WHAT DO WE STUDY?

ALEXANDER SEREBRENIK

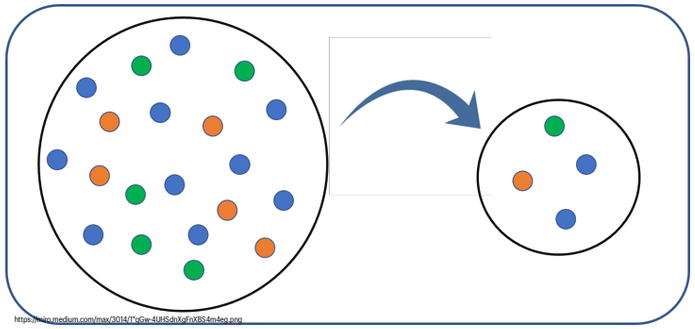
PLEASE TELL US WHAT YOU THINK

RECAP: both Nathan and I will do our best to make this course interesting for you. Please do not wait till the end of the course to provide us feedback.

DATA COLLECTION

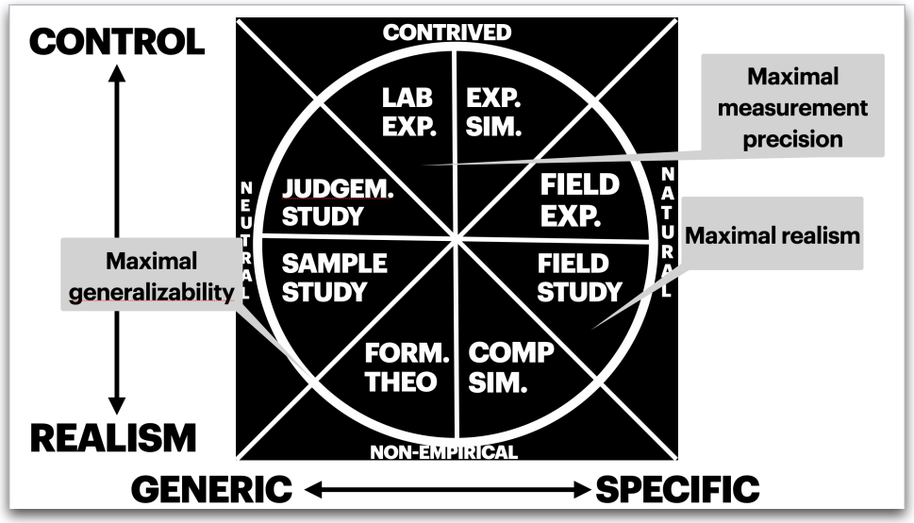
Today we start discussion data collection techniques. Today we are going to talk about **sampling**.

SAMPLING IS THE PROCESS OF SELECTING A SMALLER GROUP OF ITEMS TO STUDY (A **SAMPLE**) FROM A LARGER GROUP OF ITEMS OF INTEREST (**POPULATION**).



https://miro.medium.com/max/3014/1*qGw-4UH5drXqFnX8S4m4eg.png

SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)



WHAT RESEARCH STRATEGIES REQUIRE SAMPLING?

All! For field study and field experiment you need to decide which of the “fields” to consider, i.e., what company or project to study. For laboratory experiment you need to select participants, and ideas for the judgement study (“experts” or “members of the jury”). For simulation we need to select what process we would like to simulate, and for experimental simulations what participants to select. Even for formal theory we need to select the literature to derive the theory

Sampling is the process of selecting a smaller group of items to study (a sample) from a larger group of items of interest. The group of items of interest is called the population, and the (usually imperfect) population list is called the sampling frame.

POPULATION

SAMPLING FRAME

SAMPLE

open source initiative

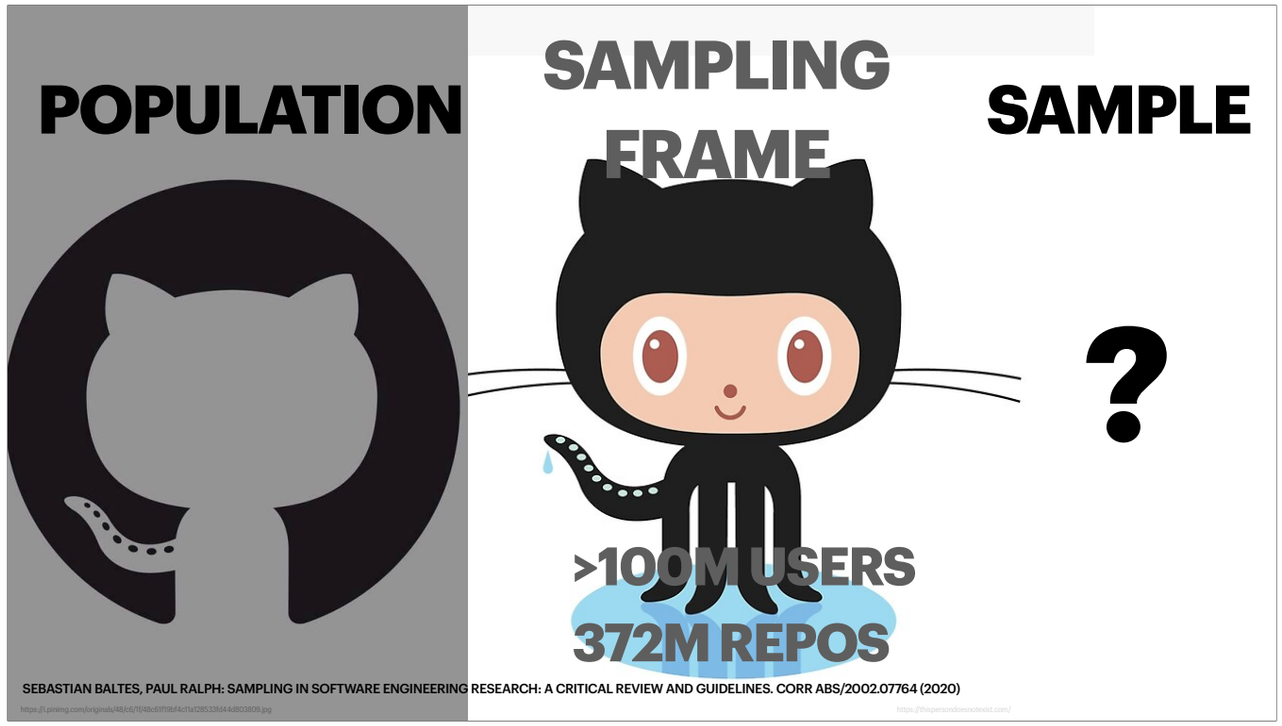
?

?

SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)

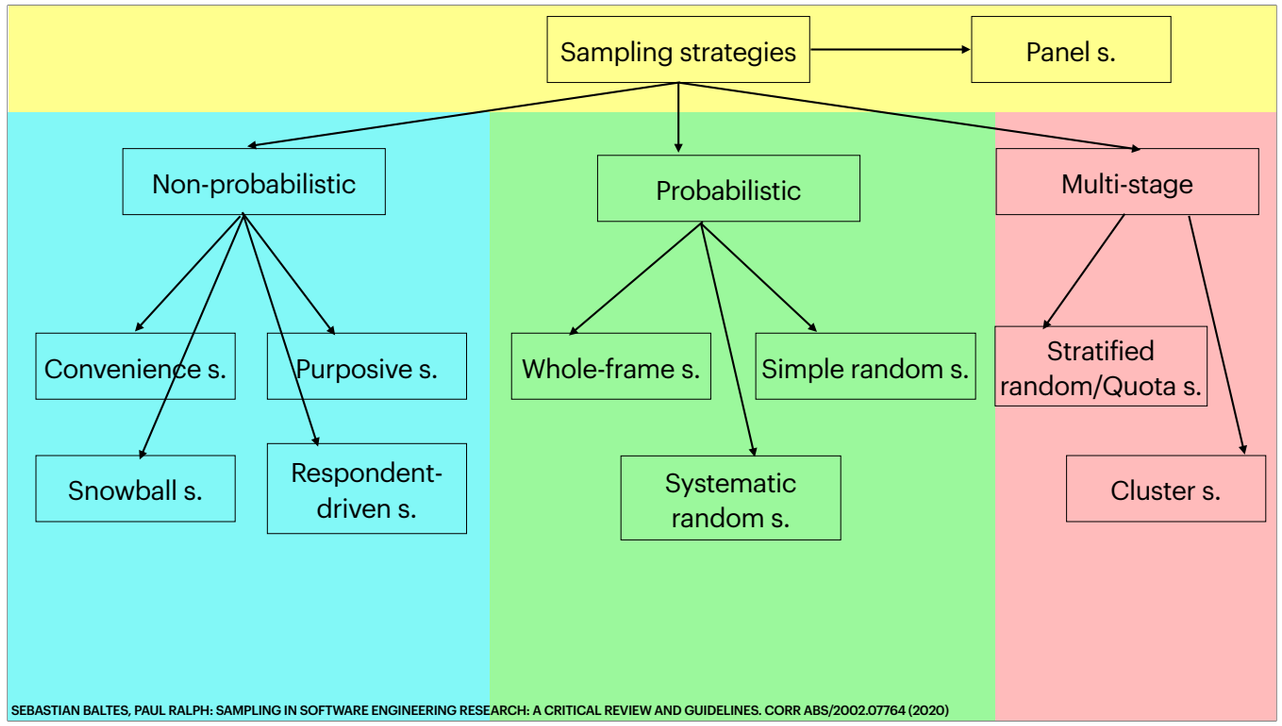
https://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/2000px-Opensource.svg.png https://www.huroncounty.ca/assess/bnames4.jpg https://thepersondoesnotexist.com/

However in software engineering it is difficult… We do not have a complete list of all open source projects



Even in GitHub this might not be easy: people might have multiple aliases; sometimes a project is stored in multiple repositories and sometimes a repository contains multiple projects

SAMPLING STRATEGIES



Non-probabilistic sampling includes all of the sampling techniques that do not employ randomness.
Probabilistic sampling includes all of the sampling techniques that employ randomness.
Methodologists often present multistage sampling as a special case where two or more sampling strategies are intentionally combined.

Convenience sampling. Items are selected based on availability or expedience. When we select people or artifacts to study arbitrarily, or based on them being nearby, available or otherwise easy to study, we adopt convenience sampling. For example, asking our friends to fill in the survey or advertising it on Twitter leads to convenience sampling.

"We recruited 27 CS students (23 males, four females) from the Department of Computer Science of our University of which 21 undergraduates, five graduates, and one post-graduate."

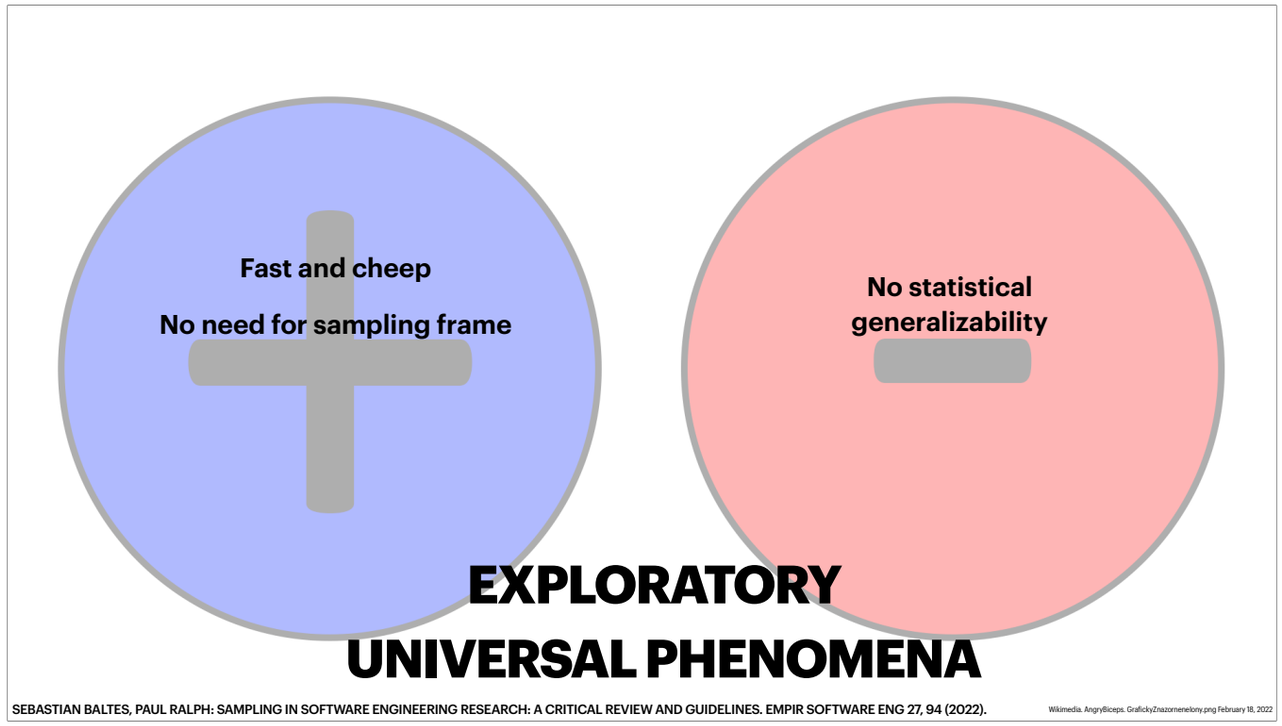
Empir Software Eng (2016) 21:449–482
DOI 10.1007/s10664-015-9360-1

Preprocessor-based variability in open-source and industrial software systems: An empirical study

