

# rOpenGov project: a community of R package developers on open government data analytics

Aalto HELDIG DH pizza seminar 6.5.2022



Pyry Kantanen, Leo Lahti



UNIVERSITY  
OF TURKU

# About rOpenGov

- Started in 2010
- Winning entry in Apps4Finland 2011 competition
- Core team: Leo Lahti, Markus Kainu, Joona Lehtomäki, Juuso Parkkinen, Pyry Kantanen
- Finnish collaborators: Janne Huovari (PTT), Jussi Paananen, Juho Kopra, Tuomo Nieminen (THL) among others

Joona, Leo and Markus representing rOpenGov in useR! 2017 conference in Brussels



# International collaborators

Scott Chamberlain (US)

- rsunlight
- rtimes
- enigma

Thomas Leeper (UK)

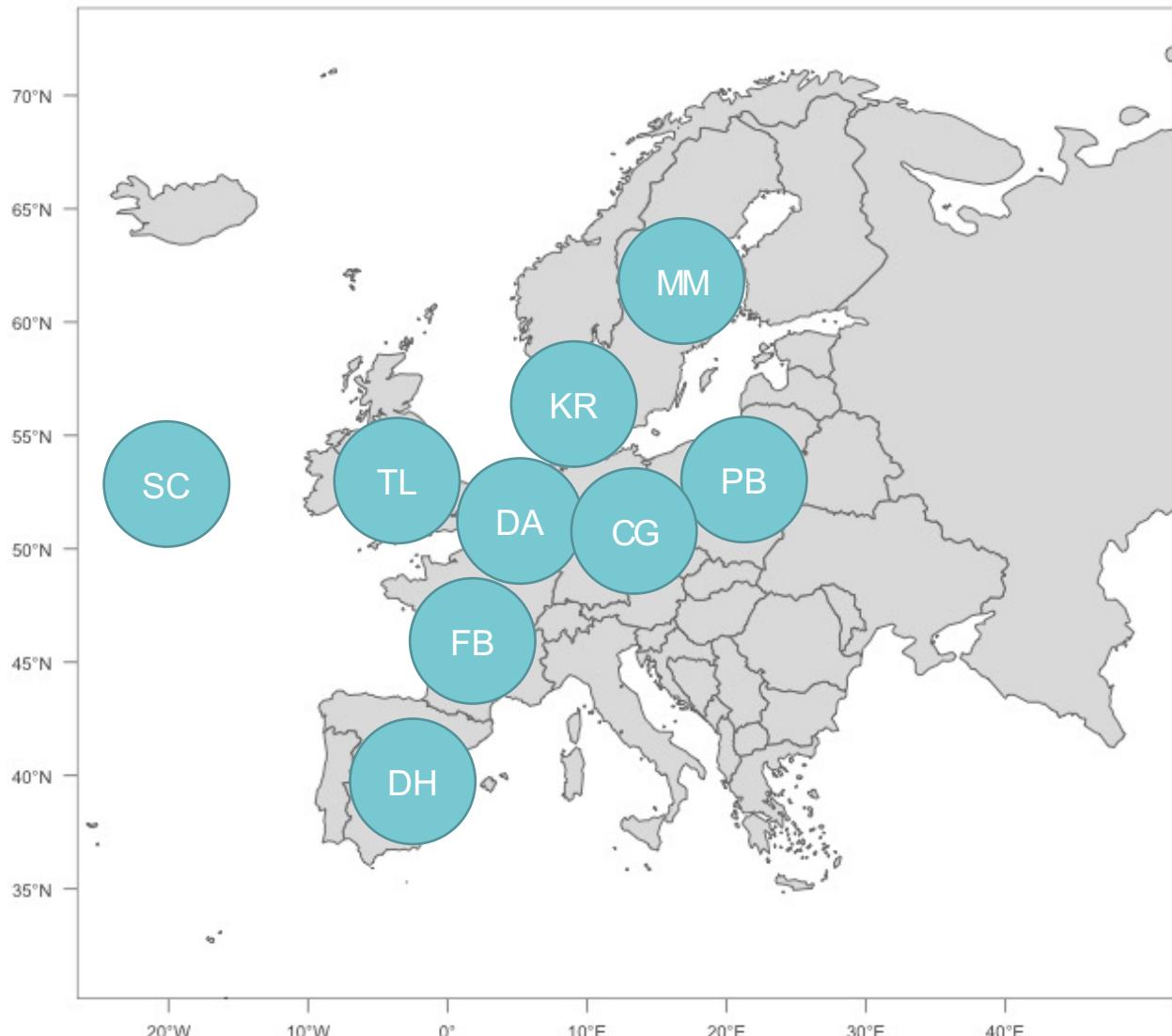
- federalregister
- mpg
- Rpublica
- pollstR

