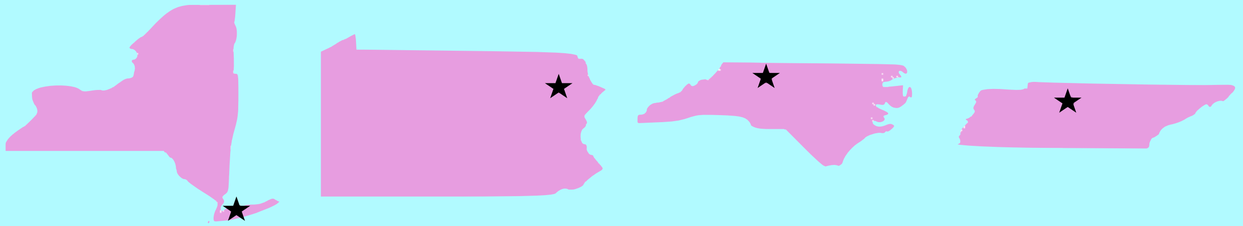
Statistics in Your Own Backyard

Day of Discovery

Ashley Mullan, M.S.

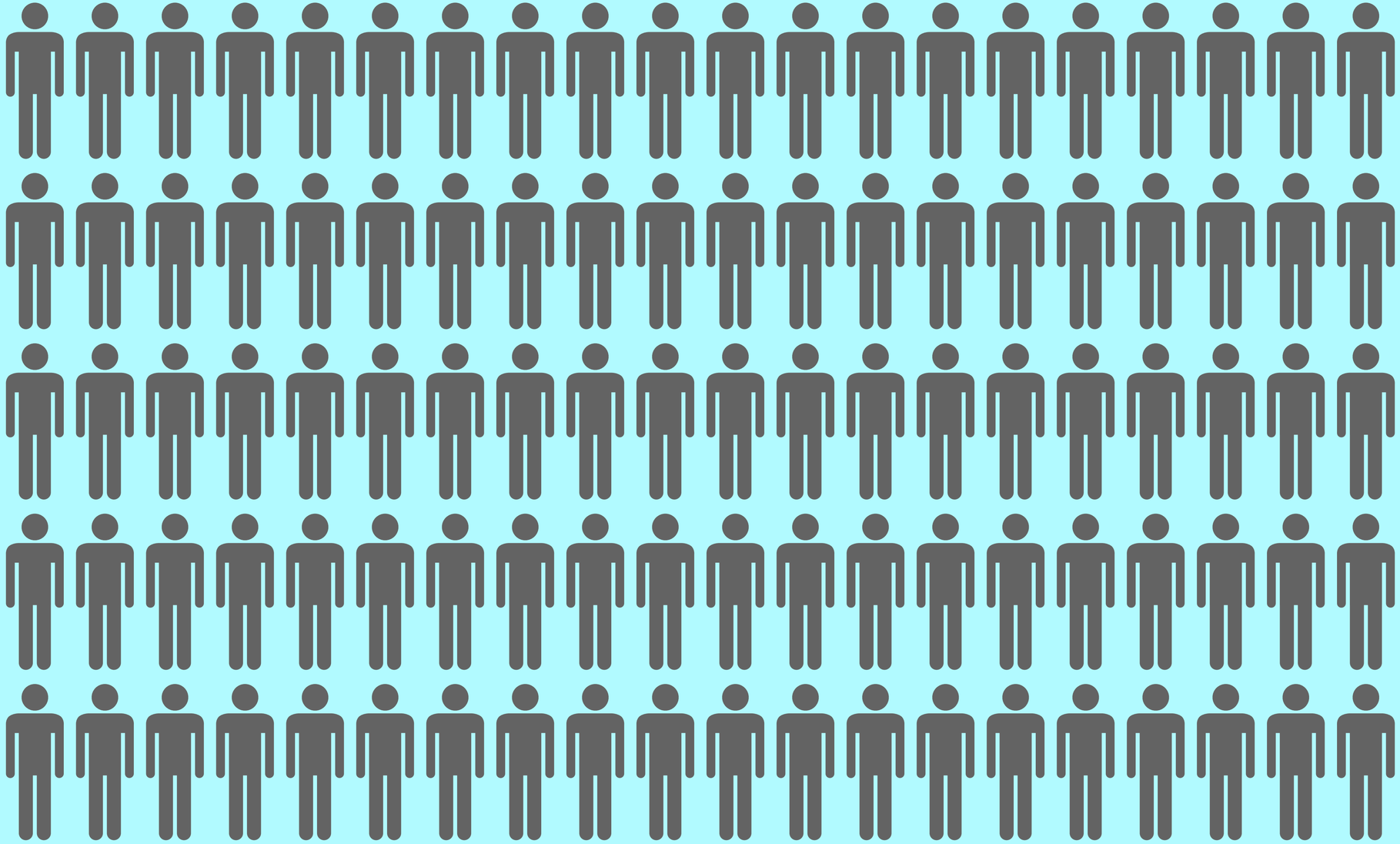
But first, who am I?

👋 I’m Ashley! I’m a PhD student in the Department of Biostatistics at Vanderbilt University. I’ve earned degrees in math, statistics, and philosophy from the University of Scranton and Wake Forest University, and now I work on solving public health problems with math, computers, and data!



What I Do

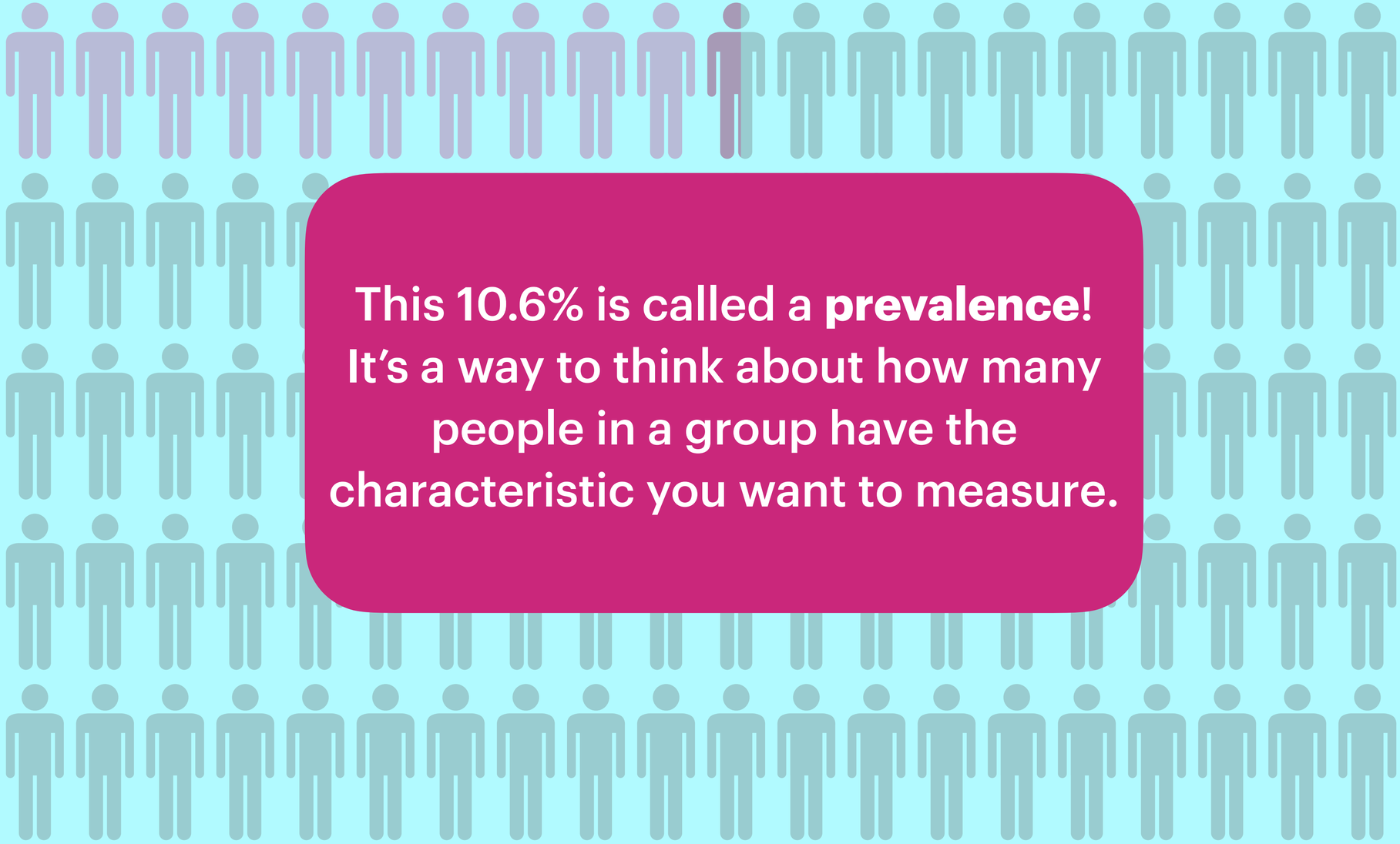
Take 100 adults from North Carolina.



