

Instructions – Creating Input files (cleanBib, gender-api.com)

1. **gender-api.com** : check it out (to get an API key, needed for cleanBib)
2. **your citation management tool** (Zotero/Mendeley/EndNote etc):
 - + Create a folder including all your references &
 - + Edit those references to a certain format
 - + Export .bib file
3. **cleanBib** :
 - + open
 - + import .bib file
 - + run code
 - + download output files (predictions.csv, cleanedBib.csv)
4. **citation-diversity-tracking** :
 - + save cleanBib output files to */data/...*
 - + check out knitted example_code .html file
 - + edit code .Rmd file to your needs
 - + knit
 - + **Done** : checkout your own knitted .html file + */output/...* for your results!

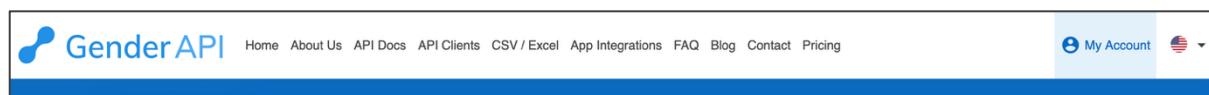
Disclaimer:

- I use Zotero (i.e., also in these instructions)
- gender-api.com has updated their policy: the last time I checked, they provided you with *500 search requests for free per month*. Should you run into problems reaching that limit, feel free to reach out to me ☺

[1]

<https://gender-api.com>

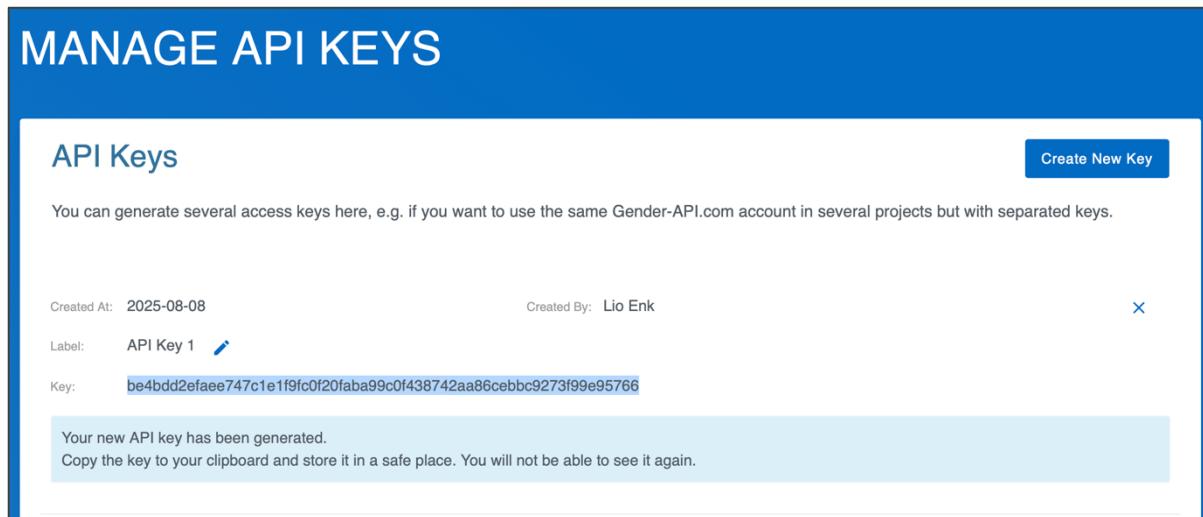
- (a) Create an account!
- (b) Log in
- (c) Click on ‘My Account’



- (d) Scroll down to box ‘Create New API Key’

A screenshot of a 'Create New API Key' form. The title 'Create New API Key' is at the top left. To the right is a 'Manage Existing Keys' button. Below the title, there's a note: 'Click on the following button to create a new API Key. Use the key for our REST API in the key parameter, as a bearer token, and in all our add-ons and extensions.' In the center is a blue 'Create New API Key' button. Below the button, there's a note: 'Add this key to your API calls:' followed by a text input field containing the URL 'https://gender-api.com/get?name=elizabeth&key=API_KEY'. The entire form is enclosed in a blue border.

- (e) Click on ‘Create New API Key’ or ‘Manage existing Keys’
 (f) Copy (and store) key, e.g.:
be4bdd2efae747c1e1f9fc0f20faba99c0f438742aa86cebbc9273f99e95766
 for later (→ cleanBib)



The screenshot shows a 'MANAGE API KEYS' interface. At the top, there's a blue header bar with the title 'MANAGE API KEYS'. Below it, a white section titled 'API Keys' contains a 'Create New Key' button. A message says: 'You can generate several access keys here, e.g. if you want to use the same Gender-API.com account in several projects but with separated keys.' Below this, a table lists a single key entry:

Created At:	2025-08-08	Created By:	Lio Enk	X
Label:	API Key 1			
Key:	<code>be4bdd2efae747c1e1f9fc0f20faba99c0f438742aa86cebbc9273f99e95766</code>			

A message at the bottom says: 'Your new API key has been generated. Copy the key to your clipboard and store it in a safe place. You will not be able to see it again.'

