# Project 8 Template

## Sofia Guo

```r
# Add to this package list for additional SL algorithms
pacman::p_load(
  tidyverse,
  ggthemes,
  ltmle,
  tmle,
  SuperLearner,
  tidymodels,
  caret,
  dagitty,
  ggdag,
  here,
  randomForest,
  ggdag)

#source the pretty dag function
source("pretty_dag.R")

heart_disease <- read_csv('heart_disease_tmle.csv')
```

```
## Rows: 10000 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (14): age, sex_at_birth, simplified_race, college_educ, income_thousands...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## Introduction

Heart disease is the leading cause of death in the United States, and treating it properly is an important public health goal. However, it is a complex disease with several different risk factors and potential treatments. Physicians typically recommend changes in diet, increased exercise, and/or medication to treat symptoms, but it is difficult to determine how effective any one of these factors is in treating the disease. In this project, you will explore SuperLearner, Targeted Maximum Likelihood Estimation (TMLE), and Longitudinal Targeted Maximum Likelihood Estimation (LTMLE). Using a simulated dataset, you will explore whether taking blood pressure medication reduces mortality risk.

## Data

This dataset was simulated using R (so it does not come from a previous study or other data source). It contains several variables:

- **blood_pressure_medication**: Treatment indicator for whether the individual took blood pressure medication (0 for control, 1 for treatment)

- **mortality**: Outcome indicator for whether the individual passed away from complications of heart disease (0 for no, 1 for yes)

- **age**: Age at time 1

- **sex_at_birth**: Sex assigned at birth (0 female, 1 male)

- **simplified_race**: Simplified racial category. (1: White/Caucasian, 2: Black/African American, 3: Latinx, 4: Asian American,
5: Mixed Race/Other)

- **income_thousands**: Household income in thousands of dollars

- **college_educ**: Indicator for college education (0 for no, 1 for yes)

- **bmi**: Body mass index (BMI)

- **chol**: Cholesterol level

- **blood_pressure**: Systolic blood pressure

- **bmi_2**: BMI measured at time 2

- **chol_2**: Cholesterol measured at time 2

- **blood_pressure_2**: BP measured at time 2

- **blood_pressure_medication_2**: Whether the person took treatment at time period 2

For the "SuperLearner" and "TMLE" portions, you can ignore any variable that ends in "_2", we will reintroduce these for LTMLE.

# SuperLearner

## Modeling

Fit a SuperLearner model to estimate the probability of someone dying from complications of heart disease, conditional on treatment and the relevant covariates. Do the following:

1. Choose a library of at least 5 machine learning algorithms to evaluate. **Note**: We did not cover how to hyperparameter tune constituent algorithms within SuperLearner in lab, but you are free to do so if you like (though not required to for this exercise).

2. Split your data into train and test sets.

3. Train SuperLearner

4. Report the risk and coefficient associated with each model, and the performance of the discrete winner and SuperLearner ensemble

5. Create a confusion matrix and report your overall accuracy, recall, and precision

```
# Fit SuperLearner Model

## Train/Test split
# initial split
# ----------
disease_split <-
  initial_split(heart_disease, prop = 3/4) # create initial split (tidymodels)
```

```r
# Training
# ----------
train <-
  # Declare the training set with rsample::training()
  training(disease_split)

# y_train
y_train <-
  train %>%
  pull(mortality)
  # pull and save as vector

# x_train
x_train <-
  train %>%
  # drop the target variable
  select(-mortality)

# Testing
# ----------
test <-
  # Declare the training set with rsample::training()
  testing(disease_split)

# y test
y_test <-
  test %>%
  pull(mortality)

# x test
x_test <-
  test %>%
  select(-mortality)
```

```r
## Train SuperLearner

# set seed
set.seed(1000)

# multiple models  = 5 models
#listWrappers() to see the list of all the SuperLearner algorithms
# ----------
sl = SuperLearner(Y = y_train,
                  X = x_train,
                  family = binomial(),
                  # notice these models are concatenated
                  SL.library = c('SL.mean',     # if you just guessed the average - serves as a baseline
                                 'SL.glmnet',
                                 'SL.bayesglm',
                                 'SL.lm',
                                 'SL.caret'))
```

```
## Loading required namespace: arm
```

```
## + Fold01: mtry= 2
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set
```

```
## Warning: Setting row names on a tibble is deprecated.