Take 100 adu a.

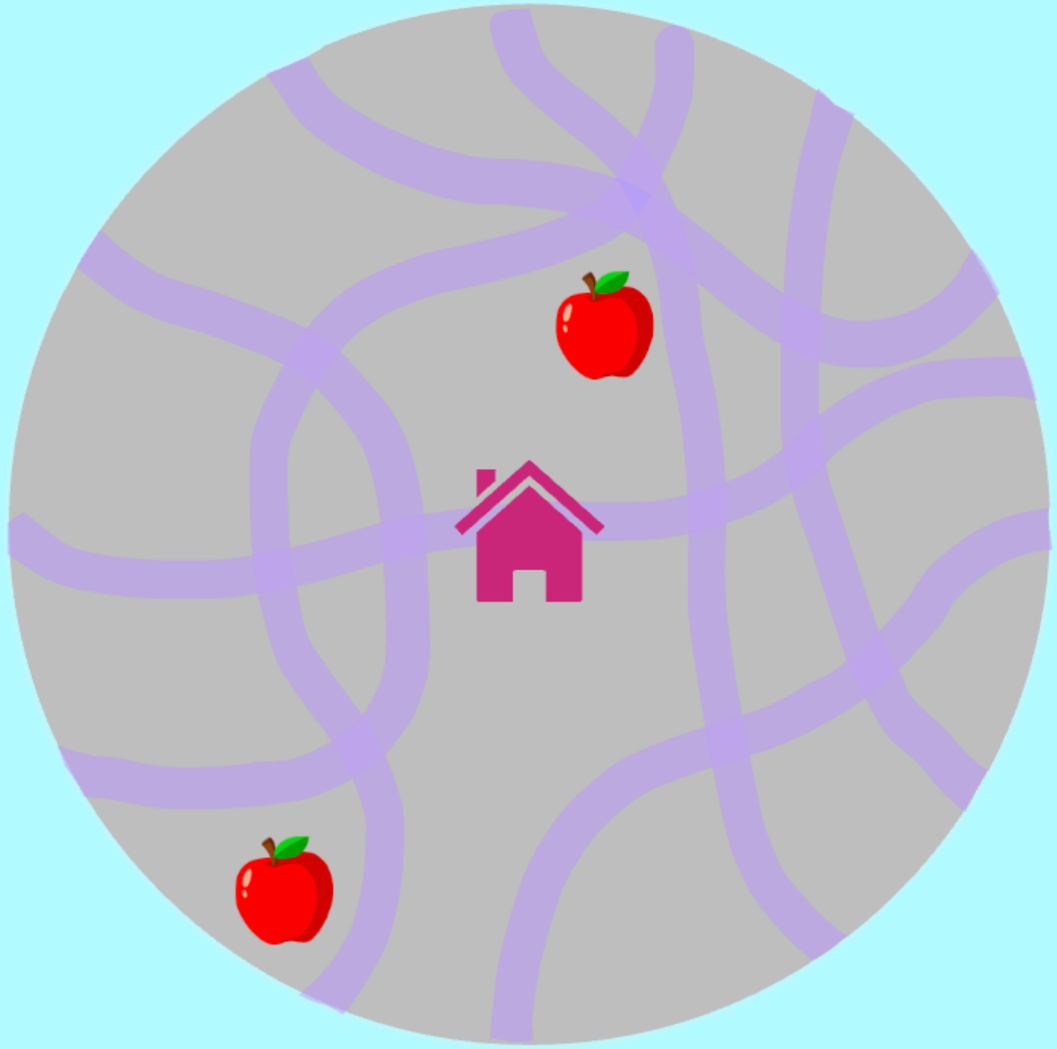
On average, 10.6% of them have diabetes.

On average, 10.6% of them have diabetes.

