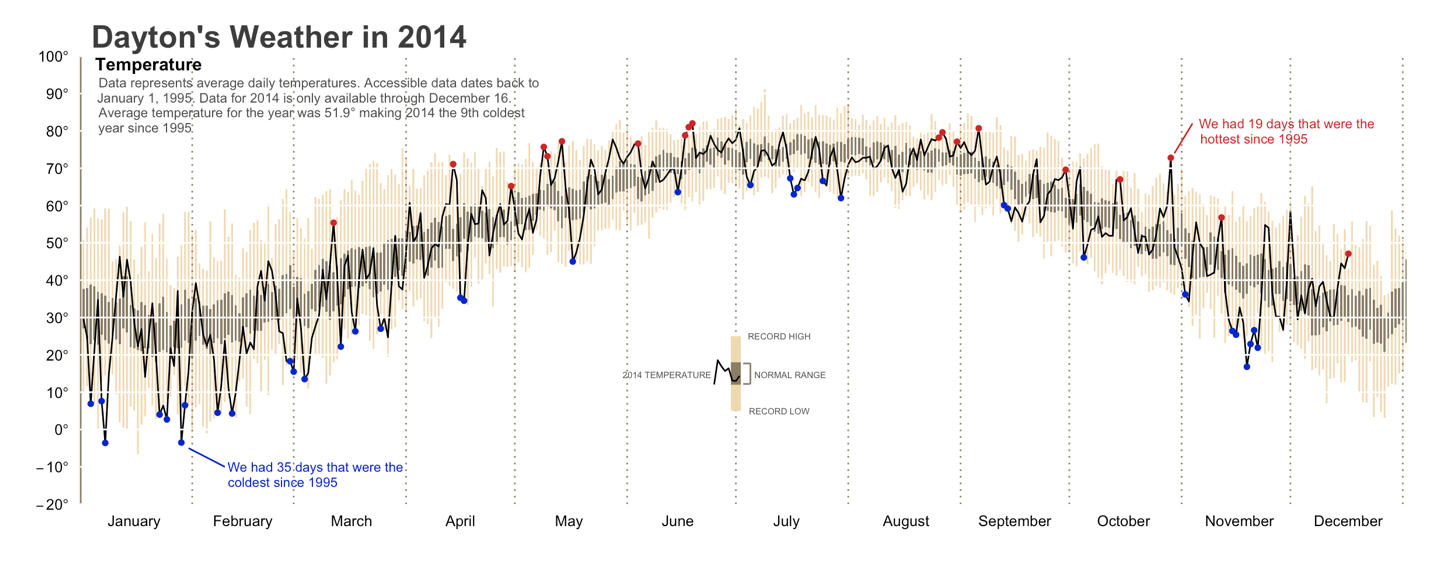
# **R Weather Graphic Challenge**

There’s an iconic weather graphic in the book Visual Display of Quantitative Information that follows many of the good design principles we’ve covered this year and last.



Many R courses have used the same underlying data set for Dayton, Ohio to create similar graphics. See a tutorial here that will form the basis for the rest of this lab: <https://rpubs.com/bradleyboehmke/weather_graphic>

The problem is that the weather data file used for the tutorial is no longer available on the university website and doesn’t seem to be available elsewhere online either. To work around this, I’ve found and formatted temperature data for two locations of interest: Philadelphia, PA and Crater Lake, OR. Philadelphia was selected because it’s similar to our local climate and Crater Lake because it’s a stark contrast to our area, with cooler temperatures and heavy snowfall. The record highs and lows as well as the averages in the spreadsheet are for the period 1894-2023 for Philadelphia and 1919-2023 for Crater Lake.

For this lab, you’ll be using the tutorial linked above to create a weather graphic similar to the one pasted above for each of the two locations, but with some necessary tweaks.

