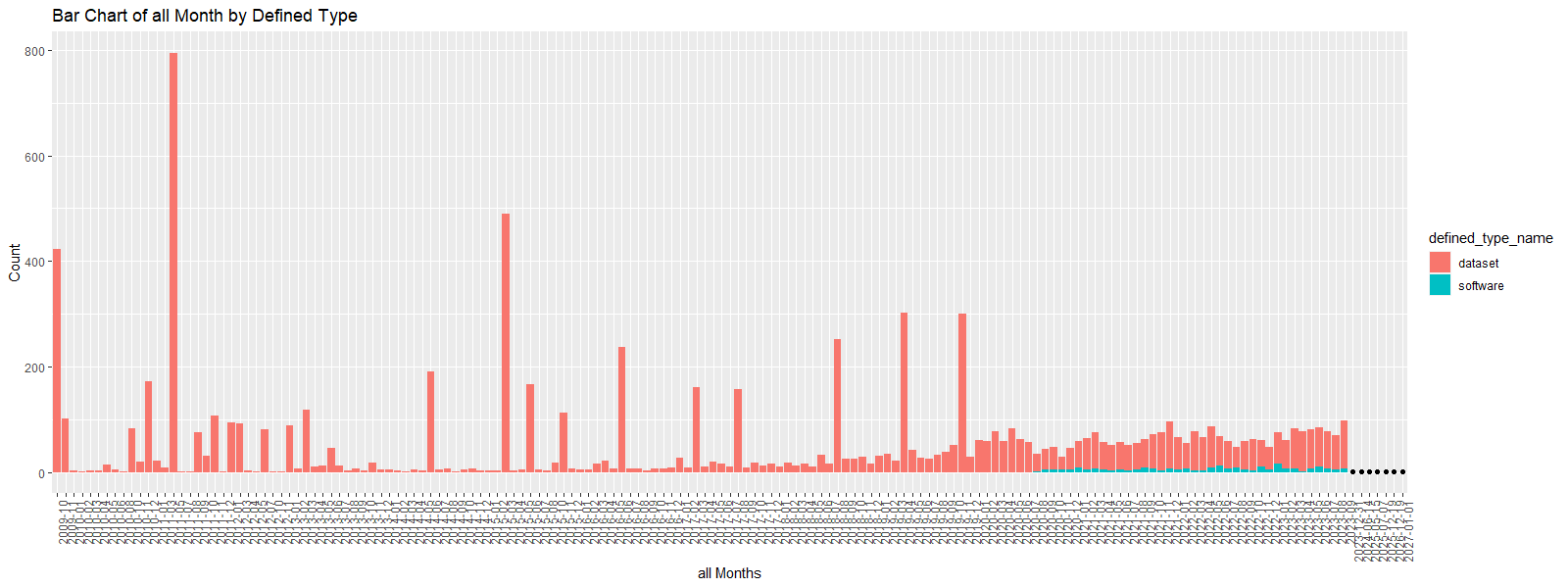
## README

7 plots in total.

## Study of the posted amount

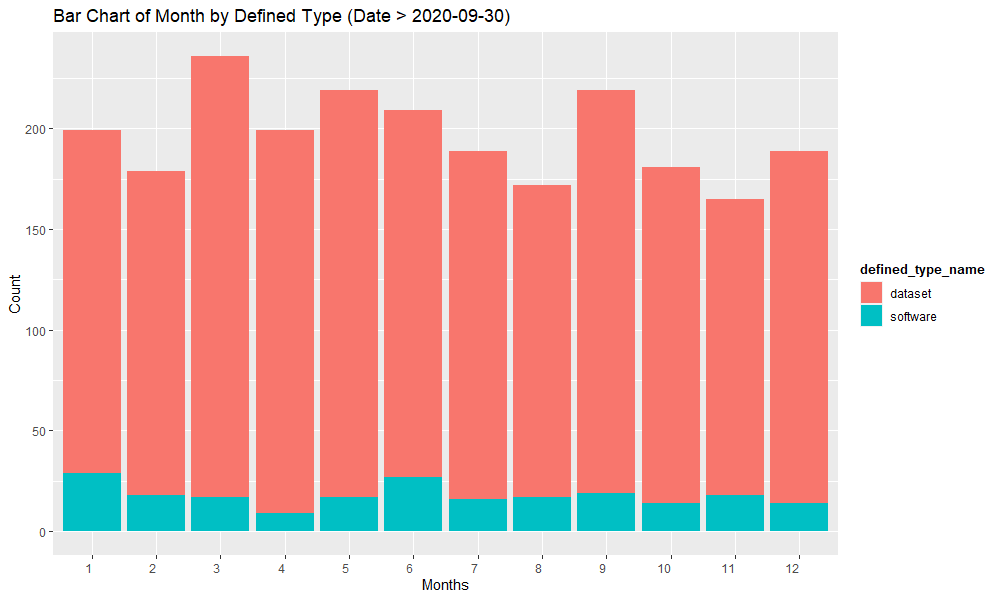
1. *Figure title: Fig1\_yearmonth\_all\_definedtype.png*

We first noticed that there’s 8598 pieces of data in the dataset. Seven of them are still under embargo, and the rest are posted for download. There are two types of posted data, 8371 pieces of “dataset” and 220 pieces of “software”. The counts of the data over the years and months are shown in the following picture. We noticed unregular updates, we suspected that the management of the database is less organized before 2020 than after 2020.