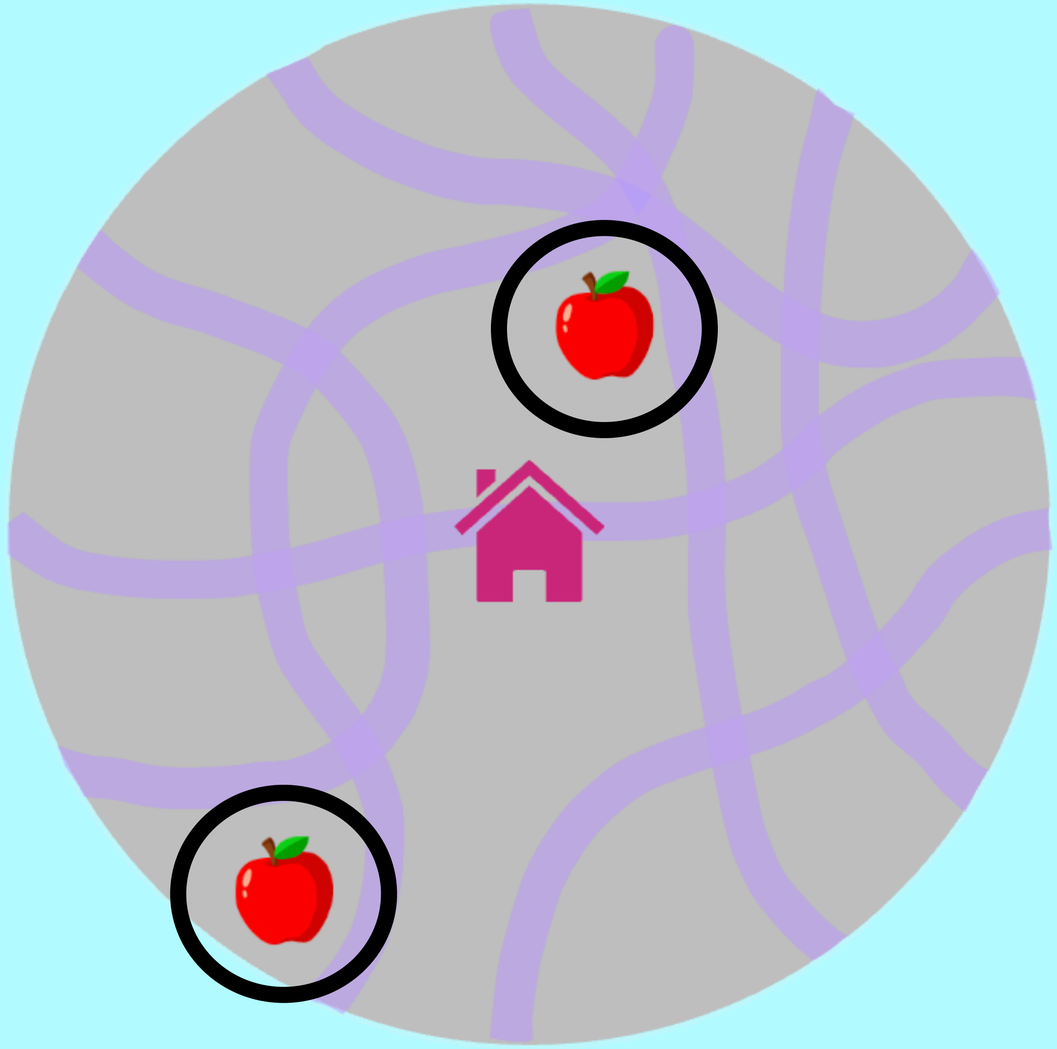


This 10.6% is called a **prevalence**! It's a way to think about how many people in a group have the characteristic you want to measure.

Let’s focus on access to healthy food!

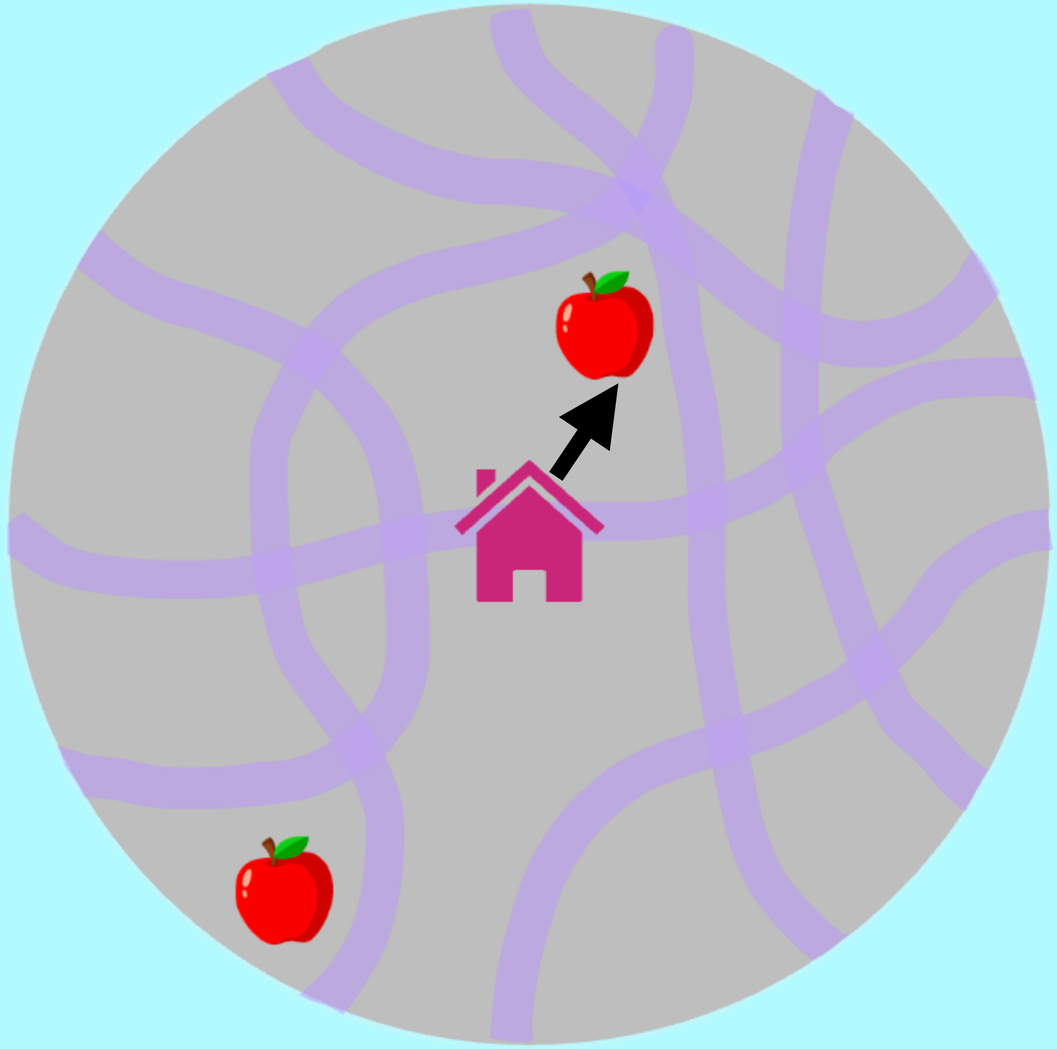


Let’s focus on access to healthy food!



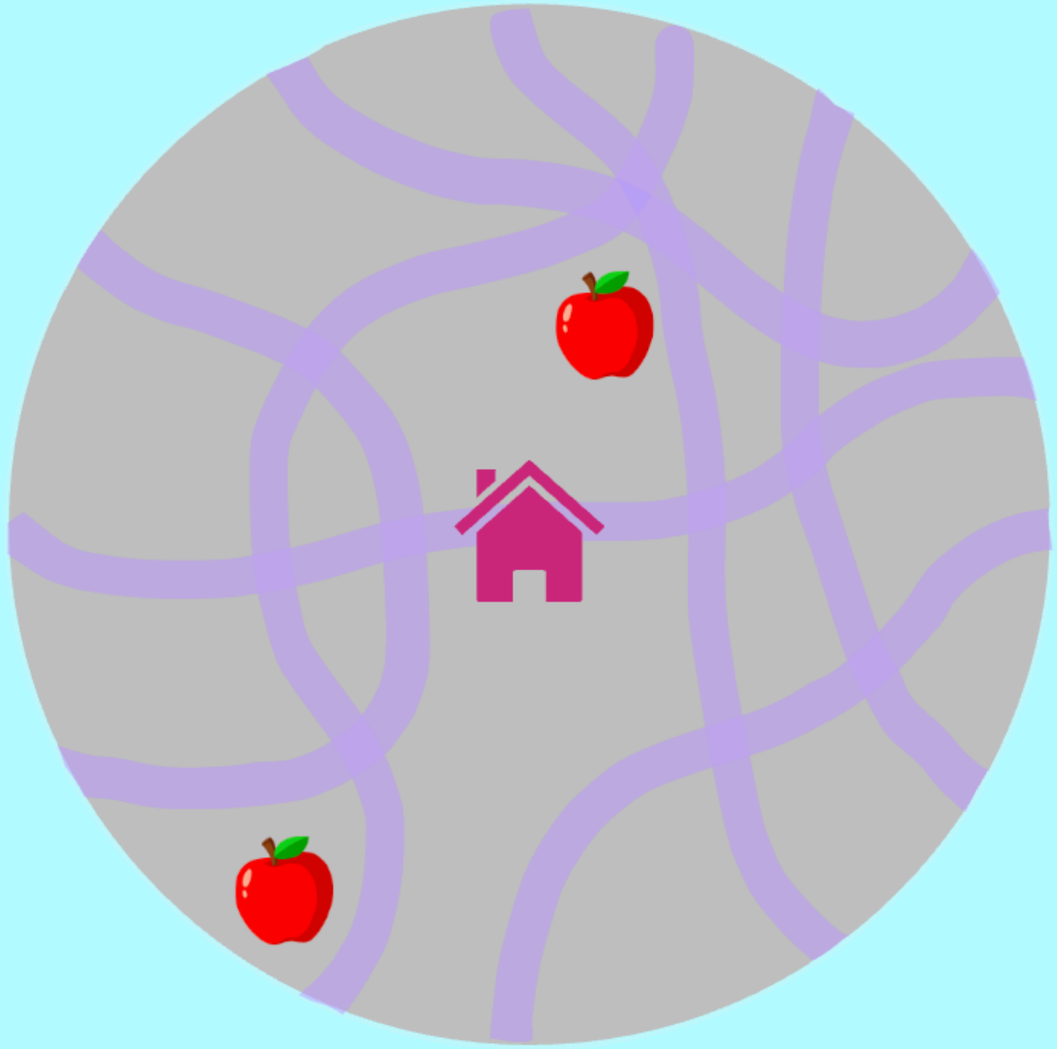
What if we count the number of stores in the neighborhood that sell healthy food?

