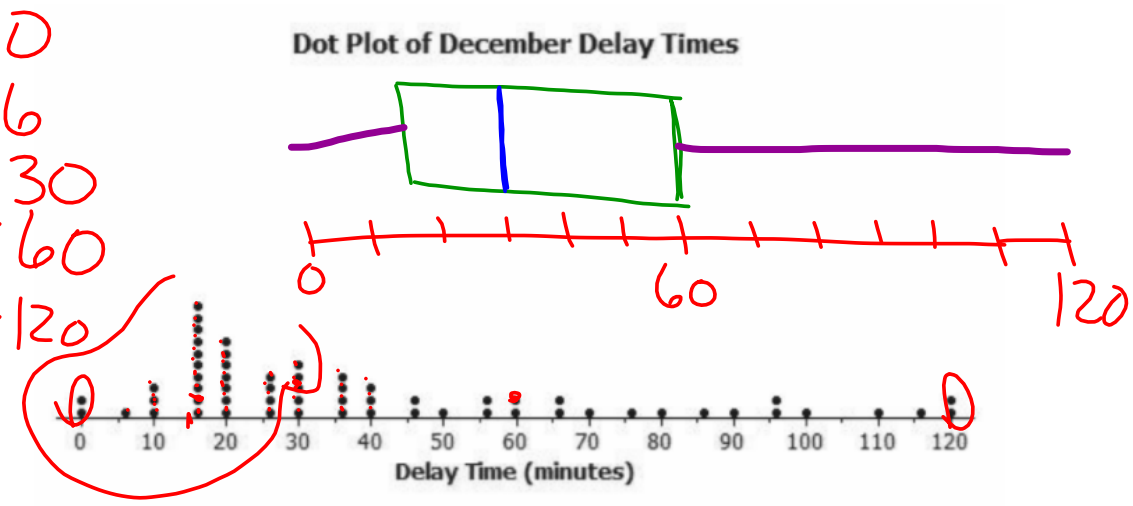
Consider the following scenario. Transportation officials collect data on flight delays (the number of minutes a flight takes off after its scheduled time).

Consider the dot plot of the delay times in minutes for 60 BigAir flights during December 2012:

- Min: 0
- Q1: 16
- Med: 30
- Q3: 60
- Max: 120



1. Find the 5-number summary and calculate the Interquartile Range.

2. Create a box plot based on the data above.

Module 2: Lesson 10

"Two variables"

Summarizing Bivariate Categorical Data with Relative Frequencies

How often something occurs

Lesson Summary

- *Categorical data* are data that take on values that are categories rather than numbers. Examples include male or female for the categorical variable of gender or the five superpower categories for the categorical variable of superpower qualities.
- A *two-way frequency table* is used to summarize bivariate categorical data.
- A *relative frequency* compares a frequency count to the total number of observations. It can be written as a decimal or percent. A two-way table summarizing the relative frequencies of each cell is called a *relative frequency table*.
- The marginal cells in a two-way relative frequency table are called the *marginal relative frequencies*, while the joint cells are called the *joint relative frequencies*.

Types of Data

1. Numerical Data - gives information about quantities

Ex
- Age
- Height
- Weight
- Distance
- Time

2. Categorical Data - tells how many things surveyed fall into a particular category

- food people like
- Sports teams
- Clothing
- Eye color
- favorite car
- Cats or dogs

Two-Way Frequency Table

A two-way frequency table is used to summarize two sets of categorical data. We use this to see how two categories are related to each other, if they are at all.

		Hair Color			
		Black	Blond	Red	Total
Eye Color	Blue	3	4	1	8
	Brown	5	2	0	7
	Green	1	1	3	5
	Total	9	7	4	20

1. How many people have blond hair and blue eyes?

4 people

2. How many people have black hair?

9 people

3. How many people have red hair and brown eyes?

0 people

Relative Frequency - Relates each frequency count to the total number of observations. This is represented by a percent, ratio, or fraction.

		Hair Color			
		Black	Blond	Red	Total
Eye Color	Blue	3	4	1	8
	Brown	5	2	0	7
	Green	1	1	3	5
	Total	9	7	4	20

Data values w/ 2 variables

Data values w/ one variable

Construct a table that shows the **joint relative frequencies** and the **marginal relative frequencies** for the data above.

		Hair Color			
		Black	Blond	Red	Total
Eye Color	Blue	$\frac{3}{20}=.15=15\%$	$\frac{4}{20}=20\%$	$\frac{1}{20}=5\%$	$\frac{8}{20}=40\%$
	Brown	$\frac{5}{20}=25\%$	$\frac{2}{20}=10\%$	$\frac{0}{20}=0\%$	$\frac{7}{20}=35\%$
	Green	$\frac{1}{20}=5\%$	5	15%	25%
	Total	$\frac{9}{20}=45\%$	$\frac{7}{20}=35\%$	$\frac{4}{20}=20\%$	100%

Example

Demographers are trying to understand the association between where a person lives and how they commute to work. They survey 100 people in three cities with the results shown below.

	Car	Train	Walk	Total
New York	5	25	10	40
Los Angeles	18	12	5	35
Chicago	8	14	3	25
Total	31	51	18	100

Fill in the table below with the relative frequencies.

	Car	Train	Walk	Total
New York	5%	25%	10%	40%
Los Angeles	18%	12%	5%	35%
Chicago	8%	14%	3%	25%
Total	31%	51%	18%	100%

Given that a person rides a train to work, what is the conditional relative frequency that they live in New York?

In other words, if someone rides a train to work what is the chance that they also live in New York? So we look at the intersection of the "Train" column and the "New York" row on our table of relative frequencies. We find that the answer is 25%.

If a person lives in Los Angeles, what is the conditional relative frequency that they drive a car?

Again, if someone lives in Los Angeles what is the chance that they also drive a car?

To answer this we look at where "Los Angeles" and "Car" intersect on our relative frequency table. When we do, we see that our answer is 18%.

What is the marginal frequency of people walking to work?

This question asks "What percent of people walk to work, regardless of what city they live in?". So we look at the percentage under the "Walk" column and the "Total" row. Since that value is located in the margins, we know that it's the correct value. Therefore, our answer is 18%.

Is a person more likely to ride a train if they live in New York or if they live in Chicago? Justify your answer.

The chance of someone living in New You and riding a train is 25%.

The chance of someone living in Chicago and riding a train is 14%.

Therefore, if a person rides a train, they are more likely to live in New York than Chicago.