## + Fold01: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 2
## + Fold01: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry= 7
## + Fold01: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold01: mtry=13
## + Fold02: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 2
## + Fold02: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry= 7
## + Fold02: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold02: mtry=13
## + Fold03: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 2
## + Fold03: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry= 7
## + Fold03: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold03: mtry=13
## + Fold04: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 2
## + Fold04: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry= 7
## + Fold04: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold04: mtry=13
## + Fold05: mtry= 2

## Warning: Setting row names on a tibble is deprecated.
```

```
## - Fold05: mtry= 2
## + Fold05: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry= 7
## + Fold05: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold05: mtry=13
## + Fold06: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 2
## + Fold06: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry= 7
## + Fold06: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold06: mtry=13
## + Fold07: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 2
## + Fold07: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry= 7
## + Fold07: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold07: mtry=13
## + Fold08: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 2
## + Fold08: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry= 7
## + Fold08: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold08: mtry=13
## + Fold09: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 2
## + Fold09: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry= 7
## + Fold09: mtry=13
```

```
## Warning: Setting row names on a tibble is deprecated.

## - Fold09: mtry=13
## + Fold10: mtry= 2

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 2
## + Fold10: mtry= 7

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry= 7
## + Fold10: mtry=13

## Warning: Setting row names on a tibble is deprecated.

## - Fold10: mtry=13
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set

## Warning: Setting row names on a tibble is deprecated.
```

```r
## Risk and Coefficient of each model
sl
```

```
##
## Call:
## SuperLearner(Y = y_train, X = x_train, family = binomial(), SL.library = c("SL.mean",
##     "SL.glmnet", "SL.bayesglm", "SL.lm", "SL.caret"))
##
##
##                       Risk        Coef
## SL.mean_All      0.2499410 0.03614453
## SL.glmnet_All    0.2352831 0.00000000
## SL.bayesglm_All  0.2355907 0.00000000
## SL.lm_All        0.2351229 0.26715275
## SL.caret_All     0.2302234 0.69670272
```

```r
# Here is the risk of the best model (discrete SuperLearner winner).
# Use which.min boolean to find minimum cvRisk in list
sl$cvRisk[which.min(sl$cvRisk)]
```

```
## SL.caret_All
##    0.2302234
```

```r
# set seed
set.seed(987)
## Discrete winner and superlearner ensemble performance
# predictions
# ----------
preds <-
  predict(sl,               # use the superlearner not individual models
          x_test,           # prediction on test set
          onlySL = TRUE)    # use only models that were found to be useful (had weights)


# start with y_test
validation <-
```

29

```r
y_test %>%
  # add our predictions - first column of predictions
  bind_cols(preds$pred[,1]) %>%
  # rename columns
  rename(obs = `...1`,      # actual observations
         pred = `...2`) %>% # predicted prob
  # change pred column so that obs above .5 are 1, otherwise 0
  mutate(pred = ifelse(pred >= .5,
                       1,
                       0))
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
```

```r
# view
head(validation)
```

```
## # A tibble: 6 x 2
##     obs  pred
##   <dbl> <dbl>
## 1     1     1
## 2     0     1
## 3     1     1
## 4     0     1
## 5     1     1
## 6     1     1
```

```r
## Confusion Matrix
# ----------
caret::confusionMatrix(as.factor(validation$pred),
                       as.factor(validation$obs))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0  314  179
##          1  876 1131
##
##                Accuracy : 0.578
##                  95% CI : (0.5584, 0.5975)
##     No Information Rate : 0.524
##     P-Value [Acc > NIR] : 3.326e-08
##
##                   Kappa : 0.1307
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.2639
##             Specificity : 0.8634
##          Pos Pred Value : 0.6369
##          Neg Pred Value : 0.5635
##              Prevalence : 0.4760
##          Detection Rate : 0.1256
##    Detection Prevalence : 0.1972
```