Diego Hernandez (ES)

- giscoR

François Briatte (FR)

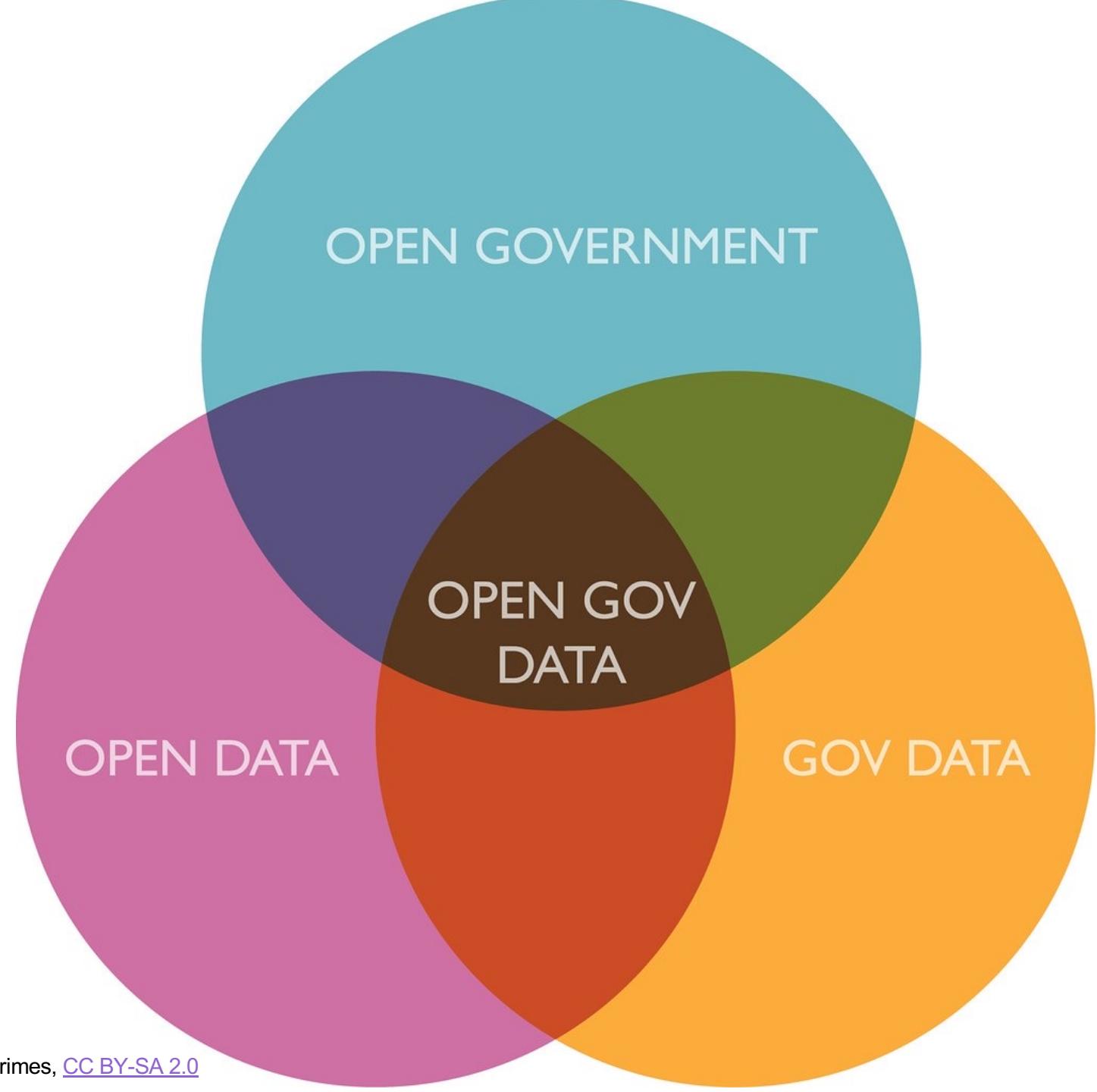
- psdata



© EuroGeographics for the administrative boundaries

# Definitions

- # Open data
- # Government Data
- # Open Government Data
- # Examples of Open Government Data



# **Open + Government + Data**

## **Open data**

**///** Data that can be freely used, re-used and distributed by anyone, only subject to (at the most) the requirement that users attribute the data and that they make their work available to be shared as well.”

