Environmental
& Statistical
Consultants

Day 2
Distance Sampling

Statistical Methods for Estimating Abundance in Ecology

Jason Carlisle – WEST, Inc.

Virtual Workshop Sponsored by Wyoming EPSCoR
12/1/2021

Outline of Topics

1) Distance Sampling in General
2) Crash Course in Hierarchical Models
3) The Hierarchical Distance Sampling (HDS) Model
4) Example HDS Analysis

1) Distance Sampling in General

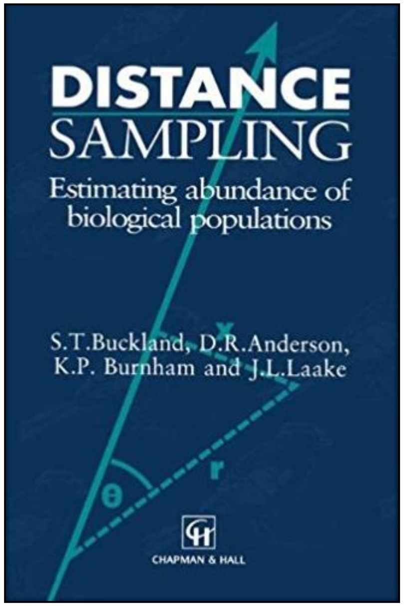
Distance Sampling

- Method to estimate the abundance (or density) of biological populations that accounts for imperfect detection of individuals

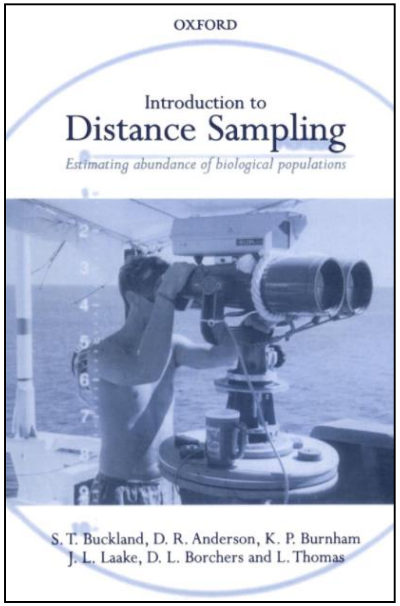
- Pop quiz:
  - Why use it – why not just count things directly?
  - Why the name?

Resources on Distance Sampling

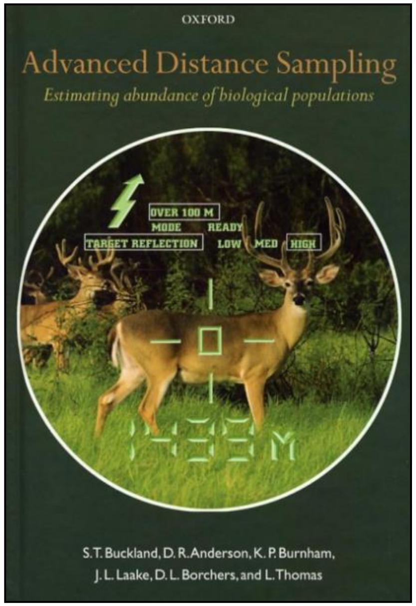
Top recommendation



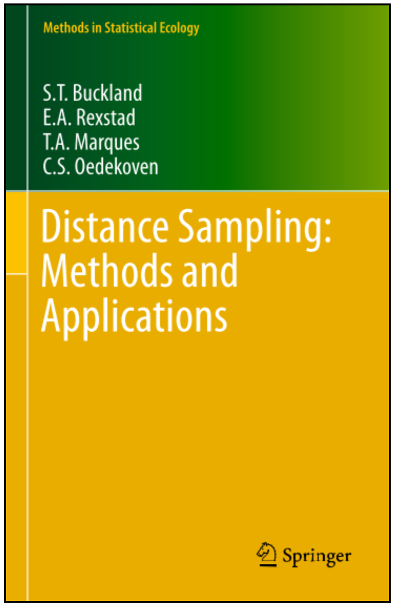
(Buckland et al. 1993)



(Buckland et al. 2001)



(Buckland et al. 2004)



(Buckland et al. 2015)

How to Estimate Density and Abundance

When probability of detection is 100%

- Don’t need distance sampling (or other methods) if the probability of detection (p) = 100%, but this scenario helps one understand how distance sampling accounts for imperfect detection

How to Estimate Density and Abundance

When probability of detection is 100%

- True abundance (N) is same as observed count

$$N = Count$$

- True density (D) is easy to calculate if you know the size of the area surveyed

$$D = \frac{Count}{Area}$$

- Area surveyed for (straight) transects is just a rectangle

$$Area = Width * Length$$

How to Estimate Density and Abundance

When probability of detection is 100%

$$\widehat{D} = \frac{Count}{2 * SearchDistance * TotalTransectLength}$$

Pop Quiz

When probability of detection is 100%

- Assume you surveyed 5 m on both sides of 100 transects that were each 1,000 m long, and you counted 250 lizards.
- Assume p = 100%

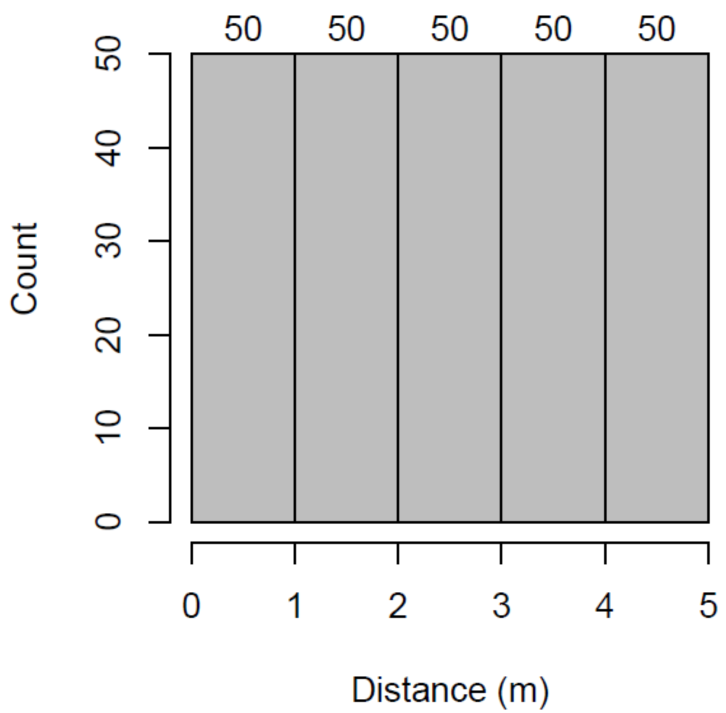
- What is the estimated density of horned lizards per hectare (ha)?
- Hint: 1 ha = 10,000 m^2

Answer?