```
##        Balanced Accuracy : 0.5636
##
##          'Positive' Class : 0
##
```

```r
#calculate recall
recall_sl <- 1109/(1109 +192) #TP/TP+FN

#calculate precision
precision_sl <- 1109/(1109+853) #TP/TP+FP

recall_sl
```

```
## [1] 0.8524212
```

```r
precision_sl
```

```
## [1] 0.5652396
```

## Discussion Questions

1. Why should we, in general, prefer the SuperLearner ensemble to the discrete winner in cross-validation? Or in other words, what is the advantage of "blending" algorithms together and giving them each weights, rather than just using the single best algorithm (with best being defined as minimizing risk)?

2. Answer: The blending of algorithms helps reduce bias from any particular method while potentially increasing the validity of the approach by integrating strengths from a diverse set of methods. It could be thought of as trying to get the best of all worlds while minimizing the risk of error.

# Targeted Maximum Likelihood Estimation

## Causal Diagram

TMLE requires estimating two models:

1. The outcome model, or the relationship between the outcome and the treatment/predictors, $P(Y|(A, W)$.

2. The propensity score model, or the relationship between assignment to treatment and predictors $P(A|W)$

Using ggdag and daggity, draw a directed acylcic graph (DAG) that describes the relationships between the outcome, treatment, and covariates/predictors. Note, if you think there are covariates that are not related to other variables in the dataset, note this by either including them as freestanding nodes or by omitting them and noting omissions in your discussion.

```r
# DAG for TMLE
dag <- dagify(
  BloodPressure ~ Age + Sex + Cholesterol,
  Education ~ Age + Income + Sex,
  Treatment ~ BloodPressure,
  BMI ~ Age + Income + Education,
  Mortality ~ Treatment + Age + BMI,
  Cholesterol ~ Age + BMI + Income,
  Income ~ Age + Sex + Education
)

