STScI | SPACE TELESCOPE SCIENCE INSTITUTE

EXPANDING THE FRONTIERS OF SPACE ASTRONOMY

JWST GO1 Data Analysis Survey

Klaus Pontoppidan

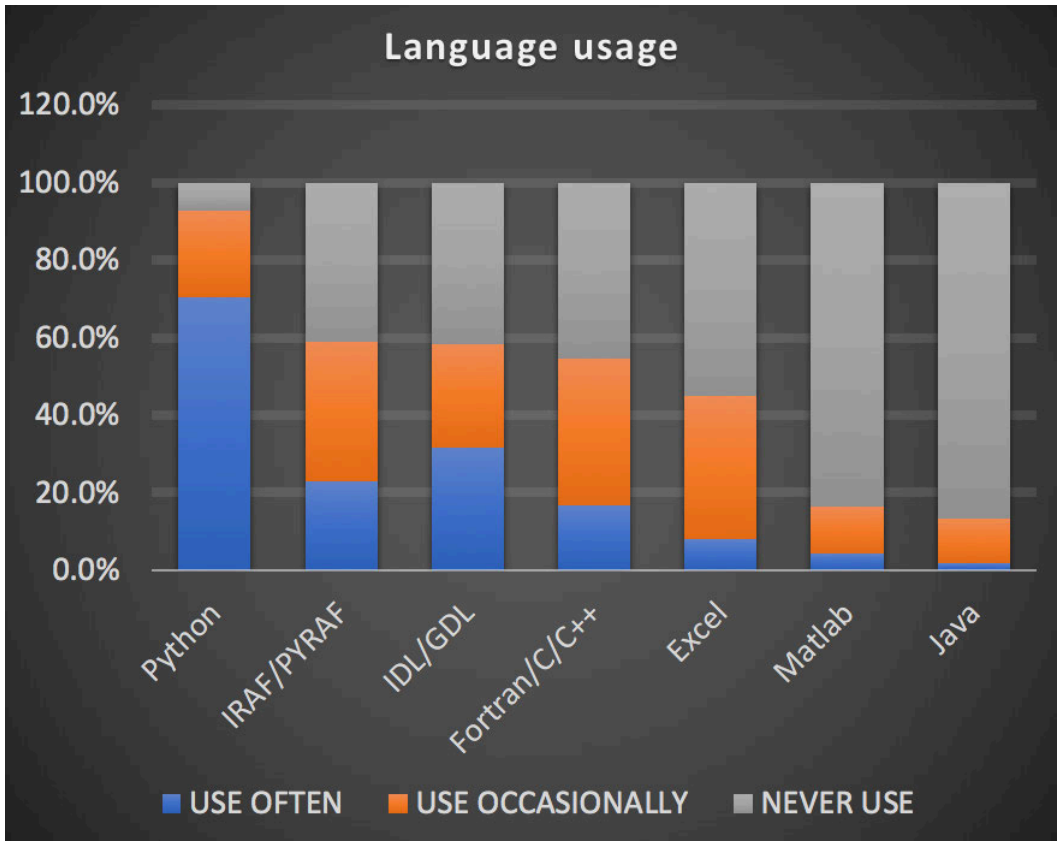
JSTUC - September 9, 2019

Basic statistics

- Survey run time: June 7, 2019 → August 2, 2019
- 504 total responses
- Average question completion: 82%
- Demographics:
  - 3 undergrads (<1%)
  - 58 grads (14%)
  - 95 postdocs (23%)
  - 242 faculty/staff (59%)
  - 13 other (3%)
- 85% of respondents are primary JWST data analysts (others rely on team members). 3% do not plan to analyze JWST at all…
- 88% responded to an email; 12% responded via social media (Twitter/FaceBook).
- Median time spent: 5m 2s.

