Lecture 3. Data visualisation

Nic Rivers

Advanced quantitative research methods, API6319
Fall 2019

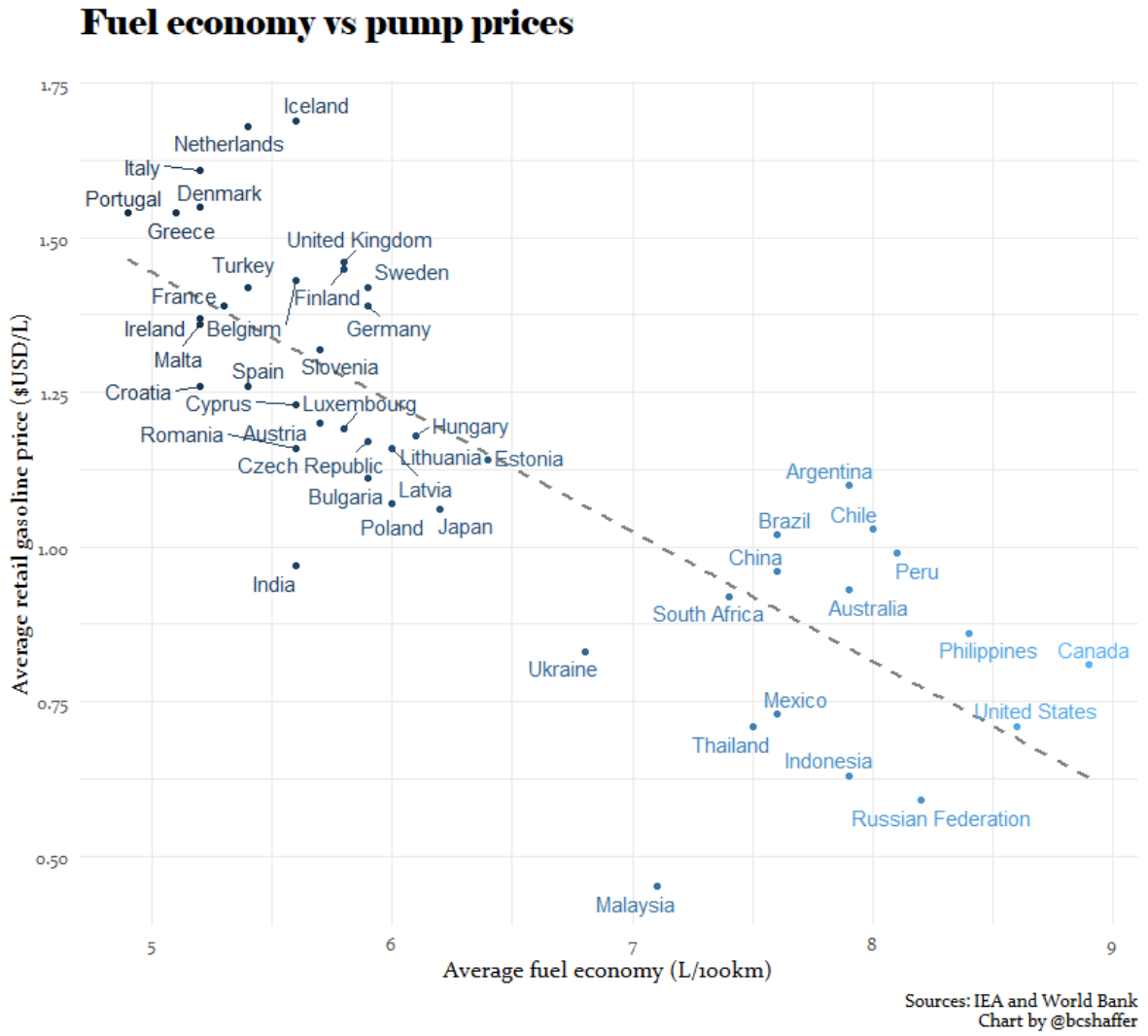
Introduction

- Effective visualisation of data can be an extremely effective way to get your point across:

  A picture is worth a thousand words.

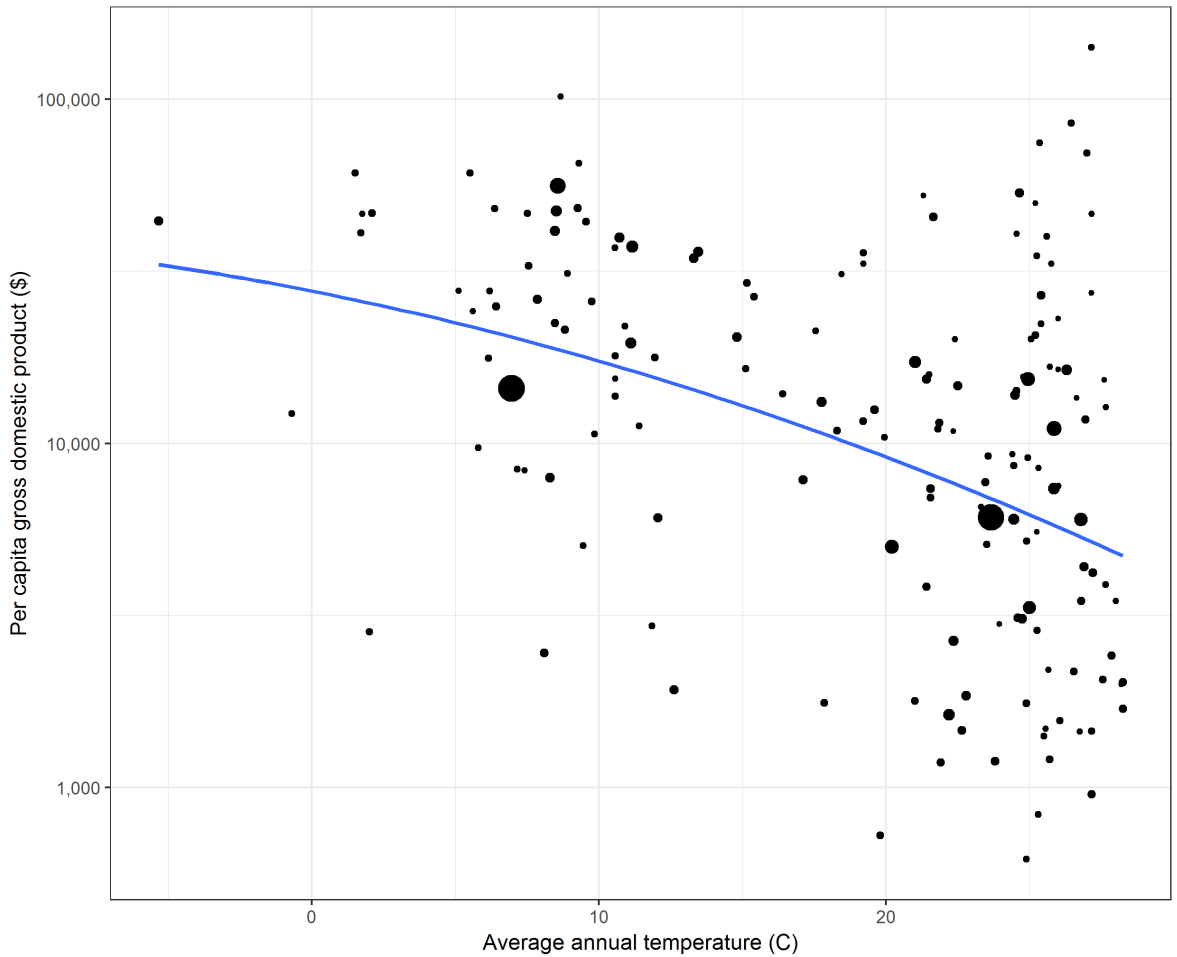
- Data visualisation can help to motivate your analysis.
- Data visualisation helps your reader draw their own conclusions from the data (which support your analysis/writing).
- Data visualisation is also very useful for you to understand patterns/problems in the data, and to think of an effective research design.

