

OSS '19, May 26th 2019, Montréal, Canada

# Why Do People Give Up FLOSSing?

A Study of Contributor Disengagement in Open Source

Courtney Miller, David Gray Widder  
Christian Kästner, Bogdan Vasilescu

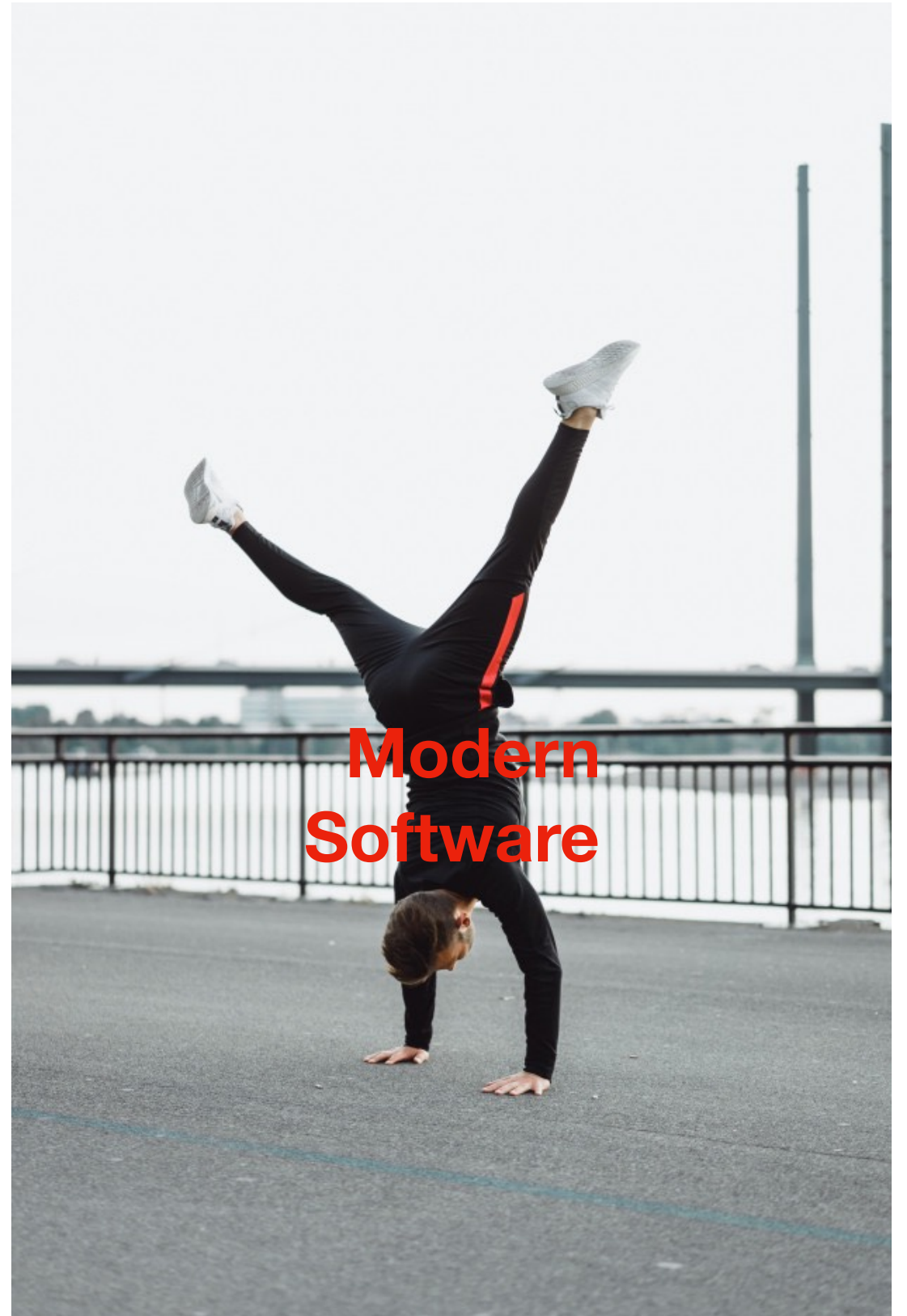
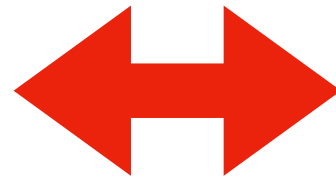


**Carnegie Mellon University**

STRIDEL  
SOCIO-TECHNICAL RESEARCH  
USING DATA EXCAVATION LAB

**isr** institute for  
SOFTWARE  
RESEARCH

# Many OSS Projects are Analogous to Critical Infrastructure





# Survival Rate Among FLOSS Projects is Very Low



**Bad**

**Having Contributors in Drivers Seat is Critical to Project Survival**

**41% of failed OSS projects cited developer issues**

Coelho, J., & Valente, M. T. (2017, August)

**Good**





# Contributor On-Boarding Has Been Thoroughly Researched



Contributor Onboarding

# The Body of Work on Established Contributors is Much Smaller



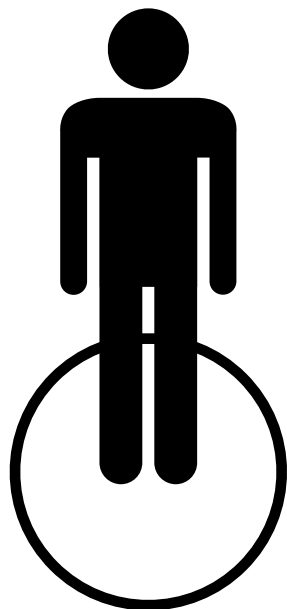
**Established Contributor**

# We're Interested in Exploring Factors to Predict Disengagement from OSS

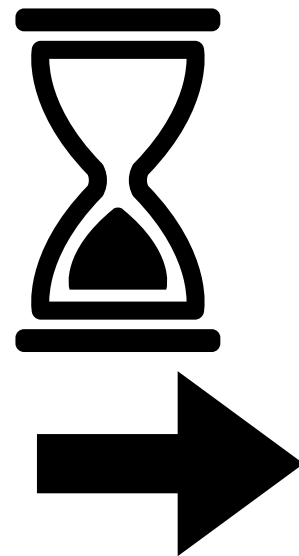
?



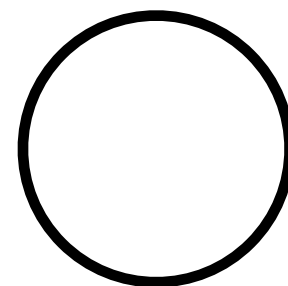
Engaged Contributor



FLOSS

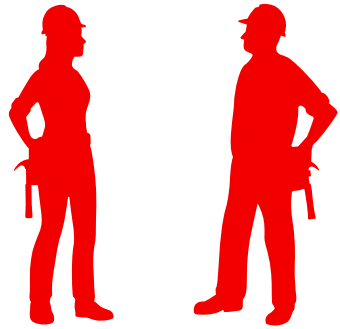


Disengaged Contributor

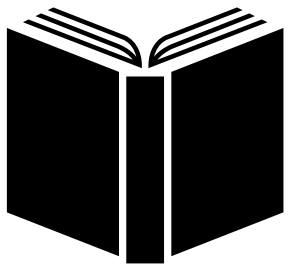
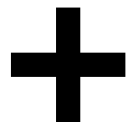


FLOSS

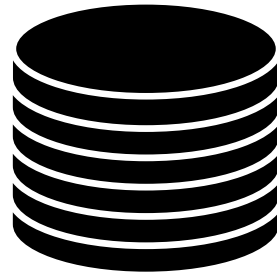
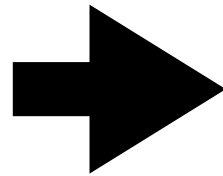
# General Methodology



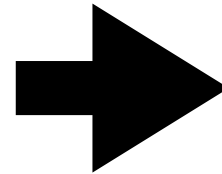
**Survey of  
Disengaged  
Contributors**