ggdag(dag)
```
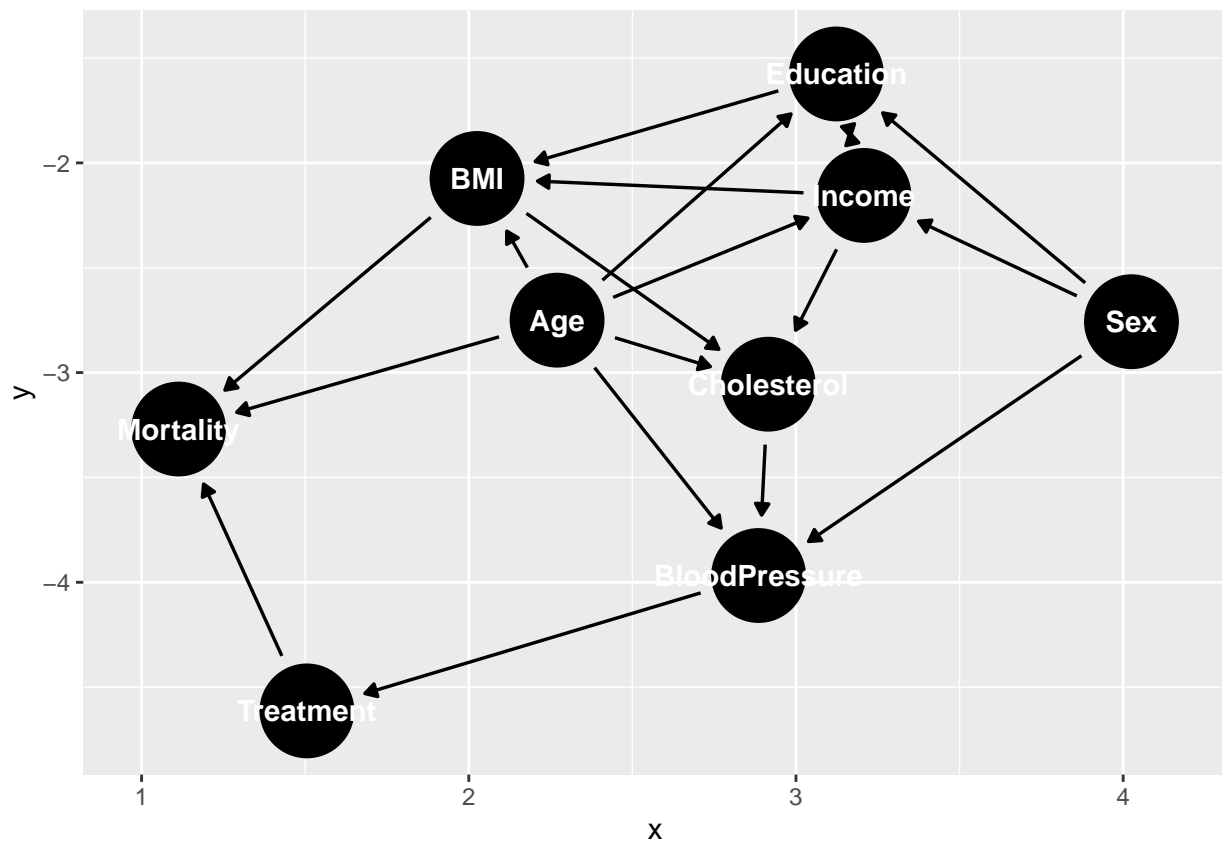
## TMLE Estimation

Use the `tmle` package to estimate a model for the effect of blood pressure medication on the probability of mortality. Do the following:

1. Use the same SuperLearner library you defined earlier

2. Use the same outcome model and propensity score model that you specified in the DAG above. If in your DAG you concluded that it is not possible to make a causal inference from this dataset, specify a simpler model and note your assumptions for this step.

3. Report the average treatment effect and any other relevant statistics

```
# Prepare data for SuperLearner/TMLE
# ----------
# Mutate Y, A for outcome and treatment, use tax, age, and crim as covariates
data_obs <-
  heart_disease %>%
  mutate(Y = mortality) %>%
  rename(A = blood_pressure_medication) %>%
  select(Y, A, age, sex_at_birth, simplified_race, college_educ, income_thousands, bmi,blood_pressure,ch

# Outcome
# ----------
Y <-
  data_obs %>%
  pull(Y)

# Covariates
```

```
# ----------
W <-
  data_obs %>%
  select(-Y)

# Treatment
A <- data_obs %>%
  pull(A)

sl_libs <- c('SL.mean','SL.glmnet','SL.bayesglm','SL.lm','SL.caret')

# set seed for reproducibility
set.seed(1000)
# implement above all in one step using tmle
# ----------
tmle_fit <-
  tmle::tmle(Y = Y,                   # outcome
             A = A,                   # treatment
             W = W,                   # baseline covariates
             Q.SL.library = sl_libs, # libraries for initial estimate
             g.SL.library = sl_libs) # libraries for prob to be in treatment

