WOMBAT 2025 Tutorial

Visualising Uncertainty

Harriet Mason, Dianne Cook
Department of Econometrics and Business Statistics

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Welcome 👋🏻

Thanks for joining to learn about making data plots today.

🦘 Di is a *Professor of Statistics*. She has more than 30 years of research and teaching of data visualisation, and open source software development.
🐨 Harriet is a *final year PhD student*, working on better representation of uncertainty in data visualisations, particularly focused on spatial data.

We are both in Econometrics and Business Statistics, at Monash University.

🧩 Feel free to ask questions any time. 🤔

🎯 The objectives for today are:

1. understand what uncertainty means, in particular in relation to data visualisation.
2. be able to plot represent uncertainty on different types of plots.
3. assess whether one representation may be better than another, based on cognitive perception principles and visual testing.
4. apply these new approaches for showing uncertainty in spatial data visualisation.

Caution: Incorporating uncertainty into plots is far from a finshed state of best practices. This is our best attempt to summarise current literature, what we find to be valuable approaches and available tools.

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Session 1
Foundations of uncertainty in data visualisations

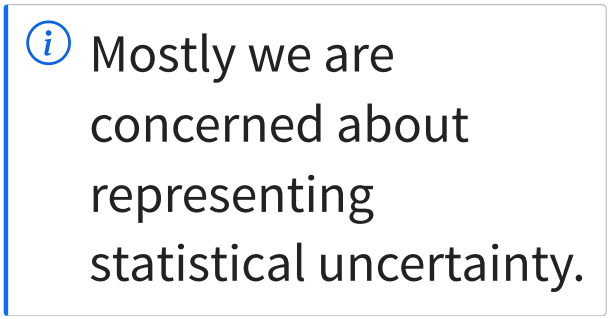
https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Introduction

What is uncertainty?

You don't know what you don't know

- Statistical (aleatory) uncertainty
  - notion of randomness
  - variability in the outcome/measurements
- Systemic (epistemic) uncertainty
  - due to bias, misunderstanding, assumptions
  - measurement error
  - handling of missing information, and pre-processing choices
  - model choices
  - incorrect comparisons
  - *can you think of any others?*



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Showing uncertainty

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Show the data (1/4)

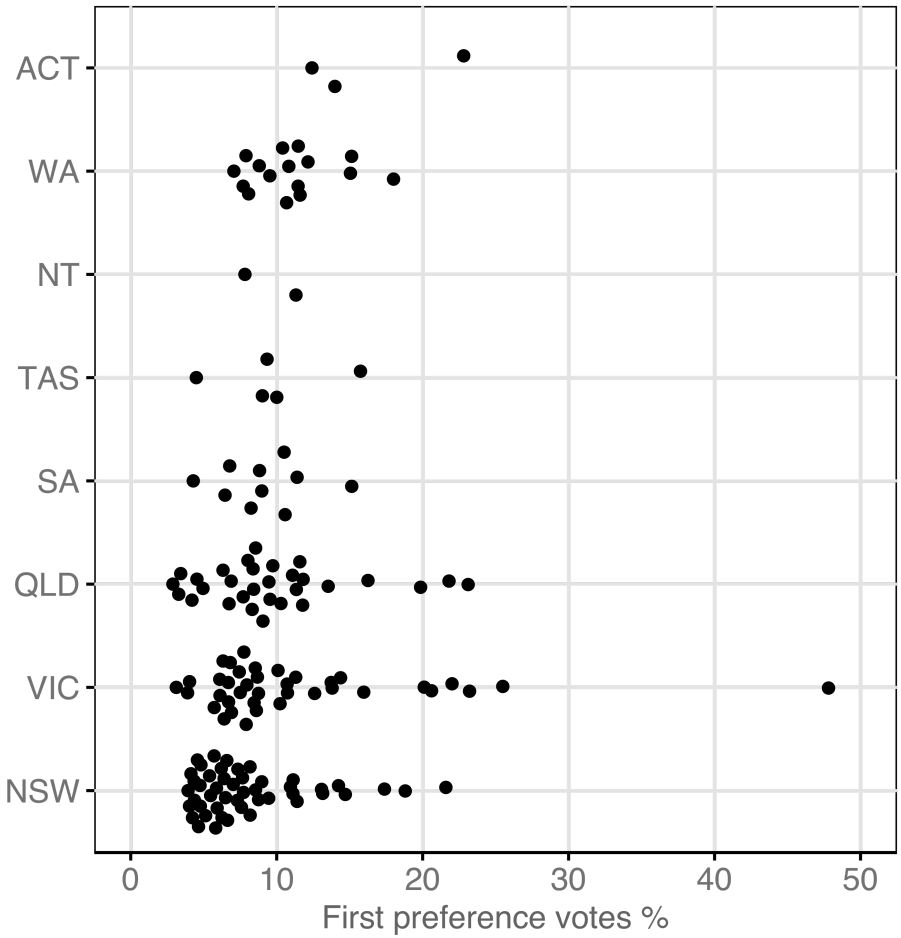
https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

> ⓘ The most valuable way to show uncertainty is to show all the data.

Plot shows first preference % for greens in the 2019 Australian Federal Election, for the 150 electorates.

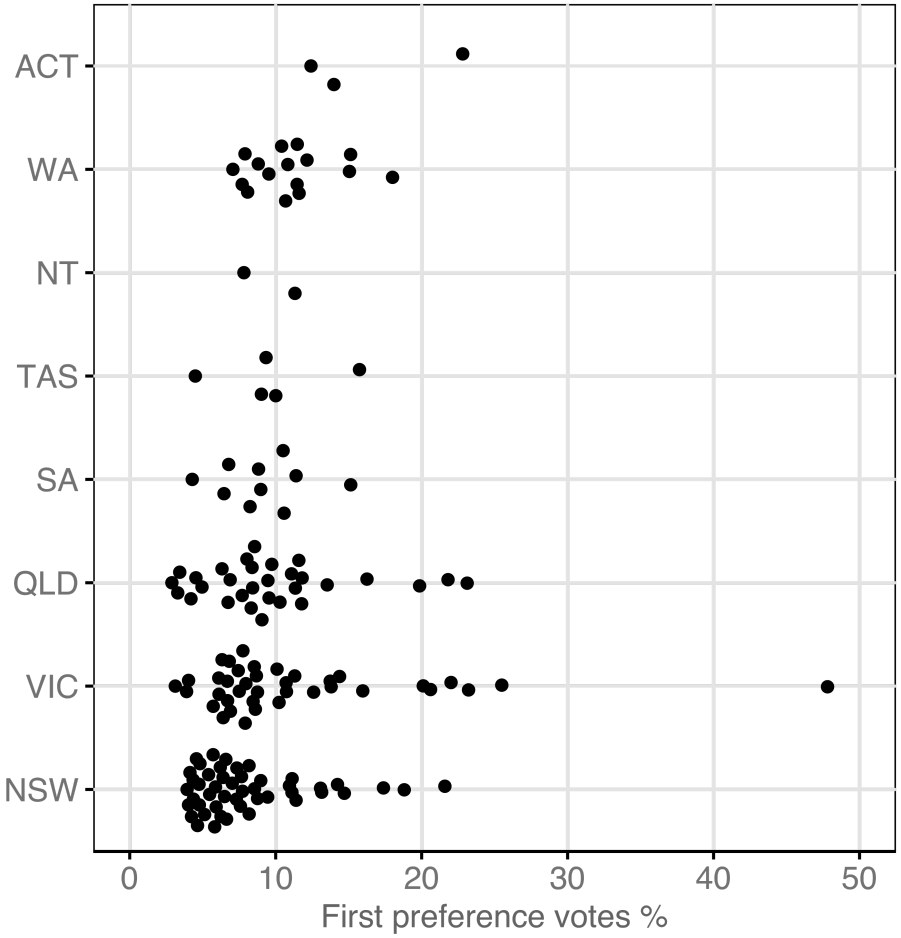
Plot of choice is the **jittered dotplot**, where points are spread vertically according to density.

▶ Code



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Show the data (2/4)



What do we learn?

- Different number of observations in each state
- One outlier in Vic
- As a group, ACT has higher %'s
- Vic has a small cluster of points with higher %'s
- %'s are mostly very low

This plot ONLY shows uncertainty!

