

How Data Can Help Us Understanding Our World

LaCuKnoS Language Booster

Dr. Yanming Di is a statistician who teaches and does research at Oregon State. He does some of his work with data we get from your SMILE clubs. Yanming was always interested in math and puzzles as a kid and he became fascinated with computers and the idea of artificial intelligence as he was growing up. He decided to study statistics so that he could combine his interests in math, computers, and solving problems. You can read more about Yanming's story later in this activity.



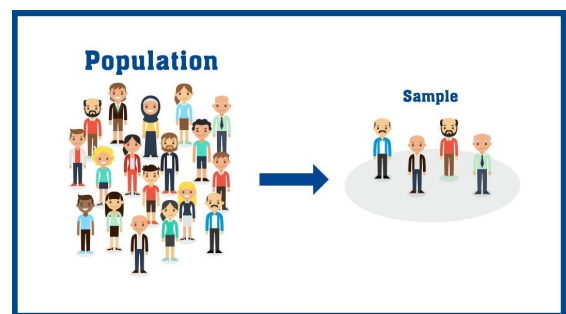
Kerrie Mengersen is another statistician who works with scientists and engineers to better understand our world. She says she has “the coolest job in the world.” [Watch the first 2 minutes of this video about Kerrie Mengersen](#) to see how she uses new technologies like drones and virtual reality headsets to answer questions we could never answer before.

Statisticians like Yanming and Kerrie summarize and represent large amounts of data using graphs or figures. This process is called **data visualization**. Most people can understand information more quickly and easily when they can visualize, or see the data represented in a graph, figure, or table.



Some of Yanming's research analyzes **surveys** of people's ideas about different topics. Have you or your parents ever been asked to answer a survey? Data visualization of surveys can help Yanming to see patterns and relationships in people's opinions. Businesses, governments, schools, and many other groups use surveys. For example, businesses give surveys to customers to understand what products people do or do not want to buy.

When the number of people we want to learn about is very large, we can't ask the questions to everyone. Instead, we can ask the questions to only a part of the whole **population**. The people we choose to ask the questions to are called a **sample**. A sample is used to represent the larger group. But how do we know if the sample really represents the whole group or if the sample is **biased**? The best way to select a sample that represents the whole group is by **random selection**. This is like tossing a coin to decide whether each person will be included or not in the sample.



Surveys of people can tell us many interesting and surprising things. For example, one recent survey showed that about half of American teens say that they are “almost constantly” on the internet. Are you?

Talk with your partner about the following questions:

1. Have you ever been asked to complete a survey? What was the survey about? Why do you think you were asked to take this survey?
2. Would you like to have Kerrie Mengersen's job? Or maybe Yanming Di's job? Why or why not?

Learning with Data LaCuKnoS Investigation

Activity 1 - Pizza! Pizza!

In the first part of this activity you will collect survey data about one question from everyone in your SMILE club and then you will represent and visualize the results of your survey.



Materials:

- Paper plates (1 per student)
- Colored pencils
- Ruler or straight edge

Procedure:

Survey Question for your club: What is your favorite pizza topping?

- a. Cheese
- b. Pepperoni
- c. Pineapple
- d. Mushroom
- e. Other topping

1. Draw a data table on the white board for collecting your club data. Each student should add to the total with their favorite topping. Here's an example:

Favorite Pizza topping	Cheese	Pepperoni	Pineapple	Mushroom	Other
student responses	5	7	3	2	5

2. Work with a partner or small group. Use a paper plate to create a “Pizza pie graph” to represent your data.
3. How big should each slice of the pie graph be on your paper plate? You can use different colors to represent the different pizza toppings on your pie graph.

Talk with your partner about these questions, then write your answers.

Do you think your SMILE club data on favorite pizza toppings is similar to the results for students in other SMILE clubs in Oregon? Why or why not?

What kind of business might be interested in knowing people's favorite Pizza toppings? Why?

