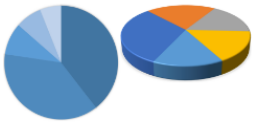
Data stories: Perception, reasoning, & credibility

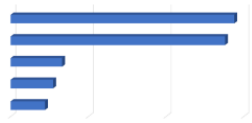
Richard Layton

Session 1, 2022–02–16

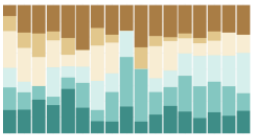
Limitations of common graph types: Four main topics



Effective alternatives to pie charts



Effective alternatives to bar charts



Aligning the design to the story

Advice from experts

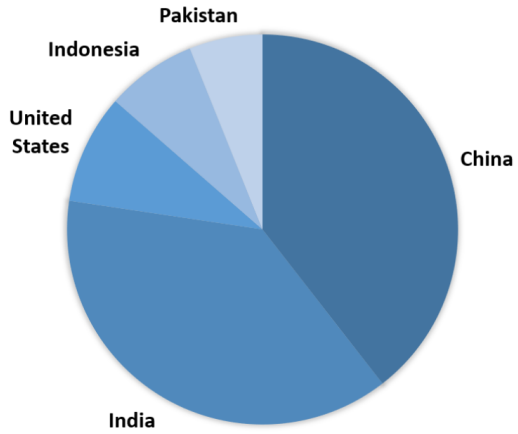
I suggest you have a printed copy of these worksheets to write in during the workshop. We have a number of think-write-share activities that for many people work best when thoughts are written down.

§ *Effective alternatives to pie charts*

Judging pie slices is a low-accuracy task[1]

[1] Data source: World Bank [2022-01]

- Fill in the blanks with your visual estimates of each pie slice.
- Total should be 100%



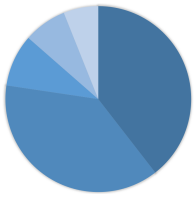
Estimate the percentage of each pie slice (fill in the blanks).

Country	Percentage
China	
India	
United States	
Indonesia	
Pakistan	

The total should be 100%.

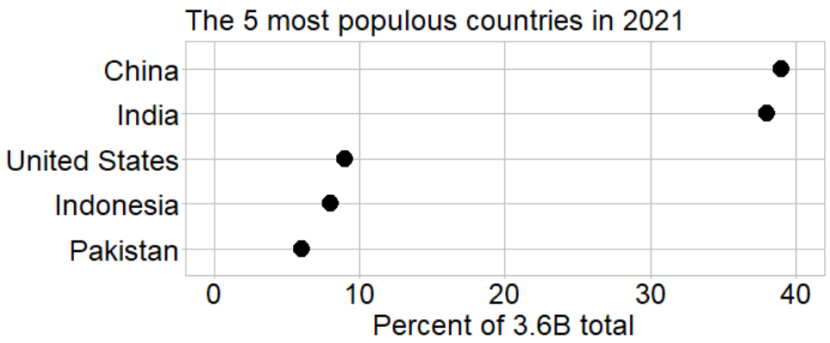
Judging values along a common axis is a high-accuracy task

- The same data is displayed along a common scale.
- Make new visual estimates (fill in the blanks).



The data from the pie chart is shown below as dots along a common scale.

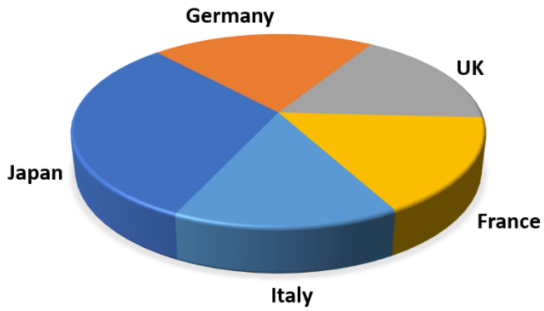
Country	Percentage
China	
India	
United States	
Indonesia	
Pakistan	



3D effects distort our judgment even further[2]

[2] Data source: World Bank [2022-01]

- Fill in the blanks with your visual estimates of each pie slice
- Total should be 100%



Estimate the percentage of each pie slice (fill in the blanks).

