

Henry McCoy

Lab Tutorial

Ggplot tutorial:

In this tutorial, we will go over how to use the ggplot package in R. This package helps users create graphs in a simple and efficient manner, creating histograms, boxplots, dot plots, and more.

We will be using the ggplot package to analyze the “air quality” data set in R.

1. First, if you don't have the ggplot package, you want to download it into Rstudio by selecting “tools”, then “install packages”, then type “ggplot2” in the search box and download the package.
2. Now that you have the package downloaded, you can implement it using the following code:

```
4  
5 library(ggplot2)  
6
```

3. Next, enter and view the “air quality” data set with the following code:

```
3  
4 data("airquality")  
5 view(airquality)  
6
```

Make sure when using the “View” function to include a capital V. A common error is using a lowercase v, which will not work.

4. We can create a vector named “data” and assign the data from the “air quality” package to the that vector.

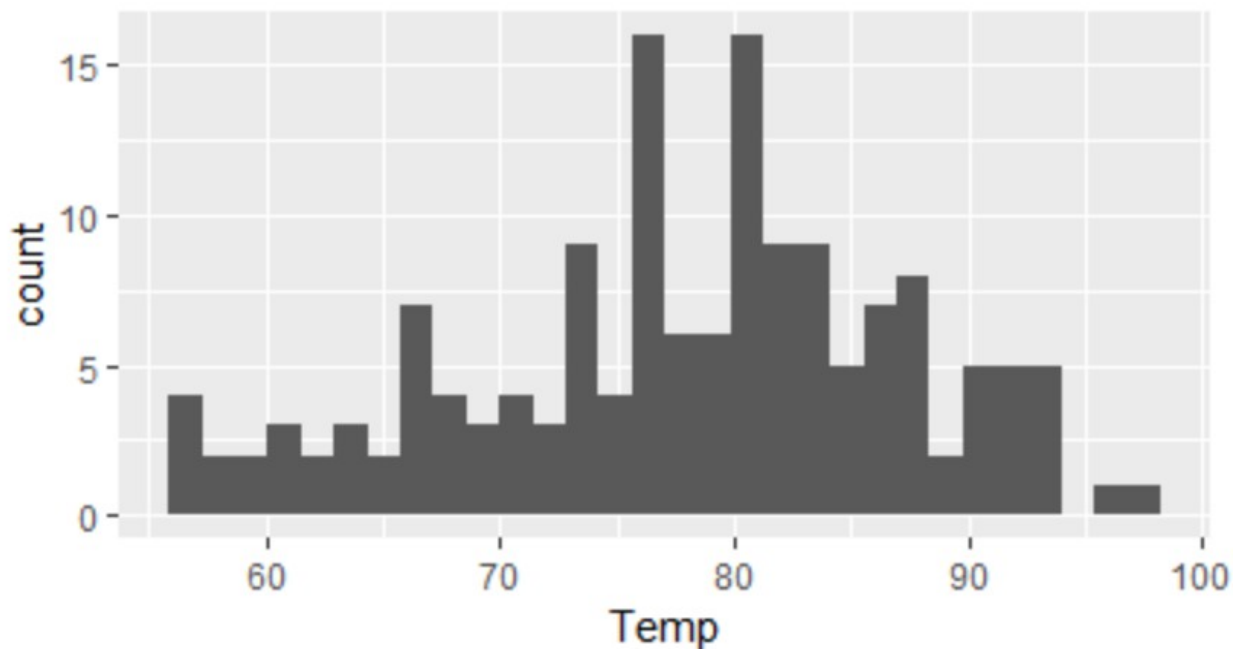
```
11  
12 data <- data("airquality")  
13
```

We will use this vector to represent our data when making graphs with ggplot.

- Using ggplot, we can make simple graphs, but we can expand on that and edit the graphs to include more detail. For example, let's use a histogram to measure the most frequent air temperature. This graph will have no added details or adjustments.

