Fundamentals of Data Science

Data Visualisation

Southampton Data Science Academy

Storytelling

Southampton Data Science Academy

Storytelling

"The conveying of events in words, sound and/or images, often by improvisation or embellishment."
(Wikipedia)

Southampton Data Science Academy

Storytelling with data visualisation

Recall that the aim of visualisation is to effectively convey some information from data.

You want decision makers to be able to see the **results** of your analysis.

- These should be conveyed in a manner that is suitable for the intended audience to understand.

Southampton Data Science Academy

Considerations for the story

Who are your audience?

- Prior knowledge? Time available?

What is the story?

- What does the data show? What is the **result**?
- Why do you want people to look at your visualisation?
- Common technique in journalism is to “simplify, then exaggerate”.

What do you want people to *do* after seeing your visualisation?

- How do you want their behaviour to change?

Southampton Data Science Academy

Author- or reader-driven?

- The type of story will determine the type of visualisation you need to create.
- An author-driven story has a linear ordering, which you walk the reader through, with little or no interactivity.
- For reader-driven stories, there is no prescribed ordering to the content, allowing the reader to interact and find their own story.

Southampton Data Science Academy

Balancing author- and reader-driven stories

As a compromise between the two approaches, we can merge them both

1. Initial author-driven sequence, that opens up into an interactive narrative for exploration.
2. Author-driven sequence with interaction allowed ‘mid-narrative’ – user can explore a particular point more before moving ahead.
3. General theme presented allowing reader-driven interaction, but they can then choose particular elements to reveal additional author-driven details.

Southampton Data Science Academy

Communicating your message

Highlight and emphasise

- The important information is key to the story, so it should be seen clearly without distractions
- Remove chart junk! Increase data:ink ratio.

Organise

- Using knowledge about gestalt theory and visual perception, information should be structured to guide readers through in a way that will produce the most understanding.

