ALGORITHMIC BIAS THROUGH DATASETS

Using Google's Quick, Draw! neural net to introduce and apply machine learning concepts

Activity designed and delivered by:
Emily Springer, PhD

Creative Common License (CC-BY)

- References:

Quick, Draw! – [the game]

Quick, Draw! – [the dataset]

McConchie, Alan. [PopvsSoda.com]

The course and the activity:

- This hands-on activity was taught at Arizona State University in Fall 2020 and Spring 2021 semesters for a sociology course I created and delivered titled *Technology and Society*.

- Comments from undergraduate students, Arizona State University (2021):

"It's fun. I just played the game but going over it now, it's eye opening to see how it can apply."

"I love that they give us the data set. It's helpful to see how we match up against a crowdsource dataset. Often, these are things we don't have visibility into."

"I thought this was a pretty good representation of data points and how they work in an algorithm."

Quick, Draw! activity in context:

- This brief hands-on activity serves as a segue between two units. The first unit builds student knowledge of algorithms and introduces machine learning concepts. The second unit covers algorithmic bias in society along race, class, and gender dimensions using both U.S. and global examples.
- This activity serves as the entry point to conceptualizing how algorithmic bias may happen from a socio-techno perspective.
- This activity can be done in 60 minutes in an undergraduate virtual classroom.

UNIT 1: Machine learning basics → THIS LESSON: Quick, Draw! Activity → UNIT 2: Algorithmic bias

Activity Objectives

1. **Reinforce machine learning concepts and components from Unit 1 (data points, datasets, algorithmic models, prediction abilities) in an applied, hands-on format.**
   - 1.1: To consider the associations that can be created across numerous datapoints.
   - 1.2: To demonstrate that algorithms can predict differences in ways that are not readily apparent to humans.
   - 1.3: To apply the importance of the dataset to how an algorithm defines “success.”

2. **Introduce concepts around bias and discrimination in a friendly, accessible format.**
   - 2.1: To demonstrate that consensus is not always present in datasets.
   - 2.2: To demonstrate there is not consensus in data labeling.
   - 2.3: To demonstrate that not everyone is equally seen by an algorithm.

In advance of activity:

Requirements for Instructor:

- Be familiar with sharing functionality of your online platform.
- Ensure teacher and students have access to share screen.
- Be prepared to draw a soda/pop/coke can live.
- Build a poll in advance: What did you call the drawing? 1. Soda 2. Pop 3. Coke 4. Other.

Requirements for Students:

- As homework, prior to coming to class, students are asked to play Quick, Draw! with Google's neural net. In one round of play, users are asked by a machine learning algorithm to draw 6 different words, offering the user 20 seconds to draw the image, ending when either 20 seconds has passed or when the algorithm correctly guesses the prompted word.

Logistical notes:

- During the activity, students are shown smaller subsets of the larger Quick, Draw! Datasets by data label, usually around 80 datapoints, to make discussion easier. See their website for the full dataset by label.
- In this deck, students are only shown the slides which feature Quick, Draw! datasets (cats, broccoli/tree, telephone, traffic/streetlight, hurricane, and tornado). All text slides are for instructional and demonstration purposes only.

ACTIVITY OPENER

- OPENER: Now you all played Quick, Draw! with a machine learning algorithm. Was anyone able to successfully get the algorithm to guess their drawing in all 6 tries?

- QUESTION: When we play Quick, Draw! what are we, as users or players, creating? What is our role?

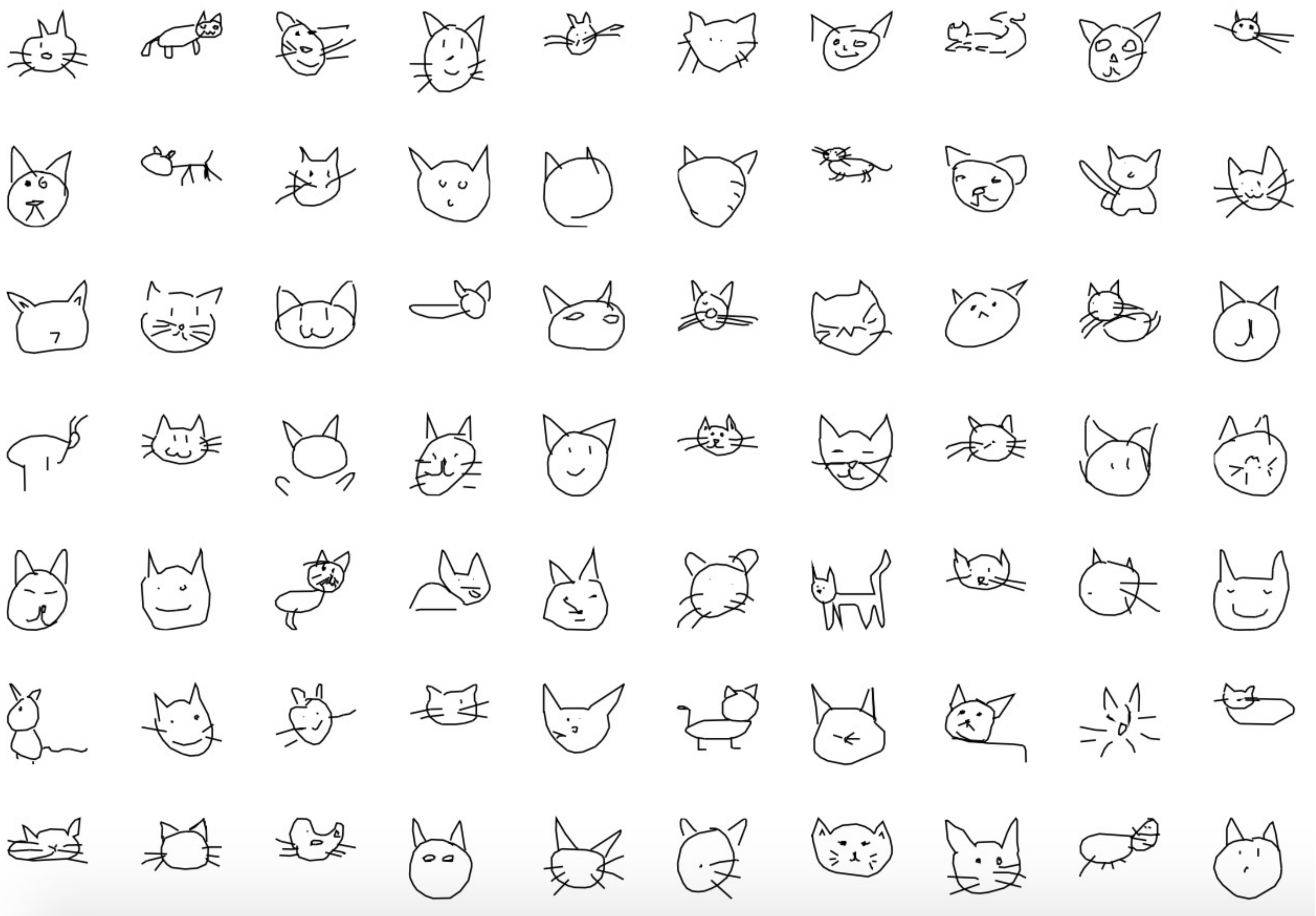
- ANSWER: *[The word we are prompted with is the <u>data label</u>. When we participate, we are creating the <u>underlying visual data</u> for any given label. This is important because everyone has different drawing abilities, no two people will draw exactly alike. Yet, there are similarities across our drawings. When I draw, I create a <u>single datapoint</u>.]*

- QUESTION: What exactly is being predicted here? Does that mean this is a classification or regression algorithm?

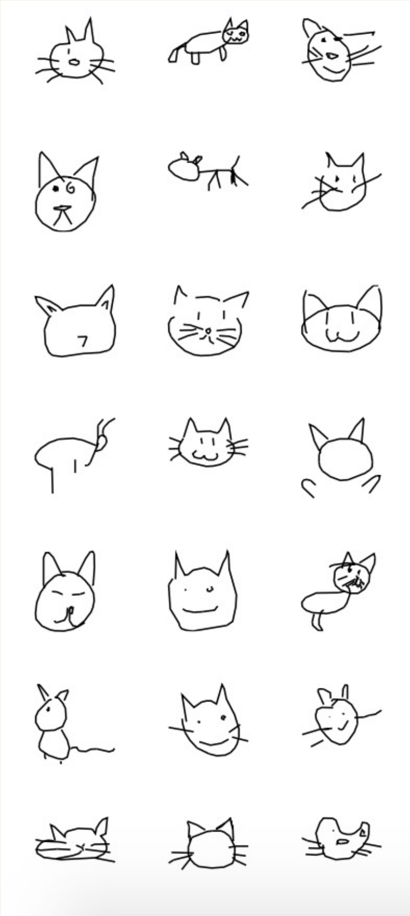
- ANSWER: *[The algorithm is trying to predict what we are drawing. It is <u>classification</u> because we are drawing things, the algorithm is assessing across categories or buckets to get closer and closer to our image. <u>Regression</u> provides numeric predictions and values.]*

Objective 1.1: To consider the associations that can be created across numerous datapoints.

- OPENER: Great, now let’s dive into an example. Let’s take a look at some drawings people created from an internet fav—cats. I’m going to show you just a small subset of a massive dataset. On the site, instead of drawing, you can *actually see* the entire dataset. When we click into the dataset, we see Quick, Draw! has amassed 103,031 drawings of cats through user participation. That is a huge dataset! I don’t know how often they update the dataset, but all of us just participated as well, so we just added to the dataset.



