Class 2

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Agenda:

Briefly Review Math

Briefly Review Chapter 1

Lecture Chapter 2.1-2.4

Review Math

1. Summation Notation

$$\sum_{i=1}^{n} f(x_i) = f(x_1) + f(x_2) + \ldots + f(x_n)$$

2. Factorials

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

3. Computations

x=20, y=14, s=16, w=-2, m=15, n=10

Compute $x + y \cdot \dfrac{\sqrt{s}}{n} = 25.6$

4. Simple Linear Equations

$2 - 2x = 3x + 3 \qquad x = -1/5$

Recap Chapter 1

Chapter 1: Statistics

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

1: Statistics

1.1 Americans Here's Looking at you

Statistics is all around us!

How much time between Internet usage?

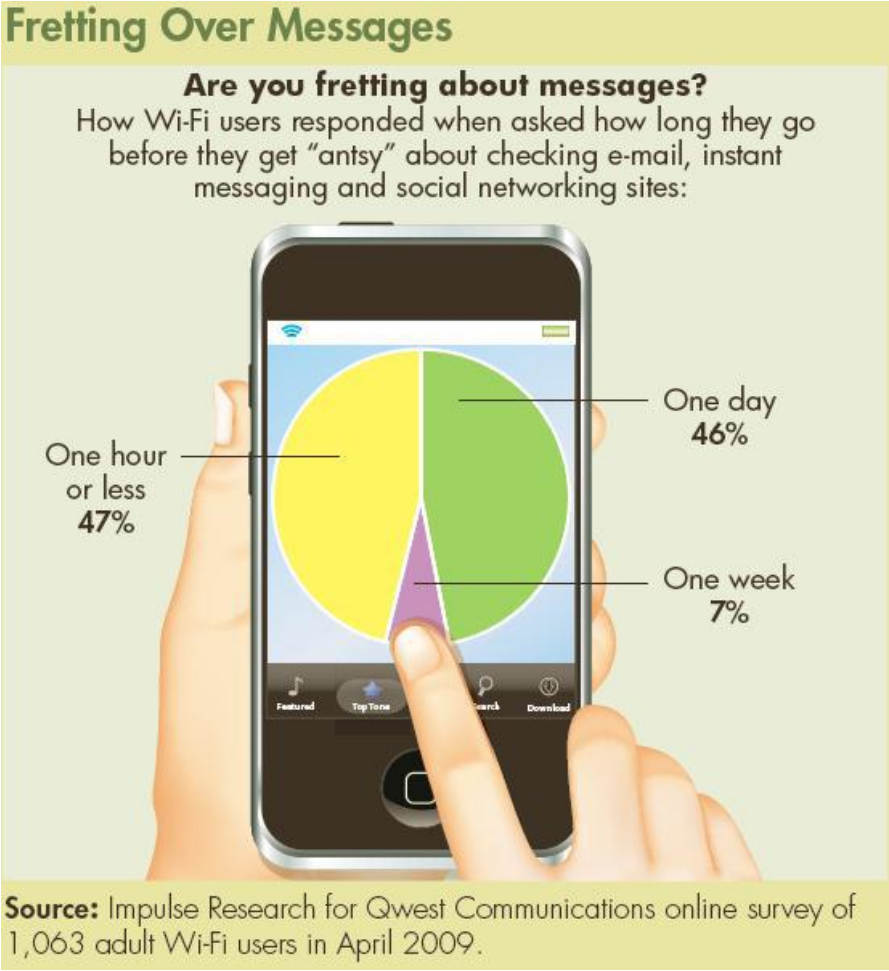


Figure from Johnson & Kuby, 2012.

1: Statistics

1.1 What is Statistics?

Population: A collection, or set, of individuals, objects, or events whose properties are to be analyzed.

Sample: Subset of the population.

Variable: A characteristic of interest about each individual element of a population or sample.

Data value: The value of the variable associated with one element of a population or sample.

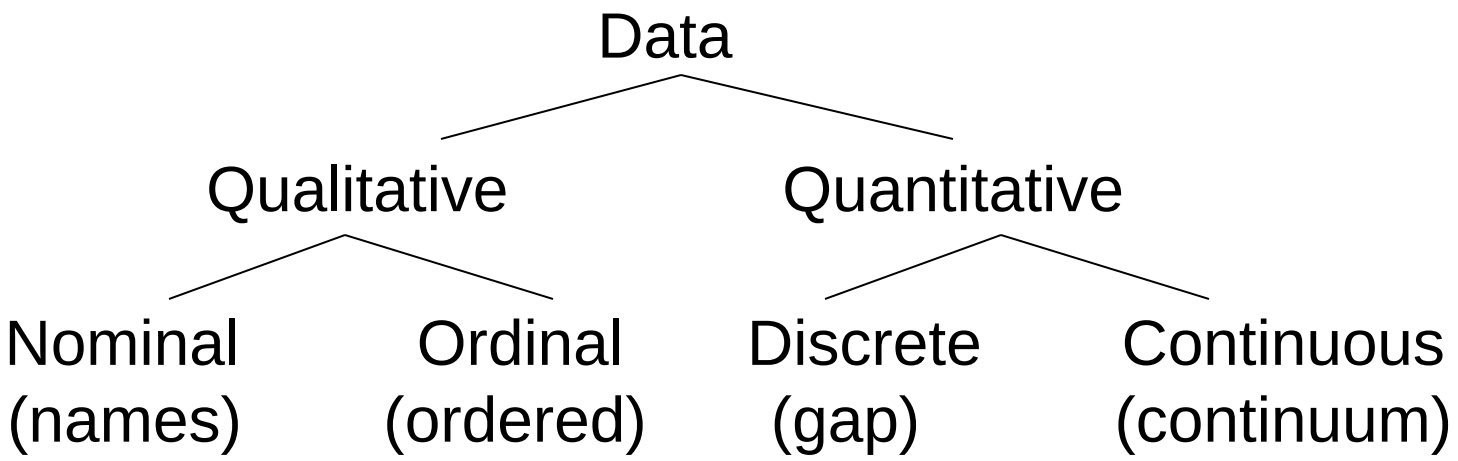
Parameter: A numerical value summarizing all the data of an entire population.

Statistic: A numerical value summarizing the sample data.

1: Statistics

1.1 What is Statistics?

Data: The set of values collected from the variable from each of the elements that belong to the sample.



