# Name:

# Classwork 13: How does NFL punish domestic violence?

1. Do you remember that 2014 video of a football player named Ray Rice who punched his then girlfriend (now wife) Janay Palmer? He was suspended for 2 games because of the notoriety of that incident. Does that punishment seem to fit the situation?
2. To many people, 2 games seemed low. NFL commissioner Roger Goodell said at a press conference (shortly after this incident) that the suspension has to be consistent with other cases. So let’s take a look at other cases of suspension in the NFL. Examine the data frame **nfl\_suspensions**. Circle the outcome we are interested in:

**nfl\_suspensions** is a data frame with 269 rows representing National Football League players and 7 variables:

Variables:

**name** first initial.last name

**team** team at time of suspension

**games** number of games suspended (one regular season = 16 games)

**category** personal conduct, substance abuse, performance enhancing drugs (PED), or in-game violence

**description** description of suspension

**year** year of suspension

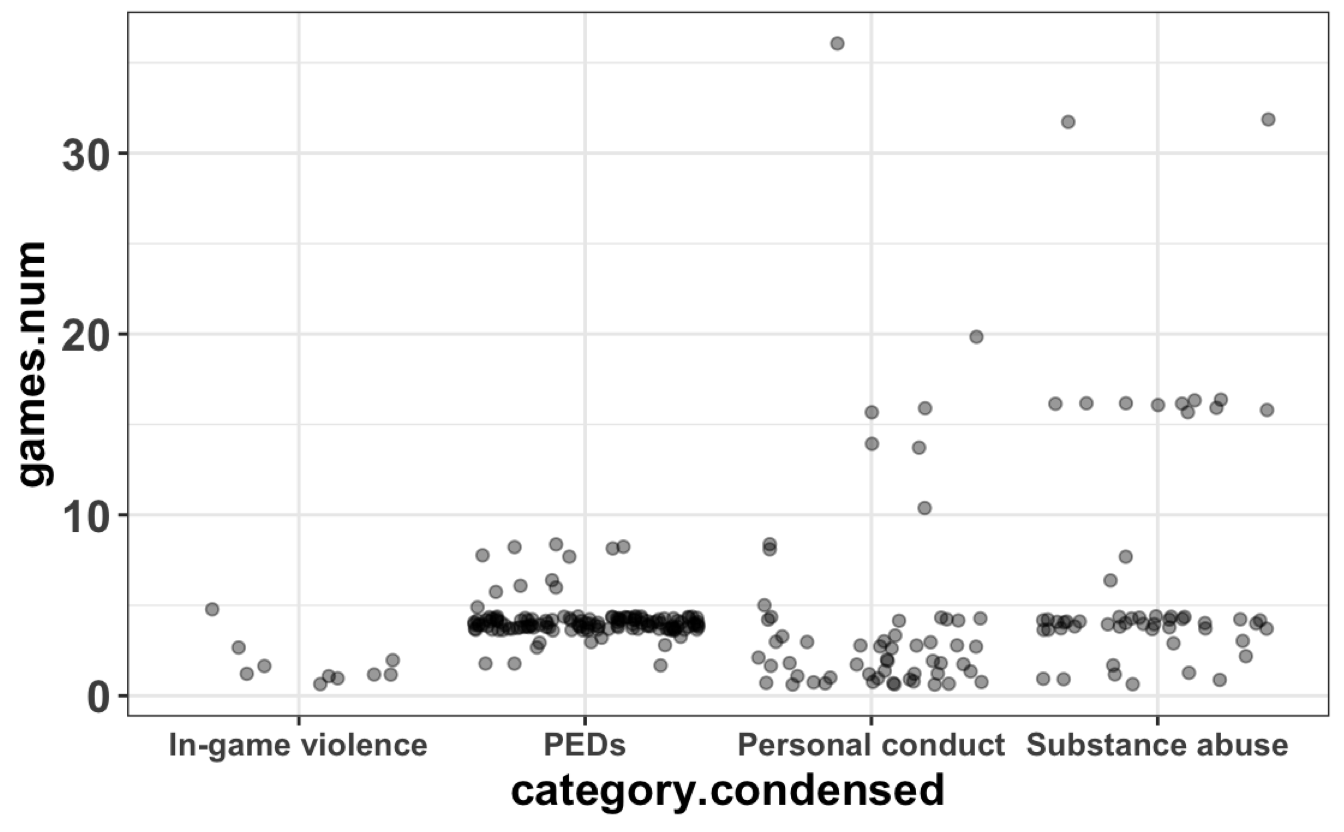
