

## **Final Report:**

Does the weather, my mood & Manchester United's form have an impact on my Youtube consumption?

By: Adé Obayomi

To: Professor Marc Stober et Neel

## Context

During the winter term which ranged from January to March of 2023, I took an Intro to Programming class where we were tasked with creating a data visualization using javascript and its d3 library. We were tasked with creating a cause and relation hypothesis and collecting data based on our hypothesis to visualize this data that will enable us to see if the data we collected leans towards supporting or disagreeing with said hypothesis.

While at first it was hard to come up with a hypothesis; I decided to stay true to things that interested me and frankly, data I would be eager to observe upon the completion of this project. For my project, I chose to collect data on the following variables and my rationales behind them can be seen below.

<b>Data</b>	<b>Rationale</b>
<b>Daily consumption of Youtube content in minutes</b>	I spend a lot of my free time on Youtube consuming content ranging from psychology, to sport highlights and most recently Y Combinator videos! I was interested in seeing what factors led to me spending more minutes on Youtube on certain days as opposed to other days.
<b>Mood Impact</b> Recorded using a scale of 0-5, with 5 being happiest	This might be a personal flaw, but I don't dabble too much with asking myself daily, "hey how am i feeling today?" I thought it would be important to also see how my daily emotional state has an impact on my Youtube consumption.
<b>Manchester United Impact</b>	About 50% of the reason I am on Youtube is to watch content relating to Manchester United. As aforementioned, sport highlights are a big part of the content I consume on Youtube. For the purpose of this research, I was curious to see on days where Manchester United played, how did the <b>result</b> affect my Youtube consumption and on days they don't play does that decrease my Youtube consumption?
<b>Job search outcomes</b>	As it's my last term at Northeastern, I have been applying for full time roles since the genesis of 2023. Job searches are definitely an emotional roller coaster and I was eager to see how this wild card could disrupt data for certain days.
<b>Daily Average Temperature in °F</b>	Winter term spans from Jan to March. February is typically very cold and in March we get some flashes of spring here and there. I wanted to see if

	on days where there was relatively warmer weather produced lower or higher Youtube consumption rates.
--	---

**So what was my hypothesis?**

By collecting these pieces of data, I wanted to find an answer to *whether the weather, my mood & Manchester United’s form* have an impact on my Youtube consumption? And if so, what is that impact? It was my *hypothesis* that **on days where Manchester United played a game and won, would also be the days where my Youtube consumption and my mood were at their peak.**

**How was this data collected?**

**Youtube consumption:**

While time wasn’t a factor in my data I counted a day as 12 am -11:59 pm. This was done because it made it easier to track my Youtube consumption as that is how Youtube records videos users watch within 24 hours. I collected data from February 11th to March 28th; which totals to about 40+ days worth of data.

In terms of tracking time on Youtube, Youtube provides a very intuitive *history feature* that allows users to essentially see the videos they consumed within a given day as well as the duration of these videos. I counted my time on Youtube based on the duration of each video watched; so essentially for the Youtube data my goal was to find how *many minutes* I spent on Youtube within a 24 hour time window.

**Mood Impact:**

While this is subjective data, I tried to stay true to my emotional states while recording this and collected data using a scale of **0-5**. To me, these are what each number and its emotional state represent:

<b>Mood Number</b>	<b>Mood Meaning</b>
0	I unfortunately must have moved on to the next phase of life because if I have 0 happiness, what is life?
1	The world is trying to take my smile away and I’m feeling super vulnerable right now
2	You know, things could be worse
3	Meh, Comme ci, Comme Ça
4	Ok, I can get used to days like this

**Manchester United:**

Days where both Manchester United played and didn't play were going to be tracked in this research but for days where Manchester United played, I also kept a record of *whether they won or didn't win* (Using boolean logic, in this case, a loss and a draw are equal).

