

02-07: Plots 1

1 - Purpose

- Create a scatter plot and a line plot
- Adding lines to a plot
- Style the plot and change axis ranges
- Include a legend in your plot

2 - Concepts

3 - A New Data Frame

For this lesson we will be using an expanded weather data set that adds the **dayOfYear** and **humidity**. **dayOfYear** assigns a numerical value for each day of the year from 1-365 (or 366 on a leap year). Copy the data below and save it as **LansingWeather3.csv** in your **data** folder.

```
1  date,dayOfYear,highTemp, lowTemp, precipitation,humidity
2  27-Mar,86,57,45,0.01,88
3  28-Mar,87,50,43,0.005,87
4  29-Mar,88,54,42,0.04,80
5  30-Mar,89,40,38,1.11,74
6  31-Mar,90,39,37,0.12,67
7  1-Apr,91,58,45,0.84
8  2-Apr,92,60,46,0.005,89
9  3-Apr,93,53,50,0.49,59
10 4-Apr,94,55,48,0.45,55
11 5-Apr,95,44,40,0.3,72
12 6-Apr,96,39,36,1.13,82
13 7-Apr,97,53,43,0.004,83
14 8-Apr,98,61,45,0.87
15 9-Apr,99,75,63,0.58
```

4 - Your first plot

We will first plot a time series of the high temperatures from the data above.

The R function we will use to plot is **plot()** and there are many parameters you can include that control the plot's properties and appearance. There are numerous ways in which you can modify the plot and we will learn some of these in this and future lessons.

For now we will simply add the vector **highTempData** as a parameter for **plot()**.

```
1 {
```

```

2 rm(list=ls()); options(show.error.locations = TRUE);
3
4 weatherData = read.csv("data/Lansingweather3.csv");
5 highTempData = weatherData[, "highTemp"];
6
7 plot(highTempData);
8 }

```

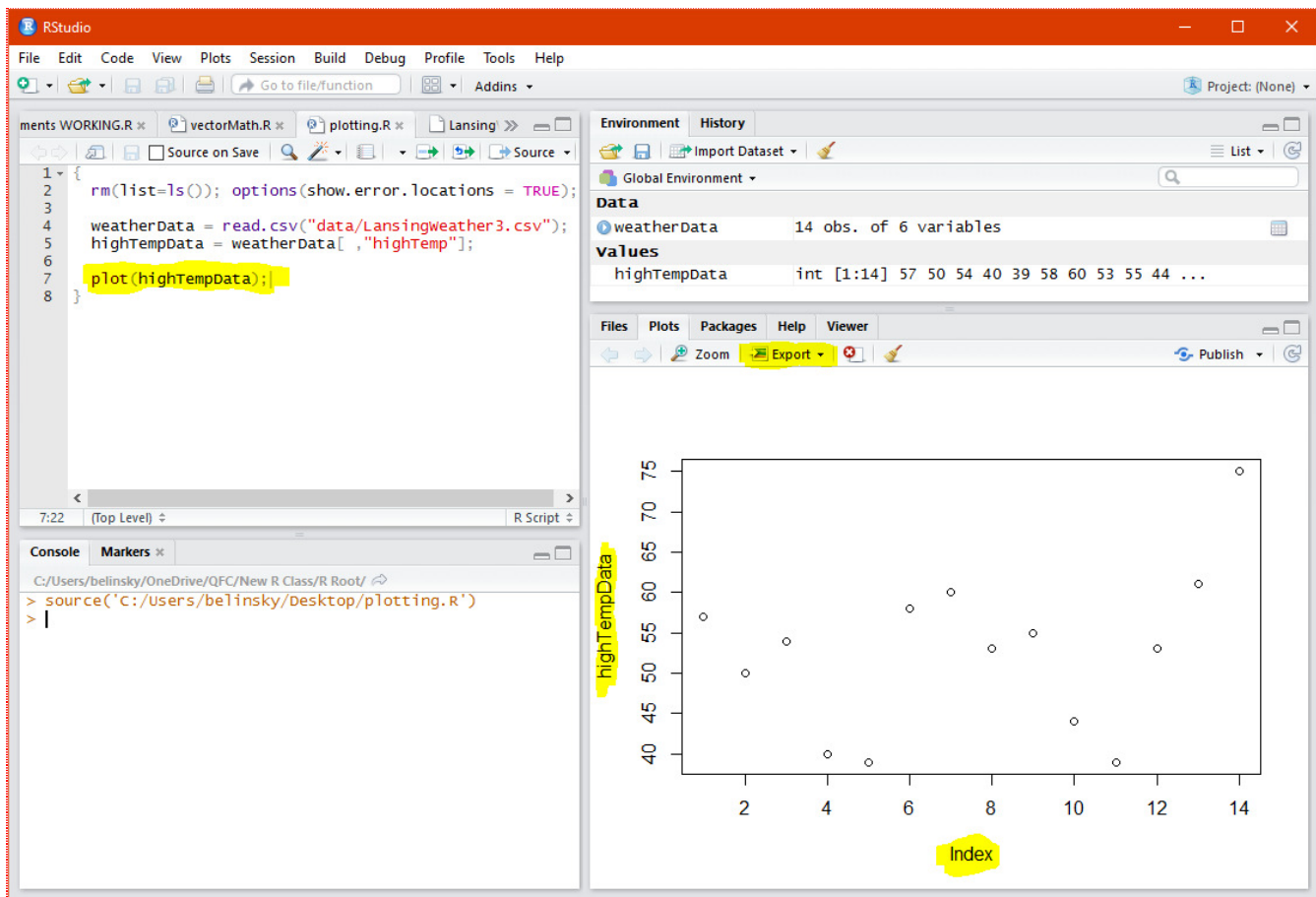


Fig 1: Scatterplot of the high temperature from **weatherData**

In the Plot Window, we see a scatterplot of the high temperature values plotted against high temperature's vector index.

The plot can be resized by dragging the gray area surrounding the Plot Window. *Extension: Full-sized window for plots.*

You can also use the **Export** button to save the plot as an image, a PDF, or to the clipboard (to be pasted to a document).

Trap: A very common, and very unintuitive, plot error

4.1 - Plotting X and Y variables

The plot above just plots the high temperature vs. the index number of the values. Instead of plotting against the index number, we can plot the high temperature against the **dayOfYear**. The **dayOfYear** for

March 27 (the first day in the data) is 86. Note: we use **dayOfYear** instead of **date** because **date** (i.e., **2-Apr**) does not contain numeric values.

```
1 {
2   rm(list=ls()); options(show.error.locations = TRUE);
3
4   weatherData = read.csv("data/LansingWeather3.csv");
5   highTempData = weatherData[, "highTemp"];
6   dateData = weatherData[, "dayOfYear"];
7
8   plot(formula=highTempData~dateData); # plot(y~x)
9 }
```

In the last line of the script:

```
1 plot(formula=highTempData~dateData);
```

highTempData is the dependent value (y) and **dateData** is the independent value (x)

The line can also be written in this form -- **plot(x, y)**

```
1 plot(x=dateData, y=highTempData);
```

