The Question

I want to create a query to view the popularity of a programming language over time. I originally wanted to see how long a user has a book checked out, possibly how long a book sits on a shelf vs being checked out. But first I have to start smaller. Select the number of books with a title containing a programming language, and count the number of checkouts by year.

The Approach

First I planned to search titles with a key word. I tried to reduce noise by adding spaces between the word and the wildcard character. This didn't help me with titles like *"A B C of Cooking"* or *"Ford 170 C Motor."* Another problem is common words, like python snakes.

The Dewey Decimal Class system made it much easier to search and remove a lot of static and noise. Targeting only books under the computer science category quickly narrowed down the results.

Count the number of checkouts for programming languages.

```
SELECT
sum(CASE WHEN title LIKE "%something%" then 1 else 0 end) as counter
FROM spl2.inraw WHERE (deweyClass>0 AND deweyClass<=6) AND YEAR(cout) = '2007'
```

I repeated this query multiple times, only increasing the cout year to the next, up until 2013. I noticed a slight downward trend.

I created another query with languages associated more with web technologies. Next I want to investigate how many unique titles a language has, and how many copies of those titles.

The Query

```
SELECT
sum(CASE WHEN title LIKE "%HTML%" then 1 else 0 end) as HTML,
sum(CASE WHEN title LIKE "%CSS%" then 1 else 0 end) as CSS,
sum(CASE WHEN title LIKE "%JavaScript%" then 1 else 0 end) as JS,
sum(CASE WHEN title LIKE "%SQL%" then 1 else 0 end) as sSQL,
sum(CASE WHEN title LIKE "%PHP%" then 1 else 0 end) as PHP
FROM spl2.inraw WHERE (deweyClass>0 AND deweyClass<=6) AND YEAR(cout) = '2013'
