Changes in Forest Communities of the Eastern United States

Jonathan Knott, Trenton Ford, Chathurangi Pathiravasan

Purdue University, University of Notre Dame, Southern Illinois University

knott1@purdue.edu, tford5@nd.edu, chathurangi@siu.edu

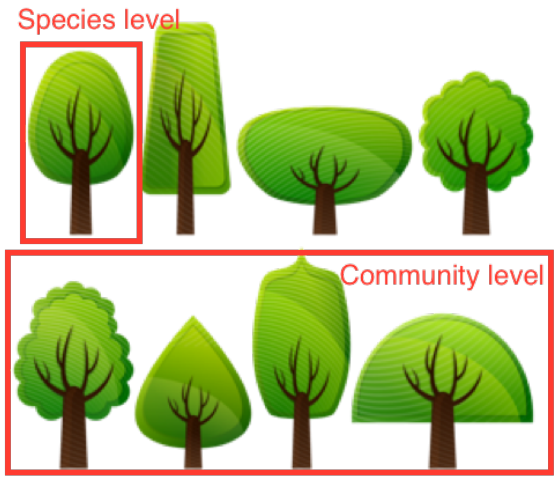
May 25, 2018

Motivation

Imagine you're walking through a forest...

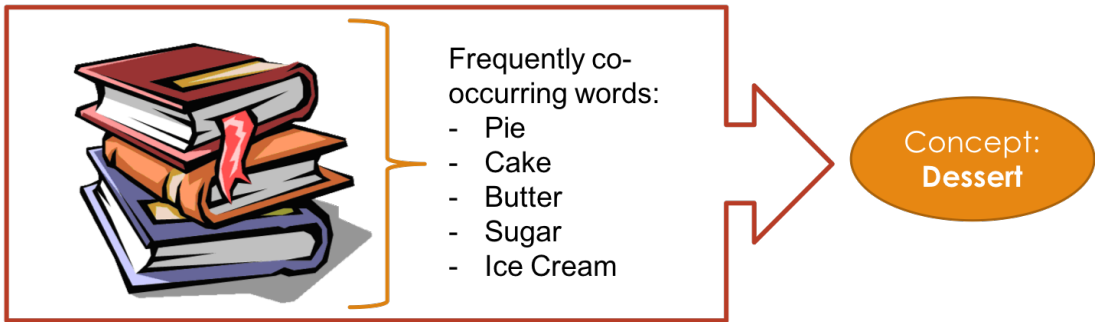
Research Goals

- Identify the main forest communities of the Eastern U.S
- Assess how they have changed based on two scales.
  1. Species Level
     (Reason: Species loss/local extinction, Species gain/invasion and Economic value)
  2. Community Level
     (Reason: Ecosystem functioning, Loss of forests/habitat types and Species interactions)

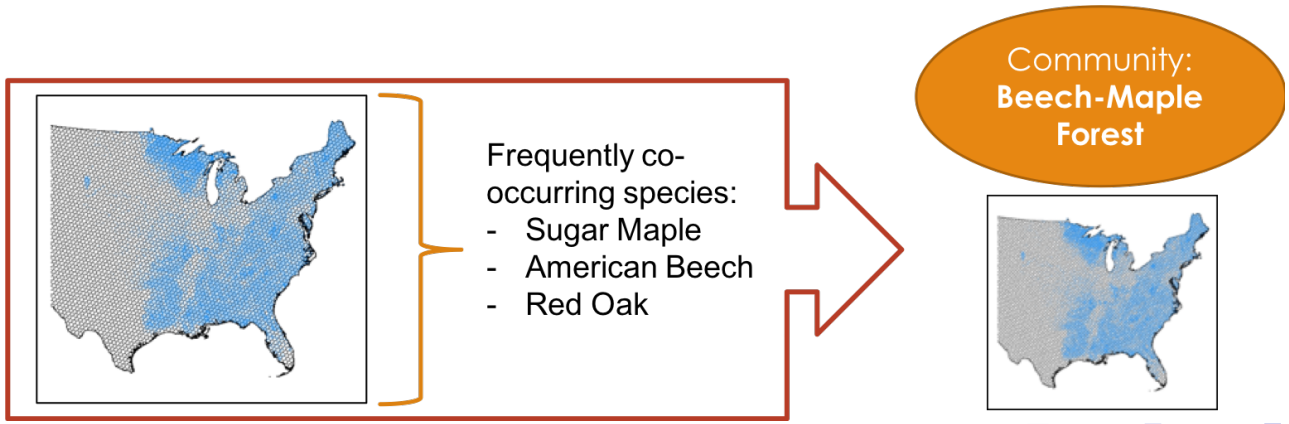


Latent Dirichlet Allocation (LDA)

