

Does the Performance of a City's Sports Teams Impact the Happiness of the Populace? (No)

Introduction

In 2012 The Barclays Center opened in Brooklyn becoming the new home to the (formally) New Jersey Nets becoming the first professional sports team to play in Brooklyn since The Brooklyn Dodgers defected to Los Angeles in 1957. At the time the owner of the Dodgers, Walter O'Malley, wanted to build a new stadium next to the Atlantic Railroad Yards and wanted the city to seize the property for him using eminent domain. When Robert Moses refused to give the Dodgers the land, O'Malley moved the team across the country to Los Angeles. Half a century later the city used eminent domain to seize property just next to the land O'Malley had wanted to make room for the Barclays Center, displacing up to 3000 people from their homes¹.

The consensus among economists is that there seems to be very little economic benefit to a city by having a professional sports team²³ but maybe there is some other benefit that can justify evicting people from their homes to make room, or the huge subsidies cities often give to sports teams? Does having a local sports team perhaps make a city's populace happier?

Happiness is of course difficult to quantify, but three different datasets were looked at which may offer some insight into how happy a city is: New York City's 311 data, Tweets geolocated in New York City, and Crime incident data in Boston.

Sports Data and Impact

For this analysis the impact of a sports team is strictly based on how well or poorly a team is doing. It could be argued that merely having a team would make a city happier, but the datasets available don't go back far enough to evaluate such a claim. How well a team is doing was considered by the team's streak which is the current number of wins or losses they've had in a row and whether the game is a playoff game.

311 Analysis

New York City's 311 system allows people to report non-urgent problems such as noise complaints or potholes. A dataset is available with each individual incident since 2010 including data reported, the type of incident, and details about it. The data was simply aggregated by date to get a count of how many complaints were made each day.

The logic in looking into 311 data is that people in a good mood may be less likely to go out of their way to report problems than if they were already unhappy. Noise complaints were also looked at separately from the 311 incidents as a whole because an inverse relationship may be expected – that is, people often make more noise when they are happy.

Looking at number of 311 complaints versus day of the week could offer some credence to the above assumptions. Total number of complaints per day is very highly correlated with the day of the week with much fewer complaints being made on weekends and then a slight uptick in complaints early in the work week. There is also a variance in number of complaints across seasons. Table 1 contains the results from doing Ordinary Least Squares

linear regression of the number of 311 complaints made on a particular day against the day of week and month of year, both treated as categorical variables with Monday and January treated as the reference categories, respectively. And looking at just noise complaints versus day of the week we see the reverse, shown in Table 2.

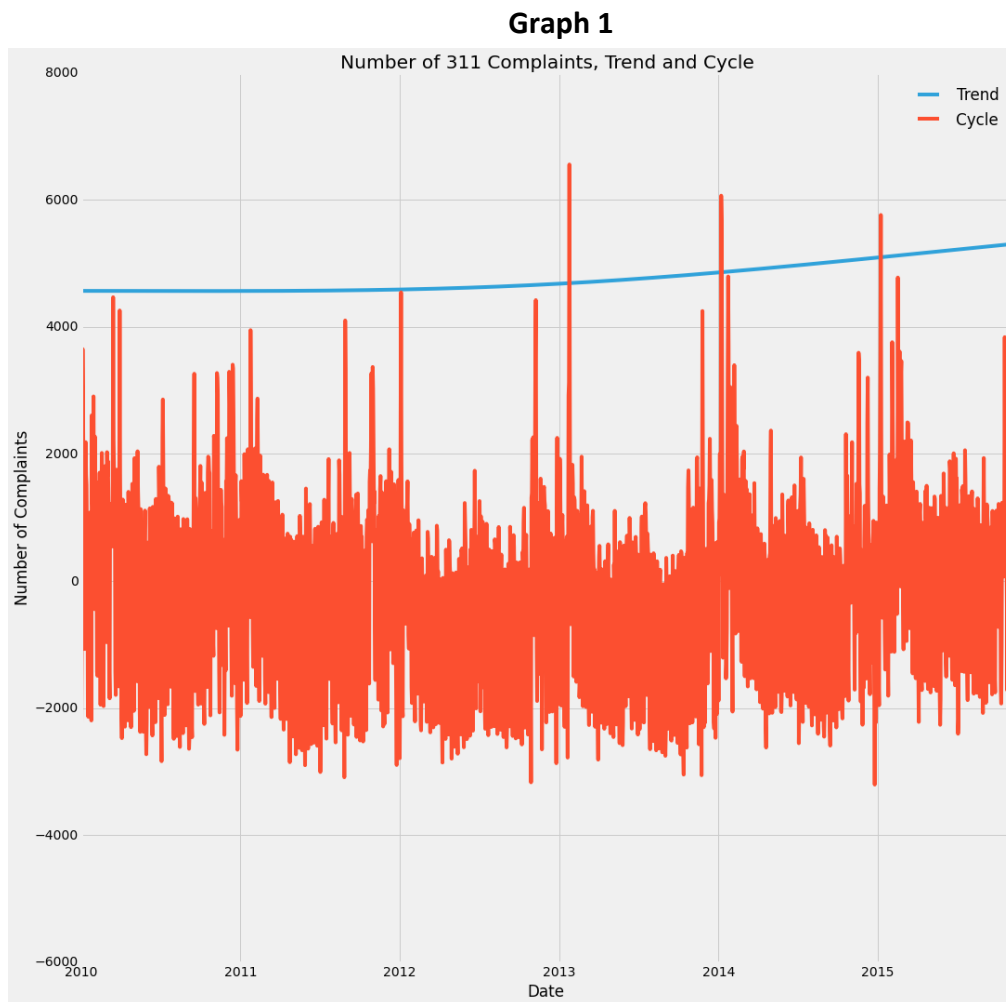
Table 1: Number of 311 Complaints Regressed versus Days of Week and Month