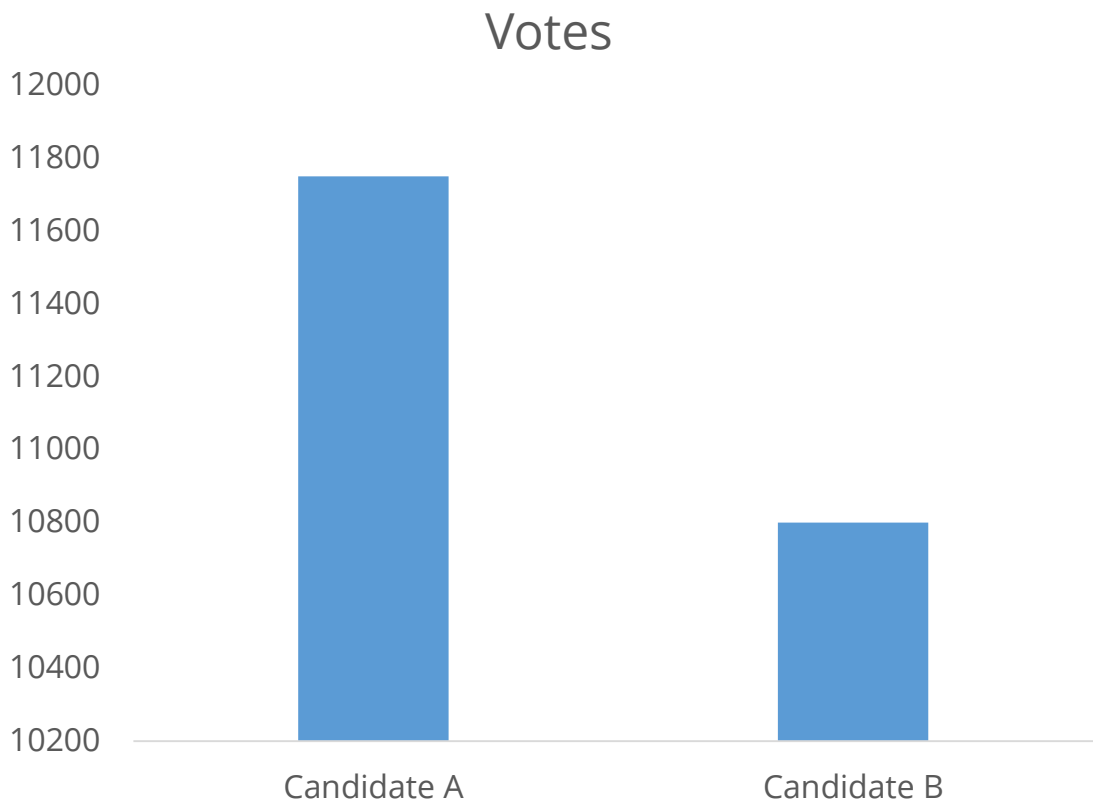
Southampton Data Science Academy

Be objective

- Tufte created a formula to calculate the ‘lie factor’ or how ‘misleading’ a graphic is
- Lie factor = $EffectSizeInGraphic / EffectSizeInData$
  - where EffectSize = $(value2 - value1) / value1$
- a reliable graphic should be between 0.95 and 1.05.

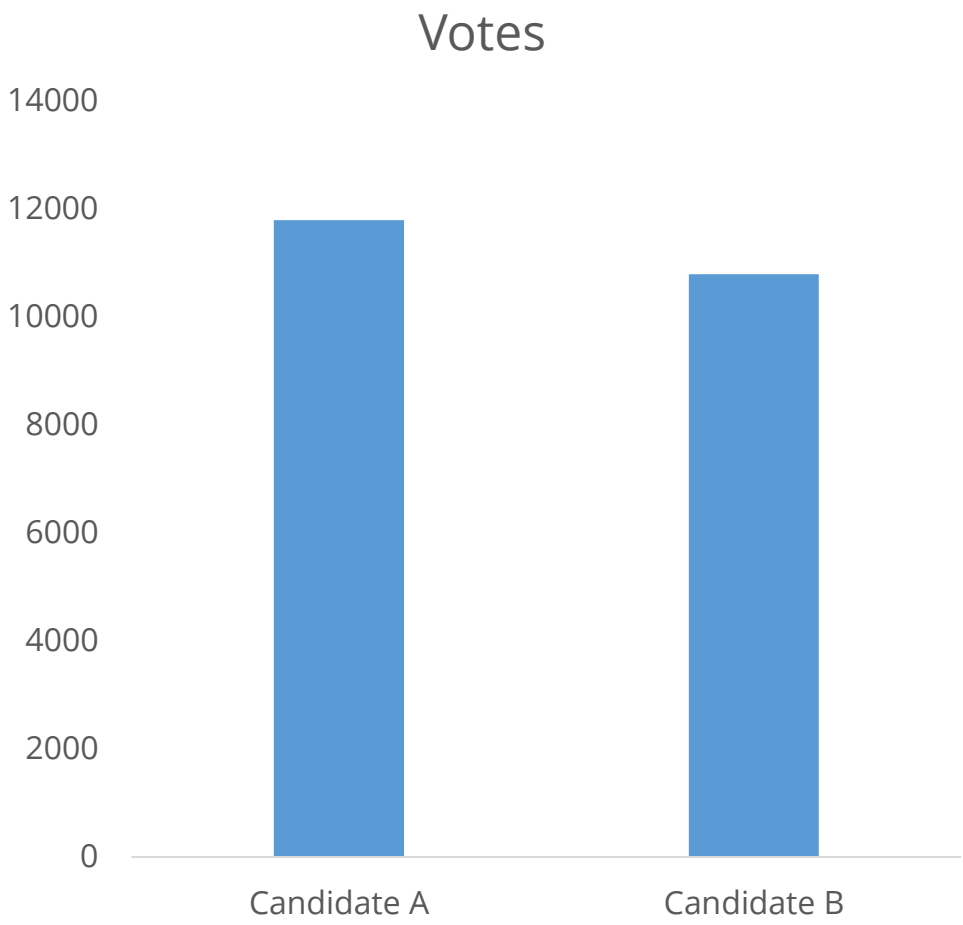
Lie Factor

This suggests that **A** received around three times as many votes as **B**.



Lie Factor

In reality, things are much closer.



Summary

Decision makers need to see the results of your analysis

- And understand the ‘why’ and ‘so what’ as quickly as possible

You need to understand the data and present it in an appropriate way

- Consider theories around visual perception and make it easy for viewers to spot the important message in your image

Southampton Data Science Academy