Introduction to Case Study 3

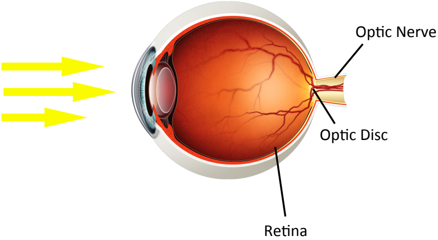
Samuel I. Berchuck
STA 440L, Duke University

February 28, 2023

Outline

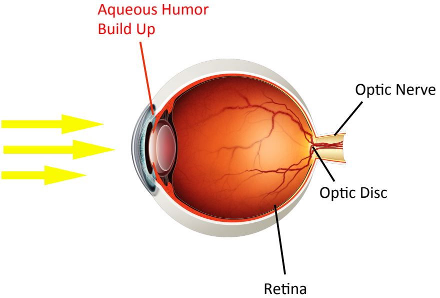
1. Glaucoma Background
   - Glaucoma and Disease Progression
   - Visual Field Data Description
2. Overview of Spatial Statistics
   - Types of Spatial Data
   - Why are Spatial Models Needed?
3. Discussion of the Case Study Goals and Expectations

Primary Open-Angle Glaucoma



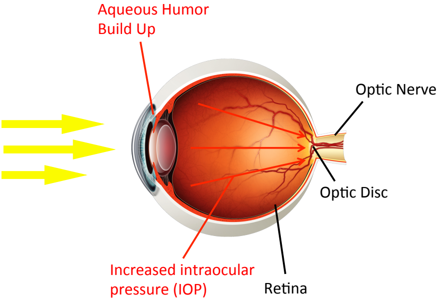
- Glaucoma is the most common cause of irreversible vision loss worldwide
- Disease that damages the eyes optic nerve
- Fluid builds in the eye leading to increased pressure
- Increased pressure causes damage to the optic nerve, resulting in vision loss
- No symptoms until vision loss occurs!

Primary Open-Angle Glaucoma



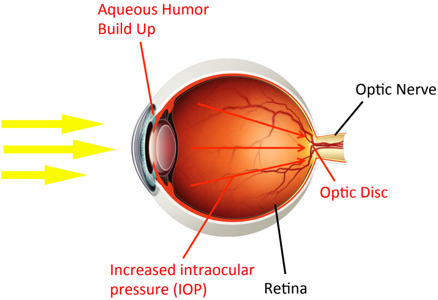
- Glaucoma is the most common cause of irreversible vision loss worldwide
- Disease that damages the eyes optic nerve
- Fluid builds in the eye leading to increased pressure
- Increased pressure causes damage to the optic nerve, resulting in vision loss
- No symptoms until vision loss occurs!

Primary Open-Angle Glaucoma



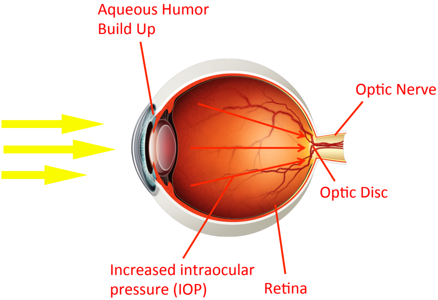
- Glaucoma is the most common cause of irreversible vision loss worldwide
- Disease that damages the eyes optic nerve
- Fluid builds in the eye leading to increased pressure
- Increased pressure causes damage to the optic nerve, resulting in vision loss
- No symptoms until vision loss occurs!

Primary Open-Angle Glaucoma



- Glaucoma is the most common cause of irreversible vision loss worldwide
- Disease that damages the eyes optic nerve
- Fluid builds in the eye leading to increased pressure
- Increased pressure causes damage to the optic nerve, resulting in vision loss
- No symptoms until vision loss occurs!

Primary Open-Angle Glaucoma



- Glaucoma is the most common cause of irreversible vision loss worldwide
- Disease that damages the eyes optic nerve
- Fluid builds in the eye leading to increased pressure
- Increased pressure causes damage to the optic nerve, resulting in vision loss
- No symptoms until vision loss occurs!