- In the Latent Dirichlet Allocation (LDA) topic model, the frequency and co-occurrence of words in text segments define concepts. [Blei et al., 2003]

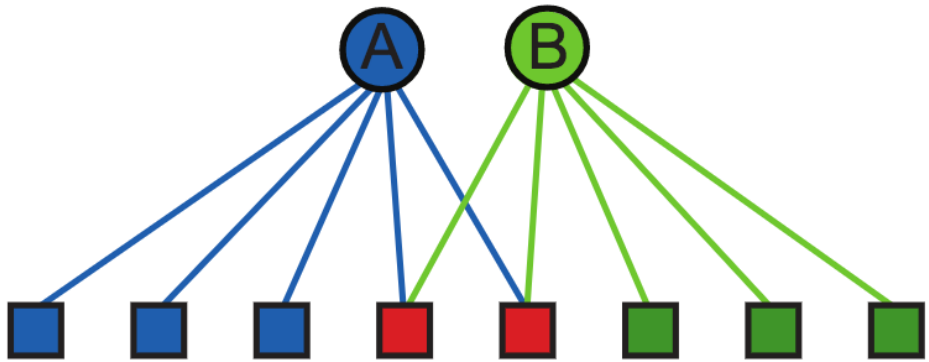


- LDA has recently been used to define communities from frequency and co-occurrence of species in sampling units [Valle et al., 2014]



BigCLAM Clustering Algorithm

- Cluster Affiliation Model for Big Networks (BigClam) on the Stanford Network Analysis Project (SNAP) [Yang and Leskovec, 2013]
- It is a popular graph mining algorithm that is capable of finding overlapping communities in networks containing millions of nodes and edges.



(a) Community Affiliation Network

- Squares = nodes = species
- Circles = clusters = communities
- Lines = cluster/community membership

Forest Inventory and Analysis (FIA)

- Approx. 80,000 plots in the eastern U.S.
  - Collected by U.S. Forest Service
  - $\geq$200 species; 79 selected for this project
- Compiled for two time periods (varies by state)
  - T1: 1980-1993
  - T2: 2013-2015
  - Date range for complete coverage
- Aggregated to a hexagon sample unit ($\sim$ 2400)
  - Reduces sampling bias
  - Accounts for fuzzed and swapped Lat/Lon from USFS

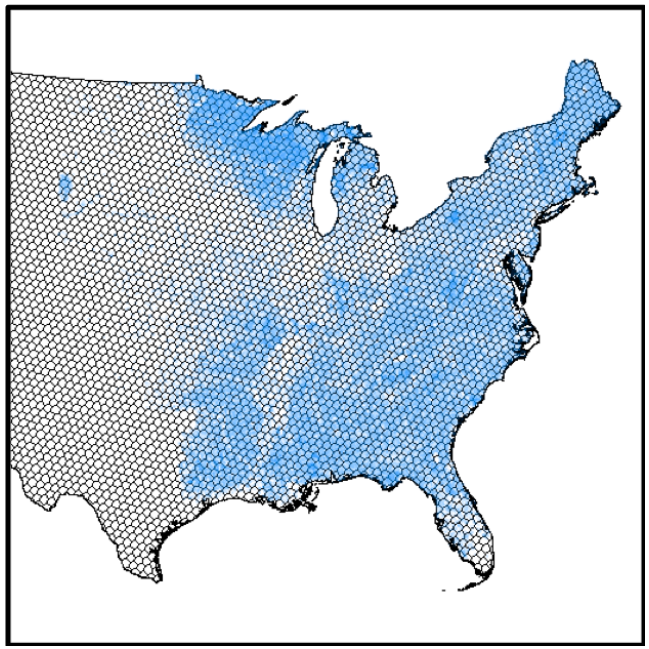


Figure: FIA plots (blue dots) and hexagon sample units

Abundance measures - For T1 and T2

LDA with Importance Value

$$\text{Importance Value (IV)} = \left(\frac{\text{rel. stem density} + \text{rel. basal area}}{2} \right)$$

LDA with Species Dominance Index [Costanza et al., 2017]