Lecture Chapter 2

Chapter 2: Descriptive Analysis and Presentation of Single-Variable Data

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

2: Descriptive Analysis and Single Variable Data

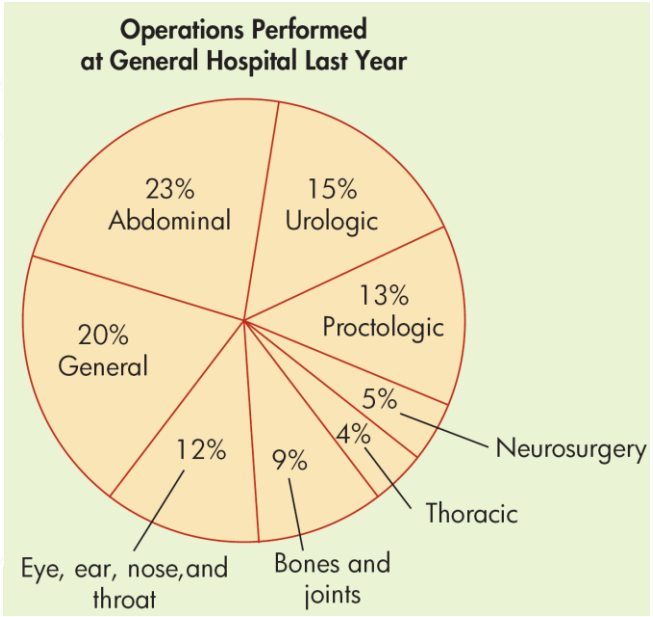
2.1 Graphs - Qualitative Data

Pie charts (circle graphs) and bar graphs: Graphs that are used to summarize **qualitative**, or attribute, or categorical **data**.

TABLE 2.1

Operations Performed at General Hospital Last Year [TA02-01]

Type of Operation	Number of Cases
Thoracic	20
Bones and joints	45
Eye, ear, nose, and throat	58
General	98
Abdominal	115
Urologic	74
Proctologic	65
Neurosurgery	23
Total	498



Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

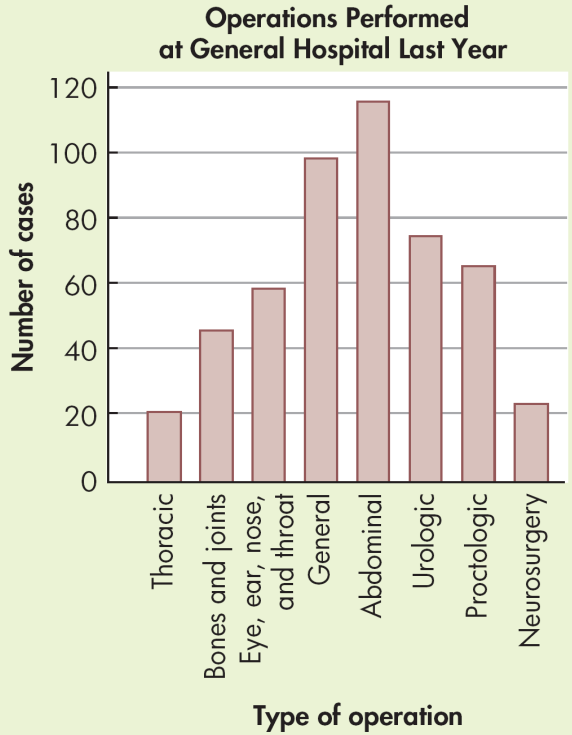
2.1 Graphs - Qualitative Data

Pie charts (circle graphs) and bar graphs: Graphs that are used to summarize **qualitative**, or attribute, or categorical **data**.

TABLE 2.1

Operations Performed at General Hospital Last Year [TA02-01]

Type of Operation	Number of Cases
Thoracic	20
Bones and joints	45
Eye, ear, nose, and throat	58
General	98
Abdominal	115
Urologic	74
Proctologic	65
Neurosurgery	23
Total	498



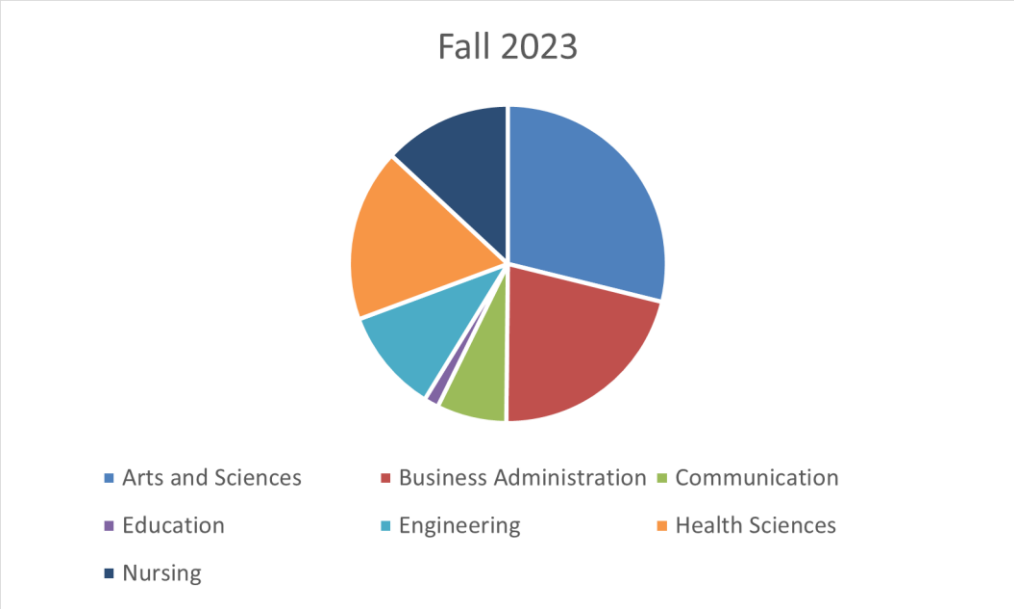
Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data

Pie charts (circle graphs) and bar graphs: Graphs that are used to summarize **qualitative**, or attribute, or categorical **data**.

First Year By College	F2023
Arts and Sciences	543
Business Administration	401
Communication	134
Education	28
Engineering	199
Health Sciences	331
Nursing	246
Total	**1882**



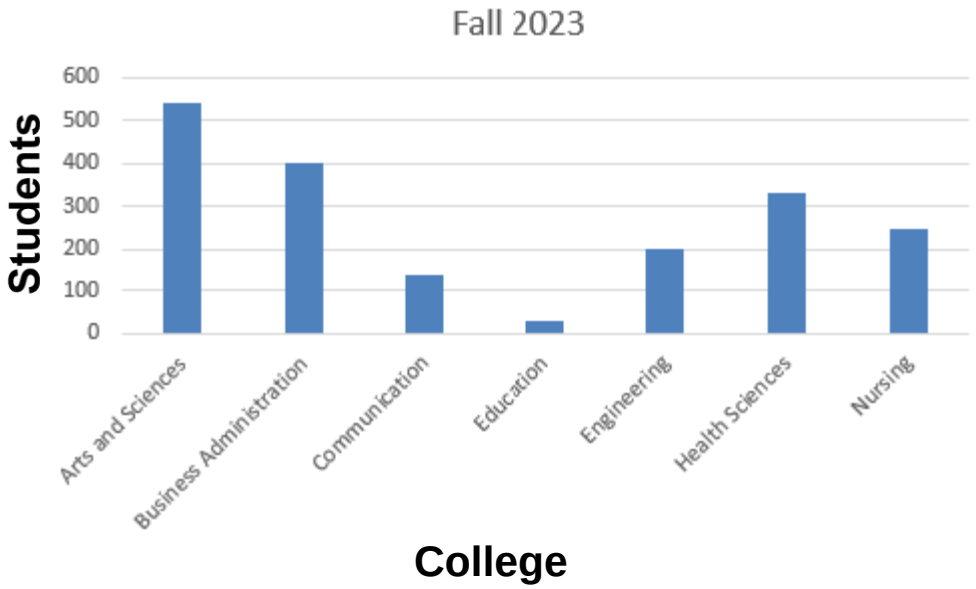
https://www.marquette.edu/institutional-research-analysis/public-reports/freshman-dash.php

2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data

Pie charts (circle graphs) and bar graphs: Graphs that are used to summarize **qualitative**, or attribute, or categorical **data**.