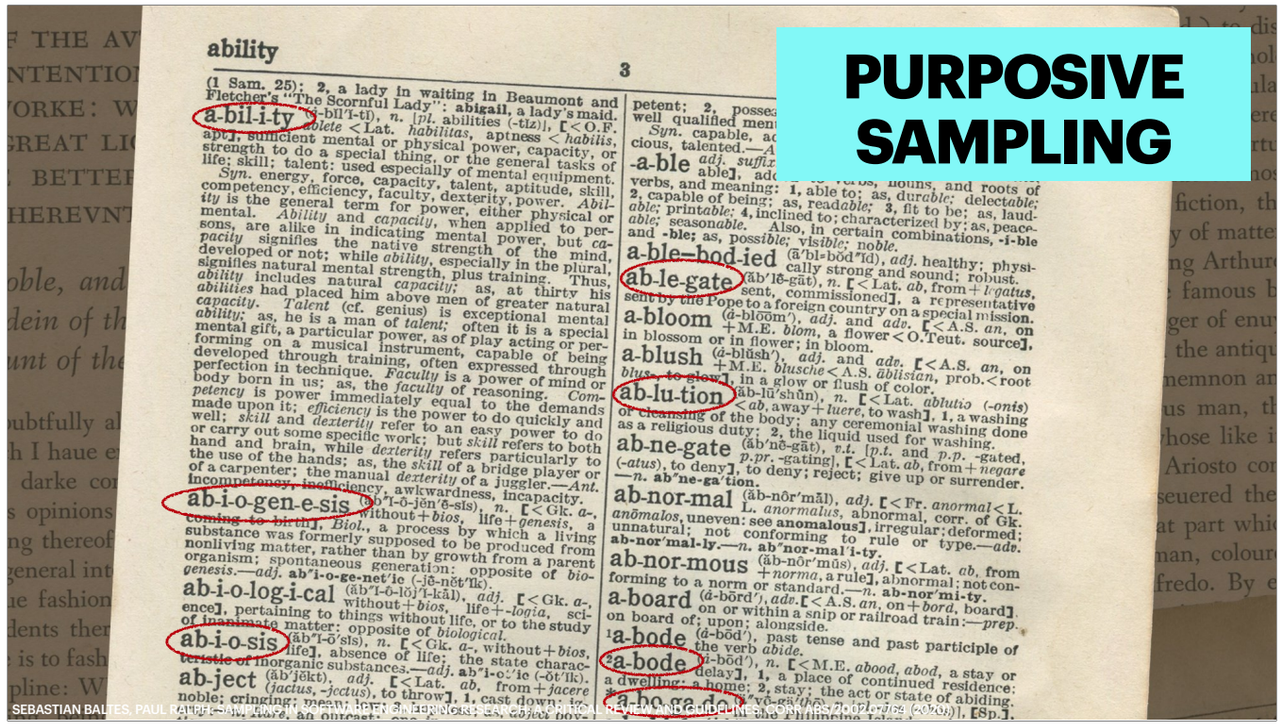
Claus Hunsen · Bo Zhang · Janet Siegmund · Christian Kästner · Olaf Leßenich · Martin Becker · Sven Apel

"For industrial systems <...>, we had less choice and had to rely on convenience sampling, which biased the selection of industrial systems toward the embedded-systems domain."

DANIELA GIRARDI, NICOLE NOVIELLI, DAVIDE FUCCI, FILIPPO LANUBILE: RECOGNIZING DEVELOPERS' EMOTIONS WHILE PROGRAMMING. ICSE 2020

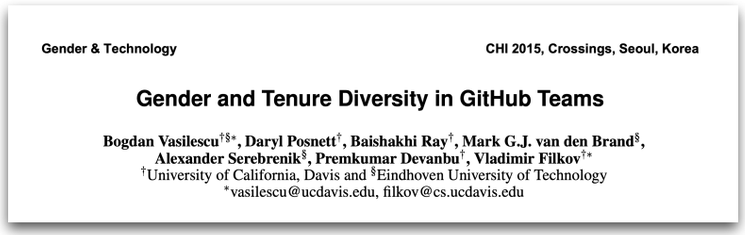


Convenience sampling is controversial because it is very popular, especially in laboratory experiments, despite threatening generalizability [cf. 1, 29]. The key advantages of convenience sampling are: (1) speed, (2) low cost and (3) no need for a sampling frame. These advantages make convenience sampling ideal for pilot studies and studying universal phenomena such as cognitive biases [48].



Purposive sampling. Items are selected based on specific characteristics, other than availability, that make them relevant to the study's objective. When we select our objects of study according to some logic or strategy, carefully but not randomly, we employ purposive or purposeful sampling. While non-probability samples can be representative of broader populations, the goal of non-probability sampling is often to find accessible, information rich cases, sites, organizations or contexts from which researchers can learn about their topic of study.

For example, we are interested in studying nouns.



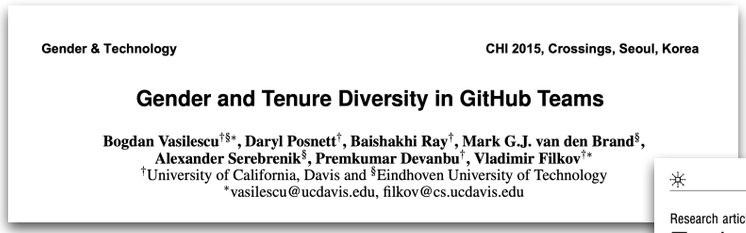
Gender & Technology CHI 2015, Crossings, Seoul, Korea

Gender and Tenure Diversity in GitHub Teams

Bogdan Vasilescu[†§*], Daryl Posnett[†], Baishakhi Ray[†], Mark G.J. van den Brand[§], Alexander Serebrenik[§], Premkumar Devanbu[†], Vladimir Filkov[†*]
[†]University of California, Davis and [§]Eindhoven University of Technology
[*]vasilescu@ucdavis.edu, filkov@cs.ucdavis.edu

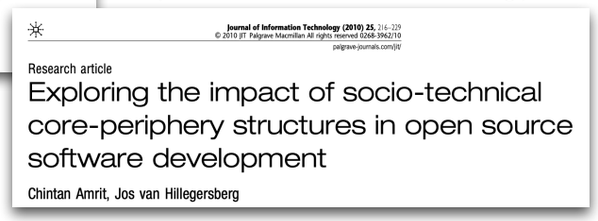
"We chose **GITHUB** for our study because it: (i) is the largest ecosystem of its kind; (ii) has a wealth of trace data available [29]; and (iii) is known to promote women in technology"

Purposive sampling includes a whole range of approaches, for example:
(1) Studying projects hosted on GitHub because it is popular and has good tool support.

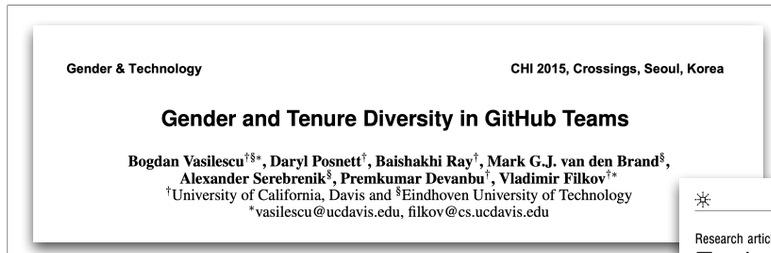


"We use this tool to study a **diverse** collection of Open Source projects."

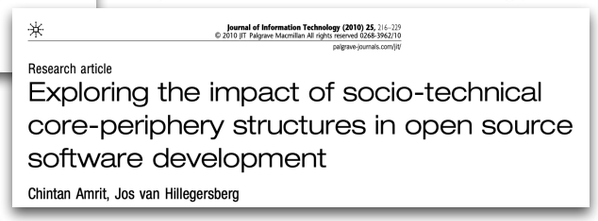
"We chose **GITHUB** for our study because it: (i) is the largest ecosystem of its kind; (ii) has a wealth of trace data available [29]; and (iii) is known to promote women in technology"



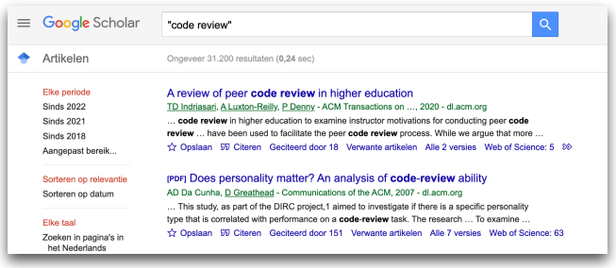
Purposive sampling includes a whole range of approaches, for example:
(1) Studying projects hosted on GitHub because it is popular and has good tool support.
(2) Selecting projects that are as diverse as possible ("heterogeneity sampling").



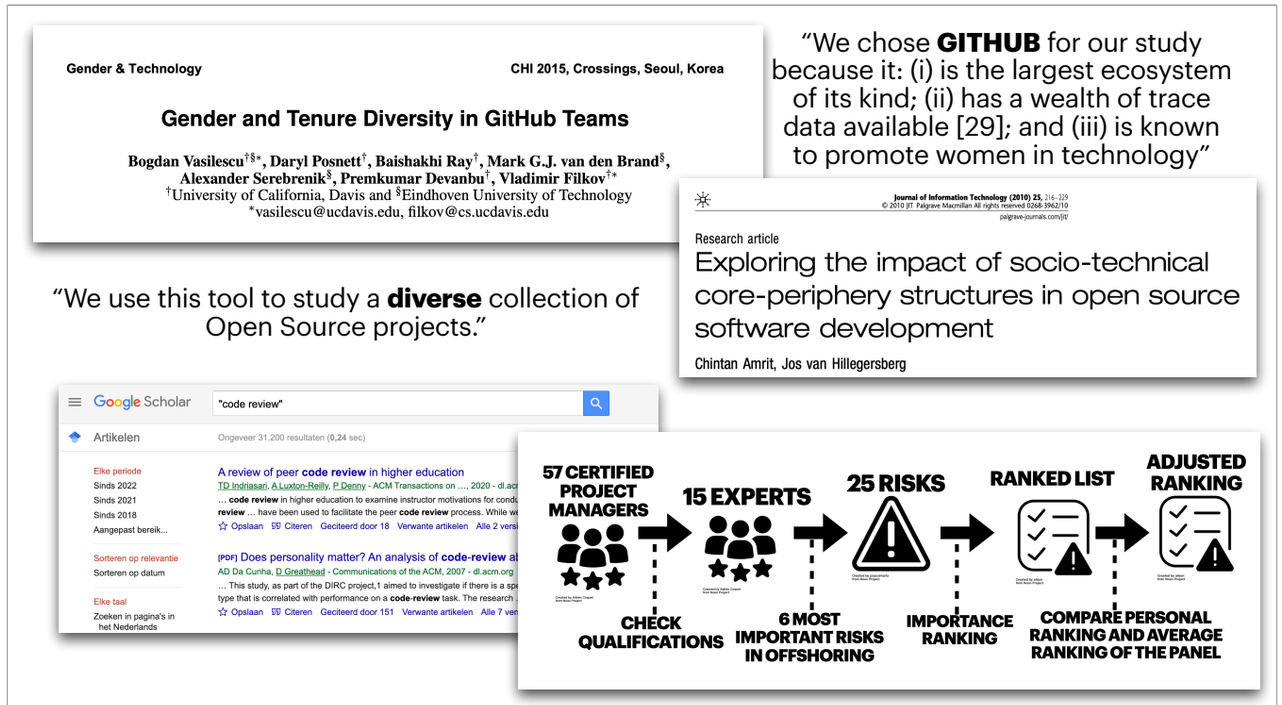
"We chose **GITHUB** for our study because it: (i) is the largest ecosystem of its kind; (ii) has a wealth of trace data available [29]; and (iii) is known to promote women in technology"



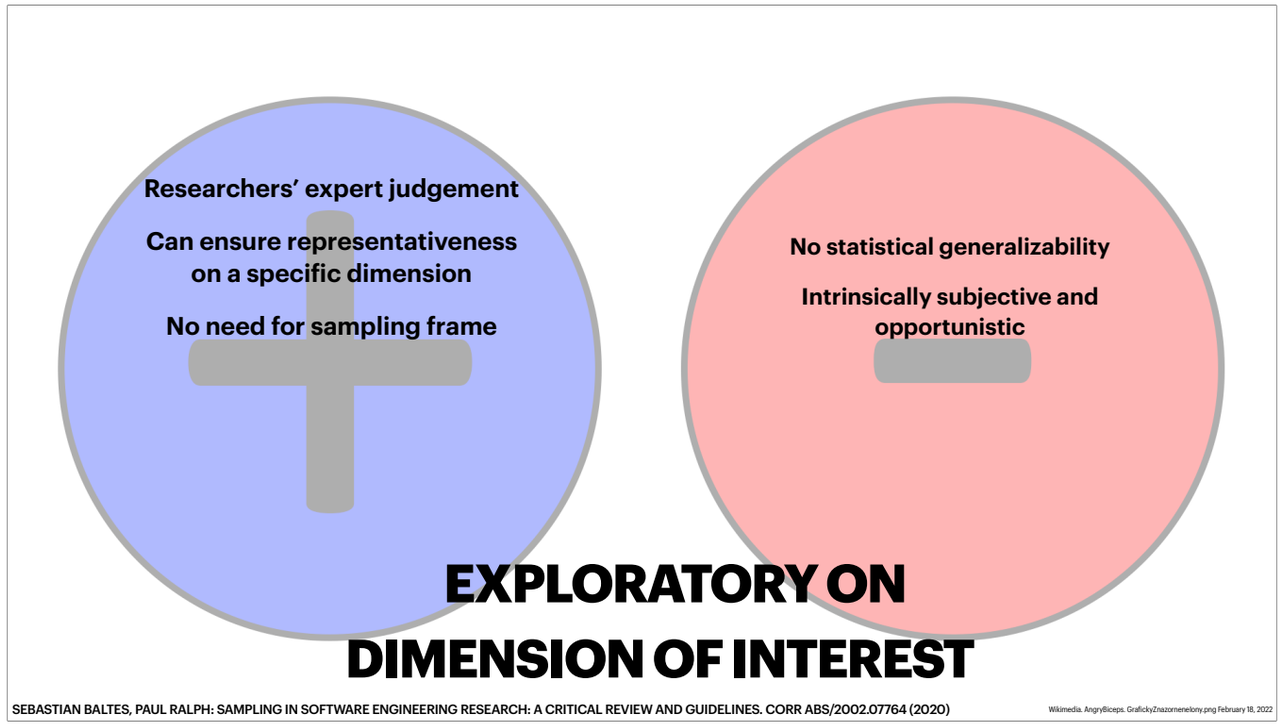
"We use this tool to study a **diverse** collection of Open Source projects."



Purposive sampling includes a whole range of approaches, for example:
(1) Studying projects hosted on GitHub because it is popular and has good tool support.
(2) Selecting projects that are as diverse as possible ("heterogeneity sampling").
(3) Extracting the sample from a repository using a specific search query, as in a systematic literature review ("query-based sampling").



