Vis II

Import the weather data

```r
weather_df =
  rnoaa::meteo_pull_monitors(
    c("USW00094728", "USW00022534", "USS0023B17S"),
    var = c("PRCP", "TMIN", "TMAX"),
    date_min = "2021-01-01",
    date_max = "2022-12-31") |>
  mutate(
    name = case_match(
      id,
      "USW00094728" ~ "CentralPark_NY",
      "USW00022534" ~ "Molokai_HI",
      "USS0023B17S" ~ "Waterhole_WA"),
    tmin = tmin / 10,
    tmax = tmax / 10) |>
  select(name, id, everything())
```

```
## using cached file: /Users/soomin.you/Library/Caches/org.R-project.R/R/rnoaa/noaa_ghcnd/USW00094728.d

## date created (size, mb): 2024-09-03 14:09:15.067935 (8.636)

## file min/max dates: 1869-01-01 / 2024-09-30

## using cached file: /Users/soomin.you/Library/Caches/org.R-project.R/R/rnoaa/noaa_ghcnd/USW00022534.d

## date created (size, mb): 2024-09-03 14:09:24.583853 (3.913)

## file min/max dates: 1949-10-01 / 2024-09-30

## using cached file: /Users/soomin.you/Library/Caches/org.R-project.R/R/rnoaa/noaa_ghcnd/USS0023B17S.d

## date created (size, mb): 2024-09-03 14:09:27.654133 (1.036)

## file min/max dates: 1999-09-01 / 2024-08-31
```

```r
weather_df
```

```
## # A tibble: 2,190 x 6
##    name           id          date         prcp  tmax  tmin
##    <chr>          <chr>       <date>       <dbl> <dbl> <dbl>
##  1 CentralPark_NY USW00094728 2021-01-01     157   4.4   0.6
```

```
##  2 CentralPark_NY USW00094728 2021-01-02      13  10.6   2.2
##  3 CentralPark_NY USW00094728 2021-01-03      56   3.3   1.1
##  4 CentralPark_NY USW00094728 2021-01-04       5   6.1   1.7
##  5 CentralPark_NY USW00094728 2021-01-05       0   5.6   2.2
##  6 CentralPark_NY USW00094728 2021-01-06       0   5     1.1
##  7 CentralPark_NY USW00094728 2021-01-07       0   5    -1
##  8 CentralPark_NY USW00094728 2021-01-08       0   2.8  -2.7
##  9 CentralPark_NY USW00094728 2021-01-09       0   2.8  -4.3
## 10 CentralPark_NY USW00094728 2021-01-10       0   5    -1.6
## # i 2,180 more rows
```

Make a scatterplot but fancy this time.

```
weather_df |>
  ggplot(aes(x = tmin, y = tmax, color = name)) +
  geom_point(alpha = .3) +
  labs(
    title = "Temperature scatterplot",
    x = "Minimum Temp(C)",
    y = "Maximum Temp (C)",
    color = "Location",
    caption = "Weather data taken from rnoaa package for three stations."
  )
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## ('geom_point()').
```
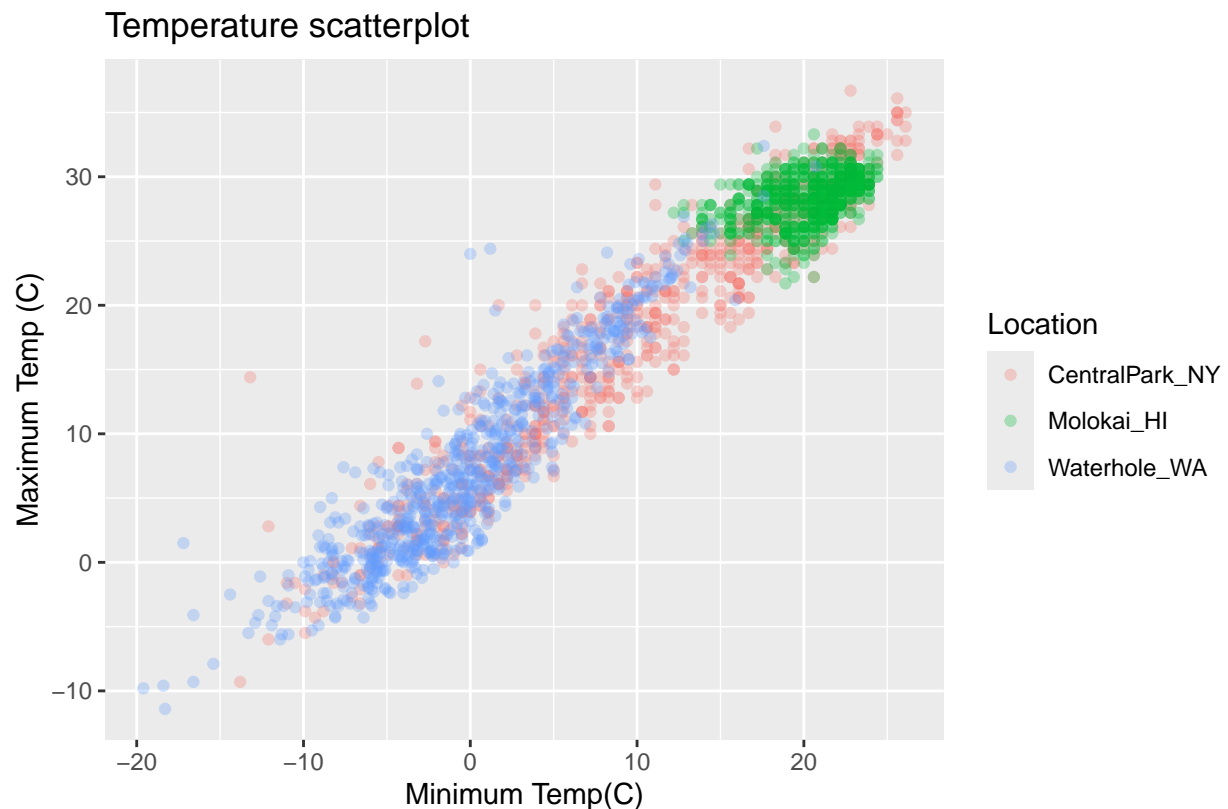

