

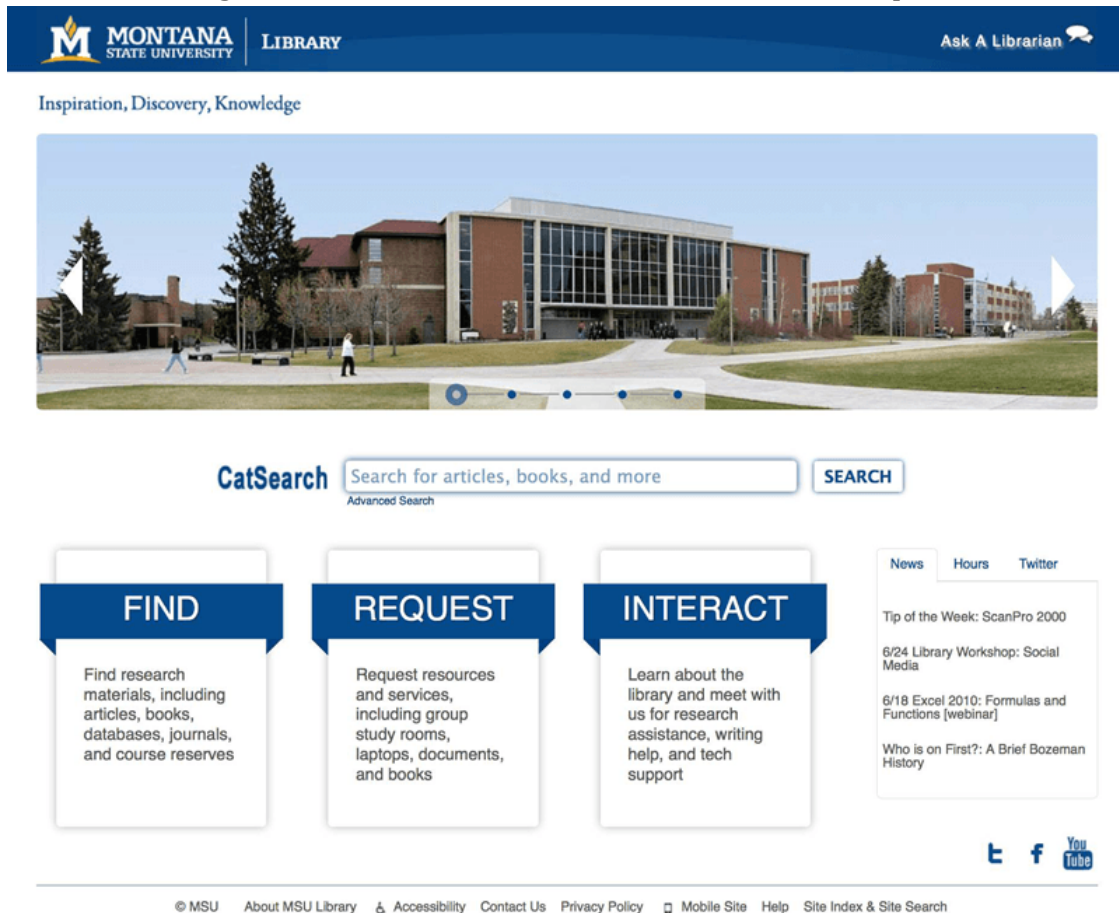
Portfolio: Web Analytics Case Study

Mitchell D. Shuey

January 17, 2019

Introduction:

A librarian and researcher, Scott H. W. Young, working with Montana State University noticed that user services were not used frequently by users. Young used this opportunity to design an A/B test and publish both the results and the data collected. In addition to the test itself, Young conducted user interviews and recorded the responses.



The original index page design, with user services linked via the subtitle “Interact”

About the Data:

Several variations of the MSU library index page were made in the pursuit of A/B Testing their front page. Specifically, the altered piece of the website was the headline linking to user services the library provides. The original, “Interact”, and four alternatives were considered: “Connect”, “Learn”, “Help”, and “Services”.

Google Analytics and a click-tracking service (“CrazyEgg”) were used to generate the data. Users, when visiting the site, were randomly sent to one of the five variations. Our goal is to determine which led to more clicks (the “conversion rate” or click-through rate) overall.

