3.5.5 Proof of Concept

Characterization of the OSS community and measurement of impact and diffusion at the country, sector, and institution levels require detailed information about the contributors. Here we show the feasibility of one of the approaches using a pilot study on R packages. As a proof of concept, we use the data collection strategy shown in Fig. 2 to obtain information about over 11K packages from CRAN published between 2005 and 2018. The information includes the published date, authors and their roles (creator or maintainer, contributor, copyright holder), dependencies, and URL's to the repositories. CRAN also requires that the maintainer must provide an email address. We use the email addresses and an internet country domains list (e.g., [41]) to obtain information about the location and institutions of the creators. The left table in Fig. 5 summarizes the most common top-level domains, and the corresponding number of projects and creators. The largest share of maintainers have .com email addresses; these provide little information about geography or sector. However, more than a third (36%) have country-specific email addresses, and 17% have .edu domains, giving us a lower boundary on university contributions. The right table in Fig. 5 presents the top 10 countries (out of 88) with the highest number of contributions (packages).

		# Packages		# Maintainers	
Top-Level Domains (Maintainer emails)	.com	4,964	42%	2,770	40%
	.edu	1,981	17%	1,202	17%
	.org	481	4%	184	3%
	.net	168	1%	89	1%
	.gov	69	1%	43	1%
	.name	33	0%	3	0%
	.info	8	0%	6	0%
	.biz	6	0%	3	0%
	.(country)	4,124	35%	2,495	36%
	Total	**11,886**		**6,697**	

Country (top-level domain)	Number of packages (%)		Number of Maintainers (%)	
Germany (.de)	687	(5.8%)	427	(6.2%)
United Kingdom (.uk)	434	(3.7%)	267	(3.9%)
France (.fr)	398	(3.3%)	235	(3.4%)
Canada (.ca)	335	(2.8%)	160	(2.3%)
Australia (.au)	198	(1.7%)	109	(1.6%)
Italy (.it)	198	(1.7%)	129	(1.9%)
Switzerland (.ch)	172	(1.4%)	102	(1.5%)
Spain (.es)	166	(1.4%)	102	(1.5%)
Netherlands (.nl)	151	(1.3%)	89	(1.3%)
Austria (.at)	123	(1.0%)	56	(0.8%)

Figure 5: **(left)** Top-level Domains; **(right)** Country-level Domains of R Package Maintainers

To focus on specific sectors and identify specific organizations, we further explore .edu email domains. Fig. 6 shows the top contributors, based on .edu email addresses on CRAN [22].

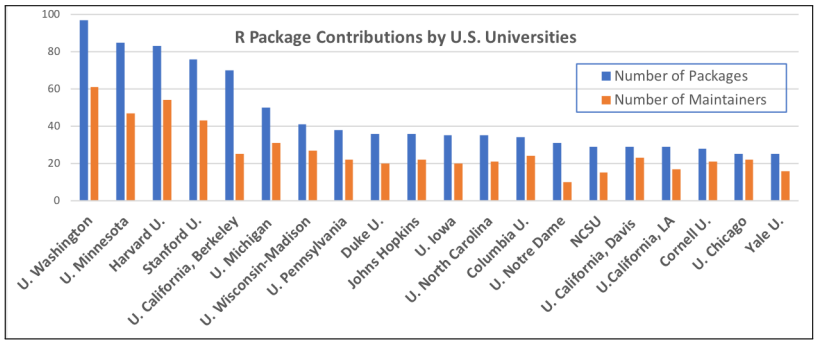

Figure 6: Top R Contributors on CRAN [22] by university. Based on 17% of the 11.8K packages that have the maintainer email domain given as .edu.

Having identified the universities associated with OSS projects, one can study the patterns of collaborations, identify the impact of these packages, and analyze the diffusion across these organizations. This pilot study provides evidence of the availability of the data to support our research and also shows some of the limitations to be addressed during the study.