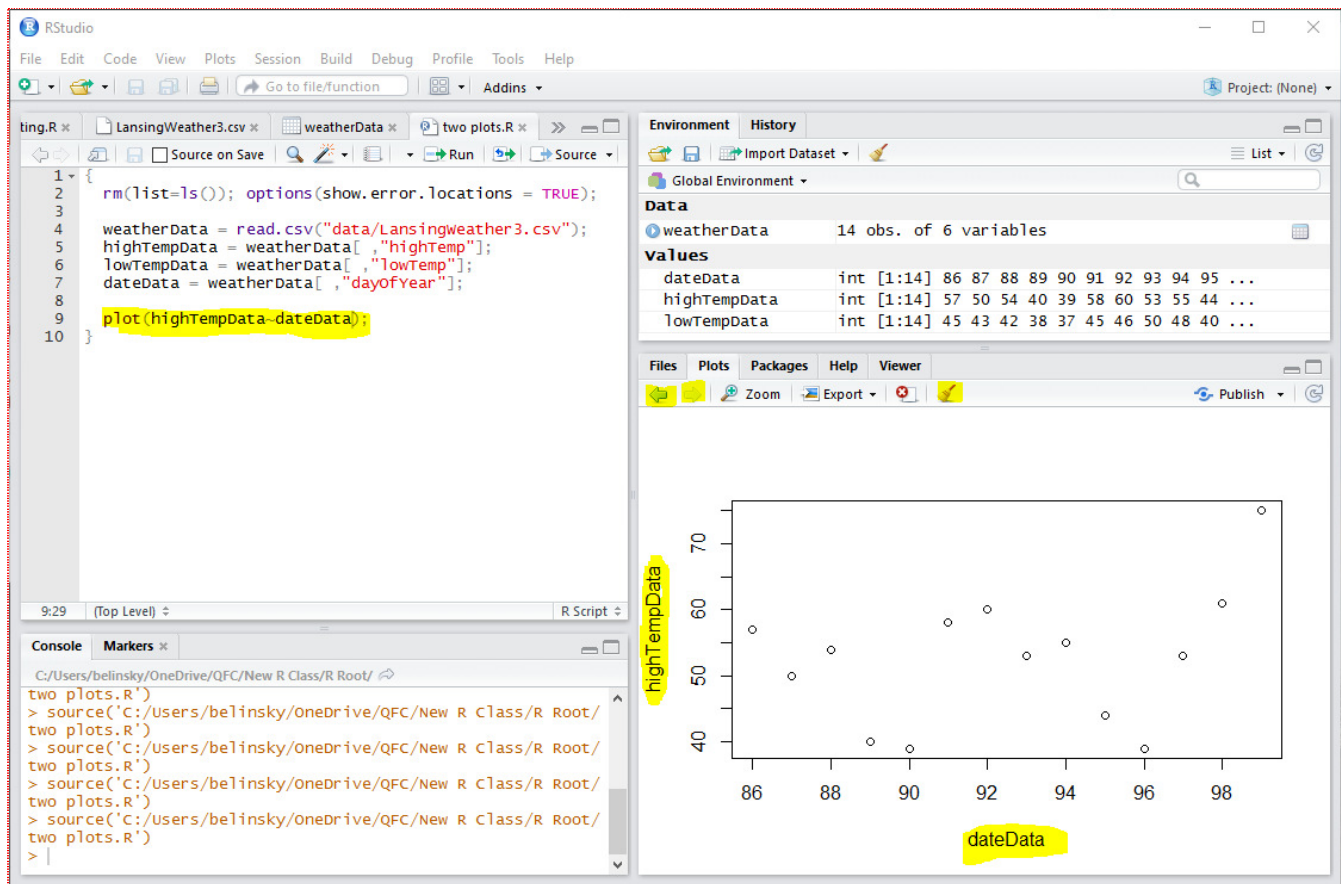


Fig 2: Plotting high temperature vs. the day-of-year

Note that the Plot Window stores all the previous plots and the plots can be cycled through by clicking the arrow buttons. You can use the brush button to remove all of the plots.

4.2 - Plotting lines

If we want to connect the points -- in other words, we want a line graph, we need to add an parameter to `plot()` called ***type***. Note: the points are connected according to their order in the vector.

the parameter ***type*** can have the following values:

- "p": points only
- "l": line only
- "o": points and lines
- "n": neither points nor lines (empty plot)

```
1 {
2   rm(list=ls()); options(show.error.locations = TRUE);
3
4   weatherData = read.csv("data/LansingWeather3.csv");
5   highTempData = weatherData[, "highTemp"];
6   dateData = weatherData[, "dayOfYear"];
7
8   plot(highTempData~dateData, type = "o");
9 }
```

Don't forget to put a comma between the parameters in `plot()`.

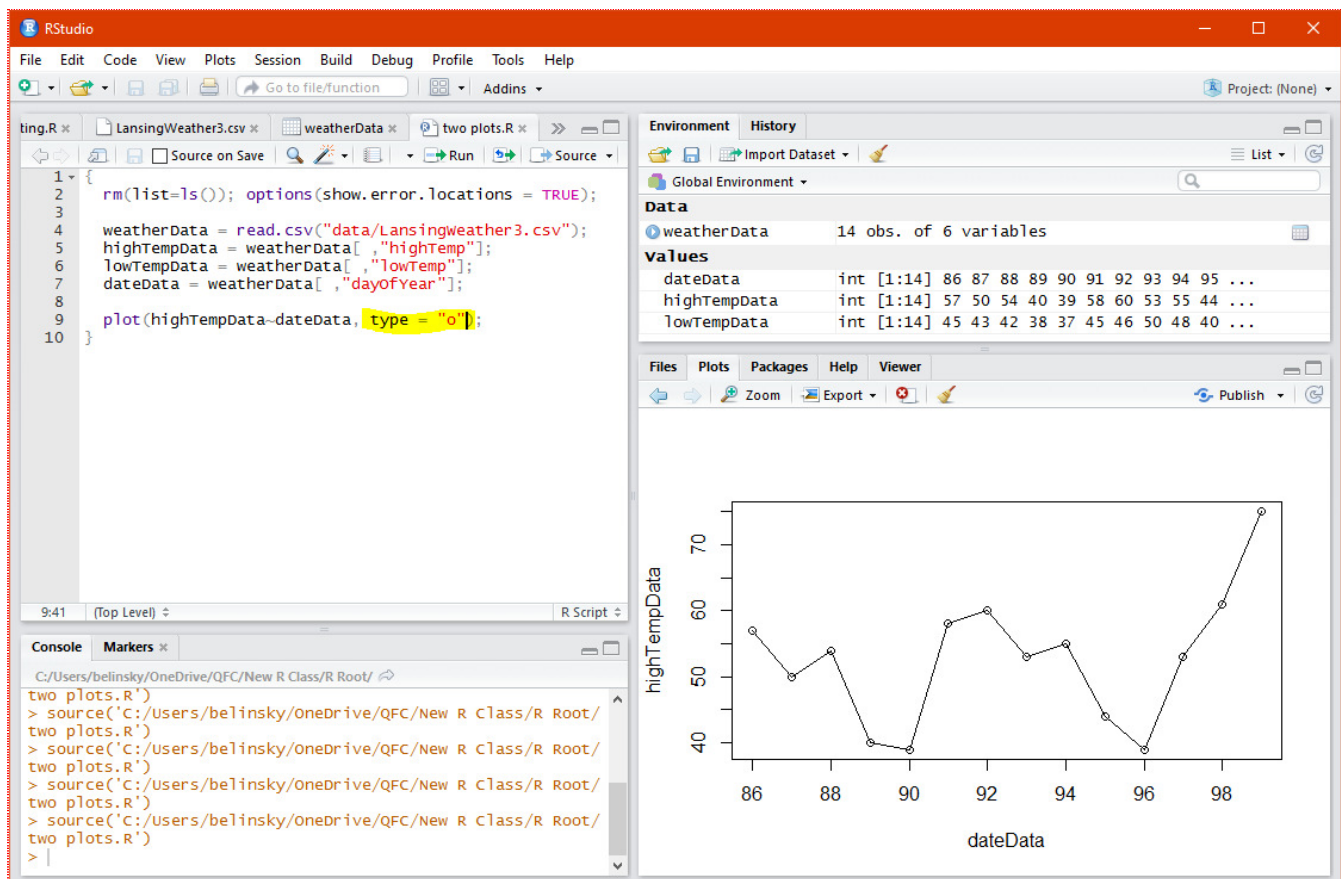


Fig 3: Connecting the points on a scatterplot.

