Data dimensionality

- Dimensionality of the data often causes problem when running LFI methods
  - In high dimensions it becomes increasing improbable to generate data close to the observed data
- The (current) standard approach is still to use summary statistics $S(\cdot)$
  - E.g $d(x, x^o) \approx d(S(x), S(x^o)) \quad (\approx \rho(x, x^o))$
- Sufficient statistics usually do not exist
- How to choose them?

Selecting summary statistics

- An open problem
- Often we use bespoke summary statistics
  - Use **domain expertise** if available
  - **Explore** the simulator prior to inference
  - **Diagnose** the inference results
- Automatic algorithms for **selecting/constructing** the summary statistics