The first dataset, titled Lib_msu, shows all the pages of the whole website that were clicked on between May 29th and June 18th, 2013 (the length of the trial), from Google Analytics. The variables of interest include the total page views, average time on the page in seconds, bounce rate, exit rate, and the click-through rate.

Definitions

Exit Rate- Leaving the site entirely after reaching one of the home page variations. May have visited other pages prior to ending at this page.

Bounce Rate- Leaving the site after reaching one of the home page variations. Entering and leaving, nothing more.

The next five sets of data, ctrl and v1-v4, are the index pages’ click counts over those three weeks for each variation. These datasets and their respective heat maps were generated by CrazyEgg. The only key field of interest on each is the click count for the user services section, and after this observation is collected the datasets will no longer be used.

```
## -- Attaching packages -----
tidyverse 1.2.1 --

## v ggplot2 3.1.0      v purrr  0.3.0
## v tibble  2.0.1      v dplyr  0.7.8
## v tidyr   0.8.2      v stringr 1.3.1
## v readr   1.3.1      v forcats 0.3.0

## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Now we’ll begin taking apart the data. The piece of interest is the clicks on each subtitle, compared to the total overall from the Lib_msu file.

```
clicks = rbind(ctrl[10,], v1[7,], v2[11,], v3[8,], v4[8,])
index = Lib_msu[3:7,]
#Re-ordering our columns
index=index[c(4,5,2,3,1),]
clicks=cbind(clicks[, -6], index)
```

#Calculating click rate

`clicks$Clickthrough_Rate=clicks$`No..clicks`/clicks$Pageviews`

`clicks$ctr_share=clicks$No..clicks/sum(clicks$No..clicks)`

	Name	Page	Pageviews	Unique Pageviews	Avg. Time on Page	Entrances	Bounce Rate	% Exit	Clickthrough_Rate	ctr_share
1	INTE	/	2206	1750	57.83	1694	0.420	0.388	0.0190390	0.211
0	RACT				543		8973	4859		0553
7	CONN	/index	2099	1618	69.40	1583	0.456	0.414	0.0252501	0.266
	ECT	2.php			000		0960	0067		3317
1	LEAR	/index	2213	1727	74.00	1663	0.467	0.430	0.0094894	0.105
1	N	3.php			000		8292	6371		5276
8	HELP	/index	2210	1701	98.77	1657	0.499	0.449	0.0171946	0.190
		4.php			056		0947	7738		9548
8	SERVI	/index	2284	1806	73.93	1759	0.486	0.447	0.0197023	0.226
1	CES	5.php			903		0716	0228		1307

We see that with a rate of 2.5%, “Connect” has the highest click-through rate, but we need to determine the significance of this with a test. Visuals including error bars will also help judge the differences of results.

`library(ggplot2)`

`plotset=gather(clicks, RateType, percent, 11, 12, 14)`

#Calculating standard deviations for each rate versus the variation's , bernuolli

`plotset$sd=sqrt(plotset$percent*(1-plotset$percent)/plotset$Pageviews)`

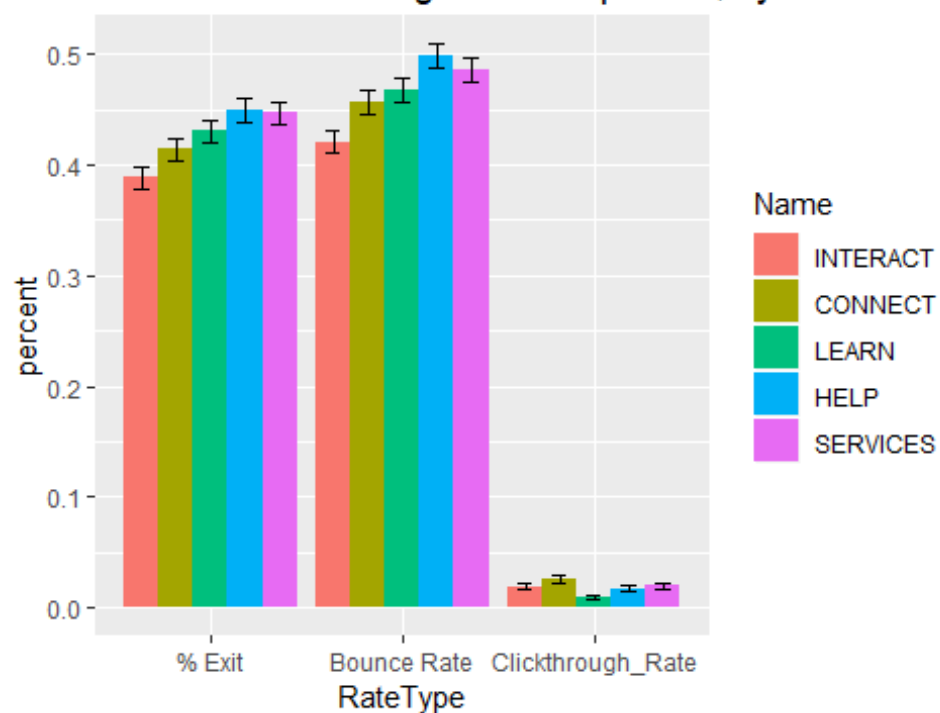
#plotting

`rates=ggplot(plotset, aes(RateType, percent, fill=Name))`

`bars=rates+geom_col(position="dodge")`

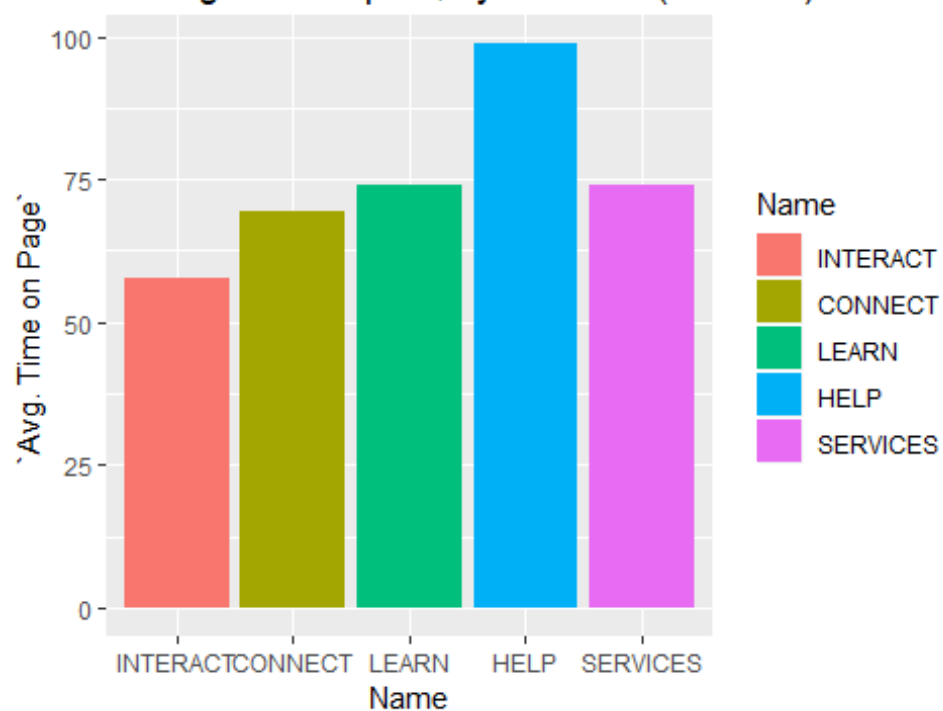
`bars+ geom_errorbar(aes(ymin=percent-sd, ymax=percent+sd), width=.4, position=position_dodge(.9))+ggtitle("Rates of Click-throughs and Departure, by Variation")`

Rates of Click-throughs and Departure, by Variation



```
ggplot(plotset, aes(Name, `Avg. Time on Page`, fill=Name))+
  geom_col(position="dodge")+ggtitle("Average Time Spent, by Variation
(seconds)")
```

Average Time Spent, by Variation (seconds)



```
share=ggplot(clicks, aes(No..clicks, fill=Name))
```

While it's hard to interpret whether a higher or lower exit rate means users found what they were looking for, we know that a lower bounce rate is better because the users found what they were looking for.

Hypothesis Testing

We initially assume the null: All treatments (versions) have an equal conversion rate The alternative: they are unequal in some degree.

We can then dive into comparing all five click-through rates against each other, with the p-values in the matrix representing the chance that the difference between, say, the control (1) and "Help" (4) was due to the randomness of the data:

```
prop.test(clicks$No..clicks, clicks$pageviews)

##
## 5-sample test for equality of proportions without continuity
## correction
##
## data: clicks$No..clicks out of clicks$pageviews
## X-squared = 15.836, df = 4, p-value = 0.003248
## alternative hypothesis: two.sided
## sample estimates:
##      prop 1      prop 2      prop 3      prop 4      prop 5
## 0.019038985 0.025250119 0.009489381 0.017194570 0.019702277

#The p-value is significant enough to plan further testing

pairwise.prop.test(clicks$No..clicks, clicks$pageviews, p.adjust.method =
"bonferroni")

##
## Pairwise comparisons using Pairwise comparison of proportions
##
## data: clicks$No..clicks out of clicks$pageviews
##
##      1      2      3      4
## 2 1.0000 -      -      -
## 3 0.1075 0.0011 -      -
## 4 1.0000 0.8322 0.3553 -
## 5 1.0000 1.0000 0.0646 1.0000
##
## P value adjustment method: bonferroni
```

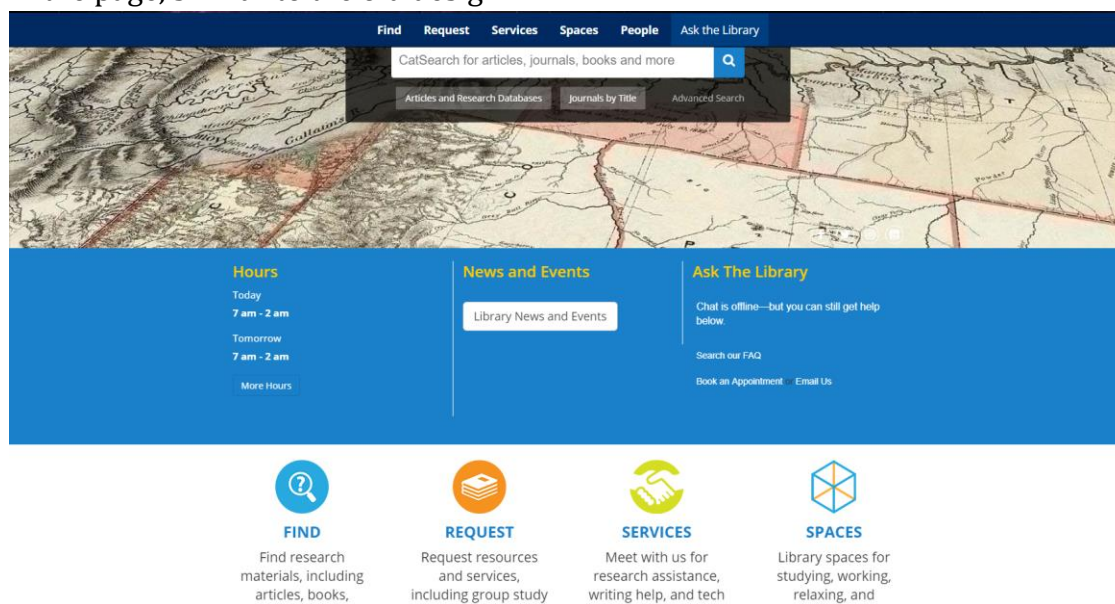
Results

According to our proportion test, despite the Click-through rates being different, using the standard 5% p-value cutoff, few have a significant difference from other variations 1-on-1. The results of the original research state that “Learn” has the worst click-through performance. I conclude no significant difference was found within the data here; however this does not mean we are unable to make a recommendation.

Looking at the average time chart, we can see that the control, “Interact”, has the shortest average time on page. In an information services context such as a library, where finding what you want on the page quickly is important, this could mean that the experiment implementation led students to take more time on the home page to make sure they were going to select what they needed. They might be used to looking for “Interact”, versus interpreting the new variations. This is an important effect to consider when changing sections of a website. #Conclusion Given the chance to repeat the test, less variations, or limiting testing to one alternative at a time would be a good idea. It’s worth noting that comparing all of these proportions to each other at once calls for adjustments to the p-value. In this case, using the “Bonferroni” method, all p-values were multiplied by 15, the number of proportions being compared.

The timing of the experiment is also crucial. The number of users would likely be higher during a semester, rather than late May into early June. Another variation worth testing is the placement of the link within the website. It may be better served within the sidebar, or as a floating option in a corner.

It’s worth noting that on the MSU library’s current page, the choice was made to use “Services” and placed links in a hovering toolbar as well as alongside the other options later in the page, similar to the old design.



A section of the website's new design. Notice the placement of links both at the top and side by side on the page.

The original paper made different choices in calculating the click rates and decided to forego testing for significance; I have contacted the author and asked about the calculation methodology and hope to better replicate Young's results. The original paper also conducted user surveys, which offer far more contextual insight into a user's behavior and final click decisions. A good insight campaign uses all available sources of consumer experience information.

Works Cited:

Young, Scott W. H. "Improving Library User Experience with A/B Testing: Principles and Process." Weave, Michigan Publishing, 2014, dx.doi.org/10.3998/weave.12535642.0001.101.

Young SWH (2014) A/B Testing Web Analytics Data [dataset]. Montana State University ScholarWorks. <http://doi.org/10.15788/m2rp42>

Appendix: Heat Maps



MONTANA
STATE UNIVERSITY

LIBRARY

Ask A Librarian

Inspiration, Discovery, Knowledge



CatSearch

Search for articles, books, and more

Advanced Search

SEARCH

FIND

Find research materials, including articles, books, databases, journals, and course reserves

REQUEST

Request resources and services, including group study rooms, laptops, documents, and books

INTERACT

Learn about the library and meet with us for research assistance, writing help, and tech support

News Hours Twitter

6/17 Library Workshop: Zotero Basics

6/17 Library Workshop: EndNote Web Basics

Tip of the Week: Get MS Office Help

New Database: Access World News





MONTANA
STATE UNIVERSITY

LIBRARY

Ask A Librarian 

Inspiration, Discovery, Knowledge



CatSearch

Search for articles, books, and more

SEARCH

[Advanced Search](#)

FIND

Find research materials, including articles, books, databases, journals, and course reserves

REQUEST

Request resources and services, including group study rooms, laptops, documents, and books

CONNECT

Learn about the library and meet with us for research assistance, writing help, and tech support

[News](#)

[Hours](#)

[Twitter](#)

6/17 Library Workshop: Zotero Basics

6/17 Library Workshop: EndNote Web Basics

Tip of the Week: Get MS Office Help

New Database: Access World News





MONTANA
STATE UNIVERSITY

LIBRARY

Ask A Librarian 

Inspiration, Discovery, Knowledge



CatSearch

Search for articles, books, and more

SEARCH

[Advanced Search](#)

FIND

Find research materials, including articles, books, databases, journals, and course reserves

REQUEST

Request resources and services, including group study rooms, laptops, documents, and books

LEARN

Learn about the library and meet with us for research assistance, writing help, and tech support

[News](#)

[Hours](#)

[Twitter](#)

6/17 Library Workshop: Zotero Basics

6/17 Library Workshop: EndNote Web Basics

Tip of the Week: Get MS Office Help

New Database: Access World News





MONTANA
STATE UNIVERSITY

LIBRARY

Ask A Librarian 

Inspiration, Discovery, Knowledge



CatSearch

Search for articles, books, and more

SEARCH

[Advanced Search](#)

FIND

Find research materials, including articles, books, databases, journals, and course reserves

REQUEST

Request resources and services, including group study rooms, laptops, documents, and books

HELP

Learn about the library and meet with us for research assistance, writing help, and tech support

[News](#)

[Hours](#)

[Twitter](#)

6/17 Library Workshop: Zotero Basics

6/17 Library Workshop: EndNote Web Basics

Tip of the Week: Get MS Office Help

New Database: Access World News





MONTANA
STATE UNIVERSITY

LIBRARY

Ask A Librarian 

Inspiration, Discovery, Knowledge



CatSearch

Search for articles, books, and more

SEARCH

Advanced Search

FIND

Find research materials, including articles, books, databases, journals, and course reserves

REQUEST

Request resources and services, including group study rooms, laptops, documents, and books

SERVICES

Learn about the library and meet with us for research assistance, writing help, and tech support

News

Hours

Twitter

6/17 Library Workshop: Zotero Basics

6/17 Library Workshop: EndNote Web Basics

Tip of the Week: Get MS Office Help

New Database: Access World News

