R Course: Lesson 3

Paolo Cavatore

19 giugno 2016

```
rm(list=ls())
load("rcourse_lesson3_environment.RData")
```

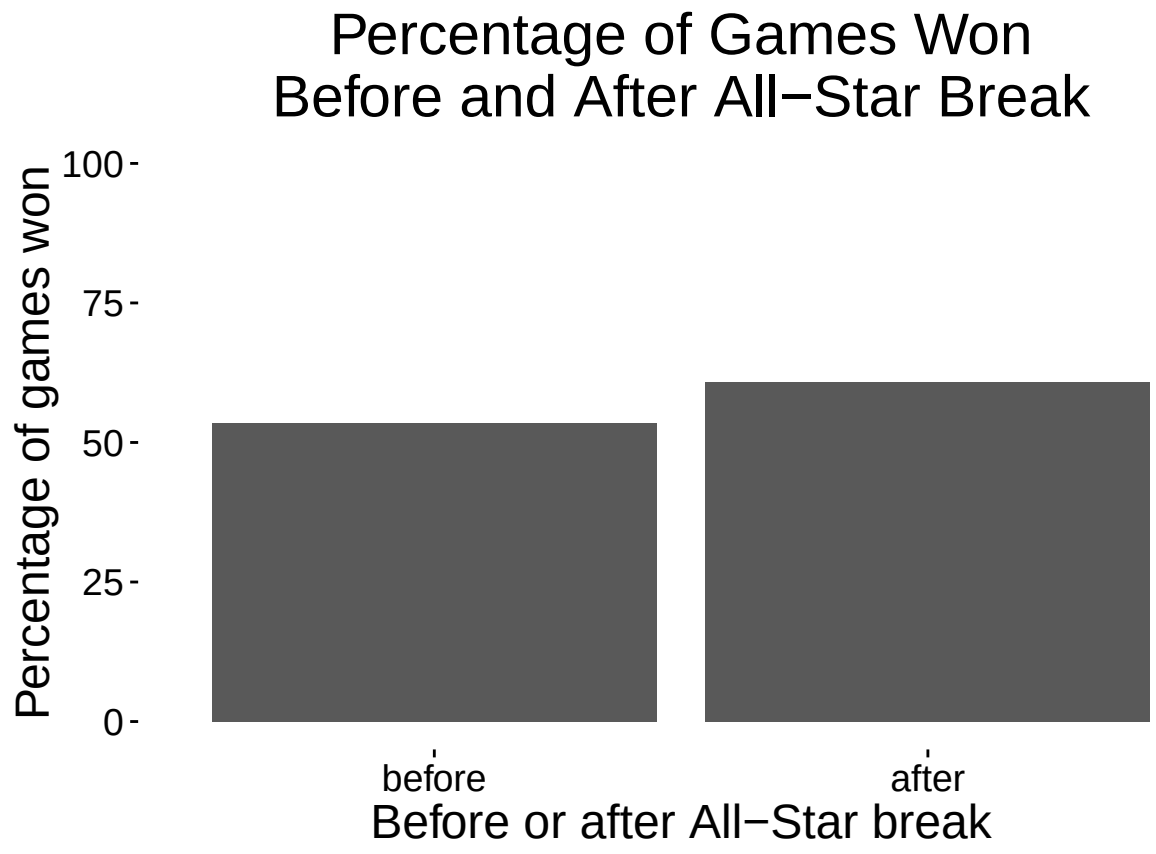
Introduction

I analyzed the Giants' 2010 World Series winning season to see what could significantly predict games they won. I looked at both full season data (all 162 games) and games specific to when Buster Posey was playing.

Results

Full Season Data

For the full season data I tested for an effect of whether the Giants had more wins after the All-Star break or before the All-Star break. Initial visual examination of the data suggests that numerically they won a higher percentage of games after the All-Star break, but the effect looks very small.



