Cherry Blossom Prediction

Exploratory Data Analysis (EDA)

Alex Yinan Guo

Contents

Setup

Load packages and read in data sets.

```
library(GGally)
library(rjson)
library(tidyverse)
library(lubridate)
library(cowplot)
library(ggpubr)

washingtondc <- read.csv(file = '../data/raw/washingtondc.csv')
south_korea <- read.csv(file = '../data/raw/south_korea.csv')
kyoto <- read.csv(file = '../data/raw/kyoto.csv')
japan <- read.csv(file = '../data/raw/japan.csv')
liestal <- read.csv(file = '../data/raw/liestal.csv')
meteoswiss <- read.csv(file = '../data/raw/meteoswiss.csv')
USA_status_intensity <-
  read.csv(file = '../data/raw/USA-NPN_status_intensity_observations_data.csv') %>%
  filter(!State %in% c('Okinawa', 'Liaoning Sheng', 'Hyogo', 'Chongqing Shi')) %>%
  mutate(Observation_Date = mdy(Observation_Date),
         Year = format(Observation_Date, format = "%Y"),
         Date = format(Observation_Date, format = "%m-%d"))
USA_individual_phenometrics <-
```

```
  read.csv(file = '../data/raw/USA-NPN_individual_phenometrics_data.csv') %>%
  filter(!State %in% c('Okinawa', 'Liaoning Sheng', 'Hyogo', 'Chongqing Shi'))
```