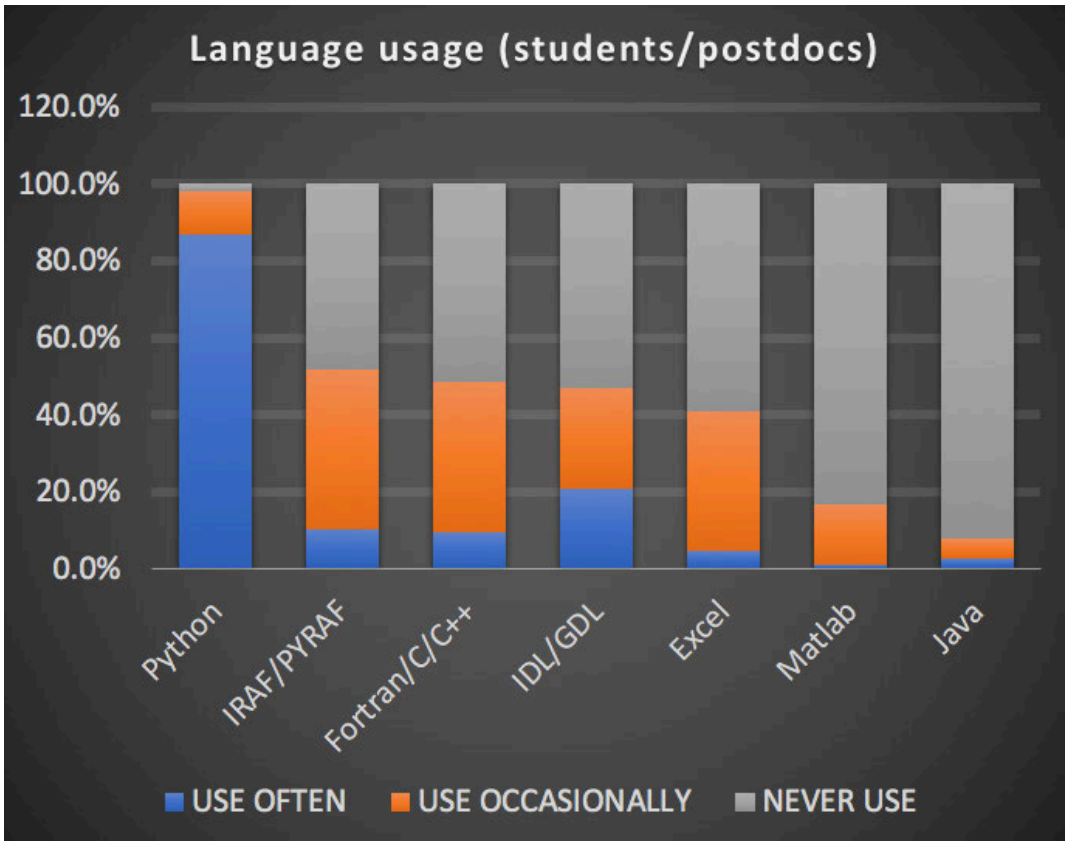
Programming language usage

- Python most often used – only 8% never use it.
- But >50% use IRAF/IDL/Fortran+C. Even Excel is used occasionally by >40% of respondents.
- Comments mention R and Mathematica often.
- Also to a lesser degree Julia, SuperMongo, GILDAS/CASA/MIRIAD, Perl, Igor, …
- Clear generational evolution – low numbers of students+postdocs rely on IRAF (10%) or IDL (20%).

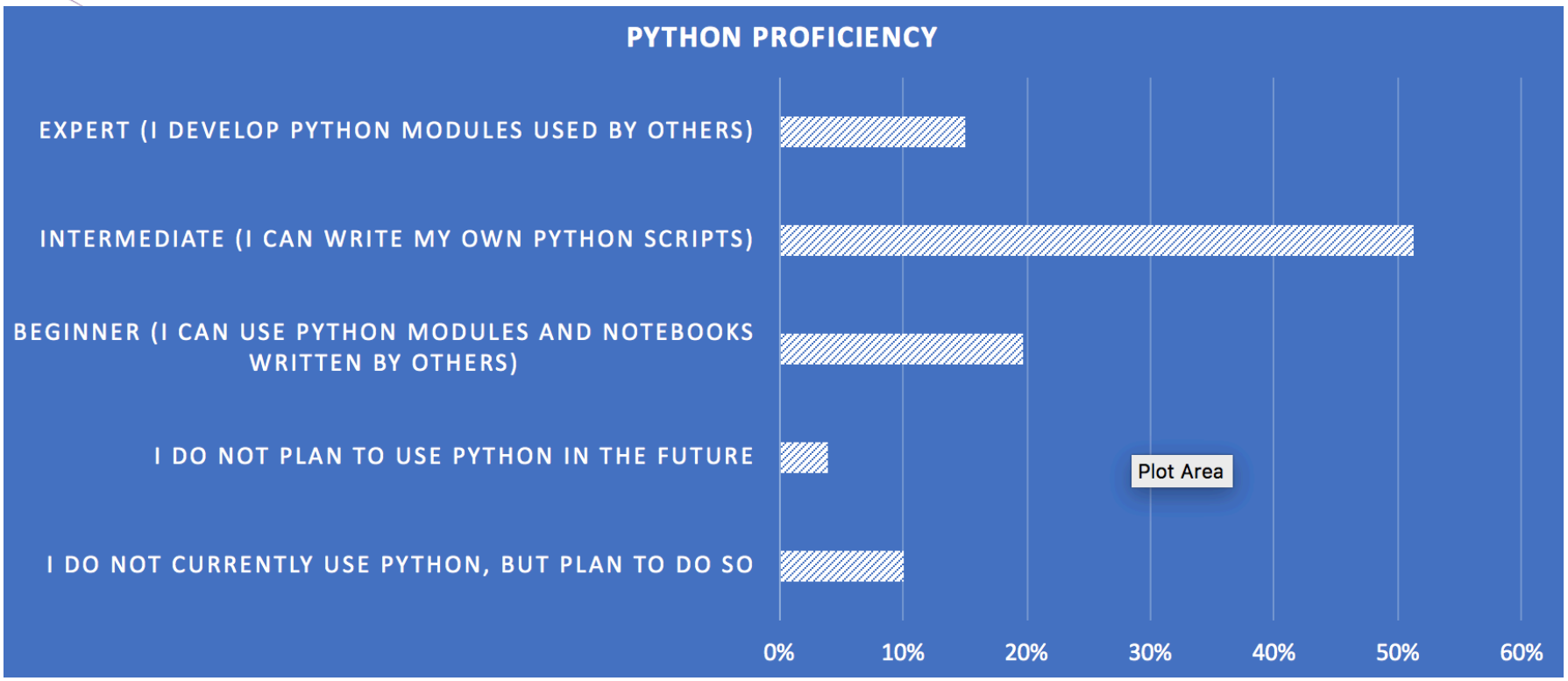


Programming language usage

- Python most often used – only 8% never use it.
- But >50% use IRAF/IDL/Fortran+C. Even Excel is used occasionally by >40% of respondents.
- Comments mention R and Mathematica often.
- Also to a lesser degree Julia, SuperMongo, GILDAS/CASA/MIRIAD, Perl, Igor, …
- Clear generational evolution – low numbers of students+postdocs rely on IRAF (10%) or IDL (20%).