Glaucoma Progression

- Once a patient is diagnosed, clinicians must balance the risks and expenses of advancing levels of medical and surgical intervention with the risks of further vision loss due to **disease progression**
- Determining if the disease is progressing remains one of the most difficult tasks in the clinical setting

Methods to Detect Progression

Structural changes of the optic nerve head or retinal nerve fiber layer (RNFL) or **functional** changes in the visual field (VF)

Demonstrating the Visual Field

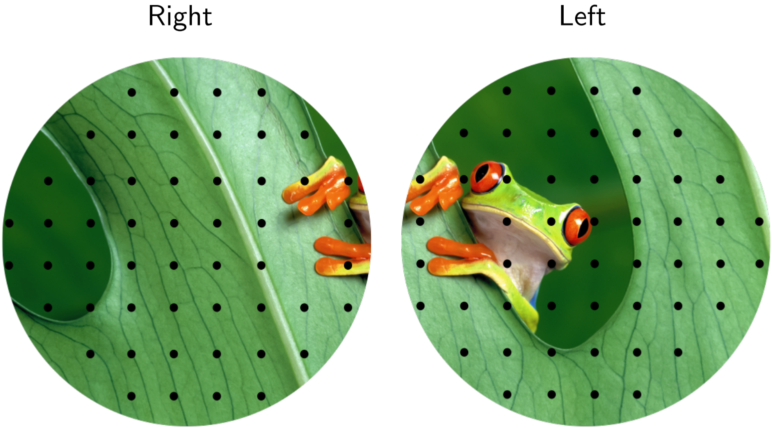
Visual field

Demonstrating the Visual Field

Right

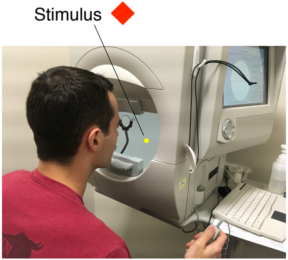
Left

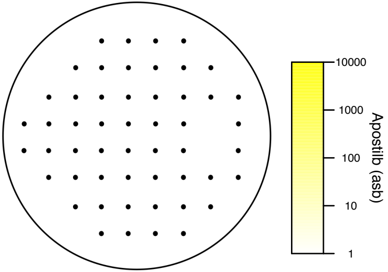
Demonstrating the Visual Field



Generating Visual Field Data

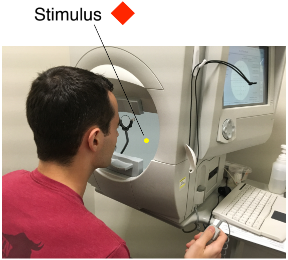
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - 1 ≈ Background (no constrast)
  - 10,000 = Bright (large contrast)

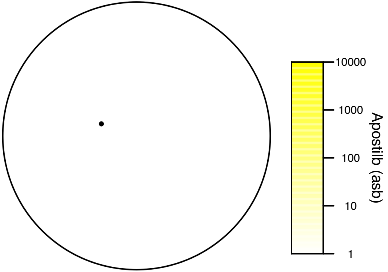




Generating Visual Field Data

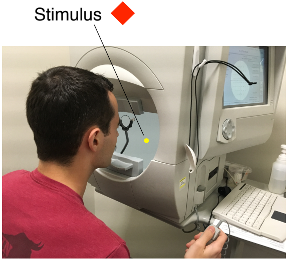
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - 1 ≈ Background (no constrast)
  - 10,000 = Bright (large contrast)

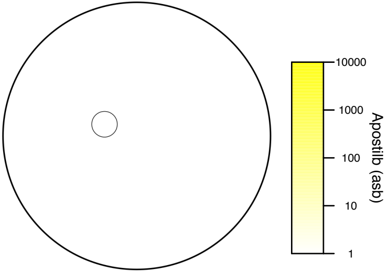




Generating Visual Field Data

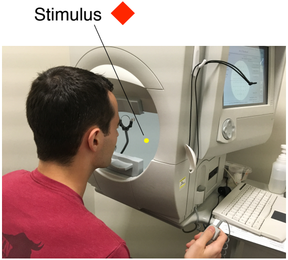
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - 1 ≈ Background (no constrast)
  - 10,000 = Bright (large contrast)

