

Peer-review - Bâtiments de Troyes Champagne Métropole

Maman j'ai raté la vizu

Introduction

Étude des bâtiments de Troyes

Données

Le jeu de données « Bâtiments de Troyes Champagne Métropole » recense l'ensemble des bâtiments situés dans l'agglomération Troyes Champagne Métropole, à des fins variées telles que l'urbanisme, l'aménagement du territoire, les analyses foncières ou encore la gestion des services publics. Il a été publié en août 2024 sur la plateforme data.gouv.fr.

Ce jeu de données provient de l'unification des bases BDTOPO® et BD PARCELLAIRE® de l'IGN. Il comprend 107 946 observations, avec très peu de valeurs manquantes, ce qui en fait une base fiable et complète pour l'analyse spatiale.

Nous avons choisi ce jeu de données, car il nous semble pertinent d'étudier la répartition géographique des bâtiments dans la ville de Troyes. Afin d'enrichir notre analyse, nous avons décidé d'y associer un second jeu de données : la Base Adresse Nationale (BAN), qui recense les adresses géolocalisées des bâtiments sur l'ensemble du territoire français (nous utiliserons uniquement les adresses du département nous concernant).

Origines et format des données

La base de données bâtiments est disponible ici. Ce jeu de données a été commandé par “Troyes Champagne Métropole” et a été mise à jour dernièrement en août 2024 dans le but de regrouper les entités géographiques de la métropole troyenne. Les données proviennent elles-mêmes de deux sources différentes qui ont été fusionnées : BDTOPO® et BD PARCELLAIRE®, les deux étant produits par IGN.

- La BD PARCELLAIRE® fournit l'information cadastrale numérique, géo-référencée et continue sur l'ensemble du territoire français. Elle est réalisée à partir de l'assemblage du plan cadastral dématérialisé. Cependant, elle n'est plus entretenue depuis 2019.
- La BDTOPO® est maintenu depuis 2019, elle couvre de manière cohérente l'ensemble des entités géographiques et administratives du territoire national.

Et la base de données des adresses, un csv par département, est ici (adresses-10.csv que nous avons renommé adresses_aube.csv). Les données proviennent de la BAN. Ces données sont initialement destiné aux services d'urgence pour qu'ils puissent se diriger. Elle est aussi destiné au raccord aux réseaux d'énergies ou de communication ou encore pour des analyses cartographiques précises. Sa constitution est copilotée par l'ANCT, la DINUM et l'IGN.

Les liens, quant à eux sont disponibles sur ce lien. Produits aussi par IGN, c'est une base de données qui permet de lier la BAN à son environnement géographique.

Les variables

Dans notre étude, nous allons considérer 16 variables.

BATI Nous avons sélectionné 14 variables dans le 1er jeu de données.

Nom	Type	Description
NATURE	nominal	architecture du bâtiment
USAGE1	nominal	utilisation du bâtiment (agricole, résidentiel, religieux...)
USAGE2	nominal (optionnel)	utilisation du bâtiment
LEGER	discrète	structure légère ou pas
DATE_APP	discrète	date d'apparition/construction (pas toujours présent)
ACQU_PLANI	nominal	de quelle base provient l'information
NB_LOGTS	discrète	nombre de logements dans le bâtiment
NB_ETAGES	discrète	nombre d'étages du bâtiment
MAT_MURS	discrète	code des matériaux des murs (à croiser avec le csv materiaux_murs.csv)
MAT_TOITS	discrète	code des matériaux de la toiture (à croiser avec le csv materiaux_toits.csv)
HAUTEUR	continue	hauteur du bâtiment
Z_SOL	continue	altitude du sol
Z_TOIT	continue	altitude du toit
ETAT	nominal	état du bâtiment (en projet, en construction, en service, en ruine)

ADRESSES Pour la localisation, nous utiliserons 2 variables : la longitude et la latitude.

Nom	Type	Description
LONGITUDE	continue	coordonnée GPS
LATITUDE	continue	coordonnée GPS

Pourquoi avoir choisi ces variables Nous avons enlevé les variables qui nous semblaient inutiles ainsi que celles comportant peu de valeurs. Nous avons aussi enlevé les variables que nous n'avons pas trouvées pertinentes.

Plan d'analyse

1. Comment la ville de Troyes s'est étendue géographiquement au fil des années ?

Il s'agit ici de visualiser l'évolution spatiale de la ville de Troyes au fil du temps. Pour cela, nous utiliserons la variable `date_app` et les variables de coordonnées géographiques : `latitude` et `longitude`.

Comme types de visualisations à envisager, nous avons :

- Une carte avec les bâtiments colorés selon leur date (période) de construction

Les problèmes potentiels :

- l'absence de date de construction pour certains bâtiments

2. Y a-t-il un lien entre les matériaux de construction et les dates de construction ?

Il s'agit ici de déterminer s'il existe une relation entre les variables `date_app`, `mat_murs` et `mat_toits`. L'objectif est de vérifier si les matériaux de construction des bâtiments varient en fonction des époques.

Pour cela, nous envisageons deux types de visualisations :

- Un line chart pour suivre l'évolution du nombre de bâtiments par matériau au fil du temps :
 - L'axe des abscisses représenterait les années
 - L'axe des ordonnées représenterait le nombre de bâtiments
 - Chaque ligne représenterait un matériau
- Un violon plot pour étudier la distribution des dates selon chaque matériau :
 - L'axe des abscisses représenterait les années
 - L'axe des ordonnées représenterait les matériaux de construction

Les problèmes potentiels sont :

- un grand nombre de catégories de matériaux (risque de graphe illisible)

3. Y a-t-il une corrélation entre les matériaux de construction et la hauteur du bâtiment ?

L'objectif ici est de comprendre si l'évolution des matériaux a un impact sur la hauteur des bâtiments. Nous aimeraisons comparer la distribution des hauteurs en fonction des matériaux et vérifier si certains matériaux sont associés à des hauteurs spécifiques. Nous utiliserions de nouveau les variables `mat_murs` et `mat_toits` que nous mettrions en relation avec la variable `hauteur`.

Pour cela, nous envisageons d'utiliser :

- Un violon plot ou box plot :
 - L'axe des abscisses représentant le type de matériau
 - L'axe des ordonnées représentant la hauteur des bâtiments

4. Existe-t-il une relation entre la localisation géographique et l'usage des bâtiments ?

L'objectif ici est de déterminer si certains types d'usage des bâtiments sont spécifiquement localisés dans certaines zones de la ville de Troyes. Pouvons-nous identifier des zones résidentielles, des zones industrielles et des zones commerciales ? Nous utiliserons les variables de coordonnées géographiques (`longitude` et `latitude`) et les variables `usage1` et `usage2`.

Pour cela, nous envisageons :

- Une carte avec les bâtiments colorés selon leur usage

Les problèmes potentiels :

- il sera peut-être nécessaire de regrouper les usages

5. Y a-t-il une liaison entre la hauteur des bâtiments et leur usage ?

Il s'agit de déterminer s'il existe une relation entre la hauteur d'un bâtiment et son usage. Les variables utilisées sont : la variable `hauteur` et la variable `usage1`.

Pour cela, nous envisageons d'utiliser :

- Un diagramme à barres empilées pour étudier la répartition des usages :
 - L'axe des abscisses représenterait les tranches de hauteur
 - L'axe des ordonnées représenterait le nombre de bâtiments
 - Chaque barre avec une répartition des usages

Exploration

Import des packages et des données

```
library(dplyr, warn.conflicts = FALSE)
library(ggplot2)
library(patchwork)
library(readr)
library(sf)

## Linking to GEOS 3.13.0, GDAL 3.10.1, PROJ 9.5.1; sf_use_s2() is TRUE

library(viridis, quietly = TRUE)

batiments = read_csv("../data/batiments_troyes_coords.csv", col_types = cols(
  ID = col_character(),
  NATURE = col_character(),
  USAGE1 = col_character(),
  USAGE2 = col_character(),
  LEGER = col_character(),
  ETAT = col_character(),
  DATE_CREAT = col_datetime(),
  DATE_MAJ = col_datetime(),
  DATE_APP = col_date(),
  DATE_CONF = col_date(),
  SOURCE = col_character(),
  ID_SOURCE = col_character(),
  ACQU_PLANI = col_character(),
  PREC_PLANI = col_double(),
  ACQU_ALTI = col_character(),
  PREC_ALTI = col_double(),
  NB_LOGTS = col_double(),
  NB_ETAGES = col_double(),
  MAT_MURS = col_character(),
  MAT_TOITS = col_character(),
  HAUTEUR = col_double(),
  Z_MIN_SOL = col_double(),
  Z_MIN_TOIT = col_double(),
```

```

    Z_MAX_TOIT = col_double(),
    Z_MAX_SOL = col_double(),
    ORIGIN_BAT = col_character(),
    APP_FF = col_character(),
    LONGITUDE = col_double(),
    LATITUDE = col_double()
))

materiaux_murs = read_csv("../data/materiaux_murs.csv", col_types = cols(
  code = col_character(),
  valeur = col_character()
))
materiaux_toits = read_csv("../data/materiaux_toits.csv", col_types = cols(
  code = col_character(),
  valeur = col_character()
))

batiments = batiments %>%
  mutate(LEGER = if_else(LEGER == "Oui", TRUE, FALSE))

batiments = batiments %>%
  left_join(materiaux_murs, by = c("MAT_MURS" = "code"),
            relationship = "many-to-one") %>%
  select(-MAT_MURS) %>%
  rename(MAT_MURS = valeur)

batiments = batiments %>%
  left_join(materiaux_toits, by = c("MAT_TOITS" = "code"),
            relationship = "many-to-one") %>%
  select(-MAT_TOITS) %>%
  rename(MAT_TOITS = valeur)

batiments = batiments %>%
  select(c(NATURE, USAGE1, USAGE2, LEGER, DATE_APP, ACQU_PLANI, NB_LOGTS,
          NB_ETAGES, MAT_MURS, MAT_TOITS, HAUTEUR, Z_MIN_SOL, Z_MAX_TOIT,
          ETAT, LONGITUDE, LATITUDE))

head(batiments)

## # A tibble: 6 x 16
##   NATURE  USAGE1 USAGE2 LEGER DATE_APP ACQU_PLANI    NB_LOGTS NB_ETAGES MAT_MURS
##   <chr>   <chr>  <chr>  <lgl> <date>   <chr>        <dbl>      <dbl> <chr>
## 1 Indiffé~ Indif~ <NA>   TRUE   NA       BDParcellai~      NA        NA <NA>
## 2 Indiffé~ Indif~ <NA>   TRUE   NA       BDParcellai~      NA        NA <NA>
## 3 Indiffé~ Indif~ <NA>   TRUE   NA       BDParcellai~      NA        NA <NA>
## 4 Industr~ Indus~ <NA>  FALSE  NA       BDParcellai~      NA        NA <NA>
## 5 Indiffé~ Indif~ <NA>  FALSE  NA       BDParcellai~      NA        NA <NA>
## 6 Indiffé~ Indif~ <NA>   TRUE   NA       BDParcellai~      NA        NA <NA>
## # i 7 more variables: MAT_TOITS <chr>, HAUTEUR <dbl>, Z_MIN_SOL <dbl>,
## #   Z_MAX_TOIT <dbl>, ETAT <chr>, LONGITUDE <dbl>, LATITUDE <dbl>
```

