Data visualization with ggplot2

Graphical Analysis of Biological Data

This web page is available as a PDF file

Read

R4ds: Chapter 3

Resources

Why visualize?

As a young scientist, you should read A protocol for data exploration to avoid common statistical problems (free access) by Zuur et al. 2010. The article describes the importance of visualizing your data before you begin your statistical analysis.

Many of the graphs you will make for this course will relate to the protocol outlined in Figure 1 of this paper.

This paper is part of the underlying philosophy of this course. Read it, even though you may not (yet) understand all of what they are discussing.

ggplot2

`ggplot2` is one of the `tidyverse` packages. `ggplot2` adopts the principle of a layered grammar of graphics, first developed by Leland Wilkinson in The Grammar of Graphics. The layered grammar of graphics allows you to build up graphs in layers, as you will learn in the assignment.

ggplot2 cheatsheet

Dipanjan Sarkar has a nice web page on the layered grammar of graphics used by ggplot2.

Designing good figures

Edward Tufte is a pioneer of innovative graphic design. His graphs maximize the data-ink ratio by minimizing chartjunk. Excel and PowerPoint are notoriously bad for chartjunk. Some of his ideas are unusual and debated but the overall theme of reducing unneeded "ink" to increase data signal is widely accepted and one that we will follow.

Another perspective on ChartJunk by Stephen Few.

Designing effective tables and graphs by Stephen Few of the Perceptual Edge. He has several articles that are worth looking through.

Data visualization: a practical introduction by Kieran Healy is a great reference for graph design with ggplot2.

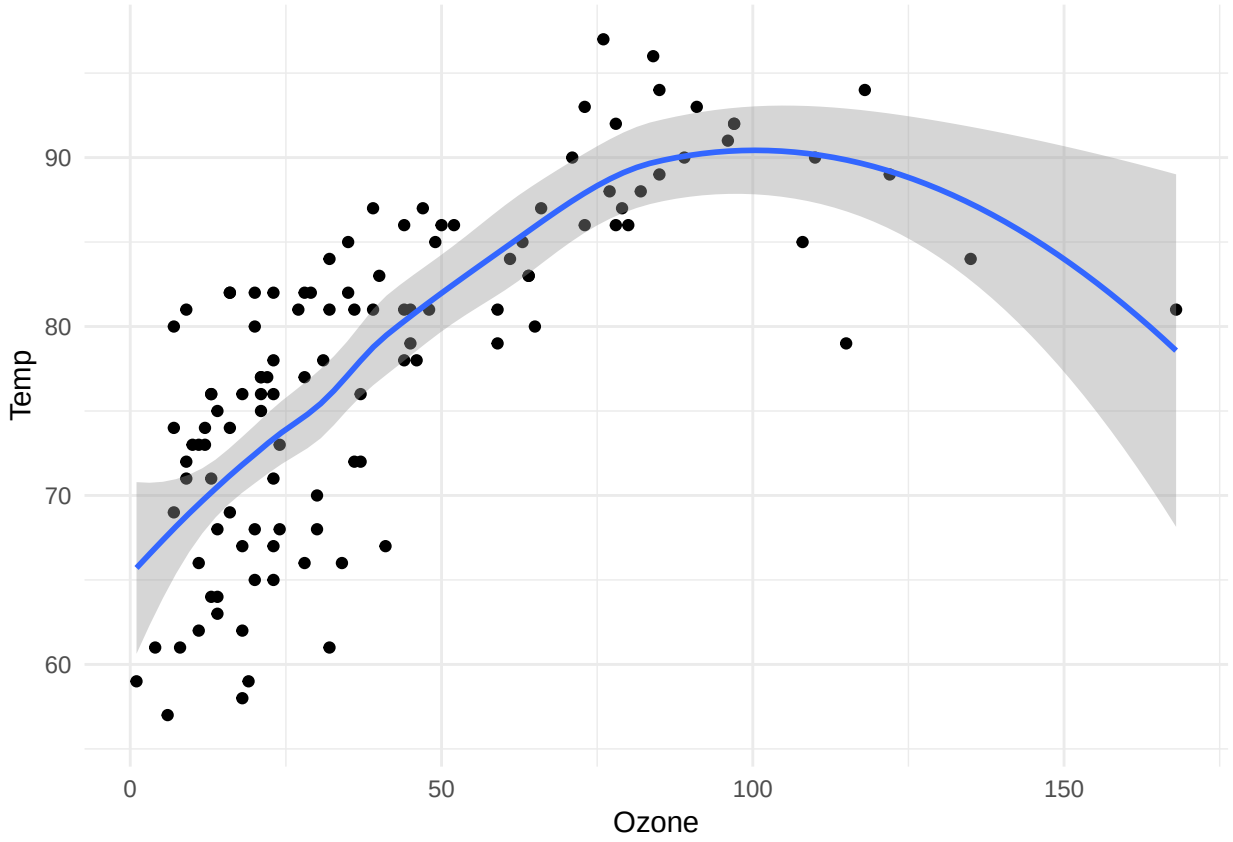
Fundamentals of Data Visualisation by Claus O. Wilke is another great reference for graph design. Although Dr. Wilke used ggplot2 for the figures, his book focuses on the elements of good design and not the code.

A hint or two...

- Use `? <function name` to get help on a particular function. For example, `?geom_point` (space or no space after the question `?` works) will show you the help file for `geom_point` in `ggplot2`.

- Place your cursor on a function name in the RStudio console or in a code chunk, and then press the F1 function key (on Mac laptops, you may have to press the "fn" key when you press F1.) This will also bring up the help for that function.
- Type your code. Do not copy and paste from the assigned reading. You will do yourself a disserve if you copy and paste. Type your code!
- `ggplot2` builds the graphs in layers. Make sure your line breaks between layers ends with the `+` so that your code runs properly, as shown below.

```
library(ggplot2)
ggplot(data = airquality, aes(x = Ozone, y = Temp)) +
  geom_point() +
  geom_smooth() +
  theme_minimal()
```



Now, go make some graphs.