

# Descriptive Statistics - POIR614 Final Project

Bijean Ghafouri

11/11/2021

## Data, concepts and measures

First, I describe the data used and how I prepared it. I use a dataset published online by Chen, Deb and Ferrara (2021) that collected all tweets related to the 2020 US Presidential election from early 2020 until the summer of 2021. These tweets were collected through keyword search using Twitter's API.

From this data, I sample all tweets posted in June 2020. In total, this represents 67,181,551 tweets related to the election (in English).

The goal of this report is to show preliminary results related to my term project. In this project, I seek to describe the persuasive efficiency of extremist content on social media. Using complex contagion, I hope to demonstrate that extremist behavior requires less exposure to adopt relative to other types of political behavior. My hypothesis is that holding exposure constant, users on Twitter are more likely to adopt extremist behavior than moderate behavior.

## Exposure and adoption

In this project, I measure the exposure of tweets and subsequent adoption of #qanon. This hashtag is meant to conceptualize extremist behavior. I consider users who use #qanon in one of their tweets as adopting extremist political behavior. I consider users exposed to #qanon when they retweet, quote tweet or reply to a tweet using #qanon. Ideally, exposure would be measure implicitly, in which we measure exposure when a user follows another that used the hashtag previously. Instead, I use an explicit measure of exposure where we know a user is consciously exposed to #qanon since they engaged with the tweet in some capacity. A user adopts an extremist behavior when they use #qanon in a tweet they authored.

To model complex contagion, I count the number of exposures a user had of #qanon before themselves using the hashtag. Out of the 67,181,551 tweets posted in June 2020 related to the US Presidential election, 351,108 tweets contained #qanon without 83,867 unique users. Out of these tweets, only 4024 tweets were original, meaning that 4024 users demonstrate the adoption of extremist behavior.

Out of these 4024 users who wrote at least one tweet with #qanon, only 1624 users were exposed to #qanon prior to adopting the behavior. This means that 2400 users in the sample used #qanon without prior exposure.

Among users who required prior exposure to adopt extremist behavior (1624 users), we observe an average of 5.467 exposures prior to behavior adoption. Among the entire sample of 4024 users, the average is 2.206 exposures prior to behavior adoption. In other words, Twitter users who authored a tweet containing #qanon were exposed to tweets authored by other users they saw on their timeline about 2.2 times before they authored that tweet.

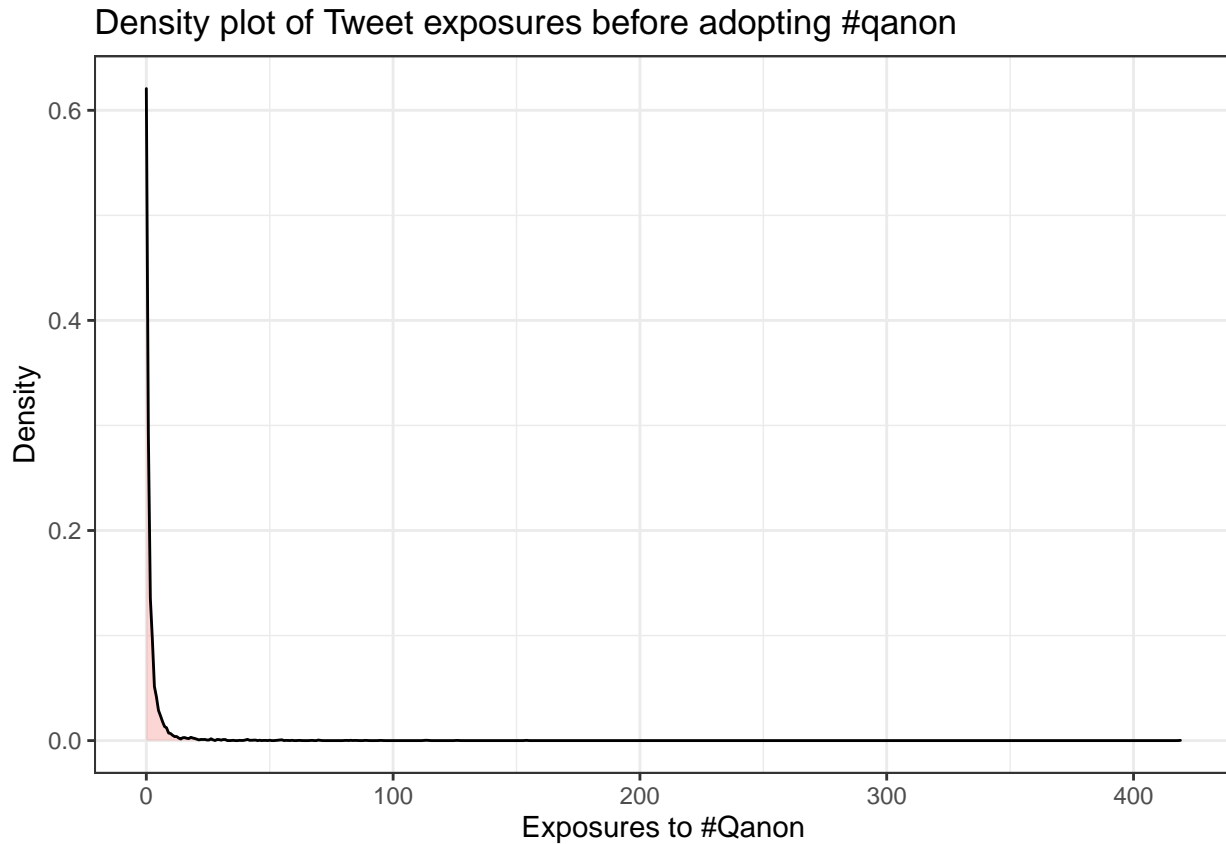
It is important to mention that I do not measure exposures to #qanon after the first time a user adopts the behavior. I mark a difference between behavior adoption and extremist expression. That is, adoption measures the first time a user tweets with #qanon. Extremist expression includes the number of times a

user tweets #qanon. It is possible that a single exposure enables a user to initially adopt the behavior, while they subsequently require far most exposures to express political extremism frequently. On the contrary, it is possible for a user to require multiple exposures to initially adopt the behavior, but subsequently Although this may be interesting to measure, I do not capture these possibilities in this report.

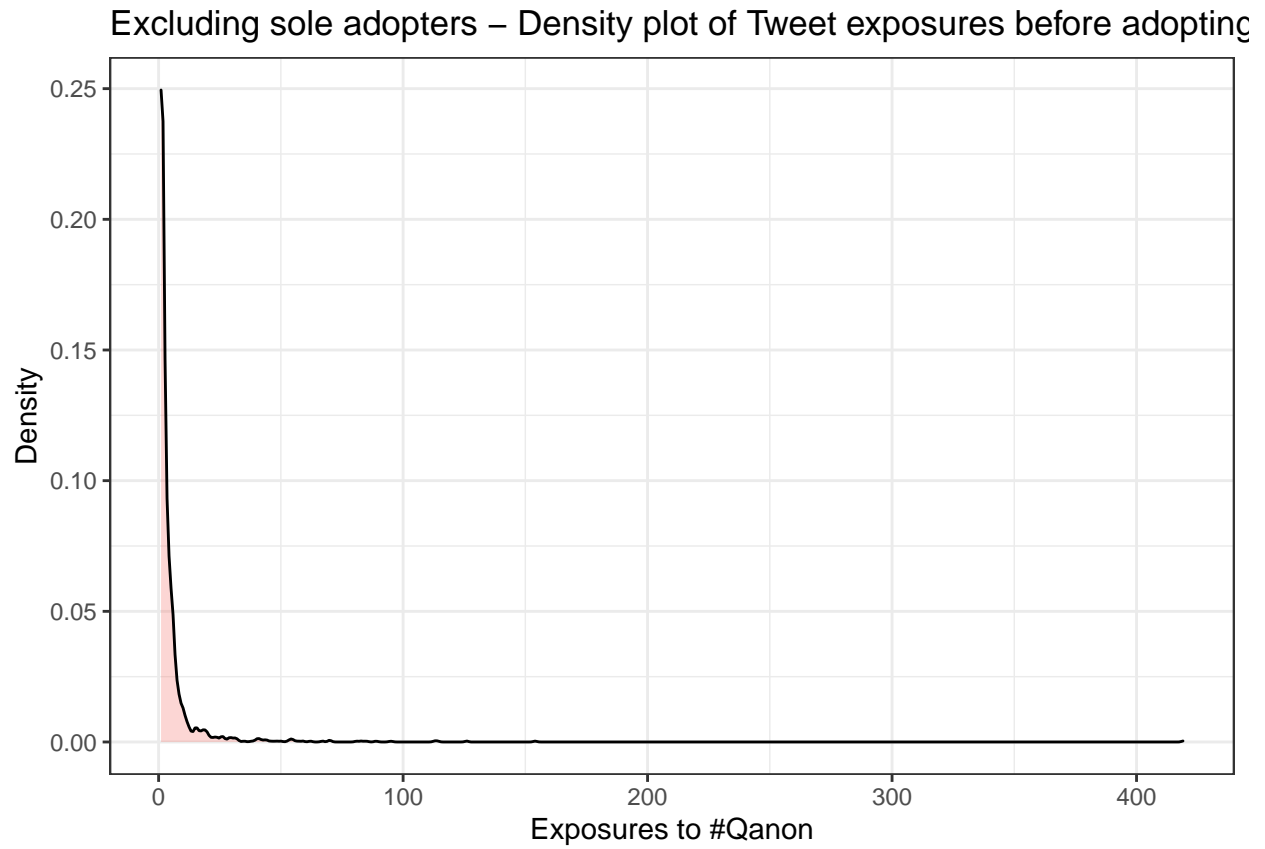
## Data visualization

Here, I show some density plots that demonstrate my preliminary results.

First, we can plot the density distribution of the number of exposures before adoption. This includes users who require 0 exposures, which represent 60% of the sample.

