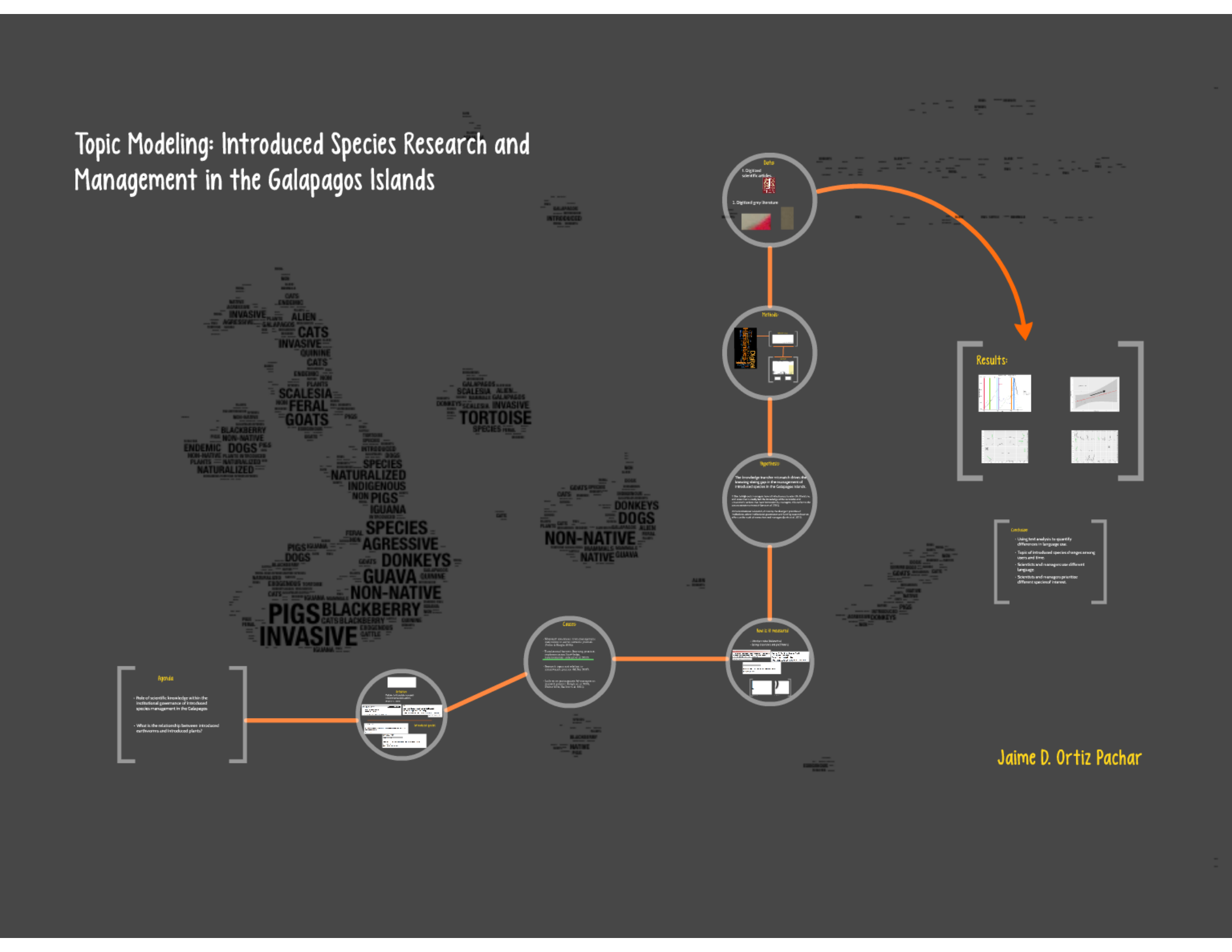
# Topic Modeling: Introduced Species Research and Management in the Galapagos Islands



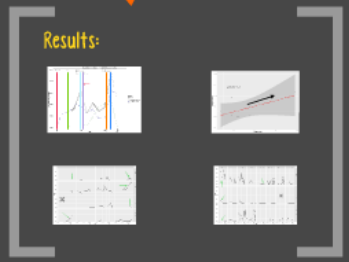
# Topic Modeling: Introduced Species Research and Management in the Galapagos Islands

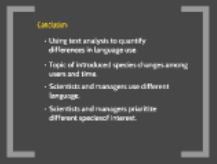









Jaime D. Ortiz Pachar

# Agenda

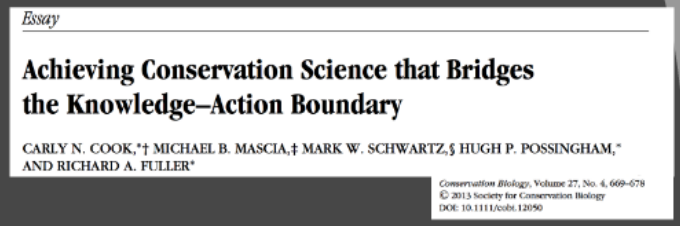
- Role of scientific knowledge within the institutional governance of introduced species management in the Galapagos

- What is the relationship between introduced earthworms and introduced plants?


## Definition:

Failure to translate research into conservation action.

(Knight et al., 2008)


## Introduced species


# Causes:

- Monetary constraints from management institutions to access scientific journals (Pullin & Knight 2005).
- Translational barriers (learning, practice, implementation, knowledge, measurements) (Arlettaz et al. 2010).
- Research topics not relevant to conservation practice (McNie 2007).
- Little or no participation by managers in research projects (Knight et al. 2008; Hulme 2014; Barlow et al. 2016).

