LING/C SC/PSYC 438/538

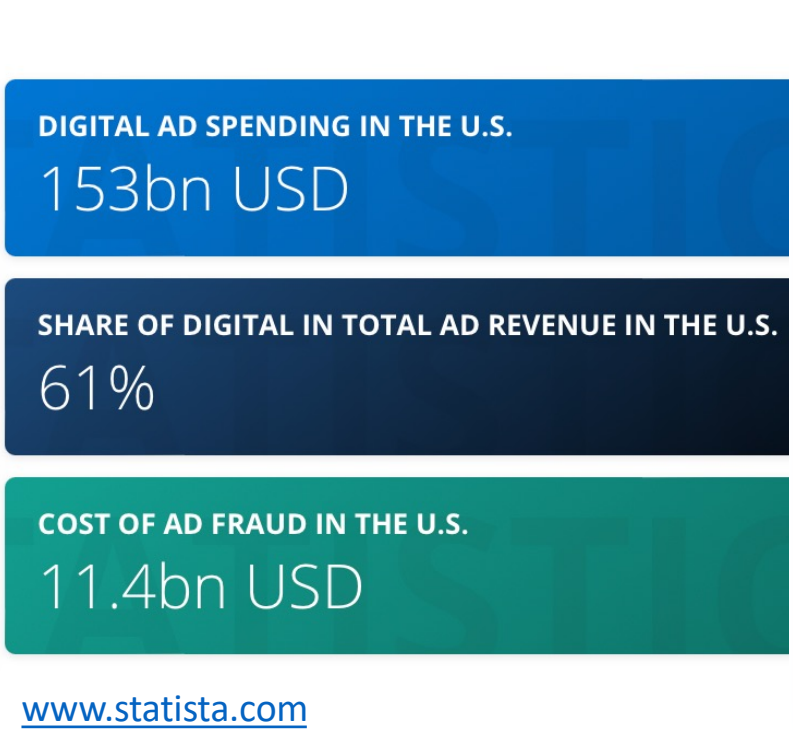
Lecture 9

Sandiway Fong

Today's Topics

- Homework 7: *a fun one*
- Odds and ends…
  - Perl references
  - Time and Date

Digital advertising



www.statista.com

SEARCH

Digital search ad spend
88.1bn USD

Digital search ad spend gr
12.2%

Bing ad revenue
11.59 bn USD

Google ad revenue
209bn USD

SOCIAL MEDIA

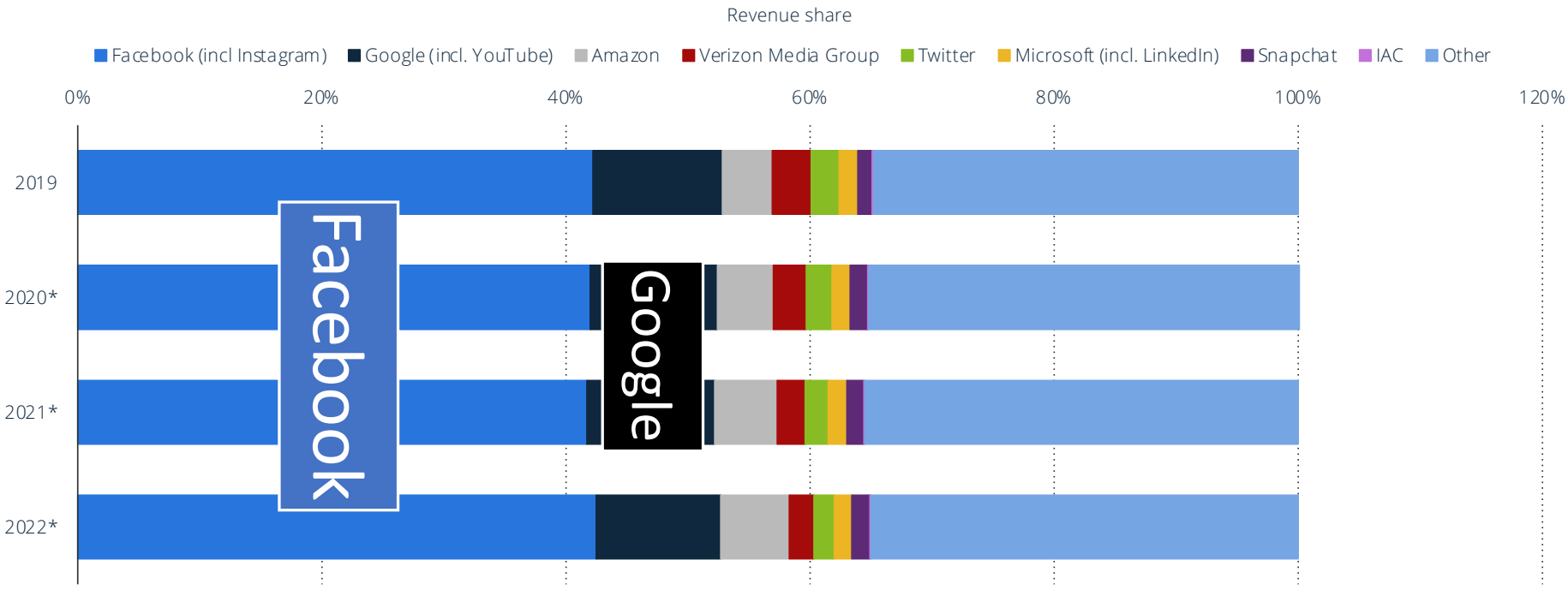
Social media ad spend
80.7bn USD

