

# How (not) to bring Wikipedia into the classroom

JSM 2020

Amelia McNamara [@AmeliaMN](#)  
University of St Thomas

# **Communicating data**

Courses I've taught:

- STAT 98T: Data Visualization (UCLA, 2015)
- SDS 136: Communicating with Data (Smith, 2017)
- SDS 236: Data Journalism (Smith, 2018)
- STAT 336: Data Communication and Visualization (St Thomas, 2020)

**Visualizing data**

**Writing** about data

**Speaking** about data

# Writing about data

- “One number story” (Data Journalism, Data Communication)
- Science reporting (Data Journalism)
- “Standard story” (Data Journalism)
- ...
- Wikipedia article? (Data Journalism, Data Communication)



WIKIPEDIA  
The Free Encyclopedia

[Main page](#)  
[Contents](#)  
[Current events](#)  
[Random article](#)  
[About Wikipedia](#)  
[Contact us](#)  
[Donate](#)

Contributo

Help

[Community portal](#)  
[Recent changes](#)  
[Upload file](#)

Tools

[What links here](#)  
[Related changes](#)  
[Special pages](#)  
[Page information](#)

# (Learning) goals for writing Wikipedia articles

Editing Amanda Cox

## Notice about sources

This article must adhere to the [biographies of living persons policy](#), even if it is not a biography, because it contains material about living persons. Take extra care to use high-quality sources. Material about living persons should not be added when the journalism; see [more information on sources](#). Never use self-published sources about a living person unless written or published by the subject: see [WP:BLPSPS](#) and [WP:BLPSELF PUB](#).

Contentious material about living persons that is unsourced or [poorly sourced](#) **must be removed immediately** from the article and its talk page, especially if potentially [libellous](#). If such material is repeatedly inserted, or if you have other concerns, please [noticeboard](#). If you are connected to one of the subjects of this article and need help, see [this page](#).

*Content that violates any copyrights will be deleted. Encyclopedic content must be verifiable. Work submitted to Wikipedia can be edited, used, and redistributed—by anyone—subject to certain terms and conditions.*

[B](#) [I](#) [G](#) [M](#) [W](#) [H](#) [F](#) [A](#) [Advanced](#) [Special characters](#) [Help](#) [C](#)

}}  
'''Amanda Cox''' is an American journalist and the editor of the ''[[New York Times]]'' [[data journalism|Journalism.<ref name="NYU18">{{cite web | title=Amanda Cox | website=NYU Journalism | date=Oct 2018}}]]

— Life and education —

Cox was born in Michigan in 1980, and raised by her accountant parents. [\[1\]](#)  
She received her master's degree in statistics from the University of Washington. [\[2\]](#)

—Garrison and Wm. C. Brown—

**== Career and research ==**

She began her career at the "New York Times" as a summer intern while in graduate school.at-the-new|publisher=SimplyStatistics|date=June 1, 2012}} Cox worked at the Federal Reserve Times – Breaking News, World News & Multimedia | date=August 28, 2018 | url=https://www.nytimes.com/2018/08/28/us/politics/federal-reserve-times.html

In her years at the Times, Cox has worked on many stories using statistics and data visualization, including "The Power of Visualization,"url=https://hbr.org/2013/03/power-of-visualizations-aha-moment|title=The Power of Visualization and "From Service Desk to Standalone News Desk: How the New York Times' Graphics Department Has Grown Up."last1=Doctor|first1=Ken|publisher=NiemanLab|date=March 7, 2016|url=http://www.niemanlab.org/2016/03/from-service-desk-to-standalone-news-desk-how-the-new-york-times-graphic-department-has-grown-up/

On April 22, 2014 the "New York Times" website launched{cite web|url=https://www.nytimes.com/2014/04/23/technology/the-new-york-times-launches-data-journalism-site-the-upshot.html|title=New York Times launches data journalism site|date=April 22, 2014|author-link=David Leonhardt}} its data journalism section{cite news|last=Baquet|first=Dean|title=Amanda Cox Named Editor, The Upshot|url=https://qz.com/185922/the-upshot-is-the-new-york-times-replacement-for-nate-silvers-five-new-editors/|date=March 10, 2014|work=Quartz}}Cox was named editorname="named">{cite news|last=Baquet|first=Dean|title=Amanda Cox Named Editor, The Upshot|url=https://qz.com/185922/the-upshot-is-the-new-york-times-replacement-for-nate-silvers-five-new-editors/|date=March 10, 2014|work=Quartz}}

Cite your sources: [<ref></ref>](#)

**Edit summary** (Briefly describe your changes)

This is a minor edit  Watch this page

By publishing changes, you agree to the [Terms of Use](#), and you irrevocably agree to release your contribution under the [CC BY-SA](#) license.

