Technical questions: Where to ask and how to get answers

Bari J. Ballew, PhD

September 16, 2019

- Have you ever had a question regarding bioinformatics/data science?
- Where did you go for answers?
- Did you get a reply?

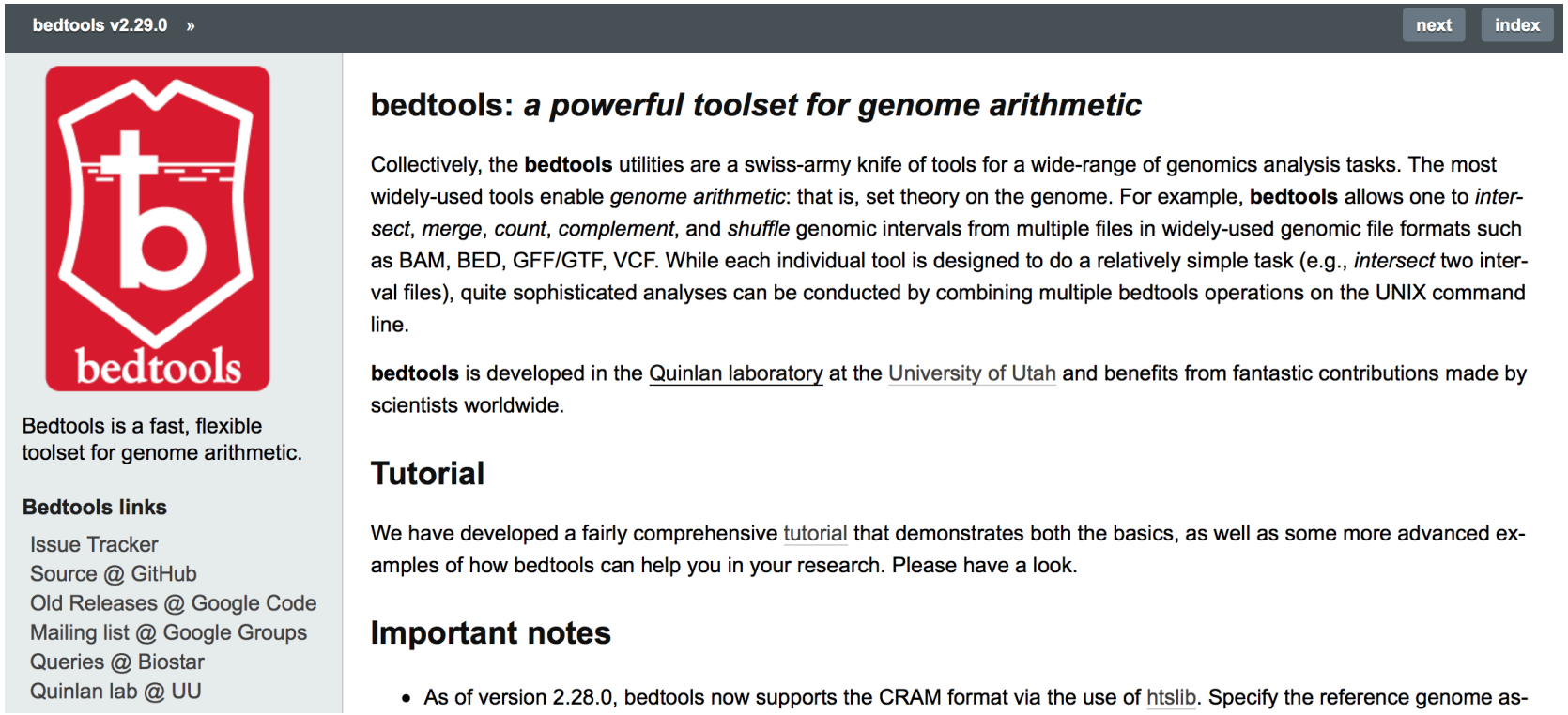
- This talk will cover:
  1. Where/whom to ask
  2. How to ask

Where/whom can I ask?