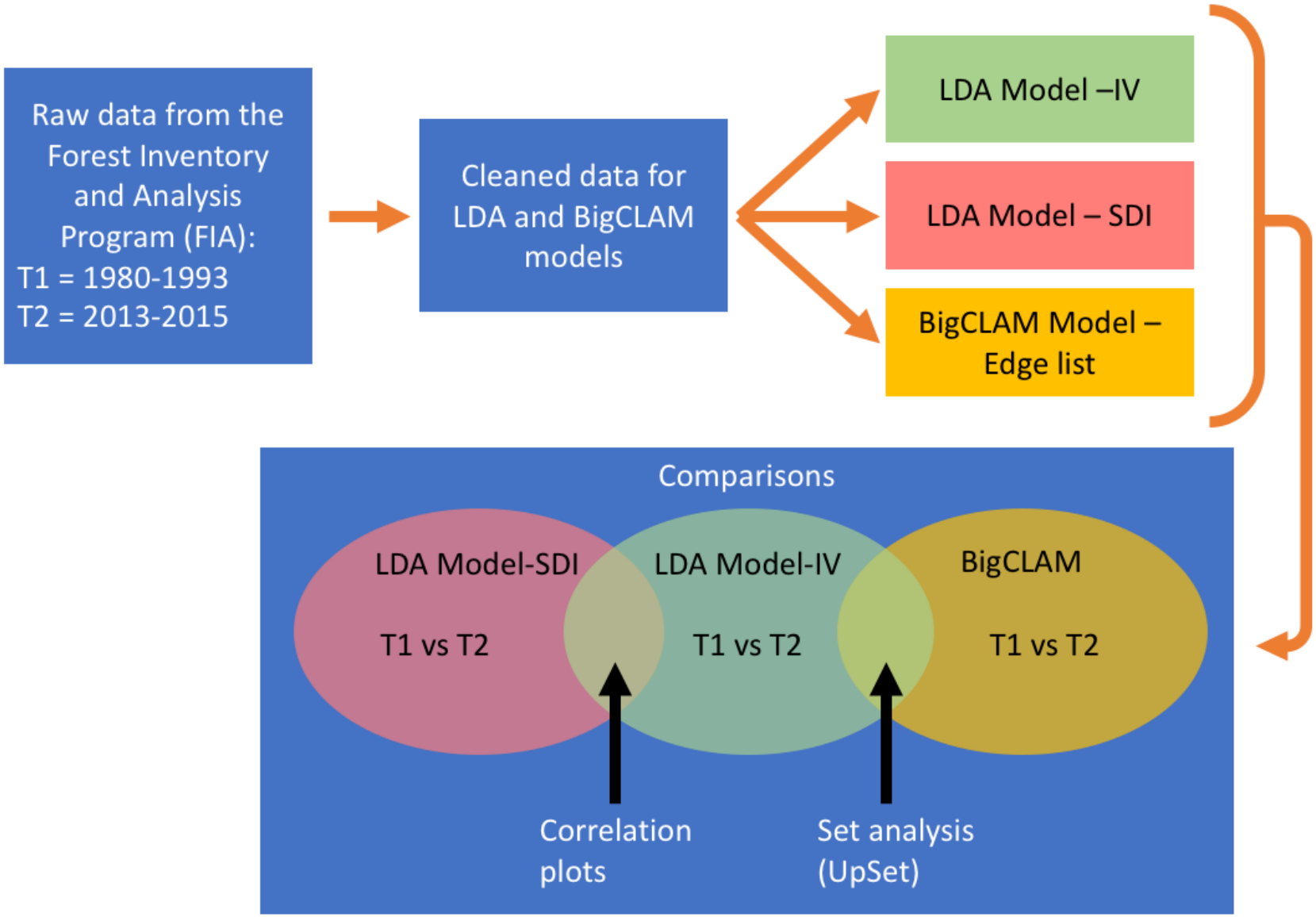
$$\text{Species Dominance Index} = \left(\frac{IV + \frac{1}{no.species\ in\ hex} + THC}{3} \right)$$

$$THC(\text{the tendency toward high cover}) = \begin{cases} 1 & \text{for IV} \geq 0.25 \ \ \& \ \max(\text{IV}) \ in \text{ the hexagon} \\ 0 & \text{otherwise} \end{cases}$$

