RENEWING THE EUSTAT TOURISM SURVEY: NEW COLLECTION METHODS AND DESIGN FOR MORE DETAILED ESTIMATES

Elena Goni
Jorge Aramendi
Marta Salvador
Anjeles Iztueta
Fernando Tusell

European Establishment Statistics Workshop

Nuremberg - September 2013

Eustat
EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADÍSTICA

EUSKO JAURLARITZA
GOBIERNO VASCO

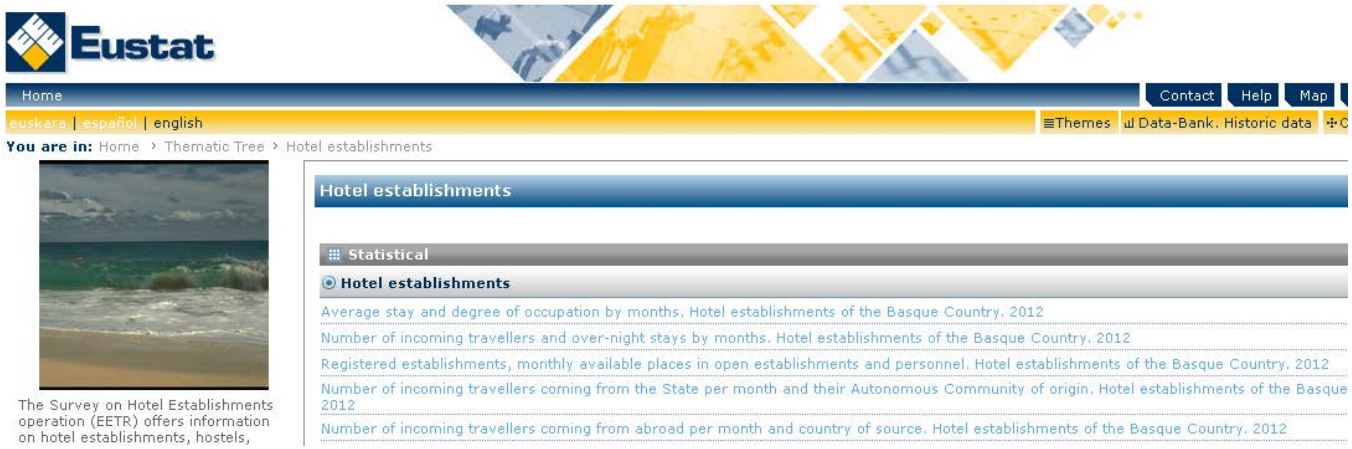
eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

Overview

- Introduction
- The data
- Previous method
- New method: an imputation approach
- Methodology
- XML
- Results and Conclusions