## **Government Data**

**///** Any data and information produced or commissioned by public bodies.”

Source: Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", OECD Working Papers on Public Governance, No. 22, OECD Publishing, Paris, <https://doi.org/10.1787/5k46bj4f03s7-en>

# Open Government Data

- Any data and information produced or commissioned by public bodies that can be freely used, re-used and distributed by anyone.

# Examples of Open Government Data from OECD<sup>1</sup>

## ## business information

- Chamber of Commerce information, official business

## ## registers, patent and trademark information, public tender databases

## ## geographic information

- address information, aerial photos, buildings, cadastral or land registry data, geodetic networks, hydrographical data, topographic information

## ## legal information

## ## meteorological information

- climate data, climate models, weather forecasts

## ## social data

- various types of statistics on economics, employment, health, population, public administration

## ## transport information

- traffic congestion, work on roads, public transport, vehicle registration

[1] Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", OECD Working Papers on Public Governance, No. 22, OECD Publishing, Paris, <https://doi.org/10.1787/5k46bj4f03s7-en>

# rOpenGov packages: Things already covered

- Geographic information
  - geofi
  - giscoR
  - regions
  - helsinki
- Meteorological information
  - fmi2
- Social data
  - eurostat
- pxweb
- sotkanet
- hetu & sveidnumbr
- openthl<sup>2</sup> (with THL collaborators)
- *vipunen*<sup>1</sup>
- Transport information
  - *turku-package*<sup>2</sup>
  - *digitransit*<sup>1</sup>

[1] currently unmaintained

[2] unreleased, work-in-progress

# rOpenGov packages: Things yet to be done, in Open Government Data

- ✖ business information
- ✖ registers, patent and trademark information, public tender databases
- ✖ legal information

# Use case: Tilastot kartalle<sup>1</sup>



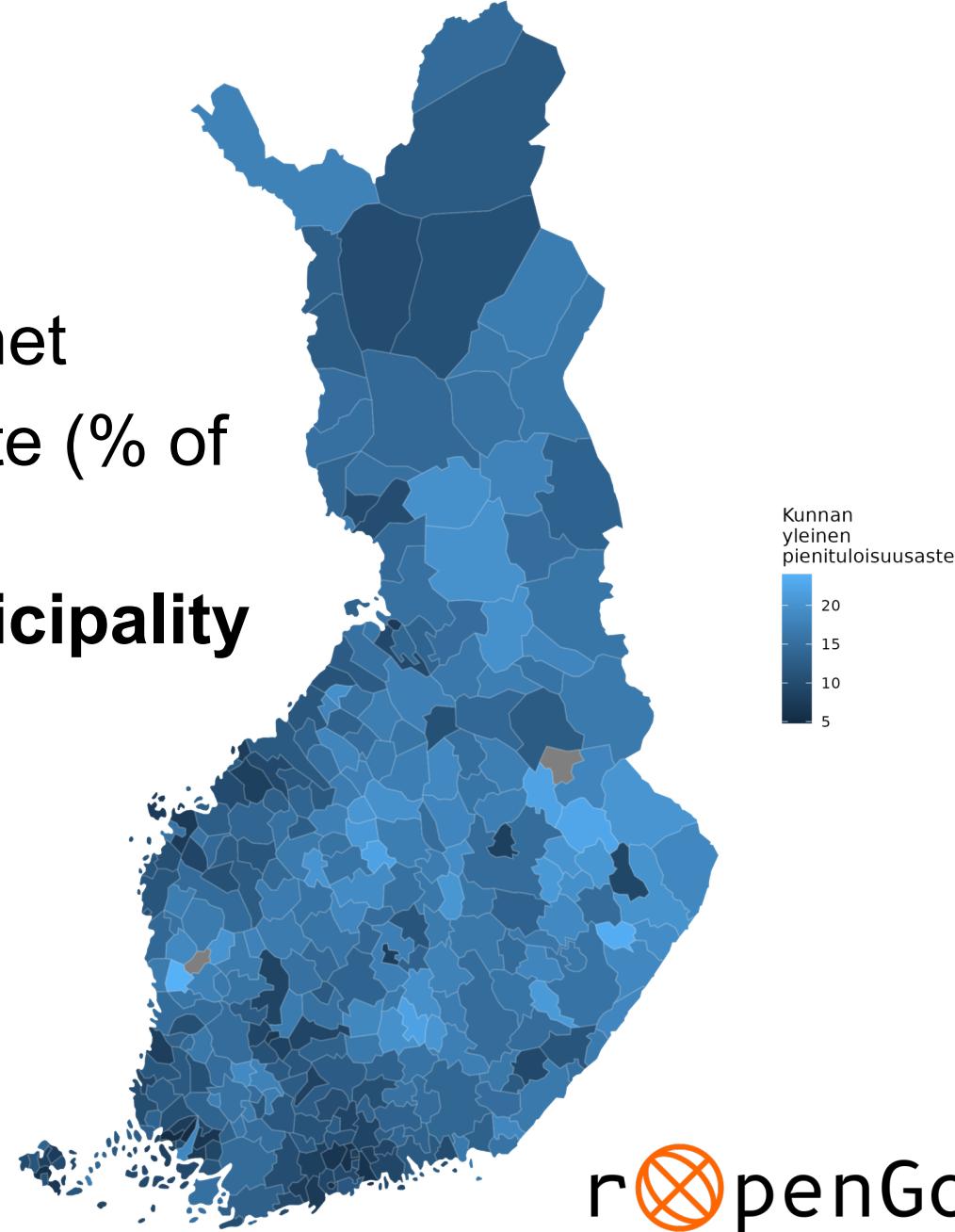
- Entry to Statistic Finland's Datamenestyjät-competition<sup>2</sup> 2021
  - Easily produce choropleth maps with **leaflet** by combining **geofi**, **pxweb** and **sotkanet** packages with **Shiny** interface
  - Output working R code that the user can further edit and refine

