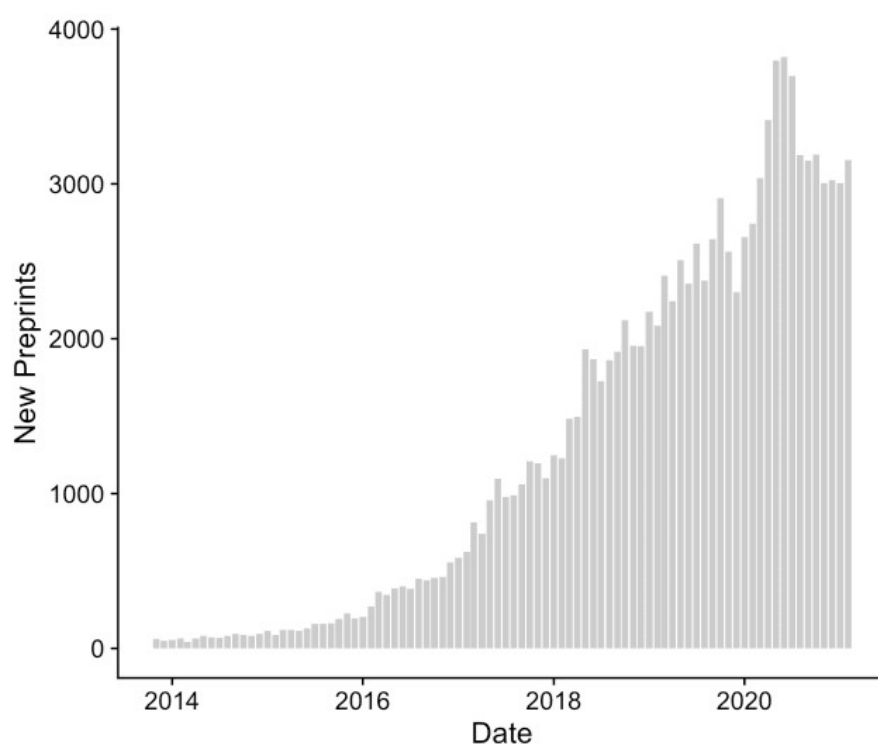


As the COVID-19 pandemic continues, different countries are experiencing various restrictions including lockdowns. Some of these restrictions alter our ability to do science: by hindering lab access or taking time away from researchers for homeschooling. So, what impact has the pandemic had on scientific output?

One way to look at this – for biology – is to look at newly deposited papers on bioRxiv. Which I'll do here using the [rbiorxiv](#) package written by Nicholas Fraser. There are three advantages of looking at preprints as a readout of scientific activity:

- Preprints, rather than journal publications, are immediately released so they give a close to real-time readout of completed scientific activity.
- The number of papers being written, (still) far exceeds the number made available as preprints, so the number deposited on bioRxiv is not limited by anything other than the desire to use the bioRxiv server.
- bioRxiv does not allow secondary literature (review articles), so we are looking at reports of experimental science.



The number of new preprints each month

Pre-pandemic, the trend for preprinting in biology was clearly upward. The graph shows that the rise of first submissions to bioRxiv has increased month-on-month from launch until early 2020. There then appears to be an acceleration in submissions which coincides with the major lockdowns in the USA and Western Europe, followed by a relaxation and possible plateau.

```
# Install package
install.packages("devtools")
devtools::install_github("nicholasfraser/rbiorxiv")
# Load packages
library(rbiorxiv)
library(tidyverse)
library(cowplot)
```

```
df_summary <- biorxiv_summary(interval = "m", format = "df")

p4 <- df_summary %>% mutate(month = as.Date(paste0(month, "-01"),
format = "%Y-%m-%d")) %>%
  ggplot() +
  geom_bar(aes(x = month, y = new_papers), fill = "#cccccc", stat =
"identity") +
  labs(x = "Date", y= "New Preprints") +
  theme_cowplot()
p4
```

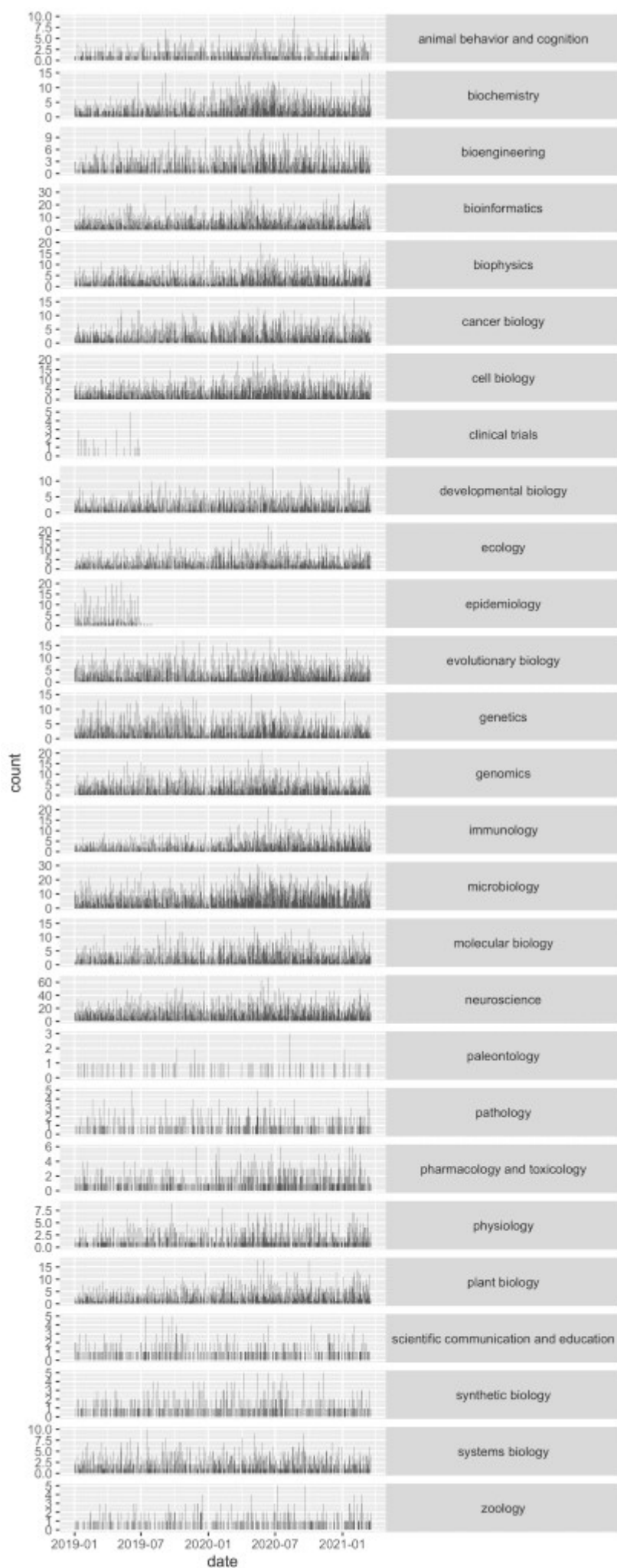
Rise and fall

The rise and fall seen from March 2020 onwards might be explained by:

- a boom coming from people in lockdown taking the opportunity to write up already completed labwork.
- a slowdown due to reduced lab access etc.

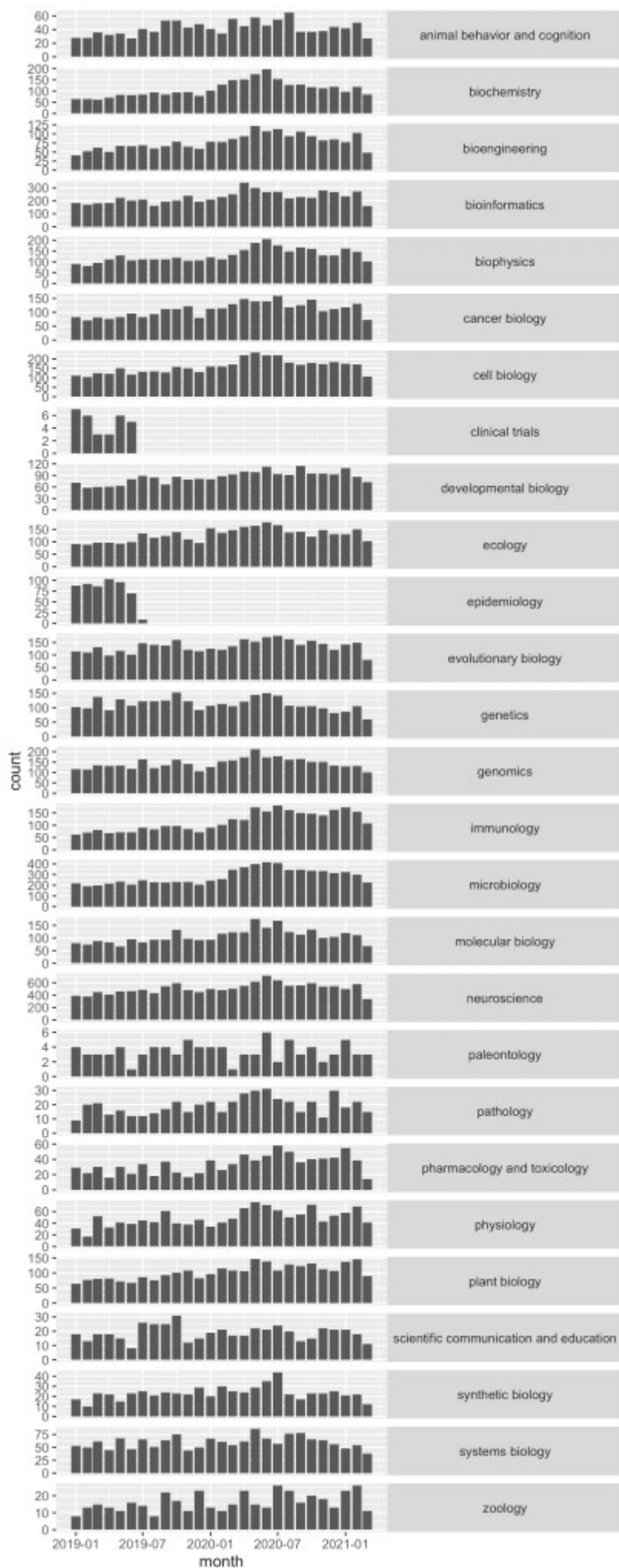
Of course there are other possible explanations. The boom may have been due to [COVID-19 related research](#), but even by mid-April 2020 the share of COVID-19 preprints on medRxiv was 78% compared to 22% at bioRxiv. Plus bioRxiv restricted pandemic-related content early in the pandemic. Other scenarios such as authors being more willing to preprint given the uncertainty of journal publication during a pandemic might also be a factor.

