WASABY

Work Package 6

Tina Žagar, Sara Korat, Vesna Zadnik

2nd Steering Committee Meeting
16th December 2020

CANCER REGISTRY
REGISTER RAKA
SLOVENIJA

ONKOLOŠKI INŠTITUT
INSTITUTE OF ONCOLOGY
LJUBLJANA

Work Package 6

- Objectives

  To determine methodologies suitable to perform spatial analysis with cancer data, using breast cancer incidence as an example.

- Deliverables
  - ✓ D6.1 – Report on methods for spatial analysis of the cancer data
    D6.2 – Practical manual for CR's personnel
    D6.3 – Report on spatial analysis of breast cancer incidence data from interested CRs

- Milestones
  - ✓ M6.1 – Creation of WP-6 group of experts on spatial analysis

M6.1 – WASABY group of experts on spatial analysis

- The experts on spatial analysis were identified through already established work and also according to the participation in the WASABY project or other past international collaborating actions on spatial analysing of the cancer registries' data.

Name	Affiliation	Country	e-mail adress
Tina Žagar	Slovenian Cancer Registry Inctitute of Oncology Ljubljana Zaloška 2 1000 Ljubljana	Slovenia	tzagar@onko-i.si
Vesna Zadnik	Slovenian Cancer Registry Inctitute of Oncology Ljubljana Zaloška 2 1000 Ljubljana	Slovenia	vzadnik@onko-i.si
Andreja Kukec	Faculty of Medicine University of Ljubljana Vrazov trg 2 1000 Ljubljana	Slovenia	andreja.kukec@mf.uni-lj.si
Ron Pritzkuleit	Institute for Cancer Epidemiology Universitat zu Lubeck Ratzeburger Allee 160 23538 Lubeck	Germany	ron.pritzkuleit@uksh.de
Marc Colonna	Isere Cancer Registry CHU de Grenoble – Pavillon E BP 217 38043 Grenoble Cedex 9	France	mcolonna.registre@wanadoo.fr
Joséphine Bryere	Universite de Caen Normandie 3 Avenue Général Harris 14076 Caen Cedex 05	France	josephine.bryere@gmail.com
Ludivine Launay	Universite de Caen Normandie 3 Avenue Général Harris 14076 Caen Cedex 05	France	ludivine.launay@inserm.fr

Name	Affiliation	Country	e-mail adress
Fortunato Bianconi	Umbria Cancer Registry University of Perugia Via del Giochetto 06126 Perugia	Italy	fortunato.bianconi@gmail.com
Maurizio Zarcone	Palermo Cancer Registry Palermo University Hospital Via del Vespro 133 90127 Palermo	Italy	registrotumoripalermo@unipa.it
Martina Bertoldi	Varese Province Cancer Registry Fondazione "Istituto Nazionale Tumori" Via Venezian 1 20133 Milan	Italy	martina.bertoldi@istitutotumori.mi.it
Eero Pukkala	Finnish Cancer Registry Institute for Statistical and Epidemiological Cancer Research Unioninkatu 22 FI-00130 Helsinki	Finland	eero.pukkala@cancer.fi
Marc Saez	Universitat de Girona Plaça Sant Domènec, 3 Edifici Les Àligues 17004 Girona	Spain	marc.saez@udg.edu

D6.1 – Report on methods for spatial analysis of the cancer data

Table of contents

WASABY

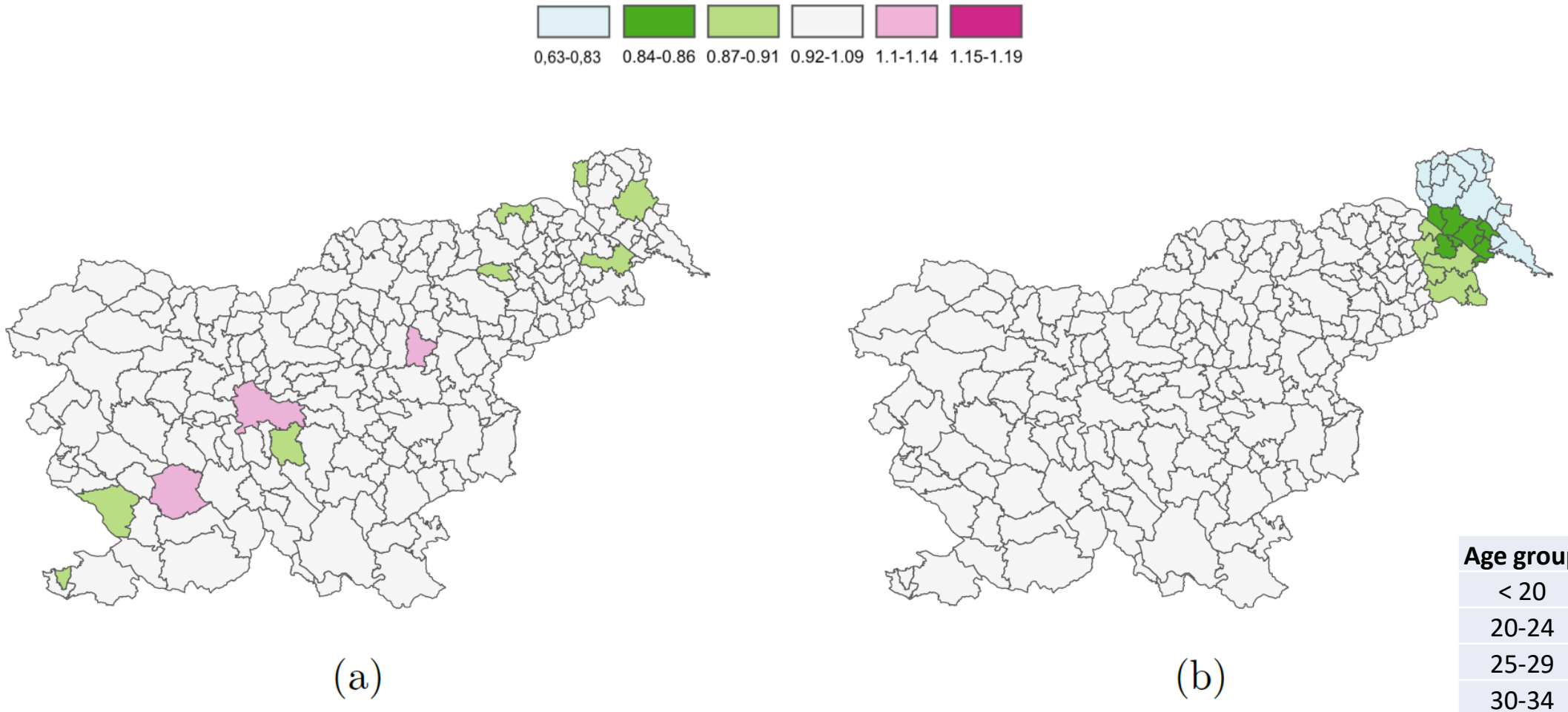
“INLA study“ – comparison with Gibbs

- We perfomed the comparisson of two Bayesian spatial smoothing methods: INLA and Gibbs (package *NIMBLE* in R, uses MCMC).
- We used the BYM model with the CAR distribution for the spatial component.

One approach to calculating the smoothed values of relative risk is with the established MCMC methods, using the Gibbs sampler, whereas the other is to use the approximative method INLA.

- We studied properties and application on data of the Slovenian Cancer Registry.

INLA and Gibbs

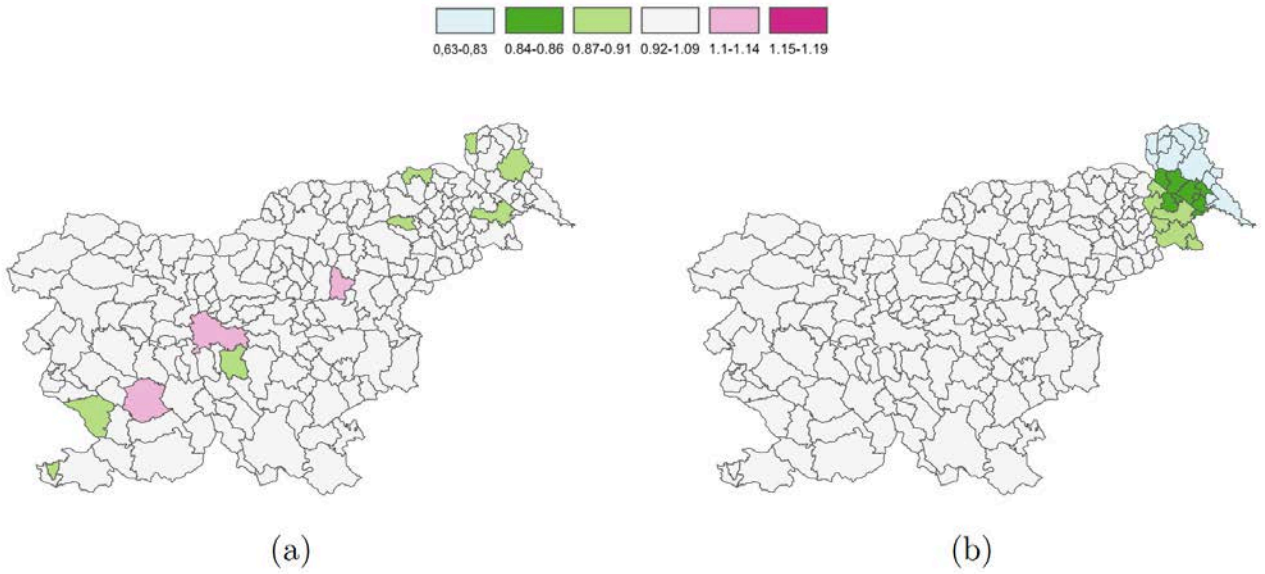


Smoothed SIR of breast cancer by Slovenian municipalities among women, aged 0 to 49. (a) with INLA (b), with Gibbs (2006–2015).

Age group	No of cases
< 20	0
20-24	9
25-29	51
30-34	199
35-39	368
40-44	712
45-49	1125
Total	**2464**

INLA and Gibbs

- When using the Gibbs sampler to calculate smoothed values, the calculations take a long time, unlike when using INLA, as INLA is an approximative method.
- In the example: INLA shows no clusters and estimates higher values.



WASABY

INLA and Gibbs

Based on real data we can‘t decide which method performs better. In the second part we made a comparison of both methods on generated data, trying to generally answer the question on how the methods behave in different situations. When using the methods on real and generated data, we was able to observe minimal differences between them.

SIMULATED DATA:

- Generated and permutated municipalities‘ population data
- Based on chosen relative risk (known expected incidence for each municipality) the incidence was generated (Poisson distribution)
- EDI and shapes of Slovenia and municipalities remained unchanged.

- INLA and Gibbs are compared using:
  - squared error of deviation from known relative risk for each municipality
  - Squared error for Slovenia
  - Average of squared errors

D6.2 – Practical manual for CR's personnel

- „Based on the report on analysed breast cancer incidence data a practical handbook is prepared (including details on data preparation) for CR's personnel and other researchers aiming to promote spatial analysis of CR data in general.“

- The practical manual will :
  - follow the content of D6.1 – Report on methods for spatial analysis of the cancer data,
  - include theoretical background,
  - include practical instruction for preparing the datasets,
  - also be the manual for using the program.

D6.3 – Report on spatial analysis of breast cancer incidence data from interested CRs

- Not all CRs are producing the maps themselves.
- We can help those CRs by making the program that is free and easy to use.
    - For demonstration, we will prepare reports for CRs submitting the datasets using this program.
    - But the CRs will be able also to use the programe themselves for other datasets with the help of Practical manual for CR's personnel.

D6.3 – Report on spatial analysis of breast cancer incidence data from interested CRs

- Is written in R and Shiny.

  Shiny is an R package that makes it easy to build interactive web apps straight from R.

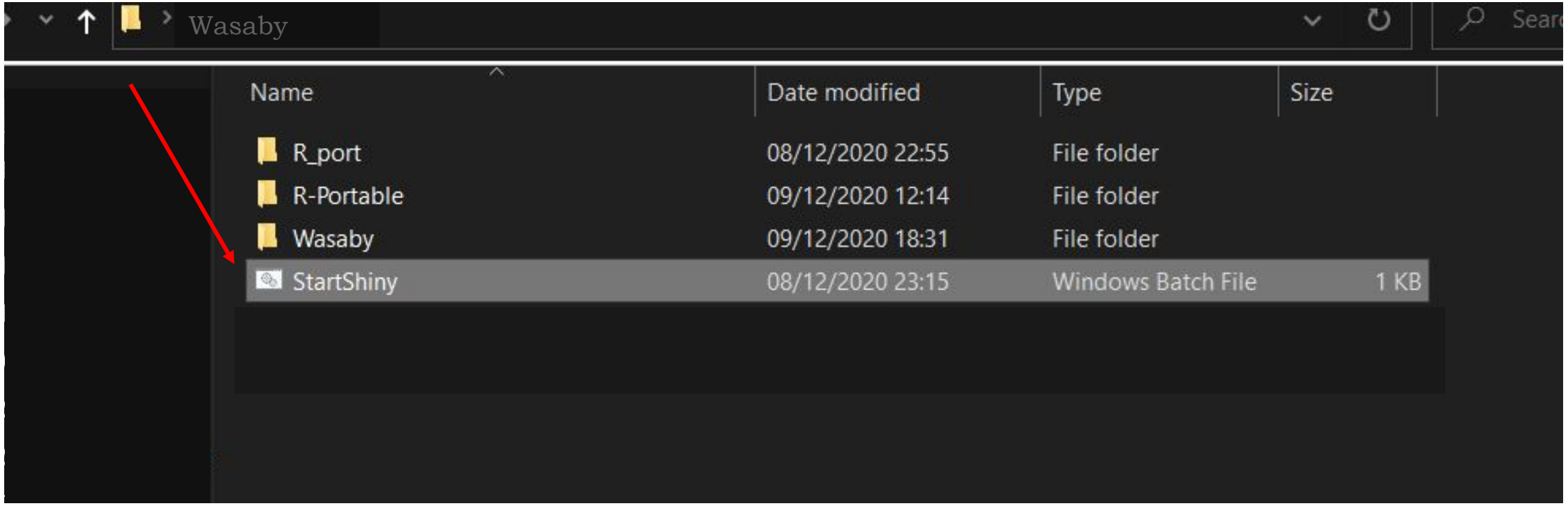
- The code is open – anyone can change it.

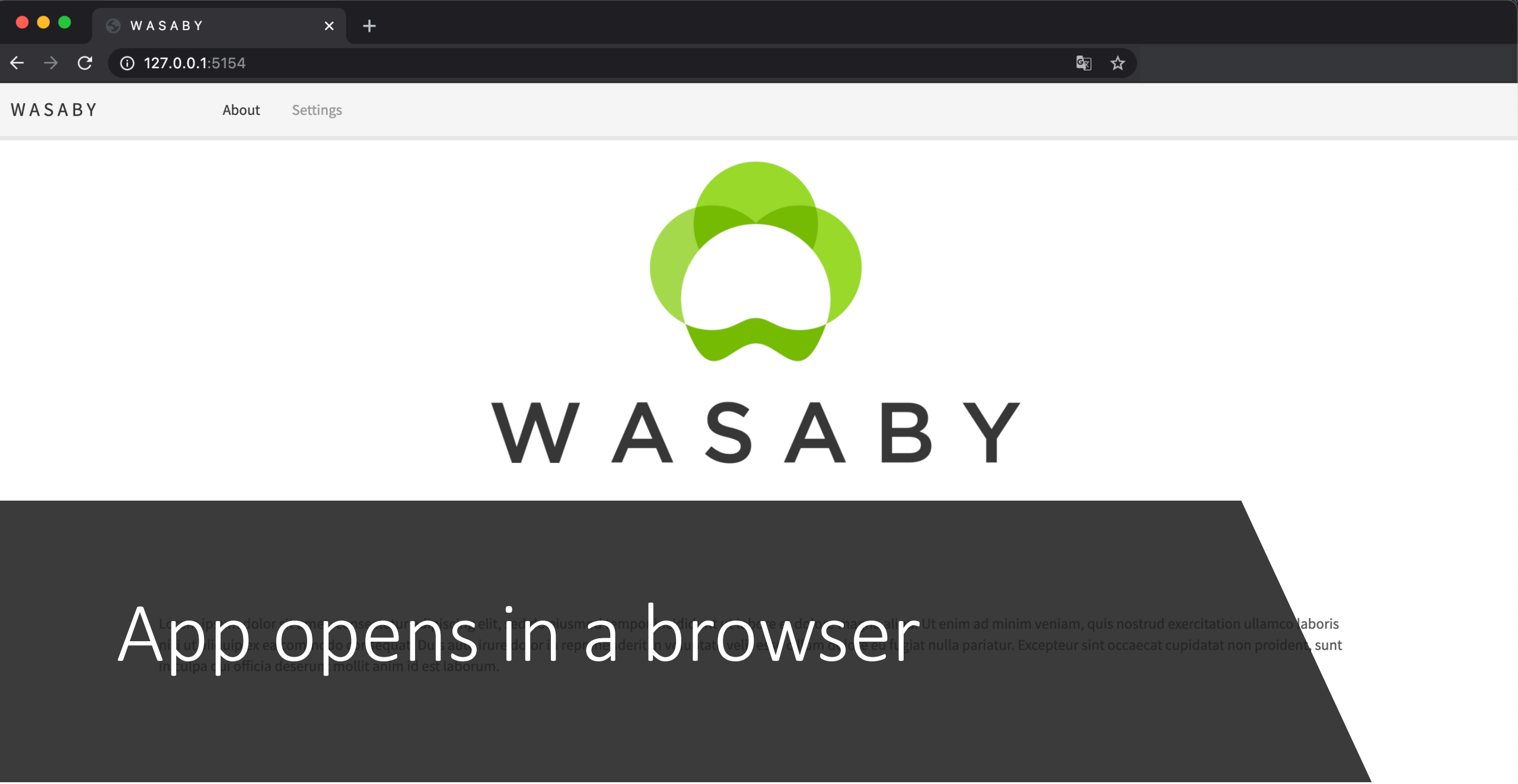
- Most of the development till now was put in the generality:
  - different datasets can be used,
  - errors are easy to understand and are meant to guide the user.