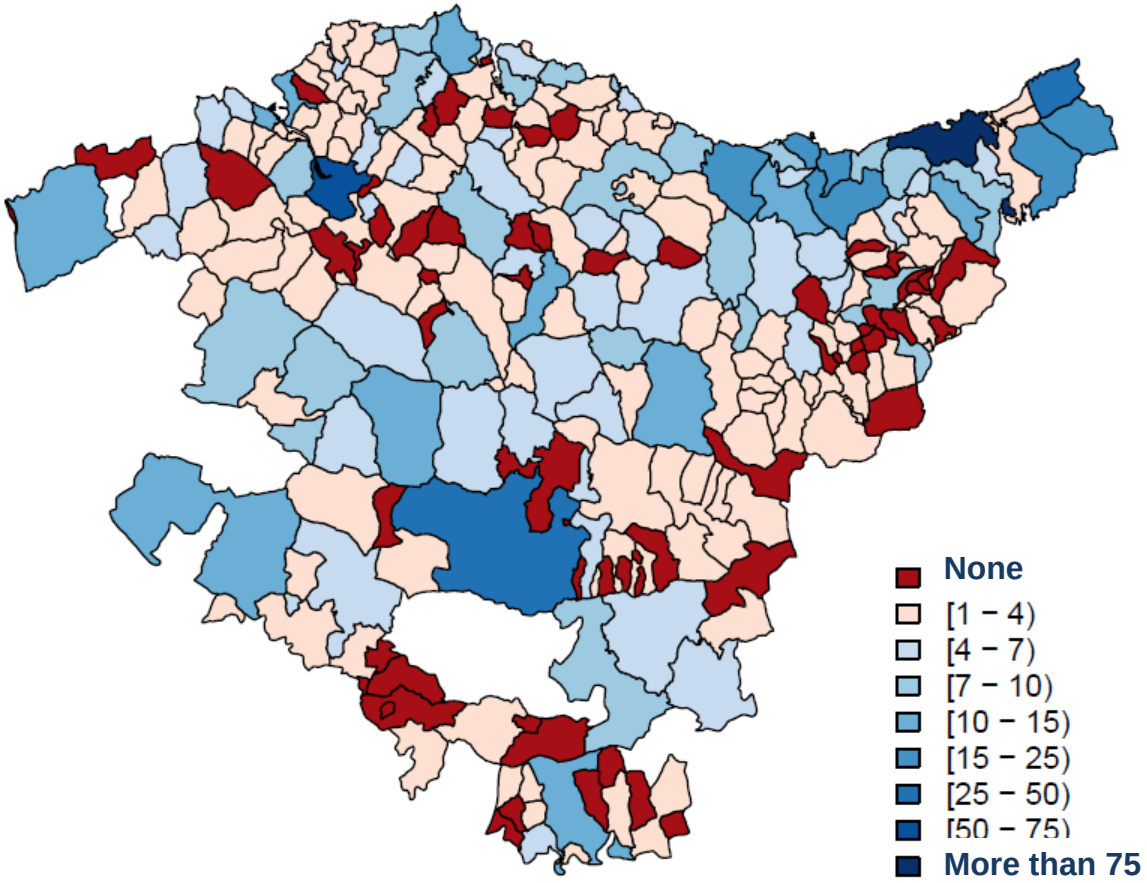
Introduction

- **The Survey of Tourism (EETR) carried out by EUSTAT is one of the pillars of the tourism information system in the Basque Country**
- **Our population: around 1000 establishments, from large hotels to rural houses or inns**
- **Variables investigated: night-stays, arrivals, occupancy rates (in terms of beds and rooms) and length of stay, all broken down by strata and geographical origin of the visitors**



