Lab 4 | Refining Visualizations & Visualizing Uncertainty

ST 437 Data Visualization

Student Name

April 23, 2025

Getting Started

First, ensure you have the necessary packages installed and loaded.

> **!** Downloading R-packages
>
> Use `install.packages('Name of Package')` to install any R packages you don't have.

```
library(tidyverse)
library(ggridges)
library(viridis)
```

Before moving on to the remainder of this activity, try rendering the document to both html and pdf. It's generally a good idea to render your document periodically so that if there's an issue, you can spot it more easily.

Load the Data

Make sure you have the `salem_weather_2024.csv` file downloaded from Canvas and saved in the same location as this .qmd file. The dataset contains temperature and wind speed information for Salem, Oregon in 2024.

It's helpful to have your code chunks named. Give the following chunk a label by clicking on the small, faint gear icon in the upper right corner of the chunk. In the Chunk Name field, specify a name for the code chunk below. I recommend doing this for all remaining chunks!

```
salem24 <- read_csv("salem_weather_2024.csv")
```

The dataset contains four variables

- `DATE`: the date
- `TMAX`: the maximum temperature in tenths of degrees Celsius
- `TMIN`: the minimum temperature in tenths of degrees Celsius
- `AWND`: the average daily wind speed in tenths of meters per second

Clean the Data

Currently the units that temperature and wind speed are reported in area a little unusual. At least not units we commonly communicate in. Use the `mutate` function to convert the temperature variables to degrees Fahrenheit and the wind speed variable to miles per hour. Use the following conversions:

- one tenth degrees Celsius can by converted to degrees Fahrenheit by multiplying by 0.18 then adding 32
- one tenth meters per second is approximately 0.2237 miles per hour

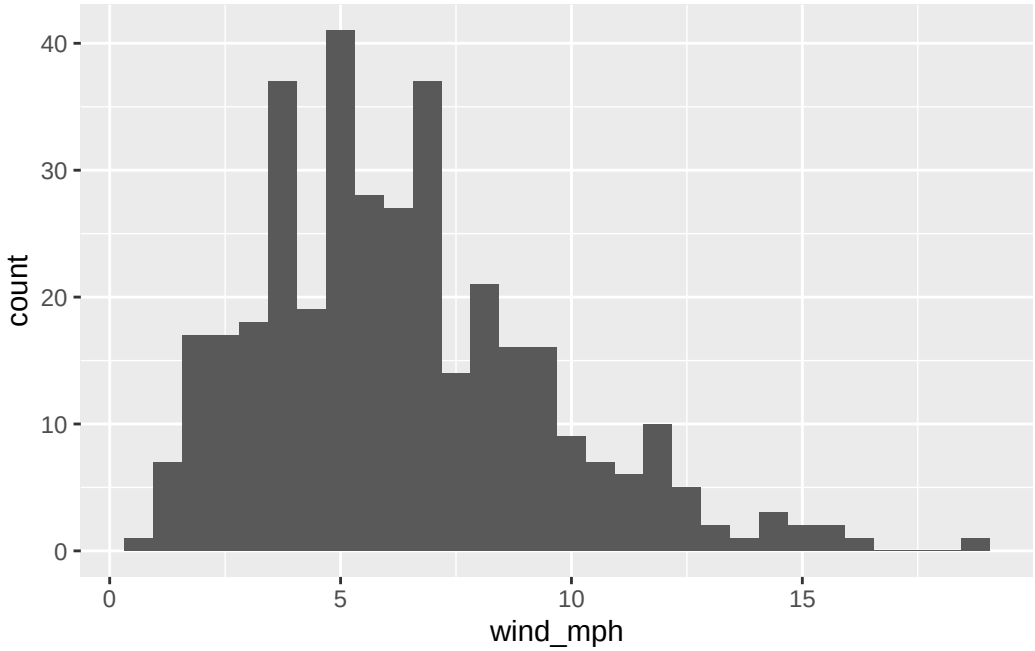
Additionally, remove any rows with `NA` values (hint: use `drop_na()` at the end of your pipe sequence.

```
salem24 <- salem24 |>
  mutate(tmax_f = (TMAX*0.18)+32,
         tmin_f = (TMIN*0.18)+32,
         wind_mph = AWND*0.2237) |>
  drop_na()
```

Histograms

Create a basic (not too polished) histogram that displays the distribution of average daily wind speed (in miles per hour).