Purposive sampling includes a whole range of approaches, for example:
(1) Studying projects hosted on GitHub because it is popular and has good tool support.
(2) Selecting projects that are as diverse as possible (“heterogeneity sampling”).
(3) Extracting the sample from a repository using a specific search query, as in a systematic literature review (“query-based sampling”).
(4) Recruiting a panel of experts on a particular topic for a focus group (“expert sampling”).



The key advantages of purposive sampling are: (1) the researcher can exercise expert judgment; (2) the researcher can ensure representativeness on a specific dimension (see Section 2.6); (3) no sampling frame is needed. The main challenge with purposive sampling is that it is intrinsically subjective and opportunistic.

QUESTION

Islam and Zibran use the largest (as of February 2016) data-set from Boa, which <...> consists of more than 7.8 million projects collected from GitHub before September 2015. From this large data-set, we select the top 50 projects having the highest number of commits. We study all the commit messages in these projects, which constitute 490,659 commit comments.

(A) CONVENIENCE SAMPLING

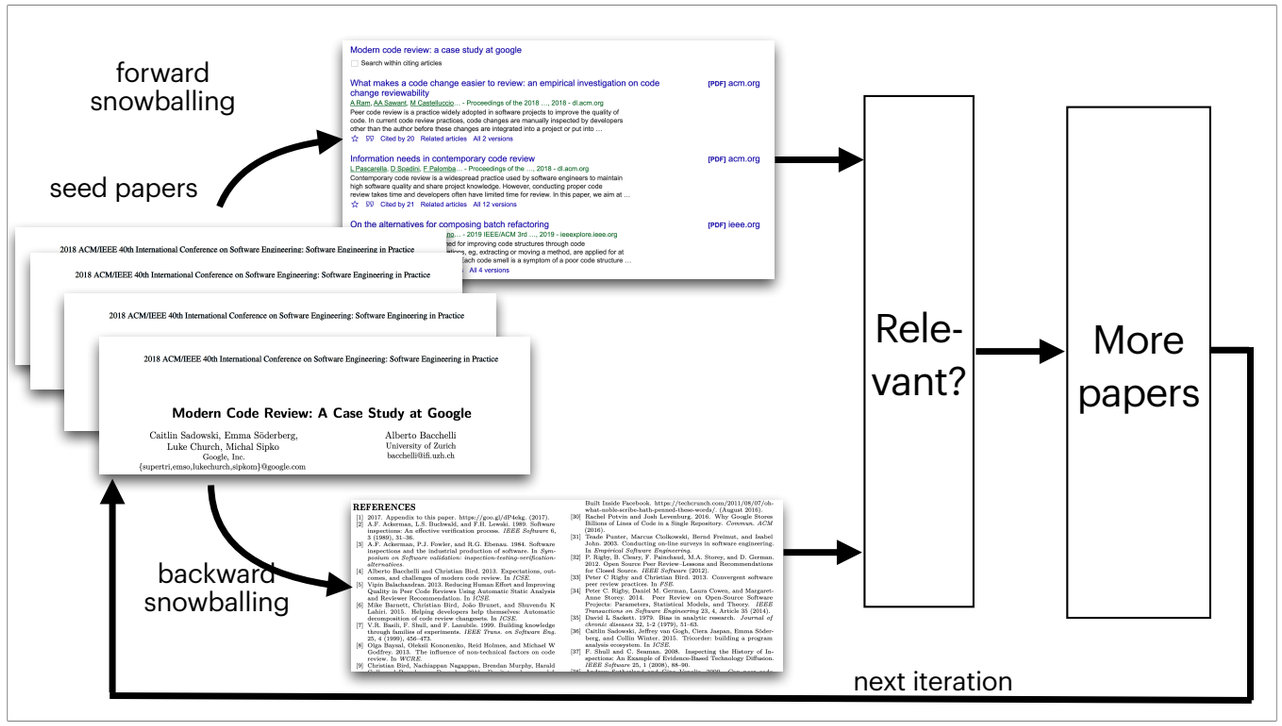
(B) PURPOSIVE SAMPLING

Purposive sampling

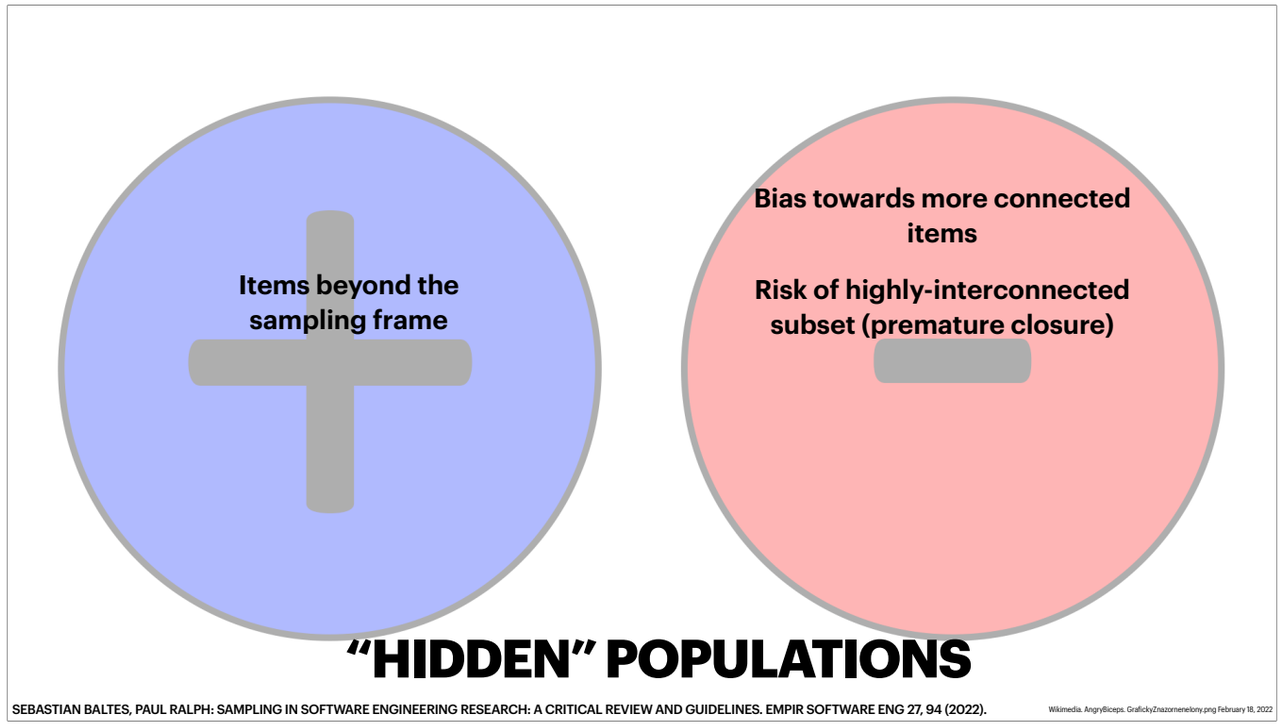
SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)

Referral-chain (snowball) sampling. Items are selected based on their relationship to previously selected items. Referral-chain sampling (also called snowball sampling) is useful when there is no good sampling frame for the population of interest. For example, there is no comprehensive list of black-hat hackers or software developers who have experienced sexual harassment. However, members of such “hidden populations” often know each other.

Snowball sampling with human participants therefore works by finding a few individuals in the population, studying them, and then asking them to refer other members of the population whom they know.



In SE, snowball sampling is commonly used in systematic literature reviews to supplement query-based sampling. When we begin with an article A, searching the papers A cites is sometimes called backward snowballing while searching the papers that cite A is sometimes called forward snowballing. We can study software libraries, methods and services (as in service-oriented architectures), in much the same way.

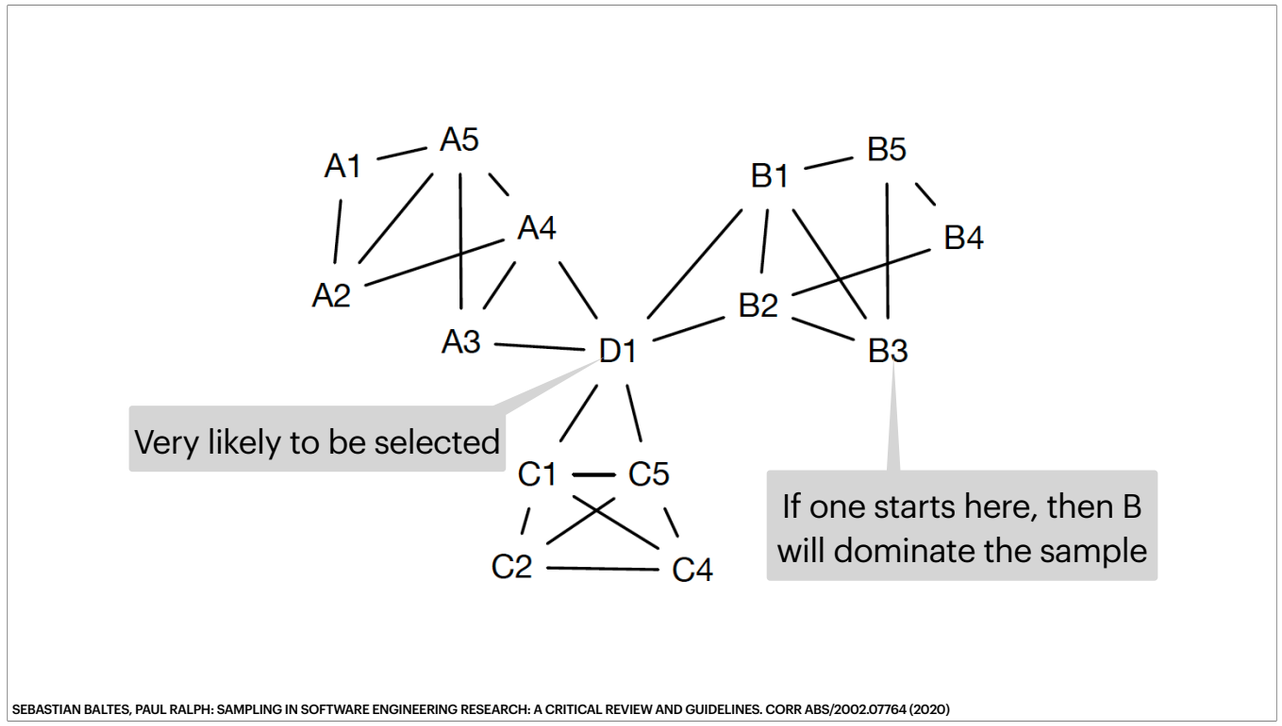


"HIDDEN" POPULATIONS

SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. EMPIR SOFTWARE ENG 27, 94 (2022).

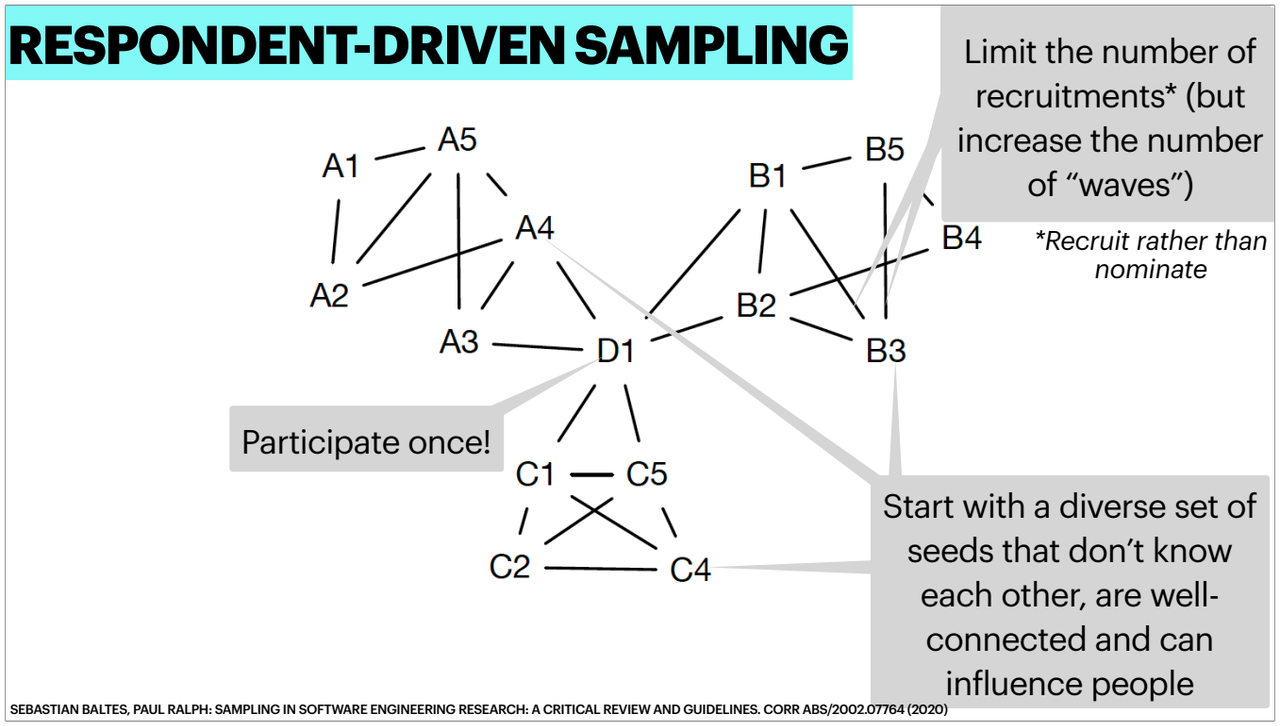
Wikimedia. AngryBiceps. GrafickyZnazornenelony.png February 18, 2022

The advantage of snowball sampling is that it helps us to identify items that are not in our sampling frame. However, snowball sampling has two major limitations: 1) it biases results toward more connected people (or artifacts); 2) it can lead to sampling a small, highly-interconnected subset of a larger population (see next slide)

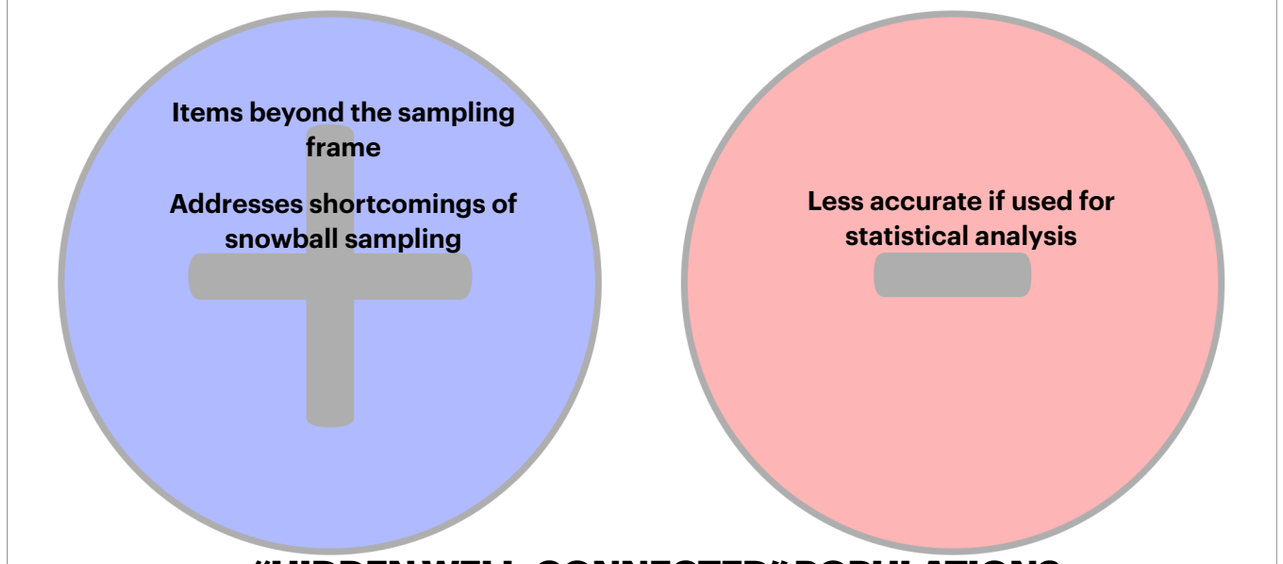


SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)

Snowball sampling limitations: (i) item **D1** is more likely to be selected because it has more connections and has higher network centrality; (ii) beginning with item **B3**, for example, the sample may be dominated by items from the B-cluster.