Summaries of Data Sets

```
summary(washingtondc)
```

```
##    location             lat             long             alt         year
##  Length:101         Min.   :38.89   Min.   :-77.04   Min.   :0   Min.   :1921
##  Class :character   1st Qu.:38.89   1st Qu.:-77.04   1st Qu.:0   1st Qu.:1946
##  Mode  :character   Median :38.89   Median :-77.04   Median :0   Median :1971
##                     Mean   :38.89   Mean   :-77.04   Mean   :0   Mean   :1971
##                     3rd Qu.:38.89   3rd Qu.:-77.04   3rd Qu.:0   3rd Qu.:1996
##                     Max.   :38.89   Max.   :-77.04   Max.   :0   Max.   :2021
##   bloom_date          bloom_doy
##  Length:101         Min.   : 74.00
##  Class :character   1st Qu.: 89.00
##  Mode  :character   Median : 95.00
##                     Mean   : 93.64
##                     3rd Qu.: 99.00
##                     Max.   :108.00
```

```
summary(liestal)
```

```
##    location             lat             long            alt
##  Length:128         Min.   :47.48   Min.   :7.731   Min.   :350
##  Class :character   1st Qu.:47.48   1st Qu.:7.731   1st Qu.:350
##  Mode  :character   Median :47.48   Median :7.731   Median :350
##                     Mean   :47.48   Mean   :7.731   Mean   :350
##                     3rd Qu.:47.48   3rd Qu.:7.731   3rd Qu.:350
##                     Max.   :47.48   Max.   :7.731   Max.   :350
##       year       bloom_date          bloom_doy
##  Min.   :1894   Length:128         Min.   : 75.00
##  1st Qu.:1926   Class :character   1st Qu.: 92.75
##  Median :1958   Mode  :character   Median :102.00
##  Mean   :1958                      Mean   :101.23
##  3rd Qu.:1989                      3rd Qu.:108.25
##  Max.   :2021                      Max.   :124.00
```

```
summary(kyoto)
```

```
##    location             lat             long            alt          year
##  Length:833         Min.   :35.01   Min.   :135.7   Min.   :44   Min.   : 812
##  Class :character   1st Qu.:35.01   1st Qu.:135.7   1st Qu.:44   1st Qu.:1329
##  Mode  :character   Median :35.01   Median :135.7   Median :44   Median :1586
##                     Mean   :35.01   Mean   :135.7   Mean   :44   Mean   :1552
##                     3rd Qu.:35.01   3rd Qu.:135.7   3rd Qu.:44   3rd Qu.:1808
##                     Max.   :35.01   Max.   :135.7   Max.   :44   Max.   :2021
##   bloom_date          bloom_doy
##  Length:833         Min.   : 85.0
##  Class :character   1st Qu.:100.0
##  Mode  :character   Median :104.0
##                     Mean   :104.5
##                     3rd Qu.:109.0
```

```
##                     Max.   :124.0
```

```
summary(south_korea)
```

```
##    location              lat             long            alt        
##  Length:994         Min.   :33.25   Min.   :126.4   Min.   :  3.94  
##  Class :character   1st Qu.:35.17   1st Qu.:126.9   1st Qu.: 34.84  
##  Mode  :character   Median :36.10   Median :127.7   Median : 58.70  
##                     Mean   :36.08   Mean   :127.7   Mean   : 82.98  
##                     3rd Qu.:37.16   3rd Qu.:128.5   3rd Qu.: 85.67  
##                     Max.   :38.25   Max.   :130.9   Max.   :772.57  
##       year       bloom_date          bloom_doy     
##  Min.   :1980   Length:994         Min.   : 62.0  
##  1st Qu.:1984   Class :character   1st Qu.: 95.0  
##  Median :1988   Mode  :character   Median :102.0  
##  Mean   :1991                      Mean   :101.2  
##  3rd Qu.:1993                      3rd Qu.:107.0  
##  Max.   :2020                      Max.   :166.0
```

```
unique(south_korea$location)
```

```
##  [1] "South Korea/Sokcho"         "South Korea/Daegwallyeong"
##  [3] "South Korea/Chuncheon"      "South Korea/Gangneung"
##  [5] "South Korea/Seoul"          "South Korea/Incheon"
##  [7] "South Korea/Wonju"          "South Korea/Ulleungdo"
##  [9] "South Korea/Suwon"          "South Korea/Chungju"
## [11] "South Korea/Seosan"         "South Korea/Uljin"
## [13] "South Korea/Cheongju"       "South Korea/Daejeon"
## [15] "South Korea/Chupungnyeong" "South Korea/Pohang"
## [17] "South Korea/Gunsan"         "South Korea/Daegu"
## [19] "South Korea/Jeonju"         "South Korea/Ulsan"
## [21] "South Korea/Gwangju"        "South Korea/Busan"
## [23] "South Korea/Tongyeong"      "South Korea/Mokpo"
## [25] "South Korea/Yeosu"          "South Korea/Wando"
## [27] "South Korea/Jeju"           "South Korea/Seongsan"
## [29] "South Korea/Seogwipo"       "South Korea/Jinju"
## [31] "South Korea/Ganghwa"        "South Korea/Yangpyeong"
## [33] "South Korea/Icheon"         "South Korea/Inje"
## [35] "South Korea/Hongcheon"      "South Korea/Jecheon"
## [37] "South Korea/Boeun"          "South Korea/Cheonan"
## [39] "South Korea/Boryeong"       "South Korea/Buyeo"
## [41] "South Korea/Geumsan"        "South Korea/Buan"
## [43] "South Korea/Imsil"          "South Korea/Jeongeup"
## [45] "South Korea/Jangheung"      "South Korea/Haenam"
## [47] "South Korea/Goheung"        "South Korea/Yeongju"
## [49] "South Korea/Mungyeong"      "South Korea/Yeongdeok"
## [51] "South Korea/Uiseong"        "South Korea/Gumi"
## [53] "South Korea/Yeongcheon"     "South Korea/Geochang"
## [55] "South Korea/Hapcheon"       "South Korea/Miryang"
## [57] "South Korea/Sancheong"      "South Korea/Geoje"
## [59] "South Korea/Namhae"
```

```
summary(japan)
```

```
##    location              lat             long            alt        
##  Length:6573        Min.   :24.34   Min.   :123.0   Min.   :  0.50
```

```
##  Class :character   1st Qu.:34.15   1st Qu.:133.5   1st Qu.:  5.32
##  Mode  :character   Median :35.52   Median :137.0   Median : 19.20
##                     Mean   :36.22   Mean   :136.7   Mean   : 55.89
##                     3rd Qu.:38.26   3rd Qu.:140.5   3rd Qu.: 42.30
##                     Max.   :45.41   Max.   :145.6   Max.   :610.00
##       year       bloom_date          bloom_doy
##  Min.   :1953   Length:6573        Min.   :  8.0
##  1st Qu.:1970   Class :character   1st Qu.: 93.0
##  Median :1986   Mode  :character   Median : 99.0
##  Mean   :1986                      Mean   :101.1
##  3rd Qu.:2001                      3rd Qu.:111.0
##  Max.   :2021                      Max.   :160.0
```

```
unique(japan$location)
```

```
##   [1] "Japan/Wakkanai"        "Japan/Rumoi"           "Japan/Asahikawa"
##   [4] "Japan/Abashiri"        "Japan/Sapporo"         "Japan/Iwamizawa"
##   [7] "Japan/Obihiro"         "Japan/Kushiro"         "Japan/Nemuro"
##  [10] "Japan/Muroran"         "Japan/Urakawa"         "Japan/Esashi"
##  [13] "Japan/Hakodate"        "Japan/Kutchan"         "Japan/Mombetsu"
##  [16] "Japan/Hiroo"           "Japan/Shinjo"          "Japan/Aomori"
##  [19] "Japan/Hachinohe"       "Japan/Akita"           "Japan/Morioka"
##  [22] "Japan/Miyako"          "Japan/Sakata"          "Japan/Yamagata"
##  [25] "Japan/Sendai"          "Japan/Fukushima"       "Japan/Shirakawa"
##  [28] "Japan/Onahama"         "Japan/Wajima"          "Japan/Aikawa"
##  [31] "Japan/Niigata"         "Japan/Kanazawa"        "Japan/Toyama"
##  [34] "Japan/Nagano"          "Japan/Takada"          "Japan/Utsunomiya"
##  [37] "Japan/Fukui"           "Japan/Takayama"        "Japan/Matsumoto"
##  [40] "Japan/Maebashi"        "Japan/Kumagaya"        "Japan/Mito"
##  [43] "Japan/Tsuruga"         "Japan/Gifu"            "Japan/Nagoya"
##  [46] "Japan/Iida"            "Japan/Kofu"            "Japan/Choshi"
##  [49] "Japan/Tsu"             "Japan/Hamamatsu"       "Japan/Shizuoka"
##  [52] "Japan/Tokyo"           "Japan/Owase"           "Japan/Yokohama"
##  [55] "Japan/Tateyama"        "Japan/Oshima"          "Japan/Miyakejima"
##  [58] "Japan/Hachijojima"     "Japan/Saigo"           "Japan/Matsue"
##  [61] "Japan/Yonago"          "Japan/Tottori"         "Japan/Toyooka"
##  [64] "Japan/Maizuru"         "Japan/Hamada"          "Japan/Kyoto"
##  [67] "Japan/Hikone"          "Japan/Shimonoseki"     "Japan/Hiroshima"
##  [70] "Japan/Okayama"         "Japan/Kobe"            "Japan/Osaka"
##  [73] "Japan/Sumoto"          "Japan/Wakayama"        "Japan/Shionomisaki"
##  [76] "Japan/Nara"            "Japan/Izuhara"         "Japan/Fukuoka"
##  [79] "Japan/Saga"            "Japan/Oita"            "Japan/Nagasaki"
##  [82] "Japan/Kumamoto"        "Japan/Nobeoka"         "Japan/Kagoshima"
##  [85] "Japan/Miyazaki"        "Japan/Yakushima"       "Japan/Tanegashima"
##  [88] "Japan/Fukue"           "Japan/Matsuyama"       "Japan/Takamatsu"
##  [91] "Japan/Uwajima"         "Japan/Kochi"           "Japan/Tokushima"
##  [94] "Japan/Naze"            "Japan/Naze/Funchatoge" "Japan/Yonagunijima"
##  [97] "Japan/Iriomotejima"    "Japan/Ishigakijima"    "Japan/Miyakojima"
## [100] "Japan/Kumejima"        "Japan/Naha"            "Japan/Nago"
## [103] "Japan/Minamidaitojima"
```

```
summary(meteoswiss)
```

```
##    location              lat             long              alt
##  Length:6374        Min.   :45.86   Min.   : 6.014   Min.   : 200.0
```

```
##  Class :character   1st Qu.:46.60   1st Qu.: 7.410   1st Qu.: 480.0
##  Mode  :character   Median :46.98   Median : 8.444   Median : 620.0
##                     Mean   :46.94   Mean   : 8.261   Mean   : 702.7
##                     3rd Qu.:47.29   3rd Qu.: 9.072   3rd Qu.: 900.0
##                     Max.   :47.76   Max.   :10.465   Max.   :1650.0
##       year       bloom_date          bloom_doy
##  Min.   :1951   Length:6374        Min.   : 70.0
##  1st Qu.:1981   Class :character   1st Qu.:104.0
##  Median :1995   Mode  :character   Median :114.0
##  Mean   :1994                      Mean   :114.4
##  3rd Qu.:2008                      3rd Qu.:124.8
##  Max.   :2021                      Max.   :175.0
```

```
unique(meteoswiss$location)
```

```
##   [1] "Switzerland/Enges"
##   [2] "Switzerland/Fiesch"
##   [3] "Switzerland/Liddes"
##   [4] "Switzerland/Les Ponts-de-Martel"
##   [5] "Switzerland/Orvin"
##   [6] "Switzerland/Posieux"
##   [7] "Switzerland/Schiers"
##   [8] "Switzerland/Seewis Dorf"
##   [9] "Switzerland/Trient"
##  [10] "Switzerland/Vallorbe"
##  [11] "Switzerland/Zofingen"
##  [12] "Switzerland/Biel"
##  [13] "Switzerland/Flawil"
##  [14] "Switzerland/Liestal"
##  [15] "Switzerland/Murg"
##  [16] "Switzerland/Rafz"
##  [17] "Switzerland/Seon"
##  [18] "Switzerland/Versoix"
##  [19] "Switzerland/Wattwil, SG"
##  [20] "Switzerland/Wiliberg"
##  [21] "Switzerland/ZÃ¼rich-AlbisgÃ¼etli"
##  [22] "Switzerland/Andeer"
##  [23] "Switzerland/Orbe / Bochuz"
##  [24] "Switzerland/Simplon-Dorf"
##  [25] "Switzerland/Cartigny"
##  [26] "Switzerland/Couvet"
##  [27] "Switzerland/Sarnen"
##  [28] "Switzerland/Wald, ZH"
##  [29] "Switzerland/L'Abergement"
##  [30] "Switzerland/Aurigeno"
##  [31] "Switzerland/Brusio-Piazzo"
##  [32] "Switzerland/Disentis"
##  [33] "Switzerland/Escholzmatt"
##  [34] "Switzerland/Elm"
##  [35] "Switzerland/Gadmen"
##  [36] "Switzerland/Gryon"
##  [37] "Switzerland/Le Locle"
##  [38] "Switzerland/Meiringen I"
##  [39] "Switzerland/MÃ¶hlin"
##  [40] "Switzerland/Sargans II"
```

```
##  [41] "Switzerland/Thusis"
##  [42] "Switzerland/Vals"
##  [43] "Switzerland/Vergeletto"
##  [44] "Switzerland/La Valsainte"
##  [45] "Switzerland/WÃ¤denswil"
##  [46] "Switzerland/ZÃ¼rich-MeteoSchweiz"
##  [47] "Switzerland/Bauma"
##  [48] "Switzerland/Les Plans-sur-Bex"
##  [49] "Switzerland/Prato-Sornico"
##  [50] "Switzerland/Entlebuch"
##  [51] "Switzerland/Zweisimmen"
##  [52] "Switzerland/Longirod"
##  [53] "Switzerland/Merishausen"
##  [54] "Switzerland/Sihlbrugg"
##  [55] "Switzerland/Chur"
##  [56] "Switzerland/Oeschberg"
##  [57] "Switzerland/Grossdietwil"
##  [58] "Switzerland/Moutier"
##  [59] "Switzerland/Davos-Dorf"
##  [60] "Switzerland/Adelboden"
##  [61] "Switzerland/Changins"
##  [62] "Switzerland/Basel-Binningen"
##  [63] "Switzerland/Cevio-Cavergno"
##  [64] "Switzerland/Horgen"
##  [65] "Switzerland/Locarno"
##  [66] "Switzerland/Moudon"
##  [67] "Switzerland/Winterthur"
##  [68] "Switzerland/ZÃ¼rich-Witikon"
##  [69] "Switzerland/Les Diablerets"
##  [70] "Switzerland/Leysin"
##  [71] "Switzerland/Cernier"
##  [72] "Switzerland/Herzogenbuchsee"
##  [73] "Switzerland/Appenzell"
##  [74] "Switzerland/Bondo GR"
##  [75] "Switzerland/Domat / Ems"
##  [76] "Switzerland/Fanas"
##  [77] "Switzerland/GrÃ¼sch"
##  [78] "Switzerland/Jenaz"
##  [79] "Switzerland/Laufenburg"
##  [80] "Switzerland/Linthal"
##  [81] "Switzerland/NÃ¤fels"
##  [82] "Switzerland/SchÃ¶nenwerd"
##  [83] "Switzerland/Sent"
##  [84] "Switzerland/St. Luc"
##  [85] "Switzerland/Stampa"
##  [86] "Switzerland/Villnachern"
##  [87] "Switzerland/Wynau"
##  [88] "Switzerland/Diessenhofen"
##  [89] "Switzerland/Edlibach"
##  [90] "Switzerland/Zizers"
##  [91] "Switzerland/Eschen-Boja"
##  [92] "Switzerland/Neuhausen"
##  [93] "Switzerland/DÃ¶ttingen"
##  [94] "Switzerland/Einsiedeln"
```

```
##  [95] "Switzerland/Heiden"
##  [96] "Switzerland/Luzern"
##  [97] "Switzerland/Muri, AG"
##  [98] "Switzerland/Oberehrendingen"
##  [99] "Switzerland/Vira / Gambarogno"
## [100] "Switzerland/Altdorf"
## [101] "Switzerland/Leytron"
## [102] "Switzerland/Martina"
## [103] "Switzerland/Morgins VS"
## [104] "Switzerland/Osterfingen"
## [105] "Switzerland/Rorschach"
## [106] "Switzerland/Oberlangenegg"
## [107] "Switzerland/Olivone"
## [108] "Switzerland/Les Rangiers"
## [109] "Switzerland/Romanshorn"
## [110] "Switzerland/Hallau"
## [111] "Switzerland/Stein, AR II"
## [112] "Switzerland/Azmoos"
## [113] "Switzerland/Chardonne"
## [114] "Switzerland/Birmensdorf"
## [115] "Switzerland/Worb BE"
## [116] "Switzerland/Gundetswil"
## [117] "Switzerland/Scuol"
## [118] "Switzerland/Sagno TI"
## [119] "Switzerland/Blonay"
## [120] "Switzerland/Estavayer-le-Lac"
## [121] "Switzerland/Comprovasco / Motto"
## [122] "Switzerland/Bellelay"
## [123] "Switzerland/Locarno-Monti"
## [124] "Switzerland/Frauenfeld"
## [125] "Switzerland/Hard b. Weinfelden"
## [126] "Switzerland/HÃ¶fen"
## [127] "Switzerland/Lanterswil"
## [128] "Switzerland/Raperswilen, TG"
## [129] "Switzerland/Silenen"
## [130] "Switzerland/Visp"
## [131] "Switzerland/Wil, SG"
## [132] "Switzerland/Wolhusen"
## [133] "Switzerland/Wengen"
## [134] "Switzerland/Wetzikon"
## [135] "Switzerland/Chaumont"
## [136] "Switzerland/Le SÃ©pey"
## [137] "Switzerland/La BrÃ©vine"
## [138] "Switzerland/Dornach"
## [139] "Switzerland/Sta. Maria (Val Mustair)"
## [140] "Switzerland/Buchs"
## [141] "Switzerland/Wildhaus"
## [142] "Switzerland/Boudry"
## [143] "Switzerland/Grellingen"
## [144] "Switzerland/Gstaad"
## [145] "Switzerland/Therwil"
## [146] "Switzerland/Hochdorf"
## [147] "Switzerland/Seewil"
## [148] "Switzerland/Unterseen"
```

```
## [149] "Switzerland/Jegenstorf II"
## [150] "Switzerland/Rancate"
## [151] "Switzerland/AlchenflÃ¼h "
## [152] "Switzerland/Langnau i.E."
## [153] "Switzerland/NeuchÃ¢tel"
## [154] "Switzerland/Ballens"
## [155] "Switzerland/Osogna"
## [156] "Switzerland/Ried / Frutigen"
## [157] "Switzerland/Faido"
## [158] "Switzerland/Payerne"
## [159] "Switzerland/Reinach, BL"
## [160] "Switzerland/Mugena"
## [161] "Switzerland/Echandens"
```

```
summary(USA_status_intensity)
```

```
##  Observation_ID     Update_Datetime       Site_ID         Latitude
##  Min.   :  137803   Length:3270        Min.   :    3   Min.   :35.18
##  1st Qu.: 5996755   Class :character   1st Qu.: 8901   1st Qu.:37.99
##  Median :13342592   Mode  :character   Median :19713   Median :39.73
##  Mean   :13826421                      Mean   :21202   Mean   :39.84
##  3rd Qu.:21595651                      3rd Qu.:32789   3rd Qu.:40.95
##  Max.   :28116281                      Max.   :45262   Max.   :53.93
##    Longitude       Elevation_in_Meters    State              Species_ID
##  Min.   :-123.33   Min.   :   1.0      Length:3270        Min.   :  87.0
##  1st Qu.: -84.40   1st Qu.:  26.0      Class :character   1st Qu.: 227.0
##  Median : -78.74   Median : 136.0      Mode  :character   Median : 228.0
##  Mean   : -86.37   Mean   : 597.2                         Mean   : 503.3
##  3rd Qu.: -74.00   3rd Qu.:1452.0                         3rd Qu.:1189.0
##  Max.   : -71.05   Max.   :2017.0                         Max.   :1753.0
##     Genus             Species          Common_Name          Kingdom
##  Length:3270        Length:3270        Length:3270        Length:3270
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Individual_ID    Phenophase_ID Phenophase_Description Observation_Date
##  Min.   :  4385   Min.   :501   Length:3270            Min.   :2009-03-29
##  1st Qu.: 48481   1st Qu.:501   Class :character       1st Qu.:2015-05-18
##  Median : 82589   Median :501   Mode  :character       Median :2018-03-22
##  Mean   :116063   Mean   :501                          Mean   :2017-09-24
##  3rd Qu.:183138   3rd Qu.:501                          3rd Qu.:2020-02-26
##  Max.   :276655   Max.   :501                          Max.   :2021-12-31
##   Day_of_Year  Phenophase_Status Intensity_Category_ID Intensity_Value
##  Min.   :  8   Min.   :-1.0000   Min.   :-9999         Length:3270
##  1st Qu.: 90   1st Qu.: 0.0000   1st Qu.:   50         Class :character
##  Median :131   Median : 0.0000   Median :   50         Mode  :character
##  Mean   :157   Mean   : 0.1667   Mean   :-1318
##  3rd Qu.:229   3rd Qu.: 0.0000   3rd Qu.:   50
##  Max.   :365   Max.   : 1.0000   Max.   :   50
##  Abundance_Value      AGDD           AGDD_in_F        Tmax_Winter
##  Min.   :-9999   Min.   :-9999.0   Min.   :-9999.0   Min.   :-9999.00
##  1st Qu.:-9999   1st Qu.:  242.9   1st Qu.:  437.1   1st Qu.:    2.85
##  Median :-9999   Median :  552.1   Median :  993.8   Median :    5.89
```

```
##  Mean   :-9999   Mean   : -232.8   Mean   :  718.4   Mean   :-1416.78
##  3rd Qu.:-9999   3rd Qu.: 1896.5   3rd Qu.: 3413.7   3rd Qu.:    7.77
##  Max.   :-9999   Max.   : 5724.2   Max.   :10303.6   Max.   :   16.54
##   Tmax_Spring        Tmin_Winter        Tmin_Spring         Prcp_Winter
##  Min.   :-9999.00   Min.   :-9999.00   Min.   :-9999.000   Min.   :-9999.0
##  1st Qu.:   11.99   1st Qu.:   -7.14   1st Qu.:    0.880   1st Qu.:  168.0
##  Median :   15.39   Median :   -2.94   Median :    3.935   Median :  288.4
##  Mean   :-1408.45   Mean   :-1424.89   Mean   :-1418.067   Mean   :-1139.1
##  3rd Qu.:   17.79   3rd Qu.:   -1.79   3rd Qu.:    6.780   3rd Qu.:  370.0
##  Max.   :   23.25   Max.   :    7.51   Max.   :   10.380   Max.   :  902.0
##   Prcp_Spring        Daylength         Year               Date
##  Min.   :-9999.0   Min.   :-9999   Length:3270        Length:3270
##  1st Qu.:  180.0   1st Qu.:38362   Class :character   Class :character
##  Median :  305.0   Median :43950   Mode  :character   Mode  :character
##  Mean   :-1127.8   Mean   :37341
##  3rd Qu.:  406.7   3rd Qu.:49075
##  Max.   :  779.0   Max.   :57370
```

```
summary(USA_individual_phenometrics)
```

```
##     Site_ID         Latitude       Longitude       Elevation_in_Meters
##  Min.   :    3   Min.   :35.18   Min.   :-123.33   Min.   :   1.0
##  1st Qu.: 9044   1st Qu.:38.91   1st Qu.:-104.96   1st Qu.:  31.0
##  Median :28788   Median :39.10   Median : -79.01   Median :  61.5
##  Mean   :23710   Mean   :40.40   Mean   : -89.52   Mean   : 451.0
##  3rd Qu.:30309   3rd Qu.:41.71   3rd Qu.: -76.94   3rd Qu.: 458.8
##  Max.   :44322   Max.   :53.93   Max.   : -71.05   Max.   :2017.0
##     State              Species_ID        Genus              Species
##  Length:204         Min.   :  87.0   Length:204         Length:204
##  Class :character   1st Qu.: 227.0   Class :character   Class :character
##  Mode  :character   Median : 228.0   Mode  :character   Mode  :character
##                     Mean   : 439.7
##                     3rd Qu.: 228.0
##                     Max.   :1753.0
##  Common_Name          Kingdom           Individual_ID    Phenophase_ID
##  Length:204         Length:204         Min.   :  4616   Min.   :501
##  Class :character   Class :character   1st Qu.: 49730   1st Qu.:501
##  Mode  :character   Mode  :character   Median :151437   Median :501
##                                        Mean   :129404   Mean   :501
##                                        3rd Qu.:162213   3rd Qu.:501
##                                        Max.   :270388   Max.   :501
##  Phenophase_Description First_Yes_Year First_Yes_Month  First_Yes_Day
##  Length:204             Min.   :2009   Min.   : 2.000   Min.   : 1.00
##  Class :character       1st Qu.:2016   1st Qu.: 4.000   1st Qu.: 8.00
##  Mode  :character       Median :2018   Median : 4.000   Median :16.00
##                         Mean   :2017   Mean   : 4.186   Mean   :15.31
##                         3rd Qu.:2019   3rd Qu.: 5.000   3rd Qu.:23.00
##                         Max.   :2021   Max.   :10.000   Max.   :30.00
##  First_Yes_DOY   First_Yes_Julian_Date NumDays_Since_Prior_No Last_Yes_Year
##  Min.   : 58.0   Min.   :2454920       Min.   :-9999          Min.   :2009
##  1st Qu.: 93.0   1st Qu.:2457494       1st Qu.:-9999          1st Qu.:2016
##  Median :104.0   Median :2458213       Median :    6          Median :2018
##  Mean   :110.9   Mean   :2458038       Mean   :-2738          Mean   :2017
##  3rd Qu.:135.0   3rd Qu.:2458628       3rd Qu.:    7          3rd Qu.:2019
##  Max.   :279.0   Max.   :2459363       Max.   :   51          Max.   :2021
```

```
##  Last_Yes_Month    Last_Yes_Day    Last_Yes_DOY   Last_Yes_Julian_Date
##  Min.   : 3.000   Min.   : 1.00   Min.   : 73.0   Min.   :2454920
##  1st Qu.: 4.000   1st Qu.: 9.00   1st Qu.:104.0   1st Qu.:2457499
##  Median : 4.000   Median :17.00   Median :114.0   Median :2458230
##  Mean   : 4.471   Mean   :16.24   Mean   :120.5   Mean   :2458048
##  3rd Qu.: 5.000   3rd Qu.:24.00   3rd Qu.:138.0   3rd Qu.:2458636
##  Max.   :11.000   Max.   :31.00   Max.   :317.0   Max.   :2459369
##  NumDays_Until_Next_No      AGDD           AGDD_in_F        Tmax_Winter
##  Min.   :-9999         Min.   :-9999.0   Min.   :-9999.0   Min.   :-9999.00
##  1st Qu.:-9999         1st Qu.:  376.8   1st Qu.:  678.3   1st Qu.:    2.67
##  Median :    3         Median :  479.0   Median :  862.2   Median :    7.18
##  Mean   :-4552         Mean   : -859.8   Mean   : -449.8   Mean   :-1367.20
##  3rd Qu.:    7         3rd Qu.:  649.2   3rd Qu.: 1168.5   3rd Qu.:    7.78
##  Max.   :  141         Max.   : 3518.8   Max.   : 6333.8   Max.   :   15.31
##   Tmax_Spring        Tmin_Winter         Tmin_Spring         Prcp_Winter
##  Min.   :-9999.00   Min.   :-9999.000   Min.   :-9999.000   Min.   :-9999
##  1st Qu.:   12.12   1st Qu.:   -7.140   1st Qu.:    1.413   1st Qu.:  168
##  Median :   15.85   Median :   -2.695   Median :    5.420   Median :  211
##  Mean   :-1358.48   Mean   :-1375.386   Mean   :-1368.197   Mean   :-1121
##  3rd Qu.:   17.97   3rd Qu.:   -1.668   3rd Qu.:    7.240   3rd Qu.:  342
##  Max.   :   23.25   Max.   :    6.630   Max.   :   10.130   Max.   :  902
##   Prcp_Spring
##  Min.   :-9999.0
##  1st Qu.:  175.8
##  Median :  306.0
##  Mean   :-1087.6
##  3rd Qu.:  370.0
##  Max.   :  686.0
```

```
# Species in the U.S.
unique(USA_individual_phenometrics%>%select(Common_Name))
```

```
##                    Common_Name
## 1               Yoshino cherry
## 2                bitter cherry
## 10  Japanese flowering cherry
## 22                  pin cherry
## 155                sour cherry
```

```
# Species in Washington, D.C.
unique(USA_individual_phenometrics%>%filter(State=='DC')%>%select(Common_Name))
```

```
##                  Common_Name
## 1             Yoshino cherry
## 4 Japanese flowering cherry
```