Instagram ad revenue
25.05bn USD

Snapchat ad revenue
1.82bn USD

Share of consumers whose purchasing decisions were influenced by social media
51%

Digital advertising: biggest sellers

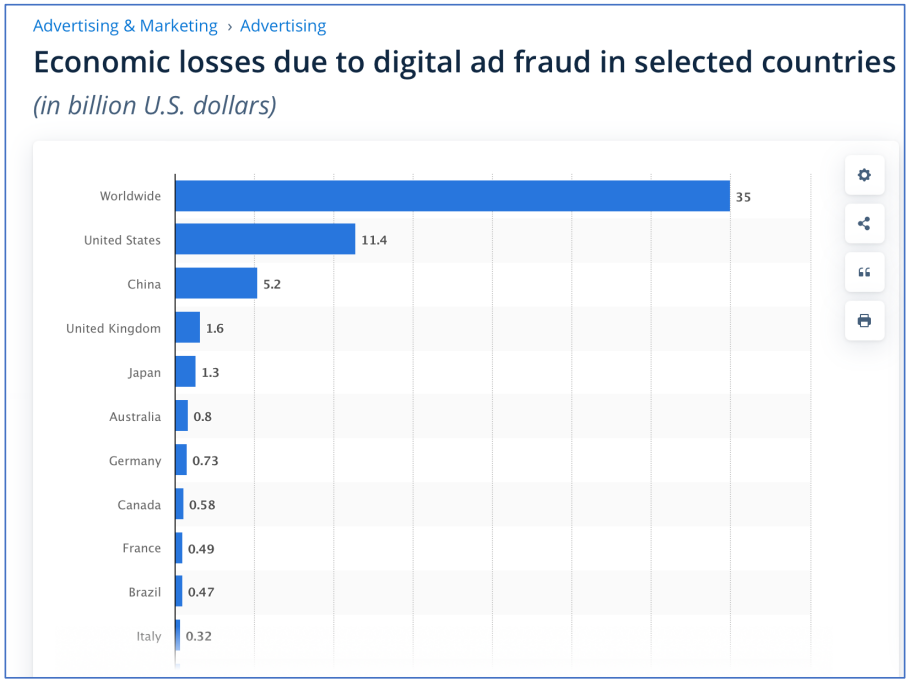


Note(s): United States; 2019

Further information regarding this statistic can be found on page 8.

Source(s): eMarketer; ID 237208

Digital advertising



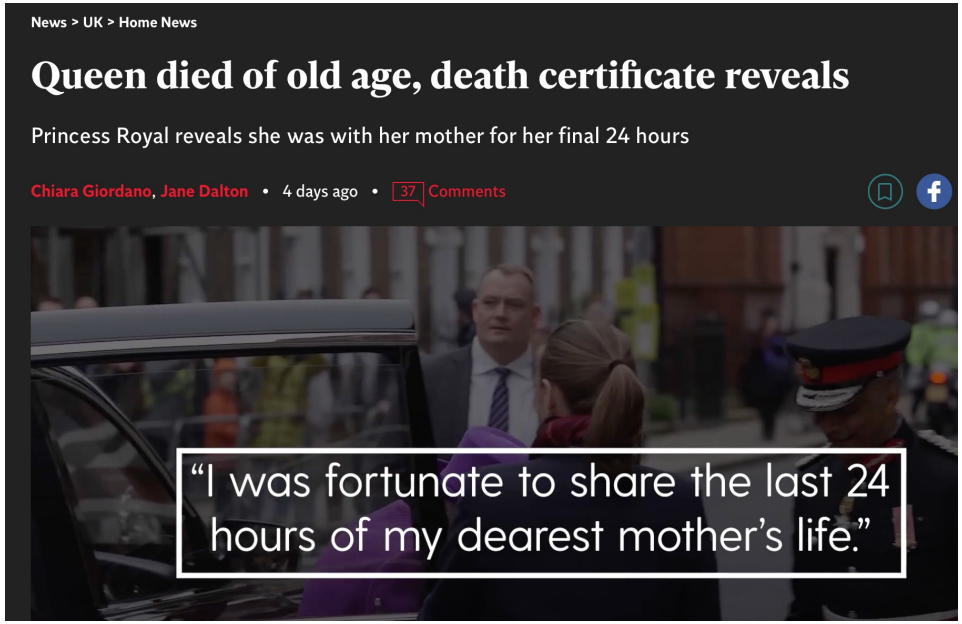
- digital ad fraud
  - click fraud
  - (bots)