(1) Begin with **diverse initial participants** (seeds) who (i) have large social networks, (ii) represent different sub-populations, (iii) do not know each other, and (iv) can influence peers to participate.
(2) Have participants **recruit, rather than identify,** peers. This reduces selection bias by the researcher.
(3) **Limit recruitment** such that each participant can only recruit a small number of peers (typically three). This prevents highly-connected participants from biasing the sample.
(4) Require **many (e.g. 20) recruitment waves**. This generates longer referral chains, decreasing the risk of oversampling from a highly connected subset of the population.
(5) Prevent individuals from participating **more than once.**
(6) Continue recruitment until the sample reaches **equilibrium**, the point where the distribution of variables of interest is stable.
(7) Apply a mathematical model to account for sampling bias [28, 34].



“HIDDEN WELL-CONNECTED” POPULATIONS

SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. EMPIR SOFTWARE ENG **27**, 94 (2022).

Wikimedia. AngryBiceps. GrafickyZnazornenelony.png February 18, 2022

QUESTION

"we invited industrial developers to complete an on-line questionnaire about BMs (*Brain Methods*) in the application JGroups. We choose JGroups because it is one of the applications with the largest number of BMs and also because it is well documented. The invitations were sent via e-mail and Java development groups of Linkedin."

(A) CONVENIENCE SAMPLING

(C) SNOWBALL SAMPLING

(B) PURPOSIVE SAMPLING

(D) RESPONDENT-DRIVEN SAMPLING

SANTIAGO A. VIDAL, IÑAKI BERRA, SANTIAGO ZULLIANI, CLAUDIA A. MARCOS, J. ANDRÉS DÍAZ PACE: ASSESSING THE REFACTORING OF BRAIN METHODS. ACM TRANS. SOFTW. ENG. METHODOL. 27(1): 2:1-2:43 (2018)

(B) Purposive sampling

QUESTION

“we interviewed 16 participants (5 women, 11 men) from eight different organizations at Microsoft. <...>

First, we identified presenters at data-driven engineering meetups and technical community meetings, since these have been responsible internally for sharing best practices.

Next, we selected additional data scientists by word-of-mouth, asking each participant to introduce us to other data scientists or other key stakeholders whom they knew. ”

(A) CONVENIENCE SAMPLING

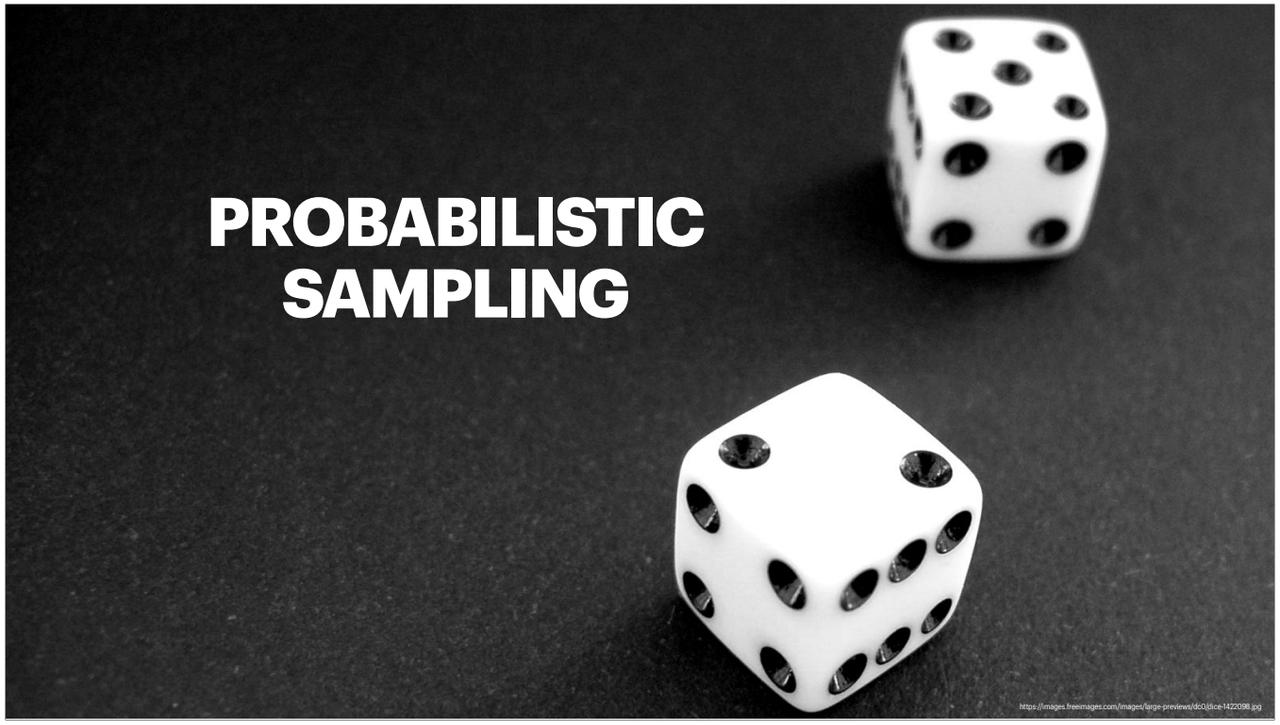
(C) SNOWBALL SAMPLING

(B) PURPOSIVE SAMPLING

(D) RESPONDENT-DRIVEN SAMPLING

MIRYUNG KIM, THOMAS ZIMMERMANN, ROBERT DELINE, ANDREW BEGEL: THE EMERGING ROLE OF DATA SCIENTISTS ON SOFTWARE DEVELOPMENT TEAMS. ICSE 2016: 96-107

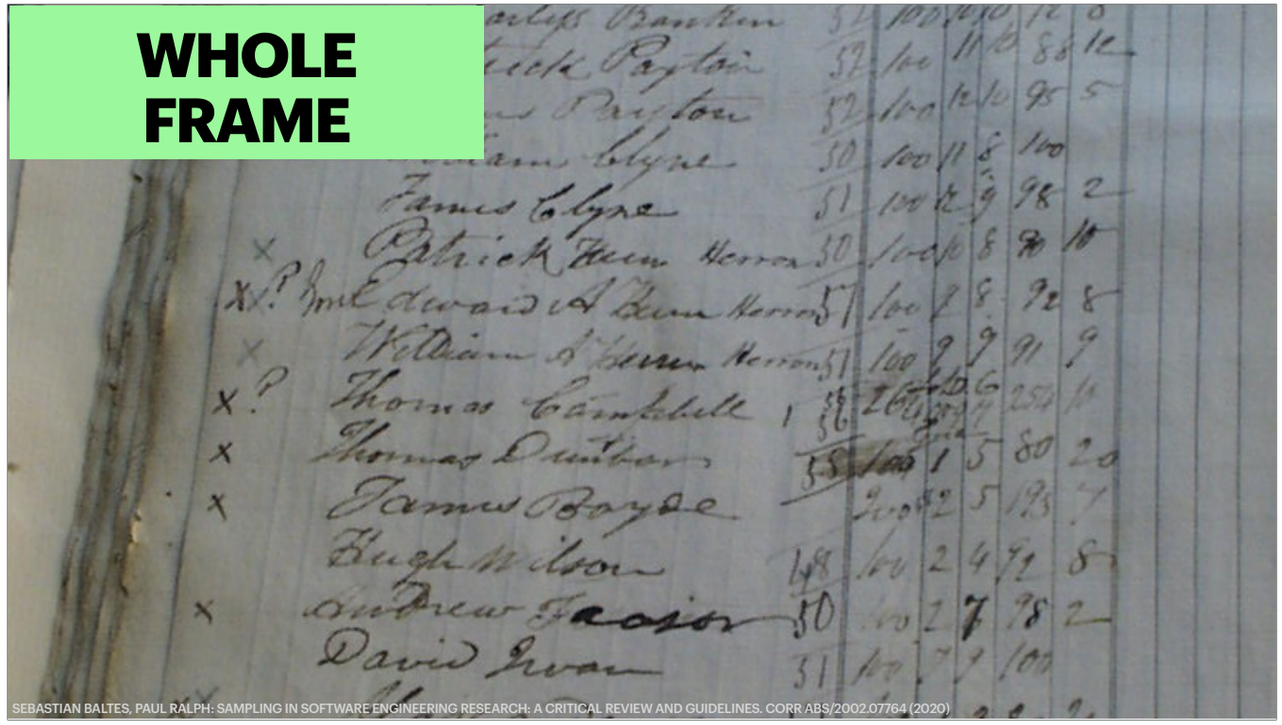
(B) Snowball sampling (word-of-mouth).
No indication that special care has been taken to go for a respondent-driven sampling



Since examining the entire population is usually impractical, the researcher selects a subset of the population (a sample) and attempts to estimate a property of the population by statistically analyzing the sample. Probability sampling ostensibly facilitates such statistical generalization.

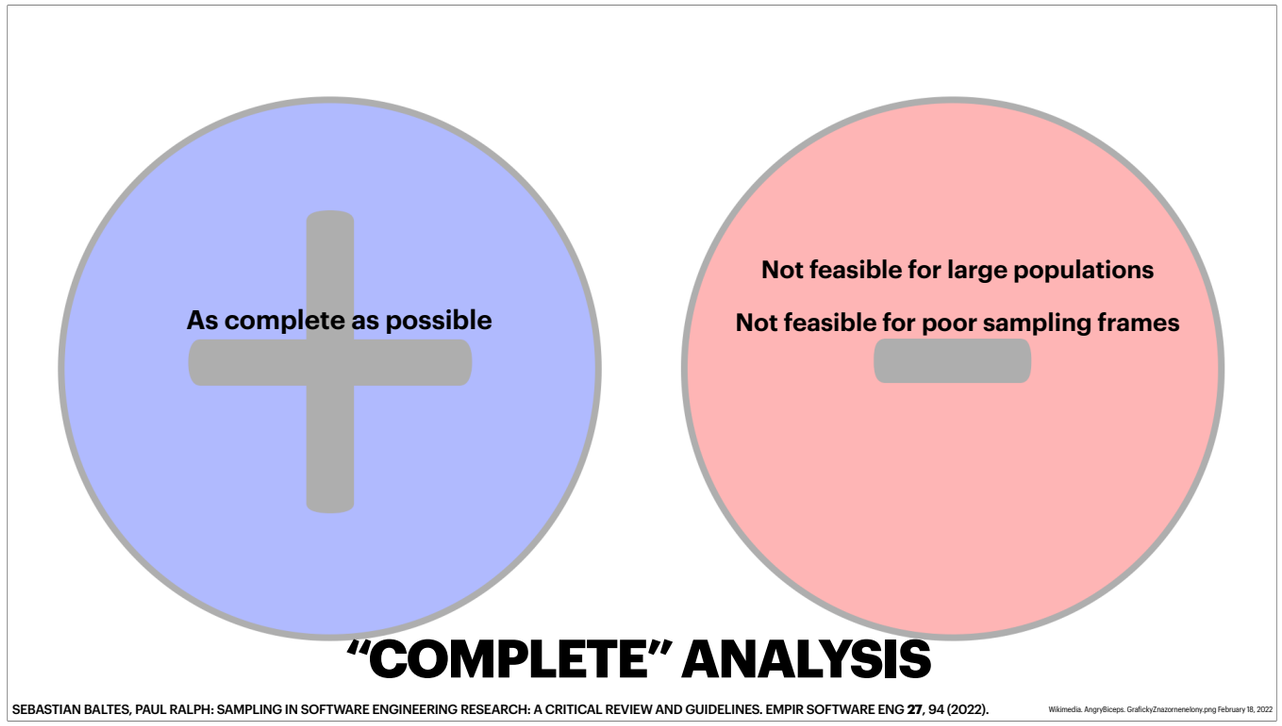
Standing on a street corner interviewing “random” pedestrians is not random in the statistical sense: you miss cyclists, car drivers and people using pubic transport. Time of the day, day of the week and period of the year influence who is outside; location, of course, influences as well. Recruiting participants using email or advertising on social networks is not random: due to some people not having access to these particular social networks (Facebook in China) or self-selection. Assigning participants to experimental conditions in the order in which they arrive at a laboratory is not random. Practically speaking, any selection without using a random number generator, probably is not random.