Program for drawing maps for CRs

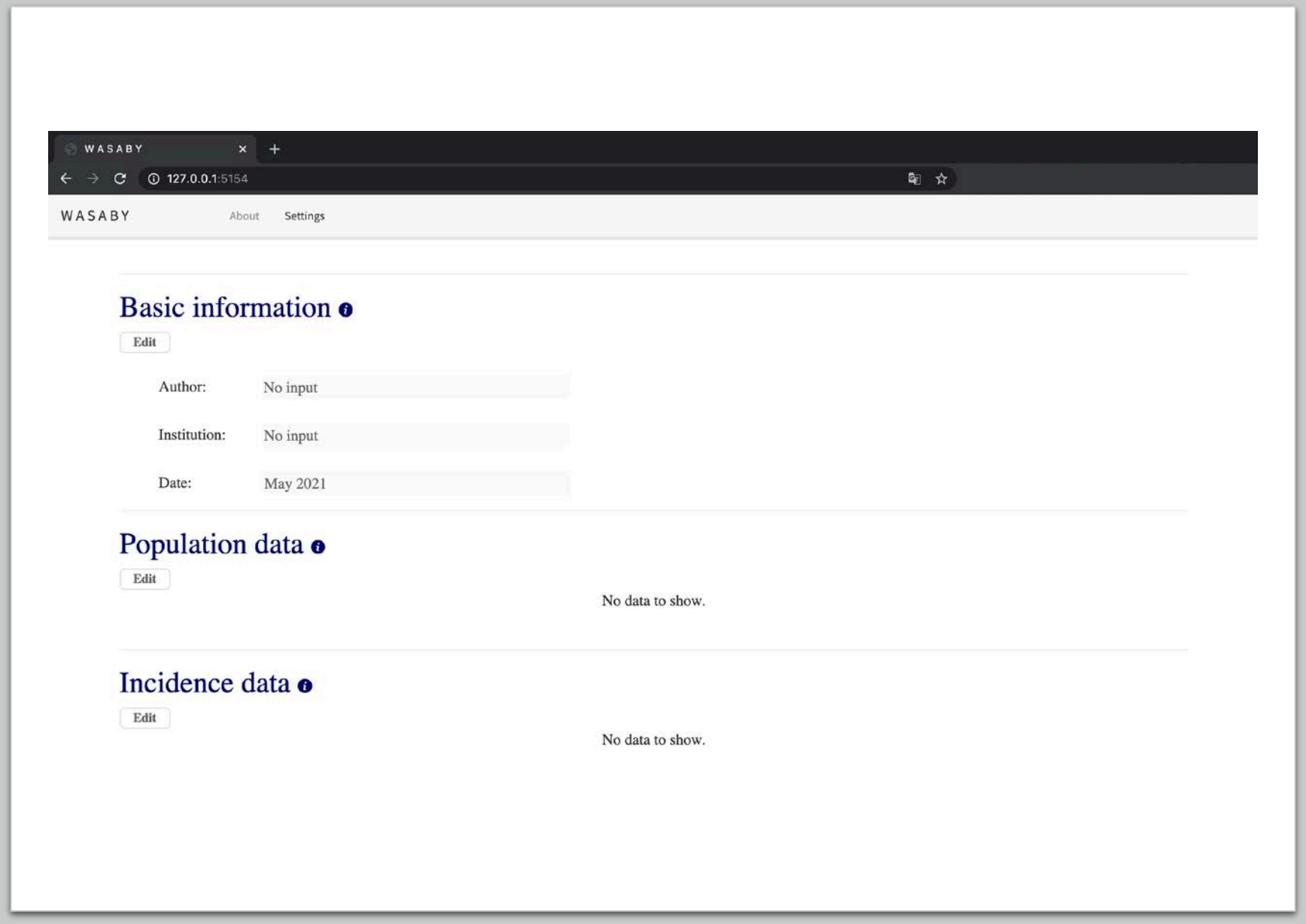
Install and run

Run StartShiny

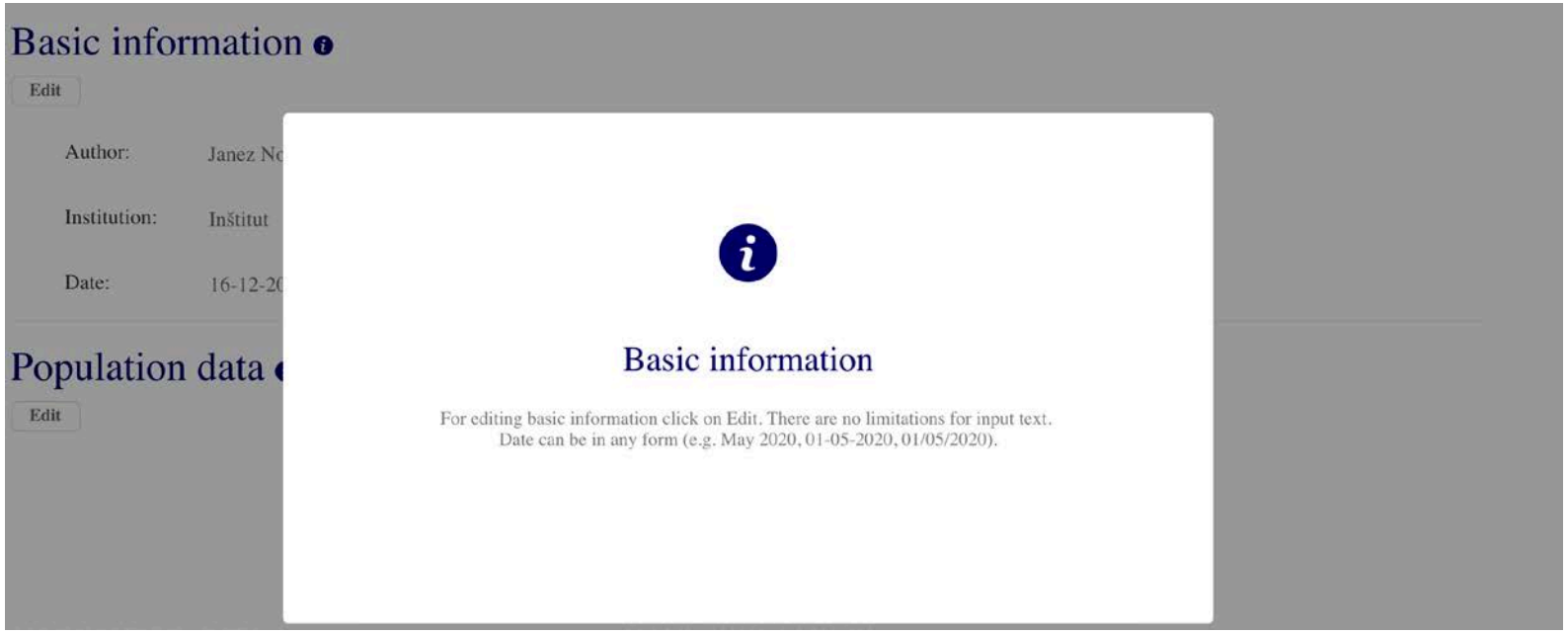



WASABY
127.0.0.1:5154
WASABY
About
Settings
WASABY
App opens in a browser

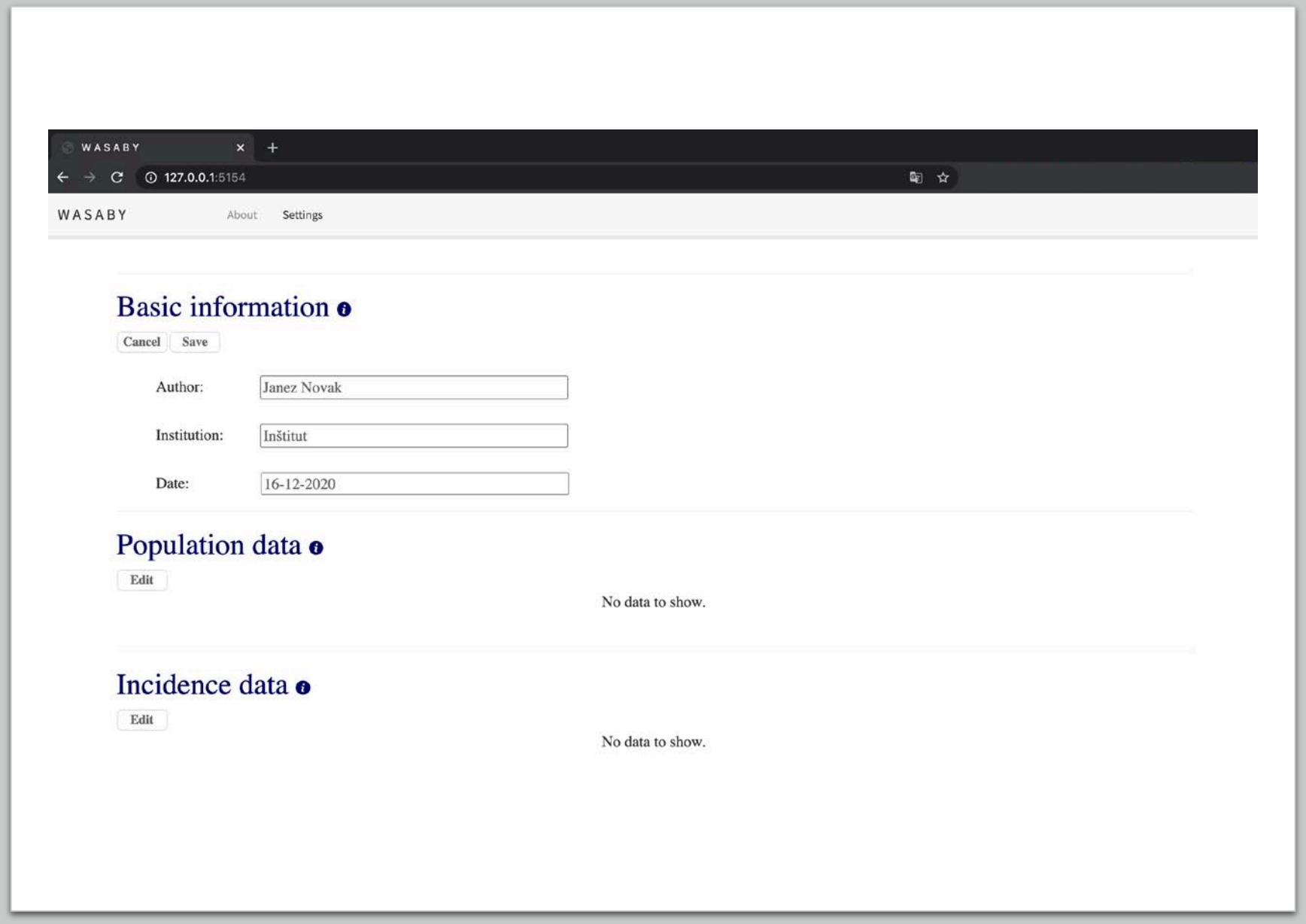
In the Settings tab you can change the required parameters.



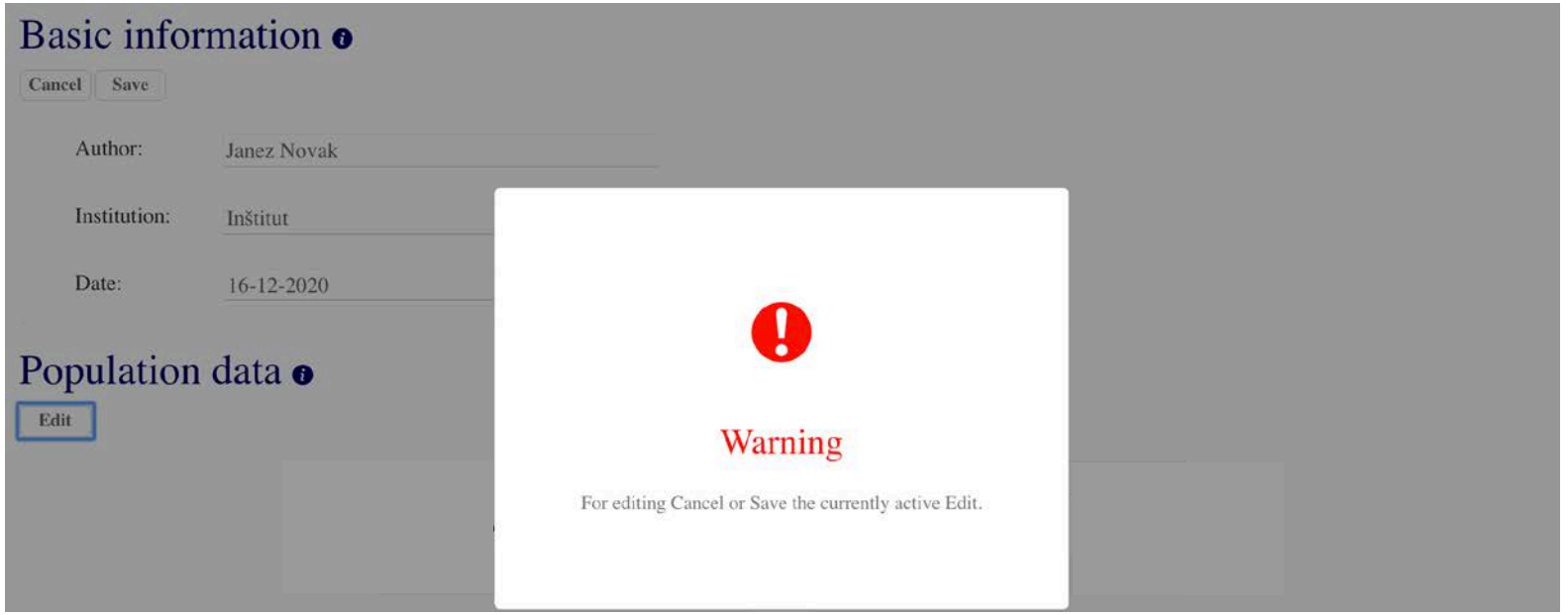
For more information about each category, click the blue “i” button.



