



Open Source Software as Intangible Capital: Measuring the Cost and Impact of Free Digital Tools

**Carol A. Robbins¹, Gizem Korkmaz², José Bayoán Santiago Calderón³, Daniel Chen²,
Claire Kelling⁴, Stephanie Shipp², Sallie Keller²**

Measuring Economic Welfare in the Digital Age: What and How?
IMF Statistical Forum
Washington, D.C.
November 19–20, 2018

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

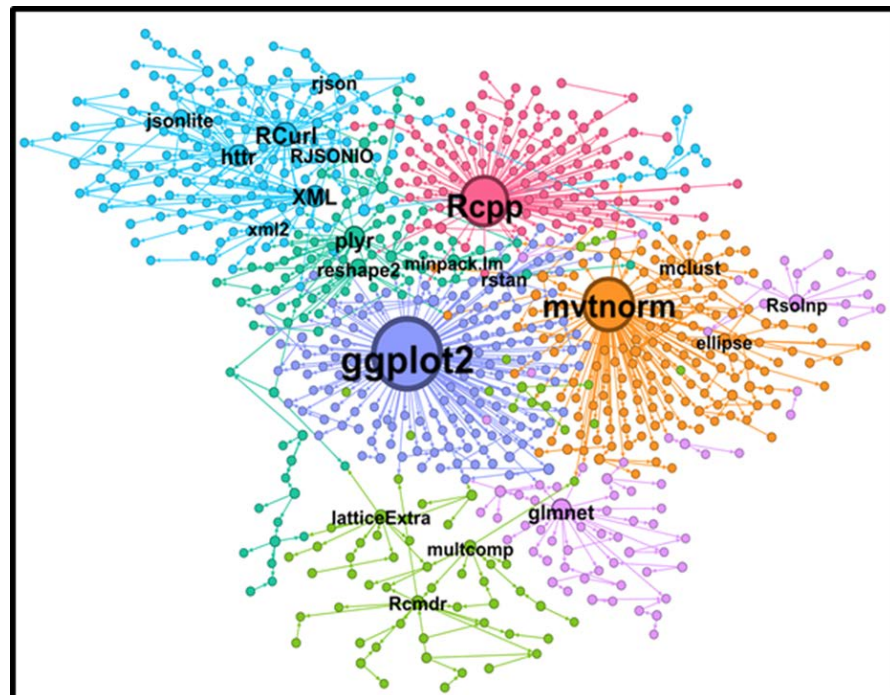
Measuring Open Source Software

How (continued):

