

insights__writeup

Benji Lu

December 13, 2015

Miscellaneous Insights

There's a lot to be learned from the data beyond the topic content analysis. For reference, here (again) is a glimpse of the data that we have on hand:

```
## Observations: 5,188
## Variables: 7
## $ id          (dbl) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ...
## $ title       (chr) "How to Run the World in 10 Easy Steps", "A Hea...
## $ content     (chr) "Every so often a serious but accessible intern...
## $ published_date (date) 2011-03-23, 2011-03-24, 2011-03-27, 2011-03-24...
## $ section_id  (int) 3, 4, 3, 3, 4, 3, 1, 4, 4, 4, 1, 2, 1, 2, 2, 2,...
## $ profile_id  (int) 2, 38, NA, 14, 20, 6, 27, 20, 28, 24, 158, 153,...
## $ author_name (fctr) Matt Wolfson, Tender Buttons, NA, Terri Martin...
```

With a little bit of wrangling and visualization, we were able to arrive at some interesting findings that might be of some value to the staff at *TSL*. For example, it's no secret that the newspaper struggles with retaining staff writers. But just how bad is it? By tracking the date of each writer's first published article and the writer's last published article, we were able to get a sense of how long they wrote for the paper.

```
# break down date published into component parts, determine semester published
articles <- articles %>%
  mutate(yearPublished = year(published_date)) %>%
  mutate(monthPublished = month(published_date)) %>%
  mutate(dayPublished = day(published_date)) %>%
  mutate(semester = ifelse(monthPublished < 6, 'spring', 'fall'))

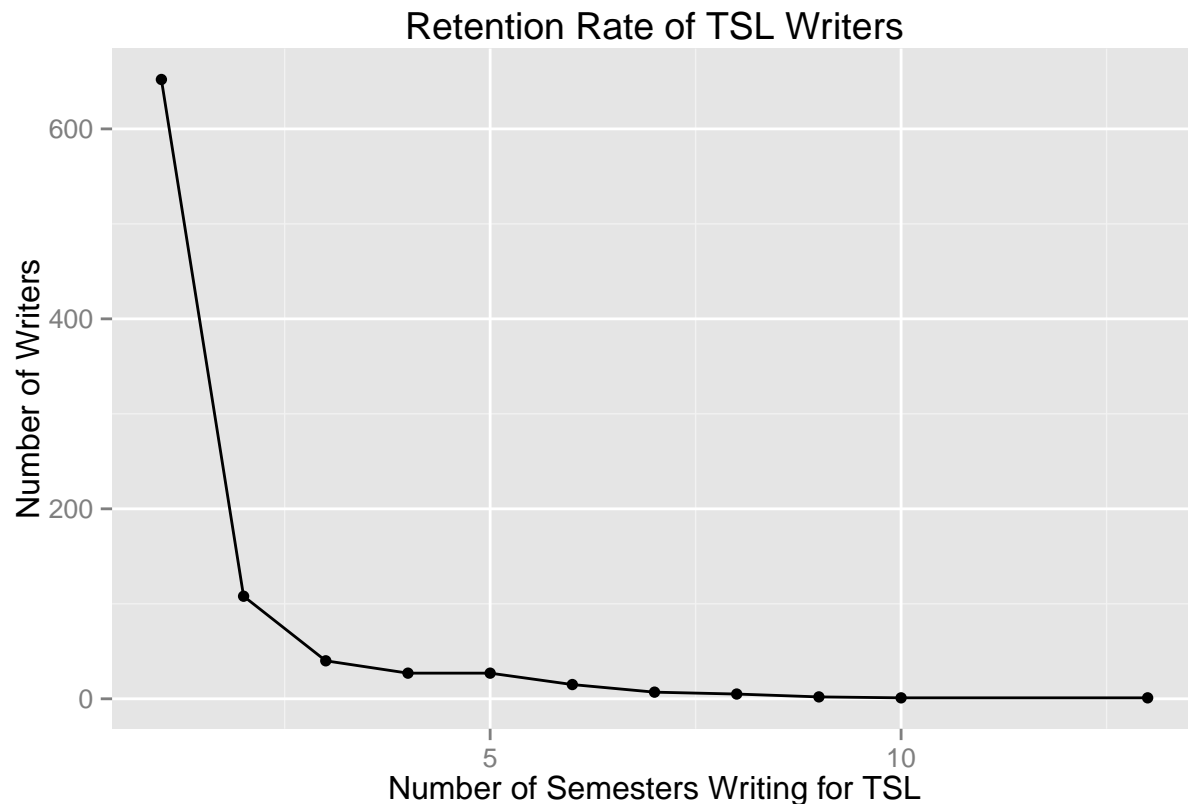
# calculate retention by semester for each writer
retention <- articles %>%
  group_by(profile_id) %>%
  mutate(maxYear = max(yearPublished),
         minYear = min(yearPublished),
         semesterOff = ifelse(maxYear==yearPublished & semester=='fall', 'fall', 'spring'),
         semesterJoin = ifelse(minYear==yearPublished &
                                semester=='spring', 'spring', 'fall'),
         semestersTotal = ifelse(semesterOff==semesterJoin, (maxYear-minYear)*2+1,
                                (maxYear-minYear)*2))

retention <- retention[!duplicated(retention[,6]),]
```