When probability of detection is 100%

Segue

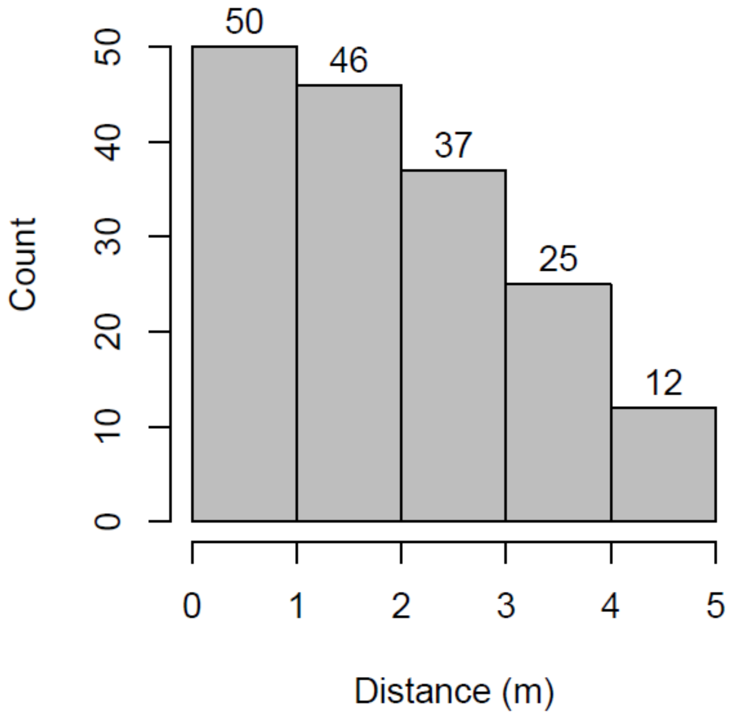
- If p =100%, under the assumptions of line transect sampling (i.e., random placement of transects), we would obtain (on average) a distance histogram like this:



- Bars same height
- The 250 lizards counted were distributed equally across distance bins
  - Pop quiz: what distance is being measured here?

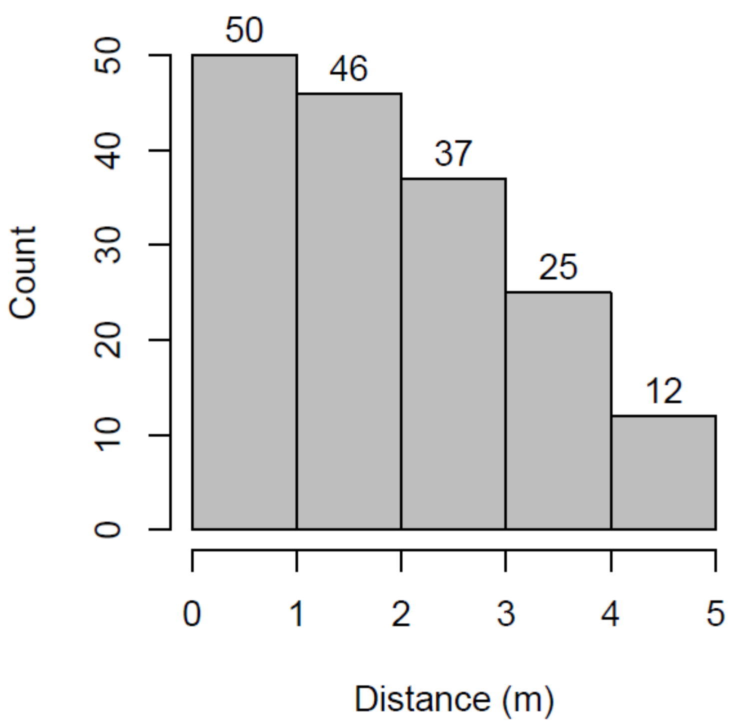
Segue

- But what if you repeated the same survey, but now only counted 170 lizards, and your histogram looked like this?



Segue

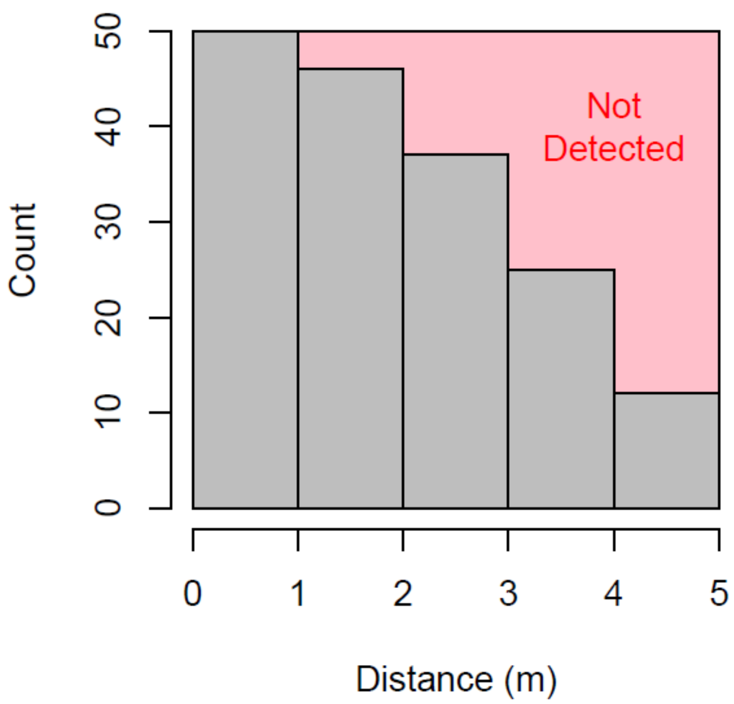
- But what if you repeated the same survey, but now only counted 170 lizards, and your histogram looked like this?



- Lower counts at farther distances is a sign that $p < 100\%$
- Specifically that detectability decreases as distance increases

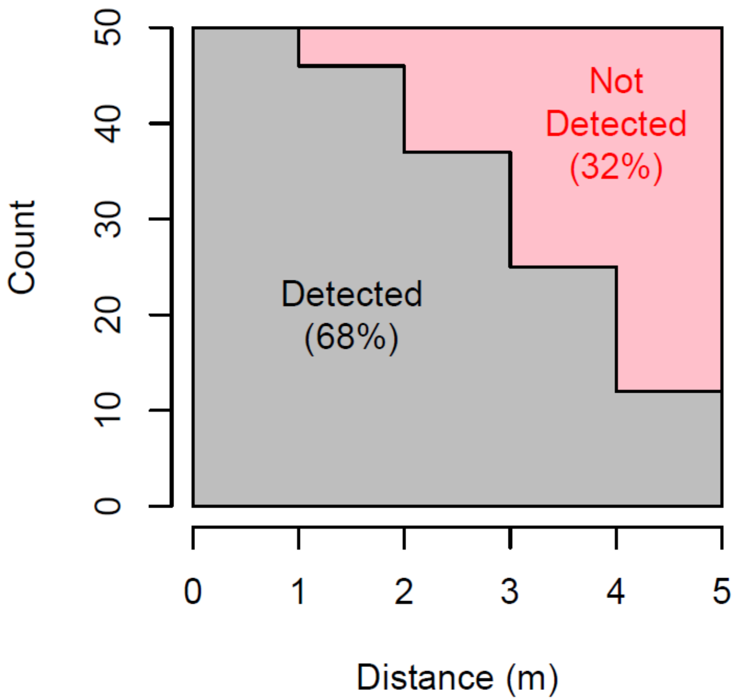
Segue

- Some proportion of lizards were not detected
- But what proportion?



Segue

- Some proportion of lizards were not detected
- But what proportion?



- We know true N is 250
- We counted 170
- 170/250 = 68% detected
- 1 – 0.68 = 32% missed

How to Estimate Density and Abundance

When probability of detection < 100%

- Need distance sampling (or other method) to account for imperfect detection when estimating density or abundance

How to Estimate Density and Abundance

When probability of detection < 100%

- True abundance (N) estimated as

$$N = \frac{Count}{p}$$

- True density (D) is easy to calculate if you know the size of the area surveyed

$$D = \frac{\left(\frac{Count}{p}\right)}{Area} = \frac{Count}{p * Area}$$

- Area surveyed for (straight) transects is just a rectangle

$$Area = Width * Length$$

How to Estimate Density and Abundance

When probability of detection < 100%

$$\hat{D} = \frac{\left(\frac{Count}{p}\right)}{2 * SearchDistance * TotalTransectLength}$$

equivalent to

$$\hat{D} = \frac{Count}{2 * p * SearchDistance * TotalTransectLength}$$

Pop Quiz

When probability of detection < 100%

- Assume you surveyed 5 m on both sides of 100 transects that were each 1,000 m long, and you counted 170 lizards.
- Assume $p = 68\%$

- What is the estimated density of horned lizards per hectare?

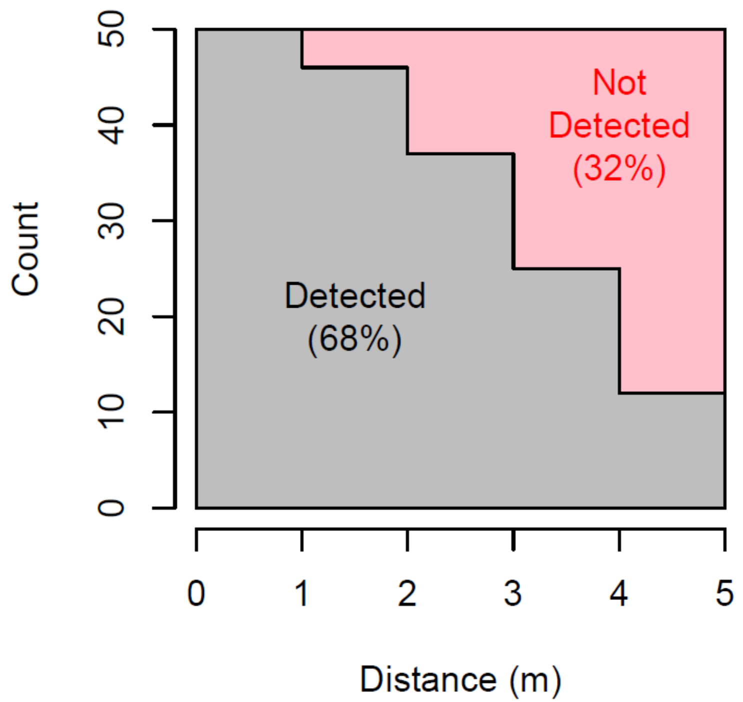
Answer?

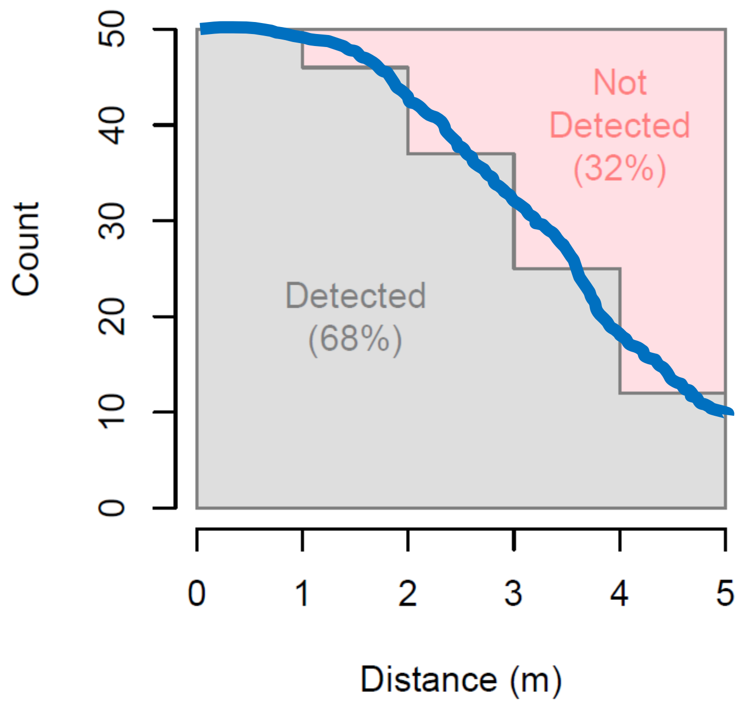
When probability of detection < 100%

Summary

- The shape of the distance histogram is used to estimate the probability of detection

- Typically, we fit a model to describe the shape (i.e., the detection function)