1. The Dayton data file had weather data from 1995-2013. In the tutorial, they had to calculate the record highs and lows over that period (PastLows and PastHighs). The file I’m providing already has the record highs and lows for each day, calculated across the entire historical period. We will use those instead. The record highs/lows are the yellow bars in the graphic above.
2. Since the Dayton, Ohio data spanned 18 years, they could calculate a 95% confidence interval (avg\_upper and avg\_lower) by day. We don’t have data for all of these years available, so we will have to use an approximation instead. This can be tricky and you’ll have to think about this. You have the averages and records to help with the approximation. The confidence intervals are the black bars in the graphic above.

Some recommendations:

* We have two sheets of data in Excel for the two locations. Read only one in to start. Make sure the numeric columns are being interpreted as such. Rename the columns to make them easier to work with.
* Create a dataframe called Past, like in the tutorial. Store the record highs and lows as well as your approximated confidence intervals. You don’t need a filter. I fixed the bad data points and there aren’t any past years of data that we need to look at. The Present dataframe shouldn’t need many tweaks.
* You’ll obviously have to change a lot of the caption text. Feel free to add your own additional information, such as info about where the locations are, their climate classifications, any particularly interesting findings from the 2022 data, etc.
* Put your own unique style on it. Change some colors, font types, etc. if you think that would improve upon the original graphic above.

Paste all code and output for each of the two visuals (Philadelphia and Crater Lake) below, along with any comments about what went well or didn’t.

> Philadelphia <- read\_excel("data/R Weather Graphic Input.xlsx", 2)

> Crater Lake <- read\_excel("data/R Weather Graphic Input.xlsx", 3)

> library(stats)

> library(janitor)

> Philadelphia <- Philadelipha |>

+ clean\_names()

> Past <- Philadelphia |>

+ select(month, day, record\_high\_f, record\_low\_f, average\_high\_f, average\_low\_f)

> print(Past)

# A tibble: 365 × 6

month day record\_high\_f record\_low\_f average\_high\_f

*<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*

1 1 1 64 0 43.5

2 1 2 64 2 43.4

3 1 3 63 -6 43.3

4 1 4 68 -7 43.2

5 1 5 66 -16 42.9

6 1 6 73 -12 42.7

7 1 7 65 -4 42.3

8 1 8 69 -5 42

9 1 9 66 -3 41.8

10 1 10 65 -3 41.9

Past <- Past |>

mutate(newDay = seq(1, length(day))) # label days as 1:365 (will represent x-axis)

p <- ggplot(Past, aes(newDay, average\_high\_f)) +

theme(plot.background = element\_blank(),

panel.grid.minor = element\_blank(),

panel.grid.major = element\_blank(),

panel.border = element\_blank(),

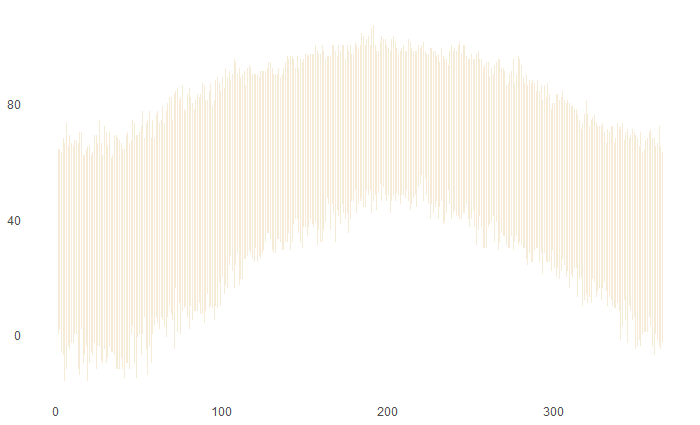
panel.background = element\_blank(),

axis.ticks = element\_blank(),

axis.title = element\_blank()) +

geom\_linerange(Past, mapping=aes(x=newDay, ymin=record\_low\_f, ymax=record\_high\_f),alpha=.5 , color = "wheat2")

print(p)



