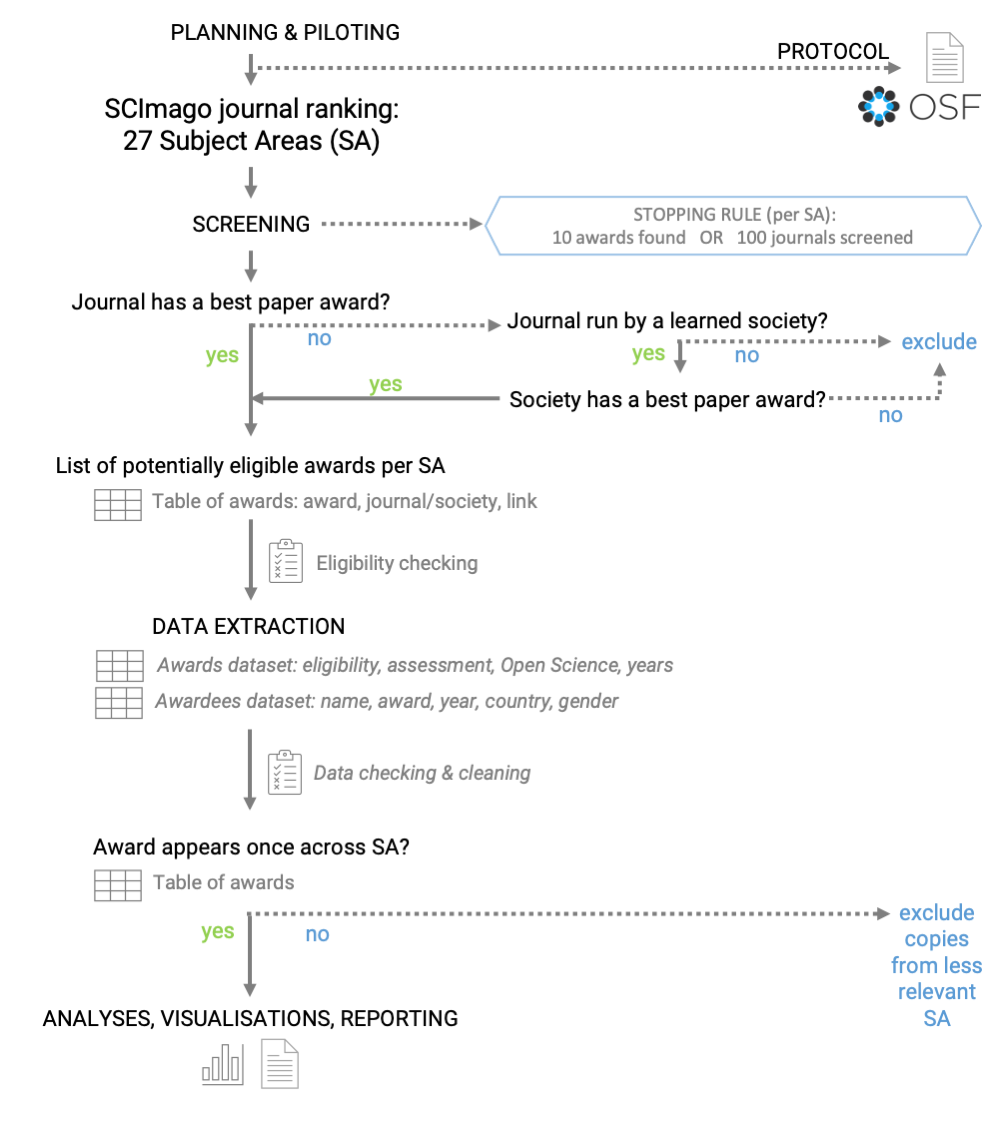
SI\_methods\_analyses

ML

03/08/2023

# Supplementary Methods

## Project workflow

 #### Figure S1.  
*Project workflow.*

### Project aim and general approach.

“Best paper” (or equivalent) awards are usually associated with specific journals / societies / publishers. We aimed to conduct a systematic-like search to collect a sample of highly regarded awards (from top-ranking journals and societies) across a broad range of disciplines. We conducted the process collaboratively and transparently using shared Google Sheets documents (file copies were provided and re-integrated for participants who were not able to access online Google Docs).

The initial project description and contribution guidelines were publicly available as GitHub pages: <https://mlagisz.github.io/survey_best_paper_awards/>. Project contributors were recruited via Open Science communities and conferences. They helped to pilot the project procedures and provided feedback on the project protocol, which was then updated to it second version. roject contributors self-assigned to the individual project tasks. Their actual contributions (completed tasks) were recorded on Google Sheets. Detailed project procedures were shared as a dedicated Project Manual on Google Docs. This document, with incorporated users’ feedback, is available at <https://docs.google.com/document/d/1aw_HhKawpi264njGi--Ms-x-bOs_uQJBiOepQKE89pQ/edit?usp=sharing>.

### Award inclusion criteria

We followed pre-specified criteria outlined in the registered protocol (<https://osf.io/7ez28>) to increase data consistency and enable comparisons across disciplines:  
1. We only included awards for a single published research contribution in a format of a research article (theses / dissertations are allowed if published as a journal article).  
2. W only included awards managed by a journal, publisher, or a learned society.  
3. We excluded awards that are discontinued, and awards that are restricted to applicants from underrepresented groups, e.g., women-only / minorities-only awards.

### Search and screening

1. Our starting point was from searching academic journals and societies/associations to identify those that award “Best Paper” prizes. To do so, we first used journal lists from 27 SCImago Subject Areas rankings (SCImago Journal & Country Rank (SJR) website at <https://www.scimagojr.com/journalrank.php>). To represent SAs evenly, we considered the top 100 journals from each SA list, or we will stop if we find 10 relevant awards before reaching this threshold.
2. During the intial screening, for each journal list representing a SSCImago Subject Area, we checked if journal websites or announcements include “best paper” awards , and if not, we checked if journals are run by learned societies (i.e. are offciel journals of a given learned society/ies). If so, we also checked if a relevant learned society awards a “best paper” prize that fits in a given subject area being assessed. If a society awards multiple relevant awards in journals in the same subject area, we choose the one from our top 100 SA list or an award that appears to be most prestigious (e.g., with highest monetary value, oldest).
3. For each subject area, we collated a preliminary list of 10-15 awards that meet our selection criteria for data extraction. If some of these awards got excluded during data extraction, we went back to the screening stage for that subject area to find replacement awards until we had 10 awards that are eligible for inclusion, or we ran out of journals to screen within the top 100 Subject Area list positions.

### Screening cross-checking

1. All screened journal lists from 27 SCImago Subject Areas rankings (<https://www.scimagojr.com>) were independently cross-checked to detect potential mistakes (it is easy to miss a note about awards on journals or organizational webpages, misinterpret eligibility of the award, etc.).
2. Project contributors self-assigned to one (or more) SCImago Subject Area - always different from the one they did initial Subject Areas screening.
3. The checking person confirmed if screening information and decision recorded by the initial screener is correct and accurate.
4. The checking person recorded in a new column (Award\_name\_extract) the name of the eligible award to be extracted and recorded any screening comments in additional columns.
5. Obvious mistakes were fixed and any unclear or disputable cases have been commented on and discussed before being resolved.

### Data extraction

1. For each 27 Scimago Subject Areas (SA) ranking (<https://www.scimagojr.com>), we extracted data on max. 10 awards that meet our selection criteria.
2. For each award, its eligibility (as outlined in “Selection criteria”) for inclusion was confirmed during data extraction, and if deemed not eligible, no further data was be extracted.
3. Using a pre-piloted Google Form, we extracted relevant data on each award from the websites (e.g., journals, societies, publishers) or other publicly available documents (e.g., instructions for applicants).
4. The extracted pre-specified data included information on the granting journal / publisher / society, award eligibility criteria, award assessment criteria, and source of information on the past awardees (full metadata is provided alongside the data files).
5. We archived web pages with award descriptions by saving them to .pdf files named after award names (as used in data extraction documents).
6. The second stap of data extraction focused on the identities and characteristics of the past participants. For collecting data on individual winners, we had a data collection directly on a Google Sheet. We only collect data on winners from years 2001-2022. This data was used to infer patterns related to individual winners’ gender and country of affiliation. Specifically: Gender (binary: F = female, M = male, ? - unassigned) was be assigned to the past awardee names using information (pronouns, images, names) available on award websites, personal or institutional websites, or using <https://gender-api.com/>, with an accuracy threshold of >95 for assigning gender from first names. We noted instances of non-binary gender (not applicable) and also when we are unable to assign binary gender. Affiliation information (institution, country) was be assigned to past winners’ names for the year when the award was received using information available on award websites, personal or institutional websites, or information in the awarded paper. Additional information was also recorded: links to the websites with relevant information, type of source of information (award page, award announcement, article, other), whether award is shared (more than one author from one winning paper), whether photo or bio are provided on award page or award announcement, reference information and DOI for winning article, any additional notes.

### Data extraction cross-checking

1. Extracted main award characteristics data (awarding body, description, etc.) has been independently cross-checked by a second researcher (i.e. one that did not extract the data). Obvious mistakes were fixed and any unclear or disputable cases have been commented on and discussed before being resolved.
2. Extracted additional data (gender, affiliation country, and additional details on the award and winners) has been independently cross-checked by a second researcher (i.e. one that did not extract the data). Obvious mistakes were fixed and any unclear or disputable cases have been commented on and discussed before being resolved.

### Deviations from the protocol

The pre-registered protocol is available on OSF (<https://osf.io/8mxpq/>).  
We acknowledge the following deviations from the protocol:

1. We coded whether the award goes to an individual author (explicitly named) or to an article (without mentioning individual author/s). Some awards only lists winning articles and other focus on specific authors who were the prizewinners (sometimes even not mentioning for which exact article they got an award).
2. In the protocol, we planned only to extract awardee names, gender and affiliations for winners from 2001 onwards. However, such data is only meaningful when an award is given to a specific individual researcher - thus, we did not extract such information for awards that only list winning papers (i.e. were awarded to the whole authorship team, not a particular person within the team).
3. We coded whether the individual winners got their photo and/or bio shared online by the journals/societies. Photographs and bios make assigning gender easier and provide some extra recognition and visibility for the winners.
4. We recorded the information (reference and DOI) on the winning article identity, where provided or findable. However, we did not use this information in the analyses, as it was solely for documentation and cross-checking purposes, since we extracted information on winner’s affiliation from the winning article when affiliation it was not reported on the award webpage or announcement.
5. We recorded whether award criteria mention article impact indices (e.g. number of downloads or citations) and whether impact indices were the only criteria. Although impact indices give an impression of objectivity they are known to be biased and not necessarily reflect the quality of the published research. As such, we decided to include this information in our assessment of the award characteristics.
6. For the analyses, we did not remove Subject Areas that had data extracted from fewer than 5 awards, as initially planned. This is because our main results are presented as totals across all Subject Areas. Also, the two Subject Areas to be potentially excluded were Medicine and Immunology, and we felt that removing them would make our findings less complete.
7. For text mining of award descriptions, we ad hoc grouped the most common terms (truncated to word stems) into terms representing generic excellence and terms explicitly related to transparency and robustness.
8. We were not able to seek clarifications from the award committees or contact person, because most of the award descriptions lacked contact details (see our additional results on the availability of the contact details).
9. In our plots by decade, we also showed the incomplete decade 2021-2022 - in our protocol we stated that only decades 2001-2010 and 2011-2020 will be shown.

### Data processing and analyses - an overview

The main steps were as follows:

1. Uploading cross-checked meta-data, screening data and extracted data into R.
2. Creating meta-data tables (variable descriptions) for each data set.
3. Data pre-processing (e.g. merging SA into a single data frame, removing excluded awards).
4. Summarising data sets via tabulation and filtering.
5. Creating figures for the supplementary materials - shown as in-text figures.
6. Creating figures for the main manuscript text - saved as stand-alone files.

NOTE: Steps 3-6 were done for each data set in turn.  
All code and results are provided below. Code for figures used in the main text is shown, but not evaluated for this supplementary document, because these figures are presented in the main text.

## Project set up

NOTE: project github page can be found at <https://github.com/mlagisz/survey_best_paper_awards>.

* Loading packages and general settings.

## Loading data

* Load SCImago Subject Area -level data set and associated meta-data table.

# accessing all the sheets   
#sheet\_names <- excel\_sheets(here("data", "scimagojr 2021 Subject Areas.xlsx"))  
sheet\_names <- excel\_sheets(here("data", "scimagojr 2021 Subject Areas\_screening.xlsx"))  
sheet\_names <- sheet\_names[-1] #remove first sheet with meta-data  
SA\_list\_all <- lapply(sheet\_names, function(x)   
 {as.data.frame(read\_excel(here("data", "scimagojr 2021 Subject Areas.xlsx"), sheet = x))}) #read all sheets  
names(SA\_list\_all) <- sheet\_names #rename list elements   
#lapply(SA\_list\_all, names) #extract column names  
#unique(do.call(rbind, lapply(SA\_list\_all, names))) #making sure all have same column names  
SAdata <- do.call(rbind, SA\_list\_all)  
names(SAdata) <- gsub(" ","\_", names(SAdata)) #change spaces to \_ in the column names  
names(SAdata) <- gsub("\\.","", names(SAdata)) #remove . the column names  
SAdata <- SAdata %>% drop\_na(Subject\_area)  
#table(is.na(SAdata$Title)) #note no empty journal names  
  
SAmeta <- read\_excel(here("data", "scimagojr 2021 Subject Areas\_screening.xlsx"),  
 sheet = 1, skip = 1) #load SCImago SA meta-data  
#dim(SAmeta)  
#names(SAmeta)

* Load award-level data set and associated meta-data table.

#BP load and prepare award-level meta-data  
BPmeta <- read\_excel(here("data", "Survey-Best\_paper\_awards (Responses)\_SHAREDCOPY\_checked.xlsx"),  
 sheet = 2, skip = 1) #load and skip first line  
  
## load and prepare main extracted award-level data  
BPdata <- read\_excel(here("data","Survey-Best\_paper\_awards (Responses)\_SHAREDCOPY\_checked.xlsx"),  
 sheet = 1) #load main award data  
  
names(BPdata) <- gsub(" ","\_", names(BPdata)) #change spaces to \_ in the column names  
  
#rename selected data columns  
BPdata <- BPdata %>%   
rename(Extractor\_name = "Name\_of\_the\_extracting\_person",  
 Award\_name = "Full\_name\_of\_the\_award",   
 Journal\_name = "Full\_name\_of\_the\_awarding\_journal",   
 Award\_description = "Paste\_the\_information\_text\_describing\_the\_award",   
 Eligible = "Confirm\_eligibility\_of\_the\_award",  
 Awarding\_journal = "Full\_name\_of\_the\_awarding\_journal",  
 Awarding\_society = "Full\_name\_of\_the\_awarding\_society",  
 Awarding\_other = "Full\_name\_of\_the\_awarding\_publisher/other\_body",   
 Career\_stage = "Target\_career\_stage\_of\_eligible\_applicants",  
 Flexible\_eligibility = "Flexibility\_of\_the\_eligibility\_criteria",   
 Assessors\_transparency = "Assessor\_transparency",   
 Open\_science = "Valuing\_Open\_Science",   
 Self\_nomination = "Self-nomination\_allowed"  
 )  
  
##check for rows with empty "Scimago Subject Area" values  
table(is.na(BPdata$Scimago\_Subject\_Area)) #4 rows from pilot screening

##   
## FALSE TRUE   
## 264 4

##remove rows with pilot extractions and empty "Scimago Subject Area" values  
BPdata <- BPdata[!is.na(BPdata$Scimago\_Subject\_Area), ]  
  
#remove awards that were duplicate-extracted and excluded at extraction stages  
BPdata <- BPdata[BPdata$Row\_excluded == 0, ]  
dim(BPdata) #222 rows left

## [1] 222 54

##NOTE: total of 41 rows removed as duplicates   
##(19 awards were doubly- or triply-extracted)  
  
#create new variable for awards with description or without  
BPdata$Descr\_available <- fct\_collapse(BPdata$Open\_science,  
 available = c("no", "yes"),  
 "not available" = "not available"  
)

* Load winners-level data set and associated meta-data table.

#load individual winners data  
BPindiv <- read\_csv(here("data", "BP\_awards\_lists\_SHAREDCOPY - indiv\_winners\_20230915.csv"),  
 skip = 1) #load individual winners data

## New names:  
## Rows: 1079 Columns: 23  
## ── Column specification  
## ──────────────────────────────────────────────────────── Delimiter: "," chr  
## (21): award\_SA, award\_name, award\_link, individual, awardee\_name, awarde... dbl  
## (2): record\_count, award\_year  
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ  
## Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## • `` -> `...23`

#load individual winners meta-data  
BPindiv\_meta <- read\_csv(here("data", "BP\_awards\_lists\_SHAREDCOPY - indiv\_winners\_meta-data.csv"),  
 skip = 1) #load individual winners data

## Rows: 22 Columns: 3  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (3): column name, description, data type [options]  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

* Load country-level productivity SCImago data set and create associated meta-data table.

#load SCImago 2021 country productivity (documents) data downloaded from https://www.scimagojr.com/countryrank.php?year=2021&min=0&min\_type=it  
COprod <- read\_csv(here("data", "scimagojr country rank 2021.csv"), skip = 0) #load data

## Rows: 235 Columns: 9  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): Country, Region  
## dbl (7): Rank, Documents, Citable documents, Citations, Self-citations, Cita...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#create meta-data with columns: data type [options]  
COprod\_meta <- tibble("column name" = names(COprod),   
 "description" = c("Rank of a given country across all Scimago Subject Areas.",  
 "Country name.",  
 "Country location classification into one of the eight SCImago Major World Regions.",  
 "Number of documents associated with a goven country published in 2021.",  
 "Number of citable documents associated with a goven country published in 2021.",  
 "Number of citations to documents associated with a goven country published in 2021.",  
 "Number of self-citations to documents associated with a goven country published in 2021.",  
 "Number of citations per document associated with a goven country published in 2021.",  
 "H index of a country in 2021." ),  
 "data type [options]" = c("integer",  
 "free text",  
 "free text",  
 "integer",  
 "integer",  
 "integer",  
 "integer",  
 "numeric",  
 "integer"  
 ))  
#dim(COprod\_meta)  
#names(COprod\_meta)

## Meta-data tables

#### Table S1.

*Meta-data for the SCImago journal rankings by Subject Areas dataset*.

# making a scrollable table  
kable(SAmeta, "html") %>%  
 kable\_styling("striped", position = "left") %>%  
 scroll\_box(width = "100%", height = "1500px")

column name

description

data type [options]

Subject\_area

Name of the Scimago Subject Area (SA) representing 27 major thematic categories.

categorical [one of 27 Subject Areas]

Rank

Rank of a given Journal within a given Scimago Subject Area.

integer

Sourceid

Unique ID for the journal.

integer

Title

Name of the journal.

free text

Reviewer\_name

Name of a person who reviews the journal trying to find associated eligible awards.

free text

Link\_journal

Year when award was won/announced for a give individual winner, as reported on the award page or other documents with winner information.

web link

Society\_name

Name of a person who won the award, as reported on the award page or other documents with winner information.

free text

Has\_relevant\_award

Gender assigned to an individual past awardee names using information from pronouns, images, names) available on award websites, personal or institutional websites, or from first names.

categorical [yes; no]

Link\_award

Main source of information for assigning gender. Order of preference: pronouns, photo, name.

web link

Notes\_award

Affiliation institution (usually university) assigned to a winner when the award was received using information available on award websites, personal or institutional websites, or information in the awarded paper.

free text

Checker\_name

Name of the person who cross-checked a given row of screening data.

free text

Check\_pass

Main source of the winner’s affiliation information (institution, country) for the year when the award was received or announced.

categorical [yes; no]

Check\_notes

Any comments for awardee, e.g. award sub-category, or link to additional info from Internet searches.

free text

Award\_name\_extract

Whether the award is shared with other authors of the same article.

categorical [yes; no]

Extract

Whether a bio/profile of the winner is shown on the award page/announcement (with more extended information than affiliation information only).

categorical [yes; no]

Award\_extractor\_name

Name of the person who extracted award data .

categorical [yes; no]

Extracted

Whether a given award has been extracted.

categorical [yes; no]

Comments

Anny comments regarding screened and extracted data in a given row.

free text

Archived

DOI of the winning paper. DOI code in a format starting with “10.”.

DOI

Issn

ISSN number of teh journal

free text

SJR

SCImago Journal Rank indicator expresses the average number of weighted citations received in the selected year by the documents published in the selected journal in the three previous years, i.e. weighted citations received in year X to documents published in the journal in years X-1, X-2 and X-3

free text

SJR Best Quartile

Anny comments regarding extracted data in a given row.

categorical [Q1; Q2; Q3; Q4]

H index

The h index expresses the journal’s number of articles (h) that have received at least h citations. It quantifies both journal scientific productivity and scientific impact and it is also applicable to scientists, countries, etc.

number

Total Docs. (2021)

Output of the selected period. All types of documents are considered, including citable and non citable documents.

integer

Total Docs. (3years)

Published documents in the three previous years (selected year documents are excluded), i.e.when the year X is selected, then X-1, X-2 and X-3 published documents are retrieved. All types of documents are considered, including citable and non citable documents.

integer

Total Refs.

Total number of all the bibliographical references in a journal in the selected period.

integer

Total Cites (3years)

Number of citations received in the seleted year by a journal to the documents published in the three previous years, –i.e. citations received in year X to documents published in years X-1, X-2 and X-3. All types of documents are considered.

integer

Citable Docs. (3years)

Number of citable documents published by a journal in the three previous years (selected year documents are excluded). Exclusively articles, reviews and conference papers are considered.

integer

Cites / Doc. (2years)

Average citations per document in a 2 year period. It is computed considering the number of citations received by a journal in the current year to the documents published in the two previous years, –i.e. citations received in year X to documents published in years X-1 and X-2.

number

Ref. / Doc.

References per Document is the average number of references per document in the selected year.

number

Country

Country where journal is indexed by Scopus.

categorical

Region

Eight Major World Regions in used to facilitate sectorial analyses.

categorical

Publisher

name of teh journal publisher.

free text

Coverage

Year range for which publication data is available on Scopus (?).

free text

Categories

Tags for relevant 309 specific subject categories according to Scopus® Classification.

categorical

#### Table S2.

*Meta-data for the Best Paper awards dataset*.

