

Edit Wars

When Stars and Condoms Collide

Lilly, Joe, Adriana, Mike, Stephanie

April 10, 2018

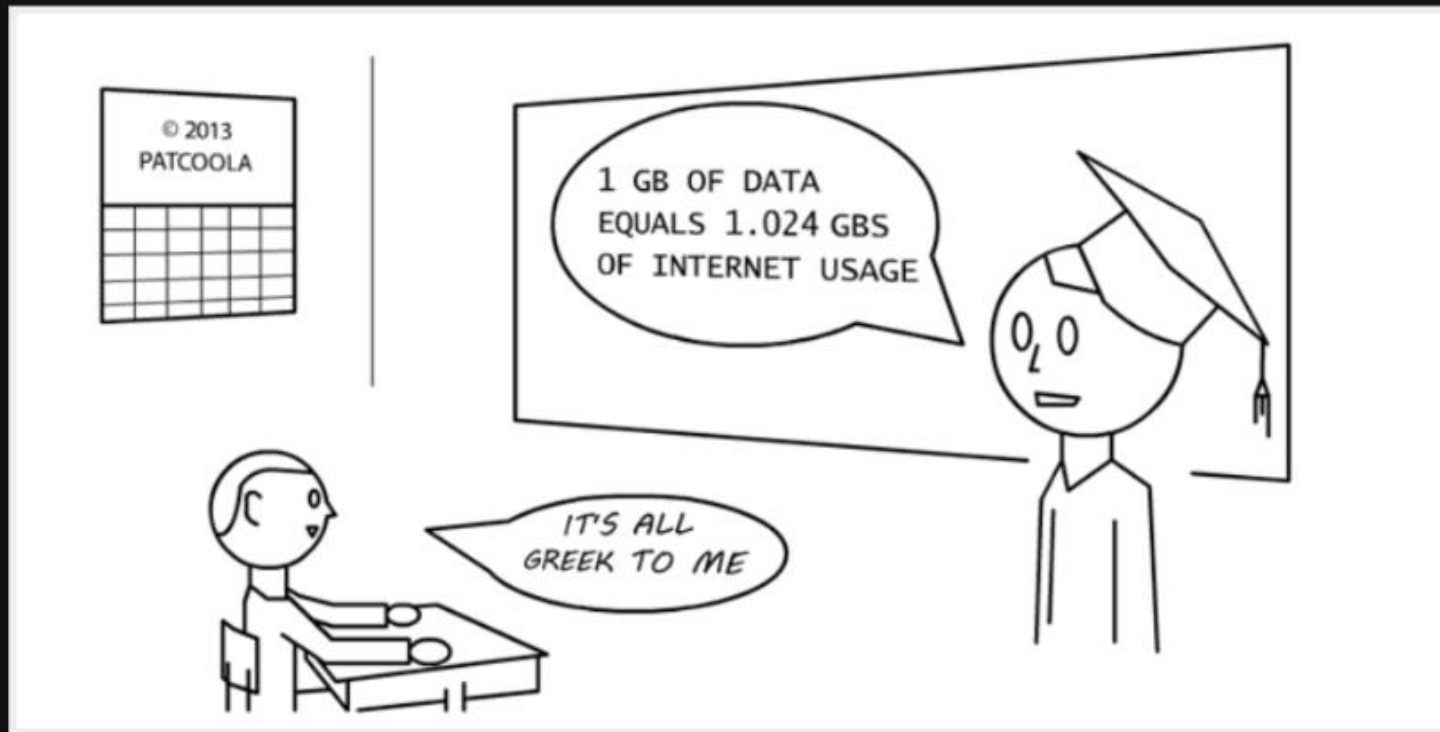


Introduction



Original Data

- Complete edit history of all Wikipedia (2001-2008).
- Original zip file > 18 GB, unzipped raw data file > 291 GB.



Our project

Using Wikipedia edit history, our interest was:

