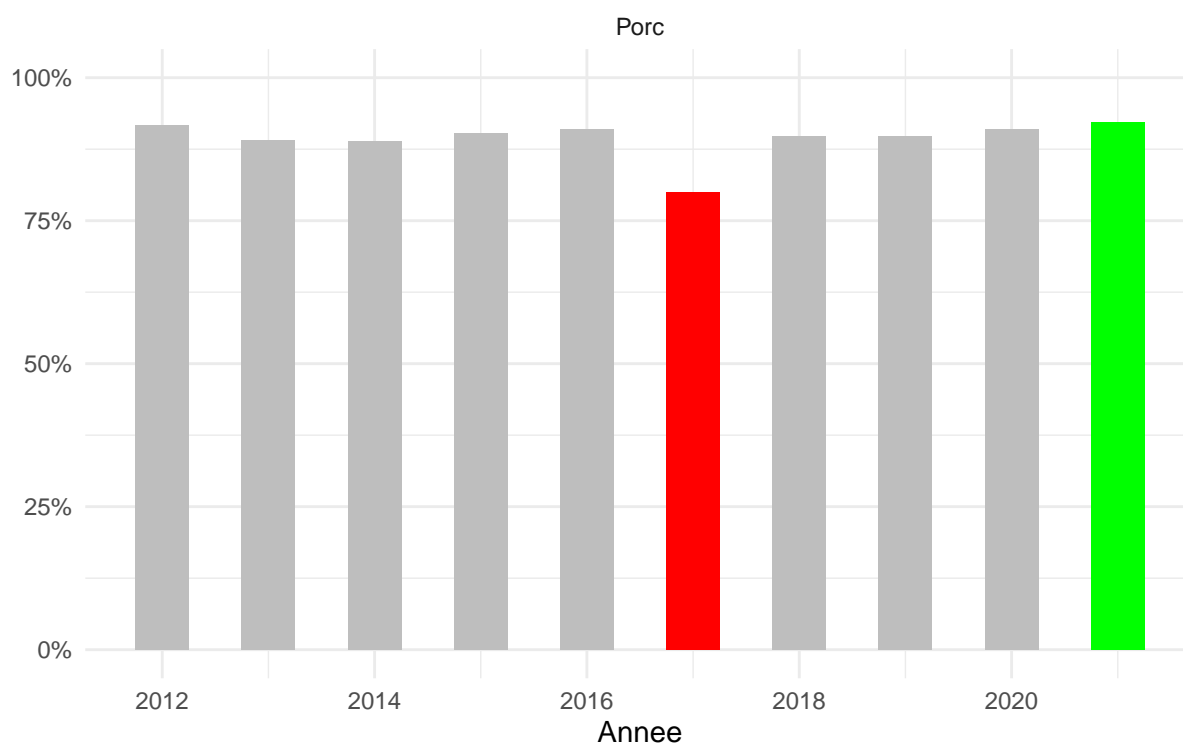
Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 79.26%



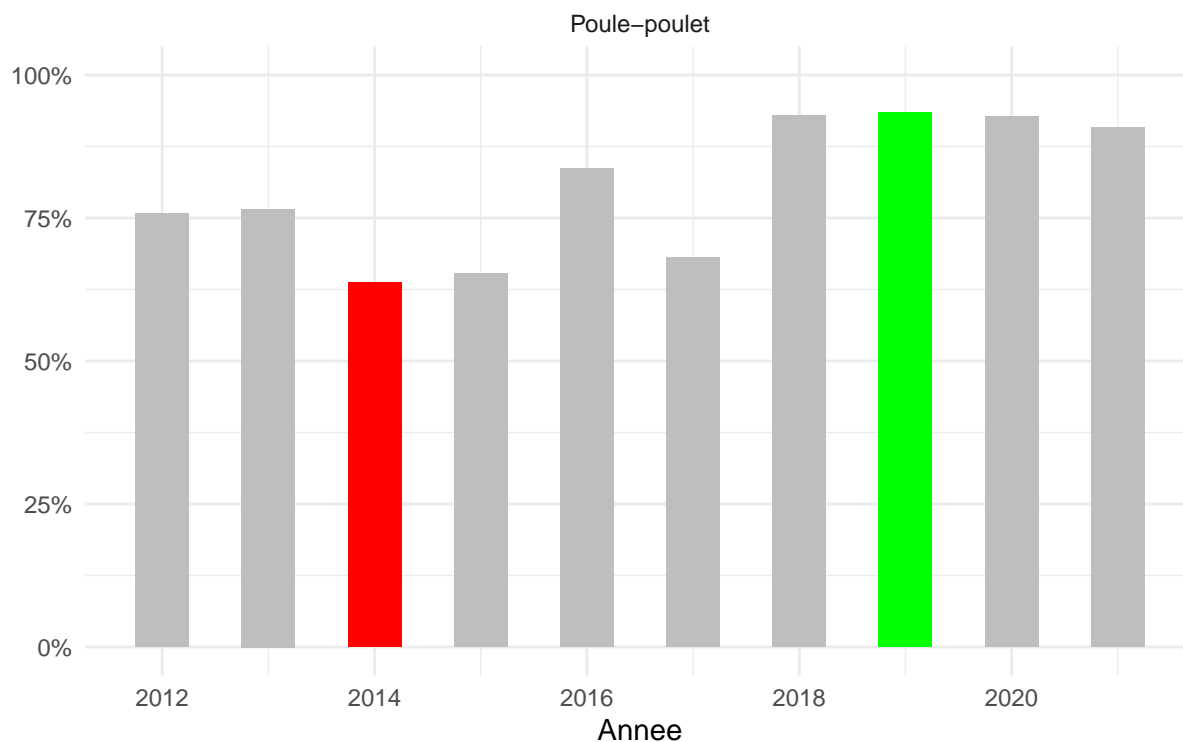
Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 60.1%



