

What Motivates Users to Contribute to Wiki Travel Content Online?
A Study of User Behavior on WikiVoyage and Wikipedia

Liz Legg
June 2015
UW CommLead
Community Data Science

What's the Problem?

As the world becomes increasingly accessible, more people are choosing to plan and book international trips on their own, without the help of travel agencies or tour providers. This affords people the opportunity to save money and experience a greater sense of adventure by traveling independently. However, it also puts a greater burden on the individual traveler to make decisions about how to spend her time and money in destinations she has little or no experience with. The decisions a traveler makes before getting on an airplane will influence his experiences while abroad. Thus, it's important to understand how accurate and up-to-date the resources that travelers are using to make these decisions are.

Travelers have a plethora of options when it comes to researching and planning a trip. We can read travel guides from big name publishers like Lonely Planet or Rough Guides, we can talk to friends and family who have previously traveled to places we're planning to go, and we can turn to the Internet for recommendations from Trip Advisor and other travel forums. Unfortunately, each of these solutions has inherent problems. Physical guidebooks quickly become out-of-date and people reading them flock to the same sights. Recommendations from friends and family may not be specific enough to allow us to create a complete itinerary. The Internet is filled with recommendations from strangers that, while often useful, don't always reflect our own values and travel style.

Any savvy traveler will likely utilize a combination of all three resources. However, according to a 2012 Google study¹, most travelers research upcoming trips online, which means that people are increasingly likely to rely on the Internet to make their travel decisions even though they know the least about how the content they find there is sourced and updated.

This study looks at online destination content revision data from Wikipedia and WikiVoyage.org – an online travel wiki that allows users to revise and create travel-related content for any destination – in an attempt to better understand how, when, and why people update travel content online.

¹ https://ssl.gstatic.com/think/docs/the-2012-traveler_research-studies.pdf

My Hypotheses

I hypothesize that the more established, popular, and newsworthy a destination is, the more likely it is that people will update content about it. I also theorize that a lack of content for a newly popular destination will influence people to update the content, as will a major political or natural event in the region.

More specifically, when looking at total revisions, number of unique edits/users, revision size, and article length, I speculate that the following will be true:

- Destinations **growing** (vs. declining) in tourism will have...
 - more total edits
 - more unique edits/users
 - more total content
- Destinations that are **visited most** (vs. least) will have...
 - more total edits
 - more unique edits/users
 - more total content
- Destinations that are **newly** opened to tourism (vs. those that have been visited for decades) will have...
 - fewer total edits
 - fewer unique edits/users
 - larger revision sizes
- Destinations that have experienced a significant, **negative event** will have...
 - more total edits around the time of the event
 - more unique edits/users around the time of the event

Methods and Approach

I first identified a list of destinations that fall into the different travel classifications I wanted to compare: destinations that are growing in tourism vs. declining in tourism, very popular vs. least popular, new vs. established, and destinations that have had significant political unrest, terrorist attacks, or a major natural disaster within the past 5.5 years.

I used the following criteria to select each destination...

Decline vs. Growth

In order to compare destinations that are growing vs. declining in tourism, I looked at MasterCard's Global Destination Cities Index², which lists the top 20 cities by visitor spend and their corresponding international overnight visitors on a yearly basis.

TABLE 1 Global Top 20 Top Destination Cities by International Overnight Visitors (2014)

| | Destination City | Country | Visitors (millions) | | | | | | 2014 Visitor Spend (US\$ bn) |
|----|------------------|----------------|---------------------|-------|-------|-------|--------------|----------------|------------------------------|
| | | | 2010 | 2011 | 2012 | 2013 | 2014 | %Δ 2013 & 2014 | |
| 1 | London | United Kingdom | 14.71 | 15.29 | 15.46 | 17.30 | 18.69 | 8.0% | \$19.3 |
| 2 | Bangkok | Thailand | 10.44 | 13.80 | 15.82 | 18.46 | 16.42 | -11.0% | \$13.0 |
| 3 | Paris | France | 13.27 | 13.88 | 14.33 | 15.29 | 15.57 | 1.8% | \$17.0 |
| 4 | Singapore | Singapore | 8.80 | 10.14 | 11.11 | 12.10 | 12.47 | 3.1% | \$14.3 |
| 5 | Dubai | UAE | 8.41 | 9.20 | 10.16 | 11.12 | 11.95 | 7.5% | \$10.9 |
| 6 | New York | USA | 9.43 | 10.27 | 10.60 | 11.08 | 11.81 | 6.6% | \$18.6 |
| 7 | Istanbul | Turkey | 6.45 | 7.51 | 8.82 | 9.87 | 11.60 | 17.5% | \$9.4 |
| 8 | Kuala Lumpur | Malaysia | 8.90 | 8.99 | 9.26 | 9.56 | 10.81 | 13.1% | \$8.1 |
| 9 | Hong Kong | China | 8.13 | 8.43 | 8.37 | 8.26 | 8.84 | 7.0% | \$8.3 |
| 10 | Seoul | South Korea | 6.06 | 6.56 | 7.51 | 8.24 | 8.63 | 4.7% | \$11.5 |
| 11 | Barcelona | Spain | 6.18 | 6.89 | 6.91 | 7.18 | 7.37 | 2.7% | \$11.2 |
| 12 | Amsterdam | Netherlands | 5.86 | 6.07 | 6.10 | 6.74 | 7.23 | 7.2% | \$4.4 |
| 13 | Milan | Italy | 5.83 | 6.59 | 6.88 | 6.85 | 6.82 | -0.4% | \$5.3 |
| 14 | Rome | Italy | 6.65 | 6.66 | 6.82 | 6.63 | 6.79 | 2.5% | \$5.6 |
| 15 | Taipei | Chinese Taipei | 3.52 | 3.96 | 4.70 | 5.80 | 6.29 | 8.4% | \$10.8 |
| 16 | Shanghai | China | 6.67 | 6.18 | 6.04 | 5.66 | 6.09 | 7.6% | \$5.3 |
| 17 | Vienna | Austria | 4.64 | 5.08 | 5.38 | 5.67 | 6.05 | 6.8% | \$5.6 |
| 18 | Riyadh | Saudi Arabia | 1.82 | 4.16 | 4.83 | 5.52 | 5.59 | 1.3% | \$4.1 |
| 19 | Tokyo | Japan | 4.47 | 2.94 | 4.07 | 5.05 | 5.38 | 6.5% | \$7.4 |
| 20 | Lima | Peru | 2.07 | 2.94 | 3.94 | 4.91 | 5.11 | 4.1% | \$1.8 |

² http://newsroom.mastercard.com/wp-content/uploads/2014/07/Mastercard_GDCI_2014_Letter_Final_70814.pdf

I selected three cities that have grown steadily in overnight international visitors year over year since 2010 to compose the “growth” classification.

Although I was unable to find any cities from the MasterCard Index that were steadily declining year over year, I selected five cities that saw at least one year of significant decline in international overnight visitors (Bangkok was down 11% from 2013 to 2014) or multiple years of less significant decline (Hong Kong was down .7% from 2011 to 2012 and another 1.3% from 2012 to 2013).

Most Popular vs. Least Popular

In order to find the seven destinations I selected for the “most popular” classification, I cross-referenced the MasterCard Index (discussed above) with Euromonitor International’s Top City Destinations Ranking³ from 2013 (the most recent data available). I intentionally selected destinations from multiple continents.

I was unable to find a reliable source for the least visited cities in the world, so I selected the largest cities from a list of least visited countries compiled by travel blogger Gunnar Garfors⁴. Garfors’ data range from 2007 to 2012 and are gathered from multiple sources, so there is a larger margin for error with these destinations.

New vs. Established

I chose two “new” travel destinations based largely on my personal knowledge.

In 2011 the Myanmar government relaxed travel restrictions for U.S. citizens and other international travelers and President Obama eased sanctions against the country⁵. Foreign tourists to the country more than doubled from 2012 (1 million) to 2013 (2.04 million)⁶.

Similarly, Cuba is beginning to see an increase in American tourists as relations between the two countries continue to improve. Although there are still restrictions on U.S. tourists entering Cuba – you must be visiting family, on an educational trip, or entering the country for some reason other than strictly tourism⁷ – Cuba has seen a 36% increase in American visitors from Jan. through early May 2015 as compared to the same time period in 2014. Visits by international travelers in general are up 14% from 2014⁸.

³ <http://blog.euromonitor.com/2015/01/top-100-city-destinations-ranking.html>

⁴ <http://www.garfors.com/2013/01/the-25-least-visited-countries-in-world.html>

⁵ http://www.nytimes.com/2012/08/05/travel/visiting-myanmar-its-complicated.html%3F_r=0

⁶ http://en.wikipedia.org/wiki/Tourism_in_Burma

⁷ http://wikitravel.org/en/Americans_in_Cuba

⁸ <http://www.usnews.com/news/world/articles/2015/06/02/thawing-of-us-cuba-relations-inspires-havana-tourism-dreams>

Again, although the data I found are at the country level, I chose the largest cities for Myanmar and Cuba – Yangon and Havana respectively – in order to remain consistent in my analysis.

I compare these “new” destinations to the previous list of “most visited” destinations, as those are also well-established tourist destinations with few or no travel restrictions for U.S. visitors.

Negative Events

For this classification I, again, largely relied on my personal knowledge of major political unrest, terrorist attacks, and natural disasters that have occurred over the past 5.5 years. I also found that two of the least visited destinations are well known for ongoing political unrest.