- QUESTION: Imagine for a moment that you do not know what a cat is. All you have are these drawings. What are some features of cats?

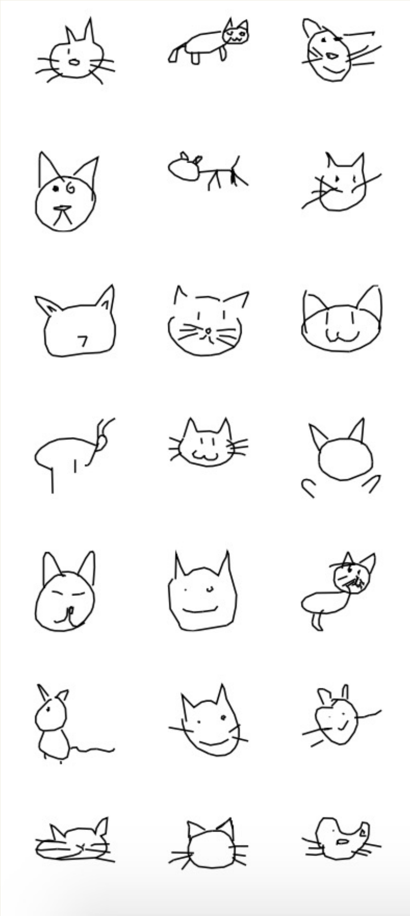


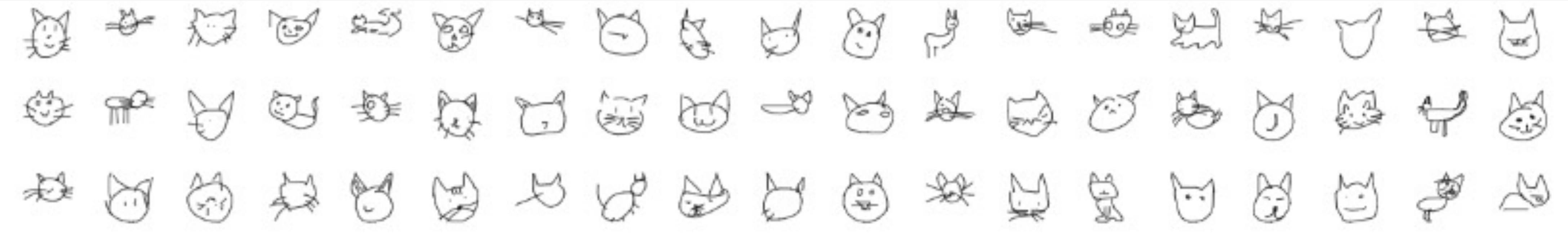
- ANSWER: *[Students typically respond with pointy ears, whiskers, nose, eyes, circle face]*

- DEEPEN: In other words, the algorithm can learn associations between these different features. Anyone remember from the example about an algorithm classifying dogs or wolves. What association did the algorithm make that the researchers weren't expecting?

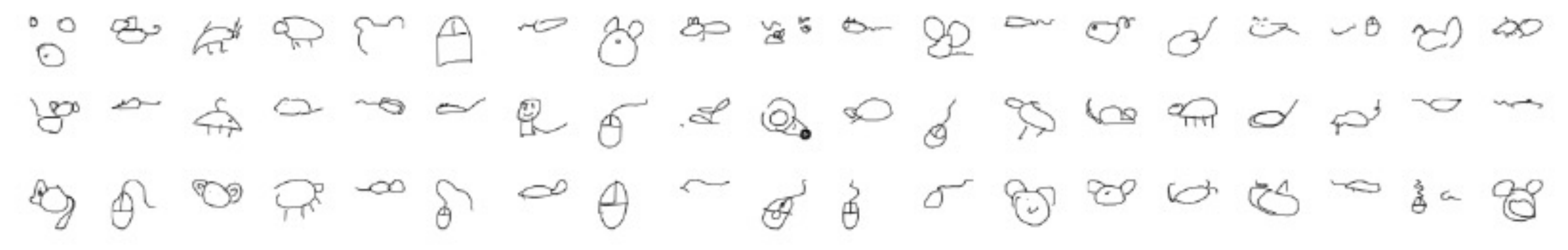
- ANSWER: *[Even though they labelled some photos dogs and some wolves, the researchers didn't realize that all the photos of wolves had snow in the background. So whenever it was presented with a photo of a dog in snow, it said it was a wolf.]*

- AFFIRM: Absolutely! That’s an association the algorithm made. So, the algorithm learned to form as association between wolf and snow. The background was a feature that it was not taught about but it learned.
- QUESTION: Okay, so let’s imagine again we don’t know what cats are. What can we learn from where the ears are placed? Where are the ears in relation to the circle face? What about whiskers? What other animals might come to mind with whiskers?
- ANSWER: *[Students typically respond mouse]*

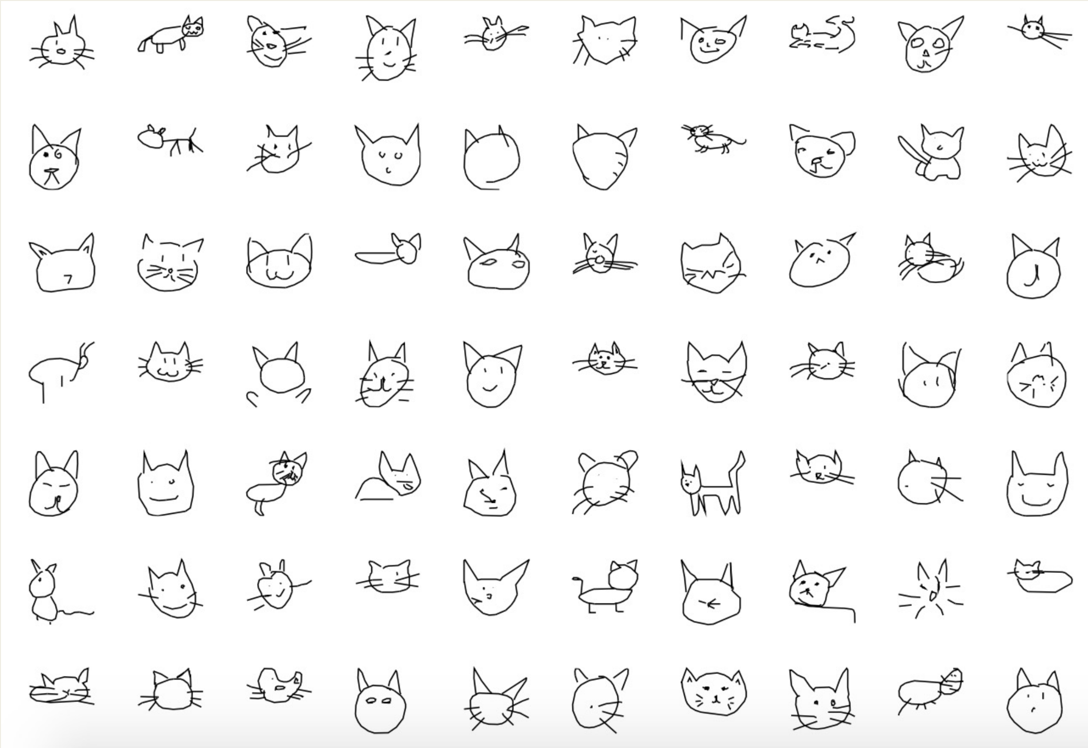


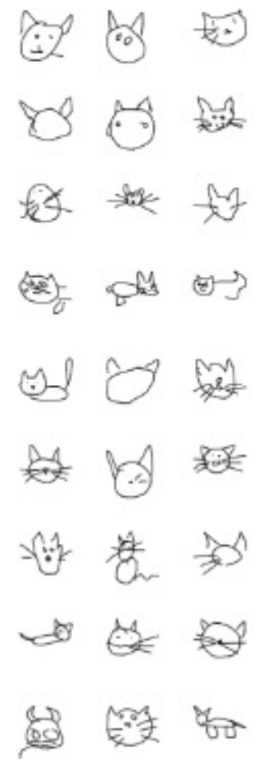


- QUESTION: Sure yes! So how did you know it was a cat and not a mouse? What's different about a mouse face versus a cat face? Or maybe: how would you have drawn a mouse differently?
- ANSWER: *[Students typically laugh and ponder for a bit. They say mouses have pointier noses and smaller ears. Some say they might draw the body of a mouse or draw it with cheese. As a percentage more people draw the body of a mouse than a cat.]*

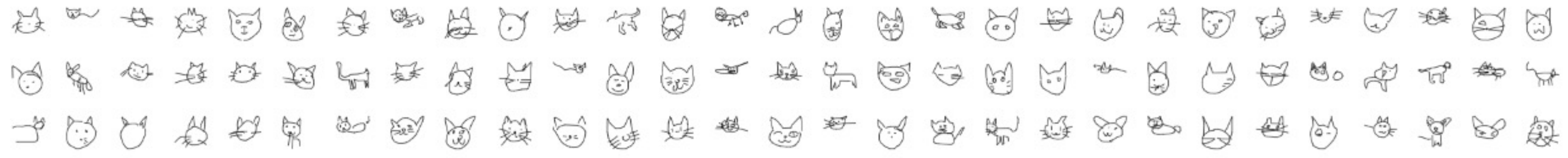


- AFFIRM: Fascinating! Raise your hands if you would have drawn the entire body of the mouse. Let's take a peek at this Quick, Draw! Sample and see how many people took that approach when it comes to cat. Interesting, only 14 people out of 70 people.
- QUESTION: Did anyone get cat? Did it correctly guess yours?
- ANSWER: *[Most students say yes, others laugh at how bad at drawing they are].*





- CLOSER: So, the algorithm gets to learn from across the diversity of 103,031 different people's drawings of and perceptions of cats, some with faces, some with the whole body. The neural net gets to think across them *all*. But consider, when users are prompted with a word, we begin drawing. The algorithm does not yet know what our drawing will be—it is taking snapshots in time of our drawing and returning, fairly instantaneously, likely labels for our drawing. At any moment in time, the algorithm offers a probabilistic guess, or, prediction of the label. It feels like the algorithm is interacting with us. But keep in mind, it's not that the algorithm is watching you and responding, it is just running really really fast. It is taking a snapshot and assessing what the drawing is likely to be.