1. Lien entre la hauteur des bâtiments et leur usage.

Il s'agit de déterminer s'il existe une relation entre la hauteur d'un bâtiment et son usage. Les variables utilisées sont : la variable `hauteur` et la variable `usage1`.

Résultats attendus :

Nous pensons que :

- Les bâtiments résidentiels seraient majoritairement de petite ou moyenne hauteur
- Les bâtiments à usage commercial ou industriel pourraient être de hauteur élevée

Visualisation

```
batiments1 <- batiments

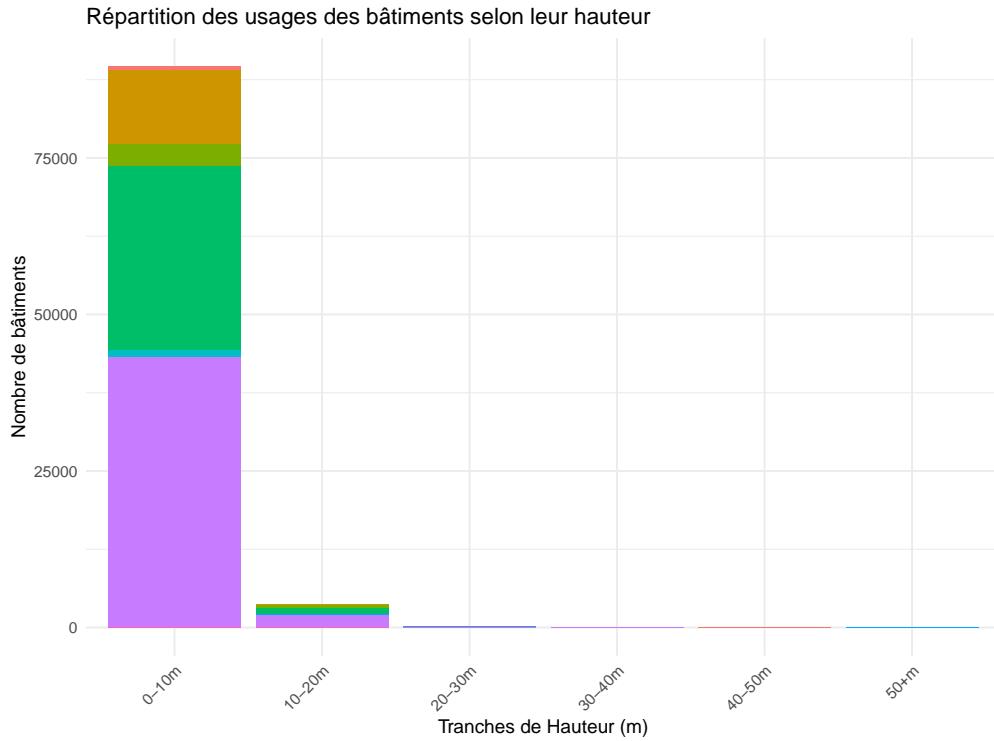
# visualisation

##créer les tranches de hauteur
batiments1$hauteur_tranche <- cut(batiments$HAUTEUR,
                                      breaks = c(0, 10, 20, 30, 40, 50, 100),
                                      labels = c("0-10m", "10-20m", "20-30m",
                                                "30-40m", "40-50m", "50+m"),
                                      right = FALSE)

#
# Calculer la répartition des usages par tranche de hauteur
repartition_usage <- batiments1 %>%
  group_by(hauteur_tranche, USAGE1) %>%
  summarise(nombre_batiments = n(), .groups = "drop")

#
repartition_usage <- repartition_usage %>%
  filter(!is.na(hauteur_tranche))

#diagramme
ggplot(repartition_usage, aes(x = hauteur_tranche, y = nombre_batiments, fill = USAGE1)) +
  geom_bar(stat = "identity") +
  labs(title = "Répartition des usages des bâtiments selon leur hauteur",
       x = "Tranches de Hauteur (m)",
       y = "Nombre de bâtiments") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Interprétation des résultats

On remarque que la plupart des bâtiments ont une hauteur inférieure à 10 m. Cela peut être lié au fait que la ville de Troyes n'a pas subi une urbanisation dense. Ensuite, on remarque que la plupart des bâtiments sont à usage résidentiel.

Contrairement à l'hypothèse de départ, la plupart des bâtiments à usage commercial ont une hauteur inférieure à 10 m. Cela s'explique par le fait que les commerces sont intégrés en rez-de-chaussée des bâtiments résidentiels.

2. Expansion de la ville de Troyes au fil du temps

Nous voulons voir si les bâtiments ont été construits de plus en plus loin du centre de Troyes au fil des années ou si cela a peu d'influence.

Les variables utilisées sont : `date_app`, `longitude` et `latitude`.

Nous pensons que :

- Les bâtiments dans le centre sont les plus vieux
- Il y a des vieux bâtiments aussi autour de Troyes

Visualisation

