Galaxy Small-Scale-Admin February 2024 Poll Results

Table of Contents

I About the document

I.A How to read the document/how it was made

A) Overall, I didn't change the question names nor the data
B) But I did change a few values when I saw they were incoherent (ex : an answer of « 0 » that was meant to be « -1 » (no answer) judging by the comment which clearly said « I don't know »). (However I did not systematically do this, I assume the answers to be correct).
C) I also renamed and re-ordered a few questions for better clarity
D) I removed a line entirely (« Comments » associated to end-user support) because it had 0 answers
E) Next to each graph that had a « Comment » section, I put the (value/comments) combinations for only the values that had comments.
F) The report includes Hans' answer from Feb. 20
G) (Most of the graphs were generated by a R script I wrote)

I.B Personal info :

A) I removed the names/emails before analysing
B) I deleted the answers from the Framaforms website (after analysing them) since they contained names/emails.

I.C Limitations

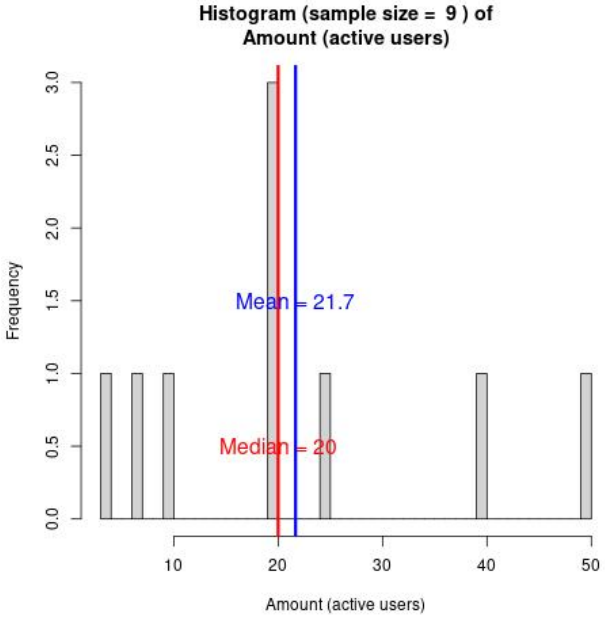
A) « Computing » and « Object Store » should have been multiple-choice
B) The « end-user support » question and comment was vague
    - Accidentally called it end-user assistance
        - But the numerical «avg. nb hours of end-user support per month » still got seemingly valid answers
    - The associated« Comment » field was misleading (it said « Comment (tool packaging) » because of a copy/paste mistake)
        - It got no answers because of this
C) For the « Other recurrent admin. Tasks » question, I should have specified monthly/yearly.
    - As such, the apparent outliers might not actually be if divided by 12
    - I was intentionally vague since the question was meant to be open-ended, but I should've still been more precise.
D) Some arbitrariness in the time-length of some questions (week/month/year)
E) Some arbitrariness in some questions' location
    - « Do you use Ansible ? » could've been grouped together with « Do you group Gravity »
    - « Nb users » could have been in the « Users » section
F) No doubt some missing questions
G) Some potential overlap (eg: "end-user support" and "other monthly admin.")
H) Overall feeling that, question-wise, some sections were too detailed while others not enough, but I don't have enough Galaxy expertise to really fine-tune this.

I.D Interpretation of the results

The results should be specific and detailed enough to be self-explanatory.

