

## **Zoommeeting Digitization Board 2.6.2022, 8:30-9:20**

KHuK, Jan Steinheimer, KET, Günter Quast, Christian Zeitnitz, KFN Susan Schorr, Tobias Richter, KFSI, KAT, Andreas Haungs, KFS, KfB, Erik Bründermann, RDS, FI, BDA, Gregor Kasieczka, RDM, UI, Pierre Schnizer, KD Dirk Lützenkirchen-Hecht, RPB, ErUM-Data-Hub Angela Warkentin (Gast), SP, Martin Erdmann

### **Focus Topic: Big Data Analytics (Gregor)**

The steering group of Big Data Analytics meets every two weeks, all ErUM committees participate (KFSI representation still open). The group has developed a survey to find out which topics are suited for interdisciplinary research work, and which are specific questions to be solved individually. From the 15 answers so far, most interesting information can already be captured (see next pages). There is a broad coverage / application / development of modern methods, and quite some overlap in the techniques used. Cross-disciplinary online seminars are favoured over in-person workshops. A wiki is needed as well as a match-making page (see Eriks <https://indico.desy.de/event/28766>). PT.DESY has sent an email to all PIs who applied to the ErUM-Data call, inviting to the big-data-analytics email list.

#### **Requested action items:**

- Please **distribute the survey to digital experts** you know, to email lists of digital experts of your ErUM committees (e.g. KET, KAT,... have these lists), and to NFDI colleagues.  
<https://campus.lamapoll.de/Big-Data-Analytic-Kopie>
- A **wiki** is needed! The selection of the underlying system is still ongoing: The latest attempt using a git concept was not convincing (see minutes of the topic group chair meeting 19.5.).
- For **future surveys** we need another system beyond LLamapoll.

The **ErUM-Data-Hub** has a first web presence (Angela), please feel free to comment:

<https://www.erumdatahub.de>

We will try now again to complete the team of the ErUM-Data-Hub, trying to attract very good people.

#### **Knowledge Distribution (Dirk) on the program 2022:**

Preparations for the Train-the-Trainer workshop next week June 9/10 in Wuppertal are well advanced: <https://indico.scc.kit.edu/event/2728>

The locations for the following schools/workshops have been booked:

08.08.22 - 12.08.22 Deep Learning School "Basic Concepts", Haus Nordhelle in Meinerzhagen

12.09.22 - 16.09.22 Expert Workshop on Autonomous Deep Learning, Hotel zur Post in Wiehl

28.11.22 - 01.12.22 Deep Learning School "Advanced Concepts", Kloster Damme

**User Interface (Pierre):** workshop planning is ongoing, will have a date/place in the next 2 weeks.

#### **Research Data Management**

The workshop "ErUM-Data and NFDI" can be planned for September 2022. After conversation with BMBF there will be no PRISMA strategy talk on ErUM-Data at that time.

