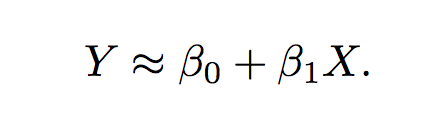
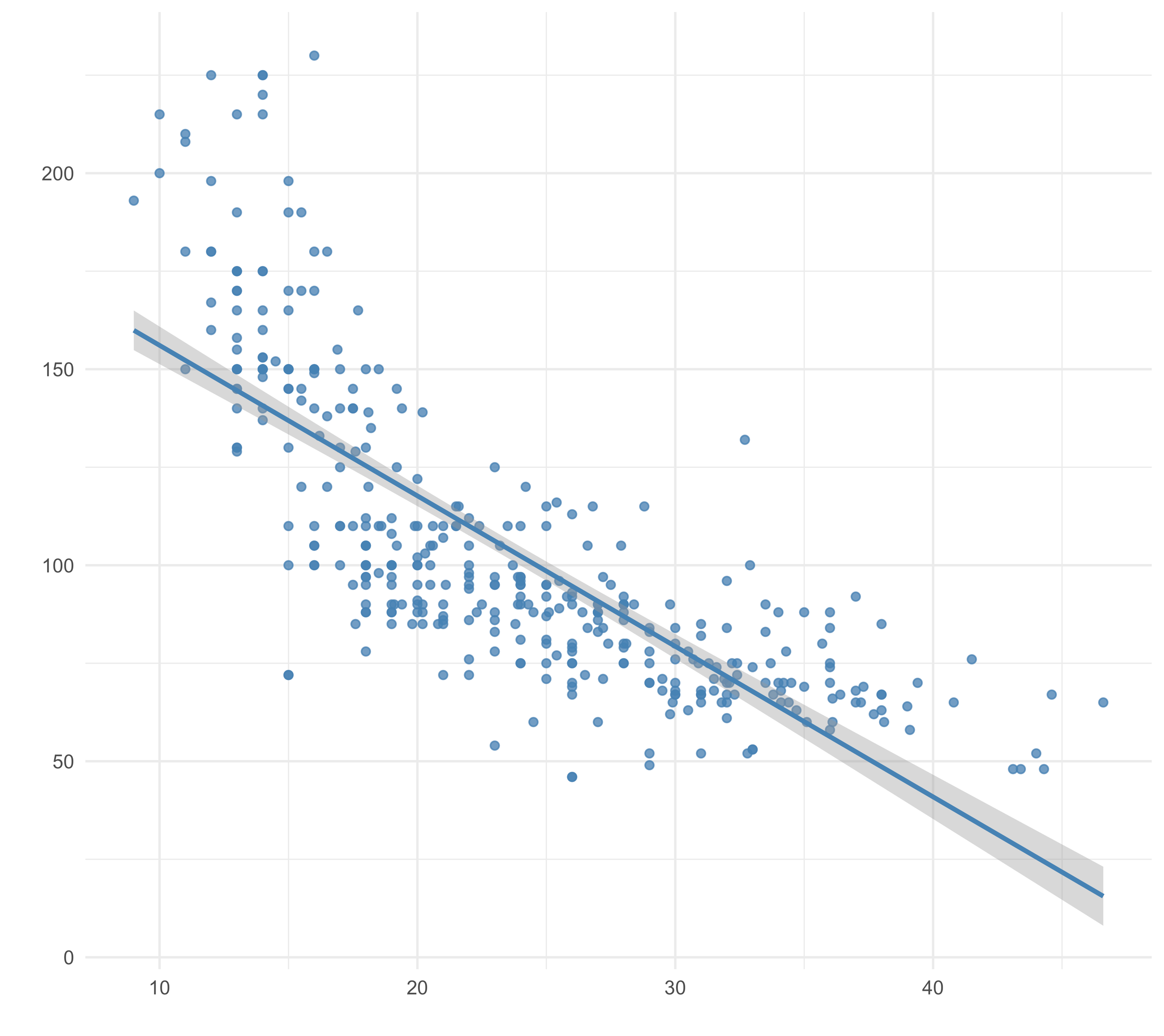
**CHEAT SHEET 1: LINEAR REGRESSION**

*Formula:*

**

Where Bo = intercept, and b1 equals slope of the line.

*Plot:*

**

*Suitable for*: Predicting the value of a quantitative variable when there exists a linear relationship with the predictor variable(s).

*Procedure*: **(1) Explore the relationships** using visualisation and correlation matrices. **(2) Run the model**. **(3)** **Examine** the model fit, using the plot() function.

*Interpreting the model*. **Estimate** is the estimated value of x, the **standard error** is the 95% confidence interval boundaries (+ or -) of the estimate; the t value is the coefficient divided by its standard error; the **p value** gives an indication of whether or not the slope of the line differs significantly from 0; the **residual standard error** indicates how far, on average, the points are from the regression line; the **r2** is an indication of how much variance in *y* in explained by *x*, and the **F statistic** gives an indication of whether or not the model differs significantly from an intercept-only model (and hence whether or not it has been worth running the model).

*Actions* *for improving models*. **(1) Transform** the variables using the log or quadratic functions (note specific rules exist around when to use transformation). **(2)** **Add or remove variables** and re-run the model.

*Considerations*. **(1) Collinearity**. Can be resolved by removing variables or by adopting other. **(2) Outliers/high leverage points**. Can be an issue because they can exert considerable leverage on the model. Need to use own judgement re: when to remove. **(3) Sample size** Small samples can lead to underpowered models. **(4) Qualitative predictors**. Qualitative predictors need to be dealt with carefully, as they can influence the coefficients. One way is by dummy coding. **(5) Error terms** need to be uncorrelated.