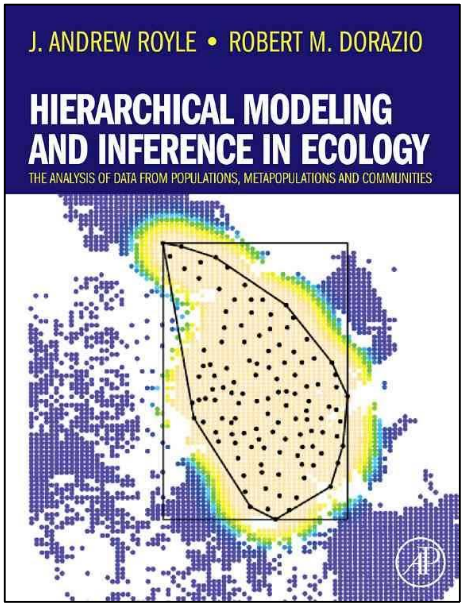




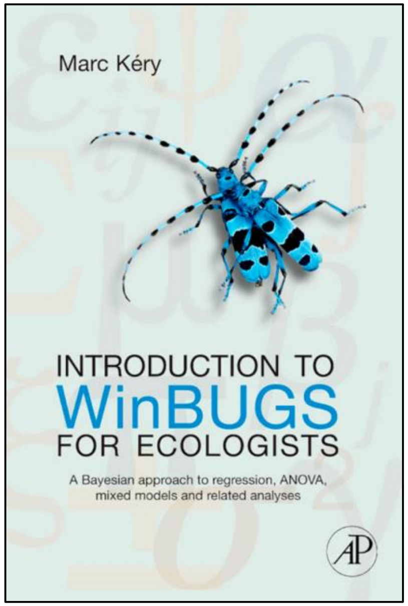
2) Crash Course in Hierarchical Models

Resources on Hierarchical Models

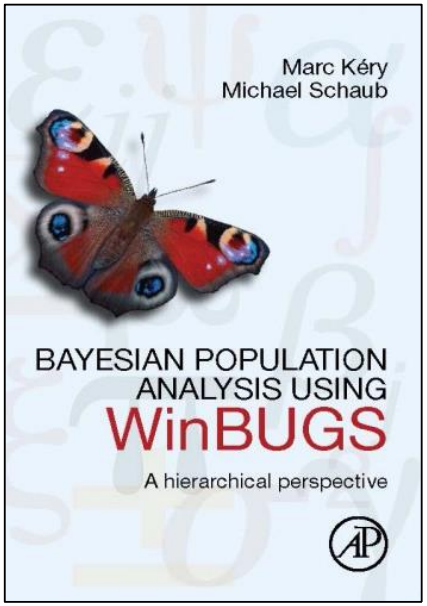
Top recommendation



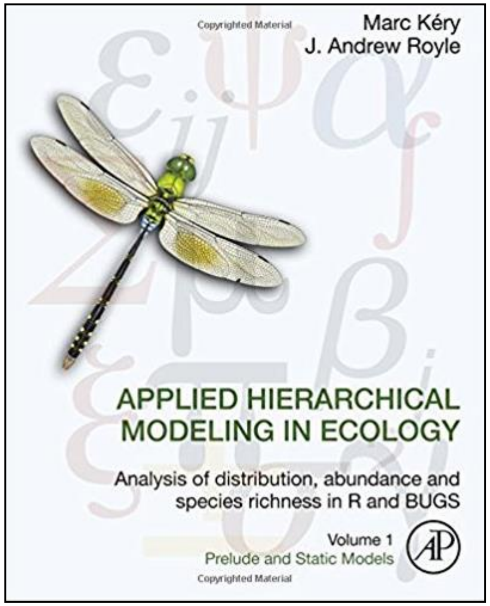
(Royle and Dorazio 2008)



(Kéry 2010)



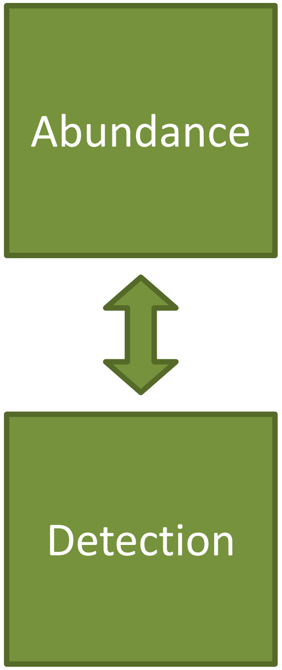
(Kéry and Schaub 2012)



(Kéry and Royle 2016)

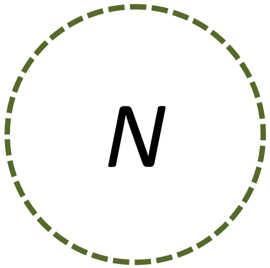
Hierarchical Models, Defined

- A sequence of probability models that are ordered by their conditional probability structure
  - i.e., they describe conditionally dependent random variables
- Describe the true state of nature that is not observable (or only partly so) and also describe the measurement error



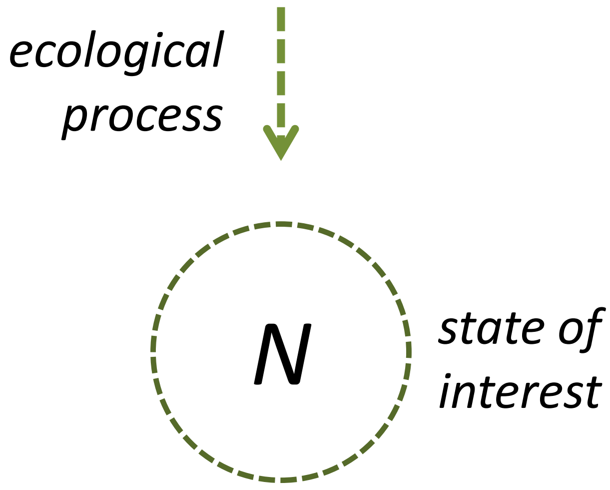
(Kéry and Schaub 2012, Kéry and Royle 2016)

Ecology and Hierarchical Models

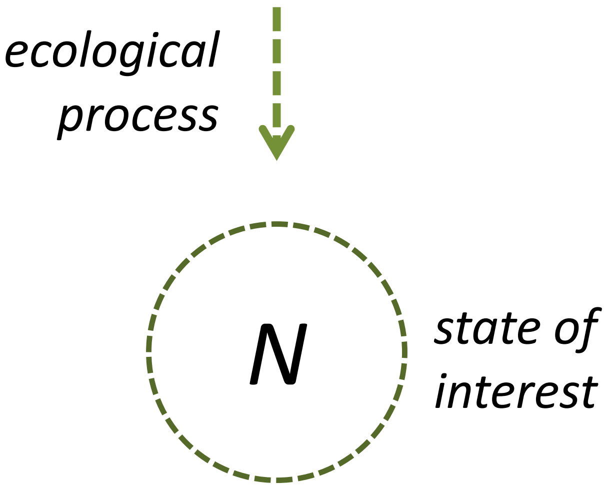


- Ecology is primarily concerned with ecological states
  - e.g., abundance (*N*)
- Describing ecological states (or why they change in space or time) is central to ecological research and natural resource management