Let’s focus on access to healthy food!



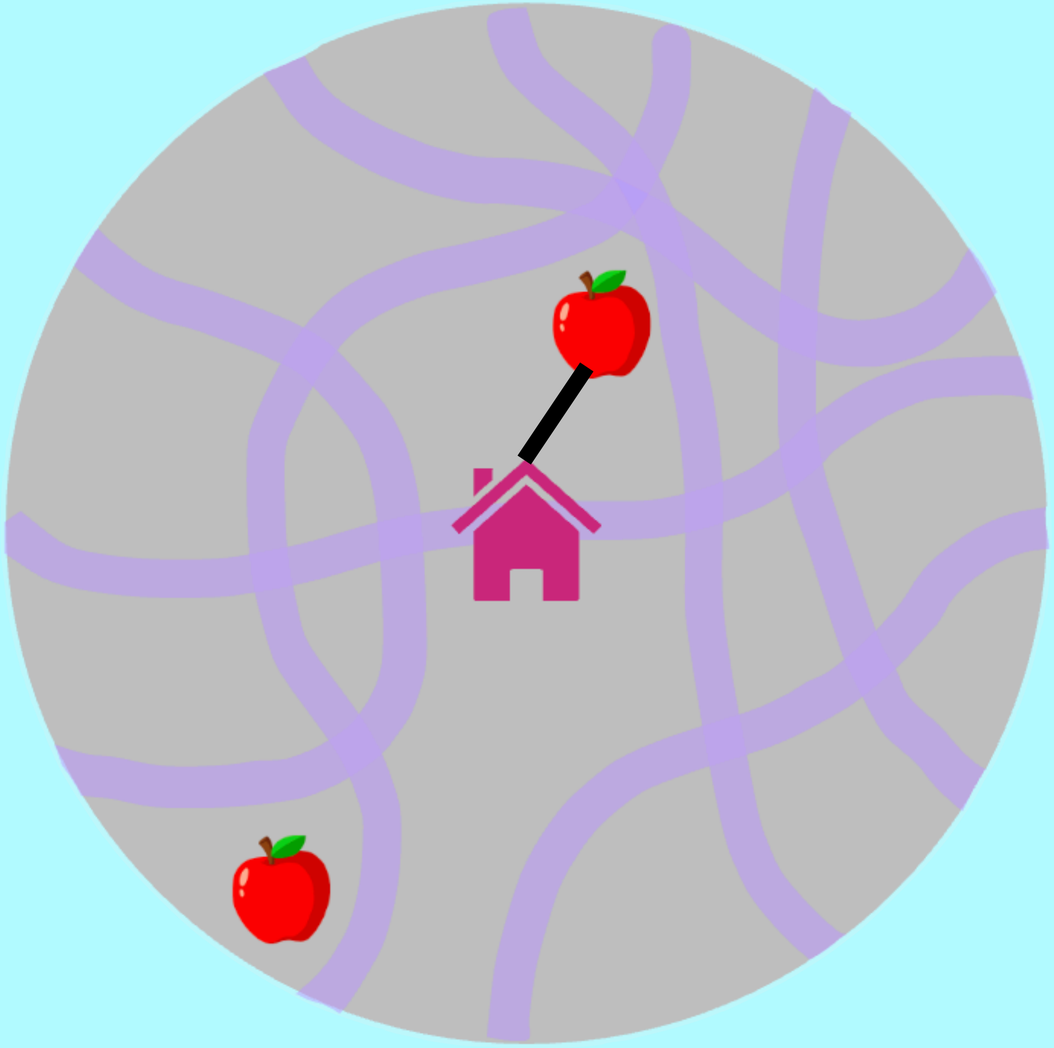
What if we find the closest store that sells healthy food and see how far it is from us?

Let’s focus on access to healthy food!



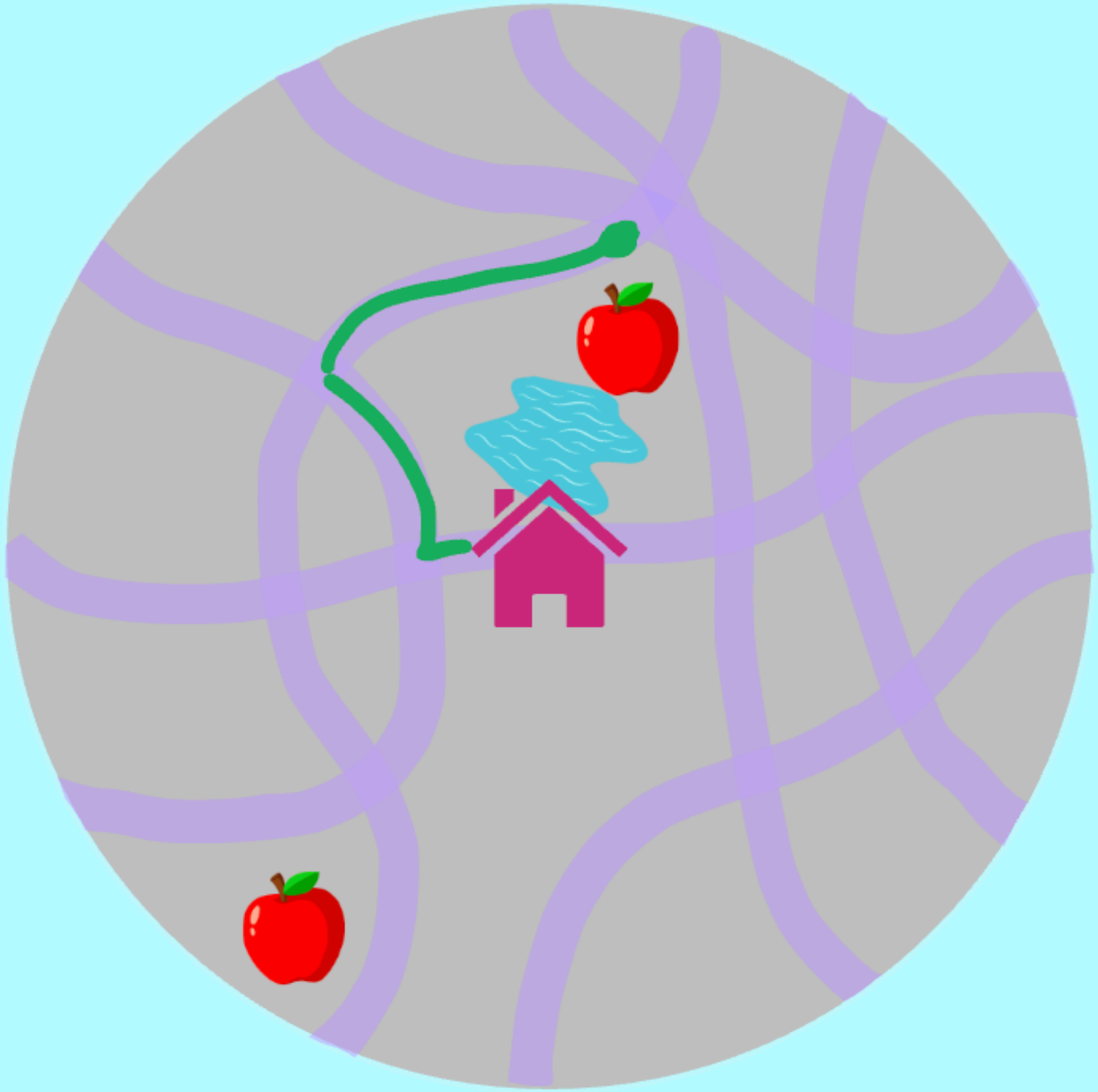
What if we ask, “Is there at least one place to buy healthy food in my neighborhood?”

