PM 566 Lab 4

AUTHOR
Ziquan 'Harrison' Liu

Step 1 Read in the data

```
library(ggplot2)
library(data.table)
library(dplyr)
```

```
Attaching package: 'dplyr'

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

    between, first, last

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

    filter, lag

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

    intersect, setdiff, setequal, union
```

```
library(leaflet)
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ forcats   1.0.0     ✔ stringr   1.5.2
✔ lubridate 1.9.4     ✔ tibble    3.3.0
✔ purrr     1.1.0     ✔ tidyr     1.3.1
✔ readr     2.1.5
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::between()     masks data.table::between()
✖ dplyr::filter()      masks stats::filter()
✖ dplyr::first()       masks data.table::first()
✖ lubridate::hour()    masks data.table::hour()
✖ lubridate::isoweek() masks data.table::isoweek()
✖ dplyr::lag()         masks stats::lag()
✖ dplyr::last()        masks data.table::last()
✖ lubridate::mday()    masks data.table::mday()
✖ lubridate::minute()  masks data.table::minute()
✖ lubridate::month()   masks data.table::month()
✖ lubridate::quarter() masks data.table::quarter()
✖ lubridate::second()  masks data.table::second()
```

```
✖ purrr::transpose()   masks data.table::transpose()
✖ lubridate::wday()    masks data.table::wday()
✖ lubridate::week()    masks data.table::week()
✖ lubridate::yday()    masks data.table::yday()
✖ lubridate::year()    masks data.table::year()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to
become errors
```

```
library(R.utils)
```

```
Loading required package: R.oo
Loading required package: R.methodsS3
R.methodsS3 v1.8.2 (2022-06-13 22:00:14 UTC) successfully loaded. See ?R.methodsS3 for
help.
R.oo v1.27.1 (2025-05-02 21:00:05 UTC) successfully loaded. See ?R.oo for help.

Attaching package: 'R.oo'

The following object is masked from 'package:R.methodsS3':

    throw

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

    getClasses, getMethods

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

    attach, detach, load, save

R.utils v2.13.0 (2025-02-24 21:20:02 UTC) successfully loaded. See ?R.utils for help.

Attaching package: 'R.utils'

The following object is masked from 'package:tidyr':

    extract

The following object is masked from 'package:utils':

    timestamp

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

    cat, commandArgs, getOption, isOpen, nullfile, parse, use, warnings
```

```
if (!file.exists("met_all.gz"))
  download.file(
    url = "https://raw.githubusercontent.com/USCbiostats/data-science-data/master/02_met/
    destfile = "met_all.gz",
```

```
    method   = "libcurl",
    timeout  = 60
    )
met <- data.table::fread("met_all.gz")
```

```
head(met)
```

```
   USAFID  WBAN  year month   day  hour   min   lat      lon  elev wind.dir
    <int> <int> <int> <int> <int> <int> <int> <num>    <num> <int>    <int>
1: 690150 93121  2019     8     1     0    56  34.3 -116.166   696      220
2: 690150 93121  2019     8     1     1    56  34.3 -116.166   696      230
3: 690150 93121  2019     8     1     2    56  34.3 -116.166   696      230
4: 690150 93121  2019     8     1     3    56  34.3 -116.166   696      210
5: 690150 93121  2019     8     1     4    56  34.3 -116.166   696      120
6: 690150 93121  2019     8     1     5    56  34.3 -116.166   696       NA
   wind.dir.qc wind.type.code wind.sp wind.sp.qc ceiling.ht ceiling.ht.qc
        <char>         <char>   <num>     <char>      <int>         <int>
1:          5              N     5.7          5      22000             5
2:          5              N     8.2          5      22000             5
3:          5              N     6.7          5      22000             5
4:          5              N     5.1          5      22000             5
5:          5              N     2.1          5      22000             5
6:          9              C     0.0          5      22000             5
   ceiling.ht.method sky.cond vis.dist vis.dist.qc vis.var vis.var.qc  temp
               <char>   <char>    <int>      <char>  <char>     <char> <num>
1:                 9        N    16093           5       N          5  37.2
2:                 9        N    16093           5       N          5  35.6
3:                 9        N    16093           5       N          5  34.4
4:                 9        N    16093           5       N          5  33.3
5:                 9        N    16093           5       N          5  32.8
6:                 9        N    16093           5       N          5  31.1
   temp.qc dew.point dew.point.qc atm.press atm.press.qc       rh
    <char>     <num>       <char>     <num>        <int>    <num>
1:       5      10.6            5    1009.9            5 19.88127
2:       5      10.6            5    1010.3            5 21.76098
3:       5       7.2            5    1010.6            5 18.48212
4:       5       5.0            5    1011.6            5 16.88862
5:       5       5.0            5    1012.7            5 17.38410
6:       5       5.6            5    1012.7            5 20.01540
```

```
summary(met)
```

```
     USAFID           WBAN            year          month        day
 Min.   :690150   Min.   :  116   Min.   :2019   Min.   :8   Min.   : 1
 1st Qu.:720928   1st Qu.: 3706   1st Qu.:2019   1st Qu.:8   1st Qu.: 8
 Median :722728   Median :13860   Median :2019   Median :8   Median :16
 Mean   :723099   Mean   :29539   Mean   :2019   Mean   :8   Mean   :16
 3rd Qu.:725090   3rd Qu.:54767   3rd Qu.:2019   3rd Qu.:8   3rd Qu.:24
 Max.   :726813   Max.   :94998   Max.   :2019   Max.   :8   Max.   :31
```

```
      hour            min              lat             lon
 Min.   : 0.00   Min.   : 0.00   Min.   :24.55   Min.   :-124.29
 1st Qu.: 5.00   1st Qu.:23.00   1st Qu.:33.97   1st Qu.: -98.02
 Median :11.00   Median :50.00   Median :38.35   Median : -91.71
 Mean   :11.34   Mean   :39.56   Mean   :37.94   Mean   : -92.15
 3rd Qu.:17.00   3rd Qu.:55.00   3rd Qu.:41.94   3rd Qu.: -82.99
 Max.   :23.00   Max.   :59.00   Max.   :48.94   Max.   : -68.31

      elev           wind.dir      wind.dir.qc        wind.type.code
 Min.   : -13.0   Min.   :  3      Length:2377343     Length:2377343
 1st Qu.: 101.0   1st Qu.:120      Class :character   Class :character
 Median : 252.0   Median :180      Mode  :character   Mode  :character
 Mean   : 415.8   Mean   :185
 3rd Qu.: 400.0   3rd Qu.:260
 Max.   :9999.0   Max.   :360
                  NA's   :785290
    wind.sp        wind.sp.qc          ceiling.ht     ceiling.ht.qc
 Min.   : 0.000   Length:2377343     Min.   :    0    Min.   :1.000
 1st Qu.: 0.000   Class :character   1st Qu.: 3048    1st Qu.:5.000
 Median : 2.100   Mode  :character   Median :22000    Median :5.000
 Mean   : 2.459                      Mean   :16166    Mean   :5.027
 3rd Qu.: 3.600                      3rd Qu.:22000    3rd Qu.:5.000
 Max.   :36.000                      Max.   :22000    Max.   :9.000
 NA's   :79693                       NA's   :121275
 ceiling.ht.method    sky.cond            vis.dist      vis.dist.qc
 Length:2377343     Length:2377343     Min.   :     0   Length:2377343
 Class :character   Class :character   1st Qu.: 16093   Class :character
 Mode  :character   Mode  :character   Median : 16093   Mode  :character
                                       Mean   : 14921
                                       3rd Qu.: 16093
                                       Max.   :160000
                                       NA's   :80956
   vis.var           vis.var.qc             temp          temp.qc
 Length:2377343     Length:2377343     Min.   :-40.00   Length:2377343
 Class :character   Class :character   1st Qu.: 19.60   Class :character
 Mode  :character   Mode  :character   Median : 23.50   Mode  :character
                                       Mean   : 23.59
                                       3rd Qu.: 27.80
                                       Max.   : 56.00
                                       NA's   :60089
   dew.point      dew.point.qc         atm.press        atm.press.qc
 Min.   :-37.20   Length:2377343     Min.   : 960.5    Min.   :1.000
 1st Qu.: 13.80   Class :character   1st Qu.:1011.8    1st Qu.:5.000
 Median : 18.10   Mode  :character   Median :1014.1    Median :9.000
 Mean   : 17.02                      Mean   :1014.2    Mean   :7.728
 3rd Qu.: 21.70                      3rd Qu.:1016.4    3rd Qu.:9.000
 Max.   : 36.00                      Max.   :1059.9    Max.   :9.000
 NA's   :66288                       NA's   :1666274
       rh
 Min.   :  0.833
```