With the above, we were able to obtain the following graph:

```
retention %>%
  group_by(semestersTotal) %>%
  summarise(n = n()) %>%
  ggplot(aes(x = semestersTotal)) + geom_line(aes(y=n)) + geom_point(aes(y=n)) +
```

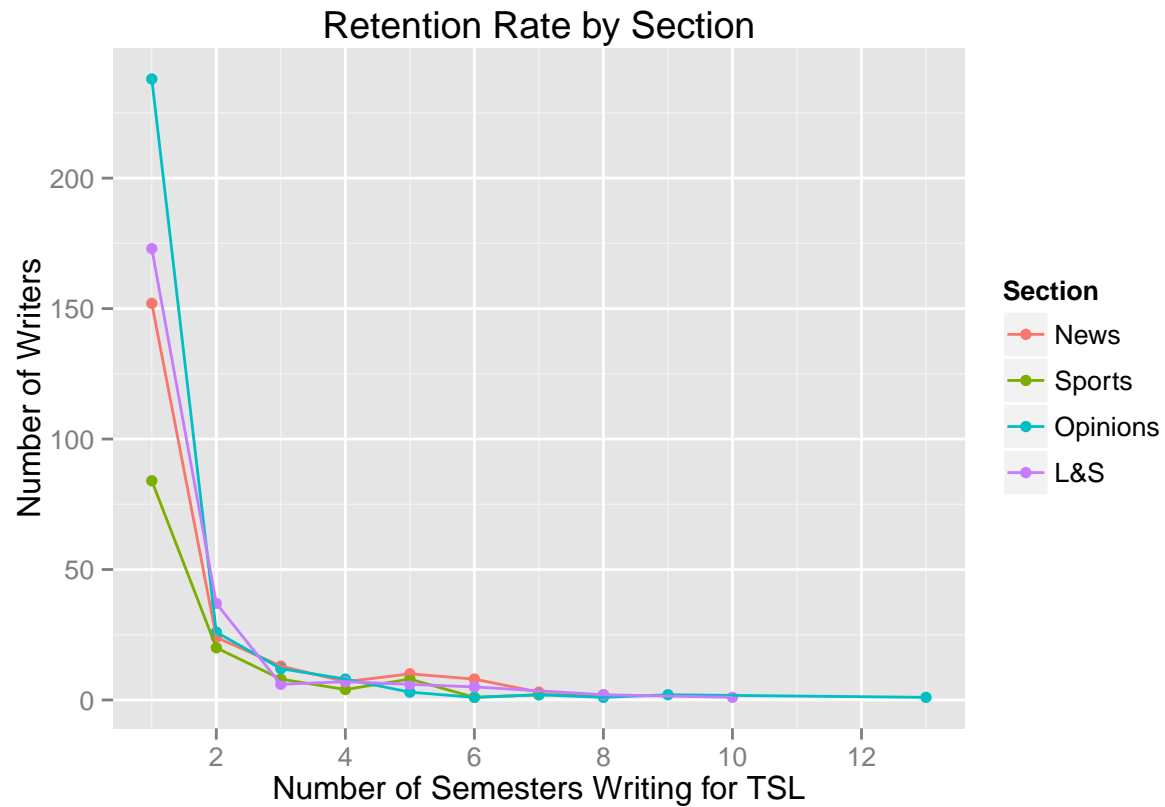
```
xlab('Number of Semesters Writing for TSL') + ylab('Number of Writers') +
ggtitle('Retention Rate of TSL Writers')
```