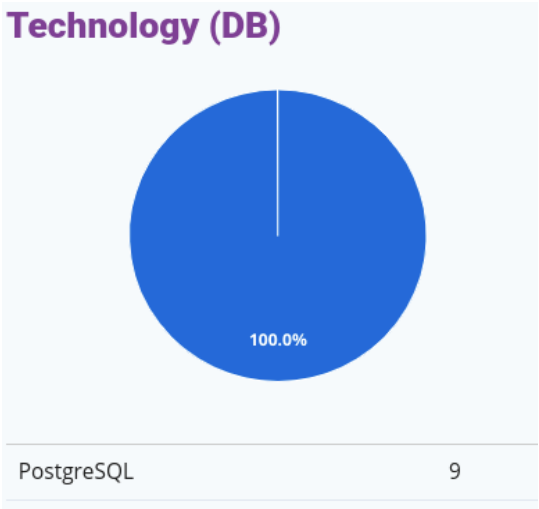
II Galaxy instance characteristics

II.A Active users



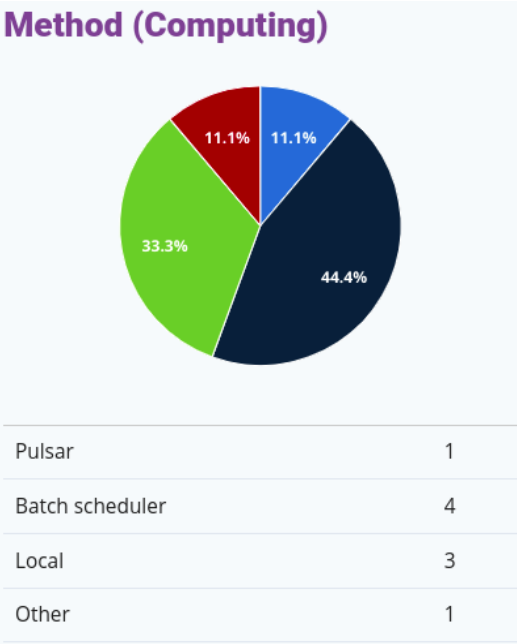
Amount (active users)	20	20	50	25	10
Comment (active users)	This number varies a lot, since there is a constant turnover at our institute...and not everybody who has attended the internal training sessions keeps on using Galaxy	Total 200	they're all students, generally they start/stop using the galaxy in batches of 20-30	There are about 5 times as many registered users as active users	total 84

II.B DB



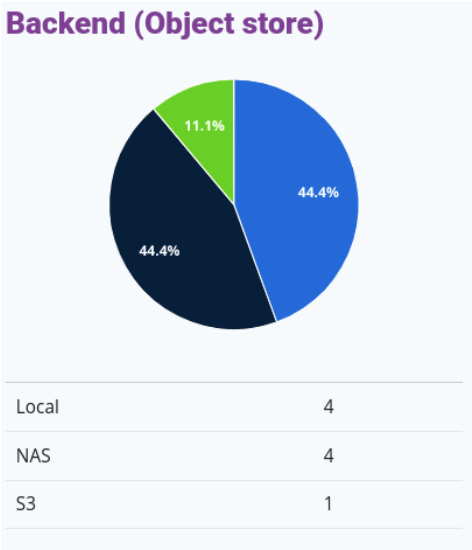
Technology (DB)	PostgreSQL
Comment (DB Technology)	Original, our server was run by a MySQL database, I have managed to do the transition (without loosing data) to PostgreSQL (see: https://galaxyproject.org/blog/2015-07-mysql-2-postgresql/)

II.C Computing



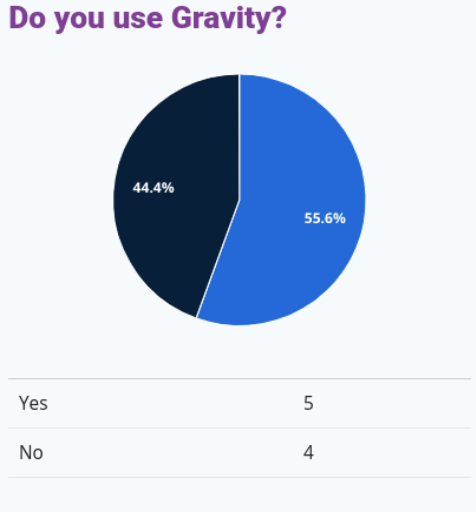
Method (Computing)	Local	Batch scheduler	Local	Batch scheduler	Pulsar	Batch scheduler	Other
Comment (Computing)	everything runs on a 28cores (double threaded) box	SLURM	sbatch scheduler on the same machine	Local HTCondor Deployment	HTCondor again, in addition to pulsar	Jobs run on a compute cluster provided by the university's central research IT team using a local custom job runner based around shared folders, to separate Galaxy from the cluster	mixture of custom job runner deploying to external resources and local

II.D Object store



Backend (Object store)	S3	NAS
Comment (Object store backend)	I mounted a s3 bucket locally that is accessible through the data libraries.	Wish this was a multi-select, NAS + local files :)

II.E Gravity



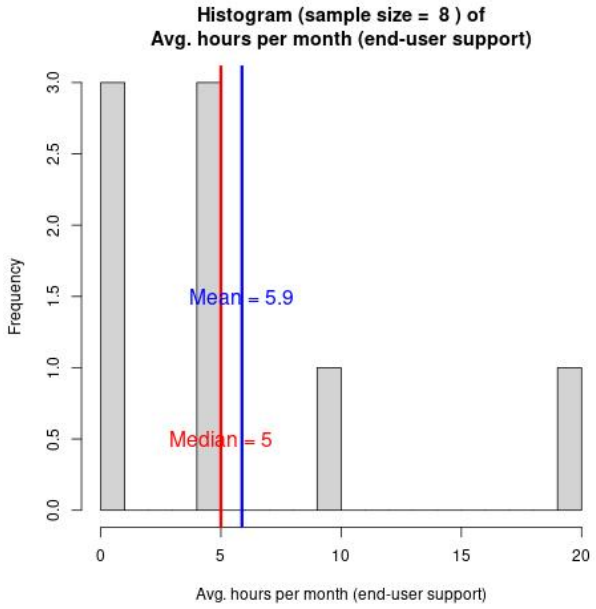
Do you use Gravity?	Yes	Yes	No
Details/ Comment (Gravity)	Yes but not directly, via ansible.	indirectly via ansible	Not yet upgraded to a Galaxy version that requires use of Gravity