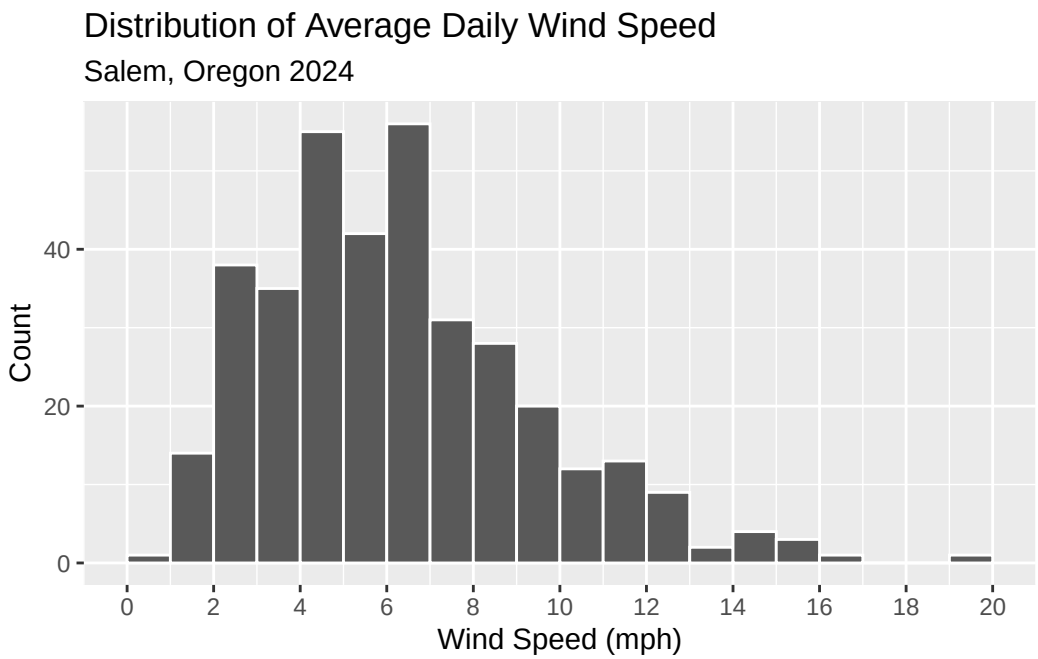
```
# create a basic histogram
ggplot(salem24, aes(x = wind_mph)) +
  geom_histogram()
```



Now create a polished version of the histogram. Polishing checklist:

- Plot Title
- Informative labels
- Set a binwidth
- Set the `boundary` argument in `geom_histogram` to 0 so that the bins are aligned with whole numbers
- Add white outlines to the bins so that they can be seen more distinctly
- Add x-axis breaks at each even number within the range of the windspeeds (hint: use `scale_x_continuous(breaks = seq(0, 20, by = 2))`)

```
ggplot(salem24, aes(x = wind_mph)) +
  geom_histogram(binwidth = 1,
                 boundary = 0,
                 color = "white") +
  scale_x_continuous(breaks = seq(0,20,by=2)) +
  labs(title = "Distribution of Average Daily Wind Speed",
       subtitle = "Salem, Oregon 2024",
       x = "Wind Speed (mph)",
       y = "Count")
```



Facet by Month

Create a new column in the dataset that contains the month of the observation. Within `mutate` use the `month` function to extract the month from `DATE`. To have the displayed by its full name, use the `label=TRUE` and `abbr=FALSE` arguments in the `month` function.