**Literature  
Review**



**Operationalization of  
Predictive Factors**

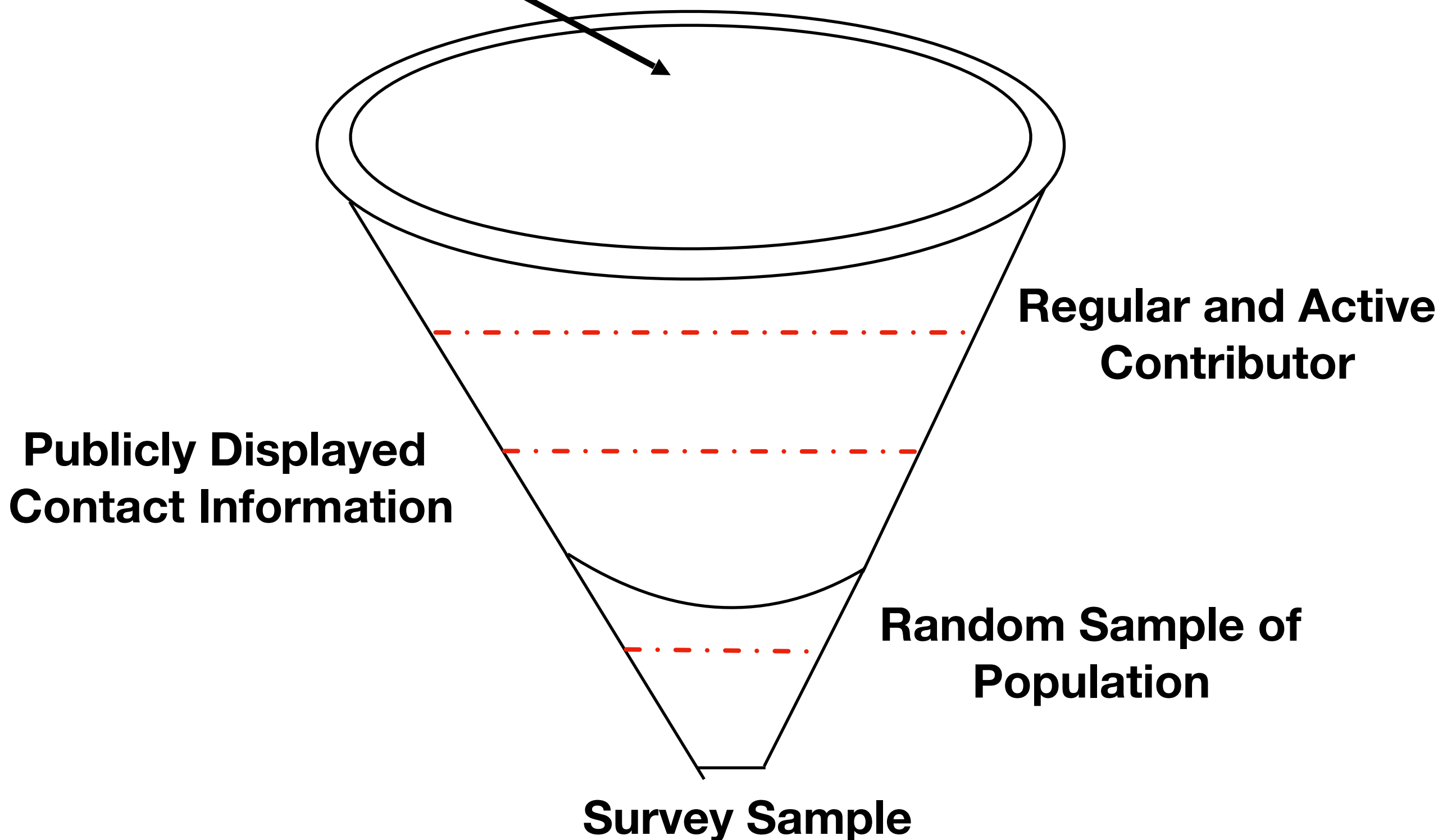


**Survival  
Model**



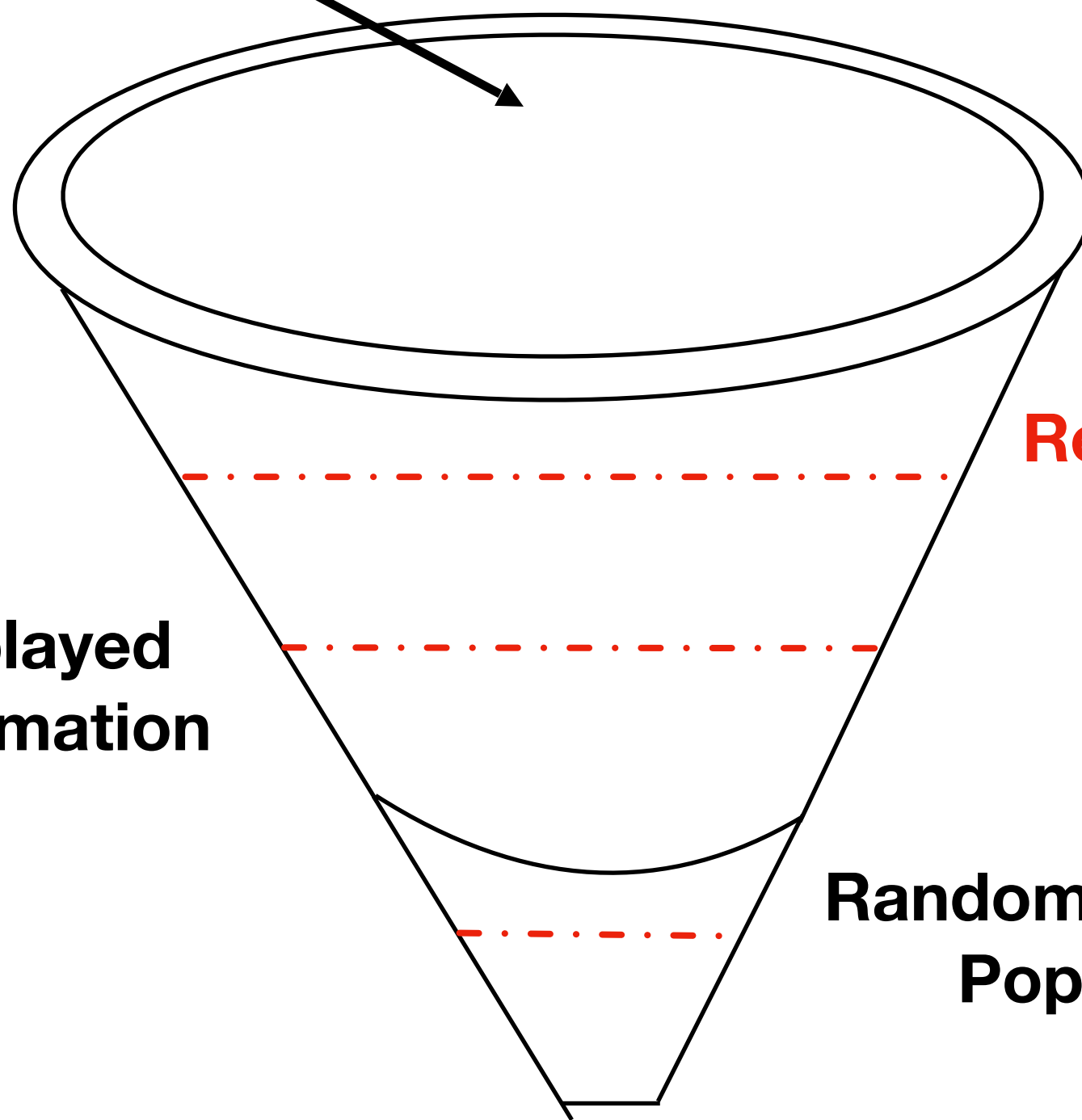
# Used GHTorrent to Collect Survey Sample

**All Public GH Contributors**



# Used GHTorrent to Collect Survey Sample

All Public GH Contributors



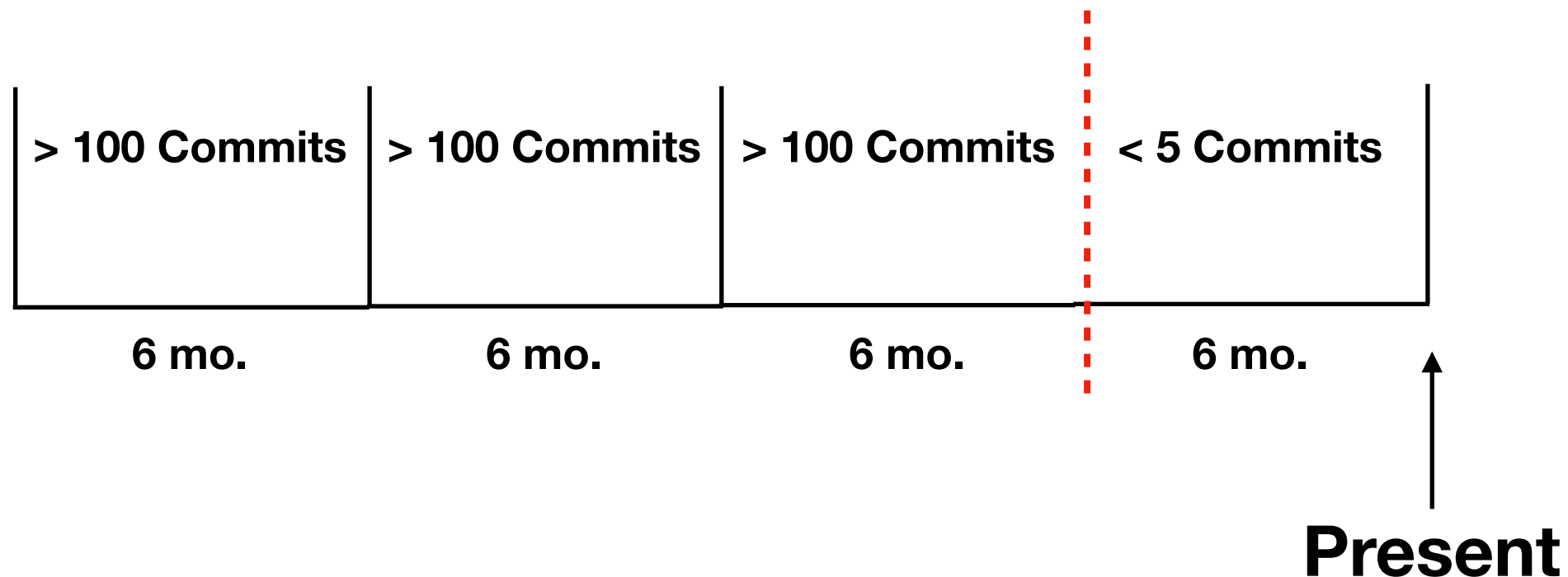
Regular and Active  
Contributor

Publicly Displayed  
Contact Information

Random Sample of  
Population

Survey Sample

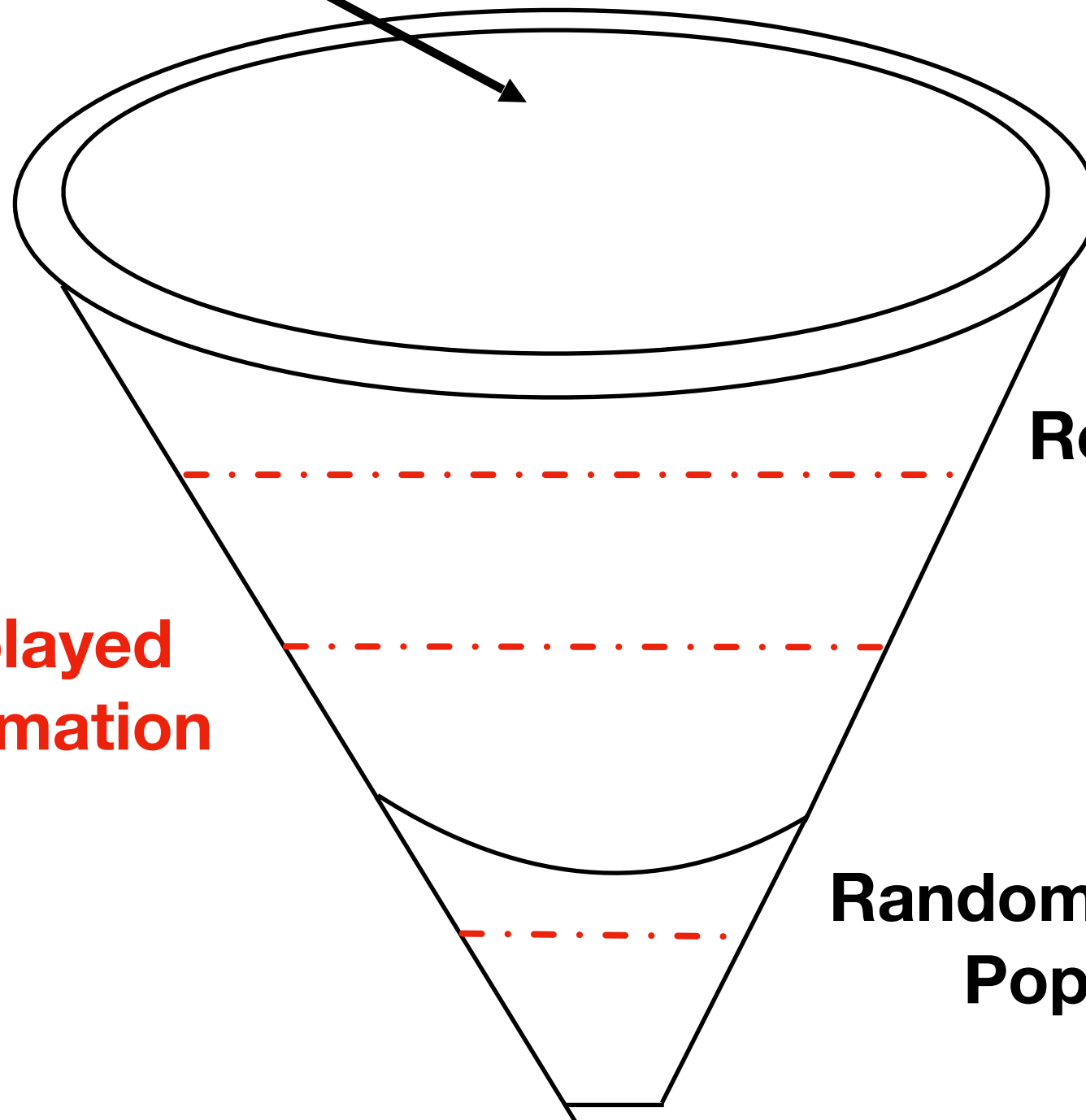
# Identifying Regular and Active Contributors Who Disengaged