It seems like *TSL*'s reputation for shedding writers isn't exaggerated. Out of the pool of writers over the past five years, 652 writers stay for one semester or less; 108 writers stay for two semesters; 40 stay for three; and 85 stay for four or more semesters. There are a few caveats, though. First, since the database only includes articles published within the last five years, it's possible that some writers who were seniors when the data first began being collected had been writing for *TSL* for many semesters but are represented as only writing for one semester because their earlier articles aren't included in the database. This might lead to an overestimate of the number of writers who only stay for one or two semesters. Second, the method we applied above to calculate retention reports the difference between the first semester they began writing for *TSL* and the last semester they wrote for *TSL*. Some writers, however, take a semester or more off from writing for *TSL* to study abroad or simply pursue other interests before returning to write again. In these cases, the number of semesters they spent writing for *TSL* are overestimated. These cases probably consistute only a small minority of the data points, however, so the overall trend still holds.

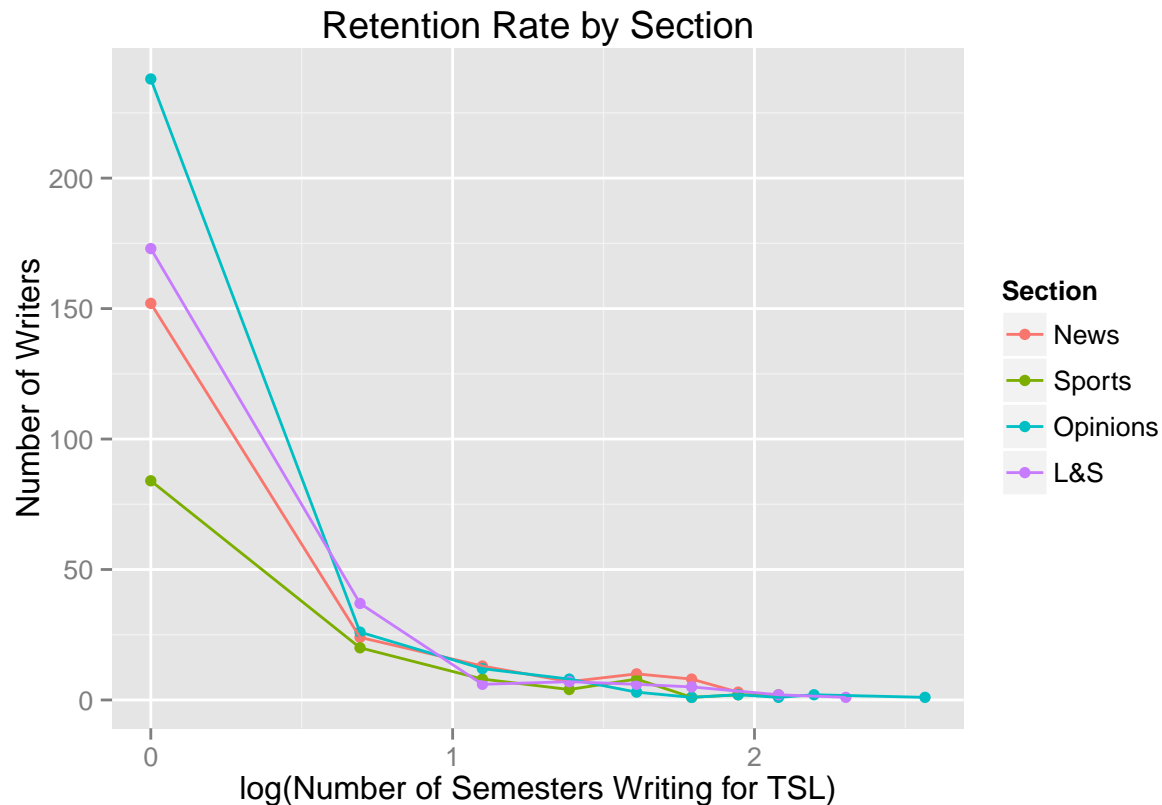
We can easily track the retention rate by section as well:

```
retention %>%
  filter(section_id != 5) %>%
  group_by(semestersTotal, section_id) %>%
  summarise(n = n()) %>%
  ggplot(aes(x = semestersTotal, y=n, color=factor(section_id))) + geom_point() +
  geom_line() + scale_color_discrete(name="Section", labels =
  c('News','Sports','Opinions','L&S')) + xlab('Number of Semesters Writing for TSL') +
  ylab('Number of Writers') + ggtitle('Retention Rate by Section') +
  scale_x_continuous(limits=c(1,13),breaks=seq(0,13,2))
```