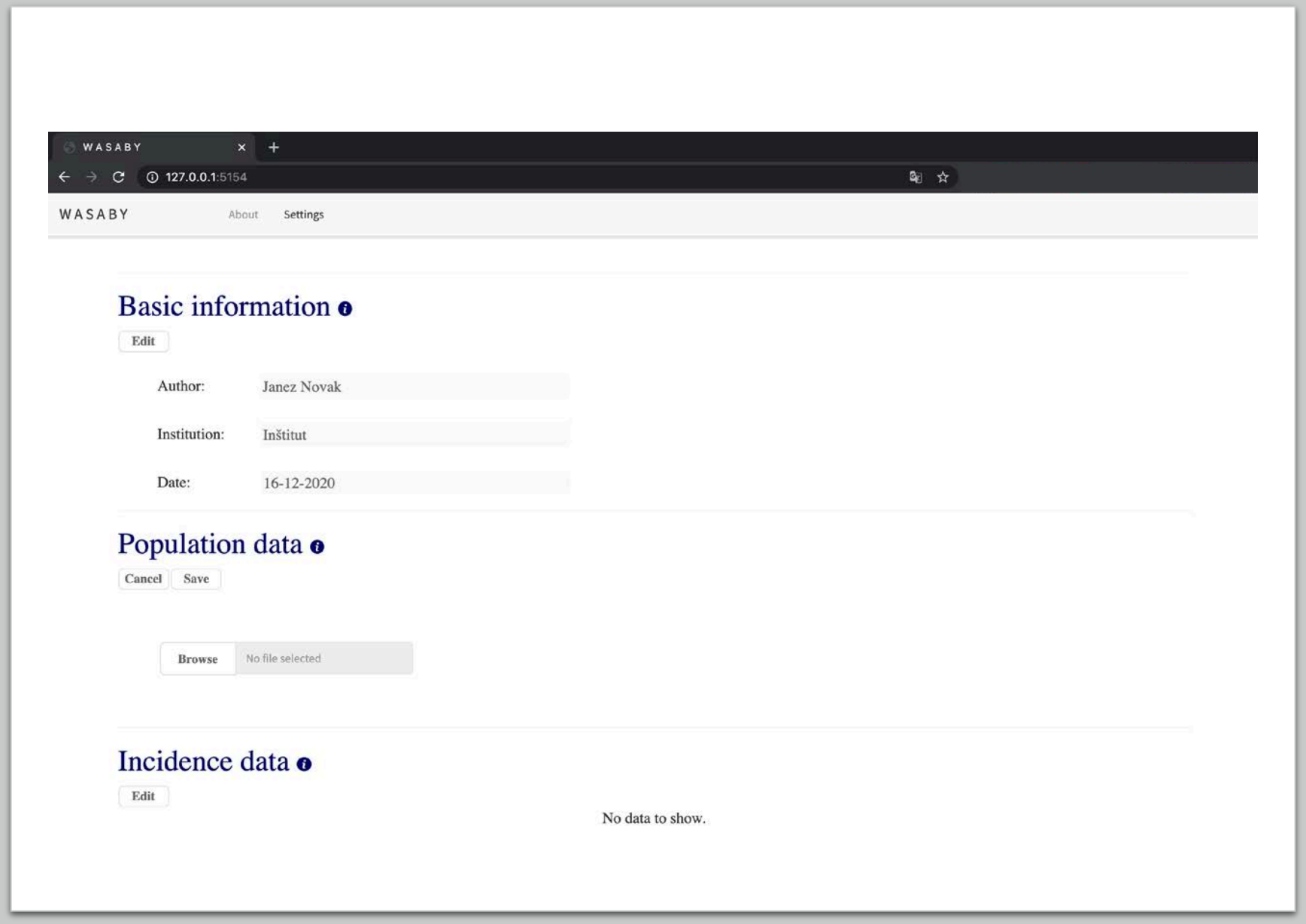
For editing data in each section click on Edit. After editing click Save.



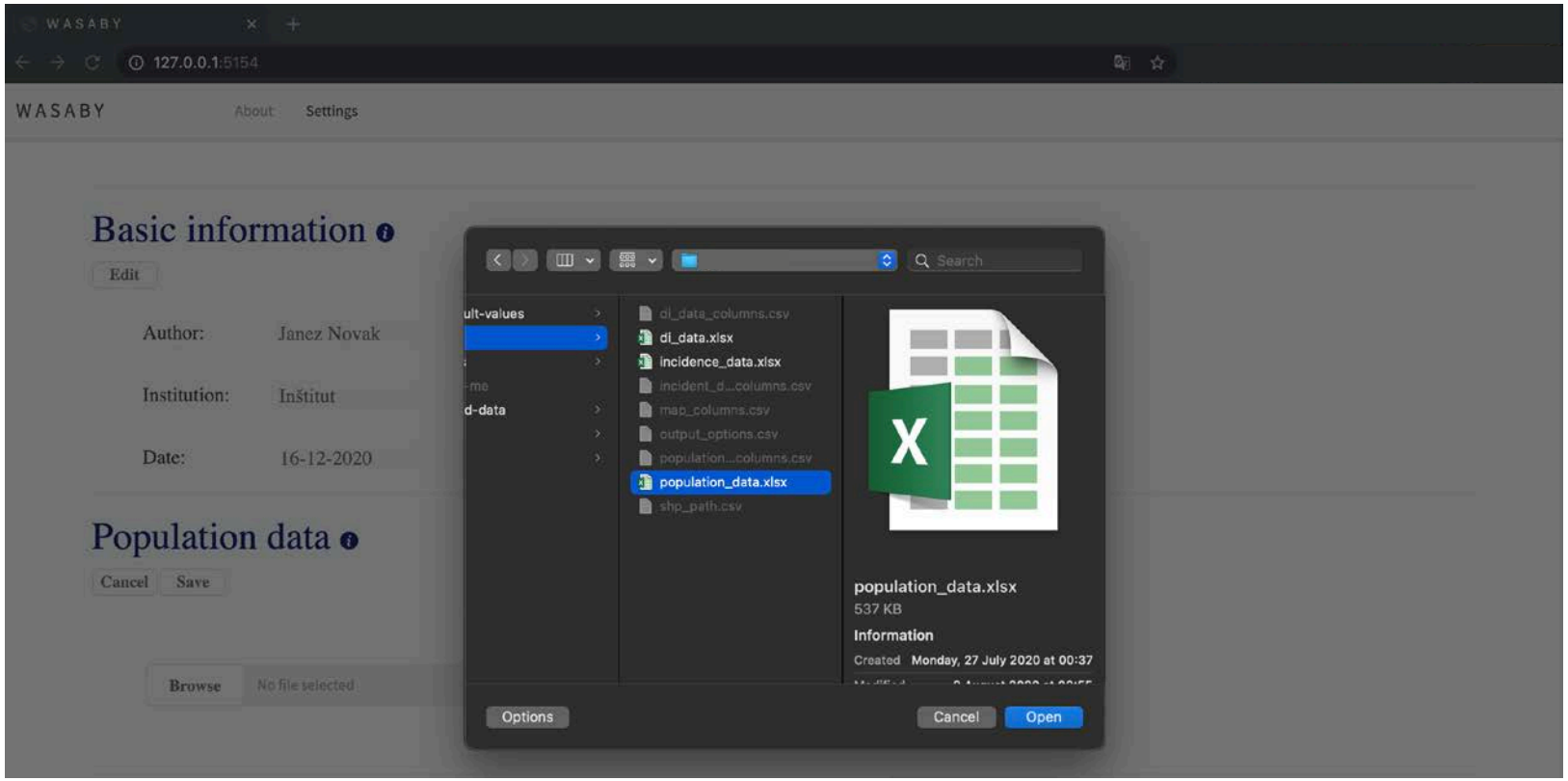
You can edit only one section at a time.



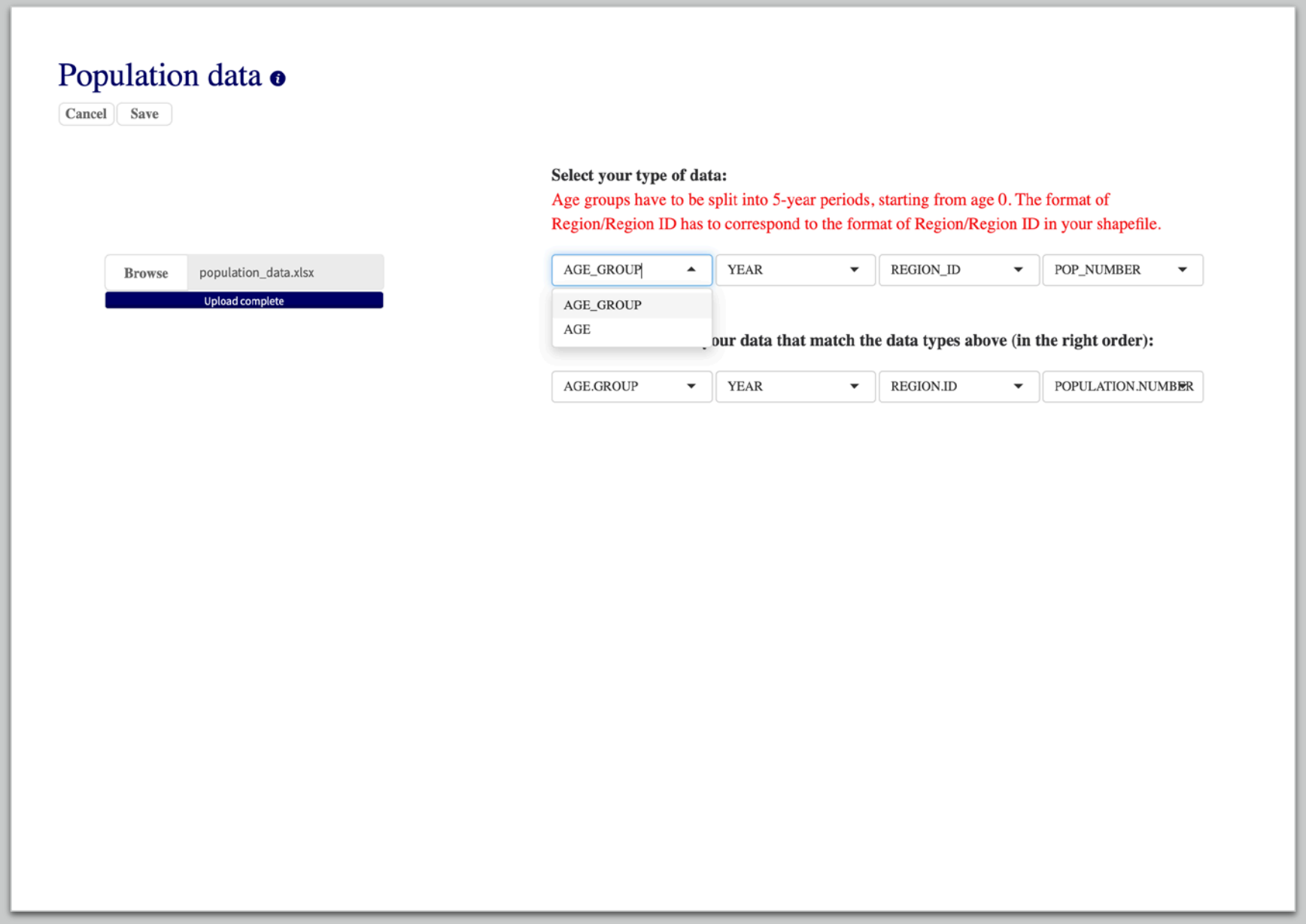
Click on a browse will open a new window from where you can select your file.



