

Econ 216 Presentation Figures

Kyler Rosen, Raunak Bhumsaria, Gordon Bradley, and Alex Illchev

2024-04-16

Data Wrangling

```
# Load necessary libraries
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## vforcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyverse  1.3.1
## v purrr    1.0.2
## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lubridate)
library(ggmap)

## Warning: package 'ggmap' was built under R version 4.3.1

## i Google's Terms of Service: <https://mapsplatform.google.com>
##   Stadia Maps' Terms of Service: <https://stadiamaps.com/terms-of-service/>
##   OpenStreetMap's Tile Usage Policy: <https://operations.osmfoundation.org/policies/tiles/>
## i Please cite ggmap if you use it! Use 'citation("ggmap")' for details.

library(dplyr)
library(patchwork)

us_map <- map_data("state")

# Load the data
data <- read.csv("uswtdb_v4_3_20220114.csv")

# Inspect the first few rows of the data
head(data)
```

```

##   case_id    faa_ors          faa_asn usgs_pr_id eia_id t_state    t_county
## 1 3072661                               5149  52161    CA Kern County
## 2 3072695                               5143  52161    CA Kern County
## 3 3072704                               5146  52161    CA Kern County
## 4 3063272 19-028134 2014-WTE-4084-0E      NA     NA    IA Story County
## 5 3053390 19-028015 2015-WTE-6386-0E      NA     NA    IA Boone County
## 6 3063269 19-028130 2016-WTE-5934-0E      NA     NA    IA Story County
##   t_fips          p_name p_year p_tnum p_cap t_manu    t_model t_cap
## 1   6029        251 Wind   1987    194 18.43 Vestas      95
## 2   6029        251 Wind   1987    194 18.43 Vestas      95
## 3   6029        251 Wind   1987    194 18.43 Vestas      95
## 4 19169 30 MW Iowa DG Portfolio 2017     10 30.00 Nordex AW125/3000 3000
## 5 19015 30 MW Iowa DG Portfolio 2017     10 30.00 Nordex AW125/3000 3000
## 6 19169 30 MW Iowa DG Portfolio 2017     10 30.00 Nordex AW125/3000 3000
##   t_hh t_rd    t_rsa t_ttlh retrofit retrofit_year t_conf_atr t_conf_loc
## 1   NA   NA     NA     NA     0      NA      2      3
## 2   NA   NA     NA     NA     0      NA      2      3
## 3   NA   NA     NA     NA     0      NA      2      3
## 4 87.5 125 12271.85    150     0      NA      3      3
## 5 87.5 125 12271.85    150     0      NA      3      3
## 6 87.5 125 12271.85    150     0      NA      3      3
##   t_img_date    t_img_srce      xlong      ylat
## 1 5/8/2018 Digital Globe -118.36376 35.07791
## 2 5/8/2018 Digital Globe -118.36441 35.07744
## 3 5/8/2018 Digital Globe -118.36420 35.07764
## 4 4/24/2017 Digital Globe -93.43037 42.02823
## 5 6/1/2017 Digital Globe -93.70042 41.97761
## 6 7/23/2017 Digital Globe -93.63284 41.88248

```

```

# Summarize the data to understand its structure
summary(data)

```

```

##   case_id    faa_ors          faa_asn usgs_pr_id
## Min. :3000001 Length:70808    Length:70808    Min. : 1
## 1st Qu.:3032230 Class :character  Class :character 1st Qu.:18626
## Median :3050978 Mode  :character  Mode  :character Median :28598
## Mean   :3058490                           Mean   :27524
## 3rd Qu.:3090448                           3rd Qu.:38720
## Max.   :3118671                           Max.   :49135
##                                     NA's   :32545
##   eia_id      t_state    t_county    t_fips
## Min. : 90 Length:70808 Length:70808 Min. : 2013
## 1st Qu.:56763 Class :character Class :character 1st Qu.:19081
## Median :57752 Mode  :character Mode  :character Median :35057
## Mean   :57878                           Mean   :32244
## 3rd Qu.:60338                           3rd Qu.:48141
## Max.   :65270                           Max.   :72133
## NA's   :5793
##   p_name      p_year    p_tnum    p_cap
## Length:70808 Min.   :1981  Min.   : 1.0  Min.   : 0.05
## Class :character 1st Qu.:2008  1st Qu.: 56.0  1st Qu.: 99.00
## Mode  :character Median :2012  Median : 85.0  Median :158.00
##                           Mean   :2012  Mean   :104.4  Mean   :170.18
##                           3rd Qu.:2018 3rd Qu.:121.0 3rd Qu.:211.22