# Used GHTorrent to Collect Survey Sample

All Public GH Contributors



Regular and Active  
Contributor

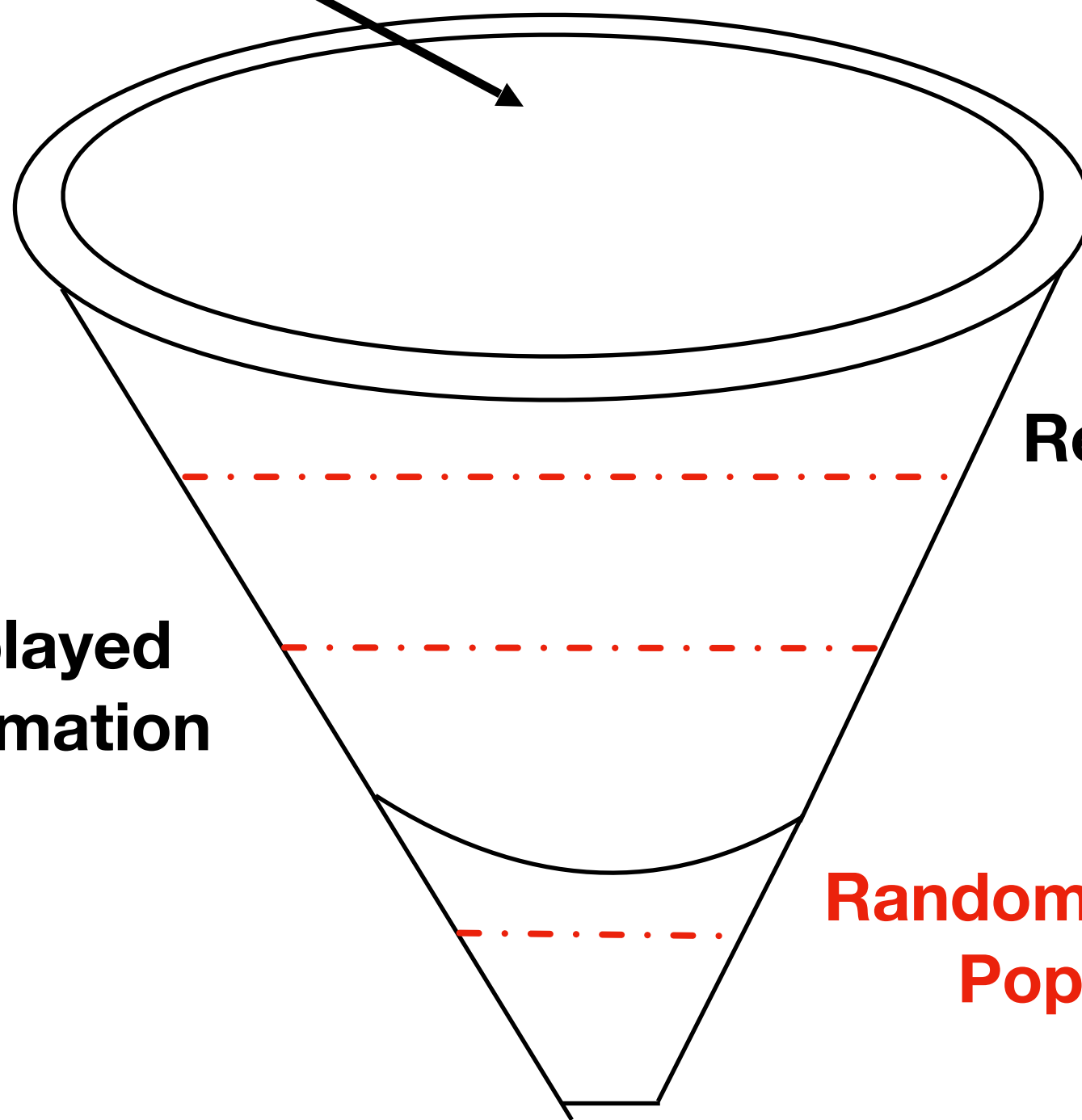
Publicly Displayed  
Contact Information

Random Sample of  
Population

Survey Sample

# Used GHTorrent to Collect Survey Sample

**All Public GH Contributors**



**Regular and Active  
Contributor**

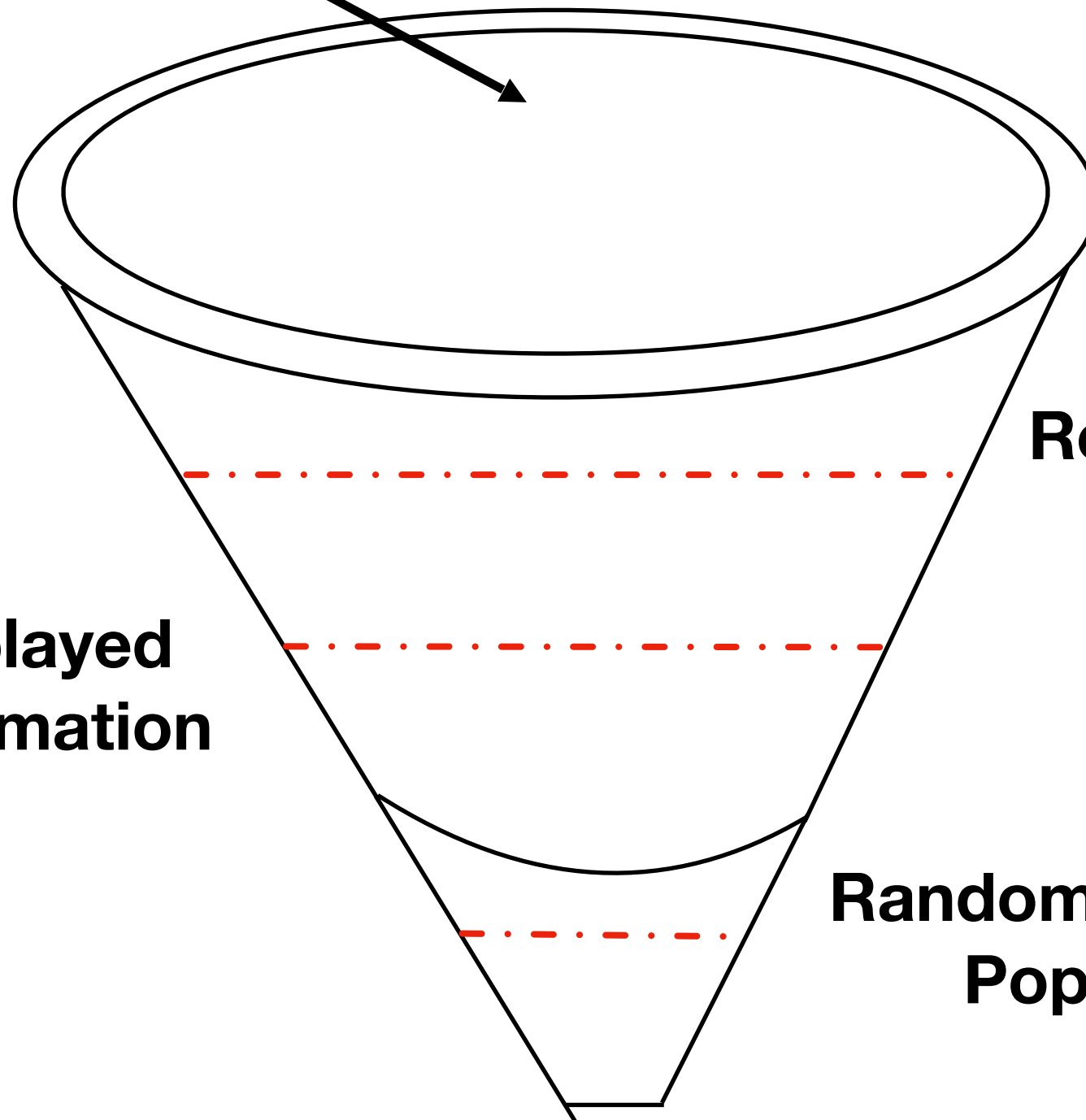
**Publicly Displayed  
Contact Information**

