Reddit User Analysis

Husam Abdul-Kafi, Hari Devaraj, Ethan Liu

Guiding Question:
Who supports Donald Trump on reddit and can we observe a change in his user base over time?

Who supports Donald Trump on reddit and can we observe a change in his user base over time?

- ○ Reddit core user base is young, college-educated and liberal
- ○ The_Donald subreddit is almost a completely separate entity on reddit
- ○ Huge number of people following the_donald
- ○ Different from what pollsters show?

Who supports Donald Trump on reddit and can we observe a change in his user base over time?

- Google Big Query as our data source
- Reddit activity from June 2016 - March 2017
- Total 111 GB data
- The_Donald: 3.81 GB of data

Who supports Donald Trump on reddit and can we observe a change in his user base over time?
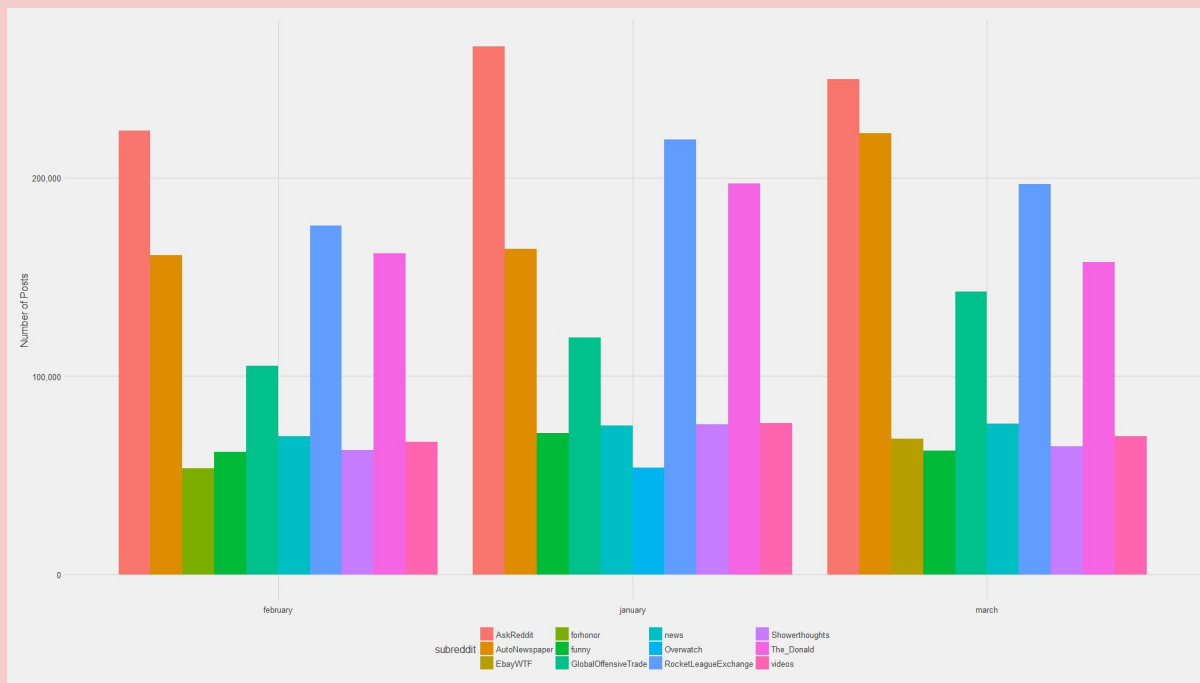
How interesting is this question?

- Top subreddit for March 2017!
  (by posts)
- Top 4 subreddit for March 2017!
  (by score)

Row	subreddit	n_posts
1	AskReddit	249725
2	AutoNewspaper	222477
3	RocketLeagueExchange	196859
4	The_Donald	157405
5	GlobalOffensiveTrade	142488
6	news	75836
7	videos	69694
8	EbayWTF	68342
9	Showerthoughts	64722
10	funny	62515
11	gaming	51026
12	DankMemeArchive	50345
13	Overwatch	49383
14	politics	44355
15	aww	41309
16	▆▆▆▆▆▆	37681
17	worldnews	36948
18	me_irl	36709

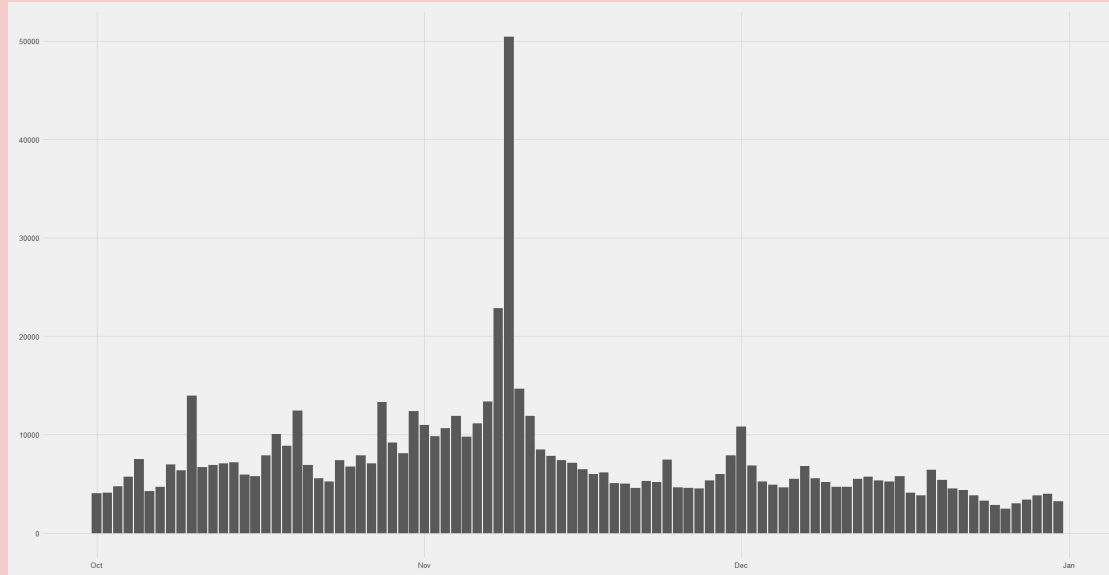| Row	subreddit	total_score
1	The_Donald	42814052
2	funny	19863476
3	politics	14811710
4	aww	12985297
5	pics	12312256
6	gifs	10675949
7	gaming	9558790
8	todayilearned	7616023
9	me_irl	5689820
10	mildlyinteresting	5537912
11	▆▆▆▆▆▆	5226234
12	worldnews	5124352
13	dankmemes	4767430
14	Showerthoughts	4679056
15	videos	4532086
16	Overwatch	4223069
17	AskReddit	4015491
18	AdviceAnimals	3769368

Who supports Donald Trump on reddit and can we observe a change in his user base over time?

Number of posts on top 10 subreddits for 2017

Who supports Donald Trump on reddit and can we observe a change in his user base over time?
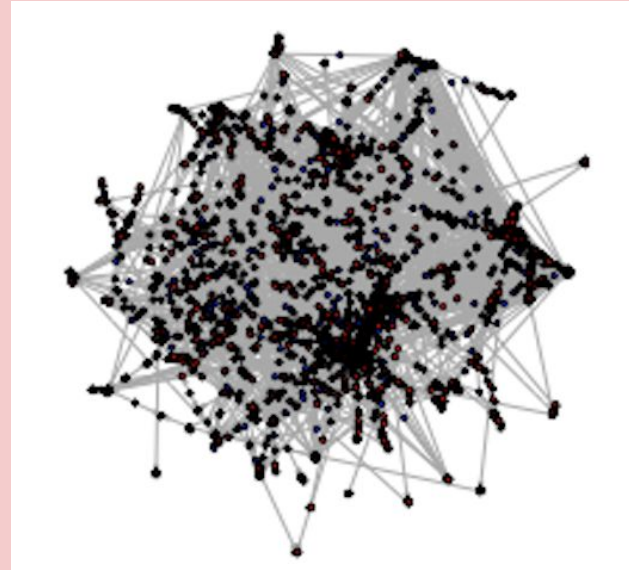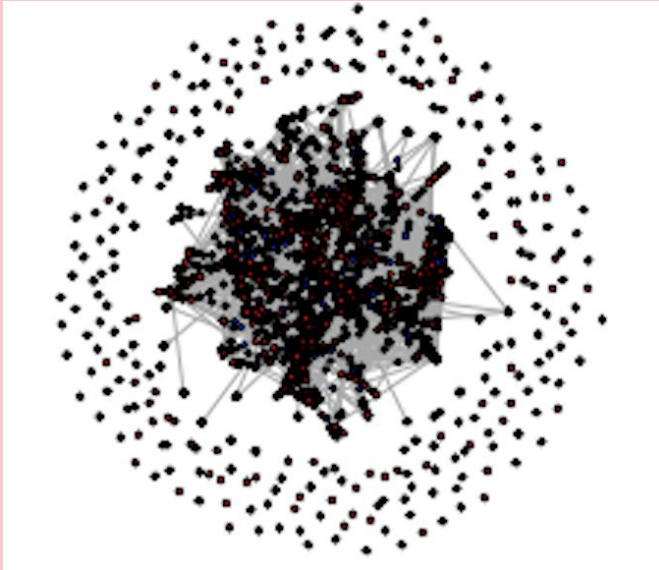
Filter The_Donald by month and day

