Lab 4

Jazmin Hernandez

1. Read in the data

```r
met <- data.table::fread("met_all.gz")
```

```
Warning in data.table::fread("met_all.gz"): Discarded single-line footer:
<<721042,486,2019,8,20,15,35,28.228,-82.156,27,100,5,N,3.6,5,1036,5,M,N,16093,5,N,5,30>>
```

2. Prepare the data

```r
met <- met[met$temp >= -17, ]
met[met$temp %in% c(9999, 999)] <- NA
met[met$wind.sp %in% c(9999, 999)] <- NA
met[met$dew.point %in% c(9999, 999)] <- NA
met[met$date %in% c(9999, 999)] <- NA
met[met$elev %in% c(9999, 999)] <- NA
met[met$wind.dir %in% c(9999, 999)] <- NA
```

```r
met$date <- as.Date(paste(met$year, met$month, met$day, sep = "-"), format = "%Y-%m-%d")
```

```r
library(data.table)
library(lubridate)
```

```
Attaching package: 'lubridate'
```

```
The following objects are masked from 'package:data.table':

    hour, isoweek, mday, minute, month, quarter, second, wday, week,
    yday, year


The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```r
met[, week_of_year := week(date)]
met <- met[day(date) <= 7]
```

```r
mean_by_station <- met[, .(
  mean_temp = mean(temp, na.rm = TRUE),
  mean_rh = mean(rh, na.rm = TRUE),
  mean_wind_sp = mean(wind.sp, na.rm = TRUE),
  mean_vis_dist = mean(vis.dist, na.rm = TRUE),
  mean_dew_point = mean(dew.point, na.rm = TRUE),
  mean_lat = mean(lat, na.rm = TRUE),
  mean_lon = mean(lon, na.rm = TRUE),
  mean_elev = mean(elev, na.rm = TRUE)
), by = USAFID]
head(mean_by_station, 10)  # Displays the first 10 rows
```

```
    USAFID mean_temp  mean_rh mean_wind_sp mean_vis_dist mean_dew_point
     <int>     <num>    <num>        <num>         <num>          <num>
 1: 690150  34.07560 15.67014    3.7410714     16016.375       3.402381
 2: 720110  31.19246 52.32596    1.7515873     15984.442      19.321429
 3: 720113  24.40505 60.28099    1.5630934     16068.355      15.386064
 4: 720120  25.64310 88.30706    1.7730640     15315.468      23.451178
 5: 720137  23.16885 68.98459    1.4380952     15965.278      16.398016
 6: 720151  28.76667 34.36229    2.8007937     16057.877      10.327381
 7: 720169  23.79150 78.64946    0.9726721     15386.128      19.285425
 8: 720170  24.92540 77.99549    0.8345238      8254.121      20.404365
 9: 720172  28.21000 78.65215    0.5964824     15149.545      23.974490
10: 720175  27.38596 72.99504    1.2292135     15372.000      21.652247
    mean_lat   mean_lon mean_elev
       <num>      <num>     <num>
 1: 34.29995 -116.16595  695.1548
 2: 30.78400  -98.66200  336.0000
 3: 42.54300  -83.17800  222.0000
```