# view results
tmle_fit
```

Note: The cell above did run but it was taking so long to run that I decided not to include it in my PDF.

## Discussion Questions

1. What is a "double robust" estimator? Why does it provide a guarantee of consistency if either the outcome model or propensity score model is correctly specified? Or in other words, why does mispecifying one of the models not break the analysis? **Hint**: When answering this question, think about how your introductory statistics courses emphasized using theory to determine the correct outcome model, and in this course how we explored the benefits of matching.

2. Answer: A double robust estimator uses models of both expected outcomes and probability of treatment to estimate a treatment effect assuming that either or both of those models are specified correctly. It provides a guarantee of consistency because as the sample size increases, bias will tend towards zero for both of the models - assuming that it is unlikely both models will be biased. This assumption is likely to hold because the benefits of matching include controlling for unmeasured confounding which is what misspecification can lead to (spurious correlations); with double robust ML estimators that include propensity score matching, one can essentially control for unmeasured confounding.

# LTMLE Estimation

Now imagine that everything you measured up until now was in "time period 1". Some people either choose not to or otherwise lack access to medication in that time period, but do start taking the medication in time period 2. Imagine we measure covariates like BMI, blood pressure, and cholesterol at that time for everyone in the study (indicated by a "_2" after the covariate name).

## Causal Diagram

Update your causal diagram to incorporate this new information. **Note**: If your groups divides up sections and someone is working on LTMLE separately from TMLE then just draw a causal diagram even if it does not match the one you specified above.

**Hint**: Check out slide 27 from Maya's lecture, or slides 15-17 from Dave's second slide deck in week 8 on matching.

**Hint**: Keep in mind that any of the variables that end in "_2" are likely affected by both the previous covariates and the first treatment when drawing your DAG.

```
# DAG for TMLE
dag_2 <- dagify(
  BloodPressure ~ Age + Sex + Cholesterol + BMI,
  Education ~ Age + Income + Sex,
  Treatment ~ BloodPressure,
  BMI ~ Age + Income + Education,
  Mortality ~ Treatment + Age + BMI,
  Cholesterol ~ Age + BMI + Income,
  Income ~ Age + Sex + Education,
  BloodPressure_2 ~ Treatment_2 + BMI_2 + Cholesterol_2,
  Treatment_2 ~ BloodPressure
)

ggdag(dag_2)
```
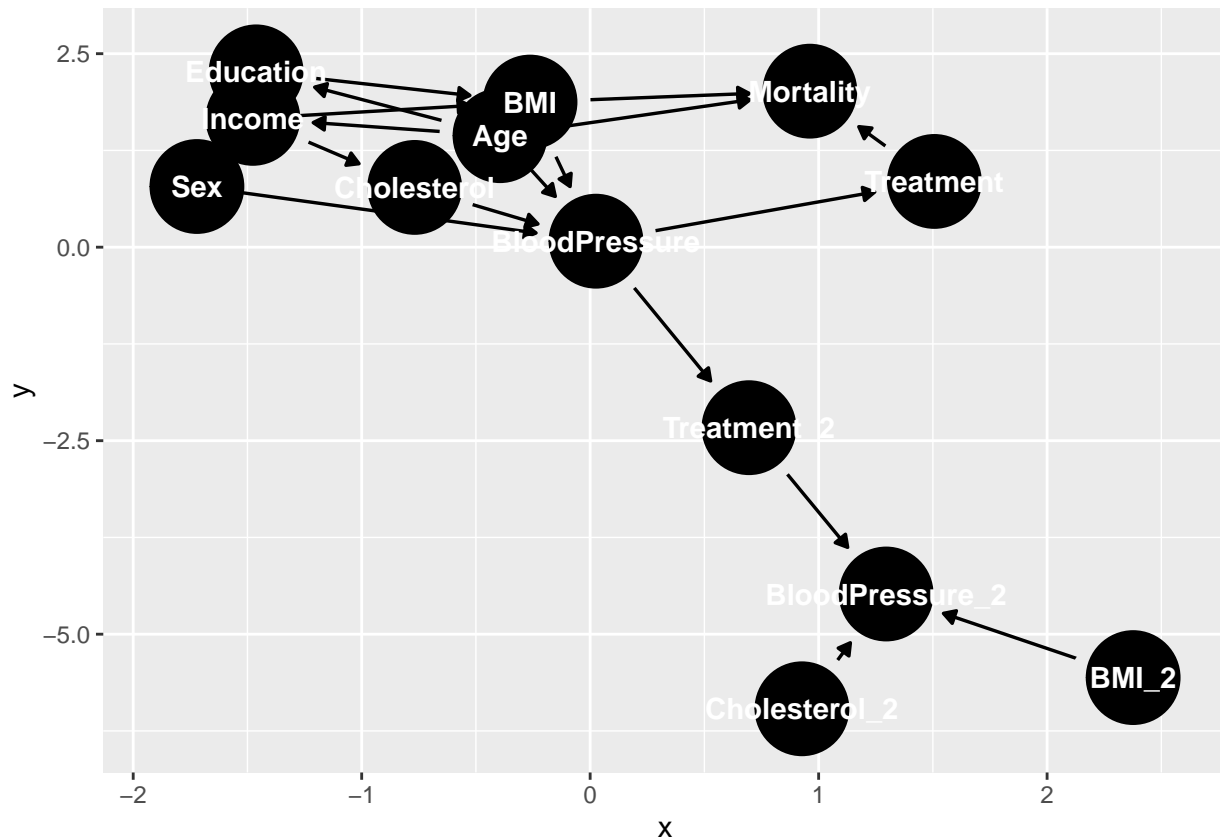


## LTMLE Estimation

Use the `ltmle` package for this section. First fit a "naive model" that **does not** control for the time-dependent confounding. Then run a LTMLE model that does control for any time dependent confounding. Follow the same steps as in the TMLE section. Do you see a difference between the two estimates?

