

Arbres-gbif

François Pelletier

Une forêt de données:

Enrichir ses données à l'aide des catalogues de données ouvertes et des interfaces de programmation publiques

Ce projet utilise des données provenant du [catalogue de données de la Ville de Québec](#). Ces données sont disponibles sous la [version 4.0 de la licence Creative Commons Attribution](#).

L'objectif de ce projet est de tirer un maximum d'information du jeu de données [Arbres répertoriés](#).

Chargement du jeu de données

Le jeu de données contient 122825 observations de 16 variables.

Classe de chaque variable

Le tableau suivant décrit la classe attribuée par R à chaque variable.

| Variable | Classe |
|-------------|-----------|
| TYPE_LIEU | character |
| NOM_LAT | character |
| NOM_FR | character |
| TYPE_ARBRE | character |
| DIAMETRE | integer |
| POS_MESURE | character |
| MULTI_TRONC | character |
| DATE_PLANTE | POSIXct |
| TYPE_PROP | character |
| LONGITUDE | numeric |

| Variable | Classe |
|------------|-----------|
| LATITUDE | numeric |
| GENERIQUE | character |
| LIAISON | character |
| SPECIFIQUE | character |
| DIRECTION | character |
| NOM_TOPO | character |

Exploration de base du jeu de données

Description des variables qualitatives

Table 2: TYPE_LIEU

| Valeur | Fréquence | NA |
|----------------|-----------|-------|
| Voie publique | A | 73126 |
| Lieu public | B | 49698 |
| Non disponible | C | 1 |

Table 3: NOM_LAT

| Valeur | Fréquence | NA |
|---------------------------------|-----------|-------|
| Ulmus americana | A | 14351 |
| Acer platanoides | B | 10901 |
| Acer saccharinum | C | 7743 |
| Fraxinus pennsylvanica | D | 7363 |
| Quercus rubra | E | 4920 |
| Tilia cordata | F | 3509 |
| Syringa reticulata 'Ivory Silk' | G | 3487 |
| Picea glauca | H | 3004 |
| Ulmus 'Morton Accolade' | I | 2848 |
| Acer rubrum | J | 2628 |

Table 4: NOM_FR

| Valeur | Fréquence | NA |
|-----------------------------|-----------|-------|
| orme d'Amérique | A | 14351 |
| érable de Norvège | B | 10901 |
| érable argenté | C | 7743 |
| frêne rouge | D | 7363 |
| chêne rouge | E | 4920 |
| tilleul à petites feuilles | F | 3509 |
| lilas Japonais 'Ivory Silk' | G | 3487 |
| épinette blanche | H | 3004 |
| orme 'Morton Accolade' | I | 2848 |
| érable rouge | J | 2628 |

Table 5: TYPE_ARBRE

| Valeur | Fréquence | NA |
|----------------|-----------|--------|
| Feuille | A | 109765 |
| Conifère | B | 13043 |
| NON DISPONIBLE | C | 17 |

Table 6: POS_MESURE

| Valeur | Fréquence | NA |
|--------|-----------|-------|
| DHP | A | 87549 |
| DHS | B | 27426 |
| M | C | 7835 |
| | D | 15 |

Table 7: MULTI_TRONC

| Valeur | Fréquence | NA |
|--------|-----------|--------|
| N | A | 115222 |
| O | B | 7526 |
| ND | C | 77 |

Table 8: TYPE_PROP

| Valeur | Fréquence | NA |
|-------------------------------|-----------|-------|
| Public | A | 93804 |
| Privés | B | 14544 |
| Public:Terrain Privé | C | 3917 |
| Privés:OMHQ | D | 2102 |
| Public:Arbre mitoyen | E | 1681 |
| Privés:Frais Partagé | F | 1331 |
| Privés:C.S. de la Capitale | G | 900 |
| Public:Parc-École | H | 713 |
| Cas Spéciaux | I | 600 |
| Public:Entretenu Par La Ville | J | 578 |

Table 9: GENERIQUE

| Valeur | Fréquence | NA |
|---------------------|-----------|-------|
| Rue | A | 43252 |
| Parc | B | 21339 |
| Îlot | C | 18276 |
| Avenue | D | 14968 |
| Boulevard | E | 7651 |
| Chemin | F | 2276 |
| Piste cyclable | G | 2210 |
| Parc linéaire | H | 1944 |
| Bassin de rétention | I | 1507 |
| Parc-école | J | 1259 |

Table 10: LIAISON

| Valeur | Fréquence | NA |
|--------|-----------|-------|
| | A | 81197 |
| des | B | 10566 |
| de la | C | 9546 |
| du | D | 8353 |
| de | E | 7244 |
| de l' | F | 5240 |
| d' | G | 469 |
| aux | H | 60 |

| Valeur | Fréquence | NA |
|--------|-----------|----|
| l' | I | 56 |
| les | J | 51 |

Table 11: SPECIFIQUE

| Valeur | Fréquence | NA |
|------------------------|-----------|------|
| Rivière Saint-Charles | A | 1940 |
| Robert-Bourassa | B | 1214 |
| Corridor des Cheminots | C | 931 |
| Plage-Jacques-Cartier | D | 918 |
| Victoria | E | 917 |
| Exposition | F | 880 |
| Arboretum | G | 775 |
| Sainte-Foy | H | 723 |
| Saint-Louis | I | 688 |
| Pointe-aux-Lièvres | J | 571 |

Table 12: DIRECTION

| Valeur | Fréquence | NA |
|--------|-----------|--------|
| | A | 116985 |
| O | B | 3031 |
| E | C | 2597 |
| N | D | 159 |
| S | E | 53 |

Table 13: NOM_TOPO

| Valeur | Fréquence | NA |
|---|-----------|------|
| Parc linéaire de la Rivière Saint-Charles | A | 1940 |
| Ilot Robert-Bourassa | B | 1214 |
| Piste cyclable Corridor des Cheminots | C | 931 |
| Parc Victoria | D | 917 |
| Parc de la Plage-Jacques-Cartier | E | 905 |
| Parc de l'Exposition | F | 880 |
| Parc de l'Arboretum | G | 775 |
| Chemin Saint-Louis | H | 660 |

| Valeur | Fréquence | NA |
|-------------------------------|-----------|-----|
| Domaine de Maizerets | I | 539 |
| Parc de la Pointe-aux-Lièvres | J | 523 |

Description des variables quantitatives

| Valeur | Fréquence |
|--------|-----------|
| moyDIA | 24.84 |
| minDIA | 0 |
| maxDIA | 958 |
| minLON | -71.54 |
| minLAT | 46.72 |
| maxLON | -71.15 |
| maxLAT | 46.94 |

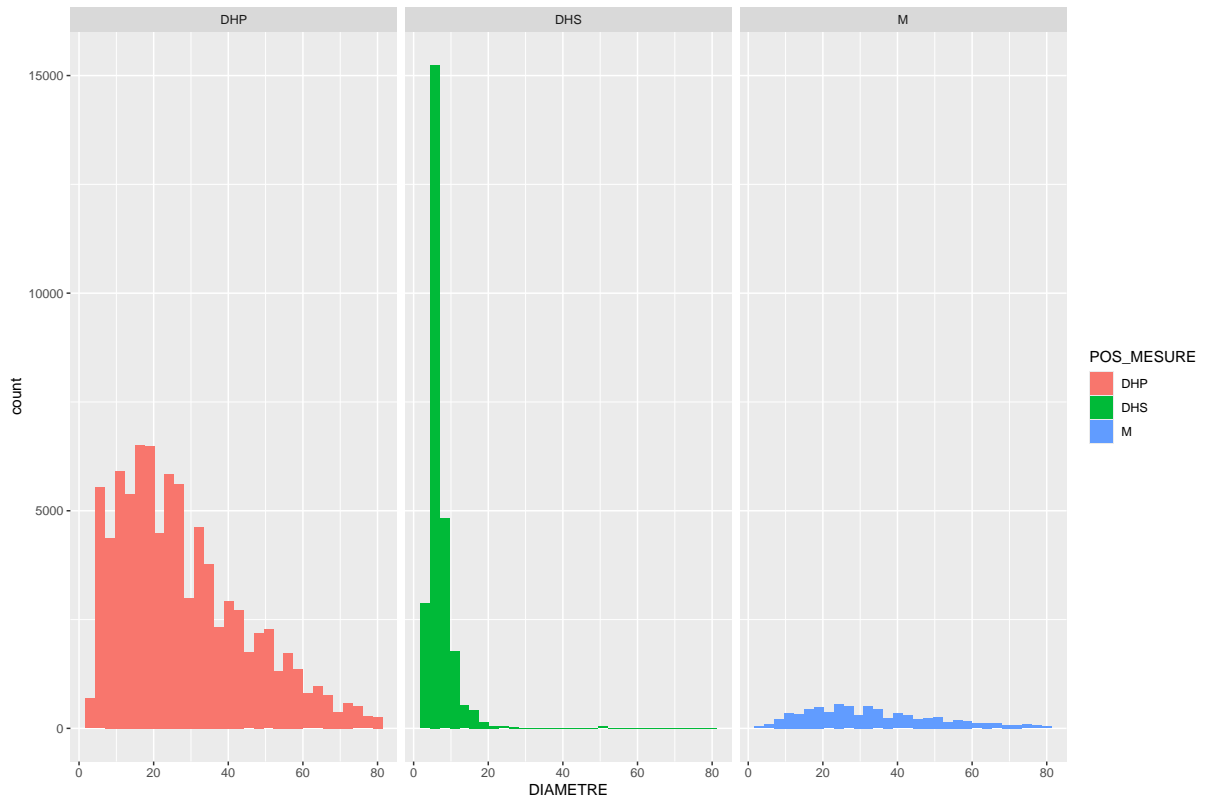
| minDATE | maxDATE |
|------------|------------|
| 1951-12-31 | 2016-11-13 |

On remarque entre autres qu'il y a des erreurs dans les données de diamètres. Certaines valeurs semblent inscrites en millimètres.

Distributions

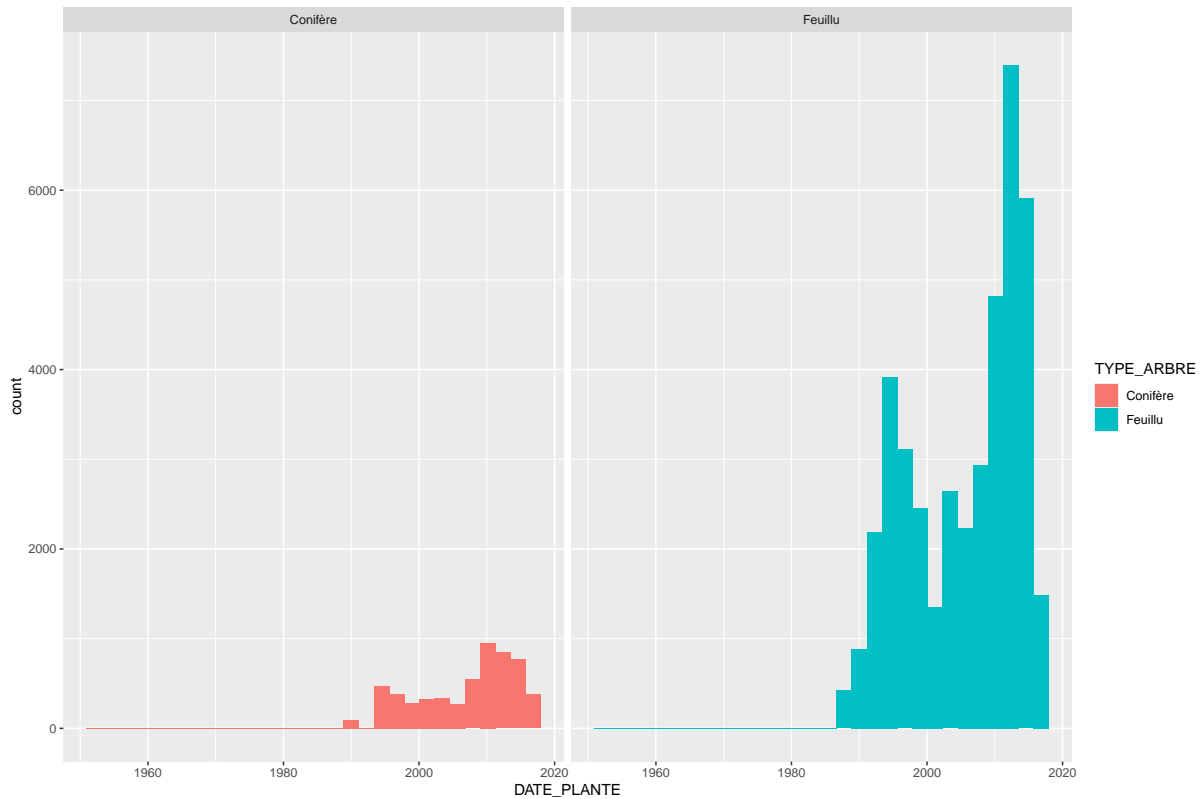
Diamètre des arbres

```
range_vals <- quantile(dA$DIAMETRE, probs = c(0.02, 0.98), na.rm = TRUE)
ggplot(
  data = dA %>% filter(range_vals[1] <= DIAMETRE & DIAMETRE < range_vals[2] & POS_MESURE != 'I')
  mapping = aes(x = DIAMETRE, fill = POS_MESURE)
) + geom_histogram() + facet_wrap("POS_MESURE")
```



Date de plantation

```
ggplot(
  data = dA %>% filter(TYPE_ARBRE != "NON DISPONIBLE"),
  mapping = aes(x = DATE_PLANTE, fill = TYPE_ARBRE)
) + geom_histogram() + facet_wrap("TYPE_ARBRE")
```



Enrichissement des variétés d'arbres

J'utilise les données provenant du [Système Mondial d'Informations sur la Biodiversité GFIB](#).

J'extrais d'abord les noms latins des espèces présentes dans la table avec `build_name` pour construire les URL de requêtes. Puis, je vais les requêtes en lot avec un `mapply` sur la fonction `get_url_gfib`.

```
get_url_gfib <- function(x)
  httr::GET(url = paste0("http://api.gbif.org/v1/species/match/?name=", x))

build_name <- function(x)
  gsub(pattern = " ",
        replacement = "+",
        x %>%
          strsplit("") %>%
          unlist %>%
          '['(1) %>%
```

```

    trimws())

## Noms uniques (incluant la variété locale)
nomsUniques <-
  dA %>%
  select(NOM_LAT) %>%
  distinct() %>%
  mutate(nom_url = sapply(NOM_LAT, build_name) %>% tolower())

## Noms uniques pour construire les URL (excluant la variété locale qui ne se trouve pas dans
nomsUrlUniques <-
  nomsUniques %>%
  select(nom_url) %>%
  distinct()

```

Je transforme ensuite les données recueillies dans le format JSON en une table que je pourrai joindre aux données source.

```

load("data_json_gfib.RData")
json_content <- sapply(data_json_gfib %>%
  as.data.frame %>%
  pull(content), rawToChar)
json_content2 <- data.frame(
  nom_url = names(json_content),
  json_content, row.names = NULL, stringsAsFactors = FALSE
)
json_content3 <- json_content2$json_content %>%
  lapply(fromJSON, flatten = TRUE) %>%
  lapply(as.data.frame) %>%
  bind_rows() %>%
  cbind(nom_url = json_content2$nom_url)
json_content4 <- merge(json_content3, nomsUniques, by = c("nom_url"))

```

Je peux maintenant joindre ces nouvelles informations aux données source

```

dA2 <- merge(dA, json_content4, by = c("NOM_LAT"))

```

Médias par espèces

```

get_url_media <- function(x)
  httr::GET(url = paste0("http://api.gbif.org/v1/species/", x, "/media"))

disct_speciesKey <- dA2 %>% select(speciesKey) %>% filter(!is.na(speciesKey)) %>% distinct()

load("json_media.RData")

json_medial <- data.frame(
  json_content = sapply(json_media %>%
                        as.data.frame %>%
                        pull(content), rawToChar),
  stringsAsFactors = FALSE
) %>%
  mutate(json_content1 = json_content %>%
         lapply(fromJSON, flatten = TRUE) %>%
         sapply('[', "results"))

json_medial$speciesKey <- disct_speciesKey$speciesKey

json_media2 <- json_medial[sapply(json_medial$json_content1, function(x) is.data.frame(x) &&
# Combine all media data.frames with their speciesKey
media_dfs <- list()
for (i in seq_len(nrow(json_media2))) {
  df <- json_media2$json_content1[[i]]
  if (is.data.frame(df) && nrow(df) > 0) {
    df$speciesKey <- json_media2$speciesKey[i]
    media_dfs[[length(media_dfs) + 1]] <- df
  }
}
json_media3 <- bind_rows(media_dfs)

# Keep only columns that actually exist
keep_cols <- intersect(
  c("value", "type", "format", "identifier", "references", "title",
    "description", "source", "creator", "publisher", "license", "speciesKey"),
  names(json_media3)
)
json_media3 <- json_media3 %>% select(all_of(keep_cols))

```

Joindre les données médias

```
dA3 <- merge(dA2, json_media3, all.x = TRUE)
```

Ajout du quartier et de l'arrondissement

```
qrtqc <- sf::read_sf("QUARTIERS/QUARTIER.shp") %>% sf::st_transform(4326)
arrqc <- sf::read_sf("ARROND/ARROND.shp") %>% sf::st_transform(4326)

# Rename columns to avoid conflicts
names(qrtqc)[names(qrtqc) != "geometry"] <- paste0(
  names(qrtqc)[names(qrtqc) != "geometry"], "_QRT"
)
names(arrqc)[names(arrqc) != "geometry"] <- paste0(
  names(arrqc)[names(arrqc) != "geometry"], "_ARR"
)

# Convert points to sf and perform spatial join
dA3_sf <- sf::st_as_sf(dA3, coords = c("LONGITUDE", "LATITUDE"), crs = 4326)
dA4.1 <- sf::st_join(dA3_sf, qrtqc)
dA4 <- sf::st_join(dA4.1, arrqc)

# Add coordinates back and remove geometry
dA4 <- dA4 %>%
  mutate(
    LONGITUDE = sf::st_coordinates(.)[, 1],
    LATITUDE = sf::st_coordinates(.)[, 2]
  ) %>%
  sf::st_drop_geometry()

save(dA4, file = "dA4.RData")
readr::write_csv(dA4, "arbres-augmented.csv")
```

Arbre le plus courant par quartier (ayant une photo disponible)

```
count_arbre_arr <- dA4 %>%
  filter(identifiant != "" & !is.na(NOM_QRT)) %>%
```

```

select(NOM_QRT, scientificName, identifiier) %>%
group_by(NOM_QRT, scientificName, identifiier) %>%
summarise(freq = n(), .groups = "drop") %>%
group_by(NOM_QRT) %>%
top_n(n = 1)

pandoc.table(count_arbre_arr %>% select(NOM_QRT, scientificName, freq))

```

| NOM_QRT | scientificName | freq |
|----------------------------|-------------------------------|------|
| Cap-Rouge | Ulmus americana L. | 978 |
| Chutes-Montmorency | Acer saccharinum L. | 429 |
| Chutes-Montmorency | Acer saccharinum L. | 429 |
| Cité Universitaire | Ulmus americana L. | 484 |
| Des Châtel | Acer platanoides L. | 1211 |
| Duberger-Les Saules | Acer platanoides L. | 3132 |
| Jésuites | Acer saccharinum L. | 399 |
| Jésuites | Acer saccharinum L. | 399 |
| L'Aéroport | Ulmus americana L. | 649 |
| Lac-Saint-Charles | Ulmus americana L. | 317 |
| Lairet | Acer platanoides L. | 1075 |
| Loretteville | Acer platanoides L. | 370 |
| Maizerets | Fraxinus pennsylvanica Marsh. | 1465 |
| Montcalm | Ulmus americana L. | 610 |
| Neufchâtel-Est/Lebourgneuf | Acer platanoides L. | 2431 |
| Notre-Dame-des-Laurentides | Ulmus americana L. | 364 |
| Plateau | Acer platanoides L. | 319 |
| Pointe-de-Sainte-Foy | Acer platanoides L. | 443 |
| Quartier 4-2 | Ulmus americana L. | 317 |
| Quartier 4-3 | Acer platanoides L. | 169 |
| Quartier 4-5 | Acer platanoides L. | 467 |
| Quartier 4-6 | Acer platanoides L. | 257 |
| Quartier 5-1 | Acer platanoides L. | 141 |
| Quartier 5-2 | Acer platanoides L. | 306 |
| Quartier 5-4 | Ulmus americana L. | 509 |
| Saint-Jean-Baptiste | Ulmus americana L. | 292 |
| Saint-Louis | Acer platanoides L. | 513 |
| Saint-Roch | Acer platanoides L. | 436 |
| Saint-Sacrement | Acer platanoides L. | 575 |
| Saint-Sauveur | Ulmus americana L. | 693 |
| Saint-Émile | Acer platanoides L. | 326 |
| Sillery | Ulmus americana L. | 1422 |

| NOM_QRT | scientificName | freq |
|---|-------------------------------|------|
| Val-Bélair | Acer platanoides L. | 460 |
| Vanier | Fraxinus pennsylvanica Marsh. | 418 |
| Vieux-Limoilou | Acer platanoides L. | 634 |
| Vieux-Moulin | Acer platanoides L. | 202 |
| Vieux-Québec/Cap-Blanc/Colline parlementaire | Ulmus americana L. | 718 |

```
count_arbre_arr %>%
  mutate(image = paste0("[", scientificName, "](", identifier, ")")) %>%
  select(NOM_QRT, image) %>%
  t %>%
  pandoc.table()
```

Table 17: Table continues below

| | |
|---------|------------------------------------|
| NOM_QRT | Cap-Rouge |
| image | Ulmus americana L. |

Table 18: Table continues below

| | |
|---------|-------------------------------------|
| NOM_QRT | Chutes-Montmorency |
| image | Acer saccharinum L. |

Table 19: Table continues below

| | |
|---------|-------------------------------------|
| NOM_QRT | Chutes-Montmorency |
| image | Acer saccharinum L. |

Table 20: Table continues below

| | |
|---------|------------------------------------|
| NOM_QRT | Cité Universitaire |
| image | Ulmus americana L. |

Table 21: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Des Châtels Acer platanoides L. |
|------------------|--|

Table 22: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Duburger-Les Saules Acer platanoides L. |
|------------------|--|

Table 23: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Jésuites Acer saccharinum L. |
|------------------|---|

Table 24: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Jésuites Acer saccharinum L. |
|------------------|---|

Table 25: Table continues below

| | |
|------------------|--|
| NOM_QRT image | L'Aéroport Ulmus americana L. |
|------------------|--|

Table 26: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Lac-Saint-Charles Ulmus americana L. |
|------------------|---|

Table 27: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Lairt Acer platanoides L. |
|------------------|--|

Table 28: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Loretteville Acer platanoides L. |
|------------------|---|

Table 29: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Maizerets Fraxinus pennsylvanica Marsh. |
|------------------|--|

Table 30: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Montcalm Ulmus americana L. |
|------------------|--|

Table 31: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Neufchâtel-Est/Lebourgneuf Acer platanoides L. |
|------------------|---|

Table 32: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Notre-Dame-des-Laurentides Ulmus americana L. |
|------------------|--|

Table 33: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Plateau Acer platanoides L. |
|------------------|--|

Table 34: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Pointe-de-Sainte-Foy Acer platanoides L. |
|------------------|---|

Table 35: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Quartier 4-2 Ulmus americana L. |
|------------------|--|

Table 36: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Quartier 4-3 Acer platanoides L. |
|------------------|---|

Table 37: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Quartier 4-5 Acer platanoides L. |
|------------------|---|

Table 38: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Quartier 4-6 Acer platanoides L. |
|------------------|---|

Table 39: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Quartier 5-1 Acer platanoides L. |
|------------------|---|

Table 40: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Quartier 5-2 Acer platanoides L. |
|------------------|---|

Table 41: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Quartier 5-4 Ulmus americana L. |
|------------------|--|

Table 42: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Saint-Jean-Baptiste Ulmus americana L. |
|------------------|---|

Table 43: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Saint-Louis Acer platanoides L. |
|------------------|--|

Table 44: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Saint-Roch Acer platanoides L. |
|------------------|---|

Table 45: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Saint-Sacrement Acer platanoides L. |
|------------------|--|

Table 46: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Saint-Sauveur Ulmus americana L. |
|------------------|---|

Table 47: Table continues below

| | |
|------------------|--|
| NOM_QRT image | Saint-Émile Acer platanoides L. |
|------------------|--|

Table 48: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Sillery Ulmus americana L. |
|------------------|---|

Table 49: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Val-Bélair Acer platanoides L. |
|------------------|---|

Table 50: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Vanier Fraxinus pennsylvanica Marsh. |
|------------------|---|

Table 51: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Vieux-Limoilou Acer platanoides L. |
|------------------|---|

Table 52: Table continues below

| | |
|------------------|---|
| NOM_QRT image | Vieux-Moulin Acer platanoides L. |
|------------------|---|

| | |
|------------------|--|
| NOM_QRT image | Vieux-Québec/Cap-Blanc/Colline parlementaire Ulmus americana L. |
|------------------|--|

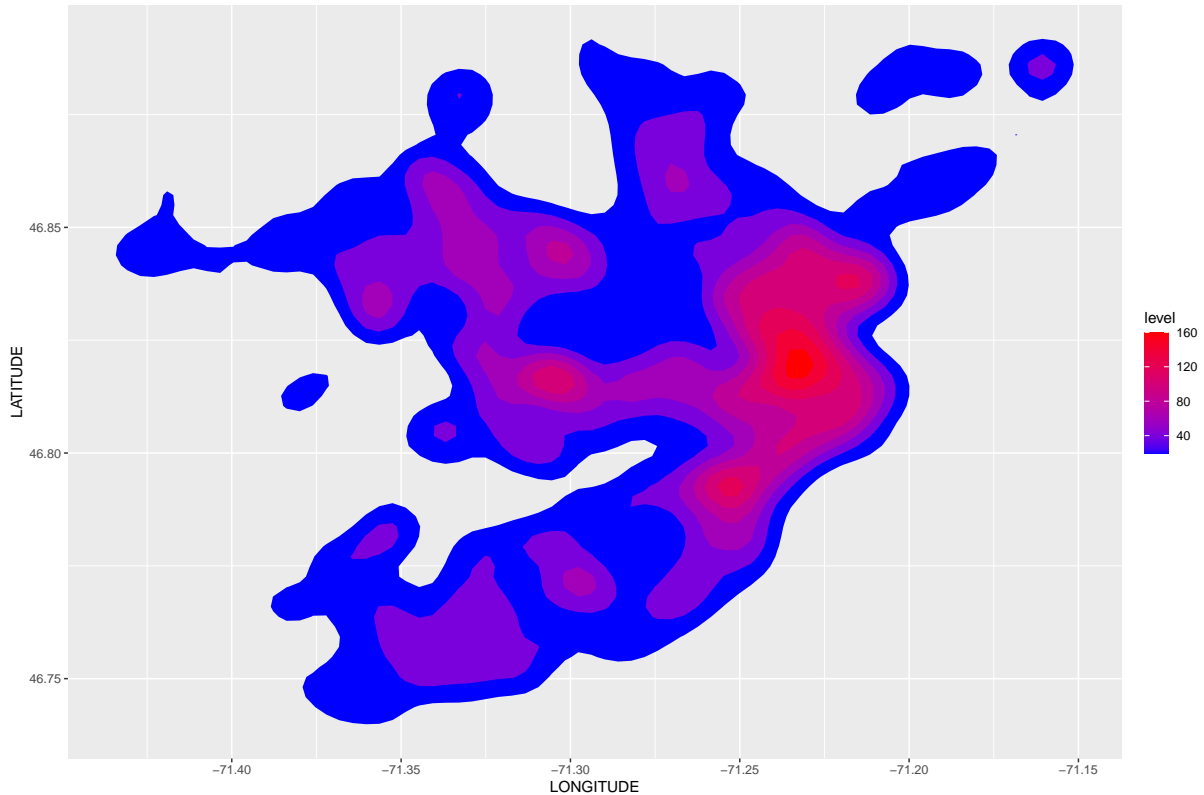
Localisation sur une carte

```
select_for_map <- dA4 %>%
  select(LONGITUDE, LATITUDE, order)
range_long <- range(select_for_map$LONGITUDE)
range_lat <- range(select_for_map$LATITUDE)

# Without ggmap (not available), use a clean ggplot2 background
QuebecMap <- ggplot(data = select_for_map, aes(x = LONGITUDE, y = LATITUDE))
map1 <- QuebecMap + scale_fill_gradient(low = "blue", high = "red")
```

Carte Simple

```
map1 +
  stat_density2d(aes(x = LONGITUDE, y = LATITUDE, fill = after_stat(level)),
    geom = "polygon", data = select_for_map)
```

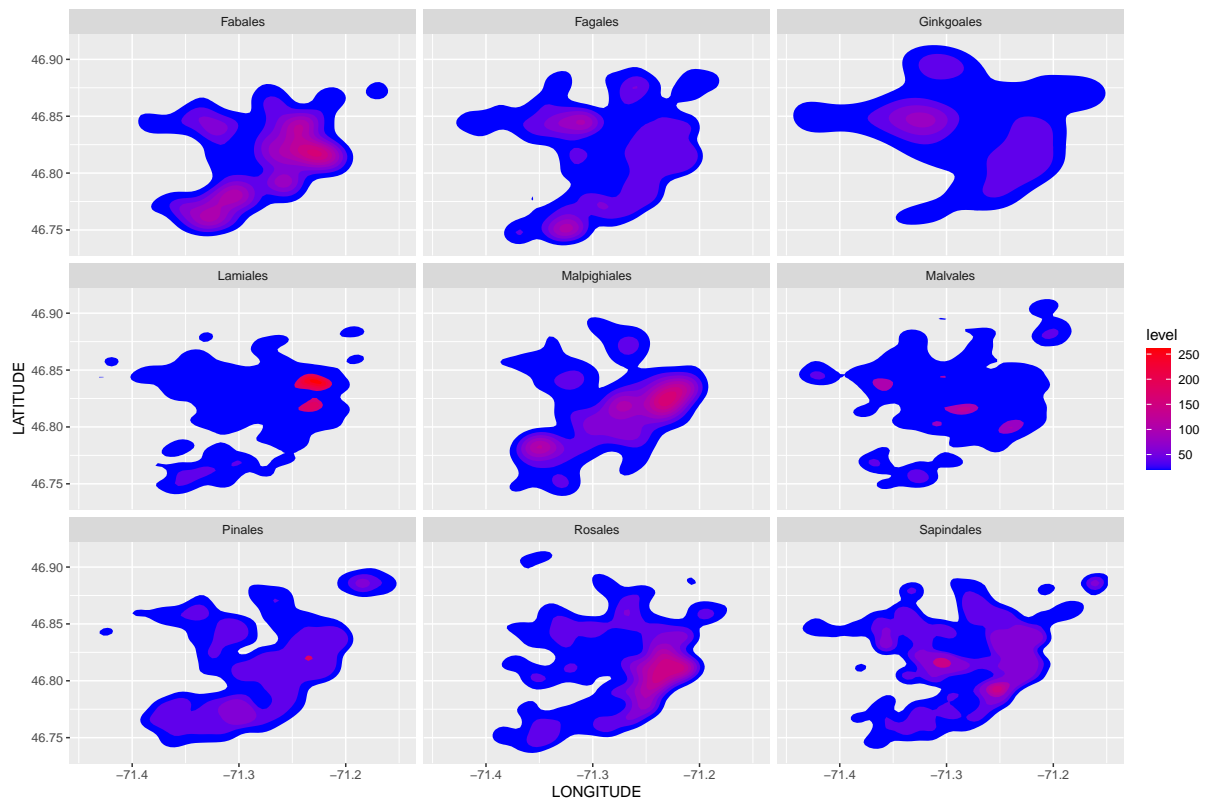


Carte Composée

```
table_order <- with(select_for_map, table(order))
select_for_map2 <- table_order %>%
  as.data.frame() %>%
  merge(select_for_map, all.y = TRUE) %>%
  filter(Freq > 500)
# Garder seulement les niveaux actifs
select_for_map2$order2 <- factor(select_for_map2$order)

map1 +
  stat_density2d(aes(x = LONGITUDE, y = LATITUDE, fill = after_stat(level)),
```

```
geom = "polygon", data = select_for_map2) +
facet_wrap(facets = "order2")
```