II.F Other comments

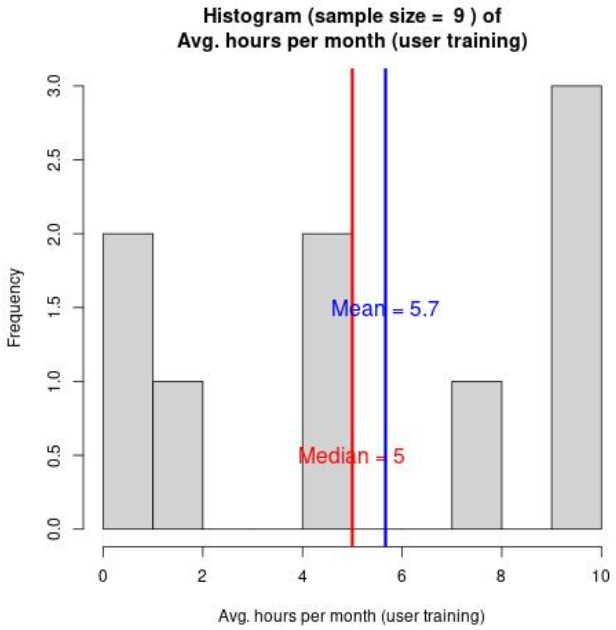
Other comments (cluster characteristics)	Hardware: 16 cores (32 CPUs), 384GB RAM, 16TB storage	Mostly single-node machines with some shared storage.	it was 3 nodes, one test, one prod, and one spare that ran other services + shared computational load

III Users - various

III.A End-user support



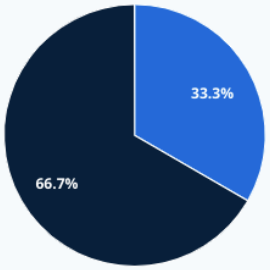
III.B User training



On average, hours per month (user training)	5	10	10	0	5
Comment (user training)	I give 2-3 courses (half day) a year using the GTN material.	This is the majority of my time. They do a tutorial, but then they need to do something different or special and I am the only one who can help debug workflow issues they encounter	So much student training required and support questions	No user training offered	A rarely do training classes. Preferred method is direct user assistance whenever help is required.

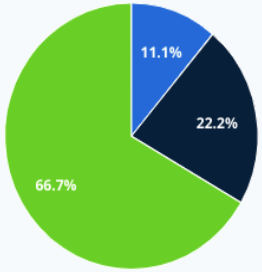
III.C User tool development

Are there any users developing tools themselves locally?



Yes	3
No	6

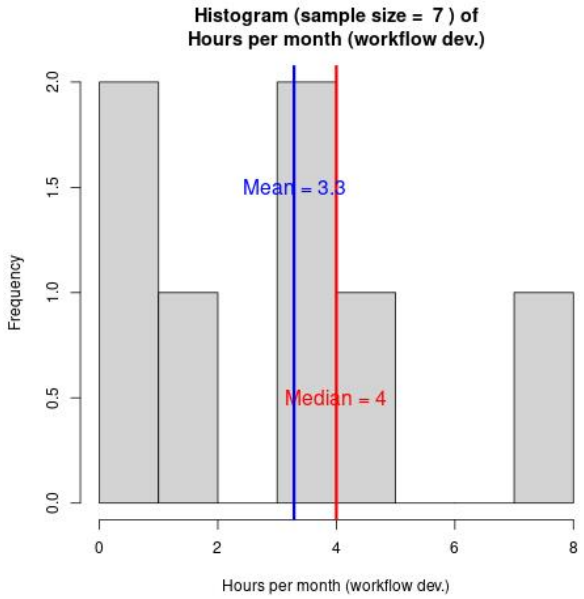
If so, do they use Planemo?



Yes	1
No	2
No answer	6

IV Admin - developement

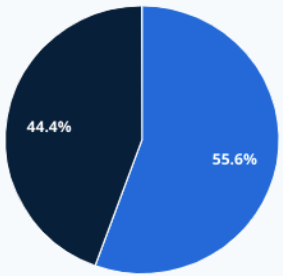
IV.A Admin workflow development



Hours per month (workflow dev.)	2
Comment (workflow dev.)	Very variable it can be 0 for months and then days...

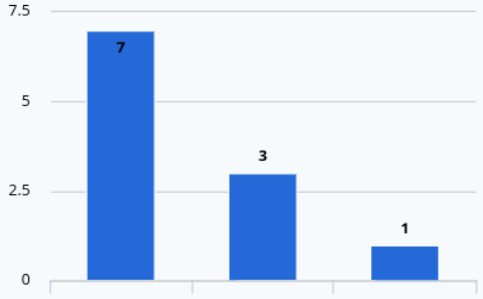
IV.B Admin tool development

Do you use Planemo?

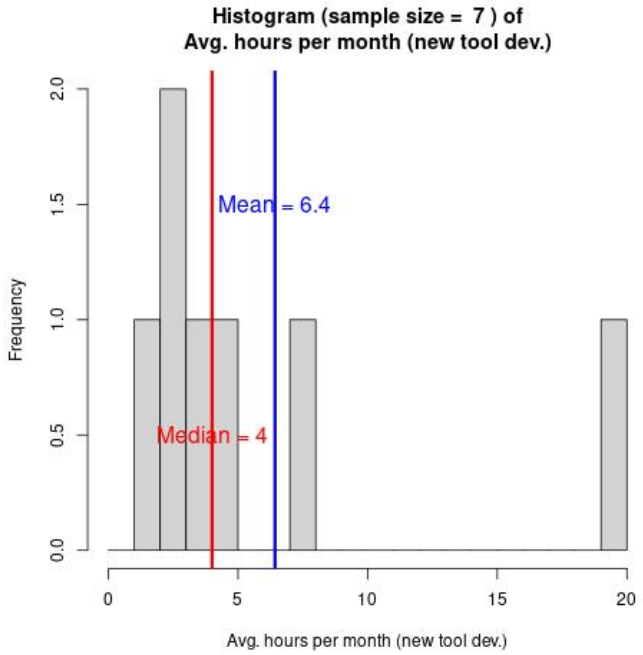


Yes	5
No	4

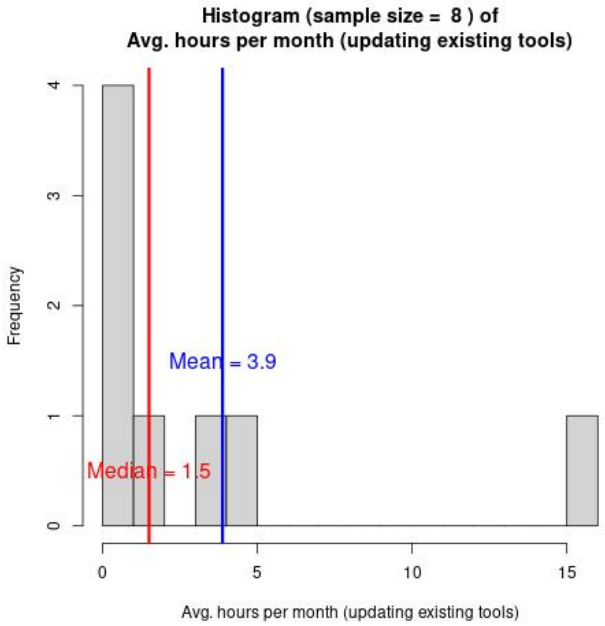
Tools Storage



Local Filesystem	7
Public Mercurial toolshed	3
No answer	1



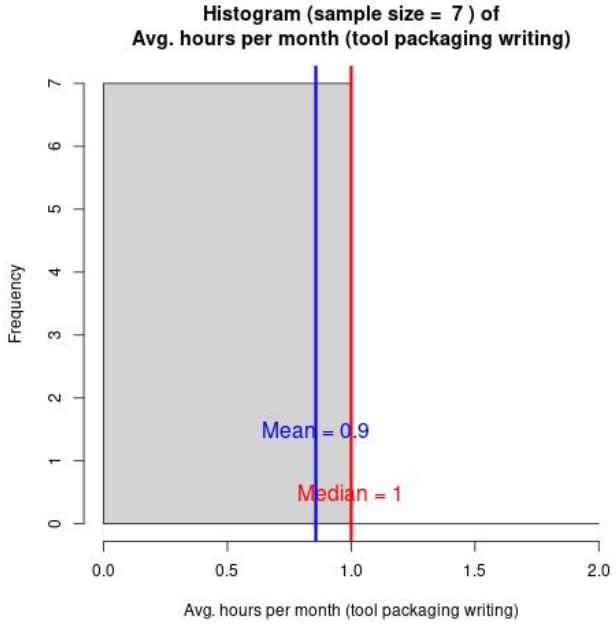
On average, hours per month (new tool dev.)	20	3
Comment (new tool dev.)	I do not understand the question on Tools Storage? I only use tools via the public toolshed.	Very variable it can be 0 for months and then days...



On average, hours per month (updating existing tools)	1	1
Cause/Comment (tool updates)	Could be 0 if I would automate it.	It's more likely to be 12 hours once a year than 1 hour a month every month for a year

IV.C Tool packaging

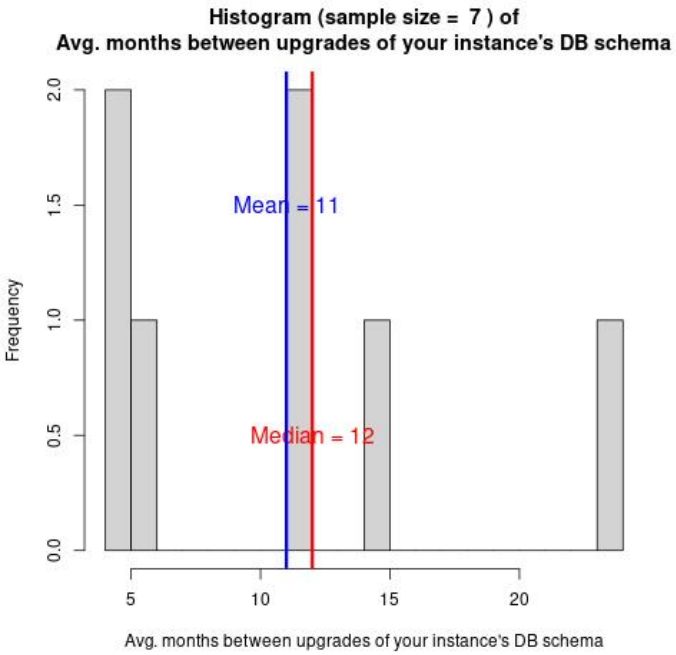
Tool Packaging Method	Conda + Container
Comment (tool packaging method)	Conda only historic. I try to get rid of it.

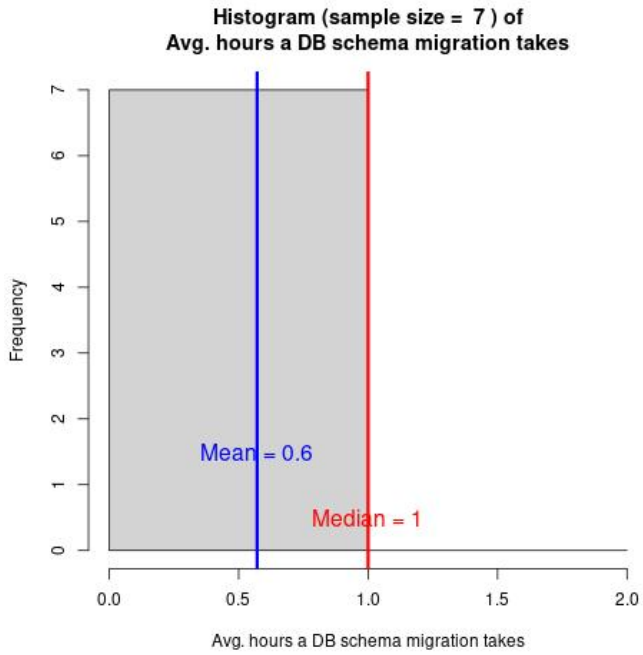


Avg. hours per month (tool packaging writing (conda recipes/container config/...))	1	1
Comment (tool packaging cost)	We mostly used already packaged software	Should actually be "<1"

