

The "No Points Left Behind" Cheatsheet

Getting points taken off for silly mistakes? It's okay, we all are! Use this cheatsheet to remind yourself of all you can do to make the best graphs possible!

Order those categories!

Our brains process things in order first! We might confuse the audience if the categorical data is shown out of order.

Ex: Instead of "Fair", "Poor", "Good", do "Poor, Fair, Good"



Having trouble arranging the order? Here's a fix:

Without ggplot, you can refactor a column, replacing the levels attribute with the correct order of your categories:

```
data$column= factor(data$column, levels = c("normal", "moderate", "severe"))
```

Running a test? Make sure you interpret those p-values!

The p-value tells you how likely your event happened randomly. A higher p-value tells you that your data isn't very significant, just random, where a smaller one is more indicative that there is something significant about your data.

If the p-value < 0.05:

We reject the null hypothesis. The data is statistically significant and has strong evidence against the null hypothesis.



If the p-value > 0.05:

We fail to reject the null hypothesis. The data is not statistically significant and is not enough evidence against the null hypothesis. In place of this sentence, put this in context of your data.

Need a ggplot reminder?

Here are the basics:

```
ggplot(data, aes(x="column1", y="column2")) +  
<GEOM_FUNCTION> +  
<LABELS> +  
<THEME>
```

Forgot which geom function to use for your plot?

For Line Segments:

geom_abline, geom_hline, geom_vline

For Discrete:

geom_bar

For Continuous Linear:

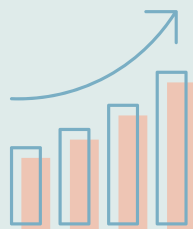
geom_histogram, geom_density, geom_area, geom_freqpoly

For Two Variable Both Continuous:

geom_label, geom_smooth(), geom_text(), geom_quantile

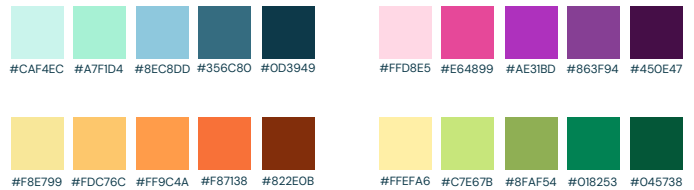
For Discrete and Continuous:

geom_col, box_plot()



Lets get a color scheme going!

People like to look at a pretty graph. Use color schemes that are easy to discern, and nice to look at. Here are my favorites!



Remember, the dependent variable goes on the y-axis!

Whether its a scatterplot or a mosaic plot, the dependent variable still goes on the y-axis!

The dependent variable is what changes based on the independent variable, which doesn't change.

You can generally just set the y column to the dependent variable's column.

In a mosaic plot, it goes: Dependent ~ Independent

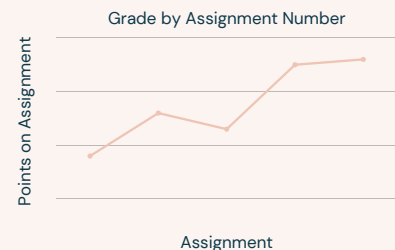


Have a uniform theme with good titles and labels!

After you make your plot, adding titles and axis labels make your visualization easier to digest, and makes the purpose of your plot very clear!

The following code lets you add a label, and center it! You can customize the theme function to your liking.

```
ggplot(data, aes()) +  
<GEOM_FUNCTION> +  
labs(title="Title", x = "x_label", y = "y_label", cex = 1) +  
theme(plot.title = element_text(hjust = 0.5), legend.position = "right", axis.text.x = element_text(angle = 60, hjust = 1, vjust = 1))
```



Make sure to write out all your assumptions

Sometimes you have to make some decisions that don't really have a right or wrong answer. In that case, make sure you state this decision, to make it clear to the audience of your interpretation!

Ex: We choose to include N/A results because it is an important aspect that contributes to this visualization.