[Publish changes](#) [Show preview](#) [Show changes](#) [Cancel](#)

- A new form of typesetting.
  - Quality reference-finding skills.
  - A neutral writing style.
  - Some form of versioning.
  - Deep knowledge about a particular person.
  - A new Wikipedia article!

# How ~~not~~ to do this assignment

- ~~First assignment of class~~
- ~~Guidance supplied by professor~~
- ~~Guided by Wiki Education~~
- ~~Scaffolded by professor~~
- Scaffolded using deadlines and milestones from Wiki Education
- ~~Students think of people to write articles about~~
- Professor generates list of potential people

The screenshot shows a web browser window with the URL <https://dashboard.wikiedu.org/courses>. The page title is "Data Communication and Visual". The navigation bar includes links for Home, Timeline (which is active), Students, Articles, Uploads, Activity, and Resources. A "Log in" link is also present. The main content area is divided into sections for each week of the assignment timeline:

- Week 2** (02/09 - 02/15):
  - Introduction to the Wikipedia assignment** (In Class): Welcome message, assignment details, and resources for editing and evaluating Wikipedia.
  - Get started on Wikipedia** (Assignment): Training modules for Wikipedia policies and sandboxes.
  - Everyone has a Wikipedia account** (Milestone): A reminder for students to create their accounts.
- Week 3** (02/16 - 02/22)
- Week 4** (02/23 - 02/29)
- Week 5** (03/01 - 03/07)
- Week 6** (03/08 - 03/14)
- Week 7** (03/15 - 03/21)
- Week 8** (03/22 - 03/28)
- Week 9** (03/29 - 04/04)
- Week 10** (04/05 - 04/11)
- Week 11** (04/12 - 04/18)
- Week 12** (04/19 - 04/20)

# List + brief bio helps students get started

Wikipedia-people (page 23 of 24)

Luke Tierney

Maxine Pfannkuch

Nadieh Bremer

Regina Nuzzo

Rob Simon

Roger Peng

18

19

20

21

22

23

# Roger Peng

- Biostatistics professor at Johns Hopkins
- Co-host of Not So Standard Deviations podcast
- One of the creators of the Johns Hopkins Data Science Specialization on Coursera  
<https://qz.com/342712/you-dont-have-to-go-to-wharton-to-get-access-to-tech-giants/>
- See <http://www.biostat.jhsph.edu/~rpeng/>  
<https://scholar.google.com/citations?user=h5wUydwAAAAJ&hl=en&oi=ao>



# Articles that made it

Spring 2018

Amanda Cox

Giorgia Lupi

Janaya Khan

Jeff Leek

Jenny Bryan

Jer Thorp

Kim Rees

Lena Groeger

Lynn Cherny

Mark Hansen

Meredith Broussard

Stefanie Posavec

Victoria Stodden

Spring 2020

Adam Harvey

Andreas Buja

Antony Unwin

Catherine D'Ignazio

Colin Ware

David Robinson

Jen Christiansen

Karl Broman

Kyle McDonald

Nadieh Bremer

Regina Nuzzo

Robert Simmon

Roger Peng

Relevant to  
today's session!

16

19

353

19<sup>i</sup>

15.8K

249<sup>i</sup>

143K

2<sup>i</sup>

Articles Created

Articles Edited

Total Edits

Student Editors

Words Added

References Added

Article Views

Commons Uploads

Last statistics update: 2 months ago. [See more](#)

## Data Communication and Visualization

[Edit Description](#)

This is a data communication course. Students will be learning to visualize data, write about data, and speak about data. I'm hoping to have students write/contribute to articles (as a writing assignment) as well as create visualizations from Wikipedia tables (as a visualization assignment). I'm not sure if the visualization piece will make sense to do through this platform.

## Details

[Edit Details](#)

Instructors: [AmeliaMN](#) (Amelia McNamara / amelia.mcnamara@stthomas.edu — Instructor)

- 16 articles created
- 19 articles edited
- 353 total edits
- 19 student editors
- 15.8k words added
- 249 references added
- 143k article views
- 2 commons uploads

# Articles that made it

Spring 2018

Amanda Cox

Giorgia Lupi

Janaya Khan

Jeff I am super proud of my students' work. They clearly made valuable additions to Wikipedia!  
Jenn And, perhaps added to the diversity of the site

Jer Thorp

Kim Rees

Lena Groeger

Lynn Cherny

Mark Hansen

Meredith Broussard

Stefanie Posavec

Victoria Stodden

Spring 2020

Adam Harvey

Andreas Buja

Antony Unwin

David Robinson

Jen Christiansen

Karl Broman

Kyle McDonald

Nadieh Bremer

Regina Nuzzo

Robert Simmon

Roger Peng

**But...**

# **Bringing Wikipedia into the classroom isn't always fun**

- Editors are often brusque and/or rude
- Notability guidelines are both clear-cut and very open to interpretation
- Students do unexpected things
- It can be hard to see something you've worked on get deleted
- The instructor needs to be extremely plugged in to watch and manage pages, even when using the Wiki Education framework

# Wikipedia:Conflict of interest

Wikipedia has guidelines about conflicts of interest. You shouldn't write an article about yourself, your friends, your employer, etc.

I come up with a list of suggested people for my students to write articles about (I don't think this is a conflict of interest?) and often, that is the first time a student has heard the person's name

Still, a common criticism of articles my students write is “self promotion”

“the page is nauseating in terms of its self promotion”

“overtly promotional”

# Wikipedia:Notability (people)

People are presumed **notable** if they have received significant coverage in **multiple published secondary sources** that are **reliable**, **intellectually independent** of each other, and **independent of the subject**.

- If the depth of coverage in any given source is not substantial, then multiple independent sources may be combined to demonstrate notability; trivial coverage of a subject by secondary sources may not be sufficient to establish notability.
- Primary sources may be used to support content in an article, but they do not contribute toward proving the notability of a subject.

People who meet the basic criteria may be considered notable without meeting the additional criteria below. Articles may still not be created for such people if they fall under exclusionary criteria, such as being notable only for a single event, or such as those listed in Wikipedia:What Wikipedia is not.

# Wikipedia:Notability (academics)

Academics meeting **any one** of the following conditions, as substantiated through reliable sources, are notable. Academics meeting none of these conditions *may* still be notable if they meet the conditions of WP:BIO or other notability criteria. The merits of an article on the academic will depend largely on the extent to which it is verifiable. **Before applying these criteria, see the General notes and Specific criteria notes sections, which follow.**

1. The person's research has had a significant impact in their scholarly discipline, broadly construed, as demonstrated by independent reliable sources.
2. The person has received a highly prestigious academic award or honor at a national or international level.
3. The person has been an elected member of a highly selective and prestigious scholarly society or association (e.g., a National Academy of Sciences or the Royal Society) or a fellow of a major scholarly society which reserves fellow status as a highly selective honor (e.g., Fellow of the Institute of Electrical and Electronics Engineers).
4. The person's academic work has made a significant impact in the area of higher education, affecting a substantial number of academic institutions.
5. The person has held a named chair appointment or distinguished professor appointment at a major institution of higher education and research, or an equivalent position in countries where named chairs are uncommon.
6. The person has held a highest-level elected or appointed administrative post at a major academic institution or major academic society.
7. The person has had a substantial impact outside academia in their academic capacity.
8. The person has been the head or chief editor of a major, well-established academic journal in their subject area.

# Notability

“He seems to be notable, based on the very highly cited papers shown in Google Scholar. Seeing them, after I had determined that the book was self-published, I was rather skeptical, & checked the contents, because statisticians are frequently joint authors for papers in biomedicine, but contributed to the work as a statistician only, not as the person who conceived the project. In earlier years, the statistician was usually acknowledged at the end of the paper; under current practice, the statistician is a co-authors [sic]. That seems to be the case for some of the papers, but the 4 most highly cited are papers about biostatistics. The mist [sic] highly cited is a rather pedagogical review article, & such articles have extremely high & uncharacteristic citation counts. The next three however are research papers, and are highly cited because they are the foundations for very important & widespread biological work.”

