Working with multiple groups: demo notes

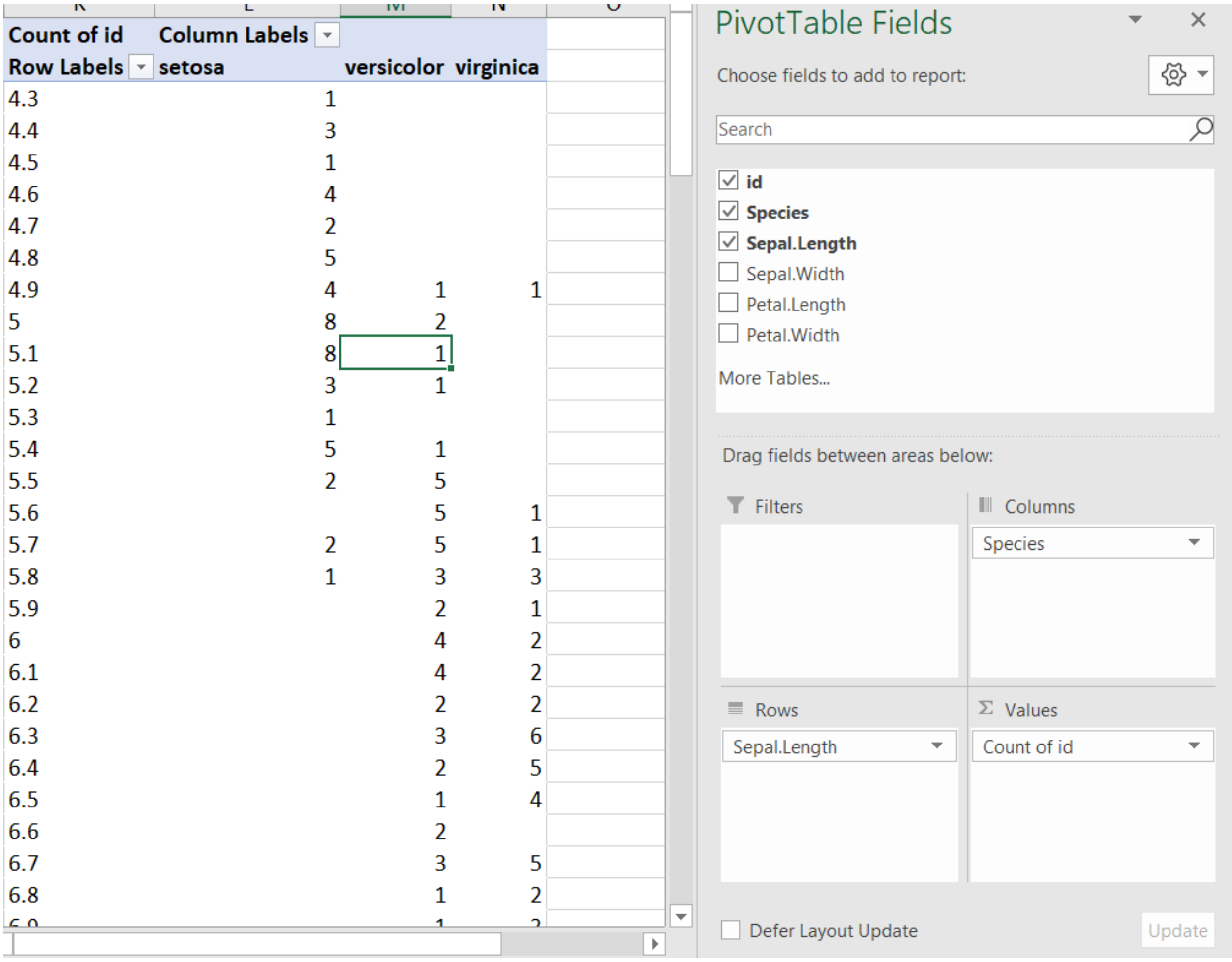
Visualizing distributions: histograms and box plots

File: `iris-viz.xlsx`

Histograms

We will visualize the distribution of sepal length for each species.

1. Insert a PivotTable from the source data. Place `Species` in the Columns section, `Sepal.Length` in the Rows section and `Count of id` in the Values section.



Count of id	Column Labels		
Row Labels	setosa	versicolor	virginica
4.3	1		
4.4	3		
4.5	1		
4.6	4		
4.7	2		
4.8	5		
4.9	4	1	1
5	8	2	
5.1	8	1	
5.2	3	1	
5.3	1		
5.4	5	1	
5.5	2	5	
5.6		5	1
5.7	2	5	1
5.8	1	3	3
5.9		2	1
6		4	2
6.1		4	2
6.2		2	2
6.3		3	6
6.4		2	5
6.5		1	4
6.6		2	
6.7		3	5
6.8		1	2
6.9		1	3

PivotTable Fields

Choose fields to add to report:

Search

- ☑ id
- ☑ Species
- ☑ Sepal.Length
- □ Sepal.Width
- □ Petal.Length
- □ Petal.Width

More Tables...

Drag fields between areas below:

Filters

Columns: Species

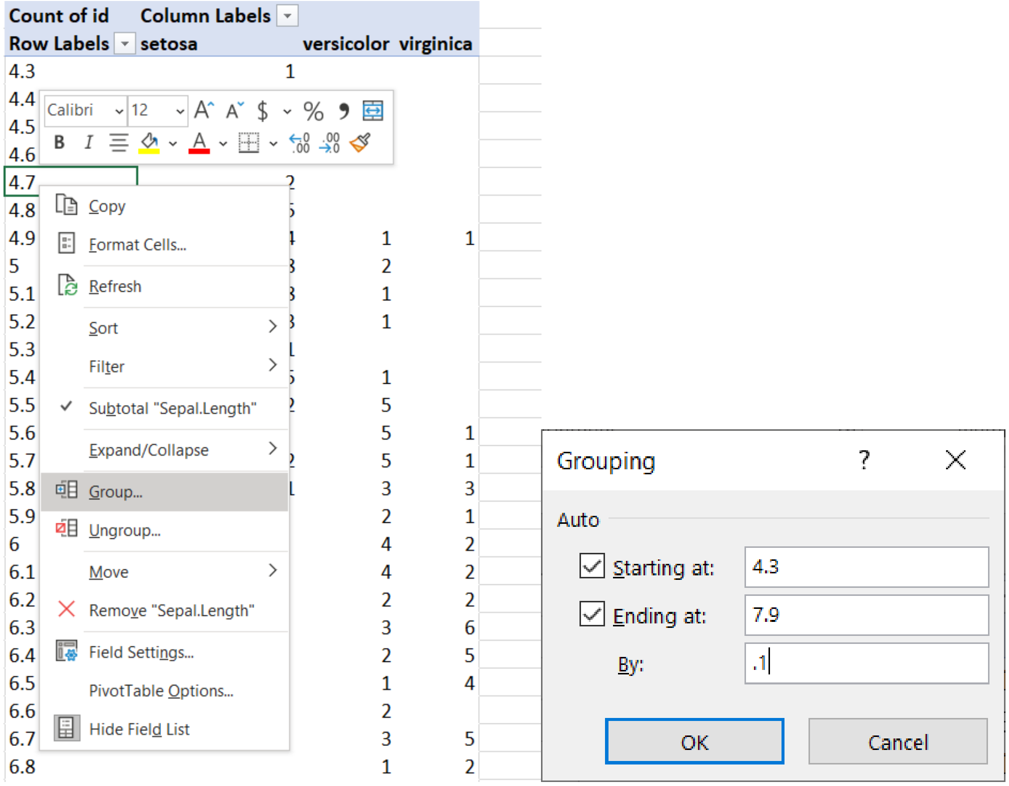
Rows: Sepal.Length

Values: Count of id

□ Defer Layout Update

Update

2. Right-click on the Row Labels and select Group. Group the variable at intervals of .1.



Count of id	Column Labels		
Row Labels	setosa	versicolor	virginica
4.3	1		
4.4			
4.5			
4.6			
4.7	2		
4.8	5		
4.9	4	1	1
5	8	2	
5.1	8	1	
5.2	3	1	
5.3	1		
5.4	5	1	
5.5	2	5	
5.6		5	1
5.7	2	5	1
5.8	1	3	3
5.9		2	1
6		4	2
6.1		4	2
6.2		2	2
6.3		3	6
6.4		2	5
6.5		1	4
6.6		2	
6.7		3	5
6.8		1	2

