Basics to R

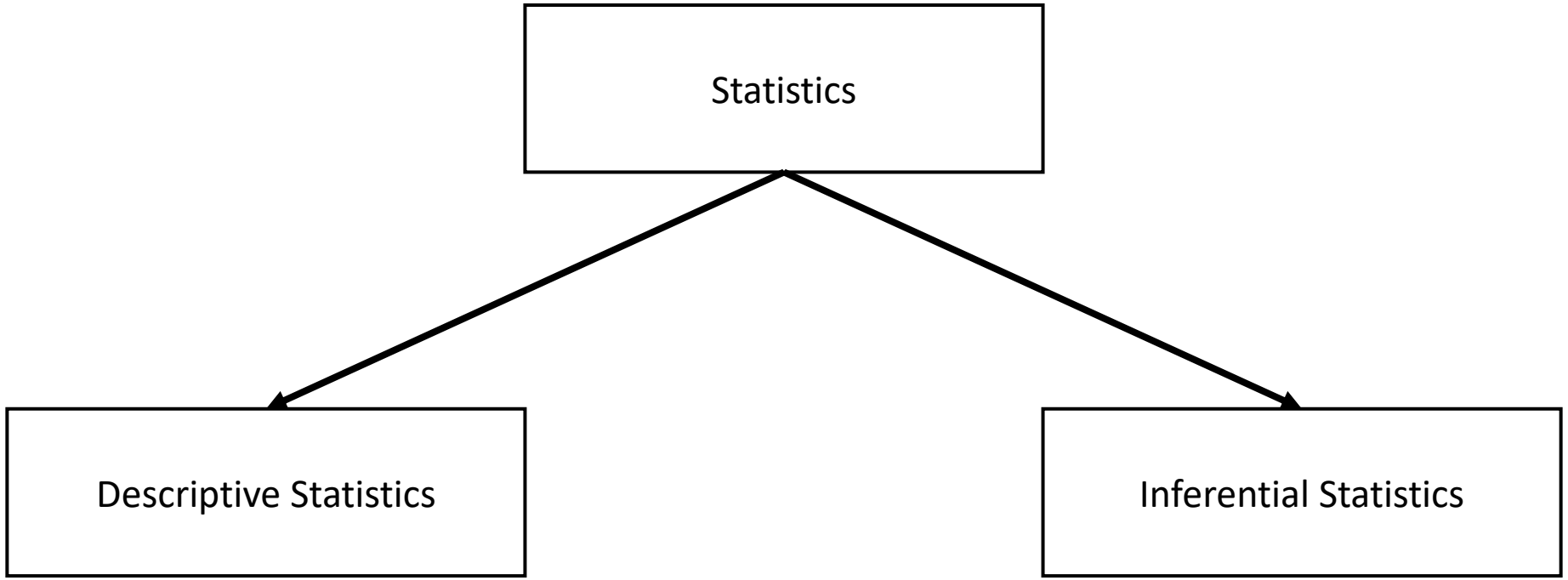
Correlation, Linear Regression, and ANOVA