- **Action item: we need a date very soon.**

#### **Future monthly meetings,**

Next Meeting 16-June-2022, 8:30-9:30 **Topic Groups**

Next Meeting 30-June-2022, 8:30-9:30 **Digitization Board**

# **Big Data Analytics - Short Update**

**GK, 2.6.2022**

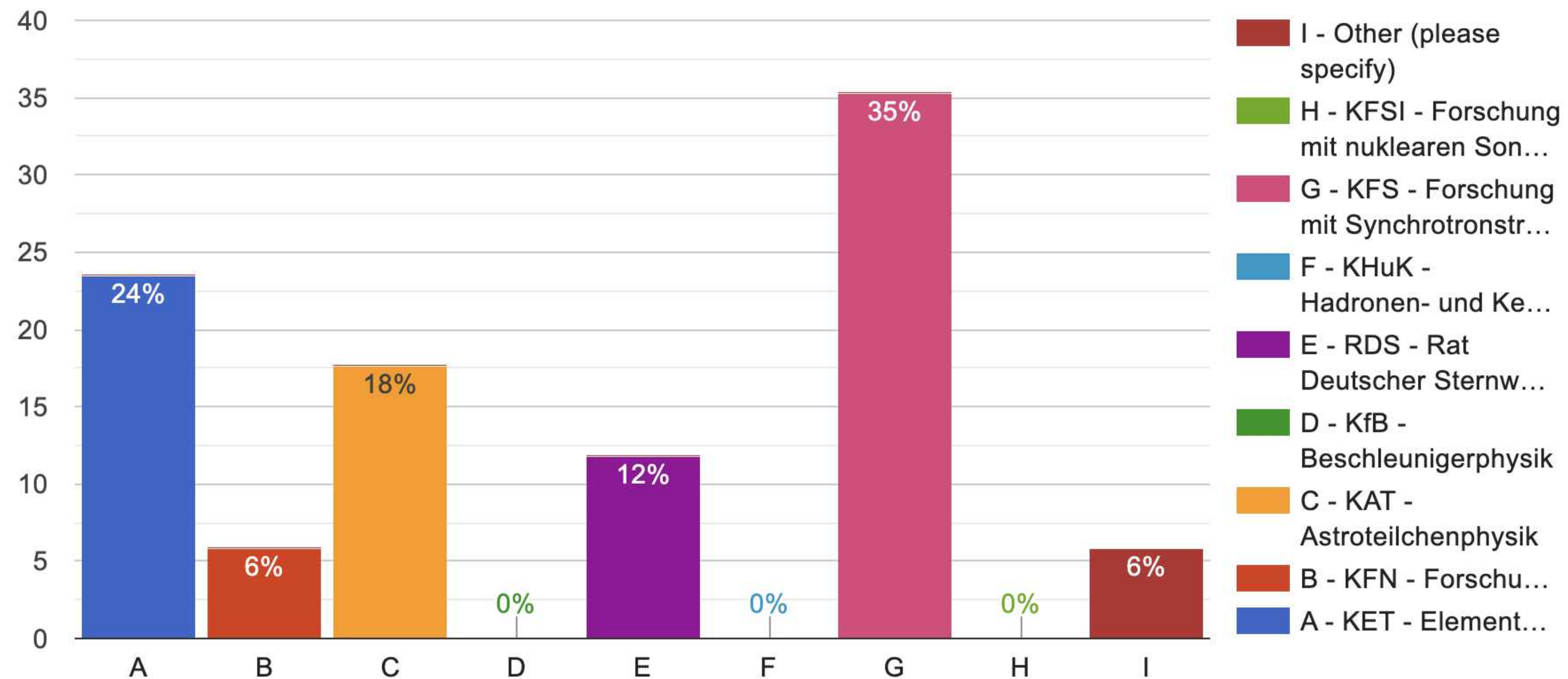
# Overall

- **Topics:**  
The development, support and use of data analysis and algorithms, including AI.
- Currently meeting bi-weekly
- BDA was main topic of recent ErUM-Data *Software & Algorithms* call.  
Connecting the projects as one key activity. Awaiting results of 2nd review phase.
- **People:**
  - KET: GK
  - KFN: Christian Stieghorst
  - KAT: Jonas Glombitza
  - KfB: Michael Bussmann
  - RdS: Jochen Weller
  - KHuK: Jan Steinheimer
  - KFS: Andrea Thorn
  - KFSI:

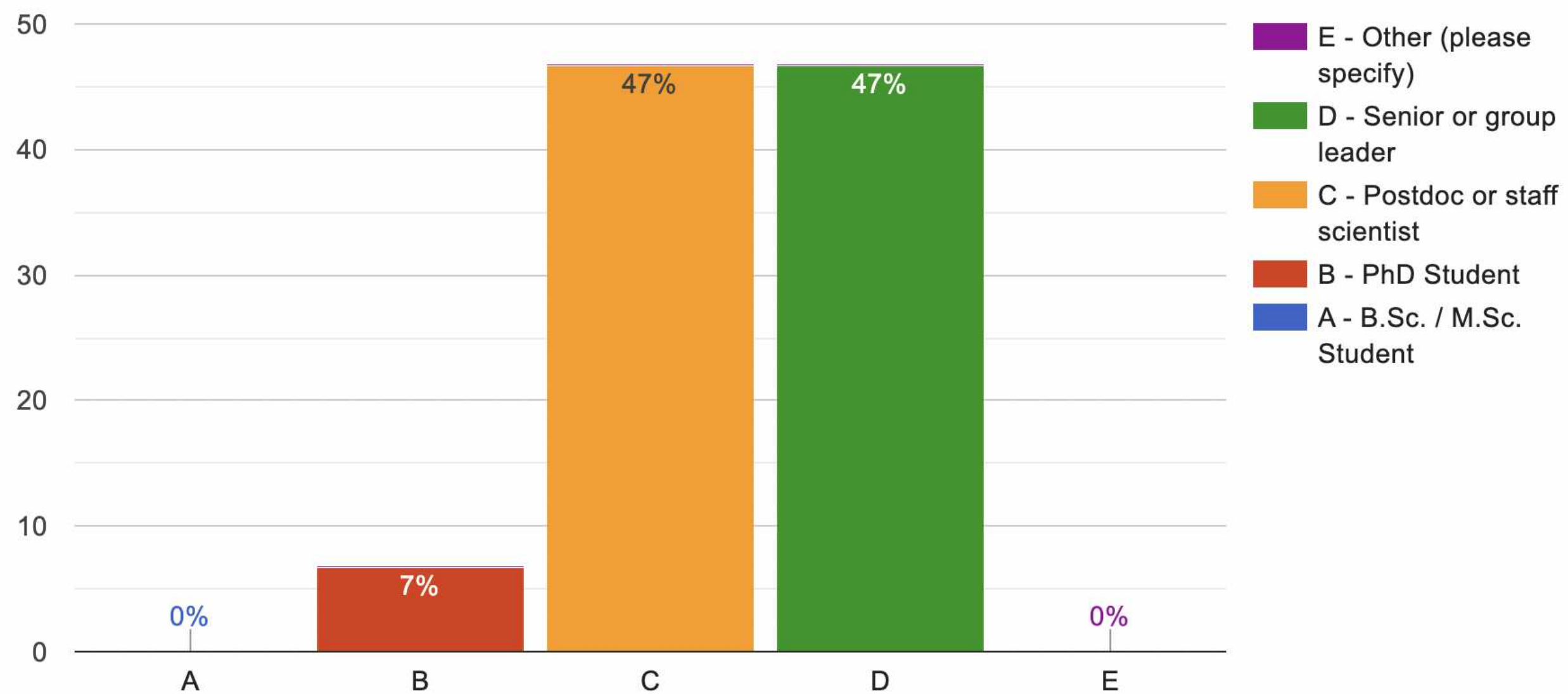
# Survey

- Identify topics relevant across communities
- Survey sent out last week to BDA mailing list.  
15 responses so far.
- Will give a first overview.  
Detailed analysis to follow when survey concluded
- Organisational:  
BDA collected questions as a text document  
Many thanks to Angela Warkentin / ErUM-Data Hub for turning into a poll
- Technical:  
LLamapoll was used. German commercial platform.  
Might benefit from another solution longer term (or start paying)