First Year By College	F2023
Arts and Sciences	543
Business Administration	401
Communication	134
Education	28
Engineering	199
Health Sciences	331
Nursing	246
Total	**1882**

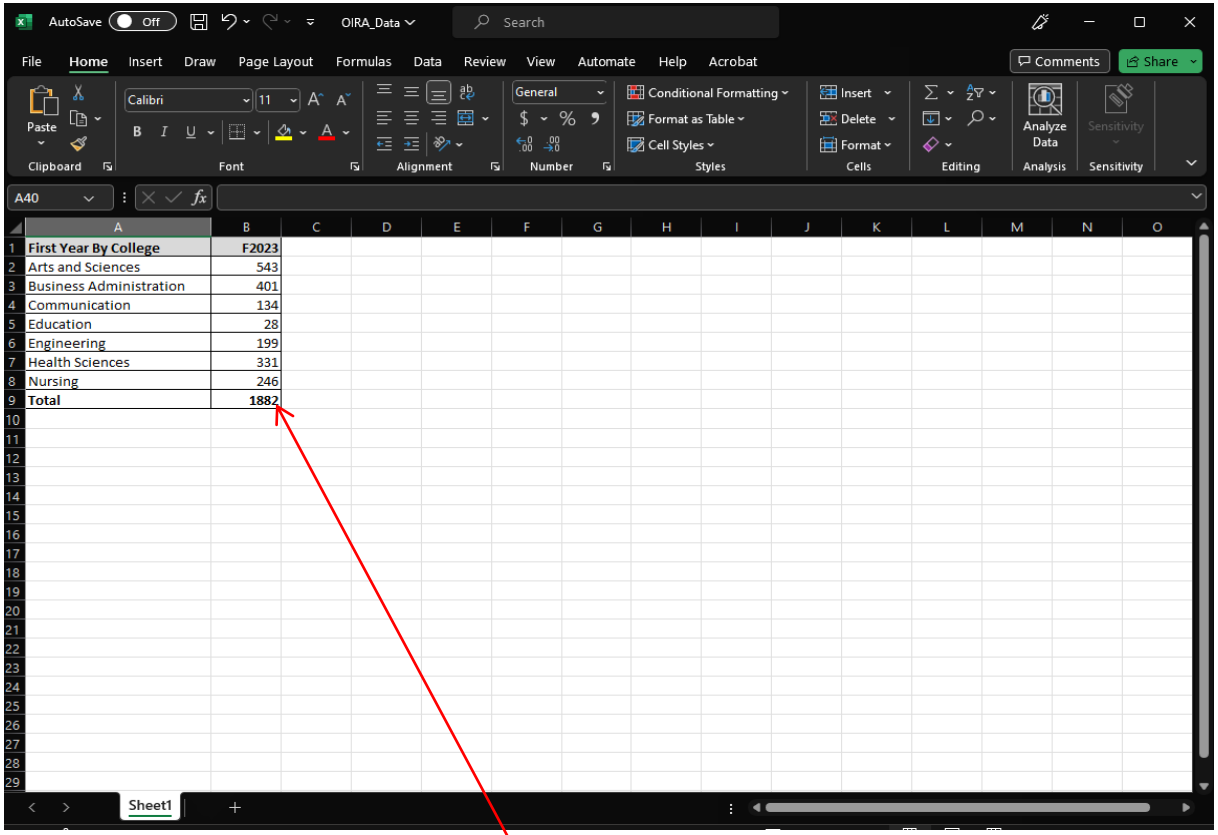


https://www.marquette.edu/institutional-research-analysis/public-reports/freshman-dash.php

2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data

Example:



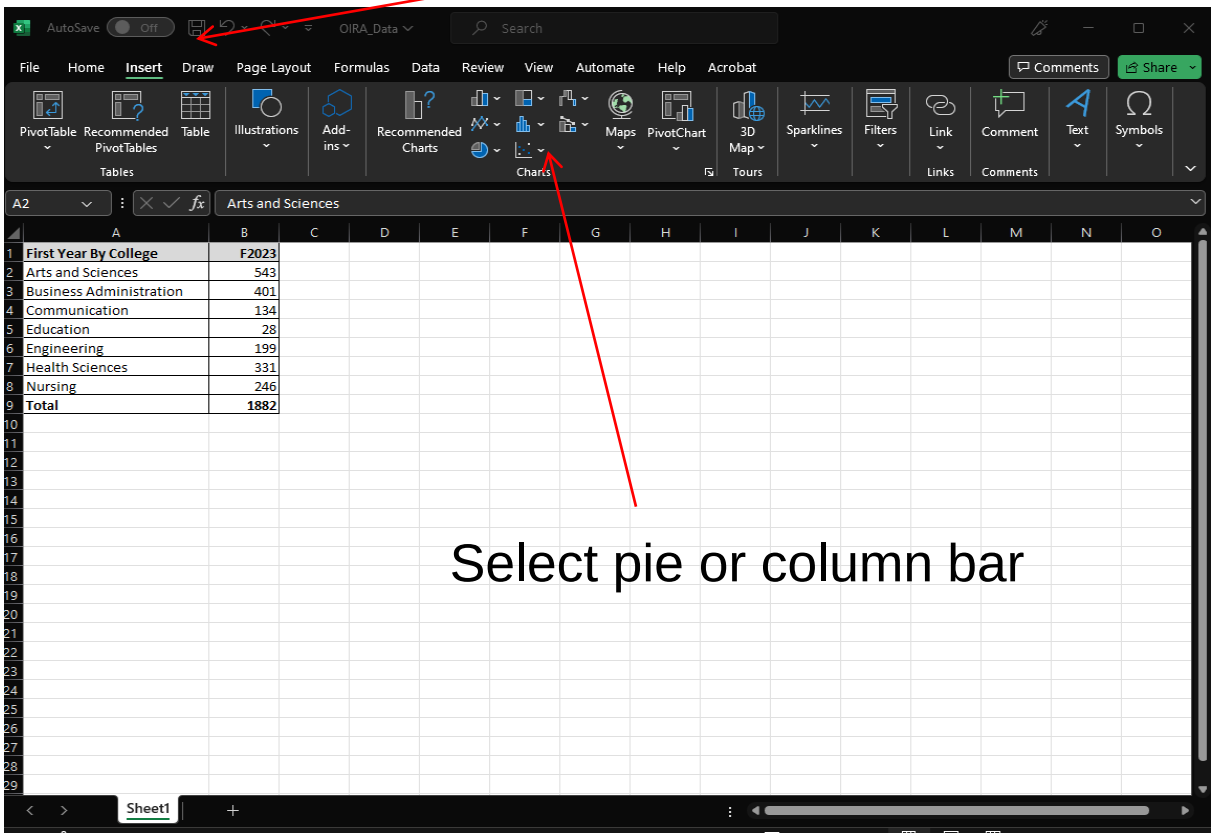
	First Year By College	F2023
2	Arts and Sciences	543
3	Business Administration	401
4	Communication	134
5	Education	28
6	Engineering	199
7	Health Sciences	331
8	Nursing	246
9	Total	1882