# Notability

“He seems to be notable, **based on the very highly cited papers shown in Google Scholar**. Seeing them, after I had determined that the book was self-published, I was rather skeptical, & checked the contents, because statisticians are frequently joint authors for papers in biomedicine, but contributed to the work as a statistician only, not as the person who conceived the project. In earlier years, the statistician was usually acknowledged at the end of the paper; under current practice, the statistician is a co-authors [sic]. That seems to be the case for some of the papers, but the 4 most highly cited are papers about biostatistics. The mist [sic] highly cited is a rather pedagogical review article, & such articles have extremely high & uncharacteristic citation counts. The next three however are research papers, and are highly cited because they are the foundations for very important & widespread biological work.”

# Notability

“He seems to be notable, based on the very highly cited papers shown in Google Scholar. Seeing them, after I had determined that the book was self-published, I was rather skeptical, & checked the contents, because **statisticians are frequently joint authors for papers in biomedicine, but contributed to the work as a statistician only**, not as the person who conceived the project. In earlier years, the statistician was usually acknowledged at the end of the paper; under current practice, the statistician is a co-authors [sic]. That seems to be the case for some of the papers, but the 4 most highly cited are papers about biostatistics. The mist [sic] highly cited is a rather pedagogical review article, & such articles have extremely high & uncharacteristic citation counts. The next three however are research papers, and are highly cited because they are the foundations for very important & widespread biological work.” ??

# Notability

“He seems to be notable, based on the very highly cited papers shown in Google Scholar. Seeing them, after I had determined that the book was self-published, I was rather skeptical, & checked the contents, because statisticians are frequently joint authors for papers in biomedicine, but contributed to the work as a statistician only, not as the person who conceived the project. In earlier years, the statistician was usually acknowledged at the end of the paper; under current practice, the statistician is a co-authors [sic]. That seems to be the case for some of the papers, but the 4 most highly cited are papers about biostatistics. The mist [sic] highly cited is a rather **pedagogical review article, & such articles have ??? extremely high & uncharacteristic citation counts.** The next three however are research papers, and are highly cited because they are the foundations for very important & widespread biological work.”



## Jeff Leek

FOLLOW

Professor of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Verified email at jhsph.edu - [Homepage](#)

Statistics Computing Genomics Personalized Medicine Scientific Communication

Cited by

[VIEW ALL](#)

All

Since 2015

	All	Since 2015
Citations	12498	9550
h-index	40	37
i10-index	64	63

TITLE

CITED BY

YEAR

**The sva package for removing batch effects and other unwanted variation in high-throughput experiments**

JT Leek, WE Johnson, HS Parker, AE Jaffe, JD Storey  
Bioinformatics 28 (6), 882-883

1568

2012

**Capturing heterogeneity in gene expression studies by surrogate variable analysis**

JT Leek, JD Storey  
PLoS Genet 3 (9), e161

1348

2007

**Tackling the widespread and critical impact of batch effects in high-throughput data**

JT Leek, RB Scharpf, HC Bravo, D Simcha, B Langmead, WE Johnson, ...  
Nature Reviews Genetics 11 (10), 733-739

1338

2010

**Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown**

M Pertea, D Kim, GM Pertea, JT Leek, SL Salzberg  
Nature protocols 11 (9), 1650

1235

2016

**Significance analysis of time course microarray experiments**

JD Storey, W Xiao, JT Leek, RG Tompkins, RW Davis  
Proceedings of the National Academy of Sciences 102 (36), 12837-12842

644

2005

**Temporal dynamics and genetic control of transcription in the human prefrontal cortex**

C Colantuoni, BK Lipska, T Ye, TM Hyde, R Tao, JT Leek, EA Colantuoni, ...  
Nature 478 (7370), 519-523

591

2011

**Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies**

AE Jaffe, P Murakami, H Lee, JT Leek, MD Fallin, AP Feinberg, RA Irizarry  
International journal of epidemiology 41 (1), 200-209

