SCIU4T4: Why study statistics?

Why study statistics?

- Importance of good statistics
- Statistical revolution
- Populations and samples
- Expectation and uncertainty
- Types of variables

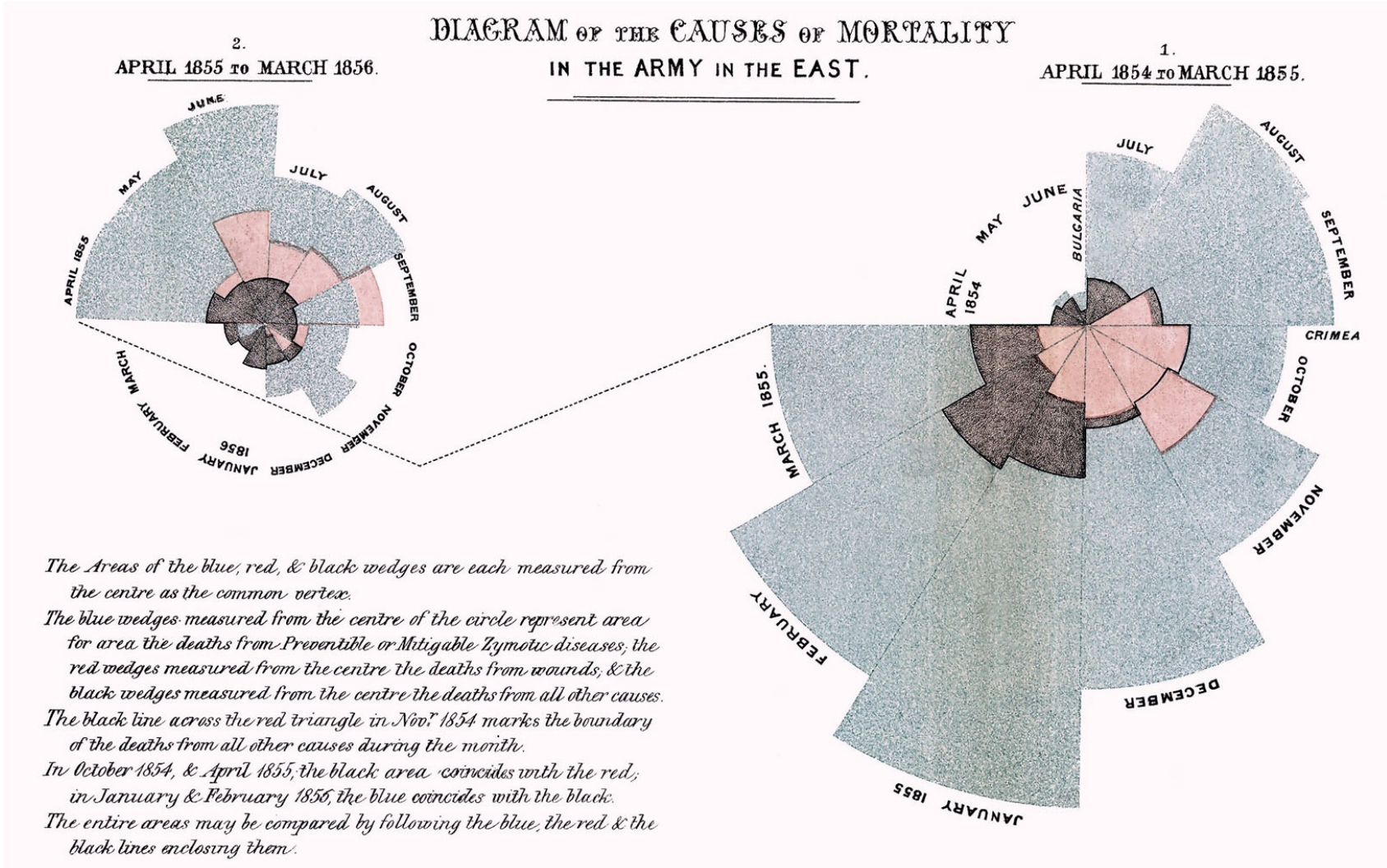
Florence Nightingale

- Pioneer of modern nursing and statistics
- Founded first secular nursing school 1860

Nurse during Crimean war (1853-1856)

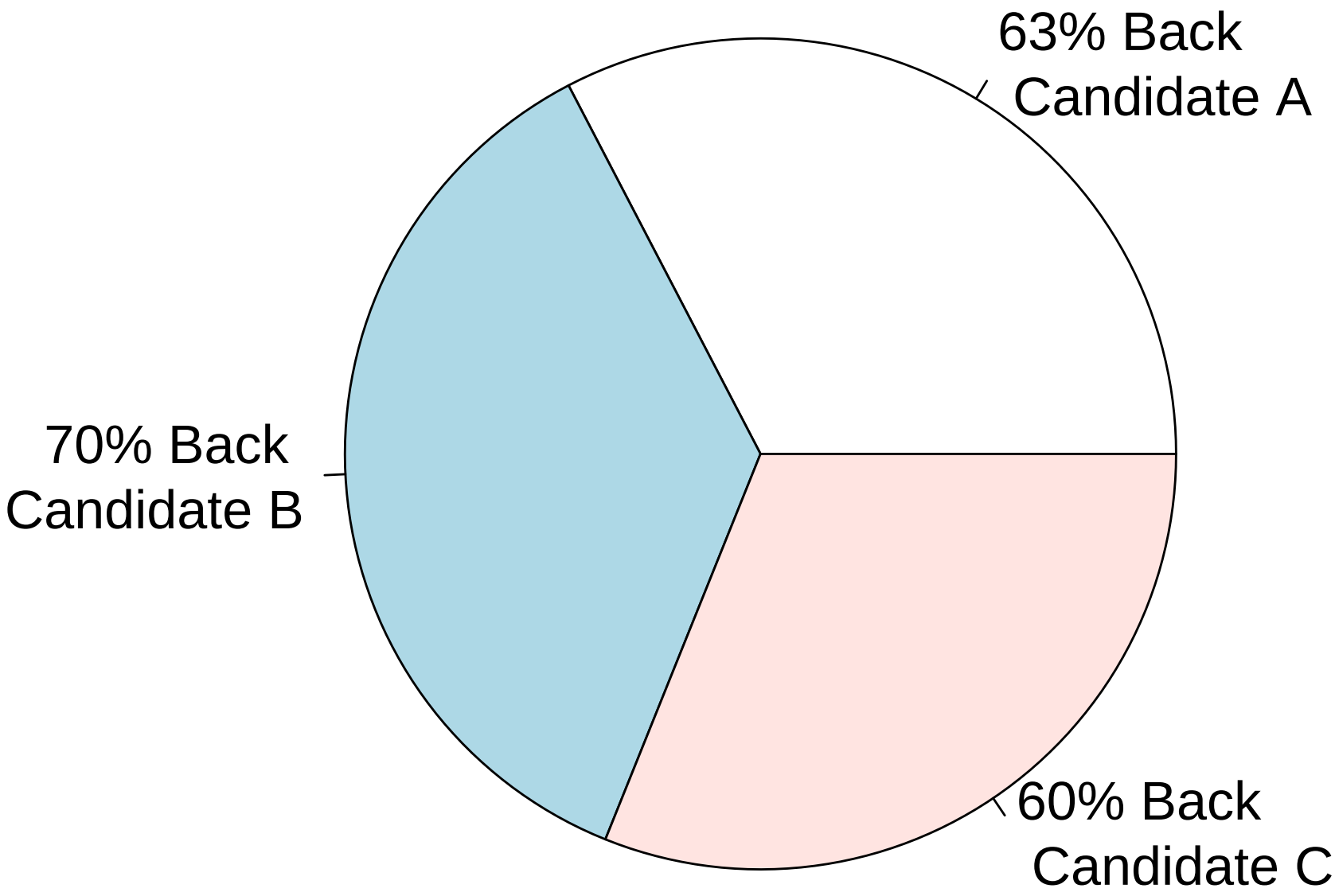
- 16k of 18k deaths due to sanitation
- Fatalities plummeted after improvements
- Persuaded parliament to reform hospitals

