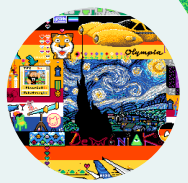
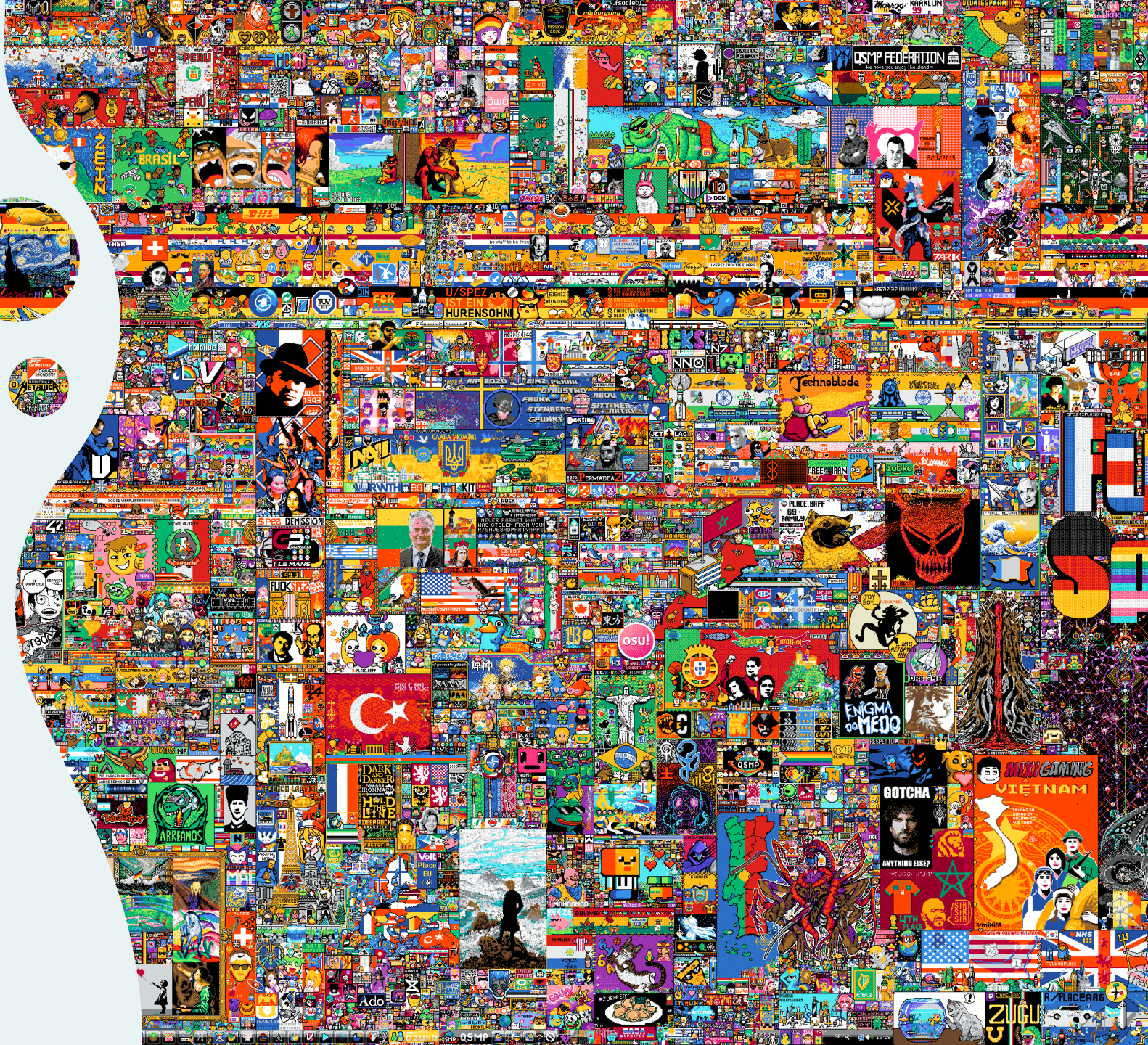
reddit r/place .v2