4. Look at the data visualization below. It represents people's favorite pizza toppings in each state in the U.S.



Talk with your partner about these questions, then write your answers.

Does this data visualization help you to see patterns in the data? What patterns do you see?

Why might different pizza toppings be more or less popular in different parts of the country?

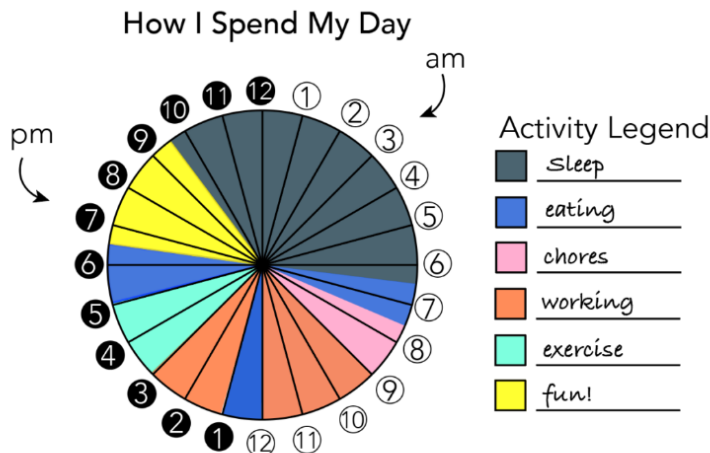
The figure says that the most popular pizza topping in Oregon is pineapple? Do you believe that? Why or why not? Is there a way we could test this?

Activity 2 - How I spend my day

In the second part of this activity you will compare how you use the hours in your day with how one of your family members uses the hours in their day.

Materials:

- Paper plates (2 per student)
- Colored pencils
- ruler or straight edge



Procedure:

Survey Question: On a normal school day, how much time do you spend doing each activity?

- Sleep
- School/ job
- Chores/ Homework
- Eat
- Exercise/ physical activity
- Fun/ relax/ entertainment

- First, complete the data table on the next page to record what you normally do for each hour of a typical school day.
- Take a paper plate and a ruler or straight edge and divide your plate into 24 slices to represent the hours in the day. Number them like a 24 hour clock. See this picture above for an example.
- Use the paper plate to create a pie graph to represent the data for your day. Use different colors to represent the different activities. See this picture above for an example.
- Next, you will compare how you spend your day with an adult in your family.
- Create a 2nd paper plate like the one you made in step 2. Give your 2nd data table and 2nd plate to someone else in your family. Explain the directions for how to visualize their time. Show them your plate as an example.
- Bring both of your plates back to your SMILE club.

Typical Day Data Table

Hour of the day	Sleep	School/ job	Chores/ homework	Eating	Exercise	Fun/ entertainment
12 AM (midnight)						
1 AM						
2 AM						
3 AM						
4 AM						
5 AM						
6 AM						
7 AM						
8 AM						
9 AM						
10 AM						
11 AM						
12 PM (noon)						
1 PM						
2 PM						
3 PM						
4 PM						
5 PM						
6 PM						
7 PM						
8 PM						
9 PM						
10 PM						
11 PM						

Follow up:

In small groups, share both of your plates and talk about them.

Talk with your partner about these questions, then write your answers on the back of your paper plate.

What are the biggest differences between my two pie graphs (student and adult)?

What are the biggest differences between the pie graphs for students and the pie graphs for adults in my SMILE club?

What differences do you see between the pie graphs for different families?

We want to see your pie graphs!

1. Write your name and the answers to the questions on the back of your plate.
2. Staple your 2 plates (yours and the adult you asked) together and put them in the envelope to send back to SMILE.
3. If your club does the Extension #1 activity below (An Oregon map of pizza toppings) send your club's pizza topping data back to us in the same envelope.
4. If more that half of your SMILE club members send back your data, your Club will

Want to do More? Here are some extension ideas

Extension idea #1 - An Oregon map of pizza toppings

In Activity 1 we learned that the most common pizza topping in Oregon is pineapple. We can use our SMILE clubs around the state to create a more detailed map of favorite pizza toppings in different regions of the state. Here's what to do:

1. Send your data about favorite toppings from your club back to SMILE (include it with your pie charts of how you spend your time).
2. We will organize the data into regions of the state and send it back to all the SMILE clubs
3. See if your club can find interesting ways to visualize the data for all the clubs.
4. Send your data visualizations back to us!