[2] Your citation management tool (Zotero/Mendeley/EndNote etc.)

- (a) Make sure to have a folder (in Zotero, called ‘(sub)collection’) including all your references. Especially, if you are using the gender-api.com account with 500 free search requests, it may be good not having to re-run the process again.

My example: Subcollection ‘Marshall et al. 2024’

Marshall et al 2024	Title	Creator
	▶ Cardiac cycle gated cognitive-emotional control in superior frontal cortices	Adelhöfer et al.
	▶ 'Bodily precision': a predictive coding account of individual differences in interoceptive accuracy	Ainley et al.
	▶ Heart-brain interactions shape somatosensory perception and evoked potentials	Al et al.
	▶ Afferent cardiac signals modulate attentional engagement to low spatial frequency fearful faces	Azevedo et al.
	▶ Cardiac afferent activity modulates the expression of racial stereotypes	Azevedo et al.
	▶ An Inventory for Measuring Depression	Beck et al.
	▶ Modulation of Human Muscle Spindle Discharge by Arterial Pulsations – Functional Effects and ...	Birznieks et al.
	▶ Auditory Feedback Differentially Modulates Behavioral and Neural Markers of Objective and Su...	Canales-Johnson et al.
	▶ The man who feels two hearts: the different pathways of interoception	Couto et al.
	▶ The influence of physiological signals on cognition	Critchley and Garfinkel
	▶ Visceral Influences on Brain and Behavior	Critchley and Harrison
	▶ Neurocognitive mechanisms of statistical-sequential learning: what do event-related potential...	Daltrozzo and Conway
	▶ Posterior Medial Frontal Cortex Activity Predicts Post-Error Adaptations in Task-Related Visual...	Danielmeier et al.
	▶ Global Impairment of Cardiac Autonomic Nervous Activity Late After Repair of Tetralogy of Fallot	Davos et al.
	▶ Heartbeat-Induced Axial Motion Artifacts in Optical Coherence Tomography Measurements of ...	de Kinkelder et al.
	▶ Aroused with heart: Modulation of heartbeat evoked potential by arousal induction and its osci...	Di Bernardi Luft and Bhat...
	▶ Is the P300 component a manifestation of context updating?	Donchin and Coles
	▶ Interceptive rhythms in the brain	Engelen et al.
	▶ Interceptive predictions in the brain	Feldman Barrett and Sim...
	▶ The influence of physical exercise on the relation between the phase of cardiac cycle and shoo...	Gallicchio et al.
	▶ Active sampling in visual search is coupled to the cardiac cycle	Galvez-Pol et al.
	▶ The influence of interoceptive signals on the processing of external sensory stimuli	Galvez-Pol et al.
	▶ Fear from the Heart: Sensitivity to Fear Stimuli Depends on Individual Heartbeats	Garfinkel et al.

- (b) **Before exporting it as .bib file, please go through your references and check the first names!**

You need:

- + full first names (A. → Anja) **# finding those may take time!**
- + preferred (first/middle) name at the start
(M. Veronika → Veronika)
- (P. Read → Read)
- + first names only with letters a-z,
no (!!!) non-ASCII characters (such as áàış)
(Hélène → Helene)
- (Aslı → Asli)
- (Paweł → Paweł)

The screenshot shows a Zotero item editor window. At the top, there are tabs: Info (which is selected), Notes, Tags, and Related. Below the tabs, the item details are listed:

- Item Type: Book
- Title: Book of examples
- Author 1: Muster, A. Robyn → *Robyn* (with edit, minus, plus buttons)
- Author 2: Muster, K. → *Kim* (with edit, minus, plus buttons)
- Author 3: Muster, Åsa → *Asa* (with edit, minus, plus buttons)
- Abstract
- Series

Why is this important? Gender-API does not recognize the names otherwise, and cleanBib (v2.2.3) is not non-ASCII character compatible (it just deletes them).

- (c) Once done, export the subcollection as .bib file.
That is, right click on the subcollection in the Zotero toolbar (to the left), a dropdown menu pops up, select 'Export Library'. On the next screen, choose format BibTex (default) and click OK. Save.

[3]
cleanBib
<https://github.com/dalejn/cleanBib>

The screenshot shows the GitHub repository page for 'cleanBib' by dalejn. The repository has 6 branches and 4 tags. The commit history shows several commits from dalejn, including patches for optional .aux file use, adding Florida data, cleaning up diversityStatement, fixing lossy img, trying third env, and updating Python and Seaborn. Other commits include breaking first cell into fn units, creating a LICENSE, and adding functions for finding unused citations. The repository has 155 stars, 32 forks, and 4 releases, with the latest being v1.1.3. The README and MIT license are also present.

- (a) Go to <https://github.com/dalejn/cleanBib>
 (b) Scroll down to Instructions, Point 2.

Instructions

The goal of the coding notebook is to analyze the predicted gender and race of first and last authors in reference lists of *manuscripts in progress*. The code will clean your `.bib` file to only contain references that you have cited in your manuscript. This cleaned `.bib` will then be used to generate a data table of names that will be used to query the probabilistic gender ([Gender API](#)) and race ([ethnicolr](#)) database. Proportions of the predicted gender for first and last author pairs (man/man, man/woman, woman/man, and woman/woman) and predicted race (white and author of color) will be calculated using the database probabilities.