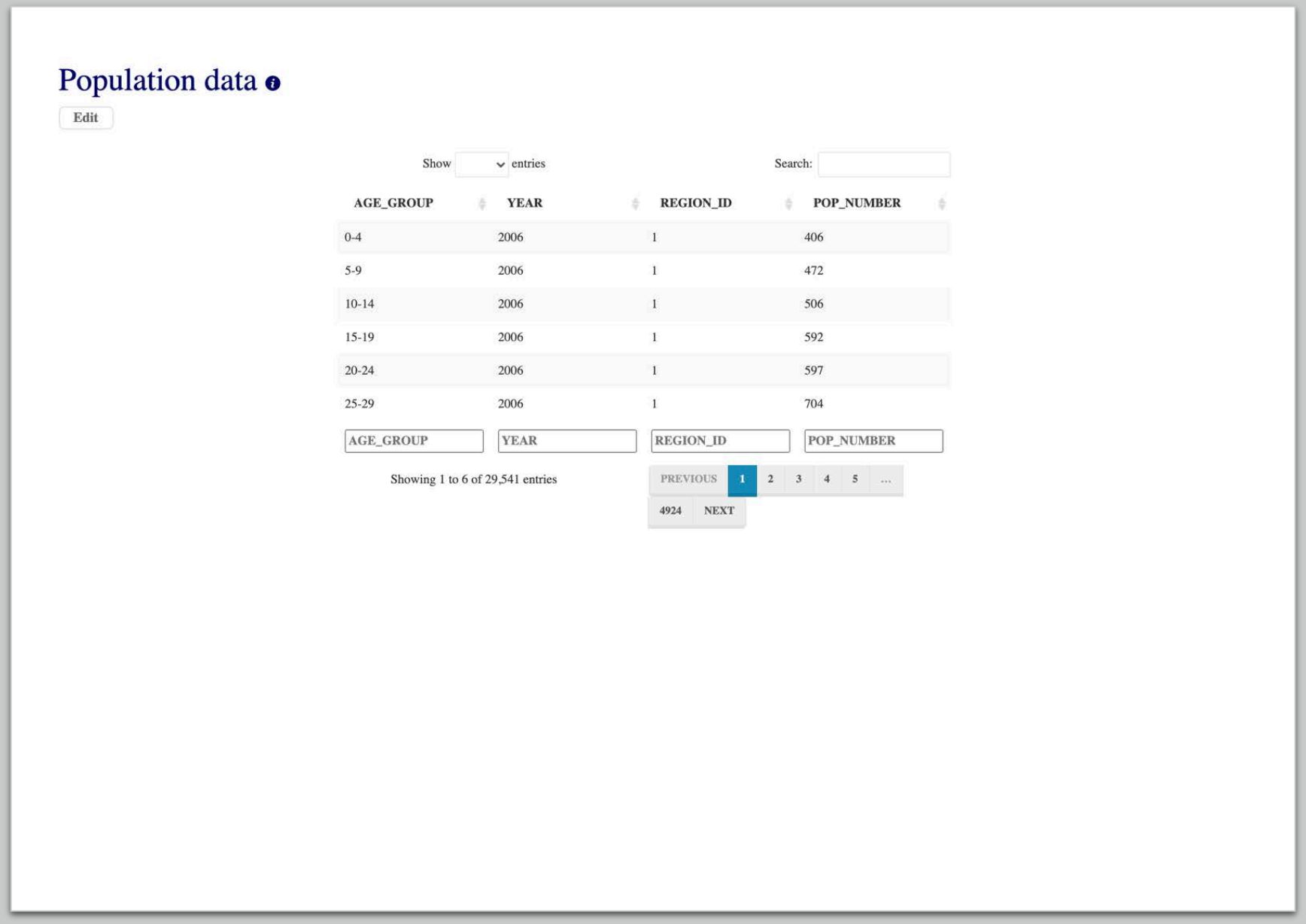
Upload .xlsx file.



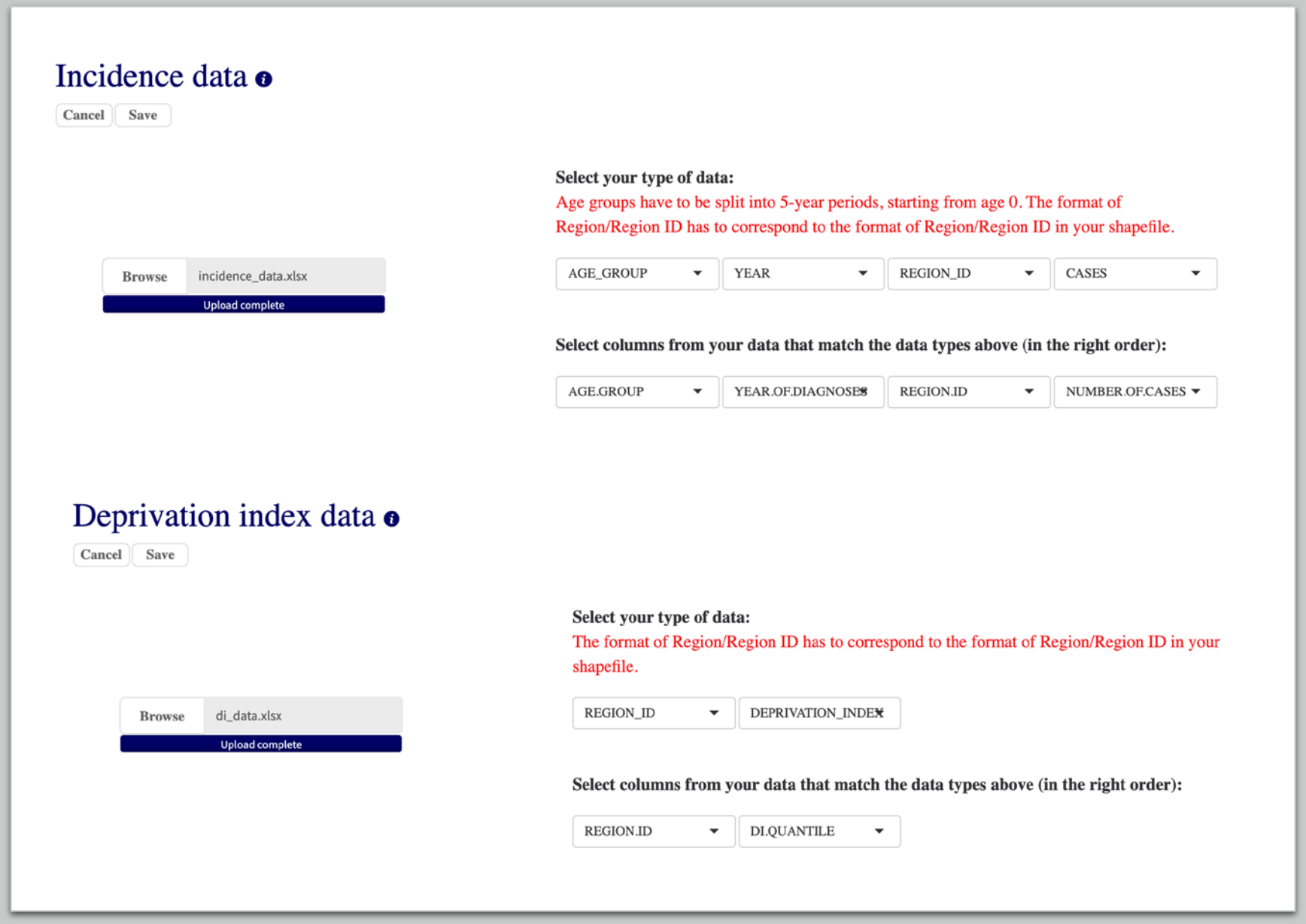
Select the type of data and the corresponding column names in your data.



