STAT 108 Lecture 2 Th, Aug 27, 2015

1.5 Cross-Sectional Data Versus Time Series Data

Depending on how data are collected over time, they can be classified as either **cross-sectional** or **time series.**

Definition. **Cross-sectional data** are collected by taking measurements on all sample units at the same point in time.

Definition. **Time series data** (or **repeated measures** or **longitudinal data**) are collected by taking measurements on the same sample units at different time periods.

Examples.

- Data consist of average gas prices for different regions on August 27, 2015 (cross-sectional data).
- Data consist of average monthly gas prices for Long Beach for the past 18 months (time series data).

More examples of time series data

- Stock market behavior over time
- Clinical trials (data collected during follow-up doctor's visits)
- Pollution over time in the port of LA

1.7 Summation Notation

Suppose our observations are

$$x_1 = 5, x_2 = 3, x_3 = 8, x_4 = 0, x_5 = 6$$

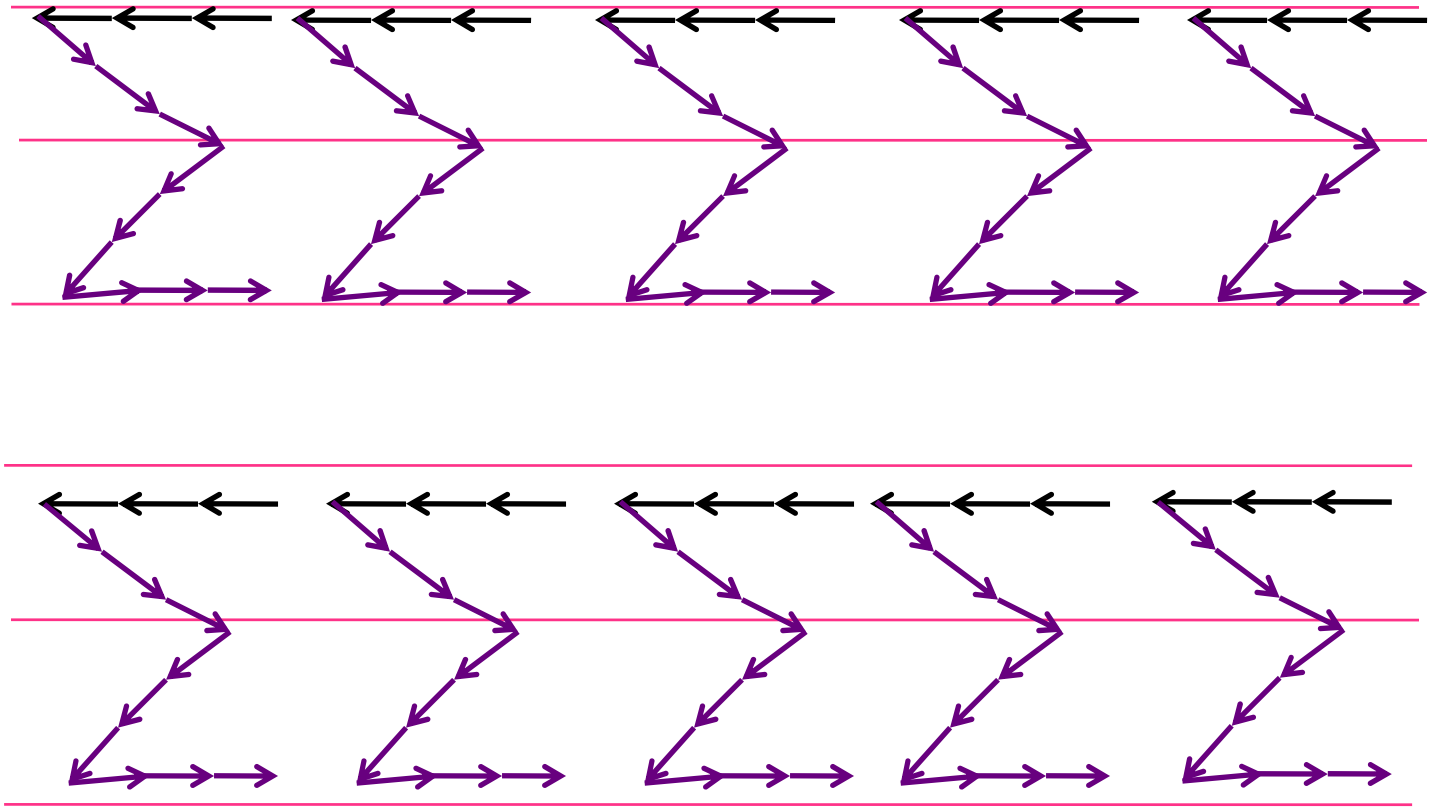
To compute the sum of these values, we write

$$x_1 + x_2 + x_3 + x_4 + x_5$$
$$= 5 + 3 + 8 + 0 + 6 = 22$$

For a large number of observations, to simplify the notation, we use a **summation sign**, that is, we write

$$\sum x = x_1 + x_2 + \cdots + x_n$$

This summation sign is the capital Greek letter "sigma" (pronounced *sig-muh*). It is written as follows:



Examples. The following table lists three pairs of m and f values:

m	2	5	3
f	5	1	3

Compute

- $\sum m = 2 + 5 + 3 = 10$
- $\sum f = 5 + 1 + 3 = 9$
- $\sum mf = 2 \cdot 5 + 5 \cdot 1 + 3 \cdot 3 = 24$
- $(\sum m)(\sum f) = 10 \cdot 9 = 90$
- $(\sum m^2)(\sum f)^2 = (2^2 + 5^2 + 3^2) \cdot 9^2 = 3078$

2.1 Raw Data

Definition. **Raw data** are observations perhaps in the sequence in which they were collected and before they are processed in any way.

Definition. Data that are not raw, may be referred to as **processed data**.

Examples.

- Raw data – ages of 5 students are recorded. They are 21 19 24 25 20
- Ordered data – the ages are put in increasing order: 19 20 21 24 25
- Summarized data – min, max and the arithmetic average of the ages are computed: min = 19, max = 25, average = (19+20+21+24+25)/5=21.8

2.1 Organizing and Graphing Qualitative Data

Recall that a qualitative variable has values that fall into one of several categories.

Definition. **Qualitative data** are observations of a qualitative variable recorded for each observational unit.

Definition. A **frequency of a category** is the number of observations that fall into that category.

Definition. A **frequency distribution** is the list of all categories and their respective frequencies.

Example. Type of employment students intend to engage in after graduation and respective frequencies are given in the table below.

Category	Frequency
Industry	44
Government	16
Own business	20
Academia	5

This table presents the frequency distribution.

Definition. A **relative frequency of a category** is the frequency of that category divided by the sum of frequencies in all categories.

Definition. A **relative frequency distribution** is the list of all categories and their respective relative frequencies.

Example.

Category	Relative Frequency
Industry	44/(44+16+20+5)=44/85=0.518
Government	16/(44+16+20+5)=16/85=0.188
Own business	20/(44+16+20+5)=20/85=0.235
Academia	5/(44+16+20+5) = 5/85 =0.059

This table presents the relative frequency distribution.

Note that relative frequencies add up to one (up to a round-off error).

Definition. A **percentage of a category** is the relative frequency multiplied by 100%.

Definition. A **percentage distribution** is the list of all categories and their respective percentages.

Example.

Category	Percentage
Industry	(0.518)(100%)=51.8%
Government	(0.188)(100%)=18.8%
Own business	(0.235)(100%)=23.5%
Academia	(0.059)(100%)= 5.9%

This table presents the percentage distribution.

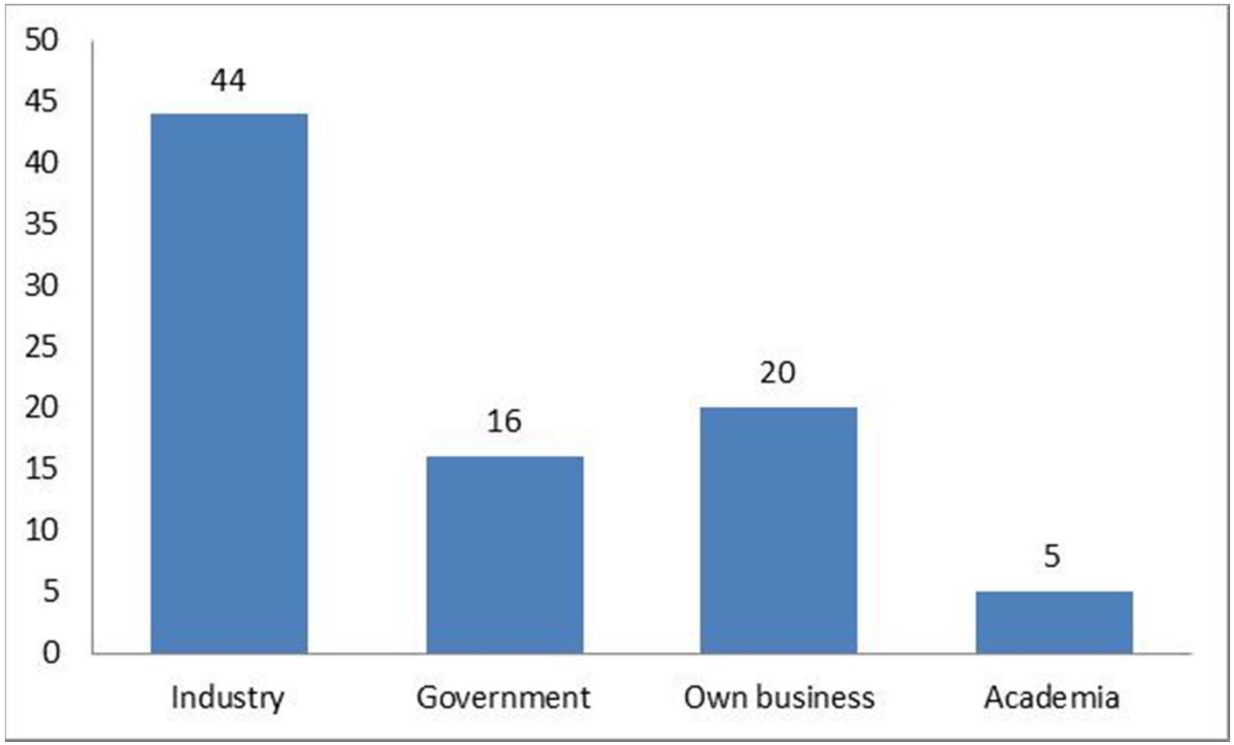
Note that percentages add up to 100% (up to a round-off error).

Graphical Presentation of Qualitative Data

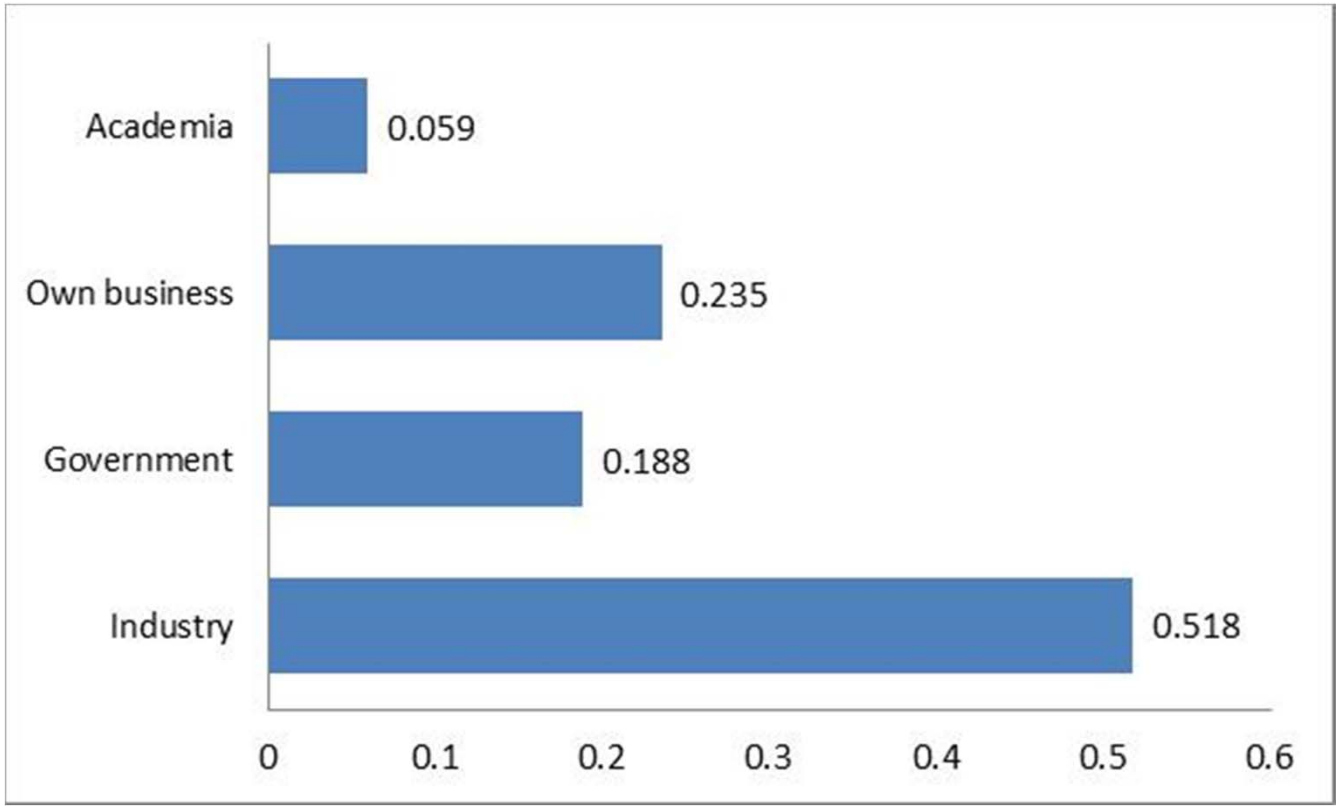
Two types of graphs are used to display qualitative data. They are a bar graph and a pie chart.

Definition. A **bar graph** is a graph made of vertical (or sometimes horizontal) bars which heights represent the frequencies (or relative frequencies or percentages) of the respective categories.

Example. Here is a vertical bar graph of frequencies for our data.



Here is a horizontal bar graph of relative frequencies for our data.



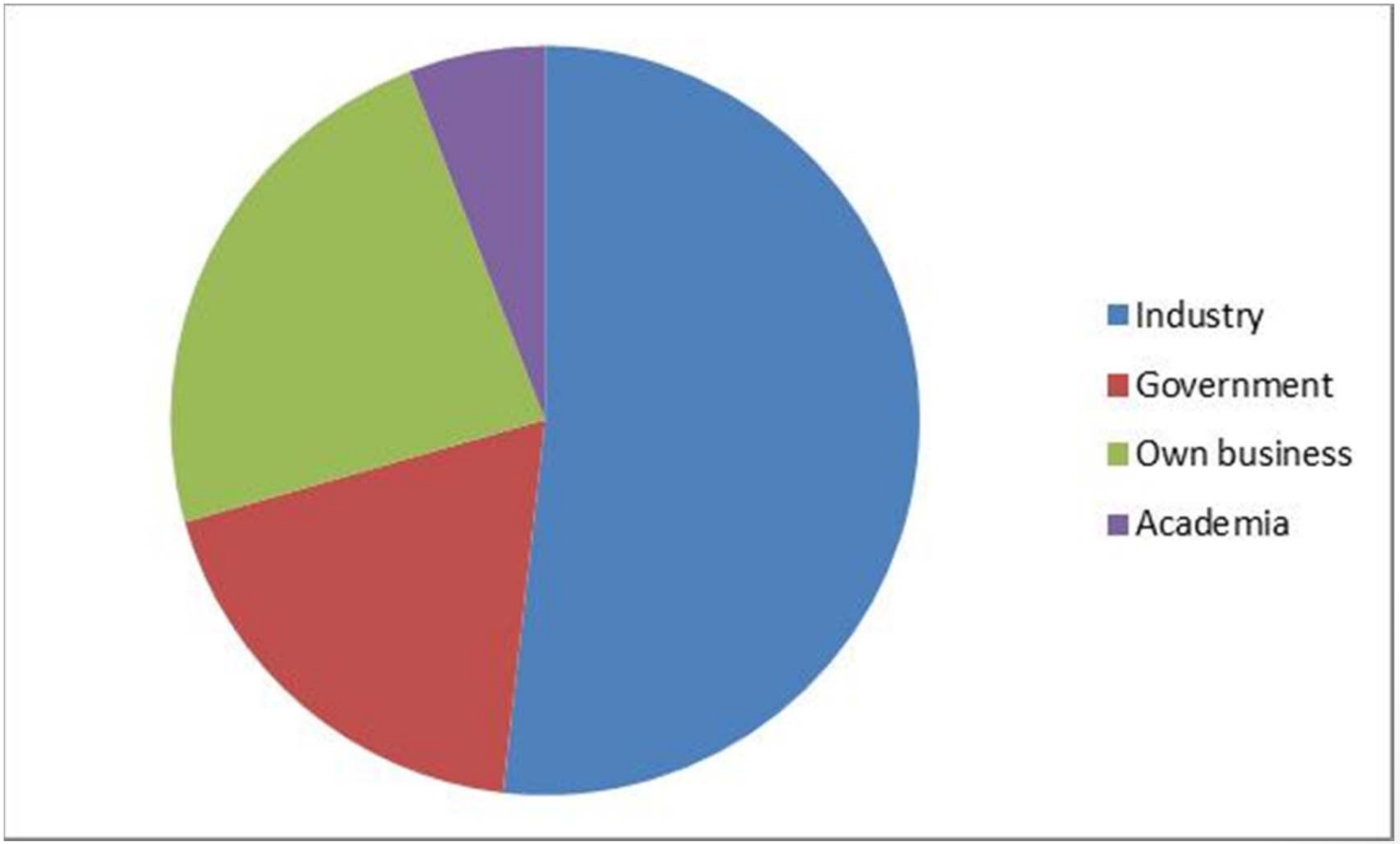
Definition. A **pie chart** is a circle divided into portions that represent the relative frequencies (or sometimes percentages) for each category.

Since a complete circle is 360 degrees, to find an exact angle of a certain portion, we have to multiply the relative frequency by 360.

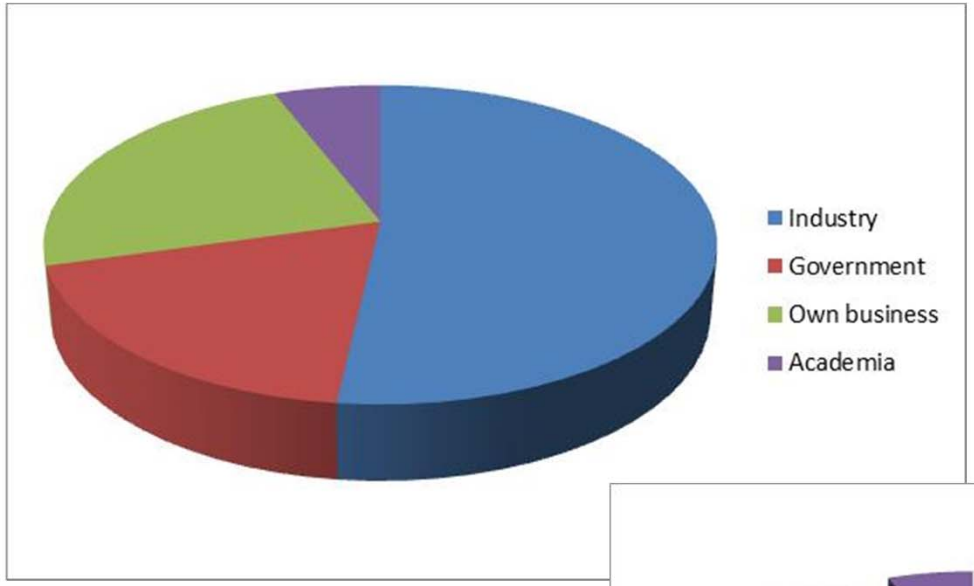
Example.

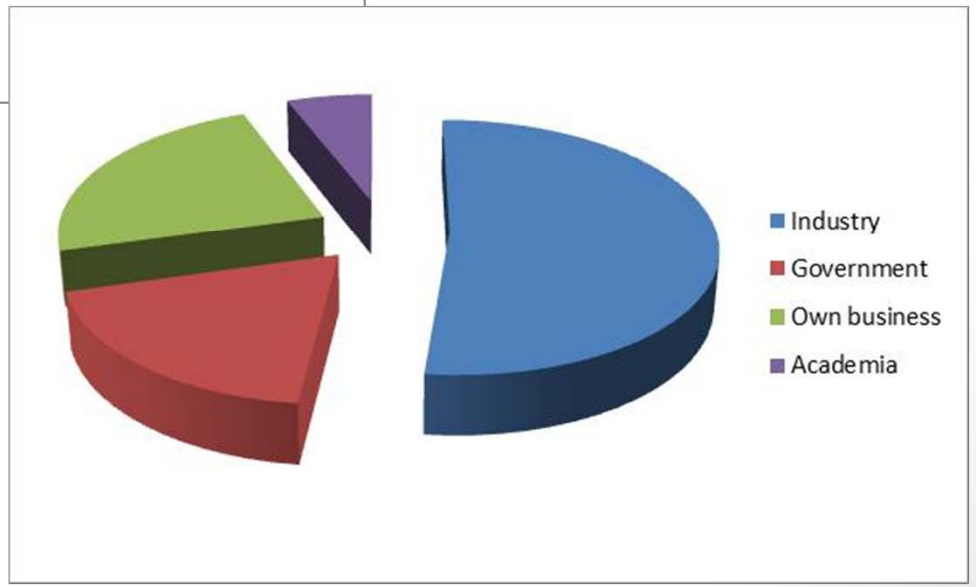
Category	Relative Frequency	Angle (in degrees)
Industry	0.518	(0.518)(360)=186.48
Government	0.188	(0.188)(360)= 67.68
Own business	0.235	(0.235)(360)= 84.60
Academia	0.059	(0.059)(360)= 21.24

Here is a two-dimensional pie chart for our data.



Here are three-dimensional pie charts for our data.





"Let's break for lunch. I'll order pizza."


This pie chart shows how much pie I ate while making this chart.
BIZARRO.COM
DIST. BY KING FEATURES
©DAN PIRARO
12-1-04


Okay - wait. What were we studying, again?
Other
Don't Know
search ID: bsm474
© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com