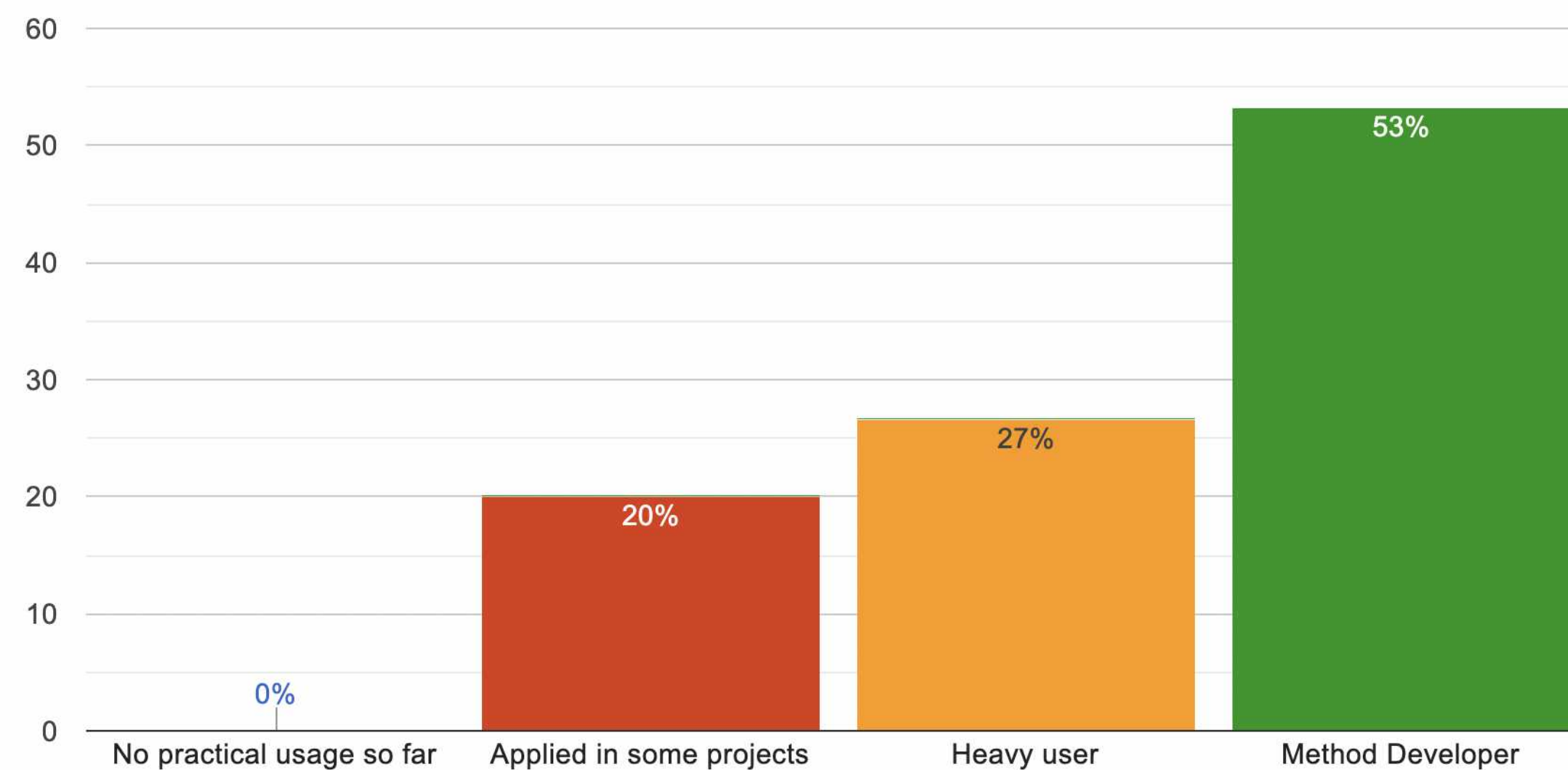
### What is your primary Area of Research?



