Capacity evolution of wind turbines in Canada

This is the dataset we will be working with:

```
repo <- "https://raw.githubusercontent.com/rfordatascience/tidytuesday/master"
file <- "/data/2020/2020-10-27/wind-turbine.csv"

wind_turbine <-
  readr::read_csv(
    paste(repo,file, sep = "")
  )
```

More information about the dataset can be found here: https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-10-27 and https://open.canada.ca/data/en/dataset/79fdad93-9025-49ad-ba16-c26d718cc070.

Question: Which are the two Canada's provinces with the majority of wind turbines installed, and how has the capacity of wind turbines installed in those provinces changed over the last ten years ?

Introduction: We are working with the `wind_turbine` dataset, which contains 6698 records of geographic location and key technology details for wind turbines installed in Canada. Each row corresponds to the details of each wind turbine, and there are 15 columns that provide information about the precise latitude and longitude of every turbine, along with details like its dimensions, its power output, its manufacturer and the date it was commissioned.

To answer the question, we will extract only five variables from the dataset. The variables are the date the turbine was commissioned (column `commissioning_date`), the turbine capacity in kilowatts (column `turbine_rated_capacity_k_w`), the latitude of where the turbine is located (column `latitude`), the longitude of where the turbine is located (column `longitude`), and province territory where the turbine is installed (column `province_territory`). The commissioning_date is provided as a character value reporting only the year the turbine was commissioned. The province_territory is provided as a string. The turbine_rated_capacity_k_w, latitude and longitude are provided as a numerical value.

Approach: Our approach to answer the question is to first do data wrangling to extract the five columns necessary for the analysis from the dataset. This can be done using `select()`. Then, for the variable `commissioning_date` we have to clean values like "2013/2014". For simplicity, we use the first year reported as to when the majority of turbines corresponding to a certain project started to function. We don't count the turbines for 2013 and 2014 because it'll augment the original total number of turbines. Then, we convert the values from character to numeric with `as.numeric()`. We preferred numeric over date type because only the year has been given. Finally, we plot a histogram for all our variables to check the distribution for each one and we compute a summary statistic with `summary()`.

To answer the question, we can use a choropleth map using the function `geom_sf()` which will plot a map of Canada with Provincial/Territorial boundaries. We have to first obtain geospatial data from this source https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2011-eng.cfm to draw the map. We load the data from the file that was already processed. Then, we can use the latitude and longitude values of each turbine to localize where the turbine was installed. Finally, to answer the question about the evolution of capacity we can use a ridgeline plot with `geom_density_ridges()` and we will treat the year as a factor to show the distributions of capacity along the past ten years.(2012-2022)