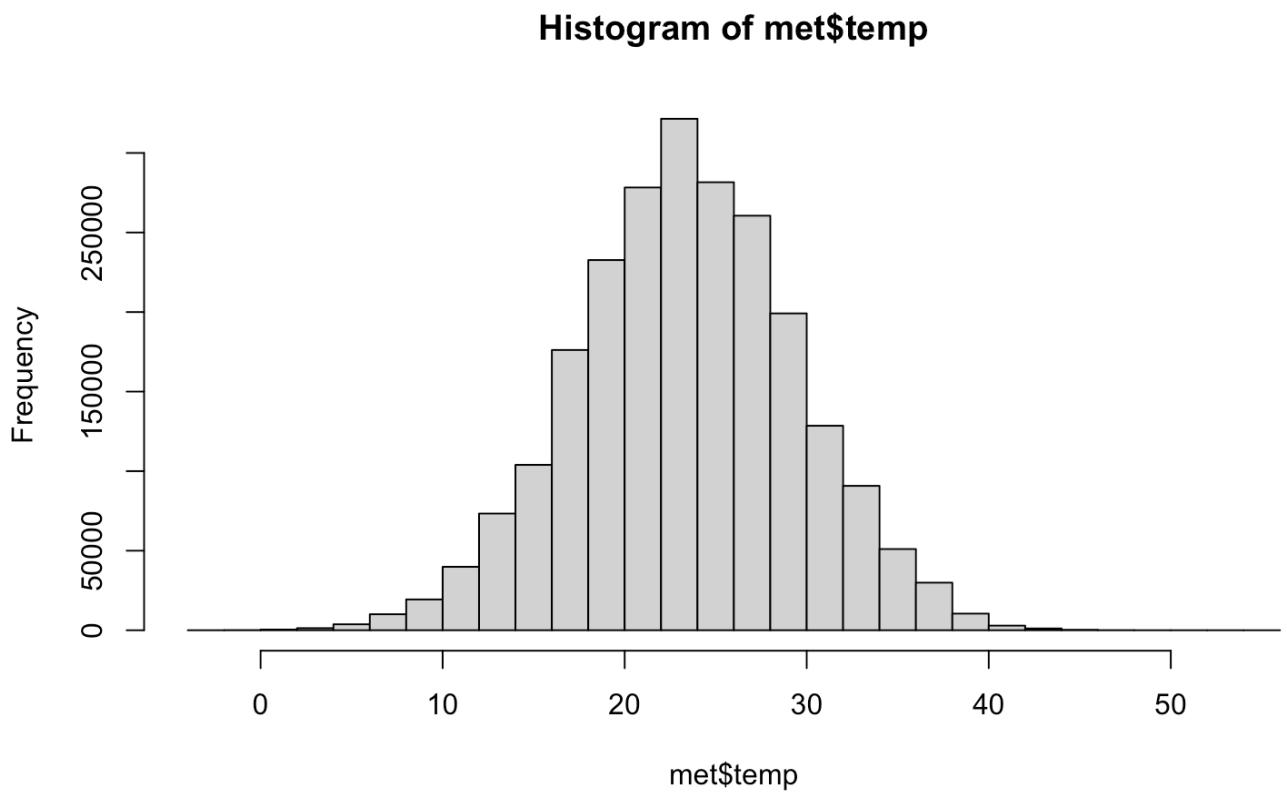
```
1st Qu.: 55.790
Median : 76.554
Mean   : 71.641
3rd Qu.: 90.629
Max.   :100.000
NA's   :66426
```