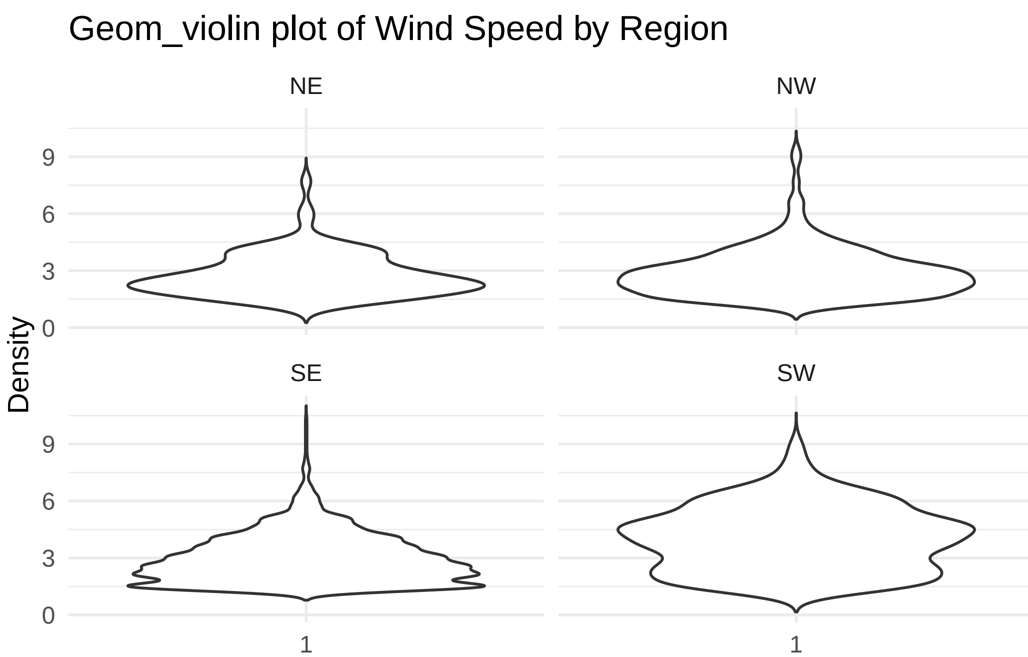
```
 4: 32.21733  -80.69986    6.0000
 5: 41.42500  -88.41900  178.0000
 6: 30.38300 -103.68300 1315.0000
 7: 38.58300  -91.00000  149.0000
 8: 37.18600  -88.75100  117.0000
 9: 34.54500  -94.20300  329.0000
10: 33.63600  -91.75600   82.0000
```

```
met[, region := fifelse(lat >= 39.71 & lon < -98, "NW",
                 fifelse(lat >= 39.71 & lon >= -98, "NE",
                 fifelse(lat < 39.71 & lon < -98, "SW", "SE")))]
```

```
met[, elev_cat := fifelse(elev > 252, "high", "low")]
```

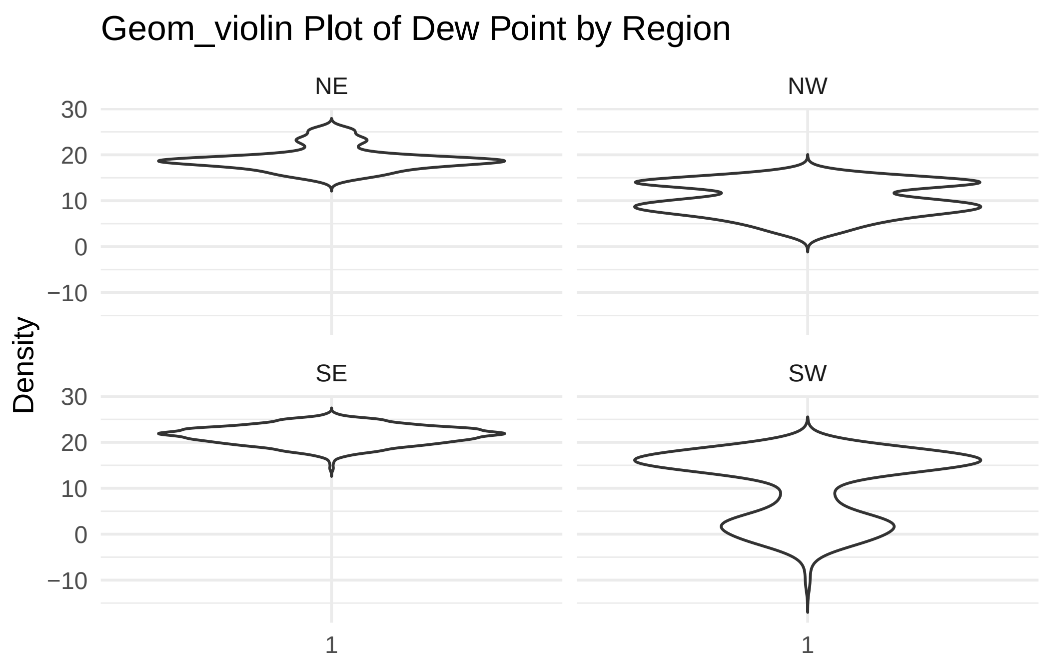
3. Use geom_violin to examine the wind speed and dew point by region

```
library(ggplot2)
ggplot(na.omit(met), aes(x = factor(1), y = wind.sp)) +
  geom_violin(trim = FALSE) +
  facet_wrap (~region) +
  labs(x = "wind speed", y = "Density", title = "Geom_violin plot of Wind Speed by Region") +
  theme_minimal() +
  theme(axis.title.x = element_blank())
```