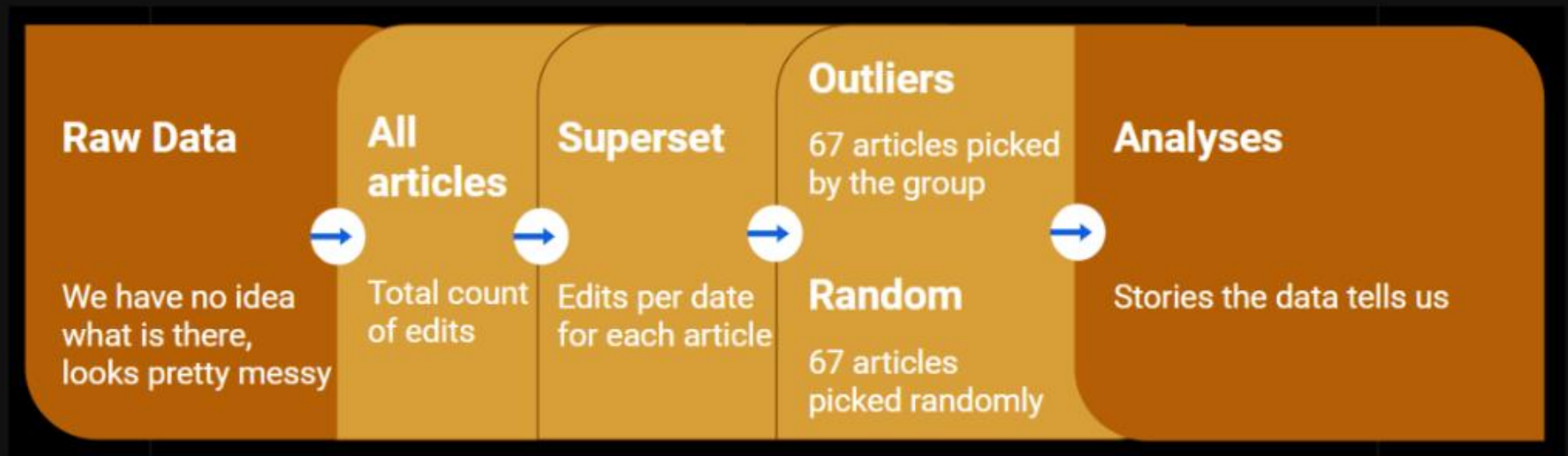
- Edit Wars?
- How we found them?
- A closer look into the articles for Barnard's Star and Condoms.



What are “Edit Wars”, and how to find them?



Process Flow



Raw Data - Munging

```
cat(raw.txt)
```

```
## REVISION 12 122974 Anarchism 2002-07-23T01:12:22Z Brion_VIBBER 51
## CATEGORY IMAGE MAIN Syndicalism Nihilism Gustave_de_Molinari Benjamin_Tucker
## Benjamin_Tucker Noam_Chomsky Peter_Kropotkin Leo_Tolstoy Leo_Tolstoy
## Classical_liberalism Individualist_anarchism Individualist_anarchism
## Individualist_anarchism Individualist_anarchism Libertarianism Libertarianism
## Anomie Anomie Mutual_aid Mutual_aid General_strike Atheism Free_Spirit
## Libertarian_socialism Libertarian_socialism Libertarian_socialism
## Lysander_Spooner Capitalism Mikhail_Bakunin Self-defense Political Minarchism
## Punk_rock William_McKinley Anarcho-syndicalists Karl_Marx Anarchism/ToDo
## Coercion John_Locke Anarcho-capitalism Anarcho-capitalism Anarcho-capitalism
## Anarcho-capitalism Regicide Government Michael_Bakunin Property_damage
## Max_Stirner Rudolph_Rocker Peter_Kropotkin Non-violent_resistance Proudhon
## David_Friedman Voltairine_de_Cleyre TALK USER USER_TALK OTHER Pl:Anarchizm
## Fr:Anarchisme Eo:Anarkismo EXTERNAL TEMPLATE COMMENT lang links: +fr, pl
## MINOR 1 TEXTDATA 1169
```



First look at the structured data

```
cat(output.txt)
```

```
## REVISION 12 18201 Anarchism 2002-02-25T15:00:22Z ip:Conversion_script
## p:Conversion_cript
## CATEGORY
## IMAGE
## MAIN Syndicalism Nihilism Gustave_de_Molinari Benjamin_Tucker Benjamin_Tucker
## inarchism Noam_Chomsky Punk_rock Fascism Anarchism/ToDo Coercion John_Locke
## eo_Tolstoy Anarcho-capitalism Anarcho-capitalism Anarcho-capitalism
## TALK Anarchy Anarchism
## USER
## USER_TALK
## OTHER
## EXTERNAL http://dwardmac.pitzer.edu/anarchist\_archives/kropotkin.html
## http://recollectionbooks.com/bleed/sinners/RockerRudolf.htm
## http://www.wikipedia.com/wiki.cgi?action=history&id=Anarchy
## TEMPLATE
## COMMENT Automated conversion
## MINOR 1
## TEXTDATA 1214
```



What does 291 GBs of data give us?

- 2,953,385 Unique Articles
- 116,590,855 Edits



All articles: *What does it look like?*

```
pandoc.table(head(initial_output_example), justify = 'right')
```

```
##
## -----
##                               Name    Count
## -----
##                               !!!      302
##
##                !!!Fuck_You!!!      88
##
##    !!!Fuck_You!!!_And_Then_Some      1
##
##    !!!Fuck_You!!!_and_Then_Some      81
##
##                !!!_(album)      71
##
##                               !!      41
## -----
```

As you can see, looks pretty ugly



Top 10 overall edited

```
pandoc.table(top_most_edited, justify = 'right')
```

```
##
## -----
##              Name      Count
## -----
##      George_W._Bush    38948
##
##      Wikipedia         26740
##
##      United_States     22467
##
##      World_Wrestling_Entertainment_roster    20838
##
##              Wii       18984
##
##              Jesus     18818
##
##      Deaths_in_2007    18124
##
##      World_War_II       17906
##
##      Adolf_Hitler       17180
##
```



Superset: Making sense of the data!

