Visualisation

It was the best of plots, it was the worst of plots

Sam Clifford

2019-10-31

# Introduction

## About this workshop

In the lecture we introduced visualisation with the histogram, - plots and other scatter plot techniques, and touched on Tufte’s principles of graphical excellence.

This prac will investigate the visual display of data and what makes a good and a bad graph.

* Assumed skills
  + Writing R code into a script file
  + Identifying things that are visually pleasing
* Learning objectives
  + Identifying things that are informative
  + Being able to critique a graph
  + Understanding why and how data is encoded and decoded visually
  + Understanding the subjectivity of what is aesthetically pleasing
* Professional skills
  + Creating high quality graphics

A reminder of expectations in the prac:

* Keep a record of the work being completed, both the R script and this document
* Allow everyone a chance to participate in the discussion
* All opinions are valued provided they do not harm others
* Everyone is expected to help out with completing the work, learning seldom occurs solely by watching someone else do work

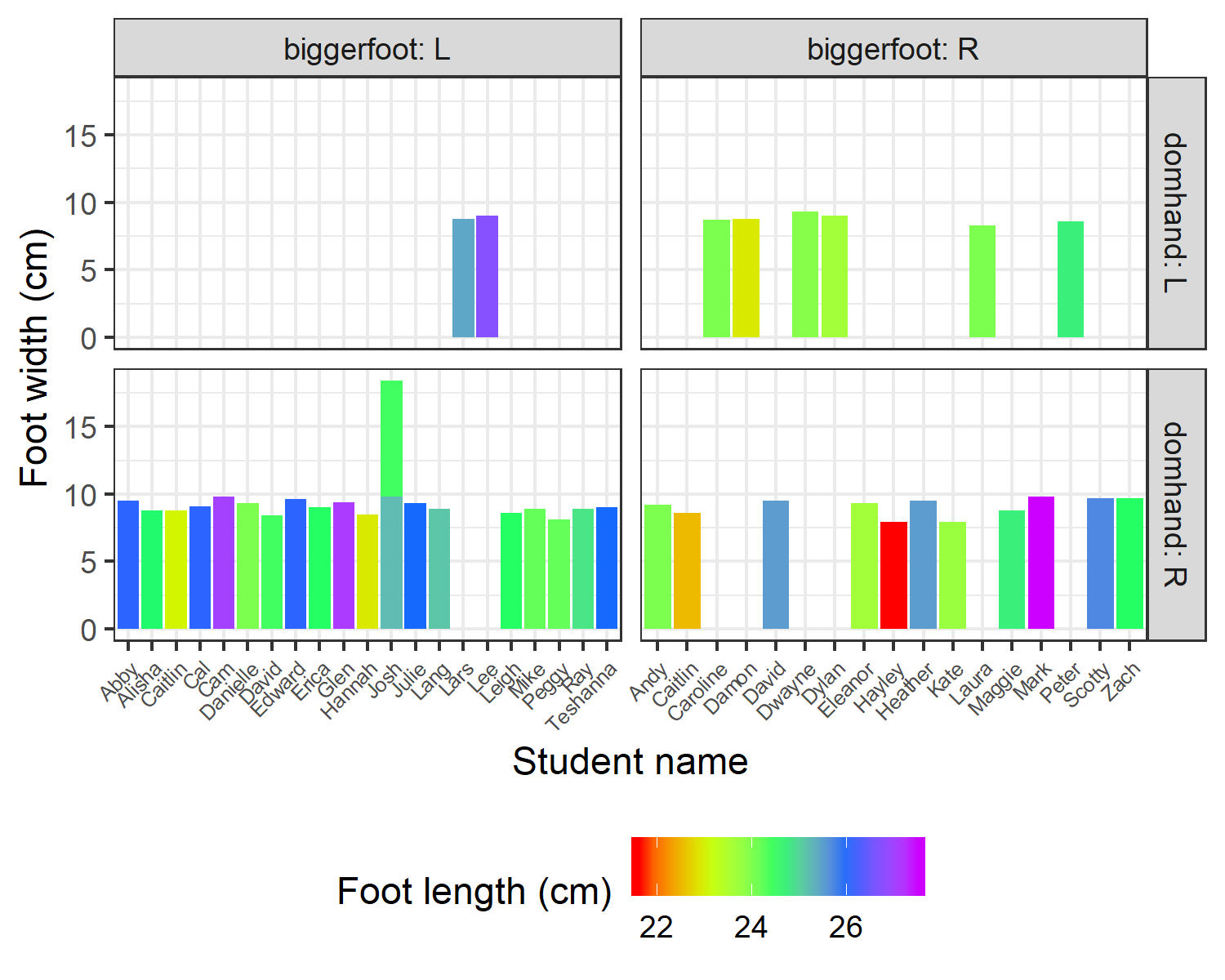
# Activity 1 - Building an attempt at a plot

We will be looking at a data set of the length and width of the feet of a class of fourth-graders in Ann Arbor, Michigan, collected in October 1997 (Meyer 2006). The experiment was performed to investigate whether boys do indeed have wider feet than girls.

library(tidyverse)  
library(mosaicData)  
  
data(KidsFeet)

