**OSD - 665 dataset:**

**1st plot:**

* Total sequences: There is no such a difference between environments, but we can notice that in the GC environment the variation between the samples is higher than other environments.
* Total deduplicated percentage: Appears to be not affected by the environment but we can notice a little bit higher total sequences in the GC environment, it is important to note that a little difference is considered a great one in the context of biology.
* Percent duplicates: It seems like it is not affected by the environment but is differs from one sample to another even in the same environment.

**2nd plot:**

This plot aims to show the relationship between the environment and the categorical columns, where each unique value of each categorical column is presented as the frequency in each environment.

* Per tile sequence quality: This feature has equal in all aspects which indicates no relationship with the environment.
* Per ace sequency content: The same as the previous feature but there is a difference only in the FLT ENVIRONMENT, ALTHOUGH WE CANNOT say that there is a relationship from this little information.
* Per sequence GC content: While the ‘warn’ value dominated all environments, the VIV environment has the most pass samples.
* Adapter content: While the GC environment had the most fails, the VIV environment appeared to have the most pass sample, and the FLT environment is between them.
* Percent GC: The 3 environments have very similar GC percents, but the VIV has the higher percentage, followed by the GC, then the FLT.
* Percentage fails: While the fails has only 4 values across all the dataset, the VIV environment appeared to have ethe lowest failure percentage.

**3rd plot:**

-Shows the mean value in millions for each environment for the same columns in the previous plot, the purpose of this plot is to display the data in simpler, and summarized way through one measure.

* Total sequences: It is very clear that the VIV environment has higher mean value in the context of total sequences, while the GC has lower value, and at the bottom comes the FLT, which lead us to the conclusion that the environment in space may lead to less total duplicated in the RNA.
* Total deduplicated percentage: When it comes to the total deduplicated percentage it seems that the GC environment has higher mean value, while the VIV and FLT are similar, it seems like this feature does not have a direct relationship with the environment which emphasis on the insight derived from the previous plot.
* Percent duplicates: Looks like there is no clear relationships between it and the environment.

**OSD 379 dataset:**

**1st plot:**

* Shows that there is no such a clear relation between the mice type and sequences, deduplicates, or duplicates.

**2nd plot:**

* The quality scores are similar for each environment, but their frequency differ a little bit across all environments
* Across all environments the content is failing more than warn
* GC content pass the most in FLT, and fail the most in BSL
* The adapter content only passed in FLT, and most warn cases occurred in BSL while other environments involved a great portion of fail cases.
* Highest GC percentage exists in BSL while the lowest GC percentage exist in FLT

**3rd plot:**

* There is no big different when it comes to the place and sequences or duplicates but the in the ISS has more sequences, more deduplicates and less duplicates.
* When it comes to the age, older mice seem to have more sequences, deduplicates, and less duplicated compared to young mice.