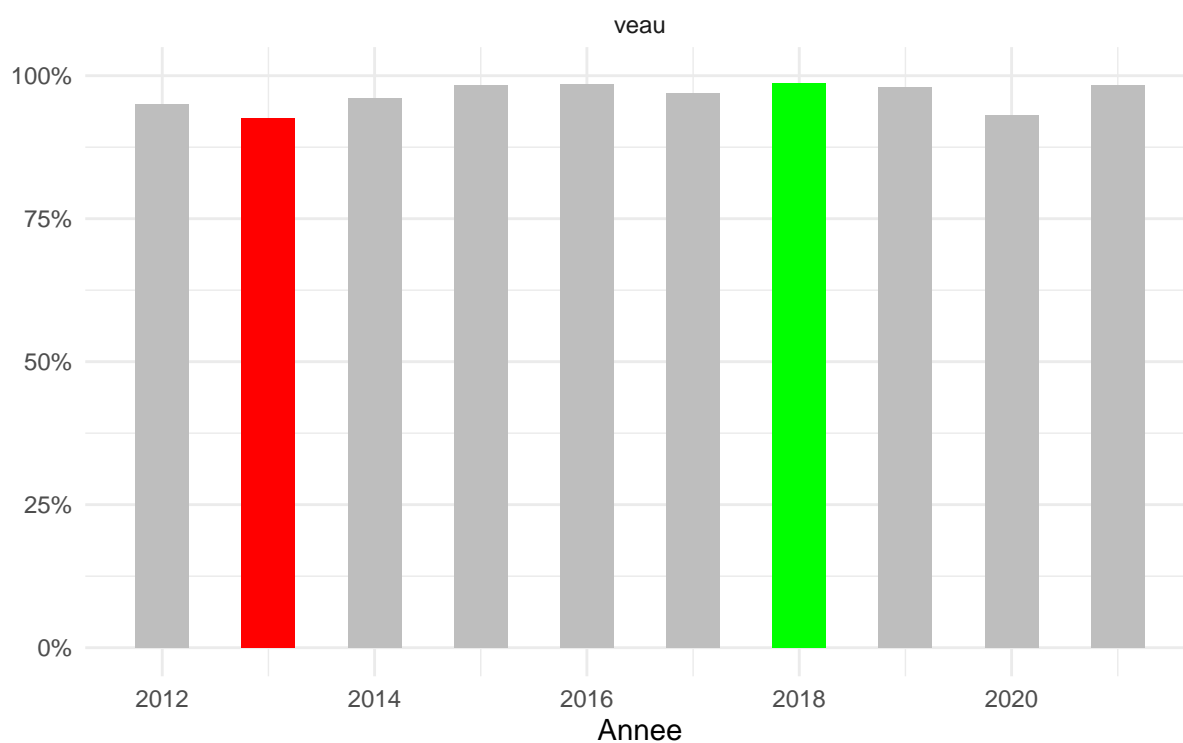
Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 89.38%



Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 80.31%



Fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu être calculée
En moyenne, le taux est de 96.59%