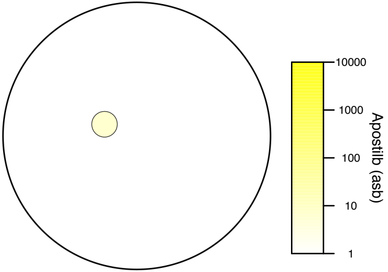




Generating Visual Field Data

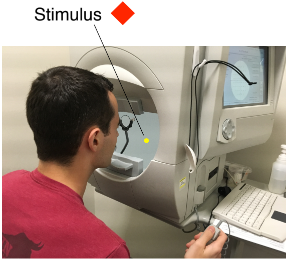
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - 1 ≈ Background (no constrast)
  - 10,000 = Bright (large contrast)

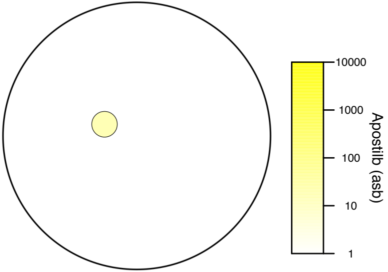




Generating Visual Field Data

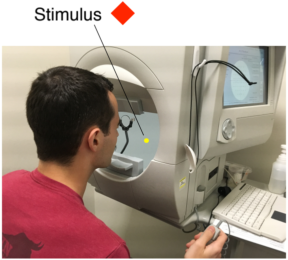
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - 1 $\approx$ Background (no constrast)
  - 10,000 = Bright (large contrast)

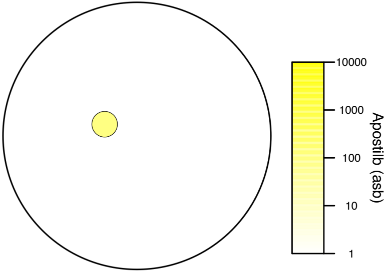




Generating Visual Field Data

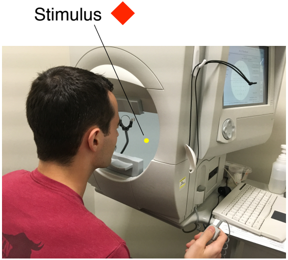
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - $1 \approx$ Background (no constrast)
  - $10{,}000 =$ Bright (large contrast)

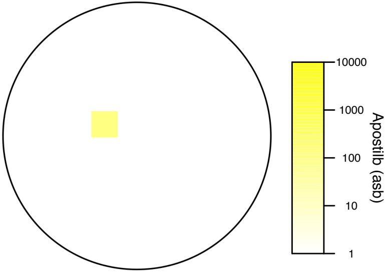




Generating Visual Field Data

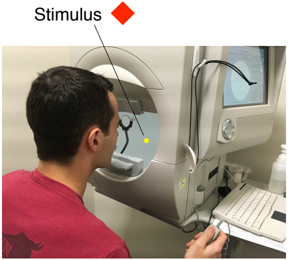
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - 1 $\approx$ Background (no constrast)
  - 10,000 = Bright (large contrast)

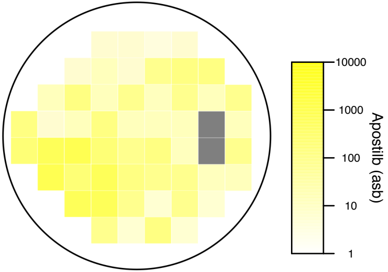




Generating Visual Field Data

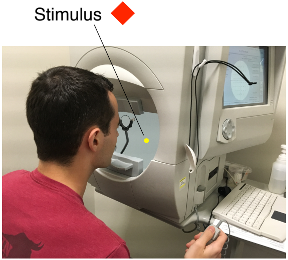
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - 1 ≈ Background (no constrast)
  - 10,000 = Bright (large contrast)

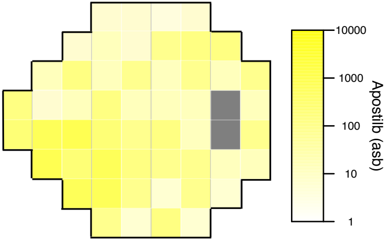




Generating Visual Field Data

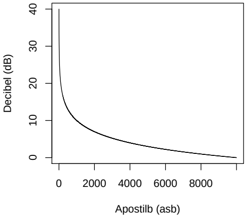
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - 1 $\approx$ Background (no constrast)
  - 10,000 $=$ Bright (large contrast)

