09.26 data visualization

Guangling Xu

9/26/2019

Set up

```
knitr::opts_chunk$set(
  fig.width = 6,
  fig.asp = .6,
  out.width = "90%"
)
library(tidyverse)
library(ggridges)
```

Use weather_df data

```
weather_df =
  rnoaa::meteo_pull_monitors(c("USW00094728", "USC00519397", "USS0023B17S"),
                      var = c("PRCP", "TMIN", "TMAX"),
                      date_min = "2017-01-01",
                      date_max = "2017-12-31") %>%
  mutate(
    name = recode(id, USW00094728 = "CentralPark_NY",
                      USC00519397 = "Waikiki_HA",
                      USS0023B17S = "Waterhole_WA"),
    tmin = tmin / 10,
    tmax = tmax / 10) %>%
  select(name, id, everything())
```

```
## Registered S3 method overwritten by 'crul':
##   method                 from
##   as.character.form_file httr

## Registered S3 method overwritten by 'hoardr':
##   method           from
##   print.cache_info httr

## file path:          C:\Users\xu499\AppData\Local\rnoaa\rnoaa\Cache/ghcnd/USW00094728.dly

## file last updated:  2019-09-26 10:23:06

## file min/max dates: 1869-01-01 / 2019-09-30

## file path:          C:\Users\xu499\AppData\Local\rnoaa\rnoaa\Cache/ghcnd/USC00519397.dly

## file last updated:  2019-09-26 10:23:20
```

```
## file min/max dates: 1965-01-01 / 2019-09-30

## file path:          C:\Users\xu499\AppData\Local\rnoaa\rnoaa\Cache/ghcnd/USS0023B17S.dly

## file last updated:  2019-09-26 10:23:25

## file min/max dates: 1999-09-01 / 2019-09-30
```

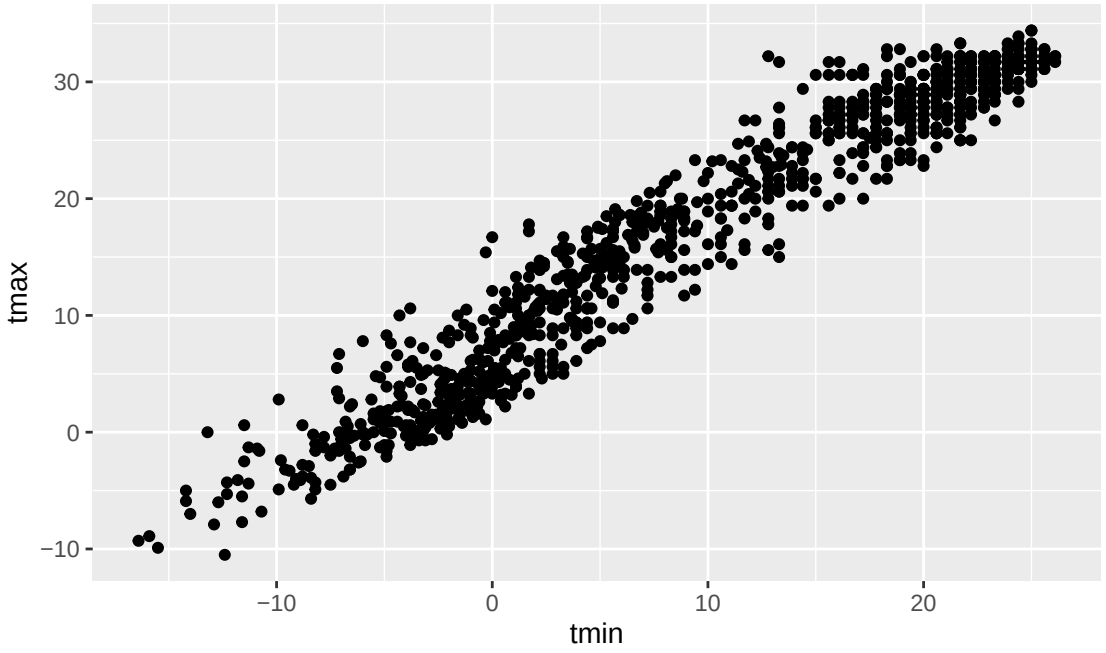
```
weather_df
```

```
## # A tibble: 1,095 x 6
##    name           id          date        prcp  tmax  tmin
##    <chr>          <chr>       <date>     <dbl> <dbl> <dbl>
##  1 CentralPark_NY USW00094728 2017-01-01     0   8.9   4.4
##  2 CentralPark_NY USW00094728 2017-01-02    53   5     2.8
##  3 CentralPark_NY USW00094728 2017-01-03   147   6.1   3.9
##  4 CentralPark_NY USW00094728 2017-01-04     0  11.1   1.1
##  5 CentralPark_NY USW00094728 2017-01-05     0   1.1  -2.7
##  6 CentralPark_NY USW00094728 2017-01-06    13   0.6  -3.8
##  7 CentralPark_NY USW00094728 2017-01-07    81  -3.2  -6.6
##  8 CentralPark_NY USW00094728 2017-01-08     0  -3.8  -8.8
##  9 CentralPark_NY USW00094728 2017-01-09     0  -4.9  -9.9
## 10 CentralPark_NY USW00094728 2017-01-10     0   7.8  -6
## # ... with 1,085 more rows
```

Basic Scatterplot

```
ggplot(weather_df, aes(x= tmin, y = tmax)) +
  geom_point()
```

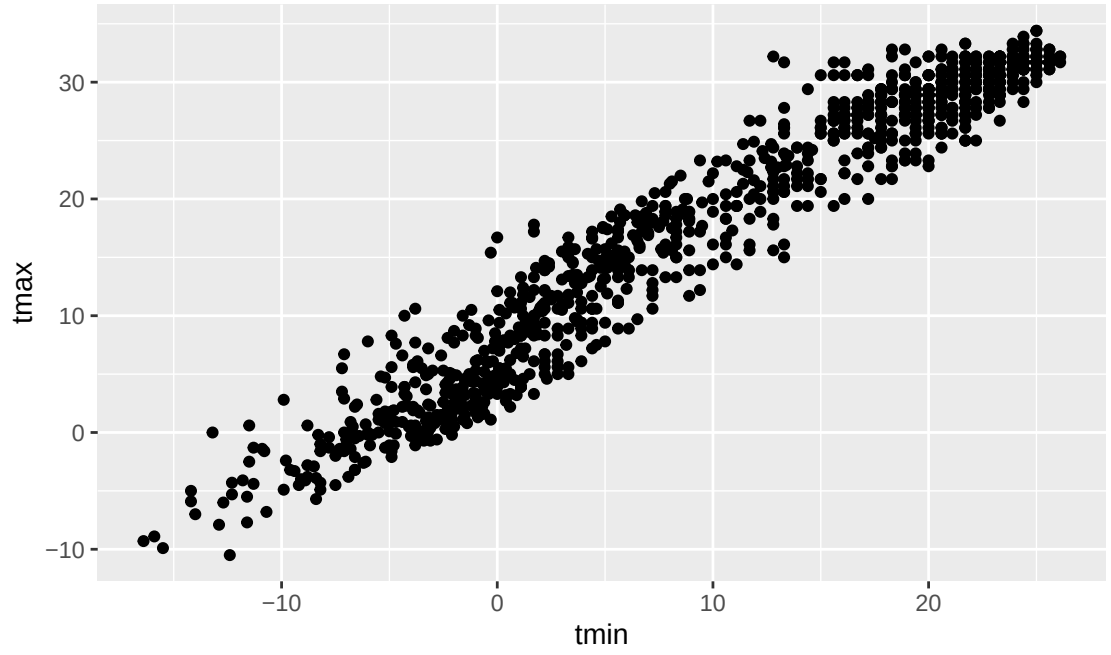
```
## Warning: Removed 15 rows containing missing values (geom_point).
```



alteranate way of making this plot

```
weather_df %>%
  ggplot(aes(x = tmin, y = tmax)) +
  geom_point()
```