[1] "statistics on the map"

[2] "data achievers" -competition

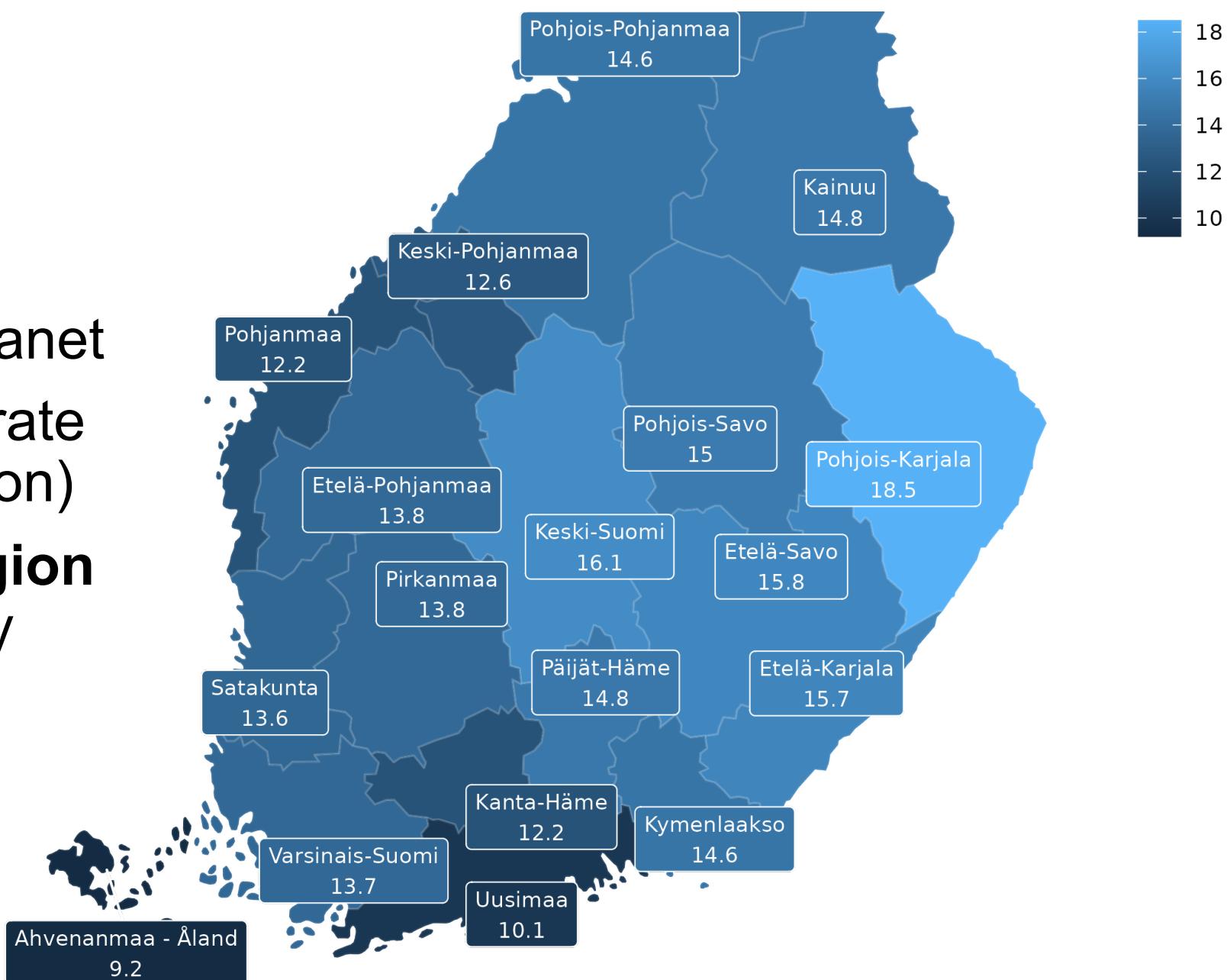
# Example 1/3

- \* Data source: Sotkanet
- \* At-risk-of-poverty rate (% of total population)
- \* Boundaries by **municipality**
- \* Year: 2018



