Lecture 1

DEFINITION **Statistics**

Statistics is the science and art of collecting, analyzing, and drawing conclusions from data.

1. Find a problem
2. Understand the **nature** of the problem
3. How to measure it
4. Data collection
5. Data summarization
6. Formal Analysis
7. Interpretation of results

Lecture 1

V a r i a b l e

Activity

Talk about your holiday about :

1. travel or not (yes or no)
2. # of days used for travelling
3. female or male

Name	Travel or not	# of traveling days	gender

Individuals are objects described by a set of data.

A variable is a characteristic that changes from one individual to another.

A categorical/qualitative variable take on category names or group labels

A quantitative/numerical variable takes on numerical values

Examples

Quantitative Variables	Qualitative Variables
Number of students in a class	Eye color
Number of square feet in a house	Gender
Population size of a city	Breed of dog
Age of an individual	Level of Education
Height of an individual	Marital status

PROBLEM: Census At School is an international project that collects data about primary and secondary school students using surveys. Hundreds of thousands of students from Australia, Canada, Ireland, Japan, New Zealand, South Africa, South Korea, the United Kingdom, and the United States have taken part in the project. Data from the surveys are available online. We used the site's "Random Data Selector" to choose 10 Canadian students who completed the survey in a recent year. The table displays the data.

Variables

10 individuals

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circumference (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

Not every variable that takes number values is quantitative.

zip code, phone number, ID card number .....

Grade level	GPA	Children in family	Homework last night (min)	Android or iPhone?
9	2.3	3	0–14	iPhone
11	3.8	6	15–29	Android
10	3.1	2	15–29	Android
10	4.0	1	45–59	iPhone
10	3.4	4	0–14	iPhone
10	3.0	3	30–44	Android
9	3.9	2	15–29	iPhone
12	3.5	2	0–14	iPhone

Definition

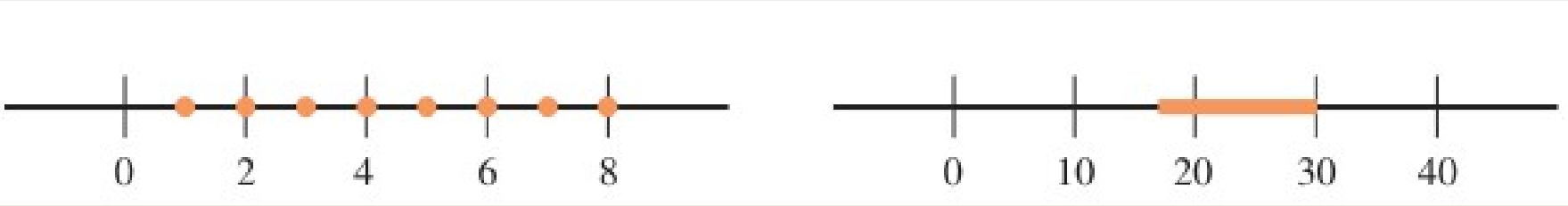
Quantitative variable
- discrete variable
- continuous variable

Discrete variable:

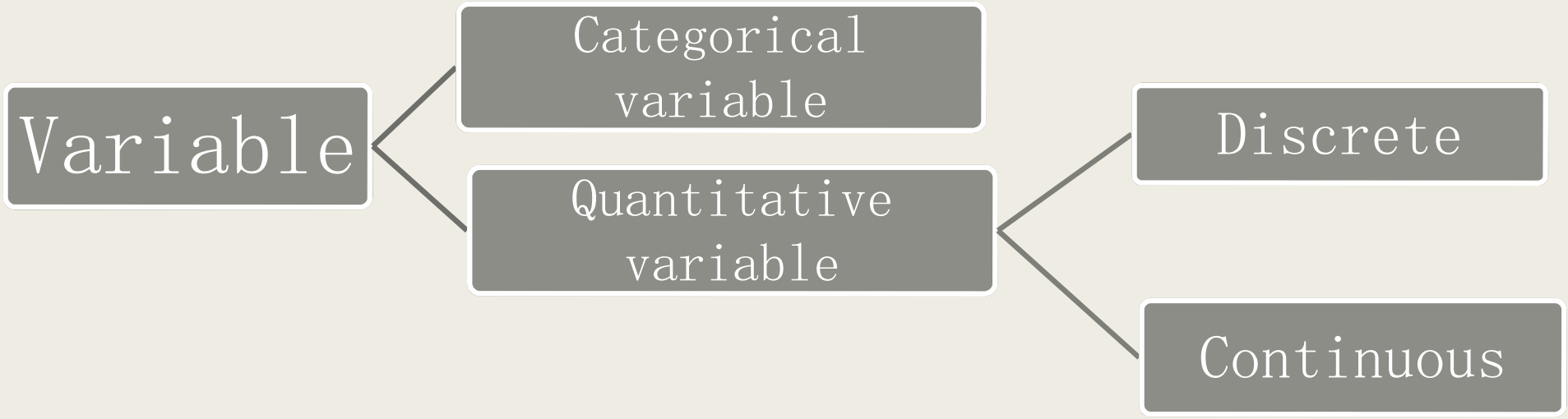
the possible values of the variable correspond to **isolated** points on the number line.

Continuous variable:

the possible values form an **entire interval** on the number line.



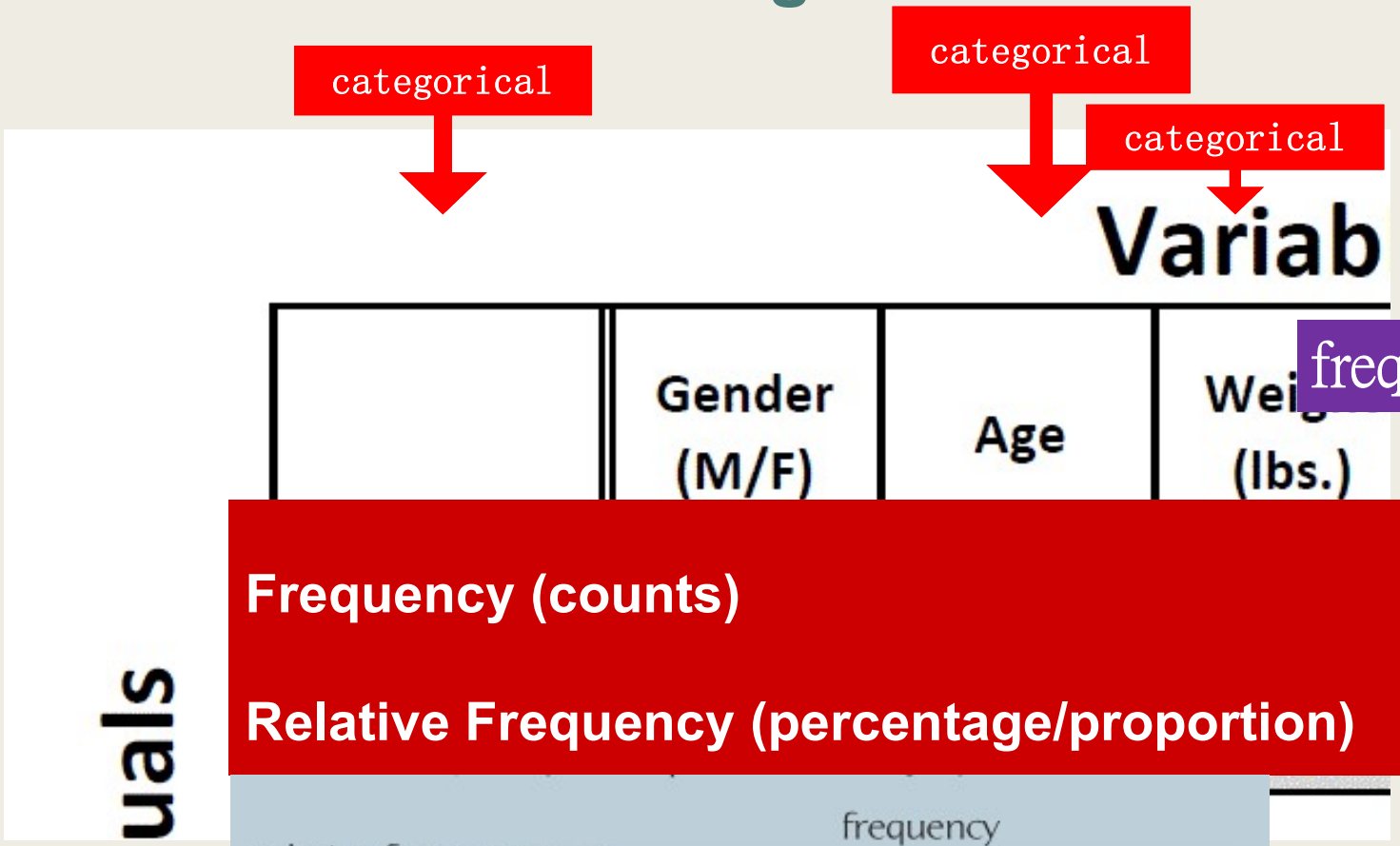
Summary


Variable
Categorical variable
Quantitative variable
Discrete
Continuous

Lecture 2

Displaying Categorical Data with Tables and Graphs

How to measure categorical variables?



Frequency (counts)

Relative Frequency (percentage/proportion)

$$\text{relative frequency} = \frac{\text{frequency}}{\text{number of obervations in the data set}}$$

A **frequency table** shows the number of individuals having each value.

frequency distribution table

A **relative frequency table** shows the proportion or percent of individuals having each value.

Relative frequency distribution table

Here is a small part of the data set that describes the students in a class. The data come from anonymous responses to a questionnaire filled out on the first day of class.

Gender	Grade level	GPA	Children in family	Homework last night (min)	Android or iPhone?
F	9	2.3	3	0–14	iPhone
M	11	3.8	6	15–29	Android
M	10	3.1	2	15–29	Android
F	10	4.0	1	45–59	iPhone
F	10	3.4	4	0–14	iPhone
F	10	3.0	3	30–44	Android
M	9	3.9	2	15–29	iPhone
M	12	3.5	2	0–14	iPhone

Note that the frequencies and relative frequencies listed in these tables are not data !!!!!!!

Homework last night (min)
0–14
15–29
15–29
45–59
0–14
30–44
15–29
0–14

HW last night	Frequency
0-14	3
15-29	3
30-44	1
45-59	1

Note that the frequencies and relative frequencies listed in these tables are not data !!!!!!!

Homework last night (min)
0–14
15–29
15–29
45–59
0–14
30–44
15–29
0–14

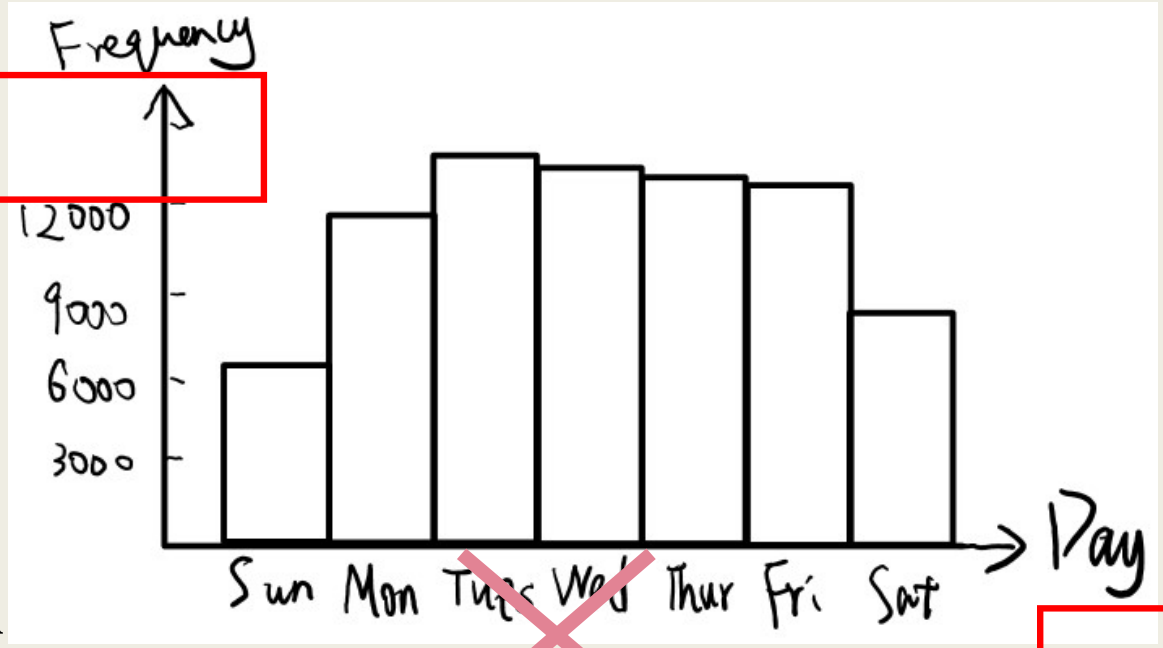
HW last night	Frequency	Relative Frequency Proportion	Relative Frequency Percent
0-14	3	0.375	37.5%
15-29	3	0.375	37.5%
30-44	1	0.125	12.5%
45-59	1	0.125	12.5%

Frequency tables sometimes are difficult to read.

Sometimes it is easier to analyze a distribution by displaying it with a bar graph or pie chart.

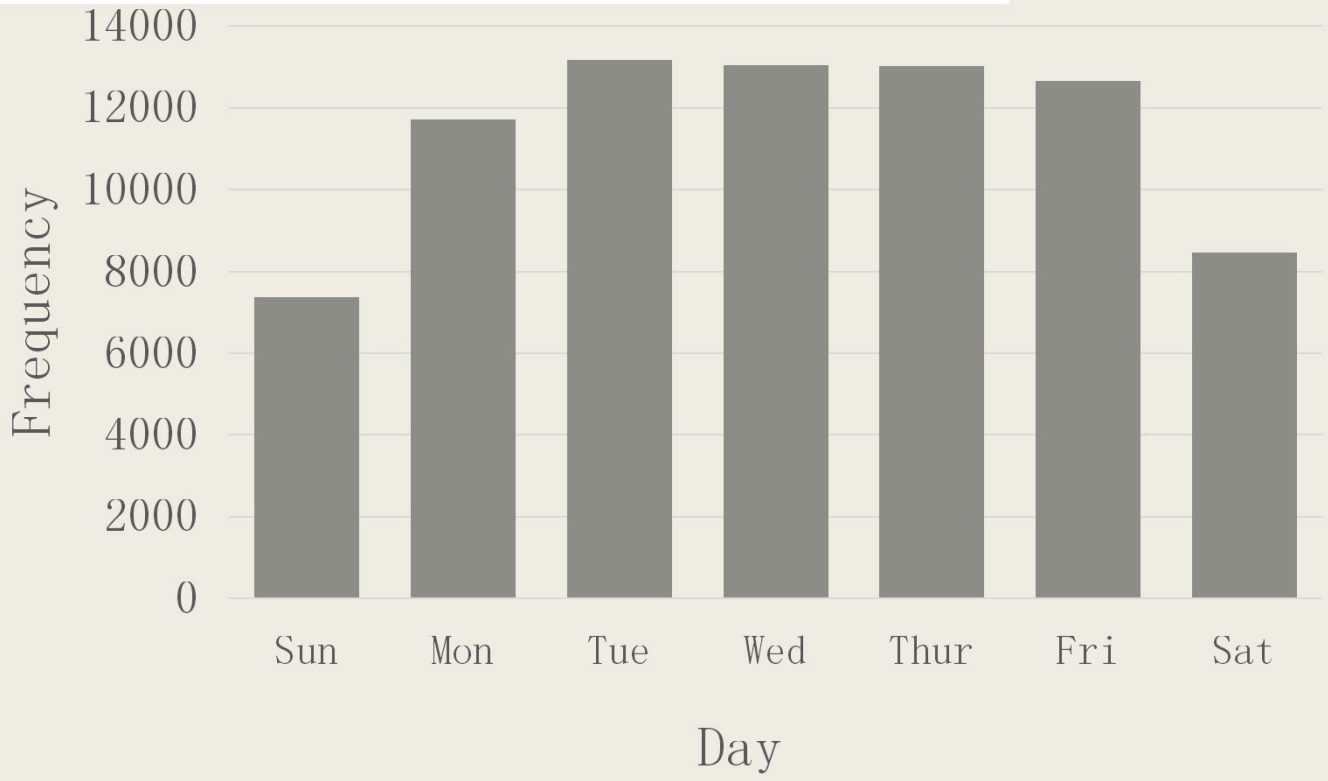
Bar Graph

1. Draw a horizontal axis, and write the category names or labels below the line at **regularly spaced** intervals.
2. Draw a vertical axis, and label the scale using either **frequency or relative frequency**.
3. Place a rectangular bar above each category label. The **heights** of bars that represent the frequency or relative frequency, so all bars should have **the same width**. With the same width, both the height and the area of the bar are proportional to frequency and relative frequency

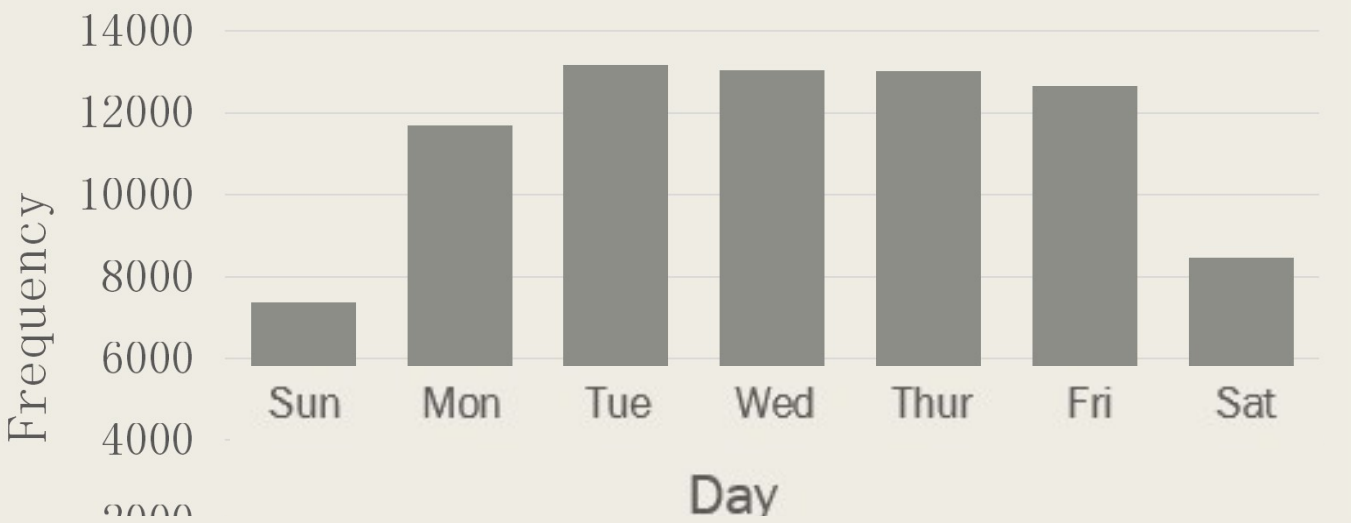


Bar Graph

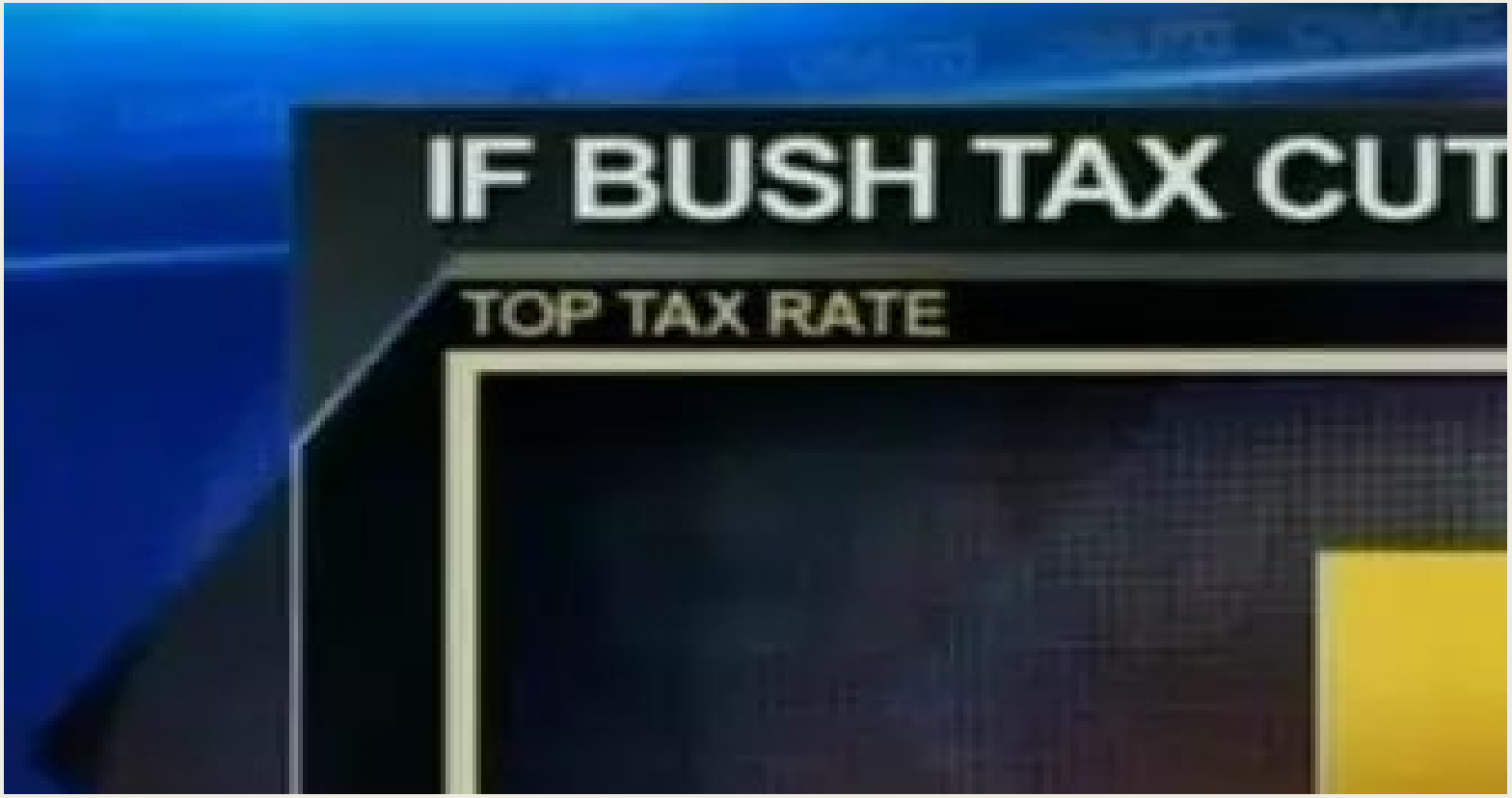
- ✓ bars equally wide.
- ✓ Spaced intervals
- ✓ Name of the horizontal and vertical axis

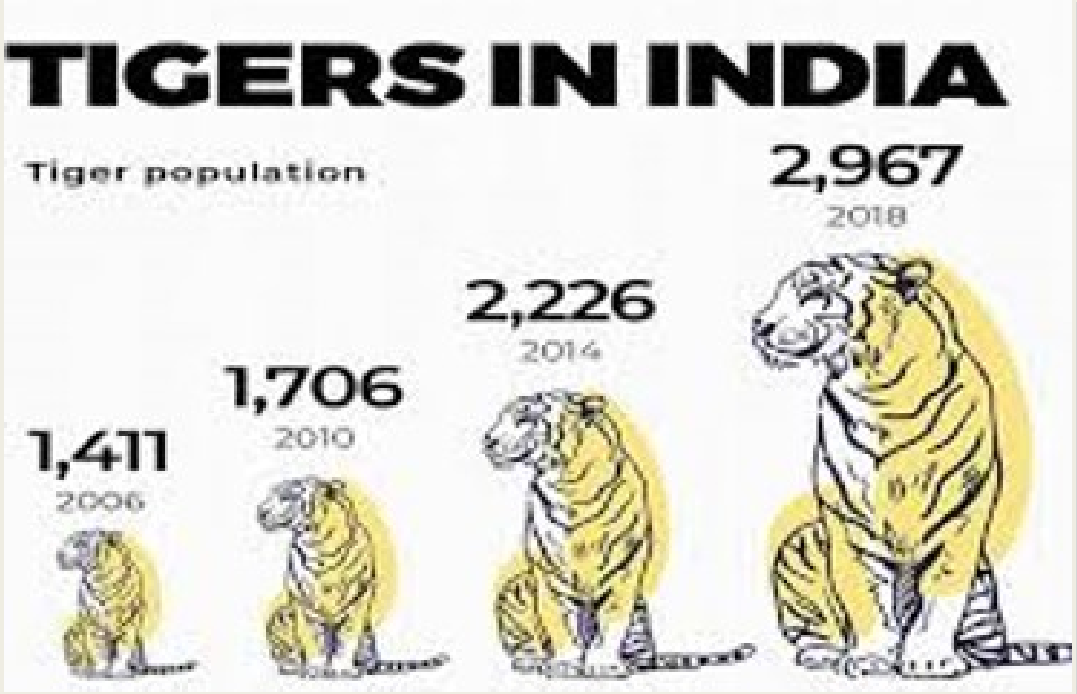


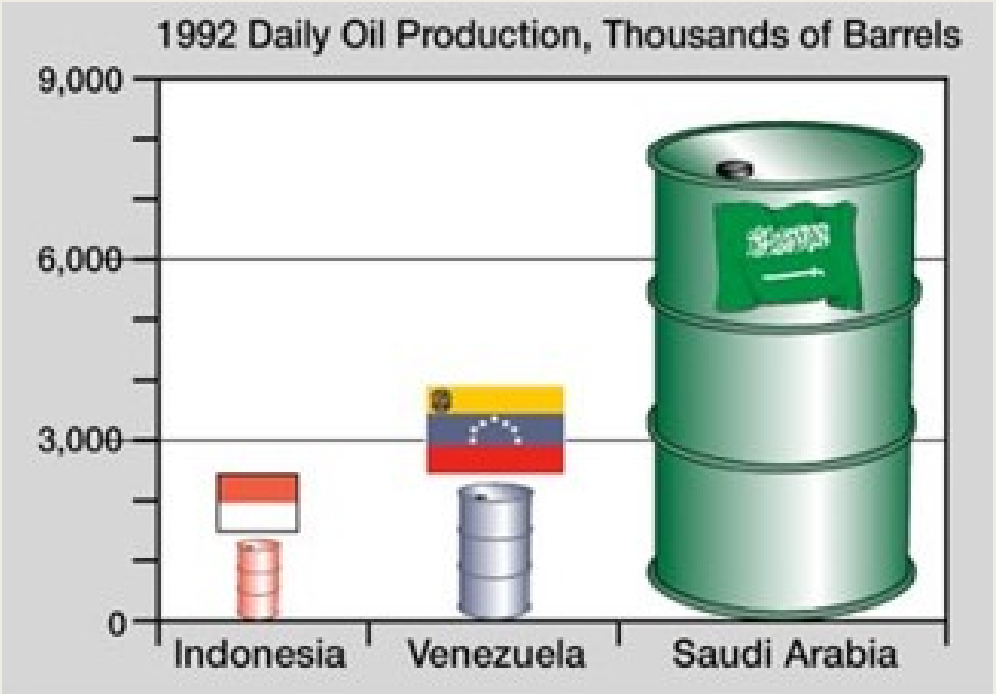
Bar charts - beware of vertical axis...



39.6% never looked so much larger than 35% !!!





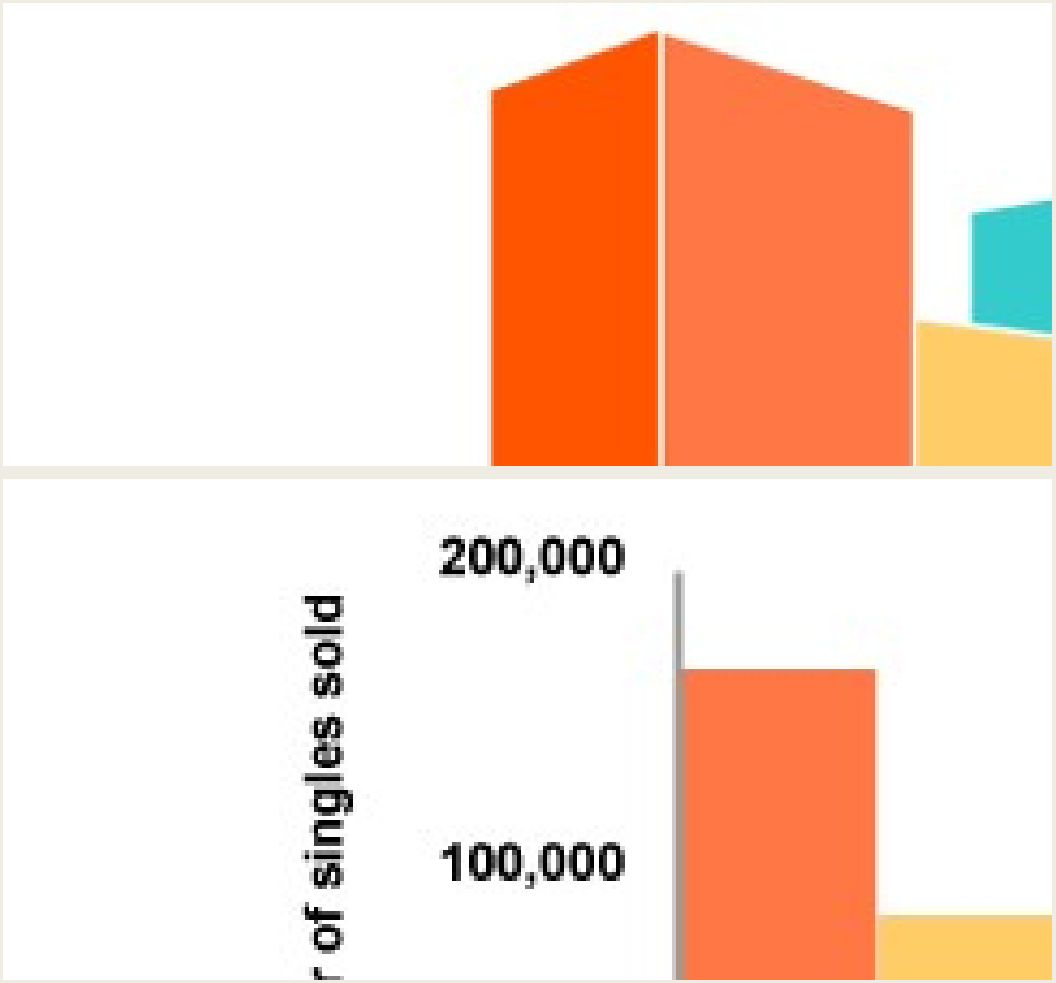


- pictograms are in 2-dimension

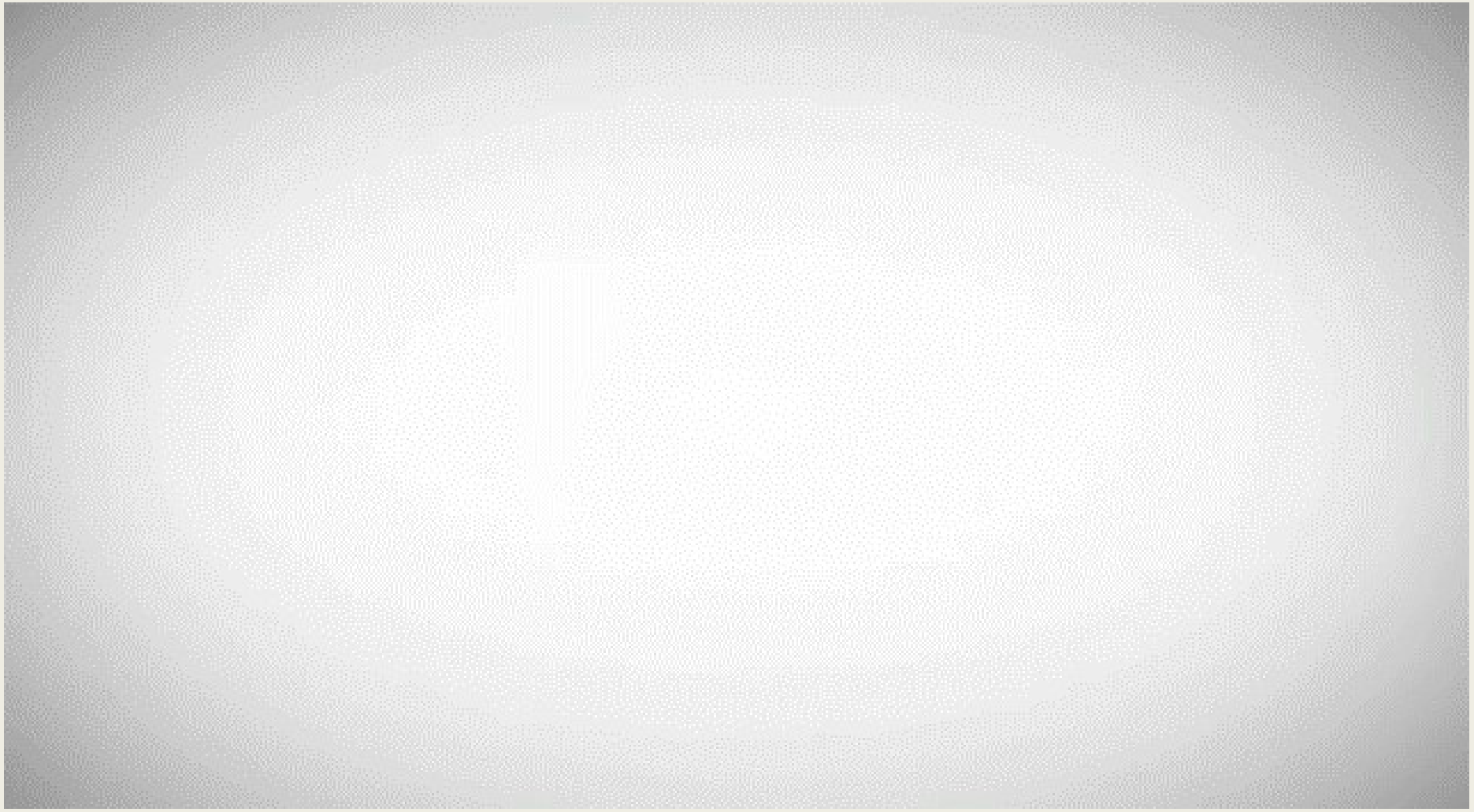
- making the increase/decrease seem much more dramatic.

Avoid pictograms!

Avoid perspective (3D) bar charts - they are misleading



Video Time!



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

- ✓ Make and interpret frequency table, relative frequency table and two-way table
- ✓ Calculate marginal and joint relative frequencies from a two-way table
- ✓ Calculate conditional relative frequencies from a two-way table

Data Analysis: Making Sense of Data

- ✓ DISPLAY categorical data with a bar graph
- ✓ IDENTIFY what makes soe graphs of categorical data deceptive
- ✓ CALCULATE and DISPLAY the marginal distribution of a categorical variable from a two-way table
- ✓ CALCULATE and DISPLAY the conditional distribution of a categorical variable for a particular value of the other categorical variable in a two-way table
- ✓ DESCRIBE the association between two categorical variables

Check your understanding

The American Statistical Association sponsors a web-based project that collects data about primary and secondary school students using surveys. We used the site's "Random Sampler" to choose 40 U.S. high school students who completed the survey in a recent year.
One of the questions asked:
Which would you prefer to be? Select one.
_____Rich _____Happy _____Famous _____Healthy
Here are the responses from the 40 randomly selected students:

Famous	Healthy	Healthy	Famous	Happy	Famous	Happy	Happy	Famous
Rich	Happy	Happy	Rich	Happy	Happy	Happy	Rich	Happy
Famous	Healthy	Rich	Happy	Happy	Rich	Happy	Happy	Rich
Healthy	Happy	Happy	Rich	Happy	Happy	Rich	Happy	Famous
Famous	Happy	Happy	Happy					

Make a relative frequency bar graph of the data.