Probability & Statistics Workbook Solutions

Visualizing data

krista king
MATH

ONE-WAY DATA

■ 1. Identify the variables in the following data description and classify the variables as categorical or quantitative. If the variable is quantitative, list the units.

> "The Indianapolis 500 is a car race that's been taking place since 1911 and is often scheduled to take place over Memorial Day weekend. The race takes place at the Indianapolis Motor Speedway and a driver needs to complete 200 laps that cover a distance of 500 miles. Race results are reported by driver number, the driver's name, the type of car the driver uses, and the time to the nearest ten-thousandth of a second. If a driver doesn't finish the race, instead of the time to complete the race, their number of laps completed is recorded."

Solution:

Remember that categorical variables can be represented as numbers. But they just don't measure anything, and you can't use them to perform a calculation. The driver's number is a categorical variable because it's not a measurement, but a way of keeping track of a person. The driver's name and the type of car are also categorical.

The quantitative variables are measurements, like the time it takes a driver to finish the race, or the number of laps completed.

Categorical variables	Quantitative variables
Driver number	Time
Driver name	Number of laps
Type of car	

2. Casey is taking a survey of her senior class. She plans to ask the seniors this question:

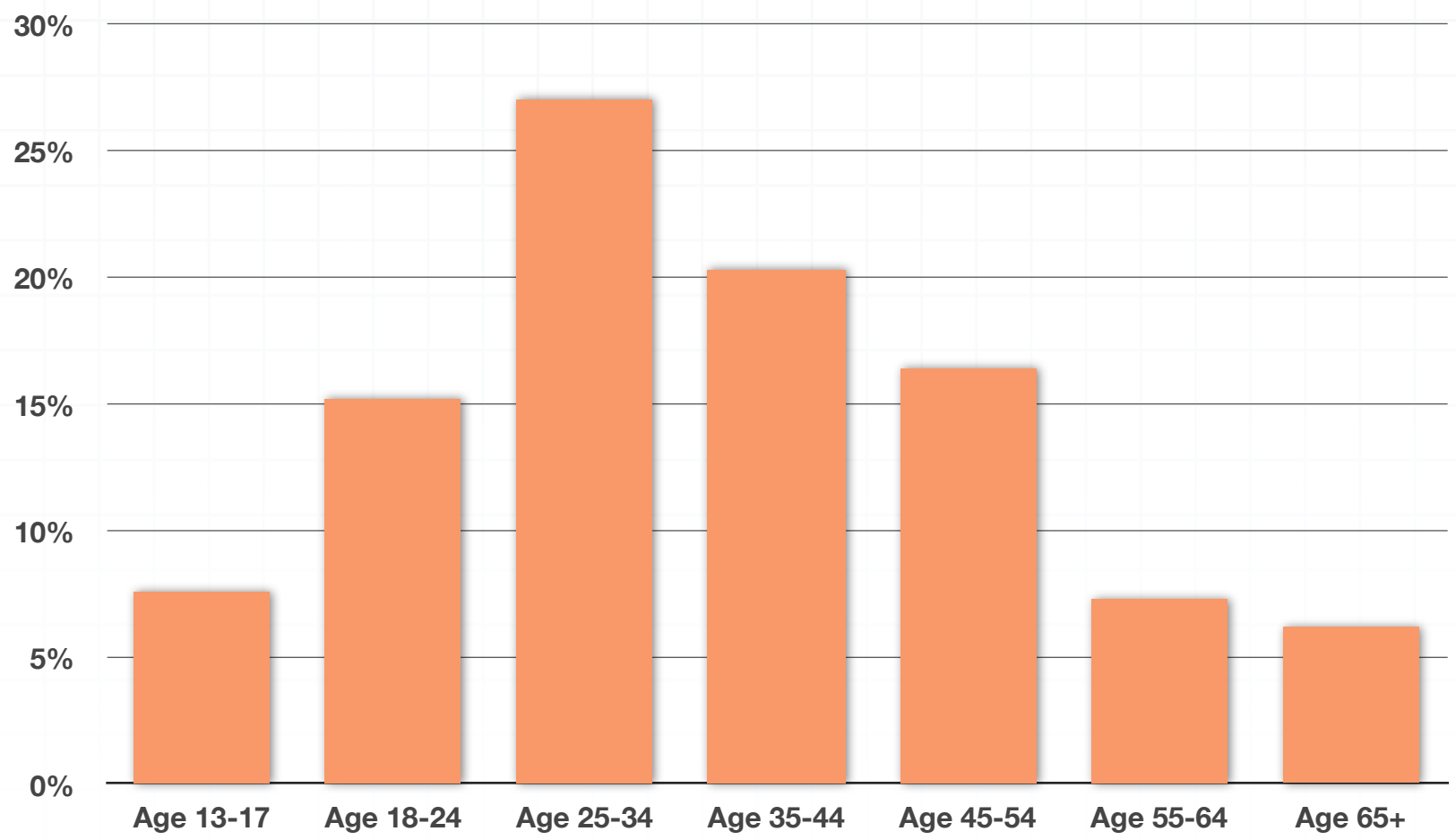
> "In general do you think things have gotten better or worse for our students over the course of the year?"

Her survey has a checklist with these responses: Better, Worse, Stayed the same, and Don't know. Who are the individuals in the survey? What type of response variable is Casey looking for? Is it categorical or quantitative?

Solution:

Casey is surveying the senior class, so those students are the individuals of interest in the survey. Casey's data is categorical because there's not a unit of measurement included in the survey. Instead, she's organizing each respondent's answer into a category of Better, Worse, Stayed the same, or Don't know.

■ 3. The graph below shows the age breakdown of Apple iPad owners in the United States in February, 2011. Who are the individuals in the data? What is the variable? Is it categorical or quantitative?



Source: www.statista.com

Solution:

The individuals in the data are the respondents, by age group. The percentage of each age group is a quantitative variable because it's a measurement.

■ 4. The table below shows the number of rejected products by worker and shift. Is the data below one-way data? Why or why not?

Worker ID	1st shift	2nd shift	3rd shift
1123	42	45	42
2256	45	74	32
6435	36	78	41

Solution:

This data is not an example of one-way data. In order to know the number of rejected products, you'd need to know two things: the individual worker ID, and the shift.

This means the data is now dependent on two independent things, not just one. In order to get to an answer to the question: "How many rejected products?," you'd need to ask something like "How many rejected products were there for worker 1123 during the first shift?" Since you need more than one reference point, this is not one-way data.

■ 5. Why is this table an example of one-way data?

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

Solution:

Even though this table has three different variables, if I'm given one individual and a category, I can answer a question about the data. For example I could ask: "How many scoops of vanilla were sold in July?" and I would know right away that the answer was 300.

If the data isn't one-way data, I'd need to answer a question about both categories. For example, to answer a question like "How many scoops were sold?" you might need to ask something like "What flavor and which store?".

■ 6. A botany student wants to test the claim of a diaper company that their product may be used in a compost pile. He creates 12 identical gardens and plants a random selection of 7 tomato plants in each one. He plans to have a fellow student use traditional compost on 6 of the garden plots and the compost from the diapers on the other 6. He does this so he doesn't know which plot is which. He plans to check the tomato plants for

disease every two days for a month, and record the number of tomato plants with disease after each check. Would this experiment result in one-way data? Why or why not?

Solution:

The experiment does not result in one-way data.

We can think of how the botany student would need to record his data to see whether or not this is an example of one-way data. He could create a table with the checks on the plants and the number of plots to record the number of tomato plants with disease.

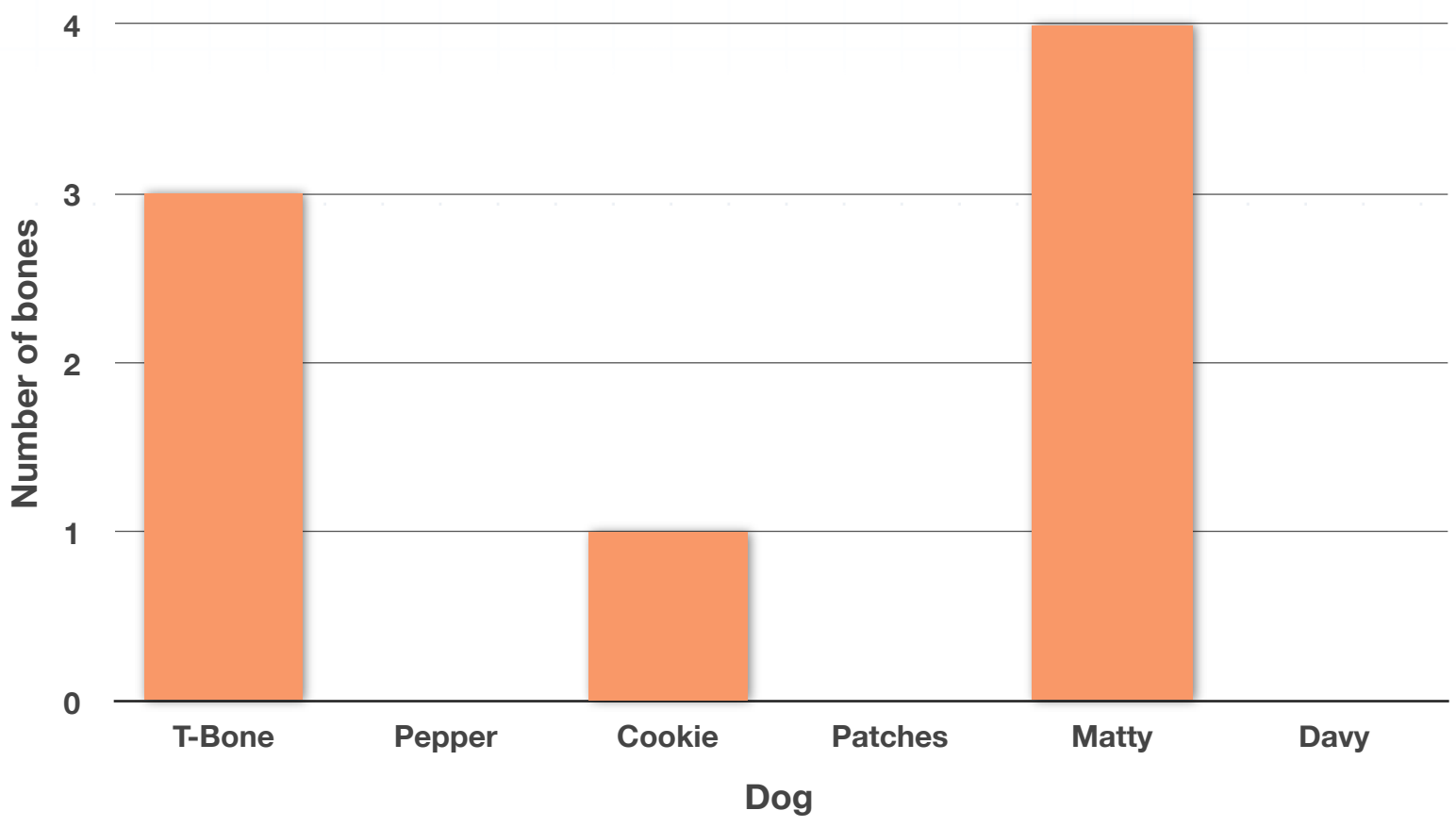
	Check 1	Check 2	Check 3	Check 4	...	Check 15
Plot 1						
Plot 2						
Plot 3						
Plot 4						
...						
Plot 12						

From the table, you can see that to answer a question like “How many tomato plants ended up with a disease?” you would need to know two things: which plot, and which check. This means the data is not an example of one-way data.

BAR GRAPHS AND PIE CHARTS

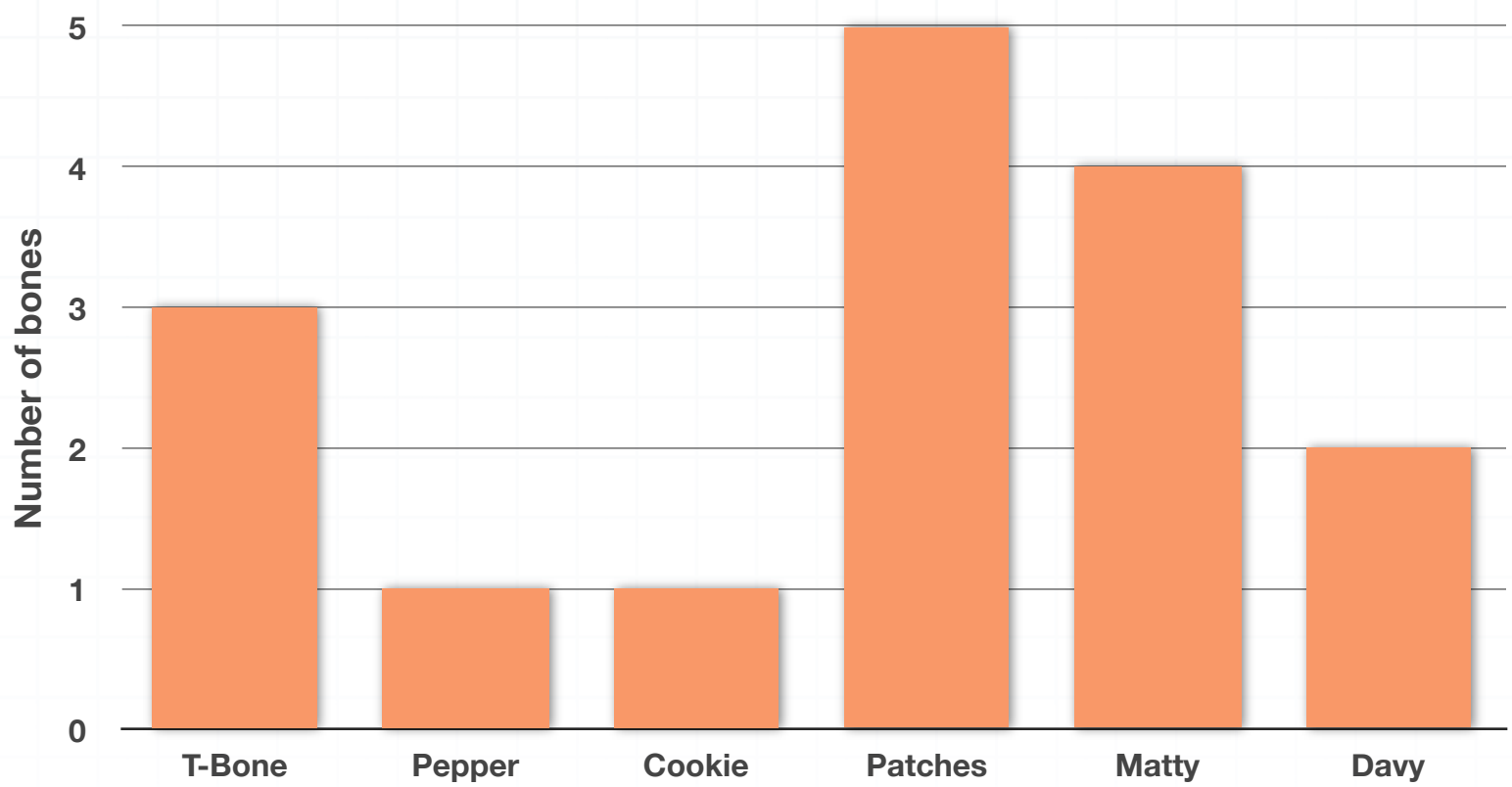
1. Both the bar graph and the table have missing information about the number of bones each dog consumed at doggie daycare. Use the graph and table together to fill in the missing pieces.

Dog	Number of bones
T-Bone	
Pepper	1
Cookie	
Patches	5
Matty	
Davy	2



Solution:

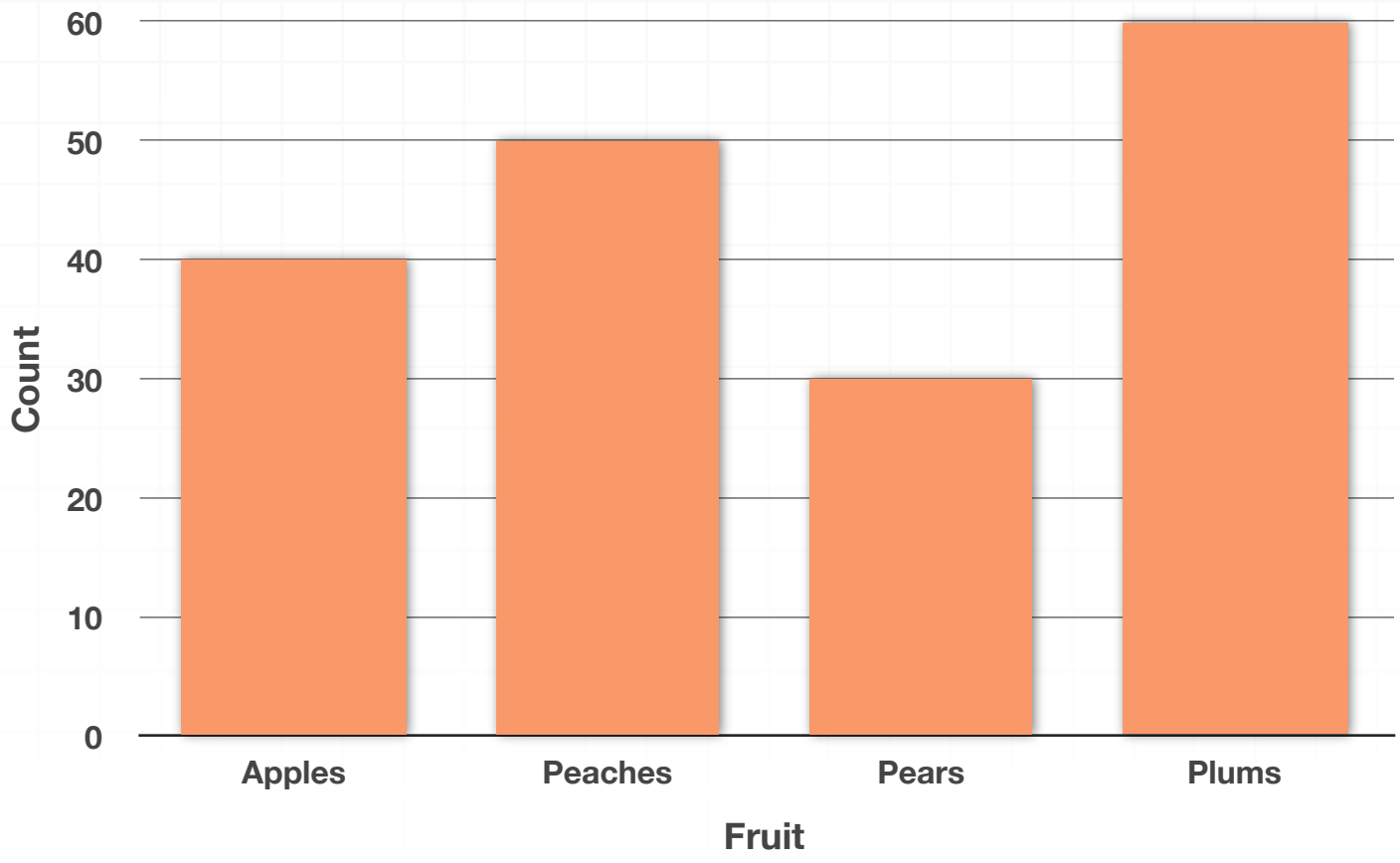
You can read from the table that Pepper ate 1 bone, Patches ate 5 bones, and Davy ate 2 bones. Therefore, the completed bar graph is



You can read from the bar graph that T-Bone ate 3 bones, Cookie ate 1 bone, and Matty ate 4 bones. Therefore, the completed table is

Dog	Number of bones
T-Bone	3
Pepper	1
Cookie	1
Patches	5
Matty	4
Davy	2

2. Eric's class went on a trip to an orchard. At the end of the trip they counted how many pieces of fruit came from each type of tree and graphed it in the bar graph shown below. Use the bar graph to create a pie chart of the data.



Solution:

To create a pie chart you can divide the circle into fractional parts. We can see the students picked 40 apples, 50 peaches, 30 pears and 60 plums. That makes the total amount of fruit the class picked

$$40 + 50 + 30 + 60 = 180$$

The nice thing about this data is that it's all divisible by 10. We can therefore divide the pie chart into $180 \div 10 = 18$ equal pieces, and then shade in the appropriate number of pieces for each fruit.