```
batiments_sf = st_as_sf(batiments, coords = c("LONGITUDE", "LATITUDE"), crs = 4326)
batiments_sf = mutate(batiments_sf, DATE_APP = as.double(substr(DATE_APP, 1, 4)))
periodes = c("Avant 1600", "1600-1700", "1700-1800",
```

```

        "1800-1900", "1900-2000", "Depuis 2000")
batiments_sf = mutate(batiments_sf, PERIODE = case_when(
    DATE_APP < 1600 ~ "Avant 1600",
    1600 <= DATE_APP & DATE_APP < 1700 ~ "1600-1700",
    1700 <= DATE_APP & DATE_APP < 1800 ~ "1700-1800",
    1800 <= DATE_APP & DATE_APP < 1900 ~ "1800-1900",
    1900 <= DATE_APP & DATE_APP < 2000 ~ "1900-2000",
    2000 <= DATE_APP ~ "Depuis 2000",
    .default = "Inconnu"
))
couleurs = c(
    "Avant 1600" = "#b2182b",
    "1600-1700" = "#ef8a62",
    "1700-1800" = "#fddbc7",
    "1800-1900" = "#d1e5f0",
    "1900-2000" = "#67a9cf",
    "Depuis 2000" = "#2166ac"
)
batiments_sf = filter(batiments_sf, PERIODE != "Inconnu")
batiments_sf = mutate(batiments_sf, PERIODE = factor(PERIODE, levels = periodes))
bbox = st_bbox(batiments_sf)

plot_périodes = function(df, period_list) {
    df = filter(df, PERIODE %in% period_list)

    ggplot(df) +
    geom_sf(aes(color = PERIODE), size = 0.2) +
    scale_color_manual(values = couleurs) +
    coord_sf(
        xlim = c(bbox["xmin"], bbox["xmax"]),
        ylim = c(bbox["ymin"], bbox["ymax"]))
    ) +
    theme_minimal() +
    theme(
        legend.position = "none",
        axis.text.x = element_text(angle = 45, hjust = 1),
        aspect.ratio = 0.8
    )
}

plots = lapply(1:length(periodes), function(i) {
    plot_périodes(batiments_sf, periodes[1:i])
})

final_plot = wrap_plots(plots, ncol = 2) +
    plot_layout(guides = "collect") +
    plot_annotation(
        title = "Évolution spatiale de Troyes par périodes de construction",
        subtitle = "Les axes sont la longitude et la latitude",
        caption = "Données : Troyes Champagne Métropole"
    ) +
    labs(color = "Période de construction") +
    theme(

```