Category	Coefficient	95% Confidence Interval	
Tuesday	195.9771	51.588	340.366
Wednesday	84.5197	-59.871	228.91
Thursday	-66.8906	-211.402	77.62
Friday	-467.505	-611.896	-323.113
Saturday	-2261.73	-2406.12	-2117.34
Sunday	-2617.83	-2762.22	-2473.44
February	-118.364	-307.853	71.124
March	-367.626	-552.522	-182.73
April	-891.666	-1078.1	-705.235
May	-920.161	-1105.05	-735.269
June	-615.115	-801.547	-428.684
July	-651.215	-836.111	-466.319
August	-868.564	-1053.46	-683.67
September	-874.653	-1061.09	-688.222
October	-528.615	-713.509	-343.72
November	-363.857	-558.122	-169.591
December	-574.497	-768.417	-380.576

Table 1: Number of 311 Noise Complaints Regressed versus Days of Week and Month

Category	Coefficient	95% Confidence Interval	
Tuesday	0.1162	-19.896	20.128
Wednesday	5.8707	-14.141	25.883
Thursday	13.2183	-6.81	33.247
Friday	44.7673	24.755	64.779
Saturday	168.5864	148.574	188.598
Sunday	113.2616	93.25	133.273
February	11.0968	-15.166	37.359
March	67.0907	41.465	92.717
April	151.7579	125.919	177.597
May	254.8118	229.186	280.437
June	250.192	224.353	276.031
July	179.338	153.712	204.964
August	219.8238	194.198	245.449
September	241.8218	215.983	267.66
October	186.7548	161.129	212.38
November	68.8689	41.944	95.793
December	26.8339	-0.043	53.711

Of course there are many other ways to explain the correlation, such as people simply not being out as much on weekends so they notice fewer problems which they then see come Monday; and there would be more noise complaints on weekends since there are more parties then. But regardless of the reason for the correlation, this means we need to control for day of the week and month in the following analysis.



Lastly the number of 311 complaints has not been constant over time as show in Graph

1. So the cycle was detrended from the trend using an HP filter with a lambda of

104976000000. The following regressions were run against the cycle rather than the total number of complaints to take the underlying increasing trend into account.

Taking all of this into account, the fluctuations of 311 complaints per day was regressed versus the day of the week, month of year, current streak of the team and also whether it was a playoff game or not. This was done for the Mets 2010-2015 and the Yankees 2010-2015 for both all 311 complaints and for just noise complaints and the results are in Table 3. The only coefficient that doesn't include 0 in the 95% confidence interval is the coefficient of the Mets' win streak. But that is still so close to 0 that we can conclude there is no effect. And in fact the value is greater than 0, opposite of the effect we hypothesized which if anything would mean New Yorkers are less happy when the Mets are doing well (which frankly sounds about right).

Table 3: 311 vs Mets and Yankees Performance Regression Results

			Coefficient	95% Confidence Interval	
All 311 Complaints	Mets	Streak	17.0954	0.920	33.271
		Is Postseason	71.5393	-471.124	616.203
	Yankees	Streak	-5.3705	-22.737	11.996
		Is Postseason	-178.8692	-649.880	292.142
311 Noise Complaints	Mets	Streak	1.1870	-1.484	3.858
		Is Postseason	23.5082	-66.109	113.125
	Yankees	Streak	1.9611	-0.907	4.829
		Is Postseason	10.4681	-67.316	88.252

Crime Data

The New York Police Department doesn't publish incident level records, only summaries at a weekly level, so instead Boston's crime dataset was used, which does publish a dataset of each crime as an individual entry.

The methodology for analysis was very similar to that described in the 311 section – records were aggregated at a daily level and then the daily count regressed versus the performance of, in this case, The Boston Red Sox while controlling for day of week and month of year. The results are in table 4. Again the 95% confidence intervals of the coefficients include 0, so we can't reject that there is no effect.