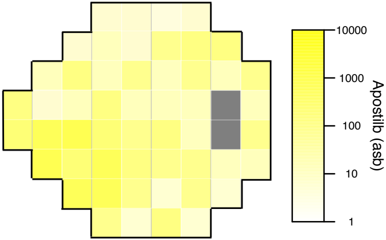




Generating Visual Field Data

- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
    - 1 $\approx$ Background (no constrast)
    - 10,000 = Bright (large contrast)

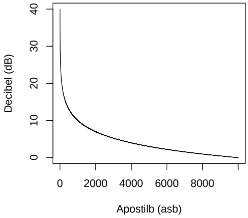


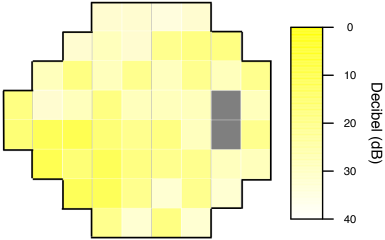


- Converting from Apostilbs to Decibels

Generating Visual Field Data

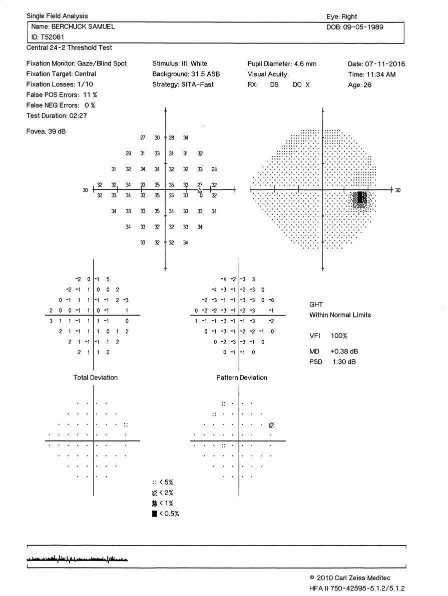
- Standard automated perimetry: Humphrey Field Analyzer-II
- Estimating differential light sensitivity (DLS) across the VF
- Intensity: measured in Apostilbs
  - $1 \approx$ Background (no constrast)
  - $10{,}000 =$ Bright (large contrast)