Extension #2 - Exploring more about careers in statistics

Read Yanming Di's Story – How I Became a Statistician

How did you become interested in statistics?

I liked math and puzzle books when I was a kid. In my fourth grade, I saw computers for the first time when I joined an afterschool club, and I was immediately intrigued. Like many kids, I liked computer games, but I was also fascinated by how you can write computer programs to solve many puzzle games. Then I was exposed to the words like “artificial intelligence” and “machine learning” and they were mind blowing to me at that time. I tried to learn more about computers and coding, and to understand how computers can learn and have intelligence.

Why did you decide to pursue this career and then why the specific job of statistics professor (rather than another job using statistics)?

I majored in math in college. I knew I wanted to become a professor because I always admired my teachers and professors and liked the environment of a college campus. So I continued to pursue a graduate degree after college. When choosing a research area, I eventually decided on statistics: I feel that statistics is a subject where I can combine my math skills and computer programming skills. I chose the PhD program at University of Washington because there were a few professors there that were working on machine learning. Once I entered the program, I learned that there are so many other areas where we can use statistics. I eventually did my PhD research in statistical genetics: I studied how we can find genes that underlie complex diseases. I became a college professor after graduation.

The Era of Data Science

I now have a better understanding of what is machine learning: it means **understanding our world through data**. To achieve the understanding, we have to be able to collect high-quality data that faithfully represent the world through survey samples, designed experiments (e.g., clinical trials).

Nowadays, a lot of what we statisticians and many computer scientists do are collectively called **data science**. Students in statistics or data science can often find jobs in banks, pharmaceutical companies, government agencies, IT companies. Basically, any field where data are collected will need help from data scientists.

Here are several more videos about statisticians and what they do. Groups of students can pick one of the videos to watch and then share something interesting that they learned with the rest of the club.

Super Bowl statistician - <https://www.youtube.com/watch?v=G8msa3YEBKQ>

Statisticians - <https://www.youtube.com/watch?v=gXCgwnkNL0w>

Hans Rosling - The Joy of Stats - <https://www.youtube.com/watch?v=jbkSRLYSojo>

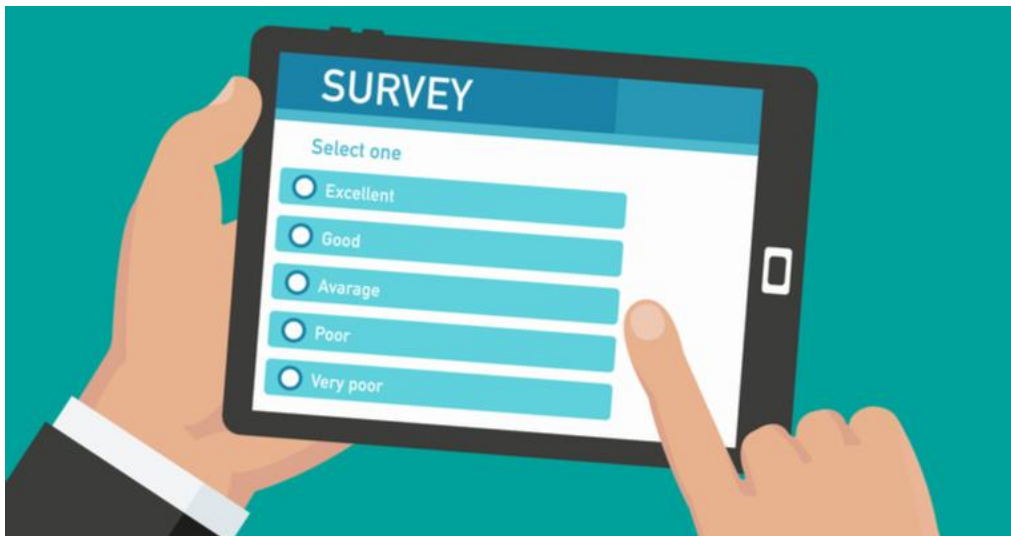
Why you should study statistics - [Why You Need to Study Statistics - YouTube](#)

Learning with Data
LaCuKnoS Concept Cards

Survey/Encuesta

A set of questions about a topic that people are asked to complete.

Un conjunto de preguntas sobre un tema que se les pide a las personas que completen.



The shoppers took a survey about their favorite brand of shoes.

Concept Card

Population/ Población

The people who live in an area, considered as a group.

Las personas que viven en un área, consideradas como un grupo.



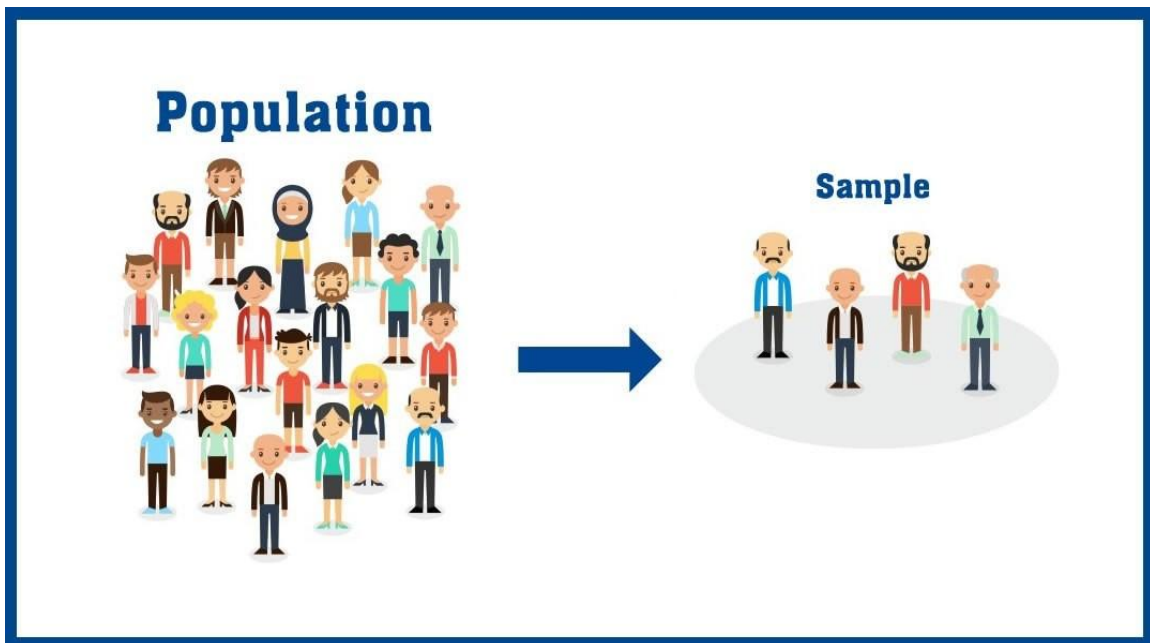
The population of Oregon has been growing quickly in the past few years.

Concept Card

Sample/ Muestra

A part or subset of individuals from the whole population.

Una parte o subconjunto de individuos de toda la población.



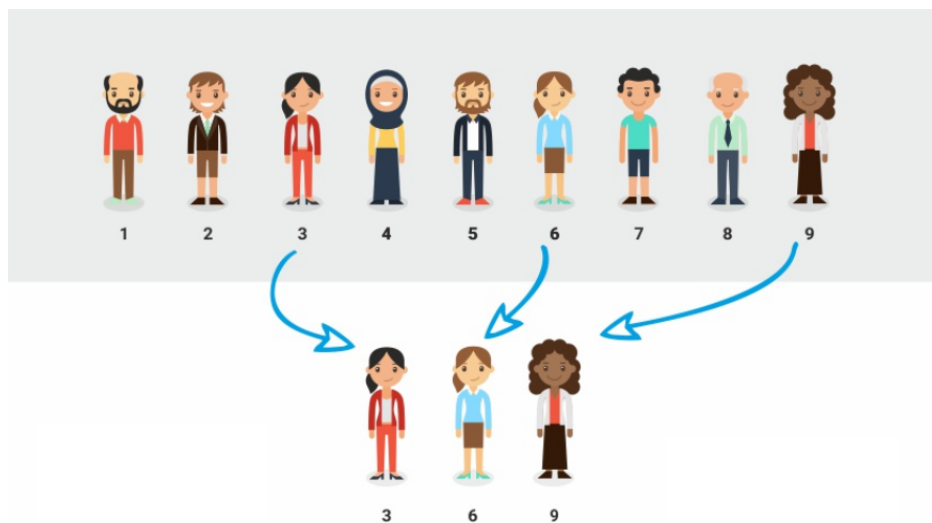
Researchers who use surveys typically pick a sample from a larger population.

Concept Card

Random Selection/ Selección Aleatoria

When every individual in a population has the same chance to be picked as part of a sample.

Cuando todos los individuos de una población tienen la misma oportunidad de ser elegidos como parte de una muestra.



Experiments in science often use random selection.

Concept Card

Bias/Sesgo

An opinion or decision that is unfair. A result that is not representative of the whole group.

Una opinión o decisión que es injusta. Un resultado que no es representativo de todo el grupo.



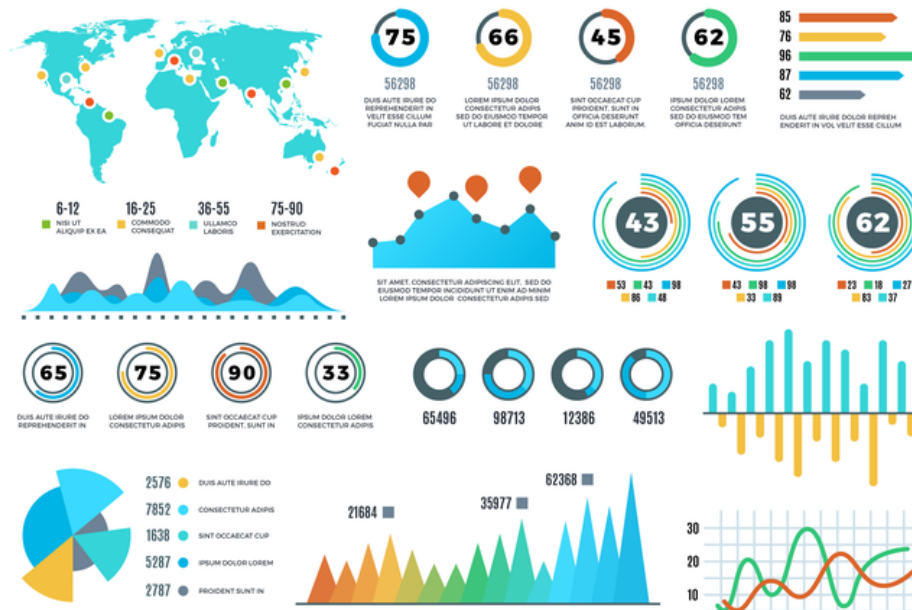
Part of thinking like a scientist is to try to identify our own biases.

Concept Card

Data visualization/ Visualización de datos

Using figures or tables to summarize your data and highlight interesting findings.

Usar figuras o tablas para resumir tus datos y resaltar hallazgos interesantes.



Data visualization has become an important part of statistics.

Concept Card

You can find the lesson here!



LaCuKnoS