Example: scatter plot



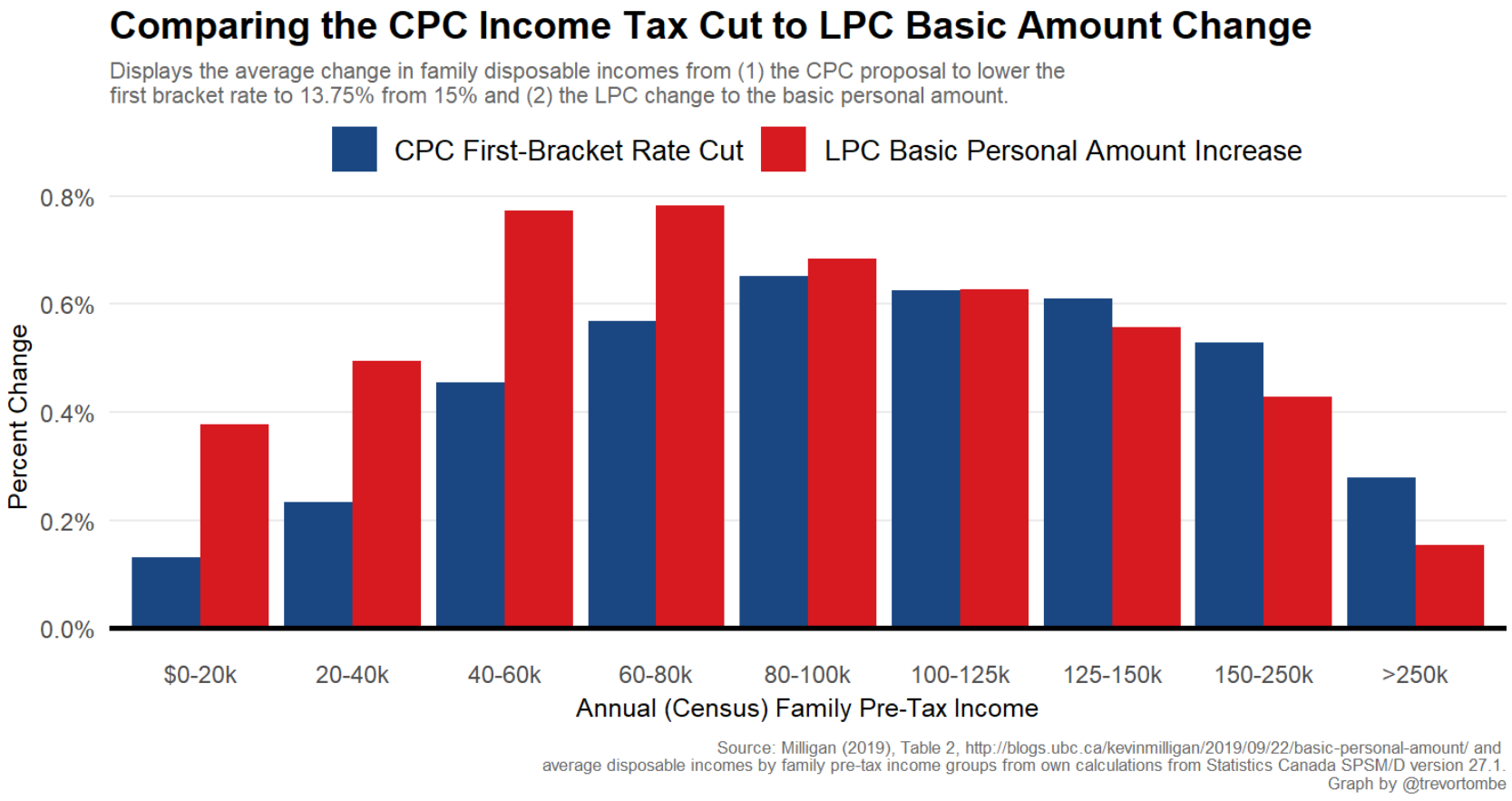
Example: scatter plot

Figure: Higher temperatures are associated with