```
## Naive Model (no time-dependent confounding) estimate
# process data
```

```
# ----------
data_obs_ltmle <-
  data_obs %>%
  # need to specify W1, W2, etc
  rename(W1 = age, W2 = sex_at_birth, W3 = simplified_race, W4 = income_thousands, W5 = college_educ, W
  select(W1, W2, W3, A, Y)

# implement ltmle
# ----------
result <- ltmle(data_obs_ltmle, # dataset
                Anodes = "A",   # vector that shows treatment
                Ynodes = "Y",   # vector that shows outcome
                abar = 1)
```

## Qform not specified, using defaults:

## formula for Y:

## Q.kplus1 ~ W1 + W2 + W3 + A

##

## gform not specified, using defaults:

## formula for A:

## A ~ W1 + W2 + W3

##

## Estimate of time to completion: < 1 minute

```
# view
result
```

## Call:
## ltmle(data = data_obs_ltmle, Anodes = "A", Ynodes = "Y", abar = 1)
##
## TMLE Estimate:  0.2485843

```
#with time dependent controls
# simulate data
# ----------
data_obs_2 <-
  heart_disease %>%
  dplyr::mutate(Y = mortality) %>%
  dplyr::rename(A1 = blood_pressure_medication,
        A2 = blood_pressure_medication_2,
        L = income_thousands) %>%
 dplyr:: select(Y, A1, A2, L, sex_at_birth, simplified_race, college_educ, bmi,blood_pressure,chol)
# implement ltmle
# ----------
ltmle(data_obs_2,
      Anodes=c("A1", "A2"),  # two treatment variables
      Lnodes="L",            # L indicator
      Ynodes="Y",            # outcome
      abar=c(1, 1),          # treatment indicator in Anodes vector
      SL.library = sl_libs)
```

Unfortunately I couldn't figure out this second LTMLE part but I think I understand that basically it integrates (using the "L" argument) time dependent variables like income and age which could be correlated with both the treatment and the outcome of interest. I just couldn't figure out exactly how to fix the function...

## Discussion Questions

1. What sorts of time-dependent confounding should we be especially worried about? For instance, would we be concerned about a running variable for age the same way we might be concerned about blood pressure measured at two different times?

2. Answer: I should be concerned about time-dependent confounding that affects both the treatment and the outcome simultaneously and at different levels depending on the time period. The concern about the age variable is that it affects the likelihood of a person taking the blood pressure medication and dying over time, creating bias. For example, younger folks may just be less likely to take blood pressure medication in general due to their relative health, and that youth related health could also mean they are less likely to die with or without the treatment. There is less of a similar type of concern with measuring blood pressure at two different times because those observations are measuring the same phenomenon and does not necessarily confound selection into treatment and the outcome simultaneously.