- **Intramural resoures:** many groups at NIH that provide support, including (*not an exhaustive list!*):
  - CBIIT
  - NIH Library
  - NIH HPC Group
  - Additional resources listed here
  - NIH bioinformatics slack channel: https://nih-byob.slack.com
- **Non-interactive resources**
- **Interactive resources**

Non-interactive resources

- Manual/readme/documentation



bedtools v2.29.0 »

next index

Bedtools is a fast, flexible toolset for genome arithmetic.

Bedtools links

Issue Tracker
Source @ GitHub
Old Releases @ Google Code
Mailing list @ Google Groups
Queries @ Biostar
Quinlan lab @ UU

bedtools: *a powerful toolset for genome arithmetic*

Collectively, the **bedtools** utilities are a swiss-army knife of tools for a wide-range of genomics analysis tasks. The most widely-used tools enable *genome arithmetic*: that is, set theory on the genome. For example, **bedtools** allows one to *intersect*, *merge*, *count*, *complement*, and *shuffle* genomic intervals from multiple files in widely-used genomic file formats such as BAM, BED, GFF/GTF, VCF. While each individual tool is designed to do a relatively simple task (e.g., *intersect* two interval files), quite sophisticated analyses can be conducted by combining multiple bedtools operations on the UNIX command line.

bedtools is developed in the Quinlan laboratory at the University of Utah and benefits from fantastic contributions made by scientists worldwide.

Tutorial

We have developed a fairly comprehensive tutorial that demonstrates both the basics, as well as some more advanced examples of how bedtools can help you in your research. Please have a look.

Important notes

- As of version 2.28.0, bedtools now supports the CRAM format via the use of htslib. Specify the reference genome as-

Non-interactive resources

- Manual/readme/documentation
- FAQs/vignettes/tutorials

Introduction to dplyr

When working with data you must:

- Figure out what you want to do.
- Describe those tasks in the form of a computer program.
- Execute the program.

The dplyr package makes these steps fast and easy:

- By constraining your options, it helps you think about your data manipulation challenges.
- It provides simple "verbs", functions that correspond to the most common data manipulation tasks, to help you translate your thoughts into code.
- It uses efficient backends, so you spend less time waiting for the computer.

This document introduces you to dplyr's basic set of tools, and shows you how to apply them to data frames. dplyr also supports databases via the dbplyr package, once you've installed, read `vignette("dbplyr")` to learn more.