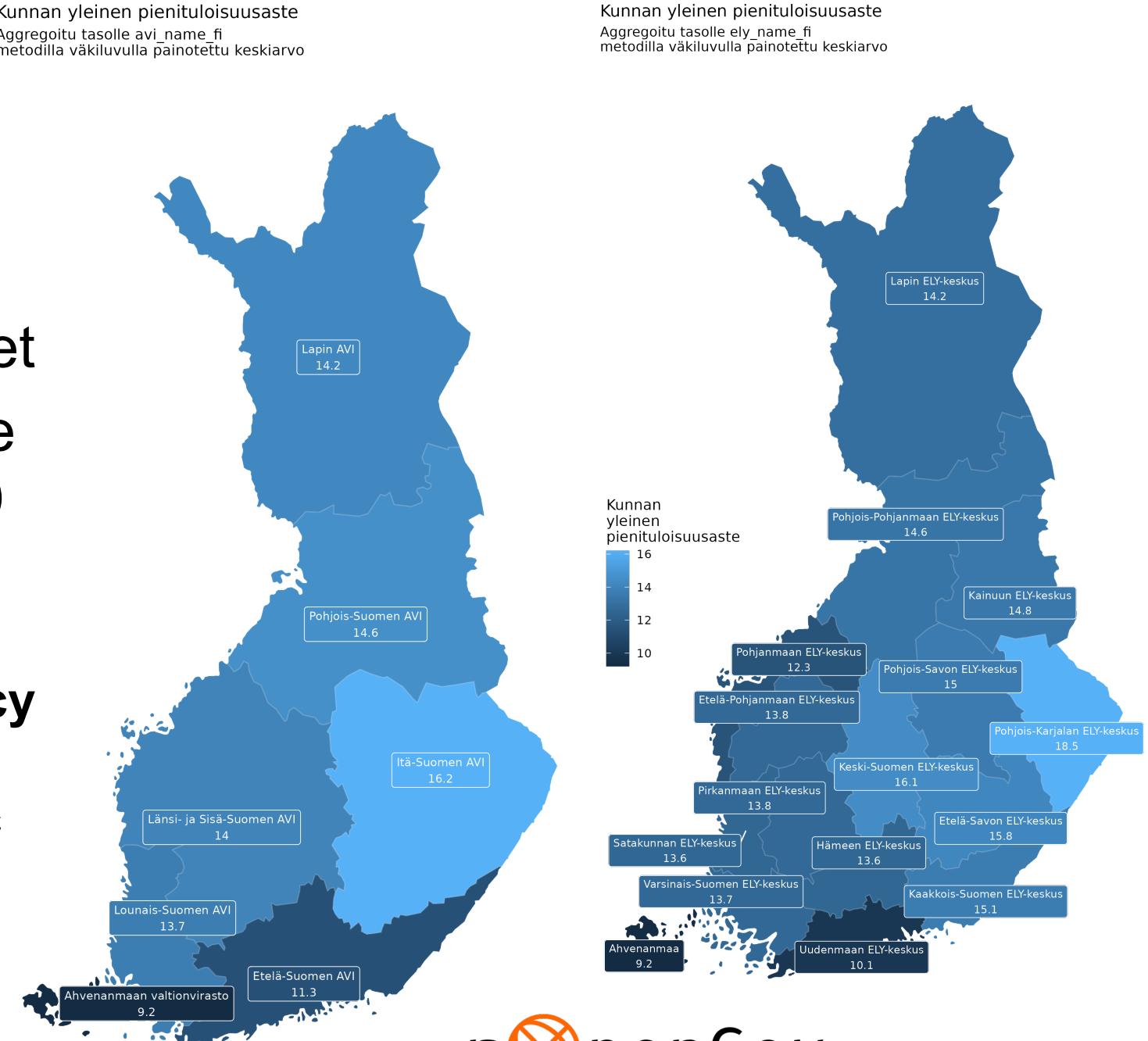
# Example 2/3

- \* Data source: Sotkanet
- \* At-risk-of-poverty rate (% of total population)
- \* Boundaries by **region** (maakunta) - a very common level of aggregation!
- \* Year: 2018



# Example 3/3

- \* Data source: Sotkanet
- \* At-risk-of-poverty rate (% of total population)
- \* Boundaries by
  - \* regional state administrative agency areas (avi-alue) and
  - \* centre for economic development areas (ely-alue)
- \* Year: 2018



A large, round pizza is shown in a black metal pan. The pizza is topped with melted cheese, sliced mushrooms, diced red onions, and green bell pepper strips. The crust is golden brown and slightly charred at the edges. The toppings are evenly distributed across the surface.

# Open Government Data and Digital Humanities

# rOpenGov packages: Things yet to be done

- ✖ business information
- ✖ registers, patent and trademark information, public tender databases
- ✖ legal information

➤ What about categories not covered by OECD?

# rOpenGov packages: Things yet to be done, in Digital Humanities 1/3

- For social scientists statistical information produced by different (government) institutions for describing the structures and quantifiable qualities of people living in modern societies may be sufficient
- For DH scientists studying languages, literature, history, ethics etc. abovementioned things are not sufficient for studying “how people document and process the human experience”
  - May be helpful as explanatory variables