Clickbait headlines: *all about the curiosity gap*

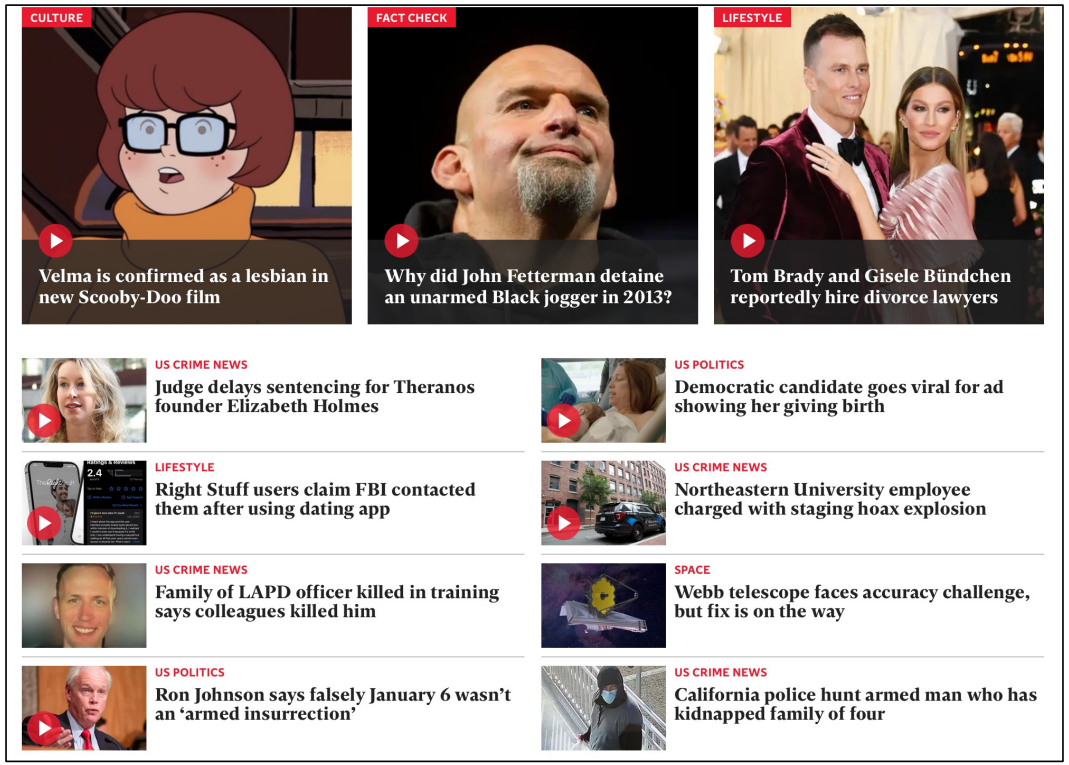
uses the word
revealed
without *revealing*
anything of note …

Clickbait headlines: *all about the curiosity gap*

- The Independent:
  - https://www.independent.co.uk/news/uk/home-news/queen-elizabeth-death-age-cause-b2178021.html