The overwhelming challenge for applying any kind of probability sampling in SE is the absence of comprehensive sampling frames for common units of analysis/ observation



SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)

All items in the sampling frame are selected. Suppose a researcher wants to assesses morale at a specific software development company. The company provides a complete list of developers and their contact information. The researcher creates a survey with questions about job satisfaction, views of the company, employees' future plans, etc. They send the questionnaire to all of the developers—the entire sampling frame.
A similar analysis would consider all commits of a GitHub repository, or all issues in the issue tracker of a project or all questions on Stack Overflow.



Whether this is technically “sampling” is debatable, but it is an important option to consider, especially when data collection and analysis are largely automated. “Poor” means incomplete (lots of items are missing and we do not know what is missing) or inadequate (we know what is missing but it is clear that the sampling frame is different from the population, e.g., all GitHub repositories (population) vs all public GitHub repositories (sampling frame)).

Simple Random Sampling. Items are selected entirely by chance, such that each item has equal chance of inclusion. Now suppose the results of the above morale survey are less than spectacular. The researcher decides to follow up with some in-depth interviews. However, interviewing all 10,000 developers is clearly impractical, so the researcher assigns each developer a number between 1 and 10,000, uses a random number generator to select 20 numbers in the same range, and interviews those 20 developers. This is simple random sampling because the researcher simply chooses n random elements from the population.

A Large-Scale Evaluation of Automated Unit Test Generation Using EvoSuite

GORDON FRASER, University of Sheffield
ANDREA ARCURI, Simula Research Laboratory

"We randomly selected 100 Java projects from SourceForge, a well-established open-source repository"

A Large-Scale Evaluation of Automated Unit Test Generation Using EvoSuite

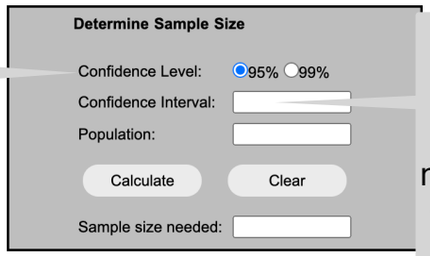
GORDON FRASER, University of Sheffield
ANDREA ARCURI, Simula Research Laboratory

"We randomly selected 100 Java projects from SourceForge, a well-established open-source repository"

IS 100 ENOUGH?

https://www.surveysystem.com/sscalc.htm

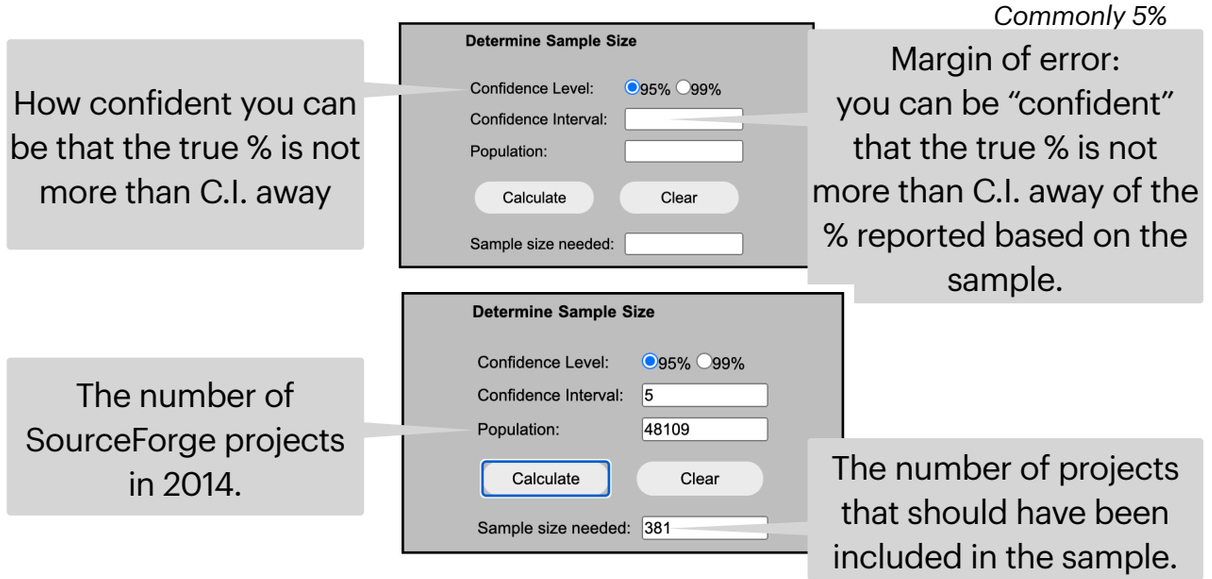
How confident you can be that the true % is not more than C.I. away



Commonly 5%

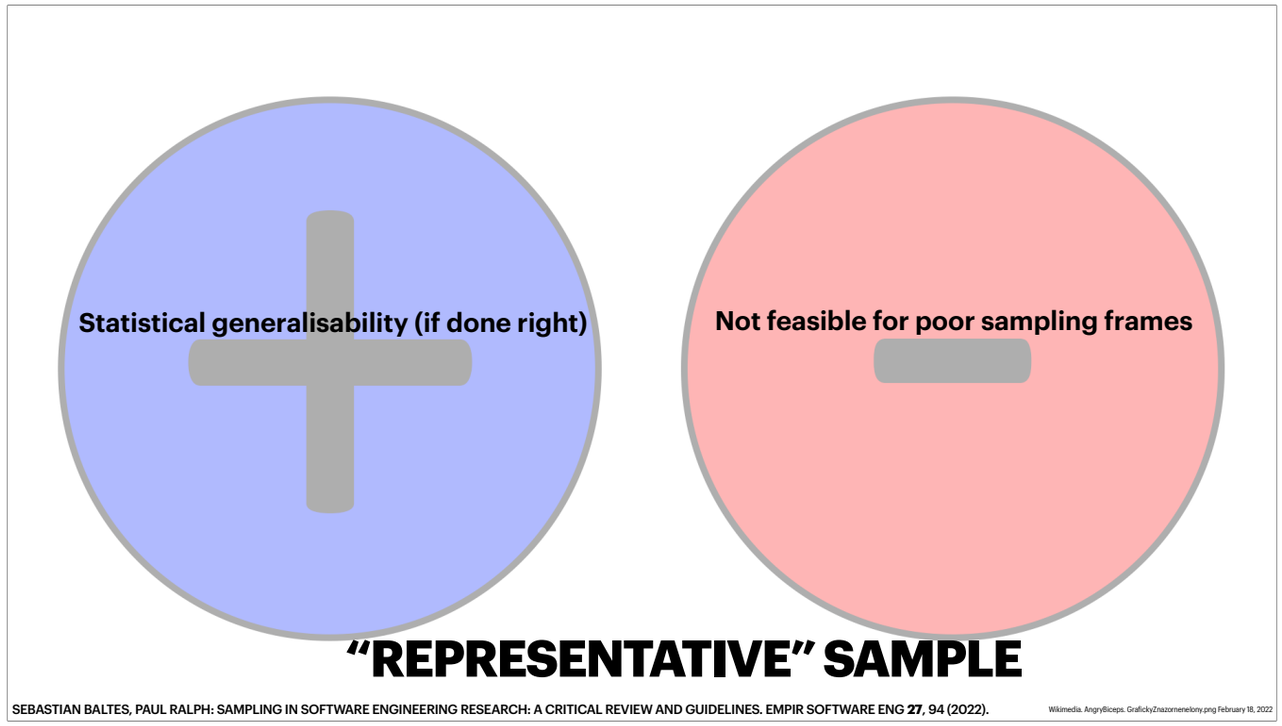
Margin of error:
you can be “confident” that the true % is not more than C.I. away of the % reported based on the sample.

https://www.surveysystem.com/sscalc.htm

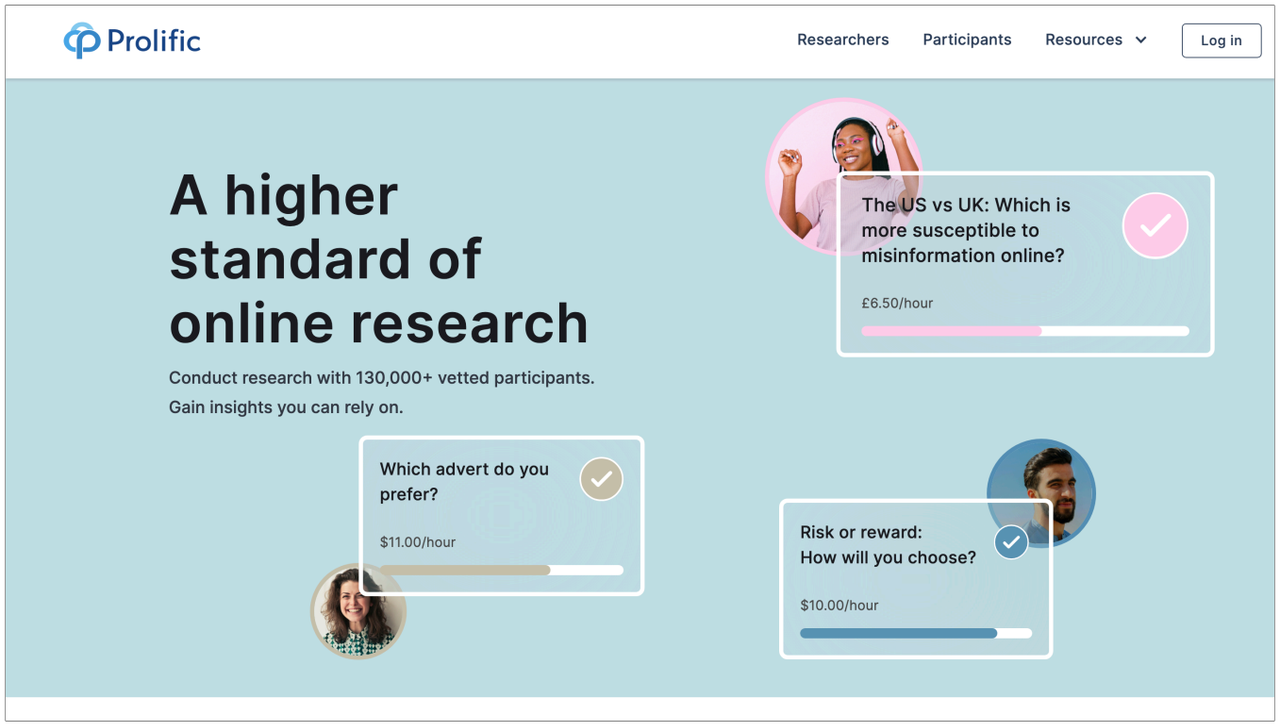


NO, 100 IS NOT ENOUGH

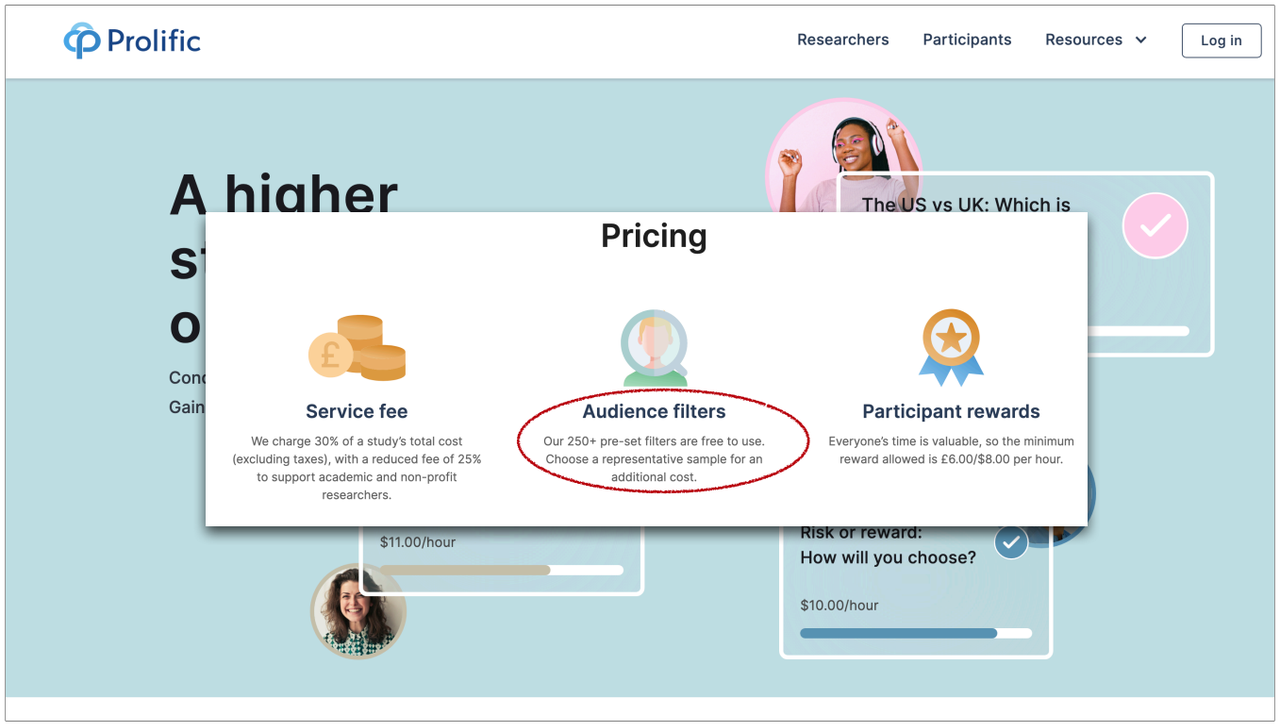
We will discuss how to determine this number in one of the following lectures



Representativeness is a difficult matter. Here, representativeness is the **degree to which a sample’s properties (of interest) resemble those of a target population**. A sample can be representative with respect to one dimension and not with respect to another. In SE people tend to assume that representativeness with respect to an easily observable property (e.g., LOC) implies representativeness with respect to a much more difficultly observable property (e.g., complexity of discussion). Moreover, probabilistic sampling does not guarantee representativeness and a non-probabilistic sample can be more representative with respect to the property of interest.