# How is it measured

- Literature review (bibliometrics)
- Surveys (researchers and practitioners)

LETTER

**Closing the Knowing–Doing Gap in Invasive Plant Management: Accessibility and Interdisciplinarity of Scientific Research**

Virginia Matzek[1], Justin Covino[1], Jennifer L. Funk[2], & Martin Saunders[1]

Conservation Letters, May/June 2014, 7(3), 208–215

**Knowing But Not Doing: Selecting Priority Conservation Areas and the Research–Implementation Gap**

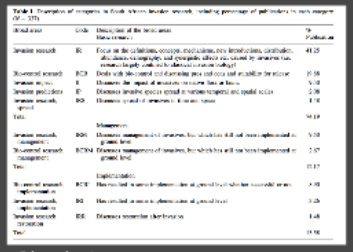
ANDREW T. KNIGHT,*†† RICHARD M. COWLING,* MATHIEU ROUGET,† ANDREW BALMFORD,‡ AMANDA T. LOMBARD,*§ AND BRUCE M. CAMPBELL**

Biol Invasions (2010) 12:4065–4075
DOI 10.1007/s10530-010-9812-x

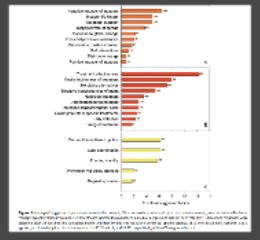
ORIGINAL PAPER

**How wide is the "knowing-doing" gap in invasion biology?**

Karen J. Esler · Heidi Prozesky · Gyan P. Sharma · Melodie McGeoch



Esler et al. 2010



Matzek et al. 2014

· Melodie McGeoch

**Table 1** Description of categories in South African invasion research, including percentage of publications in each category ($N = 337$)

| Broad areas | Code | Description of the broad areas | % Publication |
|---|---|---|---|
| | | Basic research | |
| Invasion research | IR | Focus on the definitions, concepts, mechanisms, new introductions, distribution, abundance, demography, and synergistic affects etc. caused by invasives (i.e. research largely confined to classical invasion biology) | 41.25 |
| Bio-control research | BCR | Deals with bio-control and discussing pros and cons and suitability for release | 19.88 |
| Invasion impact | II | Discusses the impact of invasives on native flora or fauna | 9.50 |
| Invasion predictions | IP | Discusses invasive species spread at various temporal and spatial scales | 2.08 |
| Invasion research, spread | IRS | Discusses spread of invasives in time and space | 1.48 |
| Total | | | 74.19 |
| | | Management | |
| Invasion research, management | IRM | Discusses management of invasives, but which has still not been implemented at ground level | 9.50 |
| Bio-control research, management | BCRM | Discusses management of invasives, but which has still not been implemented at ground level | 2.67 |
| Total | | | 12.17 |
| | | Implementation | |
| Bio-control research, implementation | BCRI | Has resulted in some implementation at ground level whether successful or not | 8.90 |
| Invasion research, implementation | IRI | Has resulted in some implementation at ground level | 3.26 |
| Invasion research restoration | IRR | Discusses restoration after invasion | 1.48 |
| Total | | | 13.58 |

Esler et al. 2010

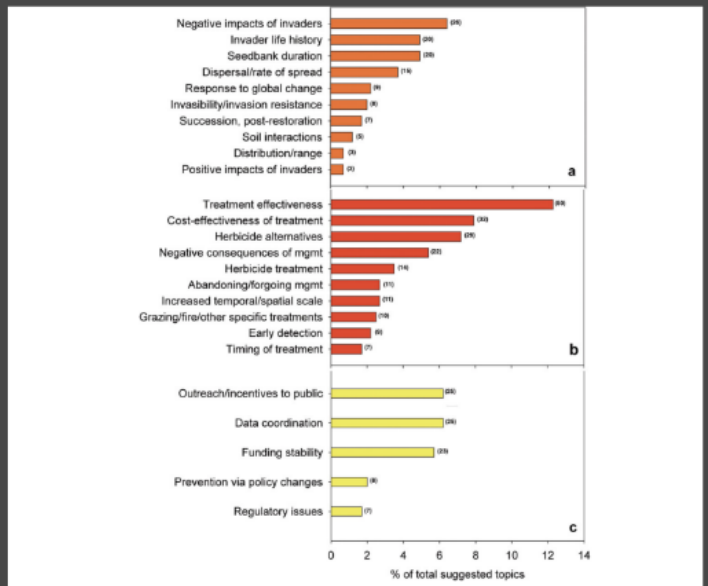



**Figure 3** Managers' suggested topics in response to the prompt, "What research questions do you most need answered, in order to be effective at managing plant invasions?" Numbers in parentheses indicate the quantity of individual suggestions made in each category. Open-ended responses were classified after the fact into the categories shown and then divided into (a) basic science, (b) applied science, or (c) interdisciplinary research. Basic, applied, and interdisciplinary topics accounted for 30.1%, 48.1%, and 21.7%, respectively, of the 405 suggested topics.

Matzek et al. 2014

# Hypothesis:

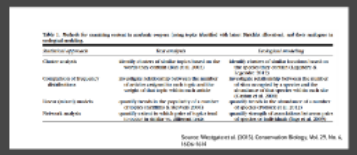
## The knowledge transfer mismatch drives the knowing-doing gap in the management of introduced species in the Galapagos Islands.

1) Due to high costs managers have a limited access to scientific literature, and researchers usually lack the knowledge of the successful and unsuccessful actions that have been taken by managers, this conforms the *communication mismatch* (Jarvis *et al.* 2015).