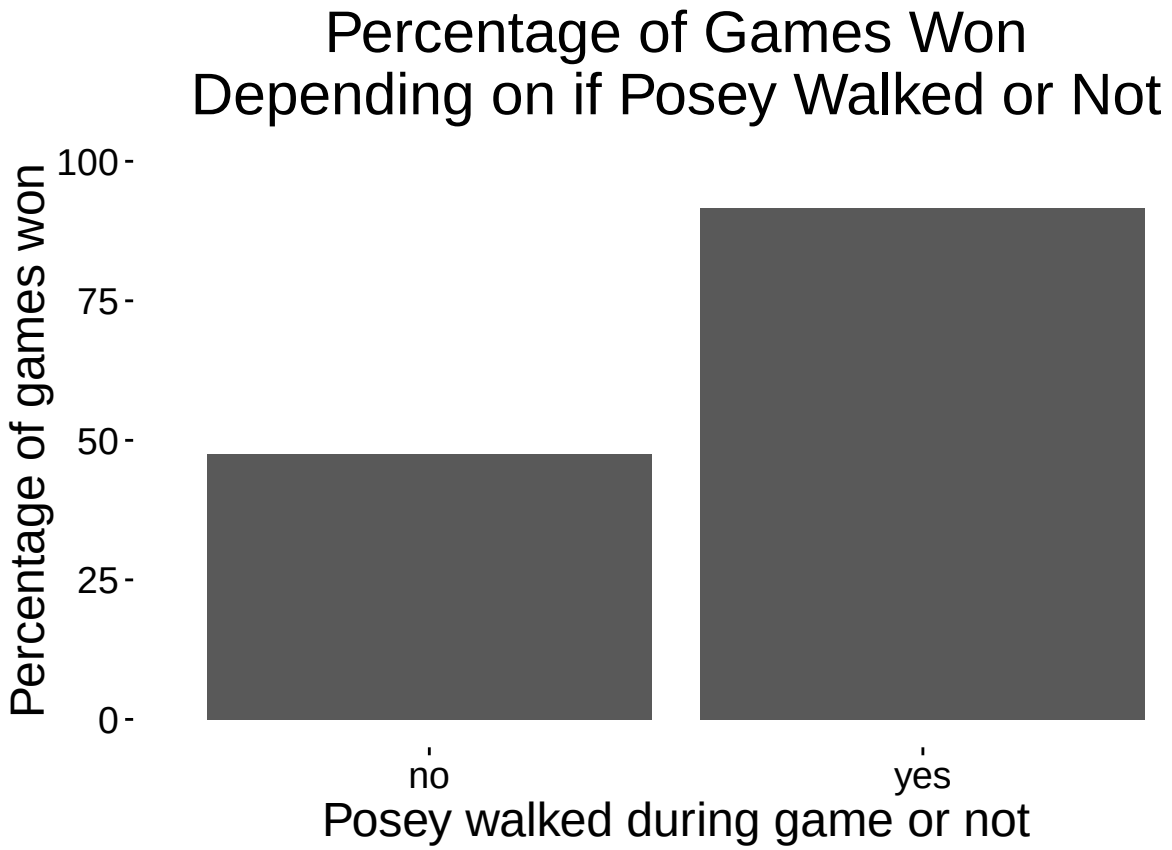
To test this effect I ran a logistic regression with win or loss as the dependent variable and before or after the All-Star break as the independent variable. There was no significant effect of the All-Star break.

```
allstar.glm_sum
```

```
##
## Call:
## glm(formula = win ~ allstar_break, family = "binomial", data = data_stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3688  -1.2359   0.9974   1.1200   1.1200
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)           0.4394     0.2381   1.845    0.065 .
## allstar_breakbefore  -0.3028     0.3200  -0.946    0.344
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 221.58  on 161  degrees of freedom
## Residual deviance: 220.68  on 160  degrees of freedom
## AIC: 224.68
##
## Number of Fisher Scoring iterations: 4
```

Buster Posey Data

The second analysis focused on games where Buster Posey played. I wanted to see if whether Posey was walked or not had any effect on if the Giants won or lost the game. Visual inspection of the data suggests the Giants were much more likely to win if Posey had been walked at least once in the game.



To test this effect I ran a logistic regression with win or loss as the dependent variable and Posey walked or not as the independent variable. There was a significant effect of Posey being walked, with the Giants being more likely to win if Posey had been walked in games where he had played.

```
posey_walked.glm_sum
```

```
##
## Call:
## glm(formula = win ~ walked, family = "binomial", data = data_posey_stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2293  -1.1372   0.4172   1.2181   1.2181
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.09531    0.21847  -0.436  0.66264
## walkedyes    2.49321    0.77018   3.237  0.00121 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 147.34  on 107  degrees of freedom
## Residual deviance: 130.03  on 106  degrees of freedom
## AIC: 134.03
##
```

```
## Number of Fisher Scoring iterations: 5
```

Conclusion

The All-Star break had no significant effect on the Giants' road to the 2010 World Series, but Posey being walked did.