lower per capita output

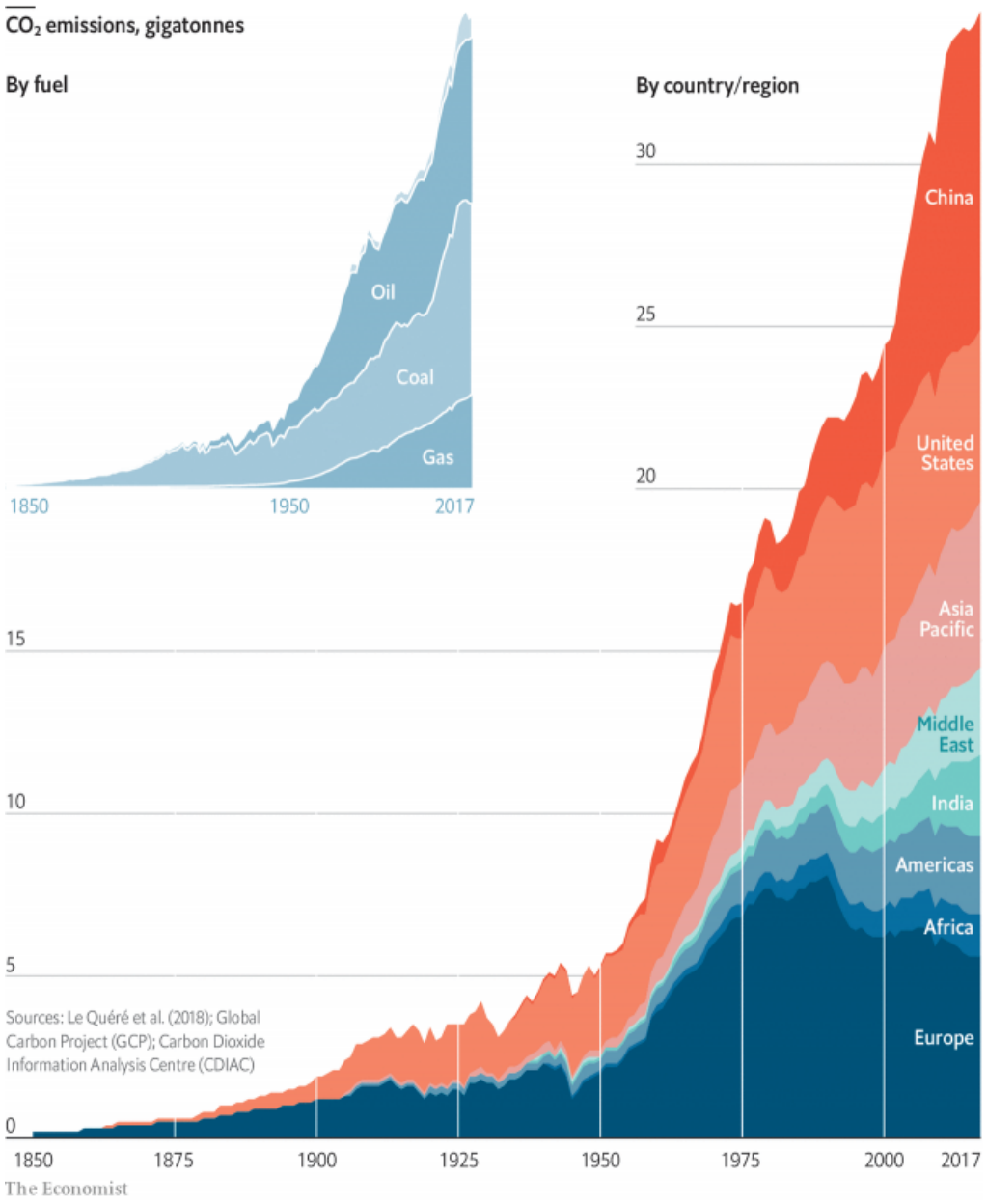


Point size is proportional to population

Example: bar plot