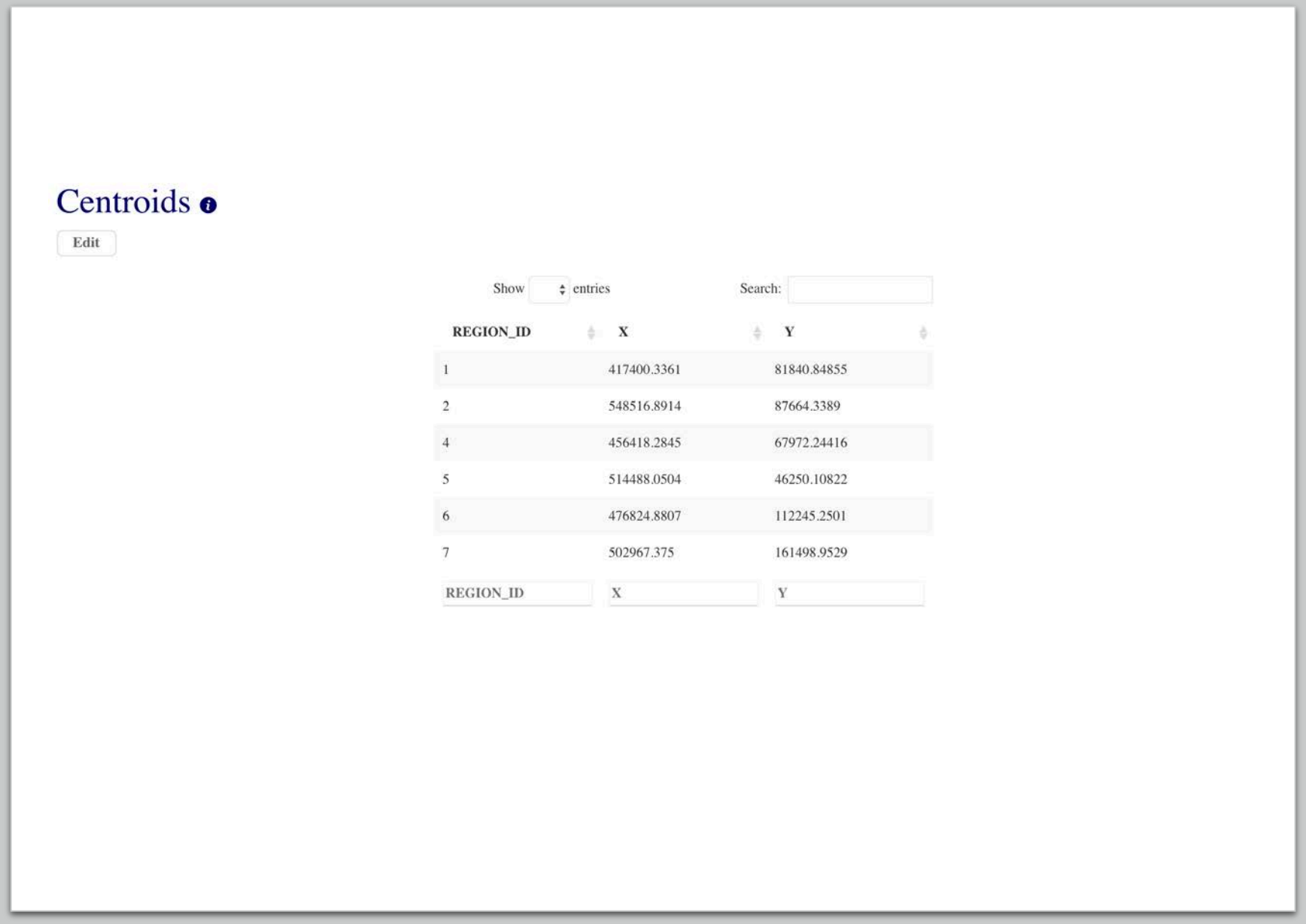
After saving the uploaded data, the preview of your data will be shown.



Do the same for incidence data and deprivation index data.



Centroids are optional – depending on whether you want to draw a Finnish smoothed maps.



Centroids

Edit

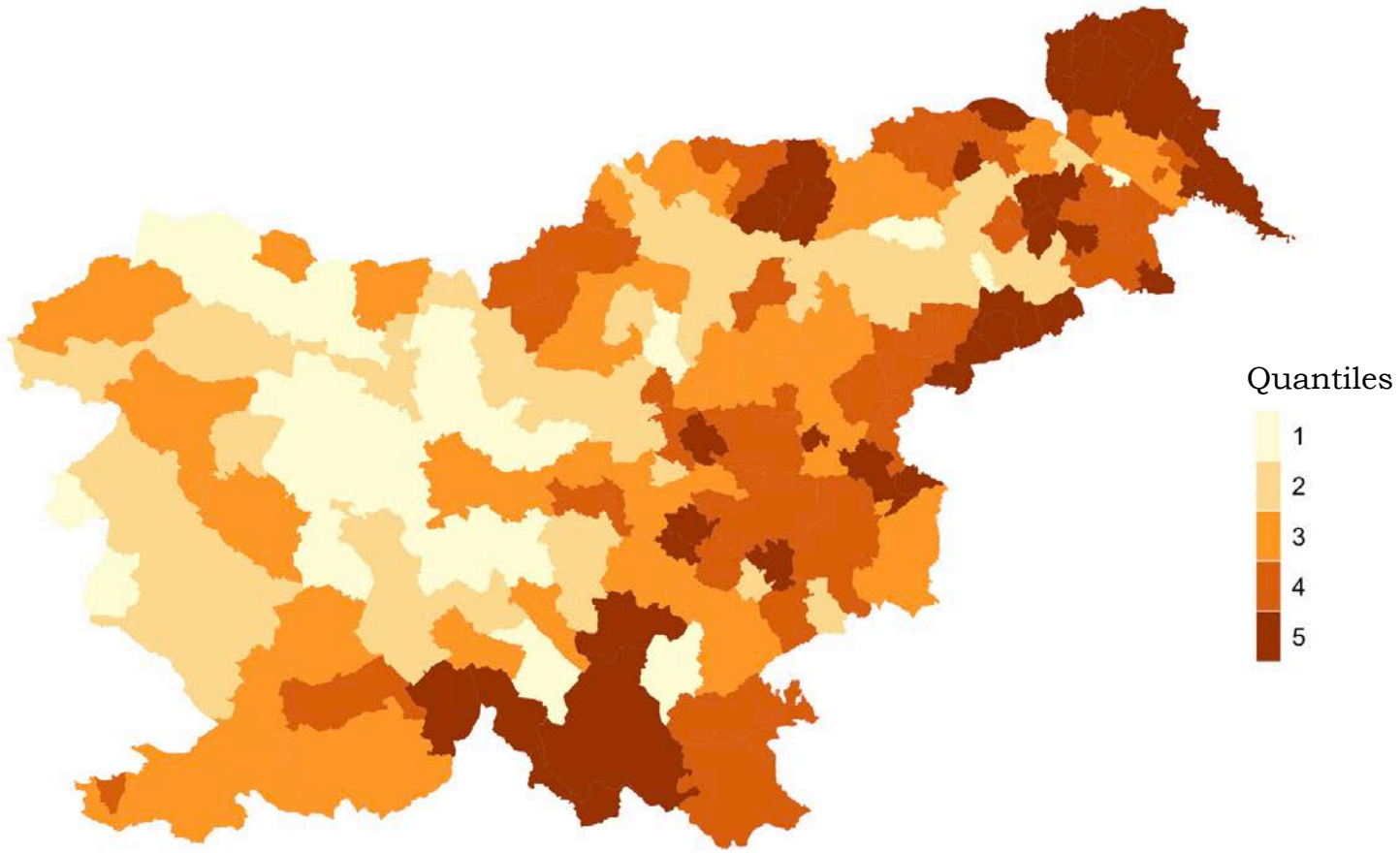
Show entries Search:

REGION_ID	X	Y
1	417400.3361	81840.84855
2	548516.8914	87664.3389
4	456418.2845	67972.24416
5	514488.0504	46250.10822
6	476824.8807	112245.2501
7	502967.375	161498.9529
REGION_ID	X	Y

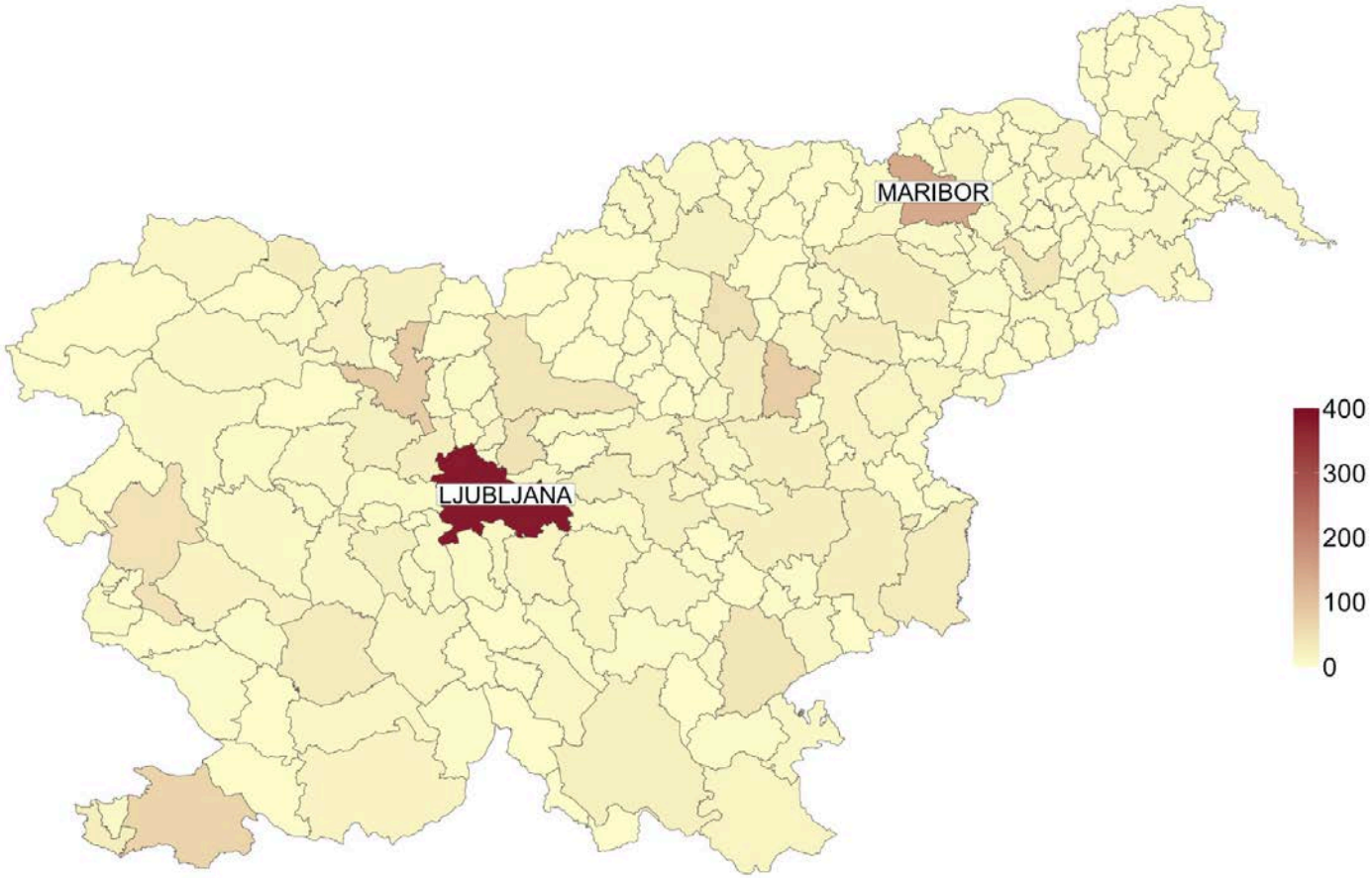
After choosing a destination folder with a shapefile, application checks if a folder you selected contains the right files. After selecting a destination folder, chose a variable, or variables, that correspond with region data in previously uploaded data.