Country	Percentage
Japan	
Germany	
UK	
France	
Italy	

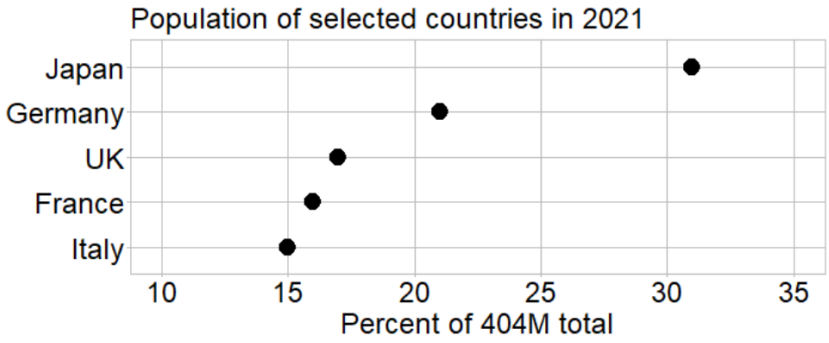
The total should be 100%.

Again, a common scale improves our visual judgments

- The same data is displayed along a common scale.
- Make new visual estimates (fill in the blanks).

The data from the pie chart is shown below as dots along a common scale.

Country	Percentage
Japan	
Germany	
UK	
France	
Italy	



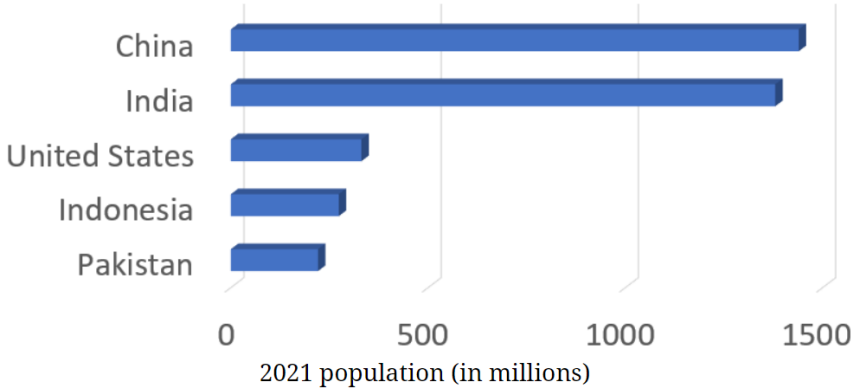
§ *Effective alternatives to bar charts*

3D effects always distort our judgment[3]

[3] Data source: World Bank [2022-01]

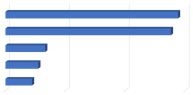
- Fill in the blanks with your visual estimates of each bar length

Country	Millions
China	
India	
United States	
Indonesia	
Pakistan	



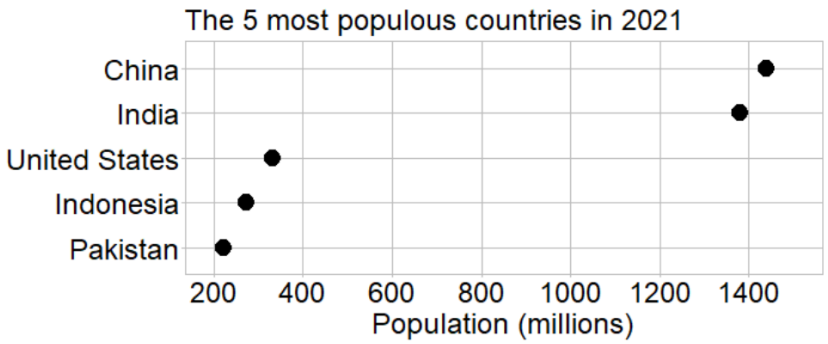
Same data—without 3D effects—along a common scale

- The same data is displayed along a common scale
- Make new visual estimates (fill in the blanks).



The data from the 3D bar chart is shown below as dots along a common scale.

Country	Millions
China	
India	
United States	
Indonesia	
Pakistan	

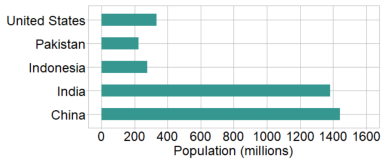


With a zero baseline and no 3D effects, bars are OK

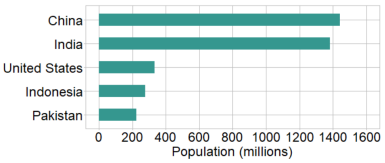
- Bar charts must have a zero baseline to avoid deception.
- Ordering rows by the data values facilitates visual comparisons.
- The only information in the bar is the position of its end point. The bar itself is superfluous.
- Dot charts allow direct visual comparison of quantities.
- Dot charts are effective replacements for pie charts and bar charts.

Notes

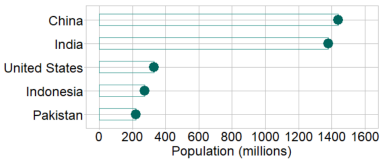
Default bar chart:



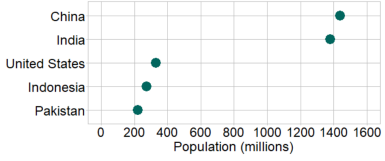
Ordered by magnitude:



Omitting the fill color:



Produces a dot chart:

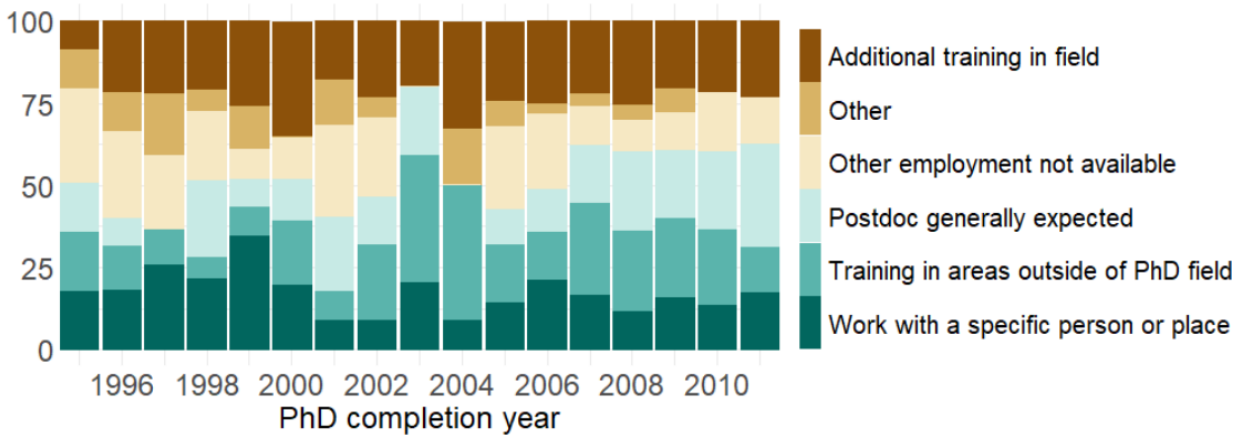


§ *Aligning the design to the story*

Survey: "What was your reason for taking this postdoc?"[4]

[4] Data adapted from Main et al. [2021]

Before we can talk about what the chart *says*, we have to agree on what it *shows*.



Write your responses below.

- What does a color represent?

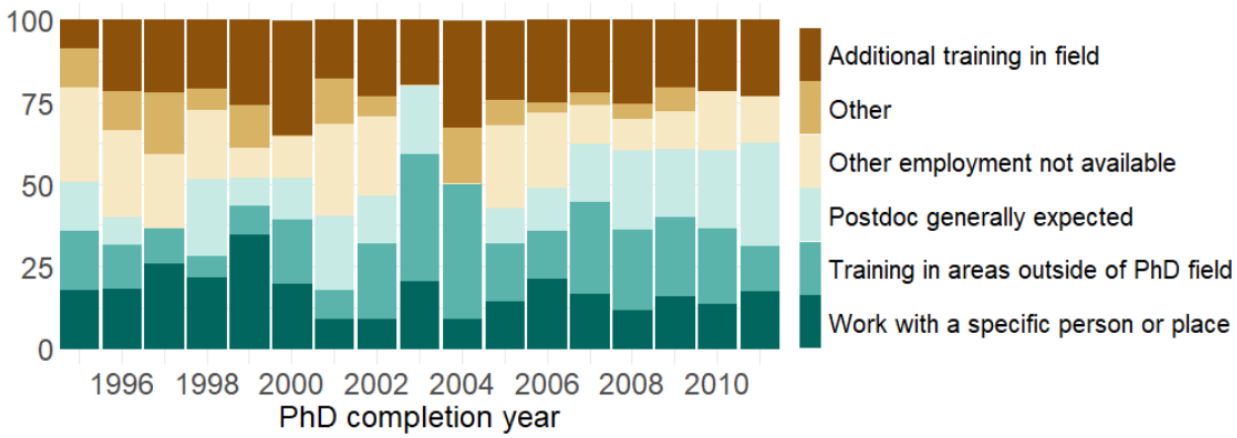
- What does a single color-segment of a bar represent?

- What does the changing height over time of a segment represent?

What ideas are conveyed by the chart?[5]

[5] Data adapted from Main et al. [2021]

We agree on what the chart *shows*; now we can consider what it *says*.



Write your responses below.

- Describe a trend for one of the six reasons for obtaining postdoc training.

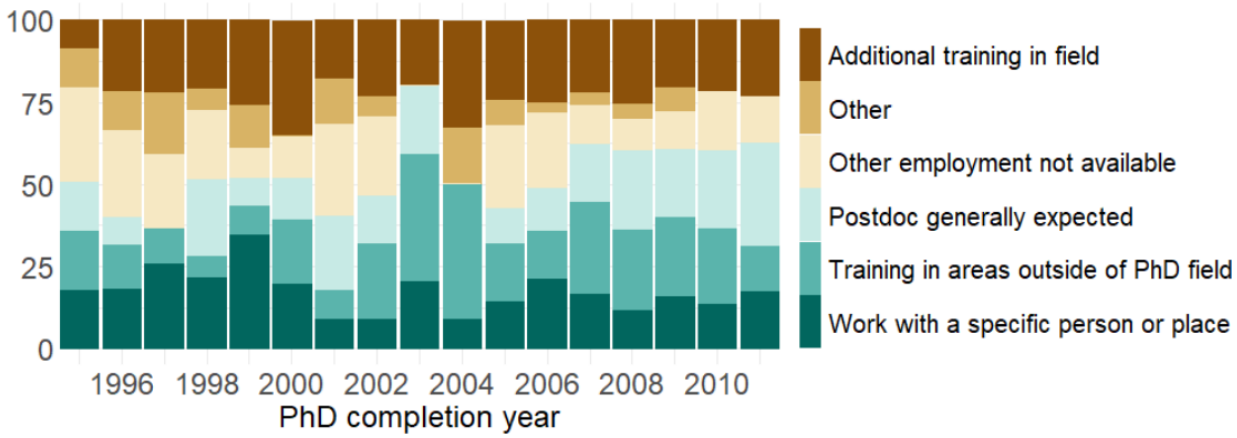
- Compare two of the reasons over time.

- Describe the main idea this chart conveys to you.

What can we say about the variables?[6]

[6] Data adapted from Main et al. [2021]

Choosing an effective chart design depends in part on what variables you have.



FILL IN THE BLANKS to begin summarizing the data structure.

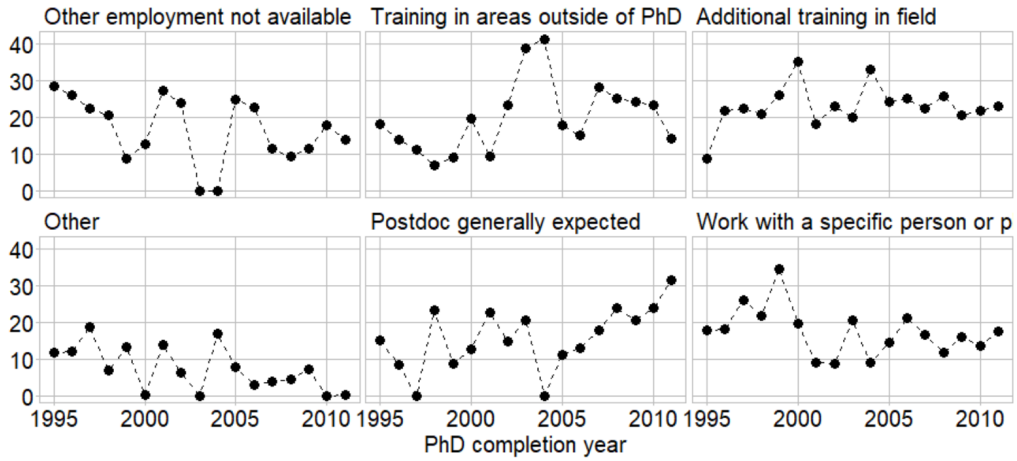
1. Time (discrete years) is one categorical variable.
2. The other categorical variable is _____ .
3. The quantitative variable is _____ .
4. Which is the independent variable? _____ .