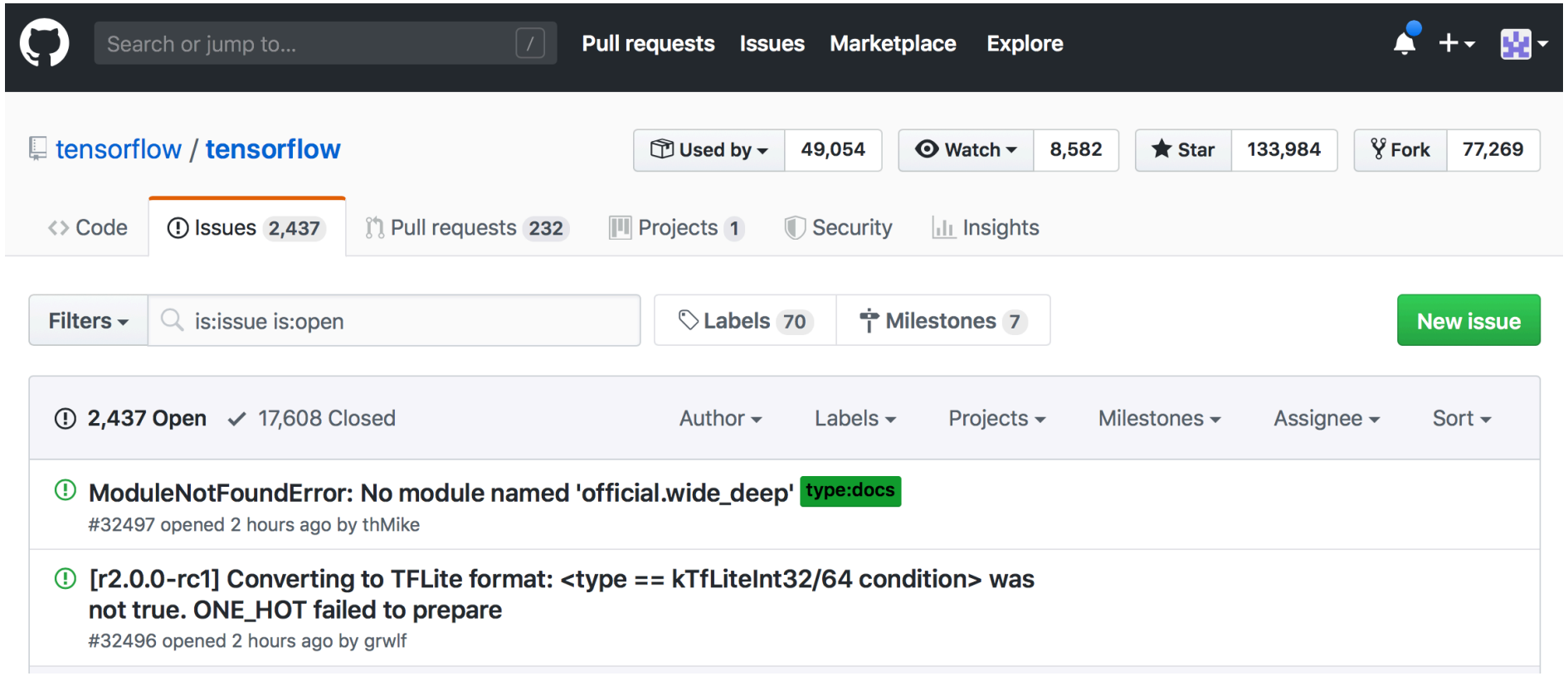
Data: nycflights13

To explore the basic data manipulation verbs of dplyr, we'll use `nycflights13::flights`. This dataset contains all 336776 flights that departed from New York City in 2013. The data comes from the US Bureau of Transportation Statistics, and is documented in `?nycflights13`

```
library(nycflights13)
dim(flights)
#> [1] 336776     19
flights
#> # A tibble: 336,776 x 19
```