# making a scrollable table  
kable(BPmeta, "html") %>%  
 kable\_styling("striped", position = "left") %>%  
 scroll\_box(width = "100%", height = "2500px")

column name

description

data type [options]

Timestamp

System-recorded data and time of when the given row of data was submitted via the Google Form.

time

Name of the extracting person

Name of the extracting person: Please use your name in one consistent form (should be matching the name you gave us on EOI and recorded on GoogleSheet for award screening). It is best to always use your first and last name, to avoid confusion.

free text

Scimago Subject Area

Scimago Subject Area: Record name of the Scimago Subject Area (SA) you are doing extractions for - as in the Subject\_area column in the GoogleSheet with journal and awards lists. Should be exactly as used in the original list of Subject Areas.

categorical [one of 27 Subject Areas]

Full name of the award

Full name of the award: Record specific award name, as in the Award\_name\_extract column in the GoogleSheet with journal and awards lists.

free text

Source of the information on the award

Source of the information on the award: Ideally, enter a link to an online page/document with information on award eligibility and assessment criteria. If not available, could be also a link to any document describing the award. If you cannot find any information about the award enter “NA”. You can paste more than one link separated by “;”.

weblink

Paste the information text describing the award

Paste the information text describing the award: Ideally, copy and paste the section of the online page/document with information on award eligibility and assessment criteria only. This description text will be used for text-mining analyses. Do not copy and paste list of past winners or generic information about the journal or society. Do not copy and paste any images. If you cannot find any information about the award enter “NA”.

free text

Confirm eligibility of the award

Confirm eligibility of the award: · This award is for a single published research contribution in a format of a research article (theses / dissertation are allowed if published as a journal article). · This award is managed by a journal, publisher, or a learned society. · This award is NOT discontinued or restricted to applicants from underrepresented groups, e.g., women-only / minorities-only. Options: yes (eligible award), no (not eligible - make note on the reasons below and enter NA to the following compulsory questions, then submit), unclear (make note on the reasons below)

categorical [yes; no; unclear]

Comment on the award eligibility

Comment on the award eligibility: Any comments, e.g. reasons for excluding this award.

free text

Full name of the awarding journal

Full name of the awarding journal: Record if award is associated with a specific journal. Enter “NA” if not associated with a specific journal.

free text

Full name of the awarding society

Full name of the awarding society: Record if award is associated with a specific society. Enter “NA” if not associated with a specific society.

free text

Full name of the awarding publisher/other body

Full name of the awarding publisher/other body: Record if award is associated with a specific publisher (group of journals) or awarding organisation other than learned society or journal. Enter “NA” if not associated with a specific publisher.

free text

Comment on the awarding body

Comment on the awarding body: Any comments, e.g. if more than one journal considered for the award.

free text

Award cash

Award cash: Code whether there is any cash coming with the award. - Select “yes” if award description mentions any cash given to the winners. - Select “no” if award description does not mention any cash given to the winners. Travel awards, publishing discounts, plenary talks, etc. do not count as cash awards. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: In the next question (“Inclusivity phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available. Options: yes, no, not available.

free text

Comment on the award cash

Comment on the award cash: Add comments on cash amount, if provided, and note any other benefits winners receive, as relevant (e.g., travel awards, publishing discounts, plenary talks)

free text

Target career stage of eligible applicants

Target career stage of eligible applicants, as stated in the award information: As stated in the award information. More than one choice is possible. If it is for any paper in a journal/s, select “any career stage”. Options: student, early-career, mid-career, any career stage, unclear.

categorical [student; early-career; mid-career; any career stage; unclear]

Comment on the career stage

Comment on the career stage: Any comments, e.g. paste the phrasing of who is eligible for the award.

free text

Flexibility of the eligibility criteria

Flexibility of the eligibility criteria: Code whether explicitly allowing for career interruptions in eligibility timeframes: - Select “not applicable” if description only states that published research has to be performed while studying for a degree OR if there is no eligibility limit in terms of years since an event (e.g. a PhD or author’s age. - Select “yes” if the description says that there is an eligibility limit of years after a degree to apply for the. award and this limit can be extended in special cases”. - Select “no” if the description says that there is an eligibility limit of years after a degree to apply for the award but it does not mention that this limit can be extended in special cases. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: In the next question (“Eligibility phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available. Options: yes, no, not applicable, not available.

categorical [yes; no; not applicable; not available]

Eligibility phrasing

Eligibility phrasing: Wording of the eligibility criteria in relation to career stage in the relevant documentation, if available.

free text

Inclusivity statement

Inclusivity statement: Code whether underrepresented/minority groups are encouraged to apply for the award (this does not mean that the award is restricted to underrepresented groups, e.g., women-only) or award information includes declarations of commitment to equity / diversity / inclusivity (EDI): - Select “yes” if award description mentions underrepresented/minority groups or EDI. - Select “no” if award description does not mention anything about underrepresented/minority groups or EDI. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: Extract this information from award descriptions only - do not include information from other documents not directly related to the award, e.g. journal/society/institutional policies or mission statements. In the next question (“Inclusivity phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available.

categorical [yes; no; not available]

Inclusivity phrasing

Inclusivity phrasing: Wording of the inclusivity statement in the relevant documentation, if available.

free text

Assessor transparency

Assessor transparency: Code whether information is provided on who will be conducting assessments of candidate papers (for example, that editors-in-chief will be deciding): - Select “yes” if information is provided, e.g. names, or stating that journal editors will be doing this, or mentioning specific existing committee that is described somewhere else (e.g. on a society webpage). - Select “no” if description does not mention any information on the assessors. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). In the next question (“Assessor phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available.

categorical [yes; no; not available]

Assessor phrasing

Assessor phrasing: Wording of the information on who will be conducting the assessments, if available.

free text

Process transparency

Process transparency: Code whether breakdown of the applicants / candidates by gender or geographic region is publicly available: - Select “yes” if such information is provided (e.g. on a society webpage). - Select “no” if such information is not provided. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: In the next question (“Process phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available. Options: yes, no, not available.

categorical [yes; no; not available]

Process phrasing

Process phrasing: Wording of the information on the transparency of the process, e.g. if breakdown of the applicants / candidates by gender or geographic region is publicly available.

free text

Feedback availability

Feedback availability: Code whether award information included an offer of constructive feedback for unsuccessful applicants: - Select “yes” if such feedback can be provided (e.g. on request). - Select “no” if such feedback is not provided or not mentioned. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: In the next question (“Feedback phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available. Options: yes, no, not available.

categorical [yes; no; not available]

Feedback phrasing

Feedback phrasing: Wording of the information on whether/how feedback will be provided, if available.

free text

Criteria transparency

Criteria transparency: Code whether assessment criteria are detailed (more than one vague sentence) or vague (often one vague sentence, e.g. “assessed on innovation and novelty”): - Select “yes” if assessment criteria are more than one vague sentence (e.g., mentions things like readability of the language, sample size, performing alternative tests/analyses to check robustness of the results, sharing data or code, registering the study/protocol - whatever is relevant to a given research field). - Select “no” if assessment criteria are only one vague sentence (e.g. award is for “best paper”, “outstanding contribution”, “excellent research”, “novel insights”, etc.). - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: In the next question (“Criteria phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available. Options: yes, no, not available.

categorical [yes; no; not available]

Criteria phrasing

Criteria phrasing: Wording of the information on the assessment criteria, if available.

free text

Valuing Open Science

Valuing Open Science: Code whether any Open Science practices (data, code, materials sharing, preregistration, transparency of reporting, etc.) are explicitly included in the assessment criteria: - Select “yes” if award description mentions valuing any of the Open Science practices (e.g., sharing data or code, registering the study/protocol, performing replication studies - whatever is relevant to a given research field). - Select “no” if the award description does not mention valuing any of the Open Science practices. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: In the next question (“Valuing Open Science phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available. Options: yes, no, not available.

categorical [yes; no; not available]

Valuing Open Science phrasing

Valuing Open Science phrasing: Wording of the information on the assessment criteria valuing Open Science practices, if available.

free text

Self-nomination allowed

Self-nomination allowed: Code whether candidates can self-nominate for the award. - Select “yes” if the award description explicitly mentions that candidates can self-nominate. - Select “no” if the award description states that candidates/papers need to be nominated by somebody else or does not explicitly mention that candidates can self-nominate. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: In the next question (“Self-nomination phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available. Options: yes, no, not available,

categorical [yes; no; not available]

Self-nomination phrasing

Self-nomination phrasing: Wording of the information on how to nominate (e.g., using a checkbox on the manuscript submission form), if available.

free text

Letter required

Letter required: Code whether candidates are required to provide nomination/recommendation letter/s: - Select “yes” if award description explicitly mentions that candidates have to provide nomination or support letters from others. - Select “no” if the award description states that candidates/papers do not need to provide letters from somebody else or does not explicitly mention anything about such letters. - Select “not available” if there is no page/document describing the award (e.g. only the list of winners is publicly published). NOTE: In the next question (“Letter requirement phrasing”) copy and paste relevant text from the award document or make a note if no such document/information is available. Options: yes, no, not available.

categorical [yes; no; not available]

Letter requirement phrasing

Letter requirement phrasing: Wording of the information on the requirement for written nominations / reference letters, if available.

free text

Awardee list source

Awardee list source: Source of the information on the past winners - paste a link to a webpage, file name, personal information, etc., showing a list of past winners (any year span). Write “not available” if no such list exists.

weblink

Awardee list most recent year

Awardee list most recent year: The most recent year for which information on past awardees is available (enter whole number, e.g. 2022). Leave empty if no such list exists.

integer

Awardee list earliest year

Awardee list earliest year: The earliest year for which information on past awardees is available (enter whole number, e.g. 1999). Leave empty if no such list exists.

integer

Comments on awardees list

Comments on awardees list: Any comments for awardees list, e.g. if Internet searches might be needed to locate announcements with the names of the past winners.

free text

Comments\_general

Comments\_general: Any other notes and comments on issues, assumptions, or seeking additional information from the award committees / contacts.

free text

Checked

Whether a given row of extracted data has been cross-checked.

categorical [yes; no]

Checker\_name

Name of the person who cross-checked a given row of extracted data.

free text

Checker\_comments

Anny comments regarding extracted data in a given row.

free text

Row\_excluded

Whether a given row of extracted data has been cross-checked.

integer [1 = yes; 0 = no]

Award\_individual

Whether a given award recognises individual (selected) authors rather than all authors of a winning paper.

categorical [yes; no]

Award\_impact\_metrics\_mentioned

Whether award criteria mention impact metrics (number of citations, downloads) as basis of winner selection.

categorical [yes; no]

Award\_impact\_metrics\_only

Whether impact metrics (number of citations, downloads) ar the sole basis of winner selection.

categorical [yes; no]

Award\_impact\_metrics\_comment

Quote relevent sections of award description if impact metrics are mentioned. Add any comments regarding award criteria and impact metrics.

free text

Award\_contact\_provided

Whether specific contact information (email) is provided on the award webpage or publicly available application/description documents.

categorical [yes; no]

Award\_integrity\_mentioned

Whether award description states how potential conflictes of interests will be managed

categorical [yes; no; not available]

Award\_integrity\_comment

Quote relevent sections of award description on how potential conflicts of interests will ba managed.

free text

Award\_cash\_max\_USD\_pperson

The maximum reported amount of cash awarded per winning paper, recalculated online into USD.

Integer

N\_winners\_extracted

Number of individual winners from years 2001-2022 extracted for this awards (in indiv\_winners data set).

Integer

Archived\_files

Whether award website and over information have been archived as pdf files in the dedicated project folder.

categorical [yes; no]

Comment\_extraction

Anny comments regarding extracted data in a given row.

free text

#### Table S3.

*Meta-data for the individual winners dataset*.

# making a scrollable table  
kable(BPindiv\_meta, "html") %>%  
 kable\_styling("striped", position = "left") %>%  
 scroll\_box(width = "100%", height = "1500px")

column name

description

data type [options]

award\_SA

Name of the Scimago Subject Area (SA) you are doing extractions for - as in the Subject\_area column in the GoogleSheet with journal and awards lists. Should be exactly as used in the original list of Subject Areas.

categorical [one of 27 Subject Areas]

record\_count

Subsequent numbers for counting extracted winners within each award.

integer

award\_name

Name of the award - as in the Award\_name column in the main GoogleSheet with awards main data extraction.

free text

award\_link

Weblink to the main Internet page with award description / information.

weblink

individual

Whether an award is individual or article-focused. Individual awards are defined as awards that are usually not shared equally between all article authors. For example the award names as a winner only the ECR authors or only the first author (unless there are more than one ECR/first authors who contributed equally).

categorical [yes; no]

award\_year

Year when award was won/announced for a give individual winner, as reported on the award page or other documents with winner information.

integer

awardee\_name

Name of a person who won the award, as reported on the award page or other documents with winner information.

free text

awardee\_gender

Gender assigned to an individual past awardee names using information from pronouns, images, names) available on award websites, personal or institutional websites, or from first names.

categorical [F = female; M = male; ? - unassigned]

gender\_source

Main source of information for assigning gender. Order of preference: pronouns, photo, name.

categorical [pronouns; photo; name]

affiliation\_institution

Affiliation institution (usually university) assigned to a winner when the award was received using information available on award websites, personal or institutional websites, or information in the awarded paper.

free text

affiliation\_country

Affiliation country assigned to a past winner for the year when the award was received using information available on award websites, personal or institutional websites, or information in the awarded paper.

free text

affiliation\_info\_source

Main source of the winner’s affiliation information (institution, country) for the year when the award was received or announced.

categorical [award page; award announcement; article; other]

awardee\_comment1

Any comments for awardee, e.g. award sub-category, or link to additional info from Internet searches.

free text

shared

Whether the award is shared with other authors of the same article.

categorical [yes; no]

awardee\_profile\_shown

Whether a bio/profile of the winner is shown on the award page/announcement (with more extended information than affiliation information only).

categorical [yes; no]

awardee\_photo\_shown

Whether a photo of the winner is shown on the award page/announcement.

categorical [yes; no]

awardee\_comment2

Any other comments on the awardee or awarded paper, e.g. award sub-category, or link to additional info from Internet searches.

free text

awarded\_paper\_ref

Bibliographic reference of the winning paper.

free text [reference in any long format]

awarded\_paper\_doi

DOI of the winning paper. DOI code in a format starting with “10.”.

DOI

checked

Whether a given row of extracted data has been cross-checked.

categorical [yes; no]

checker\_name

Name of the person who cross-checked a given row of extracted data.

free text

checker\_comment

Anny comments regarding extracted data in a given row.

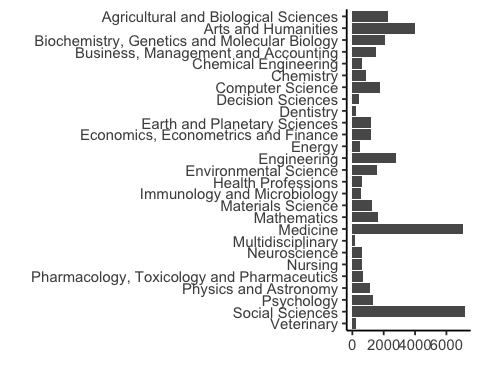
free text

# Supplementary Results

## Characteristics of SCImago Subject Area (SA) journal lists and screening data

### Number of journals per SCImago Subject Area

##count journals per SA  
# count\_j\_perSA <- SAdata %>%  
# count(Subject\_area) %>%  
# arrange(desc(n))   
  
#table(SAdata$Subject\_area)  
#min(table(SAdata$Subject\_area))  
#max(table(SAdata$Subject\_area))  
  
#all SA as a simple barplot  
SAdata %>%   
 #filter(!is.na(Subject\_area)) %>%   
 count(Subject\_area) %>%  
 arrange(Subject\_area) %>%  
 ggplot(aes(x = reorder(Subject\_area, desc(Subject\_area)), y = n)) +  
 geom\_bar(stat = "identity", position = position\_dodge(0.9)) +  
 coord\_flip()



#### Figure S2.

*Counts of journal Titles per SCImago ranking lists by Subject Area*.

The numbers of journals per Subject Area vary widely reflecting the differences in sizes of different research disciplines. Thus, random sampling of journals across Subject Areas could result in uneven representation of disciplines, potrentially biasing the findings towards the largest disciplines. By sampling similar number of journals with similar levels of “ranking” in each Subject Area, we were able to get a rough estimate of the frequency of the awards across disciplines.

### Overlap of journals across SCImago Subject Areas

#names(SAdata)  
# length(unique(SAdata$Title)) #number of unique journals  
# length(unique(SAdata$Title))/length(SAdata$Title) #60% unique  
  
#reduce to top 100 from each SA  
SAdata %>%   
 #filter(!is.na(Subject\_area)) %>%   
 group\_by(Subject\_area) %>%  
 top\_n(100, Title) -> SAdata100  
  
#count number of duplicated journal titles across all SA top 100 journal titles:  
#sum(duplicated(SAdata100$Title)) #644 duplicated journal Titles (out of 2700)  
#sum(duplicated(SAdata100$Title))/length(SAdata100$Title) #24% duplicated journal Titles (out of 2700)

Across the top 100 journals in each Subject Area ranking list there are 644 duplicated journal titles (23.9%). This indicates that one journal can often be classified into one or more related Subject Areas (e.g. “Trends in Cognitive Sciences” is included both in Neuroscience and Psychology Subject Areas). This overlap implies that the related awards will also often be relevant to more than one Subject Area. Thus, after the screening step, we checked for duplicate inclusions and data extractions and assigner each award to only one Subject Area for the analyses to avoid double-counting.

## Award-level dataset characteristics

**Overall summary:** Total number of unique names of awards: 222.

Total number of unique names of awards: 27.

Total number of unique names of awarding journals: 177.

Total number of unique names of awarding societies: 127.

Total number of unique names of awarding other bodies: 22.

Total number of awards without award description: 11 (5%)

## Award-level dataset characteristics

**Overall:** Total number of unique names of individual winners records: 1079.

##Check award names - same award could have been extracted in different SA, or different awards have same names   
# table(duplicated(BPdata$Award\_name)) #0 duplicated, 222 unique  
  
## Simple counts  
  
# #eligible awards per SA  
# count\_awards\_perSA <- BPdata %>%   
# count(Scimago\_Subject\_Area) #%>%  
# #arrange(desc(n))   
   
# count\_awards <- BPdata %>%   
# # group\_by(Scimago\_Subject\_Area) %>%  
# count(Descr\_available) %>%  
# arrange(desc(n)) #description not available for 12 awards (5%)  
  
# BPdata %>%  
# group\_by(Awarding\_journal) %>%  
# filter(n() > 1) %>%  
# select(Awarding\_journal) %>%  
# filter(!is.na(Awarding\_journal)) %>%  
# count() #no journals with more than one award across different  
  
# BPdata %>%  
# group\_by(Awarding\_society) %>%  
# filter(n() > 1) %>%  
# select(Awarding\_society) %>%  
# filter(!is.na(Awarding\_society)) %>%  
# count() #some societies with more than one award (max.4)  
  
# BPdata %>%  
# group\_by(Awarding\_other) %>%  
# filter(n() > 0) %>%  
# select(Awarding\_other) %>%  
# filter(!is.na(Awarding\_other)) %>%  
# count() #%>% View()#list of other awarding bodies mentioned with counts  
  
# #count number of unique awards per year:  
# BPindiv %>%   
# #filter(!is.na(affiliation\_country)) %>%   
# group\_by(award\_year) %>%   
# summarise(count = n\_distinct(award\_name))   
  
##How many awards per Scimago\_Subject\_Area  
# table(BPdata$Scimago\_Subject\_Area) #note many SA with <10 extracted awards   
#- this is because of many journals being shared between SA  
  
table(BPdata$Awarding\_journal == "NA") #46 not associated with a journal  
table(BPdata$Awarding\_society == "NA") #78 not associated with a learned society  
table(BPdata$Career\_stage, useNA = "always")  
table(BPdata$Award\_individual, useNA = "always") #66 yes  
table(BPdata$Flexible\_eligibility, useNA = "always") #only 5 - "yes"   
#View(BPdata[BPdata$Flexible\_eligibility == "yes", ]) #see rows with "yes"  
  
#two-way table for the Career\_stage and Flexible\_eligibility  
table(BPdata$Career\_stage, BPdata$Flexible\_eligibility, useNA = "always") #most "not applicable" is for "any career stage"  
  
#two-way table for the Flexible\_eligibility and Award\_individual  
table(BPdata$Flexible\_eligibility, BPdata$Award\_individual, useNA = "always") #flexible ones are for ECRs only  
  
#two-way table for the Career\_stage and Award\_individual  
table(BPdata$Career\_stage, BPdata$Award\_individual, useNA = "always") #individual are mostly for ECRs  
#BPdata$Award\_name[BPdata$Career\_stage == "unclear"]  
#BPdata$Award\_name[BPdata$Career\_stage == "early-career" & BPdata$Award\_individual == "no"]  
#BPdata$Award\_name[BPdata$Career\_stage == "early-career" & BPdata$Award\_individual == "no"]  
#length(BPdata$Award\_name[BPdata$Career\_stage == "any career stage" & BPdata$Award\_individual == "no"]) #150  
  
#three-way table  
table(BPdata$Career\_stage, BPdata$Flexible\_eligibility, BPdata$Award\_individual) #most "not applicable" is for "any career stage"  
  
length(BPdata$Award\_name[BPdata$Award\_individual == "no" &   
 BPdata$Career\_stage == "any career stage" &   
 BPdata$Flexible\_eligibility == "not applicable"]) #147 not applicable flexibility  
  
length(BPdata$Award\_name[BPdata$Award\_individual == "no" &   
 BPdata$Career\_stage == "any career stage" &   
 BPdata$Flexible\_eligibility == "not available"]) #3 with no description  
  
length(BPdata$Award\_name[BPdata$Award\_individual == "no" &   
 BPdata$Career\_stage == "any career stage" &   
 BPdata$Flexible\_eligibility == "yes"]) #0 with flexibility  
  
length(BPdata$Award\_name[BPdata$Award\_individual == "yes" &   
 BPdata$Career\_stage == "any career stage"]) #28 not limited to specific career stage  
  
length(BPdata$Award\_name[BPdata$Award\_individual == "yes" &   
 BPdata$Flexible\_eligibility == "not applicable"]) #38 not applicable flexibility  
  
length(BPdata$Award\_name[BPdata$Award\_individual == "yes" &   
 BPdata$Flexible\_eligibility == "not available"]) #2 with no description  
  
length(BPdata$Award\_name[BPdata$Award\_individual == "yes" &   
 BPdata$Flexible\_eligibility == "yes"]) #5 with flexibility  
  
table(BPdata$Inclusivity\_statement, useNA = "always") #only 2 with "yes"  
  
BPdata$Award\_name[BPdata$Inclusivity\_statement == "yes"] #see rows with "yes"  
  
table(BPdata$Assessors\_transparency, useNA = "always") #114 "yes"  
#View(BPdata[BPdata$Assessors\_transparency == "yes", ]) #see rows with "yes"  
  
# sum(str\_count(tolower(BPdata$Assessor\_phrasing), "editor"), na.rm = TRUE) #162 - counting all mentions  
# sum(str\_detect(BPdata$Assessor\_phrasing, "editor"), na.rm = TRUE) #36 - counting once per award  
  
table(BPdata$Process\_transparency, useNA = "always") #only 2 "yes"  
#BPdata$Award\_name[BPdata$Process\_transparency == "yes"] #see award name with "yes"  
#View(BPdata[BPdata$Process\_transparency == "yes",]) #see award name with "yes"  
  
table(BPdata$Self\_nomination, useNA = "always") #28 yes  
#View(BPdata[BPdata$Self\_nomination == "yes", ]) #see rows with "yes"  
  
table(BPdata$Letter\_required, useNA = "always") #38 yes  
#View(BPdata[BPdata$Letter\_required == "yes", ]) #see rows with "yes"  
  
table(BPdata$Letter\_required, BPdata$Self\_nomination, useNA = "always") #two-way table  
dim(BPdata[BPdata$Letter\_required == "yes" & BPdata$Self\_nomination == "yes", ])[1] #mentioning letter and self-nomination   
dim(BPdata[BPdata$Letter\_required == "no" & BPdata$Self\_nomination == "no", ])[1] #not mentioning letter and self-nomination   
  
table(BPdata$Feedback\_availability, useNA = "always") #0 "yes"  
  
table(BPdata$Award\_contact\_provided, useNA = "always") #38 "yes"  
  
table(BPdata$Criteria\_transparency, useNA = "always") #40 yes  
#View(BPdata[BPdata$Criteria\_transparency == "yes", ]) #see rows with "yes"  
  
table(BPdata$Award\_impact\_metrics\_mentioned, useNA = "always") #21 "yes"  
  
table(BPdata$Award\_impact\_metrics\_only, useNA = "always") #8 "yes"  
  
table(BPdata$Open\_science, useNA = "always") #1 "yes"  
#View(BPdata[BPdata$Open\_science == "yes", ]) #see rows with "yes": Social Sciences: Review Of Research Award by American Educational Research Association. Two journals are considered: Review of Educational Research. Review of Research in Education  
  
table(BPdata$Awardee\_list\_most\_recent\_year, useNA = "always") #a few <2020 and 19 NA  
#View(BPdata[is.na(BPdata$Awardee\_list\_most\_recent\_year), ]) #see rows with NA  
  
table(BPdata$Awardee\_list\_earliest\_year, useNA = "always") #some old, 19 NA  
#View(BPdata[is.na(BPdata$Awardee\_list\_earliest\_year), ]) #see rows with NA  
#hist(BPdata$Awardee\_list\_earliest\_year)  
length(BPdata$Awardee\_list\_earliest\_year[BPdata$Awardee\_list\_earliest\_year > 2010 & !is.na(BPdata$Awardee\_list\_earliest\_year)] ) #107 from 2011-2023

#### Figure 1.

A. Screening effort - count numbers of journals screened to get 10 shortlisted awards:

#load journal screening summary dataset:  
BPscreening <- read\_csv(here("data","BP\_awards\_lists\_SHAREDCOPY - scimagojr\_2021\_SA.csv")) #load award screening summary data

## Rows: 29 Columns: 28  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (22): Subject\_Area, Reveiwer\_name, Reviewer\_email, Screening\_status (ass...  
## dbl (4): Nr, N\_records\_imported, N\_records\_screened, Extraction1\_N\_awards\_e...  
## lgl (2): Extraction1\_comments, Extraction2\_winners\_checking\_comments  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

BPscreening <- BPscreening[!is.na(BPscreening$Subject\_Area), ] #remove 2 extra rows without Subject\_Area  
count\_awards\_perSA <- BPdata %>% count(Scimago\_Subject\_Area) #count included awards per SA  
BPscreening$N\_included <- count\_awards\_perSA$n #add included award counts per SA  
BPscreening$N\_excluded <- BPscreening$N\_records\_screened - BPscreening$N\_included   
BPscreening$Subject\_Area <- as.factor(BPscreening$Subject\_Area) #convert to factor  
  
## reshape into long dataframe format:  
BPscreening %>% select(Subject\_Area, N\_included, N\_excluded) %>% gather(key = status, count, N\_included:N\_excluded) -> BPscreening\_long  
  
#reorder status levels and edit labels:  
BPscreening\_long <- BPscreening\_long %>%   
 mutate(status = factor(status, levels = rev(c("N\_excluded", "N\_included"))))  
  
BPscreening\_long$status <- as.character(BPscreening\_long$status) #convert to character  
#BPscreening\_long$status[BPscreening\_long$count <= 5] <- "N\_included\_low" #use to mark SA with 5 or less awards included   
  
#wacolors$san\_juan #pallette to be used  
   
figure1A <- BPscreening\_long %>%  
 mutate(status = recode(status,   
 'N\_excluded' = 'not found',   
 'N\_included' = 'found and included')) %>%  
 ggplot(aes(x = reorder(Subject\_Area, desc(Subject\_Area)),   
 y = count, fill = status)) +   
 geom\_col(width = 0.8,   
 position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0,25,50,75,100)) +  
 theme\_bw() +   
 scale\_fill\_manual(values = c("#CA884C", "#BAAF9F")) +  
 labs(x = "",   
 y = "Count of screened journals",   
 fill = "Award: ") +   
 theme(legend.position = "top",   
 legend.justification = c(0, 1),   
 legend.title = element\_text(size = rel(0.7)),   
 legend.text = element\_text(size = rel(0.7)),   
 axis.title.x = element\_text(size = 10, hjust = 0.5),  
 plot.margin = unit(c(0, 0, 0, 0), "pt")) +  
 geom\_hline(yintercept = 10,   
 linetype = "dotted",   
 color = "black",   
 size = 0.7) +  
 scale\_x\_discrete(position = "top")

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

B. Include awards with and without award description:

figure1B <- BPdata %>%  
 count(Scimago\_Subject\_Area, Descr\_available) %>%  
 arrange(desc(n)) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Descr\_available)) +   
 geom\_col(width = 0.8,   
 position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0,5,10)) +  
 theme\_bw() +   
 scale\_fill\_manual(values = c("#3A5775", "#C9DCE2")) +  
 labs(x = "",   
 y = "Count of included awards",   
 fill = "Award description:") +   
 scale\_x\_discrete(labels = NULL) +   
 labs(x = "") + #used to remove vertical labels, also breaks = NULL  
 theme(legend.position = "top",   
 legend.title = element\_text(size = rel(0.7)),   
 legend.text = element\_text(size = rel(0.7)),   
 axis.title.x = element\_text(size = 10),  
 plot.margin = unit(c(0, 0, 0, 0), "pt"))

### Award descriptions

Word counts per award description:

#add new variable with counts of words in the award description:  
BPdata$Award\_description\_wordcount <- str\_count(BPdata$Award\_description, "\\w+")  
#hist(str\_count(BPdata$Award\_description, "\\w+"), breaks = 20)  
  
BPdata[is.na(BPdata$Scimago\_Subject\_Area), ] #no NA, but no description was coded as NA, so if 1 then change to 0 (no description)

## # A tibble: 0 × 56  
## # ℹ 56 variables: Timestamp <dttm>, Extractor\_name <chr>,  
## # Scimago\_Subject\_Area <chr>, Award\_name <chr>,  
## # Source\_of\_the\_information\_on\_the\_award <chr>, Award\_description <chr>,  
## # Eligible <chr>, Comment\_on\_the\_award\_eligibility <chr>,  
## # Awarding\_journal <chr>, Awarding\_society <chr>, Awarding\_other <chr>,  
## # Comment\_on\_the\_awarding\_body <chr>, Award\_cash <chr>,  
## # Comment\_on\_the\_award\_cash <chr>, Career\_stage <chr>, …

BPdata$Award\_description\_wordcount[BPdata$Award\_description\_wordcount < 2] <- 0 #if 1 then change to 0 (no description)  
length(BPdata$Award\_description\_wordcount[BPdata$Award\_description\_wordcount == 0]) #11 awards without description (0 words)

## [1] 11

#dim(BPdata[BPdata$Award\_description\_wordcount < 100, ])[1] #90 descriptions have less than 100 words! (includes 11 with no description)

Number of awards with at least 100-word description: 79 (40 % of wards with descriptions).

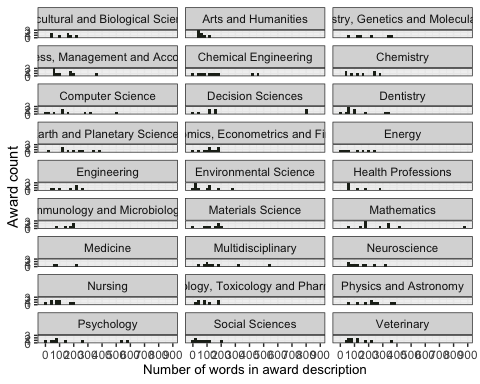
Median length of award description: 123.

Plot of award description lengths - overall:

figure1C <- BPdata %>%   
 filter(!is.na(Award\_description\_wordcount)) %>%  
 ggplot(aes(x = Award\_description\_wordcount)) +   
 geom\_histogram(binwidth = 20,   
 fill = "#21281D") +  
 theme\_bw() +   
 labs(x = "Number of words in award description",   
 y = "Award count") +   
 scale\_x\_continuous(breaks = c(seq(0, 1000, by = 100))) +  
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10))

Plot by Subject Area:

BPdata %>%   
 filter(!is.na(Award\_description\_wordcount)) %>%  
 ggplot(aes(x = Award\_description\_wordcount)) +   
 geom\_histogram(binwidth = 20,   
 fill = "#21281D") +  
 theme\_bw() +   
 labs(x = "Number of words in award description",   
 y = "Award count") +   
 scale\_x\_continuous(breaks = c(seq(0, 1000, by = 100))) +  
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10)) +   
 facet\_wrap(~Scimago\_Subject\_Area,   
 ncol = 3)



#### Figure S3

*Plot of award description lengths - by Subject Area*.

### How old are the awards?

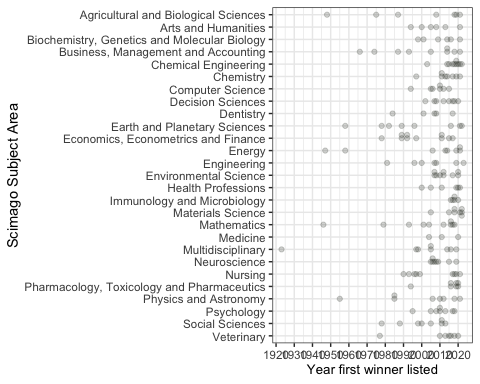
NOTE: Using the year of oldest listed past winner as a proxy.

Overall:

#table(!is.na(BPdata$Awardee\_list\_earliest\_year)) # values available  
  
#plot overall - beeswarm  
figure1D <- BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%  
 ggplot(aes(y = Awardee\_list\_earliest\_year,   
 x = Row\_excluded)) +   
 geom\_beeswarm(color = "#21281D",   
 alpha = 0.2,   
 cex = 3.3) +  
 coord\_flip() +  
 theme\_bw() +   
 labs(y = "Year first winner listed",   
 x = "") +   
 scale\_x\_discrete(labels = NULL) +   
 labs(x = "") + #used to remove vertical labels, also breaks = NULL  
 scale\_y\_continuous(breaks = c(seq(1900, 2030, by = 10))) +  
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10)) +  
 annotate("rect",   
 ymin = 2001,   
 ymax = 2022,   
 xmin = -1.5,   
 xmax = 1.5,   
 alpha = .1,  
 fill = "#008B00")

Plot by Subject Area:

#plot by SA - beeswarm  
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)), y = Awardee\_list\_earliest\_year)) +   
 geom\_beeswarm(color = "#21281D", alpha = 0.2) +  
 coord\_flip() +  
 theme\_bw() +   
 labs(y = "Year first winner listed",   
 x = "Scimago Subject Area") +   
 scale\_y\_continuous(breaks = c(seq(1900, 2030, by = 10))) +  
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10))



#### Figure S4

*Plot of award earliest extracted winner award year - by Subject Area*.

Combine 4 panels and save:

#assemble the panels using patchwork package  
figure1 <- (figure1A + figure1B) / figure1C / figure1D +   
 plot\_layout(widths = c(1, 1),   
 heights = c(3, 1, 1)) +  
 plot\_annotation(tag\_levels = "A")  
  
#ggsave(plot = figure1, here("plots", "Fig1ABCD\_screening\_descriptions\_v1.png"), width = 18.4, height = 18, units = "cm", dpi = "retina", scale = 1.2)  
#ggsave(plot = figure1, here("plots", "Fig1ABCD\_screening\_descriptions\_v1.pdf"), width = 18.4, height = 18, units = "cm", scale = 1.2)

**Summary by award age**  
Award year of the earliest listed winner used as a proxy: 1923.

Number of awards with the earliest listed winner in or before 2001: 54 (24.3 %).

Number of awards with the earliest listed winner within 2001-2023: 45 (20.3 %).

Number of awards with the earliest listed winner within 2011-2023: 107 (48.2 %).

#### Figure 2.

Figure 2A - Awarding bodies:

# table(BPdata$Awarding\_journal == "NA")  
# table(BPdata$Awarding\_society == "NA")  
# table(BPdata$Awarding\_other == "NA")  
  
#create new variable with simple categorisation of awarding body type mentioned:  
BPdata2 <- BPdata %>%  
 mutate(  
 Awarded\_by = case\_when(Awarding\_journal != "NA" & Awarding\_society == "NA" & Awarding\_other == "NA" ~ "journal",   
 Awarding\_journal == "NA" & BPdata$Awarding\_society != "NA" & BPdata$Awarding\_other == "NA" ~ "society",   
 Awarding\_journal == "NA" & BPdata$Awarding\_society == "NA" & BPdata$Awarding\_other != "NA" ~ "other",   
 Awarding\_journal != "NA" & BPdata$Awarding\_society != "NA" & BPdata$Awarding\_other == "NA" ~ "journal&society",  
 Awarding\_journal != "NA" & BPdata$Awarding\_society != "NA" & BPdata$Awarding\_other != "NA" ~ "journal&society&other",   
 Awarding\_journal != "NA" & BPdata$Awarding\_society == "NA" & BPdata$Awarding\_other != "NA" ~ "journal&other",   
 Awarding\_journal == "NA" & BPdata$Awarding\_society != "NA" & BPdata$Awarding\_other != "NA" ~ "society&other",   
 TRUE ~ "none")  
 )  
  
#table(BPdata$Awarded\_by)  
  
#create new variable for awards with description or without  
BPdata2 %>%   
 select(Award\_name, Awarding\_journal, Awarding\_society, Awarding\_other) %>%   
 mutate(  
 across(c(Awarding\_journal, Awarding\_society, Awarding\_other), ~ if\_else(. == "NA", FALSE, TRUE))  
 ) -> BPdata2  
  
#make upset plot using library(ggupset)  
figure2A <- BPdata2 %>%   
 column\_to\_rownames(var = "Award\_name") %>%   
 as\_tibble(rownames = "Award\_name") %>%  
 gather(Body, Award, -Award\_name) %>%   
 filter(Award) %>%  
 select(-Award) %>%  
 mutate(Body = recode(Body,   
 'Awarding\_journal' = 'journal',   
 'Awarding\_society' = 'society',   
 'Awarding\_other' = 'other')) %>%  
 group\_by(Award\_name) %>%  
 summarize(Awarding\_bodies = list(Body)) %>%  
 ggplot(aes(x = Awarding\_bodies)) +  
 geom\_bar(fill = "#21281D", width = 0.8) +  
 scale\_y\_continuous(limits = c(0, 200)) +  
 scale\_x\_upset(order\_by = c("freq")) +  
 theme\_combmatrix(combmatrix.panel.striped\_background = FALSE,  
 combmatrix.panel.point.color.fill = "#21281D",  
 combmatrix.panel.line.size = 0,  
 plot.title = element\_text(family = "sans", size = 20, face = "plain", color = "black"),  
 axis.title.x = element\_text(family="sans", size = 14, color = "black", face="plain"),   
 axis.title.y = element\_text(family="sans", size = 14, color = "black", face = "plain", vjust = -2),  
 ) +  
 labs(title = "A",   
 x = "Awarding bodies combination",   
 y = "Award count")

Figure 2B - target awardee career stages:

#table(BPdata$Career\_stage) #mai options: student, early-career, mid-career, any career stage, unclear  
  
#create new variable with separate career stage values as a list  
BPdata2 %>%   
 mutate(  
 Career\_stage\_list = str\_split(BPdata$Career\_stage, pattern = ", ")  
 ) -> BPdata2  
  
#BPdata$Career\_stage\_list <- str\_split(BPdata$Career\_stage, pattern = ", ") #sam as above but modifies original data frame  
  
#make upset plot using library(ggupset)  
figure2B <- BPdata2 %>%   
 ggplot(aes(x = Career\_stage\_list)) +  
 geom\_bar(fill = "#21281D",width = 0.8) +  
 scale\_y\_continuous(limits = c(0, 200)) +  
 scale\_x\_upset(order\_by = c("freq")) +  
 theme\_combmatrix(combmatrix.panel.striped\_background = FALSE,  
 combmatrix.panel.point.color.fill = "#21281D",  
 combmatrix.panel.line.size = 0,  
 plot.title = element\_text(family = "sans", size = 20, face = "plain", color = "#21281D"),  
 axis.title.x = element\_text(family="sans", size = 14, color = "#21281D", face="plain"),   
 axis.title.y = element\_text(family="sans", size = 14, color = "#21281D", face = "plain", vjust = -2),  
 ) +   
 labs(title = "B",   
 x = "Target career stages combination",   
 y = "")

Combine 2 panels and save:

#assemble the panels using patchwork package  
figure2 <- figure2A / figure2B +   
 plot\_layout(ncol = 2, nrow = 1, widths = c(1, 1)) #+ plot\_annotation(tag\_levels = "A") #does not work with this lot type  
  
#ggsave(plot = figure2, here("plots", "Fig2AB\_award\_descriptions\_v1.png"), width = 18, height = 8, units = "cm", bg = "white", dpi = "retina", scale = 1.6)  
#ggsave(plot = figure2, here("plots", "Fig2AB\_award\_descriptions\_v1.pdf"), width = 18, height = 8, units = "cm", scale = 1.6)

**Summary by awarding or funding body type:**

Total number of awards associated with a journal or a few related journals: 176 (79.3 %).

Total number of awards associated with a learned society: 144 (64.9 %).

Total number of awards associated with other organisation (e.g. a publisher, university, charity): 39 (17.6 %).

Most commonly mentioned other organisations:  
15 mention “Elsevier”, 6 mention “Wiley”).

**Summary by target career stage:**  
Total number of awards for ‘any career stage’ authors: 179 (80.6%).

Total number of awards for ‘early career’ authors: 31 (14%).

Total number of awards for ‘student’ authors: 27 (12.2%).

Total number of awards for ‘early career’ or ‘student’ authors: 20 (9%).

Total number of awards for ’mid-career’authors: 5 (2.3%).

#### Figure 3.

#BPdata %>%  
# count(Open\_science) %>%  
# ggplot(aes(x = Open\_science, y = n, fill = Open\_science)) +   
# geom\_col(width = 0.7) +  
# geom\_text(position = position\_stack(vjust = 0.5), aes(label = n)) +   
# theme\_classic() +   
# coord\_flip() +  
# labs(x = "Open Science valued", y = "Article count", fill = "Open\_science") +   
# theme(legend.position = "none", axis.title.x = element\_text(size = 10))

Figure 3A - Summary plots of key award characteristics:

BPdata %>% select(Award\_individual, Flexible\_eligibility, Inclusivity\_statement, Assessors\_transparency, Award\_integrity\_mentioned, Process\_transparency, Self\_nomination, Letter\_required, Feedback\_availability, Award\_contact\_provided) -> BPdata3A  
  
## reshape into long dataframe format:  
BPdata3A %>% gather(key = var\_name, value = value, 1:ncol(BPdata3A)) -> BPdata3Along  
  
#reorder var\_name levels  
# BPdata3Along <- BPdata3Along %>% mutate(var\_name = factor(var\_name, levels = rev(c("Award\_individual",  
# "Flexible\_eligibility",   
# "Inclusivity\_statement",   
# "Assessors\_transparency",   
# "Award\_integrity\_mentioned",   
# "Process\_transparency",  
# "Self\_nomination",   
# "Letter\_required",   
# "Feedback\_availability",   
# "Award\_contact\_provided"   
# ))))  
  
  
#reorder value levels:  
# BPdata3Along <- BPdata3Along %>% mutate(value = factor(value, levels = rev(c("no",  
# "yes",   
# "not applicable",   
# "not available"))))  
  
# colSums(BPdata3A == "no") #check how many "no" values each column has  
  
BPdata3Along <- BPdata3Along %>% mutate(var\_name = factor(var\_name, levels = rev(c("Flexible\_eligibility",   
 "Assessors\_transparency",  
 "Award\_individual",  
 "Letter\_required",   
 "Self\_nomination",   
 "Award\_contact\_provided",   
 "Award\_integrity\_mentioned",   
 "Inclusivity\_statement",   
 "Process\_transparency",  
 "Feedback\_availability"   
 ))))  
  
BPdata3Along <- BPdata3Along %>% mutate(value = factor(value, levels = rev(c("yes",   
 "not available", "not applicable",   
 "no"))))  
#wacolors$palouse #pallette to be used  
  
figure3A <- BPdata3Along %>%   
 group\_by(var\_name) %>%   
 count(var\_name, value) %>%   
 ggplot(aes(x = var\_name,   
 y = n,   
 fill = value,   
 pattern = value,   
 pattern\_angle = value)) +  
 geom\_bar\_pattern(stat = "identity",   
 pattern\_density = 0.1,  
 pattern\_spacing = 0.024,  
 pattern\_key\_scale\_factor = 0.8,  
 pattern\_alpha = 0.5,  
 pattern\_fill = 'white',  
 pattern\_colour = 'white') +  
 coord\_flip() +  
 scale\_fill\_manual(values = c("#8A6172", "#C0A43D", "#CCBA98", "#748A52")) +  
 scale\_x\_discrete(labels = rev(c("Eligibility career timeline is flexible",   
 "Assessors are revealed",   
 "Award is focused on individual authors",  
 "Nomination letter is required",   
 "Self-nomination is allowed",  
 "Award contact point is provided",  
 "Award integrity mentioned",   
 "Inclusivity statement or encouragement is provided",  
 "Process is transparent",  
 "Feedback is availabile"   
 ))) +  
 labs(x = "Award characteristics",   
 y = "Award count") +   
 theme\_classic()

Figure 3B - Summary plots of key award characteristics:

BPdata %>% select(Criteria\_transparency,   
 Award\_impact\_metrics\_mentioned,   
 Award\_impact\_metrics\_only,   
 Open\_science) -> BPdata3B  
  
## reshape into long dataframe format:  
BPdata3B %>% gather(key = var\_name, value = value, 1:ncol(BPdata3B)) -> BPdata3Blong  
#str(BPdata3long)  
  
#reorder var\_name levels  
BPdata3Blong <- BPdata3Blong %>%   
 mutate(var\_name = factor(var\_name, levels = rev(c( "Criteria\_transparency",   
 "Award\_impact\_metrics\_mentioned",   
 "Award\_impact\_metrics\_only",   
 "Open\_science"))))  
  
#reorder value levels:  
BPdata3Blong <- BPdata3Blong %>%   
 mutate(value = factor(value, levels = rev(c("yes",  
 "not available",  
 "no"))))  
figure3B <- BPdata3Blong %>%   
 group\_by(var\_name) %>%   
 count(var\_name, value) %>%   
 #ggplot(aes(x = var\_name, y = n, fill = value)) +  
 ggplot(aes(x = var\_name,   
 y = n,   
 fill = value,   
 pattern = value,   
 pattern\_angle = value)) +  
 geom\_bar\_pattern(stat = "identity",   
 pattern\_density = 0.1,   
 pattern\_key\_scale\_factor = 0.4,  
 pattern\_alpha = 0.5,  
 pattern\_fill = 'white',  
 pattern\_colour = 'white') +  
 coord\_flip() +  
 scale\_fill\_manual(values = c("#8A6172", "#CCBA98", "#748A52" )) +  
 #geom\_bar(aes(fill = value), stat = "identity") +  
 #coord\_flip() +  
 labs(x = "Award criteria",   
 y = "Award count") +   
 scale\_x\_discrete(labels = rev(c("Assessment criteria are detailed out",  
 "Impact metrics are mentioned in criteria",  
 "Impact metrics are only criteria",   
 "Open Science practices are valued"))) +  
 theme\_classic()

Combine 2 panels and save:

#assemble the panels using patchwork package  
figure3 <- figure3A / figure3B +   
 plot\_layout(ncol = 1, nrow = 2, heights = c(2, 1)) +   
 plot\_annotation(tag\_levels = "A")   
  
#ggsave(plot = figure3, here("plots", "Fig3AB\_award\_descriptions\_v1.png"), width = 18, height = 10, units = "cm", bg = "white", dpi = "retina", scale = 1.2)  
#ggsave(plot = figure3, here("plots", "Fig3AB\_award\_descriptions\_v1.pdf"), width = 18, height = 10, units = "cm", scale = 1.2)

**Summary by focus on individuals (“individual award”) or whole article (“team award”):**  
Total number of awards for individual authors: 66 (29.7%).