BigCLAM with edge list

List of species overlap in each hexagon

Methodology

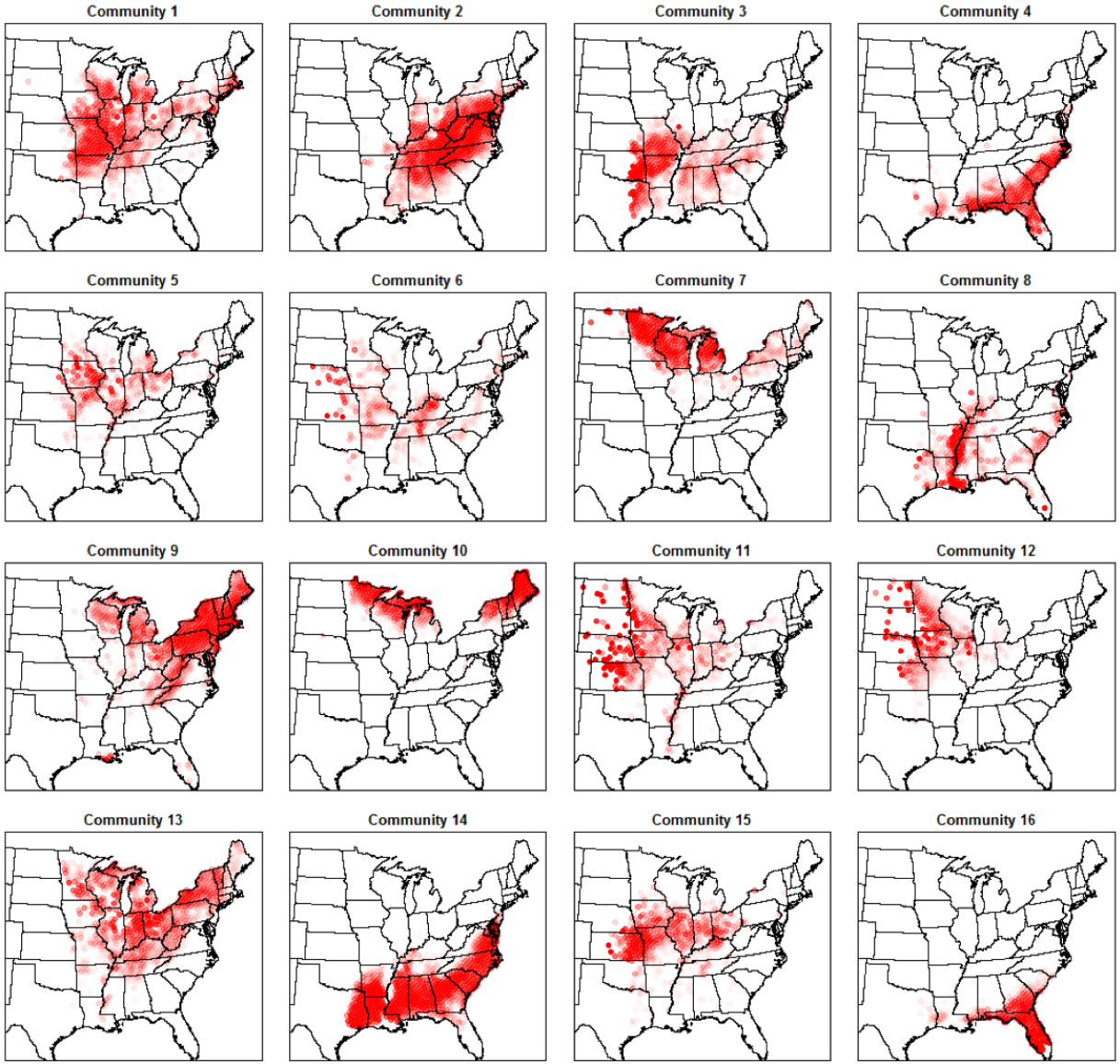


Results - Communities

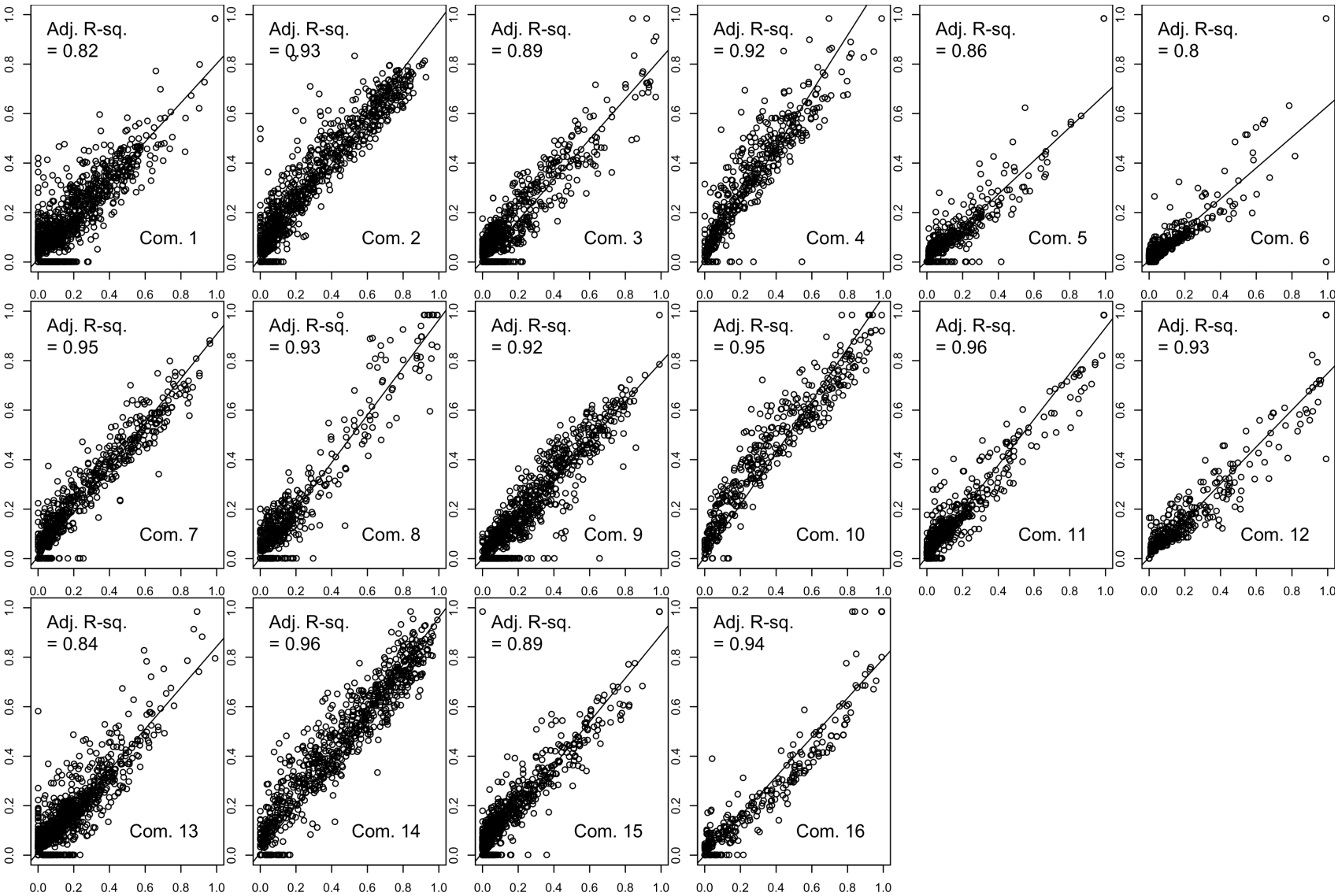
C1	C2	C3	C4	C5	C6	C7	C8
White Oak	Yellow Poplar	Post Oak	Red Maple	Silver Maple	Eastern Redcedar	Quaking Aspen	Sweetgum
Black Oak	Red Maple	Shortleaf Pine	Longleaf Pine	Pin Oak	Chinkapin Oak	Paper Birch	Red Maple
Northern Red Oak	Chestnut Oak	Blackjack Oak	Sweetbay	Swamp White Oak	Sycamore	Black Ash	Baldcypress
Sassafras	White Oak	Southern Red Oak	Water Oak	Hophornbeam	Eastern Redbud	Red Pine	Water Tupelo
Scarlet Oak	Virginia Pine	Black Oak	Slash Pine	Northern Red Oak	Scarlet Oak	Northern Red Oak	Overcup Oak

C9	C10	C11	C12	C13	C14	C15	C16
Red Maple	Balsam Fir	Boxelder	Bur Oak	Sugar Maple	Loblolly Pine	Black Walnut	Slash Pine
Eastern White Pine	Northern White Cedar	Eastern Cottonwood	Hophornbeam	Hophornbeam	Sweetgum	Honeylocust	Pondcypress
Eastern Hemlock	Paper Birch	Black Walnut	Northern Red Oak	American Beech	Shortleaf Pine	Osage Orange	Live Oak
American Beech	Black Spruce	Buckeye Horsechestnut	Boxelder	Northern Red Oak	Water Oak	Black Locust	Laurel Oak
Northern Red Oak	Red Spruce	Black Ash	Buckeye Horsechestnut	Musclewood	Southern Red Oak	Eastern Redbud	Baldcypress