**Random Sample of  
Population**

**Survey Sample**

# Used GHTorrent to Collect Survey Sample

**All Public GH Contributors**



**Regular and Active  
Contributor**

**Publicly Displayed  
Contact Information**

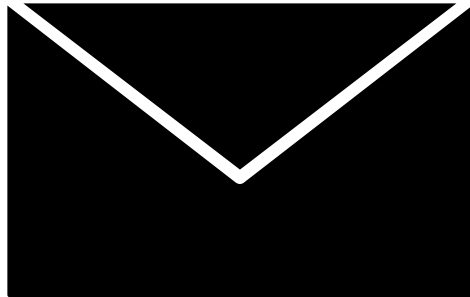
**Random Sample of  
Population**

**Survey Sample**



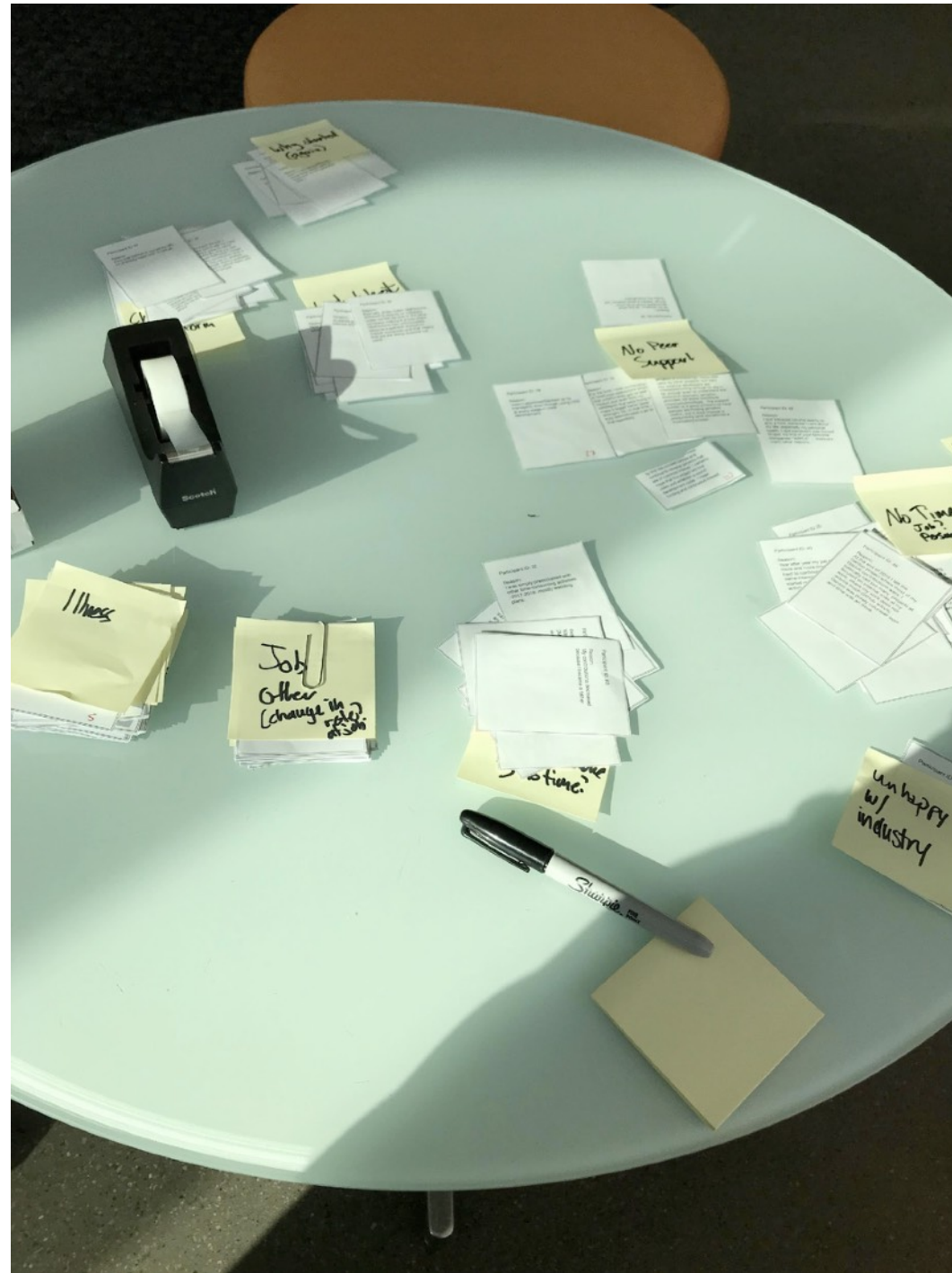
# Sent Out Simple One-Question Survey to Sample

**“Could you help us understand your reasons for reducing your contributions to GitHub projects?”**



**21.5% Response Rate**

# We Used Card Sorting Analysis to Parse Raw Survey Responses



# Macro Survey Results

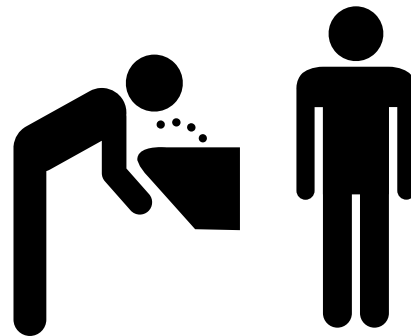
Survey Says Out of 151 Respondents...

70%



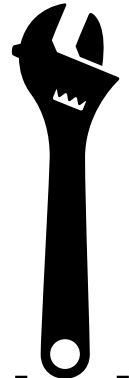
**Occupational  
Issues**

34%



**Social  
Issues**

25%



**Technical  
Issues**



# Breakdown of Survey Responses

Subgroup	Count
<i>Occupational reasons</i>	
Got new job that doesn't support FLOSS	37
Changed role/project	25
Left job where they contributed to FLOSS	16
No time: new job	15
No time: existing job	10
Left school where they contributed to FLOSS	12
No time: in school	12
FLOSS in school, now job doesn't support FLOSS	7
Too much coding at work	4
<i>Social reasons</i>	
Lost interest in FLOSS	24
No time: personal	23
Lack of peer support	16
No time: nondescript	15
<i>Technical reasons</i>	
Issues w GitHub or industry	14
Individually moved to private repos	12
Changed platform	10
Feature complete project	3

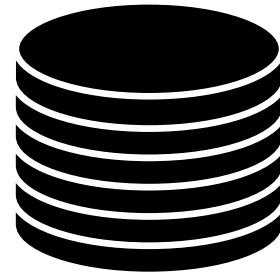
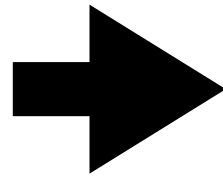
# General Methodology



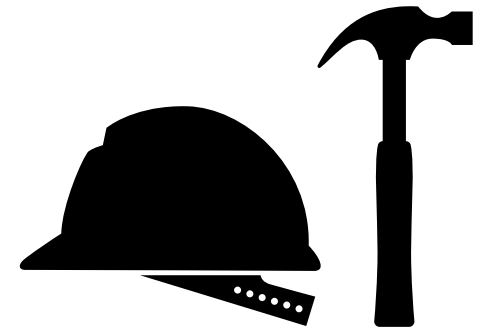
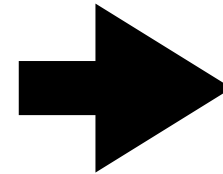
**Survey of  
Disengaged  
Contributors**



