**The Scope and Impact of Open Source Software as Intangible Capital:
A Framework for Measurement with an Application Based on the use of R Packages**

**Carol A. Robbins*(1), Gizem Korkmaz (2), José Bayoán Santiago Calderón (3),
Daniel Chen (2), Aaron Schroeder (2), Claire Kelling (4), Stephanie Shipp (2), Sallie Keller (2)**

CRIW Conference: Big Data for 21st Century Economic Statistics
Bethesda, MD
Washington, D.C.
March 15-16, 2019

1) National Center for Science and Engineering Statistics, National Science Foundation ; 2) Social & Decision Analytics Division, Biocomplexity Institute & Initiative, University of Virginia; 3) Claremont Graduate University; 4) Pennsylvania State University

The views expressed in this paper are those of the authors and not necessarily those of their respective institutions.

Measuring Open Source Software

Why:

Long-lasting benefits, zero purchase cost, creates new products

Public goods qualities, output of public as well as private spending

How:

1. Access detailed data about these new products and tools from repositories:

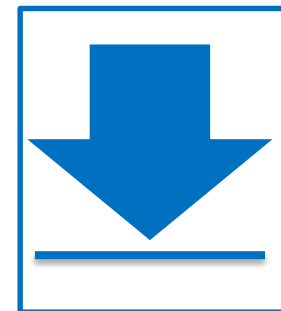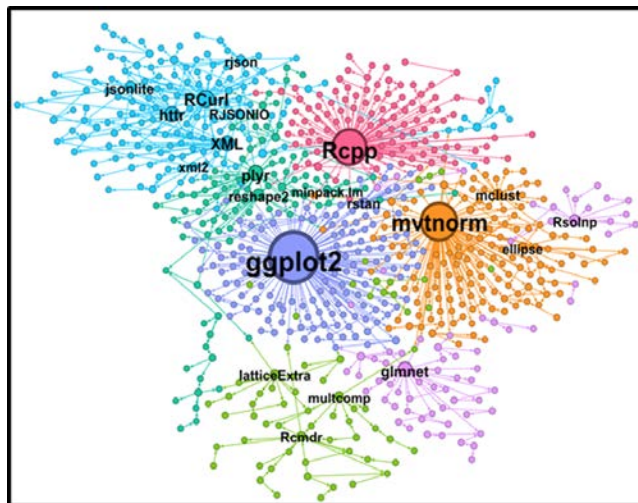2. Develop framework for measuring cost

$2017

Measuring Open Source Software

How (continued):

3. Estimate resource cost for two languages

4. Use Network Analysis and Downloads to show impact

U.S. Software Investment in 2017: $381 Billion

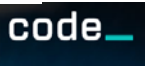| Components of Software Investment	Private Sector total in billions $352.9*	Public Sector total in billions $38.4	
Prepackaged	$147.6		
Custom	$141.1	State and Local $17.7	Federal $10.7
Own-account	64.3		

Source: BEA. Intellectual Property Products Fixed Asset Tables (private) and Investment in Government Fixed Assets (Table 7.5B).
* Difference between total and sum is a rounding error.

3

A Small Tweak to Standard Software Investment Categories can highlight OSS

Components of Software Investment	Private Sector			Public Sector			Household Sector	Rest of World
	Business	Other private nonprofits	Higher education	Higher education	Federal Government and FFRDCs	Non-federal government, ex. Higher Ed.		
Prepackaged								
Custom								
Proprietary								
Open Source (OSS)								
Own-account								
Proprietary								
Open Source (OSS)								

4

"Big Data" sources for OSS measurement

- Registry contains:
  - Basic package information
  - Package manifest file
- CRAN   The Comprehensive R Archive Network
- PyPI
- Code.gov

- Source Code Hosting Platform contains:
  - Author(s), Maintainers

  Version

  - License
- GitHub

GitHub Platform

Number of Users or Developers, in millions

Language	R	Python
Package manager	CRAN	PyPI
Number of packages	13,719	164,836
Production ready	13,350	17,482
OSI-approved & production ready	13,143	15,043
Packages on GitHub	4,407	11,016
Packages on GitHub (analysis)	4,358	9,773

6

Open Source Projects by Federal Government Organization
Top 5 by number of projects
for projects started before January 1, 2018

Organization Name	Total Projects on Code.gov	Number of Projects Linked to Github collection	Kilo-lines of code (kloc)	Commits	Number of contributors
Total	4,457	2,688	2,486,210	950,625	8,292
General Services Administration	1,501	1,368	266,860	318,676	4,631
Department of Energy	899	704	1,219,835	485,726	2,433
Consumer Financial Protection Bureau	261	243	753,447	49,781	334
National Aeronautics and Space Administration	998	141	179,917	51,936	358
Environmental Protection Agency	156	61	14,327	4,711	78

Sharing America's Code
Unlock the tremendous potential of the Federal Government's software.

Search thousands of Federal Government projects

e.g. JavaScript, NASA, web standards   Go

Or

Browse by Agency

Ready. Set. Code!
Whether you are a beginner or an experienced coder, join the open source community. Help improve America's Code.

Explore Open Tasks

7

Note: for projects started before January 1, 2018

Cost of OSS Software Package Creation

1) Kilo-lines represent effort

- Effort is a nonlinear function of complexity and lines of code
- historical software project factors

$$Effort = 2.4(KLOC)^{1.05}$$

$$Nominal\ development\ time = 2.5(Effort)^{.38}$$

$$Development\ cost = Monthly\ wage\ x\ Nominal\ development\ time$$

2) Estimate resource cost with wage equivalent for 2017

- Computer programmers, software developers
- Occupation Employment Survey, Bureau of Labor Statistics

3) Estimate non-wage costs adapting OECD and BEA methods

KLOC and Resource Cost Based on Projects in CRAN

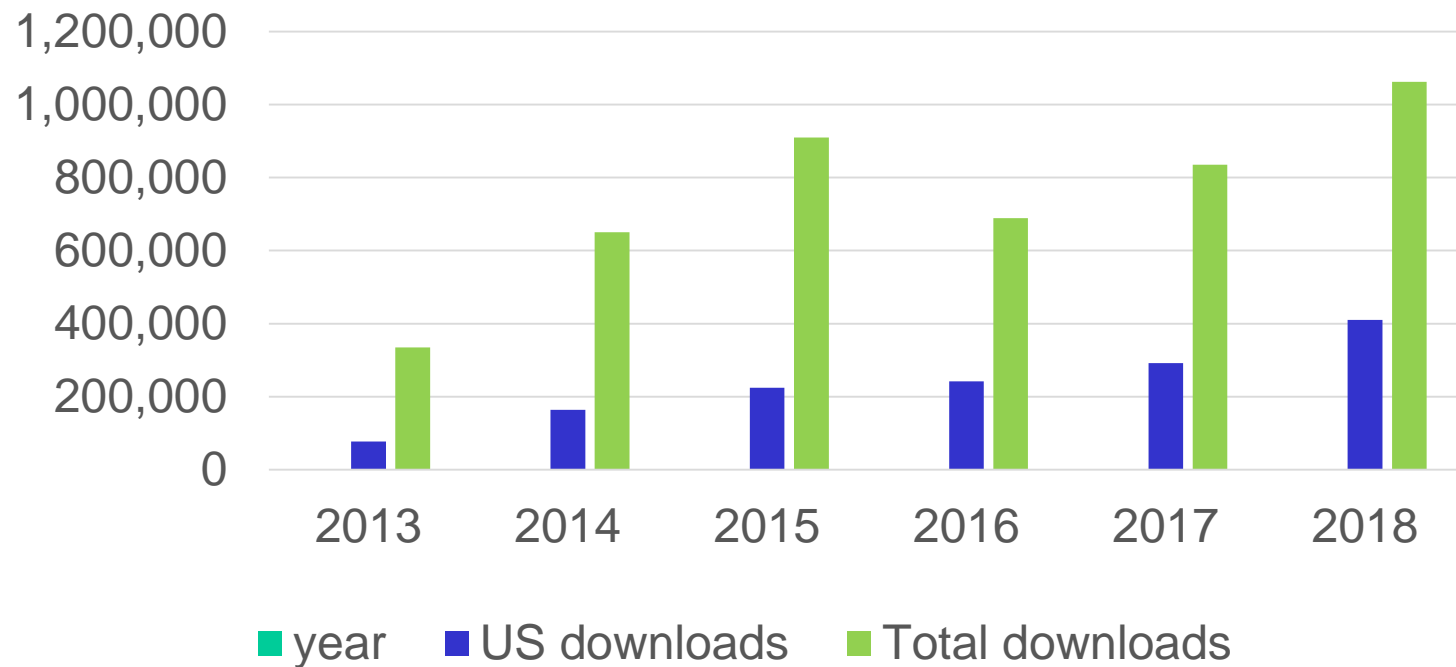| Name	Klines	Estimated Cost in Thousands of 2017$
All packages	100,216.787	1,579,689
mapdata	2,257.20	1,516
hunspell	756.9	980
edgarWebR	456.8	801
TCGA2STAT	376.9	742
igraph	364	732

KLOC and Resource Cost Based on GitHub KLOC

Package Name	KLOC	Estimated Cost in Thousands of 2017$
All packages	282,167.871	883,209
archivist	28488.639	4,169
CollessLike	15844.721	3,299
readtext	13888.309	3,130
ptwikiwords	11452.965	2,898
nasapower	10613.638	2,812

Package Name	KLOC	Estimated Cost in Thousands of 2017$
All packages	611,601.568	1,560,374
libsass	50340.53	5,233
py3-ortools	37412.424	4,648
LSD-Bubble	15270.398	3,251
IotPy	14899.252	3,219
openquake.engine	13841.578	3,126

R Downloads and Reuses as Impact

Package	# Reuse
ggplot2	105,774
plyr	101,596
digest	99,774
stringr	98,086
colorspace	93,590
RColorBrewer	81,448
reshape2	81,350
scales	73,385
proto	71,698
munsell	71,483

Package	2018 Downloads
Rcpp	3,519,510
rlang	2,893,889
stringi	2,610,184
stringr	2,511,011
ggplot2	2,495,315
digest	2,453,958
glue	2,296,688
tibble	2,242,376
pillar	2,222,364
yaml	2,207,621

Sectors of R Contributors from CRAN

Domain	Packages	Percent	Maintainers	Percent
Total	11,886.0		6,697	
.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%

Countries of R Contributors from CRAN

Domain	Packages	Percent	Maintainers	Percent
Germany (.de)	687	6%	427	6%
United Kingdom (.uk)	434	4%	267	4%
France (.fr)	398	3%	235	3%
Canada (.ca)	335	3%	160	2%
Australia (.au)	198	2%	109	2%
Italy (.it)	198	2%	129	2%
Switzerland (.ch)	172	1%	102	2%
Spain (.es)	166	1%	102	2%
Netherlands (.nl)	151	1%	89	1%
Austria (.at)	123	0.0	56.0	1%

Questions?
crobbins@nsf.gov

Grateful Acknowledgement to
Team Open Source Surfers (TOSS)

Hannah Brinkley, Daniel Chen, José Bayoán Santiago Calderón, Eirik Iversen, Keren Chen, and Gizem Korkmaz