This is the easy way to measure distance.



Fun Fact: A straight line is the fastest way to get from one place to another!

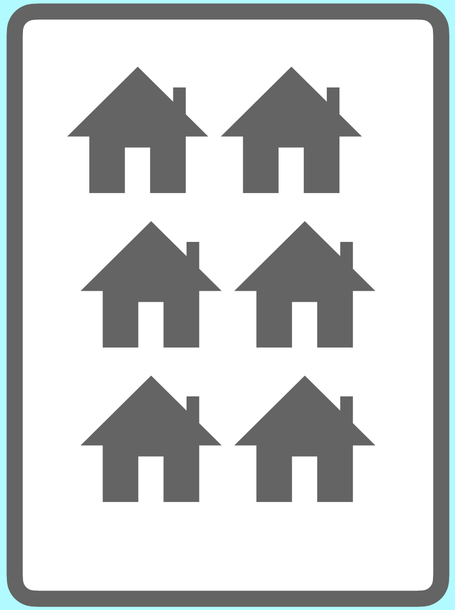
You can’t always do that!



Unless you can walk on water, you’ll have to take the road.

Which is better?

- The straight line option is **easy** to measure but **not always correct**.
- The roads option is **harder to measure** but **more accurate**.
- One way to compromise is to **use a little bit of both!**



Step 1: Measure the straight line distances for all six neighborhoods.

Step 2: Pick two neighborhoods and measure the road distances.

Now, let’s ask a question!

- Your question should be **specific** and **measurable**. Think about **who** we want to study, **what** we want to measure about them, and **how** we want to measure it.
- It’s ok if it doesn’t start out that way! We can always **refine** the question.

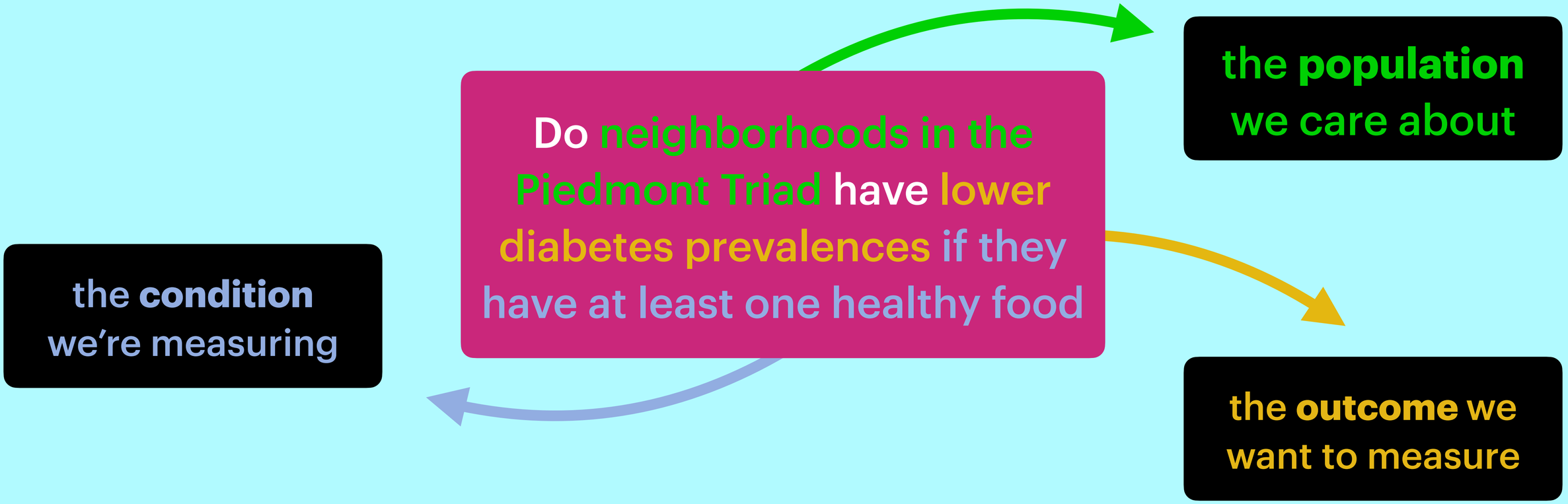
Are healthy neighborhoods less likely to have diabetes?

Now, let's ask a question!

- Your question should be **specific** and **measurable**. Think about **who** we want to study, **what** we want to measure about them, and **how** we want to measure it.
- It's ok if it doesn't start out that way! We can always **refine** the question.

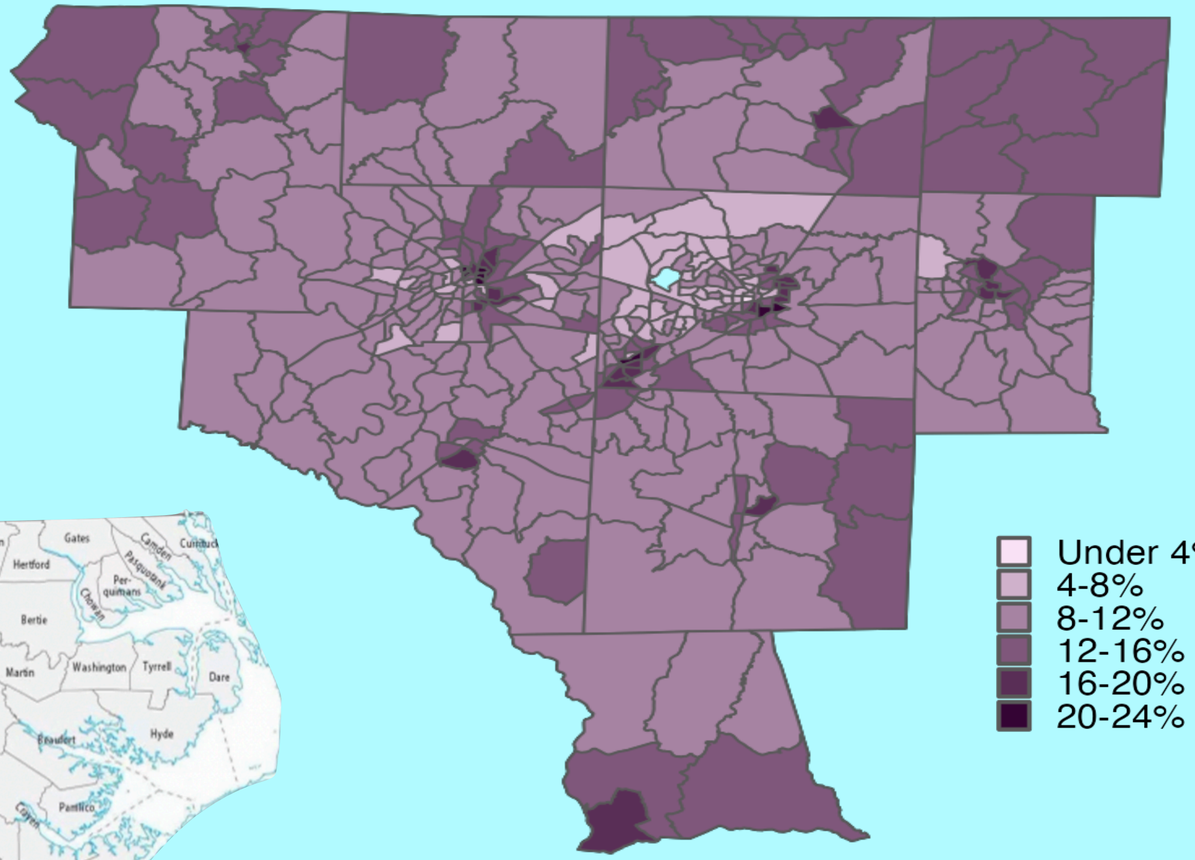
Now, let's ask a question!

