Introduction:

The WNBA worksheet will introduce the idea of predicting playoff teams from their record earlier in the season. Most of the time, it is expected that teams with better records halfway through the season will have a higher chance of making it to the playoffs, but is this always the case? How do you know that the data you are using is a valid way of predicting this? By completing this worksheet, you will be guided through various data cleaning steps and will be able to answer these questions.

Learning Objectives:

* Use various `dplyr` package functions to clean a data set for further use.
* Learn the process of cleaning data, recognizing any problems within the data, and determining whether the data can be used for the given analysis problem.
* Understand why missing data presents a problem for analysis and brainstorm ways to troubleshoot this problem.

Methods:

* Use the `distinct()` and `tally()` functions to determine if there are clear differences between groups.
* Use `select()` and `filter()` functions to clean data.
* Create new datasets using subsets of larger datasets.
* Using the internet to determine any inconsistencies between the given data and real team statistics.