Hormis pour le bovin-abattoir à l'année 2019, dans ce cas le profil FQ-R semble être plus sensible que les

autres (effectif plus grand), aucun antibiotique ne semble présenter une résistance plus forte ce qui aurait engendrer une grosse différence dans les effectifs. Il n'y a donc pas de grande perte d'effectif suite aux AB testés.

```
table_join<-table_join%>%
  mutate(taux=(eff_pansensible/eff_C3G)*100)

# Boxplot pour représenter la variation intra-groupe

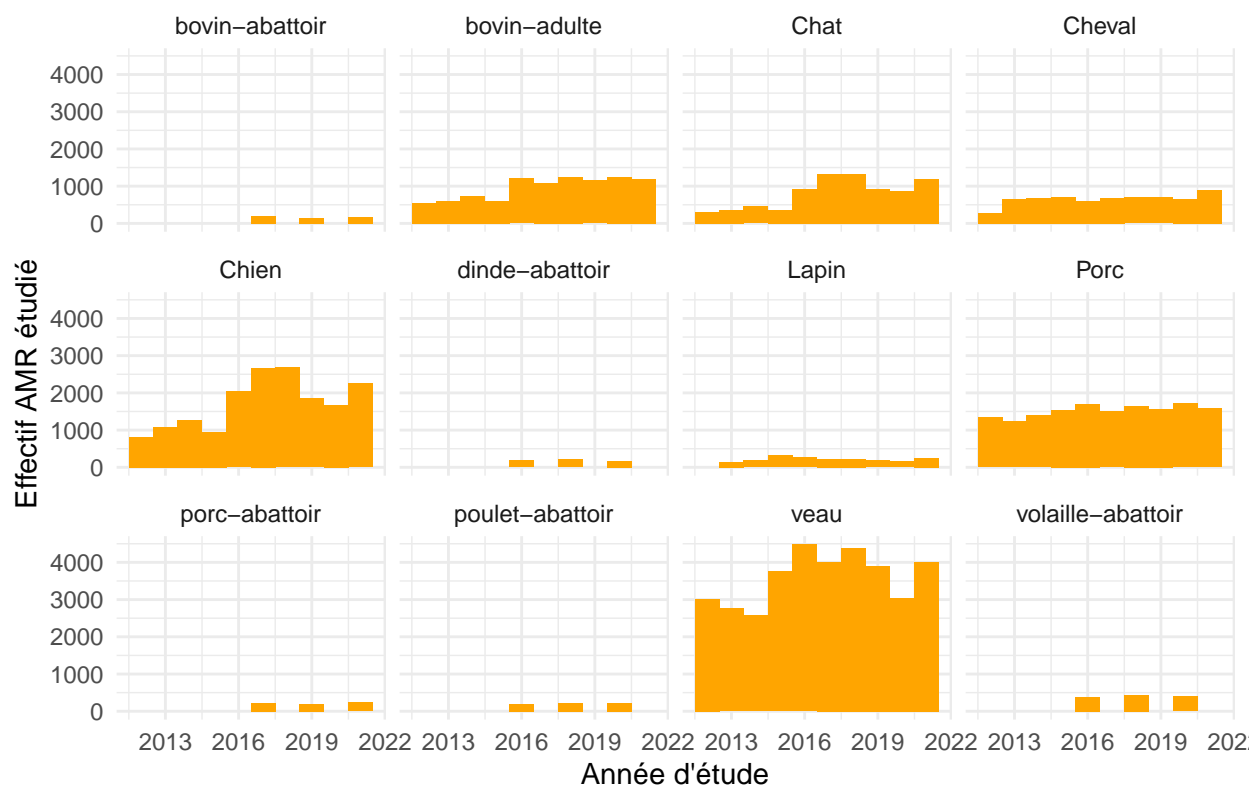
boxplot_taux<-ggplot(table_join, aes(x = reorder(population, taux), y = taux)) +
  geom_boxplot() +
  labs(title = "Variation de la fraction d'isolats bactériens pour lesquels la pan-sensibilité a pu \nê
        x = "Groupe",
        y = "taux de représentation",
        caption = str_glue("Cohorte : population animale au niveau National\nLecture : On remarque un ta
  scale_y_continuous(labels =scales::number_format(suffix=" %"),limits=c(0,100))
```

