K Nearest Neighbor Establishing shot:

Yogi Bearesque ranger station

"Welcome Rangers and friends of the parks."

Next page, same scene:

Ranger (from inside), "Welcome rangers and friends, we have hired consultants, whose reputations are... curious if not cheap, StatCat and DataDog from ACME Data Services. We have, as you know, discovered a pair of exciting new species of Flora and Fauna, what appears to be a wild Iris, and a frog.

We hope that at the end of the hour, we will all have learned something. Refreshments are available at the table along the back."

If budget and page count allows, simple shot of table with a pitcher of water, 6 plastic cups, a bag of pretzels, 2 cans of ginger ale and a bottle of ketchup.

Return to exterior.

"Welcome, Statcat!"

Stage (to be used over and over)

StatCat takes stage in most dramatic pose possible, waiting for applause.

If budget and page count allows: Shot of audience, a medium size cabin room with rows of benches and 4 or 5 people seated far apart, all eating pretzels.

Return to StatCat posing. Sound of crickets in background.

Stage with StatCat: Undeterred, he begins his lecture in full form. He first tells them how he and DataDog used the K-Nearest Neighbors algorithm to classify Iris flowers. He then lectures on the basic ideas behind this algorithm, giving a quick introduction.

StatCat: "Using the k-NN (that's K nearest neighbor to those of you who missed my last lecture, 'The statistical and social significance of abbreviations in a post truth narrative,'"

He then explains how this algorithim helps to identify species by observing the neighbor nearest them.

Question from audience, using our audience shot, or DataDog asks question.

"So what if the nearest neighbor to our undocumented frog are flamingoes."

StatCat: "Well then, obviously, of we do the science..."

Perhaps begins an absurd equation on the chalkboard or overhead projector.

"...what we would come up with would be... that this would obviously have to be some sort of strange frogmingo..."

We see picture in his mind of frog standing high on flamingo legs.

"No, of course not, that would be ridiculous, and nature, apart from the aardvark and platypus, tends to avoid the ridiculous."

"What about the sloth and the hairless cat?" asks DataDog.

Audience shot, different people speaking up:

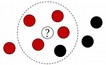
"And the probiscus monkey!"

"And how about the blobfish!"

"Oooh... and the donkey!"

"But we digress," StatCat returns to his lecture using the information in blue below.

K-Nearest Neighbors (k-NN) is an algorithm that classifies a data point based on the classification of its neighbors. This means that if a data point is surrounded by four red points and one black point, then majority vote would suggest that the data point is likely red.



The k in k-NN is a parameter referring to the number of nearest neighbors to include in the majority vote. In the example above, k equals five and includes all of the colored circles within the dashed line. Given this scenario, what color do you think the algorithm will predict for the circle with the question mark? Yes, the algorithm will predict that it’s red.

But how many neighbors should be considered? That is, what value should k be set to? Choosing the right value of k is a process known as parameter tuning, and it is critical to prediction accuracy.

"And, now well into the future The Jetsons promised, we can use A.I. to help us gather the most accurate and unbiased data possible. Allow me to introduce my trusty sidekick... DataDog.

"Sidekick" mutters DataDog as she takes the stage.

One look at the audience of 4 and she dashes off stage in terror.

Standing on stage, StatCat tries to coax DataDog back on stage, but she is not having it.

Comical way to resolve this (yet to be devised... from offstage, scenes of empty stage with DataDog's work balloon pointing off stage; DataDog unable to speak, sits on StatCat's lap like a ventriloquist dummy while StatCat reads off the information; or some better idea.

That information is also given.

The Ranger and audience is duly impressed. Indeed, he believes that the unknown flower in the park is a relative of the Iris and is certain that the algorithm will work in this case too.

StatCat, "Now let's peruse these new species of frog and flower before our return to the StatCave."

They bring out the new iris and frog.

StatCat begins to take the frog out of the jar, everyone gesturing for him to stop.

Frog leaps onto the iris.

Iris swallows it in one gulp.

StatCat: "Well, yes... I see, well, let's look at the bright side, appears I've saved you a lot of work, don't ya know, now you only have to identify this new plant... and thanks to me, you now know it was most definitely not an iris."

Shot of 4 audience members, jaws dropped.

DataDog: "Gee, I hope that wasn't the only one of its kind." Shot of 4 audience members, jaws dropped.

StatCat: "Yes, well, we'll just leave this here," he says, dropping the A.I. data collecting frogbot into the jar. "Should come in handy, don't ya know."

Shot of ranger and audience members, still jaw dropped and silent.

StatCat: "Yes, well, 'tis getting late after all. Back to the StatCave DataDog! We'll send you our bill in the post."

Back at the DataDen, simple shot(s) of DataDog talking to the students.

"Fortunately, they forgave us, and I've finally gotten the data they gathered to identify our hungry new flower. Good thing for you our frog friend was eaten, that cuts your homework in half!"

"So let's get to work analyzing this data! And... if you get all the way to the end... you get a prize! A shiny data-collection A.I. Frogbot all your own!

DataDog guided learning event.

Final page, gaudy giff of glowing mechanical frog.

They are brought out.

* Data Collection Scene

StatCat, the Ranger, and BooBoo head out to collect samples of the park’s flowers and return with a basket of what appear to be three distinct species. Meanwhile, DataDog builds a miniature A.I. frog to measure each of the flowers and then generate an Iris-like dataset.

* Data Learning Event
* A.I. frog takes measurements of the unknown species, feeds it to the algorithm, and it responds with a prediction. Everyone is excited to finally know what species this flower is.

But while focused on the computer output, the unknown flower swallows DataDog’s A.I. frog. Apparently, this new species is a carnivore, a member of the Venus Flytrap family!

A.I. Tour of Applications

Flower Story – K-Nearest Neighbors

Goldilocks Story – Overfit, Underfit, and Just Right