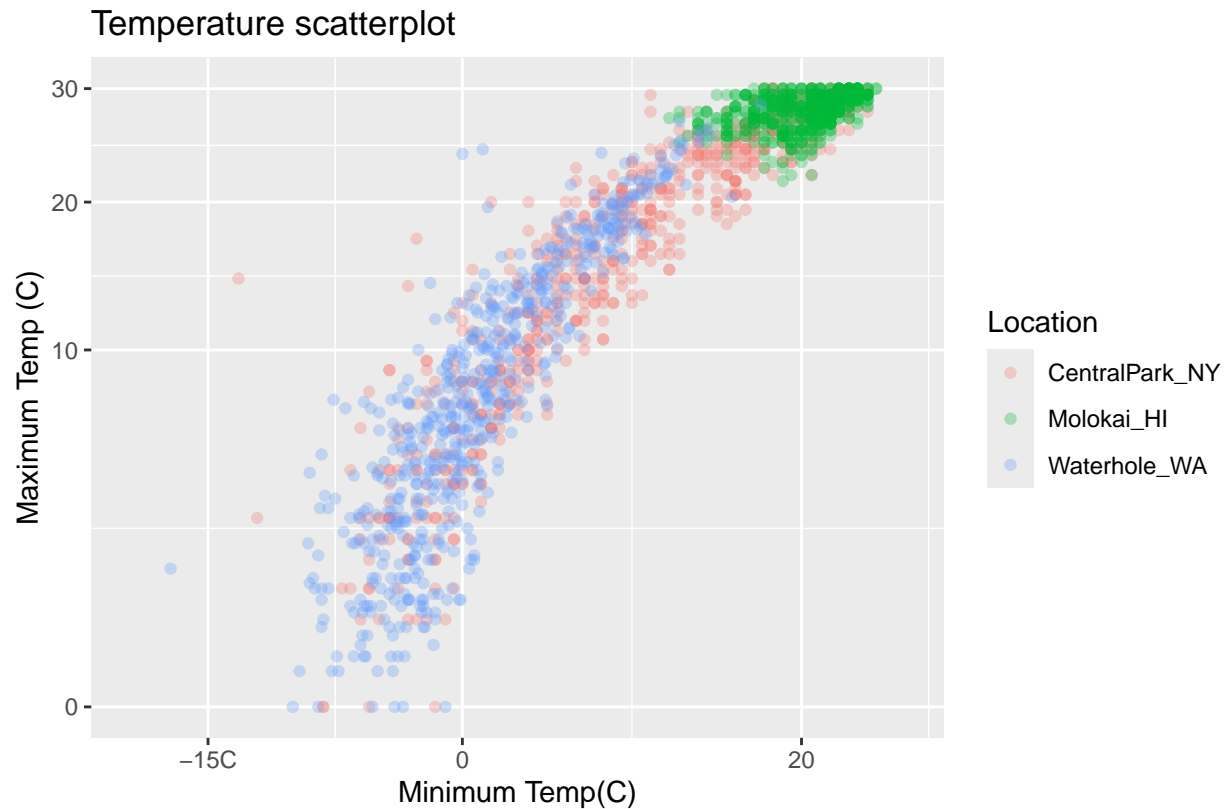Weather data taken from rnoaa package for three stations.

Scales – start with x and y and then do color.

```r
weather_df |>
  ggplot(aes(x = tmin, y = tmax, color = name)) +
  geom_point(alpha = .3) +
  labs(
    title = "Temperature scatterplot",
    x = "Minimum Temp(C)",
    y = "Maximum Temp (C)",
    color = "Location",
    caption = "Weather data taken from rnoaa package for three stations."
  ) +
  scale_x_continuous(
    breaks = c(-15, 0, 20),
    labels = c("-15C", "0","20")
  ) +
  scale_y_continuous(
    limits = c(0, 30),
    transform = "sqrt"
  )
```

```
## Warning in transformation$transform(x): NaNs produced
```

```
## Warning in scale_y_continuous(limits = c(0, 30), transform = "sqrt"): sqrt
## transformation introduced infinite values.
```

```
## Warning: Removed 302 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Weather data taken from rnoaa package for three stations.

Look at color:

```
weather_df |>
  ggplot(aes(x = tmin, y = tmax, color = name)) +
  geom_point(alpha = .3) +
  labs(
    title = "Temperature scatterplot",
    x = "Minimum Temp(C)",
    y = "Maximum Temp (C)",
    color = "Location",
    caption = "Weather data taken from rnoaa package for three stations."
  ) +
  scale_color_hue(h =c(100, 400))+
  viridis::scale_color_viridis(discrete = TRUE)
```
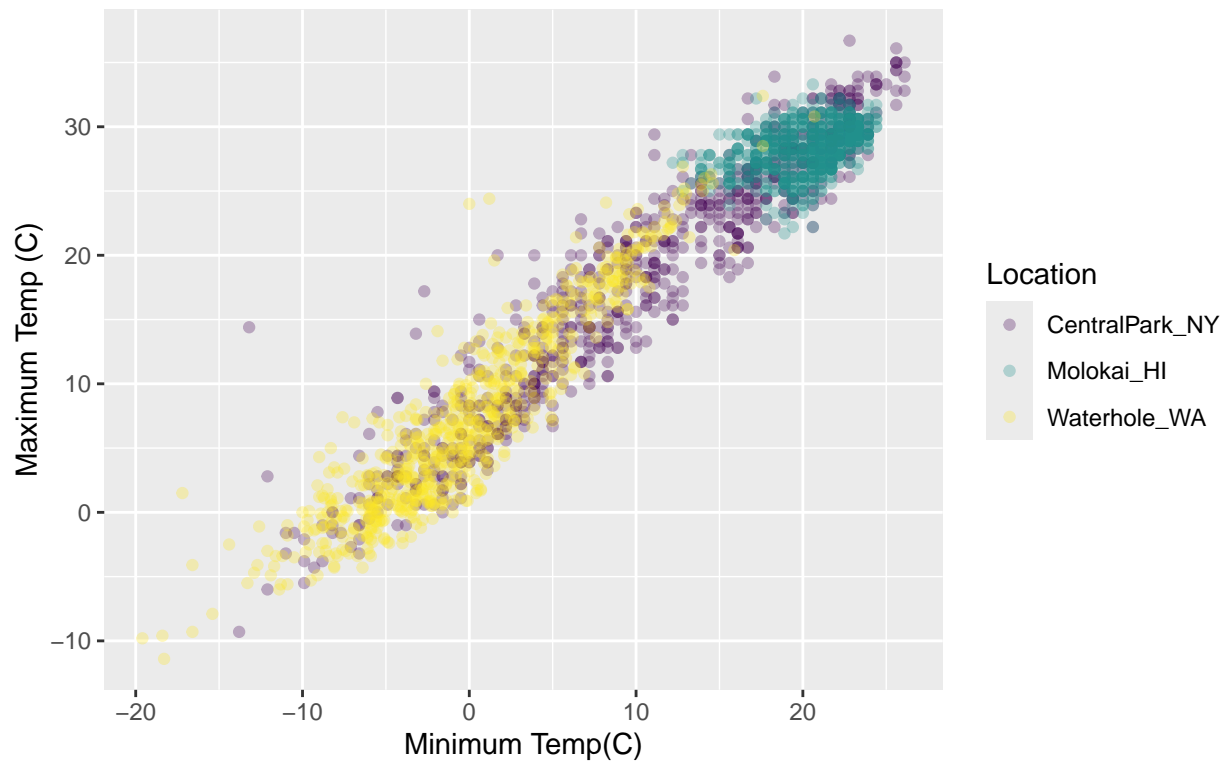
```
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Temperature scatterplot

