```
library(tidyverse)
```

Load spreadsheet

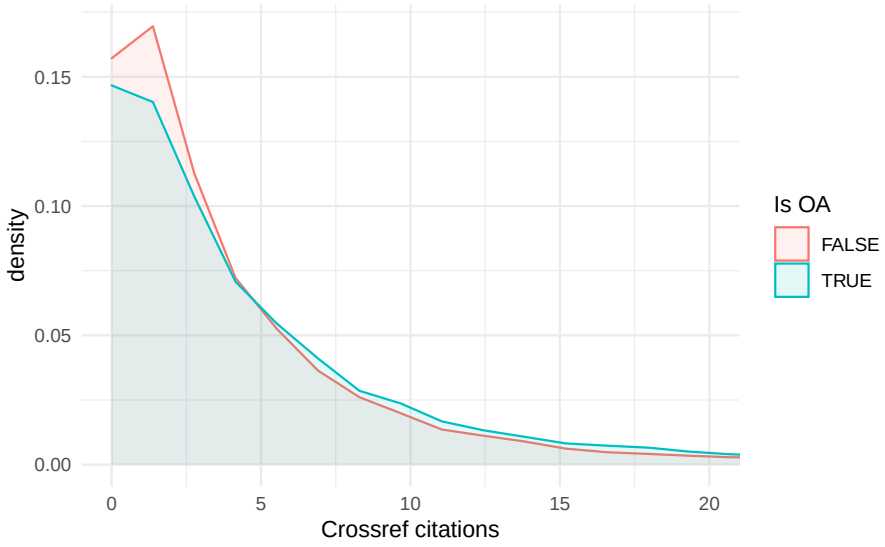
```
my_df <- readxl::read_xlsx("peerj-51535-peerj-51535-data-v2-11Dec2020.xlsx")
```

Citation distribution

```
eligible_pubs <- my_df %>%
  # only authors tagged with with no more than 20 authors
  filter(`Author count>20` == "No")
```

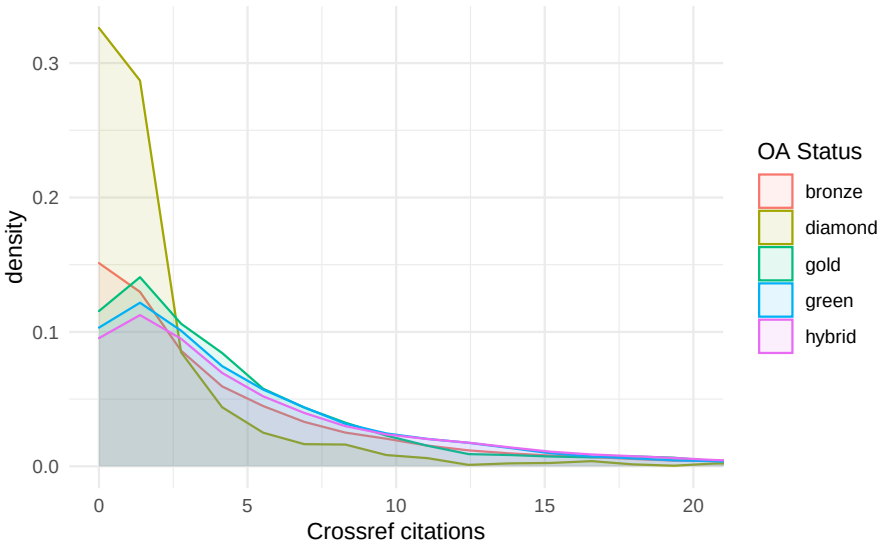
OA / Closed

```
# oa / closed
ggplot(eligible_pubs, aes(`Crossref citations`, fill = `Is OA`, color = `Is OA`)) +
  geom_density(alpha = 0.1) +
  coord_cartesian(xlim = c(0,20)) +
  theme_minimal()
```



