



# Analyzing Reddit Data

By: Ryan Hathcock,  
Michael Kintanar, Danny  
Xie, Daniel Kang, Klayton  
Trinh



# What is Reddit?

- #5 most visited site in the US
- An american social new aggregation, web content rating, and discussion website
- Founded on June 23, 2005
- Subreddits and topics for almost anything



## Data Set

- Size: 29.62GB
- Date: May 2015
- URL:  
<https://www.kaggle.com/reddit/reddit-comments-may-2015>
- Originally 1+ terabyte of data for 1.7 billion publicly available comments



## Data Set cont.

- 1 Table
- 22 Columns
- 54,504,410 Total Rows
- Reddit comments for the month of May 15
- Fields include: create\_utc, subreddit id, author, score, body, ups, downs, etc

