One-way tables

Individuals and variables

The simplest kind of information we'll work with in this course is a set of **individuals** with one or more properties, called **variables**. The individuals are the items in the data set and can be cases, things, people, etc.

In this table for "heights of people in the car," the individuals are the people, and their heights are a variable because height is a property of each individual.

Person	Height
Wendy	5'6"
Michael	5'9"
Rachael	5'3"
Allen	5'11"

Together, individuals and variables are called **data**. When we organize data into a table like this one, we call it a **data table**. Here's an example of a data table about ice cream:

Flavor	Scoops sold	Contains chocolate?	Smooth or chunky?
Vanilla	300	No	Smooth
Chocolate	450	Yes	Smooth
Cookies & Cream	275	Yes	Chunky
Mint Chocolate Chip	315	Yes	Chunky
Fudge Brownie	375	Yes	Chunky
Rocky Road	250	Yes	Chunky

The individuals are the flavors: Vanilla, Chocolate, Cookies & Cream, etc., and the variables are their properties: Scoops sold, Contains chocolate, and Smooth or chunky.

Variables can be **categorical** or **quantitative**. In the table for "heights of people in the car," there's one quantitative variable: the height. In the table for "ice cream shop data for July," there's one quantitative variable (scoops), and two categorical variables (contains chocolate and smooth or chunky).

- **Categorical variables** are non-numerical variables. Categorical variables are also called "qualitative" variables. Their values aren't represented with numbers. Whether or not the ice cream contains chocolate, and whether the ice cream is smooth or chunky are categorical variables. This is because the values there are words, not numbers.

- **Quantitative variables** are numerical variables. Their values are numbers. The height of the people in the car and the number of

scoops of ice cream sold are quantitative variables, because the values there are numbers.

We can divide quantitative variables into **discrete variables** and **continuous variables**.

- **Discrete variables** are those that we can obtain by counting. Therefore, they can take on only certain numerical values. For example, the number of scoops of ice cream sold is a discrete quantitative variable, because one cannot sell 8.3 or 5.23 scoops of ice cream.

- On the other hand, **quantitative variables** may include data as decimals, fractions, or irrational numbers. For example, the height of the people in the car is a continuous quantitative variable.

Levels of measurement

When we work with data, we need to understand the level of measurement, because not every statistical test can be used with every type of data set. There are four levels of data measurement: nominal, ordinal, interval, and ratio.

- Things like favorite food, colors, names, and “yes” or “no” responses have a **nominal** scale of measurement. Only categorical data can be measured with a nominal scale.

- Categorial data can also be **ordinal**. This type of data can be ordered. The top three national parks in California is an example of

ordinal data. Or, for example, when we ask a group of people about how they liked their trip, we may get responses like "awesome," "good," "satisfactory," or "terrible," which follow an order from best to worst.

- Data measured using an **interval** scale can be ordered like ordinal data. But interval data also gives us a known interval between measurements. For example, temperature is measured using an interval scale, because we can understand the exact interval of difference between 50 and 60 degrees.

- Data measured using a **ratio** scale is just like interval scale data, except that ratio scale data has a starting point, or absolute zero. Whereas interval data like temperature can have negative and positive values, things like time, height, and weight are examples of ratio scale data because those measures can never be negative.

When we have a single individual and a categorical or quantitative variable assigned to it, we can construct a one-way table. These kinds of tables are called "one-way" because they usually represent data for one individual only.

Constructing one-way tables

When we construct a table, we want to think about whether we have more individuals or more variables. In the two tables we looked at, we had the individuals listed down the left-hand side, and the variables listed across the top.

But if we have lots of variables but only a few individuals, it can be helpful to list the individuals across the top and the variables down the left side. For example, if we're comparing lots of information about two houses for sale in the same neighborhood, we might make a table like this:

	317 Spruce Rd	819 Lilac St
Price	$299,000	$349,000
Square footage	3652	3812
Price per square foot	$82	$92
Bedrooms	4	5
Bathrooms	2.5	3
Stories	3	2
Basement	Finished	Unfinished
Garage spaces	3	2
Lot acres	0.36	0.31
Grass backyard	Yes	yes
Year built	1974	2001
Property tax	$2,356	$2,595
Payment	$1,120	$1,045

The individuals are the addresses listed across the top of the table, 317 Spruce Road and 819 Lilac Street. All the variables are listed down the left side of the table. Since there are so many more variables than individuals, listing the variables vertically makes the table fit better on paper than if we had tried to list all the variables horizontally.