Importance of good statistics



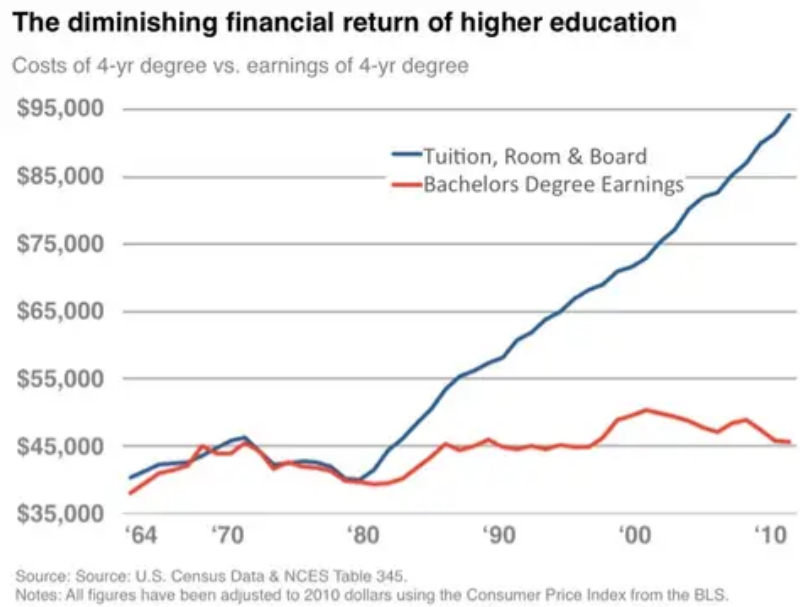
[1] **Image:** Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army (Public domain).

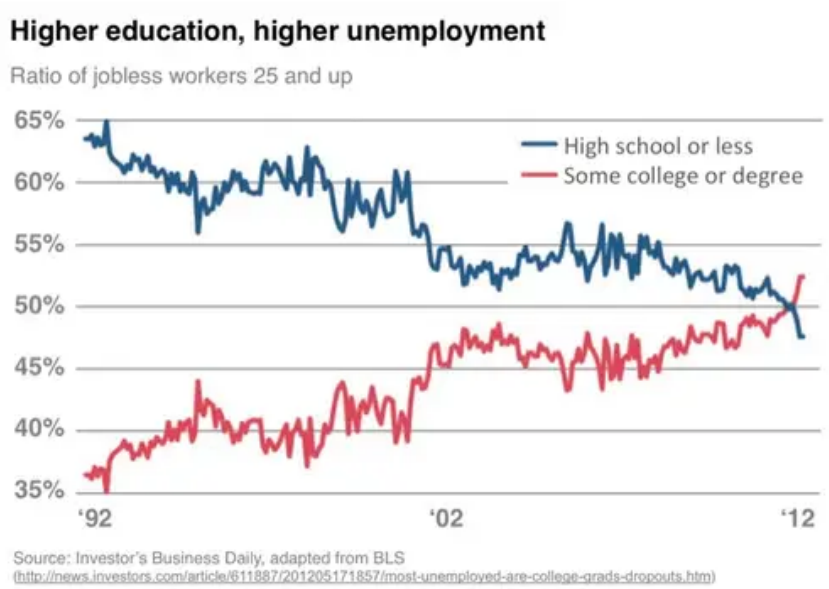
Importance of good statistics



[1]Source: A widely watched news channel

Importance of good statistics

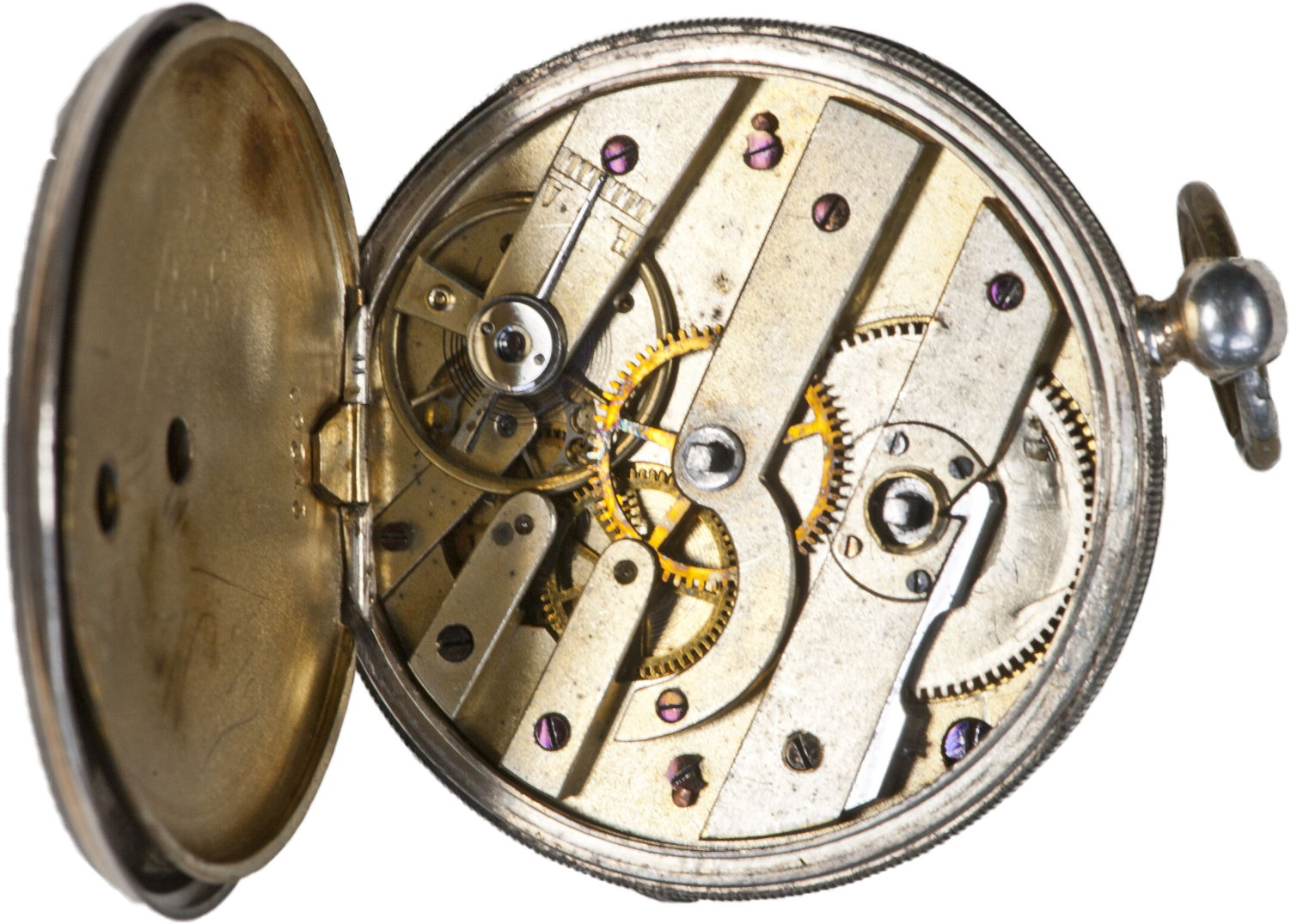




- Sometimes graphs do not tell the full story
- What about the earnings for people without a degree over time?

[1] Blake, D. 2012. These Two Charts Prove A College Education Just Isn't Worth The Money Anymore https://www.businessinsider.com/these-two-charts-prove-a-college-education-just-isnt-worth-the-money-anymore-2012-6.

Statistical revolution: Old view of a clockwork universe



[1] **Image:** Grobe, H. 2017. (Public domain).

Statistical revolution: Old view of a clockwork universe

- Newtonian physics success
- Formulas for laws of motion
- Precise measurement
- Eliminate measurement errors
- Apply to life sciences?

This old view of eliminating uncertainty was challenged in the early 1900s.[1]

[1] Salsburg, D. 2001. The lady tasting tea: How statistics revolutionized science in the twentieth century. Macmillan.