by Kaden Buckley & Brogan Pietrocini





Data Processing & Preparation

Step 1: Unzip .gzip files

Step 2: Combine .csv files into .parquet

Step 3: Analyze parquet for bot activity in AWS Glue, export to S3

Step 4: Visualize in Sagemaker

~ 13.5gb

Parquet

~ 13.5gb

aws Glue

~ 13.5gb

Amazon SageMaker

~ 2.5gb

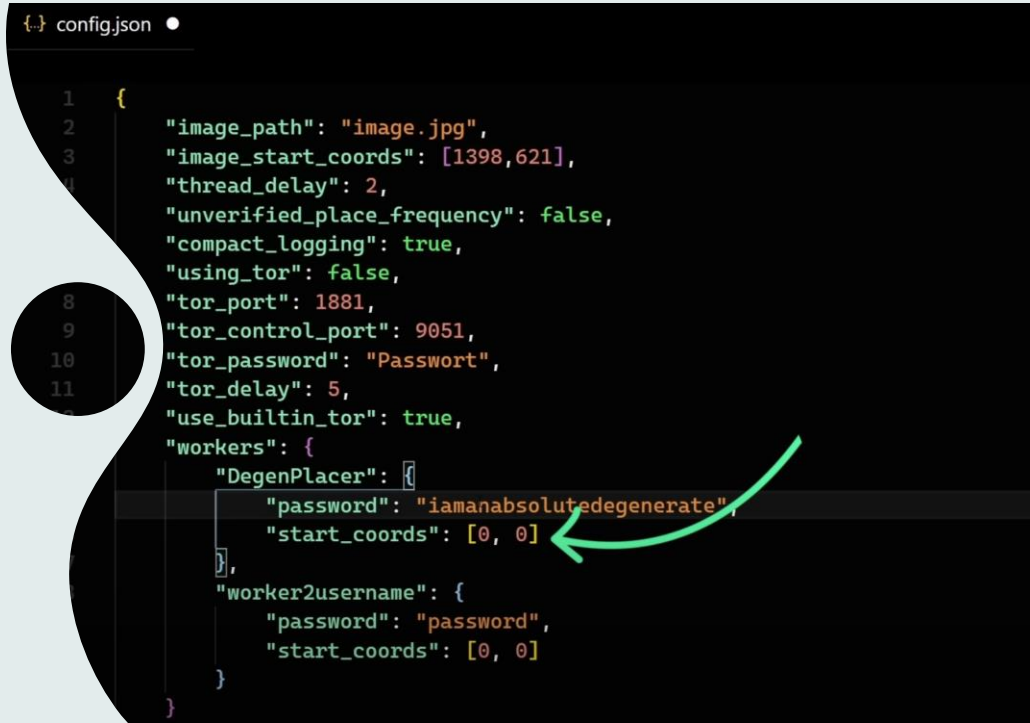
Thinking like a Bot

“workers”: accounts working together as a bot.

A single image can come from many different users, even though it is the same bot.

“start_coords”: location for each individual bot account (worker) to begin pixel placement.

Individual workers operate pixel by pixel, which is hard for a person to achieve.



Bot Criteria

Sus → **Criteria #1:** The time between pixel placement is precisely 5 minutes (tolerance of 1 second).

Also Sus → **Criteria #2:** The pixel placed is within 1-pixel from the user's last placement.

