**Spreadsheets - Visualize Data (Second Week)**

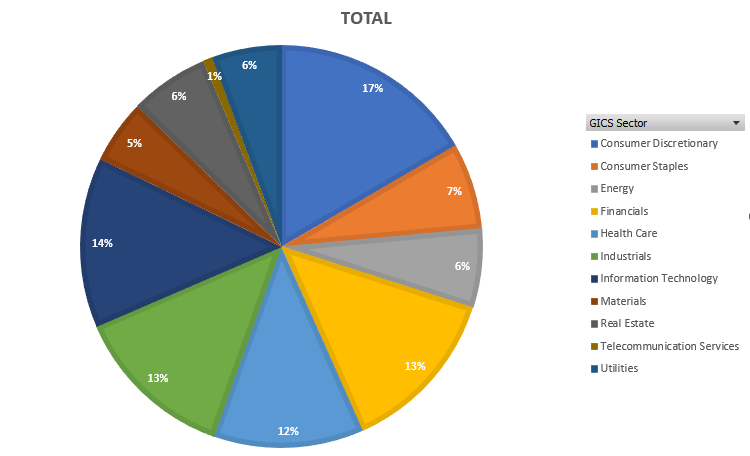
## Pie Charts

### Illustrating proportionality

*A pie chart is used to illustrate proportionality. Think of it as slicing the pie into pieces, where each piece matches a percentage of the whole list.*

In spreadsheets, this is easy, because all we need is a list of the categories and matching values such as sums or counts.

When the chart is selected, a design and format menu is available on the Excel ribbon at the top of the page. The design menu gives numerous chart options and choices, such as specific coloring or displaying percentages. You can also change the chart title.



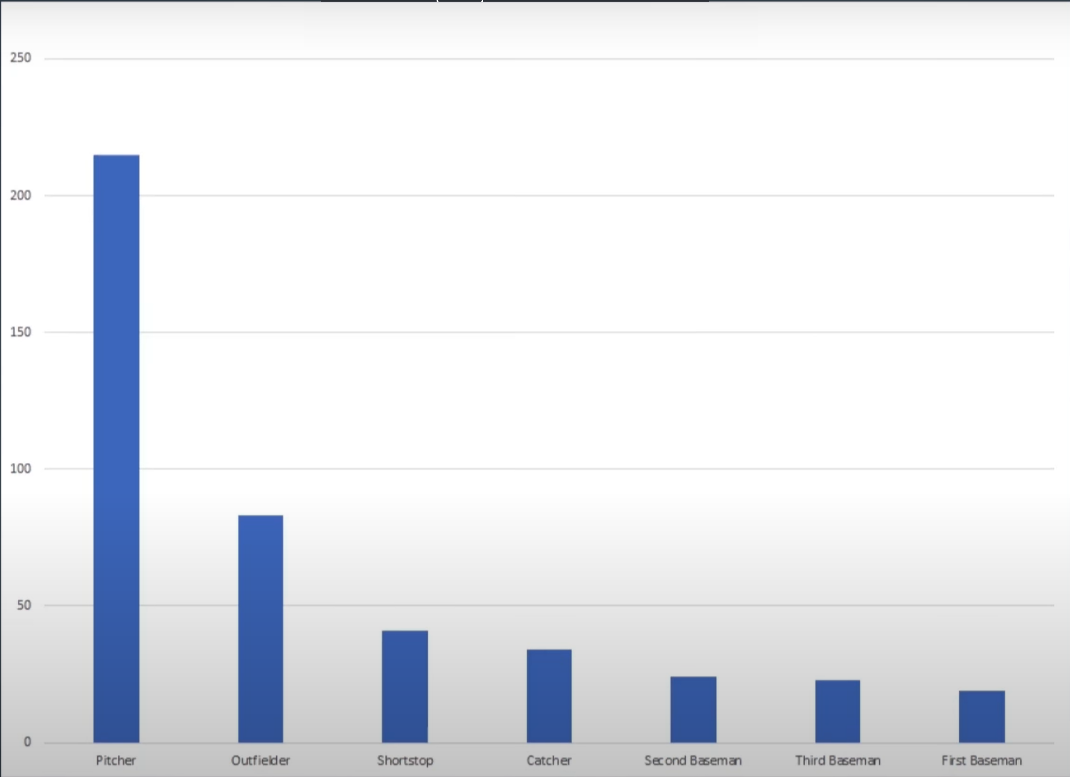
### **Pivot table**

Using the pivot table we created earlier with some careful selection, we want to highlight the position categories in the top row and the totals are in the bottom row. To do this:

1. **Select**the categories at the top of the table.
2. While holding down the control key on Windows or the Command key on Apple keyboards, **select**the bottom row with your mouse.
3. We are going to **copy** this highlighted data to another location.
4. **Paste (paste-transpose)** using the transpose feature so it creates columns instead of rows.
5. **Select** and **choose** "insert pie chart" as before.

## Bar Charts

We could use the same information as before and choose a bar or column chart instead of a pie chart. Instead of percentages, it would just show the values with longer bars or columns representing larger values.



Bar charts do not show percentages for each category

In the chart above, we're comparing the category values against each other, and we see their relative sizes. However, we do not have much sense of the whole league or the percentage of each category as we did with the pie charts.

*Choosing which kind of chart to use really depends on what patterns you want to highlight and what questions you want to answer.*

### Bar or pie chart?

* Use bar or column charts to compare category values with each other.
* Use a pie chart to show the proportionality of categories.

## Scatter and Line Plots

### Line Charts vs. Pie Charts

We use pie and bar charts to visualize categorical data. If we have a list of numerical data, such as the list of stock prices over time, a line chart gives us a better picture of the data set.

#### Simple line charts

Using a table of data downloaded from a financial website, listing prices for AAPL stock, we can explore line charts:

Chart, line chart

Description automatically generated

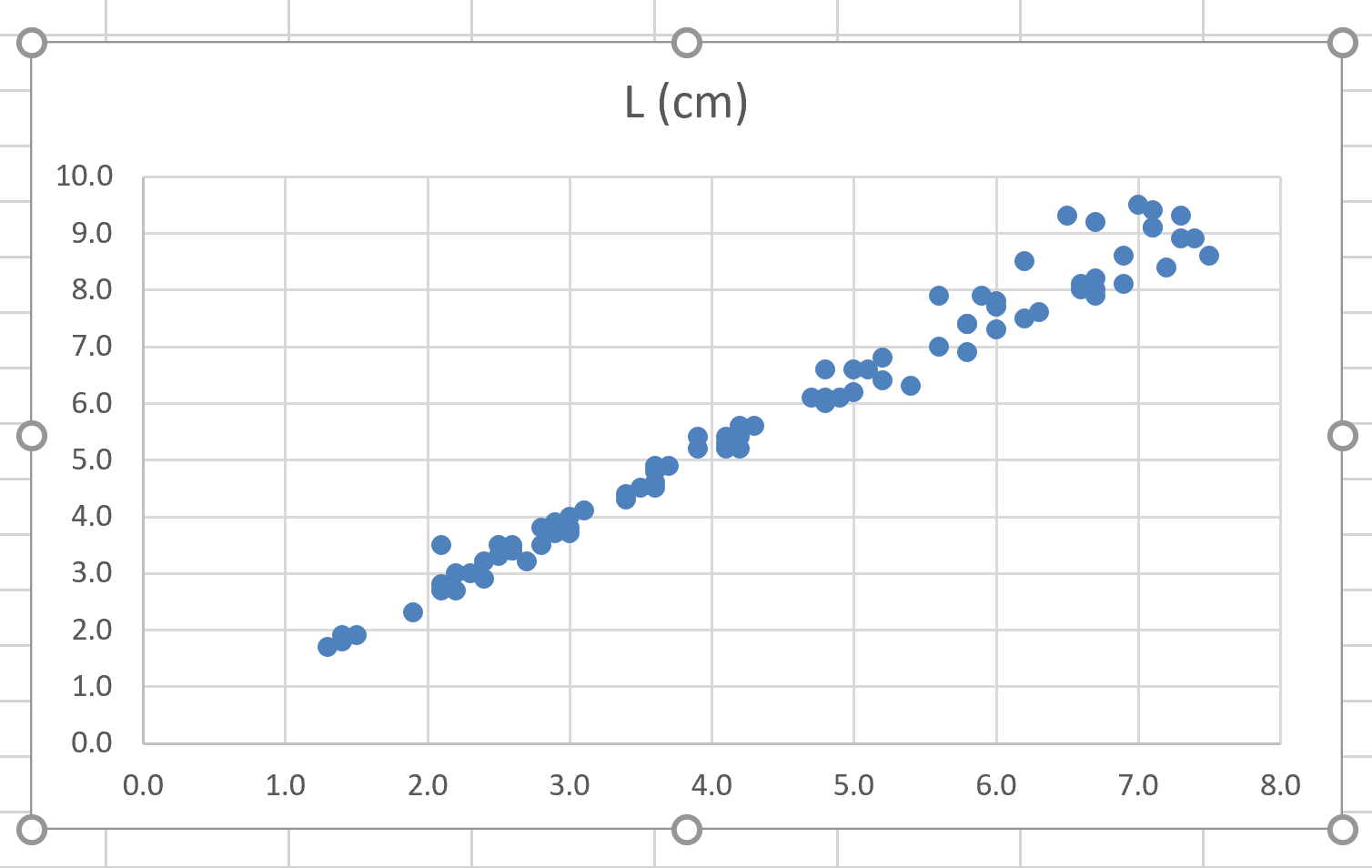
1. Notice that it has columns for date, open, high, low, close, and volume.
2. Select the date column and the close column.
3. Go to the insert menu and select a line chart.
4. Move the chart to its own sheet to see the detail better.
5. Choose a quick style fix in the design menu.
6. Change the title to AAPL Stock Price.
7. Verify that the horizontal axis shows the dates, and the vertical axis shows dollar values.
8. Observe that over the past year the stock has gone up with a little hiccup about a month ago.