### What is your position?

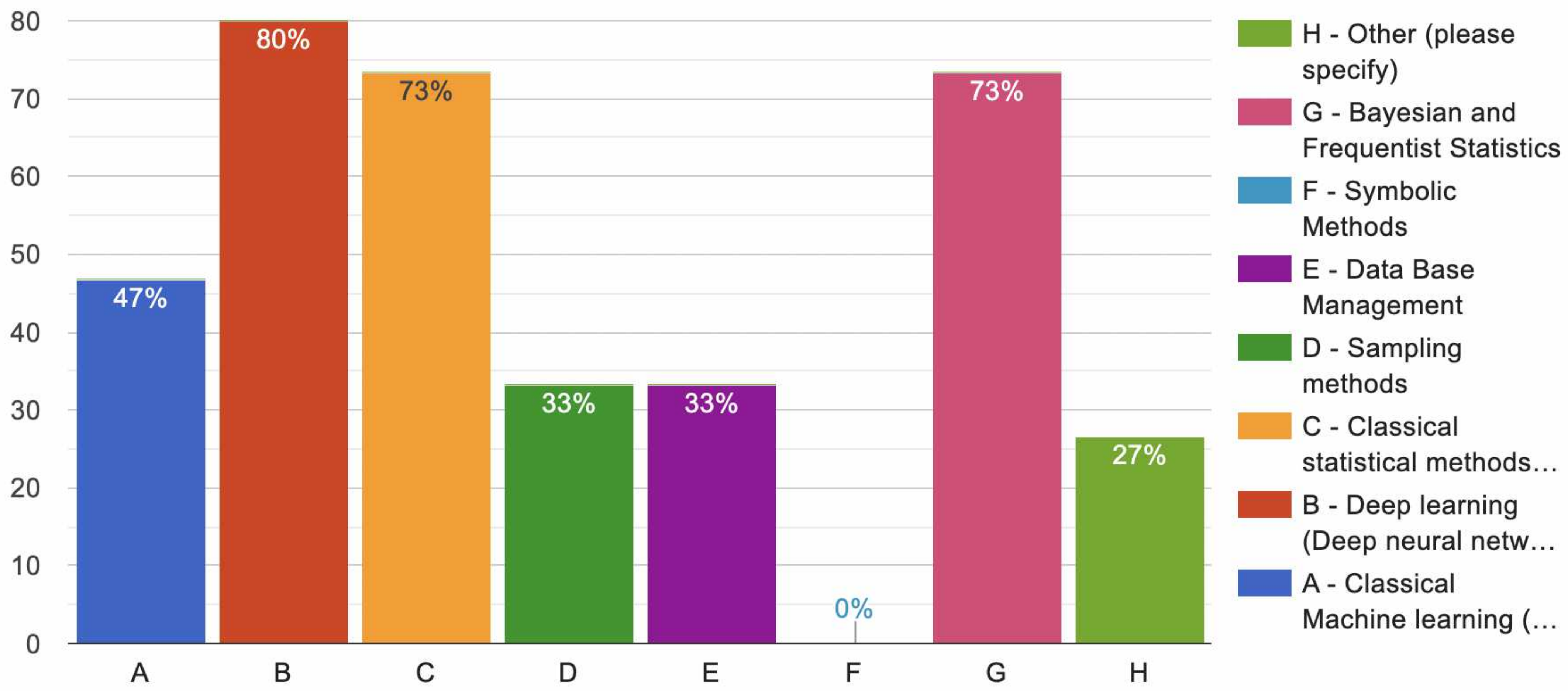


### What is your overall usage of modern data analysis techniques?

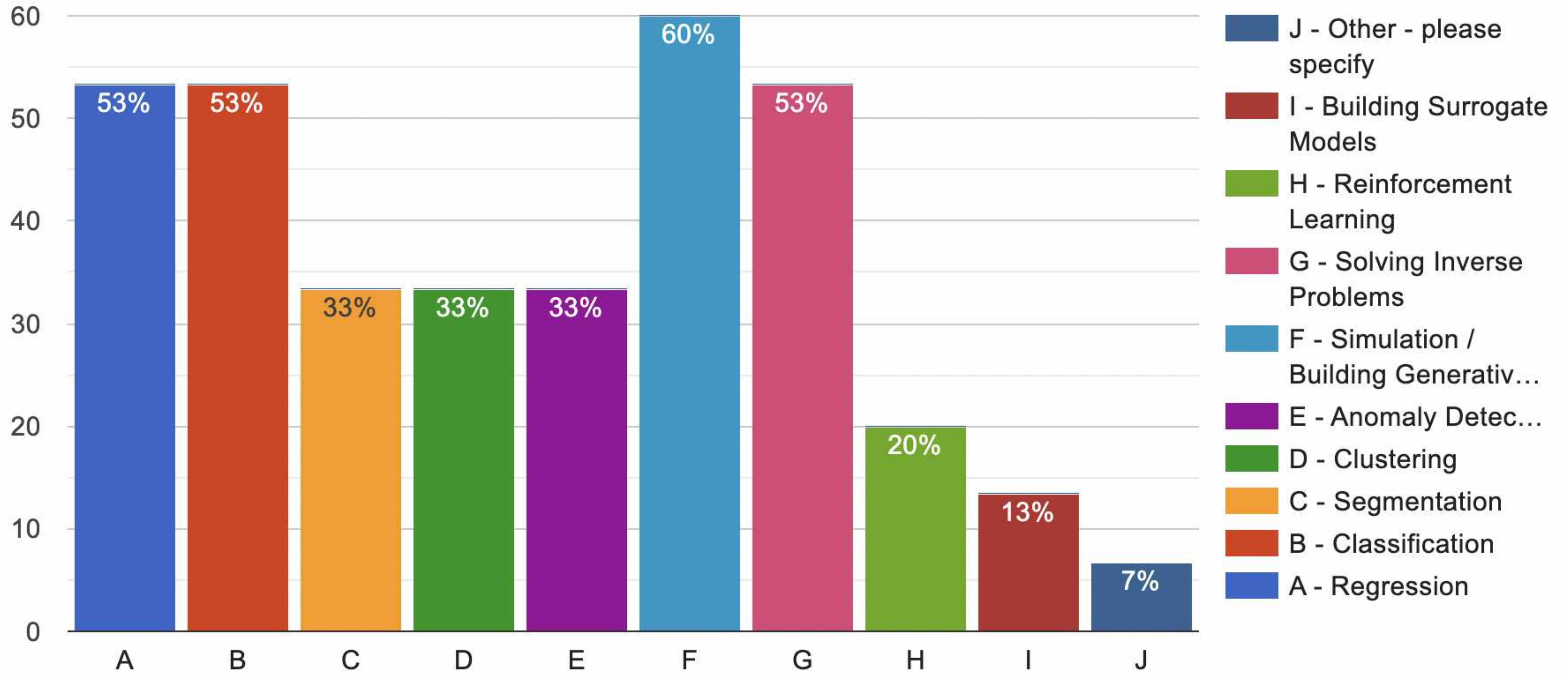


### Which techniques are you interested in?

Häufigkeit in %  
(bezogen auf Anzahl der Teilnehmer: 15)



### Which tasks are you trying to solve in your research?



# Technical & Legal questions

- Python as most widely used programming language
  - Jupyter popular
  - Hdf5, numpy and csv as widely adopted data formats
  - Even split across Keras/Pytorch/TensorFlow/ScikitLearn
  - Primary targets are CPU/GPU. No TPU/ASICs. One FPGA
- 
- Guideline for licensing interesting



# Overlaps

- Around one quarter is also active in NFDI
- 40% have submitted proposals in ErUM-Data call

# Potential offerings

- General favourable of a Wiki-page collecting results and examples of BDA topics by German researchers
- General favourable for a match-making page
- General very favourable for a cross-disciplinary online seminar
- General less favourable for in-person workshops

# Last words

- Just two samples:
- *...We need to define major research directions with the highest potential for progress and discovery: short term - medium term - long term. For this we need workshops with leading active researchers in the various ErUM fields.*
- *...I want to directly connect the three pillars in physics: experiments - mathematics - machine learning. So far, there is only slow progress in the directional work from machine learning to analytic expressions. We need physics laws in terms of mathematical expressions resulting from machine learning methods.*