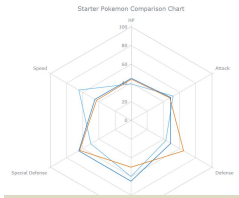
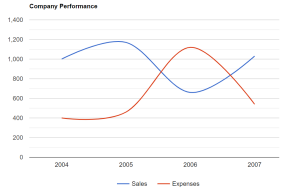
CSX4202/ITX4202: DATA MINING

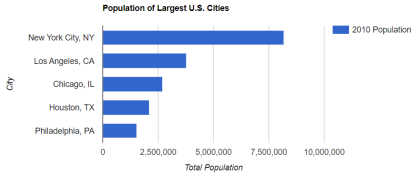
LECTURE 3'S SUPPLEMENTARY SLIDES – CREATING VISUALIZATION USING MS EXCEL



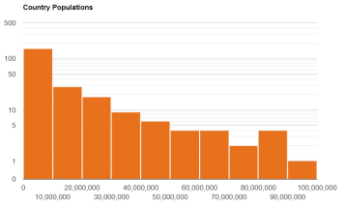
Spider Chart



Line Chart



Bar Chart



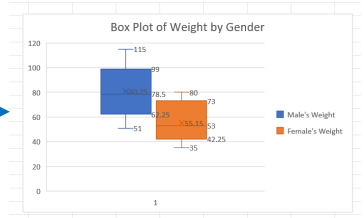
Histogram

Comparison

What would u like to show in your chart or graph?

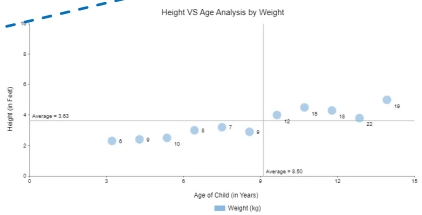
Relationship

Distribution

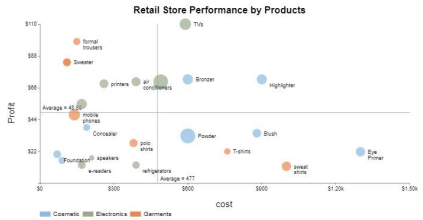


Box Plot

Scatter Plot

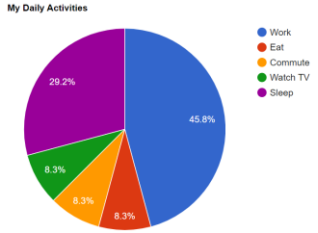


Bubble Chart

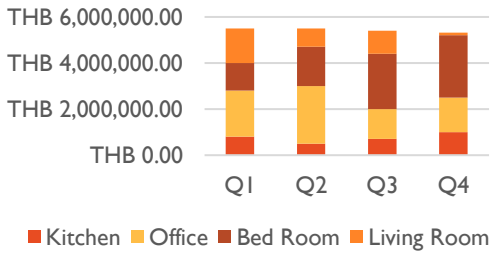


Composition

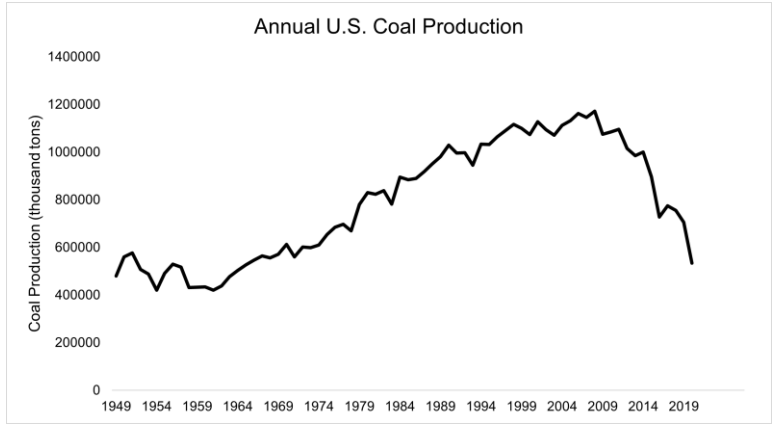
Pie Chart

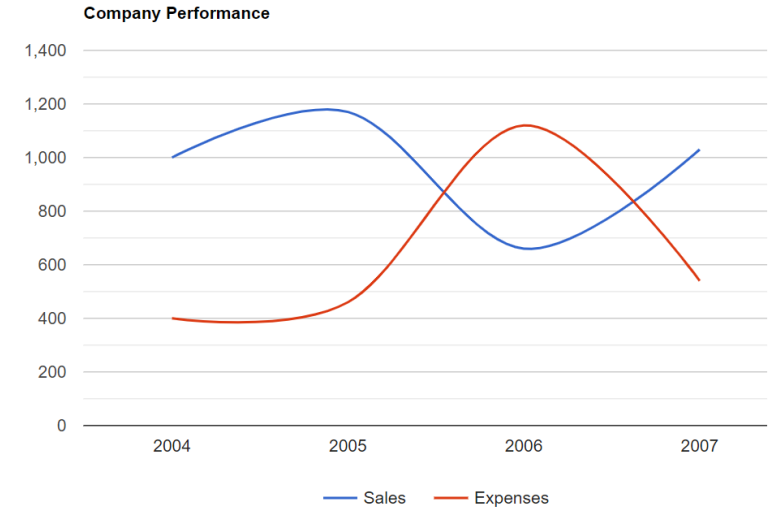


Stacked Column Chart

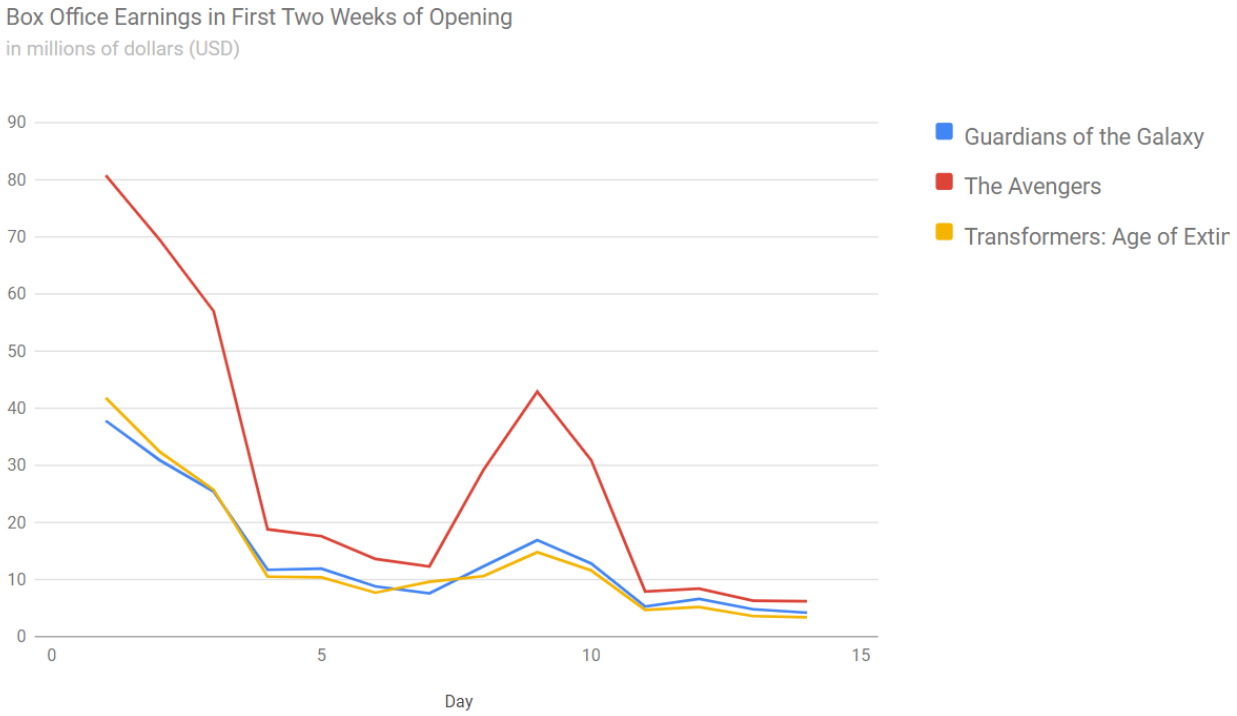


LINE CHART



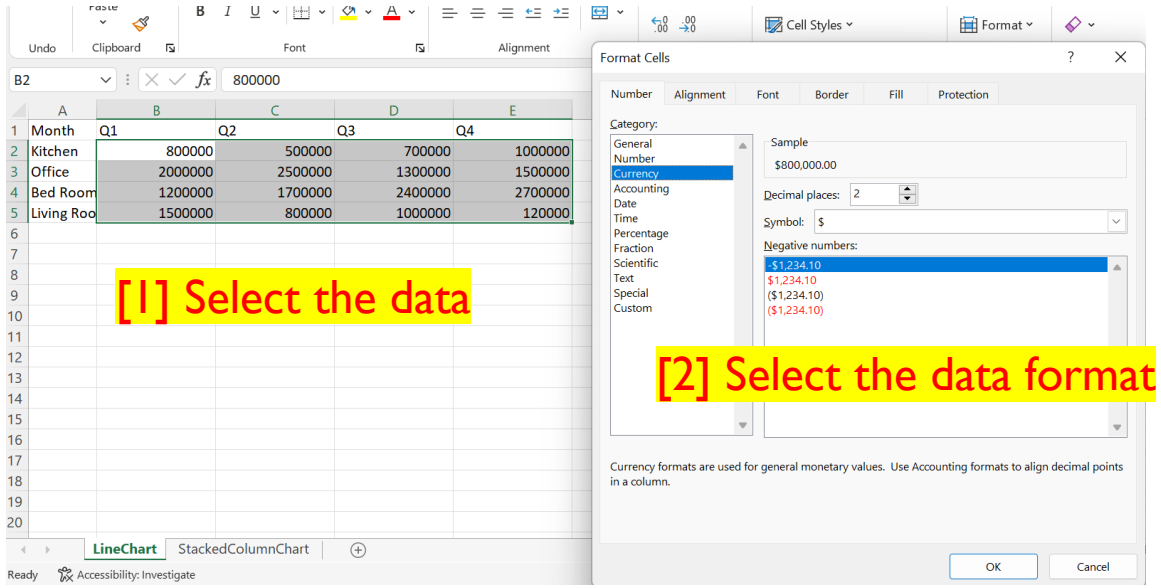


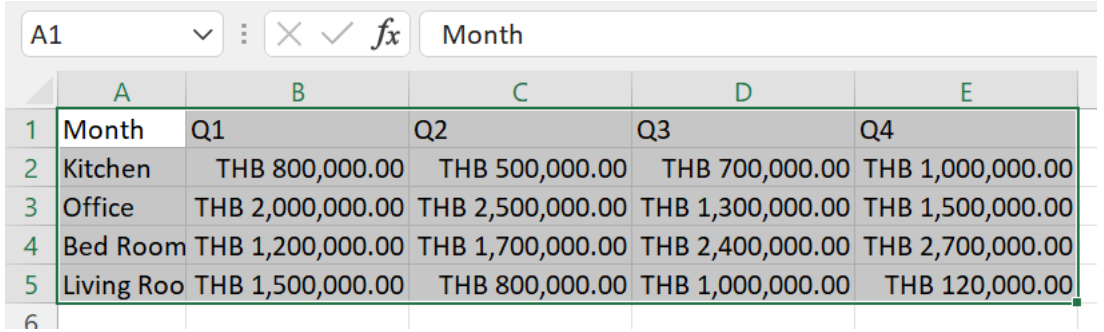
- Show the relationship between 2 variables
- Track changes or trends over time (x-axis)