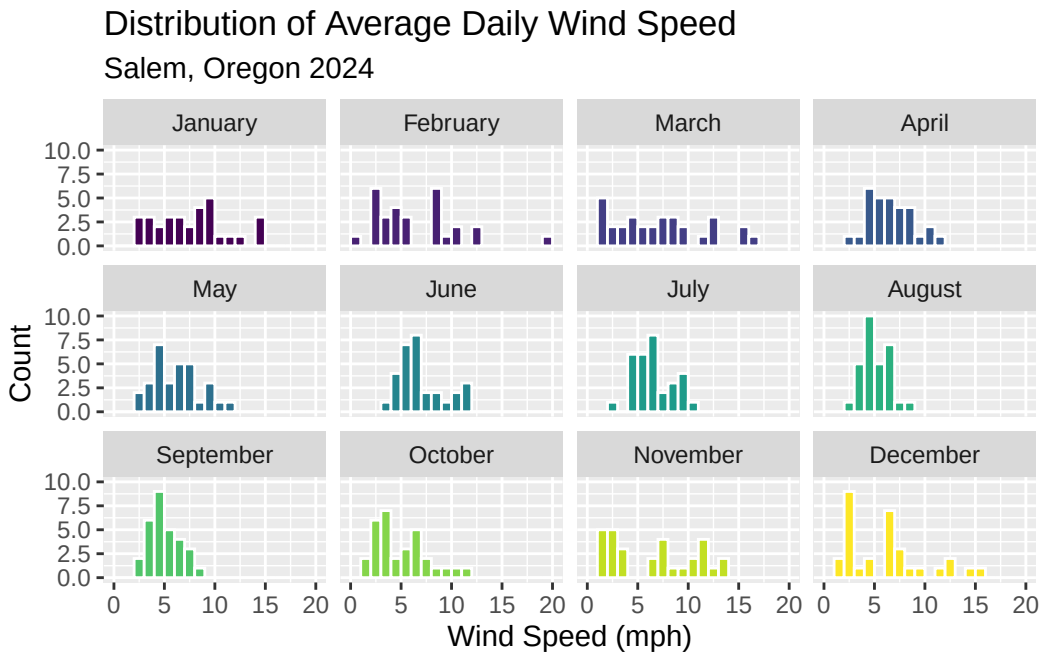
```
salem24 <- salem24 |>
  mutate(month = month(DATE, label = TRUE, abbr = FALSE))
```

Create a series of histograms, one for each month, by faceting by month. Polishing checklist:

- Plot Title
- Informative labels
- Set a binwidth
- Set the `boundary` argument in `geom_histogram` to 0 so that the bins are aligned with whole numbers
- Add white outlines to the bins so that they can be seen more distinctly
- Adjust x-axis breaks so that the labels are readable (hint: use `scale_x_continuous`)
- Add color to each month's histogram

- Remove the unnecessary legend since each facet has a label (hint: use `guides(fill = "none")`)

```
ggplot(salem24, aes(x = wind_mph)) +
  geom_histogram(aes(fill = month),
                 binwidth = 1,
                 color = "white",
                 boundary = 0) +
  scale_x_continuous(breaks = seq(0,20,by=5)) +
  guides(fill = "none") +
  facet_wrap(vars(month)) +
  labs(title = "Distribution of Average Daily Wind Speed",
       subtitle = "Salem, Oregon 2024",
       x = "Wind Speed (mph)",
       y = "Count")
```