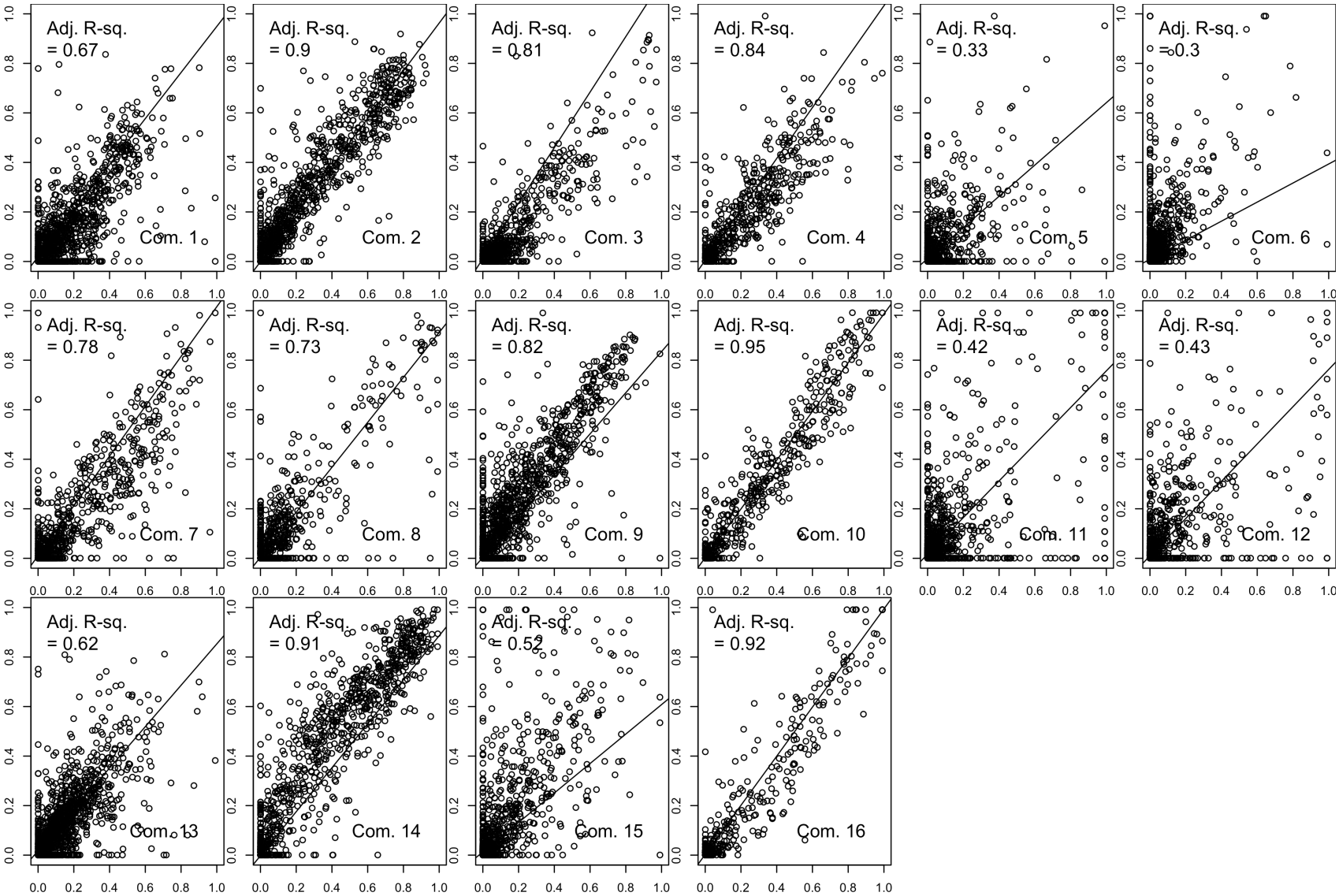
Results - Community Location



Results - IV vs. SDI at T1



Results - T1 vs. T2 (IV)