- Can I use this code to analyze published article(s) instead of manuscripts in progress?

1. Obtain a `.bib` file of your manuscript's reference list. You can do this with common reference managers. Please try to export your `.bib` in an output style that uses full first names (rather than only first initials) and using the full author lists (rather than abbreviated author lists with "et al."). If a journal only provides first initials, our code will try to automatically find the full first name using the paper title or DOI (this can typically retrieve the first name 70% of the time).
 - [Export .bib from Mendeley](#)
 - [Export .bib from Zotero](#)
 - Tips for Microsoft Word and Google Docs integration
 - [Export .bib from EndNote](#). Note: Please export full first names by either [choosing an output style that does so by default \(e.g. in MLA style\)](#)
 - [Export .bib from Read Cube Papers](#)
2. Launch the coding environment. Please refresh the page if the Binder does not load after 5-10 mins.

Option 1 (recommended): [launch binder](#)

Option 2: Visit <https://notebooks.gesis.org/binder/>, paste <https://github.com/dalejn/cleanBib> into the GitHub repository name or URL field, and press "launch."
3. Open the notebook `cleanBib.ipynb`. Follow the instructions above each code block. It can take 10 minutes.

- (c) Copy <https://github.com/dalejn/cleanBib>
and click on <https://notebooks.gesis.org/binder/>
- (d) Paste <https://github.com/dalejn/cleanBib> into the box and click on LAUNCH.

Turn a Git repo into a collection of interactive notebooks
Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.
New to Binder? Get started with a [Zero-to-Binder tutorial](#) in Julia, Python, or R.

Build and launch a repository

GitHub repository name or URL

Git ref (branch, tag, or commit)
 File to open (in JupyterLab)

