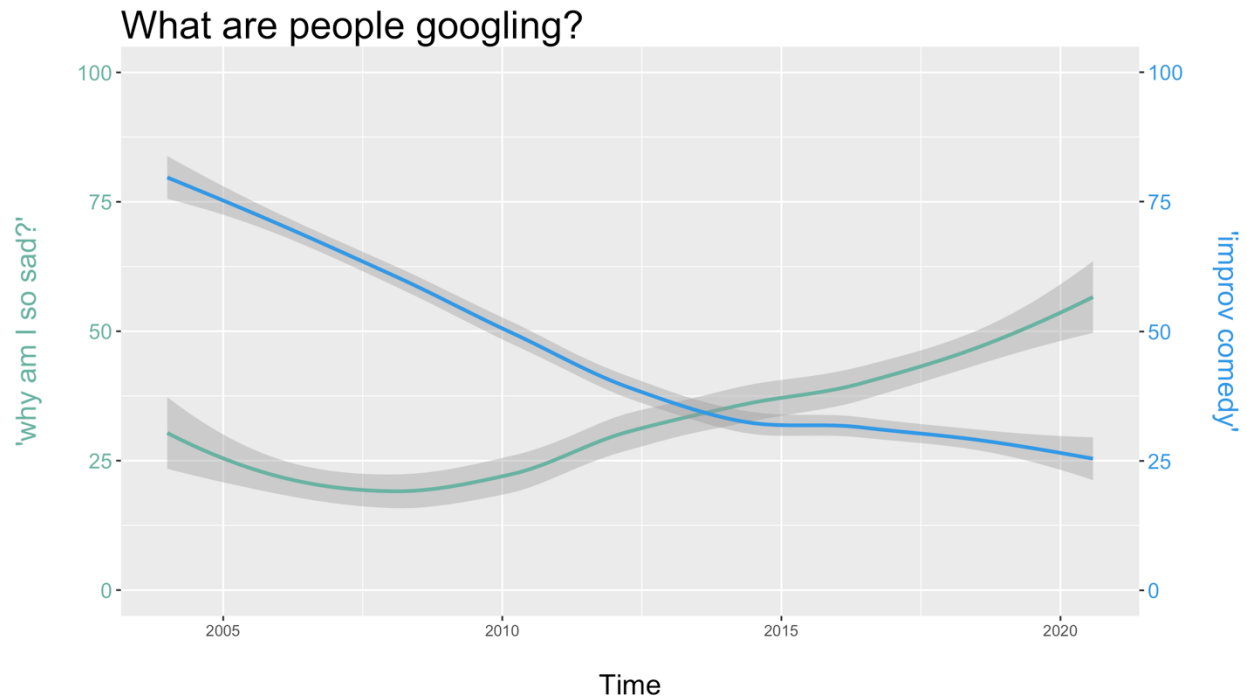# Coco's Correlations –

(Article 1)