Results

Largest Overlapping Communities with Exclusive Species

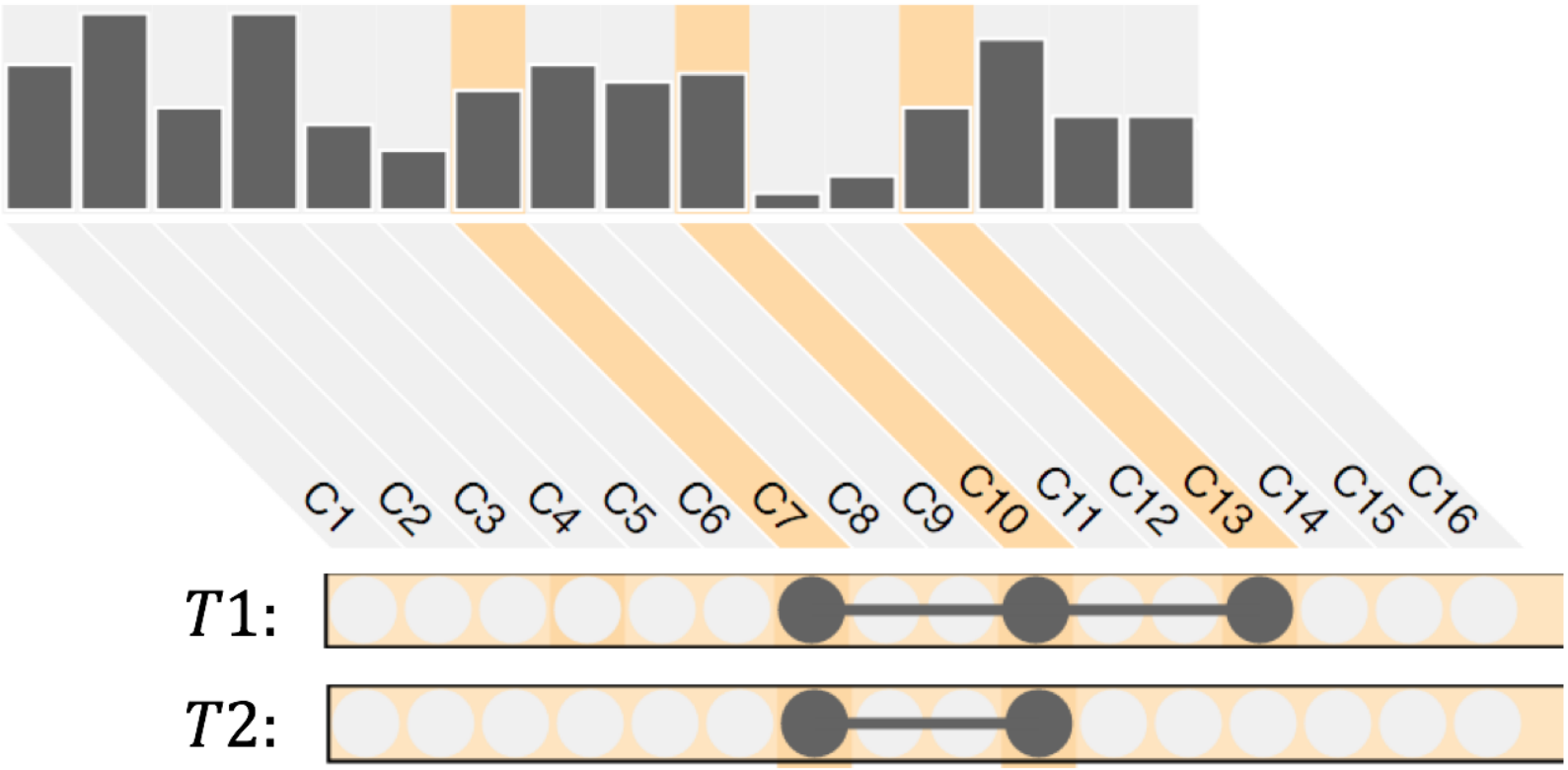
Table: T1 — LDA

Species	Communities
Balsam Poplar	2
Paper Birch	2
Quaking Aspen	2
Tamarack	2
White Spruce	2

Table: T2 — LDA

Species	Communities
Balsam Poplar	2
Black Ash	2
Paper Birch	2
Quaking Aspen	2
Tamarack	2
White Spruce	2

Results - Black Ash



Conclusions

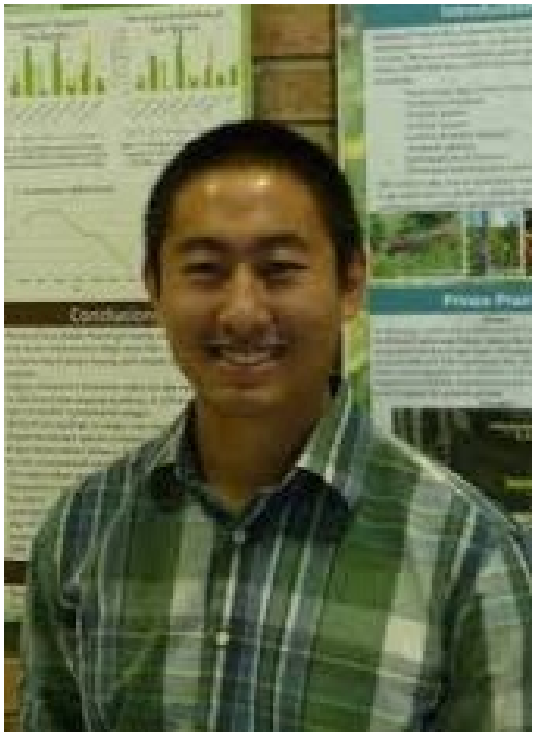
- High concordance between LDA model with IV, SDI, and BigCLAM model
- Close (but not perfect) relationship between T1 and T2: evidence of forest community change
- Possible evidence of community response to Emerald Ash Borer invasion

Future Directions

- Determine the best number of communities to describe the data set using Bootstrapping methods. (Currently k = 16 - AIC)
  - Assess "goodness-of-fit" for LDA and BigClam by incorporating silhouette or other measures for validation of consistency within clusters.
- Interpret results (such as Black Ash Reduction) in an ecological context
- Predict the forest changes using improved clustering methods (hierarchical/ k-means clustering)[Costanza et al., 2017].
- Investigate factors that affect communities (climate change, land use change, management practices, etc.)