Statistical revolution: A new perspective

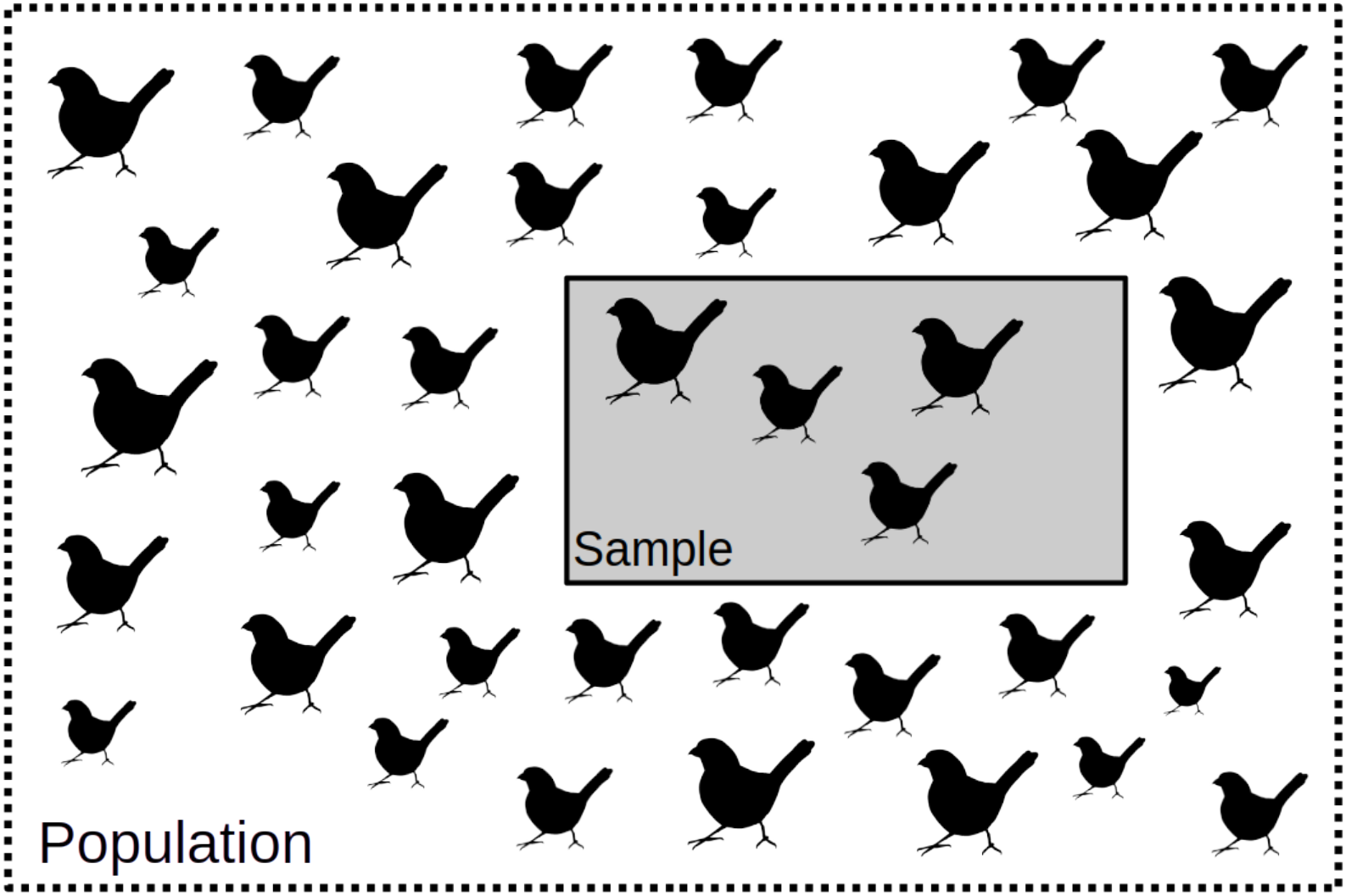


Figure 1: Samples are taken from a subset of a population

Statistical revolution: A new perspective

"In statistics, population always means the **totality of individual observations about which inferences are to be made, existing anywhere in the world or at least within a definitely specified sampling area limited in space and time**"[1]

[1] Sokal, Robert R, and F James Rohlf. 1995. Biometry. 3rd ed. W H Freeman & Company, New York, USA.

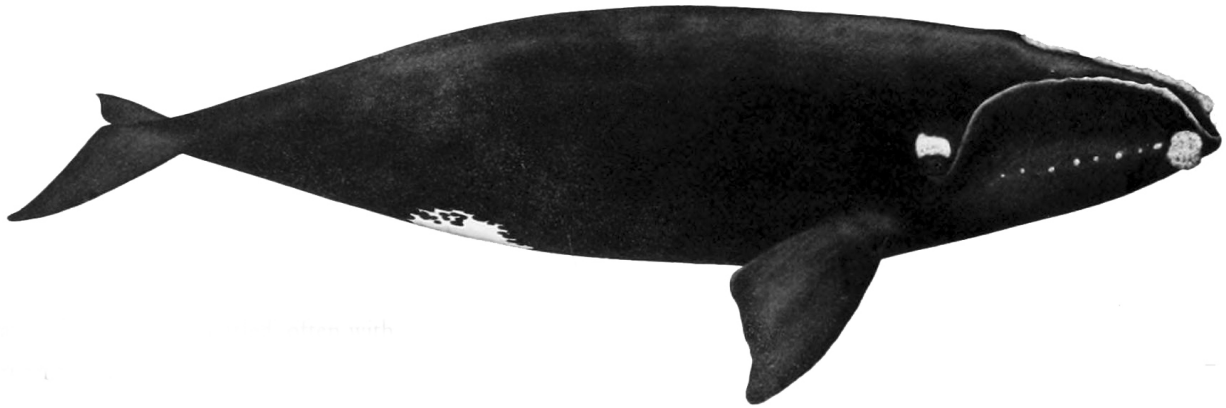
Populations and samples

Statistical population:

- Not necessarily a *biological* population
- Not necessarily an *actual* population
- Might be theoretical, not concrete

[1] **Image:** Wilson, A. 2019. (Public domain)

Populations and samples



North Atlantic right whale (*Eubalaena glacialis*)

