

ArviZ user survey results

Introduction

The purpose of conducting the ArviZ survey was to get feedback from users. The survey ran from June 14th to October 28 (2021). The survey was created using Google forms and was advertised on the ArviZ docs website, and Twitter. The survey comprised of two parts/pages. The first page was for all the users whereas the second one was for contributors of the ArviZ.

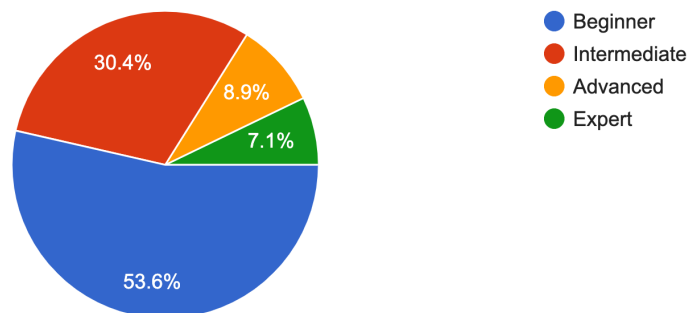
In total, we received 56 responses, out of which 20 people filled the contributor's section.

Graphical Summary

The graphical summary of close-ended questions is given below.

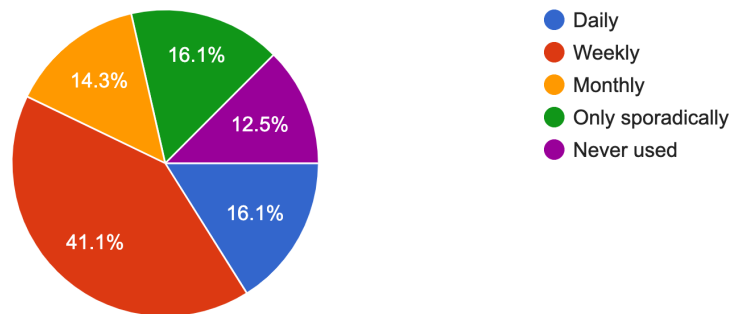
What is your ArviZ experience level?

56 responses



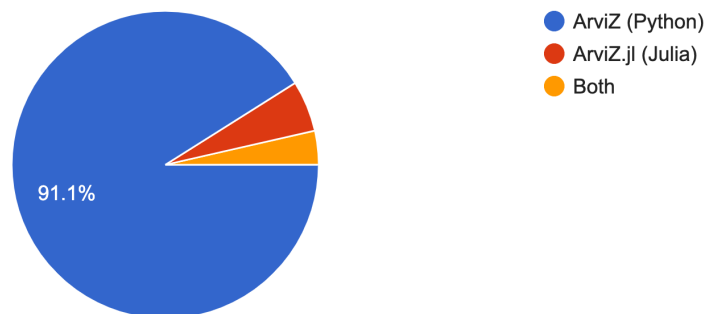
How often do you use ArviZ?

56 responses



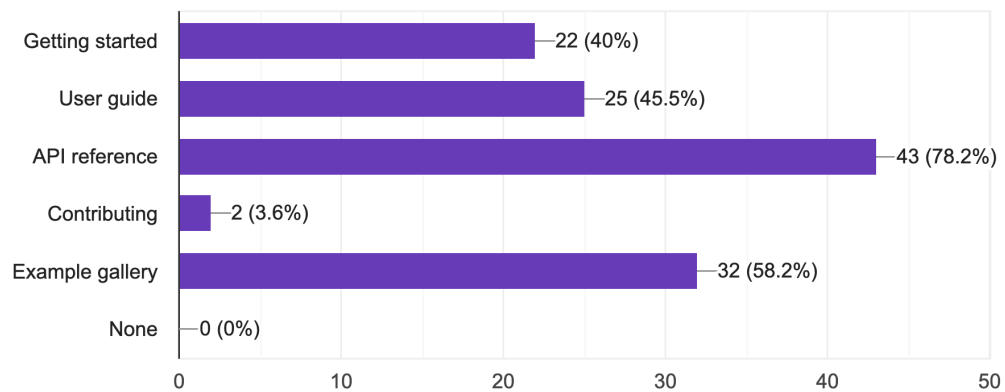
Which docs do you use the most often?