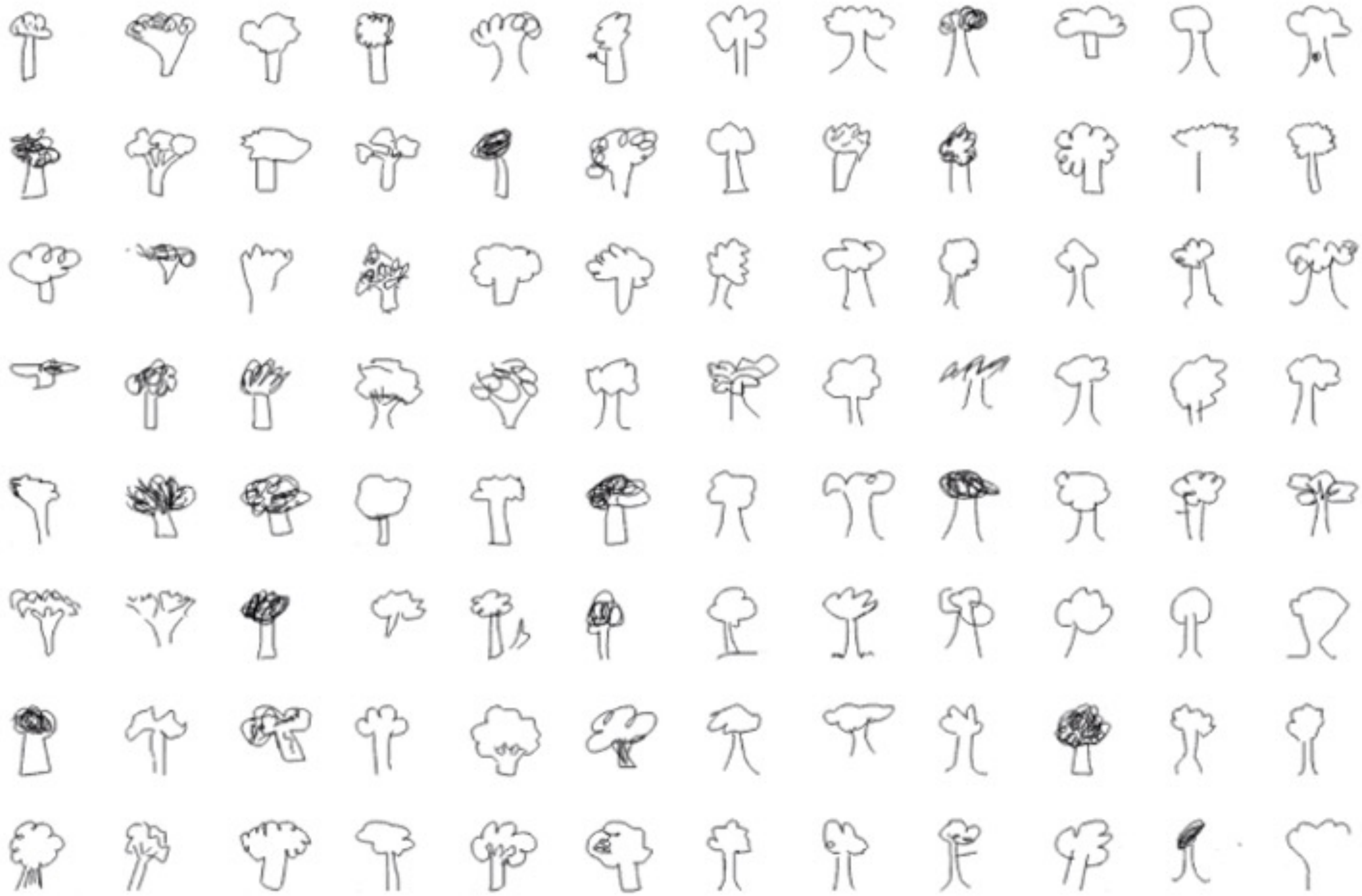


- CLOSER: Because of advancements in processing speed, this algorithm thinks across and considers not only all the cat drawings but is thinking across ALL the drawings data it holds across *all the words* to place our drawing actions into *one* bucket. When it does not correctly guess, it moves to the next most likely data label and says “Oh! Oh! I know! It’s ---"

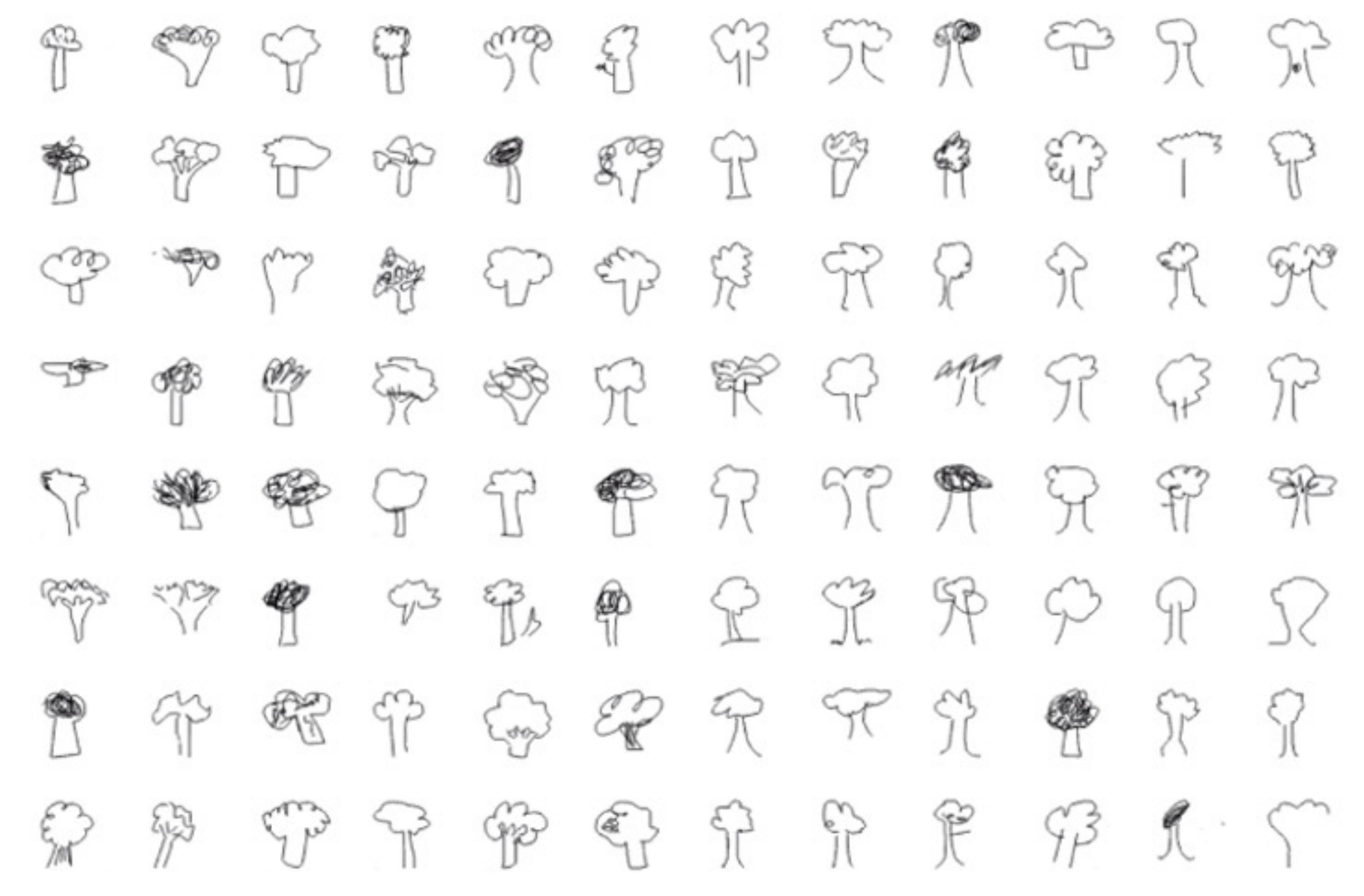
  As it is guessing your drawing, it is looking for associations between your drawing and the millions of other drawings in its existing dataset. And that is how it “guesses” but it’s really a probability of a snapshot. Alright, enough about our beloved internet cats, let look at another example.

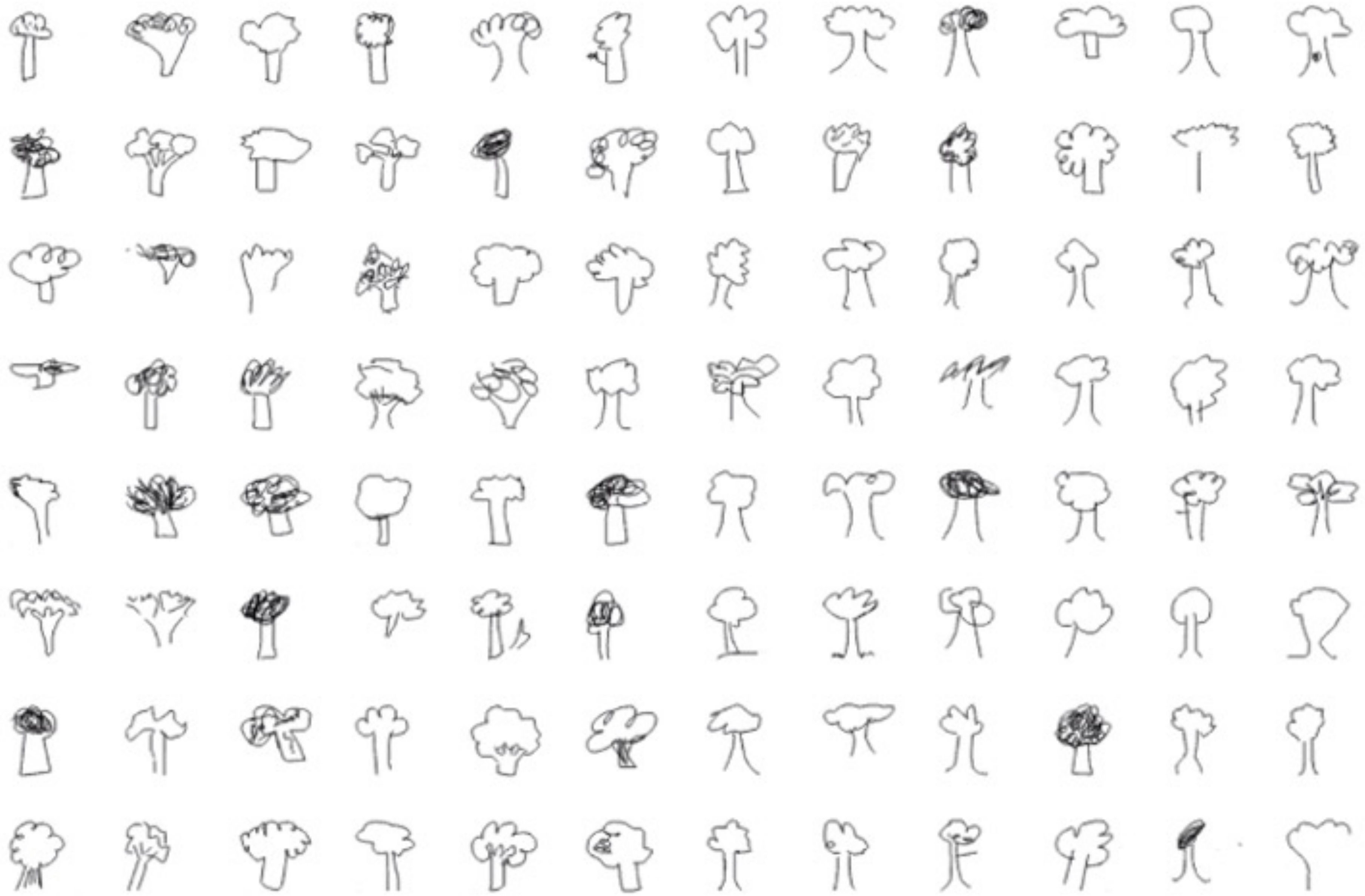
Objective 1.2: To demonstrate that algorithms can predict differences in ways that are not readily apparent to humans.

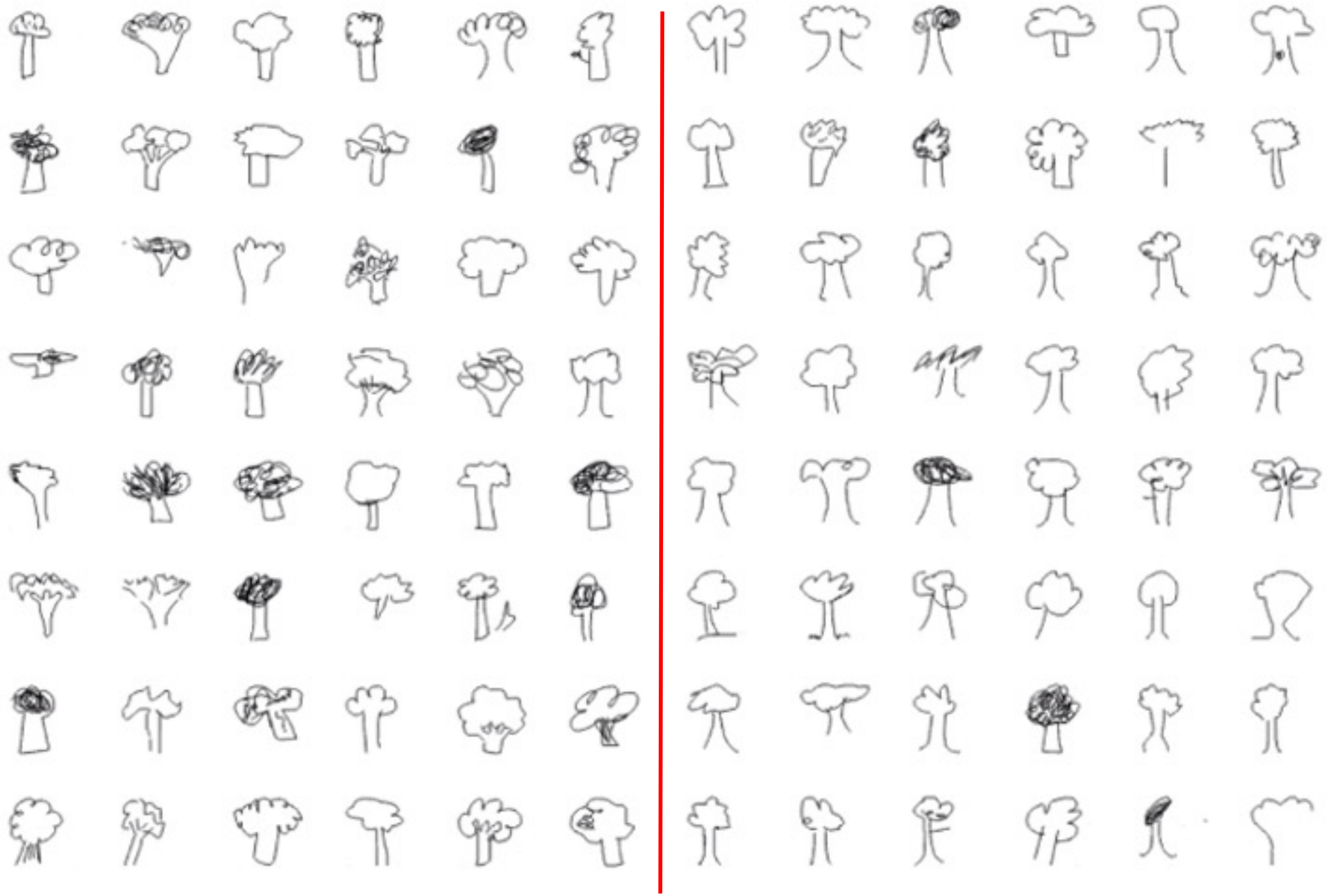
- OPENER: Okay, now I want to move onto another dataset for discussion. Here it is. What do you think it is?



- ANSWER: *[Students usually guess tree]*
- TEASER: Good guesses. What if I told you this was actually two different datasets representing two different labels? What do you think now?







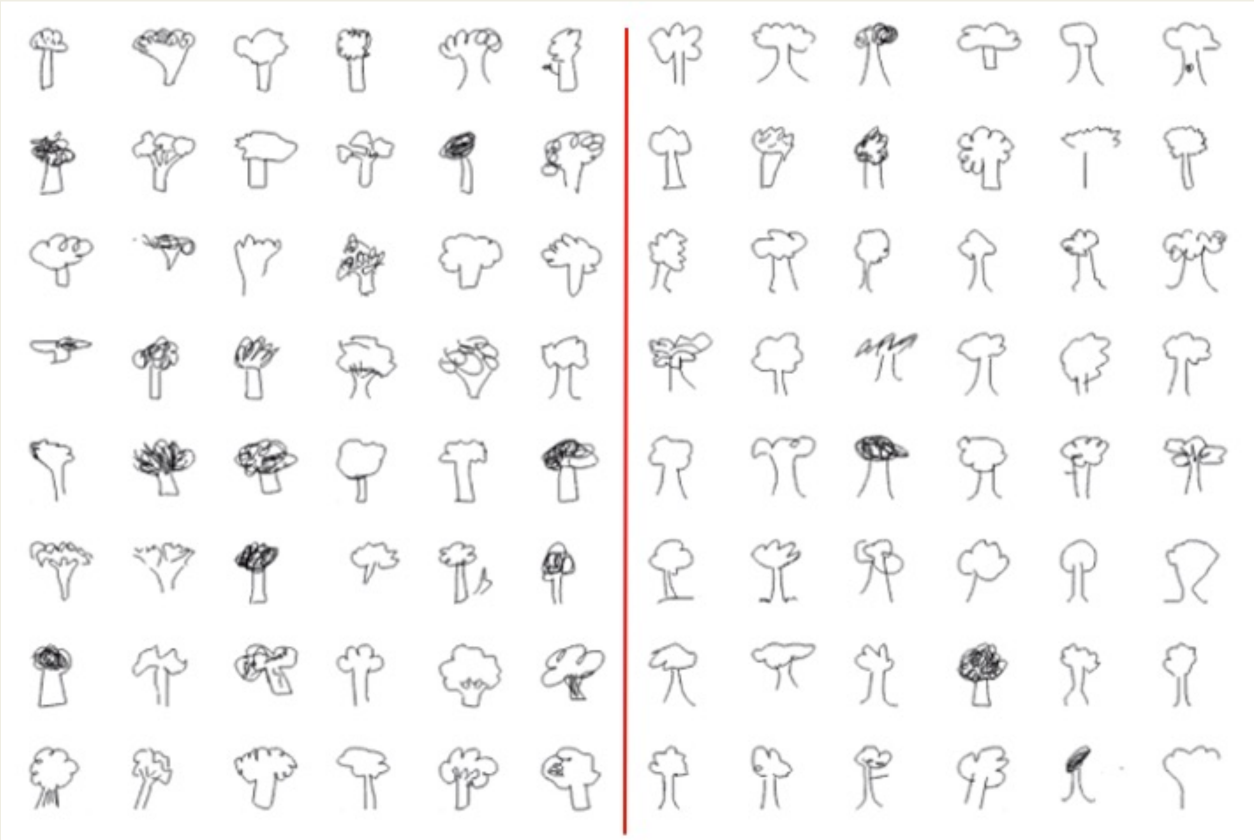
- QUESTION: What do you think the data label of the images on the left is? The right?

- ANSWER: [*Note: broccoli is on the left and tree is on the right. Students typically laugh and ponder. Someone often guesses tree. Ask which side is "tree" and why the other side is not. Let them debate and discuss why. If students make declarative statements ask them how or why they believe that to be true.]*

- DEEPEN: With only these images and nothing else, the algorithm is making a call irrespective of scale. Maybe it is seeing something we cannot see. What are some key differences?

- ANSWER: *[Students are likely to start pointing out similarities across drawings or to particular drawings. E.g. the top right datapoint is a tree. The drawer even gave it a hole in the trunk.]*

- CLOSER: Nice work identifying some of those smaller details that help clue us to other features we associate with trees or broccoli. We learn this from our life experiences, from repetition, from children’s books, photos, movies, and more. As we discussed last week, machine learning algorithms learn associations from datasets.



Objective 1.3: To apply the importance of the dataset to how an algorithm defines "success"

- PROMPT: Okay so it’s learning associations between all these datapoints. How might we break Quick, Draw!’s ability to guess the right item?
- ANSWER: *[Students usually pause and need to think here. After a minute, someone usually offers an iterative of “feed it junk”]*
- DEEPEN: Exactly! But what does it mean to have *one* person purposefully draw images incorrectly?
- ANSWER: *[Students respond saying that the dataset is too big. You cannot sway it. Typically, a student in computer or engineering field will offer: You could write a script to constantly feed it bogus data. You would need to overtake or overthrow the existing data with junk data to have it learn new associations.]*
- AFFIRM: Good thinking! So, if we wanted to “break” an algorithm’s definition of “success, we would need to feed it so many data points. One is not enough data, we have to think *at scale*.

Objective 2.1: To demonstrate that consensus is not always present in datasets.

- OPENER: Okay so now we know how to try to break what an algorithm determines is “success.” Let’s take another example. I’d like to start by asking a student to volunteer to share their screen and draw a phone.

- ACTION: *[Student volunteer will draw a phone—they may draw a mobile phone, a rotary phone, a cordless phone, etc. The goal here is to get students understanding that when people are prompted with “phone” not everyone draws the same thing.]*

- PROMPT: Great! Do others agree that this is a drawing of a phone? Now I want you to think about how you would draw a phone. What comes to mind? What are different types of phones?

- ANSWER: *[Students typically describe phones, some struggle with naming them so be ready to help out. Smartphones, feature phones, flip phones, cordless phones with base, corded phones on tables, corded wall phones, rotary phones, pay phones, phones with shoulder holders, etc. When students offer a type of phone they would draw, prompt them to describe how and where they are familiar with that generation of phone.]*

- DEEPEN: Phones are a technology that have existed since 1876 with Alexander Graham Bell. But they certainly look and operate in different ways today! The literal apparatus has changed so much over time. Ever seen those videos of millennials trying to use a rotary phone?

  [Pause for discussion]

  How many of you know how to use a rotary phone?

  [Take a quick informal poll.]

  Great, now let's take a look at the "phone" dataset in Quick, Draw!

- PROMPT: What do you notice about this dataset? Does it look like cat or tree datasets?

- ANSWER: *[Students comment there are lots of different types of phones in here.]*

- REINFORCE: So, let’s think about this from machine learning perspective. Is there a consensus around the data label “phone”?

- ANSWER: *[Students respond, no! and typically name off some of the categories of phones.]*

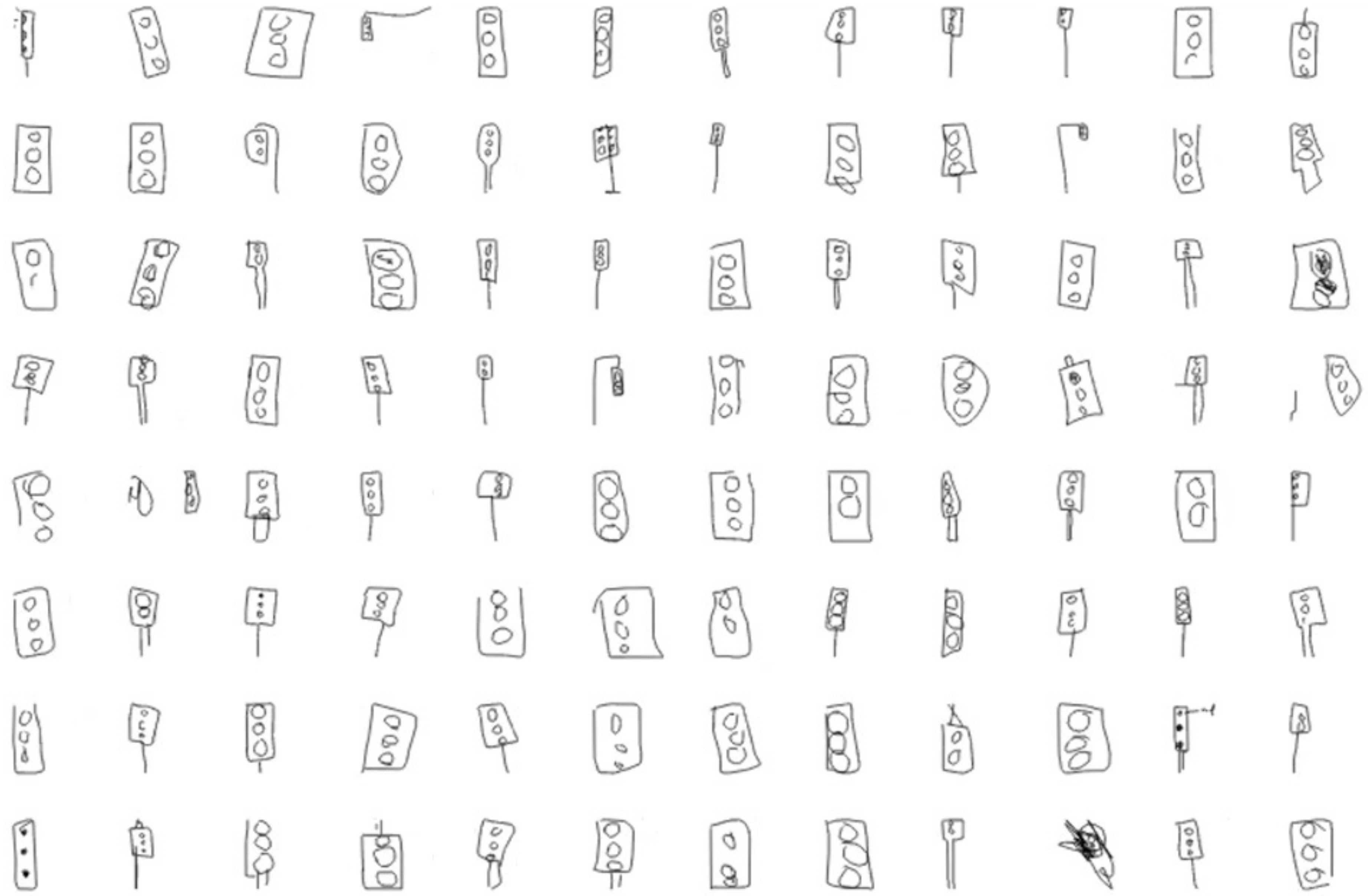
- DEEPEN: Consider that nearly ALL cat drawings had an association between pointy ears and whiskers and a round face. What do you see here? What are some associations that you might draw about what a phone apparatus looks like?

- ANSWER: *[Students typically note that there are no associations across the entire dataset, acknowledging that there is no real consensus about a phone. They naturally start to divide the images into buckets that represent different generations of phones. that sort into a higher order phone.]*

- REINFORCE: Great! So, what you are describing here is a different level of classification. While trees were drawn fairly similarly, phone drawings do not have strong consensus across the entire dataset, but we can sort them into smaller buckets or classifications that aggregate up to a prediction that a certain drawing is “phone.” The algorithm can understand these different buckets or generations of phones. So, although there may not be consensus at the largest level, there are lower levels of classifications that would create more consensus internal to that particular bucket e.g. smartphones are rectangular, corded phones have a spiral coming out of one end, rotary phones rest on a base, etc.

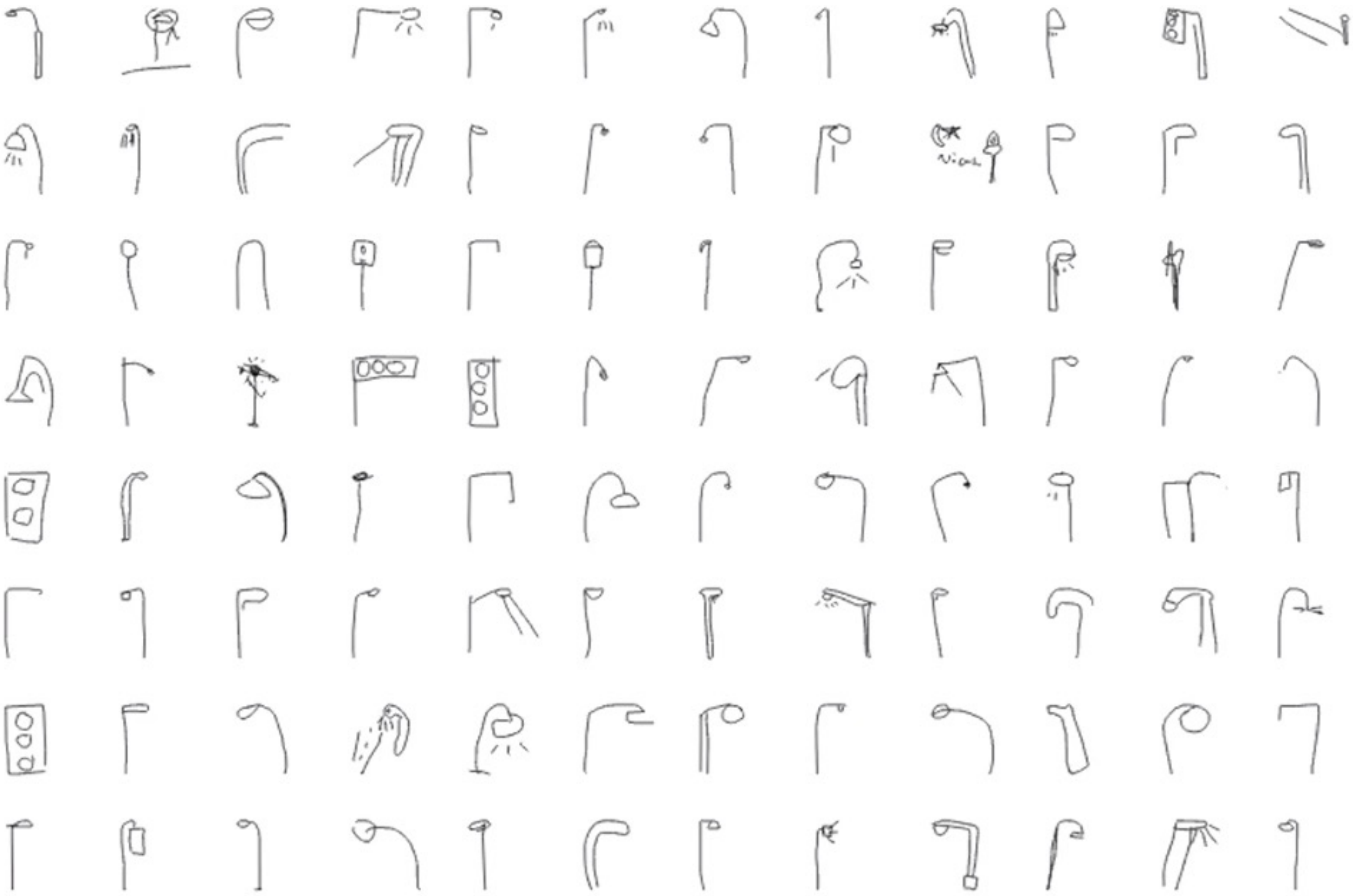
Objective 2.2: To demonstrate there is not consensus in data labeling.

- PIVOT: Now I’d like you to look across numerous data points and provide what you think the data label is. Now we’re going to explore consensus building not in the data, but in the label. What is this? What do you think the data label is for this dataset?



- ANSWER: *[Students often respond stop light, some may say traffic light. Capture and highlight the different labels]*

- PROMPT: Actually, the algorithm labels this a “traffic light.” That’s interesting. What if you don’t know the exact word? Ok so if that was a traffic light, what do you see here?

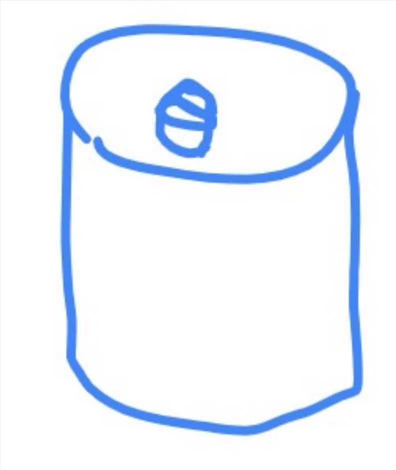


- QUESTION: Quick, Draw! has 117,000 data points of "street light." Is there consensus here about what a street light is?

- ANSWER: *[No, some people drew traffic lights but there seems to be a consistent idea of a light on a pole, often bent. The consensus is better than the phone]*

- PROMPT: Can we think of any examples where there is wide variety across the population for a particular image? Any ideas?

- DEEPEN: Alright, I'm going to open up a poll in just a moment. I want you to comment on your own label for the image I'm about to draw. [Share screen and draw an image of a pop/soda/coke/etc.]. Okay, here is the drawing. Now enter in your own data label for these images into the chat as fast as possible.



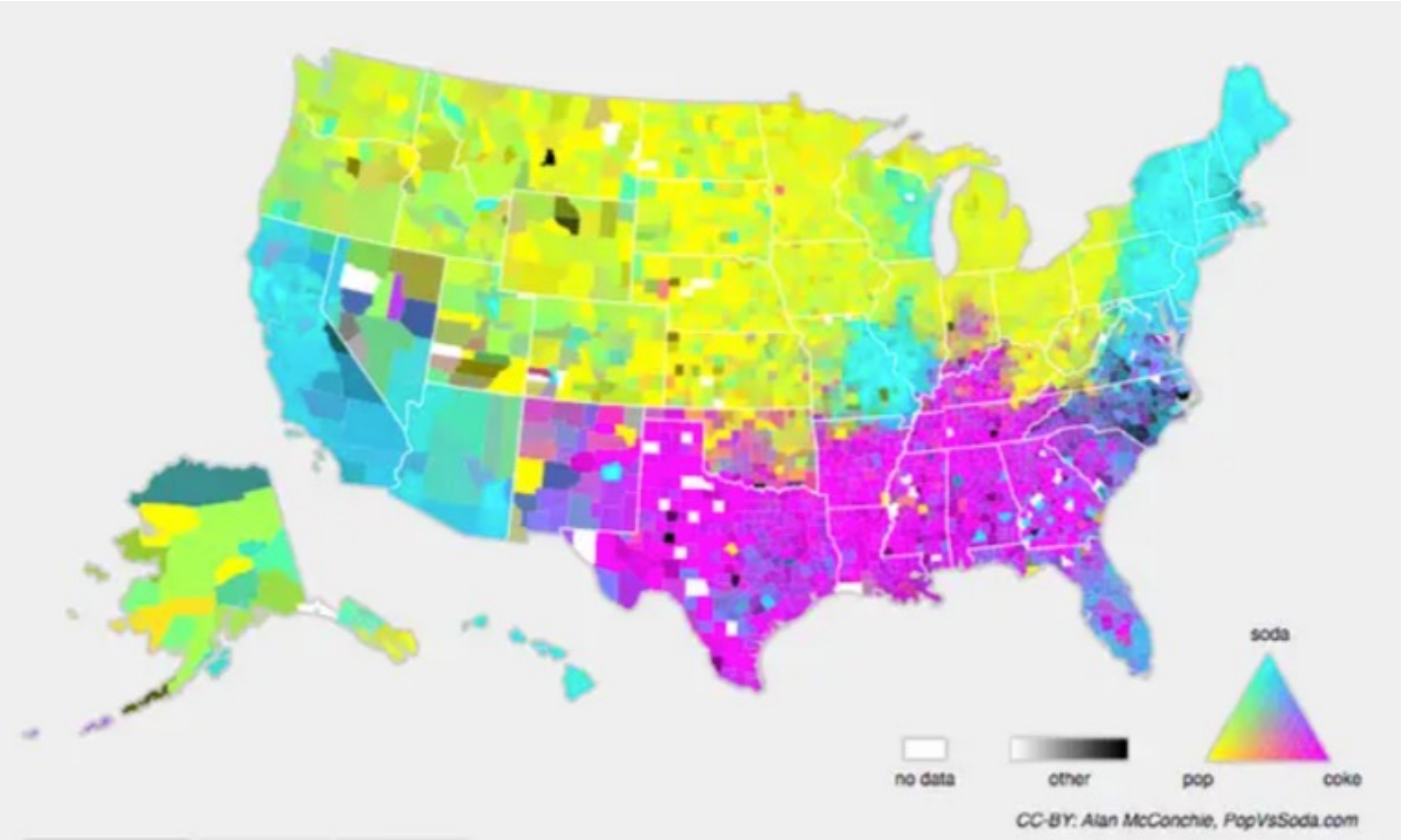
- DISCUSS: Great, let’s dive into your answers.

  [Share the poll results so students can see how their answer compares to others and move through the different answers.]

  Who said “[coke]”, anyone? Will you share where you’re from?

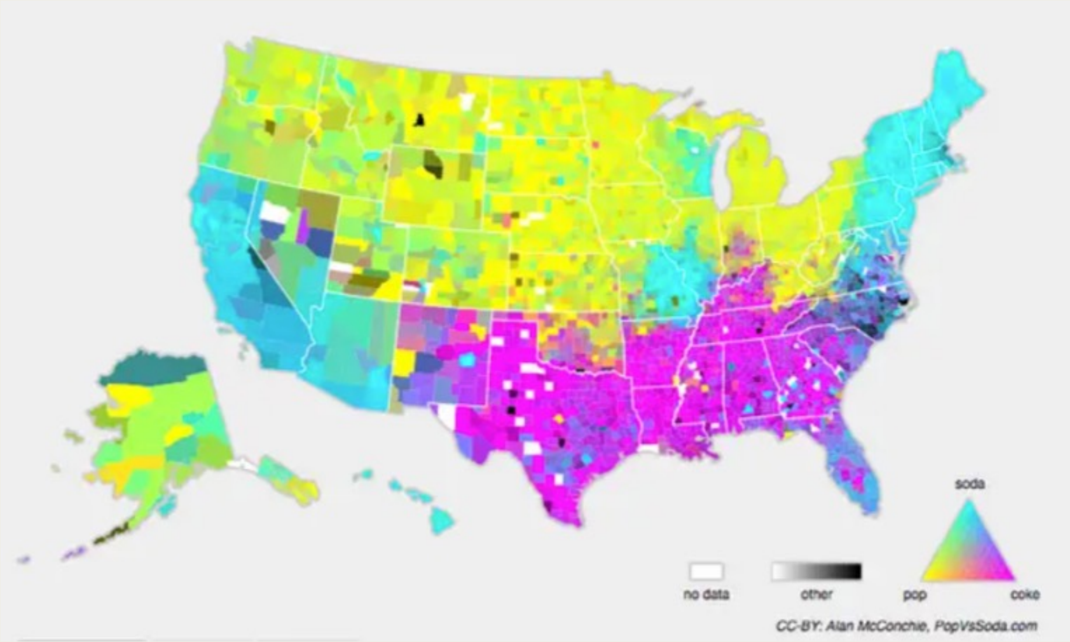
  Note: depending on your university and its student population, you may have more or less diversity in answers.]

  What can we observe about differences in our answers? These linguist differences have been researched by others and there are fairly surprising consistencies by region. Take a look at this map:


soda
no data
other
pop
coke
CC-BY: Alan McConchie, PopVsSoda.com

- QUESTION: What do you see here? Can you read the key? It's in the lower right. So blue is soda, yellow is pop, and pink is coke. What associations can you make between region and the likely data label?

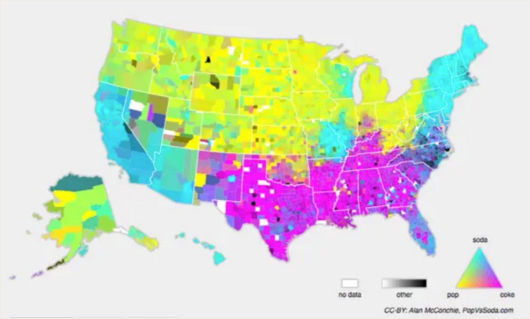
- ANSWER: *[Students typically respond by going over the different regions, naming the key ones.]*



- DEEPEN: And where are areas where there is the least consensus? Consensus matters because predictions are *probabilities*. They are not 100% accurate. The algorithm predicts an outcome based on datasets, but the prediction is a guess, a *likely* outcome, *probably* true. Where might we not be able to accurately predict how a person would label an image?

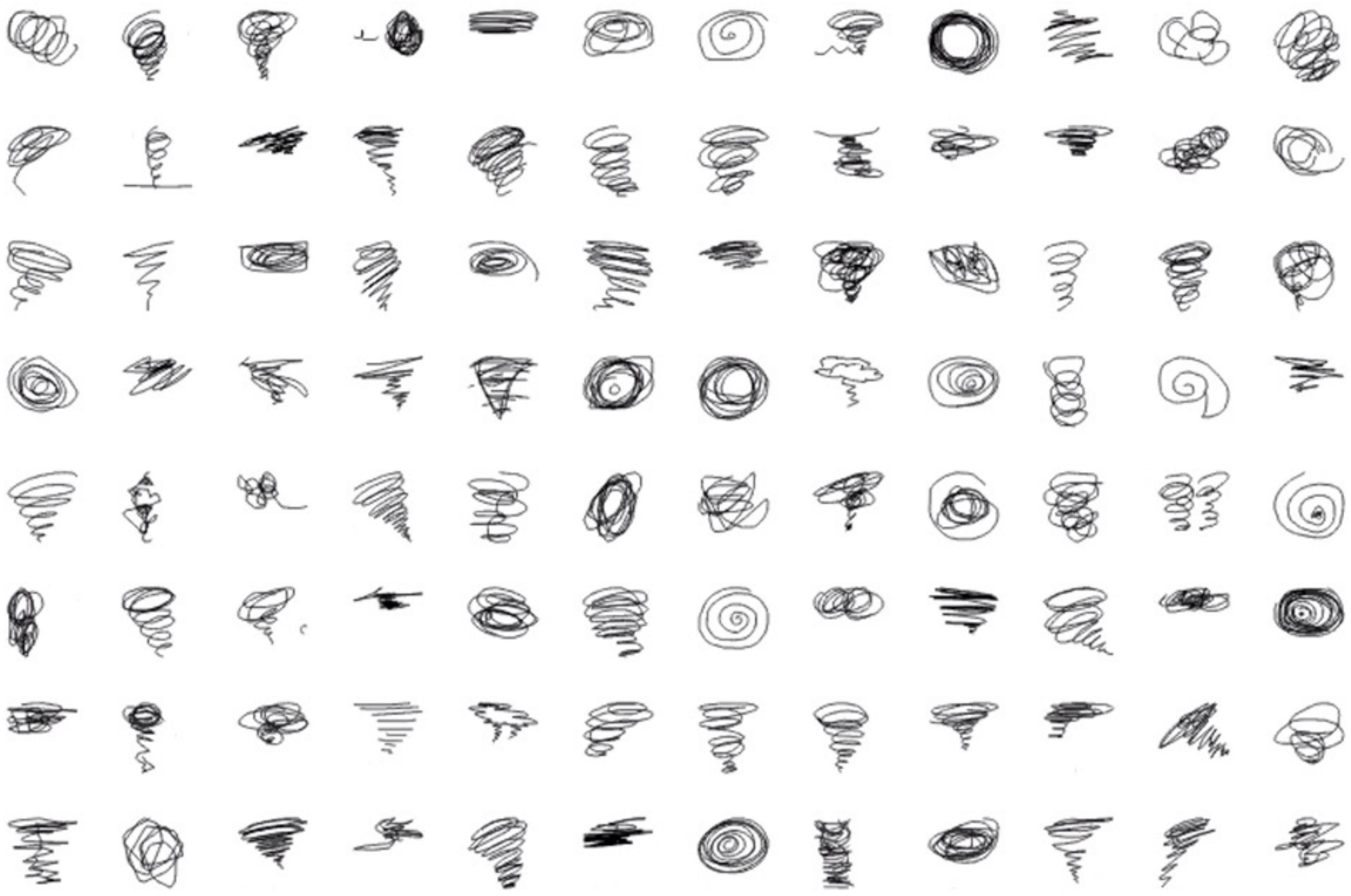
- ANSWER: *[Students should highlight green areas (Pacific NW into the West; Wisconsin; etc.); black areas (key notes these are "other") and something odd appears to be happening in Nevada, New Mexico. The blue blob in the middle of the country, in Missouri.]*

- APPLY: So, there can be many variations that will change how we, as individuals, label something. One person saying soda is not wrong, another person saying coke is not wrong, there are just differences in our perception and ways of describing the world around us. In Louisiana, an algorithm is likely to be highly accurate in its prediction that someone will label a can “coke,” whereas in Florida, the accuracy is likely to decrease because we’re not sure if someone will say “coke” or “soda.” However, see how there is no yellow (or orange or green) in Florida? Although we cannot accurately predict what someone will say, we would most likely be accurate if we said we know they will not say “pop.” We have more statistical clarity that around what they will not say, than what they will say. So, in summary, if we were employed to label images for machine learning datasets, we might label the same image differently. This has to do with our experiences, socialization, and local cultures and languages of which we are a part.



Objective 2.3: To demonstrate that not everyone is equally seen by an algorithm.

- OPENER: Alright now I want to switch gears into a thought experiment. If you correctly guess the right label for this dataset, you win $100. Cool? If you guess it right, you get the money, if you don’t guess it right, you don’t get the money. Okay, what do you think this is?

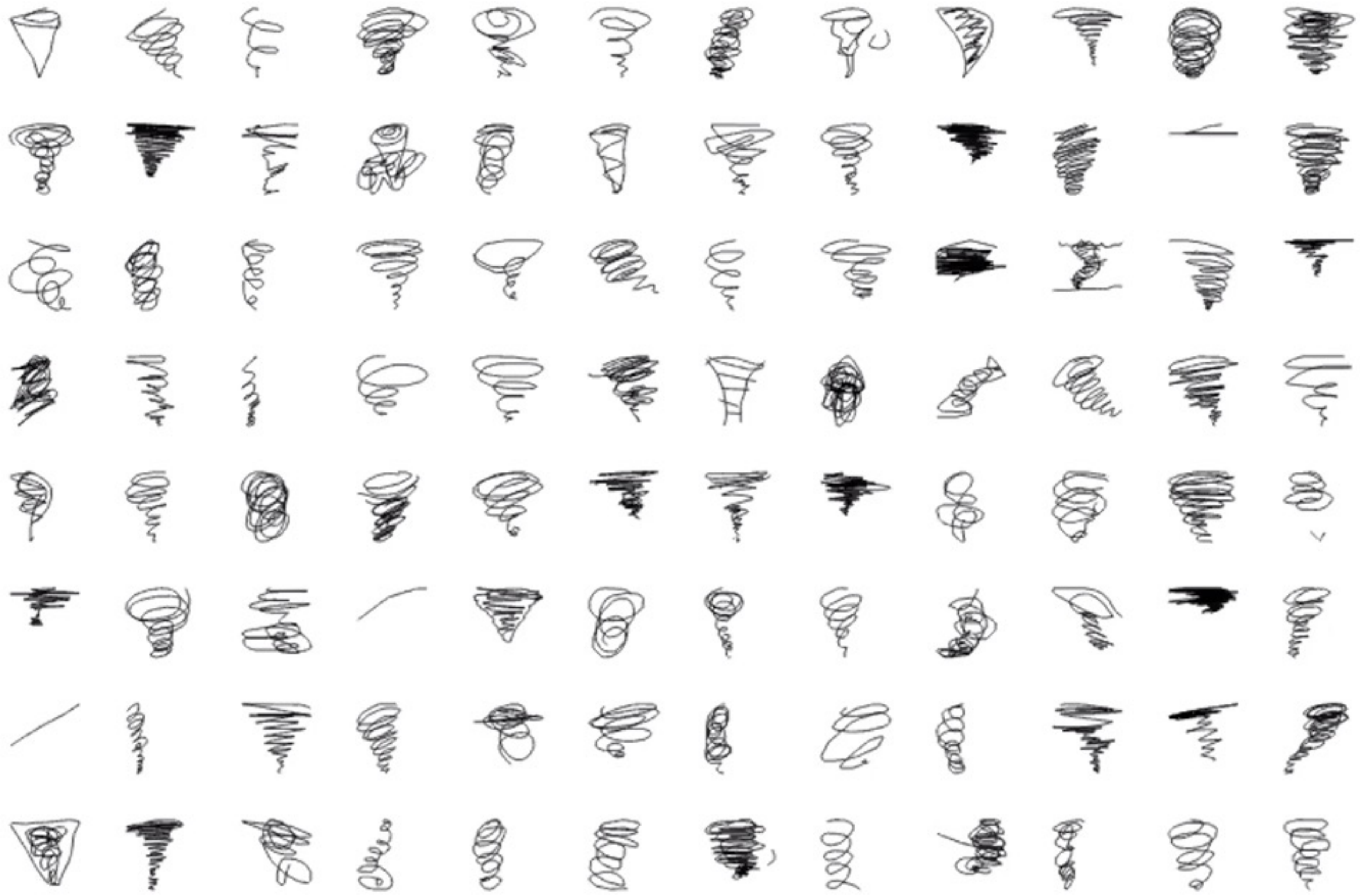


- ANSWER: *[Students answer coil, spring, spiral, tornado, hurricane]*

- QUESTION: Yes! It is hurricane. Only a few of you get the money! Here, we see a small subset of 80 datapoints. For those of you who won the money, from which data points are you *specifically* getting the idea of a hurricane from?

- ANSWER: *[Students answer the ones that are circles or spirals].*

- TEASE: Okay so for comparison's sake, let's look at the dataset for tornado. Is this instantly recognizable as a tornado? Why or why not? What about consensus in this dataset? Has anyone ever seen a tornado? Does it look like this from the ground/from our vantage point?



- ANSWER: *[Some students might say they thought it was a coil or spring. They respond favorably to this being a common view of a tornado]*

- DEEPEN: Good! Absolutely, there is more consistency in this dataset. I saw a tornado once, it was terrifying. I saw it come down from the sky in this funnel shape. So, I could instantly recognize it because it looked the same as these images. But I've never seen a hurricane, has anyone ever seen a hurricane firsthand? What does it look like from the ground?

- ANSWER: *[Students may say yes. Ask for more details and to describe what they see]*

- PROVOKE: Okay, but what about those of you who have never seen a hurricane firsthand, where did you get the idea about a hurricane? Where is that imagery coming from?

- ANSWER: *[Students answer the news or social media]*

- DEEPEN: Right! But where is the imagery coming from? What technology enables this view of a hurricane?

- ANSWER: *[Students say “from airplanes” or satellites.]*

- RECALL: Right! And we get access to satellites imagery because we watch the news. What if you don't have electricity? What if you don't have a tv? What if you get your news from the radio? Recall the digital divides that we spoke about earlier in class and that nearly 60% of the global population does not have access to the internet. Quick recall, what are the 3 levels of the digital divide?

- ANSWER: *[Students should be able to recap from previous readings: (1) internet access (2) ability to engage with online world (3) ability to turn online world into outcomes in the analog world.]*

- DEEPEN: Nice job recalling that, keep those digital divides in mind. Let’s return to the global population. So, if you don’t have a tv and get to watch the news to get weather, it’s unlikely that you know what satellite images of hurricanes look like. So just like the coke/pop/soda example, our different life experiences and our resources impact how we label images. So, if everyone around the world got the dataset and needed to properly label it “hurricane” in order to win $100, the ability to win at that game is not the same for everyone.

- APPLY: Okay so let’s run another thought experiment. Now we don’t have to guess the right data label, we need to get the Quick, Draw! algorithm to guess the right label in order for us to get, let’s raise the stakes, $500! I sure could use $500 bucks, what about you all? Now it’s not just us and datapoints, it is our ability to draw an image that an algorithm can quickly guess. We are now interacting with the algorithm and that interaction determines our ability to win $500. We win the $500 if we can get the algorithm to properly predict we are drawing a “hurricane.” What are all the factors that go into our ability to get the algorithm to guess hurricane?
- ANSWER: *[Students start listing: my ability to draw, how my drawing is similar to other drawings in the dataset, access to internet, ability to read the instructions, my knowledge of what a hurricane looks like, having seen a hurricane, having seen satellite footage of hurricanes].*

- QUESTION: Really great points. So, do we see how many different factors could go into our ability to get resources, in this case $500, in interacting with an algorithm? Some of us would win and some of us wouldn't. But our ability to win the money is not equal going in, for reasons to do with our abilities, our experiences, our access to resources before playing the game. Another way to think about this is to consider how easy or hard it is for you to be "seen" by an algorithm.

- EXTEND: So, let's return to an example at the beginning of today. I want you to imagine that you get $1,000 if you can get the algorithm to guess "phone" in 10 seconds. If only elderly people create the dataset for "phone" who is NOT going to get $1,000? Why?

- ANSWER: *[The dataset will have more older rotary or corded phones and likely not include smart phone apparatuses. Young people who draw a rectangle will not be visible to the algorithm, they will not be legible.]*
- QUESTION: Okay, now imagine that only people who are 20 and younger drew the dataset for "phone." Who is not going to get $1,000? Why?
- ANSWER: *[The dataset will likely have more rectangular shaped smart phones, lacking the older receivers with two clear knobs on either end, corded phones, rotary phones.]*
- EXTEND: Now, why is the Quick, Draw! algorithm able to identify phone across these differences?
- ANSWER: *[Because the dataset includes them all and is large, covering the variability of phone apparatuses across technological developments].*

- CLOSER: Right! So now we start to see why representative datasets are so important! Be sure to consider these ideas and concepts as we move into our material for next week—algorithmic bias. Next week we are going to look at cutting edge research by Facebook and Microsoft researchers as well as academic research.

End of Activity

The following week students read:

Excerpts from:

- Algorithms of Oppression by Safiya Umoja Nobles
- Automating Inequality by Virginia Eubanks
- Weapons of Math Destruction by Cathy O'Neill
- "Does Object Recognition Work for Everyone?" by Facebook researchers

And we begin to discuss the role of AI in producing differential outcomes along race, class, and gender dimensions with attention to representation in datasets.

Does this interest you?

- Let's make the 21st a better world together! But first that means we need to geek out on all things AI, ethics, equity, discrimination, intersectionality, and more!
- I'm always happy to chat, take feedback, and use this as a springboard for discussion!
- Please reach out! emilyspringer@gmail.com