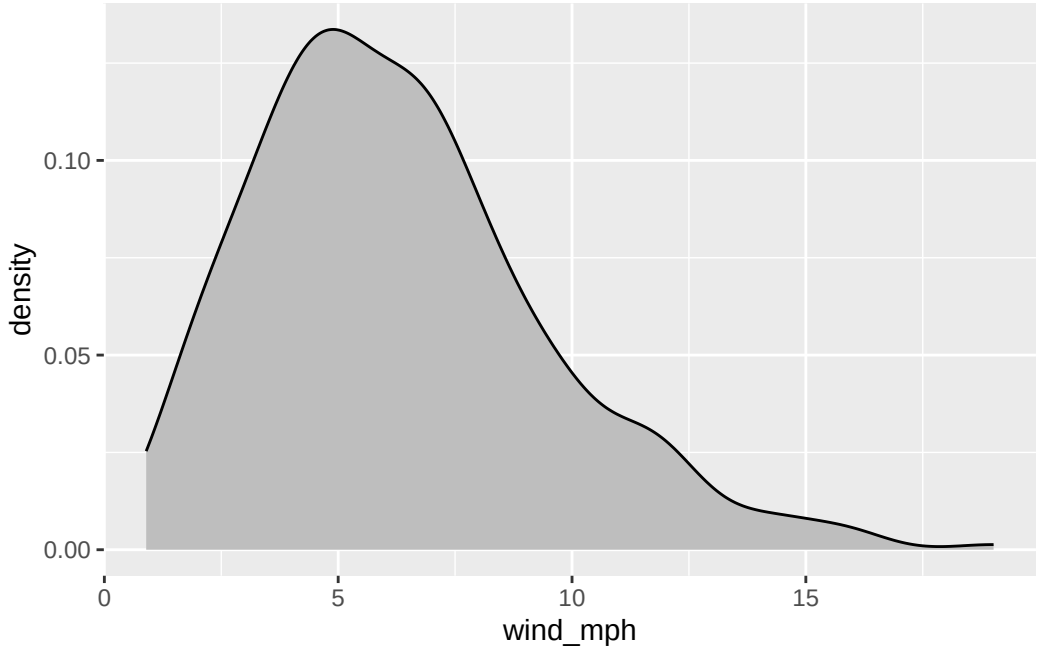


Consider: What information do the faceted histograms provide that the single histogram didn't?

Density Plots

Create a basic (not too polished) density plot that displays the distribution of average daily wind speed (in miles per hour).

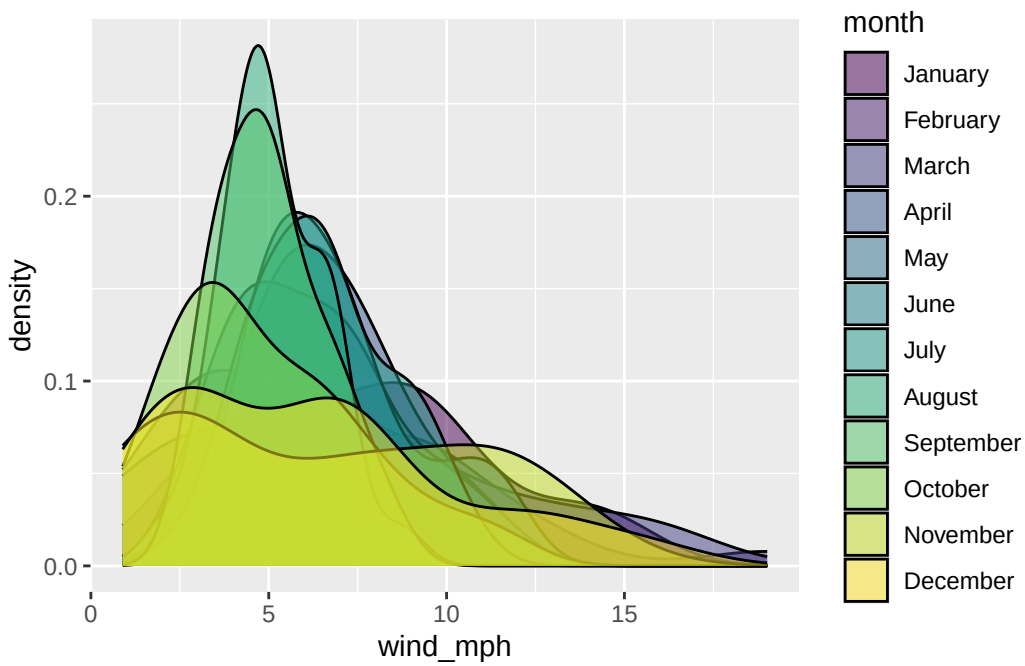
```
ggplot(salem24, aes(x = wind_mph)) +
  geom_density(fill = "grey")
```



Overlaid Density Plots

Overlay transparent density plots filled by month.

