Choose your seat

✓ Must be with at least one other person or more

✓ Must be toward the front $\frac{2}{3}$ of the class
(since class is small)

Good morning :)

Today: 8:30 - 9:30

Resources on HW help:

a) Check answers in back of textbook.

b) Check videos available to help with certain problems

c) Come in before class (7:50 -)

d) Let me know before class if you want to go over a problem in class either by...
- email or Remind App
- board

but, don't cheat yourself out of the productive struggle opportunity (perseverance)

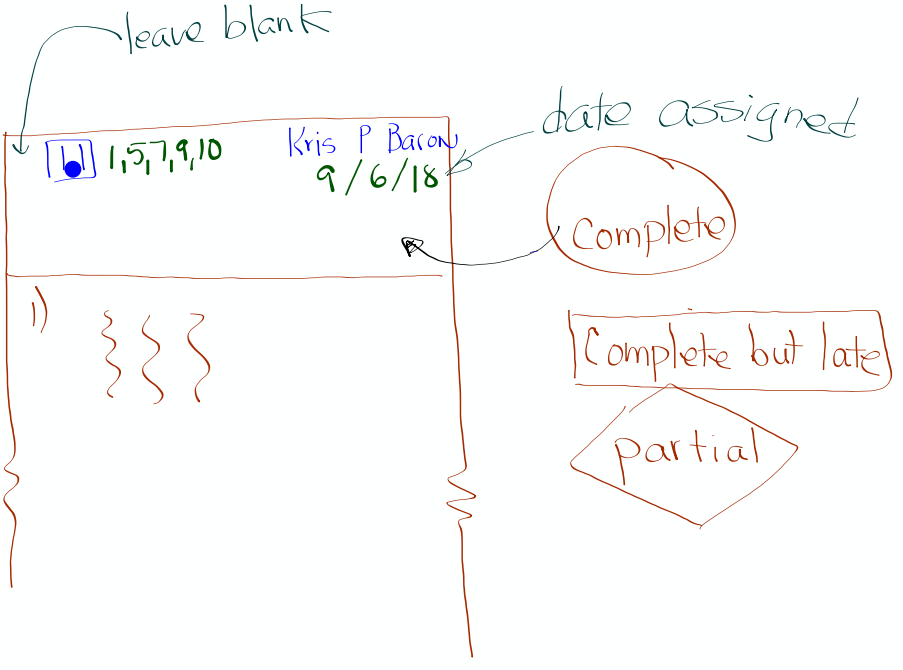
before seeking help !!

#7 Magazine College Rankings

Possible Categorical Variables: region of country
type of institution (2 yr, 4 yr)
zipcode

Possible Quantitative Variables: retention rate
graduation rate
class size (Avg)
faculty salaries
student-faculty ratio
Avg. financial aid
Percentage of alumni who give to the school

10) C



Today and Monday

Data Analysis

Section 1.1
Analyzing Categorical Data

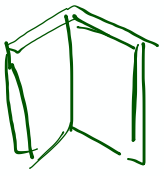
Handout

LEARNING TARGETS

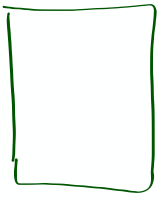
By the end of this section, you should be able to:

- ✓ MAKE and INTERPRET bar graphs for categorical data.
- ✓ IDENTIFY what makes some graphs of categorical data misleading.
- ✓ CALCULATE marginal and joint relative frequencies from a two-way table.
- ✓ CALCULATE conditional relative frequencies from a two-way table.
- ✓ Use bar graphs to COMPARE distributions of categorical data.
- ✓ DESCRIBE the nature of the association between two categorical variables.

Suggestion



textbook + Handout



Organizing Categorical Data

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Categorical variable

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Categorical variable

Values (These are the data)

Frequency table

Preferred method	Frequency
Cell phone	2
Facebook	1
In person	3
Text messaging	4

Relative frequency table

Preferred method	Relative frequency
Cell phone	2/10 = 0.20 or 20%
Facebook	1/10 = 0.10 or 10%
In person	3/10 = 0.30 or 30%
Text messaging	4/10 = 0.40 or 40%

Frequency table

Preferred method	Frequency
Cell phone	2
Facebook	1
In person	3
Text messaging	4

Count

Relative frequency table

Preferred method	Relative frequency
Cell phone	2/10 = 0.20 or 20%
Facebook	1/10 = 0.10 or 10%
In person	3/10 = 0.30 or 30%
Text messaging	4/10 = 0.40 or 40%

Frequency table

Preferred method	Frequency
Cell phone	2
Facebook	1
In person	3
Text messaging	4

Count

Relative frequency table

Preferred method	Relative frequency
Cell phone	2/10 = 0.20 or 20%
Facebook	1/10 = 0.10 or 10%
In person	3/10 = 0.30 or 30%
Text messaging	4/10 = 0.40 or 40%

Proportion

Percent

To display the distribution of categorical data, make a **bar graph**

Frequency table

Preferred method	Frequency
Cell phone	2
Facebook	1
In person	3
Text messaging	4

Count

Relative frequency table

Preferred method	Relative frequency
Cell phone	2/10 = 0.20 or 20%
Facebook	1/10 = 0.10 or 10%
In person	3/10 = 0.30 or 30%
Text messaging	4/10 = 0.40 or 40%

Proportion Percent

1. What is the difference between a data table, a frequency table, and a relative frequency table? When is it better to use relative frequency?

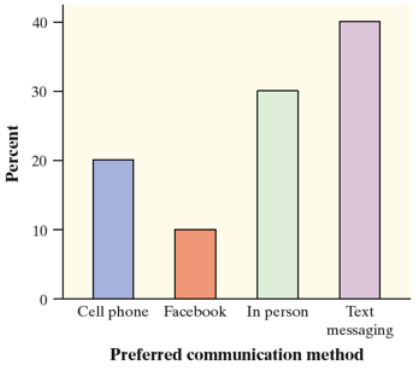
Frequency tables summarize 1 of the variables using counts.

Relative freq. tables summarize percents/proportions

Use relative freq. when comparing groups of different sizes.

For example: Does the distribution of communication method for our class differ worldwide

To display the distribution of categorical data, make a **bar graph**

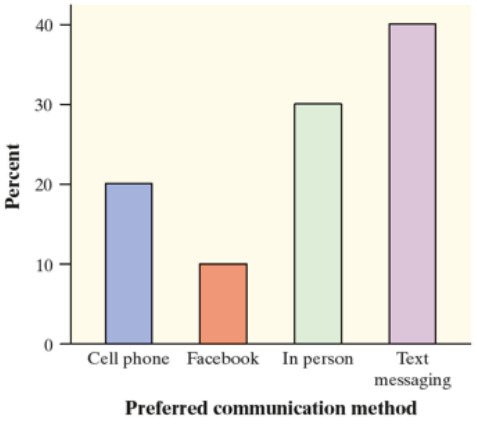


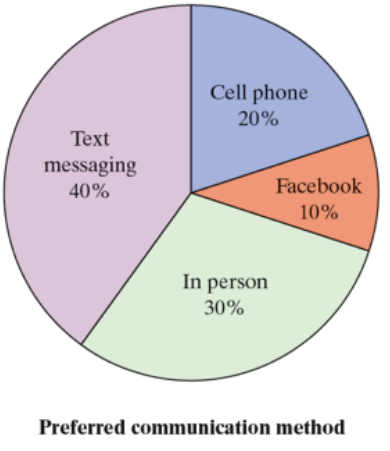
Relative frequency table

Preferred method	Relative frequency
Cell phone	2/10 = 0.20 or 20%
Facebook	1/10 = 0.10 or 10%
In person	3/10 = 0.30 or 30%
Text messaging	4/10 = 0.40 or 40%

Proportion Percent

To display the distribution of categorical data, make a **bar graph** or a **pie chart**.





2. What is the most important thing to remember when making pie charts and bar graphs? Why do statisticians prefer bar graphs?

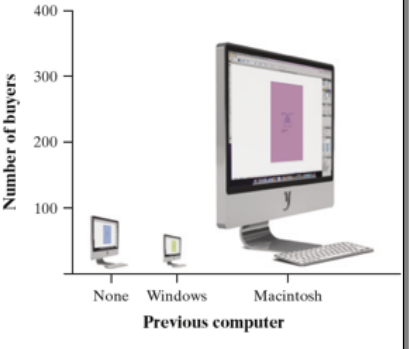
✓ Proper Labeling !

✓ Bar graphs - Easier to make and compare.
- Can compare the percent of success in multiple groups.

Graphs: Good and Bad

Bar graphs are a bit dull to look at. It is tempting to replace the bars with pictures or to use special 3-D effects to make the graphs seem more interesting.

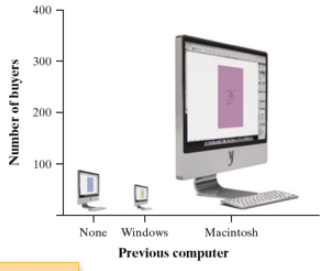
Don't do it!



Graphs: Good and Bad

Bar graphs are a bit dull to look at. It is tempting to replace the bars with pictures or to use special 3-D effects to make the graphs seem more interesting.

Don't do it!



CAUTION:

1) beware the pictograph
2) watch those scales

3. What are some common ways to make a misleading graph?

- ✓ Violating Area Principle (area should be proportional to freq.)
- ✓ Icons (bar widths consistent)
- ✓ Not starting frequency at 0

Analyzing Data on Two Categorical Variables

Analyzing Data on Two Categorical Variables

How do you analyze data do when a data set involves two categorical variables?

Respondent	Environmental club?	Snowmobile use
1	No	Own
2	No	Rent
3	Yes	Never
4	Yes	Rent
5	No	Never
⋮	⋮	⋮

How do you analyze data do when a data set involves two categorical variables?

Respondent	Environmental club?	Snowmobile use
1	No	Own
2	No	Rent
3	Yes	Never
4	Yes	Rent
5	No	Never
⋮	⋮	⋮

A **two-way table** is a table of counts that summarizes data on the relationship between two categorical variables for some group of individuals.

#4

How do you analyze data do when a data set involves two categorical variables?

Respondent	Environmental club?	Snowmobile use
1	No	Own
2	No	Rent
3	Yes	Never
4	Yes	Rent
5	No	Never
⋮	⋮	⋮

A **two-way table** is a table of counts that summarizes data on the relationship between two categorical variables for some group of individuals.

How do you analyze data do when a data set involves two categorical variables?

Respondent	Environmental club?	Snowmobile use
1	No	Own
2	No	Rent
3	Yes	Never
4	Yes	Rent
5	No	Never
⋮	⋮	⋮

A **two-way table** is a table of counts that summarizes data on the relationship between two categorical variables for some group of individuals.

Snowmobile use	Environmental club member? No	Yes
Never	445	212
Rent	497	77
Own	279	16

How do you analyze data do when a data set involves two categorical variables?

Respondent	Environmental club?	Snowmobile use
1	No	Own
2	No	Rent
3	Yes	Never
4	Yes	Rent
5	No	Never
⋮	⋮	⋮

A **two-way table** is a table of counts that summarizes data on the relationship between two categorical variables for some group of individuals.

We can include row and column totals

Snowmobile use	Environmental club No	Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

#5

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

$$\text{No: } \frac{1221}{1526} = 0.800 \text{ or } 80.0\%$$

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

No: $\frac{1221}{1526} = 0.800$ or 80.0% Yes: $\frac{305}{1526} = 0.200$ or 20.0%

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

		Environmental club		
		No	Yes	Total
nowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

No: $\frac{1221}{1526} = 0.800$ or 80.0% Yes: $\frac{305}{1526} = 0.200$ or 20.0%

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

Never: $\frac{657}{1526} = 0.431$ or 43.1%

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

Never: $\frac{657}{1526} = 0.431$ or 43.1%

Rent: $\frac{574}{1526} = 0.376$ or 37.6%

No: $\frac{1221}{1526} = 0.800$ or 80.0% Yes: $\frac{305}{1526} = 0.200$ or 20.0%

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

Never: $\frac{657}{1526} = 0.431$ or 43.1%

Rent: $\frac{574}{1526} = 0.376$ or 37.6%

No: $\frac{1221}{1526} = 0.800$ or 80.0% Yes: $\frac{305}{1526} = 0.200$ or 20.0%

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

Own: $\frac{295}{1526} = 0.193$ or 19.3%

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

Never: $\frac{657}{1526} = 0.431$ or 43.1%

Rent: $\frac{574}{1526} = 0.376$ or 37.6%

Own: $\frac{295}{1526} = 0.193$ or 19.3%

No: $\frac{1221}{1526} = 0.800$ or 80.0% Yes: $\frac{305}{1526} = 0.200$ or 20.0%

A marginal relative frequency tells you about only *one* of the variables in a two-way table.

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

A **joint relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable and a specific value for another categorical variable.

#6

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

A **joint relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable and a specific value for another categorical variable.

A joint relative frequency helps answer questions involving *both* of the variables in a two-way table.

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

A **joint relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable and a specific value for another categorical variable.

A joint relative frequency helps answer questions involving *both* of the variables in a two-way table.

What percent of people in the sample are environmental club members **and** own snowmobiles?

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

A **joint relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable and a specific value for another categorical variable.

What percent of people in the sample are environmental club members **and** own snowmobiles?

$$\frac{16}{1526} = 0.010 = 1.0\%$$

A joint relative frequency helps answer questions involving *both* of the variables in a two-way table.

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

A **joint relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable and a specific value for another categorical variable.

What percent of people in the sample are environmental club members **and** own snowmobiles?

$$\frac{16}{1526} = 0.010 = 1.0\%$$

A joint relative frequency helps answer questions involving *both* of the variables in a two-way table.

What proportion of people in the sample are not environmental club members **and** never use snowmobiles?

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

A **joint relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable and a specific value for another categorical variable.

What percent of people in the sample are environmental club members **and** own snowmobiles?

$$\frac{16}{1526} = 0.010 = 1.0\%$$

What proportion of people in the sample are not environmental club members **and** never use snowmobiles?

$$\frac{445}{1526} = 0.292$$

A joint relative frequency helps answer questions involving *both* of the variables in a two-way table.

Any questions about anything?

Assignment

1.1 ... 13, 15, 17, 19, 21, 23