(dans Pour_rapport) Pourquoi ne pas étudier les populations animales au niveau régional?

```
effectifs_totaux_animale<-effectifs_totaux%>%
  filter(famille==unique(AMR$famille)[1])

ggplot(data=effectifs_totaux_animale, aes(x = annee, y = effectif_total_national)) +
  geom_col(width = 1, fill = "orange") +          # tracer le nombre de cas sous forme de colonnes
  theme_minimal()+                                # simplifier les arrière-plans
  labs(                                             # ajouter les noms d'axes, titres ...
    x = "Année d'étude",
    y = "Effectif AMR étudié",
    title = "Effectif total AMR par population animale en fonction de l'année") +
  facet_wrap(~population)
```

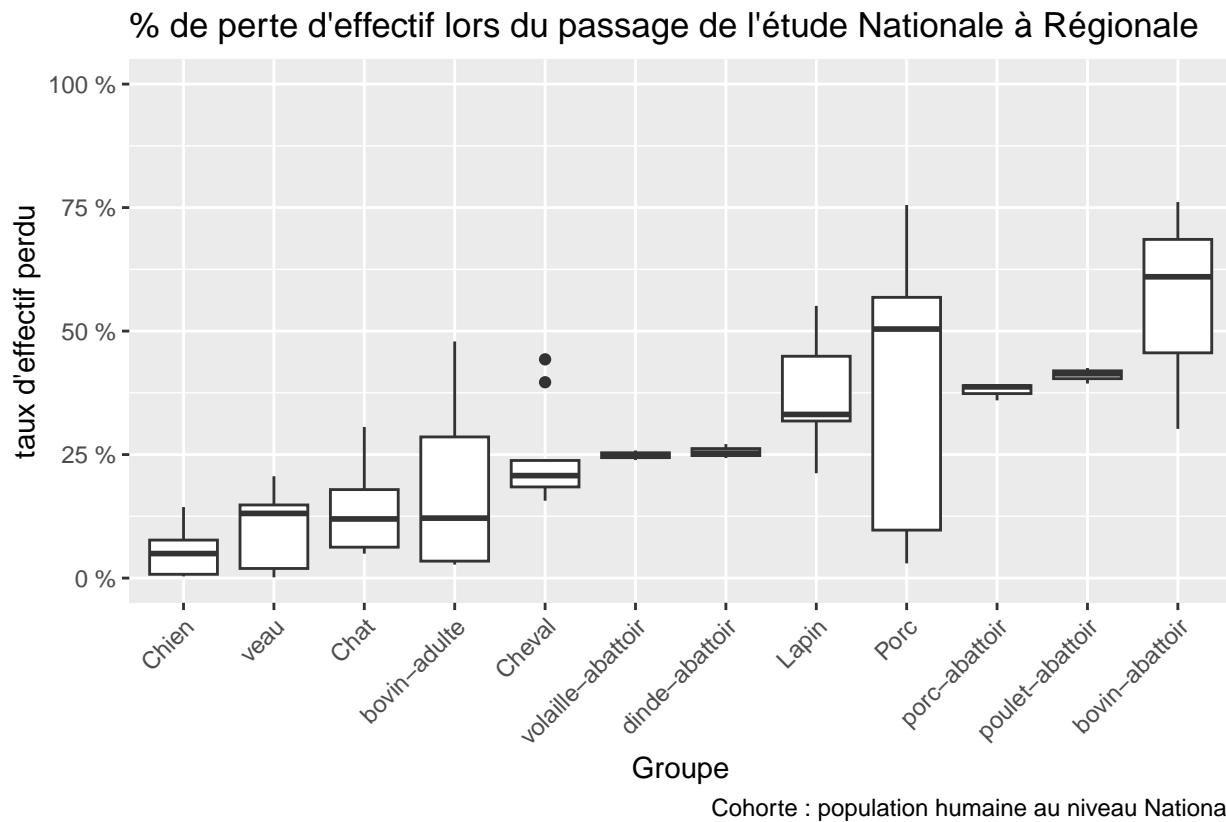
Effectif total AMR par population animale en fonction de l'année



```
effectifs_totaux_animale<-effectifs_totaux_animale%>%
  mutate(taux_eff_perte=(diff_effectif/effectif_total_national)*100)

boxplot_taux<-ggplot(effectifs_totaux_animale, aes(x = reorder(population, taux_eff_perte), y = taux_eff_perte)) +
  geom_boxplot() +
  labs(title = "% de perte d'effectif lors du passage de l'étude Nationale à Régionale",
       x = "Groupe",
       y = "taux d'effectif perdu",
       caption = str_glue("Cohorte : population humaine au niveau National\nLecture : ")) +
  scale_y_continuous(labels = scales::number_format(suffix=" %"),limits=c(0,100)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

boxplot_taux
```