Getting a list of articles edited by date:

```
pandoc.table(head(top1000_revisions_in_a_day[,1:3], 6), justify = 'right')
```

```
##
## -----
##      Date              Name      Count
## -----
## 12/29/2001      Deaths_in_2003      55
##
## 12/30/2001      Deaths_in_2003      54
##
##   6/2/2002      Han_Dynasty      75
##
## 10/2/2002      Anarchism      79
##
## 10/20/2002  Christopher_Columbus      47
##
## 10/24/2002  Politics_of_Canada      52
## -----
```



Datasets

- *Superset*
 - A list of every article's date and number of edits for that date.
- *Outliers*
 - The mean number of edits per day for each article (top 5%).
- *Random*
 - Excluding the previous 67 articles, we randomly picked another 67 articles.



Data Structure and fields we used

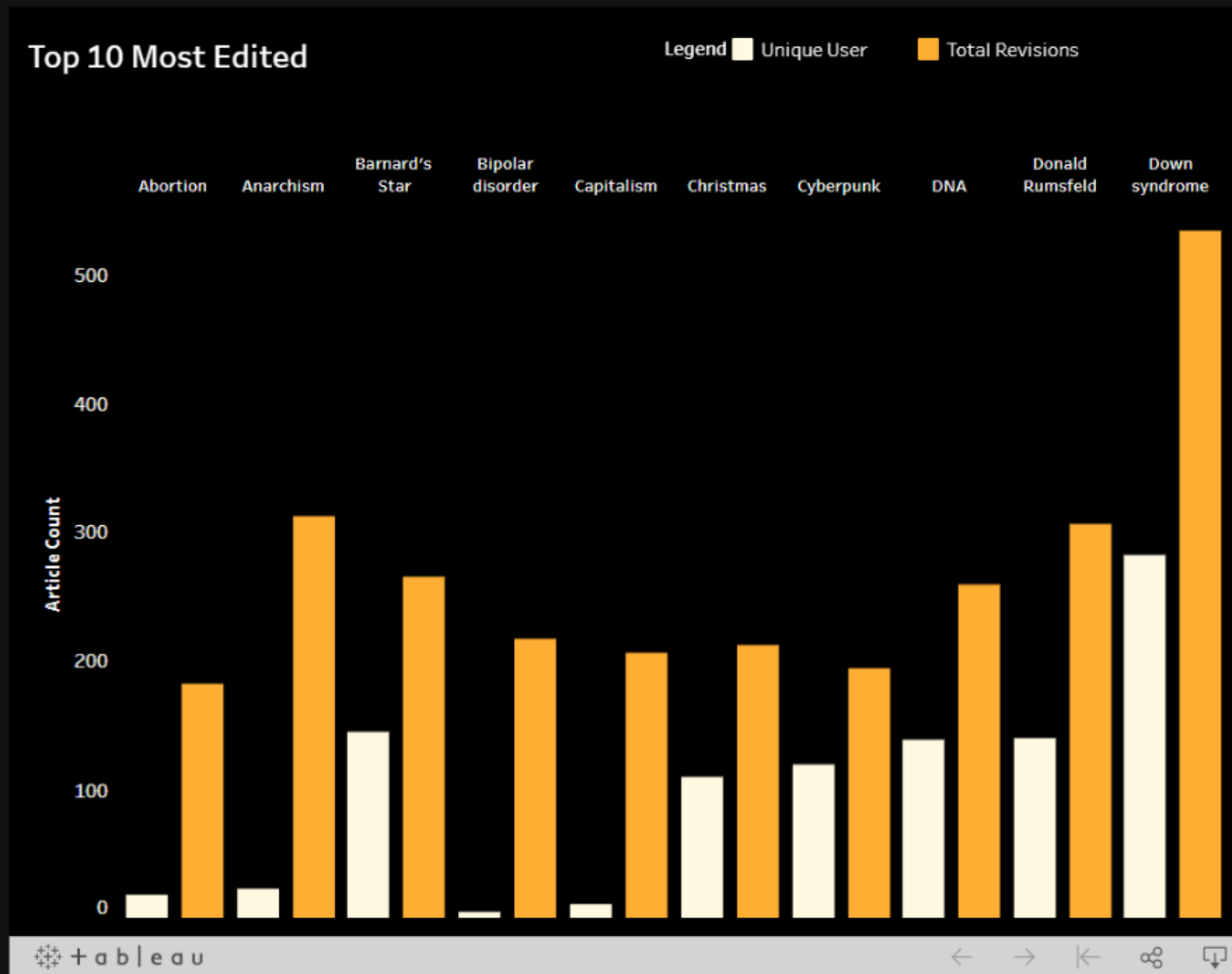
- REVISION ***
- CATEGORY
- IMAGE
- MAIN
- EXTERNAL
- TEMPLATE
- COMMENT ***
- MINOR
- TEXTDATA



**We finally have a manageable
dataset! Now what?**



Top 10 Edits:



Bickering and Vandalism

- “Edit-War” Flag: $\frac{\# \text{unique users}}{\# \text{comments}}$
- Troll Flag: Count of the vandalism tags per article



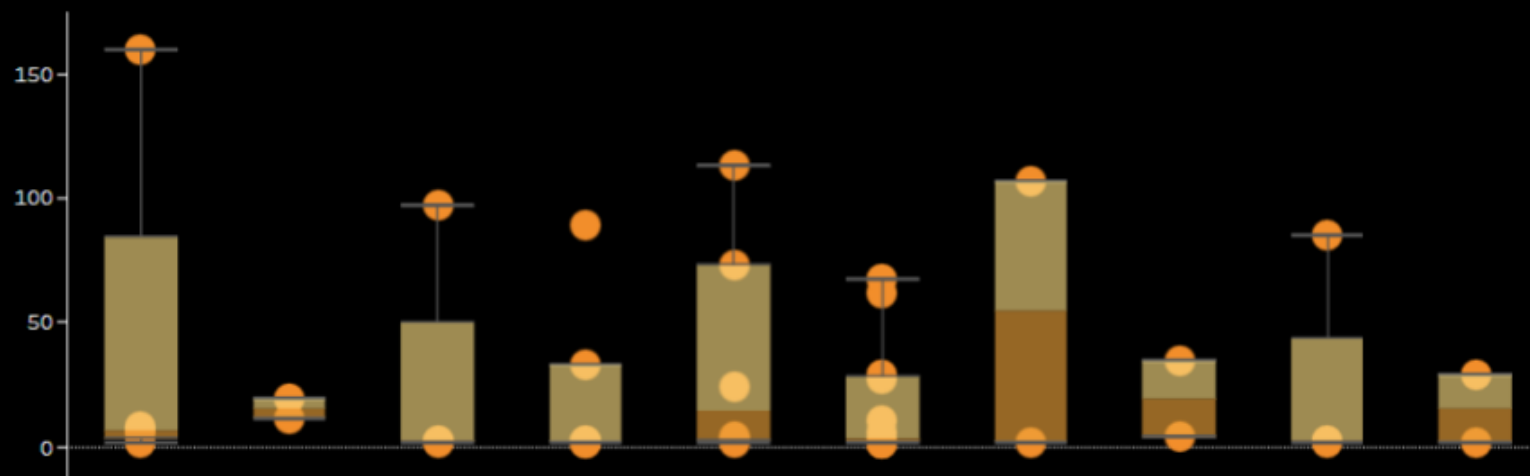
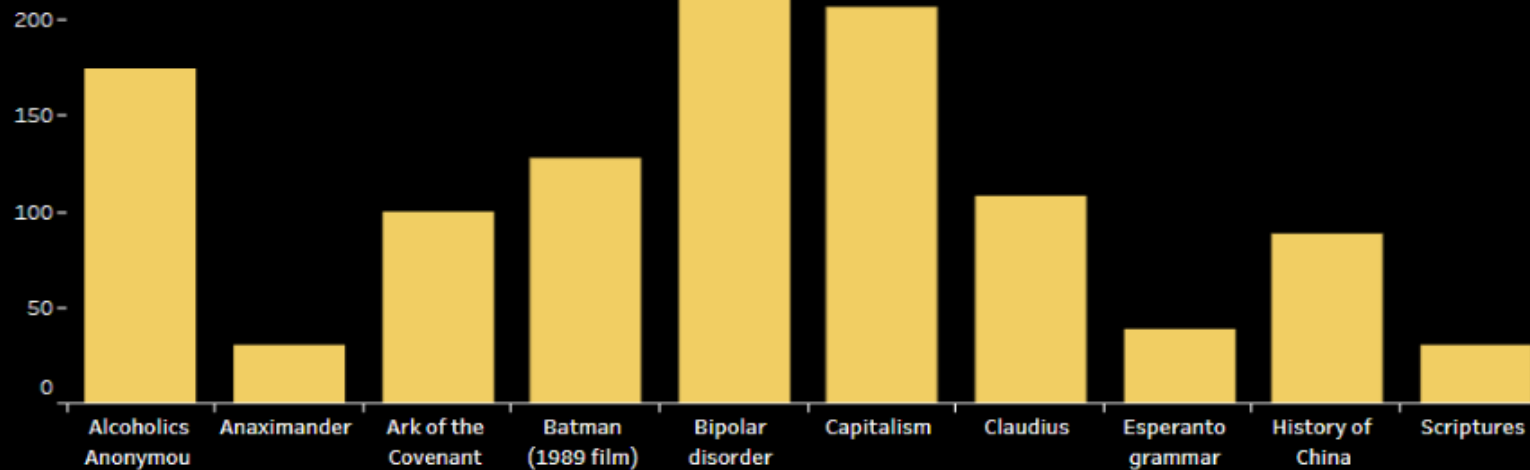
Top 5 articles with the most vandalism tags