Introduction

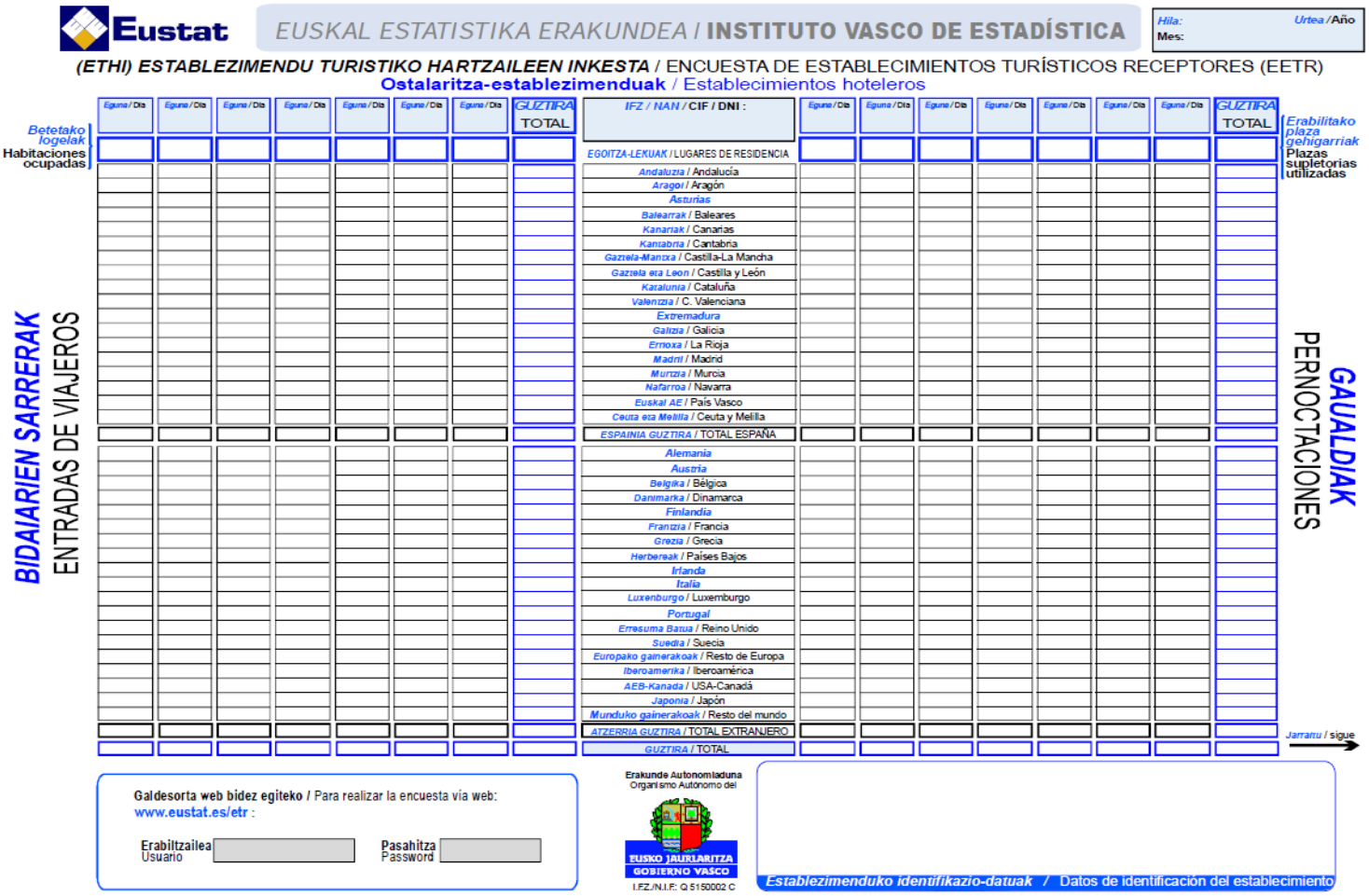
Number of hotels per municipality



The data

- In EETR, 3 or more star hotels (120 hotels which represent 60% of the offered beds) provide information for every day of the month
- The rest of them, 2 or less star ones and rural houses, are required to provide information for a period of one week
- Random choice of week: each month hotels are randomly assigned to one week
- Twenty-two strata combination of geographical area, category and type of establishment
- Sample size has been growing steadily, but still short in some strata

The data



Eustat

EUSKAL ESTATISTIKA ERAKUNDEA / INSTITUTO VASCO DE ESTADÍSTICA

Hila: / Mes: Urtea / Año

(ETHI) ESTABLEZIMENDU TURISTIKO HARTZAILEEN INKESTA / ENCUESTA DE ESTABLECIMIENTOS TURÍSTICOS RECEPTORES (EETR)

Ostalaritza-establezimenduak / Establecimientos hoteleros

Betetako logelak / Habitaciones ocupadas

Eguna / Día

GUZTIRA TOTAL

IFZ / NAN / CIF / DNI :

EGOITZA-LEKUAK / LUGARES DE RESIDENCIA

BIDAIARIEN SARRERAK / ENTRADAS DE VIAJEROS

Andalusia / Andalucía
Aragoi / Aragón
Asturias
Balearrak / Baleares
Kanariak / Canarias
Kantabria / Cantabria
Gaztela-Mantxa / Castilla-La Mancha
Gaztela eta Leon / Castilla y León
Katalunia / Cataluña
Valentzia / C. Valenciana
Extremadura
Galizia / Galicia
Errioxa / La Rioja
Madril / Madrid
Murtzia / Murcia
Nafarroa / Navarra
Euskal AE / País Vasco
Ceuta eta Melilla / Ceuta y Melilla
ESPAINIA GUZTIRA / TOTAL ESPAÑA

Alemania
Austria
Belgika / Bélgica
Danimarka / Dinamarca
Finlandia
Frantzia / Francia
Grezia / Grecia
Herbehereak / Países Bajos
Irlanda
Italia
Luxenburgo / Luxemburgo
Portugal
Erresuma Batua / Reino Unido
Suedia / Suecia
Europako gainerakoak / Resto de Europa
Iberoamerika / Iberoamérica
AEB-Kanada / USA-Canadá
Japonia / Japón
Munduko gainerakoak / Resto del mundo
ATZERRIA GUZTIRA / TOTAL EXTRANJERO
GUZTIRA / TOTAL