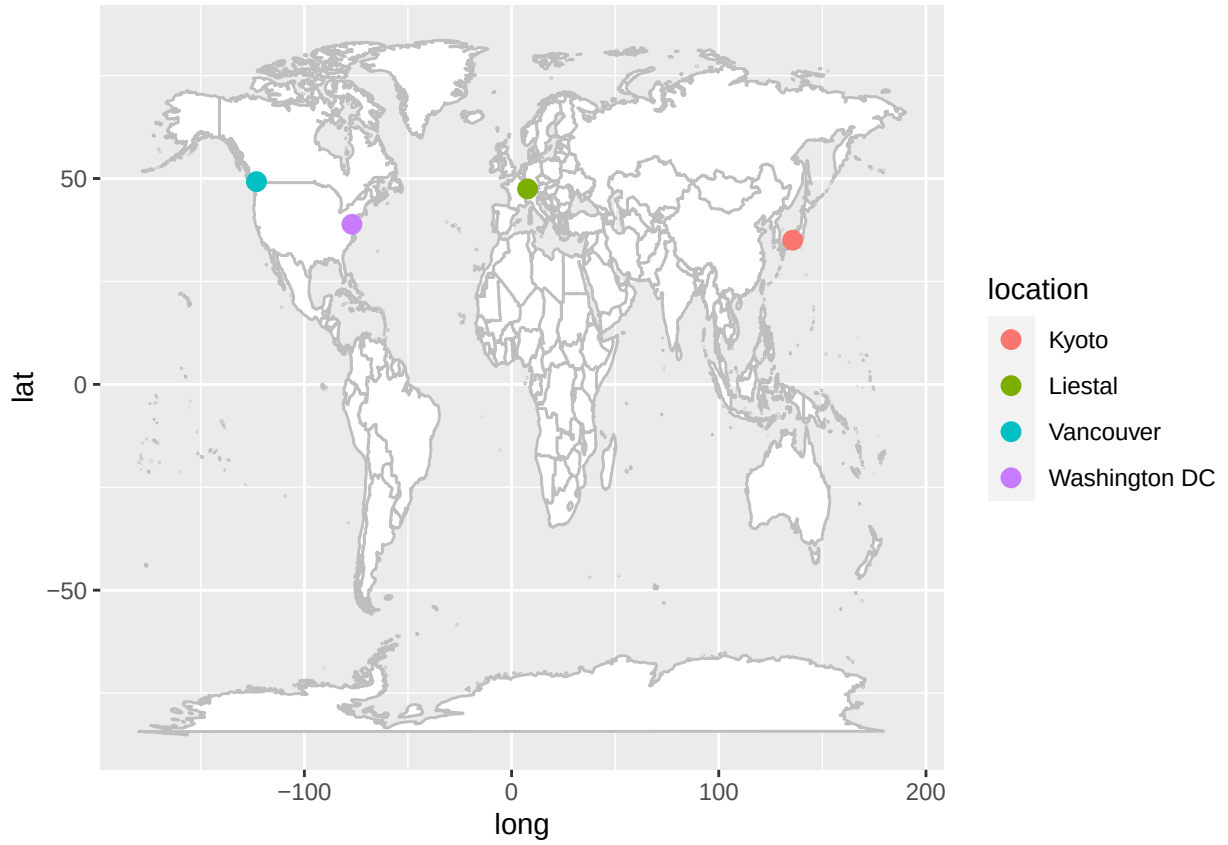
Exploration of City Data Sets

For this competition, we are going to forecast the cherry blossom peak of the following four cities:

- Washington, D.C. (USA)
- Kyoto (Japan)
- Liestal-Weideli (Switzerland)
- Vancouver, BC (Canada)

Locations of Cities

```
world <- map_data("world")
cities <- rbind(kyoto[1,],washingtondc[1,],liestal[1,]) %>%
  select(lat, long) %>%
  add_row(lat=49.246292, long=-123.116226) %>%
  add_column(location=c('Kyoto','Washington DC', 'Liestal', 'Vancouver'))

ggplot() +
  geom_map(
    data = world,
    map = world,
    aes(long, lat, map_id = region),
    color = "gray", fill = "white") +
  geom_point(
    data = cities,
    aes(long, lat, color=location), size=3)
```

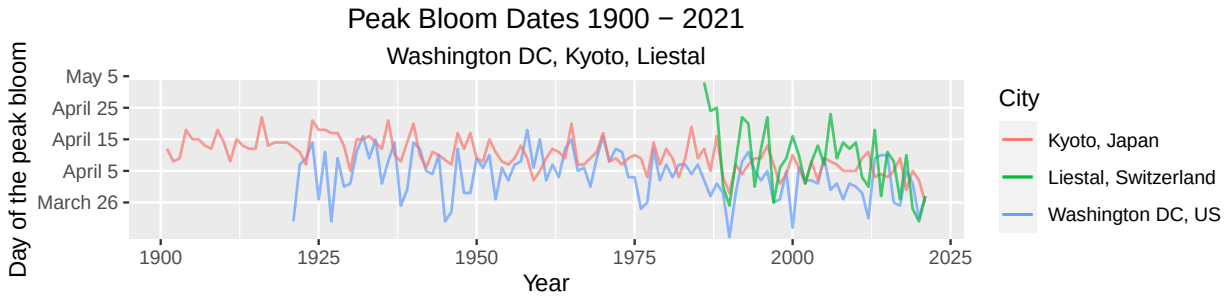


Time Series Visualization of Peak Bloom Dates

We use the provided historical bloom data of three cities where the cherry blossom peak bloom dates are going to be predicted to plot time series.

```
ggplot() +
  geom_line(data = washingtondc,
            aes(year, bloom_doy, color='Washington DC, US'),
```

```
            alpha=0.7,
            size=0.6) +
  geom_line(data = kyoto %>% filter(year>1900),
            aes(year, bloom_doy, color='Kyoto, Japan'),
            alpha=0.7,
            size=0.6) +
  geom_line(data = meteoswiss %>%
              filter(location=='Switzerland/Liestal', year>1980),
            aes(year, bloom_doy, color='Liestal, Switzerland'),
            alpha=0.7,
            size=0.6) +
  labs(title="Peak Bloom Dates 1900 - 2021",
       subtitle = "Washington DC, Kyoto, Liestal",
       x ="Year",
       y = "Day of the peak bloom",
       color = "City") +
  scale_y_discrete(limit = seq(85, 125, 10),
                   labels = c('March 26',
                              'April 5',
                              'April 15',
                              'April 25',
                              'May 5')) +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



Exploration of Country Data Sets

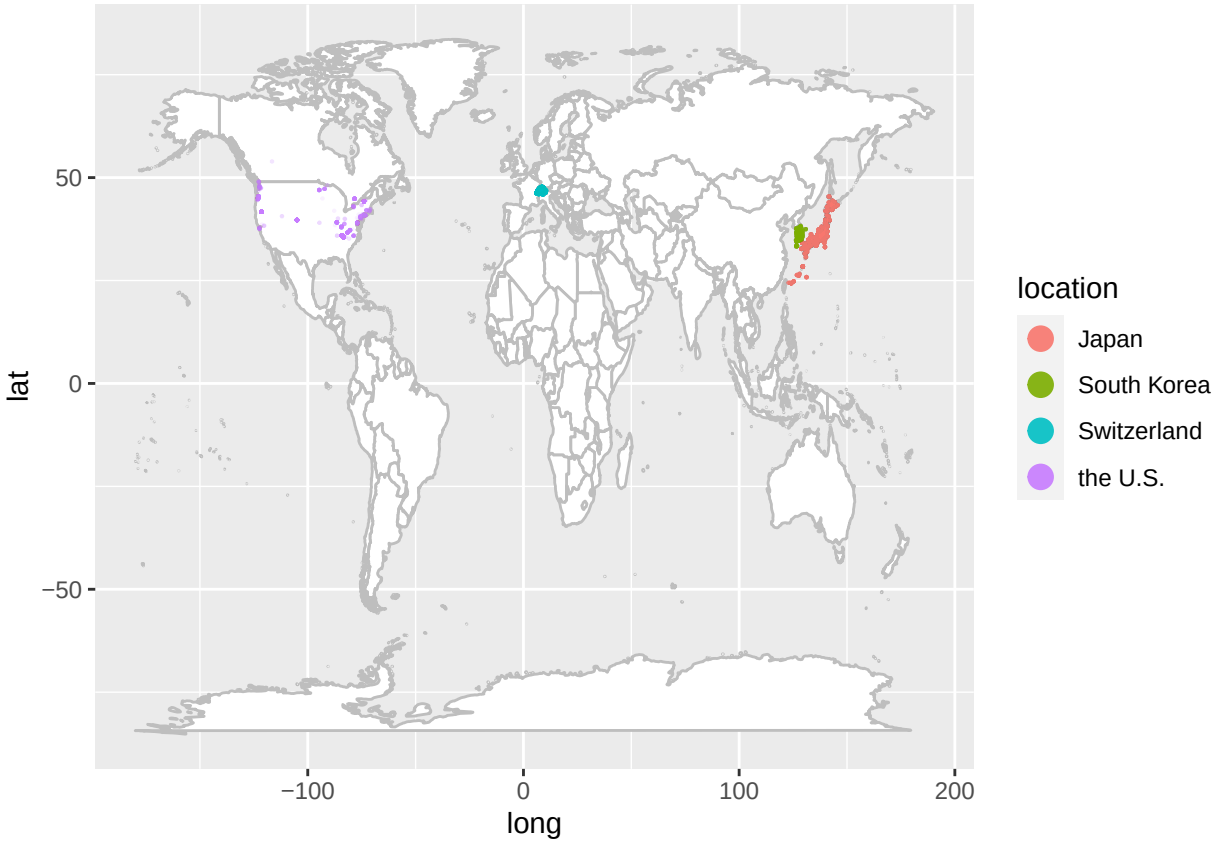
Here we explore more about the following data sets containing information regarding cherry blossom of different locations in four countries:

- south_korea.csv
- japan.csv
- meteoswiss.csv
- USA-NPN_status_intensity_observations_data.csv

Visualization of Data Locations

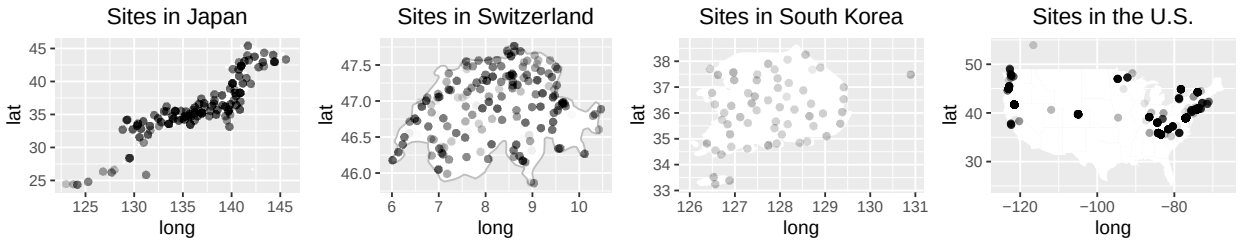
```
ggplot() +
  geom_map(
    data = world, map = world,
    aes(long, lat, map_id = region, fill = region),
    color = "gray", fill = "white") +
```

```
  geom_point(
    data = rbind(japan %>% mutate(location=sub("/.*", "",location)) %>%
                   select(location,long,lat),
                 meteoswiss %>% mutate(location=sub("/.*", "",location)) %>%
                   select(location,long,lat),
                 south_korea %>% mutate(location=sub("/.*", "",location)) %>%
                   select(location,long,lat),
                 USA_status_intensity %>%
                   select(Longitude, Latitude) %>%
                   mutate(location='the U.S.') %>%
                   rename(long=Longitude, lat=Latitude)),
    aes(long, lat, color=location),
    alpha=0.1, size=0.1) +
  guides(colour = guide_legend(override.aes = list(alpha=0.9,size=4)))
```



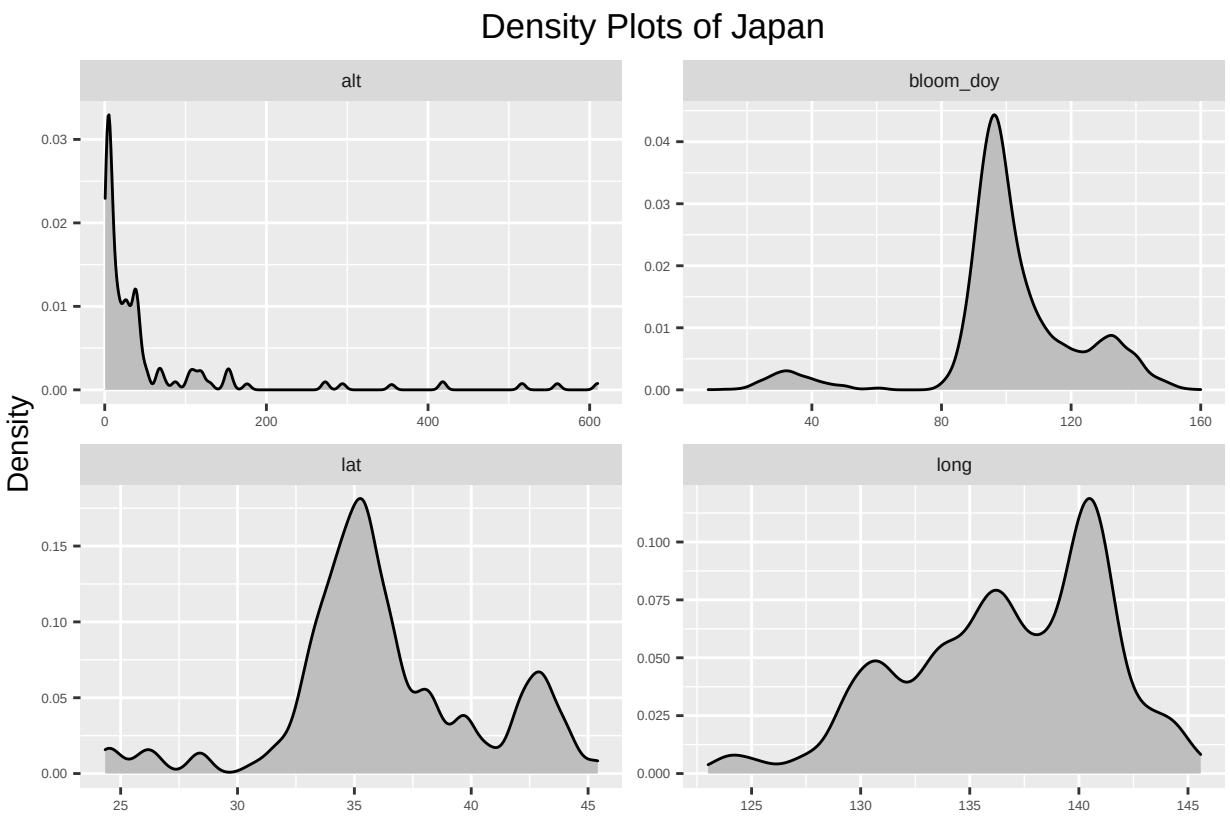