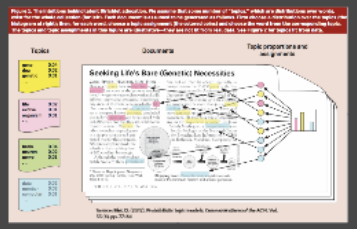
2) the *institutional mismatch*, driven by the divergent priorities of institutions, where institutional governance and funding sources have an effect on the work of researchers and managers (Jarvis *et al.* 2015).

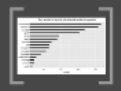
# Methods:


## Digitized text analysis



## Topic Model



# Digitized text analysis

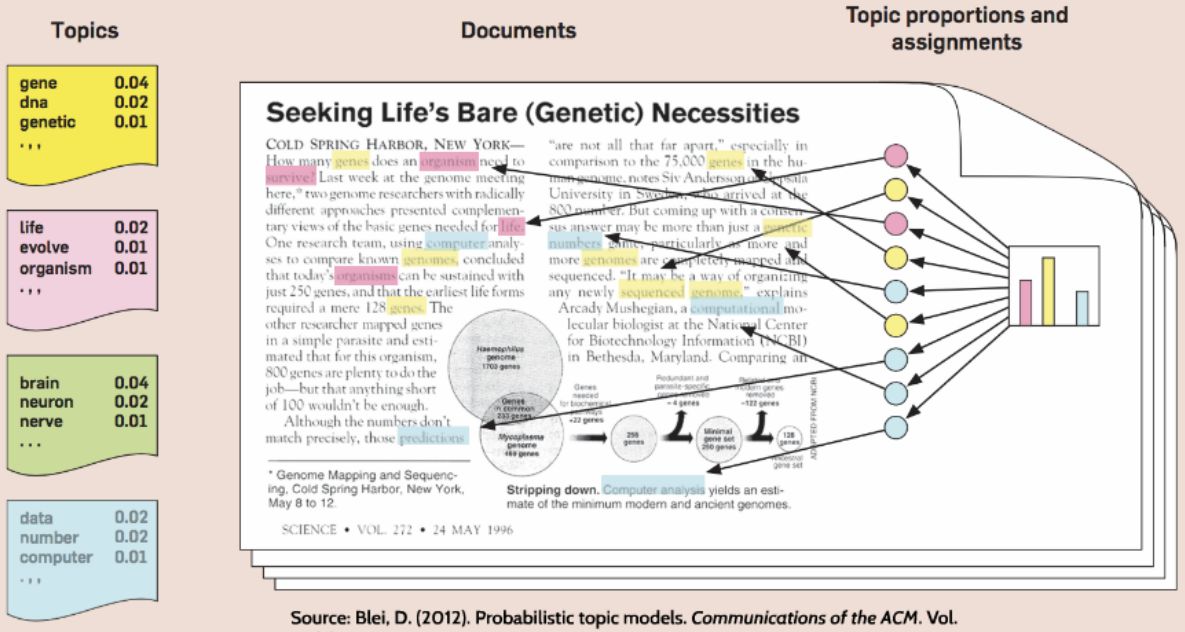
**Table 1. Methods for examining content in academic corpora (using topics identified with latent Dirichlet allocation), and their analogues in ecological modeling.**

| *Statistical approach* | *Text analysis* | *Ecological modeling* |
|---|---|---|
| Cluster analysis | identify clusters of similar topics based on the words they contain (Blei et al. 2003) | identify clusters of similar locations based on the species they contain (Legendre & Legendre 2012) |
| Comparison of frequency distributions | investigate relationship between the number of articles assigned to each topic and the weight of that topic within each article | investigate relationship between the number of sites occupied by a species and the abundance of that species within each site (Gaston et al. 2000) |
| Linear (mixed) models | quantify trends in the popularity of a number of topics (Griffiths & Steyvers 2004) | quantify trends in the abundance of a number of species (Pollock et al. 2012) |
| Network analysis | quantify extent to which pairs of topics tend to occur in similar vs. different texts | quantify strength of associations between pairs of species or individuals (Ings et al. 2009) |

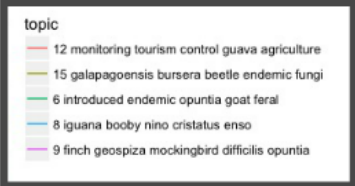
Source: Westgate et al. (2015). Conservation Biology, Vol. 29, No. 6, 1606-1614