Source. https://developers.google.com/chart/interactive/docs/gallery/linechart

CREATE A LINE CHART IN MS EXCEL – 1/3

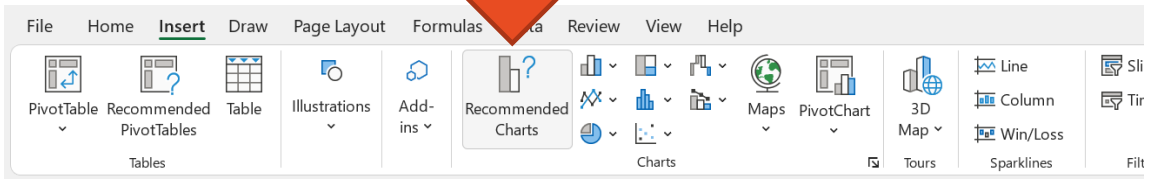




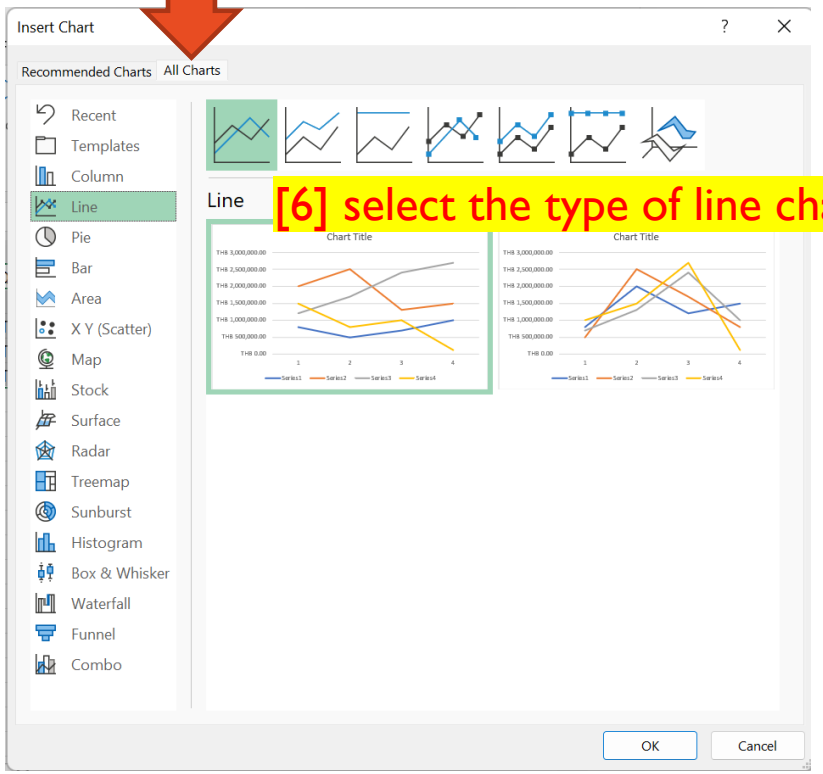
CREATE A LINE CHART IN MS EXCEL – 2/3

[3]

[4]



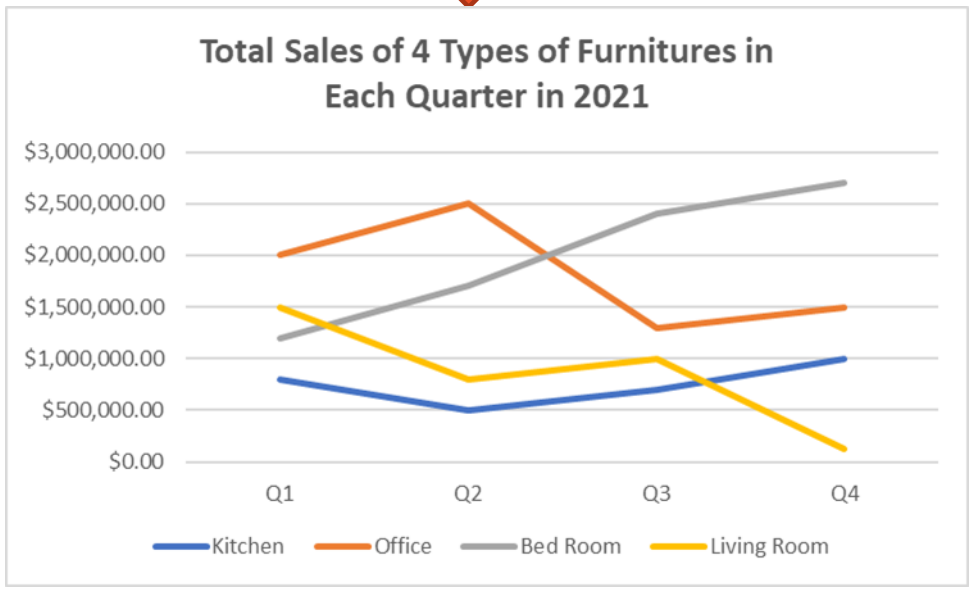
[5]



[6] select the type of line chart

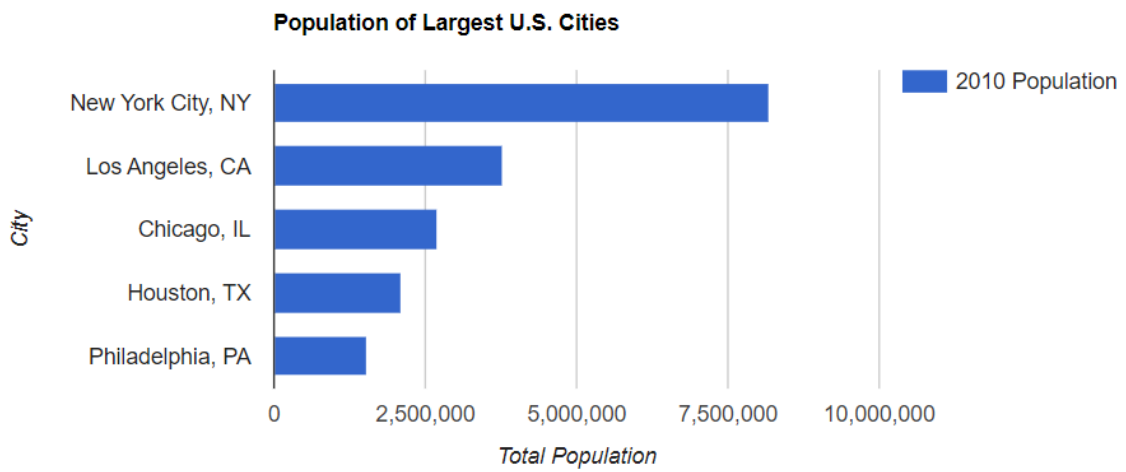
CREATE A LINE CHART IN MS EXCEL – 3/3

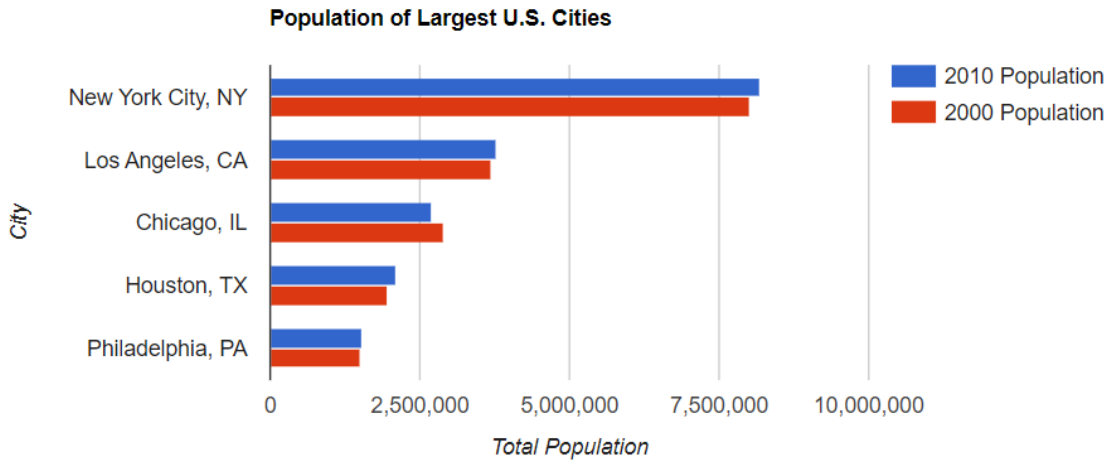
[7] Add title



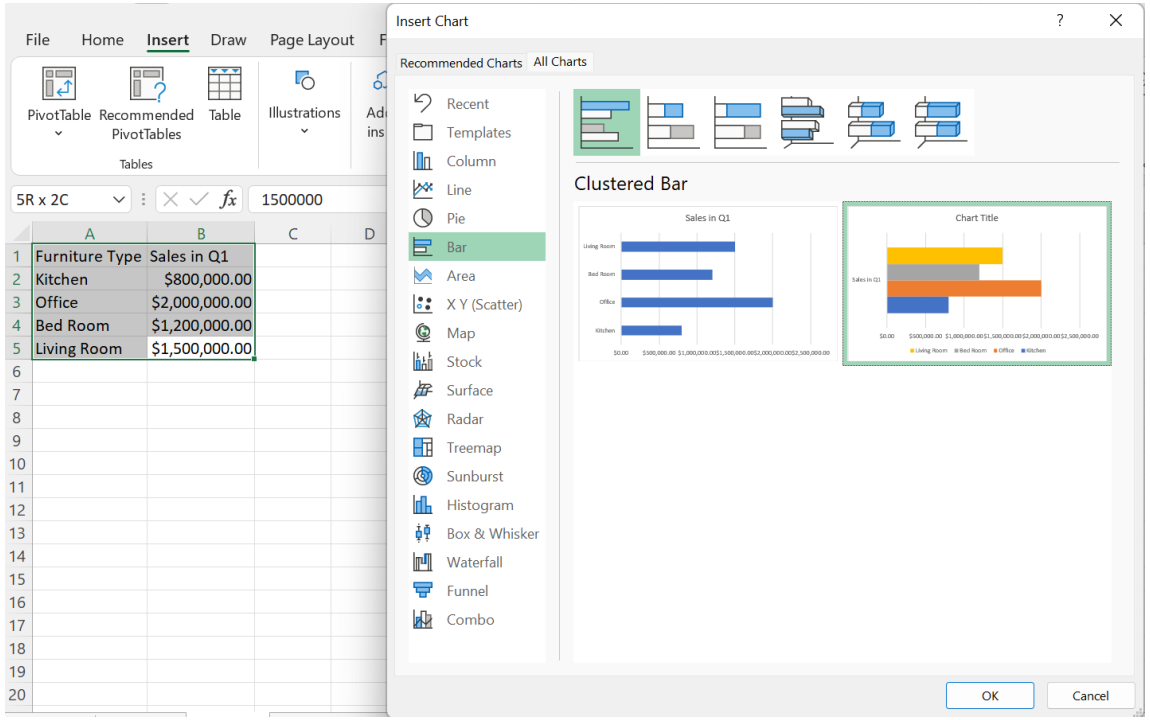
BAR CHART

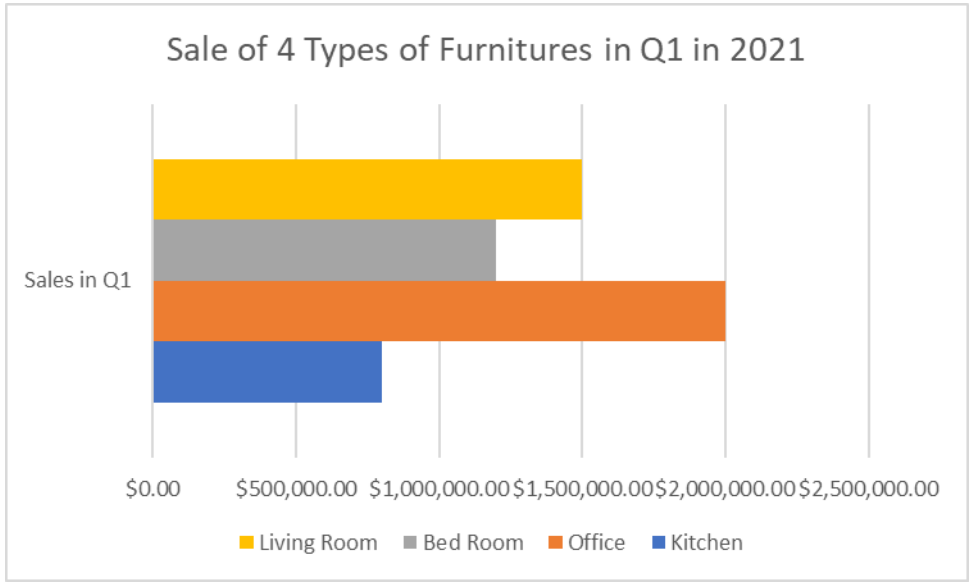
- Used with nominal data or numerical data that splits nicely into different categories.
- Compare data across multiple categories to quickly see the comparative results and trends.
- Can be either vertically or horizontally oriented.