**source** news source

1. What does the distribution of game suspensions look like? Why can’t we see this variable in a histogram? What can we do about that?
2. Let’s look at the distribution of games suspended. Describe the distribution: shape, center, spread. Does these mean that most football players will be suspended about 3 or 4 games? Why or why not?
3. There seem to be some players that were suspended over 30 games. What did they do?
4. Take a look at the different categories of wrongdoing. In your opinion, do you think some categories of wrongdoing *deserves* more game suspensions than others? If so, which ones should outrank which ones? (Note: Depending on our values, we may disagree.)

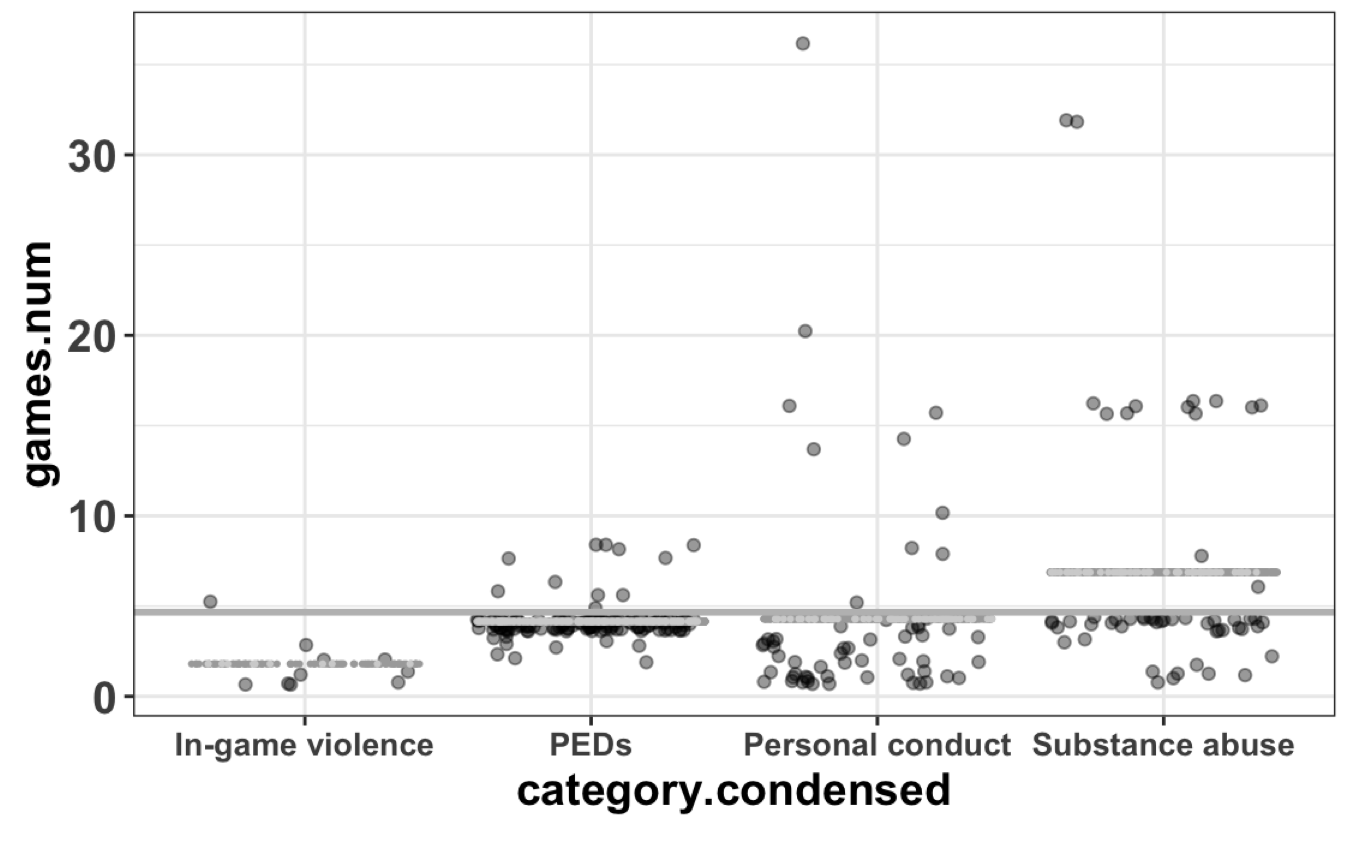
1. There are a lot of categories of wrong doing. Are there any that we can combine in order to create fewer categories? How would we do that in R?