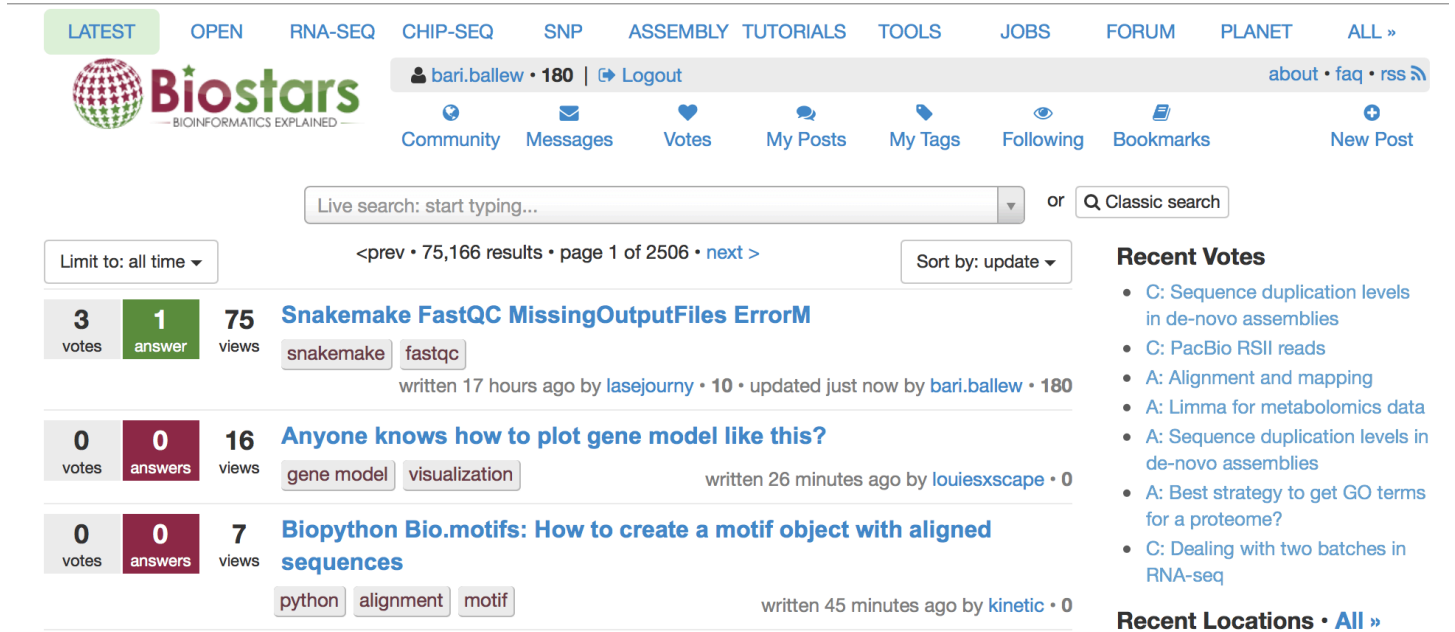
Interactive resources

- GitHub/Bitbucket issues pages



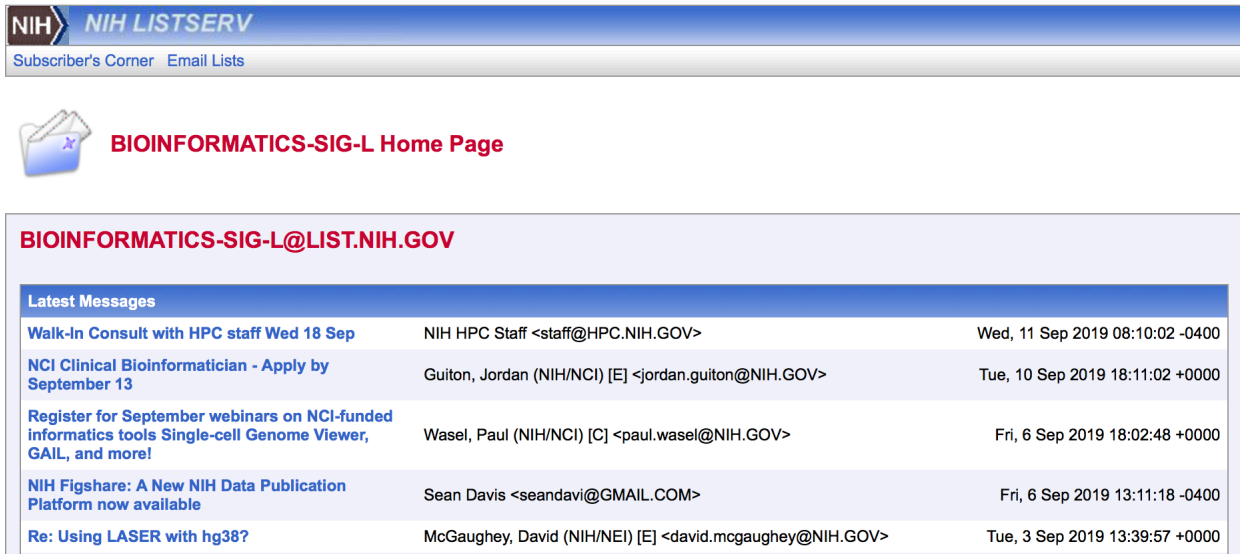
Interactive resources

- GitHub/Bitbucket issues pages
- Forums
  - https://www.biostars.org
  - https://stackoverflow.com