After clicking on the Output button, wait for the output files to be generated. Files will be generated in the Home folder.

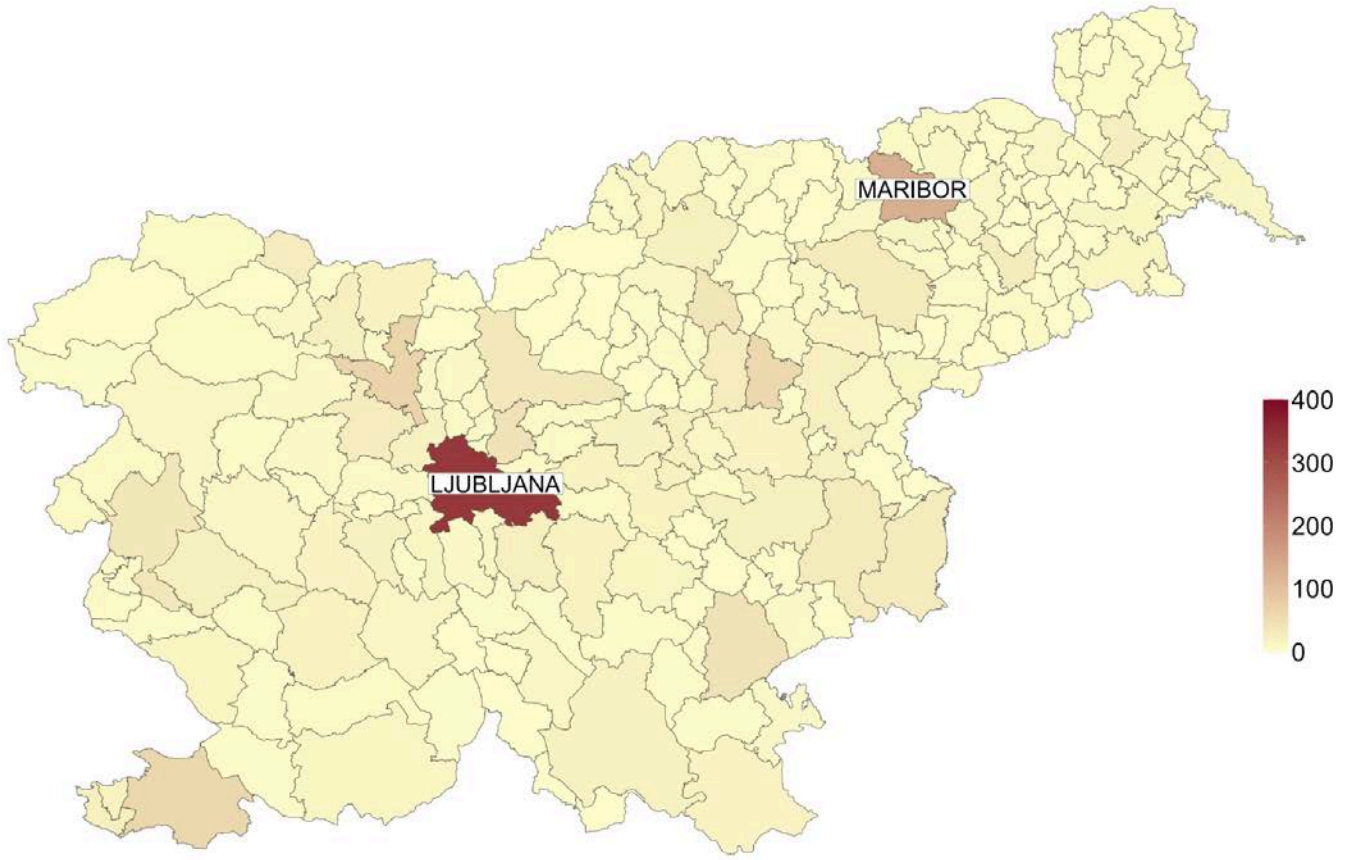
Output example – Breast cancer in Slovenia (2006-2015, age: 0-49)



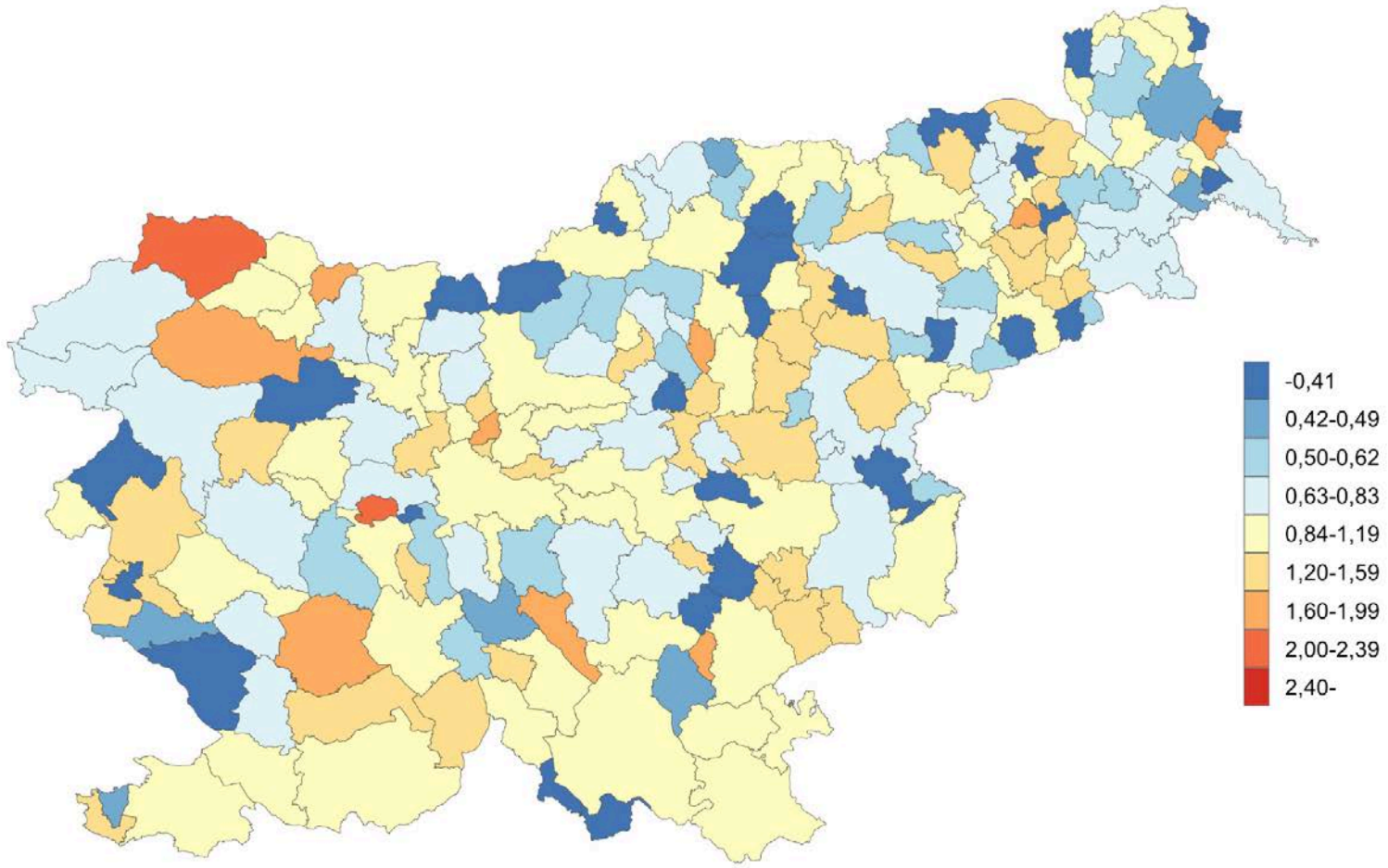
The distribution of quantile values of deprivation index by Slovenian municipalities for year 2011.



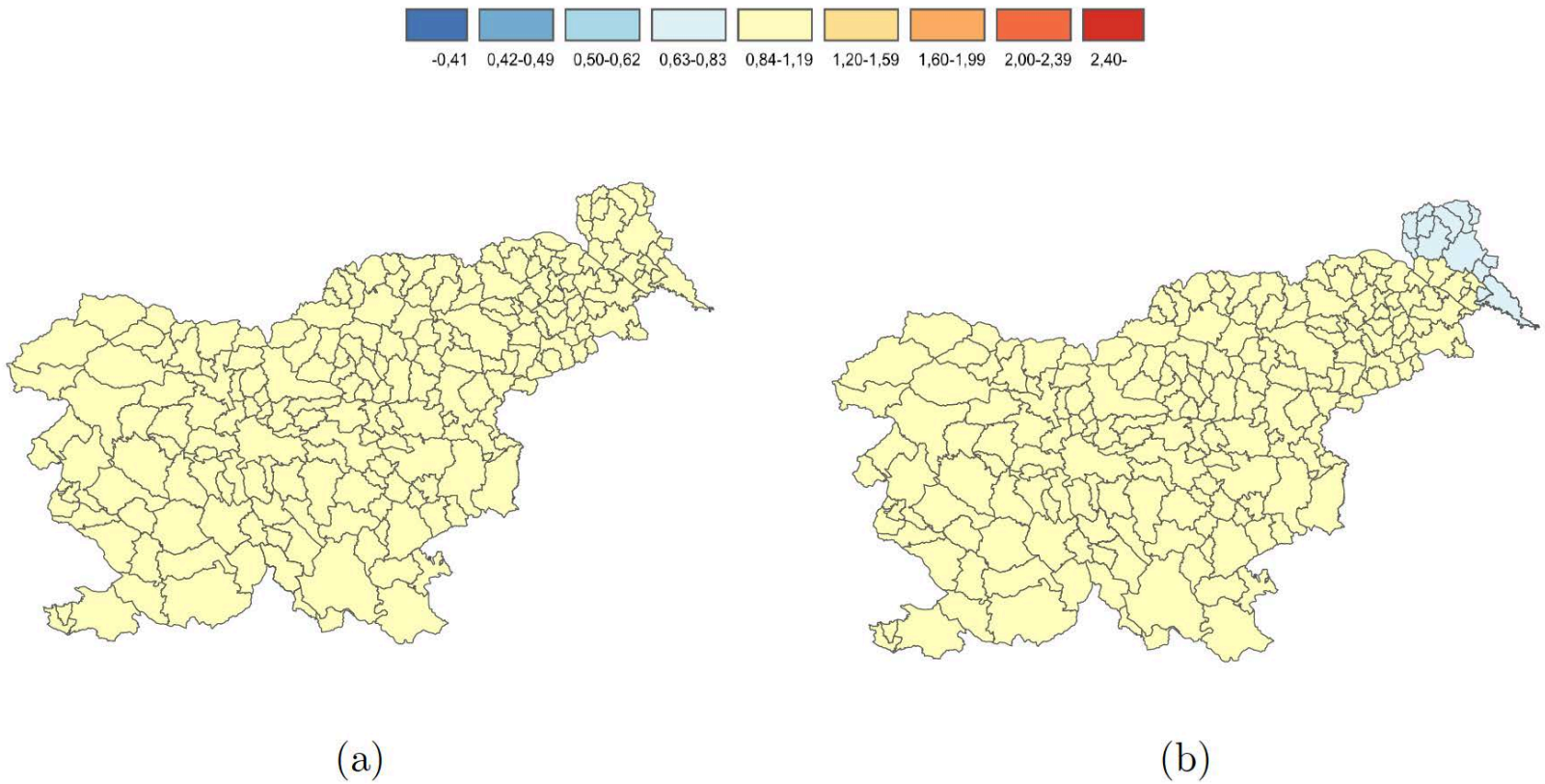
Unsmoothed absolute observed incidence of breast cancer among women, aged 0 to 49 for 211 municipalities in Slovenia (2006–2015).