> # create dataframe that identifies the days in 2014 in which the temps were lower than all previous 19 years

> PresentLows <- Philadelphia %>%

+ mutate(record = ifelse(low\_f<=record\_low\_f, "Y", "N")) %>% # identifies if current year was record low

+ filter(record == "Y") # filter for days that represent current year record lows

> # create dataframe that identifies the days in 2014 in which the temps were higher than all previous 19 years

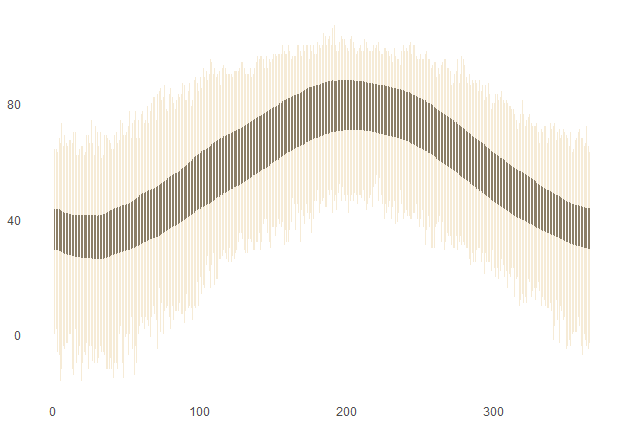
> PresentHighs <- Philadelphia %>%

+ mutate(record = ifelse(high\_f>=record\_high\_f, "Y", "N")) %>% # identifies if current year was record high

+ filter(record == "Y") # filter for days that represe current year record highs

p <- p +

geom\_linerange(Past, mapping=aes(x=newDay, ymin=average\_low\_f, ymax=average\_high\_f), colour = "wheat4")



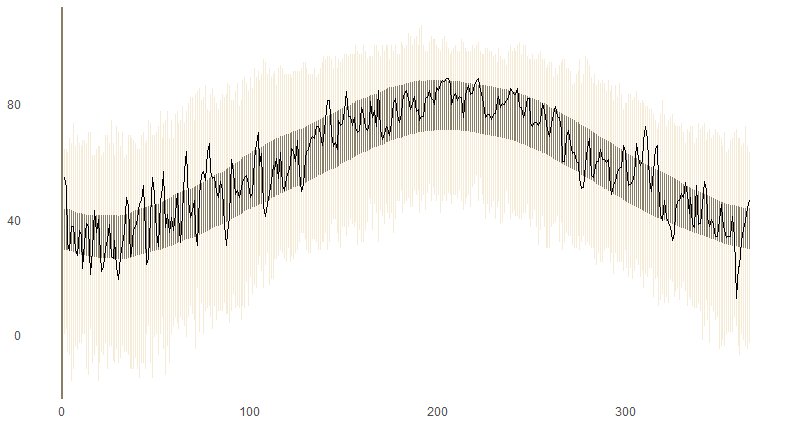
> Philadelphia <- Philadelphia |>

+ mutate(temp = (high\_f + low\_f) / 2)

p <- p +

geom\_line(Philadelphia, mapping=aes(x=newDay, y=temp, group=1)) +

geom\_vline(xintercept = 0, colour = "wheat4", linetype=1, linewidth=1)



