

Population – is the set representing all measurements of interest to the researcher.

Parameter – Numerical descriptive measure of a population.

Sample – is a subset of measurements selected from the population of interest.

Statistic – Numerical descriptive measure of the sample.

Branches of Statistics

Descriptive statistics – a branch of statistics concerned with numerical and graphical techniques for **describing** one or more characteristics of a group and comparing characteristics between groups.

Inferential statistics – a branch of statistics concerned with the acquisition of data, with sampling, and with the use of the data in making **inferences (conclusion)** about a population.

Type of Observations

Constant – occurs when the phenomenon or value of a sample or population will **not change** after successive trials. Examples: tribe of a Subanen family,

Variable – occurs when the phenomenon or value of a sample or population will **change** after successive trials. Examples: tribes of people living in Ozamiz City, family income, age, birth rate.

Types of Variable

Quantitative variable – one whose observation vary in magnitude from trial to trial. (greatness of size, degree, quantity). Examples: family income, population size, age.

- **Discrete quantitative variable** – a quantitative variable that can assume only a countable number of values. Examples: number of children for each household, number of deaths for each year due to lung cancer.
- **Continuous quantitative variable** - a quantitative variable can assume any one of the countless number of values in a line interval. Example: height of a person, weight of a lady, temperature of a certain room, time.

Qualitative variable - one whose observation vary in kind (a group united by common traits or interest) but not in degree. Example: Sex, Religion, Marital Status, Political Party.

Scales(Levels) of Measurement (Data)

Nominal Scale – defines specific categories by name.

These categories are called **levels** of the scale. **All qualitative variables are measured on a nominal scale.** Example: Political Party, Sex and others.

Ordinal Scale – incorporates the features of a nominal scale and additional feature that observations can be **ordered or ranked** from low to high. Example: categorize gross annual income, prestige ranking of occupations.

Interval Scale – incorporates all features of an ordinal (and hence nominal) scale and the additional feature that we can specify **distances** between levels on the scale. Example: IQ

Ratio Scale – incorporates all the features of interval (and hence nominal and ordinal) scales and the additional feature that ratios can be formed with levels of the scale, and there is a meaningful zero point for the scale. Example: Temperature, birth rate

Type of Numerical Descriptive Measures

1. Measures of Central Tendency
2. Measures of Variation (Dispersion)



Describing Data: Frequency Distributions and Graphic Presentation

Frequency Distribution

- A Frequency Distribution is a grouping of data into *mutually exclusive* categories showing the number of observations in each class.

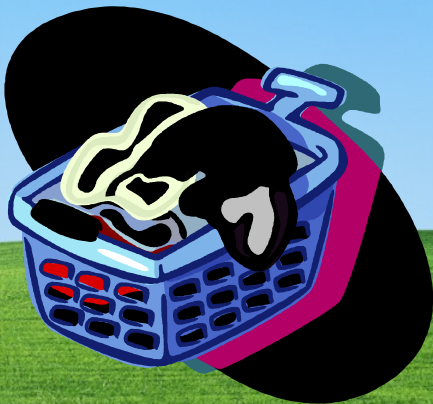
-(explanation) you are just developing categories or classes based on a characteristic and then putting elements into categories based on *that* characteristic. No element appears in more than one class.

Frequency Distribution

-Here's an analogy...We divide clothes to wash into what three categories or classification before we wash???!!!

Whites, Lights/Colors, and Darks, right?

-The freq. dist. of clothes is developed by counting how many articles of clothing are in each laundry bin.



Rule of Thumb for Developing a Frequency Dist.

- Step 1: Decide on the number of classes (k) or containers...
hint: must be more than 1, but less than a million. $2^k > n$,
where k=number of classes ,n=number of obs.
 - If obs=50, $2^6=64>50$ so we should use at least 6 classes.
- Step 2: Determine the class interval width (i)
 - Should be the same for all classes, and
 - Cover lowest (L) to highest (H) observation value
 - $i \geq (H-L)/k$

[This is a rule of thumb folks typically round up to the next convenient number for i, e.g., 8.9 becomes 10 and 94 becomes 100.]
- Step 3: Set the individual class limits. Don't overlap at all. E.g., dollars...classes like \$50-59, \$60-69, and so on. Don't have \$50-60, \$60-70...if something is \$60 is will appear in two classes. Note: if you say \$50 up to \$60, \$60 is NOT included.