December 4, 2020

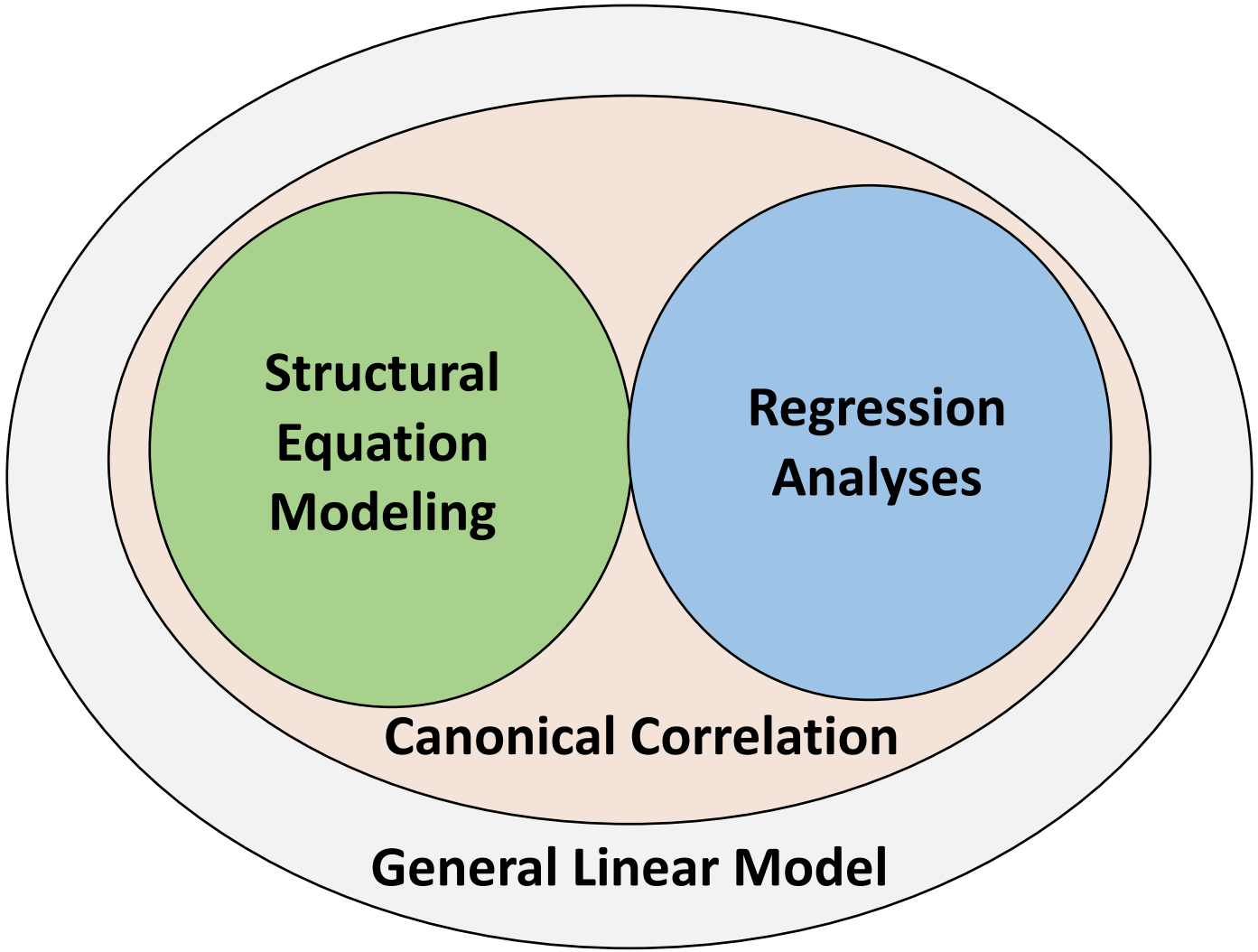
Outline for Today

- What are inferential statistics?
- The General Linear Model
- Correlation, Linear Regression, and ANOVA
- What do I need to look at?
- Basic data visualization of inferential statistics in *ggplot2*

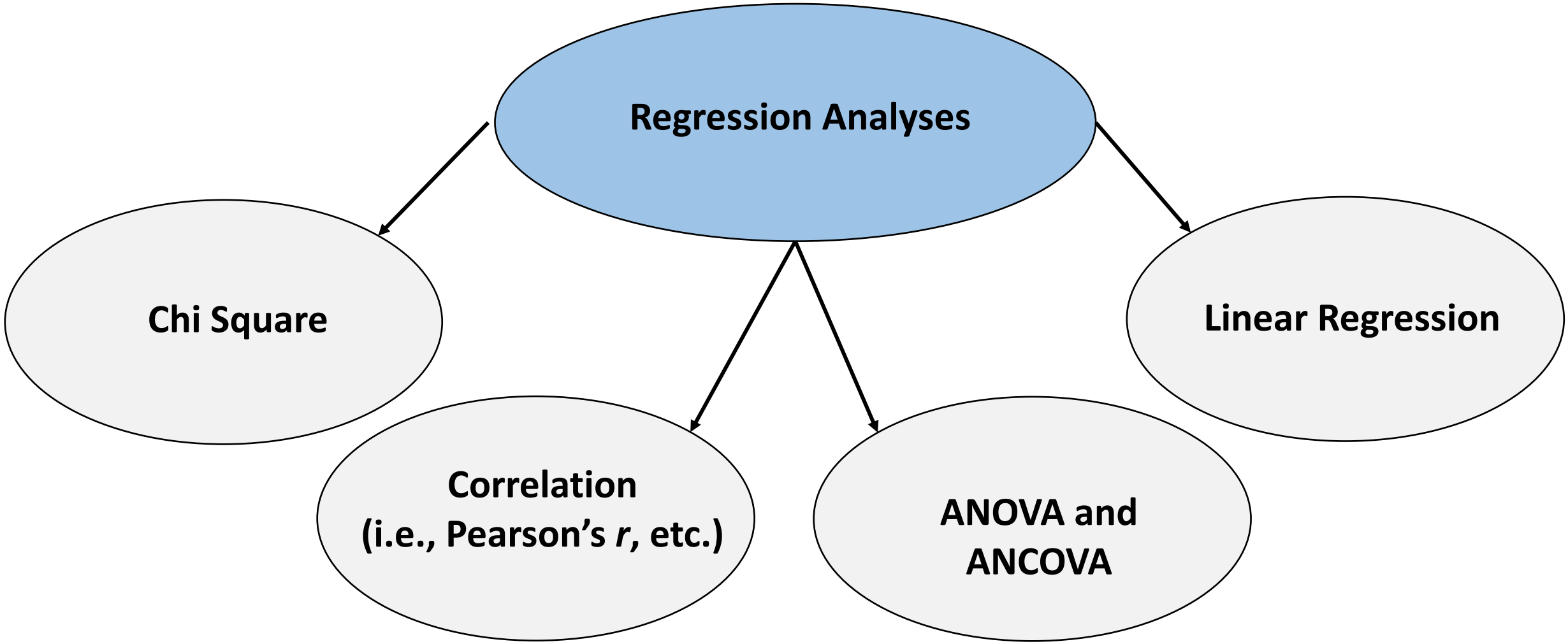
What are inferential statistics?