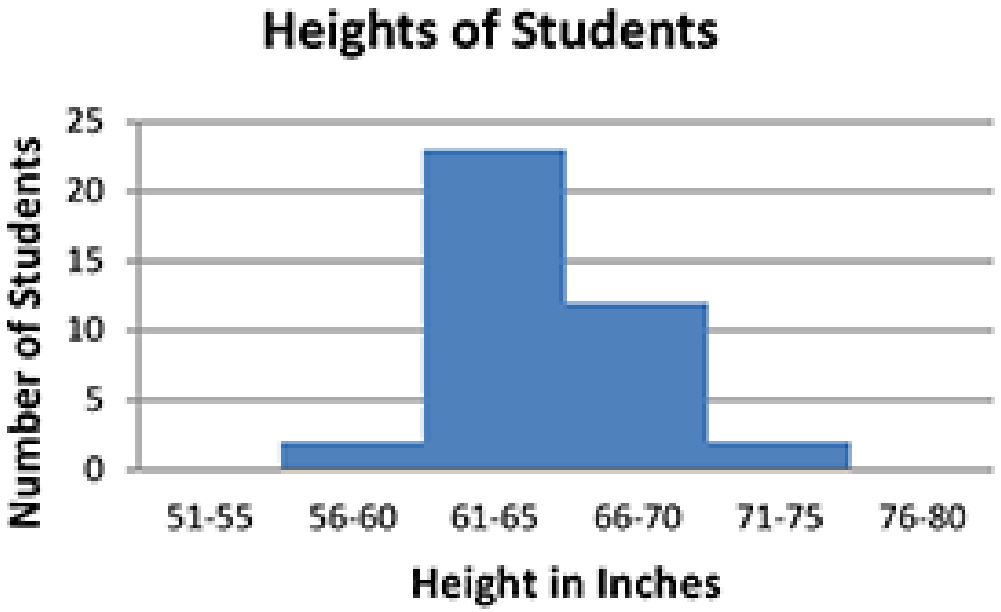
CREATE A BAR CHART IN MS EXCEL

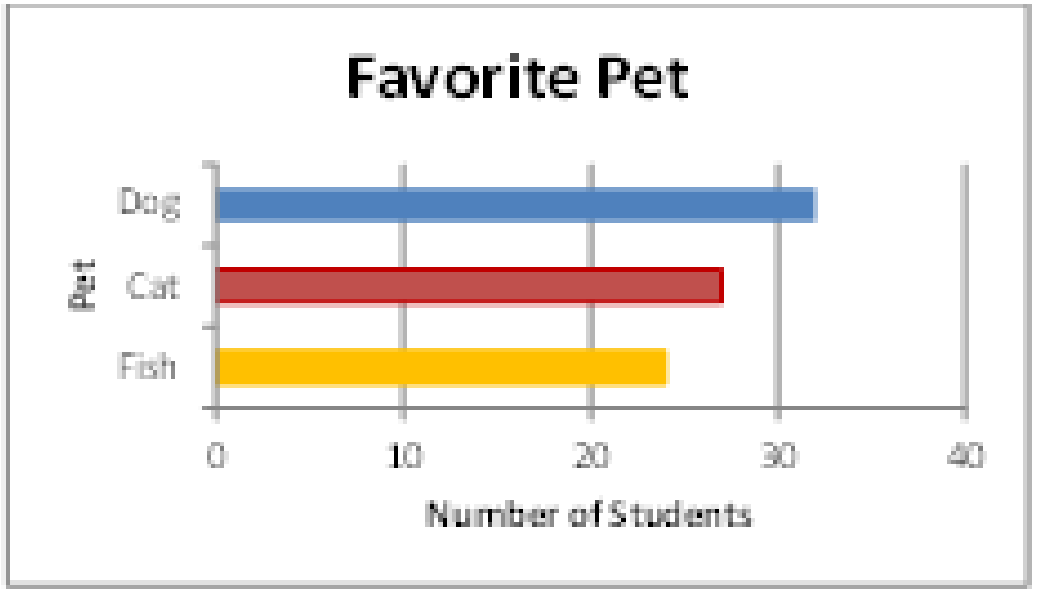




HISTOGRAM

- Is a specific type of bar chart.
- Shows the frequency distribution of a variable or several variables.
  - X-axis shows the categories or ranges.
  - Y-axis shows the measures/values/frequencies.

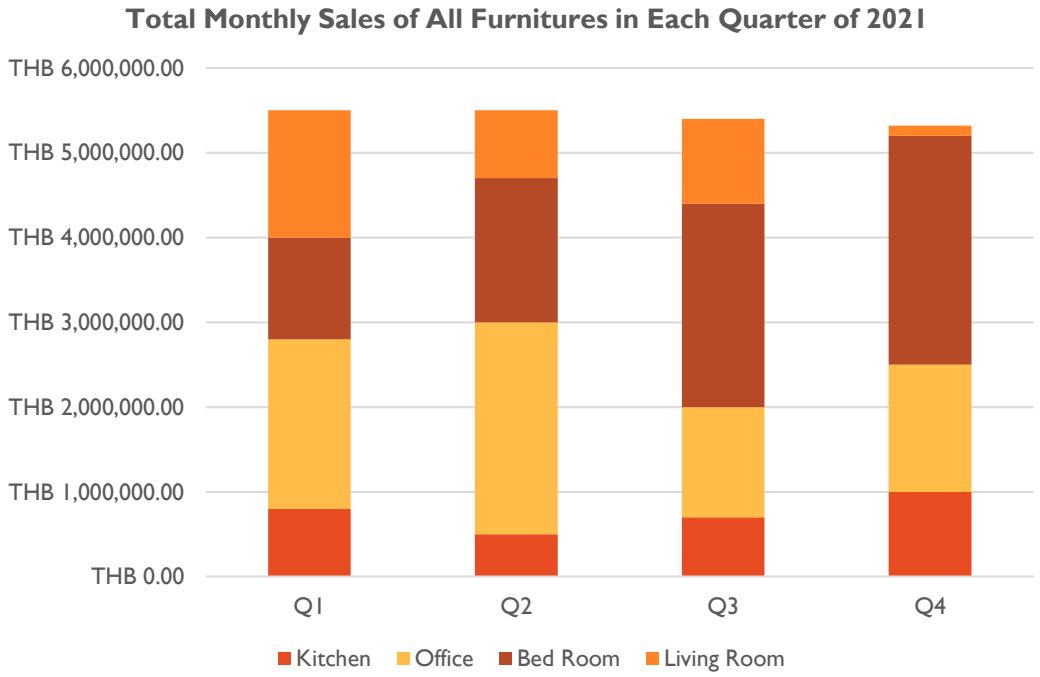


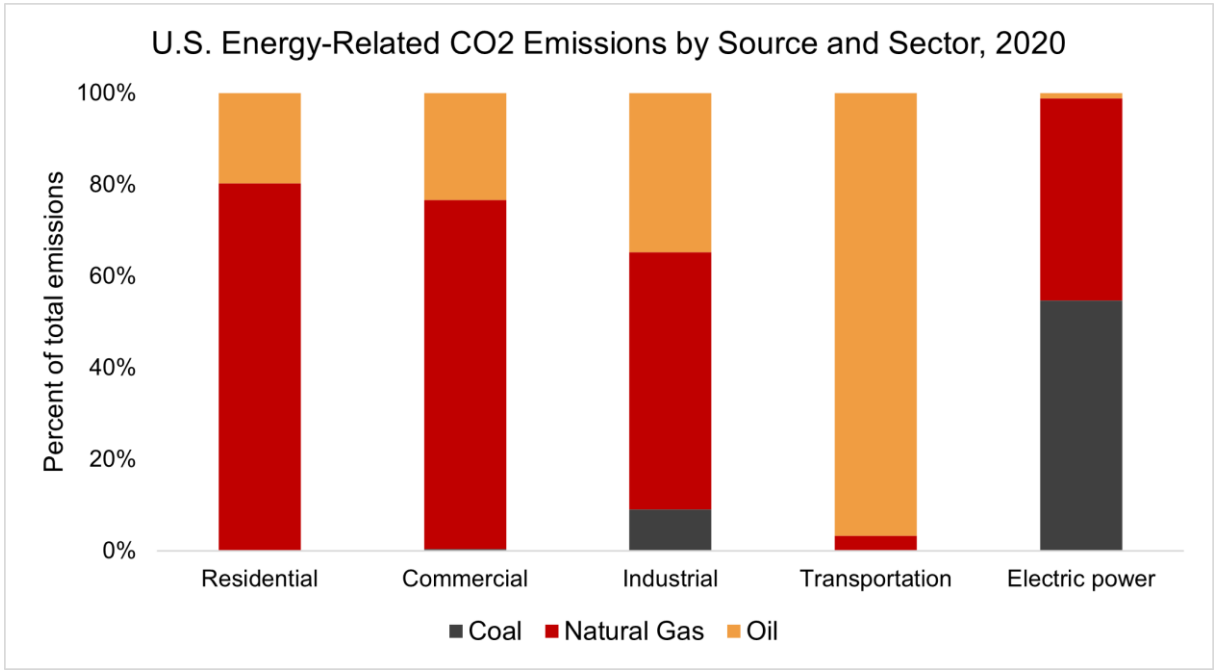


Source: http://jukebox.esc13.net/interactiveGlossary/HTML_files/histogram.html

STACKED COLUMN CHART

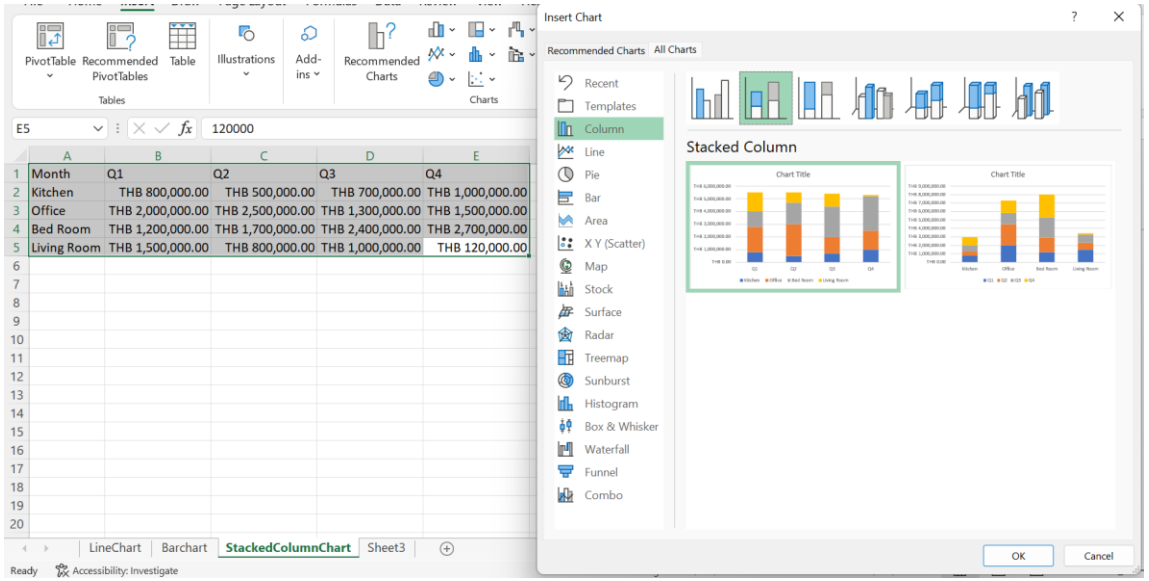
- Is a specific type of bar chart;
  - A basic Excel chart type to allow part-to-whole comparisons over time, or across categories.
- Data series are stacked one on top of the other in vertical columns.