Total number of awards for individual authors with any career stage being eligible: 28 (42.4%).

Total number of awards for individual authors with inflexible time limits for eligibility: 21 (31.8%).

Total number of awards for individual authors with flexible time limits for eligibility: 5 (7.6%).

**Summary by basic award characteristics:**  
Total number of awards with an inclusivity statement: 2 ( 0.9%).

Total number of awards with assessor transparency: 114 ( 51.4%).

Total number of awards with assessment integrity mentioned: 19 ( 8.6%).

Total number of awards with process transparency: 2 ( 0.9%).

Total number of awards with self-nomination allowed: 28 ( 12.6%).

Total number of awards with required nomination letter: 38 ( 17.1%).

Total number of awards with feedback available: 0 ( 0%).

Total number of awards with contact details available: 38 ( 17.1%).

**Summary by award criteria characteristics:**

Total number of awards with non-vague criteria: 40 ( 18%).

Total number of awards mentioning impact metrics: 21 ( 9.5%).

Total number of awards based only on impact metrics: 8 ( 3.6%).

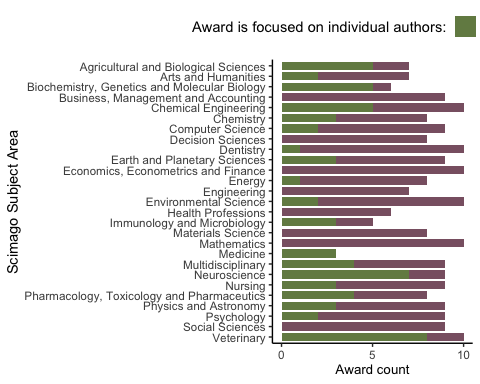
Total number of awards which value Open Science: 1 ( 0.45%).

Total number of awards with cash award: 112 ( 50.5%).

### Award characteristics by Subject Area

Plot Award\_individual by SA:

#table(BPdata$Award\_individual) #as a table - 62 are individual(28%)  
#table(BPdata$Scimago\_Subject\_Area, BPdata$Award\_individual) #as a table - 9 disciplines have 0!  
  
BPdata %>%   
 mutate(Award\_individual = factor(Award\_individual,   
 levels = c("yes",   
 "no"))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Award\_individual) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Award\_individual)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Award is focused on individual authors:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

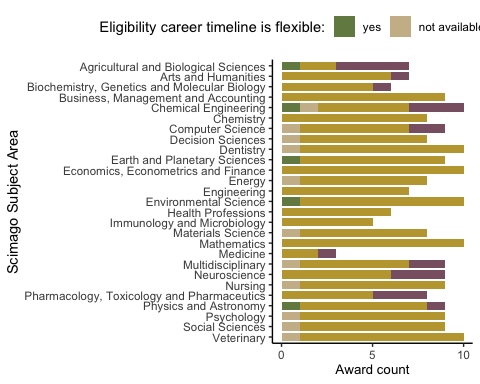


#### Figure S5

*Individual-focused awards - by Subject Area*.

Plot Flexible\_eligibility by SA:

#table(BPdata$Flexible\_eligibility, useNA = "always") #as a table  
#table(BPdata$Scimago\_Subject\_Area, BPdata$Flexible\_eligibility, useNA = "always") #as a table  
  
BPdata %>%   
 mutate(Flexible\_eligibility = factor(Flexible\_eligibility,   
 levels = c("yes",   
 "not available",   
 "not applicable",   
 "no"))) %>%  
 count(Scimago\_Subject\_Area, Flexible\_eligibility) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Flexible\_eligibility)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#C0A43D", "#8A6172")) +  
 #scale\_fill\_manual(values = c("#FA8072", "#FA807210", "#F1FFC1", "#C1FFC1")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Eligibility career timeline is flexible:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

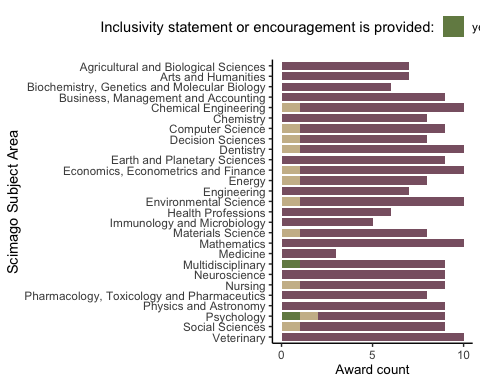


#### Figure S6

*Award eligibility timeline flexibility - by Subject Area*.

Plot Inclusivity\_statement by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Inclusivity\_statement) #as a table  
  
BPdata %>%   
 mutate(Inclusivity\_statement = factor(Inclusivity\_statement,   
 levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Inclusivity\_statement) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Inclusivity\_statement)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Inclusivity statement or encouragement is provided:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

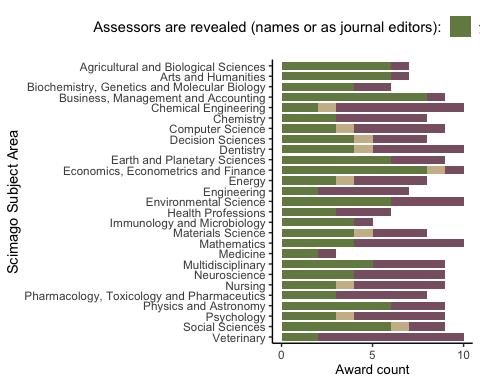


#### Figure S7

*Award inclusivity statement or encouragement - by Subject Area*.

Plot Assessors\_transparency by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Assessors\_transparency) #as a table  
  
BPdata %>%   
 mutate(Assessors\_transparency = factor(Assessors\_transparency, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Assessors\_transparency) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Assessors\_transparency)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Assessors are revealed (names or as journal editors):") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

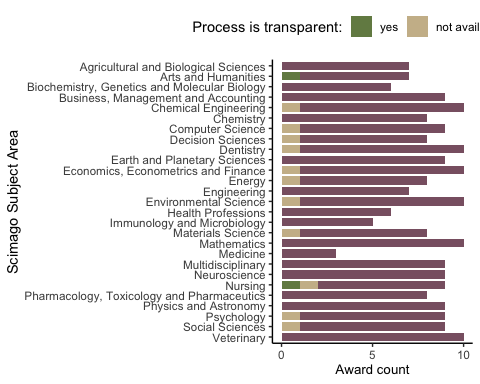


#### Figure S8

*Award assessors transparency - by Subject Area*.

Plot Process\_transparency by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Process\_transparency) #as a table  
  
BPdata %>%   
 mutate(Process\_transparency = factor(Process\_transparency, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Process\_transparency) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Process\_transparency)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Process is transparent:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

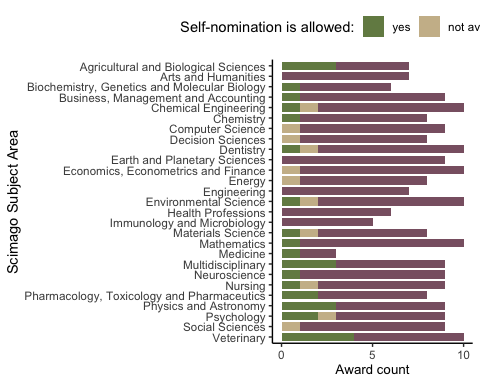


#### Figure S9

*Award process transparency - by Subject Area*.

Plot Self\_nomination by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Self\_nomination) #as a table  
  
BPdata %>%   
 mutate(Self\_nomination = factor(Self\_nomination, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Self\_nomination) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Self\_nomination)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Self-nomination is allowed:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

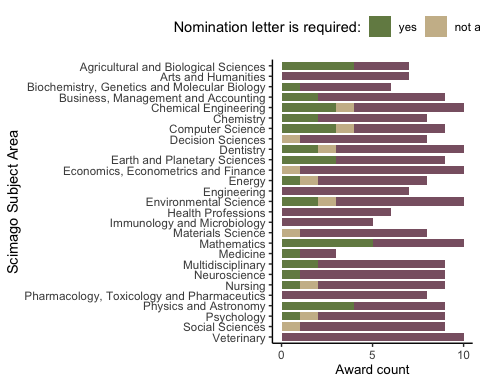


#### Figure S10

*Award self-nomination explicitly allowed - by Subject Area*.

Plot Letter\_required by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Letter\_required) #as a table  
  
BPdata %>%   
 mutate(Letter\_required = factor(Letter\_required, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Letter\_required) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Letter\_required)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Nomination letter is required:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

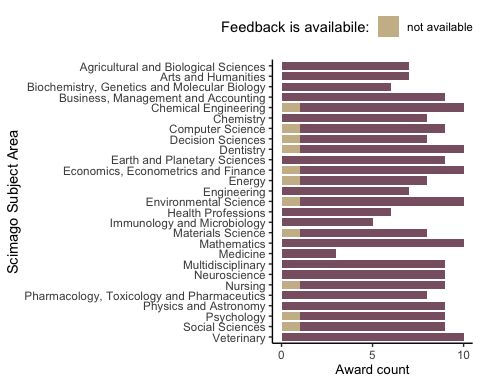


#### Figure S11

*Award nomination letter required - by Subject Area*.

Plot Feedback\_availability by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Feedback\_availability) #as a table  
  
BPdata %>%   
 mutate(Feedback\_availability = factor(Feedback\_availability, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Feedback\_availability) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),  
 y = n,   
 fill = Feedback\_availability)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Feedback is availabile:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

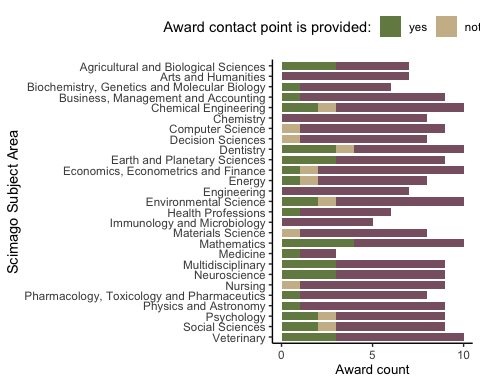


#### Figure S12

*Award feedback availability - by Subject Area*.

Plot Award\_contact\_provided by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Award\_contact\_provided) #as a table  
  
BPdata %>%   
 mutate(Award\_contact\_provided = factor(Award\_contact\_provided, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Award\_contact\_provided) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Award\_contact\_provided)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Award contact point is provided:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

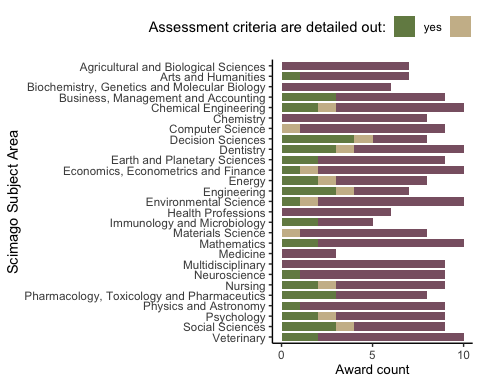


#### Figure S13

*Award contact point provided - by Subject Area*.

Plot Criteria\_transparency by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Criteria\_transparency) #as a table  
  
BPdata %>%   
 mutate(Criteria\_transparency = factor(Criteria\_transparency, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Criteria\_transparency) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Criteria\_transparency)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Assessment criteria are detailed out:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

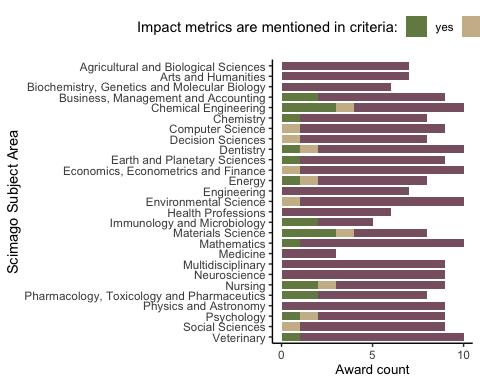


#### Figure S14

*Detailed award assessment criteria - by Subject Area*.

Plot Award\_impact\_metrics\_mentioned by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Award\_impact\_metrics\_mentioned) #as a table  
  
BPdata %>%   
 mutate(Award\_impact\_metrics\_mentioned = factor(Award\_impact\_metrics\_mentioned, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Award\_impact\_metrics\_mentioned) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Award\_impact\_metrics\_mentioned)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Impact metrics are mentioned in criteria:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

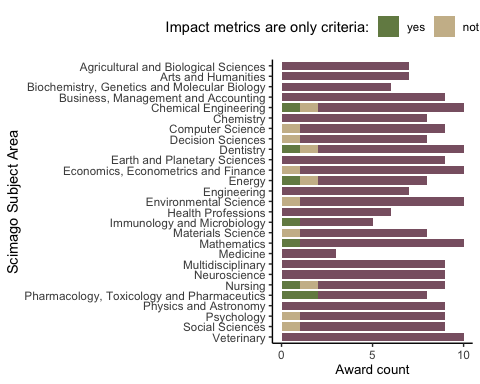


#### Figure S15

*Award assessment criteria with impact metrics - by Subject Area*.

Plot Award\_impact\_metrics\_only by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Award\_impact\_metrics\_only) #as a table  
  
BPdata %>%   
 mutate(Award\_impact\_metrics\_only = factor(Award\_impact\_metrics\_only, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Award\_impact\_metrics\_only) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Award\_impact\_metrics\_only)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Impact metrics are only criteria:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

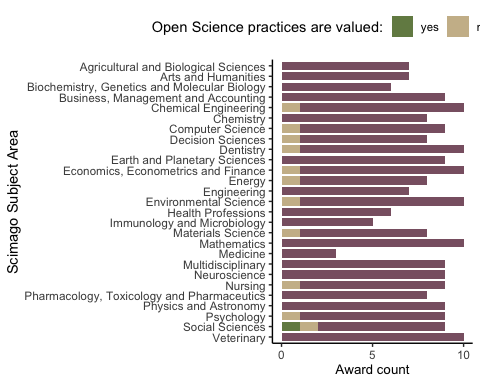


#### Figure S16

*Impact metrics as only criteria - award scores by Subject Area*.

Plot Open\_science by SA:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Open\_science) #as a table  
  
BPdata %>%   
 mutate(Open\_science = factor(Open\_science, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 count(Scimago\_Subject\_Area, Open\_science) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Open\_science)) +   
 geom\_col(width = 0.8, position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Open Science practices are valued:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))



#### Figure S17

*Open Science practices - award scores by Subject Area*.

### Text mining of award descriptions

Award descriptions - frequencies of words.

#Using packages tidytext and stopwords  
Award\_description\_txt <- tibble(txt = tolower(BPdata$Award\_description))  
Award\_description\_txt <- Award\_description\_txt %>% unnest\_tokens(output = word, input = txt, token = "words", to\_lower = TRUE) #restructure all descriptions as one-token-per-row format  
Award\_description\_txt <- Award\_description\_txt %>% anti\_join(get\_stopwords()) #remove stop words from library(stopwords)

## Joining with `by = join\_by(word)`

Award\_description\_txt <- Award\_description\_txt %>% anti\_join(get\_stopwords()) #remove stop words from library(stopwords)

## Joining with `by = join\_by(word)`

Award\_description\_txt$word <- tokenize\_word\_stems(Award\_description\_txt$word) #make all word stems lowercase  
word.freq <- Award\_description\_txt %>% count(word, sort = TRUE) #count words

Award descriptions - count mentions of specific words (stemmed).

#create list of specific words (stemmed) to count within strings  
specific.words <- c("best", "outstand", "impact", "original", "significant", "merit", "excell", "innovat", "novel", "creativ", "transparen", "reliab", "robust", "reproduc", "replica", "rigor", "rigour", "trust")  
  
#prepare award descriptions as a single lowercase string  
descriptions <- BPdata %>%   
 filter(!is.na(Award\_description)) %>%   
 select(Award\_description) %>%   
 tolower() #single lowercase string  
  
#sum of all mentions for each word  
specific.words.mentions <- specific.words %>%  
 map\_int(~ str\_count(tolower(descriptions), .x))  
  
#prepare award descriptions while keeping them separate for each award  
descriptions2 <- tolower(BPdata$Award\_description) #vector of lowercase strings  
  
#sum of mentions per award for each word (counts only one mention per award)  
specific.words.mentions2 <- specific.words %>%  
 map\_int(~ sum(str\_detect(descriptions2, .x), na.rm = TRUE))  
  
# ## doing the same as above, but manually:  
# #count all mentions of words(parts) individually, e.g.:  
# sum(str\_count(tolower(BPdata$Award\_description), "best"), na.rm = TRUE) #169  
# #counting once per award, e.g.   
# sum(str\_detect(BPdata$Award\_description, "best"), na.rm = TRUE) #82

#### Figure 4.

Plot frequencies of specific words - all mentions in award descriptions:

words.df <- tibble(Words = specific.words,   
 Count\_all = specific.words.mentions,  
 Count\_once = specific.words.mentions2)  
  
#wacolors$palouse #pallette to be used  
  
figure4A <- words.df %>%  
 ggplot(aes(x = reorder(Words, Count\_all),   
 y = Count\_all)) +   
 geom\_col(width = 0.8,   
 fill = "#2D3F4A") +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 50, 100, 150),   
 limits = c(0, 200)) +  
 theme\_bw() +   
 labs(x = "Word stem", y = "Count of all mentions") +   
 annotate(geom = "rect",xmin = 8.5, xmax = Inf, ymin = -Inf, ymax = Inf, fill = "#8A6172", alpha = 0.2) +  
 annotate(geom = "rect",xmin = -Inf, xmax = 8.5, ymin = -Inf, ymax = Inf, fill = "#748A52", alpha = 0.2) +  
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10))

Plot frequencies of specific words - first mentions in award descriptions:

figure4B <- words.df %>%  
 ggplot(aes(x = reorder(Words, Count\_all),   
 y = Count\_once)) +   
 geom\_col(width = 0.8,   
 fill = "#2D3F4A",   
 alpha = 0.7) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 50, 100, 150),   
 limits = c(0, 200)) +  
 theme\_bw() +   
 labs(x = "Word stem",   
 y = "Count of first mention per award") +   
 scale\_x\_discrete(labels = NULL) +   
 labs(x = "") + #used to remove vertical labels, also breaks = NULL  
 annotate(geom = "rect",xmin = 8.5, xmax = Inf, ymin = -Inf, ymax = Inf, fill = "#8A6172", alpha = 0.2) +  
 annotate(geom = "rect",xmin = -Inf, xmax = 8.5, ymin = -Inf, ymax = Inf, fill = "#748A52", alpha = 0.2) +  
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10))

Combine 2 panels and save:

#assemble the panels using patchwork package  
figure4 <- figure4A / figure4B +   
 plot\_layout(ncol = 2, nrow = 1) +  
 plot\_annotation(tag\_levels = "A")  
#ggsave(plot = figure4, here("plots", "Fig4AB\_words\_counts\_v1.png"), width = 18, height = 8, units = "cm", dpi = "retina", scale = 1.2)  
#ggsave(plot = figure4, here("plots", "Fig4AB\_words\_counts\_v1.pdf"), width = 18, height = 8, units = "cm", scale = 1.2)

#### Characteristics of the past winners.

length(unique(BPdata$Award\_name[BPdata$N\_winners\_extracted != 0])) #61 - number from the records in the award-level data set

## [1] 61

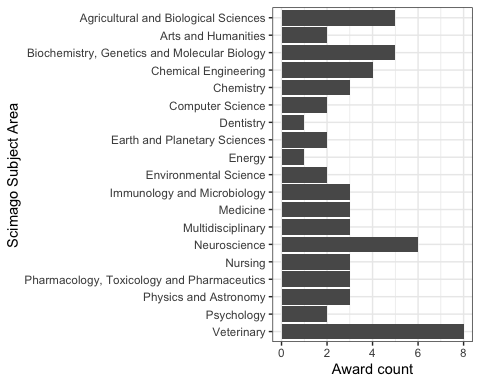
length(unique(BPindiv$award\_name)) #61 - number from records in past winner's data (matching)

## [1] 61

Total number of extracted awards: 61.

Plot numbers of awards extracted per SA:

#from BPdata:  
BPdata %>%   
 group\_by(Scimago\_Subject\_Area) %>%   
 filter(N\_winners\_extracted != 0) %>%   
 count(Scimago\_Subject\_Area) -> awards\_perSA  
  
# #from BPindiv:  
# BPindiv %>%   
# group\_by(award\_SA) %>%  
# summarise(count = n\_distinct(award\_name))  
  
# #from BPindiv:  
# BPindiv %>%  
# group\_by(award\_SA) %>%  
# summarise(count = n\_distinct(award\_name)) %>% summarise(mean(count)) #3  
  
  
#plot counts where >0 awards per SA  
p\_count.awards.SA <- BPindiv %>%   
 group\_by(award\_SA) %>%  
 summarise(count = n\_distinct(award\_name)) %>%   
 ggplot(aes(x = reorder(award\_SA, desc(award\_SA)),   
 y = count)) +  
 geom\_bar(stat = "identity",   
 position = position\_dodge(0.9)) +  
 coord\_flip() +  
 theme\_bw() +  
 labs(x = "Scimago Subject Area",   
 y = "Award count")   
  
p\_count.awards.SA



#### Figure S18

*Number of awards with extracted individual winners data - counts by Subject Area*.

Tital number of extracted records (awardee names): rnrow(BPindiv)`.