Erabilitako plaza gehigarriak / Plazas supletorias utilizadas

GAUALDIAK / PERNOCTACIONES

Jarraitu / sigue

Galdesorta web bidez egiteko / Para realizar la encuesta vía web:
www.eustat.es/etr :

Erabiltzailea / Usuario

Pasahitza / Password

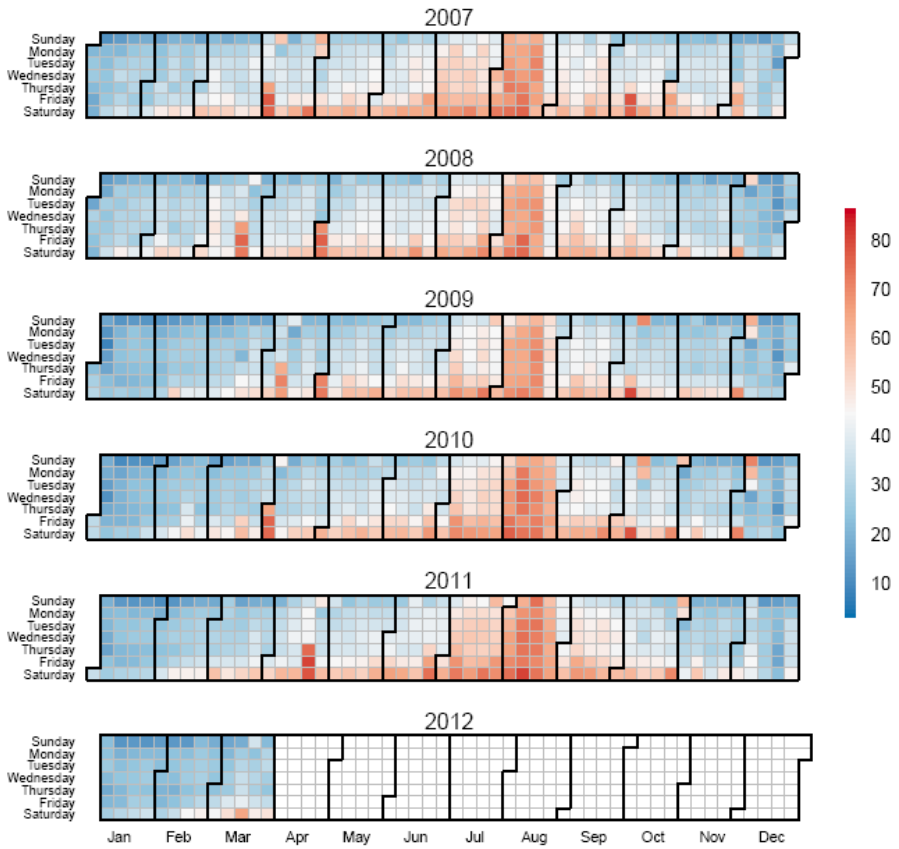
Erakunde Autonomiduna
Organismo Autónomo del
EUSKO JAURLARITZA
GOBIERNO VASCO
I.FZ./N.I.F: Q 5150002 C

Establezimenduko identifikazio-datuak / Datos de identificación del establecimiento

Universidad del País Vasco Euskal Herriko Unibertsitatea

The data

Occupancy rate in beds for the Basque Country. Years 2007-2012



Previous method

- **Sampled data was raised to strata / population totals using weights**

$$\hat{Y}_P = \frac{\sum_{(i,t)\in P} B_{it}}{\sum_{(i,t)\in O} B_{it}} \times \sum_{(i,t)\in O} y_{it}$$

- **Weights were computed using offered beds**

  ⇒ **We assumed the observed occupancy rate could be extended to non-observed beds**

- **Problem: non-standard aggregations (increasing in number) processed on demand:**

  ⇒ **Taylor-made work and extra resource consuming**
  ⇒ **Many could not be satisfied**