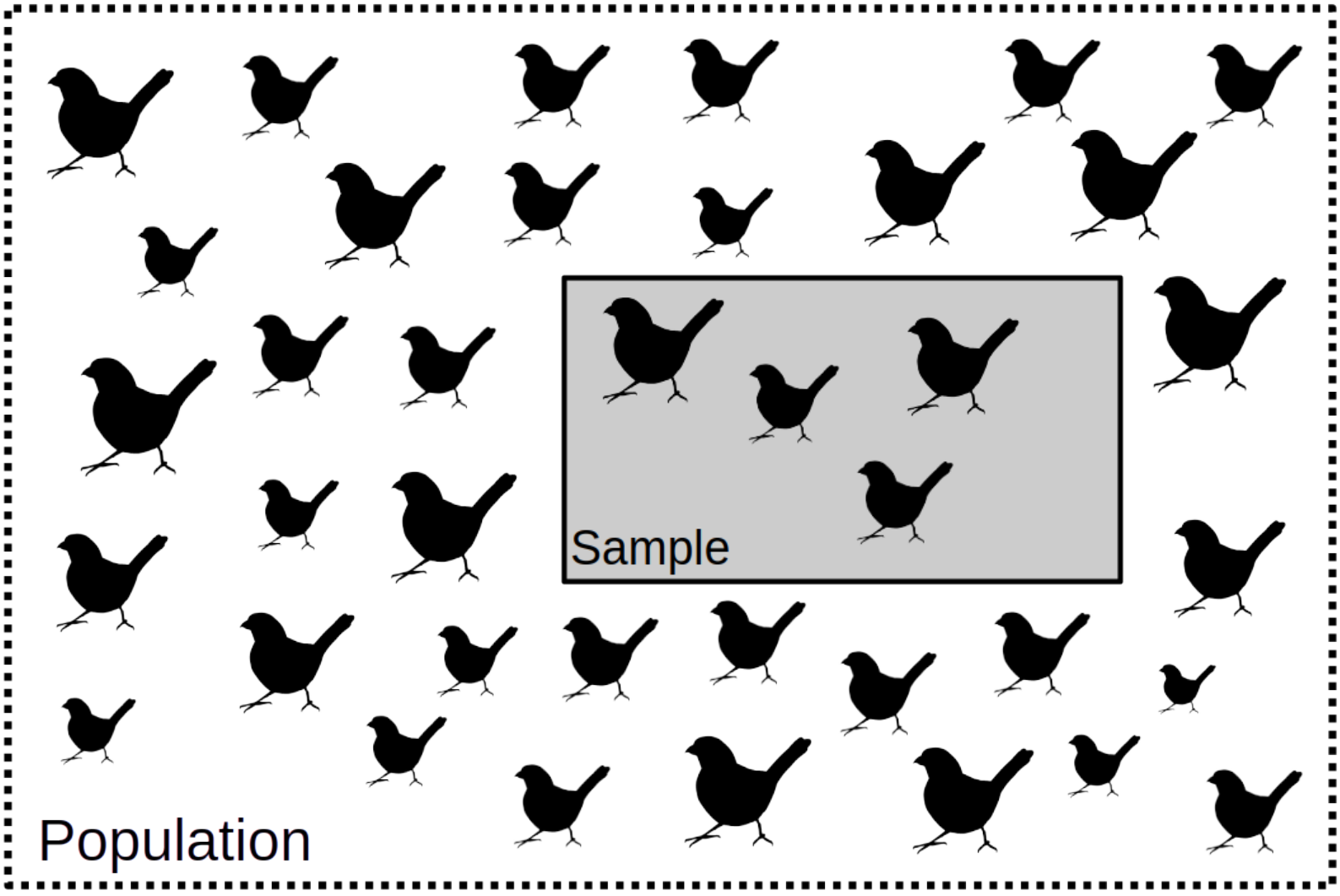
- Estimated 338 left[1]
- Conservation requires estimating traits
- Average value might not be *expected* value[2]

[1] NOAA (2023). North Atlantic right whale (*Eubalaena glacialis*): Western Atlantic Stock.

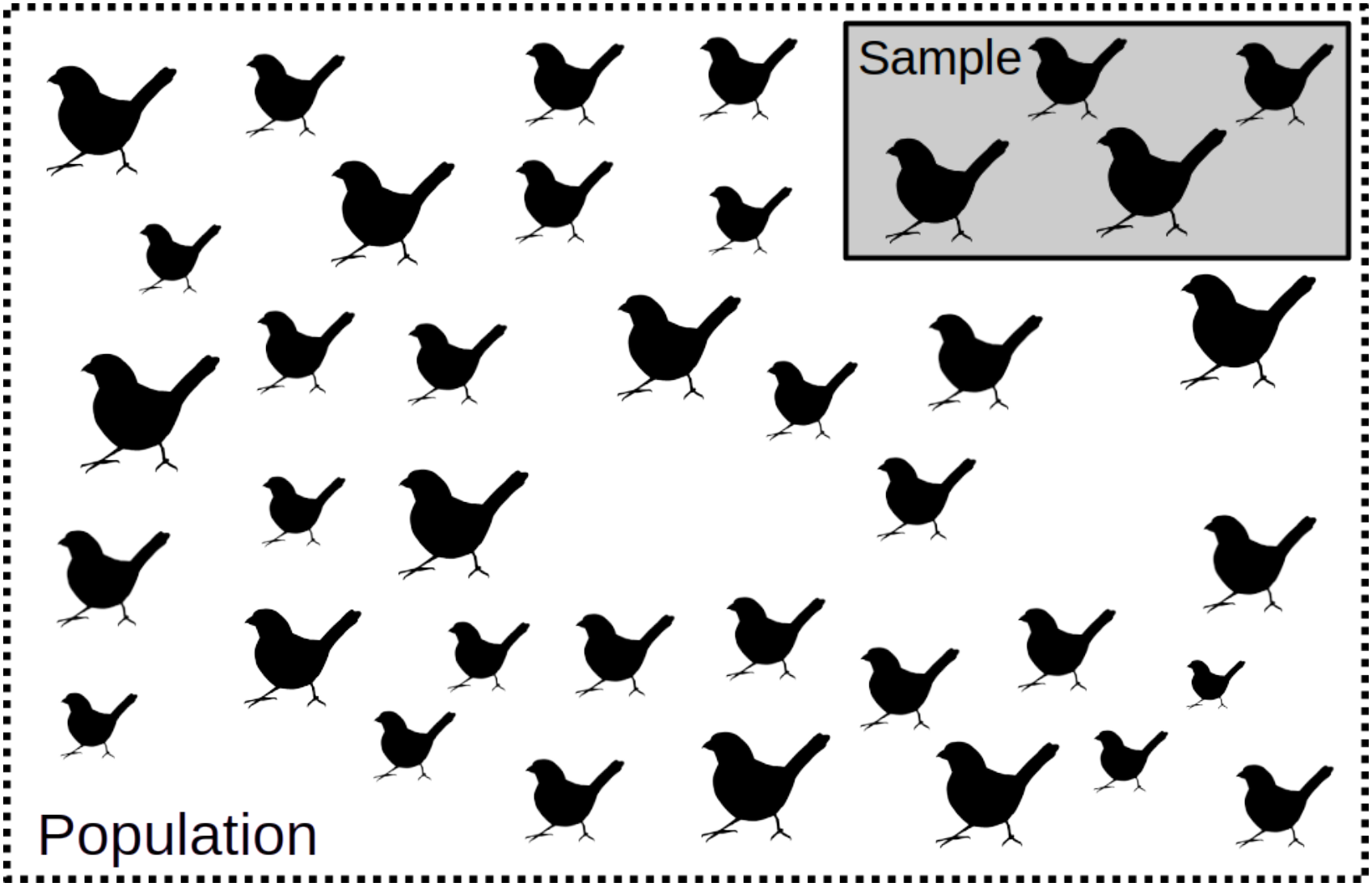
[2] van Veelen, M. (2025). The general version of Hamilton's rule. Elife, 14, RP105065.

[3] **Image:** NOAA. 1987. (Public domain).

