

Best Practices for Data Visualisation

RSS International Conference 2023

Harrogate, UK

Andreas Krause, Nicola Rennie, and Brian Tarran

Welcome

In this session we will cover... what the guide is, and how it came about, what's in the guide, and how you can contribute to the guide.

We also want to hear from you about what sort of content you'd like to see added to the guide.

Some background

A survey in 2021 asked RSS members their views on *Significance* magazine.

Respondents were asked, “What aspects of content could be improved?”

- “Better, more consistent charts... I’d like to see a house style like *The Economist*”
- “The plots sometimes look amateurish...”
- “The figures are often difficult to read...”

Help wanted

We put out a call:

“RSS publications seek data visualisation expert to develop best practice guidance”

Help wanted

The guide would:

- Help contributors develop data visualisations that are high quality, readable, effective at conveying information, and fulfil their intended purpose.
- Summarise and link to authoritative advice on chart styles and formats for different types of data.
- Show how to override software defaults in common data visualisation software and packages.

Help wanted

It would also provide basic information on figure sizes, fonts, colours, resolution, etc., used in RSS publications.

Help wanted

Andreas and Nicola answered the call. We started work in February this year, and six months later...

The guide was published



- Introduction
- How to use this guide
- The guide
- Why we visualise data
- Principles and elements of visualisations
- Choosing a visualisation type
- Styling for accessibility
- Styling for RSS publications
- References and resources
- About the authors
- Join us at RSS Conference
- Terms and conditions

Best Practices for Data Visualisation

Insights, advice, and examples (with code) to make data outputs more readable, accessible, and impactful



Coming to the RSS Conference this September? Find out more about [our conference session](#) and how to become a guide contributor!

Statistics is “the science of collecting, analyzing, presenting, and interpreting data” ([Williams, Anderson, and Sweeney 2023](#)). Presentation of data is a key means to support and guide interpretation and subsequent decision making. Techniques exist for effective display. This is what this guide is all about.

Good data visualisation requires appreciation and careful consideration of the technical aspects of data presentation. But it also involves a creative element. Authorial choices are made about the “story” we want to tell, and design decisions are driven by the need to convey that story most effectively to our audience. Software systems use default settings for most graphical elements. However, each visualisation has its own story to tell, and so we must actively consider and choose settings for the visualisation under construction.

This guide covers both aspects of data visualisation: the art and the science. It is written primarily for contributors to [Royal Statistical Society publications](#) – chiefly, [Significance magazine](#), the [Journal of the Royal Statistical Society Series A](#), and [Real World Data Science](#) – but we trust you will find the information and advice within to be of broad relevance and use to any data visualisation task.

How the guide is structured

Why we visualise data

- Motivation
- A brief history of data visualisation

Principles and elements of visualisations

- Elements of charts, including:
 - Layout
 - Aspect ratio
 - Lines, points and symbols
- Elements of tables, including:
 - Layout
 - Digits
 - Alignment

Choosing a visualisation type

- Goals
- Target audience
- Data types
- Data relationships

Styling charts for accessibility

- Colours
- Annotations
- Fonts
- Alt Text

Styling charts for RSS publications

- Styling with different tools, including:
 - R
 - Python
 - Julia
- Publication specifications, including:
 - Pages sizes and column widths
 - Font types and sizes
 - Colours

References and resources

- Primary sources
- Further reading
- Additional resources

Why visualise data?

Motivation

Visualisations are found everywhere.

They are the key medium for transporting a message.

Some are better than others.

There are largely no standards.

Motivation: Anchor the message

- Grab attention
- Improve access to information (over text)
- Increase precision (over text)
- Bolster credibility: see yourself
- Summarise content

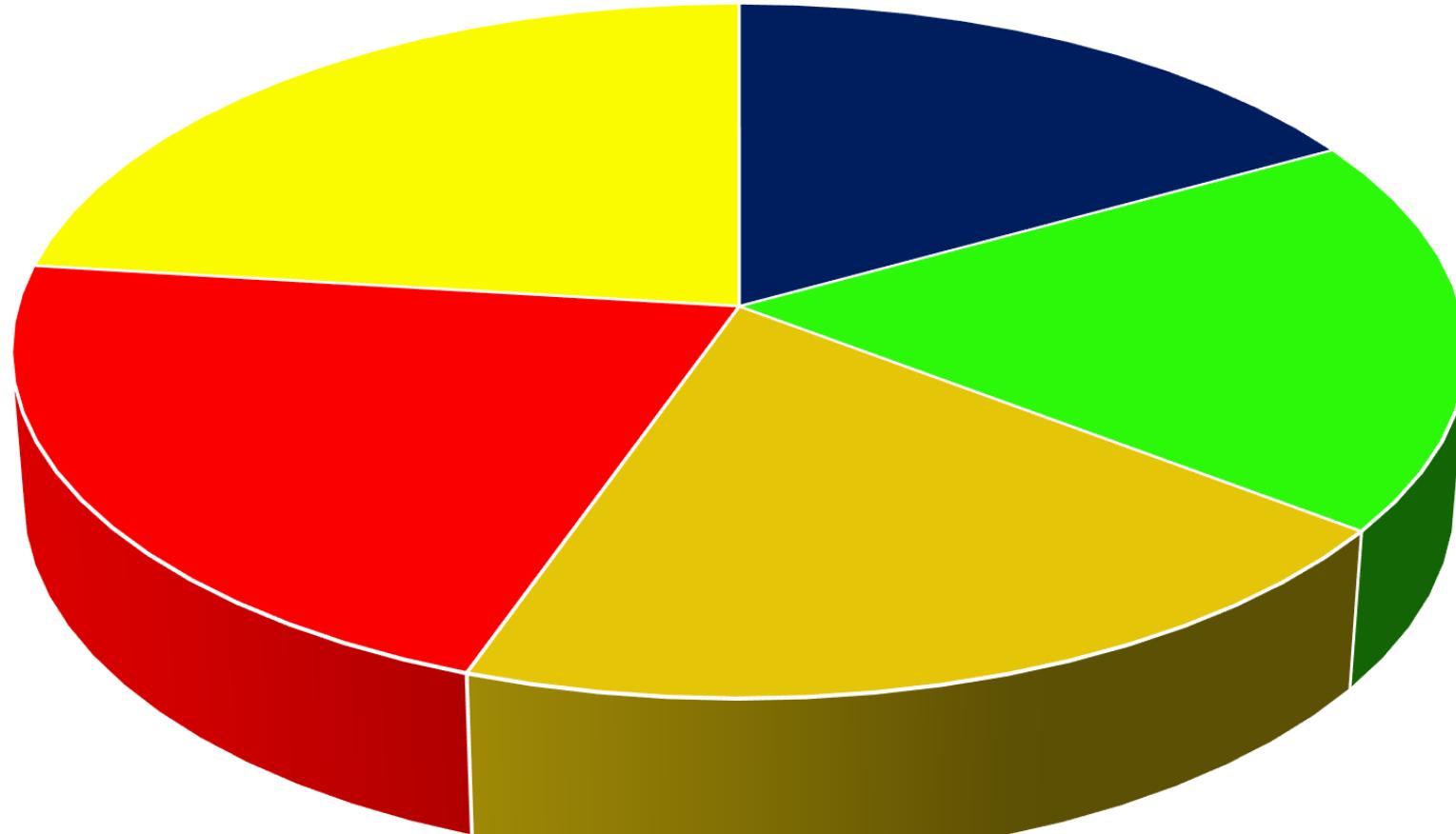
The Art of Visualisation (1)

Gauge the sizes, determine the largest piece



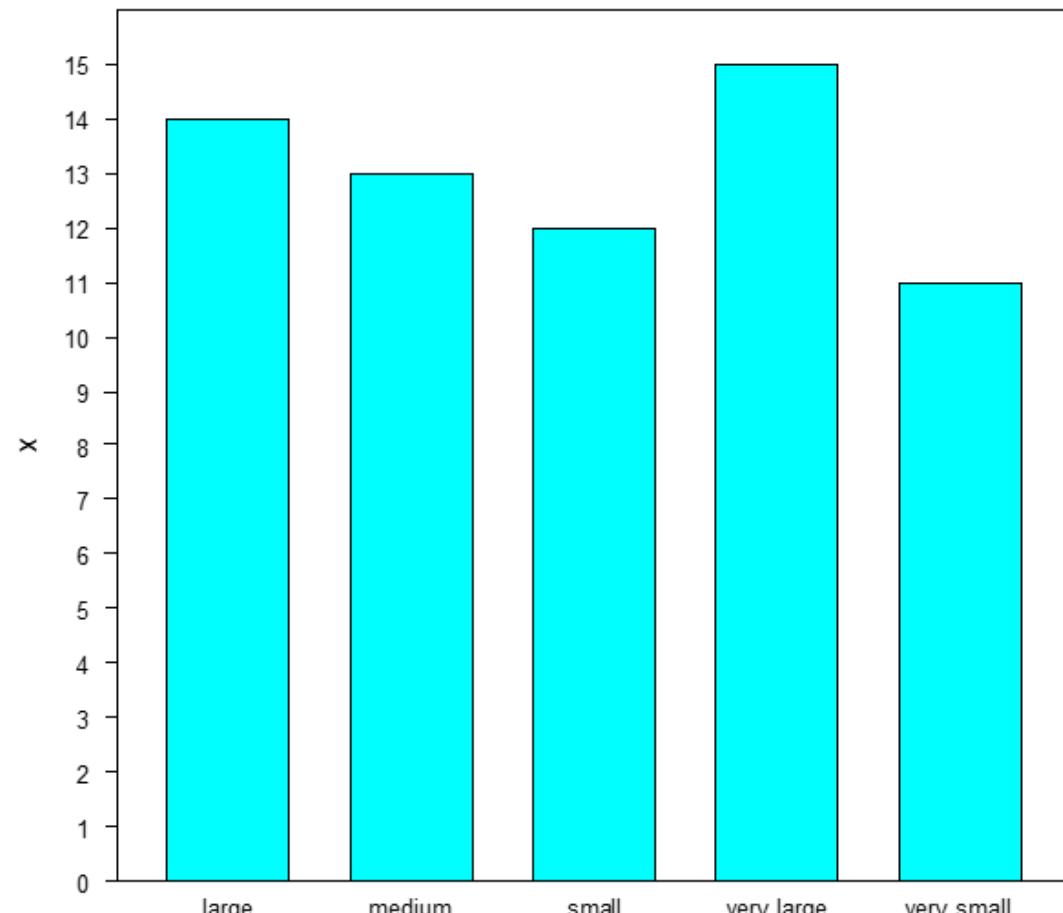
The Art of Visualisation (2)