What are inferential statistics? A very brief introduction to the General Linear Model



What are regression analyses?



What are the assumptions of regression analyses?

- Linearity
- Independence
- Normality
- Equivariance (i.e., Homoscedasticity)

Assumption #1: Linearity

The regression model is linear in parameters

$Y = \alpha + (\beta 1 * X1) + (\beta 2 * X2)$

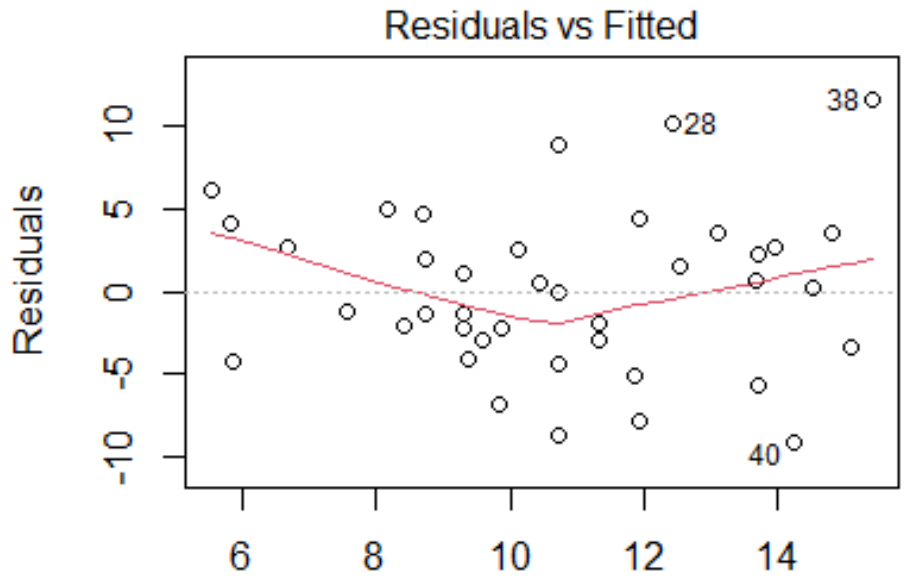
$Y = \alpha + (\beta 1 * X1) + (\beta 2 * X2^2)$