Step 2 Prepare the data

```
# Remove temperatures less than -17C
met <- met[met$temp > -17, ]
summary(met$temp)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -3.00   19.60   23.50   23.59   27.80   56.00
```

```
hist(met$temp)
```




```
# Make sure there is no missing data in the key variables coded as 9999, 999, etc.
met$elev[met$elev == 9999.0] <- NA
```

```
# Generate a date variable using the functions as.Date() (hint: You will need the followi
```

```
met[, week := week(as.Date(paste(year,month,day,sep="-")))]
```

```
# Using the data.table::week function, keep the observations of the first week of the mon
met <- met[week == min(week, na.rm = TRUE)]
```

```
# Compute the mean by station of the variables temp , rh , wind.sp , vis.dist , dew.point
met_avg <- met[,.(temp=mean(temp,na.rm=TRUE), rh=mean(rh,na.rm=TRUE),
                wind.sp=mean(wind.sp,na.rm=TRUE),
                vis.dist=mean(vis.dist, na.rm=TRUE),
                dew.point = mean(dew.point, na.rm=TRUE),
                lat=mean(lat,na.rm=TRUE), lon=mean(lon,na.rm=TRUE),
                elev=mean(elev,na.rm=TRUE)), by = "USAFID"
                ]
```

```
# Create a region variable for NW, SW, NE, SE based on lon = -98.00 and lat = 39.71 degre
met_avg$region <- ifelse(met_avg$lon > -98 & met_avg$lat >39.71, "north east",
                         ifelse(met_avg$lon > -98 & met_avg$lat < 39.71, "south east",
                                ifelse(met_avg$lon < -98 & met_avg$lat >39.71, "north wes
table(met_avg$region)
```

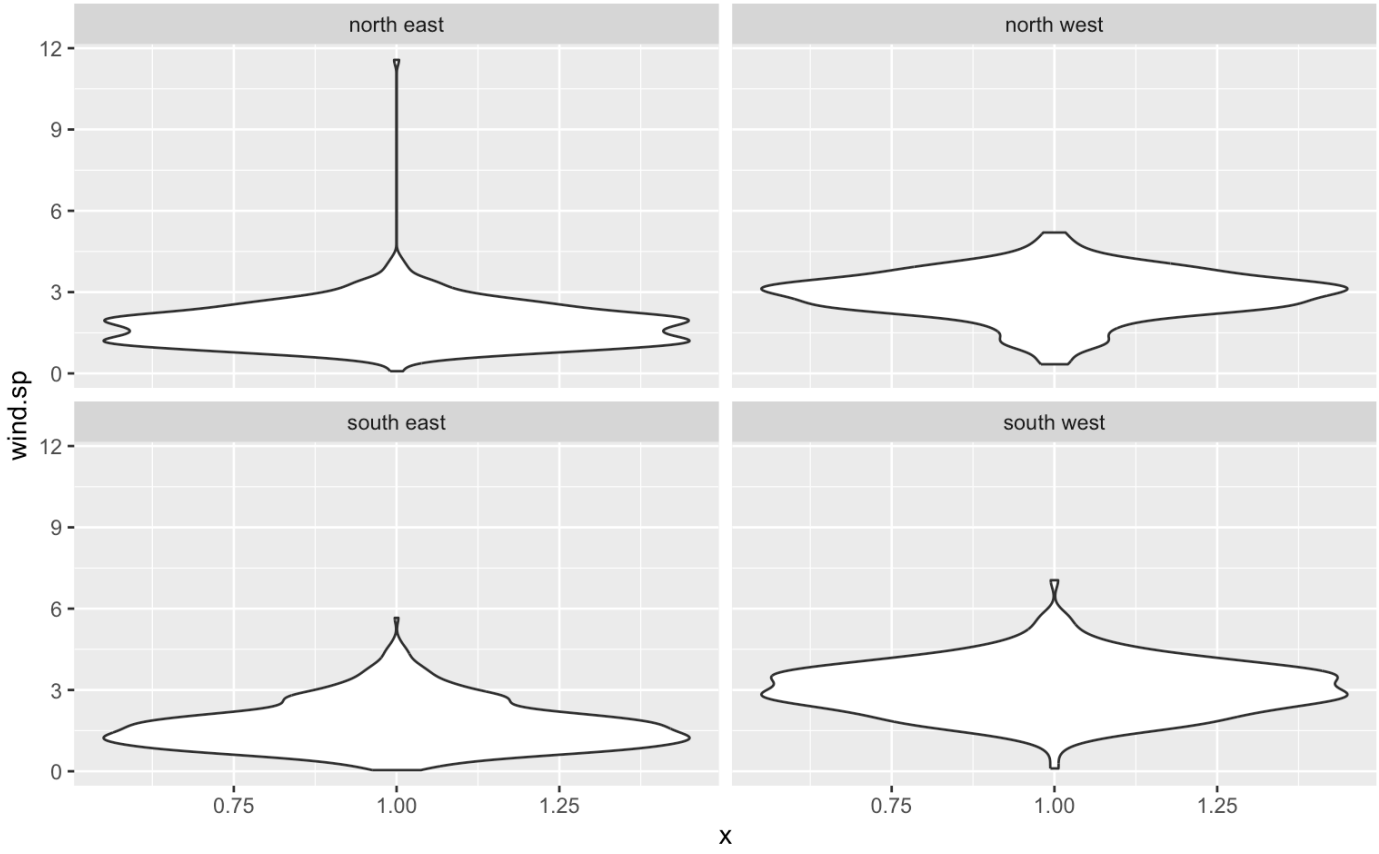
```
north east north west south east south west
       484        146        649        296
```

```
# Create a categorical variable for elevation as in the lecture slides
met_avg$elev_cat <- ifelse(met_avg$elev> 252, "high", "low")
```