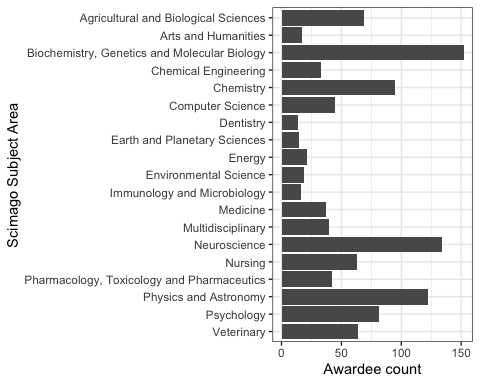
#sum(BPdata$N\_winners\_extracted, na.rm = TRUE) #number from the records in the award-level data set  
#dim(BPindiv)[1] #number of records in past winner's data (matching)

Plot numbers of records (names) extracted per SA:

#from BPdata:  
BPdata %>%   
 group\_by(Scimago\_Subject\_Area) %>%   
 select(N\_winners\_extracted) %>%   
 filter(!is.na(N\_winners\_extracted)) %>%   
 summarise(sum\_winners = sum(N\_winners\_extracted), .groups = 'drop') -> winners\_perSA1

## Adding missing grouping variables: `Scimago\_Subject\_Area`

#dim(winners\_perSA1[winners\_perSA1$sum\_winners != 0, ])[1] #number of SA with any extracted names  
#winners\_perSA1[winners\_perSA1$sum\_winners != 0, ] #table of SA with any extracted names  
  
#from BPindiv:  
#table(BPindiv$award\_SA)  
#dim(table(BPindiv$award\_SA, useNA = "always")) #19 SA with any extracted names  
BPindiv %>%  
 group\_by(award\_SA) %>% count() -> winners\_perSA1\_extracted #does not show SA with 0 extracted  
  
#median(winners\_perSA1\_extracted$n) #42  
  
#plot counts where >0 winners per SA  
p\_count.awardees.SA <- BPindiv %>%   
 group\_by(award\_SA) %>%   
 count() %>%  
 ggplot(aes(x = reorder(award\_SA, desc(award\_SA)),   
 y = n)) +  
 geom\_bar(stat = "identity",   
 position = position\_dodge(0.9)) +  
 coord\_flip() +  
 theme\_bw() +  
 #scale\_x\_discrete(labels = NULL) + labs(x = "") + #used to remove vertical labels, also breaks = NULL  
 labs(x = "Scimago Subject Area",   
 y = "Awardee count")   
  
p\_count.awardees.SA



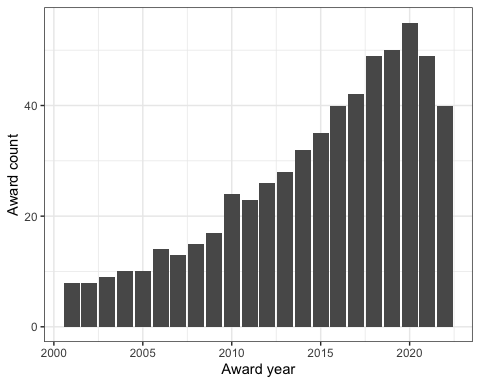
#### Figure S19

*Numbers of extracted individual winners - counts by Subject Area*.

Number of SA eithout any extracted individual winners? - 8.

Plot numbers of awards with individual winner data per year:

BPindiv %>%   
 group\_by(award\_year) %>%   
 summarise(n = n\_distinct(award\_name)) %>%  
 ggplot(aes(x = award\_year,   
 y = n)) +  
 geom\_bar(stat = "identity",   
 position = position\_dodge(0.9)) +  
 theme\_bw() +  
 labs(x = "Award year", y = "Award count")

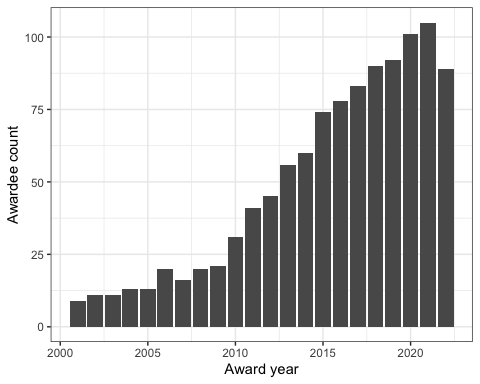


#### Figure S20

*Number of extracted awards with individual winners names per year - counts for all Subject Areas*.

Plot numbers of individual winner records per year:

#table(BPindiv$award\_year, useNA = "always")  
  
BPindiv %>%   
 group\_by(award\_year) %>%   
 count() %>%  
 ggplot(aes(x = award\_year,   
 y = n)) +  
 geom\_bar(stat = "identity",   
 position = position\_dodge(0.9)) +  
 theme\_bw() +  
 labs(x = "Award year",   
 y = "Awardee count")



#### Figure S21

*Number of extracted individual winners per year - counts for all Subject Areas*.

Number of individual winner records per decade:

#create decade variable 2001-2010, 2011-2020, 2021-2022  
BPindiv %>% mutate(Decade = case\_when(  
 award\_year >= 2001 & award\_year <= 2010 ~ "2001-2010",  
 award\_year >= 2011 & award\_year <= 2020 ~ "2011-2020",  
 award\_year >= 2021 & award\_year <= 2022 ~ "2021-2022",  
 award\_year == "Veterinary" ~ "Biology",  
 TRUE ~ NA)  
 ) -> BPindiv  
  
table(BPindiv$Decade, useNA = "always")

##   
## 2001-2010 2011-2020 2021-2022 <NA>   
## 165 720 194 0

Counts by decade: - 2001-2010: 165.  
- 2011-2020: 720.  
- 2021-2022: 194.

#table(BPindiv$shared, useNA = "always") #156 shared (14% of names), usually shared by 2 authors)  
#table(BPindiv$Decade, BPindiv$shared) #more common recently  
#table(BPindiv$award\_SA, BPindiv$shared) # sharing most frequent in Computer Science

Number of individuals sharing awards: 156 (NA%).

NOTE: during the first round of data extraction, we located around 3/4 of winning articles - these were articles where title or DOI was available on award page or announcement. However, article information was sometimes only publicly listed for recently awarded articles. Our additional Internet searches increased the number of identified articles to 796.

Number of awards with article information (reference): 880 (81.6%).

Number of awards with article information (DOI): 879 (81.5%).

Number of awards with article information (reference or DOI): 879 (81.5%).

### Awardee gender

Source of information on gender:

* Pronouns: 187 (17.3%).
* Photo: 265 (24.6%).
* Name only: 627 (24.6%).

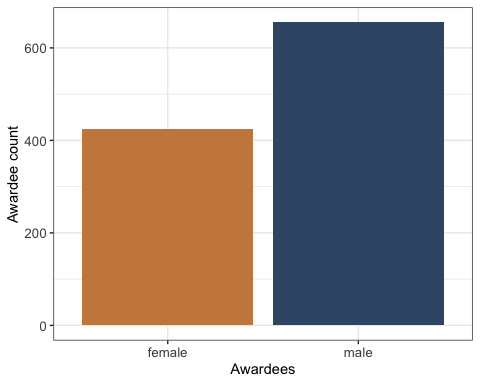
Total number of past awardees without gender assigned: 0.

Total number of past awardees assigned as a woman: 424 (39.3%).

Total number of past awardees assigned as a man: 655 (60.7%).

Plot gender counts overall:

#wacolors$san\_juan  
  
## set the levels in order we want:  
BPindiv$Gender <- factor(BPindiv$awardee\_gender)  
levels(BPindiv$Gender) <- c("female", "male")  
  
## plot for a all SA and years - not stacked, with count values:  
BPindiv %>%   
# group\_by(award\_SA) %>%  
 count(Gender) %>%  
 ggplot(aes(x = Gender,   
 y = n)) +  
 geom\_bar(aes(fill = Gender),   
 stat = "identity",   
 position = position\_dodge(0.9)) +  
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 theme\_bw() +  
 theme(legend.position = "none",   
 legend.box = "horizontal",   
 axis.text = element\_text(size = 10)) +  
 labs(fill = "Awardee gender:") +  
 labs(x = "Awardees",   
 y = "Awardee count")

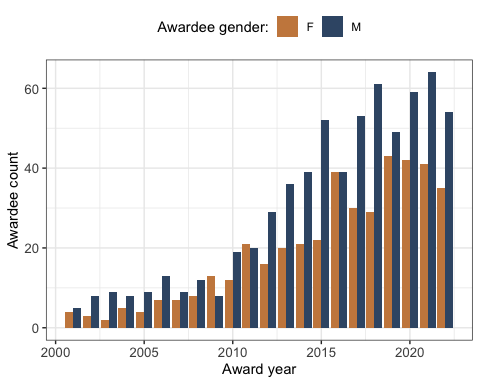


#### Figure S22

*Counts of the past awardees (across all Subject Areas) assigned as female or male based on their pronouns, photos and names. F = female / women, M = male / men*.

Plot gender counts by year:

#wacolors$san\_juan  
  
BPindiv %>%   
 group\_by(award\_year) %>%  
 count(awardee\_gender) %>%  
 ggplot(aes(x = award\_year,   
 y = n,   
 fill = awardee\_gender)) +  
 geom\_bar(stat = "identity", position = position\_dodge(0.9)) +  
 theme\_bw() +   
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 theme(legend.position = "top",   
 legend.box = "horizontal",   
 axis.text = element\_text(size = 10)) +  
 labs(fill = "Awardee gender:") +  
 labs(x = "Award year",   
 y = "Awardee count")

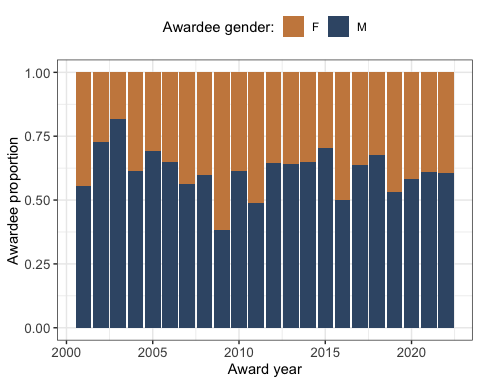


#### Figure S23

*Counts of the past awardees (across all Subject Areas) assigned as female or male based on their pronouns, photos and names - by year. F = female / women, M = male / men*.

Plot gender proportion by year:

#wacolors$san\_juan  
  
BPindiv %>%   
 group\_by(award\_year) %>%  
 count(awardee\_gender) %>%  
 ggplot(aes(x = award\_year,   
 y = n,   
 fill = awardee\_gender)) +  
 geom\_bar(stat = "identity",   
 position = "fill") +  
 theme\_bw() +   
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 theme(legend.position = "top",   
 legend.box = "horizontal",   
 axis.text = element\_text(size = 10)) +  
 labs(fill = "Awardee gender:") +  
 labs(x = "Award year",   
 y = "Awardee proportion")



#### Figure S24

*Proportions of the past awardees (across all Subject Areas) assigned as female or male based on their pronouns, photos and names - by year. F = female / women, M = male / men*.

Figure 5A - plot of gender counts by decade:

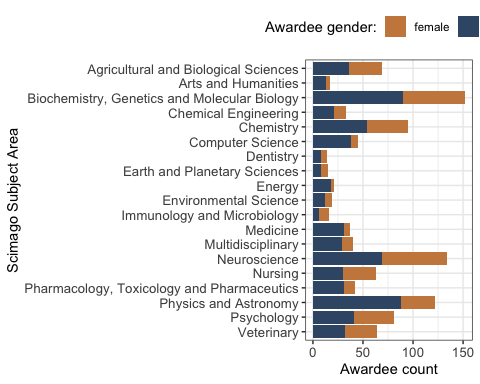
##See the actual counts  
# BPindiv %>%   
# group\_by(Decade) %>%  
# count(awardee\_gender) %>%  
# mutate(decade\_n = sum(n)) %>%  
# mutate(pct = n / decade\_n \*100)  
  
## plot for all Subject Areas by Decade - horizontal, not stacked, with count values:  
figure5A <- BPindiv %>%   
 group\_by(Decade) %>%  
 count(Gender) %>%  
 ggplot(aes(y = Decade,   
 x = n,   
 fill = Gender)) +  
 geom\_bar(aes(fill = Gender),   
 stat = "identity") +   
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 theme\_bw() +  
 theme(legend.position="top",   
 legend.text = element\_text(size = 8),   
 legend.title = element\_text(size = 8),  
 axis.text = element\_text(size = 10),   
 legend.box = "horizontal",   
 legend.margin = margin()) +  
 guides(fill = guide\_legend(nrow = 1, byrow = TRUE, reverse = TRUE)) +   
 labs(fill = "Awardee gender:") +  
 labs(y = "Decade",   
 x = "Awardee counts")

Figure 5B - plot of gender proportions by decade:

## plot for a all Subject Areas by Decade - horizontal, stacked, with proportion values:  
figure5B <- BPindiv %>%   
 group\_by(Decade) %>%  
 count(Gender) %>%  
 ggplot(aes(x = Decade,   
 y = n,   
 fill = Gender)) +  
 geom\_bar(aes(fill = Gender),   
 stat = "identity",   
 position = "fill") + # use , position = "fill" for proportion plot  
 coord\_flip() +   
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 theme\_bw() +  
 theme(legend.position = "none",   
 legend.box = "horizontal",   
 axis.text = element\_text(size = 10)) +  
 labs(fill = "Awardee gender:") +  
 labs(x = "Decade",   
 y = "Awardee proportion")

Plot gender counts by Subject Area:

##See the counts by Subject Area  
# BPindiv %>%   
# group\_by(award\_SA) %>%  
# count(awardee\_gender) %>%  
# mutate(SA\_n = sum(n)) %>%  
# mutate(pct = n / SA\_n \*100)  
  
## plot for a all awards by SA - not stacked, with count values:  
BPindiv %>%   
 group\_by(award\_SA) %>%  
 count(Gender) %>%  
 ggplot(aes(x = reorder(award\_SA, desc(award\_SA)),   
 y = n,   
 fill = Gender)) +  
 geom\_bar(aes(fill = Gender),   
 stat = "identity") + # use , position = "fill" for proportion plot  
 coord\_flip() +   
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 theme\_bw() +  
 theme(legend.position="top",   
 legend.box = "horizontal",   
 axis.text = element\_text(size = 10)) +  
 labs(fill = "Awardee gender:") +  
 labs(x = "Scimago Subject Area",   
 y = "Awardee count")

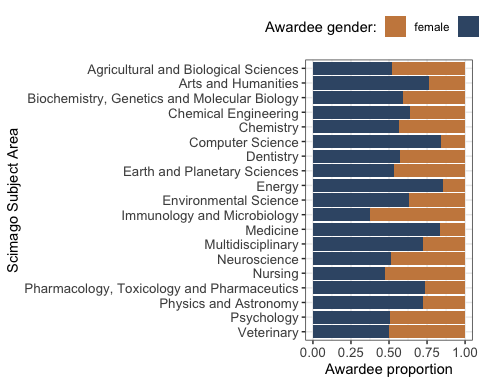


#### Figure S25

*Counts of the past awardees assigned as female or male based on their pronouns, photos and names - by Subject Area. F = female / women, M = male / men*.

Plot gender proportions by Subject Area:

## plot for a all awards by SA -stacked, with proportion values:  
BPindiv %>%   
 group\_by(award\_SA) %>%  
 count(Gender) %>%  
 ggplot(aes(x = reorder(award\_SA, desc(award\_SA)),   
 y = n,   
 fill = Gender)) +  
 geom\_bar(aes(fill = Gender),   
 stat = "identity",   
 position = "fill") +   
 coord\_flip() +   
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 theme\_bw() +  
 theme(legend.position="top",   
 legend.box = "horizontal",   
 axis.text = element\_text(size = 10)) +  
 labs(fill = "Awardee gender:") +  
 labs(x = "Scimago Subject Area",   
 y = "Awardee proportion")

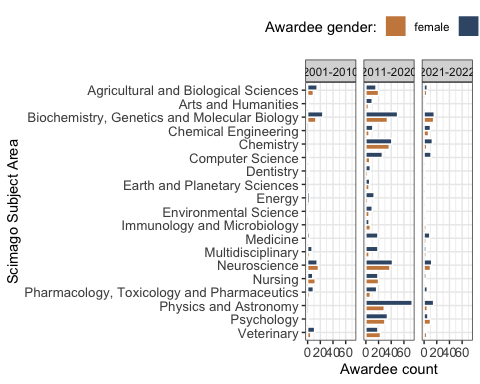


#### Figure S26

*Proportions of the past awardees assigned as female or male based on their pronouns, photos and names - by Subject Area. F = female / women, M = male / men*.

Plot awardee gender counts by Subject Area and Decade:

## plot for a all awards by SA and Decade - not stacked, with count values:  
BPindiv %>%   
 group\_by(award\_SA, Decade) %>%  
 count(Gender) %>%  
 ggplot(aes(x = reorder(award\_SA, desc(award\_SA)),   
 y = n,   
 fill = Gender)) +  
 geom\_bar(aes(fill = Gender),   
 stat = "identity",   
 position = position\_dodge(0.8),   
 width = 0.8,   
 col = "white") + # use , position = "fill" for proportion plot  
 facet\_wrap(~Decade, scales = "fixed", nrow = 1, ncol = 3) +  
 coord\_flip() +   
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 theme\_bw() +  
 theme(legend.position="top",   
 legend.box = "horizontal",   
 axis.text = element\_text(size = 10)) +  
 labs(fill = "Awardee gender:") +  
 labs(x = "Scimago Subject Area",   
 y = "Awardee count")



#### Figure S27

*Counts of the past awardees assigned as female or male based on their pronouns, photos and names - by Subject Area and Decade. F = female / women, M = male / men*.

### Affiliation country.

Number of past awardees without affiliation country info: 54 (5%).

Source of affiliation information:

* Award page: 355 (32.9%).
* Award announcement:  
  159 (14.7%).
* Winning article:  
  482 (44.7%).
* Other:  
  30 (2.8%).

#count individuals with affiliations:  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 nrow() -> affiliation\_indiv\_count  
  
#count number of unique countries overall:  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 summarise(count = n\_distinct(affiliation\_country)) -> affiliation\_country\_count  
  
#count number of unique countries per decade:  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 group\_by(Decade) %>%   
 summarise(count = n\_distinct(affiliation\_country))

## # A tibble: 3 × 2  
## Decade count  
## <chr> <int>  
## 1 2001-2010 26  
## 2 2011-2020 39  
## 3 2021-2022 28

#count number of unique countries per year:  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 group\_by(award\_year) %>%   
 summarise(count = n\_distinct(affiliation\_country))

## # A tibble: 22 × 2  
## award\_year count  
## <dbl> <int>  
## 1 2001 3  
## 2 2002 5  
## 3 2003 5  
## 4 2004 6  
## 5 2005 4  
## 6 2006 7  
## 7 2007 6  
## 8 2008 8  
## 9 2009 7  
## 10 2010 12  
## # ℹ 12 more rows

#count by countries:  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 count(affiliation\_country) %>%   
 arrange(desc(n)) -> country\_count  
  
#count for USA only:  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 filter(affiliation\_country == "USA") %>%   
 nrow() -> affiliation\_country\_USA  
  
# #count countries - view top 20:  
# BPindiv %>%   
# filter(!is.na(affiliation\_country)) %>%   
# count(affiliation\_country) %>%   
# arrange(desc(n)) %>%  
# filter(n>5) %>%  
# View()  
  
#recode top 10 countries + other  
BPindiv$Country10 <- recode(BPindiv$affiliation\_country,   
 "USA" = "USA",   
 "UK" = "UK",   
 "Australia" = "Australia",   
 "China" = "China",   
 "Germany" = "Germany",   
 "Canada" = "Canada",   
 "Japan" = "Japan",   
 "Netherlands" = "Netherlands",   
 "Italy" = "Italy",   
 "France" = "France",   
 .default = "other")  
#table(BPindiv$Country10)  
  
#recode countries with n>5 (20)  
#BPindiv$Country20 <- recode(BPindiv$affiliation\_country,   
 # "USA" = "USA",   
 # "UK" = "UK",   
 # "Australia" = "Australia",   
 # "China" = "China",   
 # "Germany" = "Germany",   
 # "Canada" = "Canada",   
 # "Japan" = "Japan",   
 # "Netherlands" = "Netherlands",   
 # "Italy" = "Italy",   
 # "France" = "France",   
 # "Austria" = "Austria",   
 # "Spain" = "Spain",   
 # "Sweden" = "Sweden",   
 # "Switzerland" = "Switzerland",   
 # "India" = "India",   
 # "South Korea" = "South Korea",   
 # "Brazil" = "Brazil",   
 # "Singapore" = "Singapore",   
 # "Taiwan" = "Taiwan",   
 # "Belgium" = "Belgium",  
 # .default = "other")  
#table(BPindiv$Country20)  
  
## make a simple table with percentages of the total number of Documents  
# BPindiv %>%   
# filter(!is.na(affiliation\_country)) %>%   
# count(Country10) %>%   
# arrange(desc(n)) %>%   
# mutate(Country\_pct = n/sum(n) \* 100)

Affiliation country composition:

Out of 1079 extracted individual winners, 1025 (95%) had information on their affiliation country.

Out of 1025 extracted individual winners with affiliation country information, 487 (47.5%) had USA as their affiliation country.

Data on Global South nations (United Nations FCFSSC) among past awardees.