4.3 - Graph labels and title

The default labels for the axes are the names of the variables used in the plot (e.g., **dateData** and **highTempData**) and the default for the title is blank. We can edit the labels using the parameters **xlab** and **ylab** and add a title using the parameter **main**.

```

1 {
2   rm(list=ls()); options(show.error.locations = TRUE);
3
4   weatherData = read.csv("data/Lansingweather3.csv");
5   highTempData = weatherData[, "highTemp"];
6   dateData = weatherData[, "dayOfYear"];
7
8   plot(highTempData~dateData, type = "o",
9         main="Lansing high temperatures",
10        xlab="Day of Year",
11        ylab="temperature (F)");
12 }

```

Note, I split up the function **plot()** into multiple lines to make it easier to see the individual parameters.

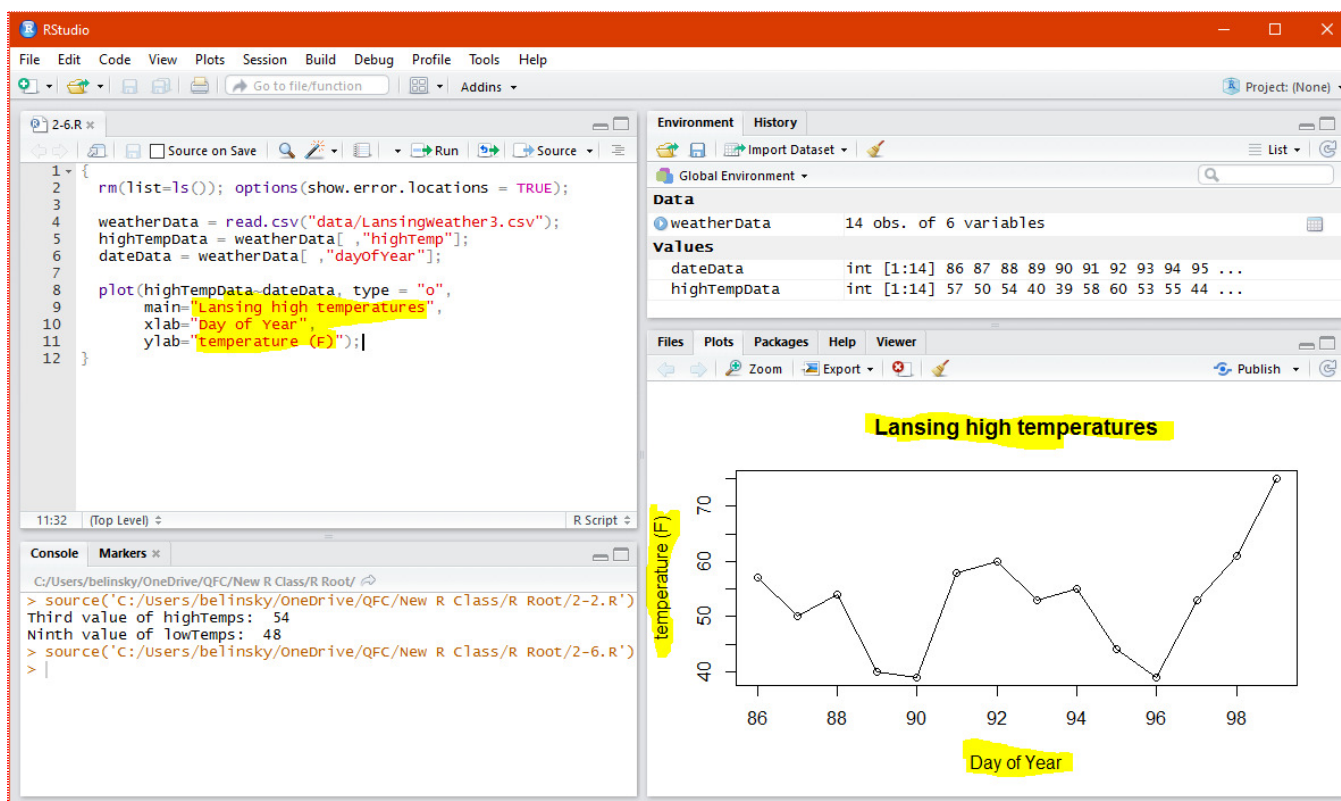


Fig 4: Adding labels to the axes and a title to the plot.

5 - Adding a second plot

The next step is to add another set of values to the plot -- specifically the low temperature data. We cannot use **plot()** to add the low temperature data because **plot()** creates a new plot so **plot()** cannot be used to

append data to an existing plot.

Instead we use the function **points()**. **points()** appends data to the current plot and the parameters for **points()** are almost the same as **plot()** -- including **type="o"** to connect the points.

```

1 {
2   rm(list=ls()); options(show.error.locations = TRUE);
3
4   weatherData = read.csv("data/LansingWeather3.csv");
5   highTempData = weatherData[ , "highTemp"];
6   lowTempData = weatherData[ , "lowTemp"];
7   dateData = weatherData[ , "dayOfYear"];
8
9   plot(highTempData~dateData, type = "o",
10        main="Lansing high temperatures",
11        xlab="Day of Year",
12        ylab="temperature (F)");
13
14   points(lowTempData~dateData, type="o");
15 }