Rule of Thumb for Developing a Frequency Dist.

- Step 4: Tally the items into correct classes.

Step 5: Count the number of items in each class.

- Now you can graphically depict the counts with a “histogram.”

Example: Hudson Auto Repair

The manager of Hudson Auto would like to have a better understanding of the cost of parts used in the engine tune-ups performed in the shop. She examines 50 customer invoices for tune-ups. The costs of parts, rounded to the nearest dollar, are listed on the next slide.



Example: Hudson Auto Repair



■ Sample of Parts Cost for 50 Tune-ups

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

- Based on the “rule of thumb,” how many classes might we use? $2^k > n$, where n is 50...

$2^6 = 64$ which is juuuust greater than 50.

- Based on the “rule of thumb,” what should the width of the classes be? $i \geq (H-L)/k$...

$(109-52)/6 = 9.5$...Let's round up to 10 to make it easy, and let's start the classes at 50 (just lower than the lowest observation)

Tabular Summary: Frequency and Relative (or Percent) Frequency

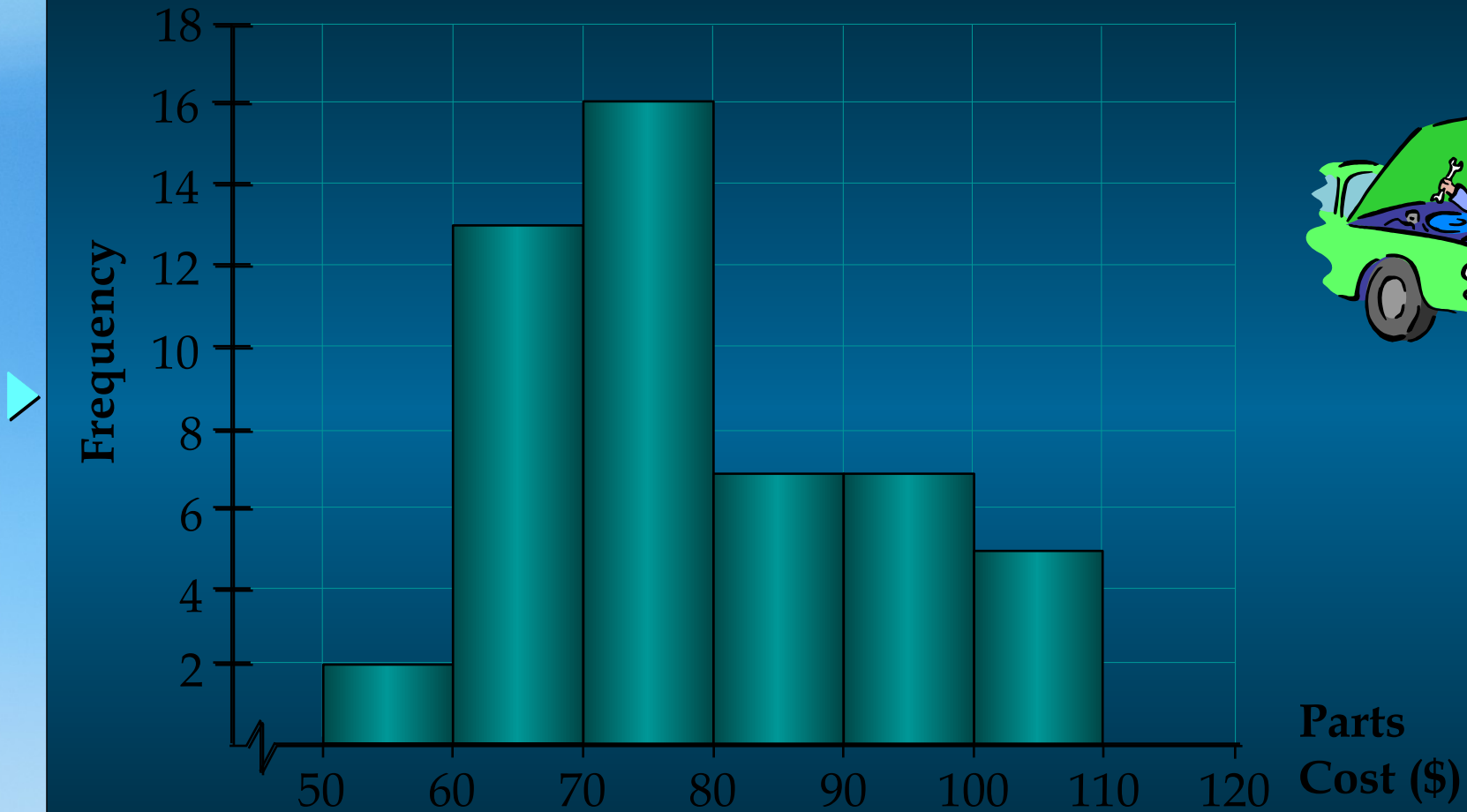
<u>Parts Cost (\$)</u>	<u>Parts Frequency</u>	<u>Relative Frequency(%)</u>
50-59	2	4
60-69	13	26
70-79	16	32
80-89	7	14
90-99	7	14
100-109	<u>5</u>	<u>10</u>
	50	100



$(2/50)100$

Graphical Summary: Histogram

Tune-up Parts Cost



Other Graphical Depictions of Data

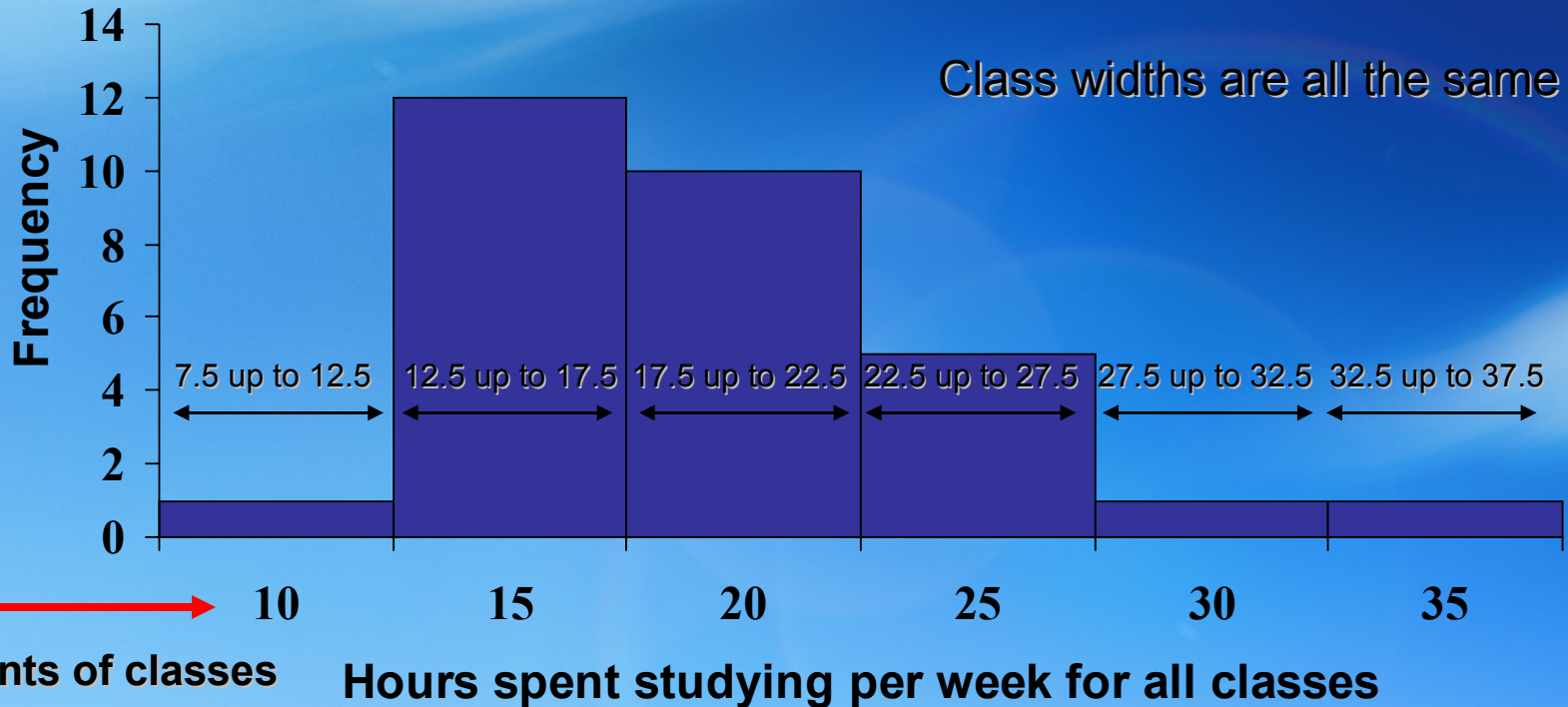
- Pie Chart-for Relative Frequencies and Shares of the Whole
- Line Graphs-for changes over time, trends, or differences between groups
- Bar Charts-Similar to line graphs in their uses. Sometimes they make for better pair-wise comparisons.

The three commonly used graphic forms are **Histograms, Frequency Polygons**, and a **Cumulative Frequency** distribution.

A **Histogram** is a graph in which the class midpoints or limits are marked on the horizontal axis and the class frequencies on the vertical axis.

The class frequencies are represented by the heights of the bars and the bars are drawn adjacent to each other.

Example: Histogram for Hours Spent Studying



How do you read this graphic?

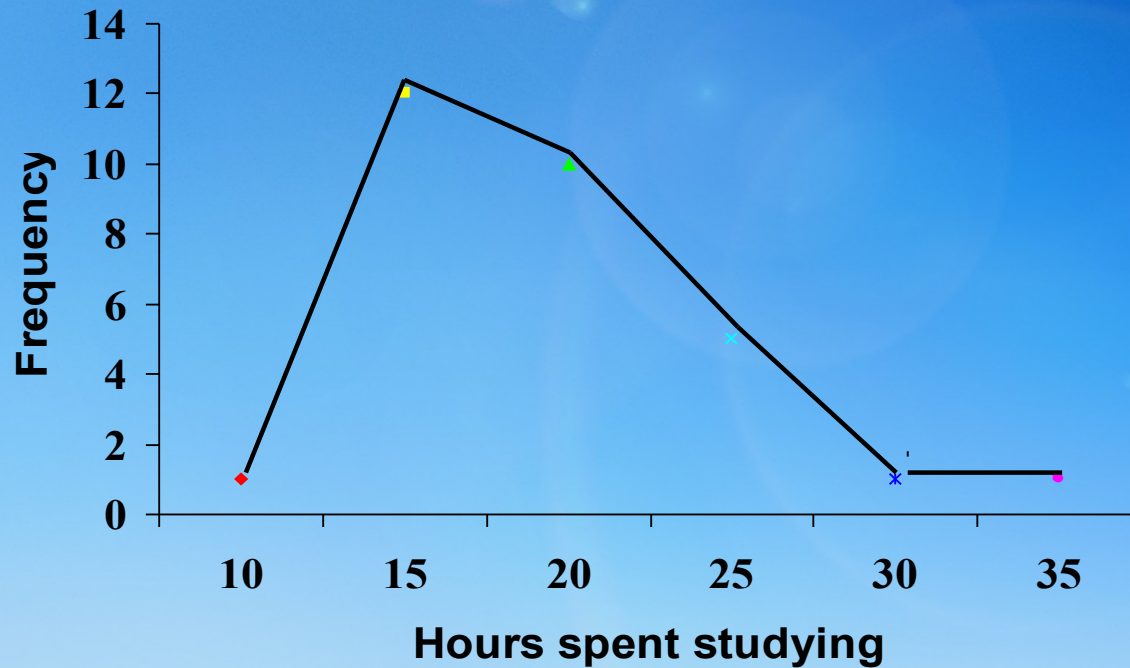
How many people study around 20 hours per week?

How many study less than 32.5 hours per week?

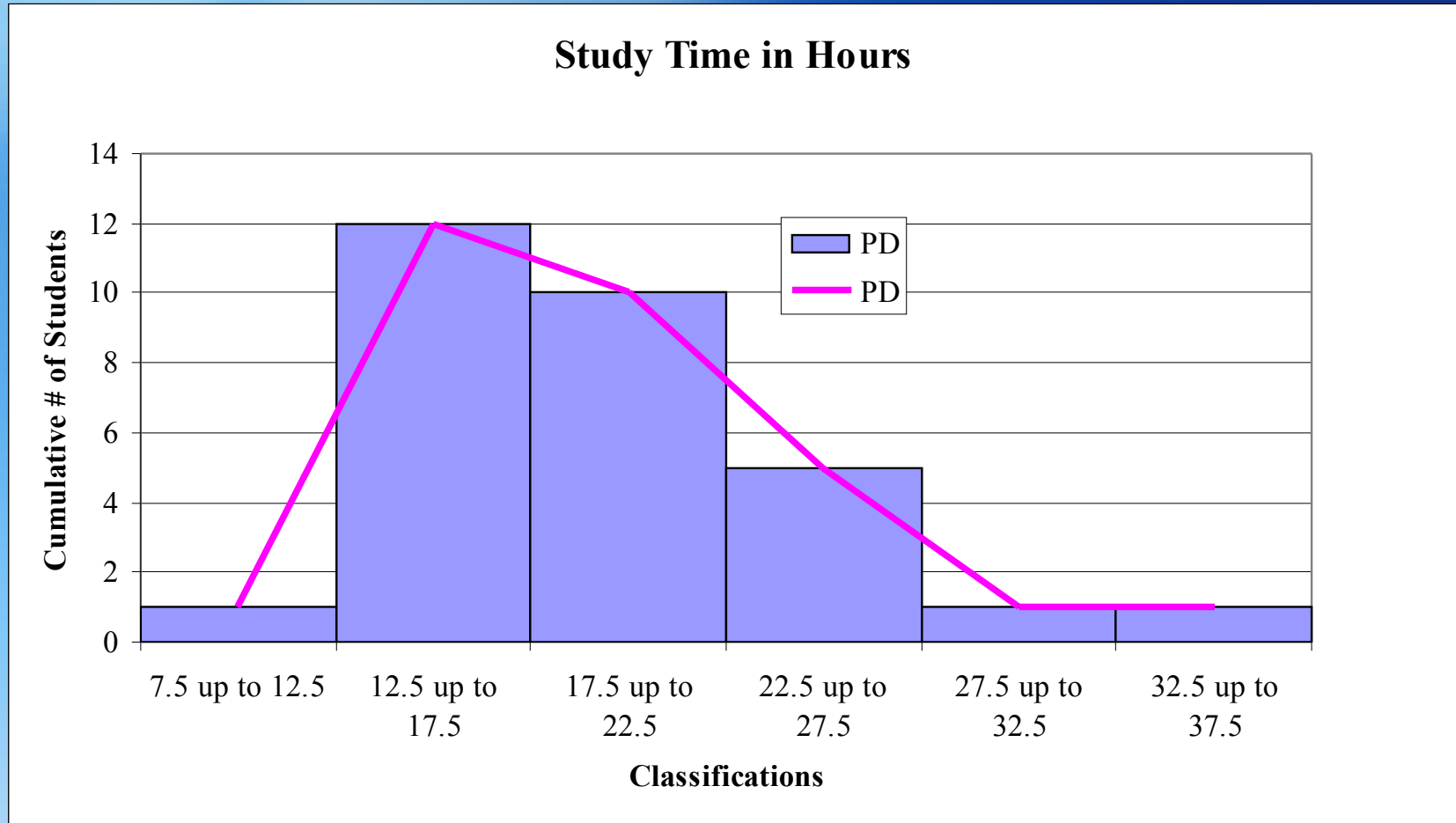
Graphic Presentation of a Frequency Distribution

A **Frequency Polygon** consists of line segments connecting the points formed by the class midpoint and the class frequency.

Frequency Polygon for Hours Spent Studying



Both on the same Chart



Cumulative Frequency Distribution

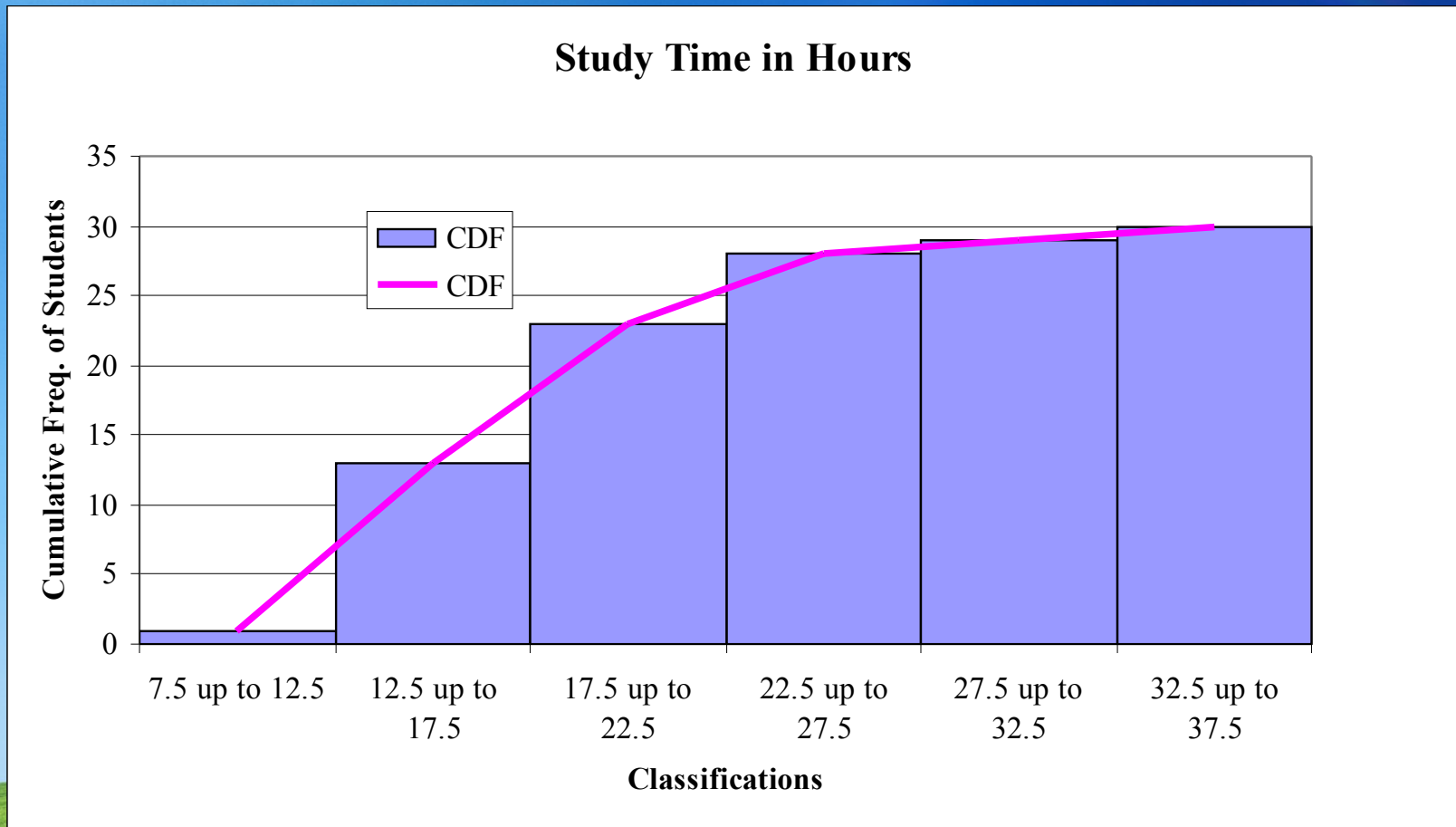
A Cumulative Frequency Distribution is used to determine how many or what proportion of the data values are below or above a certain value.

You are just adding up as you go along...

Cumulative Frequency Table for Hours Spent Studying

Hours Studying	Upper Limit	Freq	Cumulative Frequency
7.5 up to 12.5	12.5	1	1
12.5 up to 17.5	17.5	12	13 (1+12)
17.5 up to 22.5	22.5	10	23 (13+10)
22.5 up to 27.5	27.5	5	28 (23+5)
27.5 up to 32.5	32.5	1	29 (28+1)
32.5 up to 37.5	37.5	1	30 (29+1)

Cumulative Frequency Distribution For Hours Studying

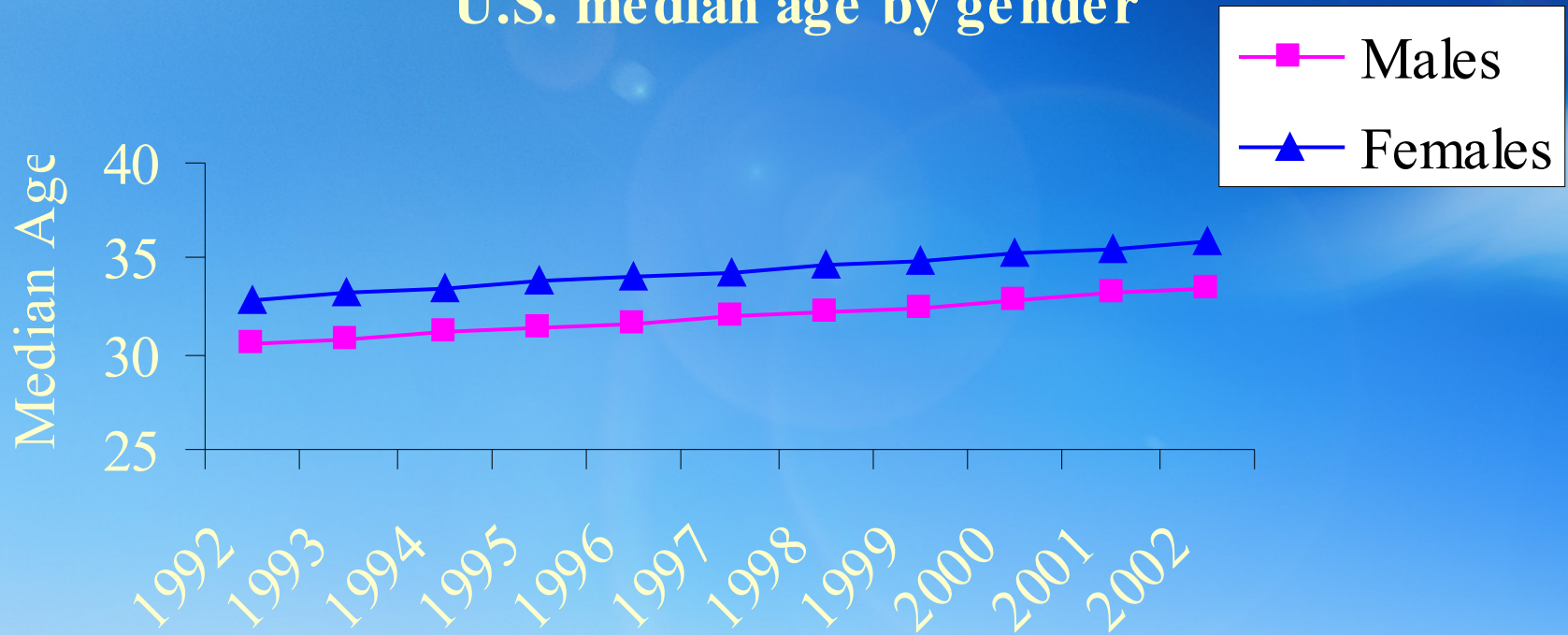


Line graphs are typically used to show the change or trend in a variable over time.

Year	Males	Females
1992	30.5	32.9
1993	30.8	33.2
1994	31.1	33.5
1995	31.4	33.8
1996	31.6	34.0
1997	31.9	34.3
1998	32.2	34.6
1999	32.5	34.9
2000	32.8	35.2
2001	33.2	35.5
2002	33.5	35.8

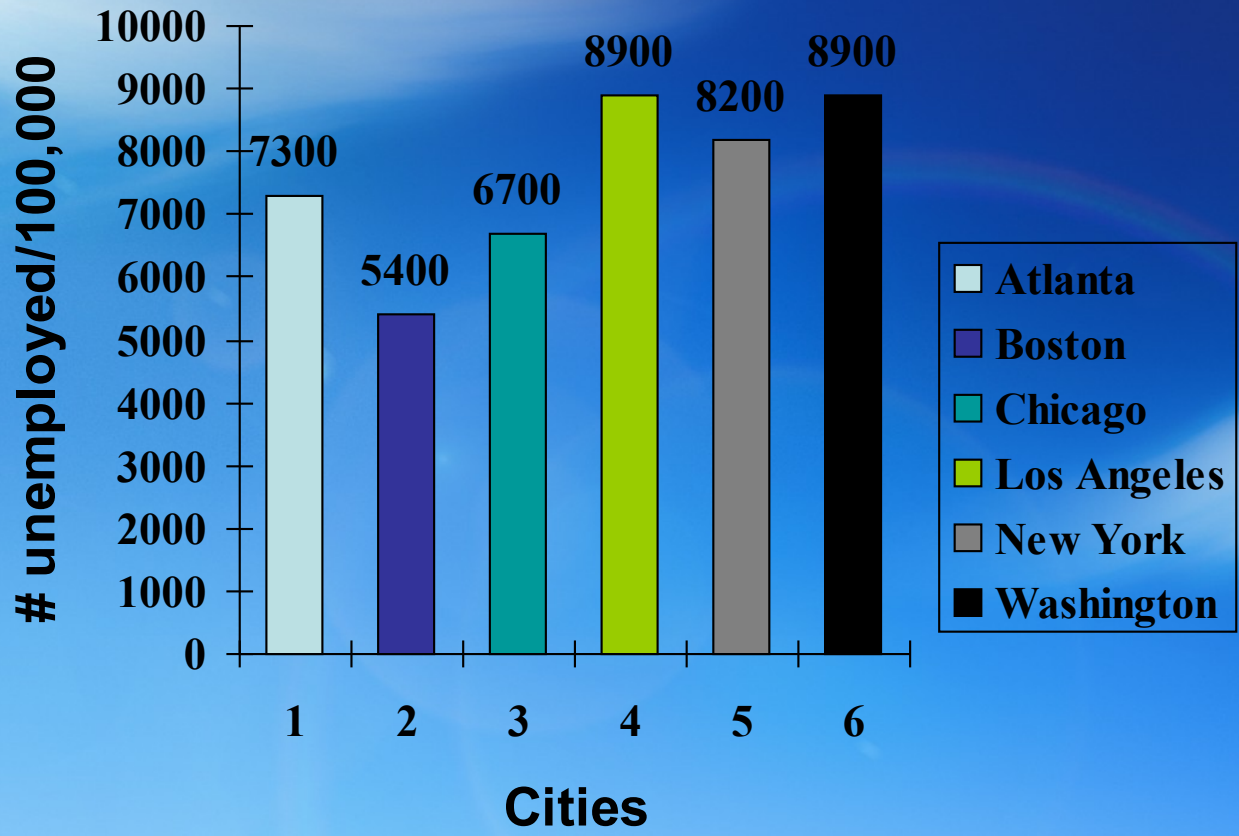
Example 3 continued

U.S. median age by gender



A **Bar Chart** can be used to depict any of the levels of measurement (nominal, ordinal, interval, or ratio).

Construct a bar chart for the number of unemployed per 100,000 population for selected cities during 2001



Bar Chart for the Unemployment Data

A **Pie Chart** is useful for displaying a relative frequency distribution. A circle is divided proportionally to the relative frequency and portions of the circle are allocated for the different groups.

A sample of 200 runners were asked to indicate their favorite type of running shoe. Draw a pie chart based on the following information.

Pie Chart for Running Shoes