Random		Outliers	
Astronomical Unit	4	Barnard's Star	29
Antartic Treaty System	3	DNA	12
Inca Empire	3	Blitzkrieg	10
Big Bang	2	Comet Shoemaker-Levy 9	6
December 10	2	Autism	5

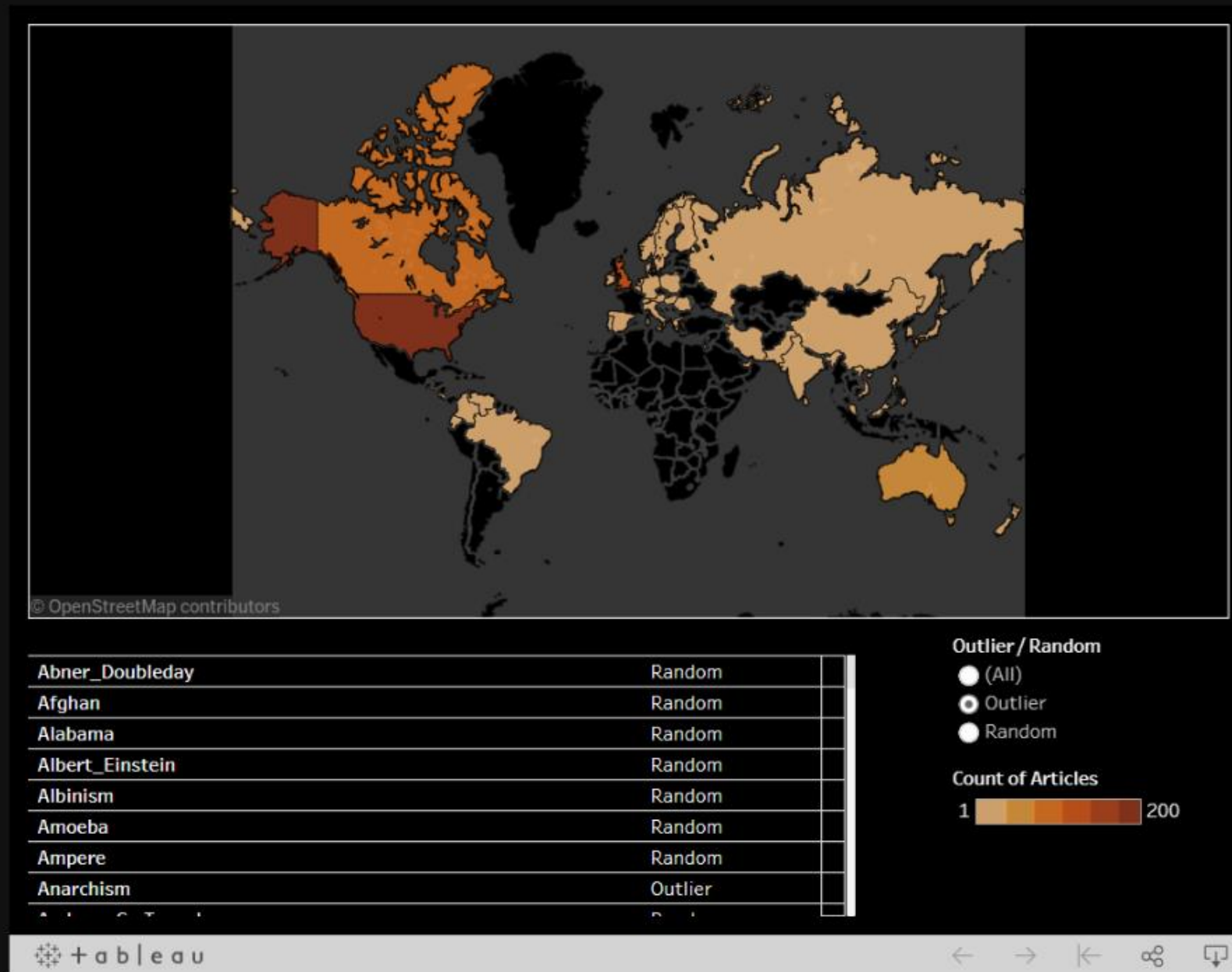


Top 10 Edit Wars

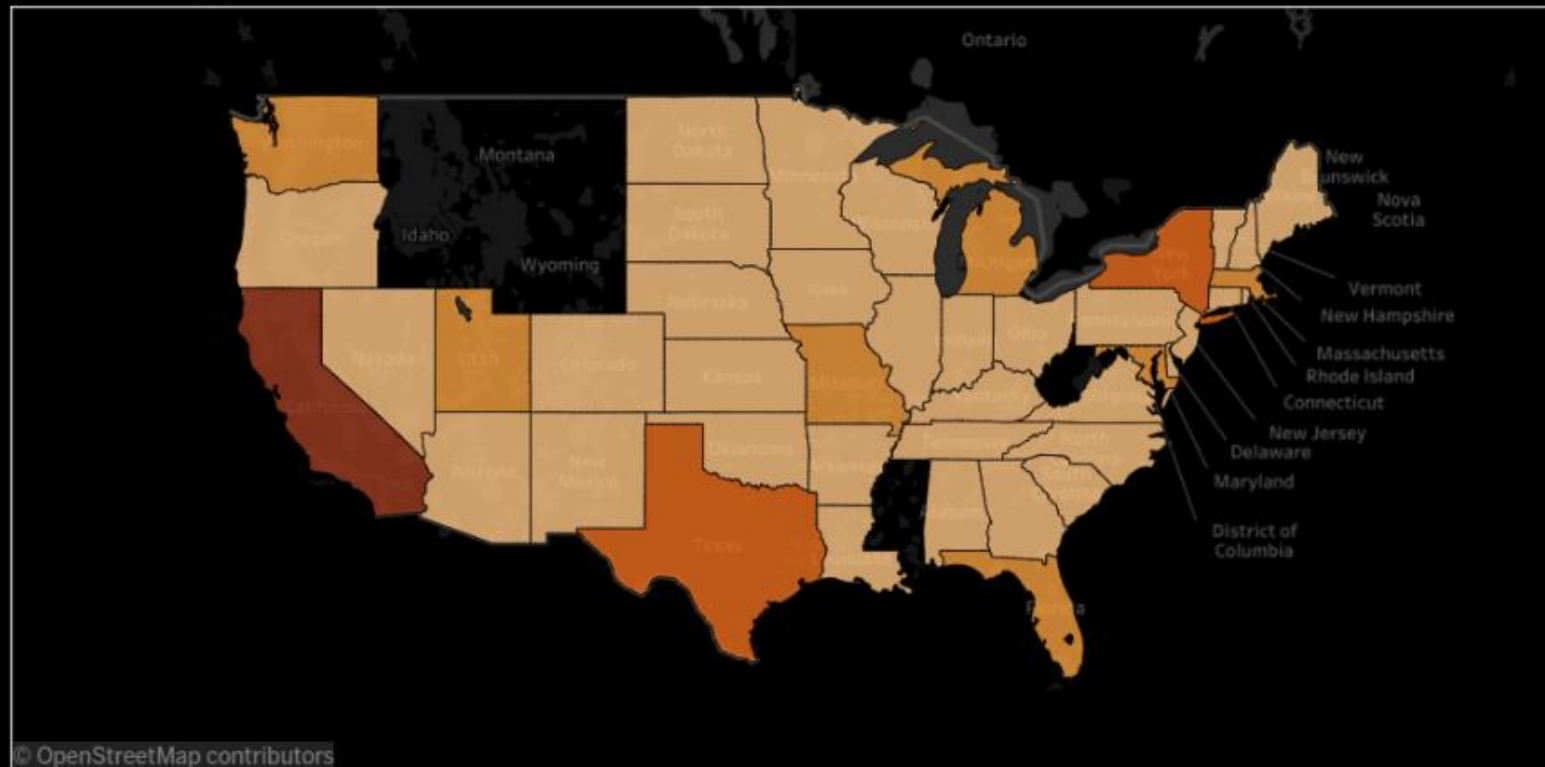
EDIT WAR FLAG : 10 lowest scores
Article total revision count v. unique contributor revision count



Where do anonymous users edit?



Where do anonymous users edit in the US?



Abner_Doubleday	Random	
Afghan	Random	
Alabama	Random	
Albert_Einstein	Random	
Albinism	Random	
Ampere	Random	
Anarchism	Outlier	
Argument from evolution	Random	

Outlier / Random

- (All)
- Outlier
- Random

Count of Articles

1 100



Our goals

- Create a methodology
- Find something interesting
- Tell a story

We found several good stories to tell, we brought you two.



Barnard's Star



Story Statistics

- Dataset: Outlier
- Article: Barnard's Star
- Date: 2007-10-18
- UniqueUser: 145
- TotalRevisions: 266
- Ratio: 0.5451
- Vandalims Flag: 29 (most in our dataset)



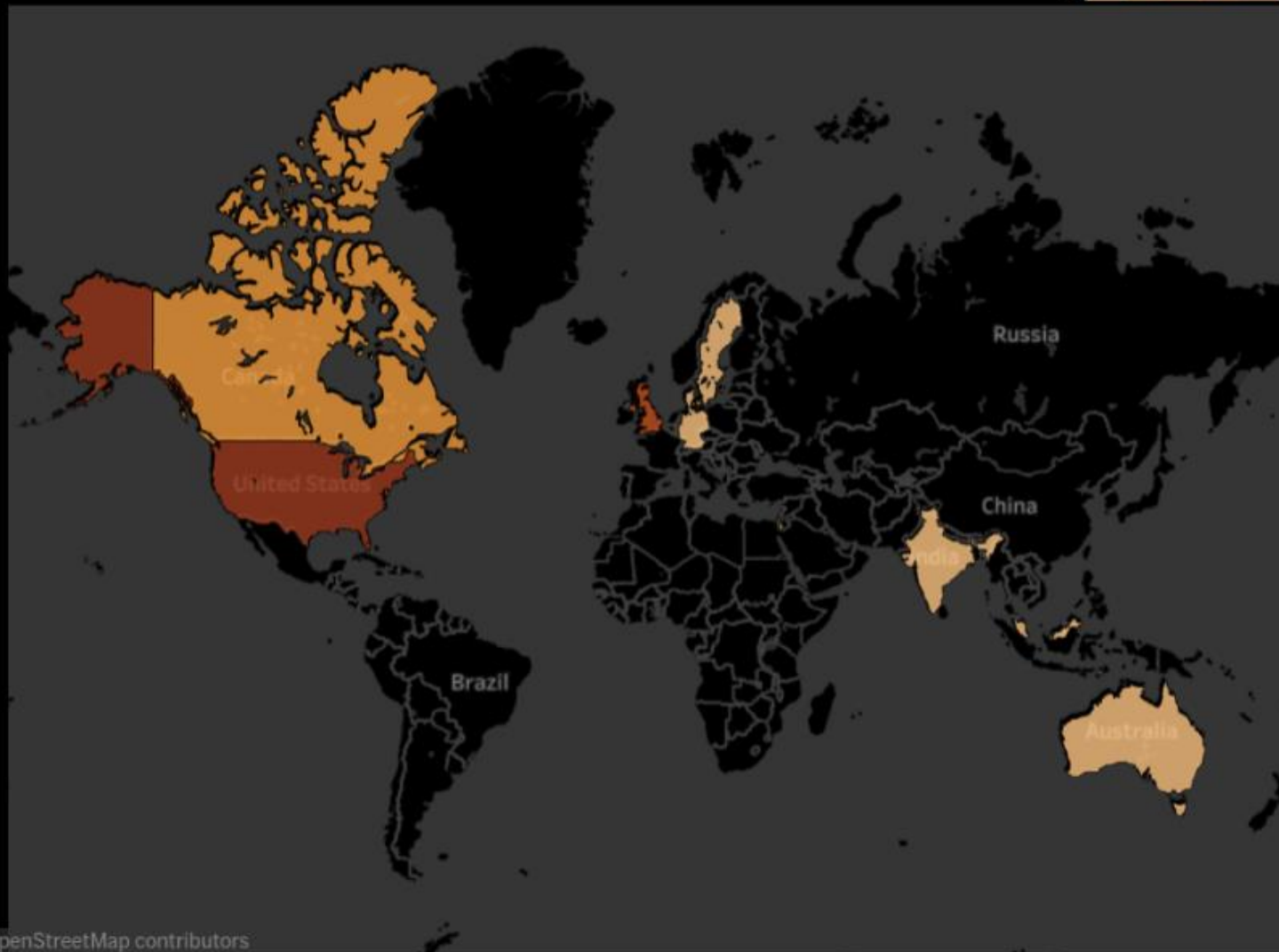
Wordcloud



Barnard's Star - Anonymous User locations

Revision Count

1 25



© OpenStreetMap contributors

tableau



Chat

Timestamp	User	Comment
02:12:59	Angilbas	<i>/* System summary */</i> Brightness of moon and planets if Barnard's Star were to replace Sun
02:16:59	ip:68.126.204.102	[[WP:AES
02:17:10	ClueBot	Reverting possible vandalism by [[Special:Contributions/68.126.204.102]]. If this is a mistake
02:21:23	ip:68.126.204.102	
02:22:07	Slakr	Reverted 1 edit by [[Special:Contributions/68.126.204.102
02:40:21	ip:69.114.14.88	<i>/* System summary */</i>
02:41:32	ip:69.114.14.88	<i>/* System summary */</i>
02:47:14	ip:71.72.216.86	
02:48:04	Assblazer	Take THAT motherfucker.
02:48:39	VoABot_II	BOT - Reverted edits by [[Special:contributions/Assblazer
02:49:36	Lorkki	Revert vandalism by 71.72.216.86
02:52:42	Analblazer	
02:53:45	Analblazer	
02:54:35	SandyGeorgia	[[WP:UNDO
02:55:19	ip:70.15.116.59	Quit the vandalism already



What triggered the mass edits?