Looking at the violin plots, it seems that NE, NW, and SE regions have a pretty low density of wind speed values looking at the section that is the widest for the respective regions. However, the SW region does seem to have the most variation in wind speed values as we can see that there are more wider sections at varying density distributions.

```
ggplot(na.omit(met), aes(x = factor(1), y = dew.point)) +
  geom_violin(trim = FALSE) +
  facet_wrap(~ region) +
  labs(x = "Dew Point", y = "Density", title = "Geom_violin Plot of Dew Point by Region") +
  theme_minimal() +
  theme(axis.title.x = element_blank())
```

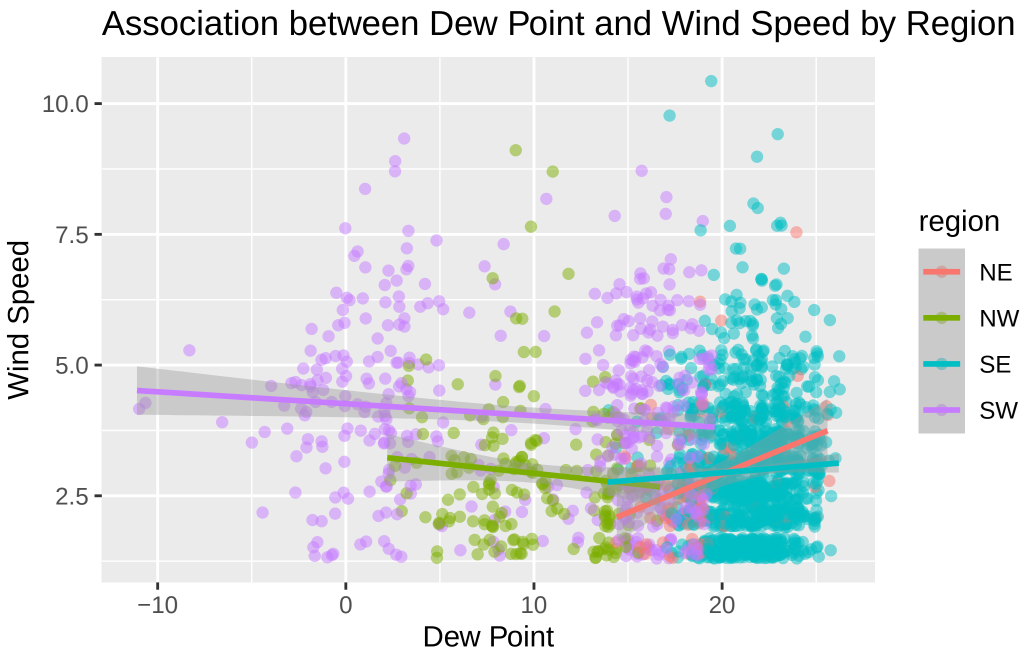


In this plot, it appears that NE and SE regions have a similar dew point concentration at around the same density distribution. Again, the SW region seems to be the one that varies the most with most of its dew point concentration located at higher density and the next majority of the concentration is in lower density. In contrast, the NE and SE regions have a higher dew point concentration at higher density and higher wind speeds at lower densities.