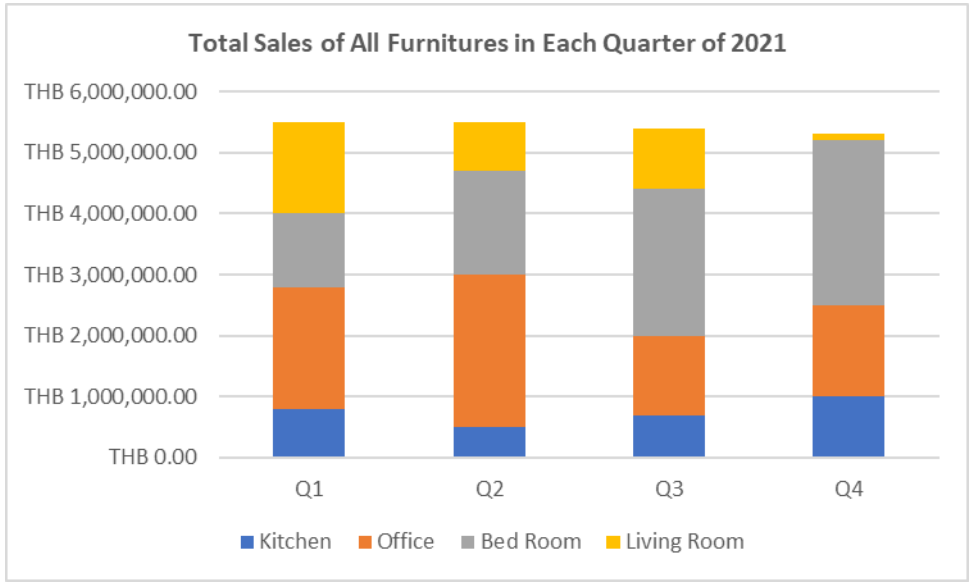




Source: https://www.eesi.org/topics/fossil-fuels/description

CREATE STACKED COLUMN CHART IN MS EXCEL

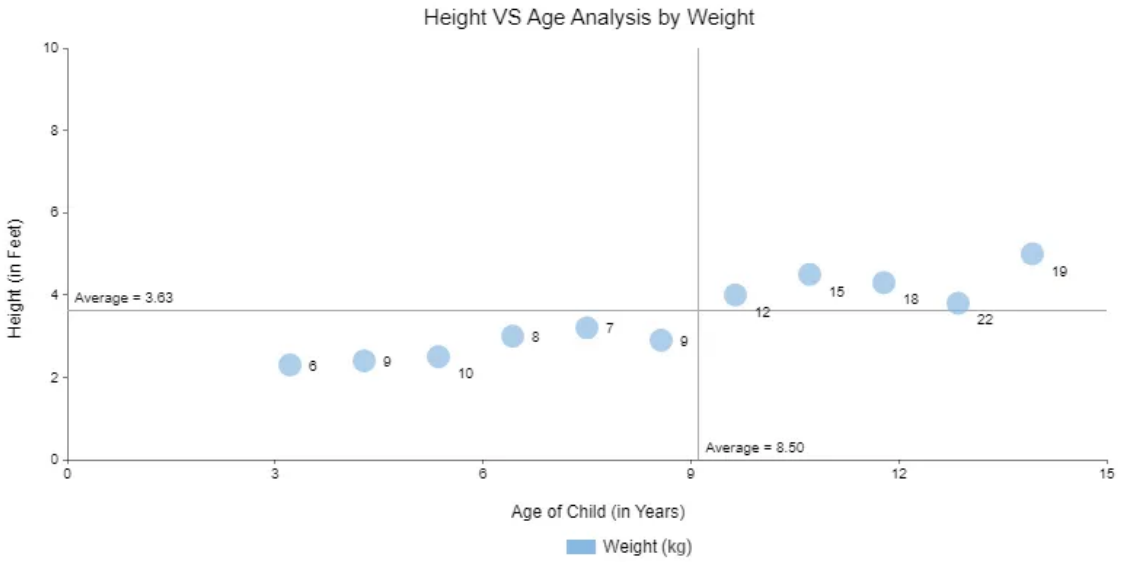




SCATTER PLOT

- Is a visualization design that uses Cartesian coordinates to display values in dots.
- Explore the relationship between a few variables

Bubble charts are enhanced versions of scatter plots by varying the size and/or color of the circles to tadd additional data dimensions.

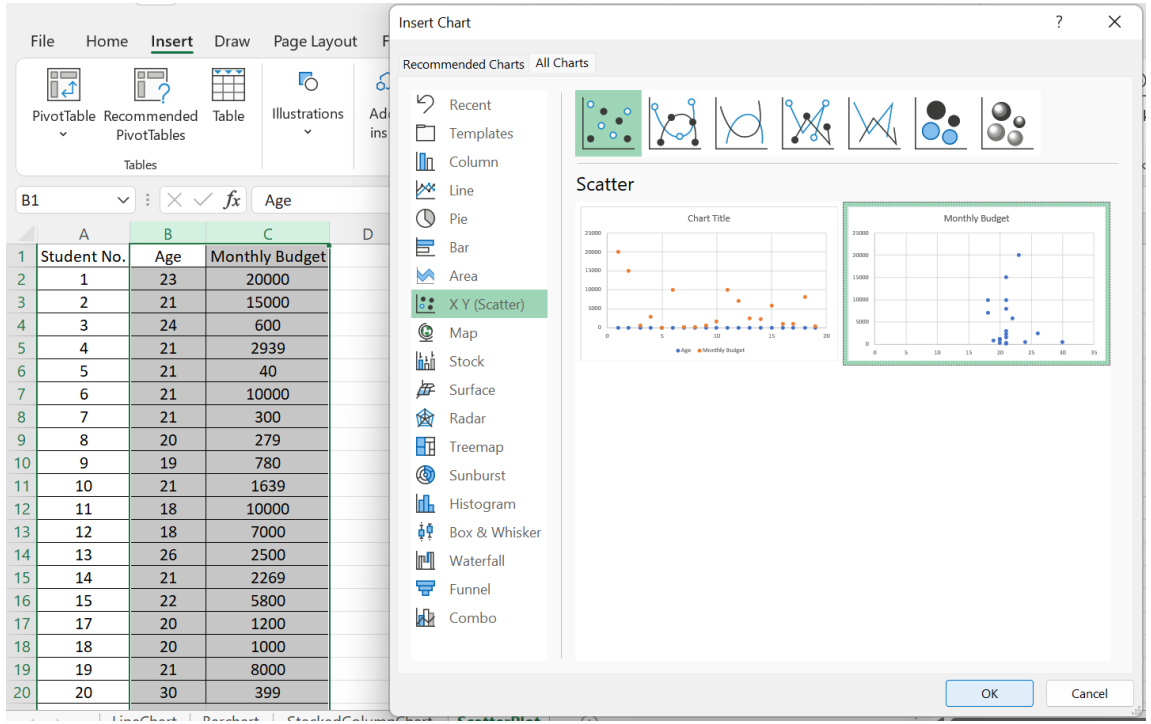


Note. Weight in Kg. is represented by size of circles.

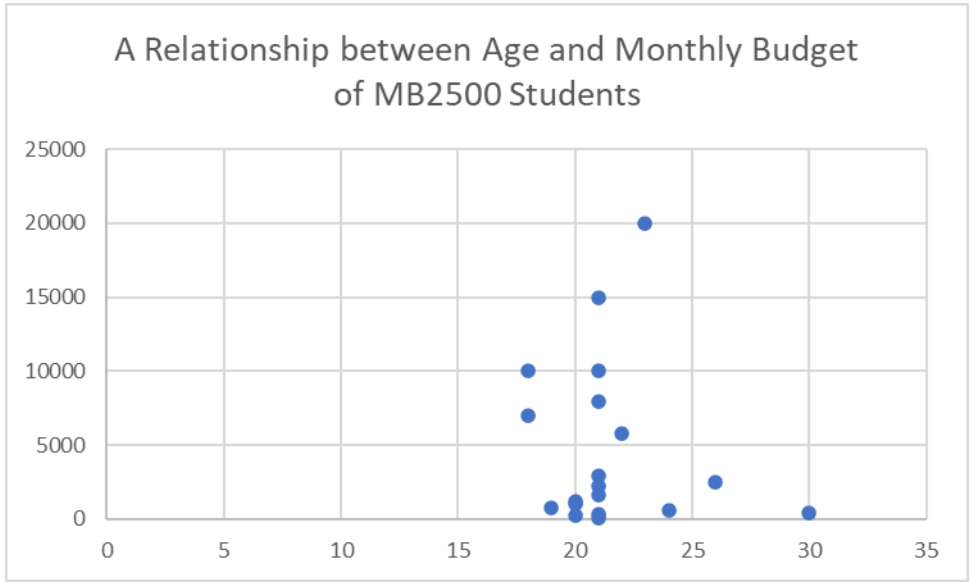
Note. No. of orders is represented by size of circles.

