# Dayton's Weather in 2014

## 12/26/2021

```
library(dplyr)
```

**Temp trend in Dayton, OH. This data for this graphic is available as a white-space separated file. We first read the data and rename the columns.**

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v stringr 1.4.0
## v tidyr   1.1.4     v forcats 0.5.1
## v readr   2.1.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
DAY <- read_table("https://bit.ly/2NGLI7N", col_names = FALSE)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   X1 = col_double(),
##   X2 = col_double(),
##   X3 = col_double(),
##   X4 = col_double()
## )
```

```
## Warning: 4 parsing failures.
##  row col  expected    actual                           file
## 4623  -- 4 columns 5 columns 'https://bit.ly/2NGLI7N'
## 4979  -- 4 columns 5 columns 'https://bit.ly/2NGLI7N'
## 5016  -- 4 columns 5 columns 'https://bit.ly/2NGLI7N'
## 5213  -- 4 columns 5 columns 'https://bit.ly/2NGLI7N'
```

```r
names(DAY) <- c("Month", "Day", "Year", "Temp")
DAY <- select(DAY, "Year", "Month", "Day", "Temp")
DAY
```

```
## # A tibble: 9,265 x 4
##     Year Month   Day  Temp
##    <dbl> <dbl> <dbl> <dbl>
##  1  1995     1     1  39
##  2  1995     1     2  19.6
##  3  1995     1     3  20.6
##  4  1995     1     4  11.3
##  5  1995     1     5   6.8
##  6  1995     1     6  23
##  7  1995     1     7  28.2
##  8  1995     1     8  24.7
##  9  1995     1     9  25
## 10  1995     1    10  23.6
## # ... with 9,255 more rows
```

```r
NEWDAY <- DAY %>%
  group_by(Year) %>%
  mutate(newDay = seq(1, length(Day))) %>%
  ungroup()

Past <- NEWDAY %>%
  filter(Temp != -99 & Year != 2014) %>% #-99 is missing
  group_by(newDay) %>%
  mutate(upper = max(Temp), #identify max value for each day
         lower = min(Temp), # identify min value for each day
         avg = mean(Temp), # calculate mean value for each day
         se = sd(Temp)/sqrt(length(Temp)),#standard error
         avg_upper = avg+(2.101*se), # 99% "confidence" interval
         avg_lower = avg-(2.101*se)) %>% ungroup()
Past
```

We then create a data frame of the historical data from 1995 to 2013. Output newDay goes 1 to 365/366 each year.

```
## # A tibble: 8,884 x 11
##     Year Month   Day  Temp newDay upper lower   avg    se avg_upper avg_lower
##    <dbl> <dbl> <dbl> <dbl>  <int> <dbl> <dbl> <dbl> <dbl>     <dbl>     <dbl>
##  1  1995     1     1  39        1  50.8  -0.3  31.4  2.39      36.5      26.4
##  2  1995     1     2  19.6      2  54.1  -3.6  30.8  2.88      36.9      24.8
```

```
## 3   1995     1    3  20.6     3  56.8   8.8  31.9  3.19      38.6      25.2
## 4   1995     1    4  11.3     4  59.2   7.1  31.6  2.79      37.5      25.7
## 5   1995     1    5   6.8     5  56.5     0  28.1  2.84      34.1      22.2
## 6   1995     1    6  23       6  56.3   2.4  28.8  2.58      34.2      23.4
## 7   1995     1    7  28.2     7  59.2   4.6  28.3  2.65      33.8      22.7
## 8   1995     1    8  24.7     8  59.1   0.8  28.6  2.66      34.2      23.0
## 9   1995     1    9  25       9  48.6  14.8  30.2  2.06      34.6      25.9
## 10  1995     1   10  23.6    10  52.6   6.2  30.0  2.33      34.9      25.1
## # ... with 8,874 more rows
```

