Data viz in R Cheat Sheet

by nikkism via cheatography.com/138448/cs/36106/

Geoms

What is a Geom?

A geom is a geometric object and is a function that controls the way in which your data is visualized.

Basic Graph Features:

geom_blank():	Creates a blank canvas
geom_path():	Data points are joined according to how they are ordered in the data
geom_line():	Data points are connected according to the order on the x axis
geom_ribbon():	A line graph that has an area highlighted above and below the line. The thickness of this highlighted part is defined by a y-min and y-max
geom_segment():	Connects 2 data points with a line segment
geom_rect():	Create rectangles
geom_polygon():	Create polygons
geom_text():	Add labels and text

Single variables

Discrete:

geom_bar():	Create a bar graph

Continuous:

geom_histogram():	Create a histogram (to show distribution of a continuous variable)
geom_density():	Create a density plot (a smoothed version of a histogram)
geom_dotplot():	Each dot represents an observation where the size of the dot is the bin width
geom_freqpoly():	A frequency polygon for when you want to compare the distribution of various elements in a category. An alternative to stacking histograms. With a histogram you display the number of observations using a bar, but with a frequency polygon you use lines.

Two variables:

Both continuous:

geom_point():	Scatterplot
geom_quantile():	Drawing a line through a regression
geom_smooth():	Add a line of best fit

Show distribution:

geom_bin2d():	Creates a heatmap - as an alternative to geom_point if too many points
geom_density2d():	Creates a 2D density plot
geom_hex():	An alternative to geom_bin2d() but the bins are hexagons

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Two variables: (cont)	
At least one discrete:	
geom_count():	When there are too many points in a specific location on your plot, you can count them and create a group. This helps when there are too many data points to plot effectively (this is to prevent overplotting)
geom_jitter():	Adds random variation (dots) at each data point
One continuous, one discrete:	
geom_bar(stat = "identity"):	geom_bar uses stat="bin" as its default making the height of each bar equal to the number of cases in each group. If you want the heights of the bars to represent values in the data, use stat="identity" and give the y aesthetic a value.
geom_boxplot():	Box plots
geom_violin():	Violin plot (like a box plot but instead of a box, you have the shape of how the data is distributed)
One time, one continuous	
geom_area():	Area plot
geom_line():	Line plot
geom_step():	Step plot - Connects data points as they change creating a line that looks like a staircase
Spatial:	
geom_map():	Create a map with geographical data

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com