3. Estimate resource cost for four OSS Languages



4. Use Network Analysis to show impact



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

Go

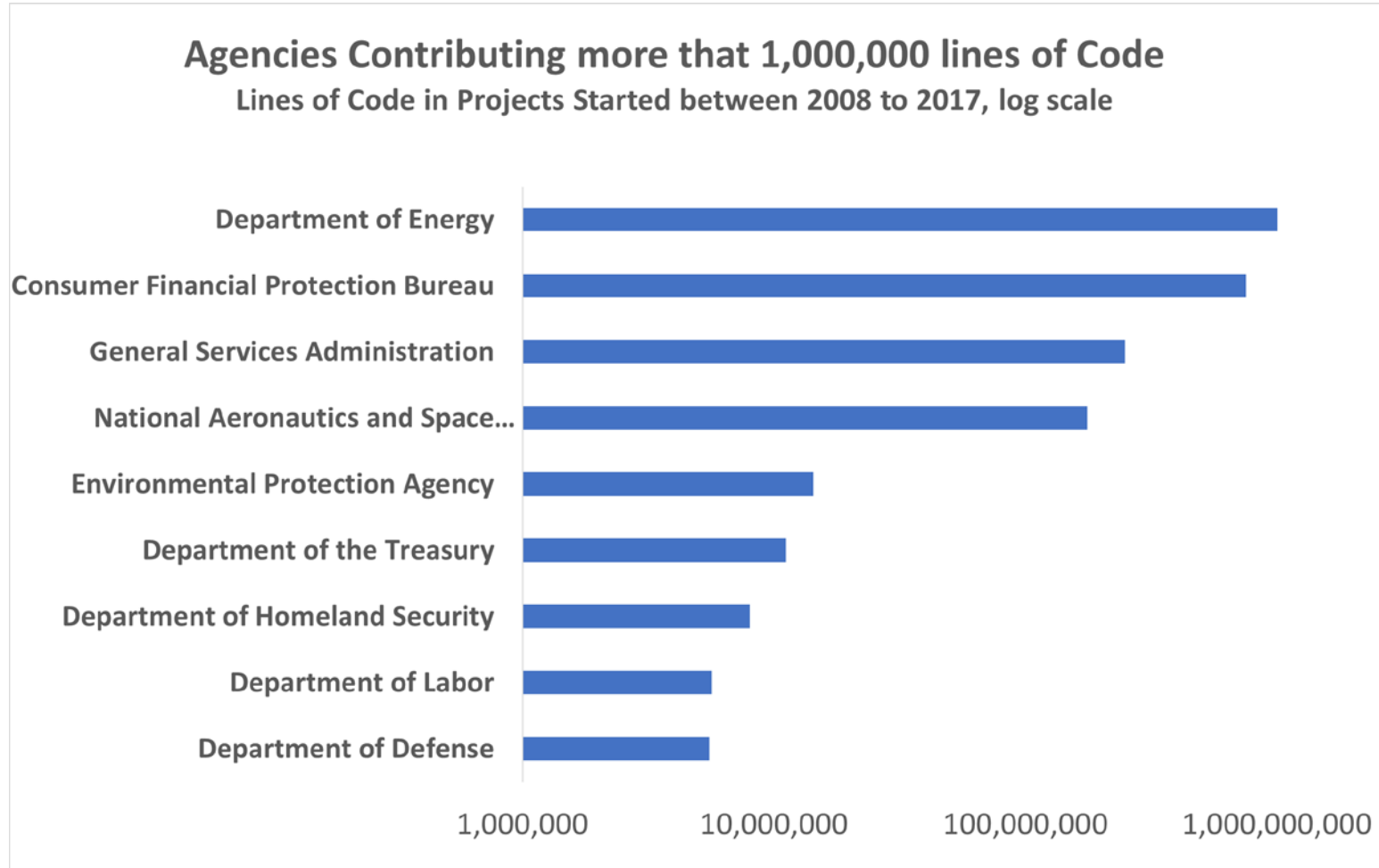
Or

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

Federal Contributions to Code.gov are Widespread, but intensity varies on a log scale

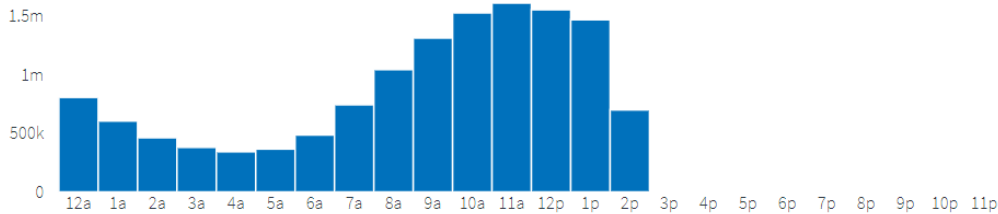


Note: Calculated from the projects on code.gov that are also posted on github.

133,017

people on government websites now

Visits Today



Eastern Time

Visits in the Past 90 Days

There were **2.83 billion** visits over the past 90 days

Devices

Desktop	53.2%
Mobile	41.1%
Tablet	5.7%

Based on rough network segmentation data, we estimate that **less than 5%** of all traffic across all agencies comes from US federal government networks

Browsers

Chrome	46.6%
Safari	28.7%
Internet Explorer	10.7%
Firefox	9.8%
Opera	0.8%
Edge	2.1%

Top Pages

Now	7 Days	30 Days
<i>People on a single, specific page now. We only count pages with at least 10 people on the page. Download the full dataset.</i>		
AIRNow - San Francisco, CA Air Quality		1,769
FAFSA®: Apply for Aid Federal Student Aid		1,538
USPS Tracking®		1,137
Medicare Plan Finder for Health, Prescription Drug and Medigap plans		974
AirNow - Highest Current AQI for CA		894
Universal Enrollment Services (UES)		755
National Weather Service		655
Welcome USPS		597
AIRNow - San Jose, CA Air Quality		576
AIRNow - Sacramento, CA Air Quality		565

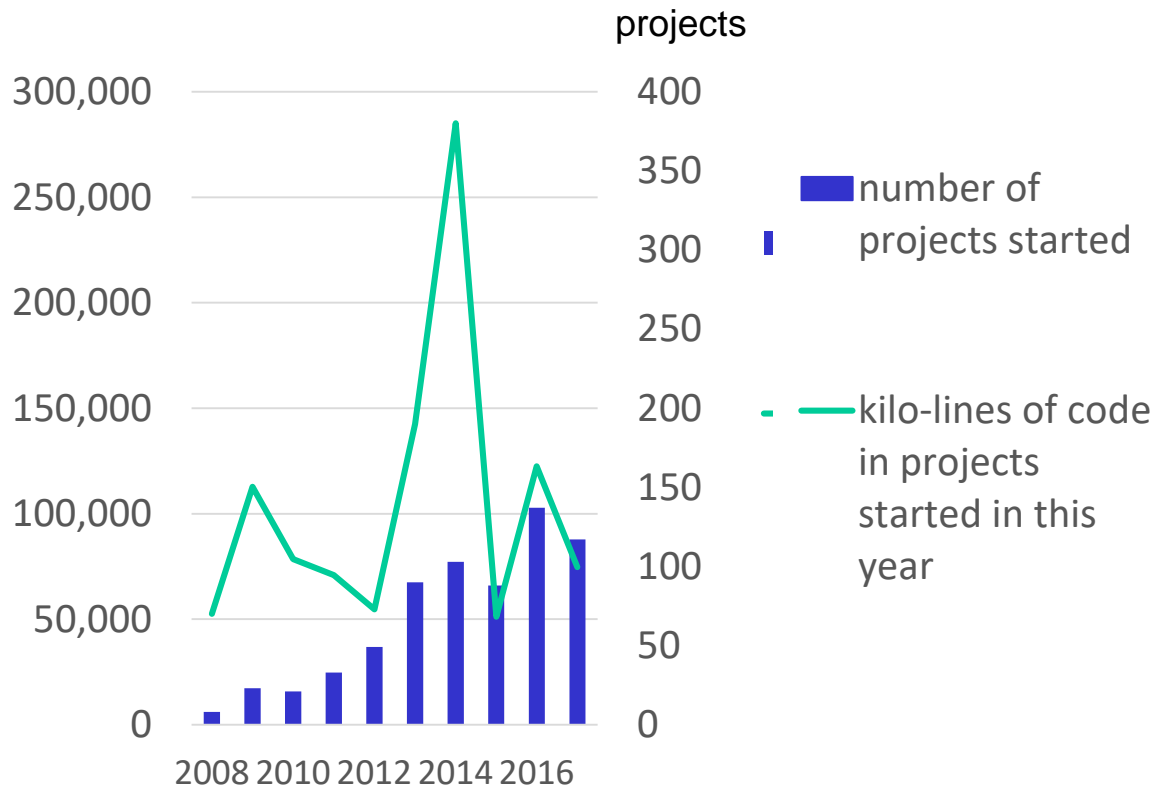
Top Pages

Now	7 Days	30 Days
<i>People on a single, specific page now. We only count pages with at least 10 people on the page. Download the full dataset.</i>		
AIRNow - San Francisco, CA Air Quality		1,769
FAFSA®: Apply for Aid Federal Student Aid		1,566
USPS Tracking®		1,168
AirNow - Highest Current AQI for CA		964
Medicare Plan Finder for Health, Prescription Drug and Medigap plans		918
Universal Enrollment Services (UES)		736

