Social Media and Art Patronage

STATS 767 Project

Saurabh Gupta

17th October 2018

Introduction

Effect of Social Media on Art Patronage

Research Question: *Is the social media presence of creators (artists) related to their online patrons?*

Patreon.com is an online platform that allows creators to seek Patrons (similar to paid subscribers) pledging to pay as little as $1 per month to more than $1000.

- The payments and the distribution technology is taken care of by Patreon in exchange for a 5% fee (and some charges).
- It hosts more than 122 thousand creators and 4 million pledges.

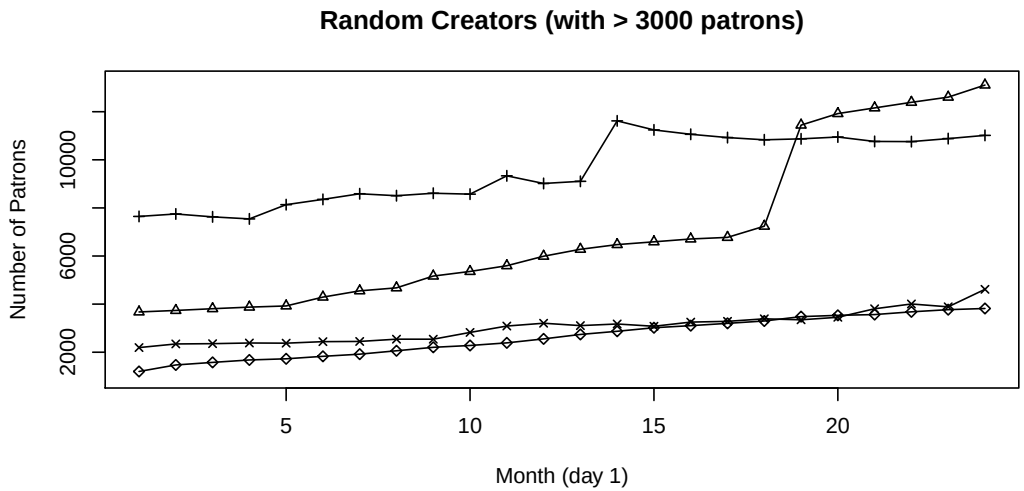
[0]Notes: Brand and creator names used in this presentation are copyrights of their respective owners. The data has been sourced from `Patreon.com` and `Graphtreon.com`. It cannot not be published without their permission.

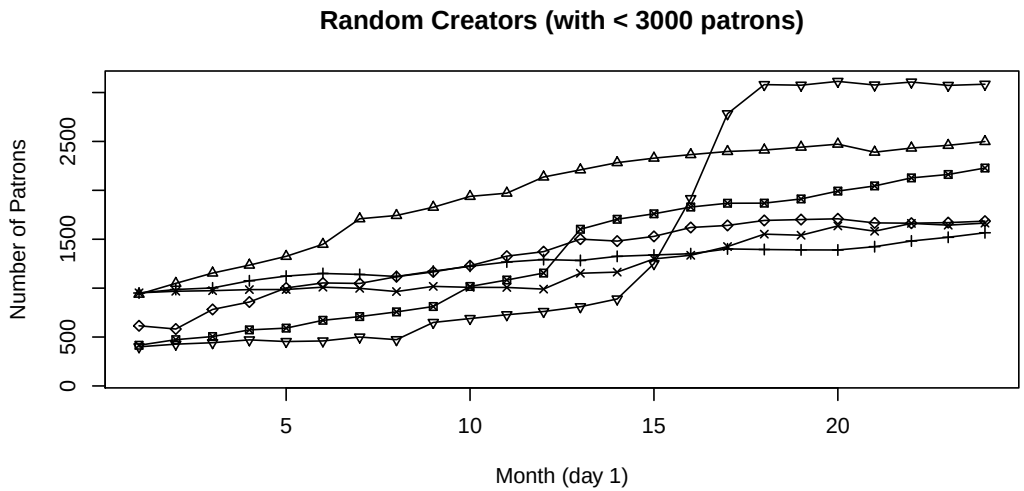
The dataset

Time series data on the number of Patrons and 5 social media statistics for top 103 creators in different categories.