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Show the data (3/4)

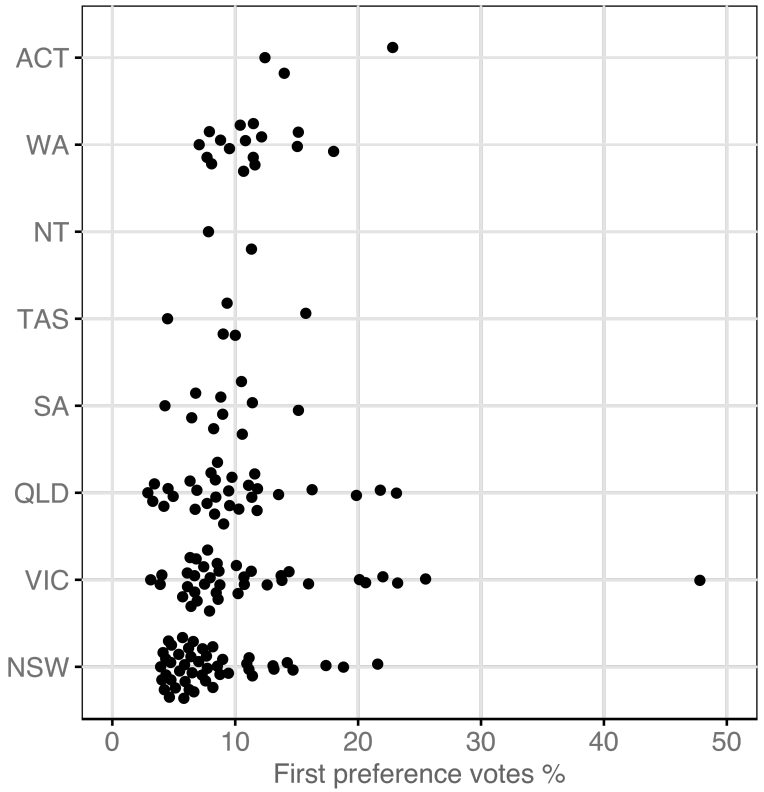
What would be other common ways to display this data?

- Side-by-side boxplots
- Side-by-side violin
- On a map of electorates

For each plot think about

- what is uncertainty, and what is estimate
- what the plot shows or hides

Dotplot | Boxplot | Violin | Map



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Show the data (4/4)

> ⓘ Even when you think you are showing the data, it is often an estimate and some representation of uncertainty.

The election data is actually estimates. The electorates are strata, so what was shown was % computed on each strata.

What is the full data? What are different strata possible?

Generally, we trust the values provided by AEC, and we explore the distribution of votes by different strata in the electorate structure. The goal being to understand the variability in the way the people have voted, identify electorates where the winner might flip next time, …

It's really difficult to concretely define uncertainty!

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Terminology

Names for main thing:

- estimate
- statistic
- signal

Names for uncertainty, needed to understand main thing:

- variation
- variability
- variance/standard deviation
- error/standard error
- IQR/MAD
- noise

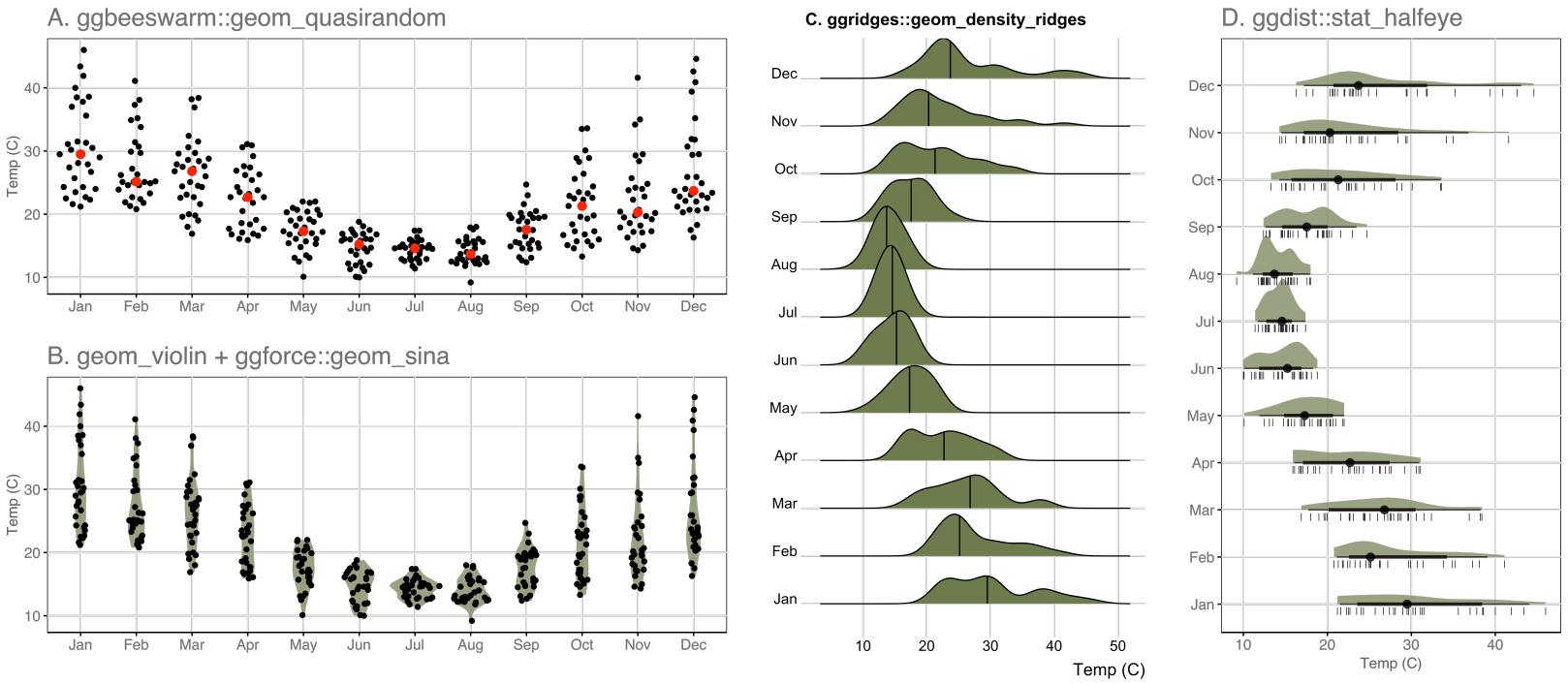
> ⓘ Displaying uncertainty is described **signal suppression**.

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Example: distributions

Variations in plot choices | Consider these things

▶ Code



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Exercise 1

Continue working with the Melbourne temperature data:

1. Decide the appropriate information about the uncertainty to include in the display.
2. Play with different options on your choice of display to make various displays. Aim to have three different designs.
3. Is there a winner, or several roughly equally good displays?

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

How this affects perception

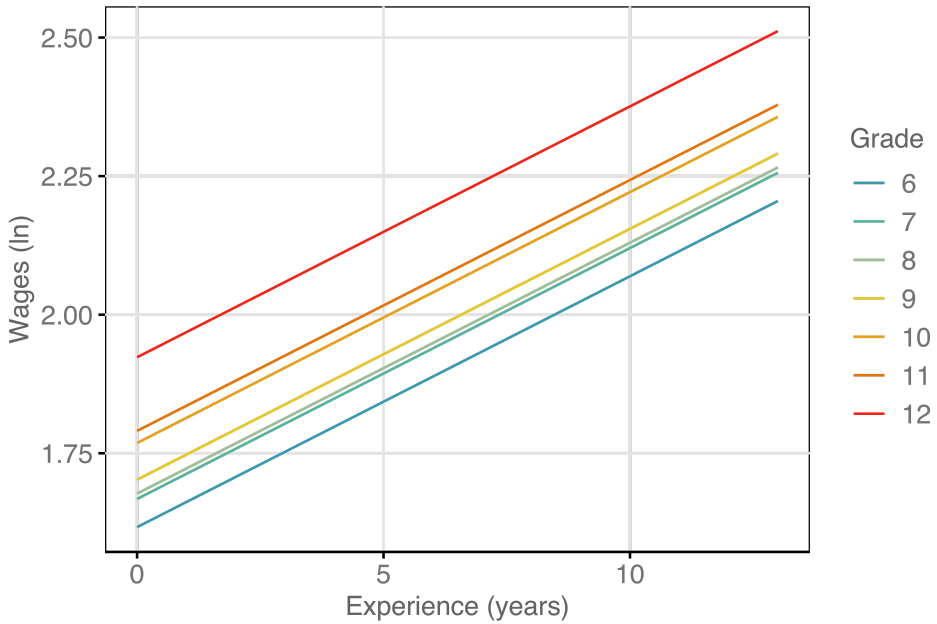
https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Why it it “signal suppression”?

model | and data | and SE | individual fits | which is honest?

Plotting the fitted model alone

▶ Code



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Perception and uncertainty

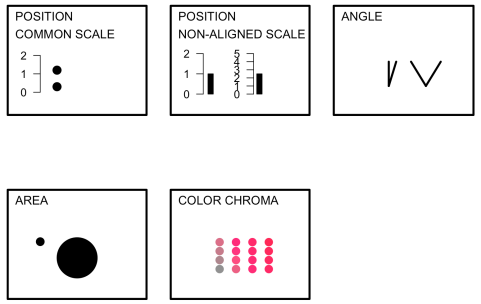
- There can be multiple levels of uncertainty:
  - **Example had two measurement levels** - fixed effect/demographic strata, random effects/individuals - but multiple ways to represent these.
  - A **classical example** is in a simple regression we have *confidence* intervals for the model estimates, and also *prediction* intervals for predicting new observations.
- Uncertainty is not just another variable, but some measures might be handled this way, and encoded in the data as an extra column, e.g. standard error, IQR. This can be used in the plot to incorporate uncertainty.
- You need to decide what is the appropriate uncertainty measure for the problem.

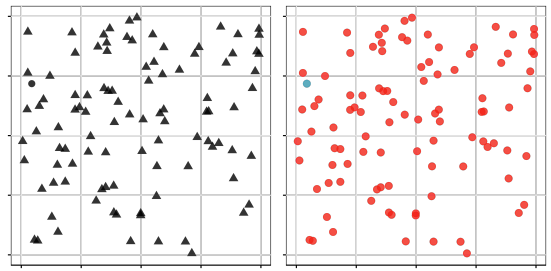
> ⓘ Including uncertainty representation adds to the complexity, and multiple elements on a plot can interfere with the perception of either one.

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Perceptual principles

- **Hierarchy** of mappings:
  1. *Position - common scale* (BEST): scatterplot, barchart
  2. *Position - nonaligned scale*: side-by-side boxplot, stacked barchart
  3. *Length, direction, angle*: piechart, rose plot, gauge plot, donut, wind direction map, starplot
  4. *Area*: treemap, bubble chart, mosaicplot
  5. *Volume*, curvature: chernoff face
  6. *Shading, color* (WORST): choropleth map
- **Pre-attentive**: noticed before you even realise it.
- **Color palettes**: qualitative, sequential, diverging.
- **Proximity**: Place elements for primary comparison close together.
- **Change blindness**: When focus is interrupted differences may not be noticed.





https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Applying these to making plots

- Make **primary information prominent**: use pre-attentive elements like colour, make larger.
- Place items to **compare first near each other**, main thing + uncertainty.
- When using colour to map a **two-ended** continuous variable use a **diverging** palette, and conversely, for *low-high* continuous variable such as confidence use a *sequential* palette.
- **Axes and labels should be faint** in the background, to be examined only when needing to interpret or quantify patterns.
- **Order items**, categories by a numerical rank.
- For separate plots, **common axes** to make comparison easier.

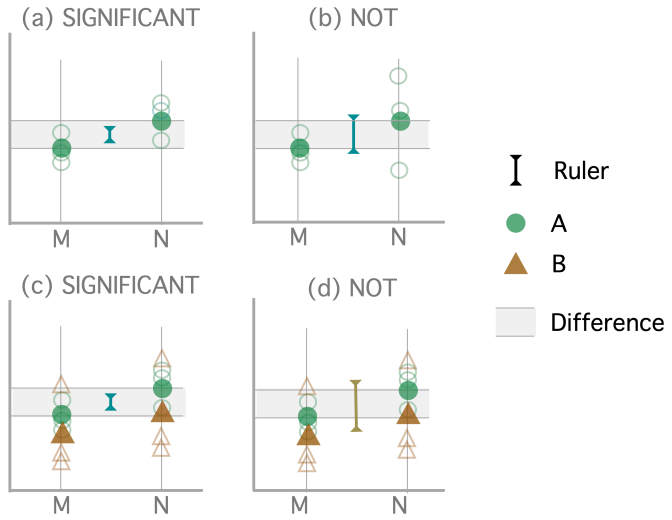
Accessibility considerations when adding uncertainty to plots are harder because there is more information provided.

- What is the focus? line, or the confidence bands?
- Does the uncertainty representation obscure the signal?
- Reading colour is already hard. For spatial plots where a 2D palette is used for signal and uncertainty is hard to read. Mapping a variable to saturation hurts accessibility. *Stay tuned for session 2!*

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Application to uncertainty visualisation

1. Make the **main thing pre-attentive**. It will draw attention to the primary information, such as trend, median, mean, estimate, first.
2. Map the **uncertainty to a lower level of attentiveness**, so it can be considered secondarily.
3. Uncertainty needs to be **placed** with the main thing in order to **make comparison**. The purpose of including uncertainty representation is to compare pattern in the main thing relative to variation remaining.



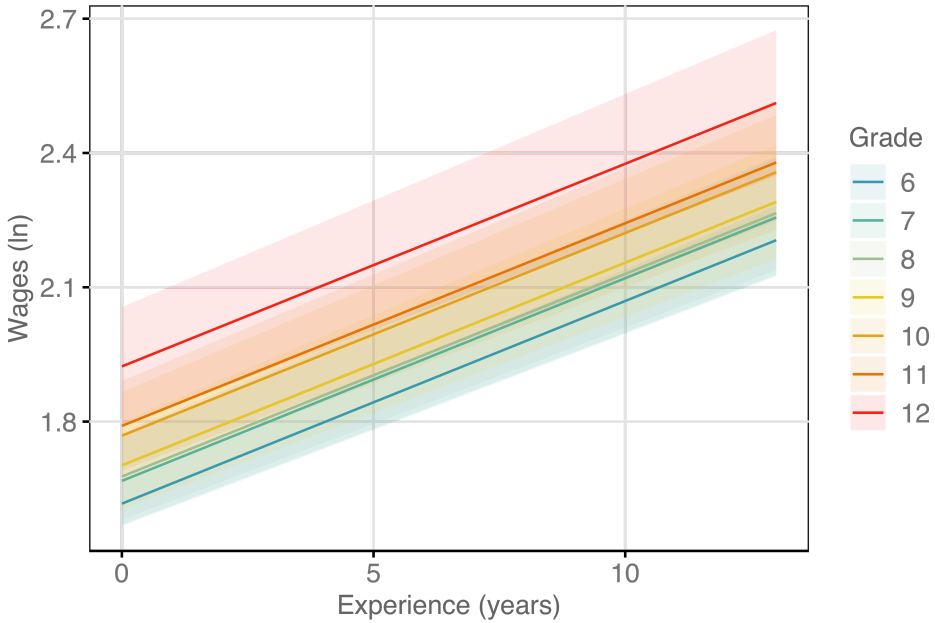
Source: http://dx.doi.org/10.4172/2153-0602.1000139

At the core of statistical reasoning is asking **compared to what.**

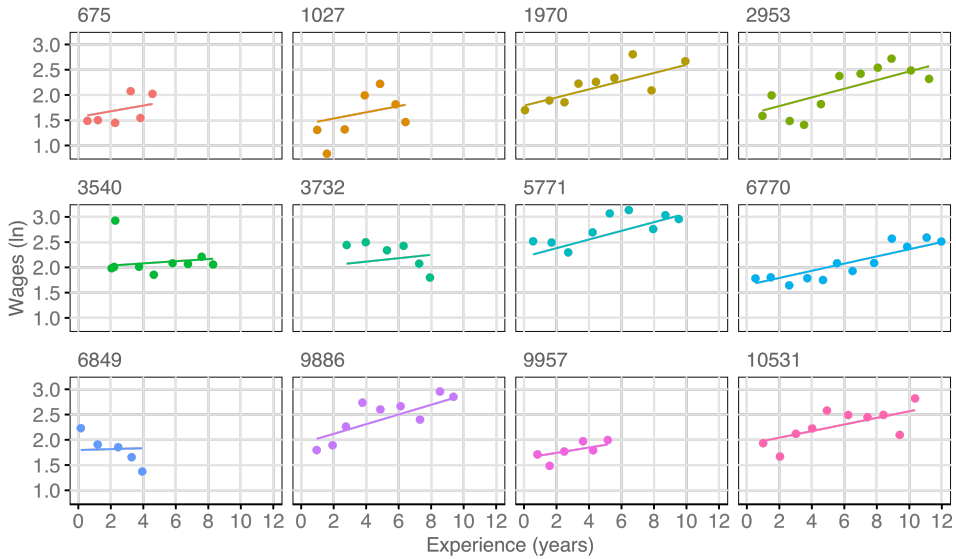
https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Example

What principle(s) is this using?



What principle(s) is this using?



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Exercise 2

1. Take a look at the plots made in Exercise 1. Ask yourself whether the main thing is pre-attentive, and the uncertainty representation is sitting a little into the background?
2. Tinker with the design of one plot to make it better fit this principle.

Inspirations | ggbeeswarm | violin + sina

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

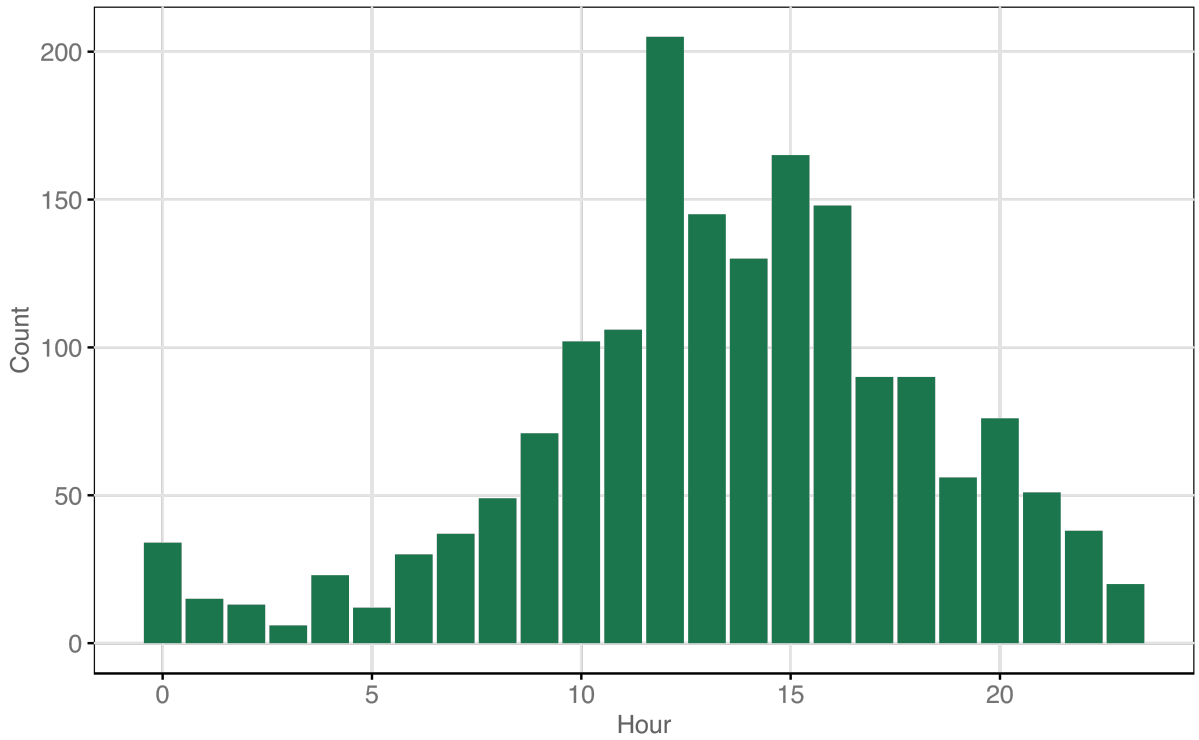
Common measures and representations

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Barcharts

Melbourne pedestrian counts at Southern Cross Station, Sunday Aug 31, 2025.

Bars | bar+CI | CI | Gradient | Ribbon | LineRibbon | Loess | Code



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Broader applicability

The approaches used on the barcharts here are the same approaches that apply to many other types of displays.

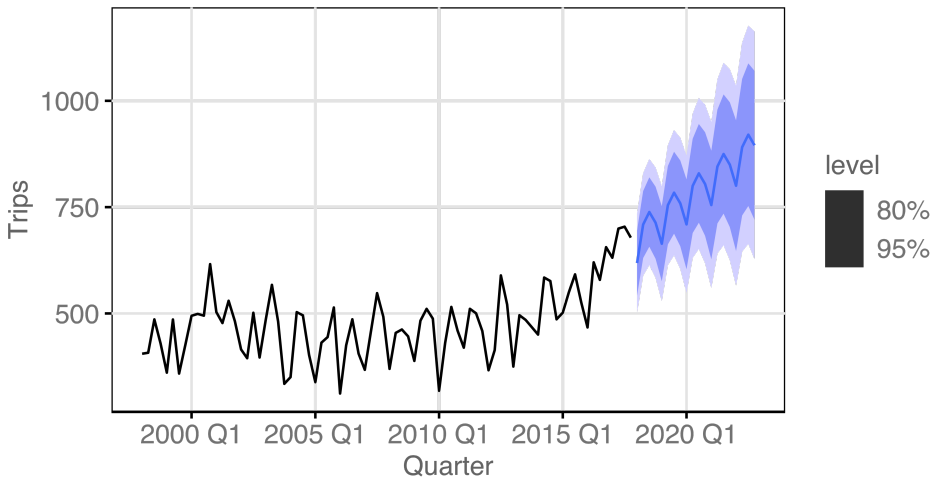
- Error bars
- Error bands
- Gradients
- Multiple samples, such as bootstrap or simulation (stay tuned!)

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Exercise 3

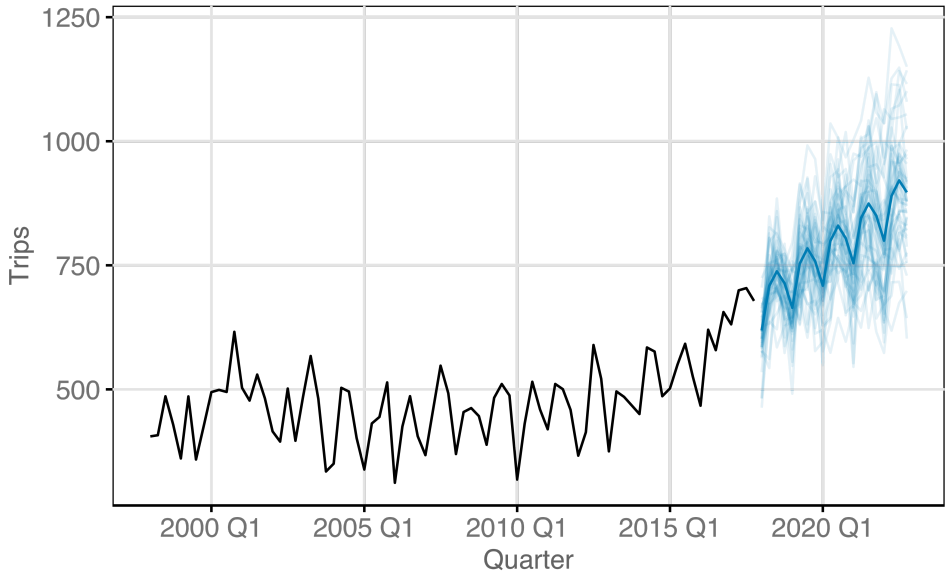
This is a forecast of business trips to Melbourne for 2018-2022, based on data from 2000-2017. How is the uncertainty represented?

▶ Code



In what ways do you think this representation is better (or worse) than the previous representation?

▶ Code



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Deciding which is the best design

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Evaluation criteria

Uncertainty visualisation should:

1. Reinforce signals that are important
2. Hide signals that are primarily noise

to enable making the better decisions and conclusions, or dare we say, inference.

Objective testing procedure:

- **Simulate** data to *control pattern* being examined.
- Use the **lineup protocol** to determine if a reader can *detect the structured plot from not structured null plots*.

We'll illustrate this with a simplified pedestrian count plot example.

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Try this

start | 1a | 1b | 2a | 2b

01:00

Process

- group **a** or **b**
- pick the plot that is different
- record choice, and why that plot looks different

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Lineup experiment

Simulated data setup | Testing protocol | Code

- **Time**: 8 instead of 24
- **Count**: pattern is large wave
- **Variance model**: uniform ranging from small to large. Poisson used to generate count.

▶ Code

End of session 1