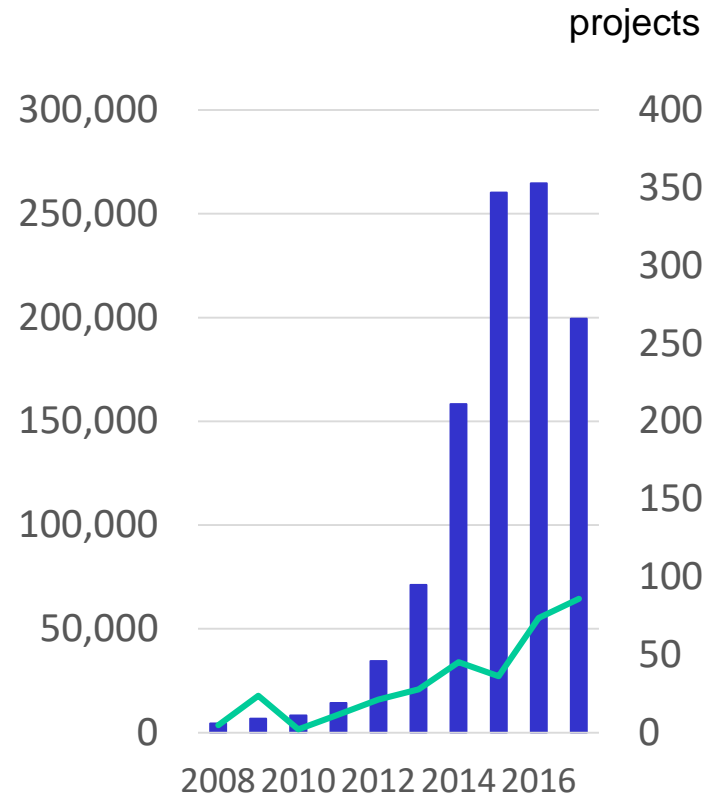
n Time

Projects and Kilo-lines of Code

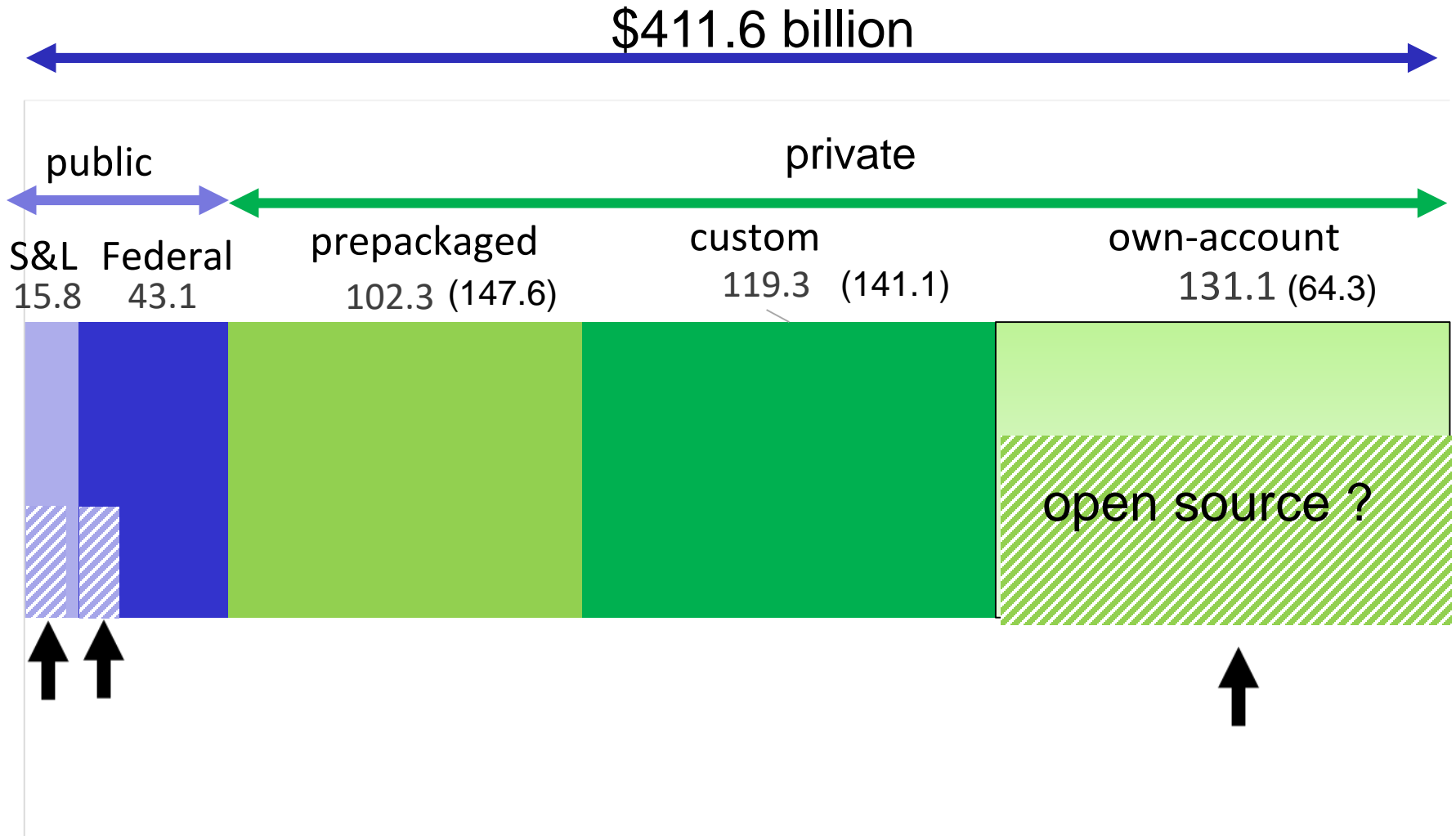
Department of Energy



General Services Administration



U.S. Software Investment 2016 (2017)



Source: BEA. Intellectual Property Products Fixed Asset Tables (private) and Investment in Government Fixed Assets (Table 7.5B).

Framework to Highlight Investment in OSS Software by Different US Sectors

Software subcategory of Intellectual Property Products 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)								

OSS Languages



Data Collection				
Language	Package manager	Number of packages	OSI-approved & production ready	Packages on Github
R	CRAN	12,614	11,886	3,396
Python	PyPi	143,047	7,392	3,804
Julia	Pkg.jl	2,040	1,324	1,324
JavaScript	CDNJS	3,367	3,367	3,213

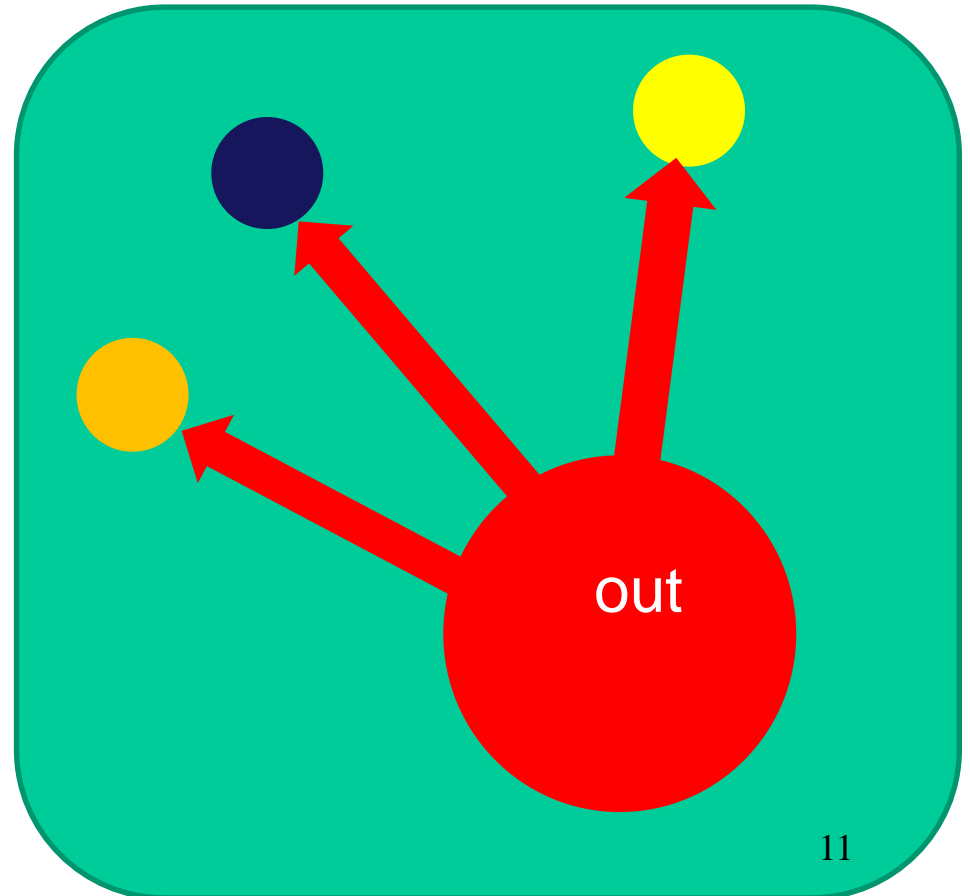
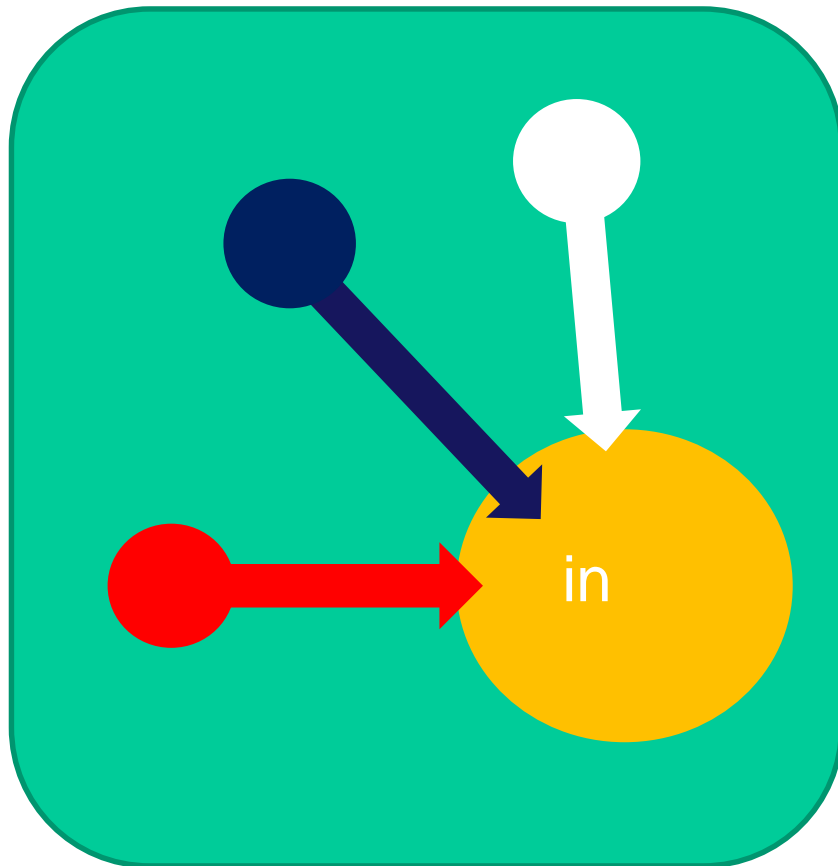


Cost of OSS Software Package Creation

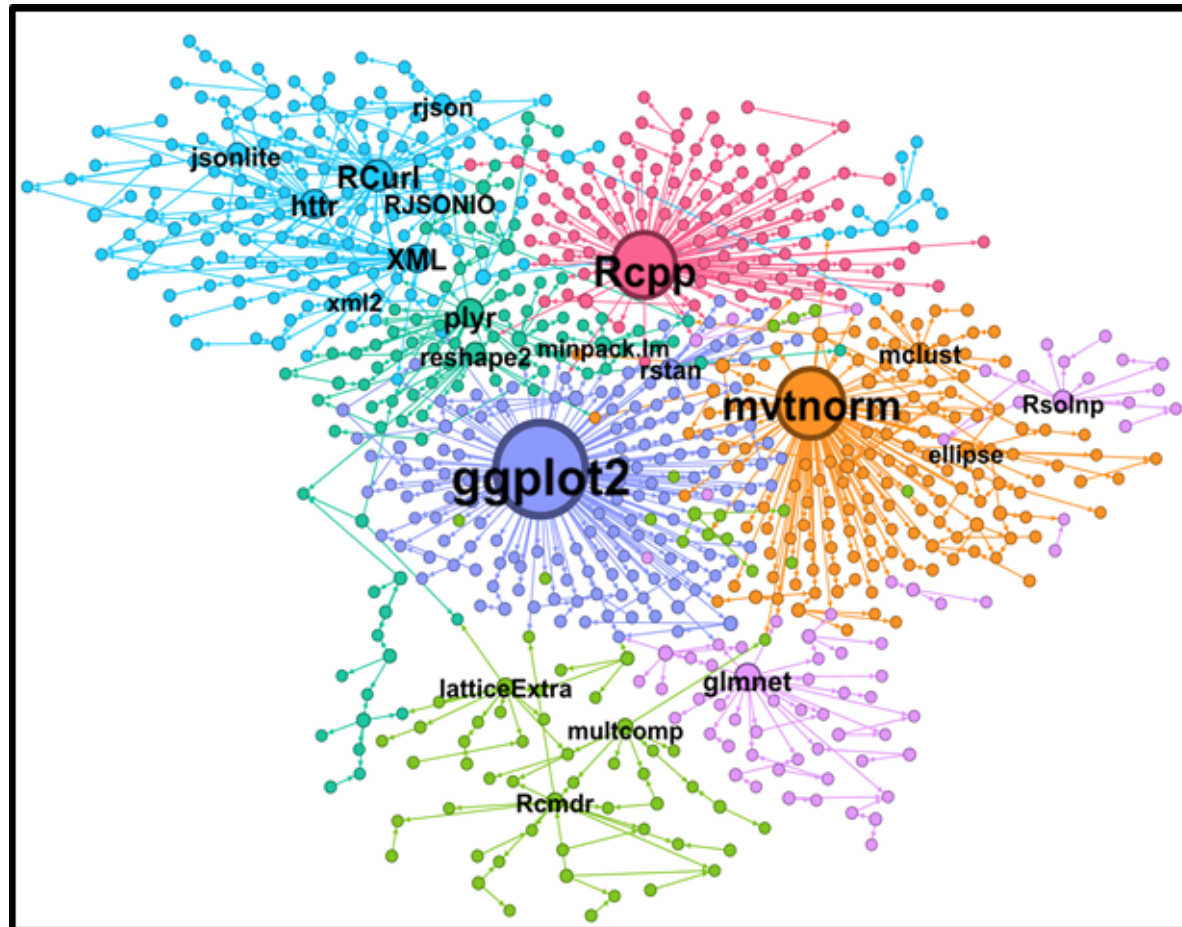
- 1) Kilo-lines of code represent effort
 - Effort is a function of complexity and lines of code
 - historical software project factors
- 2) Estimate resource cost with wage equivalent for 2017
 - Computer programmers, software developers
 - Occupation Employment Survey, Bureau of Labor Statistics
- 3) Estimate non-wage costs adapting OECD and BEA methods

Top 5 packages with the highest total cost (in USD)							
CRAN (R)		PyPi (Python)		Julia (Julia)		CDNJS (JavaScript)	
Package	Cost	Package	Cost	Package	Cost	Package	Cost
googleAnalyticsR	4.4M	Nupic	3.9M	GeoStatsImages	4.2M	Webkit.js	7.6M
Archivist	4.4M	Django-workon	3.6M	PSPlot	2.7M	Phaser-ce	4.9M
Quanteda	3.7M	D1_python	3.5M	MIToS	2.5M	Phaser	4.7M
CollessLike	3.3M	Selenium	3.3M	PDSampler	2.2M	Ag-grid	3.6M
Readtext	3.1M	Senaite.core	3.3M	GeoIP	2.1M	Libsodium.js	3.5M
TOTAL R	942M	TOTAL Python	824M	TOTAL JULIA	264M	TOTAL CDNJS	1,323M

In-degrees and out-degrees are one-way relationships



A Sub-graph of the R Network: many packages depend on ggplot2



Costs of top packages with the highest out-degree

	Package	Out-degree	In-degree	Cost (\$)
R	ggplot2	925	7	1.08M
	Rcpp	838	0	960K
	dplyr	626	10	875K
Python	requests	735	0	643K
	setuptools	182	0	555K
	scipy	131	0	2.58M
	Django	103	0	2.04M
Julia	Compat	596	0	235K
	Distributions	147	7	530K
	StatsBase	136	4	306K
CDNJS	mocha	468	0	694K
	gulp	438	2	236K
	chai	258	0	633K