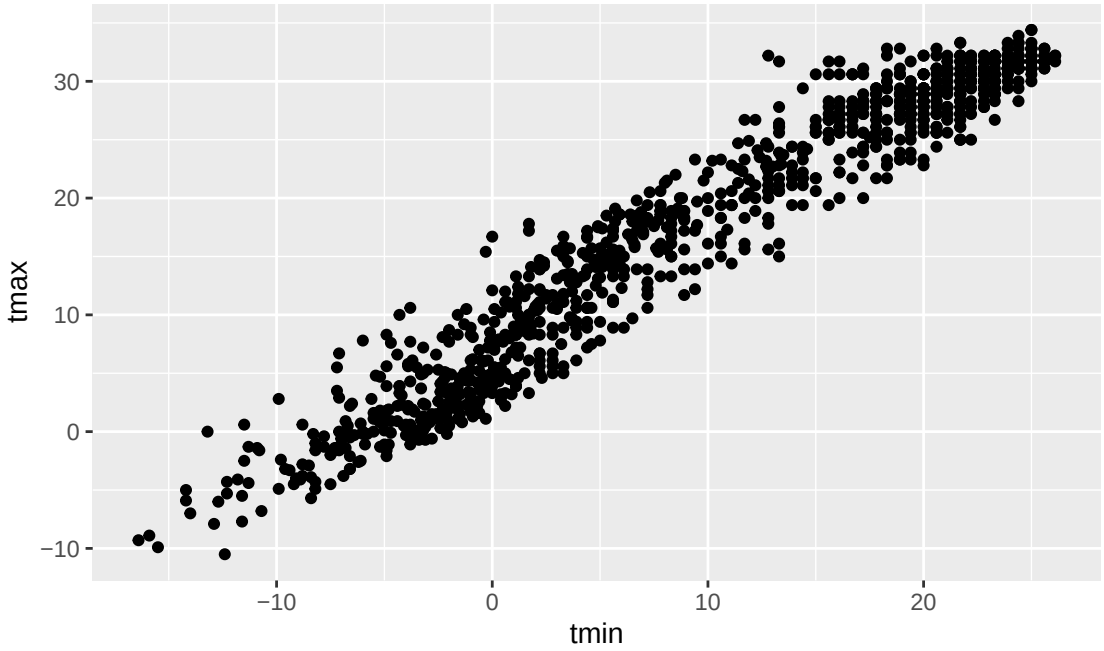
```
## Warning: Removed 15 rows containing missing values (geom_point).
```



saving intial plots

```
scatterplot =
  weather_df %>%
  ggplot(aes(x = tmin, y = tmax)) +
  geom_point()

scatterplot
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```



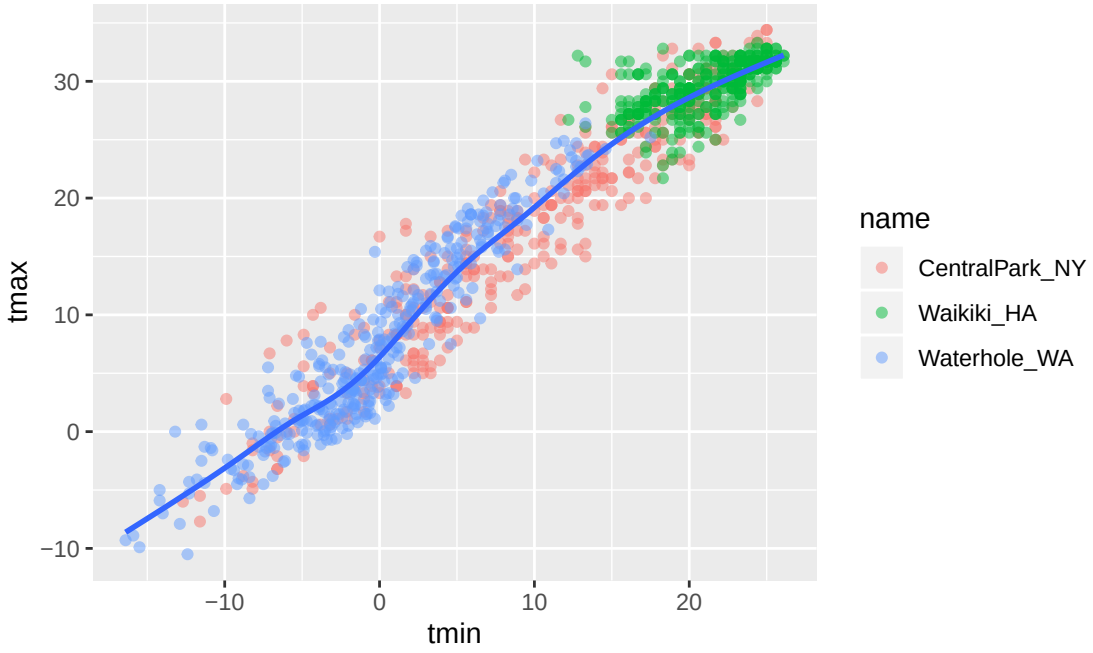
advanced scatterplot

```
ggplot(weather_df, aes(x = tmin, y = tmax)) +
  geom_point(aes(color = name), alpha = .5) +
  geom_smooth(se = FALSE)  ## alpha = transcripency ; se = standard error
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 15 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```



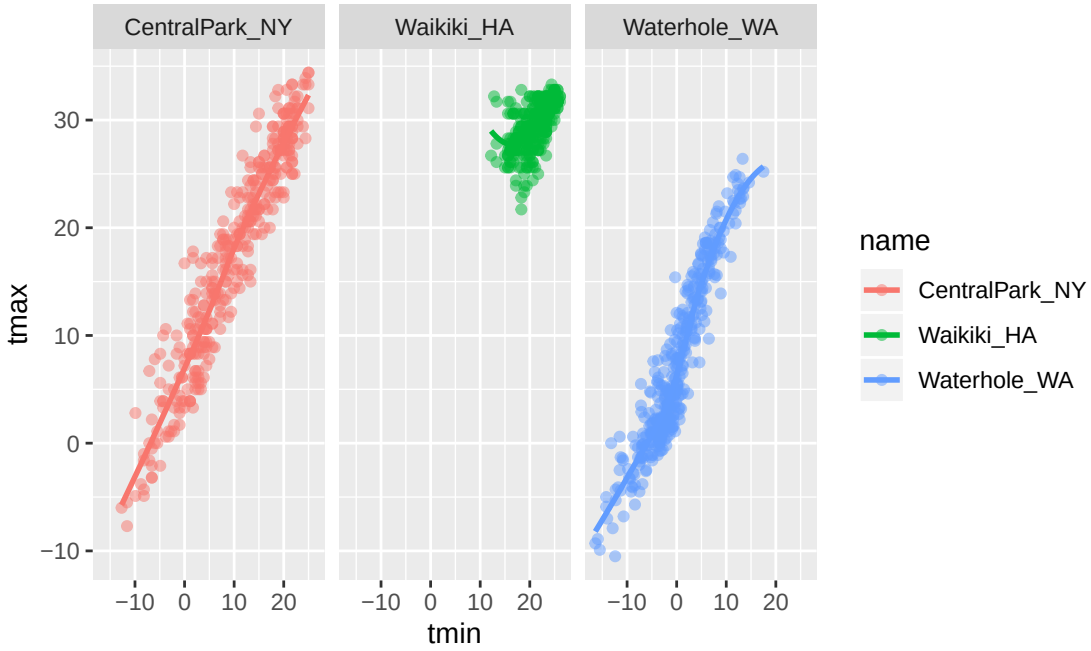
facet

```
ggplot(weather_df, aes(x = tmin, y = tmax, color = name)) +
  geom_point(alpha = .5) +
  geom_smooth(se = FALSE) +
  facet_grid(. ~ name)
```