Enter and highlight data

2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data

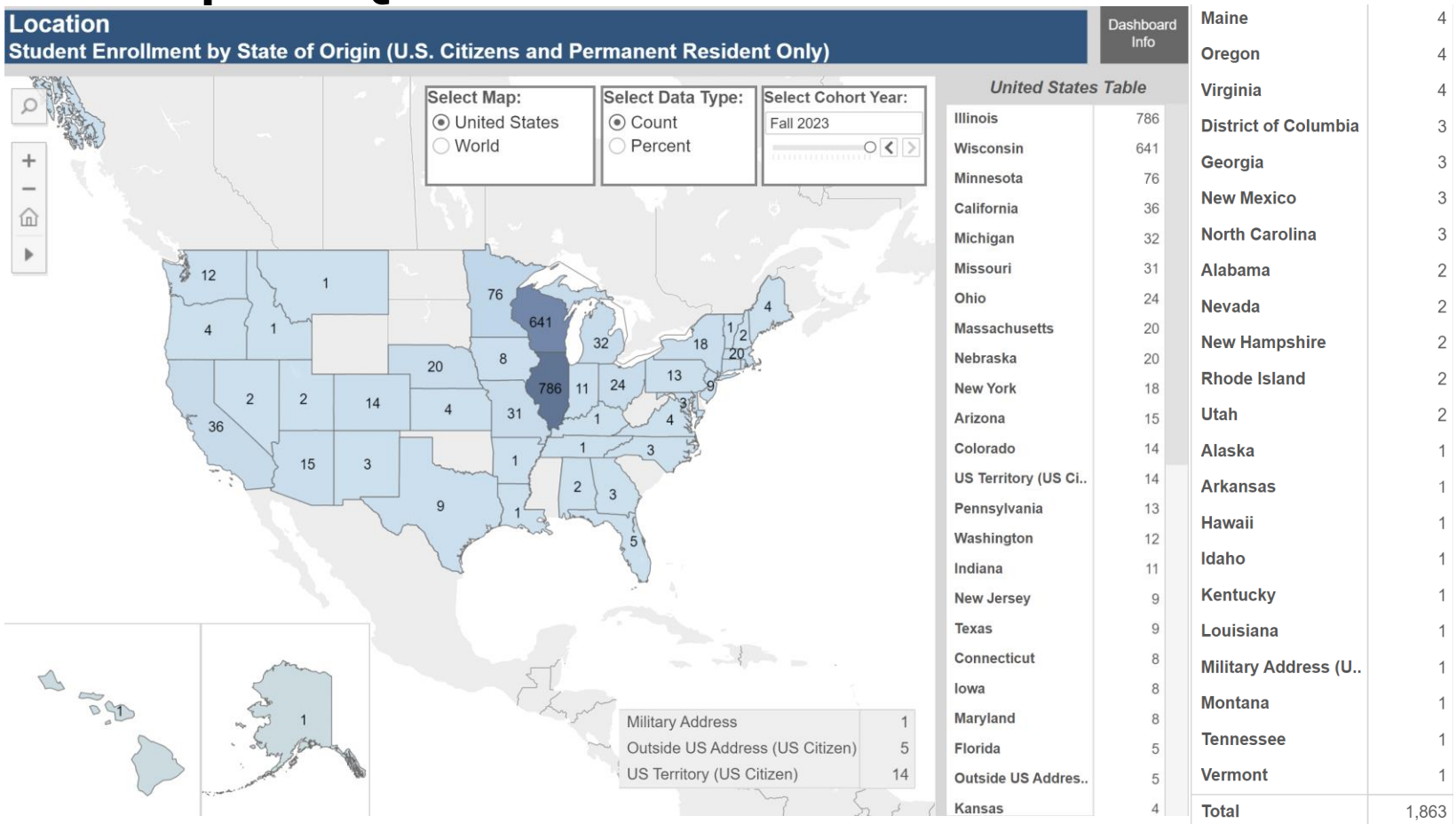
Example:



First Year By College	F2023
Arts and Sciences	543
Business Administration	401
Communication	134
Education	28
Engineering	199
Health Sciences	331
Nursing	246
Total	**1882**