Row	day_of_month	n_posts
1	2016-11-01	10960
2	2016-11-02	9840
3	2016-11-03	10641
4	2016-11-04	11895
5	2016-11-05	9791
6	2016-11-06	11148
7	2016-11-07	13366
8	2016-11-08	22842
9	2016-11-09	50398
10	2016-11-10	14651
11	2016-11-11	11875
12	2016-11-12	8501
13	2016-11-13	7809
14	2016-11-14	7418

Who supports Donald Trump on reddit and can we observe a change in his user base over time?
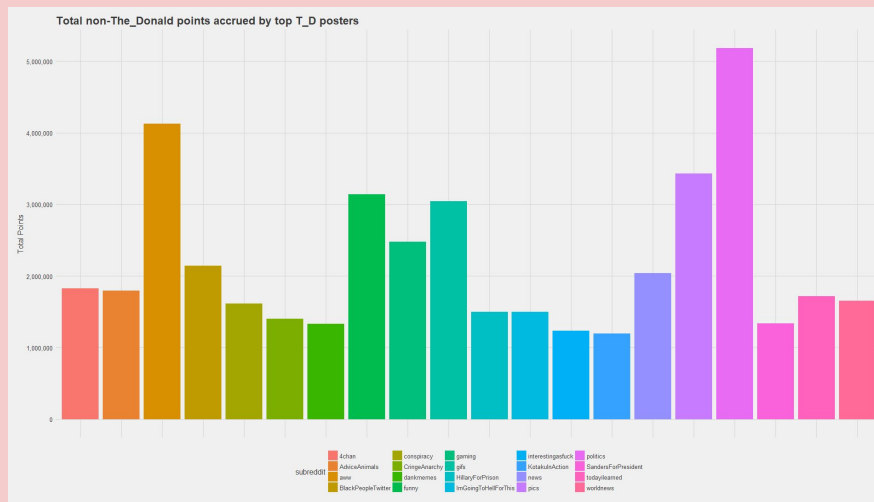
- Can we graph it? Yes we can!

- Played around with several different thresholds:
    - Top five actively used subreddits (>500pts) for active contributors to the_donald (>500pts)
        - User -> subreddit other than the_donald, weight=points on other subreddit
        - Data was too dense
    - Top five actively used subreddits (>500pts) for active contributors to the_donald (>1000pts)
        - Still too dense!

Who supports Donald Trump on reddit and can we observe a change in his user base over time?

Who supports Donald Trump on reddit and can we observe a change in his user base over time?

- Top subreddits for the_donald
    - June 2016-March 2017
    - For those users in t_d with over 1k pts
- Maybe using top subreddits isn't the best strategy?

Total non-The_Donald points accrued by top T_D posters

Row	subreddit	total_score_sum
1	politics	5179472
2	aww	4123193
3	pics	3431812
4	funny	3140245
5	gifs	3045262
6	gaming	2480118
7	BlackPeopleTwitter	2144369
8	news	2040785
9	4chan	1825582
10	AdviceAnimals	1797296
11	todayilearned	1717611
12	worldnews	1652389
13	conspiracy	1612193
14	ImGoingToHellForThis	1497246
15	HillaryForPrison	1497079
16	CringeAnarchy	1404745
17	SandersForPresident	1337849
18	dankmemes	1331856
19	interestingasfuck	1236953

Who supports Donald Trump on reddit and can we observe a change in his user base over time?

- Working on decluttering our graphs
    - We want to demonstrate nice, clean clustering around a few main subreddits
- The graph for: top five actively used subreddits (>500pts) for active contributors to the_donald (>1000pts) about 265 "connected components"
    - First component has 4464 users
    - Every other component has two vertices
    - Clearly we need a better way of decomposing the graph
    - Data size: 17933 edges vs 7172

Who supports Donald Trump on reddit and can we observe a change in his user base over time?

- Ideas for visualizing change over time:
    - Break up observances into three periods:
      June - Election, Election - Inauguration, Inauguration - Current
    - Break up observances into month - month analysis
- Judge users on top n most popular subreddits overall for active users of the_donald?
- Play around with threshold parameters for better visualization
- Can we do better than a bipartite visualization?

Who supports Donald Trump on reddit and can we observe a change in his user base over time?

- Reddit doesn't have information on its user base apart from activity
- Is reddit activity a strong indicator of preference?
- Can preference be interpreted as the sum of your top n interests?
- How do idiosyncratic behaviors of the community affect the data?
    - Both reddit specific and the_donald specific
- Something that is unquantifiable: people act differently when they are anonymous