The Treachery of Images

Exploring the Interdependence Between Graphics, Statistics, and Interaction

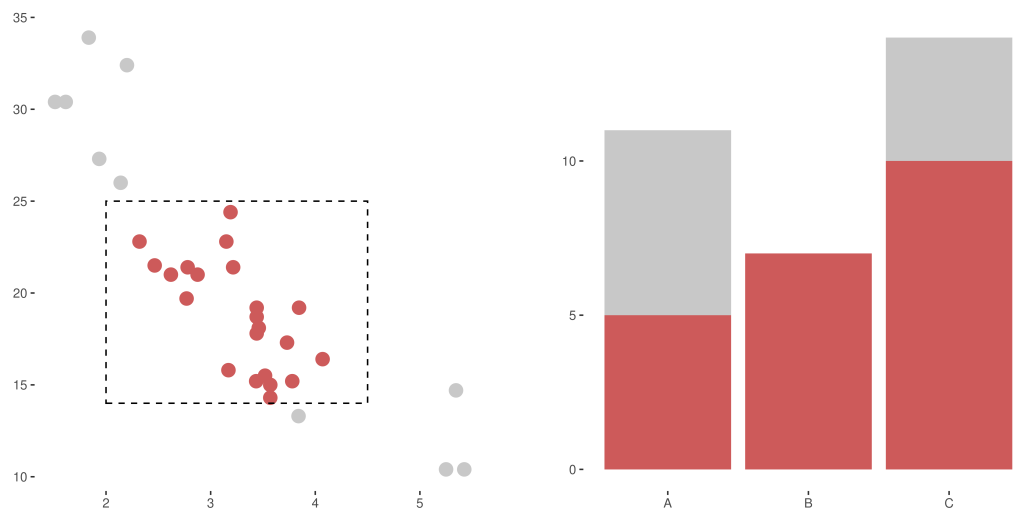
Adam Bartonicek
Dr. Simon Urbanek (primary supervisor)
Dr. Paul Murrell (co-supervisor)