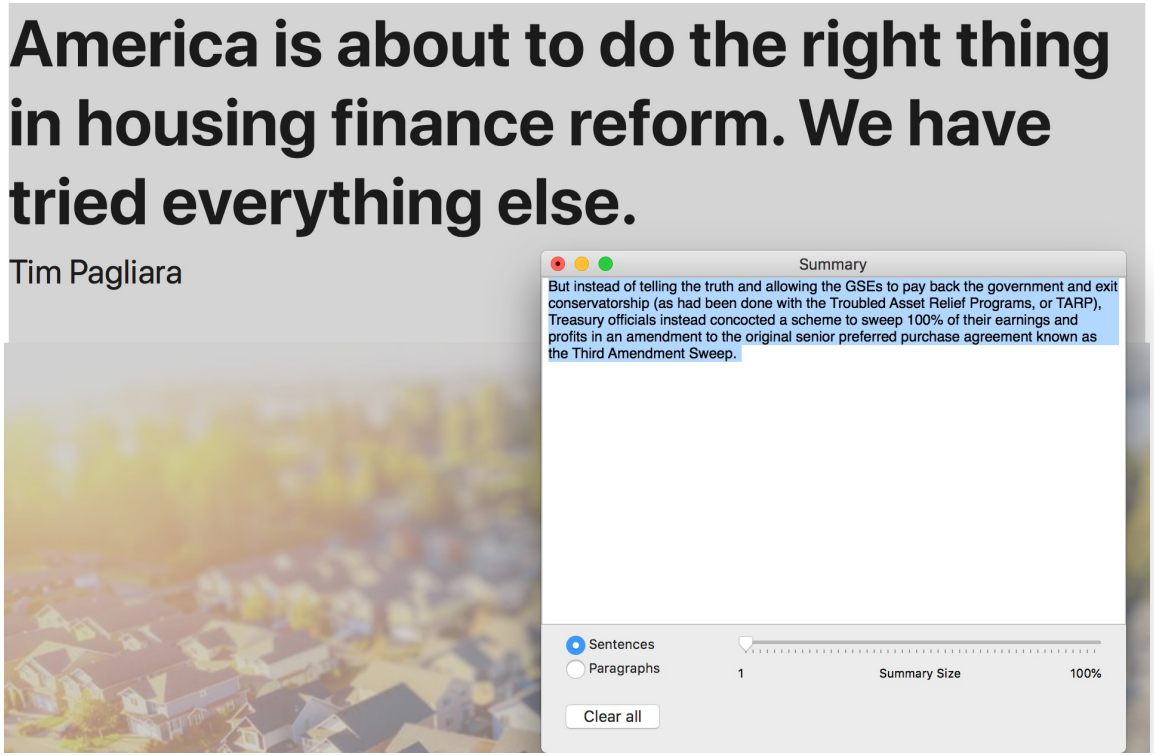


Generally not clickbait



same newspaper

Clickbait headlines: *all about the curiosity gap*



Summarization Ratio

5% 10% 20% 30% 40% 50% 60% 70% 80%

Language

en

Submit

During the time Calabria worked with Senator Richard Shelby on the Senate Banking Committee, they were guided by the belief that “well run, adequately capitalized, properly regulated financial institutions, do not fail.”

Despite President Trump’s executive order directing the Treasury to reprivatize the companies and end 11 years of U.S. government control and the emphatic smack down from the Fifth Circuit, there continues to be rogue arguments for the shut-down of the GSEs – more Jim Parrott.

Clickbait Example

- *Fannie Mae* and *Freddie Mac* are government-sponsored entities (GSEs).
- Fannie and Freddie don't make loans themselves. Instead, they … [buy] mortgages from lenders and [package] them into bonds that are sold to investors with guarantees of interest and principal.
- The GSEs are among the *most profitable companies* in the world.

https://www.americanbanker.com/articles/fannie-mae-freddie-mac-investors-fighting-profit-sweep-get-key-court-win