One way to look at this is to examine preprints in different categories to see if they follow similar patterns. If the boom is only due to COVID-19 work we wouldn't expect Neuroscience for example to see the same boom.



Rug plots of daily preprints (first submissions) by category

The daily view shows similar trends but we need to bin into months to see more clearly



Monthly first submissions by category

Most categories show a similar pattern of rise and fall since the pandemic started. Biochemistry, Bioengineering and Biophysics are good examples. Some categories are just noisy due to too few papers but even Synthetic Biology which is quite a small category, shows a similar trend.

It's hard to tell if the slowdown is real. December and January are slow months, February is short and March is only 2/3rds complete, so the 2021 plateau maybe artificial, we'll see. Secondly does it mean a scientific slowdown? Competing servers have emerged and perhaps bioRxiv's boomtime is coming to a natural end. I would hope not. We'll have to look again at this later this year.

Because the rise and fall in submissions seems to happen regardless of category, and given what we know about the impact the pandemic is having on researchers, it seems reasonable to suggest that this analysis shows we are producing less "science" than pre-pandemic.

The code

```
# get data - takes a while!
df_2021 <- biorxiv_content(from = "2021-01-01", to = "2021-03-31",
  limit = "*", format = "df")
df_2020a <- biorxiv_content(from = "2020-07-01", to = "2020-12-31",
  limit = "*", format = "df")
df_2020b <- biorxiv_content(from = "2020-01-01", to = "2020-06-30",
  limit = "*", format = "df")
df_2019a <- biorxiv_content(from = "2019-07-01", to = "2019-12-31",
  limit = "*", format = "df")
df_2019b <- biorxiv_content(from = "2019-01-01", to = "2019-06-30",
  limit = "*", format = "df")
# combine
df_all <- rbind(df_2021,df_2020a,df_2020b,df_2019a,df_2019b)
# we'll look at first submissions only
data <- filter(df_all, version == 1)
# for plotting
data$category <- as.factor(data$category)
data$date <- as.Date(data$date)
data <- data[order(as.numeric(data$date)),]
data$month <- as.Date(paste0(as.character(substr(data$date,1,7)),
  "-01", format = "%Y-%m-%d"))

# rug plot
p1 <- ggplot(data, aes(x = date)) +
  geom_bar() +
  facet_grid(category ~ ., scales = "free") +
  theme(strip.text.y = element_text(angle = 0))
p1

ggsave("Output/Plots/preprintsByCat.png", p1, height = 15, width = 6,
  dpi = 300)

p3 <- ggplot(data, aes(x = month)) +
  geom_bar() +
```

```
facet_grid(category ~ ., scales = "free") +  
theme(strip.text.y = element_text(angle = 0))  
p3  
ggsave("Output/Plots/preprintsByCatMonth.png", p3, height = 15, width =  
6, dpi = 300)
```

Conclusion

The impact of the pandemic has been a rise and fall in submissions to bioRxiv. The causes are uncertain but anecdotally, progress in my lab has definitely taken a hit by the pandemic and I imagine many other labs are also experiencing slowdown. There was no boom for us. We preprinted two papers that were written in the initial lockdown period but we were on track to do that anyway.