

The motivation behind my visualization was to help people analyze reddit posts in an attempt to make. The main goal is to be able to identify patterns in statistics of reddit posts to be able to tell when a post will make it to the front page. My research question was is it possible to identify if a post will reach the frontpage soon after the most has been made, and my hypothesis was that it would likely be possible to identify patterns that identify posts likely to make the front page but there would likely be enough outliers to make it impossible to be 100% sure.

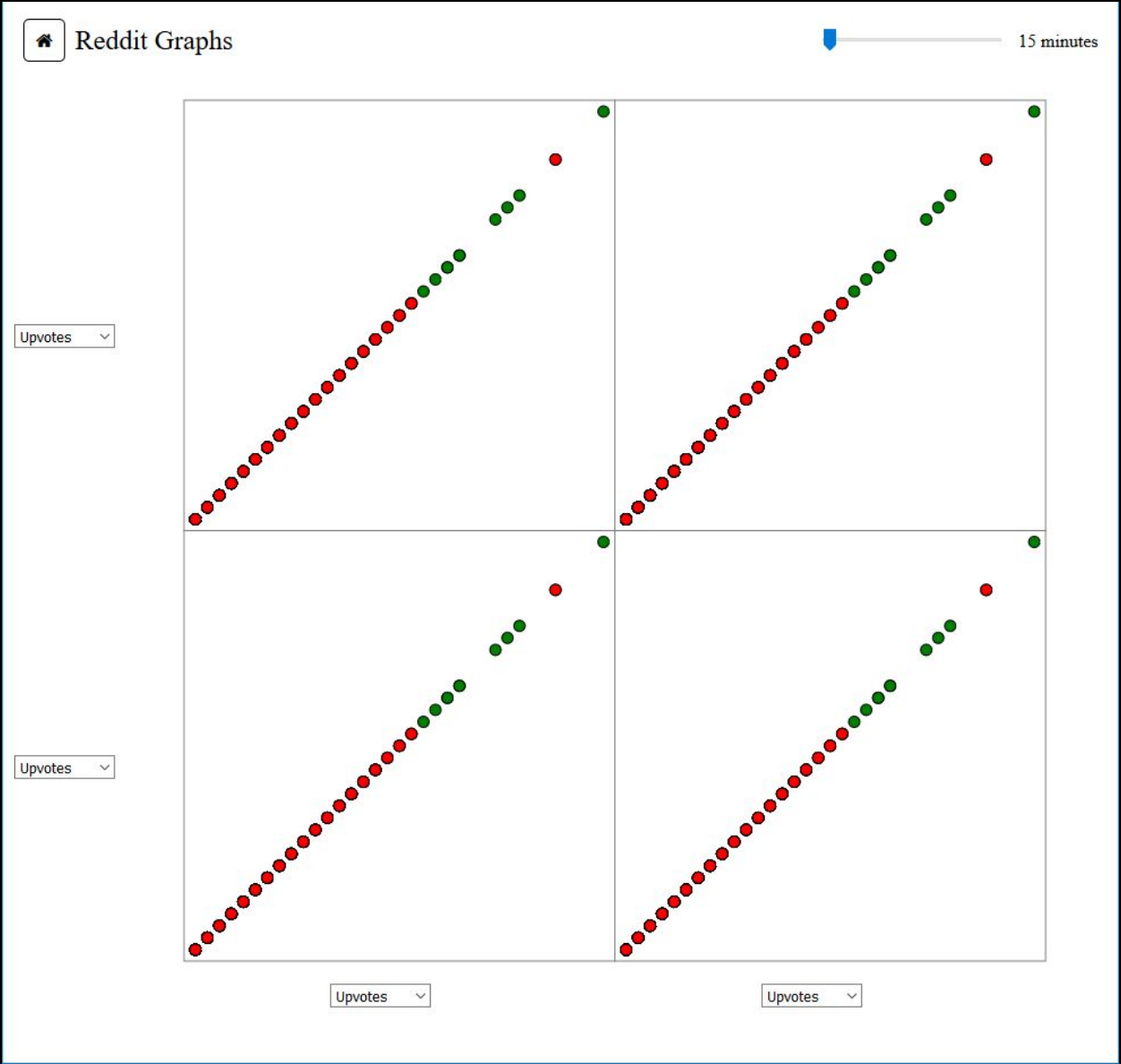
I used the related work cited below to choose which subreddit's to monitor. My data source was pure reddit post data collected every 15 minutes from the moment a post was made until 720 minutes had passed. The data was stored in a postgres database and then converted into javascript to be used by my d3 visualization.

The initial idea for this visualization was that I would allow users to choose how many rows and columns to display, as well as being able to customize what each row and column displayed as well as what time the information should be pulled from. After receiving peer feedback I decided to simplify the design, and I took this even further because due to glitches in my data collection I didn't capture quite as much different data fields as I was expecting. Instead of allowing for each row/column's time to be set independent of each other, I used a slider to control the time for the entire visualization.

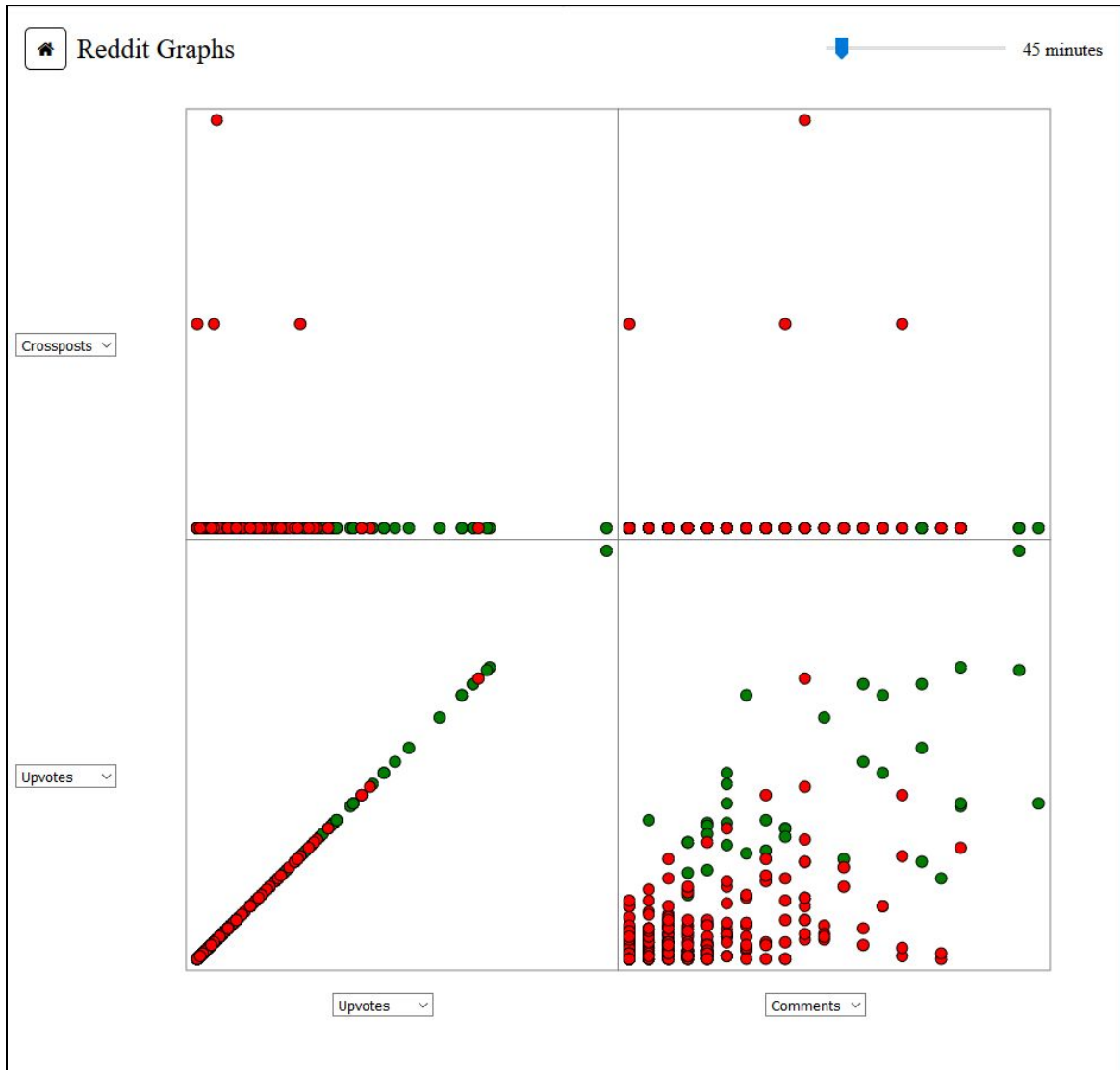
The core feature of this was being able to see multiple different graphs comparing multiple data field for a reddit post. Being able to see all of this information and being able to change the time allows users to find patterns in the data that allows them to predict whether a post will make it to the front post or not. I made this visualization using d3, and created some simple Post and Statistic objects in javascript to organize the information that was pulled from the postgres database.

This can be viewed either by going to iv.richardcollins.me or by loading the index.html in a web browser

This is the default page when you load up the website. It shows 4 different graphs showing data from posts at 15 minutes into their existence.



45 minutes into the life of a post and viewing more than just upvotes we can see that the data is becoming less uniform and more helpful



just

Much later in the life of posts the upvotes vs comments data becomes much more organic