**Job search:**

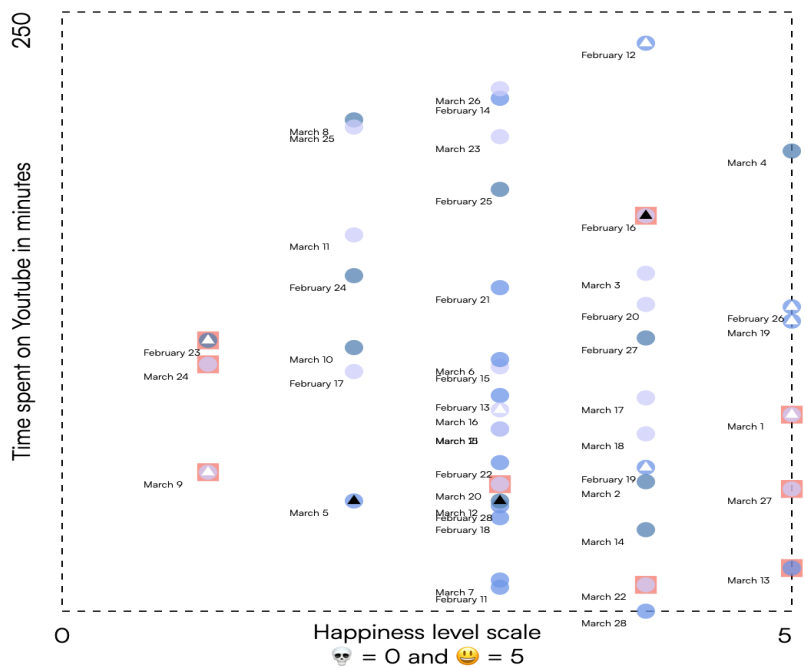
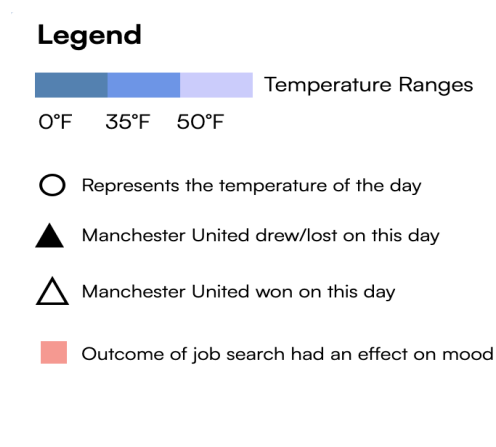
Initially this form of data wasn't being considered in my data. But after having a discussion about some of my findings with my professor and how some of those findings actually led to a paradoxical situation based off of my aforementioned hypothesis of my mood and youtube consumption being at its peak - when Manchester United won a game. I thought it would be worthwhile to document days where my job search had a positive or negative impact on my emotional state based on whether I progressed or didn't progress into further rounds of interviews I was invited to.

**Daily Average Temperature:**

I typically collected the temperature around evening time, but not with any strict regards in terms of time constraints. I relied on AccuWeather to provide me with the average daily temperature in Boston, Massachusetts in degrees Fahrenheit.