439

2012

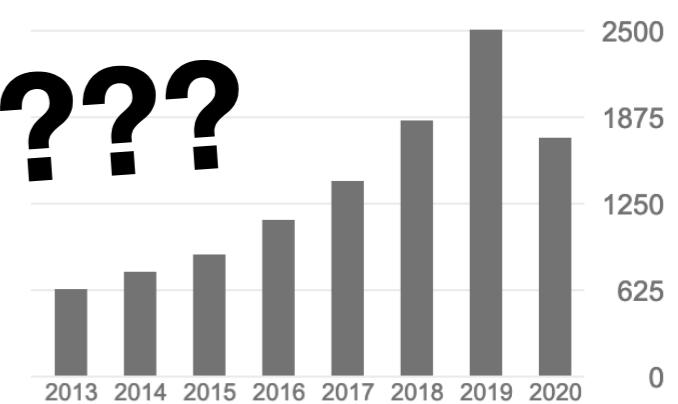
**Cloud-scale RNA-sequencing differential expression analysis with Myrna**

B Langmead, KD Hansen, JT Leek  
Genome biology 11 (8), R83

364

2010

???



# Notability

“He seems to be notable, based on the very highly cited papers shown in Google Scholar. Seeing them, after I had determined that the book was self-published, I was rather skeptical about the contents.”

In the blog post I wrote about this assignment, I characterized these comments as “charitable” which I think is some indication of the level of vitriol my students were getting on all the other articles...

...  
most highly cited are papers about biostatistics. The most [sic] highly cited is a rather pedagogical review article, & such articles have extremely high & uncharacteristic citation counts. The next three however are research papers, and are highly cited because they are the foundations for very important & widespread biological work.”

# Notability

She played a small role as a student on a widely used statistical package, sva, has high Google Scholar citations for that one thing, and won a student award. But she hasn't become an academic, and that's not enough for academic notability by itself, so we need to look for general notability instead for her post-academic career. The only evidence we have of this is this interview, which is in-depth, independent, and reliably published. But it's only one source and we need multiple sources like that.



## Hilary Parker

[FOLLOW](#)

Ph.D., Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health  
Verified email at jhsph.edu - [Homepage](#)

Biostatistics Genomics Batch Effects

### Cited by

All Since 2015

	All	Since 2015
Citations	1908	1793
h-index	8	7
i10-index	7	7

TITLE

CITED BY

YEAR

[The sva package for removing batch effects and other unwanted variation in high-throughput experiments](#)

JT Leek, WE Johnson, HS Parker, AE Jaffe, JD Storey  
Bioinformatics 28 (6), 882-883

1568

2012

[sva: Surrogate variable analysis](#)

JT Leek, WE Johnson, HS Parker, EJ Fertig, AE Jaffe, JD Storey, Y Zhang, ...  
R package version 3 (0), 882-883

180

2017

[Removing batch effects for prediction problems with frozen surrogate variable analysis](#)

HS Parker, HC Bravo, JT Leek  
PeerJ 2, e561

39

[Preserving biological heterogeneity with a permuted surrogate variable analysis for genomics batch correction](#)

HS Parker, JT Leek, AV Favorov, M Considine, X Xia, S Chavan, ...  
Bioinformatics 30 (19), 2757-2763

36

[The practical effect of batch on genomic prediction](#)

HS Parker, JT Leek  
Statistical applications in genetics and molecular biology 11 (3)

27

[sva: Surrogate Variable Analysis R package version 3.10. 0](#)

JT Leek, WE Johnson, HS Parker, AE Jaffe, JD Storey  
DOI 10. B9

24

[Sva: surrogate variable analysis. 2015](#)

JT Leek, WE Johnson, HS Parker, EJ Fertig, AE Jaffe, JD Storey, Y Zhang, ...  
R package version 3 (0), 25-27

15

[Surrogate variable analysis](#)