The University of Auckland

Exploring data interactively

https://github.com/bartonicek/plotscape/tree/master/packages/plotscaper

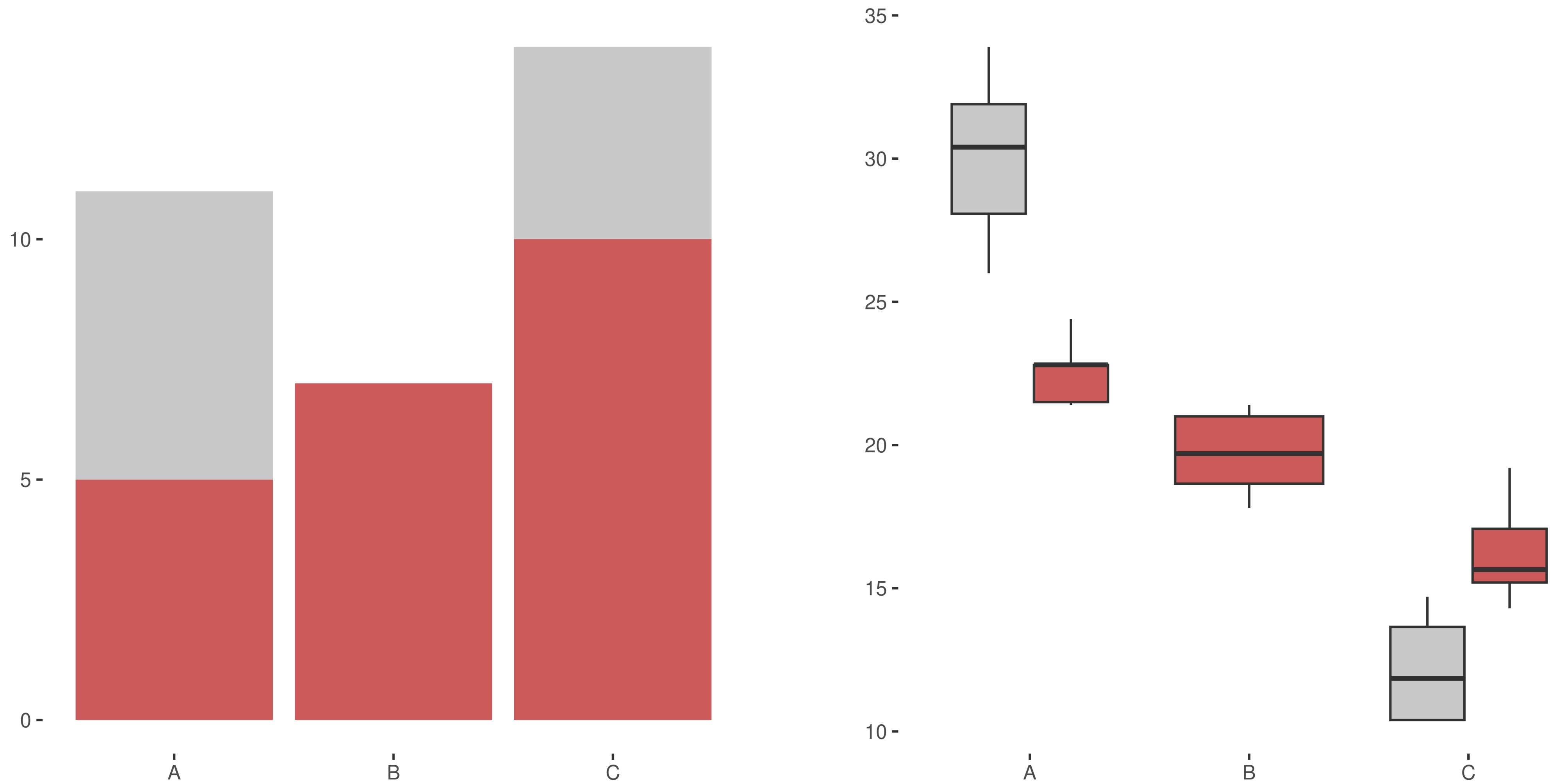
Linked selection



- Aka “brushing” or “highlighting”
- Select objects and highlight the corresponding cases
- One of the most useful interactive features (Buja, Cook, and Swayne 1996; Heer and Shneiderman 2012; Ward, Grinstein, and Keim 2015; Ware 2019)

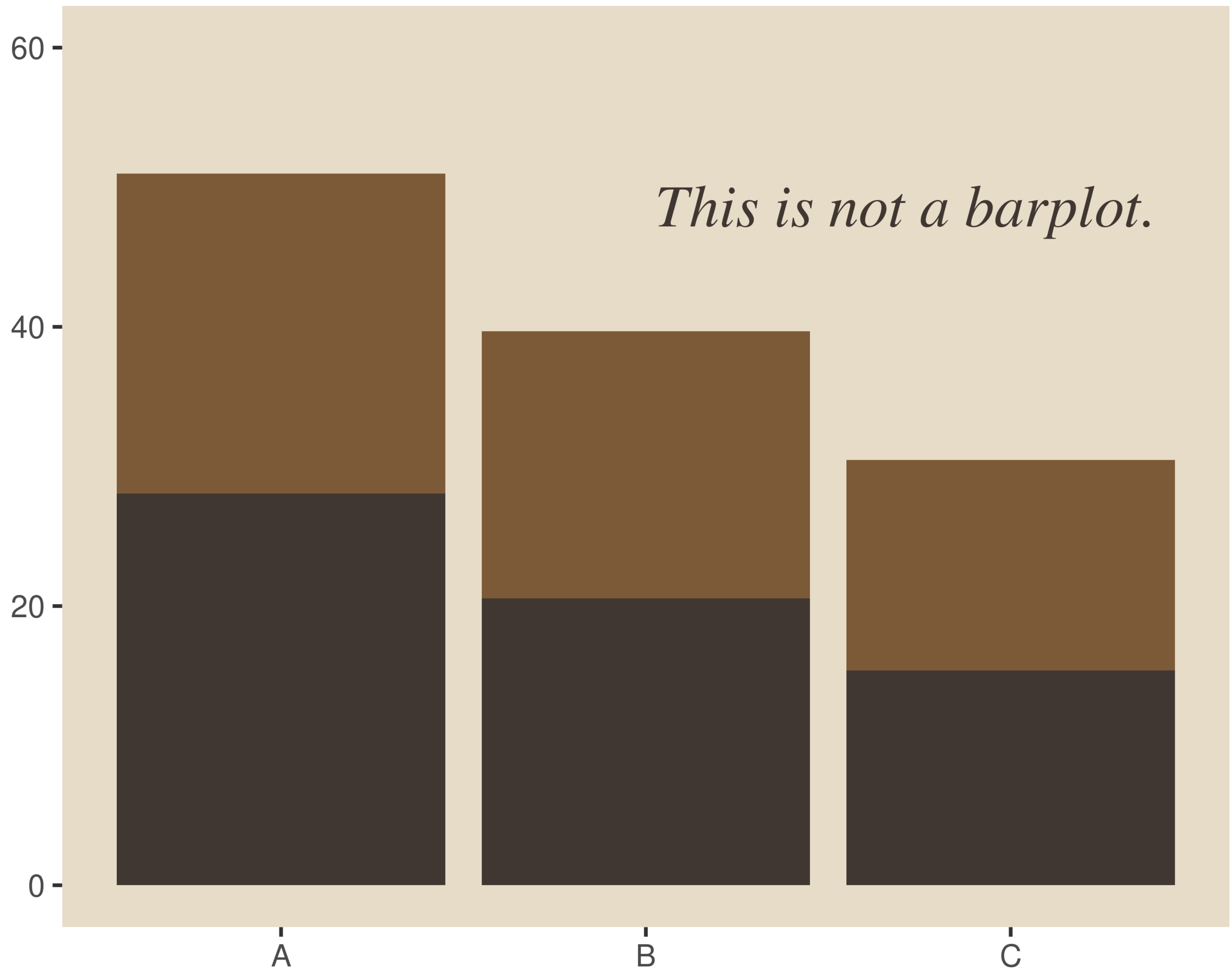
Surprisingly tricky to implement

Some plots work better than others...



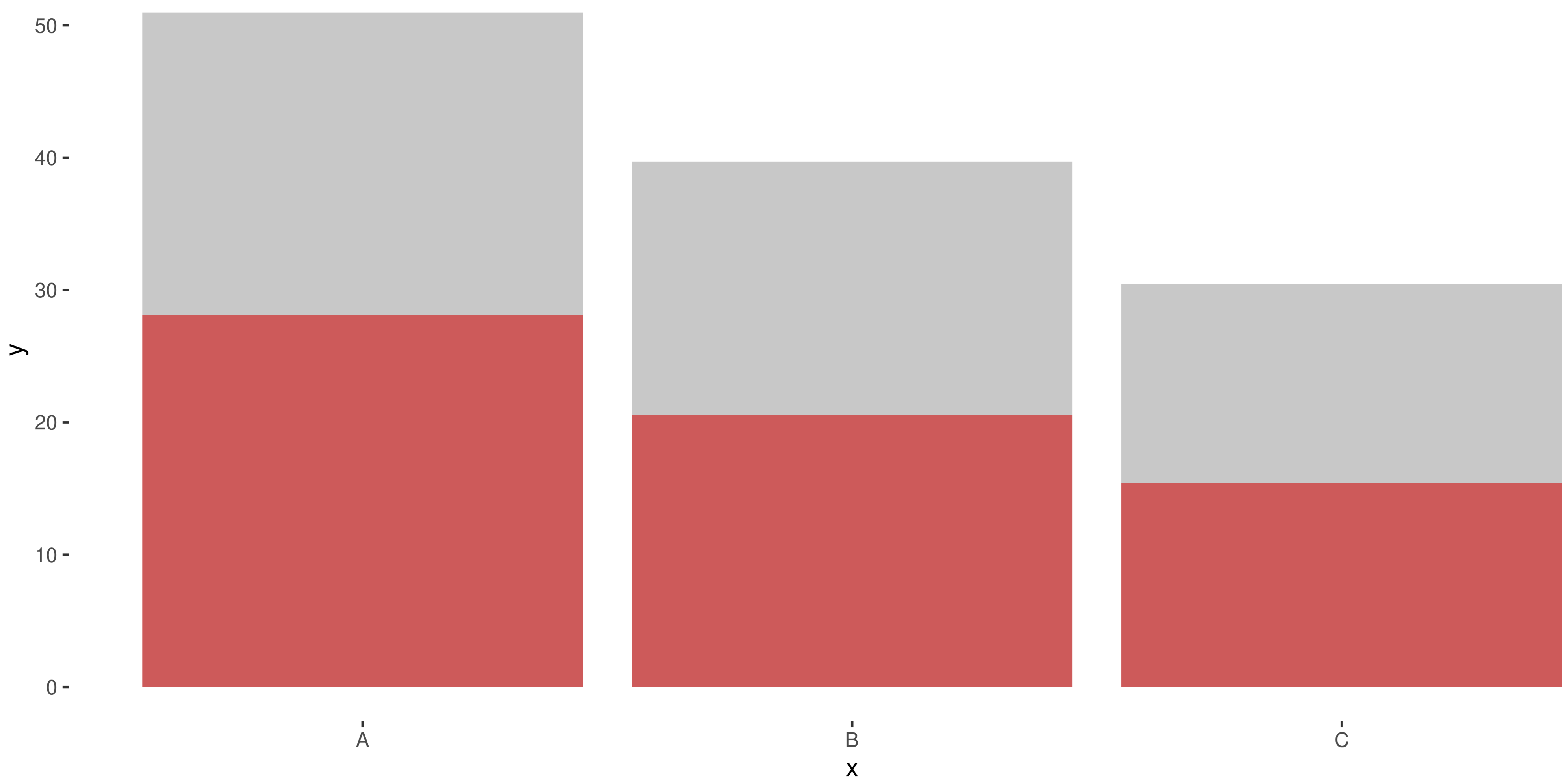
A plot is more than just geometric objects…

The Treachery of Images, René Magritte (1929)