Table 4: Boston Crime versus Red Sox Performance Regression Results

	Coefficient	95% Confidence Interval	
Streak	-0.2713	-1.149	0.607
Is Postseason	-2.3262	-25.582	20.929

Twitter Data

Using social media data is probably the most direct way we could measure happiness. Rather than trying to discern happiness from other factors such as crime or complaints, we can observe people being or saying that they are happy.

Unfortunately Twitter only offers historical datasets at a hefty price. There is a small amount of historical Twitter data available on CUSP's datahub. It contains a few hundred thousand tweets geolocated in New York City collected from the real-time Twitter feed on 17 days during winter 2014-2015. Since they were collected during the winter, basketball results were used instead of Baseball. Due to the limited amount of data, I don't expect to be able to draw any conclusions from this analysis, but present it as what may be done with more data available.

Sentiment analysis was done by feeding each tweet to a naïve bayes classifier which had been trained on the NLTK Sentiment Polarity Dataset which contains movie reviews. This produces a polarity of 0 to 1 for each tweet, 0 being negative and 1 being positive. These values were averages for each day to get an average daily sentiment. The average values ranged from .56 to .64.

These values were regressed versus the win streaks for both the Brooklyn Nets and New York Knicks. At first glance there appears to be a correlation between the Knicks' win streak and sentiment as shown in graph 2. But the three sentiment outliers are all clustered in one week during November, and controlling for that makes the effect of the Knicks go away as shown in Table 5.

Graph 2

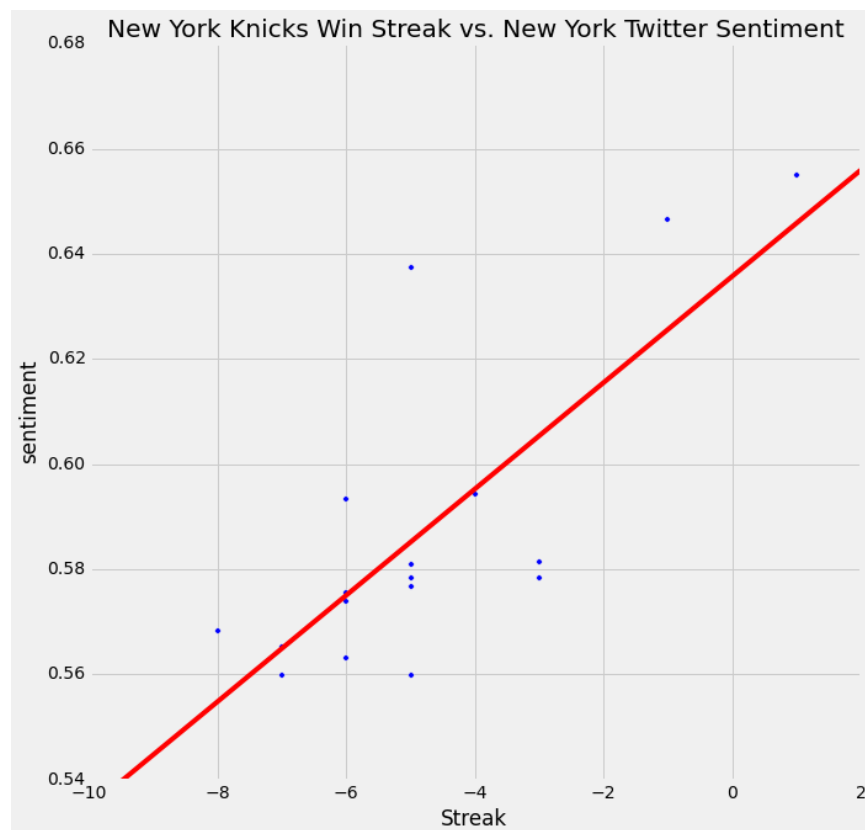


Table 5: Twitter Sentiment versus Knicks and Nets Performance Regression Results

	Coefficient	95% Confidence Interval	
Knicks	0.0024	-0.003	0.008
Nets	-0.0020	-0.006	0.002

Conclusion

None of the three aspects looked at appeared to be impacted by how well a city's sports team is doing. 311 complaints or crimes may be a bad measure of happiness but a more comprehensive sentiment analysis of social media could be fruitful. And as stated in the introduction, it may be that the mere existence of a team in a city, of which the effect was not measured here, that makes the people happier, and the performance of a team isn't a factor. But there are certainly many cities sans professional sports teams which subjectively appear plenty happy. So this just adds another reason to question the great cost a city government often goes to attract teams to play in their city.

¹ <http://dddb.net/eminentdomain/papers/appeal/AppellantBrief.pdf>

² <http://www.marketplace.org/topics/business/are-pro-sports-teams-economic-winners-cities>

³ http://college.holycross.edu/RePEc/spe/CoatesHumphreys_LitReview.pdf