56 responses



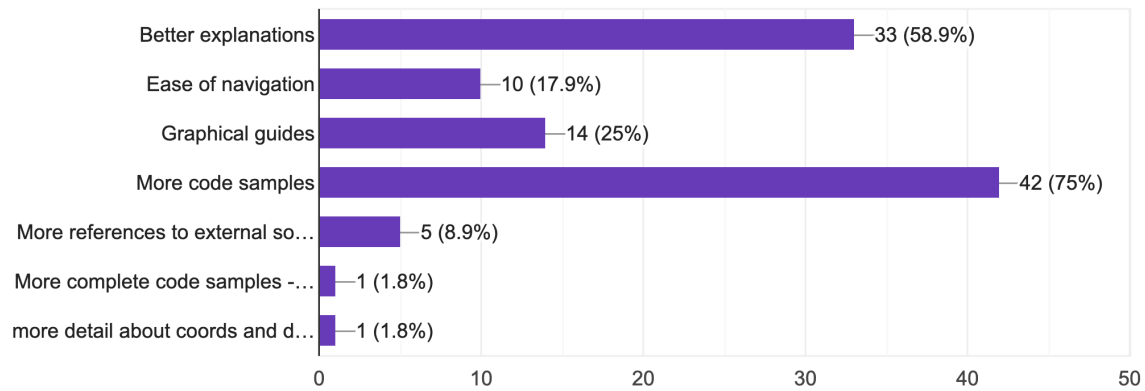
Which parts of the documentation do you usually use?

55 responses



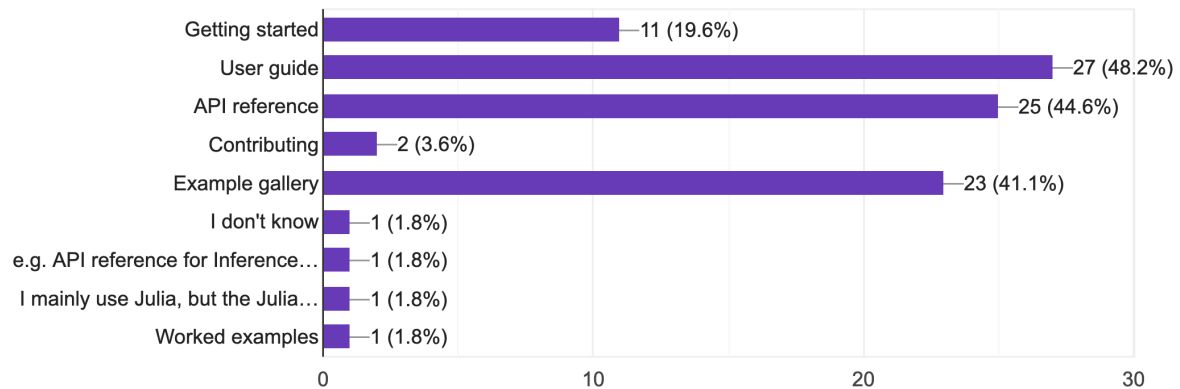
Which of the following documentation features should be improved/added?

56 responses



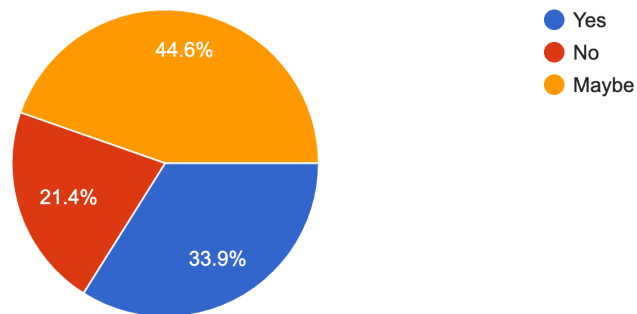
Which section of documentation do you think we should improve?

56 responses



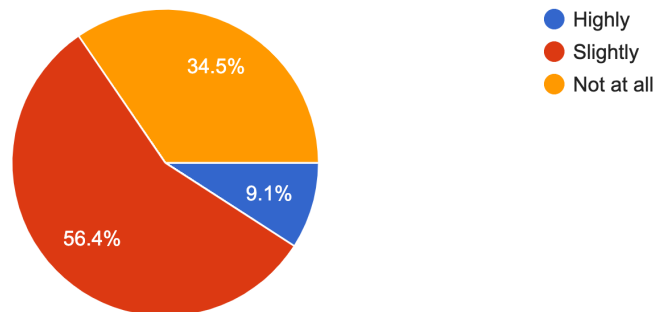
Do you think that the current documentation is enough for learning “How to use ArviZ”?