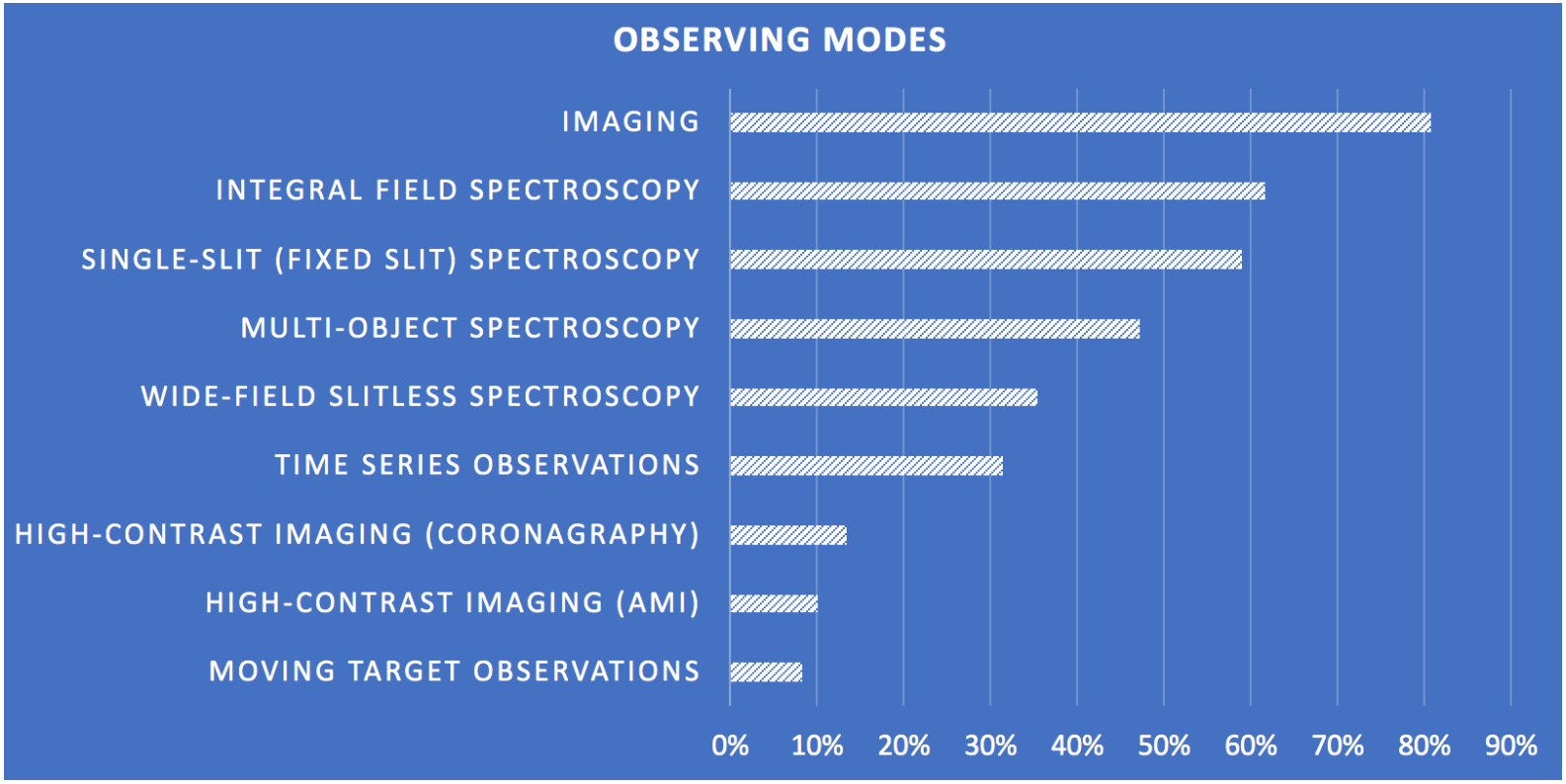


Python proficiency of respondents



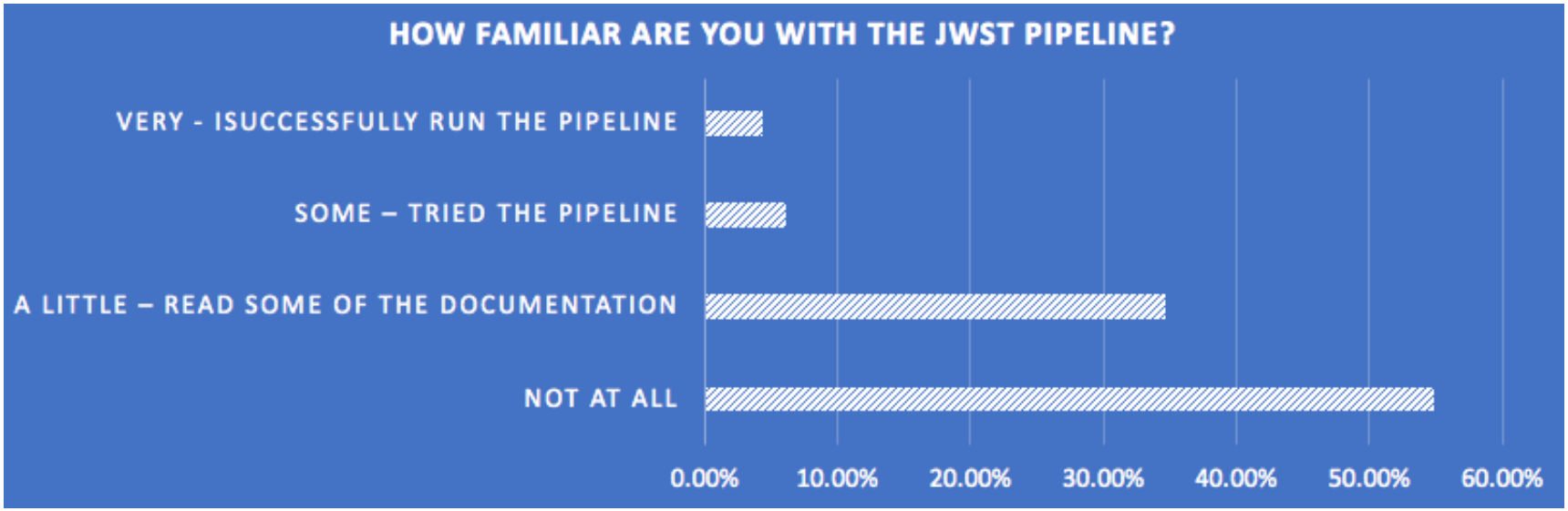
35% of respondents identify as not proficient in Python (beginners and non-users).