```
filter(Past, newDay == 1)
```

```
## # A tibble: 25 x 11
##      Year Month   Day  Temp newDay upper lower   avg    se avg_upper avg_lower
##     <dbl> <dbl> <dbl> <dbl>  <int> <dbl> <dbl> <dbl> <dbl>     <dbl>     <dbl>
## 1   1995     1     1  39         1  50.8  -0.3  31.4  2.39      36.5      26.4
## 2   1996     1     1  34.5       1  50.8  -0.3  31.4  2.39      36.5      26.4
## 3   1997     1     1  38.9       1  50.8  -0.3  31.4  2.39      36.5      26.4
## 4   1998     1     1  20.3       1  50.8  -0.3  31.4  2.39      36.5      26.4
## 5   1999     1     1  10         1  50.8  -0.3  31.4  2.39      36.5      26.4
## 6   2000     1     1  39.3       1  50.8  -0.3  31.4  2.39      36.5      26.4
## 7   2001     1     1  23.2       1  50.8  -0.3  31.4  2.39      36.5      26.4
## 8   2002     1     1  16.2       1  50.8  -0.3  31.4  2.39      36.5      26.4
## 9   2003     1     1  34.9       1  50.8  -0.3  31.4  2.39      36.5      26.4
## 10  2004     1     1  37.1       1  50.8  -0.3  31.4  2.39      36.5      26.4
## # ... with 15 more rows
```

```r
Present <- NEWDAY %>%
  filter(Temp != -99 & Year == 2014)
# Create dataframe that represents the lowest and highest
# temp for each day for the historical data from 1995 to 2013.

PastRecords <- Past %>%
  group_by(newDay) %>%
  summarise(Pastlow = min(Temp),
            Pasthigh = max(Temp))
## create dataframe that identifies the days in 2014 in which temps
## were either lower or higher than all previous 19 years.

PresentRecords <- Present %>%
  left_join(PastRecords, by = "newDay") %>%
  mutate(recordlow = ifelse(Temp < Pastlow, "Y", "N"),
         recordhigh = ifelse(Temp > Pasthigh, "Y", "N")) %>%
  filter(recordlow == "Y" | recordhigh == "Y")
```

```r
p <- ggplot(Past, mapping = aes(newDay, Temp)) +
  theme(plot.background = element_blank(),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
```

```
        panel.border = element_blank(),
        panel.background = element_blank(),
        axis.ticks = element_blank(), axis.title = element_blank())+
  geom_linerange(Past,
               mapping = aes(x = newDay, ymin = lower, ymax = upper),
               colour = "wheat2", alpha = .1)
p
```

We next extract data for the Year 2014. We then create data frames representing the lowest and highest temp for each day of the year for the historical data.
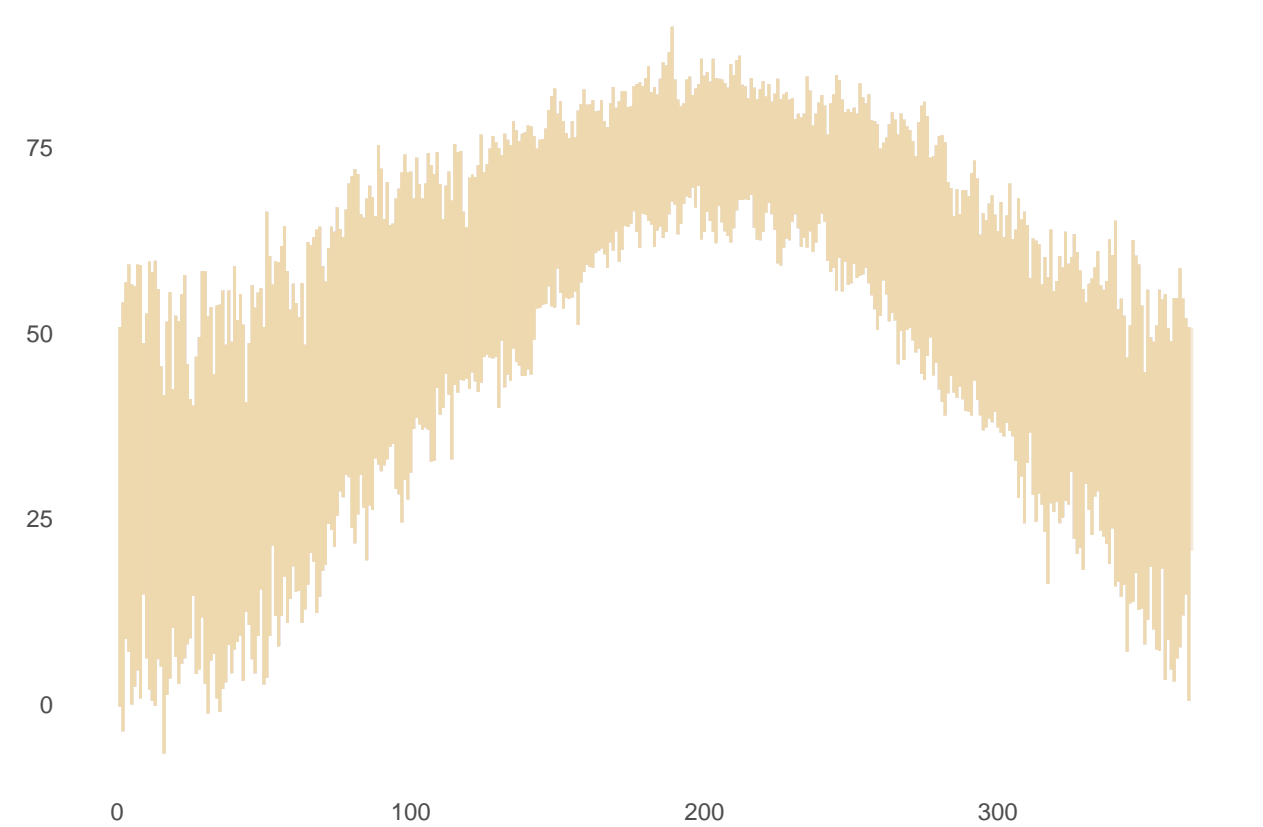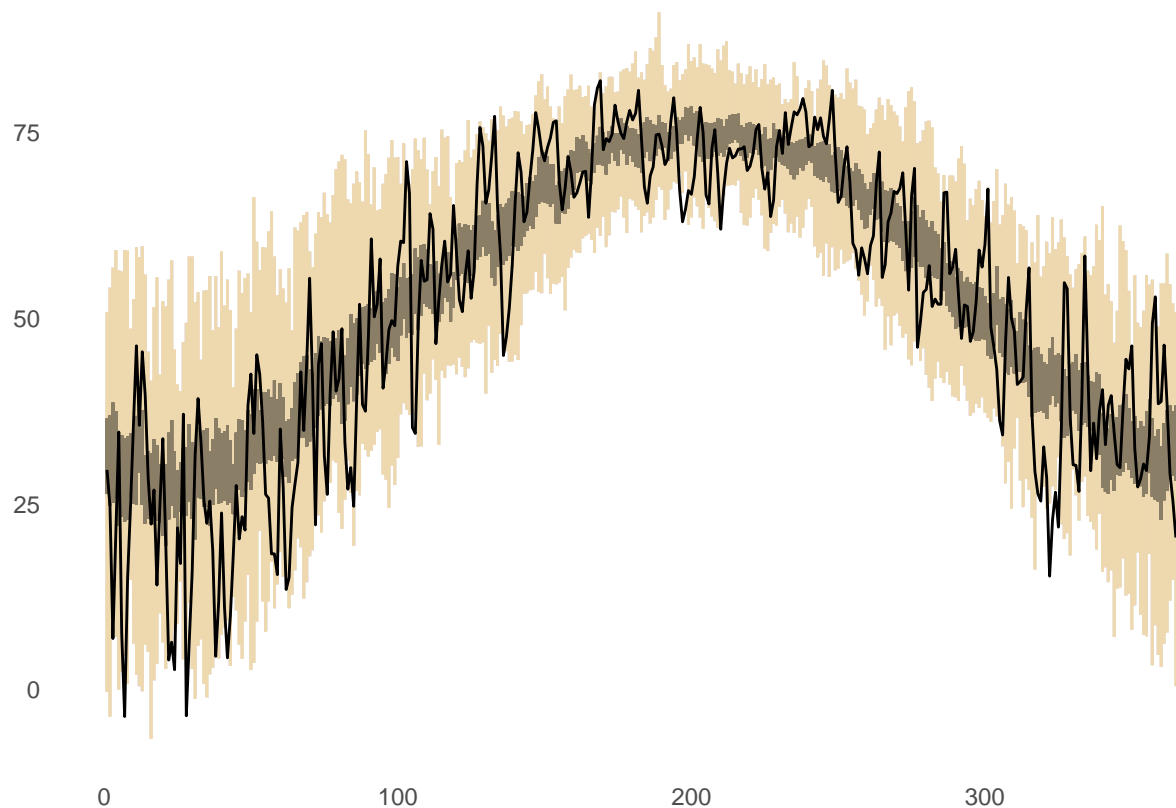


We first plot the historical data.