```
jp <- map_data('world', 'Japan')
jp_plot <- ggplot() +
  geom_map(
    data = jp, map = jp,
    aes(long, lat, map_id = region),
    color = "white", fill = "white"
  ) +
  geom_point(
    data = japan,
    aes(long, lat), alpha=0.01) +
  ggtitle("Sites in Japan") +
```

```
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

swiss <- map_data('world', 'Switzerland')
swiss_plot <- ggplot() +
  geom_map(
    data = swiss, map = swiss,
    aes(long, lat, map_id = region),
    color = "gray", fill = "white"
  ) +
  geom_point(
    data = meteoswiss,
    aes(long, lat), alpha=0.01) +
  ggtitle("Sites in Switzerland") +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

sk <- map_data('world', 'South Korea')
sk_plot <- ggplot() +
  geom_map(
    data = sk, map = sk,
    aes(long, lat, map_id = region),
    color = "white", fill = "white"
  ) +
  geom_point(
    data = south_korea,
    aes(long, lat), alpha=0.01) +
  ggtitle("Sites in South Korea") +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

usa <- map_data("state")
usa_plot <- ggplot() +
  geom_map(
    data = usa, map = usa,
    aes(long, lat, map_id = region),
    fill = "white", size = 0.1
  ) +
  geom_point(
    data = USA_status_intensity,
    aes(Longitude, Latitude), alpha=0.1) +
  ggtitle("Sites in the U.S.") +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

ggarrange(jp_plot, swiss_plot, sk_plot, usa_plot,
          ncol = 4, nrow = 1)
```

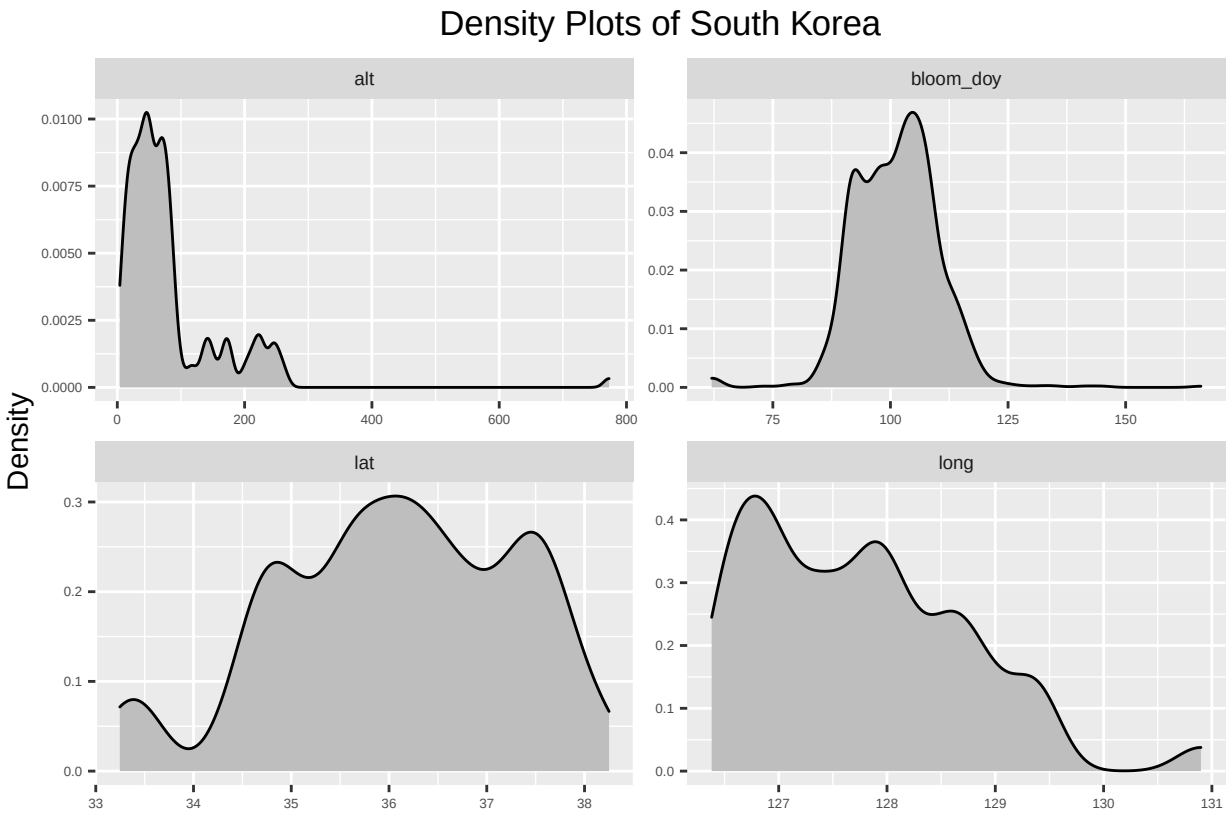


Density Plots

```
japan %>%
  select_if(is.numeric) %>%
  select(-year) %>%
  pivot_longer(everything()) %>%
  ggplot(aes(x=value)) +
  geom_density(fill='grey') +
  facet_wrap(~name, scales='free') +
  theme(strip.text = element_text(size=7),
        axis.text.x = element_text(size=5),
        axis.text.y = element_text(size=5),
        plot.title = element_text(hjust = 0.5)) +
  labs(x ="",
       y = "Density",
       title = 'Density Plots of Japan')
```

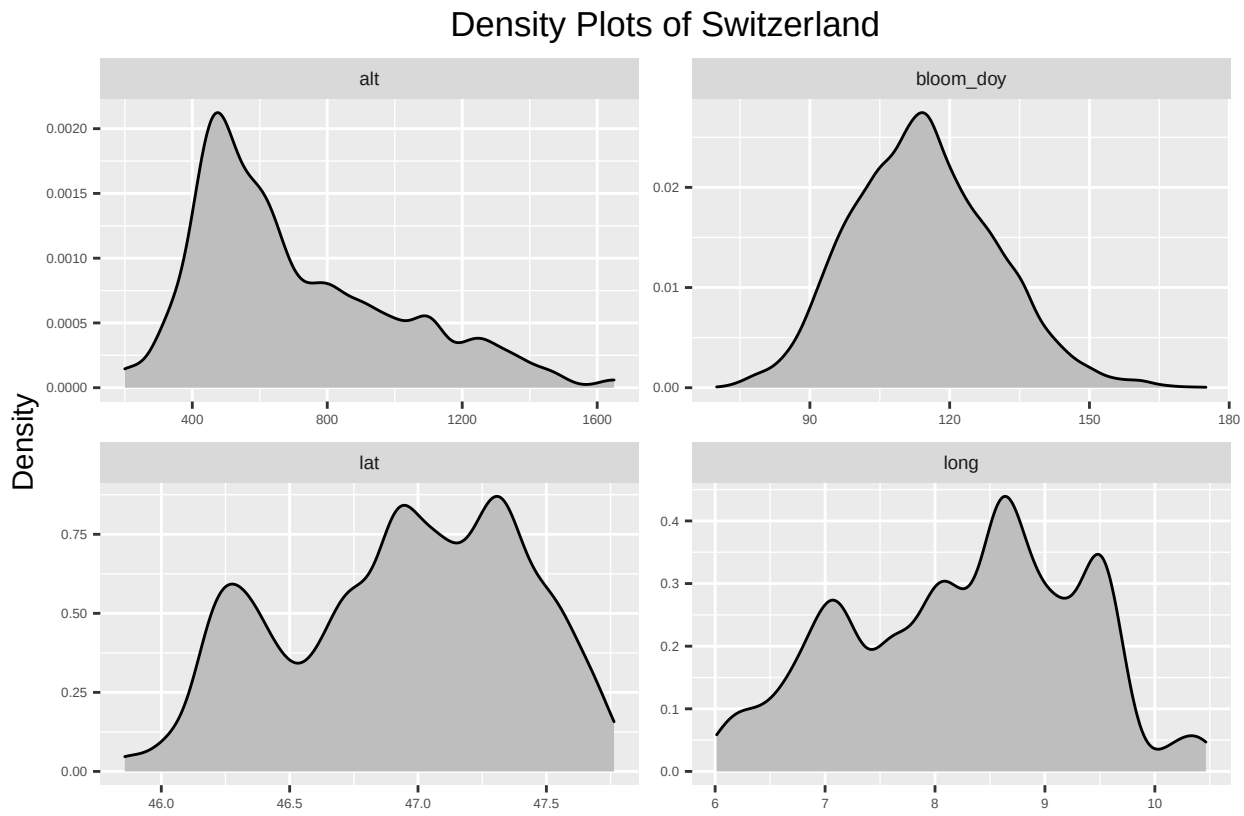


