

Interpreting a Data Visualisation

SMI105: Week 2

Dr. Calum Webb

Sheffield Methods Institute, the University of Sheffield
c.j.webb@sheffield.ac.uk

Sign in



Learning outcomes

What will I learn?

By the end of this week you will:

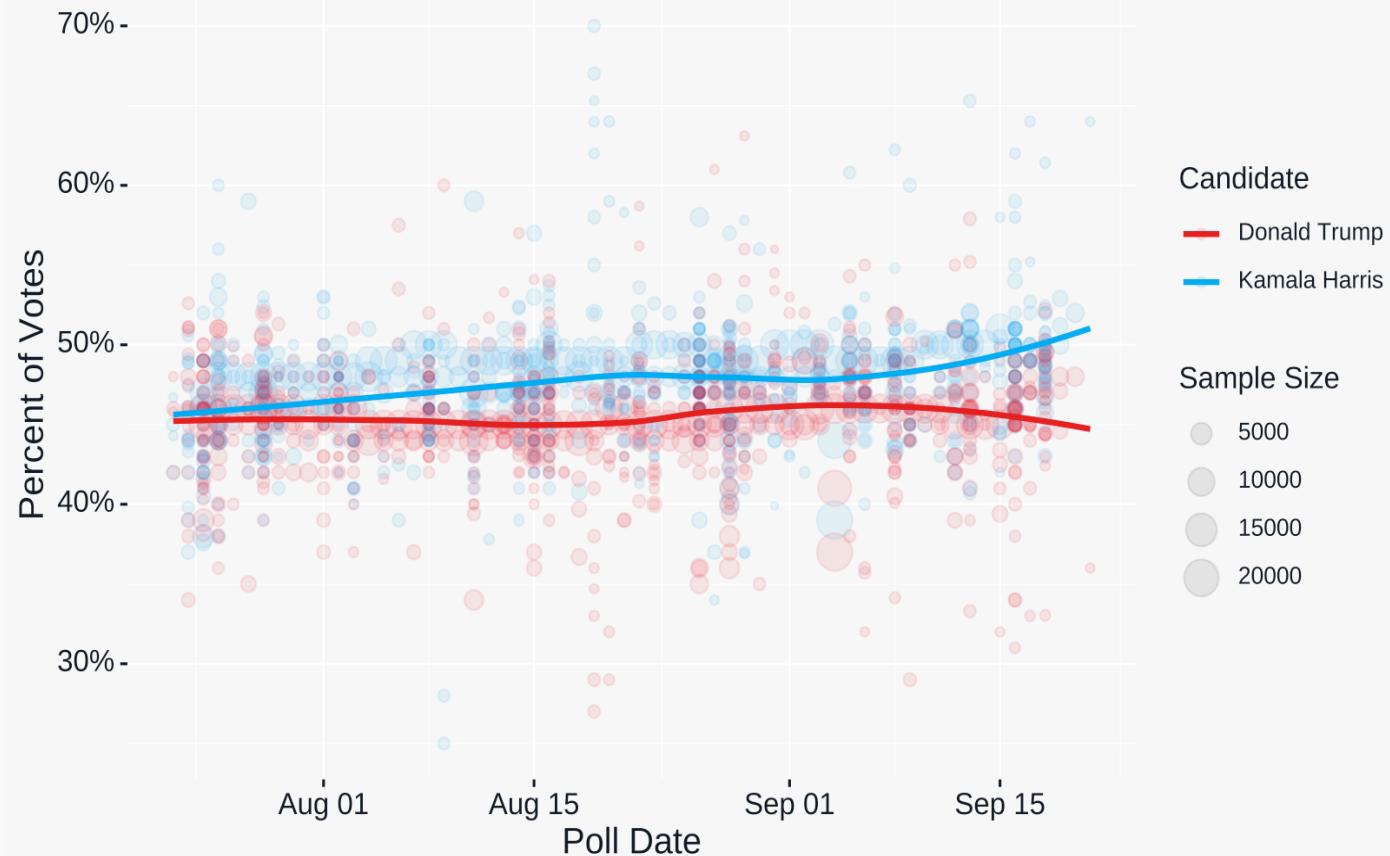
- Lecture: How to 'read' a data visualisation, how to identify key structural and aesthetic parameters in their construction, and how they can be manipulated.
- Lecture: What social and societal factors influence engagement.
- Workshop: How to create bar charts to visualise data

Part I

The anatomy of a data visualisation



Current polling data for 2024 US Presidential Election
 Line indicates average of polls, dots indicate individual poll results

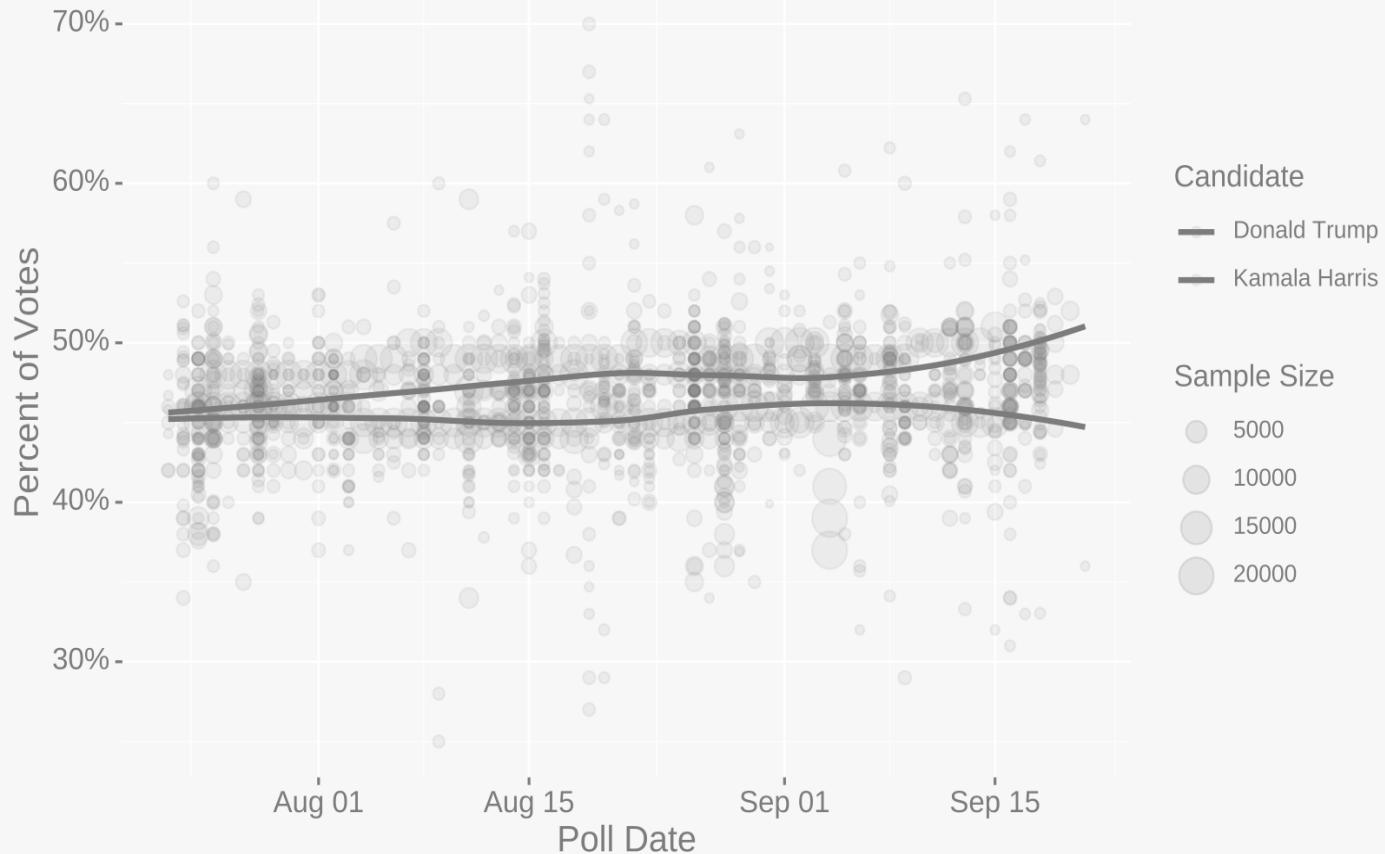


The Anatomy of a Chart

Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

Current polling data for 2024 US Presidential Election

Line indicates average of polls, dots indicate individual poll results



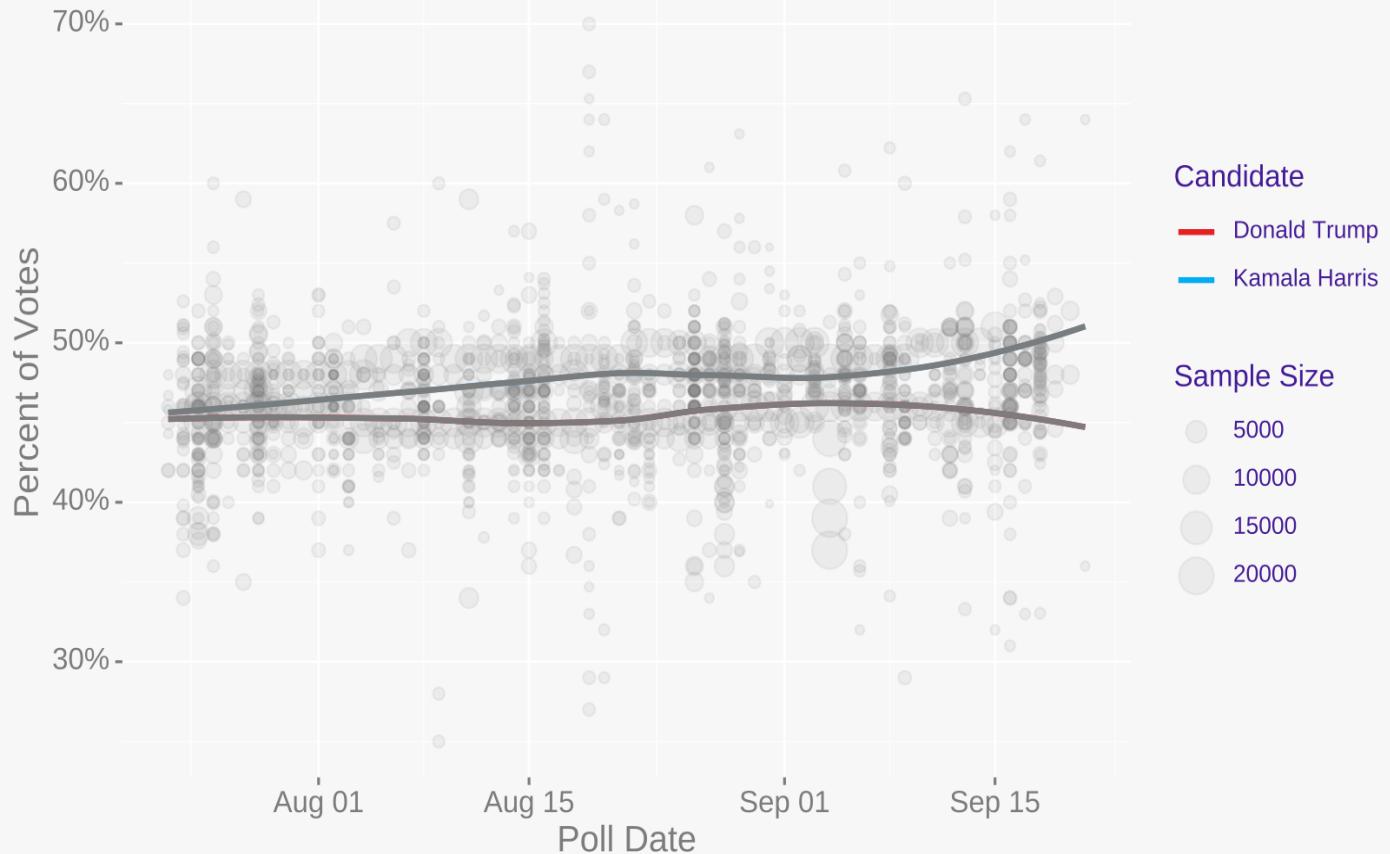
Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

The Anatomy of a Chart

- **The Title**
 - Is there a title?
 - Does the title give you information about what to expect from the chart?
 - Does the chart show or support what the title states?

