Learn about sitemaps

- Sitemaps are files that provide info about pages, videos, or other info on a site as well as their relationships
- Google's crawler reads file to "crawl" the site more efficiently
- Tells Google which pages or files are relevant to it (e.g., last update of page, alternate language versions, etc.)
- Can provide info about specific types of content (e.g., images, news, etc.)
  - Video entry can specify runtime, category, age rating
  - Image entry could include image subject, type, license
  - News entry could include title, publication date

Why is it necessary?

- If pages are properly linked (i.e., "important" pages can be accessed through a form of navigation, whether through menus or links), Google can typically crawl most of it. However, a sitemap can potentially greatly improve this for larger, complex sites and/or more specialized files.*
  - *It does not necessarily guarantee all items will be crawled and indexed, but will almost always benefit from having it

Use Cases
1. Large sites: New or recently updated data may be overlooked
2. Large archive of content pages that are isolated or improperly/not well linked: If not properly referencing one another, can be listed in a sitemap
3. New sites with few external links: Googlebot and other crawlers function by following links page to page, so if not linked by a parent, may be overlooked

*May not be needed if the site is comprehensively internally linked, meaning that Google can find all important pages through links starting from the homepage.

Building a sitemap

A basic XML sitemap example with a single URL location:

```
<?xml version="1.0" encoding="UTF-8"?>
    <urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
        <url>
                <loc>http://www.example.com/foo.html</loc>
                <lastmod>2018-06-04</lastmod>
        </url>
    </urlset>
```

Some more complex examples.

*Should follow standard sitemap protocol.
- All formats limit a single sitemap to 50MB uncompressed and 50k URLs

- Larger files will have to be broken into multiple maps or by creating a sitemap index, which points to a list of sitemaps