```

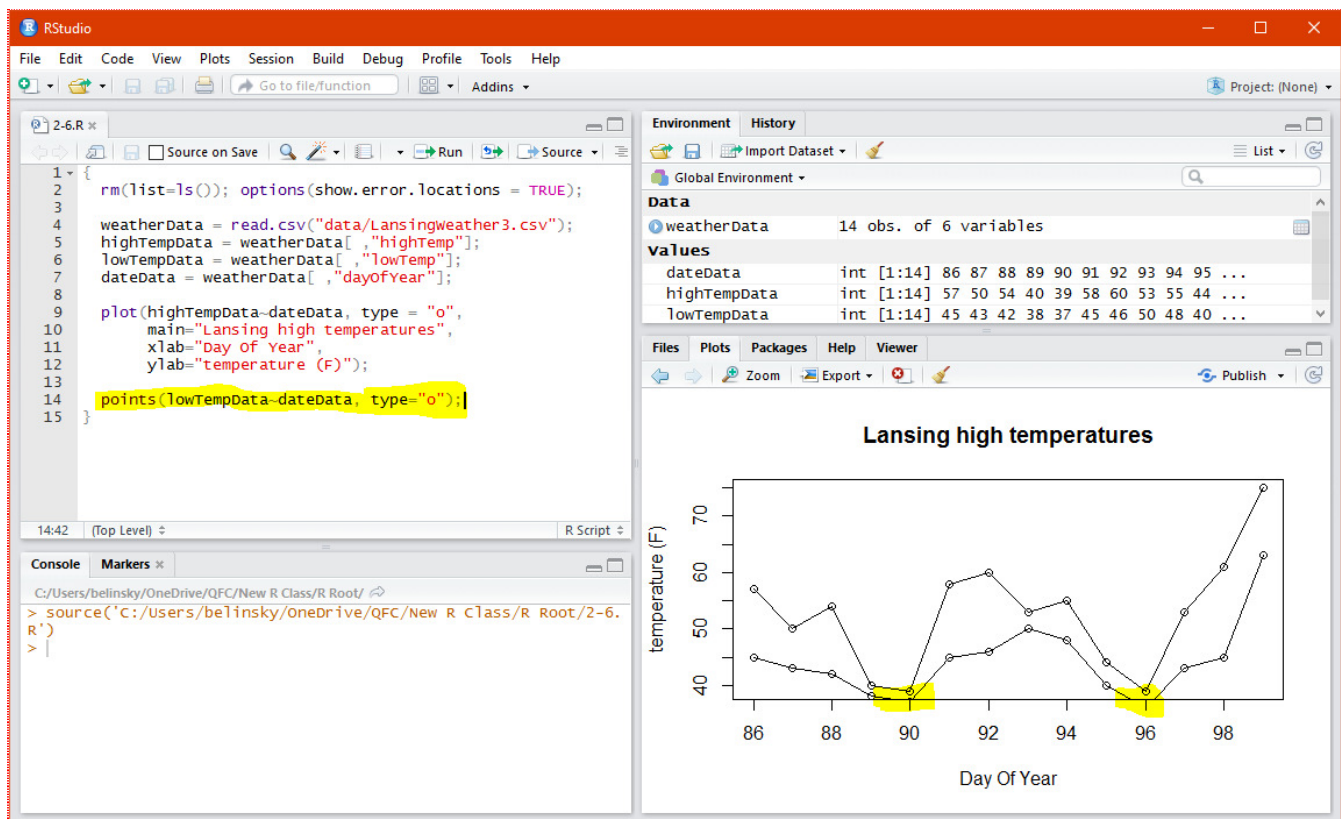


Fig 5: Plotting out both the high and low temperatures on one plot.

5.1 - Changing the limits of the axes

The low temperatures were added to the plot but some of the points are off the plot. This is because the

y-axis was not adjusted to accommodate the new values; the original y-axis limits were based on the minimum and maximum from the **highTempData** (approximately 40 to 75), but the limits do not adjust when new points are added, even if the points go beyond the y-axis limits.

We need to change the y-axis limits in the original **plot()** to accommodate the new temperature values. The parameters for changing the axis limits are **xlim** and **ylim**. In this case, we only need to change the **ylim** parameter. To change to **ylim** parameter, we set **ylim** to a vector of two values: **ylim = c(min, max)**. In this case we will use 30 and 80.

```
1 {
2   rm(list=ls()); options(show.error.locations = TRUE);
3
4   weatherData = read.csv("data/LansingWeather3.csv");
5   highTempData = weatherData[, "highTemp"];
6   lowTempData = weatherData[, "lowTemp"];
7   dateData = weatherData[, "dayOfYear"];
8
9   plot(highTempData~dateData, type = "o",
10        ylim=c(30,80),
11        main="Lansing high temperatures",
12        xlab="Day Of Year",
13        ylab="temperature (F)");
14
15   points(lowTempData~dateData, type="o");
16 }
```

Note that the y-axis now goes from 30 to 80 due to assigning values to the parameter **ylim**.

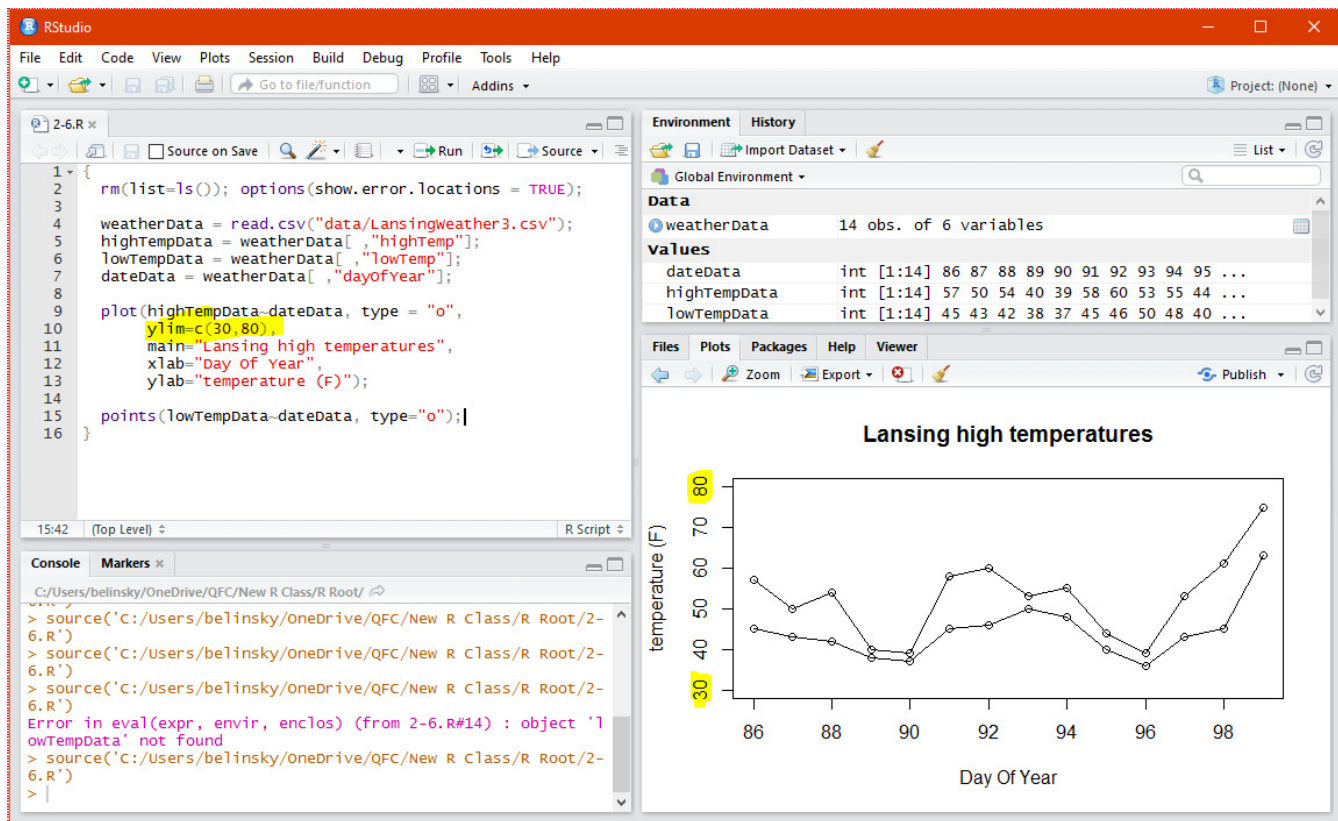


Fig 6: Changing the y-axis limits on a plot to accommodate new values

5.2 - Changing the color of the plots

We now have two lines and it would be nice to distinguish them. We are going to make the first plot blue and the second plot red using the **col** parameter.

```

1 {
2   rm(list=ls()); options(show.error.locations = TRUE);
3
4   weatherData = read.csv("data/Lansingweather3.csv");
5   highTempData = weatherData[, "highTemp"];
6   dateData = weatherData[, "dayOfYear"];
7
8   plot(highTempData~dateData, type = "o", col="red",
9        ylim=c(30,80),
10        main="Lansing high temperatures",
11        xlab="Day of Year",
12        ylab="temperature (F)");
13
14   points(lowTempData~dateData, type="o", col="blue");
15 }