## Why I depicted my data with a Multivariate Scatter Plot

Screenshots of final visualization and its legend



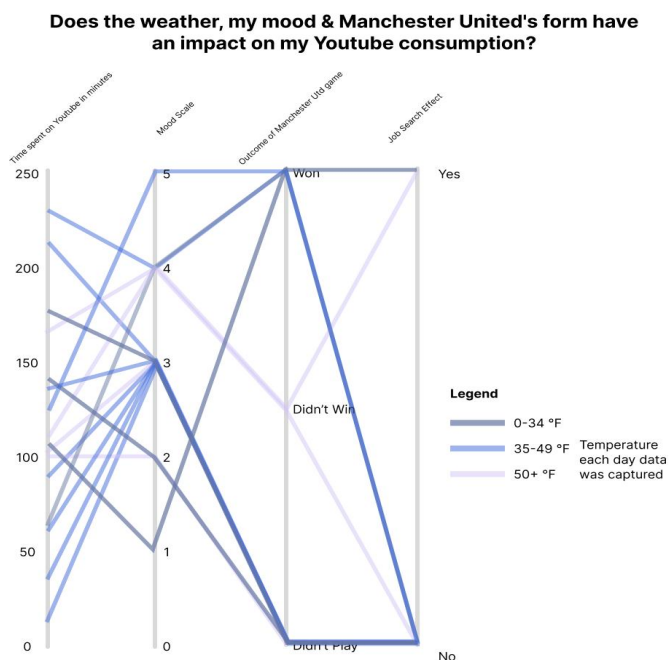
As it was my goal to know how different subjective and objective data affected my Youtube consumption, I knew employing a scatter plot would be necessary as it allowed me to show the correlation between the first two variables mapped to the x and y axes which were respectively the mood scale of 0-5 and my Youtube

consumption scale of 0-250 minutes. These variables will be depicted by their coordinated axes values. Then additional variables were mapped by employing visual characteristics such as color and shape to make it a more complex scatterplot that depicted how numerous variables contributed to the correlation of the primary variables (X,Y axis data).

Robert Kabacoff, a **data scientist** with more than 30 years of experience in multivariate statistical methods, says this method is called “Grouping” which allows one to plot data for multiple groups or contributing factors into a single graph. [1]

Even more so I thought using multivariate scatter plots would enable me to overcome one of the most common mistakes mentioned by data-to-viz .com with using scatterplots. I found that I wasn’t able to visually depict some forms of information which created interesting patterns in my data but by showing subgroups or multivariate scatter plots these hidden patterns can be revealed more intuitively.[2]

## An alternative approach: Parallel Coordinates Plot



The other data visualization approach I was intrigued by was the parallel coordinate plot. In the same way the multivariate scatter plot allows for several quantitative variables, so does the parallel coordinate plot. At first, I would say this approach helps with seeing things more *linearly* as the employment of only vertical axes within this approach actually improves readability as most humans read from left to right.

The problem with this approach though, is that quickly, as data starts to pile up, the lines begin to cover up each other’s paths. Even when you reduce the lines’ opacity it still begins to get hard to grasp the big picture of the data at a glance. I would have said this was a subjective opinion based on my visual claustrophobia but

even according to the data-to-viz article about parallel coordinates claimed, “it’s most common mistake of use cases is displaying too many samples as it just results in a cluttered and unreadable spaghetti chart.”[3] With 40+ days of data there would have been too many overlapping lines to make sense of the trends in data with ease.

When I compare the multivariate scatterplot to the parallel coordinate plot, I would say the scatterplot has the advantage of the X and Y axis and also the ability to show different variables with different shapes and the sizes & color of these shapes. The parallel plot just doesn’t have this sort of visual extensibility.

## **Analysis of my data: So does my Youtube consumption increase when I’m happier?**

(spoiler: it doesn't)

### **Youtube Consumption typically at its highest on days where I’m feeling a 3/5**

The thing about measuring your mood daily is you start to realize you have a lot of days where things for the most part, are emotionally eventless. These days, which are depicted on my scatterplot data visualization as the third column of plots, make up 40% of the 5 days where I spent the most time on Youtube.

*Embarrassingly enough, one of these days was on Valentine's day 😊.*

With that being said, the highest day of my Youtube consumption was on a day where I was feeling a 4/5 and where Manchester United won. This is the **only point** in my dataset that proves my hypothesis to be factual.

### **Manchester United winning games = bliss but that bliss isn’t indestructible due to job pursuits**

From February 11th to March 28th, Manchester United played a total of 11 games, where they won 72% or 8/11 of their games and either drew or lost the remaining 3 games. Manchester United played 24% of the time I was collecting data(46 days) and on days where Manchester United won their game, this resulted in 28% or 5/18 days of the time I felt a 4 or more out of 5 on my mood scale.