Current polling data for 2024 US Presidential Election

Line indicates average of polls, dots indicate individual poll results

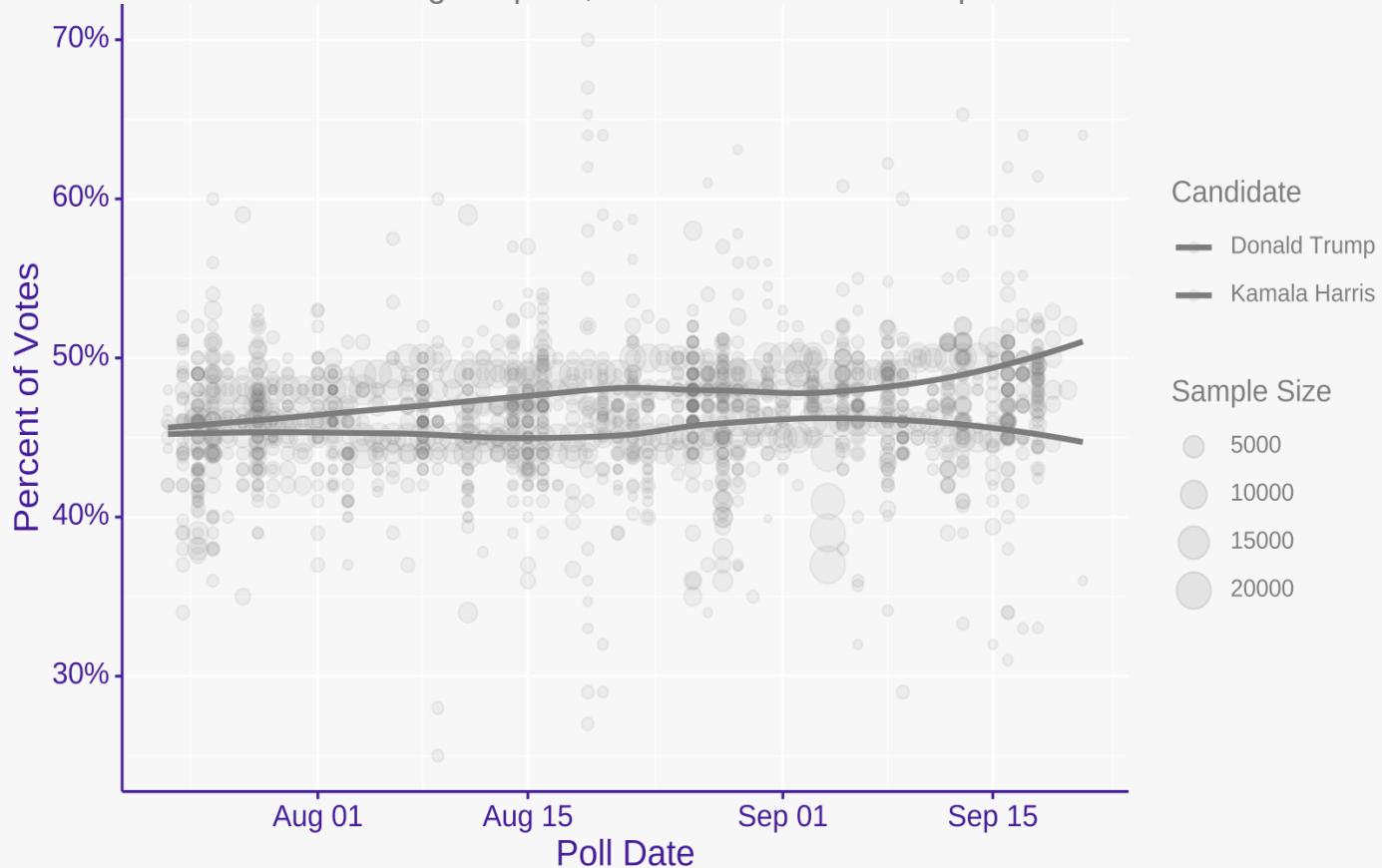


Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

The Anatomy of a Chart

- The Title
- **The Key/Legend**
 - Is there a key?
 - What does it tell you about the different colours & shapes in the chart?
 - Does it help you understand how to read it?
 - The key is not always a separate element, sometimes it can be embedded in the title (e.g. certain colours highlighted)

Current polling data for 2024 US Presidential Election
Line indicates average of polls, dots indicate individual poll results



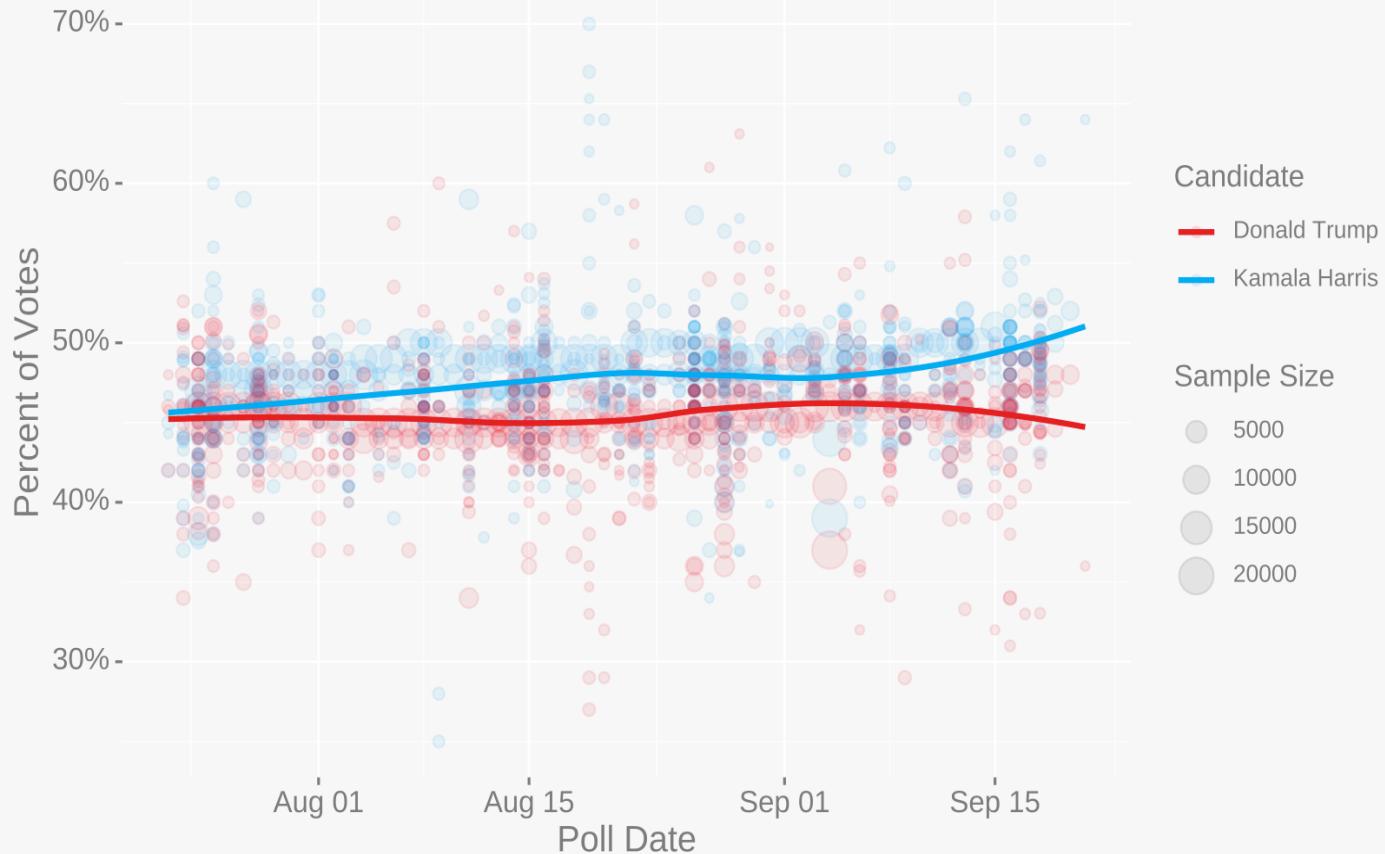
Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

The Anatomy of a Chart

- The Title
- The Key/Legend
- **The Axes & Scales**
 - What are the scales? (positions on X & Y axis, size of points, shade of colour)
 - Do the scales start at 0? If not, why not?
 - What is the range of values?
 - Which values are good or bad, high or low?
 - Have the scales been transformed in any way or are they linear? (linear means that, for example, an increase of 5 covers the same distance anywhere on the axis)

Current polling data for 2024 US Presidential Election

Line indicates average of polls, dots indicate individual poll results



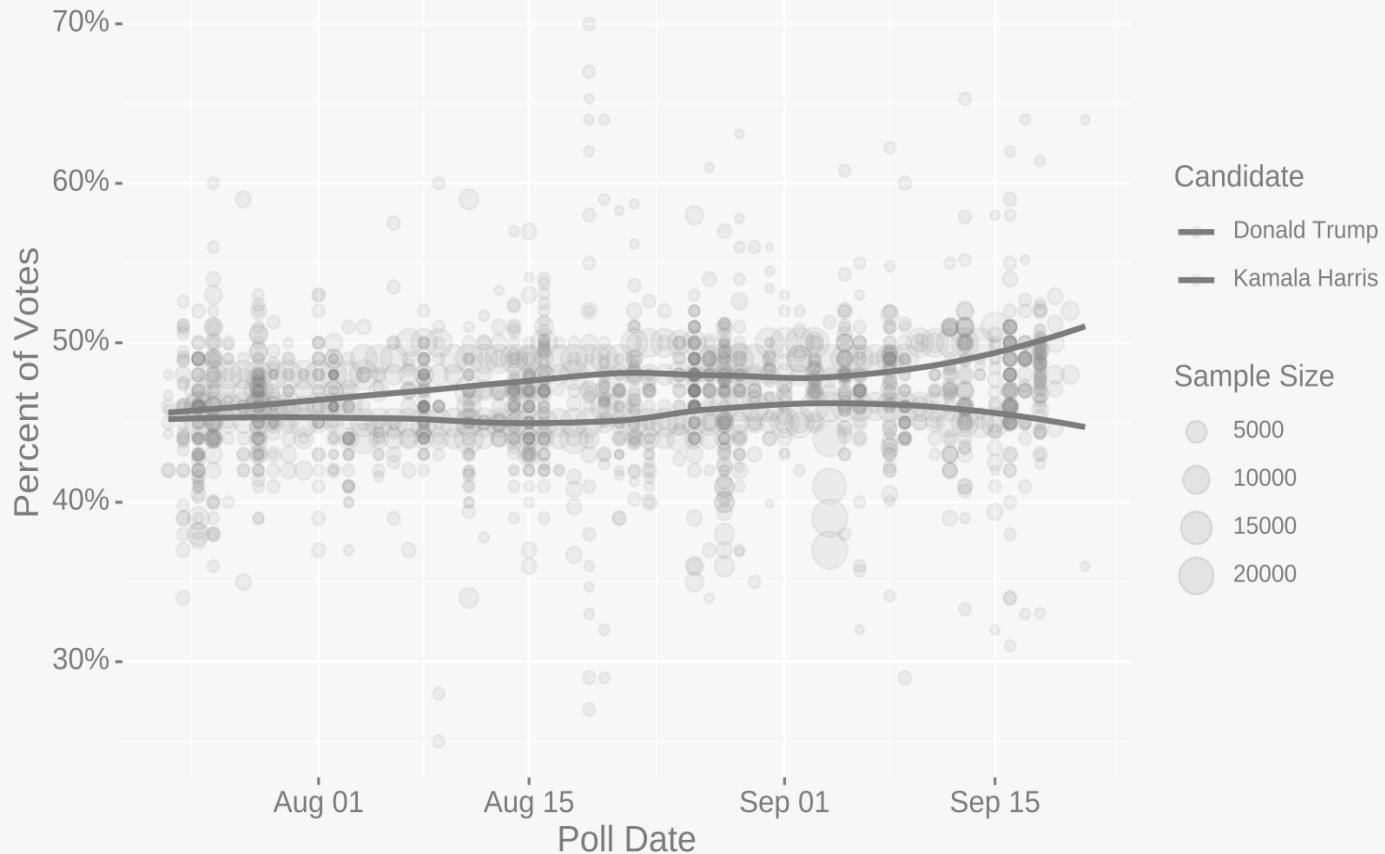
Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

The Anatomy of a Chart

- The Title
- The Key/Legend
- The Axes & Scales
- **The Data**
 - What are the biggest categories? Do they represent proportions, relative differences, or frequencies? (Bar Charts/Histograms/Area Charts)
 - What is the relationship between two variables? (e.g. Scatterplots, Heatmaps, Stacked/Grouped Bar Charts)
 - What is the relationship over time (when one axis is time, line charts)? Is upward movement good or bad? Do lines interact? Are there seasonal trends?
 - How much *variation* is there, and is variation shown at all?

Current polling data for 2024 US Presidential Election

Line indicates average of polls, dots indicate individual poll results



Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

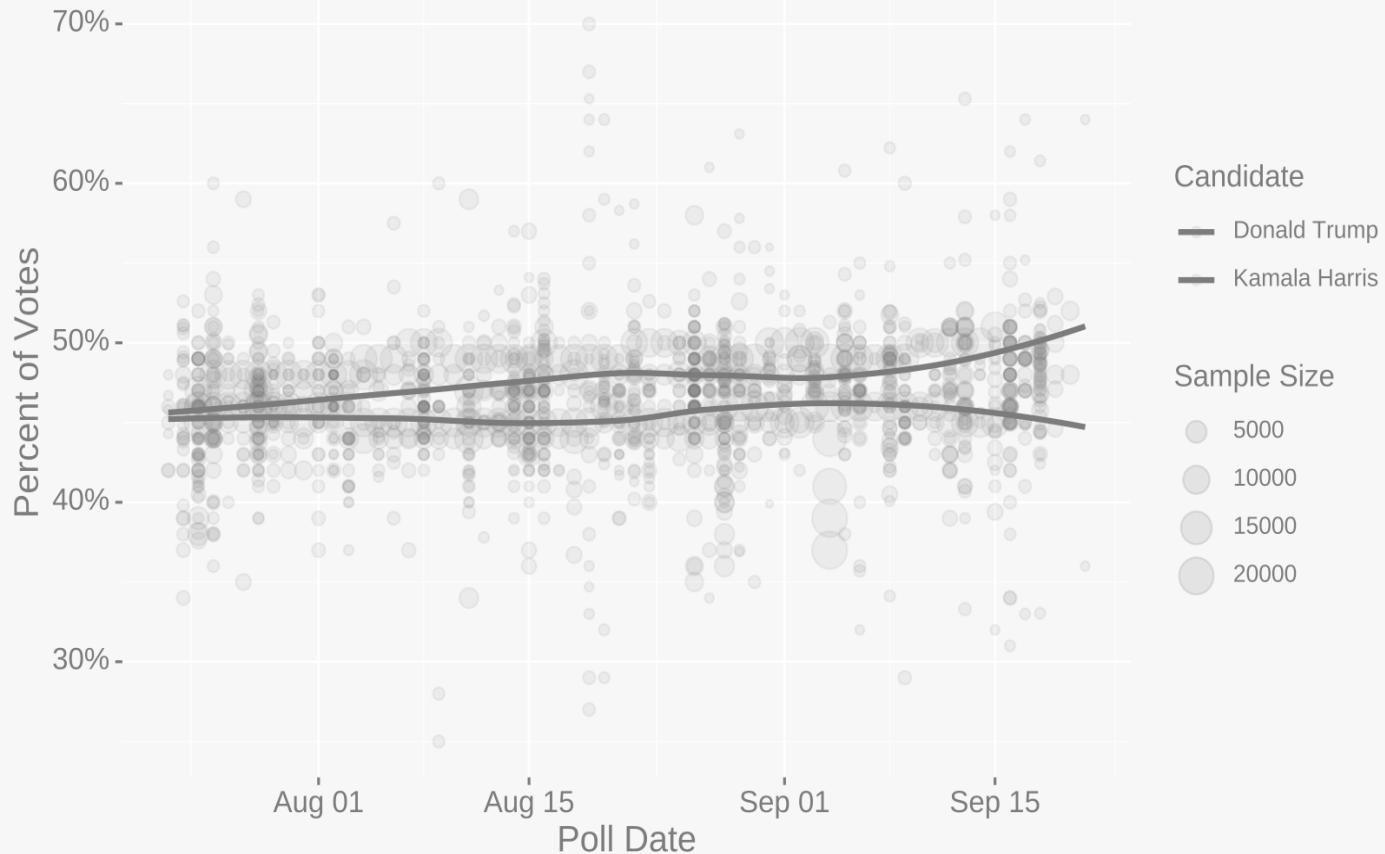
The Anatomy of a Chart

- The Title
- The Key/Legend
- The Axes & Scales
- The Data
- **The Source of the Data**
 - Is there a source provided?
 - Can the source of the data be trusted/verified?
 - If there is no source, can we be sure the data is accurate and truthful? How easy would it be to verify?
 - Has the author been willing to put their name on the visualisation? If not, why not?



Current polling data for 2024 US Presidential Election

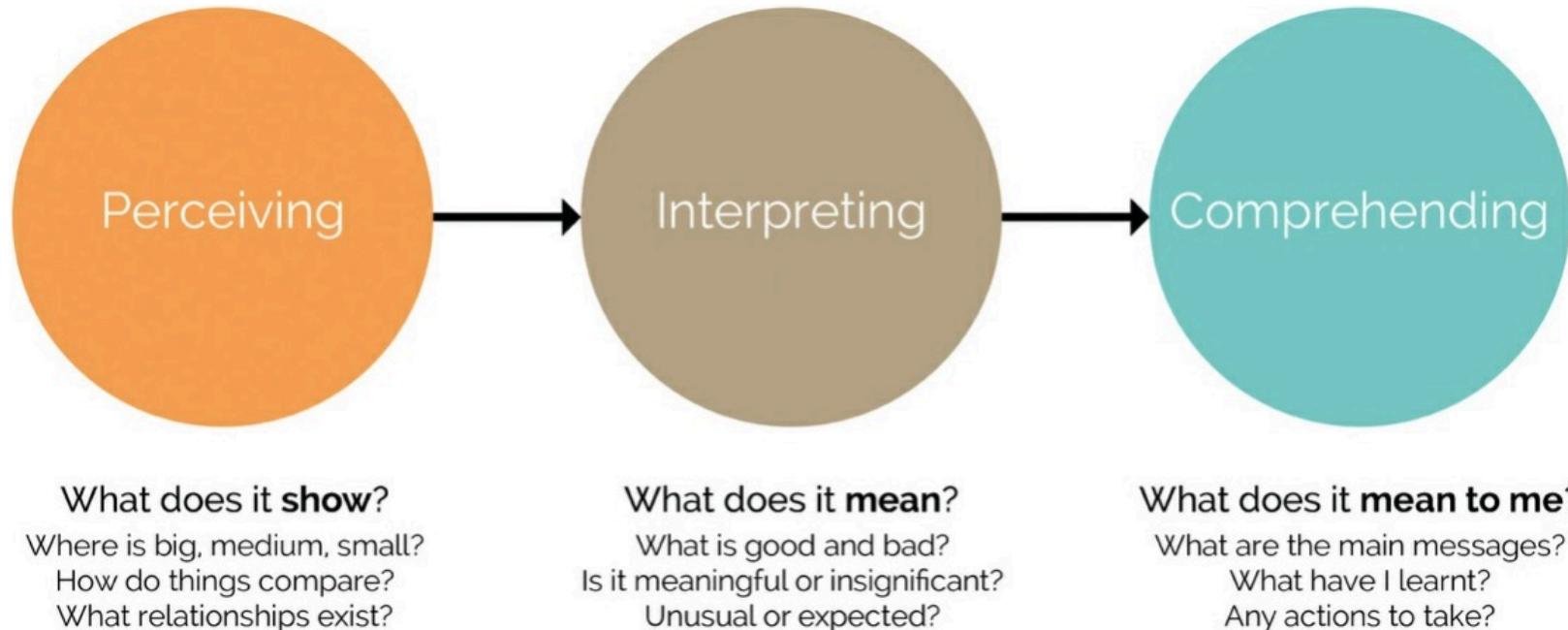
Line indicates average of polls, dots indicate individual poll results



Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

The Anatomy of a Chart

- The Title
- The Key/Legend
- The Axes & Scales
- The Data
- The Source of the Data
- **Information outside the visualisation**
 - Format and prominence
 - Titles, URL, author, sponsor, funder
 - Introductions & guides (how do I use and interpret this visualisation/interactive tool)
 - Layout
 - Data sources and treatment (how has data been transformed)
 - Interactive features



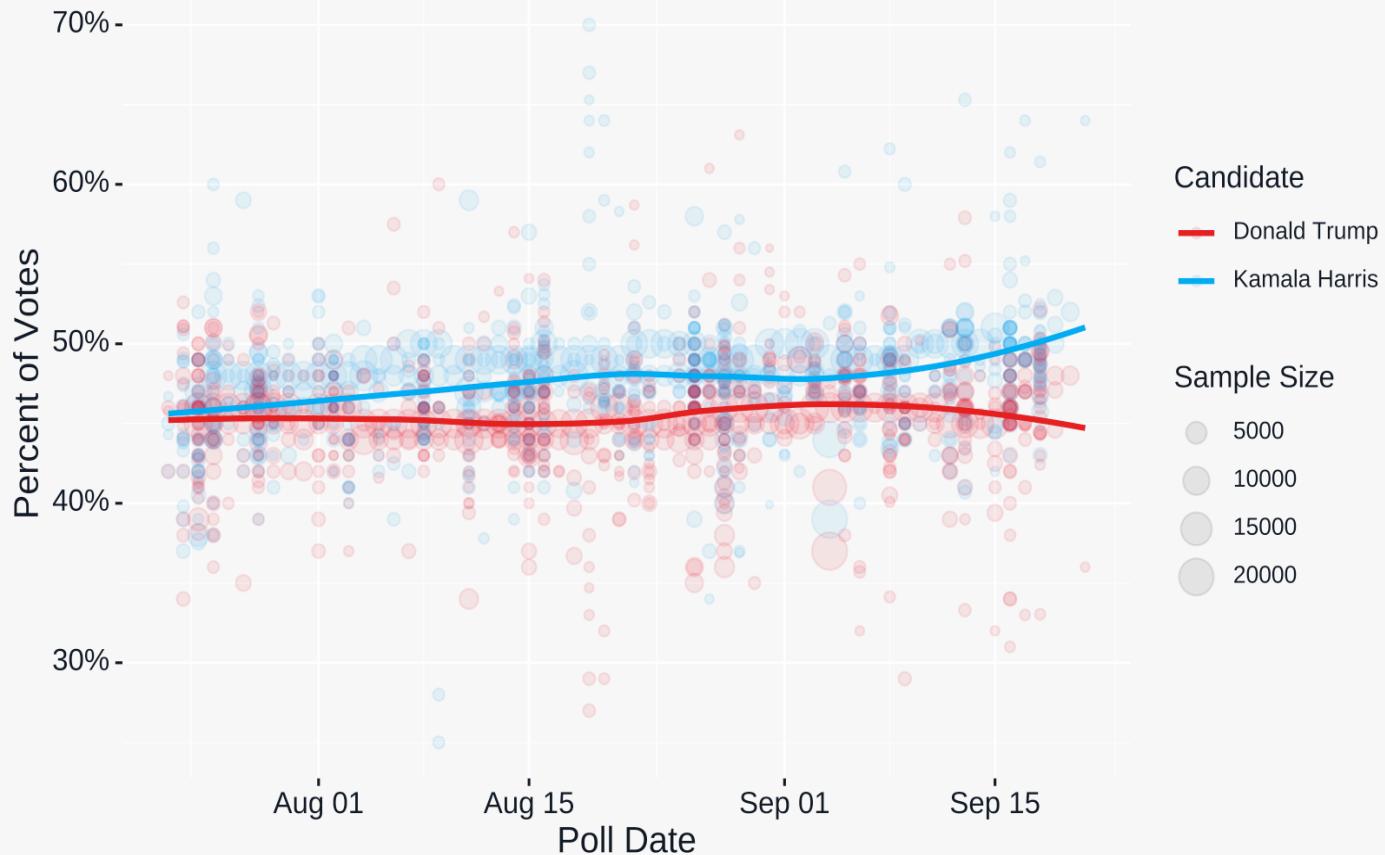
Kirk 2016, fig 1.3, The Three Stages of Understanding, p.22

How can we manipulate these parts of a data visualisation,
using bad practice to construct a specific narrative?



Current polling data for 2024 US Presidential Election

Line indicates average of polls, dots indicate individual poll results



Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

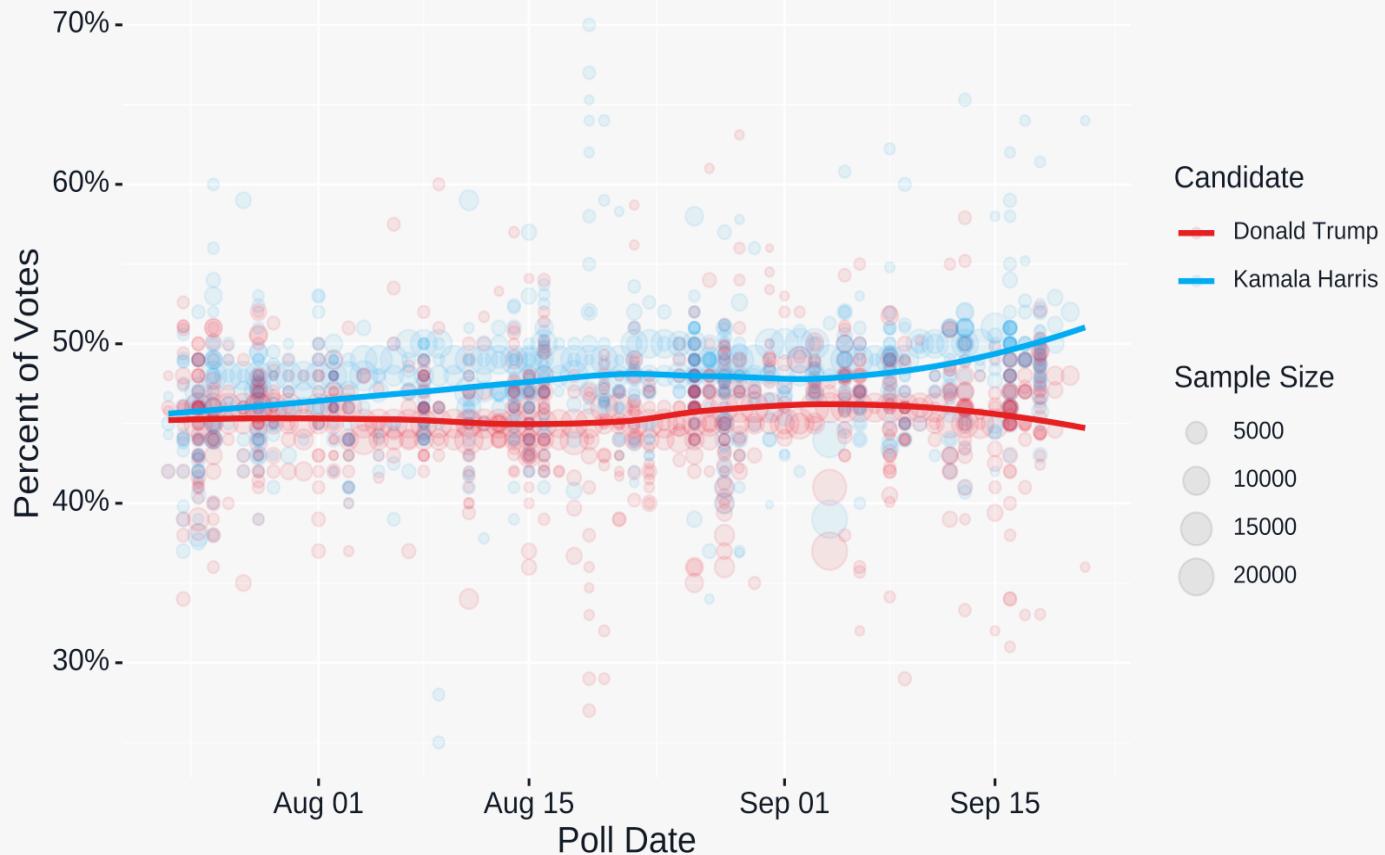
Manipulating our datavis

- You are a **Kamala Harris** Campaign Strategist. You will be manipulating this data visualisation so that it conveys a more **Pro-Harris** message.
- I will be taking the role of a Trump strategist, but also acting as your data visualiser.
- But first
 - Go to wooclap.com (or scan the QR code below)
 - Enter the Event Code: **EYEEGJ**



Current polling data for 2024 US Presidential Election

Line indicates average of polls, dots indicate individual poll results



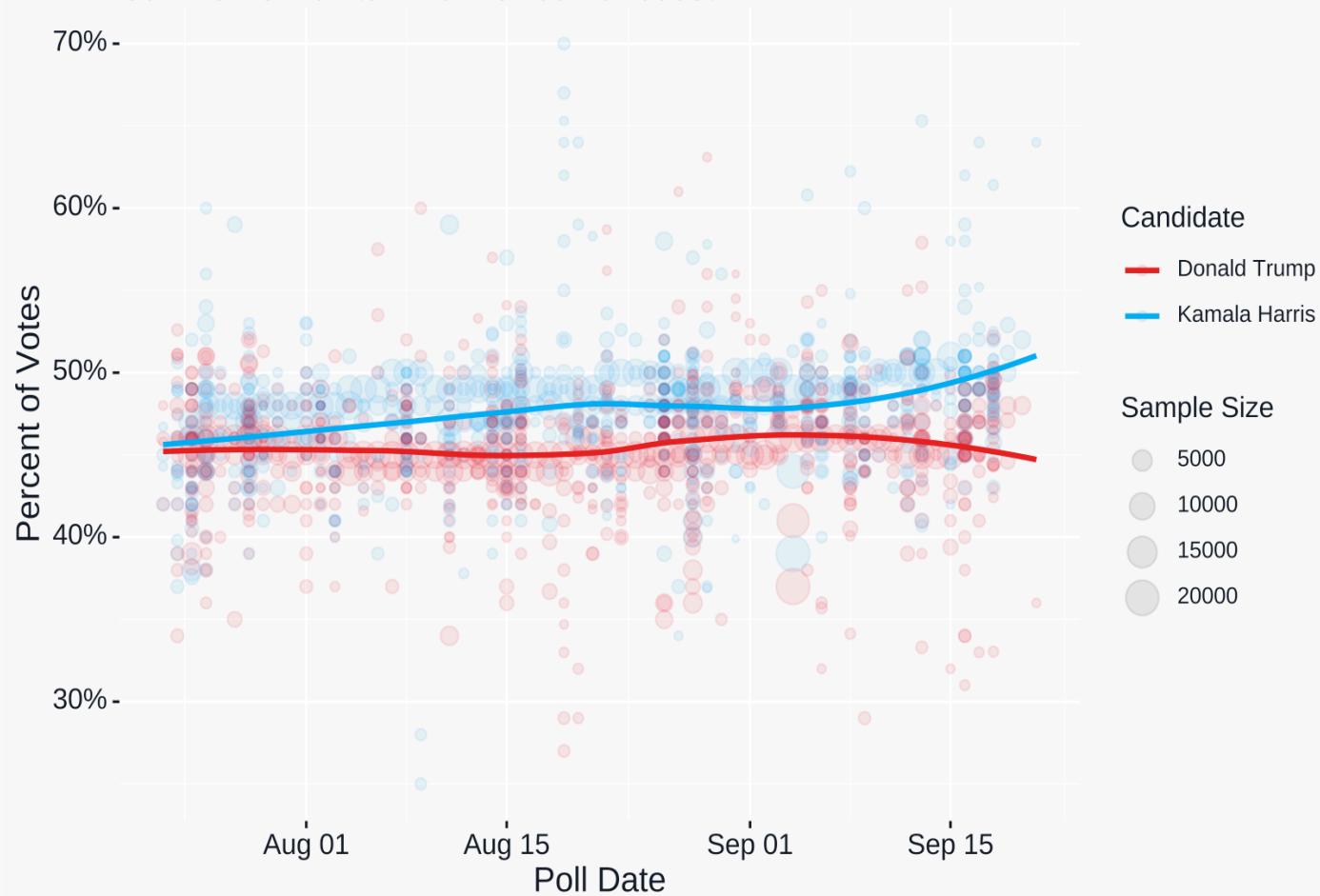
Data from fivethirtyeight.github.com/fivethirtyeight/data (23rd Sept 2024), visualisation by Dr. C Webb, University of Sheffield

Manipulating our datavis

- You are a **Kamala Harris** Campaign Strategist. You will be manipulating this data visualisation so that it conveys a more **Pro-Harris** message.
- I will be taking the role of a Trump strategist, but also acting as your data visualiser.
- But first
 - Go to wooclap.com (or scan the QR code below)
 - Enter the Event Code: **EYEEGJ**



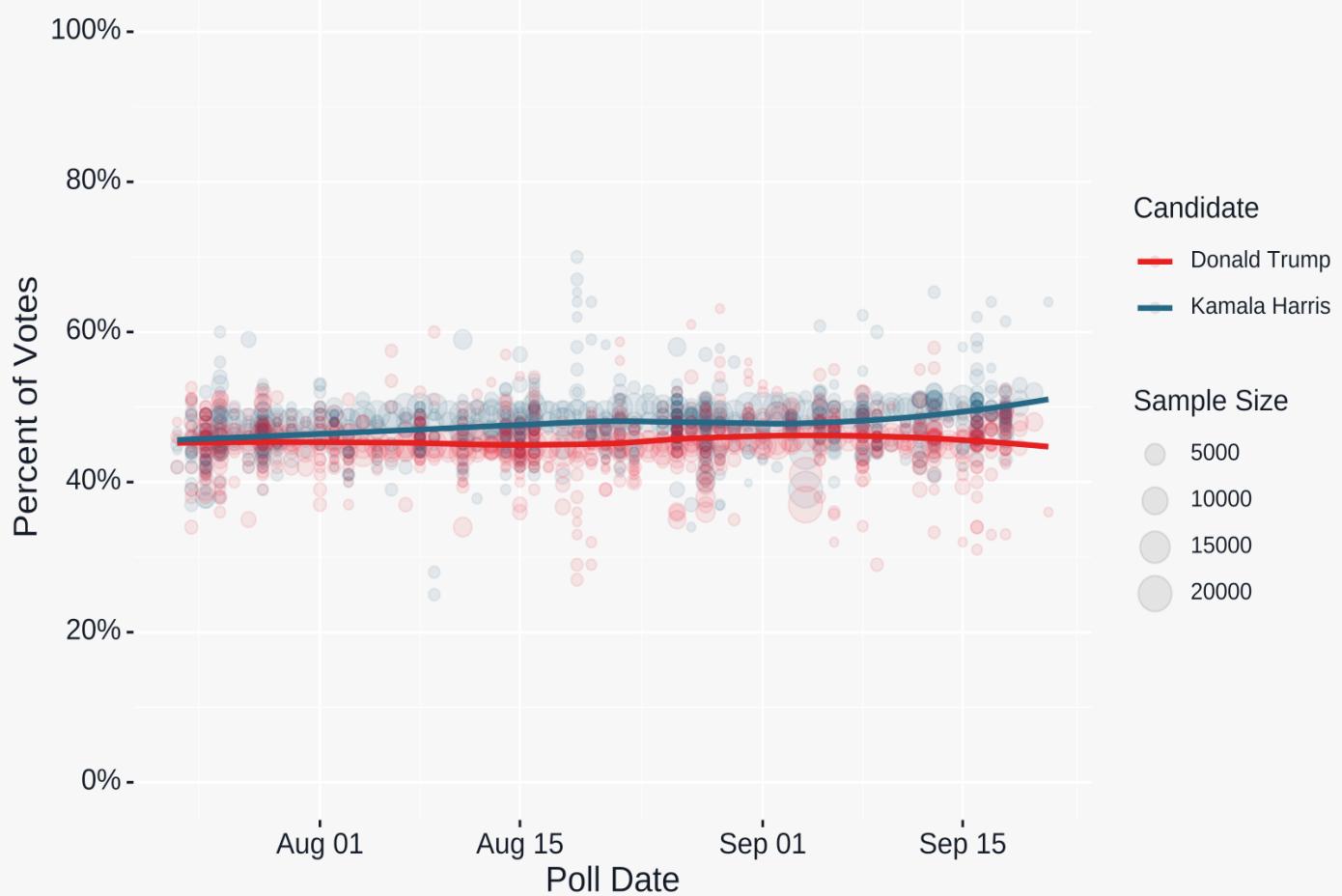
Donald Trump holds onto poll ratings despite surprise Harris entry
 Can Harris maintain her newcomer boost?



Manipulating our datavis

- The Title

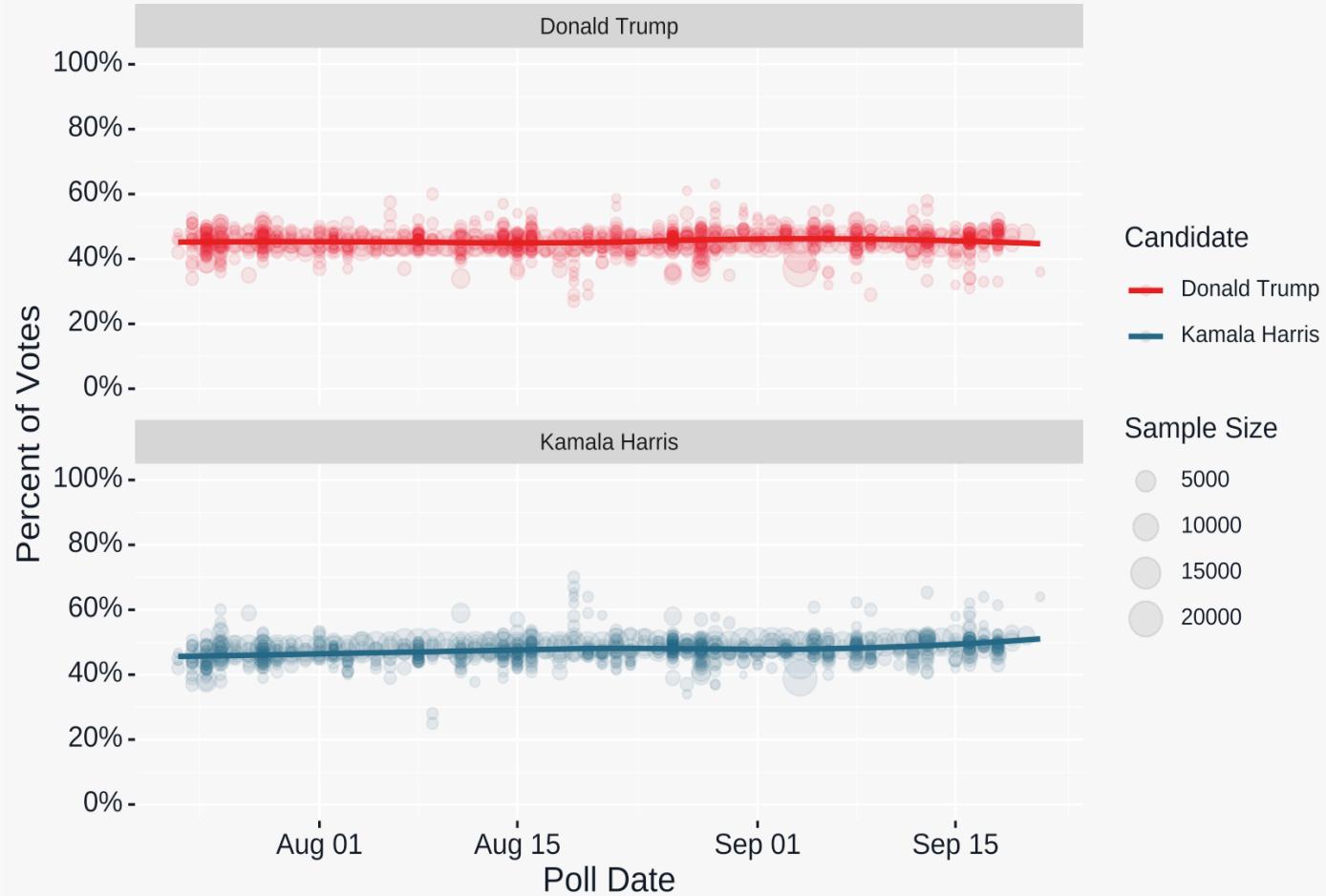
Donald Trump holds onto poll ratings despite surprise Harris entry
Can Harris maintain her newcomer boost?



Manipulating our datavis

- The Title
- The Key/Legend, axes and scales

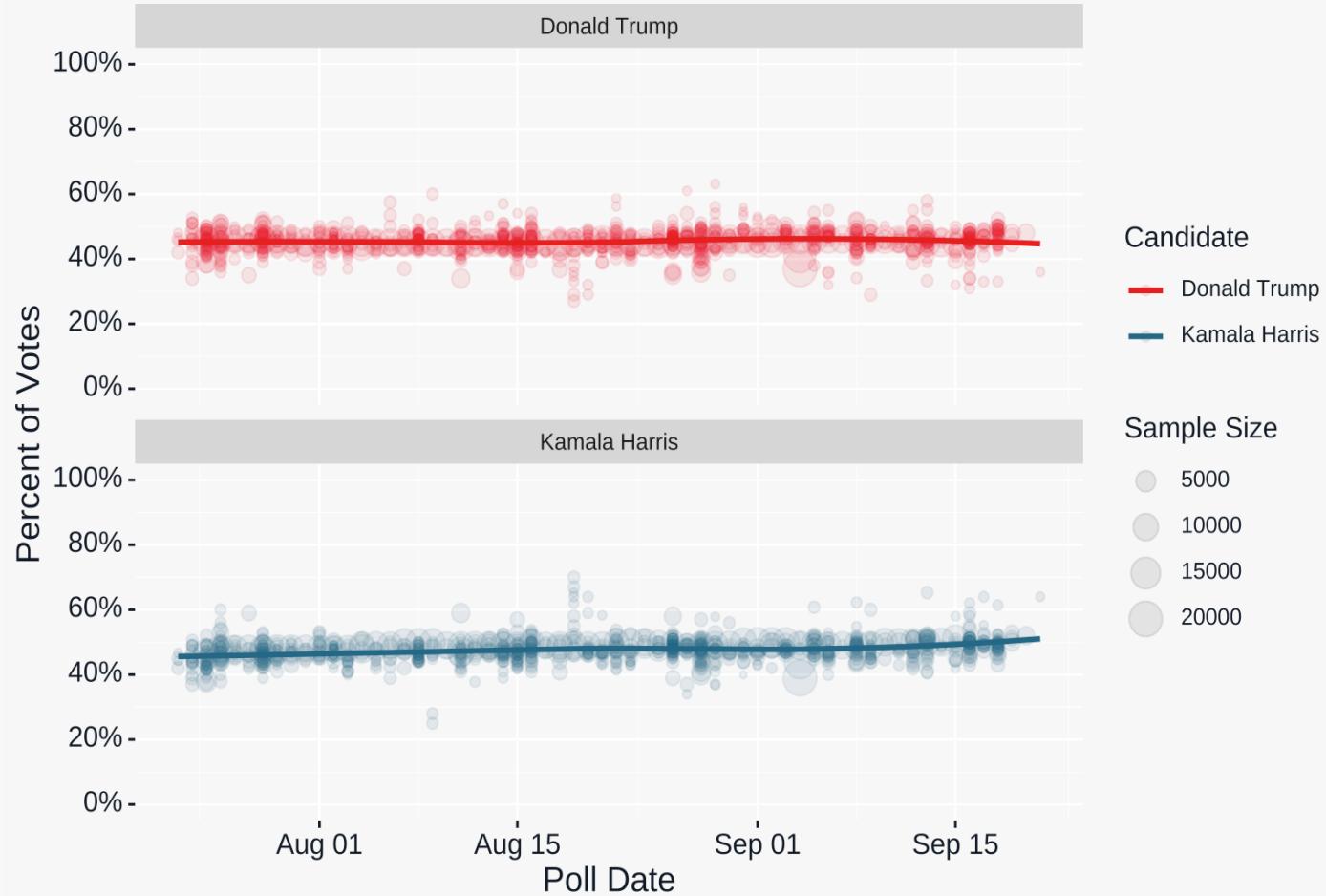
Donald Trump holds onto poll ratings despite surprise Harris entry
 Can Harris maintain her newcomer boost?



Manipulating our datavis

- The Title
- The Key/Legend, axes and scales

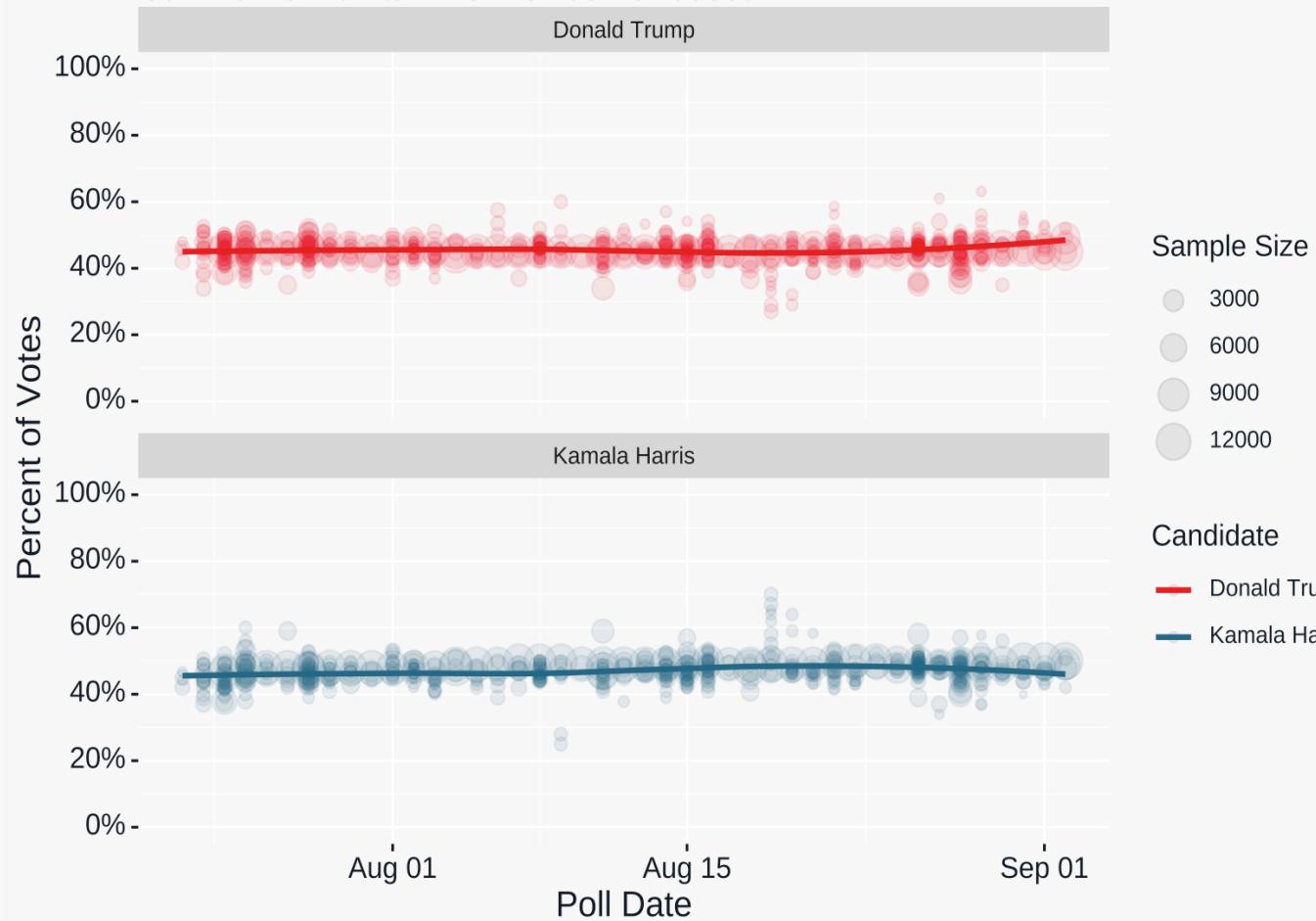
Donald Trump holds onto poll ratings despite surprise Harris entry
 Can Harris maintain her newcomer boost?