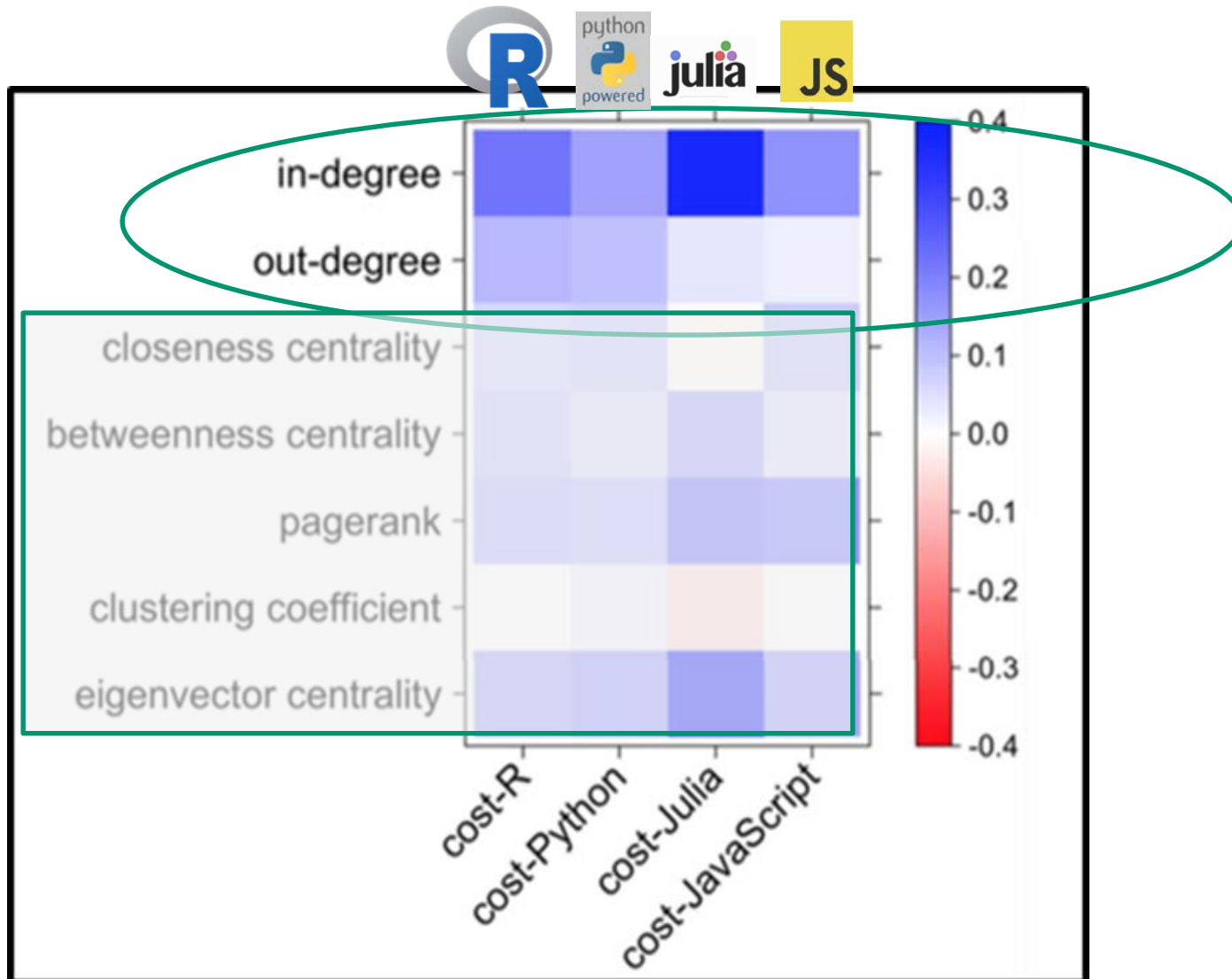
OSS Packages with the Highest Dependencies

Highest total cost (in USD)

Julia (Julia)		CDNJS (JavaScript)	
Package	Cost	Package	Cost
GeoStatsImages	4.2M	Webkit.js	7.6M
PSPlot	2.7M	Phaser-ce	4.9M
MIToS	2.5M	Phaser	4.7M
PDSampler	2.2M	Ag-grid	3.6M
GeoIP	2.1M	Libsodium.js	3.5M

TOTAL R	942M	TOTAL Python	824M	TOTAL JULIA	264M	TOTAL CDNJS	1,323M
---------	------	--------------	------	-------------	------	-------------	--------

Relationship between Cost and Dependencies



Next Steps

- 1) Extend data collection to additional repositories
- 2) Develop annual measures of release ready packages
- 3) Link contributors to sectors and countries
- 4) Refine cost estimates with additional parameters
- 5) Use network analysis to explore additional impact measures

Questions? crobbins@nsf.gov

Grateful Acknowledgement to
Team Open Source Surfers (TOSS)
Virginia Tech Social and Decision Analytics Lab



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