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

```
## Warning: Removed 15 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```

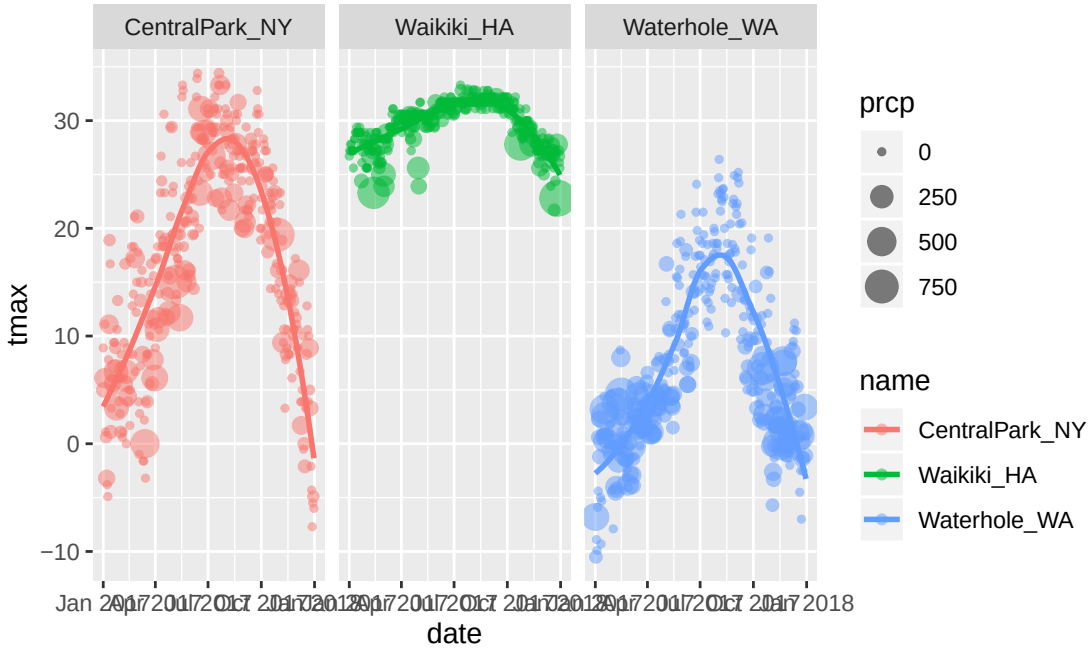



```
ggplot(weather_df, aes(x = date, y = tmax, color = name)) +
  geom_point(aes(size = prcp), alpha = .5) +
  geom_smooth(se = FALSE) +
  facet_grid(. ~ name)
```

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

```
## Warning: Removed 3 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

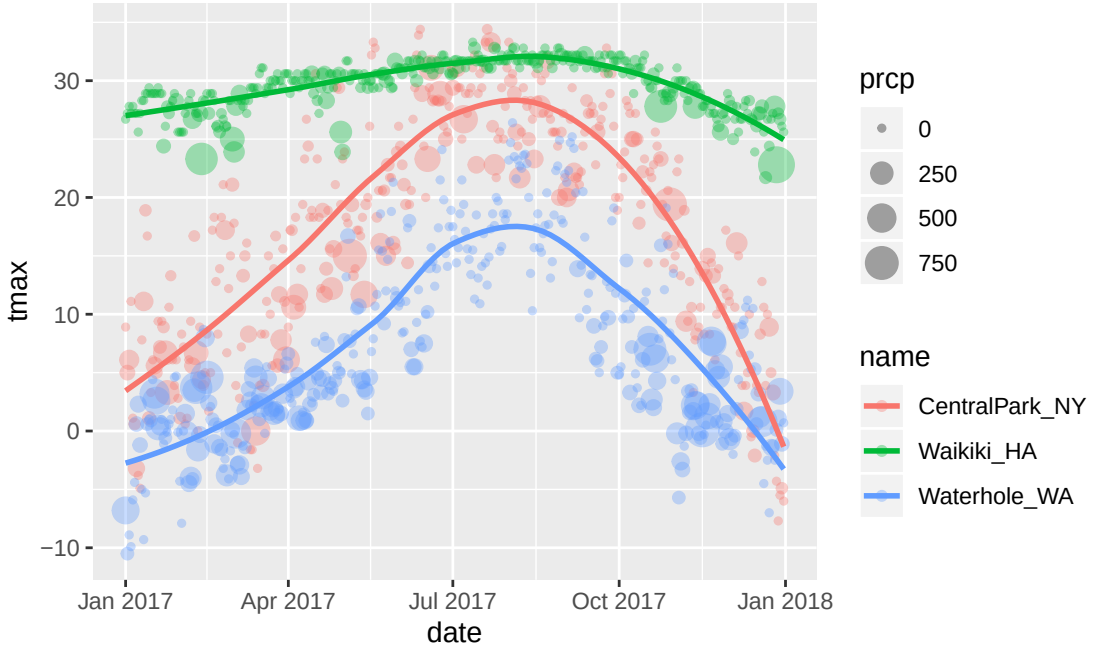


```
weather_df %>%
  ggplot(aes(x = date , y = tmax , color = name)) +
  geom_point( aes(size = prcp) , alpha = .35) +
  geom_smooth(se = FALSE)
```

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

```
## Warning: Removed 3 rows containing non-finite values (stat_smooth).
```

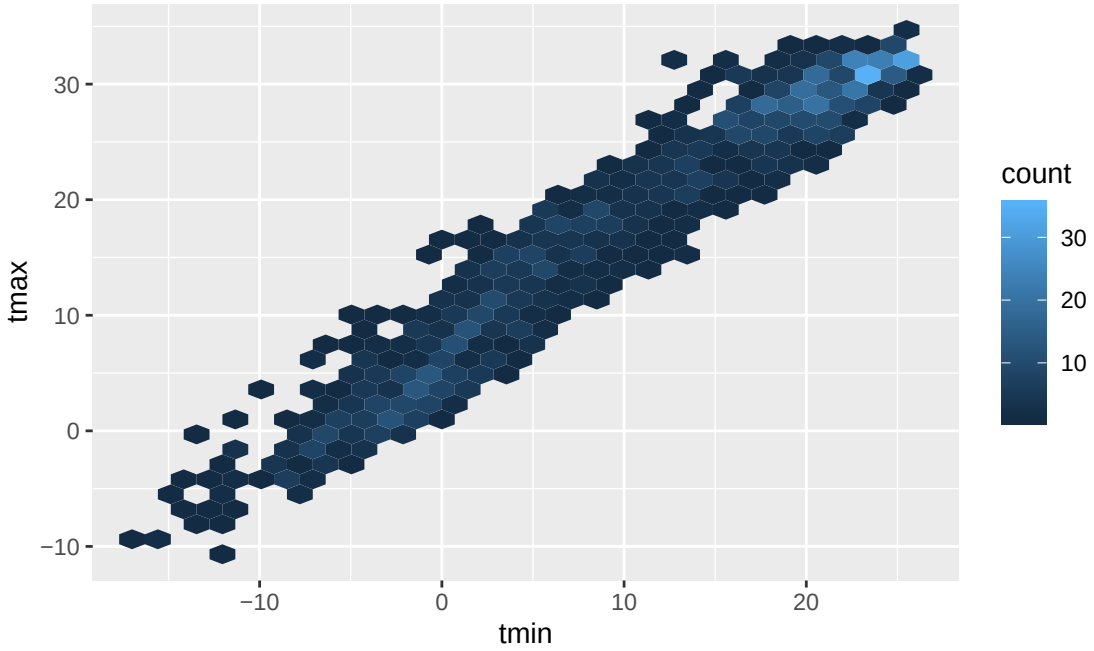
```
## Warning: Removed 3 rows containing missing values (geom_point).
```



2d density

```
weather_df %>%
  ggplot(aes(x = tmin, y = tmax)) +
  geom_hex()
```

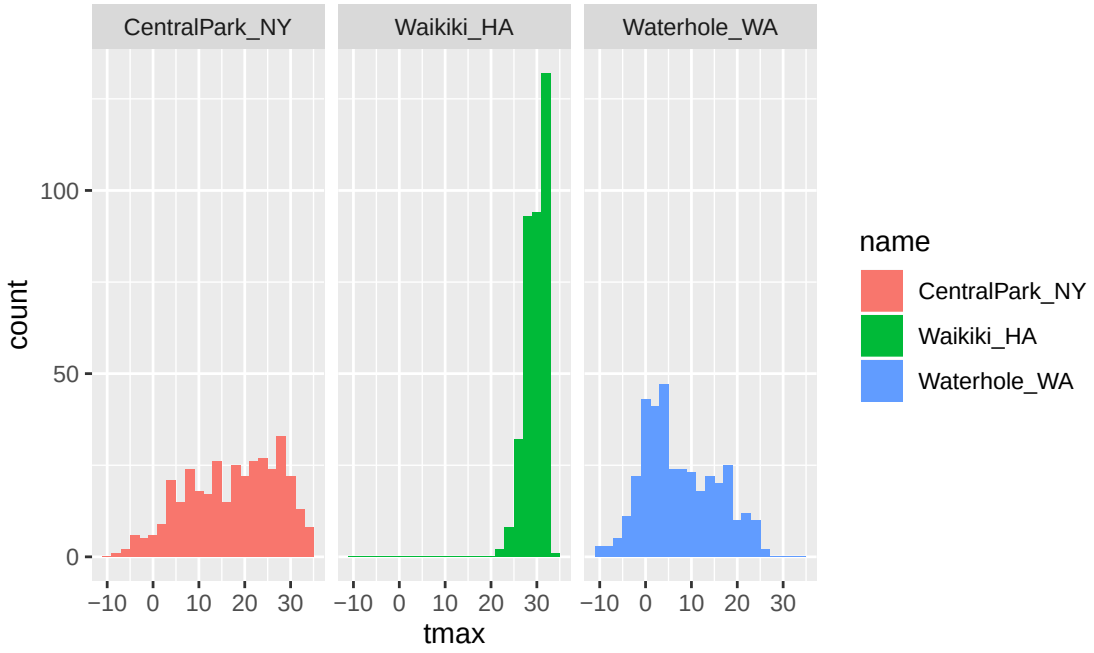
```
## Warning: Removed 15 rows containing non-finite values (stat_binhex).
```



histogram

```
ggplot(weather_df, aes(x = tmax, fill = name)) +
  geom_histogram(position = "dodge", binwidth = 2) +  ## separate bars
  facet_grid(. ~ name)
```

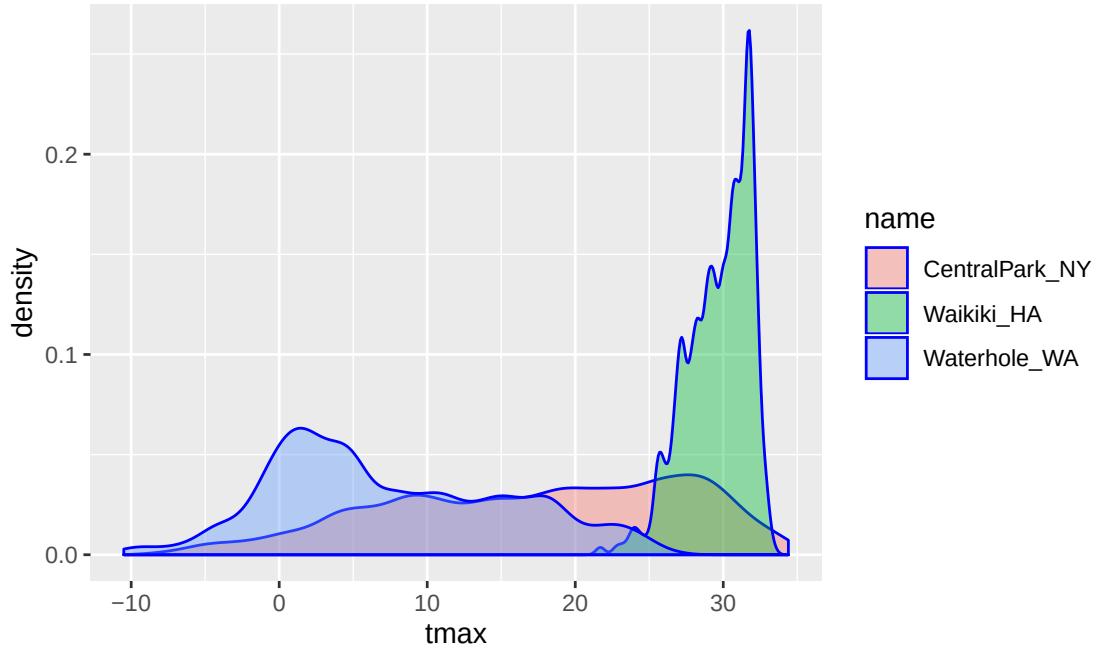
```
## Warning: Removed 3 rows containing non-finite values (stat_bin).
```



density histogram

```
ggplot(weather_df, aes(x = tmax, fill = name)) +
  geom_density(alpha = .4, adjust = .5, color = "blue")
```

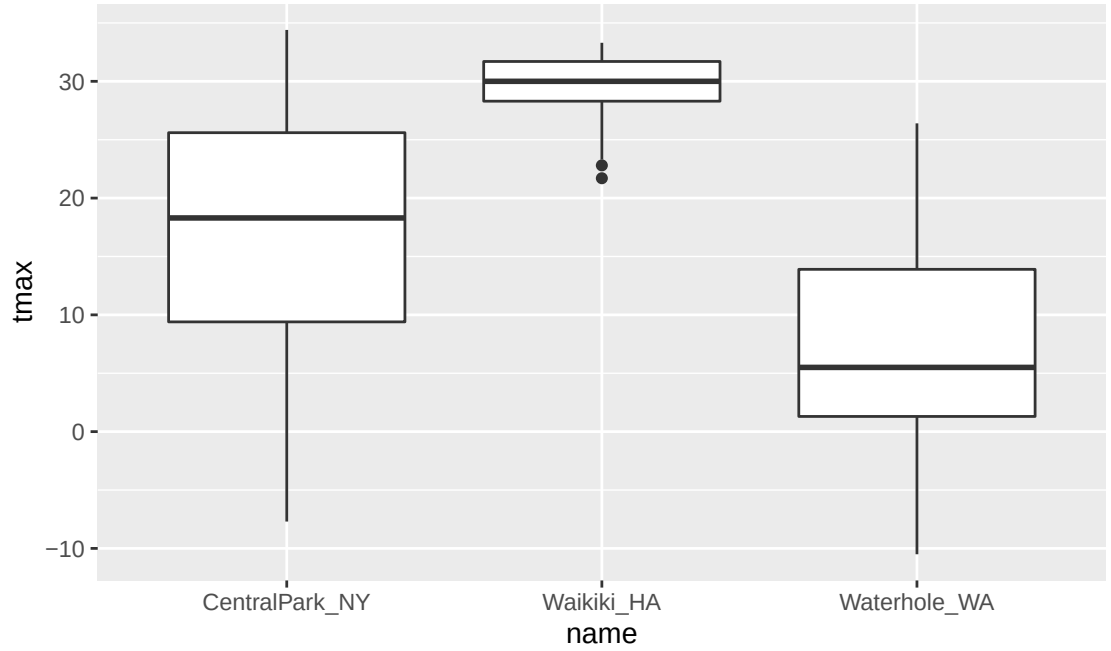
```
## Warning: Removed 3 rows containing non-finite values (stat_density).
```



boxplot vs violin plot

```
ggplot(weather_df, aes(x = name, y = tmax)) + geom_boxplot()
```

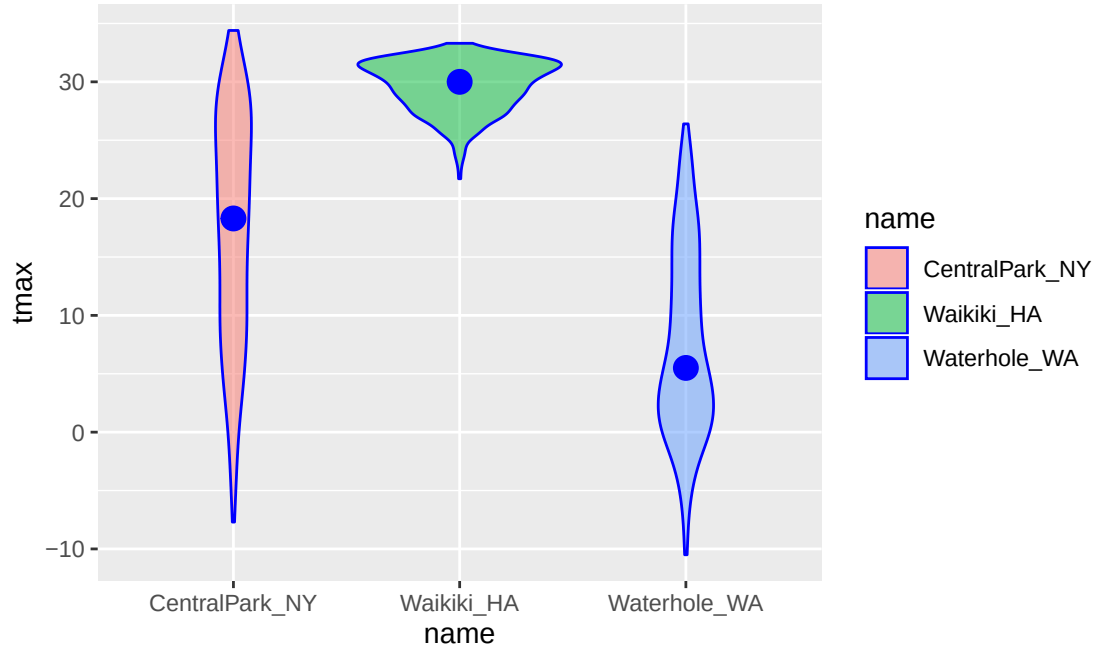
```
## Warning: Removed 3 rows containing non-finite values (stat_boxplot).
```




```
ggplot(weather_df, aes(x = name, y = tmax)) +
  geom_violin(aes(fill = name), color = "blue", alpha = .5) +
  stat_summary(fun.y = median, geom = "point", color = "blue", size = 4)
```

```
## Warning: Removed 3 rows containing non-finite values (stat_ydensity).
```

```
## Warning: Removed 3 rows containing non-finite values (stat_summary).
```

