EXPLORING SPATIAL NON-STATIONARITY THROUGH REGIONAL MODELING

A CASE STUDY OF DOG CANCER UNDERREPORTING

Gianluca Boo [1,2] gianluca.boo@geo.uzh.ch

Stefan Leyk [3] stefan.leyk@colorado.edu

Sara I. Fabrikant [1] sara.fabrikant@geo.uzh.ch

Andreas Pospischil [2] pospischil@collegium.ethz.ch

[1] Department of Geography, University of Zurich, Switzerland
[2] Collegium Helveticum, University of Zurich & Swiss Federal Institute of Technology, Zurich, Switzerland
[3] Geography Department, University of Colorado at Boulder, Boulder, CO, USA

Summary

Spatial non-stationarity occurs when a global model cannot capture statistical associations to variables changing across space [1]. For instance, models of dog cancer spatial distribution are typically affected by spatial non-stationarity because of regional variations in disease reporting practices within the study area.

Through spatial analysis of model residuals, we build regions of similar statistical performance [2]. Regional models are then fit to inform about changes in spatial non-stationarity in the statistical associations. This regional framework is meant to produce improved models of disease with regard to underreporting.

1. Introduction

Companion animals and humans share similar exposures at the community level. Dogs can thus be used as **environmental sentinels** for several human cancers [3]. However, comparative studies of dog and human cancers are still rare because underreporting impacts most canine cancer data sources.

Switzerland has the largest and most durable canine cancer registry at the country level. This exceptional data source is affected by **potential underreporting**, which is expected to vary according to regional settings [4]. For this reason, we test a regional framework for fitting dog cancer incidence rates.

2. Methods

We fit dog cancer incidence rates in Switzerland for 2008 through a **Poisson regression model**. We use the following variables: Dog Average Age, Female Dog Ratio, Average Income Tax, Human Population Density, and Distance to Veterinary Care [5]. We assess statistical performance through the McFadden pseudo-R^2 and the percentage of variance reduction of each variable [6].

Based on the model residuals, we **build regions** using a connectivity graph algorithm [7]. These regions have the maximum internal similarity and external dissimilarity [8]. We fit the Poisson regression model to each region, separately, and explore potential improvements from the global model with regard to spatial non-stationarity.

3. Results

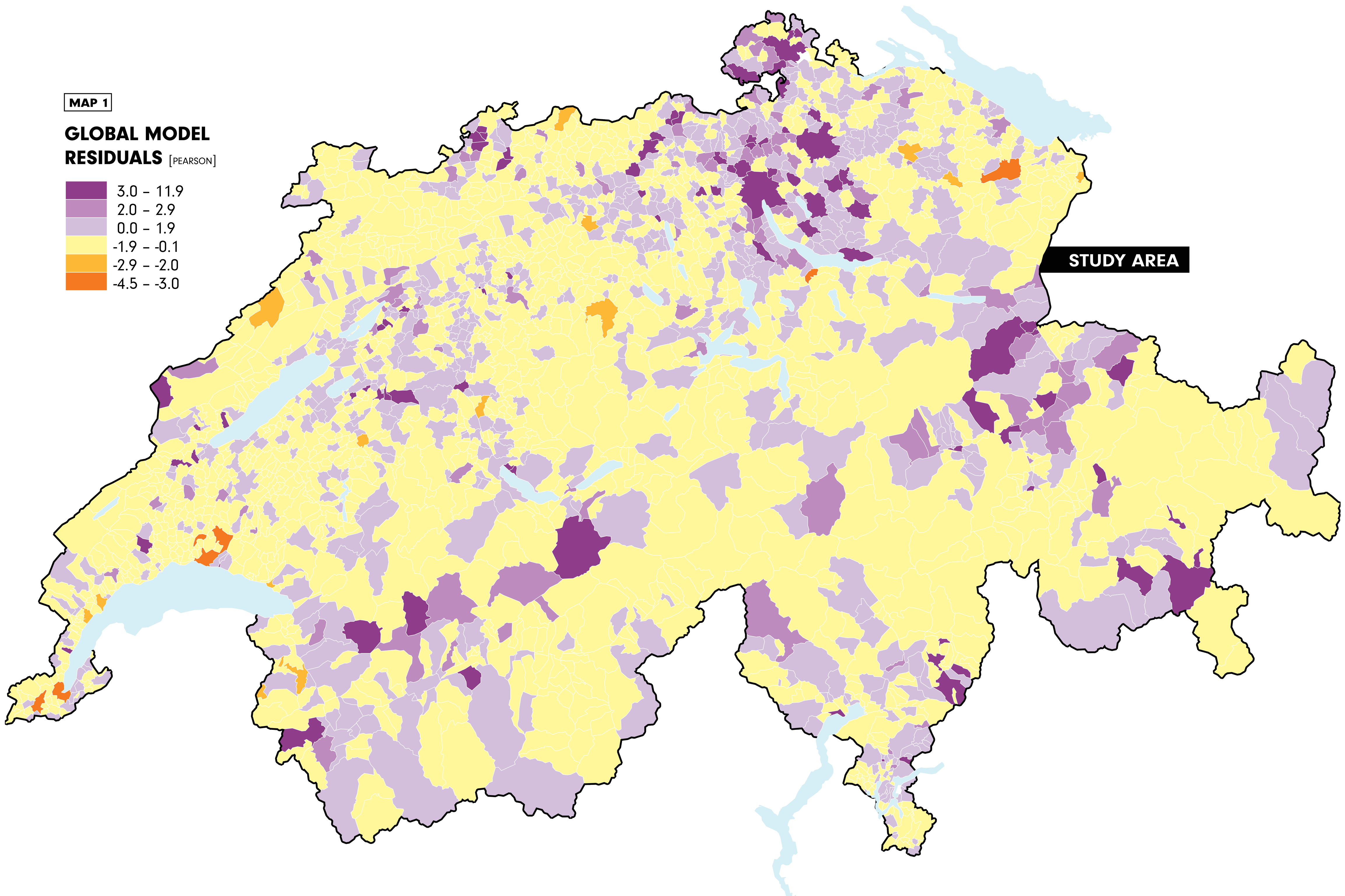
MAP 1 shows that the global model [R^2=0.50] based on the STUDY AREA produces several regions of poor model fit, characterized by units with absolute residual values above two. Most of these regions disappear when fitting the two regional models depicted in MAP 2. The improvements seem to occur mostly in REGION 1 [R^2=0.58] rather than in REGION 2 [R^2=0.46].