```

```

##                               Max.   :2021   Max.   :731.0   Max.   :525.02
##                               NA's   :613                NA's   :4482
## t_manu                  t_model                  t_cap                  t_hh
## Length:70808      Length:70808      Min.   : 50   Min.   : 19.00
## Class :character      Class :character      1st Qu.:1500   1st Qu.: 80.00
## Mode  :character      Mode  :character      Median :2000    Median : 80.00
##                               Mean    :1964   Mean    : 81.06
##                               3rd Qu.:2300   3rd Qu.: 87.00
##                               Max.   :6000    Max.   :131.00
##                               NA's   :5480    NA's   :6180
## t_rd                    t_rsa                   t_ttlh                  retrofit
## Min.   : 13.40   Min.   : 141   Min.   : 30.4   Min.   :0.00000
## 1st Qu.: 82.00   1st Qu.: 5281  1st Qu.:121.0   1st Qu.:0.00000
## Median :100.00   Median : 7854   Median :130.1   Median :0.00000
## Mean   : 95.66   Mean   : 7619   Mean   :129.1   Mean   :0.08454
## 3rd Qu.:110.00   3rd Qu.: 9503  3rd Qu.:145.1   3rd Qu.:0.00000
## Max.   :155.00   Max.   :18869  Max.   :199.6   Max.   :1.00000
## NA's   :5934    NA's   :5934   NA's   :6180
## retrofit_year      t_conf_atr      t_conf_loc      t_img_date
## Min.   :2015      Min.   :1.000   Min.   :1.000   Length:70808
## 1st Qu.:2018      1st Qu.:3.000   1st Qu.:3.000   Class :character
## Median :2019      Median :3.000   Median :3.000   Mode  :character
## Mean   :2019      Mean   :2.767   Mean   :2.884
## 3rd Qu.:2020      3rd Qu.:3.000   3rd Qu.:3.000
## Max.   :2020      Max.   :3.000   Max.   :3.000
## NA's   :64822
## t_img_srce      xlong                  ylat
## Length:70808      Min.   :-171.71  Min.   :13.39
## Class :character      1st Qu.:-103.04  1st Qu.:34.43
## Mode  :character      Median : -99.39  Median :39.05
##                               Mean   :-100.09  Mean   :38.48
##                               3rd Qu.: -95.20  3rd Qu.:42.81
##                               Max.   : 144.72  Max.   :66.84
##
```

```

# Convert date to Date format
## We want it in the Date format since it allows us to do easier modifications to the variables
data$t_img_date <- as.Date(data$t_img_date, format = "%m/%d/%Y")

## Data pre-2015 seems very unreliable pre-2015 so I am cutting off before then.
data_after_2015 <- data %>%
  filter(t_img_date > as.Date("2014-12-31"))

# Aggregate data by month
## Data is very noisy so collecting by month makes the data bit easier to parse.
monthly_data <- data_after_2015 %>%
  group_by(month = floor_date(t_img_date, "month")) %>%
  summarise(count = n())

## Need to establish bounding boxes on geographic areas for the maps
continental_bbox <- c(left = -125, bottom = 24, right = -66, top = 49)
alaska_bbox <- c(left = -170, bottom = 52, right = -130, top = 72)

```

```

hawaii_bbox <- c(left = -160, bottom = 18, right = -154, top = 23)

## Grabbing all numeric data for a rough analysis
numeric_data <- data %>% select_if(is.numeric)

register_stadiamaps("f94c64ea-35d9-425f-af7a-e139e3bd6242", write = TRUE)

## i Replacing old key (f94c64ea) with new key in /Users/kyler/.Renviron

continental_map <- get_stadiamap(bbox = continental_bbox, zoom = 5, maptype = "stamen_toner_lite")

## i © Stadia Maps © Stamen Design © OpenMapTiles © OpenStreetMap contributors.

continental_plot <- ggmap(continental_map) +
  geom_point(data = data, aes(x = xlong, y = ylat), color = "red", size = 1) +
  labs(x = "Longitude", y = "Latitude", title = "Geographic Distribution of Turbines in Continental US") +
  theme_minimal() +
  theme(legend.position = "none")

ggsave("figures/continental_us_turbines.png", continental_plot, bg = "transparent")

## Saving 6.5 x 4.5 in image

## Warning: Removed 337 rows containing missing values ('geom_point()').

alaska_map <- get_stadiamap(bbox = alaska_bbox, zoom = 4, maptype = "stamen_toner_lite")

## i © Stadia Maps © Stamen Design © OpenMapTiles © OpenStreetMap contributors.

alaska_plot <- ggmap(alaska_map) +
  geom_point(data = data, aes(x = xlong, y = ylat), color = "red", size = 1) +
  labs(x = "Longitude", y = "Latitude", title = "Geographic Distribution of Turbines in Alaska") +
  theme_minimal() +
  theme(legend.position = "none")

ggsave("figures/alaska_turbines.png", alaska_plot, bg = "transparent")

## Saving 6.5 x 4.5 in image

## Warning: Removed 70675 rows containing missing values ('geom_point()').

hawaii_map <- get_stadiamap(bbox = hawaii_bbox, zoom = 7, maptype = "stamen_toner_lite")

## i © Stadia Maps © Stamen Design © OpenMapTiles © OpenStreetMap contributors.

hawaii_plot <- ggmap(hawaii_map) +
  geom_point(data = data, aes(x = xlong, y = ylat), color = "red", size = 1) +
  labs(x = "Longitude", y = "Latitude", title = "Geographic Distribution of Turbines in Hawaii") +
  theme_minimal() +
  theme(legend.position = "none")

ggsave("figures/hawaii_turbines.png", hawaii_plot, bg = "transparent")

