

# Advanced Web Development and Web Scraping

---

Dr. Garrett Dancik

# Why learn “advanced web development”

- You can develop web pages (obviously)
- You can add features to web pages (i.e., develop browser extensions)
- You can answer questions by extracting and parsing (scraping) web page data
- You can retrieve information and answer questions by using web page application program interfaces (APIs)
- You can automate web-tasks such as logging onto a web site

# What did you guys do?

- Entertainment (video and music)
  - Streaming on Twitch since COVID-19
  - Shiny application: trending YouTube videos in the U.S
  - Tool for gathering statistics from Twitch & YouTube
  - Quarantine tunes
  - Discover Playlists: Spotify Playlist Creator
  - Artist Information Lookup
- Medical / biological / ecological applications
  - COVID-19 analysis
  - kNN / skin lesion prediction
  - A Chrome extension for improving acronym searches on PubMed
  - Comparing Country Populations
- Entertainment
  - Randomized Profile Generator

- Color legend
  - Browser extension
  - Shiny application
  - Web scraper

# What did you guys do?

- Productivity and shopping
  - Slightly more ethical alternatives to Amazon and its affiliates
  - Time tracker and site blocker
- Information access
  - Book Recommendations from The New York Times Best Seller List
  - Super ultra speed-a-tron Wikipedia quickenator 3001
  - Soccer League updates
  - Dog breed search
  - UFO Sightings in the US

- Color legend
  - Browser extension
  - Shiny application
  - Web scraper

# Other examples

- Identification of bots on Twitter:
  - <http://www.pewresearch.org/fact-tank/2018/04/19/qa-how-pew-research-center-identified-bots-on-twitter/>
- HiPub Chrome extension for knowledge discovery:
  - <https://www.ncbi.nlm.nih.gov/pubmed/27485446>
- The Cambridge Carcinoma of the Prostate App
  - <https://bioinformatics.cruk.cam.ac.uk/apps/camcAPP/>
- *shinyGEO*: analysis of gene expression data
  - <https://gdancik.github.io/shinyGEO/>
- CongressEdits: Tracks changes to Wikipedia made from IP addresses from the U.S. Congress
  - <https://source.opennews.org/articles/automating-transparency/>
- COVID-19 in CT
  - <https://gdancik.shinyapps.io/echda/>