```

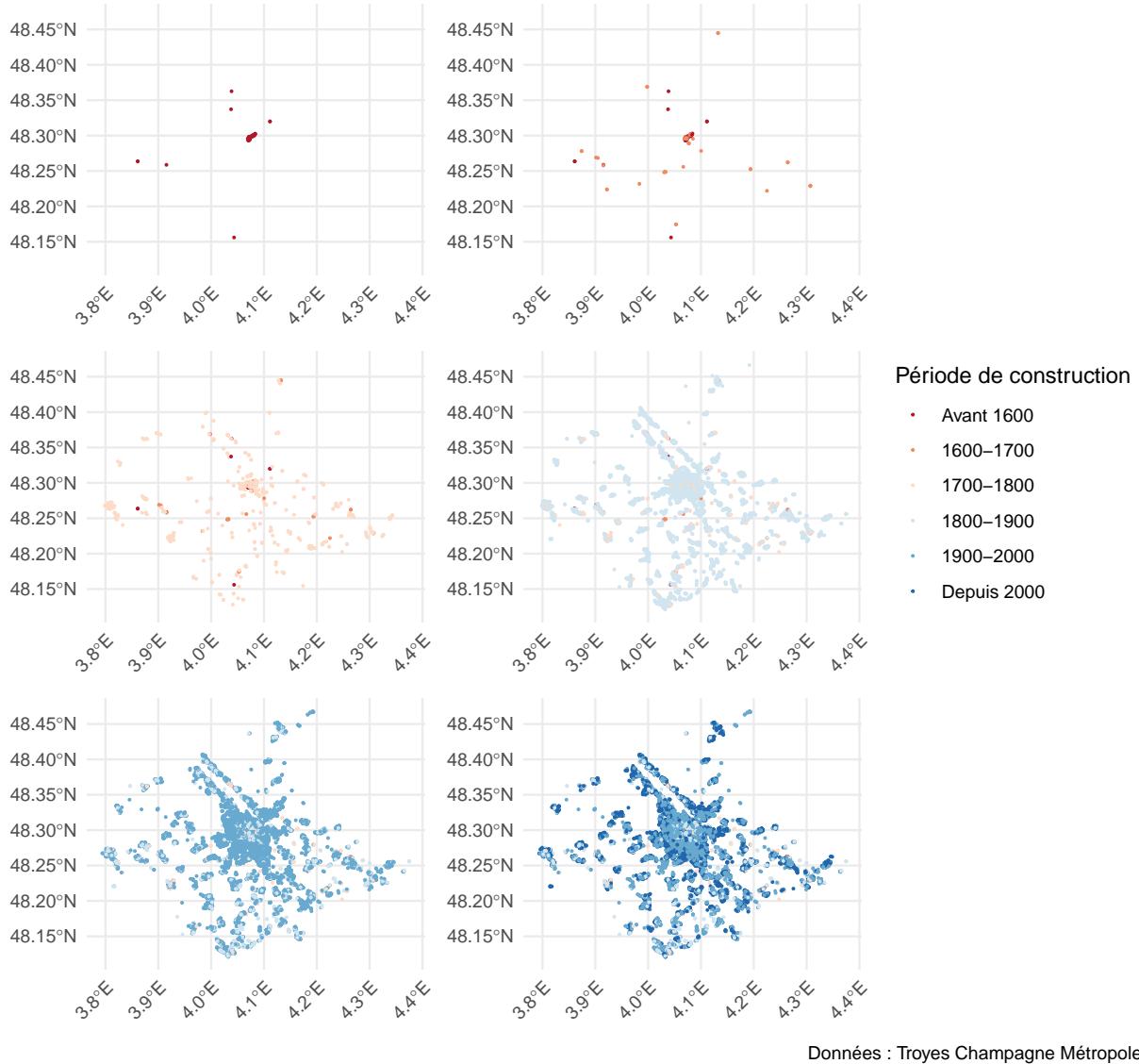
        legend.position = "right",
        legend.justification = "center"
    )

final_plot

```

Évolution spatiale de Troyes par périodes de construction

Les axes sont la longitude et la latitude



Interprétation des résultats

Il est difficile de connaître la date de construction des bâtiments qui ont été construits avant 1600, d'où le manque d'information. J'ai dû masquer tous les bâtiments dont on a pas la date de construction car cela nuisait à la lisibilité du graphique.

On remarque la construction de beaucoup de bâtiments entre 1700 et 1800, puis une assez grosse expansion

entre 1800 et 1900. On remarque aussi que l'agglomération de Troyes avait sa forme assez tôt et même si elle s'est densifiée, elle ne s'est pas vraiment étendue.

3. Y a-t-il une corrélation entre les matériaux de construction et la hauteur du bâtiment ?

L'objectif ici est de comprendre si l'évolution des matériaux, principalement des murs a un impact sur la hauteur des bâtiments. Pour cela, nous allons réaliser deux violinplot, l'un lié au matériau des murs et l'autre des toits afin de confirmer une corrélation.

Les variables utilisées sont :

- MAT_MURS : variable nominale : code des matériaux des murs, croisé avec materiaux_murs.csv
- MAT_TOITS : variable nominale : code des matériaux de la toiture, croisé avec materiaux_toits.csv
- HAUTEUR : variable continue représentant la hauteur absolue du bâtiment

Nous nous attendons à ce que les matériaux plus modernes et denses comme le béton (au niveau des murs et aussi de la toiture) composent les bâtiments les plus hauts de l'agglomération troyenne. De plus nous pensons que le matériau des murs aura plus d'influence que celui de la toiture.

Visualisation

```
rm(list = setdiff(ls(), c("batiments")))

batiments_clean <- batiments %>%
  filter(!is.na(MAT_MURS), !is.na(HAUTEUR)) %>%
  mutate(MAT_MURS = as.factor(MAT_MURS))

# Violin plot basé sur MAT_MURS
plot_murs <- ggplot(batiments_clean,
                     aes(x = MAT_MURS, y = HAUTEUR, fill = MAT_MURS)) +
  geom_violin(trim = FALSE, color = "black") +
  scale_fill_viridis_d(option = "C") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none") +
  labs(
    title = "Hauteur selon le matériau des murs",
    x = "Matériau des murs",
    y = "Hauteur (m)"
  )

# Violin plot basé sur MAT_TOITS pour varier
batiments_toits_clean <- batiments %>%
  filter(!is.na(MAT_TOITS), !is.na(HAUTEUR)) %>%
  mutate(MAT_TOITS = as.factor(MAT_TOITS))

plot_toits <- ggplot(batiments_toits_clean,
                      aes(x = MAT_TOITS, y = HAUTEUR, fill = MAT_TOITS)) +
  geom_violin(trim = FALSE, color = "black") +
  scale_fill_viridis_d(option = "D") +
```

```

theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none") +
  labs(
    title = "Hauteur selon le matériau des toits",
    x = "Matériau des toits",
    y = "Hauteur (m)"
  )

# Regroupement des deux plots
plots <- list(plot_murs, plot_spacer(), plot_toits)

# Assemblage final
final_plot = wrap_plots(plots, ncol = 1, nrow = 3) +
  plot_layout(guides = "collect", heights = c(1, 0.2, 1)) +
  plot_annotation(
    title = "Analyse des hauteurs de bâtiments selon matériaux",
    subtitle = "Ville de Troyes",
    caption = "Données : BDD PARCELAIRES - Troyes Champagne Métropole"
  )