- 103 rows wherein each creator is an observation
- 138 variables (columns) include 23 months' data for 6 variables
- Social media metrics include Facebook Likes, Twitter Followers, YouTube Subscribers, YouTube Videos and YouTube Views
- Data is measured on the 1st of each month
- The time series window is from 1st July 2016 to 1st May 2018

Number of Patrons for 10 random creators



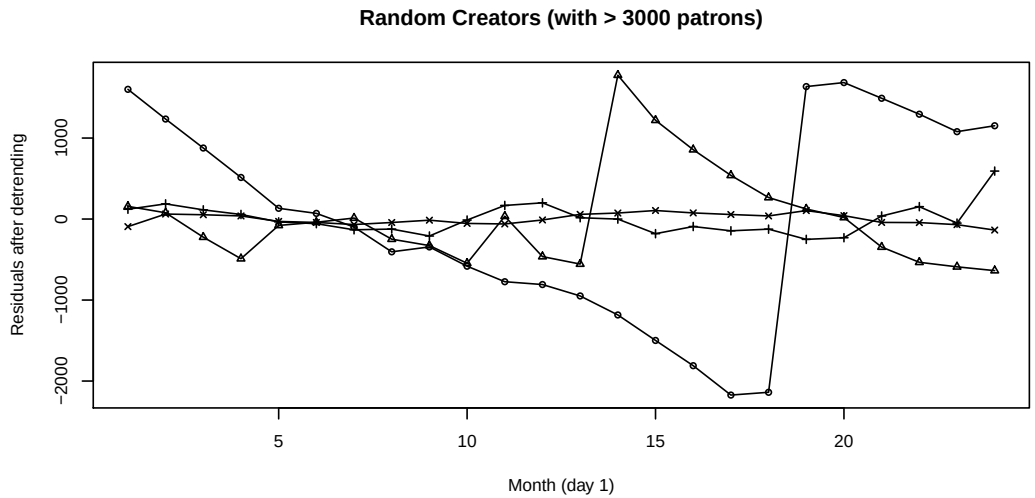


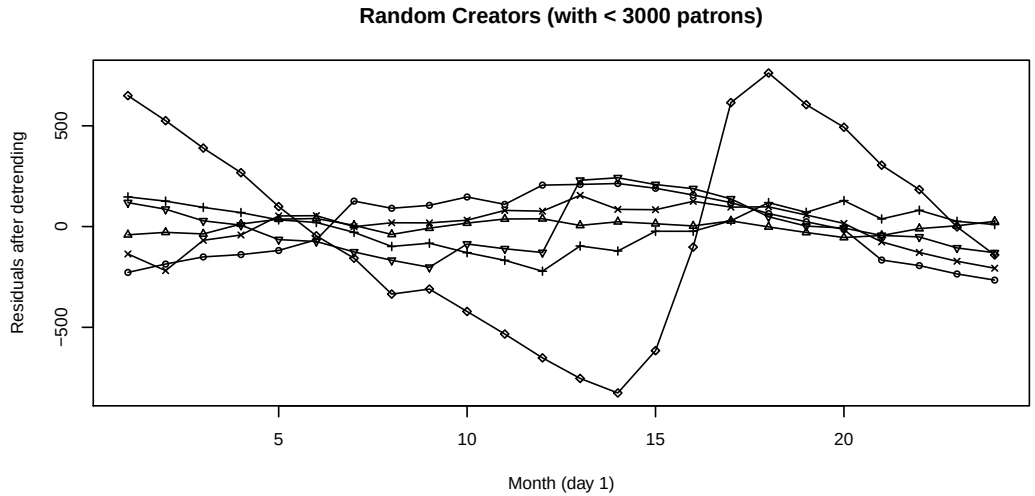
Methodology

The trend in all variables can result in spurious correlations. To take a closer and unbiased look, we may detrend the data by regressing each of the variables on time period $t = (1, 2, \ldots, 23)$.

The number of Patrons aren't expected to increase indefinitely. Hence, I used quadratic function of time to detrend the data. After detrending, we are left with the residuals. They represent the variability in the data.

Patrons after quadratic detrending





Methodology (contd.)

The number of variables, $p = 138$ is larger than the number of observations, $n = 103$. Hence, I will perform PLS Correlation to check the relationship between variability in social media metrics and the number of Patrons.

- Let, Y denote the variability in Patrons for each of the 23 months.
- X denotes the the variability in social media metrics for the same period.

The variables

```
Dimensions of Y

[1] 103  23
```

Names of first 5 Y Variables
Patrons_1
Patrons_2
Patrons_3
Patrons_4
Patrons_5

```
Dimensions of X
```

```
[1] 103 115
```

Names of first 5 X variables
Facebook.Likes_1
Twitter.Followers_1
Youtube.Subscribers_1
Youtube.Videos_1
Youtube.Views_1

Partial Least Squares Correlation

Correlation between the first pair of canonical variates

Scaling of variables is done by default in the `pls` function in **R**. The X and Y scores are stored in PLS object `variates`.