```

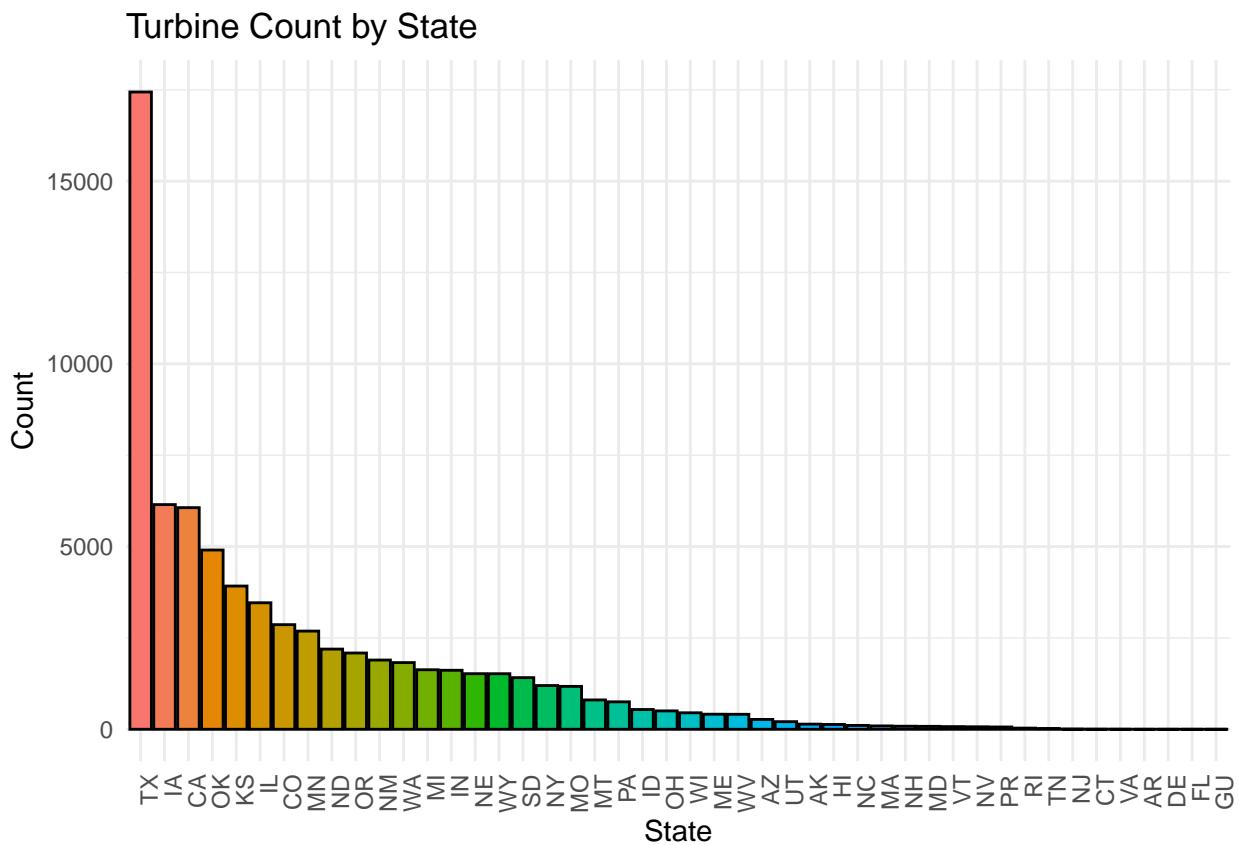
```
## Saving 6.5 x 4.5 in image  
  
## Warning: Removed 70676 rows containing missing values ('geom_point()').
```

```
state_counts <- data %>%
  group_by(t_state) %>%
  summarise(Count = n()) %>%
  ungroup() %>%
  arrange(desc(Count))

state_counts$t_state <- factor(state_counts$t_state, levels = state_counts$t_state)

bar <- ggplot(state_counts, aes(x = t_state, y = Count, fill = t_state)) +
  geom_bar(stat = "identity", color = "black") +
  labs(x = "State", y = "Count", title = "Turbine Count by State") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(legend.position = "none")

print(bar)
```



```
ggsave("figures/state_turbines_amount.png", bar, bg = "transparent")
```

```
## Saving 6.5 x 4.5 in image
```

```

gaussian_jitter <- function(data, amount=1) {
  data + rnorm(length(data), mean=0, sd=amount)
}

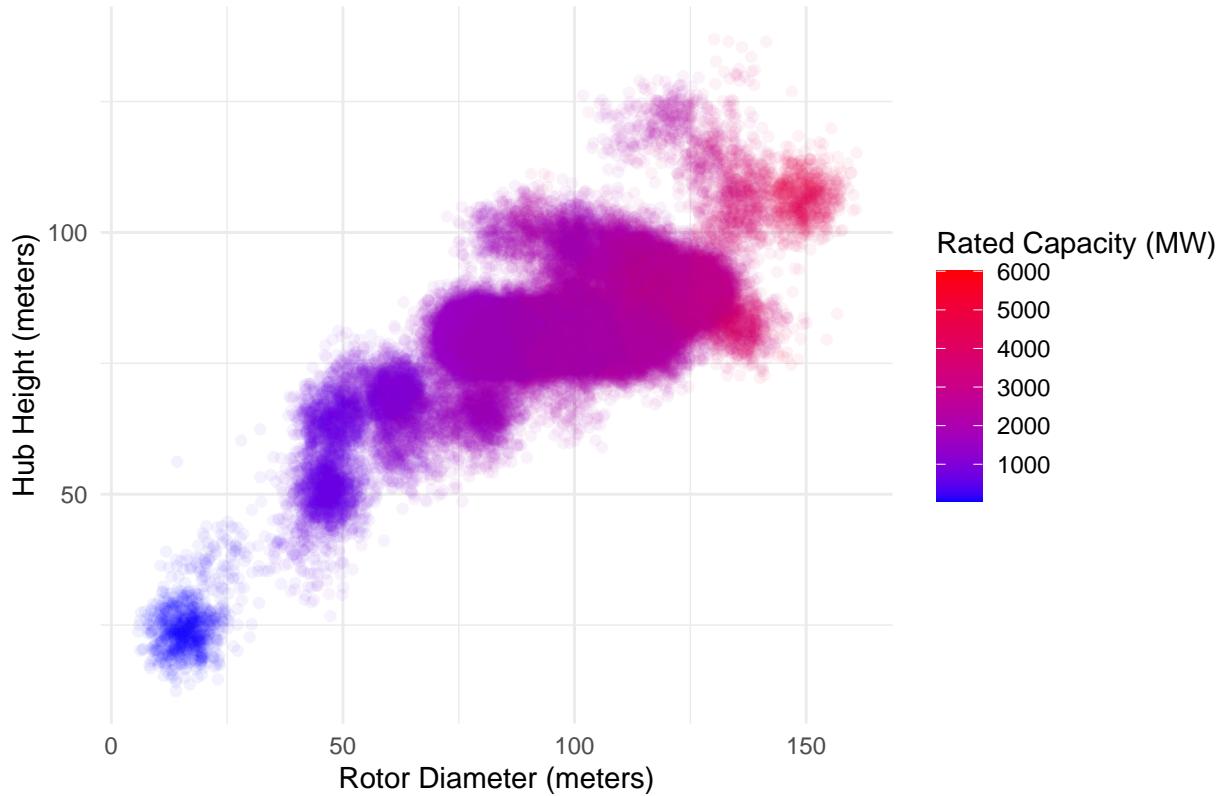
plot <- ggplot(data, aes(x = gaussian_jitter(t_rd, 3), y = gaussian_jitter(t_hh, 3), color = t_cap)) +
  geom_point(position = position_jitter(width = 3, height = 3), alpha = 0.05) +
  scale_color_gradient(low = "blue", high = "red") +
  labs(x = "Rotor Diameter (meters)",
       y = "Hub Height (meters)",
       color = "Rated Capacity (MW)",
       title = "Turbine Rotor Diameter vs Hub Height by Rated Capacity") +
  theme_minimal()

print(plot)