$Y = \alpha + (\beta 1 * X1) + (\beta 2 * X2^2) + (\beta 2 * \ln(X2))$

$Y \neq \alpha + (\beta 1 * X1) + (\beta 2 * X2)^2$

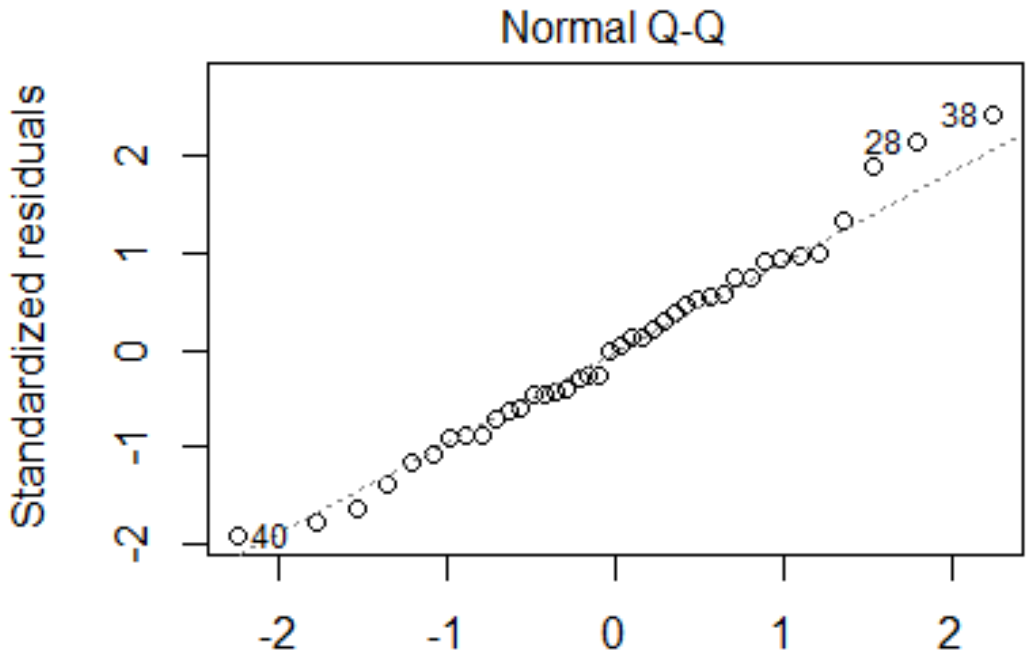
Assumption #2: Independence

Observations are independent of each other



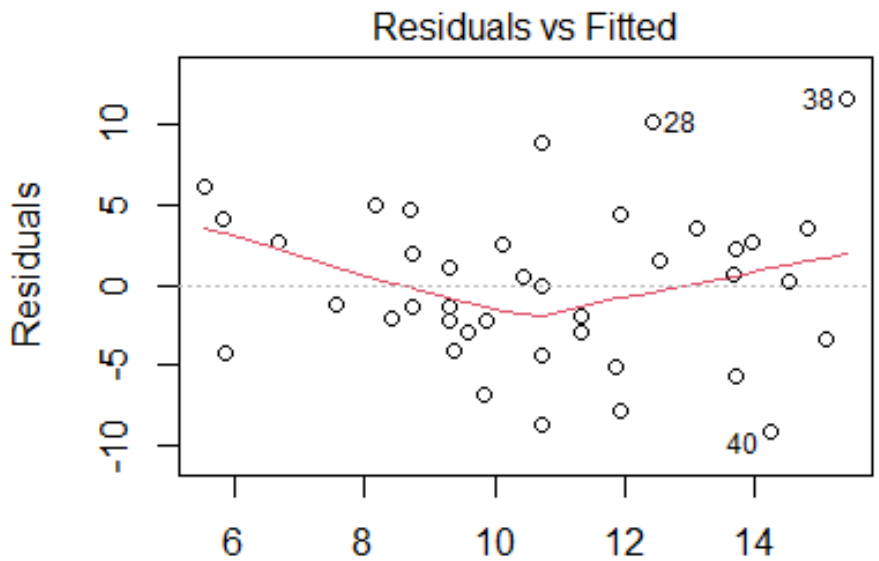
Assumption #3: Normality

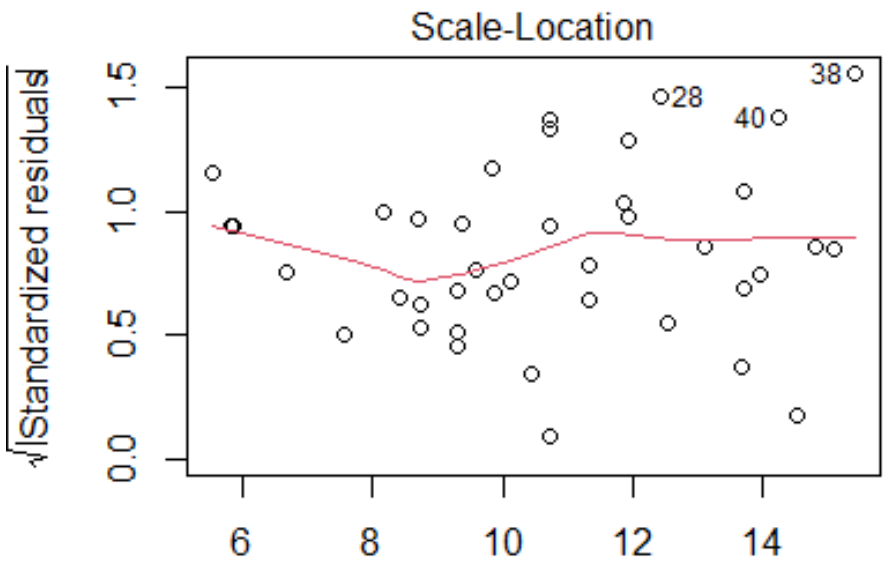
Residual errors are normally distributed



Assumption #4: Equivariance or Homoscedasticity

<u>The variance of residuals should not increase with fitted values of the response variable.</u>





Summary

- The fundamental basis of most statistics is rooted in the same mathematics
- This means that how you write your R code is pretty much the same for many common analyses
- Ryan Gosling stats memes always work