p <- p +

geom\_hline(yintercept = -20, colour = "white", linetype=1) +

geom\_hline(yintercept = -10, colour = "white", linetype=1) +

geom\_hline(yintercept = 0, colour = "white", linetype=1) +

geom\_hline(yintercept = 10, colour = "white", linetype=1) +

geom\_hline(yintercept = 20, colour = "white", linetype=1) +

geom\_hline(yintercept = 30, colour = "white", linetype=1) +

geom\_hline(yintercept = 40, colour = "white", linetype=1) +

geom\_hline(yintercept = 50, colour = "white", linetype=1) +

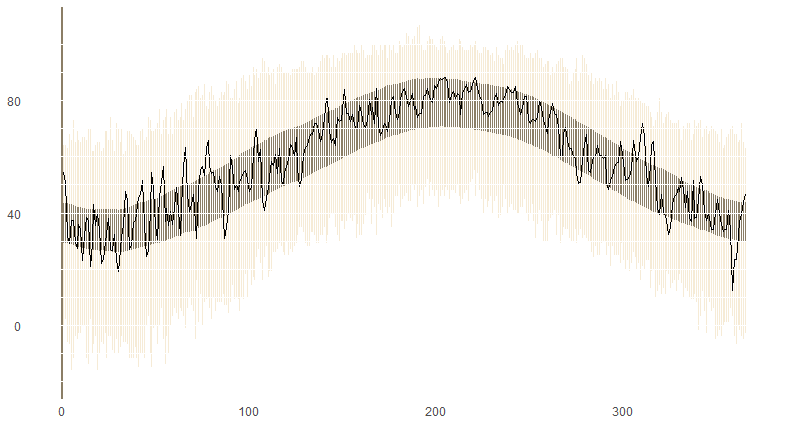
geom\_hline(yintercept = 60, colour = "white", linetype=1) +

geom\_hline(yintercept = 70, colour = "white", linetype=1) +

geom\_hline(yintercept = 80, colour = "white", linetype=1) +

geom\_hline(yintercept = 90, colour = "white", linetype=1) +

geom\_hline(yintercept = 100, colour = "white", linetype=1)



p <- p +

geom\_vline(xintercept = 31, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 59, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 90, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 120, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 151, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 181, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 212, colour = "wheat4", linetype=3, size=.5) +

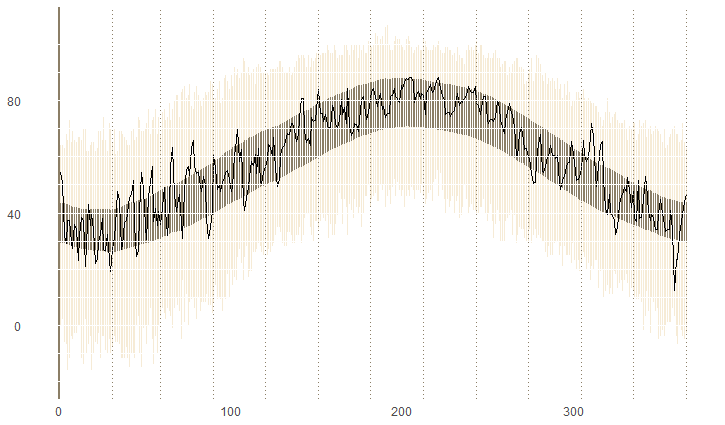
geom\_vline(xintercept = 243, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 273, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 304, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 334, colour = "wheat4", linetype=3, size=.5) +

geom\_vline(xintercept = 365, colour = "wheat4", linetype=3, size=.5)