4. Use geom_jitter with stat_smooth to examine the association between dew point and wind speed by region

```
ggplot(na.omit(met), aes(x = dew.point, y = wind.sp, color = region)) +
  geom_jitter(alpha = 0.5, width = 0.2, height = 0.2) +
  stat_smooth(method = "lm", se = TRUE) +
  labs(x = "Dew Point", y = "Wind Speed", title = "Association between Dew Point and Wind Spe
```

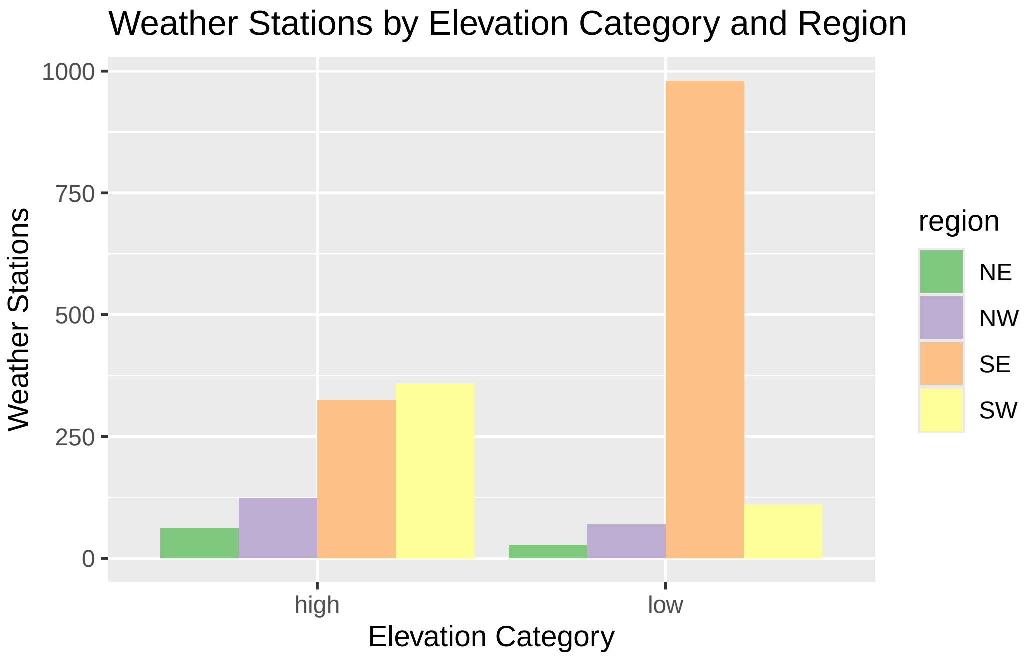
```
`geom_smooth()` using formula = 'y ~ x'
```



There seems to be a slight negative correlation between wind speed and dew point looking at NW and SW regions. However, this differs from the NE and SE regions with a slightly positive association in dew point and wind speed. It appears that for the SE region, a higher dew point is found at a lower wind speed. On the contrary, a lower dew point for the SW region appear to be related to a higher wind speed. The difference in regression lines or association between the NW and SW regions with the SE and NE regions can be due to climatic effects and even the effect of ocean currents.