Note that discrete time units are not 'continuous', so the time units here are a categorical (not quantitative) variable.

The appropriate design for a time series is a line graph[7]

[7] Data adapted from Main et al. [2021]

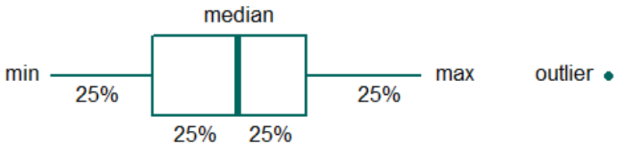
Separating the reasons into individual panels clarifies the data



- Describe the main idea this chart conveys to you.

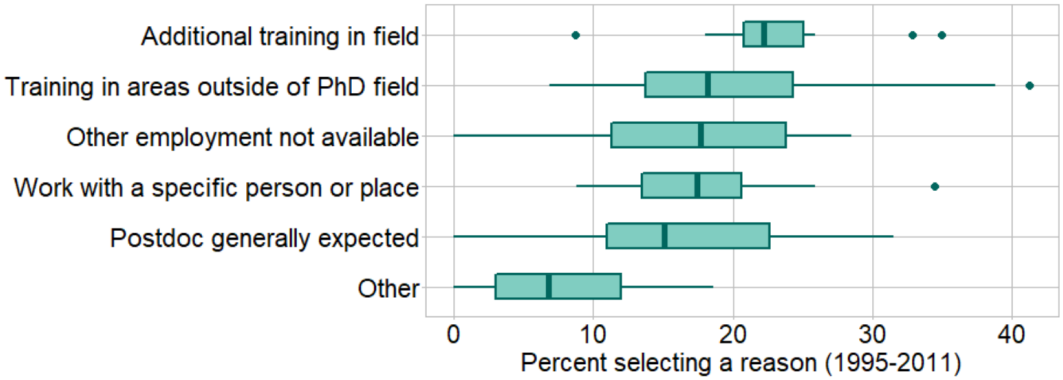
Conventions of the box-and-whisker plot

Designed to show a summary of the distribution of a single quantitative variable.



Our final design shows distributions of annual percentages[8]

[8] Data adapted from Main et al. [2021]



- Describe the main idea(s) this chart conveys to you.

§ *Advice from experts*

Match the expert to the advice.[9,10,11,12]

[9] Cairo [2019]
[10] Doumont [2009]
[11] Evergreen [2017]
[12] Tufte [1983]

FILL IN THE BLANKS with letters A–D.

Expert	Letter	Emphasizes the importance of
A. Alberto Cairo	_____	message
B. Jean-luc Doumont	_____	variables
C. Stephanie Evergreen	_____	revealing the complex
D. Edward Tufte	_____	knowing your main point
	_____	not lying to yourself

Ideas to consider

- Characterize the data structure and content
- Explore a story's context, causality, and complexity
- Align visual and verbal logic by revising iteratively
- Edit to suit the rhetorical goals for each audience
- Control every pixel—avoid thoughtless conformity
- Question are you seeing only what you want to believe?

References

Alberto Cairo. *How Charts Lie*. W.W. Norton, New York, 2019.

Jean-luc Doumont. *Trees, Maps, and Theorems*. Principiae, Belgium, 2009.

Stephanie D. H. Evergreen. *Effective Data Visualization*. Sage, Thousand Oaks, CA, 2017.

Joyce B. Main, Yanbing Wang, and Li Tan. The career outlook of engineering PhDs. *Journal of Engineering Education*, 110(4):977–1002, 2021. URL https://doi.org/10.1002/jee.20416.

Edward Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.

World Bank. Population total for United States, 2022-01. URL https://fred.stlouisfed.org/series/POPTOTUSA647NWDB.