final_plot

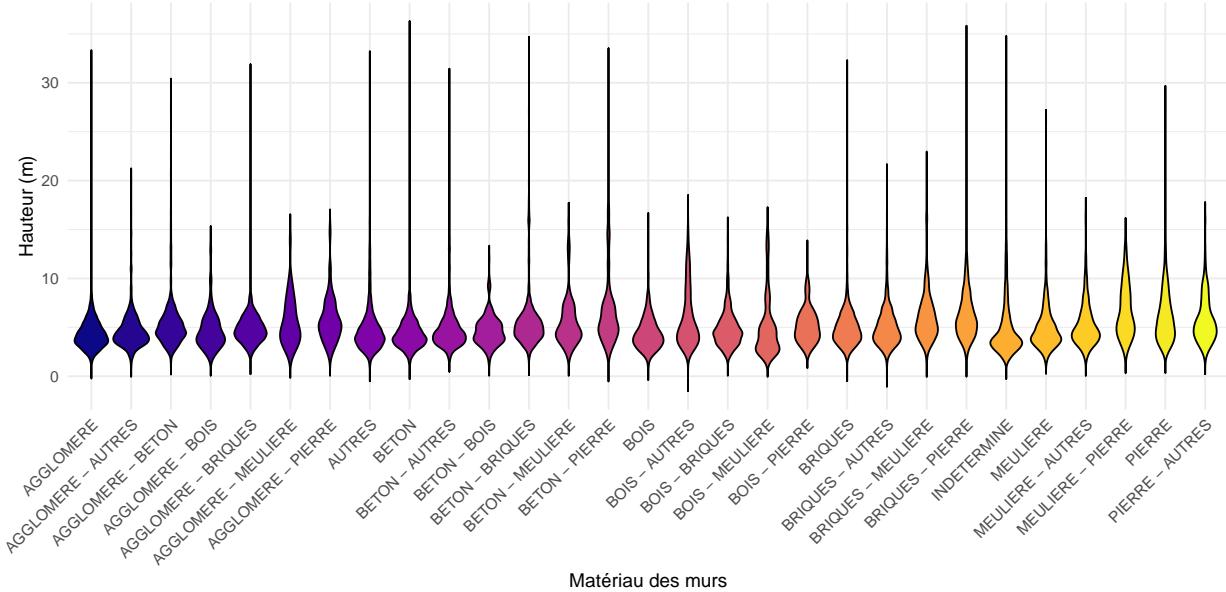
## Warning: Groups with fewer than two datapoints have been dropped.
## i Set 'drop = FALSE' to consider such groups for position adjustment purposes.

```

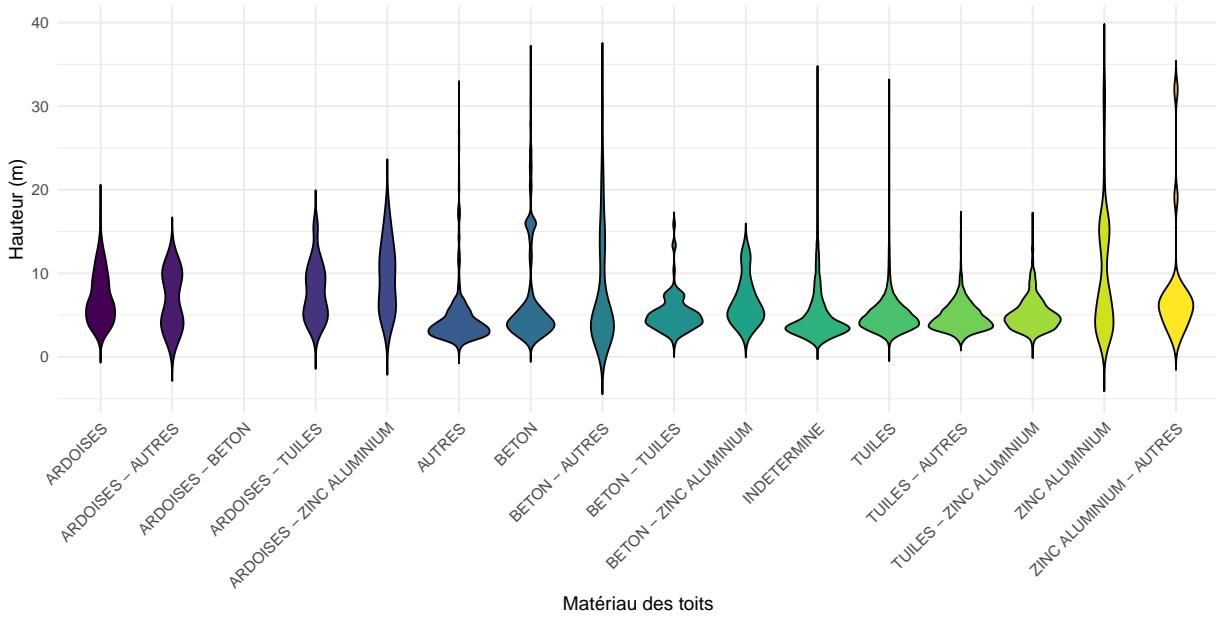
Analyse des hauteurs de bâtiments selon matériaux

Ville de Troyes

Hauteur selon le matériau des murs



Hauteur selon le matériau des toits



Données : BDD PARCELAIRE – Troyes Champagne Métropole

Interprétation des résultats

Tout d'abord, quand l'on regarde le graphique de la hauteur par rapport aux matériaux des murs, il est difficile d'observer une corrélation. En effet, on remarque de très faibles différences dans la distribution des hauteurs pour chaque combinaison de matériau. Cela pourrait amener à penser que le matériau des

murs n'a pas une influence déterminante sur la hauteur des bâtiments, du moins d'après cette visualisation. Une approche complémentaire, comme représenter la hauteur moyenne ou médiane pour chaque matériau, pourrait nous apporter des réponses plus claires et mettre en évidence des tendances qui peuvent être plus légères.

Cependant, quand l'on observe le graphique lié à la hauteur des bâtiments selon le matériau de la toiture, on peut clairement observer que les matériaux ZINC ALUMINIUM, BETON et ARDOISES recouvrent les bâtiments les plus hauts, alors que les TUILES ne dépassent que très rarement la dizaine de mètres de hauteur. Cela indique une possible corrélation entre le type de matériau utilisé pour la toiture et la hauteur des bâtiments.

En conclusion, la question est partiellement répondue et les résultats diffèrent sensiblement de nos attentes initiales. On remarque une corrélation significative entre les matériaux de toiture et la hauteur des bâtiments, alors que celle entre les matériaux des murs et la hauteur est beaucoup plus difficile à établir sans pour autant pouvoir l'écartier du champ des possibilités. Cela soulève donc une nouvelle question : pourquoi le matériau de la toiture semble-t-il mieux refléter la hauteur d'un bâtiment que celui des murs ?

4. Peut-on voir des matériaux de construction des murs de prédilection selon les époques ?

L'objectif ici est de comprendre si l'utilisation des matériaux changent selon les époques. Pour cela nous allons réaliser un violinplot lié matériau des murs.

Les variables utilisées sont :

- MAT_MURS: variable nominal: code des matériaux des murs, croisé avec materiaux_murs.csv
- DATE_APP: variable discrète d'apparition des bâtiments

Nous nous attendons à ce que les matériaux les plus solides de chaque époque soit utilisés successivement.

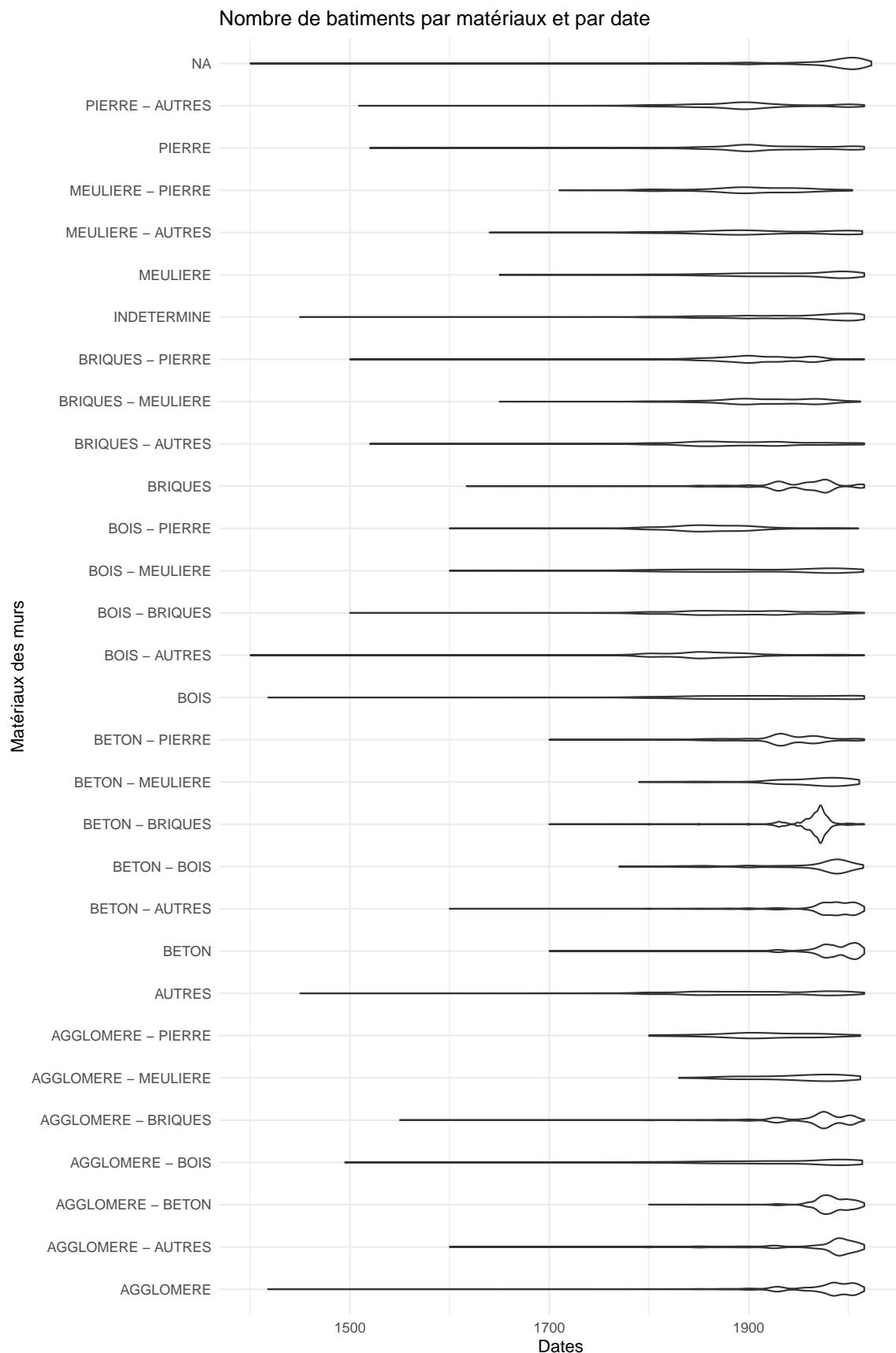
Visualisation