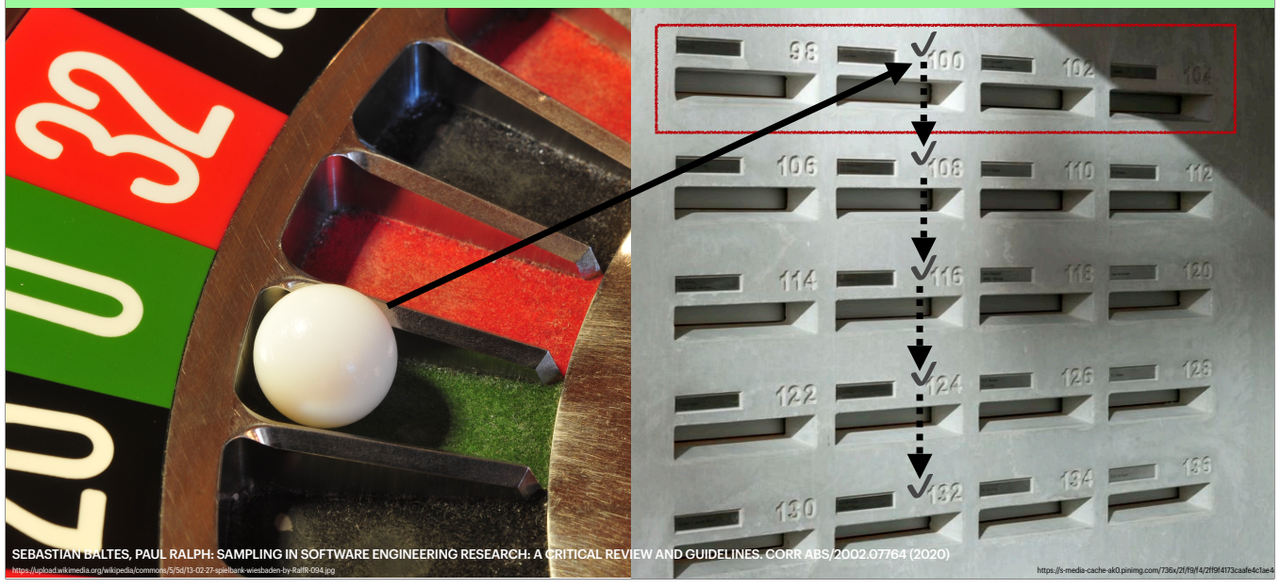


This is Prolific, a popular platform for recruiting participants.



Representativeness is problematic here: at the very least Prolific can deliver a sample that is representative for their sampling frame but we do not know whether this sampling frame is a trustworthy representation of our population. Carianne's story about a TikTok video, "go and make money".

SYSTEMATIC RANDOM SAMPLING



SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)

https://upload.wikimedia.org/wikipedia/commons/5/5d/13-02-27-spielbank-wiesbaden-by-RalfR-094.jpg

https://s-media-cache-ak0.pinimg.com/736x/2f/f9/f4/2ff9f4173caafe4c1ae4

Given an interval, x, every xth item is selected, from a starting point chosen entirely by chance. To complement the interviews, the researcher decides to review developers' posts on the company's messaging system (e.g. Slack). Suppose there is no easy way to jump to a random message and there are too many messages to read them all. So the researcher generates a random number between 1 and 100 (say, 47) and then reads message 47, 147, 247, etc. until reaching the end of the messages.

Each post still has an equal probability of inclusion; however, the consistent interval could bias the sample if there is a recurring pattern that coincides with the interval (e.g. taking annual weather data in the middle of summer vs. the middle of winter).

Exploring Regular Expression Usage and Context in Python

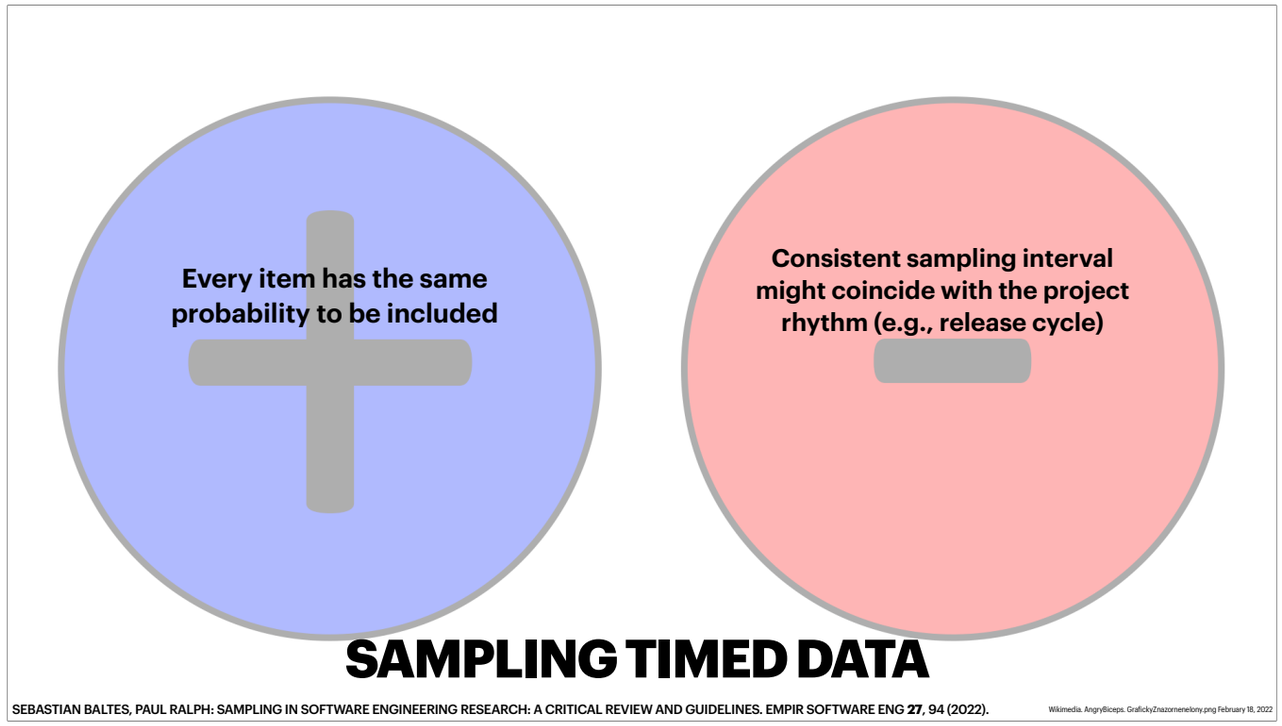
Carl Chapman
Department of Computer Science
Iowa State University
Ames, IA, USA
carl1978@iastate.edu

Kathryn T. Stolee
Departments of Computer Science
North Carolina State University
Raleigh, NC, USA
ktstolee@ncsu.edu

"We also did not scrape all commits in every project for regular expression utilizations, rather, we grabbed each project **every 20 commits**."

FEASIBILITY OF THE ANALYSIS ⟷ **LOSS OF DATA**

SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)



Each post still has an equal probability of inclusion; however, the consistent interval could bias the sample if there is a recurring pattern that coincides with the interval (e.g. taking annual weather data in the middle of summer vs. the middle of winter).

QUESTION

"The exception hierarchy of the Android APIs. Uncaught exceptions and statements throwing exceptions are a major source of faults in Android apps. We automatically crawled the official Android developer JavaDoc guide to extract the exception hierarchy and API methods throwing exceptions. We collected 5,414 items from which we sampled 360 of them for manual analysis."

(A) SIMPLE RANDOM SAMPLING

(B) SYSTEMATIC RANDOM SAMPLING

(C) SNOWBALL SAMPLING

(D) NONE OF THE ABOVE

MARIO LINARES VÁSQUEZ, GABRIELE BAVOTA, MICHELE TUFANO, KEVIN MORAN, MASSIMILIANO DI PENTA, CHRISTOPHER VENDOME, CARLOS BERNAL-CÁRDENAS, DENYS POSHYVANYK: ENABLING MUTATION TESTING FOR ANDROID APPS. ESEC/SIGSOFT FSE 2017: 233-244

(D) none of the above

MULTI-STAGE SAMPLING

SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)

Multi-stage sampling is a special case where two or more sampling strategies are intentionally combined

Stratified/Quota sampling. The sampling frame is divided into sub-frames with proportional representation. Suppose that the developer morale survey discussed above reveals significant differences between developers who identify as white and those who do not. However, further suppose that 90% of the developers are white.

To get more insight into these differences, the researcher might divide developers into two strata—white and non-white—and select 10 developers from each strata. If the developers are selected randomly, this is called stratified random sampling. If the developers are selected purposively, it is called quota sampling.

We conceptualize these strategies as multistage because the researcher purposively chooses the strata (stage 1) before selecting the people or artifacts to study (stage 2).

Stratified/Quota sampling. The sampling frame is divided into sub-frames with proportional representation. Suppose that the developer morale survey discussed above reveals significant differences between developers who identify as white and those who do not. However, further suppose that 90% of the developers are white.

To get more insight into these differences, the researcher might divide developers into two strata—white and non-white—and select 10 developers from each strata. If the developers are selected randomly, this is called stratified random sampling. If the developers are selected purposively, it is called quota sampling.

We conceptualize these strategies as multistage because the researcher purposively chooses the strata (stage 1) before selecting the people or artifacts to study (stage 2).

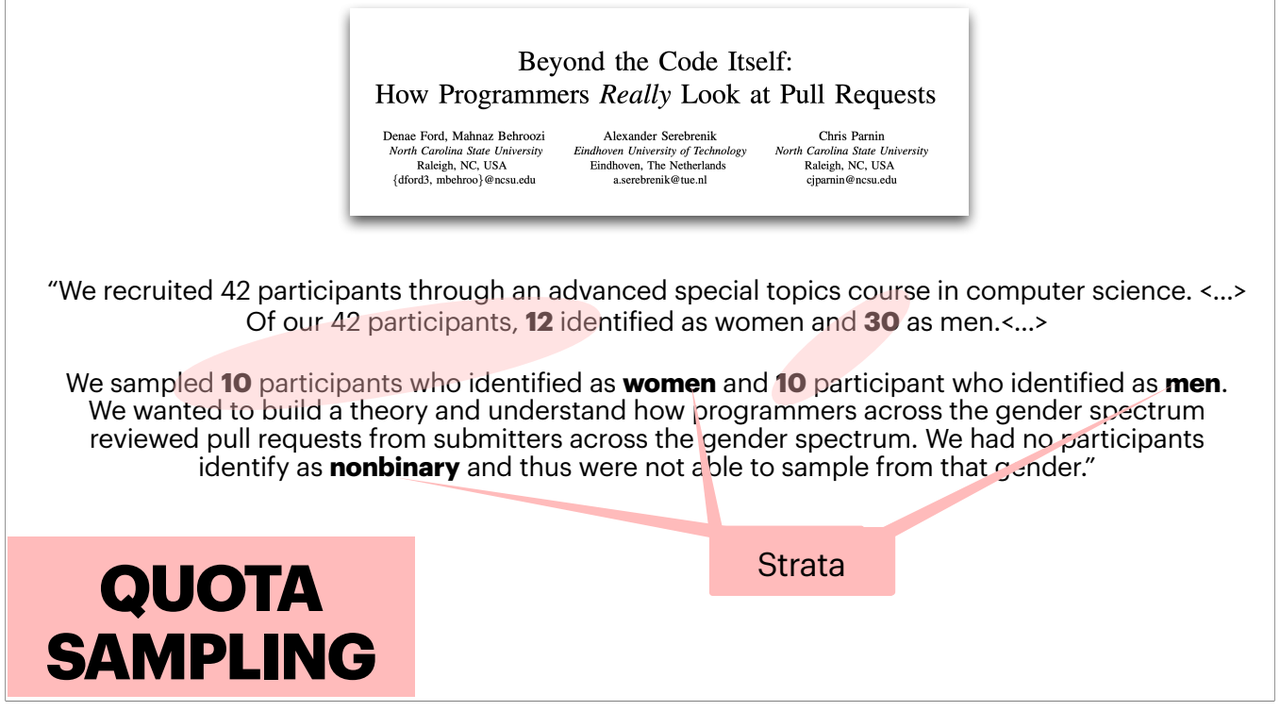
REPENT: Analyzing the Nature of Identifier Renamings

Venera Arnaoudova, Laleh M. Eshkevari, Massimiliano Di Penta, *Member, IEEE*, Rocco Oliveto, *Member, IEEE*, Giuliano Antoniol, and Yann-Gaël Guéhéneuc, *Senior Member, IEEE*

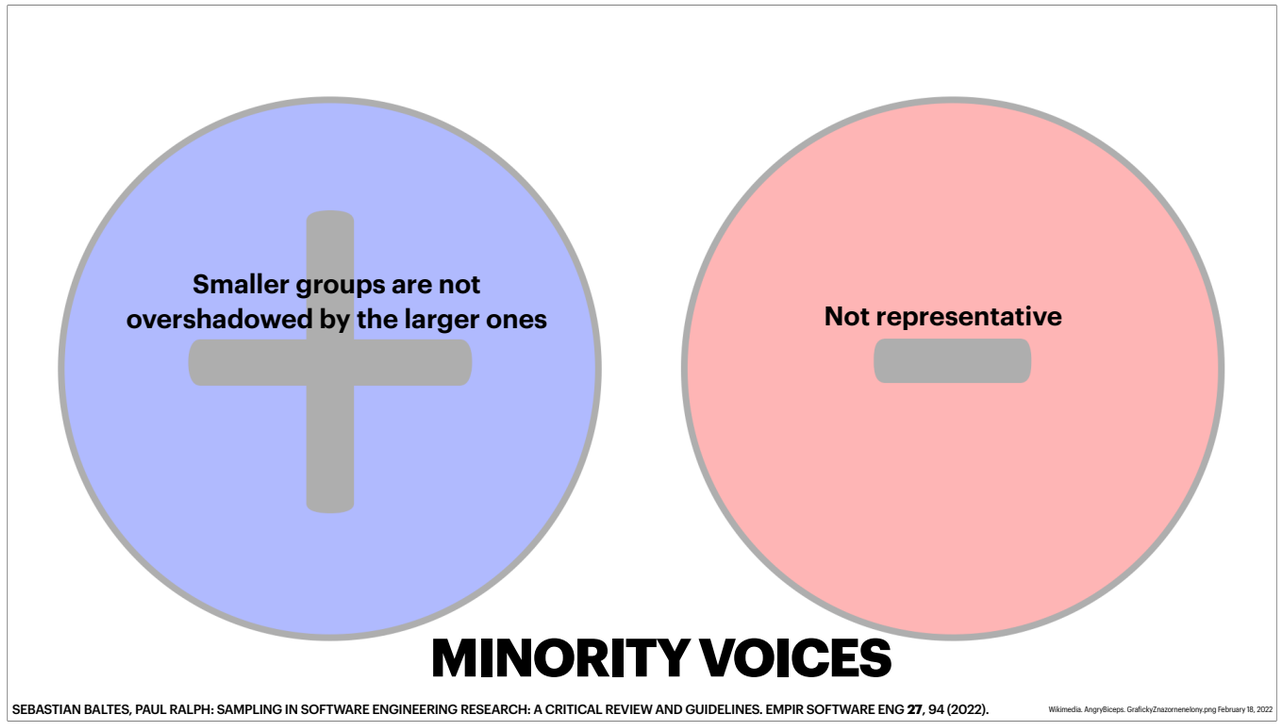
"for each program we first group renamings based on the **kind of entity** being renamed <...> (e.g., type, method). Then, we estimate the proportion of each group with respect to the total population of detected renamings for that particular program and we use the same proportion for the sample. <...> Finally, we **randomly** select the sample for each group."