Weather data taken from rnoaa package for three stations.
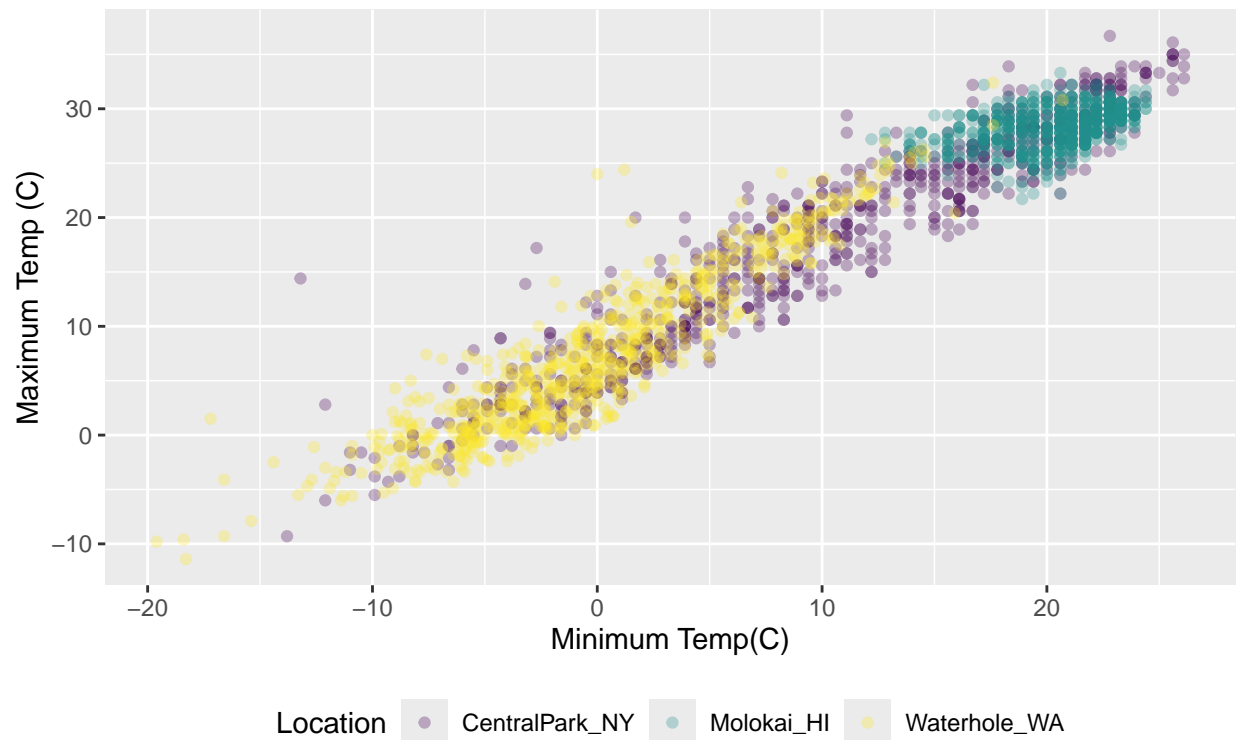
Themes

```
ggp_scatterplot =
  weather_df |>
  ggplot(aes(x = tmin, y = tmax, color = name)) +
  geom_point(alpha = .3) +
  labs(
    title = "Temperature scatterplot",
    x = "Minimum Temp(C)",
    y = "Maximum Temp (C)",
    color = "Location",
    caption = "Weather data taken from rnoaa package for three stations."
  ) +
  scale_color_hue(h =c(100, 400))+
  viridis::scale_color_viridis(discrete = TRUE)
```

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

```
ggp_scatterplot +
  theme(legend.position = "bottom")
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## ('geom_point()').
```
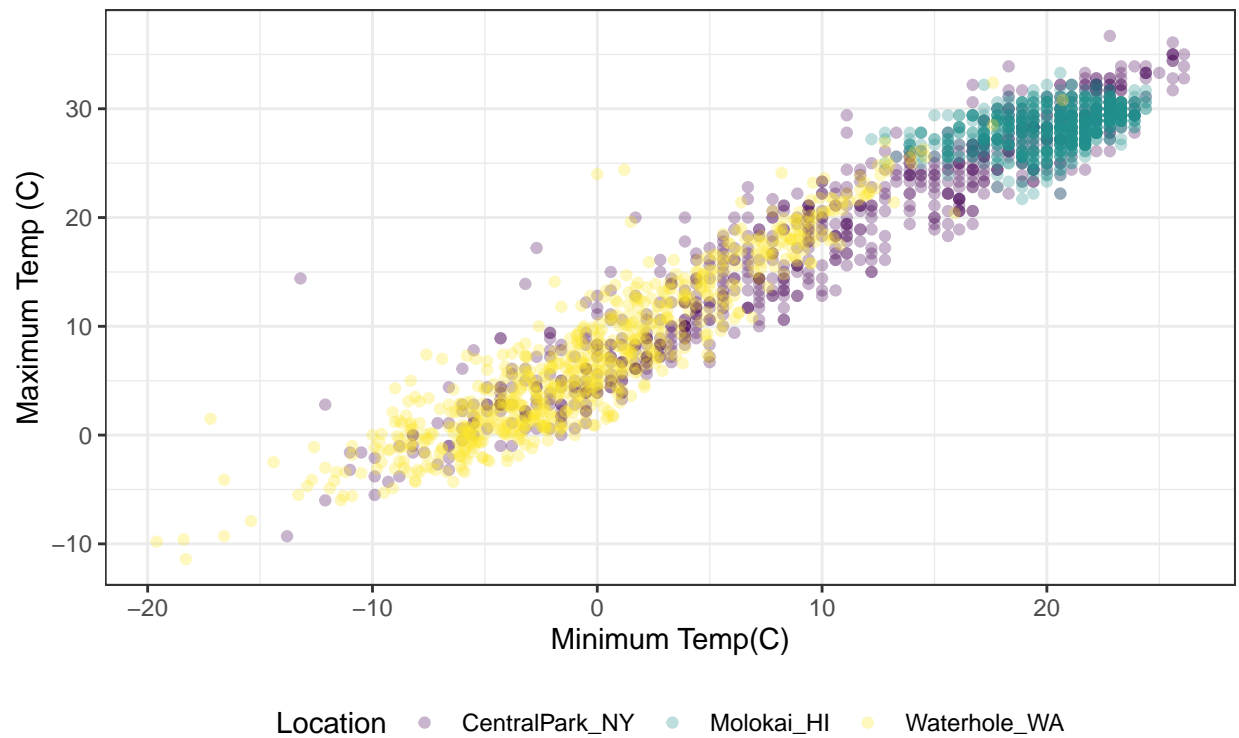
Temperature scatterplot