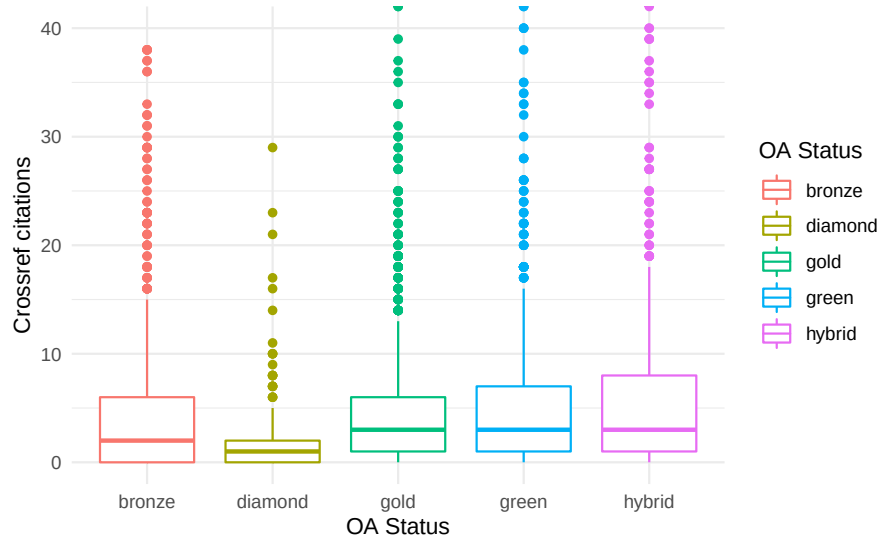
By OA Type

```
# by oa type
eligible_pubs %>%
  # remove closed
  filter(`Is OA` == TRUE) %>%
  ggplot(aes(`Crossref citations`, fill = `OA Status`, color = `OA Status`)) +
  geom_density(alpha = 0.1) +
  coord_cartesian(xlim = c(0,20)) +
  theme_minimal()
```



Boxplot

```
# by oa type
eligible_pubs %>%
  # remove closed
  filter(`Is OA` == TRUE) %>%
  ggplot(aes(`OA Status`, `Crossref citations`, color = `OA Status`)) +
  geom_boxplot() +
  coord_cartesian(ylim = c(0,40)) +
  theme_minimal()
```



Hybrid issue

10.1038/nature20805 is a highly collaborative paper, authored by the Cancer Genome Atlas Research Network. However, the spreadsheet list 9 authors. Similarly, 10.1056/NEJMoa1614362 records 195 authors, but was coded as having less.

Both are highly cited publications!

```
eligible_pubs %>%
  filter(DOI %in% c("10.1038/nature20805", "10.1056/NEJMoa1614362")) %>%
```

```
  select(DOI, `Author count>20`, `Author count`, `OA Status`, `Crossref citations`)
```

DOI	Author count>20	Author count	OA Status	Crossref citations
10.1038/nature20805	No	9	hybrid	283
10.1056/NEJMoa1614362	No	195	green	706

Mean citation with the papers

```
eligible_pubs %>%
  filter(!DOI %in% c("10.1038/nature20805", "10.1056/NEJMoa1614362")) %>%
  group_by(`OA Status`) %>%
  summarise(articles = n_distinct(DOI),
            mean_citations = mean(`Crossref citations`),
            median = median(`Crossref citations`),
            sd = sd(`Crossref citations`))
```

OA Status	articles	mean_citations	median	sd
bronze	1132	5.1	2	11.1
closed	7284	4.4	2	9.4
diamond	265	1.8	1	3.6
gold	1709	5.1	3	7.8
green	1276	6.9	3	17.2
hybrid	634	7.5	3	15.5

Mean citation without these papers

```
eligible_pubs %>%
  filter(!DOI %in% c("10.1038/nature20805", "10.1056/NEJMoa1614362")) %>%
  group_by(`OA Status`) %>%
  summarise(articles = n_distinct(DOI),
            mean_citations = mean(`Crossref citations`),
            median = median(`Crossref citations`),
            sd = sd(`Crossref citations`))
```

OA Status	articles	mean_citations	median	sd
bronze	1132	5.1	2	11.1
closed	7284	4.4	2	9.4
diamond	265	1.8	1	3.6
gold	1709	5.1	3	7.8
green	1276	6.9	3	17.2
hybrid	634	7.5	3	15.5