# Contents

1 Introduction 1

2 Background 5
   2.1 Terminology 5
   2.2 Fundamentals of Visualization 8
      2.2.1 Visual Marks and Visual Channels 11
      2.2.2 Visualizing High-Dimensional Data 13
   2.3 Visualizing Multivariate Data 17
   2.4 Visualizing Multivariate Functions 20
   2.5 Visualizing Probability Distributions 22

3 Visual Encoding of Multivariate Probability Distributions and Data 27
   3.1 Basic Definitions 29
   3.2 Facets: Derived Visualizations 30
      3.2.1 Density Facet 30
      3.2.2 Marginal Facet 35
      3.2.3 Data/Samples Facet 35
      3.2.4 Aggregation Facet 37
   3.3 Aligning Data and Model Operations 39
   3.4 Aligning Data and Model Visualizations 40
   3.5 Combining Facets 41
   3.6 Additional Visual Channels: Color, Shape, Size 42
   3.7 Design Choices for Visual Encoding 44
      3.7.1 Challenges 45
      3.7.2 Design Goals 45
      3.7.3 Design Decisions 45

4 Lumen: A Visual Interface for Probabilistic Models and Data 49
   4.1 A Life Cycle for Probabilistic Models 50
   4.2 Visual Analytics System Requirements 54
   4.3 The Visual Interface 55
      4.3.1 Schema Panel 55
      4.3.2 Specification Panel 56
      4.3.3 Dashboard 57
      4.3.4 Basic Workflow 58
   4.4 Step-by-Step Example 58
   4.5 Design Considerations 60

5 Applications 63
   5.1 Using Lumen for Data-driven Probabilistic Modeling 64
      5.1.1 CG Distributions 65
      5.1.2 Sum-Product Networks 66
      5.1.3 Simple Density Estimator 70
      5.1.4 Exploration of a CG Model 71
      5.1.5 Modeling the Penguin Data Set 72