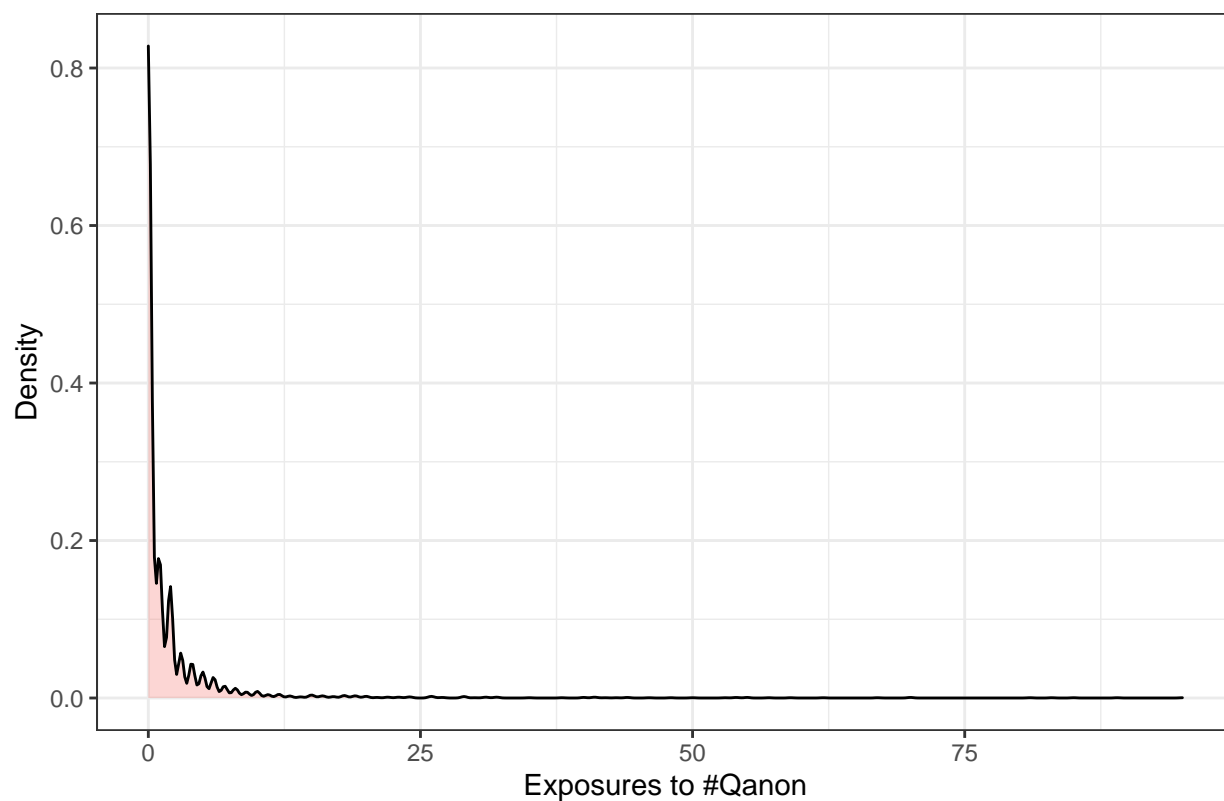


Next, let's consider users who required at least one exposure before using #qanon.



We notice that only a few users require more than 100 exposures to adopt the behavior. Let's exclude these users from our sample, considering users who required less than 100 exposures.

Reduced: Density plot of Tweet exposures before adopting #qanon



Finally, let us consider the same sample, but with users who again required at least one exposure before adoption.

Reduced: Excluding sole adopters – Density plot of Tweet exposures before