<https://mybinder.org/v2/gh/%20https://notebooks.gesis.org/binder/HEAD>

HEAD

<https://mybinder.org/v2/gh/dalejn/cleanBib/HEAD>

Badges for your README

Waiting Building Pushing Launching

- (e) Click on **cleanBib.ipynb** → opens

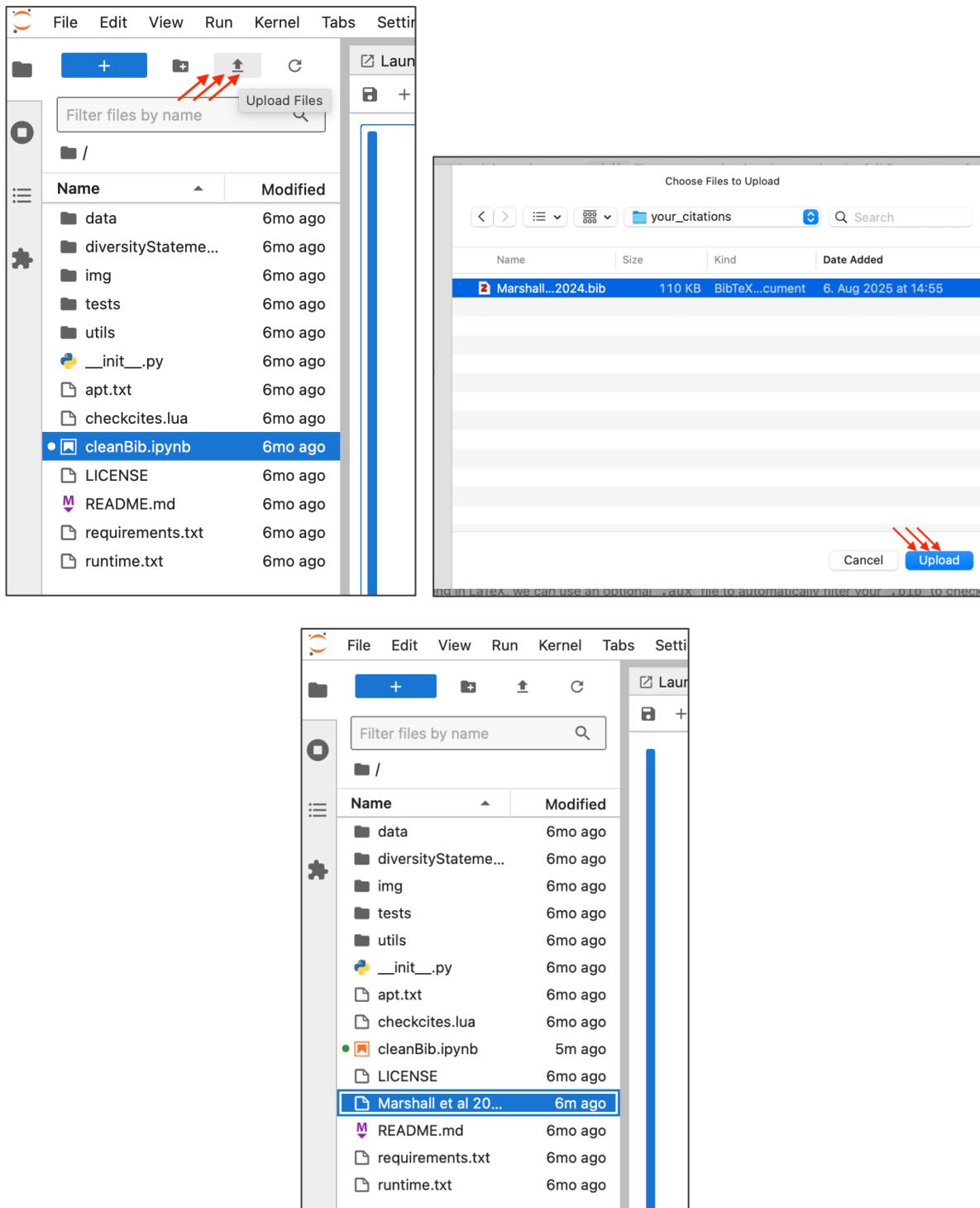
File Edit View Run Kernel Tabs Settings

+ Filter files by name

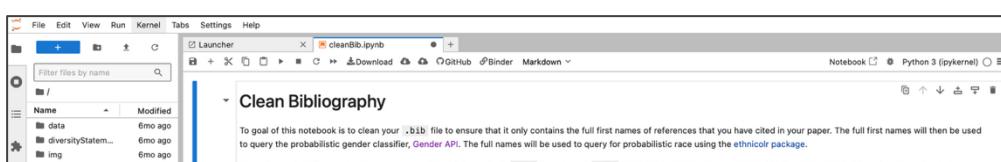
/

Name	Modified
data	6mo ago
diversityStateme...	6mo ago
img	6mo ago
tests	6mo ago
utils	6mo ago
__init__.py	6mo ago
apt.txt	6mo ago
checkcites.lua	6mo ago
cleanBib.ipynb	6mo ago
LICENSE	6mo ago
README.md	6mo ago
requirements.txt	6mo ago
runtime.txt	6mo ago

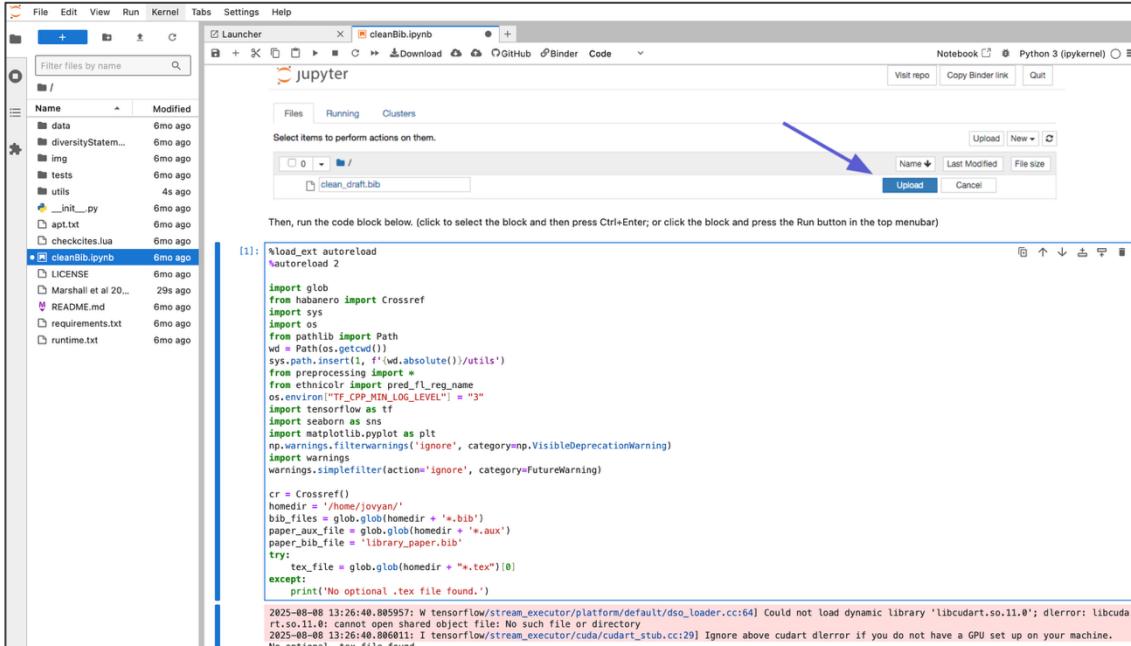
- (f) Click on **Upload button** in horizontal toolbar, browse + select your .bib file and upload it. You should be able to see it listed.



- (g) Turn to opened **cleanBib.ipynb** code
 (h) Read through it, if you like. You can also just follow my instructions in here.