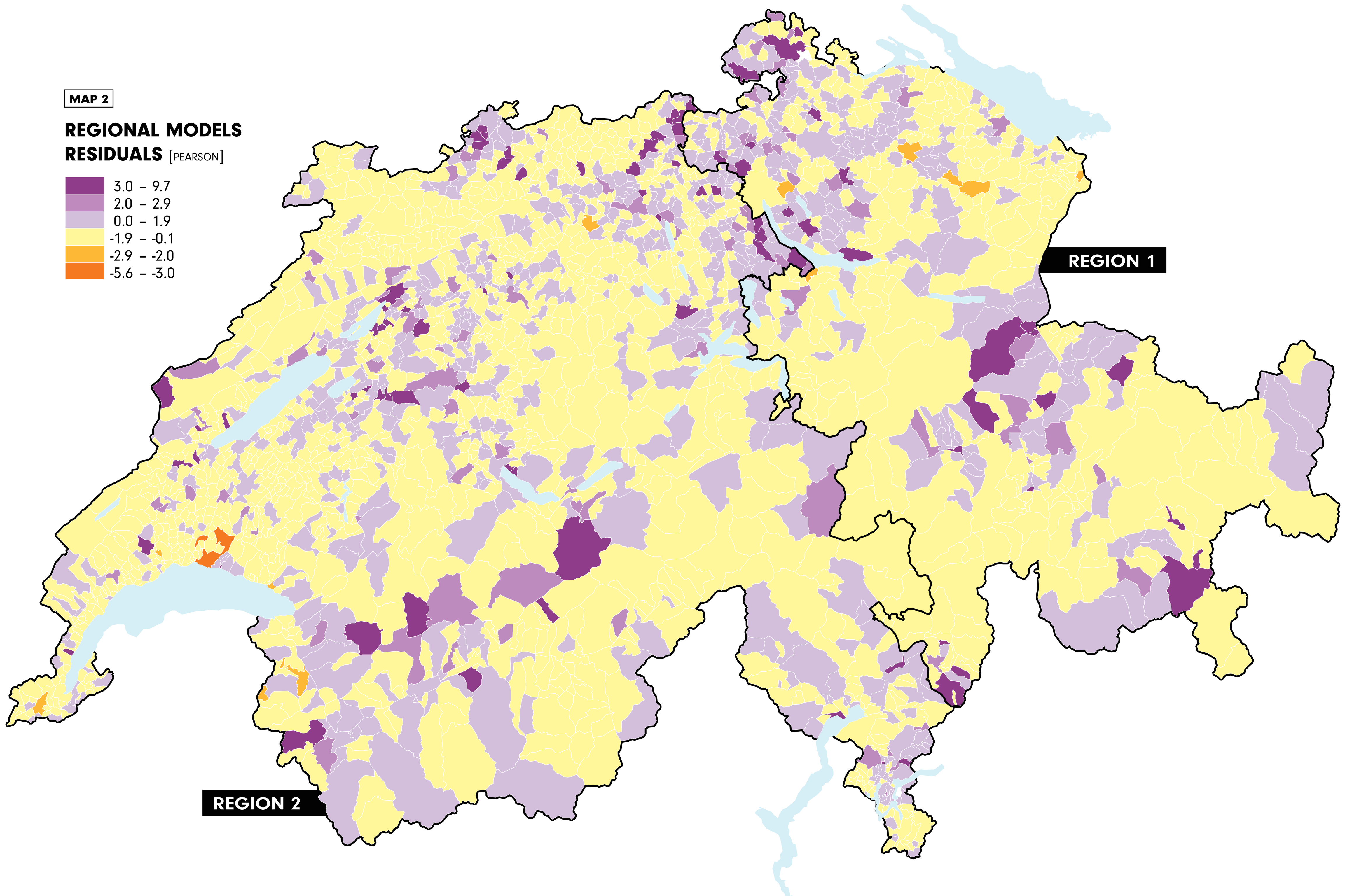
CHART 1 shows the percentage of variance reduction of the variables accounting for dog demographic characteristics [DOG AVERAGE AGE, FEMALE DOG RATIO] and potential underreporting [AVERAGE INCOME TAX, HUMAN POPULATION DENSITY, DISTANCE TO VETERINARY CARE]. The latter contribute to 86.4% of variance reduction in the global model based on the STUDY AREA, 71.6% in the model fit to REGION 1 and 95.5% in the model fit to REGION 2.

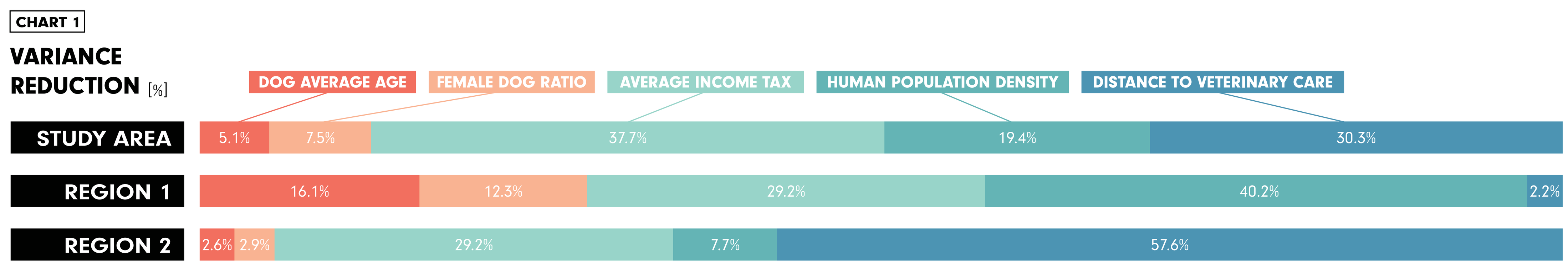
4. Conclusions

Compared to the global model, the model fit to Region 1 shows **improved statistical performance**, most likely due to reduced underreporting. Conversely, the model fit to Region 2 shows deteriorated performance, due to increased underreporting. This results in less reliable statistical associations to dog cancer incidence rates.

The implementation of this basic regional framework confirmed the need for systematic **investigations of regional changes** in statistical associations. In future studies, we intend to refine this framework to deepen the understanding of spatial non-stationarity in models of dog cancer with regard to underreporting.







References

[1] Fotheringham AS, Charlton M, Brunsdon C. 1996. The geography of parameter space: an investigation of spatial non-stationarity. Geogr Inf Syst. 10:605–627.

[2] Leyk S, Norlund PU, Nuckols JR. 2012. Robust assessment of spatial non-stationarity in model associations related to pediatric mortality due to diarrheal disease in Brazil. Spat Spatio-Temporal Epidemiol. 3:95–105.

[3] Reif JS. 2011. Animal Sentinels for Environmental and Public Health. Public Health Rep. 126:50–57.

[4] Boo G, Fabrikant SI, Leyk S. 2015. A Novel Approach to Veterinary Spatial Epidemiology: Dasymetric Refinement of the Swiss Dog Tumor Registry Data. ISPRS Ann Photogramm Remote Sens Spat Inf Sci. II-3/W5:263–269.

[5] Frome EL. 1983. The Analysis of Rates Using Poisson Regression Models. Biometrics. 39:665–674.

[6] Cameron AC, Windmeijer FAG. 1997. An R-squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models. J Econom. 77:329–342.

[7] Duque JC, Ramos R, Suriñach J. 2007. Supervised Regionalization methods: A Survey. Int Reg Sci Rev. 30:195–220.

[8] Ketchen DJ, Shook CL. 1996. The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. Strategic Manage J. 17:441–458.