5. Use geom_bar to create barplots of the weather stations by elevation category colored by region

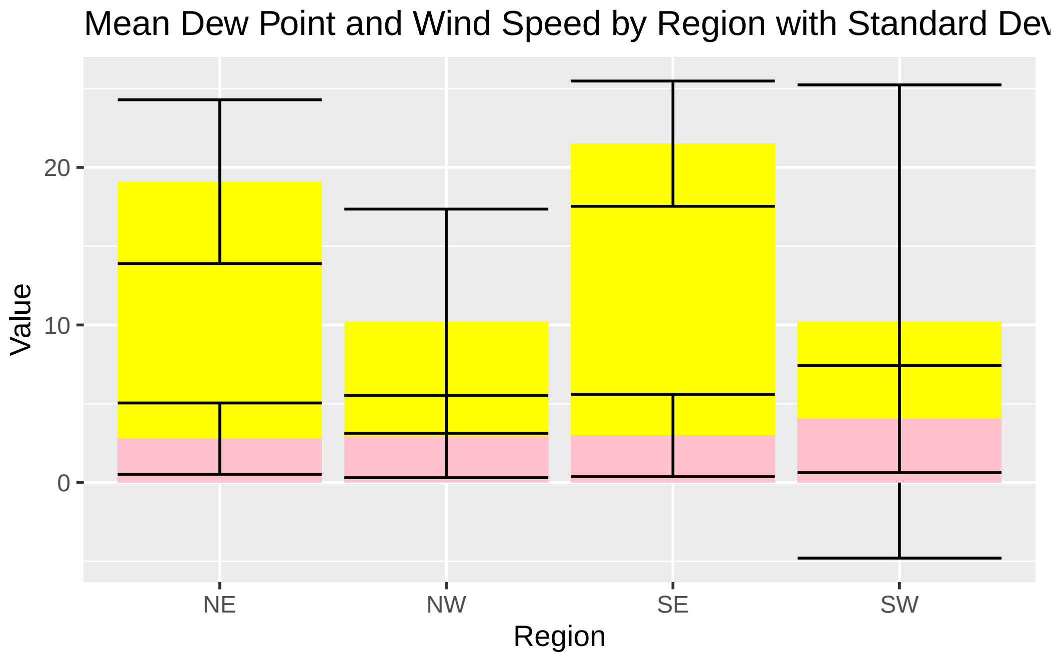
```
ggplot(na.omit(met), aes(x = elev_cat, fill = region)) +
  geom_bar(position = "dodge") +
  scale_fill_brewer(palette = "Accent") +
  labs(x = "Elevation Category", y = "Weather Stations", title = "Weather Stations by Elevat
```



From the graph, it is clear that the SE region has the most weather stations in lower elevation categories while SW has the most weather stations in the higher elevation categories. The significant number of weather stations in the SE region located at a lower elevation can indicate that major climatic occurrences are of importance at this elevation in the SE region.

6. Use stat_summary to examine mean dew point and wind speed by region with standard deviation error bars

```
ggplot(na.omit(met)) +
stat_summary(aes(x = region, y = dew.point),
fun.data = "mean_sdl", geom = "bar", fill = "yellow") +
stat_summary(aes(x = region, y = dew.point), fun.data = "mean_sdl", geom = "errorbar") +
stat_summary(aes(x = region, y = wind.sp), fun.data = "mean_sdl", geom = "bar", fill = "pink"
stat_summary(aes(x = region, y = wind.sp), fun.data = "mean_sdl", geom = "errorbar") +
labs(x = "Region", y = "Value", title = "Mean Dew Point and Wind Speed by Region with Standa
```