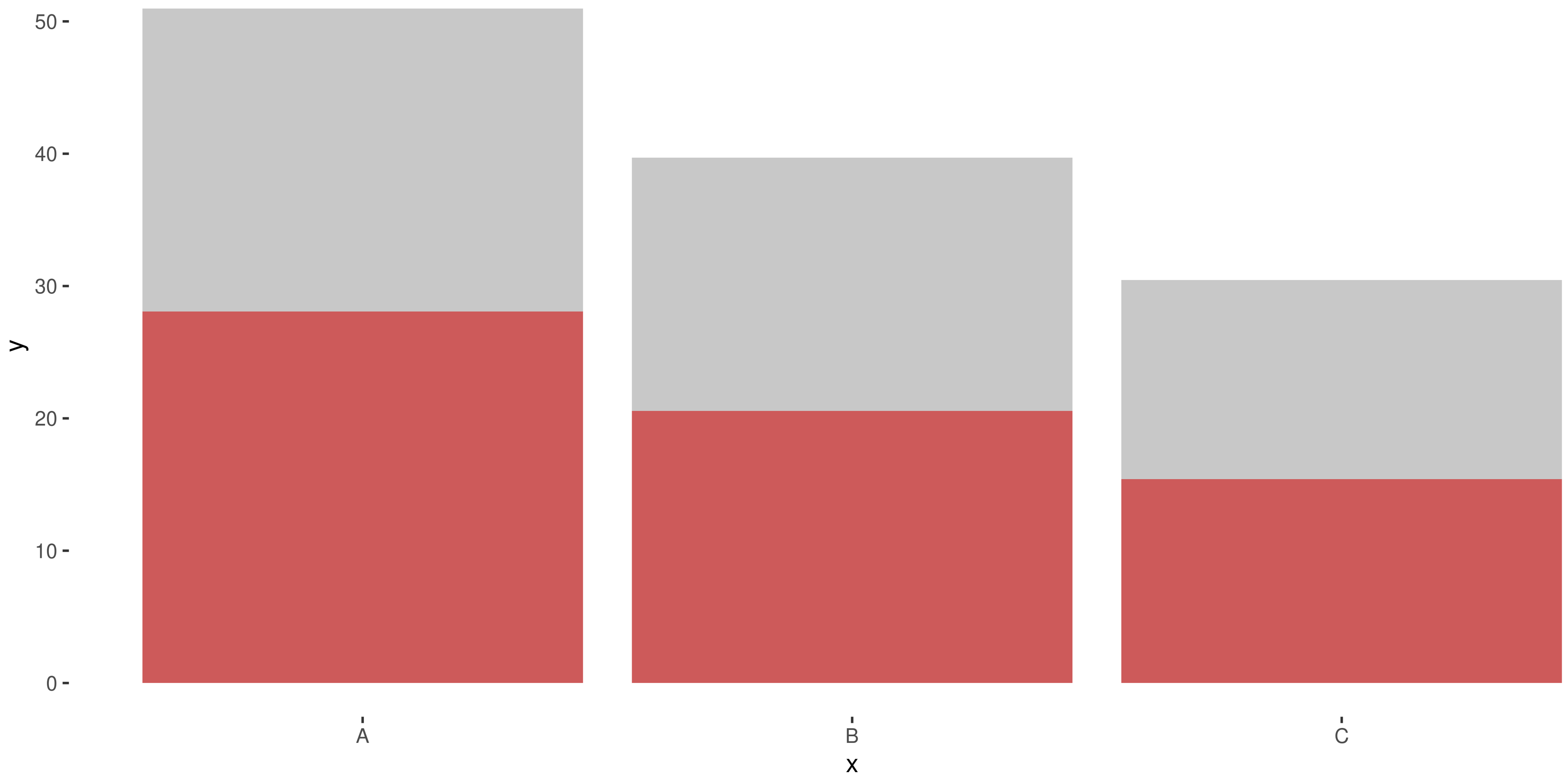


60
40
20
0
This is not a barplot.
A
B
C

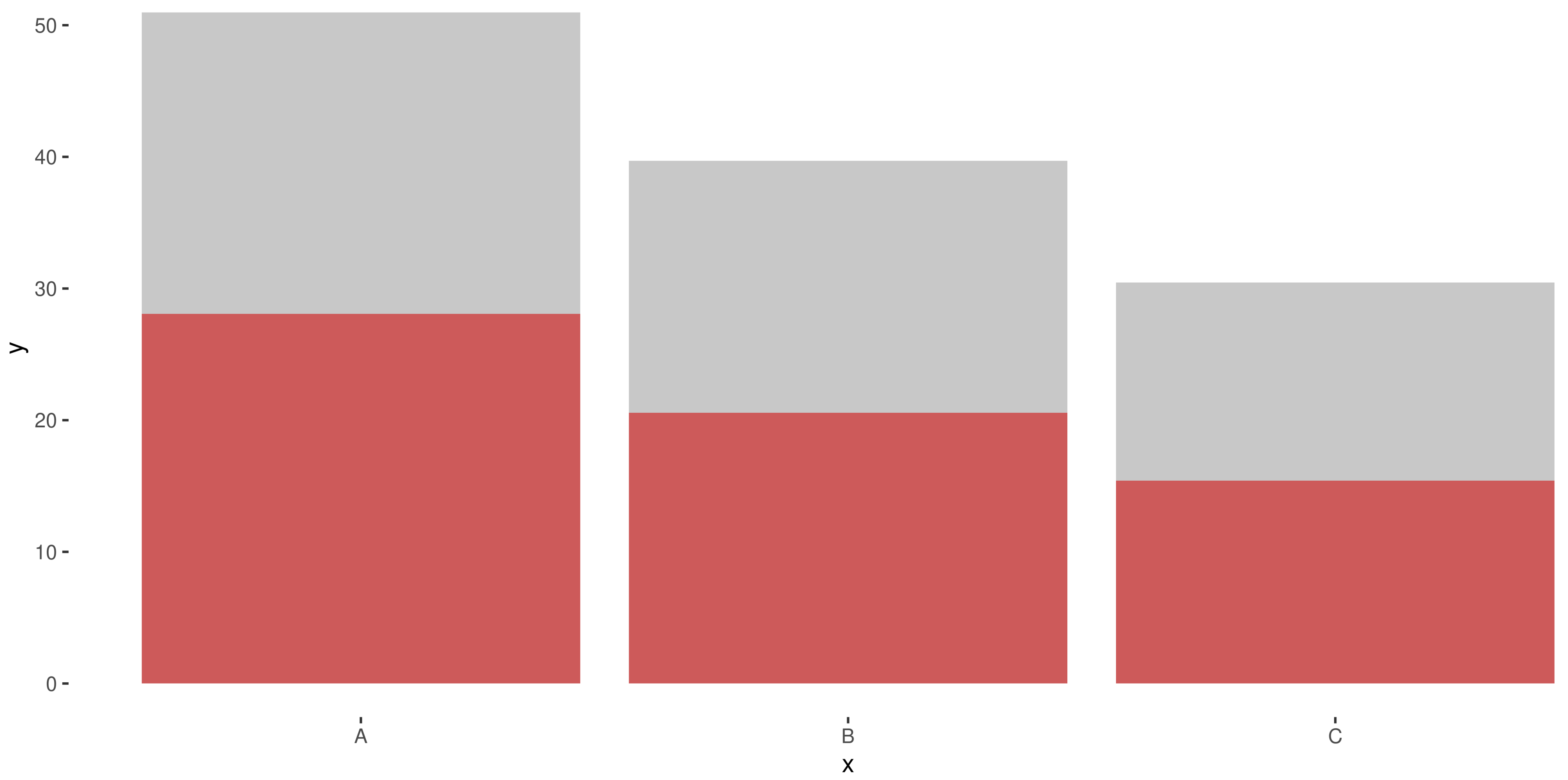
```r
ggplot(mtcars, aes(x, y, fill = z)) +
  geom_bar(stat = "summary", fun = "mean") +
  scale_fill_manual(values = c("grey80", "indianred")) +
  guides(fill = "none")
```



```r
ggplot(mtcars, aes(x, y, fill = z)) +
  geom_bar(stat = "summary", fun = "mean") +
  scale_fill_manual(values = c("grey80", "indianred")) +
  guides(fill = "none")
```



```r
ggplot(mtcars, aes(x, y, fill = z)) +
  geom_bar(stat = "summary", fun = "mean", position = "stack") +