p <- p +

coord\_cartesian(ylim = c(-20,100)) +

scale\_y\_continuous(breaks = seq(-20,100, by=10)) +

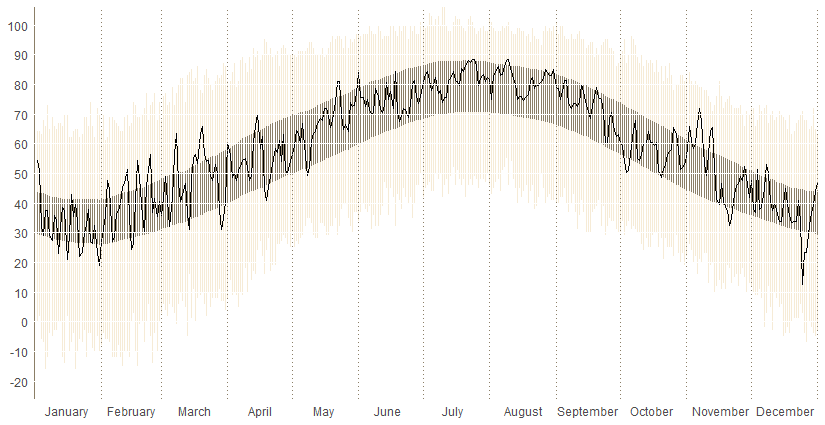
scale\_x\_continuous(expand = c(0, 0),

breaks = c(15,45,75,105,135,165,195,228,258,288,320,350),

labels = c("January", "February", "March", "April",

"May", "June", "July", "August", "September",

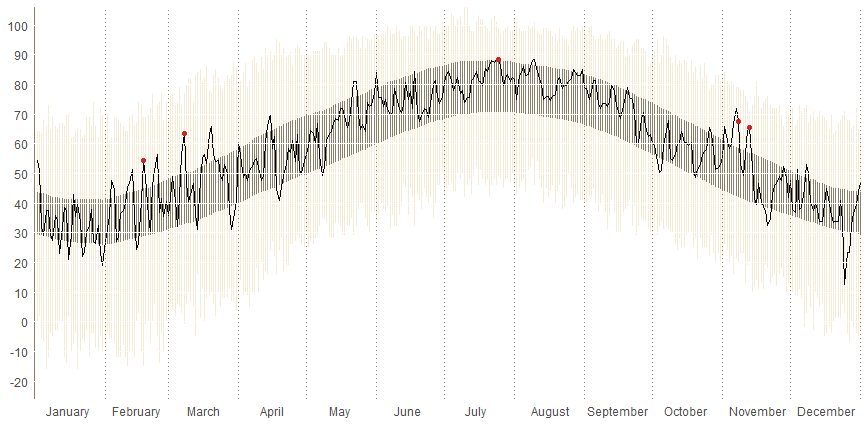
"October", "November", "December"))



p <- p +

geom\_point(data=PresentLows, aes(x=newDay, y=temp), colour="blue3") +

geom\_point(data=PresentHighs, aes(x=newDay, y=temp), colour="firebrick3")



p <- p +

ggtitle("Philadelphia's Weather in 2022") +

theme(plot.title=element\_text(face="bold",hjust=.012,vjust=.8,colour="#3C3C3C",size=20)) +

annotate("text", x = 19, y = 98, label = "Temperature", size=4, fontface="bold")

p <- p +

annotate("segment", x = 302, xend = 307, y = 74, yend = 82, colour = "firebrick3") +

annotate("text", x = 333, y = 82, label = "We had 5 days that were the", size=3, colour="firebrick3") +

annotate("text", x = 324, y = 78, label = "hottest since 1995", size=3, colour="firebrick3")

p <- p +

annotate("segment", x = 181, xend = 181, y = 5, yend = 25, colour = "wheat2", size=3) +

annotate("segment", x = 181, xend = 181, y = 12, yend = 18, colour = "wheat4", size=3) +

annotate("segment", x = 183, xend = 185, y = 17.7, yend = 17.7, colour = "wheat4", size=.5) +

annotate("segment", x = 183, xend = 185, y = 12.2, yend = 12.2, colour = "wheat4", size=.5) +

annotate("segment", x = 185, xend = 185, y = 12.2, yend = 17.7, colour = "wheat4", size=.5) +

annotate("text", x = 196, y = 14.75, label = "NORMAL RANGE", size=2, colour="gray30") +

annotate("text", x = 162, y = 14.75, label = "2022 TEMPERATURE", size=2, colour="gray30") +

annotate("text", x = 193, y = 25, label = "RECORD HIGH", size=2, colour="gray30") +

annotate("text", x = 193, y = 5, label = "RECORD LOW", size=2, colour="gray30")

print(p)