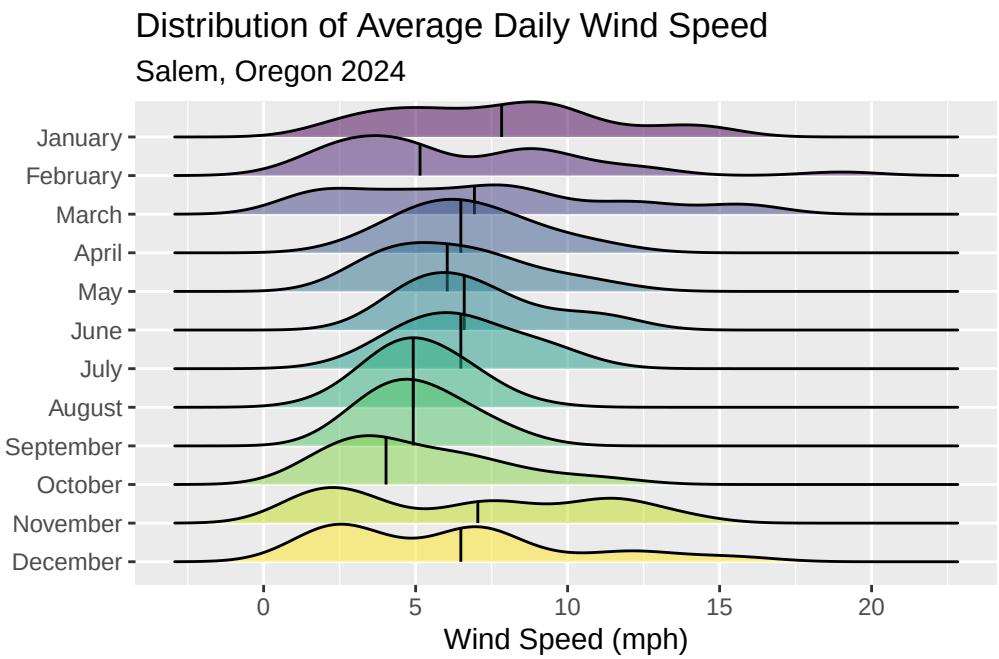
```
ggplot(salem24, aes(x = wind_mph)) +
  geom_density(aes(fill = month),
               alpha = 0.5)
```



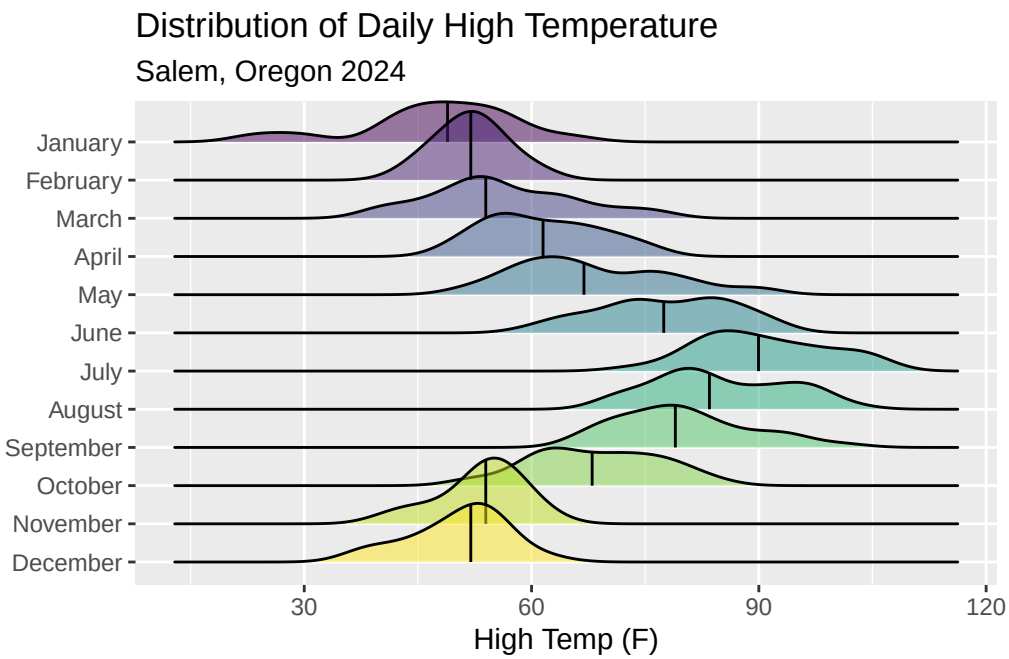
Ridge Plots

Even when we use transparent density plots, with 12 different levels of month, its difficult to interpret anything meaningful from the plot. Try using a ridge plot instead.

