

APR. 14, 2014, AT 10:21 AM

A Statistical Analysis of the Work of Bob Ross

By [Walt Hickey](#)

Filed under [Happy Spreadsheets](#)

Get the data on [GitHub](#)



Bob Ross in 1985. THE JOY OF PAINTING, THE BOB ROSS NAME AND IMAGES ARE TRADEMARKS OF BOB ROSS INC. ALL RIGHTS RESERVED. USED WITH PERMISSION.

Bob Ross was a consummate teacher. He guided fans along as he painted “happy trees,” “almighty mountains” and “fluffy clouds” over the course of his 11-year television career on his PBS show, “The Joy of Painting.” In total, Ross painted 381 works on the show, relying on a distinct set of elements, scenes and themes, and thereby providing thousands of data points. I decided to use that data to teach something myself: the important statistical concepts of conditional probability and clustering, as well as a lesson on the limitations of data.

So let’s perm out our hair and get ready to create some happy spreadsheets!

What I found — through data analysis and an interview with one of Ross’s closest collaborators — was a body of work that was defined by consistency and a fundamentally personal ideal. Ross was born in Daytona, Fla., and joined the Air Force at 17. He was stationed in Fairbanks and spent the next 20 years in Alaska. His time there seems to have had a significant impact on his preferred subjects of trees, mountains, clouds, lakes and snow.



Paintings by Bob Ross featured on PBS's "The Joy of Painting." THE BOB ROSS NAME AND IMAGES ARE TRADEMARKS OF BOB ROSS INC. ALL RIGHTS RESERVED. USED WITH PERMISSION.

Of the 403 episodes of "The Joy of Painting" — whose first run was from 1983 to 1994 and which continues to air in reruns on PBS stations nationwide — Ross painted in

381, and the rest featured a guest, most frequently his son Steve Ross. Based on images of Bob Ross's paintings available in the Bob Ross Inc. store, I coded all the episodes^x

Of the 403 total episodes, I was not able to see the completed work of three paintings: season 9 episode 10, "Country Charm"; season 15 episode 4, "Peaceful Reflections" and season 26 episode 10, "Purple Mountain Range."

using 67 keywords describing content (trees, water, mountains, weather elements and man-made structures), stylistic choices in framing the paintings, and guest artists, for a grand total of 3,224 tags.^x

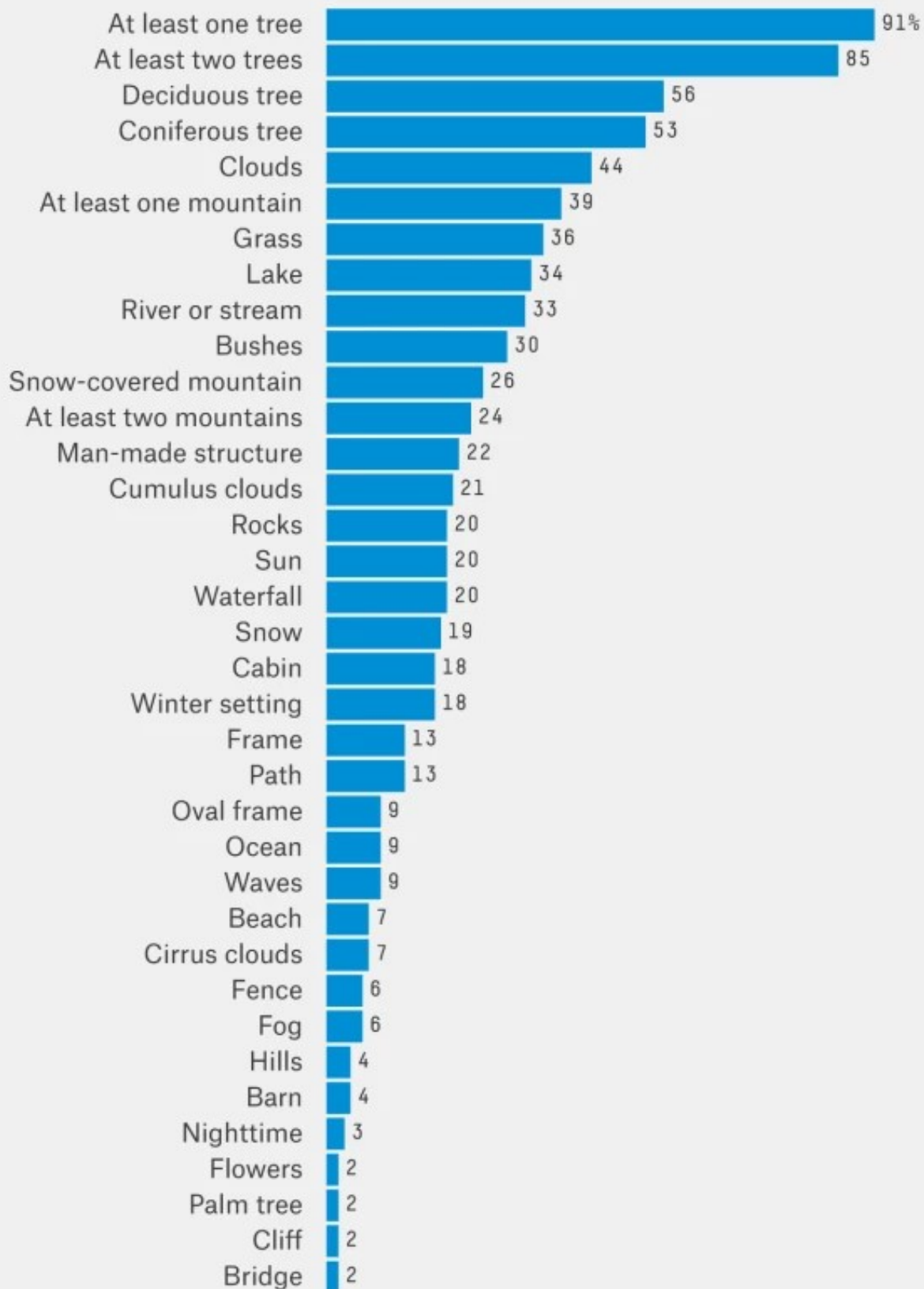
This data set remains a work in progress — it's the first of its kind — and there is of course the potential for omissions. It would take just over eight straight days to watch all of "The Joy Of Painting," so it's a task ill-suited for one person. But I'm confident that the data as it stands describes the work over Ross's career accurately and consistently.

