Advancing into Analytics Foundations: Demo notes

Frequencies

Create a PivotTable from the source data.

1. Make a one-way frequency table by selecting categories of interest in the
    Rows/Columns field, then place a Count of the ID field in the Values section.
    a. To convert a field from a Sum to a Count, double-click on that variable header,
        and select Count in the "Summarize value field by" menu.

2. Make a two-way frequency table by dragging another categorical variable along the Columns area:

	A	B	C	D	E	F	G	H	I
1									
2									
3	**Count of id**	**Column Labels** ▾							
4	**Row Labels** ▾	**no**	**yes**	**Grand Total**					
5	no		248	125	373				
6	yes		107	66	173				
7	**Grand Total**		**355**	**191**	**546**				
8									

PivotChart Fields

Choose fields to add to report:

Search

☐ stories
☐ driveway
☐ recroom
☑ **fullbase**
☐ gashw
☑ **airco**

Drag fields between areas below:

▼ Filters

▥ Legend (Series)
fullbase ▾

☰ Axis (Categori...
airco ▾

Σ Values
Count of id ▾

   a. Select Insert > Recommended Charts > Clustered Column to visualize the frequencies:
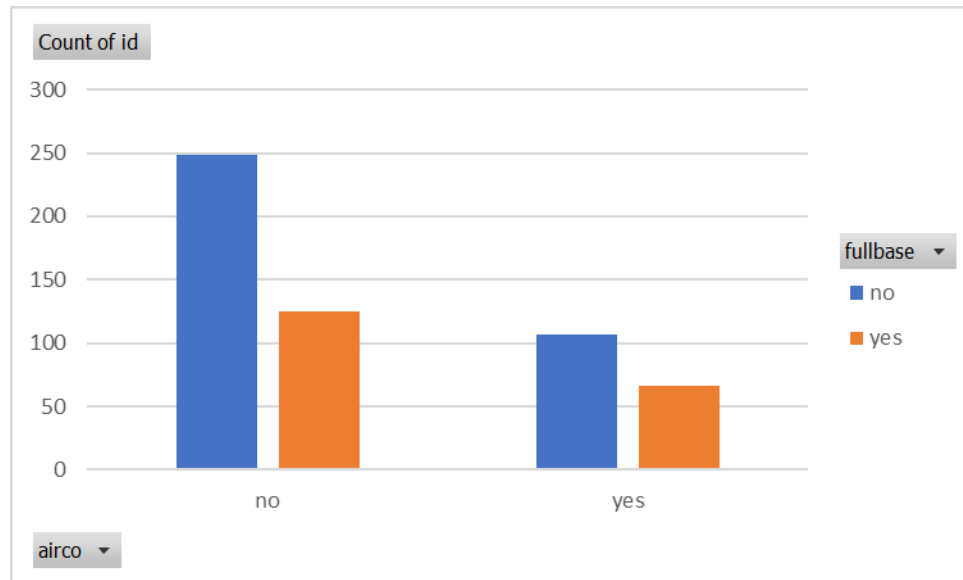
Downloading the Analysis ToolPak

See instructions from Microsoft here. Note the process is different for Windows and Mac.

Descriptive Statistics

1. Go to the Data tab on the home ribbon.

2. Select Data Analysis from the Analyze group (far right of the menu).

3. Select Descriptive Statistics from the menu.
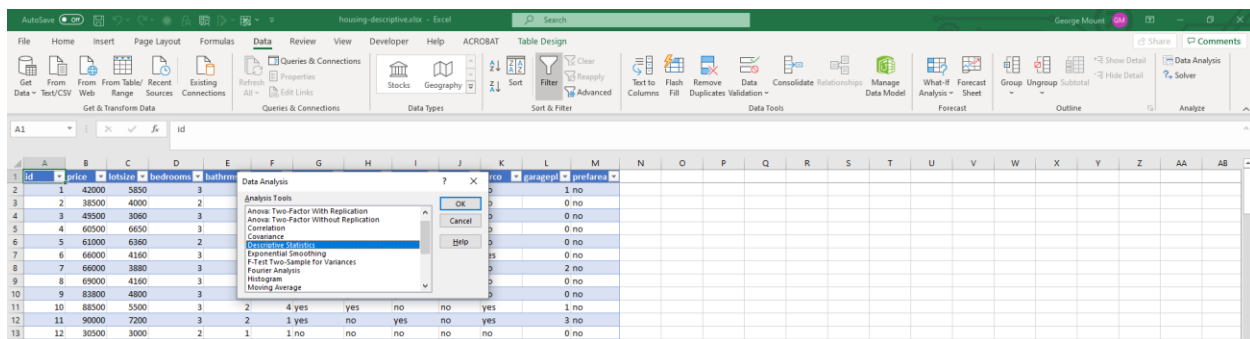
4. Select your Input Range. This will be Column B, `price`. If your selection includes a header row, make sure to check on the "Labels in First Row" option.

5. By default, the output will be placed in a new worksheet. If you want it elsewhere, click inside "Output Range." Make sure to double-click inside the dialog box before selecting a new range, otherwise the input range will be re-written.

6. Check on "Summary Statistics."

Descriptive Statistics dialog box:

- Input
  - Input Range: $B$1:$B$547
  - Grouped By: ● Columns ○ Rows
  - ☑ Labels in First Row
- Output options
  - ○ Output Range:
  - ● New Worksheet Ply:
  - ○ New Workbook
  - ☑ Summary statistics
  - ☐ Confidence Level for Mean: 95 %
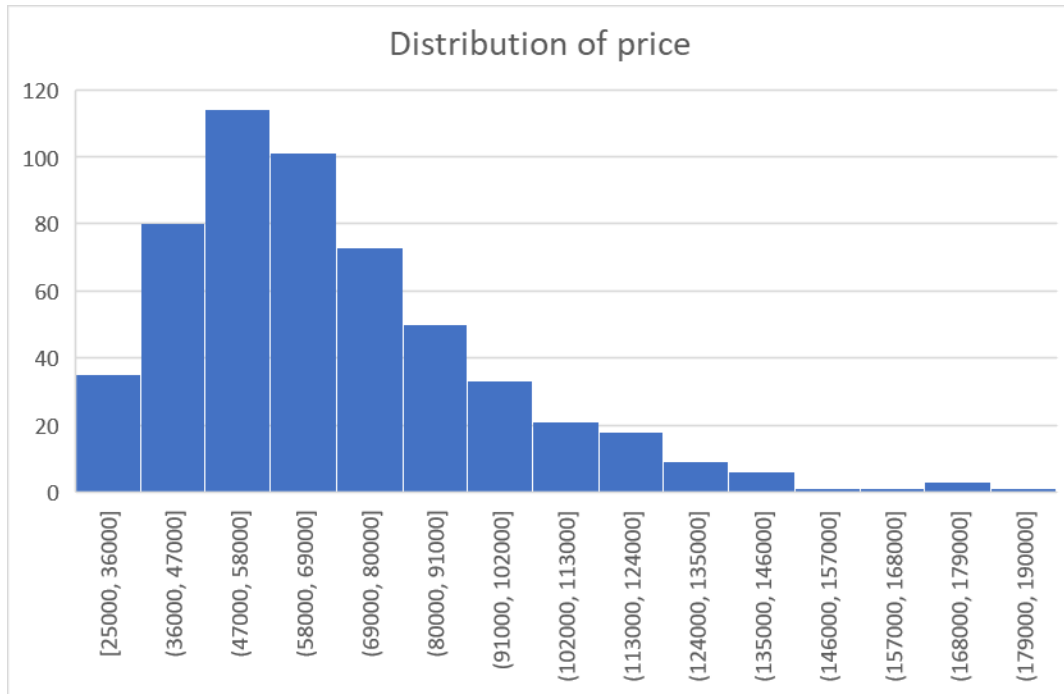 - ☐ Kth Largest: 1
  - ☐ Kth Smallest: 1
- OK
- Cancel
- Help

7. We find that the average sale price is about $68,122 with a standard deviation of $26,702.

	A	B
1	*price*	
2		
3	Mean	68121.59707
4	Standard Error	1142.76882
5	Median	62000
6	Mode	60000
7	Standard Deviation	26702.67093
8	Sample Variance	713032634.6
9	Kurtosis	1.959739991
10	Skewness	1.209829216
11	Range	165000
12	Minimum	25000
13	Maximum	190000
14	Sum	37194392
15	Count	546
16		

Histograms

1. Select your input range (the price column) and go to Insert -> Charts. Histogram should be your third option. Select that. You can cut and paste the resulting histogram elsewhere in the workbook.

Distribution of price

A histogram titled "Distribution of price" with bins on the X-axis from [25000, 36000] through (179000, 190000] and frequency on the Y-axis from 0 to 120.

2. To change the number of bins in the histogram, right-click on the X-axis and select Format Axis. You can then customize the X-axis on the side menu. *Note: these features are not available on Excel for Mac.*

Descriptive statistics for two categories

1. Create a PivotTable containing ID in the rows, airco in the Columns and Sum of price in the Values.

a. Design > Grand Totals > Off for Rows and Columns

	A	B	C
1			
2			
3	Sum of price	Column Labels	
4	Row Labels	no	yes
5	1		42000
6	2		38500
7	3		49500
8	4		60500
9	5		61000
10	6		66000
11	7		66000
12	8		69000
13	9		83800
14	10		88500
15	11		90000
16	12		30500
17	13		27000
18	14		36000
19	15		37000
20	16		37900
21	17		40500
22	18		40750
23	19		45000
24	20		45000
25	21		48500
26	22		65900
27	23		37900
28	24		38000
29	25		42000
30	26		42300

PivotTable F.. – ✕

Choose fields to add to report:

Search

☑ **id**
☑ **price**
☐ lotsize
☐ bedrooms
☐ bathrms

Drag fields between areas below:

▼ Filters

▥ Columns
airco ▼

≡ Rows
id ▼

Σ Values
Sum of p... ▼

☐ Defer Layout Upda...   Update

2. Run the descriptives for each category using the Analysis ToolPak.

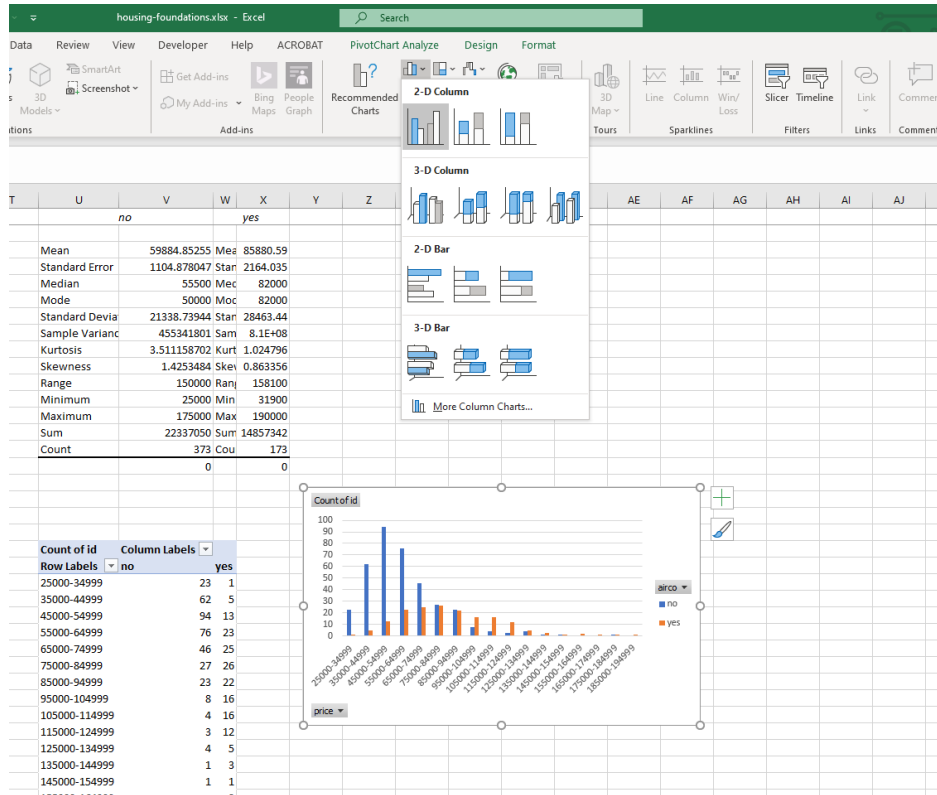## Plot a histogram for two categories

1. Create another PivotTable (copy-paste the old one)

2. Place the two categories along the Columns, the continuous variable of interest down the Rows, and the Count of the ID variable in the Values.

3. Right-click the Rows area and select Group.

4. Go to Insert on the home ribbon and select a 2-D Column chart.

5. Right-click on any of the bars in the resulting bar chart. Select Format Data Series.
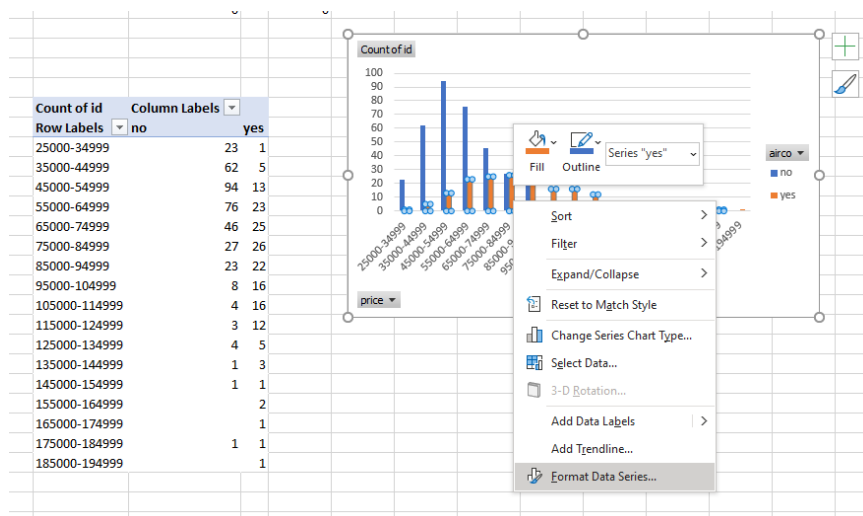
6. Adjust the Series Overlap to 100% and the Gap Width to 0%.