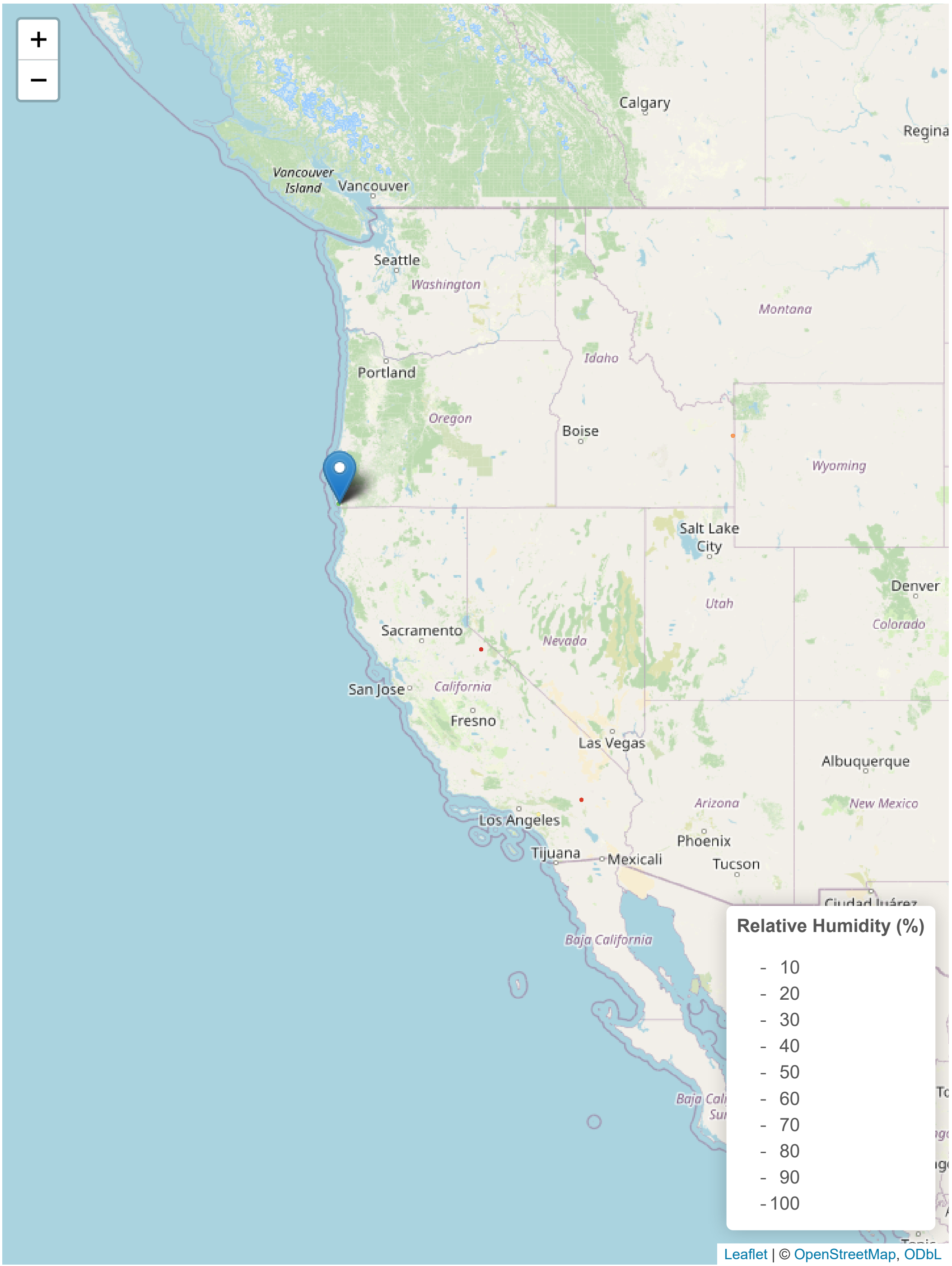


The dew point and wind speed seem to be significantly different between NE and SE regions. However, there does appear to be an overlap in error bars between dew point and wind speed in the NW and SW regions. Dew point is higher in NE and SE regions. There appears to be the most variation in dew point and wind speed in the SW region shown by the larger error bars.

7. Make a map showing the spatial trend in relative humidity in the US

```
library(leaflet)
library(RColorBrewer)
met_clean <- na.omit(met)
palette <- colorNumeric(palette = brewer.pal(11, "RdYlGn"), domain = met_clean$rh)
top_rh <- met_clean[order(-rh)][1:10]
leaflet(data = met_clean) |>
addTiles() |>
addCircles(
lng = ~lon,
lat = ~lat,
weight = 1,
radius = 500,
color = ~palette(rh),
```

```
stroke = TRUE,
fillOpacity = 0.5) |>
addMarkers(
data = top_rh,
lng = ~lon,
lat = ~lat,
popup = ~paste("USAFID:", USAFID)  ) |>
addLegend("bottomright",
pal = palette,
values = ~rh,
title = "Relative Humidity (%)",
opacity = 1)
```