Example: line/area plot



Rules for data visualisation

Rules for data visualisation

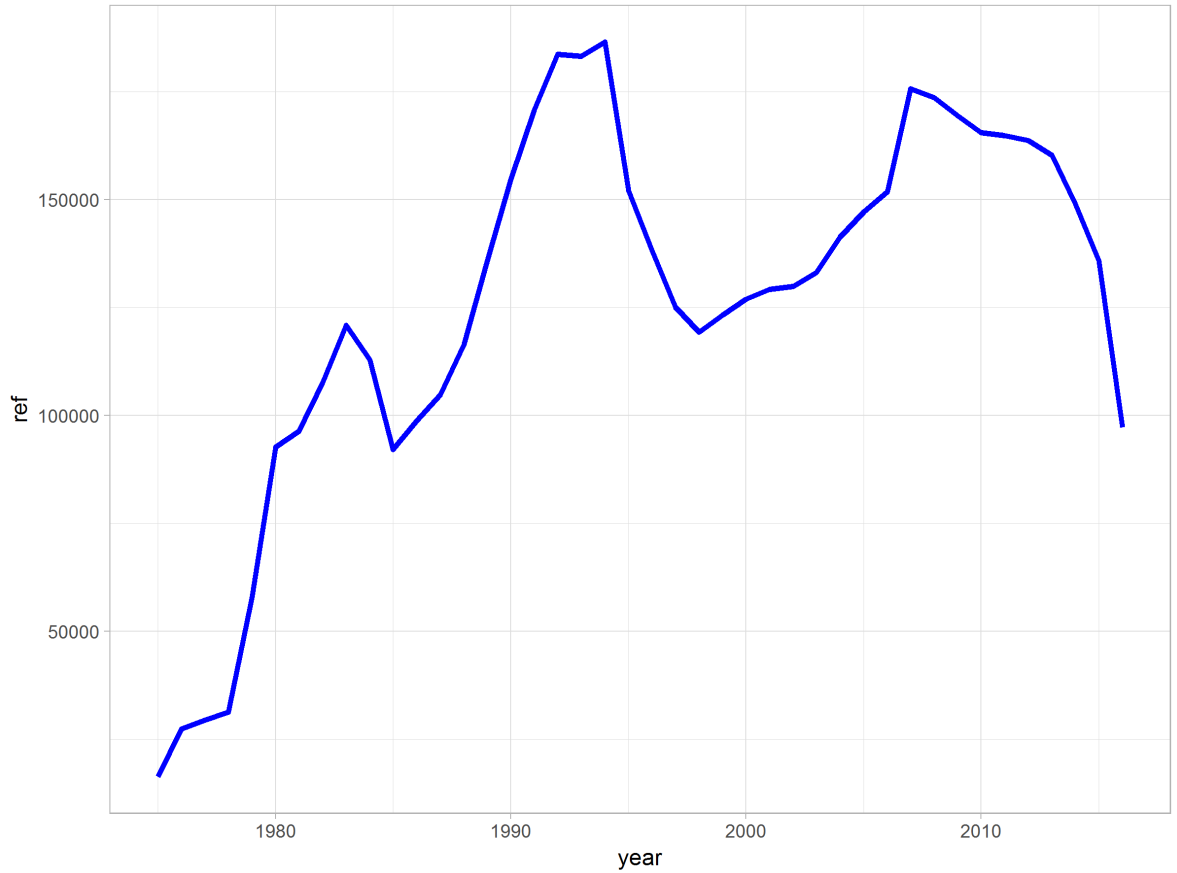
- There are no rules.