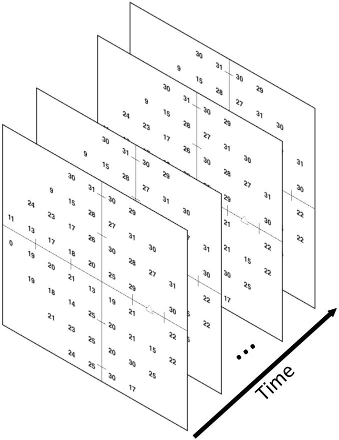


- Converting from Apostilbs to Decibels

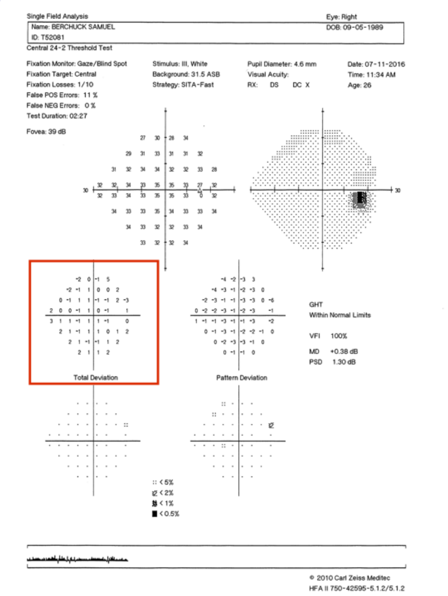
Generating Visual Field Data



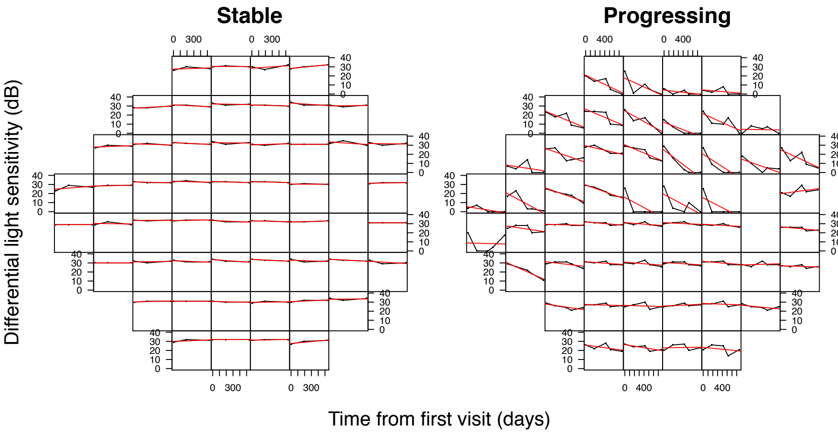
Generating Visual Field Data



Generating Visual Field Data



Generating Visual Field Data



Supplemental Reading

- Hierarchical Modeling and Analysis for Spatial Data (Banerjee, Carlin, and Gelfand)
  - Great reference for Bayesian modeling
- Statistics for Spatial Data (Cressie)
  - Frequentist version, classic and very thorough

Introduction

- Three Types of Spatial Data:

Introduction

- Three Types of Spatial Data:
  - Geostatistical Point Referenced Data

Introduction

- Three Types of Spatial Data:
  - Geostatistical Point Referenced Data
  - Lattice Data (Areal Data)

Introduction

- Three Types of Spatial Data:
  - Geostatistical Point Referenced Data
  - Lattice Data (Areal Data)
  - Spatial Point Process Data

Introduction

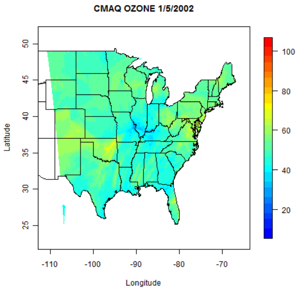
- Three Types of Spatial Data:
    - Geostatistical Point Referenced Data
    - Lattice Data (Areal Data)
    - Spatial Point Process Data
- All of these data settings can be extended to space-time

Introduction

- Three Types of Spatial Data:
  - Geostatistical Point Referenced Data
  - Lattice Data (Areal Data)
  - Spatial Point Process Data
- All of these data settings can be extended to space-time
- Our glaucoma application will require a space-time model

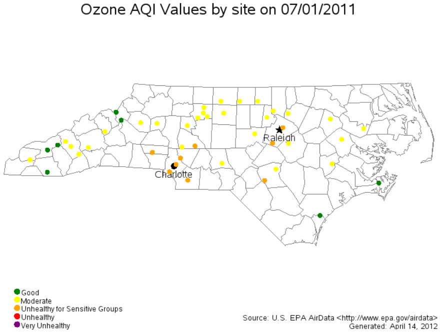
Spatial-Temporal Models

- Space-time modeling can be difficult computationally and conceptually (many options)



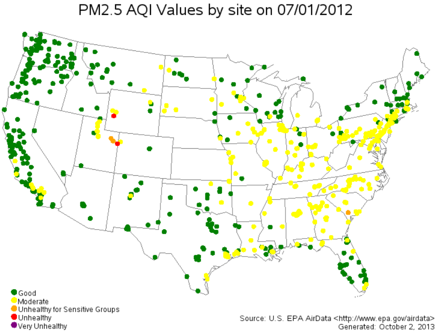
(1) Geostatistical Point Referenced Data

- Point observations of a continuously varying quantity over a region
  - Daily Concentrations of Ozone Over NC



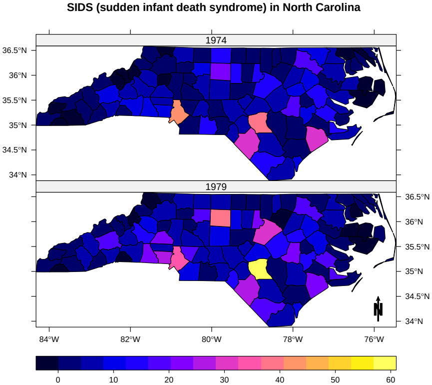
(1) Geostatistical Point Referenced Data

