The Big Picture

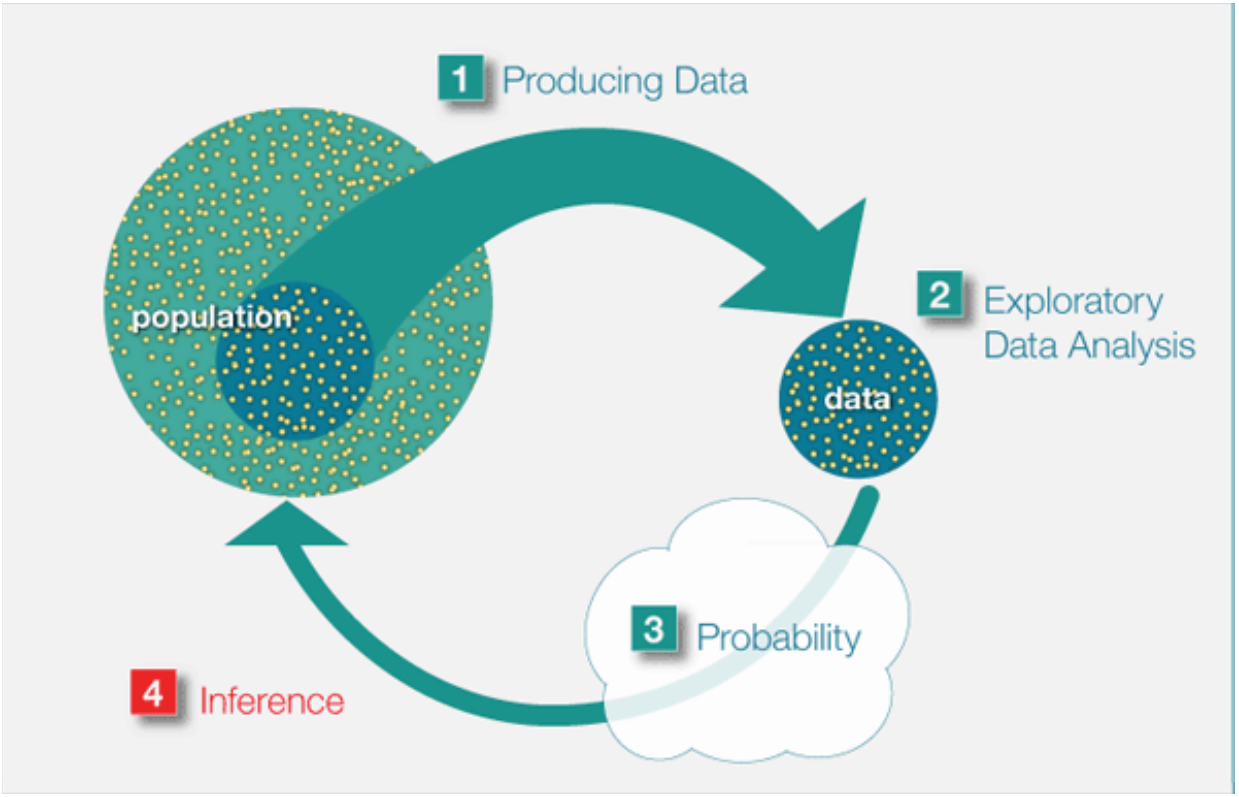

Figure: Our Subject Mostly 3 and 4

4 Phases in Data Science

1. Producing Data: collecting it as it arises, or conducting an experiment or random sample from a population
2. Exploratory Data Analysis: organising the data (computer science), producing summaries (mean, standard deviation etc)
3. Probability: understanding the chance mechanisms that generated the data:- random variables and their properties are important
4. Inference: using probability to draw inferences about the data: find estimates of population quantities, understand the uncertainty in the estimates, decide on hypotheses and predict

Probability - first half of our subject

Module 1: the basics for understanding random samples - important background:- sets, permutations and combinations, functions (1:1 and onto)

Module 2: discrete distributions and random variables - important background: series and the exponential function, differentiation

Module 3: continuous distributions and random variables - important background: integration

Module 4: distributions for two random variables, correlation - important background: two variable calculus

Module 5: transformations of random variables and limits - important background: limits, multivariable calculus

Statistical Inference - second half of our subject

Probability important because data = prob. model + residual is underlying idea of statistical inference

Module 6: data description and point estimation - important background:- Modules 1 to 5

Module 7: quantifying uncertainty through interval estimation - important background: Modules 1 to 5

Module 8: tests of hypotheses - important background: Modules 1 to 5

Module 9: special cases including regression, analysis of variance and contingency tables - important background: Modules 1 to 5