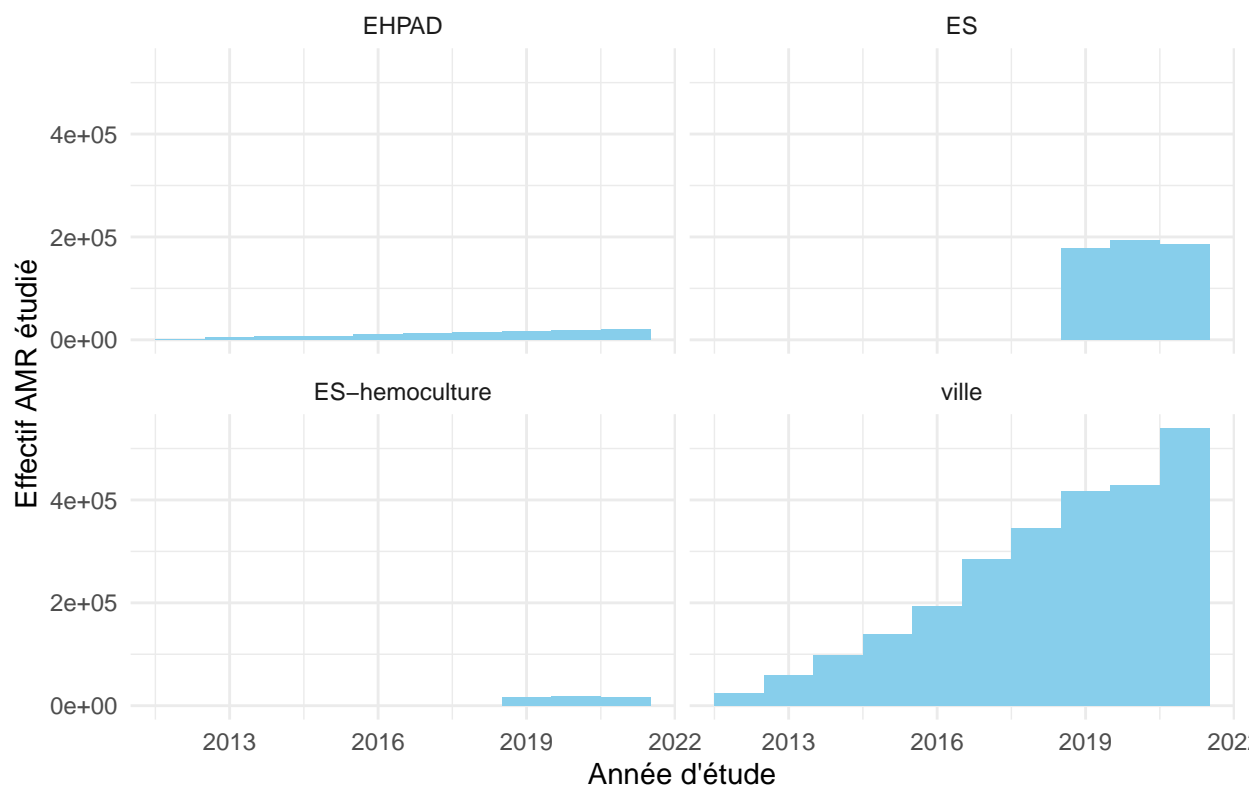
Pour les populations animales, la perte est assez élevée donc une étude régionale constitue un risque de perte d'informations.

(dans Pour_rapport) Etude de la population humaine

```
effectifs_totaux_humain<-effectifs_totaux%>%
  filter(famille==unique(AMR$famille)[2])

ggplot(data=effectifs_totaux_humain, aes(x = annee, y = effectif_total_national)) +
  geom_col(width = 1, fill = "skyblue") +      # tracer le nombre de cas sous forme de colonnes
  theme_minimal()+                             # simplifier les arrière-plans
  labs(                                         # ajouter les noms d'axes, titres ...
    x = "Année d'étude",
    y = "Effectif AMR étudié",
    title = "Effectif total AMR par population humaine en fonction de l'année") +
  facet_wrap(~population)
```

Effectif total AMR par population humaine en fonction de l'année



La différence entre les effectifs nationaux et régionaux pour les population humaines sont de l'ordre de 1% et constitue donc une part négligeable à l'étude. L'étude au niveau régionale ne constitue donc pas un risque en terme de perte d'information.

Les situations et les variables disponibles sont assez contrastées ?

L'indicateur est-il lié à l'exposition aux antimicrobiens (= la pression sélective) ?