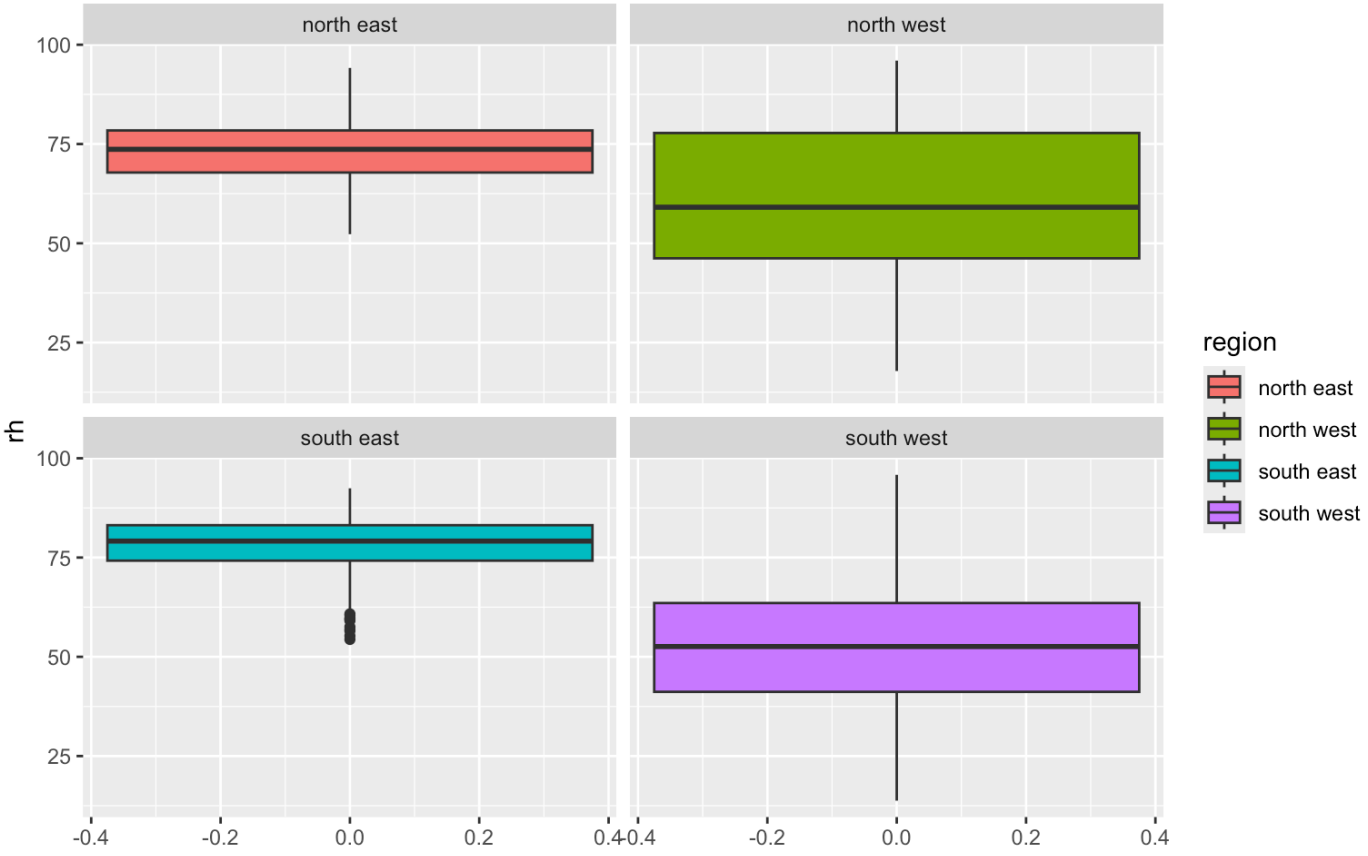
Step 3 Use geom_violin to examine the wind speed and dew point by region

```
met_avg %>%
  filter(!is.na(dew.point))%>% #  # make sure to deal with NAs
ggplot()+
  geom_violin(mapping = aes(y=wind.sp, x=1)) +
  facet_wrap(~region, nrow=2) # use facets (by region in this case)
```

```
Warning: Removed 13 rows containing non-finite outside the scale range
(`stat_ydensity()`).
```




```
met_avg %>%
  filter(!is.na(wind.sp)) %>% # make sure to deal with NAs
ggplot()+
  geom_boxplot(mapping = aes(y=rh, fill=region)) +
  facet_wrap(~region, nrow=2) # use facets
```