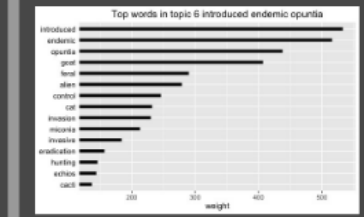
# Topic Model

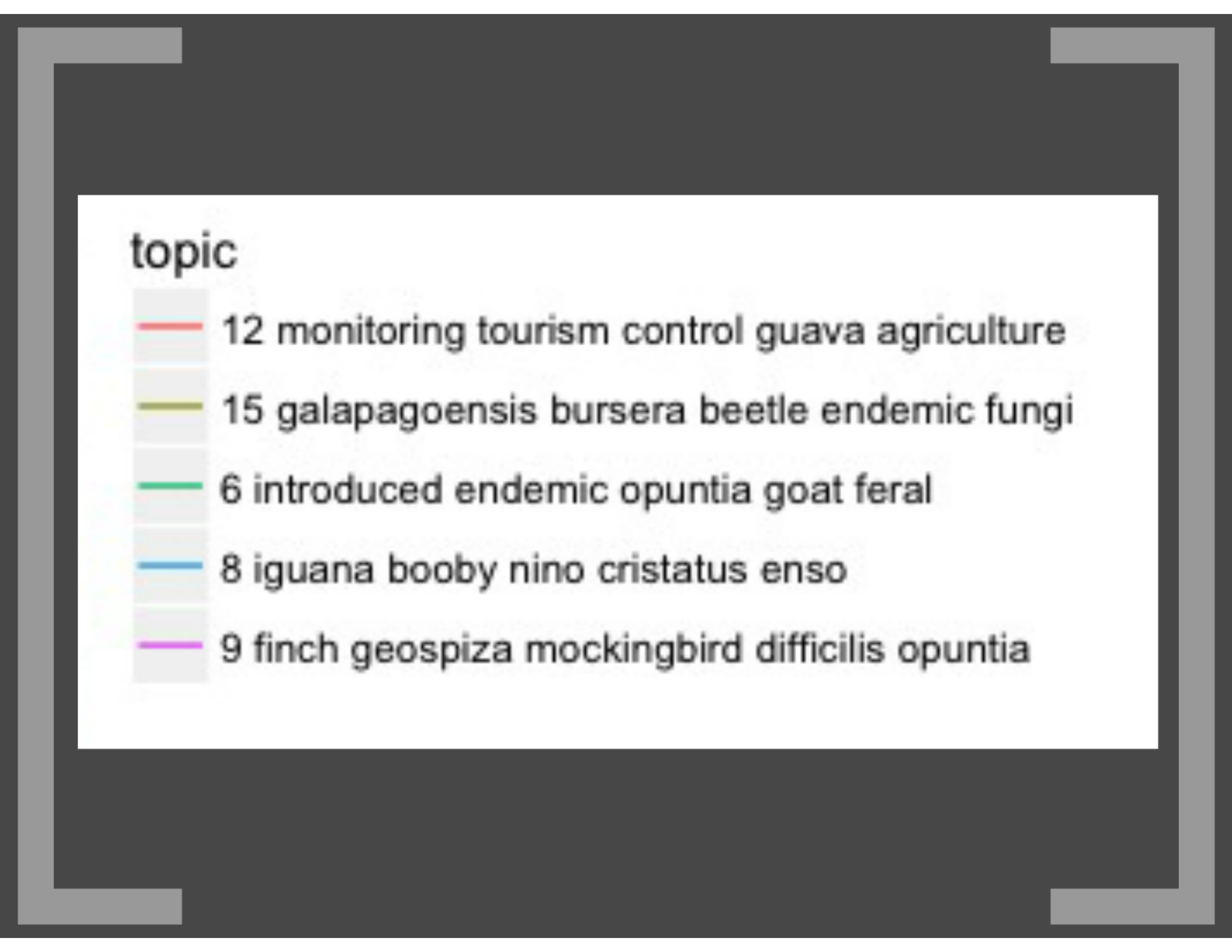
**Figure 1. The intuitions behind latent Dirichlet allocation. We assume that some number of "topics," which are distributions over words, exist for the whole collection (far left). Each document is assumed to be generated as follows. First choose a distribution over the topics (the histogram at right); then, for each word, choose a topic assignment (the colored coins) and choose the word from the corresponding topic. The topics and topic assignments in this figure are illustrative—they are not fit from real data. See Figure 2 for topics fit from data.**