Barnard's Star

Barnard's Star was the wikipedia "article of the day" when it saw its spike in edits. We attribute the spike in edits to be due to the people "trolling" this article due to its unexpected exposure as being selected as the article of the day on October 18, 2007.

Featured Article



Barnard's Star is a [featured article](#); it (or a previous version of it) has been [identified](#) as one of the best articles produced by the [Wikipedia community](#). Even so, if you can update or improve it, [please do so](#).



This article appeared on Wikipedia's Main Page as [Today's featured article](#) on October 18, 2007.

Article milestones

Date	Process	Result
October 11, 2006	Peer review	Reviewed
October 21, 2006	Featured article candidate	Promoted

Current status: **Featured article**



Condom



Story Statistics

- Dataset: Outlier
- Article: Condom
- Date: 2005-12-26
- UniqueUser: 17
- TotalRevisions: 94
- Ratio: 0.1809



Wordcloud



Chat

Timestamp	User	Comment
T00:13:42Z	SarekOfVulcan	rv - hope you don't actually think sex and surgery are related beyond alphabetically
T00:18:39Z	Chooserr	rvt...I'll list the reasons on the talk if you can wait a few seconds :)
T00:21:43Z	ip:81.102.251.15	/* Anti-Condom Trends */
T00:28:28Z	SarekOfVulcan	/* Health issues */ remove talc ref for invalid comparison between methods of entering abdominal cavity
T00:32:35Z	Chooserr	rvt...I don't really understand where you are coming from. It's used on the lubricant of the condom - when the condom is inside it (the lubricant) can rub off or do what ever it does...use talk pls
T02:13:02Z	Chooserr	/* Health issues */ There is a difference between N-nitrosamine and normal nitrosamines - Or they wouldn't have seperate names right? :D
T02:18:01Z	SarekOfVulcan	restored nitrosamine link
T03:43:10Z	Ytrewqt	RV! there are holes in that edit summary I could fly a jet through
T03:51:33Z	Ytrewqt	rv, I don't see ANY evidence that the origional charges are false, smark, smark, smark, keep smarking and some day it will come back and bite you in the ass
T03:52:26Z	FCYTravis	rv, perfectly well-sourced rebuttal.
T05:27:53Z	Rose_Garden	Rv. Restore sourced material.
T05:28:46Z	Ytrewqt	rv, I seem to have picked up a wiki stalker, that didn't take long



Chat

T08:42:42	Locke_Cole	<i>/* Proper use */ restoring proper use section</i>
T08:46:10	Chooserr	this was already voted on, bring it up on the talk page before making edits like tht
T08:57:40	Chooserr	rvt it was agreed on
T10:35:47Z	FCYTravis	<i>/* Condoms & the Environment */ Stupid. You might as well add such a section to [[McDonalds]] or [[Hamburger]] or [[Soda]].</i>
T16:36:26Z	Locke_Cole	rvt - rewrite was better
T16:46:57Z	ip:63.22.0.39	rvt - censorship
T16:50:51Z	FreplySpang	rv: Agree that FCYTravis' rewrite was better, and no, it's not censorship. Discuss on talk if you disagree.
T16:59:36Z	ip:63.22.0.39	rvt - and since you're using puppets to avoid [[WP:3RR
T17:18:46Z	ip:63.22.0.39	restore {{sprotect}} template
T17:29:11Z	FreplySpang	remove misused sprotect template; revert to FCYTravis' version as discussed on talk
T19:12:37Z	Chooserr	<i>/* Environmental impact */ put my version back - has specific sources - but left the bit about Polyethylene condoms and their landfill mass being negligible because that is important</i>
T20:40:12Z	FCYTravis	rm sensationalism "holy god no condoms are damaging the environment OH NOES!!!!111!!eleven!1!1!" Bin it.
T21:04:54Z	ip:63.22.55.169	rvt to last consensus version, abusive edit summaries won't get you anywhere
T21:11:40Z	FCYTravis	There's hardly a consensus for that version.
T21:18:13Z	ip:63.22.55.169	rv to last (pre-troll)version
T21:25:35Z	La_goutte_de_pluie	rv censorship



What triggered the mass edits?

Condom

On December 25, 2005, the Catholic Church released a statement stating that “Safe sex” in regards to sexual activity using condoms is not actually safe. The condom article saw its spike in edits on December 26, 2005 - one day after this released statement.

Statement



Conclusion



Questions?



Thank you!