```
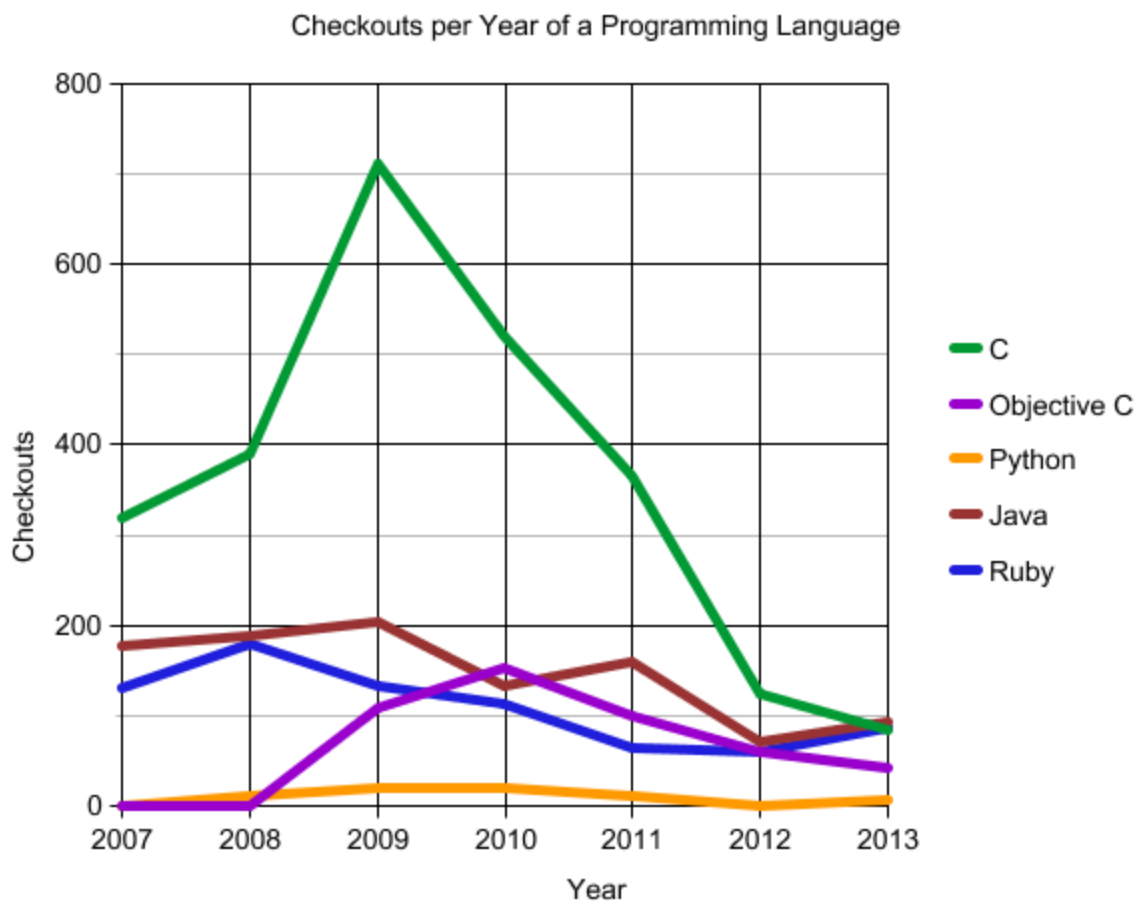
Again, changing the year with each submission. These queries took on average of 72 seconds.
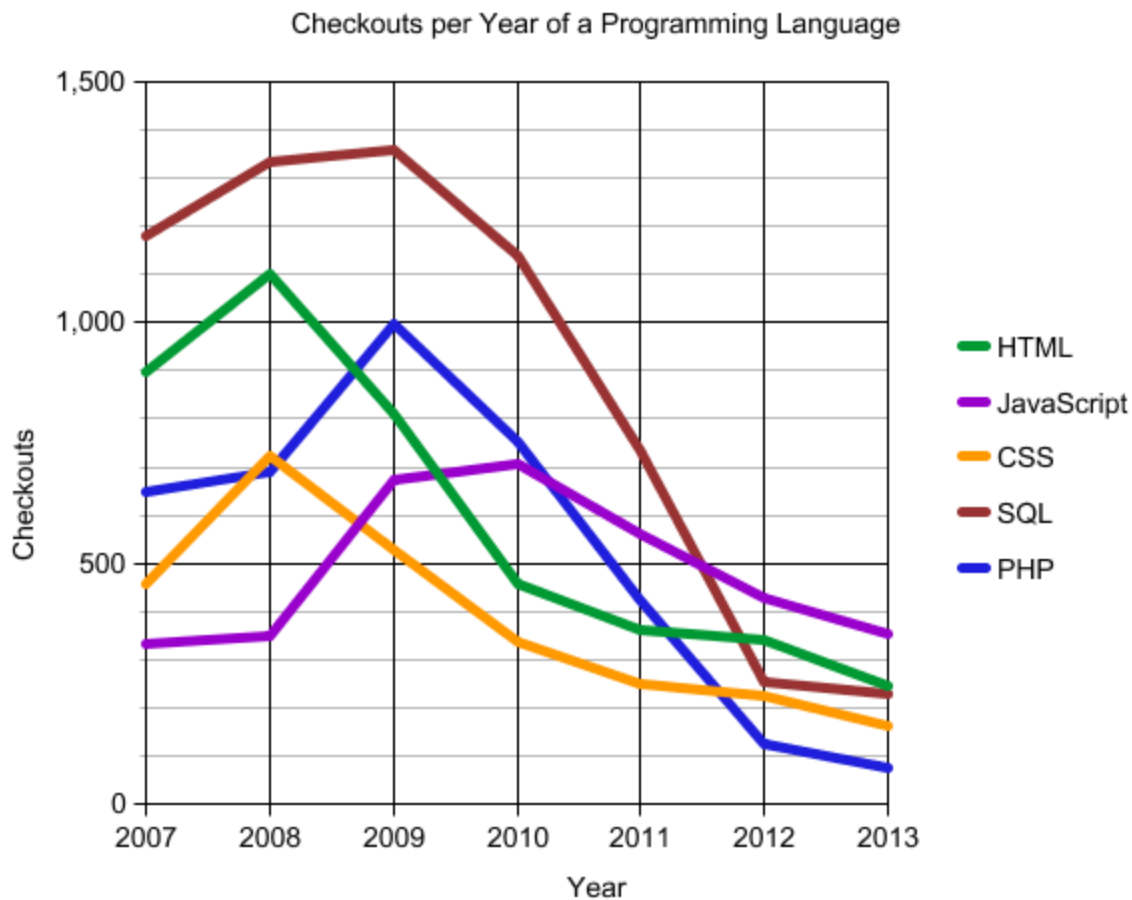
Next I want to investigate how many unique titles a language has, and how many copies of those titles.

```
SELECT COUNT(*) AS titleCount, title, itemtype, deweyClass
FROM spl2.inraw
WHERE title LIKE "% Ruby %" AND (deweyClass>0 AND deweyClass<=6) GROUP BY title
```

These queries took much longer, averaging to 278 seconds. I attempted to make the count function in its own query to increase speed, but the time remained constant. The number of rows returned tells us how many distinct books there are and the titleCount tells us how many copies of that book exists.

The Results

Checkouts per Year of a Programming Language

Checkouts per Year of a Programming Language

The lower chart of programming languages has a higher total count of checkouts per year and steeper decline because of the relation between these languages. Languages like HTML and CSS are much closer intertwined than Java and Ruby. Some titles would be counted twice because of this.
"web publishing with html and css"
"html to javascript"
"Sams teach yourself PHP mySQL and Apache all at once"

More data was collected from the additional queries
HTML 57 books
CSS 14 books
JavaScript 28 books
SQL 60 books
PHP 22 books

[More data about number of copies contained in the CSV files]

Comments and Analysis

I was expecting to see a rise, not a decline. The sudden drop around 2009 confuses me. My guess is that more people are using online resources to learn these languages. Watching videos online instead of checking out books and dvds.

JavaScript and Objective C are the only languages to have more checkouts in 2013 than 2007.