Taking the log of the x-axis, we get:

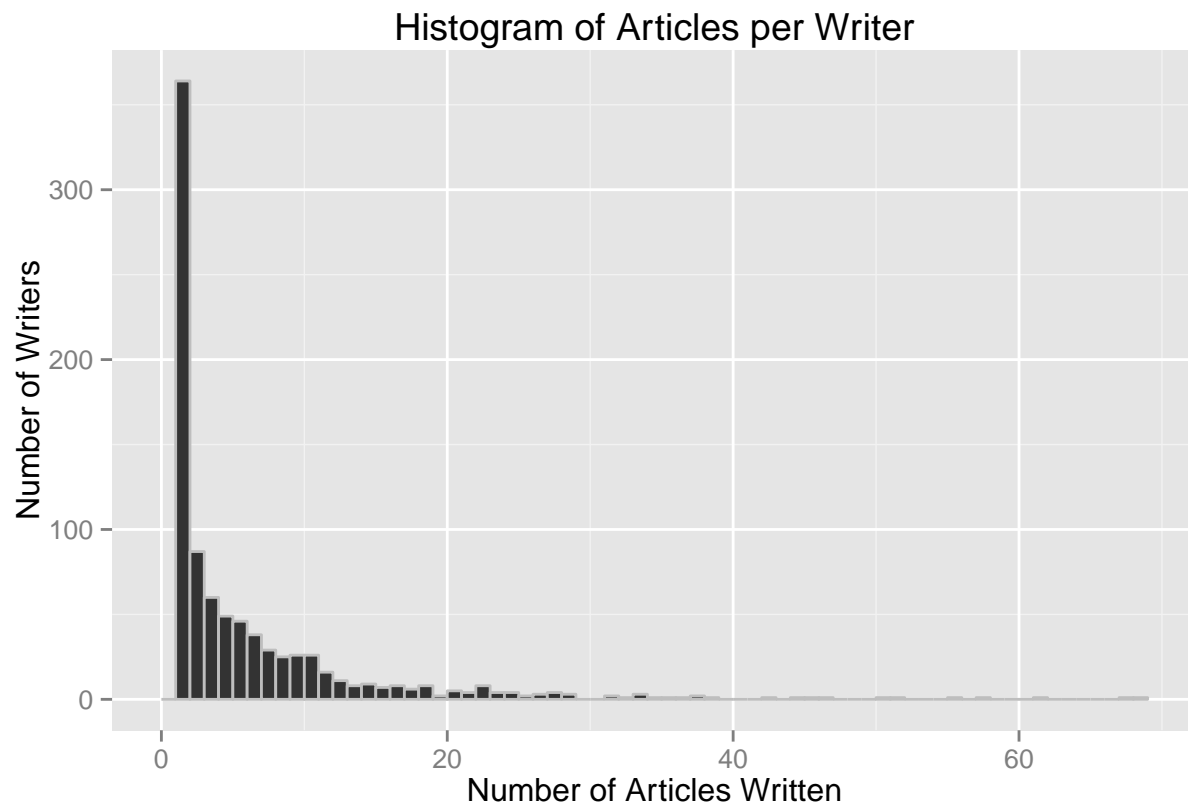
```
retention %>%
  filter(section_id != 5) %>%
  group_by(semesterTotal, section_id) %>%
  summarise(n = n()) %>%
  ggplot(aes(x = log(semesterTotal), y=n, color=factor(section_id))) + geom_point() +
  geom_line() + scale_color_discrete(name="Section", labels =
  c('News', 'Sports', 'Opinions', 'L&S')) +
  xlab('log(Number of Semesters Writing for TSL)') + ylab('Number of Writers') +
  ggtitle('Retention Rate by Section')
```



It seems like the opinions section suffers the steepest drop in returning writers. A lot of this may be attributable to the fact that anywhere from one-fourth to one-half of the articles in the opinions section each week are written by guest writers, who typically write only for a single, specific event or issue. Unfortunately, we can't filter those writers out given the data that we have.

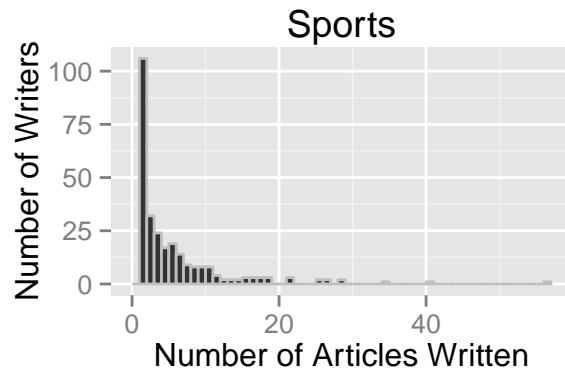
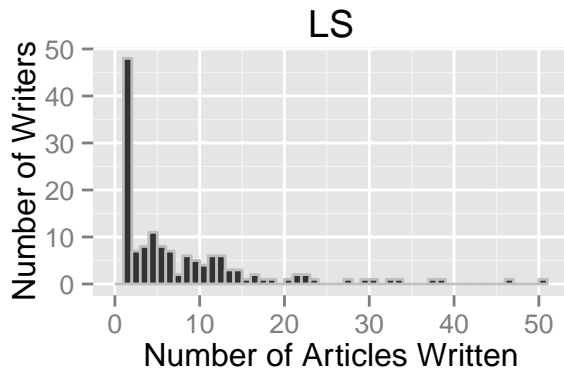
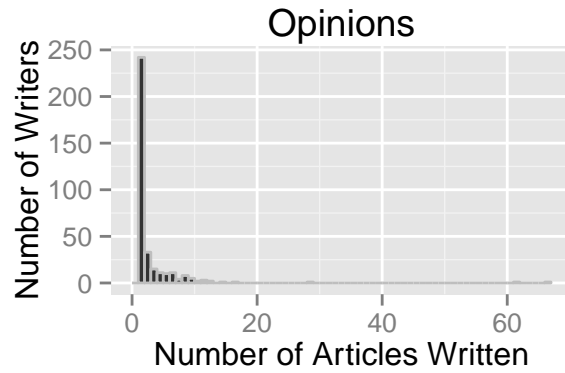
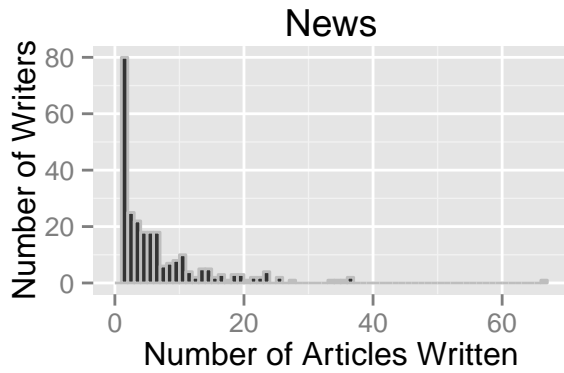
Another area of interest might be how prolific writers for *TSL* are—that is, how many articles each writer writes. It seems pretty reasonable to expect that trend to resemble the retention rates shown above. After all, a person can only write so many articles if they're only a writer for one semester. And indeed, we can see that this is the case:

```
articles_per_writer %>%
  ggplot(aes(x = total)) + geom_histogram(binwidth = 1, col='gray') + ggtitle('Histogram of Articles
```



We can also break it down by section:

```
sections = c('News', 'LS', 'Opinions', 'Sports')
for (num in 1:4) {
  section <- sections[num]
  assign(paste(section, 'section_articles_per_writer', sep='_'), articles %>%
    filter(section_id == num) %>%
    group_by(profile_id) %>%
    summarise(total = n()) %>%
    ggplot(aes(x = total)) + geom_histogram(binwidth = 1, col=I('gray')) + ggtitle(section) + xlab('Num')
}
grid.arrange(News_section_articles_per_writer, Opinions_section_articles_per_writer, LS_section_articles_per_writer, Sports_section_articles_per_writer)
```



It does seem, though, like there are a few especially prolific writers. We can isolate them to see who they are:

```
## [1] "Wes Haas "
```

	title	published_date
## 1	ASPC Moves to Paperless Elections; Vote Divided	2011-09-16
## 2	OCL Cuts RA Pay	2011-09-23
## 3	Senate Briefs 9/19/11	2011-09-23
## 4	HMC Student Named US Puzzle Champion	2011-09-30
## 5	Senate Briefs 9/26/11	2011-09-30
## 6	Food Handler Law Will Affect 5C Student Workers	2011-10-07
## 7	Senate Briefs 10/3/11	2011-10-07
## 8	Pomona Unveils \$600,000 Elemental Arts Initiative	2011-10-14
## 9	Senate Briefs 10/10/11	2011-10-14
## 10	ASPC Subcommittee Will Investigate WFJ Claims	2011-10-21

```
## [1] "Editorial Board"
```

	title
## 1	Pomona Should Renew Respect for Student Voices
## 2	Sponsor Program Should Take Time for Self-Reflection
## 3	Defending TSL's Recent Coverage
## 4	ASPC-Funded Groups Should be Allowed to Endorse Candidates
## 5	Students Overreach with Lear Petition
## 6	Newspapers Should Inform and Challenge
## 7	'Snapping' Masks Issues Worth Discussing
## 8	Policy Change Deserves Scrutiny
## 9	OCL Should Not Punish Students Retroactively

```

## 10                Pomona Should Respect SAC Alcohol Policy
##  published_date
## 1      2011-03-24
## 2      2011-04-07
## 3      2011-04-14
## 4      2011-04-22
## 5      2011-04-29
## 6      2011-09-16
## 7      2011-09-23
## 8      2011-09-30
## 9      2011-10-07
## 10     2011-10-14

```

```
## [1] "Editorial Board"
```

```

##                                     title
## 1      College Should Prioritize Academics Amid Budget Cuts
## 2      5-C Students Are Disenfranchised by the CUC Budget Process
## 3                                     March 27 Corrections
## 4 Pomona Should Reconsider How It Funds Work Study Opportunities
## 5                                     Response to Zahner on TSL
## 6                                     March 27 Corrections
## 7      Evaluating the Sponsor Program is No Easy Task
## 8      Stand with Staff Demands Should Be Improved
## 9      A New Semester, A New TSL
## 10     Corrections: October 2
##  published_date
## 1      2009-03-18
## 2      2009-04-10
## 3      2009-04-10
## 4      2009-04-17
## 5      2009-04-17
## 6      2009-04-17
## 7      2009-04-24
## 8      2009-05-01
## 9      2009-09-25
## 10     2009-10-01

```

```
## [1] "Tim Taylor"
```

```

##                                     title
## 1      Final Fantasy Goes Classical at UCLA
## 2      Competitive Gaming Evolves at the 5Cs
## 3      Gaming Gets Around
## 4      Video Game Column: Smashing It Up!
## 5      The Rise and Fall of the Exploration Platformer
## 6      iOS: Not the Next Game Boy
## 7 The Ocarina of Sublime: The Legend of Zelda 25th Anniversary Concert at the Pantages
## 8      Why the Gaming Industry Needs to Crash Again
## 9      Beleaguered Blue Blur: A Sonic the Hedgehog Retrospective
## 10     Video Game Buyer's Guide for the 2011 Holiday Season
##  published_date
## 1      2011-09-15

```

```
## 2      2011-09-23
## 3      2011-09-30
## 4      2011-10-06
## 5      2011-10-13
## 6      2011-10-21
## 7      2011-10-28
## 8      2011-11-04
## 9      2011-11-11
## 10     2011-11-18
```

```
## [1] "The Student Life Staff"
```

```
##                                     title
## 1  Get Your Fill (And Probably More) At Kickback Jack's
## 2      Bold Pomona Freshman Poops in Gender-Neutral Bathroom
## 3                                     Sagehen of the Week
## 4                                     Sagecock of the Week
## 5                                     Athena of the Week
## 6                                     Stag of the Week
## 7      Featured Matches of the Week
## 8                                     Athena of the Week
## 9                                     Stag of the Week
## 10                                    Sagecock of the Week
##      published_date
## 1      2009-03-18
## 2      2009-04-06
## 3      2009-04-10
## 4      2009-04-10
## 5      2009-04-10
## 6      2009-04-10
## 7      2009-04-10
## 8      2009-04-17
## 9      2009-04-17
## 10     2009-04-17
```

```
## [1] "Nachi Baru"
```

```
##                                     title
## 1      First Half Goal Sinks Men's Soccer
## 2      P-P Men's Soccer Falls to No. 15 UT Dallas
## 3      P-P Men's Soccer Defeats Struggling Oxy, Falls to Chapman
## 4      P-P Men's Soccer Overpowers Caltech 4-2
## 5      P-P Men's Soccer Rallies for Win Over Poets
## 6      Shekita Scores Lone Goal, Sagehens Shut Out Stags
## 7      Hens Split Weekend Games, Playoff Hopes Fizzle
## 8      Season Ends on Low Note, Sagehens Fall to Bulldogs
## 9      Sagehens Fall Short of Expectations, Look Ahead to Next Year
## 10 Moss Shines, Sagehens Tackle Nationally Ranked St. Thomas in Season Opener
##      published_date
## 1      2013-09-20
## 2      2013-09-27
## 3      2013-10-04
## 4      2013-10-11
```


## 5	2013-10-18
## 6	2013-10-25
## 7	2013-11-01
## 8	2013-11-08
## 9	2013-11-15
## 10	2013-11-22