Grouping

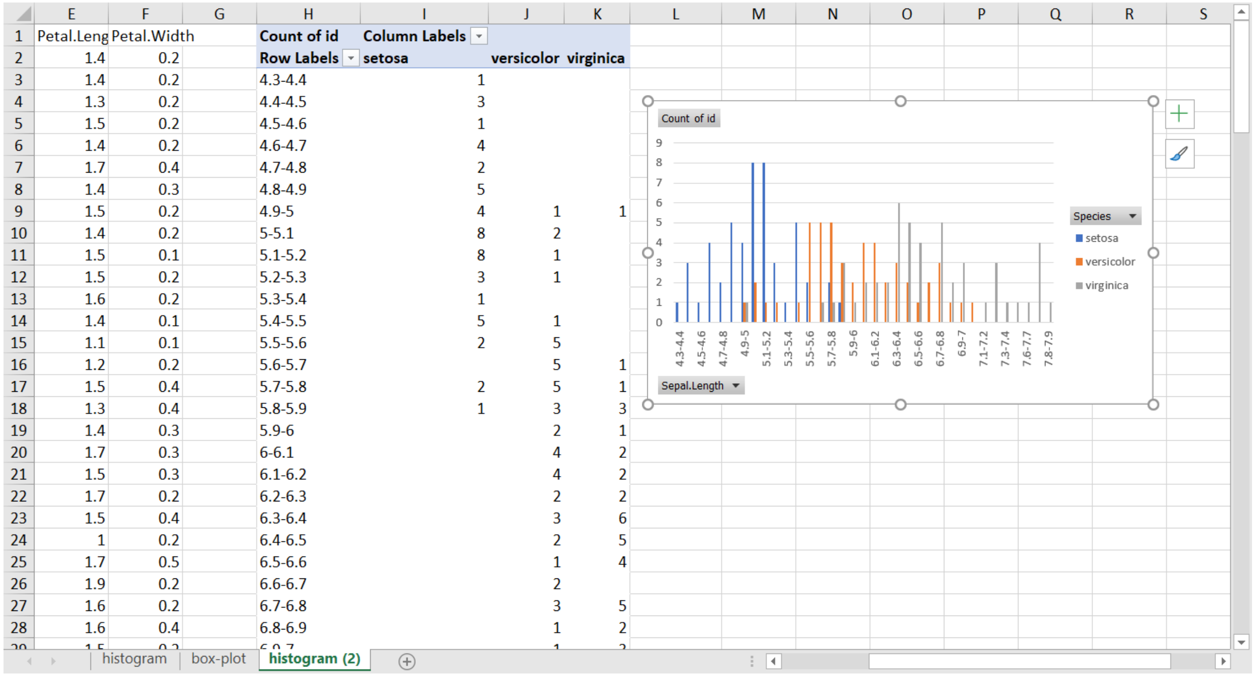
Auto

☑ Starting at: 4.3

☑ Ending at: 7.9

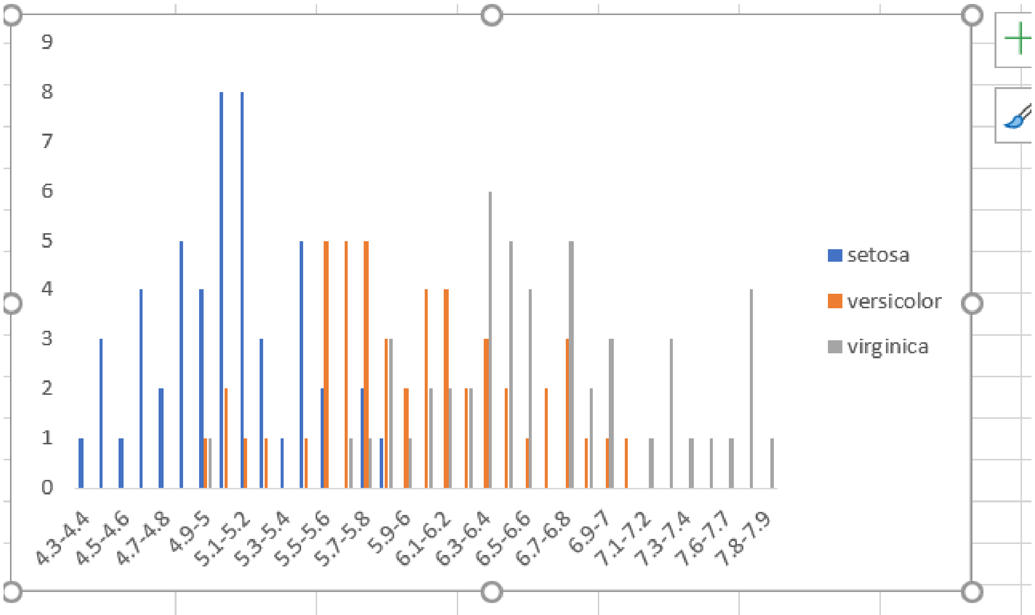
By: .1

3. Insert the recommended chart: clustered column.

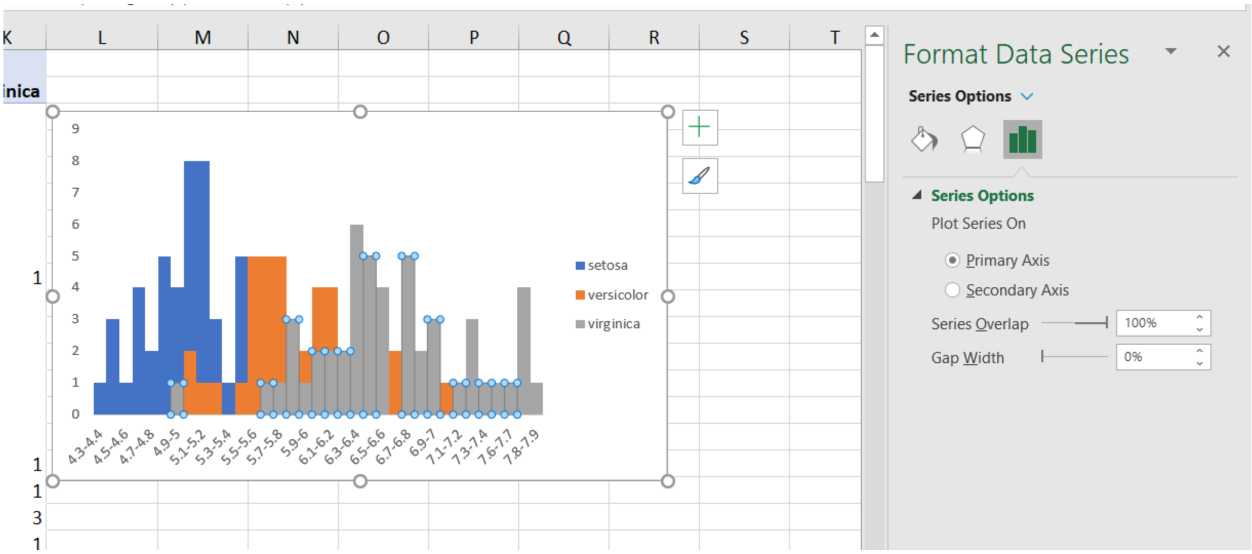


	E	F	G	H	I	J	K
1	Petal.Leng	Petal.Width		Count of id	Column Labels		
2	1.4	0.2		Row Labels	setosa	versicolor	virginica
3	1.4	0.2		4.3-4.4	1		
4	1.3	0.2		4.4-4.5	3		
5	1.5	0.2		4.5-4.6	1		
6	1.4	0.2		4.6-4.7	4		
7	1.7	0.4		4.7-4.8	2		
8	1.4	0.3		4.8-4.9	5		
9	1.5	0.2		4.9-5	4	1	1
10	1.4	0.2		5-5.1	8	2	
11	1.5	0.1		5.1-5.2	8	1	
12	1.5	0.2		5.2-5.3	3	1	
13	1.6	0.2		5.3-5.4	1		
14	1.4	0.1		5.4-5.5	5	1	
15	1.1	0.1		5.5-5.6	2	5	
16	1.2	0.2		5.6-5.7		5	1
17	1.5	0.4		5.7-5.8	2	5	1
18	1.3	0.4		5.8-5.9	1	3	3
19	1.4	0.3		5.9-6		2	1
20	1.7	0.3		6-6.1		4	2
21	1.5	0.3		6.1-6.2		4	2
22	1.7	0.2		6.2-6.3		2	2
23	1.5	0.4		6.3-6.4		3	6
24	1	0.2		6.4-6.5		2	5
25	1.7	0.5		6.5-6.6		1	4
26	1.9	0.2		6.6-6.7		2	
27	1.6	0.2		6.7-6.8		3	5
28	1.6	0.4		6.8-6.9		1	2

4. Clean up this chart by right-clicking on any of the labels and selecting "Hide All Field Buttons on Chart." You can also remove the chart gridlines by clicking on any of them and pressing the Delete key.



5. Right-click on any of the bars and select "Format Data Series." A menu will appear to the right. Set Series Overlap to 100% and Gap Width to 0%.



Format Data Series

Series Options

Series Options

Plot Series On

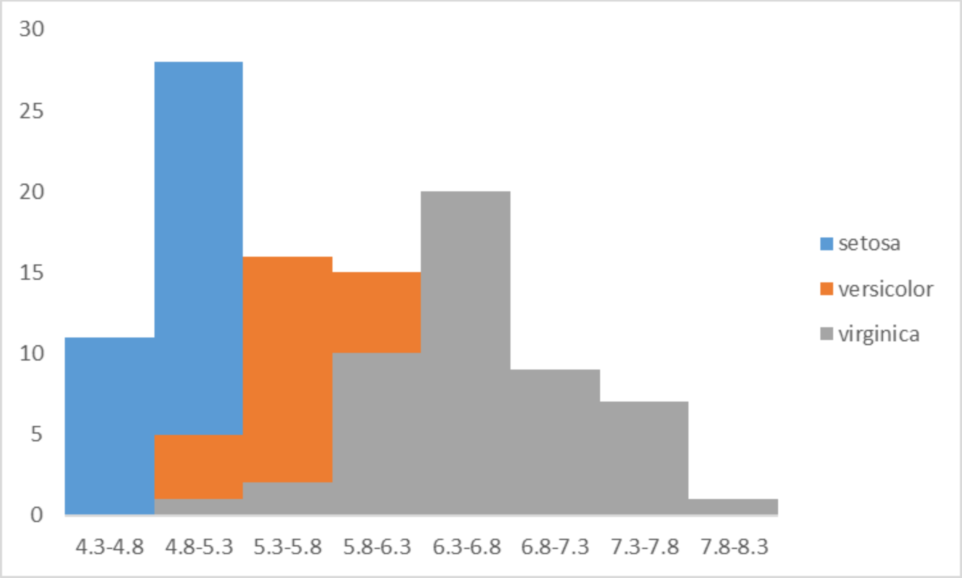
● Primary Axis

○ Secondary Axis

Series Overlap 100%

Gap Width 0%

6. You can resize the bins of the histogram by right-clicking back on the Row Labels of the PivotTable and selecting Group. What happens if we put it in intervals of .5?

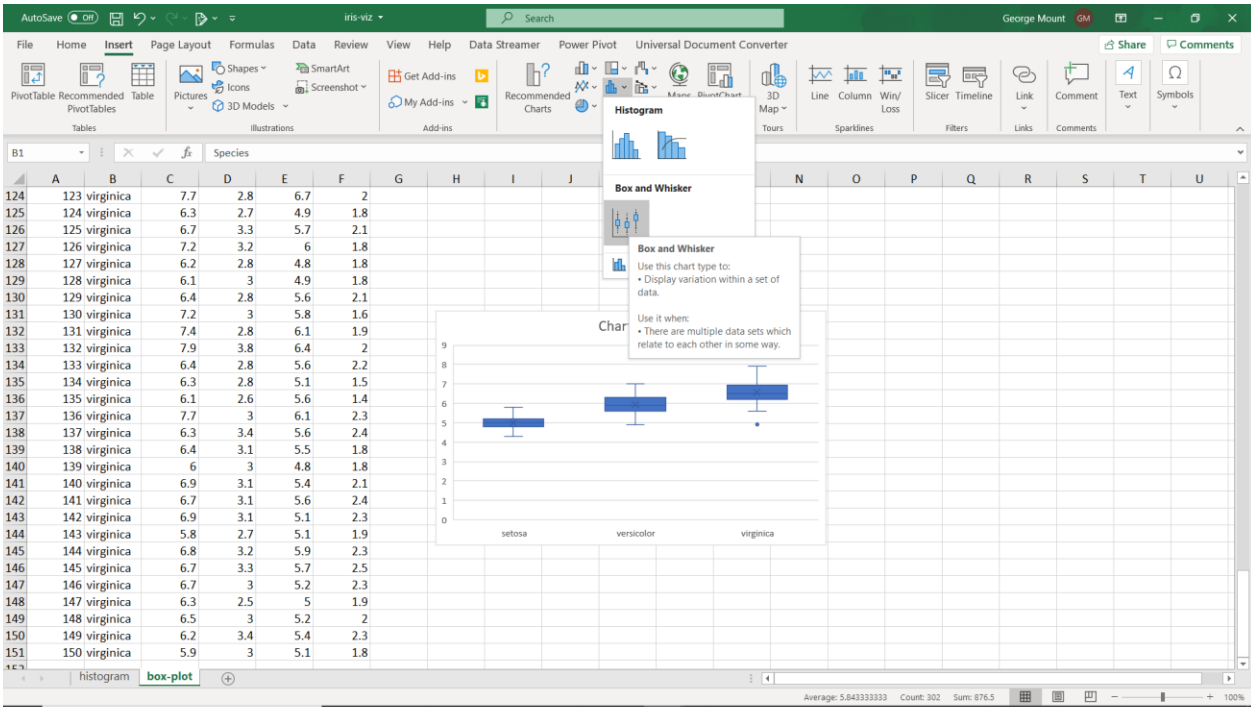


Box plots

Multiple histograms on the same chart can get messy. Let's try a different visualization: the box plot.

We will again plot the distributions of sepal length by species.

1. Select columns B-C and head to Insert > Chart. Under Histogram there will be an option, Box and Whisker.



B1 | Species

	A	B	C	D	E	F
124	123	virginica	7.7	2.8	6.7	2
125	124	virginica	6.3	2.7	4.9	1.8
126	125	virginica	6.7	3.3	5.7	2.1
127	126	virginica	7.2	3.2	6	1.8
128	127	virginica	6.2	2.8	4.8	1.8
129	128	virginica	6.1	3	4.9	1.8
130	129	virginica	6.4	2.8	5.6	2.1
131	130	virginica	7.2	3	5.8	1.6
132	131	virginica	7.4	2.8	6.1	1.9
133	132	virginica	7.9	3.8	6.4	2
134	133	virginica	6.4	2.8	5.6	2.2
135	134	virginica	6.3	2.8	5.1	1.5
136	135	virginica	6.1	2.6	5.6	1.4
137	136	virginica	7.7	3	6.1	2.3
138	137	virginica	6.3	3.4	5.6	2.4
139	138	virginica	6.4	3.1	5.5	1.8
140	139	virginica	6	3	4.8	1.8
141	140	virginica	6.9	3.1	5.4	2.1
142	141	virginica	6.7	3.1	5.6	2.4
143	142	virginica	6.9	3.1	5.1	2.3
144	143	virginica	5.8	2.7	5.1	1.9
145	144	virginica	6.8	3.2	5.9	2.3
146	145	virginica	6.7	3.3	5.7	2.5
147	146	virginica	6.7	3	5.2	2.3
148	147	virginica	6.3	2.5	5	1.9
149	148	virginica	6.5	3	5.2	2
150	149	virginica	6.2	3.4	5.4	2.3
151	150	virginica	5.9	3	5.1	1.8

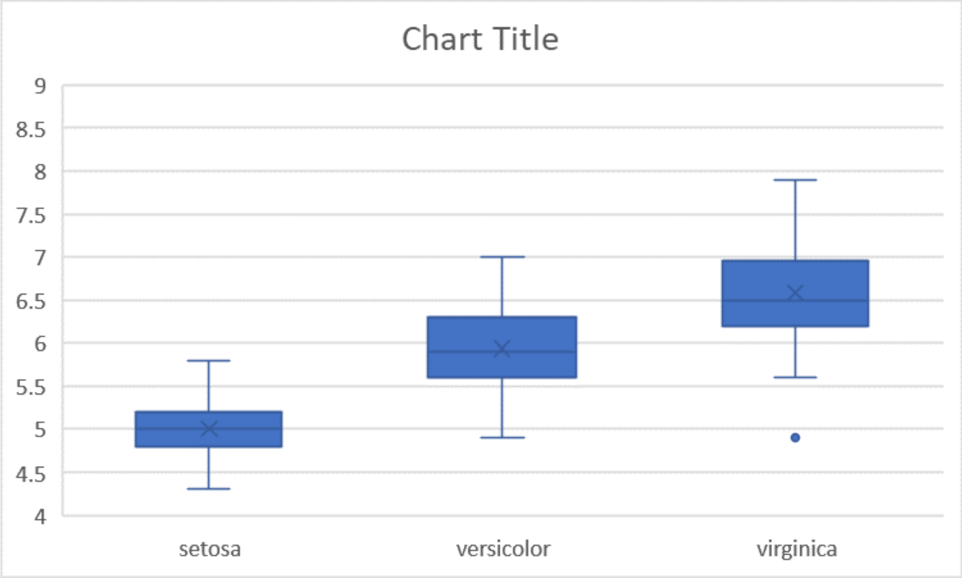
Box and Whisker

Use this chart type to:
- Display variation within a set of data.

Use it when:
- There are multiple data sets which relate to each other in some way.

Average: 5.843333333 Count: 302 Sum: 876.5 100%

2. Fortunately there is not too much more prep needed for this chart. We could re-set the y-axis to start at a value besides 0 (controversial, but sometimes useful).
   a. Right-click on the y-axis and select "Format Axis." You can now set the minimum bound to 4.



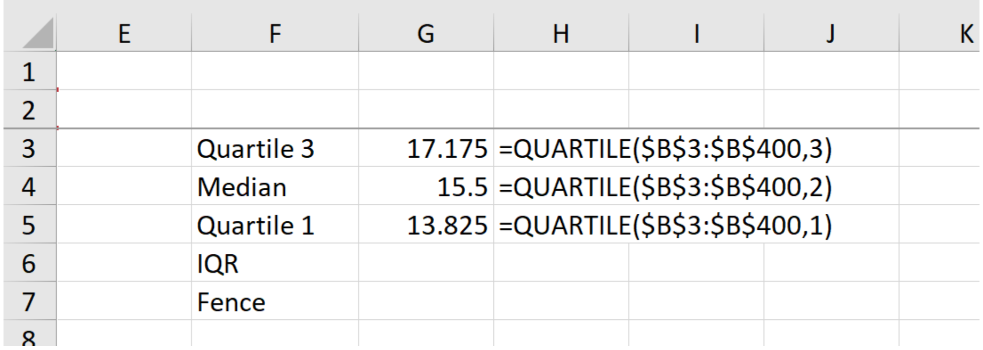
3. Take a look at the example box-and-whisker chart in the file to make sense of these distributions. What is the point under virginica doing there?
   a. Any datapoint that is 1.5 times the IQR is an outlier and excluded from the box and whisker plot.

Outlier detection

File: `outliers.xlsx`

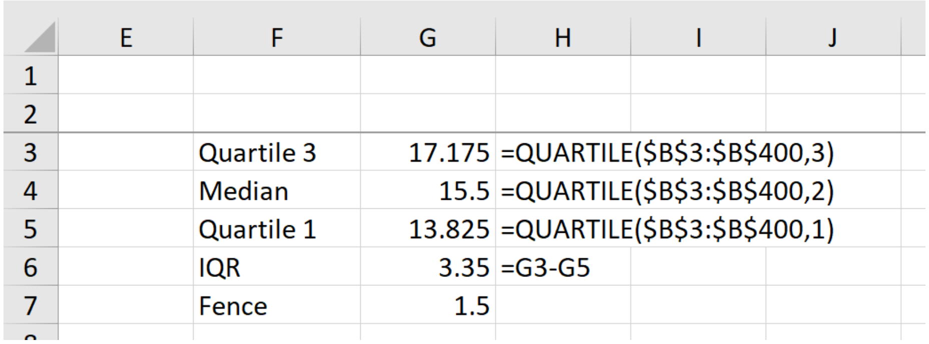
Let's calculate for ourselves the outlier range, and remove those datapoints from the box plot to see what happens.

1. To do this, we will calculate the 3rd, 2nd or *median*, and 1st quartiles of our acceleration data using the `QUARTILE()` function. We will pass in our data range, and what number quartile we want:



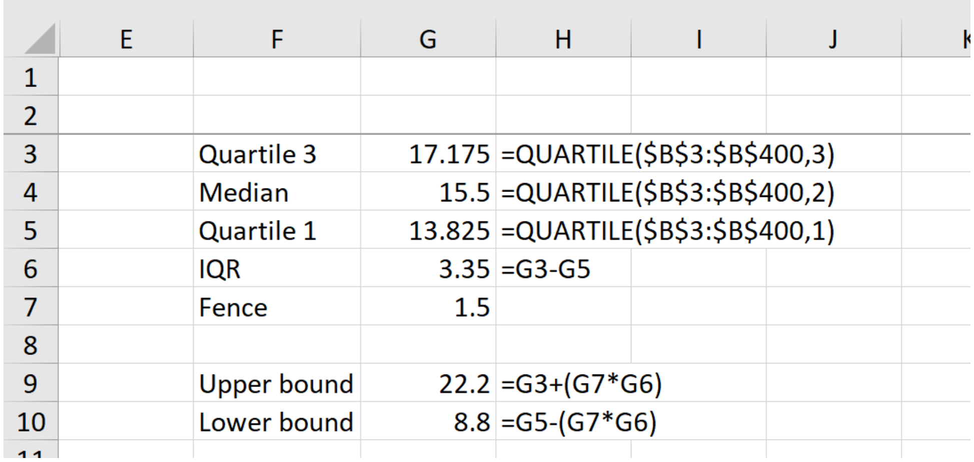
	E	F	G	H	I	J	K
1							
2							
3		Quartile 3	17.175	=QUARTILE($B$3:$B$400,3)			
4		Median	15.5	=QUARTILE($B$3:$B$400,2)			
5		Quartile 1	13.825	=QUARTILE($B$3:$B$400,1)			
6		IQR					
7		Fence					

2. We will now calculate the interquartile range (IQR) as the difference between the third and first quartiles.
    a. The fence will be 1.5. This is a hard-coded value that we will use to set the threshold for being an outlier.



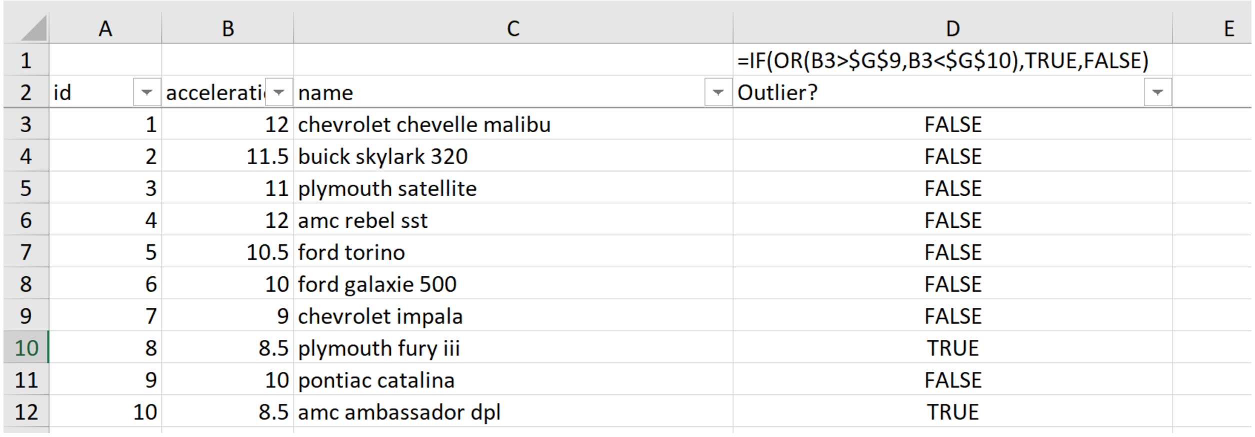
	E	F	G	H	I	J
1						
2						
3		Quartile 3	17.175	=QUARTILE($B$3:$B$400,3)		
4		Median	15.5	=QUARTILE($B$3:$B$400,2)		
5		Quartile 1	13.825	=QUARTILE($B$3:$B$400,1)		
6		IQR	3.35	=G3-G5		
7		Fence	1.5			

3. We will now calculate the outlier thresholds as 1.5 times the IQR, plus our upper bound and minus our lower bound respectively.