- Your question should be **specific** and **measurable**. Think about **who** we want to study, **what** we want to measure about them, and **how** we want to measure it.
- It's ok if it doesn't start out that way! We can always **refine** the question.



Let’s explore diabetes prevalence in the Triad.

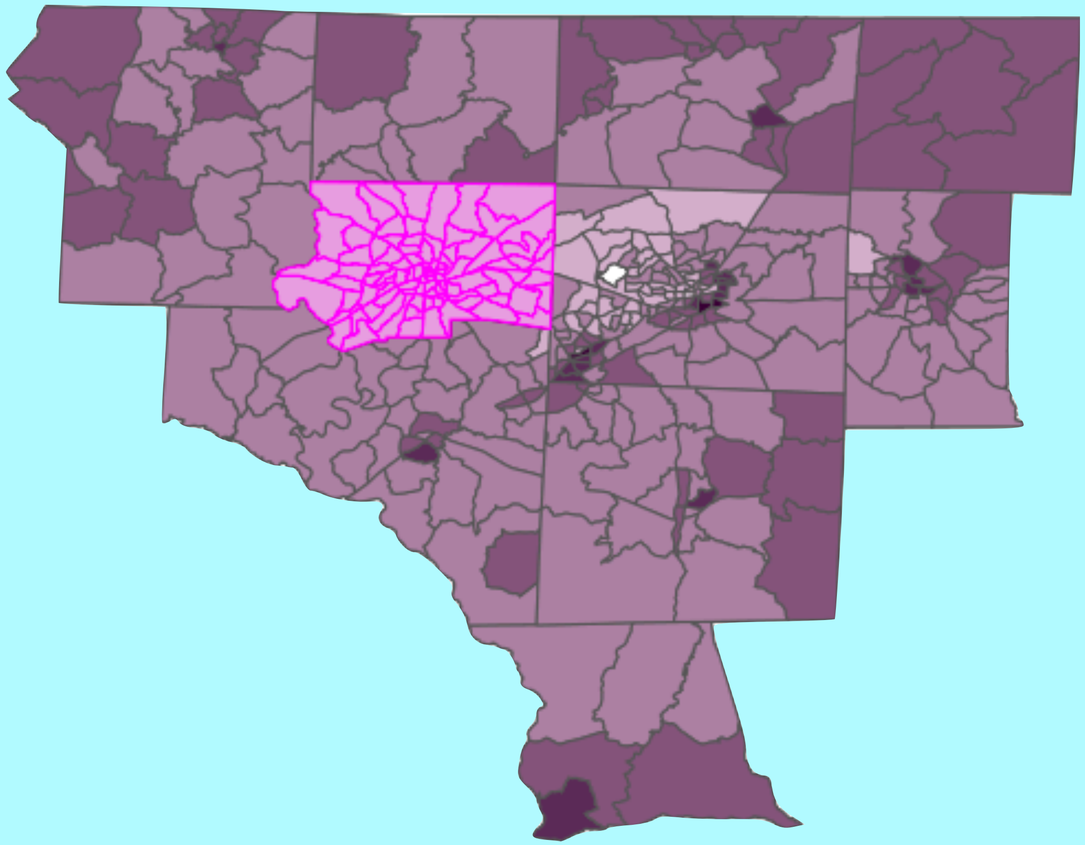
Piedmont Triad

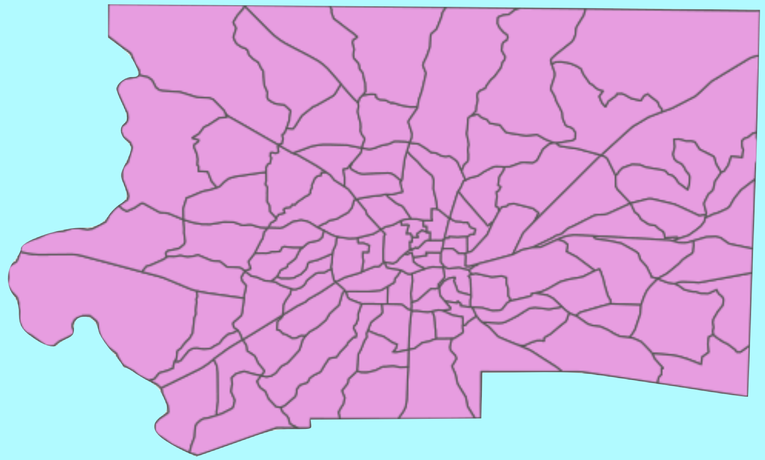


North Carolina

Can we zoom in a little further?

