Name:_____ Date:_____

LAB 4H: Finding clusters
Response Sheet

Directions: Record your responses to the lab questions in the spaces provided.

Clustering data

The k-means algorithm

(1) Write and run code creating a scatterplot of the players' `ht_inches` and `wt_lbs` and color each dot based on the `league` they play for.

Running k-means

(2) Fill in the blanks below to use k-means to cluster the same height and weight data into two groups:

```
kclusters(_____ ~ _____, data = futbol, k = _____)
```

(3) Use this code and the `mutate` function to add the values from `kclusters` to the `futbol` data. Name the variable `clusters`.

k-means vs. ground-truth

(4) Compare the clusters chosen by k-means to the ground-truth. How successful was k-means at recovering the league information?

On your own

(5) Write and run code creating a scatterplot of `homework` and `videogames` variables.

(6) Based on this graph, identify and remove any outliers by using the `filter` function.

Name:_____ Date:_____

LAB 4H: Finding clusters
Response Sheet

(7) Write and run code using `kclusters` with `k=2` for `homework` and `videogames`.

(8) Describe how the groups differ from each other in terms of how long each group spends playing `videogames` and doing `homework`.