I chose six cities that have experienced political unrest or terrorism and four that have experienced a major natural disaster:

Unrest

- Cairo – Egyptian Revolution / Arab Spring (Jan–Feb 2011)⁹
- Nairobi – 147 killed at Garissa University College (April 2015)¹⁰
- Paris – Charlie Hebdo terrorist attack (Jan 2015)¹¹
- Kiev – Ukrainian revolution (Feb 2014)¹²
- Ashgabat, Turkmenistan – tightly controlled police state (ongoing)¹³
- Mogadishu, Somalia – decades of urban warfare/suicide bombings (ongoing)¹⁴

Natural Disaster

- Christchurch – 6.3 magnitude earthquake (February 2011)¹⁵
- Tokyo – Tohoku earthquake and tsunami (March 2011)¹⁶
- Kathmandu – 7.8 magnitude earthquake (April 2015)¹⁷
- New Orleans – Hurricane Isaac (August 2012)¹⁸

⁹ http://en.wikipedia.org/wiki/Egyptian_Revolution_of_2011

¹⁰ <http://www.cnn.com/2015/04/02/africa/kenya-university-attack/>

¹¹ <http://www.cnn.com/2015/01/08/europe/charlie-hebdo-attack-timeline/>

¹² http://en.wikipedia.org/wiki/2014_Ukrainian_revolution

¹³ <http://wikitravel.org/en/Turkmenistan>

¹⁴ <http://en.wikivoyage.org/wiki/Mogadishu>

¹⁵ http://en.wikipedia.org/wiki/2011_Christchurch_earthquake

¹⁶ http://en.wikipedia.org/wiki/2011_T%C5%8Dhoku_earthquake_and_tsunami

¹⁷ mailto:http://en.wikipedia.org/wiki/April_2015_Nepal_earthquake

¹⁸ [http://en.wikipedia.org/wiki/Hurricane_Isaac_\(2012\)](http://en.wikipedia.org/wiki/Hurricane_Isaac_(2012))

Ultimately, I settled on the following list of cities:

| City | Decline | Growth | Natural Disaster | Political Unrest / Terrorism | New | Most Visited | Least Visited |
|----------------|---------|--------|------------------|------------------------------|-----|--------------|---------------|
| Bangkok | x | | | | | x | |
| Istanbul | | x | | | | x | |
| London | | x | | | | x | |
| Hong Kong | x | | | | | x | |
| Shanghai | x | | | | | | |
| Kathmandu | | | x | | | | |
| New Orleans | | | x | | | | |
| Tokyo | x | | x | | | x | |
| New York City | | | | | | x | |
| Cairo | | | | x | | | |
| Nairobi | | | | x | | | |
| Paris | | x | | x | | x | |
| Havana | | | | | x | | |
| Yangon | | | | | x | | |
| Christchurch | | | x | | | | |
| Kiev | | | | x | | | |
| Milan | x | | | | | | |
| Ashgabat | | | | x | | | x |
| Yaren District | | | | | | | x |
| Thimphu | | | | | | | x |
| Mogadishu | | | | x | | | x |

Once I established which cities to study, I began collecting data.

Collecting the Data

Note: Please refer to the code repository that accompanies this document for the full Python files referenced below.

I utilized the MediaWiki API¹⁹ to pull revision data, including size, user, and date information for revisions made since January 1, 2010 to the WikiVoyage article pages for each destination in my study.

I quickly realized that many of the destination pages on WikiVoyage also have sub-category pages that I should include in my data set. For example, the New York City page has a sub-category page for Manhattan, which has a sub-category page for Central Park. It's vital to include data for these sub-category pages; otherwise much of the essential data for each destination would be excluded.

As such, my first step was to find all WikiVoyage sub-categories up to three levels deep, for each destination page (*see file: 1_destination_sub_categories.py*). My code creates two output files, one with a list of all the top-level city and sub-category titles (*see file: 1_all_subcat_page_titles.tsv*), which I read in to future files in order to get the revision code previously discussed, and another that counts the number of sub-categories for each city (*see file: 1_destination_sub_category.tsv*).

Once I had the titles for the main destination pages and their sub-categories, I called the MediaWiki API and requested the timestamp, size, and user revision rvprop parameters for data from January 1, 2010 through late May 2015 for each title (*see file: 2_revisions_for_all_subcats.py*).

The remainder of my code interprets this revision information in different ways. I parse the revision data to count the total number of edits for each destination (*see file: 3_total_edits_by_destination.py*) as well as the unique number of edits/users per day for each destination (*see file: 3_unique_edits_by_destination.py*). I found comparing edits by day by destination did not produce a large enough variance, so I instead counted the total edits by month for each destination (*see file: 4_edits_by_month_by_destination.py*) as well as unique edits/users per day by month for each destination (*see file: 4_unique_edits_by_month_by_destination.py*).

The "revision size" parameter in MediaWiki is a little misleading. It is not, as the name suggests, the size of the revision, but the size of the entire article after the revision. I use this parameter to find the article length after each revision for top-level destination pages (*see file: 6_article_length.py*) and determine the size of each revision by subtracting the previous revision size from the current revision size in Excel (*see file: 2_revisions_for_all_subcats.xlsx*).

¹⁹ http://www.mediawiki.org/wiki/API:Main_page

After parsing the WikiVoyage data by unique edits/users per day by month (*see file: 7_total_unique_edits_by_month.py*), I felt that, although I did have enough data to draw some conclusions, adding Wikipedia revision data for the same pages would provide additional destination data that would enhance my findings. Although the Wikipedia content is not travel-specific, much of it could be useful to travelers and I believe it's reasonable to assume that users would have similar motivations for updating the page content. My results do, however, differentiate the two sources in order to avoid confusion.

In the case of Wikipedia, I did not include sub-categories because I found that they quickly veered away from destination-specific information. For example, if I had included sub-categories, data on pages like “Barbie Shanghai,”²⁰ which discusses a Barbie retail store in the city, would have been included in the data set.

The code I used to pull the Wikipedia data is nearly identical to the WikiVoyage code. The file names correspond to the WikiVoyage files with the exception that they are preceded by “no-cats-wikipedia-“ in the file names.

Finally, I wrote additional code to combine the Wikipedia data with the WikiVoyage data. These files include:

- **article_length.py** – Finds the maximum article length for both Wikipedia and WikiVoyage. The corresponding Excel file compiles current article length and average revision size data as well.
- **combo_edits_by_month.py** – Looks at four other files that find total and unique edits by month for both Wikipedia and WikiVoyage and writes them into a single file.
- **total_edits_by_destination.py** – Writes the Wikipedia and WikiVoyage data into a single file for ease of use.
- **unique_edits_by_destination.py** – Writes the Wikipedia and WikiVoyage data into a single file for ease of use.

In many cases I included the classifications I outlined for the study (growth, decline, most visited, least visited, new, political unrest, natural disaster) in the code as well so that I could compare the data across those classifications. I then used the filter feature within Excel to easily total and average the data for each comparison.

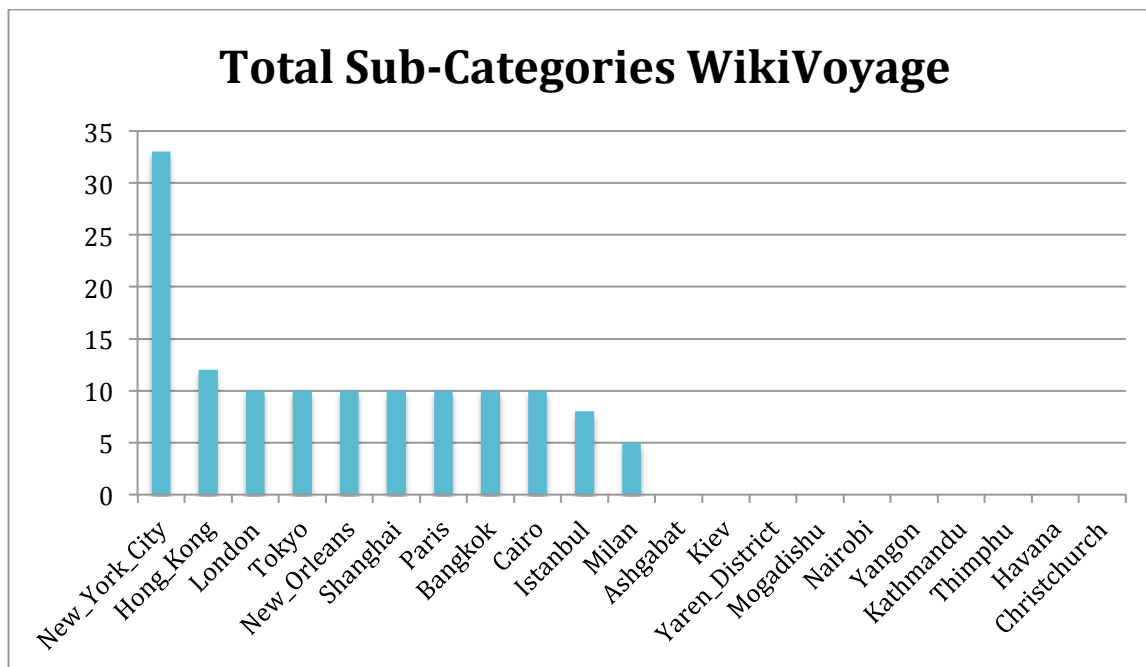
For more details, see the Python files. I've included a brief explanation at the beginning of each file of what the code does. I also specify which files I used to generate my findings in the following section.