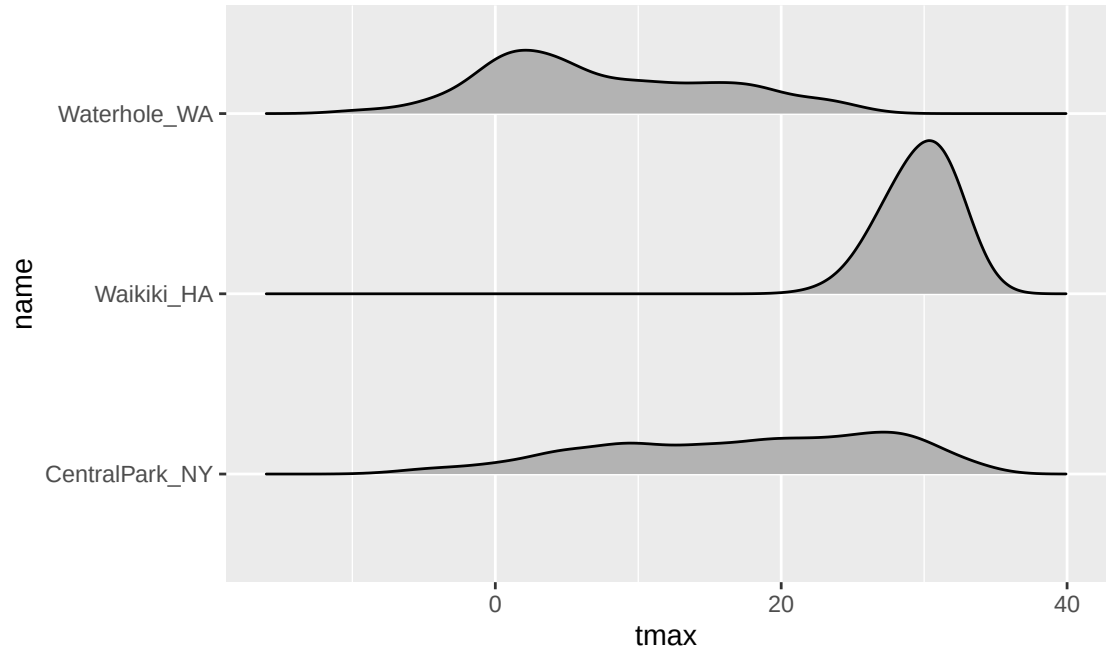


ridge plots

```
ggplot(weather_df, aes(x = tmax, y = name)) +
  geom_density_ridges(scale = .85)
```

```
## Picking joint bandwidth of 1.84
```

```
## Warning: Removed 3 rows containing non-finite values (stat_density_ridges).
```



save plots

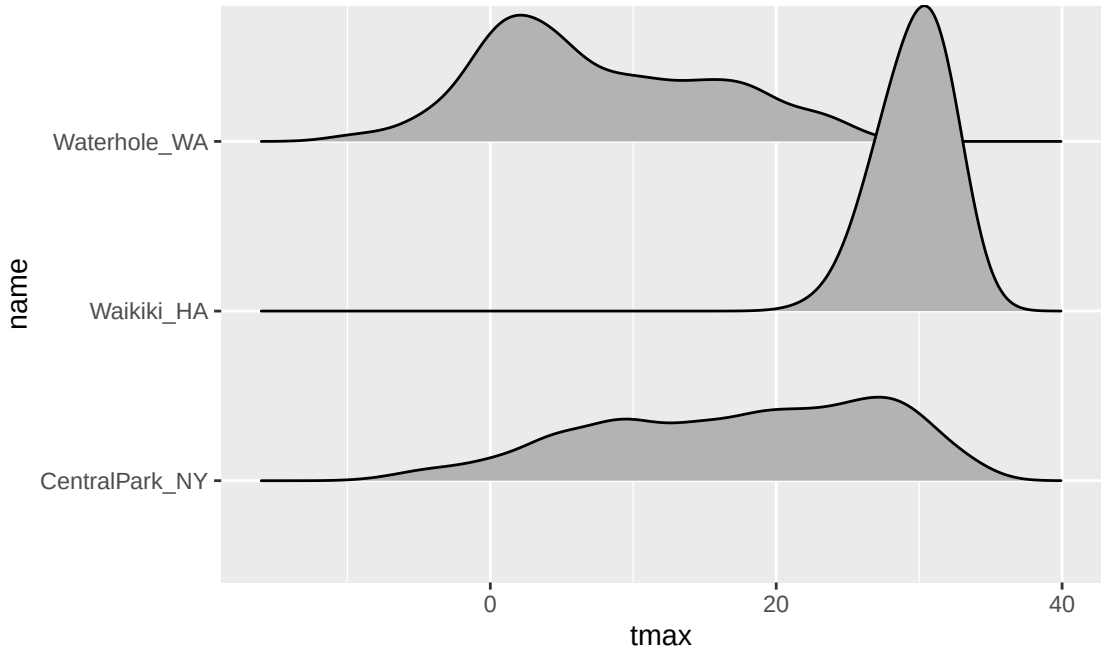
```
weather_plot = ggplot(weather_df, aes(x = tmin, y = tmax)) +
  geom_point(aes(color = name), alpha = .5)

ggsave("weather_plot.pdf", weather_plot, width = 8, height = 5)
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```

```
ggp_ridge_plot =
  weather_df %>%
  ggplot(aes(x= tmax, y = name)) +
  geom_density_ridges()
ggp_ridge_plot
```

```
## Picking joint bandwidth of 1.84
```

```
## Warning: Removed 3 rows containing non-finite values (stat_density_ridges).
```



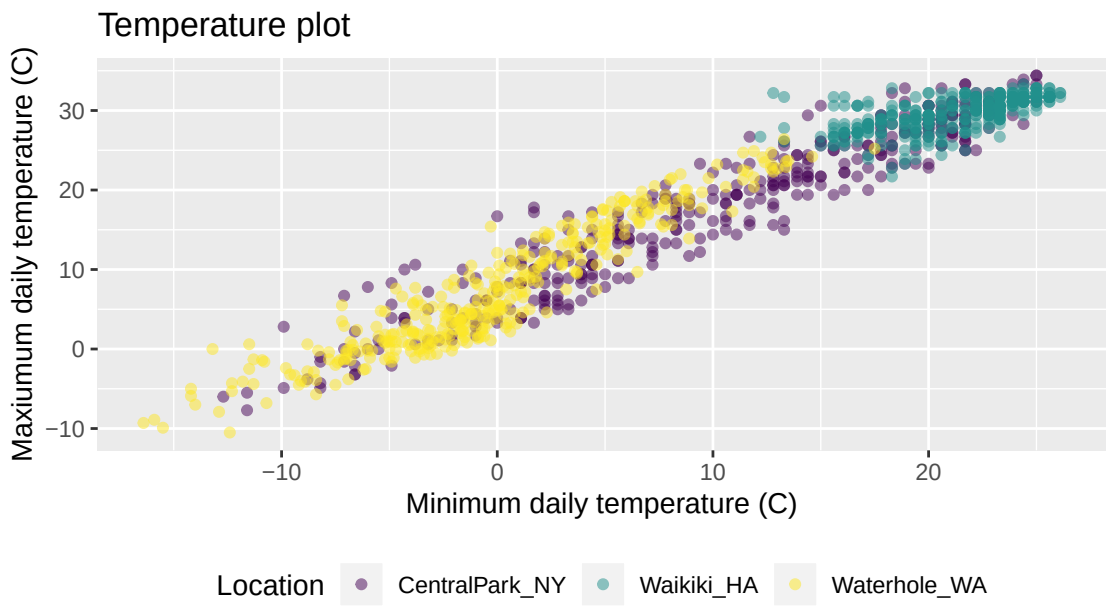