Ref. https://ppcexpo.com/blog/scatter-plot-examples

CREATE SCATTER PLOT IN MS EXCEL

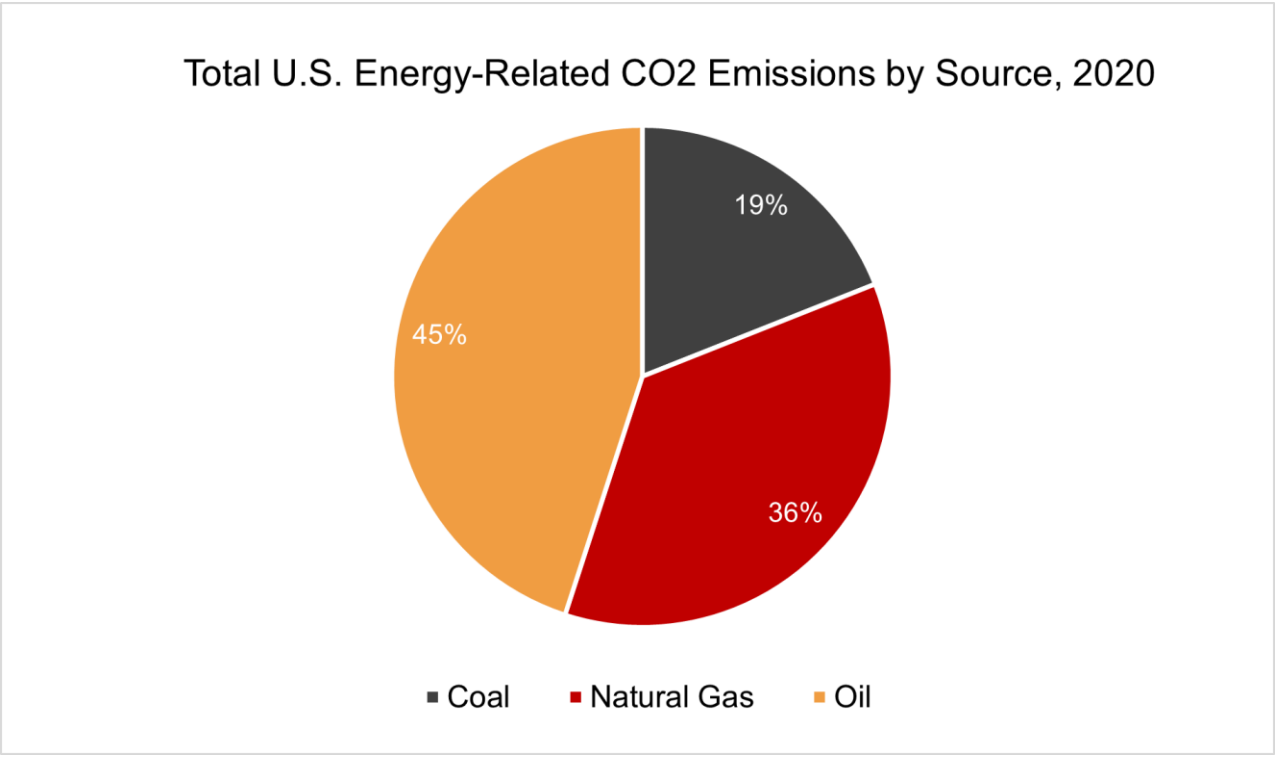


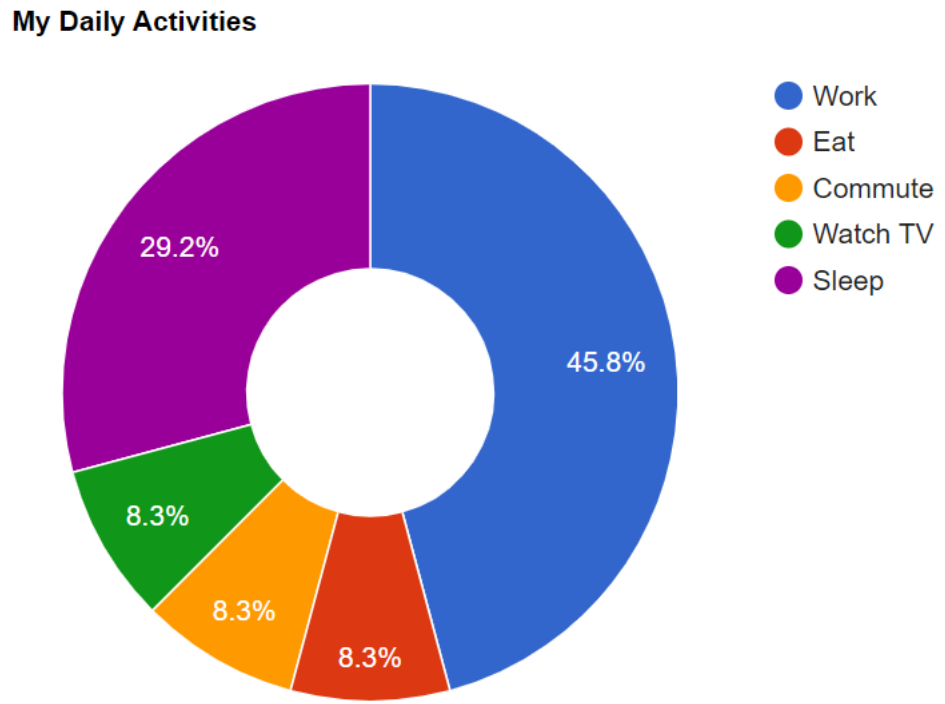
	A	B	C
1	Student No.	Age	Monthly Budget
2	1	23	20000
3	2	21	15000
4	3	24	600
5	4	21	2939
6	5	21	40
7	6	21	10000
8	7	21	300
9	8	20	279
10	9	19	780
11	10	21	1639
12	11	18	10000
13	12	18	7000
14	13	26	2500
15	14	21	2269
16	15	22	5800
17	17	20	1200
18	18	20	1000
19	19	21	8000
20	20	30	399



PIE CHART

- Illustrate relative proportions of a specific measure.





Source: https://developers.google.com/chart/interactive/docs/gallery/piechart
https://www.eesi.org/topics/fossil-fuels/description

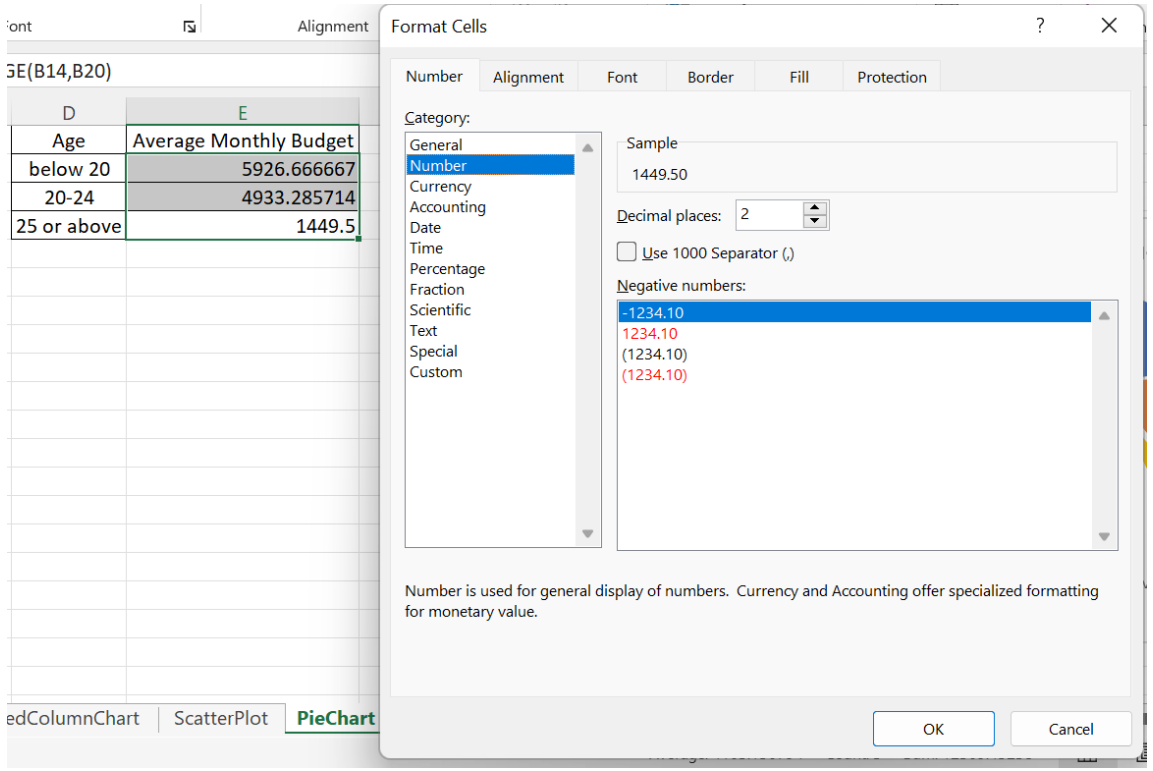
CREATE PIE CHART IN MS EXCEL – 1/2

	A	B	C	D	E
1	Age	Monthly Budget		Age	Average Monthly Budget
2	20-24	20000		below 20	5926.666667
3	20-24	15000		20-24	4933.285714
4	20-24	600		25 or above	1449.5
5	20-24	2939			
6	20-24	40			
7	20-24	10000			
8	20-24	300			
9	20-24	279			
10	below 20	780			
11	20-24	1639			
12	below 20	10000			
13	below 20	7000			
14	25 or above	2500			
15	20-24	2269			
16	20-24	5800			
17	20-24	1200			
18	20-24	1000			
19	20-24	8000			
20	25 or above	399			

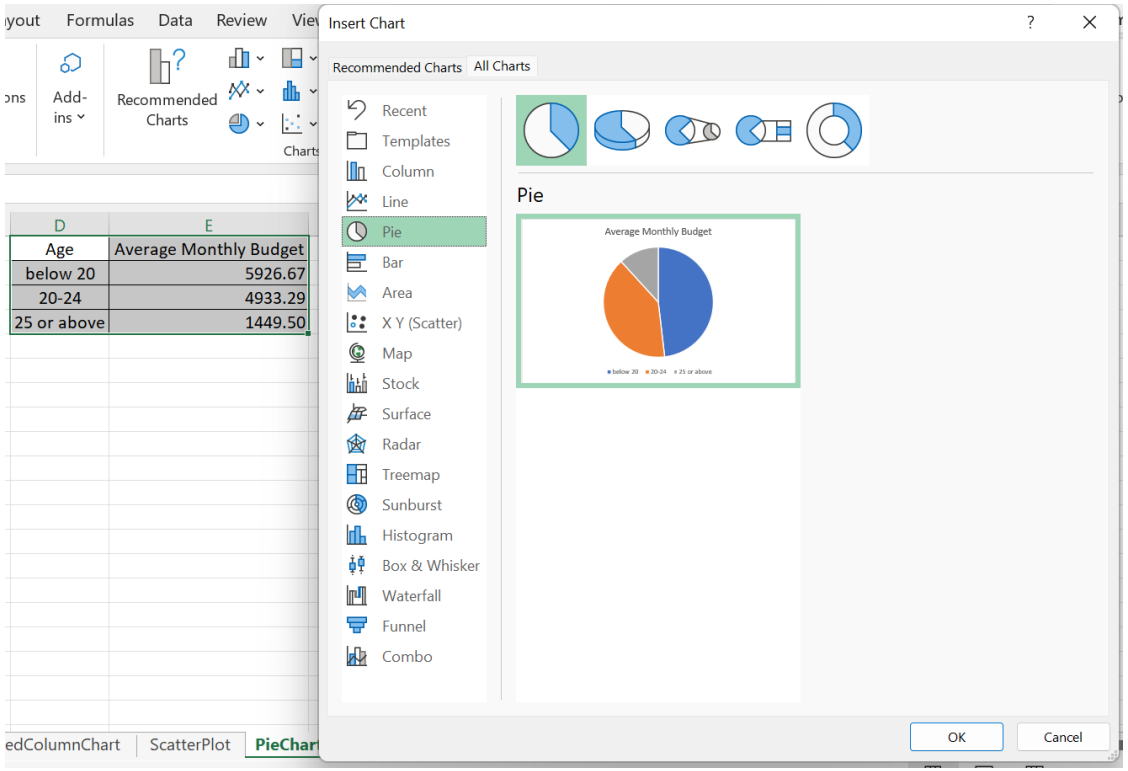
=AVERAGE(B10,B12,B13)