```

Warning: Removed 6182 rows containing missing values ('geom_point()').

Turbine Rotor Diameter vs Hub Height by Rated Capacity



```
ggsave("figures/specs_effect.png", plot, bg = "transparent")
```

Saving 6.5 x 4.5 in image

Warning: Removed 6182 rows containing missing values ('geom_point()').

```

continental_map <- get_stadiamap(bbox = continental_bbox, zoom = 5, maptype = "stamen_toner_lite")

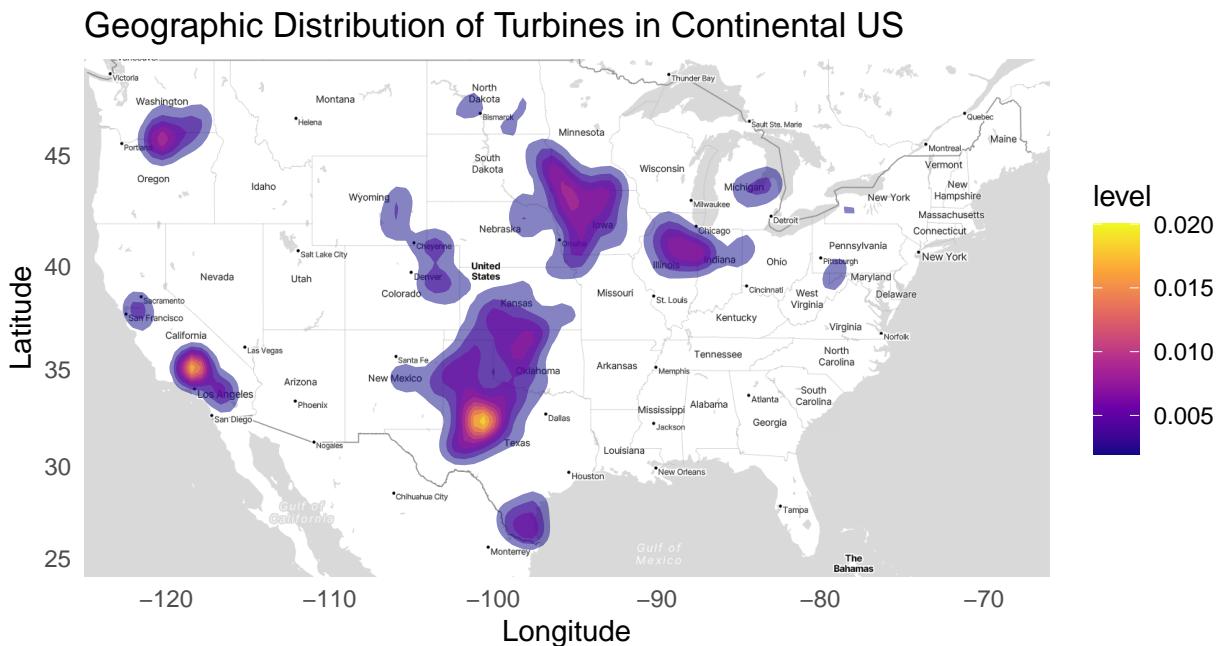
## i © Stadia Maps © Stamen Design © OpenMapTiles © OpenStreetMap contributors.

continental_plot <- ggmap(continental_map) +
  stat_density2d(data = data, aes(x = xlong, y = ylat, fill = ..level..), geom = "polygon", alpha = 0.5)
  scale_fill_viridis_c(option = "C") +
  labs(x = "Longitude", y = "Latitude", title = "Geographic Distribution of Turbines in Continental US") +
  theme_minimal() +
  theme(legend.position = "right")
print(continental_plot)

## Warning: The dot-dot notation ('..level..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(level)' instead.
## i The deprecated feature was likely used in the ggmap package.
## Please report the issue at <https://github.com/dkahle/ggmap/issues>.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

## Warning: Removed 337 rows containing non-finite values ('stat_density2d()').

```



```

ggsave("figures/continental_turbines_heatmap.png", continental_plot, bg = "transparent")

## Saving 6.5 x 4.5 in image

## Warning: Removed 337 rows containing non-finite values ('stat_density2d()').

plot2 <- ggplot(data, aes_string(x = "t_cap")) +
  geom_histogram(binwidth = 100, fill = "blue", color = "black") +
  labs(x = "Turbine Rated Capacity (kW)", y = "Frequency", title = "Distribution of Turbine Rated Capacity (kW)", subtitle = "Continents", fill = "Continent") +
  theme_minimal()

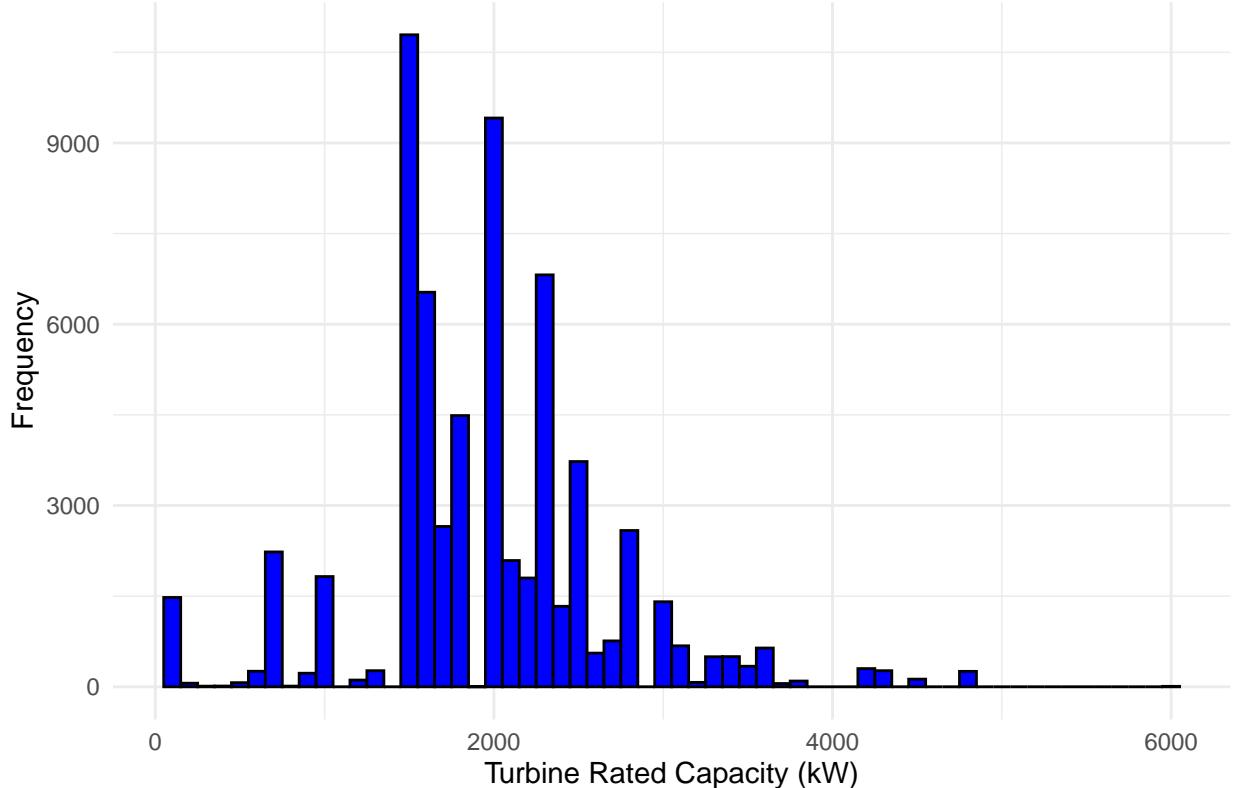
## Warning: 'aes_string()' was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with 'aes()'.
## i See also 'vignette("ggplot2-in-packages")' for more information.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

print(plot2)

```

Warning: Removed 5480 rows containing non-finite values ('stat_bin()').

Distribution of Turbine Rated Capacity (kW)



```

ggsave("figures/t_cap.png", plot = plot2, bg = "transparent")

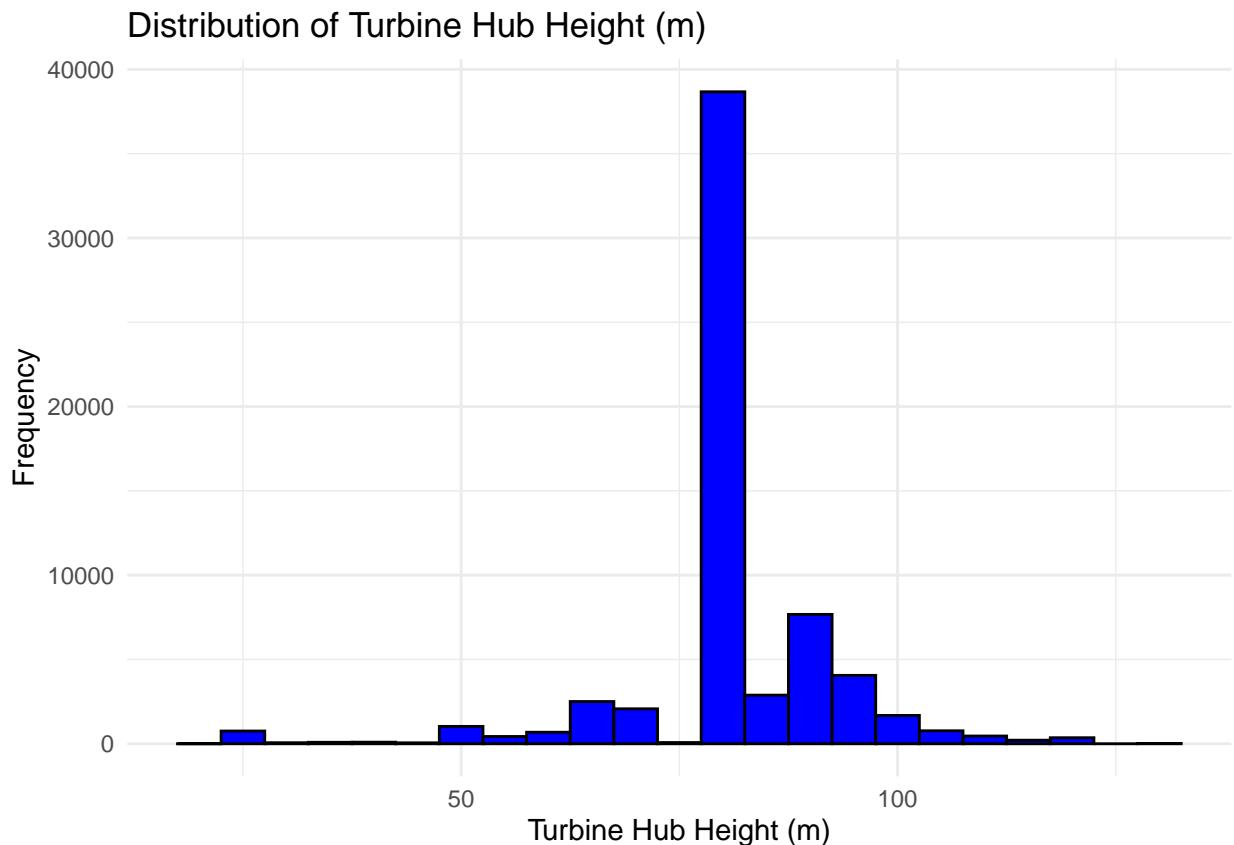
## Saving 6.5 x 4.5 in image

## Warning: Removed 5480 rows containing non-finite values ('stat_bin()').

plot3 <- ggplot(data, aes_string(x = "t_hh")) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  labs(x = "Turbine Hub Height (m)", y = "Frequency", title = "Distribution of Turbine Hub Height (m)")
print(plot3)

```

```
## Warning: Removed 6180 rows containing non-finite values ('stat_bin()').
```



```
ggsave("figures/hub_height.png", plot = plot3, bg = "transparent")
```

```
## Saving 6.5 x 4.5 in image
```

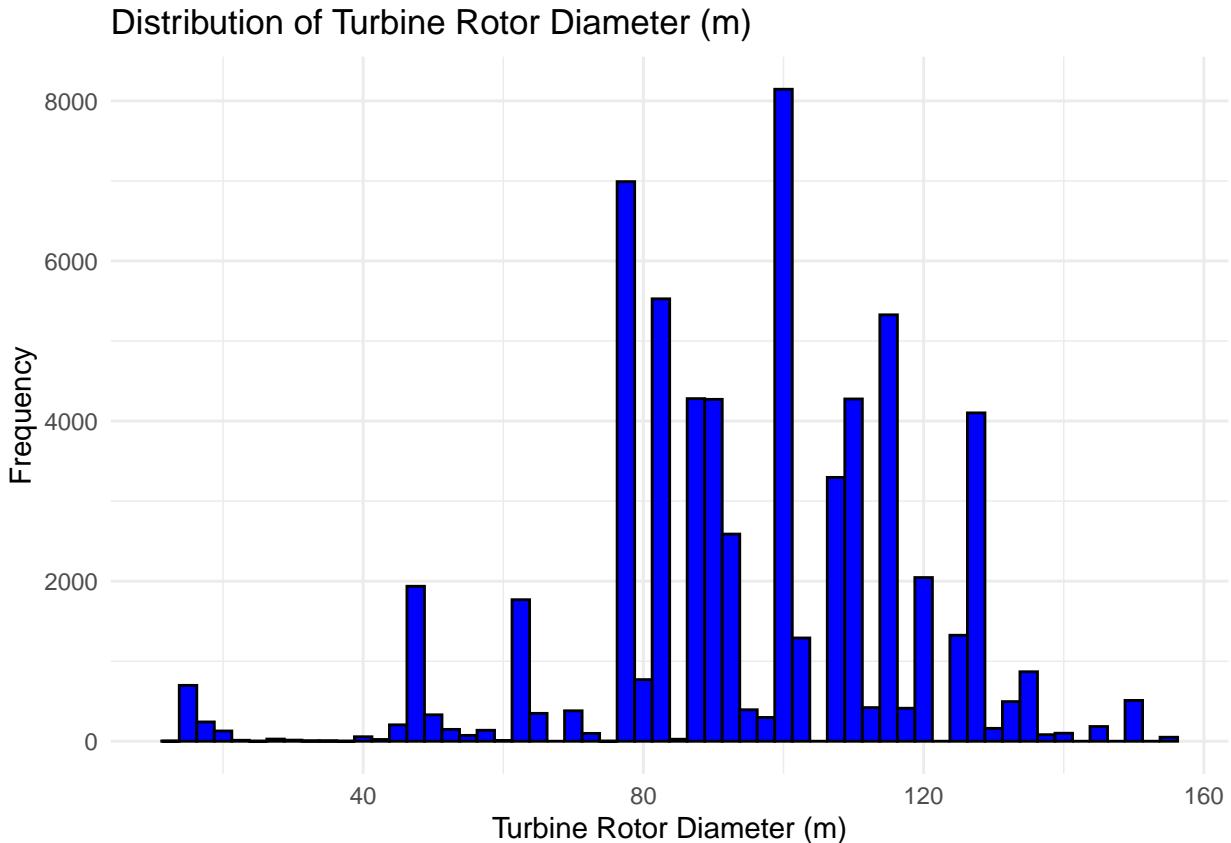
```
## Warning: Removed 6180 rows containing non-finite values ('stat_bin()').
```

```

plot4 <- ggplot(data, aes_string(x = "t_rd")) +
  geom_histogram(binwidth = 2.5, fill = "blue", color = "black") +
  labs(x = "Turbine Rotor Diameter (m)", y = "Frequency", title = "Distribution of Turbine Rotor
  theme_minimal()
print(plot4)

```

Warning: Removed 5934 rows containing non-finite values ('stat_bin()').



```
ggsave("figures/rotor_diameter.png", plot = plot4, bg = "transparent")
```

Saving 6.5 x 4.5 in image

Warning: Removed 5934 rows containing non-finite values ('stat_bin()').

```
register_stadiamaps("f94c64ea-35d9-425f-af7a-e139e3bd6242", write = TRUE)
```

i Replacing old key (f94c64ea) with new key in /Users/kyler/.Renviron

```
continental_bbox <- c(left = -125, bottom = 25, right = -66, top = 49) # Adjust as needed
continental_map <- get_stadiamap(bbox = continental_bbox, zoom = 5, maptype = "stamen_toner_lite")
```

i © Stadia Maps © Stamen Design © OpenMapTiles © OpenStreetMap contributors.

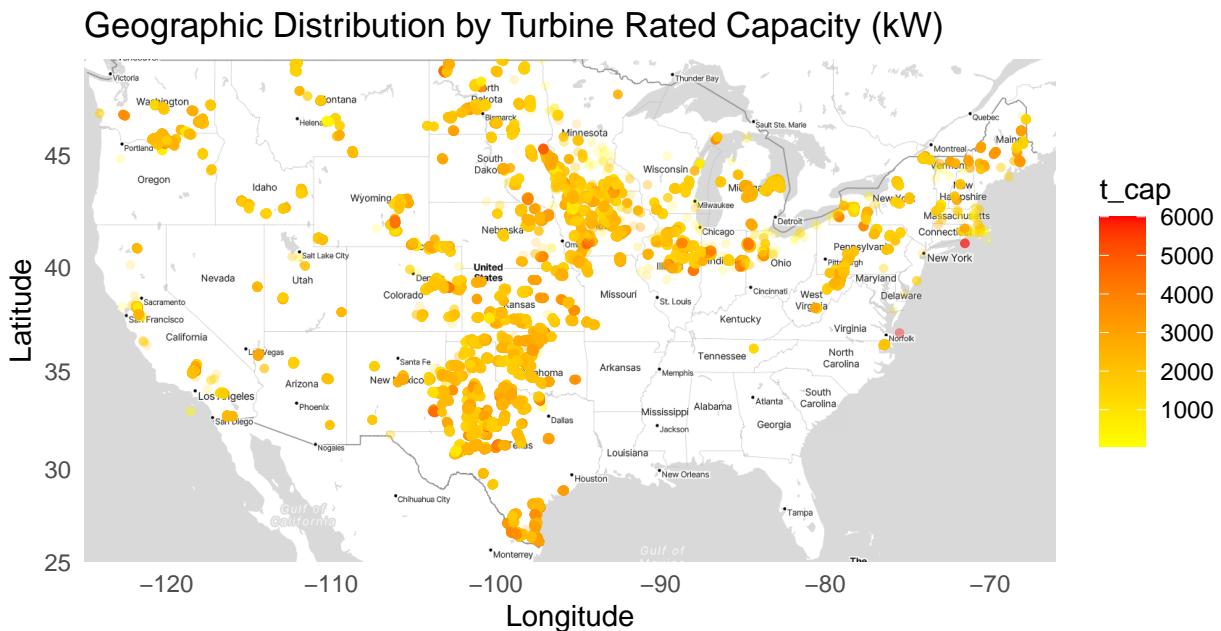
```

data_filtered <- data %>%
  filter(!is.na(p_cap), !is.na(t_cap), !is.na(t_hh), !is.na(t_rd))

plot2 <- ggmap(continental_map) +
  geom_point(data = data_filtered, aes(x = xlong, y = ylat, color = t_cap), size = 1, alpha = 0.2) +
  scale_color_gradient(low = "yellow", high = "red") +
  labs(x = "Longitude", y = "Latitude", title = "Geographic Distribution by Turbine Rated Capacity (kW") +
  theme_minimal()
print(plot2)

```

Warning: Removed 312 rows containing missing values ('geom_point()').



```
ggsave("figures/t_cap_map.png", plot = plot2, bg = "transparent")
```

Saving 6.5 x 4.5 in image

Warning: Removed 312 rows containing missing values ('geom_point()').

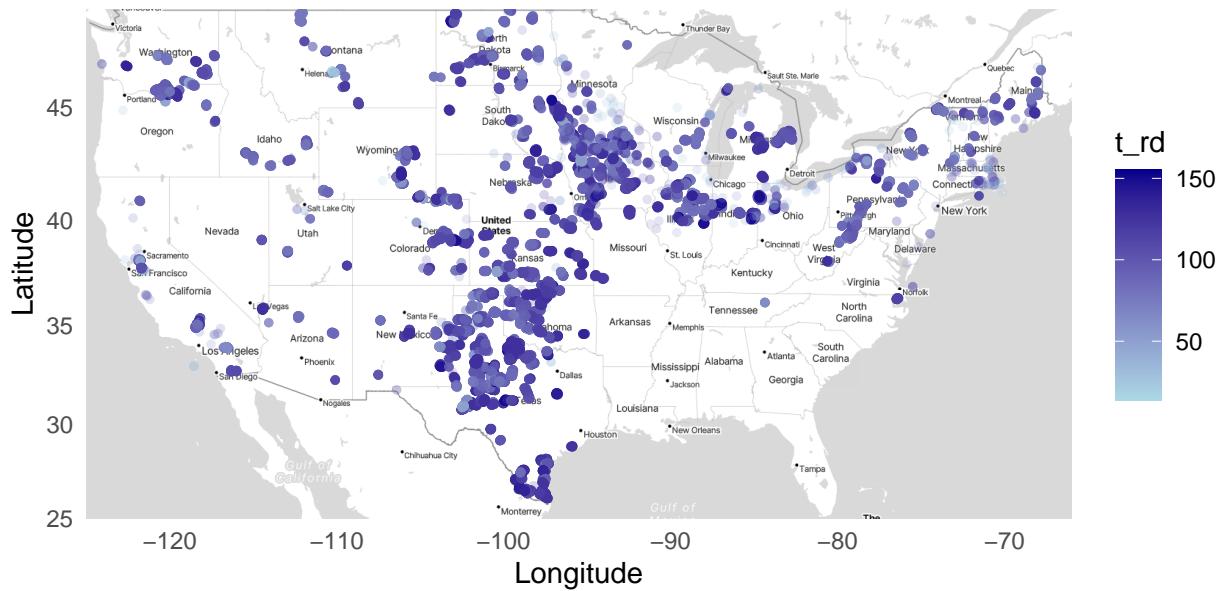
```

plot4 <- ggmap(continental_map) +
  geom_point(data = data_filtered, aes(x = xlong, y = ylat, color = t_rd), size = 1, alpha = 0.2) +
  scale_color_gradient(low = "lightblue", high = "darkblue") +
  labs(x = "Longitude", y = "Latitude", title = "Geographic Distribution by Turbine Rotor Diameter (m") +
  theme_minimal()
print(plot4)

```

```
## Warning: Removed 312 rows containing missing values ('geom_point()').
```

Geographic Distribution by Turbine Rotor Diameter (m)



```
ggsave("figures/rotor_diameter_map.png", plot = plot4, bg = "transparent")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 312 rows containing missing values ('geom_point()').
```