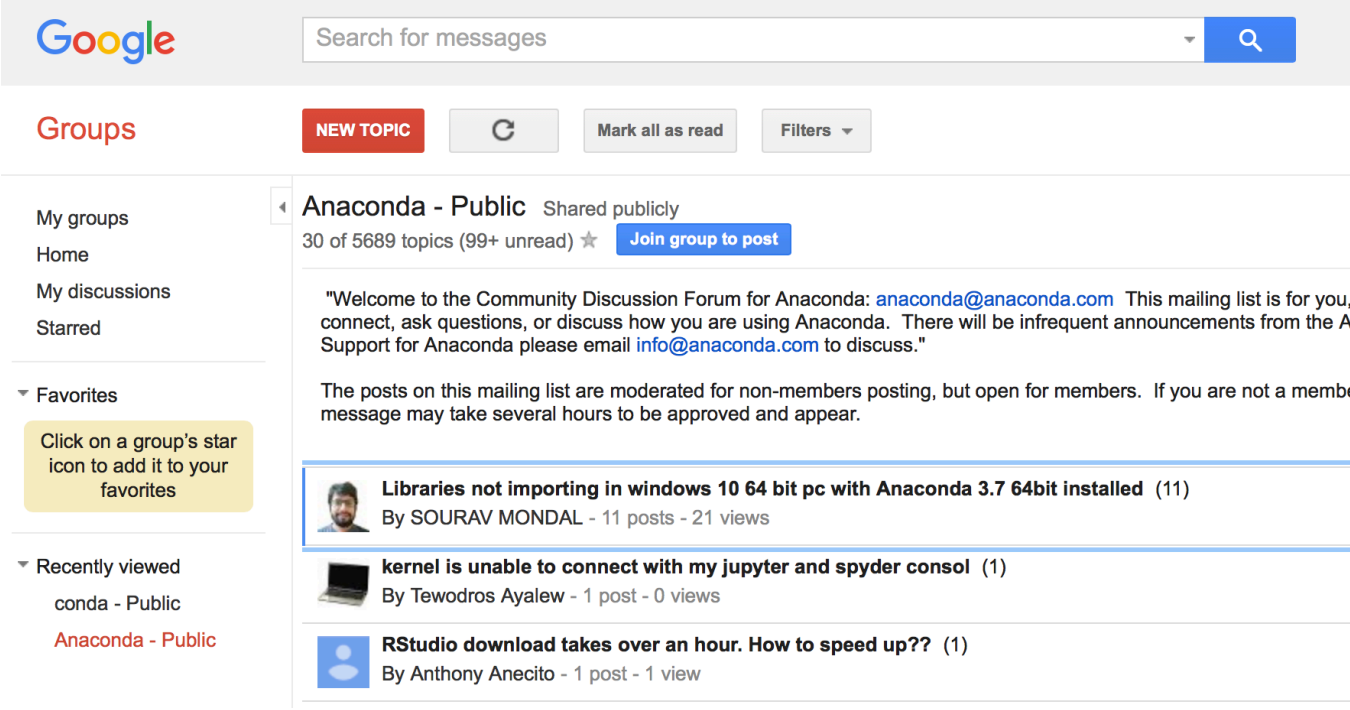
Interactive resources

- GitHub/Bitbucket issues pages
- Forums
  - https://www.biostars.org
  - https://stackoverflow.com
- NIH listserv: https://list.nih.gov
  - BIOINFORMATICS-SIG-L
  - NIH-DATASCIENCE-L



Interactive resources

- GitHub/Bitbucket issues pages
- Forums
  - https://www.biostars.org
  - https://stackoverflow.com
- NIH listserv: https://list.nih.gov
  - BIOINFORMATICS-SIG-L
  - NIH-DATASCIENCE-L
- User groups
  - Specific to tool



How to ask

- Do your due diligence
- Ask a colleague
- Post on a forum

Due diligence

- GOOGLE! Your problem is probably not unique
  - Search all non-interactive resources (manuals, FAQs, etc)
  - Search previous posts to interactive resources (forums, issues, etc)
- Use the information gleaned to do some troubleshooting

Ask a colleague

Original

I tried to run bcftools merge, but got an error. What do I do?

Better

Ask a colleague: Describe the problem

- Your environment (OS, program version, dependencies)
- The exact command/series of commands you ran

Original

I tried to run bcftools merge, but got an error. What do I do?

Better

I tried to run bcftools merge (version 1.9) on biowulf on an interactive node to combine several bgzipped VCF files as follows:

```
bcftools merge -o out.vcf file1.vcf.gz
file2.vc.gzf file3.vcf.gz
```

Ask a colleague: Describe the problem

- Your environment (OS, program, version information)
- The exact command/series of commands you ran
- Your input
- Your output (expected and observed)

Original

I tried to run bcftools merge, but got an error. What do I do?