Respondents' observing modes



- Relatively low numbers of high-contrast imaging and moving target analysts. Surprising large number of fixed-slit user
- 16.5% selected either coronagraphy or AMI (so significant overlap).

Familiarity with the JWST pipeline

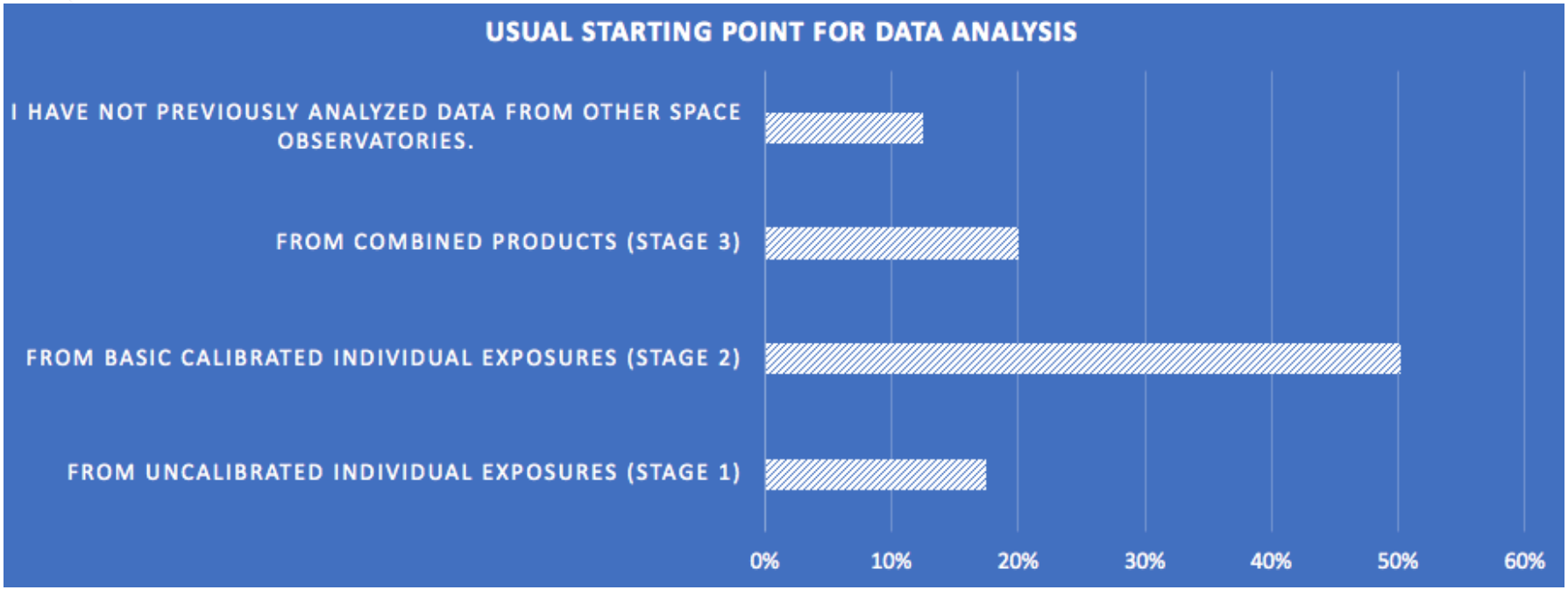


No surprises - few people have tried the pipeline; we have not yet received any practical feedback.

How to use training in JWST data analysis

- Observers want training specifically for JWST.
- Not for more general analysis purposes.
- Critical need for training in the pipeline and data products.