Weather data taken from rnoaa package for three stations.

```
ggp_scatterplot +
  theme_bw() +
  theme(legend.position = "bottom")
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

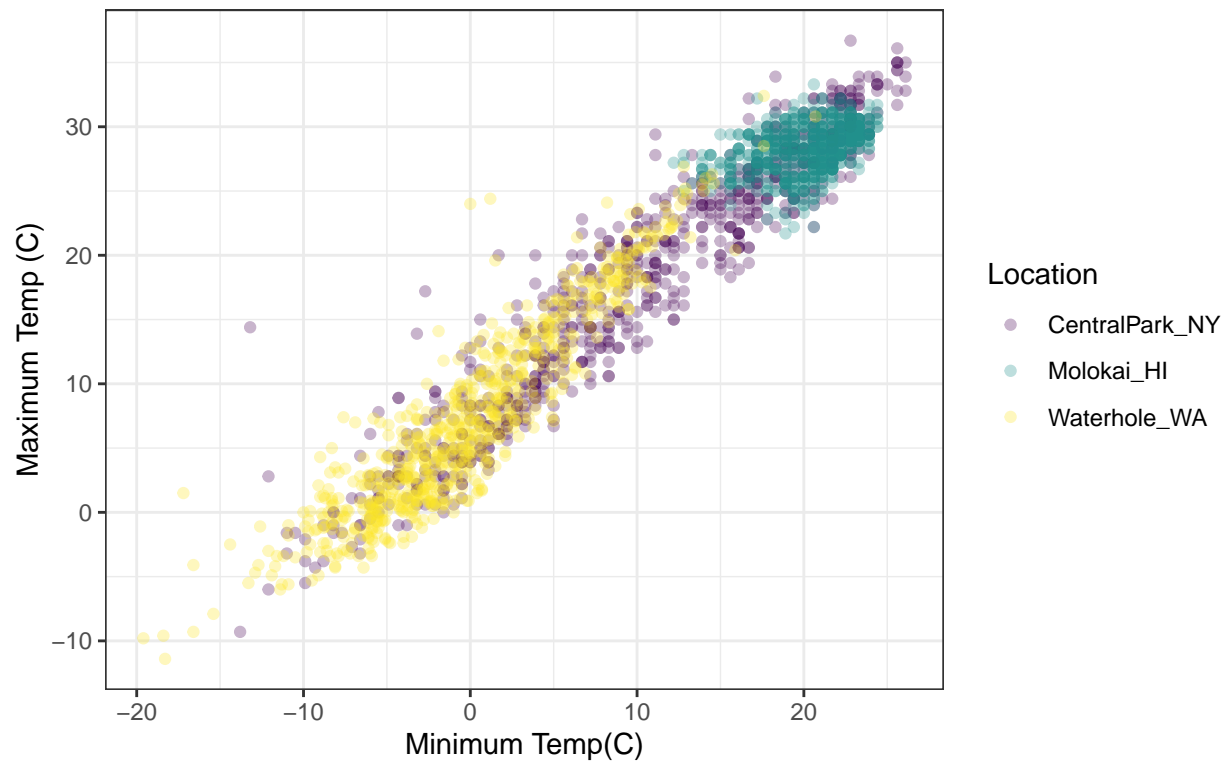## Temperature scatterplot

Weather data taken from rnoaa package for three stations.

Order matters . . .

```
ggp_scatterplot +
  theme(legend.position = "bottom") +
  theme_bw()
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## ('geom_point()').
```
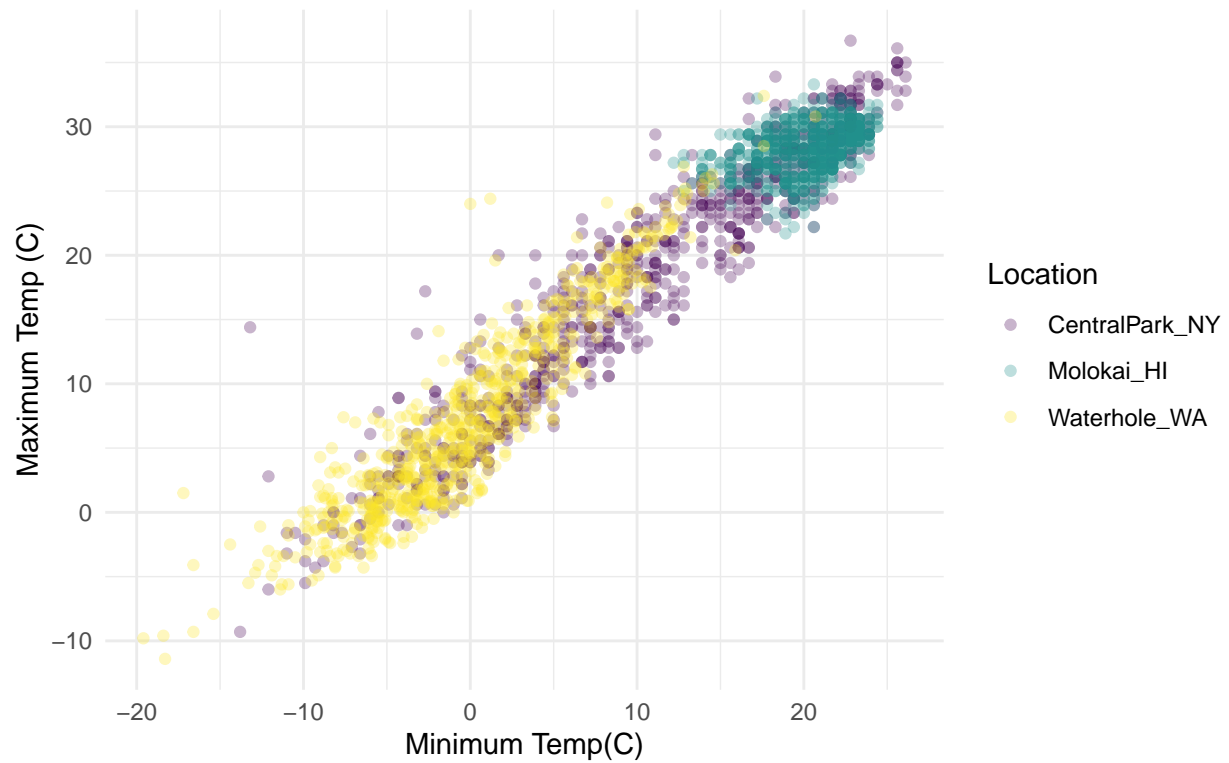
Temperature scatterplot

Weather data taken from rnoaa package for three stations.

```
ggp_scatterplot +
  theme(legend.position = "bottom") +
  theme_minimal()
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## ('geom_point()').
```
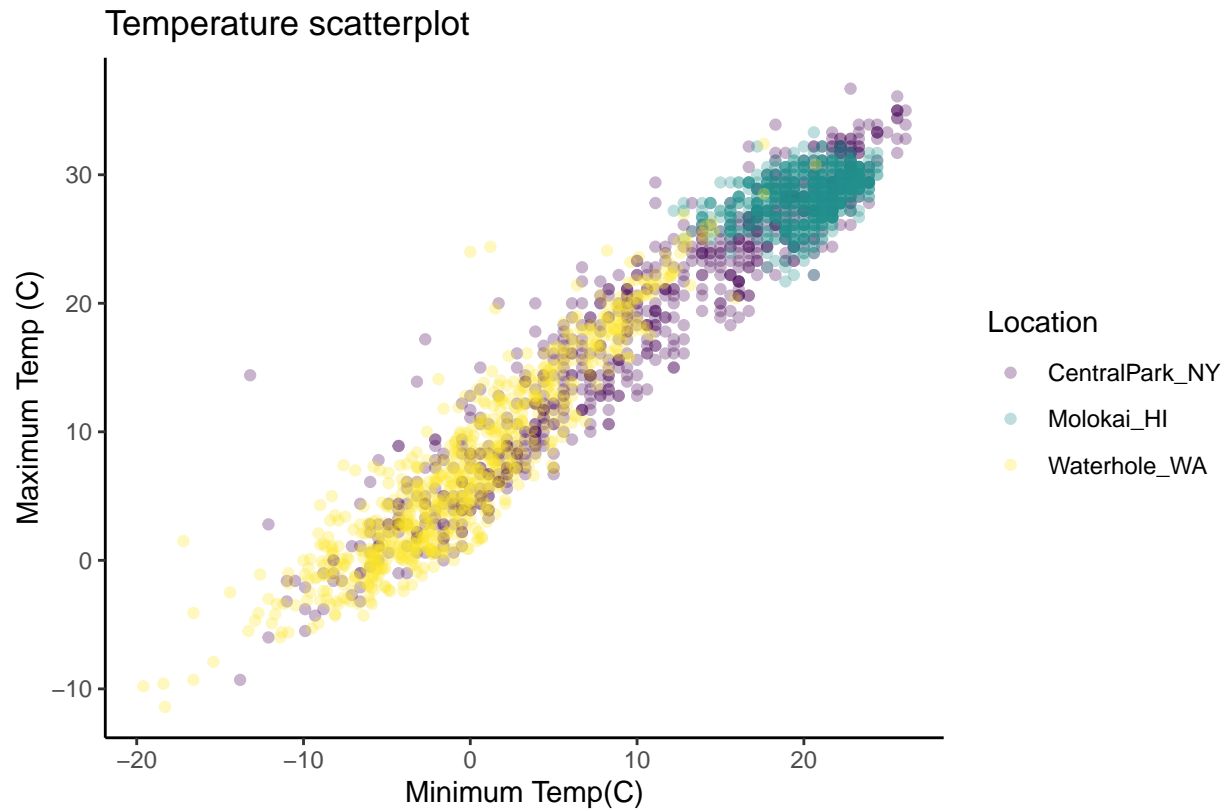
Weather data taken from rnoaa package for three stations.