Piedmont Triad

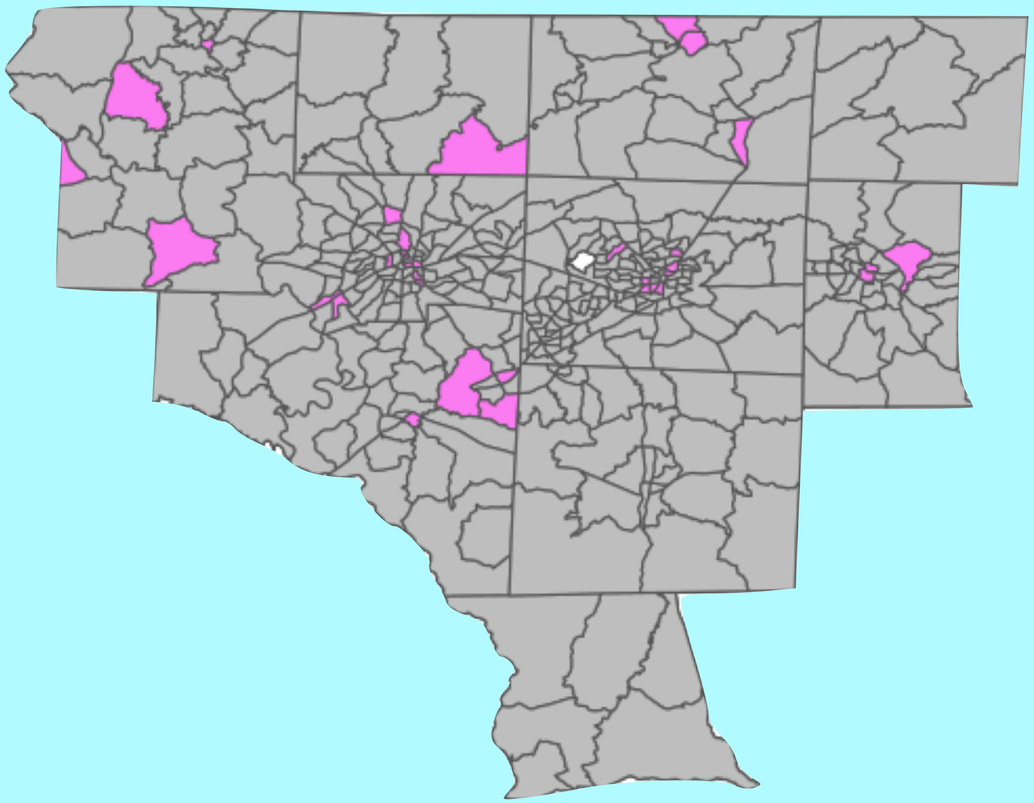




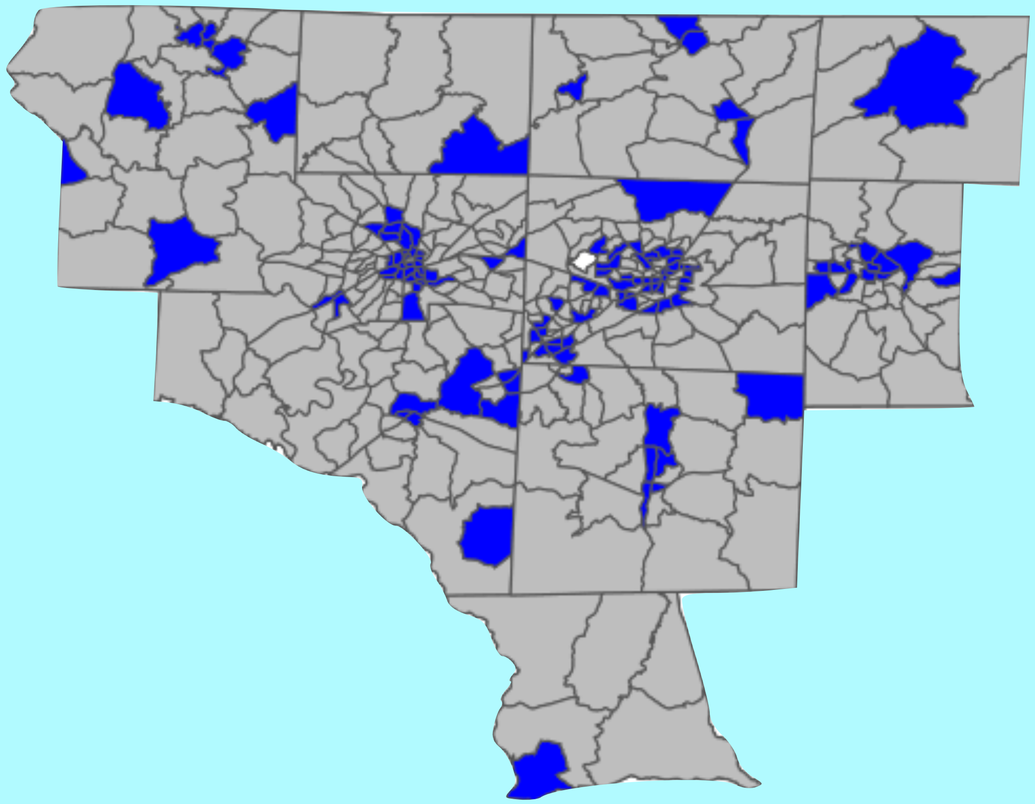
Forsyth County

A **census tract** is defined by the US Census Bureau as small, relatively permanent statistical **subdivision of a county** or statistically equivalent entity. These little zones are census tracts!

Let’s map out food access in the Triad.



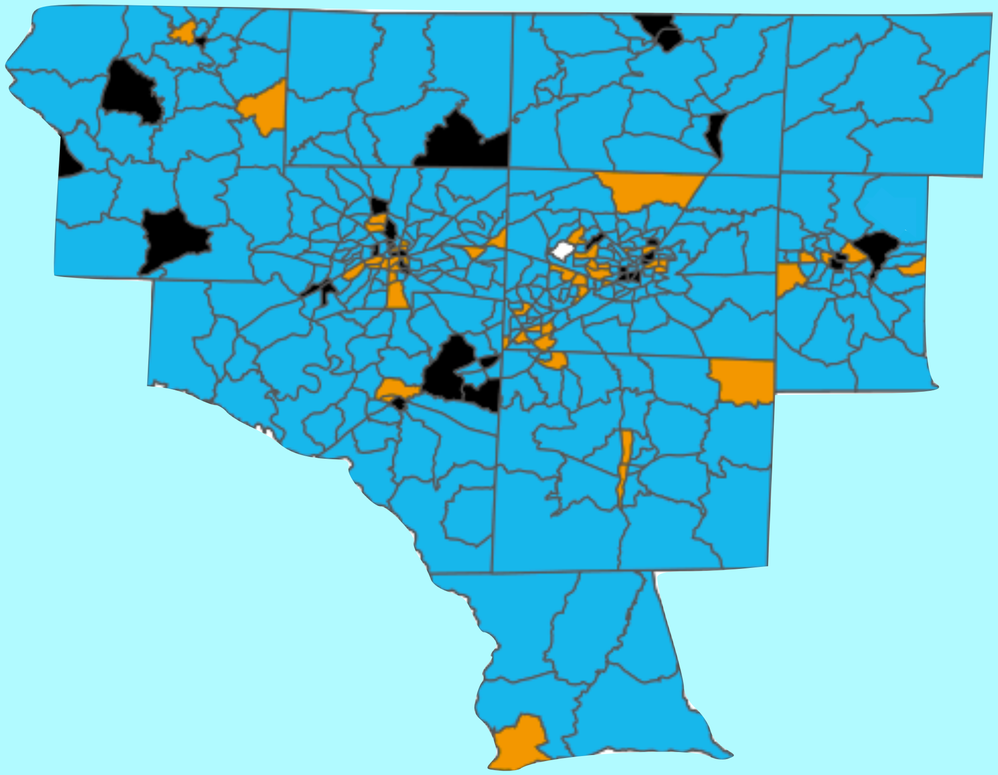
🍏 (1/2 Mile)



🍏 (1 Mile)