  scale_fill_manual(values = c("grey80", "indianred")) +
  guides(fill = "none")
```



The height of the stacked bars is not a valid summary statistic…

> “Stacking is useful when the sum of the amounts represented by the individual stacked bars is in itself a meaningful amount” (Wilke 2019, 52).

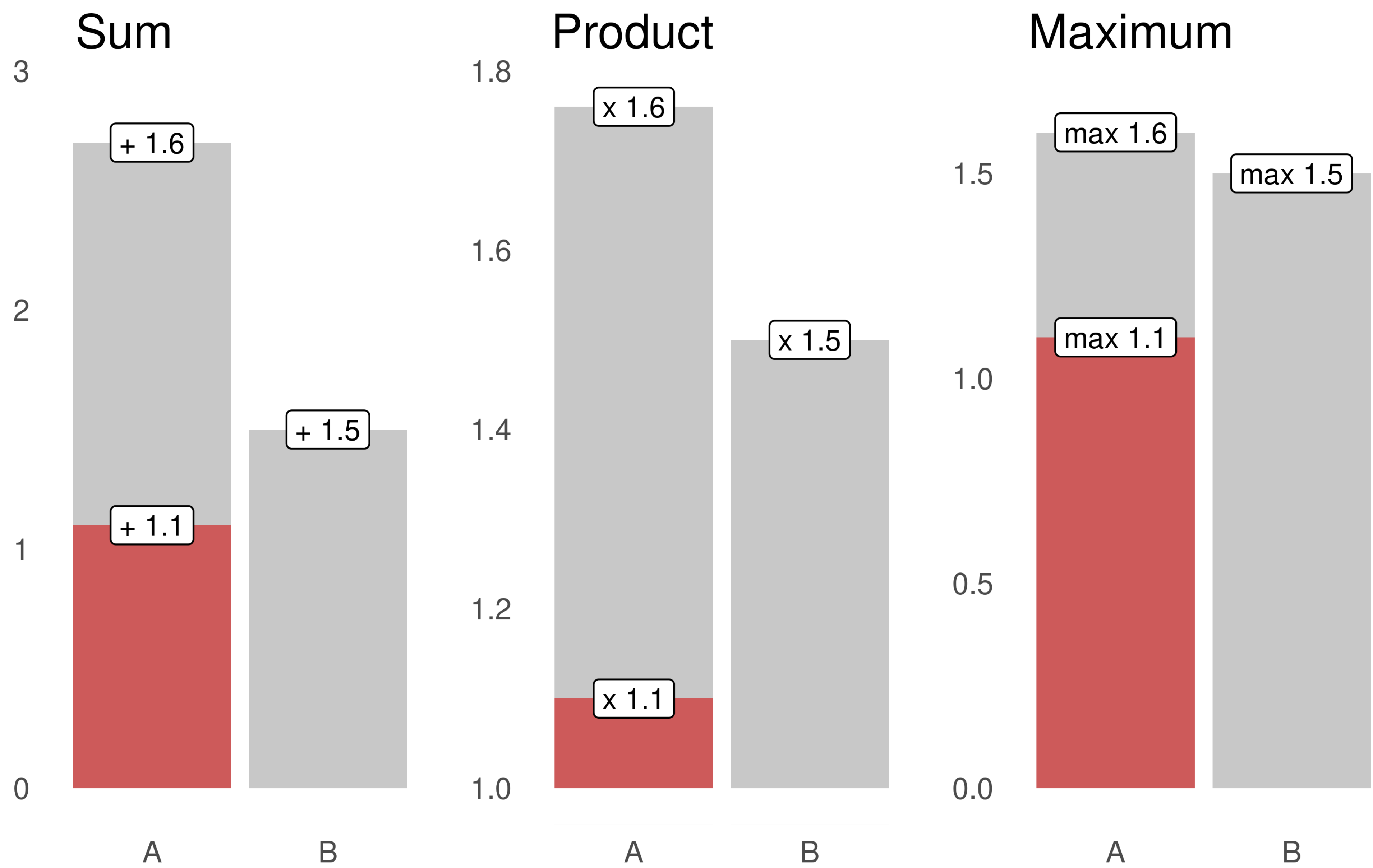
> “[…] It is very important that if the element’s size is used to display a statistic, then that statistic must be summable. Stacking bars that represent counts, sums, or percentages is fine, but a stacked bar chart where bars show average values is generally meaningless.” (Wills 2011, 112).

Can we only do highlighting on sums?

> The ***sum*** of group ***sum***s is equal to the ***sum*** on the whole

> The ***foo*** of group ***foo***s is equal to the ***foo*** on the whole

Other statistics behave this way too…



Sum
3
2
1
0
+ 1.6
+ 1.1
+ 1.5
A
B
Product
1.8
1.6
1.4
1.2
1.0
x 1.6
x 1.1
x 1.5
A
B
Maximum
1.5
1.0
0.5
0.0
max 1.6
max 1.1
max 1.5
A
B

We need some math…

Monoids

- A monoid is a tuple $(M, \otimes, e)$ consisting of of:
  - A set of objects M
  - A binary operation $\otimes : M \times M \to M$
  - A neutral object $e \in M$
- Subject to two rules…

(see e.g. Fong and Spivak 2019; Lawvere and Schanuel 2009)

Monoids: Rules

- Unitality: $x \otimes e = e \otimes x = x$
- Associativity: $x \otimes (y \otimes z) = (x \otimes y) \otimes z$

(see e.g. Fong and Spivak 2019; Lawvere and Schanuel 2009)

Example: Sums

$$1 + 0 = 0 + 1 = 1$$

$$1 + (2 + 3) = (1 + 2) + 3$$

Example: Products

$$1 \times 2 = 2 \times 1 = 2$$

$$2 \times (3 \times 4) = (2 \times 3) \times 4$$

Example: Maximum

$$\max(x, -\infty) = \max(-\infty, x) = x$$

$$\max(x, \max(y, z)) = \max(\max(x, y), z)$$

Counterexample: Exponentiation

$$(x)^1 = x \qquad \text{but} \qquad 1^x \neq x$$

$$(x^y)^z \neq x^{(y^z)}$$

Counterexample: Averages

$$\text{mean}(x, ?) = x$$

$$\text{mean}(x, \text{mean}(y, z)) \neq \text{mean}(\text{mean}(x, y), z)$$

Monoids preserve set union

- Suppose we have some subsets of the data $A, B \subseteq D$, and a monoidal summary F such that e.g.:

$$F(A) = a_1 \otimes a_2 \otimes \ldots \otimes a_n$$

- Combining summaries is the same as summarizing union:

$$\begin{aligned} F(A) \otimes F(B) &= (a_1 \otimes a_2 \otimes \ldots a_n) \otimes (b_1 \otimes b_2 \otimes \ldots b_m) \\ &= a_1 \otimes a_2 \otimes \ldots a_n \otimes b_1 \otimes b_2 \otimes \ldots b_m \\ &= F(A \cup B) \end{aligned}$$

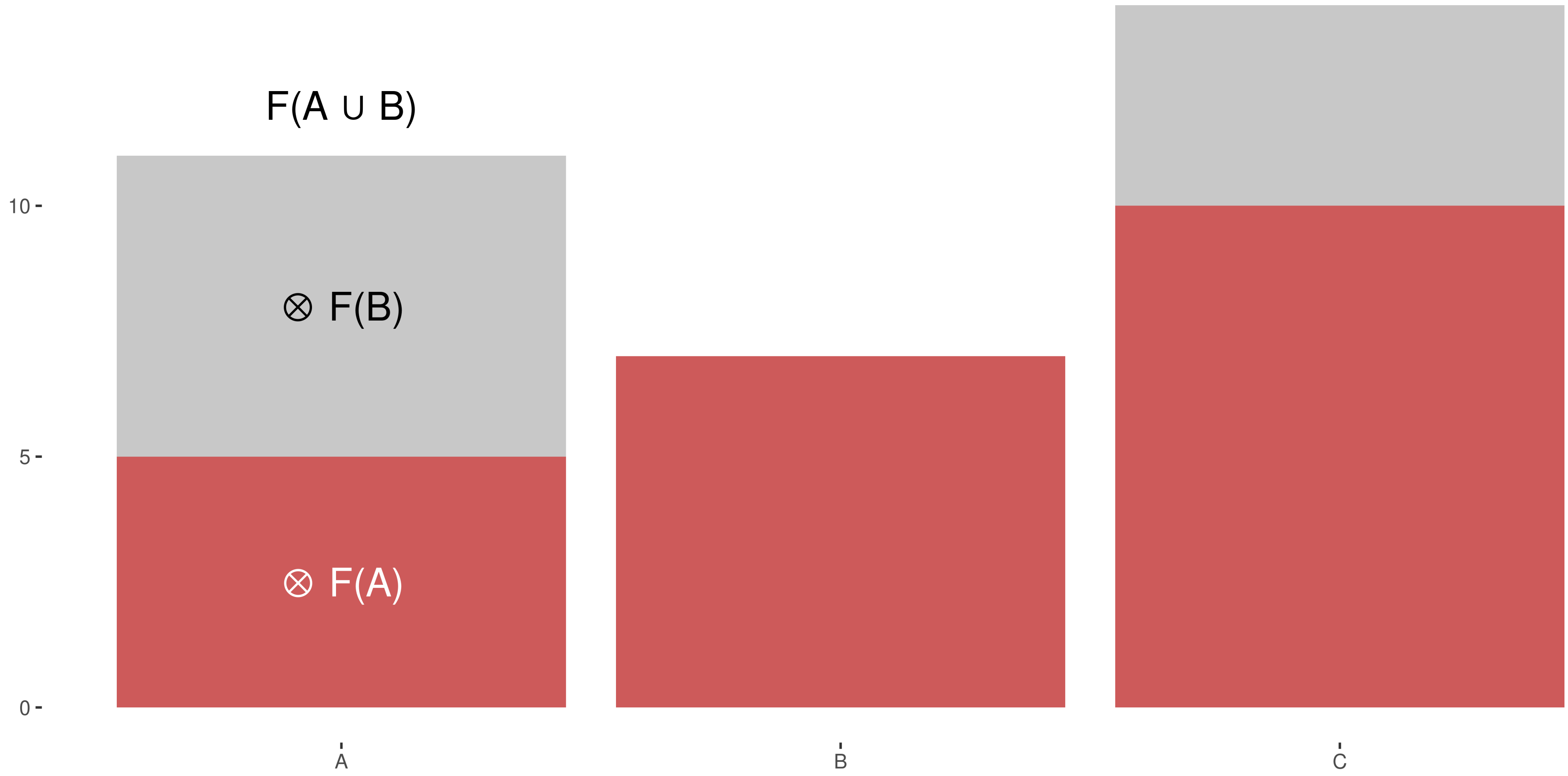
As a commutative diagram…

$$
\begin{tikzcd}
 & A \arrow[dl, "-\cup B"'] \arrow[dr, "-\cup(B\cup C)"] \arrow[dd, dashed] & \\
A \cup B \arrow[rr, "-\cup C"'] \arrow[dd, dashed] & & A \cup B \cup C \arrow[dd, dashed] \\
 & F(A) \arrow[dl, "-\otimes F(B)"'] \arrow[dr, "-\otimes F(B\cup C)"] & \\
F(A \cup B) \arrow[rr, "-\otimes F(C)"'] & & F(A \cup B \cup C)
\end{tikzcd}
$$

We can compare nested subsets!

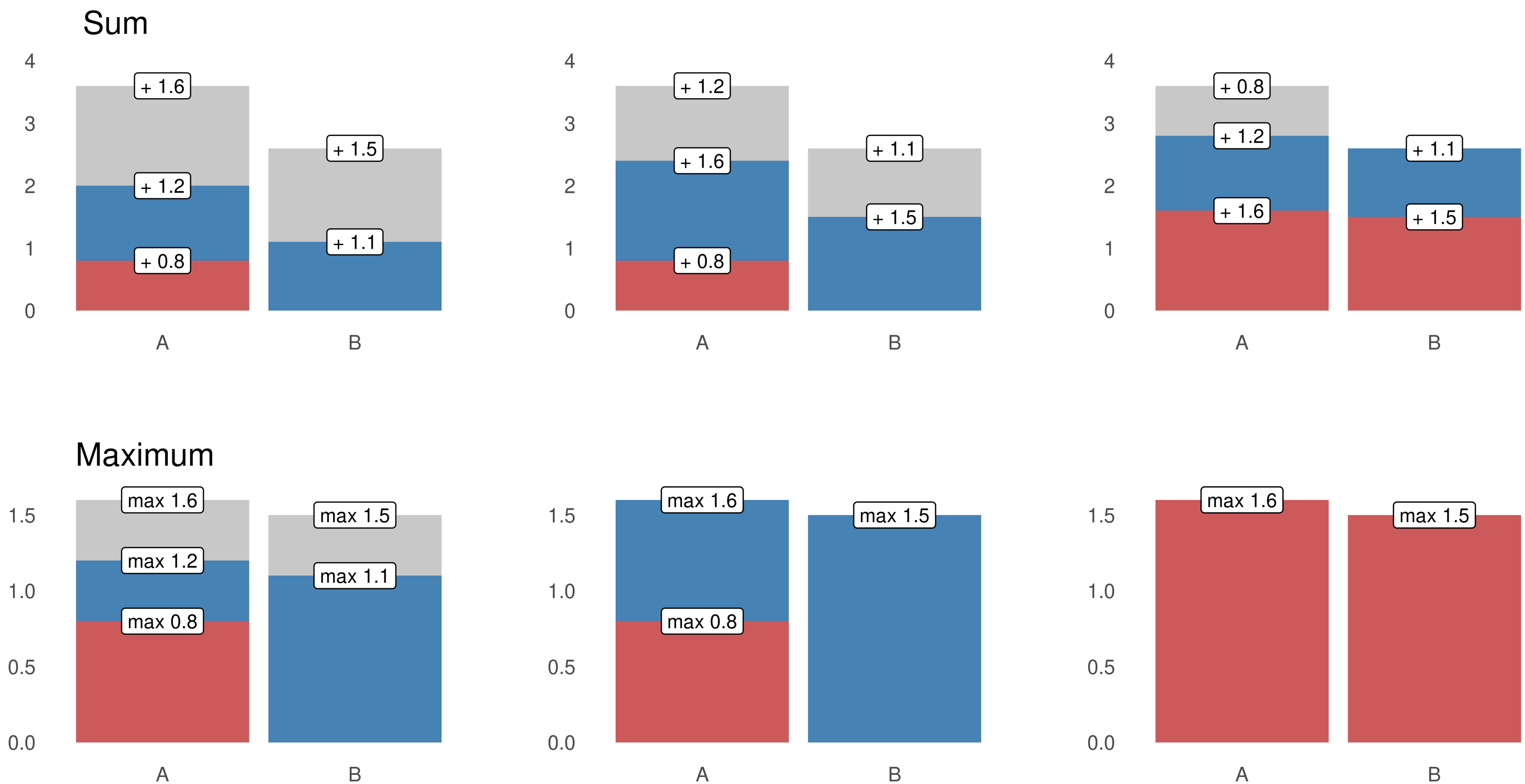
$F(A)$ vs. $F(A \cup B)$

Linked selection and monoids...



Bonus: What about disjoint subsets?

- $F(A)$ vs. $F(B)$



Bonus: Groups

- The monoidal product can “collapse” info about subsets
- To recover $F(B)$ from $F(A \cup B)$ and $F(A)$, we also need the inverse operator:

$$F(B) = F(A \cup B) \otimes^{-1} F(A)$$

- The inverse exists for e.g. sums (minus) but not for max
- Monoid + inverse operator = *Group*

Conclusion

- Interactivity is useful for exploring data
- Especially linked selection/highlighting
- We can do highlighting with statistics other than sums
- However, need to have specific algebraic properties:
  - Monoids for single-group highlighting (nested subsets)
  - Groups for multi-group highlighting (disjoint subsets)

Thank you!

https://github.com/bartonicek/plotscape/tree/master/packages/plotscaper

References

Buja, Andreas, Dianne Cook, and Deborah F Swayne. 1996. "Interactive High-Dimensional Data Visualization." *Journal of Computational and Graphical Statistics* 5 (1): 78–99.

Fong, Brendan, and David I Spivak. 2019. *An Invitation to Applied Category Theory: Seven Sketches in Compositionality*. Cambridge University Press.

Heer, Jeffrey, and Ben Shneiderman. 2012. "Interactive Dynamics for Visual Analysis: A Taxonomy of Tools That Support the Fluent and Flexible Use of Visualizations." *Queue* 10 (2): 30–55.

Lawvere, F William, and Stephen H Schanuel. 2009. *Conceptual Mathematics: A First Introduction to Categories*. Cambridge University Press.

Magritte, René. 1929. "The Treachery of Images (This is Not a Pipe) (La trahison des images [Ceci n'est pas une pipe]) | LACMA Collections." https://collections.lacma.org/node/239578.

Ward, Matthew O, Georges Grinstein, and Daniel Keim. 2015. *Interactive Data Visualization: Foundations, Techniques, and Applications*. CRC Press.

Ware, Colin. 2019. *Information Visualization: Perception for Design*. Morgan Kaufmann.

Wilke, Claus O. 2019. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media.

Wills, Graham. 2011. *Visualizing Time: Designing Graphical Representations for Statistical Data*. Springer Science & Business Media.