STAT 100 - Choosing the Right Graph

Numerical variable: Takes on values that are numbers, which you can measure and "do math" with

- *Ex: Salary ($100k, $50k, $70k, $55k) → Average salary is $68.75k*

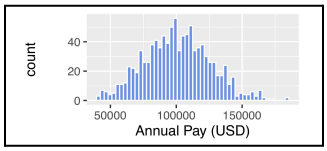
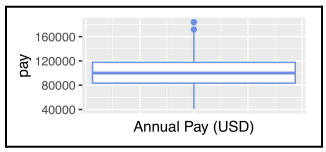
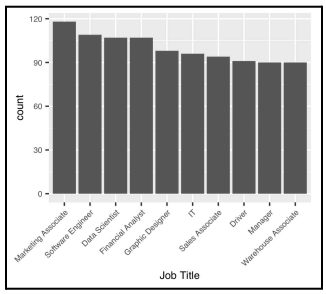
Categorical variable: Takes on values that are labels, which you use to group the data

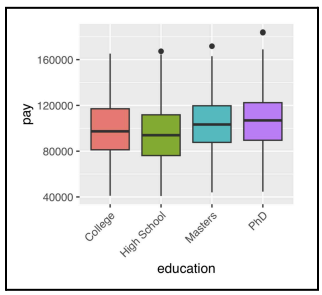
- *Ex: Income Level (low, middle, high) → Groupings that you can't "do math" with*

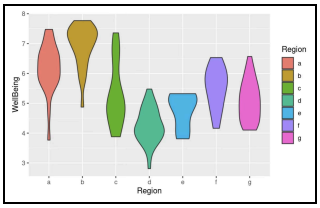
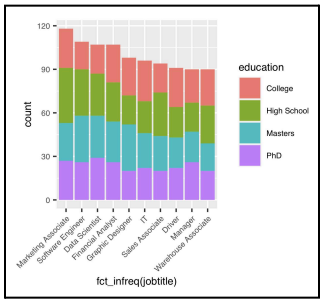
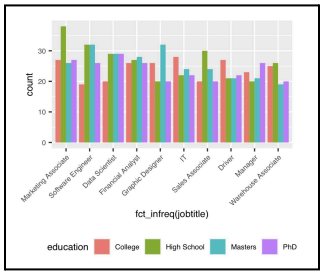
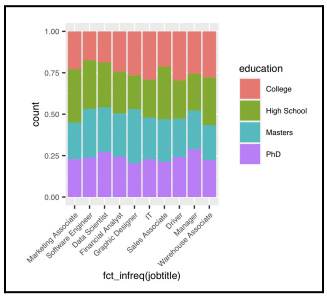
Explanatory variable: Expected cause ("input")

Response variable: Expected result of explanatory variable ("output")

- *Ex: Measuring the effect of education level (explanatory) on salary (response)*

	Response variable	Graph	Template code	Image
1 variable	Numerical	**Histogram**	ggplot(data = ---, mapping = aes(x = ---)) + geom_histogram()	
		Boxplot	ggplot(data = ---, mapping = aes(y = ---)) + geom_boxplot()	
	Categorical	**Barplot**	ggplot(data = ---, mapping = aes(x = ---)) + geom_bar()	

	Response variable	Explanatory variable	Graph	Template code	Image
2 variables	Numerical	Categorical	**Side-by-side boxplots**	ggplot(data = ---, mapping = aes(y = ---, x = ---, fill = ---)) + geom_boxplot()	

			Side-by-side violin plots	ggplot(data = ---, mapping = aes(y = ---, x = ---, fill = ---)) + geom_violin( )	
	Categorical	Categorical	**Segmented barplot (stacked)**	ggplot(data = ---, mapping = aes(x = ---, fill = ---)) + geom_bar()	
			Segmented barplot (side-by-side)	ggplot(data = ---, mapping = aes(x = ---, fill = ---)) + geom_bar(position = "dodge")	
			Segmented barplot (relative frequency)	ggplot(data = ---, mapping = aes(x = ---, fill = ---)) + geom_bar(position = "fill")	
	Numerical	Numerical	**Scatterplot**	ggplot(data = ---, mapping = aes(y = ---, x = ---)) + geom_point()