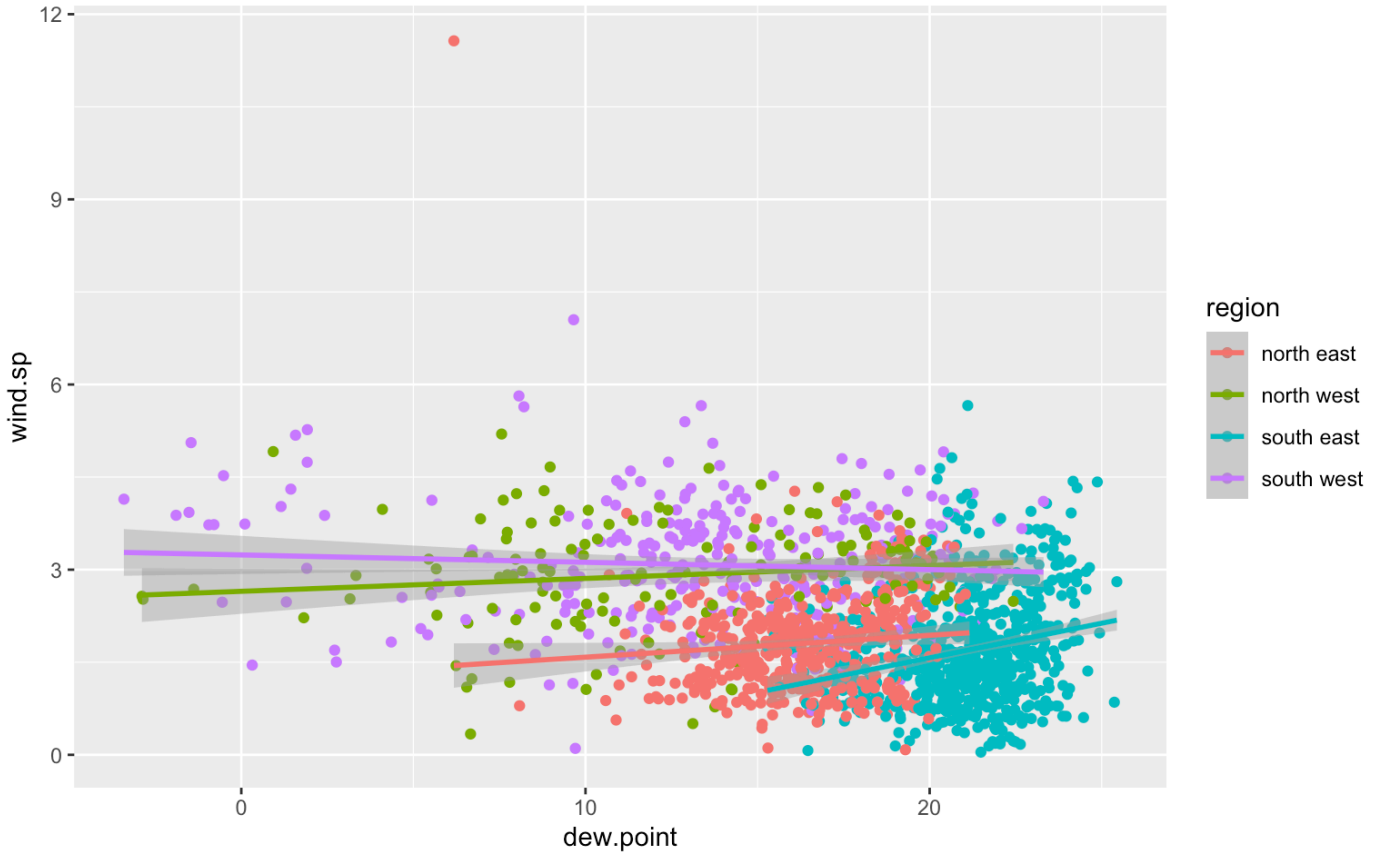


Descrption (Template): The violin plots reveal distinct regional patterns in wind speed distributions. Ihe northeast region shows..., while the northwest region displays.... The southeast region exhibits..., and the southwest region demonstrates.... Overall, the... region appears to have the highest wind speeds, while the... region shows the most concentrated distribution around... m/s.

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

```
met_avg %>%
filter(!is.na(dew.point) & !is.na(wind.sp)) %>% # make sur to deal with NAs
  ggplot(mapping = aes(x=dew.point, y=wind.sp, color=region))+ # color poiunts by region
  geom_jitter() +
  stat_smooth(method=lm) # fit a linear reg line of region
```