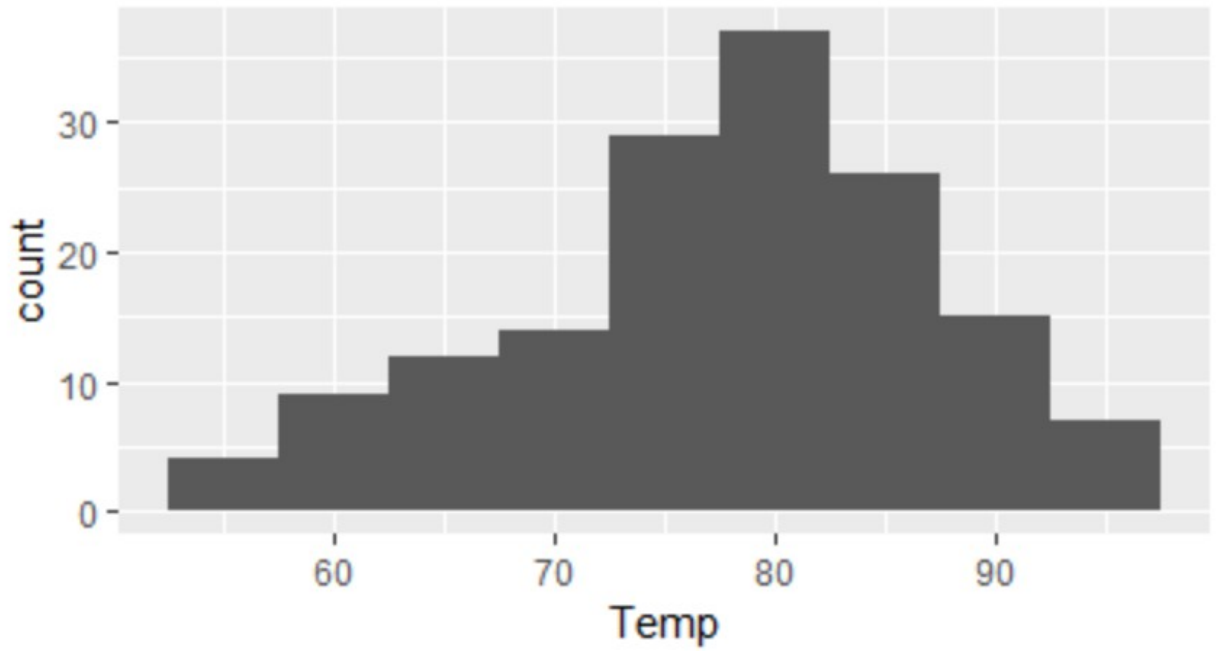
```
13  
14 ggplot(data = airquality, aes(x = Temp)) + geom_histogram()  
15
```

The graph shows as:



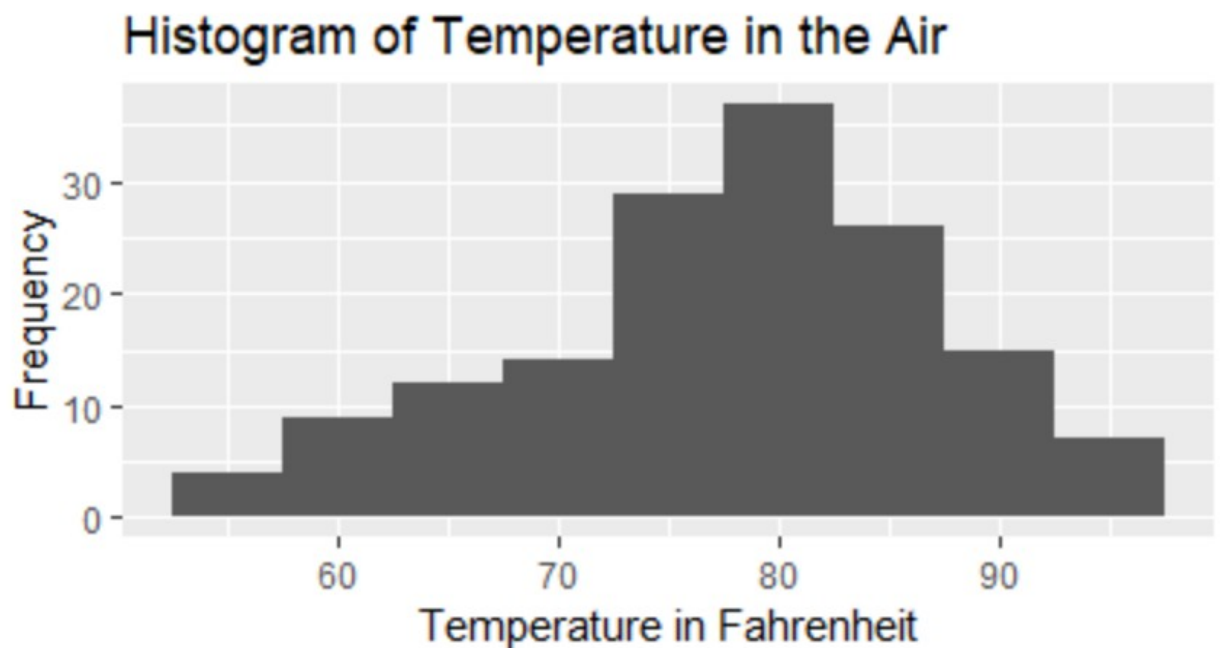
As you can see, the binwidth is at a default of 30, but that is too large. You can make the graph better by changing the binwidth to a more appropriate value, such as 5.

```
16  
17 ggplot(data = airquality, aes(x = Temp)) + geom_histogram(binwidth = 5)  
18
```