```
south_korea %>%
  select_if(is.numeric) %>%
  select(-year) %>%
  pivot_longer(everything()) %>%
  ggplot(aes(x=value)) +
  geom_density(fill='grey') +
  facet_wrap(~name, scales='free') +
  theme(strip.text = element_text(size=7),
        axis.text.x = element_text(size=5),
        axis.text.y = element_text(size=5),
        plot.title = element_text(hjust = 0.5)) +
  labs(x ="",
       y = "Density",
       title = 'Density Plots of South Korea')
```



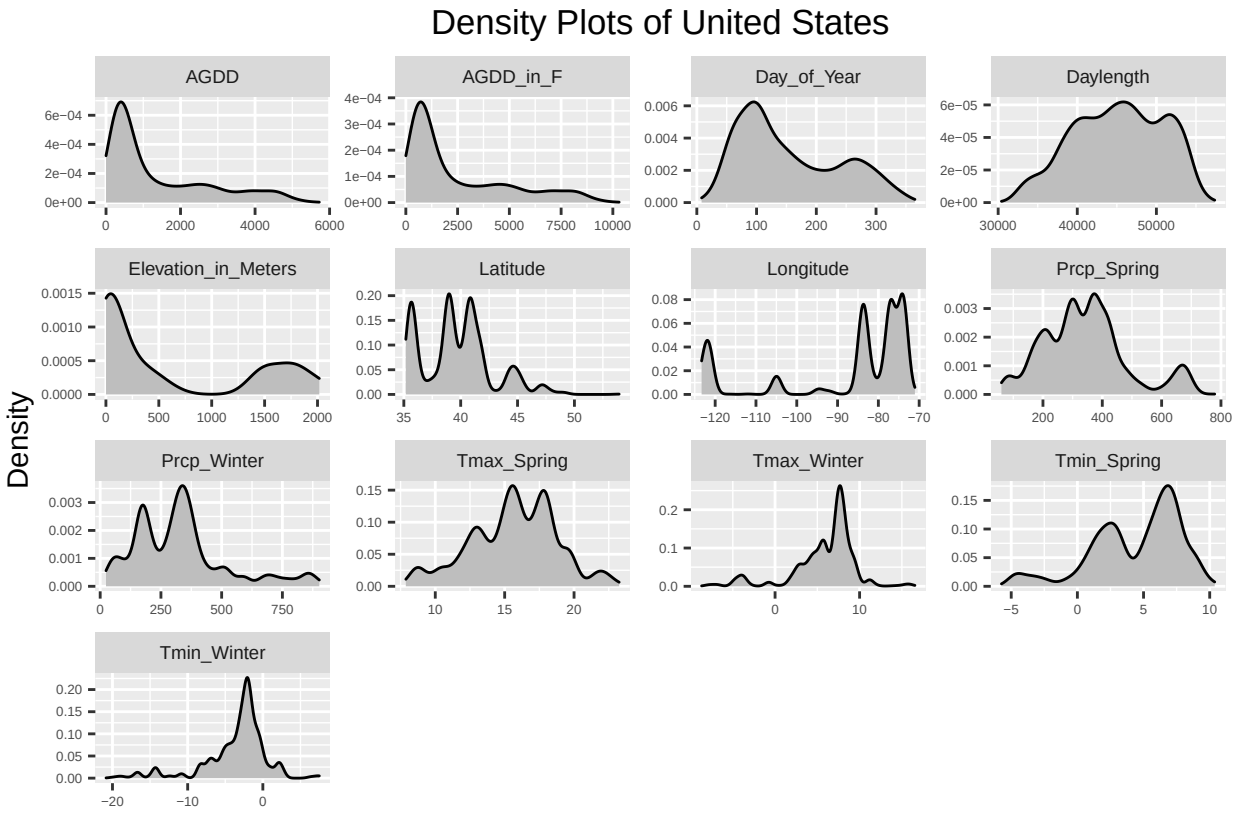

```
meteoswiss %>%
  select_if(is.numeric) %>%
  select(-year) %>%
  pivot_longer(everything()) %>%
  ggplot(aes(x=value)) +
  geom_density(fill='grey') +
  facet_wrap(~name, scales='free') +
  theme(strip.text = element_text(size=7),
        axis.text.x = element_text(size=5),
        axis.text.y = element_text(size=5),
        plot.title = element_text(hjust = 0.5)) +
```

```
  labs(x ="",
       y = "Density",
       title = 'Density Plots of Switzerland')
```




```
USA_status_intensity %>%
  select_if(is.numeric) %>%
  select(-c(Abundance_Value,
            Individual_ID,
            Intensity_Category_ID,
            Observation_ID,
            Phenophase_ID,
            Phenophase_Status,
            Site_ID,
            Species_ID)) %>%
  filter(Prcp_Spring>0,
         Prcp_Winter>0,
         Tmax_Spring != '-9999',
         Tmax_Winter != '-9999',) %>%
  pivot_longer(everything()) %>%
  ggplot(aes(x=value)) +
  geom_density(fill='grey') +
  facet_wrap(~name, scales='free') +
  theme(strip.text = element_text(size=7),
        axis.text.x = element_text(size=5),
        axis.text.y = element_text(size=5),
        plot.title = element_text(hjust = 0.5)) +
```

```
  labs(x ="",
       y = "Density",
       title = 'Density Plots of United States')
```

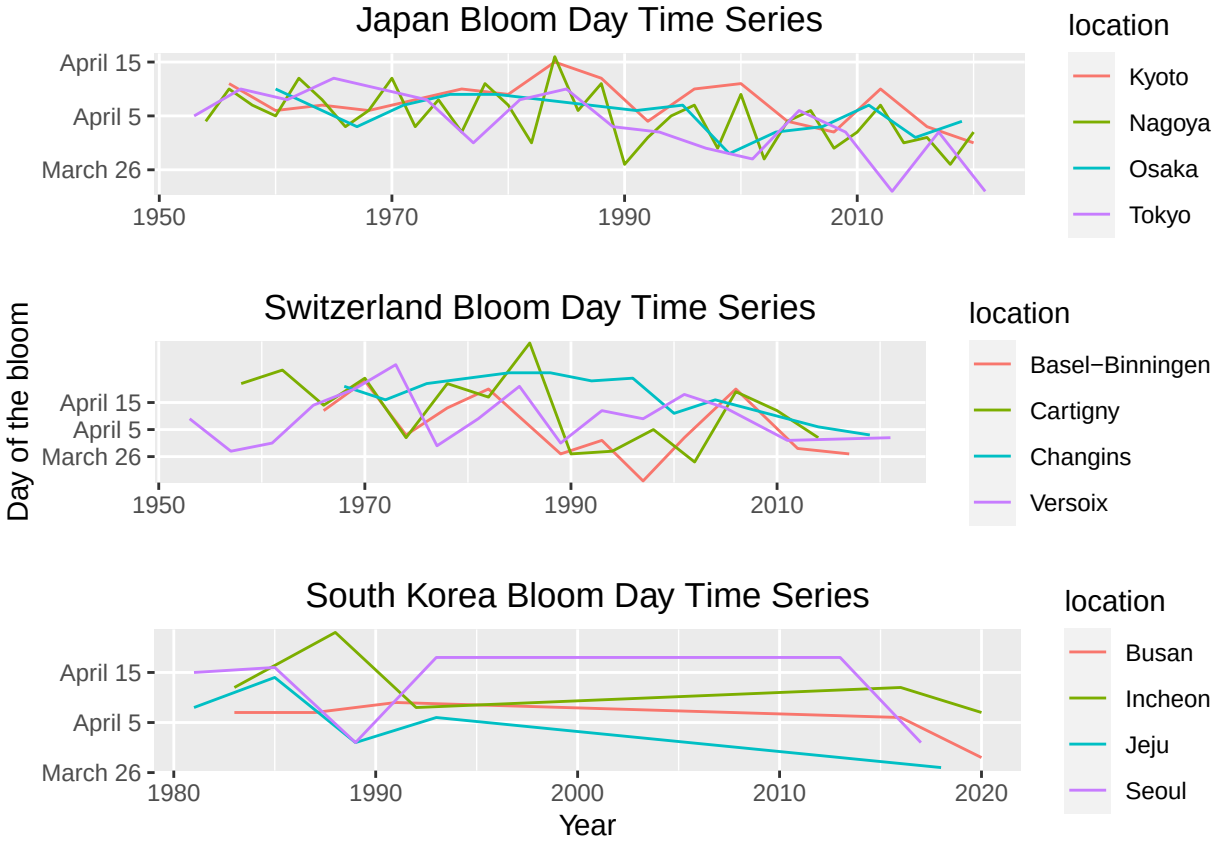


Time Series Visualization of Cities

Here we randomly selected some cities in each data set to plot the peak bloom date against year. We only create line charts for cities in Japan, Switzerland, and South Korea. The number of usable observations are too low in the U.S. data set, so we will not visualize it.

```
japan_plot <- japan %>% mutate(location=gsub("Japan/","",location)) %>%
  filter(location==c('Nagoya', 'Tokyo', 'Kyoto', 'Osaka'))%>%
  ggplot(aes(year, bloom_doy)) +
  aes(x = year, color=location) +
  geom_line() +
  labs(title="Japan Bloom Day Time Series",
       x =" ", y = " ") +
  scale_y_discrete(limit = seq(85, 105, 10),
                   labels = c('March 26', 'April 5', 'April 15')) +
  theme(plot.title = element_text(hjust = 0.5))

meteoswiss_plot <- meteoswiss %>%
  mutate(location=gsub("Switzerland/","",location)) %>%
  arrange(location) %>%
  filter(location==c('Changins', 'Basel-Binningen', 'Cartigny', 'Versoix')) %>%
  ggplot(aes(year, bloom_doy)) +
```

```
  aes(x = year, color=location) +
  geom_line() +
  labs(title="Switzerland Bloom Day Time Series",
       x =" ", y = "Day of the bloom") +
  scale_y_discrete(limit = seq(85, 105, 10),
                   labels = c('March 26', 'April 5', 'April 15')) +
  theme(plot.title = element_text(hjust = 0.5))

south_korea_plot <- south_korea %>%
  mutate(location=gsub("South Korea/","",location)) %>%
  arrange(location) %>%
  filter(location==c('Seoul', 'Incheon', 'Busan', 'Jeju')) %>%
ggplot(aes(year, bloom_doy)) +
    aes(x = year, color=location) +
    geom_line(na.rm = FALSE) +
  labs(title="South Korea Bloom Day Time Series",
       x ="Year", y = " ") +
  scale_y_discrete(limit = seq(85, 105, 10),
                   labels = c('March 26', 'April 5', 'April 15')) +
  theme(plot.title = element_text(hjust = 0.5))

ggarrange(japan_plot, meteoswiss_plot,south_korea_plot,
          ncol = 1, nrow = 3)
```

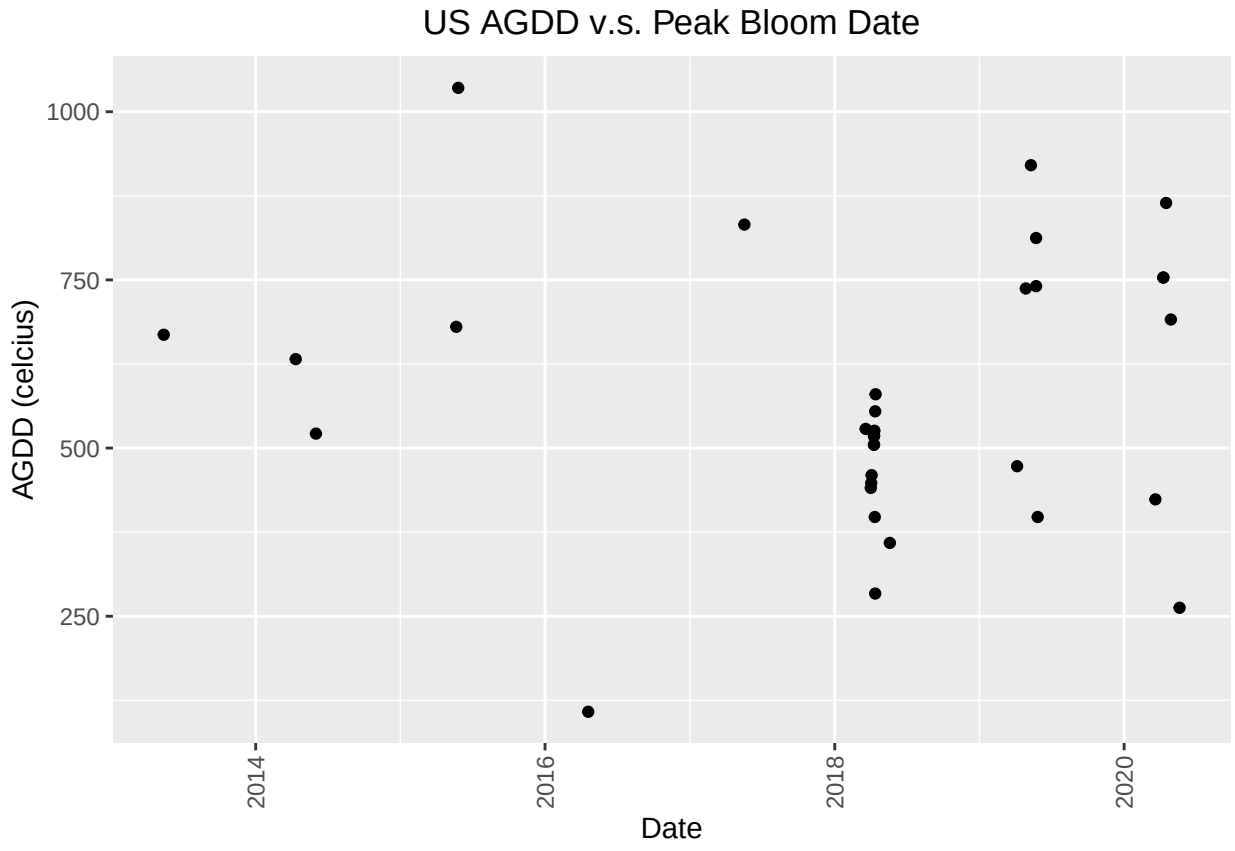


Wxploration of USA NPN Data Sets

Relationships Between AGDD and Peak Bloom Dates

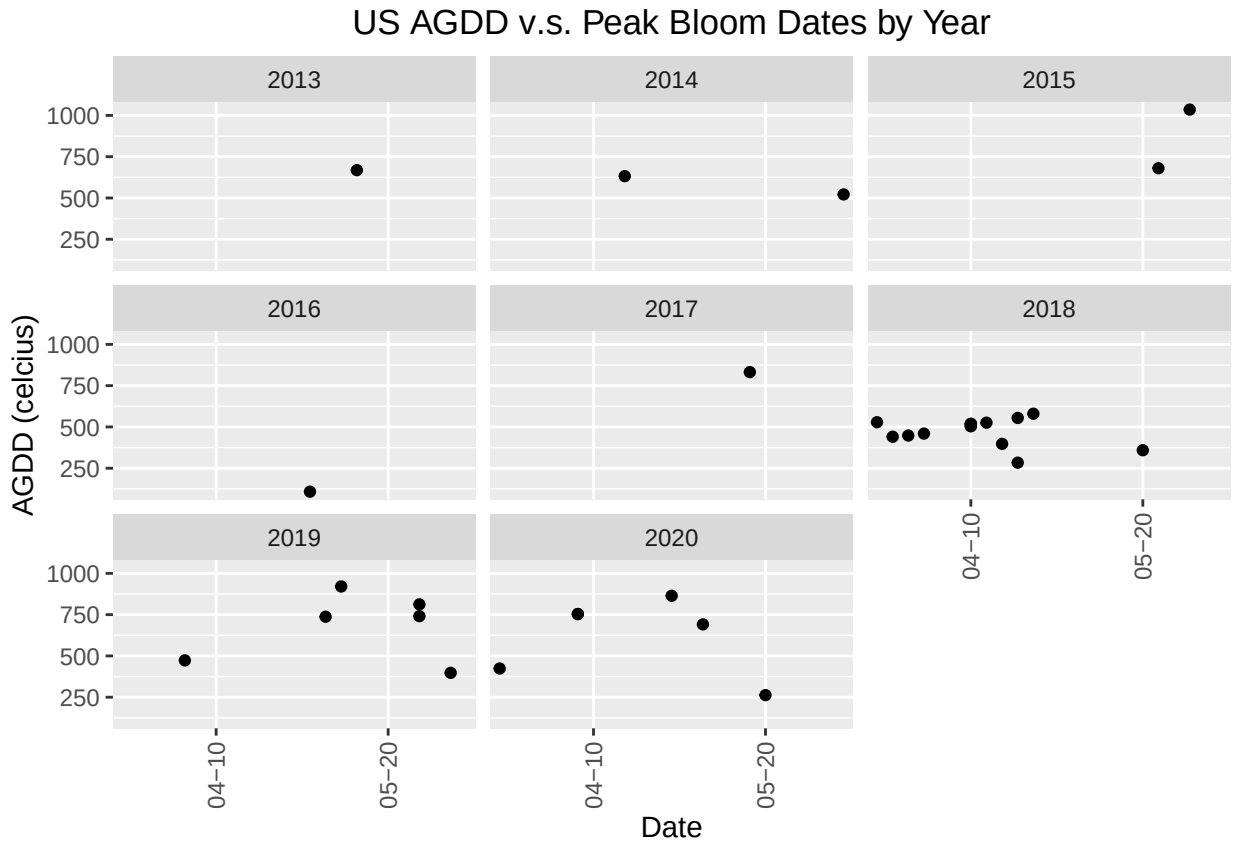
Before visualization, we filter the data by intensity value over 75% because we want to explore more about relationships between peak bloom date and other variables.

```
USA_status_intensity <- USA_status_intensity %>%
  filter(Intensity_Value=='75-94%') %>%
  group_by(Year, Site_ID) %>%
  slice(which.min(Day_of_Year))

USA_status_intensity %>%
filter(AGDD > 0) %>%
  ggplot() +
  geom_point(aes(x = Observation_Date, y = AGDD)) +
  labs(title="US AGDD v.s. Peak Bloom Date",
       x ="Date", y = "AGDD (celcius)") +
  theme(plot.title = element_text(hjust = 0.5),
       axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



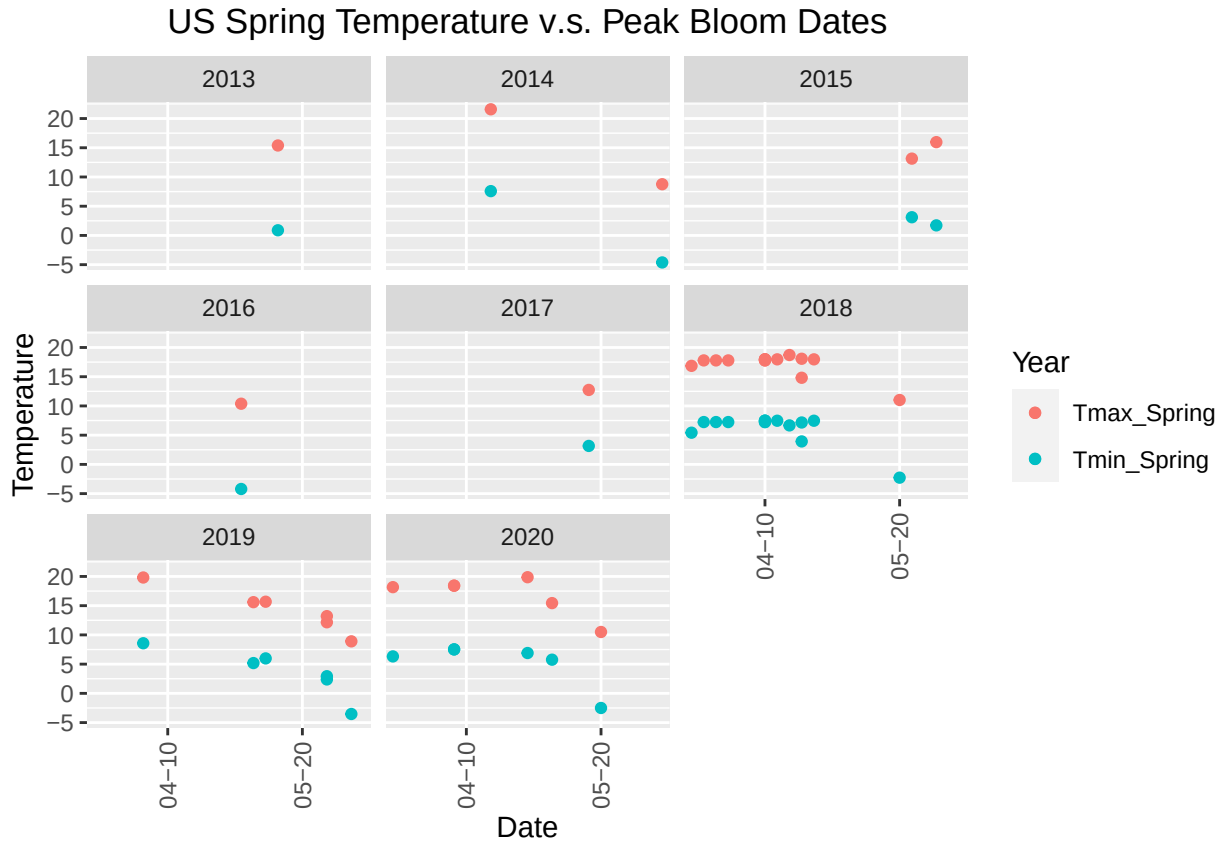

```
USA_status_intensity %>%
  filter(AGDD > 0) %>%
  ggplot() +
  geom_point(aes(x = Date, y = AGDD)) +
  facet_wrap(~Year) +
  labs(title="US AGDD v.s. Peak Bloom Dates by Year",
       x ="Date", y = "AGDD (celcius)") +
```

```
  scale_x_discrete(breaks = c('04-10','04-20','05-20')) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

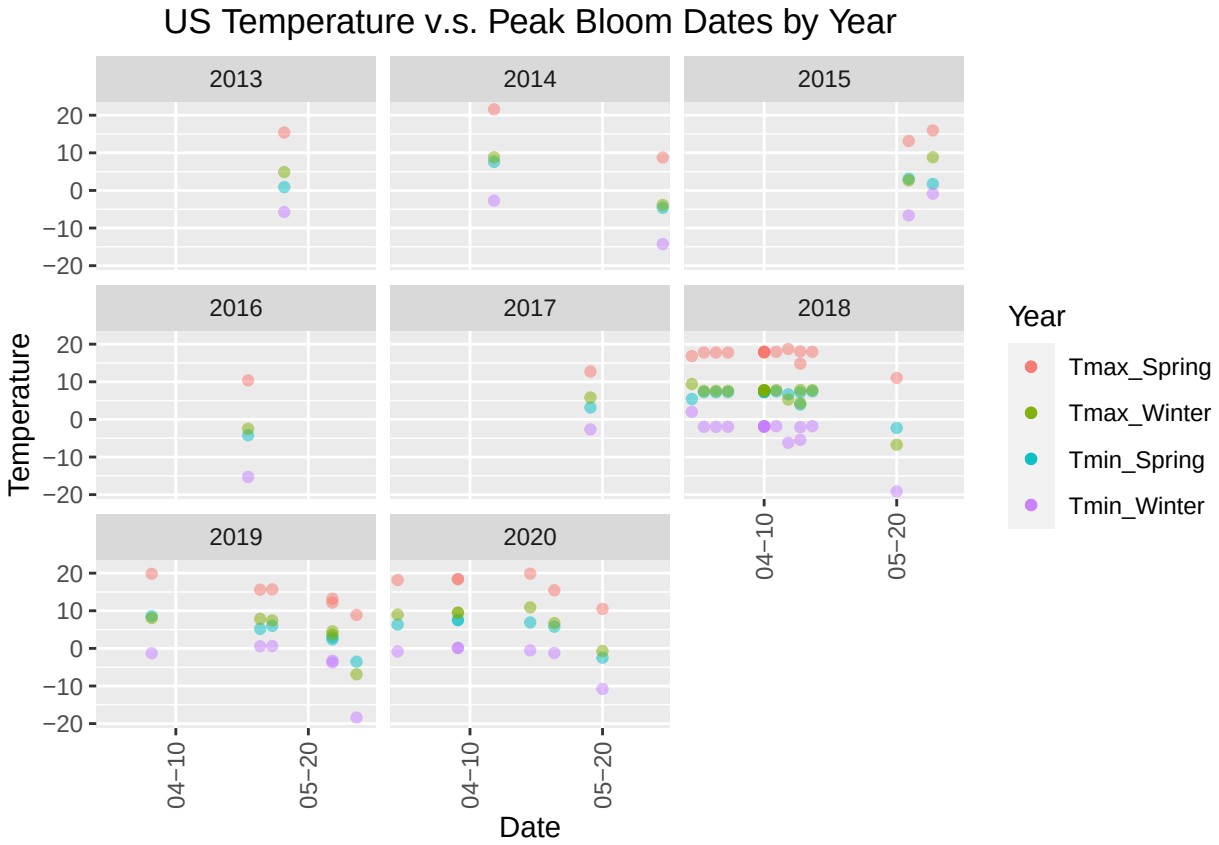


Relationships Between Temperature and Peak Bloom Dates