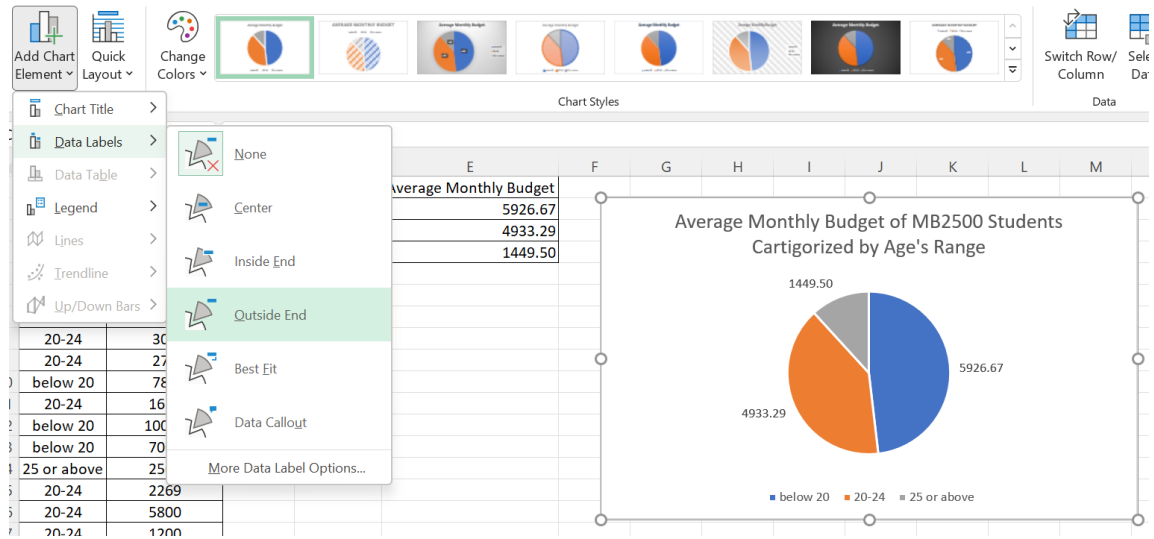
=AVERAGE(B2:B9,B11,B15:B19)

=AVERAGE(B14,B20)



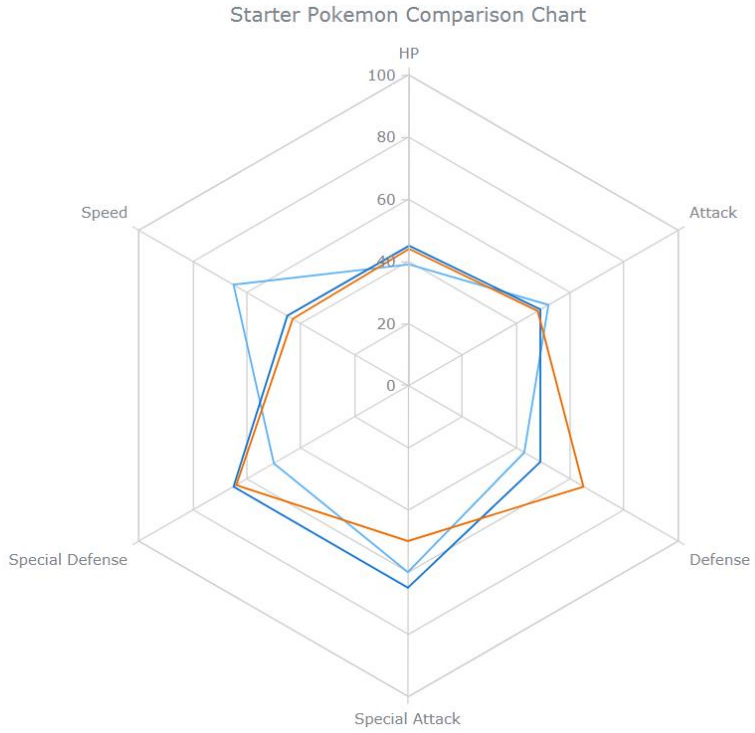
CREATE PIE CHART IN MS EXCEL – 2/2





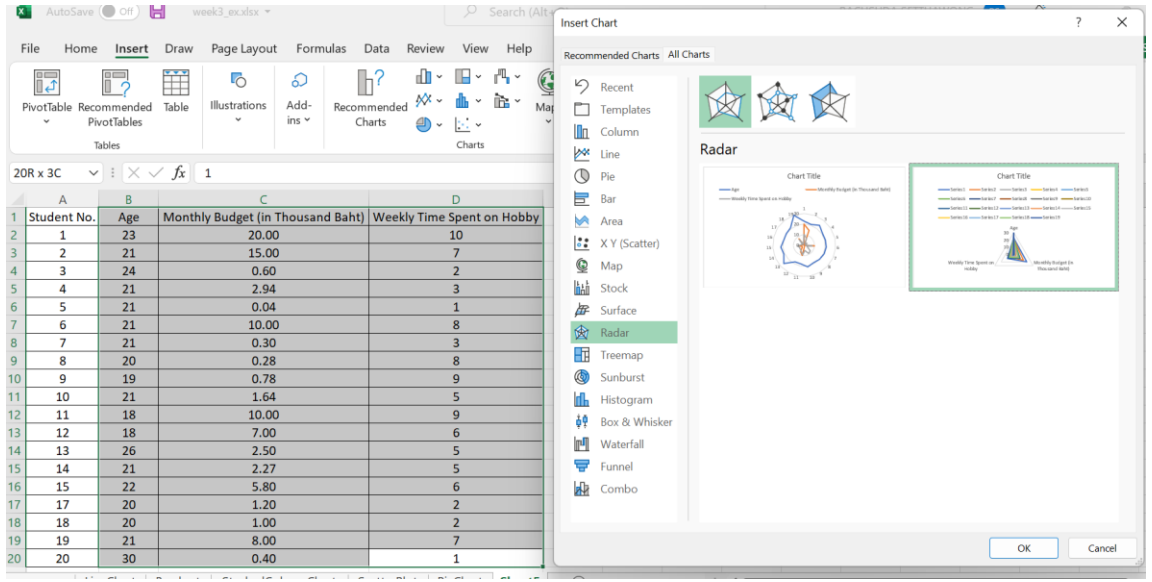
SPIDER CHART (RADAR PLOT)

- Compare observations with multiple quantitative variables
  - Each variable is encoded to a spoke which are equidistant apart
  - The higher the value, the further away from the center of the chart the point is made

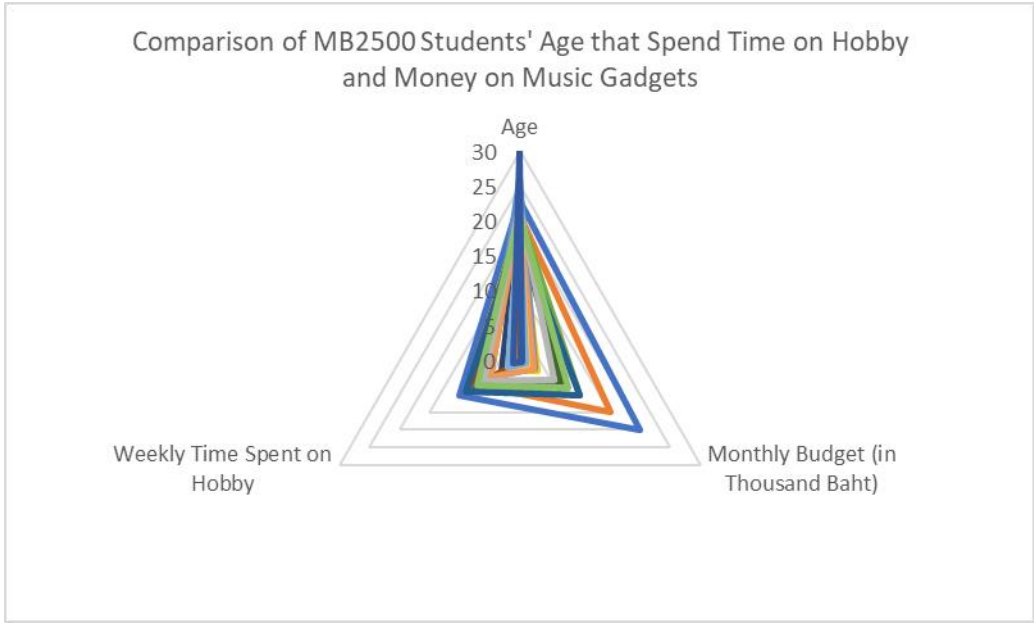


Source: https://www.anychart.com/blog/2020/07/07/radar-chart-js/

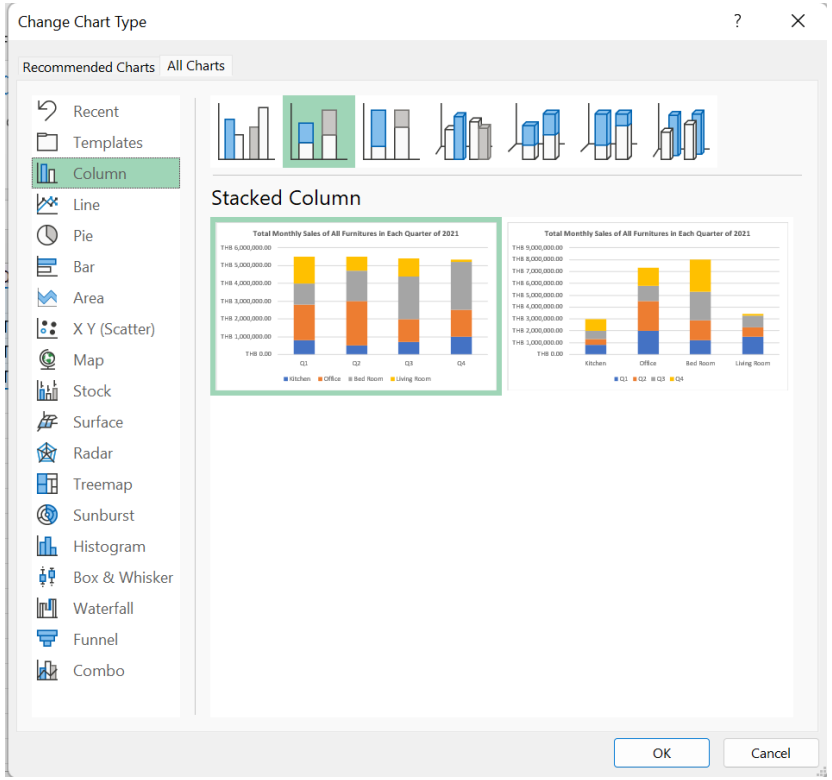
CREATE SPIDER CHART IN MS EXCEL



Student No.	Age	Monthly Budget (in Thousand Baht)	Weekly Time Spent on Hobby
1	23	20.00	10
2	21	15.00	7
3	24	0.60	2
4	21	2.94	3
5	21	0.04	1
6	21	10.00	8
7	21	0.30	3
8	20	0.28	8
9	19	0.78	9
10	21	1.64	5
11	18	10.00	9
12	18	7.00	6
13	26	2.50	5
14	21	2.27	5
15	22	5.80	6
17	20	1.20	2
18	20	1.00	2
19	21	8.00	7
20	30	0.40	1