**Exercise:** Copy and paste the code below to produce a bar plot that shows how foot length and foot width vary within the class of grade four students.

ggplot(data = KidsFeet, aes(x = name, y = width)) +  
 geom\_col(aes(fill = length)) +  
 facet\_grid(domhand ~ biggerfoot,   
 labeller = label\_both,   
 scales = "free\_x") +  
 theme\_bw() +  
 scale\_fill\_gradientn(colours = rainbow(5),  
 name = "Foot length (cm)") +  
 theme(axis.text.x = element\_text(angle = 45, size = 6,  
 hjust = 1, vjust = 1),  
 legend.position = "bottom") +  
 ylab("Foot width (cm)") +  
 xlab("Student name")



**Exercise:** Discuss, within your group, what you think is good and bad about this plot. Does it conform to Tufte’s principles of graphical excellence? Is it easy to interpret? Does it show the relationship we are interested in? List the three most important improvements that are needed for this graph to be useful.

**Answer:**

1. …
2. …
3. …

**Exercise:** As a group, discuss what you think each line of code in the above block does. You may wish to answer here or leave comments in your code (everything after a # is a comment)

**Answer:**

# Activity 2 – Making a better graph

Based on the ideas discussed, build a graph which your group believes better shows how the relationship between foot length and width varies within the class. You may choose to either modify the code given above or create your own graph from scratch. Make sure your code is written in your script file with appropriate comments.

Some things you may wish to consider:

* fixing up the axis labels
* a relevant title
* a different theme
* different plotting geometries
* different aesthetic options for colour, shape, etc.

You may wish to sketch the graph by hand before attempting to write the R code to generate it. This will help you and your group come to an agreement about the plot you want to make and will help the tutors understand what you’re aiming for when you ask them for help.

If you get stuck, look at the [ggplot2](https://ggplot2.tidyverse.org/reference/) documentation or ask a tutor.

**Exercise:** What changes did you make to improve the graph?

**Answer:**

1. …
2. …
3. …

**Delete this text and include your plot here.**

# Activity 3 – Making a worse graph

Make a new graph as in the previous activity but make it as bad as possible while still attempting to honestly show the information (i.e. don’t add things to the plot which can’t be derived from the variables in the plot.

What are the principles of graphical excellence and how can we go against them to make a truly terrible plot? Think about what was bad about the bar plot provided earlier. Consider abusing the ability to map graphical options (e.g. color, fill, line type, point size) to our variables of interest.

**Exercise:** What changes did you make to make the graph worse?

**Answer:**

1. …
2. …
3. …

**Delete this text and include your plot here.**

# Activity 4 – Group discussion

Have participants present their best (and/or worst) graph from the last activities. What did they identify as good and bad and how has each group attempted to present the relationship?

# Tidy up

Make sure you save this Word document, your R script, and any graphs you have produced and email them to the entire group. Restart your CoW when you are finished, put the chairs back neatly around the desk, leaving the room in a better condition than you found it. Email your worst graph to [Dr Sam Clifford](mailto:sam.clifford@lshtm.ac.uk).

# Further reading

A lot of the key ideas in data visualisation arose with Tufte (1983), and are summarised by Pantoliano (2012). [Tufte’s website](https://www.edwardtufte.com/tufte/) is well worth exploring, particularly the discussion on how the visual presentation of information could have helped avert the *Challenger* disaster (Tufte 1997).

Some of the history of data visualisation is summarised well by Friendly (2005) and Friendly (2006).

For some more guidance on using ggplot2 for data visualisation, check Chapter 3 of Wickham and Grolemund (2017), the RStudio cheatsheets (RStudio 2012), and Chang (2017).

Chang, Winston. 2017. *R Graphics Cookbook: Practical Recipes for Visualizing Data*. 2nd ed. O’Reilly Media.

Friendly, M. 2005. “Milestones in the History of Data Visualization: A Case Study in Statistical Historiography.” In *Classification: The Ubiquitous Challenge*, edited by C. Weihs and W. Gaul, 34–52. New York: Springer. <http://www.math.yorku.ca/SCS/Papers/gfkl.pdf>.

———. 2006. “A Brief History of Data Visualization.” In *Handbook of Computational Statistics: Data Visualization*, edited by C. Chen, W. Härdle, and A Unwin. Vol. III. Heidelberg: Springer-Verlag. <http://www.datavis.ca/papers/hbook.pdf>.

Meyer, Mary C. 2006. “Wider Shoes for Wider Feet?” *Journal of Statistics Education* 14 (1). American Statistical Association.

Pantoliano, Mike. 2012. “Data Visualization Principles: Lessons from Tufte.” 2012. <https://moz.com/blog/data-visualization-principles-lessons-from-tufte>.

RStudio. 2012. “RStudio Cheat Sheets.” 2012. <https://www.rstudio.com/resources/cheatsheets/>.

Tufte, Edward. 1983. *The Visual Display of Quantitative Information*. Graphics Press.

Tufte, E. R. 1997. “Visual and Statistical Thinking.” In *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press. <https://www.edwardtufte.com/tufte/books_textb>.

Wickham, Hadley, and Garrett Grolemund. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O’Reilly Media. <http://r4ds.had.co.nz>.