Clickbait Example

Easy answer

Top stories

Fannie Mae, Freddie Mac can keep their profits now

Curbed

22 hours ago

U.S. allows Fannie Mae, Freddie Mac to start keeping profits

Reuters

1 day ago

Fannie, Freddie to Retain Earnings

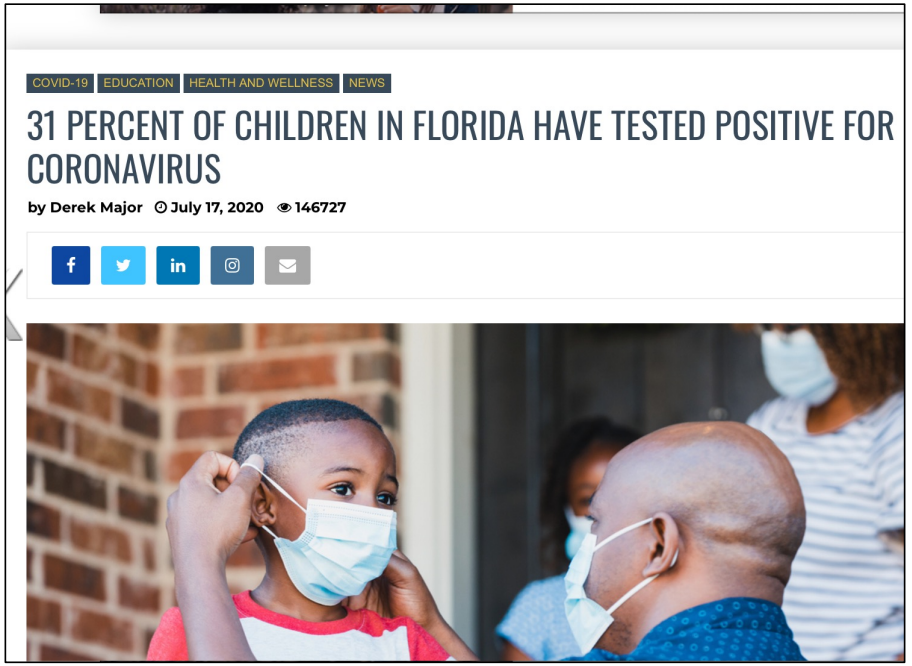
The Wall Street Journal

21 hours ago

Clickbait Examples

Others:

- *Before you do* X*, read this.*
- *You won't believe what* X *said* (*about Y*).
- *The real reason why …*
- *N ways to …*
- *A statistically sensationalized headline* ☞
- etc.



Lots of recent interest on spotting clickbait

2016

Clickbait Detection

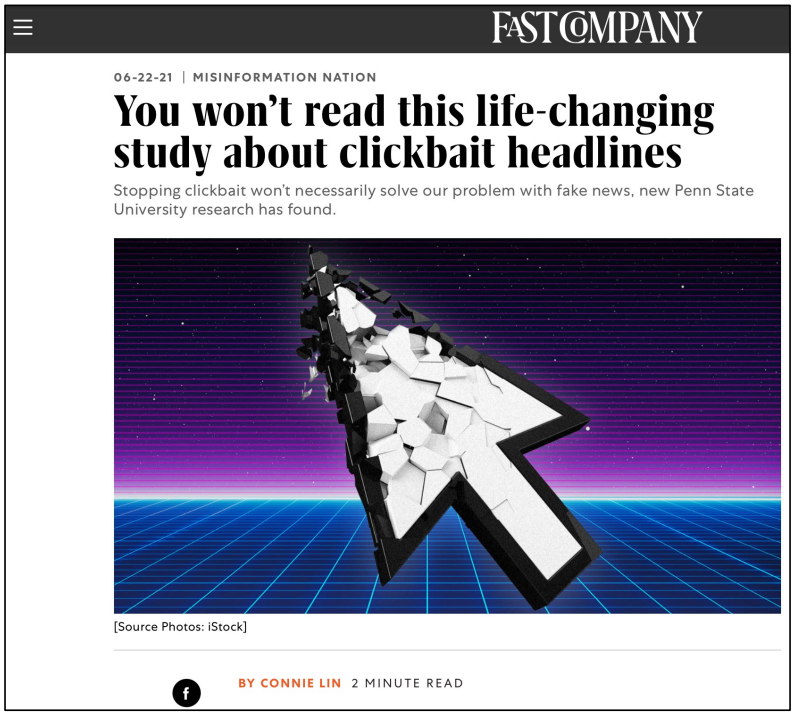
Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen

Bauhaus-Universität Weimar

<first name>.<last name>@uni-weimar.de

Abstract This paper proposes a new model for the detection of clickbait, i.e., short messages that lure readers to click a link. Clickbait is primarily used by online content publishers to increase their readership, whereas its automatic detection will give readers a way of filtering their news stream. We contribute by compiling the first clickbait corpus of 2992 Twitter tweets, 767 of which are clickbait, and, by developing a clickbait model based on 215 features that enables a random forest classifier to achieve 0.79 ROC-AUC at 0.76 precision and 0.76 recall.

- https://www.fastcompany.com/90649160/you-wont-read-this-life-changing-study-about-clickbait-headlines



FAST COMPANY

06-22-21 | MISINFORMATION NATION

You won't read this life-changing study about clickbait headlines

Stopping clickbait won't necessarily solve our problem with fake news, new Penn State University research has found.

[Source Photos: iStock]

BY CONNIE LIN 2 MINUTE READ

Homework 7

Go to https://news.yahoo.com or https://slate.com

- Find 3 clickbait headlines

Carefully:

1. Explain why they are clickbait headlines.
2. Figure out the crucial missing information from the article.
3. Can text summarization help bridge the gap?
   - (Use Open Text Summarizer tool or Summarize on your Mac or any other summarization tool you can find.)
4. Propose a way that a computer program can figure whether the article should be flagged as clickbait. Or argue that it can't be done.

- Usual rules: one PDF file. Subject: 438/538 Homework 7 *YOUR NAME*.
- Due date: Sunday midnight.

Python behavior

```
python
Python 3.9.12 (main, Jun  1 2022, 06:34:44)
[Clang 12.0.0 ] :: Anaconda, Inc. on darwin
Type "help", "copyright", "credits" or
"license" for more information.
>>> a = [1]
>>> b = a
>>> b[0] = -1
>>> b
[-1]
>>> a
[-1]
```

```
>>> a = 1
>>> b = a
>>> b = -1
>>> b
-1
>>> a
1
>>>
```

How do we reconcile these?

Perl: more complex data structures

- Arrays and hashes may only contain scalars
  - (Python is way better here)
- **Question**: How to accomplish nesting, i.e. put non-scalars inside?
- **Answer**: use references (called **pointers** in C), which happen to be scalars
- http://perldoc.perl.org/perlreftut.html

(actually a reference is just an unsigned number: a computer address)

Perl: References

- Two ways to make a reference:

Make Rule 1

If you put a `\` in front of a variable, you get a reference to that variable.

```
1.    $aref = \@array;        # $aref now holds a reference to @array
2.    $href = \%hash;         # $href now holds a reference to %hash
3.    $sref = \$scalar;       # $sref now holds a reference to $scalar
```

Make Rule 2

`[ ITEMS ]` makes a new, anonymous array, and returns a reference to that array. `{ ITEMS }` anonymous hash, and returns a reference to that hash.

```
1.    $aref = [ 1, "foo", undef, 13 ];
2.    # $aref now holds a reference to an array
3.
4.    $href = { APR => 4, AUG => 8 };
5.    # $href now holds a reference to a hash
```

Remember **bracketing** when initializing:

( ) List – used for both arrays and hashes

[ ] Reference to an array

{ } Reference to a hash

Perl: References

- Example: array of arrays

```
1.          @a = ( [1, 2, 3],
2.                 [4, 5, 6],
3.                 [7, 8, 9]
4.               );
```

Note: uses Make Rule 2: square brackets

- Let's figure out what the following mean:

```
$a[1],
${$a[1]}[1],
$a[1]->[0],
$a[1][2],
```

de-referencing arrow

Arrow Rule

In between two **subscripts**, the arrow is optional.

Instead of `$a[1]->[2]` , we can write `$a[1][2]` ;

Perl: References

- Looping (using **for/foreach**) with array/hash references:

```
1.          for my $element (@array) {
2.             ...
3.          }
```

o replace the array name, `@array` , with the reference:

```
1.          for my $element (@{$aref}) {
2.             ...
3.          }
```

`${$aref}[3]` is too hard to read, so you can write `$aref->[3]` instead.

`${$href}{red}` is too hard to read, so you can write `$href->{red}` instead.

Careful! `$aref->[3]` and `$aref[3]` are different

Perl: References

- Looping (using **for/foreach**) with array/hash references:

```
1.  for my $key (keys %hash) {
2.    print "$key => $hash{$key}\n";
3.  }
```

nd then replace the hash name with the reference:

```
1.  for my $key (keys %{$href}) {
2.    print "$key => ${$href}{$key}\n";
3.  }
```

`${$aref}[3]` is too hard to read, so you can write `$aref->[3]` instead.

`${$href}{red}` is too hard to read, so you can write `$href->{red}` instead.

Careful! `$href->{'red'}` and `$href{'red'}` are different.

Perl: References

- Perl code:

```
$a = [1, 2, 3, 4, 5];
print $a+1, "\n";
```

- What happens here?