```
p <- p + geom_linerange(Past, mapping = aes(x = newDay, ymin = avg_lower, ymax = avg_upper),
                       colour = "wheat4")
p <- p + geom_line(Present, mapping = aes(x = newDay, y = Temp))
p
```
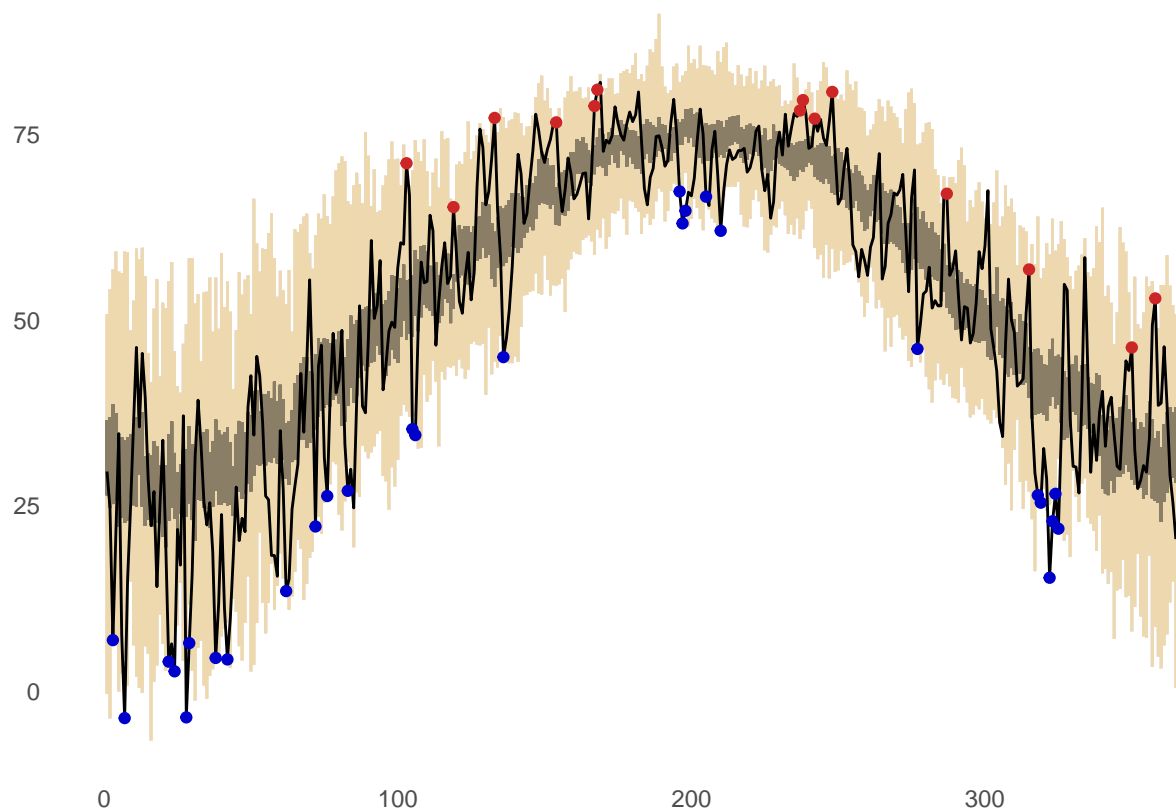
We next plot the temp. for 2014 along with the avearge temp. based on historical data.



We then plot the dates (in 2014) for which the temperature break either the lowest or the highest record.

```
p <- p + geom_point(data = filter(PresentRecords, recordlow == "Y"), aes(x = newDay, y = Temp), colour
  geom_point(data = filter(PresentRecords, recordhigh == "Y"),
            aes(x = newDay, y = Temp), colour = "firebrick3")
p
```

We then setup the axis and labels.

```r
p <- p + geom_vline(xintercept = 0, colour = "wheat4", linetype = 1, size = 1)
xseq <- c(31,59,90,120,151,181,212,243,273,304,334,365)
for(x in xseq){
  p <- p + geom_vline(xintercept = x, colour = "wheat4",
                      linetype = 3, size = .5)
}
dgr_fmt <- function(x, ...){
  parse(text = paste(x, "*degree", sep = ""))
}
# create y-axis variable
a <- dgr_fmt(seq(-20,100, by = 10))

p <- p + coord_cartesian(ylim = c(-20,100))+
  scale_y_continuous(breaks = seq(-20,100, by =10), labels = a)+
  scale_x_continuous(expand = c(0,0),
                     breaks = c(15,45,75,105,135,165,195,
                                228,258,288,320,350),
                     labels = c("Jan", "Feb", "Mar", "Apr",
                                "May", "June", "July", "Aug", "sep", "Oct", "Nov", "Dec"))
p
```
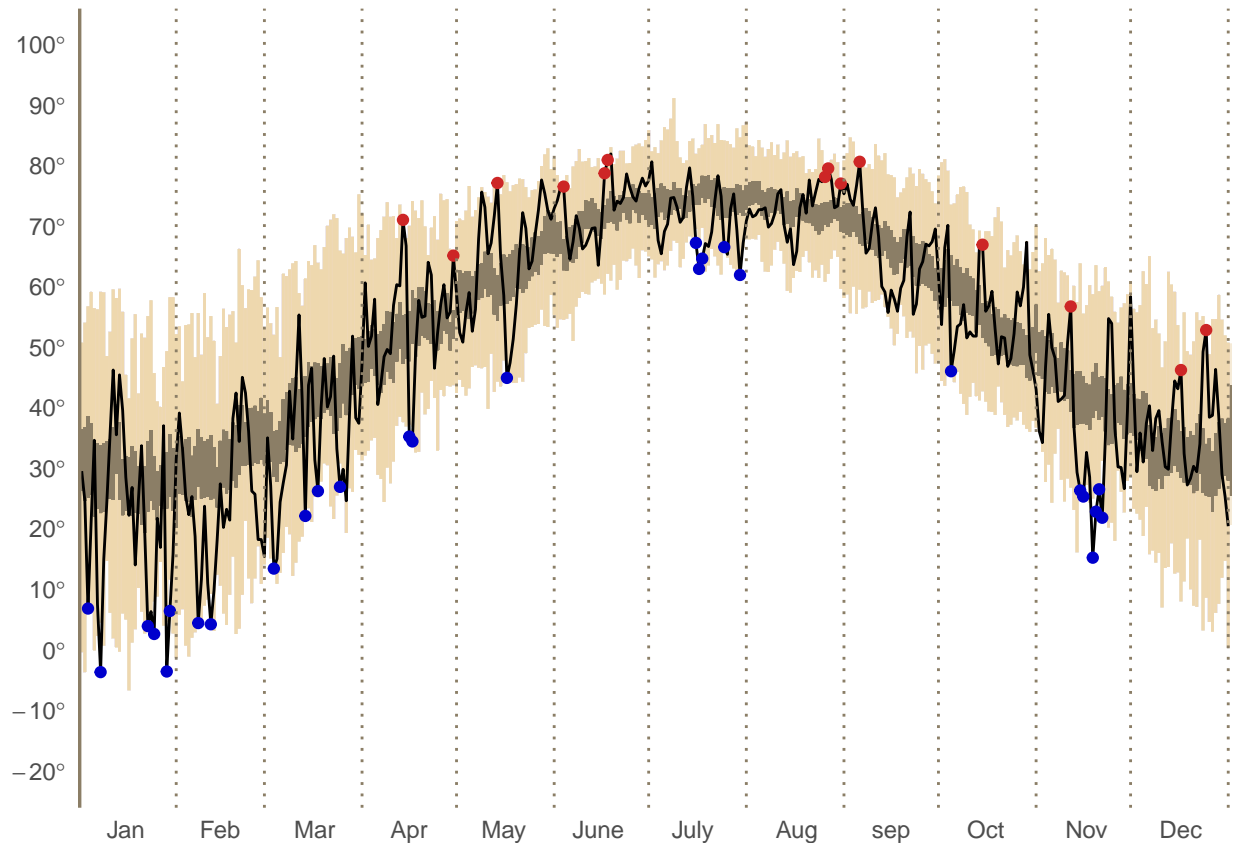
**Final output is obtained by adding annotate commands for the annotations**

```
p <- p + ggtitle("Dayton's weather in 2014") +
  theme(plot.title = element_text(face = "bold", hjust = 0.012, vjust = 0.8, colour = "#3C3C3C", size =
  annotate("text", x = 119, y = 98, label = "Temperature", size = 4, fortface = "bold")
```

```
## Warning: Ignoring unknown parameters: fortface
```

```
p <- p + annotate("text", x = 126, y = 93,
                  label = paste("Data represents average daily",
                                "temperatures. Accessible data dates",
                                "back to"),
                  size = 3, colour = "gray30") +
  annotate("text", x = 122, y = 89,
           label = paste("January 1, 1995. Data for 2014 is",
                         "only available through December 16."),
           size = 3, colour = "gray30") +
  annotate("text", x = 124, y = 85,
           label = paste("Average temperature for the year was",
                         "51.9 degree making 2014 the 9th coldest"),
           size = 3, colour = "gray30")+
  annotate("text", x = 128, y = 81, label = "year since 1995",
           size = 3, colour = "gray30")
```

```
p <- p +
  annotate("segment", x = 30, xend = 40, y = -5,
           yend = -10, colour = "blue3") +
  annotate("text", x = 65, y = -10,
           label = "We had 35 days that were the",
           size = 3, colour = "blue3") +
  annotate("text", x = 56, y = -14,
           label = "coldest since 1995",
           size = 3, colour = "blue3") +
  annotate("segment", x = 302, xend = 307, y = 74,
           yend = 82, colour = "firebrick3") +
  annotate("text", x = 333, y = 92,
           label = "we had 19 days that were the",
           size = 3, colour = "firebrick3") +
  annotate("text", x = 324, y = 88,
           label = "hottest since 1995",
           size = 3, colour = "firebrick3")
p
```