2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data



United States Table

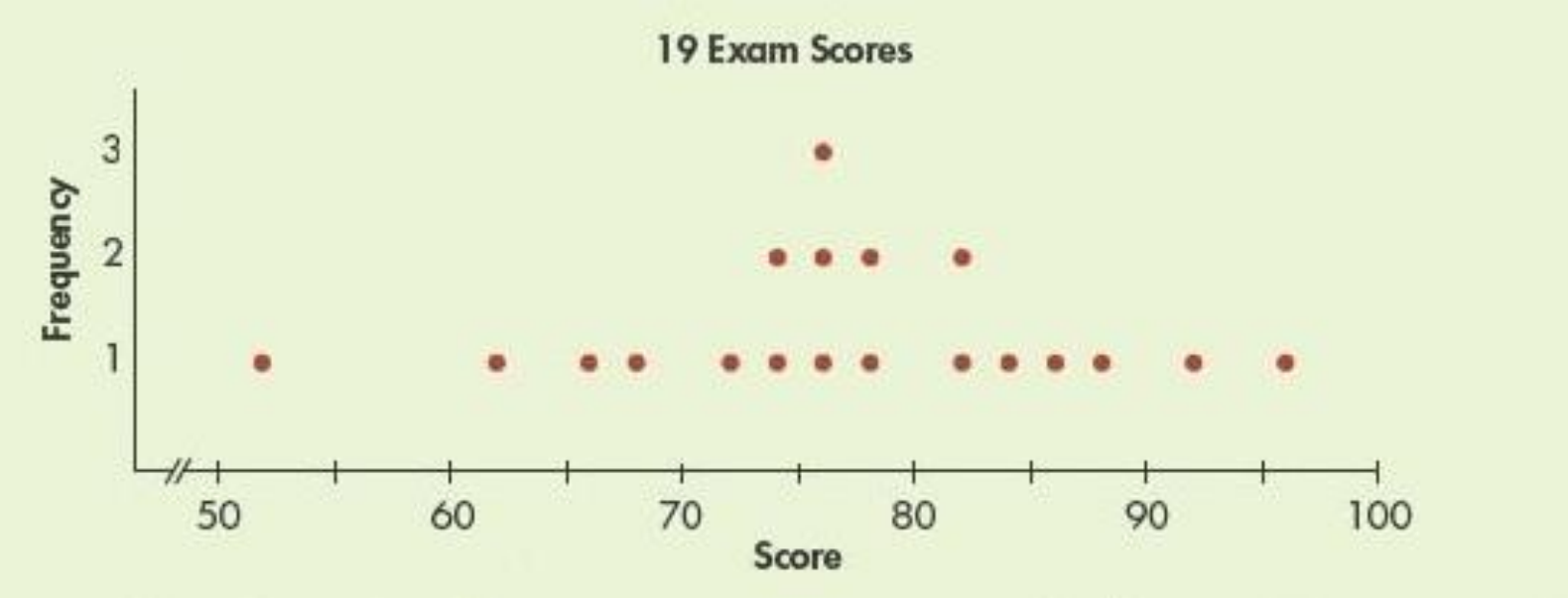
Illinois	786
Wisconsin	641
Minnesota	76
California	36
Michigan	32
Missouri	31
Ohio	24
Massachusetts	20
Nebraska	20
New York	18
Arizona	15
Colorado	14
US Territory (US Ci..	14
Pennsylvania	13
Washington	12
Indiana	11
New Jersey	9
Texas	9
Connecticut	8
Iowa	8
Maryland	8
Florida	5
Outside US Addres..	5
Kansas	4
Maine	4
Oregon	4
Virginia	4
District of Columbia	3
Georgia	3
New Mexico	3
North Carolina	3
Alabama	2
Nevada	2
New Hampshire	2
Rhode Island	2
Utah	2
Alaska	1
Arkansas	1
Hawaii	1
Idaho	1
Kentucky	1
Louisiana	1
Military Address (U..	1
Montana	1
Tennessee	1
Vermont	1
Total	1,863

https://www.marquette.edu/institutional-research-analysis/public-reports/freshman-dash.php

2: Descriptive Analysis and Single Variable Data

2.1 Graphs – Quantitative Data

Dotplot Display: Displays the data of a sample by representing each data value with a dot positioned above the scale.



Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Quantitative Data

Distribution: The Pattern of variability displayed by the data of a variable. The distribution displays the frequency of each value of the variable.

2.2 Frequency Distributions and Histograms

Frequency distribution: A listing, often expressed in chart form, that pairs values of a variable with their frequency.

2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

Statistics Exam Scores

This is somewhat of an art!

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

1. Identify the high score (H=98) and the low score (L=39).

   range = $H - L$ = 98 – 39 = 59

Figure from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

Statistics Exam Scores

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

2. Select the number of classes (m=7) and a class width (c=10)
   (These are subjective and depend on how you feel.
   But the larger n, the more classes and smaller c you should have,
   the smaller n, the fewer classes you should have and larger c.)

   mc=70 a little larger than the range=59.

Figure from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

Statistics Exam Scores

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

3. Pick a starting point and set up class boundaries

$$35 \le x < 45,\ \ 45 \le x < 55,\ \ 55 \le x < 65$$

$$65 \le x < 75,\ \ 75 \le x < 85,\ \ 85 \le x < 95,\ \ 95 \le x < 105$$

Figure from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

Statistics Exam Scores

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

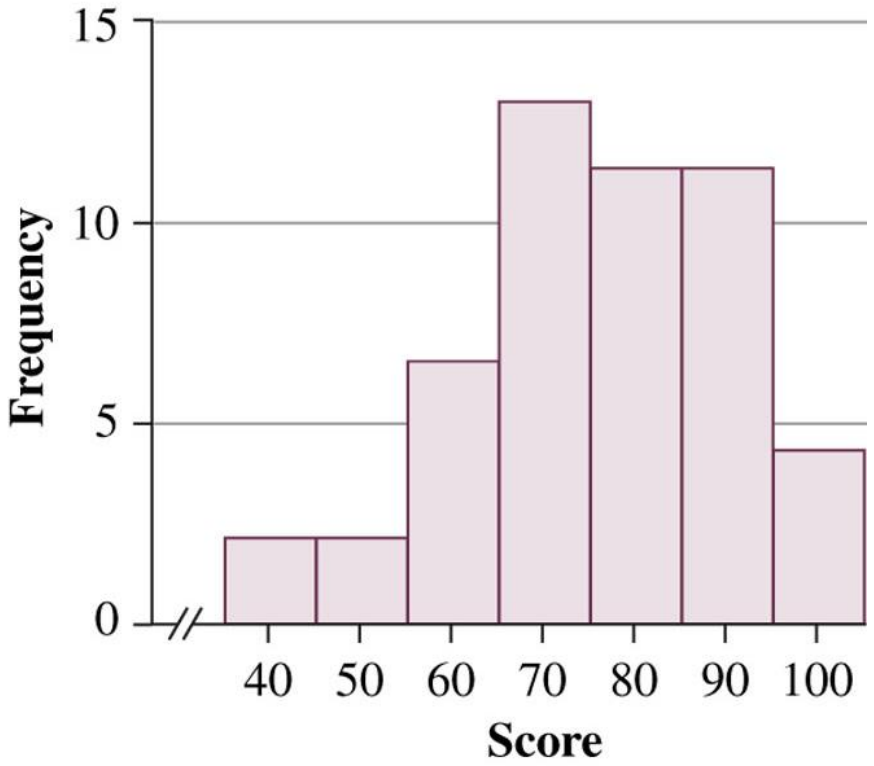
Class Number	Class Tallies	Boundaries	Frequency
1	||	$35 \le x < 45$	2
2	||	$45 \le x < 55$	2
3	||||| ||	$55 \le x < 65$	7
4	||||| ||||| |||	$65 \le x < 75$	13
5	||||| ||||| |	$75 \le x < 85$	11
6	||||| ||||| |	$85 \le x < 95$	11
7	||||	$95 \le x \le 105$	4
			50

Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

Boundaries	Frequency
$35 \le x < 45$	2
$45 \le x < 55$	2
$55 \le x < 65$	7
$65 \le x < 75$	13
$75 \le x < 85$	11
$85 \le x < 95$	11
$95 \le x \le 105$	4
	50



