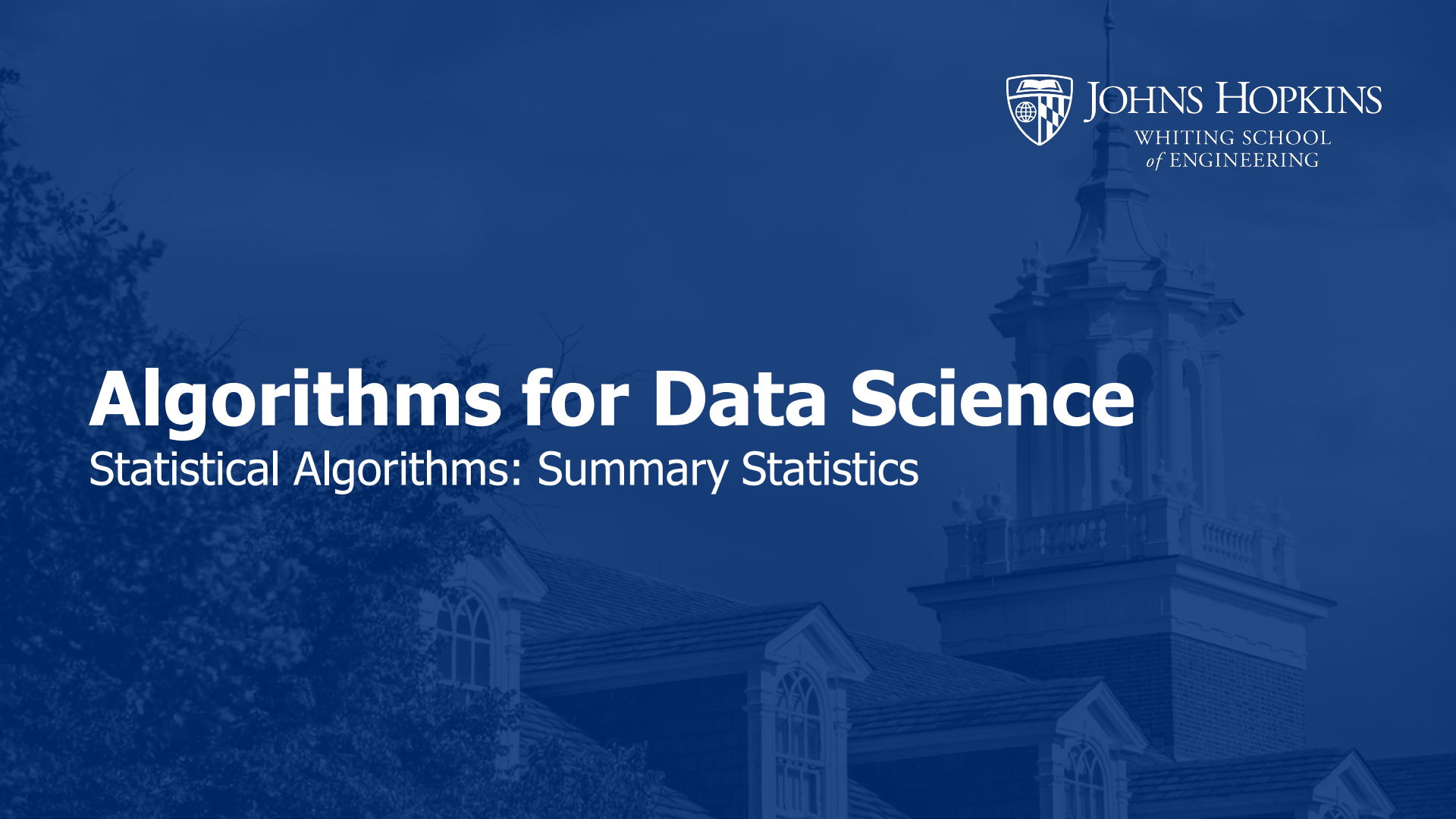

Algorithms for Data Science

Statistical Algorithms: Summary Statistics

Summary Statistics

Numerical measures that summarize and describe the main features of a dataset.

- **Purpose:**
  - Simplify large datasets into understandable metrics.
  - Provide insights into data distribution and variability.
  - Serve as inputs for statistical algorithms.

X1	X2	X3	X4	X5	
1	2	3	4	5	…
2	3	4	5	6	…
⋮	⋮	⋮	⋮	⋮	

	X1	X2	X3	X4	X5
Min	1	2	3	4	5
Max	2	3	4	5	6
Mean	1.5	2.5	3.5	4.5	5.5

Common Summary Statistics

Central Tendency

- **Mean:** The average of all data points.
- **Median:** The middle value when data is ordered.
- **Mode:** The most frequently occurring value.

Spread

- **Range:** Difference between the maximum and minimum values.
- **Variance:** Average squared deviation from the mean.
- **Standard Deviation:** Square root of variance; measures spread around the mean.

Other

- **Skewness:** Indicates asymmetry in the data distribution.
- **Kurtosis:** Measures the "tailedness" or extremes of the distribution.

Visualizing Summary Statistics