STRATIFIED RANDOM SAMPLING

If the data points are selected randomly, this is called stratified random sampling.



If the data points are selected purposively, it is called quota sampling. This example is not the best one but the best one I could find: the problem here is that the authors do not indicate how 10 women among 12, and 10 men among 30 have been selected. However since I am one of the authors of this paper, I happen to remember that the selection tried to diversify their background and minority status.

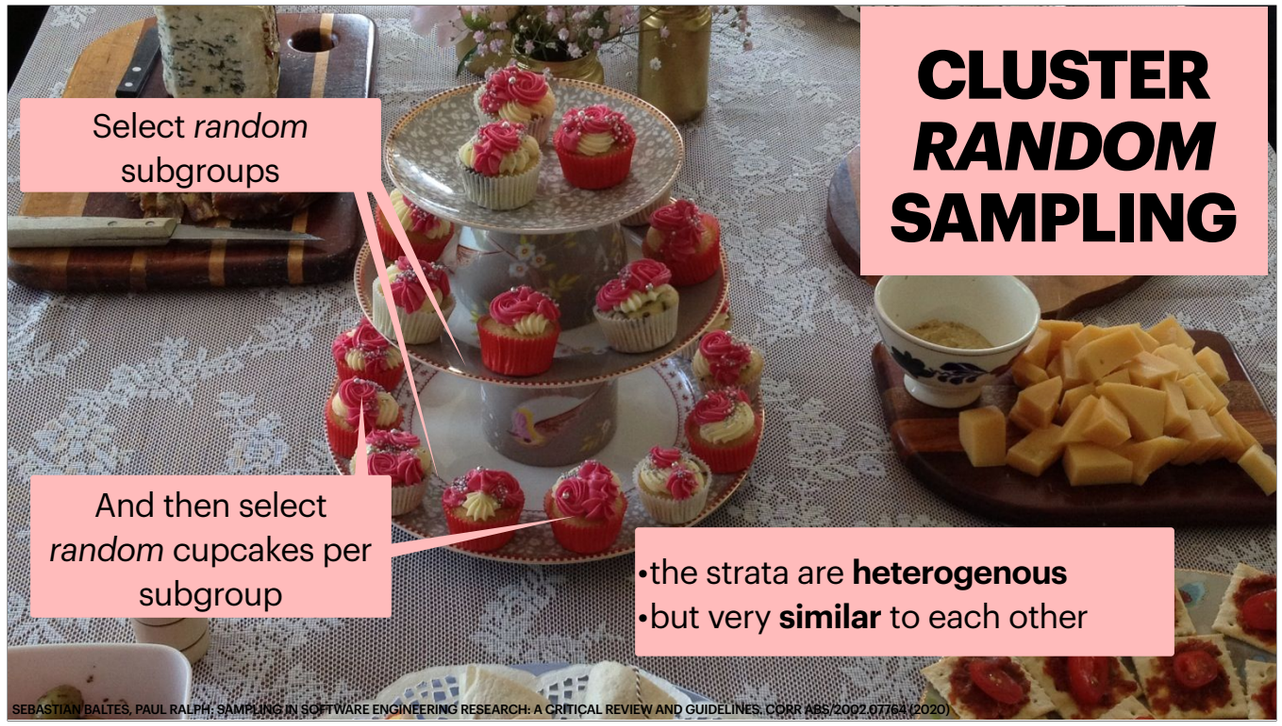


This sampling strategy is interesting because it is intentionally non-representative [72].

Cluster sampling. The sampling frame is divided into groups and items are drawn from a subset of groups. Suppose that the company from our morale survey example has 20 offices spread around the world. If the researcher wants to conduct face-to-face interviews, traveling to all 20 offices could be prohibitively expensive. Instead, the researcher selects three offices (stage 1) and then selects 7 participants in each of these offices (stage 2). This is called cluster sampling.

If and only if both selections are random, it is cluster random sampling. Cluster sampling works best when the groups (clusters) are similar to each other but internally diverse on the dimensions of interest.

Suppose that the researcher finds that the seven developers at one office seem much happier than developers in the rest of the company. If the researcher decides to conduct extra interviews at that office, in hopes of unraveling the sources of improved morale, this is called adaptive cluster sampling [69, 73].



Cluster sampling. The sampling frame is divided into groups and items are drawn from a subset of groups. Suppose that the company from our morale survey example has 20 offices spread around the world. If the researcher wants to conduct face-to-face interviews, traveling to all 20 offices could be prohibitively expensive. Instead, the researcher selects three offices (stage 1) and then selects 7 participants in each of these offices (stage 2). This is called cluster sampling.

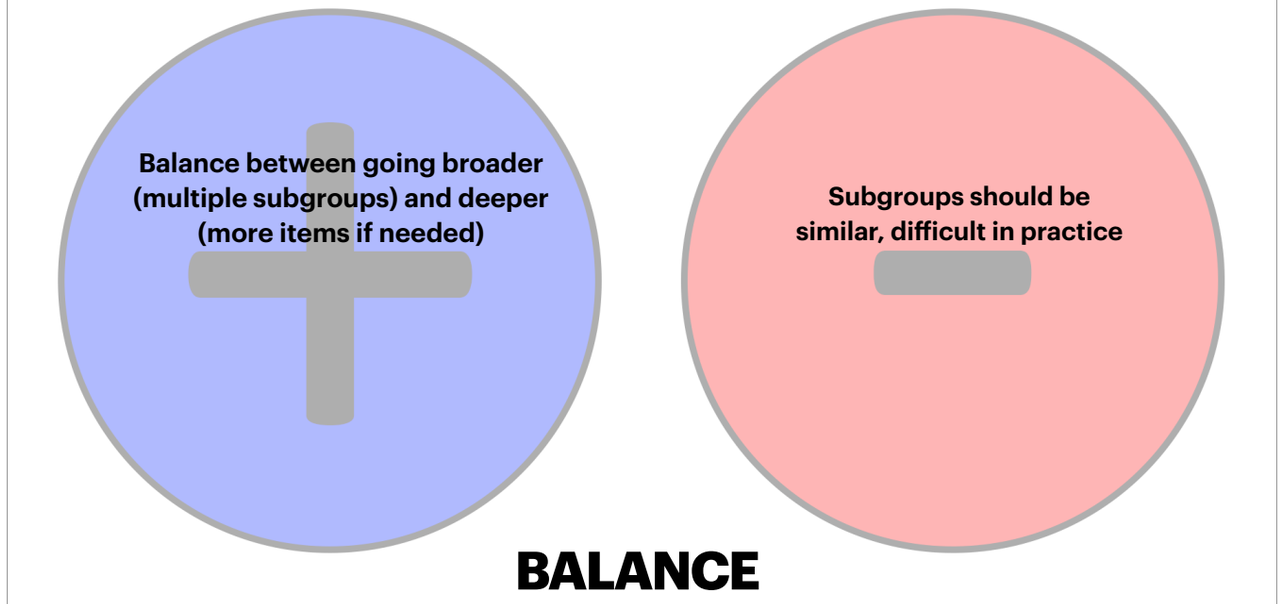
If and only if both selections are random, it is cluster random sampling. Cluster sampling works best when the groups (clusters) are similar to each other but internally diverse on the dimensions of interest.

Suppose that the researcher finds that the seven developers at one office seem much happier than developers in the rest of the company. If the researcher decides to conduct extra interviews at that office, in hopes of unraveling the sources of improved morale, this is called adaptive cluster sampling [69, 73].

Cluster sampling. The sampling frame is divided into groups and items are drawn from a subset of groups. Suppose that the company from our morale survey example has 20 offices spread around the world. If the researcher wants to conduct face-to-face interviews, traveling to all 20 offices could be prohibitively expensive. Instead, the researcher selects three offices (stage 1) and then selects 7 participants in each of these offices (stage 2). This is called cluster sampling.

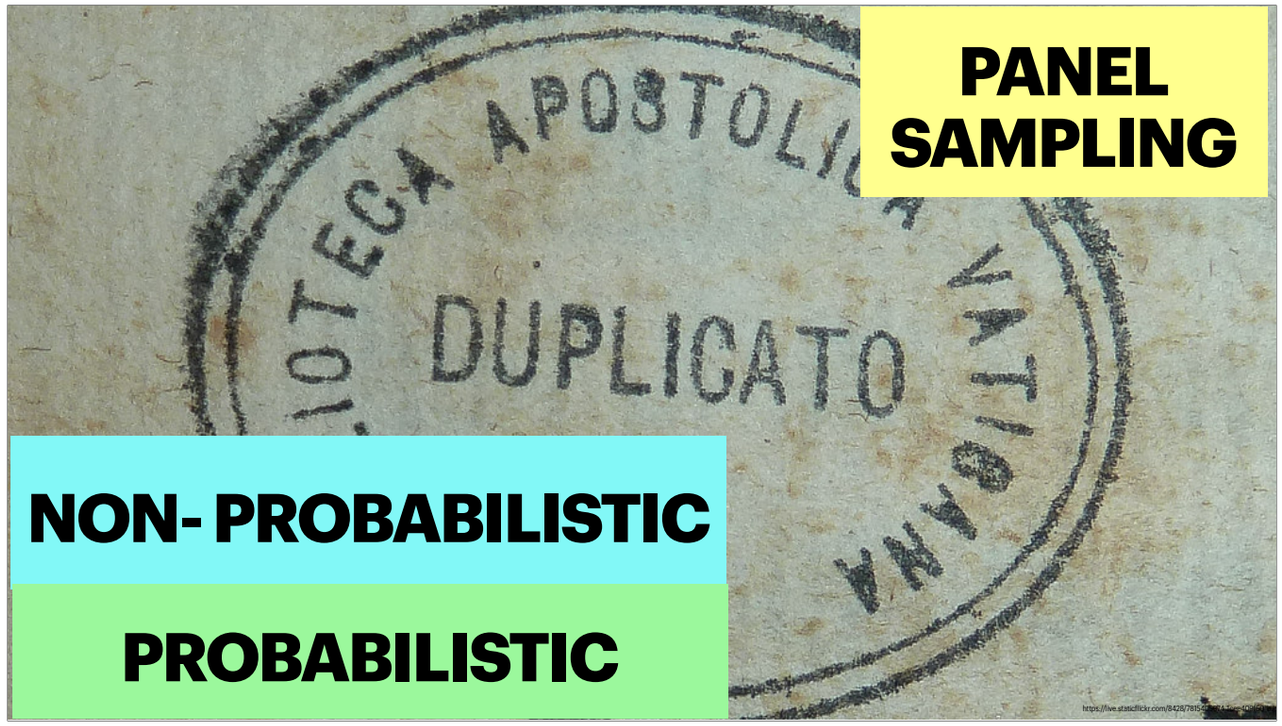
If and only if both selections are random, it is cluster random sampling. Cluster sampling works best when the groups (clusters) are similar to each other but internally diverse on the dimensions of interest.

Suppose that the researcher finds that the seven developers at one office seem much happier than developers in the rest of the company. If the researcher decides to conduct extra interviews at that office, in hopes of unraveling the sources of improved morale, this is called adaptive cluster sampling [69, 73].

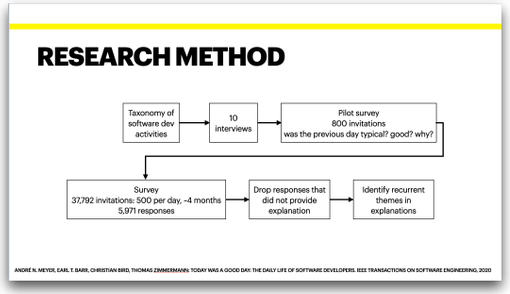


SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. EMPIR SOFTWARE ENG **27**, 94 (2022).

Wikimedia. AngryBiceps. GrafickyZnazornenelony.png February 18, 2022



Panel sampling. The same sample is studied two or more times. Now suppose the researcher implements a program for improving morale, and a year later, re-interviews the same 20 developers to see if their attitudes have changed. This is called panel sampling because the same panel of developers is sampled multiple times. Panel sampling is probability sampling if the panel is selected randomly, and non-probability sampling otherwise.



QUESTION

"We then sent out 37,792 invitations to complete the survey by sending approximately 500 invitations on a daily basis over the course of roughly 4 months. Developers were selected randomly with replacement, meaning that it was possible that a developer would receive the survey multiple times over the course of the study (though never more than once on a given day)."

(A) PANEL SAMPLING

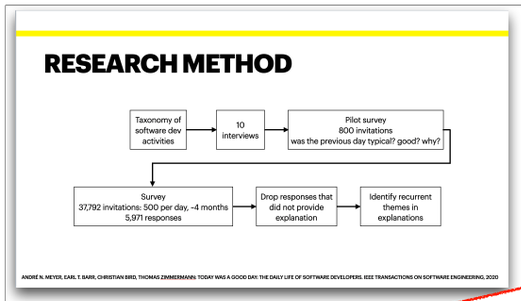
(C) QUOTA SAMPLING

(B) CLUSTER SAMPLING

(D) NONE OF THE ABOVE

ANDRÉ N. MEYER, EARL T. BARR, CHRISTIAN BIRD, THOMAS ZIMMERMANN: TODAY WAS A GOOD DAY: THE DAILY LIFE OF SOFTWARE DEVELOPERS. IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, 2020

(D) None of the above This is not quota/cluster since there are no subcategories. This is not panel since these are not the same people



QUESTION

NON-STANDARD SAMPLING

"We then sent out 37[illegible] survey by sending approximately 5[illegible]e course of roughly 4 months. Develop[illegible]mly with replacement, meaning that it was possibl[illegible]veloper would receive the survey multiple times over the course of the study (though never more than once on a given day)."