# rOpenGov packages: Things yet to be done, in Digital Humanities 2/3

- Creating an easily accessible dataset of historical administrative divisions in Finland, for example:
  - municipalities before municipal mergers
  - hundreds (kihlakunta)
  - fiefs (lääni)
  - dioceses (hiippakunta)
  - sockens (pitääjä)
- Making historical demographic data (e.g. births, marriages, deaths from parish records (HisKi)) more easily accessible

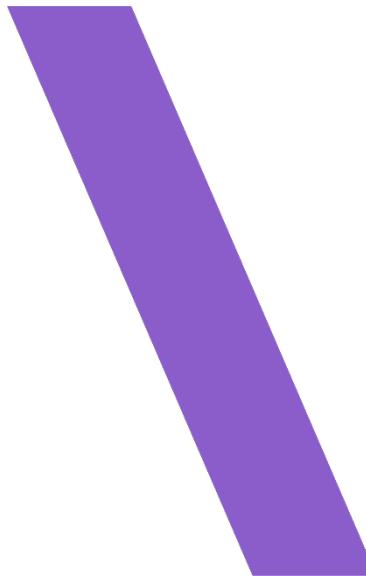
# rOpenGov packages: Things yet to be done, in Digital Humanities 3/3

- Building data pipelines and reusable infrastructure on top of CSC services
  - Current example: bibliographic data harmonization
  - Currently relying heavily on GitHub tools and GitHub Actions
- Solving the problem of long-term storage of statistical and geospatial data

# What's next for rOpenGov, as a collaborative project

- We are always looking to **grow our community of R developers** interested in open (government) data
  - Special focus in fostering the use of R and in part building a sense of community between R users in Finland (not excluding international connections!)
- We want to encourage researchers and data scientists to **open up their R code** to public and use **permissive software licenses**
- We are also open to new **joint research projects**

# Thank you!



**Check out our websites:**



<http://ropengov.org>

[https://ropengov.shinyapps.io/  
tilastot\\_kartalle/](https://ropengov.shinyapps.io/tilastot_kartalle/)



**Get engaged in GitHub:**

<https://github.com/rOpenGov>



**Follow us in Twitter:**

@ropengov

@antagomir (Leo Lahti)



UNIVERSITY  
OF TURKU

rOpenGov