**Literature  
Review**

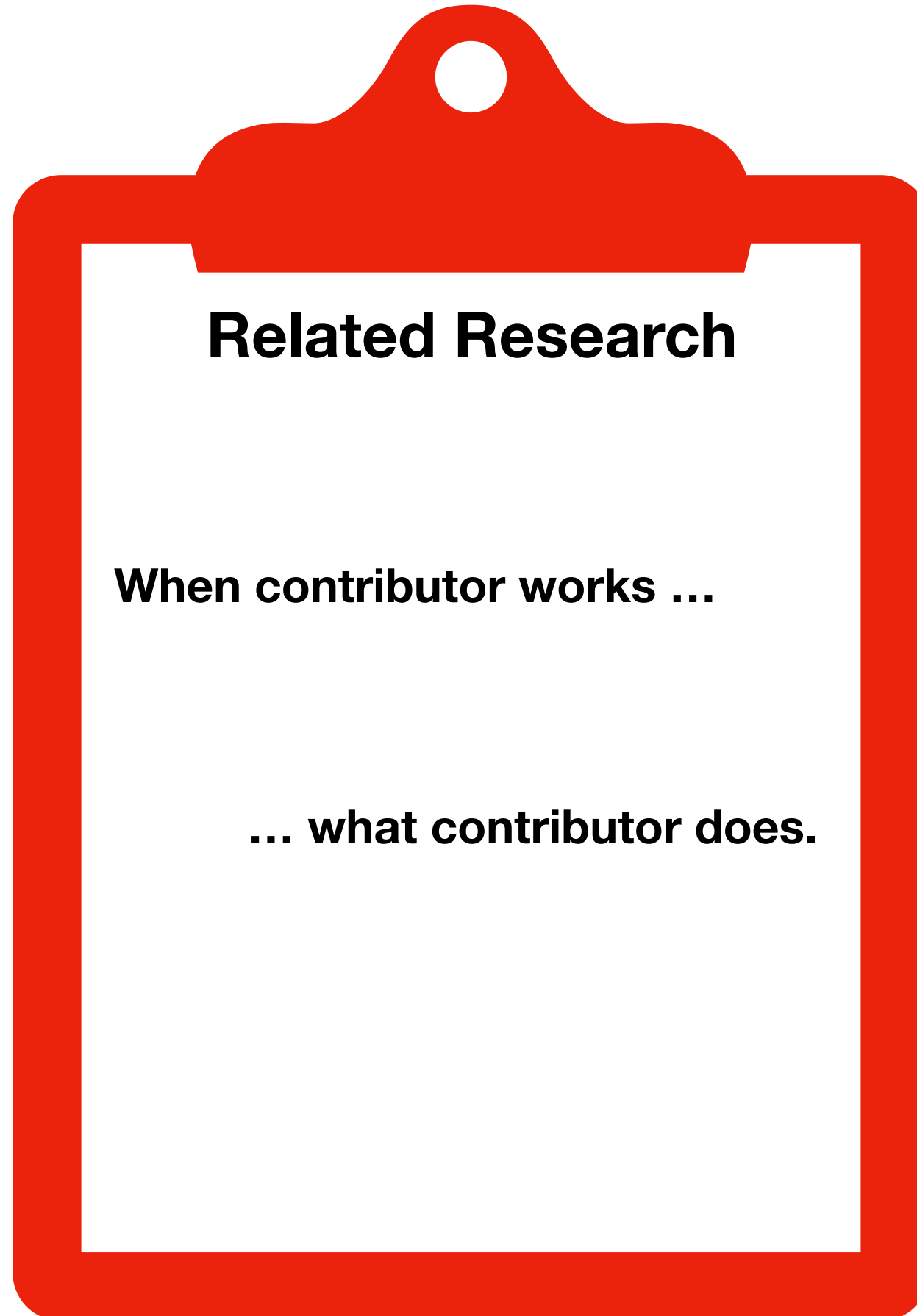


**Operationalization of  
Predictive Factors**



**Survival  
Model**

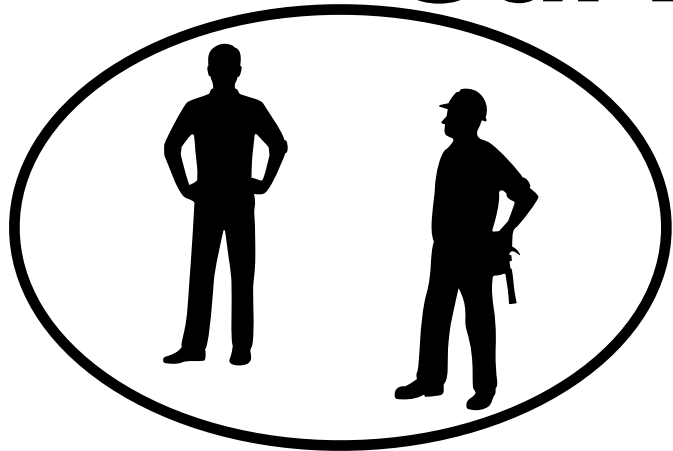
# **We Performed Literature Review**





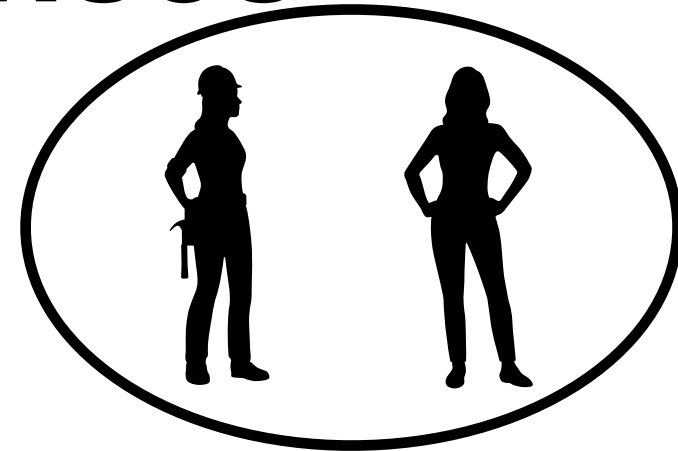
# We Were Interested In Stratifying Survey Responses

**H1:**



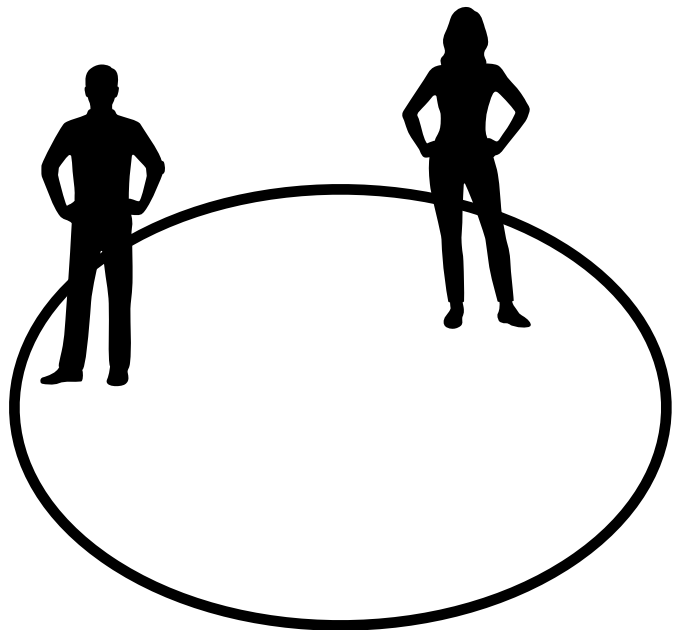
**Office Hours**

**vs.**



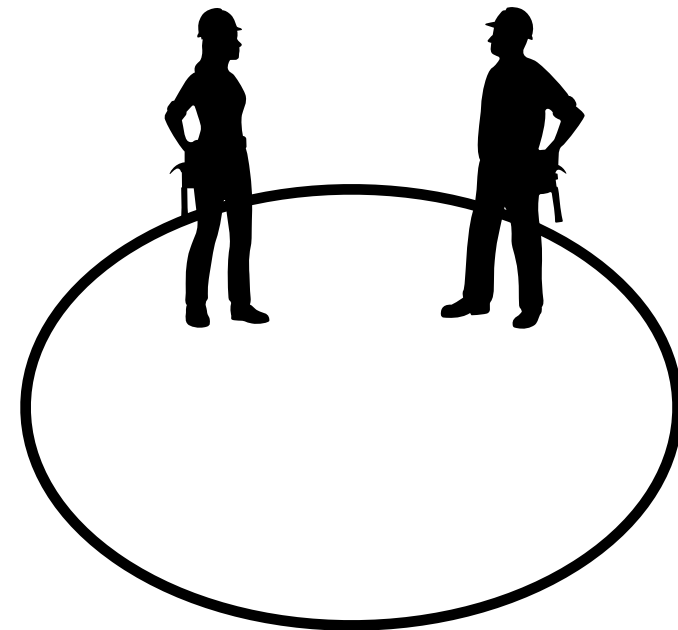
**Nights & Weekends**

**H2:**



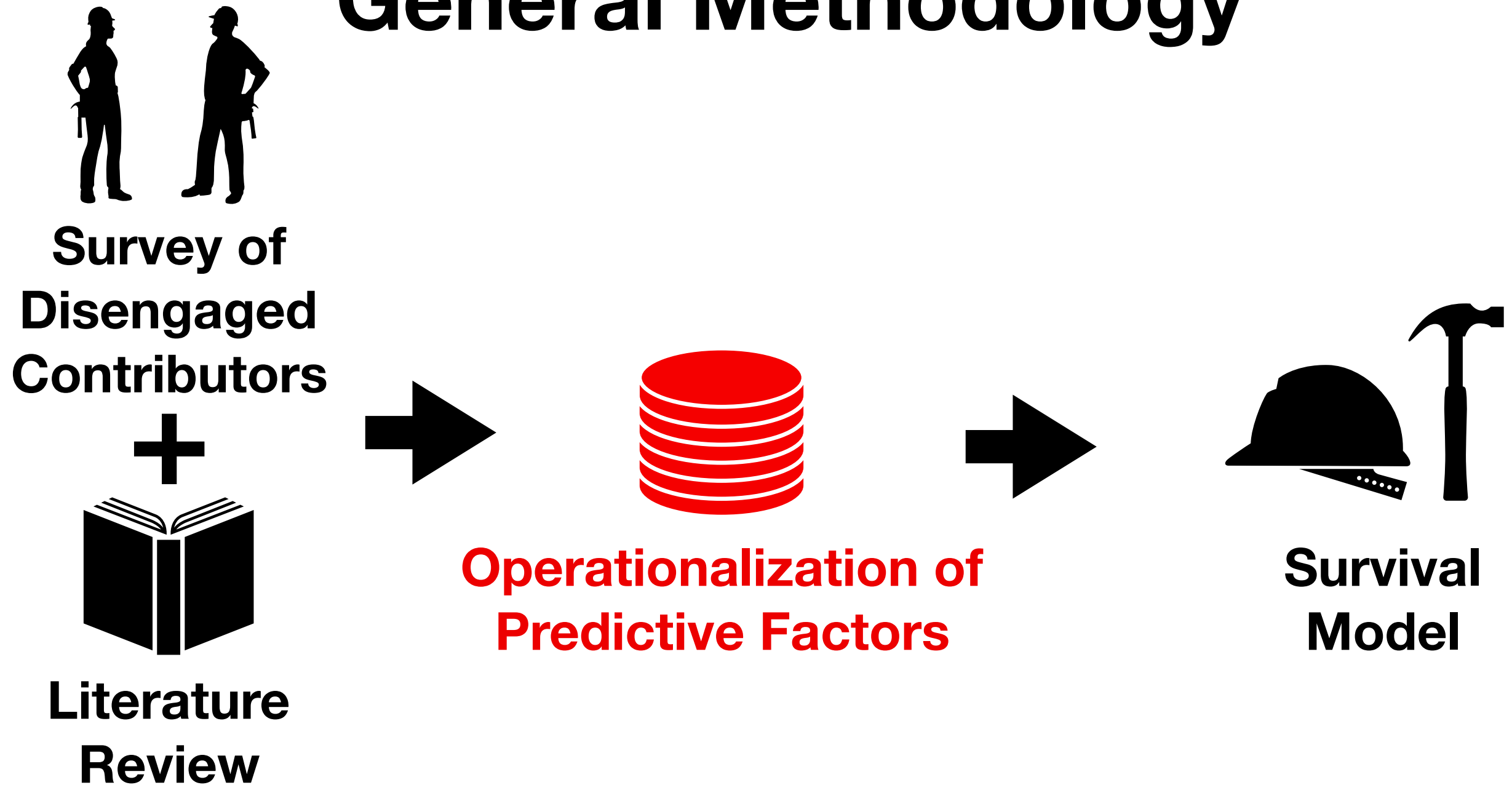
**More Support Work**

**vs.**



**Less Support Work**

# General Methodology

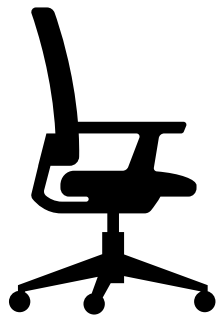


# workingHours

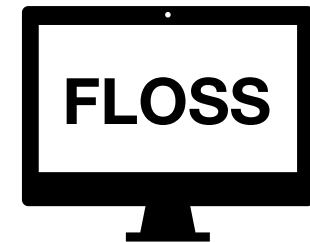
07:00

19:00

Office Hours



Nights & Weekends



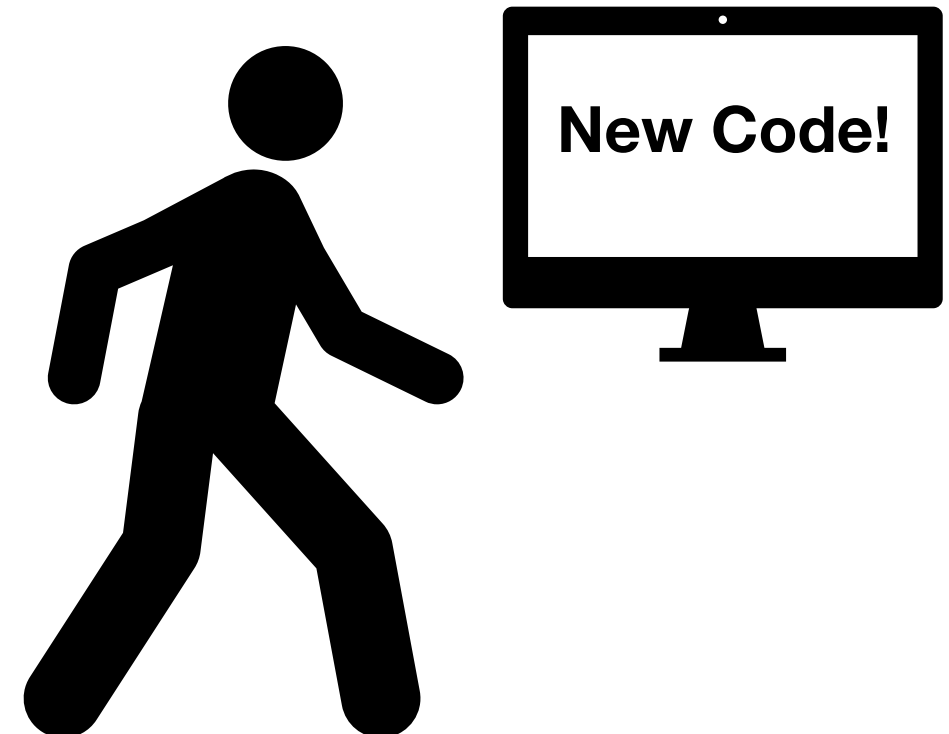
# supportWork

**More Support Work**




**Pull Requests  
Issues  
Etc.**

**Less Support Work**





**Commits**

# Stratified Survey Results

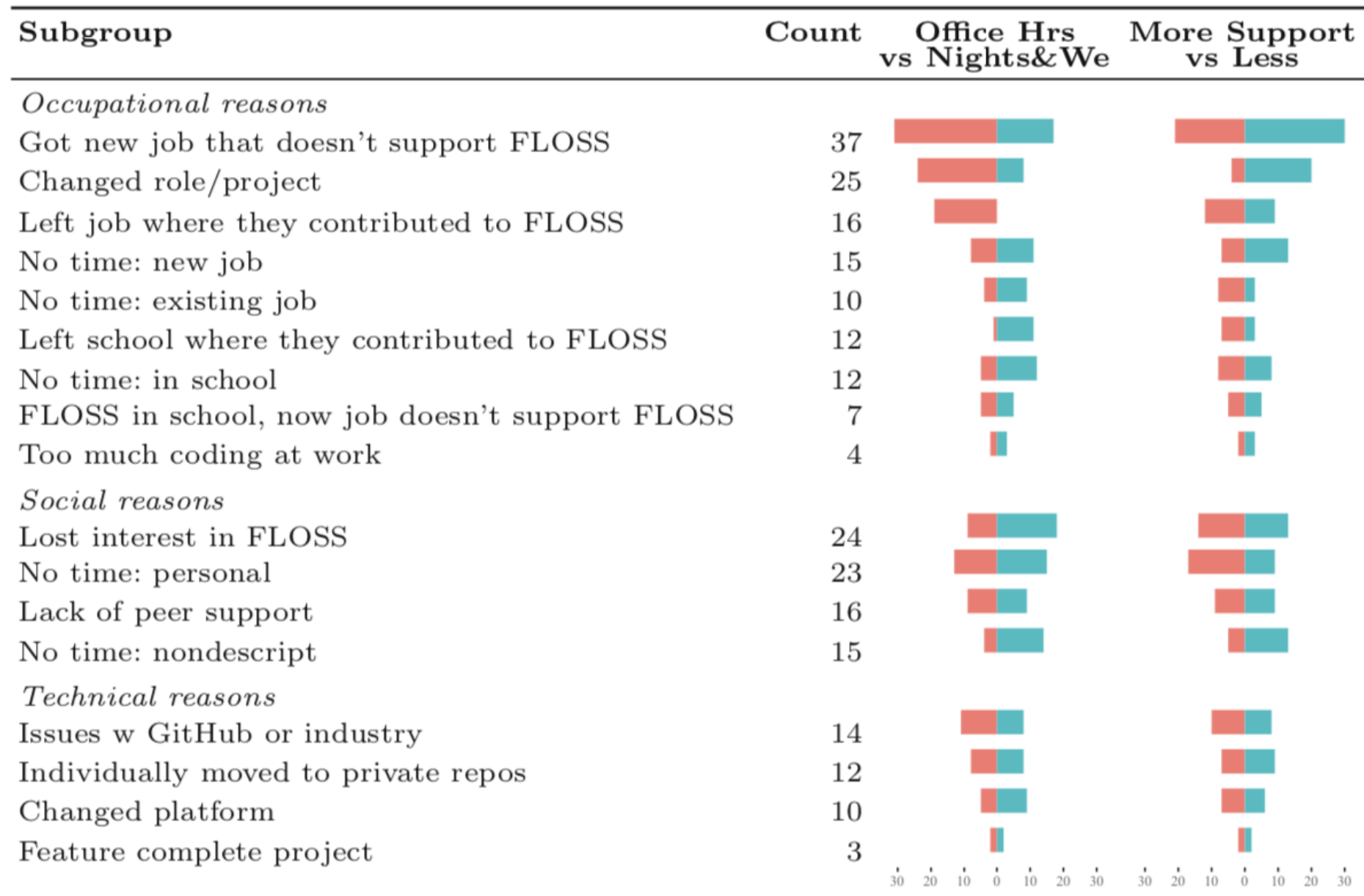
Subgroup	Count	Office Hrs vs Nights&We
<i>Occupational reasons</i>		
Got new job that doesn't support FLOSS	37	



# Stratified Survey Results

Subgroup	Count	Office Hrs vs Nights&We	More Support vs Less
<i>Occupational reasons</i>			
Got new job that doesn't support FLOSS	37		

# Stratified Survey Results



# TransitionFound

## A Good Example: Brett Cannon



### Brett Cannon

Principal software engineer at Microsoft

Vancouver, BC, Canada

[brettsky](#)

[github.com/brettcannon](#)

<http://snarky.ca>

Last seen on Stack Overflow 3 days ago



#### TECHNOLOGIES

Preferred technologies

[python](#) [dart](#) [typescript](#)

Non-preferred technologies

[c++](#) [java](#)

Top Percentiles

Top 10% [python](#) [visual-studio-code](#)

#### EXPERIENCE (10)



Principal software engineer

[Microsoft](#)

Jul 2015 → Current (3 years, 11 months)



Fellow

[Python Software Foundation](#)

Mar 2003 → Current (16 years, 3 months)



Software Engineer

[Google](#)

May 2011 → Jun 2015 (4 years, 2 months)

# TransitionFound

## Identifying Transitions

---

### EXPERIENCE (10)



Principal software engineer

Microsoft

Jul 2015 → Current (3 years, 11 months)



Fellow

Python Software Foundation

Mar 2003 → Current (16 years, 3 months)



Software Engineer

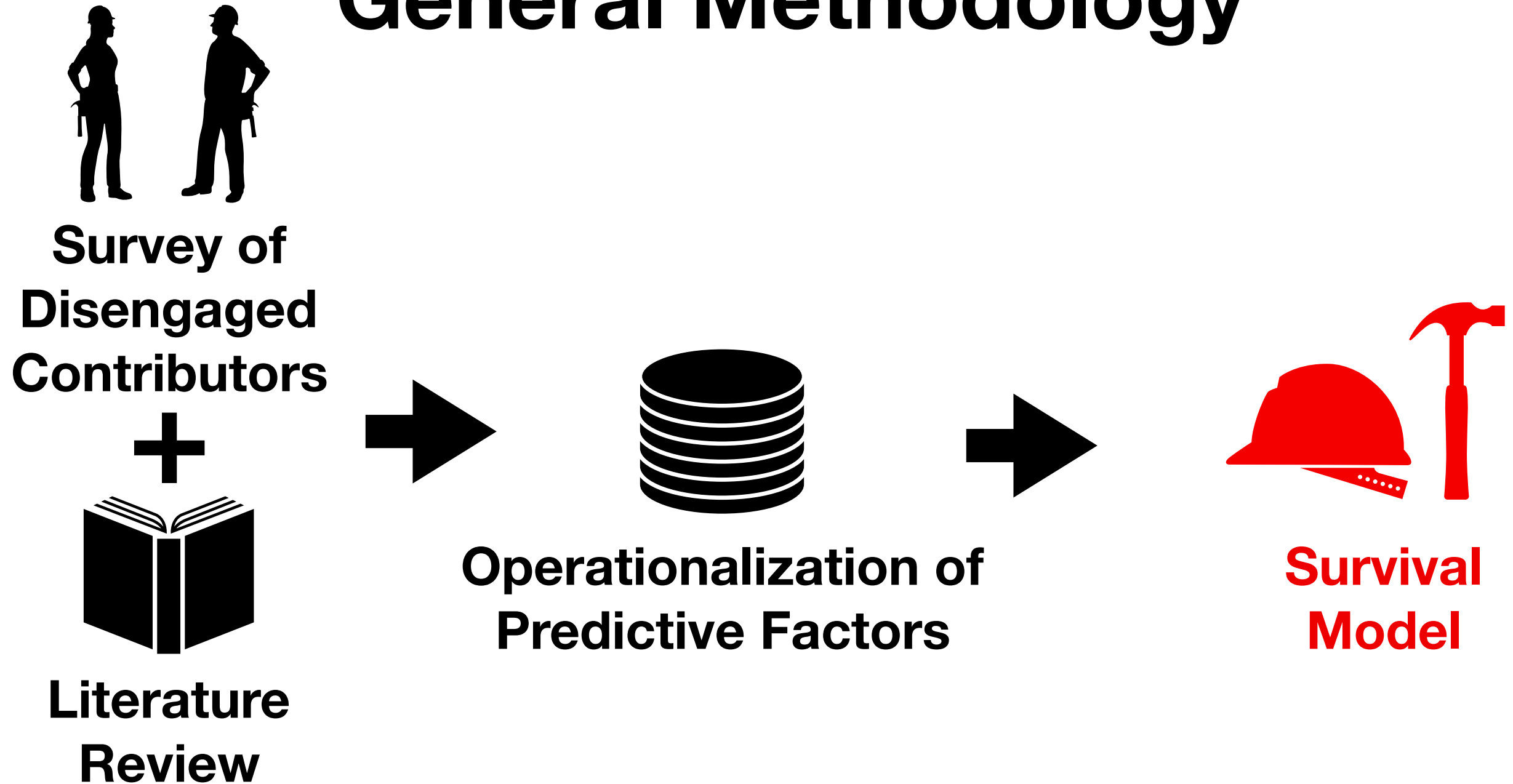
Google

May 2011 → Jun 2015 (4 years, 2 months)