- (i) Run first piece of code # nothing needs to be added
 (click onto the code cell, shortcut **%command+Enter** runs the code – as indicated by the asterisks in square brackets next to the top left corner)
 (when run, the asterisks in square brackets turns into number 1)



```
[1]: %load_ext autoreload
%autoreload 2

import glob
from homedir import Crossref
import sys
import os
from pathlib import Path
wd = Path(os.getcwd())
sys.path.insert(1, f'{wd.absolute()}/utils')
from ete3 import NCBITaxa
from ete3.ncbi import NCBIRegName
os.environ['TF_CPP_MIN_LOG_LEVEL'] = "3"
import tensorflow as tf
import seaborn as sns
import matplotlib.pyplot as plt
np.warnings.filterwarnings('ignore', category=np.VisibleDeprecationWarning)
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)

cr = Crossref()
homadir = '/home/jovyan/'
bib_files = glob.glob(homadir + '*.bib')
paper_aux_file = glob.glob(homadir + '*.aux')
paper_bib_file = 'library_paper.bib'
try:
    tex_file = glob.glob(homadir + '*.tex')[0]
except:
    print('No optional .tex file found.')

2025-08-08 13:26:48.984957: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0'; dlsym error: libcudart.so.11.0: cannot open shared object file: No such file or directory
2025-08-08 13:26:48.886011: I tensorflow/stream_executor/cuda/cudart_stub.cc:20] Ignore above cudart dlsym error if you do not have a GPU set up on your machine.
```

(You can ignore the error message)

- (j) The next *code cell* allows you to add the names of the first authors (FA) and last author (LA). If you do, the code automatically deletes those references that include those names at first or last positions from the reference list (in your .bib file).
- To have a better overview and possibly analyse self-citation behaviour later on, I would not recommend doing that. Instead, I would suggest to add any other irrelevant fictive name.
- Insert + run the piece of code.

Common issues include:

- Bibliography entry did not include a last author because the author list was truncated by "and Others" or "et al."
- Some older journals articles only provide first initial and not full first names, in which case you will need to go digging via Google to identify that person.
- In rare cases where the author cannot be identified even after searching by hand, replace the first name with "UNKNOWNNAMES" so that the classifier will estimate the gender as unknown.

```
[1]: yourFirstAuthor = 'LastName, FirstName OptionalMiddleInitial' 'Muster, Kim'
yourLastAuthor = 'LastName, FirstName OptionalMiddleInitial' 'Master, Robyn'
optionalEqualContributors = ['LastName, FirstName OptionalMiddleInitial', 'LastName, FirstName OptionalMiddleInitial']
checkingPublishedArticle = False

if paper_aux_file:
    unused_keys = [find_unused_cites([file]) for file in paper_aux_file]
    unused_keys = [item for sublist in unused_keys for item in sublist]
else:
    unused_keys = None

bib_data = get_bib_data(bib_files[0])
if checkingPublishedArticle:
    get_names_published(homadir, bib_data, cr)
else:
    # find and print duplicates
    bib_data = get_duplicates(bib_data, bib_files[0])

    # get names, remove CDS, find self cites
    get_names(homadir, bib_data, yourFirstAuthor, yourLastAuthor, optionalEqualContributors, cr, unused_keys)

bib_check(homadir)
```

→ [1]: yourFirstAuthor = 'Muster, Kim'
 yourLastAuthor = 'Master, Robyn OptionalMiddleInitial'
 optionalEqualContributors = ['LastName, FirstName Optiona
 checkingPublishedArticle = False

This process triggers a check of each of the references listed in the .bib file (watch as references are checked step-by-step). When done, the code informs you about specific references you should check another time.

```

bib_check(noneatty)
1: al_heartbeat_2021
2: motyka_interactions_2019
3: al_heartbrain_2020
4: whiteside_five_2001
5: wilkinson_electrocutaneous_2013
6: seth_interoceptive_2013
7: suzuki_multisensory_2013
8: schulte-korne_motion-onset_2004
9: verbruggen_fictitious_2013
10: verbruggen_proactive_2014
11: di_bernardi_luft_aroused_2015
12: macefield_cardiovascular_2003
13: makowski_heart_2020
14: marshall_exteroceptive_2017
15: marshall_cardiac_2018
16: marshall_inside_2022
17: ohl_microsaccades_2016
18: palser_relationship_2021
19: petzschner_focus_2019
20: poreh_bpd_2006
21: potts_when_2006
22: rae_response_2018
23: rae_impact_2020
24: ren_response_2022
25: richter_phase-amplitude_2017
26: russeler_implicit_2000
27: soder_role_2016
28: san_martin_size_2010
29: salomon_insula_2016
30: de_kinkelder_heartbeat-induced_2011
31: edwards_pain-related_2008
32: gallicchio_influence_2019
33: galvez-pol_active_2020
34: garfinkel_fear_2014
35: gray_baroreceptor_2010
36: jarosz_what_2014
37: kunzendorf_active_2019
38: larra_disenantangling_2020
39: azevedo_cardiac_2017
40: azevedo_afferent_2018
41: feldman_barrett_interoceptive_2015
42: birznieks_modulation_2012
43: canales-johnson_auditory_2015
44: couto_man_2014
45: critchley_influence_2018
46: daltrozzo_neurocognitive_2014
47: danielmeier_posterior_2011
48: engelen_interoceptive_2023
49: spielberger_manual_1983
50: beck_inventory_1961
51: adelhofer_cardiac_2020
52: maris_nonparametric_2007
53: pramme_cardiac_2016
54: pramme_cardiac_2014

STOP: Please revise incomplete full first names or empty cells. Then, re-run step 2.
Here are some suggestions to check for with the following citation keys in your .bib
file:
['potts_when_2006', 'feldman_barrett_interoceptive_2015']

```

Example: go check 'potts_when_2006' and 'feldman_barrett_interoceptive_2015' once more!

You do this by clicking on your .bib file (listed to the left). This opens the .bib file in a new tab. Use shortcut **⌘+F** to search for the specific references (type in + Enter).

Examples:

In both cases, authors preferred their middle name and we forgot to delete the initial (P. / W.) of their actual first first name. Please delete that initial and move on. #You could, of course also search for their full first name, if you want

```
.pdf:/Users/liobaenk/Zotero/storage/S46UFIX1 potts
Assessment of Borderline .pdf:application/pdf
368 }
369
370 @article{potts_when_2006,
371   title = {When {Things} {Are} {Better} or {Worse} than {Expected}: {The} {Medial}
372   {Frontal} {Cortex} and the {Allocation} of {Processing} {Resources}},
373   volume = {18},
374   issn = {0898-929X, 1530-8898},
375   shorttitle = {When {Things} {Are} {Better} or {Worse} than {Expected}},
376   url = {https://direct.mit.edu/jocn/article/18/7/1112/4204/When-Things-Are-Better-or-Worse-
377   than-Expected-The},
378   doi = {10.1162/jocn.2006.18.7.1112},
379   abstract = {Access to limited-capacity neural systems of cognitive control must be
380   restricted to the most relevant information. How the brain identifies and selects items for
381   preferential processing is not fully understood. Anatomical models often place the selection
382   mechanism in the medial frontal cortex (MFC), and one computational model proposes that the
383   mesotelencephalic dopamine (DA) system, via its reward prediction properties, provides a
384   "gate" through which information gains access to limited-capacity systems. There is a medial
385   frontal event-related potential (ERP) index of attention selection, the anterior positivity
386   (P2a), associated with DA reward system input to the MFC for the identification of task-
387   relevant perceptual representations. The P2a has a similar spatio-temporal distribution as
388   the medial frontal negativity (MFN), elicited to error responses or choices resulting in
389   monetary loss. The MFN has also been linked to DA projections to the MFC but for action
390   monitoring rather than attention selection. This study proposes that the P2a and the MFN
391   reflect the same MFC evaluation function and use a passive reward prediction design
392   containing neither instructed attention nor response to demonstrate that the ERP over medial
393   frontal leads at the P2a/MFN latency is consistent with activity of midbrain DA neurons,
394   positive to unpredicted rewards and negative when a predicted reward is withheld. This result
395   suggests that MFC activity is regulated by DA reward system input and may function to
396   identify items or actions that exceed or fail to meet motivational prediction.},
397   language = {en},
398   number = {7},
399   urldate = {2023-12-14},
400   journal = {Journal of Cognitive Neuroscience},
401   author = {Potts, Geoffrey F. and Martin, Laura E. and Burton, Philip and Montague, P.-
402   Read},
403   month = jul,
404   year = {2006},
405   pages = {1112--1119},
406   file = {Potts et al. - 2006 - When Things Are Better or Worse than Expected
407   The.pdf:/Users/liobaenk/Zotero/storage/L3VN8R6M/Potts et al. - 2006 - When Things Are Better
408   or Worse than Expected The.pdf:application/pdf},
409 }
410
signals modulate attentional enga.pdf:appli feldman
737 }
738
739 @article{feldman_barrett_interoceptive_2015,
740   title = {Interoceptive predictions in the brain},
741   volume = {16},
742   issn = {1471-003X, 1471-0048},
743   url = {https://www.nature.com/articles/nrn3950},
744   doi = {10.1038/nrn3950},
745   abstract = {Intuition suggests that perception follows sensation and therefore bodily
746   feelings originate in the body. However, recent evidence goes against this logic:
747   interoceptive experience may largely reflect limbic predictions about the expected state of
748   the body that are constrained by ascending visceral sensations. In this Opinion article, we
749   introduce the Embodied Predictive Interoception Coding model, which integrates an anatomical
750   model of corticocortical connections with Bayesian active inference principles, to propose
751   that agranular visceromotor cortices contribute to interoception by issuing interoceptive
752   predictions. We then discuss how disruptions in interoceptive predictions could function as a
753   common vulnerability for mental and physical illness.},
754   language = {en},
755   number = {7},
756   urldate = {2023-12-14},
757   journal = {Nature Reviews Neuroscience},
758   author = {Feldman Barrett, Lisa and Simmons, W.-Kyle},
759   month = jul,
760   year = {2015},
761   pages = {419--429},
762   file = {Barrett and Simmons - 2015 - Interoceptive predictions in the
763   brain.pdf:/Users/liobaenk/Zotero/storage/BJQE6AFQ/Barrett and Simmons - 2015 - Interoceptive
764   predictions in the brain.pdf:application/pdf},
765 }
```

Important : Don't forget to save the file (File → Save File / ⌘command+S), when you are done editing!

Re-run the piece of code... repeat until you receive the following note:
Only continue if you've run step 2, and this code no longer returns error or instructions to revise the .bib file. # see below

```
bib_check(homedir)
1: al_heartbeat_2021
2: motyka_interactions_2019
3: al_heartbrain_2020
4: whiteside_five_2001
5: wilkinson_electrocuteaneous_2013
6: seth_interoceptive_2013
7: suzuki_multisensory_2013
8: schulte-korne_motion-onset_2004
9: verbruggen_fictitious_2013
10: verbruggen_proactive_2014
11: di_bernardi_luft_aroused_2015
12: macefield_cardiovascular_2003
13: makowski_heart_2020
14: marshall_exteroceptive_2017
15: marshall_cardiac_2018
16: marshall_inside_2022
17: ohl_microsaccades_2016
18: palser_relationship_2021
19: petzschner_focus_2019
20: poreh_bpq_2006
21: potts_when_2006
22: rae_response_2018
23: rae_impact_2020
24: ren_response_2022
25: richter_phase-amplitude_2017
26: russeler_implicit_2000
27: soder_role_2016
28: san_martin_size_2010
29: salomon_insula_2016
30: de_kinkelder_heartbeat-induced_2011
31: edwards_pain-related_2008
32: gallicchio_influence_2019
33: galvez-pol_active_2020
34: garfinkel_fear_2014
35: gray_baroreceptor_2010
36: jarosz_what_2014
37: kunzendorf_active_2019
38: larra_disentangling_2020
39: azzevedo_cardiac_2017
40: azzevedo_afferent_2018
41: feldman_barrett_interoceptive_2015
42: birznieks_modulation_2012
43: canales-johnson_auditory_2015
44: couto_man_2014
45: critchley_influence_2018
46: daltrozzo_neurocognitive_2014
47: danielmeier_posterior_2011
48: engelen_interoceptive_2023
49: spelberger_manual_1983
50: beck_inventory_1961
51: adelhofer_cardiac_2020
52: maris_nonparametric_2007
53: pramme_cardiac_2016
54: pramme_cardiac_2014

Only continue if you've run step 2, and this code no longer returns error or instructions to revise the .bib file.
```

- (k) Scroll further down. You now arrive at section ‘3. Estimate gender and race of authors from cleaned bibliography’. Please insert your Gender API key in the code cell, and run the code.

When done so (see below), you receive a notification on how many more search requests (or ‘credits’) you have available. #In this example, it is N=4146.

Then, run the code blocks below to estimate how many credits we will need to use. (click to select the block and then press Ctrl+Enter; or click the block and press the Run button in the top menubar)

```
[ ]: genderAPI_key = '&key='
      # Check your credit balance
      check_genderAPI_balance(genderAPI_key, homedir)

[7]: genderAPI_key = '&key=9e639492d0967f683ff2306e4400b83f061441288a1ba655c8a675b739b1941d'
      # Check your credit balance
      check_genderAPI_balance(genderAPI_key, homedir)
      Remaining credits: 4146
      This should use (at most) 82 credits, saving you approx 26 credit(s) by storing queries.
```

- (l) Run the following piece of code. You can watch as the references are being checked. → 100% 54/54

4. Describe the proportions of genders in your reference list and compare it to published base rates in neuroscience.

NOTE: your free GenderAPI account has 500 queries per month. This box contains the code that will use your limited API credits/queries if it runs without error. Re-running all code repeatedly will repeatedly use credits.

Run the code blocks below. (click to select the block and then press Ctrl+Enter; or click the block and press the Run button in the top menubar)

```
[ ]: mm, wm, mw, ww, aw, aa, citation_matrix, paper_df = get_preg_demos((yourFirstAuthor+' '+yourLastAuthor).replace(' ', '_'), homedir, bib_data, genderAPI_key, unused_keys)
statement, statementLatex = print_statements(mm, wm, mw, ww, aw, aa)
first author is Kim Muster
last author is Robyn Master
we don't count these, but check the predictions file to ensure your names did not slip through!
looping through your references, predicting gender and race
  0%| [ 0/54 [00:00:00-?, ?it/s]2025-08-08 14:06:35.006270: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcuda.so.1'
  cuda.so.1: cannot open shared object file: No such file or directory
2025-08-08 14:06:35.006337: W tensorflow/stream_executor/cuda/cuda_driver.cc:269] failed call to cuInit: UNKNOWN ERROR (303)
2025-08-08 14:06:35.006373: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel driver does not appear to be running on this host (jupyter-dalejn-cleanbib-5cdej90
ver/nvidia/version does not exist
2025-08-08 14:06:35.008383: I tensorflow/core/platform/cpu_feature_guard.cc:151] This TensorFlow binary is optimized with oneAPI Deep Neural Network Library (oneDNN) to use the
U instructions in performance-critical operations: AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
85%|██████████ | 46/54 [00:08<00:01, 6.98it/s]
```