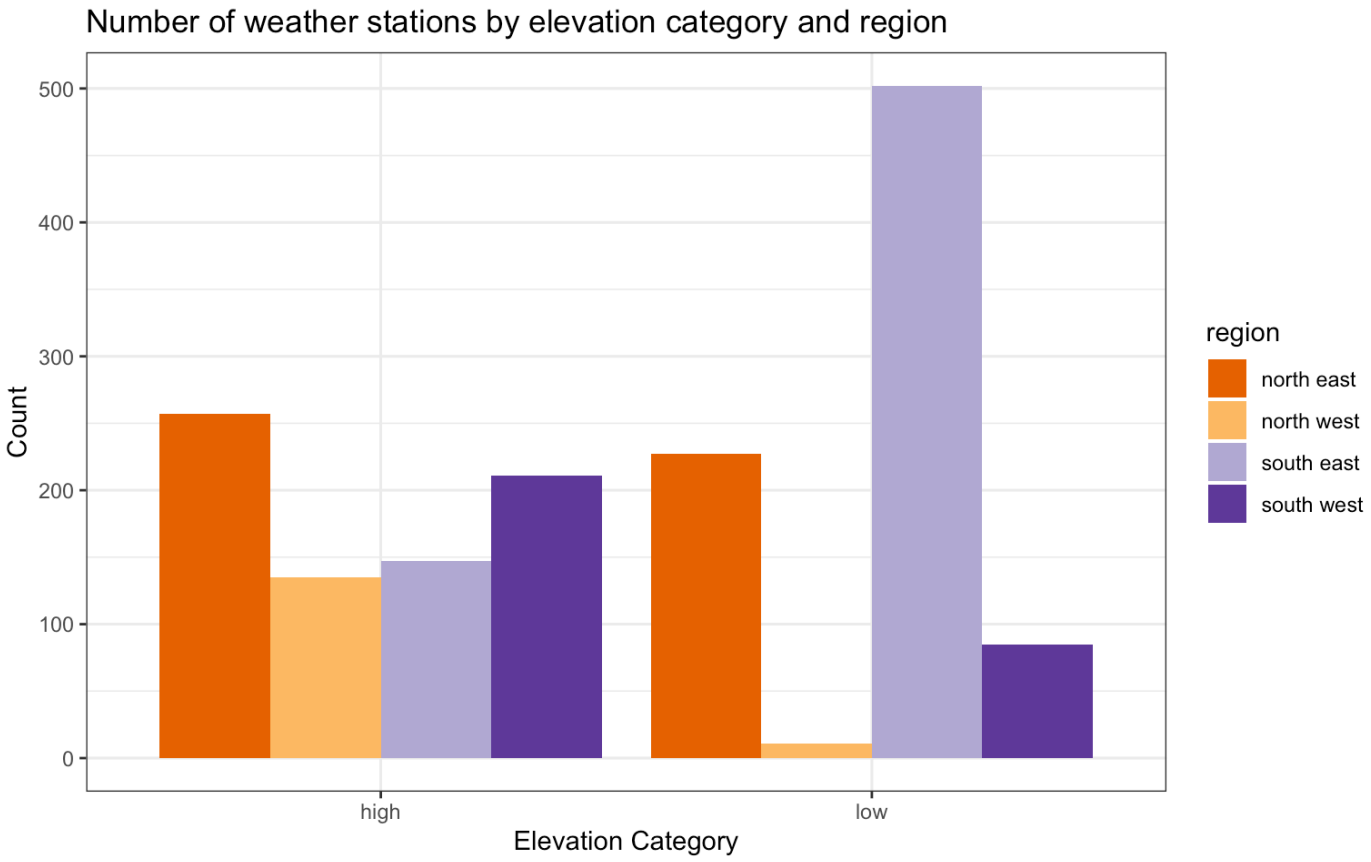
```
`geom_smooth()` using formula = 'y ~ x'
```



Description (Template): The scatter plot demonstrates a... relationship between dew point and relative humidity across all regions. All regional regression lines show... slopes, indicating that as dew point increases, relative humidity.... The relationship appears strongest in the... regions, while the... regions show more.... This pattern makes meteorological sense because...

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

```
met_avg %>%
filter(!(region %in% NA)) %>% # make sure to deal with NA values
  ggplot()+
  geom_bar(mapping=aes(x=elev_cat,fill=region), position = "dodge")+ # Bars by elevation
  scale_fill_brewer(palette = "PuOr")+ # change colors from the default. (Color region us
  labs(title="Number of weather stations by elevation category and region", x="Elevation
  theme_bw()
```