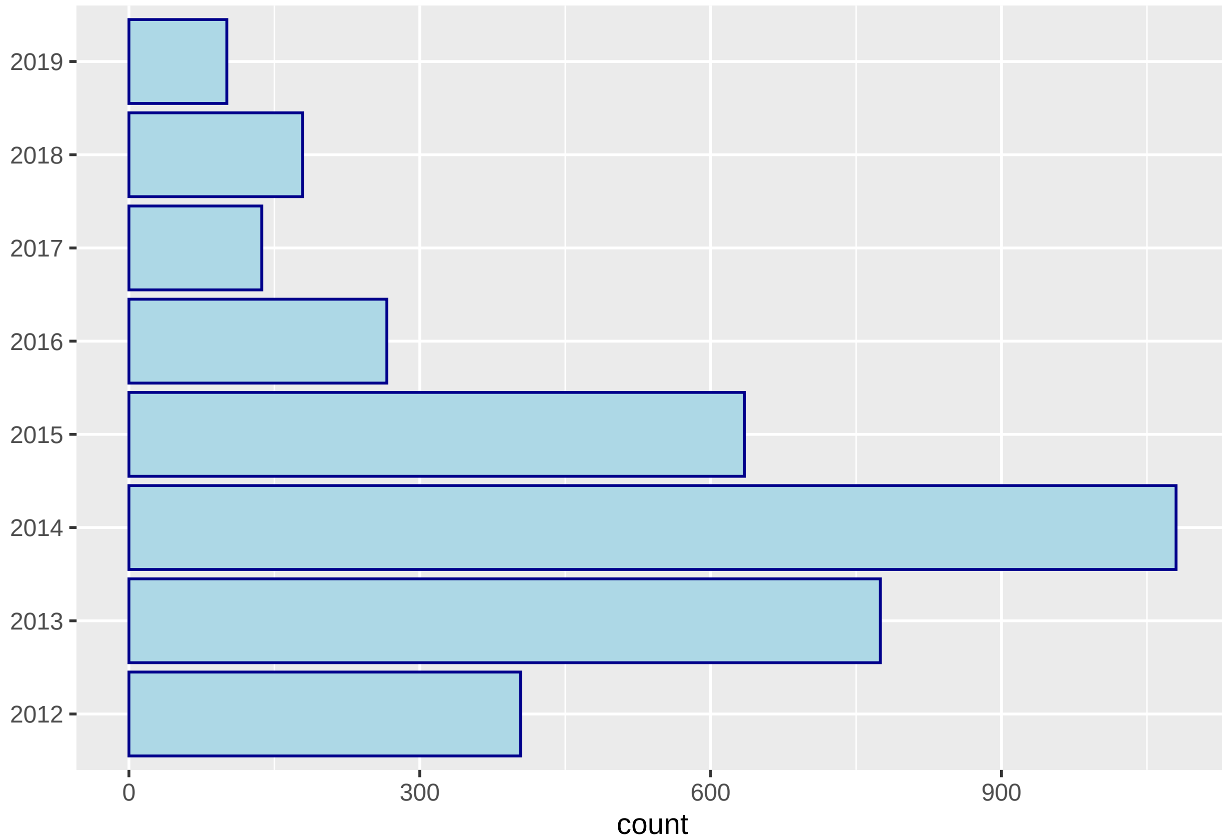
Analysis:

Filtering only the columns that we use for the analysis.

```
# Filtering columns useful to answer the question
# (turbine_rated_capacity_k_w, commissioning_date, latitude, ideal_diet_coded)

wind_turbine_filtered <- wind_turbine %>%
select(turbine_rated_capacity_k_w, commissioning_date, latitude, longitude, province_territory)
```

After filtering, we have to clean the data.

```
#Cleaning data.

# ************ commissioning_date ************

# Replacing values like "2014/2015" with "2014" only we'll take the first value.
wind_turbine_filtered$commissioning_date <-
  sub("\\/.*", "", wind_turbine_filtered$commissioning_date)

# After replacing invalid strings, we can convert the year to numeric
wind_turbine_filtered$commissioning_date <-
  as.numeric(wind_turbine_filtered$commissioning_date)

# Histogram
ggplot(wind_turbine_filtered %>%
         filter(commissioning_date > 2011), aes(y =factor(commissioning_date))) +
  geom_bar(color="darkblue", fill="lightblue")+
  ylab(NULL)
```