Manipulating our datavis

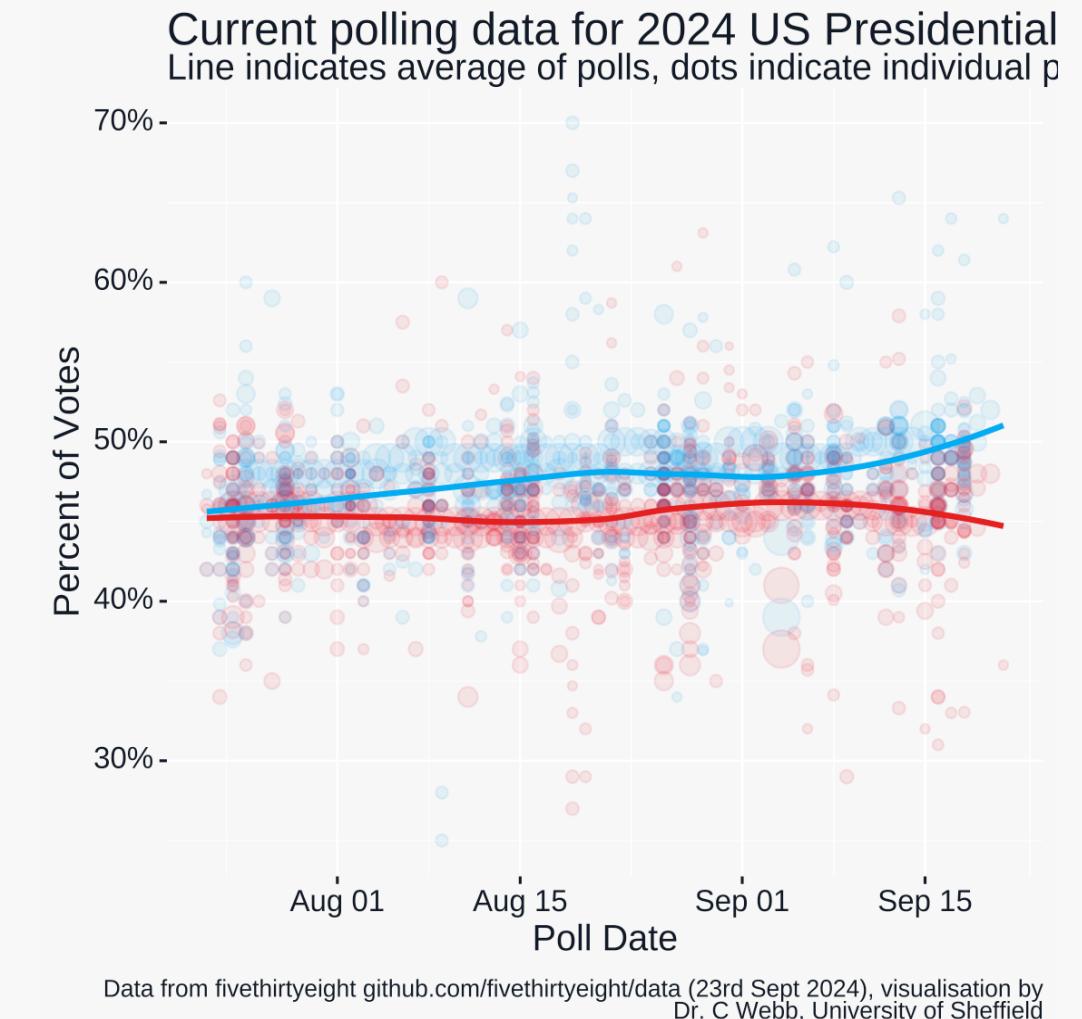
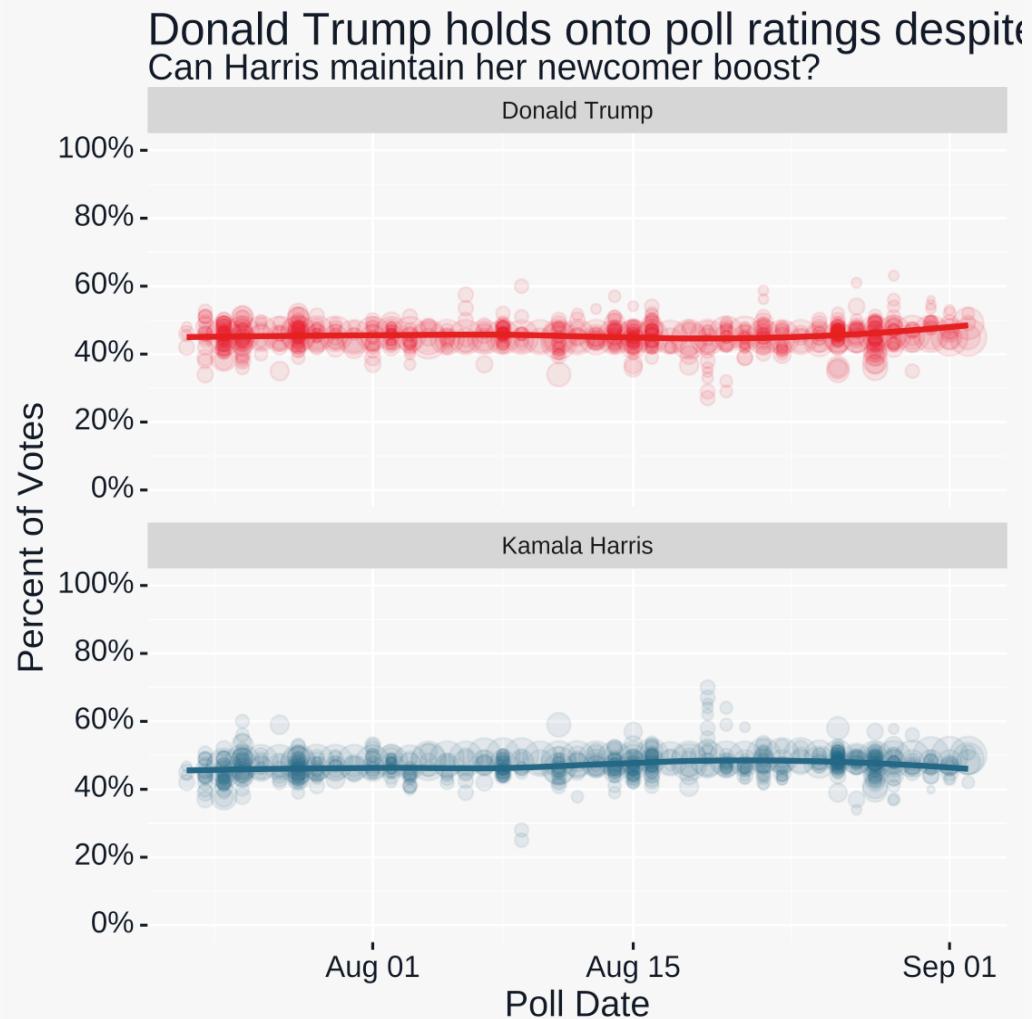
- The Title
- The Key/Legend, axes and scales
- The Data

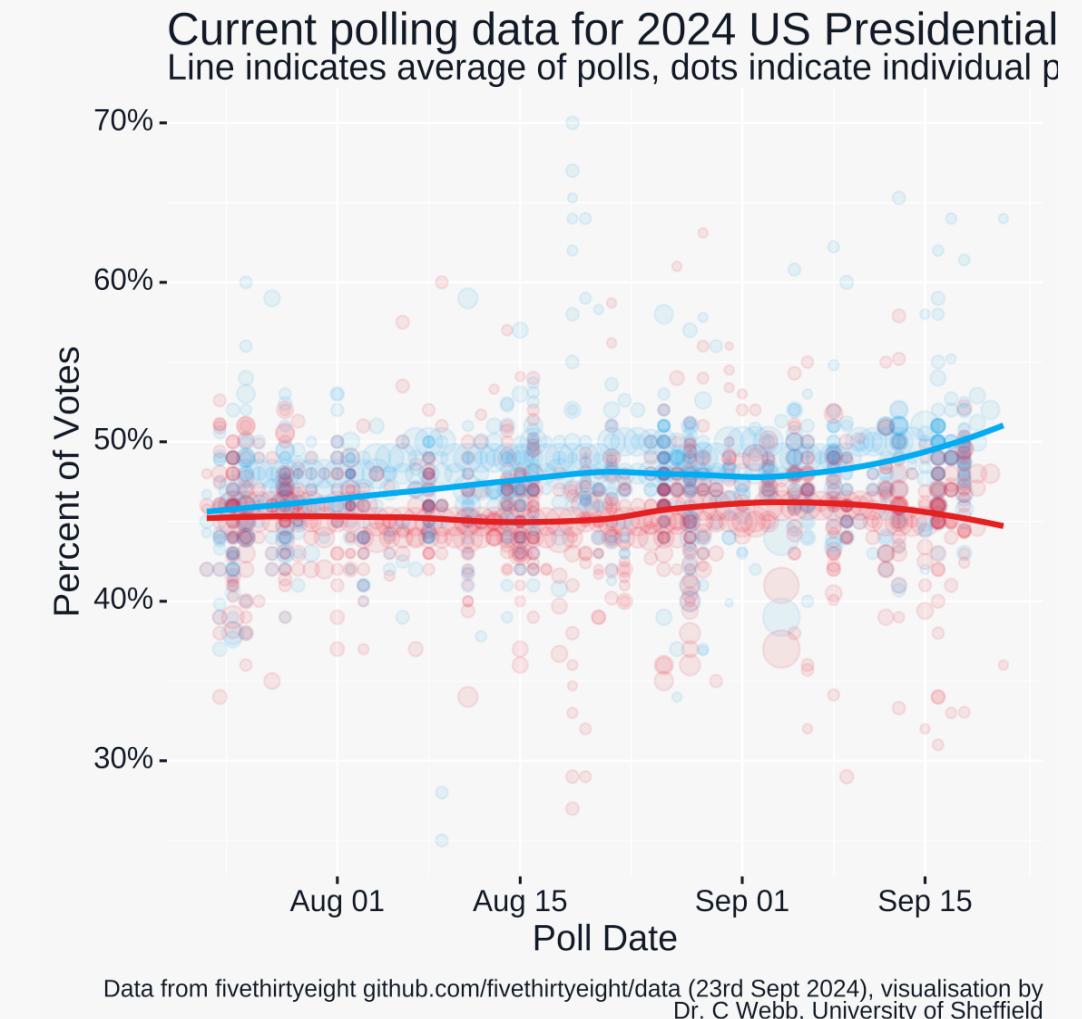
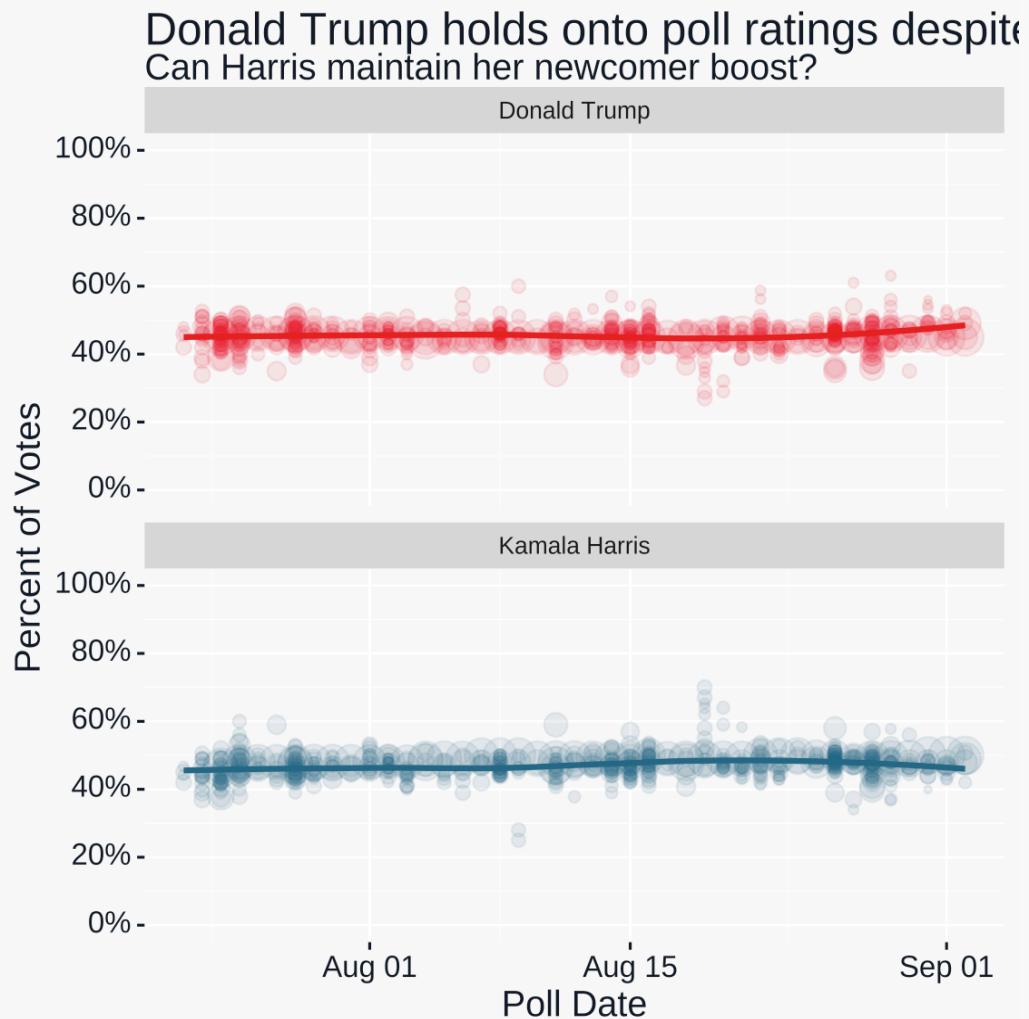
Donald Trump holds onto poll ratings despite surprise Harris entry
Can Harris maintain her newcomer boost?



Manipulating our datavis

- The Title
- The Key/Legend, axes and scales
- The Data



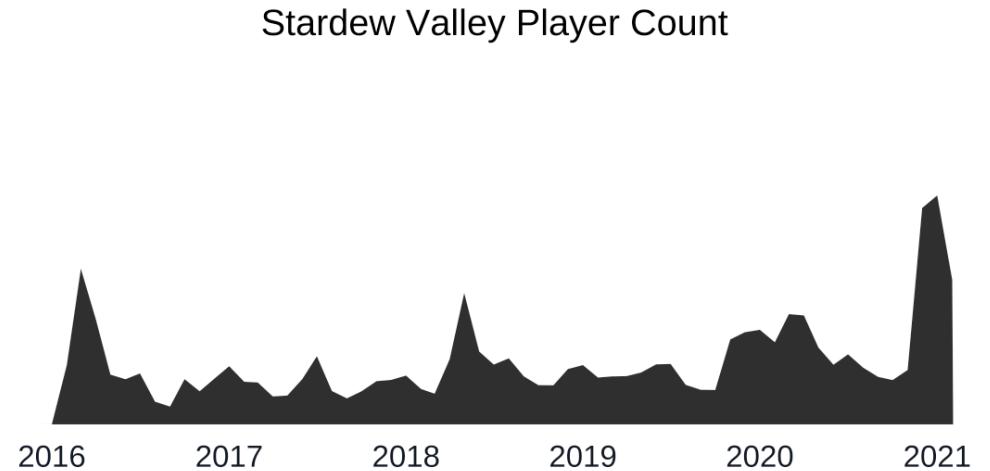


I am showing you this not so you can do it, but so that you can identify when data visualisations have been manipulated to support a specific narrative.



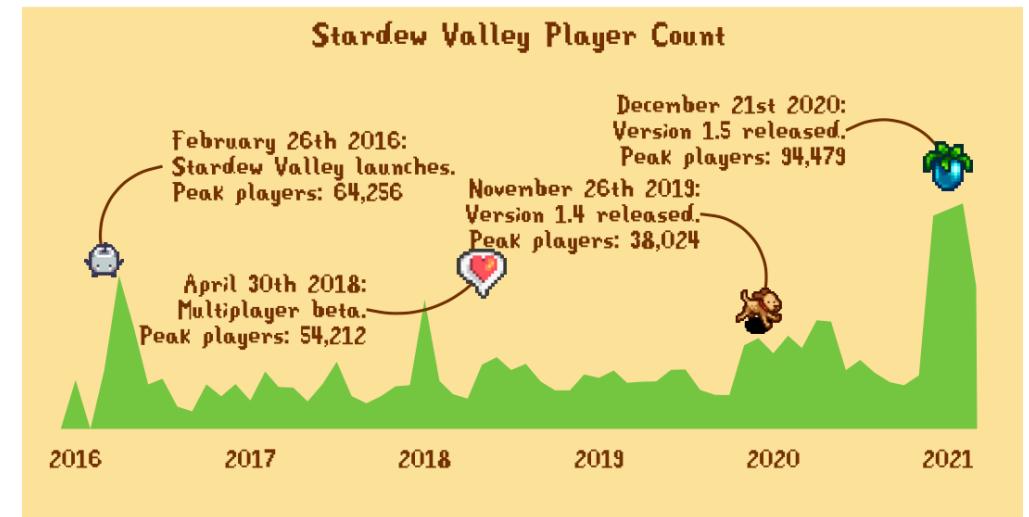
Problematising 'Neutral'

- Most good practice guidance (e.g. Tufte) emphasises the importance of being 'neutral' and 'objective'.
- There is a difference between being 'neutral'/'objective' and being 'misleading'.
- But there is also a difference between being 'neutral'/'objective' and being **boring!**
- You want people to engage with your data visualisation, so you can pick colours, themes, and designs that you find appealing or that communicate the kind of emotions you want to communicate!

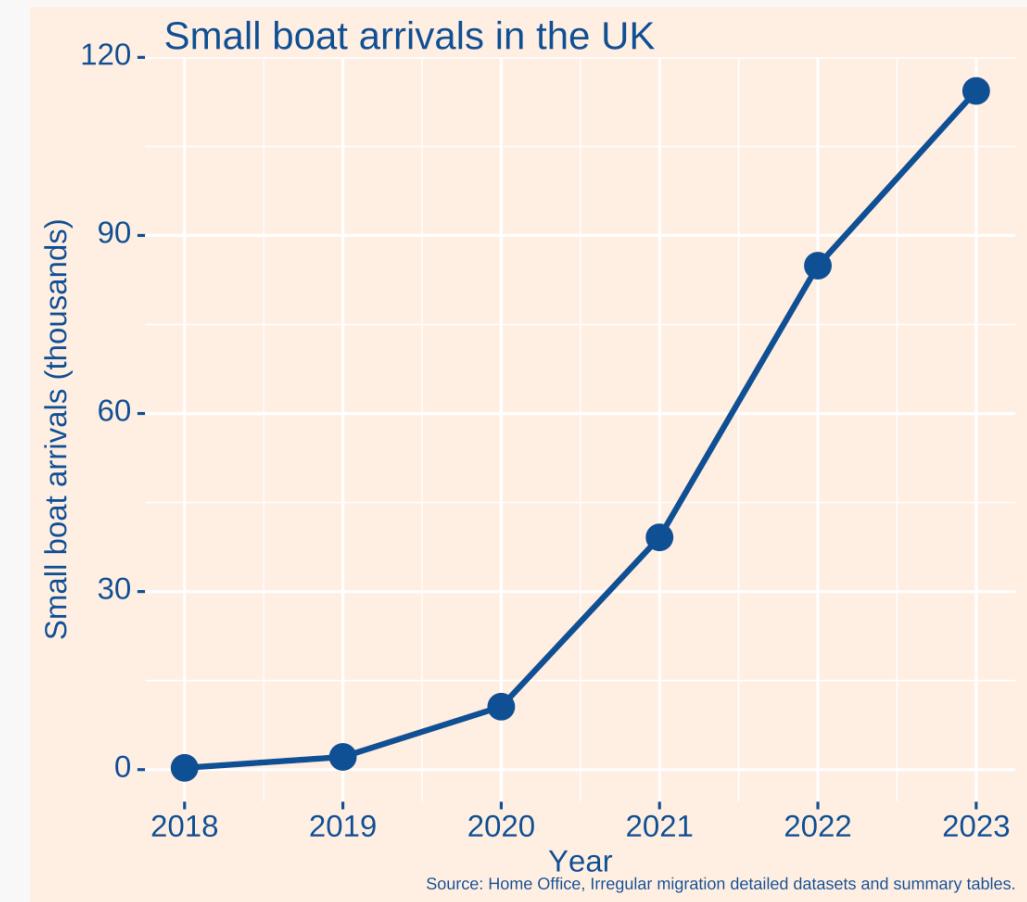
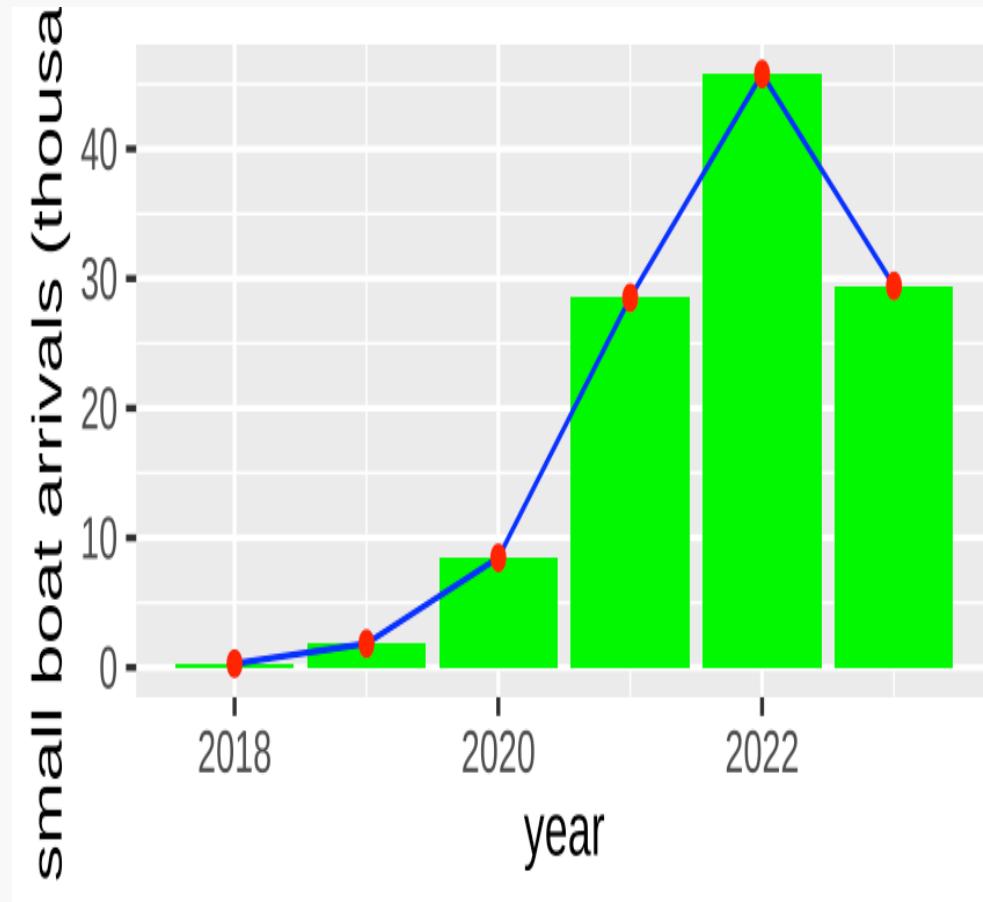


Problematising 'Neutral'

- Most good practice guidance (e.g. Tufte) emphasises the importance of being 'neutral' and 'objective'.
- There is a difference between being 'neutral'/'objective' and being 'misleading'.
- But there is also a difference between being 'neutral'/'objective' and being **boring!**
- You want people to engage with your data visualisation, so you can pick colours, themes, and designs that you find appealing or that communicate the kind of emotions you want to communicate!



Which data visualisation do you trust more?



Problematising 'Neutral'

'Neutral' or 'objective' presentation can also be used to mask the trustworthiness of the data and/or source by leveraging the **optics of scientific data visualisation**.

Problematising 'Neutral'

'Neutral' or 'objective' presentation can also be used to mask the trustworthiness of the data and/or source by leveraging the **optics of scientific data visualisation.**

The data visualisation equivalent of academic jargon:

"The systemic realignment of intersubjective modalities generates a paradigm shift in the ontological discourse, fostering emergent epistemic fluidities across multilayered socio-dynamic frameworks."

Prompt: chatGPT: generate some academic sounding jargon that actually means nothing

Part II

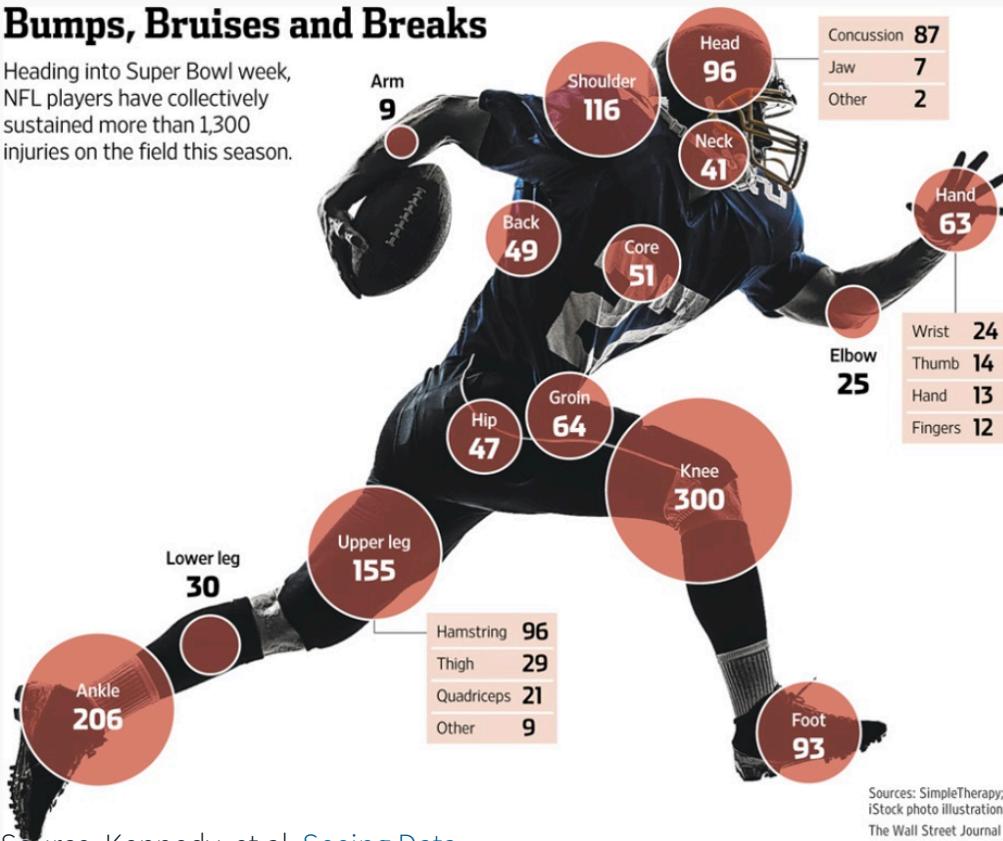
Factors that influence engagement



Factors that influence engagement

Bumps, Bruises and Breaks

Heading into Super Bowl week, NFL players have collectively sustained more than 1,300 injuries on the field this season.



Source: Kennedy, et al. [Seeing Data](#)

- Subject matter & the source

“I’m not really interested in American Football, but the dynamic image of the footballer and the big red circles drew me in by getting me thinking about the impacts of sport on the body.”

Factors that influence engagement

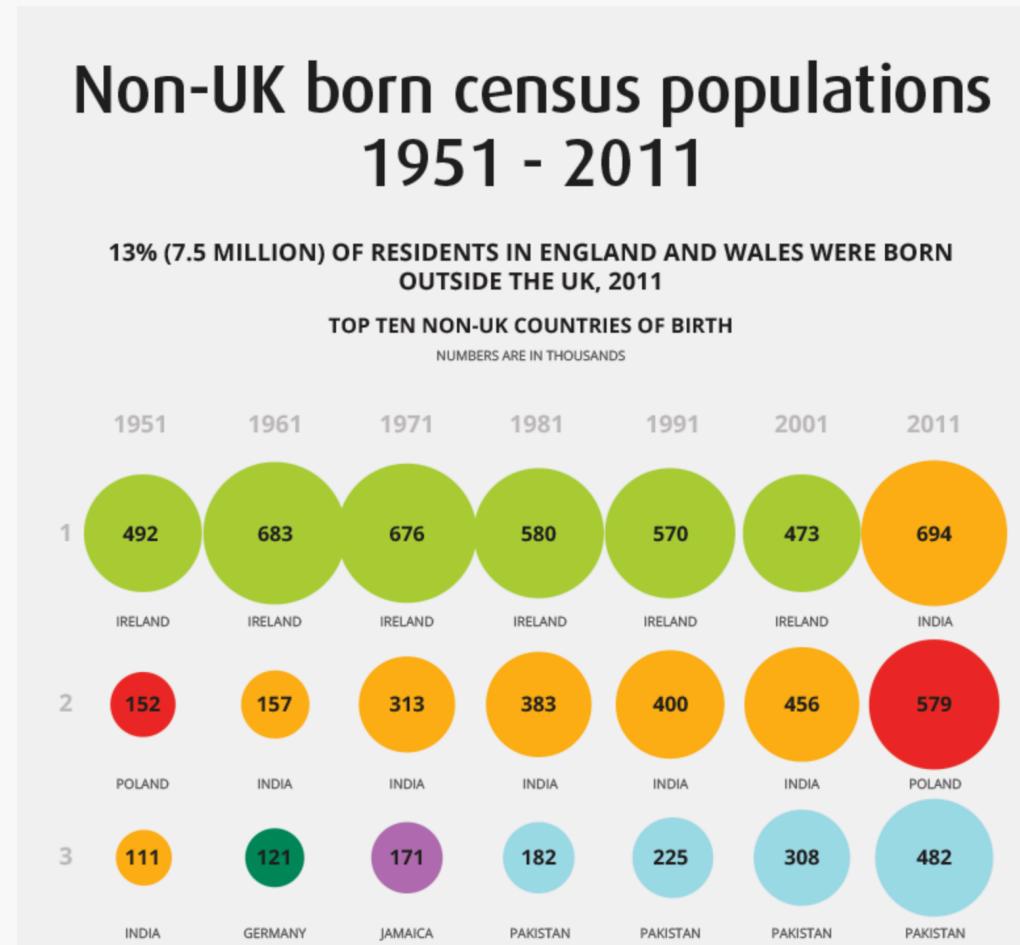


- Subject matter & the source
- **Beliefs and opinions**
 - Confirmation

“I would say it reinforced ... how I feel. It’s not great the way that refugees and migrants are portrayed. But I wouldn’t say [my view] changed, because I had that view before; it just reinforced it I think”

Source: Kennedy, et al. [Seeing Data](#)

Factors that influence engagement



- Subject matter & the source
- **Beliefs and opinions**
 - Confirmation
 - Challenged or changed

“I was surprised that Irish immigrants were the most common in the U.K. ... it was something I hadn't even thought of and it was like, ‘Wow!’. For all the talk of immigration things, the fact that Irish immigrants are the most common, for a lot of the time until ten years ago, it was something I didn't expect”

Factors that influence engagement



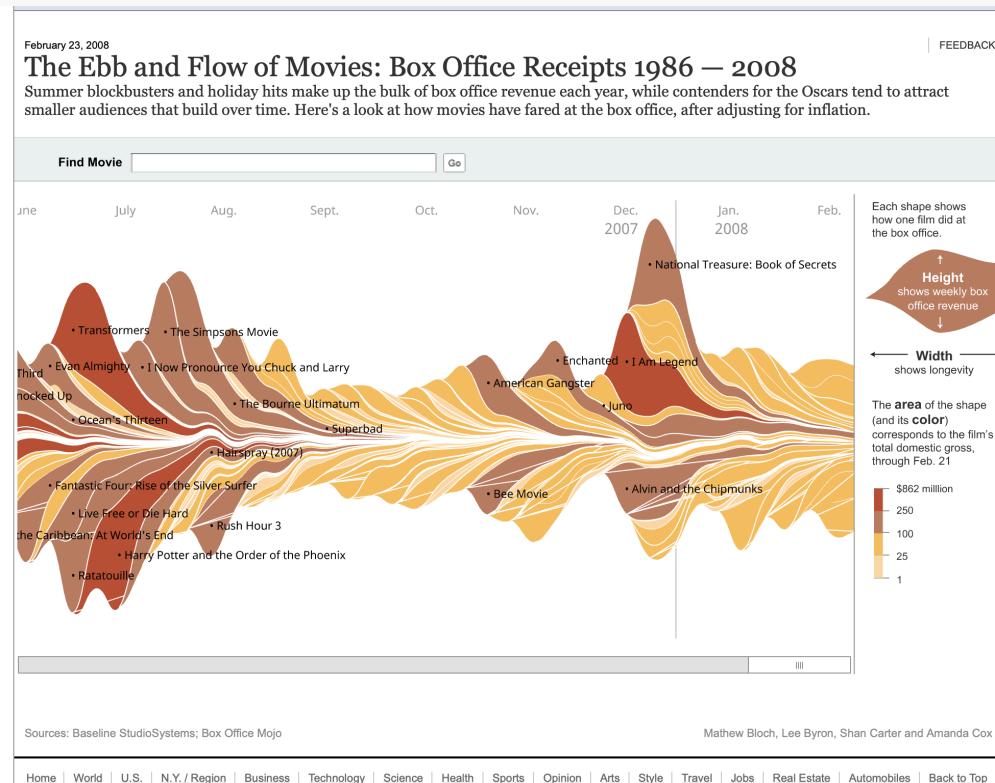
- Subject matter & the source
- Beliefs and opinions
- **Time/cognitive load**

“Because I don’t have a lot of time to like read things and what have you, so if it’s kept simple and easy to read, then I’m more likely to be interested in it and reading it all and, and you know, to look at it, have a good look at it really.”

“I could probably spend an hour on it”

Source: Kennedy, et al. [Seeing Data](#)

Factors that influence engagement



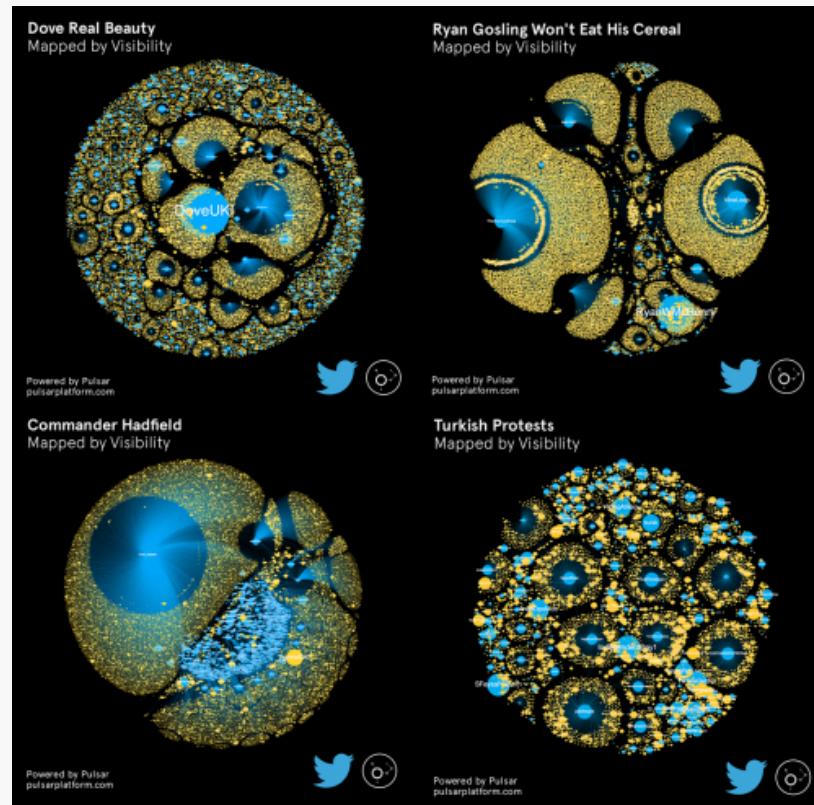
- Subject matter & the source
- Beliefs and opinions
- Time/cognitive load
- **Emotions**

“It was a pleasure to look at this visual presentation because of the co-ordination between the image and the message it carries”

“Frustrated. It was an ugly representation to start with, difficult to see clearly, no information, just a mess”

Source: Kennedy, et al. [Seeing Data](#)

Factors that influence engagement



- Subject matter & the source
- Beliefs and opinions
- Time/cognitive load
- Emotions
- **Confidence** (does it *feel* like it is worth engaging with)

“It was all these circles and colours and I thought that looks like a bit of hard work; don’t know if I understand.”

Source: Kennedy, et al. [Seeing Data](#)

Factors that influence engagement

- Subject matter & the source
- Beliefs and opinions
- Time/cognitive load
- Emotions
- Confidence

Remember what it feels like to be on the receiving end of making sense of other people's data visualisation — to be the audience — before you become an expert in data visualisation!

The rest of this week:

This week's workshop:

Let's extend bar charts. Last week we drew our first bar chart, but we can do better. What if we want to make the categories easier to interpret? What if we want to put the categories in descending order? What if we want to have bar charts that have the average of different categories, rather than just the number of categories that there are?

The rest of this week:

This week's workshop:

Let's extend bar charts. Last week we drew our first bar chart, but we can do better. What if we want to make the categories easier to interpret? What if we want to put the categories in descending order? What if we want to have bar charts that have the average of different categories, rather than just the number of categories that there are?

Before week 3:

Core tasks:

- **Before lecture:** Give yourself two minutes. See how many different ways you can visualise this dataset: 75, 37 (yes, it is a dataset that only has two numbers in it). Use pen and paper to sketch all the different ways you can think of to visually compare these two numbers. Share (on Piazza) the total number of ways you find.
- **Before lecture:** Read the first half of the second chapter, pages 17-42, A Narrative Approach to Data Storytelling, in Feigenbaum, A. and Alamalhodaei, A. (2020) The Data Storytelling Workbook, Routledge
- **Before workshop:** Work through the task at the end of the workshop handout from week 2