```
USA_status_intensity %>%
  filter(Tmin_Winter !=-9999) %>%
  ggplot() +
  geom_point(aes(x = Date, y = Tmax_Spring, color='Tmax_Spring')) +
  geom_point(aes(x = Date, y = Tmin_Spring, color='Tmin_Spring')) +
#  geom_point(aes(x = Date, y = Tmax_Winter, color='Tmax_Winter')) +
#  geom_point(aes(x = Date, y = Tmin_Winter, color='Tmin_Winter')) +
  facet_wrap(~Year) +
  labs(title="US Spring Temperature v.s. Peak Bloom Dates",
       x ="Date", y = "Temperature", color='Year') +
  scale_x_discrete(breaks = c('04-10','05-01','05-20')) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

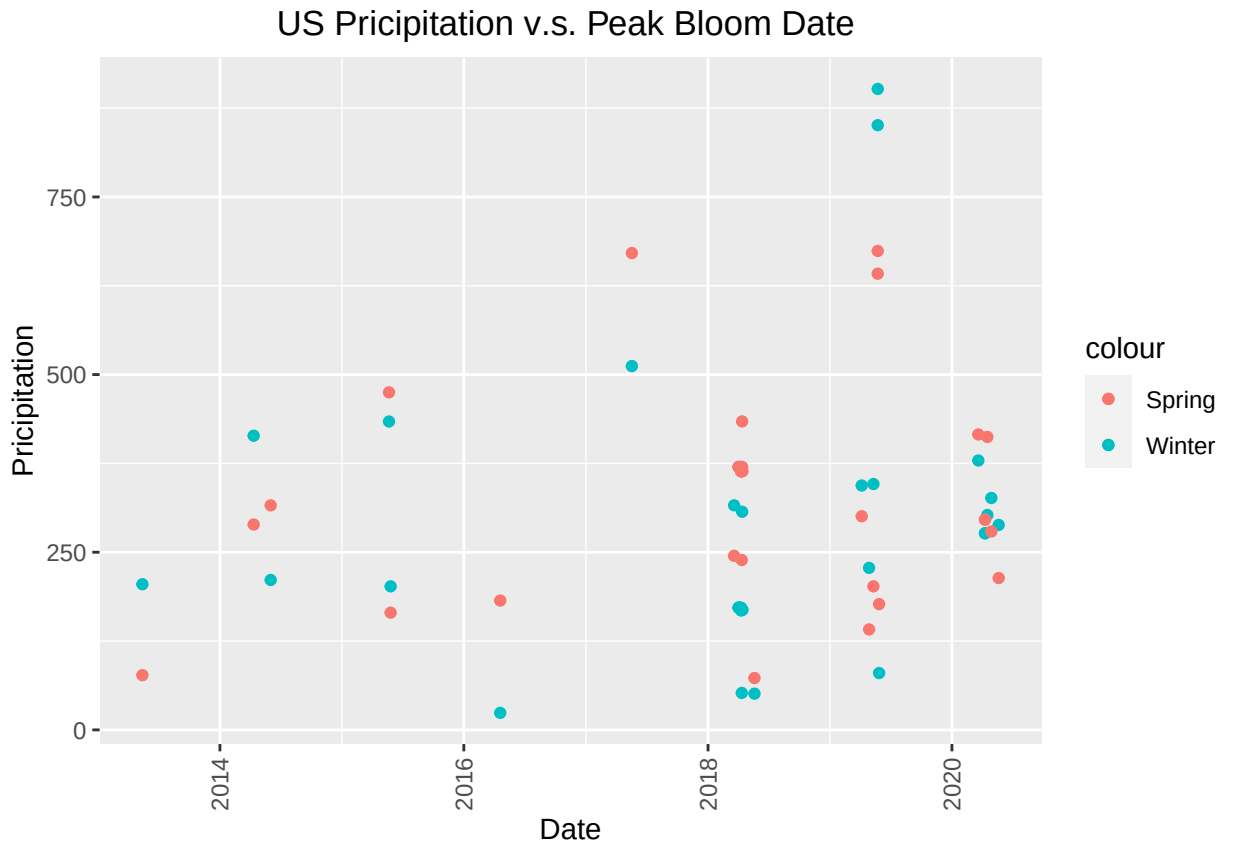



```
USA_status_intensity %>%
  filter(Tmin_Winter !=-9999) %>%
  ggplot() +
  geom_point(aes(x = Date, y = Tmax_Spring, color='Tmax_Spring'), alpha=0.5) +
  geom_point(aes(x = Date, y = Tmin_Spring, color='Tmin_Spring'), alpha=0.5) +
  geom_point(aes(x = Date, y = Tmax_Winter, color='Tmax_Winter'), alpha=0.5) +
  geom_point(aes(x = Date, y = Tmin_Winter, color='Tmin_Winter'), alpha=0.5) +
  facet_wrap(~Year) +
  labs(title="US Temperature v.s. Peak Bloom Dates by Year",
        x ="Date", y = "Temperature", color='Year') +
  scale_x_discrete(breaks = c('04-10','05-01','05-20')) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

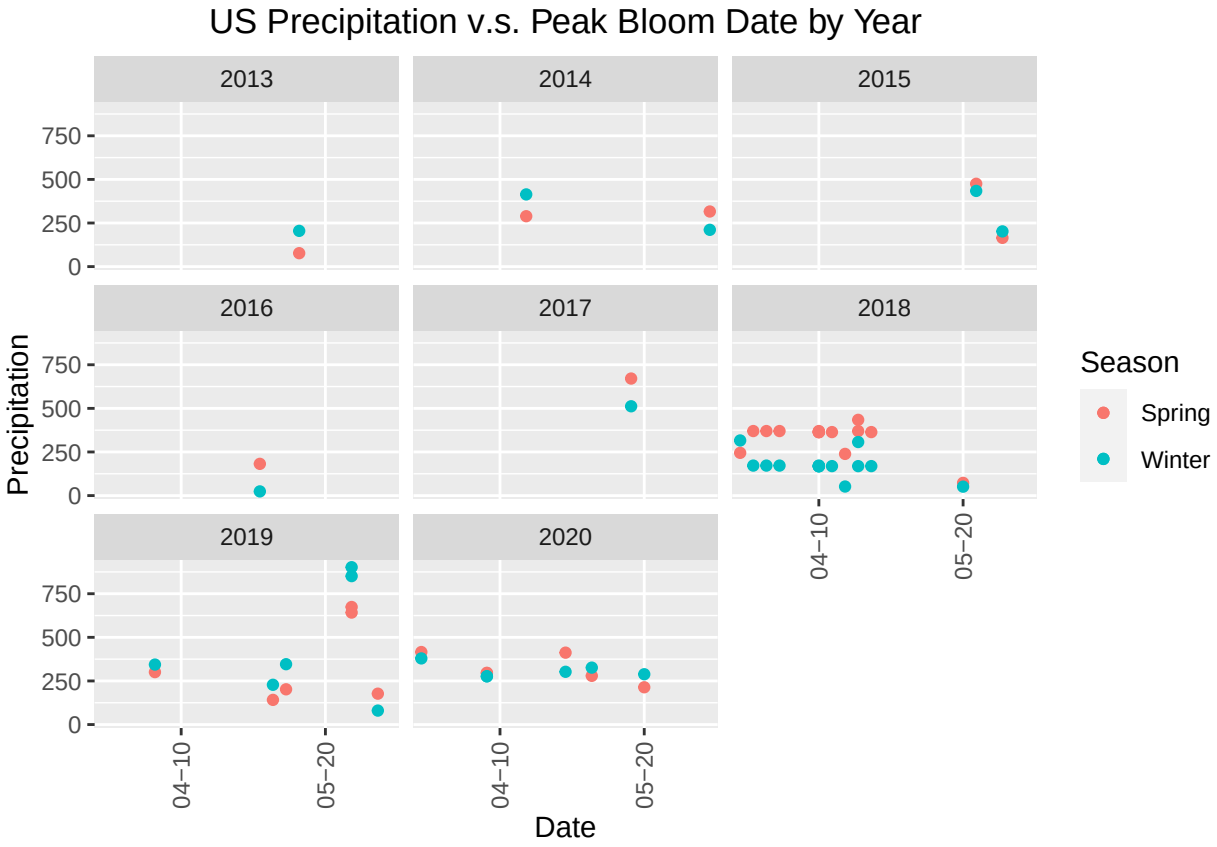


Relationships Between Pricipitation and Peak Bloom Dates

```
USA_status_intensity %>%
  filter(Prcp_Spring > 0, Prcp_Winter > 0) %>%
  ggplot() +
  geom_point(aes(x = Observation_Date, y = Prcp_Winter, color='Winter')) +
  geom_point(aes(x = Observation_Date, y = Prcp_Spring, color='Spring')) +
  labs(title="US Pricipitation v.s. Peak Bloom Date ",
        x ="Date", y = "Pricipitation") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```




```
USA_status_intensity %>%
  filter(Prcp_Spring > 0, Prcp_Winter > 0) %>%
  ggplot() +
  geom_point(aes(x = Date, y = Prcp_Spring, color = 'Spring')) +
  geom_point(aes(x = Date, y = Prcp_Winter, color = 'Winter')) +
  facet_wrap(~Year) +
  labs(title="US Precipitation v.s. Peak Bloom Date by Year",
        x ="Date", y = "Precipitation", color='Season') +
  scale_x_discrete(breaks = c('04-10','05-01','05-20')) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```


US Precipitation v.s. Peak Bloom Date by Year
2013
2014
2015
2016
2017
2018
2019
2020
Precipitation
750
500
250
0
04–10
05–20
Date
Season
Spring
Winter