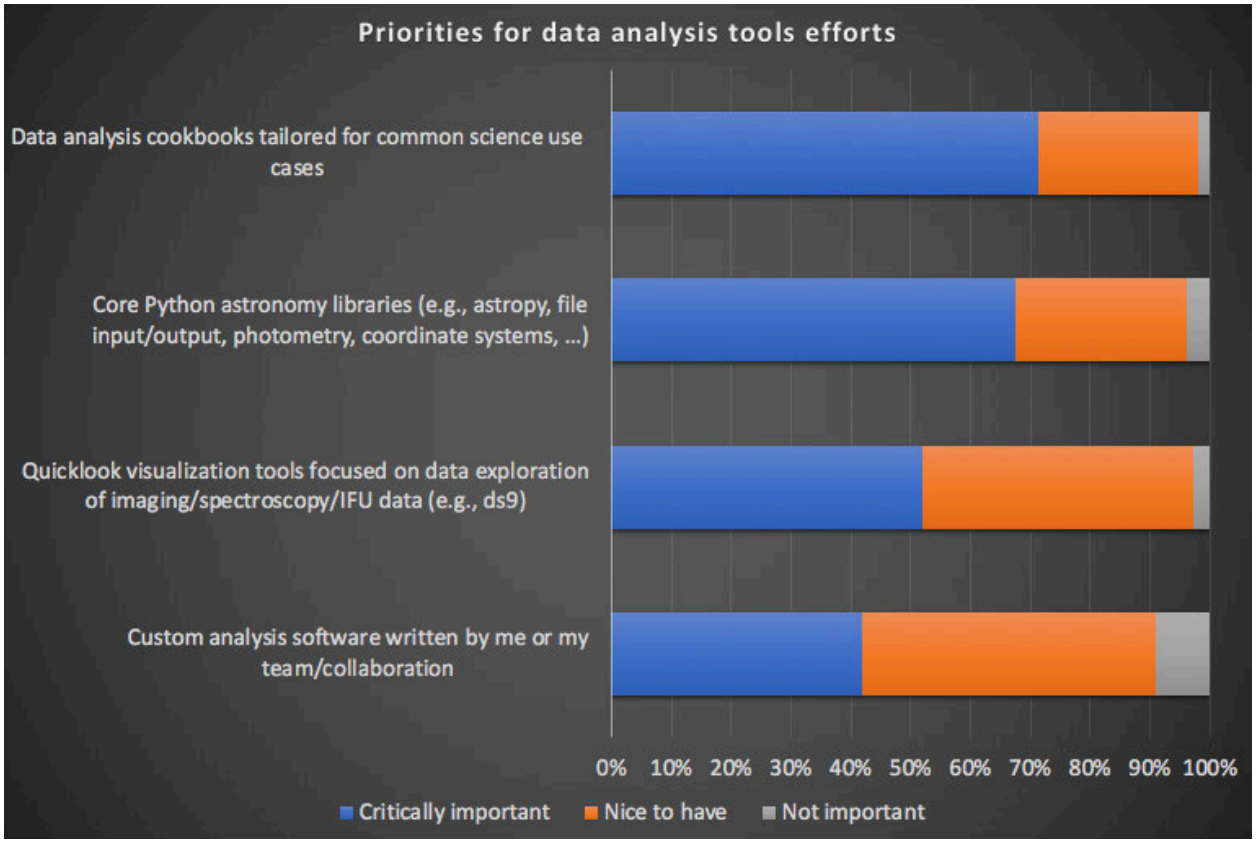
Expectations for highly processed data



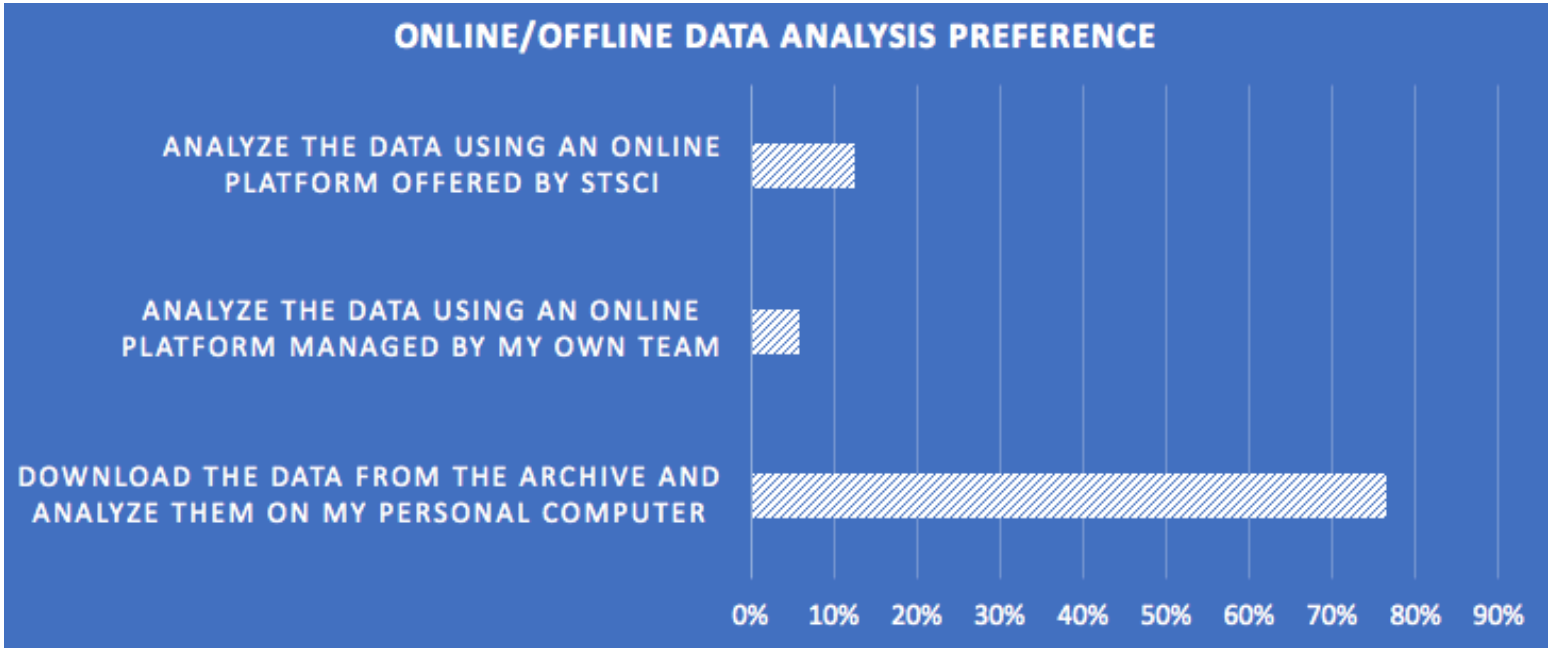
Observers do not have an expectation for delivery of highly calibrated data products. >70% with an opinion expect to do their own data reduction, either from scratch or from individual exposures.

Priorities for data analysis tools development

- Users want data analysis cookbooks – likely to better understand the pipeline.
- Users want core libraries
- Visualization is still critically important for about 50% of users, if slightly lower priority than cookbooks and core libraries.

