Courseware   **Course Info**   **Discussion**   **Syllabus**   **Download R and RStudio**   **R Tutorials**   **Readings**   **Contact Us**

Progress   **Office Hours**

GRAPHING A SINGLE VARIABLE
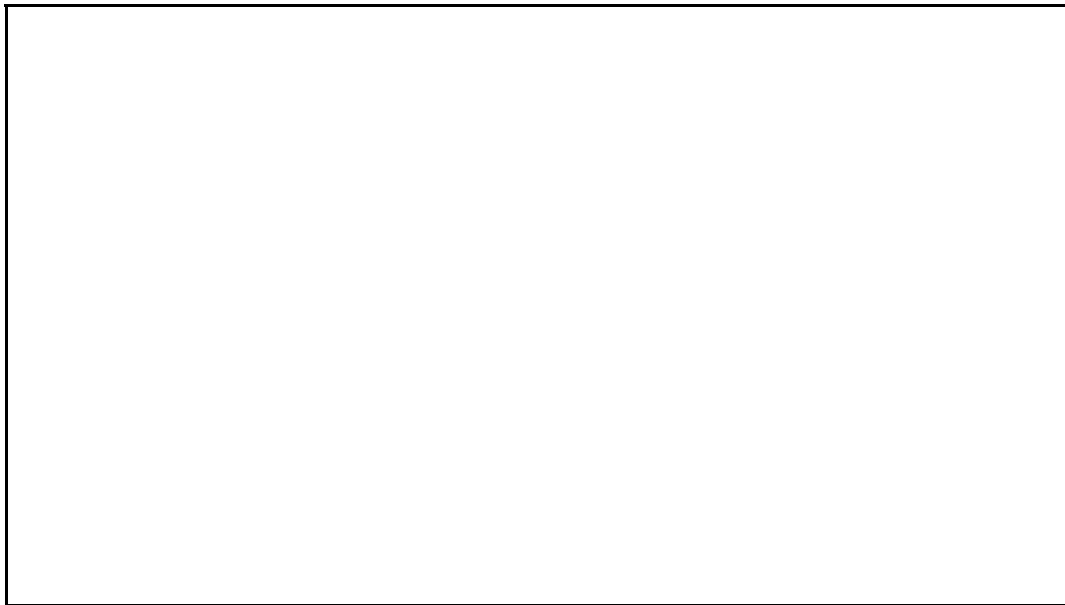
SPEAKER: MICHAEL J. MAHOMETA, Ph.D.

We've seen the importance of a graph - a visualization

to help tell a story about our data or to help

answer a question that we might have.

But to first get to that visualization, we

have to summarize our data in some way.

And the easiest way for us to do this is through aggregation into a table.

4:49 / 4:49   1.0x

11/17/2014 01:12 PM

Now the variable type that we're talking about

- whether it be categorical or quantitative

- will help dictate a type of table that we use and, ultimately,

the type of graph that we can generate.

Here's how.

As always, let's start with a question.

So we've got data on driving preference during inclement weather,

and here's a description of each of the variables.

Choice to drive is a categorical variable, and our choice of graph

is restricted by the variable type.

Now for a table, the main thing we need to recognize

is that we will be counting rows that fit a certain criteria of our variable.

Our variable of "Choice to Drive" ranges from "Always" to "Never."

The number of rows that have an "Always" in it

in the column of "Choice to Drive" is 8, while the number

of rows that have an "Almost Always" in the column of "Choice to Drive" is 7.

If we keep at this, we can summarize the

"Choice to Drive" variable in a table.

Because the variable is categorical, we call this a "Table of Counts."

To go from a table to a visualization, we use a bar chart.

In the end we go from this - the data.

To this - the data in summary form.

To this a visual representation of that summary.

And here's the great thing, the summary does

a great job telling the story of the variable and answering the question.

Now this is where looking at data gets really interesting.

"Choice to Drive" is an action question: what does the subject actually DO?

But let's take a look at the variable of "Accident" - the percent of accidents

THOUGHT to occur from driving in inclement weather.

This is a belief question, and it's a whole lot more interesting.

Again, we'll start with a question: "Can we visually

summarize the participant's belief about accidents during inclement weather?"

Well sure - with another table.

Help

Since this new variable has a large span (from 0 to 100),

we'll want to use what's called a Grouped Frequency Table.

Grouped because the variable's range is so large, and we want to chunk

or bin the data into smaller, more manageable ranges.

After all, since we KNOW we want to summarize, does it make sense for us

to make a table for each row or possible percentage value?

Our dataset has only 61 rows in total.

If we were to take all of that information

and put it into a table representing one row for each percentage point,

that would be over a table of 100 rows.

That's not really a summary; that's just a really bad copy.

Most of the time we want to shoot for about 10 bins in our frequency tables

- or 10 rows.

So let's chunk our data by 10.

Just like we did with categorical data, we

want to count the number of rows that fit a certain criteria.

The number of rows that fit the 0 to 10% bin

is 3; the number of rows that fit the 10 to 20% bin is 2;

the number of rows that fit the 20 to 30% bin is 9; and so on.

And now, we can make what's called a histogram.

At the first glance, it looks like a bar graph, and in a way it is.

Both the bar graph and the histogram COUNT the number of rows

(or frequency) that a variable had a certain value afforded.

For the bar graph, that value is a category,

and we show that by the separation of the bars in the graph.

But for a histogram, that value is numerical, and it's usually continuous.

And we show this by removing that separation between the bars.

And there you have it.

Two of the most common, and useful, graphs in univariate statistics:

the bar graph and the histogram.

Both of which are visualizations of our data that come from aggregate summaries

through tables.

Help

Comprehension Check

(3/3 points)

What's the first step in creating a graph of our data?

- ○ Plotting the mean of the data
- ○ Binning our variables into more manageable sizes
- ● Coming up with an aggregate table ✔
- ○ Plotting the spread of our data

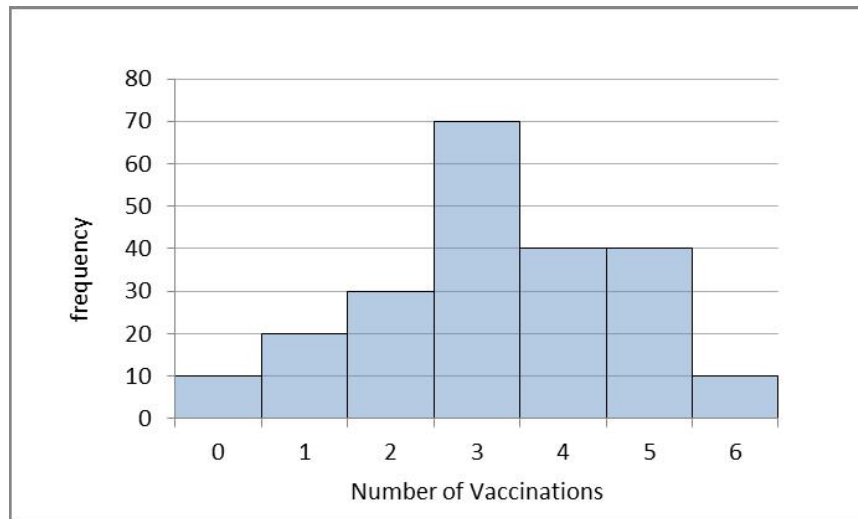The aggregated table when examining a categorical variable is:

- ◉ Table of Counts ✔

- ◯ Frequency Table

- ◯ Grouped Frequency Table

What would be a standard "good" number of rows in a frequency table?

- ◯ 1 to 5

- ◉ 10 to 15 ✔

- ◯ 20 to 30

- ◯ As many rows as data values

| Check | Show Answer |

Below is a histogram showing the number of vaccinations received by a sample of dogs in the local animal shelter.

(6/6 points)

According to the histogram, how many dogs did not receive any vaccinations at all while in the shelter?

10

10

How many dogs were in this sample?

220

220

What percentage of the dogs received three or more vaccinations? *(Report to one decimal, e.g. 45.6.* **DO NOT include percentage symbol "%.")**

72.7

72.7

What is the bin size for the vaccination histogram?

- ● 1 ✔
- ○ 2
- ○ 6
- ○ Varies for each number of vaccinations

The common graph for categorical data to represent counts for each category is a:

- ● Bar Plot ✔
- ○ Dot Plot
- ○ Histogram
- ○ Line Graph

The common graph for numerical data to represent the frequency for each value (or bin) is a:

Help

○ Bar Plot

○ Dot Plot

◉ Histogram ✔

○ Line Graph

Check   Show Answer

edX

EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities, law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

About & Company Info

About

News

Contact

FAQ

edX Blog

Follow Us

Twitter

Facebook

Meetup

LinkedIn

11/17/2014 01:12 PM

Terms of Service and Honor Code

Privacy Policy (Revised 4/16/2014)

Help

Donate to edX
https://courses.edx.org/courses/UTAustinX/UT.7.01x/3T2014/c...

Jobs at edX

Google+

11/17/2014 01:12 PM