```
# Récupération des variables utilisées
batiments1 = batiments %>%
  select(DATE_APP,MAT_MURS)

# Nettoyage des données
batiments1 = batiments1 %>%
  filter(!is.na(MAT_MURS) | MAT_MURS != "" | !is.na(DATE_APP) | DATE_APP != "")

# Création du graphique
ggplot(batiments, aes(x = DATE_APP, y=MAT_MURS)) +
  geom_violin() +
  labs(title = "Nombre de bâtiments par matériaux et par date",
       x = "Dates",
       y = "Matériaux des murs") +
  theme_minimal()

## Warning: Removed 51774 rows containing non-finite outside the scale range
## ('stat_ydensity()' ).
```



Interprétation des résultats

Tout d'abord, il faut préciser que nous manquant de beaucoup d'information, particulièrement sur les années avant 1800. Il est donc impossible de généraliser les informations du graphe. De plus, un certain nombre de bâtiments n'ont pas de matériaux connus. Nous passons donc de 107 946 observations à 59097 observations. Cependant, nous pouvons voir des tendances se dégager. Certains matériaux sont présent depuis avant 1600, comme l'agglomere et le bois. D'autres ont connu des pics de popularité récents, comme le béton, la brique et l'agglomere. On peut aussi remarqué l'utilisation massive de la pierre autours des années 1900 mais s'est rarifié depuis. Même si certains matériaux s'imposent à certaines époques, on peut voir qu'il y a une diverité de ceux-ci et beaucoup de mélange de matériaux.

Pour aller plus loin, on aurait pu faire le même graphique mais en ayant une distribution des matériaux par période en pourcentage, ce qui éviterait "l'écrasement" des données plus anciennes et moins nombre, mais ferait perdre du détail sur la chronologie.