Better

I tried to run bcftools merge (version 1.9) on biowulf on an interactive node to combine several bgzipped VCF files as follows:

```
bcftools merge -o out.vcf file1.vcf.gz
file2.vc.gzf file3.vcf.gz
```

I got an output file containing the headers and some variant rows, but many fewer variants than were in my input files.

Ask a colleague: Describe the problem

- Your environment (OS, program, version information)
- The exact command/series of commands you ran
- Your input
- Your output (expected and observed)
- Any error messages

Original

I tried to run bcftools merge, but got an error. What do I do?

Better

I tried to run bcftools merge (version 1.9) on biowulf on an interactive node to combine several bgzipped VCF files as follows:

```
bcftools merge -o out.vcf file1.vcf.gz
file2.vc.gzf file3.vcf.gz
```

I got an output file containing the headers and some variant rows, but many fewer variants than were in my input files. I got an error message that says "Error at chr1:123: wrong number of fields in ANN_phred."

Ask a colleague: Describe the problem

- Your environment (OS, program, version information)
- The exact command/series of commands you ran
- Your input
- Your output (expected and observed)
- Any error messages

What resources should I have looked at prior to asking this question?

Original

I tried to run bcftools merge, but got an error. What do I do?

Better

I tried to run bcftools merge (version 1.9) on biowulf on an interactive node to combine several bgzipped VCF files as follows:

```
bcftools merge -o out.vcf file1.vcf.gz
file2.vc.gzf file3.vcf.gz
```

I got an output file containing the headers and some variant rows, but many fewer variants than were in my input files. I got an error message that says "Error at chr1:123: wrong number of fields in ANN_phred."

Ask a colleague: Provide context

- Briefly explaining *why* you're doing something can often yield more helpful answers
  - **Original answer:** The steps to upload the files are…
  - **Better answer:** If you are looking at differential expression, it's best not to look by eye, but to used an established pipeline like xyz.

Original

How do you upload BAMs to the UCSC Genome Browser?

Better

I am trying to identify differently expressed transcripts from my RNA-seq data. I would like to visualize the read abundance in each of my samples in the UCSC Genome Browser to find differences. How do I upload my BAMs to the browser?

Ask a colleague: Show what you've tried

- Describing what you've tried so far to solve your problem makes others more likely to help you
- It also helps others better understand your issue and give more helpful replies

Original

I'm using bedtools v2.28.0 to intersect a BED with a VCF but when I run it as follows it gives me an empty file.

```
bedtools intersect –a x.vcf –b y.bed > outFile
```

Better

I'm using bedtools v2.28.0 to intersect a BED with a VCF but when I run it as follows it gives me an empty file.

```
bedtools intersect –a x.vcf –b y.bed > outFile
```

I've checked that both my input files are not empty, and used vcf-checker to make sure the VCF file is properly formed. I tried running the same command in v2.19. I also tried switching the order (`–a y.bed –b x.vcf`) but my output file is still empty.

Ask a colleague: Be specific

- There are a lot of different ways to do things…and they don't necessarily have the same results.
- The person providing an answer is often going to try to follow your steps and replicate the problem, so they can fix it and tell you how. Help them out!

Original

I installed python and pandas as required by the program.

Better

I am using python 3.7 via biowulf's module system (i.e. `module load python3`). I installed pandas with anaconda (`conda install pandas`).

Post to a forum

- All the earlier things, plus…
  - Title
  - Text formatting (if applicable)
  - Avoid cross-posting
  - Focus on one question per post

Original
Title: Help with bedtools

Better
Title: Bedtools intersect VCF with multiple databases: empty output file

Follow-up

- Try to reply to specific questions that are asked of you (e.g. “Can you post the output of conda list?”)
- If you figure out the solution yourself, consider replying to your own post, for future answer-seekers

Asking a question seems like a lot of work!

- Yes, it is! But, doing the background reading, the troubleshooting, and framing your question correctly will often help you better understand the situation, if not solve it on your own.
- Absent this work, your question may come off as expecting others to do the heavy lifting for you. In this case it is likely your question will not be answered.

Thanks!