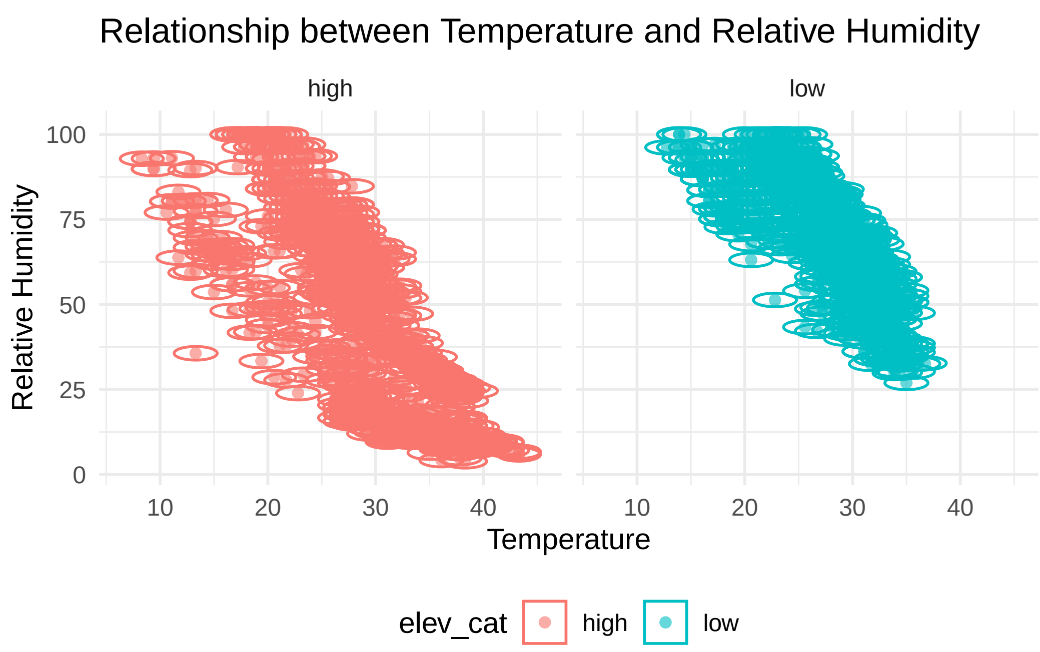


+
−
Calgary
Regina
Vancouver Island
Vancouver
Seattle
Washington
Montana
Portland
Idaho
Oregon
Boise
Wyoming
Salt Lake City
Denver
Utah
Colorado
Sacramento
Nevada
San Jose
California
Fresno
Las Vegas
Albuquerque
Arizona
New Mexico
Los Angeles
Phoenix
Tijuana
Mexicali
Tucson
Baja California
Relative Humidity (%)
- 10
- 20
- 30
- 40
- 50
- 60
- 70
- 80
- 90
- 100
Leaflet | © OpenStreetMap, ODbL

Relative humidity seems to be the lowest in California and in Texas and in areas west in general. However, moving towards the east region, we can see that relative humidity increases, especially looking at the NE and SE regions.

8. Use a ggplot extension

```r
library(ggforce)
ggplot(met_clean, aes(x = temp, y = rh, color = elev_cat)) +
  geom_point(alpha = 0.6) +
  geom_circle(aes(x0 = temp, y0 = rh, r = 2), alpha = 0.2) +
  facet_wrap(~ elev_cat) +
  labs(title = 'Relationship between Temperature and Relative Humidity',
       x = 'Temperature',
       y = 'Relative Humidity') +
  theme_minimal() +
  theme(legend.position = "bottom")
```




```r
library(ggplot2)
library(patchwork)
p1 <- ggplot(met_clean) + geom_point(aes(rh, wind.sp))
p2 <- ggplot(met_clean) + geom_boxplot(aes(elev_cat, wind.sp, group = elev_cat))
```

```
p3 <- ggplot(met_clean) + geom_smooth(aes(dew.point, wind.sp))
p4 <- ggplot(met_clean) + geom_bar(aes(region))
(p1 | p2 | p3) /
      p4
```

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