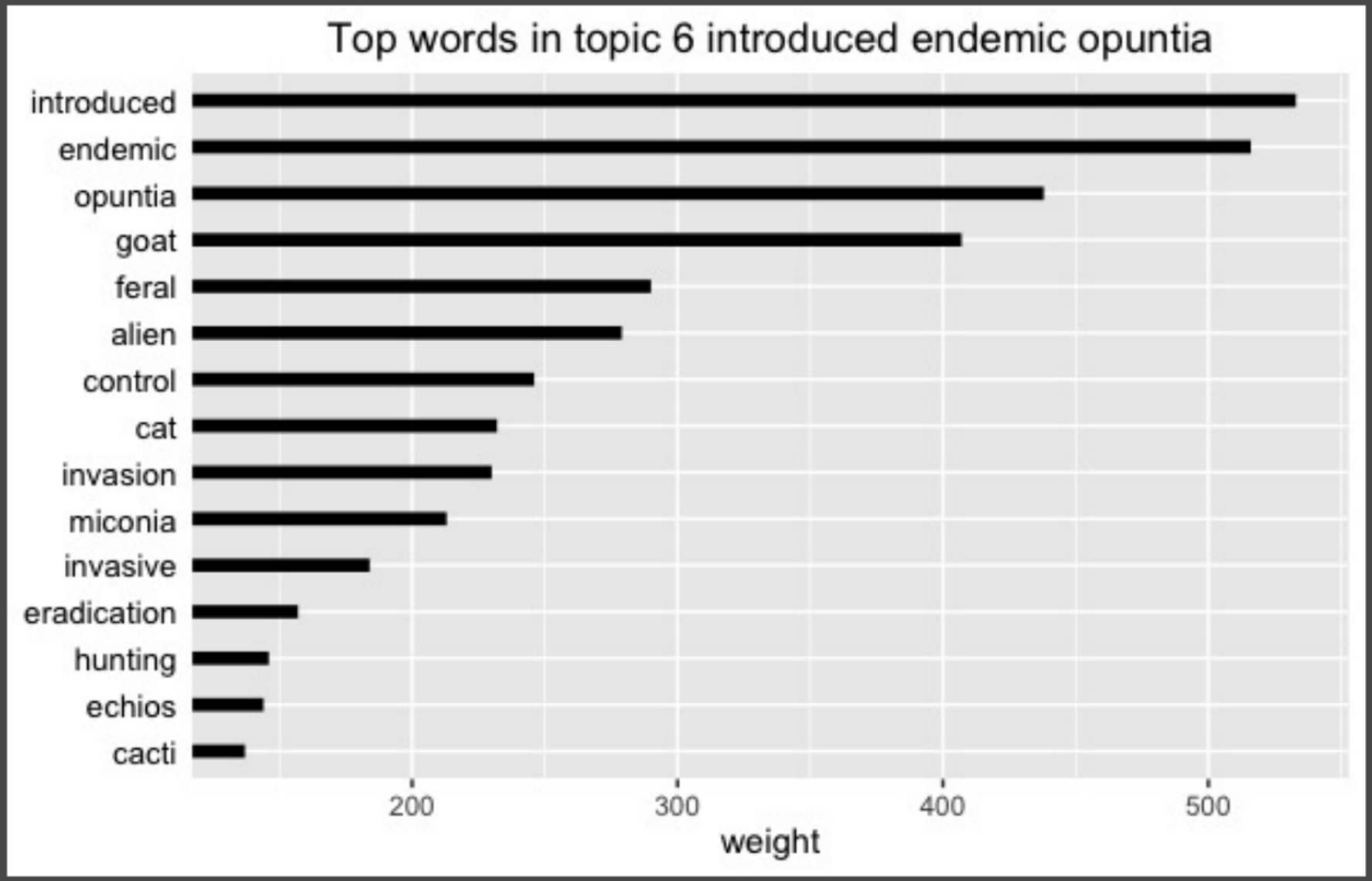



Source: Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*. Vol. 55 (4). pp. 77-84




topic
12 monitoring tourism control guava agriculture
15 galapagoensis bursera beetle endemic fungi
6 introduced endemic opuntia goat feral
8 iguana booby nino cristatus enso
9 finch geospiza mockingbird difficilis opuntia

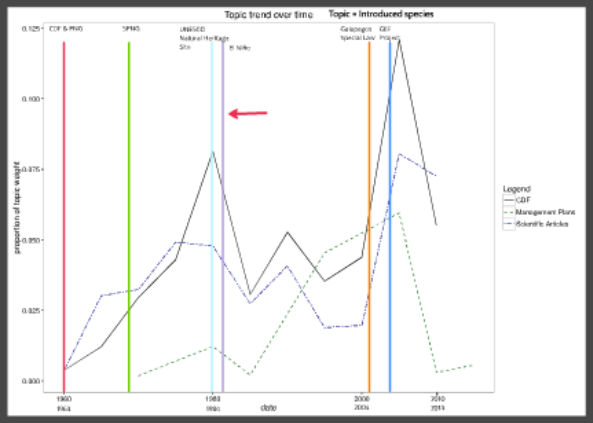

Top words in topic 6 introduced endemic opuntia
introduced
endemic
opuntia
goat
feral
alien
control
cat
invasion
miconia
invasive
eradication
hunting
echios
cacti
200
300
400
500
weight

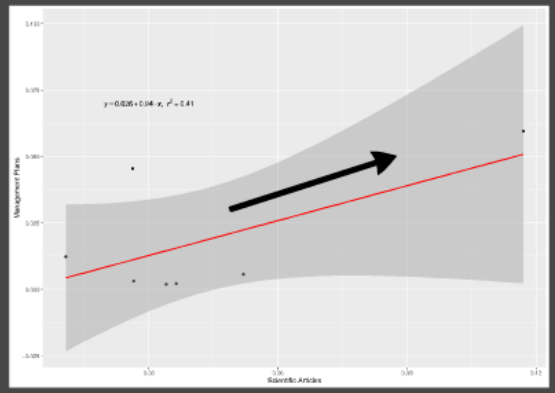
# Data:

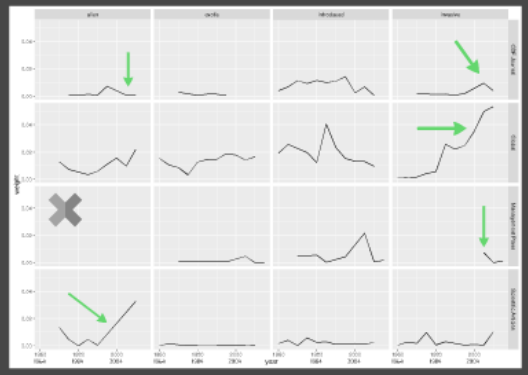
1. Digitized scientific articles.

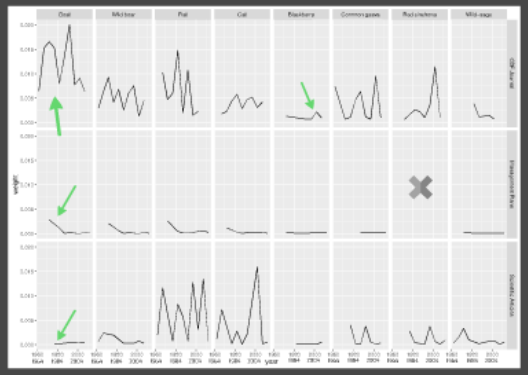
2. Digitized grey literature

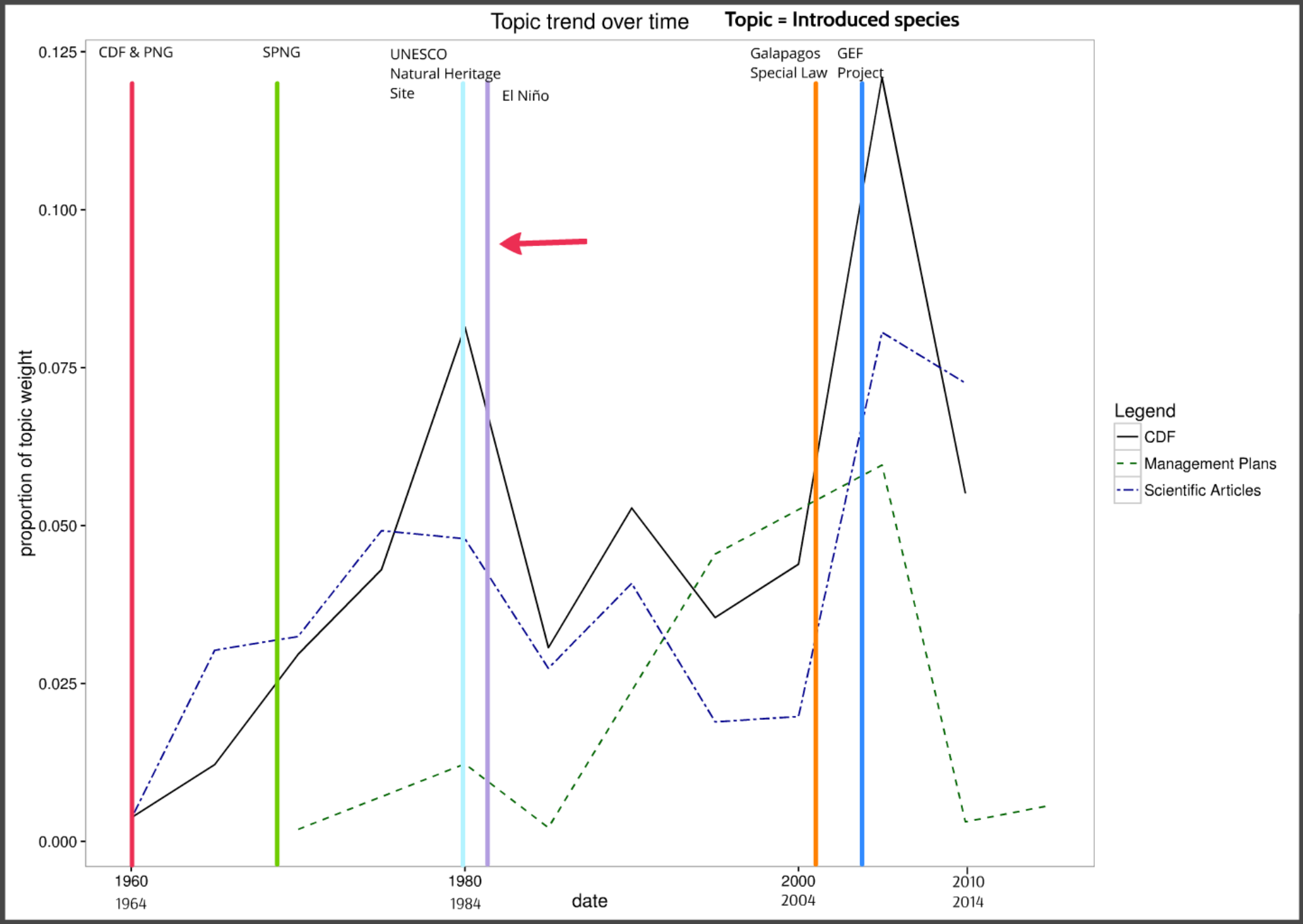
# Results:

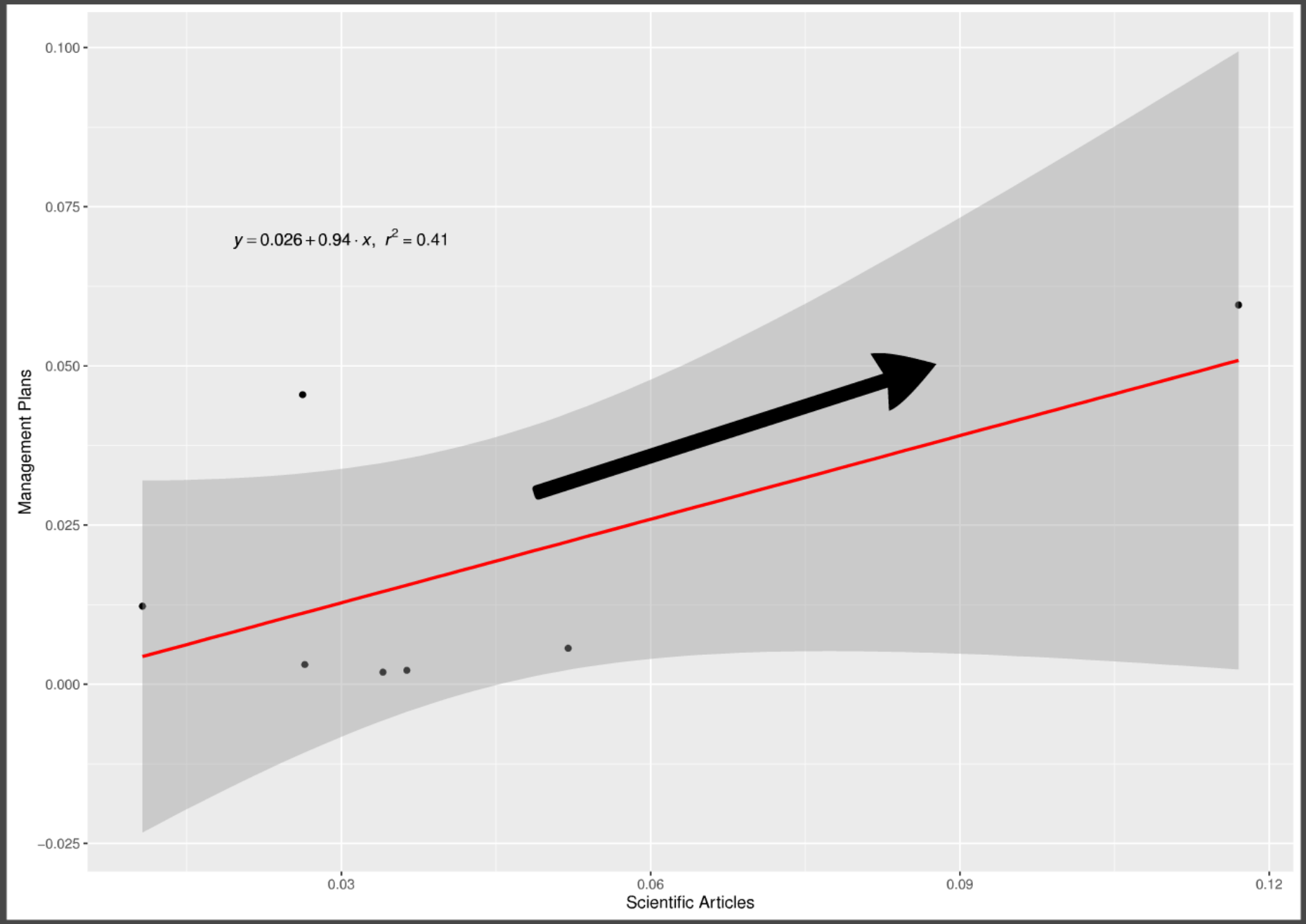

Topic trend over time
Topic = Introduced species
CDF & PMS
SPMS
UNESCO
National Heritage
Site
E Nifty
Galapagos
Special Law
CEF
PSMS
proportion of topic weight
0.125
0.100
0.075
0.050
0.025
0.000
1900
1964
1980
1985
2000
2004
2020
2024
date
Legend
CDF
Management Plans
Scientific Articles

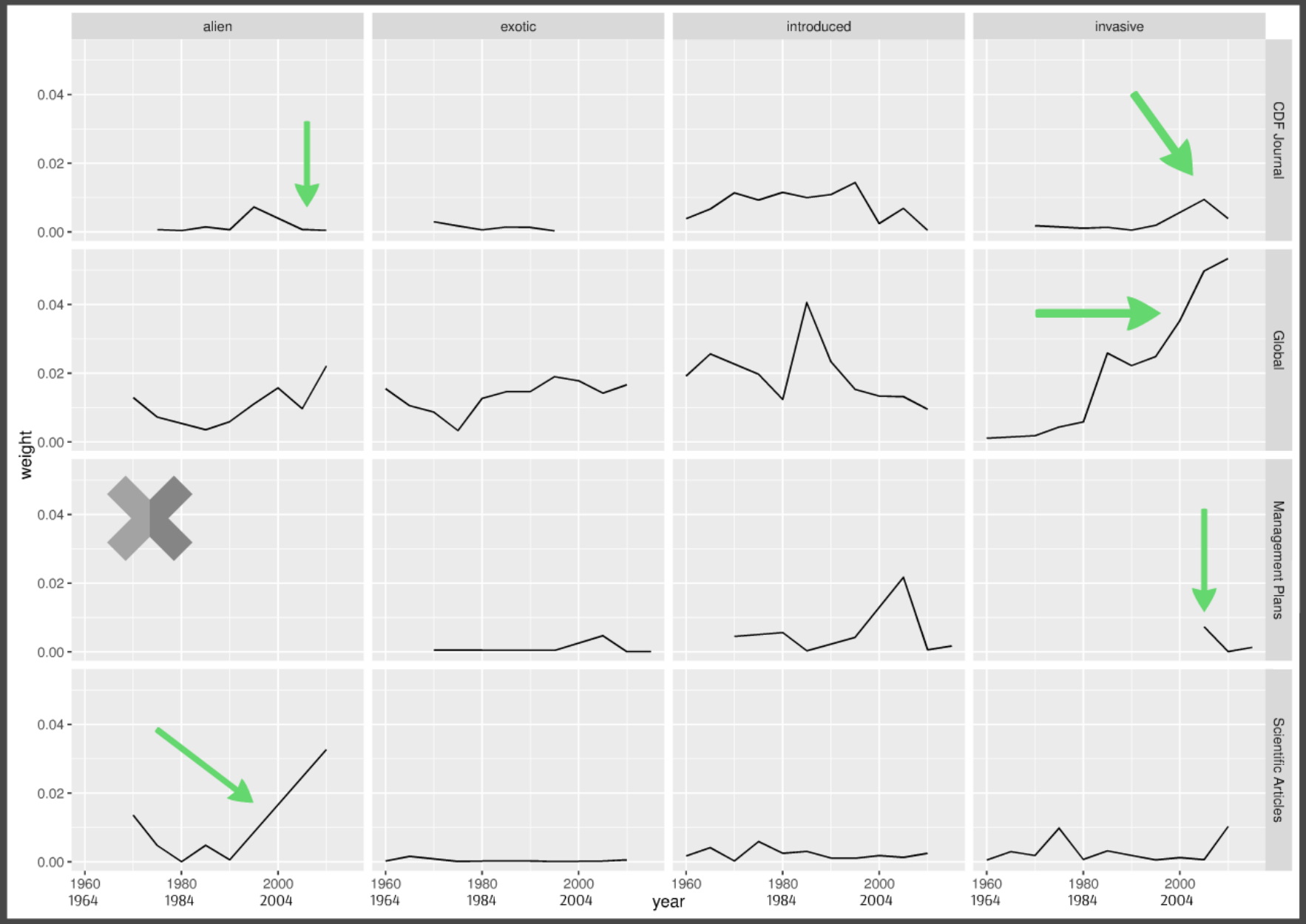

y = 0.035 + 0.04 · x, r² = 0.41
Management Plans
Scientific Articles

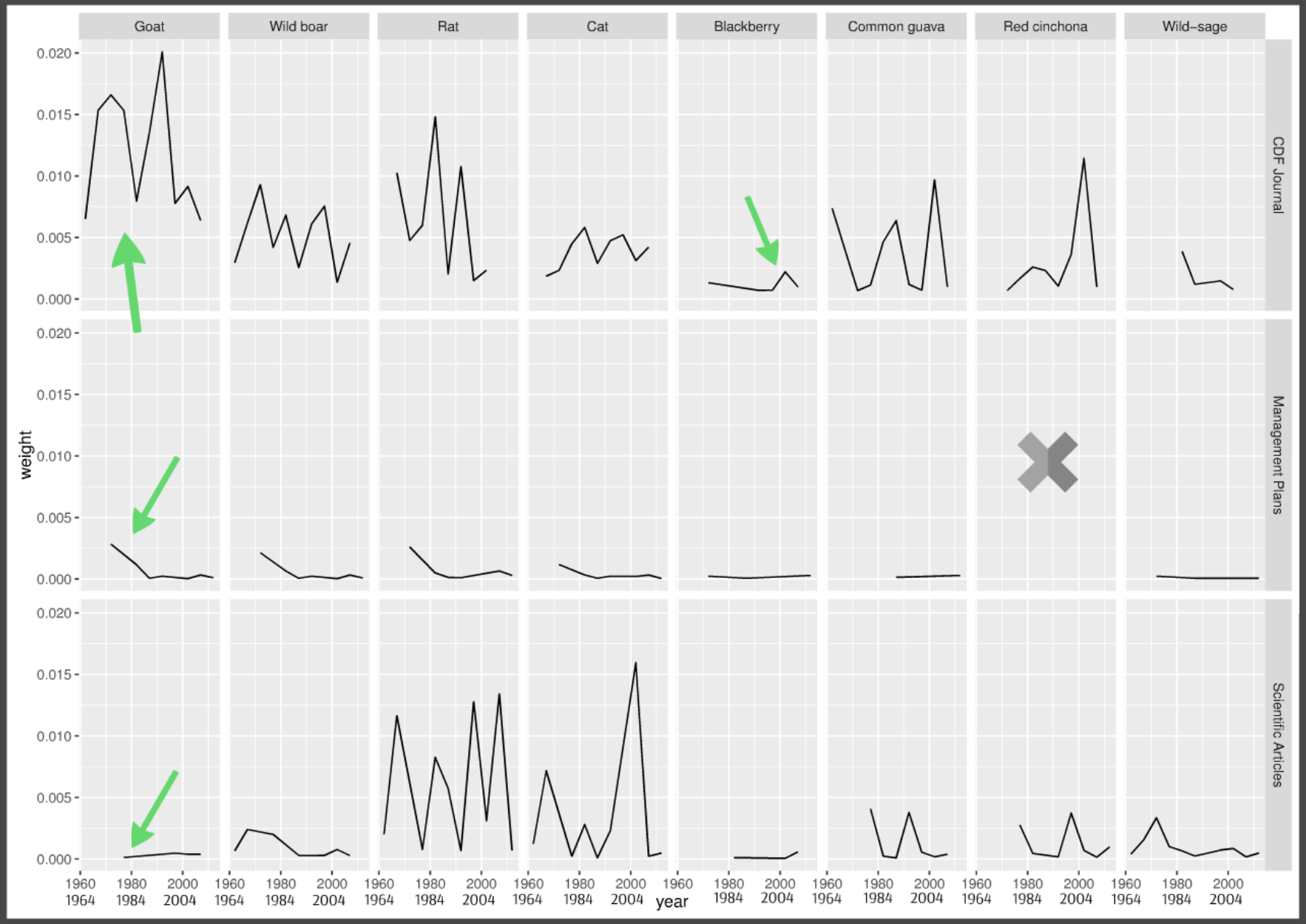

plan
cecils
Introduced
species
CDF by var
Species
Management Plans
Scientific Articles
weight
year


Goat
Wild boar
Pig
Cat
Blackberry
Common grasses
Plant orchems
Wild sugar
CDF by var
Species
Management Plans
Scientific Articles
weight
year


Topic trend over time
Topic = Introduced species
CDF & PNG
SPNG
UNESCO
Natural Heritage
Site
El Niño
Galapagos
Special Law
GEF
Project
0.125
0.100
0.075
0.050
0.025
0.000
proportion of topic weight
1960
1964
1980
1984
2000
2004
2010
2014
date
Legend
CDF
Management Plans
Scientific Articles


0.100
0.075
0.050
0.025
0.000
−0.025
Management Plans
$y = 0.026 + 0.94 \cdot x,\ r^2 = 0.41$
0.03
0.06
0.09
0.12
Scientific Articles


alien
exotic
introduced
invasive
CDF Journal
Global
Management Plans
Scientific Articles
weight
0.04
0.02
0.00
1960
1964
1980
1984
2000
2004
year


Goat
Wild boar
Rat
Cat
Blackberry
Common guava
Red cinchona
Wild−sage
CDF Journal
Management Plans
Scientific Articles
weight
0.020
0.015
0.010
0.005
0.000
1960
1964
1980
1984
2000
2004
year

## Conclusion:

- Using text analysis to quantify differences in language use.
- Topic of introduced species changes among users and time.
- Scientists and managers use different language.
- Scientists and managers priaritize different speciesof interest.