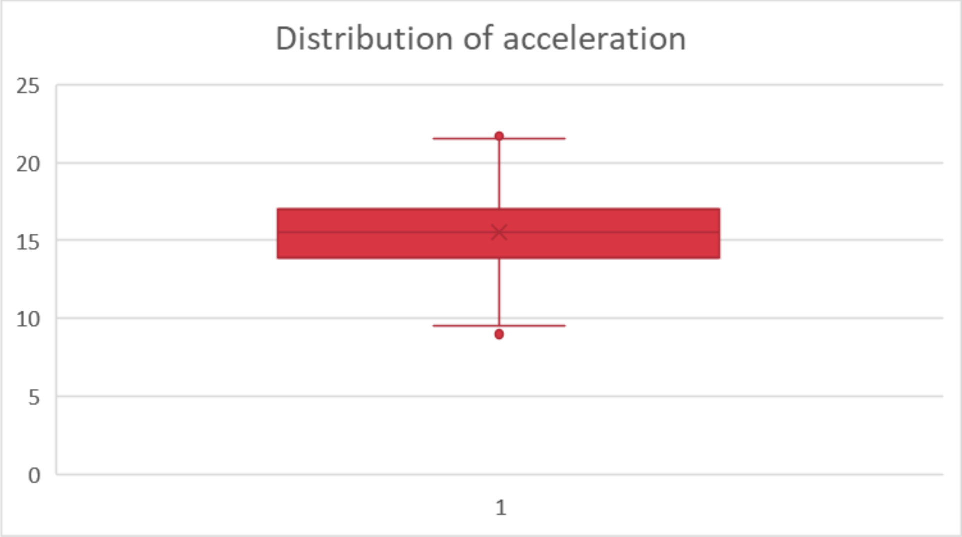
	E	F	G	H	I	J	K
1							
2							
3		Quartile 3	17.175	=QUARTILE($B$3:$B$400,3)			
4		Median	15.5	=QUARTILE($B$3:$B$400,2)			
5		Quartile 1	13.825	=QUARTILE($B$3:$B$400,1)			
6		IQR	3.35	=G3-G5			
7		Fence	1.5				
8							
9		Upper bound	22.2	=G3+(G7*G6)			
10		Lower bound	8.8	=G5-(G7*G6)			

4. We now know that any datapoint less than 8.8 or 22.2 is considered an outlier. We can use conditional logic to flag each value as TRUE or FALSE as being an outlier.



	A	B	C	D	E
1				=IF(OR(B3>$G$9,B3<$G$10),TRUE,FALSE)	
2	id	accelerati	name	Outlier?	
3	1	12	chevrolet chevelle malibu	FALSE	
4	2	11.5	buick skylark 320	FALSE	
5	3	11	plymouth satellite	FALSE	
6	4	12	amc rebel sst	FALSE	
7	5	10.5	ford torino	FALSE	
8	6	10	ford galaxie 500	FALSE	
9	7	9	chevrolet impala	FALSE	
10	8	8.5	plymouth fury iii	TRUE	
11	9	10	pontiac catalina	FALSE	
12	10	8.5	amc ambassador dpl	TRUE	

5. We can now filter our source data to exclude datapoints where `Outlier?` equals `TRUE`.
    a. The outliers have been removed from our box plot, with the exceptions of some datapoints that are right on the cusp of being outliers.

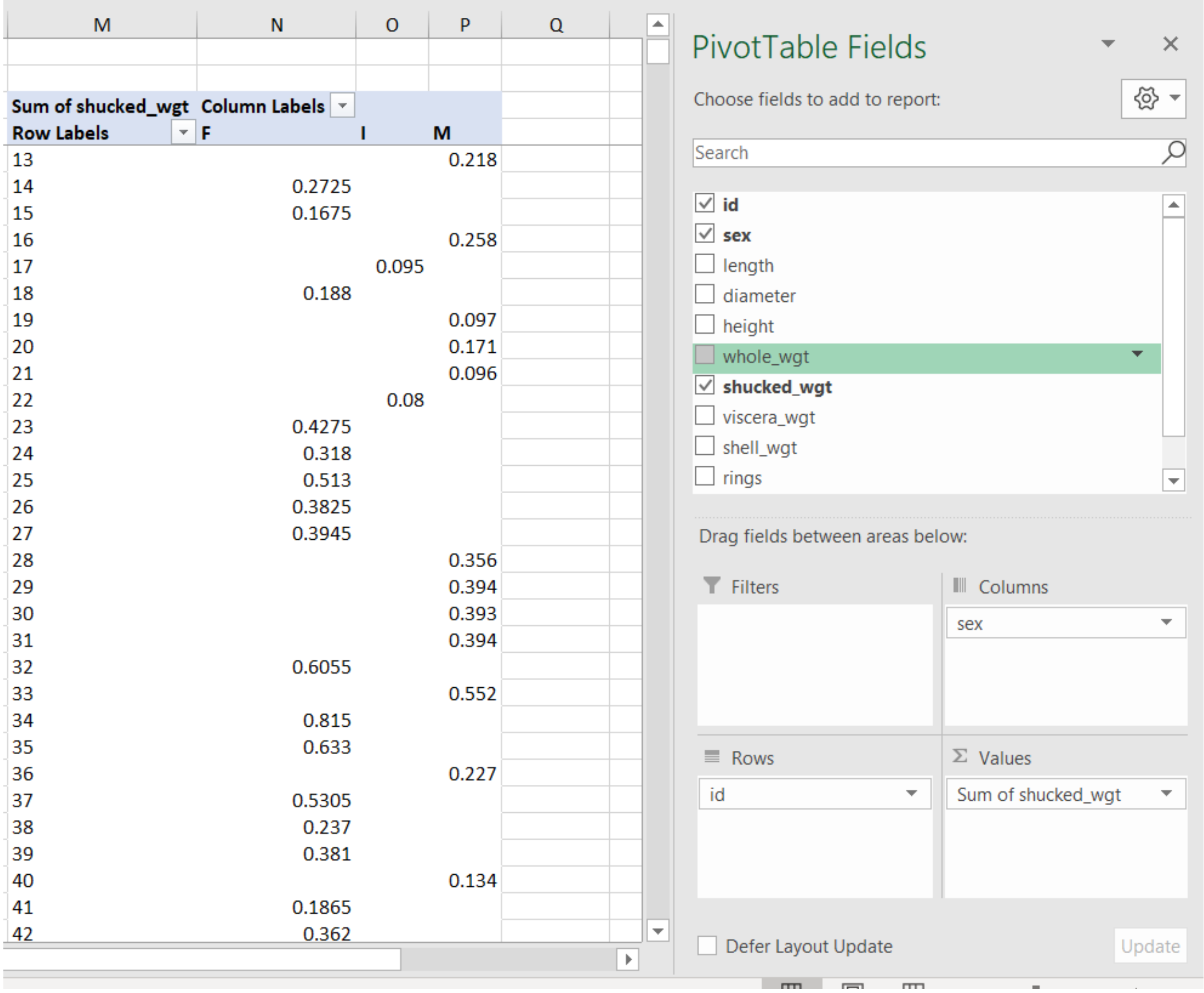


Analysis of variance (ANOVA)

File: `abalone-anova.xlsx`

Let’s check for a significant difference in shucked weights across male, female and infant snails.

1. Insert a PivotTable. Put `id` in the Rows section, `sex` in the Columns section and `Sum of shucked_wgt` in the Values section.
   a. Turn off the totals by clicking inside the PivotTable and selecting Design > Grand Totals > Off for Rows and Columns.



Sum of shucked_wgt	Column Labels		
Row Labels	F	I	M
13			0.218
14	0.2725		
15	0.1675		
16			0.258
17		0.095	
18	0.188		
19			0.097
20			0.171
21			0.096
22		0.08	
23	0.4275		
24	0.318		
25	0.513		
26	0.3825		
27	0.3945		
28			0.356
29			0.394
30			0.393
31			0.394
32	0.6055		
33			0.552
34	0.815		
35	0.633		
36			0.227
37	0.5305		
38	0.237		
39	0.381		
40			0.134
41	0.1865		
42	0.362		

PivotTable Fields

Choose fields to add to report:

☑ id
☑ sex
□ length
□ diameter
□ height
□ whole_wgt
☑ shucked_wgt
□ viscera_wgt
□ shell_wgt
□ rings

Drag fields between areas below:

Filters

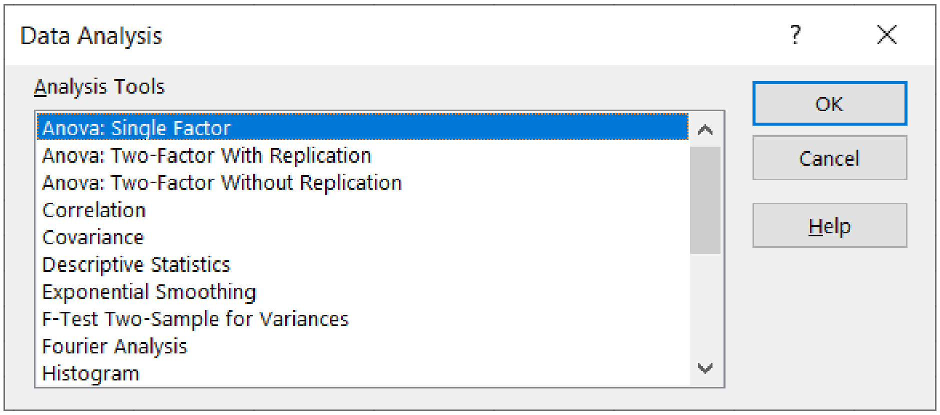
Columns: sex

Rows: id

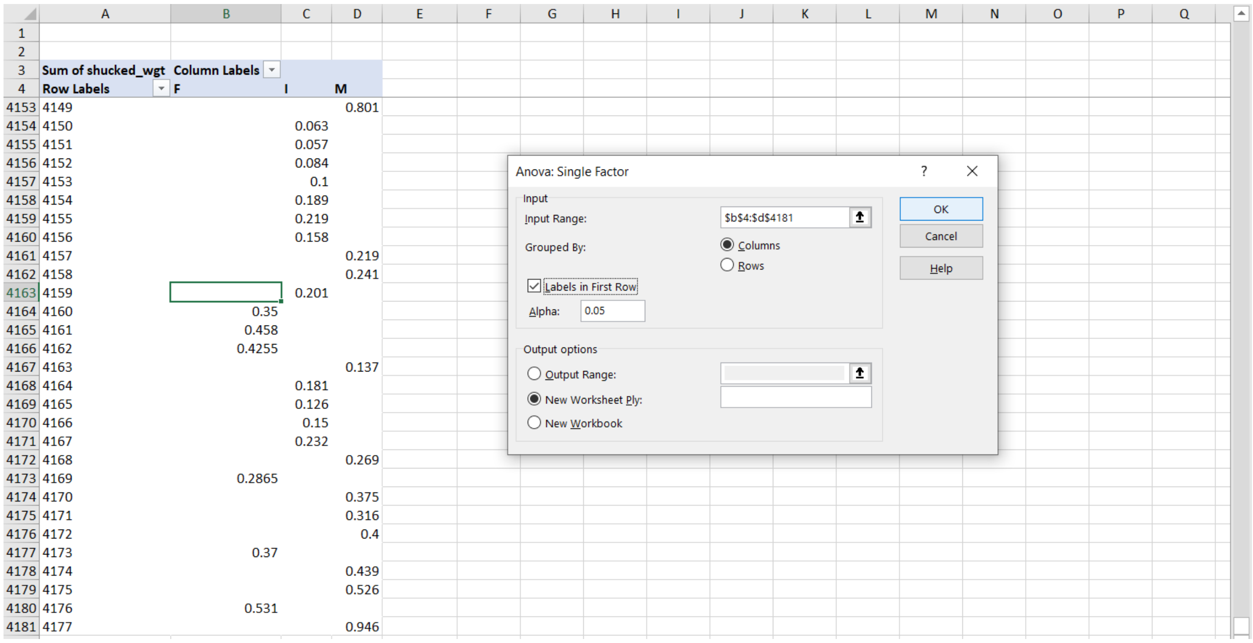
Values: Sum of shucked_wgt

□ Defer Layout Update

2. In the Analysis ToolPak, select Anova: Single Factor.



a. The input range is the three columns for each category: F, I and M.



	A	B	C	D
1				
2				
3	Sum of shucked_wgt	Column Labels		
4	Row Labels	F	I	M
4153	4149			0.801
4154	4150		0.063	
4155	4151		0.057	
4156	4152		0.084	
4157	4153		0.1	
4158	4154		0.189	
4159	4155		0.219	
4160	4156		0.158	
4161	4157			0.219
4162	4158			0.241
4163	4159		0.201	
4164	4160	0.35		
4165	4161	0.458		
4166	4162	0.4255		
4167	4163			0.137
4168	4164		0.181	
4169	4165		0.126	
4170	4166		0.15	
4171	4167		0.232	
4172	4168			0.269
4173	4169	0.2865		
4174	4170			0.375
4175	4171			0.316
4176	4172			0.4
4177	4173	0.37		
4178	4174			0.439
4179	4175			0.526
4180	4176	0.531		
4181	4177			0.946

Anova: Single Factor

Input

Input Range: b4:d4181

Grouped By: ● Columns ○ Rows

☑ Labels in First Row

Alpha: 0.05

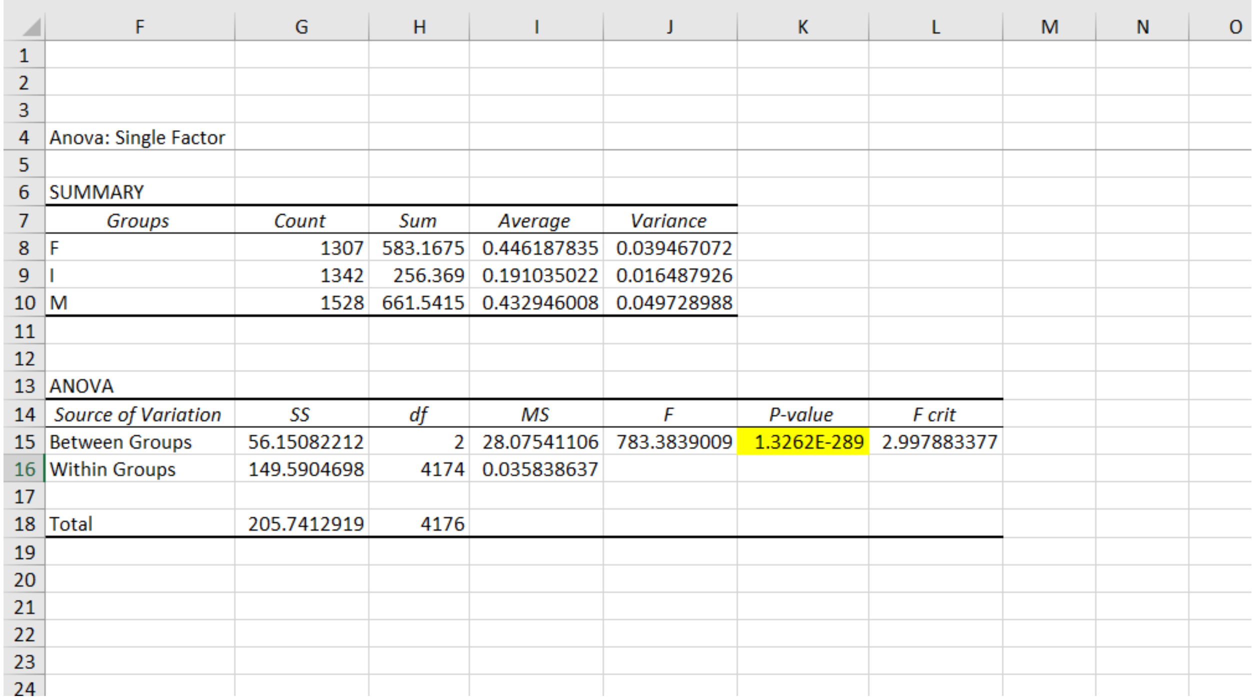
Output options

○ Output Range:

● New Worksheet Ply:

○ New Workbook

3. The results of the ANOVA are available in the second box of outputs. The p-value for between-groups variation tells us if there is a significant difference across group means.



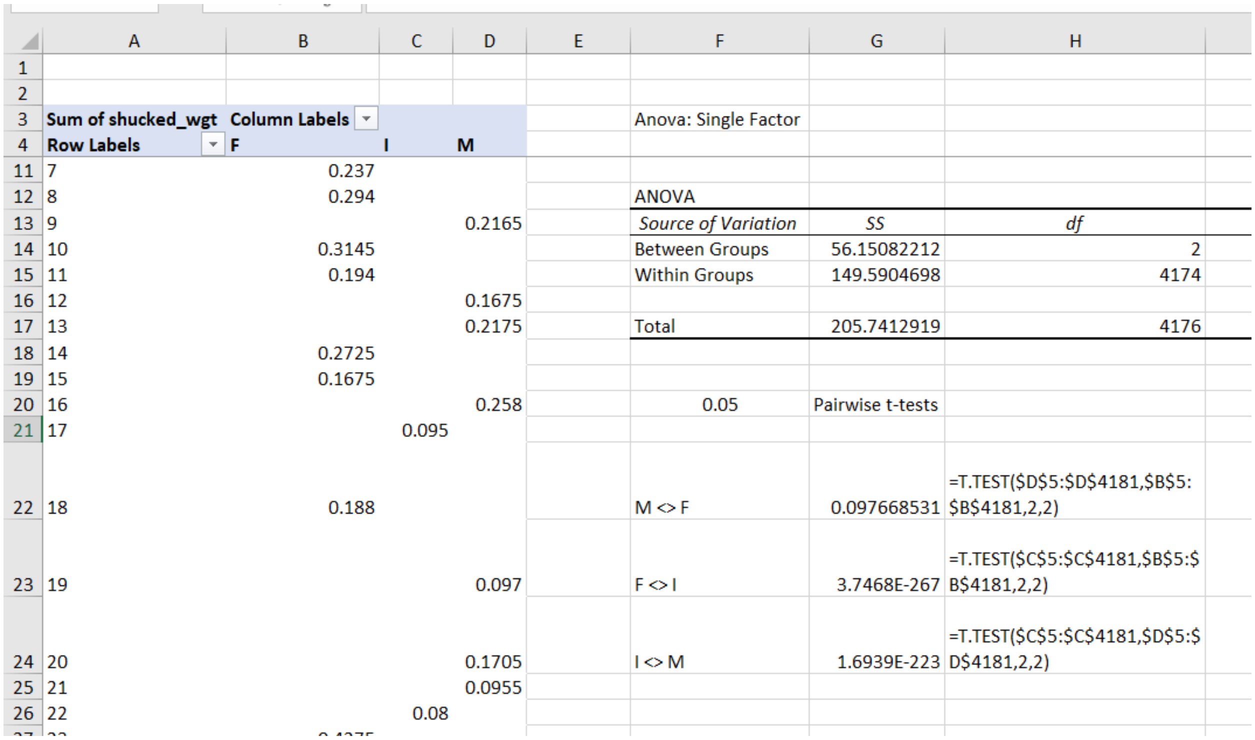
	F	G	H	I	J	K	L	M	N	O
1										
2										
3										
4	Anova: Single Factor									
5										
6	SUMMARY									
7	*Groups*	*Count*	*Sum*	*Average*	*Variance*					
8	F	1307	583.1675	0.446187835	0.039467072					
9	I	1342	256.369	0.191035022	0.016487926					
10	M	1528	661.5415	0.432946008	0.049728988					
11										
12										
13	ANOVA									
14	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*			
15	Between Groups	56.15082212	2	28.07541106	783.3839009	1.3262E-289	2.997883377			
16	Within Groups	149.5904698	4174	0.035838637						
17										
18	Total	205.7412919	4176							
19										
20										
21										
22										
23										
24										

ANOVA post-hoc tests: pairwise comparisons with Bonferroni correction

File: `abalone-post-hoc.xlsx`

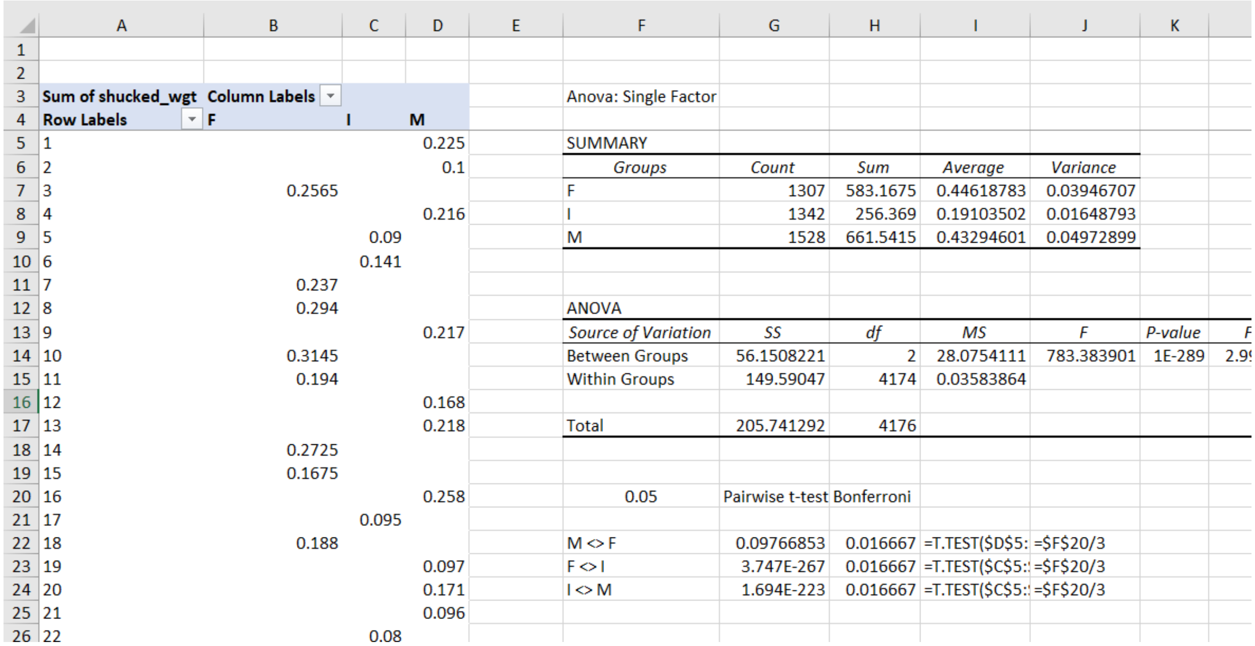
The ANOVA in itself does *not* tell us *which* groups in particular are significantly higher/lower than the others. To do that, we will run *post-hoc* tests while adjusting for *experimentwise error*.

1. Conduct a *pairwise t-test* for to compare each pair of categories using the T.TEST() function.
    a. This will take four arguments:
        i. The range containing the first category to compare
        ii. The range containing the second category to compare
        iii. Whether this is a one- or two-tail test. We are using two-tail tests, so the argument is 2.
        iv. The type of t-test. Since these are independent samples, this is not a paired t-test. We will assume equal variances as that is an assumption of the ANOVA. So the argument here is also 2.
        v. The result of `T.TEST()` is the test's p-value. We will compare it against the adjusted alpha next.



	A	B	C	D	E	F	G	H
1								
2								
3	Sum of shucked_wgt	Column Labels ▾				Anova: Single Factor		
4	Row Labels ▾	F	I	M				
11	7	0.237						
12	8	0.294				ANOVA		
13	9			0.2165		*Source of Variation*	*SS*	*df*
14	10	0.3145				Between Groups	56.15082212	2
15	11	0.194				Within Groups	149.5904698	4174
16	12			0.1675				
17	13			0.2175		Total	205.7412919	4176
18	14	0.2725						
19	15	0.1675						
20	16			0.258		0.05	Pairwise t-tests	
21	17		0.095					
22	18	0.188				M <> F	0.097668531	=T.TEST($D$5:$D$4181,$B$5:$B$4181,2,2)
23	19			0.097		F <> I	3.7468E-267	=T.TEST($C$5:$C$4181,$B$5:$B$4181,2,2)
24	20			0.1705		I <> M	1.6939E-223	=T.TEST($C$5:$C$4181,$D$5:$D$4181,2,2)
25	21			0.0955				
26	22		0.08					

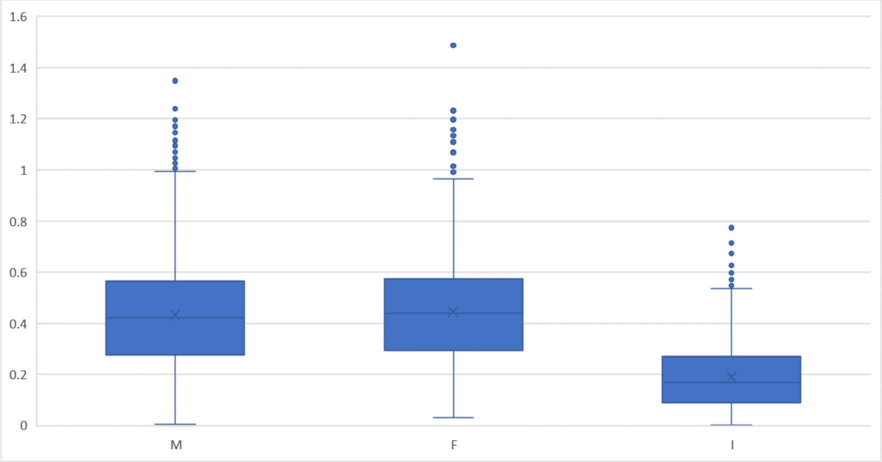
2. We will now compare these p-values to a Bonferroni-adjusted alpha. This number will be our original alpha (.05) divided by the number of groups we are comparing (3). This makes our new alpha .0017.



	A	B	C	D	E	F	G	H	I	J	K	
1												
2												
3	Sum of shucked_wgt	Column Labels ▾				Anova: Single Factor						
4	Row Labels ▾	F	I	M								
5	1			0.225		SUMMARY						
6	2			0.1		*Groups*	*Count*	*Sum*	*Average*	*Variance*		
7	3	0.2565				F	1307	583.1675	0.44618783	0.03946707		
8	4			0.216		I	1342	256.369	0.19103502	0.01648793		
9	5		0.09			M	1528	661.5415	0.43294601	0.04972899		
10	6		0.141									
11	7	0.237										
12	8	0.294				ANOVA						
13	9			0.217		*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F*
14	10	0.3145				Between Groups	56.1508221	2	28.0754111	783.383901	1E-289	2.9
15	11	0.194				Within Groups	149.59047	4174	0.03583864			
16	12			0.168								
17	13			0.218		Total	205.741292	4176				
18	14	0.2725										
19	15	0.1675										
20	16			0.258		0.05	Pairwise t-test	Bonferroni				
21	17		0.095									
22	18	0.188				M <> F	0.09766853	0.016667	=T.TEST($D$5:	=$F$20/3		
23	19			0.097		F <> I	3.747E-267	0.016667	=T.TEST($C$5:	=$F$20/3		
24	20			0.171		I <> M	1.694E-223	0.016667	=T.TEST($C$5:	=$F$20/3		
25	21			0.096								
26	22		0.08									

a. Based on these results, there is no significant difference in weights between male and female snails, but there is a significant difference between female and infant snails, and a significant difference between male and infant snails.

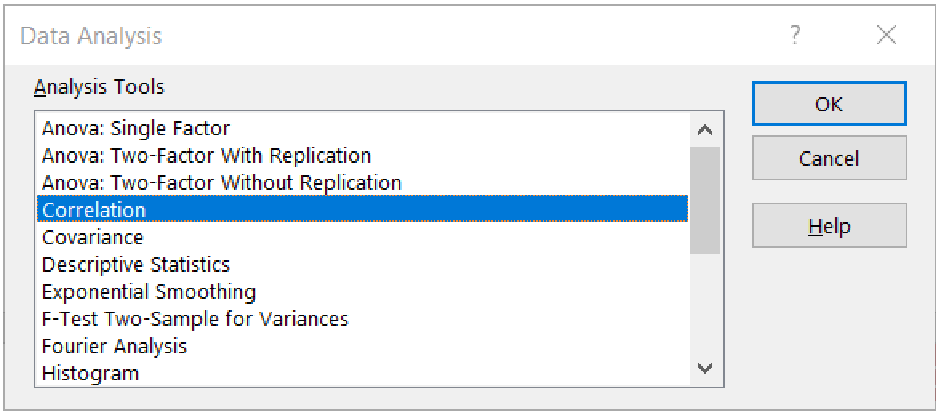
i. This is a good time to refer back to the box plots for a visual understanding of the analysis.



Pearson correlations

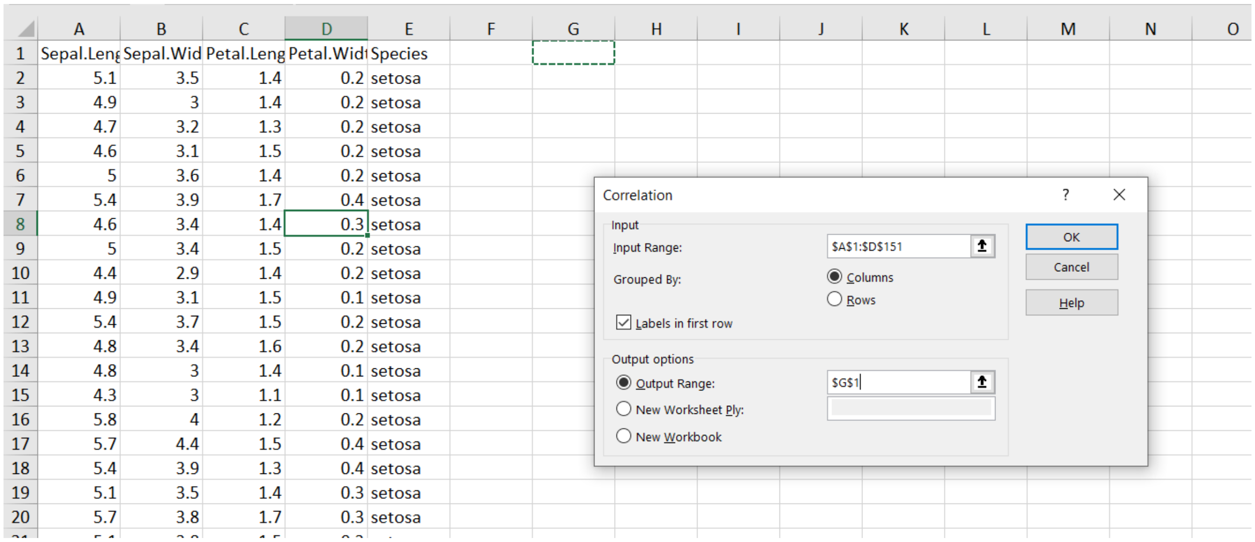
File: `iris-corr.xlsx`

1. To insert a correlation matrix, go to the Analysis ToolPak and select Correlation.



Anova: Single Factor
Anova: Two-Factor With Replication
Anova: Two-Factor Without Replication
Correlation
Covariance
Descriptive Statistics
Exponential Smoothing
F-Test Two-Sample for Variances
Fourier Analysis
Histogram

a. The input range will be all the *numeric* fields. String fields cannot be included in the correlation analysis.



	A	B	C	D	E
1	Sepal.Leng	Sepal.Wid	Petal.Leng	Petal.Widt	Species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	1.4	0.2	setosa
4	4.7	3.2	1.3	0.2	setosa
5	4.6	3.1	1.5	0.2	setosa
6	5	3.6	1.4	0.2	setosa
7	5.4	3.9	1.7	0.4	setosa
8	4.6	3.4	1.4	0.3	setosa
9	5	3.4	1.5	0.2	setosa
10	4.4	2.9	1.4	0.2	setosa
11	4.9	3.1	1.5	0.1	setosa
12	5.4	3.7	1.5	0.2	setosa
13	4.8	3.4	1.6	0.2	setosa
14	4.8	3	1.4	0.1	setosa
15	4.3	3	1.1	0.1	setosa
16	5.8	4	1.2	0.2	setosa
17	5.7	4.4	1.5	0.4	setosa
18	5.4	3.9	1.3	0.4	setosa
19	5.1	3.5	1.4	0.3	setosa
20	5.7	3.8	1.7	0.3	setosa

Correlation

Input

Input Range: $A$1:$D$151

Grouped By: ● Columns ○ Rows

☑ Labels in first row

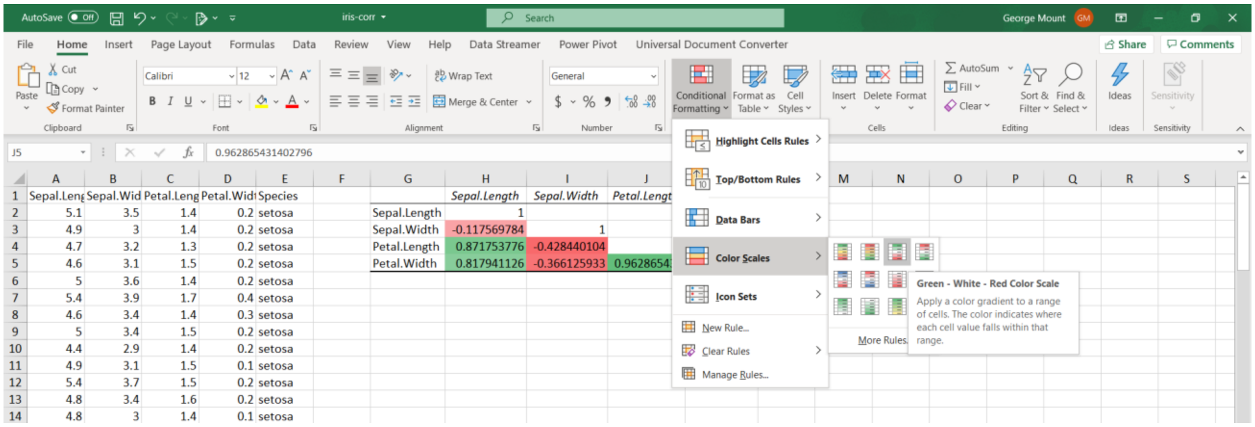
Output options

● Output Range: $G$1

○ New Worksheet Ply:

○ New Workbook

b. For ease of interpretation, select all the correlation values and select Home > Conditional Formatting > Color Scales > Green – White – Red Color Scales.



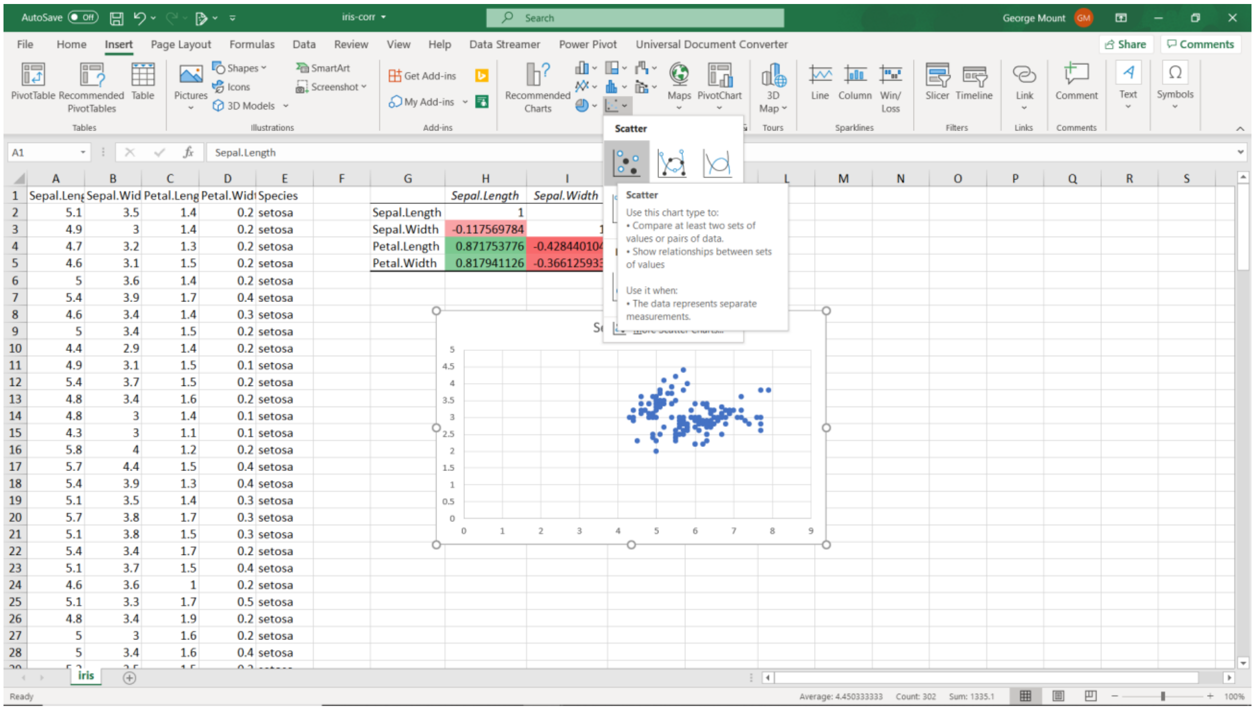
J5 0.962865431402796

	G	H	I	J
1		Sepal.Length	Sepal.Width	Petal.Lengt
2	Sepal.Length	1		
3	Sepal.Width	-0.117569784	1	
4	Petal.Length	0.871753776	-0.428440104	
5	Petal.Width	0.817941126	-0.366125933	0.9628654

Green - White - Red Color Scale

Apply a color gradient to a range of cells. The color indicates where each cell value falls within that range.

2. To plot the relationship between two variables (in this case, sepal length and sepal width), highlight the data and select Insert > Scatter.



A1 | Sepal.Length

	A	B	C	D	E
1	Sepal.Leng	Sepal.Wid	Petal.Leng	Petal.Widt	Species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	1.4	0.2	setosa
4	4.7	3.2	1.3	0.2	setosa
5	4.6	3.1	1.5	0.2	setosa
6	5	3.6	1.4	0.2	setosa
7	5.4	3.9	1.7	0.4	setosa
8	4.6	3.4	1.4	0.3	setosa
9	5	3.4	1.5	0.2	setosa
10	4.4	2.9	1.4	0.2	setosa
11	4.9	3.1	1.5	0.1	setosa
12	5.4	3.7	1.5	0.2	setosa
13	4.8	3.4	1.6	0.2	setosa
14	4.8	3	1.4	0.1	setosa
15	4.3	3	1.1	0.1	setosa
16	5.8	4	1.2	0.2	setosa
17	5.7	4.4	1.5	0.4	setosa
18	5.4	3.9	1.3	0.4	setosa
19	5.1	3.5	1.4	0.3	setosa
20	5.7	3.8	1.7	0.3	setosa
21	5.1	3.8	1.5	0.3	setosa
22	5.4	3.4	1.7	0.2	setosa
23	5.1	3.7	1.5	0.4	setosa
24	4.6	3.6	1	0.2	setosa
25	5.1	3.3	1.7	0.5	setosa
26	4.8	3.4	1.9	0.2	setosa
27	5	3	1.6	0.2	setosa
28	5	3.4	1.6	0.4	setosa

G	H	I
	Sepal.Length	*Sepal.Width*
Sepal.Length	1	
Sepal.Width	-0.117569784	1
Petal.Length	0.871753776	-0.42844010
Petal.Width	0.817941126	-0.36612593

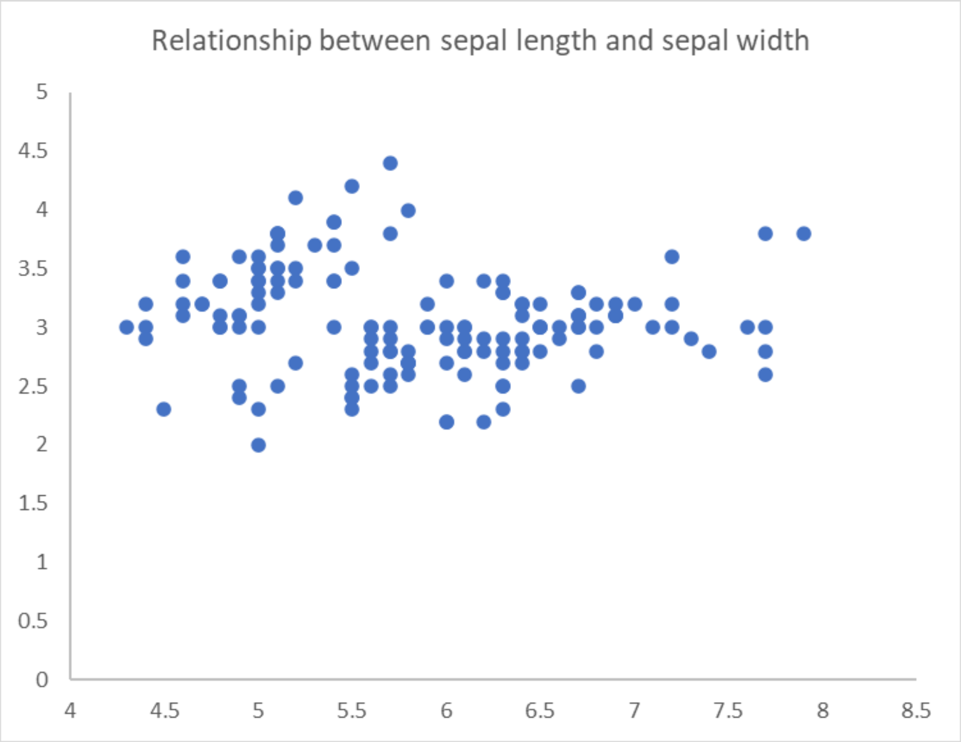
Scatter

Use this chart type to:
- Compare at least two sets of values or pairs of data.
- Show relationships between sets of values

Use it when:
- The data represents separate measurements.

Average: 4.450333333 Count: 302 Sum: 1335.1 100%

a. To clean up this scatter chart, set the X-axis to 4 and remove the gridlines. It's also a good idea to label this chart more clearly.



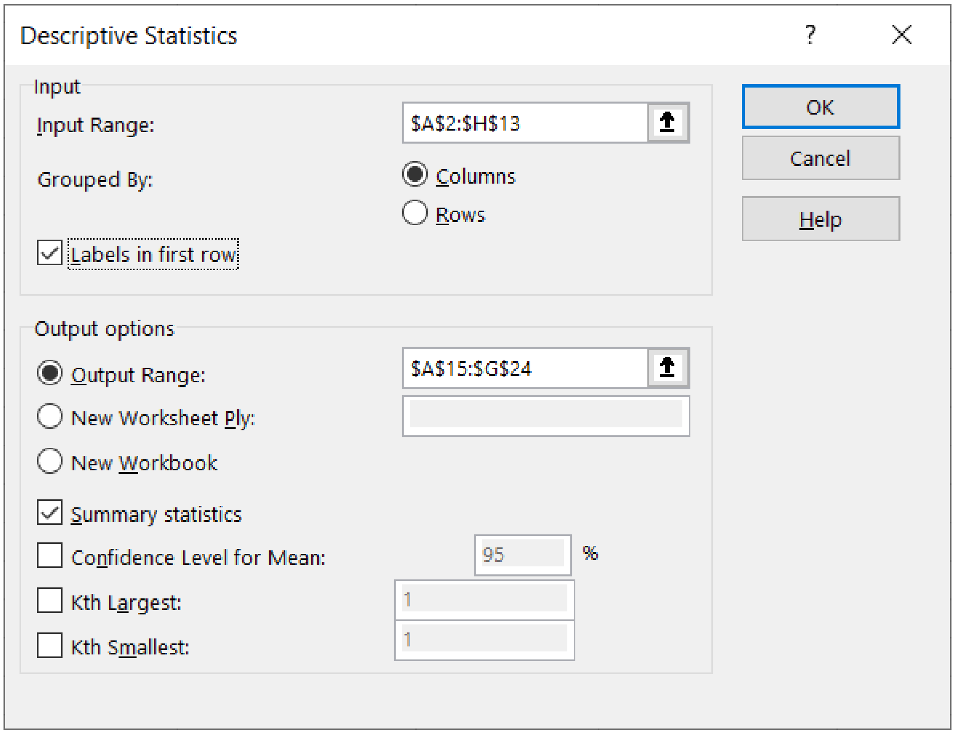
b. What does the scatterplot of sepal length and petal length look like?

c. By custom, you want to put the *independent* variable on the X-axis, and the *dependent* on the Y-axis.

Careful with correlations!

File: `anscombe.xlsx`

1. Perform descriptive statistics on all variables by selecting Descriptive Statistics from the ToolPak. Make sure you select “Summary statistics” from the Output Options.



Descriptive Statistics

Input

Input Range: $A$2:$H$13

Grouped By: ● Columns ○ Rows

☑ Labels in first row

Output options

● Output Range: $A$15:$G$24

○ New Worksheet Ply:

○ New Workbook

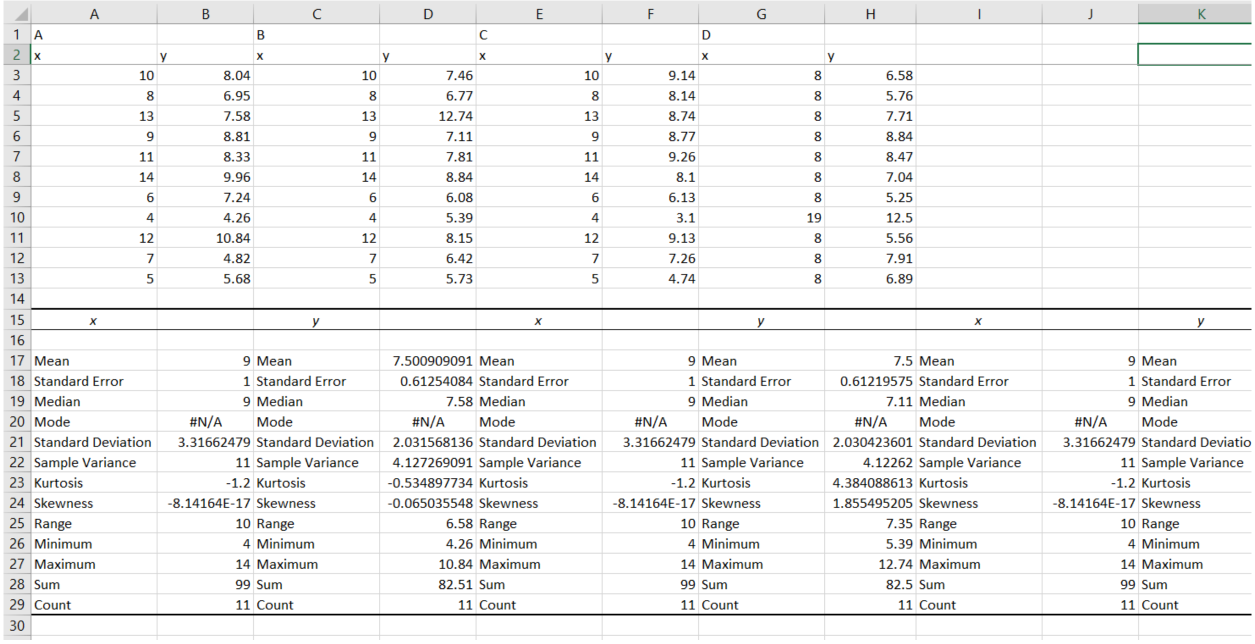
☑ Summary statistics

□ Confidence Level for Mean: 95 %

□ Kth Largest: 1

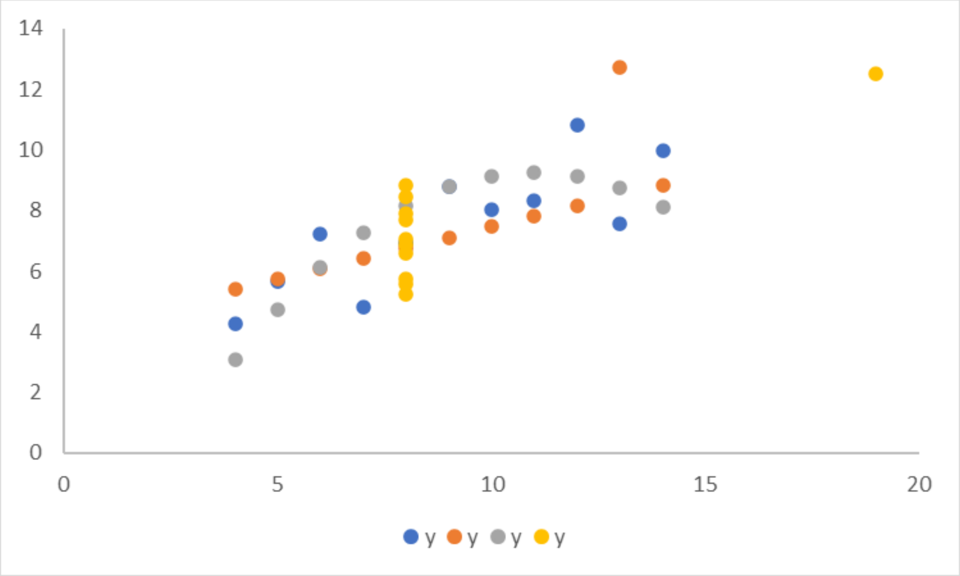
□ Kth Smallest: 1

   a. From the results of these descriptive statistics, each X-Y pair is *very* similar.



	A	B	C	D	E	F	G	H	I	J	K
1	A		B		C		D				
2	x	y	x	y	x	y	x	y			
3	10	8.04	10	7.46	10	9.14	8	6.58			
4	8	6.95	8	6.77	8	8.14	8	5.76			
5	13	7.58	13	12.74	13	8.74	8	7.71			
6	9	8.81	9	7.11	9	8.77	8	8.84			
7	11	8.33	11	7.81	11	9.26	8	8.47			
8	14	9.96	14	8.84	14	8.1	8	7.04			
9	6	7.24	6	6.08	6	6.13	8	5.25			
10	4	4.26	4	5.39	4	3.1	19	12.5			
11	12	10.84	12	8.15	12	9.13	8	5.56			
12	7	4.82	7	6.42	7	7.26	8	7.91			
13	5	5.68	5	5.73	5	4.74	8	6.89			
14											
15	*x*		*y*		*x*		*y*		*x*		*y*
16											
17	Mean	9	Mean	7.500909091	Mean	9	Mean	7.5	Mean	9	Mean
18	Standard Error	1	Standard Error	0.61254084	Standard Error	1	Standard Error	0.61219575	Standard Error	1	Standard Error
19	Median	9	Median	7.58	Median	9	Median	7.11	Median	9	Median
20	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode
21	Standard Deviation	3.31662479	Standard Deviation	2.031568136	Standard Deviation	3.31662479	Standard Deviation	2.030423601	Standard Deviation	3.31662479	Standard Deviatio
22	Sample Variance	11	Sample Variance	4.127269091	Sample Variance	11	Sample Variance	4.12262	Sample Variance	11	Sample Variance
23	Kurtosis	-1.2	Kurtosis	-0.534897734	Kurtosis	-1.2	Kurtosis	4.384088613	Kurtosis	-1.2	Kurtosis
24	Skewness	-8.14164E-17	Skewness	-0.065035548	Skewness	-8.14164E-17	Skewness	1.855495205	Skewness	-8.14164E-17	Skewness
25	Range	10	Range	6.58	Range	10	Range	7.35	Range	10	Range
26	Minimum	4	Minimum	4.26	Minimum	4	Minimum	5.39	Minimum	4	Minimum
27	Maximum	14	Maximum	10.84	Maximum	14	Maximum	12.74	Maximum	14	Maximum
28	Sum	99	Sum	82.51	Sum	99	Sum	82.5	Sum	99	Sum
29	Count	11	Count	11	Count	11	Count	11	Count	11	Count
30											

2. What happens if we graph this data using a scatter plot? Highlight the range and select Insert > Recommended Charts > Scatter.
   a. Only from the visualization can we easily see that these datasets are quite different!


14
12
10
8
6
4
2
0
0
5
10
15
20
● y ● y ● y ● y