Approximately 10-11% permutations have larger correlation than observed for the first axis and 11-12% for the second axis. This implies that there is weak evidence of a relationship between Patrons and social media metrics.

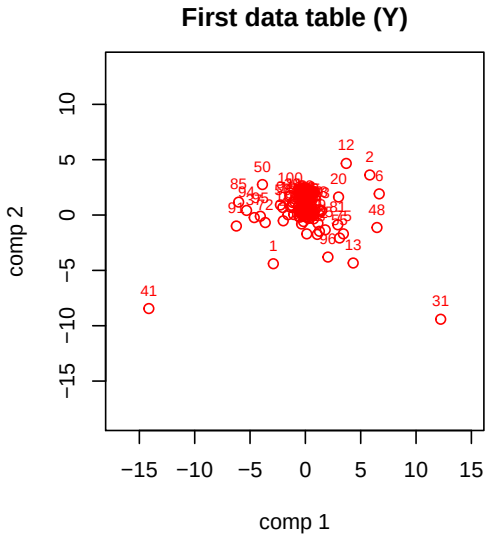
```
Observed Correlation: 0.63
```

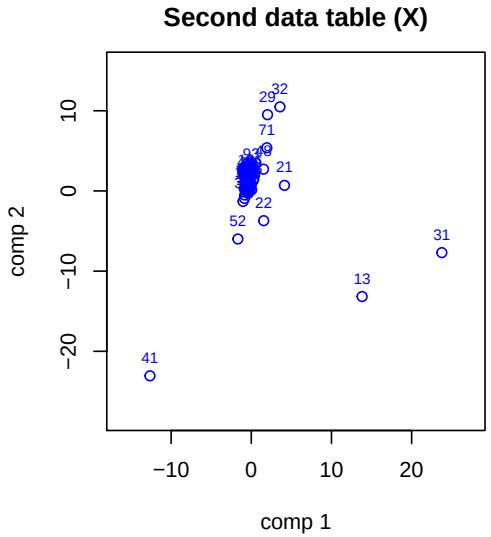
```
p-value of correlation based on 1000 permutations: 0.108
```

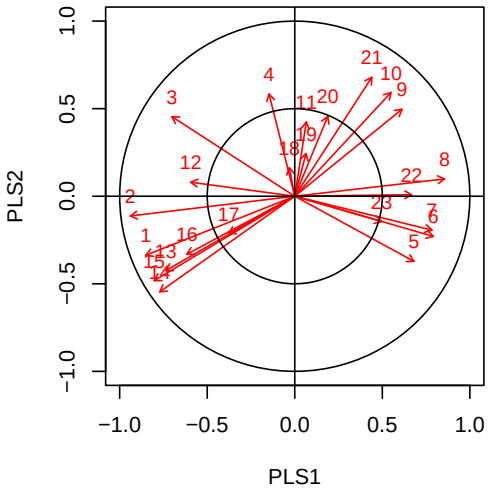
Correlation between the second pair

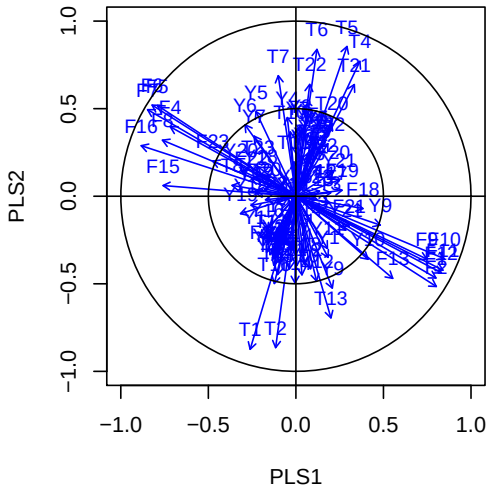
```
Observed Correlation: 0.62

p-value of correlation based on 1000 permutations: 0.114
```

Biplots


First data table (Y)
comp 2
comp 1


Second data table (X)
comp 2
comp 1


PLS2
PLS1


PLS2
PLS1

Months well represented by PLS axes of Y

The biplot for **Y** (bottom left on previous slide) indicates:

- Months 5 to 8 have high positive correlation with axis 1. These were the periods with apparently lower variability (not tested statistically).
- First 3 months and 13 to 16 have high negative correlation with axis 1. These were the periods with high variability.
- Axis 2 doesn't have high correlation with most months (except 14, 21).