```

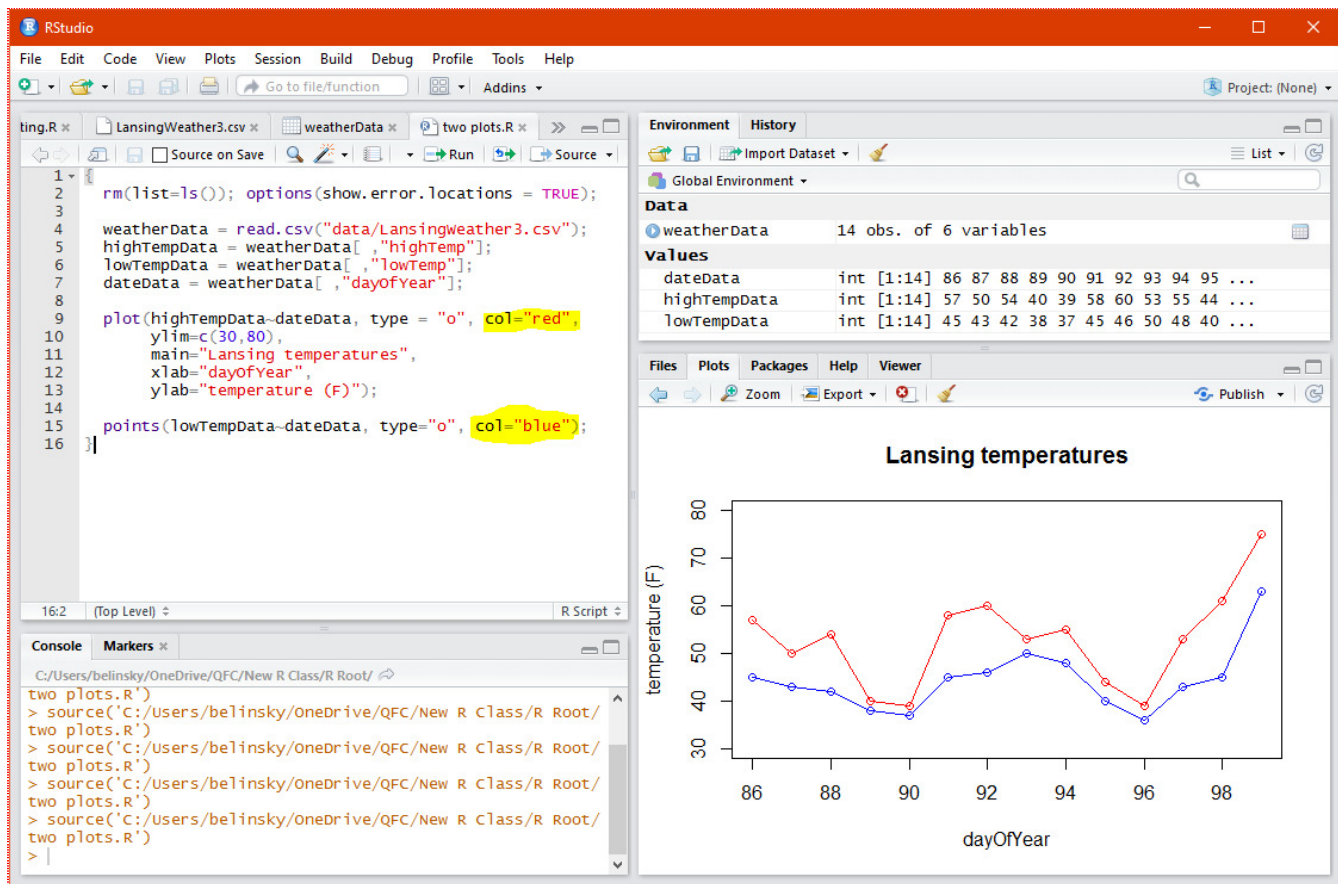



Fig 7: Changing the color of the lines in a plot

6 - Adding a legend

Lastly, we would like a way to annotate the different lines and the best way is to use a legend. In R the function to add a legend to a plot is **legend()**. **legend()**, like **plot()** and **points()** has numerous properties that can be modified. We can find the help page for **legend** by switching to the Help tab in the Plot Window and typing "legend" in the search bar (Fig 8).

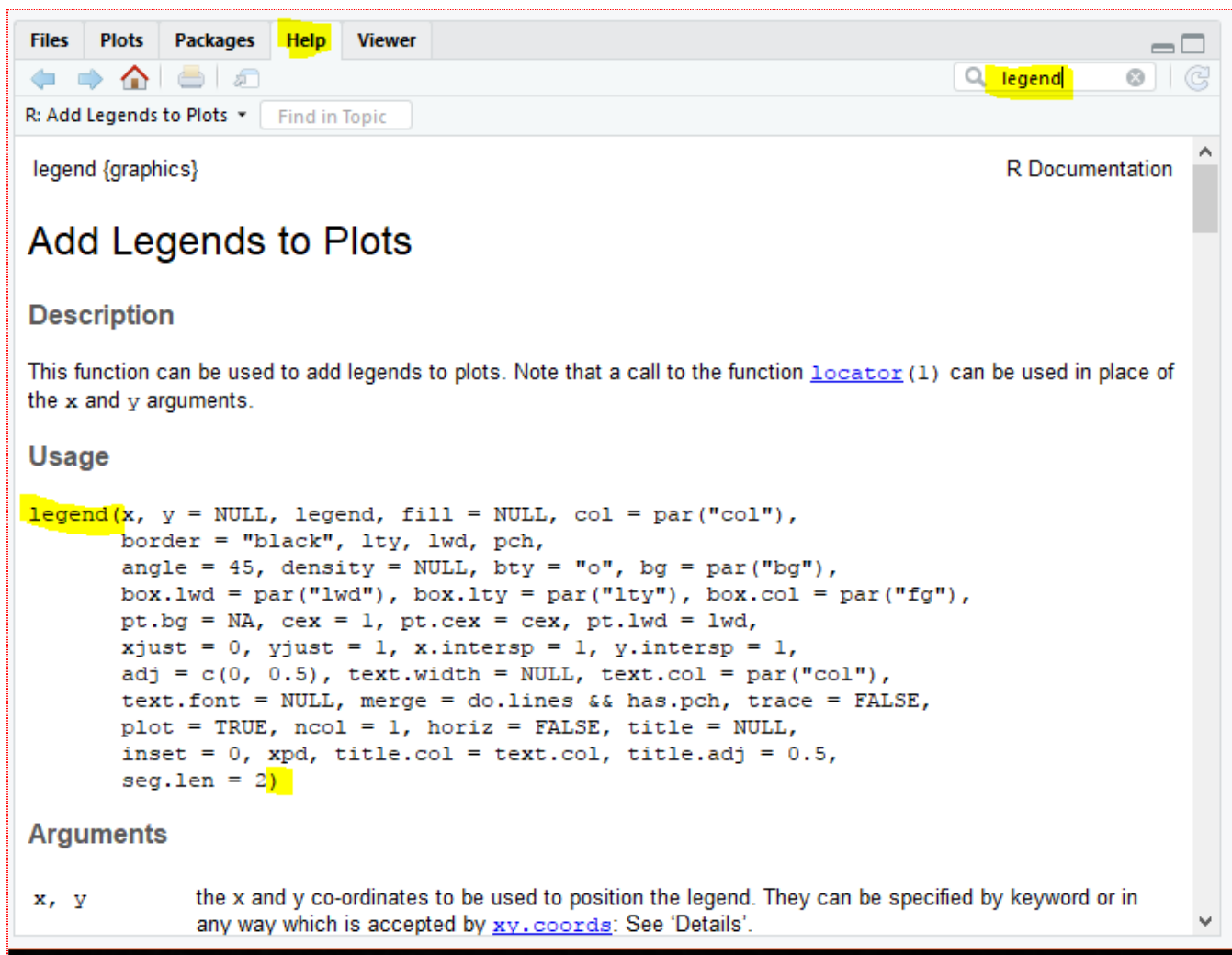


Fig 8: Searching for "legend" in the Help Window.

For this example we are only going to use the parameters that are needed to get the legend up and functional. They are:

- **x**: set the position of the legend
- **legend**: labels in the legend
- **lty**: line type (needed to put lines in the legend)
- **pch**: point type (needed to put points in the legend)
- **col**: color

Note: R does not provide the most intuitive parameter names!

We will start by positioning the legend and putting the labels in:

```
1 {
2   rm(list=ls()); options(show.error.locations = TRUE);
3
4   weatherData = read.csv("data/LansingWeather3.csv");
5   highTempData = weatherData[ , "highTemp"];
6   lowTempData = weatherData[ , "lowTemp"];
7   dateData = weatherData[ , "dayOfYear"];
8 }
```

```

9 plot(highTempData~dateData, type = "o", col="red",
10      ylim=c(30,80),
11      main="Lansing temperatures",
12      xlab="Day of Year",
13      ylab="temperature (F)");
14
15 points(lowTempData~dateData, type="o", col="blue");
16
17 legend(x="topleft", _____ # position
18        legend=c("High Temp", "Low Temp")); # labels
19 }

```

I positioned the legend in the topleft corner:

```
1 x="topleft"
```

There are nine values you can put in for position: "topleft", "topright", "top", "center", "right", "left", "bottomleft", and "bottomright".

The labels

```
1 legend=c("High Temp", "Low Temp")
```

are put in a vector and R recognizes each value in the vector ("High Temp", "Low Temp") as a separate label.

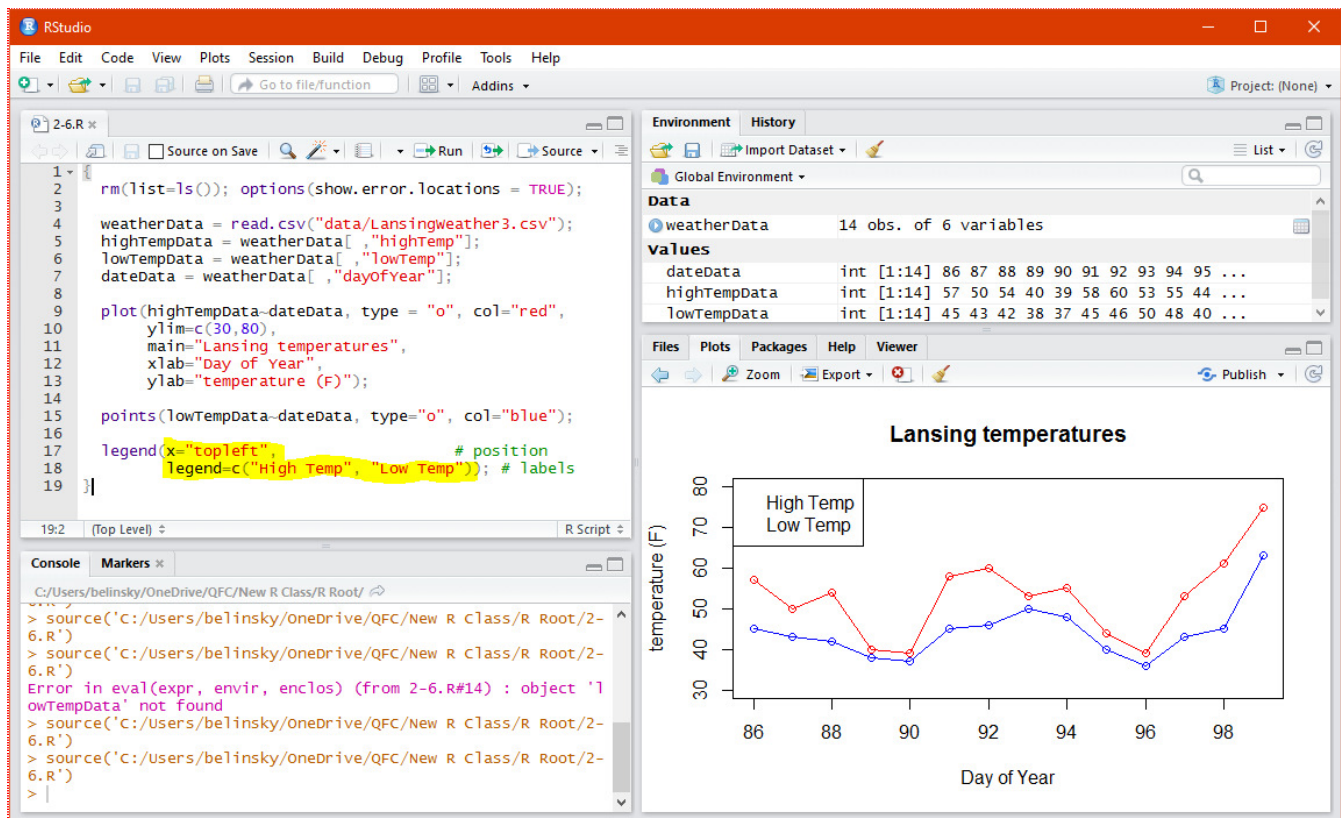


Fig 9: Legend placed in the plot

Now we want to put lines and points in the legend that look like the lines and points in the plot.

```

1 {
2   rm(list=ls()); options(show.error.locations = TRUE);
3
4   weatherData = read.csv("data/Lansingweather3.csv");
5   highTempData = weatherData[ , "highTemp"];
6   lowTempData = weatherData[ , "lowTemp"];
7   dateData = weatherData[ , "dayOfYear"];
8
9   plot(highTempData~dateData, type = "o", col="red",
10    ylim=c(30,80),
11    main="Lansing temperatures",
12    xlab="Day Of Year",
13    ylab="temperature (F)");
14
15   points(lowTempData~dateData, type="o", col="blue");
16
17   legend(x="topleft",          # position
18         legend=c("High Temp", "Low Temp"), # labels
19         lty=c(1,1),           # line type
20         pch=c(1,1),           # point type
21         col=c("red", "blue")); # colors
22 }

```

lty and **pch** need to have values if you want a line and/or a point next to the label.

```

1 lty = c(1,1),
2 pch = c(1,1)

```

means we will use the first line type and the first point type for both the **HighTemp** and **LowTemp** labels.