GS\_nations <- c("India",   
 "China",   
 "Indonesia",   
 "Pakistan",   
 "Nigeria",   
 "Brazil",   
 "Bangladesh",   
 "Ethiopia",   
 "Philippines",   
 "Egypt",   
 "DR Congo",   
 "Vietnam",   
 "Iran",   
 "Thailand",   
 "Tanzania",   
 "South Africa",   
 "Kenya",   
 "Myanmar",   
 "Colombia",   
 "Uganda",   
 "Sudan",   
 "Argentina",   
 "Algeria",   
 "Iraq",   
 "Afghanistan",   
 "Morocco",   
 "Saudi Arabia",   
 "Angola",   
 "Yemen",   
 "Peru",   
 "Malaysia",   
 "Ghana",   
 "Mozambique",   
 "Nepal",   
 "Madagascar",   
 "Ivory Coast",   
 "Venezuela",   
 "Cameroon",   
 "Niger",   
 "North Korea",   
 "Mali",   
 "Burkina Faso",   
 "Syria",   
 "Sri Lanka",   
 "Malawi",   
 "Zambia",   
 "Chile",   
 "Chad",   
 "Ecuador",   
 "Somalia",   
 "Guatemala",   
 "Senegal",   
 "Cambodia",   
 "Zimbabwe",   
 "Guinea",   
 "Rwanda",   
 "Benin",   
 "Burundi",   
 "Tunisia",   
 "Bolivia",   
 "Haiti",   
 "Jordan",   
 "Dominican Republic",   
 "Cuba",   
 "South Sudan",   
 "Honduras",   
 "Papua New Guinea",   
 "Tajikistan",   
 "United Arab Emirates",   
 "Togo",   
 "Sierra Leone",   
 "Laos",   
 "Nicaragua",   
 "Libya",   
 "Paraguay",   
 "Turkmenistan",   
 "El Salvador",   
 "Republic Of The Congo",   
 "Singapore",   
 "Central African Republic",   
 "Liberia",   
 "Palestine",   
 "Lebanon",   
 "Costa Rica",   
 "Mauritania",   
 "Oman",   
 "Panama",   
 "Kuwait",   
 "Eritrea",   
 "Mongolia",   
 "Uruguay",   
 "Bosnia And Herzegovina",   
 "Jamaica",   
 "Gambia",   
 "Qatar",   
 "Botswana",   
 "Namibia",   
 "Gabon",   
 "Lesotho",   
 "Guinea Bissau",   
 "Equatorial Guinea",   
 "Trinidad And Tobago",   
 "Bahrain",   
 "Timor Leste",   
 "Mauritius",   
 "Eswatini",   
 "Djibouti",   
 "Fiji",   
 "Comoros",   
 "Guyana",   
 "Bhutan",   
 "Solomon Islands",   
 "Suriname",   
 "Cape Verde",   
 "Maldives",   
 "Brunei",   
 "Bahamas",   
 "Belize",   
 "Vanuatu",   
 "Barbados",   
 "Sao Tome And Principe",   
 "Samoa",   
 "Saint Lucia",   
 "Kiribati",   
 "Grenada",   
 "Micronesia",   
 "Tonga",   
 "Seychelles",   
 "Saint Vincent And The Grenadines",   
 "Antigua And Barbuda",   
 "Dominica",   
 "Saint Kitts And Nevis",   
 "Marshall Islands",   
 "Nauru"  
)  
  
  
#count Global South countries in the data set:  
# BPindiv %>%   
# filter(!is.na(affiliation\_country)) %>%   
# filter(affiliation\_country %in% GS\_nations) %>%  
# count(affiliation\_country) %>%   
# arrange(desc(n)) #%>% View() #table of GS nations and winner counts  
  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>% nrow() -> n\_winners\_affil #number of winners with affiliation country  
  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 filter(affiliation\_country %in% GS\_nations) %>%  
 count(affiliation\_country) %>%   
 arrange(desc(n)) %>% count() -> n\_winners\_GS\_countries #number of countries from GS  
  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 filter(affiliation\_country %in% GS\_nations) %>%  
 count(affiliation\_country) %>%   
 arrange(desc(n)) %>% select(n) %>% sum() -> n\_winners\_GS\_indiv #number of individual winners from GS  
  
  
# Table of Count of winners by their country of affiliation   
# country\_count %>%   
# mutate(Global\_South = ifelse(affiliation\_country %in% GS\_nations, "yes", "no")) %>%   
# kable("html") %>%  
# kable\_styling("striped", position = "left") %>%  
# scroll\_box(width = "100%", height = "1500px")

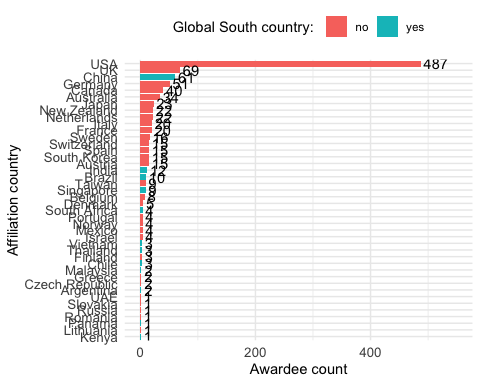
Number of awardees from the Global South:

Out of 1079 extracted individual winners, 111 (10.8%) were affiliated to a country classified as the Global South.

The number of countries classified as the Global South: 12 (27.9% of the all detected countries).

Plot awardee counts by country affiliations overall for all countries:

#all countries as a simple barplot  
BPindiv %>%   
 filter(!is.na(affiliation\_country)) %>%   
 count(affiliation\_country) %>%  
 arrange((n)) %>%  
 mutate(Global\_South = ifelse(affiliation\_country %in% GS\_nations,   
 "yes",   
 "no")) %>%   
 ggplot(aes(x = reorder(affiliation\_country, n),   
 y = n,  
 fill = Global\_South)) +  
 geom\_bar(stat = "identity",   
 position = position\_dodge(0.9)) +  
 theme\_minimal() +   
 geom\_text(aes(label = n),   
 vjust = 0.5,   
 hjust = 0,   
 nudge\_y = 5) + #add counts #add GS status  
 scale\_y\_continuous(limits = c(0, 550)) +  
 coord\_flip() +  
 theme(legend.position="top",   
 legend.box = "horizontal",   
 axis.text = element\_text(size = 10)) +  
 labs(fill = "Global South country: ") +  
 labs(x = "Affiliation country",   
 y = "Awardee count")

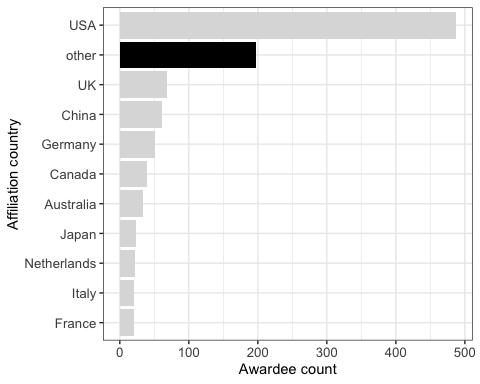


#### Figure S28

*Countries of first affiliation of the past awardees*.

Plot awardee counts for only top 10 countries with most awardees:

#top 10 countries - two-colour barplot  
BPindiv %>%   
 filter(!is.na(Country10)) %>%   
 count(Country10) %>%  
 arrange((n)) %>%  
 ggplot(aes(x = reorder(Country10, n), y = n)) +  
 geom\_bar(aes(fill= Country10 == "other"),   
 stat = "identity",   
 position = position\_dodge(0.9)) +  
 scale\_fill\_manual(values = c("#DCDCDCFF", "#000000")) +  
 coord\_flip() +   
 theme\_bw() +  
 theme(legend.position="none",   
 axis.text = element\_text(size = 10)) +  
 labs(x = "Affiliation country",   
 y = "Awardee count")

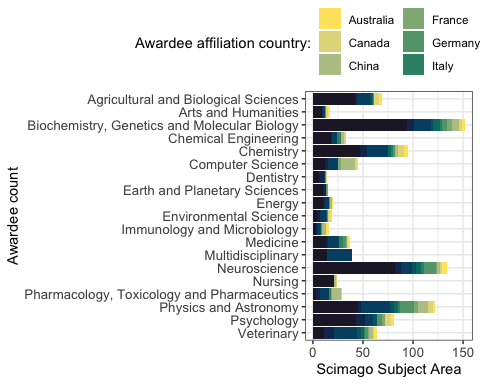


#### Figure S29

*Countries of first affiliation of the past awardees - top 10 countries only (remaining countries grouped as “other” and coloured black)*.

Plot country of affiliation counts by Subject Area:

#top 10 countries by SA  
BPindiv %>%   
 filter(!is.na(Country10)) %>%   
 count(award\_SA, Country10) %>%   
 ggplot(aes(y = reorder(award\_SA, desc(award\_SA)),   
 x = n,   
 fill = Country)) +  
 #geom\_bar(stat = "identity")  
 geom\_bar(aes(fill = Country10),   
 stat = "identity") +  
 scale\_fill\_manual(values = wa\_pal("ferries", 11, reverse = TRUE)) +  
 theme\_bw() +   
 theme(legend.position="top",   
 axis.text = element\_text(size = 10),   
 legend.box = "horizontal",   
 legend.margin = margin()) +  
 labs(fill = "Awardee affiliation country:") +  
 labs(x = "Scimago Subject Area",   
 y = "Awardee count")



#ggsave(here("plots", "awardee\_country\_SA\_v1.png"), width = 18, height = 8, units = "cm", dpi = "retina")  
#ggsave(here("plots", "awardee\_country\_SA\_v1.pdf"), width = 18, height = 8, units = "cm")

#### Figure S30

*Countries of first affiliation of the past awardees - by Subject Area (remaining countries grouped as “other”)*.

Figure 5C - plot of counts of top 10 country affiliations by decade:

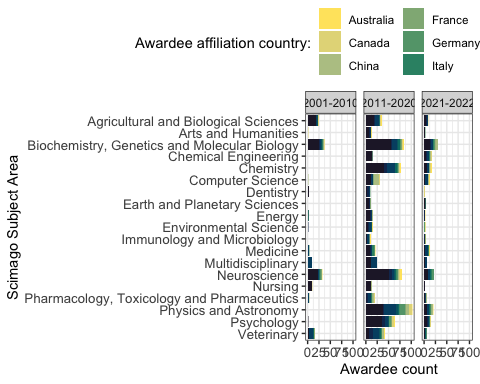
#top 10 countries by Decade - horizontal, stacked, with count values  
figure5C <- BPindiv %>%   
 filter(!is.na(Country10)) %>%   
 count(Decade, Country10) %>%   
 ggplot(aes(y = Decade,   
 x = n,   
 fill = Country10)) +  
 geom\_bar(aes(fill = Country10),   
 stat = "identity") + # use , position = "fill" for proportion plot  
 scale\_fill\_manual(values = wa\_pal("ferries", 11, reverse = TRUE)) +  
 theme\_bw() +  
 theme(legend.position="top",   
 legend.text = element\_text(size = 8),   
 legend.title = element\_text(size = 8),  
 axis.text = element\_text(size = 10),   
 legend.box = "horizontal",   
 legend.margin = margin()) +  
 guides(fill = guide\_legend(nrow = 2, byrow = TRUE, reverse = TRUE)) +   
 labs(fill = "Awardee affiliation country:") +  
 labs(y = "Decade",   
 x = "Awardee counts")

Figure 5D - proportions of top 10 affiliation countries by decade:

#top 10 countries by Decade - horizontal, stacked, with proportion values  
figure5D <- BPindiv %>%   
 filter(!is.na(Country10)) %>%   
 count(Decade, Country10) %>%   
 ggplot(aes(y = Decade,   
 x = n,   
 fill = Country10)) +  
 geom\_bar(aes(fill = Country10),   
 stat = "identity",   
 position = "fill") + # use , position = "fill" for proportion plot  
 scale\_fill\_manual(values = wa\_pal("ferries", 11, reverse = TRUE)) +  
 theme\_bw() +  
 theme(legend.position = "none",   
 axis.text = element\_text(size = 10),   
 legend.box = "horizontal",   
 legend.margin = margin()) +  
 labs(fill = "Awardee affiliation country:") +  
 labs(y = "Decade",   
 x = "Awardee proportion")

Plot awardee country of affiliation counts by decade and by Subject Area:

#top 10 countries by SA and decade  
BPindiv %>%   
 filter(!is.na(Country10)) %>%   
 count(award\_SA, Decade, Country10) %>%   
 ggplot(aes(y = reorder(award\_SA, desc(award\_SA)),   
 x = n,   
 fill = Country10)) +  
 geom\_bar(aes(fill = Country10),   
 stat = "identity") +   
 scale\_fill\_manual(values = wa\_pal("ferries", 11, reverse = TRUE)) +  
 facet\_wrap(~Decade, scales = "fixed", nrow = 1, ncol = 3) +  
 theme\_bw() +   
 theme(legend.position="top",   
 axis.text = element\_text(size = 10),   
 legend.box = "horizontal",   
 legend.margin = margin()) +  
 labs(fill = "Awardee affiliation country:") +  
 labs(x = "Awardee count",   
 y = "Scimago Subject Area")



#ggsave(here("plots", "awardee\_country\_SA\_Decade\_v1.png"), width = 18, height = 8, units = "cm", dpi = "retina")  
#ggsave(here("plots", "awardee\_country\_SA\_Decade\_v1.pdf"), width = 18, height = 8, units = "cm")

#### Figure S31

*Counts of awardees countries of first affiliation - by decade and Subject Area*.

Figure 5 - combine 4 panels and save:

#assemble the panels using patchwork package  
figure5 <- figure5A / figure5B / figure5C / figure5D +   
 plot\_layout(ncol = 1, nrow = 4) +  
 plot\_annotation(tag\_levels = "A")  
#ggsave(plot = figure5, here("plots", "Fig5ABCD\_gender\_countries\_v1.png"), width = 18, height = 14, units = "cm", dpi = "retina", scale = 1.2)  
#ggsave(plot = figure5, here("plots", "Fig5ABCD\_gender\_countries\_v1.pdf"), width = 18, height = 14, units = "cm", scale = 1.2)

### Country productivity context

Load and process SCImago Country document productivity from 2021:

COprod <- read\_csv(here("data", "scimagojr country rank 2021.csv"), skip = 0) #load data

## Rows: 235 Columns: 9  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): Country, Region  
## dbl (7): Rank, Documents, Citable documents, Citations, Self-citations, Cita...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#top10 contries  
COprod$Country %in% unique(BPindiv$Country10) %>% sum() #check overlap - 8

## [1] 8

#setdiff(unique(BPindiv$Country10), COprod$Country) #missing: "United States" and "United Kingdom"  
COprod$Country <- gsub("United States", "USA", COprod$Country) #replace with matching name  
COprod$Country <- gsub("United Kingdom", "UK", COprod$Country) #replace with a matching name  
#all countries  
COprod$Country %in% unique(BPindiv$affiliation\_country) %>% sum() #check overlap - 40

## [1] 39

setdiff(unique(BPindiv$affiliation\_country), COprod$Country) #missing: "Vietnam" "UAE" "Russia"

## [1] "Vietnam" NA "UAE" "Russia"

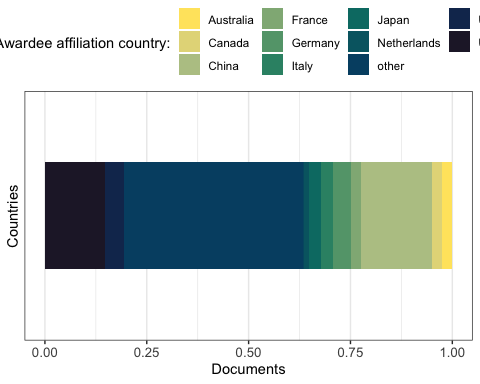
COprod$Country <- gsub("Viet Nam", "Vietnam", COprod$Country) #replace with matching name  
COprod$Country <- gsub("United Arab Emirates", "UAE", COprod$Country) #replace with matching name  
COprod$Country <- gsub("Russian Federation", "Russia", COprod$Country) #replace with a matching name  
  
## Plot for all countries  
#recode top 10 countries + other  
COprod$Country\_affil <- recode(COprod$Country,   
 "USA" = "USA",   
 "UK" = "UK",   
 "Australia" = "Australia",   
 "China" = "China",   
 "Germany" = "Germany",   
 "Canada" = "Canada",   
 "Japan" = "Japan",   
 "Netherlands" = "Netherlands",   
 "Italy" = "Italy",   
 "France" = "France",   
 .default = "other")  
#table(COprod\_affil$Country\_affil)  
names(COprod)

## [1] "Rank" "Country" "Region"   
## [4] "Documents" "Citable documents" "Citations"   
## [7] "Self-citations" "Citations per document" "H index"   
## [10] "Country\_affil"

# ## make a simple table with percentages of the total number of Documents  
# COprod %>%   
# group\_by(Country\_affil) %>%   
# summarise(Documents = sum(Documents)) %>%   
# mutate(Documents\_pct = Documents/sum(Documents) \* 100)

Plot for the top 10 most productive countries:

## Plot productivity of top 10 countries and other - color productivity (Documents) for Countries that are in BPindiv data  
COprod %>%   
 ggplot(aes(x = 1,   
 y = Documents,   
 fill = Country\_affil)) +   
 geom\_bar(stat = "identity",   
 position = "fill") +  
 coord\_flip() +   
 scale\_fill\_manual(values = wa\_pal("ferries", 11, reverse = TRUE)) +  
 theme\_bw() +   
 theme(legend.position="top",   
 axis.text = element\_text(size = 10),   
 legend.box = "horizontal",   
 legend.margin = margin()) +  
 labs(x = "Countries",   
 y = "Documents",   
 fill = "Awardee affiliation country:") +   
 scale\_x\_discrete(labels = NULL) #used to remove vertical labels, also breaks = NULL

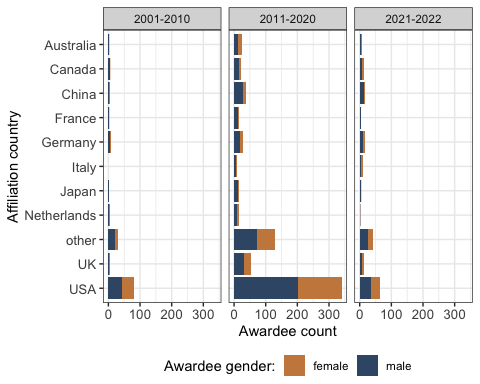


#### Figure S32

*SCImago Country document productivity distribution for top 10 countries in the data set (remaining countries are grouped as “other”)*.

Plot of awardee gender counts by country affiliation by decade:

#top 10 countries by SA and decade  
BPindiv %>%   
 filter(!is.na(Country10)) %>%   
 count(Gender, Decade, Country10) %>%   
 ggplot(aes(y = reorder(Country10, desc(Country10)),   
 x = n,   
 fill = Gender)) +  
 #geom\_bar(stat = "identity")  
 geom\_bar(aes(fill = Gender),   
 stat = "identity") + # use , position = "fill" for proportion plot  
 scale\_fill\_manual(values = c("#CA884C", "#3A5775")) +  
 facet\_wrap(~Decade, scales = "fixed", nrow = 1, ncol = 3) +  
 theme\_bw() +   
 theme(legend.position="bottom",   
 axis.text = element\_text(size = 10),   
 legend.box = "horizontal",   
 legend.margin = margin()) +  
 labs(fill = "Awardee gender:") +  
 labs(x = "Awardee count",   
 y = "Affiliation country")



#### Figure S33

*Counts of awardees by gender - by their country of affiliation and decade. F = female / women, M = male / men*.

### Analyses of past awardees first names

What are the most common first names of the past award winners?

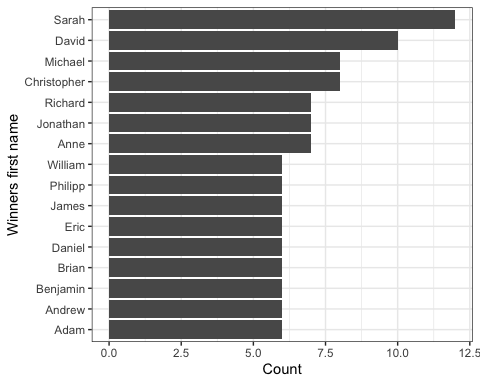
BPindiv$first\_names <- word(BPindiv$awardee\_name, 1) #extract first word  
#length(unique(BPindiv$first\_names)) #815 unique first names, note some are initials  
  
# #count number of unique first names per decade:  
# BPindiv %>%   
# filter(!is.na(first\_names)) %>%   
# group\_by(Decade) %>%   
# summarise(count = n\_distinct(first\_names))  
  
# #count number of unique first names per year:  
# BPindiv %>%   
# filter(!is.na(first\_names)) %>%   
# group\_by(award\_year) %>%   
# summarise(count = n\_distinct(first\_names)) %>% View()  
  
  
#count first\_names - view 16 names with >5 counts (16 names):  
BPindiv %>%   
 #filter(!is.na(first\_names)) %>%   
 count(first\_names) %>%   
 arrange(desc(n)) %>%  
 filter(n > 5)

## # A tibble: 16 × 2  
## first\_names n  
## <chr> <int>  
## 1 Sarah 12  
## 2 David 10  
## 3 Christopher 8  
## 4 Michael 8  
## 5 Anne 7  
## 6 Jonathan 7  
## 7 Richard 7  
## 8 Adam 6  
## 9 Andrew 6  
## 10 Benjamin 6  
## 11 Brian 6  
## 12 Daniel 6  
## 13 Eric 6  
## 14 James 6  
## 15 Philipp 6  
## 16 William 6

The total number of unique past awardee first names in the dataset: 12.

Plot counts of first awardees names that appear more than 5 times:

#plot first names that appear more than 5 times  
BPindiv %>%   
 #filter(!is.na(first\_names)) %>%   
 count(first\_names) %>%   
 arrange(desc(n)) %>%  
 filter(n > 5) %>%  
 ggplot(aes(x = reorder(first\_names, n),   
 y = n)) +  
 geom\_bar(stat = "identity",   
 position = position\_dodge(0.9)) +  
 coord\_flip() +  
 theme\_bw() +  
 labs(x = "Winners first name", y = "Count")



#### Figure S34

*Counts of awardee first names. Only names that appear more than 5 times are shown*.

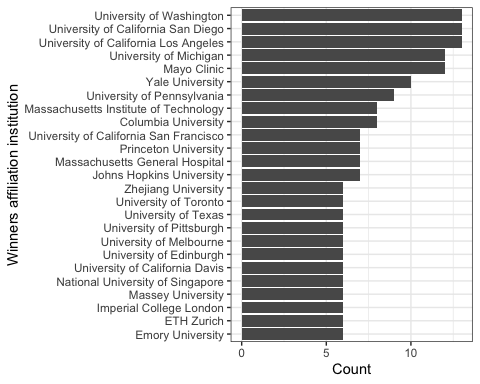
### Affiliation institutions of past awardees

#length(unique(BPindiv$affiliation\_institution)) #672 unique first names, note some require cleaning from extra information after ,  
#do some cleaning  
BPindiv$institution <- BPindiv$affiliation\_institution  
BPindiv$institution <- gsub(",.\*", "", BPindiv$affiliation\_institution) #remove everything (usually a city) after first comma  
  
# #count number of unique institution names per decade:  
# BPindiv %>%  
# filter(!is.na(institution)) %>%  
# group\_by(Decade) %>%  
# summarise(count = n\_distinct(institution))  
  
# #count number of unique institution names per year:  
# BPindiv %>%   
# filter(!is.na(institution)) %>%   
# group\_by(award\_year) %>%   
# summarise(count = n\_distinct(institution)) %>% View()  
  
# #count institutions - view institution names with >5 counts (16 names):  
# BPindiv %>%   
# filter(!is.na(institution)) %>%   
# count(institution) %>%   
# arrange(desc(n)) %>%  
# filter(n > 5)

The total number of unique affiliation institution names in the dataset: 574.

Plot counts of institution names that appear more than 5 times:

#plot institution names that appear more than 5 times  
BPindiv %>%   
 filter(!is.na(institution)) %>%   
 count(institution) %>%   
 arrange(desc(n)) %>%  
 filter(n > 5) %>%  
 ggplot(aes(x = reorder(institution, n),   
 y = n)) +  
 geom\_bar(stat = "identity",   
 position = position\_dodge(0.9)) +  
 coord\_flip() +  
 theme\_bw() +  
 labs(x = "Winners affiliation institution",   
 y = "Count")



#### Figure S35

*Counts of awardee first affiliation institutions. Only institutions that appear more than 5 times are shown*.

## Cash and perks for the winners

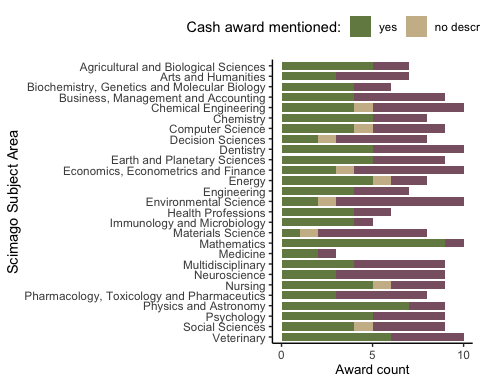
Cash award mentioned in 112 (50.5%) award descriptions.