Rules for data visualisation

- There are no rules.
- But, there are some useful guidelines.
  - Provide clear labels that make it easy to understand (title, axis labels, labels for different colours/lines/etc.)
  - Use annotations to help tell a story.
  - Avoid chart junk (minimize ink used). Avoid 3d figures.
  - Don't make the figure too complex.
  - Use the right data for the job.
- Emulate figures you find helpful.

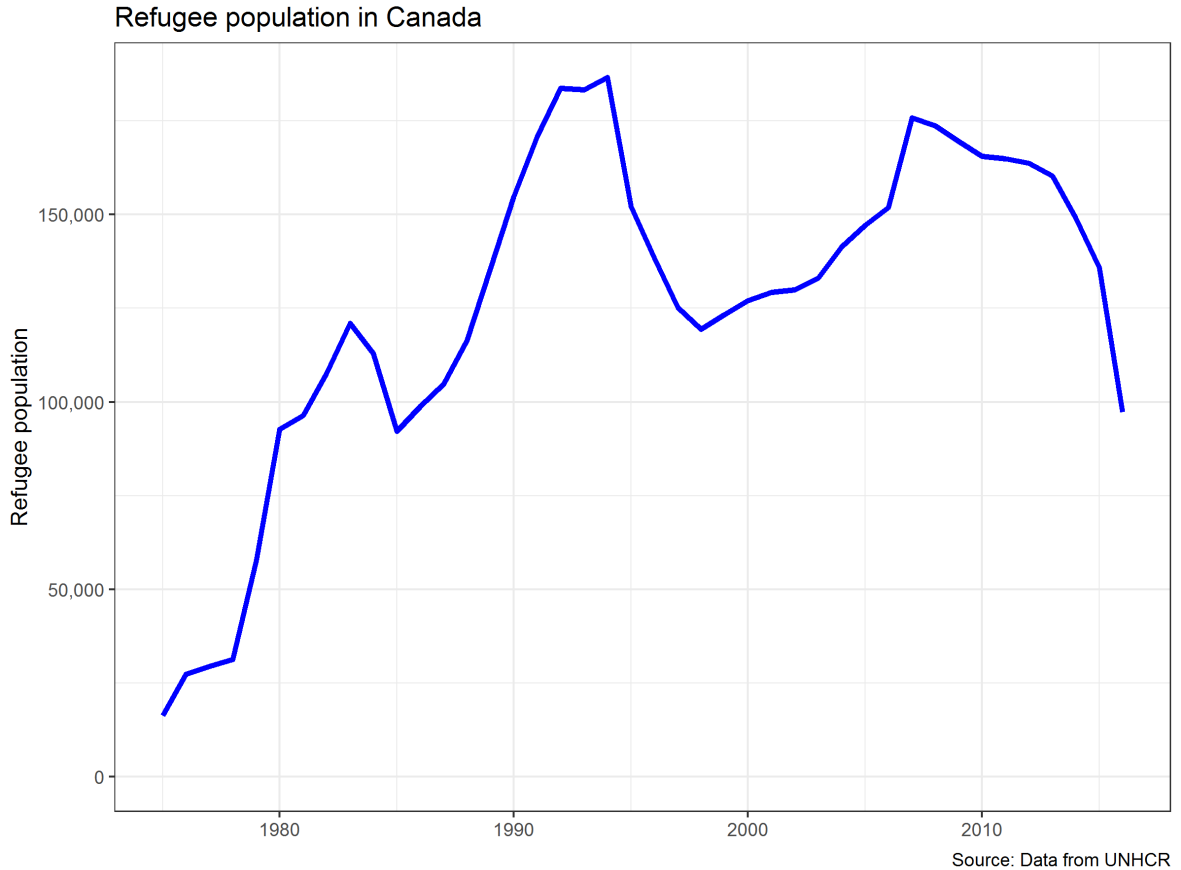
Labeling: don't accept the defaults!