I analyzed the data to find out exactly what Ross, who died in 1995, painted for more than a decade on TV. The top-line results are to be expected — wouldn't you know, he did paint a bunch of mountains, trees and lakes! — but then I put some numbers to Ross's classic figures of speech. He didn't paint oaks or spruces, he painted "happy trees." He favored "almighty mountains" to peaks. Once he'd painted one tree, he didn't paint another — he painted a "friend."

Here's how often each tag that appeared more than five times showed up over the 381 episodes:

The Paintings of Bob Ross

Percentage containing each element



Now that we know the basic probabilities of individual tags, we can also find the *joint probabilities* of some of these events. For instance, how often do a deciduous tree and a coniferous tree appear in the same painting? We know that 57 percent of paintings contain a deciduous tree and 53 percent of paintings contain a coniferous tree. According to our data set, 20 percent of paintings contain at least one of each.

What's more, we can also find the probability that Ross painted something given that he painted something else, a statistic that's called *conditional probability*.

Conditional probability can be a bit tricky. We know that 44 percent of Ross's paintings contain clouds, 9 percent contain the beach and 7 percent contain both the clouds and the beach. We can use this information to figure out two things: the probability that Ross painted a cloud *given* that he painted a beach, and the probability that he painted a beach *given* that he painted a cloud. You divide the joint probability — 7 percent in this case — by the probability of the *given* — 44 percent or 9 percent, depending on whether you want to know the probability of a beach given a cloud or a cloud given a beach.

The biggest pitfall people often face is assuming the two probabilities are the same. The probability that Ross painted a cloud given that he painted the beach — essentially, how many beach paintings have clouds — is $(0.07)/(0.09)$, which is 78 percent. The vast majority of beach scenes contain clouds. However, the probability that Ross painted a beach given that he painted a cloud — or, how many cloud paintings contain a beach — is $(0.07)/(0.44)$, or 16 percent. So the vast majority of cloud paintings don't have beaches.

I figured out the conditional probability of every Bob Ross tag against every other tag to answer the following pressing questions.

What is the probability, given that Ross painted a happy tree, that he then painted a friend for that tree?

There's a 93 percent chance that Ross paints a second tree given that he has painted a first.

What percentage of Bob Ross paintings contain an almighty mountain?

About 39 percent prominently feature a mountain.

What percentage of those paintings contain several almighty mountains?

Ross was also amenable to painting friends for mountains. Sixty percent of paintings with one mountain in them have at least two mountains.

In what percentage of those paintings is a mountain covered with snow?

Given that Ross painted a mountain, there is a 66 percent chance there is snow on it.

What about footy little hills?

Hills appear in 4 percent of Ross's paintings. He clearly preferred almighty mountains.

How about happy little clouds?

Excellent question, as 44 percent of Ross's paintings prominently feature at least one cloud. Given that there is a painted cloud, there's a 47 percent chance it is a distinctly cumulus one. There's only a 14 percent chance that a painted cloud is a distinctly cirrus one.

What about charming little cabins?

About 18 percent of his paintings feature a cabin. Given that Ross painted a cabin, there's a 35 percent chance that it's on a lake, and a 40 percent chance there's snow on the ground. While 72 percent of cabins are in the same painting as conifers, only 63 percent are near deciduous trees.

How often did he paint water?

All the time! About 34 percent of Ross's paintings contain a lake, 33 percent contain a river or stream, and 9 percent contain the ocean.

Sounds like he didn't like the beach.

Much to the contrary. You can see the beach in 75 percent of Ross's seaside paintings, but the sun in only 31 percent of them. If there's an ocean, it's probably choppy: 97 percent of ocean paintings have waves. Ross's 36 ocean paintings were also more likely to feature cliffs, clouds and rocks than the average painting.

What about Steve Ross?

Steve seemed to prefer lakes far more than Bob. While only 34 percent of Bob's

paintings have a lake in them, 91 percent of Steve's paintings do.

One useful lens we can apply to this sort of data — where we're comparing vectors of information — is a clustering tool. The idea behind clustering is to determine how close certain groups of data are to other points in the data set. Researchers use clustering analysis in all sorts of areas — from biology to consumer marketing — as a way of segmenting a population of, say, plants or people. It allows us to find interesting subsets of data based on how similar or different certain subgroups are from the rest of the set.

I used an algorithm to divide the entire set of 403 paintings from “The Joy of Painting” into clusters of similar paintings. I wanted to know whether it was possible to identify the 10 basic paintings featured on the PBS series. To do this, I ran a [k-means](#) clustering analysis of the paintings.³ The results were mixed.

First, let's look at the clusters that make intuitive sense. The clear winners are:

- A cluster of 50 paintings tagged “snow” and “winter”
- A cluster of 28 paintings each with an oval white-space frame
- A cluster of 35 paintings of ocean scenes.

These were the kinds of clear clusterings we were hoping to find. Each has a common theme and falls under the banner of iconic Bob Ross images. He painted about one beach scene and one oval-framed image per season, and about two scenes with snow in the foreground per season. It makes sense.

Here are some clusters that also make sense, but don't tell us a whole lot about Ross's favorite kind of painting:

- A cluster of 13 paintings by guest host Steve Ross
- A cluster of 7 paintings containing a bridge
- A cluster of 11 paintings containing flowers
- A cluster of 30 paintings containing a fence or a barn
- A cluster of 33 paintings containing a waterfall.

These clusters identify some tags that appear in only a few paintings, but the groupings are not supremely helpful in defining what Ross painted. For example,

flowers were very rarely the main focus of a painting, and we already knew how many times Steve Ross appeared on the program.

The final two clusters were the most broad:

- A cluster of 95 paintings that had trees and at least one mountain
- A cluster of 103 paintings that had trees but no mountains.

Not supremely helpful, but still quite interesting. Clustering analysis is an appealing tool for this kind of data but hardly has all the answers.

To learn more about Ross and his work beyond what I already knew from the data, I called Annette Kowalski, who founded Bob Ross Inc. with the painter and remains the steward of his work.⁴ She confirmed something I had discovered in my review of hundreds of Ross's landscapes: His work isn't defined by what is included in his paintings, but by what's excluded.

"I can think of two times he painted people," Kowalski said. "There was a man by a campfire,⁵ and two people walking through the woods."⁶ Indeed, our data shows that Ross only painted a person — in silhouette against a tree near a campfire — one time.



One of the few Bob Ross paintings in PBS's "The Joy of Painting" that featured a person. THE BOB ROSS NAME AND IMAGES ARE TRADEMARKS OF BOB ROSS INC. ALL RIGHTS RESERVED. USED WITH PERMISSION.

When we analyze the structures he painted, it appears Ross preferred the simple to the elaborate. He painted 69 cabins, 25 fences in various states of disrepair and 17 barns. More complex man-made structures are remarkably rare in his work. Bridges appear only seven times. Boats and mills, twice each. Ross painted one dock, one lighthouse and one windmill over his 381 episodes.

There's something about the structures Ross painted that has gone almost entirely unnoticed by fans, according to Kowalski.

"I will tell you Bob's biggest secret. If you notice, his cabins never had chimneys on them," she said. "That's because chimneys represented people, and he didn't want any sign of a person in his paintings. Check the cabins. They have no chimneys."

She immediately added, "I'm sure you're going to call me tomorrow and say you found a chimney." And I did! But it took a lot of hunting. In season 7 episode 1, "Winter Cabin," there's a chimney on the cabin (featured above in the third row, center column). But the fact that a chimney appeared once in 381 paintings doesn't really

diminish Kowalski's point.

When it comes down to it, "The Joy of Painting" was never really about painting. Even Kowalski, who runs a company that sells Bob Ross-branded painting supplies, believes most viewers aren't in it for the art.

"The majority of people who watch Bob Ross have no interest in painting," she said. "Mostly it's his calming voice."

Footnotes

1.

Of the 403 total episodes, I was not able to see the completed work of three paintings: season 9 episode 10, "Country Charm"; season 15 episode 4, "Peaceful Reflections" and season 26 episode 10, "Purple Mountain Range."

2.

This data set remains a work in progress — it's the first of its kind — and there is of course the potential for omissions. It would take just over eight straight days to watch all of "The Joy Of Painting," so it's a task ill-suited for one person. But I'm confident that the data as it stands describes the work over Ross's career accurately and consistently.

3.

The *k*-means algorithm is what we call non-deterministic. This means you'll get a slightly different result each time, because of the randomness the algorithm factors in when determining the points that define the centers of the clusters.

4.

Kowalski was also a guest artist for season 29 episode 10, "Pot o' Posies.💎💎💎

5.

Season 3, episode 10 "Campfire"

6.

This second painting did not appear in my data set, but it's entirely possible I missed it when tagging.