AGAINST ALL ODDS
EPISODE 2 – “STEMPLOTS”
TRANSCRIPT

FUNDER CREDITS

Funding for this program is provided by Annenberg Learner.

INTRO

Pardis Sabeti
Hello, I'm Pardis Sabeti and this is *Against All Odds*, where we make statistics count.

Statistics is all about the data. And frankly it's easy to feel overwhelmed by an avalanche of numbers if we don't figure out good ways to organize it all.

One of the best places to start is with a picture. You've seen informative pictures like these before… pie charts, bar graphs. Visualizing data like this can be a good first step toward organizing it and understanding it. And it provides a framework to contextualize a single number you might encounter. For instance, my friend told me her new baby weighed in at 7lbs and 8oz… I'm not a baby expert, but I can use these graphs to get a sense of where on the continuum from tiny to gigantic this one baby fits in. Looks like he's right about in the middle!

There are a variety of ways to visualize raw numbers…. And tons of real world data sets to work with! Let's step into the Army's boots for a moment to see the data it collected to help outfit each and every soldier with the right size uniform and gear.

Since the Civil War, the Army has kept tabs on the size of its soldiers, tracking changes over the decades. Throughout the 1900s, Americans grew taller due to better nutrition and medical care. Once that countrywide growth spurt had maxed out, the Army thought it had a handle on its soldiers' dimensions.

But in 2003, as Army leaders started outfitting soldiers for Operation Iraqi Freedom, they noticed they had a gear shortage on their hands.

> **Cynthia Blackwell**
> There were some indications that we didn't have the right numbers of sizes of equipment and it was critical equipment because back then you might remember we were talking about dirty bombs and that sort of thing so we needed to make sure they had the right chemical equipment, and ballistic equipment and we found that we did not have that in the correct numbers of the correct sizes. And so that was like ding ding ding bell kind of situation because we had had them in the past.
>
> **Claire Gordon**
> And this was a clue to us that our database was out of date. Now we'd been tracking the increase in body size in civilians, but you don't necessarily expect the same increase in body size in the military because we have weight for height standards.

Pardis Sabeti

Stockpiles hadn't been adjusted to reflect the changing shape of today's all-volunteer Army. It was time to call in the anthropometric team – they're the scientists who conduct the survey to update the size database. They would figure out how big today's soldiers are… and how big the equipment designed for them needs to be.

Researchers traveled the country to multiple bases and randomly sampled almost 12,000 active duty soldiers, National Guard, and Reserve Members. The goal was to learn the complete size range of today's soldiers and how common each size is.

> **Cynthia Blackwell**
> Each person we took 94 dimensions, that's heights, weights, breadths, depths, those kinds of dimensions and three scans: One of the whole body, one of the head and face, and one of the foot. / But there are a variety of ways in which you can gather that data. You use the right tool for the right job so to speak. / Sometimes when they were just doing, for example, circumferences they would just use a regular tape measure.

Pardis Sabeti
The measuring team confirmed that soldiers are bigger now than they had been in the past. From 1966 to 1988 they had already seen a weight increase due to more muscular physiques, thanks to the popularization of weight training. Now they were seeing an increase in body circumferences related to where body fat is stored. As part of our national obesity epidemic, soldiers are bulking up right along with their civilian counterparts.

Whatever a soldier's size, it's crucial that his gear fits well so that he can do his job and be protected. The survey's precise measurements become a valuable resource for designers working on the next generation of everything from body armor and boots, to workstations and vehicles. This was the first big survey that created 3D scans of each soldier that can be used when designing a new cockpit, for instance.

> **Claire Gordon**
> To do that without physically creating a cockpit model and putting people in it, which is very costly and time-consuming, what you would do is you would take the dimensions of the cockpit, put it into a virtual space on the computer, then you would need a human of particular dimensions that you want to fit to also put virtually in that space and then to go through a series of tasks. / Now the dimensions of that digital human have to reflect the dimensions of the users.

Pardis Sabeti
And with a better sense of the most frequently-found dimensions nowadays, the army now knows which sizes to keep well-stocked and which sizes are rare

enough that it's cheaper to custom order them. But saving taxpayer dollars isn't the only concern.

> **Cynthia Blackwell**
> It's a particular / underlying consideration that we who develop and design equipment always have in the back of our minds…. It's soldiers' lives, it's soldiers' lives, it's soldiers' lives.

Pardis Sabeti
Here's the foot measurement data in centimeters for 30 of the soldiers. When you see a bunch of unorganized numbers it's hard to tell if they have any rhyme or reason. Let's organize these data into a stemplot to get a better sense of how widely foot size varied.

First, using software or a calculator, we can sort them into numerical order, which already gives us a little bit better idea of what's going on.

Now let's separate each measurement into a stem and a leaf. The stem is the first digits and the leaf is the final digit. Line up the stems in order vertically… then fill in the leaves on the appropriate stems. In this case, the final digit is the tenths of a centimeter. Always include all possible stems in your data range, even those that don't have leaves to go with them. The final step is to organize the leaves in numerical order.