Previous method

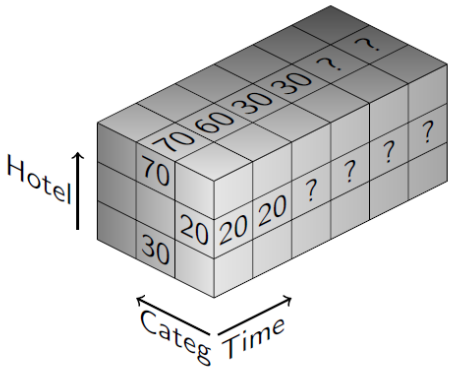
Possible solutions:

- If values were fully observed (every day and every hotel), the computation of any aggregate would only require aggregation so that every demanded aggregation could be satisfied

- If hotels could send us directly their daily booking figures without filling in our questionnaires, the response burden would be reduced and extra work needed would be reduced as well

New method: an imputation approach

- **We have create a macrotable for every variable investigated that contains the data for each hotel for each day of the month**



- **Some values are observed and some are imputed, so that the estimation of the aggregates are now calculated by simple aggregation:**

$$\tilde{\tilde{Y}}_P = \sum_{(i,t)\in O} y_{it} + \sum_{(i,t)\in P} \hat{y}_{it}$$

New method: an imputation approach

$$\tilde{Y}_P = \sum_{(i,t)\in O} y_{it} + \sum_{(i,t)\in P} \hat{y}_{it}$$

- **The first summation extends to observed values and the second summation extends to the imputed values which have not been observed and are approximated**
- **The imputation method is donor-based**
- **Constrain: coherency among imputed variable values**

  e.g.: $Night\ stays_{t} - Departures_{t+1} + Arrivals_{t+1} = Night\ stays_{t+1}$

- **Donor should be a similar hotel, so we need a definition of proximity or "similarity"**

Methodology

- **Each respondent assumes the form of a time series, so a "similarity" notion must be defined among time series**
- **One approach is a distance between the time series of each hotel i and j, for instance, the Euclidean distance:**

$$d^2(i,j) = \sum_t \left| y_{it} - y_{jt} \right|^2$$

- **What we have done is fit trajectories to each series and compute distances between the fitted trajectories**

Methodology

A model for occupancy series of hotel *i* is:

$$y_{ij} = \beta_{i,Trend(t)} + \beta_{i,DayOfYear(t)} + \beta_{i,DayOfWeek(t)} + \beta_{i,Easter(t)} + \varepsilon_{it}$$

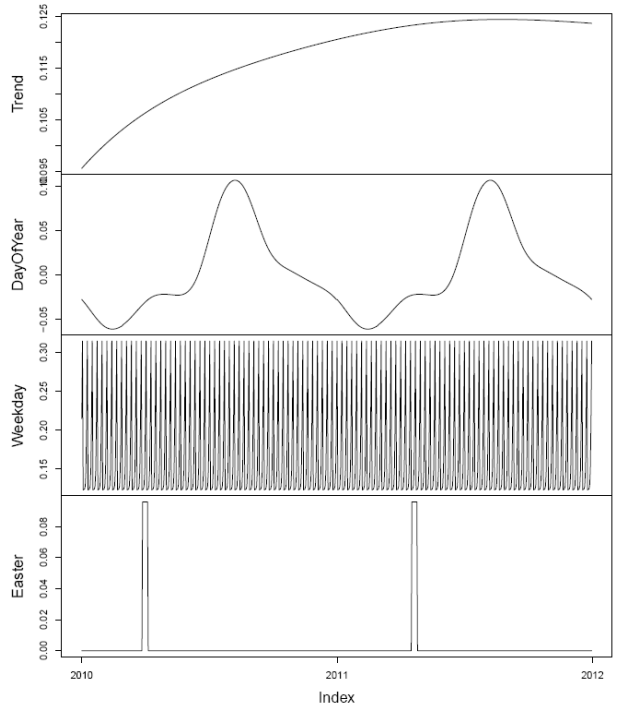
$\beta_{i,Trend(t)}$ **is a smooth function of time to capture variation over the years.**

$\beta_{i,DayOfYear(t)}$ **is a function of the day within the year associated with *t***