The number of pixels per pie depends on its position



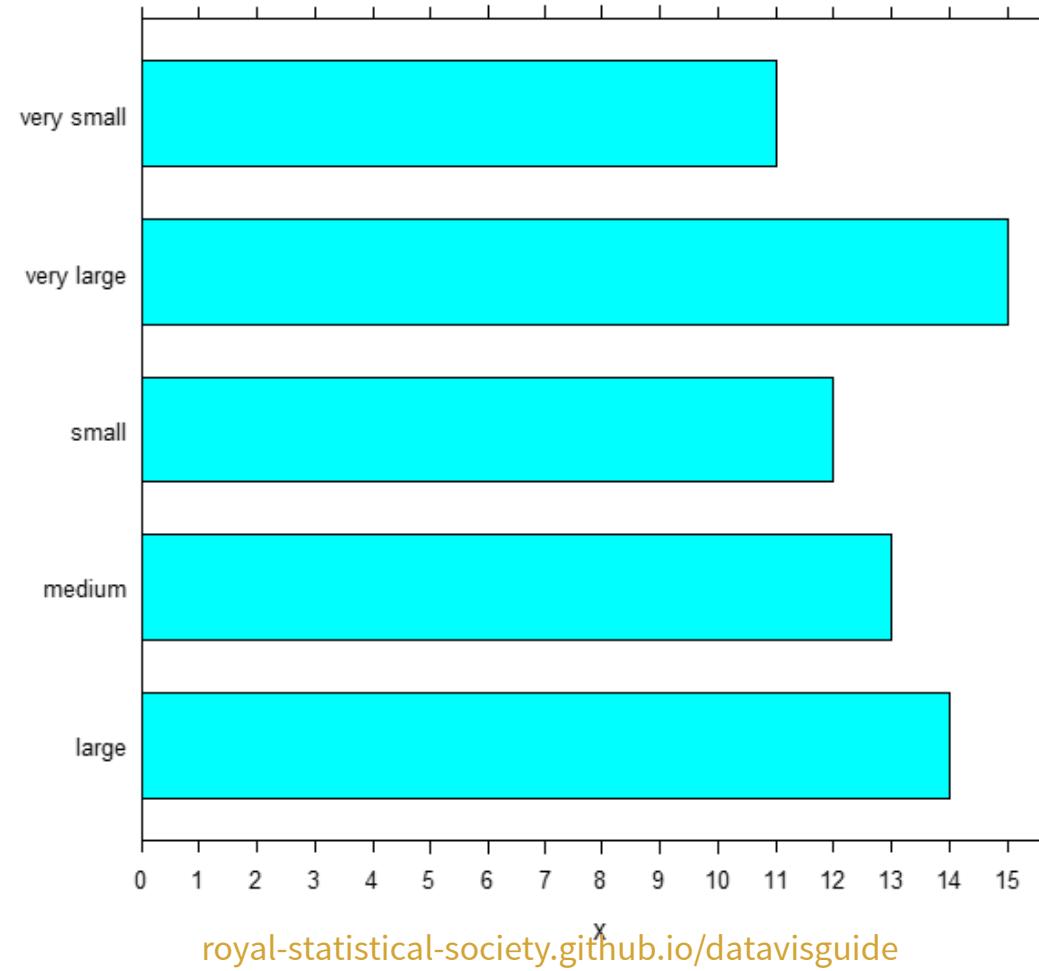
The Art of Visualisation (3)

Sorting the bars by height is easy



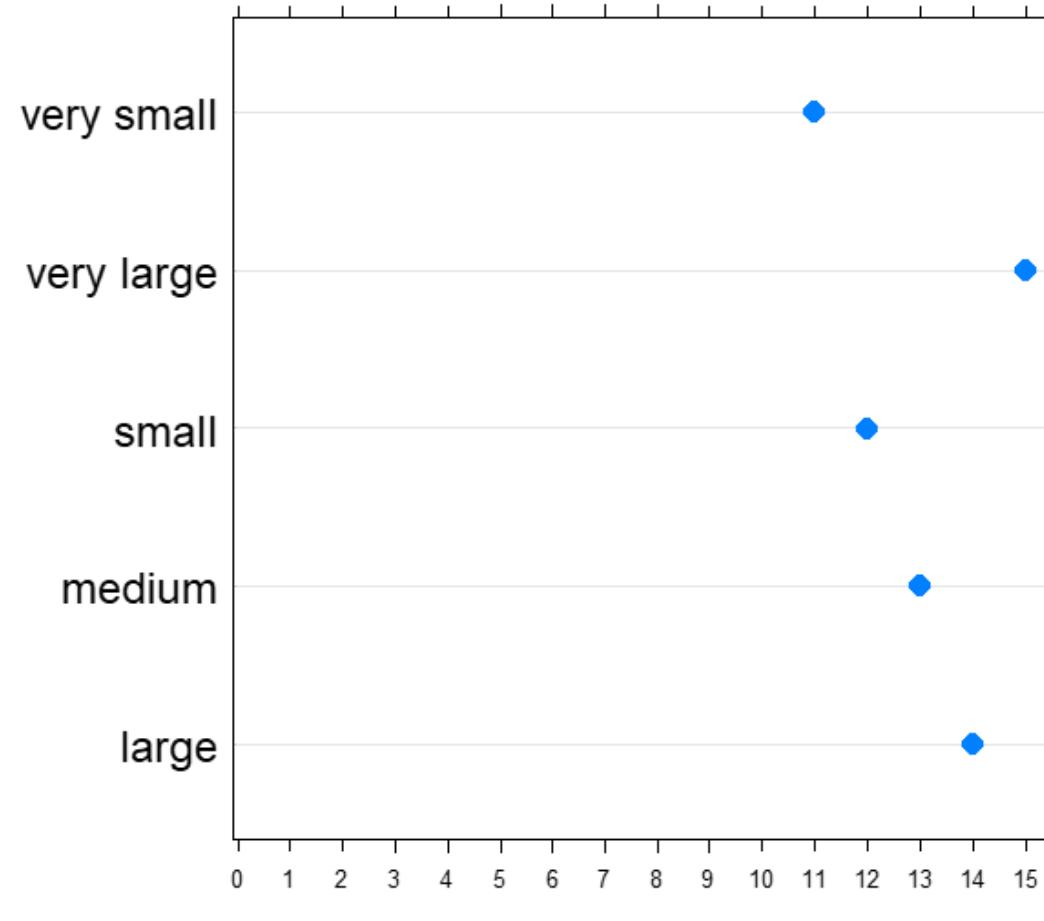
The Art of Visualisation (4)

... and arguably even easier with horizontal layout



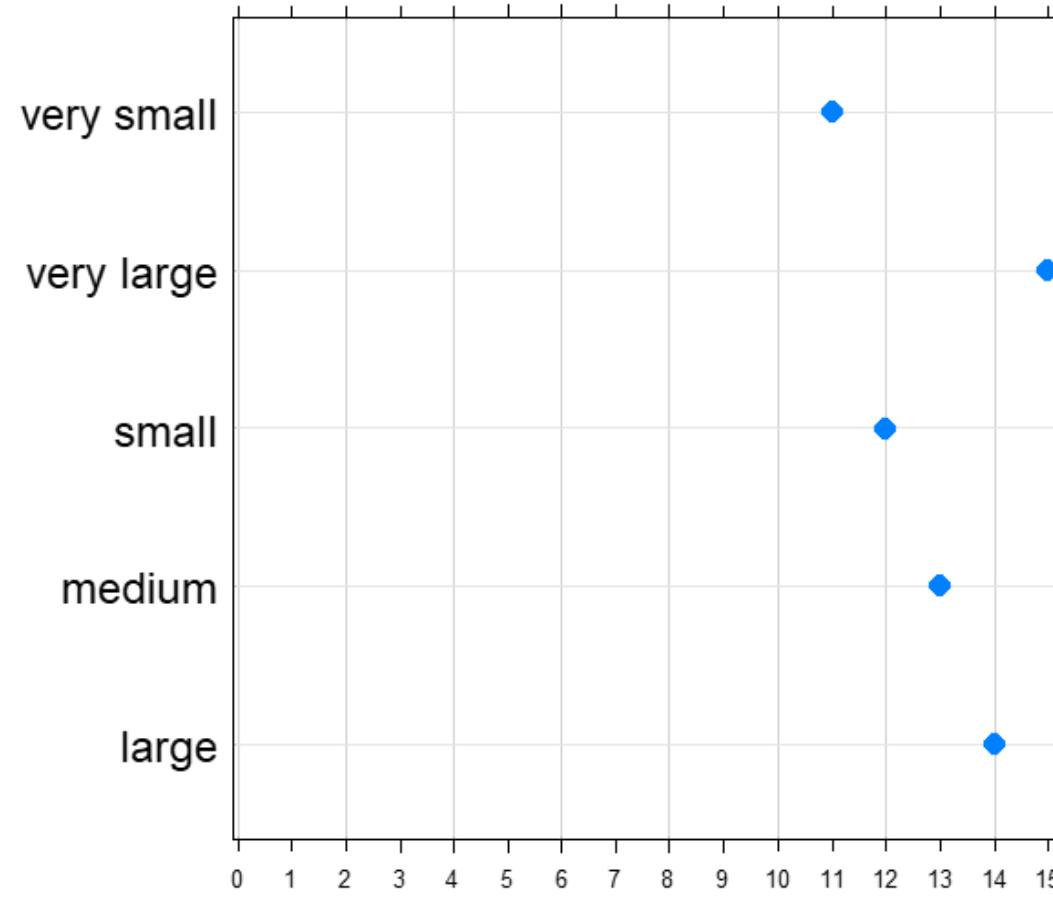
The Art of Visualisation (5)

A single pixel carries the same information as a large bar



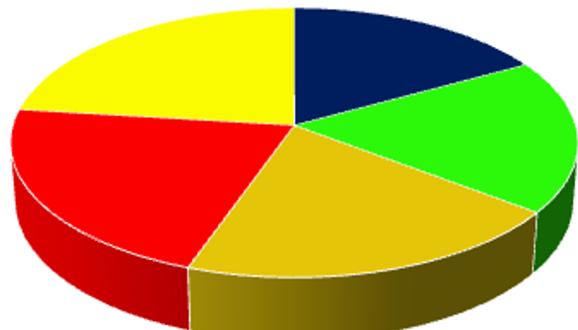
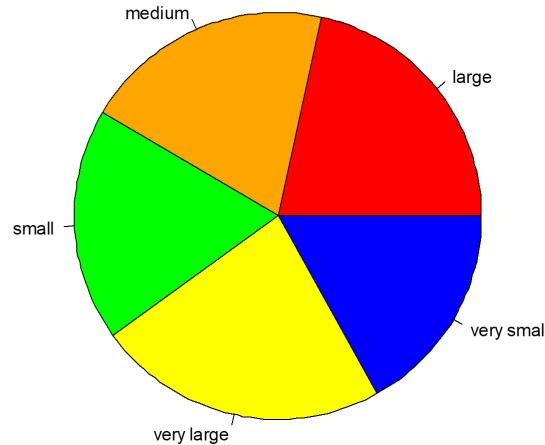
The Art of Visualisation (6)

Faint gridlines help reading off values precisely

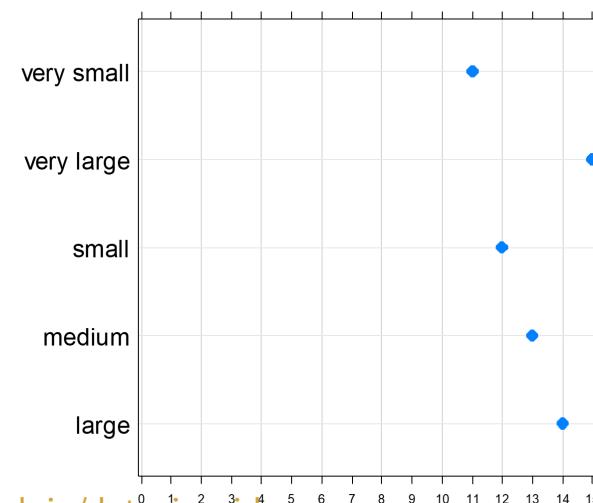


The Art of Visualisation (7)

There are options in designing a visualisation!



Size	Value
Very small	11
Small	12
Medium	13
Large	14
Very large	15



Principles and elements of visualisation

Purpose

Data visualisations must serve a purpose.

Frequent aim: comparison.

Ask yourself:

- What is the purpose?
- Does the visualisation support the purpose?
- Quickly, Accurately, and Intuitively?

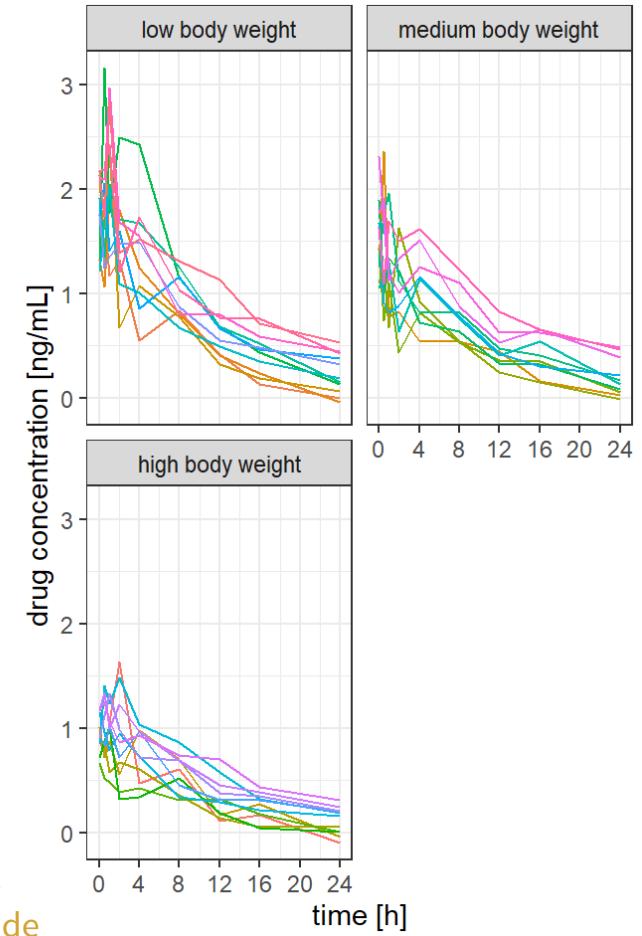
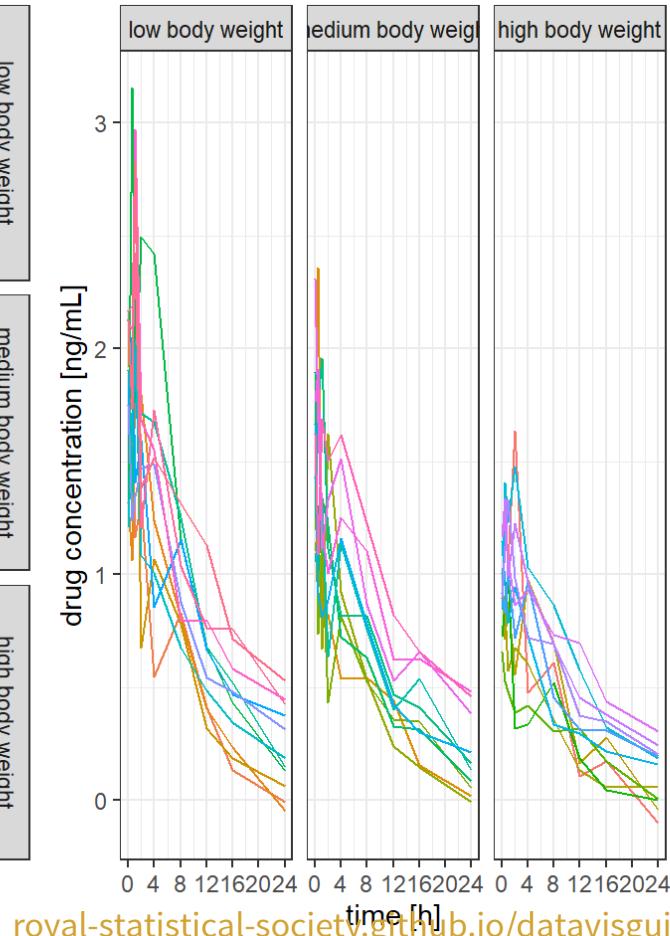
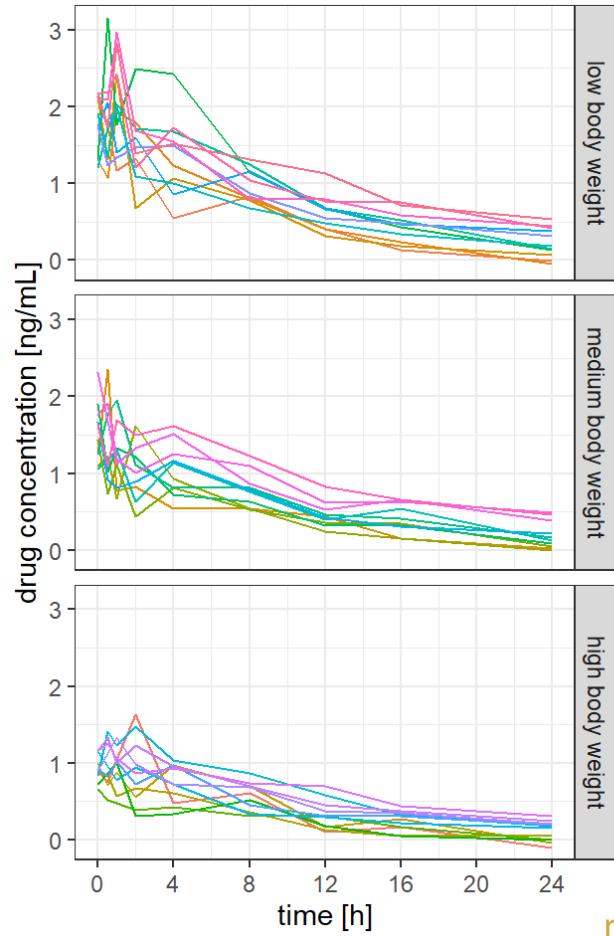
Elements of Charts

- Layout
- Aspect ratio
- Lines
- Points
- Colours
- Axes
- Symbols
- Legends
- Orientation
- Auxiliary elements
- Dimensionality

Note the defaults: the boxplot function in R has 27 of them.

Layout

- Which axes are to be compared?



Lines

- Lines introduce an order
- No order, no lines

Line types: map style elements to order

- Line thickness
- Dash density for dashed lines
- Brightness (black to white)

Points

If data points overlay (which they generally do)

- Open circles still allow discerning data points
- Smaller dots can be considered

If data points overlay perfectly (example: integer data)

- Consider jittering

Colours

- Colours should serve a purpose: discerning groups of data

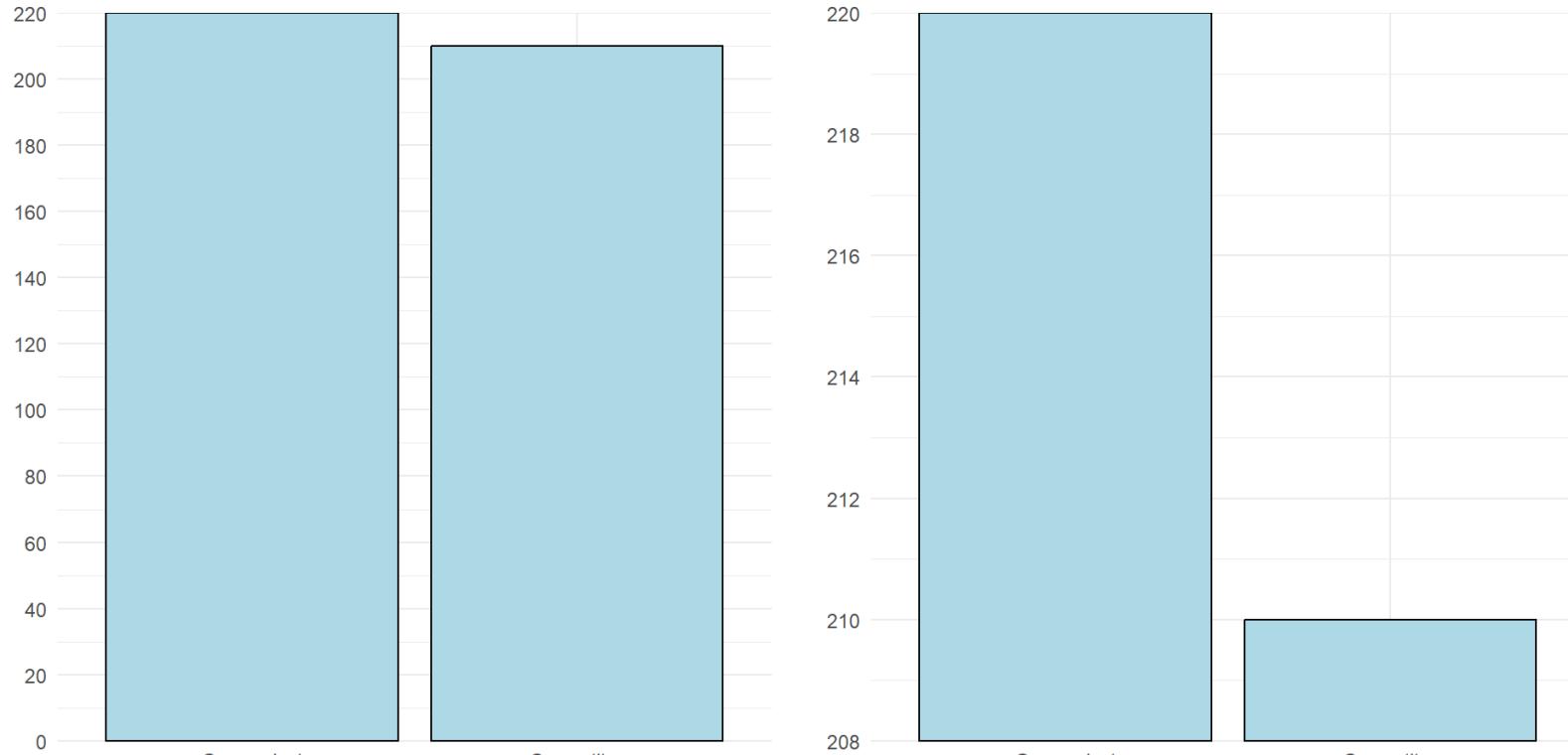
Shades of gray

- Have a natural visual hierarchy
- Show varying quantities better than color
- Provide an easily comprehended order to the data measures
- This is the key

Edward Tufte (2001), p. 154

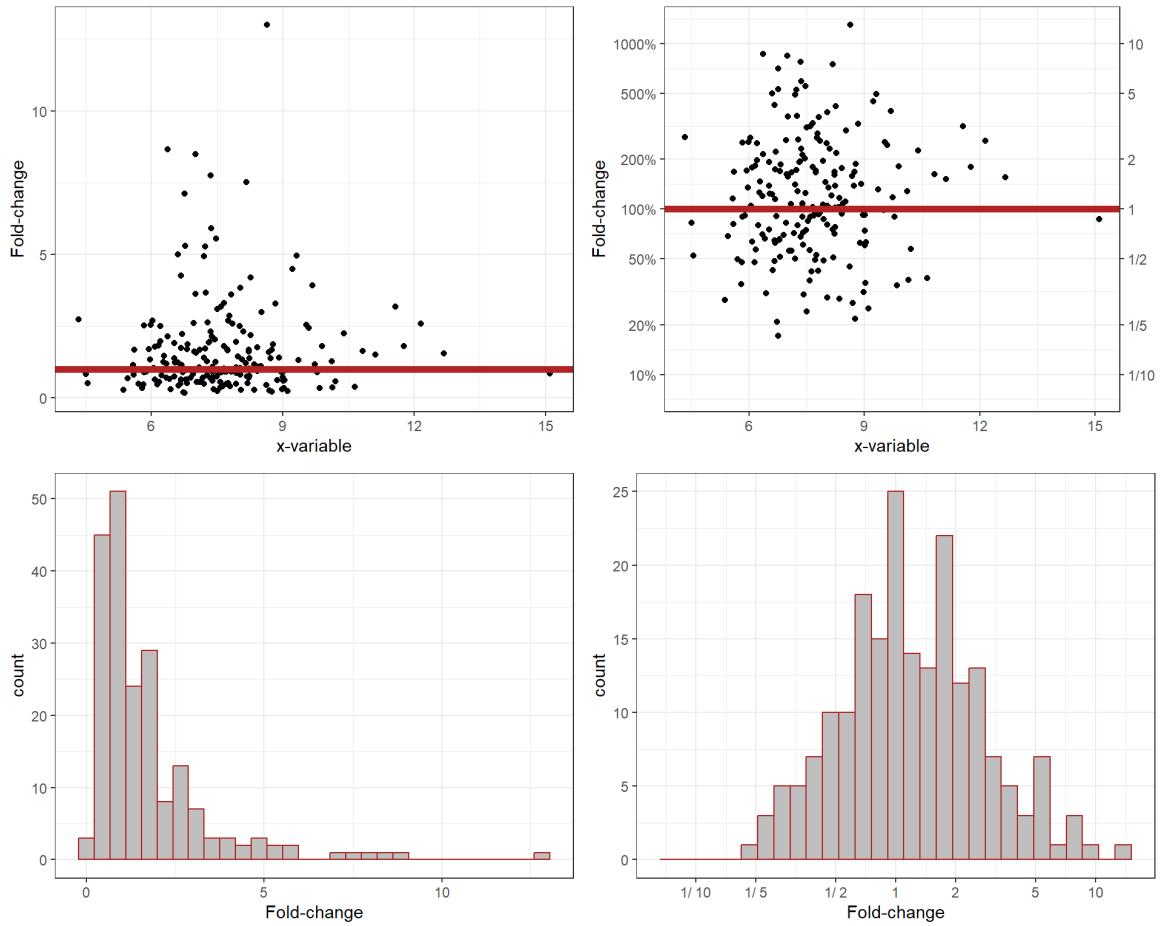
Axes

- Should generally start at 0
- Should not have negative numbers with none in the data



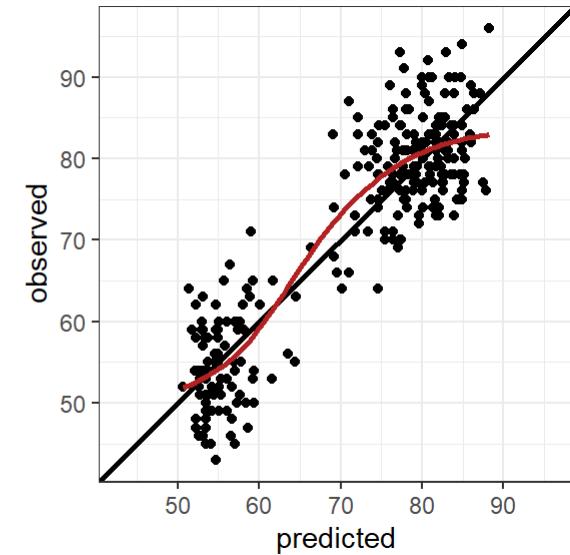
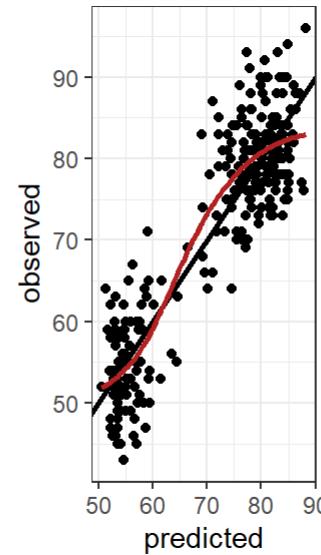
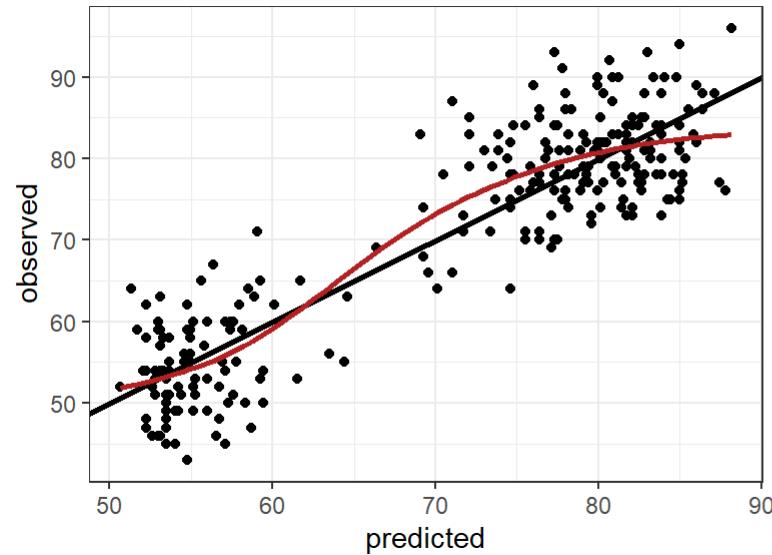
Axes (2): Relative changes

- Log-axes for symmetry
- Tickmarks 1/4 and 4
- Line at “no change”
- Faint gray grid



Aspect ratio

- Number of pixels allocated to 1 measurement unit in y vs x
- Comparing x and y (example: predicted vs observed data)
- Identical axis limits, aspect ratio = 1 \rightarrow square figure



Symbols

Should be intuitive:

- Good: +, bad:-, neutral: 0
- Consider using letters (example: “L”ow, “M”edium, “H”igh)

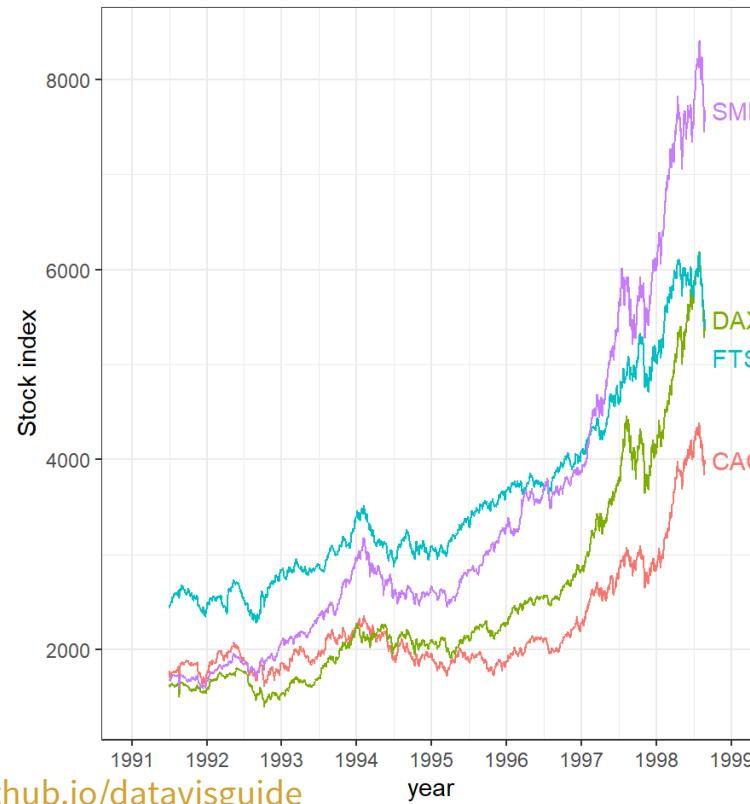
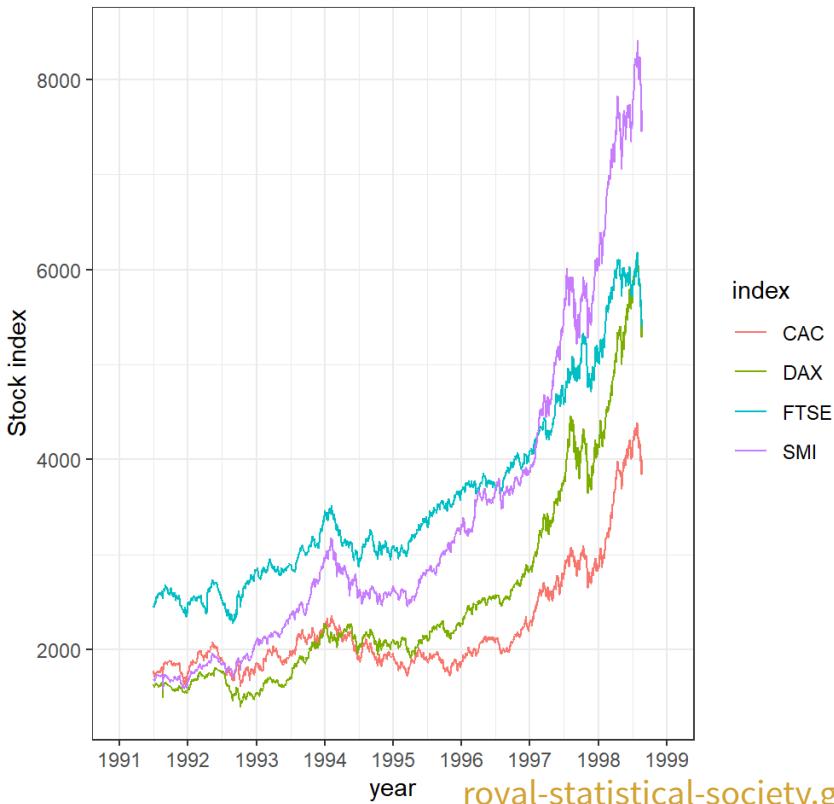
Ideal case: Single look at the legend to memorize the mapping

Not intuitive: triangles, circles, squares -> repeated looks

(unless the order - number of vertices - carries a meaning)

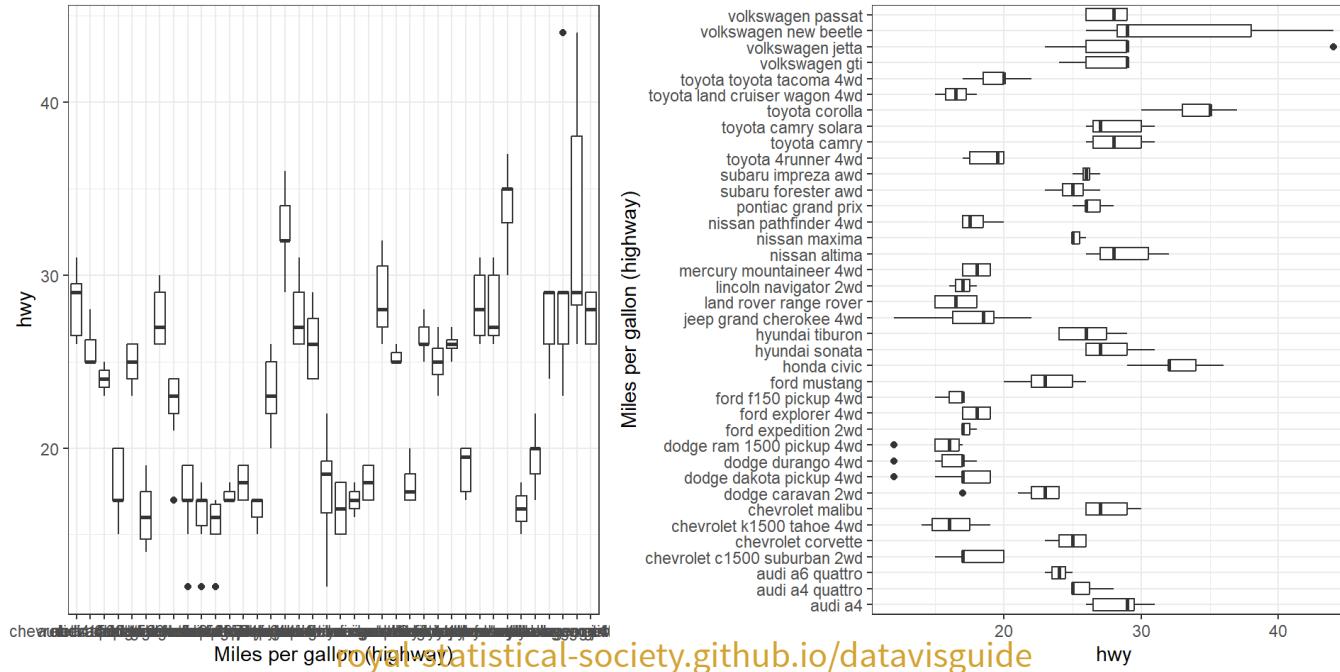
Legends

- Should not use up valuable space for data
- May be integrated into the figure



Orientation

- Order: y-axis from low to high, x-axis from left to right
- Time flows from left to right (past to future)
- Longer labels best at the y-axis, horizontally

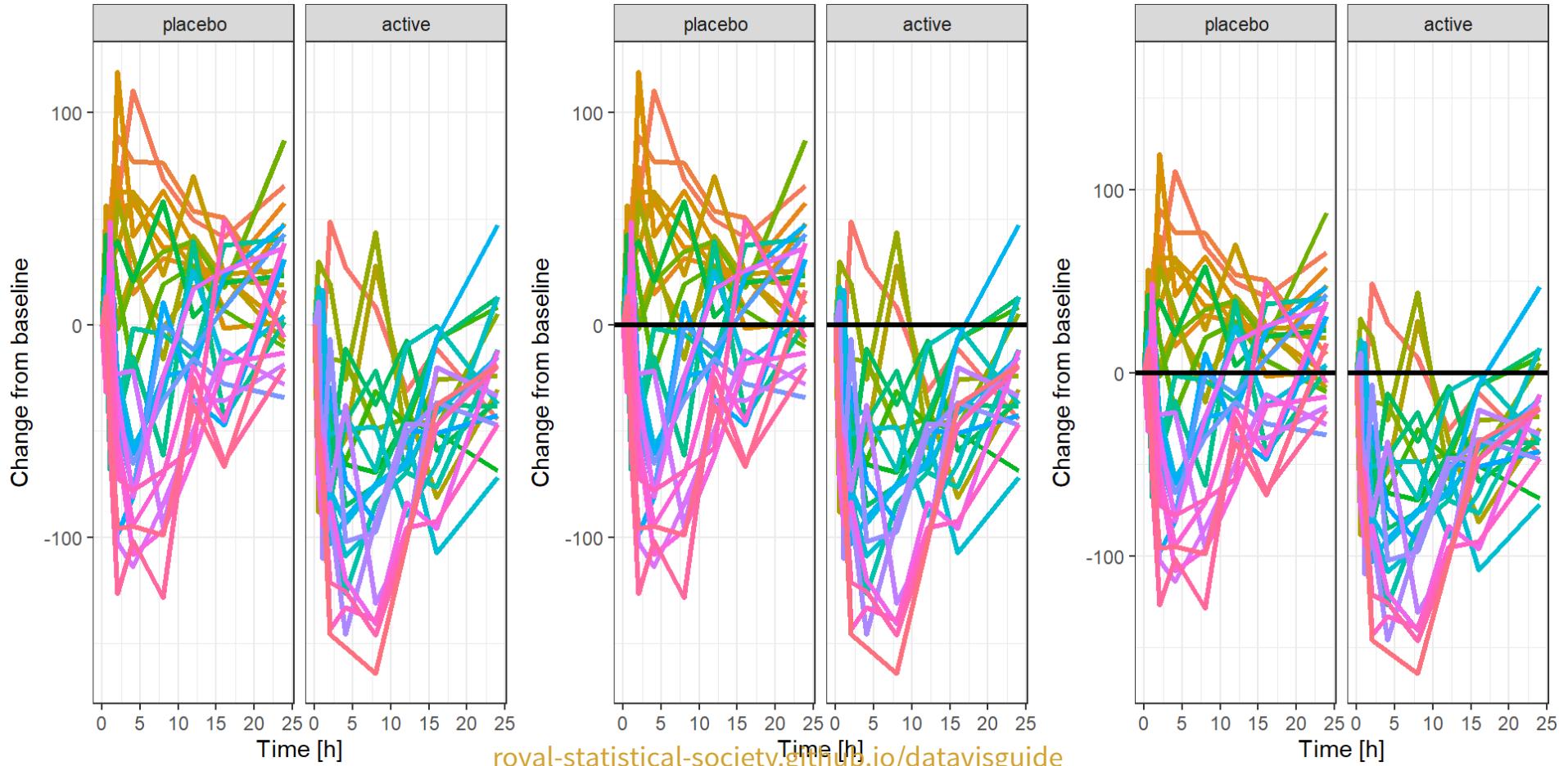


Auxiliary elements

- Tufte: avoid “chart junk”, elements without information
- My example: the gray background in ggplot2 figures
- No relevant contribution: remove (repeated identical axes)
- Helpful lines (examples: $y=0$, faint gray lines)
- Smoother to support trend identification
- Confidence band only if relevant (default with R loess)

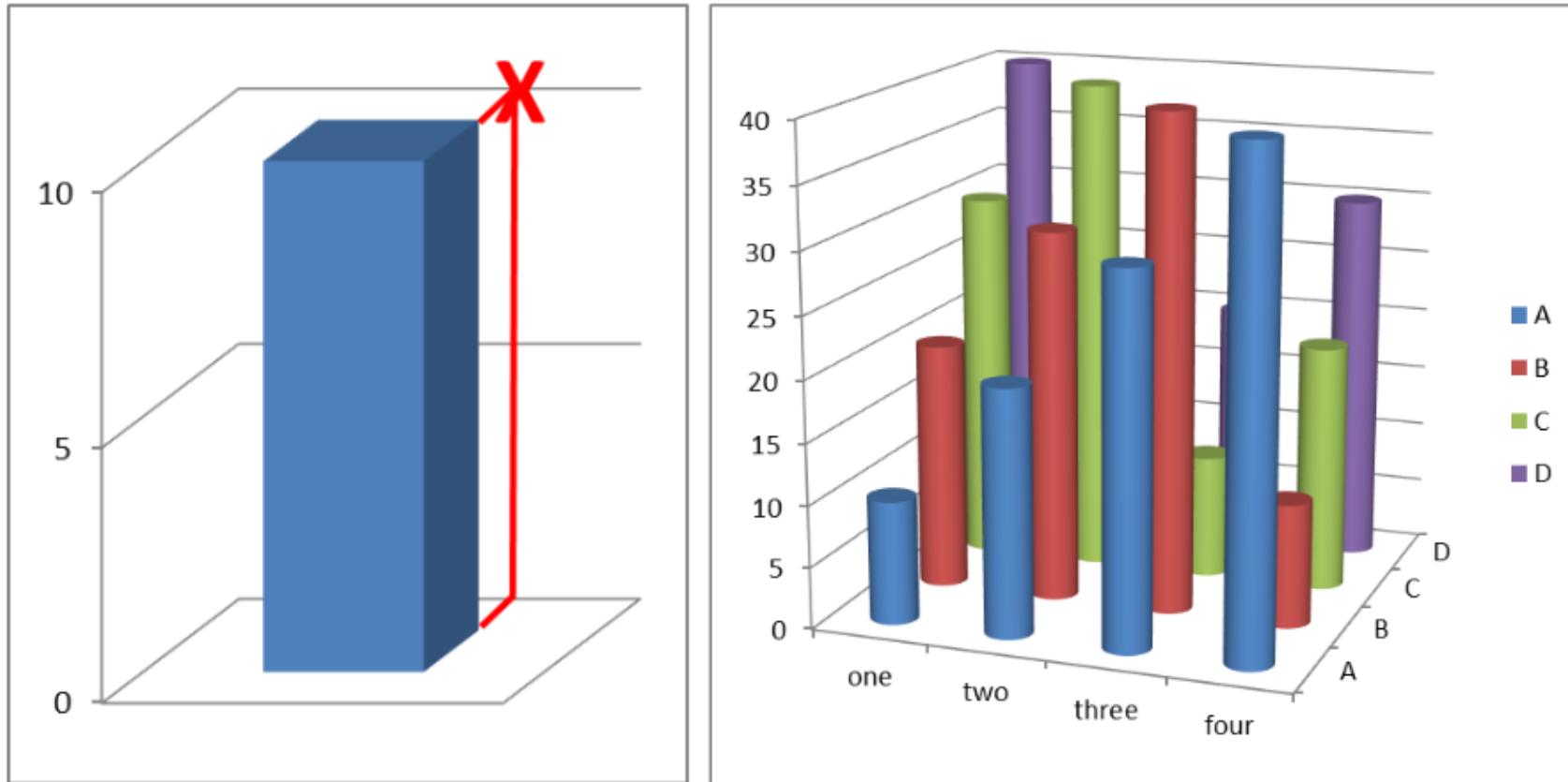
Auxiliary elements: change

- Symmetric y-axis, line at $y=0$ -> focus on data



Dimensionality

- Accurate interpretation is not straightforward. Don't.



Elements of Tables

- Layout
- Digits
- Alignment
- Multiple numbers per cell
- Orientation
- Fonts
- Colours

Tables: Layout

- Choose rows and columns consciously
- Numbers are easier to compare vertically than horizontally
- Generally: variables in columns, observations in rows

Name	Age	Weight
Alex	55	123.45
Sandy	33	77.07

Name	Alex	Sandy
Age	55	33
Weight	123.45	77.07

Tables: Alignment

Generally helpful:

- Decimal points aligned vertically (monospace fonts?)
- Right adjustment (larger numbers “stick out”)
- Difficult if numbers are very different (e.g., 953 and 0.07)

Name	Age	Weight
Alex	55	123.45
Sandy	33	77.07

Name	Alex	Sandy
Age	55.00	33.00
Weight	123.45	77.07

Tables: Digits

- Unnecessary precision is to be avoided
- R: Digits argument

```
1 print(1:7 + 1/(1:7), digits=2)
```

```
[1] 2.0 2.5 3.3 4.2 5.2 6.2 7.1
```

```
1 print(1:7 + 1/(1:7), digits=3)
```

```
[1] 2.00 2.50 3.33 4.25 5.20 6.17 7.14
```

```
1 format(10+1:7 + 1/(1:7), digits=4)
```

```
[1] "12.00" "12.50" "13.33" "14.25" "15.20" "16.17" "17.14"
```

```
1 format(c(1234, 1/1234), digits=3)
```

```
[1] "1.23e+03" "8.10e-04"
```

Tables: Multiple numbers per cell

- Hard to read, better separate columns

Variable	Mean (%CV)
Age	55 (9)
Weight	88 (25)

Variable	Mean	%CV
Age	55	9
Weight	88	25

Tables: Orientation

- Single landscape pages are a pain
- Consider splitting the table into two

Tables: Fonts and colours

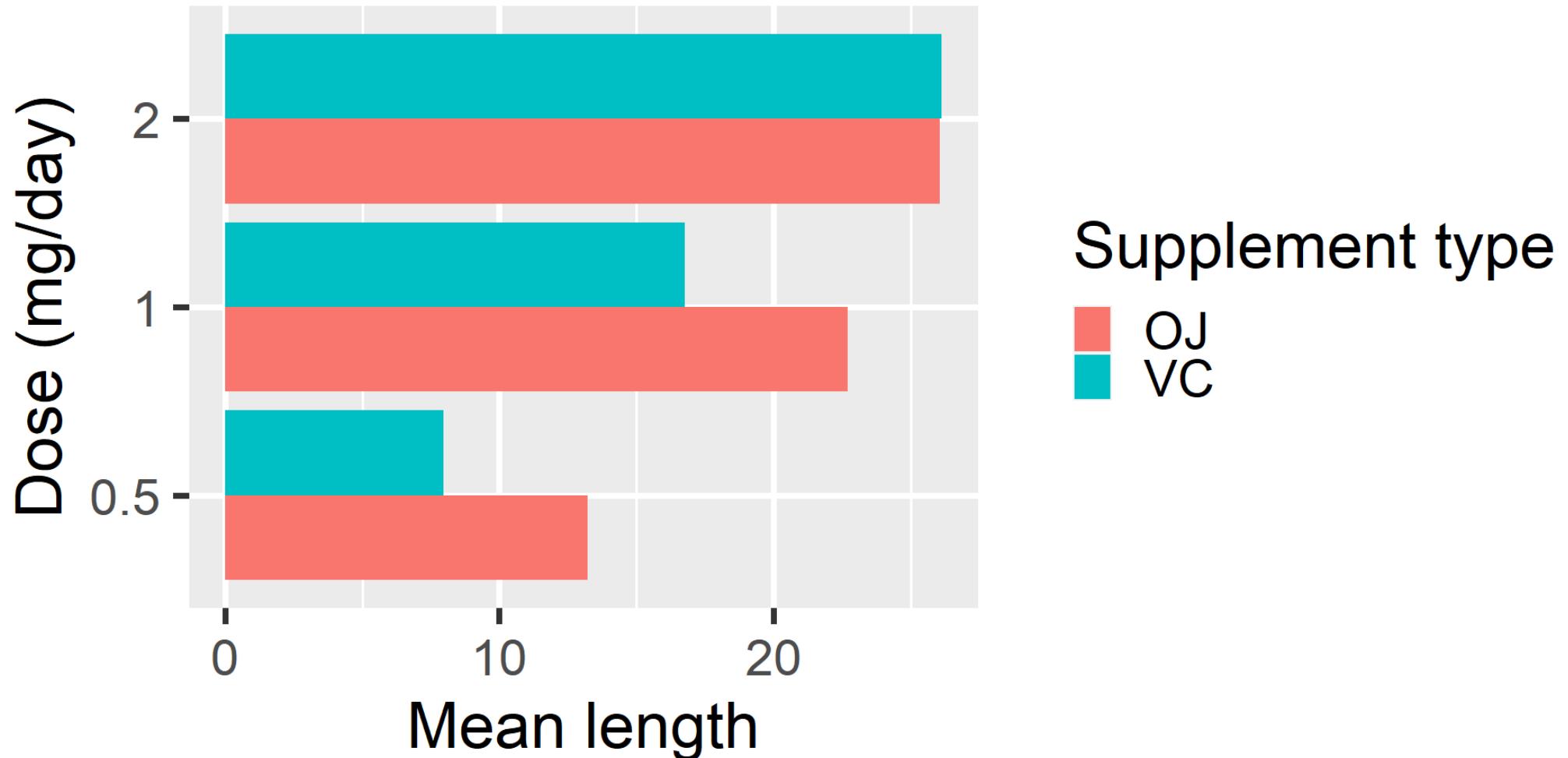
- Some fonts are easier to read than others
- These are usually the standard fonts
- Use of different fonts and colors only for a purpose
- Example: extreme values in boldface or red

Recap: Creating Charts and Tables

- Actively designing charts is recommended
- What is the question?
- Does the visual enable answering the question efficiently?
- Good visualisations increase $P(\text{paper gets accepted})$
- Visual abstracts come into fashion with journals

Styling charts

What's wrong with this chart?



Colours

Why use colours in data visualisation?

- Colours can highlight or emphasise parts of your data.
- Not always the most effective for e.g. communicating differences between variables.

Colours: types of palette

Sequential



Diverging

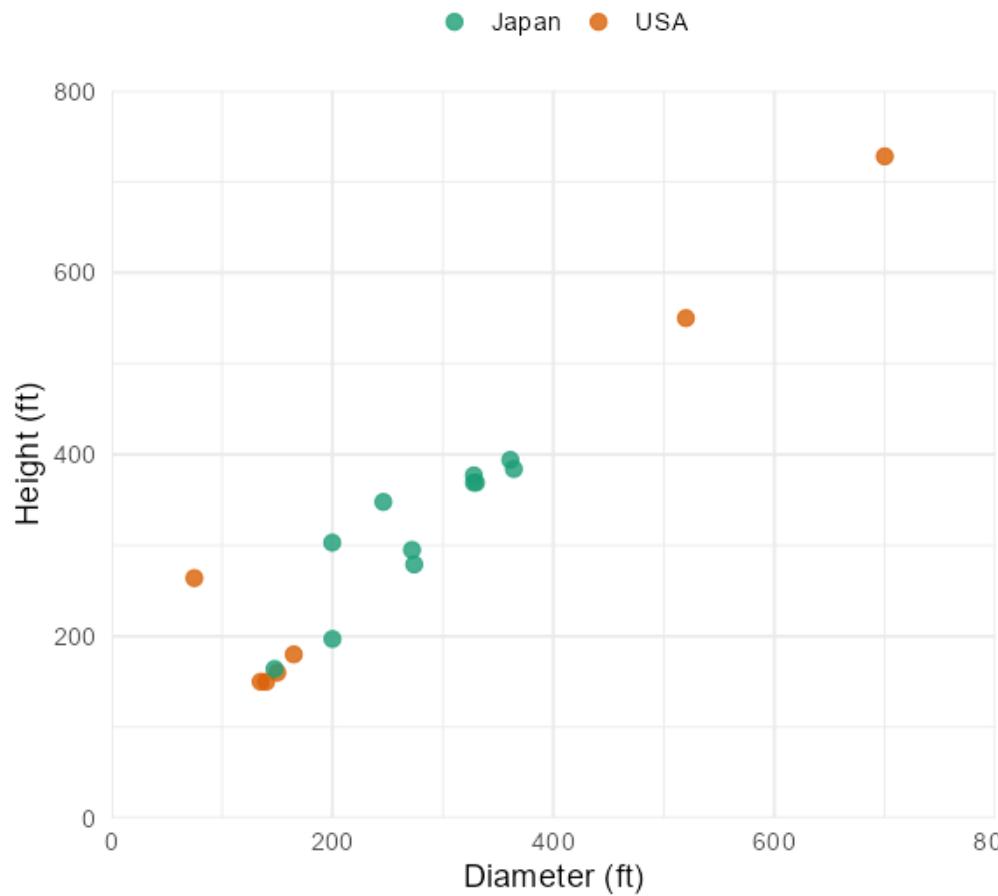


Qualitative

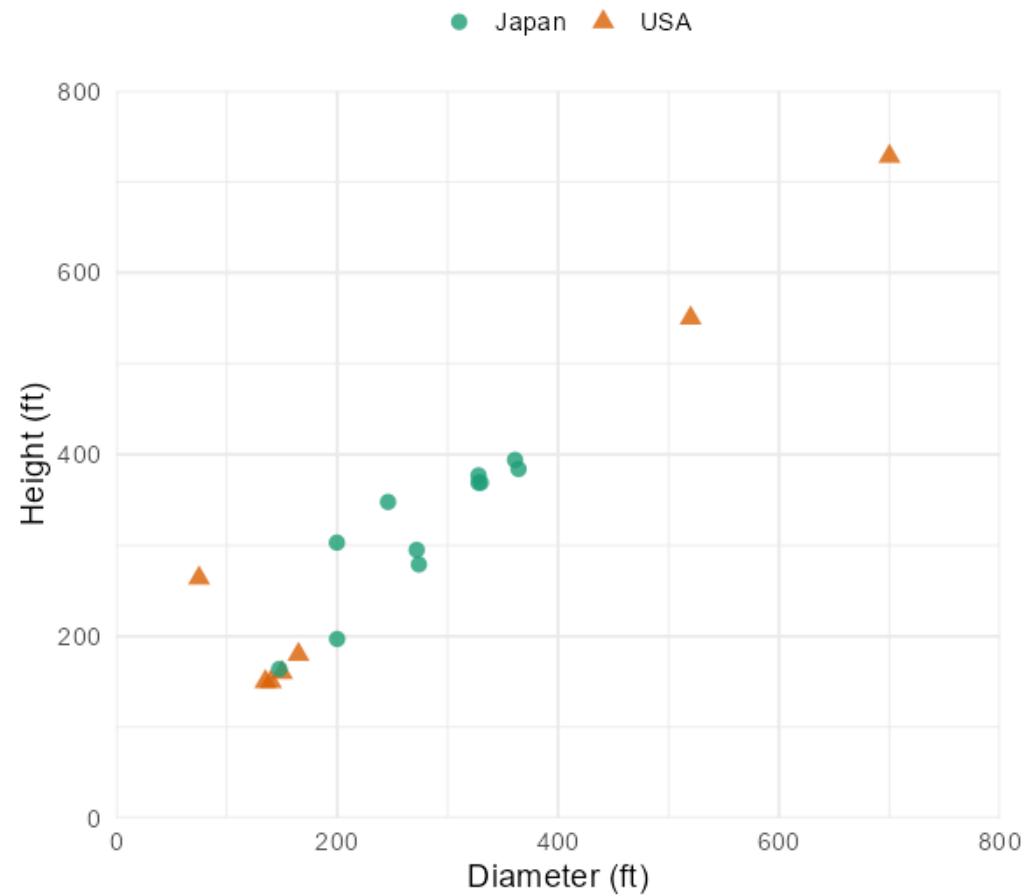


Colours: don't rely on colour

Ferris wheels



Ferris wheels



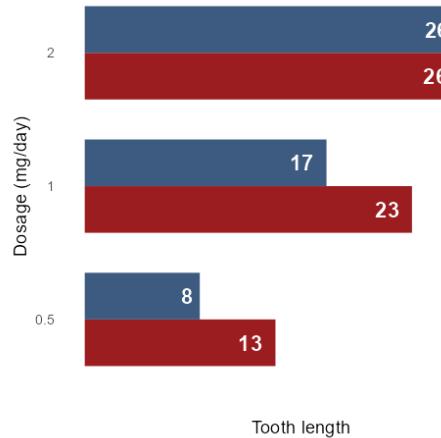
Colours: check for accessibility

- Different types of colourblindness
- Monochrome printing
- Ink usage

Colours: check for accessibility

Tooth Growth

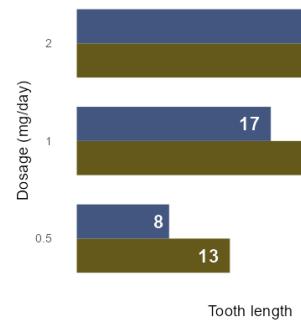
Each of 60 guinea pigs received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods: **orange juice** or **ascorbic acid**.



Deutanomaly

Tooth Growth

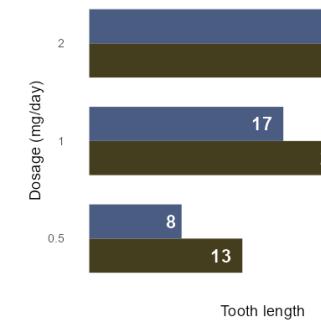
Each of 60 guinea pigs received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods: **orange juice** or **ascorbic acid**.



Protanomaly

Tooth Growth

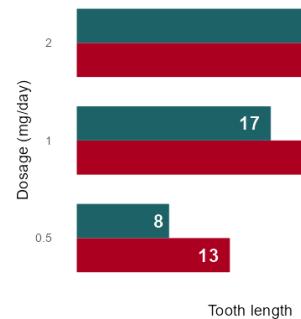
Each of 60 guinea pigs received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods: **orange juice** or **ascorbic acid**.



Tritanomaly

Tooth Growth

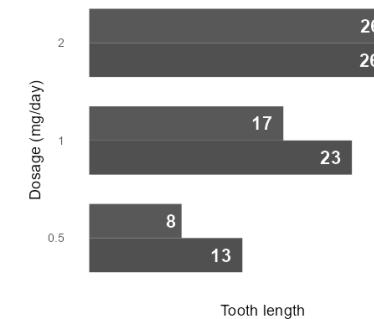
Each of 60 guinea pigs received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods: **orange juice** or **ascorbic acid**.



Desaturated

Tooth Growth

Each of 60 guinea pigs received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods: **orange juice** or **ascorbic acid**.

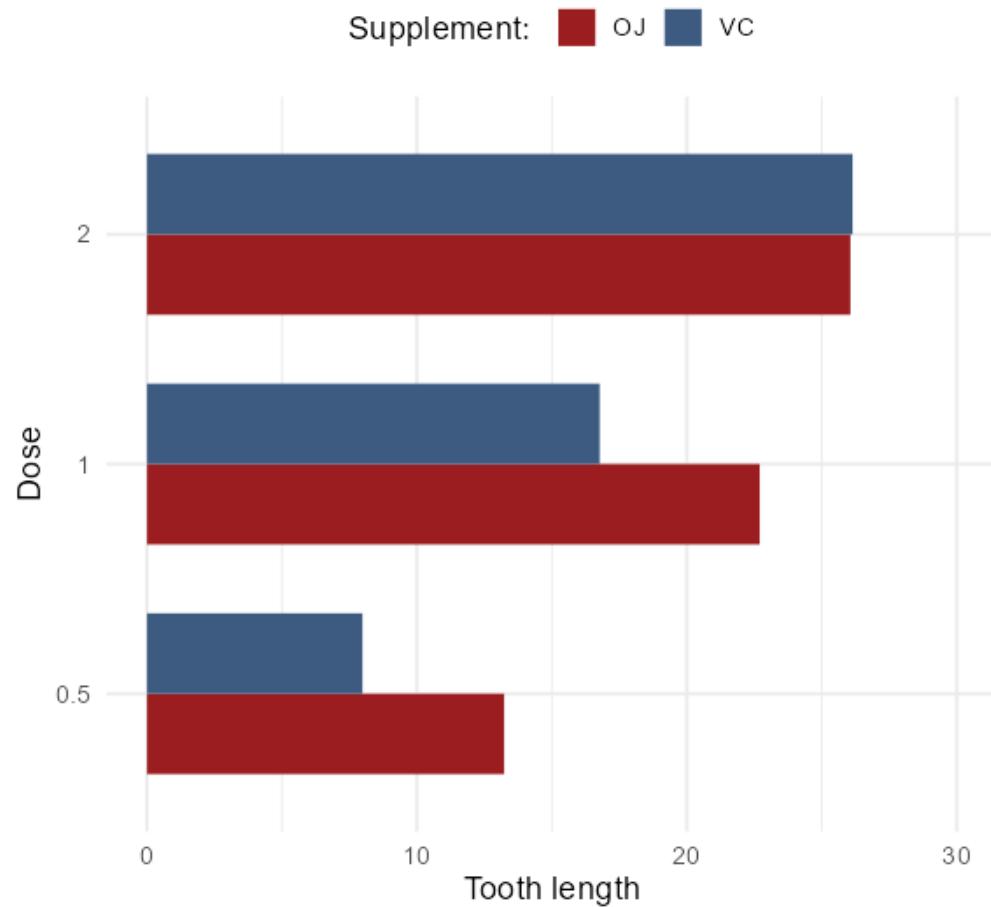


Annotations

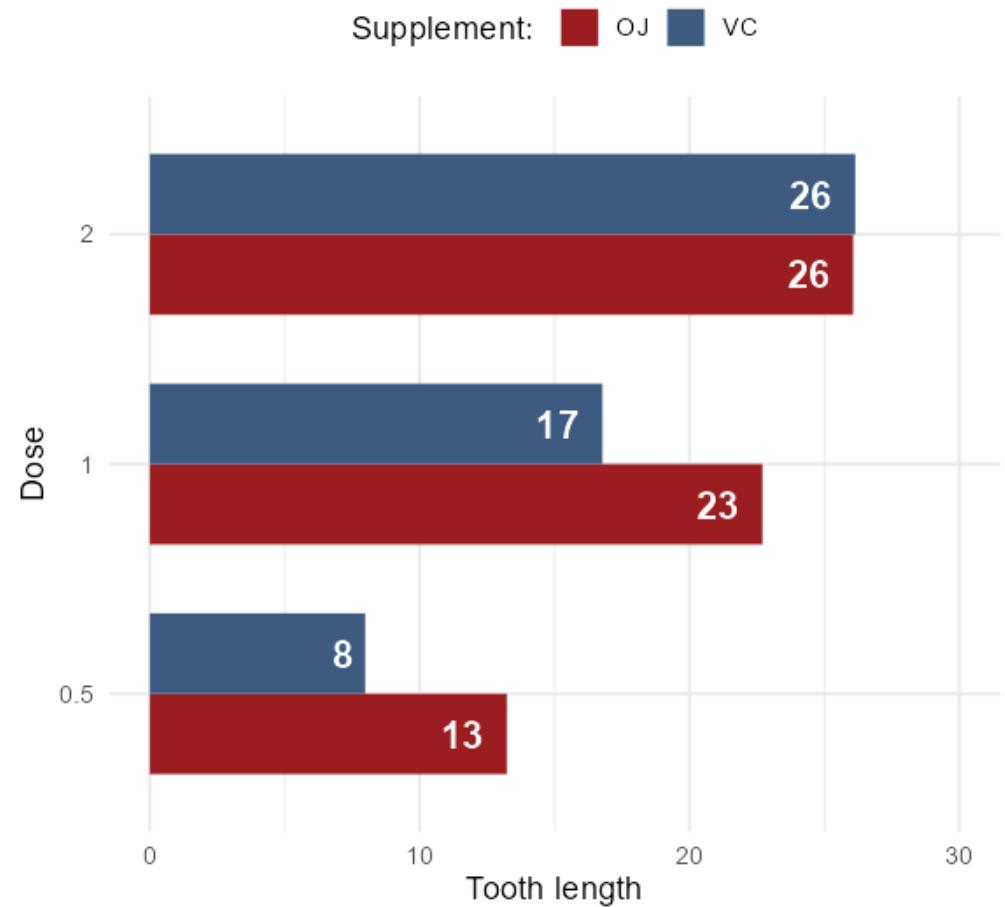
- Add clarification or detail
- Highlight an interesting data point
- Labelling data points (sometimes!)

Annotations

Tooth Growth



Tooth Growth



Fonts

- Font size: larger fonts are (usually) better
- Font colour: ensure sufficient contrast
- Font face: highlight text using bold font, avoid italics
- Font family: choose a clear font with distinguishable features

Fonts: font family

Arial: Does it pass the 1II test?

Times New Roman: Does it pass the 1I1 test?

Courier New: Does it pass the 1I1 test?

Alt Text

Alt text (AKA alternative text) is text that describes the visual aspects and purpose of an image – including charts.

Though alt text has various uses, its primary purpose is to aid visually impaired users in interpreting images when the alt text is read aloud by screen readers.

Alt Text

Writing Alt Text for Data Visualization

You probably can't write text that conveys the entire meaning of a chart. But, that doesn't mean that you shouldn't try.

alt= “**Chart type** of **type of data**
where **reason for including chart**”

Include a **link to data source**
somewhere in the text

Source: medium.com/nightingale/writing-alt-text-for-data-
visualization (Amy Cesal)
59 royal-statistical-society.github.io/datavisguide

Styling for RSS publications

Styling for Significance Magazine

- Following data visualisation guidelines
- Consistent font and colour choices
- Easy to implement!



{RSSthemes} R package

- Colour palettes
- Base R helper functions
- {ggplot2} helper functions

Installation

Install from CRAN:

```
1 install.packages("RSSthemes")
```

Installing from GitHub (development version):

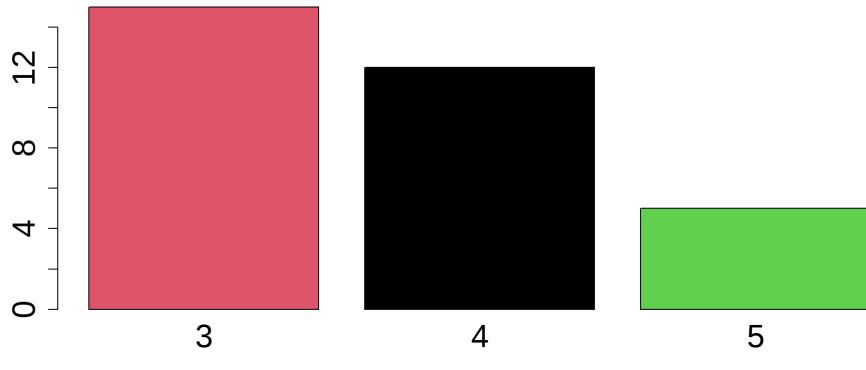
```
1 remotes::install_github("nrennie/RSSthemes")
```

Load package:

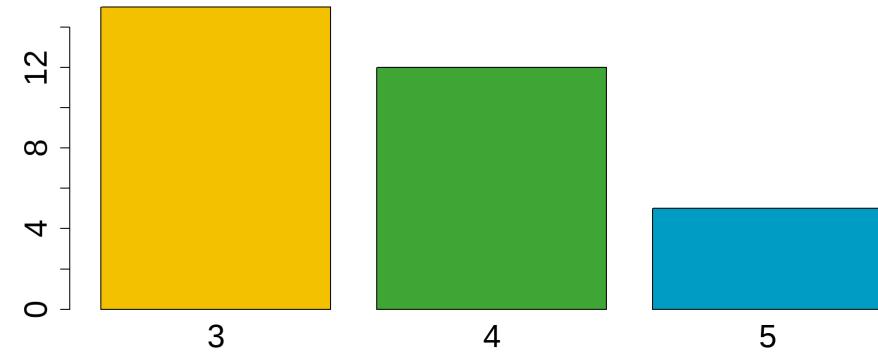
```
1 library(RSSthemes)
```

Example: plotting with base R

```
1 barplot(  
2   height = table(mtcars$gear),  
3   col = factor(  
4     unique(mtcars$gear)  
5   ),  
6   cex.axis = 4, cex.names = 4  
7 )
```



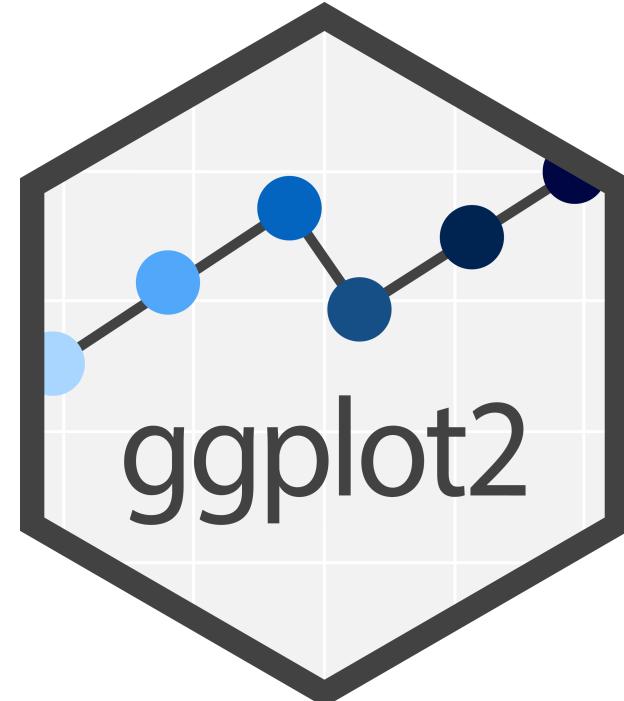
```
1 set_rss_palette("signif_qual")  
2 barplot(  
3   height = table(mtcars$gear),  
4   col = factor(  
5     unique(mtcars$gear)  
6   ),  
7   cex.axis = 4, cex.names = 4  
8 )
```



Example: plotting with ggplot2}

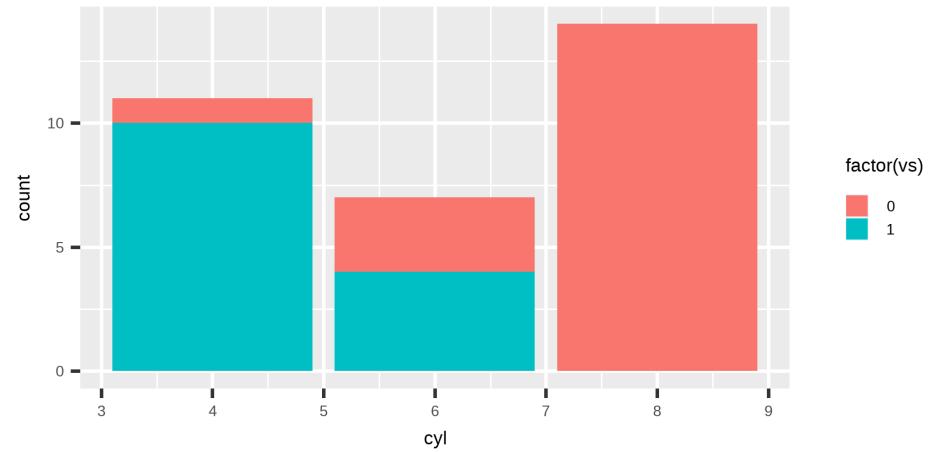
- {ggplot2} is an R package that provides functionality for drawing graphics.

```
1 install.packages("ggplot2")
```



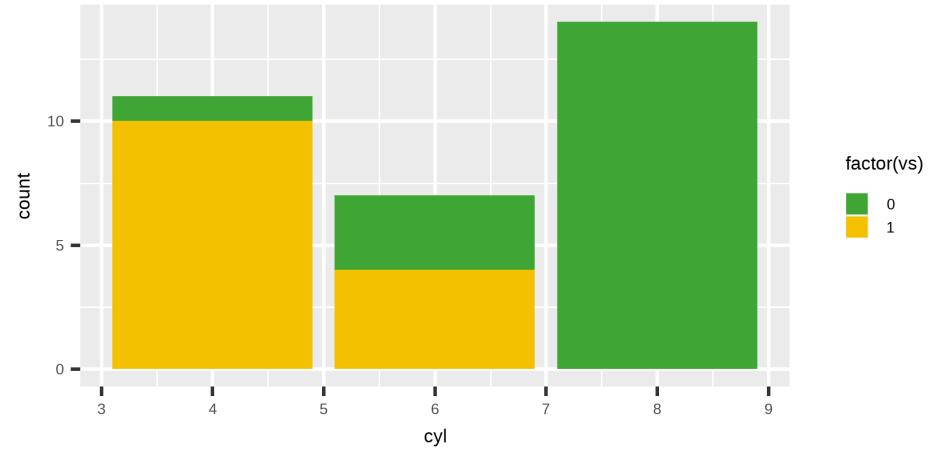
{ggplot2}: basic plot

```
1 library(ggplot2)
2 g <- ggplot(data = mtcars) +
3   geom_bar(
4     mapping = aes(
5       x = cyl,
6       fill = factor(vs)
7     )
8   )
9 g
```



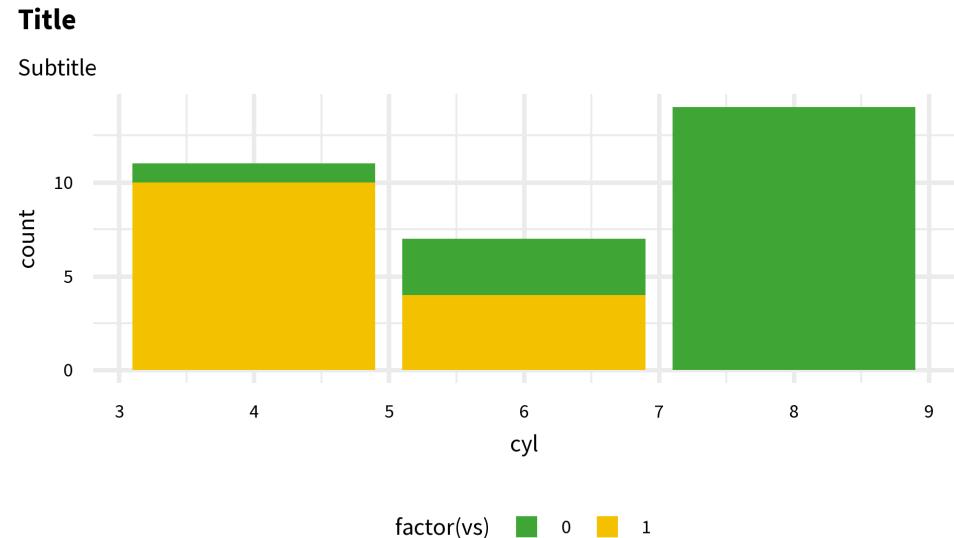
{ggplot2}: scales

```
1 g +  
2   scale_fill_rss_d("signif_qual")
```



{ggplot2}: theme

```
1 g +  
2   scale_fill_rss_d("signif_qual")  
3   labs(  
4     title = "Title",  
5     subtitle = "Subtitle") +  
6   theme_significance(base_size = 36)
```



I don't use R...

The guide also has examples of charts with:

- **Julia**: [datavisguide/docs/tools.html#julia](#) (community contribution)
- **Python**: [datavisguide/docs/tools.html#python](#)
 - Python package: [github.com/nrennie/RSSpythemes](#)
 - Currently contains colour palettes but not themes (yet)

Use something else?

- Help us add to the guide!

Questions?

Contributing to the guide

royal-statistical-society.github.io/datavisguide

Quarto

Quarto is an open-source scientific and technical publishing system that allows you to combine text, images, code, plots, and tables in a fully-reproducible document.

Quarto has support for multiple languages including R, Python, Julia, and Observable. It works for a range of output formats such as PDFs, HTML documents, websites, presentations,...



GitHub

The source code for the guide is stored on GitHub.

If you want to contribute to the guide, the easiest way is via GitHub.

- GitHub link: github.com/royal-statistical-society/datavisguide
- Contributor guide: royal-statistical-society.github.io/datavisguide/howto.html#how-to-contribute-to-this-guide

To ask a question or make a suggestion

Create or add to a GitHub discussion

The screenshot shows a GitHub Discussions page for the repository 'royal-statistical-society / datavisguide'. The top navigation bar includes links for Code, Issues (1), Pull requests (2), Discussions (selected), Actions, Projects, Wiki, Security, and Insights. A search bar at the top right contains the placeholder 'Type to search'. Below the navigation is a purple header bar with the title 'Visualization for ToothGrowth' and a green profile icon. Underneath the header, it says 'General · Scientist328455'. The main content area features a search bar with the query 'is:open', sorting options ('Sort by: Latest activity'), and filters ('Label', 'Filter: Open'). A prominent green button labeled 'New discussion' is visible. On the left, there's a sidebar titled 'Categories' with links for 'View all discussions' (which is selected and highlighted in blue), Announcements, General, Ideas, Polls, Q&A, and Show and tell. To the right of the sidebar, a discussion titled 'Visualization for ToothGrowth' by 'Scientist328455' is listed, showing it was started 3 weeks ago in the General category. At the bottom of the page, there are links for 'Most helpful', 'Code of conduct', 'royal-statistical-society.github.io/datavisguide', and 'Community insights'.



To report a bug or add a feature

Create an issue and describe:

- what the bug or error is, and add the issue tag *bug*
- what feature you want to include, and add the issue tag *enhancement*

Make a fork

Screenshot of a GitHub repository page for `royal-statistical-society / datavisguide`. The page shows a list of commits, an activity timeline, and various repository details.

The top navigation bar includes links for Code, Issues (2), Pull requests, Discussions, Actions, Projects, Wiki, Security, and Insights. The repository name `datavisguide` is shown as Public. The Fork button, which has 11 forks, is highlighted with a red box.

Key statistics on the page include:

- main branch
- 2 branches
- 0 tags
- 192 commits
- 457a50b (last commit, 2 days ago)
- 11 forks
- 67 stars

The repository description states: "Introductory guide to the art and science of data visualisation. Insights, advice, and examples (with code) to make data outputs more readable, accessible, and impactful."

Tags associated with the repository include: visualization, charts, data, graphics, and tables.

Links on the right side of the page include:

- Readme
- View license
- Code of conduct
- Activity
- 67 stars
- 5 watching
- 11 forks
- Report repository

Contributors are listed at the bottom, showing six individuals with their profile icons.

The footer of the page displays the URL `royal-statistical-society.github.io/datavisguide`.

Clone the repository

The screenshot shows a GitHub repository page for 'rssdatavisguide'. The 'Code' button in the top navigation bar is highlighted with a red box. A larger red box highlights the 'Clone' section of the dropdown menu, which contains links for HTTPS, SSH, and GitHub CLI, along with a web URL and an 'Open with GitHub Desktop' option.

This branch is 1 commit ahead of royal-statistical-society:main.

File	Description	Last Commit
.github/workflows	return to original workflow	3 weeks ago
docs	Minard img filename corrected	3 weeks ago
images	favicon added	3 weeks ago
.gitignore	update gitignore	3 weeks ago
.nojekyll	ghpages prep royal-statistical-	3 weeks ago
CODE_OF_CONDUCT.md	text edits	3 weeks ago
LICENCE.md	text edits	3 weeks ago
README.md	text updates to documentation	3 weeks ago
RSS-data-vis-guide.pdf	Render PDF	43 minutes ago
_make_pdf.qmd	added rss_conf to pdf	2 weeks ago
_quarto.yml	favicon added	4 days ago
about.qmd	dash to endash	3 days ago
howto.qmd	Add PDF link	3 weeks ago

About

Introductory guide to the art and science of data visualisation. Insights, advice, and examples (with code) to make data outputs more readable, accessible, and impactful.

[royal-statistical-society.github.io/datavis...](#)

[Readme](#) [View license](#) [Code of conduct](#) [Activity](#) [0 stars](#) [0 watching](#) [11 forks](#)

Releases

No releases published [Create a new release](#)

Packages

No packages published [Publish your first package](#)

Then make your changes and commit them...
79 royal-statistical-society.github.io/datavisguide

Create a pull request

Open a pull request, describe what changes it contains, reference any issues it describes, and wait for review.

The screenshot shows a GitHub pull request page for a repository named 'royal-statistical-society / datavisguide'. The pull request is titled 'added footnote to reader comment #40' and is currently 'Open'. A yellow banner at the top indicates that 'brtarran requested your review on this pull request'. The main content area shows a conversation between 'brtarran' and 'nrennie'. 'brtarran' commented 2 days ago, suggesting to link to a reader comment via a footnote. 'nrennie' responded by requesting a review from 'brtarran' 2 days ago. Below the conversation, there's a note about pushing commits to the 'link-ToothGrowth-comment' branch. The pull request summary on the right side includes sections for 'Reviewers' (nrennie), 'Assignees' (No one—assign yourself), 'Labels' (None yet), 'Projects' (None yet), 'Milestone' (No milestone), 'Development' (Successfully merging this pull request may close these issues), 'Notifications' (Customize, Unsubscribe), and '1 participant' (brtarran). At the bottom, there's a 'Leave a comment' section and a note that maintainers are allowed to edit the pull request.

Need help with GitHub?

Ask us!

RSS Conference session: GitHub: Version control for research,
teaching and industry, Thu 7th @ 11:40

Questions?

Links

 royal-statistical-society.github.io/datavisguide

 github.com/royal-statistical-society/datavisguide

 nrennie.github.io/rss-conference-2023-data-viz-slides

royal-statistical-society.github.io/datavisguide