When People Stop Learning Improv, they Become Sad

By Niki (Coco) Kalmus

Quick note: This week I used an open source program called RStudio to create my graphs – perhaps in the future I can teach a little R, Python, and Tableau. In the future I am absolutely going to use code to put a gif on a graph!!!! As we start, I want to bring you slowly into my nerdy world of loving data and making graphs. At the end of the article there is a link to my Github where you can copy my R code and make the graphs the way I did. Also, please let me know if you have any data or ideas that you'd like me to explore!
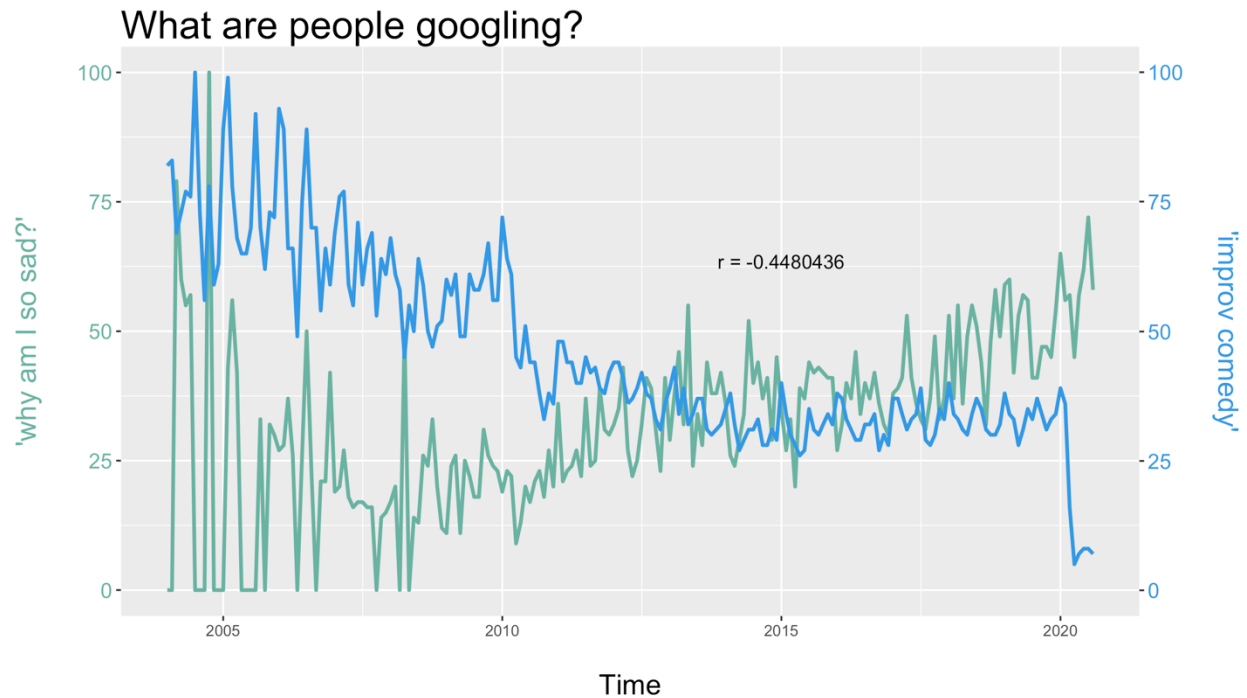
What are people googling?

Numbers represent search interest relative to the highest point on the chart for the given region and time.
A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.

Especially to the members of this Save Our Stage FCI group, the statement "When people stop learning about improv, they become sad," will come as no surprise. We all know the joy that improv brings us, the laughter… the fun!

Imagine that this article (the graph and the title) just came up on a social media platform. That's all you see. You'll go "yeah, duh! Everyone should do improv" and smash that share button. (Do people say that?) But let's take a step back… what is this graph really saying?

You know that people are Googling two phrases "why am I so sad?" and "improv comedy." It looks like people are Googling improv less over time and they are looking up reasons for their sadness more over time. But this is graph is extremely misleading for a million reasons.

Okay, just kidding it's only a few reasons.

The very first thing that is misleading is that this data is over time. This is called "time series" because it is a **series** of data points displayed over **time.** With time there are often a lot of units of observation – seconds, minutes, days, seasons, years, etc.! Google receives over 63,000 searches per second on any given day… this means that the graph I showed you has been **smoothed out**. It's more of an average or a sum over time (I won't go into detail just yet on what I did… you're just dipping your toes in today!)… What does it look like if I show you **actual daily data over the past 16 years?**

What are people googling?

r = -0.4480436

y-axis (left): 'why am I so sad?'
y-axis (right): 'improv comedy'

Time

Numbers represent search interest relative to the highest point on the chart for the given region and time.
A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.

Oh dear. Okay, that's… very different. You can see that the trend overall is the same – people are Googling "why am I so sad" more and "improv comedy less." But there are huge ups and downs as time passes. That's a ton of **variation**.

Do you see the "r= - 0.4480436" on the graph? This is called a **correlation coefficient.** It is a number between -1 and 1 that tells you how strong the relationship is between the two variables (in our graph, improv and sadness). 1 is a very strong correlation and 0 is pretty weak. If this is negative that means that as one goes up the other goes down (like in this graph) and if it is positive it means the two variables move in the same direction. Negative 0.4 is a pretty WEAK negative correlation! We can see why it's not very strong by looking at the daily data – there is too much **variation.** They really aren't moving together over time at all. So, it is absolutely misleading to call this article "*When People Stop Learning Improv, they Become Sad.*" But I did it anyways because… freedom of speech?

