1. Project Overview

- **Client**
  - Dr. James D. Ivory
  - Virginia Tech Gaming and Media Effects Lab

- **Problem:** Unknown data on extremist behavior in online game communities

- **Deliverables**
  - Forum scraper
  - Cleaned data and analysis
  - Visualizations
2. Timeline: February

- Research scraping
- Develop blog scraper
- Understand/Modify previous scraper
- Determine media to scrape
- Test and begin scraping
- Report #1
Timeline: March

- Presentation #2
- Finish scraping
- Create storage of clean dataset
- Begin cleaning data
- Determine what we want to visualize
Timeline: April

- Finish cleaning data
- Sentiment analysis with client
- Finalize visualization
- Begin/finish final report
- Final presentation
Timeline: May

- Report #2
- VTechWorks Submission
3. Deliverables

- YouTube Scraper Script
- Forum Scraper Script
- Raw Data
- Cleaned Data
- Visualizations
4. YouTube Scraper Script

```javascript
function YoutubeScraper(iterations, search.query, activeSheet, offset)
    var next = null
    const start_date = new Date("January 1, 2013 00:00:00 -0000");
    for (let i = 0; i < iterations; i++) {
        var search = YouTube.Search.list("snippet", id, q: search.query, maxResults: 5, relevanceLanguage: "en", regionCode: "US", startDate: start_date, endDate: end_date);
        else {
            var search = YouTube.Search.list("snippet", id, q: search.query, maxResults: 5, pageToken: next, relevanceLanguage: "en", regionCode: "US", startDate: start_date, endDate: end_date);
        }
        var results = search.items.map(item => [item.id, item.snippet.title, item.snippet.publishedAt]);
        next = search.nextPageToken
        var ids = results.map([id] => id);
        var stats = YouTube.VideoList.list('statistics', [id: ids]);
        var videoStats = stats.items.map(item => [item.statistics.viewCount, item.statistics.likeCount, item.statistics.commentCount]);
        activeSheet.getRange(1 + offset, 1).setValues(results);
        activeSheet.getRange(1 + offset, 4, videoStats.length, videoStats[0].length).setValues(videoStats)
    }
}
```

```javascript
function scrapeComments(x, y)
    var ss = SpreadsheetApp.getActiveSpreadsheet();
    var result = [{Name: 'Comment', 'Time', 'Likes', 'Reply Count', 'Reply Author', 'Reply', 'Published', 'Updated'}];
    var vid = ss.getSheets()[0].getRange(x, y).getValue();
    var nextPageToken = undefined;
    for (let i = 0; i < 10; i++) {
        var data = YouTube.CommentThreads.list('snippet', {videoId: vid, maxResults: 100, pageToken: nextPageToken})
        nextPageToken = data.nextPageToken;
        for (var row = 0; row < data.items.length; row++)
            result.push([data.items[row].snippet.topLevelComment.snippet.authorDisplayName, data.items[row].snippet.topLevelComment.snippet.textDisplay, data.items[row].snippet.topLevelComment.snippet.publishedAt, data.items[row].snippet.topLevelComment.snippet.likeCount, data.items[row].snippet.topLevelComment.likeCount, data.items[row].snippet.topLevelComment.replyCount, 'null', 'null', 'null']);
        console.log(vid + x + y + PublishedAt);
        //console.log(nextPageToken);
    }
```
5. Forum Scraper Script

- Files
  - category-thread-id-scraper.py
  - post-scraper.py
  - randomizeThreads.py
5. Forum Scraper Script cont.

- Getting Categories
- Getting Threads
- Randomizing Threads
- Scrape Posts
6. Raw Data

- Forum Data
  - SQLite3 database
- YouTube Data
  - CSV file
  - SQLite3 database
6. Cleaned Data

- CSV files
  - YouTube
  - Forum
7. Visualizations

- War Thunder Forum Data
- YouTube Data
- Previous Team’s Dictionary
War Thunder Forum Visualizations
Document Leak

Eurofighter Typhoon DA7 fighter jet

Lockheed F-117 Nighthawk stealth attack aircraft

Boeing AH-64D Apache helicopter

Norinco VT-4 tank

M2A2 Bradley IFV

Number of posts on each respective leak date
Our scraped data from the War Thunder Forms

- August 31, 2023: 238 posts
- September 12, 2023: 85 posts
- September 15, 2023: 124 posts
- December 11, 2023: 377 posts
- December 12, 2023: 410 posts

Number of posts
YouTube Visualizations
Comparing word frequency

Forum Word Cloud

YouTube Word Cloud
8. Future Work

● Scrape Old War Thunder Forum
9. Acknowledgements

- **Client:**
  - Dr. James D. Ivory
  - Michael Senters (Grad Researcher)

- **Professor:**
  - Dr. Edward A. Fox
10. References

- James D. Ivory, College Of Liberal Arts and Human Sciences, Accessed January 2024
- Niko Olson, Mauricio Lovelace, Earl Carter, “Gaming and Political Extremism”, 2023, CS4624 team term project, Accessed January 2024
  http://hdl.handle.net/10919/115018
- Gaijin Games Kft, War Thunder Forums, 2024, Accessed January 2024
  https://forum.warthunder.com/
- Randie Pathirage, “YouTube Comment Scraping Made Easy with Apps Script”, May 16, 2023, Medium, Accessed March 2024
10. References P2

- Soham, Script to export all sheets in a Google Sheet to CSV, Stackoverflow, Accessed March 2024
  https://stackoverflow.com/questions/28707169/google-spreadsheet-download-all-the-sheets-at-once/28711961#28711961
- Google, Google Sheets Official API Documentation, April 3, 2024, Accessed March 2024
  https://developers.google.com/sheets/api/guides/concepts
Questions