- (m) Run the remaining bit of code. This will also produce the output needed:

predictions.csv
cleanedBib.csv

(race_gender_citations.pdf) # see my comments in the Q&A section of the README.md file in <https://github.com/liobaenk/citation-diversity-tracking>

```
the top menubar)
```

```
[9]: print('Plain text template:')
print(statement)
print('\n')
print('LaTeX template:')
print(statementLatex)

paper_df.to_csv('/home/jovyan/predictions.csv')

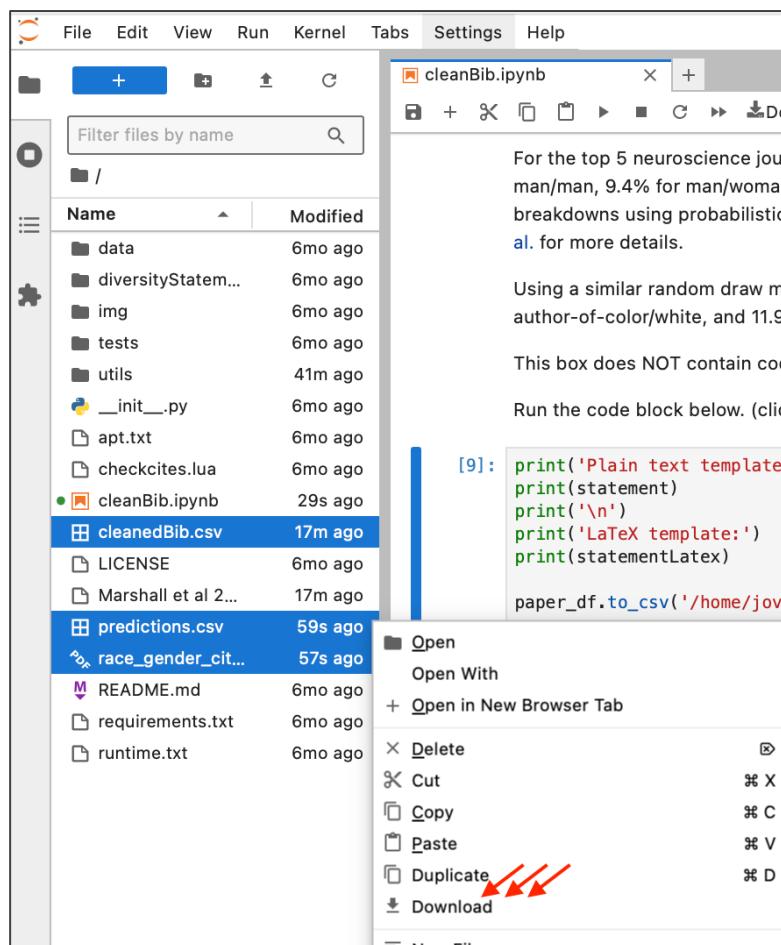
plot_heatmaps(citation_matrix, homedir)

plot_histograms()

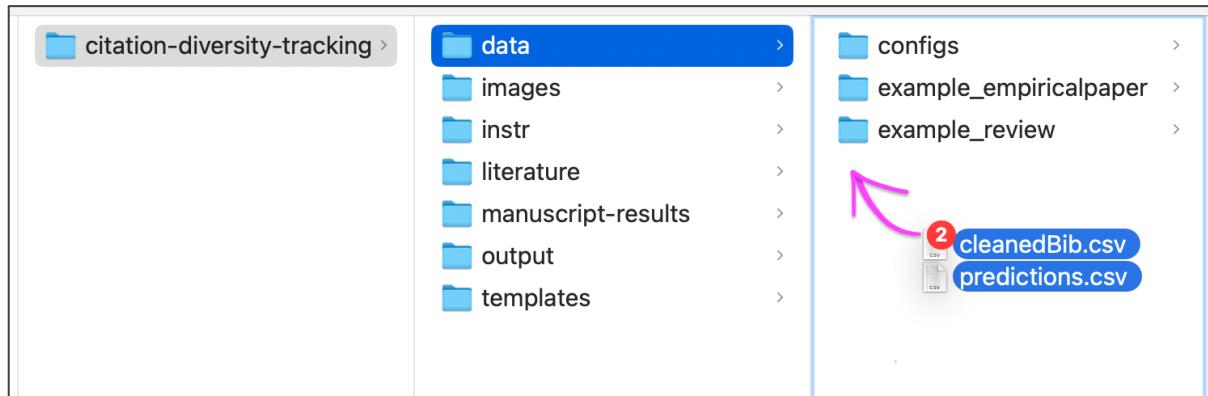
Plain text template:
Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minority scholars are under-cited relative to the number of such papers in the field (1–9). Here we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, race, ethnicity, and other factors. First, we obtained the predicted gender of the first and last author of each reference by using databases that store the probability of a first name being carried by a woman (5, 10). By this measure and excluding self-citations to the first and last authors of our current paper, our references contain 3.7% woman(first)/woman(last), 11.11% man/woman, 35.19% woman/man, and 50.0% man/man. This method is limited in that a) names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity and b) it cannot account for intersex, non-binary, or transgender people. Second, we obtained predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of color (11, 12). By this measure (and excluding self-citations), our references contain 7.75% author of color(first)/author of color(last), 14.30% white author/author of color, 21.49% author of color/white author, and 56.46% white author/white author. This method is limited in that a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity, and b) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

LaTeX template:
Recent work in several fields of science has identified a bias in citation practices such that papers from wom
```

(n) Download the output files



- (o) Move the downloaded output files into folder `/data/...` of *citation-diversity-tracking*



DONE !!!

All further explanations are provided in **citation-diversity-tracking**.
example_code .html
code .Rmd file