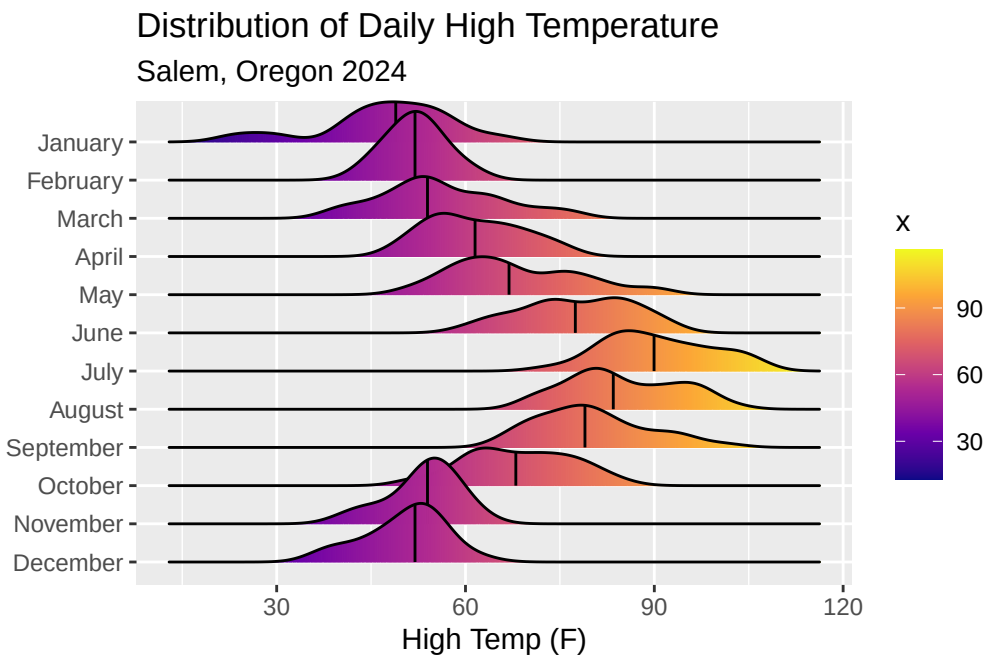
```
ggplot(salem24, aes(x = wind_mph, y = fct_rev(month))) +
  geom_density_ridges(aes(fill = month),
                      alpha = 0.5,
                      quantile_lines = TRUE,
                      quantiles = 2) +
  coord_cartesian(clip = "off") +
  scale_x_continuous(breaks = seq(0,20,by=5)) +
  guides(fill = "none") +
  labs(title = "Distribution of Average Daily Wind Speed",
       subtitle = "Salem, Oregon 2024",
       x = "Wind Speed (mph)",
       y = "")
```




```
ggplot(salem24, aes(x = tmax_f, y = fct_rev(month))) +
  geom_density_ridges(aes(fill = month),
                      alpha = 0.5,
                      quantile_lines = TRUE,
                      quantiles = 2) +
  coord_cartesian(clip = "off") +
  guides(fill = "none") +
  labs(title = "Distribution of Daily High Temperature",
       subtitle = "Salem, Oregon 2024",
       x = "High Temp (F)",
       y = "")
```

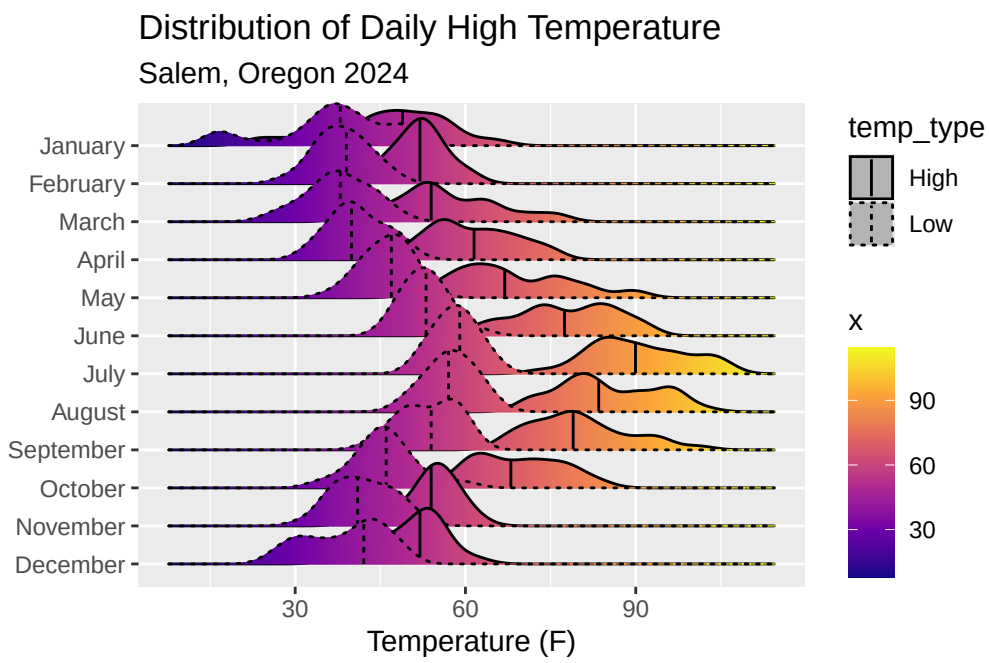



```
ggplot(salem24, aes(x = tmax_f, y = fct_rev(month))) +
  geom_density_ridges_gradient(aes(fill = after_stat(x)),
                       alpha = 0.5,
                       quantile_lines = TRUE,
                       quantiles = 2) +
  scale_fill_viridis_c(option = "plasma") +
  coord_cartesian(clip = "off") +
  labs(title = "Distribution of Daily High Temperature",
       subtitle = "Salem, Oregon 2024",
       x = "High Temp (F)",
       y = "")
```




```
salem24_long <- salem24 |>
  pivot_longer(cols = c(tmin_f, tmax_f),
               names_to = "temp_type",
               values_to = "temp") |>
  mutate(temp_type = recode(temp_type,
                            tmin_f = "Low",
                            tmax_f = "High")) |>
  select(DATE, temp_type, temp, month)
```