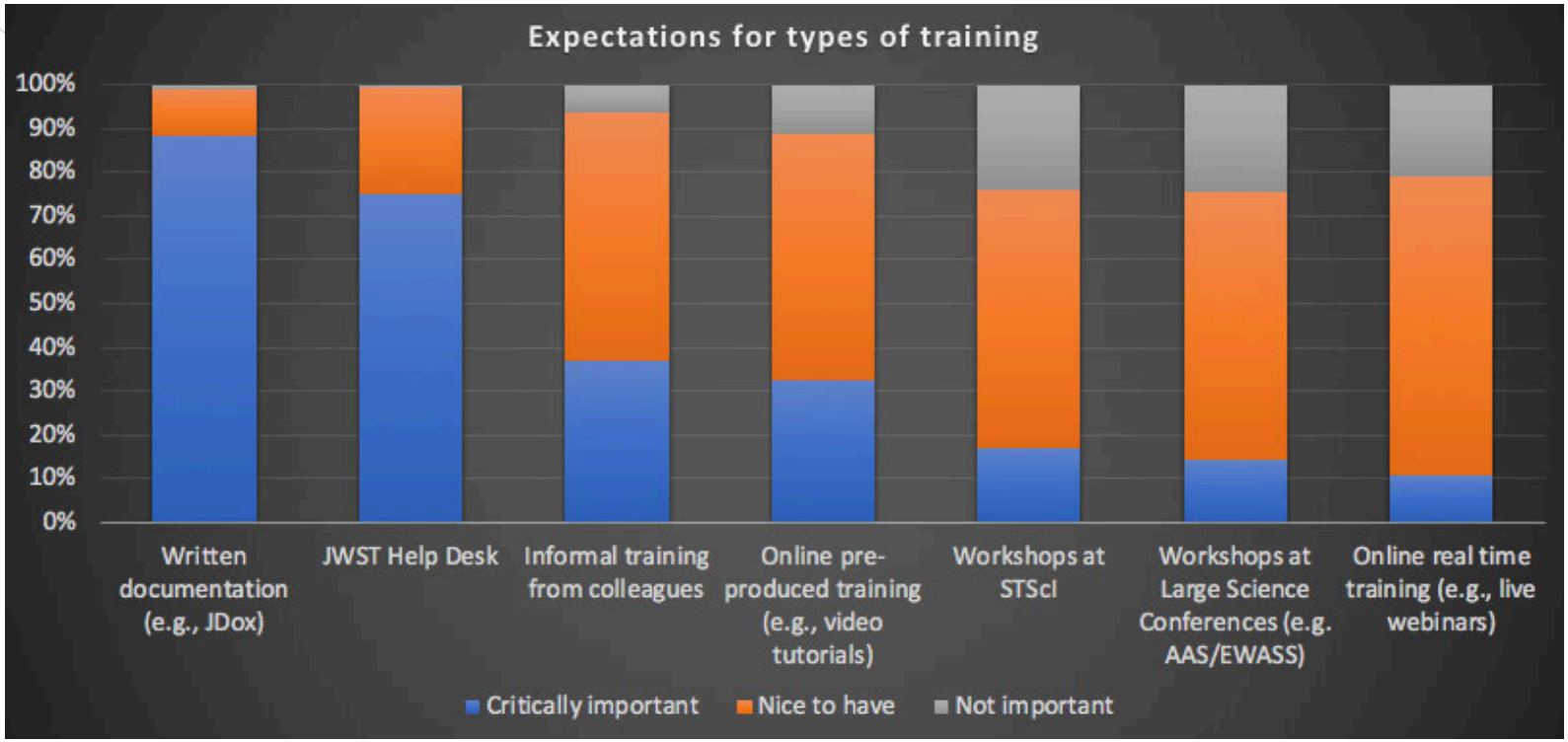


Expectations for online analysis



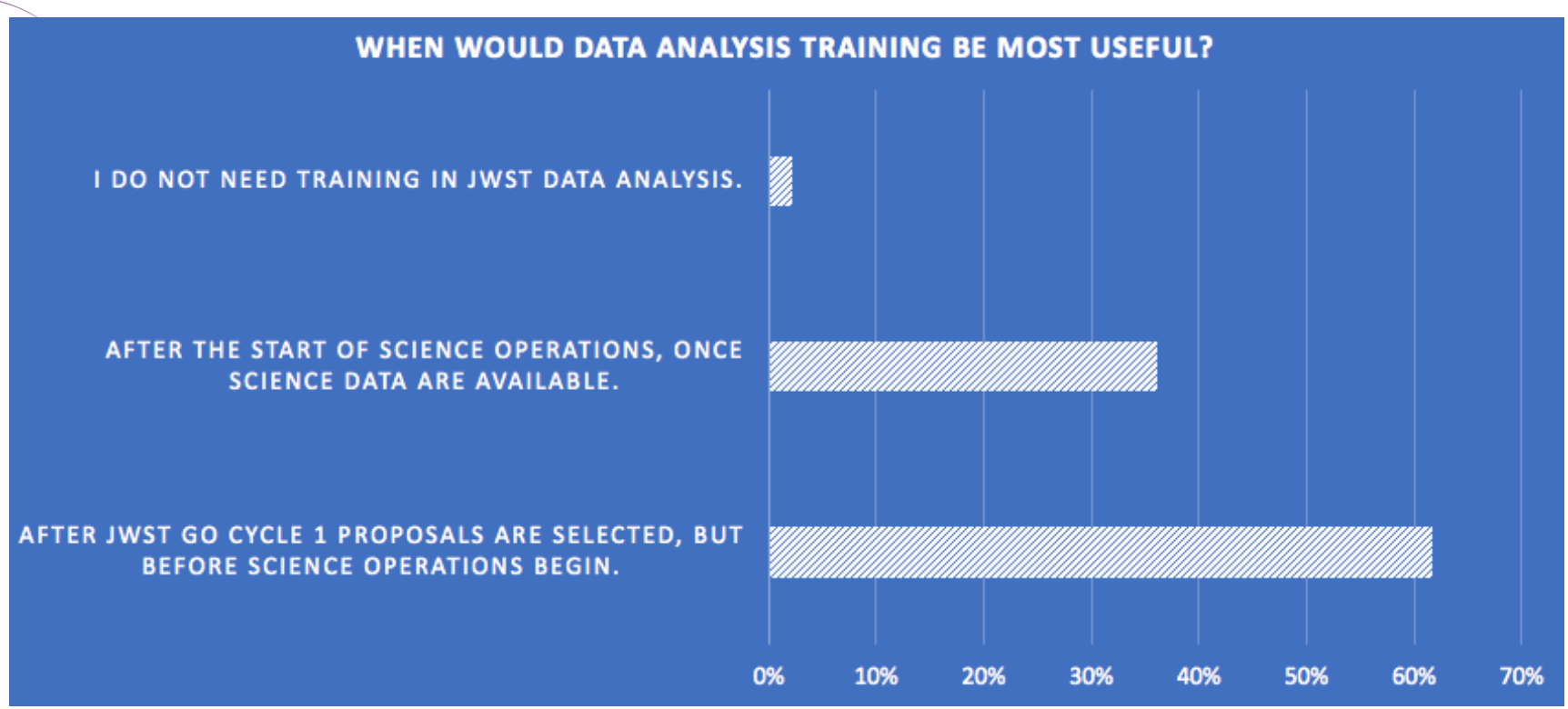
- Users overwhelmingly prefer to download their data and analyze them locally.
- Selling an online solution requires a major communication effort.