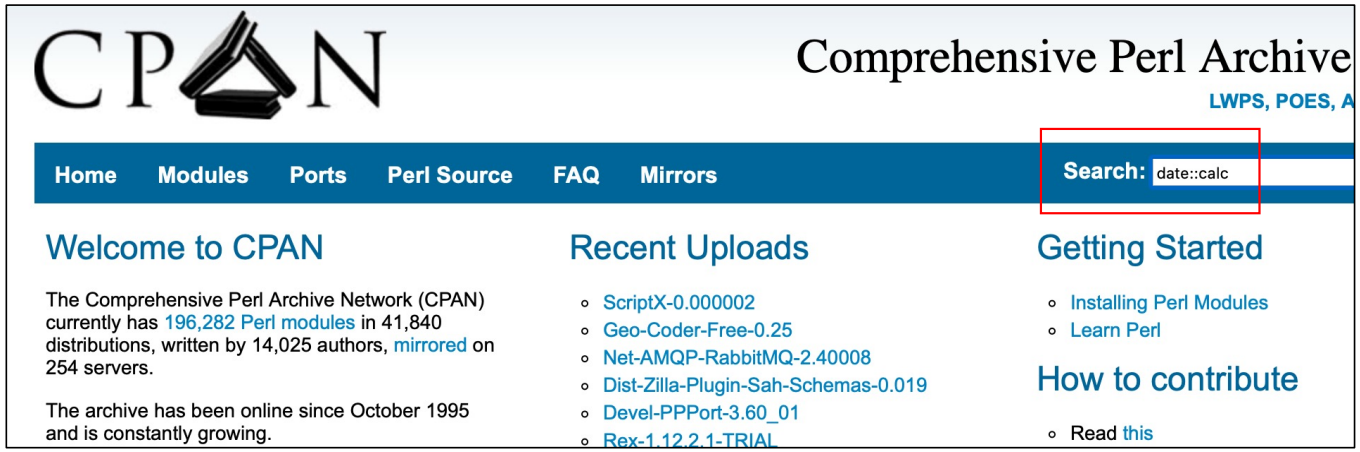
```
~$ perl -le '$a = [1, 2, 3, 4, 5]; print $a+1'
140353424343217
~$ perl -le '$a = [1, 2, 3, 4, 5]; print $a+1'
140552158855345
~$ perl -le '$a = [1, 2, 3, 4, 5]; print $a+1'
140576645202097
~$
```

Time and Date

```
1use Date::Calc qw(:all);¶
2die "usage: month day year\n" if $#ARGV != 2;¶
3($month, $day, $year) = @ARGV;¶
4¶
5$dow = Day_of_Week_to_Text(Day_of_Week($year,$month,$day));¶
6print "$month/$day/$year falls on a $dow\n";¶
7¶
8($year, $month, $day) = Today();¶
9print "$month/$day/$year is today\n";¶
```

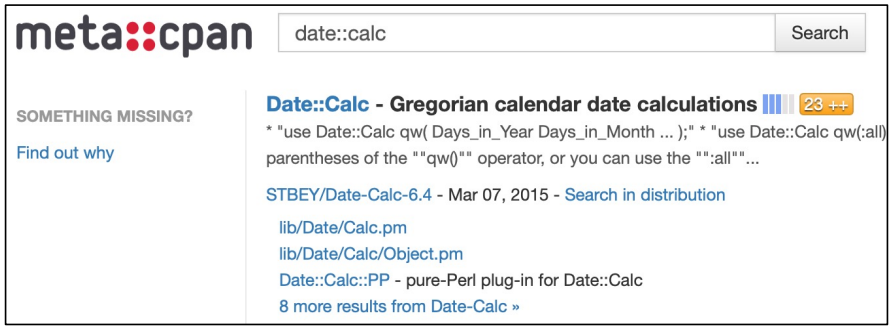
```
(base) ling538-22$ perl dow.perl
usage: month day year
(base) ling538-22$ perl dow.perl 10 5 2022
10/5/2022 falls on a Wednesday
10/4/2022 is today
(base) ling538-22$
```

CPAN date::calc



- *Looking for* date::calc
- https://www.cpan.org/modules/INSTALL.html
  - can use command cpan on the command line
  - see also alternatives

CPAN date::calc



SYNOPSIS

```
use Date::Calc qw(
    Days_in_Year
    Days_in_Month
    Weeks_in_Year
    leap_year
    check_date
    check_time
    check_business_date
    Day_of_Year
    Date_to_Days
    Day_of_Week
    Week_Number
    Week_of_Year
    Monday_of_Week
    Nth_Weekday_of_Month_Year
    Standard_to_Business
    Business_to_Standard
    System_Clock
    Today
    Now
    Today_and_Now
    This_Year
    Gmtime
    Localtime
    Mktime
    Timezone
    Date_to_Time
    Time_to_Date
    Easter_Sunday
    Decode_Month
    Decode_Day_of_Week
    Decode_Language
```