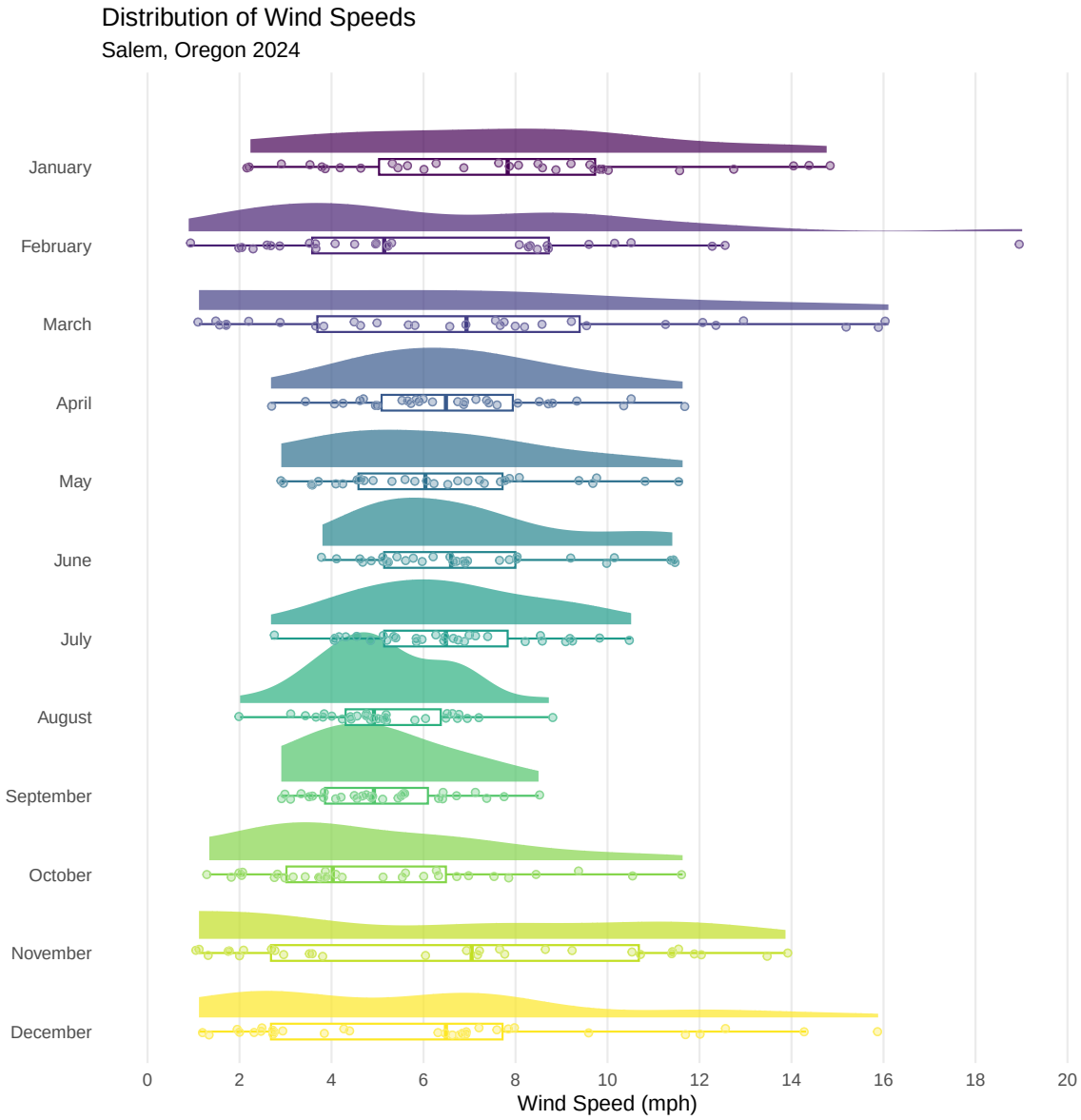
```
ggplot(salem24_long, aes(x = temp,
                         y = fct_rev(month),
                         fill = after_stat(x),
                         linetype = temp_type)) +
  geom_density_ridges_gradient(
                      alpha = 0.5,
                      quantile_lines = TRUE,
                      quantiles = 2) +
  scale_fill_viridis_c(option = "plasma") +
  coord_cartesian(clip = "off") +
  labs(title = "Distribution of Daily High Temperature",
       subtitle = "Salem, Oregon 2024",
       x = "Temperature (F)",
       y = "")
```



Rain Cloud Plots

```
salem24 |>
  ggplot(
    aes(x = fct_rev(month), y = wind_mph)
  ) +
  ggdist::stat_halfeye(
    aes(color = month, fill = month),
    point_color = NA, .width = 0, #adjust = -0.25,
    #width = 0.75,
    justification = -0.2,
    alpha = 0.7
  ) +
  geom_boxplot(
    aes(color = month),
    outlier.shape = NA,
    width = 0.2
  ) +
  geom_point(
```

```
  aes(color = month),
  shape = 21,
  size = 1.5,
  position = position_jitter(seed = 1, width = 0.05)
) +
geom_point(
  aes(fill = month),
  color = "transparent",
  shape = 21,
  size = 1.5,
  alpha = 0.3,
  position = position_jitter(seed = 1, width = 0.05)
) +
coord_flip(xlim = c(1.2, NA), clip = "off") +
scale_y_continuous(
  limits = c(0, 20),
  breaks = seq(0, 20, by = 2)
) +
guides(fill = "none", color = "none") +
labs(x = NULL,
     y = "Wind Speed (mph)",
     title = "Distribution of Wind Speeds",
     subtitle = "Salem, Oregon 2024") +
theme_minimal() +
theme(
  panel.grid.minor = element_blank(),
  panel.grid.major.y = element_blank(),
  axis.ticks = element_blank()
)
```


Distribution of Wind Speeds
Salem, Oregon 2024
January
February
March
April
May
June
July
August
September
October
November
December
0
2
4
6
8
10
12
14
16
18
20
Wind Speed (mph)