Types of training



- Written documentation and just-in-time resources is much more important than real time workshops.
- 18.4% thinks a workshop at either ST or elsewhere is critically important.
- Help desk assistance is considered critically important by >70% of users.

Preferred timing for data analysis training



- Note that the previous slide indicates a strong preference for just-in-time training resources (i.e., not workshops).
- This could be interpreted as an early need for documentation (including Jdox, videos, cookbooks,…).

Selected free form comments (Tools)

- *Many of us will want to continue to use the already existing data analysis tools for IFU cubes (e.g QFitsView, DS9 etc) rather than learn a new one.*
- *It would be good to have a roadmap on visualisation tools (DS9 and QFITSview equivalents). This is something that astropy does not yet do well (need fast UIs).*
- ***You are going to find that working with something like multi-object spectroscopy, IFU data, or high-resolution imaging, where the data are complex and it's helpful to look at the actual pixels, is going to be extremely clunky in a notebook style format.***
- *STScI must prepare to support users in low-level interactions with the data, understanding that in the early phase of the mission, canned recipes will often not provide usable results.*
- *I think [question] 9 above is very important.* ***I see no reason nowadays why one can not run reduction and analysis software online to process and obtain the final data products.***
- *Would [there] be some link with laboratory database infrastructure such as SSHADE (solids), VAMDC (gases)?*
- *I work on transiting exoplanets and in the past our community has had to write data reduction pipelines ourselves because we use unusual observing modes and/or we need to carefully control for systematics. It would be great to change this paradigm with JWST.*
- *We need line fitting applications.*
- *I plan on working with graduate students and post-docs who have expertise in Python. It is critically important to have 'non-expert' intro software and a hierarchy of tools [...]* ***The present suite of software is far too 'expert' oriented.***
- ***please invest in common-use functionality (e.g. in astropy)*** *rather than attempt to build complex analysis tools.*
- *The data products provided by the STScI archive must be modular in the sense that all steps of the calibration must be available and each module must be clearly explained.*
- *The most important aspect of JWST data analysis are* ***having updated tools that do the *basics*.*** *For example, can JWST data be visualized with DS9? If not, then tools like Ginga need to be completed.*
- ***The entire Jypiter notebook Python environment that STScI now uses is a mystery to me****. Installation is too complex, I've never gotten it to work*
- *Line fitting with statistical analysis and error propagation from variances an absolute requirement (little exists within Python yet!).*

Selected free form comments (Online platform)

- *[Download] offers more flexibility on what to do and check the data. Sharing the data and having multiple collaborators to check them is important but that's an easy task, that does not require a specific space or tool.*
- ***I can only imagine the bottle neck in accessing STScI computers to analyze "hot" data.***
- *Online platform offered by STScI sounds interesting but I don't understand how well this platform will support development of custom data analysis tools.*
- ***Often have to work on airplanes**, etc, and having the data with me makes this possible.*
- *[STScI platform] would be the ideal, but in practice such services are rarely (if ever) useable. Hence [download] is give the user the most control and confidence in the results.*
- ***Traditionally I do everything thing on my own** (super)computer but I am willing to try other collaborative models.*
- ***I prefer to download to my own computer** as I worry about being able to configure a remote/online platform to my specific needs (custom and/or personal code libraries, etc).*
- *Since I do not have much experience with Python and don't have it set up on my computer, if STScI offered an online platform with documentation/cookbook, I would consider using it and it may be very helpful.*
- ***In Australia, latency for online would be too high.** We could consider a local mirror, this is something I could potentially help you with as I have considerable JWST funding (Karl Glazebrook)*
- ***JWST products should be small,** downloading and analysing them at home would be the most flexible option*
- *Historically I've downloaded & analyzed data, but with larger teams and new classes of software, the online platform approach, which we are using for some of our ground-based data, is the way for me to go in the future.*
- ***Already have lot of legacy codes from Spitzer, SOFIA for moving target analysis that do not rely on Python.***

Selected free form comments (Policy)

- *Access to non-ITAR-controlled spacecraft ephemeris, telemetry and housekeeping data can be important for science and research.*
- *As important as learning how to handle the JWST pipelines is funding support, e.g. a small amount of funding is important for prep work for accepted proposals such that data reduction work can start as soon as data are available.*
- ***Don't forget about moving objects.*** *I've definitely had to do significant work at other observatories to get the data products that I need.*

Selected free form comments (Python vs IRAF/IDL)

- *I'm a beginner in Python but expert in too many other programming languages to care that much.*
- ***Don't assume that python will still be what's cool in 10 years.*** *We need access to a simple arithmetic system to handle basic data needs; adding images, median filtering, etc. This is what IRAF provided for 40 years.*
- ***Keeping iraf/pyraf alive into the JWST era is critically important****: the python modules are insufficient, and the interactive tools are woefully inadequate. JWST should NOT be used as an excuse to force the community to deal with the limitations imposed by astropy.*
- *I use IDL. I have translated IDL to python and the code worked. However, the more senior a person is (i.e me) the less cost effective it is for said person to code.*
- *As I am a beginner at* ***python, it seems that it is an immature data analysis tool at present****. There is a vital need for a modern iraf/pyraf package.*
- *Personally* ***I am no fan of python****, whose merits are that it is free, but the downside is that even with Anaconda, frequently there are conflicts between modules or missing modules […] I find the use of python notebooks somewhat clunky.*
- ***Python is slow, not backwards compatible, and horrible to maintain****. […] We don't all work at places where IT manages multiple virtual environments. […] And no, I am not ignorant, out of touch, or whatever other epithet you want to throw my way. […] You get about what you'd expect from github's latest buggy releases maintained by unpaid volunteers. Going forward, Jupyter is for Julia, python, and R.*
- *Make sure JWST data reduction and analysis remains easy to use at small universities. We don't have python everywhere. So it is mandatory to keep IRAF as a backup.*
- ***Language agnostic tools would be grately preferred,*** *i.e. a library of programs or routines with bindings for scripting languages such as python being offered as add-ons.*
- ***There needs to be a uniform basic sofware package that does the trivial things*** *- add images, subtract images, caculates statistics of mean counts etc. This is what IRAF did, and the utility of IRAF was not that it was a great language, but it offered a basic tool set*

Selected free form comments (Documentation)

- ***Extremely disappointed in STScI's ongoing failure to explain the planned OV and calibration program activities to the community.***
- *Data analysis tools are only as useful as the documentation.* ***Replicate Spitzer level documentation****, not Kepler level.*
- *Having something similar to Stack Overflow (stackoverflow.com) would be extremely helpful for supplementing the documentation.*
- ***Instructional videos/JDocs on common tasks are very useful.*** *e.g. downloading a raw IFU data set, running the calibration pipeline and explaining each input, displaying and extracting final spectra, etc.*
- *I sincerely hope that the institute will consider putting out static IHB/DHB documents starting with Cycle 1. […]* ***the organization of JDox is hopelessly tangled****. […] my comments come after years of using JDox.*
- *In documentation, please try not to skip over explaining concepts by saying things like "...which is just like how it works for HST" or the like. Not everyone is familiar with the previous facilities.*
- *For a quick reference, I would appreciate a compact version along with the extensive data manual. Given my experience with space telescopes (mostly HST and SPITZER), going through the user/data manuals for trivial information is excruciating.*
- ***A cookbook reduction tutorial is absolutely critical to have****. I am worried that there is too much "just find it on github" form of data reduction software.*
- *It is important that written documentation be kept up-to-date. Much of the current astrodrizzle documentation is woefully out-of-date and therefore wrong and confusing.*
- *Cookbooks that are well-written and annotated. And a CENTRAL PLACE where all these are linked.*
- *Documentation with a TOC in a searchable pdf is critical.*
- ***Detailed written documentation (both "cookbooks" and more detailed manuals) is by far the most important thing to have.*** *In person training or workshops are nice in the short term, but it's hard to remember things from them*
- ***Workshops always seem to reduce to the "lowest common denominator,"*** *and spend time teaching the least familiar people in the room how to do things.*

Selected free form comments (Training)

- *You should have trained local experts around the country in major institutions.*
- *Workshops and documentation that are specific (it is not useful to work with documentation that is only looking at distant galaxies when you are looking at a bright planet).*
- *We need data analysis training now, if we're going to write our own software to analyze JWST data in real time as it becomes available.*
- *You must consider the way that workshops can lead to geographical advantages to proposers. Many of the astronomers in the US now are geographically distributed and cannot readily attend the workshops.*
- *I think that JWSTMO is sending confusing messages about who is responsible for preparing the community for Webb. The JWST MasterClass effort for proposal preparation is effectively farming out the job of training the community to others, via a train-the-trainer approach. In contrast, this data analysis training effort seems to be led from within STScI. I hope the effort is not outsourced later.*
- *I lack funding to travel for training at STSci or conferences; I also prefer to minimize travel for environmental reasons.*
- *A Help Desk that is patient and happy to deal with students who are learning and not just professionals.*
- *Workshops were people can work through an analysis workflow with experts in the room are critical. Webinars are OK for those who are not able to attend a workshop, but not nearly as effective.*
- *if an analysis tool requires a video tutorial, it is too complicated IMO.*
- *If you provide video tutorials, please provide transcripts of the tutorials and time stamps for when different topics begin.*
- *Regional JWST workshops! I live in Brazil, far from the centers where these workshops are typically organized at (i.e., Europe, US)...*

Summary notes

- Strong indications
  - A majority of the community still uses non-Python platforms for at least some of their data analysis. A significant minority relies on them. Younger researchers are more likely to use Python.
  - Spectroscopy and imaging are dominant modes, and the need for spectroscopy tools very strong. High contrast imaging is used by a smaller fraction of the community. This is consistent with other indicators.
  - Familiarity with the JWST pipeline is very low, as expected at this phase of the mission.
  - Training is most needed for the JWST pipeline and data products, not for general data analysis.
  - There is little expectation for high-level data products. The community is not used to high-quality data products immediately after launch and anticipates starting from the exposure level data products (stage 2). Effective communication is needed to increase confidence in high-level JWST data products.
  - There is an overwhelming expectation to download JWST data and analyze them locally.
  - There is a strong preference for documentation, help desk, and just-in-time training resources over in-person workshops.
  - Comments indicate a strong preference for supporting basic data analysis tools and documentation over complex tools and novel platforms.
- Less strong indications
  - There is a small, but significant preference for prioritizing cookbooks and core astropy functionality over quicklook visualization and data exploration tools. This distinction seems stronger in the free form comments. Half the community still considers data exploration tools critical.
  - There is a significant preference for training between the GO1 deadline and science ops, compared to after the start of science ops. The strong preference for just-in-time training indicates that pre-science operations training should de-emphasize workshops.
  - There is significant concerns about STScI “needlessly” requiring the community to spend extra time learning new languages and tools when the most pressing need is understanding the specifics of JWST data.