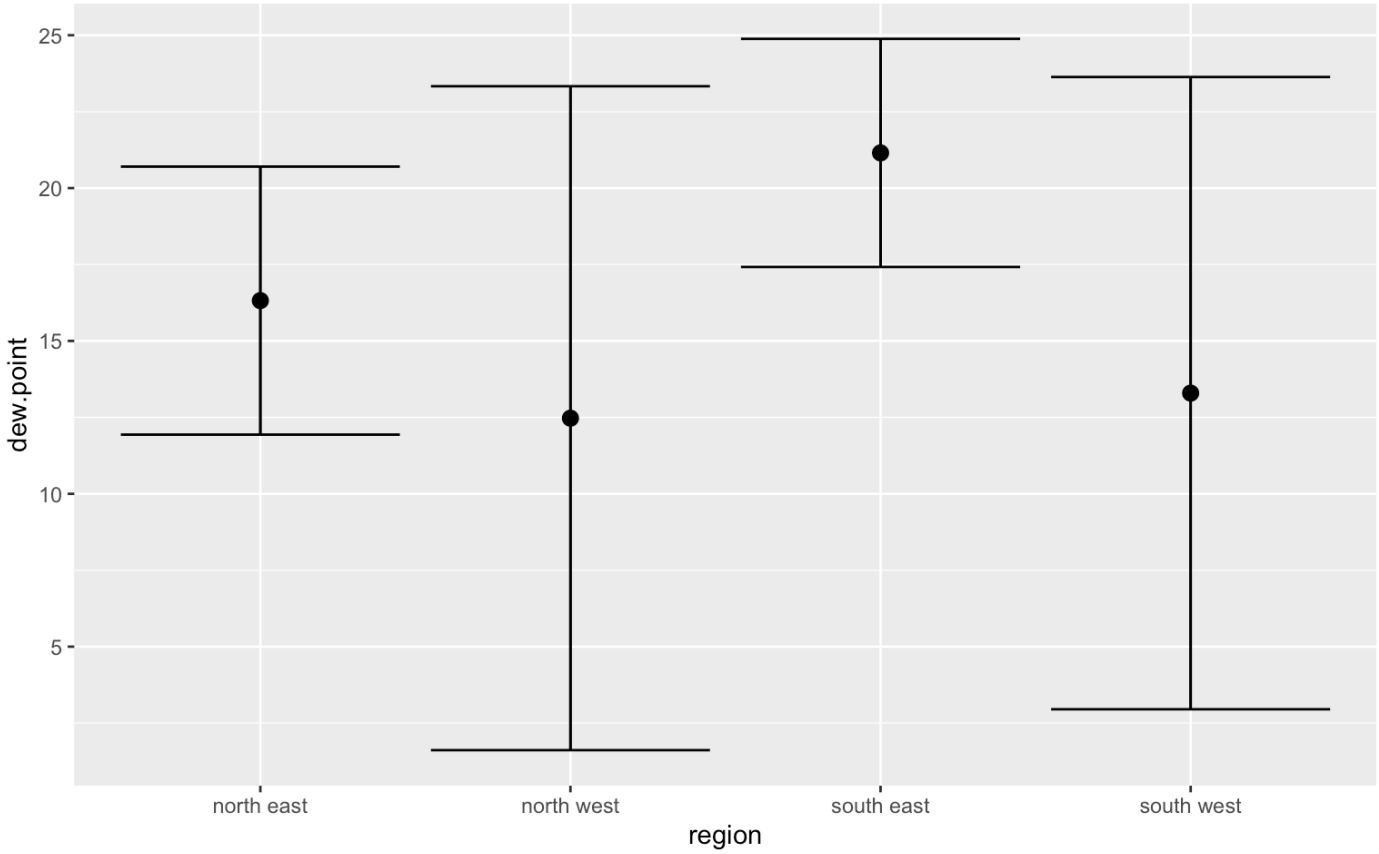


Description (Template): The bar chart reveals significant differences in weather station distribution across regions and elevation categories. The... region has the highest number of stations overall, particularly concentrated at... elevations. The... region shows more stations at high elevations compared to..., which likely reflects.... The... region has the fewest total stations but shows.... This distribution pattern suggests

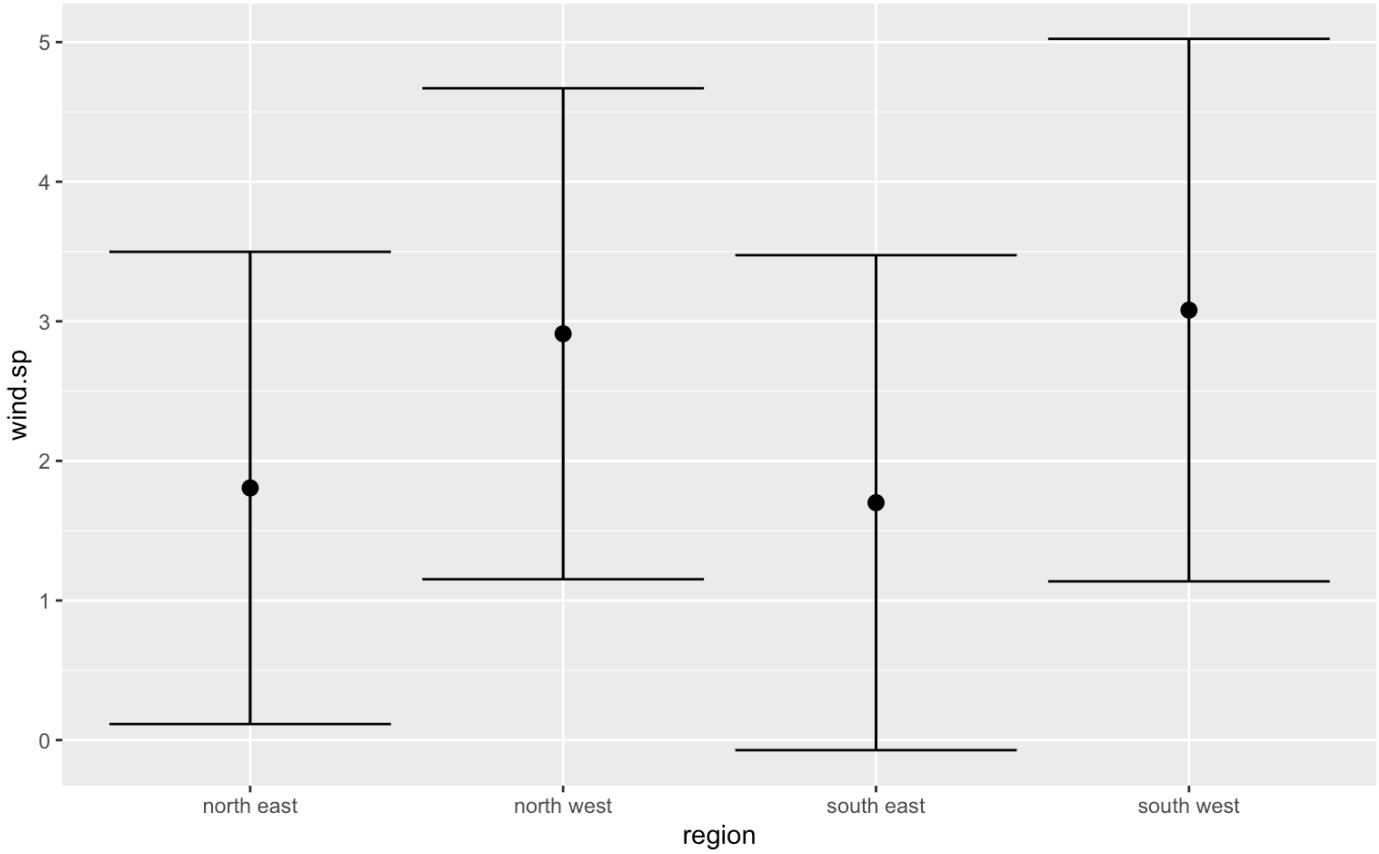
---

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

```
met_avg %>%
filter(!is.na(dew.point)) %>% # make sure to deal with NA values
  ggplot(mapping=aes(x=region, y=dew.point)) +
  stat_summary(fun.data="mean_sdl", geom="errorbar") +
  stat_summary(fun.data="mean_sdl")
```




```
met_avg %>%
filter(!is.na(wind.sp)) %>% # make sure to deal with NA values
  ggplot(mapping=aes(x=region, y=wind.sp)) +
  stat_summary(fun.data="mean_sdl", geom="errorbar") +
  stat_summary(fun.data="mean_sdl")
```



Description (Template):The region with the highest mean dew point (approximately...°C), indicating.... The... region displays the lowest mean dew point around... °C, suggesting.... The error bars reveal that... region has the most variable conditions, while... region shows...; Regional differences in wind speed are... compared to dew point patterns. The... regions show higher mean wind speeds around... m/s with... confidence intervals, indicating.... The... regions display lower mean wind speeds suggesting...