```
ggsave("ggp_ridge_plot.pdf",ggp_ridge_plot, width = 8, height = 5 )
```

```
## Picking joint bandwidth of 1.84
```

```
## Warning: Removed 3 rows containing non-finite values (stat_density_ridges).
```

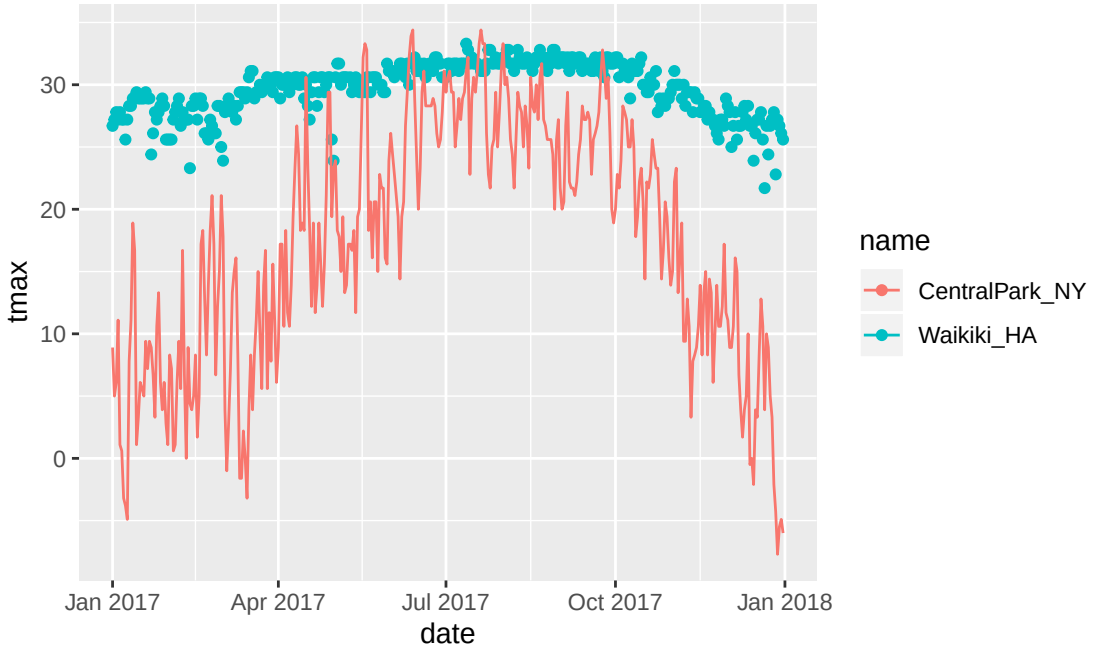
```
ggp_temp_plot =
  weather_df %>%
  ggplot(aes(x = tmin, y = tmax)) +
  geom_point(aes(color = name), alpha = .5) +
  labs(
    title = "Temperature plot",
    x = "Minimum daily temperature (C)",
    y = "Maxiumum daily temperature (C)",
    caption = "Data from the rnoaa package"
  ) +
  viridis::scale_color_viridis(
    name = "Location",
    discrete = TRUE
  )+theme(legend.position = "bottom")

ggp_temp_plot
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```



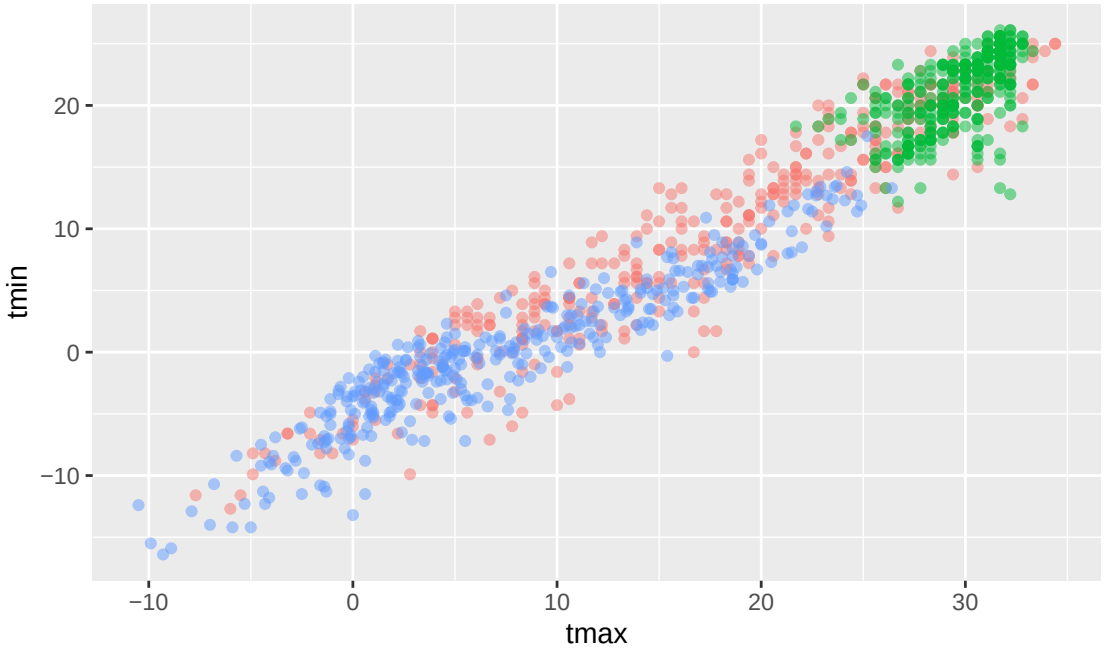

```
central_park =
  weather_df %>%
  filter(name == "CentralPark_NY")

waikiki =
  weather_df %>%
  filter(name == "Waikiki_HA")

ggplot(data = waikiki, aes(x = date, y = tmax, color = name)) +
  geom_point() +
  geom_line(data = central_park)
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

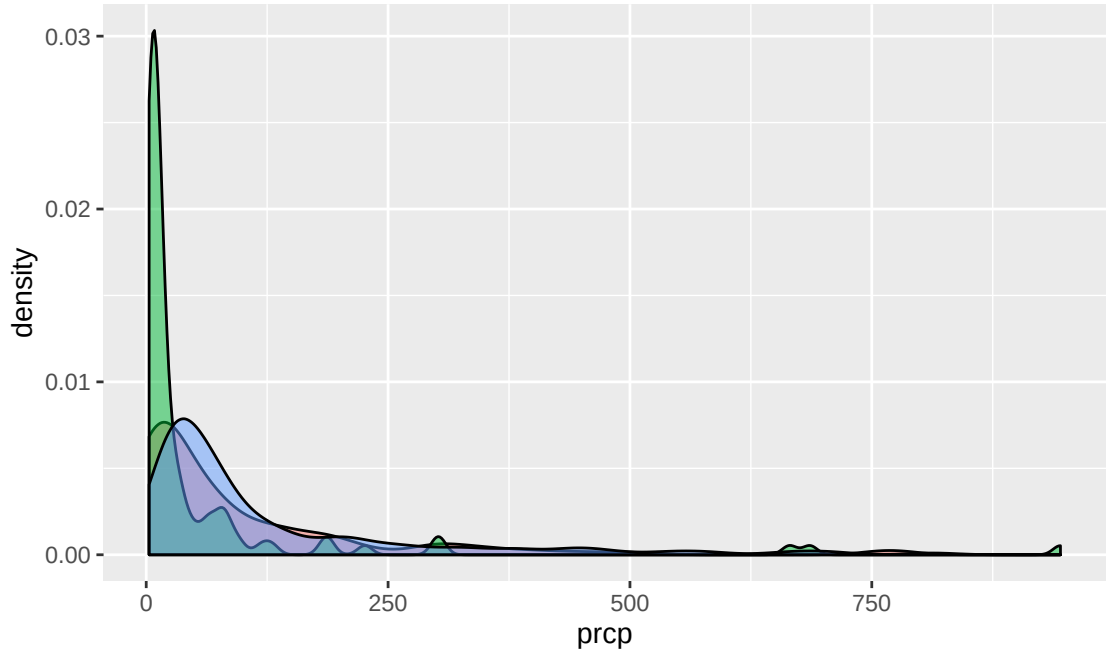