Time and Date

```
sudo cpan Date::Calc

Password:

Loading internal logger. Log::Log4perl recommended for better logging

Reading '/Users/sandiway/.cpan/Metadata'

  Database was generated on Tue, 07 Sep 2021 20:41:03 GMT

Running install for module 'Date::Calc'

Fetching with HTTP::Tiny:

http://www.cpan.org/authors/id/S/ST/STBEY/Date-Calc-6.4.tar.gz

Warning (usually harmless): 'YAML' not installed, cannot parse
'/Users/sandiway/.cpan/FTPstats.yml'

Fetching with HTTP::Tiny:

http://www.cpan.org/authors/id/S/ST/STBEY/CHECKSUMS

Checksum for /Users/sandiway/.cpan/sources/authors/id/S/ST/STBEY/Date-
Calc-6.4.tar.gz ok

'YAML' not installed, will not store persistent state

Configuring S/ST/STBEY/Date-Calc-6.4.tar.gz with Makefile.PL

…
```

```
 All tests successful.

Files=51, Tests=3381,  1 wallclock secs ( 0.16 usr  0.06 sys +  0.72
cusr  0.16 csys =  1.10 CPU)

Result: PASS

STBEY/Date-Calc-XS-6.4.tar.gz

  /usr/bin/make test -- OK

Running make install

"/opt/local/bin/perl5.28" -MExtUtils::Command::MM -e 'cp_nonempty' -- XS.bs
blib/arch/auto/Date/Calc/XS/XS.bs 644

Manifying 1 pod document

Files found in blib/arch: installing files in blib/lib into architecture
dependent library tree

Installing /opt/local/lib/perl5/site_perl/5.28/darwin-thread-multi-
2level/auto/Date/Calc/XS/XS.bundle

Installing /opt/local/lib/perl5/site_perl/5.28/darwin-thread-multi-
2level/Date/Calc/XS.pod

Installing /opt/local/lib/perl5/site_perl/5.28/darwin-thread-multi-
2level/Date/Calc/XS.pm

Installing /opt/local/share/perl5.28/siteman/man3/Date::Calc::XS.3pm

Appending installation info to /opt/local/lib/perl5/5.28/darwin-thread-
multi-2level/perllocal.pod

  STBEY/Date-Calc-XS-6.4.tar.gz

  /usr/bin/make install  -- OK
```

Time and Date

- Perl function time:

time

Returns the number of non-leap seconds since whatever time the system considers to be the epoch, and `localtime`. On most systems the epoch is 00:00:00 UTC, January 1, 1970; a prominent except

```
perl –le 'print time'
1601523969
```

- To pause for *n* seconds, use `sleep` *n*
  - `perl –le '$t1 = time; sleep 10; $t2 = time; print $t2–$t1'`
  - `10`

Time and Date

- Perl function localtime:

localtime EXPR

localtime

Converts a time as returned by the time function to a 9-element list with the time analyzed for the local time zone. T

```
$perl -le 'print localtime'
32582030812032730
$perl -le '$x = localtime; print $x'
Wed Sep 30 20:58:43 2020
$perl -le '@a = localtime; print "@a"'
47 58 20 30 8 120 3 273 0
$
```

```
#     0    1    2     3     4    5     6     7     8
my ($sec,$min,$hour,$mday,$mon,$year,$wday,$yday,$isdst) =
                                1900          localtime(time);
```

Note: $mon, $wday (beginning on Sunday) and $yday indexed from 0

Time and Date

Reference:
https://docs.python.org/3/library/datetime.html#strftime-and-strptime-format-codes

```
python3
Python 3.9.12 (main, Jun  1 2022, 06:34:44)
[Clang 12.0.0 ] :: Anaconda, Inc. on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> import datetime
>>> d = datetime.date(2022,10,5)
>>> d
datetime.date(2022, 10, 5)
>>> print(d.strftime("%A"))
Wednesday
>>> print(d.strftime("%B %-d %Y"))
October 5 2022
>>> datetime.datetime.now()
datetime.datetime(2022, 10, 4, 22, 1, 22, 576812)
>>>
```

Directive	Meaning	Example
%a	Weekday as locale's abbreviated name.	Sun, Mon, …, Sat (en_US); So, Mo, …, Sa (de_DE)
%A	Weekday as locale's full name.	Sunday, Monday, …, Saturday (en_US); Sonntag, Montag, …, Samstag (de_DE)
%w	Weekday as a decimal number, where 0 is Sunday and 6 is Saturday.	0, 1, …, 6
%d	Day of the month as a zero-padded decimal number.	01, 02, …, 31
%b	Month as locale's abbreviated name.	Jan, Feb, …, Dec (en_US); Jan, Feb, …, Dez (de_DE)
%B	Month as locale's full name.	January, February, …, December (en_US); Januar, Februar, …, Dezember (de_DE)