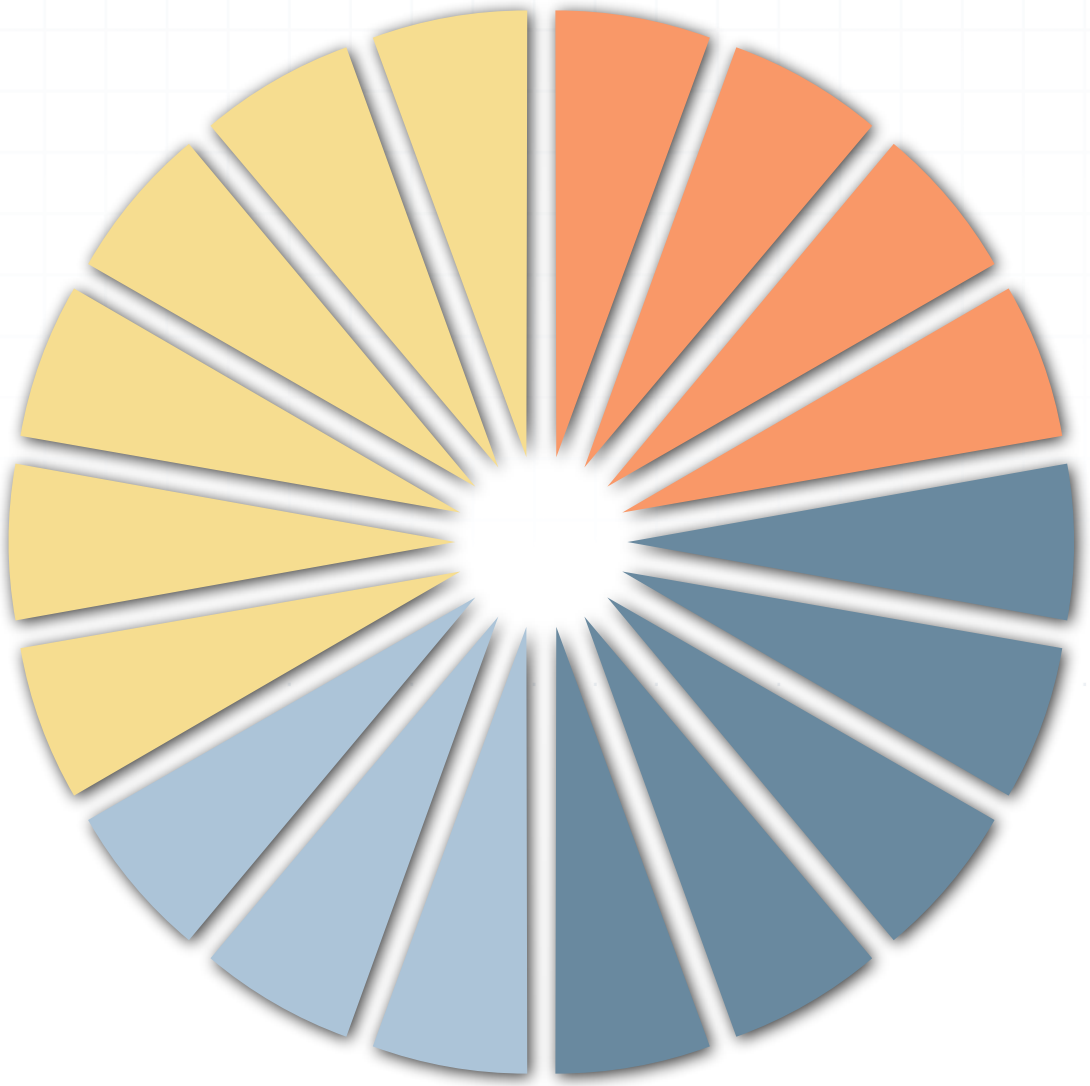
For apples: $40 \div 10 = 4$

For peaches: $50 \div 10 = 5$

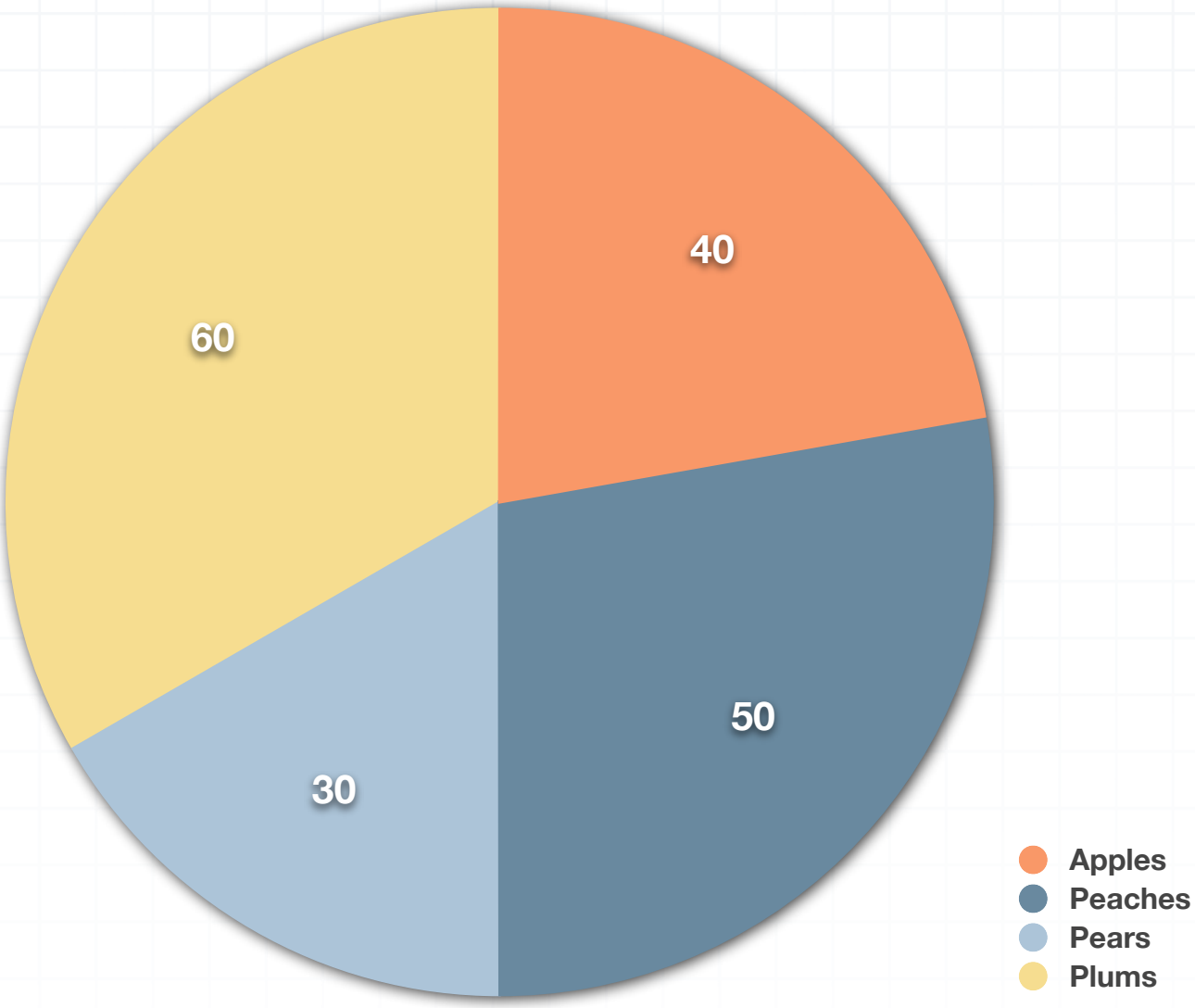
For pears: $30 \div 10 = 3$

For plums: $60 \div 10 = 6$

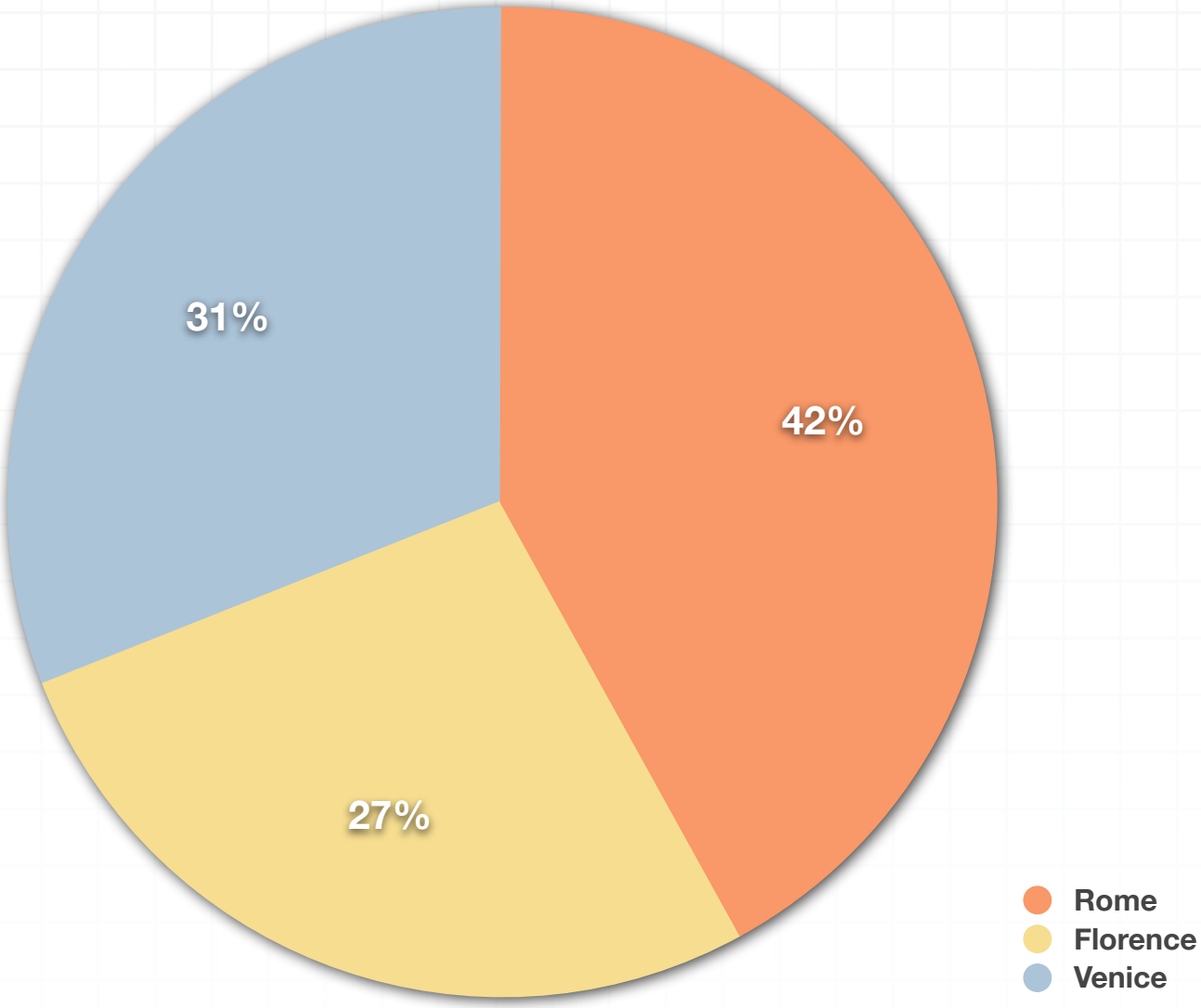
So if we use red for apples, dark blue for peaches, light blue for pears, and yellow for plums, we would shade 18 equal slices this way:



Then the finished pie chart is



3. A tourist company took a survey of 600 clients and asked them which Italian city they were most interested in visiting. How many clients said they wanted to visit Rome?

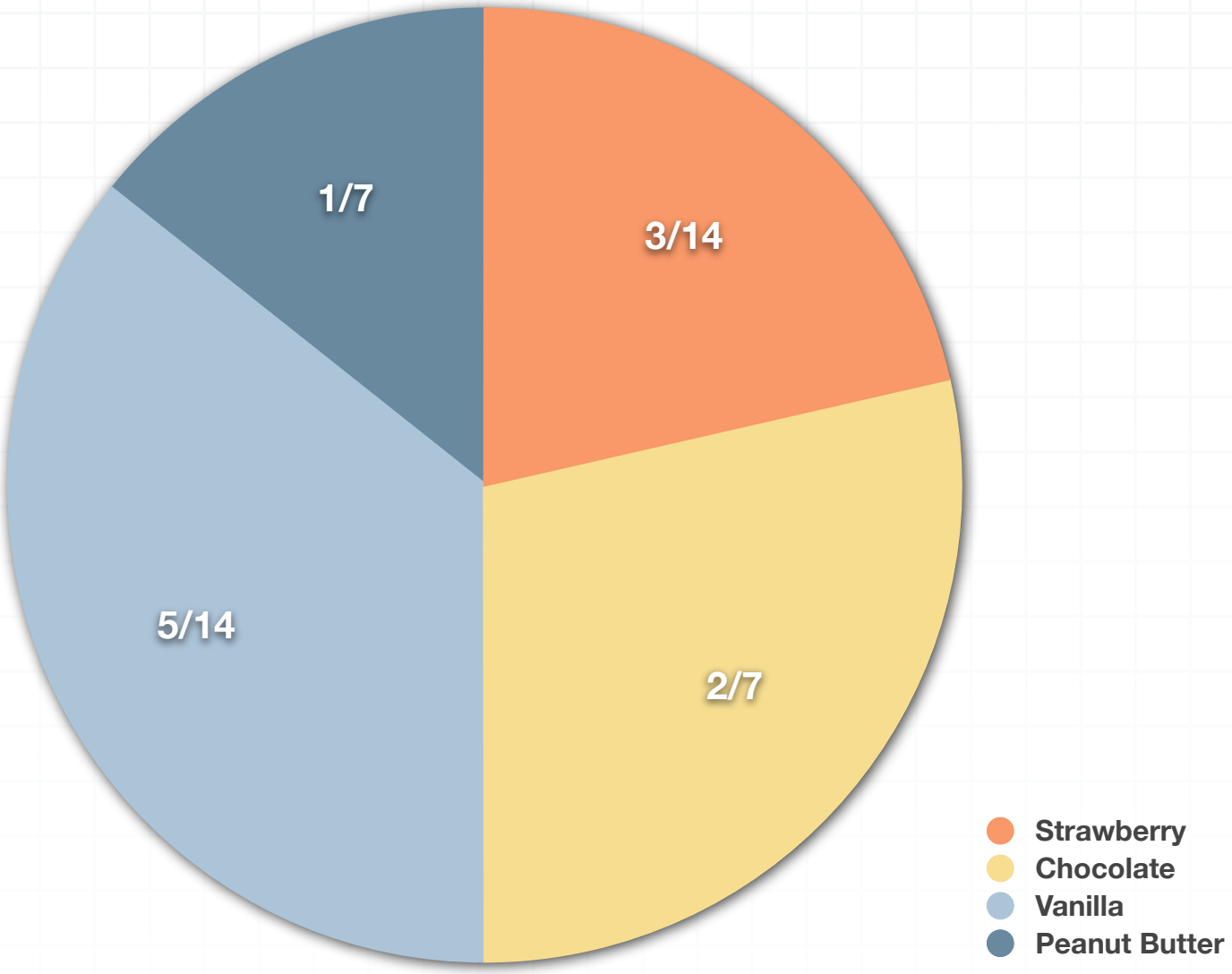


Solution:

Out of the 600 clients surveyed, $42\,\%$ of them said they wanted to visit Rome. $42\,\%$ of 600 is

$$600 \cdot 0.42 = 252 \text{ clients}$$

■ 4. The pie chart shows how many ice cream cones of each flavor were sold. Assuming 280 total ice cream cones were sold in August, convert the pie chart to a bar graph.



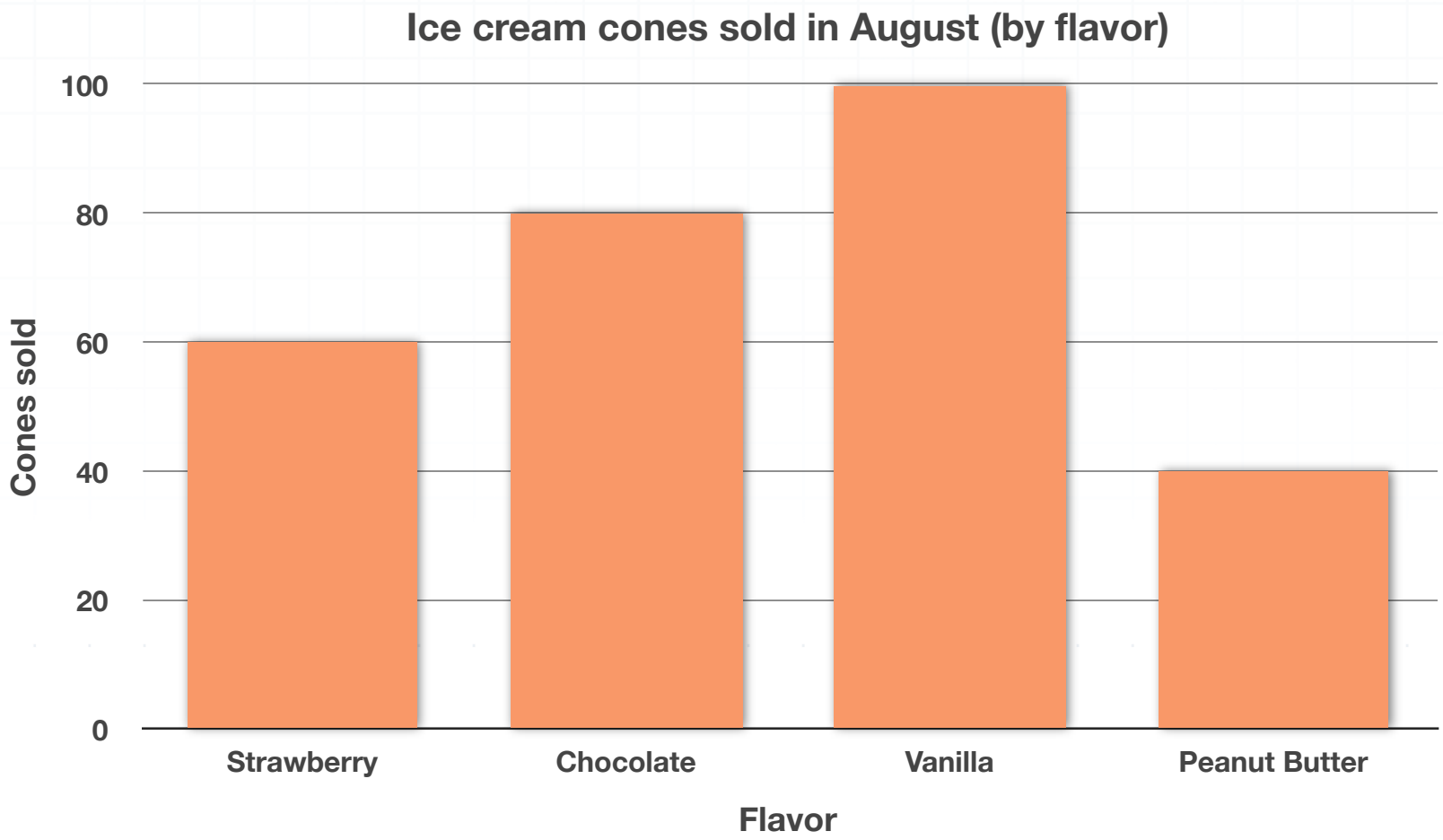
Solution:

We know that 280 ice cream cones were sold, and we have the amount of each flavor sold, as a fraction. For example, we know that $3/14$ of the scoops of ice cream sold in August were strawberry.

Let's convert the information into a table first and also find the number of scoops sold of each type.

Flavor	Fraction	Cones sold
Strawberry	3/14	(3/14)(280)=60
Chocolate	2/7	(2/7)(280)=80
Vanilla	5/14	(5/14)(280)=100
Peanut Butter	1/7	(1/7)(280)=40

Now we can create a bar chart with the flavors on the horizontal axis and the number of cones sold on the vertical axis.

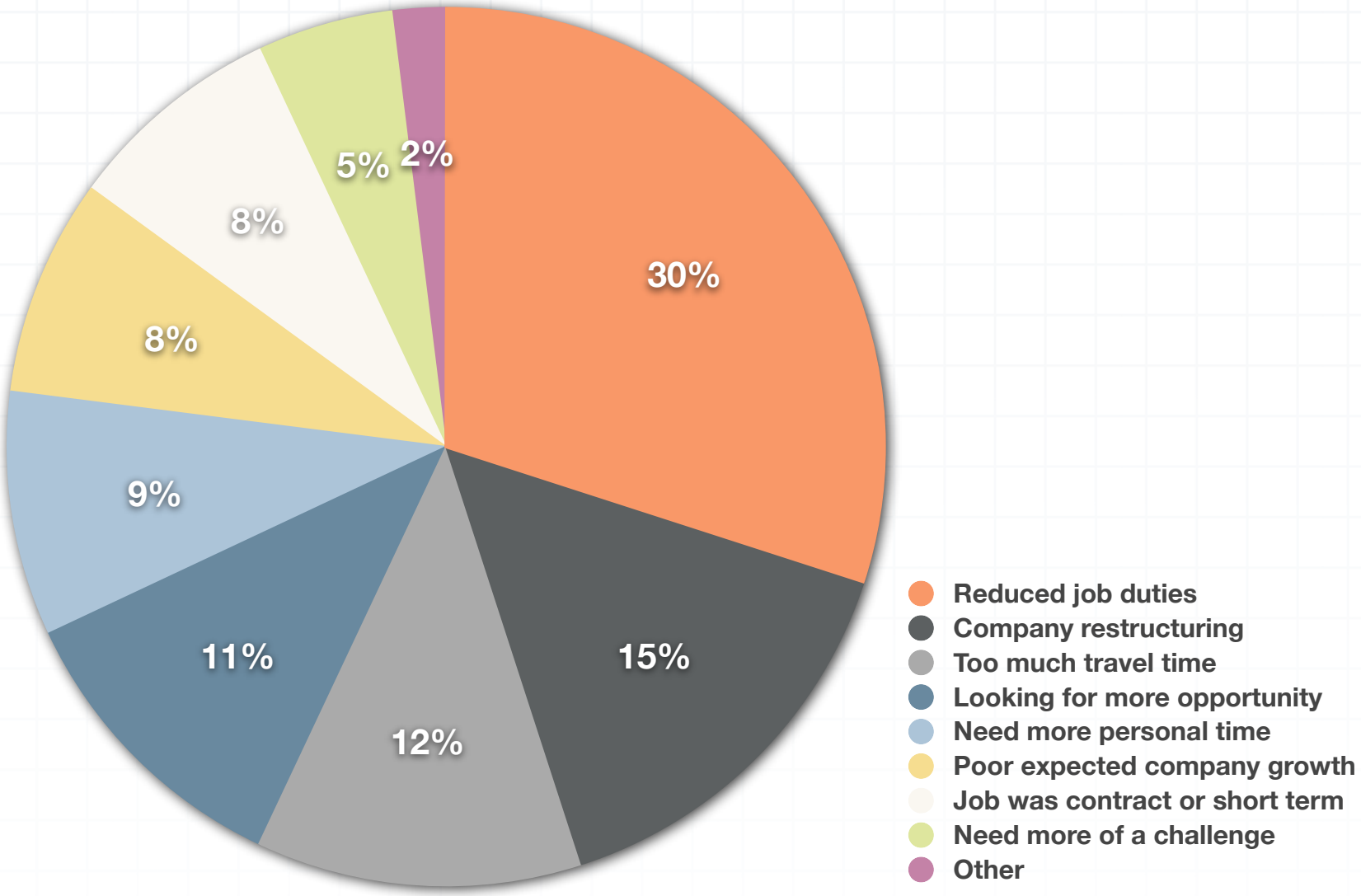


■ 5. A company is analyzing the results from a recent survey about why people left their employment. The results are shown in the data table below. In general, is a bar graph or a pie chart a better choice to display the data? Why?

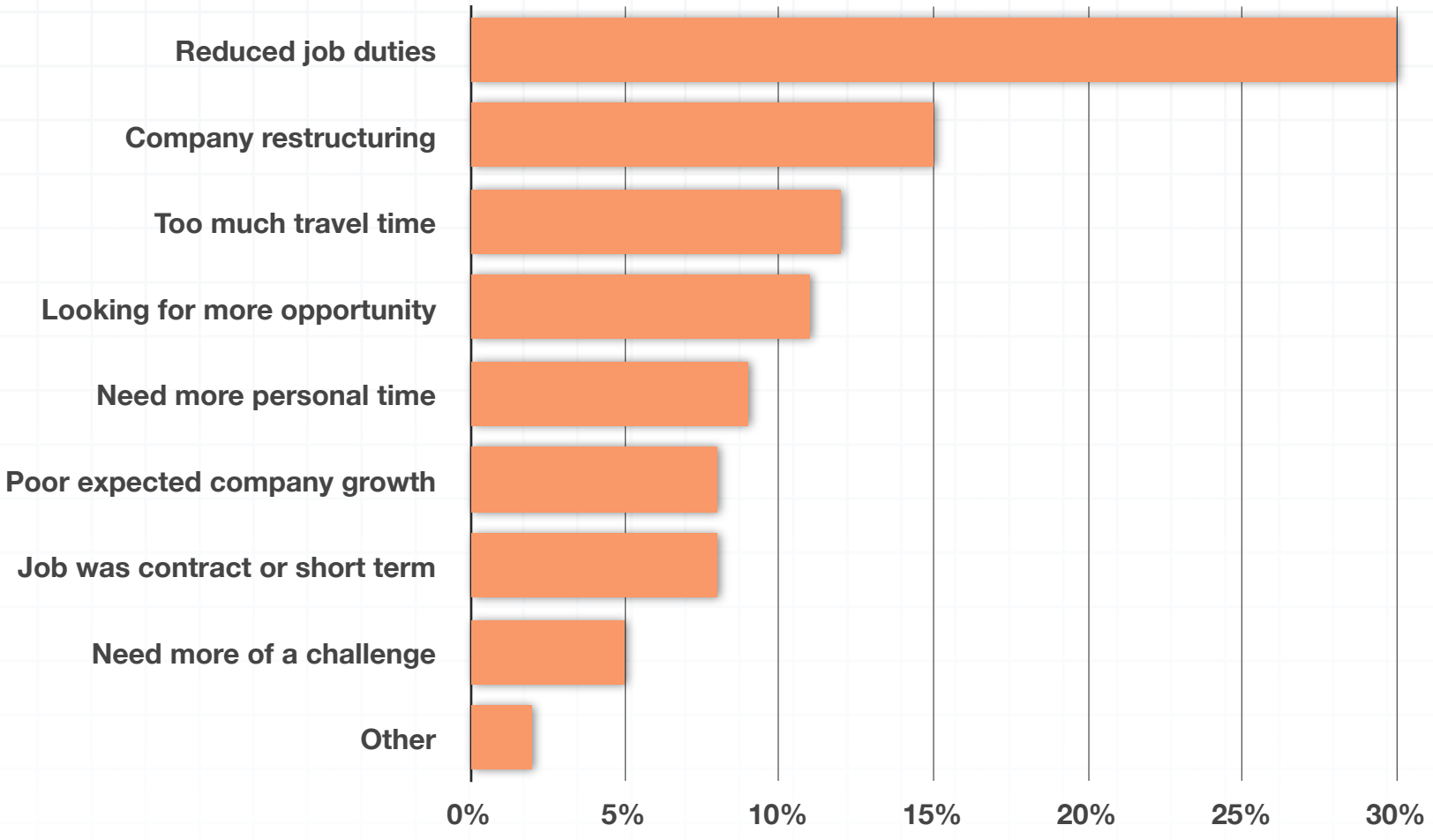
Reasons for leaving job	
Reduced job duties	30%
Company restructuring	15%
Too much travel time	12%
Looking for more opportunity	11%
Need more personal time	9%
Poor expected company growth	8%
Job was contract or short term	8%
Need more of a challenge	5%
Other	2%

Solution:

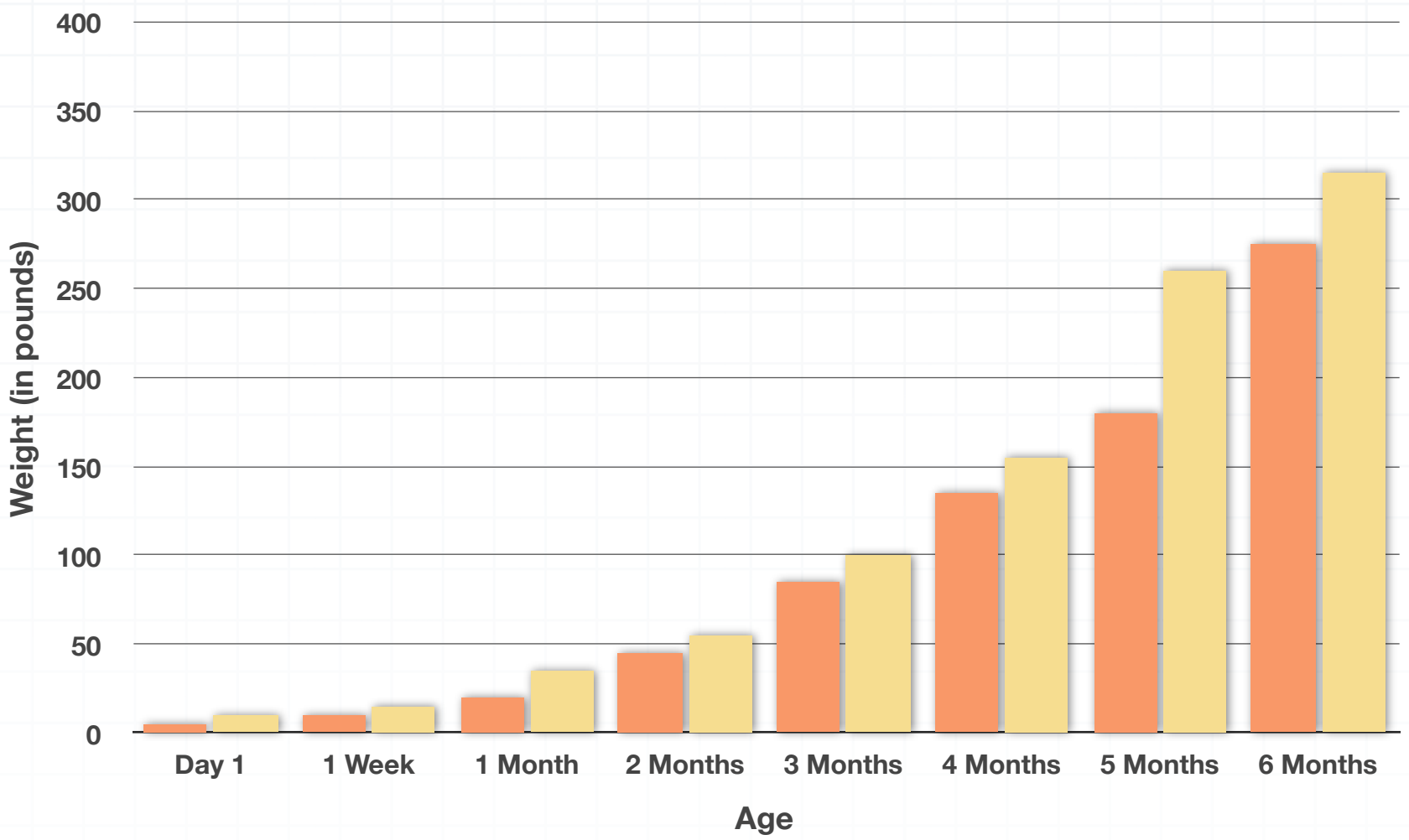
A bar graph is a better choice to display the data because there are so many different categories. A pie chart can get cluttered when there are a lot of categories,



but a bar graph will remain fairly easy to read. Also notice that this is a good time to use a horizontal bar graph because the category titles are lengthy.



6. The comparison bar graph shows the growth of two pigs over their first 6 months of life. Which pig grew the most between 4 and 5 months?



Solution:

The largest pig in the litter grew from approximately 155 pounds to approximately 260 pounds, a change of about $260 - 155 = 105$ pounds. The runt of the litter grew from approximately 135 pounds to approximately 175 pounds, a change of about $175 - 135 = 40$ pounds. The largest pig in the litter grew much more than the runt.

LINE GRAPHS AND OGIVES

■ 1. Bethany started a sit-up program so that she can do 200 sit-ups in a day. At the end of week 6 she'll have completed $1{,}685$ sit-ups. Create an ogive of the data.

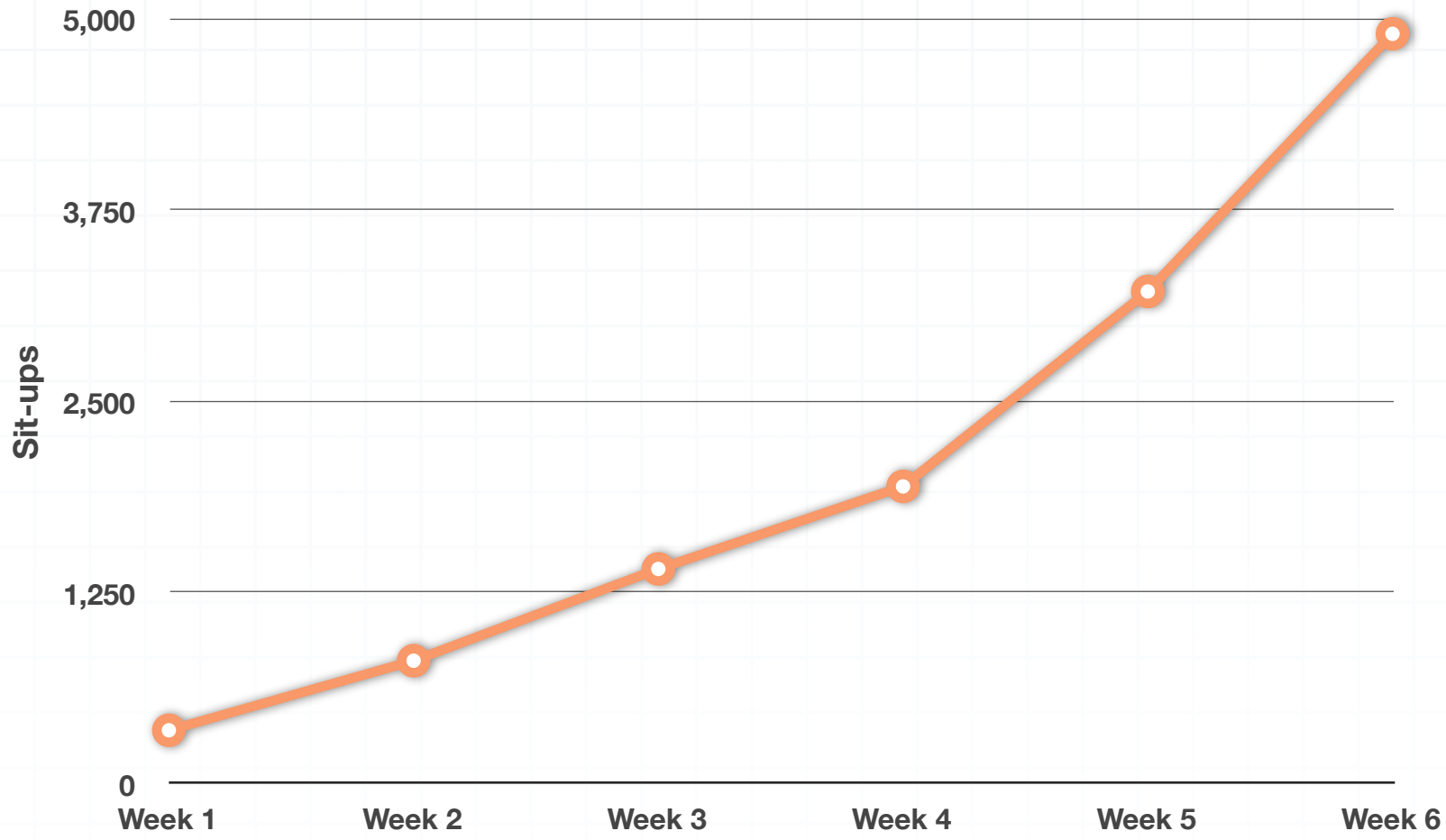
Week	Number of sit-ups
Week 1	350
Week 2	455
Week 3	600
Week 4	540
Week 5	1,275
Week 6	1,685

Solution:

We'll first add a total column to the table.

Week	Number of sit-ups	Total
Week 1	350	350
Week 2	455	805
Week 3	600	1,405
Week 4	540	1,945
Week 5	1,275	3,220
Week 6	1,685	4,905

Now, from the total column, we can create the ogive.

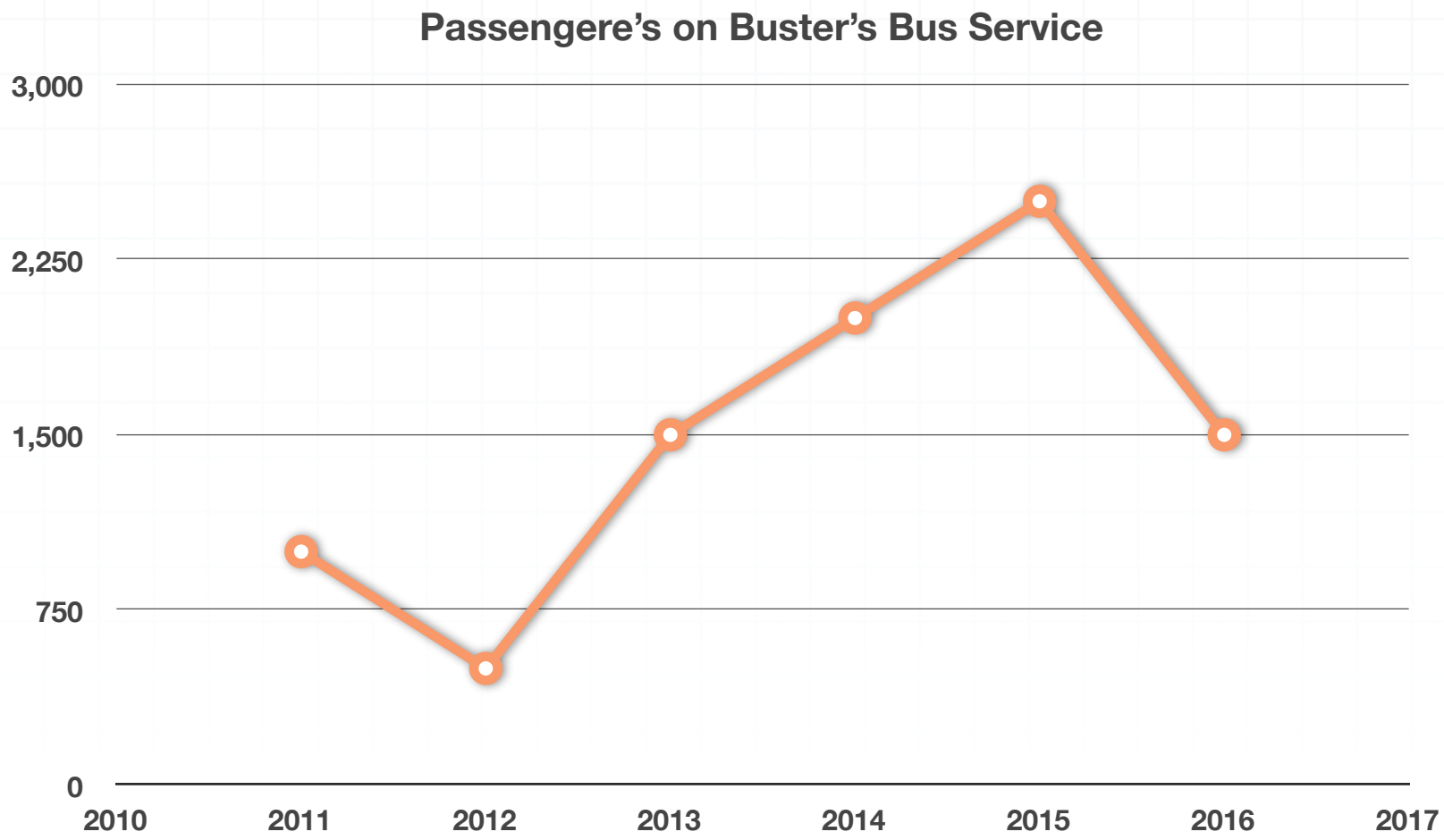


2. The table shows passengers by year for Buster's Bus Service. Create a line graph of the data in the table.

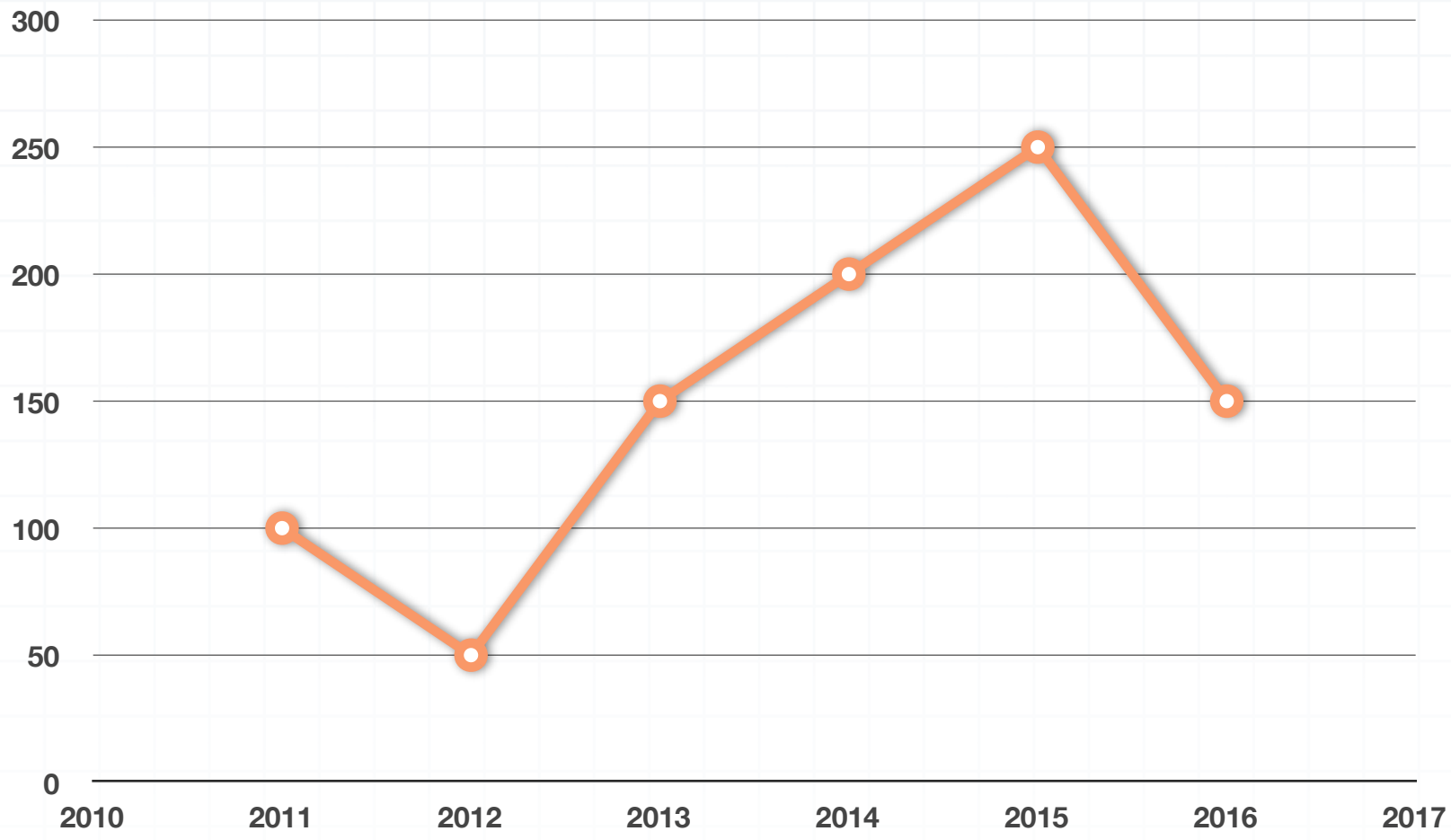
Year	Passengers
2011	1,000
2012	500
2013	1,500
2014	2,000
2015	2,500
2016	1,500

Solution:

Start the line graph at the year 2010 and go by year along the horizontal axis to 2017. Make sure you choose units on the vertical axis that make it easy to graph, as well as read.



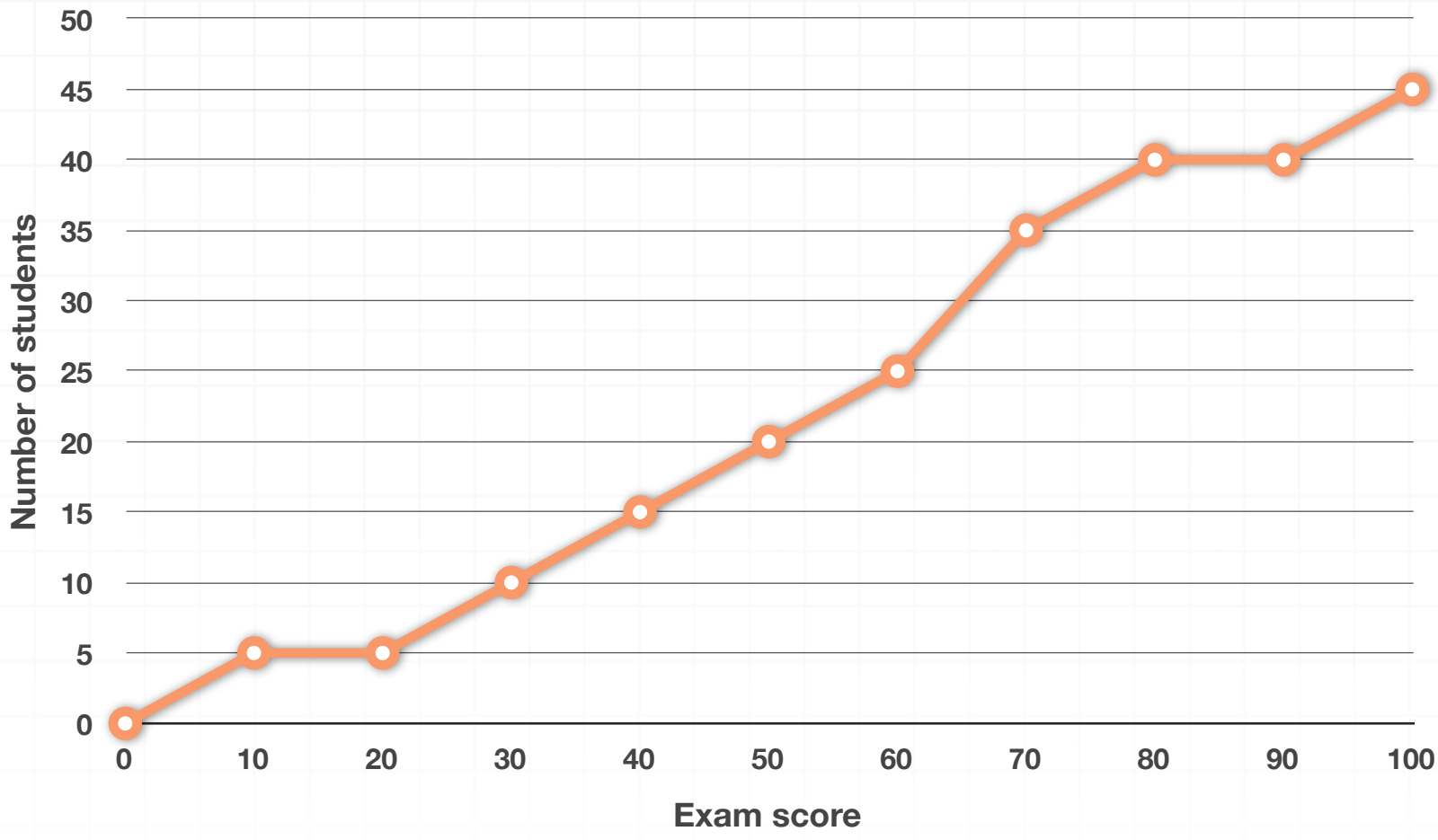
3. Between what two consecutive years was there the largest increase in car sales?



Solution:

The greatest increase in card sales was between 2012 and 2013. If you look at the line graph, you can see that the line increases at the sharpest rate between 2012 and 2013, these are the years car sales increased the most.

■ 4. Mrs. Moore gave her students a midterm exam, then she created this ogive of the 45 exam scores. How many students got a score between $70\,\%$ and $90\,\%$?



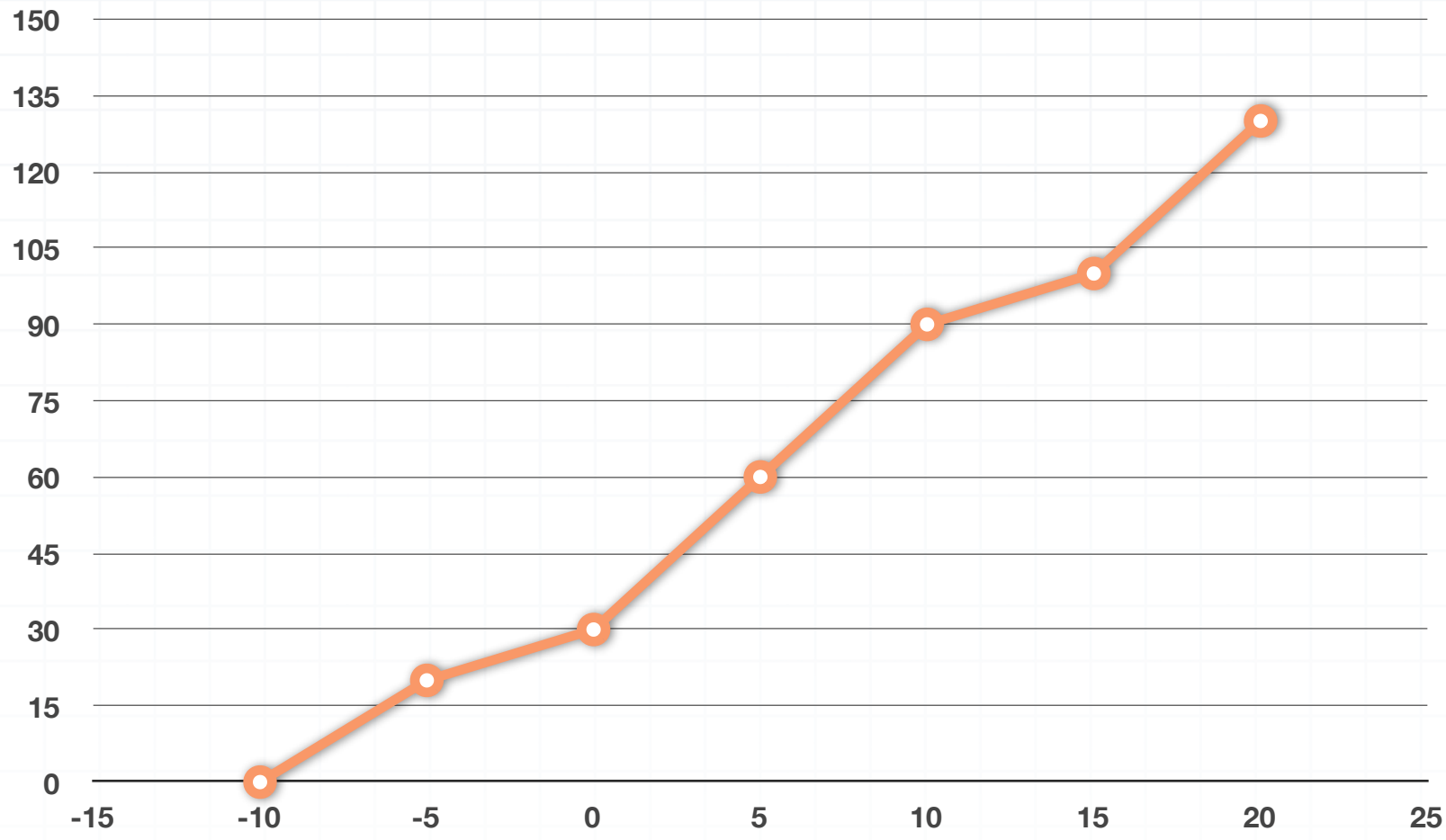
Solution:

We can tell from the ogive that 35 students scored lower than $70\,\%$, and that 40 students scored lower than $90\,\%$. Which means

$$40 - 35 = 5 \text{ students}$$

must have scored between $70\,\%$ and $90\,\%$.

■ 5. Draw the line graph that corresponds to the ogive below.



Solution:

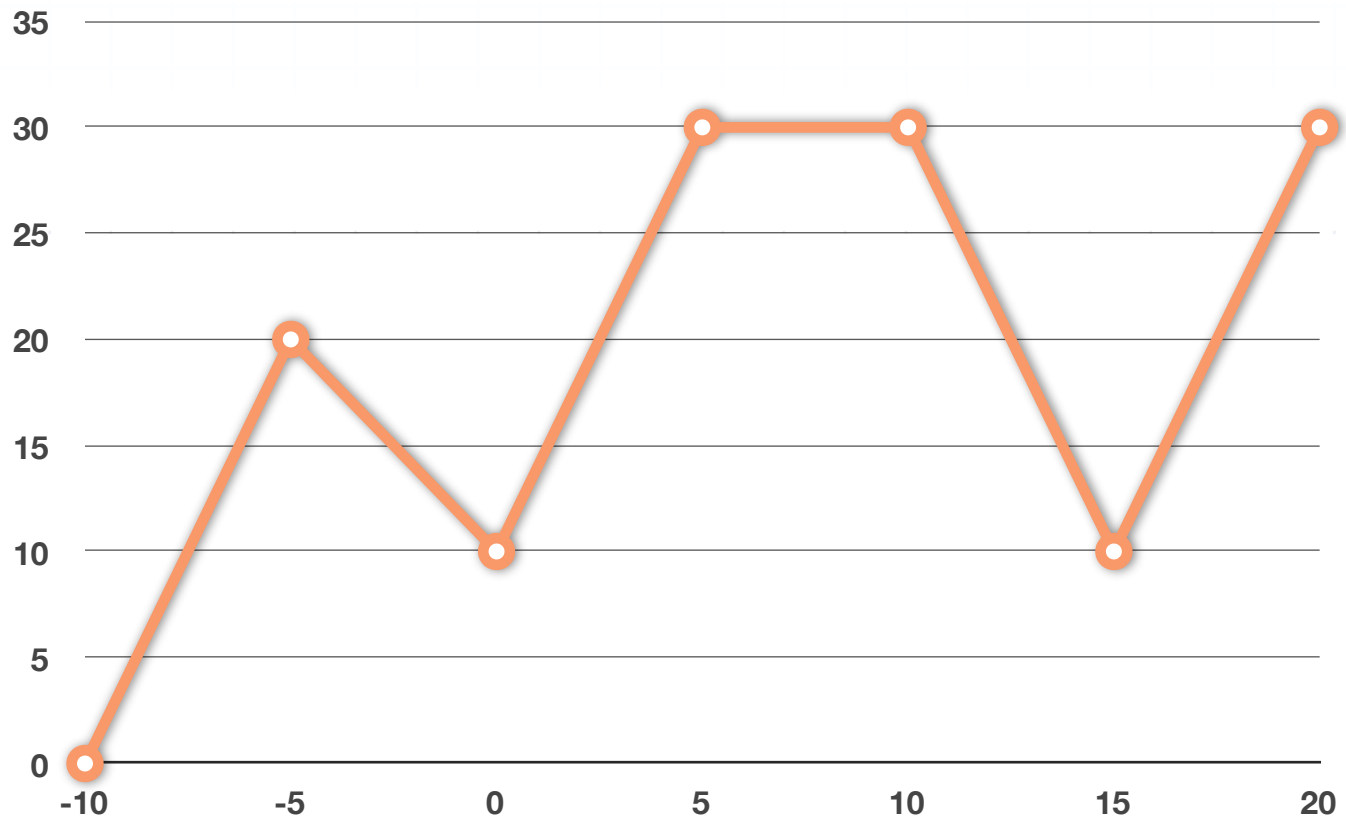
First we can create a table of the information from the ogive. The statistical name for the "total" is the "cumulative frequency."

Horizontal value	Cumulative frequency
-10	0
-5	20
0	30
5	60
10	90
15	100
20	130

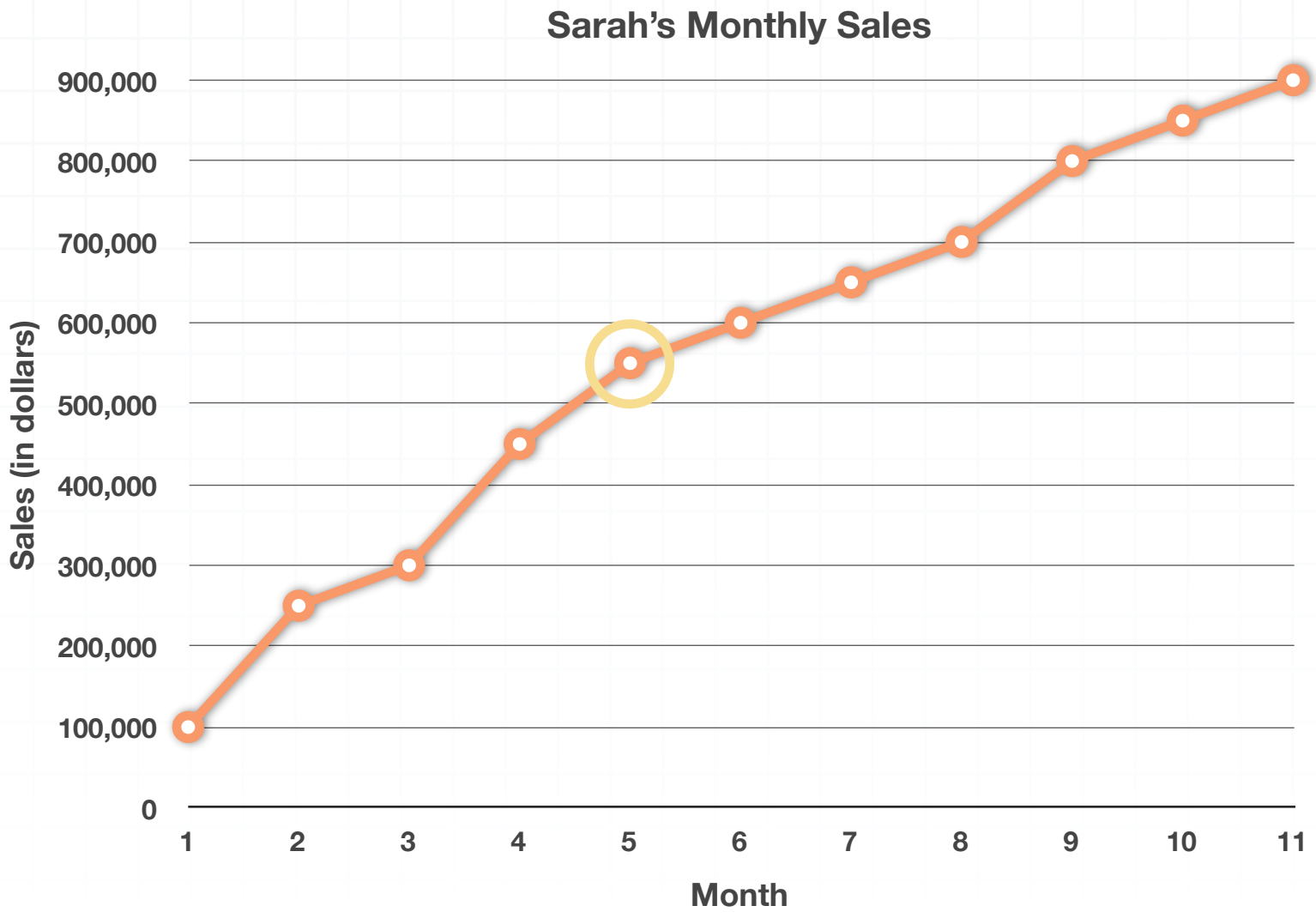
You can use the difference between the cumulative frequencies to find the frequency of each value and create a line graph.

Horizontal value	Cumulative frequency	Frequency
-10	0	0
-5	20	20-0=20
0	30	30-20=10
5	60	60-30=30
10	90	90-60=30
15	100	100-90=10
20	130	130-100=30

Now we'll create a line graph from the "frequency" column.



6. Sarah's monthly sales to date are shown in the ogive. What is the meaning of the circled point?



Solution:

By the time Sarah completed her fifth month at the company, she'd sold about $\$550{,}000$ worth of cars. This amounts to a little more than $60\,\%$ of her total sales since she's worked at the company, because $550/900 \approx 0.61$.

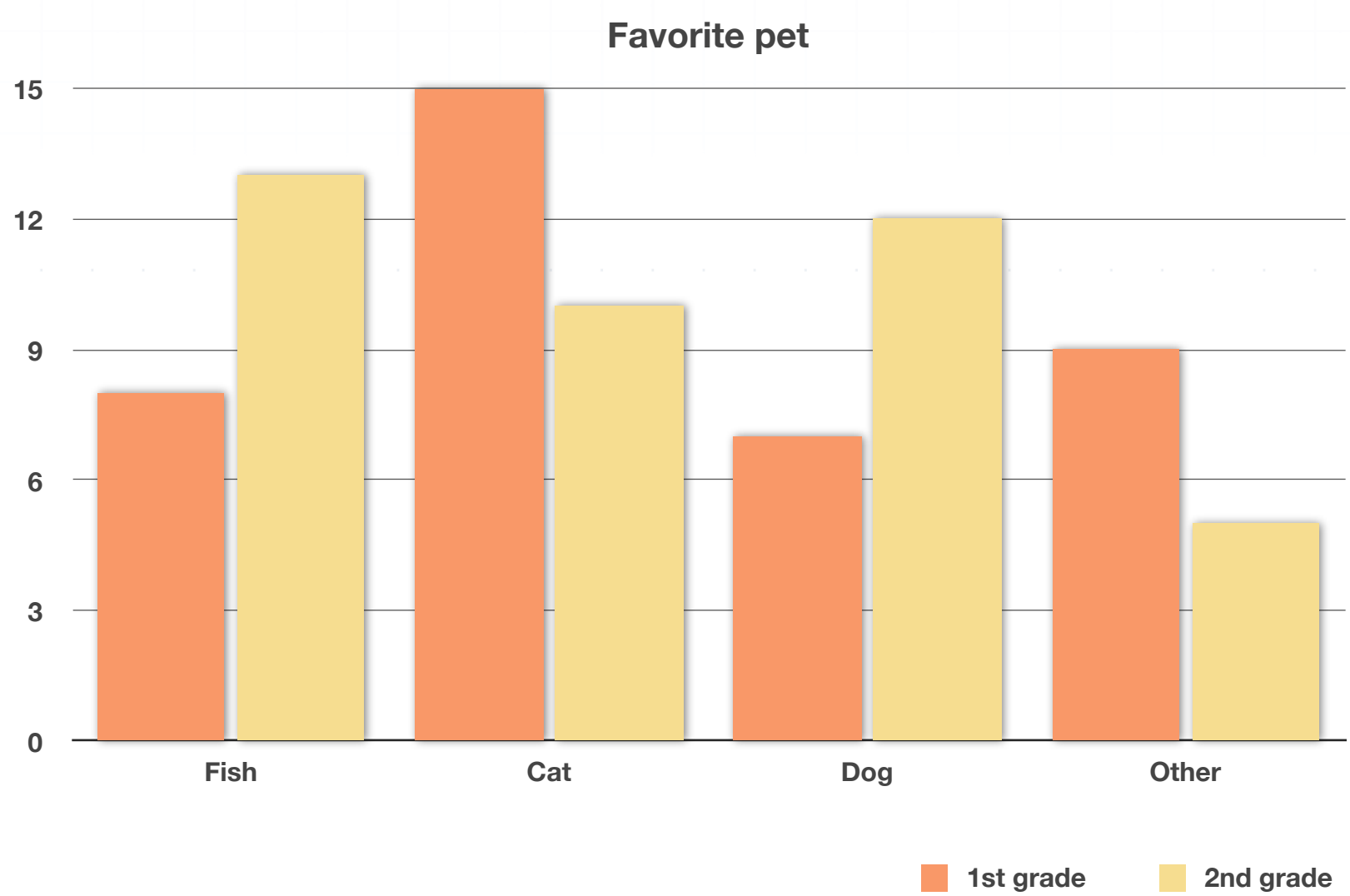
TWO-WAY DATA

1. Create a comparison bar graph for the two-way table.

Favorite pet	Fish	Cat	Dog	Other
1st grade	8	15	7	9
2nd grade	13	10	12	5

Solution:

We'll include a title, a key, start the vertical axis at 0, and plot the data.



2. A pizza parlor wants to know if the age range of their customers affects pizza preferences. The pizza parlor asks each customer two questions:

1. Which type of pizza is your favorite: pepperoni, cheese, supreme or veggie?
2. What is your age range: Under 18, or 18 and over?

The results of the survey are as follows:

Of the 50 customers who prefer pepperoni pizza, 25 are under 18.

Of the 20 customers who prefer cheese pizza, 18 are under 18.

Of the 30 customers who prefer supreme pizza, 24 are over 18.

Of the 25 customers who prefer veggie pizza, 19 are over 18.

What type of data is the pizza parlor collecting, one-way or two-way? Create the best type of frequency table for the data.

Solution:

This is an example of data that can be organized into a two-way table because the data has two types of categories that can be organized together: age range and favorite pizza.

Since we're given the totals in each response, it can be easiest to start our table by filling in the totals of the people who like each type of pizza.

Favorite pizza	Pepperoni	Cheese	Supreme	Veggie	Total
Under 18					
18 and over					
Total	**50**	**20**	**30**	**25**	

Now we can use the rest of the information from each statement.

Favorite pizza	Pepperoni	Cheese	Supreme	Veggie	Total
Under 18	25	18			
18 and over			24	19	
Total	**50**	**20**	**30**	**25**	**125**

Now subtract each of these values from the totals to find the missing information.

Favorite pizza	Pepperoni	Cheese	Supreme	Veggie	Total
Under 18	25	18	30-24=6	25-19=6	
18 and over	50-25=25	20-18=2	24	19	
Total	**50**	**20**	**30**	**25**	**125**

Now find the totals to complete the table.

Favorite pizza	Pepperoni	Cheese	Supreme	Veggie	Total
Under 18	25	18	6	6	**25+18+6+6=55**
18 and over	25	2	24	19	**25+2+24+19=70**
Total	**50**	**20**	**30**	**25**	**55+70=125**

Therefore, the finished table is

Favorite pizza	Pepperoni	Cheese	Supreme	Veggie	Total
Under 18	25	18	6	6	**55**
18 and over	25	2	24	19	**70**
Total	**50**	**20**	**30**	**25**	**125**

■ 3. An elementary school creates the following two-way table. What is the best name for the row variable and what is the best name for the column variable?

	Walk	School bus	Day care vehicle	Carpool
Pre-school	1	10	20	26
First	5	12	14	19
Second	10	22	5	15
Third	8	33	3	10

Solution:

The idea of a row or column variable is that it's how you could explain each section of the two-way table. The row variable describes the data in each row, and the column variable explains the data in each column. "Method of transportation" is one possible description for the column variable. "Student grade" is a possible description for the row variable.

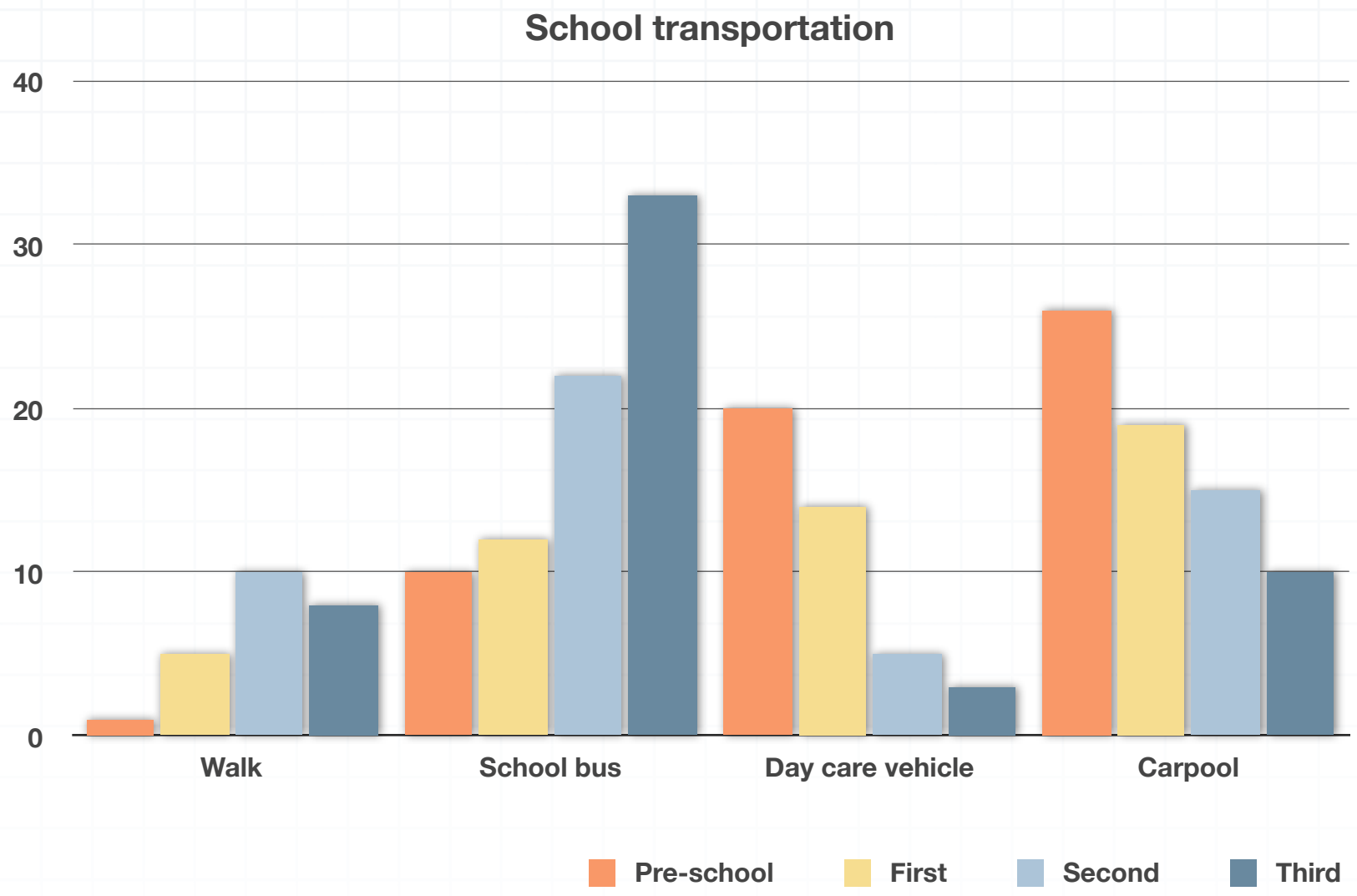
		Method of transportation			
		Walk	School bus	Day care vehicle	Carpool
Grade in school	Pre-school	1	10	20	26
	First	5	12	14	19
	Second	10	22	5	15
	Third	8	33	3	10

■ 4. Which graph would be a better choice to display the data from the two-way table: a comparison bar graph or a comparison line graph? Create your chosen graph.

		Method of transportation			
		Walk	School bus	Day care vehicle	Carpool
Grade in school	Pre-school	1	10	20	26
	First	5	12	14	19
	Second	10	22	5	15
	Third	8	33	3	10

Solution:

A comparison bar graph is the best choice for the data because a comparison line graph is used to show changes over time. Here we're comparing grades in school, so the comparison bar graph is the best choice. Remember when you create a comparison bar graph you need to include the title, key and a reasonable scale on the vertical axis.



■ 5. Eric creates a survey asking students who ate a snack in the morning between classes if they felt sleepy or not. Here are his survey results:

Snack	Yes	Yes	No	No	No	No	Yes	No	Yes	No	Yes	Yes	No	Yes	No
Sleepy	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	Yes	Yes	No	No	Yes

Create a two-way data table for Eric's survey.

Solution:

We could set up the table this way:

		Do you feel sleepy?		
		Yes	No	Total
Did you eat a snack?	Yes			
	No			
	Total			

There are 5 people who ate a snack and feel sleepy. There are 2 people who ate a snack but don't feel sleepy. There are 3 people who didn't eat a snack and feel sleepy. And there are 5 people who didn't eat a snack and don't feel sleepy.

		Do you feel sleepy?		
		Yes	No	Total
Did you eat a snack?	Yes	5	2	
	No	3	5	
	Total			

Now we just total everything up.

		Do you feel sleepy?		
		Yes	No	Total
Did you eat a snack?	Yes	5	2	**7**
	No	3	5	**8**
	Total	**8**	**7**	**15**

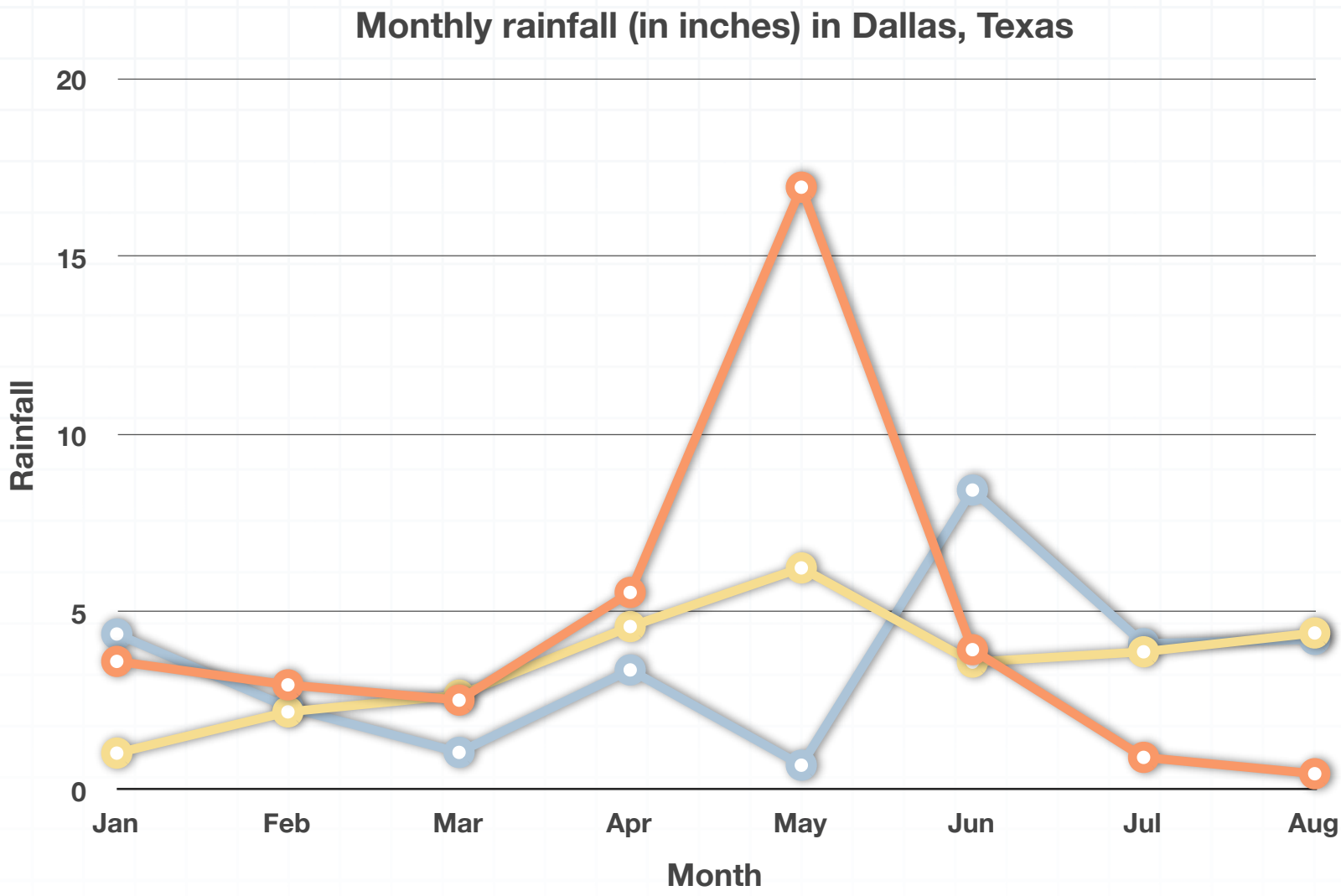
■ 6. Is a comparison line graph an appropriate visual display for the data table, which shows monthly rainfall (in inches) for Dallas, Texas, January -

August? Why or why not? If it's an appropriate display, create a comparison line graph. If it's not an appropriate display for the data, create a comparison bar graph.

	2015	2016	2017
January	3.62	1.04	4.39
February	2.96	2.20	2.33
March	2.53	2.67	1.06
April	5.56	4.60	3.38
May	16.96	6.25	0.70
June	3.95	3.60	8.44
July	0.92	3.89	4.12
August	0.46	4.42	4.24

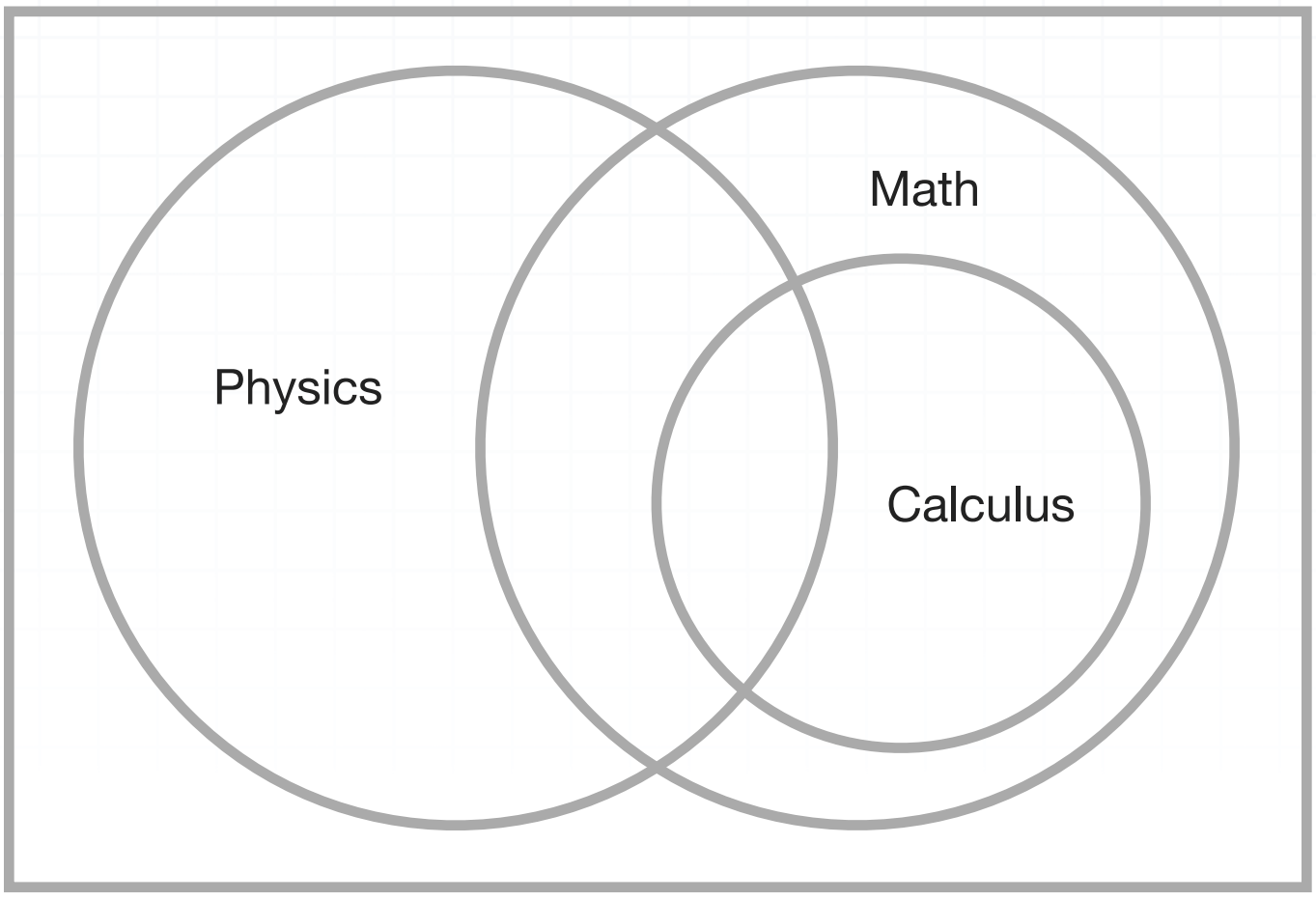
Solution:

Yes, a comparison line graph is an appropriate visual display for the data because it would be useful to track rainfall in Dallas over a given time period.



VENN DIAGRAMS

■ 1. What does the Venn diagram show about how Calculus is related to Physics and Mathematics?



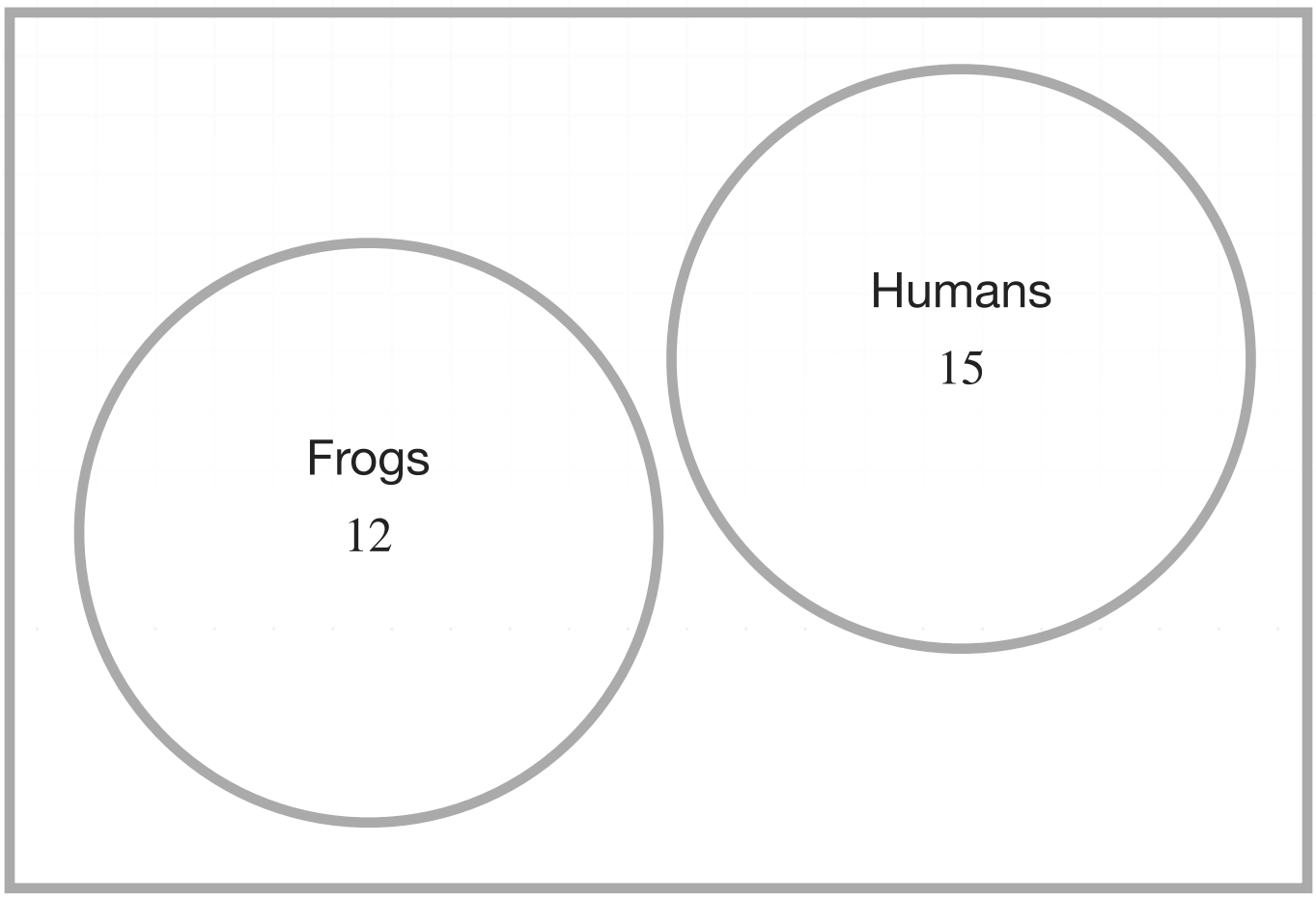
Solution:

We can see from the Venn diagram that all of Calculus is a subset of Mathematics. Some Calculus is part of Physics, although there's also some Mathematics in Physics that does not include Calculus.

■ 2. Draw the Venn diagram for the number of humans in a room and the number of frogs in a room, if the room has 12 frogs and 15 humans.

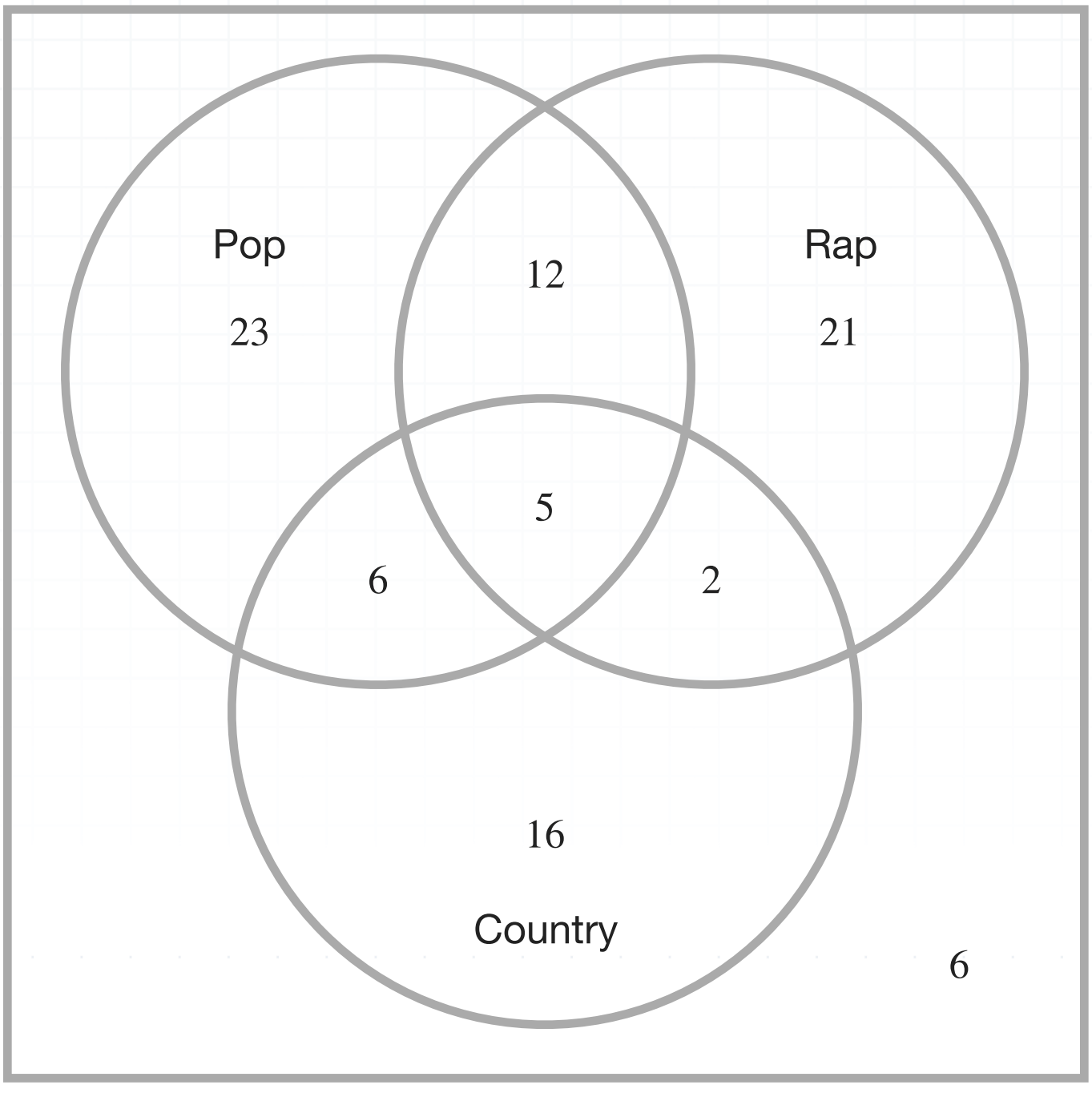
Solution:

Notice that it's possible for a Venn diagram to have no overlapping parts. In this case, the Frogs and Humans do not share any characteristics we're interested in, so the two circles do not overlap.



■ 3. Students at Green Bow High School conducted a survey during lunch time to see what kind of music the students at the school liked. They recorded their results in a Venn diagram. How many students participated

in the survey? What percentage of the students who participated did not like Pop Music?



Solution:

Add up all of the data in the Venn diagram, and don't forget the 6 on the outside.

$$23 + 6 + 5 + 12 + 21 + 2 + 16 + 6 = 91$$

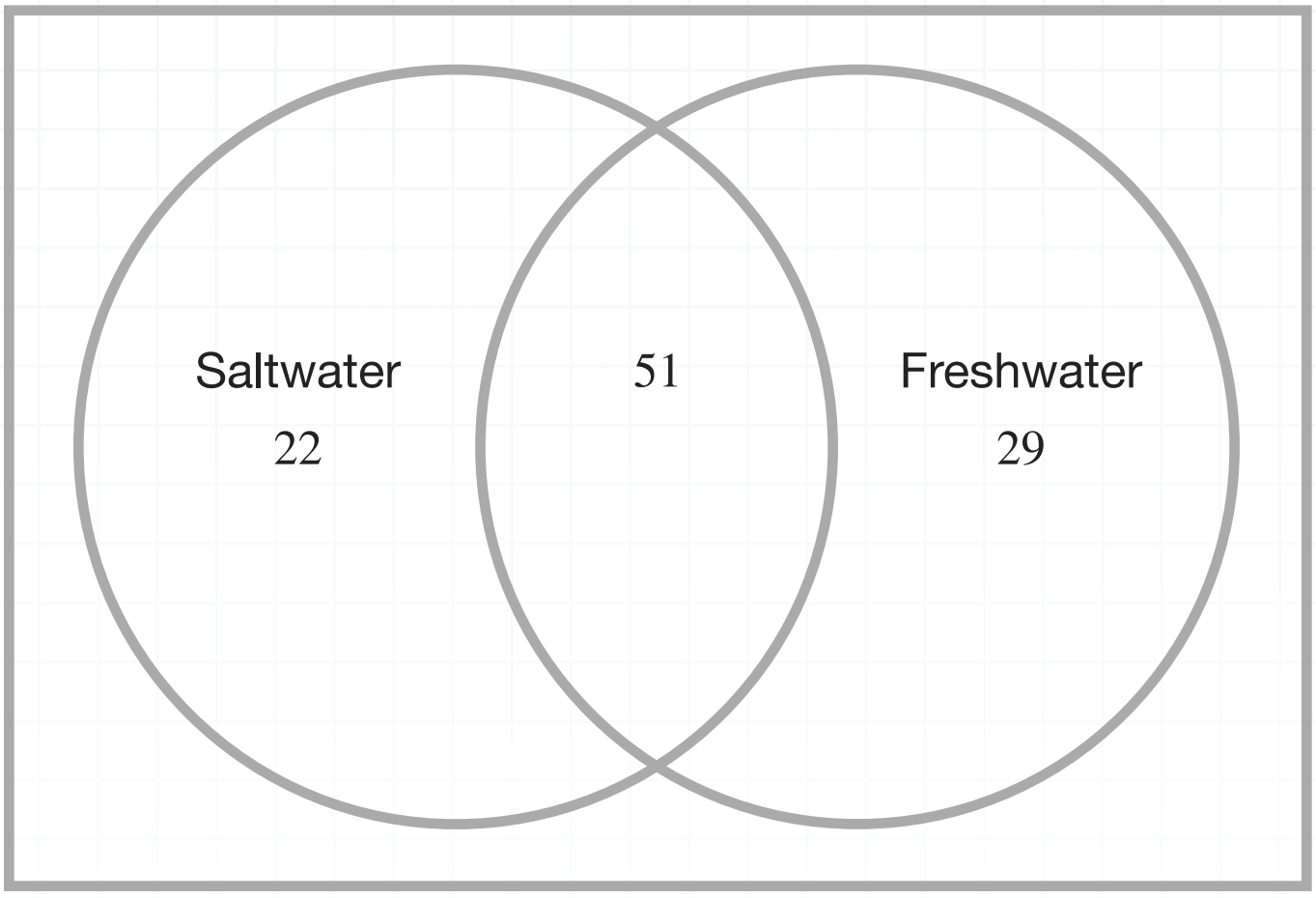
This is the number of students who participated in the survey. The students who did not like Pop music are those who only liked Rap (21), Country (16), Country and Rap (2), or something else (6). That adds to

$$21 + 2 + 16 + 6 = 45$$

The total number of students who participated in the survey was 91. Therefore, the percentage who didn't like Pop is then

$$\frac{45}{91} = 49\,\%$$

■ 4. A survey team is collecting data on a type of minnow that lives where a river meets the sea. They place nets in the river, where the river and sea meet and where there is only sea. They count the minnows caught in each net. What percent of the minnows were living in the brackish water? Brackish water is water that's a combination of fresh and saltwater.



Solution:

$50\,\%$ of the minnows were from brackish water. You can find the total number of minnows caught in the sample by adding $22 + 51 + 29 = 102$. We can read in the overlap that 51 of the minnows were caught in the brackish water since it's a combination of the saltwater and freshwater. Now we can calculate the percent as $51/102 = 0.50 = 50\,\%$. This means $50\,\%$ of the minnows were from brackish water.

■ 5. Fill in the Venn diagram using the following information.

18 people's favorite exercise was swimming.

13 people's favorite exercise was running.

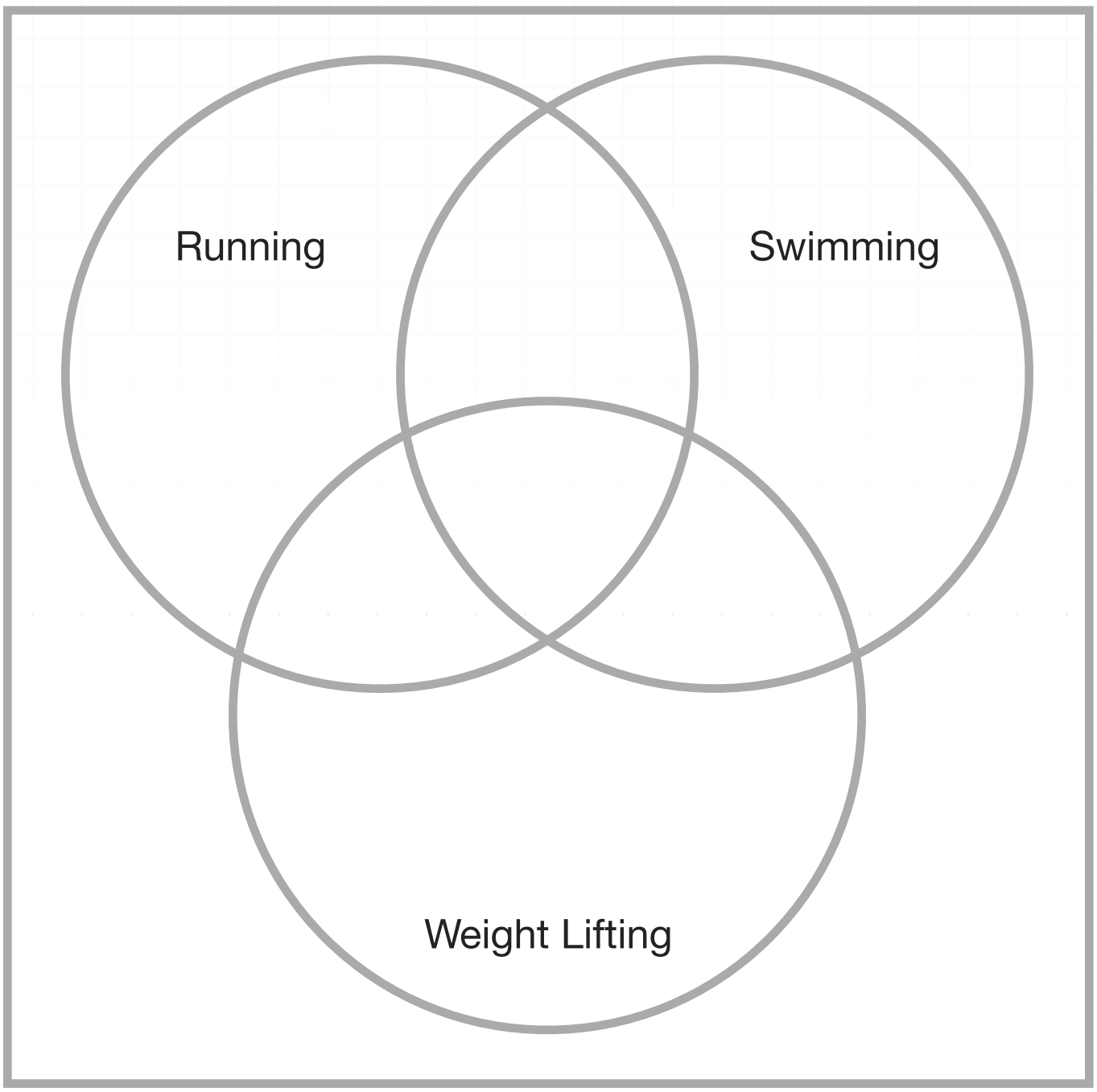
10 people only liked weight lifting.

5 people liked swimming and weight lifting equally.

4 people liked running and weight lifting equally, but not swimming.

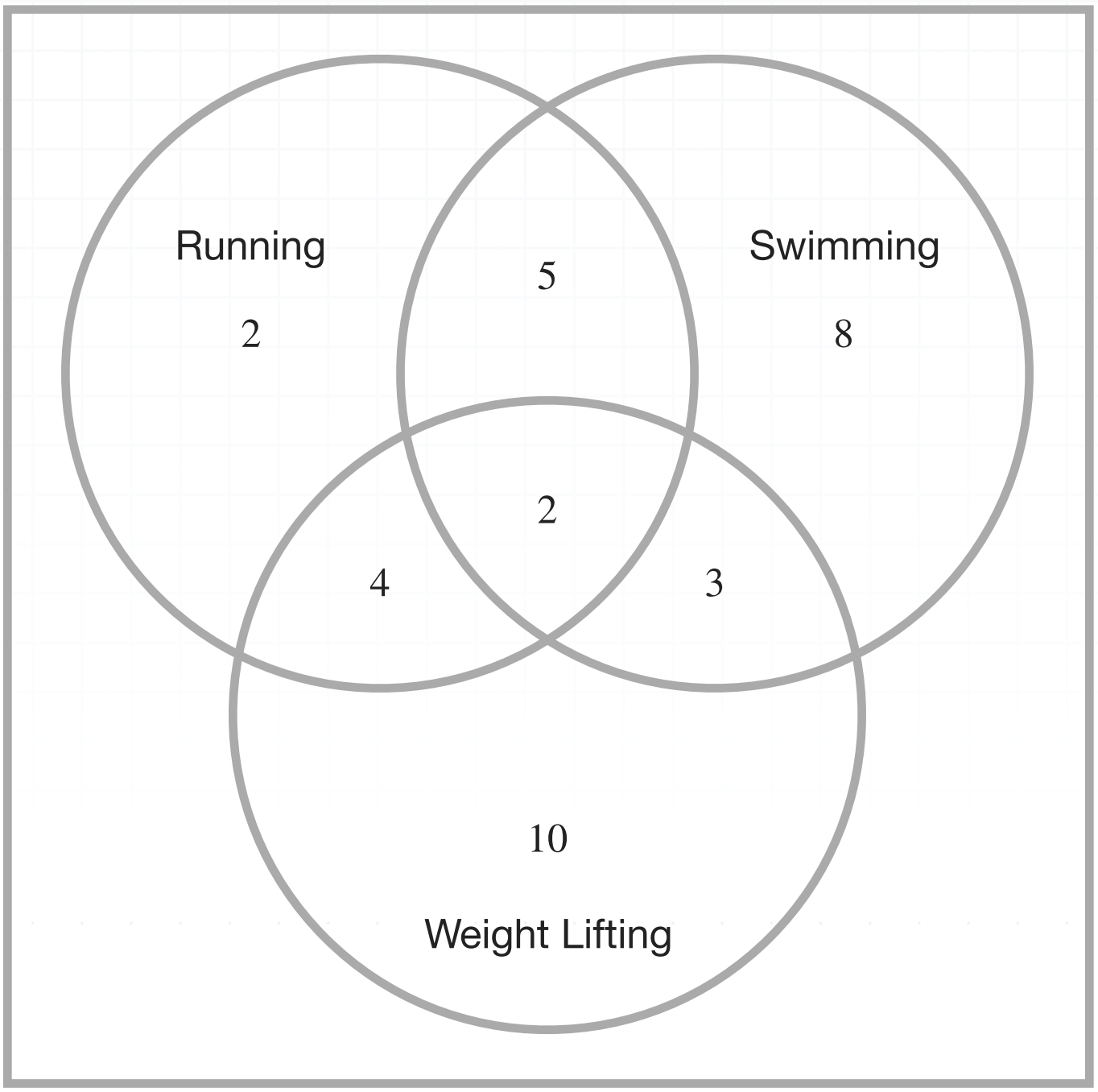
5 people liked running and swimming equally, but not weight lifting.

2 people liked all three equally.



Solution:

From the information we were given, this is the Venn diagram:

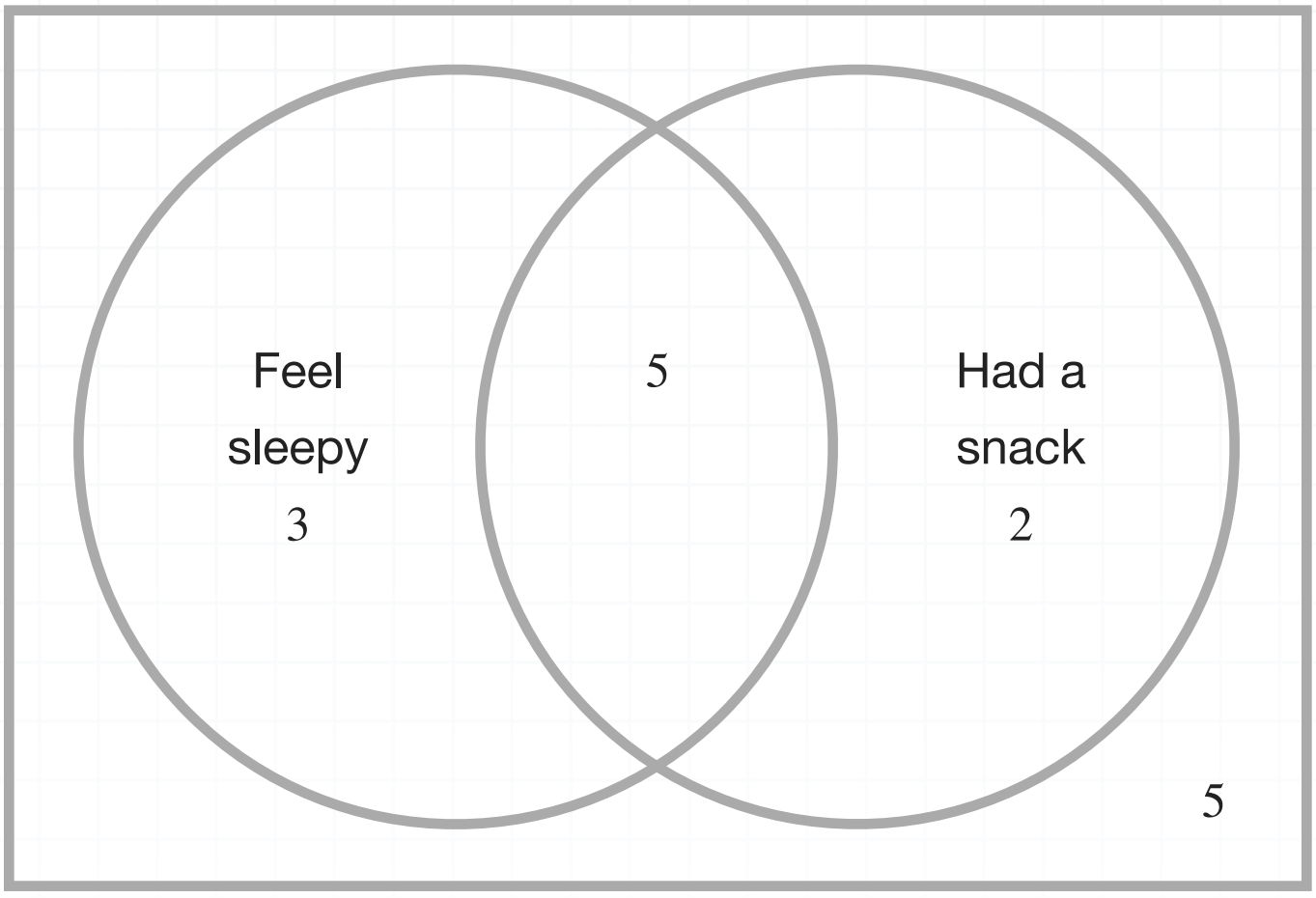


■ 6. Eric creates a survey asking students who ate a snack in the morning between classes if they felt sleepy or not. He organizes his survey results into a two-way data table. Draw a Venn diagram for Eric's survey results.

		Do you feel sleepy?		
		Yes	No	Total
Did you eat a snack?	Yes	5	2	7
	No	3	5	8
	Total	8	7	15

Solution:

There are 5 students who ate a snack and feel sleepy, so we'll put a 5 in the middle. There are 3 students who didn't eat a snack but feel sleepy, so we'll put a 3 in the "feel sleepy" circle. There are 2 students who had a snack but don't feel sleepy, so we'll put a 2 in the "had a snack" circle. And there are 5 students who didn't have a snack and don't feel sleepy, so we'll put a 5 outside of both circles.



RELATIVE FREQUENCY TABLES

1. Blake is surveying students in his class (made up of juniors and seniors) about whether or not they play video games on a daily basis. What type of relative frequency table is shown? Finish filling in the table.

	Play at least one video game daily	Don't play any video games daily	Total
Junior	23%		**75%**
Senior		14%	
Total			**100%**

Solution:

This is a total-relative frequency table because there's a $100\,\%$ in the grand total box. We can use the information in the table to fill out the rest of the information.

	Play at least one video game daily	Don't play any video games daily	Total
Junior	23%	52%	**75%**
Senior	11%	14%	**25%**
Total	**34%**	**66%**	**100%**

■ 2. Create the row-relative frequency table for the frequency table below displaying 9th grade students who participate in an after school activity, and then answer the question: What percent of female 9th grade students do not participate in an after school activity?

	Participate	Don't participate
Male	62	40
Female	57	38

Solution:

40 % of female 9th grade students don't participate in an after school activity. To figure this out, we need to create a row-relative frequency.

We'll start by finding row totals.

	Participate	Don't participate	Total
Male	62	40	**62+40=102**
Female	57	38	**57+38=95**

Now we can turn the table into a row-relative frequency table.

	Participate	Don't participate	Total
Male	62/102=61%	40/102=39%	**102/102=100%**
Female	57/95=60%	38/95=40%	**95/95=100%**

From the finalized table, we can see that $40\,\%$ of female 9th grade students don't participate in an after school activity.

	Participate	Don't participate	Total
Male	61%	39%	**100%**
Female	60%	40%	**100%**

■ 3. Create the column-relative frequency table for this data table and then answer the question: What percentage of those who participate in an after school activity are male?

	Participate	Don't participate
Male	62	40
Female	57	38

Solution:

The first thing we need to do is to find the totals for each column.

	Participate	Don't participate
Male	62	40
Female	57	38
Total	**62+57=119**	**40+38=78**

Now use the column totals to calculate the column-relative frequencies for each column.

	Participate	Don't participate
Male	62/119=52%	40/78=51%
Female	57/119=48%	38/78=49%
Total	**119/119=100%**	**78/78=100%**

From the finalized table, we can see that $52\,\%$ of 9th grade students who participate in an after school activity are male.

	Participate	Don't participate
Male	52%	51%
Female	48%	49%
Total	**100%**	**100%**

■ 4. Create the total-relative frequency table for the data, and then answer this question: Carl is in charge of creating an activity for the students in his college dorm. If Carl wants the highest possible turnout, which activity should he choose? Why?

	Movie	Bowling	Pizza Party
Male	20	40	55
Female	35	50	62

Solution:

The largest percentage of students preferred a pizza party ($45\,\%$), so that is the event that Carl should choose. To figure this out, we need to create a total relative frequency table.

	Movie	Bowling	Pizza Party	Total
Male	20	40	55	**20+40+55=115**
Female	35	50	62	**35+50+62=147**
Total	**20+35=55**	**40+50=90**	**55+62=117**	**115+147=262**

So here is the finished frequency table.

	Movie	Bowling	Pizza Party	Total
Male	20	40	55	**115**
Female	35	50	62	**147**
Total	**55**	**90**	**117**	**262**

Now the total relative frequency table is:

	Movie	Bowling	Pizza Party	Total
Male	20/262=8%	40/262=15%	55/262=21%	**115/262=44%**
Female	35/262=13%	50/262=19%	62/262=24%	**147/262=56%**
Total	**55/262=21%**	**90/262=34%**	**117/262=45%**	**262/262=100%**

	Movie	Bowling	Pizza Party	Total
Male	8%	15%	21%	**44%**
Female	13%	19%	24%	**56%**
Total	**21%**	**34%**	**45%**	**100%**

It looks like the largest percentage of students preferred a pizza party ($45\,\%$), so that's the event that Carl should choose.

■ 5. A city hall is looking into a dangerous intersection that has caused many bicycle accidents over the past month, due to rerouted traffic. They have counted the number of bicycle accidents and put them into a frequency table like the one below. Create the relative frequency table for the data and answer the following question: What day had the highest percentage of bicycle accidents?

Day of the week	Number of crashes
Sunday	13
Monday	10
Tuesday	8
Wednesday	6
Thursday	2
Friday	11
Saturday	14

Solution:

The highest percentage of bicycle accidents ($22\,\%$) happened on Saturday. We know this is true simply because the largest number of bicycle accidents happened on Saturday, but we could also create a relative frequency table, first by finding a total.

Day of the week	Number of crashes
Sunday	13
Monday	10
Tuesday	8
Wednesday	6
Thursday	2
Friday	11
Saturday	14
Total	**64**

Now we'll find the crashes on each day as a percentage of the total, and we can see that the highest percentage of accidents is still occurring on Saturday.

Day of the week	Number of crashes
Sunday	13/64=20%
Monday	10/64=16%
Tuesday	8/64=13%
Wednesday	6/64=9%
Thursday	2/64=3%
Friday	11/64=17%
Saturday	14/64=22%
Total	**64/64=100%**

■ 6. Addie took a poll of the children in her neighborhood. She found that 15 of them watch 2 hours or more of cartoons per day. Out of the 15 that watch 2 hours or more, 10 watched the cartoons on a device other than

the television. There were also 12 children who watched less than 2 hours of cartoons per day. For those 12 children, 2 of them watched cartoons on a device other than a television. Construct a two-way table to summarize the data and then construct a total-relative frequency table for the data.

Solution:

Given what we know, we can fill in the table with this information:

	< 2 hours	> 2 hours	Total
Watched on T.V.			
Watched on a different device	2	10	
Total	**12**	**15**	

And then we can fill in the rest of the table:

	< 2 hours	> 2 hours	Total
Watched on T.V.	10	5	**15**
Watched on a different device	2	10	**12**
Total	**12**	**15**	**27**

And then we can convert this to a total-relative frequency table by dividing by the total in the lower right.

	< 2 hours	> 2 hours	Total
Watched on T.V.	10/27=37%	5/27=19%	**15/27=56%**
Watched on a different device	2/27=7%	10/27=37%	**12/27=44%**
Total	**12/27=44%**	**15/27=56%**	**27/27=100%**

So the total-relative frequency table is:

	< 2 hours	> 2 hours	Total
Watched on T.V.	37%	19%	**56%**
Watched on a different device	7%	37%	**44%**
Total	**44%**	**56%**	**100%**

JOINT DISTRIBUTIONS

1. To study the relationship between votes for a new park and people who have children, a community group surveyed voters. What percentage of those surveyed had children? Is this part of the joint, conditional, or marginal distribution?

	For	Against	No opinion
Children	125	50	30
No children	40	150	60

Solution: About $45\,\%$ of those surveyed had children. This is part of the marginal distribution.

	For	Against	No opinion	Total
Children	125	50	30	**205**
No children	40	150	60	**250**
Total	**165**	**200**	**90**	**455**

Now we can calculate the percentage of the voters who had children:

$$\frac{205}{455} \approx 45\,\%$$

Since this calculation was done from the total column it is part of the marginal distribution.

■ 2. To study the relationship between votes for a new park and people who have children, a community group surveyed voters. What percentage of those surveyed were for the park and had children? Is this part of the joint, conditional, or marginal distribution?

	For	Against	No opinion
Children	125	50	30
No children	40	150	60

Solution: About $27\,\%$ of those surveyed voted for the park and had children. This is part of the joint distribution.

	For	Against	No opinion	Total
Children	125	50	30	**205**
No children	40	150	60	**250**
Total	**165**	**200**	**90**	**455**

Now we can calculate the percentage of the voters who had children and voted for the park:

$$\frac{125}{455} \approx 27\,\%$$

Since this is dependent on the grand total, this is part of the joint distribution.

■ 3. To study the relationship between votes for a new park and people who have children, a community group surveyed voters. What percentage of those with no children had no opinion? Is this part of the joint, conditional, or marginal distribution?

	For	Against	No opinion
Children	125	50	30
No children	40	150	60

Solution: About $24\,\%$ of those with no children had no opinion. This is part of a conditional distribution. Here we need to only look at the 250 people surveyed who had no children. Out of those we want to know who had "no opinion" on the park.

	For	Against	No opinion	Total
Children	125	50	30	**205**
No children	40	150	60	**250**
Total	**165**	**200**	**90**	**455**

Since we're interested in a subset of those surveyed, this is a conditional distribution. The percentage of those with no children who had no opinion is:

$$\frac{60}{250} \approx 24\,\%$$

■ 4. Carl is in charge of creating an activity for the students in his college dorm, and he records their preferences by activity and gender. What percentage of the female students prefer pizza? To answer the question, did you use a marginal, joint, or conditional distribution?

	Movie	Bowling	Pizza Party
Male	20	40	55
Female	35	50	62

Solution: To answer the question, you need to use a conditional distribution. You're interested in the percentage of female students who prefer pizza. So, we're interested only in the conditional distribution for the row of female students.

	Movie	Bowling	Pizza Party	Total
Female	35	50	62	**147**

The percentage of female students who preferred a pizza party was

$$\frac{62}{147} \approx 42\,\%$$

■ 5. A pharmaceutical company is testing heart burn as a side effect of its new pain reliever. What conclusions can you draw from the marginal distributions of the study?

	Pain reliever	Placebo	Total
Minor heartburn	4	171	**175**
Major heartburn	102	25	**127**
No heartburn	10,568	10,478	**21,046**
Total	**10,674**	**10,674**	**21,348**

Solution: Remember that for the marginal distributions, we're just looking at the total column and total row. $10{,}674$ of the $21{,}348$ people in the study took the pain reliever, and $10{,}674$ took the placebo. This means

$$\frac{10{,}674}{21{,}348} \approx 50\,\%$$

took the pain reliever and about

$$\frac{10{,}674}{21{,}348} \approx 50\,\%$$

took the placebo. We also know 175 of those in the study experienced minor heartburn, or

$$\frac{175}{21{,}348} \approx 0.8\,\%$$

127 experienced major heartburn,

$$\frac{127}{21{,}348} \approx 0.6\,\%$$

and $21{,}046$ or

$$\frac{21{,}046}{21{,}348} \approx 99.5\,\%$$

experienced no heartburn. Without calculating the conditional probabilities, we can't say much more about heartburn as a side effect of the pain reliever.

■ 6. Consider the same data as the previous question. What do the conditional distributions (given the participant experienced minor heartburn, major heartburn, or no heartburn) tell us about the study?

	Pain reliever	Placebo	Total
Minor heartburn	4	171	**175**
Major heartburn	102	25	**127**
No heartburn	10,568	10,478	**21,046**
Total	**10,674**	**10,674**	**21,348**

Solution: The conditional distributions described here are the row-relative frequencies. Of the 175 people in the study who had minor heartburn, 4 took the pain reliever and 171 took the placebo.

	Pain reliever	Placebo	Total
Minor heartburn	4/175=2.3%	171/175=97.7%	**100%**

More people who took the placebo suffered from minor heartburn than those that took the pain reliever. It's probably safe to say that taking the pain reliever doesn't cause minor heartburn.

Of the 127 people in the study who had major heartburn, 102 took the pain reliever and 25 took the placebo.

	Pain reliever	Placebo	Total
Major heartburn	102/127=80.3%	25/127=19.7%	**100%**

More people who took the pain reliever suffered from major heartburn than those that took the placebo.

	Pain reliever	Placebo	Total
No heartburn	10,568/21,046=50.2%	10,478/21,046=49.8%	**100%**

Those who took the pain reliever and placebo were symptom free at roughly the same rate. Due to the discrepancy between major and minor heartburn symptoms, it could be worthwhile to look at the study again to see if there was a problem with the placebo used in the minor heartburn part of the study.

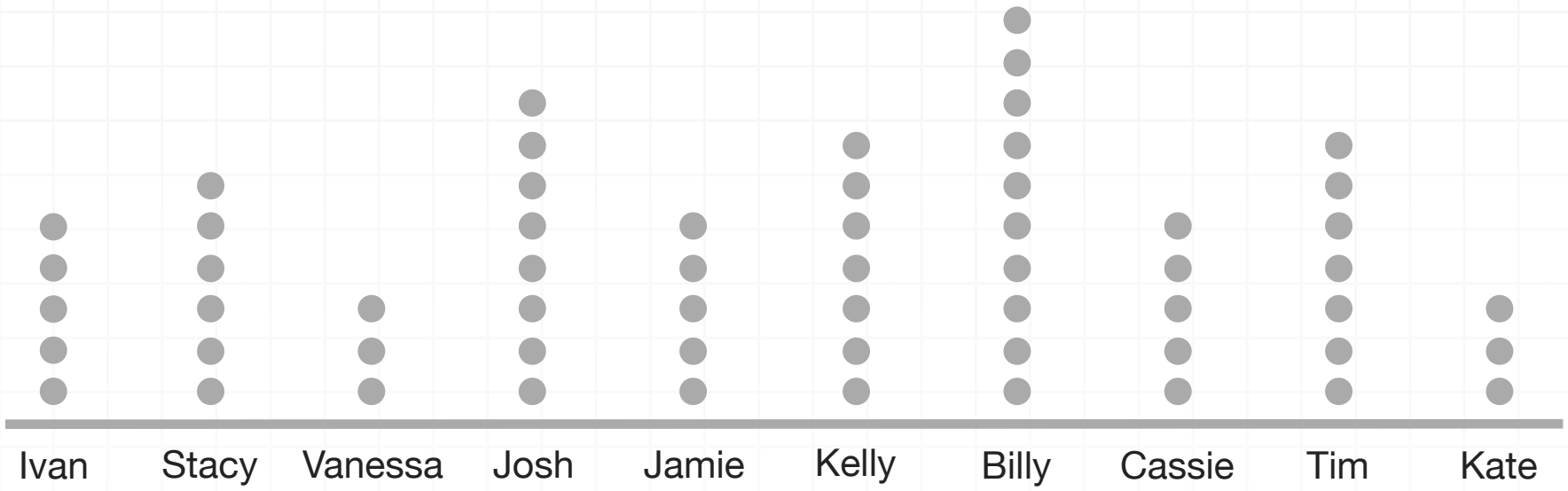
FREQUENCY TABLES AND DOT PLOTS

■ 1. The frequency table shows the number of seed packets sold by each child during a pre-school fundraiser. Create a dot plot from the frequency table.

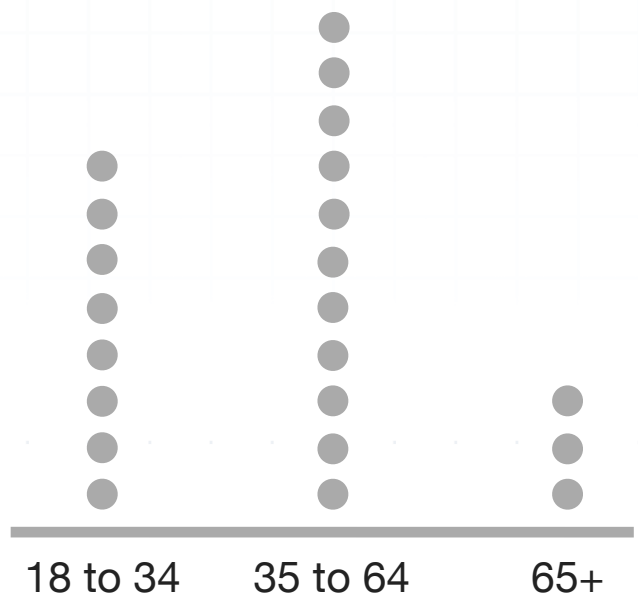
Name	Packets sold
Ivan	5
Stacy	6
Vanessa	3
Josh	8
Jamie	5
Kelly	7
Billy	10
Cassie	5
Tim	7
Kate	3

Solution:

Label the bottom of the dot plot with the names of the preschoolers. Put a dot to represent each packet sold. The dot plot would then look like this:



■ 2. The dot plot shows the age of people who bought a bag of kale at a grocery store. Create a frequency table from the dot plot.



Solution:

Count the number of dots in each column. The number of dots are the frequency of each age group, so the frequency table would look like this:

Age of purchaser	Count
18 to 34	8
35 to 64	11
65+	3

■ 3. The following data shows the number of red marbles drawn in a class lottery. Create a frequency table for the data.

0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 5, 5, 5, 7, 7

Solution:

Count the number of times the same amount of red marbles appeared and organize them into the table.

Red marbles	Frequency
0	3
1	5
2	5
5	3
7	2

■ 4. The following data shows the favorite color of the students in Sebastian's kindergarten class. Create a frequency table for the data.

pink, pink, pink, pink, purple, purple, blue, blue, blue, blue, blue, red, red, red, yellow, orange, orange, green, green, green, black

Solution:

To create the frequency table, count how many of each color you have and record the data in the table.

Color	Frequency
Pink	4
Purple	2
Blue	5
Red	3
Yellow	1
Orange	2
Green	3
Black	1

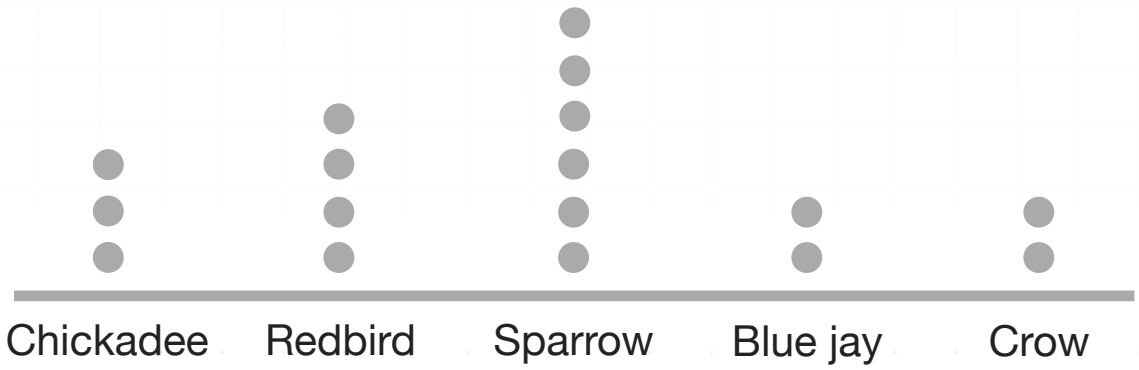
■ 5. Kevin watches birds from his window and records what kind he sees. Create a dot plot from the data.

chickadee, redbird, redbird, redbird, chickadee, sparrow, sparrow, sparrow, sparrow, blue jay, crow, crow, redbird, chickadee, sparrow, sparrow, blue jay

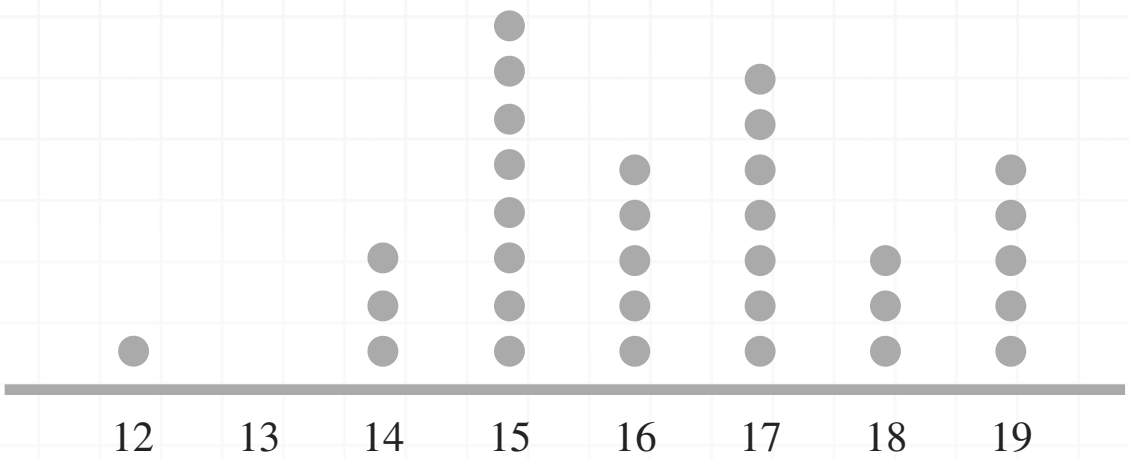
It can help to make a frequency table first. The birds are not grouped according to kind so make sure you count how many of each type you have.

Bird type	Frequency
Chickadee	3
Redbird	4
Sparrow	6
Blue jay	2
Crow	2

Now create the dot plot with the same number of dots for each category as the table has frequencies.



■ 6. The dot plot shows the ages of people in a lifeguard class at the local recreation center. How many people are enrolled in the class who are either 16, 17, or 18 years old?



Solution:

Use the dot plot to count how many lifeguards are 16, 17, and 18. There are 5 lifeguards who are 16, 7 lifeguards who are 17, and 3 lifeguards who are 18, which means there are 15 lifeguards in this age range.

$$5 + 7 + 3 = 15$$

HISTOGRAMS AND STEM-AND-LEAF PLOTS

■ 1. A doctor recorded the weight of all the babies that visited her clinic last week. How many babies weighed no more than 24 pounds?

1	5 5 7 8
2	2 4 6
3	5 6
4	
5	2 6
6	0

$1|5 = 15$

Solution:

"No more than" means we include all of the babies that weigh 24 pounds or less in our count. That means 6 babies weighed no more than 24 pounds.

■ 2. The stem plot shows the number of clothing pieces on each rack at a clothing store. Create a histogram from the stem plot, and use buckets of size 10.

1	0 1 2 8
2	8 8 8
3	2 6 8 9
4	4 4 4
5	2 6
6	0

$$1 \mid 0 = 10$$

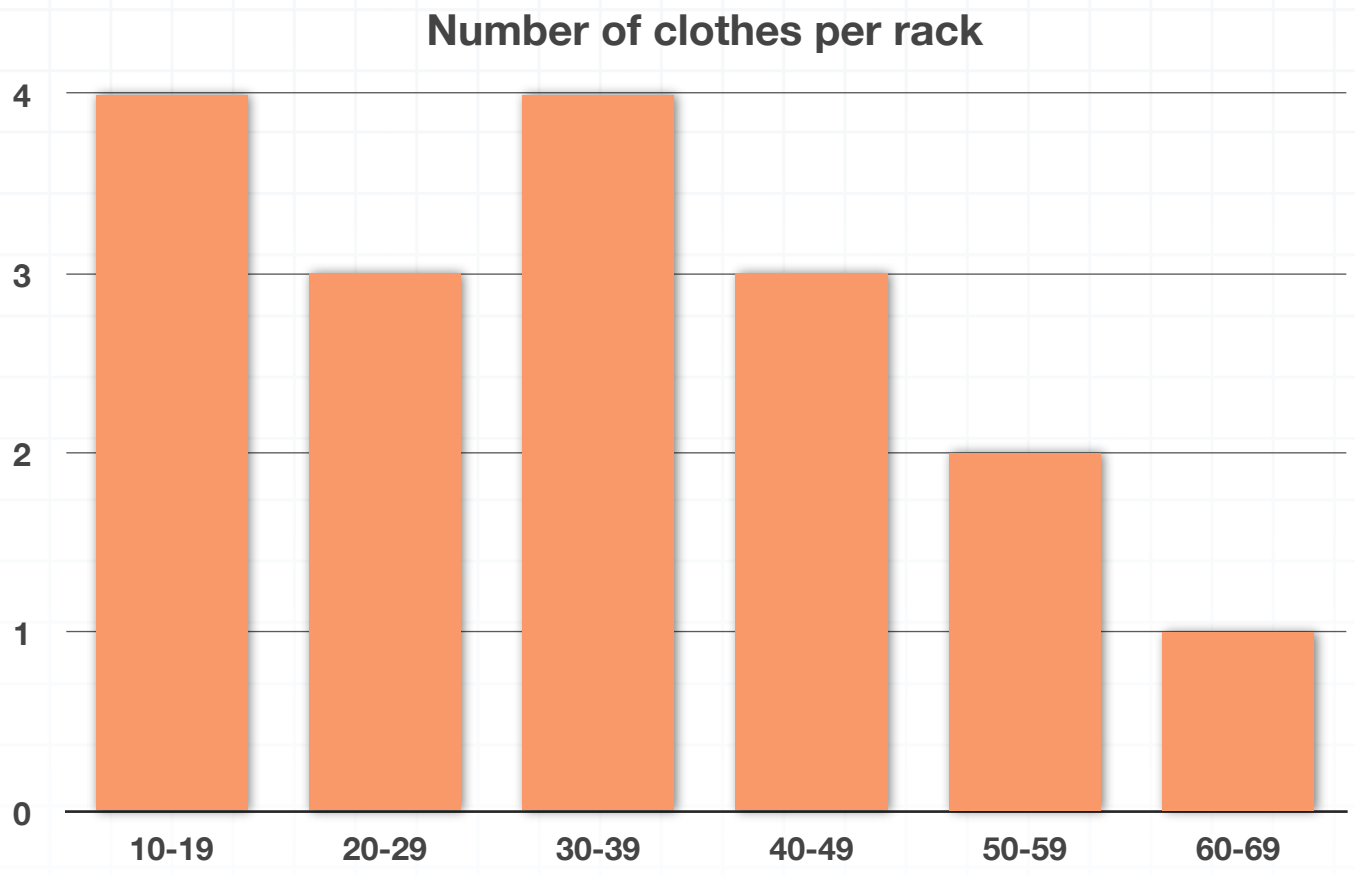
Solution:

The stem-and-leaf plot counts by 10s along the left-hand side. You can then see how many data points should go into each bucket because we're using buckets of size 10.

In this way, the stems become the buckets and the number of leaves become the frequencies graphed in the histogram.

Buckets	Number of leaves
10-19	0 1 2 8
20-29	8 8 8
30-39	2 6 8 9
40-49	4 4 4
50-59	2 6
60-69	0

Now you can turn this frequency table into a histogram.

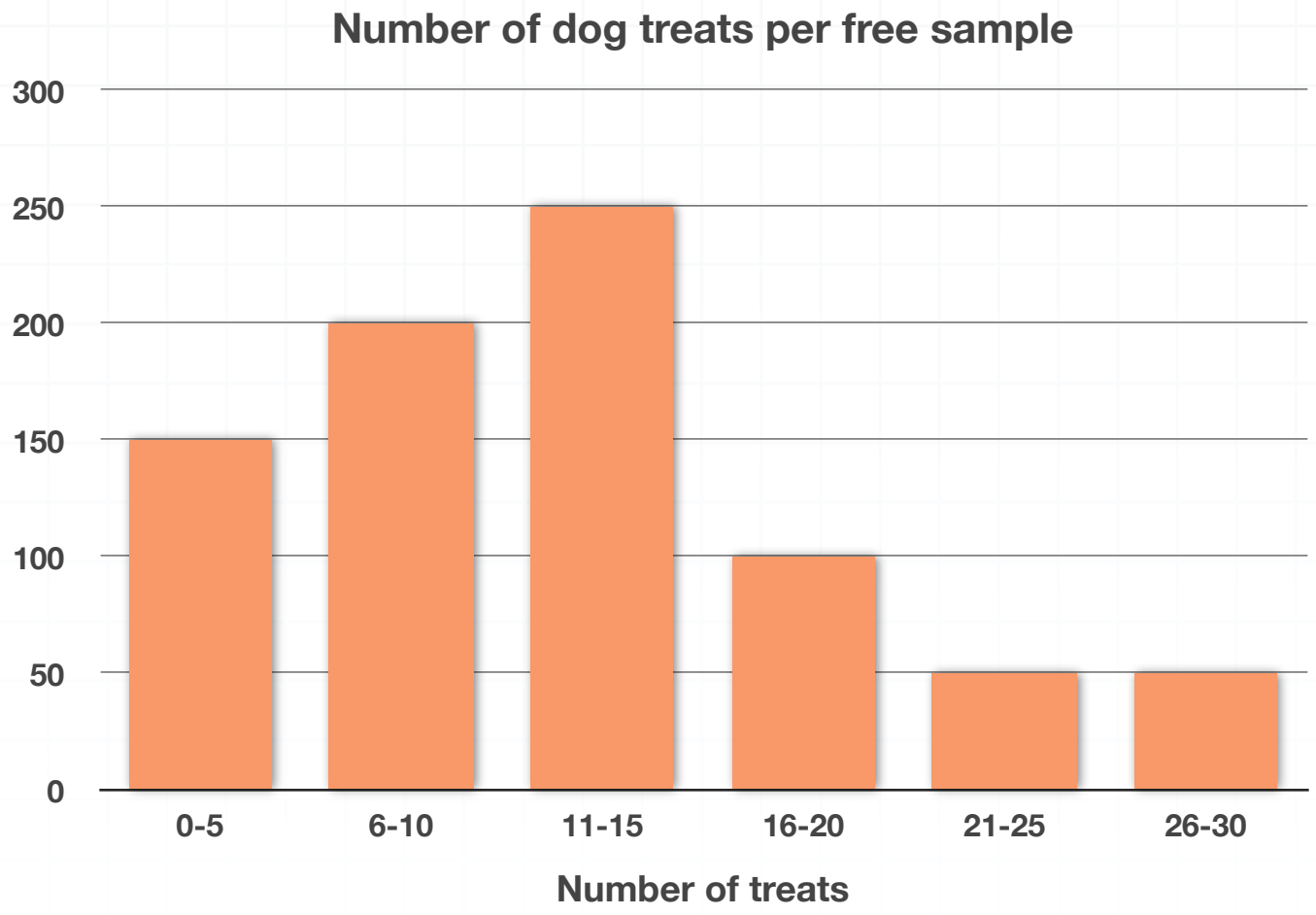


3. Is it possible to create a stem-and-leaf plot from a histogram? Why or why not?

Solution:

You can't make a stem-and-leaf plot from a histogram, because a stem-and-leaf records each data point, while a histogram records how many data points occur in a certain range. This means that a histogram doesn't contain specific enough information to create a stem-and-leaf plot.

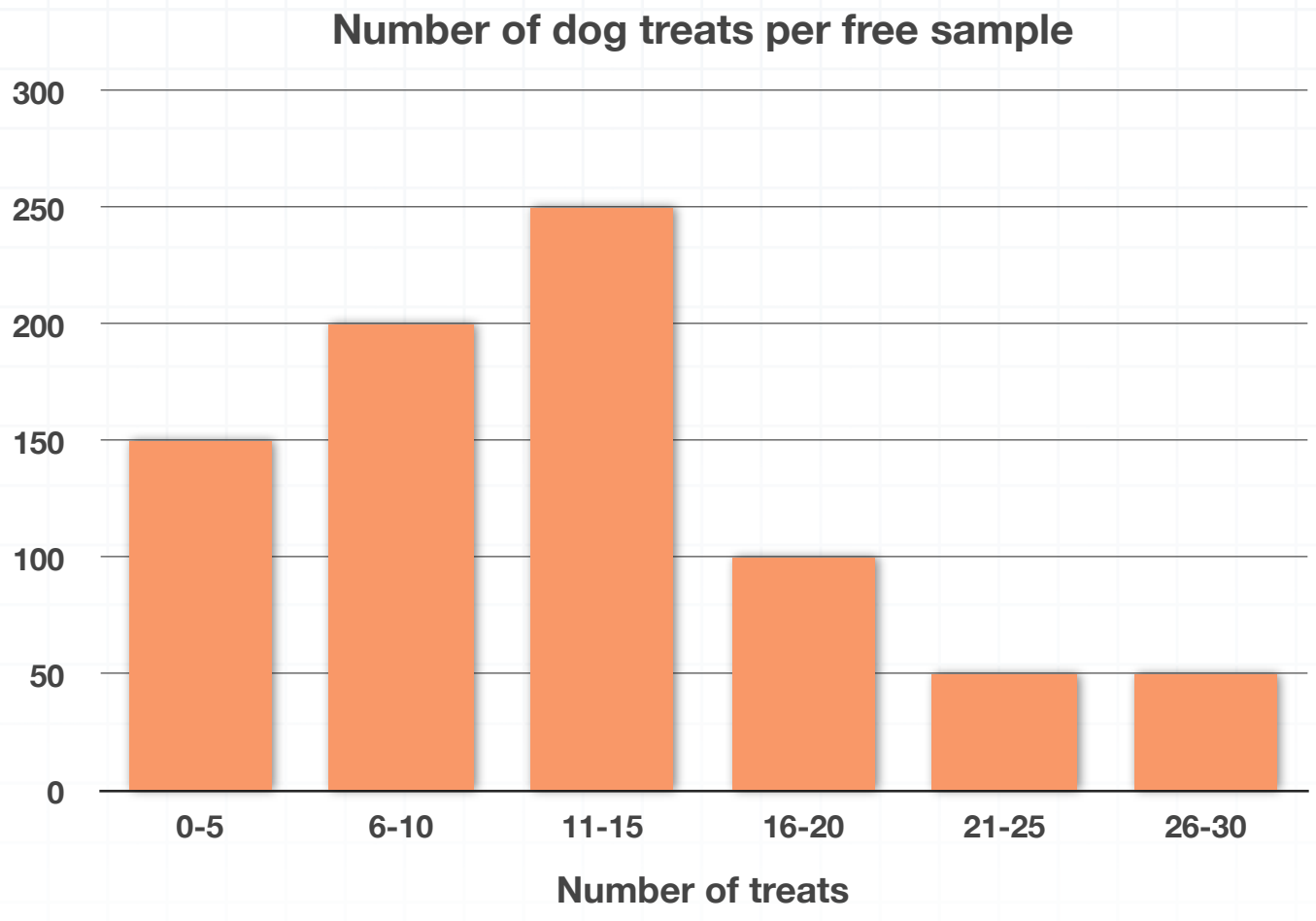
■ 4. A company mails out packets of dog treat samples based on a consumer's previous dog food purchases. How many times did the company mail a packet of $11 - 15$ treats?



Solution:

The bar that is labeled $11 - 15$ has a frequency of 250. This means the company mailed out bags with $11 - 15$ treats 250 times. So 250 samples contained $11 - 15$ treats.

■ 5. A company mails out packets of dog treat samples based on a consumer's previous dog food purchases. How many packets of dog treat samples did the company give out?



Solution:

To find the total number of treats the company mailed out, add up the frequencies in the histogram.

$$150 + 200 + 250 + 100 + 50 + 50 = 800 \text{ samples}$$

The company mailed out 800 packets of treat samples.

■ 6. Create a stem-and-leaf chart from the list of student test scores.

$$60,\ 65,\ 80,\ 80,\ 81,\ 82,\ 88,\ 89,\ 90,\ 97,\ 98,\ 100,\ 100$$

Solution:

To make a stem-and-leaf plot, it's helpful to make sure all of your data is in order from smallest to largest. In this case, it makes sense to choose a stem of tens and leaves of ones.

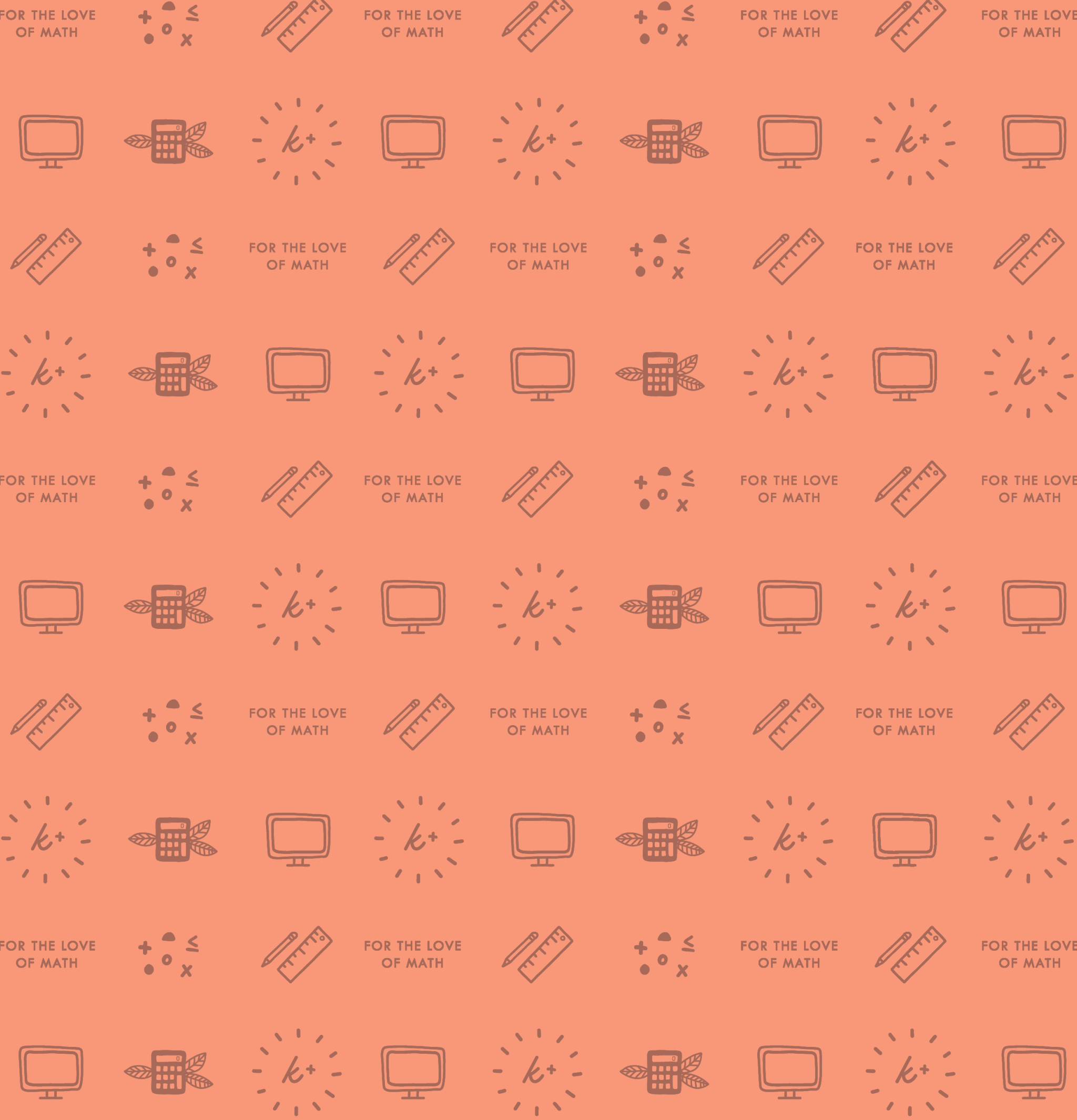
$$60,\ 65,\ 80,\ 80,\ 81,\ 82,\ 88,\ 89,\ 90,\ 97,\ 98,\ 100,\ 100$$

Notice the data does not have any numbers in the 70s, so that stem is left blank. When you get to 100, the leaf needs to be 10 to represent ten 10s, or 100.

The stem-and-leaf plot for student test scores is

6	0 5
7	
8	0 0 1 2 8 9
9	0 7 8
10	0 0

$$6 \mid 0 = 60$$


FOR THE LOVE
OF MATH