(adapted from Kéry and Schaub 2012)

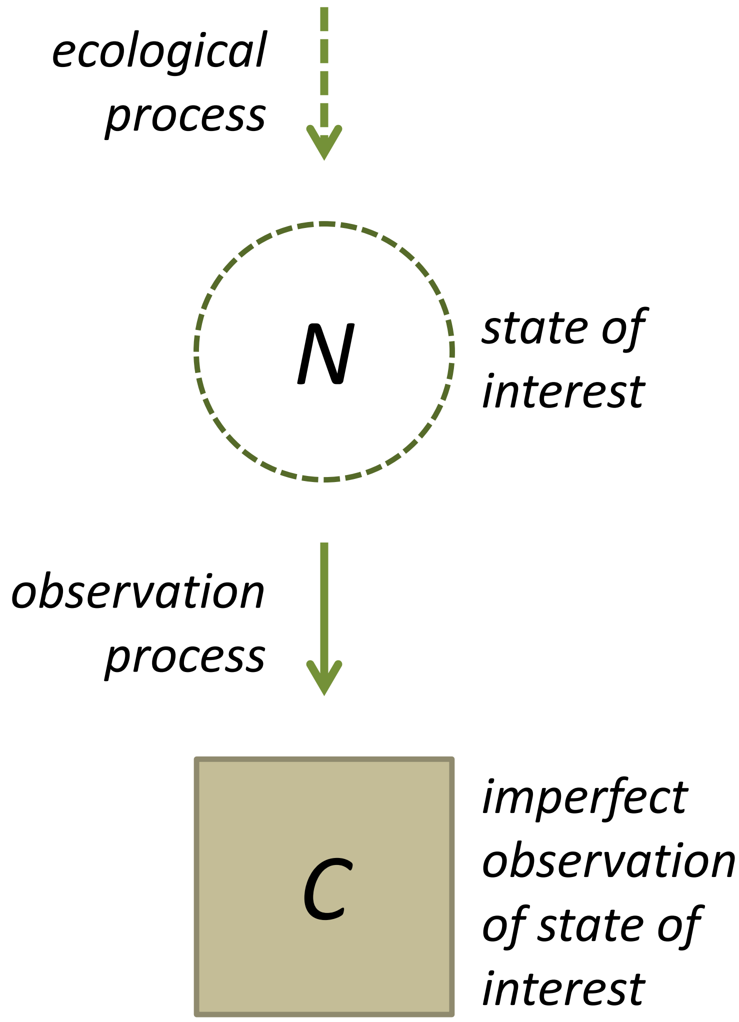


N is the result of an ecological process



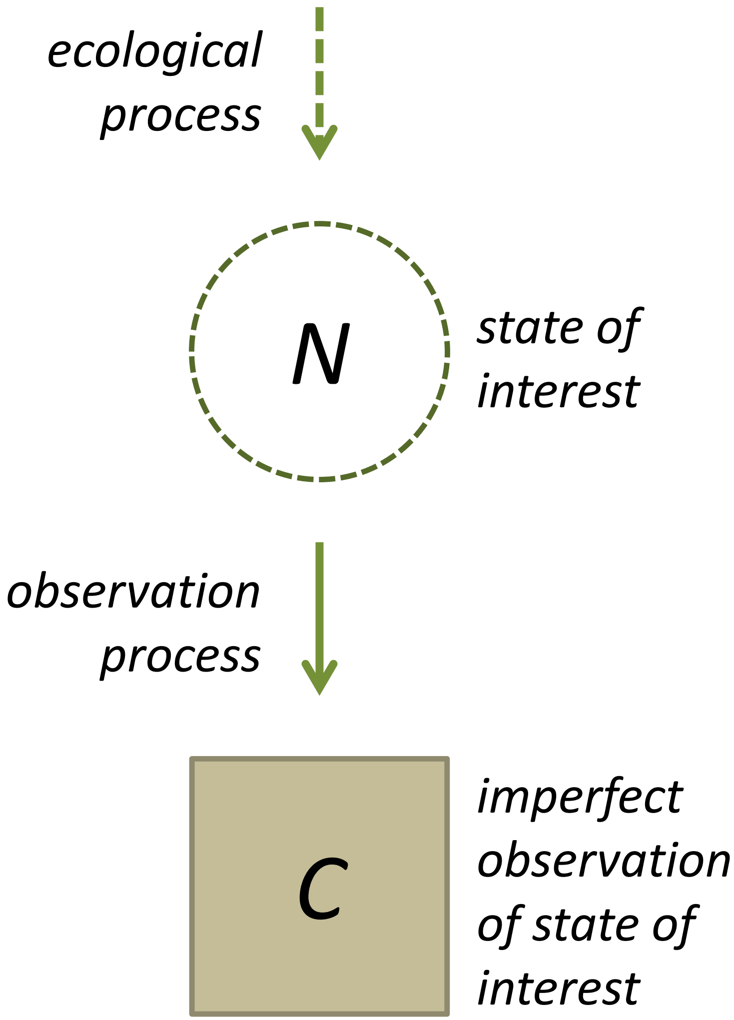
N is the result of an ecological process

Problem:
N is not directly observable
(*N* is latent)



N is the result of an ecological process

Solution:
Try to observe *N* anyways, resulting in counts (*C*)
$C \neq N$



N is the result of an ecological process

C is the result of an ecological process <u>AND an error-prone observational process</u>

- Ecologists need the following to make inference on *N*, when all they really have is *C*

  - Special data collection designs

  - Special models

(Kéry and Schaub 2012)

- Ecologists need the following to make inference on *N*, when all they really have is *C*
  - Special data collection designs → recording the distance to each individual in *C* (aka distance-sampling)
  - Special models → a hierarchical model that accounts for the two processes that gave rise to *C:* ecological and observational

(Kéry and Schaub 2012)

3) The Hierarchical Distance Sampling Model

Resources on Hierarchical Distance Sampling

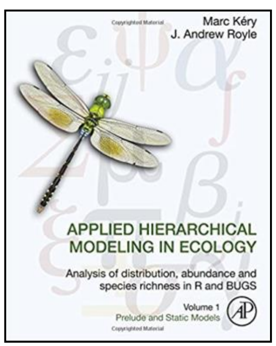
- Seminal paper: Royle et al. 2004

MODELING ABUNDANCE EFFECTS IN DISTANCE SAMPLING

J. Andrew Royle,[1,4] Deanna K. Dawson,[2] and Scott Bates[3]

[1]*U.S. Fish and Wildlife Service, Division of Migratory Bird Management, 11510 American Holly Drive, Laurel, Maryland 20708 USA*
[2]*USGS Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, Maryland 20708 USA*
[3]*National Park Service, Center for Urban Ecology, 4598 MacArthur Boulevard, NW, Washington, D.C. 20007-4227 USA*

- Thorough coverage: Ch. 8 in Kéry and Royle 2016



CHAPTER 8

MODELING ABUNDANCE USING HIERARCHICAL DISTANCE SAMPLING

Hierarchical Distance Sampling (HDS)

- One type of multinomial N-mixture model
- Straightforward to implement in R package *unmarked*

Session	Model	Detection Process	Abundance Process	*unmarked* Function
1	(Binomial) N-Mixture	Binomial	Poisson (or other)	*pcount*
2	Distance Sampling	Multinomial	Poisson (or other)	*distsamp*

(Royle et al. 2004, Fiske and Chandler 2011)

HDS Data Structure

- Counts are binned by distance
- For each of (I) sites (transects), we have observed counts in each of (J) distance classes (here, 3). And optional site-level covariate(s).