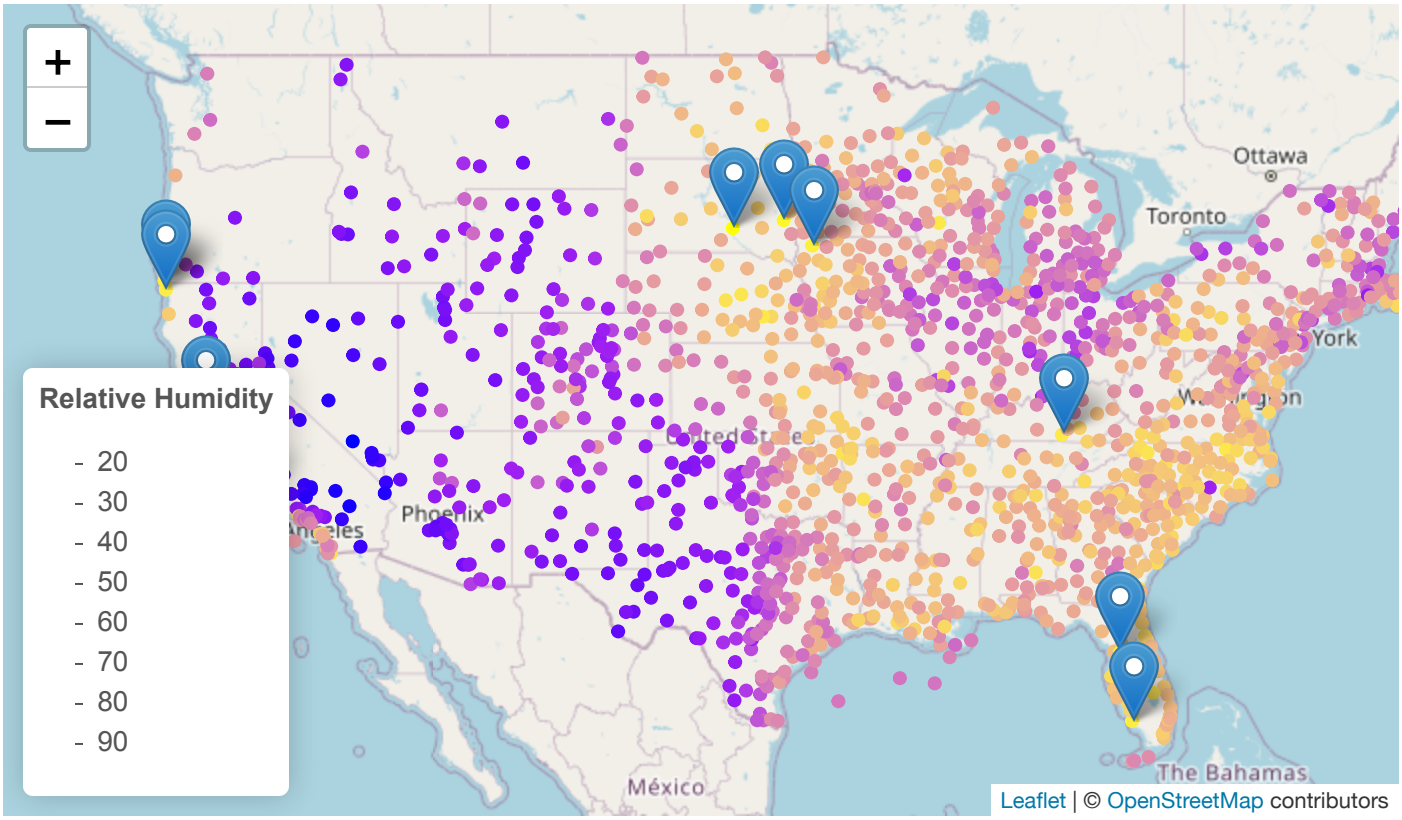
Step 7 Make a map showing the spatial trend in relative humidity in the US

```
met_avg2<-met_avg[!is.na(rh)]

# Top five
top5 <- met_avg2[rank(-rh) <= 10]

rh_pal = colorNumeric(c('blue','purple','yellow'), domain=met_avg2$rh)
leaflet(met_avg2) %>%
  addProviderTiles('OpenStreetMap') %>%
```

```
  addCircles(lat=~lat, lng=~lon, color=~rh_pal(rh), label=~paste0(round(rh,2), ' rh'), op
  addMarkers(lat=~lat, lng=~lon, label=~paste0(round(rh,2), ' rh'), data = top5) %>%
  addLegend('bottomleft',pal=rh_pal, values=met_avg2$rh, title="Relative Humidity", opaci
```



Description (Template): The relative humidity map reveals a cleart. gradient across the United States. The eastern regions show... relative humidity values (...%), represented by... colors. The central regions display... humidity levels, while the western regions exhibit... values. The top 10 highest relative humidity locations are predominantly located in..., which reflects the influence of.... This pattern demonstrates how... and... factors affect regional humidity distributions.

Step 8 Use a ggplot extension

```
# cloud plot
v8 <- ggplot(
  data = met_avg %>% filter(!is.na(wind.sp)),
  aes(x = region, y = wind.sp, fill = region)
) +
  scale_fill_viridis_d(name = "") +
  ggdist::stat_halfeye(
    adjust = .5,
```

```
    width = .6,
    justification = -.2,
    .width = 0,
    point_colour = NA
  ) +
  geom_boxplot(
    width = .12,
    outlier.color = NA
  ) +
  ggdist::stat_dots(
    side = "left",
    justification = 1.1,
    binwidth = .004) +
  coord_cartesian(xlim = c(1.2, NA),)
v8+scale_fill_manual(values=c("#669900","#FF99CC","#3399FF","#6633FF"))
```

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