²⁰ http://en.wikipedia.org/wiki/Barbie_Shanghai

Findings

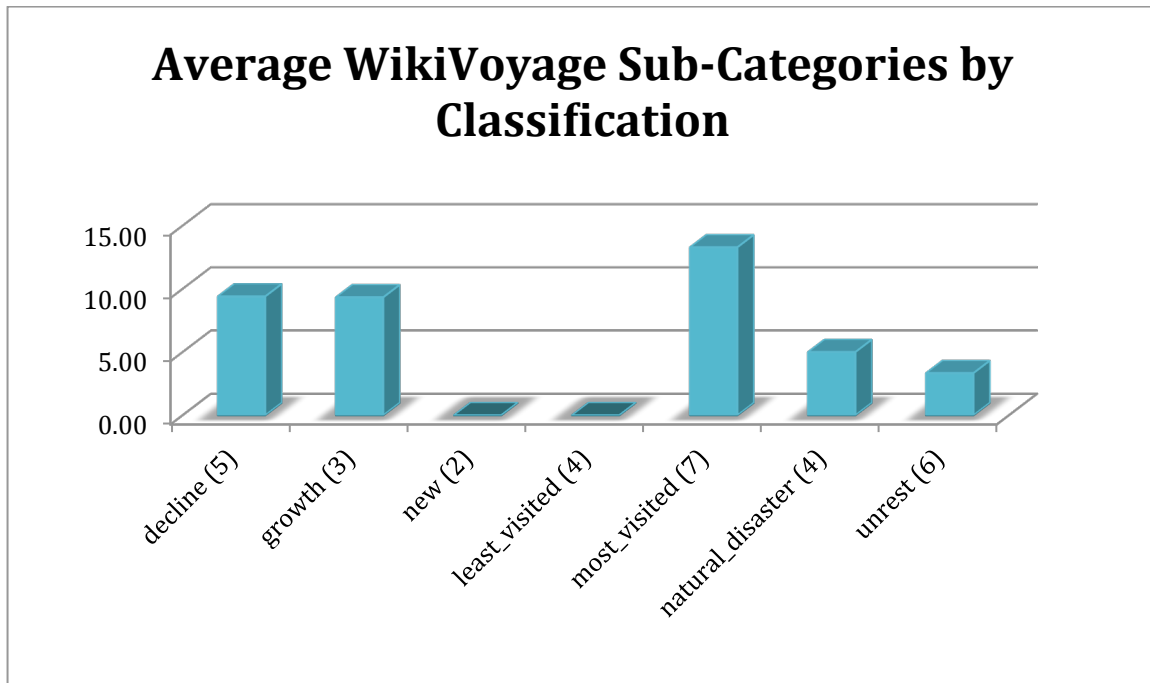
WikiVoyage Sub-Categories

My findings confirm that it's essential to include sub-categories when analyzing destination data on WikiVoyage. Of the 21 destinations in the study, 11 have sub-category pages and 82% of those have 10 or more sub-categories. There were 128 WikiVoyage sub-categories across all destinations in this study.



(see file: 1_destination_sub_category.xls)

After sorting these data by classification (decline, growth, etc.) and taking the average, I found that the most visited destinations have the most sub-categories (13.29 on average) while the newest and least visited destinations have none.



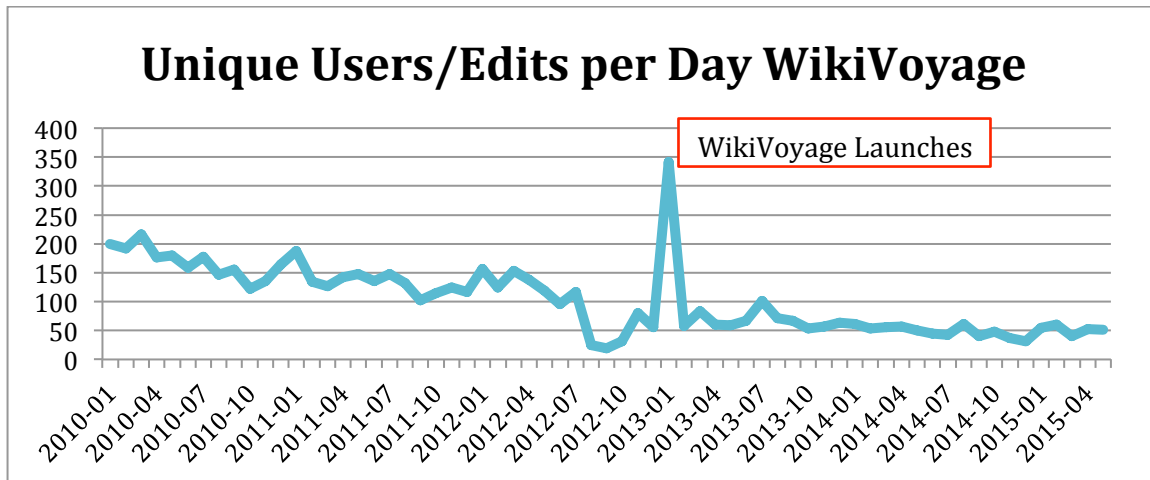
(see file: 1_destination_sub_category.xls)

*Note: the number next to the classification in all charts indicates the number of destinations the classification comprises.

WikiVoyage Data

WikiVoyage has some inherent limitations as a data source, which I will cover in more detail in the limitations section. However, one limitation that is worth covering here is that the site re-launched in the middle of the date range this study analyzes. The site began as WikiTravel.org, but when the community became dissatisfied with the site's host, users decided to fork the project and create a new site: WikiVoyage.org. The site was forked in December 2012 and officially re-launched as WikiVoyage on January 15, 2013²¹. The chart below (which is composed of data from the 21 destinations identified in this study) reflects this move.

²¹ <http://en.wikipedia.org/wiki/Wikivoyage>



(see file: 7_total_unique_edits_by_month.py)

The number of unique edits/users per day, charted by month from 2010 through 2012 illustrates a steady decline in site usage – possibly due to the community's dissatisfaction. The obvious dip in late 2012 coincides with the site move and the sharp spike in January 2013 corresponds to the re-launch of the site. Digging into the data, I found that a lot of content was imported from WikiTravel during that spike.

Since the re-launch, the unique edits/users continued to decline through 2013 and have settled at an average rate of about 50 unique edits per month in 2014 and 2015.

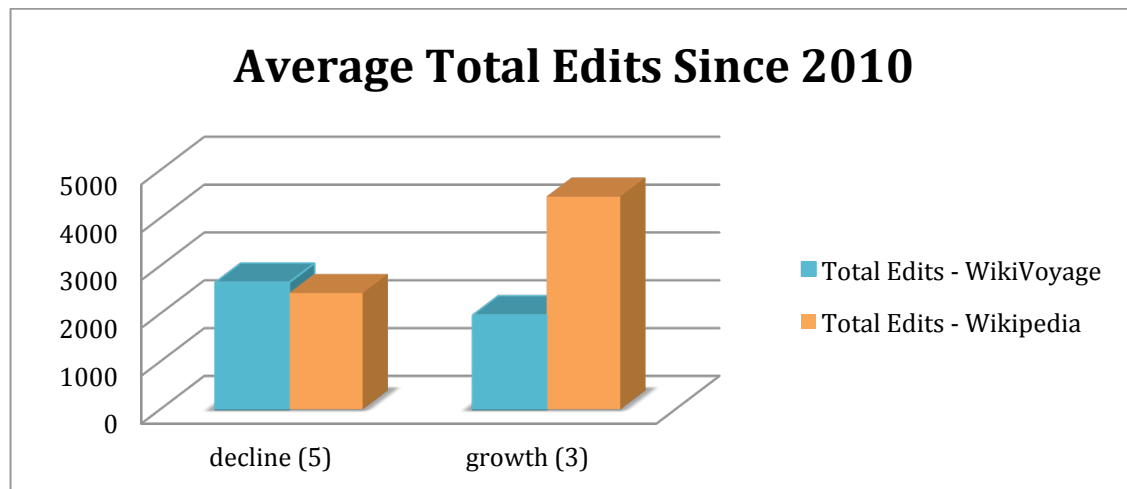
While these data do tell a story, it's clear there are external factors at play. Due to this and the limited amount of data available from WikiVoyage in general (on average there are only 4.9 unique edits/users per month for each destination), I opted to include corresponding Wikipedia data as well.

Findings – Destinations Growing vs. Declining

I hypothesize that destinations that are steadily growing in international overnight visitors vs. those that have seen a decline at least once year over year in the past 5.5 years will have more total edits, unique edits/users, and more total content.

Total Edits – Growth vs. Decline

When looking at average total edits for each classification, we see that the hypothesis holds true for Wikipedia content – there are more total edits for growing destinations. However, we find the opposite for WikiVoyage.



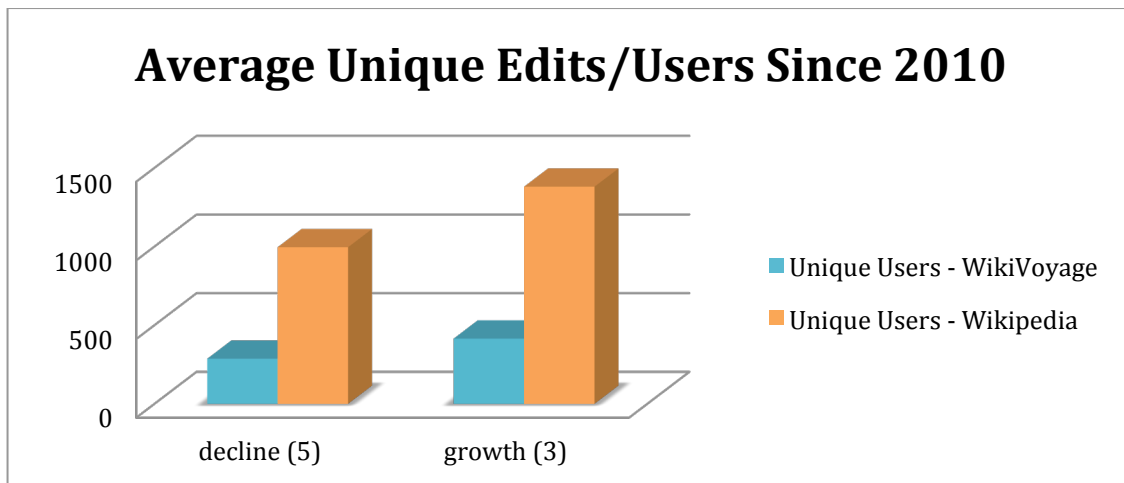
(see file: *total_edits_by_destination.xlsx*)

For WikiVoyage, the average total edits for the decline classification are unexpectedly higher than those in the growth category. The three cities causing this are Bangkok, Hong Kong, and (less so) Shanghai. All destinations, including growth destinations, have a similar number of sub-categories, so sub-category data should not be a factor.

In the following chart we will see that the number of unique edits/users for these classifications do support my hypothesis, which indicates that individual users are, for some reason, making more revisions per user to the content for certain destinations in the decline classification. More research is required to determine if this is a trend, or outlier data (for example, the user Globe-trotter made 4,013 total edits to the Bangkok WikiVoyage page during the timeframe analyzed).

Unique Edits/Users – Growth vs. Decline

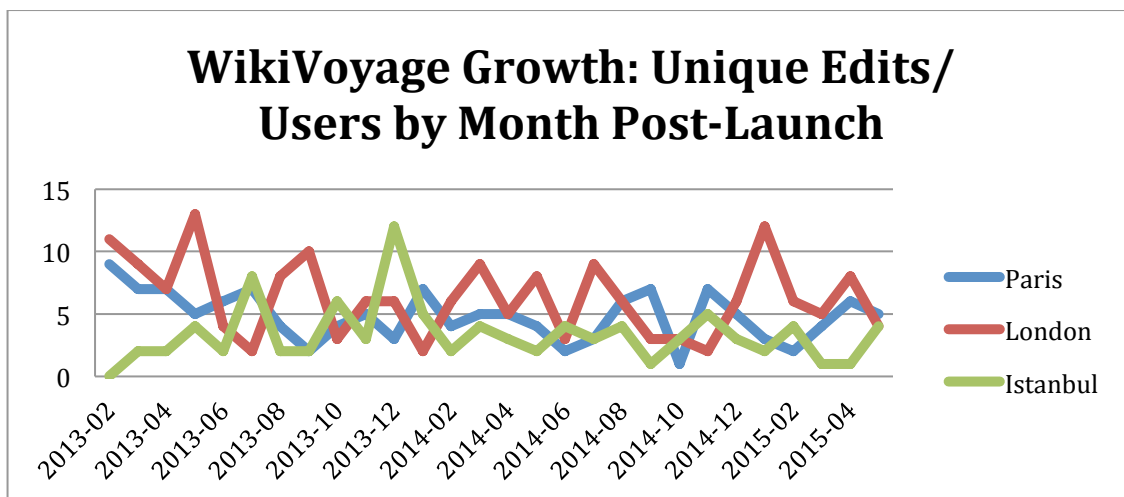
The data support my hypothesis that growing destinations will have more unique edits/users per day than those in decline. This holds true for both Wikipedia and WikiVoyage.



(see file: *unique_edits_by_destination_2.xlsx*)

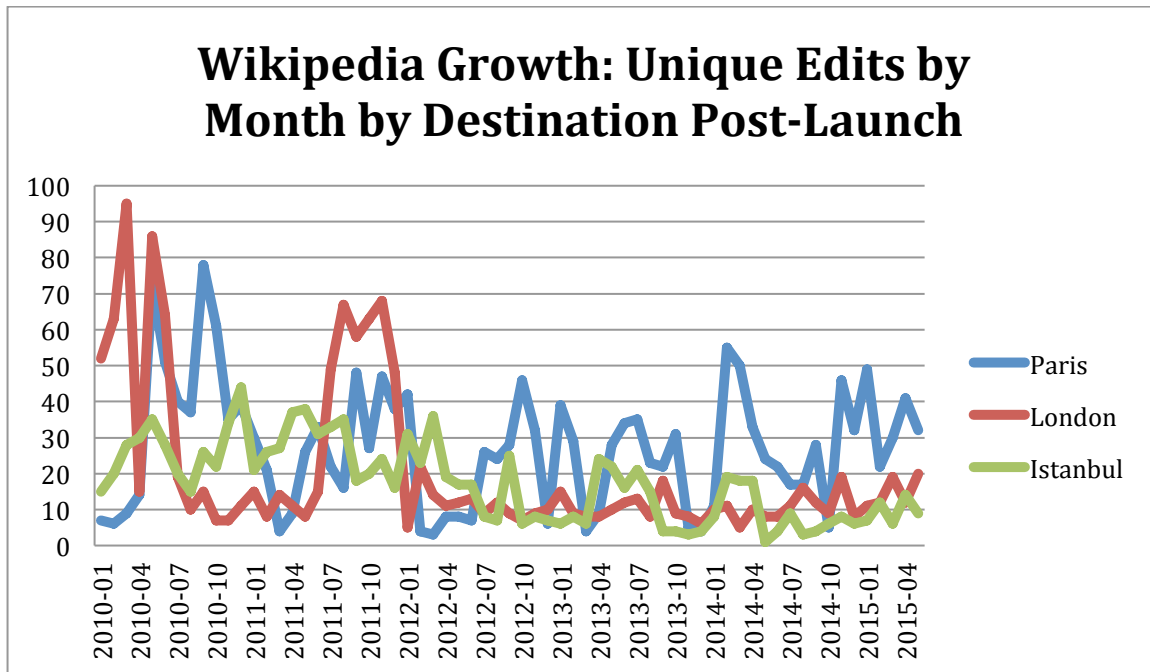
I would also expect to see the overall unique edits/users display a generally increasing trend for growing destinations and a decreasing trend for declining destinations.

I excluded the data prior to the WikiVoyage re-launch in order to remove dissatisfaction with the site as a factor. However, when comparing unique edits/users per day by month across the three growth destinations, we do not find the expected growth trend on WikiVoyage.



(see file: *combo_edits_by_month_final.xlsx*)

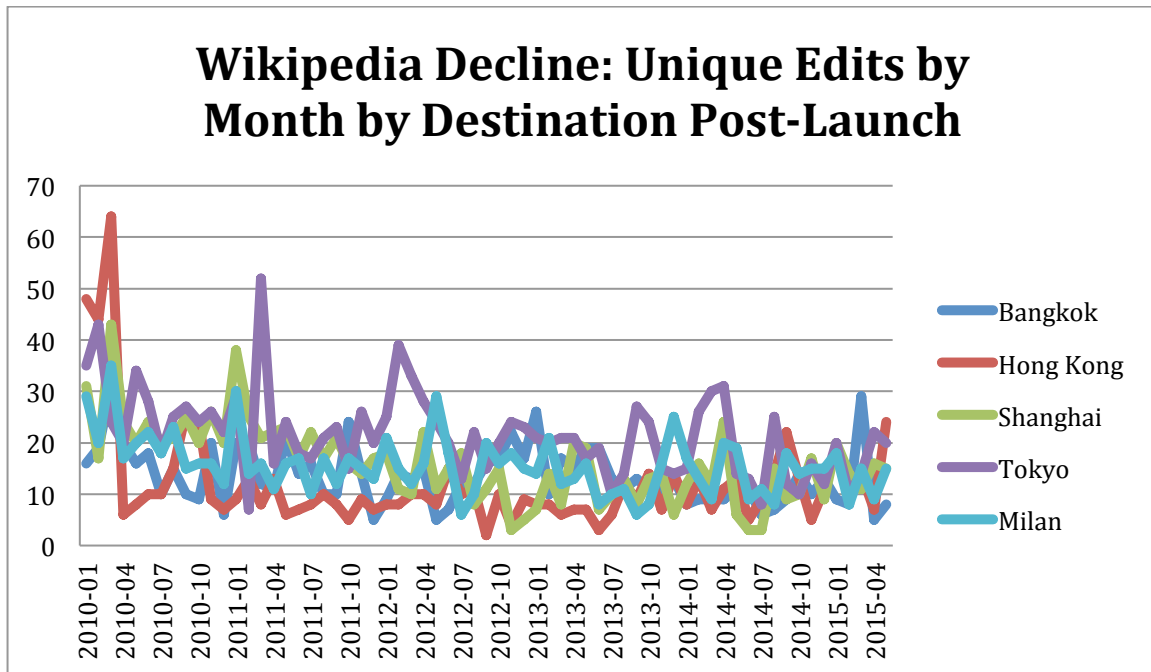
In the corresponding Wikipedia graph for growth (which does include data back to 2010), we see a steady, or even slightly declining trend where we would expect to see growth.



(see file: *combo_edits_by_month_final.xlsx*)

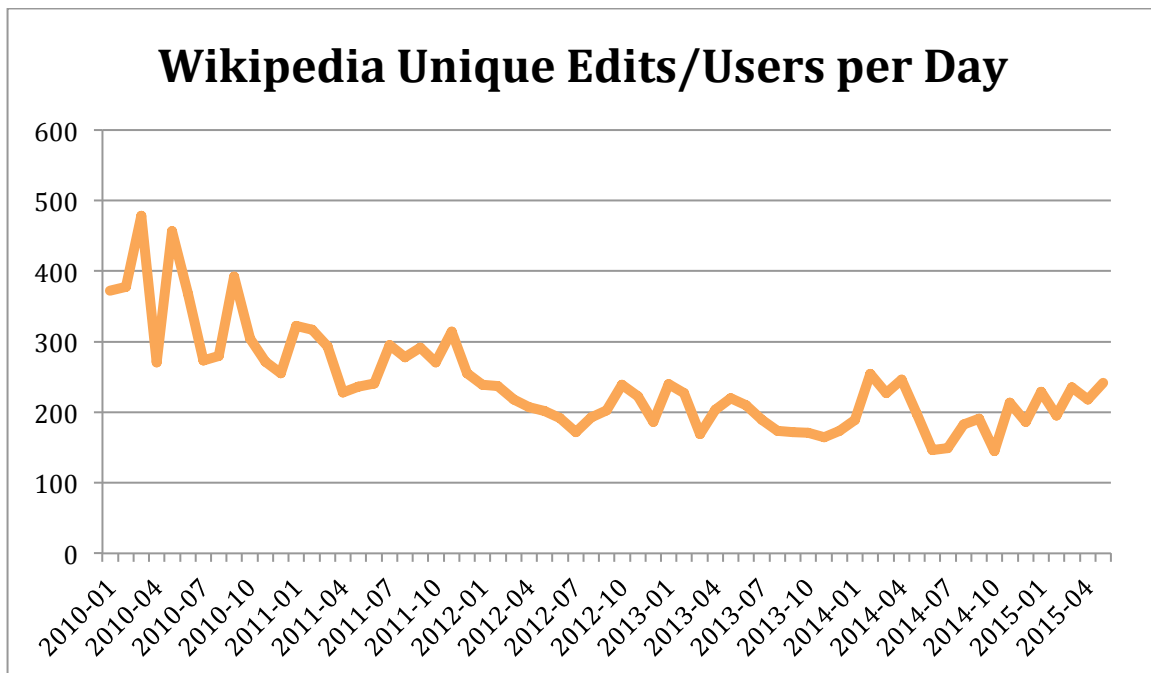
Since the year over year decline in international overnight visitors occurred for most destinations in the data set over the same timeframe where we saw an overall decline in WikiVoyage usage (2010–2012), the WikiVoyage data will not offer useful insights and thus I will not include it.

When we look at the Wikipedia data for unique edits/users per day by month for destinations that have seen a decline in international overnight visitors, we do see a slight downward trend from 2010–2012 (during the time the destinations – other than Bangkok, which saw decline from 2013–2014 – were in decline).



(see file: *combo_edits_by_month_final.xlsx*)

However, the general Wikipedia trend over the same timeframe has been a decline in unique edits/users as well.



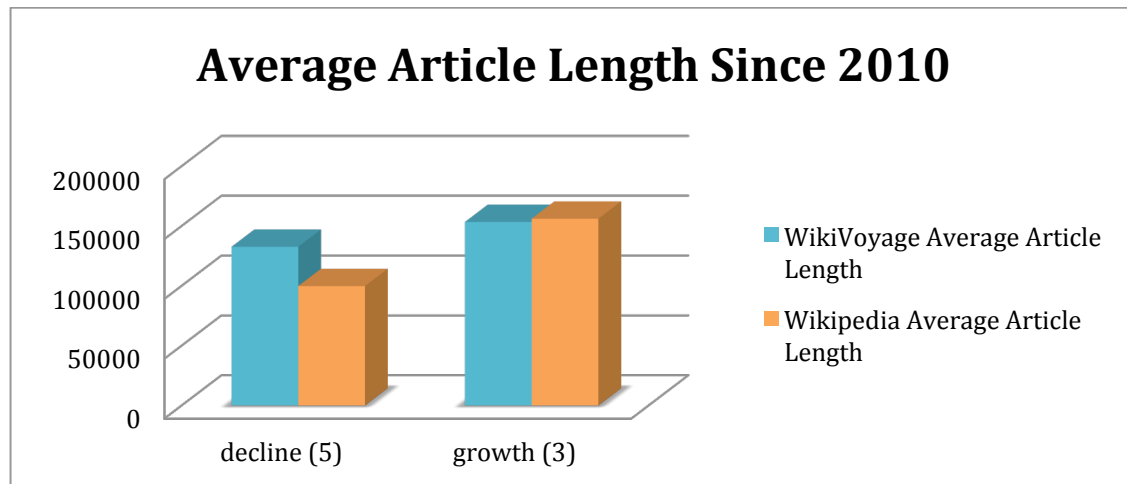
(see file: *no_cats_wikipedia_7_total_unique_edits_by_month.py*)

Therefore, I am unable to successfully isolate the data in such a way that will provide a meaningful view of the revision trends for growing vs. declining

destinations. More destinations need to be reviewed and the general downward trend for Wikipedia as a whole should also be accounted for in order to clean up the data.

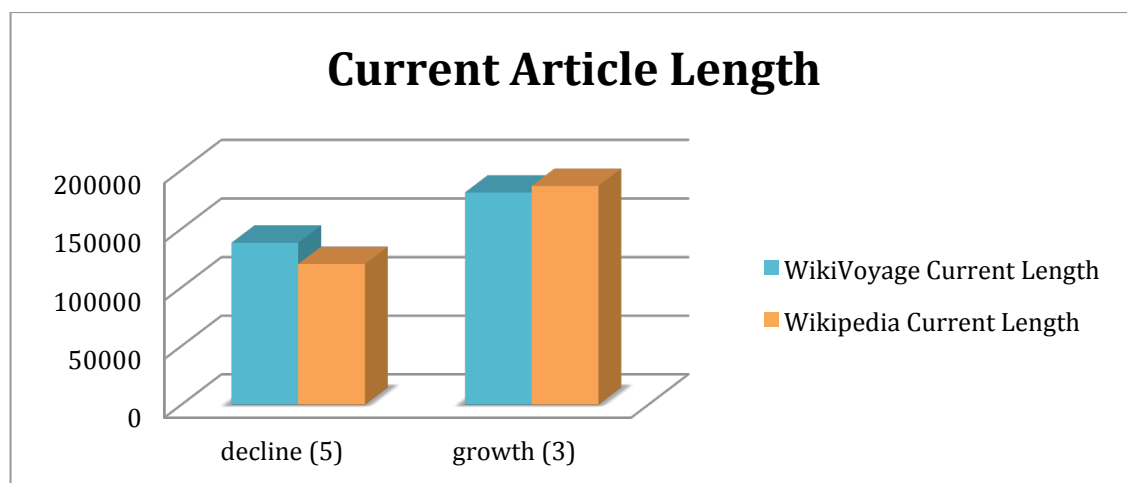
Total Content – Growth vs. Decline

When comparing average article length for top-level pages (no sub-categories) over time, we see that, as expected, growing destinations have more content than those in decline on both Wikipedia and WikiVoyage.



(see file: *article_length.xlsx*)

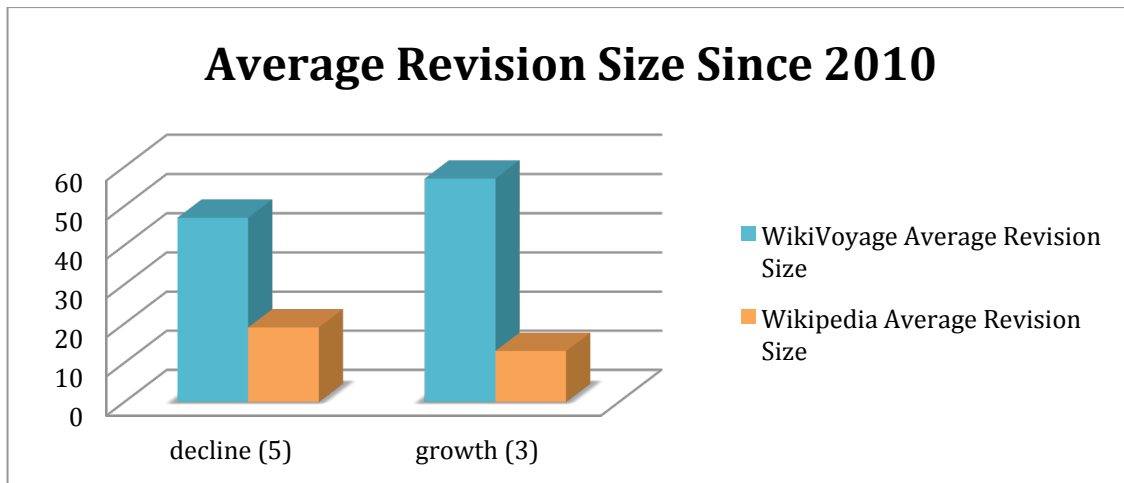
Looking at the current article length shows the same trend.



(see file: *article_length.xlsx*)

When reviewing average revision size, however, we do see some unexpected results. Although the WikiVoyage revision trend supports the hypothesis that users will make longer revisions to growing destinations, it's surprising that the average WikiVoyage revision size is significantly larger than the average Wikipedia revision

size. This holds true in all subsequent sections. Users are more likely to make larger revisions on WikiVoyage than Wikipedia.



(see file: article_length.xlsx)

Additionally, declining destinations have slightly larger revisions than growing destinations on Wikipedia.

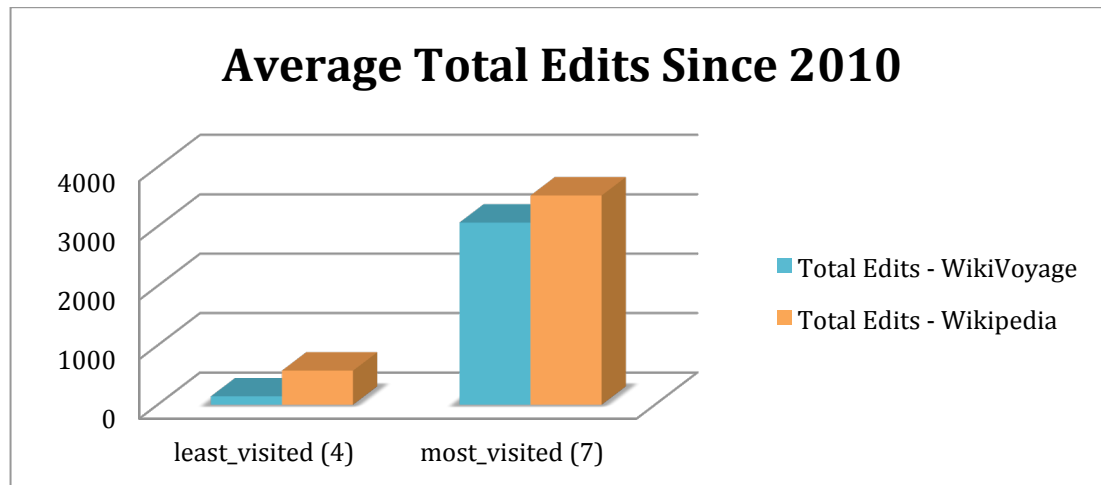
A larger data set and a deeper look at why Wikipedia users behave differently are warranted.

Findings – Most vs. Least Visited Destinations

When comparing most visited destinations to least visited destinations, I hypothesize that the most visited destinations will see more total edits, more unique edits/users, and more total content.

Total Edits – Most vs. Least Visited

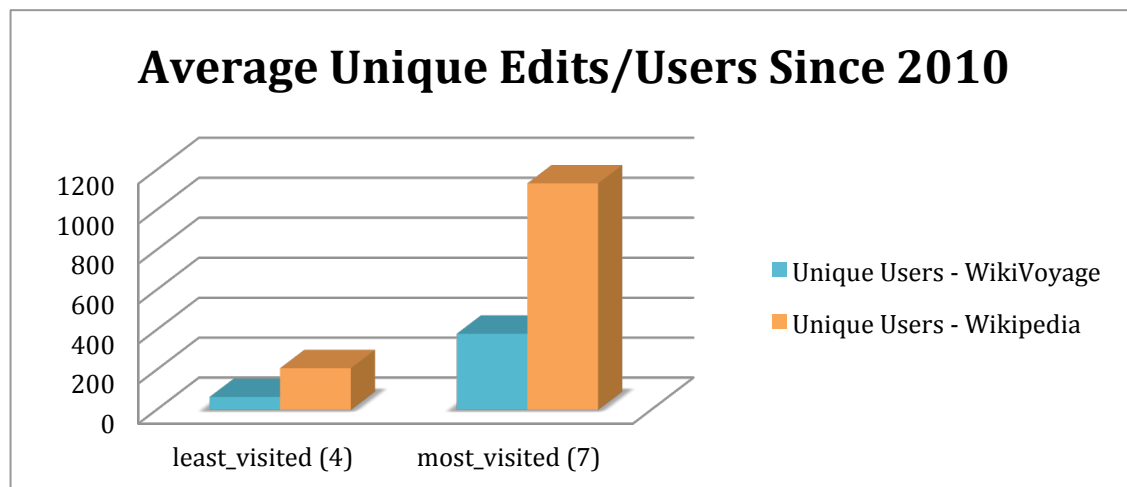
The data confirm my hypothesis that the most visited destinations will see the most total edits. This is true for both Wikipedia and WikiVoyage.



(see file: *total_edits_by_destination.xlsx*)

Unique Edits/Users – Most vs. Least Visited

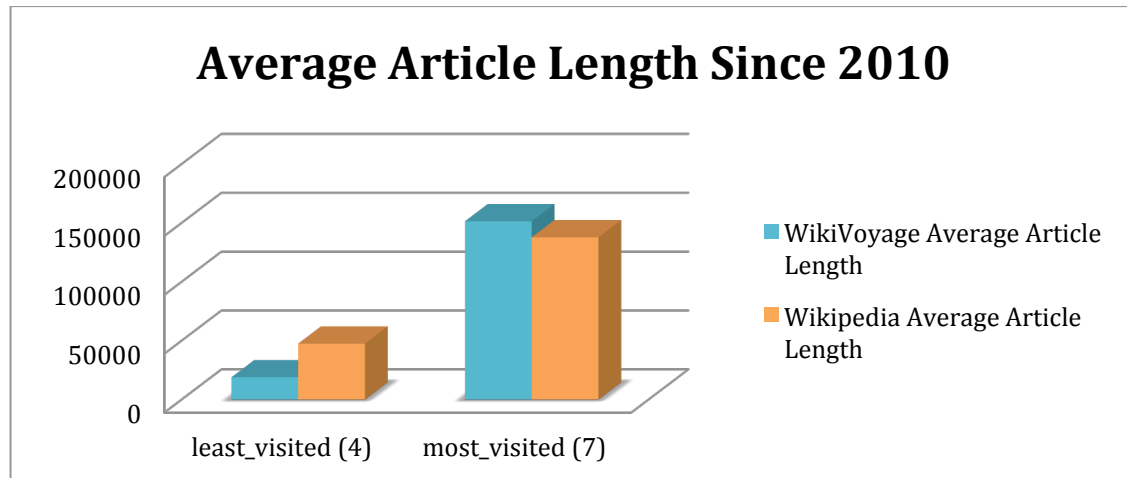
The data also validate the theory that the most visited destinations will have more unique edits/users.



(see file: *unique_edits_by_destination_2.xlsx*)

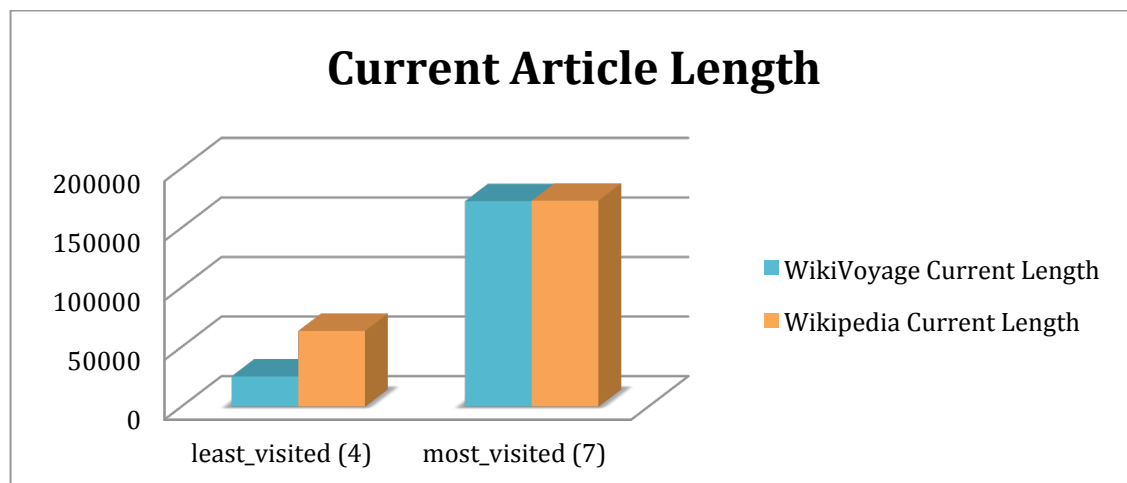
Total Content – Most vs. Least Visited

Similarly, we see longer articles on both WikiVoyage and Wikipedia, on average, for destinations that are visited most.



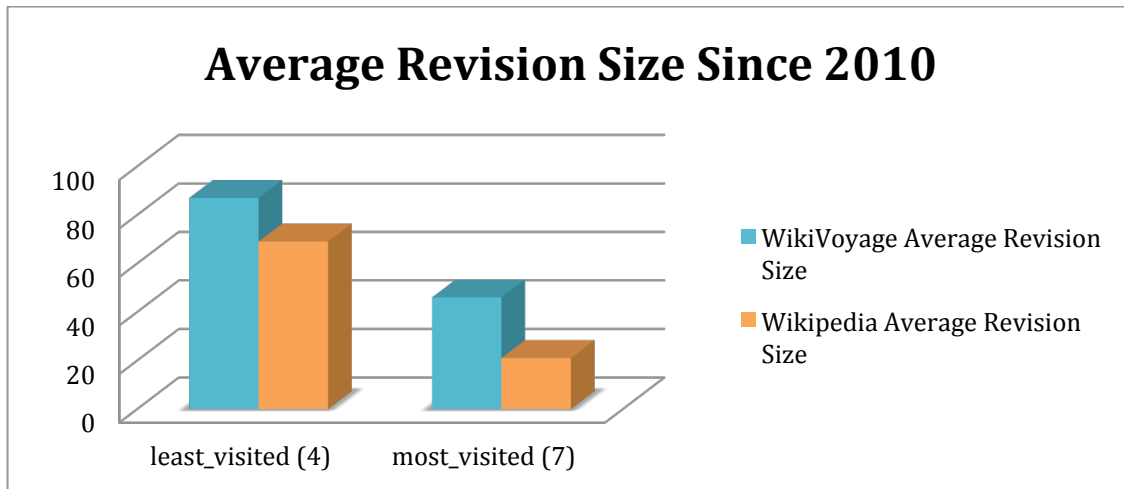
(see file: *article_length.xlsx*)

This is also true when comparing current article lengths for both sites.



(see file: *article_length.xlsx*)

The only unforeseen outcome when comparing most visited to least visited destinations is the average revision size. Not only are WikiVoyage revisions consistently longer than Wikipedia revisions (which we saw previously), the least visited destinations are significantly more likely to have larger revisions for both Wikipedia and WikiVoyage than the most visited destinations.



(see file: *article_length.xlsx*)

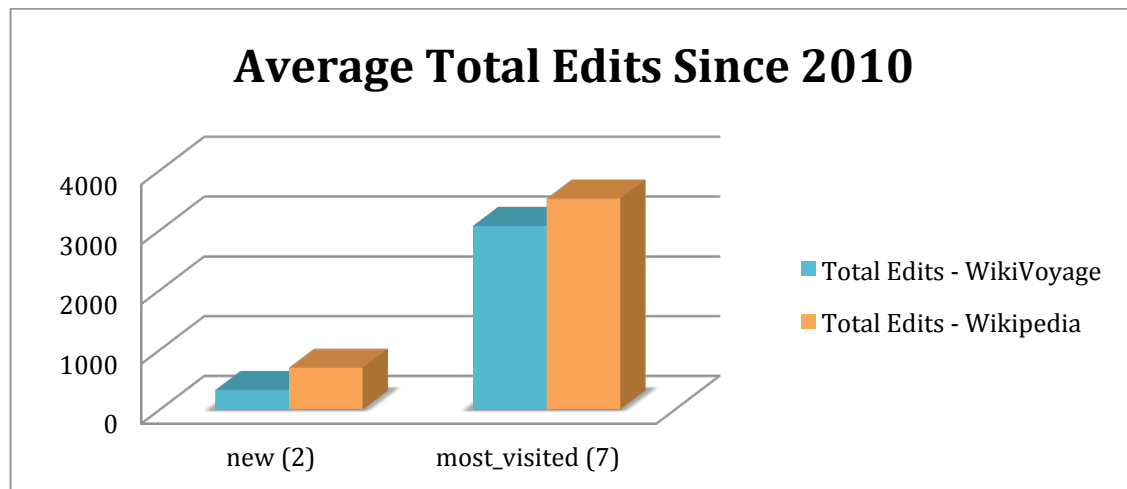
The average WikiVoyage revision size is nearly twice as big for least visited destinations and the Wikipedia revision size is more than three times as large for destinations that are visited least. This finding suggests that articles with less content may encourage users to make larger revisions than articles that already have a lot of content.

Findings – New vs. Established Destinations

For new vs. established (which, in this study, is the same as most visited) destinations, I hypothesize that new destinations will see fewer total edits and fewer unique edits/users than well-established destinations. However, I posit that new destinations will see larger revision sizes than established destinations.

Total Edits – New vs. Established

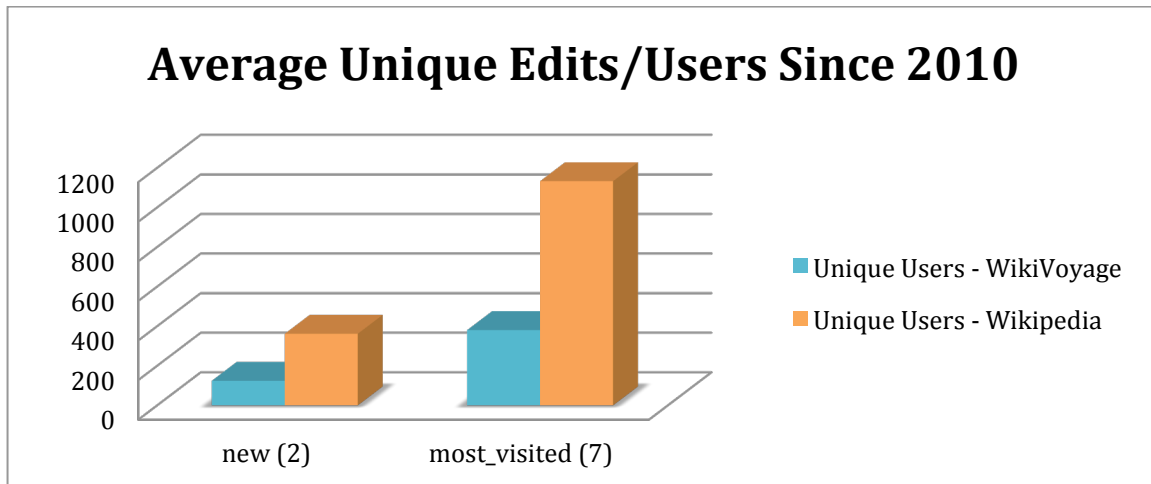
Indeed, the average total edits for established destinations are much higher than those for new destinations for both Wikipedia and WikiVoyage.



(see file: *total_edits_by_destination.xlsx*)

Unique Edits/Users – New vs. Established

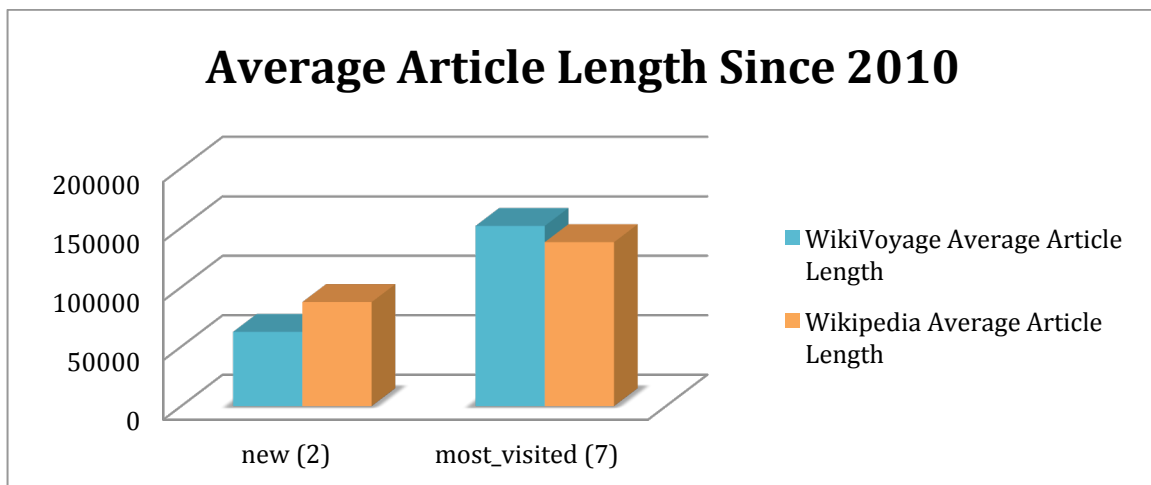
We encounter similar findings when analyzing average unique edits/users: established destinations have more unique edits/users than new destinations.



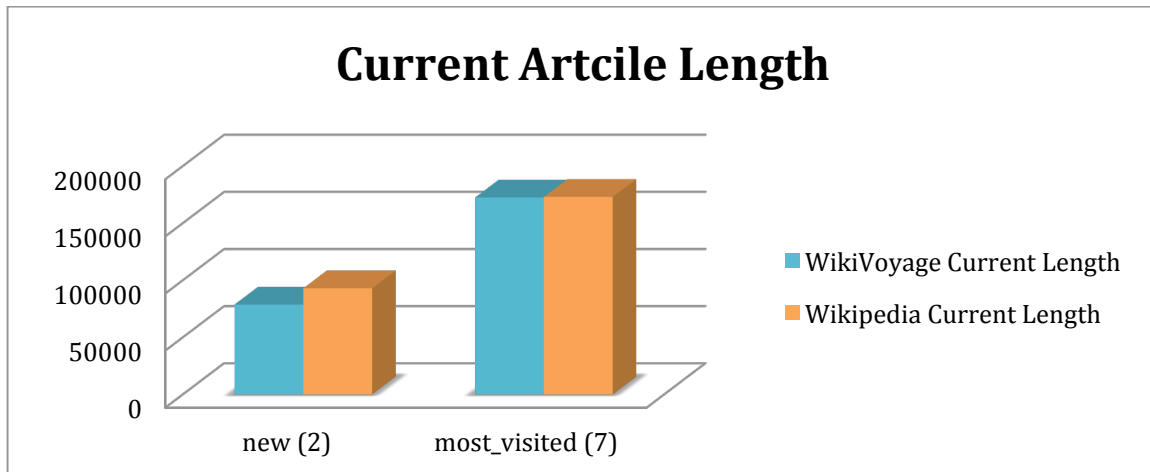
(see file: *unique_edits_by_destination_2.xlsx*)

Total Content – New vs. Established

Comparing average and current article lengths, we also see that established destinations tend to have longer articles.

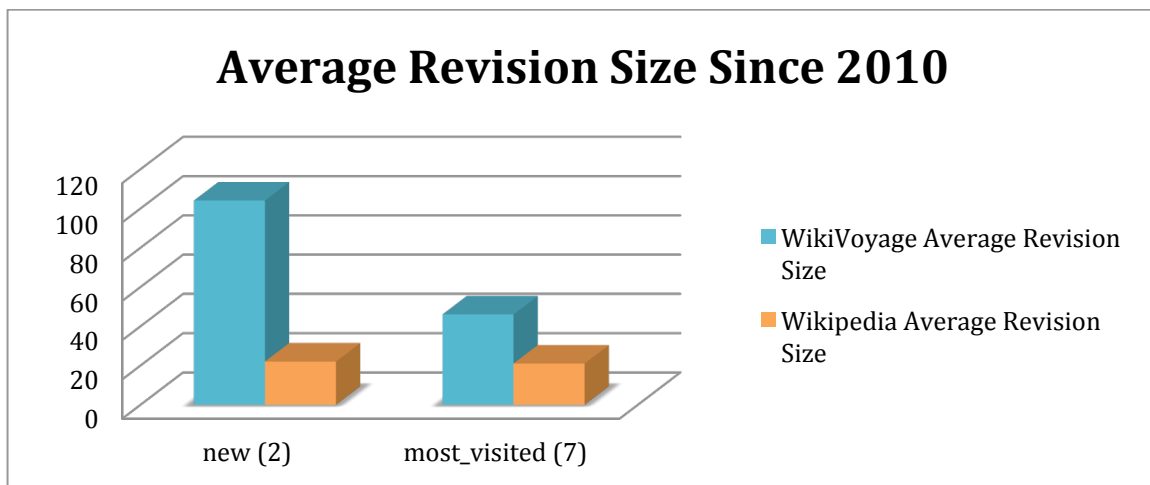


(see file: *6_article_length.xlsx*)



(see file: 6_article_length.xlsx)

Average revision size is again the comparison that offers the most interesting results. Just as we saw when comparing average revision sizes for the least visited destinations to the most visited, new destination average revision sizes, at least on WikiVoyage, are likely to be more than twice as large as revisions for established destinations.



(see file: 6_article_length.xlsx)

In this case, the Wikipedia revisions are almost identical in size, which would require further analysis to explain.

Nevertheless, these findings bolster the theory that less content in a travel article motivates users to make larger revisions to it. Since fewer people are traveling to least visited and new destinations, the people that have traveled to those destinations likely feel more obliged to contribute more exhaustive revisions.

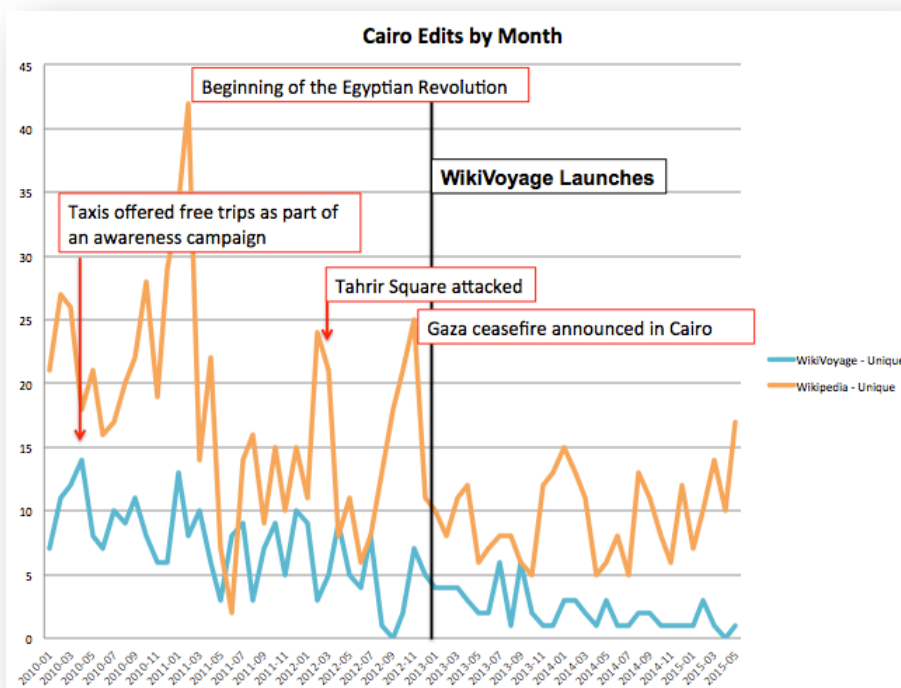
And, again, we see that the WikiVoyage average revision size is larger than that of Wikipedia.

Findings – Negative Events

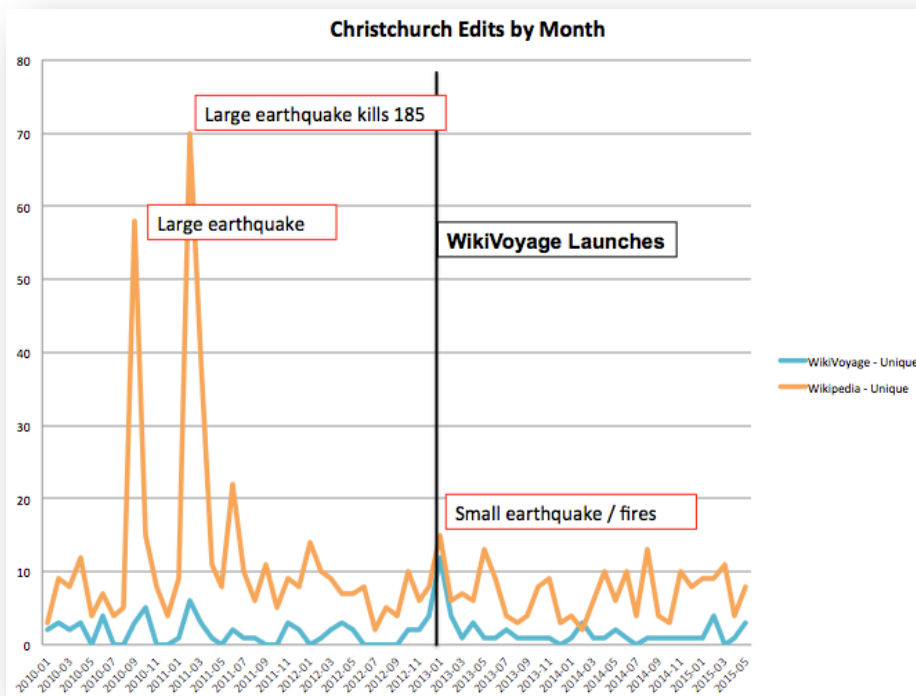
A detailed look at the unique edits/user data for individual destinations suggests that negative events, especially political unrest, terrorist attacks, and natural disasters do encourage more people to revise destination content.

The following graphs for Cairo, Christchurch, and Bangkok (which is not classified as a destination that experienced negative events) demonstrate that unique edits/users per day spike shortly after a negative event has occurred.

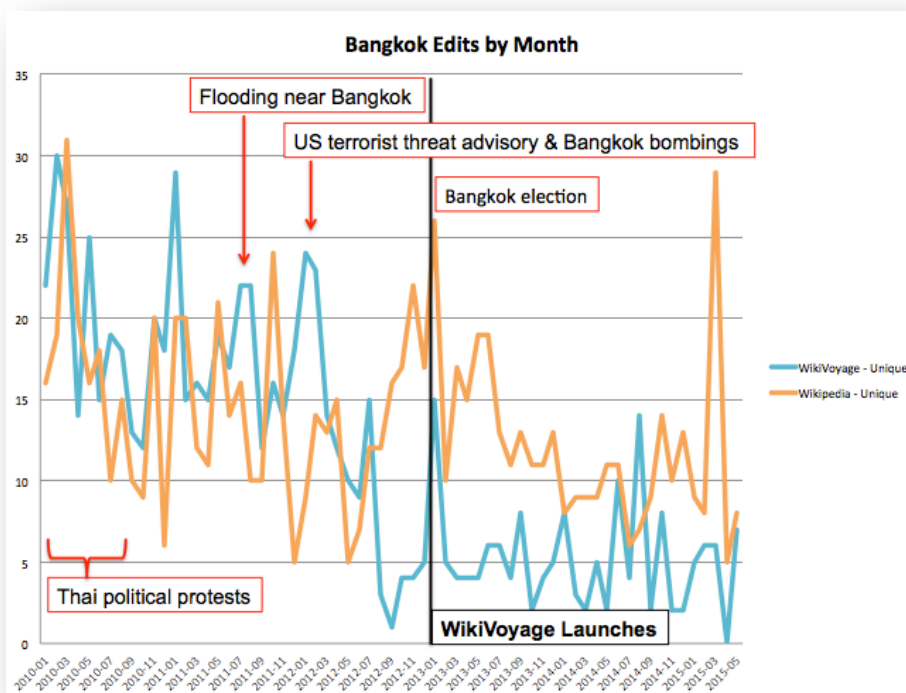
Notice that many of the negative event-related spikes appear on both Wikipedia and WikiVoyage.



(see file: *combo_edits_by_month_final.xlsx*)



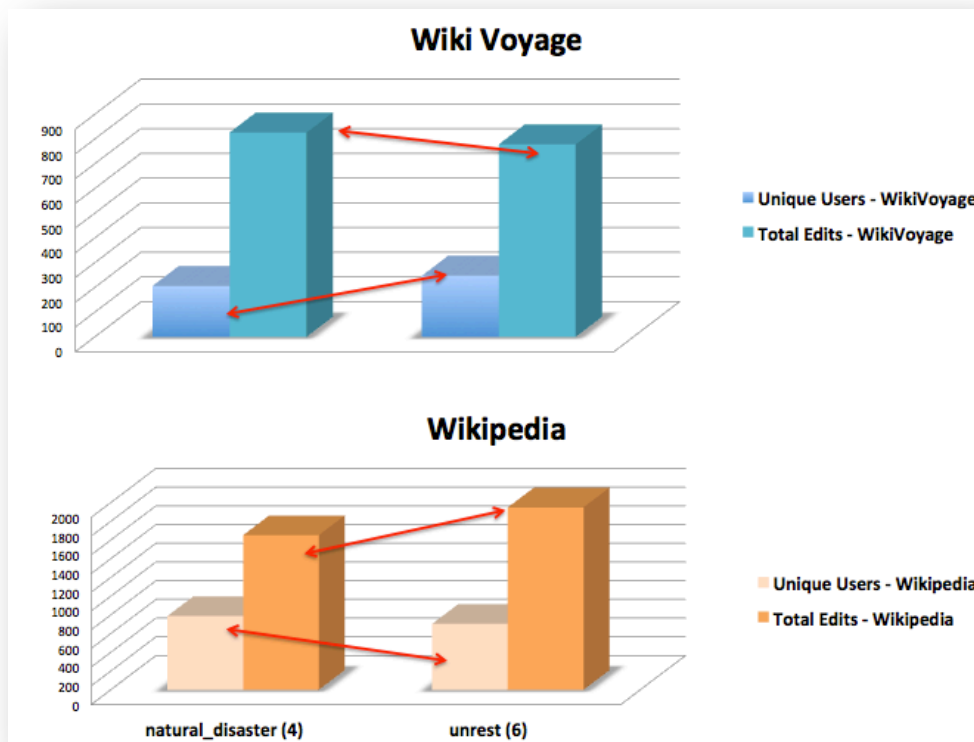
(see file: *combo_edits_by_month_final.xlsx*)



(see file: *combo_edits_by_month_final.xlsx*)

The Bangkok graph, in particular, demonstrates that Wikipedia and WikiVoyage users do not behave the same when it comes to editing negative event content across sites. For example, the number of unique edits/users on WikiVoyage spikes beyond unique Wikipedia edits/users after a large flood and after bombings in Bangkok.

Based on this, I looked more closely at how Wikipedia users compare to WikiVoyage users when revising content about a natural disaster vs. political unrest. Although the data is very close, user behavior seems to be opposite for the two sites.



(see files: *total_edits_by_destination.xlsx* and *unique_edits_by_destination_2.xlsx*)

For a natural disaster, WikiVoyage has fewer unique users making more total edits. For political unrest, WikiVoyage sees more unique users making fewer total edits. Wikipedia users have the exact opposite behavior.

More research is essential to confirm this finding. However, it generates a new hypothesis:

I surmise that, when political unrest occurs, WikiVoyage users simply need to know if it's safe to travel to the region or not (fewer edits). However, when a natural disaster occurs, it can change the infrastructure of the destination over time (more edits). Conversely, for Wikipedia users, political unrest is

often an ongoing situation where its impact continues to morph (more edits), while a natural disaster is a one-time event that, while still important, does not garner as much interest over time (fewer edits).

Conclusions

Although I found it difficult to compare data from growing vs. declining destinations because too many external factors, especially the overall decline in edits over time on Wikipedia and WikiVoyage, impacted the results, the data I compiled did generally corroborate my hypotheses that more visited, better established destinations have more total edits, more unique edits/users, and more content than lesser visited and newer destinations.

The data support the theory that new destinations will have larger revision sizes than established destinations. Along with the unexpected, but logical finding that this is also true for destinations that are visited the least, it seems probable that the less content a destination has, the more likely its users are to contribute larger revisions.

It was surprising to find that, although overall and average content lengths didn't differ as significantly, WikiVoyage users were likely to have much larger revision sizes than Wikipedia users across all classifications.

Finally, users have opposite behaviors when revising natural disaster vs. political unrest content on WikiVoyage when compared to Wikipedia. In order to understand and confirm this phenomenon, more research is required.

As travelers researching destinations online, we can be confident that we'll have more information from more sources on better-established, more-often-visited destinations. We can also rely on users to quickly update content shortly after a major political or terrorist event or natural disaster. As destinations become less popular, the amount of content available and the number of people making updates will decrease. However, the people that are updating that content will make longer revisions.

Limitations and How this Study Can be Improved

This project is limited by my initial assumptions about which factors are likely to motivate users to contribute to a travel article. I specifically look at international overnight visitors year over year, most and least visited destinations, major local events for each destination, and how recently a destination has opened to tourism. Although I made an effort to correct for the general decline in Wikipedia edits over time, it's possible I missed other factors. Perhaps users contribute revisions differently for destinations in Asia as compared to the United States; this study does not determine that.

It's also difficult to isolate destinations into a single classification. One-third of the destinations in this study fall into more than one classification (i.e. least visited and political unrest). Therefore, the comparisons are not absolute. The results of this study may be skewed based on how the classifications work in conjunction with each other.

Another limitation of this study is that the impact of each factor is difficult to quantify and could affect how likely a user is to revise an article. I do not try to measure how people perceive the impact of the recent earthquake in Kathmandu vs. that of the 2011 earthquake in Christchurch, for example.

Finally, the limited data available from WikiVoyage makes it difficult to achieve statistical significance for these findings. Including Wikipedia data helps combat this, but introduces data that is not travel-specific.

Future analyses should find more destinations for each classification across a wider variety of countries and continents. They should also attempt to isolate factors studied by comparing destinations that fall into the same classification during the same timeframe, which would reduce the effects of external influences on the data.