V Admin – Galaxy upgrades

V.A DB schema migration

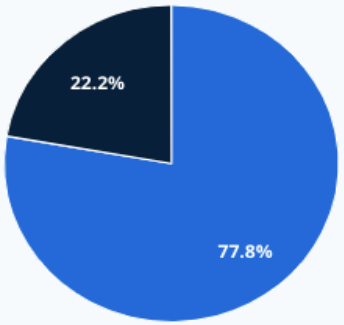




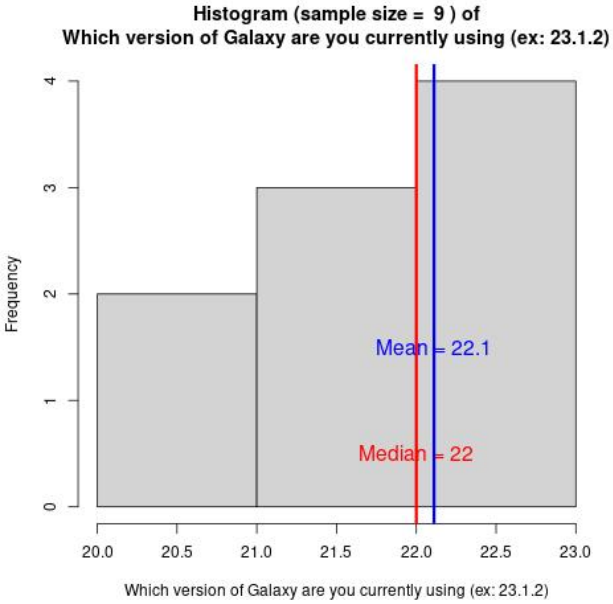
Avg. months between upgrades of your instance's DB schema	15	6	-1	4	4	24	-1
Avg. hours DB schema migration takes	0	0	-1	1	1	1	-1
Comment (DB Migration)	upgrading the database schema has not been a problem (except very long time ago) and happens within minutes (or even seconds)	I usually update to the previous version when the next release is published. DB migration is completely automatized .. don't think that admins need to worry about this.	The database migration is usually linked to the upgrade, I cannot dissociate.	It's very fast thanks to ansible. Upgrades of schemas only happen during galaxy releases which are like, kinda sorta 3x per year.	it's fine, thanks ansible	Actually <1 (done as part of Galaxy version upgrades)	can't tell about migration time, sorry. This job is done by a sys admin

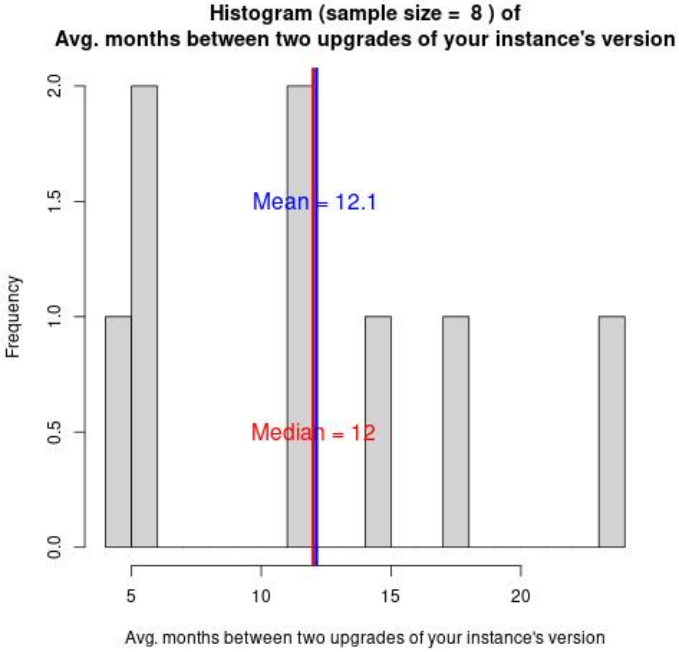
V.B Galaxy version upgrades

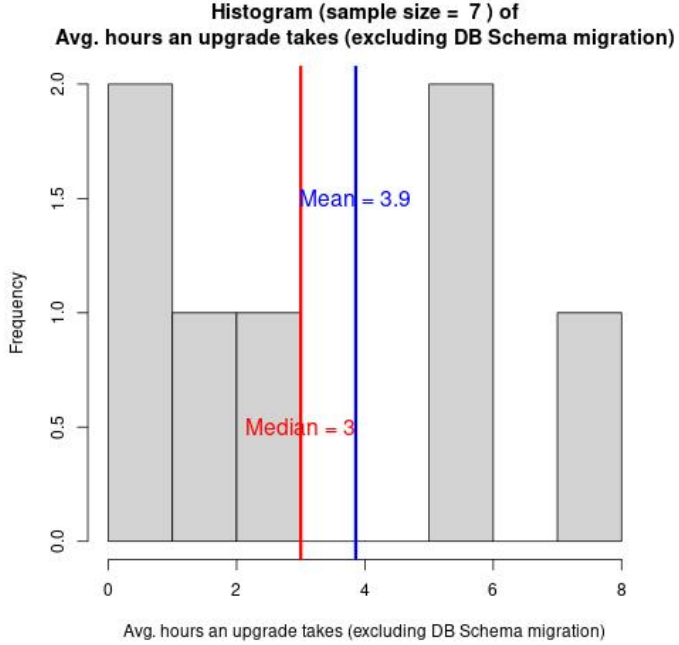
Ansible



I use it	7
I don't use it	2



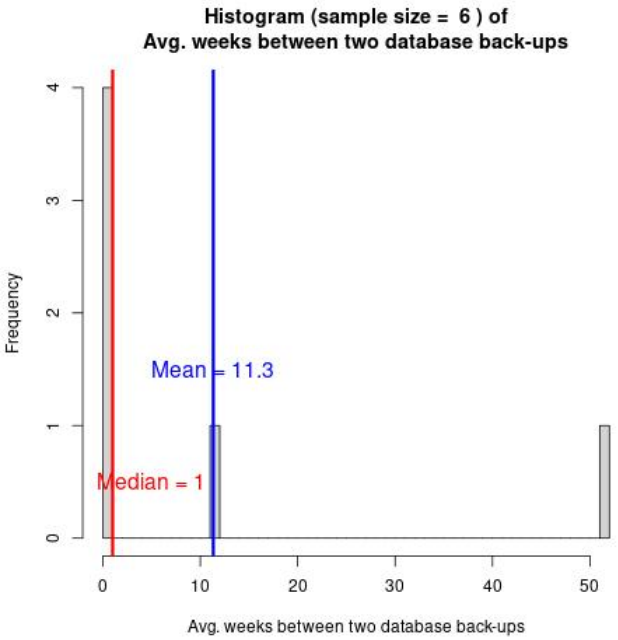




What are your motivations for an upgrade?	- new features - trying to be up-to-date	Get the newest features and bug fixes. Some tools will only run with recent versions. The effort to upgrade is usually really small when not skipping releases..	I would like to have all new features / new datatypes into my instance. This is my motivation to upgrade. From 20.09 to 22.01 everything went smoothly I could upgrade in few minutes, simply by modifying the galaxy_commit_id in the playbook. But as 22.05 has a lot of changes I need to test the upgrade on a VM before running it for real. This requires a lot of time that I could not dedicate yet.	bugfixes mostly, occasionally new features.	we try not to change it during the school year, so need to get new features + bugfixes during the summer break.	Re motivations: keeping Galaxy current (easier to get help with newer version), some tool versions not available for older Galaxy versions), get security and bug fixes, get new features for users that are available on public instances Re average time for upgrade: this is the time taken to perform an upgrade and includes notifying users etc. Time to prepare for upgrade (e.g. testing on Vagrant & test instance) can be several days or longer depending on changes introduced in the target Galaxy release.	can't tell about upgrade time as well. This job is done by a sys admin	Be up to date	better to do it voluntarily than be forced at some point as the latter will be more painful; access to new features