Circling back to my hypothesis; most of the days where Manchester United won and I was at my happiest, were days where my Youtube consumption was *actually average or lower than average*. When throwing the impacts of me trying to partake in the capitalistic rate race, there were 3 out of the 8 days where Manchester United won that my mood was altered by events from the job search as 2 of these 3 days were given a mood ranking of 1 with my Youtube consumption also being average. Meaning effects of the job search can also overturn my aforementioned hypothesis

### **Weather Impact on Youtube consumption could mean I need to start going out more**

You’d consider yourself an outgoing person till you find out the most days you are on Youtube are on the days where it’s relatively warmer. While on days where the weather was 50°F or more equated to 3/5 of the

days I spent the most time on Youtube, if I add the next highest day of my Youtube consumption to the mix, it increases that rate to be 3/6 of the days. In other words, 50% of the 6 days with the most recorded Youtube consumption were on days where the weather was at a warmer time of my data collection time period.

### **I use Youtube to elevate my emotions**

It also could be, now that I look at the dates, that I do have less workload as it's the last few days of the term. So even though the weather has been warmer, it's less the weather having an impact but me having more free time to spend on Youtube. But free time doesn't equate to happiness. Which means I most likely use Youtube to cover up some emotional void. If you could please look at the scatter plot 😊, you can see on March 24th my mood is low and the subsequent days after my Youtube consumption increases but as it increases so does my mood. And as my mood is increasing from the 26th of March till the 28th my Youtube consumption is on a downward trend. I have been working a lot on and procrastinating not only this report but my other capstone project so on the 28th I was very productive. The upward curve of my mood could also be from the remaining effects of a positive interview as seen on the 27th of March-or it's just from a feeling of checking things off my to-do list.

### **What other ideas could I have added to amplify this data: Sleep data and size of circles**

In retrospect, perhaps I could also have recorded the amount of sleep I had each day but I do think that might have been a challenge due to my unconventional sleeping patterns. I'm a power napper which means I take intermittent naps at night as I feel I'm most productive past midnight. If I were to include my sleep pattern in the data I would have represented that data by manipulating the size of the circles based on a 3 way threshold scale of hours of sleep to circle size ratio.

### **What I learnt and took from this project**

As aforementioned, I don't dabble too much with gauging my daily emotions but due to this project, I have begun to appreciate keeping track of this kind of data as analyzing it does bring some moments of self reflection-which is always a beneficial thing. It's funny, outside of class, I've presented my findings twice to friends but that's mostly because I like public speaking. But visualizing my habits and analyzing them could make for a good bimonthly Substack newsletter that I can share amongst my mutuals; that's if they'll care enough to analyze scatter plots and my findings every now and then. Perhaps I'll call it Adéta. Till next time.

## **Works Cited**

- 1.) Kabacoff, Robert. "Data Visualization with R - Github Pages." *Rkabacoff.github.io*, Rob Kabacoff, 1 Dec. 2020, <https://rkabacoff.github.io/datavis/Multivariate.html>.

2.) Healy, Yan Holtz and Conor. “Scatter Plot.” *Scatter Plot – from Data to Viz*,  
<https://www.data-to-viz.com/graph/scatter.html>.

3.) Healy, Yan Holtz and Conor. “Parallel Coordinates Plot.” *Parallel Coordinates Plot – from Data to Viz*, <https://www.data-to-viz.com/graph/parallel.html>.

### **Additional Resources**

1.) Cheekkallur, Aravind. “Adding a Title Attribute to SVG:G Element in d3.js.” *Adding a Title Attribute to Svg:g Element in D3.js*, Stack Overflow, 22 Aug. 2013,  
<https://stackoverflow.com/questions/11462029/adding-a-title-attribute-to-svg-element-in-d3-js>.

2.) “D3.js Symboltriangle Symbol.” *GeeksforGeeks*, GeeksforGeeks, 7 Sept. 2020,  
<https://www.geeksforgeeks.org/d3-js-symboltriangle-symbol/>.