JT Leek

8

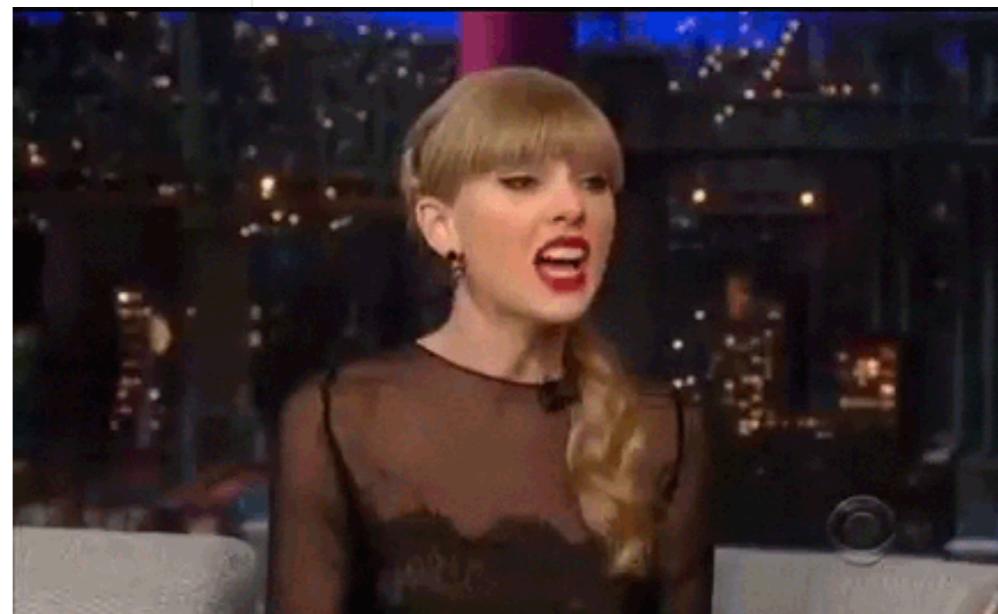
2007

[Opinionated analysis development](#)

H Parker  
PeerJ Preprints 5, e3210v1

6

2017



# I could go on and on

Many articles required intervention (either by me or others) to make it to the mainspace

**Spring 2018:**

Amanda Cox

Jenny Bryan

Kim Rees

Lena Groeger

Lynn Cherny

Meredith Broussard

**Still needs work:**

Julia Silge

**Spring 2020:**

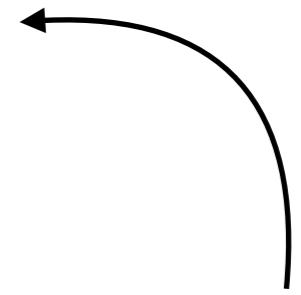
Nadieh Bremer

**Still need work:**

Hilary Parker

Irene Ros

This list is shorter.  
Maybe that's good?



# Wikipedia:No amount of editing can overcome a lack of notability

When articles are listed for deletion on the grounds of the topic's notability, the creators of such articles often ask how they could write better articles about that topic. But they are, in fact, asking the wrong question. There are many notability guidelines for different types of articles, but when a notability issue is invoked, no matter what the topic is, it always boils down to this question: *should we even have an article about that topic?*

This means that unless information is added to an article to show that its topic meets the relevant notability guideline, or unless the notability issue was invoked in error, there is nothing that can be done to save the article. Not a better writing style. Not a more neutral wording. Not surrendering the redaction to another person to circumvent conflict of interest guidelines. Not the removal of material potentially regarded as promotional. Not a more explicit referencing from primary sources, press releases, or interviews. Not even a promise that, soon, the subject will meet the notability guidelines. Nothing. None of these things address the problem. The problem is not with the article itself: the problem is what the article is about.

Editors who protest against deletion nominations of articles they create are often closely related to the subject. Our conflict of interest guidelines do not prevent anyone from creating or editing articles about themselves, but the very act of creating an article often arises from such a relationship, and the creator often overestimates the notability of the subject. However, people who create articles about themselves or projects they are involved in can do something when their articles are deleted on notability grounds. But it can only be done off-wiki. They can get others to notice them or their projects *first*.

## **Wikipedia:No amount of editing can overcome a lack of notability**



the grounds of the topic's notability, the creators  
should write better articles about that topic. But they  
There are many notability guidelines for different  
issue is invoked, no matter what the topic is, it  
should we even have an article about that topic?  
added to an article to show that its topic meets  
less the notability issue was invoked in error, there  
article. Not a better writing style. Not a more  
redaction to another person to circumvent conflict  
of material potentially regarded as promotional.  
many sources, press releases, or interviews. Not

even a promise that, soon, the subject will meet the notability guidelines. Nothing. None of these things address the problem. **The problem is not with the article itself: the problem is what the article is about.**

Editors who protest against deletion nominations of articles they create are often closely related to the subject. Our conflict of interest guidelines do not prevent anyone from creating or editing articles about themselves, but the very act of creating an article often arises from such a relationship, and the creator often overestimates the notability of the subject. However, people who create articles about themselves or projects they are involved in can do something when their articles are deleted on notability grounds. *But it can only be done off-wiki.* They can get others to notice them or their projects first.

