library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

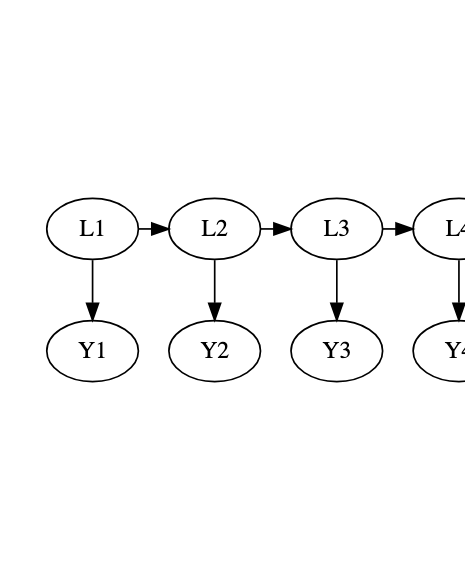
library(DiagrammeR)

# A Quick Introduction to Knowledge Tracing

The knowledge tracing model [(Corbet and Anderson, 1995)](CorbettAnderson1995.pdf) posits that there is a latent skill, whose value at Time is . Each time the student practices the skill there is a chance that the student learns the skill (). The diagram below shows this struction, where are the latent skills, and are the observed correct/incorrect scores for the items.

DiagrammeR::grViz('  
digraph KT {  
 rankdir="BT"  
{rank="max"  
 L1 -> L2 -> L3 -> L4 -> L5 -> L6 -> L7 -> L8  
}  
{rank="min"  
 Y1; Y2; Y3; Y4; Y5; Y6; Y7; Y8  
}  
  
L1 -> Y1  
L2 -> Y2  
L3 -> Y3  
L4 -> Y4  
L5 -> Y5  
L6 -> Y6  
L7 -> Y7  
L8 -> Y8  
  
}')

## PhantomJS not found. You can install it with webshot::install\_phantomjs(). If it is installed, please make sure the phantomjs executable can be found via the PATH variable.



Following [Van de Sande, 2013](VandeSande2013.pdf), this model has four parameters.

* – Probability student has the skill at Time 0 (before the first exercise).
* – Probability that a student who is a non-master makes the transition to mastery at a practice occasion
* – Probability that a non-master gets the item correct (guessing)
* – Probabiity that a master gets the item incorrect (slipping)

After some algebra, it can be seen that the students probability of success at Time can be written as

where (and is a complex function of all of the parameters. Thus we can fit this model by using logistic regression.

# The Data

The data in <homework.csv> is simulated data from a homework system. There are 100 students and there are four variables.

* SID the student ID
* group a demographic group, which takes the values Focal and Reference. (Focal and reference are names used in differential item functioning analysis, usually the “focal” group is one which has some history of discrimination, and the reference group is a higher status comparison group.)
* attempt the trial number (or time in the above discussion)
* success A logical value indicating whether the student was success at that trial or not.

homework <- read.csv("homework.csv")  
homework$group <- as.factor(homework$group)  
summary(homework)

## SID group attempt success   
## Length:800 Focal :400 Min. :1.00 Mode :logical   
## Class :character Reference:400 1st Qu.:2.75 FALSE:278   
## Mode :character Median :4.50 TRUE :522   
## Mean :4.50   
## 3rd Qu.:6.25   
## Max. :8.00

# The assignment

## 1. Learning curve

Calculate the fraction of successful trials for each time point (1 through 8). Plot the trial number versus the success probability. This is known as the *learning curve*.

## 2. Plot the data with an estimated curve.

Note: you will want to jitter the values so they can be seen, and also add a lowess curve.

## 3. Fit a logistic regression

Use the model success ~ attempt Write up a summary of the results.

## 4. Interpret the slope

Explain in words what the slope means.

## 5. Add color to the plot according to the group

Describe the difference if any between the groups.

## 6. Pick the right model with the group

Fit both the model success ~ attempt + group (no interaction with slope) and success ~ attempt \* group (adds an interaction) and compare them to the no group model success ~ attempt which you fit before. Which model do you like better and why?

## 7. Describe the model you picked in words and equations.

## 8. Make a graph of the final model

Hint: Predict the success rate for people in both the reference and focal groups at attempts 1 through 8 and plot that.