Figure 6A - mentions of cash awards:

#table(BPdata$Award\_cash) #103 no, 111 yes  
  
figure6A <- BPdata %>%   
 mutate(Award\_cash = factor(Award\_cash, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 mutate(Award\_cash = recode(Award\_cash,   
 'no' = 'no',   
 'yes' = 'yes',   
 'not available' = 'no description')) %>% #change level names  
 count(Award\_cash) %>%  
 ggplot(aes(x = 1,   
 y = n,   
 fill = Award\_cash,   
 pattern = Award\_cash,   
 pattern\_angle = Award\_cash)) +  
 geom\_bar\_pattern(stat = "identity",   
 pattern\_density = 0.1,   
 pattern\_size = 0.5,  
 pattern\_spacing = 0.1,  
 pattern\_key\_scale\_factor = 0.2,  
 pattern\_alpha = 0.5,  
 pattern\_fill = 'white',  
 pattern\_colour = 'white') +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 50, 100, 150, 200, 250)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Cash",   
 y = "Award count") +   
 scale\_x\_discrete(labels = NULL) + #used to remove vertical labels, also breaks = NULL  
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

Plot of cash awards mentioned per Subject Area:

#table(BPdata$Scimago\_Subject\_Area, BPdata$Award\_cash) #as a table by SA  
  
BPdata %>%   
 mutate(Award\_cash = factor(Award\_cash, levels = (c("yes",   
 "not available",   
 "no")))) %>% #reorder value levels  
 mutate(Award\_cash = recode(Award\_cash,   
 'no' = 'no',   
 'yes' = 'yes',   
 'not available' = 'no description')) %>% #change level names  
 count(Scimago\_Subject\_Area, Award\_cash) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area, desc(Scimago\_Subject\_Area)),   
 y = n,   
 fill = Award\_cash)) +   
 geom\_col(width = 0.8,   
 position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10)) +  
 theme\_classic() +   
 scale\_fill\_manual(values = c("#748A52", "#CCBA98", "#8A6172")) +  
 labs(x = "Scimago Subject Area",   
 y = "Award count",   
 fill = "Cash award mentioned:") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))



#### Figure S36

*Counts of mentions of cash awards in award descriptions - by Subject Area*.

Figure 6B - disclosed values of cash awards:

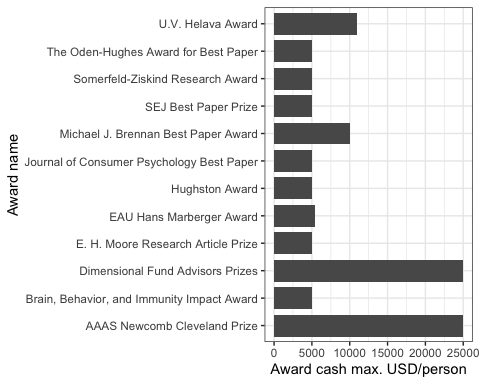
#table(!is.na(BPdata$Award\_cash\_max\_USD\_pperson)) #96 values available  
#median(BPdata$Award\_cash\_max\_USD\_pperson[!is.na(BPdata$Award\_cash\_max\_USD\_pperson)]) #1086 median  
#table(!is.na(BPdata$Award\_cash\_max\_USD\_pperson), BPdata$Award\_individual) #of 96 values available, 39 are for individual awards  
  
#wacolors$uw #pallette to be used  
   
#plot overall - histogram  
figure6B <- BPdata %>%   
 filter(!is.na(Award\_cash\_max\_USD\_pperson)) %>%  
 ggplot(aes(y = Award\_cash\_max\_USD\_pperson)) +   
 geom\_histogram(binwidth = 500,   
 fill = "#748A52",   
 col = "white") +  
 coord\_flip() +  
 theme\_bw() +   
 labs(y = "Award cash max. USD/person",   
 x = "Award count") +   
 scale\_y\_continuous(breaks = c(seq(0, 30000, by = 5000))) +  
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10))

Figure 6 - Combine 2 panels and save:

#assemble the panels using patchwork package  
figure6 <- figure6A / figure6B +  
 plot\_layout(nrow = 2, heights = c(1, 3)) +  
 plot\_annotation(tag\_levels = "A")  
#ggsave(plot = figure6, here("plots", "Fig6AB\_perks\_v1.png"), width = 18, height = 8, units = "cm", dpi = "retina", scale = 1.2)  
#ggsave(plot = figure6, here("plots", "Fig6AB\_perks\_v1.pdf"), width = 18, height = 8, units = "cm", scale = 1.2)

Plot of awards with highest cash amounts:

BPdata %>%   
 filter(!is.na(Award\_cash\_max\_USD\_pperson)) %>%  
 filter(Award\_cash\_max\_USD\_pperson >= 5000) %>%  
 ggplot(aes(x = Award\_name,   
 y = Award\_cash\_max\_USD\_pperson)) +   
 geom\_col(width = 0.8,   
 position = position\_stack(reverse = TRUE)) +  
 coord\_flip() +  
 theme\_bw() +   
 labs(y = "Award cash max. USD/person",   
 x = "Award name")

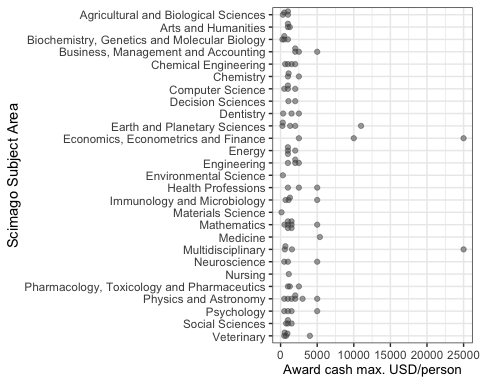


#### Figure S37

*Awards with monetary values of at least 5,000 USD*.

Plot disclosed cash amounts by Subject Area:

BPdata %>%   
 filter(!is.na(Award\_cash\_max\_USD\_pperson)) %>%  
 ggplot(aes(x = reorder(Scimago\_Subject\_Area,   
 desc(Scimago\_Subject\_Area)),   
 y = Award\_cash\_max\_USD\_pperson)) +   
 geom\_beeswarm(color = "grey30",   
 alpha = 0.5) +  
 coord\_flip() +  
 theme\_bw() +   
 labs(y = "Award cash max. USD/person",   
 x = "Scimago Subject Area") +   
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10))



#### Figure S37

*Awards with monetary values of at least 5,000 USD - by Subject Area*.

## Text mining of the description of the perks and benefits

Comment\_on\_the\_award\_cash - process description text:

#Using packages tidytext and stopwords  
Comment\_on\_the\_award\_cash\_txt <- tibble(txt = tolower(BPdata$Comment\_on\_the\_award\_cash))  
Comment\_on\_the\_award\_cash\_txt <- Comment\_on\_the\_award\_cash\_txt %>% unnest\_tokens(output = word, input = txt, token = "words", to\_lower = TRUE) #restructure all descriptions as one-token-per-row format  
Comment\_on\_the\_award\_cash\_txt <- Comment\_on\_the\_award\_cash\_txt %>% anti\_join(get\_stopwords()) #remove stop words from library(stopwords)

## Joining with `by = join\_by(word)`

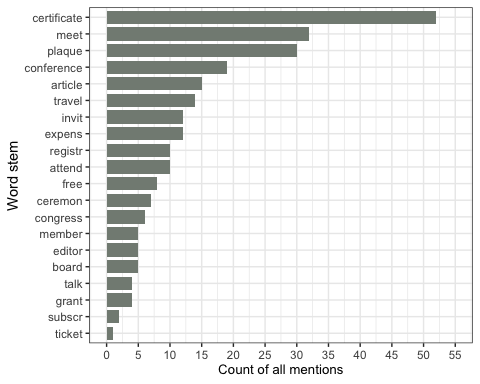
Comment\_on\_the\_award\_cash\_txt$word <- tokenize\_word\_stems(Comment\_on\_the\_award\_cash\_txt$word) #make all word stems lowercase  
word.freq <- Comment\_on\_the\_award\_cash\_txt %>% count(word, sort = TRUE) #count words  
word.freq$word <- gsub("[[:digit:]]", "", word.freq$word) #remove numbers  
word.freq$word <- gsub("[[:punct:][:blank:]]+", " ", word.freq$word) #remove punctuation  
word.freq <- word.freq %>% drop\_na() %>% filter(word != "")

Count mentions of specific words (stemmed) in Comment\_on\_the\_award\_cash:

#create list of specific words (stemmed) to count within strings  
specific.words2 <- c("certificate", "grant", "attend", "expens", "plaque", "talk", "invit", "registr", "meet", "conference", "member", "travel", "subscr", "board", "editor", "ticket", "article", "free", "ceremon", "congress")  
  
#prepare award descriptions as a single lowercase string  
descriptions2 <- BPdata %>%   
 filter(!is.na(Comment\_on\_the\_award\_cash)) %>%   
 select(Comment\_on\_the\_award\_cash) %>%   
 tolower() #single lowercase string  
  
#sum of all mentions for each word  
specific.words.mentions2 <- specific.words2 %>%  
 map\_int(~ str\_count(tolower(descriptions2), .x))  
  
#prepare award descriptions while keeping them separate for each award  
descriptions3 <- tolower(BPdata$Comment\_on\_the\_award\_cash) #vector of lowercase strings  
  
#sum of mentions per award for each word (counts only one mention per award)  
specific.words.mentions3 <- specific.words2 %>%  
 map\_int(~ sum(str\_detect(descriptions3, .x), na.rm = TRUE))  
  
# ## doing the same as above, but manually:  
# #count all mentions of words(parts) individually, e.g.:  
# sum(str\_count(tolower(BPdata$Comment\_on\_the\_award\_cash), "certificate"), na.rm = TRUE)  
# #counting once per award, e.g.   
# sum(str\_detect(BPdata$Comment\_on\_the\_award\_cash, "certificate"), na.rm = TRUE)

Plot frequencies of specific words - all mentions in cash descriptions:

words.df <- tibble(Words = specific.words2,   
 Count\_all = specific.words.mentions2,  
 Count\_once = specific.words.mentions3)  
  
words.df %>%  
 ggplot(aes(x = reorder(Words, Count\_all),   
 y = Count\_all)) +   
 geom\_col(width = 0.8,   
 fill = "#838B83") +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55),   
 limits = c(0, 55)) +  
 theme\_bw() +   
 labs(x = "Word stem",   
 y = "Count of all mentions") +   
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10))

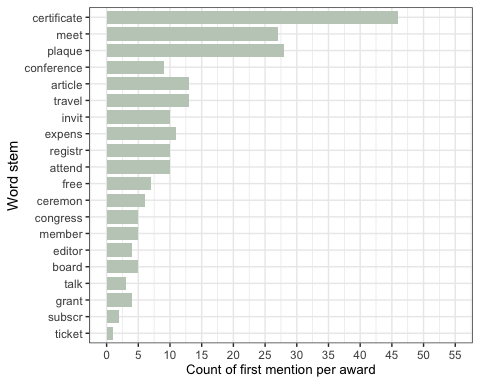


#### Figure S38

*Total counts of the top 50 most mentioned words (stemmed) in the parts of award description related to the prize value and related perks*.

Plot frequencies of specific words - first mentions in perks and benefits descriptions:

words.df %>%  
 ggplot(aes(x = reorder(Words, Count\_all),   
 y = Count\_once)) +   
 geom\_col(width = 0.8,   
 fill = "#C1CDC1") +  
 coord\_flip() +  
 scale\_y\_continuous(breaks = c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55),   
 limits = c(0, 55)) +  
 theme\_bw() +   
 labs(x = "Word stem",   
 y = "Count of first mention per award") +   
 #scale\_x\_discrete(labels = NULL) + labs(x = "") + #used to remove vertical labels, also breaks = NULL  
 theme(legend.position = "none",   
 axis.title.x = element\_text(size = 10))



#### Figure S39

*Total counts of the top 50 most mentioned words (stemmed) in the parts of award description related to the prize value and related perks - only one mention per award counted*.

Figure 7A - Individual winners data - presence of a personal profile:

#note: based on individual winners data  
#table(BPindiv$awardee\_profile\_shown, useNA = "always")  
  
#plot overall  
figure7A <- BPindiv %>%   
 mutate(awardee\_profile\_shown = factor(awardee\_profile\_shown, levels = (c("yes",   
 "no")))) %>% #reorder value levels  
 group\_by(award\_year) %>%  
 count(awardee\_profile\_shown) %>%  
 ggplot(aes(x = award\_year,   
 y = n,   
 fill = awardee\_profile\_shown,   
 pattern = awardee\_profile\_shown,   
 pattern\_angle = awardee\_profile\_shown)) +  
 geom\_bar\_pattern(stat = "identity",   
 pattern\_density = 0.1,   
 pattern\_size = 0.4,  
 pattern\_spacing = 0.015,  
 pattern\_key\_scale\_factor = 0.5,  
 pattern\_alpha = 0.5,  
 pattern\_fill = 'white',  
 pattern\_colour = 'white') +  
 scale\_fill\_manual(values = c("#748A52", "#8A6172")) +  
 theme\_bw() +   
 labs(x = "Year",   
 y = "Winner count") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

Figure 7B - Individual winners data - presence of a personal photo:

#table(BPindiv$awardee\_photo\_shown, useNA = "always")  
  
#plot overall  
figure7B <- BPindiv %>%   
 mutate(awardee\_photo\_shown = factor(awardee\_photo\_shown, levels = (c("yes",   
 "no")))) %>% #reorder value levels  
 group\_by(award\_year) %>%  
 count(awardee\_photo\_shown) %>%  
 ggplot(aes(x = award\_year,   
 y = n,   
 fill = awardee\_photo\_shown,   
 pattern = awardee\_photo\_shown,   
 pattern\_angle = awardee\_photo\_shown)) +  
 geom\_bar\_pattern(stat = "identity",   
 pattern\_density = 0.1,   
 pattern\_size = 0.4,  
 pattern\_spacing = 0.015,  
 pattern\_key\_scale\_factor = 0.5,  
 pattern\_alpha = 0.5,  
 pattern\_fill = 'white',  
 pattern\_colour = 'white') +  
 scale\_fill\_manual(values = c("#748A52", "#8A6172")) +  
 theme\_bw() +   
 labs(x = "Year",  
 y = "Winner count") +   
 theme(legend.position = "top",   
 axis.title.x = element\_text(size = 10))

Figure 7 - Combine 2 panels and save:

#assemble the panels using patchwork package  
figure7 <- figure7A / figure7B +  
 plot\_layout(ncol = 2, nrow = 1) +  
 plot\_annotation(tag\_levels = "A")  
ggsave(plot = figure7, here("plots", "Fig7AB\_perks\_v1.png"), width = 18, height = 8, units = "cm", dpi = "retina", scale = 1.2)  
#ggsave(plot = figure7, here("plots", "Fig7AB\_perks\_v1.pdf"), width = 18, height = 8, units = "cm", scale = 1.2)

## Statistical models

### Models for individual winners gender across years

#Scale the award\_year variable to have the mean of 0 and SD of 1  
BPindiv$Award\_year\_scaled <- scale(BPindiv$award\_year)

Statistical analyses of the gender differences

#table(BPindiv$awardee\_gender, useNA = "always")  
BPindiv %>%   
 mutate(Gender = case\_when(  
 endsWith(awardee\_gender, "F") ~ 1,  
 endsWith(awardee\_gender, "M") ~ 0  
 )) -> BPindiv  
#table(BPindiv$Gender)  
  
#Fit generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
  
#without year  
model\_gender <- glmer(Gender ~ 1 + (1|award\_SA) + (1|award\_name),   
 family = "binomial",   
 data = BPindiv) #without year as a predictor  
summary(model\_gender)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Gender ~ 1 + (1 | award\_SA) + (1 | award\_name)  
## Data: BPindiv  
##   
## AIC BIC logLik deviance df.resid   
## 1422.4 1437.3 -708.2 1416.4 1076   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.4256 -0.7983 -0.6339 1.0853 2.0280   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## award\_name (Intercept) 0.2275 0.4770   
## award\_SA (Intercept) 0.1647 0.4058   
## Number of obs: 1079, groups: award\_name, 61; award\_SA, 19  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.5093 0.1454 -3.503 0.000461 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plogis(summary(model\_gender)$coef[1,1]) \*100 #calculate % difference between genders at the intercept: 38% female names]]

## [1] 37.53664

#with year  
model\_gender <- glmer(Gender ~ Award\_year\_scaled + (1|award\_SA) + (1|award\_name),   
 family = "binomial",   
 data = BPindiv) #with year as a predictor  
summary(model\_gender) #year not significant

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Gender ~ Award\_year\_scaled + (1 | award\_SA) + (1 | award\_name)  
## Data: BPindiv  
##   
## AIC BIC logLik deviance df.resid   
## 1423.5 1443.4 -707.7 1415.5 1075   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.4546 -0.8070 -0.6205 1.0708 2.1288   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## award\_name (Intercept) 0.2133 0.4618   
## award\_SA (Intercept) 0.1824 0.4271   
## Number of obs: 1079, groups: award\_name, 61; award\_SA, 19  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.52084 0.14784 -3.523 0.000427 \*\*\*  
## Award\_year\_scaled 0.06841 0.07214 0.948 0.342979   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_yr\_scl -0.082

plogis(summary(model\_gender)$coef[1,1])\*100 #calculate % difference between genders at the intercept: 37% female names

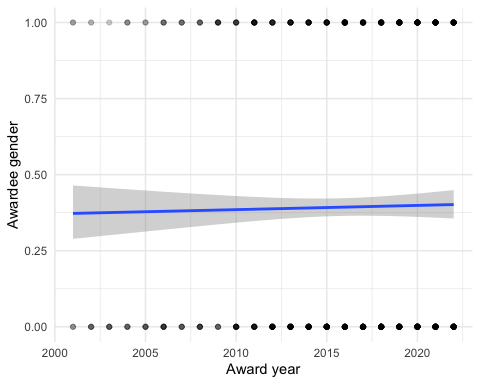
## [1] 37.26562

#use the reports package (Makowski et al. 2021) to summarize the analysis  
# report::report(model\_gender)

Logistic plot for awardee gender by year:

#plot  
BPindiv %>%   
 ggplot(aes(x = award\_year,   
 y = Gender)) +   
 geom\_point(alpha = 0.1) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Award year") +   
 ylab("Awardee gender") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S40

*Logistic plot of awardee gender, by award year*.

### Models for individual winners online profile across years

#table(BPindiv$awardee\_profile\_shown, useNA = "always")  
BPindiv %>%   
 mutate(Profile = case\_when(  
 endsWith(awardee\_profile\_shown, "yes") ~ 1,  
 endsWith(awardee\_profile\_shown, "no") ~ 0  
 )) -> BPindiv  
  
#Fit generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package). NOte: does not work with SA as a random effect:  
model\_profile <- glmer(Profile ~ Award\_year\_scaled + (1|award\_name),   
 family = "binomial",   
 data = BPindiv) #with year as a predictor  
summary(model\_profile) #slope 1.11 signif

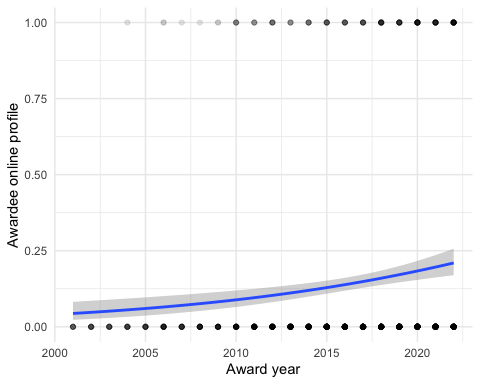
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Profile ~ Award\_year\_scaled + (1 | award\_name)  
## Data: BPindiv  
##   
## AIC BIC logLik deviance df.resid   
## 383.9 398.8 -188.9 377.9 1076   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.7555 -0.0157 -0.0097 -0.0052 10.5452   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## award\_name (Intercept) 94.32 9.712   
## Number of obs: 1079, groups: award\_name, 61  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -9.0425 1.4326 -6.312 2.75e-10 \*\*\*  
## Award\_year\_scaled 1.1087 0.2231 4.970 6.70e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_yr\_scl -0.155

#use the reports package (Makowski et al. 2021) to summarize the analysis  
# report::report(model\_profile)

Logistic plot for awardee online profile by year:

#plot  
BPindiv %>%   
 ggplot(aes(x = award\_year,   
 y = Profile)) +   
 geom\_point(alpha = 0.1) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Award year") +   
 ylab("Awardee online profile") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S41

*Logistic plot of awardee online profile, by award year*.

### Models for individual winners online photo across years

#table(BPindiv$awardee\_photo\_shown, useNA = "always")  
BPindiv %>%   
 mutate(Photo = case\_when(  
 endsWith(awardee\_photo\_shown, "yes") ~ 1,  
 endsWith(awardee\_photo\_shown, "no") ~ 0  
 )) -> BPindiv  
  
#Fit generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package). Note: does not work with SA as a random effect:  
model\_photo <- glmer(Photo ~ Award\_year\_scaled + (1|award\_name),   
 family = "binomial",   
 data = BPindiv) #with year as a predictor  
summary(model\_photo) #slope 1.18 signif

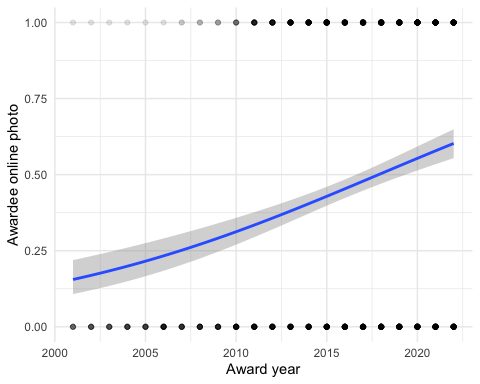
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Photo ~ Award\_year\_scaled + (1 | award\_name)  
## Data: BPindiv  
##   
## AIC BIC logLik deviance df.resid   
## 426.5 441.5 -210.3 420.5 1076   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -13.2419 -0.0148 -0.0037 0.0755 2.0921   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## award\_name (Intercept) 151.6 12.31   
## Number of obs: 1079, groups: award\_name, 61  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -8.9908 1.5487 -5.806 6.41e-09 \*\*\*  
## Award\_year\_scaled 1.2244 0.2215 5.528 3.24e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_yr\_scl -0.170

#use the reports package (Makowski et al. 2021) to summarize the analysis  
# report::report(model\_photo)

Logistic plot for awardee online photo by year:

#plot  
BPindiv %>%   
 ggplot(aes(x = award\_year,   
 y = Photo)) +   
 geom\_point(alpha = 0.1) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Award year") +   
 ylab("Awardee online photo") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S42

*Logistic plot of awardee online photo, by award year*.

## Models for award characteristics vs. how old are the awards

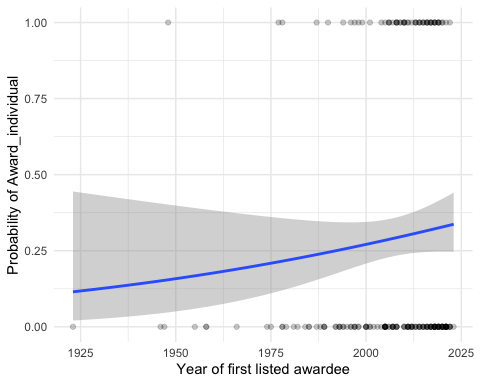
NOTE: Using the year of oldest listed past winner as a proxy of award establishment year:.