56 responses



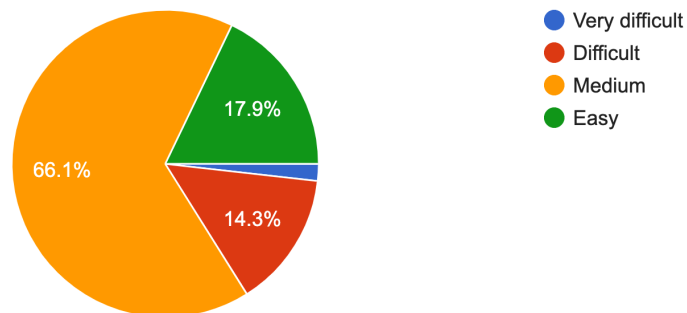
How strongly would you suggest adding a section “How to read ArviZ documentation”?

55 responses

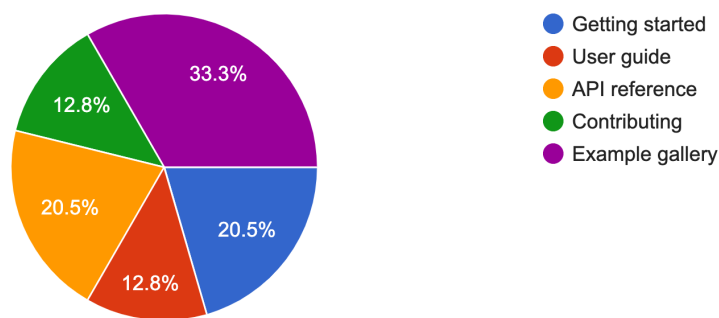


How would you rate the language of ArviZ documentation on difficulty level?

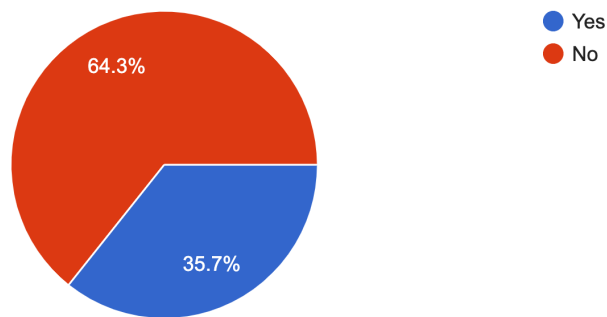
56 responses



What is one thing you would not change about the current documentation?
39 responses



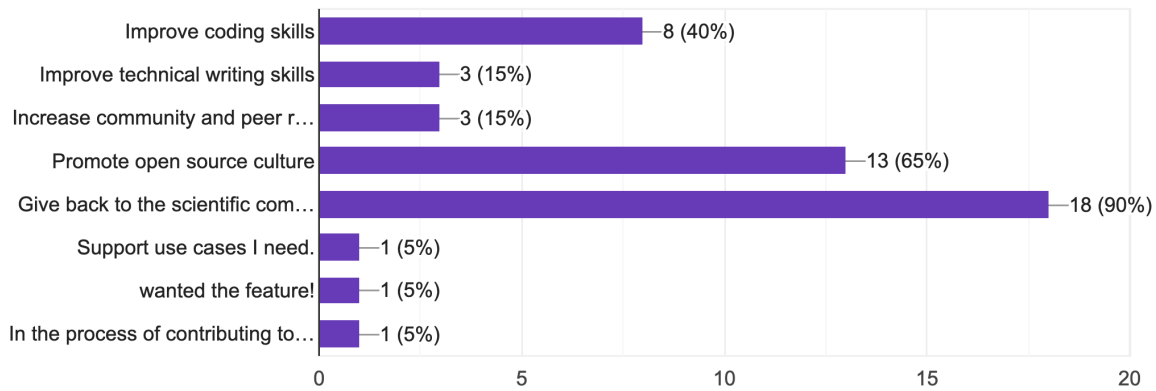
Have you contributed to ArviZ or considered doing so?
56 responses



