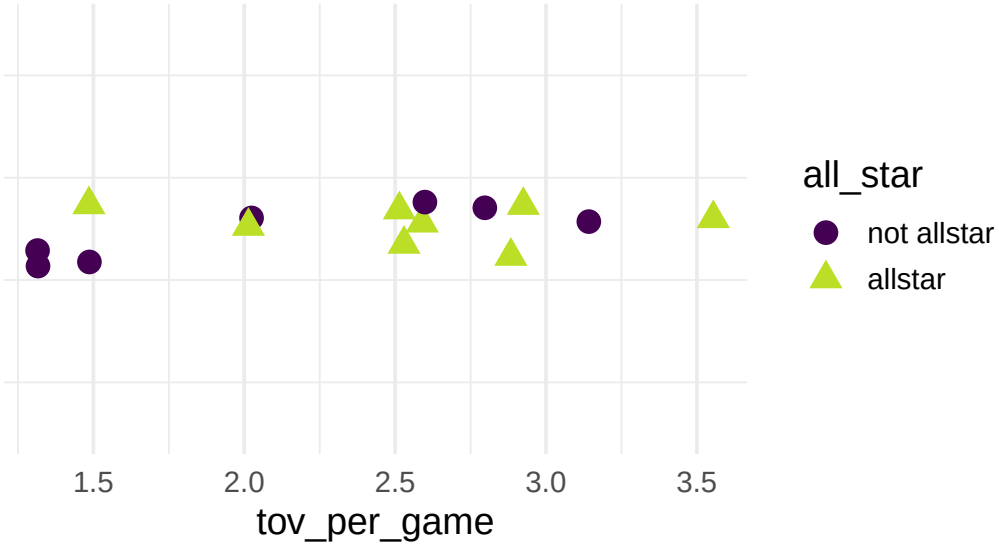
Worksheet on k-Nearest-Neighbors

Part 1 (After First Video on Introducing knn)

In the video, you saw how the k-nearest-neighbors algorithm works with a `pts_per_game` predictor to predict whether or not NBA players made an all-star team.

Here, you will apply the algorithm using a different predictor: turnovers per game (called `tov_per_game`).



Exercise 1. With $k = 1$, predict whether or not the following 10 players make the all-star team. Write in either `allstar` or `not allstar` as a new column called `pred_1` in the printout below (at the turnover value of 1.5, assume that the blue circle has a higher turnover value than the green triangle. Additionally, you might have to make an educated guess about which observation is "closer" for some of the players below, which is just fine :)).

```
# A tibble: 10 x 2
   player             tov_per_game
   <chr>                     <dbl>
 1 Keita Bates-Diop            0.8
 2 Kelly Olynyk                2.5
 3 Kenyon Martin Jr.           1.1
 4 Kevin Durant                3.3
 5 Kevon Looney                0.5
 6 Lauri Markkanen             1.9
 7 LeBron James                3.2
 8 Markelle Fultz              2.3
 9 Mason Plumlee               1.5
10 Michael Porter Jr.          1.1
```

Exercise 2. With $k = 3$, predict whether or not the 10 players above make the all-star team. Write in either `allstar` or `not allstar` as a new column called `pred_3` in the printout above.

Exercise 3. From the graph, it looks like a lot of players with a large amount of turnovers still made the all-star team, even though giving up a turnover is "bad." Come up with a hypothesis for why this might be.

Exercise 4. Recall that there are 15 total NBA players shown on the graph. What would happen if you set $k = 15$, the total number of players shown on the graph? How would the algorithm classify each of the 10 players in the data printout?

Exercise 5. Consider the $k = 1$ setting again. How might you break ties if two players were exactly the same distance away from the player you are trying to classify?

Part 2 (After Second Video on Classification Rates)

Exercise 1. Using the turnovers variable as the predictor with $k = 1$ and your predictions from Part 1, construct a confusion matrix by filling in the table below.

Recall from the slides that, of the 10 players in the printout, only Durant, Markkanen, and James were actually selected as all-stars (the other 7 players were not selected as allstars that year). You might want to make a new column called `all_star` in the printout on the previous page that gives the value `allstar` for those 3 players and `not allstar` for the other 7 players to help with your construction of the confusion matrix).

	Predicted Allstar	Predicted Not
Allstar		
Not		

Exercise 2. From your confusion matrix, calculate the classification rate.

Exercise 3. Using $k = 15$ (the max that k can be), construct a confusion matrix using the turnovers variable as the predictor. **Hint**: Look back at Exercise 4 on the previous page.

	Predicted Allstar	Predicted Not
Allstar		
Not		

Exercise 4. From your confusion matrix, calculate the classifciation rate.

Exercise 5. Which k is better for the turnovers predictor? $k = 1$ or $k = 15$? Why?