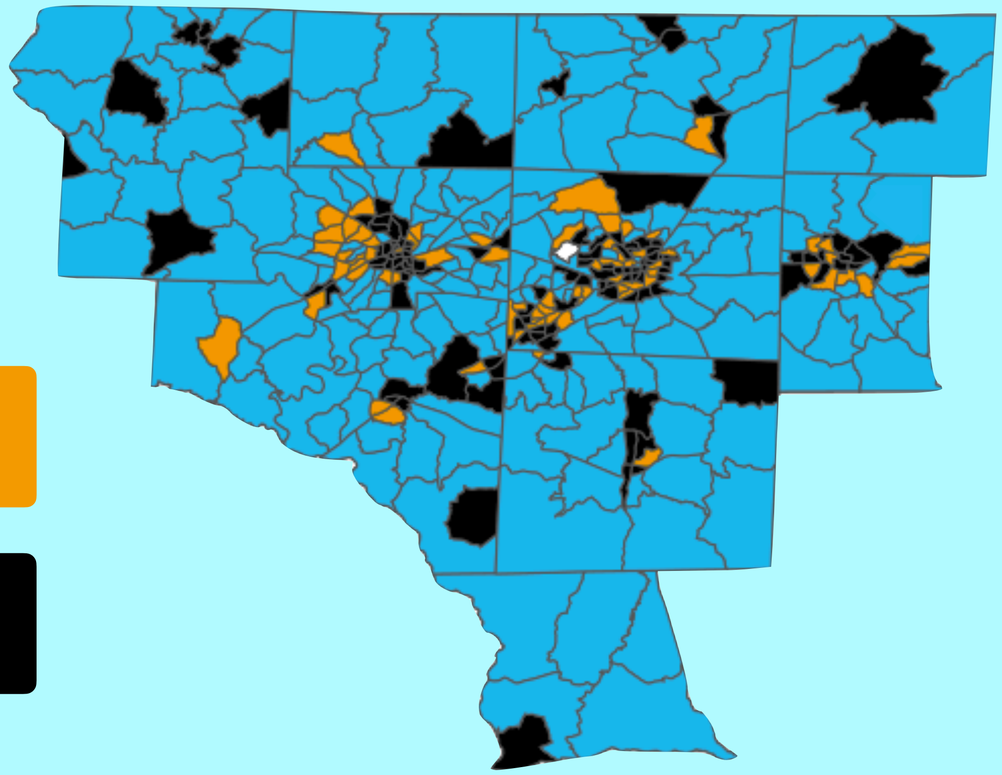
Our data aren’t always so perfect!

1/2 Mile



X

1 Mile



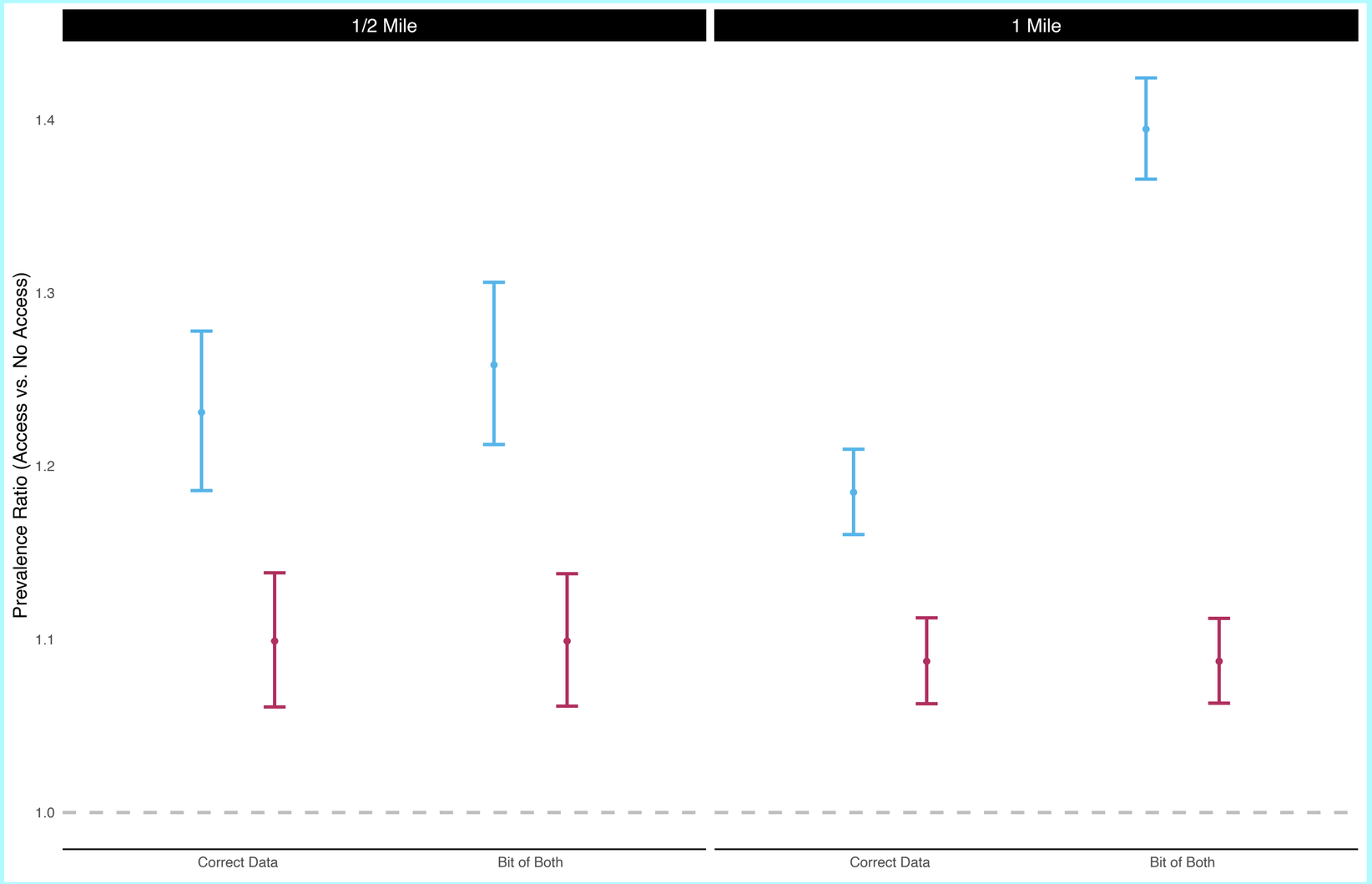
✔

Let’s put diabetes and food access together!

- We’re going to **model** what we expect diabetes prevalences to look like in a tract with food access at that radius when compared to one without it.
- We’re going to account for whether the tract is in a **city** or not, since we think it might **change our guess**!
- If the model spits out “1.XYZ”, that means that we expect the prevalence of diabetes to be **XYZ% higher** in tracts with access to healthy food than in those without it.
- We’re lucky enough to have all of the data we need, but we’re going to try the “bit of both” model too and see if they **agree**.
- We’re also going to give a **95% confidence interval**, which gives our guess a little wiggle room!

What do we expect to see?

The models say:



City

Not City

no relationship

Two Bonus Questions

Everybody, stand up!

1. Look around for the nearest bag of candy to you. There should be three bags!
2. Walk to the nearest bag of candy, counting the number of steps you took, and then return to your seat quietly.
3. Enter how many steps you took and how tall you are (in inches, it’s ok if you have to make a guess) in the spreadsheet!

What other data might we want to collect?

Tell us what other data you might be interested in analyzing! Some options are:

- Sports
- Fashion
- Music
- Community
- Movies
- Business

Please list them one suggestion per cell, but I'll give you a column. We're going to create a **bar chart**!

Here's a quick recap!

- We can use **data** to learn about the world around us by asking **specific questions** that we hope to **generalize**.
- Data can be **hard to get**, so sometimes we have to be **clever** about how we collect and analyze them!
- All models are **wrong**, but some are **useful**.
- We can often use **statistical models** to learn about how knowing something about a person or place can help us guess something else!
- Often, things that are **near each other** are **similar**.

Here's a quick recap!

- We can use **data** to learn about the world around us by asking **specific questi**
- Data ca ut how we coll
- All mod
- We can something about a person or place can help us guess something else!
- Often, things that are **near each other** are **similar**.

THANK YOU!