If we wanted to use the 3rd line type for High Temp and the 5th line type for Low Temp we would use:

```

1 lty = c(3, 5)

```

You can change the numbers and see how the lines and points change. A complete list of line and point types can be found at <http://www.statmethods.net/advgraphs/parameters.html>. Scroll down to the Plotting Symbols and Lines section of the webpage.

```

1 col = c("red", "blue")

```

mean the High Temp lines and points are red and the Low Temp lines and points are blue

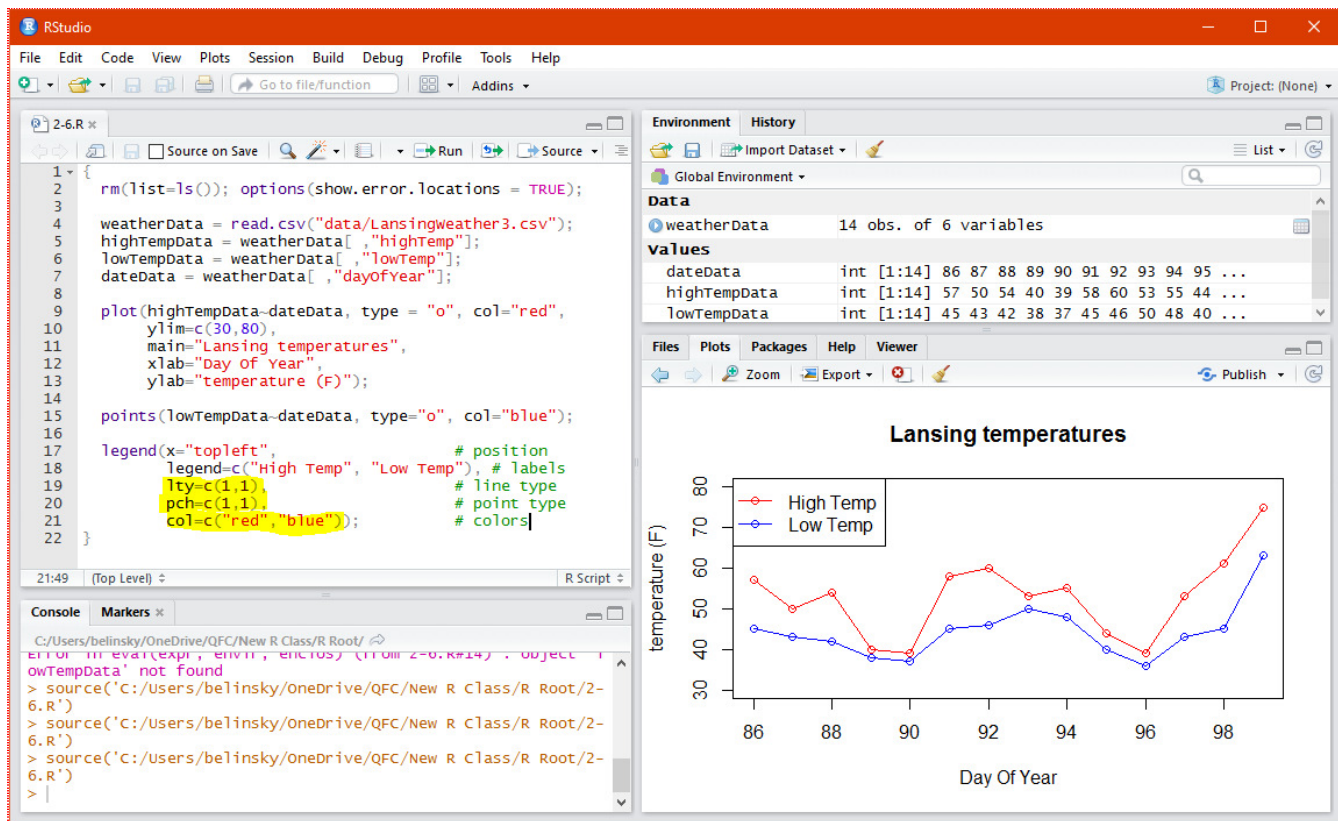


Fig 10: Legend added with lines and points

7 - Application

A) Produce a plot of **precipitation** vs **humidity** using the data from **LansingWeather3.csv**

- label the axes: humidity (%) and precipitation (inches)
- title the plot: Precipitation vs. Humidity
- make the plot points-only (no lines)

B) Add a **changeInTemp** vector to the **highTemp** and **lowTemp** plot created earlier in the lesson (Fig 10)

- **changeInTemp** is **highTemp** minus **lowTemp**
- adjust the axis so all points from all three plots can be seen
- add Change In Temp to the legend
- make the color of the Change In Temp line purple in the plot and the legend

Note: The legend will probably cover part of the plot -- we will deal with this in the next plotting lesson.

8 - Trap: A very common plot error

If you have plots in your script then an error you will commonly get in RStudio is:

Error in plot.new() (from <some file>.R##) : figure margins too large

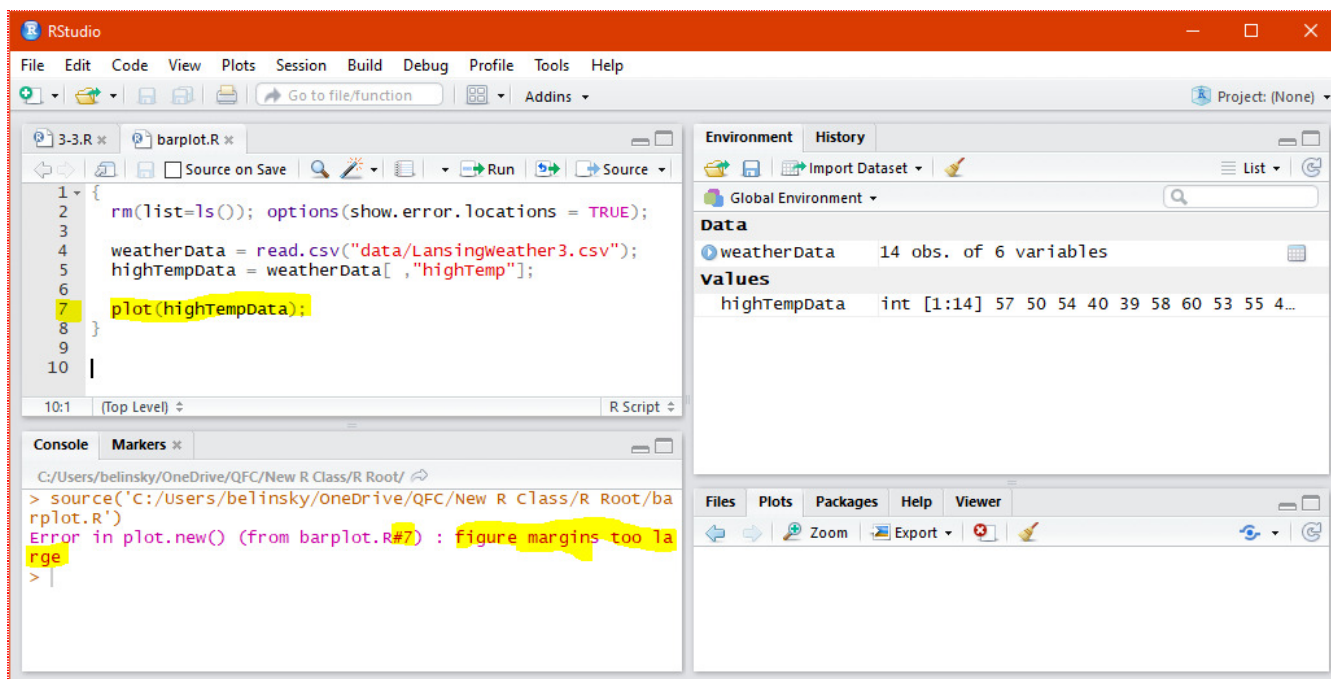


Fig 11: Figure margin too large error.

This is not an error in your script!

The error occurs because the RStudio Plot Window is too small to hold the plot. Technically speaking, the error is saying that the figure margins on your plot are larger than the Plot Window.

To fix the problem, just make the Plot Window larger.