Site	Dist Class 1 (0 – 20 m)	Dist Class 2 (20 – 40 m)	Dist Class 3 (40 – 60 m)	…	Covariate
Transect 1	2	1	0	…	v_1
Transect 2	3	0	1	…	v_2
Transect 3	1	1	0	…	v_3
…	…	…	…	…	…
Transect I	2	1	1	…	v_I

The HDS Process Model (Abundance)

- The process model is a Poisson GLM

Latent, transect-level abundance (N) at site i is a random variable from the Poisson distribution:

$$N_i \sim Poisson(\lambda_i)$$

And covariates (v) may influence expected abundance (λ):

$$\log(\lambda_i) = \beta_0 + \beta_1 v_i$$

The HDS Observation Model (Detection)

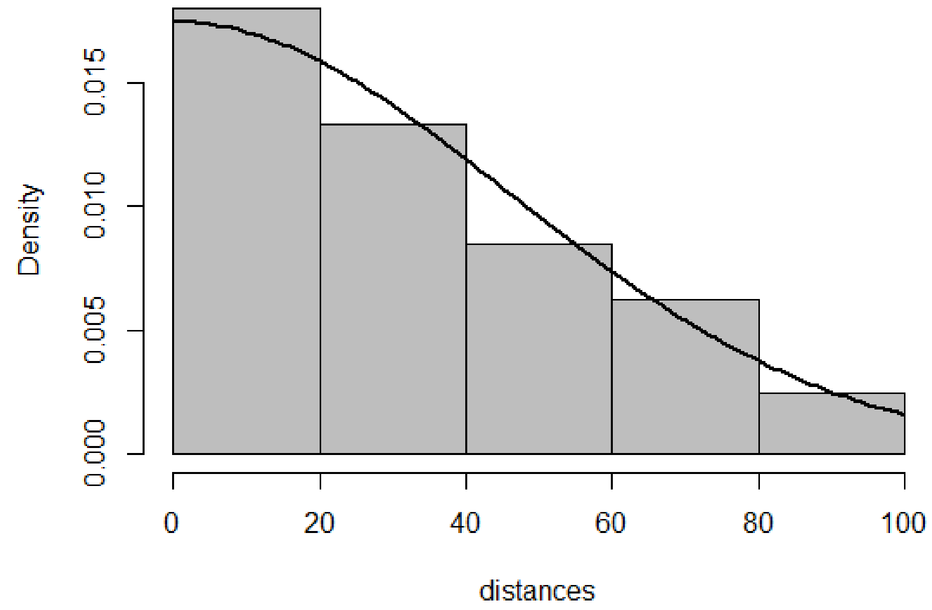
The observed count (C_{ij}) in each of (J) distance classes are conditional on the population size (N_i), and have a multinomial distribution:

$$C_{ij}\,|N_i \sim Multinomial(N_i, \pi_{ij})$$

	Dist Class 1 (0 – 20 m)	Dist Class 2 (20 – 40 m)	Dist Class 3 (40 – 60 m)
Probability (π_i)	60%	30%	10%

The HDS Observation Model (Detection)

A detection function (here, half-normal) is fit over the bins

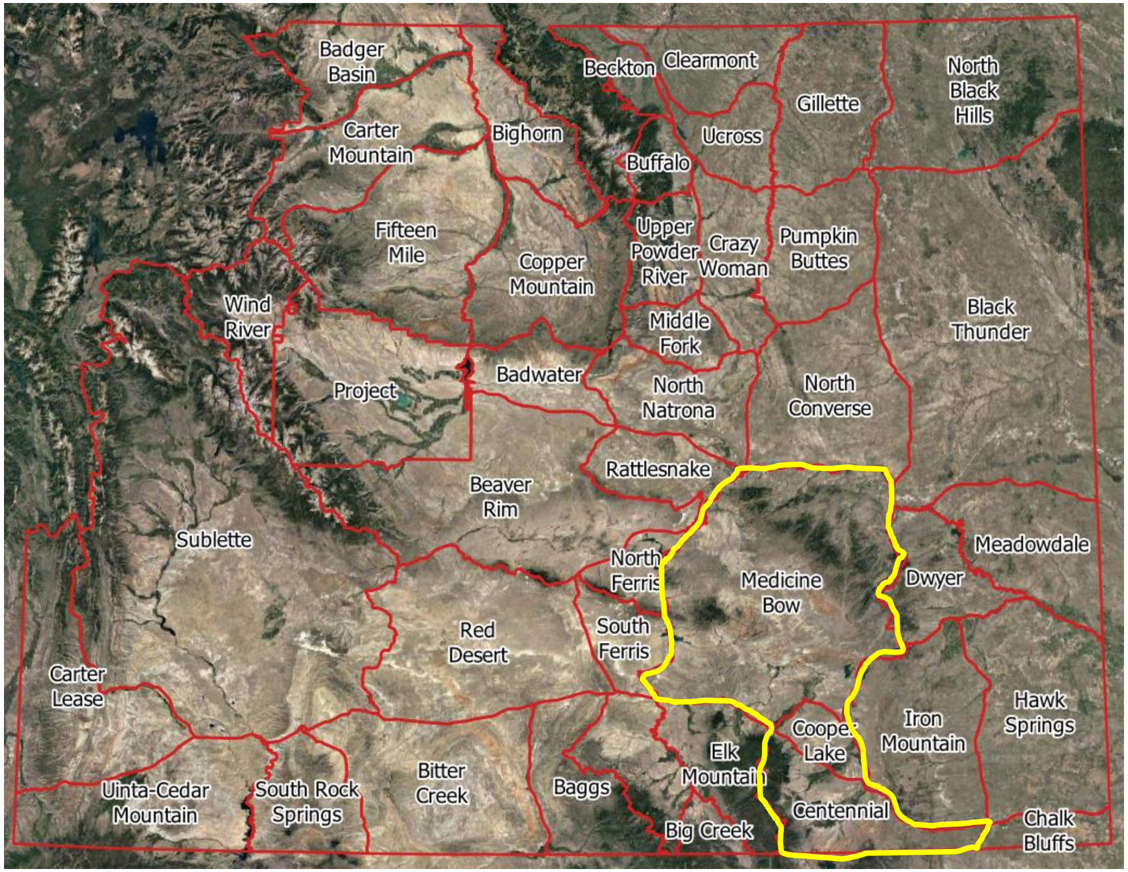


As before, covariates (v) may influence the shape of the detection function fit over distance bins (J)

