

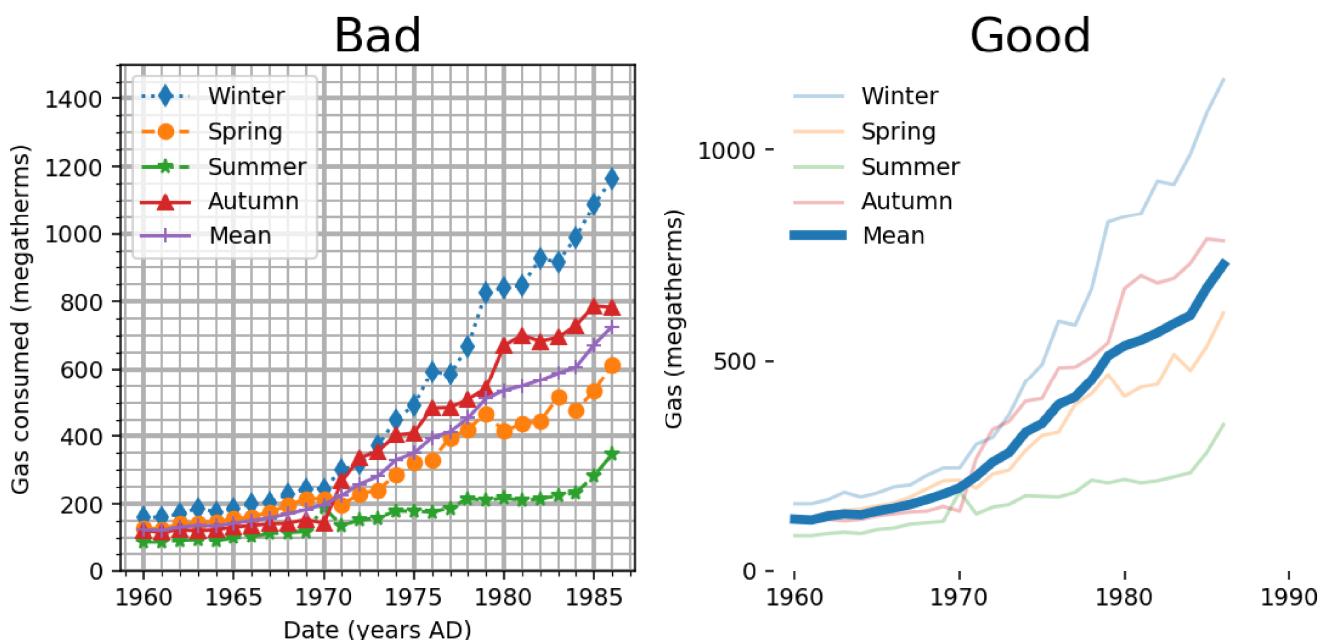
Criticising visualisations

Avoid chartjunk

Tufte proposed the principle of the **data-to-ink ratio**. As little "ink" (that is, markings) should be used to communicate as much data as possible. A simple, clean visualisation is less confusing and more aesthetically pleasing than a jumbled mess of unnecessary **chartjunk**.

Typical chartjunk includes:

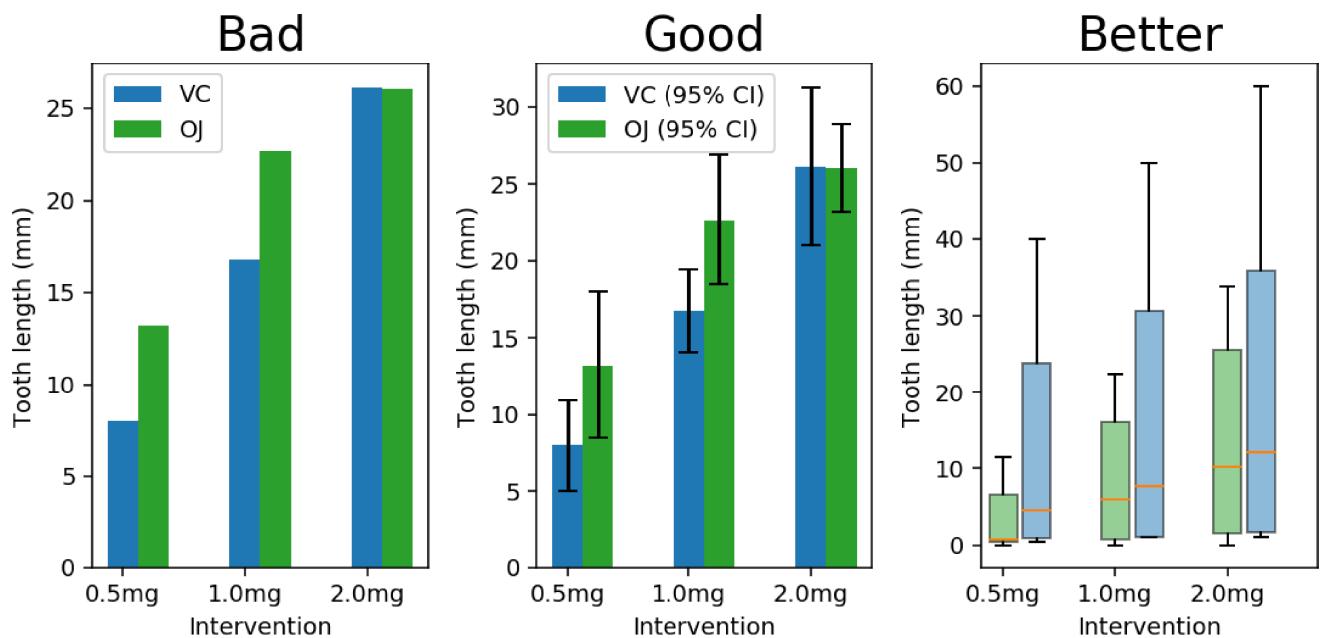
- Frames around plots
- Obtrusive grids
- Excessive marker or line styles
- Lack of visual distinction between important and unimportant elements
- Excessive tick marks
- Excessive annotations
- Clipart (just say no to clipart)



[**Left:** a dense, messy plot with lots of extraneous elements, including a dense grid, excessive tick marks and unnecessary markers. **Right** the same plot, in a clean, style with just enough elements to make the graph clear and readable.]

Represent uncertainty clearly and honestly

- Include measures of spread of a sample if these are relevant to comparing values.
- Bar charts should have error bars if they represent measures of central tendency (e.g. means, medians).
- Use Box plots or similar techniques (e.g. violin plots) if the distribution of values is useful in making judgements about differences between groups.
- Ribbon plots can be used to extend line geoms to include uncertainty.

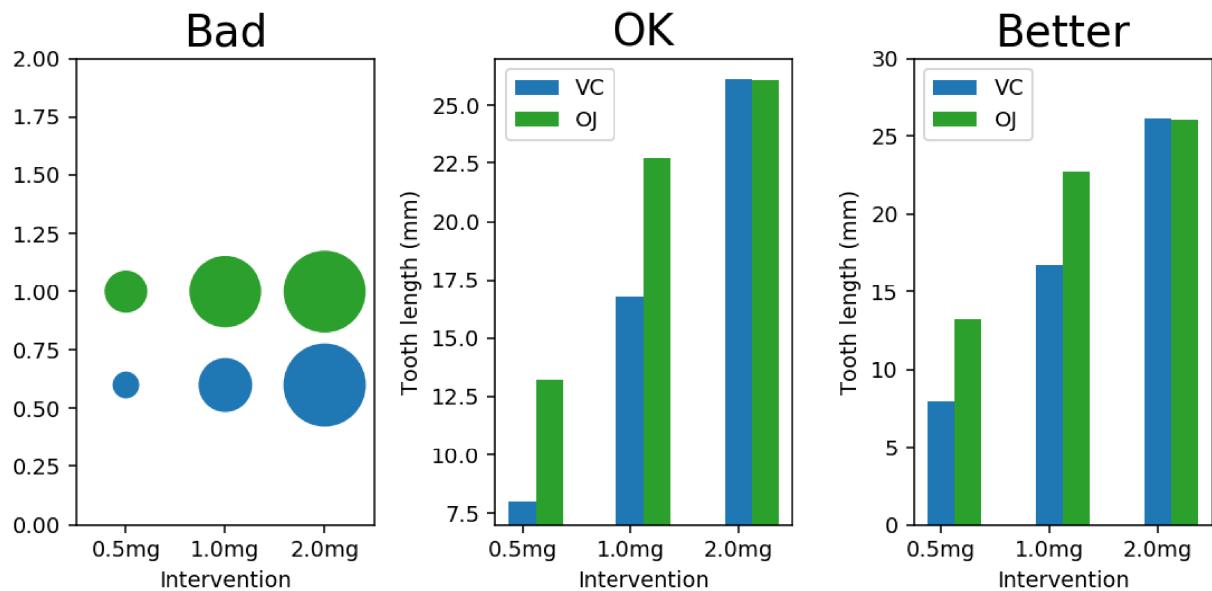


[**Left** blue and green represent means of two groups, but no uncertainty is shown. It is impossible to tell how big the differences between the blue and green results really are. **Centre** error bars give some indication of the spread of values. **Right** a Box plot is more complete summary and should be used whenever it is appropriate to do so]

Don't use deceptive scales

Use axis scales that inform the reader without distorting the message.

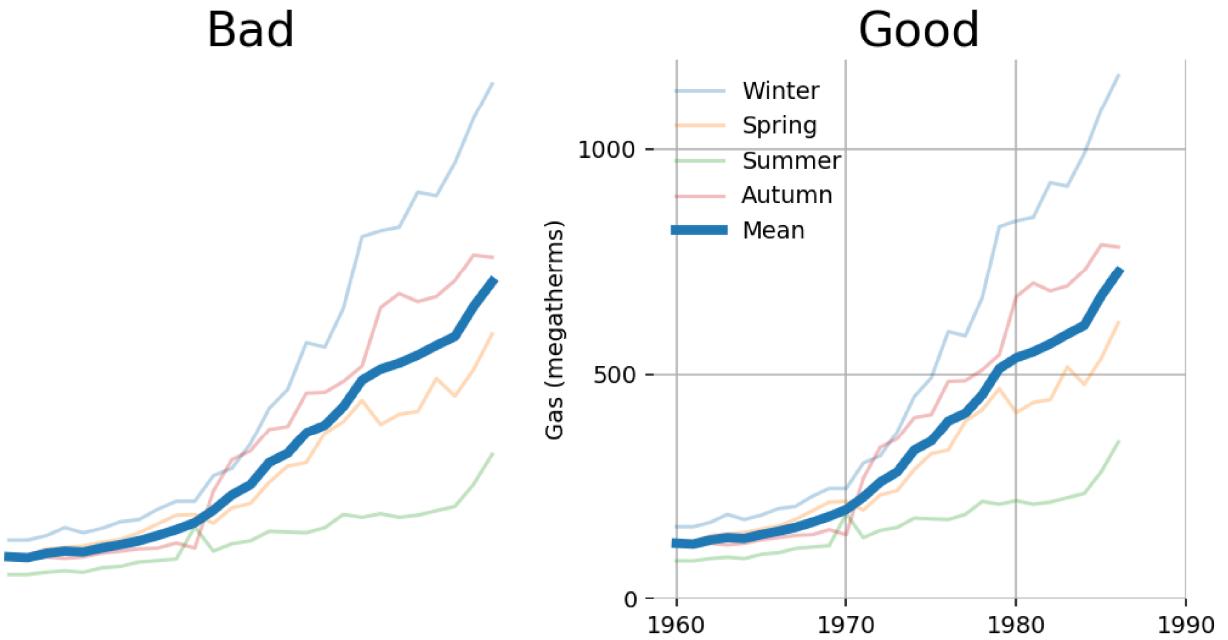
- Don't cut off data when plotting, except in extreme cases, and there you must be careful to explicitly annotate what has been cut off.
- Avoid broken axes unless absolutely necessary. Especially avoid axes that are broken somewhere other than at the origin!
- Use appropriate linear or logarithmic transformations to make the relationship clear.
- Avoid area-based (and especially volume-based) representations; you must make sure that the apparent area/volume is proportional to the value displayed and not the radius/diameter of the object and such graphs are still hard to read.



[**Left** radii of points are proportional to value; this massively inflates differences and no guide is provided to establish the units. **Centre** Visual units are now linear with data units, but the y axis scale is truncated, and the difference between the first two items is exaggerated; **Right** the y axis scale is no longer misleading]

Provide guides

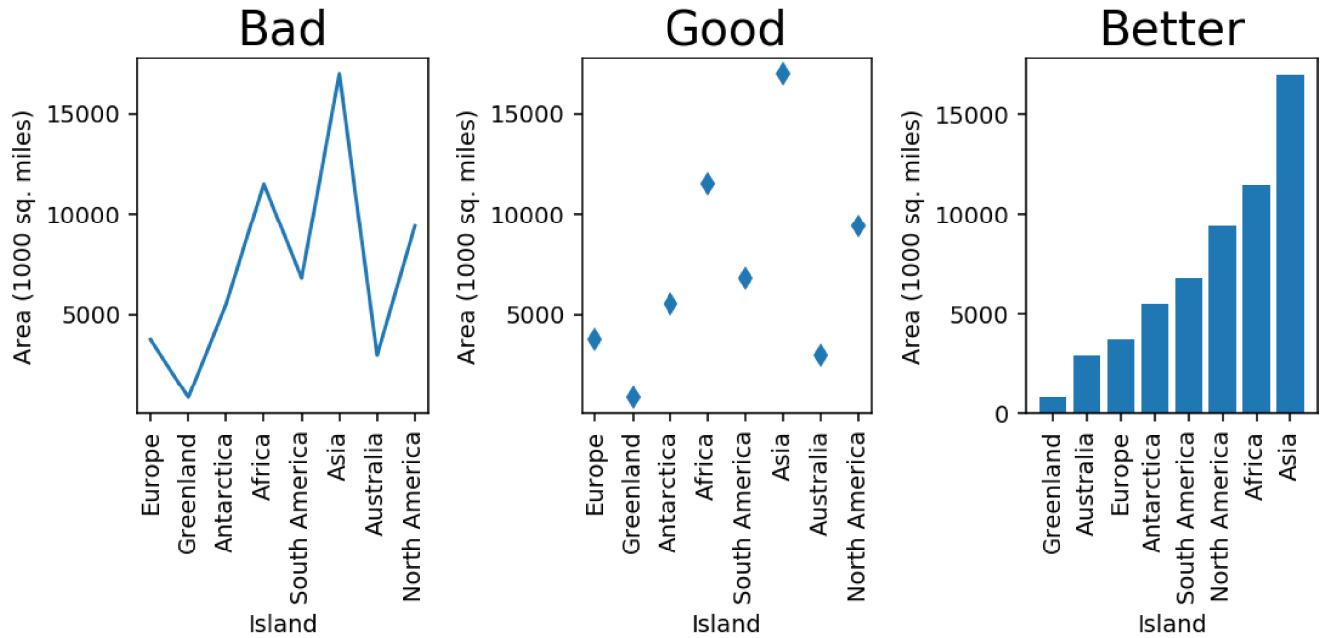
- Always provide guides for visual units (i.e. tick marks) and label axes clearly, with real-world units wherever possible.
- Always use legends to identify different layers in a plot
- Always provide a descriptive title to a plot, and a caption which explains what the figure is and what the reader should gain from looking at it.



[**Left** No guides at all (no axes, no grid, no legend) **Right** axis labels present, with real world units. Legend identifies layers. Grid makes it easy to look up specific values]

False connectivity

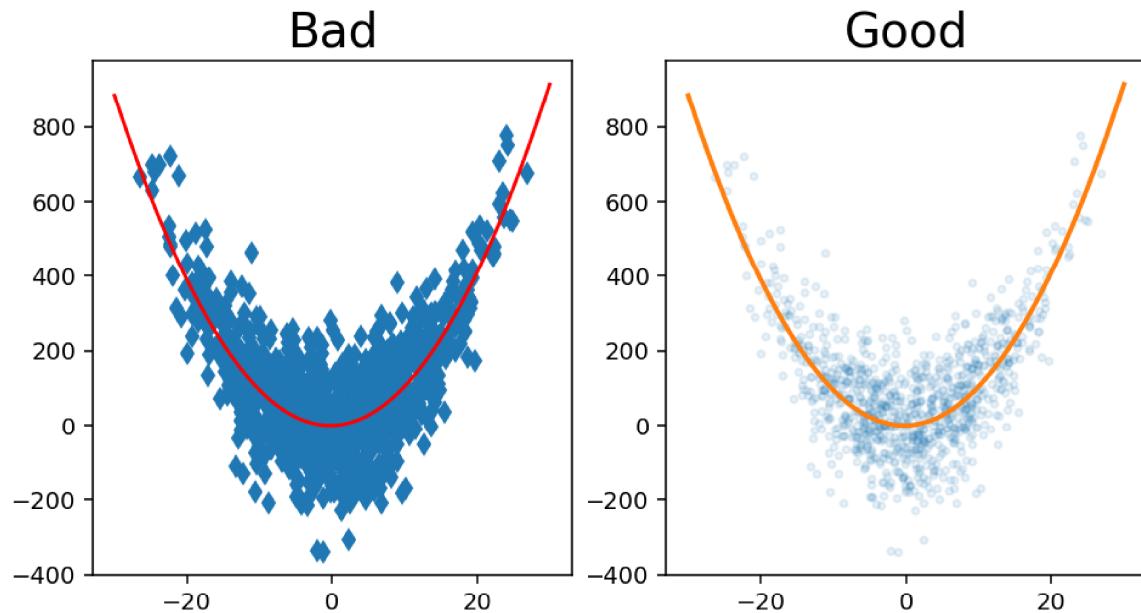
Don't connect data points with lines or curves if they don't form a continuous function. If the x axis has no defined order (categorical variable), it is even less meaningful to draw lines in between data points.



[**Left** Lines connect landmasses, but what does it mean to be in between "Greenland" and "Antartica"? The data do not form a continuous function. **Centre** No false connectivity between distinct items. **Right** A bar chart is easier to read, and by sorting the elements before plotting the graph becomes more organised and easier to read.]

Don't clutter

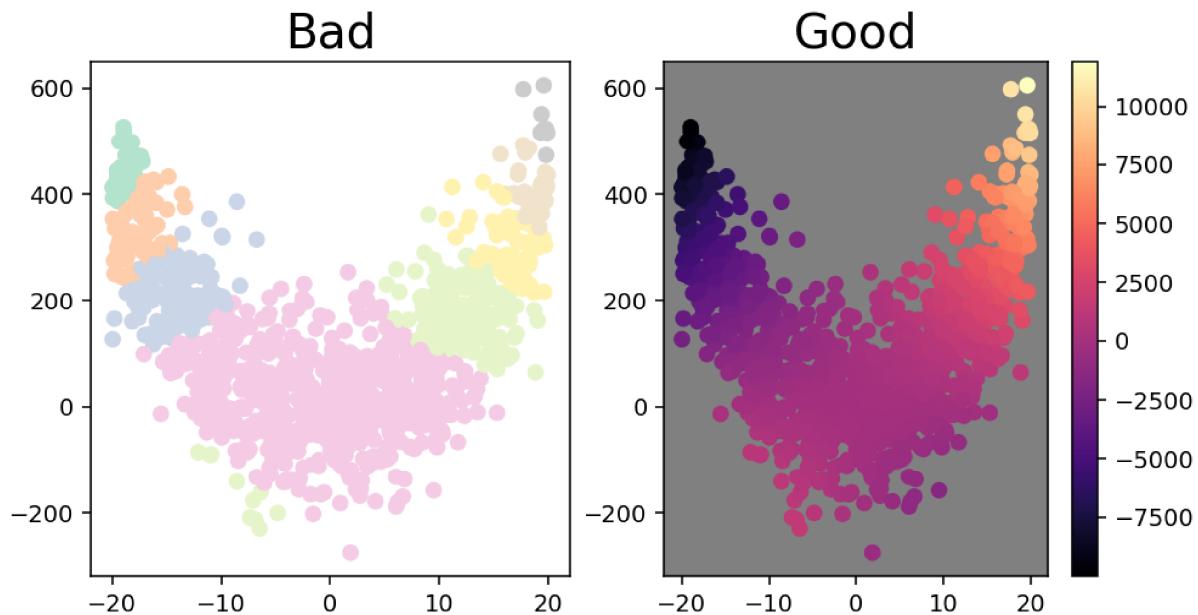
- If you have huge numbers of markers or other geoms present, consider using a lower opacity to make the plot easier to read.
- If this isn't sufficient, consider using a 2D histogram or kernel density estimate to show the **density** of points accurately (at the expense of losing the precise location of individual data points).
- Emphasise important curves or datapoints with thicker lines, larger points, but don't go overboard.



[**Left** Extremely dense scatterplot with large markers. The points become indistinct and the whole area turns into a blue mass. **Right** Appropriate use of alpha blending (opacity) can reveal the density of points more accurately and declutter the plot.]

Use appropriate colour scales

- If your data is unsigned (positive only), use a perceptually linear colour map with monotonically increasing brightness.
- If your data is signed (and the sign matters), use a colour map which diverges around 0.
- Scale data to colorscales appropriately, and always provide a colour bar.

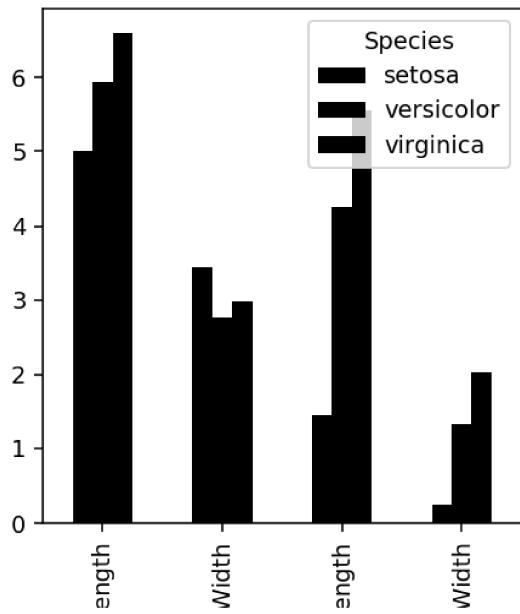


[**Left** The colour scale is wholly inappropriate, with discrete steps in colour introducing false contours, and a very low contrast result which is not perceptually linear in brightness. There is no guide to indicate how to interpret the colour values. **Right** Perceptually linear, monotonically increasing brightness colour map. A colour bar is provided as a guide to the true values being displayed.]

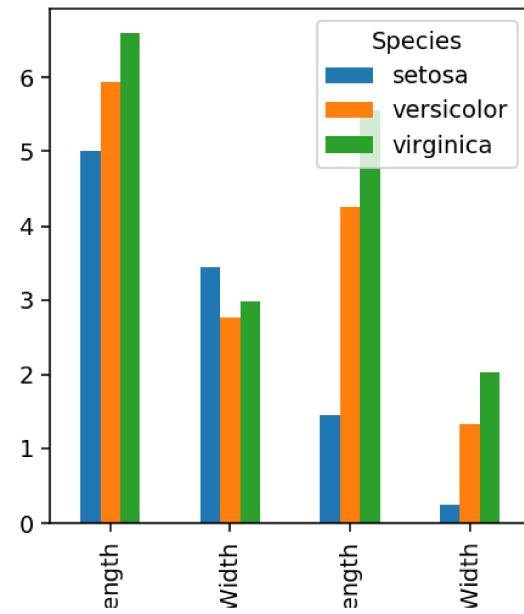
Use appropriate styling

- Make sure your geoms are distinguishable in black and white if they might appear in black and white.
- Use dashed styles, but sparingly; only 2 or 3 can easily be distinguished on a plot.
- Mark actual measurements with markers if the reader might need to know where the measurements are.

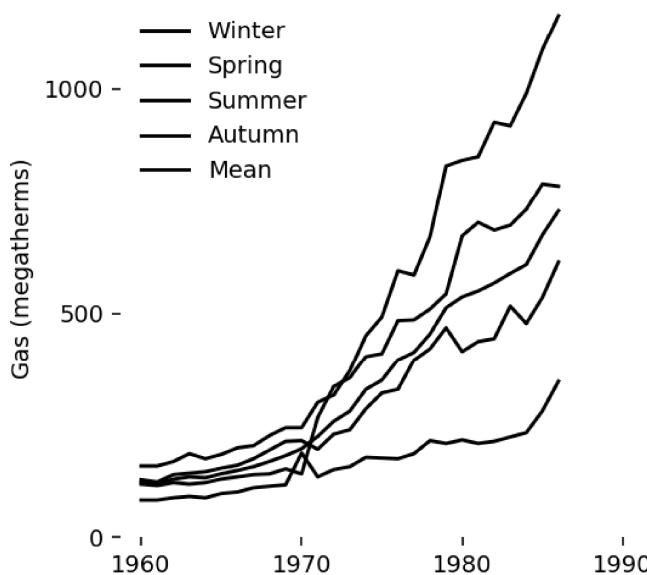
Bad



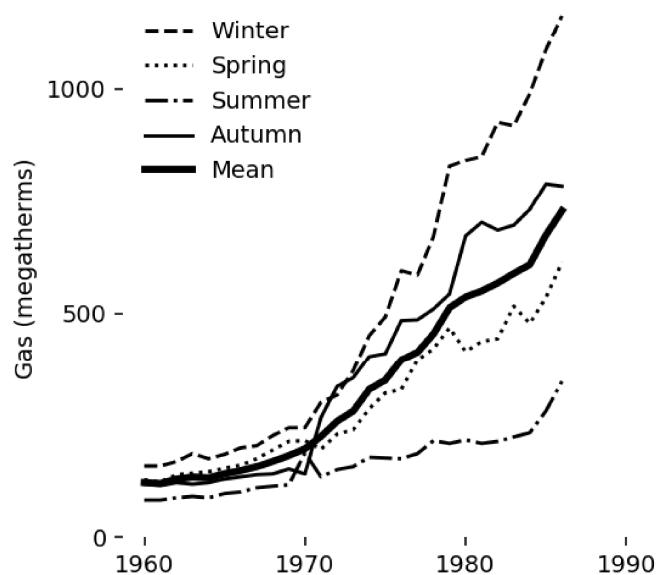
Good



Bad



Good



[Left All geoms are plotted in the same colour, and same style. It is impossible to tell the different layers of the plot apart. **Right** Different colours (above) and different line styles (below) are applied to geoms to visually distinguish different layers of the plot]