As you can see, the histogram looks much better. Using ggplot, we can add a title and labels to the x and y axis using “labs”.

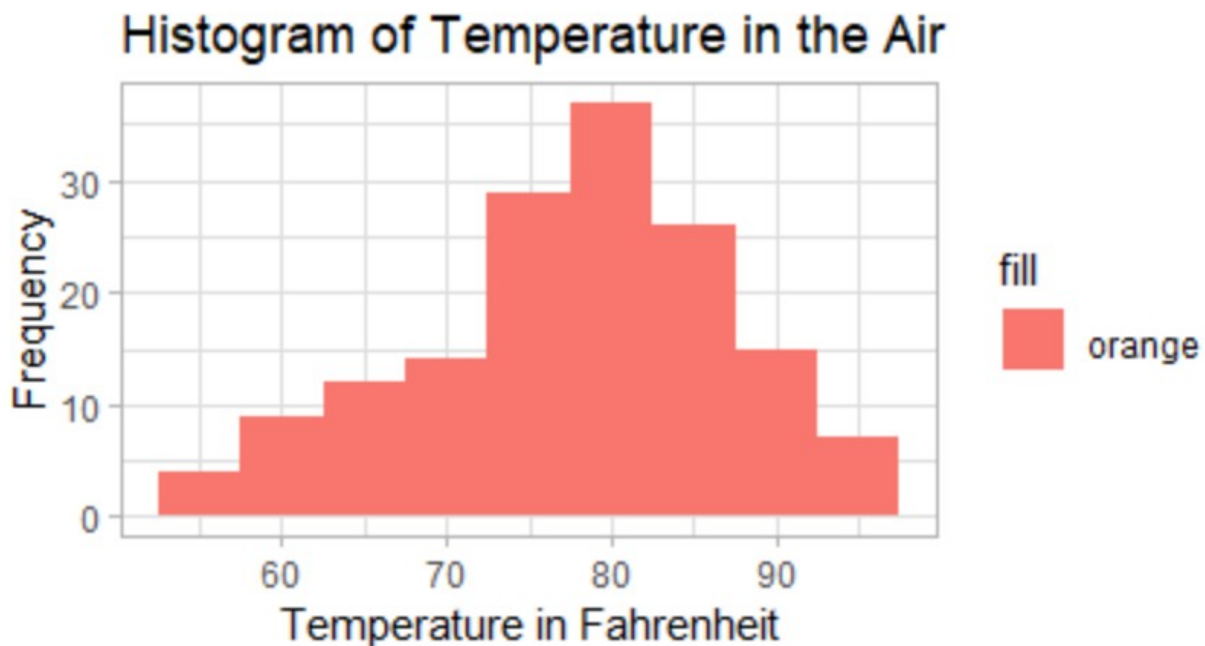
```
19
20 ggplot(data = airquality, aes(x = Temp)) +
21   geom_histogram(binwidth = 5, bins = 5)+
22   labs(title = "Histogram of Temperature in the Air",
23        x = "Temperature in Fahrenheit", y = "Frequency")
24
25
```



If you want to change the color and theme/background of the graph, you can do that through the “aesthetics” section of the code.

Next, we can change the color of the bars on the graph by using “fill” in the aes section of the code, and the theme for the background by using “theme”.

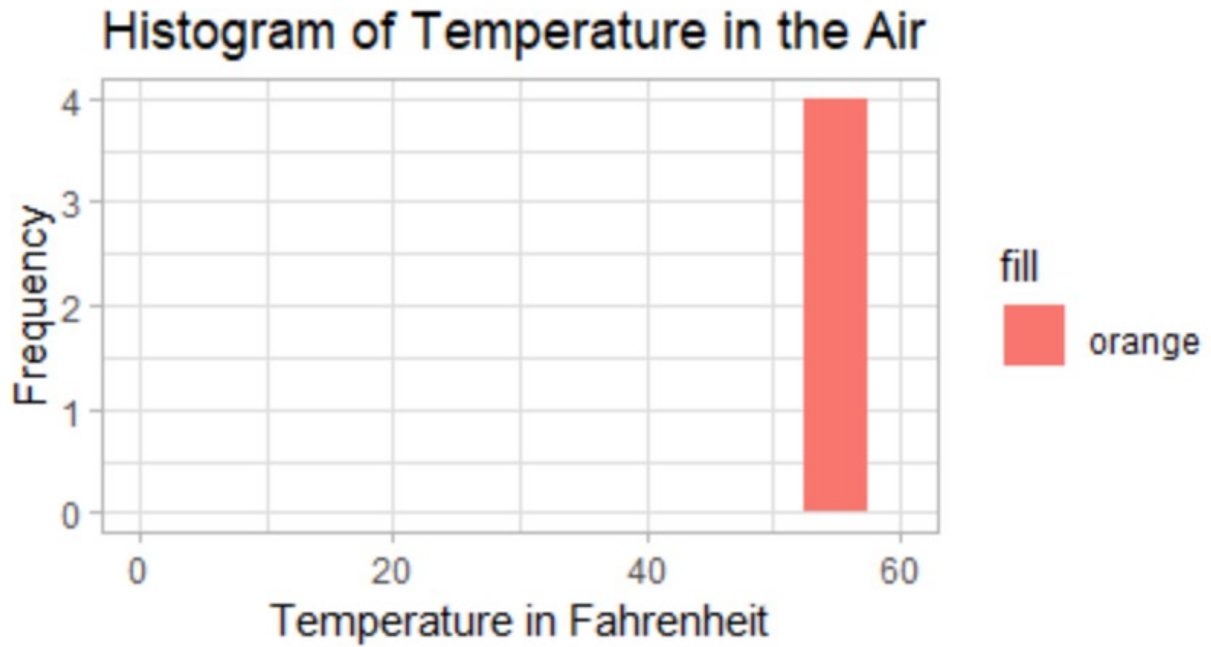
```
25  
26  
27 ggplot(data = airquality, aes(x = Temp, fill = "orange")) +  
28   geom_histogram(binwidth = 5, bins = 5) +  
29   labs(title = "Histogram of Temperature in the Air",  
30         x = "Temperature in Fahrenheit", y = "Frequency") +  
31   theme_light()  
32
```



Say we wanted to compare the temperatures from this city to another city that has colder weather.

We can change the range of the x-axis to correspond with the colder temperatures using “xlim”.

```
33  
34  
35 ggplot(data = airquality, aes(x = Temp, fill = "orange")) +  
36   geom_histogram(binwidth = 5, bins = 5) +  
37   labs(title = "Histogram of Temperature in the Air",  
38         x = "Temperature in Fahrenheit", y = "Frequency") +  
39   theme_light() + xlim(0, 60)  
40  
41  
42
```



The graph doesn't look good since the range of the x-axis is fit for colder cities. The temperatures in the city we are analyzing are generally warm, so the only bin with a frequency is one of the warmest bins.