```
tmax_tmin_p =
  weather_df %>%
  ggplot(aes(x = tmax, y = tmin, color = name)) +
  geom_point(alpha = .5) +
  theme(legend.position = "none")
tmax_tmin_p
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```



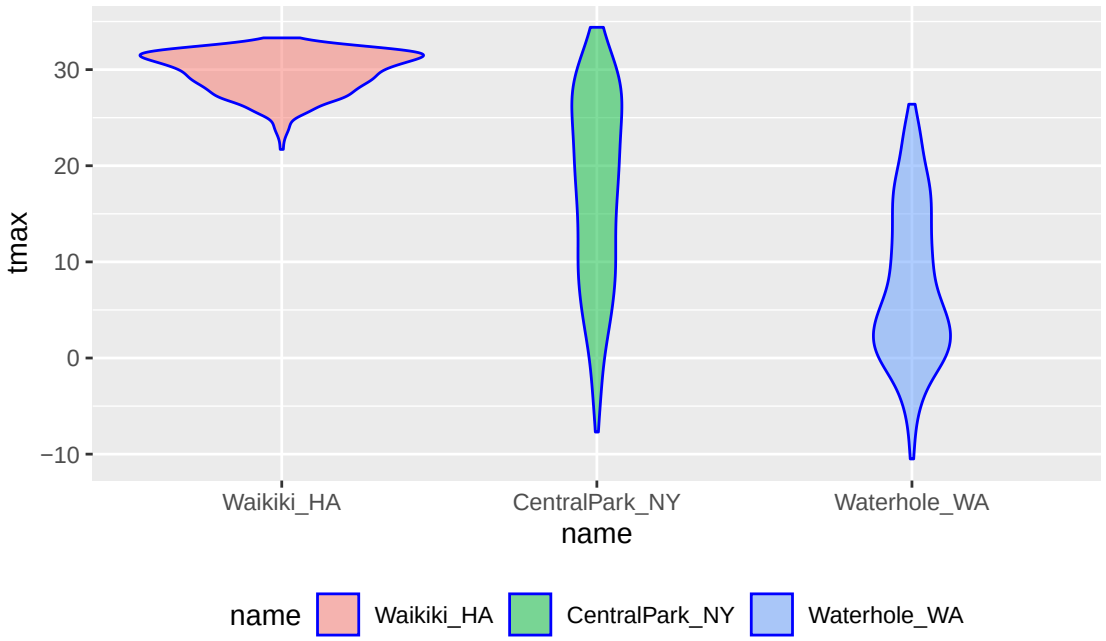
```
prcp_dens_p =
  weather_df %>%
  filter(prcp > 0) %>%
  ggplot(aes(x = prcp, fill = name)) +
  geom_density(alpha = .5) +
  theme(legend.position = "none")
prcp_dens_p
```




```
tmax_date_p =
  weather_df %>%
  ggplot(aes(x = date, y = tmax, color = name)) +
  geom_point(alpha = .5) +
  geom_smooth(se = FALSE) +
  theme(legend.position = "bottom")
```

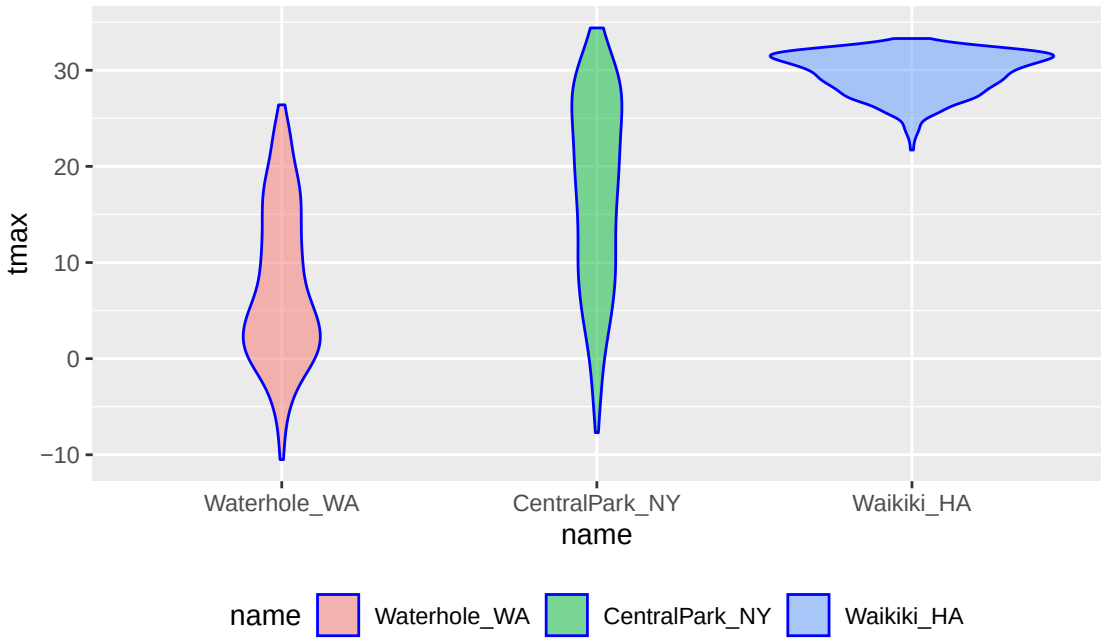
```
weather_df %>%
  mutate(name = forcats::fct_relevel(name, c("Waikiki_HA", "CentralPark_NY", "Waterhole_WA"))) %>%
  ggplot(aes(x = name, y = tmax)) +
  geom_violin(aes(fill = name), color = "blue", alpha = .5) +
  theme(legend.position = "bottom")
```

```
## Warning: Removed 3 rows containing non-finite values (stat_ydensity).
```



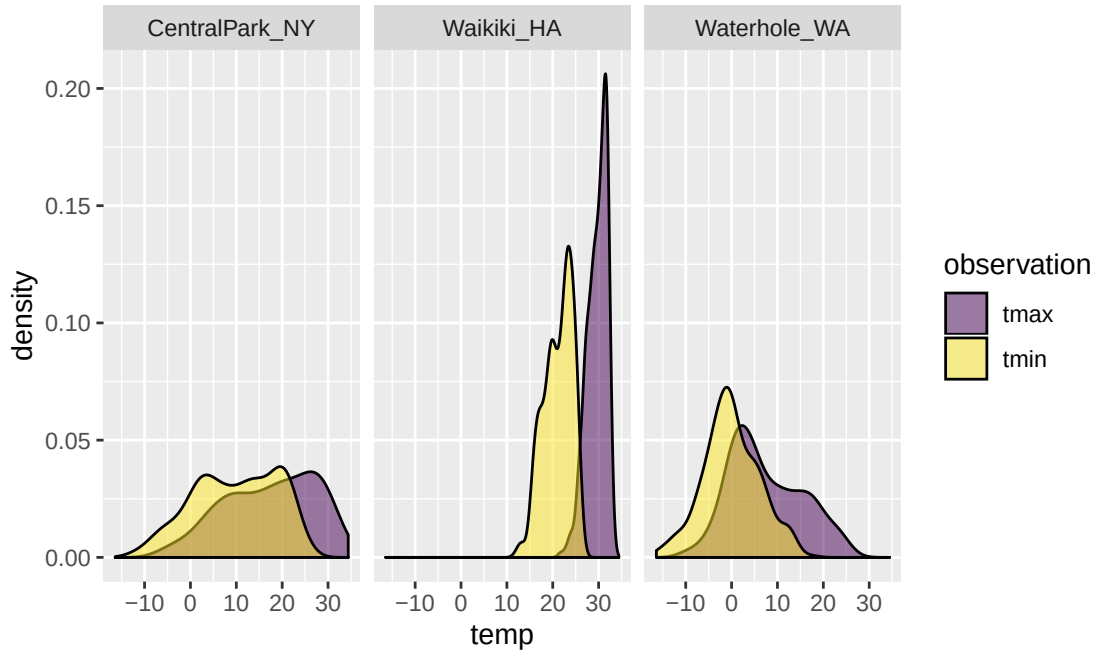

```
weather_df %>%
  mutate(name = forcats::fct_reorder(name, tmax)) %>%
  ggplot(aes(x = name, y = tmax)) +
  geom_violin(aes(fill = name), color = "blue", alpha = .5) +
  theme(legend.position = "bottom")
```

```
## Warning: Removed 3 rows containing non-finite values (stat_ydensity).
```



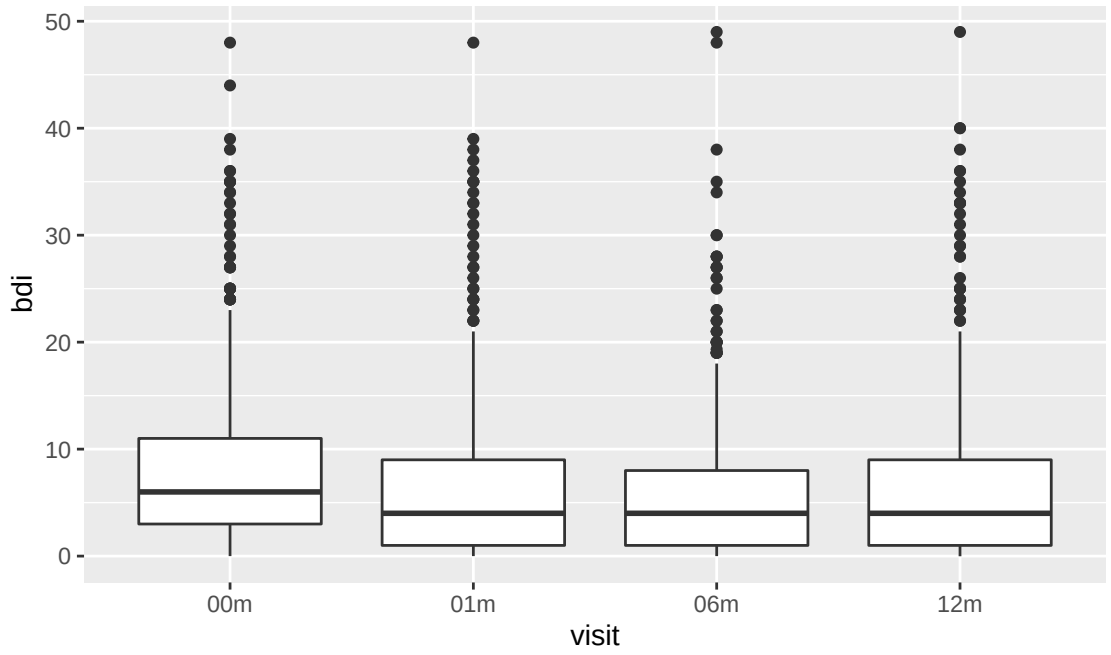
```
weather_df %>%
  select(name, tmax, tmin) %>%
  pivot_longer(
    tmax:tmin,
    names_to = "observation",
    values_to = "temp") %>%
  ggplot(aes(x = temp, fill = observation)) +
  geom_density(alpha = .5) +
  facet_grid(~name) +
  viridis::scale_fill_viridis(discrete = TRUE)
```

```
## Warning: Removed 18 rows containing non-finite values (stat_density).
```

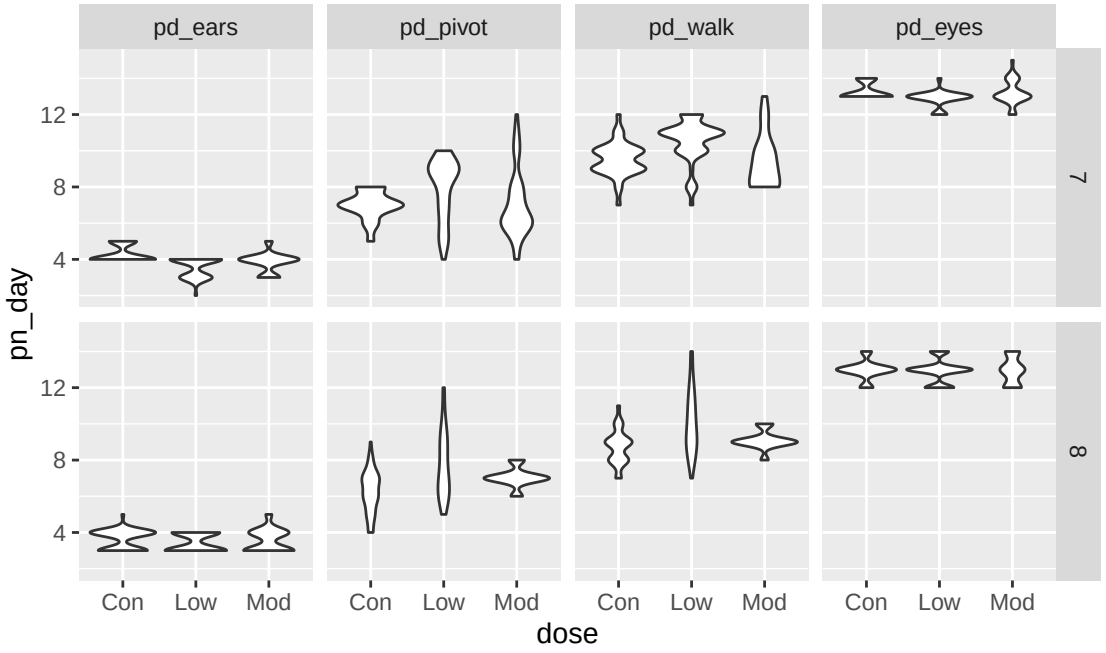



```
pulse_data =
  haven::read_sas("./public_pulse_data.sas7bdat") %>%
  janitor::clean_names() %>%
  pivot_longer(
    bdi_score_bl:bdi_score_12m,
    names_to = "visit",
    names_prefix = "bdi_score_",
    values_to = "bdi") %>%
  select(id, visit, everything()) %>%
  mutate(visit = recode(visit, "bl" = "00m"),
         visit = factor(visit, levels = str_c(c("00", "01", "06", "12"), "m"))) %>%
  arrange(id, visit)

ggplot(pulse_data, aes(x = visit, y = bdi)) +
  geom_boxplot()
```

```
## Warning: Removed 879 rows containing non-finite values (stat_boxplot).
```




```
pup_data =
  read_csv("./FAS_pups.csv", col_types = "ciiiii") %>%
  janitor::clean_names() %>%
  mutate(sex = recode(sex, `1` = "male", `2` = "female"))

litter_data =
  read_csv("./FAS_litters.csv", col_types = "ccddiiii") %>%
  janitor::clean_names() %>%
  select(-pups_survive) %>%
  separate(group, into = c("dose", "day_of_tx"), sep = 3) %>%
  mutate(wt_gain = gd18_weight - gd0_weight,
         day_of_tx = as.numeric(day_of_tx))

fas_data = left_join(pup_data, litter_data, by = "litter_number")

fas_data %>%
  select(sex, dose, day_of_tx, pd_ears:pd_walk) %>%
  pivot_longer(
    pd_ears:pd_walk,
    names_to = "outcome",
    values_to = "pn_day") %>%
  drop_na() %>%
  mutate(outcome = forcats::fct_reorder(outcome, day_of_tx, median)) %>%
  ggplot(aes(x = dose, y = pn_day)) +
  geom_violin() +
  facet_grid(day_of_tx ~ outcome)
```


pd_ears
pd_pivot
pd_walk
pd_eyes
7
8
pn_day
12
8
4
Con
Low
Mod
dose