# **A short list of things that don't confer notability**

- Giving keynote talks at major conferences
- Having many listeners to your podcast
- Being a journalist who has worked on front-page stories

# **One thing that really does confer notability**

- Being an ASA fellow

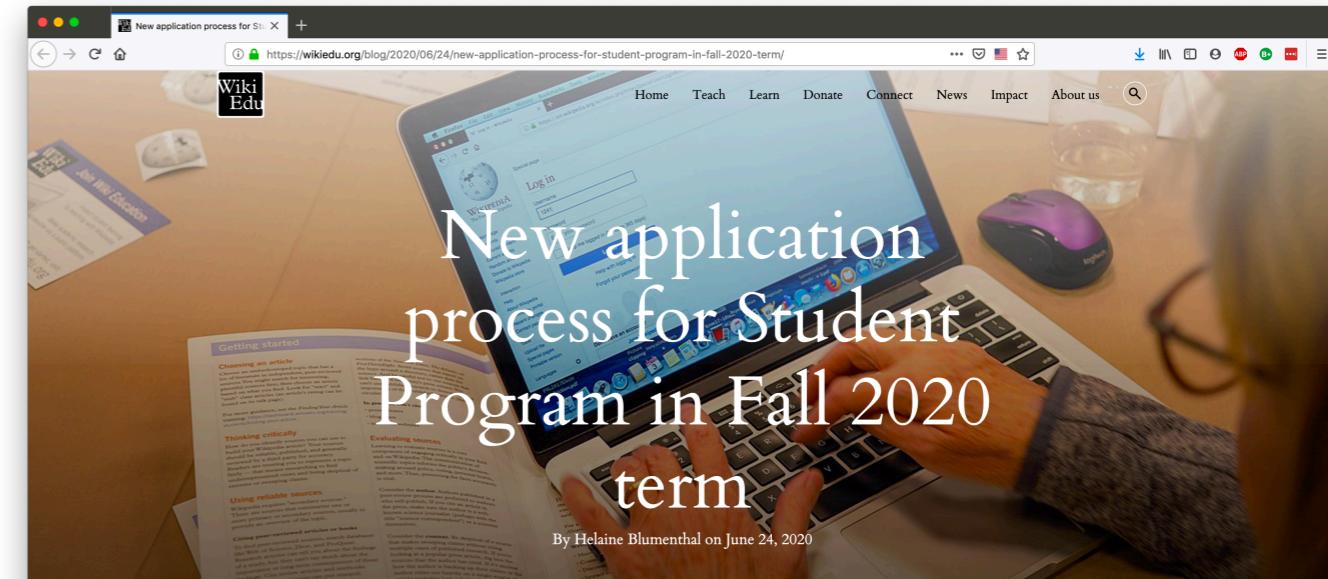
# So what?

If you want to do this assignment

Practice by writing some articles yourself. It's harder than it looks!

Prep by generating a list of ideas of people. It's harder than it looks!

Sign up to be part of the Wiki Education project (deadline has passed for fall semester, but they may accept on a rolling basis)



New application  
process for Student  
Program in Fall 2020  
term

By Helaine Blumenthal on June 24, 2020

As higher education plans for a potential virtual term in Fall 2020 in the United States and Canada, Wiki Education is expecting to have an extremely high demand for our Student Program, where we support faculty who assign Wikipedia editing as a class assignment.

While we would love to support everyone who's interested in teaching in our program, the financial realities of being a small, grant-funded nonprofit means we have [limited staff capacity](#) to support classes. Therefore, for the first time ever, we are instituting an [application process for Fall 2020](#). If you are interested in teaching with Wikipedia during the Fall 2020 term, you must submit your course page by July 26, 2020.

We anticipate that we will be able to support the vast majority of courses that apply and are instituting these new procedures to ensure that we can provide meaningful support to all of our students and instructors.

Who does this apply to?

Any instructor who wishes to teach with Wikipedia through Wiki Education's Student Program in fall 2020 must apply to participate. This application process applies to both returning instructors who have taught with our program in the past, and those new to the program.

# So what?

If you want to increase the representation of stats folks on Wikipedia

- Nominate people as ASA fellows!
- Work to get journalistic coverage of things your colleagues do
- Watch pages of people, and jump into “Articles for deletion” discussions

**Thank you**