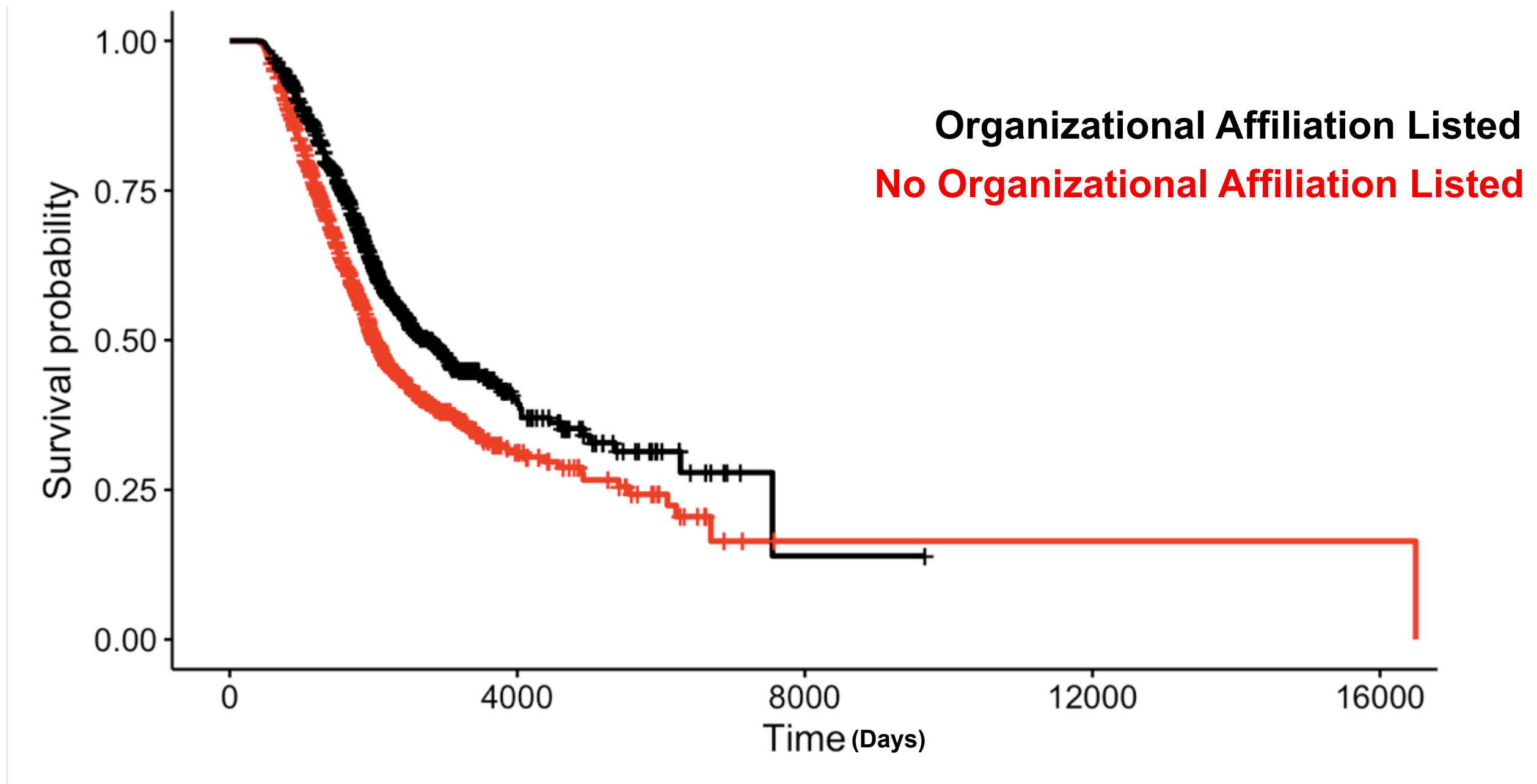
python go c++ java google-app-engine google-chrome-extension android

- Charged with finishing the master/slave to HRD migration tool for App Engine for public release
- Designed a novel algorithm to scrape web pages for *[redacted]* and implemented it as an entirely client-side Chrome extension which included live image cropping
- Heavily participated in and solely maintained an email parsing pipeline for purchase receipts to collect purchase-related data (e.g., shipment tracking numbers), which included adding support for [schema markup](#) and working with ontologists to add missing structured data support to [schema.org](#)
- Took the Google Now in-store card from a UX concept to implementation and through to launch which required writing both the Android client code along with the server code that worked with disparate backends

