Project 5

Read in the dataset you will be working with:

```
wind_turbine <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-10-27/wind-turbine.csv')

head(wind_turbine)
```

```
## # A tibble: 6 x 15
##   objectid province_territory project_name     total_project_c~ turbine_identif~
##      <dbl> <chr>              <chr>                       <dbl> <chr>
## 1        1 Alberta            Optimist Wind E~             0.9  OWE1
## 2        2 Alberta            Castle River Wi~            44    CRW1
## 3        3 Alberta            Waterton Wind T~             3.78 WWT1
## 4        4 Alberta            Waterton Wind T~             3.78 WWT2
## 5        5 Alberta            Waterton Wind T~             3.78 WWT3
## 6        6 Alberta            Waterton Wind T~             3.78 WWT4
## # ... with 10 more variables: turbine_number_in_project <chr>,
## #   turbine_rated_capacity_k_w <dbl>, rotor_diameter_m <dbl>,
## #   hub_height_m <dbl>, manufacturer <chr>, model <chr>,
## #   commissioning_date <chr>, latitude <dbl>, longitude <dbl>, notes <chr>
```

We are working with `The Canadian Wind Turbine Database` which records the information about wind turbines installed. It belongs to Government of Canada and is collected by Will Noel, Tim Weis and Andrew Leach. It consists of the geographic location (ie. `Latitude` , `Longitude` , `province_territory`) and core technology information (e.g. `rotor_diameter` , `hub_height` , `turbine_rated_capacity_k_w` , `manufacturer` , `model` etc) for wind turbines installed in Canada. For more information regarding the dataset, please refer https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-10-27/readme.md (https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-10-27/readme.md)

Question: *What is the count and distribution of turbines in each province ? Where are these turbines exactly located in geo-spatial space ? How has the average capacity of the turbines changed by provinces and commissioning year ? What do you observe ?*

Introduction:

We are working with Canada Wind Turbine Database. It consists of total 6699 rows, and 15 rows. Each row records the exact latitudinal and longitudinal location of the wind-turbine along with details such as `province_territory` , `turbine_rated_capacity_k_w` , `rotor_diameter` , `hub_height` , `project_name` , `manufacturer` , `model` , `total_project_capacity_mw` etc. For this project, we will count the number of turbines in each province, and also visualize the exact location of turbines in geo-spatial space and also explore how the capacity of turbines changes by province and the comissioning year.

To carry out this objective, we will make use of 5 columns from the wind turbine dataset. 1. `province_territory` : It's a categorical variable that lists the provinces of Canada where the turbine is installed 2. `Latitude` : It lists the exact latitudinal location of the turbine in Canada 3. `Longitude` : It lists the longitudinal location of the turbine in Canada 4. `comissioning_date` : Year during when the equipment is installed to verify correct operation 5. `turbine_rated_capacity_k_w` : It lists the turbine capacity in kilowatts.

Moreover, we will also import an additional dataset for loading the map boundaries of provinces in canada to view the no. of turbines in the geo-spatial space. We have also provided this dataset along with this file. We will use the path "canada-maps-r-main/data/canada_cd_sim.geojson" from to load the map boundaries. For

more information regarding this dataset, please visit https://github.com/fifty8/canada-maps-r/tree/main/data (https://github.com/fifty8/canada-maps-r/tree/main/data)

Approach:

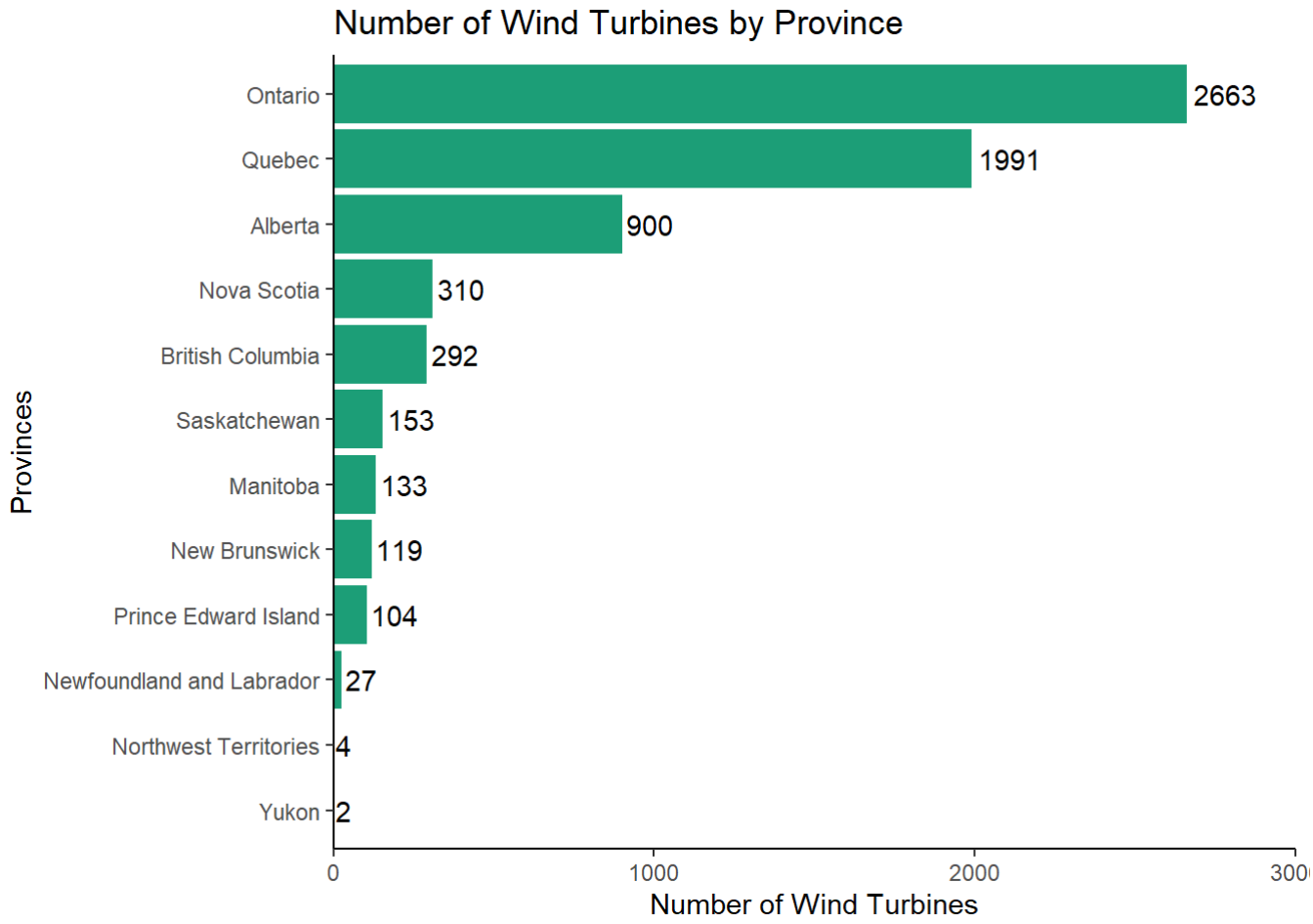
Count of Turbines can be visualized with the help of a bar graph which will show the total count of turbines on x-axis and province-name on Y-Axis. We have used bar graph since it presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent

For displaying the map-boundaries of canada, we have imported an additional dataset of canada-maps. Then, we will call st_read() to read the geojson file from the imported dataset. We have provided the local path, and the data is also provided locally. Then, we will mutate the dataset wind_turbine in such a way that it adds additional column of geometry from `latitude` *and* `Longitudal` *fields. Reasoning behind adding geometry column is to make it work well while using geom_sf package. Then, we will visualize the geo-spatial data with the help of geom_sf.*

For understanding how the turbine capacity has evolved by provinces and commissioing date, we will mutate the wind_turbine dataset. For mutation, we will group it by commissioning date and provinces. Then, we will average the turbine_rated_capacity_k_w and remove the null values if any. After that, we will mutate the comissioning date to character variable and then mutate the character variable to date variable and just format it by year. Then, we will show a scatter-plot of average turbine capacity vs comissioning year faceted by provinces. We will use scatter-plot since its one of the best ways to explore the relationship and to visualize the progress of turbine capacity by year and we have used facet_wrap to understand how it changes by provinces

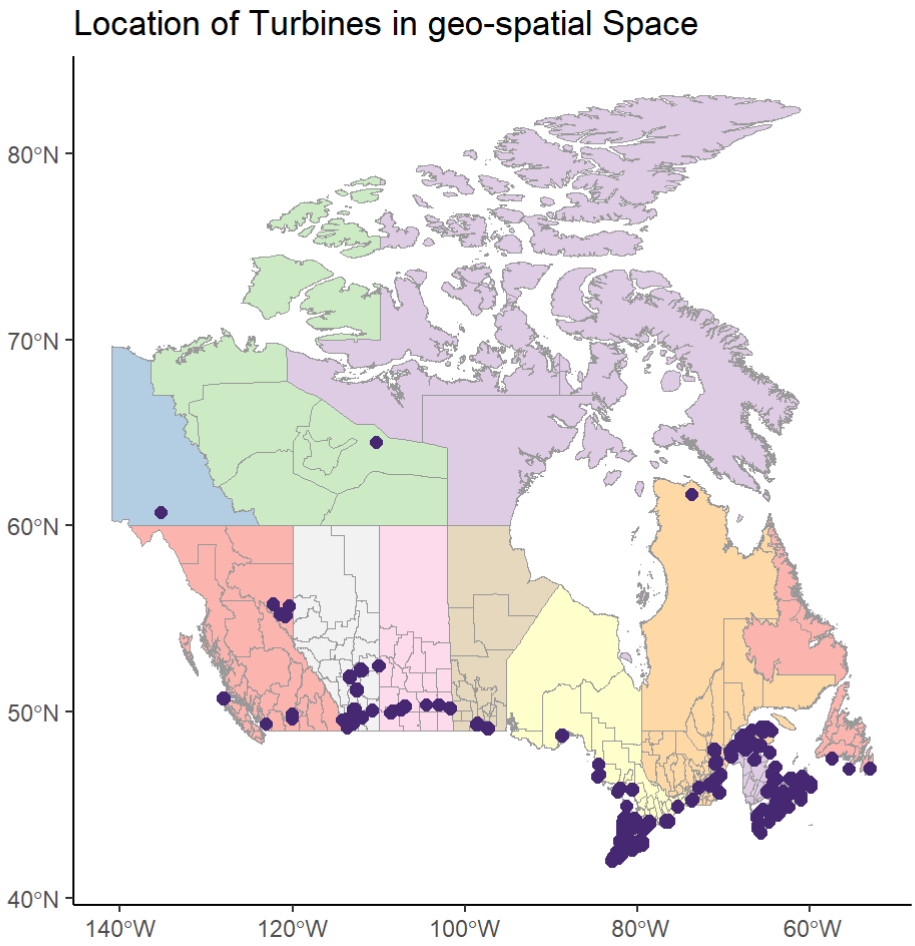
Analysis:

```
# Number of wind Turbines by Province

wind_turbine %>%
  count(province_territory, sort = TRUE, name="Count") %>%
  ggplot(aes(Count, fct_reorder(province_territory, Count))) +
  geom_col(fill="#1c9e77") +
  geom_text(aes(label=Count),hjust = -.1)+
  labs(title="Number of Wind Turbines by Province",x="Number of Wind Turbines",y="Provinces")
+
  theme_classic()+
  scale_x_continuous(expand=c(0,0),limits = c(0,3000))
```




```
# Location of Turbines in geo-spatial space

#Load the data provided along with this file.
canada_cd <- st_read("canada-maps-r-main/data/canada_cd_sim.geojson",quiet = TRUE)

wind_turbine_map <- wind_turbine |>
  st_as_sf(
    coords = c("longitude", "latitude"),
    crs = 4326,
    agr = "constant"
    )

map_colors <- RColorBrewer::brewer.pal(9, "Pastel1") %>% rep(37)

ggplot(wind_turbine_map) +
    geom_sf(aes(fill = PRUID), color = "gray60", size = 0.1, data = canada_cd) +
    geom_sf(data = wind_turbine_map, color = "#462872" , size = 2) +
    # coord_sf(crs = crs_string) +
    scale_fill_manual(values = map_colors) +
    guides(fill = FALSE) +
  labs(title="Location of Turbines in geo-spatial Space")+
    theme_classic() +
    theme(panel.grid.major = element_line(color = "white"),
          legend.key = element_rect(color = "gray40", size = 0.1))
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

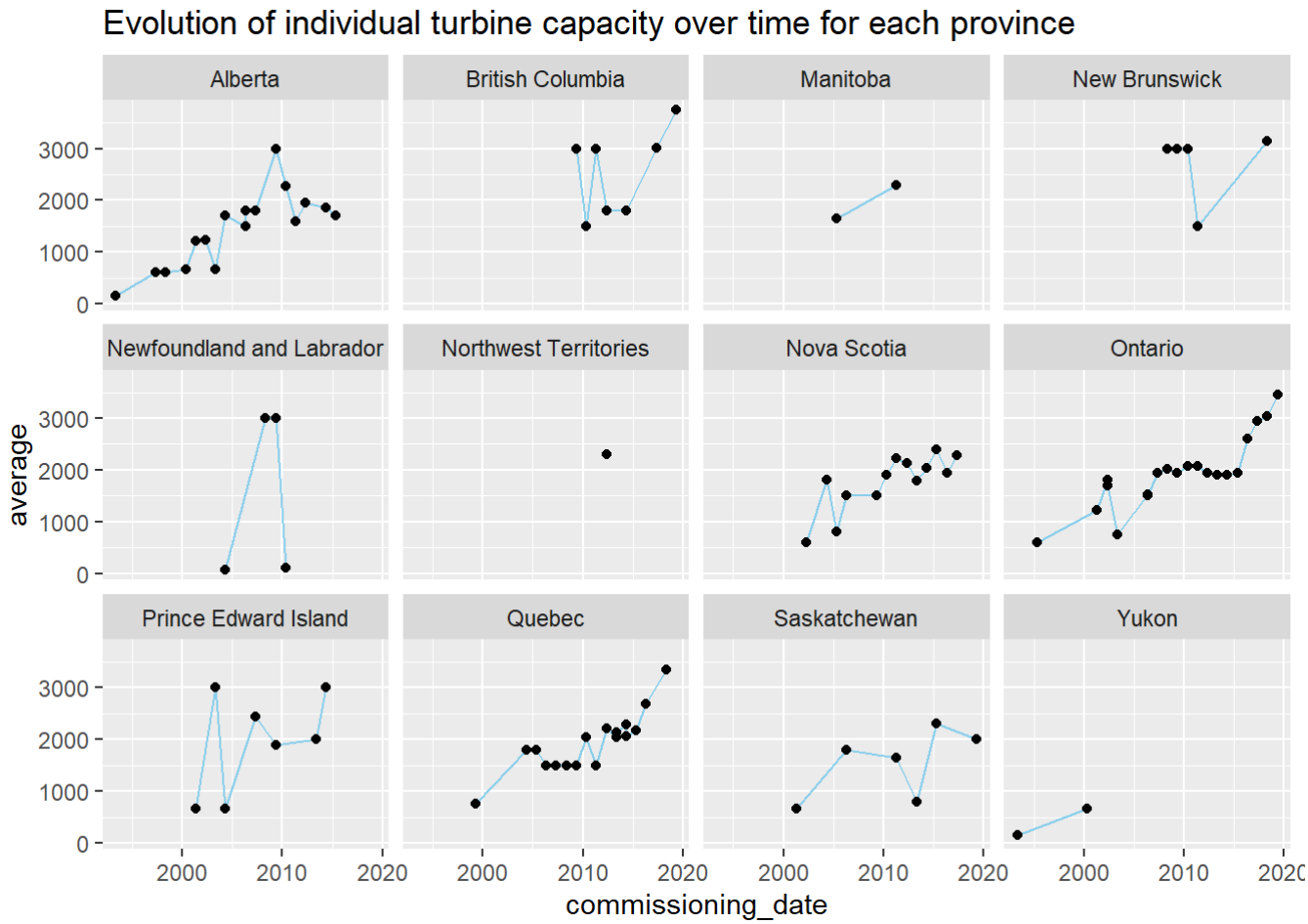



```
# Evolution of turbine capacity by time for each province

wind_turbine_date <- wind_turbine %>%
  group_by(commissioning_date, province_territory) %>%
  summarise(average = mean(turbine_rated_capacity_k_w, na.rm = T)) %>%
  mutate(commissioning_date = as.Date(as.character(commissioning_date), format = "%Y"))
```

```
## `summarise()` has grouped output by 'commissioning_date'. You can override using
## the `.groups` argument.
```

```
ggplot(wind_turbine_date, aes(x = commissioning_date,
             y = average)) +
  geom_line(color = "skyblue", size = 0.5) +
  geom_point() +
  facet_wrap(province_territory~.) +
  labs(title = "Evolution of individual turbine capacity over time for each province")
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```



Discussion:

First graph shows the count of wind turbines by province. It clearly shows that Ontario has the highest number of wind turbines followed by Quebec and Alberta. Moreover, Yukon and North West Territories has the least number of wind turbines

Second graph shows the number of turbines in geo-spatial space. It bolsters the analysis concluded by the first graph but also it shows the exact place where the turbine is installed. For e.g. In Ontario, we can see that huge number of wind turbines are installed near Toronto. It also shows where the turbine is not installed.(ie. where there are no blue points). We can see that there are very less number of wind turbines installed along the north-coast. We can also speculate about the un-tapped locations, where there is potential to install wind turbines in future

Third graph For Newfoundland and Labrador, Prince Edward Island, British Columbia, New Brunswick, one can observe that even though they have relatively less number of wind turbines, still the turbine capacity at some specific years is more.In Ontario, there is an increasing trend in terms of wind-capacity after 2004. Yukon not only has relatively very less no. of wind turbines but also don't have more wind-capacity.