Chathurangi Pathiravasan
Southern Illinois University
chathurangi@siu.edu

Trenton Ford
University of Notre Dame
tford5@nd.edu



Jonathan Knott
Purdue University
knott1@purdue.edu

References

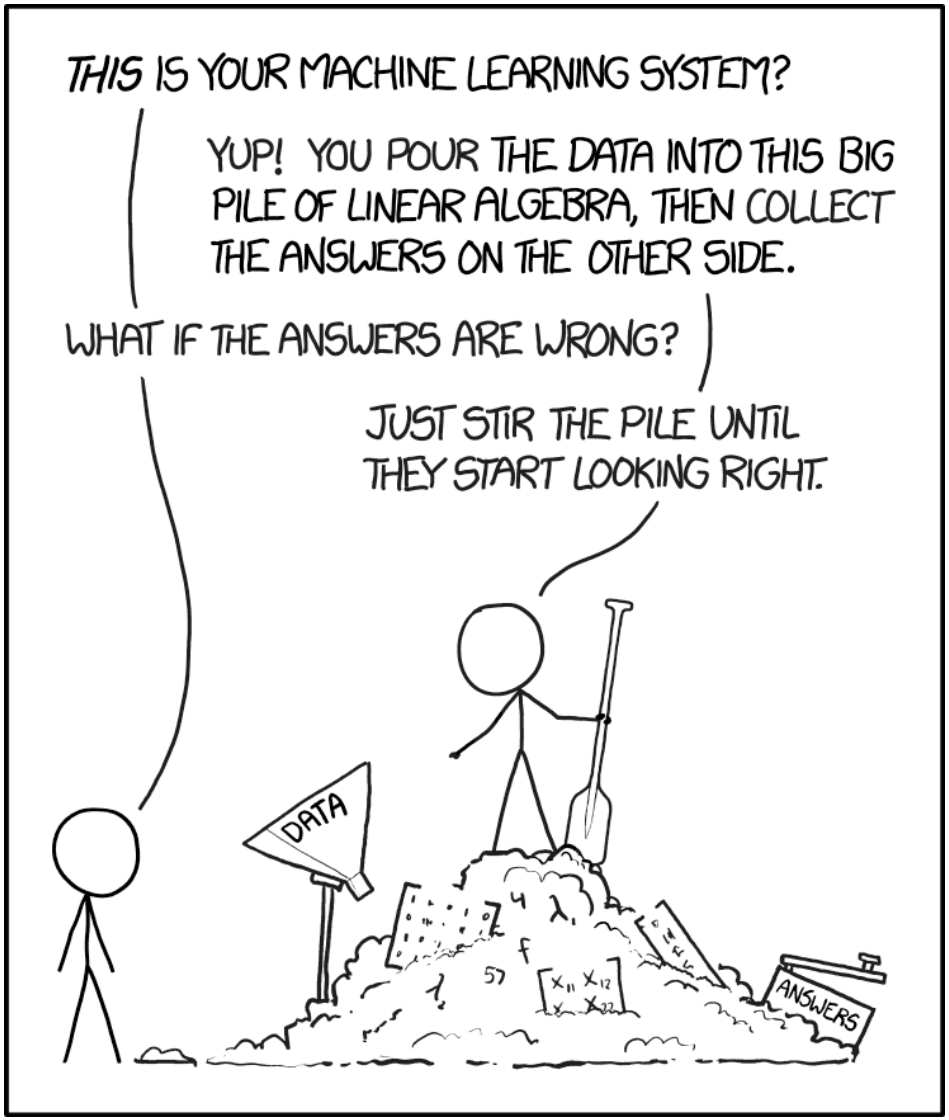

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003).
Latent dirichlet allocation.
Journal of machine Learning research, 3(Jan):993–1022.


Costanza, J. K., Coulston, J. W., and Wear, D. N. (2017).
An empirical, hierarchical typology of tree species assemblages for assessing forest dynamics under global change scenarios.
PloS one, 12(9):e0184062.


Valle, D., Baiser, B., Woodall, C. W., and Chazdon, R. (2014).
Decomposing biodiversity data using the latent dirichlet allocation model, a probabilistic multivariate statistical method.
Ecology letters, 17(12):1591–1601.


Yang, J. and Leskovec, J. (2013).
Overlapping community detection at scale: a nonnegative matrix factorization approach.
In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM.

Questions?


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS