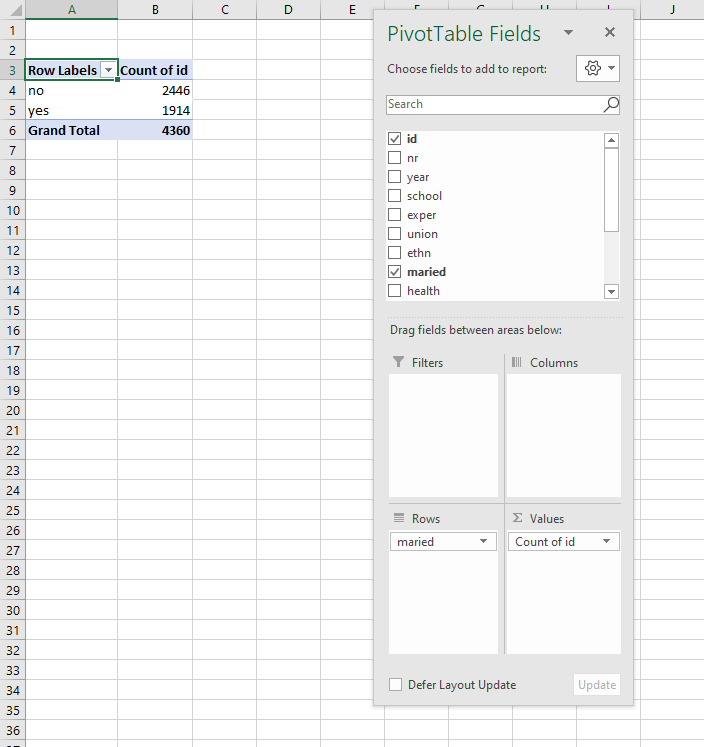
**EXPLORATORY DATA ANALYSIS IN EXCEL– DEMO NOTES**

**Frequencies**

Suppose you have 3 groups and one measurement. Then there will be some variation within each group and some between the groups. ANOVA compares the variation *between* to the variation *within*. If the former is large compared to the latter, then there is a lot of evidence of a group difference. If the former is small relative to the latter, then there is much less evidence.

Create a PivotTable from the source data.

1. Make a frequency table by selecting categories of interest in the Rows/Columns field, then place a Count of the ID field in the Values section.
   1. 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.

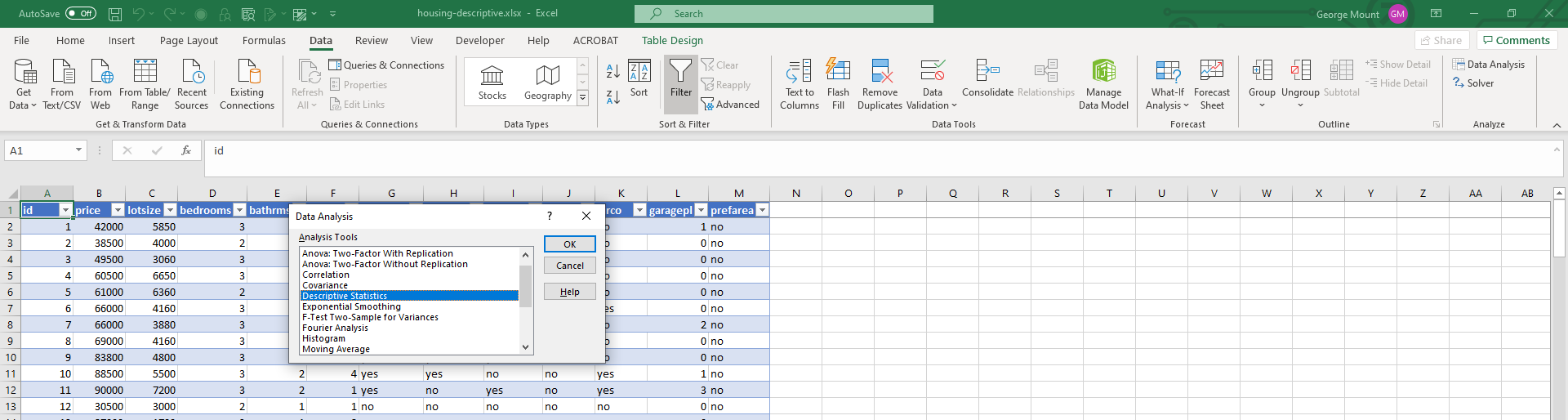


**Downloading the Analysis ToolPak**

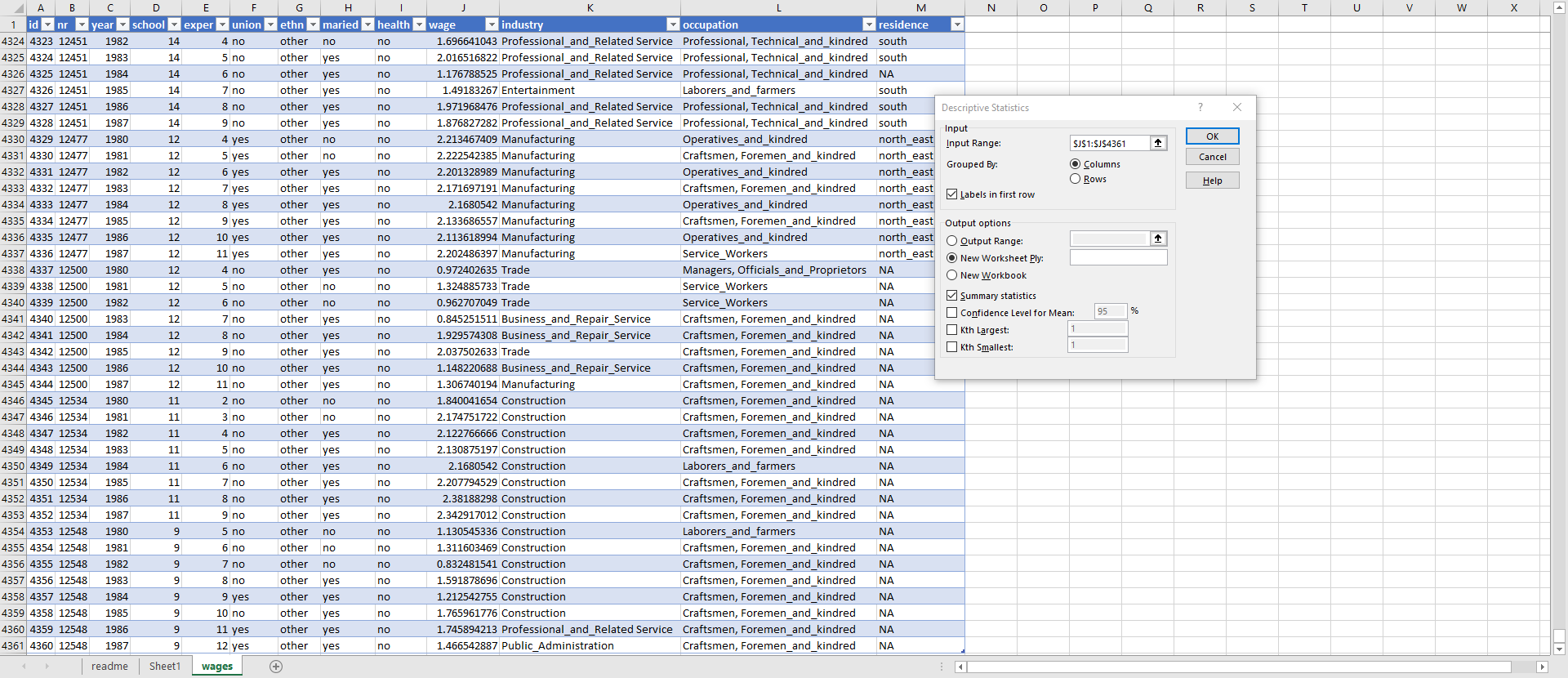
[See instructions from Microsoft here](https://support.office.com/en-us/article/Load-the-Analysis-ToolPak-in-Excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4). Note the process is different for Windows and Mac.

**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.

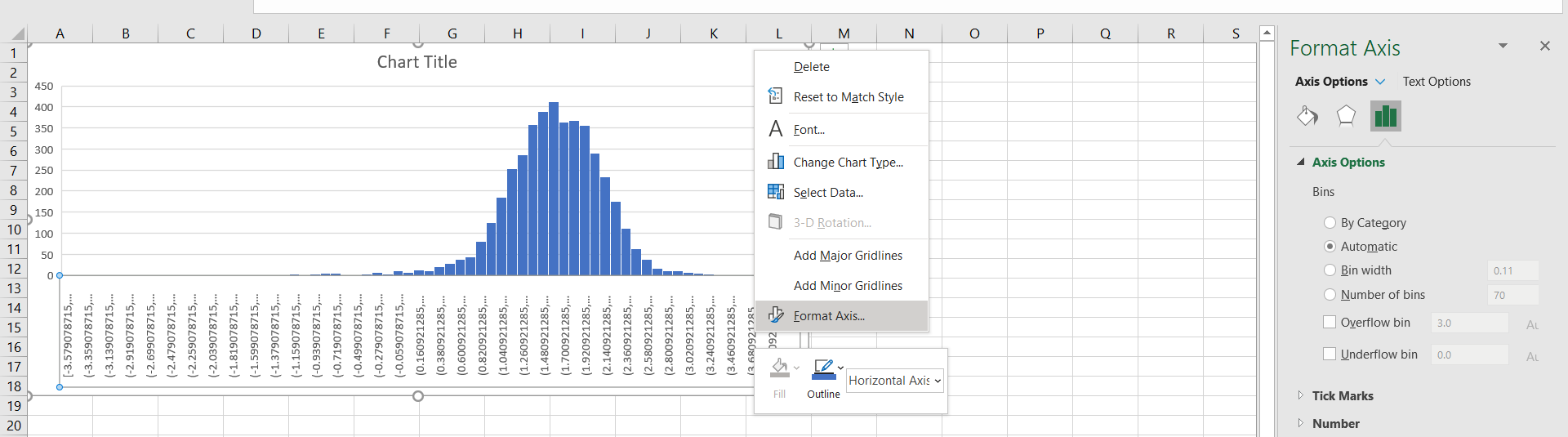


1. Select your Input Range. This will be Column J, year. If your selection includes a header row, make sure to check on the “Labels in First Row” option.
2. By default, the output will be placed in a new worksheet. If you want it elsewhere, click inside “Output Range.” Make sure to double-click inside the dialog box before selecting a new range, otherwise the input range will be re-written.
3. Check on “Summary Statistics.”



**Histograms**

1. Select your input range 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.
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.*



**Central limit theorem**

1. Simulate 500 rounds of a roulette spin using RANDBETWEEN(0,36)
2. Plot the resulting frequency distribution as a histogram.
3. Use F9 while in your workbook to refresh it. A screenshot of a cell phone

   Description automatically generated
4. This is a *uniform* distribution.
5. Now simulate a roulette spin 100 times and take the average spin. Do this 500 times and plot the resulting distribution of *sample* means.

A screenshot of a cell phone

Description automatically generated

1. This time we get a normal distribution, due to the central limit theorem.

**Law of large numbers: large-numbers.xlsx**

1. Simulate a roulette toss 500 times in Column B: RANDBETWEEN(0,36)
2. Take a running total in Column C: SUM($B$2:B2)
3. Take a running total in Column D: C2/A2
4. Plot Column D as a line chart. Press F9 to recalculate.
   1. The line converges to the expected mean due to the law of large numbers.