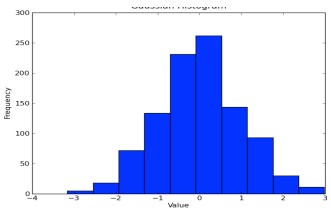
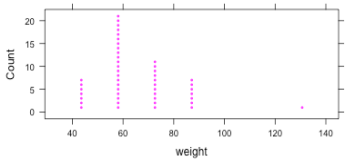
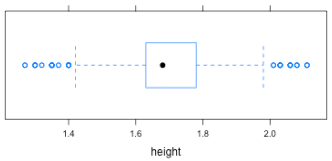
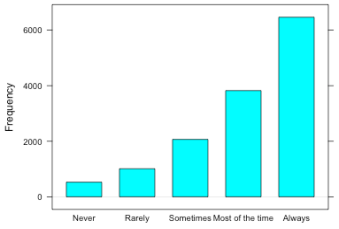
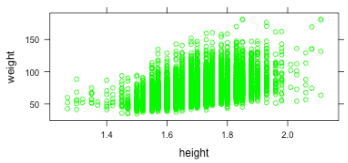
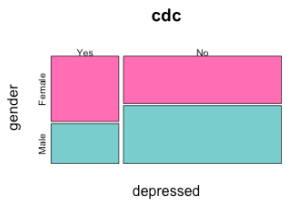
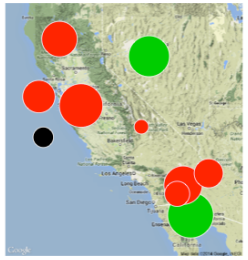
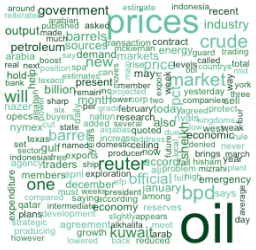
Plot Analysis Guide

Plot Type	Snapshot	What to look for	Questions Toolkit	Fun Facts
 Histogram	*A graphical summary for ***numerical*** data. *Visualizes the distribution of a numerical variable. *Observations are grouped into bins, and bars are drawn to show how many observations (or what proportion of the observations) lie in each bin.	***Shape:*** symmetric or skewed ***Center:*** the typical value if unimodal, typical values by group if bimodal or multimodal ***Spread:*** the amount of variation in the data ("spread" refers to horizontal width)	*What type of plot is this? *How many variables are displayed in this plot? Name the variable(s). *What do the x and y-axes represent? *What type of variable does this plot show? *What shape does this distribution have? What does that tell us? *What is the typical value of this data set? Are there any extreme values (i.e. outliers)?	If made with large bins, we see a very crude picture of the distribution. If we make small bins, we get a very "choppy" picture. There is an art to finding the right width for bins, and most software programs make a good choice, but it can be fun to examine other widths. Statisticians call this activity (varying bin widths) an example of "***smoothing***."
 Dot Plot	*A graphical summary for ***numerical*** data. *Visualizes the distribution of a numerical variable. *Observations are represented individually as a dot.	***Shape:*** symmetric or skewed ***Center:*** the typical value if unimodal, typical values by group if bimodal or multimodal ***Spread:*** the amount of variation in the data ("spread" refers to horizontal width)	*What type of plot is this? *How many variables are displayed in this plot? Name the variable(s). *What do the x and y-axes represent? *What type of variable does this plot show? *What shape does this distribution have? What does that tell us? *What is the typical value of this data set? Are there any extreme values (i.e. outliers)?	Dot plots are usually best with small sample sizes. If the sample gets too large, the dots become blurred together and make the plot much more difficult to read. Dot plots can be helpful for beginners, because each observation is represented explicitly.

Plot Type	Snapshot	What to look for	Questions Toolkit	Fun Facts
 Box Plot	*A graphical summary for numerical data. *The 5-number summary is used to create the box by using the median and the first and third quartiles.	***Shape:*** symmetric or skewed ***Center:*** median value ***Spread:*** the amount of variation in the data ("spread" refers to horizontal width)	*What type of plot is this? *How many variables are displayed in this plot? Name the variable(s). *What type of variable does this plot show? *What shape does this distribution have? What does that tell us? *Can you determine modality from this plot? *What is the typical value of this data set? *Are there any extreme values (i.e. outliers)? How do you know?	Boxplots are great when we have skewed data because they clearly identify potentially unusual points by single points. However, the overall shape of the distribution is impossible to tell in a boxplot. We may know the data is right skewed, but there is no way to tell if the data is unimodal, bimodal, etc. Boxplots also tend to send many false alarms for unusual points: think of them as points that *potentially* do not fit with the pattern Boxplots are useful for comparing two groups of data.
 Bar Graph	*A graphical summary for ***categorical*** data. *Observations are grouped into bars to show how many times each value occurred.	*Heights of bars (frequency/count)	*What type of plot is this? *How many variables are displayed in this plot? Name the variable(s). *What do the x and y-axes represent? *What type of variable does this plot show? *Which category has the highest frequency? Is there more than one such category? * Is there a "typical" outcome, or do people seem to be in disagreement?	You can order the bars from shortest to tallest or tallest to shortest. You can also simply order them by the values in the x-axis. For example, if we have high school students, it might make sense to order them as Freshmen, Sophomores, Juniors, and Seniors, rather than ordering them by bar size.

Plot Type	Snapshot	What to look for	Questions Toolkit	Fun Facts
 Scatter Plot	*A graphical summary for ***two numerical*** variables. *Each observation provides two numerical values, and these provide an x and y coordinate.	***Form:*** Linear, non-linear, exponential, quadratic, etc. ***Strength:*** Are the points vertically close together or far apart? ***Direction:*** Positive or negative trend	*What type of plot is this? *How many variables are displayed in this plot? Name the variable(s). *What do the x and y-axes represent? *What type of variable(s) does this plot show? *What type of trend do you see? Is there a positive or negative trend? *Is the data linear? If not, what type of shape does it have? *Are the points really close together or are they far apart? What does that mean? • What does this tell us about the relationship, if any, between x and y?	Scatterplots can show us the relationship between two numerical variables. They are especially useful when trying to make predictions about an unobserved y value when the x value is known.
 Mosaic Plot	*A graphical summary for ***two categorical*** variables. *Observations are grouped into subsets and their box sizes are adjusted accordingly.	Areas of individual boxes.	*What type of plot is this? *How many variables are displayed in this plot? Name the variable(s). *What do the x and y-axes represent? *What type of variable does this plot show? *Which subset has the highest frequency? *Can you get exact values from this plot? *What does this plot tell us about the relationship, if any, between x and y?	Mosaic plots are similar to scatterplots in that they are able to show relationships between two variables. However, instead of numerical variables, we are able to find relationships between categorical variables. The width represents the proportional representation of one variable, the height represents the other.

Plot Type	Snapshot	What to look for	Questions Toolkit	Fun Facts
 Maps	*A graphical display of data overlaid on top of a geographical map. *The maps are taken from GoogleMaps and data is then transposed onto them.	*Placement of points on the map.	*What type of plot is this? *What type of variable does this plot show? *What do the points represent on the map? *What patterns do we see? *What do clusters mean?	Maps can help identify "hotspots" in certain locations.
 Word Cloud	*A graphical summary of word frequency from an article of text. *The thickness of each word implies how often the word appeared in the text.	*Thickness of each word.	*What type of plot is this? *What type of variable does this plot show? *Why are some words larger than others? *Can you tell how many times a word appeared in the text?	Word clouds are useful tools for word frequency analysis. Words that are used most often throughout a text are larger, and words that are rarely used in the text are smaller.