DATA VISUALIZATION TIPS
DATA VISUALIZATION EXAMPLES

- These examples are from Data Visualization experts
  - Alberto Cairo
  - Randy Olson
FIRST EXAM

- Minimum: 0
- Maximum: 100
- Average score: 72

SECOND EXAM

- Minimum: 0
- Maximum: 137
- Average score: 96

Based on Richard Thaler’s “Misbehaving: The Making of Behavioral Economics”
Data from thermometers (red) and from tree rings, corals, ice cores and historical records (blue).
Many in Developing Nations Struggle to Afford Food

Source: GDP per capita (PPP) from IMF World Economic Outlook Database, April 2013. Data not available for Palestinian territories. PEW RESEARCH CENTER

Districts with Fewest Eligible Voters Have Large Hispanic Populations and Are Represented by Democrats

% of congressional district population that is...

Note: Eligible voters are defined as U.S. citizens, ages 18+. Source: Pew Research Center analysis of U.S. Census Bureau’s 2013 American Community Survey
PEW RESEARCH CENTER

2014 Partisan Advantages Among Whites, by Year of Birth

Source: Merged Pew Research Center surveys conducted in 2014.
PEW RESEARCH CENTER
The core principles
Good graphics...

1. They are based on good data
2. They attract readers’ attention
3. They don’t frustrate readers
4. They show the right amount of data
Less regulation = More Industry Investment

Following 1992 Cable Regulations

$14 billion

1993-1996

After regulations were relaxed

$56 billion

1999-2003

Based on a chart by the National Cable & Telecommunications Association

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo
Less regulation = More Industry Investment

Following 1992 Cable Regulations

$14 billion

1993-1996

$56 billion After regulations were relaxed

1999-2003

Cable Industry Infrastructure Expenditures

In billions

Cable Television Consumer Protection and Competition Act ("Regulation")

Telecommunications Act is passed ("Deregulation")

$16

$12

$8

$4

$16

1990 92 96 00 02 04 06 08 10 2013

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo
Minimizing misunderstanding
Robert Reich’s “Saving Capitalism: For the Many, not for the Few”
Robert Reich’s “Saving Capitalism: For the Many, not for the Few”

Visual design shouldn’t be an afterthought
Multi-scale Modeling and Assessment of Malaria Risk in Northern South America

Alimi, T. O.¹; Fuller, D. O.¹,² and Beier, J.C.¹,³

¹ Abess Center for Ecosystem Science and Policy; ² Department of Geography and Regional Studies; ³ Department of Epidemiology and Public Health, University of Miami

1. Introduction

The public health problem posed by malaria has made it a top priority for control efforts and the general consensus globally is that its elimination is crucial for continued international development. Consequently, there is ongoing research in different regions including South America (SA) to better understand the disease dynamics with the intent that findings may establish scientific frameworks that would support the development of new intervention strategies for malaria elimination in areas with seasonal malaria. One of such investigations is undertaken by the International Centers of Excellence in Malaria Research (ICEMR) under a National Institutes of Health (NIH) grant.

While only about 3% of the global malaria burden is borne by SA¹, undertaking malaria research in the region is currently important because an estimated 23 million people are at risk² and approximately about 80% of clinical cases are found in Northern South America (NSA)³. A key factor limiting effective control is lack of data and uneven implementation of control measures, including use of bednets, sprays, early diagnosis, and treatment. As part of the ICEMR investigation, this paper seeks to model the spatial patterns of malaria risk in NSA through vector distribution and land-use changes. Furthermore, I intend to investigate the perceptions of malaria risk in order to identify barriers to adoption and how they can be circumvented.

2. Significance

Spatial distribution of malaria risk is still perceived as broadly categorized by the WHO’s traditional risk maps which are highly generalized, of low resolution and have broad categories with uncertain boundaries (see da Nunes-Silva et al. 2012). There is need for up-to-date high resolution risk maps which can aid malaria control efforts. Second, modeling distribution of principal malaria vectors and land use changes which may explain the observed distribution and risk are useful tools which could guide future management strategies. Finally, understanding the perception of risk populations may help address barriers to adoption of interventions and influence policies. Overall, findings will empower NMCPs to achieve effective control and move them closer to elimination.

3. Specific Aims

3.1 Specific Aim 1: Model the spatial patterns of malaria risk through vector distribution and land use changes

3.2 Specific Aim 2: Model the spatial patterns of malaria risk through vector distribution and land use changes

4. Materials and Methods

4.1 Study Area: NSA comprising of ten countries: Bolivia, Brazil, Colombia, Ecuador, French Guiana, Guyana, Panama, Peru, Suriname and Venezuela. These countries account for approximately 90% of clinical cases in the region hence, the choice as study area (Fig. 1).

4.2 Study Area: NSA comprising of ten countries: Bolivia, Brazil, Colombia, Ecuador, French Guiana, Guyana, Panama, Peru, Suriname and Venezuela. These countries account for approximately 90% of clinical cases in the region hence, the choice as study area (Fig. 1).

4.3 Research Approach: Due to the complexity of malaria problem, I’m employing an interdisciplinary approach to address the problem (Fig. 2).

5. Preliminary Results

5.1 Areas of high to moderate risk corresponded with locations of some of the anophelines collected.

5.2 Risk scores for mosquito occurrence points were significantly higher than those generated randomly (Fig. 4).

6. Conclusion

Findings from preliminary results suggest that the MCE approach is a viable method to modeling spatial risk. The high resolution risk map produced aligned well with sampled vector points and may therefore be used to plan control of malaria vectors. Further analysis is planned to generate and validate risk maps with actual measures of malaria transmission, results of which could be used in the plan containment of future outbreaks.

References

Multi-scale Modeling and Assessment of Malaria Risk in Northern South America

Alimi, T. O.¹; Fuller, D. O.¹,² and Beier, J.C.¹,³

1 Introduction

Malaria as a public health problem has become a priority for control efforts worldwide. The global consensus is that its elimination is crucial for continual development. Ongoing research projects in different regions, including South America (SA), try to improve our understanding of the disease dynamics. Their goal is to establish a new framework that would lead to new intervention strategies for malaria elimination in areas where the disease is seasonal. One of such investigations is undertaken by the International Centers of Excellence in Malaria Research (ICEMR) under a National Institutes of Health grant.

While only about 3% of the global malaria burden is borne by SA, undertaking malaria research in the region is currently important because an estimated 23 million people are still at risk and approximately about 80% of clinical cases are found in Northern South America (NSA). A key factor limiting effective control is lack of data and uneven implementation of control measures, including use of bednets, sprays, early diagnosis, and treatment. As part of the ICEMR investigation, this project seeks to model the spatial patterns of malaria risk in NSA through vector distribution and land-use changes. Furthermore, I intend to investigate the perceptions of malaria risk in order to identify barriers to adoption and how they can be circumvented.

2 Significance

Spatial distribution of malaria risk is still perceived as broadly categorized by the WHO’s traditional risk maps which are highly generalized, of low resolution and have broad categories with uncertain boundaries (see da Nunez-Silva et al. 2012). There is need for up-to-date high resolution risk maps which can aid malaria control efforts. Secondly, modeling distribution of principal malaria vectors and land use changes which may explain the observed distribution and risk are useful tools which would guide future management strategies. Finally, understanding the perceptions of at-risk populations may help address barriers to adoption of interventions and influence policies. Overall, findings will empower NIMPs to achieve effective control and move them closer to elimination.

3 Aims

Specific Aim 1: Model the spatial patterns of malaria risk through vector distribution and land use changes.

- Hypothesis 1.1: GIS-based Multi-Criteria Evaluation (MCE) model can accurately predict spatial extent of malaria risk areas. Objective: Generate risk maps that represent risk of malaria transmission.
- Hypothesis 1.3: Land-use changes can explain the variations in predicted malaria risk. Objective: Characterize land use and land cover (LULC) and investigate changes in areas of risk.

Specific Aim 2: Investigate the perceptions of malaria risk in order to identify barriers to adoption and how they can be circumvented.

- Hypothesis 2.1: Knowledge of perception of malaria risk can aid design of malaria control strategies. Objective: Obtain and analyze data on subjective perceptions of risk.
- Hypothesis 2.2: Identification of barriers to adoption of malaria control interventions provide means of tackling them. Objective: Analyse data addressing perceived barriers and policy implications.

*Only ongoing work on Hypothesis 1.1 in presented here

4 Materials and methods

Materials: Raster data layers of environmental, climatic and anthropogenic parameters from satellite imagery, weather monitoring stations, global land cover and population data were collected from Worldclim, Digital Charts of the World, Globcover and Landiscor Vector data was collected from Field sampling by our collaborators and the Walter Reed Biosystematics Unit. Sociological data would be collected through questionnaires to be administered in one of the study areas. Other data will be collected as needed.

Procedures: To test hypothesis 1.1, raster data of parameters that influence mosquito distribution (rivers, wetlands, urban areas, roads, population and elevation) were combined using a Multi-Criteria Evaluation in IDRISI GIS package. This produced a map of potential exposure to malaria vectors which is used as a proxy for risk of malaria transmission. All the data layers were gridded at 1km spatial resolution. A set of distance layers had been created for discrete factors using standard GIS operations. All factors were subsequently standardized into a continuous common numeric range on a byte 0-255 probability scale using a fuzzy function based on knowledge of mosquito interaction with the factor. Weightings were generated for each factor based on the importance of the factor to malaria transmission by expert opinions and then assigned using Analytical Hierarchic Process. The risk maps produced were validated statistically using data on An. darlingi distribution and malaria case data from some parts of the study area. See preliminary results.

5 Preliminary results

Areas of high to moderate risk corresponded with locations of some of the anophelines collected.

Risk scores for mosquito occurrence points were significantly higher than those generated randomly.

6 Conclusion

Findings from preliminary results suggest that the MCE approach is a viable method to modeling spatial risk. The high resolution risk map produced aligned well with sampled vector points and may therefore be used to plan control of malaria vectors. Further analysis is planned to generate and validate risk maps with actual measures of malaria transmission, results of which could be used to plan containment of future outbreaks.

References

COMMON PITFALLS IN DATA VISUALIZATION
UNLABELED CHARTS

- Most common mistake
- Can make your visualization useless
UNLABELED CHARTS

- Most common mistake
- Can make your visualization useless

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com
USING 3-DIMENSIONAL CHARTS

Google tutorial searches
- Python
- JavaScript
- R
- Other

Google tutorial searches
- Python
- JavaScript
- R
- Other

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com
USING 3-DIMENSIONAL CHARTS

% Google searches

<table>
<thead>
<tr>
<th>Year</th>
<th>Python</th>
<th>JavaScript</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com
TOO MANY CATEGORIES

Google tutorial searches

Java

PHP

Python

C#

C

Javascript

Objective-C

Ruby

Swift

Matlab

VBA

Perl

lua
TOO MANY CATEGORIES

Google tutorial searches

Java
Other
PHP
Python
BAR CHARTS NOT STARTING AT ZERO

Google tutorial searches for JavaScript

Google tutorial searches for JavaScript

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com
FAILING TO NORMALIZE DATA

1000s of Python users in...

<table>
<thead>
<tr>
<th>City</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orlando</td>
<td>20</td>
</tr>
<tr>
<td>NYC</td>
<td>60</td>
</tr>
<tr>
<td>Chicago</td>
<td>30</td>
</tr>
</tbody>
</table>

% of Python users in...

<table>
<thead>
<tr>
<th>City</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orlando</td>
<td>8%</td>
</tr>
<tr>
<td>NYC</td>
<td>2%</td>
</tr>
<tr>
<td>Chicago</td>
<td>4%</td>
</tr>
</tbody>
</table>

Note: Fake data

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com
GOOD PRACTICES IN DATA VISUALIZATION
WHEN APPROPRIATE, SHOW THE DATA

Programming language popularity

StackOverflow Rank vs GitHub Rank
WHEN APPROPRIATE, SHOW THE DATA

Programming language popularity

StackOverflow Rank

GitHub rank

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com
Google tutorial searches

Python  JavaScript  R
Google tutorial searches

- Python
- JavaScript
- R

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com
LABEL OBJECTS AND DATA DIRECTLY

Google tutorial searches

- Python: 10%
- JavaScript: 7%
- R: 3%

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com
LABEL OBJECTS AND DATA DIRECTLY

Google tutorial searches

- Python: 10.4%
- JavaScript: 7.0%
- R: 2.6%

Dr. Randy Olson | Email: rso@randalolson.com | Twitter: @randal_olson | Web: randalolson.com