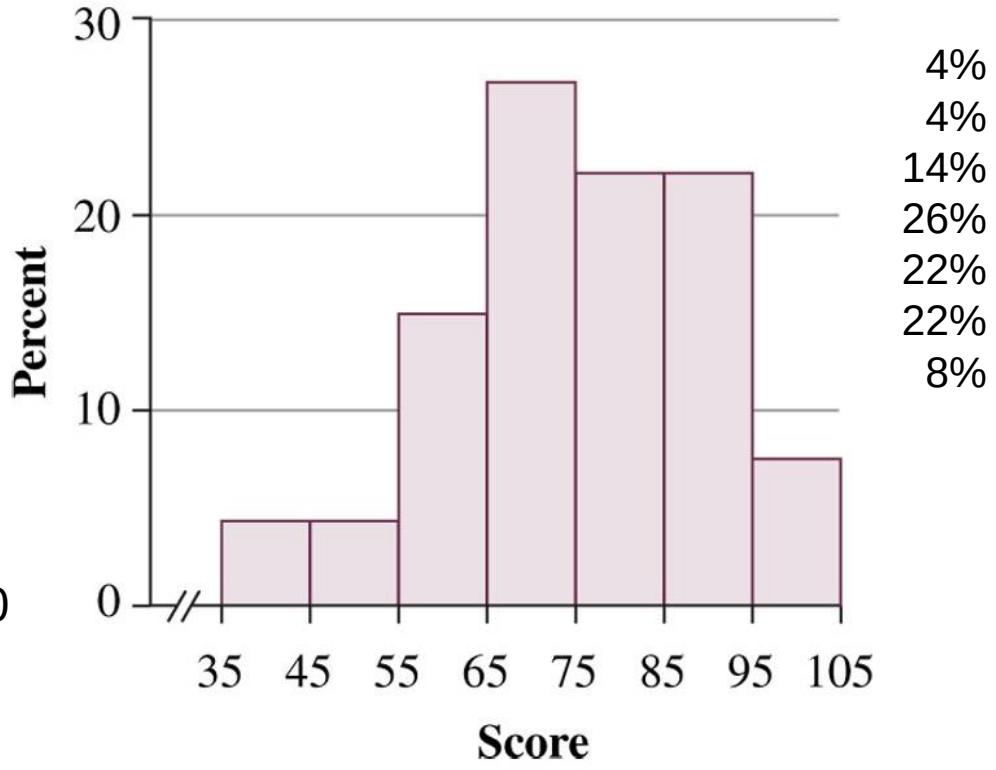
Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

Boundaries	Frequency
$35 \le x < 45$	2
$45 \le x < 55$	2
$55 \le x < 65$	7
$65 \le x < 75$	13
$75 \le x < 85$	11
$85 \le x < 95$	11
$95 \le x \le 105$	4
	50

Divide all by 50 to get percent



Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.3 Measures of Central Tendency

We were able to present the information contained in this sample of data using graphical methods.

Now let's summarize the information contained in the sample of data using numerical summary measures.

Describe the measures of central tendency
(**sample mean, sample median, sample mode**),
then describe some the measures of dispersion,
then use them in a toy example.

2: Descriptive Analysis and Single Variable Data

2.3 Measures of Central Tendency

Sample Mean: The usual average you are familiar with.
Represented by $\bar{x}$ called "*x-bar.*" p. 63

Simply add up all the values and divide by the number values.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Remember the sigma notation we reviewed?

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

Round-off Rule: When rounding a number, let's keep one more decimal place than the original numbers.

2: Descriptive Analysis and Single Variable Data

2.3 Measures of Central Tendency

Sample Median: The thing in the middle of the road! LOL.

Statistics humor.

Middle value when data ordered. 50% above, 50% below

Represented by $\tilde{x}$ called "*x-tilde.*" p. 64

$$\tilde{x} = \text{middle} \quad \text{value}$$

Order data from smallest to largest.

If n odd, $d(\tilde{x}) = \dfrac{n+1}{2}$ value

$d(\tilde{x})$ called depth

If n even, $d(\tilde{x})$ avg. of $\dfrac{n}{2}$ and $\dfrac{n}{2}+1$ values

2: Descriptive Analysis and Single Variable Data

2.3 Measures of Central Tendency

Sample Mode: The value that happens most often in sample.
Represented by $\hat{x}$ called "*x-hat.*" p. 66

Order data from smallest to largest.
Count how many time each value occurs.
Take the one with the highest count for $\hat{x}$.

If two or more values in a sample are tied for the highest frequency, we say that there is **no mode**.

2: Descriptive Analysis and Single Variable Data

2.3 Measures of Central Tendency

The measures of central tendency characterize the center of the distribution of data values.

There are other measures called measures of dispersion that characterize the spread or variability in the data.

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

Range: The difference between the highest data value (H) and lowest data values (L). p. 74

$$range = \text{high value - low value}$$

$$range = H - L$$

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

Deviation from the mean: The difference between the data value x_i and the sample mean $\overline{x}$. p. 74

$$i^{th}\ deviation\ \ from\ mean = x_i - \overline{x}$$

There can be n of these because we have $x_1, x_2, ..., x_n$.

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

Sample Variance: The mean of the squared deviations using n-1 as a divisor. p. 75

There are two equivalent formulas that can be used.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \left[\left(\sum_{i=1}^{n} x_i\right)^2 / n\right]\right\}$$

where x_i is i^{th} data value, $\bar{x}$ is sample mean, n is sample size.

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

Sample Standard Deviation: Square root of the sample variance. Has same units data values and sample mean.

$$s = \sqrt{s^2}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Note:

Sample variance s^2 uses the entire sample and a denominator n-1!
Population variance σ^2 the entire population and a denominator N!

n items in the sample
N items in the population
$n \leq N$, $n < N$ for a sample

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Central Tendency

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example:

Data values: 1,2,2,3,4

Sample Mean=?

$\overline{x} = ?$

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Central Tendency

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example:

Data values: 1,2,2,3,4

Sample Mean=?

$$\bar{x} = \frac{1+2+2+3+4}{5}$$

Sample Mean=?

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Central Tendency

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example:

Data values: 1,2,2,3,4

Sample Mean=?

$$\bar{x} = \frac{1+2+2+3+4}{5}$$

Sample Mean=?

$$\bar{x} = 2.4$$

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Central Tendency

$$\tilde{x} = \text{middle} \quad \text{value}$$

Example:

Data values: 1,2,2,3,4

Sample Median=?

$$\tilde{x} = ?$$

Order data from smallest to largest.
If the number of data values is odd, take the middle value as the median.
If the number of data values is even, take the average of the middle two.

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Central Tendency

$$\tilde{x} = \text{middle value}$$

Example:

Data values: 1,2,2,3,4

Sample Median=?

$$\tilde{x} = 2$$

Order data from smallest to largest.
If the number of data values is odd, take the middle value as the median.
If the number of data values is even, take the average of the middle two.

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Central Tendency

$\hat{x} =$ most often value

Example:

Data values: 1,2,2,3,4

Order data from smallest to largest. Count how many time each value occurs. Take the one with the highest count.

Sample Mode=?

$\hat{x} = ?$

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Central Tendency

$$\hat{x} = \text{most often value}$$

Example:

Data values: 1,2,2,3,4

Order data from smallest to largest. Count how many time each value occurs. Take the one with the highest count.

Sample Mode=?

$$\hat{x} = 2$$

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Example:

Data values: 1,2,2,3,4

Sample Variance=?

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Example: $x_1\, x_2\, x_3\, x_4\, x_5$

Data values: 1,2,2,3,4 $\quad \bar{x} = 2.4$

Sample Variance=?

$$s^2 = \frac{1}{5-1}\left[(1-2.4)^2 + (2-2.4)^2 + (2-2.4)^2 + (3-2.4)^2 + (4-2.4)^2\right]$$

(x_1, x_2, x_3, x_4, x_5 point down; n points up; $\bar{x}$ points up)

Sample Variance=?

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Example:

Data values: 1,2,2,3,4 $\quad \bar{x} = 2.4$

Sample Variance=?

$$s^2 = \frac{1}{5-1}\left[(1-2.4)^2 + (2-2.4)^2 + (2-2.4)^2 + (3-2.4)^2 + (4-2.4)^2\right]$$

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
1	2.4	-1.4	1.96
2	2.4	-0.4	0.16
2	2.4	-0.4	0.16
3	2.4	0.6	0.36
4	2.4	1.6	2.56
12			5.20

Sample mean is $\frac{12}{5} = 2.4$

Sample variance is $\frac{5.2}{4} = 1.3$

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right\}$$

Example:

Data values: 1,2,2,3,4

$\bar{x} = 2.4$

Sample Variance=?

Beter way to compute

$$s^2 = \frac{1}{5-1}\left\{34 - \frac{1}{5}(12)^2\right\}$$

x	x^2
1	1
2	4
2	4
3	9
4	16
$\sum x = 12$	$34 = \sum x^2$

Sample mean is $\frac{12}{5} = 2.4$

Sample variance is $\frac{5.2}{4} = 1.3$

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right\}$$

$$s = \sqrt{s^2}$$

Example:

Data values: 1,2,2,3,4

$$s^2 = 1.3$$

Sample Standard Deviation=?

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$s = \sqrt{s^2}$$

Example:

Data values: 1,2,2,3,4

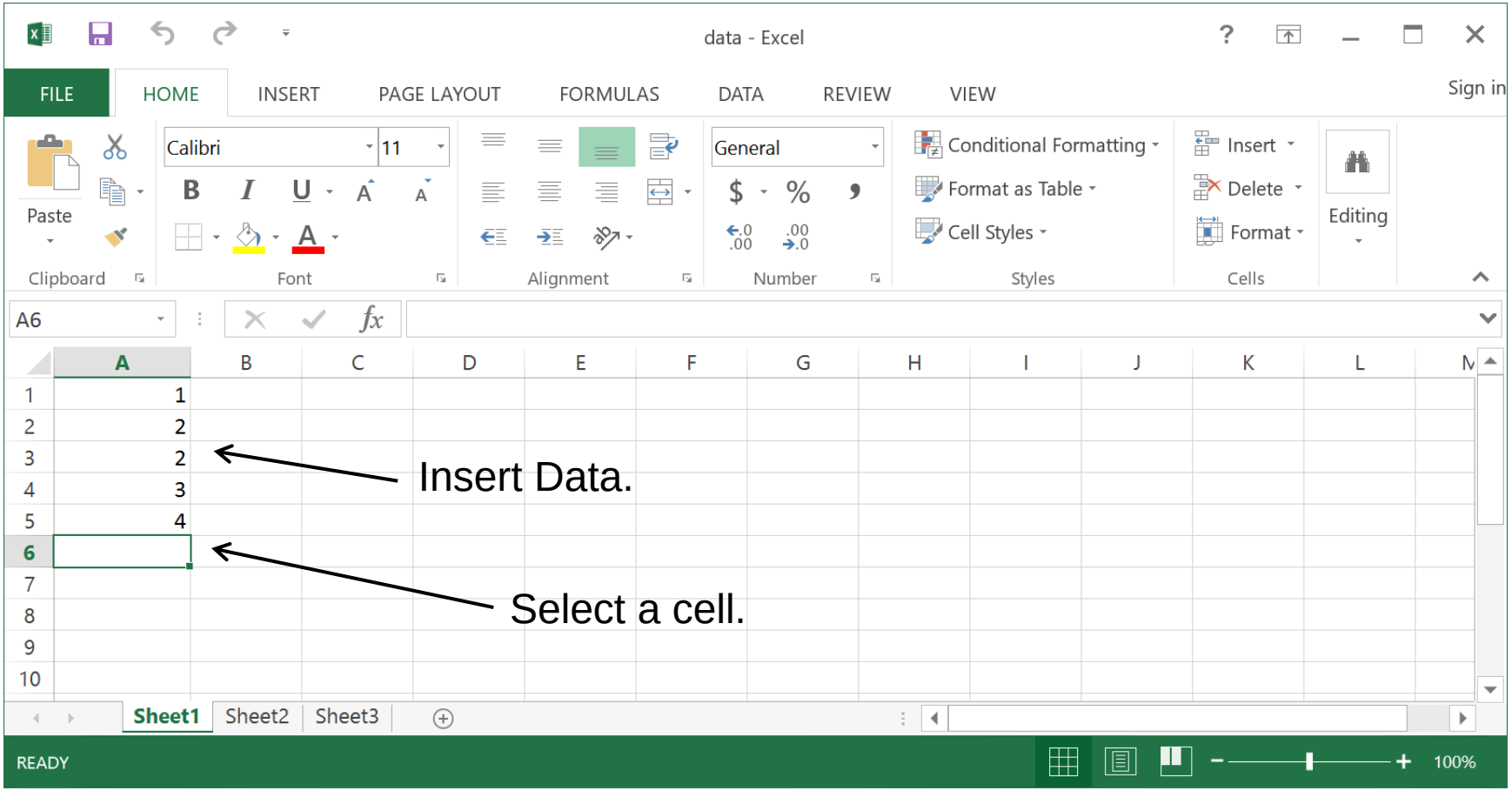
$$s^2 = 1.3$$

Sample Standard Deviation=?

$$s = 1.14$$

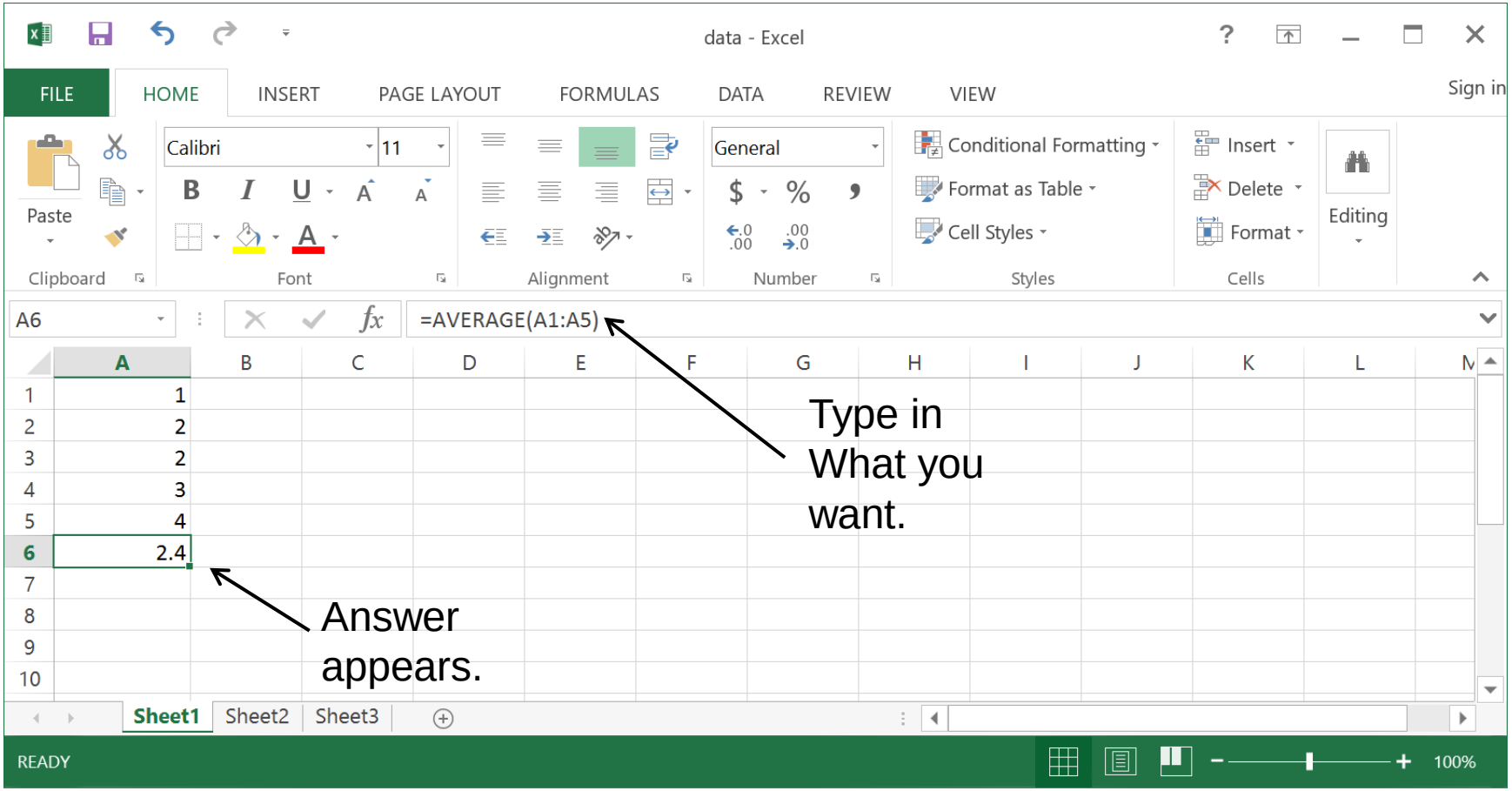
2: Descriptive Analysis and Single Variable Data

Example: In Excel:



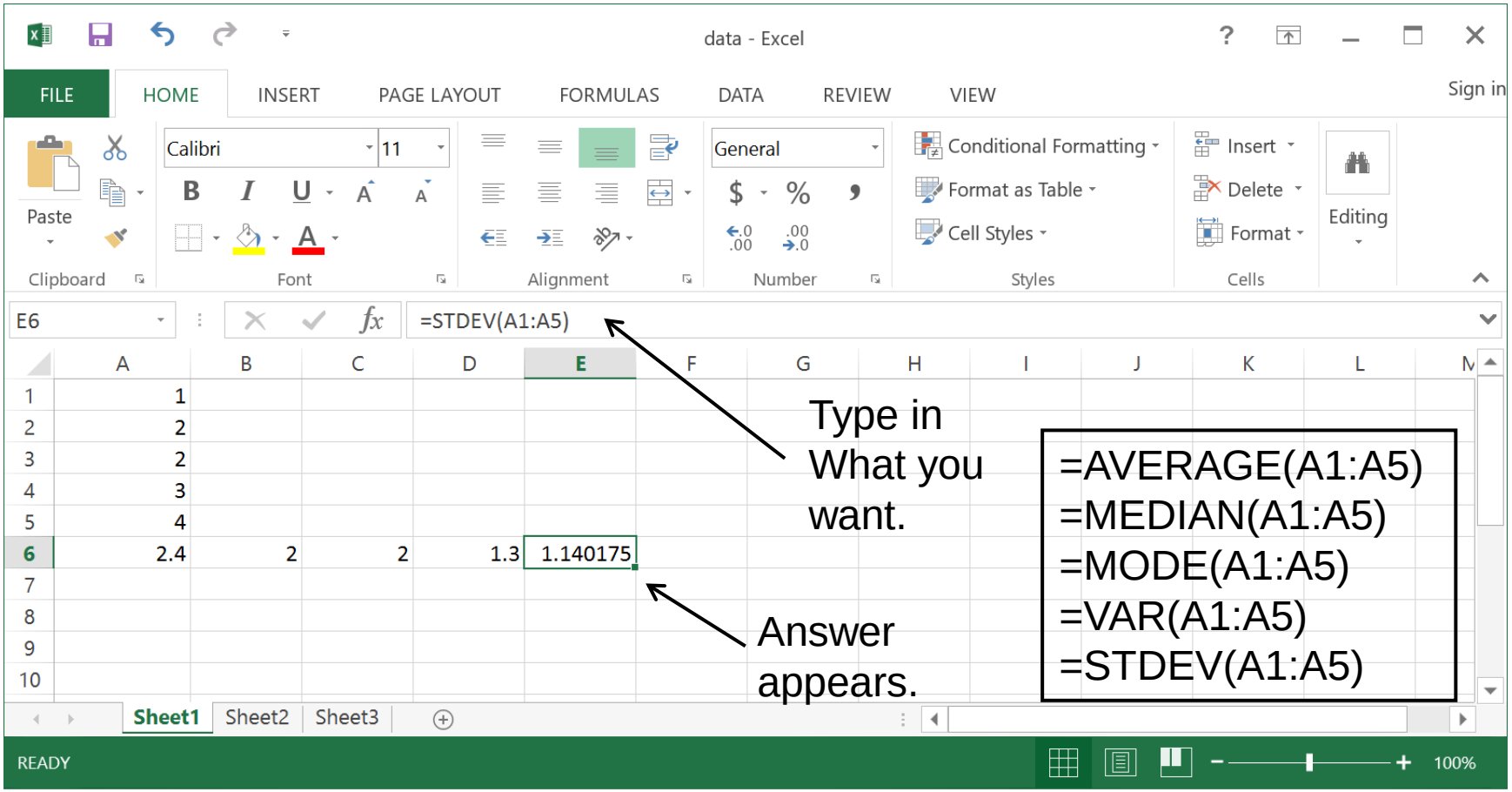
2: Descriptive Analysis and Single Variable Data

Example: In Excel:



2: Descriptive Analysis and Single Variable Data

Example: In Excel:



2: Descriptive Analysis and Single Variable Data

Questions?

Homework: Read Chapter 2.1-2.4,
WebAssign
Problems: 2.8, 2.35, 2.39, 2.75, 2.97, 2.105