```r
ggp_scatterplot +
  theme(legend.position = "bottom") +
  theme_classic()
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).
```
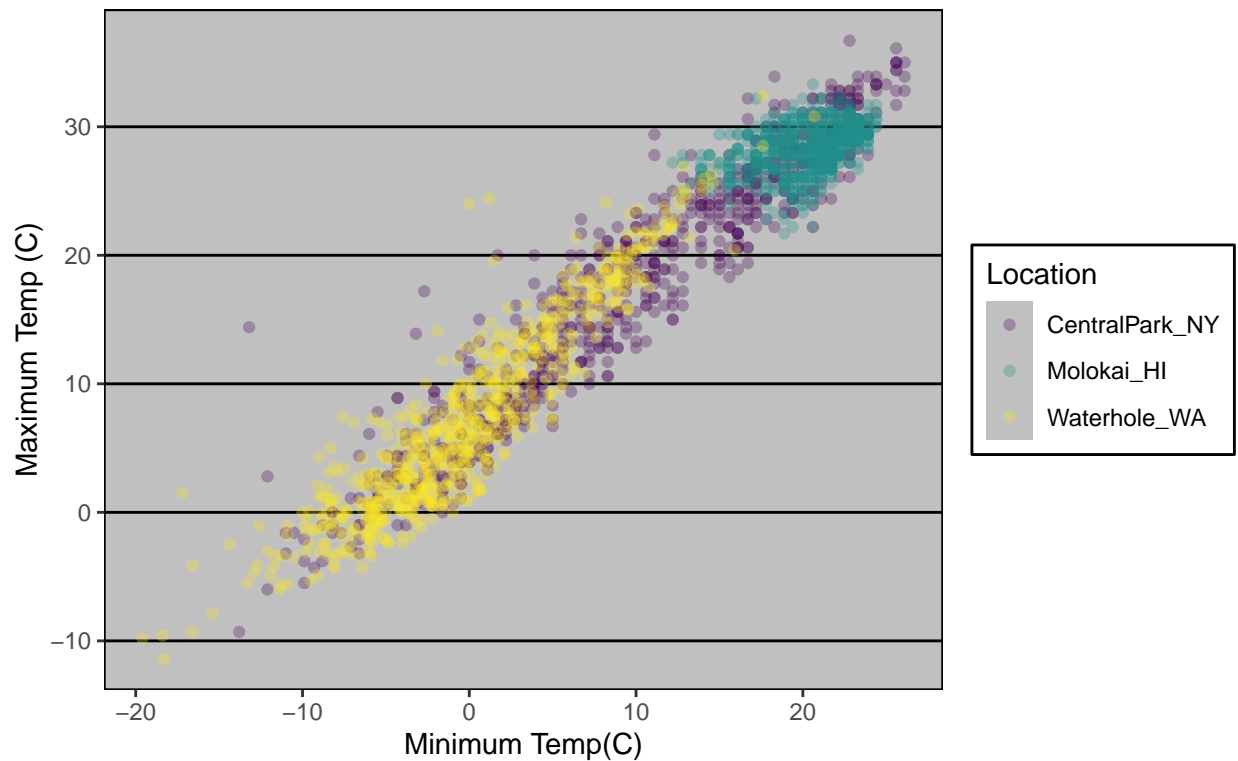
Temperature scatterplot

Weather data taken from rnoaa package for three stations.

```
ggp_scatterplot +
  theme(legend.position = "bottom") +
  ggthemes::theme_excel()
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## ('geom_point()').
```

Temperature scatterplot

Weather data taken from rnoaa package for three stations.

Learning Assessment

```r
weather_df |>
  ggplot(aes(x = date, y = tmin, color = name, size = prcp)) +
  geom_point(alpha = .3) +
  geom_smooth(se = FALSE) +
  viridis::scale_color_viridis(discrete = TRUE) +
  labs(
    title = "Seasonal variation in Min Temp",
    x = "Date",
    y = "Minimum Temp(C)",
    color = "Location",
    size = "Precipitation",
    caption = "Weather data taken from rnoaa package for three stations."
  ) +
  theme_minimal()
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```

```
## Warning: Removed 17 rows containing non-finite outside the scale range
## ('stat_smooth()').
```

```
## Warning: The following aesthetics were dropped during statistical transformation: size.
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a 'group' aesthetic or to convert a numerical
##   variable into a factor?
```

```
## Warning: Removed 19 rows containing missing values or values outside the scale range
## ('geom_point()').
```
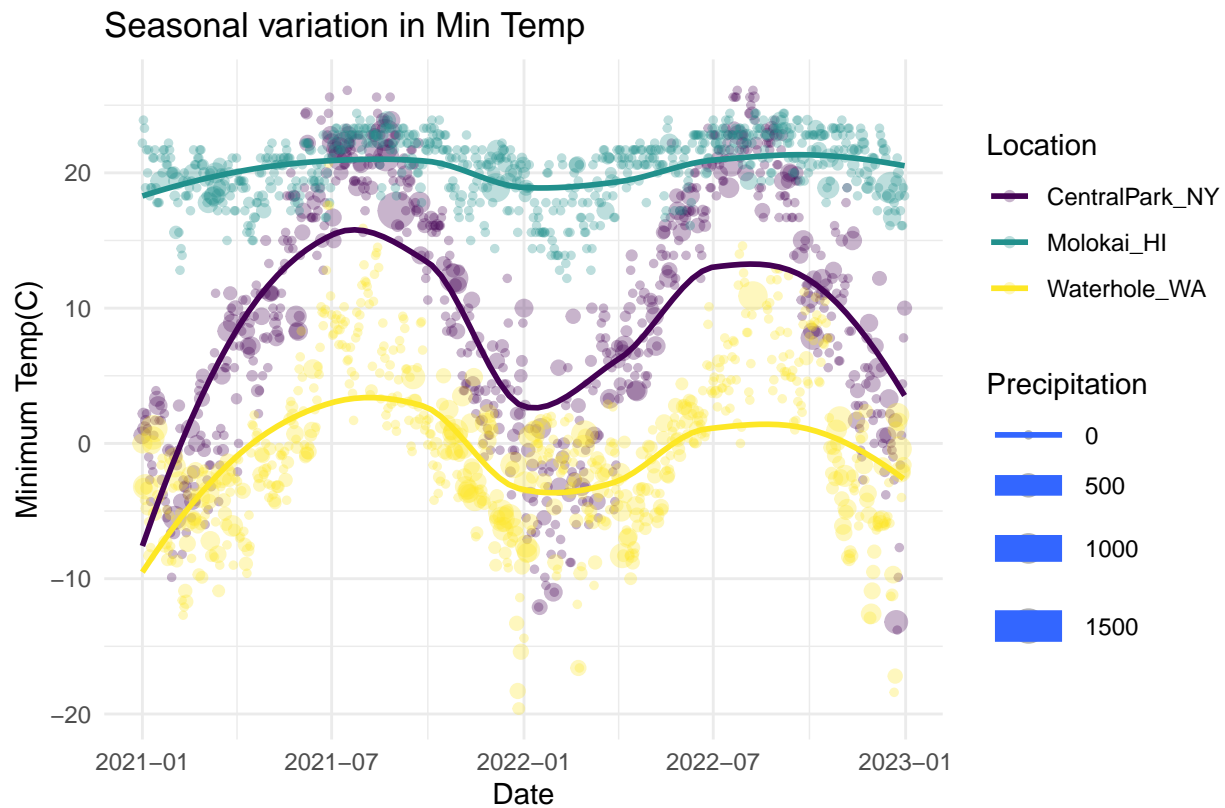