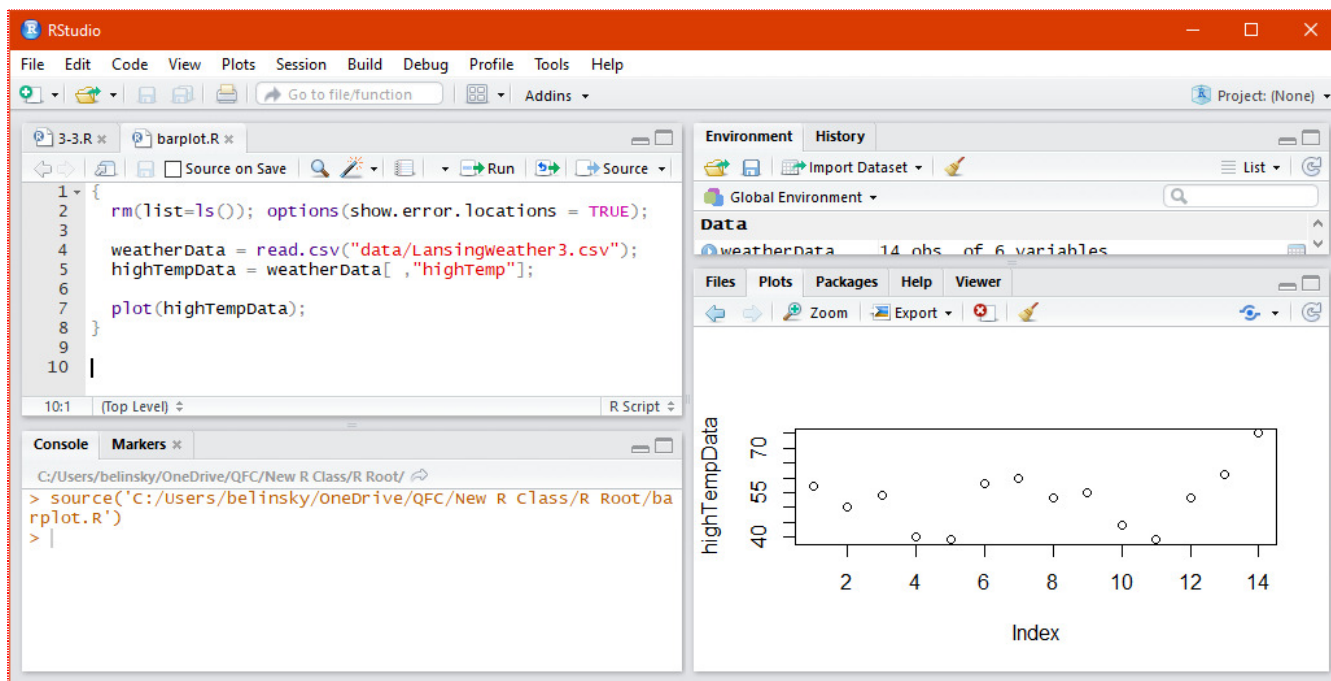


Fig 12: Increased Plot Window size and the error goes away.

Note: Just like with any error, all code after the error will not be executed. So, if your script is 100 lines long and you are plotting something on line 10, then lines 11-100 will not be executed.

One possible workaround for this error is described below in *Extension: Full-sized window for plots*.

9 - Extension: Full-sized window for plots

The Plot Window in RStudio can often be too small or cumbersome for our needs. Sometimes it is easier to open up a separate window the holds any plots created in the script.

We can do that by adding to the script:

```
1 windows(); # if this is a windows machine
```

or

```
1 quartz(); # if this is a MAC machine
```

Note: These functions must be executed before the plots in your script.

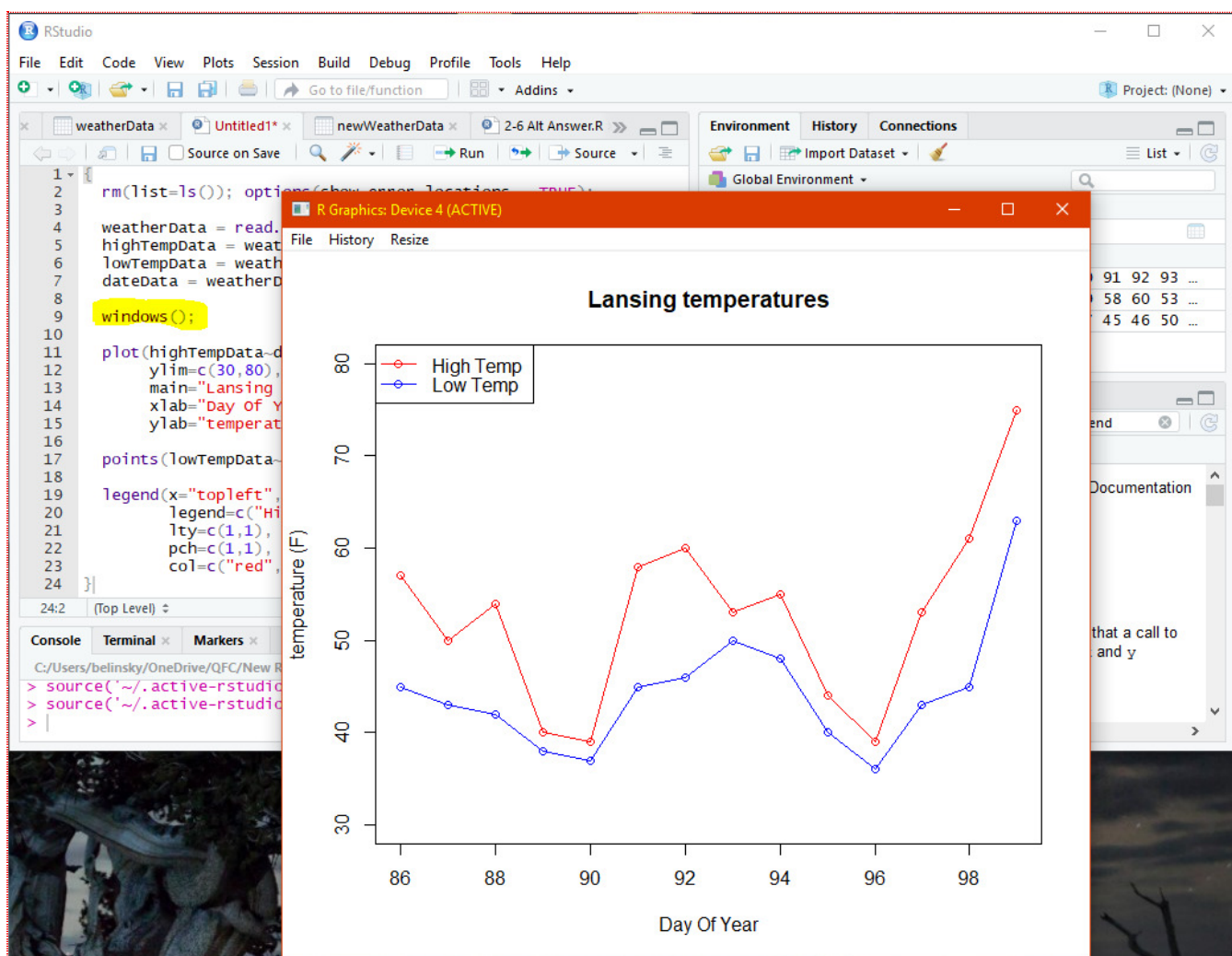


Fig 13: Opening plots in a separate window using *windows()*.