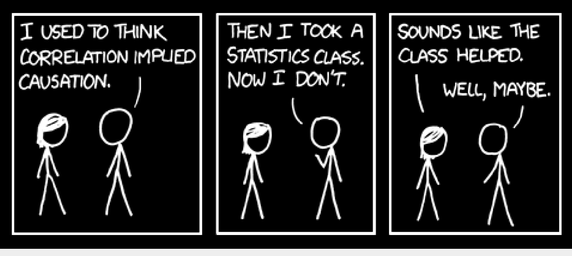
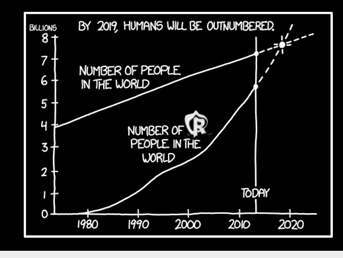


R USERS ALL AROUND THE WORLD

GERGELY DAROCZI

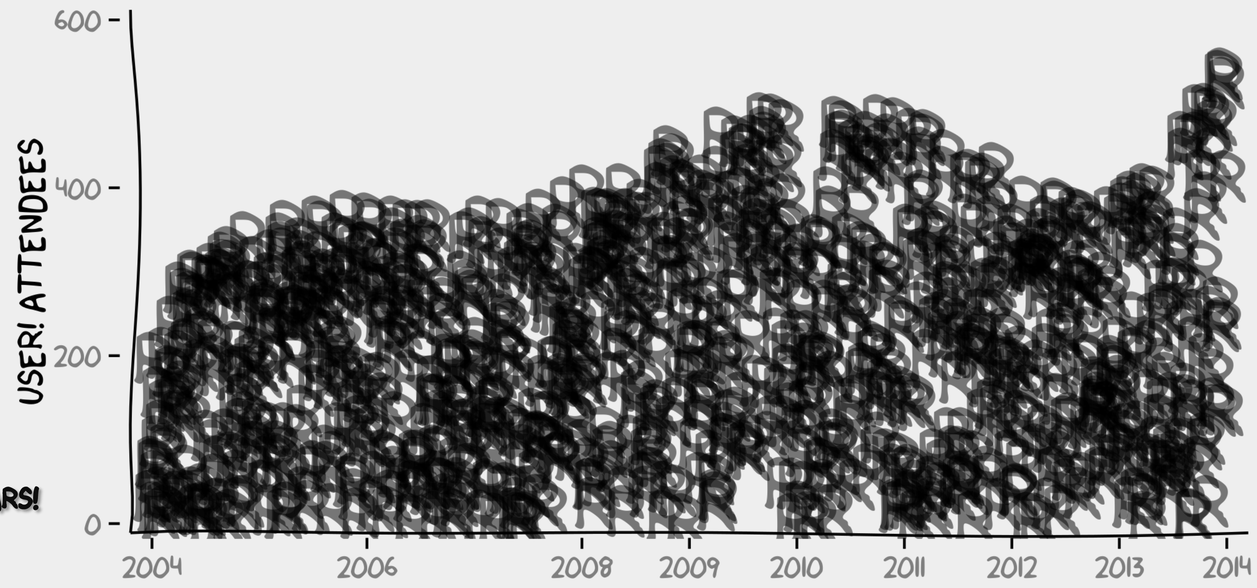
RAPPORTER.NET



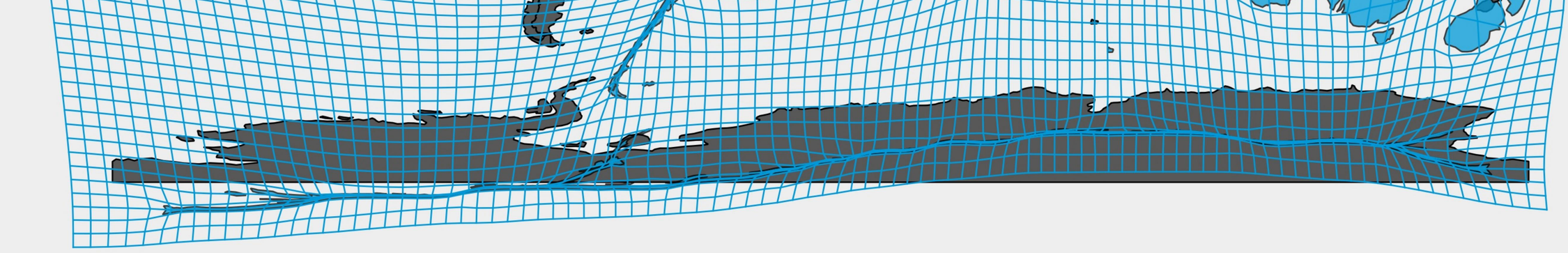
OVERVIEW

This poster was inspired by a recent blog post on „Where is the R activity?" and a few follow-up articles, where the authors tried to determine the **number of R users all around the world**.
The first version of my related experiment included various data sources on the number of **R Foundation members**, attendee-lists of the **useR! conferences**, the number and size of **R User Groups** registered on meetup.com, **R package downloads** from Rstudio cloud-CRAN servers, R-related **search queries via Google** and the geographical distribution of **GitHub users** with at least one R repository. The results and a combined R score were published on an interactive D3.js application.
Now we focus on the attendees of the useR! conferences along with the fresh dataset of **600+ registrants** here (aka *useR! 2014*). First, we visualized the number of paricipants in time and space, then identified the number of attendance on a useR basis, so that we could **estimate** the size of the unkown population with **loglinear models**. How many guys and girls are using R? 2 or 3M? Even more?
Although this poster was not be able to answer this question after all, it still features some related statistical methods along with a few **XKCD** webcomic-styled plots of the results. Further analysis was also done on the proportion of genders, professional titles and headlines of the attendees etc.

THE NUMBER OF USER! CONFERENCE ATTENDEES FROM YEAR TO YEAR

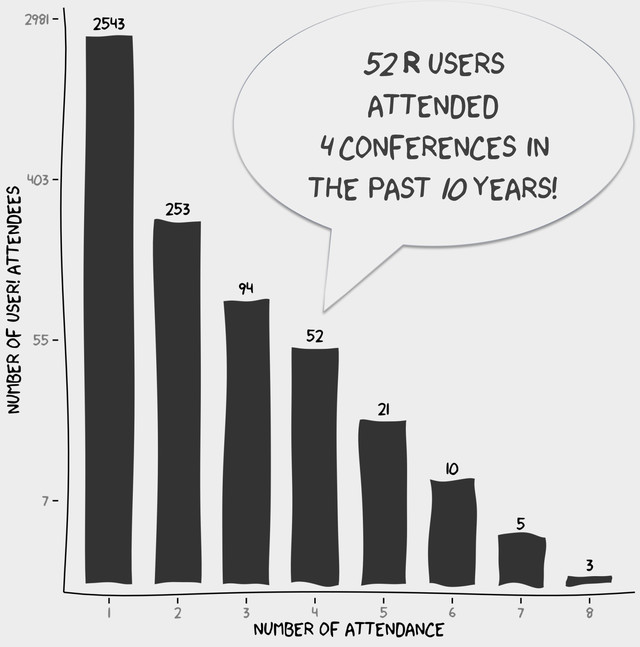


OVERALL NUMBER OF ATTENDEES FOR ALL USER! CONFERENCES BETWEEN 2004 AND 2014 – 10 YEARS!



	UNITED STATES	ARGENTINA	AUSTRALIA	AUSTRIA	BANGLADESH	BELGIUM	BRAZIL	BULGARIA	CANADA	CHILE	CHINA	CROATIA	CYPRUS	CZECH REPUBLIC	DENMARK	ECUADOR	EGYPT	ESTONIA	FINLAND	FRANCE	GERMANY	HAITI	HONG KONG	HUNGARY	ICELAND	INDIA	INDONESIA	IRAN	IRELAND	ISRAEL	ITALY	JAPAN	KENYA	KOREA	KOREA SOUTH	LATVIA	LUXEMBOURG	MEXICO	NETHERLANDS	NEW ZEALAND	NIGERIA	NORWAY	PERU	PHILIPPINES	POLAND	PORTUGAL	ROMANIA	RUSSIAN FEDERATION	SAUDI ARABIA	SERBIA	SINGAPORE	SLOVAKIA	SLOVENIA	SOUTH AFRICA	SPAIN	SWEDEN	SWITZERLAND	TAIWAN	THAILAND	TUNISIA	TURKEY	UKRAINE	UNITED KINGDOM
VIENNA	28	0	2	35	0	7	0	0	2	0	0	0	0	3	4	0	0	0	3	8	37	0	0	0	1	0	0	0	0	0	7	2	0	0	0	0	0	0	3	5	0	3	0	0	3	0	0	0	0	0	0	0	2	0	1	0	0	0	0	17
VIENNA AGAIN	57	0	3	50	0	7	1	0	9	1	0	0	0	11	6	0	1	0	6	9	70	0	1	2	0	0	0	0	0	0	14	7	1	0	0	0	0	0	2	1	0	1	2	0	3	9	1	0	0	0	0	0	2	1	3	4	1	0	0	2	0	25
AMES																																TODO!!!																														
DORTMUND	41	1	4	15	0	8	1	0	8	0	1	0	1	4	4	0	0	0	3	15	178	0	0	3	0	1	0	1	1	0	9	6	0	0	2	0	2	1	12	1	0	2	0	0	3	4	0	0	0	0	0	0	0	0	5	1	16	5	1	0	0	1	26
RENNES	58	0	4	19	0	17	2	0	6	0	0	1	0	2	6	0	0	0	0	165	55	0	0	2	0	0	0	0	3	1	13	2	0	0	0	0	2	3	17	3	0	0	0	0	4	4	0	0	0	0	0	1	2	2	11	3	13	6	2	0	1	0	33
GAITHERSBURG	362	0	2	7	0	5	2	0	6	0	1	0	0	0	0	0	0	0	1	9	15	0	0	0	0	0	0	1	0	2	1	6	2	0	0	0	0	0	2	4	1	1	0	0	0	3	2	0	0	0	0	0	1	1	6	2	3	1	1	0	0	14
COVENTRY	56	0	8	12	0	14	2	0	5	0	1	0	0	3	9	0	0	0	0	16	49	0	0	0	1	0	0	1	8	2	8	2	1	0	3	1	0	3	6	6	0	4	0	0	2	2	1	0	0	0	0	0	5	3	14	6	9	2	0	0	1	0	163
NASHVILLE	397	0	2	1	0	2	0	0	8	0	0	0	0	0	3	0	0	0	0	1	14	0	0	1	0	0	0	0	0	1	2	0	0	0	2	0	0	2	3	0	0	1	0	0	1	0	0	0	0	0	0	0	1	3	2	2	3	0	0	0	0	17
ALBACETE	58	0	6	14	1	13	0	0	3	0	0	0	0	2	8	0	0	1	3	11	24	0	0	1	0	0	0	0	1	1	12	4	0	0	6	1	0	1	11	2	0	6	0	0	11	5	0	3	0	1	1	0	3	1	66	2	6	1	0	0	3	31
HERE	490		2	2		4	3	1	12	0	0	0	0	0	5	1	0	0	2	8	10	1	0	3	0	0	0	0	0	1	2	10	0	1	2	0	0	3	3	6	0	2	0	1	1	0	0	0	0	1	2	0	2	0	6	3	5	0	0	0	0	7
THE EARTH	1547	1	33	155	1	77	11	1	59	1	3	1	1	25	45	1	1	1	18	242	452	1	1	13	2	1	1	4	18	7	73	35	2	1	15	2	5	16	61	25	1	19	2	3	29	29	1	3	2	2	4	1	17	12	121	27	73	17	7	1	8	1	333

MOST STATISTICIANS CAN READ, ANALYSE AND INTERPRET HUGE TABLES EASIER AND QUICKER COMPARED TO REALIZING WHAT'S GOING ON IN A CARTOGRAM!



Rcapture::closedp

HYPHOTESIS, METHODS, RESULTS

1. The name of all **useR! 2004-2014** attendees was compiled in a list, where some (left figure) **recurring participants** were identified after manual data cleaning.
2. These results can be used as the "number of units captured *i* time" in a "mark and recapture" research msethod for **estimating the unknown population size**.
3. Building open-population models by estimating the **observation probability**, the **survival rate** and the **number of new arrivals** based on historical data.
4. Building 6x6 **closed population** models based on 3-years long time intervals (right figure): the *lower bound* estimate of Chao showed an increase from 2,500 to 7,000 even with and without a time effect, while Daroch's model estimates tend to spread between 15,000 and 70,000. The Gamma and Poisson models of Rivest and Baillargeon (2012) returns some restrained (10,000-20,000) and some over-exaggerated Gamma results (30,000-250,000).

REFERENCES

- James Cheshire (2013). Where is the R Activity? http://spatial.ly
- Gergely Daroczi (2013). The attendants of useR! 2013 around the world. http://blog.rapporter.net
- Gergely Daroczi (2014). R activity around the world. http://blog.rapporter.net
- Dominique Andrieu, Christian Kaiser and André Ourednik: ScapeToad. v1.1. http://scapetoad.choros.ch
- Gastner, M.T. and Newman, M. (2004). Diffusion-based method for producing density equalizing maps. PNAS 101(20): 7499-7504.
- Sophie Baillargeon, Louis-Paul Rivest (2012). Rcapture: Loglinear Models for Capture-Recapture Experiments. R pkg v1.3-1. CRAN.
- Emilio Torres Manzanera (2014). xkcd: Plotting ggplot2 graphics in a XKCD style. R pkg v0.0.3. CRAN.

LIMITATIONS

- UseR! attendees do not represent R users.
- Further data cleaning is required.
- Missing data for 2007.
- Chinese R Conference is not included.
- Closed population model is not appropriate.
- Improved model selection is desired.
- We need more & independent data sources.

CONTACT: DAROCZIG AT RAPPORTER DOT NET