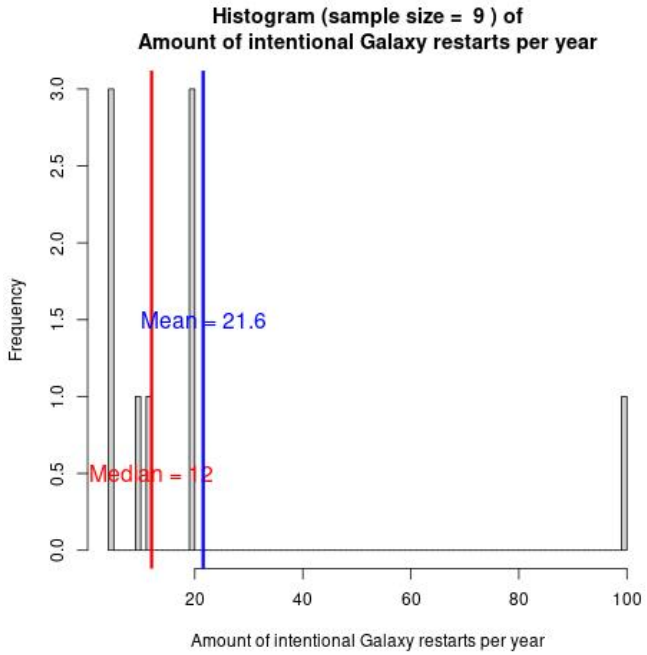
VI Admin – other recurrent tasks

VI.A DB back-up



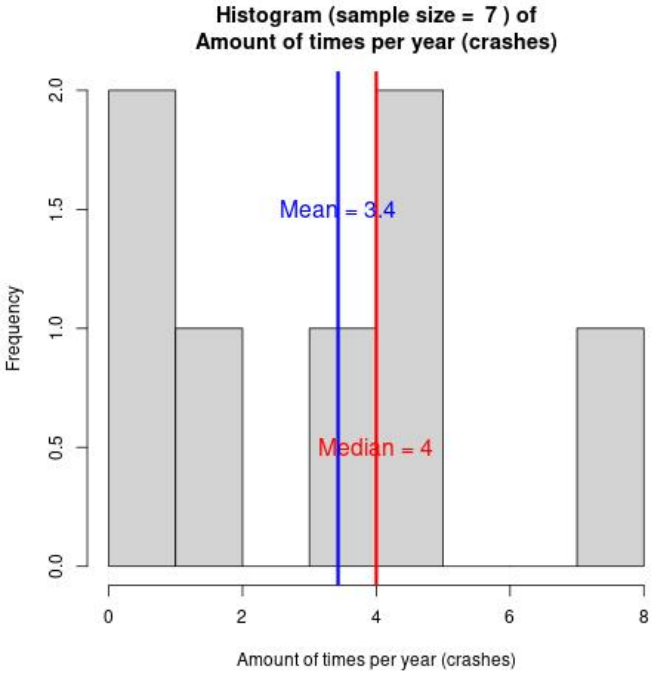
On average, weeks between two database back-ups	1	-1	1	1	1	12	-1	52
Method/ Comment (DB back-ups)	I do daily backups (run as cronjob) using 'pg_dump'	Not sure. Told my IT dept to do backups and never checked :)	The DB back-ups are automatically in the ansible playbook I am running. The cron job is set to run every week.	ansible automatically installed database backups are what we use.	ansible default	Dump SQL to flat file and gzip. Generally only done prior to upgrades	I can't tell exactly but I would expect that it is routinously as the other datasets on file system on daily basis.	rsync (yes, I know I need to do it more often)

VI.B Intentional Galaxy process restarts (for maintenance purposes)



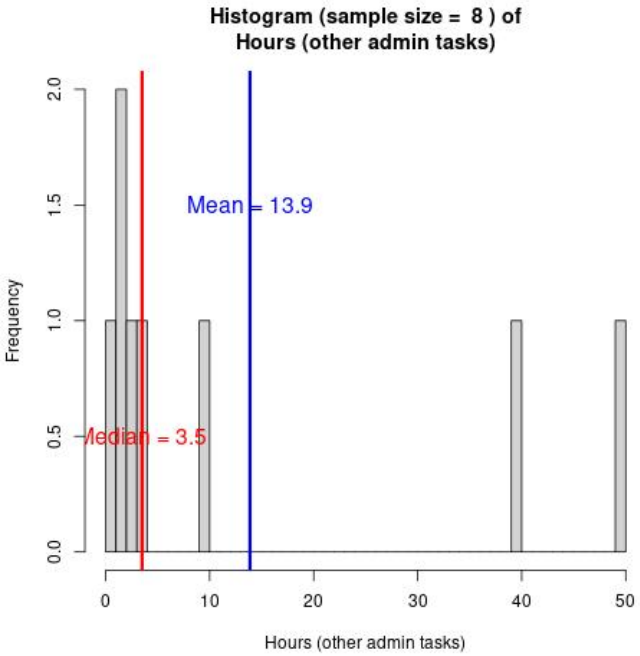
Amount of intentional Galaxy restarts per year	12	4	4	10	20	20	4
Reason/Comment (restarts)	- change in tool version (not coming from toolshed) - adjustment to Galaxy code (e.g.: tool filtering) - re-ordering of tools - adding new reference data (not covered by Data Manager)	We have maintenance windows for our HPC every 3 month which I use for upgrades. I sometimes restart also unplanned which is no problem with a few dozen users (I just announce on short notice).	Server reboot for updates .	mostly for config tweaks or tools that didn't get picked up	config changes, testing out new features, etc.	Generally done in response to addressing issues, don't normally restart otherwise	actually we have far more restarts (e.g. due to constant issues with the tus service for upload)

VI.C Crashes



Amount of times per year (crashes)	8	2	4	5
Indentified cause/Comment (crashes)	Usually caused by problems (file system) of the HPC	It was always storage issues. Then postgres fails to write and locks. Then Galaxy gets in a restart loop and generates logs filling whatever space was left.	Usually either disks full up or NFS problems	TUS upload service failing, storage issues, users overloading the ressources (BLAST search with 500 MB sequence as query)

VI.D Other recurrent admin tasks (not including tool dev, user assistance, upgrades, DB-backups, …)



Hours (other admin tasks)	10	3	2	4	50	2
Description/ Comment (other admin tasks)	-interacting with our sysadmins who are responsible for storage and system software of our server - following up on tool failure - cleaning up 'paused' Galaxy jobs - testing new galaxy versions on our test and development servers	Working on automatization on .. networking with other admins	autom ations	automation again (e.g. spent trying to set up automated "your account is too full" emails.)	I think over the year this might be a conservative estimate, tasks include: - Managing user quotas - Adding new users - Adding/updating tools from toolshed - Investigating issues reported by users (tool failures, running out of space)	Not sure whether this is included in user assistance already: setting up/updating Galaxys institutes custome welcome page with tutorials, overview of plant reference genome ressources (all non-standard), setting up specialised workflows for users.