#### Multiple columns of data

To handle more than one column of data for the same dates:

1. Observe lines for each on the same chart can be shown.
2. Select the date plus the high and low values for AAPL stock.
3. Since the high and low aren't all that far apart, change the range for the dollar amount on the left to start at 100.
4. Select the vertical axis, right-click, and format axis.
5. Observe both the high and low-value lines now and see the spreads between them.

#### Scatterplot



1. To plot two different variables, closing price, and volume, for AAPL stock, choose the scatterplot.
2. Observe a graph with the closing price on the horizontal axis and the volume of trade that day on the vertical axis.
3. Observe that the prices seem to cluster in a couple of areas, and that they have about the same volume generally, though there are some high-volume days at the lower price.

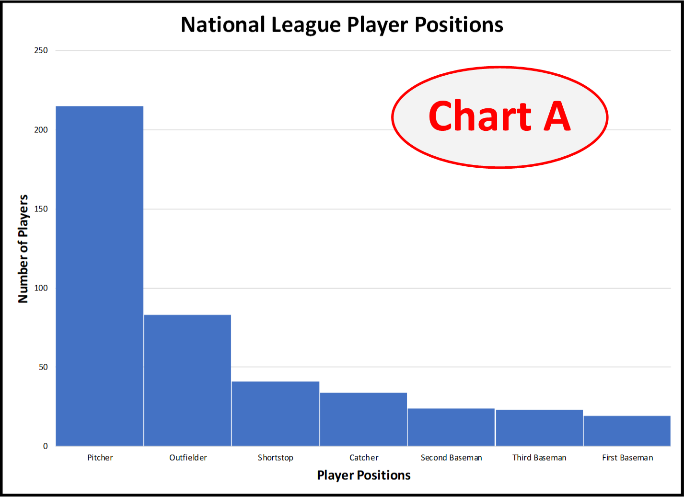
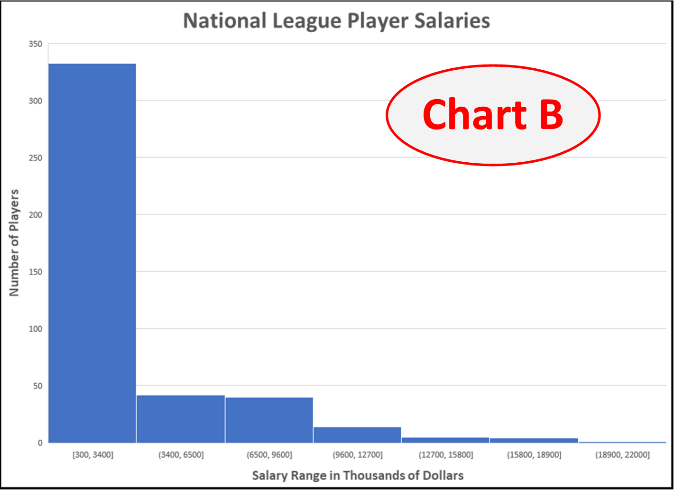
## Histograms

A Histogram is a column chart that measures the frequency of data in a data set and specifically groups numerical values into bins we define.

### Column Charts vs Histogram

* Recall that we previously created a **column chart**to compare counts of categories within a data set. This kind of chart answers a question like: how many players are there in each playing position in the league?
* But what if we want to ask the question: how many players made under $1 million in salary, and between $1 and $2 million, and between $2 and $3 million in salary? This kind of chart is called a histogram, and the groupings we choose such as, 1) all salaries between $1 and $2 million, and 2) salaries between $2 and $3 million, are the bins.

There are two ways to do this in Excel.

A is a Column Chart but B is a Histogram

**Analysis tool pack add-in:**

1. We'll start with a method that works on both Windows and Mac using the histogram tool in the analysis tool pack add-in. Instructions for loading the analysis tool pack add-in are given in the Getting Started instructions.
2. To create the histogram:
   1. Choose data analysis from the data menu on Windows or from the tool’s menu on Mac. Choose histogram, which opens a dialog.
   2. For the input range, select the data from the salary’s column.
   3. For the bin range, select the bin intervals you've created.
   4. If you have a label at the top of your columns, click labels.
   5. For the output options, select new worksheet and chart output.
   6. Press OK.

**Insert chart**

* Available in Excel 2016 for PC. The tool pack histogram requires two columns of data. One for data you want to analyze, and one for bin levels that represent the intervals for the bins. In the video example, I started at $1 million, then $2 million, et cetera, up to $15 million. When I created the histogram, the number of values in the salaries lists that are below $1 million will be in the first bin. The numbers of salaries between $1 and $2 million will be in the second bin, and so on.
* To create the histogram:
  1. Select your data and click insert, recommended charts, and choose the histogram chart.
  2. To configure details about the bins, right-click the horizontal axis of the chart, click format axis and then click axis options.
  3. The dialog provides options for choosing categorical data like the player positions or automatic for numerical data. You can specify the number of bins that you might choose to experiment with a bit. Note: If you choose bins that are too narrow, the result can be noisy. On the other hand, too few bins will hide details.
  4. As with other charts, the design and layout can be further customized from the design menu when the chart is selected.

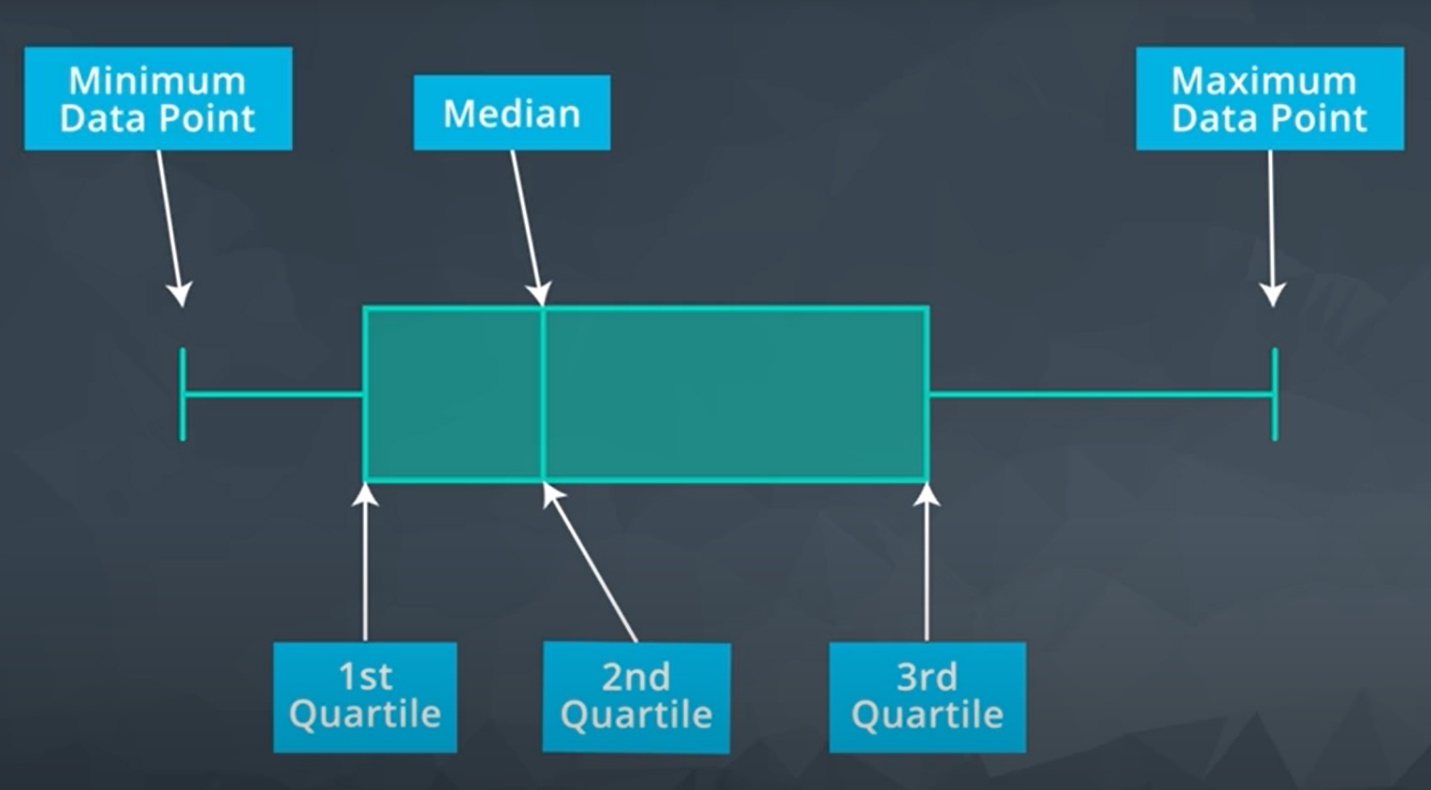
## Box Plots

*A box plot, which in our case is really a box and whisker plot, is the visualization of statistical spread in a data set of values.*

**The five numbers summary**

A traditional box plot is built using the five numbers summary. The five numbers summary consists of five values.

* maximum
* minimum
* 1st quartile
* 2nd quartile, aka 'median'
* 3rd quartile



Box plot description

Where we make a box and whisker plot:

* Maximum becomes the tip of the upper whisker.
* Minimum becomes the tip of the lower whisker.
* The box represents the middle half of the data with a line where the median is.

**Note**: Excel will give us a bonus of six numbers in the summary by placing an X at the mean or average value of the set.

1. Creating a box plot in Windows Excel 2016 is as easy as any other chart.
2. Select the appropriate columns of data.
3. Click insert in recommended charts.
4. Click the box in the whisker chart. Remember that a box plot represents statistics for a single list of numbers. So, each list you select will be represented by its own box plot.
5. Observe that the box plot visually gives a sense of the spread of the value list.
6. Adjust the range so that you can see the plots a little better, if needed.
7. Give the chart a title.