(A) PANEL SAMPLING

(C) QUOTA SAMPLING

(B) CLUSTER SAMPLING

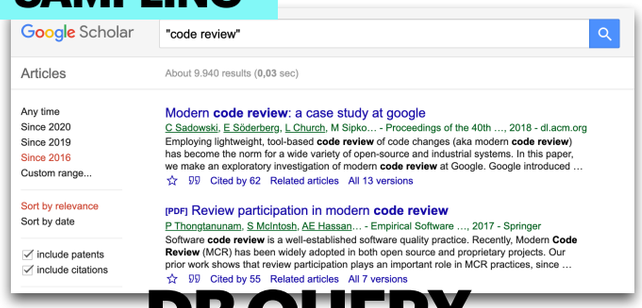
(D) NONE OF THE ABOVE

ANDRÉ N. MEYER, EARL T. BARR, CHRISTIAN BIRD, THOMAS ZIMMERMANN: TODAY WAS A GOOD DAY: THE DAILY LIFE OF SOFTWARE DEVELOPERS. IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, 2020

(D) None of the above This is not quota/cluster since there are no subcategories. This is not panel since these are not the same people

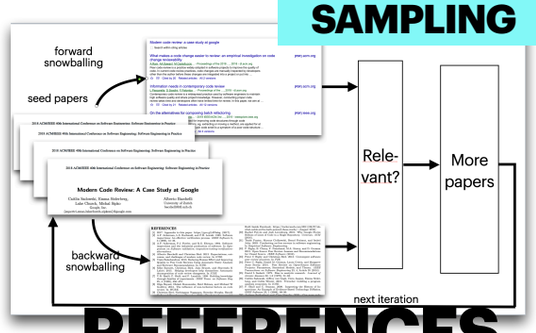
ANOTHER NON-STANDARD SAMPLING: LITERATURE REVIEW

PURPOSIVE SAMPLING



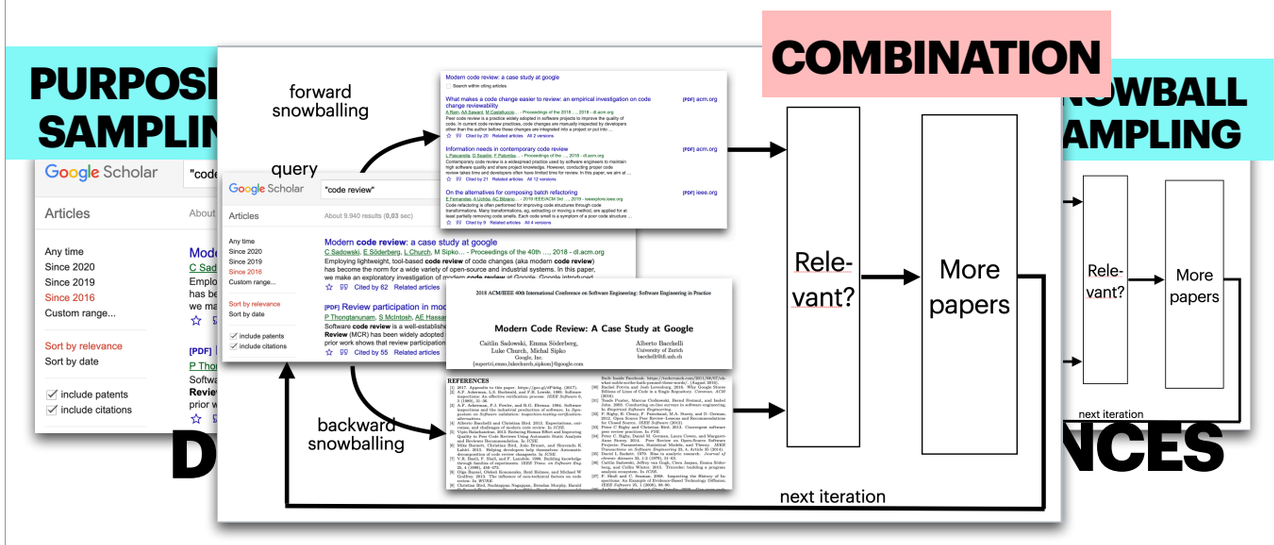
DB QUERY

SNOWBALL SAMPLING

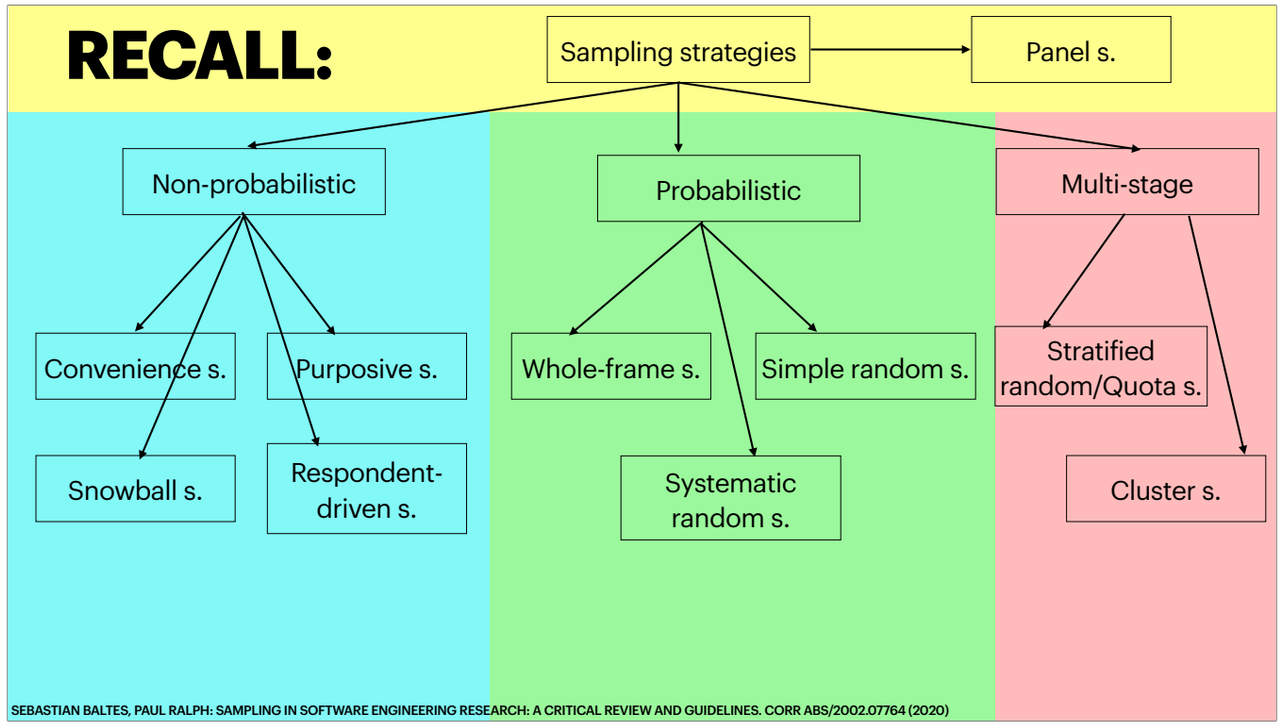


REFERENCES

ANOTHER NON-STANDARD SAMPLING: LITERATURE REVIEW



HOW COMMON ARE COMBINATION STRATEGIES?

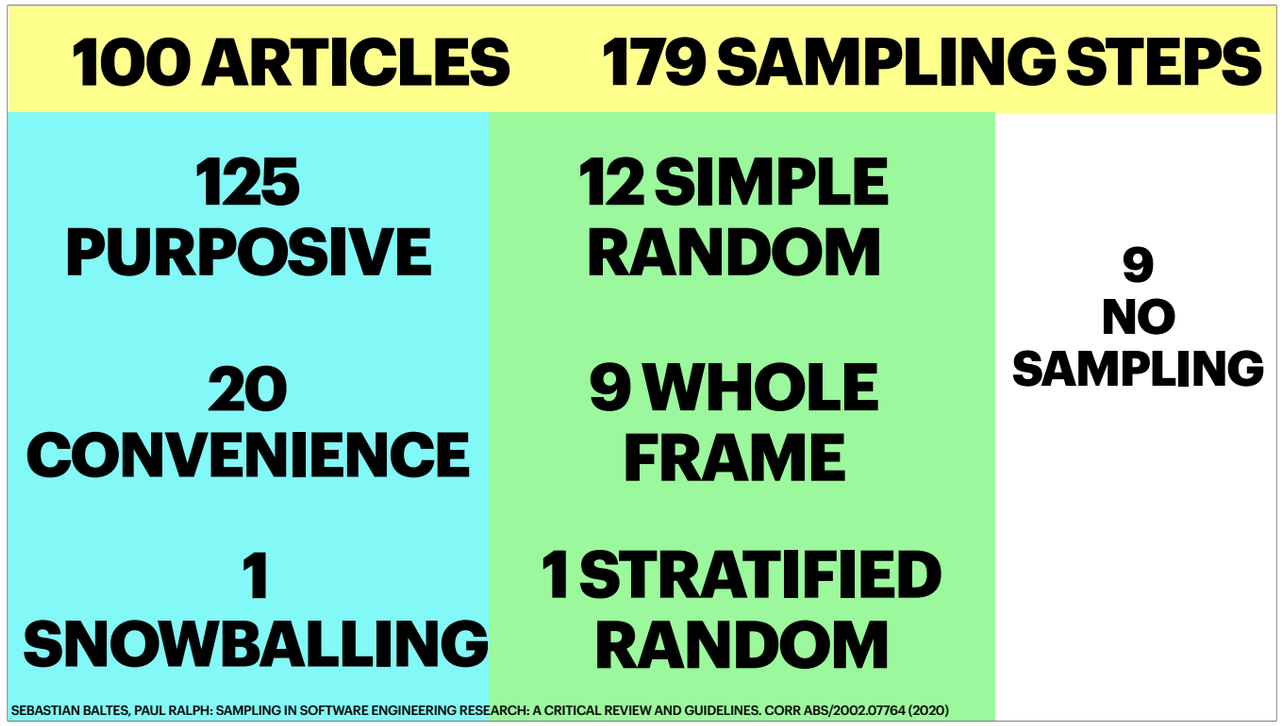


Non-probability sampling includes all of the sampling techniques that do not employ randomness.
Probability sampling includes all of the sampling techniques that employ randomness.
Methodologists often present multistage sampling as a special case where two or more sampling strategies are intentionally combined.

COMBINATIONS OF DIFFERENT STRATEGIES

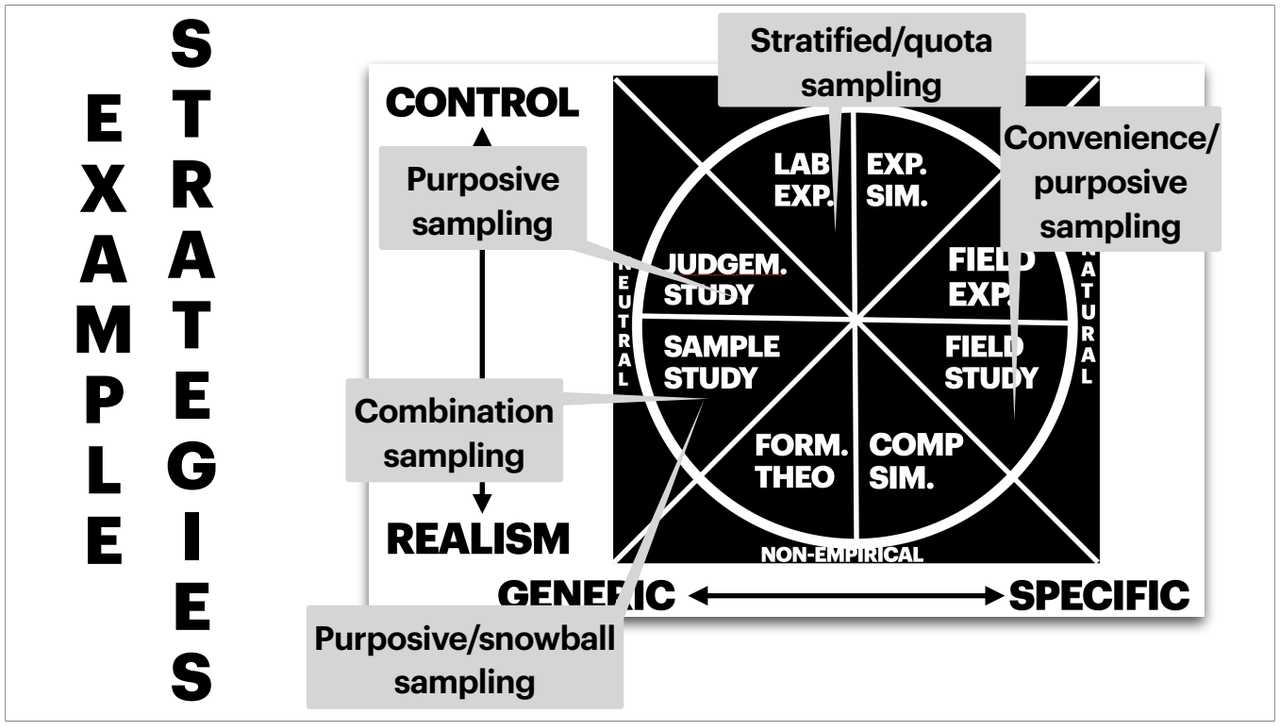
100 ARTICLES ≠ 179 SAMPLING STEPS

SEBASTIAN BALTES, PAUL RALPH: SAMPLING IN SOFTWARE ENGINEERING RESEARCH: A CRITICAL REVIEW AND GUIDELINES. CORR ABS/2002.07764 (2020)



9 no sampling

EXAMPLE STRATEGIES



Cleaning by Nick Youngson CC BY-SA 3.0 Pix4free.org

Whatever sampling strategy is used we need to ensure that the data is reliable: for example that Prolific respondents claiming to be developers are actually developers and that GitHub repositories actually contain software projects rather than websites or cookbooks.
We might need to **clean the data, i.e., exclude data points that are seen as unreliable, complete missing data etc.**

SAMPLING

SAMPLING IS THE PROCESS OF SELECTING A SMALLER GROUP OF ITEMS TO STUDY (A **SAMPLE**) FROM A LARGER GROUP OF ITEMS OF INTEREST (**POPULATION**).