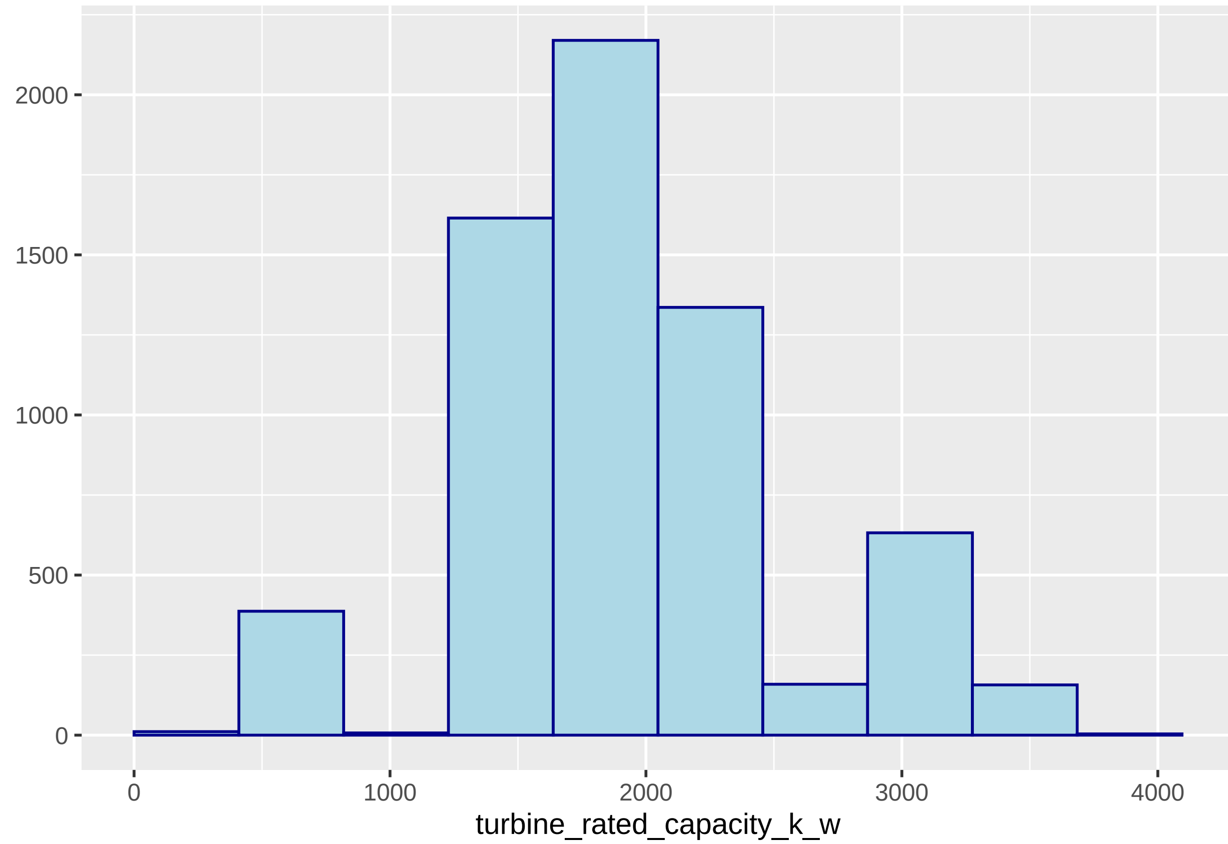



```
# Summary statistics.
summary(wind_turbine_filtered$commissioning_date)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1993    2009    2012    2011    2014    2019
```

```
#Cleaning data.

# ************ turbine_rated_capacity_k_w ************

# Histogram
ggplot(wind_turbine_filtered , aes(x =turbine_rated_capacity_k_w)) +
  geom_histogram(bins = 10,  boundary = 0,  color="darkblue", fill="lightblue") +
  ylab(NULL)
```

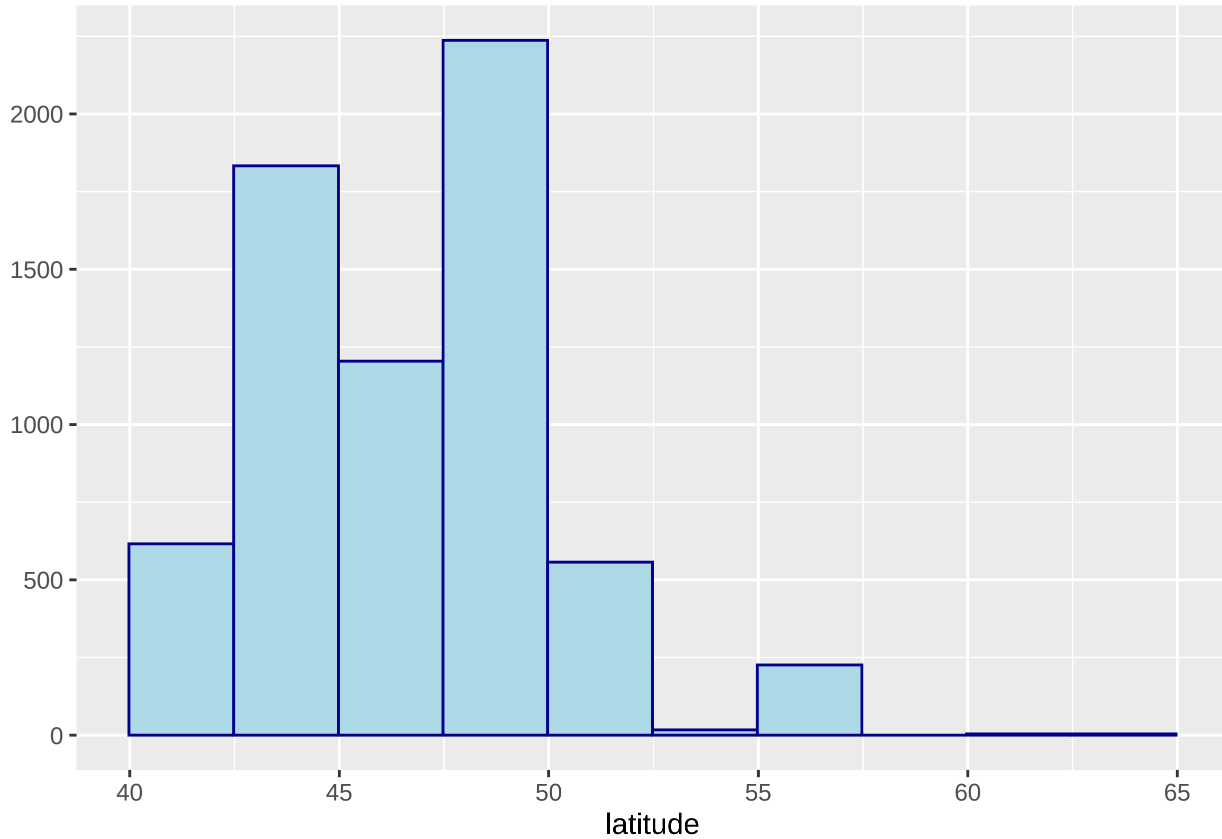
```
## Warning: Removed 220 rows containing non-finite values (`stat_bin()`).
```




```
# Summary statistics.
summary(wind_turbine_filtered$turbine_rated_capacity_k_w)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##      65    1600    1880    1967    2300    3750     220
```

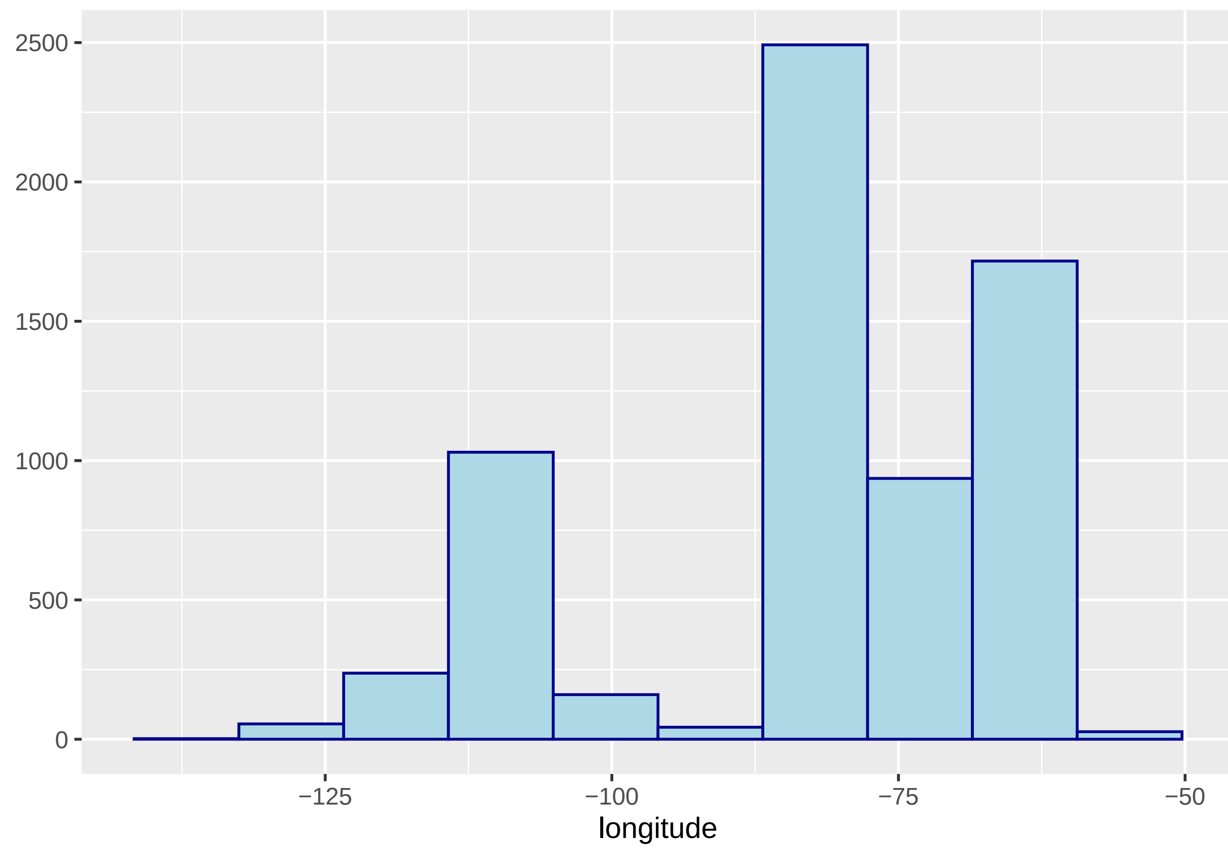
```
# ************ latitude and longitude ************

# Histogram
ggplot(wind_turbine_filtered , aes(x =latitude)) +
  geom_histogram(bins = 10,  boundary = 0,  color="darkblue", fill="lightblue") +
  ylab(NULL)
```




```
# Summary statistics.
summary(wind_turbine_filtered$latitude)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   42.00   43.98   46.67   46.76   49.17   64.49
```

```
# Histogram
ggplot(wind_turbine_filtered , aes(x =longitude)) +
  geom_histogram(bins = 10, color="darkblue", fill="lightblue") +
  ylab(NULL)
```




```r
# Summary statistics.
summary(wind_turbine_filtered$longitude)
```

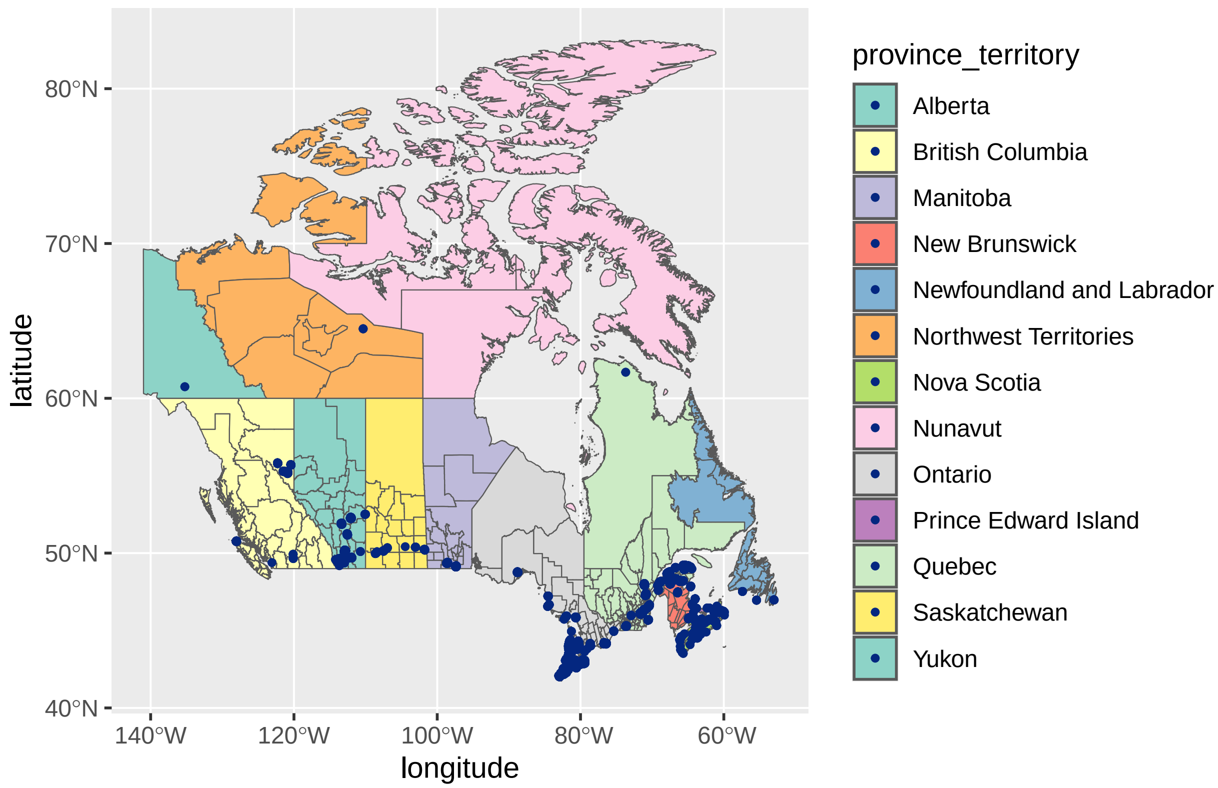
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -135.23  -84.41  -80.67  -83.03  -67.85  -52.97
```

Plot the choropleth map of Canada filling territory boundaries and localize the turbins that were installed in each province.

```r
canada_cd <- st_read("data/canada_cd_sim.geojson", quiet = TRUE)

canada_cd$PRNAME <- sub("\\/.*", "", canada_cd$PRNAME)
canada_cd$PRNAME<- trimws(canada_cd$PRNAME, which = c("right"))

canada_cd <- canada_cd %>%
  rename(
    province_territory = PRNAME
    )

## Draw the map
map_colors <-  RColorBrewer::brewer.pal(12, "Set3")
map_colors <- rep(map_colors, 60)
p <- ggplot(data = canada_cd, mapping = aes(fill = province_territory)) +
    geom_sf(size = 0.1, lwd = 0) +
    scale_fill_manual(values = map_colors)
```

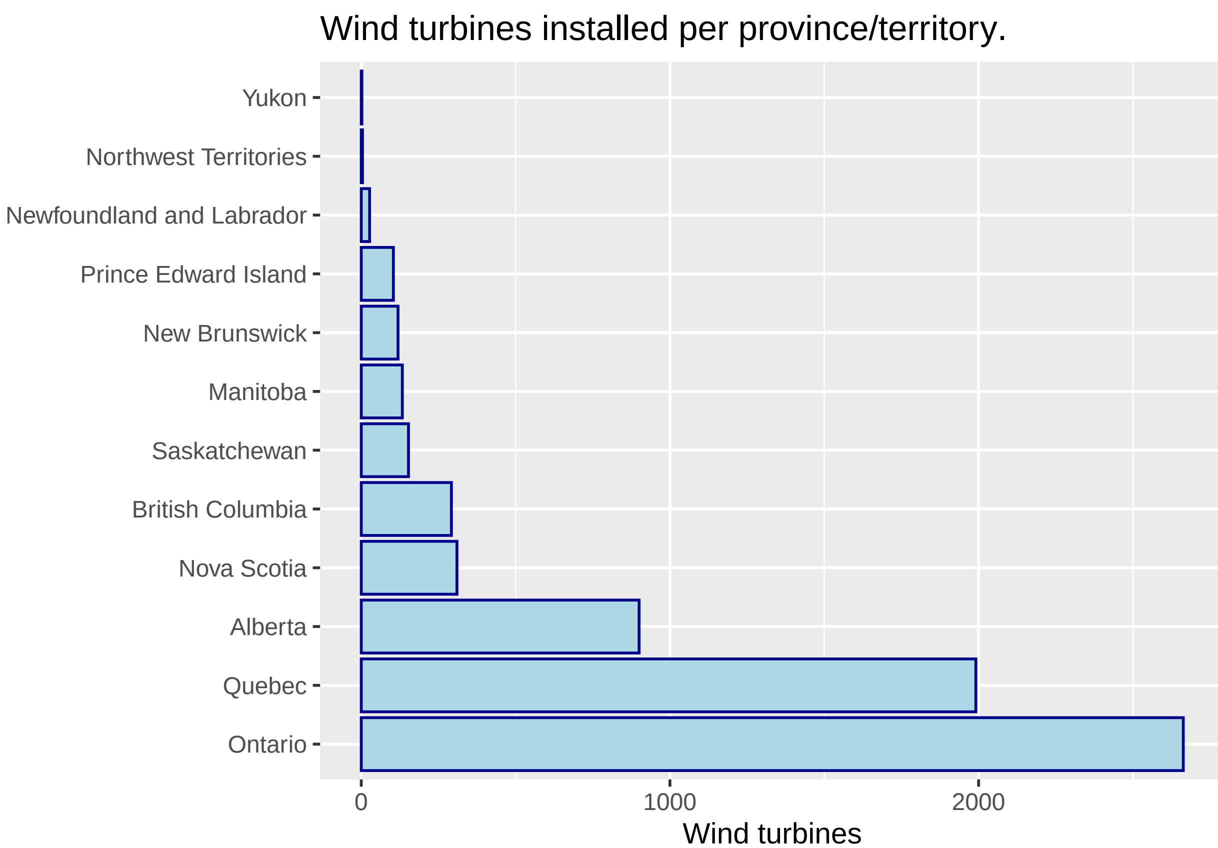
```
## Warning: Duplicated aesthetics after name standardisation: size
```

```r
p_out <- p+ geom_point(data=wind_turbine_filtered,
             aes(longitude, latitude),
             color = "#042780",
             size = 0.9,
             alpha=1)
p_out
```



The bar chart below shows the number of wind turbines installed per province/territory to find the two provinces/territories with the majority number of turbines installed to inspect their evolution over the last ten years.

```r
# Bar chart of wind turbines installed in each provicence/territory
ggplot(wind_turbine_filtered, aes(y = fct_infreq(factor(province_territory)))) +
  geom_bar(color="darkblue", fill="lightblue")+
  ggtitle("Wind turbines installed per province/territory.")+
  ylab(NULL)+
  xlab("Wind turbines")
```

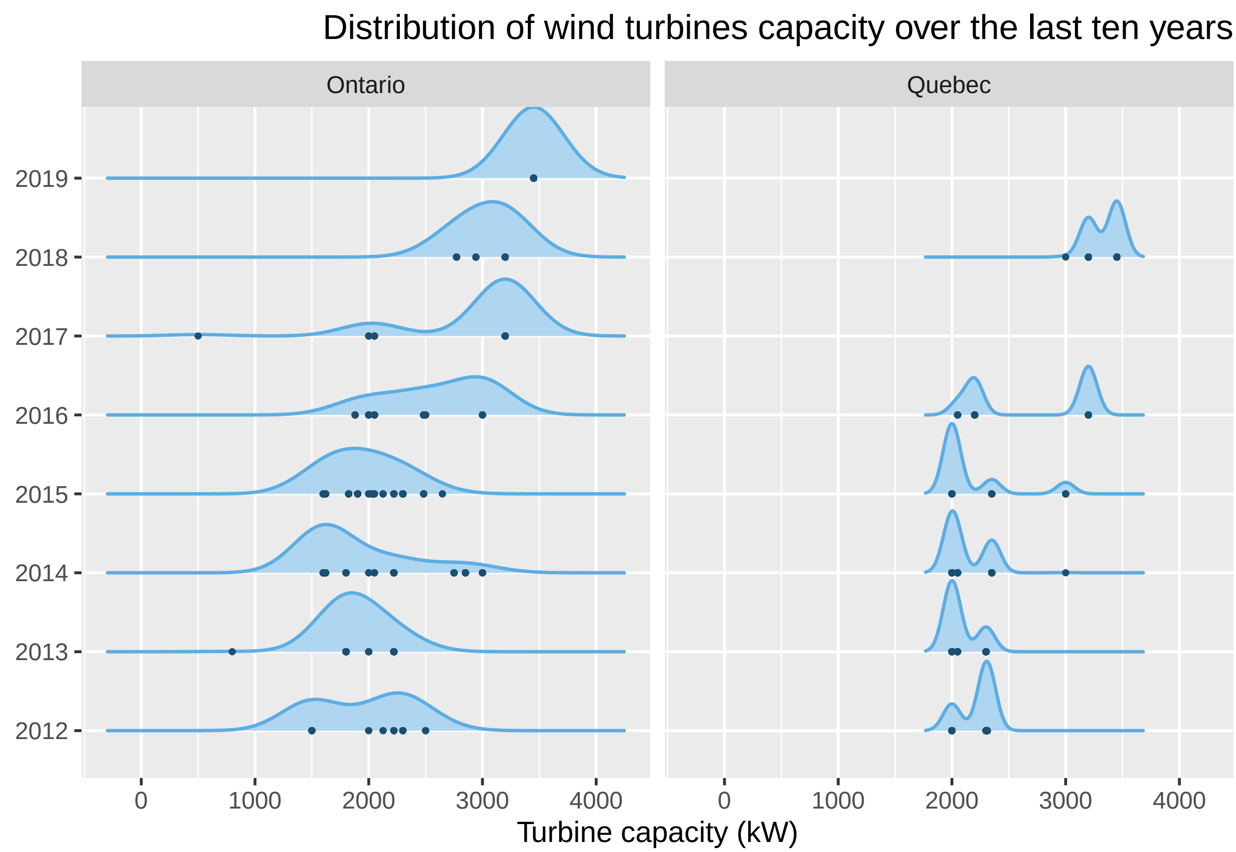



```
# Filtering data for Ontario and Quebec turbines from the last ten years.
ontario_quebec_filtered <- wind_turbine_filtered %>%
  filter(commissioning_date > 2011 &
           (province_territory == "Ontario" |
              province_territory == "Quebec"))

density_ontario_quebec <- ggplot(data=ontario_quebec_filtered,
                                 aes(x=turbine_rated_capacity_k_w,
                                     y=factor(commissioning_date)))

density_ontario_quebec <- density_ontario_quebec +
                    geom_density_ridges(fill = "#AED6F1",
                      color = "#5DADE2",
                      lwd = 0.6,
                      scale = 0.9,
                      na.rm = TRUE
                    )+
                  geom_point(color = "#1B4F72", size = 0.6, na.rm = TRUE)+
                  ylab(NULL)+
                  ggtitle("Distribution of wind turbines capacity over the last ten years")+
                  xlim(min(ontario_quebec_filtered$turbine_rated_capacity_k_w+1),
                       max(ontario_quebec_filtered$turbine_rated_capacity_k_w+1)) +
                  theme(plot.title = element_text(hjust = 1))+
                              facet_wrap(~province_territory)+
                  scale_x_continuous(name ="Turbine capacity (kW)")
```

```
## Scale for x is already present.
## Adding another scale for x, which will replace the existing scale.
```

```
density_ontario_quebec
```

```
## Picking joint bandwidth of 266
## Picking joint bandwidth of 77.5
```



Discussion: The two provinces with more turbines installed are Ontario and Quebec. This can be seen in the choropleth map of Canada and reaffirmed in the bar chart that shows the number of wind turbines installed per province/territory. Most of the turbines (blue points) are concentrated in the borders that join Ontario and Quebec provinces.

The capacity of the turbines has been stable between 1500 and 3000 kilowatts in Ontario and Quebec, but from 2017 to 2019 turbines with more capacity has been installed. In 2018 there were some turbines installed in Quebec with a capacity greater than 3000 kilowatts. However, there have been more projects to install turbines in Ontario than Quebec. For example, in 2019 and 2017 no turbines were installed in Quebec, but from 2012 to 2016 in Ontario more turbines were installed and they didn't exceed the capacity of 3000 kilowatts.