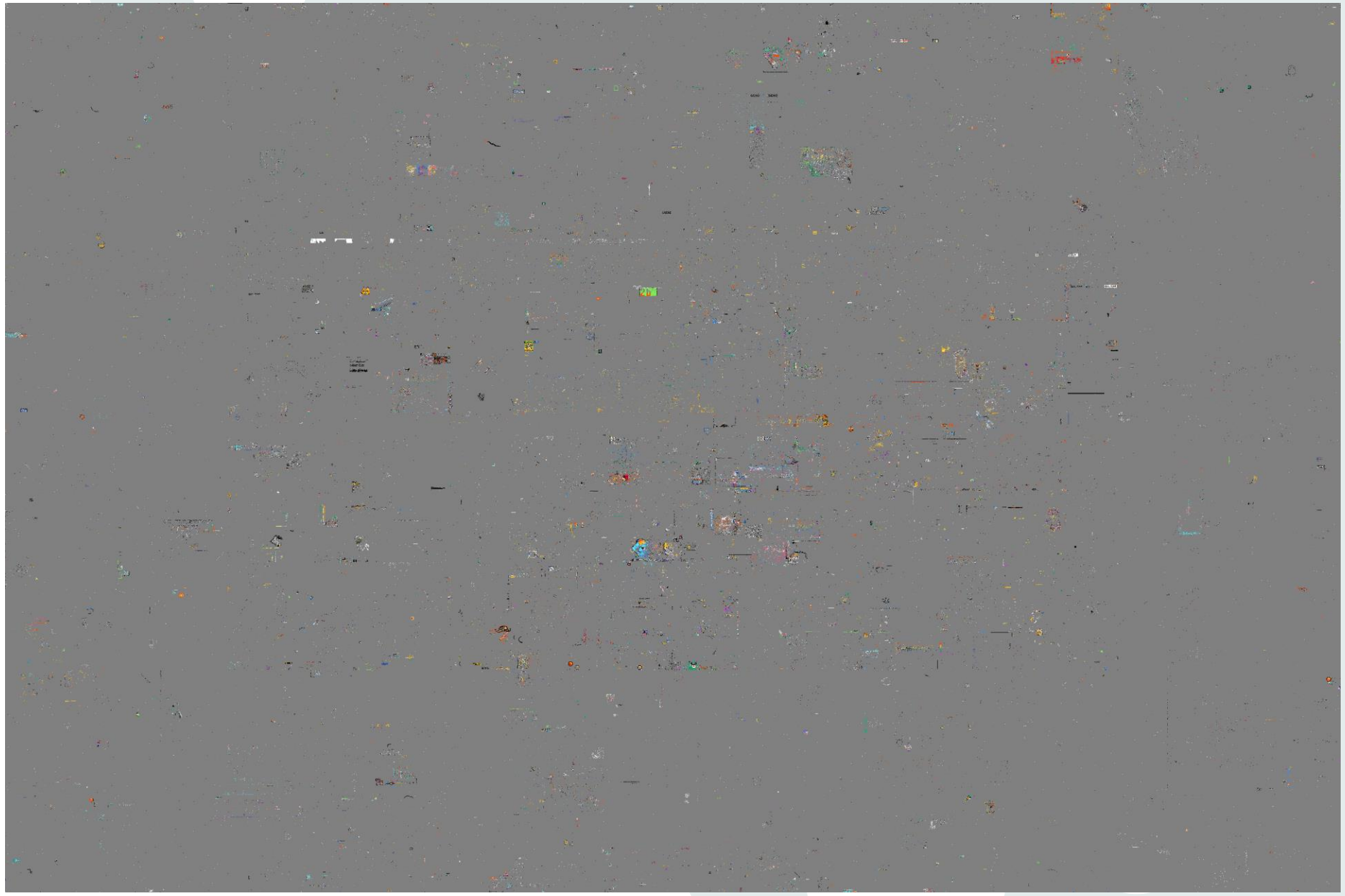
Probably bot → **Criteria #3:** A user has at least 5 instances of both criteria #1 and #2.

Adjustable Thresholds

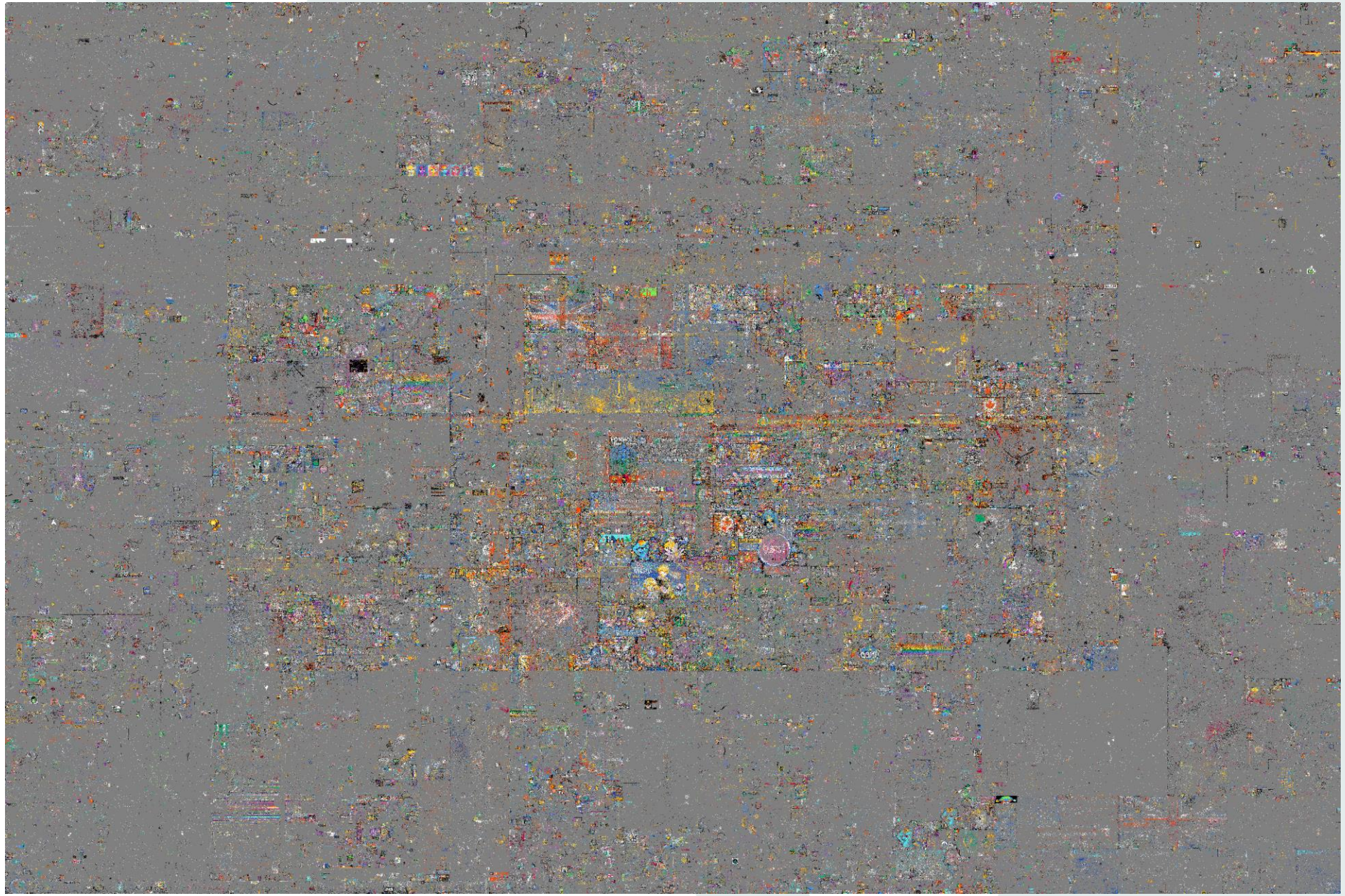
- **Strict** – Time Tolerance = 1 Second & Pixel Placement Distance Tolerance = 1 Pixel
  - Result: 118,249 pixels placed by bots

- **Relaxed** – Time Tolerance = 2 Seconds & Pixel Placement Distance Tolerance = 3 Pixels
  - Result: 1,200,000 pixels placed by bots

Bot Activity Visualization – Strict Thresholds



Bot Activity Visualization – Relaxed Thresholds



Our Bot Policy Recommendations

- Implement a Karma requirement
  - **Reddit Karma**: User score based on contribution to Reddit Community
  - Brand new Reddit accounts start with 1 Karma
- Account age requirement
  - Ex.) Accounts must be at least two weeks old
- Our Bot Detection:
  - Temporarily ban users flagged by our bot criteria (e.g., 24 hours)
  - Permanently ban the account if suspicious pixel placements continue after the temporary ban.
  - Potential additions:
    - Include new thresholds to consider different bot types
    - Include Picture Color