Maybe you have heard the phrase "correlation does not equal causation." Great! If not – no worries… now you have! It means that just because correlation exists (or no correlation at all) it **does not mean** that one definitively causes the other. Causation is pretty damn hard to prove (but there are lots of very smart people that can theorize their way to causation). So even if I didn't point that sneaky trick I used to make you think when people learn about improv via Google that they are less sad (as shown by their Google search)… we **cannot say** (statistically at least) that improv CAUSES less sadness. *Of course, we all know that improv sparks joy within us… but there are just so many things that could cause sadness besides a lack of improv, such as losing your favorite sweatshirt or the end of Schitt's Creek.*

In conclusion for this week: **be careful when you see a graph** – it might be misleading *on purpose*. Ask yourself, "Is this a time series graph? How did they aggregate the data? Who made this? Do they have any reason to try to mislead me?" The best thing to do is read on, try to learn their methodology, and think critically! Boom! You're a savvy consumer of graphs now!

Before I leave, though… check this out:

What are people googling?
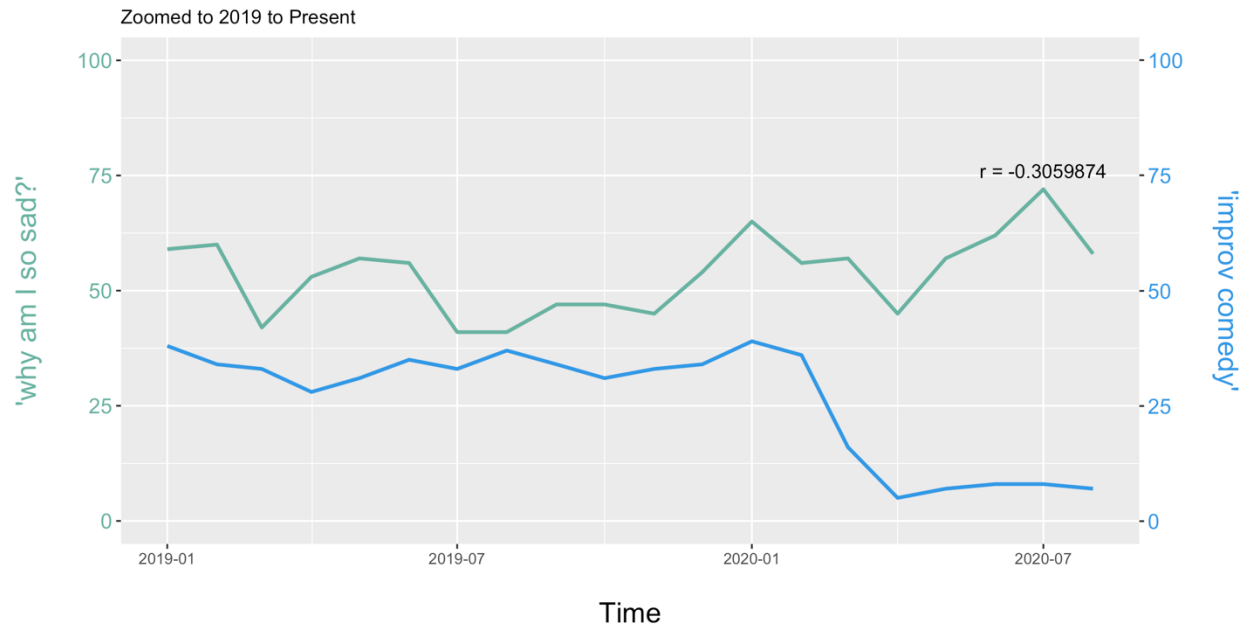
Zoomed to 2019 to Present

r = -0.3059874

Time

Numbers represent search interest relative to the highest point on the chart for the given region and time.
A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.

Around March 2020 there is a spike in Googling "why am I so sad?" and there is a huge decrease in Googling "improv comedy." Don't forget… correlation does not equal causation. But… what else happened March 2020? QUARANTINE. Sadness spike, improv decrease… I don't know… draw your own conclusions. ☺

RESOURCES:

- See what people are googling: https://trends.google.com/trends
- You can see the code I used to create these graphs here: https://github.com/nikikalmus/FCI-SaveOurStage2020 (I made them in R – more on that in later articles).