1. *Figure title: Fig2\_year\_compare*

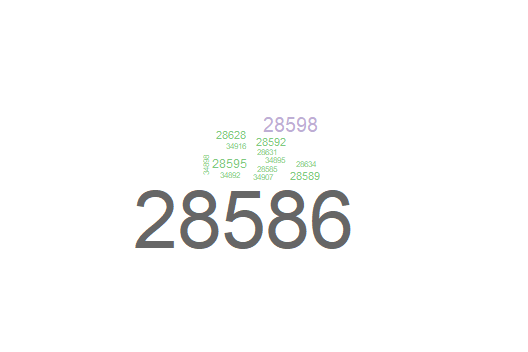
Then, we curious about at which year, the researchers usually uploaded their data. In general, it’s quite even over the year, yet around the end of the year, November till February is less than the upload amounts in the other months.



## Study of Group\_id

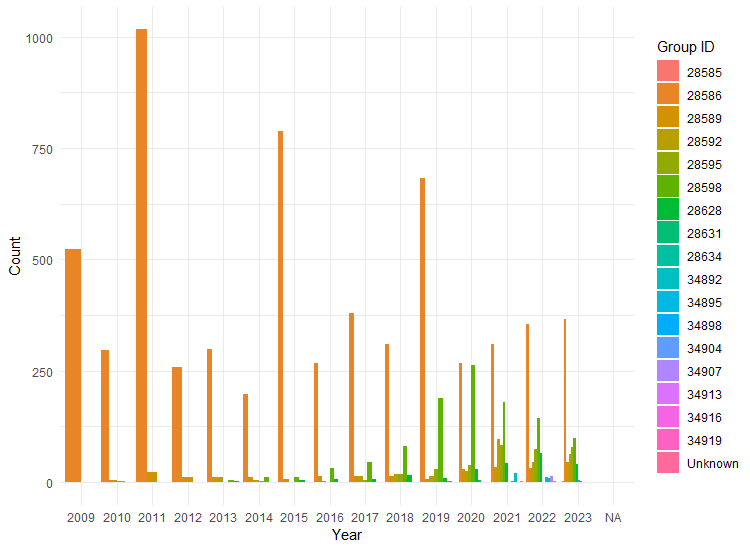
1. *Figure title: Fig3\_worldcloud\_groupid*

There’re 18 groups, and one of them “28586” account for the dominant amount than the others.



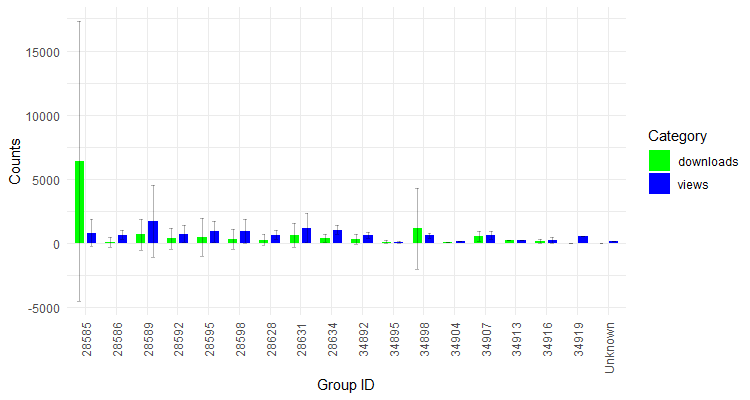
1. *Figure title: Fig4\_year\_groupid\_posted count*

The published amount for each group\_id over the years is shown in the following picture.



1. *Figure title: Fig5\_groupid\_views and downloads*

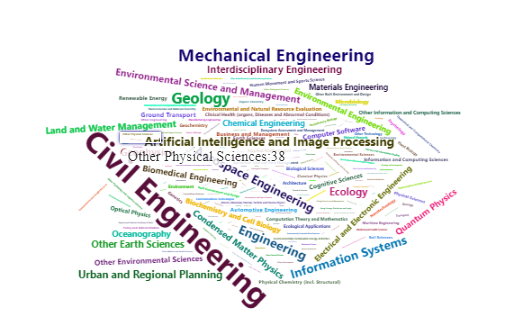
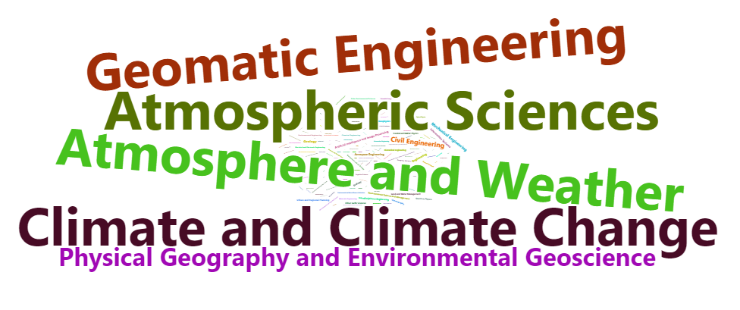
With the information of “views” and “downloads” we also investigated the means and the standard deviation of the “views” and “downloads” of each group\_id.



## Use data category information

1. *Figure title: Fig6\_wordcloud\_category*

The distribution of the data category was investigated by extracting the information via url of the giving information. There are five leading fields who’s posted counts has more than 1000 counts. (The leading fields were excluded in the second figure due to the space limitation for visualization)



1. *Figure title: Fig7\_topviewsdownloads\_category*

Review the mean views and downloads, the results of the leading categories is in the following figure.