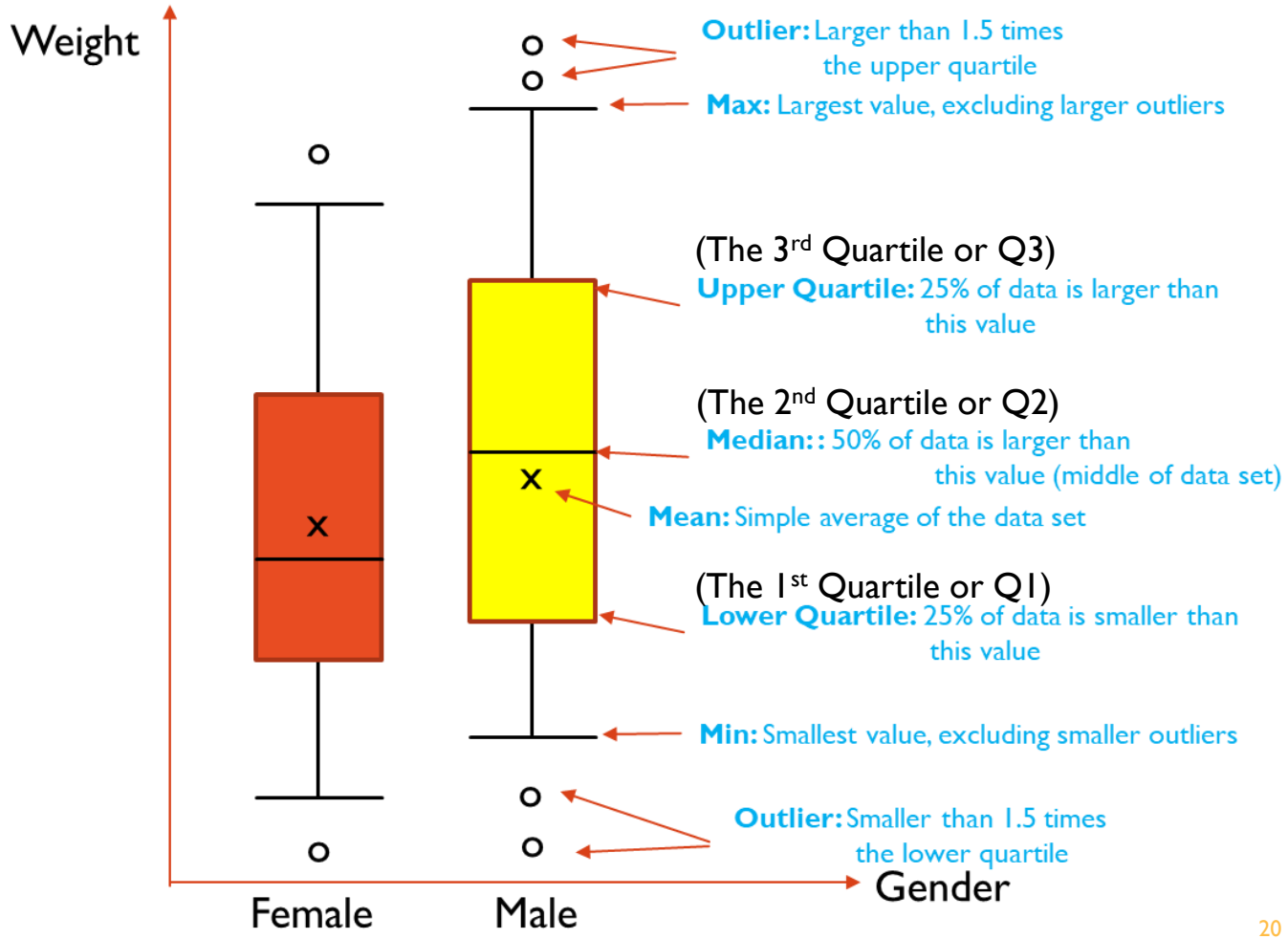


MORE ON CHARTS/GRAPHS



BOX-AND-WHISKERS PLOT

- A graphical illustration of several descriptive statistics about a given dataset.
  - Centrality
  - Dispersion
  - Minimum and maximum ranges

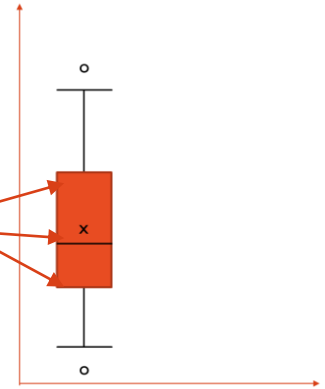


CALCULATE QUARTILES (Q1, Q2, Q3)

- **Position**: 1, 2, 3, 4, 5, 6, 7, 8, 9
- **Dataset**: {1, 2, 2, 3, 3, 4, 5, 5, 10}
  - Data must be sorted in ascending order first!

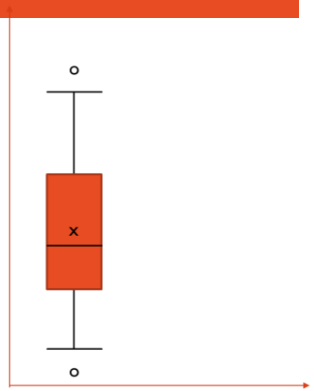
Position of the value in the sorted data

- **The 1st quartile (Q1)** = $[(1/4)*(n+1)]^{th}$ = $[(1/4)*(9+1)]^{th}$ = 2.5^{th} ∴ (2+2)/2 = 2
- **The 2nd quartile (Q2)** = $[(2/4)*(n+1)]^{th}$ = $[(2/4)*(9+1)]^{th}$ = 5^{th} = 3
- **The 3rd quartile (Q3)** = $[(3/4)*(n+1)]^{th}$ = $[(3/4)*(9+1)]^{th}$ = 7.5^{th} ∴ (5+5)/2 = 5



A calculator program for all quartiles: https://www.calculatorsoup.com/calculators/statistics/quartile-calculator.php

DETERMINE (WEAK) OUTLIERS



- **Dataset**: {1, 2, 2, 3, 3, 4, 5, 5, 10}
- **Interquartile** = 3rd quartile - 1st quartile = 5 - 2 = 3
- Multiplying the interquartile range (IQR) by 1.5 to determine whether a certain value is an outlier.
  - If we **subtract** 1.5 x IQR from the **first quartile**, any data values that are **less than** this number are considered outliers.
    - 2 – (1.5 * 3) = 2 - 4.5 = -2.5
  - If we **add** 1.5 x IQR from the **third quartile**, any data values that are **greater than** this number are considered outliers.
    - 5 + (1.5 * 3) = 5 + 4.5 = 9.5

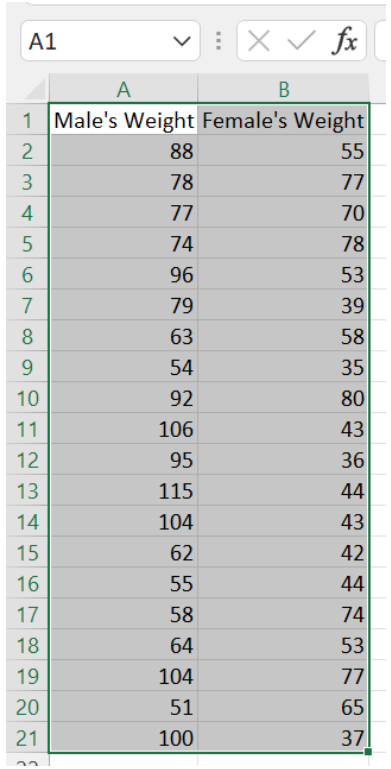
∴ The value 10 in the dataset is an outlier.

*Note: Multiplying the interquartile range (IQR) by **3.0** to determine **strong outliers***

A calculator program for all quartiles: https://www.calculatorsoup.com/calculators/statistics/quartile-calculator.php

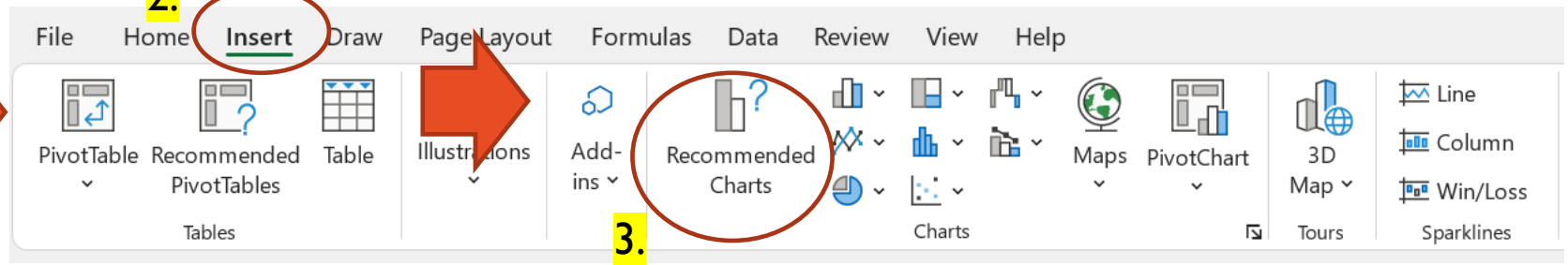
CREATE BOX-AND-WHISKERS PLOT USING MS EXCEL – 1/5

1. Select Data

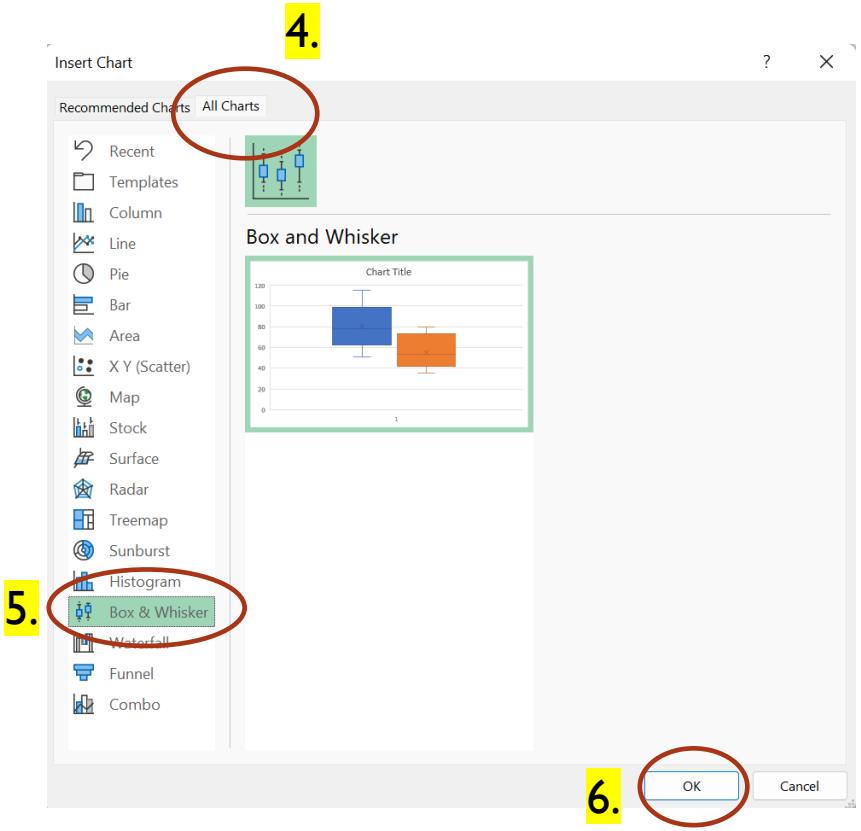


	A	B
1	Male's Weight	Female's Weight
2	88	55
3	78	77
4	77	70
5	74	78
6	96	53
7	79	39
8	63	58
9	54	35
10	92	80
11	106	43
12	95	36
13	115	44
14	104	43
15	62	42
16	55	44
17	58	74
18	64	53
19	104	77
20	51	65
21	100	37

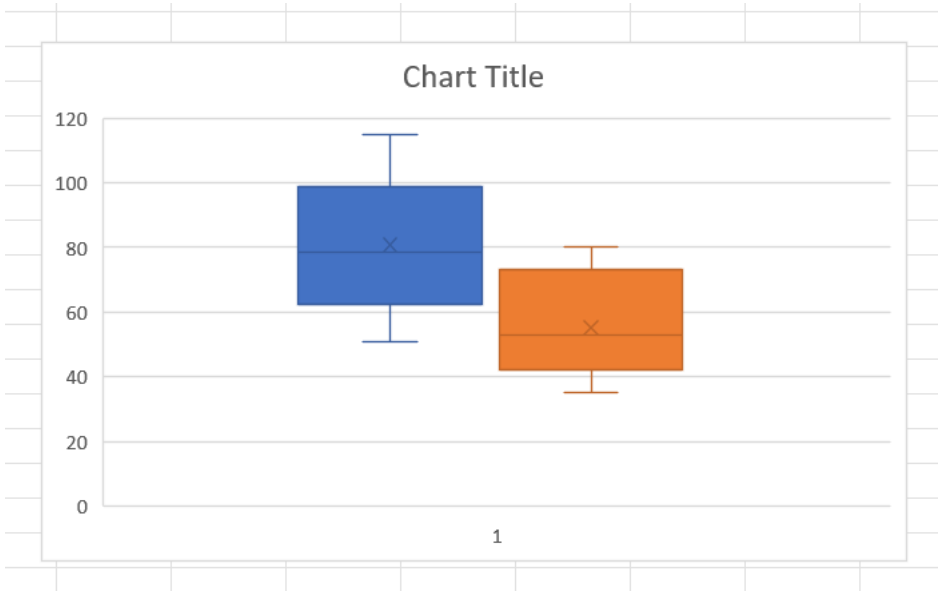
2.