#Scale the Awardee\_list\_earliest\_year variable to have the mean of 0 and SD of 1  
BPdata$Awardee\_list\_earliest\_year\_scaled <- scale(BPdata$Awardee\_list\_earliest\_year)  
#table(!is.na(BPdata$Awardee\_list\_earliest\_year)) # 203 values available

### Award\_individual

#table  
# table(BPdata$Award\_individual, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 mutate(Award\_individual\_num = recode(Award\_individual,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Award\_individual\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Award\_individual") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S43

*Logistic plot of whether award is individual-focused, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Award\_individual <- glmer(Award\_individual\_num ~ Awardee\_list\_earliest\_year\_scaled + (1|Scimago\_Subject\_Area) + (1|Awarding\_society),   
 family = "binomial",   
 data = BPdata4) #with year as a predictor  
summary(model\_Award\_individual) #slope ns

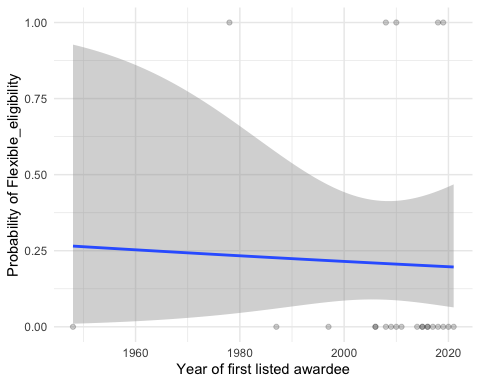
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Award\_individual\_num ~ Awardee\_list\_earliest\_year\_scaled + (1 |   
## Scimago\_Subject\_Area) + (1 | Awarding\_society)  
## Data: BPdata4  
##   
## AIC BIC logLik deviance df.resid   
## 229.9 243.1 -110.9 221.9 199   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.3911 -0.5661 -0.2933 0.6713 2.0516   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Awarding\_society (Intercept) 0.2334 0.4831   
## Scimago\_Subject\_Area (Intercept) 2.3012 1.5170   
## Number of obs: 203, groups: Awarding\_society, 118; Scimago\_Subject\_Area, 27  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.0512 0.3854 -2.727 0.00639 \*\*  
## Awardee\_list\_earliest\_year\_scaled 0.2523 0.2318 1.088 0.27650   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_lst\_\_\_ 0.040

# report::report(model\_Award\_individual) #use the reports package (Makowski et al. 2021) to summarize the analysis  
  
##alternative model reporting (as a table)  
#library(sjPlot)  
#tab\_model(model\_Award\_individual)

## Flexible\_eligibility

#table  
# table(BPdata$Flexible\_eligibility, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Flexible\_eligibility == "yes" | Flexible\_eligibility == "no") %>%   
 mutate(Flexible\_eligibility\_num = recode(Flexible\_eligibility,   
 yes = 1,   
 no = 0)) -> BPdata4  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Flexible\_eligibility\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Flexible\_eligibility") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S44

*Logistic plot of whether award has flexible eligibility, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
#model\_Flexible\_eligibility <- glmer(Flexible\_eligibility\_num ~ Awardee\_list\_earliest\_year\_scaled + (1|Scimago\_Subject\_Area) + (1|Awarding\_society), family = "binomial", data = BPdata4) #with year as a predictor - NOT CONVERGING with random effects included  
model\_Flexible\_eligibility <- glm(Flexible\_eligibility\_num ~ Awardee\_list\_earliest\_year\_scaled,   
 family = "binomial",   
 data = BPdata4) #with year as a predictor, without random effects included  
summary(model\_Flexible\_eligibility) #slope ns

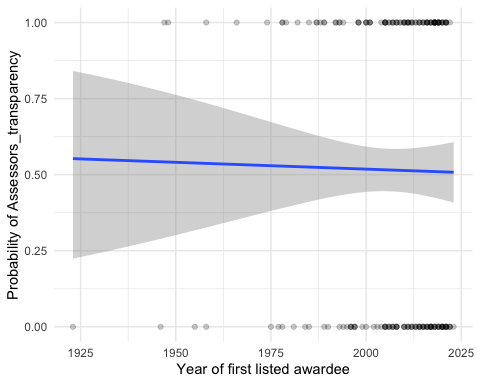
##   
## Call:  
## glm(formula = Flexible\_eligibility\_num ~ Awardee\_list\_earliest\_year\_scaled,   
## family = "binomial", data = BPdata4)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.7852 -0.6813 -0.6704 -0.6629 1.7989   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.32990 0.50324 -2.643 0.00823 \*\*  
## Awardee\_list\_earliest\_year\_scaled -0.08686 0.48362 -0.180 0.85747   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 24.564 on 23 degrees of freedom  
## Residual deviance: 24.532 on 22 degrees of freedom  
## AIC: 28.532  
##   
## Number of Fisher Scoring iterations: 4

# report::report(model\_Flexible\_eligibility) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Assessors\_transparency

#table  
# table(BPdata$Assessors\_transparency, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Assessors\_transparency == "yes" | Assessors\_transparency == "no") %>%   
 mutate(Assessors\_transparency\_num = recode(Assessors\_transparency,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Assessors\_transparency\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Assessors\_transparency") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S45

*Logistic plot of whether award has assessor transparency, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Assessors\_transparency <- glmer(Assessors\_transparency\_num ~ Awardee\_list\_earliest\_year\_scaled + (1|Scimago\_Subject\_Area),   
 family = "binomial",   
 data = BPdata4) #with year as a predictor, + (1|Awarding\_society) causes "boundary (singular)"  
summary(model\_Assessors\_transparency) #slope ns

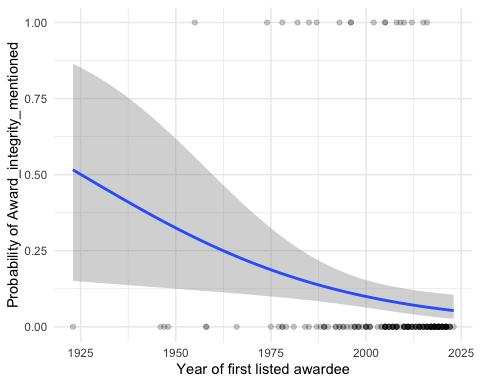
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Assessors\_transparency\_num ~ Awardee\_list\_earliest\_year\_scaled +   
## (1 | Scimago\_Subject\_Area)  
## Data: BPdata4  
##   
## AIC BIC logLik deviance df.resid   
## 272.5 282.3 -133.3 266.5 191   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.4059 -0.9351 0.7128 0.9220 1.2300   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Scimago\_Subject\_Area (Intercept) 0.2852 0.534   
## Number of obs: 194, groups: Scimago\_Subject\_Area, 27  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.06852 0.18168 0.377 0.706  
## Awardee\_list\_earliest\_year\_scaled 0.02803 0.15627 0.179 0.858  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_lst\_\_\_ 0.012

# report::report(model\_Assessors\_transparency) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Award\_integrity\_mentioned

#table  
# table(BPdata$Award\_integrity\_mentioned, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Award\_integrity\_mentioned == "yes" | Award\_integrity\_mentioned == "no") %>%   
 mutate(Award\_integrity\_mentioned\_num = recode(Award\_integrity\_mentioned,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Award\_integrity\_mentioned\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Award\_integrity\_mentioned") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S46

*Logistic plot of whether award mentions assessment integrity, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Award\_integrity\_mentioned <- glmer(Award\_integrity\_mentioned\_num ~ Awardee\_list\_earliest\_year\_scaled + (1|Awarding\_society),  
 family = "binomial",   
 data = BPdata4) #with year as a predictor, + (1|Scimago\_Subject\_Area) causes "boundary (singular)"  
summary(model\_Award\_integrity\_mentioned) #slope ns, but close

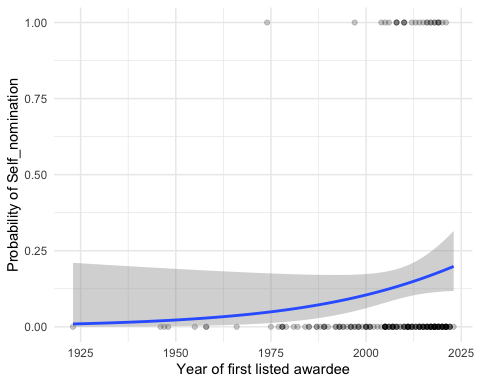
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Award\_integrity\_mentioned\_num ~ Awardee\_list\_earliest\_year\_scaled +   
## (1 | Awarding\_society)  
## Data: BPdata4  
##   
## AIC BIC logLik deviance df.resid   
## 85.6 95.4 -39.8 79.6 190   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.1129 -0.0979 -0.0037 -0.0023 4.6810   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Awarding\_society (Intercept) 559.1 23.64   
## Number of obs: 193, groups: Awarding\_society, 115  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -11.6213 2.6581 -4.372 1.23e-05 \*\*\*  
## Awardee\_list\_earliest\_year\_scaled -1.1068 0.7514 -1.473 0.141   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_lst\_\_\_ 0.657

# report::report(model\_Award\_integrity\_mentioned) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Self\_nomination

#table  
# table(BPdata$Self\_nomination, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Self\_nomination == "yes" | Self\_nomination == "no") %>%   
 mutate(Self\_nomination\_num = recode(Self\_nomination,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Self\_nomination\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Self\_nomination") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S47

*Logistic plot of whether award explicitly allows self-nominations, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Self\_nomination <- glmer(Self\_nomination\_num ~ Awardee\_list\_earliest\_year\_scaled + (1|Scimago\_Subject\_Area),   
 family = "binomial",   
 data = BPdata4) #with year as a predictor  
summary(model\_Self\_nomination) #slope ns, but trend for newer awards to mention self-nominations

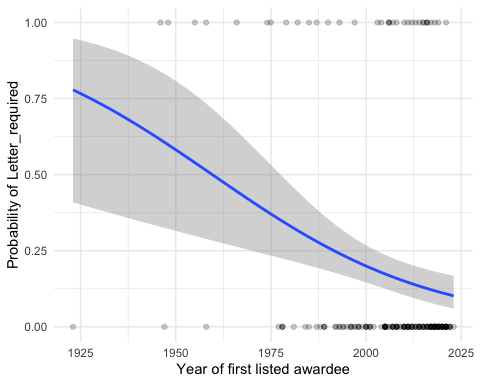
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Self\_nomination\_num ~ Awardee\_list\_earliest\_year\_scaled + (1 |   
## Scimago\_Subject\_Area)  
## Data: BPdata4  
##   
## AIC BIC logLik deviance df.resid   
## 154.0 163.8 -74.0 148.0 190   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -0.6315 -0.4125 -0.3501 -0.2592 4.8781   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Scimago\_Subject\_Area (Intercept) 0.3627 0.6023   
## Number of obs: 193, groups: Scimago\_Subject\_Area, 27  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.0653 0.3095 -6.673 2.51e-11 \*\*\*  
## Awardee\_list\_earliest\_year\_scaled 0.5658 0.3225 1.754 0.0793 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_lst\_\_\_ -0.343

# report::report(model\_Self\_nomination) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Letter\_required

#table  
# table(BPdata$Letter\_required, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Letter\_required == "yes" | Letter\_required == "no") %>%   
 mutate(Letter\_required\_num = recode(Letter\_required,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Letter\_required\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Letter\_required") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S48

*Logistic plot of whether award nomination requires a letter, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Letter\_required <- glmer(Letter\_required\_num ~ Awardee\_list\_earliest\_year\_scaled + (1|Scimago\_Subject\_Area) + (1|Awarding\_society),   
 family = "binomial",   
 data = BPdata4) #with year as a predictor  
summary(model\_Letter\_required) #slope significant negative -0.5431 - less common over the years

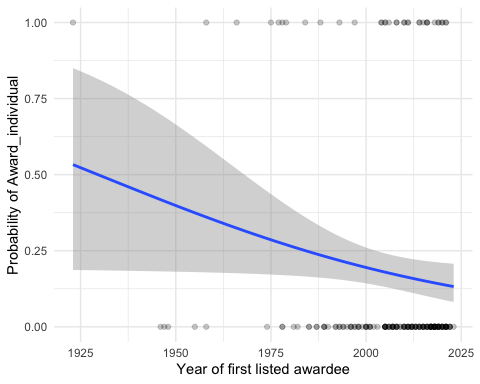
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Letter\_required\_num ~ Awardee\_list\_earliest\_year\_scaled + (1 |   
## Scimago\_Subject\_Area) + (1 | Awarding\_society)  
## Data: BPdata4  
##   
## AIC BIC logLik deviance df.resid   
## 176.8 189.9 -84.4 168.8 189   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.6478 -0.4311 -0.3374 -0.2687 2.6569   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Awarding\_society (Intercept) 0.1654 0.4067   
## Scimago\_Subject\_Area (Intercept) 0.5556 0.7454   
## Number of obs: 193, groups: Awarding\_society, 115; Scimago\_Subject\_Area, 27  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.7578 0.3110 -5.653 1.58e-08 \*\*\*  
## Awardee\_list\_earliest\_year\_scaled -0.5425 0.1929 -2.812 0.00492 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_lst\_\_\_ 0.208

# report::report(model\_Letter\_required) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Award\_contact\_provided

#table  
# table(BPdata$Award\_contact\_provided, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Award\_contact\_provided == "yes" | Award\_contact\_provided == "no") %>%   
 mutate(Award\_contact\_provided\_num = recode(Award\_contact\_provided,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Award\_contact\_provided\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Award\_individual") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S49

*Logistic plot of whether award contact details are provided, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Award\_contact\_provided <- glm(Award\_contact\_provided\_num ~ Awardee\_list\_earliest\_year\_scaled,   
 family = "binomial",   
 data = BPdata4) #with year as a predictor - but fails with random effects  
summary(model\_Award\_contact\_provided) #slope -0.3314

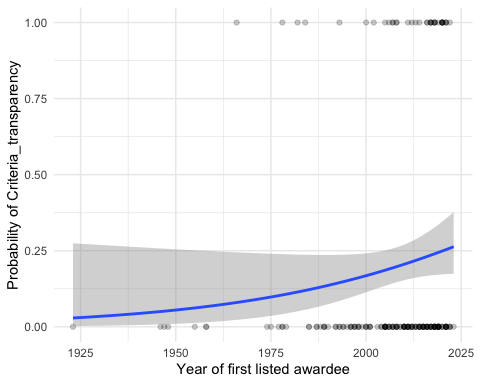
##   
## Call:  
## glm(formula = Award\_contact\_provided\_num ~ Awardee\_list\_earliest\_year\_scaled,   
## family = "binomial", data = BPdata4)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0404 -0.6292 -0.5738 -0.5426 1.9943   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.5473 0.1923 -8.048 8.41e-16 \*\*\*  
## Awardee\_list\_earliest\_year\_scaled -0.3283 0.1625 -2.021 0.0433 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 182.74 on 192 degrees of freedom  
## Residual deviance: 178.87 on 191 degrees of freedom  
## AIC: 182.87  
##   
## Number of Fisher Scoring iterations: 4

# report::report(model\_Award\_contact\_provided) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Criteria\_transparency

#table  
# table(BPdata$Criteria\_transparency, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Criteria\_transparency == "yes" | Criteria\_transparency == "no") %>%   
 mutate(Criteria\_transparency\_num = recode(Criteria\_transparency,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Criteria\_transparency\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Criteria\_transparency") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S50

*Logistic plot of whether award has transparent assessment criteria, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Criteria\_transparency <- glm(Criteria\_transparency\_num ~ Awardee\_list\_earliest\_year\_scaled,   
 family = "binomial",   
 data = BPdata4) #with year as a predictor  
summary(model\_Criteria\_transparency) #slope ns

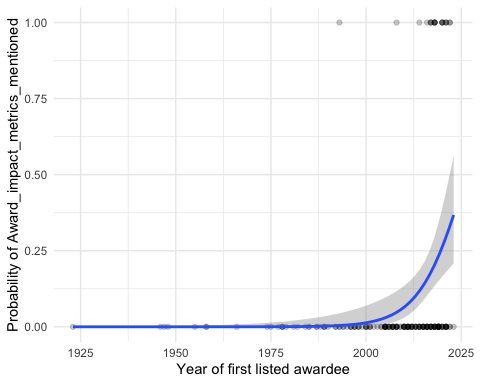
##   
## Call:  
## glm(formula = Criteria\_transparency\_num ~ Awardee\_list\_earliest\_year\_scaled,   
## family = "binomial", data = BPdata4)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.7817 -0.7168 -0.6488 -0.4688 2.2484   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.4415 0.1891 -7.623 2.48e-14 \*\*\*  
## Awardee\_list\_earliest\_year\_scaled 0.4048 0.2412 1.678 0.0934 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 191.04 on 191 degrees of freedom  
## Residual deviance: 187.57 on 190 degrees of freedom  
## AIC: 191.57  
##   
## Number of Fisher Scoring iterations: 4

# report::report(model\_Criteria\_transparency) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Award\_impact\_metrics\_mentioned

#table  
# table(BPdata$Award\_impact\_metrics\_mentioned, useNA = "always")  
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Award\_impact\_metrics\_mentioned == "yes" | Award\_impact\_metrics\_mentioned == "no") %>%   
 mutate(Award\_impact\_metrics\_mentioned\_num = recode(Award\_impact\_metrics\_mentioned,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Award\_impact\_metrics\_mentioned\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Award\_impact\_metrics\_mentioned") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S51

*Logistic plot of whether award description mentions impact metrics, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Award\_impact\_metrics\_mentioned <- glmer(Award\_impact\_metrics\_mentioned\_num ~ Awardee\_list\_earliest\_year\_scaled + (1|Awarding\_society),   
 family = "binomial",   
 data = BPdata4) #with year as a predictor  
summary(model\_Award\_impact\_metrics\_mentioned) #slope signif 4.930 - recently mentioned more often

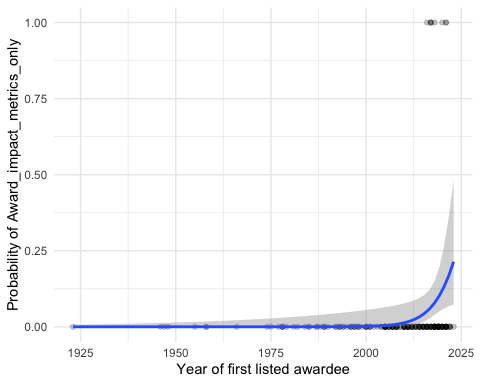
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula:   
## Award\_impact\_metrics\_mentioned\_num ~ Awardee\_list\_earliest\_year\_scaled +   
## (1 | Awarding\_society)  
## Data: BPdata4  
##   
## AIC BIC logLik deviance df.resid   
## 90.1 99.9 -42.1 84.1 190   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -0.99218 -0.01514 -0.00158 -0.00009 2.46961   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Awarding\_society (Intercept) 916.8 30.28   
## Number of obs: 193, groups: Awarding\_society, 115  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -14.264 3.262 -4.373 1.23e-05 \*\*\*  
## Awardee\_list\_earliest\_year\_scaled 4.870 2.256 2.158 0.0309 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## Awrd\_lst\_\_\_ -0.570

# report::report(model\_Award\_impact\_metrics\_mentioned) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Award\_impact\_metrics\_only - seem to be mostly from the last 5 years

#table  
# table(BPdata$Award\_impact\_metrics\_only, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Award\_impact\_metrics\_only == "yes" | Award\_impact\_metrics\_only == "no") %>%   
 mutate(Award\_impact\_metrics\_only\_num = recode(Award\_impact\_metrics\_only,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%   
 ggplot(aes(x = Awardee\_list\_earliest\_year,   
 y = Award\_impact\_metrics\_only\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Award\_impact\_metrics\_only") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S52

*Logistic plot of whether award description mentions impact metrics as only criteria, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Award\_impact\_metrics\_only <- glm(Award\_impact\_metrics\_only\_num ~ Awardee\_list\_earliest\_year\_scaled,   
 family = "binomial",   
 data = BPdata4) #with year as a predictor, fails with random effects  
summary(model\_Award\_impact\_metrics\_only) #slope signif 3.782 - more common recently

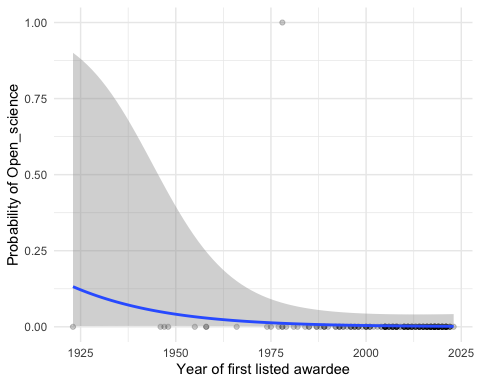
##   
## Call:  
## glm(formula = Award\_impact\_metrics\_only\_num ~ Awardee\_list\_earliest\_year\_scaled,   
## family = "binomial", data = BPdata4)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.69414 -0.36191 -0.16275 -0.03612 2.44021   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.154 1.289 -3.998 6.4e-05 \*\*\*  
## Awardee\_list\_earliest\_year\_scaled 3.784 1.705 2.219 0.0265 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 66.596 on 192 degrees of freedom  
## Residual deviance: 54.872 on 191 degrees of freedom  
## AIC: 58.872  
##   
## Number of Fisher Scoring iterations: 8

# report::report(model\_Award\_impact\_metrics\_only) #use the reports package (Makowski et al. 2021) to summarize the analysis

## Open\_science

#table  
# table(BPdata$Open\_science, useNA = "always")  
  
#filter and make binomial   
BPdata %>%   
 filter(!is.na(Awardee\_list\_earliest\_year)) %>%   
 filter(Open\_science == "yes" | Open\_science == "no") %>%   
 mutate(Open\_science\_num = recode(Open\_science,   
 yes = 1,   
 no = 0)) -> BPdata4  
  
#plot  
BPdata4 %>%  
 ggplot(aes(x = Awardee\_list\_earliest\_year, y = Open\_science\_num)) +   
 geom\_point(alpha = 0.2) +   
 stat\_smooth(method = "glm",   
 method.args = list(family = binomial),   
 se = TRUE) +  
 xlab("Year of first listed awardee") +   
 ylab("Probability of Open\_science") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



#### Figure S53

*Logistic plot of whether award description mentions Open Science practices, by award age*.

Model:

#model - fit a generalised mixed model with binomial error family and with logit link function, award as a random effect (using glmer from lme4 package):  
model\_Open\_science <- glm(Open\_science\_num ~ Awardee\_list\_earliest\_year\_scaled,   
 family = "binomial",   
 data = BPdata4) #with year as a predictor, fails with random effects  
summary(model\_Open\_science) #slope ns

##   
## Call:  
## glm(formula = Open\_science\_num ~ Awardee\_list\_earliest\_year\_scaled,   
## family = "binomial", data = BPdata4)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.53186 -0.08916 -0.07063 -0.05999 2.98784   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.7776 1.3438 -4.299 1.71e-05 \*\*\*  
## Awardee\_list\_earliest\_year\_scaled -0.7611 0.5281 -1.441 0.15   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 12.520 on 192 degrees of freedom  
## Residual deviance: 10.936 on 191 degrees of freedom  
## AIC: 14.936  
##   
## Number of Fisher Scoring iterations: 8

# report::report(model\_Open\_science) #use the reports package (Makowski et al. 2021) to summarize the analysis