# General Methodology



# But What Exactly Is A Survival Model?

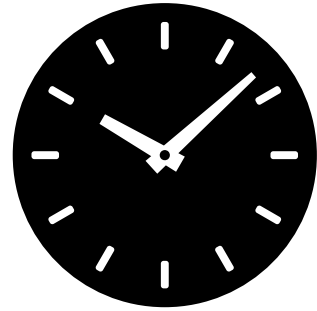




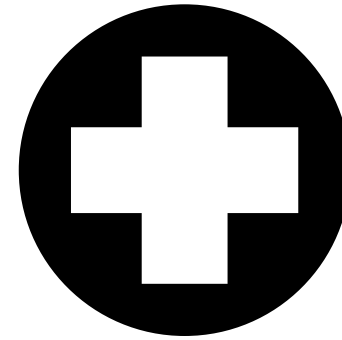
# Survival Model Factors



**Activity**



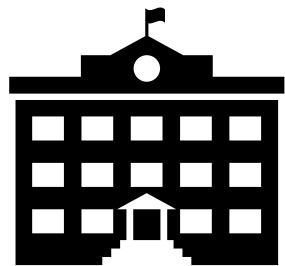
**WorkingHours**



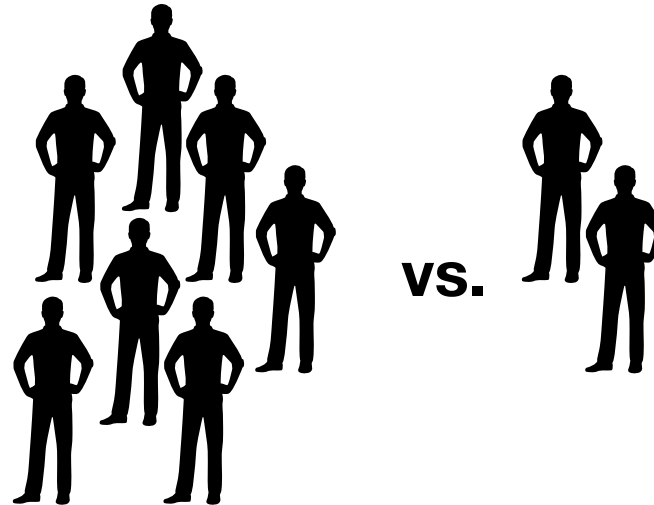
**SupportWork**



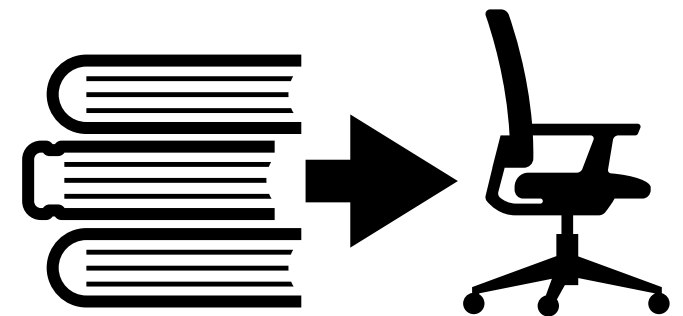
**Project  
Popularity**



**Organizational  
Affiliation**

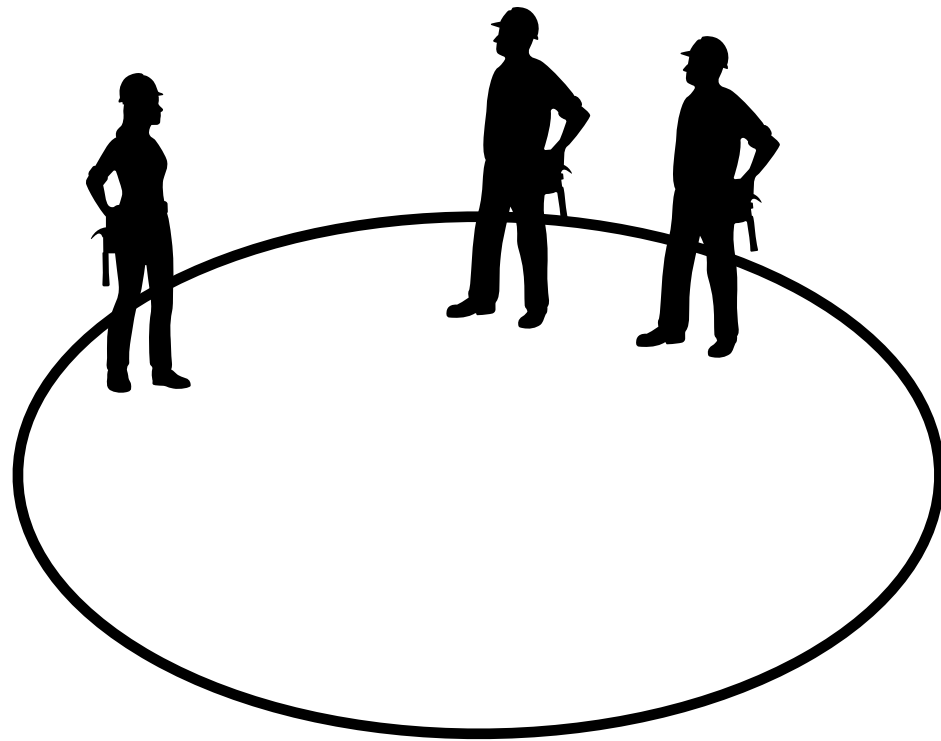


**Team  
Size**



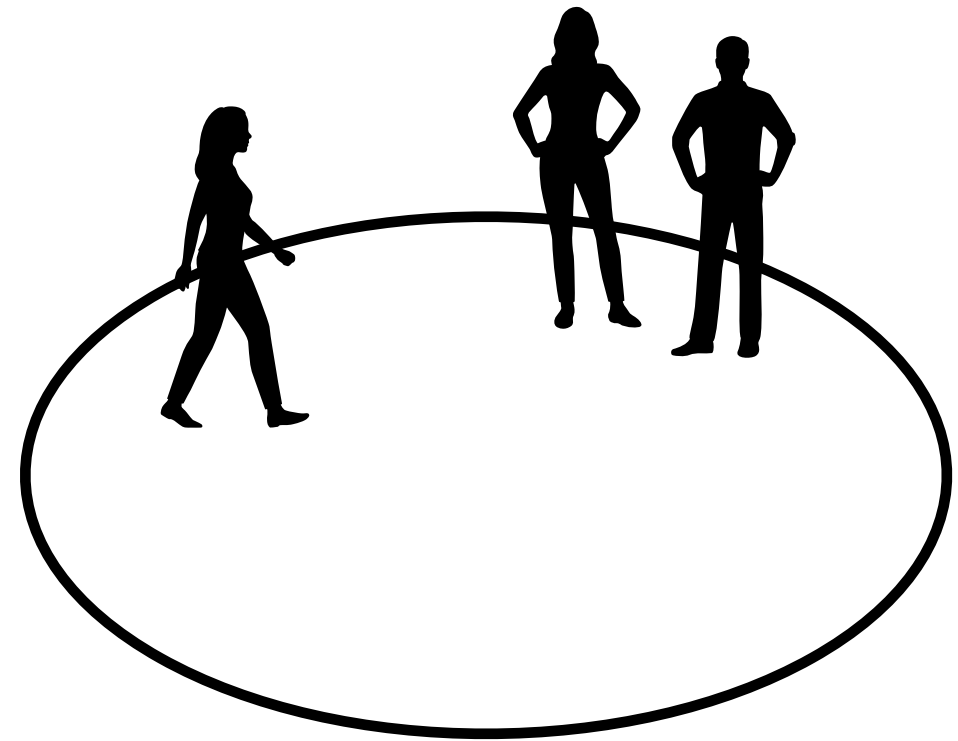
**Transition  
Found**

# Survival Model Participants



**Survey Respondents  
with Transition Data**

+



**Control Group**

# Participants with Transition 2.48 Times More Likely to Disengage



**“I started employment with a company that **didn’t really like open sourcing their projects.**” (P25)**

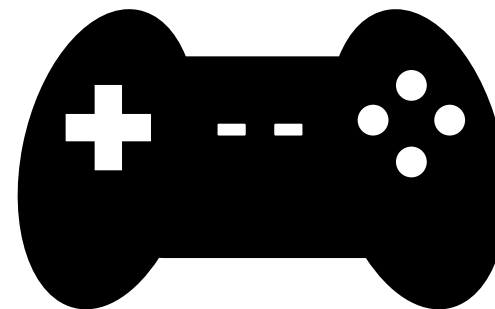
**“I’ve moved to more of a product owner role in my position, **so less time to develop.**”(P50)**

# Working Predominantly During Office Hours Increased Risk of Disengagement by Factor of 2.20



**Mandatory  
Participation**

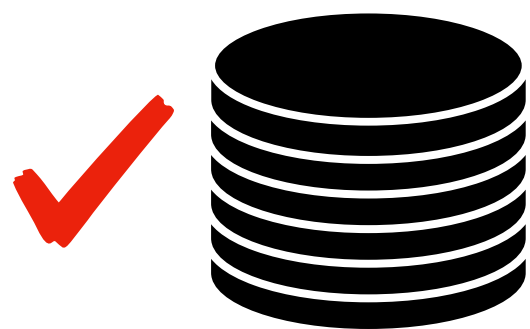
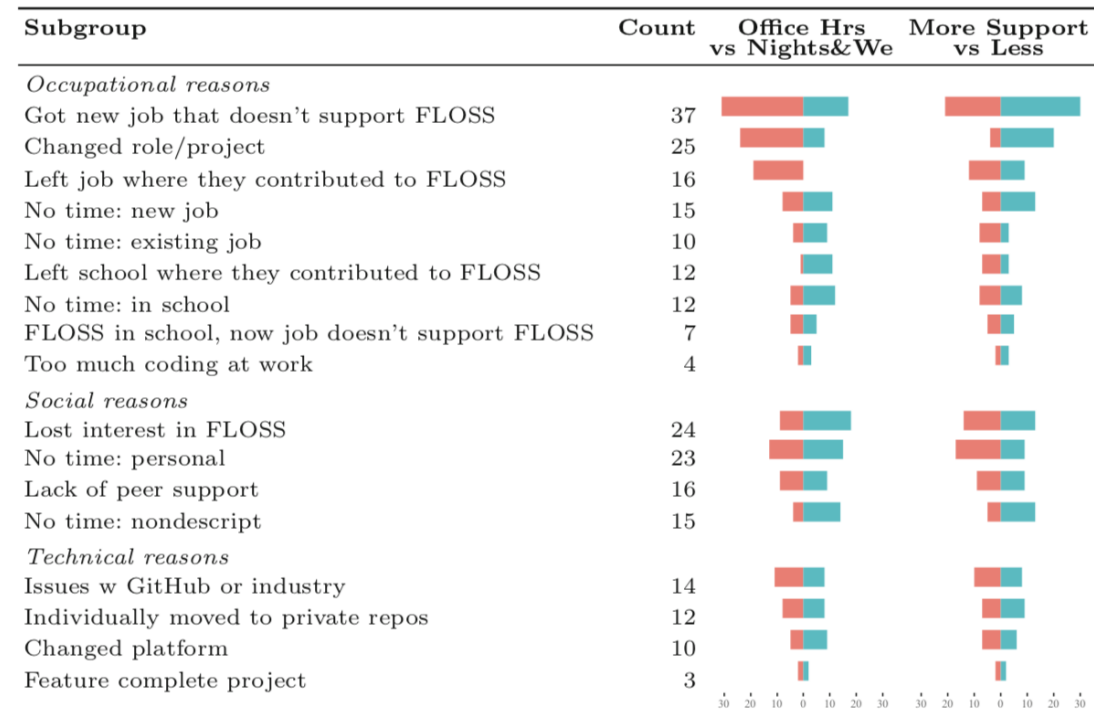
**vs.**



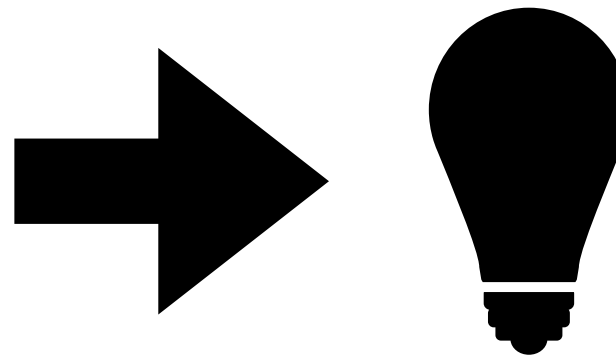
**Elective  
Participation**

**“[It’s] quite simple. I was active because the project was sponsored by my employer, i.e., my job. I'm not active on GitHub because I've moved to another project.”(P75)**

# Operationalization of Predictive Factors Provided Promising Results



**Data Driven  
Operationalization  
Of Predictive Factors**

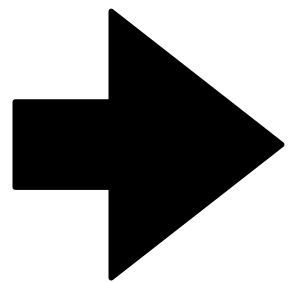



**Valuable Insights  
About Disengaged OSS  
Contributors**

# To learn more about sustained participation of contributors in FLOSS...

Friday, May 31st at 11:20, Sophie Qiu

**Going Farther Together:  
The Impact of Social Capital on Sustained Participation in Open Source**



11:00 - 12:30: <b>Papers</b> - Developer Biases and Trust at <b>Viger</b>		Chair(s): <b>Kelly Blincoe</b> University of Auckland	
11:00 - 11:20	★	<b>FLOSS Participants' Perceptions about Gender and Inclusiveness: A Survey</b> Amanda Lee University of Alabama, Jeff Carver University of Alabama <a href="#">Pre-print</a>	INDUSTRY PROGRAM TECHNICAL TRACK
11:20 - 11:40	★	<b>Going Farther Together: The Impact of Social Capital on Sustained Participation in Open Source</b> Huilian Sophie Qiu Carnegie Mellon University, Pittsburgh, Pennsylvania, United States, Alexander Nolte University of Tartu, Anita Brown Bryn Mawr College, Bryn Mawr, Pennsylvania, United States, Alexander Serebrenik Eindhoven University of Technology, Bogdan Vasilescu Carnegie Mellon University <a href="#">Pre-print</a>	INDUSTRY PROGRAM TECHNICAL TRACK 
11:40 - 12:00	★	<b>Investigating the Effects of Gender Bias on GitHub</b> Nasif Imtiaz North Carolina State University, Justin Middleton , Joymallya Chakraborty , Neill Robson NC State University, Gina Bai North Carolina State University, Emerson Murphy-Hill Google <a href="#">Pre-print</a>	INDUSTRY PROGRAM TECHNICAL TRACK



# Key Takeaways

**Occupational reasons were most commonly cited in survey, specifically occupational transitions and novel transition**  
**Found operationalization proved successful factor in survival model**

**Used operationalizations of potential disengagement factors to (1) categorize survey respondents; and (2) predict disengagement**

**Established contributors who do different kinds of work and perform said work at different times tend to cite different reasons for their disengagement.**

**Factors such as project popularity, when contributors work, and how much they work can also be used to predict established contributor disengagement from FLOSS**

## Questions?

**Courtney Elta Miller**

 **@CourtneyElta**

**Courtney.Miller17@ncf.edu**