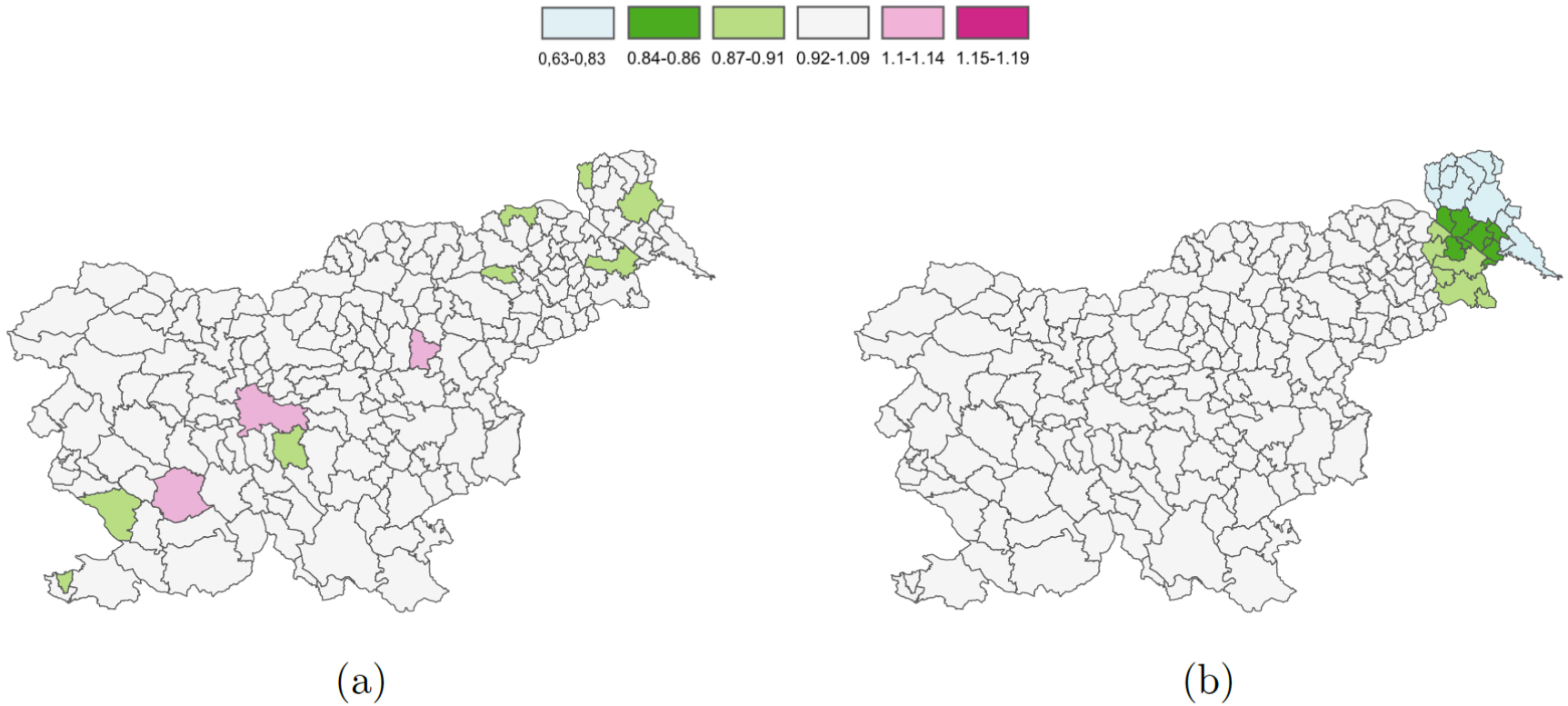
Unsmoothed absolute expected incidence of breast cancer among women, aged 0 to 49 for 211 municipalities in Slovenia (2006–2015).



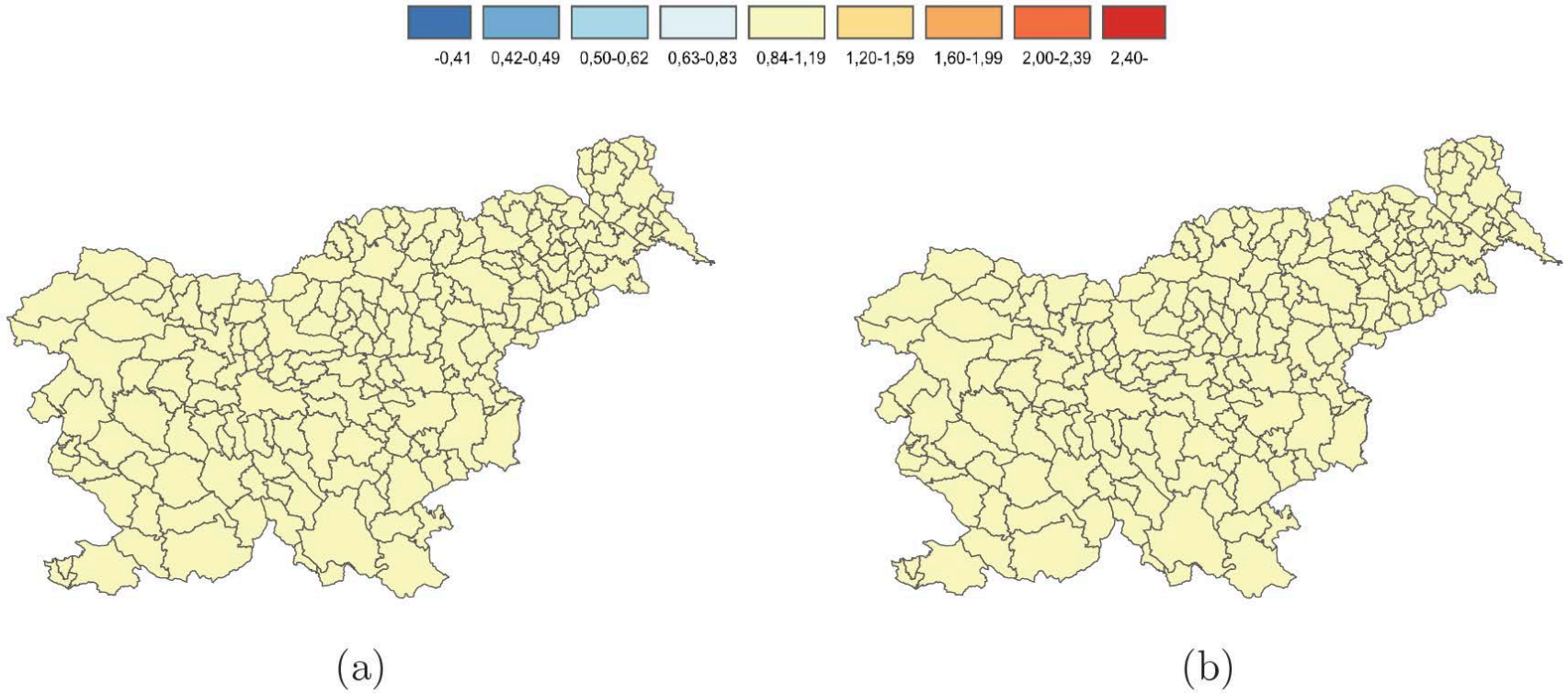
Standard incidence ratio (SIR) for breast cancer among women, aged 0 to 49, Slovenia (2006–2015).



Smoothed SIR of breast cancer by Slovenian municipalities among women, aged 0 to 49. (a) with INLA (b), with Gibbs (2006–2015).



Smoothed SIR of breast cancer by Slovenian municipalities among women, aged 0 to 49. (a) with INLA (b), with Gibbs (2006–2015).
Results shown on a different scale (0, 63–1, 19).



Smoothed SIR of breast cancer by Slovenian municipalities among women, aged 0 to 49, conditioned on the deprivation index. (a) with INLA (b), with Gibbs (2006–2015).

It‘s timely …

- INLA is fast but Buggs is slow. In datasets with most areas (around 2000) Gibbs takes 7-10 minutes.
- But plotting the graphs takes even longer time. Plotting all the graphs takes more time than all calculations together.
- The user chooses which results he wishes to produce, so he doesn't need to wait for long every time.
- Errors for checking the source data are instant.

To be included in the final version:

- Choosing age standardization.
- Choosing deprivation index centiles (10%, 20% ...)
- Selection of output destination (currently everything is saved in *Home*).
- Output preferences: select which data, which map you want to save, select the colour scale for maps ...
- Smoothing: INLA, Gibbs, floating weighted averages method (Finnish smoothing method).
- Option for uploading a .csv file.
- Notification when output is created.
- …

The next steps

- Sending the beta version of the program to the group of experts by the end of 2020:
  - proposals for improvement,
  - finding the name for the program (for example *Mapping for CRs*, but nicer).
- Preparing final version of the program.

- Writing practical manual for CRs.

- Distributing to the participating CRs:
  - practical manual,
  - results from the program for their CR,
  - the program.

- Putting the program and practical manual on the WASABY‘s page.
- On-line course on the program.

WASABY