X variables well represented by PLS

The biplots for **X** (bottom right) indicate:

- Facebook Page Likes (F1 - F23 in the biplots) are well respresented by PLS axis 1
- Twitter Followers (T1 - T23 in the biplots) are well represented by PLS axis 2

This implies that their relationship is strongest with variability in Patrons. The values are printed for clarity on the next few slides.

Outliers

The component plots in the top row indicate:

- For **Y** variables, there are atleast 2 outliers - creator 41 and 31
- For **X** variables, there are more outliers
- We also observed from the time series plots of residuals that there have been some extreme months.

The outliers may adversely affect our results and interpretation such as reducing correlations for other creators.

PLS axis 1: Highest correlations with X

	Positive Corr.
Facebook.Likes_10	0.85
Facebook.Likes_11	0.84
Facebook.Likes_12	0.83
Facebook.Likes_2	0.80
Facebook.Likes_1	0.80
Facebook.Likes_9	0.75

	Negative Corr.
Facebook.Likes_16	-0.88
Facebook.Likes_7	-0.85
Facebook.Likes_6	-0.82
Facebook.Likes_5	-0.79
Facebook.Likes_8	-0.76
Facebook.Likes_15	-0.76

The first PLS axis for **X** has the highest correlations with `Facebook.Likes`.

- Months 5 to 8 have high negative correlation. This implies if `Facebook.Likes` don't increase, the score will be higher. For the corresponding first axis of Y, it indicated low variability.
- Months 9 to 12 have high positive correlation. They are not represented as clearly by axis 1 of Y and difficult to interpret. It could be a reason why the p-value was borderline.

PLS axis 2: Highest correlations with X

	Positive Corr.
Twitter.Followers_5	0.86
Twitter.Followers_6	0.84
Twitter.Followers_4	0.77
Twitter.Followers_7	0.69
Twitter.Followers_22	0.64
Twitter.Followers_21	0.64

	Negative Corr.
Twitter.Followers_1	-0.87
Twitter.Followers_2	-0.86
Twitter.Followers_13	-0.69
Youtube.Views_9	-0.52
Facebook.Likes_2	-0.52
Twitter.Followers_16	-0.50

The highest correlation of the second PLS axis is with `Twitter.Followers`.

- Months 4 to 7 and 21, 22 are positively correlated. The corresponding axis of Y is positively correlated with months 9, 10 and 21. It could imply a time lag in the relationship between Twitter and Patrons.
- Negative correlations are weaker (closer to 0.5) and more random.
- It could be another reason for a borderline p-value.

PLS axes: Lowest absolute correlations with social metrics

PLS axis 1

	Abs. Corr.
Youtube.Views_14	0.00
Youtube.Videos_16	0.00
Twitter.Followers_14	0.01
Youtube.Views_1	0.01
Youtube.Videos_21	0.01
Youtube.Views_7	0.01

PLS axis 2

	Abs. Corr.
Youtube.Subscribers_23	0.00
Youtube.Videos_16	0.00
Youtube.Subscribers_13	0.02
Youtube.Subscribers_17	0.02
Facebook.Likes_17	0.02
Twitter.Followers_9	0.02

Lowest absolute correlations

YouTube dominates variables least correlated with the first two PLS axes.

- Videos are expected to have maximum viewer impact. `Youtube.Videos` may be considered to be a proxy for number of publications. In this context, it is a bit counter-intuitive.
- However, given that about half the creators aren't video publishers, it seems reasonable.
- Secondly, the ability to post videos on Facebook and Twitter seem to have taken their toll on YouTube.

Wrap Up

Conclusions

- Overall, there is a weak correlation between variability in social metrics and Patrons.
- Twitter and Facebook have higher correlation with Patrons.
- It reinforces popular belief that Twitter and Facebook are good platforms for engagement. However, it can have a negative impact as well.
- Time lag in the effect can be studied further.
- Effect of You Tube can be investigated separately for Video Creators.

What more can be done

- The variables have an almost linear trend.
- If we regress them over time, $t = (1, 2, \ldots, 23)$, we can get the slope and intercept term for each of the variables.
- PLS regression can be performed with slope and intercept terms of each metric as variables.
- This will let us know the effect of social media on the trend in patrons rather than the variability.

Thank You