1. Do you think by knowing what a play did wrong, we can explain some of the variation in their game suspension? What does it mean to “explain variation”?
2. If we didn’t know anything about the player or about the wrongdoing, we might simply guess that they would be suspended an average amount. Create an empty model for game suspension and write it in GLM notation.
3. Here is a jitter plot made with this R code (below). Draw the empty model and a few residuals here.

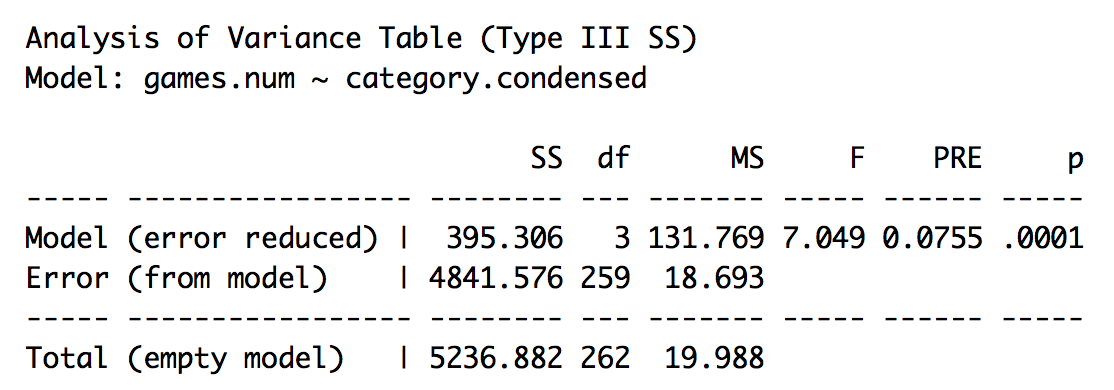
gf\_jitter(games.num ~ category2, data = nflsuspend, alpha = .4)



1. Now find the best fitting estimates for the category model (use the more condensed version of category). Write the GLM for your best model. What do the numbers mean?
2. Which model (category model versus empty model) do you think explains more of the variation seen in game suspension? Why?
3. What would the category model predict as the number of games suspended for someone who committed domestic violence (that would fall under “Personal conduct”)?
4. What would the empty model predict for someone who committed domestic violence?
5. Here we have depicted the best fitting category model (using **category.condensed** to predict **games.num**) to the scatterplot as well as the empty model. Highlight and label each below. Then draw in three different types of deviations using three different colors:
   1. Residuals from the simple model: is this variation explained or unexplained?
   2. Residuals from the complex model: is this variation explained or unexplained?
   3. Variation explained by the complex model



1. If we square and add up these deviations, which one would we give us SS Total? SS Model? SS Error?
2. Some players’ game suspensions are closer to the empty model than the category model. But if we run supernova, the complex model has a smaller leftover error than the empty model. Why?



1. What is MS? What does it stand for and what does it mean? Why are there three of them?
2. What does the PRE mean?
3. What does the F mean?
4. Why is the degrees of freedom for this category model 3? What do the other degrees of freedom (262 and 259) mean?