Contributors Section

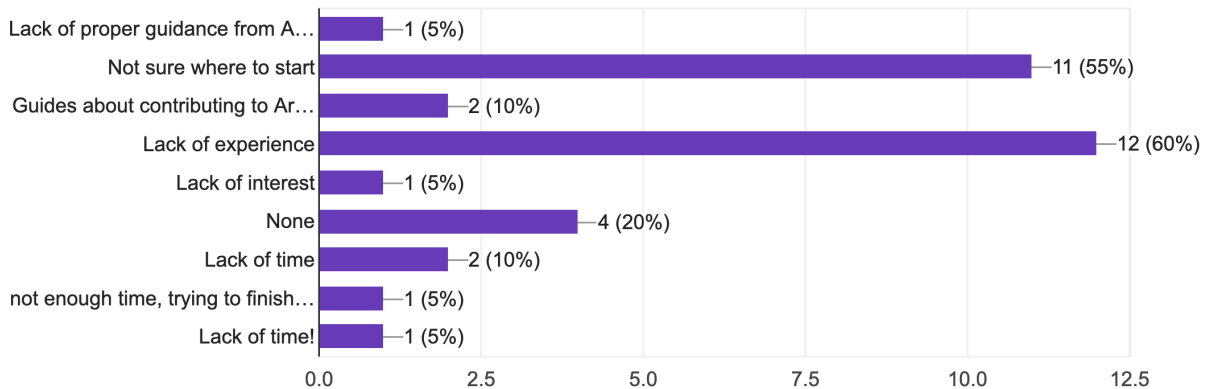
What are your motivations for contributing to ArviZ?

20 responses



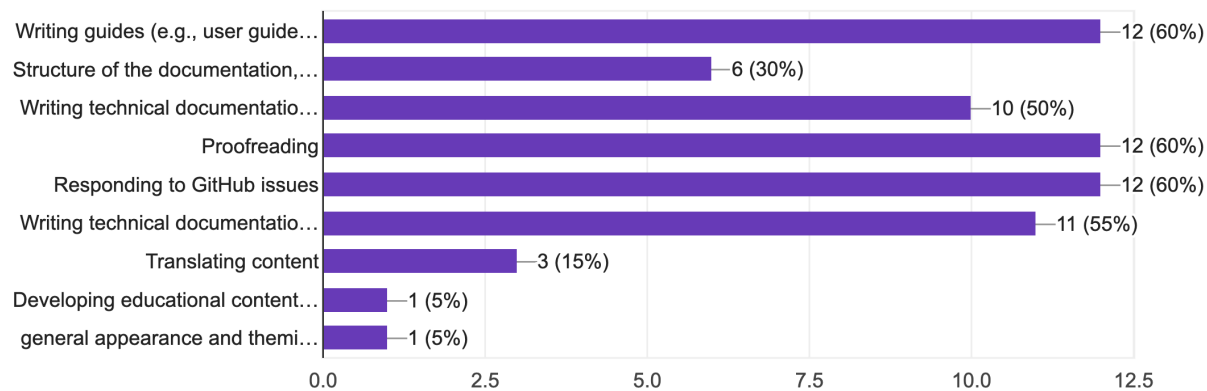
What are the factor(s) stopping you from contributing to ArviZ's documentation? If you have already contributed, choose those (if any) that made contributing harder.

20 responses



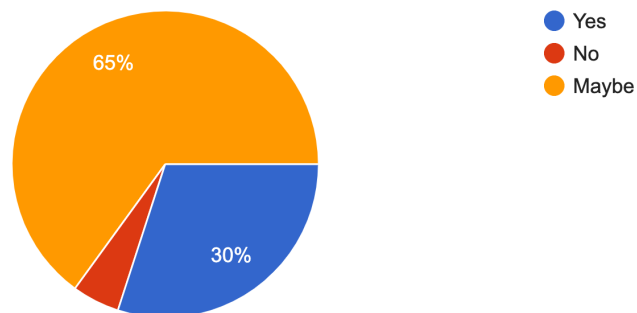
In what ways would you be interested in contributing to ArviZ's documentation?

20 responses



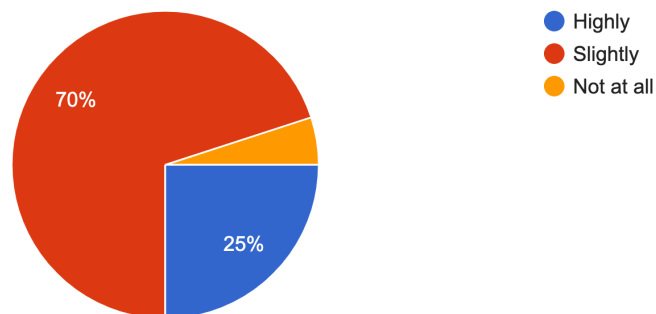
Do you think that the current documentation is enough for learning “How to contribute to ArviZ documentation”?

20 responses



How strongly would you suggest adding a section “How to write ArviZ documentation”?

20 responses



Open-Ended Questions

There was one open-ended question in the survey for people to share their feedback with us. We got 22 responses which are given below.

Question: "Is there something you have in mind about how to improve our documentation and you want to share?"

22 responses

- It might not be something about documentation, but I misunderstood the meaning of "hdi_3%" or "hdi_97%" when I first used `arviz.summary`. I thought 3% of samples exist at the left tail and the other 3% exists at the right tail of the distribution. People would think this would be a very beginner mistake, but I think many beginners might make the same mistake...
- what are the goals and non-goals of the project.
- You should remove the Python dependency from `ArviZ.jl` and make it Julia-only.
- Something about how to set up with PyStan 3 to get the most out of it (i.e. to generate the prior, the posterior predictive, including observed data, etc.)
- I like when there are short snippets of code along with the API documentation like numpy does it.
- Use of the ``coords`` argument to various plot functions to select a subset of coordinates to use in plots is currently a bit sparse - the documentation indicates this is passed to ``Dataset.sel`` but no link is provided to this documentation (and it not necessarily immediately clear this is a reference to a type defined outside of ArviZ, assuming it refers to `xarray.Dataset`) and there was not obvious examples of how to use this argument.
- I think the contributing guide is great. Maybe having a shorter version available only with the necessary git commands can be helpful.
- It would be neat to see a flow chart of the Bayesian workflow annotated with the relevant ArviZ functions at each step. E.g., in the step where you think about priors have a list of things useful for looking at the prior predictive, etc. I think this would be useful for helping newcomers build a mental map of what ArviZ functions to use at what step (the example gallery is great but perhaps it could be structured to reflect this ordering?). Anyways, I'm a big fan so good luck with Summer of Docs!

- A lot of important explanations are missing. For example, `arviz.summary()` returns values such as `mcse_mean`, `mcse_sd`, `ess_bulk`, `ess_tail`, and `r_hat`. It is however not explained what these abbreviations are, or where to find more info about them.
- Two points:
 - 1. "Getting Started" from the navigation should take you to Quickstart, not a bulleted list.
 - 2. Needs a page explaining **why** model checks are important, rather than just how to code them
- More explanation of the spec for a data structure for mcmc output - I guess the `InferenceData` object. Foreground the fact that `xarray` doesn't support multidimensional groupby. Examples showing how to augment `arviz` plots with arbitrary `matplotlib`. Generally focus on interoperability - I think most features `arviz` explicitly provides can be discovered already but even though `arviz` essentially is compatible with standard tools but it takes quite some effort to find out how to eg draw a vertical line on a plot with `ax.hline` or connect some output with an existing `pandas` dataframe.
- Documentation only appears available for the development version, not the stable version. This can lead to confusion when functionality appears to be missing, for example: <https://discourse.pymc.io/t/no-module-named-arviz-labels/6963>
- A further explanation or links to external explanations for the statistical values (sometimes it is hard for me to find out what the values actually mean).
- More clear and solid tutorials. The ones implemented right now are messy and incomplete.
- I think I have already shared most of my comments as issues
- Not `arviz`' "fault". But when linked from the `pymc3` docs very often the page cannot be found
- Please add more example models than just 8 schools!
- Usage of ``InferenceData`` with `PyMC3` needs more examples because there are many bugs.
- I'd appreciate if the documentation did not assume I know what everything is, and only need to know how to do it in `Arviz`. I'm learning details of Bayesian inference at the same time.

- Nothing made sense until I read the "terminology" section in the user guide. The docs could start there, instead of showing how to convert data into an InferenceData object (which meant nothing to me because i didn't know the terminology). I'm still confused about some of the "groups" defined in the schema, and which I should put my data into. It might be useful to have a guide for someone not coming from a supported mcmc python library. To be honest, it looks super cumbersome to work with these InferenceData objects. Suppose I want to take a subset of an InferenceData object based on the values in one of the fields (e.g. take the 5% with the best log_likelihood)? Well it looks like one needs to manually extract each of the variables (making sure to do so only for the "related" variables) and put them into a fresh InferenceData object.
- Dimension and Coordinate are hard concepts for me and how to use them in practice
- More examples of the numba and xarray integration

Conclusion

The survey helped us in the current evaluation of the docs. It could also help us in the future for documentation development and improvement.