- Daily Concentrations of PM2.5 Over the US



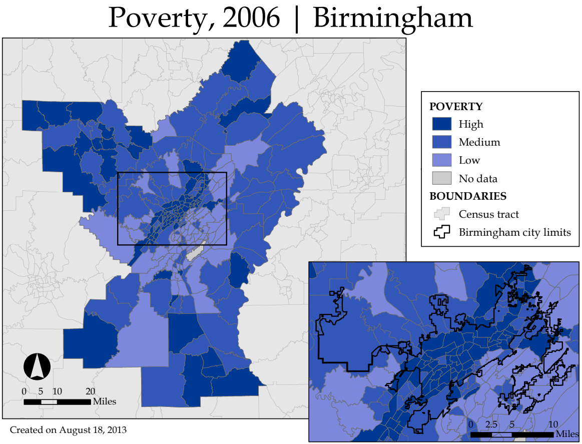
(2) Lattice Data (Areal Data)

- Data observed at the level of an areal unit
  - County Level Sudden Infant Death Syndrome Counts



(2) Lattice Data (Areal Data)

- Birmingham Tract Level Poverty Levels



(3) Spatial Point Process Data

- Analyzing the clustering of random locations

(3) Spatial Point Process Data

- Analyzing the clustering of random locations
  - Locations of a certain tree type in a forest

(3) Spatial Point Process Data

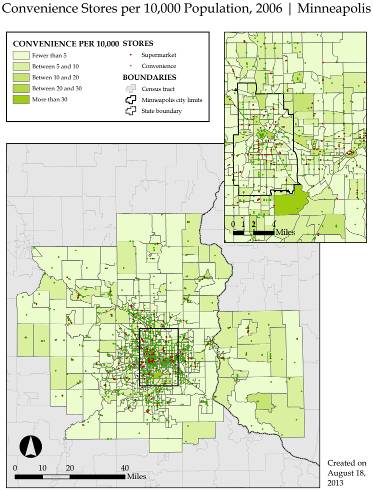
- Analyzing the clustering of random locations
  - Locations of a certain tree type in a forest
  - Epicenter of earthquakes

(3) Spatial Point Process Data

- Analyzing the clustering of random locations
  - Locations of a certain tree type in a forest
  - Epicenter of earthquakes
- Sometimes difficult to differentiate from point referenced geostatistical data (visually)

(3) Spatial Point Process Data

- Minneapolis Convenience Store Locations



Spatial Data Analysis: When, Why, and How?

- When?:

Spatial Data Analysis: When, Why, and How?

- When?:
  - Anytime you have spatial information collected with your data

Spatial Data Analysis: When, Why, and How?

- When?:
  - Anytime you have spatial information collected with your data
  - Increasing availability of some level of spatial information (latitude/longitude; county; state; etc.)

Spatial Data Analysis: When, Why, and How?

- When?:
  - Anytime you have spatial information collected with your data
  - Increasing availability of some level of spatial information (latitude/longitude; county; state; etc.)
- Why?:

Spatial Data Analysis: When, Why, and How?

- When?:
  - Anytime you have spatial information collected with your data
  - Increasing availability of some level of spatial information (latitude/longitude; county; state; etc.)
- Why?:
  - Correct statistical inference (conditional independence may not be a valid assumption!)

Spatial Data Analysis: When, Why, and How?

- When?:
  - Anytime you have spatial information collected with your data
  - Increasing availability of some level of spatial information (latitude/longitude; county; state; etc.)
- Why?:
  - Correct statistical inference (conditional independence may not be a valid assumption!)
  - Specific goals will depend on the type of spatial data you have and the objective of your analysis

Spatial Data Analysis: When, Why, and How?

- When?:
  - Anytime you have spatial information collected with your data
  - Increasing availability of some level of spatial information (latitude/longitude; county; state; etc.)
- Why?:
  - Correct statistical inference (conditional independence may not be a valid assumption!)
  - Specific goals will depend on the type of spatial data you have and the objective of your analysis
  - Producing maps with valid inference

Spatial Data Analysis: When, Why, and How?

- How?:

Spatial Data Analysis: When, Why, and How?

- How?:
  - Bayesian Hierarchical Modeling:

Spatial Data Analysis: When, Why, and How?

- How?:
  - Bayesian Hierarchical Modeling:
    - Flexible framework to handle multiple levels of uncertainty

Spatial Data Analysis: When, Why, and How?

- How?:
  - Bayesian Hierarchical Modeling:
    - Flexible framework to handle multiple levels of uncertainty
    - Markov chain Monte Carlo (MCMC) offers computationally convenient solution to make inference

Spatial Data Analysis: When, Why, and How?

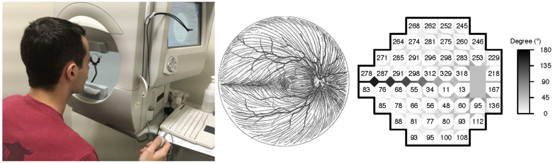
- How?:
  - Bayesian Hierarchical Modeling:
    - Flexible framework to handle multiple levels of uncertainty
    - Markov chain Monte Carlo (MCMC) offers computationally convenient solution to make inference
  - Frequentist methods also available through the EM algorithm

Spatial Data Analysis: When, Why, and How?

- How?:
    - Bayesian Hierarchical Modeling:
        - Flexible framework to handle multiple levels of uncertainty
        - Markov chain Monte Carlo (MCMC) offers computationally convenient solution to make inference
    - Frequentist methods also available through the EM algorithm
        - Original frequentist methods ignore some of the uncertainty in estimating these spatial models

Learning Objectives

Case Study 3: Disease progression in glaucoma

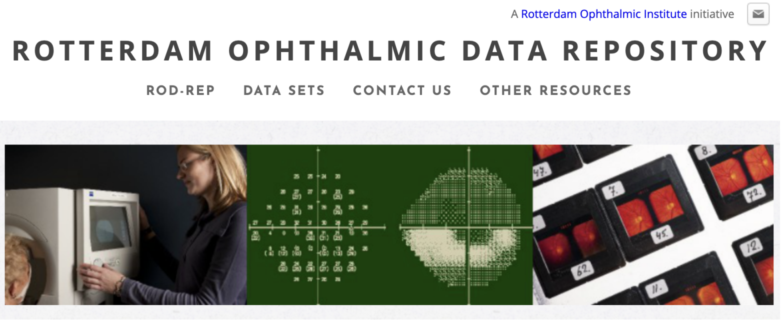


- Fit a spatial model to complex spatiotemporal data
- Gain proficiency handling datasets that have both longitudinal and spatial structure
- Learn the difference between labeled and unlabeled longitudinal data
- Translate an important clinical question into a valid research question
- Interpret model results in clear language accessible to general readers

Case Study Goals

1. Develop a statistical model that can be used to predict a future visual field for a patient given an existing longitudinal series of visual fields for that patient.
2. Determine whether using a spatial model that accounts for the neighborhood dependencies of the visual field improves the prediction performance.
3. Quantify the heterogeneity in prediction performance across regions of the visual field.

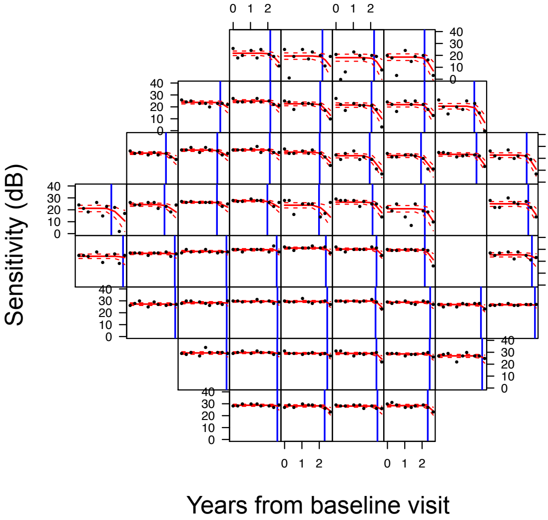
Data

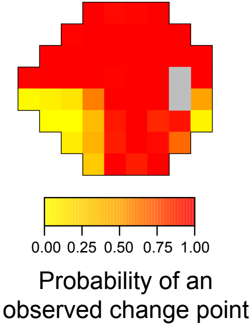


- Data Set: Longitudinal Glaucomatous Visual Field data
- Eye Level Data: intraocular pressure (IOP), mean deviation (MD), age, and sex
- Visual Field Level Data: threshold sensitivities, total deviation at 54 locations

Goal 1: Labeled and unlabeled longitudinal data

Why are we focusing on prediction?





Goal 1: Prediction evaluation

In the absence of a gold standard assessment of disease progression, we will use prediction to evaluate our model performance. For example...

- Train your models using the first 5 visual fields and assess prediction in the 10th visual field.
- Train your models using the visual fields in the first 2 years to predict a visual field at 5 years.
- Whatever you choose, be consistent and motivate it statistically and clinically.
- Summarize prediction across the visual field globally and regionally.

Note: Models are typically defined for each series of visual fields, so you would fit a unique model for each eye.

Goal 2: Baseline Comparison

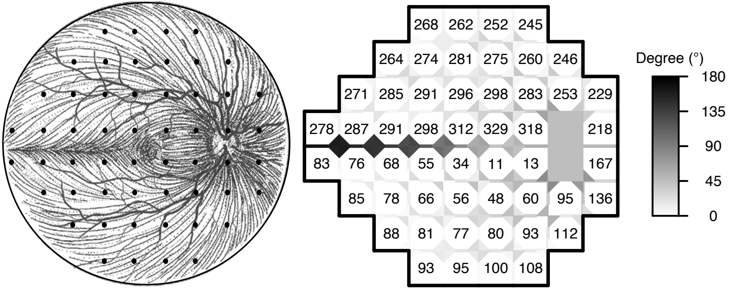
To demonstrate that the spatial model has utility, you will want to compare prediction performance to a non-spatial model. For example...

Pointwise linear regression

Believe it or not, but pointwise linear regression (PLR) is commonly used as a baseline model for prediction of visual fields in glaucoma. Just like it sounds, PLR performs ordinary least squares regression independently at each location on the visual field.

You are welcome to define a more interesting baseline model.

Goal 3: Regions of the visual field



- **Recall:** Glaucoma damages the optic disc, so VF deterioration corresponds to underlying fibers that enter the damaged regions
- Visual Field/RNFL connection: angle that each test location's RNFL fibers enter the optic disc

Goal 3: Regions of the visual field

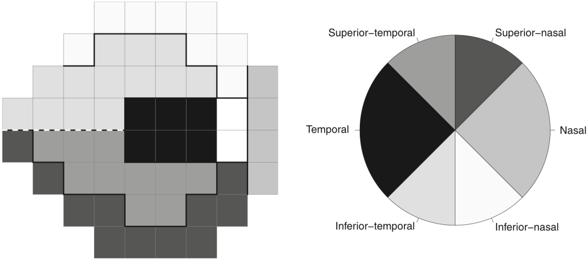


FIGURE 1. Visual field regions mapped to the optic disc.[16]

- **Recall:** Glaucoma damages the optic disc, so VF deterioration corresponds to underlying fibers that enter the damaged regions
- Visual Field/RNFL connection: angle that each test location's RNFL fibers enter the optic disc

Data suggestions

- Convert all eyes to right eyes.
- Consider converting all data to be consistent with `womblR`

Some useful software

R packages developed by me using visual field data: `womblR`, `spCP`, `spBFA`.

R packages for areal data: `CARBayes`, `CARBayesST`.

R packages for point-referenced data: `spBayes`.

Stan and `brms` can easily accommodate Bayesian spatial models.

- For example: http://paul-buerkner.github.io/brms/reference/car.html, https://mc-stan.org/users/documentation/case-studies/icar_stan.html

Spatial Statistics

- The study of spatially referenced data observations
- Types of spatial data
  1. Geostatistical (or point-referenced)
  2. Areal (or lattice)
  3. Point-process

- The foundational assumption in spatial statistics states that dependence between observations weakens as the distance between locations increases