Confession time: years ago, I was skeptical of podcasts. I was a music-only listener on commutes. Can you imagine? But around 2016, I gave in and finally took the plunge into podcasts. And I’m so glad that I did.

Since then, I’ve seen enormous benefits, all attributable at least in part to the podcasts I’ve listened to. I’ve improved my programming, learned new skills, and started multiple income-generating businesses.

Naturally, I’ve been interested in data science podcasts. Initially, I found my data science podcasts through people I followed on Twitter whose shows I knew about. But I’m always on the lookout for new podcasts, so I took to the internet to find recommendations.

If you’re a podcast listener, you’ve probably learned that the current state of podcast discovery leaves much to be desired. It’s very hard to find new podcasts, especially for niche fields like data science and analytics.

So of course, when I tried searching for interesting data science podcasts, I kept finding the same shows recommended everywhere. And usually they were the podcasts I was already listening to!

I decided that if I wanted a bigger list of data science podcasts, I needed to go deep. I searched nearly 100 lists for recommended podcasts, and used them to compile **the most complete list of data science podcasts** you will find.

[Click here for the full list of 41 podcasts](https://mailchi.mp/d18f2f50ca14/data-science-podcasts)

In this post I’m going to talk about how I collected this list and analyze the results.

**Gathering a list of data science podcasts**

I knew I wanted to build a big list. But I also knew that not all podcasts are created equal, and that I needed some way to differentiate the truly great podcasts on the list. I decided I would use podcast recommendations as a form of social proof, so that the most recommended podcasts would bubble to the top of the list. Here’s the method I decided to follow:

1. I’d generate a list of search terms
2. I’d perform a google search for each term on the list
3. I’d open each of the top 10 Google links for that search term and note all the results
4. I’d aggregate the complete results by podcast and use the total number of recommendations to create a “recommendation score”. Higher scores should, in theory, represent better podcasts. Or at least more well known podcasts.

**Generating a list of search phrases (Step 1)**

This step was relatively easy. I knew I wanted a list of data science podcasts, but I also wanted to include more specific subgenres like data visualization, as well as more broad but related topics like Python programming or SQL & Databases. I settled on this list of phrases:

* Data Science Podcasts
* Data Engineering Podcasts
* Data Visualization Podcasts
* Analytics Podcasts
* SQL Podcasts
* R Podcasts -reddit
* Python Podcasts

Pretty straightforward. The only curious bit is with the R Podcasts query. The Reddit url structure is such that reddit.com/r/podcasts leads to the podcasts subreddit. Without the explicit -reddit, all links were to that subreddit and unrelated to R programming.

**Searching Google and aggregating (Steps 2-4)**

Next, I’d search each of these phrases in Google and open up the first 10 sites that popped up in my search results.

Often times these were lists others had created of, for example, “top 5 data science podcasts”. I copied down the list of podcasts and kept an ongoing tally for how many times I’d seen a particular podcast represented.

Occasionally, the Google search would yield links to specific podcasts rather than to lists of podcasts. In this case, I would record this as well. My thinking is that ranking on Google for a particular search phrase is at least as reliable an indicator of podcast quality as inclusion on a “best data science podcasts” list. That said, this did particularly benefit those podcasts whose names closely matched my search query, as was the case with [Data Engineering Podcast](https://www.dataengineeringpodcast.com/) and the [The R-Podcast](https://r-podcast.org/).

After going through all of the search queries, I had a list of podcast recommendations along with a tally of how often the podcast had been recommended in search results.

Finally, I removed any podcasts that had only 1 appearance in the results. There were a large number of these, and I felt it would dilute the value of the list to include them. This was somewhat arbitrary, but I think it makes for a stronger list overall.

**Analyzing the list of best data science podcasts**

library(hrbrthemes)

library(tidyverse)

extrafont::loadfonts(quiet = TRUE) # Needed for hrbrthemes / mac dependency issue

podcasts <- read\_csv('~/Desktop/best\_podcasts.csv')

First I read in my compiled list of podcasts, then I used ggplot to graph the results.