Weather data taken from rnoaa package for three stations.

```
theme(legend.position = "bottom")
```

```
## List of 1
##  $ legend.position: chr "bottom"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

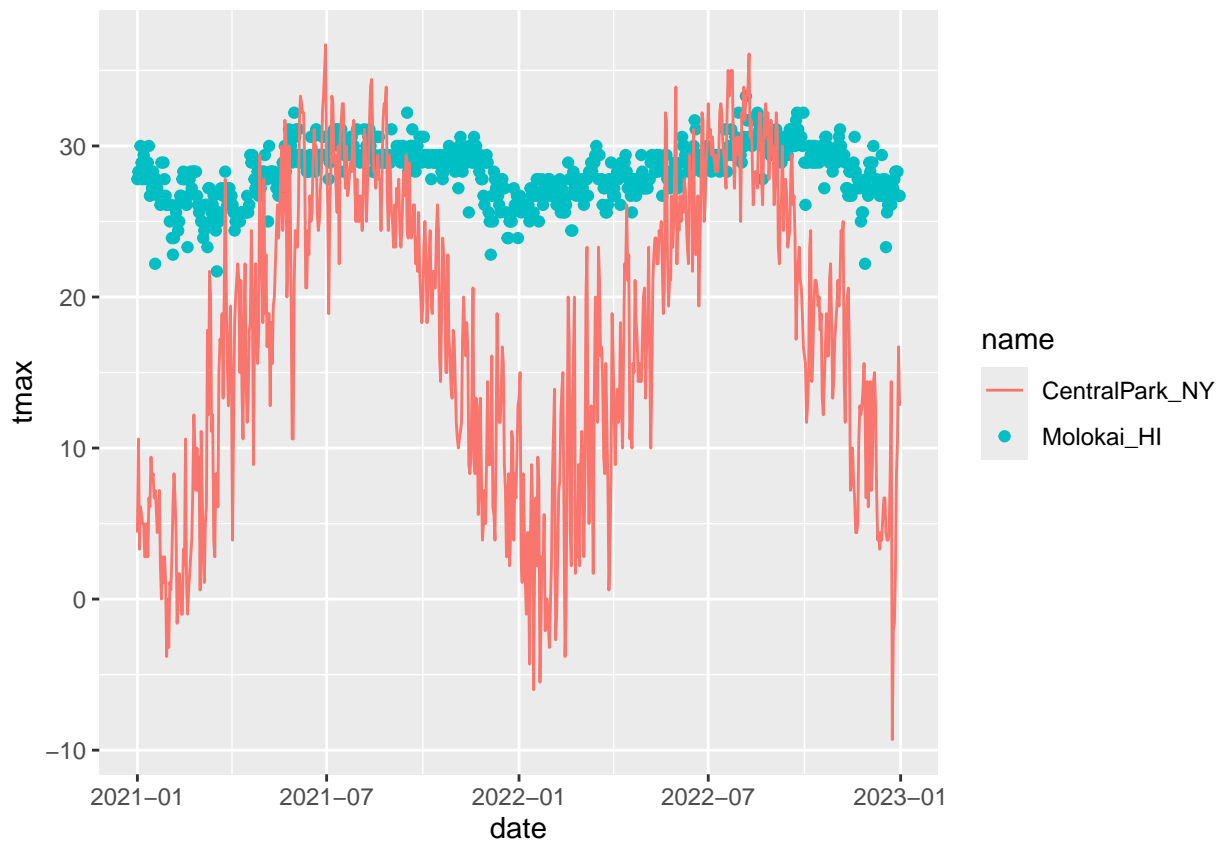Extra bonus stuff in ggplot

Use different datasets in different geoms

```
central_park_df =
  weather_df |>
  filter(name == "CentralPark_NY")

molokai_df =
  weather_df |>
  filter(name == "Molokai_HI")

molokai_df |>
  ggplot(aes(x = date, y = tmax, color = name)) +
  geom_point() +
  geom_line(data = central_park_df)
```
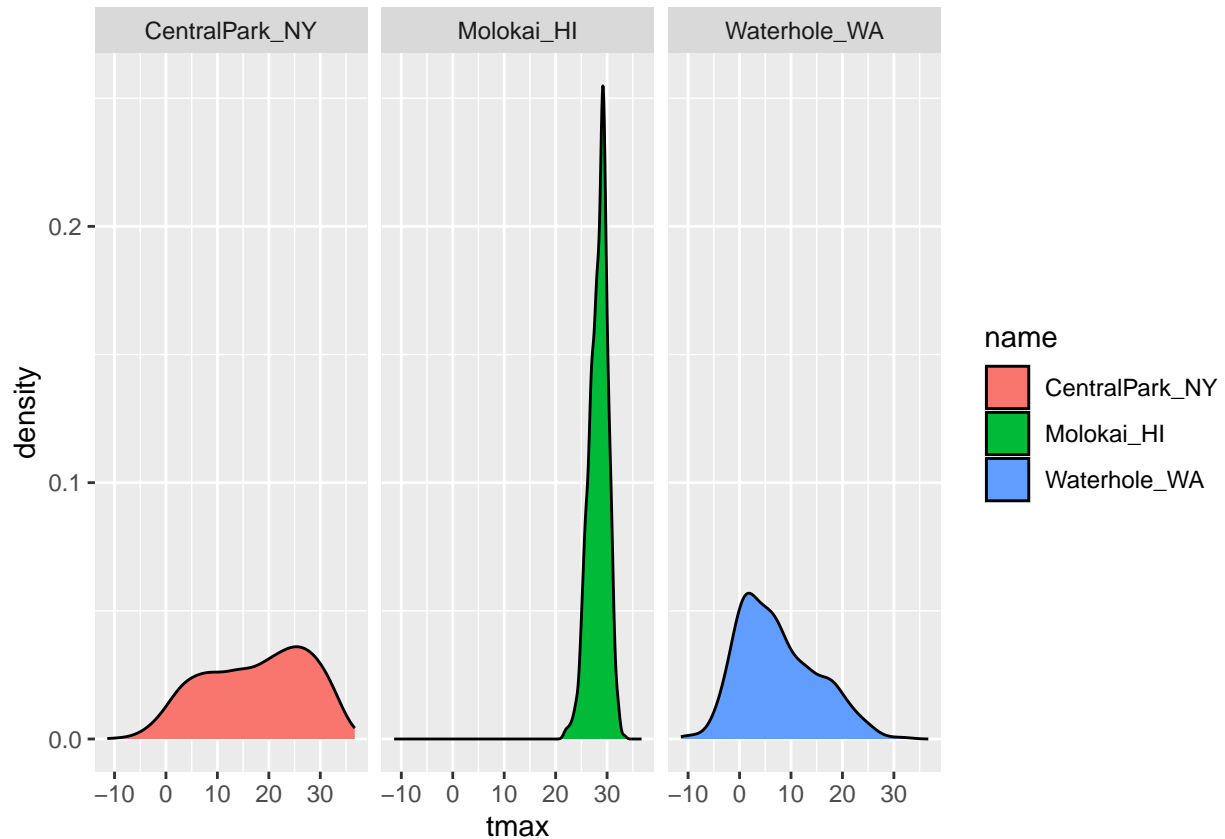
```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
```


Multiple panels

```
weather_df |>
  ggplot(aes(x = tmax, fill = name)) +
  geom_density() +
  facet_grid(. ~ name)
```

```
## Warning: Removed 17 rows containing non-finite outside the scale range
## ('stat_density()').
```



```
ggp_tmax_tmin =
  weather_df |>
  ggplot(aes(x = tmin, y = tmax, color = name)) +
  geom_point(alpha = 0.3)

ggp_tmax_density =
  weather_df |>
  ggplot(aes(x = tmax, fill = name)) +
  geom_density(alpha = 0.3)

ggp_tmax_date =
  weather_df |>
  ggplot(aes(x = date, y = tmax, color = name)) +
  geom_point() +
  geom_smooth(se = FALSE)

(ggp_tmax_tmin + ggp_tmax_density) / ggp_tmax_date
```
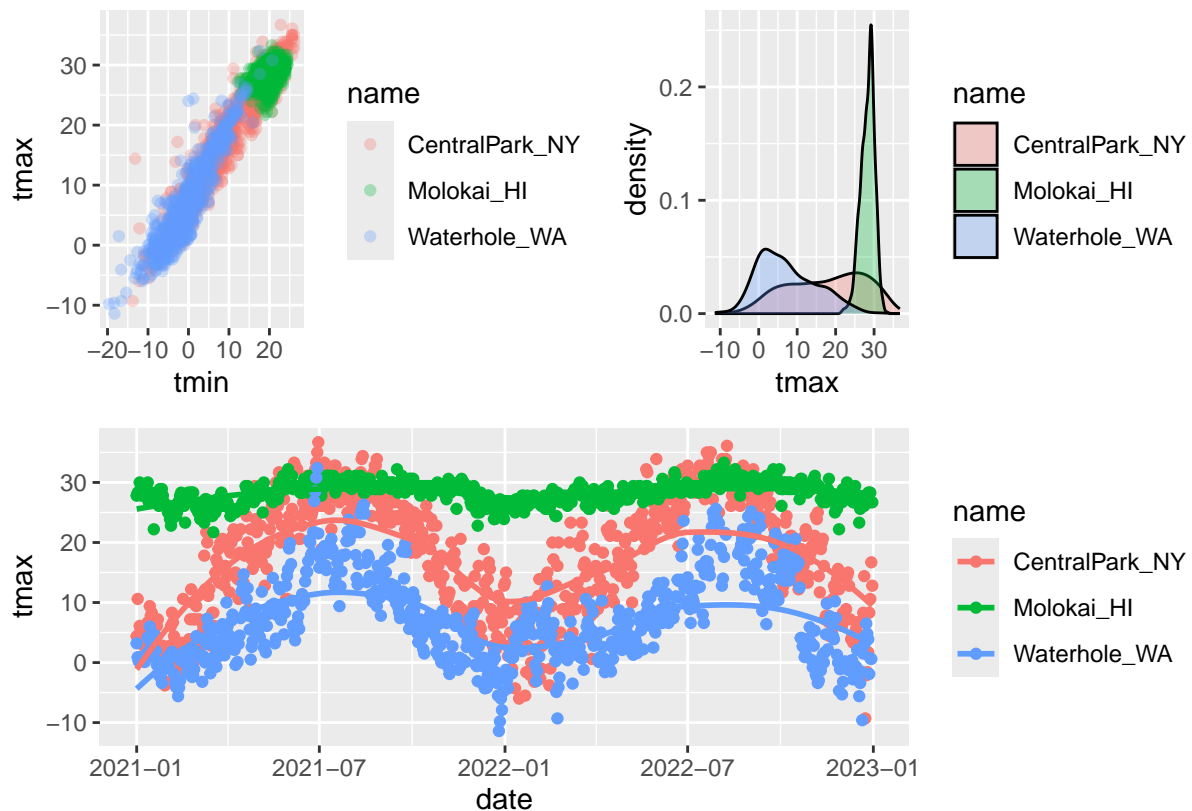
```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## ('geom_point()').
```

```
## Warning: Removed 17 rows containing non-finite outside the scale range
## ('stat_density()').
```

```
## ‘geom_smooth()‘ using method = ’loess’ and formula = ’y ~ x’
```

```
## Warning: Removed 17 rows containing non-finite outside the scale range
## (‘stat_smooth()‘).
```
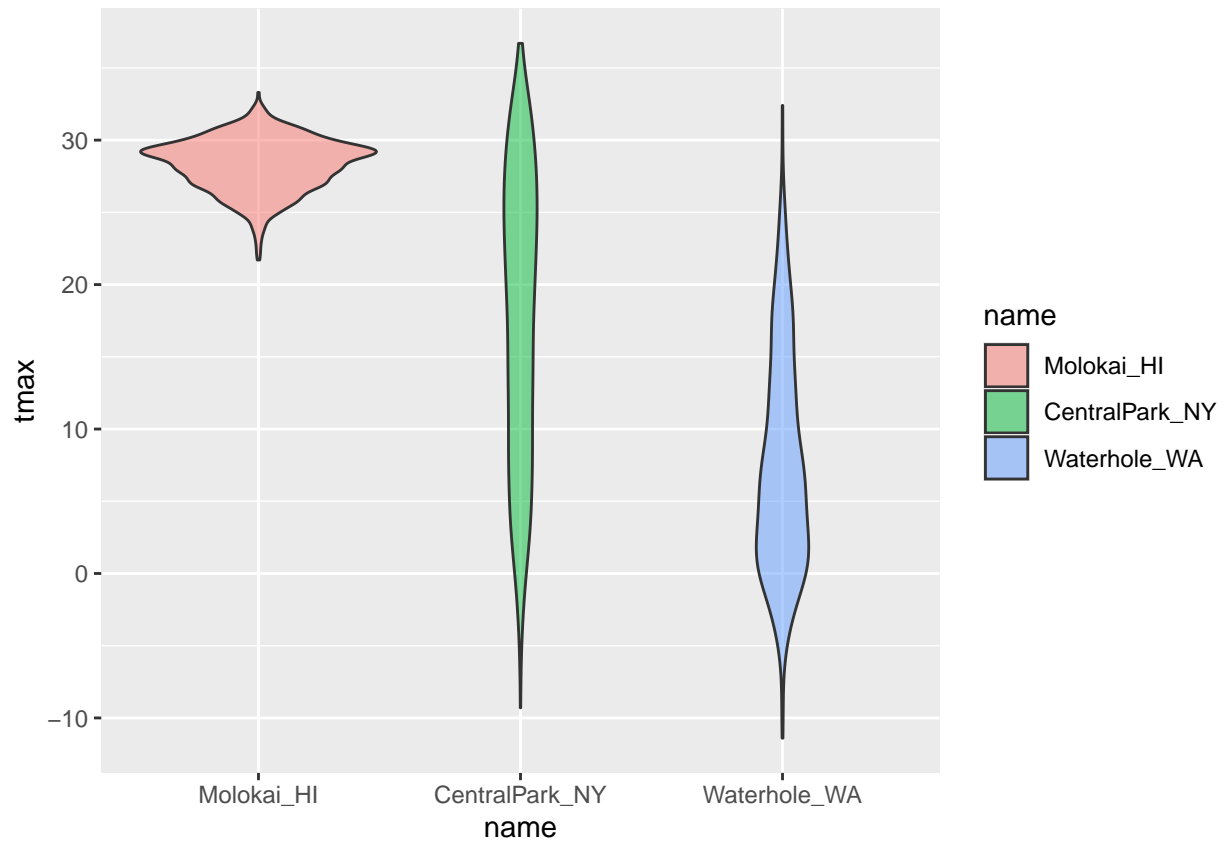
```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## (‘geom_point()‘).
```


Data manipulation

```r
weather_df |>
  mutate(name = fct_relevel(name, c("Molokai_HI", "CentralPark_NY", "Waterhole_WA"))) |>
  ggplot (aes(x = name, y = tmax, fill = name)) +
  geom_violin(alpha = .5)
```

```
## Warning: Removed 17 rows containing non-finite outside the scale range
## (‘stat_ydensity()‘).
```
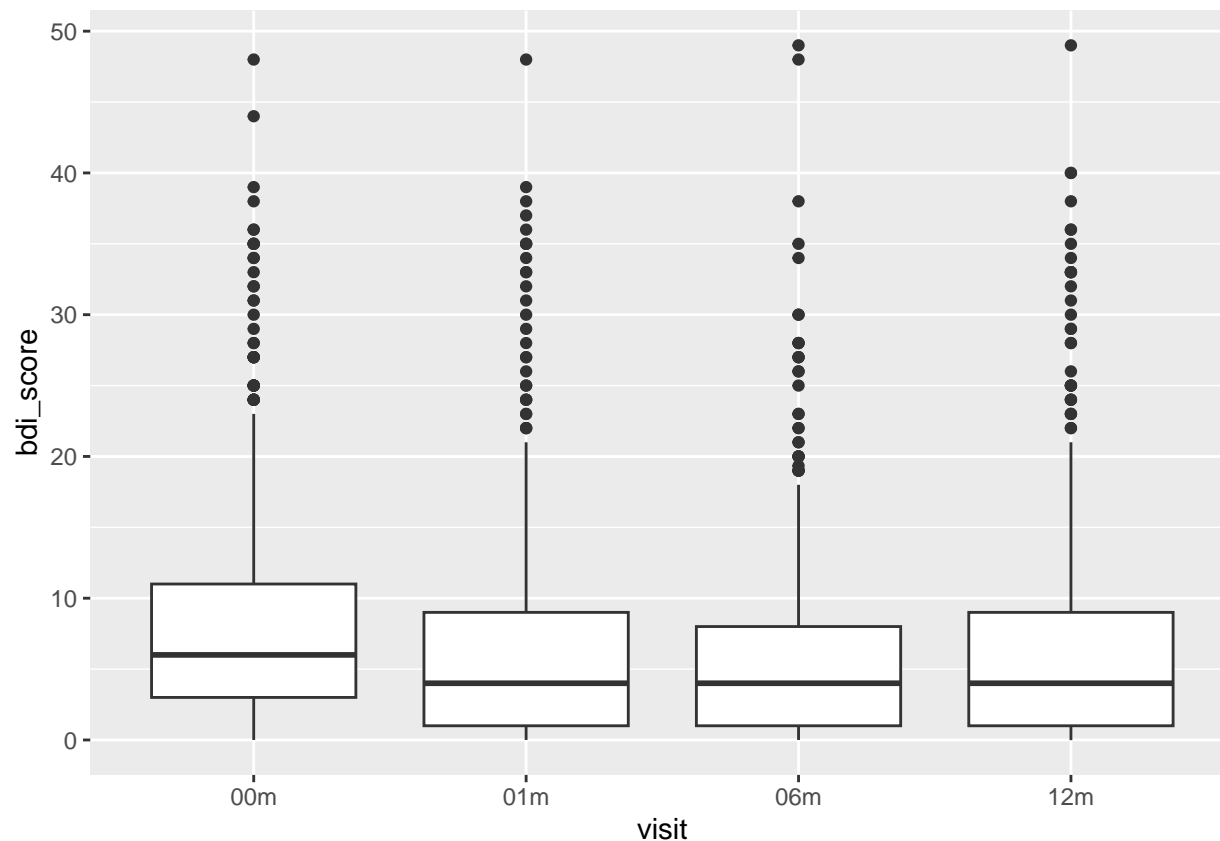
PULSE data next.

```r
pulse_df =
  read_sas("./data/public_pulse_data.sas7bdat") |>
  janitor::clean_names() |>
  pivot_longer(
    cols = bdi_score_bl:bdi_score_12m,
    names_to = "visit",
    values_to = "bdi_score",
    names_prefix = "bdi_score_"
    ) |>
  mutate(visit = ifelse(visit == "bl", "00m", visit))

pulse_df |>
  ggplot(aes(x = visit, y = bdi_score)) +
  geom_boxplot()
```

```
## Warning: Removed 879 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

Make an FAS plot

(Make a data plot with day 7 data for control, low and medium for when the ears unfold, pivot, walk and eyes open.)

```
pups_df =
  read_csv("./data/FAS_pups.csv", na = c("NA", ".", "")) |>
  janitor::clean_names() |>
  pivot_longer(
    cols = c(pd_ears:pd_walk),
    names_to = "outcome",
    values_to = "pn_day",
    names_prefix = "pd_"
  )
```

```
## Rows: 313 Columns: 6
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr (1): Litter Number
## dbl (5): Sex, PD ears, PD eyes, PD pivot, PD walk
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
litters_df =
  read_csv("./data/FAS_litters.csv", na = c("NA", ".", "")) |>
```

```
  janitor::clean_names() |>
  separate(group, into = c("dose", "tx_day"), 3)
```

```
## Rows: 49 Columns: 8
## -- Column specification -------------------------------------------------
## Delimiter: ","
## chr (2): Group, Litter Number
## dbl (6): GD0 weight, GD18 weight, GD of Birth, Pups born alive, Pups dead @ ...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
fas_df =
  left_join(pups_df, litters_df, by = "litter_number")

fas_df |>
  drop_na(tx_day) |>
  ggplot(aes(x = dose, y = pn_day)) +
  geom_boxplot() +
  facet_grid(tx_day ~ outcome)
```

```
## Warning: Removed 42 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```