What is this telling us?



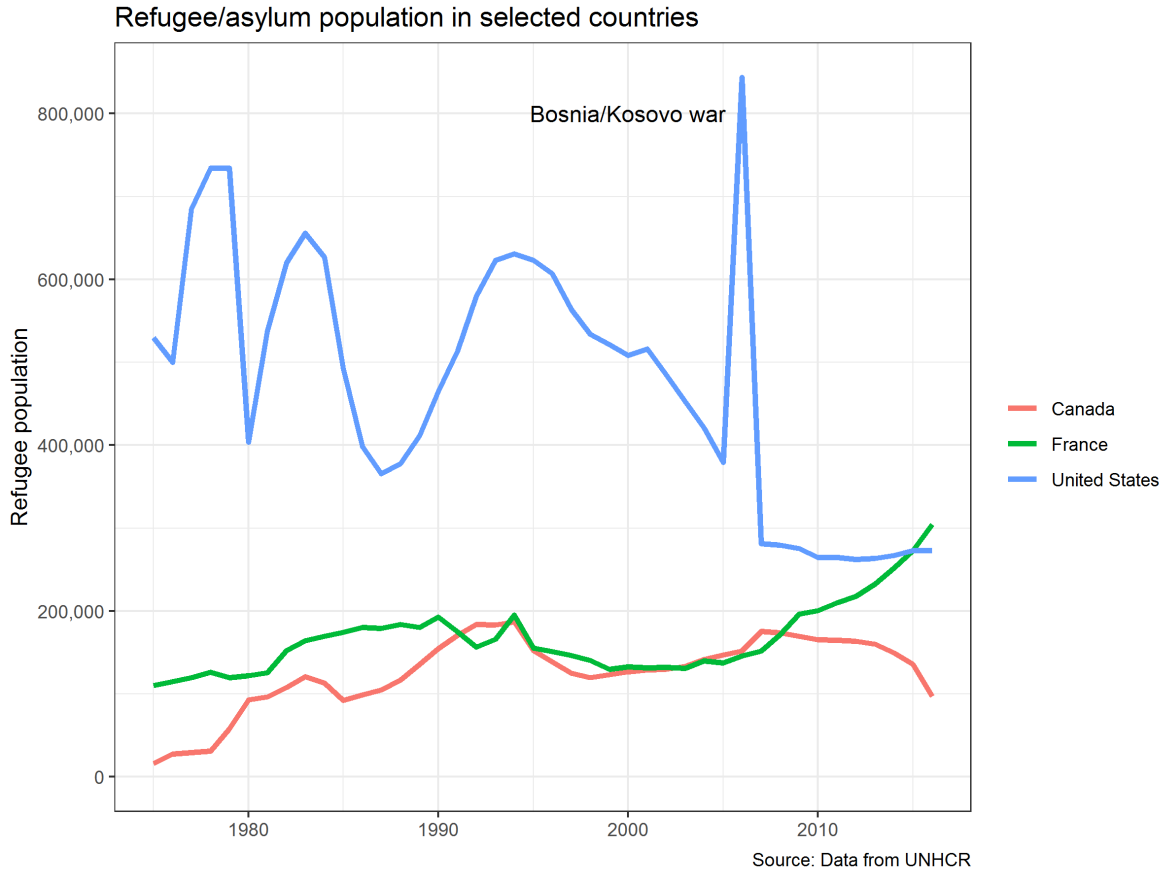
Labeling

Much better



Labeling

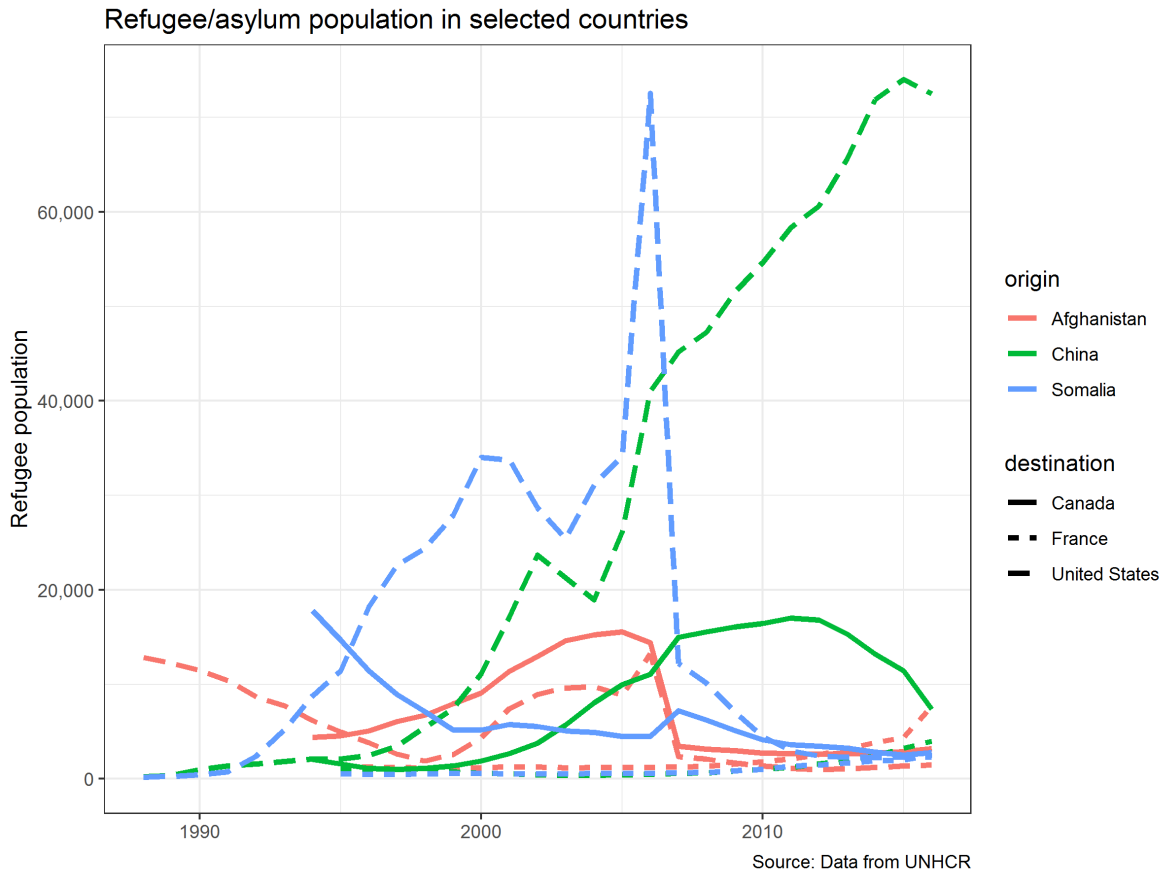
Use annotations to draw attention.



No chart junk

Not too much complexity

It's hard to process lots of different dimensions (scales) of a graph. Here there are 4: number of refugees (y), year (x), origin country (colour), destination country (linetype). Somewhat hard to process.



Emulate useful figures

Lots of sources of good figures. E.g., NYT (https://www.nytimes.com/interactive/2015/02/23/business/economy/the-changing-nature-of-middle-class-jobs.html)
R gallery (includes code) (https://www.r-graph-gallery.com/)
Another R gallery (with code) (http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html)

Types of figures

Type of figure is (partly) determined by:

- Type of data (numeric, categorical (factor))
- Number of variables

Some key graphs:

- 1 numeric variable: density, histogram
- 1 categorical variable: barplot
- 2 numeric variables: scatterplot
- 1 numeric/1 categorical: bar, boxplot