Arbres recensés par quartier

Source: [plotting polygon shapefiles](#)

```
ggmap_quartiers <- ggplot(qrtqc) +
  aes(fill = NOM_QRT) +
  geom_sf(color = "white")

ggmap_arrond <- ggplot(arrqc) +
  aes(fill = NOM_ARR) +
  geom_sf(color = "white")

plot_data_quartiers <-
  dA4 %>%
  group_by(NOM_QRT, order) %>%
```

```

summarise(freq = n(), .groups = "drop") %>%
filter(!is.na(order) & freq >= 500)

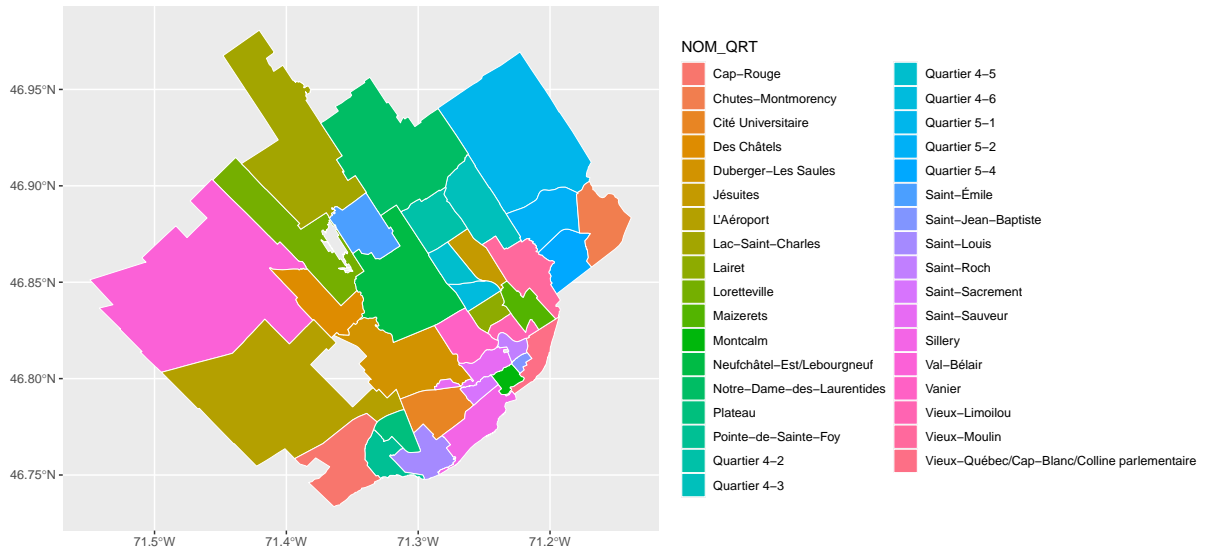
plot_data_arrond <-
dA4 %>%
group_by(NOM_ARR, order) %>%
summarise(freq = n(), .groups = "drop") %>%
filter(!is.na(order) & freq >= 500)

gg_freq_ordre_quartier <- ggplot(data = plot_data_quartiers,
                                aes(x = order, y = freq, fill = order)) +
geom_bar(position = "stack", stat = "identity") +
facet_wrap(facets = "NOM_QRT", ncol = 4) +
scale_x_discrete(breaks = NULL) +
xlab("Ordre") +
ylab("Fréquence") +
ggtitle("Ordre par quartier")

gg_freq_ordre_arrond <- ggplot(data = plot_data_arrond,
                                aes(x = order, y = freq, fill = order)) +
geom_bar(position = "stack", stat = "identity") +
facet_wrap(facets = "NOM_ARR") +
scale_x_discrete(breaks = NULL) +
xlab("Ordre") +
ylab("Fréquence") +
ggtitle("Ordre par arrondissement")

ggmap_quartiers

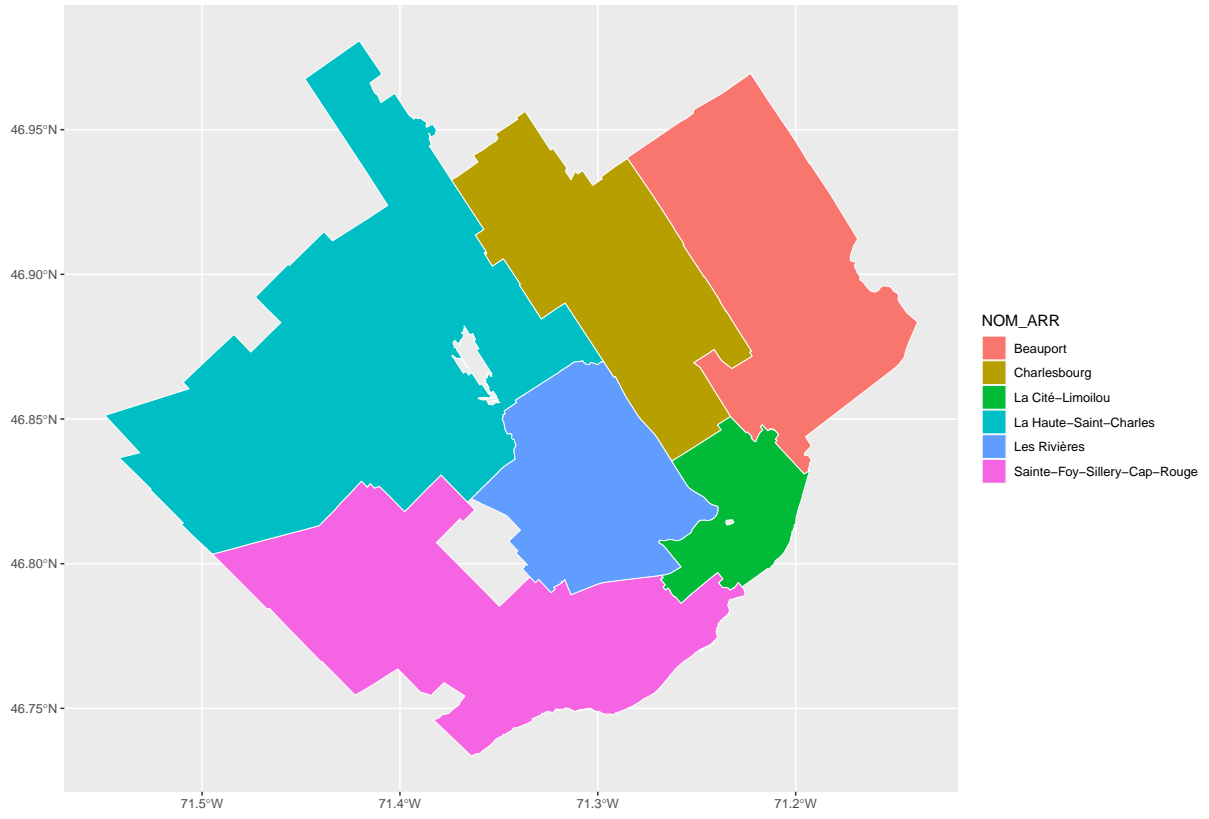
```



gg_freq_ordre_quartier



ggmap_arrond



gg_freq_ordre_arrond