Displayed as a stemplot, we can see the overall pattern of our data: the 26 centimeter values are the most common, and values on either side of that single peak are less common.

In this case, the peak looks to be around the center of the distribution… which we can calculate. There are 30 measurements, half of 30 is 15. Counting up 15 observations, starting at the smallest, lands us at 26.8, so that's roughly the center or midpoint.

A stemplot also lets you see at a glance how spread out the distribution is. The data points range from the smallest at 24.6 centimeters to the largest at 32.8 centimeters.

Check out the overall shape too. For the most part the stemplot looks pretty symmetric… except for that value out there at 32.8. An individual measurement like this one that falls outside the overall pattern of the data is called an outlier.

It's important to try to figure out what's going on when you see an outlier. Is it a mistake? Did someone just type the digits in wrong? Did someone measure in the wrong units? Or deliberately add a joke number? Sometimes outliers are in fact 'real data' but it's good to take a second look to make sure.

Stemplots work best with data sets that are relatively small… you wouldn't want to make a stemplot of foot measurements for all 12,000 soldiers who were sized up – it would be too unwieldy to help us understand more about the raw data. But we can use the stemplot to contextualize where a particular number fits in with the rest of a distribution. If we take a single measurement from the big group – lets say 25.1 – where it fits in with our stemplot gives us a sense whether a particular soldier's foot is on the larger or smaller size – something that would be harder to determine just looking at the raw numbers.

Here's another dataset where a stemplot can help us visualize the numbers – fuel economy info for Toyota's vehicle line. We arrange the miles per gallon numbers in order, this time from highest at the top to lowest at the bottom. Then we build our stem. Since the MPG data are all whole numbers, we'll use the tens place as the stem and the leaves will be the ones place. Include stems even for numbers that don't have leaves to go with them.

Take a look at the overall pattern. Most of the MPGs are clustered here at the lower end. We can expand the stem to change the resolution of our picture. If we zoom in on the data, will it help us understand the numbers even better? Let's break each stem into two, so the low digit leaves 0,1,2,3,4 are on a different stem than the high digit leaves 5,6,7,8,9.

We've got outliers again… but this time an explanation is obvious. The high numbers are due to the super fuel-efficient hybrid vehicles Toyota makes.

Stemplots can be used to compare two different data sets as well. Say we wanted to compare Toyota's 2012 numbers with those from their 1984 line. We can make a back-to-back stemplot to see how mileage numbers have changed over the decades. Use the same stem… for two separate sets of leaves.

Interestingly, today we have more vehicles way down at the low end, and a few more up at the high end. These extremes are easy to explain when you think about what you see on the roads these days – modern car buyers are interested in the not-so-fuel-efficient SUVs and trucks as well as the uber-efficient hybrids.

So you can see how stemplots start to wrangle some meaning out of the disorder of raw data. They are useful for visualizing the shape of your data's distribution, and figuring out how frequently particular data classes pop up in your sea of numbers. And stemplots are just the first step… there are plenty of other statistical ways to visualize data.

Stay tuned! I'm Pardis Sabeti for *Against All Odds*.

PRODUCTION CREDITS

Host – Dr. Pardis Sabeti

Writer/Producer/Director – Maggie Villiger

Associate Producer – Katharine Duffy

Editors – Jared Morris
- – Seth Bender
- – Bret Upham

Director of Photography – Dan Lyons

Additional Camera – Griffin Nash

Audio – Timothy Wessel

Sound Mix – Richard Bock

Animation – Jason Tierney

Title Animation – Jeremy Angier

Web + Interactive Developer – Matt Denault / Azility, Inc.

Website Designer – Dana Busch

Teleprompter – Sue Willard-Kiess

Hair/Makeup – Emily Damron

Additional Footage and Stills
- iStock/WebSubstance
- Department of Defense
- Library of Congress, Civil War Collection, LC-DIG-cwpb-01675
- Library of Congress, Panoramic Photographs Collection, Army no. 130
- Library of Congress, FSA/OWI Collection, LC-USW3-017521-D & LC-USW33-000354-ZC
- Rob Hawley, Natick Soldier Research, Development & Engineering Center
- WDAM-TV
- Army Research Laboratory
- Toyota Motor Corp.

Music

DeWolfe Music Library

Based on the original Annenberg/CPB series *Against All Odds*,
Executive Producer Joe Blatt

Annenberg Learner Program Officer – Michele McLeod

Project Manager – Dr. Sol Garfunkel

Chief Content Advisor – Dr. Marsha Davis

Executive Producer – Graham Chedd

FUNDER CREDITS

Funding for this program is provided by Annenberg Learner.

For information about this, and other Annenberg Learner programs, call 1-800-LEARNER, and visit us at www.learner.org.