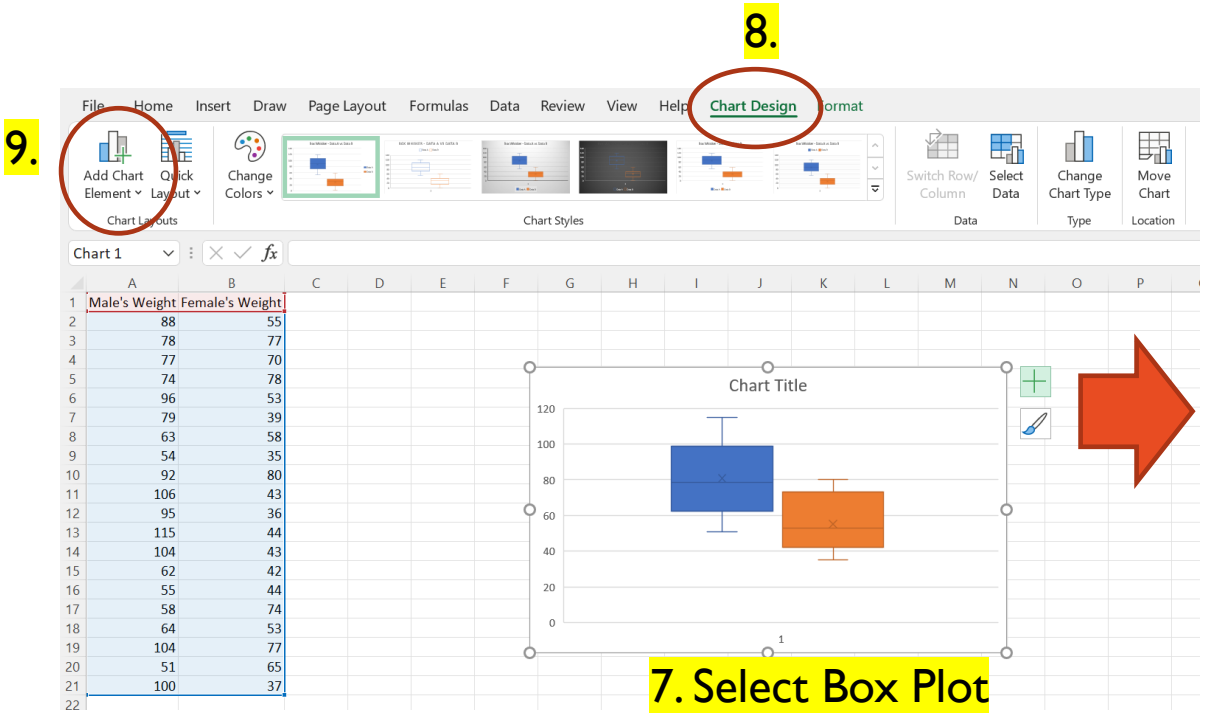
CREATE BOX-AND-WHISKERS PLOT USING MS EXCEL – 2/5

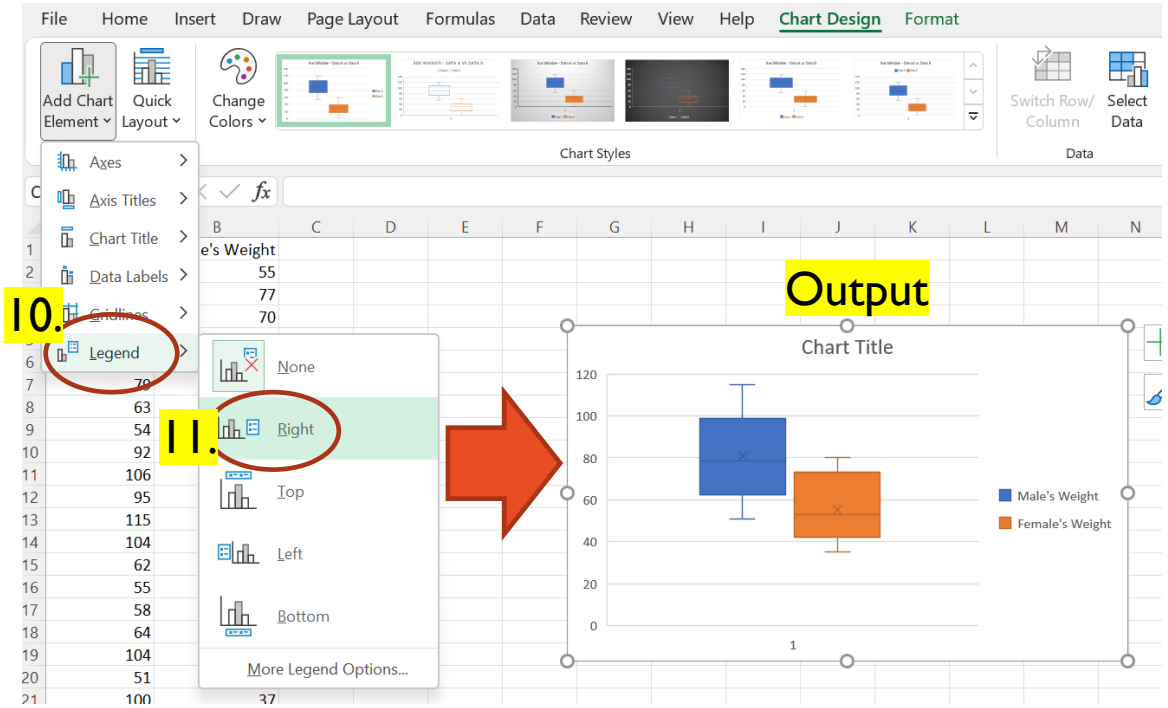


Output

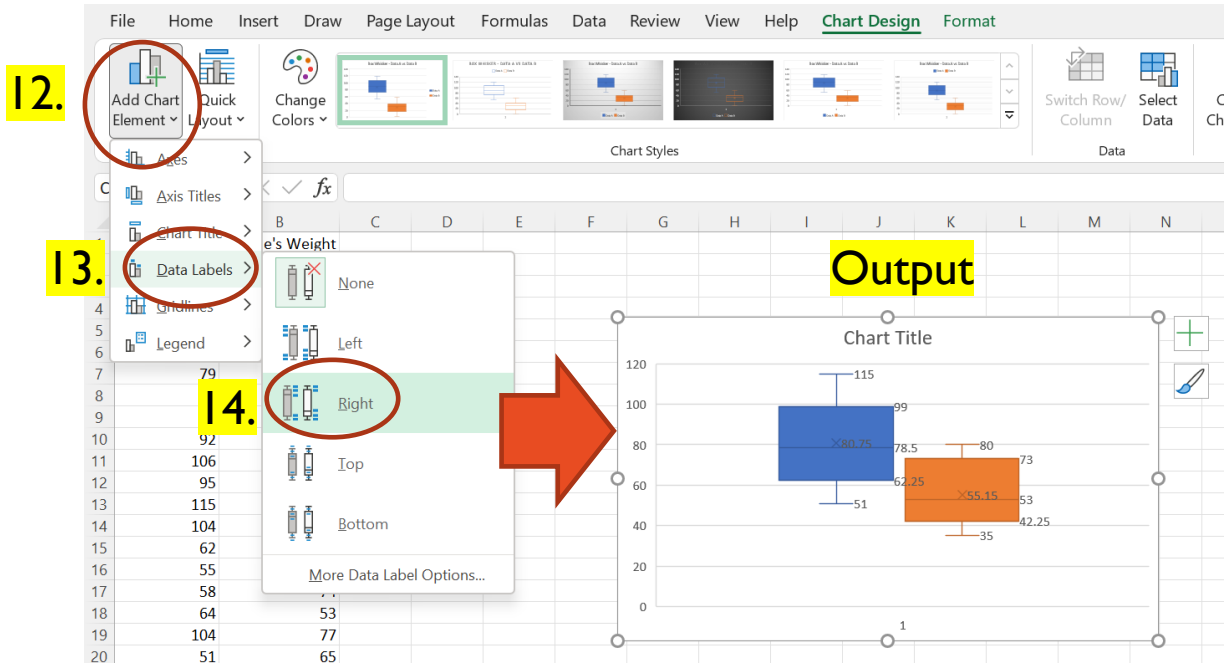


CREATE BOX-AND-WHISKERS PLOT USING MS EXCEL – 3/5



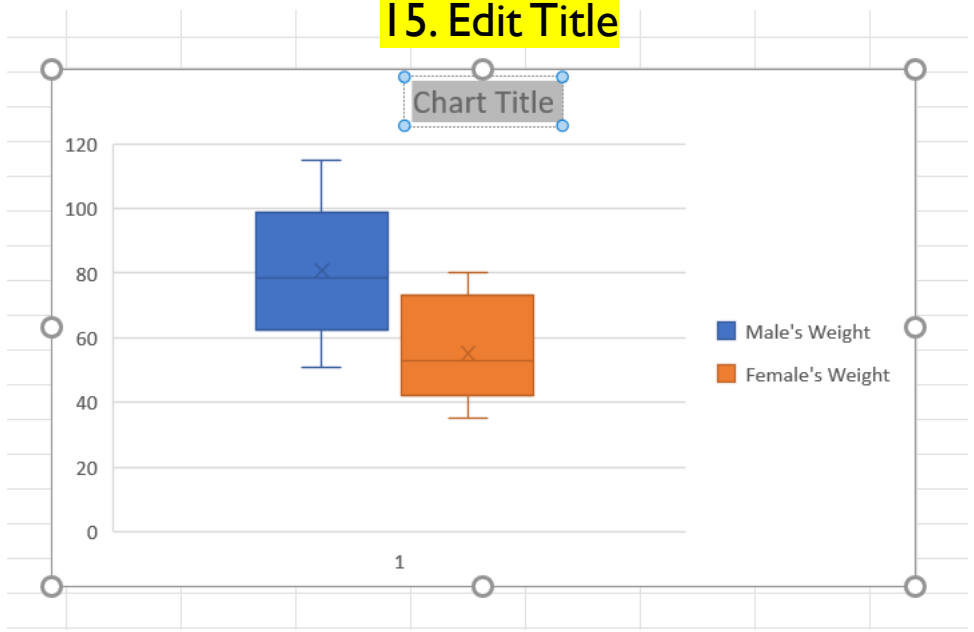


CREATE BOX-AND-WHISKERS PLOT USING MS EXCEL – 4/5



CREATE BOX-AND-WHISKERS PLOT USING MS EXCEL – 5/5

15. Edit Title



Output