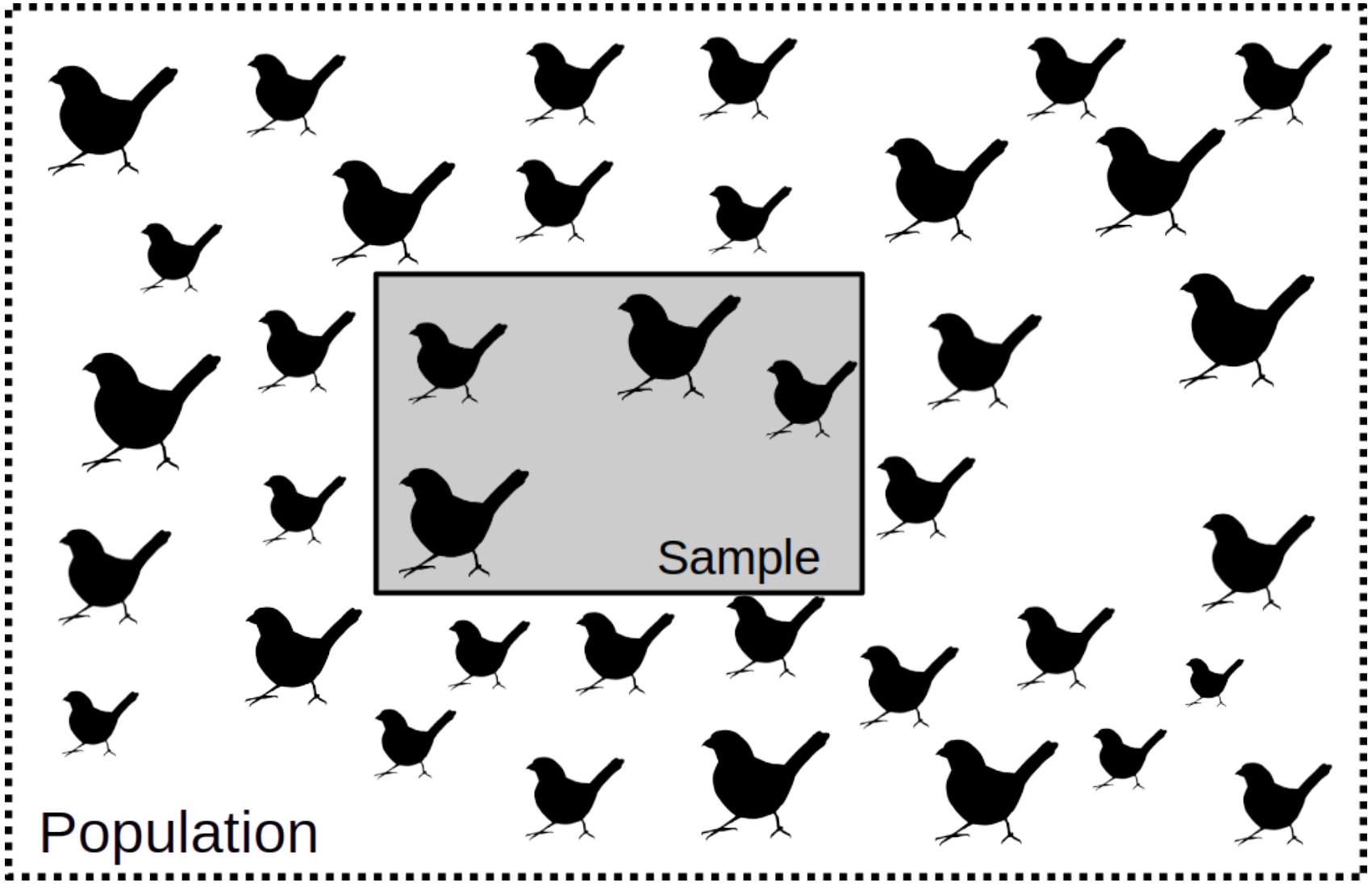
General idea of frequency statistics



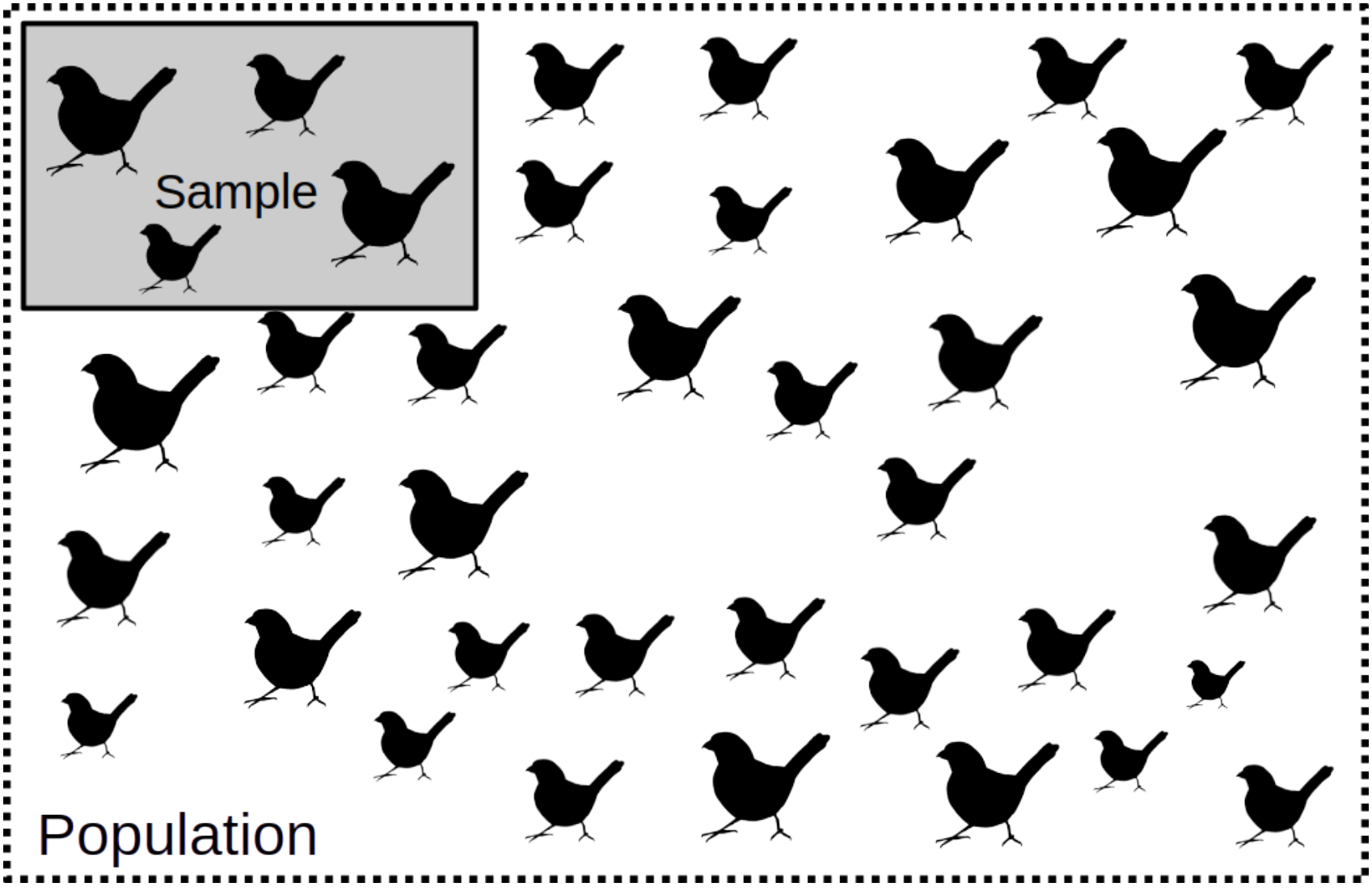
General idea of frequency statistics



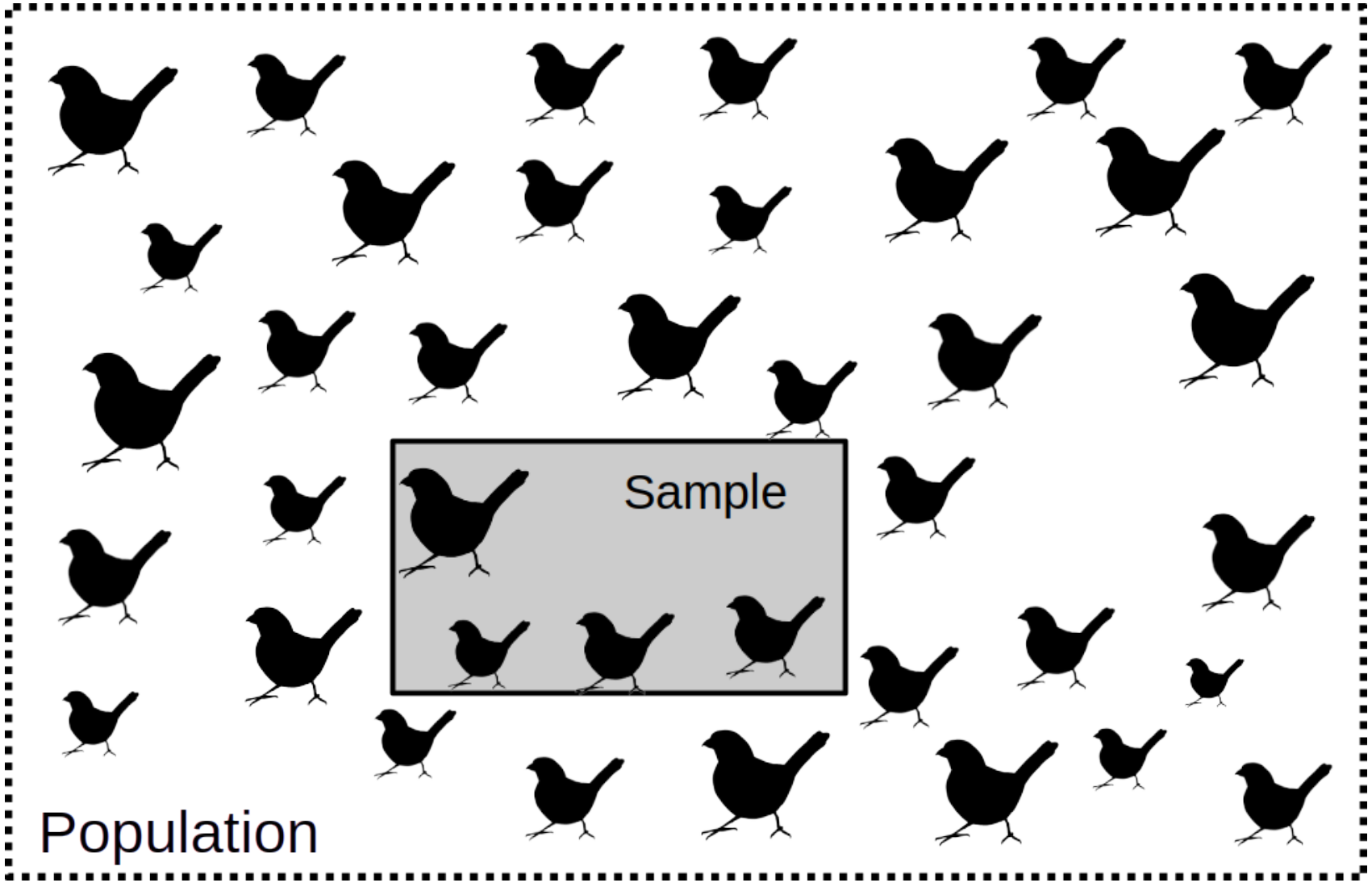
General idea of frequency statistics



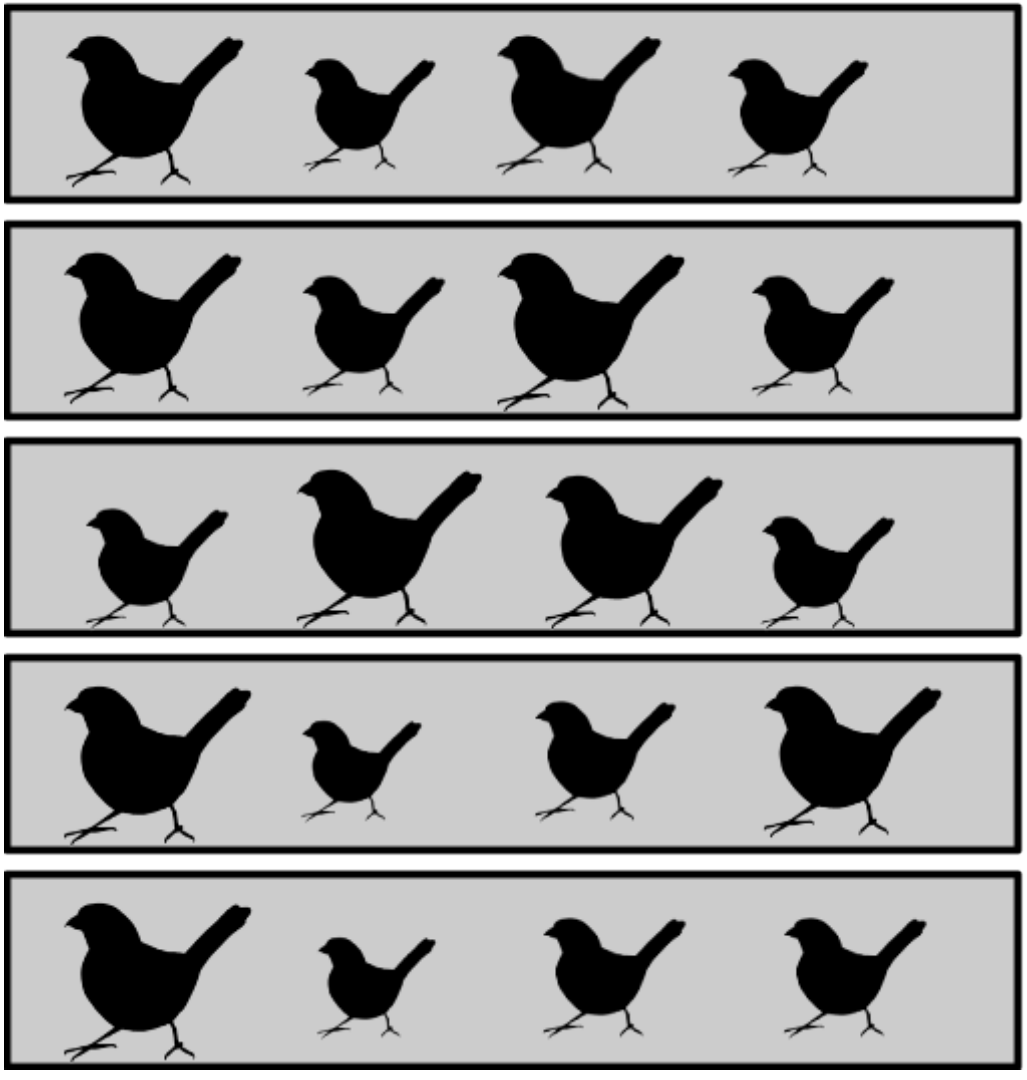
General idea of frequency statistics



General idea of frequency statistics



General idea of frequency statistics

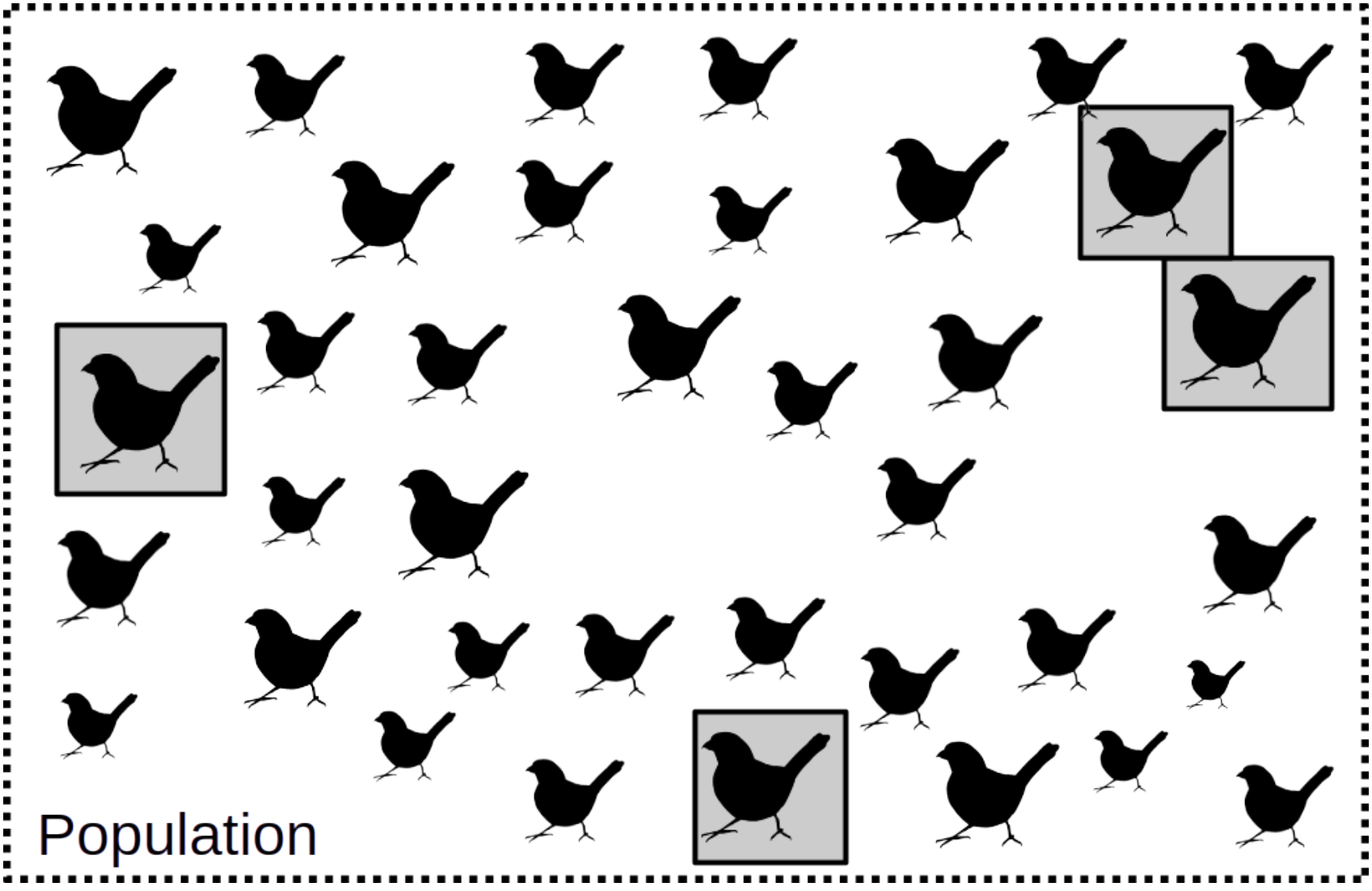


Repeated re-sampling

- What is our expectation?
- What is our uncertainty?

Want to get best estimate of *true* expected measurement and contain uncertainty.

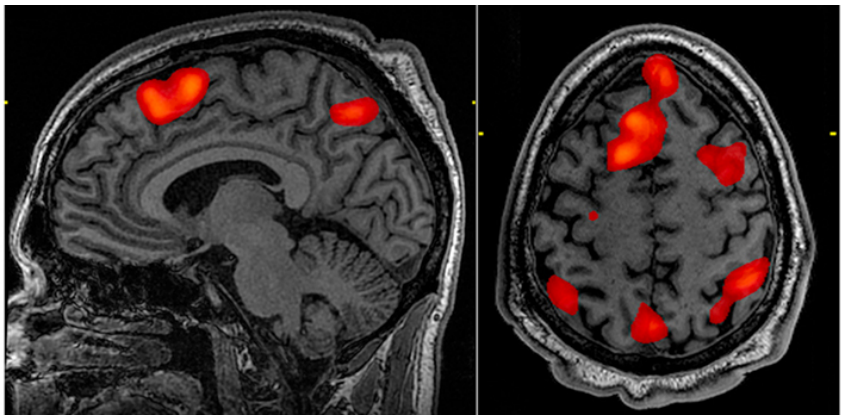
Bias and unrepresentative sample



Bias and unrepresentative sample

Neuroscience: fMRI scans before & after task

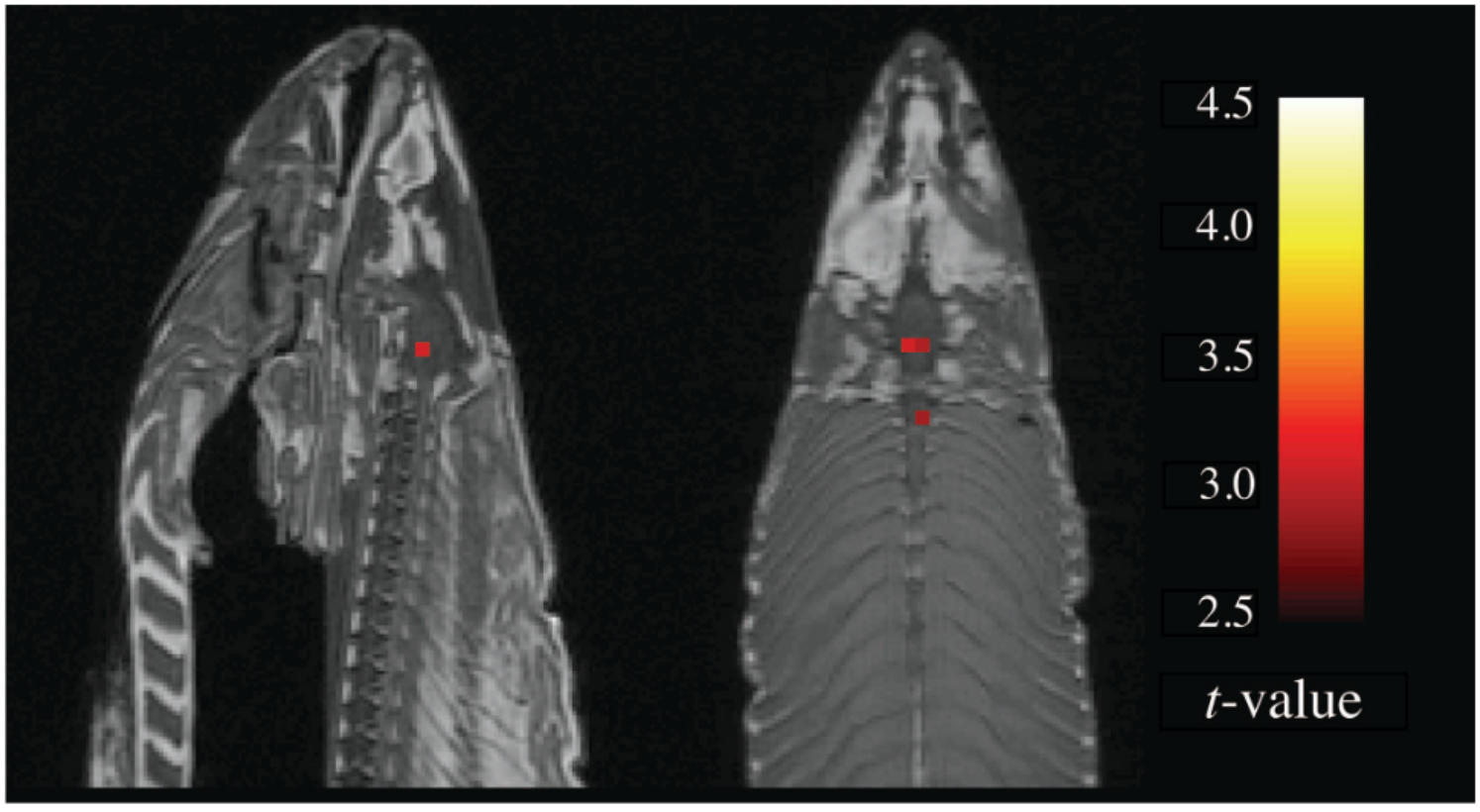
- What parts of brain most active during task?
- Compare thousands of areas of the brain
- Subject shown pictures of human social activity
- Cluster of brain areas showed high activity
- Typical analysis approach highly significant



[1] Reinhart, A. 2015. Statistics done wrong: The woefully complete guide. No starch press.

[2] **Image:** Graner, J. 2010. (Public domain).

Bias and unrepresentative sample



207 different analysis methods in 241 fMRI studies

---

[1]Reinhart, A. 2015. Statistics done wrong: The woefully complete guide. No starch press.

[2]Bennett, CM, et al. 2010. J of Serendipitous and Unexpected Results. 1:1-5.

[3]Carp, J. 2012. The secret lives of experiments: methods reporting in the fMRI literature. Neuroimage, 63:289-300. 10.1016/j.neuroimage.2012.07.004

Predicting one variable from another

- **Dependent variable**
  - Response variable
  - Shown on Y axis
- **Independent variable**
  - Explanatory variable
  - Shown on X axis

Predicting one variable from another

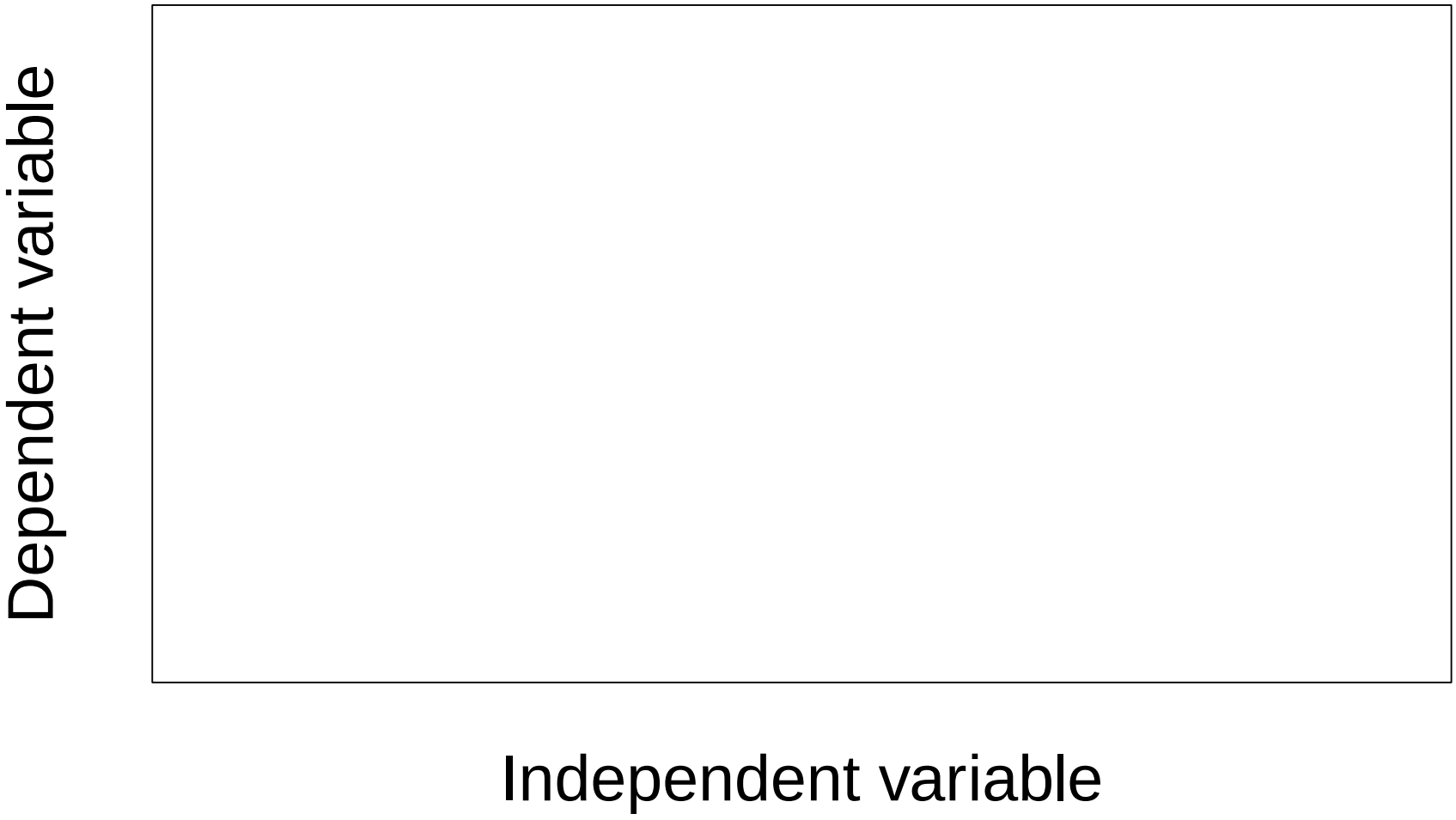
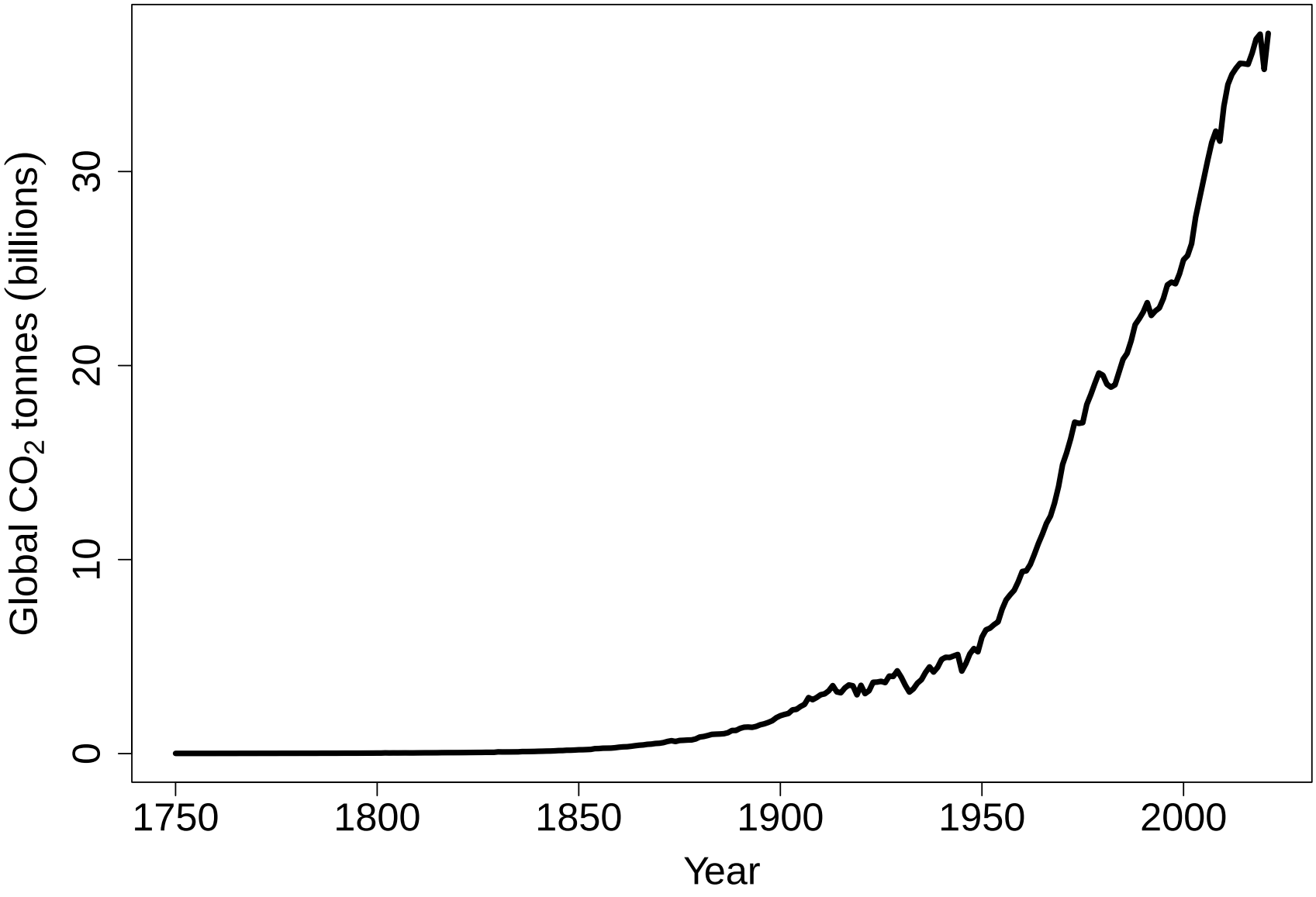


Figure 2: Dependent variable on the vertical (Y) axis, and independent variable on the horizontal (X) axis.

Predicting one variable from another



Types of variables: definitions

- **Categorical:** Fixed number of options
  - Nominal: No inherent order
  - Ordinal: Inherent order
- **Quantitative:** Numbers meaningful
  - Discrete: Limited number of values
  - Continuous: Any real number

Types of variables: examples

- **Categorical:** Fixed number of options
  - Nominal: Tundra, Boreal, Tropical
  - Ordinal: Disagree, Neutral, Agree
- **Quantitative:** Numbers meaningful
  - Discrete: Number of offspring
  - Continuous: Body mass

Types of variables: fig wasps collected from fruit

Species	Fruit	Count	Fruit length (mm)
Het1	F1	0	15
Het1	F2	0	14
Het1	F3	0	16
Het1	F4	1	15
Het1	F5	0	14
Het2	F1	0	15
Het2	F2	2	14
Het2	F3	3	16

Types of variables: implications

- Different variables handled in different ways
- Need to use the right plot or test for a variable type