Genomics Survey Analysis

Before the course, participants were asked to rank their level of comfort with, or understanding of, a series of topics. Later, following completion of the course, they completed the same exercise.

- `The existence or role of metadata in a sequencing project`

- `The organisation and format of sequence data (any)`

- `Organising a file system for a bioinformatics project`

- `Cloud computing`

- `Connecting to a cloud server such as AWS, Microsoft Azure or Google Cloud Platform`

- `Using  a shell and commands such as ls, pwd, mkdir, cat`

- `Working directories, paths and avigating a file system from the command line`

- `Viewing and changing file permissions in the shell`

- `The search tool grep`

- `Using a text editor like nano to write shell scripts`

- `Assessing read quality of sequence data`

- `Trimming and filtering sequence data`

- `Variant calling`

For the purposes of analysis, each of these topics was simplified as follows. The topics were also categorised by theme.

Original name	Shortened name	Theme
The existence or role of metadata in a sequencing project	metadata	project-org
The organisation and format of sequence data (any)	sequence_data_org_format	project-org
Organising a file system for a bioinformatics project	file_system_org	project-org
Cloud computing	cloud_comp	command-line
Connecting to a cloud server such as AWS, Microsoft Azure or Google Cloud Platform	cloud_server_connection	command-line
Using a shell and commands such as ls, pwd, mkdir, cat	basic_shell_commands	command-line
Working directories, paths and avigating a file system from the command line	file_system	command-line
Viewing and changing file permissions in the shell	file_permissions	command-line
The search tool grep	grep	command-line

Original name	Shortened name	Theme
Using a text editor like nano to write shell scripts	nano	command-line
Assessing read quality of sequence data	read_quality	genomics
Trimming and filtering sequence data	trimming_filtering	genomics
Variant calling	variant_calling	genomics

The participants responded to each topic by choosing one of six statements. For purposes of analysis, these statements were codified as numerical values, with 1 being least comfortable and 6 being most comfortable.

| Understanding Statement	Numerical Value
I don't understand what this involves	1
I would recognise this but don't really use it	2
I've used a bit but am not comfortable with it	3
Fairly comfortable in some aspects	4
Comfortable most of the time	5
Confident	6

The mean rating given to each topic before and after the course can be compared using a dumbbell plot.
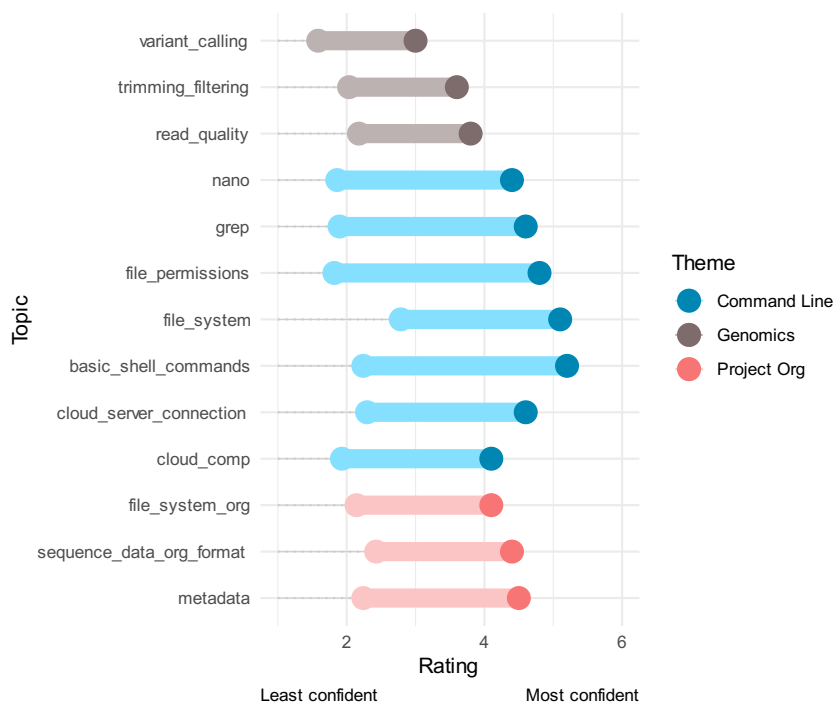
We can see that in general participants felt they improved their confidence in all topics measured, by an average of around two points.

This means most people went from a score of around 2 ("I would recognise this but don't really use it") to 4 ("Fairly comfortable in some aspects") or even 5 ("Comfortable most of the time").

On average, participants felt they improved most at using the command line, followed by project organisation. This is even clearer if we look at the overall trends for each theme.
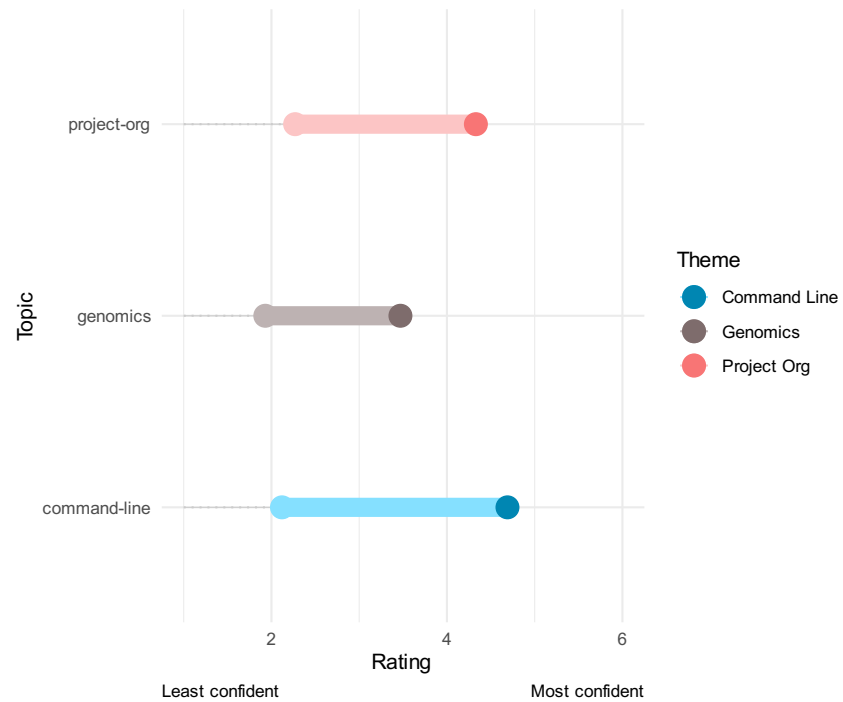
This result is unsurprising given the amount of practice that participants had at using the command line throughout the four day course, and taking into account the rushed nature of the genomics content.

Participants finished the course feeling either fairly or mostly comfortable in topics around the command line and project management, but were not yet comfortable with genomic topics.