$$\log(\sigma_i) = \alpha_0 + \alpha_1 v_i$$

4) Example Hierarchical Distance Sampling Analysis

Pronghorn Aerial Line Transect Surveys

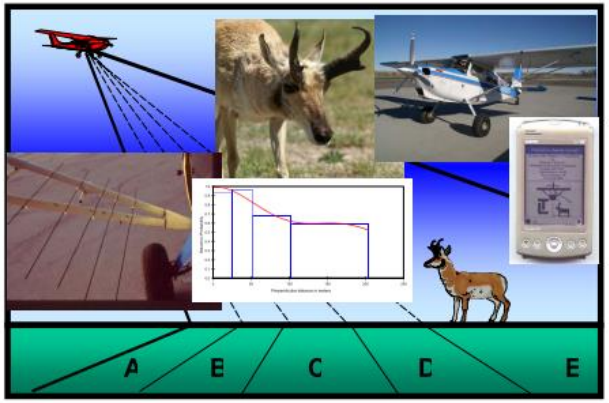


DRAFT

Procedures for Estimating Pronghorn Abundance in Wyoming Using Aerial Line Transect Sampling

By

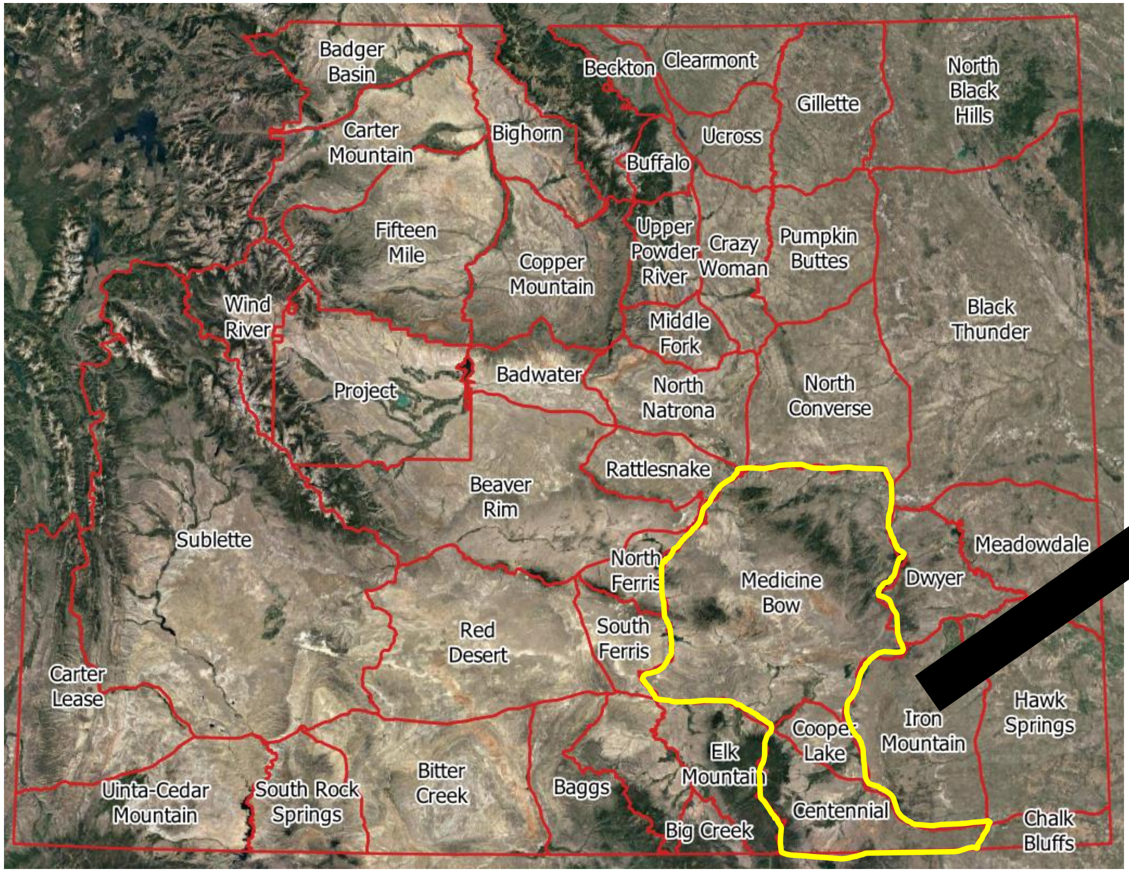
Richard J. Guenzel

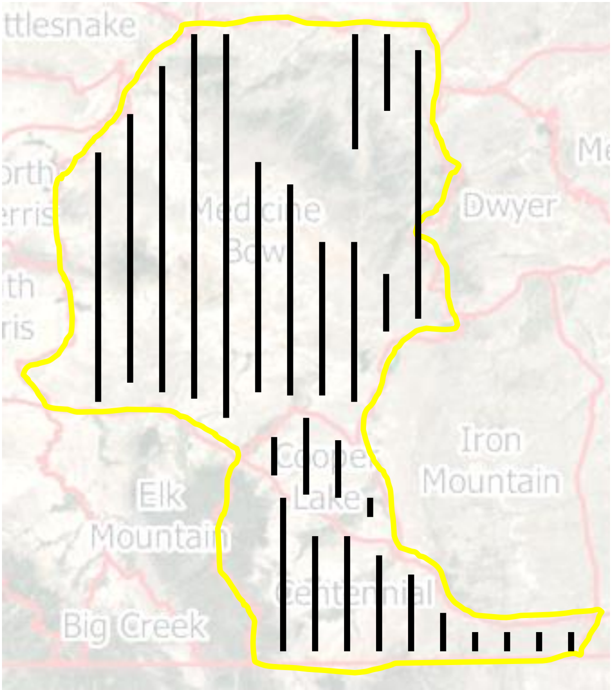


Wyoming Game and Fish Department
5400 Bishop Boulevard
Cheyenne, WY 82006
USA

2007

Pronghorn Aerial Line Transect Surveys



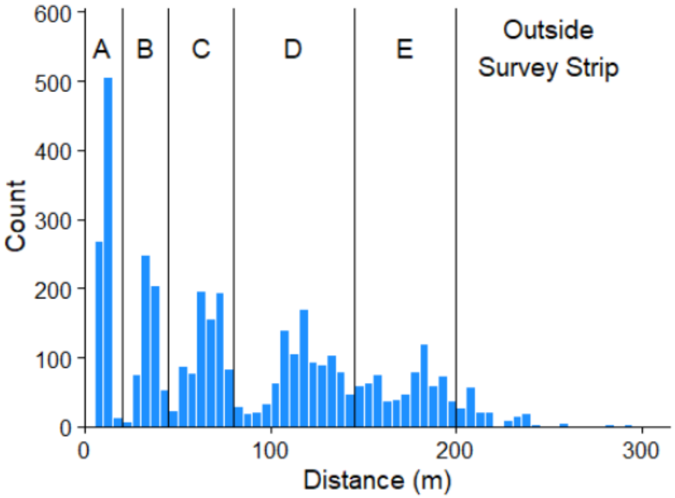


Systematic sample of North-South transects in each herd unit

Example Data: Pronghorn

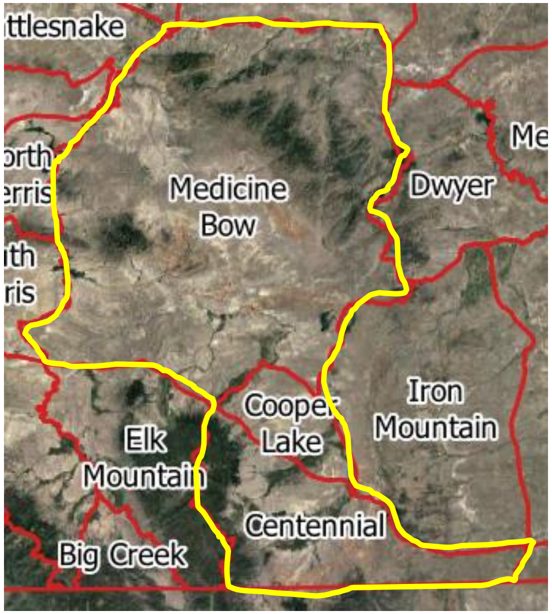
- `pronghornSiteData`
  - 186 sites
  - Lengths vary (2 – 46 km)
  - Only 1 side of transect surveyed