$\beta_{i,DayOfWeek(t)}$ **is a term capturing the effect of the day within the week corresponding to *t***

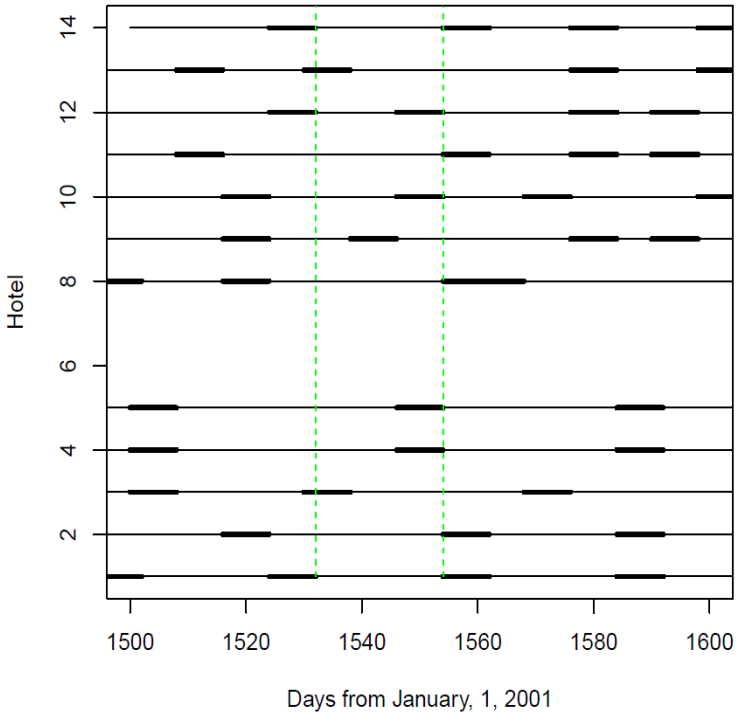
$\beta_{i,Easter(t)}$ **is a dummy variable which picks up the effect of the Easter holiday**

ε_{it} **is a random term**



Methodology

- In many cases we need more than one single donor due to the sampling scheme used (7 days per month)
- The closest candidate may be able to provide values for some days of the month, not for all needed
- We will chose new donors sequentially in order of increasing distance



Methodology

- **New hotels: the model cannot be fitted correctly if a hotel have no previous data**

- **We define an alternative "similarity" based on geographical distance:**

  - ⇒ **Hotels in the same municipality → distance "0"**
  - ⇒ **Hotels in bordering municipalities → distance "1"**

    ...and so on

- **This distance is further corrected by increasing 0,5 for pairs of hotels with different categories**

Methodology

DONOR SELECTION → EUCLIDEAN DISTANCE

1) computed from a model when data permits the fitting of model

2) geographical distance

- **Notice!! the donor and the recipient hotel will have similar occupancy profiles, HOWEVER, this may happen with hotels of very different size, so directly imputing arrivals, night-stays, etc. is not possible**

- **We adjust for the relative sizes using the ratio of their respective numbers of offered beds (B). For example, arrivals (A) for hotel i at time t:**

$$\hat{A}_{i,t} = \frac{B_{i,t}}{B_{DONOR,t}} \times A_{DONOR,t}$$

- **This ensures the imputed values stay within the allowable range**

XML files

XML files → ELECTRONIC DATA COLLECTION

- EUSTAT implemented de XML files in order to reduce the response burden
- 59 hotels / rural houses use this method to supply us with monthly data
- It consists on a software implemented in the user´s PC that generates an XLM file with the information demanded by our EETR
- This information arrives in a totally compatible format with the application used in Eustat to process the data
- Problem: nowadays the software is not free of charge

Results and Conclusions

- **A simple, donor-based method has been devised and implemented to ease the task of producing customized estimates of the EETR**
- **The approach is simple and appears to work well!!!**
- **Simple method: model fitting only once in a while**
- **The method processes data daily, thus it is immune to intra-month sample fluctuations**

Danke Schön!!!

www.eustat.es

Eustat
EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADÍSTICA

EUSKO JAURLARITZA
GOBIERNO VASCO

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea