

Summary statistics %>%

Distance function %>%

Threshold %>%

Things that make you go *hmmm*?

# Issues to be aware of

- Summary statistics may not catch relevant features of the data
  - can decrease dimension too much - lose information
  - don't decrease dimension enough - didn't solve the problem
  - can be correlated - redundant dimensions
  - etc
- How multiple summary statistics/elements of a summary statistic contribute to the distance?
  - E.g. Multiple summary statistics can each have wildly different scales