  - 1 covariate (herd unit)

- `pronghornDetectionData`
  - 4,110 detected individuals

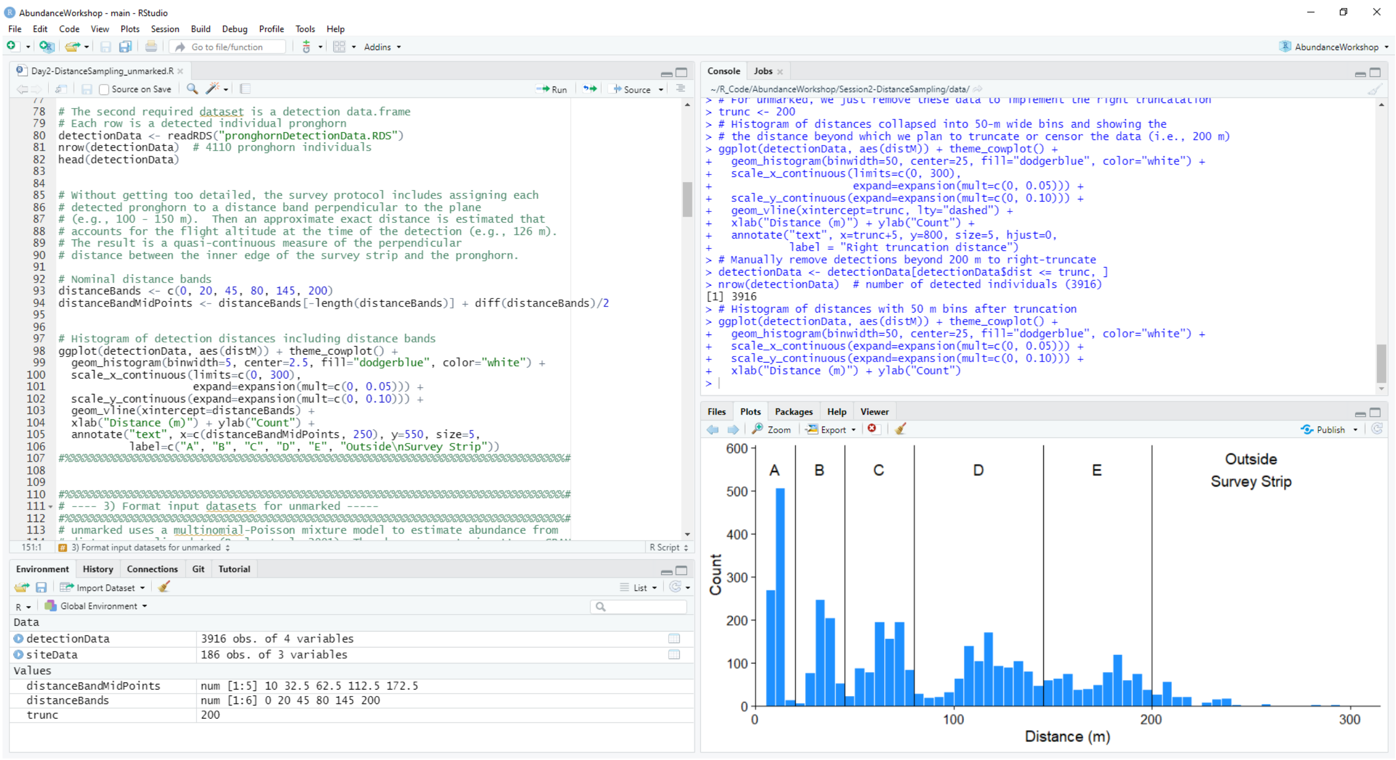


Research Question

- Which pronghorn herd in southeast Wyoming has the highest population density?
  - Medicine Bow
  - Cooper Lake
  - Centennial



Breakout Rooms: Work through R Script



Questions

1. What sorts of studies have you seen distance sampling applied to?
2. What similarities do you see between the N-mixture and distance sampling models?
3. What differences do you see?
4. Why does the distance sampling model use the Multinomial distribution to model detection?
5. Should you include the same covariate in the detection and abundance portions of the model?

west-inc.com

Corporate Headquarters

415 West 17th Street, Suite 200, Cheyenne, WY 82001
307.634.1756