ggplot(podcasts, aes(x = reorder(Title, Score), y = Score)) +

geom\_bar(stat = 'identity') +

coord\_flip() +

labs(title = 'Top Data Science Podcasts',

subtitle = 'Rankings crowdsourced from Google search results',

x = '',

y = 'Recommendation Score',

caption = 'michaeltoth.me / @michael\_toth') +

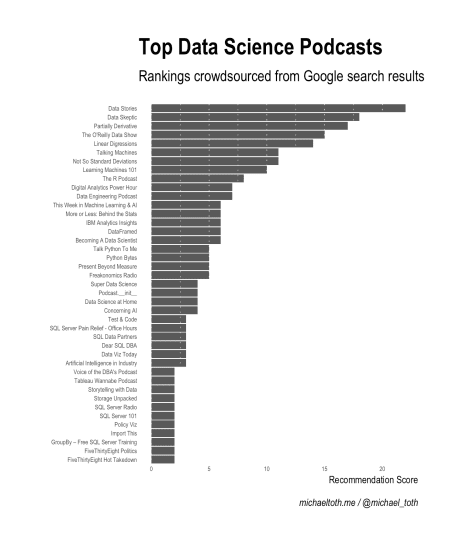
theme\_ipsum(base\_size = 5, caption\_size = 8, axis\_title\_size = 8) +

theme(panel.grid.major.y = element\_blank(),

panel.grid.major.x = element\_line(colour = 'white', linetype = 'dotted'),

panel.grid.minor.x = element\_line(colour = 'white', linetype = 'dotted'),

panel.ontop = TRUE)



This is awesome! This list of 41 podcasts is more than I found anywhere else online while going through this exercise, and I’m confident it’s the most extensive list of data science podcasts you can find.

I had heard of many of the podcasts in the top 10–Data Stories, Data Skeptic, Partially Derivative, The O’Reilly Data Show, and Not So Standard Deviations–but others were new to me! I hadn’t heard of Linear Digressions, or Talking Machines, or Learning Machines 101. While I haven’t yet had a chance to listen to these, I’m excited to check them out!

So now we have a list of 41 data science podcasts to review. This is great, but I think we can do better! While these podcasts are all loosely related to data science, they’re quite different from one another in focus. For example, FiveThirtyEight Politics is a relatively mainstream podcast, Data Stories deals with topics in data visualization, and Data Skeptic is an all-around instructional data science podcast.

I thought some form of categorization would help guide my listening, so I created a broad list of categories and grouped them as best as I could. I decided I would group the podcasts into 8 categories:

* General Data Science and Analytics
* Relevant Mainstream Podcasts
* Machine Learning & AI
* Data Visualization
* Data Engineering
* SQL & Databases
* R Programming
* Python Programming

I went through this list and categorized them as best as I could according to topic. Equipped with these new categories, let’s take a look at the list of recommendations:

ggplot(podcasts, aes(x = reorder(Title, Score), y = Score, fill = Category)) +

geom\_bar(stat = 'identity') +

coord\_flip() +

scale\_fill\_manual(values = c("#a6cee3","#1f78b4","#b2df8a","#33a02c","#fb9a99","#e31a1c","#fdbf6f","#ff7f00")) +

labs(title = 'Top Data Science Podcasts',

subtitle = 'Rankings crowdsourced from Google search results',

x = '',

y = 'Recommendation Score',

caption = 'michaeltoth.me / @michael\_toth') +

theme\_ipsum(base\_size = 5, caption\_size = 8, axis\_title\_size = 8) +

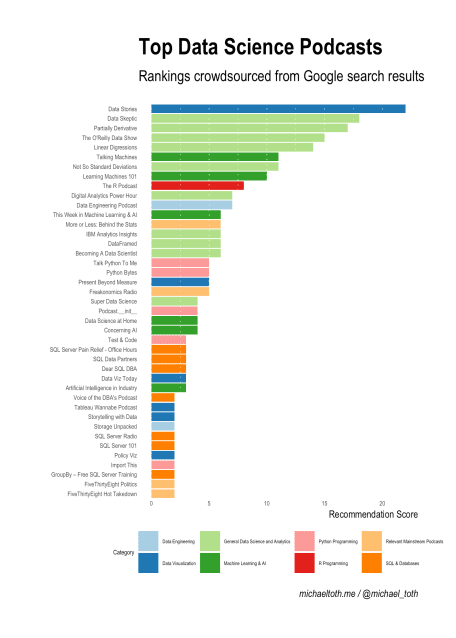
theme(panel.grid.major.y = element\_blank(),

panel.grid.major.x = element\_line(colour = 'white', linetype = 'dotted'),

panel.grid.minor.x = element\_line(colour = 'white', linetype = 'dotted'),

panel.ontop = TRUE,

legend.position = 'bottom')



While most of the top 10 are made up of what I’m calling general data science and analytics podcasts, Data Stories stands alone as the number one podcast and the only data visualization podcast in the top 10. Most of the language-specific podcasts are lower on the list, but that’s likely due to the fact that they are more specific in nature and not necessarily due to lower quality. The R podcast was the only language-specific podcast to crack the top 10, so I’m definitely going to check that one out!

Do you have favorite podcasts that aren’t included here? Let me know in the comments. Also let me know how you discover new podcasts, I’d love to improve my discovery! If you’re interested, you can get the full list of data science podcasts below: