

Exploring Data

1.1 Displaying Distributions with Graphs

YMS3e

AP Stats at LSHS

Mr. Molesky

Case Study

- Nielsen Ratings
 - Read the study on page 37.
 - What do you observe? Does one network appear to “win” the ratings race?
 - How can we get a better sense of which network has the best ratings?
 - How can Statistics help us understand this data?

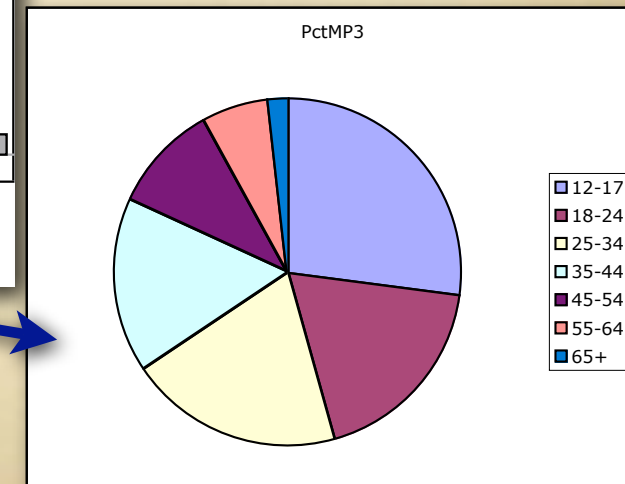
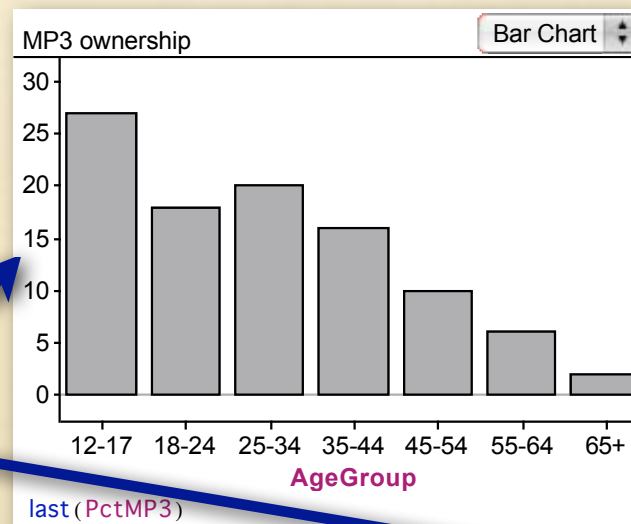
Exploratory Data Analysis

- **Exploratory Data Analysis:**
 - Statistical practice of analyzing distributions of data through graphical displays and numerical summaries.
- **Distribution:**
 - Description of the values a variable takes on and how often the variable takes on those values.
- An EDA allows us to identify patterns and departures from patterns in distributions.

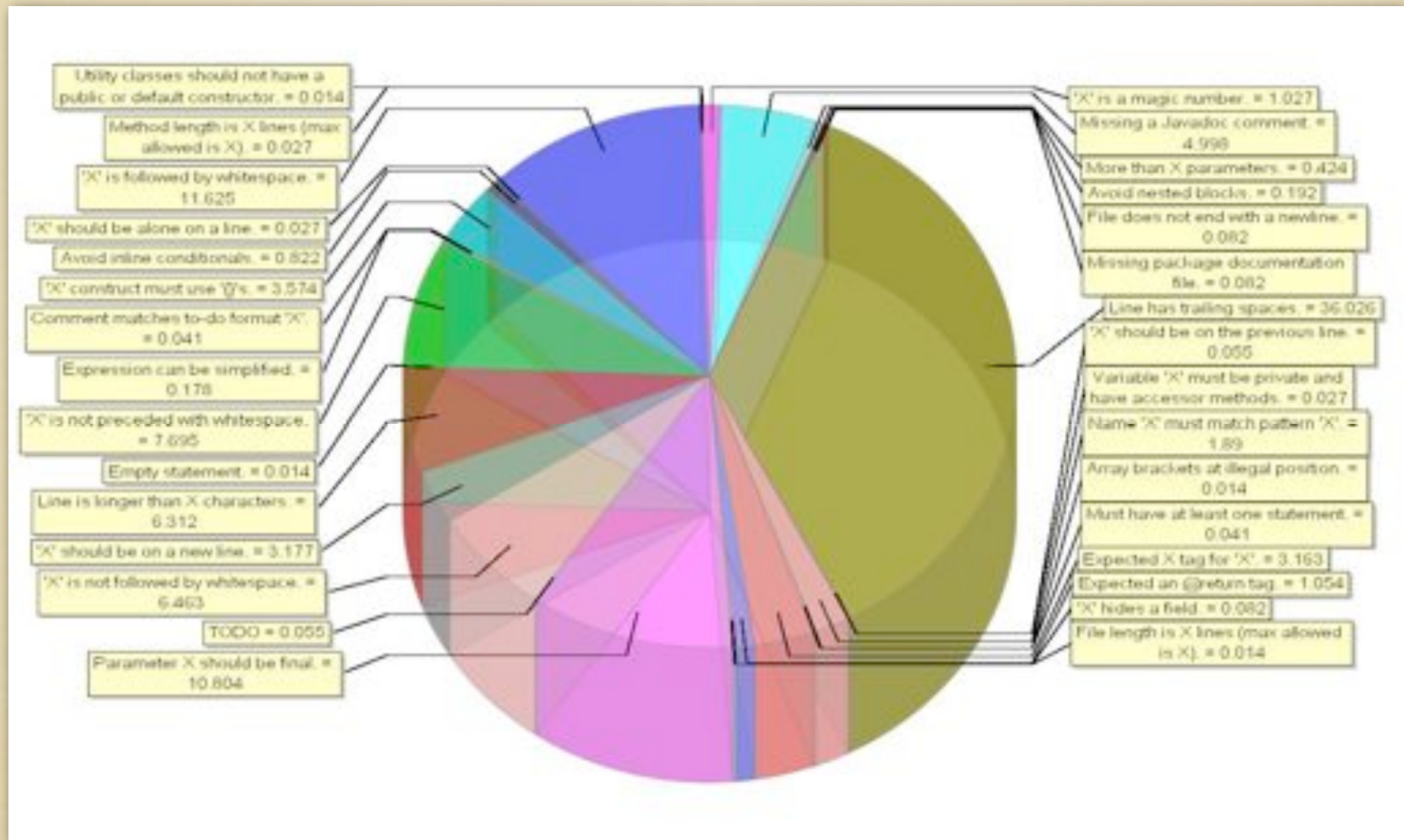
Categorical Data

- **Categorical Variable:**
 - Values are labels or categories.
 - Distributions list the categories and either the count or percent of individuals in each.
- **Displays: BarGraphs and PieCharts**

	AgeGroup	PctMP3
1	12-17	27
2	18-24	18
3	25-34	20
4	35-44	16
5	45-54	10
6	55-64	6
7	65+	2



Beware of Bad Graphs!



What is wrong with this graph? Let us count the ways...

Quantitative Data

- **Quantitative Variable:**
 - Values are numeric - arithmetic computation makes sense (average, etc.)
 - Distributions list the values and number of times the variable takes on that value.

- **Displays:**
 - **Dotplots**
 - **Stemplots**
 - **Histograms**
 - **Boxplots**

**Only organized Data can
Illuminate!**

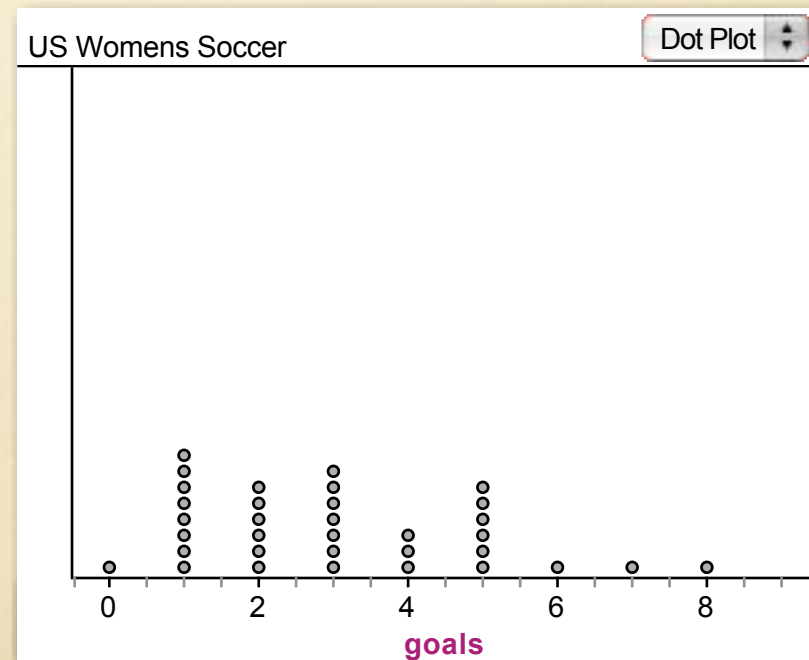
**Your goal is to make neat,
organized, labeled graphs that
display the distribution of
data effectively and provide
an insight into patterns and
departures from patterns.**

Dotplots

- Small datasets with a small range (max-min) can be easily displayed using a **dotplot**.
- ☑ Draw and label a number line from min to max.
- ☑ Place one dot per observation above its value.
- ☑ Stack multiple observations evenly.

US Womens Soccer	
	goals
1	3
2	0
3	2
4	7
5	8
6	2
7	4
8	3
9	5
10	1
11	1
12	4
13	5
14	2

34 values
ranging from
0 to 8.



Stemplots

- A **stemplot** gives a quick picture of the shape of a distribution while including the numerical values.
 - ☑ Separate each observation into a **stem** and a **leaf**.
 - ☑ eg. 14g \rightarrow 1 | 4 256 \rightarrow 25 | 6 32.9oz \rightarrow 32 | 9
 - ☑ Write stems in a vertical column and draw a vertical line to the right of the column.
 - ☑ Write each leaf to the right of its stem.

Stemplots

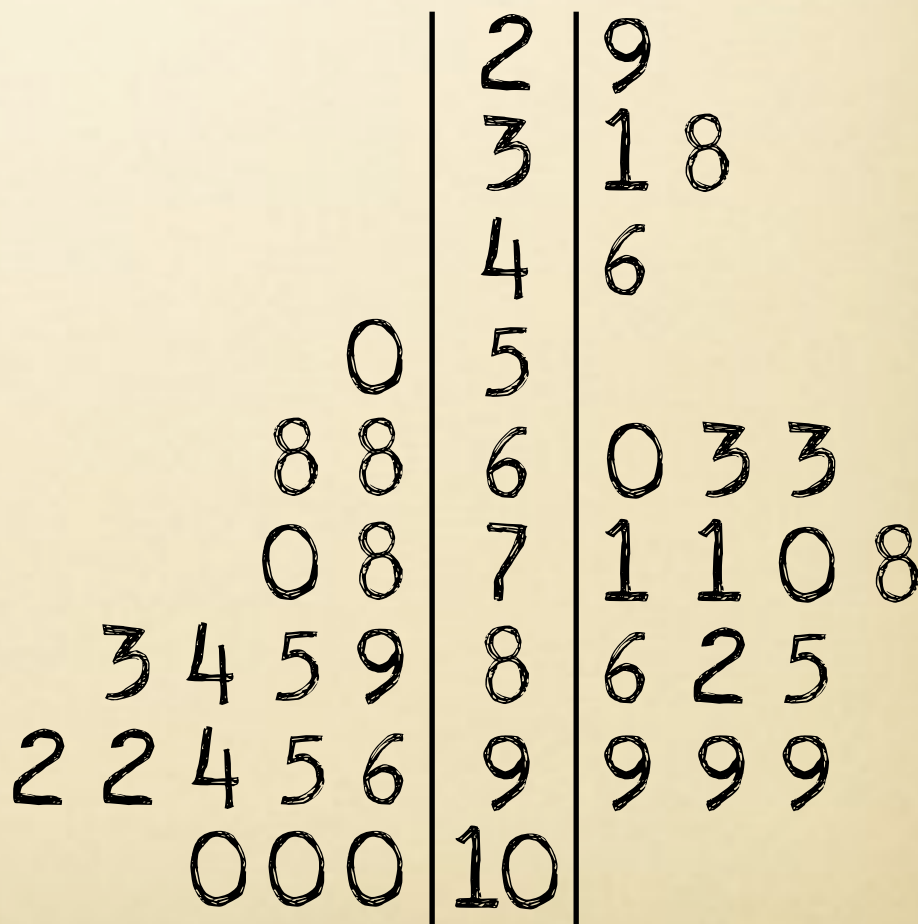
- Example 1.4, pages 42-43
- Literacy Rates in Islamic Nations

Literacy			
	Country	FemPct	MalePct
1	Algeria	60	78
2	Bangladesh	31	50
3	Egypt	46	68
4	Iran	71	85
5	Jordan	86	96
6	Kazakhstan	99	100
7	Lebanon	82	95
8	Libya	71	92
9	Malaysia	85	92
10	Morocco	38	68
11	Saudi Arabia	70	84
12	Syria	63	89
13	Tajikistan	99	100
14	Tunisia	63	83
15	Turkey	78	94
16	Uzbekistan	99	100
17	Yemen	29	70



Male %

Female %



Key: 2|9=29%

Stemplots

- Note: Stemplots do not work well for large data sets
- **Back-to-Back Stemplots:** Compare datasets
- **Splitting Stems:** Double the number of stems, writing 0-4 after the first and 5-9 after second.

2		0112234889
3		1223678
4		06779
5		125
6		01798



2		0112234
2		889
3		1223
3		678
4		0
4		6779
5		12
5		5
6		01
6		798

Histograms

- **Histograms** break the range of data values into classes and displays the count / % of observations that fall into that class.
 - ☑ Divide the range of data into equal-width classes.
 - ☑ Count the observations in each class - “**frequency**”
 - ☑ Draw bars to represent classes - height = frequency
 - ☑ Bars should touch (unlike bar graphs).

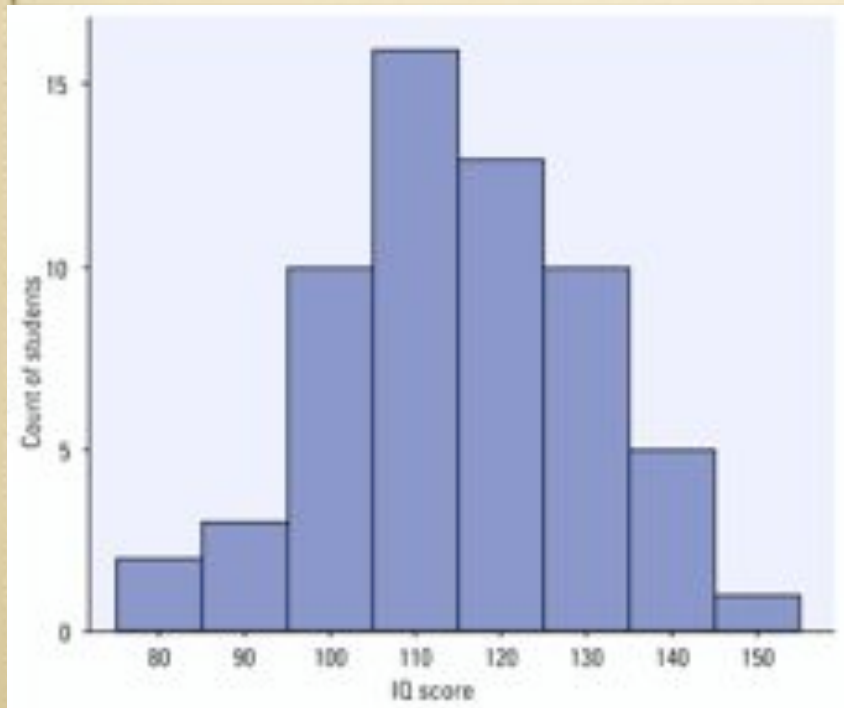
Histograms

Table 1.3 IQ test scores for 60 randomly chosen fifth-grade students

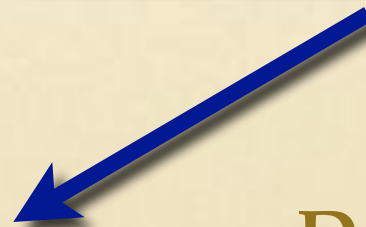
145	139	126	122	125	130	96	110	118	118
101	142	134	124	112	109	134	113	81	113
123	94	100	136	109	131	117	110	127	124
106	124	115	133	116	102	127	117	109	137
117	90	103	114	139	101	122	105	97	89
102	108	110	128	114	112	114	102	82	101

Source: James T. Fleming, "The measurement of children's perception of difficulty in reading materials," *Research in the Teaching of English*, 1 (1967), pp. 136–156.

- Example 1.6, page 49
- IQ Scores for 5th Graders



Class	Count	Class	Count
75 to 84	2	115 to 124	13
85 to 94	3	125 to 134	10
95 to 104	10	135 to 144	5
105 to 114	16	145 to 154	1



- Describe the SOCS
- What do these data suggest?

EDA Summary

- The purpose of an Exploratory Data Analysis is to organize data and identify patterns / departures.

 **PLOT YOUR DATA** - Choose an appropriate graph

 Look for overall pattern and departures from pattern

 **Shape** *{mound, bimodal, skewed, uniform}*

 **Outliers** *{points clearly away from body of data}*

 **Center** *{What number “typifies” the data?}*

 **Spread** *{How “variable” are the data values?}*