AE 09: Outliers

Baseball game times

Driver: _____, Reporter: _____, Gopher: _____

> **!** Important
>
> - Open RStudio and create a subfolder in your AE folder called "AE-09".
> - Go to the Canvas and locate your `AE-09` assignment to get started.
> - Upload the `ae-09.qmd` file into the folder you just created. The `.qmd` and PDF responses are due in Canvas. You can check the due date on the Canvas assignment.

```
library(tidyverse)
library(ggformula)
library(yardstick)
library(Stat2Data)
library(mosaic)
library(broom)
library(knitr)
library(patchwork) #arrange plots in a grid
```

Data

The data set for this assignment is from the `Stat2Data` R package which is the companion package for this course's textbook. The data contains data from all MLB games played on August, 11, 2017. On this day there were no extra-innings games or rain delays. You can find information here by searching for the *Baseball Game TImes of One Day in 2017* dataset.

```
data("BaseballTimes2017") # Loads the data from the package
```

We are interested in predicting `Time` the time in minutes to play the game, from either `Runs`, the number of runs scores by the two teams combined, or `Pitchers`, the number of pitchers sued total for the two teams.

Exercise 1

Argue that `Runs` is a better predictor of `Time` than `Pitchers`.

Exercise 2

Argue whether you think the CIN-MIL game would be considered a high leverage and/or high influence point.

Exercise 3

Remove the CIN-MIL game from the data set. Which model is better now?

To submit the AE:

> **!** Important
>
> - Render the document to produce the PDF with all of your work from today's class.
> - Upload your QMD and PDF files to the Canvas assignment.